aws

Guia do desenvolvedor

Amazon Kinesis Data Streams

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon Kinesis Data Streams: Guia do desenvolvedor

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

As marcas comerciais e imagens de marcas da Amazon não podem ser usadas no contexto de nenhum produto ou serviço que não seja da Amazon, nem de qualquer maneira que possa gerar confusão entre os clientes ou que deprecie ou desprestigie a Amazon. Todas as outras marcas comerciais que não pertencem à Amazon pertencem a seus respectivos proprietários, que podem ou não ser afiliados, patrocinados pela Amazon ou ter conexão com ela.

Table of Contents

O que é o Amazon Kinesis Data Streams? 1
O que posso fazer com o Kinesis Data Streams? 1
Benefícios do uso do Kinesis Data Streams 2
Serviços relacionados 3
Terminologia e conceitos 4
Analisar a arquitetura de alto nível do Kinesis Data Streams 4
Familiarização com a terminologia do Kinesis Data Streams 5
Fluxo de dados do Kinesis 5
Registro de dados 5
Modo de capacidade 5
Período de retenção 5
Produtor 6
Consumidor 6
Aplicativo do Amazon Kinesis Data Streams 6
Fragmento 6
Chave de partição 7
Número de sequência 7
Kinesis Client Library 7
Nome da aplicação 8
Criptografia do lado do servidor 8
Cotas e limites 9
Limites da API 12
Limites de API do plano de controle do KDS 12
Limites de API do plano de dados do KDS 18
Aumento de cotas 21
Pré-requisitos completos para configurar o Amazon Kinesis Data Streams 22
Inscreva-se para AWS 22
Fazer download de bibliotecas e ferramentas 22
Configurar seu ambiente de desenvolvimento 23
Use o AWS CLI para realizar operações do Amazon Kinesis Data Streams 24
Tutorial: Instalar e configurar o AWS CLI para Kinesis Data Streams 25
Instale o AWS CLI 25
Configurar o AWS CLI 26
Tutorial: Execute operações básicas do Kinesis Data Streams usando o AWS CLI 26

Etapa 1: criar um fluxo ... 27
Etapa 2: colocar um registro ... 28
Etapa 3: obter o registro ... 29
Etapa 4: limpar ... 32
Tutoriais de conceitos básicos ... 34
Tutorial: Processar dados de ações em tempo real usando a KCL e a KCL 2.x ... 34
Concluir os pré-requisitos ... 35
Criar um fluxo de dados ... 36
Criar um usuário e uma política do IAM ... 37
Fazer download e criar o código ... 42
Implementar o produtor ... 43
Implementar o consumidor ... 48
(Opcional) Estender o consumidor ... 52
Limpar os recursos ... 54
Tutorial: Processar dados de ações em tempo real usando a KPL e a KCL 1.x ... 55
Concluir os pré-requisitos ... 56
Criar um fluxo de dados ... 57
Criar um usuário e uma política do IAM ... 59
Fazer download e compilação do código de implementação ... 64
Implementar o produtor ... 65
Implementar o consumidor ... 70
(Opcional) Estender o consumidor ... 74
Limpar os recursos ... 75
Tutorial: analise dados do mercado de ações em tempo real usando o Amazon Managed Service for Apache Flink ... 77
Pré-requisitos ... 78
Etapa 1: Configurar uma conta da ... 78
Etapa 2: configurar o AWS CLI ... 82
Etapa 3: criar um aplicativo ... 83
Tutorial: Use AWS Lambda com o Amazon Kinesis Data Streams ... 100
Use a solução AWS de streaming de dados para o Amazon Kinesis ... 101
Criar e gerenciar fluxos de dados do Kinesis ... 102
Escolha o modo certo para transmitir ... 103
Quais são os diferentes modos no Kinesis Data Streams? ... 103
Atributos e casos de uso do modo On-demand Standard ... 104
Atributos e casos de uso do modo On-demand Advantage ... 106

Casos de uso e atributos do modo provisionado 106
Alternar entre os modos 108
Crie um stream usando o Console de gerenciamento da AWS 108
Crie um stream usando o APIs 109
Criar o cliente do Kinesis Data Streams 109
Criar o fluxo 110
Atualizar um fluxo 111
Usar o console 112
Usar a API 113
Use o AWS CLI 113
Listar fluxos 113
Listar fragmentos 115
Excluir um fluxo 118
Refragmentar um fluxo 118
Decidir uma estratégia para refragmentar 119
Dividir um fragmento 120
Mesclar dois fragmentos 122
Concluir a ação de refragmentação 123
Alterar o período de retenção de dados 126
Marcar seus recursos com tag 127
Revisar conceitos básicos de tags 128
Monitorar custos usando tags 129
Compreender as restrições de tags 129
Atribuir tags a fluxos usando o console do Kinesis Data Streams 130
Marque fluxos usando o AWS CLI 132
Marque streams usando o Kinesis Data Streams APIs 132
Marque os consumidores usando o AWS CLI 133
Marque consumidores usando o Kinesis Data Streams APIs 133
Gerenciar registros grandes 134
Atualizar seu fluxo para usar registros grandes 134
Otimizar o desempenho do seu fluxo com registros grandes 135
Reduzir o controle de utilização com registros grandes 135
Gerencie registros grandes usando o Kinesis Data Streams APIs 135
AWS componentes compatíveis com registros grandes 136
Regiões em que grandes registros são compatíveis 140
Realize testes de resiliência com AWS Fault Injection Service 142

Erros de exceção de throughput provisionada 143
Erros de exceção do iterador expirado 146
Gravar dados no Kinesis Data Streams 148
Desenvolver produtores usando a Amazon Kinesis Producer Library (KPL) 149
Analisar a função da KPL 150
Perceber as vantagens de usar a KPL 150
Entender quando não usar a KPL 152
Instalar a KPL 152
Migrar para a KPL 1.x 153
Transição para certificados Amazon Trust Services (ATS) para a KPL 157
Plataformas compatíveis com a KPL 157
Principais conceitos da KPL 158
Integrar a KPL com o código de produtor 160
Gravar no fluxo de dados do Kinesis usando a KPL 163
Configurar a KPL 165
Implementar a desagregação de consumidores 166
Usar a KPL com o Amazon Data Firehose 169
Use o KPL com o Registro do AWS Glue Esquema 170
Definir a configuração do proxy da KPL 170
Política de ciclo de vida da versão da KPL 171
Desenvolva produtores usando a API Kinesis Data Streams com o AWS SDK para Java 172
Adicionar dados a um fluxo 173
Interaja com os dados usando o AWS Glue Schema Registry 179
Gravar no Amazon Kinesis Data Streams usando o Kinesis Agent 180
Concluir os pré-requisitos do Kinesis Agent 181
Fazer download e instalar o agente 181
Configuração e inicialização do agente 182
Especificar as definições da configuração do agente 183
Monitorar vários diretórios de arquivos e gravação em vários fluxos 186
Uso do agente para pré-processar dados 187
Usar comandos da CLI do agente 192
Perguntas frequentes 192
Grave no Kinesis Data Streams usando outros serviços AWS 194
Grave no Kinesis Data Streams usando AWS Amplify 194
Gravar no Kinesis Data Streams usando o Amazon Aurora 195
Escreva para o Kinesis Data Streams usando a Amazon CloudFront 195

Grave no Kinesis Data Streams CloudWatch usando o Amazon Logs 195
Gravar no Kinesis Data Streams usando o Amazon Connect 195
Grave no Kinesis Data Streams usando AWS Database Migration Service 196
Gravar no Kinesis Data Streams usando o Amazon DynamoDB 196
Escreva para o Kinesis Data Streams usando a Amazon EventBridge 196
Grave no Kinesis Data Streams usando AWS IoT Core 197
Gravar no Kinesis Data Streams usando o Amazon Relational Database Service 197
Gravar no Kinesis Data Streams usando o Amazon Pinpoint 197
Gravar no Kinesis Data Streams usando o Amazon Quantum Ledger Database (Amazon QLDB) 198
Gravar no Kinesis Data Streams usando integrações de terceiros 198
Apache Flink 198
Fluentd 199
Debezium 199
Oráculo GoldenGate 199
Kafka Connect 199
Adobe Experience 199
Striim 199
Solução de problemas de produtores do Kinesis Data Streams 200
Meu aplicativo produtor está gravando a uma taxa menor que a esperada 200
Eu recebo um erro de permissão de chave mestra do KMS não autorizada 202
Solucionar outros problemas comuns para produtores 202
Otimizar os produtores do Kinesis Data Streams 203
Personalize o comportamento de limites de taxas e novas tentativas da KPL 203
Aplique as práticas recomendadas à agregação da KPL 204
Leitura de dados do Kinesis Data Streams 206
Desenvolver consumidores de distribuição avançada com throughput dedicado 207
Diferenças entre o consumidor de throughput compartilhado e o consumidor de distribuição avançada 209
Regiões com suporte para até 50 consumidores de fan-out aprimorados (somente sob demanda Advantage) 140
Gerenciar consumidores de distribuição avançada 212
Usar o visualizador de dados no console do Kinesis 213
Consultar seus fluxos de dados no console do Kinesis 214
Usar a Kinesis Client Library 215
O que é a Kinesis Client Library? 215

Principais atributos e benefícios da KCL ... 216
Conceitos da KCL ... 217
Tabelas de metadados do DynamoDB e balanceamento de carga na KCL ... 218
Desenvolver consumidores com a KCL ... 223
Processamento de vários fluxos com a KCL ... 234
Use o registro do AWS Glue esquema com o KCL ... 237
Permissões do IAM necessárias para aplicativos de consumo da KCL ... 238
Configurações da KCL ... 244
Política de ciclo de vida da versão da KCL ... 261
Migrar de versões anteriores da KCL ... 262
Documentação da versão anterior da KCL ... 278
Desenvolva consumidores com o AWS SDK para Java ... 364
Desenvolva consumidores com produtividade compartilhada com o AWS SDK para Java ... 365
Desenvolva consumidores expandidos aprimorados com o AWS SDK para Java ... 371
Interaja com os dados usando o AWS Glue Schema Registry ... 373
Desenvolver consumidores usando o AWS Lambda ... 374
Desenvolver consumidores usando o Managed Service for Apache Flink ... 374
Desenvolver consumidores usando o Amazon Data Firehose ... 374
Use outros AWS serviços para ler dados do Kinesis Data Streams ... 375
Leitura de dados do Kinesis Data Streams usando Amazon EMR ... 375
Leia dados do Kinesis Data Streams EventBridge usando o Amazon Pipes ... 375
Leia dados do Kinesis Data Streams usando AWS Glue ... 376
Leitura de dados do Kinesis Data Streams usando Amazon Redshift ... 376
Leitura a partir do Kinesis Data Streams usando integrações de terceiros ... 376
Apache Flink ... 377
Adobe Experience Platform ... 377
Apache Druid ... 377
Apache Spark ... 377
Databricks ... 378
Kafka Confluent Platform ... 378
Kinesumer ... 378
Talend ... 378
Solução de problemas de consumidores do Kinesis Data Streams ... 378
Erro de compilação com o construtor LeaseManagementConfig ... 379
Alguns registros do Kinesis Data Streams são ignorados quando a Kinesis Client Library é usada ... 380

Registros pertencentes ao mesmo fragmento são processados por processadores de registros diferentes ao mesmo tempo 380
O aplicativo consumidor está lendo a uma taxa menor que a esperada 381
GetRecords retorna uma matriz de registros vazia mesmo quando há dados no fluxo 382
Iterador do fragmento expira inesperadamente 383
Processamento de registros de consumidores ficando atrasados 383
Erro de permissão de chave do KMS não autorizada 384
DynamoDbException: o caminho do documento fornecido na expressão de atualização é inválido para atualização 384
Solucionar outros problemas comuns para consumidores 385
Otimizar consumidores do Kinesis Data Streams 385
Melhorar o processamento de baixa latência 386
Processe dados serializados usando AWS Lambda com a Amazon Kinesis Producer Library 387
Use refragmentação, escalonamento e processamento paralelo para alterar o número de fragmentos 387
Tratamento de registros duplicados 389
Tratamento do startup, do desligamento e do controle de utilização 391
Monitorar o Kinesis Data Streams 394
Monitore o serviço Kinesis Data Streams com CloudWatch 394
Métricas e dimensões do Amazon Kinesis Data Streams 395
Acesse CloudWatch as métricas da Amazon para Kinesis Data Streams 413
Monitore a integridade do Kinesis Data Streams Agent com CloudWatch 414
Monitor com CloudWatch 414
Registre as chamadas de API do Amazon Kinesis Data Streams com AWS CloudTrail 415
Informações do Kinesis Data Streams em CloudTrail 415
Exemplo: entradas do arquivo de log do Kinesis Data Streams 417
Monitore o KCL com CloudWatch 421
Métricas e namespace 421
Dimensões e níveis de métricas 421
Configuração da métrica 423
Lista de métricas 423
Monitore o KPL com CloudWatch 441
Métricas, dimensões e namespaces 442
Granularidade e nível de métrica 442
Acesso local e CloudWatch upload da Amazon 443

Lista de métricas ... 444
Segurança ... 448
Proteção de dados no Kinesis Data Streams ... 449
O que é criptografia no lado do servidor para o Kinesis Data Streams? ... 449
Custos, regiões e considerações sobre desempenho ... 451
Como começo a usar a criptografia no lado do servidor? ... 452
Criar e usar chaves do KMS geradas pelo usuário ... 453
Permissões para usar as chaves do KMS geradas pelo usuário ... 454
Verificar e solucionar problemas de permissões de chaves do KMS ... 456
Usar o Kinesis Data Streams com endpoints da VPC de interface ... 456
Controlar o acesso aos recursos do Kinesis Data Streams usando o IAM ... 460
Sintaxe da política ... 461
Ações para o Kinesis Data Streams ... 462
Nomes de recursos da Amazon (ARNs) para Kinesis Data Streams ... 462
Exemplo de políticas para o Kinesis Data Streams ... 463
Compartilhamento do fluxo de dados com outra conta ... 466
Configurar uma AWS Lambda função para ler do Kinesis Data Streams em outra conta ... 472
Compartilhamento de acesso usando políticas baseadas em recursos ... 472
Validação de conformidade para o Kinesis Data Streams ... 474
Resiliência no Kinesis Data Streams ... 475
Recuperação de desastres no Kinesis Data Streams ... 476
Segurança de infraestrutura no Kinesis Data Streams ... 477
Práticas recomendadas de segurança para o Kinesis Data Streams ... 477
Implemente o acesso de privilégio mínimo ... 478
Usar funções do IAM ... 478
Implementação da criptografia do lado do servidor em recursos dependentes ... 478
Use CloudTrail para monitorar chamadas de API ... 478
Trabalhando com AWS SDKs ... 480
Exemplos de código ... 482
Conceitos básicos ... 483
Conheça os conceitos básicos ... 483
Ações ... 487
Exemplos sem servidor ... 546
Invocar uma função do Lambda em um trigger do Kinesis ... 546
Relatando falhas de itens em lote para funções do Lambda com um trigger do Kinesis ... 557
Histórico do documento ... 571

..... dlxxv

O que é o Amazon Kinesis Data Streams?

É possível usar o Amazon Kinesis Data Streams para coletar e processar grandes fluxos de registros de dados em tempo real. É possível criar aplicações de processamento de dados, conhecidas como aplicações do Kinesis Data Streams. Uma aplicação típica do Kinesis Data Streams lê dados de um fluxo de dados como registros de dados. Esses aplicativos podem usar a Kinesis Client Library e podem ser executados em instâncias da Amazon EC2. É possível enviar os registros processados para painéis, usá-los para gerar alertas, alterar dinamicamente as estratégias de preços e publicidade ou enviar dados para vários outros serviços da AWS . Para obter informações sobre os recursos e preços do Kinesis Data Streams, consulte Amazon Kinesis Data Streams.

O Kinesis Data Streams é parte da plataforma de dados de fluxo Kinesis, juntamente com o Kinesis Data Firehose, Kinesis Video Streams e o Managed Service for Apache Flink.

Para obter mais informações sobre soluções de AWS big data, consulte Big Data on AWS. Para obter mais informações sobre as soluções de dados de fluxo da AWS , consulte O que são dados em fluxo?

Tópicos

- O que posso fazer com o Kinesis Data Streams?
- Benefícios do uso do Kinesis Data Streams
- Serviços relacionados

O que posso fazer com o Kinesis Data Streams?

É possível usar o Kinesis Data Streams para entrada e agregação de dados de forma rápida e contínua. O tipo de dados usado pode incluir dados de log de infraestrutura de TI, logs de aplicativo, mídias sociais, feeds de dados de mercado e dados de sequência de cliques da web. Como o tempo de resposta para a entrada e o processamento de dados é em tempo real, o processamento geralmente é leve.

Estes são alguns cenários típicos de uso do Kinesis Data Streams:

Log acelerado e consumo e processamento de dados

> É possível ter aplicações de produção que gerem dados diretamente em um fluxo. Por exemplo, gere logs de sistemas e aplicativos e eles estarão disponíveis para processamento em segundos.

Isso evitará que os dados de log sejam perdidos se o servidor de front-end ou de aplicações falhar. O Kinesis Data Streams fornece uma entrada acelerada de dados, pois os dados não são organizados em lotes nos servidores antes de enviá-los.

Métricas e relatórios em tempo real

É possível usar dados coletados no Kinesis Data Streams para análise simples de dados e geração de relatórios em tempo real. Por exemplo, seu aplicativo de processamento de dados pode funcionar em métricas e geração de relatórios para logs do sistema e de aplicativos à medida que os dados passam por ele em vez de esperar receber lotes de dados.

Análise de dados em tempo real

Ela combina o poder do processamento paralelo com o valor de dados em tempo real. Por exemplo, processar clickstreams em tempo real e, em seguida, analisar o envolvimento da usabilidade do site usando várias aplicações diferentes do Kinesis Data Streams executadas em paralelo.

Complexo processamento de stream

Você pode criar gráficos acíclicos direcionados (DAGs) de aplicativos e fluxos de dados do Kinesis Data Streams. Normalmente, isso envolve colocar dados de várias aplicações do Kinesis Data Streams em outro fluxo para processamento downstream por outro aplicação desse serviço.

Benefícios do uso do Kinesis Data Streams

Embora seja possível usar o Kinesis Data Streams para resolver vários problemas de dados em fluxo, um uso comum é agregar dados em tempo real e carregar os dados agregados em um data warehouse ou cluster de redução de mapa.

Os dados são colocados em fluxos de dados do Kinesis, o que garante durabilidade e elasticidade. O atraso entre o momento em que um registro é colocado no stream e o tempo em que ele pode ser recuperado (put-to-get atraso) geralmente é inferior a 1 segundo. Em outras palavras, uma aplicação do Kinesis Data Streams pode começar a consumir os dados do fluxo quase que imediatamente após sua adição. Como é um serviço gerenciado, o Kinesis Data Streams cuida da carga operacional de criar e executar um pipeline de entrada de dados. É possível criar aplicações de redução de mapa de fluxo. A elasticidade do Kinesis Data Streams permite escalar o fluxo. Assim, registros de dados nunca são perdidos antes que exirem.

Como várias aplicações do Kinesis Data Streams podem consumir dados de um fluxo, múltiplas ações como arquivamento e processamento podem ocorrer de maneira simultânea e independente.

Por exemplo, dois aplicativos podem ler dados do mesmo fluxo. A primeira aplicação calcula agregados em execução e atualiza uma tabela do Amazon DynamoDB, e a segunda compacta e arquiva dados em um datastore, como o Amazon Simple Storage Service (Amazon S3). A tabela do DynamoDB com agregados em execução é então lida por um painel para relatórios. up-to-the-minute

A Kinesis Client Library permite o consumo de dados tolerante a falhas de fluxos e fornece suporte à escalabilidade para aplicações do Kinesis Data Streams.

Serviços relacionados

Para obter informações sobre como usar clusters do Amazon EMR para ler e processar fluxos de dados diretamente do Kinesis, consulte Conetor do Kinesis.

Terminologia e conceitos do Amazon Kinesis Data Streams

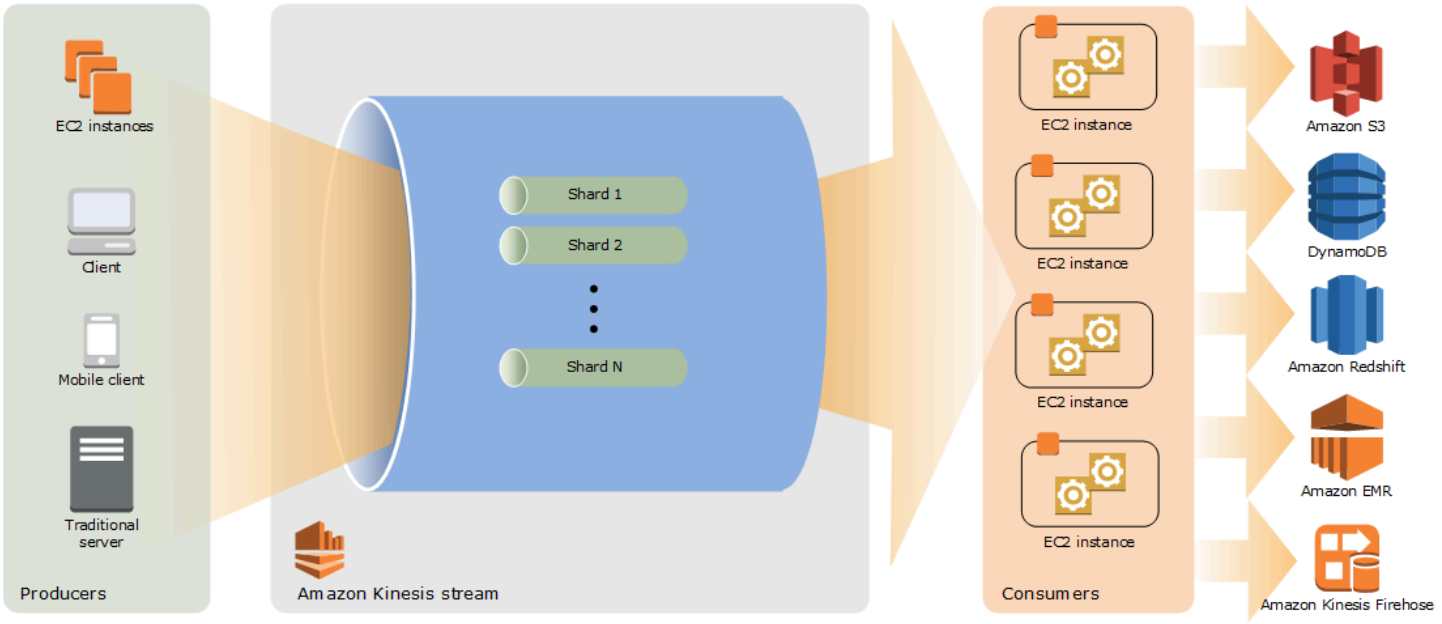
Antes de começar a usar o Amazon Kinesis Data Streams, conheça sua arquitetura e terminologia.

Tópicos

- Analisar a arquitetura de alto nível do Kinesis Data Streams
- Familiarização com a terminologia do Kinesis Data Streams

Analisar a arquitetura de alto nível do Kinesis Data Streams

O diagrama a seguir ilustra a arquitetura de alto nível do Kinesis Data Streams. Os produtores enviam dados por push continuamente ao Kinesis Data Streams, que os consumidores processam em tempo real. Os consumidores (como um aplicativo personalizado executado no Amazon EC2 ou em um stream de entrega do Amazon Data Firehose) podem armazenar seus resultados usando um AWS serviço como Amazon DynamoDB, Amazon Redshift ou Amazon S3.



Familiarização com a terminologia do Kinesis Data Streams

Fluxo de dados do Kinesis

Um fluxo de dados do Kinesis é um conjunto de fragmentos. Cada fragmento tem uma sequência de registros de dados. Cada registro de dados tem um número de sequência atribuído pelo Kinesis Data Streams.

Registro de dados

Um registro de dados é a unidade de dados armazenada em um fluxo de dados do Kinesis. Os registros de dados são compostos de um número de sequência, uma chave de partição e um blob de dados, que é uma sequência de bytes imutável. O Kinesis Data Streams não inspeciona, interpreta nem altera dados no blob. Um blob de dados pode ter até 1 MB.

Modo de capacidade

O modo de capacidade de um fluxo de dados determina como a capacidade é gerenciada e como são geradas cobranças pelo seu uso. No Kinesis Data Streams você atualmente pode escolher entre o modo sob demanda e o modo provisionado para seus fluxos de dados. Para obter mais informações, consulte Escolha o modo certo para transmitir.

No modo sob demanda, o Kinesis Data Streams gerencia automaticamente os fragmentos para fornecer o throughput necessário. Apenas o throughput real usado é cobrado, e o Kinesis Data Streams acomoda automaticamente as necessidades de throughput das cargas de trabalho à medida que elas aumentam ou diminuem. Para obter mais informações, consulte Atributos e casos de uso do modo On-demand Standard.

No modo provisionado, é necessário especificar o número de fragmentos para o fluxo de dados. A capacidade total de um fluxo de dados é a soma das capacidades de seus fragmentos. É possível aumentar ou diminuir o número de fragmentos em um fluxo de dados conforme necessário e recebe uma cobrança pelo número de fragmentos a uma taxa horária. Para obter mais informações, consulte Casos de uso e atributos do modo provisionado.

Período de retenção

O período de retenção é o tempo em que os registros de dados permanecem acessíveis depois de serem adicionados ao fluxo. O período de retenção de um fluxo é definido para um padrão de 24

horas após a criação. Você pode aumentar o período de retenção em até 8760 horas (365 dias) usando a IncreaseStreamRetentionPeriodoperação e diminuir o período de retenção para um mínimo de 24 horas usando a DecreaseStreamRetentionPeriodoperação. Encargos adicionais incidem sobre streams com período de retenção definido acima de 24 horas. Para obter mais informações, consulte Definição de preço do Amazon Kinesis Data Streams.

Produtor

Os produtores colocam registros no Amazon Kinesis Data Streams. Por exemplo, um servidor web que envia dados de log para um fluxo é um produtor.

Consumidor

Os consumidores obtêm registros do Amazon Kinesis Data Streams e os processam. Esses consumidores são conhecidos como Aplicativo do Amazon Kinesis Data Streams.

Aplicativo do Amazon Kinesis Data Streams

Uma aplicação do Amazon Kinesis Data Streams é um consumidor de fluxo normalmente executado em uma frota de instâncias do EC2.

Há dois tipos de consumidores que podem ser desenvolvidos: consumidores avançados compartilhados e consumidores avançados aprimorados. Para saber mais sobre as diferenças entre eles, e para ver como é possível criar cada tipo de consumidor, consulte Leitura de dados do Amazon Kinesis Data Streams.

A saída de uma aplicação do Kinesis Data Streams pode ser a entrada de outro fluxo, permitindo a criação de topologias complexas que processam dados em tempo real. Um aplicativo também pode enviar dados para vários outros AWS serviços. Pode haver vários aplicativos para um fluxo, e cada aplicativo pode consumir dados do fluxo de forma independente e simultaneamente.

Fragmento

Um fragmento é uma sequência de registros de dados identificada de forma exclusiva em um fluxo. Um fluxo é composto de um ou mais fragmentos, sendo que cada um deles fornece uma unidade fixa de capacidade. Cada fragmento é compatível com até 5 transações por segundo para leituras, até a taxa máxima total de leitura de dados de 2 MB por segundo, e até 1.000 registros por segundo para gravações, até a taxa máxima total de gravação de dados de 1 MB por segundo (incluindo chaves de

partição). A capacidade de dados do seu fluxo é uma função do número de fragmentos especificados para o fluxo. A capacidade total do fluxo é a soma das capacidades de seus fragmentos.

Se a taxa de dados aumenta, é possível aumentar ou diminuir o número de fragmentos alocados para seu fluxo. Para obter mais informações, consulte Refragmentar um fluxo.

Chave de partição

A chave de partição é usada para agrupar os dados por fragmento dentro de um fluxo. O Kinesis Data Streams segrega os registros de dados pertencentes a um fluxo em vários fragmentos. Ele usa a chave de partição associada a cada registro de dados para determinar a qual fragmento um determinado registro de dados pertence. As chaves de partição são strings Unicode, com um limite de tamanho máximo de 256 caracteres para cada chave. Uma função MD5 hash é usada para mapear chaves de partição para valores inteiros de 128 bits e para mapear registros de dados associados a fragmentos usando os intervalos de chaves de hash dos fragmentos. Quando um aplicativo insere dados em um fluxo, ele deve especificar uma chave de partição.

Número de sequência

Cada registro de dados tem um número de sequência exclusivo por chave de partição dentro do fragmento. O Kinesis Data Streams atribuirá o número de sequência depois que um registro é gravado no fluxo com `client.putRecords` ou `client.putRecord`. Geralmente, os números de sequência da mesma chave de partição aumentam ao longo do tempo. Quanto maior for o período entre as solicitações de gravação, maiores serão os números de sequência.

> **Note**
>
> Os números de sequência não podem ser usados como índices para conjuntos de dados dentro do mesmo fluxo. Para separar logicamente conjuntos de dados, use chaves de partição ou crie um fluxo separado para cada conjunto de dados.

Kinesis Client Library

A Kinesis Client Library é compilada em sua aplicação para permitir o consumo de dados tolerante a falhas do fluxo. A Kinesis Client Library garante que para cada fragmento haja um processador de registros em execução e processando o fragmento. A biblioteca também simplifica a leitura de dados do fluxo. A Kinesis Client Library usa tabelas do Amazon DynamoDB para armazenar

metadados relacionados ao consumo de dados. Ela cria três tabelas para cada aplicação que esteja processando dados. Para obter mais informações, consulte Usar a Kinesis Client Library.

Nome da aplicação

O nome identifica uma aplicação do Amazon Kinesis Data Streams. Cada um dos seus aplicativos deve ter um nome exclusivo que tenha como escopo a AWS conta e a região usadas pelo aplicativo. Esse nome é usado como um nome para a tabela de controle no Amazon DynamoDB e o namespace para as métricas da Amazon. CloudWatch

Criptografia do lado do servidor

O Amazon Kinesis Data Streams pode criptografar automaticamente dados confidenciais à medida que um produtor os insere em um fluxo. O Kinesis Data Streams usa chaves mestras do AWS KMS para criptografia. Para obter mais informações, consulte Proteção de dados no Amazon Kinesis Data Streams.

> **Note**
>
> Para ler ou gravar um em um fluxo criptografado, aplicativos produtores e consumidores devem ter permissão para acessar a chave mestra. Para obter informações sobre a concessão de permissões para aplicativos produtores e consumidores, consulte the section called “Permissões para usar as chaves do KMS geradas pelo usuário”.

> **Note**
>
> O uso da criptografia do lado do servidor incorre AWS Key Management Service em custos ().AWS KMS Para obter mais informações, consulte AWS Key Management Service Pricing.

Cotas e limites

A tabela a seguir descreve as cotas de fluxos e fragmentos e os limites do Amazon Kinesis Data Streams.

Quota	Modo sob demanda	Modo provisionado
Número de fluxos de dados	Não há cota máxima no número de streams em sua AWS conta. Por padrão, pode-se criar até 50 fluxos de dados usando o modo de capacidad e sob demanda. Se houver necessidade de uma cota maior, abra um tíquete de suporte.	Não há uma cota máxima para o número de fluxos com o modo provisionado em uma conta.
Número de fragmentos	Não há limite máximo. O número de fragmentos depende da quantidade de dados ingeridos e do nível de throughput necessário. O Kinesis Data Streams escala automaticamente o número de fragmentos em resposta a mudanças no volume e no tráfego de dados.	Não há limite máximo. A cota de fragmentos padrão é de 20.000 fragmentos por cada Conta da AWS , para o seguinte: Regiões da AWS • Leste dos EUA (Norte da Virgínia) • Oeste dos EUA (Oregon) • Europa (Irlanda) Em todas as outras regiões, a cota padrão é de 1.000 ou 6.000 fragmentos por Conta da AWS. Você pode ver a cota de fragmentos e a utilizaçã

Quota	Modo sob demanda	Modo provisionado
		o da sua conta por meio do console Service Quotas em. https://console.aws.amazon.com/servicequotas/ Para solicitar um aumento na cota de fragmentos, use o console Service Quotas ou. AWS CLI Consulte Solicitar um aumento de cotapara ter mais informações.

Quota	Modo sob demanda	Modo provisionado
Throughput do fluxo de dados	Por padrão, os novos fluxos de dados criados com o modo de capacidade sob demanda têm 4 MB/s de taxa de transferência de gravação e 8 MB/s de leitura. No Leste dos EUA (Norte da Virgínia), Oeste dos EUA (Oregon) e Europa (Irlanda) Regiões da AWS, os fluxos de dados com o modo de capacidade sob demanda aumentam até 10% GB/s da taxa de transferência de gravação e 20% de leitura. GB/s Para outras regiões, os fluxos de dados com o modo de capacidade sob demanda aumentam para até 200 MB/s de taxa de transferência de gravação e 400 de MB/s leitura. Se você precisar de um aumento de até 10 capacidades de GB/s gravação e 20 de GB/s leitura para essas regiões, envie um ticket de suporte.	Não há limite máximo. O throughput máximo depende do número de fragmentos provisionados para o fluxo. Cada fragmento pode suportar até 1 MB/sec ou 1.000 taxas de transferência de records/sec gravação ou até 2 MB/sec ou 2.000 taxas de transferência de records/sec leitura. Se precisar de mais capacidade de ingestão, você pode facilmente aumentar o número de fragmentos no stream usando a Console de gerenciamento da AWS ou a UpdateShardCountAPI.
Tamanho da carga útil de dados	O tamanho máximo da carga útil de dados de um registro antes da `base64-encoding` é de até 10 MiB. O Kinesis foi projetado para lidar com registros grandes intermitentes (tamanho de 1 a 10 MiB) usando a capacidade de expansão.	

Quota	Modo sob demanda	Modo provisionado
Tamanho da transação `GetRecords`	GetRecordspode recuperar até 10 MB de dados por chamada de um único fragmento e até 10.000 registros por chamada. Cada chamada para `GetRecords` é contada como uma transação de leitura. Cada fragmento pode oferecer suporte a até cinco transações de leitura por segundo. Cada transação de leitura pode fornecer até 10.000 registros com uma cota máxima de 10 MB por transação.	
Taxa de leitura de dados por fragmento	Cada fragmento pode suportar até uma taxa máxima total de leitura de dados de 2 MB por segundo via GetRecords. Se uma chamada para `GetRecords` retornar 10 MB, as chamadas subsequentes feitas nos próximos 5 segundos lançarão uma exceção.	
Número de consumidores registrados por fluxo de dados	Com o modo Kinesis On-demand Advantage, você pode criar até 50 consumidores registrados (Enhanced Fan-out). Com os modos Kinesis On-Demand Standard e Kinesis Provisioned, você pode criar até 20 consumidores registrados (limite de fan-out aprimorado) para cada stream de dados.	
Alternar entre os modos provisionado e sob demanda	Para cada fluxo de dados em sua AWS conta, você pode alternar entre os modos de capacidade sob demanda e provisionada duas vezes em 24 horas.	

Limites da API

Como a maioria AWS APIs, as operações da API Kinesis Data Streams têm taxas limitadas. Os limites a seguir se aplicam por conta da AWS e por região. Para obter mais informações sobre o Kinesis APIs Data Streams, consulte a Amazon Kinesis API Reference.

Limites de API do plano de controle do KDS

A seção a seguir descreve os limites do plano APIs de controle KDS. O plano de controle KDS APIs permite que você crie e gerencie seus fluxos de dados. Esses limites se aplicam por conta da AWS e por região.

Limites de API do plano de controle

solicitações de	Limite de chamada de API	Por conta/fluxo	Description
AddTagsToStream	5 transações por segundo (TPS)	Por conta	50 tags por fluxo de dados
CreateStream	5 TPS	Por conta	Não há um cota máxima para o número de fluxos em uma conta. É enviada uma `LimitExceededException` ao fazer uma solicitação `CreateStream` tentando executar um dos seguintes procedimentos: • Ter mais de cinco fluxos no estado `CREATING` em qualquer momento. • Criar mais fragmentos do que o autorizado para sua conta.
DecreaseStreamRetentionPeriod	5 TPS	Por fluxo	O valor mínimo do período de retenção de um fluxo de dados é de 24 horas.

solicitações de	Limite de chamada de API	Por conta/fluxo	Description
DeleteResourcePolicy	5 TPS	Por conta	Se precisar de um limite maior, abra um Tíquete de suporte.
DeleteStream	5 TPS	Por conta	
DeregisterStreamConsumer	5 TPS	Por fluxo	
DescribeAccountSettings	5 TPS	Por conta	
DescribeLimits	1 TPS	Por conta	
DescribeStream	10 TPS	Por conta	
DescribeStreamConsumer	20 TPS	Por fluxo	
DescribeStreamSummary	20 TPS	Por conta	
DisableEnhancedMonitoring	5 TPS	Por fluxo	
EnableEnhancedMonitoring	5 TPS	Por fluxo	
GetResourcePolicy	5 TPS	Por conta	Se precisar de um limite maior, abra um Tíquete de suporte.

solicitações de	Limite de chamada de API	Por conta/fluxo	Description
IncreaseStreamRetentionPeriod	5 TPS	Por fluxo	O valor máximo do período de retenção de um fluxo é de 8.760 horas (365 dias).
ListShards	1000 TPS	Por fluxo	
ListStreamConsumers	5 TPS	Por fluxo	
ListStreams	5 TPS	Por conta	
ListTagsForStream	5 TPS	Por fluxo	
MergeShards	5 TPS	Por fluxo	Aplicável somente no modo provisionado.
PutResourcePolicy	5 TPS	Por conta	Se precisar de um limite maior, abra um Tíquete de suporte.

solicitações de	Limite de chamada de API	Por conta/fluxo	Description
RegisterStreamConsumer	5 TPS	Por fluxo	É possível registrar até 20 aplicaçõs de consumo por fluxo de dados. Uma aplicação de consumo só pode ser registrada em um fluxo de dados de cada vez. Apenas cinco aplicações de consumo podem ser criadas simultaneamente. Em outras palavras, não é possível ter mais de cinco aplicações de consumo com status `CREATING` ao mesmo tempo.
RemoveTagsFromStream	5 TPS	Por fluxo	
SplitShard	5 TPS	Por fluxo	Aplicável somente no modo provisionado
StartStreamEncryption		Por fluxo	Você pode aplicar com sucesso uma nova chave AWS KMS para criptografia do lado do servidor 25 vezes em um período contínuo de 24 horas.

solicitações de	Limite de chamada de API	Por conta/fluxo	Description
StopStreamEncryption		Por fluxo	É possível desabilitar com êxito a criptografia no lado do servidor 25 vezes em um período contínuo de 24 horas.
UpdateShardCount		Por fluxo	Aplicável somente no modo provisionado. O limite padrão do número de fragmentos é 10.000. Há limites adicionais nesta API. Para obter mais informações, consulte UpdateShardCount.
UpdateStreamMode		Por fluxo	Para cada fluxo de dados em sua AWS conta, você pode alternar entre os modos de capacidade sob demanda e provisionada duas vezes em 24 horas.

solicitações de	Limite de chamada de API	Por conta/fluxo	Description
UpdateStreamWarmThroughput	5 TPS	Por conta	O throughput máximo aquecido que pode ser configurado é o limite de throughput do fluxo de dados do modo sob demanda para a conta e a região.
UpdateAccountSettings	5 TPS	Por conta	Habilite ou desabilite configurações da conta, como o modo On-Demand Advantage.

Limites de API do plano de dados do KDS

A seção a seguir descreve os limites do plano APIs de dados KDS. O plano de dados KDS APIs permite que você use seus fluxos de dados para coletar e processar registros de dados em tempo real. Esses limites se aplicam por fragmento dentro dos fluxos de dados.

Limites de API do plano de dados

solicitações de	Limite de chamada de API	Limite de carga útil	Outros detalhes
GetRecords	5 TPS	O número máximo de registros que podem ser retornados por chamada é 10.000. O tamanho máximo de dados que `GetRecords` pode retornar é 10 MB.	Se uma chamada retornar essa quantidade de dados, as chamadas subsequentes feitas nos próximos cinco segundos gerarão `Provision`

solicitações de	Limite de chamada de API	Limite de carga útil	Outros detalhes
			`edThrough putExceed edException` . Se o throughpu t provisionado no fluxo for insuficie nte, as chamadas subsequentes feitas no próximo segundo gerarão `Provision edThrough putExceed edException` .
GetShardIterator	5 TPS		Um iterador de fragmentos expira cinco minutos depois que é retornado ao solicitante. Se uma GetShardIterator solicitação for feita com muita frequênci a, você recebe uma ProvisionedThrough putExceededExcepti on.

solicitações de	Limite de chamada de API	Limite de carga útil	Outros detalhes
PutRecord	1000 TPS	Cada fragmento pode oferecer suporte a gravações de até 1.000 registros por segundo, até um total máximo de gravação de dados de 10 MiB por segundo.	O Kinesis foi projetado para lidar com registros grandes intermitentes (tamanho de 1 a 10 MiB) usando a capacidade de expansão.
PutRecords		Cada PutRecords solicitação pode suportar até 500 registros. Cada registro na solicitação pode ter no máximo 10 MiB, até um limite de 10 MiB para toda a solicitação, incluindo as chaves de partição. Cada fragmento pode oferecer suporte a gravações de até 1.000 registros por segundo, até um total máximo de gravação de dados de 1 MB por segundo.	O Kinesis foi projetado para lidar com registros grandes intermitentes (tamanho de 1 a 10 MiB) usando a capacidade de expansão.

solicitações de	Limite de chamada de API	Limite de carga útil	Outros detalhes
SubscribeToShard	Você pode fazer uma chamada Subscribe ToShard por segundo por consumido r registrado por fragmento.		Se você ligar SubscribeToShard novamente com o mesmo ConsumerArn ShardId e dentro de 5 segundos após uma chamada bem-suced ida, você receberá uma. ResourceI nUseException

Aumento de cotas

É possível usar o Service Quotas para solicitar um aumento para uma cota, se a cota for ajustável. Algumas solicitações são resolvidas automaticamente, enquanto outras são enviadas ao AWS Support. Você pode acompanhar o status de uma solicitação de aumento de cota enviada ao AWS Support. As solicitações para aumentar as Service Quotas não recebem suporte prioritário. Se você tiver uma solicitação urgente, entre em contato com o AWS Support. Para obter mais informações, consulte What is Service Quotas?

Para solicitar um aumento de cota de serviço, siga o procedimento descrito em Solicitar um aumento de cota.

Pré-requisitos completos para configurar o Amazon Kinesis Data Streams

Antes de usar o Amazon Kinesis Data Streams pela primeira vez, realize as tarefas a seguir para configurar seu ambiente.

Tarefas

- Inscreva-se para AWS
- Fazer download de bibliotecas e ferramentas
- Configurar seu ambiente de desenvolvimento

Inscreva-se para AWS

Quando você se inscreve no Amazon Web Services (AWS), sua AWS conta é automaticamente inscrita em todos os serviços AWS, incluindo o Kinesis Data Streams. A cobrança incorrerá apenas pelos serviços utilizados.

Se você já tiver uma AWS conta, vá para a próxima tarefa. Se ainda não possuir uma conta da AWS , use o procedimento a seguir para criar uma.

Para se inscrever em uma AWS conta

1. Abra a https://portal.aws.amazon.com/billing/inscrição.
2. Siga as instruções online.

   Parte do procedimento de inscrição envolve receber uma chamada telefônica ou uma mensagem de texto e inserir um código de verificação pelo teclado do telefone.

   Quando você se inscreve em um Conta da AWS, um Usuário raiz da conta da AWSé criado. O usuário-raiz tem acesso a todos os Serviços da AWS e recursos na conta. Como prática recomendada de segurança, atribua o acesso administrativo a um usuário e use somente o usuário-raiz para executar tarefas que exigem acesso de usuário-raiz.

Fazer download de bibliotecas e ferramentas

Estas bibliotecas e ferramentas ajudam a trabalhar com o Kinesis Data Streams:

- A Referência de APIs do Amazon Kinesis Data Firehose lista o conjunto básico de operações às quais o Kinesis Data Firehose oferece suporte. Para obter mais informações sobre a execução de operações básicas usando código Java, consulte o seguinte:
  - Desenvolva produtores usando a API Amazon Kinesis Data Streams com o AWS SDK para Java
  - Desenvolva consumidores com o AWS SDK para Java
  - Criar e gerenciar fluxos de dados do Kinesis
- As versões AWS SDKs para Go, Java, .NET JavaScript, PHP, Python e Ruby incluem suporte e amostras do Kinesis Data Streams. Se sua versão do AWS SDK para Java não incluir amostras do Kinesis Data Streams, você também poderá baixá-las em. GitHub
- A Kinesis Client Library (KCL) fornece um modelo de easy-to-use programação para processamento de dados. A KCL ajuda a começar a usar rapidamente o Kinesis Data Streams em Java, Node.js, .NET, Python e Ruby. Para obter mais informações, consulte Ler dados de fluxos.
- A AWS Command Line Interface é compatível com o Kinesis Data Streams. AWS CLI Isso permite que você controle vários AWS serviços a partir da linha de comando e os automatize por meio de scripts.

Configurar seu ambiente de desenvolvimento

Para usar a KCL, confirme que seu ambiente de desenvolvimento Java atende aos seguintes requisitos:

- Java 1.7 (Java SE 7 JDK) ou posterior. É possível fazer o download do software Java mais recente em Java SE Downloads no site da Oracle.
- Pacote Apache Commons (código, cliente HTTP e logs)
- Processador Jackson JSON

Observe que o AWS SDK para Java inclui o Apache Commons e o Jackson na pasta de terceiros. No entanto, o SDK para Java funciona com o Java 1.6, enquanto a Kinesis Client Library exige o Java 1.7.

Use o AWS CLI para realizar operações do Amazon Kinesis Data Streams

Esta seção mostra como realizar operações básicas do Amazon Kinesis Data Streams usando o. AWS Command Line Interface Serão apresentados os princípios fundamentais do fluxo de dados do Kinesis Data Streams e as etapas necessárias para colocar e obter dados de um fluxo de dados do Kinesis.

Se não há experiência prévia com o Kinesis Data Streams, comece familiarizando-se com os conceitos e a terminologia apresentados em Terminologia e conceitos do Amazon Kinesis Data Streams.

Tópicos

- Tutorial: Instalar e configurar o AWS CLI para Kinesis Data Streams
- Tutorial: Execute operações básicas do Kinesis Data Streams usando o AWS CLI

Para acesso à CLI, é necessário ter um ID de chave de acesso e de uma chave de acesso secreta. Use credenciais temporárias em vez de chaves de acesso de longo prazo quando possível. As credenciais temporárias incluem um ID de acesso, uma chave de acesso secreta e um token de segurança que indica quando as credenciais expiram. Para obter mais informações, consulte Uso de credenciais temporárias com AWS recursos no Guia do usuário do IAM.

Você pode encontrar instruções detalhadas de configuração step-by-step do IAM e da chave de segurança em Criar um usuário do IAM.

Nesta seção, os comandos específicos discutidos são fornecidos textualmente, exceto onde valores específicos são necessariamente diferentes para cada execução. Além disso, os exemplos estão usando a região Oeste dos EUA (Oregon), mas as etapas desta seção funcionam em qualquer região onde o Kinesis Data Streams é compatível.

Tutorial: Instalar e configurar o AWS CLI para Kinesis Data Streams

Instale o AWS CLI

Para obter etapas detalhadas sobre como instalar o AWS CLI para Windows e para os sistemas operacionais Linux, OS X e Unix, consulte Instalando a AWS CLI.

Use o comando a seguir para listar as opções e os serviços disponíveis:

```
aws help
```

Você usará o serviço Kinesis Data Streams para poder AWS CLI revisar os subcomandos relacionados ao Kinesis Data Streams usando o seguinte comando:

```
aws kinesis help
```

Esse comando gera uma saída que inclui os comandos disponíveis do Kinesis Data Streams:

```
AVAILABLE COMMANDS

       o add-tags-to-stream

       o create-stream

       o delete-stream

       o describe-stream

       o get-records

       o get-shard-iterator

       o help

       o list-streams

       o list-tags-for-stream

       o merge-shards
```

```
        o put-record

        o put-records

        o remove-tags-from-stream

        o split-shard

        o wait
```

A lista de comandos corresponde à API do Kinesis Data Streams documentada na Referência de APIs do serviço Amazon Kinesis. Por exemplo, o comando `create-stream` corresponde à ação `CreateStream` da API.

O agora AWS CLI está instalado com sucesso, mas não está configurado. Isso é mostrado na próxima seção.

Configurar o AWS CLI

Para uso geral, o `aws configure` comando é a maneira mais rápida de configurar sua AWS CLI instalação. Para obter mais informações, consulte Configurando a AWS CLI.

Tutorial: Execute operações básicas do Kinesis Data Streams usando o AWS CLI

Esta seção descreve o uso básico de um fluxo de dados do Kinesis na linha de comando usando a AWS CLI. Certifique-se de estar familiarizado com os conceitos discutidos em Terminologia e conceitos do Amazon Kinesis Data Streams.

> **Note**
>
> Depois de criar um stream, sua conta incorre em cobranças nominais pelo uso do Kinesis Data Streams porque o Kinesis Data Streams não está qualificado para o nível gratuito. AWS Ao concluir este tutorial, exclua seus AWS recursos para parar de incorrer em cobranças. Para obter mais informações, consulte Etapa 4: limpar.

Tópicos

- Etapa 1: criar um fluxo
- Etapa 2: colocar um registro
- Etapa 3: obter o registro
- Etapa 4: limpar

Etapa 1: criar um fluxo

A primeira etapa é criar um fluxo e verificar se ele foi criado com êxito. Use o comando a seguir para criar um fluxo denominado "Foo":

```
aws kinesis create-stream --stream-name Foo
```

Em seguida, emita o comando a seguir para verificar o andamento da criação do fluxo:

```
aws kinesis describe-stream-summary --stream-name Foo
```

É necessário obter uma saída semelhante ao exemplo a seguir:

```
{
    "StreamDescriptionSummary": {
        "StreamName": "Foo",
        "StreamARN": "arn:aws:kinesis:us-west-2:123456789012:stream/Foo",
        "StreamStatus": "CREATING",
        "RetentionPeriodHours": 48,
        "StreamCreationTimestamp": 1572297168.0,
        "EnhancedMonitoring": [
            {
                "ShardLevelMetrics": []
            }
        ],
        "EncryptionType": "NONE",
        "OpenShardCount": 3,
        "ConsumerCount": 0
    }
}
```

Neste exemplo, o fluxo tem o status CREATING, o que significa que ainda não está pronto para uso. Verifique novamente em alguns instantes para ver uma saída semelhante ao exemplo a seguir:

```
{
    "StreamDescriptionSummary": {
        "StreamName": "Foo",
        "StreamARN": "arn:aws:kinesis:us-west-2:123456789012:stream/Foo",
        "StreamStatus": "ACTIVE",
        "RetentionPeriodHours": 48,
        "StreamCreationTimestamp": 1572297168.0,
        "EnhancedMonitoring": [
            {
                "ShardLevelMetrics": []
            }
        ],
        "EncryptionType": "NONE",
        "OpenShardCount": 3,
        "ConsumerCount": 0
    }
}
```

Há informações nessa saída que não são necessárias neste tutorial. A informação importante agora é `"StreamStatus": "ACTIVE"`, que mostra que o fluxo está pronto para uso, e as informações sobre o único fragmento solicitado. Também é possível verificar a existência do novo fluxo usando o comando `list-streams`, como mostrado aqui:

```
aws kinesis list-streams
```

Saída:

```
{
    "StreamNames": [
        "Foo"
    ]
}
```

Etapa 2: colocar um registro

Agora que há um fluxo ativo, é possível colocar alguns dados. Neste tutorial, será usado o comando mais simples possível, `put-record`, que coloca um único registro de dados contendo o texto "testdata" no fluxo:

```
aws kinesis put-record --stream-name Foo --partition-key 123 --data testdata
```

Se bem-sucedido, esse comando gerará uma saída semelhante ao seguinte exemplo:

```
{
    "ShardId": "shardId-000000000000",
    "SequenceNumber": "49546986683135544286507457936321625675700192471156785154"
}
```

Parabéns, você adicionou dados a um fluxo! Em seguida, veja como obter dados do fluxo.

Etapa 3: obter o registro

GetShardIterator

Antes de obter dados do fluxo, é necessário obter o iterador referente ao fragmento do seu interesse. Um iterador de fragmentos representa a posição do fluxo e do fragmento da qual o consumidor (neste caso, o comando `get-record`) lerá. Use o comando `get-shard-iterator` da seguinte forma:

```
aws kinesis get-shard-iterator --shard-id shardId-000000000000 --shard-iterator-type
 TRIM_HORIZON --stream-name Foo
```

Lembre-se de que os comandos `aws kinesis` têm o suporte de uma API do Kinesis Data Streams, portanto, caso tenha curiosidade sobre algum parâmetro mostrado, leia o tópico de referência da API `GetShardIterator`. A execução bem-sucedida resultará em uma saída semelhante ao seguinte exemplo:

```
{
    "ShardIterator": "AAAAAAAAAAHSywljv0zEgPX4NyKdZ5wryMzP9yALs8NeKbUjp1IxtZs1Sp
+KEd9I6AJ9ZG4lNR1EMi+9Md/nHvtLyxpfhEzYvkTZ4D9DQVz/mBYWRO6OTZRKnW9gd
+efGN2aHFdkH1rJl4BL9Wyrk+ghYG22D2T1Da2EyNSH1+LAbK33gQweTJADBdyMwlo5r6PqcP2dzhg="
}
```

A string longa de caracteres aparentemente aleatórios é o iterador de fragmentos (a sua será diferente). Você deve inserir copy/paste o iterador de fragmento no comando get, mostrado a seguir. Os iteradores de fragmento têm uma vida útil válida de 300 segundos, o que deve ser tempo suficiente para você inserir copy/paste o iterador de fragmento no próximo comando. É

necessário remover quaisquer novas linhas do seu iterador de fragmentos antes de colá-lo no próximo comando. Ao receber uma mensagem de erro de que o iterador de fragmentos não é mais válido, execute o comando `get-shard-iterator` novamente.

GetRecords

O comando `get-records` recebe dados do fluxo e resulta em uma chamada a `GetRecords` na API do Kinesis Data Streams. O iterador de fragmentos especifica a posição, no fragmento, de onde iniciará a leitura dos registros de dados em sequência. Se não houver registros disponíveis na parte do fragmento para onde o iterador aponta, `GetRecords` retornará uma lista vazia. Poderá demorar várias chamadas para se chegar a uma parte do fragmento que contenha registros.

No exemplo do comando `get-records` a seguir:

```
aws kinesis get-records --shard-iterator
 AAAAAAAAAAHSywljv0zEgPX4NyKdZ5wryMzP9yALs8NeKbUjp1IxtZs1Sp+KEd9I6AJ9ZG4lNR1EMi
+9Md/nHvtLyxpfhEzYvkTZ4D9DQVz/mBYWRO6OTZRKnW9gd+efGN2aHFdkH1rJl4BL9Wyrk
+ghYG22D2T1Da2EyNSH1+LAbK33gQweTJADBdyMwlo5r6PqcP2dzhg=
```

Se estiver executando este tutorial a partir de um processador de comando do tipo Unix, como bash, poderá automatizar a aquisição de iterador de fragmentos usando um comando aninhado, como este:

```
SHARD_ITERATOR=$(aws kinesis get-shard-iterator --shard-id shardId-000000000000 --
shard-iterator-type TRIM_HORIZON --stream-name Foo --query 'ShardIterator')

aws kinesis get-records --shard-iterator $SHARD_ITERATOR
```

Se você estiver executando este tutorial em um sistema que oferece suporte PowerShell, você pode automatizar a aquisição do iterador de fragmentos usando um comando como este:

```
aws kinesis get-records --shard-iterator ((aws kinesis get-shard-iterator --shard-id
 shardId-000000000000 --shard-iterator-type TRIM_HORIZON --stream-name Foo).split('"')
[4])
```

O resultado bem-sucedido do comando `get-records` solicitará registros do seu fluxo para o fragmento especificado quando o iterador de fragmentos foi obtido, como no exemplo a seguir:

```
{
  "Records":[ {
```

```
    "Data":"dGVzdGRhdGE=",
    "PartitionKey":"123",
    "ApproximateArrivalTimestamp": 1.441215410867E9,
    "SequenceNumber":"49544985256907370027570885864065577703022652638596431874"
  } ],
  "MillisBehindLatest":24000,

 "NextShardIterator":"AAAAAAAAAAEDOW3ugseWPE4503kqN1yN1UaodY8unE0sYslMUmC6lX9hlig5+t4RtZM0/
tALfiI4QGjunVgJvQsjxjh2aLyxaAaPr
+LaoENQ7eVs4EdYXgKyThTZGPcca2fVXYJWL3yafv9dsDwsYVedI66dbMZFC8rPMWc797zxQkv4pSKvPOZvrUIudb8UkH3V
}
```

Observe que `get-records` é descrito acima como uma solicitação, ou seja, pode-se receber zero ou mais registros mesmo que haja registros em seu fluxo. Os registros retornados podem não representar todos os registros atualmente em seu fluxo. Isso é normal, e o código de produção pesquisará o fluxo em busca de registros em intervalos apropriados. Essa velocidade de pesquisa variará dependendo dos requisitos específicos de design do aplicativo.

Em seu registro nesta parte do tutorial, perceba que os dados parecem ser lixo e não o texto limpo `testdata` que foi enviado. Isso ocorre devido ao modo como `put-record` usa a codificação Base64 para permitir o envio de dados binários. No entanto, o suporte do Kinesis Data Streams no não fornece decodificação AWS CLI em Base64 porque a decodificação em Base64 em conteúdo binário bruto impresso em stdout pode causar comportamentos indesejados e possíveis problemas de segurança em determinadas plataformas e terminais. Ao usar um decodificador Base64 (por exemplo, https://www.base64decode.org/) para decodificar `dGVzdGRhdGE=` manualmente, verá que ele é, na verdade, `testdata`. Isso é suficiente para fins deste tutorial porque, na prática, raramente AWS CLI é usado para consumir dados. Mais frequentemente, ela é usada para monitorar o estado do fluxo e obter informações, conforme mostrado anteriormente (`describe-stream` e `list-streams`). Para obter mais informações sobre a KCL, consulte Developing Custom Consumers with Shared Throughput Using KCL.

Nem sempre `get-records` retornará todos os registros no fluxo/fragmento especificado. Quando isso acontecer, use o `NextShardIterator` a partir do último resultado para obter o próximo conjunto de registros. Se mais dados estavam sendo colocados no fluxo, o que é a situação normal em aplicativos de produção, pode-se continuar sondando dados usando `get-records` todas as vezes. No entanto, se `get-records` não for chamado usando o próximo iterador de fragmentos dentro do tempo de vida de 300 segundos do iterador de fragmentos, uma mensagem de erro será gerada, e o comando `get-shard-iterator` deverá ser usado para obter um novo iterador de fragmentos.

Essa saída também fornece `MillisBehindLatest`, que é o número de milissegundos em que a resposta da operação GetRecords está da ponta do stream, indicando o atraso do consumidor em relação ao tempo atual. Um valor zero indica que o processamento de registros foi alcançado e não há nenhum registro novo para processar no momento. No caso deste tutorial, se o meterial estiver sendo lido conforme durante o progresso, um número muito grande pode ser visto. Por padrão, os registros de dados permanecerão em um fluxo por 24 horas aguardando serem recuperados. Esse período, chamado de período de retenção, pode ser configurado para até 365 dias.

Um resultado bem-sucedido de `get-records` sempre terá um `NextShardIterator`, mesmo que não haja mais nenhum registro atualmente no fluxo. Este é um modelo de sondagem que assume que um produtor colocará mais registros no fluxo em um determinado momento. É possível criar rotinas de sondagem próprias. Porém, ao usar a KCL mencionada anteriormente para desenvolver aplicativos de consumidor, ela se encarregará dessa sondagem.

Ao chamar `get-records` até que não haja mais registros no fluxo e o fragmento do qual se esteja sondando, se obterá uma saída com registros vazios, semelhante ao exemplo a seguir:

```
{
    "Records": [],
    "NextShardIterator": "AAAAAAAAAAGCJ5jzQNjmdhO6B/YDIDE56jmZmrmMA/r1WjoHXC/
kPJXc1rckt3TFL55dENfe5meNgdkyCRpUPGzJpMgYHaJ53C3nCAjQ6s7ZupjXeJGoUFs5oCuFwhP+Wul/
EhyNeSs5DYXLSSC5XCapmCAYGFjYER69QSdQjxMmBPE/hiybFDi5qtkT6/PsZNz6kFoqtDk="
}
```

Etapa 4: limpar

Exclua seu fluxo para liberar recursos e evitar cobranças indesejadas à sua conta. Faça isso sempre que tiver criado um fluxo que não será usado, pois as cobranças incidem por fluxo mesmo que ele não seja usado para colocar e obter dados. O comando de limpeza é o seguinte:

```
aws kinesis delete-stream --stream-name Foo
```

O êxito do comando não gera saída. Use `describe-stream` para verificar o andamento da exclusão:

```
aws kinesis describe-stream-summary --stream-name Foo
```

Ao executar esse comando imediatamente após o comando de exclusão, haverá uma saída semelhante ao exemplo a seguir:

```
{
    "StreamDescriptionSummary": {
        "StreamName": "samplestream",
        "StreamARN": "arn:aws:kinesis:us-west-2:123456789012:stream/samplestream",
        "StreamStatus": "ACTIVE",
```

Após a exclusão total do fluxo, `describe-stream` gerará um erro "não encontrado":

```
A client error (ResourceNotFoundException) occurred when calling the
 DescribeStreamSummary operation:
Stream Foo under account 123456789012 not found.
```

Tutoriais de conceitos básicos do Amazon Kinesis Data Streams

O Amazon Kinesis Data Streams fornece várias soluções diferentes para ingerir e consumir dados dos fluxos de dados do Kinesis. Os tutoriais desta seção são projetados para ajudar na compreensão da funcionalidade e dos conceitos do Amazon Kinesis Data Streams e identificar a solução que atenda a cada necessidade.

Tópicos

- Tutorial: Processar dados de ações em tempo real usando a KCL e a KCL 2.x
- Tutorial: Processar dados de ações em tempo real usando a KPL e a KCL 1.x
- Tutorial: analise dados do mercado de ações em tempo real usando o Amazon Managed Service for Apache Flink
- Tutorial: Use AWS Lambda com o Amazon Kinesis Data Streams
- Use a solução AWS de streaming de dados para o Amazon Kinesis

Tutorial: Processar dados de ações em tempo real usando a KCL e a KCL 2.x

O cenário deste tutorial envolve consumir negociações do mercado de ações em um fluxo de dados e criar uma aplicação básica do Amazon Kinesis Data Streams para realizar cálculos no fluxo. Será explicado como enviar um fluxo de registros para o Kinesis Data Streams e implementar uma aplicação que consome e processa os registros em tempo quase real.

> **Important**
>
> Depois de criar um stream, sua conta incorre em cobranças nominais pelo uso do Kinesis Data Streams porque o Kinesis Data Streams não está qualificado para o nível gratuito. AWS Depois de iniciada, a aplicação de consumo também incorre em cobranças nominais pelo uso do Amazon DynamoDB. A aplicação de consumo usa o DynamoDB para monitorar o estado do processamento. Ao terminar de usar esta aplicação, exclua seus recursos da AWS para parar de gerar cobranças. Para obter mais informações, consulte Limpar os recursos.

O código não acessa os dados reais da bolsa de valores, ele simula o fluxo de negociações de ações. Isso é feito com o uso de um gerador de negociações de ações aleatórias cujo ponto de partida são dados do mercado real referente às 25 principais ações por capitalização de mercado em fevereiro de 2015. Se houver acesso a um fluxo de negociações de ações em tempo real, pode ser interessante derivar estatísticas úteis e em tempo hábil desse fluxo. Por exemplo, talvez convenha executar uma análise de janela deslizante na qual se determine a ação mais popular que foi adquirida nos últimos 5 minutos. Ou talvez convenha uma notificação sempre que uma ordem de venda for muito grande (ou seja, tenha muitas quotas). É possível estender o código nesta série para oferecer essa funcionalidade.

É possível executar as etapas deste tutorial no desktop ou laptop e executar o código de produtor e de consumidor na mesma máquina ou em qualquer plataforma que ofereça suporte aos requisitos definidos.

Os exemplos mostrados usam a região Oeste dos EUA (Oregon), mas funcionam em qualquer região da AWS que oferece suporte ao Kinesis Data Streams.

Tarefas

- Concluir os pré-requisitos
- Criar um fluxo de dados
- Criar um usuário e uma política do IAM
- Fazer download e criar o código
- Implementar o produtor
- Implementar o consumidor
- (Opcional) Estender o consumidor
- Limpar os recursos

Concluir os pré-requisitos

É necessário atender aos seguintes requisitos para concluir este tutorial:

Criar e usar uma conta da Amazon Web Services

Antes de começar, familiarize-se com os conceitos abordados em Terminologia e conceitos do Amazon Kinesis Data Streams, especialmente com fluxos, fragmentos, produtores e consumidores. Também é útil concluir as etapas no seguinte guia: Tutorial: Instalar e configurar o AWS CLI para Kinesis Data Streams.

Você deve ter uma AWS conta e um navegador da web para acessar Console de gerenciamento da AWS o.

Para acessar o console, use seu nome de usuário e senha do IAM para fazer login no Console de gerenciamento da AWS da página de login do IAM. Para obter informações sobre credenciais AWS de segurança, incluindo acesso programático e alternativas às credenciais de longo prazo, consulte as credenciais de AWS segurança no Guia do usuário do IAM. Para obter detalhes sobre como fazer login no seu Conta da AWS, consulte Como fazer login AWS no Guia do Início de Sessão da AWS usuário.

Para obter mais informações sobre o IAM e instruções de configuração da chave de segurança, consulte Criar um usuário do IAM.

Preencher os requisitos de software do sistema

O sistema usado para executar o aplicativo deve ter o Java 7 ou posterior instalado. Para fazer download e instalar o Java Development Kit (JDK) mais recente, acesse o Site de instalação do Java SE da Oracle.

É necessário ter a versão mais recente do AWS SDK para Java.

O aplicativo consumidor requer a Kinesis Client Library (KCL) versão 2.2.9 ou superior, que você pode obter em /tree/master. GitHub https://github.com/awslabs/ amazon-kinesis-client

Próximas etapas

Criar um fluxo de dados

Criar um fluxo de dados

Primeiro, é necessário criar o fluxo de dados que será usado nas etapas seguintes deste tutorial.

Para criar um fluxo

1. Faça login no Console de gerenciamento da AWS e abra o console do Kinesis em https://console.aws.amazon.com /kinesis.
2. Selecione Fluxos de dados no painel de navegação.
3. Na barra de navegação, expanda o seletor de região e escolha uma região.
4. Selecione Criar fluxo do Kinesis.
5. Insira um nome para seu fluxo de dados (por exemplo, **`StockTradeStream`**).

6. Digite **1** como o número de fragmentos, mas deixe Estimar o número de fragmentos necessários recolhido.
7. Selecione Criar fluxo do Kinesis.

Na página de lista Fluxos do Kinesis, o status do fluxo aparece como `CREATING` enquanto ele está sendo criado. Quando o fluxo fica pronto para uso, o status é alterado para `ACTIVE`.

Se o nome do fluxo for escolhido, na página exibida, a guia Detalhes exibirá um resumo da configuração do fluxo de dados. A seção Monitoramento exibe informações de monitoramento do fluxo.

Próximas etapas

Criar um usuário e uma política do IAM

Criar um usuário e uma política do IAM

Práticas recomendadas de segurança para AWS ditar o uso de permissões refinadas para controlar o acesso a diferentes recursos. AWS Identity and Access Management (IAM) permite gerenciar usuários e permissões de usuários no AWS. Uma Política do IAM lista explicitamente as ações permitidas e os recursos aos quais as ações são aplicáveis.

Veja a seguir as permissões mínimas normalmente necessárias para produtores e consumidores do Kinesis Data Streams.

Produtor

Ações	Recurso	Finalidade
`DescribeStream` , `DescribeStreamSummary` , `DescribeStreamConsumer`	Fluxo de dados do Kinesis	Antes de tentar ler registros, o consumidor verifica se o flux está ativo e se os fragmentos estão contidos no fluxo de da
`SubscribeToShard` , `RegisterStreamConsumer`	Fluxo de dados do Kinesis	Assina e registra os consumidores em um fragmento.
`PutRecord` , `PutRecords`	Fluxo de dados do Kinesis	Grava registros no Kinesis Data Streams.

Consumidor

Ações	Recurso	Finalidade
`DescribeStream`	Fluxo de dados do Kinesis	Antes de tentar ler registros, o consumidor verifica se o flux está ativo e se os fragmentos estão contidos no fluxo de da
`GetRecords` , `GetShardIterator`	Fluxo de dados do Kinesis	Lê registros de um fragmento.
`CreateTable` , `DescribeTable` , `GetItem`, `PutItem`, `Scan`, `UpdateItem`	Tabela do Amazon DynamoDB	Se for desenvolvido usando a Kinesis Client Library (KCL) v consumidor precisará de permissões para uma tabela do D monitorar o estado de processamento da aplicação.
`DeleteItem`	Tabela do Amazon DynamoDB	Para quando o consumidor realiza split/merge operações n Kinesis Data Streams.
`PutMetricData`	CloudWatch Registro da Amazon	O KCL também carrega métricas para CloudWatch, que sã aplicativo.

Neste tutorial, será criada uma única política do IAM que concede todas as permissões acima. Na produção, talvez convenha criar duas políticas, uma para produtores e outra para consumidores.

Para criar uma política do IAM

1. Localize o nome do recurso da Amazon (ARN) para o novo fluxo de dados criado na etapa anterior. Esse ARN pdoe ser encontrado listado como ARN do fluxo na parte superior da guia Detalhes. O formato do ARN é o seguinte:

   ```
   arn:aws:kinesis:region:account:stream/name
   ```

   region

   O código AWS da região; por exemplo,`us-west-2`. Para obter mais informações, consulte Conceitos de região e zona de disponibilidade.

account

O ID da AWS conta, conforme mostrado nas Configurações da conta.

name

O nome do fluxo de dados criado na etapa anterior, que é `StockTradeStream`.

2. Determine o ARN da tabela do DynamoDB a ser usada pelo consumidor (e a ser criado pela primeira instância de consumidor). Ele deve estar no seguinte formato:

```
arn:aws:dynamodb:region:account:table/name
```

A região e o ID da conta são idênticos aos valores do ARN do fluxo de dados sendo usado neste tutorial, mas o nome é o nome da tabela do DynamoDB criada e usada pela aplicação de consumo. A KCL usa o nome do aplicativo como nome da tabela. Nesta etapa, use `StockTradesProcessor` como o nome da tabela do DynamoDB, pois esse é o nome da aplicação usada nas etapas posteriores do tutorial.

3. No console do IAM, em Políticas (https://console.aws.amazon.com/iam/home #policies), escolha Create policy. Se este for o primeiro contato com políticas do IAM, escolha Conceitos básicos e Criar política.
4. Escolha Selecionar ao lado de Gerador de políticas.
5. Escolha o Amazon Kinesis como serviço. AWS
6. Selecione `DescribeStream`, `GetShardIterator`, `GetRecords`, `PutRecord`e `PutRecords` como ações permitidas.
7. Insira o ARN do fluxo de dados sendo usado neste tutorial.
8. Use Adicionar instrução para cada um dos seguintes:

AWS Serviço	Ações	ARN
Amazon DynamoDB	`CreateTable` , `DeleteItem` , `DescribeTable` , `GetItem`, `PutItem`, `Scan`, `UpdateItem`	O ARN da tabela do DynamoDB criado na Etapa 2 deste procedimento.
Amazon CloudWatch	`PutMetricData`	*

O asterisco (*) é usado quando não é necessário especificar um ARN. Nesse caso, é porque não há nenhum recurso específico CloudWatch no qual a `PutMetricData` ação seja invocada.

9. Escolha Próxima etapa.
10. Altere Nome da política para `StockTradeStreamPolicy`, revise o código e selecione Criar política.

O documento de política resultante deve ser semelhante a:

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "Stmt123",
            "Effect": "Allow",
            "Action": [
                "kinesis:DescribeStream",
                "kinesis:PutRecord",
                "kinesis:PutRecords",
                "kinesis:GetShardIterator",
                "kinesis:GetRecords",
                "kinesis:ListShards",
                "kinesis:DescribeStreamSummary",
                "kinesis:RegisterStreamConsumer"
            ],
            "Resource": [
                "arn:aws:kinesis:us-west-2:111122223333:stream/StockTradeStream"
            ]
        },
        {
            "Sid": "Stmt234",
            "Effect": "Allow",
            "Action": [
                "kinesis:SubscribeToShard",
                "kinesis:DescribeStreamConsumer"
            ],
            "Resource": [
```

```
                "arn:aws:kinesis:us-west-2:111122223333:stream/StockTradeStream/
*"
            ]
        },
        {
            "Sid": "Stmt456",
            "Effect": "Allow",
            "Action": [
                "dynamodb:*"
            ],
            "Resource": [
                "arn:aws:dynamodb:us-west-2:111122223333:table/
StockTradesProcessor"
            ]
        },
        {
            "Sid": "Stmt789",
            "Effect": "Allow",
            "Action": [
                "cloudwatch:PutMetricData"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

Para criar um usuário do IAM

1. Abra o console do IAM em https://console.aws.amazon.com/iam/.
2. Na página Usuários, selecione Adicionar usuário.
3. Em Nome de usuário, digite `StockTradeStreamUser`.
4. Em Tipo de acesso, selecione Aceso programático e, em seguida, selecione Próximo: permissões.
5. Escolha Anexar políticas existentes diretamente.
6. Pesquise por nome a política criada no procedimento anterior (`StockTradeStreamPolicy`). Selecione a caixa à esquerda do nome da política e selecione Próximo: revisão.
7. Revise os detalhes e o resumo e, em seguida, selecione Criar usuário.

8. Copie o ID da chave de acesso e salve-o de forma privada. Em Chave de acesso secreta, selecione Mostrar e salve a chave de forma privada também.
9. Cole as chaves de acesso e chaves secretas em um arquivo local em lugar seguro de acesso restrito. Para esse aplicativo, crie um arquivo denominado `~/.aws/credentials` (com permissões restritas). O arquivo deverá estar no seguinte formato:

```
[default]
aws_access_key_id=access key
aws_secret_access_key=secret access key
```

Para anexar uma política do IAM a um usuário

1. No console do IAM, abra Políticas e selecione Ações da política.
2. Selecione `StockTradeStreamPolicy` e Anexar.
3. Selecione `StockTradeStreamUser` e Anexar política.

Próximas etapas

Fazer download e criar o código

Fazer download e criar o código

Este tópico fornece um exemplo de código de implementação para o exemplo de ingestão de transações de ações no fluxo de dados (produtor) e o processamento desses dados (consumidor).

Como fazer download e criar o código

1. Baixe o código-fonte do https://github.com/aws-samples/amazon-kinesis-learning GitHub repositório para o seu computador.
2. Crie um projeto no IDE com o código-fonte, respeitando a estrutura de diretório fornecida.
3. Adicione as seguintes bibliotecas ao projeto:
   - Amazon Kinesis Client Library (KCL)
   - AWS SDK
   - Apache HttpCore
   - Apache HttpClient

- Apache Commons Lang
- Apache Commons Logging
- Guava (Google Core Libraries For Java)
- Jackson Annotations
- Jackson Core
- Jackson Databind
- Jackson Dataformat: CBOR
- Joda Time

4. Dependendo do seu IDE, o projeto pode ser compilado automaticamente. Se não, compile o projeto usando as etapas apropriadas para o seu IDE.

Se essas etapas foram concluídas com êxito, é possível agora ir para a próxima seção, the section called “Implementar o produtor”.

Próximas etapas

Implementar o produtor

Este tutorial usa o cenário do mundo real de monitoramento de transações da bolsa de valores. Os princípios a seguir explicam brevemente como este cenário é mapeado para o produtor e a estrutura de código de suporte.

Consulte o código-fonte e analise as informações a seguir.

StockTrade classe

Uma negociação de ação individual é representada por uma instância da classe StockTrade. Essa instância contém atributos como o símbolo ticker, o preço, o número de ações, o tipo da negociação (compra ou venda) e um ID que identifica a negociação com exclusividade. Essa classe é previamente implementada.

Registro de fluxo

Um fluxo é uma sequência de registros. Um registro é uma serialização de uma instância `StockTrade` no formato JSON. Por exemplo:

```
{
  "tickerSymbol": "AMZN",
  "tradeType": "BUY",
  "price": 395.87,
  "quantity": 16,
  "id": 3567129045
}
```

StockTradeGenerator classe

StockTradeGenerator tem um método chamado `getRandomTrade()` que retorna uma nova negociação de ações gerada aleatoriamente toda vez que é invocada. Essa classe é previamente implementada.

StockTradesWriter classe

O método `main` do produtor, StockTradesWriter, recupera continuamente uma negociação aleatória e a envia ao Kinesis Data Streams executando as seguintes tarefas:

1. Lê o nome do fluxo de dados e o nome da região como entrada.
2. Usa o `KinesisAsyncClientBuilder` para definir região, credenciais e configuração do cliente.
3. Verifica se o stream existe e está ativo (se não, ele será encerrado com um erro).
4. Em um loop contínuo, chama o método `StockTradeGenerator.getRandomTrade()` e o método `sendStockTrade` para enviar a negociação ao stream a cada 100 milissegundos.

O método `sendStockTrade` da classe `StockTradesWriter` tem o seguinte código:

```
private static void sendStockTrade(StockTrade trade, KinesisAsyncClient
 kinesisClient,
            String streamName) {
        byte[] bytes = trade.toJsonAsBytes();
        // The bytes could be null if there is an issue with the JSON serialization
 by the Jackson JSON library.
        if (bytes == null) {
            LOG.warn("Could not get JSON bytes for stock trade");
            return;
        }

        LOG.info("Putting trade: " + trade.toString());
```

```
        PutRecordRequest request = PutRecordRequest.builder()
                .partitionKey(trade.getTickerSymbol()) // We use the ticker symbol
as the partition key, explained in the Supplemental Information section below.
                .streamName(streamName)
                .data(SdkBytes.fromByteArray(bytes))
                .build();
        try {
            kinesisClient.putRecord(request).get();
        } catch (InterruptedException e) {
            LOG.info("Interrupted, assuming shutdown.");
        } catch (ExecutionException e) {
            LOG.error("Exception while sending data to Kinesis. Will try again next
cycle.", e);
        }
    }
```

Consulte o desmembramento do código a seguir:

- A API `PutRecord` espera uma matriz de bytes, e é necessário converter a transação para o formato JSON. Essa única linha de código executa a seguinte operação:

  ```
  byte[] bytes = trade.toJsonAsBytes();
  ```

- Antes de enviar a transação, crie uma nova instância de `PutRecordRequest` (chamada solicitação neste caso). Cada `request` exige o nome do fluxo, uma chave de partição e um blob de dados.

  ```
  PutPutRecordRequest request = PutRecordRequest.builder()
      .partitionKey(trade.getTickerSymbol()) // We use the ticker symbol as the
   partition key, explained in the Supplemental Information section below.
      .streamName(streamName)
      .data(SdkBytes.fromByteArray(bytes))
      .build();
  ```

  O exemplo usa um tíquete de ações como uma chave de partição, que mapeia o registro para um determinado fragmento. Na prática, deve haver centenas ou milhares de chaves de partição por fragmento, de forma que os registros sejam uniformemente disseminados no fluxo. Para

obter mais informações sobre como adicionar dados a um fluxo, consulte Gravar dados no Amazon Kinesis Data Streams.

Agora, `request` está pronto para enviar para o cliente (operação put):

```
   kinesisClient.putRecord(request).get();
```

- A verificação e o registro de erros são sempre inclusões úteis. Este código registra condições de erro:

```
if (bytes == null) {
    LOG.warn("Could not get JSON bytes for stock trade");
    return;
}
```

Adicione o try/catch bloco ao redor da `put` operação:

```
try {
  kinesisClient.putRecord(request).get();
} catch (InterruptedException e) {
            LOG.info("Interrupted, assuming shutdown.");
} catch (ExecutionException e) {
            LOG.error("Exception while sending data to Kinesis. Will try again
 next cycle.", e);
}
```

Isso ocorre porque uma operação put do Kinesis Data Streams pode falhar devido a erro de rede ou porque o fluxo de dados pode atingir o limite de throughput e ficar limitado. É recomendado considerar cuidadosamente sua política de tentativa para operações `put` a fim de evitar perda de dados, por exemplo, usando como uma nova tentativa.

- O registro de status é útil mas opcional:

```
LOG.info("Putting trade: " + trade.toString());
```

O produtor mostrado aqui usa a funcionalidade de registro único da API do Kinesis Data Streams, `PutRecord`. Na prática, se um produtor individual gerar muitos registros, costuma ser mais eficiente usar a funcionalidade de vários registros de `PutRecords` e enviar lotes de registros por vez. Para obter mais informações, consulte Gravar dados no Amazon Kinesis Data Streams.

Como executar o produtor

1. Verifique se a chave de acesso e o par de chaves secretas recuperados em Criar um usuário e uma política do IAM estão salvos no arquivo `~/.aws/credentials`.
2. Execute a classe `StockTradeWriter` com os seguintes argumentos:

```
StockTradeStream us-west-2
```

   Se o fluxo foi criado em uma região diferente de `us-west-2`, será necessário especificar esta região.

A saída deve ser semelhante a:

```
Feb 16, 2015 3:53:00 PM
com.amazonaws.services.kinesis.samples.stocktrades.writer.StockTradesWriter
 sendStockTrade
INFO: Putting trade: ID 8: SELL 996 shares of BUD for $124.18
Feb 16, 2015 3:53:00 PM
com.amazonaws.services.kinesis.samples.stocktrades.writer.StockTradesWriter
 sendStockTrade
INFO: Putting trade: ID 9: BUY 159 shares of GE for $20.85
Feb 16, 2015 3:53:01 PM
com.amazonaws.services.kinesis.samples.stocktrades.writer.StockTradesWriter
 sendStockTrade
INFO: Putting trade: ID 10: BUY 322 shares of WMT for $90.08
```

Suas negociações de ações agora estão sendo ingeridas pelo Kinesis Data Streams.

Próximas etapas

Implementar o consumidor

Implementar o consumidor

O aplicativo consumidor neste tutorial processa continuamente as transações de ações em seu fluxo de dados. Em seguida, ele produz as ações mais populares compradas e vendidas a cada minuto. A aplicação é compilada com base na Kinesis Client Library (KCL), que faz grande parte do trabalho pesado comum às aplicações de consumo. Para obter mais informações, consulte Informações da KCL 1.x e 2.x.

Consulte o código-fonte e analise as informações a seguir.

StockTradesProcessor classe

A principal classe do consumidor fornecida e que executa as seguintes tarefas:

- Lê o aplicativo, o fluxo de dados e os nomes de região passados como argumentos.
- Cria uma instância de `KinesisAsyncClient` com o nome da região.
- Cria uma instância de `StockTradeRecordProcessorFactory` que veicula instâncias de `ShardRecordProcessor`, implementadas por uma instância de `StockTradeRecordProcessor`.
- Cria uma instância de `ConfigsBuilder` com a instância de `KinesisAsyncClient`, `StreamName`, `ApplicationName` e `StockTradeRecordProcessorFactory`. Isso é útil para criar todas as configurações com valores padrão.
- Cria um agendador da KCL (anteriormente, nas versões 1.x da KCL, era conhecido como o operador da KCL) com a instância de `ConfigsBuilder`.
- O agendador cria uma nova thread para cada fragmento (atribuído a essa instância de consumidor), que faz loop continuamente para ler registros do fluxo de dados. Em seguida, ele invoca a instância de `StockTradeRecordProcessor` para processar cada lote de registros recebidos.

StockTradeRecordProcessor classe

Implementação da instância de `StockTradeRecordProcessor`, que, por sua vez, implementa cinco métodos necessários: `initialize`, `processRecords`, `leaseLost`, `shardEnded` e `shutdownRequested`.

Os métodos `initialize` e `shutdownRequested` são usados pela KCL para permitir que o processador de registros saiba quando ele deve estar pronto para começar a receber registros e quando ele deve esperar parar de receber registros, respectivamente, para que ele possa

executar qualquer configuração específica do aplicativo e tarefas de encerramento. `leaseLost` e `shardEnded` são usados para implementar qualquer lógica para o que fazer quando um contrato de aluguel é perdido ou um processamento chegou ao fim de um fragmento. Neste exemplo, simplesmente registramos em log mensagens indicando esses eventos.

O código para esses métodos é fornecido para você. O processamento principal ocorre no método `processRecords`, que, por sua vez, usa `processRecord` para cada registro. Esse último método é fornecido como o código esqueleto quase todo vazio, para que seja implementado na próxima etapa, onde é explicado em mais detalhes.

Observe também a implementação dos métodos de suporte de `processRecord`: `reportStats` e `resetStats`, que estão vazios no código-fonte original.

O método `processRecords`, implementado previamente, executa as seguintes etapas:

- Para cada registro passado, ele chama `processRecord`.
- Se pelo menos 1 minuto houver decorrido após o último relatório, chamará `reportStats()`, que imprime as estatísticas mais recentes e, em seguida, `resetStats()`, que limpa as estatísticas para que o próximo intervalo inclua apenas registros novos.
- Define o próximo horário para geração de relatórios.
- Se houver decorrido pelo menos 1 minuto após o último ponto de verificação, chamará `checkpoint()`.
- Define o próximo horário do ponto de verificação.

Este método usa intervalos de 60 segundos como taxa de geração de relatórios e definição de pontos de verificação. Para obter mais informações sobre pontos de verificação, consulte Using the Kinesis Client Library.

StockStats classe

Essa classe fornece retenção de dados e rastreamento de estatísticas em relação às ações mais populares ao longo do tempo. Esse código é fornecido e contém os seguintes métodos:

- `addStockTrade(StockTrade)`: injeta o `StockTrade` conhecido nas estatísticas correntes.
- `toString()`: retorna as estatísticas em uma string formatada.

Essa classe rastreia as ações mais populares mantendo uma contagem corrente do número total de negociações de cada ação e a contagem máxima. Ela atualiza essas contagens sempre que chega uma negociação de ação.

Adicione código aos métodos da classe `StockTradeRecordProcessor`, como mostrado nas etapas a seguir.

Como implementar o consumidor

1. Implemente o método `processRecord` instanciando um objeto `StockTrade` de tamanho correto e adicionando a ele os dados do registro, registrando um aviso caso ocorra problema.

```
byte[] arr = new byte[record.data().remaining()];
record.data().get(arr);
StockTrade trade = StockTrade.fromJsonAsBytes(arr);
    if (trade == null) {
        log.warn("Skipping record. Unable to parse record into StockTrade.
 Partition Key: " + record.partitionKey());
        return;
        }
stockStats.addStockTrade(trade);
```

2. Implemente um método `reportStats`. Modifique o formato de saída para se adequar às suas preferências.

```
System.out.println("****** Shard " + kinesisShardId + " stats for last 1 minute
 ******\n" +
stockStats + "\n" +
"*****************************************************************\n");
```

3. Implemente o método `resetStats`, que cria uma nova instância de `stockStats`.

```
stockStats = new StockStats();
```

4. Implemente os seguintes métodos exigidos pela interface `ShardRecordProcessor`:

```
@Override
public void leaseLost(LeaseLostInput leaseLostInput) {
    log.info("Lost lease, so terminating.");
```

```
}

@Override
public void shardEnded(ShardEndedInput shardEndedInput) {
    try {
        log.info("Reached shard end checkpointing.");
        shardEndedInput.checkpointer().checkpoint();
    } catch (ShutdownException | InvalidStateException e) {
        log.error("Exception while checkpointing at shard end. Giving up.", e);
    }
}

@Override
public void shutdownRequested(ShutdownRequestedInput shutdownRequestedInput) {
    log.info("Scheduler is shutting down, checkpointing.");
    checkpoint(shutdownRequestedInput.checkpointer());
}

private void checkpoint(RecordProcessorCheckpointer checkpointer) {
    log.info("Checkpointing shard " + kinesisShardId);
    try {
        checkpointer.checkpoint();
    } catch (ShutdownException se) {
        // Ignore checkpoint if the processor instance has been shutdown (fail
 over).
        log.info("Caught shutdown exception, skipping checkpoint.", se);
    } catch (ThrottlingException e) {
        // Skip checkpoint when throttled. In practice, consider a backoff and
 retry policy.
        log.error("Caught throttling exception, skipping checkpoint.", e);
    } catch (InvalidStateException e) {
        // This indicates an issue with the DynamoDB table (check for table,
 provisioned IOPS).
        log.error("Cannot save checkpoint to the DynamoDB table used by the Amazon
 Kinesis Client Library.", e);
    }
}
```

Como executar o consumidor

1. Execute a aplicação de produção escrita em para injetar registros de negociações de ações no fluxo.
2. Verifique se o par de chave de acesso e chave secreta recuperado anteriormente (durante a criação do usuário do IAM) foi salvo no arquivo `~/.aws/credentials`.
3. Execute a classe `StockTradesProcessor` com os seguintes argumentos:

   ```
   StockTradesProcessor StockTradeStream us-west-2
   ```

   Observe que, ao criar o fluxo em uma região diferente de `us-west-2`, é necessário especificar essa região aqui.

Depois de um minuto, deverá aparecer uma saída como a seguir, atualizada a cada minuto a partir de então:

```
****** Shard shardId-000000000001 stats for last 1 minute ******
Most popular stock being bought: WMT, 27 buys.
Most popular stock being sold: PTR, 14 sells.
************************************************************
```

Próximas etapas

(Opcional) Estender o consumidor

(Opcional) Estender o consumidor

Esta seção opcional mostra como estender o código de consumidor para um cenário um pouco mais elaborado.

Para saber mais sobre os maiores pedidos de venda a cada minuto, pode-se modificar a classe `StockStats` em três locais para acomodar essa nova prioridade.

Para estender o consumidor

1. Adicione novas variáveis de instância:

```
 // Ticker symbol of the stock that had the largest quantity of shares sold
 private String largestSellOrderStock;
 // Quantity of shares for the largest sell order trade
 private long largestSellOrderQuantity;
```

2. Adicione o seguinte código a `addStockTrade`:

```
if (type == TradeType.SELL) {
     if (largestSellOrderStock == null || trade.getQuantity() >
 largestSellOrderQuantity) {
         largestSellOrderStock = trade.getTickerSymbol();
         largestSellOrderQuantity = trade.getQuantity();
     }
 }
```

3. Modifique o método `toString` para imprimir as informações adicionais:

```
public String toString() {
    return String.format(
        "Most popular stock being bought: %s, %d buys.%n" +
        "Most popular stock being sold: %s, %d sells.%n" +
        "Largest sell order: %d shares of %s.",
        getMostPopularStock(TradeType.BUY),
 getMostPopularStockCount(TradeType.BUY),
        getMostPopularStock(TradeType.SELL),
 getMostPopularStockCount(TradeType.SELL),
        largestSellOrderQuantity, largestSellOrderStock);
}
```

Se o consumidor for executado agora (lembre-se de executar o produtor também), deverá aparecer uma saída semelhante a esta:

```
****** Shard shardId-000000000001 stats for last 1 minute ******
Most popular stock being bought: WMT, 27 buys.
```

```
  Most popular stock being sold: PTR, 14 sells.
  Largest sell order: 996 shares of BUD.
  ****************************************************************
```

Próximas etapas

Limpar os recursos

Limpar os recursos

Como a utilização do fluxo de dados do Kinesis é paga, certifique-se de excluí-la e de excluir a tabela do Amazon DynamoDB correspondente ao concluir. As cobranças nominais ocorrerão em um fluxo ativo mesmo quando não houver envio e recebimento de registros. Isso ocorre porque um fluxo ativo usa recursos por meio da "escuta" contínua de registros recebidos e solicitações para obter registros.

Para excluir o fluxo e tabela

1. Desligue os produtores e consumidores que possam estar em execução.
2. Abra o console do Kinesis em https://console.aws.amazon.com /kinesis.
3. Escolha o fluxo criado para esta aplicação (`StockTradeStream`).
4. Escolha Excluir fluxo.
5. Abra o console do DynamoDB em. https://console.aws.amazon.com/dynamodb/
6. Exclua a tabela `StockTradesProcessor`.

Resumo

Para processar uma grande quantidade de dados quase em tempo real, não é preciso escrever nenhum código complicado nem desenvolver uma imensa infraestrutura. Isso é tão básico quanto escrever lógica para processar uma pequena quantidade de dados (como escrever `processRecord(Record)`), mas usando o Kinesis Data Streams para escalar e ter um processo que funciona para uma grande quantidade de dados de fluxo. Não há necessidade de se preocupar com a escalabilidade do processamento, porque o Kinesis Data Streams cuida de tudo. Basta enviar os registros de fluxo ao Kinesis Data Streams e escrever a lógica para processar cada novo registro recebido.

Veja aqui alguns aprimoramentos potenciais para este aplicativo.

Agregar em todos os fragmentos

Atualmente, obtém-se estatísticas resultantes da agregação de registros de dados recebidos por um único operador proveniente de um único fragmento. (Um fragmento não pode ser processado por mais de um operador em um aplicativo ao mesmo tempo). Naturalmente, ao escalar havendo mais de um fragmento, pode ser desejável agregar em todos os fragmentos. É possível fazer isso tendo uma arquitetura de pipeline em que a saída de cada operador é alimentada em outro fluxo com um único fragmento, o qual é processado por um operador que agrega as saídas do primeiro estágio. Como os dados do primeiro estágio são limitados (um exemplo por minuto por fragmento), eles podem ser facilmente tratados por um fragmento.

Escalar o processamento

Quando o fluxo é expandido para ter muitos fragmentos (porque muitos produtores estão enviando dados), a maneira de escalar o processamento é adicionando mais operadores. É possível executar os operadores em instâncias do Amazon EC2 e usar grupos do Auto Scaling.

Use conectores para o Amazon S3/ DynamoDB/Amazon Redshift/Storm

Como um fluxo é processado continuamente, sua saída pode ser enviada para outros destinos. AWS fornece conectores para integrar o Kinesis Data Streams com AWS outros serviços e ferramentas de terceiros.

Tutorial: Processar dados de ações em tempo real usando a KPL e a KCL 1.x

O cenário deste tutorial envolve consumir negociações do mercado de ações em um fluxo de dados e criar uma aplicação simples do Amazon Kinesis Data Streams para realizar cálculos no fluxo. Será explicado como enviar um fluxo de registros para o Kinesis Data Streams e implementar uma aplicação que consome e processa os registros em tempo quase real.

Important

Depois de criar um stream, sua conta incorre em cobranças nominais pelo uso do Kinesis Data Streams porque o Kinesis Data Streams não está qualificado para o nível gratuito. AWS Depois de iniciada, a aplicação de consumo também incorre em cobranças nominais pelo uso do Amazon DynamoDB. A aplicação de consumo usa o DynamoDB para monitorar o

estado do processamento. Ao terminar de usar esta aplicação, exclua seus recursos da AWS para parar de gerar cobranças. Para obter mais informações, consulte Limpar os recursos.

O código não acessa os dados reais da bolsa de valores, ele simula o fluxo de negociações de ações. Isso é feito com o uso de um gerador de negociações de ações aleatórias cujo ponto de partida são dados do mercado real referente às 25 principais ações por capitalização de mercado em fevereiro de 2015. Se houver acesso a um fluxo de negociações de ações em tempo real, pode ser interessante derivar estatísticas úteis e em tempo hábil desse fluxo. Por exemplo, talvez convenha executar uma análise de janela deslizante na qual se determine a ação mais popular que foi adquirida nos últimos 5 minutos. Ou talvez convenha uma notificação sempre que uma ordem de venda for muito grande (ou seja, tenha muitas quotas). É possível estender o código nesta série para oferecer essa funcionalidade.

É possível executar as etapas deste tutorial no desktop ou laptop. O código de produtor e o de consumidor podem ser executados na mesma máquina ou em qualquer plataforma que ofereça suporte aos requisitos definidos, como o Amazon Elastic Compute Cloud (Amazon EC2).

Os exemplos mostrados usam a região Oeste dos EUA (Oregon), mas funcionam em qualquer região da AWS que oferece suporte ao Kinesis Data Streams.

Tarefas

- Concluir os pré-requisitos
- Criar um fluxo de dados
- Criar um usuário e uma política do IAM
- Fazer download e compilação do código de implementação
- Implementar o produtor
- Implementar o consumidor
- (Opcional) Estender o consumidor
- Limpar os recursos

Concluir os pré-requisitos

Estes são os requisitos para concluir o Tutorial: Processar dados de ações em tempo real usando a KPL e a KCL 1.x.

Criar e usar uma conta da Amazon Web Services

Antes de começar, familiarize-se com os conceitos abordados em Terminologia e conceitos do Amazon Kinesis Data Streams, especialmente fluxos, fragmentos, produtores e consumidores. Também é útil ter concluído Tutorial: Instalar e configurar o AWS CLI para Kinesis Data Streams.

Você precisa de uma AWS conta e de um navegador da web para acessar Console de gerenciamento da AWS o.

Para acessar o console, use seu nome de usuário e senha do IAM para fazer login no Console de gerenciamento da AWS da página de login do IAM. Para obter informações sobre credenciais AWS de segurança, incluindo acesso programático e alternativas às credenciais de longo prazo, consulte as credenciais de AWS segurança no Guia do usuário do IAM. Para obter detalhes sobre como fazer login no seu Conta da AWS, consulte Como fazer login AWS no Guia do Início de Sessão da AWS usuário.

Para obter mais informações sobre o IAM e instruções de configuração da chave de segurança, consulte Criar um usuário do IAM.

Preencher os requisitos de software do sistema

O sistema usado para executar o aplicativo precisa ter o Java 7 ou superior instalado. Para fazer download e instalar o Java Development Kit (JDK) mais recente, acesse o Site de instalação do Java SE da Oracle.

Com um Java IDE, como o Eclipse,é possível abrir o código-fonte, editá-lo, compilá-lo e executá-lo.

É necessário ter a versão mais recente do AWS SDK para Java. Usando o Eclipse como IDE, é possível instalar o kit de ferramentas da AWS para Eclipse em vez desse SDK.

O aplicativo consumidor requer a Kinesis Client Library (KCL) versão 1.2.1 ou superior, que você pode obter na GitHub Kinesis Client Library (Java).

Próximas etapas

Criar um fluxo de dados

Criar um fluxo de dados

Na primeira etapa do Tutorial: Processar dados de ações em tempo real usando a KPL e a KCL 1.x, crie o fluxo que será usado em etapas subsequentes.

Para criar um fluxo

1. Faça login no Console de gerenciamento da AWS e abra o console do Kinesis em https://console.aws.amazon.com /kinesis.
2. Selecione Fluxos de dados no painel de navegação.
3. Na barra de navegação, expanda o seletor de região e escolha uma região.
4. Selecione Criar fluxo do Kinesis.
5. Insira um nome para seu fluxo (por exemplo, **`StockTradeStream`**).
6. Digite **1** como o número de fragmentos, mas deixe Estimar o número de fragmentos necessários recolhido.
7. Selecione Criar fluxo do Kinesis.

Na página de lista Fluxos do Kinesis, o status do fluxo é `CREATING` enquanto ele está sendo criado. Quando o fluxo fica pronto para uso, o status é alterado para `ACTIVE`. Escolha o nome do fluxo. Na página exibida, a guia Detalhes exibe um resumo da configuração do fluxo. A seção Monitoramento exibe informações de monitoramento do fluxo.

Informações adicionais sobre fragmentos

Ao começar a usar o Kinesis Data Streams fora deste tutorial, pode ser necessário planejar o processo de criação de fluxos mais cuidadosamente. É necessário planejar para a demanda máxima esperada ao provisionar fragmentos. Usando este cenário como exemplo, o tráfego de negociações da bolsa de valores dos EUA atinge o pico durante o dia (fuso horário do leste dos EUA) e, a partir desse horário, é preciso tirar amostras das estimativas de demanda. Em seguida, pode-se opcionalmente provisionar para a máxima demanda esperada ou expandir e reduzir o fluxo em resposta às variações de demanda.

Um fragmento é uma unidade de capacidade de throughput. Na página Criar fluxo do Kinesis, expanda Estime o número de fragmentos necessários. Digite o tamanho médio do registro, o máximo de registros gravados por segundo e o número de aplicativos de consumo usando as seguintes diretrizes:

Tamanho médio do registro

> Uma estimativa do tamanho médio calculado dos registros. Se esse valor não for conhecido, use o tamanho de registro máximo estimado.

Máximo de registros gravados

Considere o número de entidades que fornecem dados e o número aproximado de registros por segundo produzidos por cada um. Por exemplo, ao receber dados de negociações de ações provenientes de 20 servidores mercantis, com cada um gerando 250 negociações por segundo, o número total de transações (registros) por segundo é 5.000 por segundo.

Número de aplicativos de consumo

Número de aplicativos que fazem leitura do fluxo de forma independente para processar o fluxo de outro modo e produzir saída diferente. Cada aplicativo pode ter várias instâncias em execução em máquinas diferentes (ou seja, execução em um cluster) para dar conta de um fluxo de alto volume.

Se o número estimado de fragmentos mostrado exceder o limite atual de fragmentos, poderá ser necessário enviar uma solicitação para aumentar esse limite para poder criar um fluxo com esse número de fragmentos. Para solicitar um aumento do limite de fragmentos, use o formulário de limites do Kinesis Data Streams. Para obter mais informações sobre fluxos e fragmentos, consulte Criar e gerenciar fluxos de dados do Kinesis.

Próximas etapas

Criar um usuário e uma política do IAM

Criar um usuário e uma política do IAM

Práticas recomendadas de segurança para AWS ditar o uso de permissões refinadas para controlar o acesso a diferentes recursos. AWS Identity and Access Management (IAM) permite que você gerencie usuários e permissões de usuários no AWS. Uma Política do IAM lista explicitamente as ações permitidas e os recursos aos quais as ações são aplicáveis.

Veja a seguir as permissões mínimas normalmente necessárias para um produtor e um consumidor do Kinesis Data Streams.

Produtor

Ações	Recurso	Finalidade
`DescribeStream` , `DescribeStreamSumm`	Fluxo de dados do Kinesis	Antes de tentar gravar registros, a aplicação de produção v existe e está ativo, se os fragmentos estão contidos no flux consumidor.

Ações	Recurso	Finalidade
`ary` , `DescribeStreamConsumer`		
`SubscribeToShard` , `RegisterStreamConsumer`	Fluxo de dados do Kinesis	Faz a inscrição e registra um consumidor em um fragmento Kinesis.
`PutRecord` , `PutRecords`	Fluxo de dados do Kinesis	Gravar registros no Kinesis Data Streams.

Consumidor

Ações	Recurso	Finalidade
`DescribeStream`	Fluxo de dados do Kinesis	Antes de tentar ler registros, a aplicação de consumo verifi está ativo e se os fragmentos estão contidos no fluxo.
`GetRecords` , `GetShardIterator`	Fluxo de dados do Kinesis	Ler registros em um fragmento do Kinesis Data Streams.
`CreateTable` , `DescribeTable` , `GetItem`, `PutItem`, `Scan`, `UpdateItem`	Tabela do Amazon DynamoDB	Se for desenvolvido usando a Kinesis Client Library (KCL), de permissões para uma tabela do DynamoDB a fim de mo processamento da aplicação. O primeiro consumidor iniciad
`DeleteItem`	Tabela do Amazon DynamoDB	Para quando o consumidor realiza split/merge operações n Kinesis Data Streams.
`PutMetricData`	CloudWatch Registro da Amazon	O KCL também carrega métricas para CloudWatch, que sã aplicativo.

Para essa aplicação, crie uma única política do IAM que conceda todas as permissões anteriores. Na prática, talvez convenha considerar a criação de duas políticas, uma para produtores e uma para consumidores.

Para criar uma política do IAM

1. Localize o Nome de recurso da Amazon (ARN) para o novo fluxo. Esse ARN pdoe ser encontrado listado como ARN do fluxo na parte superior da guia Detalhes. O formato do ARN é o seguinte:

   ```
   arn:aws:kinesis:region:account:stream/name
   ```

   region

   Código da região; por exemplo, `us-west-2`. Para obter mais informações, consulte Conceitos de região e zona de disponibilidade.

   account

   O ID da AWS conta, conforme mostrado nas Configurações da conta.

   name

   Nome do fluxo de Criar um fluxo de dados, que é `StockTradeStream`.

2. Determine o ARN da tabela do DynamoDB a ser usada pelo consumidor (e criada pela primeira instância de consumidor). Ele deve estar no seguinte formato:

   ```
   arn:aws:dynamodb:region:account:table/name
   ```

   A região e a conta são do mesmo local que a etapa anterior, mas desta vez name é o nome da tabela criada e usada pelo aplicativo de consumidor. A KCL usada pelo consumidor usa o nome do aplicativo como o nome da tabela. Use o nome do aplicativo que será usado mais tarde, `StockTradesProcessor`.

3. No console do IAM, em Políticas (https://console.aws.amazon.com/iam/home #policies), escolha Create policy. Se este for o primeiro contato com políticas do IAM, escolha Conceitos básicos e Criar política.
4. Escolha Selecionar ao lado de Gerador de políticas.
5. Escolha o Amazon Kinesis como serviço. AWS
6. Selecione `DescribeStream`, `GetShardIterator`, `GetRecords`, `PutRecorde PutRecords` como ações permitidas.
7. Digite o ARN criado na Etapa 1.
8. Use Adicionar instrução para cada um dos seguintes:

AWS Serviço	Ações	ARN
Amazon DynamoDB	`CreateTable` , `DeleteItem` , `DescribeTable` , `GetItem`, `PutItem`, `Scan`, `UpdateItem`	O ARN criado na etapa 2
Amazon CloudWatch	`PutMetricData`	*

O asterisco (*) é usado quando não é necessário especificar um ARN. Nesse caso, é porque não há nenhum recurso específico CloudWatch no qual a `PutMetricData` ação seja invocada.

9. Escolha Próxima etapa.
10. Altere Nome da política para `StockTradeStreamPolicy`, revise o código e selecione Criar política.

O documento de política resultante deve ser algo como o exemplo a seguir:

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "Stmt123",
            "Effect": "Allow",
            "Action": [
                "kinesis:DescribeStream",
                "kinesis:PutRecord",
                "kinesis:PutRecords",
                "kinesis:GetShardIterator",
                "kinesis:GetRecords",
                "kinesis:ListShards",
                "kinesis:DescribeStreamSummary",
                "kinesis:RegisterStreamConsumer"
            ],
            "Resource": [
```

```
                "arn:aws:kinesis:us-west-2:111122223333:stream/StockTradeStream"
            ]
        },
        {
            "Sid": "Stmt234",
            "Effect": "Allow",
            "Action": [
                "kinesis:SubscribeToShard",
                "kinesis:DescribeStreamConsumer"
            ],
            "Resource": [
                "arn:aws:kinesis:us-west-2:111122223333:stream/StockTradeStream/
*"
            ]
        },
        {
            "Sid": "Stmt456",
            "Effect": "Allow",
            "Action": [
                "dynamodb:*"
            ],
            "Resource": [
                "arn:aws:dynamodb:us-west-2:111122223333:table/
StockTradesProcessor"
            ]
        },
        {
            "Sid": "Stmt789",
            "Effect": "Allow",
            "Action": [
                "cloudwatch:PutMetricData"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

Para criar um usuário do IAM

1. Abra o console do IAM em https://console.aws.amazon.com/iam/.

2. Na página Usuários, selecione Adicionar usuário.
3. Em Nome de usuário, digite `StockTradeStreamUser`.
4. Em Tipo de acesso, selecione Aceso programático e, em seguida, selecione Próximo: permissões.
5. Escolha Anexar políticas existentes diretamente.
6. Pesquise a política criada por nome. Selecione a caixa à esquerda do nome da política e selecione Próximo: revisão.
7. Revise os detalhes e o resumo e, em seguida, selecione Criar usuário.
8. Copie o ID da chave de acesso e salve-o de forma privada. Em Chave de acesso secreta, selecione Mostrar e salve a chave de forma privada também.
9. Cole as chaves de acesso e chaves secretas em um arquivo local em lugar seguro de acesso restrito. Para esse aplicativo, crie um arquivo denominado `~/.aws/credentials` (com permissões restritas). O arquivo deverá estar no seguinte formato:

```
[default]
aws_access_key_id=access key
aws_secret_access_key=secret access key
```

Para anexar uma política do IAM a um usuário

1. No console do IAM, abra Políticas e selecione Ações da política.
2. Selecione `StockTradeStreamPolicy` e Anexar.
3. Selecione `StockTradeStreamUser` e Anexar política.

Próximas etapas

Fazer download e compilação do código de implementação

Fazer download e compilação do código de implementação

O código esqueleto é fornecido para o the section called “Tutorial: Processar dados de ações em tempo real usando a KPL e a KCL 1.x”. Ele contém uma implementação de stub para o consumo do fluxo de negociações de ações (produtor) e para o processamento dos dados (consumidor). O procedimento a seguir mostra como concluir a implementação.

Para fazer download e compilação do código de implementação

1. Faça download do código-fonte no computador.
2. Crie um projeto no IDE favorito com o código-fonte, respeitando a estrutura de diretório fornecida.
3. Adicione as seguintes bibliotecas ao projeto:
   - Amazon Kinesis Client Library (KCL)
   - AWS SDK
   - Apache HttpCore
   - Apache HttpClient
   - Apache Commons Lang
   - Apache Commons Logging
   - Guava (Google Core Libraries For Java)
   - Jackson Annotations
   - Jackson Core
   - Jackson Databind
   - Jackson Dataformat: CBOR
   - Joda Time
4. Dependendo do seu IDE, o projeto pode ser compilado automaticamente. Se não, compile o projeto usando as etapas apropriadas para o seu IDE.

Se essas etapas foram concluídas com êxito, é possível agora ir para a próxima seção, the section called “Implementar o produtor”. Se a compilação gerar erros em qualquer estágio, investigue e os corrija antes de continuar.

Próximas etapas

Implementar o produtor

O aplicativo no Tutorial: Processar dados de ações em tempo real usando a KPL e a KCL 1.x usa o cenário real de monitoramento de negociações em bolsa de valores. Os princípios a seguir explicam brevemente como este cenário é mapeado para o produtor e a estrutura de código de apoio.

Consulte o código-fonte e analise as informações a seguir.

StockTrade classe

Uma negociação de ação individual é representada por uma instância da classe `StockTrade`. Essa instância contém atributos como o símbolo ticker, o preço, o número de ações, o tipo da negociação (compra ou venda) e um ID que identifica a negociação com exclusividade. Essa classe é previamente implementada.

Registro de fluxo

Um fluxo é uma sequência de registros. Um registro é uma serialização de uma instância `StockTrade` no formato JSON. Por exemplo:

```
{
  "tickerSymbol": "AMZN",
  "tradeType": "BUY",
  "price": 395.87,
  "quantity": 16,
  "id": 3567129045
}
```

StockTradeGenerator classe

`StockTradeGenerator` tem um método denominado `getRandomTrade()`, que retorna uma nova negociação de ações gerada aleatoriamente sempre que ela é invocada. Essa classe é previamente implementada.

StockTradesWriter classe

O método `main` do produtor, `StockTradesWriter`, recupera continuamente uma negociação aleatória e a envia ao Kinesis Data Streams executando as seguintes tarefas:

1. Lê o nome do fluxo e o nome da região como entrada.
2. Cria um `AmazonKinesisClientBuilder`.
3. Usa o criador do cliente para definir região, credenciais e configuração do cliente.
4. Cria um cliente `AmazonKinesis` usando o criador do cliente.
5. Verifica se o stream existe e está ativo (se não, ele será encerrado com um erro).
6. Em um loop contínuo, chama o método `StockTradeGenerator.getRandomTrade()` e o método `sendStockTrade` para enviar a negociação ao stream a cada 100 milissegundos.

O método `sendStockTrade` da classe `StockTradesWriter` tem o seguinte código:

```
private static void sendStockTrade(StockTrade trade, AmazonKinesis kinesisClient,
 String streamName) {
    byte[] bytes = trade.toJsonAsBytes();
    // The bytes could be null if there is an issue with the JSON serialization by
 the Jackson JSON library.
    if (bytes == null) {
        LOG.warn("Could not get JSON bytes for stock trade");
        return;
    }

    LOG.info("Putting trade: " + trade.toString());
    PutRecordRequest putRecord = new PutRecordRequest();
    putRecord.setStreamName(streamName);
    // We use the ticker symbol as the partition key, explained in the Supplemental
 Information section below.
    putRecord.setPartitionKey(trade.getTickerSymbol());
    putRecord.setData(ByteBuffer.wrap(bytes));

    try {
        kinesisClient.putRecord(putRecord);
    } catch (AmazonClientException ex) {
        LOG.warn("Error sending record to Amazon Kinesis.", ex);
    }
}
```

Consulte o desmembramento do código a seguir:

- A API de `PutRecord` espera uma matriz de bytes, e é necessário converter `trade` para o formato JSON. Essa única linha de código executa a seguinte operação:

  ```
  byte[] bytes = trade.toJsonAsBytes();
  ```

- Antes de enviar a negociação, crie uma nova instância de `PutRecordRequest` (denominada `putRecord` neste caso):

  ```
  PutRecordRequest putRecord = new PutRecordRequest();
  ```

  Cada chamada a `PutRecord` requer o nome do fluxo, uma chave de partição e um blob de dados. O código a seguir preenche esses campos no objeto `putRecord` usando seus métodos `setXxxx()`:

```
putRecord.setStreamName(streamName);
putRecord.setPartitionKey(trade.getTickerSymbol());
putRecord.setData(ByteBuffer.wrap(bytes));
```

O exemplo usa um tíquete de ações como uma chave de partição, que mapeia o registro para um determinado fragmento. Na prática, deve haver centenas ou milhares de chaves de partição por fragmento, de forma que os registros sejam uniformemente disseminados no fluxo. Para obter mais informações sobre como adicionar dados a um fluxo, consulte Adicionar dados a um fluxo.

Agora `putRecord` está pronto para enviar para o cliente (operação `put`):

```
kinesisClient.putRecord(putRecord);
```

- A verificação e o registro de erros são sempre inclusões úteis. Este código registra condições de erro:

```
if (bytes == null) {
    LOG.warn("Could not get JSON bytes for stock trade");
    return;
}
```

Adicione o try/catch bloco ao redor da `put` operação:

```
try {
       kinesisClient.putRecord(putRecord);
} catch (AmazonClientException ex) {
       LOG.warn("Error sending record to Amazon Kinesis.", ex);
}
```

Isso ocorre porque uma operação `put` do Kinesis Data Streams pode falhar devido a um erro de rede ou porque o fluxo de dados atinge o limite de throughput e tem sua utilização controlada. Recomendamos considerar cuidadosamente sua política de tentativa para operações `put` a fim de evitar perda de dados, usando como uma nova tentativa.

- O registro de status é útil mas opcional:

```
LOG.info("Putting trade: " + trade.toString());
```

O produtor mostrado aqui usa a funcionalidade de registro único da API do Kinesis Data Streams, `PutRecord`. Na prática, se um produtor individual gerar muitos registros, costuma ser mais eficiente usar a funcionalidade de vários registros de `PutRecords` e enviar lotes de registros por vez. Para obter mais informações, consulte Adicionar dados a um fluxo.

Como executar o produtor

1. Verifique se o par de chave de acesso e chave secreta recuperado anteriormente (durante a criação do usuário do IAM) foi salvo no arquivo `~/.aws/credentials`.
2. Execute a classe `StockTradeWriter` com os seguintes argumentos:

   ```
   StockTradeStream us-west-2
   ```

   Se o fluxo foi criado em uma região diferente de `us-west-2`, é necessário especificar essa região aqui.

A saída deve ser semelhante a:

```
Feb 16, 2015 3:53:00 PM
com.amazonaws.services.kinesis.samples.stocktrades.writer.StockTradesWriter
 sendStockTrade
INFO: Putting trade: ID 8: SELL 996 shares of BUD for $124.18
Feb 16, 2015 3:53:00 PM
com.amazonaws.services.kinesis.samples.stocktrades.writer.StockTradesWriter
 sendStockTrade
INFO: Putting trade: ID 9: BUY 159 shares of GE for $20.85
Feb 16, 2015 3:53:01 PM
com.amazonaws.services.kinesis.samples.stocktrades.writer.StockTradesWriter
 sendStockTrade
INFO: Putting trade: ID 10: BUY 322 shares of WMT for $90.08
```

Seu fluxo de negociações de ações agora está sendo ingerido pelo Kinesis Data Streams.

Próximas etapas

Implementar o consumidor

Implementar o consumidor

A aplicação de consumo no Tutorial: Processar dados de ações em tempo real usando a KPL e a KCL 1.x processa continuamente o fluxo de negociações de ações criado em . Em seguida, ele produz as ações mais populares compradas e vendidas a cada minuto. A aplicação é compilada com base na Kinesis Client Library (KCL), que faz grande parte do trabalho pesado comum às aplicações de consumo. Para obter mais informações, consulte Desenvolver aplicações de consumo da KCL 1.x.

Consulte o código-fonte e analise as informações a seguir.

StockTradesProcessor classe

Principal classe do consumidor fornecida e que executa as seguintes tarefas:

- Lê o aplicativo, o fluxo e os nomes de região passados como argumentos.
- Lê credenciais de `~/.aws/credentials`.
- Cria uma instância de `RecordProcessorFactory` que veicula instâncias de `RecordProcessor`, implementadas por uma instância de `StockTradeRecordProcessor`.
- Cria um operador da KCL com a instância `RecordProcessorFactory` e uma configuração padrão que inclui o nome do fluxo, as credenciais e o nome da aplicação.
- O operador cria um novo thread para cada fragmento (atribuído a essa instância de consumidor), que opera em loops contínuos para ler registros do Kinesis Data Streams. Em seguida, ele invoca a instância de `RecordProcessor` para processar cada lote de registros recebidos.

StockTradeRecordProcessor classe

Implementação da instância de `RecordProcessor`, que, por sua vez, implementa três métodos necessários: `initialize`, `processRecords` e `shutdown`.

Como os nomes sugerem, `initialize` e `shutdown` são usados pela Kinesis Client Library para permitir que o processador de registros saiba quando deve estar pronto para começar a receber registros e quando deve esperar parar de receber registros, respectivamente, para poder realizar tarefas de configuração e encerramento específicas da aplicação. Este código é fornecido para você. O processamento principal ocorre no método `processRecords`, que, por sua vez, usa `processRecord` para cada registro. Esse último método é fornecido como um código esqueleto quase todo vazio, para implementação na próxima etapa, onde é melhor explicado.

Observe também a implementação dos métodos de suporte de `processRecord`: `reportStats` e `resetStats`, que estão vazios no código-fonte original.

O método `processRecords`, implementado previamente, executa as seguintes etapas:

- Para cada registro passado, chama `processRecord`.
- Se pelo menos 1 minuto houver decorrido após o último relatório, chamará `reportStats()`, que imprime as estatísticas mais recentes e, em seguida, `resetStats()`, que limpa as estatísticas para que o próximo intervalo inclua apenas registros novos.
- Define o próximo horário para geração de relatórios.
- Se houver decorrido pelo menos 1 minuto após o último ponto de verificação, chamará `checkpoint()`.
- Define o próximo horário do ponto de verificação.

Este método usa intervalos de 60 segundos como taxa de geração de relatórios e definição de pontos de verificação. Para obter mais informações sobre definição de pontos de verificação, consulte Informações adicionais sobre o consumidor.

StockStats classe

Essa classe fornece retenção de dados e rastreamento de estatísticas em relação às ações mais populares ao longo do tempo. Esse código é fornecido e contém os seguintes métodos:

- `addStockTrade(StockTrade)`: injeta o `StockTrade` conhecido nas estatísticas correntes.
- `toString()`: retorna as estatísticas em uma string formatada.

Essa classe rastreia as ações mais populares mantendo uma contagem corrente do número total de negociações de cada ação e a contagem máxima. Ela atualiza essas contagens sempre que chega uma negociação de ação.

Adicione código aos métodos da classe `StockTradeRecordProcessor`, como mostrado nas etapas a seguir.

Como implementar o consumidor

1. Implemente o método `processRecord` instanciando um objeto `StockTrade` de tamanho correto e adicionando a ele os dados do registro, registrando um aviso caso ocorra problema.

```
StockTrade trade = StockTrade.fromJsonAsBytes(record.getData().array());
```

```
if (trade == null) {
    LOG.warn("Skipping record. Unable to parse record into StockTrade. Partition
 Key: " + record.getPartitionKey());
    return;
}
stockStats.addStockTrade(trade);
```

2. Implemente um método `reportStats` simples. Sinta-se à vontade para modificar o formato de saída conforme suas preferências.

```
System.out.println("****** Shard " + kinesisShardId + " stats for last 1 minute
 ******\n" +
                   stockStats + "\n" +

 "****************************************************************\n");
```

3. Finalmente, implemente o método `resetStats`, que cria uma nova instância de `stockStats`.

```
stockStats = new StockStats();
```

Como executar o consumidor

1. Execute a aplicação de produção escrita em para injetar registros de negociações de ações no fluxo.
2. Verifique se o par de chave de acesso e chave secreta recuperado anteriormente (durante a criação do usuário do IAM) foi salvo no arquivo `~/.aws/credentials`.
3. Execute a classe `StockTradesProcessor` com os seguintes argumentos:

```
StockTradesProcessor StockTradeStream us-west-2
```

   Observe que, ao criar o fluxo em uma região diferente de `us-west-2`, é necessário especificar essa região aqui.

Depois de um minuto, deverá aparecer uma saída como a seguir, atualizada a cada minuto a partir de então:

```
****** Shard shardId-000000000001 stats for last 1 minute ******
Most popular stock being bought: WMT, 27 buys.
Most popular stock being sold: PTR, 14 sells.
```

```
**********************************************************
```

Informações adicionais sobre o consumidor

Se já houver familiaridade com as vantagens da Kinesis Client Library, abordada em Desenvolver aplicações de consumo da KCL 1.x e em outros documentos, poderá haver algum quesitonamento sobre usá-la aqui. Mesmo usando apenas um fluxo de fragmento e uma instância de consumidor para processá-lo, é mais fácil implementar o consumidor usando a KCL. Compare as etapas de implementação do código na seção do produtor para o consumidor para ver a facilidade comparativa para implementar um consumidor. Isso se deve, em grande parte, aos serviços que a KCL fornece.

Nessa aplicação, cencentre-se na implementação de uma classe de processador de registros, capaz de processar registros individuais. Não é necessário se preocupar com a forma como os registros são obtidos do Kinesis Data Streams. A KCL obtém os registros e invoca o processador de registros sempre que há novos registros disponíveis. Além disso, não é necessário se preocupar com a quantidade de fragmentos e de instâncias de consumidor. Se o fluxo for escalonado, não é necessário reescrever o aplicativo para lidar com mais de um fragmento ou com uma instância de uma aplicação de consumo.

O termo ponto de verificação significa registrar o ponto no fluxo até os registros de dados que foram consumidos e processados até o momento. Se o aplicativo falhar, o fluxo será lido a partir desse ponto e não do início do fluxo. O assunto da definição de pontos de verificação e os vários padrões de design e melhores práticas relativos estão fora do escopo deste capítulo. No entanto, é algo que pode ser encontrado em ambientes de produção.

Conforme visto no , as operações `put` na API do Kinesis Data Streams usam uma chave de partição como entrada. O Kinesis Data Streams usa uma chave de partição como um mecanismo para dividir registros em vários fragmentos (quando há mais de um fragmento no fluxo). A mesma chave de partição sempre roteia para o mesmo fragmento. Isso permite que o consumidor que processa um determinado fragmento seja projetado com a premissa de que os registros com a mesma chave de partição só sejam enviados a esse consumidor, e nenhum registro com a mesma chave de partição termine em qualquer outro consumidor. Portanto, o operador de um consumidor pode agregar todos os registros com a mesma chave de partição sem se preocupar com a ausência de dados necessários.

Nesta aplicação, como o processamento de registros do consumidor não é intensivo, é possível usar um fragmento e fazer o processamento no mesmo thread da KCL. No entanto, na prática, considere primeiro escalar o número de fragmentos. Em alguns casos, talvez convenha

mudar o processamento para outro thread ou usar um grupo de threads se for esperado que o processamento de registros seja intensivo. Dessa forma, a KCL pode obter novos registros mais rapidamente, enquanto outros threads podem processar os registros em paralelo. O design multithread não é trivial e deve ser planejado com técnicas avançadas, portanto, aumentar a contagem de fragmentos costuma ser a maneira mais eficiente de escalar.

Próximas etapas

(Opcional) Estender o consumidor

(Opcional) Estender o consumidor

O aplicativo no Tutorial: Processar dados de ações em tempo real usando a KPL e a KCL 1.x já pode ser suficiente para os seus propósitos. Esta seção opcional mostra como estender o código de consumidor para um cenário um pouco mais elaborado.

Para saber mais sobre os maiores pedidos de venda a cada minuto, pode-se modificar a classe `StockStats` em três locais para acomodar essa nova prioridade.

Para estender o consumidor

1. Adicione novas variáveis de instância:

```
// Ticker symbol of the stock that had the largest quantity of shares sold
private String largestSellOrderStock;
// Quantity of shares for the largest sell order trade
private long largestSellOrderQuantity;
```

2. Adicione o seguinte código a `addStockTrade`:

```
if (type == TradeType.SELL) {
    if (largestSellOrderStock == null || trade.getQuantity() >
largestSellOrderQuantity) {
        largestSellOrderStock = trade.getTickerSymbol();
        largestSellOrderQuantity = trade.getQuantity();
    }
}
```

3. Modifique o método `toString` para imprimir as informações adicionais:

```
public String toString() {
```

```
    return String.format(
            "Most popular stock being bought: %s, %d buys.%n" +
            "Most popular stock being sold: %s, %d sells.%n" +
            "Largest sell order: %d shares of %s.",
            getMostPopularStock(TradeType.BUY),
getMostPopularStockCount(TradeType.BUY),
            getMostPopularStock(TradeType.SELL),
getMostPopularStockCount(TradeType.SELL),
            largestSellOrderQuantity, largestSellOrderStock);
}
```

Se o consumidor for executado agora (lembre-se de executar o produtor também), deverá aparecer uma saída semelhante a esta:

```
****** Shard shardId-000000000001 stats for last 1 minute ******
 Most popular stock being bought: WMT, 27 buys.
 Most popular stock being sold: PTR, 14 sells.
 Largest sell order: 996 shares of BUD.
 ************************************************************
```

Próximas etapas

Limpar os recursos

Limpar os recursos

Como a utilização do fluxo de dados do Kinesis é paga, certifique-se de excluí-la e de excluir a tabela do Amazon DynamoDB correspondente ao concluir. As cobranças nominais ocorrerão em um fluxo ativo mesmo quando não houver envio e recebimento de registros. Isso ocorre porque um fluxo ativo usa recursos por meio da "escuta" contínua de registros recebidos e solicitações para obter registros.

Para excluir o fluxo e tabela

1. Desligue os produtores e consumidores que possam estar em execução.
2. Abra o console do Kinesis em https://console.aws.amazon.com /kinesis.
3. Escolha o fluxo criado para esta aplicação (`StockTradeStream`).
4. Escolha Excluir fluxo.
5. Abra o console do DynamoDB em. https://console.aws.amazon.com/dynamodb/
6. Exclua a tabela `StockTradesProcessor`.

Resumo

Para processar uma grande quantidade de dados quase em tempo real, não é preciso escrever nenhum código complicado nem desenvolver uma imensa infraestrutura. Isso é tão básico quanto escrever lógica para processar uma pequena quantidade de dados (como escrever `processRecord(Record)`), mas usando o Kinesis Data Streams para escalar e ter um processo que funciona para uma grande quantidade de dados de fluxo. Não há necessidade de se preocupar com a escalabilidade do processamento, porque o Kinesis Data Streams cuida de tudo. Basta enviar os registros de fluxo ao Kinesis Data Streams e escrever a lógica para processar cada novo registro recebido.

Veja aqui alguns aprimoramentos potenciais para este aplicativo.

Agregar em todos os fragmentos

Atualmente, obtém-se estatísticas resultantes da agregação de registros de dados recebidos por um único operador proveniente de um único fragmento. (Um fragmento não pode ser processado por mais de um operador em um aplicativo ao mesmo tempo). Naturalmente, ao escalar havendo mais de um fragmento, pode ser desejável agregar em todos os fragmentos. É possível fazer isso tendo uma arquitetura de pipeline em que a saída de cada operador é alimentada em outro fluxo com um único fragmento, o qual é processado por um operador que agrega as saídas do primeiro estágio. Como os dados do primeiro estágio são limitados (um exemplo por minuto por fragmento), eles podem ser facilmente tratados por um fragmento.

Escalar o processamento

Quando o fluxo é expandido para ter muitos fragmentos (porque muitos produtores estão enviando dados), a maneira de escalar o processamento é adicionando mais operadores. É possível executar os operadores em instâncias do Amazon EC2 e usar grupos do Auto Scaling.

Use conectores para o Amazon S3/ DynamoDB/Amazon Redshift/Storm

Como um fluxo é processado continuamente, sua saída pode ser enviada para outros destinos. AWS fornece conectores para integrar o Kinesis Data Streams com AWS outros serviços e ferramentas de terceiros.

Próximas etapas

- Para obter mais informações sobre o uso das operações de API do Kinesis Data Streams, consulte Desenvolva produtores usando a API Amazon Kinesis Data Streams com o AWS SDK para Java,

Desenvolva consumidores com produtividade compartilhada com o AWS SDK para Java e Criar e gerenciar fluxos de dados do Kinesis.

- Para obter mais informações sobre a Kinesis Client Library, consulte Desenvolver aplicações de consumo da KCL 1.x.
- Para obter mais informações sobre como otimizar seu aplicativo, consulte Otimizar consumidores do Amazon Kinesis Data Streams.

Tutorial: analise dados do mercado de ações em tempo real usando o Amazon Managed Service for Apache Flink

O cenário deste tutorial envolve consumir negociações do mercado de ações em um fluxo de dados e criar uma aplicação simples do Amazon Managed Service for Apache Flink para realizar cálculos no fluxo. Será explicado como enviar um fluxo de registros para o Kinesis Data Streams e implementar uma aplicação que consome e processa os registros em tempo quase real.

Com o Amazon Managed Service for Apache Flink, você pode usar Java, Scala ou SQL para processar e analisar dados do fluxo. O serviço permite que você crie e execute código Java ou Scala em comparação com origens de fluxos para fazer analytics de séries temporais, alimentar painéis e criar métricas em tempo real.

É possível criar aplicações Flink no Managed Service for Apache Flink usando bibliotecas de código aberto baseadas no Apache Flink. O Apache Flink é uma estrutura popular e um mecanismo para o processamento de fluxos de dados.

> **Important**
>
> Depois de criar dois fluxos de dados e um aplicativo, sua conta incorre em cobranças nominais pelo Kinesis Data Streams e pelo Managed Service for Apache Flink porque eles não estão qualificados para o nível gratuito. AWS Quando você terminar de usar esse aplicativo, exclua seus AWS recursos para parar de incorrer em cobranças.

O código não acessa os dados reais da bolsa de valores, ele simula o fluxo de negociações de ações. Isso é feito com o uso de um gerador de negociações de ações aleatórias. Se houver acesso a um fluxo de negociações de ações em tempo real, pode ser interessante derivar estatísticas úteis e em tempo hábil desse fluxo. Por exemplo, talvez convenha executar uma análise de janela

deslizante na qual se determine a ação mais popular que foi adquirida nos últimos 5 minutos. Ou talvez convenha uma notificação sempre que uma ordem de venda for muito grande (ou seja, tenha muitas quotas). É possível estender o código nesta série para oferecer essa funcionalidade.

Os exemplos mostrados usam a região Oeste dos EUA (Oregon), mas funcionam em qualquer região da AWS que oferece suporte ao Managed Service for Apache Flink.

Tarefas

- Pré-requisitos para concluir os exercícios
- Configurar uma AWS conta e criar um usuário administrador
- Configure o AWS Command Line Interface (AWS CLI)
- Criar e executar uma aplicação do Managed Service for Apache Flink

Pré-requisitos para concluir os exercícios

Para concluir as etapas neste guia, é necessário ter o seguinte:

- Java Development Kit (JDK) versão 8. Defina a variável do ambiente`JAVA_HOME` para apontar para o local de instalação do JDK.
- Recomenda-se o uso de um ambiente de desenvolvimento (como Eclipse Java Neon ou IntelliJ Idea) para desenvolver e compilar seu aplicativo.
- Cliente do Git. Instale o cliente do Git, se isso ainda não foi feito.
- Apache Maven Compiler Plugin. Maven deve estar em seu caminho de trabalho. Para testar a instalação do Apache Maven, insira o seguinte:

```
$ mvn -version
```

Para começar a usar, acesse Configurar uma AWS conta e criar um usuário administrador.

Configurar uma AWS conta e criar um usuário administrador

Antes de usar o Amazon Managed Service for Apache Flink pela primeira vez, conclua as seguintes tarefas:

1. Inscreva-se para AWS
2. Criar um usuário do IAM

Inscreva-se para AWS

Quando você se inscreve no Amazon Web Services (AWS), sua AWS conta é automaticamente cadastrada em todos os serviços AWS, incluindo o Amazon Managed Service para Apache Flink. A cobrança incorrerá apenas pelos serviços utilizados.

Com o Managed Service for Apache Flink, pague apenas pelos recursos usados. Novos clientes da AWS podem começar a usar o Managed Service for Apache Flink gratuitamente. Para obter mais informações, consulte Nível gratuito da AWS.

Se você já tiver uma AWS conta, vá para a próxima tarefa. Se não tiver uma conta da AWS , siga as etapas a seguir para criar uma.

Para criar uma AWS conta

1. Abra a https://portal.aws.amazon.com/billing/inscrição.
2. Siga as instruções online.

   Parte do procedimento de inscrição envolve receber uma chamada telefônica ou uma mensagem de texto e inserir um código de verificação pelo teclado do telefone.

   Quando você se inscreve em um Conta da AWS, um Usuário raiz da conta da AWSé criado. O usuário-raiz tem acesso a todos os Serviços da AWS e recursos na conta. Como prática recomendada de segurança, atribua o acesso administrativo a um usuário e use somente o usuário-raiz para executar tarefas que exigem acesso de usuário-raiz.

Anote o ID da sua AWS conta porque você precisará dele para a próxima tarefa.

Criar um usuário do IAM

Serviços em AWS, como o Amazon Managed Service para Apache Flink, exigem que você forneça credenciais ao acessá-los. Dessa maneira, o serviço pode determinar se há permissões para acessar os recursos próprios desse serviço. O Console de gerenciamento da AWS exige que você digite sua senha.

Você pode criar chaves de acesso para sua AWS conta para acessar o AWS Command Line Interface (AWS CLI) ou a API. No entanto, não recomendamos que você acesse AWS usando as credenciais da sua AWS conta. Em vez disso, recomendamos que você use AWS Identity and Access Management (IAM). Crie um usuário do IAM, adicione o usuário a um grupo do IAM com

permissões administrativas e, em seguida, conceda permissões administrativas ao usuário do IAM criado. Em seguida, pode-se acessar a AWS usando uma URL especial e as credenciais desse usuário do IAM.

Se você se inscreveu AWS, mas não criou um usuário do IAM para si mesmo, você pode criar um usando o console do IAM.

Os exercícios de conceitos básicos deste guia pressupõem a existência de um usuário (`adminuser`) com permissões de administrador. Siga o procedimento para criar `adminuser` na conta.

Para criar um grupo de administradores

1. Faça login no Console de gerenciamento da AWS e abra o console do IAM em https://console.aws.amazon.com/iam/.
2. No painel de navegação, escolha Grupos e Criar novo grupo.
3. Em Nome do grupo, digite um nome para o grupo, como **`Administrators`** e escolha Próxima etapa.
4. Na lista de políticas, marque a caixa de seleção ao lado da AdministratorAccesspolítica. É possível usar o menu Filtro e a caixa Pesquisar para filtrar a lista de políticas.
5. Selecione Próximo passo e, em seguida, Criar grupo.

O grupo novo é listado em Nome do grupo.

Para criar um usuário do IAM para si mesmo, adicioná-lo ao grupo de administradores e criar uma senha

1. No painel de navegação, escolha Usuários e depois Adicionar usuário.
2. Na caixa Nome de usuário, insira um nome de usuário.
3. Escolha tanto Acesso programático como Acesso ao console de Gerenciamento da AWS .
4. Escolha Próximo: Permissões.
5. Marque a caixa de seleção ao lado do grupo Administradores. Então, escolha Próximo: Análise.
6. Selecione Criar usuário.

Como fazer login como o novo usuário do IAM

1. Saia do Console de gerenciamento da AWS.

2. Use o seguinte formato de URL para fazer login no console:

   `https://aws_account_number.signin.aws.amazon.com/console/`

   *aws_account_number*É o ID da sua AWS conta sem hífens. Por exemplo, se o ID da sua AWS conta for 1234-5678-9012, substitua por. *aws_account_number* **123456789012** Para obter informações sobre como encontrar o número da sua conta, consulte Seu ID de AWS conta e seu alias no Guia do usuário do IAM.
3. Insira o nome e a senha de usuário do IAM que você acabou de criar. Quando você está conectado, a barra de navegação exibe *your_user_name* @*your_aws_account_id*.

> **Note**
>
> Se você não quiser que o URL da sua página de login contenha o ID da sua AWS conta, você pode criar um alias de conta.

Para criar ou remover um alias de conta

1. Abra o console do IAM em https://console.aws.amazon.com/iam/.
2. No painel de navegação, selecione Painel.
3. Encontre o link de login dos usuários do IAM.
4. Para criar o alias, escolha Personalizar. Insira o nome que deseja usar para o alias e escolha Sim, criar.
5. Para remover o alias, selecione Personalizar e, em seguida, selecione Sim, excluir. O URL de login é revertido para o uso do ID da sua AWS conta.

Para fazer o login depois de criar o alias de uma conta, use o seguinte URL:

`https://your_account_alias.signin.aws.amazon.com/console/`

Para verificar o link de cadastro para usuários do IAM para a conta, abra o console do IAM e marque IAM users sign-in link no painel.

Para obter mais informações sobre IAM, consulte o seguinte:

- AWS Identity and Access Management (IAM)

- Introdução ao IAM
- Guia do usuário do IAM

Próxima etapa

Configure o AWS Command Line Interface (AWS CLI)

Configure o AWS Command Line Interface (AWS CLI)

Nesta etapa, você baixa e configura o AWS CLI para uso com o Amazon Managed Service para Apache Flink.

> **Note**
>
> Os exercícios de conceitos básicos neste guia pressupõem o uso de credenciais de administrador (`adminuser`) em sua conta para executar as operações.

> **Note**
>
> Se você já tem o AWS CLI instalado, talvez seja necessário fazer o upgrade para obter a funcionalidade mais recente. Para obter mais informações, consulte Instalando a interface de linha de AWS comando no Guia AWS Command Line Interface do usuário. Para verificar a versão do AWS CLI, execute o seguinte comando:
>
> ```
> aws --version
> ```
>
> Os exercícios deste tutorial exigem a seguinte AWS CLI versão ou posterior:
>
> ```
> aws-cli/1.16.63
> ```

Para configurar o AWS CLI

1. Faça download e configure a AWS CLI. Para obter instruções, consulte os seguintes tópicos no Guia do usuário do AWS Command Line Interface :
   - Instalar a AWS Command Line Interface

- Configurando a AWS CLI

2. Adicione um perfil nomeado para o usuário administrador no arquivo de AWS CLI configuração. Você usa esse perfil ao executar os AWS CLI comandos. Para obter mais informações sobre perfis nomeados, consulte Perfis nomeados no Guia do usuário da AWS Command Line Interface .

```
[profile adminuser]
aws_access_key_id = adminuser access key ID
aws_secret_access_key = adminuser secret access key
region = aws-region
```

   Para obter uma lista das AWS regiões disponíveis, consulte AWS Regiões e endpoints no Referência geral da Amazon Web Services.

3. Verifique a configuração digitando o seguinte comando no prompt de comando:

```
aws help
```

Depois de configurar uma AWS conta e a AWS CLI, você pode tentar o próximo exercício, no qual você configura um aplicativo de amostra e testa a end-to-end configuração.

Próxima etapa

Criar e executar uma aplicação do Managed Service for Apache Flink

Criar e executar uma aplicação do Managed Service for Apache Flink

Neste exercício, será criado um aplicativo Managed Service for Apache Flink com fluxos de dados como origem e coletor.

Esta seção contém as seguintes etapas:

- Criar dois fluxos de dados do Amazon Kinesis
- Gravação de registros de amostra no fluxo de entrada
- Baixar e examinar o código Java de fluxo do Apache Flink
- Compilar o código da aplicação
- Upload do código Java de fluxo do Apache Flink
- Criar e executar a aplicação do Managed Service for Apache Flink

Criar dois fluxos de dados do Amazon Kinesis

Antes de criar uma aplicação do Amazon Managed Service for Apache Flink para este exercício, crie dois fluxos de dados do Kinesis (`ExampleInputStream` e `ExampleOutputStream`). O aplicativo usa esses fluxos para os fluxos de origem e de destino do aplicativo.

É possível criar esses fluxos usando o console do Amazon Kinesis ou o comando da AWS CLI a seguir. Para instruções do console, consulte Criar e atualizar fluxos de dados.

Como criar os fluxos de dados (AWS CLI)

1. Para criar o primeiro stream (`ExampleInputStream`), use o seguinte comando do Amazon Kinesis `create-stream` AWS CLI .

```
$ aws kinesis create-stream \
--stream-name ExampleInputStream \
--shard-count 1 \
--region us-west-2 \
--profile adminuser
```

2. Para criar o segundo fluxo que o aplicativo usa para gravar a saída, execute o mesmo comando, alterando o nome da transmissão para `ExampleOutputStream`.

```
$ aws kinesis create-stream \
--stream-name ExampleOutputStream \
--shard-count 1 \
--region us-west-2 \
--profile adminuser
```

Gravação de registros de amostra no fluxo de entrada

Nesta seção, será usado um script Python para gravar registros de amostra no fluxo para o aplicativo processar.

> **Note**
>
> Essa seção requer AWS SDK para Python (Boto).

1. Crie um arquivo denominado `stock.py` com o conteúdo a seguir:

```
import datetime
import json
import random
import boto3

STREAM_NAME = "ExampleInputStream"


def get_data():
    return {
        "EVENT_TIME": datetime.datetime.now().isoformat(),
        "TICKER": random.choice(["AAPL", "AMZN", "MSFT", "INTC", "TBV"]),
        "PRICE": round(random.random() * 100, 2),
    }


def generate(stream_name, kinesis_client):
    while True:
        data = get_data()
        print(data)
        kinesis_client.put_record(
            StreamName=stream_name, Data=json.dumps(data),
 PartitionKey="partitionkey"
        )


if __name__ == "__main__":
    generate(STREAM_NAME, boto3.client("kinesis"))
```

2. Mais adiante neste tutorial, será executado o script `stock.py` para enviar dados para o aplicativo.

```
$ python stock.py
```

Baixar e examinar o código Java de fluxo do Apache Flink

O código do aplicativo Java para esses exemplos está disponível em GitHub. Para fazer download do código do aplicativo, faça o seguinte:

1. Duplique o repositório remoto com o seguinte comando:

```
git clone https://github.com/aws-samples/amazon-kinesis-data-analytics-java-
examples.git
```

2. Navegue até o diretório `GettingStarted`.

O código do aplicativo está localizado nos arquivos `CustomSinkStreamingJob.java` e `CloudWatchLogSink.java`. Observe o seguinte sobre o código do aplicativo:

- A aplicação usa uma origem do Kinesis para ler o fluxo de origem. O trecho a seguir cria o coletor do Kinesis:

```
return env.addSource(new FlinkKinesisConsumer<>(inputStreamName,
                new SimpleStringSchema(), inputProperties));
```

Compilar o código da aplicação

Nesta seção, será usado o compilador do Apache Maven para criar o código Java para o aplicativo. Para obter informações sobre como instalar o Apache Maven e o Java Development Kit (JDK), consulte Pré-requisitos para concluir os exercícios.

Seu aplicativo Java requer os seguintes componentes:

- Um arquivo Project Object Model (pom.xml). este arquivo contém informações sobre a configuração e as dependências da aplicação, incluindo as bibliotecas do Amazon Managed Service for Apache Flink.
- Um método `main` que contém a lógica do aplicativo.

> **Note**
>
> Para usar o conector do Kinesis para a aplicação a seguir, você precisa baixar o código-fonte do conector e compilá-lo como descrito na documentação do Apache Flink.

Como criar e compilar o código do aplicativo

1. Crie um Java/Maven aplicativo em seu ambiente de desenvolvimento. Para obter informações sobre como criar um aplicativo, consulte a documentação do seu ambiente de desenvolvimento:

- Como criar seu primeiro projeto Java (Eclipse Java Neon)
- Como criar, executar e empacotar seu primeiro aplicativo Java (IntelliJ Idea)

2. Use o código a seguir para um arquivo chamado `StreamingJob.java`.

```
package com.amazonaws.services.kinesisanalytics;

import com.amazonaws.services.kinesisanalytics.runtime.KinesisAnalyticsRuntime;
import org.apache.flink.api.common.serialization.SimpleStringSchema;
import org.apache.flink.streaming.api.datastream.DataStream;
import org.apache.flink.streaming.api.environment.StreamExecutionEnvironment;
import org.apache.flink.streaming.connectors.kinesis.FlinkKinesisConsumer;
import org.apache.flink.streaming.connectors.kinesis.FlinkKinesisProducer;
import
 org.apache.flink.streaming.connectors.kinesis.config.ConsumerConfigConstants;

import java.io.IOException;
import java.util.Map;
import java.util.Properties;

public class StreamingJob {

    private static final String region = "us-east-1";
    private static final String inputStreamName = "ExampleInputStream";
    private static final String outputStreamName = "ExampleOutputStream";

    private static DataStream<String>
 createSourceFromStaticConfig(StreamExecutionEnvironment env) {
        Properties inputProperties = new Properties();
        inputProperties.setProperty(ConsumerConfigConstants.AWS_REGION, region);

 inputProperties.setProperty(ConsumerConfigConstants.STREAM_INITIAL_POSITION,
 "LATEST");

        return env.addSource(new FlinkKinesisConsumer<>(inputStreamName, new
 SimpleStringSchema(), inputProperties));
    }

    private static DataStream<String>
 createSourceFromApplicationProperties(StreamExecutionEnvironment env)
            throws IOException {
```

```
        Map<String, Properties> applicationProperties =
KinesisAnalyticsRuntime.getApplicationProperties();
        return env.addSource(new FlinkKinesisConsumer<>(inputStreamName, new
SimpleStringSchema(),
                applicationProperties.get("ConsumerConfigProperties")));
    }

    private static FlinkKinesisProducer<String> createSinkFromStaticConfig() {
        Properties outputProperties = new Properties();
        outputProperties.setProperty(ConsumerConfigConstants.AWS_REGION, region);
        outputProperties.setProperty("AggregationEnabled", "false");

        FlinkKinesisProducer<String> sink = new FlinkKinesisProducer<>(new
SimpleStringSchema(), outputProperties);
        sink.setDefaultStream(outputStreamName);
        sink.setDefaultPartition("0");
        return sink;
    }

    private static FlinkKinesisProducer<String>
createSinkFromApplicationProperties() throws IOException {
        Map<String, Properties> applicationProperties =
KinesisAnalyticsRuntime.getApplicationProperties();
        FlinkKinesisProducer<String> sink = new FlinkKinesisProducer<>(new
SimpleStringSchema(),
                applicationProperties.get("ProducerConfigProperties"));

        sink.setDefaultStream(outputStreamName);
        sink.setDefaultPartition("0");
        return sink;
    }

    public static void main(String[] args) throws Exception {
        // set up the streaming execution environment
        final StreamExecutionEnvironment env =
StreamExecutionEnvironment.getExecutionEnvironment();

        /*
         * if you would like to use runtime configuration properties, uncomment the
         * lines below
         * DataStream<String> input = createSourceFromApplicationProperties(env);
         */

        DataStream<String> input = createSourceFromStaticConfig(env);
```

```
        /*
         * if you would like to use runtime configuration properties, uncomment the
         * lines below
         * input.addSink(createSinkFromApplicationProperties())
         */

        input.addSink(createSinkFromStaticConfig());

        env.execute("Flink Streaming Java API Skeleton");
    }
}
```

Observe o seguinte sobre o exemplo de código anterior:

- Este arquivo contém o método `main` que define a funcionalidade do aplicativo.
- Seu aplicativo cria conectores de origem e de destino para acessar recursos externos usando um objeto `StreamExecutionEnvironment`.
- O aplicativo cria conectores de origem e de destino usando propriedades estáticas. Para usar as propriedades dinâmicas do aplicativo, use os métodos `createSourceFromApplicationProperties` e `createSinkFromApplicationProperties` para criar os conectores. Esses métodos leem as propriedades do aplicativo para configurar os conectores.

3. Para usar o seu código de aplicativo, compile-o e empacote-o em um arquivo JAR. Há duas formas de compilar e empacotar o código:

   - Use a ferramenta de linha de comando do Maven. Crie seu arquivo JAR executando o seguinte comando no diretório que contém o arquivo `pom.xml`:

     ```
     mvn package
     ```

   - Use o ambiente de desenvolvimento. Consulte a documentação de seu ambiente de desenvolvimento para obter mais detalhes.

   É possível carregar o pacote como um arquivo JAR, ou pode compactar o pacote e carregá-lo como um arquivo ZIP. Se você criar seu aplicativo usando o AWS CLI, especifique o tipo de conteúdo do código (JAR ou ZIP).

4. Se houver erros durante a compilação, verifique se sua variável de ambiente `JAVA_HOME` está definida corretamente.

Se o aplicativo for compilado com êxito, o arquivo a seguir é criado:

```
target/java-getting-started-1.0.jar
```

Upload do código Java de fluxo do Apache Flink

Nesta seção, será criado um bucket do Amazon Simple Storage Service (Amazon S3) e realizado o upload do código do aplicativo.

Para fazer upload do código do aplicativo

1. Abra o console do Amazon S3 em https://console.aws.amazon.com/s3/.
2. Selecione Criar bucket.
3. Insira **`ka-app-code-<username>`** no campo Nome do bucket. Adicione um sufixo para o nome do bucket, como o nome do usuário, para torná-lo globalmente exclusivo. Selecione Next (Próximo).
4. Na etapa Configurar opções, mantenha as configurações como estão e selecione Próximo.
5. Na etapa Definir permissões, mantenha as configurações como estão e selecione Próximo.
6. Selecione Criar bucket.
7. No console do Amazon S3, escolha o ka-app-code- ***<username>*** bucket e escolha Upload.
8. Na etapa Selecionar arquivos, selecione Adicionar arquivos. Navegue até o arquivo `java-getting-started-1.0.jar`, criado na etapa anterior. Escolha Próximo.
9. Na etapa Definir permissões, mantenha as configurações como estão. Escolha Próximo.
10. Na etapa Definir propriedades, mantenha as configurações como estão. Escolha Carregar.

O código passa a ser armazenado em um bucket do Amazon S3 que pode ser acessado pelo aplicativo.

Criar e executar a aplicação do Managed Service for Apache Flink

É possível criar e executar um aplicativo Managed Service for Apache Flink usando o console ou a AWS CLI.

> **Note**
>
> Quando você cria o aplicativo usando o console, seus recursos AWS Identity and Access Management (IAM) e do Amazon CloudWatch Logs são criados para você. Ao criar o aplicativo usando o AWS CLI, você cria esses recursos separadamente.

Tópicos

- Criar e executar a aplicação (console)
- Criar e executar a aplicação (AWS CLI)

Criar e executar a aplicação (console)

Siga estas etapas para criar, configurar, atualizar e executar o aplicativo usando o console.

Criar a aplicação

1. Abra o console do Kinesis em https://console.aws.amazon.com /kinesis.
2. No painel do Amazon Kinesis, escolha Criar aplicativo de analytics.
3. Na página Kinesis Analytics – Criar aplicativo, forneça os detalhes do aplicativo da seguinte forma:
   - Em Nome do aplicativo, insira **`MyApplication`**.
   - Em Descrição, insira **`My java test app`**.
   - Em Runtime, escolha Apache Flink 1.6.
4. Em Permissões de acesso, escolha Criar/atualizar o perfil do IAM **`kinesis-analytics-MyApplication-us-west-2`**.
5. Selecione Criar aplicativo.

> **Note**
>
> Ao criar uma aplicação do Amazon Managed Service for Apache Flink usando o console, você tem a opção de ter um perfil do IAM e uma política criados para sua aplicação. O aplicativo usa essa função e política para acessar os recursos dependentes. Esses recursos do IAM são nomeados usando o nome do aplicativo e a região da seguinte forma:

- Política: `kinesis-analytics-service-`*`MyApplication`*`-`*`us-west-2`*
- Função: `kinesis-analytics-`*`MyApplication`*`-`*`us-west-2`*

Editar a política do IAM

Edite a política do IAM para adicionar permissões de acesso aos fluxos de dados do Kinesis.

1. Abra o console do IAM em https://console.aws.amazon.com/iam/.
2. Selecione Políticas. Selecione a política **`kinesis-analytics-service-MyApplication-us-west-2`** que o console criou na seção anterior.
3. Na página Resumo, selecione Editar política. Selecione a guia JSON.
4. Adicione a seção destacada do exemplo de política a seguir à política. Substitua a conta de amostra IDs (*`012345678901`*) pelo ID da sua conta.

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "ReadCode",
            "Effect": "Allow",
            "Action": [
                "s3:GetObject",
                "s3:GetObjectVersion"
            ],
            "Resource": [
                "arn:aws:s3:::ka-app-code-username/java-getting-
started-1.0.jar"
            ]
        },
        {
            "Sid": "ListCloudwatchLogGroups",
            "Effect": "Allow",
            "Action": [
                "logs:DescribeLogGroups"
            ],
            "Resource": [
```

```
                "arn:aws:logs:us-west-2:012345678901:log-group:*"
            ]
        },
        {
            "Sid": "ListCloudwatchLogStreams",
            "Effect": "Allow",
            "Action": [
                "logs:DescribeLogStreams"
            ],
            "Resource": [
                "arn:aws:logs:us-west-2:012345678901:log-group:/aws/kinesis-
analytics/MyApplication:log-stream:*"
            ]
        },
        {
            "Sid": "PutCloudwatchLogs",
            "Effect": "Allow",
            "Action": [
                "logs:PutLogEvents"
            ],
            "Resource": [
                "arn:aws:logs:us-west-2:012345678901:log-group:/aws/kinesis-
analytics/MyApplication:log-stream:kinesis-analytics-log-stream"
            ]
        },
        {
            "Sid": "ReadInputStream",
            "Effect": "Allow",
            "Action": "kinesis:*",
            "Resource": "arn:aws:kinesis:us-west-2:012345678901:stream/
ExampleInputStream"
        },
        {
            "Sid": "WriteOutputStream",
            "Effect": "Allow",
            "Action": "kinesis:*",
            "Resource": "arn:aws:kinesis:us-west-2:012345678901:stream/
ExampleOutputStream"
        }
    ]
}
```

Configure o aplicativo

1. Na MyApplicationpágina, escolha Configurar.
2. Na página Configurar aplicativo, forneça o Local do código:
   - Em Bucket do Amazon S3, insira **ka-app-code-*<username>***.
   - Em Caminho do objeto do Amazon S3, insira **java-getting-started-1.0.jar**.
3. Na seção Acesso aos recursos do aplicativo, em Permissões de acesso, selecione Criar/atualizar o perfil do IAM **kinesis-analytics-MyApplication-us-west-2**.
4. Em Propriedades, ID do grupo, insira **ProducerConfigProperties**.
5. Insira as seguintes propriedades e valores de aplicativo:

Chave	Valor
flink.inputstream.initpos	**LATEST**
aws:region	**us-west-2**
AggregationEnabled	**false**

6. Em Monitoramento, confirme se Nível de monitoramento de métricas está definido como Aplicativo.
7. Para CloudWatch registrar, marque a caixa de seleção Ativar.
8. Selecione Atualizar.

> **Note**
>
> Quando você opta por ativar o CloudWatch registro, o Managed Service for Apache Flink cria um grupo de registros e um fluxo de registros para você. Os nomes desses recursos são os seguintes:
>
> - Grupo de logs: `/aws/kinesis-analytics/MyApplication`
> - Fluxo de logs: `kinesis-analytics-log-stream`

Execute o aplicativo

1. Na MyApplicationpágina, escolha Executar. Confirme a ação.
2. Quando o aplicativo estiver em execução, atualize a página. O console mostra o Gráfico do aplicativo.

Interromper a aplicação

Na MyApplicationpágina, escolha Parar. Confirme a ação.

Atualizar o aplicativo

Usando o console, é possível atualizar configurações do aplicativo, como as propriedades do aplicativo, as configurações de monitoramento e a localização ou o nome do arquivo JAR do aplicativo. Também é possível recarregar o JAR do aplicativo do bucket do Amazon S3 se for necessário atualizar o código do aplicativo.

Na MyApplicationpágina, escolha Configurar. Atualize as configurações do aplicativo e selecione Atualizar.

Criar e executar a aplicação (AWS CLI)

Nesta seção, você usa o AWS CLI para criar e executar o aplicativo Managed Service for Apache Flink. O Managed Service for Apache Flink usa o `kinesisanalyticsv2` AWS CLI comando para criar e interagir com o Managed Service for Apache Flink aplicativos.

Criar uma política de permissões

Primeiro, crie uma política de permissões com duas instruções: uma que concede permissões para a ação `read` no fluxo de origem, e outra que concede permissões para ações `write` no fluxo de destino. Em seguida, anexe a política a um perfil do IAM (que será criado na próxima seção). Assim, ao assumir o perfil, o serviço Managed Service for Apache Flink terá as permissões necessárias para ler o fluxo de origem e gravar no fluxo de coleta.

Use o código a seguir para criar a política de permissões `KAReadSourceStreamWriteSinkStream`. Substitua *`username`* pelo nome de usuário usado para criar o bucket do Amazon S3 e armazenar o código do aplicativo. Substitua o ID da conta nos nomes de recursos da Amazon (ARNs) (*`012345678901`*) pelo ID da sua conta.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "S3",
            "Effect": "Allow",
            "Action": [
                "s3:GetObject",
                "s3:GetObjectVersion"
            ],
            "Resource": ["arn:aws:s3:::ka-app-code-username",
                "arn:aws:s3:::ka-app-code-username/*"
            ]
        },
        {
            "Sid": "ReadInputStream",
            "Effect": "Allow",
            "Action": "kinesis:*",
            "Resource": "arn:aws:kinesis:us-west-2:012345678901:stream/
ExampleInputStream"
        },
        {
            "Sid": "WriteOutputStream",
            "Effect": "Allow",
            "Action": "kinesis:*",
            "Resource": "arn:aws:kinesis:us-west-2:012345678901:stream/
ExampleOutputStream"
        }
    ]
}
```

Para step-by-step obter instruções sobre como criar uma política de permissões, consulte Tutorial: Criar e anexar sua primeira política gerenciada pelo cliente no Guia do usuário do IAM.

> **Note**
>
> Para acessar outros AWS serviços, você pode usar AWS SDK para Java o. O Managed Service for Apache Flink define automaticamente as credenciais exigidas pelo SDK como as

credenciais do perfil do IAM associado a seu aplicativo. Não é necessária nenhuma etapa adicional.

Criar um perfil do IAM

Nesta seção, você cria um perfil do IAM que o Managed Service for Apache Flink pode assumir para ler um fluxo de origem e gravar no fluxo de coleta.

O Managed Service for Apache Flink não pode acessar seu fluxo sem permissões. Essas permissões são concedidas usando um perfil do IAM. Cada perfil do IAM tem duas políticas anexadas. A política de confiança concede ao Managed Service for Apache Flink permissão para assumir o perfil, e a política de permissões determina o que o serviço pode fazer depois de assumir a função.

Anexe a política de permissões que criou na seção anterior a essa função.

Para criar uma perfil do IAM

1. Abra o console do IAM em https://console.aws.amazon.com/iam/.
2. No painel de navegação, selecione Funções e Criar função.
3. Em Selecionar tipo de identidade de confiança, selecione Serviço da AWS . Em Selecionar o serviço que usará esta função, selecione Kinesis. Em Selecionar seu caso de uso, selecione Kinesis Analytics.

   Selecione Next: Permissions (Próximo: permissões).
4. Na página Attach permissions policies, selecione Next: Review. É possível anexar políticas de permissões depois de criar a função.
5. Na página Criar função, insira **`KA-stream-rw-role`** para o Nome da função. Selecione Criar função.

   Foi criado um perfil do IAM chamado `KA-stream-rw-role`. Em seguida, atualize as políticas de confiança e de permissões para a função.
6. Anexe a política de permissões à função.

> **Note**
>
> Para este exercício, o Managed Service for Apache Flink assume esse perfil para ler dados de um fluxo de dados do Kinesis (origem) e gravar a saída em outro fluxo de

dados do Kinesis. Depois, anexe a política criada na etapa anterior, the section called "Criar uma política de permissões".

a. Na página Resumo, selecione a guia Permissões.
b. Selecione Attach Policies.
c. Na caixa de pesquisa, insira **KAReadSourceStreamWriteSinkStream** (a política criada na seção anterior).
d. Selecione a política KAReadInputStreamWriteOutputStream e selecione Anexar política.

Agora você criou a função de execução de serviço que seu aplicativo usa para acessar os recursos. Anote o ARN da nova função.

Para step-by-step obter instruções sobre como criar uma função, consulte Como criar uma função do IAM (console) no Guia do usuário do IAM.

Criar o aplicativo do Managed Service for Apache Flink

1. Salve o seguinte código JSON em um arquivo chamado `create_request.json`. Substitua o ARN da função de amostra pelo ARN da função criada anteriormente. Substitua o sufixo do ARN do bucket (*username*) pelo sufixo selecionado na seção anterior. Substitua o ID da conta de exemplo (*012345678901*) na função de execução do serviço pelo ID da conta.

```
{
    "ApplicationName": "test",
    "ApplicationDescription": "my java test app",
    "RuntimeEnvironment": "FLINK-1_6",
    "ServiceExecutionRole": "arn:aws:iam::012345678901:role/KA-stream-rw-role",
    "ApplicationConfiguration": {
        "ApplicationCodeConfiguration": {
            "CodeContent": {
                "S3ContentLocation": {
                    "BucketARN": "arn:aws:s3:::ka-app-code-username",
                    "FileKey": "java-getting-started-1.0.jar"
                }
            },
            "CodeContentType": "ZIPFILE"
        },
        "EnvironmentProperties":  {
         "PropertyGroups": [
```

```
            {
               "PropertyGroupId": "ProducerConfigProperties",
               "PropertyMap" : {
                    "flink.stream.initpos" : "LATEST",
                    "aws.region" : "us-west-2",
                    "AggregationEnabled" : "false"
               }
            },
            {
               "PropertyGroupId": "ConsumerConfigProperties",
               "PropertyMap" : {
                    "aws.region" : "us-west-2"
               }
            }
         ]
      }
   }
}
```

2. Execute a ação `CreateApplication` com a solicitação anterior para criar o aplicativo:

```
aws kinesisanalyticsv2 create-application --cli-input-json file://
create_request.json
```

O aplicativo agora é criado. Inicie o aplicativo na próxima etapa.

Iniciar o aplicativo

Nesta seção, a ação `StartApplication` será usada para iniciar o aplicativo.

Para iniciar o aplicativo

1. Salve o seguinte código JSON em um arquivo chamado `start_request.json`.

```
{
    "ApplicationName": "test",
    "RunConfiguration": {
        "ApplicationRestoreConfiguration": {
         "ApplicationRestoreType": "RESTORE_FROM_LATEST_SNAPSHOT"
         }
    }
```

```
}
```

2. Execute a ação `StartApplication` com a solicitação anterior para iniciar o aplicativo:

```
aws kinesisanalyticsv2 start-application --cli-input-json file://start_request.json
```

O aplicativo agora está em execução. Você pode verificar as métricas do Managed Service for Apache Flink no CloudWatch console da Amazon para verificar se o aplicativo está funcionando.

Interromper o aplicativo

Nesta seção, a ação `StopApplication` será usada para interromper o aplicativo.

Como interromper o aplicativo

1. Salve o seguinte código JSON em um arquivo chamado `stop_request.json`.

```
{"ApplicationName": "test"
}
```

2. Execute a ação `StopApplication` com a seguinte solicitação para interromper o aplicativo:

```
aws kinesisanalyticsv2 stop-application --cli-input-json file://stop_request.json
```

O aplicativo agora está interrompido.

Tutorial: Use AWS Lambda com o Amazon Kinesis Data Streams

Neste tutorial, uma função do Lambda será criada para consumir eventos de um fluxo de dados do Kinesis. Neste cenário de exemplo, um aplicativo personalizado grava registros em um stream de dados do Kinesis. AWS Lambda em seguida, pesquisa esse fluxo de dados e, quando detecta novos registros de dados, invoca sua função Lambda. AWS Lambda em seguida, executa a função Lambda assumindo a função de execução que você especificou ao criar a função Lambda.

Para obter instruções detalhadas passo a passo, consulte Tutorial: Usando o AWS Lambda com o Amazon Kinesis.

> **Note**
>
> Este tutorial presume algum conhecimento de operações básicas do Lambda e do console do AWS Lambda . Se ainda não o fez, siga as instruções em Introdução ao AWS Lambda para criar sua primeira função do Lambda.

Use a solução AWS de streaming de dados para o Amazon Kinesis

A solução AWS de dados de streaming para Amazon Kinesis configura automaticamente os AWS serviços necessários para capturar, armazenar, processar e entregar dados de streaming com facilidade. A solução oferece várias opções para resolver casos de uso de dados de streaming que usam vários AWS serviços, incluindo Kinesis Data AWS Lambda Streams, Amazon API Gateway e Amazon Managed Service para Apache Flink.

Cada solução inclui os seguintes componentes:

- Um CloudFormation pacote para implantar o exemplo completo.
- Um CloudWatch painel para exibir as métricas do aplicativo.
- CloudWatch alarmes sobre as métricas de aplicação mais relevantes.
- Todos os perfis do IAM e políticas necessários.

A solução pode ser encontrada em Solução de dados de transmissão para o Amazon Kinesis

Criar e gerenciar fluxos de dados do Kinesis

O Amazon Kinesis Data Streams ingere uma grande quantidade de dados em tempo real, armazena os dados de forma durável e os torna disponíveis para consumo. Um registro de dados é a unidade de dados armazenada pelo Kinesis Data Streams. Um fluxo de dados representa um grupo de registros de dados. Os registros de dados em um fluxo de dados são distribuídos em fragmentos.

Um fragmento tem uma sequência de registros de dados em um fluxo. Ele serve como uma unidade base de throughput em um fluxo de dados do Kinesis. Um fragmento suporta 1 MB/s e 1.000 registros por segundo para gravações e 2 MB/s para leituras nos modos de capacidade sob demanda e provisionada. Os limites de fragmentos garantem prever a performance, facilitando o design e a operação de um fluxo de trabalho de fluxo de dados altamente confiável.

Nesta seção, você aprende como definir o modo de capacidade do stream e como criar um stream usando o Console de gerenciamento da AWS ou APIs. Depois, podem ser realizadas ações adicionais no fluxo.

Tópicos

- Escolha o modo certo para transmitir
- Crie um stream usando o Console de gerenciamento da AWS
- Crie um stream usando o APIs
- Atualizar um fluxo
- Listar fluxos
- Listar fragmentos
- Excluir um fluxo
- Refragmentar um fluxo
- Alterar o período de retenção de dados
- Marcar com tag seus recursos do Amazon Kinesis Data Streams
- Gerenciar registros grandes
- Realize testes de resiliência com AWS Fault Injection Service

Escolha o modo certo para transmitir

Os tópicos a seguir explicam como escolher o melhor modo de capacidade para sua aplicação e como alternar entre os modos, se necessário.

Tópicos

- Quais são os diferentes modos no Kinesis Data Streams?
- Atributos e casos de uso do modo On-demand Standard
- Atributos e casos de uso do modo On-demand Advantage
- Casos de uso e atributos do modo provisionado
- Alternar entre os modos

Quais são os diferentes modos no Kinesis Data Streams?

Um modo determina como a capacidade de um fluxo de dados é gerenciada e como são geradas cobranças pelo uso do fluxo de dados. No Amazon Kinesis Data Streams, você pode escolher On-demand Standard, On-demand Advantage e provisionado como modo para os fluxos de dados.

- On-demand Standard: os fluxos de dados no modo sob demanda não exigem planejamento de capacidade e escalam automaticamente para lidar com gigabytes de throughput de gravação e leitura por minuto. No modo sob demanda, o Kinesis Data Streams gerencia automaticamente os fragmentos para fornecer o throughput necessário.
- On-demand Advantage: um modo de conta que permite mais recursos e fornece uma estrutura de preços mais simples para fluxos sob demanda. Nesse modo, você pode aquecer proativamente a capacidade de throughput de gravação de um stream a qualquer momento. Em termos de preços, não há mais uma cobrança fixa por fluxo, e a ingestão de dados, a recuperação de dados e o uso de retenção estendida em todos os fluxos sob demanda são pelo menos 60% menores do que os do On-demand Standard.
- Provisionado: para fluxos de dados com um modo provisionado, é necessário especificar o número de fragmentos para o fluxo de dados. A capacidade total de um fluxo de dados é a soma das capacidades de seus fragmentos. É possível aumentar ou diminuir o número de fragmentos em um fluxo de dados de acordo com a necessidade.

Você pode usar o Kinesis Data `PutRecords` APIs Streams e gravar `PutRecord` dados em seus streams de dados em qualquer modo. Para recuperar dados, todos os três oferecem suporte aos

consumidores padrão, que usam a API `GetRecords`, e aos consumidores de Distribuição avançada (EFO), que usam a API `SubscribeToShard`.

Todos os recursos do Kinesis Data Streams, incluindo modo de retenção, criptografia, métricas de monitoramento e outros, são compatíveis com os modos sob demanda e provisionado. O Kinesis Data Streams fornece alta durabilidade e disponibilidade nos modos de capacidade sob demanda e provisionada.

Atributos e casos de uso do modo On-demand Standard

Os fluxos de dados no modo sob demanda não exigem planejamento de capacidade e escalam automaticamente para lidar com gigabytes de throughput de gravação e leitura por minuto. O modo sob demanda simplifica a ingestão e o armazenamento de grandes volumes de dados com baixa latência, pois elimina o provisionamento e o gerenciamento de servidores, armazenamento ou throughput. É possível pode ingerir bilhões de registros por dia sem nenhuma sobrecarga operacional.

O modo sob demanda é ideal para atender às necessidades de tráfego de aplicações altamente variável e imprevisível. Não é mais necessário provisionar essas cargas de trabalho na capacidade máxima, o que poderia resultar em custos mais altos devido à baixa utilização. O modo sob demanda é adequado para cargas de trabalho com padrões de tráfego imprevisíveis e altamente variáveis.

Com o modo de capacidade sob demanda, a cobrança é feita por GB de dados gravados e lidos em seus fluxos de dados. Não é necessário especificar o throughput de leitura e gravação que espera que a aplicação execute. O Kinesis Data Streams ajusta-se instantaneamente ao crescimento e à redução das workloads. Para obter mais informações, consulte Definição de preço do Amazon Kinesis Data Streams.

Um fluxo de dados no modo sob demanda acomoda até o dobro do pico de throughput de gravação observado nos 30 dias anteriores. Quando o throughput de gravação do fluxo de dados atinge um novo pico, o Kinesis Data Streams escala automaticamente a capacidade do fluxo de dados. Por exemplo, se seu fluxo de dados tiver uma taxa de transferência de gravação que varia entre 10 MB/s e 40% MB/s, then Kinesis Data Streams ensures that you can easily burst to double your previous peak throughput, or 80 MB/s. If the same data stream sustains a new peak throughput of 50 MB/s, Kinesis Data Streams ensures that there is enough capacity to ingest 100 MB/s da taxa de transferência de gravação. No entanto, poderá ocorrer controle de utilização se o tráfego aumentar para mais que o dobro do pico anterior em um período de 15 minutos. É necessário repetir as solicitações em controle de utilização.

A capacidade de leitura de agregados de um fluxo de dados no modo sob demanda aumenta proporcionalmente com o throughput de gravação. Isso ajuda a garantir que as aplicações de consumo sempre tenham um throughput de leitura adequado para processar os dados recebidos em tempo real. Obtém-se pelo menos o dobro do throughput de gravação em comparação com os dados de leitura usando a API `GetRecords`. Recomenda-se o uso de uma aplicação de consumo com a API `GetRecord`, para permitir espaço suficiente quando a aplicação precisar se recuperar de tempo de inatividade. É recomendável usar o recurso de distribuição avançada do Kinesis Data Streams em cenários que exijam a adição de mais de uma aplicação de consumo. A distribuição avançada permite adicionar até 20 aplicações de consumo a um fluxo de dados usando a API `SubscribeToShard`, com um throughput dedicado para cada aplicação.

Tratar exceções de throughput de leitura e gravação

Com o modo sob demanda (da mesma forma que com o modo de capacidade provisionada), é necessário especificar uma chave de partição com cada registro para gravar dados no fluxo de dados. O Kinesis Data Streams usa suas chaves de partição para distribuir dados entre fragmentos. O Kinesis Data Streams monitora o tráfego de cada fragmento. Quando o tráfego de entrada excede 500 KB/s por fragmento, ele divide o fragmento em 15 minutos. Os valores da chave de hash do fragmento pai são redistribuídos uniformemente entre os fragmentos filho.

Se o tráfego de entrada exceder o dobro do pico anterior, poderão ocorrer exceções de leitura ou gravação por cerca de 15 minutos, mesmo quando os dados forem distribuídos uniformemente entre os fragmentos. Recomenda-se repetir todas essas solicitações para que todos os registros sejam armazenados adequadamente no Kinesis Data Streams.

As exceções de leitura e gravação podem ocorrer ao usar uma chave de partição que causa uma distribuição desigual de dados, e os registros atribuídos a um fragmento específico excedem seus limites. Com o modo sob demanda, o fluxo de dados se adapta automaticamente para lidar com padrões desiguais de distribuição de dados, a menos que uma única chave de partição exceda os limites de 1 MB/s taxa de transferência e 1.000 registros por segundo de um fragmento.

No modo sob demanda, o Kinesis Data Streams divide os fragmentos uniformemente quando detecta um aumento no tráfego. No entanto, ele não detecta nem isola as chaves de hash que estão direcionando uma parte maior do tráfego de entrada para um fragmento específico. Ao usar chaves de partição altamente desiguais, as exceções de gravação poderão continuar ocorrendo. Para esses casos de uso, é recomendável usar o modo de capacidade provisionada que oferece suporte a divisões granulares de fragmentos.

Atributos e casos de uso do modo On-demand Advantage

On-demand Advantage é uma configuração de conta que permite mais recursos e fornece uma estrutura de preços diferentes para todos os fluxos sob demanda na região. Nesse modo, os fluxos sob demanda retêm sua funcionalidade e continuam a escalar automaticamente a capacidade com base no uso real dos dados. Se você quiser aquecer proativamente a capacidade de throughput de gravação de um fluxo, você pode configurar o throughput aquecido. Por exemplo, se seu fluxo de dados tiver uma taxa de transferência de gravação entre 10 MB/s e 40 MB/s, you can expect it to handle up to 80MB/s of instant throughput increases without throttling. However, if you forecast an upcoming event to peak around 200MB/s of traffic, you can configure the stream with a warm throughput of 200MB/s para garantir que a capacidade esteja disponível quando a taxa de transferência de dados chegar. O uso de um throughput aquecido não acarreta custos adicionais.

Outro benefício do modo On-demand Advantage é que os fluxos sob demanda fazem a transição para uma estrutura de preços mais simples. Com o modo habilitado, a conta não verá mais cobranças fixas por fluxo e você lidará apenas com a ingestão de dados, a recuperação de dados e as cobranças opcionais de retenção estendida. Cada dimensão de preço também tem um desconto significativo em comparação com a dimensão correspondente no On-demand Standard. Para obter mais informações, consulte Definição de preço do Amazon Kinesis Data Streams.

As recuperações de dados com distribuição avançada também não têm um preço superior em comparação com as recuperações de dados padrão nesse modo. Além disso, com o modo On-demand Advantage, você pode registrar até 50 consumidores por stream para usar o fan-out aprimorado. A ativação do On-Demand Advantage compromete a conta com pelo menos 25% MiB/s of data ingest and 25MiB/s da recuperação de dados em todos os fluxos sob demanda. Para contas que atendem ao requisito mínimo de uso, o console do Kinesis Data Streams verifica se os padrões de uso da sua conta são adequados para usar o modo On-demand Advantage.

Se o uso de dados da sua conta estiver abaixo do exigido, você pagará a diferença como um déficit, mas ainda assim terá a mesma taxa de desconto. A ativação do On-demand Advantage também tem um período mínimo de 24 horas antes que você possa desativar o modo. No geral, o On-demand Advantage é a melhor opção para fluxos com o Kinesis Data Streams se você tiver um uso consistente de throughput próximo ou acima do compromisso mínimo, se precisar de muitos consumidores distribuídos ou se operar com centenas de fluxos de dados.

Casos de uso e atributos do modo provisionado

Com o modo provisionado, depois de criar o fluxo de dados, você pode aumentar ou reduzir dinamicamente sua capacidade de fragmentos usando a ou a Console de gerenciamento da AWS

API. UpdateShardCount É possível fazer atualizações enquanto uma aplicação de produção ou de consumo do Kinesis Data Streams grava ou lê dados do fluxo.

O modo provisionado é adequado para tráfego com requisitos de capacidade fáceis de prever. É possível usar o modo provisionado quando quiser ter um controle refinado da distribuição dos entre os fragmentos.

No modo provisionado, é necessário especificar o número de fragmentos para o fluxo de dados. Para determinar o tamanho inicial de um fluxo de dados, os seguintes valores de entrada são necessários:

- O tamanho médio do registro de dados gravado no fluxo em kilobytes (KB), arredondado para o 1 KB mais próximo (`average_data_size_in_KB`).
- O número de registros de dados gravados e lidos no fluxo por segundo (`records_per_second`).
- O número de consumidores, que são as aplicações do Kinesis Data Streams que consomem dados de forma simultânea e independente do fluxo (`number_of_consumers`).
- A largura de banda de gravação de entrada em KB (`incoming_write_bandwidth_in_KB`), que é igual a `average_data_size_in_KB` multiplicado por `records_per_second`.
- A largura de banda de leitura de saída em KB (`outgoing_read_bandwidth_in_KB`), que é igual a `incoming_write_bandwidth_in_KB` multiplicado por `number_of_consumers`.

É possível calcular o número dos fragmentos (`number_of_shards`) necessários para o fluxo usando os valores de entrada na seguinte fórmula:

```
number_of_shards = ceiling(max(incoming_write_bandwidth_in_KiB/1024,
 outgoing_read_bandwidth_in_KiB/2048))
```

Ainda será possível ocorrer exceções de throughput de leitura e gravação no modo provisionado se o fluxo de dados não for configurado para lidar com o throughput máximo. Nesse caso, será preciso escalar manualmente o fluxo para acomodar o tráfego de dados.

As exceções de leitura e gravação também podem ocorrer ao usar uma chave de partição que causa uma distribuição desigual de dados, e os registros atribuídos a um fragmento excedem seus limites. Para resolver esse problema no modo provisionado, identifique esses fragmentos e divida-os manualmente para acomodar melhor o tráfego. Para obter mais informações, consulte Resharding a Stream.

Alternar entre os modos

Para cada fluxo de dados em seu Conta da AWS, você pode alternar entre os modos sob demanda e provisionado duas vezes em 24 horas. Alternar entre os modos não causa nenhuma interrupção nas aplicações que usam esse fluxo de dados. É possível continuar gravando e lendo o fluxo de dados. Durante a alternância entre os modos, de sob demanda para provisionado ou vice-versa, o status do fluxo é definido como Atualizando. É necessário esperar que o status do fluxo de dados passe a Ativo antes de tentar modificar suas propriedades novamente.

Ao mudar do modo de capacidade provisionada para o modo de capacidade sob demanda, o fluxo de dados retém inicialmente a quantidade de fragmentos que tinha antes da transição. A partir desse momento, o Kinesis Data Streams monitora o tráfego de dados e escala a contagem de fragmentos do fluxo de dados sob demanda de acordo com o throughput de gravação. Ao alternar do modo sob demanda para o provisionado, o fluxo de dados também retém inicialmente a quantidade de fragmentos que tinha antes da transição. Mas, a partir desse momento, você passa a ser responsável por monitorar e ajustar a contagem de fragmentos do fluxo de dados para acomodar o throughput de gravação da forma apropriada.

Você pode alternar do modo On-demand Standard para o modo On-demand Advantage ativando uma configuração na conta. Quando ativada, a conta se compromete com pelo menos 25% MiB/s of data ingest and 25MiB/s do uso da recuperação de dados em todos os fluxos sob demanda na região. Depois de ativado, é necessário esperar pelo menos 24 horas antes de poder desativar o On-demand Advantage, mas você pode solicitar a alteração a qualquer momento. Se quiser mudar de On-demand Advantage para On-demand Standard, você deve primeiro remover qualquer throughput aquecido configurado com fluxos sob demanda.

Crie um stream usando o Console de gerenciamento da AWS

É possível criar um fluxo usando o console do Kinesis Data Streams, a API do Kinesis Data Streams ou a AWS Command Line Interface (AWS CLI).

Para criar um fluxo de dados usando o console

1. Faça login no Console de gerenciamento da AWS e abra o console do Kinesis em https://console.aws.amazon.com /kinesis.
2. Na barra de navegação, expanda o seletor de região e escolha uma região.
3. Selecione Criar fluxo de dados.

4. Na página Criar fluxo do Kinesis, insira um nome para o fluxo de dados e escolha o modo de capacidade Sob demanda ou Provisionado. O modo Sob demanda está selecionado por padrão. Para obter mais informações, consulte Escolha o modo certo para transmitir.

   No modo sob demanda, pode-se, em seguida, escolher Criar fluxo do Kinesis para criar o fluxo de dados. No modo provisionado, é necessário especificar o número de fragmentos necessários e, em seguida, escolher Criar fluxo do Kinesis.

   Na página Fluxos do Kinesis, o Status do fluxo é Criando enquanto o fluxo está sendo criado. Quando o fluxo estiver pronto para ser usado, o Status mudará para Ativo.
5. Escolha o nome do fluxo. A página Detalhes do fluxo exibe um resumo da configuração do fluxo com informações de monitoramento.

Para criar um fluxo usando a API do Kinesis Data Streams

- Para obter informações sobre a criação de um fluxo usando a API do Kinesis Data Streams, consulte Crie um stream usando o APIs.

Para criar um stream usando o AWS CLI

- Para obter informações sobre como criar um stream usando o AWS CLI, consulte o comando create-stream.

Crie um stream usando o APIs

Use as etapas a seguir para criar o fluxo de dados do Kinesis.

Criar o cliente do Kinesis Data Streams

Para trabalhar com fluxos de dados do Kinesis, é necessário criar um objeto de cliente. O seguinte código Java cria uma instância de um criador de cliente e a usa para definir a região, as credenciais e a configuração do cliente. Em seguida, ele cria um objeto do cliente.

```
AmazonKinesisClientBuilder clientBuilder = AmazonKinesisClientBuilder.standard();

clientBuilder.setRegion(regionName);
clientBuilder.setCredentials(credentialsProvider);
clientBuilder.setClientConfiguration(config);
```

```
AmazonKinesis client = clientBuilder.build();
```

Para obter mais informações, consulte Kinesis Data Streams Regions and Endpoints na Referência geral da AWS.

Criar o fluxo

Depois de criar o cliente do Kinesis Data Streams, é possível criar um fluxo no console ou de forma programática. Para criar um fluxo de forma programática, instancie um objeto `CreateStreamRequest`, depois especifique um nome para o fluxo. Se desejar usar o modo provisionado, especifique o número de fragmentos para o uso do fluxo de dados.

- Sob demanda:

```
CreateStreamRequest createStreamRequest = new CreateStreamRequest();
createStreamRequest.setStreamName( myStreamName );
```

- Provisionado:

```
CreateStreamRequest createStreamRequest = new CreateStreamRequest();
createStreamRequest.setStreamName( myStreamName );
createStreamRequest.setShardCount( myStreamSize );
```

O nome do fluxo identifica o fluxo. O nome tem como escopo a AWS conta usada pelo aplicativo. Ele também é delimitado por região. Ou seja, dois fluxos em duas AWS contas diferentes podem ter o mesmo nome, e dois fluxos na mesma AWS conta, mas em duas regiões diferentes, podem ter o mesmo nome, mas não dois fluxos na mesma conta e na mesma região.

O throughput do fluxo depende do número de fragmentos. Para obter um throughput mais provisionado, mais fragmentos são necessários. Mais fragmentos também aumentam o custo AWS cobrado pela transmissão. Para obter mais informações sobre como calcular um número apropriado de fragmentos para o aplicativo, consulte Escolha o modo certo para transmitir.

Depois de configurar o objeto `createStreamRequest`, crie um fluxo chamando o método `createStream` para o cliente. Após chamar `createStream`, aguarde o fluxo alcançar o estado `ACTIVE` antes de executar qualquer operação nele. Para verificar o estado do fluxo, chame o método `describeStream`. Se o fluxo não existir, `describeStream` lançará uma exceção, portanto, coloque a chamada a `describeStream` em um bloco `try/catch`.

```
client.createStream( createStreamRequest );
DescribeStreamRequest describeStreamRequest = new DescribeStreamRequest();
describeStreamRequest.setStreamName( myStreamName );

long startTime = System.currentTimeMillis();
long endTime = startTime + ( 10 * 60 * 1000 );
while ( System.currentTimeMillis() < endTime ) {
  try {
    Thread.sleep(20 * 1000);
  }
  catch ( Exception e ) {}

  try {
    DescribeStreamResult describeStreamResponse =
 client.describeStream( describeStreamRequest );
    String streamStatus =
 describeStreamResponse.getStreamDescription().getStreamStatus();
    if ( streamStatus.equals( "ACTIVE" ) ) {
      break;
    }
    //
    // sleep for one second
    //
    try {
      Thread.sleep( 1000 );
    }
    catch ( Exception e ) {}
  }
  catch ( ResourceNotFoundException e ) {}
}
if ( System.currentTimeMillis() >= endTime ) {
  throw new RuntimeException( "Stream " + myStreamName + " never went active" );
}
```

Atualizar um fluxo

É possível atualizar os detalhes de um fluxo usando o console do Kinesis Data Streams, a API do Kinesis Data Streams ou a AWS CLI.

> **Note**
>
> Pode-se ativar a criptografia no lado do servidor para streams existentes ou para os recém-criados.

Usar o console

Para atualizar um fluxo de dados usando o console

1. Abra o console do Amazon Kinesis em. https://console.aws.amazon.com/kinesis/
2. Na barra de navegação, expanda o seletor de região e escolha uma região.
3. Escolha o nome do fluxo na lista. A página Detalhes do fluxo exibe um resumo das informações de configuração e monitoramento do fluxo.
4. Para alternar entre os modos de capacidade sob demanda e provisionada para um fluxo de dados, escolha Editar modo de capacidade na guia Configuração. Para obter mais informações, consulte Escolha o modo certo para transmitir.

   > **Important**
   >
   > Para cada fluxo de dados em sua AWS conta, você pode alternar entre os modos sob demanda e provisionado duas vezes em 24 horas.

5. Para editar o número de fragmentos de um fluxo de dados no modo provisionado, escolha Editar fragmentos provisionados na guia Configuração e, em seguida, insira uma nova contagem de fragmentos.
6. Para ativar a criptografia de registros de dados no lado do servidor, selecione Editar na seção Criptografia no lado do servidor. Escolha uma chave do KMS para usar como a chave mestra de criptografia ou use a chave mestra padrão, aws/kinesis, gerenciada pelo Kinesis. Se você habilitar a criptografia para um stream e usar sua própria chave AWS KMS mestra, certifique-se de que seus aplicativos de produtor e consumidor tenham acesso à chave AWS KMS mestra que você usou. Para atribuir permissões a uma aplicação para acessar uma chave do AWS KMS gerada pelo usuário, consulte the section called “Permissões para usar as chaves do KMS geradas pelo usuário”.
7. Para editar o período de retenção de dados, selecione Editar na seção Período de retenção de dados e insira um novo período de retenção de dados.

8. Se métricas personalizadas estiverem habilitadas na conta, selecione Editar na seção Métricas em nível de fragmento e, em seguida, especifique as métricas de seu fluxo. Para obter mais informações, consulte the section called “Monitore o serviço Kinesis Data Streams com CloudWatch”.

Usar a API

Para atualizar os detalhes do fluxo usando a API, consulte os seguintes métodos:

- AddTagsToStream
- DecreaseStreamRetentionPeriod
- DisableEnhancedMonitoring
- EnableEnhancedMonitoring
- IncreaseStreamRetentionPeriod
- RemoveTagsFromStream
- StartStreamEncryption
- StopStreamEncryption
- UpdateShardCount

Use o AWS CLI

Para obter informações sobre como atualizar um stream usando o AWS CLI, consulte a referência da Kinesis CLI.

Listar fluxos

Os streams têm como escopo a AWS conta associada às AWS credenciais usadas para instanciar o cliente Kinesis Data Streams e também a região especificada para o cliente. Uma conta da AWS pode ter vários fluxos ativos ao mesmo tempo. É possível listar os fluxos no console do Kinesis Data Streams ou de forma programática. O código nesta seção mostra como listar todos os streams da sua AWS conta.

```
ListStreamsRequest listStreamsRequest = new ListStreamsRequest();
listStreamsRequest.setLimit(20);
ListStreamsResult listStreamsResult = client.listStreams(listStreamsRequest);
```

```
List<String> streamNames = listStreamsResult.getStreamNames();
```

Este exemplo de código primeiro cria uma nova instância de `ListStreamsRequest` e chama seu método `setLimit` para especificar que um máximo de 20 fluxos devem ser retornados para cada chamada a `listStreams`. Se um valor não for especificado para `setLimit`, o Kinesis Data Streams retornará um número de fluxos menor ou igual ao número na conta. Em seguida, o código passa `listStreamsRequest` ao método `listStreams` do cliente. O valor de retorno `listStreams` é armazenado em um objeto `ListStreamsResult`. O código chama o método `getStreamNames` para esse objeto e armazena os nomes de fluxo retornados na lista `streamNames`. Observe que o Kinesis Data Streams pode retornar menos fluxos do que o limite especificado, mesmo quando houver um número maior de fluxos na conta e na região. Para garantir a recuperação de todos os fluxos, use o método `getHasMoreStreams` como descrito no próximo exemplo de código.

```
while (listStreamsResult.getHasMoreStreams())
{
    if (streamNames.size() > 0) {
      listStreamsRequest.setExclusiveStartStreamName(streamNames.get(streamNames.size()
 - 1));
    }
    listStreamsResult = client.listStreams(listStreamsRequest);
    streamNames.addAll(listStreamsResult.getStreamNames());
}
```

Esse código chama o método `getHasMoreStreams` para `listStreamsRequest` a fim de verificar se há fluxos adicionais disponíveis além dos que foram retornados na chamada inicial a `listStreams`. Se houver, o código chamará o método `setExclusiveStartStreamName` com o nome do último fluxo que foi retornado na chamada anterior a `listStreams`. O método `setExclusiveStartStreamName` faz com que a próxima chamada a `listStreams` comece depois desse fluxo. Em seguida, o grupo de nomes de fluxo retornados pela chamada é adicionado à lista `streamNames`. Esse processo continua até que todos os nomes de fluxo tenham sido coletados na lista.

Os streams retornados por `listStreams` podem estar em um dos seguintes estados:

- `CREATING`
- `ACTIVE`
- `UPDATING`

- `DELETING`

É possível verificar o estado de um fluxo usando o método `describeStream`, como mostrado na seção anterior, Crie um stream usando o APIs.

Listar fragmentos

Um fluxo de dados pode ter um ou mais fragmentos. O método recomendado para listar ou recuperar os fragmentos de um stream de dados é usar a ListShardsAPI. O exemplo a seguir mostra como obter uma lista de fragmentos em um fluxo de dados. Para obter uma descrição completa da operação principal usada neste exemplo e de todos os parâmetros que você pode definir para a operação, consulte ListShards.

```
import software.amazon.awssdk.services.kinesis.KinesisAsyncClient;
import software.amazon.awssdk.services.kinesis.model.ListShardsRequest;
import software.amazon.awssdk.services.kinesis.model.ListShardsResponse;

import java.util.concurrent.TimeUnit;

public class ShardSample {

    public static void main(String[] args) {

        KinesisAsyncClient client = KinesisAsyncClient.builder().build();

        ListShardsRequest request = ListShardsRequest
                .builder().streamName("myFirstStream")
                .build();

        try {
            ListShardsResponse response = client.listShards(request).get(5000,
 TimeUnit.MILLISECONDS);
            System.out.println(response.toString());
        } catch (Exception e) {
            System.out.println(e.getMessage());
        }
    }
}
```

Para executar o exemplo de código anterior, é possível usar um arquivo POM, como o seguinte.

```
<?xml version="1.0" encoding="UTF-8"?>
<project xmlns="http://maven.apache.org/POM/4.0.0"
         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
         xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.org/
xsd/maven-4.0.0.xsd">
    <modelVersion>4.0.0</modelVersion>

    <groupId>kinesis.data.streams.samples</groupId>
    <artifactId>shards</artifactId>
    <version>1.0-SNAPSHOT</version>
    <build>
        <plugins>
            <plugin>
                <groupId>org.apache.maven.plugins</groupId>
                <artifactId>maven-compiler-plugin</artifactId>
                <configuration>
                    <source>8</source>
                    <target>8</target>
                </configuration>
            </plugin>
        </plugins>
    </build>
    <dependencies>
        <dependency>
            <groupId>software.amazon.awssdk</groupId>
            <artifactId>kinesis</artifactId>
            <version>2.0.0</version>
        </dependency>
    </dependencies>
</project>
```

Com a `ListShards` API, você pode usar o ShardFilterparâmetro para filtrar a resposta da API. Só é possível especificar um filtro de cada vez.

Se você usar o `ShardFilter` parâmetro ao invocar a ListShards API, essa `Type` é a propriedade necessária e deve ser especificada. Se os tipos `AT_TRIM_HORIZON`, `FROM_TRIM_HORIZON` ou `AT_LATEST` forem especificados, não há necessidade de especificar as propriedades opcionais `ShardId` e `Timestamp`.

Se o tipo `AFTER_SHARD_ID` for especificado, também é necessário fornecer o valor para a propriedade opcional `ShardId`. A `ShardId` propriedade é idêntica em termos de funcionalidade ao `ExclusiveStartShardId` parâmetro da ListShards API. Quando a propriedade `ShardId`

é especificada, a resposta inclui os fragmentos a partir daquele cuja ID segue imediatamente a `ShardId` fornecida.

Se os tipos `AT_TIMESTAMP` ou `FROM_TIMESTAMP_ID` forem especificados, também é necessário fornecer o valor para a propriedade opcional `Timestamp`. Se o tipo `AT_TIMESTAMP` for especificado, todos os fragmentos abertos no timestamp fornecido serão retornados. Se o tipo `FROM_TIMESTAMP` for especificado, todos os fragmentos do timestamp fornecido até a extremidade serão retornados.

Important

`DescribeStreamSummarye ListShard` APIs forneça uma forma mais escalável de recuperar informações sobre seus fluxos de dados. Mais especificamente, as cotas da DescribeStream API podem causar limitação. Para obter mais informações, consulte Cotas e limites. Observe também que as `DescribeStream` cotas são compartilhadas em todos os aplicativos que interagem com todos os fluxos de dados em sua AWS conta. As cotas da ListShards API, por outro lado, são específicas para um único fluxo de dados. Portanto, você não apenas obtém um TPS mais alto com a ListShards API, mas a ação é melhor dimensionada à medida que você cria mais fluxos de dados.
Recomendamos que você migre todos os produtores e consumidores que chamam a DescribeStream API para, em vez disso, invocar o DescribeStreamSummary e o. ListShard APIs Para identificar esses produtores e consumidores, recomendamos usar o Athena para analisar CloudTrail registros, pois os agentes de usuário para KPL e KCL são capturados nas chamadas de API.

```
SELECT useridentity.sessioncontext.sessionissuer.username,
useridentity.arn,eventname,useragent, count(*) FROM
cloudtrail_logs WHERE Eventname IN ('DescribeStream')  AND
eventtime
    BETWEEN ''
        AND ''
GROUP BY
 useridentity.sessioncontext.sessionissuer.username,useridentity.arn,eventname,useragent
ORDER BY  count(*) DESC LIMIT 100
```

Também recomendamos que as integrações do AWS Lambda e do Amazon Firehose com o Kinesis Data Streams que invocam a API sejam reconfiguradas para que, em vez disso, `DescribeStream` as integrações invoquem e. `DescribeStreamSummary ListShards`

Especificamente, para o AWS Lambda, você deve atualizar o mapeamento da fonte do evento. No Amazon Firehose, as permissões correspondentes do IAM precisam ser atualizadas para que incluam a permissão `ListShards` do IAM.

Excluir um fluxo

É possível excluir um fluxo no console do Kinesis Data Streams ou de forma programática. Para excluir um fluxo de maneira programática, use `DeleteStreamRequest`, conforme mostrado no código a seguir.

```
DeleteStreamRequest deleteStreamRequest = new DeleteStreamRequest();
deleteStreamRequest.setStreamName(myStreamName);
client.deleteStream(deleteStreamRequest);
```

Desative todos os aplicativos que estejam operando no fluxo antes de excluí-lo. Se um aplicativo tentar operar em um fluxo excluído, ele receberá exceções `ResourceNotFound`. Além disso, se um novo fluxo for criado com o mesmo nome do fluxo anterior e os aplicativos que operavam nele ainda estiverem em execução, esses aplicativos poderão tentar interagir com o fluxo novo como se fosse o anterior, com resultados imprevisíveis.

Refragmentar um fluxo

Important

Você pode refragmentar seu stream usando a UpdateShardCountAPI. Caso contrário, é possível continuar executando divisões e mesclagens, como explicado aqui.

O Amazon Kinesis Data Streams oferece suporte à refragmentação, o que permite ajustar o número de fragmentos no fluxo para se adaptar a alterações na taxa de dados no fluxo. A refragmentação é considerada uma operação avançada. Se esta é a primeira experiência com o Kinesis Data Streams, volte a este tópico depois de se familiarizar com todos os outros aspectos do serviço.

Há dois tipos de operações de refragmentação: divisão de fragmento e mesclagem de fragmento. Na divisão de fragmento, um único fragmento é dividido em dois. Na mesclagem de fragmento, dois fragmentos são combinados em um. A refragmentação sempre ocorre em pares, ou seja, não é possível dividir em mais de dois fragmentos em uma única operação, e não é possível mesclar mais

de dois fragmentos em uma única operação. O fragmento (ou o par de fragmentos) que é objeto da operação de refragmentação é chamado de fragmento pai. O fragmento (ou o par de fragmentos) resultante da operação de refragmentação é chamado de fragmento filho.

A divisão aumenta o número de fragmentos no fluxo e, portanto, aumenta a capacidade de dados do fluxo. Como a cobrança é feita por fragmento, a divisão aumenta o custo do fluxo. Comparativamente, a mesclagem reduz o número de fragmentos no fluxo e, portanto, diminui a capacidade de dados e o custo do fluxo.

A refragmentação costuma ser executada por um aplicativo administrativo, que é diferente dos aplicativos de produtor (put) e dos aplicativos de consumidor (get). Esse aplicativo administrativo monitora o desempenho geral do stream com base nas métricas fornecidas pela Amazon CloudWatch ou com base nas métricas coletadas dos produtores e consumidores. O aplicativo administrativo também precisa de um conjunto mais amplo de permissões do IAM do que os consumidores ou produtores, porque os consumidores e produtores geralmente não precisam acessar o APIs usado para refragmentação. Para obter mais informações sobre as permissões do IAM para o Kinesis Data Streams, consulte Controlar o acesso aos recursos do Amazon Kinesis Data Streams usando o IAM.

Para obter mais informações sobre refragmentação, consulte How do I change the number of open shards in Kinesis Data Streams?

Tópicos

- Decidir uma estratégia para refragmentar
- Dividir um fragmento
- Mesclar dois fragmentos
- Concluir a ação de refragmentação

Decidir uma estratégia para refragmentar

A finalidade da refragmentação no Amazon Kinesis Data Streams é permitir que o fluxo se adapte a alterações na taxa do fluxo de dados. Fragmentos são divididos para aumentar a capacidade (e o custo) do fluxo. Fragmentos são mesclados para reduzir o custo (e a capacidade) do fluxo.

Uma abordagem de refragmentação pode ser dividir cada fragmento do fluxo, o que dobraria sua capacidade. No entanto, isso pode fornecer mais capacidade adicional do que o realmente necessário e, portanto, gerar um custo desnecessário.

Também é possível usar métricas para identificar os fragmentos quentes ou frios, ou seja, os fragmentos que estão recebendo muito mais ou muito menos dados do que o esperado. Em seguida, é possível seletivamente dividir os fragmentos quentes para aumentar a capacidade das chaves de hash que almejam esses fragmentos. Comparativamente, pode-se mesclar os fragmentos frios para dar uma melhor serventia à capacidade não usada.

Você pode obter alguns dados de desempenho do seu stream a partir das CloudWatch métricas da Amazon que o Kinesis Data Streams publica. No entanto, também é possível coletar algumas métricas dos seus fluxos. Uma abordagem é registrar em log os valores de chave de hash gerados pelas chaves de partição dos seus registros de dados. Lembre-se de que a chave de partição é especificada no momento em que o registro é adicoinado ao fluxo.

```
putRecordRequest.setPartitionKey( String.format( "myPartitionKey" ) );
```

O Kinesis Data MD5Streams usa para calcular a chave de hash a partir da chave de partição. Como você especifica a chave de partição para o registro, você pode MD5 usá-la para calcular o valor da chave de hash desse registro e registrá-lo.

Você também pode registrar os IDs fragmentos aos quais seus registros de dados estão atribuídos. O ID do fragmento é obtido usando-se o método `getShardId` do objeto `putRecordResults` retornado pelo método `putRecords` e o objeto `putRecordResult` retornado pelo método `putRecord`.

```
String shardId = putRecordResult.getShardId();
```

Com os valores do fragmento IDs e da chave de hash, você pode determinar quais fragmentos e chaves de hash estão recebendo mais ou menos tráfego. Em seguida, é possível usar a refragmentação para fornecer mais ou menos capacidade, conforme apropriado para essas chaves.

Dividir um fragmento

Para dividir um fragmento no Amazon Kinesis Data Streams, é necessário especificar como os valores de chave de hash dos fragmentos pai devem ser redistribuídos para os fragmentos filho. Ao adicionar um registro de dados a um fluxo, ele é atribuído a um fragmento com base em um valor de chave de hash. O valor da chave de hash é o MD5hash da chave de partição que você especifica para o registro de dados no momento em que adiciona o registro de dados ao fluxo. Os registros de dados que têm a mesma chave de partição também têm o mesmo valor de chave de hash.

Os valores possíveis de chave de hash de um determinado fragmento constituem um conjunto de números inteiros contíguos, não negativos e ordenados. Esse intervalo de possíveis valores de chave de hash é determinado pelo seguinte:

```
shard.getHashKeyRange().getStartingHashKey();
shard.getHashKeyRange().getEndingHashKey();
```

Ao dividir o fragmento, um valor é especificado neste intervalo. Esse valor de chave de hash e todos os valores de chave de hash maiores são distribuídos para um dos fragmentos filhos. Todos os valores de chave de hash menores são distribuídos para os outros fragmentos filhos.

O seguinte código demonstra uma operação de divisão de fragmento que redistribui as chaves de hash uniformemente entre cada um dos fragmentos filhos, basicamente, dividindo o fragmento pai no meio. Essa é apenas uma das maneiras possíveis de dividir o fragmento pai. É possível, por exemplo, dividir o fragmento de maneira que o terço inferior das chaves do pai vá para um fragmento filho e os dois terços superiores das chaves vão para o outro fragmento filho. No entanto, para muitos aplicativos, dividir os fragmentos no meio é uma abordagem eficiente.

O código pressupõe que `myStreamName` contém o nome do seu fluxo e a variável de objeto `shard` contém o fragmento a dividir. Comece instanciando um novo objeto `splitShardRequest` e definindo o nome do fluxo e o ID do fragmento.

```
SplitShardRequest splitShardRequest = new SplitShardRequest();
splitShardRequest.setStreamName(myStreamName);
splitShardRequest.setShardToSplit(shard.getShardId());
```

Determine o valor da chave de hash que é meio caminho entre o maior valor e o menor valor no fragmento. Trata-se do valor de chave de hash inicial para o fragmento filho que conterá a metade superior das chaves de hash do fragmento pai. Especifique esse valor no método `setNewStartingHashKey`. Basta especificar esse valor. O Kinesis Data Streams distribui automaticamente as chaves de hash abaixo desse valor para os outros fragmentos filho criados pela divisão. A última etapa é chamar o método `splitShard` no cliente do Kinesis Data Streams.

```
BigInteger startingHashKey = new
 BigInteger(shard.getHashKeyRange().getStartingHashKey());
BigInteger endingHashKey   = new
 BigInteger(shard.getHashKeyRange().getEndingHashKey());
String newStartingHashKey  = startingHashKey.add(endingHashKey).divide(new
 BigInteger("2")).toString();
```

```
splitShardRequest.setNewStartingHashKey(newStartingHashKey);
client.splitShard(splitShardRequest);
```

A primeira etapa após este procedimento é mostrada em Aguardar um fluxo ficar ativo novamente.

Mesclar dois fragmentos

Uma operação de mesclagem de fragmentos usa dois fragmentos especificados e combina-os em um único fragmento. Após a mesclagem, o único fragmento filho recebe dados para todos os valores de chave de hash cobertos pelos dois fragmentos pais.

Adjacência de fragmento

Para mesclar dois fragmentos, eles precisam estar adjacentes. Dois fragmentos são considerados adjacentes quando a união dos intervalos de chave de hash dos dois fragmentos forma um conjunto contíguo sem lacunas. Por exemplo, suponha que haja dois fragmentos, um com o intervalo de chaves de hash 276...381 e o outro com o intervalo de chaves de hash 382...454. É possível mesclar esses dois fragmentos em um só, que teria um intervalo de chaves de hash 276...454.

Para usar outro exemplo, suponha que haja dois fragmentos, um com um intervalo de chaves de hash 276...381 e o outro com um intervalo de chaves de hash 455...560. Não seria possível mesclar esses dois fragmentos porque haveria um ou mais fragmentos entre eles que estariam no intervalo 382...454.

O conjunto de todos os `OPEN` fragmentos em um fluxo, como um grupo, sempre abrange toda a faixa de valores da chave de hash. MD5 Para obter mais informações sobre esses estados de fragmento (como `CLOSED`), consulte Considerar o roteamento de dados, a persistência de dados e o estado do fragmento após uma refragmentação.

Para identificar os fragmentos candidatos para mesclagem, deve-se filtrar todos os fragmentos que estão no estado `CLOSED`. Os fragmentos `OPEN` (ou seja, não `CLOSED`) têm um número de sequência final `null`. É possível testar o número sequencial de término de um fragmento usando:

```
if( null == shard.getSequenceNumberRange().getEndingSequenceNumber() )
{
  // Shard is OPEN, so it is a possible candidate to be merged.
}
```

Após filtrar os fragmentos fechados, classifique os fragmentos restantes pelo valor de chave de hash mais alto aceito por cada fragmento. É possível recuperar esse valor usando:

```
shard.getHashKeyRange().getEndingHashKey();
```

Se dois fragmentos são adjacentes nessa lista filtrada e classificada, eles podem ser mesclados.

Código da operação de mesclagem

O código a seguir mescla dois fragmentos. O código pressupõe que `myStreamName` contém o nome do seu fluxo e as variáveis de objeto `shard1` e `shard2` contém os dois fragmentos adjacentes a mesclar.

Para a operação de mesclagem, comece instanciando um novo objeto `mergeShardsRequest`. Especifique o nome do fluxo com o método `setStreamName`. Em seguida, especifique os dois fragmentos a mesclar usando os métodos `setShardToMerge` e `setAdjacentShardToMerge`. Por fim, chame o método `mergeShards` no cliente do Kinesis Data Streams para executar a operação.

```
MergeShardsRequest mergeShardsRequest = new MergeShardsRequest();
mergeShardsRequest.setStreamName(myStreamName);
mergeShardsRequest.setShardToMerge(shard1.getShardId());
mergeShardsRequest.setAdjacentShardToMerge(shard2.getShardId());
client.mergeShards(mergeShardsRequest);
```

A primeira etapa após este procedimento é mostrada em Aguardar um fluxo ficar ativo novamente.

Concluir a ação de refragmentação

Após qualquer tipo de procedimento de refragmentação no Amazon Kinesis Data Streams e antes de retomar o processamento normal de registros, é necessário realizar outros procedimentos e fazer algumas considerações. As seções a seguir descrevem esses itens.

Tópicos

- Aguardar um fluxo ficar ativo novamente
- Considerar o roteamento de dados, a persistência de dados e o estado do fragmento após uma refragmentação

Aguardar um fluxo ficar ativo novamente

Depois de chamar uma operação de refragmentação, `splitShard` ou `mergeShards`, deve-se esperar o fluxo ficar ativo novamente. O código a ser usado é o mesmo de quando espera-se que um fluxo se torne ativo após a criação de um fluxo. Esse código é o seguinte:

```
DescribeStreamRequest describeStreamRequest = new DescribeStreamRequest();
describeStreamRequest.setStreamName( myStreamName );

long startTime = System.currentTimeMillis();
long endTime = startTime + ( 10 * 60 * 1000 );
while ( System.currentTimeMillis() < endTime )
{
  try {
    Thread.sleep(20 * 1000);
  }
  catch ( Exception e ) {}

  try {
    DescribeStreamResult describeStreamResponse =
 client.describeStream( describeStreamRequest );
    String streamStatus =
 describeStreamResponse.getStreamDescription().getStreamStatus();
    if ( streamStatus.equals( "ACTIVE" ) ) {
      break;
    }
   //
    // sleep for one second
    //
    try {
      Thread.sleep( 1000 );
    }
    catch ( Exception e ) {}
  }
  catch ( ResourceNotFoundException e ) {}
}
if ( System.currentTimeMillis() >= endTime )
{
  throw new RuntimeException( "Stream " + myStreamName + " never went active" );
}
```

Considerar o roteamento de dados, a persistência de dados e o estado do fragmento após uma refragmentação

O Kinesis Data Streams é um serviço de fluxo de dados em tempo real. Seus aplicativos devem pressupor que os dados fluem continuamente pelos fragmentos do fluxo. Ao refragmentar, os registros de dados que estavam fluindo para os fragmentos pais são re-roteados para fluírem para os fragmentos filhos com base nos valores de chave de hash para as quais são mapeadas as chaves

de partição de registro de dados. No entanto, os registros de dados que estavam nos fragmentos pais antes da refragmentação permanecem nesses fragmentos. Os fragmentos principais não desaparecem quando a refragmentação ocorre. Eles persistem com os dados que continham antes da refragmentação. Os registros de dados nos fragmentos pai podem ser acessados usando as operações `getShardIterator` e `getRecords` na API do Kinesis Data Streams ou por meio da Kinesis Client Library.

> **Note**
>
> Os registros de dados podem ser acessados a partir do momento em que são adicionados ao fluxo até o período de retenção atual. Isso é verdadeiro independentemente de quaisquer alterações aos fragmentos no fluxo durante esse período. Para obter mais informações sobre o período de retenção de um stream, consulte Alterar o período de retenção de dados.

No processo de refragmentação, um fragmento pai passa de um estado `OPEN` para um estado `CLOSED` para um estado `EXPIRED`.

- OPEN: antes de uma operação de refragmentação, um fragmento pai está no estado `OPEN`, o que significa que os registros de dados podem ser adicionados ao fragmento e recuperados do fragmento.
- CLOSED: após uma operação de refragmentação, o fragmento pai passa para um estado `CLOSED`. Isso significa que os registros de dados não são mais adicionados ao fragmento. Os registros de dados que foram adicionados a esse fragmento agora são adicionados a um fragmento filho. No entanto, os registros de dados ainda podem ser recuperados do fragmento por tempo limitado.
- EXPIRED: após a expiração do período de retenção do fluxo, todos os registros de dados no fragmento pai expiraram e não estão mais acessíveis. Neste momento, o próprio fragmento muda para um estado `EXPIRED`. As chamadas a `getStreamDescription().getShards` para enumerar os fragmentos no fluxo não incluem os fragmentos `EXPIRED` na lista de fragmentos retornados. Para obter mais informações sobre o período de retenção de um stream, consulte Alterar o período de retenção de dados.

Depois da refragmentação, com o fluxo novamente em um estado `ACTIVE`, é possível iniciar imediatamente a leitura de dados dos fragmentos filhos. No entanto, os fragmentos principais que permanecem após a refragmentação ainda podem conter dados que ainda não foram lidos e foram adicionados ao fluxo antes da refragmentação. Ao ler dados de fragmentos filhos antes ler todos os dados dos fragmentos pais, pode-se ler dados de uma determinada chave de hash fora da ordem

determinada pelos números sequenciais dos registros de dados. Portanto, pressupondo que a ordem dos dados é importante, é necessário, após uma refragmentação, sempre continuar lendo dados dos fragmentos pais até esgotá-los. Só depois deve-se começar a ler dados dos fragmentos filhos. Quando `getRecordsResult.getNextShardIterator` retorna `null`, indica que todos os dados no fragmento pai foram lidos.

Alterar o período de retenção de dados

O Amazon Kinesis Data Streams oferece suporte a alterações do período de retenção do registro de dados no fluxo de dados. Um fluxo de dados do Kinesis é uma sequência ordenada de registros de dados projetada para gravação e leitura em tempo real. Os registros de dados são, portanto, armazenados em fragmentos no fluxo temporariamente. O período entre o momento de adição de um registro e o momento em que ele deixa de estar acessível é chamado de período de retenção. Por padrão, um fluxo de dados do Kinesis armazena registros de 24 horas até 8.760 horas (365 dias).

Você pode atualizar o período de retenção por meio do console do Kinesis Data Streams ou IncreaseStreamRetentionPeriodusando as DecreaseStreamRetentionPeriodoperações e. No console do Kinesis Data Streams, pode-se editar em massa o período de retenção de mais de um fluxo de dados ao mesmo tempo. Você pode aumentar o período de retenção até um máximo de 8760 horas (365 dias) usando a IncreaseStreamRetentionPeriodoperação ou o console do Kinesis Data Streams. Você pode reduzir o período de retenção para um mínimo de 24 horas usando a DecreaseStreamRetentionPeriodoperação ou o console do Kinesis Data Streams. A sintaxe de solicitação das duas operações inclui o nome do fluxo e o período de retenção em horas. Finalmente, você pode verificar o período de retenção atual de um stream chamando a operação DescribeStream.

Este é um exemplo de alteração do período de retenção que usa a AWS CLI:

```
aws kinesis increase-stream-retention-period --stream-name retentionPeriodDemo --
retention-period-hours 72
```

O Kinesis Data Streams para de tornar inacessíveis os registros no período de retenção antigo vários minutos após o aumento do período. Por exemplo, alterar o período de retenção de 24 horas para 48 horas significa que os registros adicionados ao fluxo 23 horas 55 minutos antes ainda estarão disponíveis depois de 24 horas.

O Kinesis Data Streams torna inacessíveis os registros mais antigos que o novo período de retenção quase imediatamente após a diminuição do período. Portanto, deve-se tomar muito cuidado ao chamar a operação DecreaseStreamRetentionPeriod.

Defina o período de retenção dos dados para garantir que os consumidores possam ler dados antes de expirar, se ocorrerem problemas. Deve-se considerar cuidadosamente todas as possibilidades, como um problema com a lógica de processamento de registro ou a inatividade de uma dependência de downstream por um longo período. Ideia do período de retenção como uma rede de segurança para dar mais tempo para os consumidores de dados se recuperarem. As operações da API de período de retenção permitem que isso seja configurado de forma proativa ou responda a eventos operacionais reativamente.

Encargos adicionais incidem sobre streams com período de retenção definido acima de 24 horas. Para obter mais informações, consulte Definição de preço do Amazon Kinesis Data Streams.

Marcar com tag seus recursos do Amazon Kinesis Data Streams

É possível atribuir seus próprios metadados aos fluxos e aos consumidores de distribuição avançada criados no Amazon Kinesis Data Streams na forma de tags. Tag é um par de chave-valor definida para um fluxo. Usar tags é uma maneira simples, porém poderosa, de gerenciar AWS recursos e organizar dados, incluindo dados de faturamento.

Sumário

- Revisar conceitos básicos de tags
- Monitorar custos usando tags
- Compreender as restrições de tags
- Atribuir tags a fluxos usando o console do Kinesis Data Streams
- Marque fluxos usando o AWS CLI
- Marque streams usando o Kinesis Data Streams APIs
- Marque os consumidores usando o AWS CLI
- Marque consumidores usando o Kinesis Data Streams APIs

Revisar conceitos básicos de tags

Os recursos do Kinesis Data Streams que você pode marcar com tags incluem fluxos de dados e consumidores de distribuição avançada. Você usa o AWS CLI console do Kinesis Data Streams ou a API do Kinesis Data Streams para concluir as seguintes tarefas:

- Criar um recurso com tags
- Adicionar tags a um recurso
- Listar as tags para seus recursos
- Remover as tags de um recurso

> **Note**
>
> Você não pode aplicar tags a consumidores de distribuição avançada usando o console do Kinesis Data Streams. Para aplicar tags aos consumidores, use AWS CLI nossa API Kinesis Data Streams.

É possível usar tags para categorizar os recursos do . Por exemplo, você pode categorizar recursos por finalidade, proprietário ou ambiente. Como você define a chave e o valor para cada marca, é possível criar um conjunto de categorias personalizado para atender às suas necessidades específicas. Por exemplo, é possível definir um conjunto de tags que ajude a monitorar os recursos por proprietário e por aplicação associada. Aqui estão alguns exemplos de tags:

- Projeto: nome do projeto
- Proprietário: nome
- Objetivo: testes de carga
- Aplicação: nome da aplicação
- Ambiente: produção

> **Important**
>
> - Para adicionar tags ao criar um fluxo, inclua as permissões `kinesis:CreateStream` e `kinesis:AddTagsToStream` para esse fluxo. Você não pode usar a permissão `kinesis:TagResource` para marcar fluxos ao criá-los.

- Para adicionar tags durante o registro do consumidor, inclua as permissões `kinesis:TagResource` e `kinesis:RegisterStreamConsumer`.

Monitorar custos usando tags

Você pode usar tags para categorizar e monitorar seus AWS custos. Quando você aplica tags aos seus recursos do Kinesis Data Streams AWS , seu relatório de alocação de custos inclui o uso e os custos agregados por tags. É possível aplicar tags que representem categorias de negócios, como centros de custos, nomes de aplicações ou proprietários, para organizar seus custos em vários serviços. Para obter mais informações, consulte Usar tags de alocação de custos para relatórios de faturamento personalizados no Manual do usuário do AWS Billing .

Compreender as restrições de tags

As restrições a seguir se aplicam às tags:

Restrições básicas

- O número máximo de tags para cada recurso é 50.
- As chaves e os valores de marcas diferenciam maiúsculas de minúsculas.
- Não é possível alterar nem editar as tags de um recurso excluído.

Restrições de chaves de marcas

- Cada chave de tag deve ser exclusiva. Se uma tag for adicionada com uma chave que já estiver em uso, a nova tag substituirá o par de chave-valor existente.
- Você não pode iniciar uma chave de tag com `aws:` porque esse prefixo é reservado para uso por AWS. AWS cria tags que começam com esse prefixo em seu nome, mas você não pode editá-las nem excluí-las.
- As chaves de tag devem ter entre 1 e 128 caracteres Unicode.
- As chaves de tag devem conter os seguintes caracteres: letras Unicode, dígitos, espaço em branco e os seguintes caracteres especiais: `_ . / = + - @`.

Restrições de valor das tags

- Os valores das tags devem ter entre 0 e 255 caracteres Unicode.

- Os valores das tags podem estar em branco. Caso contrário, eles devem conter os seguintes caracteres: letras Unicode, dígitos, espaço em branco e qualquer um dos seguintes caracteres especiais: `_ . / = + - @`.

Atribuir tags a fluxos usando o console do Kinesis Data Streams

É possível adicionar, atualizar, listar e remover tags em seus fluxos usando o console do Kinesis Data Streams.

Para visualizar as tags de um fluxo

1. Faça login no Console de gerenciamento da AWS e abra o console do Kinesis em https://console.aws.amazon.com /kinesis.
2. No painel de navegação à esquerda, selecione Fluxos de dados.
3. Na página Fluxos de dados, selecione o fluxo que deseja marcar com tag.
4. Na página Detalhes do fluxo, selecione Configuração.
5. Na seção Tags, visualize as tags aplicadas ao fluxo.

Como criar um fluxo de dados com uma tag

1. Abra o console do Kinesis Data Streams.
2. No painel de navegação à esquerda, selecione Fluxos de dados.
3. Selecione Criar fluxo de dados.
4. Na página Criar fluxo de dados, insira um nome para seu fluxo de dados.
5. Para Capacidade de fluxo de dados, selecione o modo de capacidade Sob demanda ou Provisionada.

   Para obter mais informações sobre modos de capacidade, consulte Escolha o modo certo para transmitir.
6. Na seção Tags, faça o seguinte:

   a. Selecione Adicionar nova tag.

   b. Para Chave, insira a tag e, opcionalmente, especifique um valor no campo Valor.

      Se aparecer um erro, a chave ou o valor da tag especificado não atende às restrições da tag. Para obter mais informações, consulte Compreender as restrições de tags.

7. Selecione Criar fluxo de dados.

Para adicionar ou atualizar uma tag em um fluxo

1. Abra o console do Kinesis Data Streams.
2. No painel de navegação à esquerda, selecione Fluxos de dados.
3. Na página Fluxos de dados, selecione o fluxo ao qual você deseja adicionar ou atualizar as tags.
4. Na página Detalhes do fluxo, selecione Configuração.
5. Na seção Tags, escolha Gerenciar tags.
6. Em Tags, faça uma das seguintes opções:

   - Para adicionar uma tag, selecione Adicionar nova tag e insira os dados da tag Chave e Valor. Repita esta etapa quantas vezes for necessário.

     O número máximo de tags que podem ser adicionadas a cada fluxo é 50.
   - Para atualizar uma tag atual, insira um novo valor de tag no campo Valor da Chave dessa tag.

   Se aparecer um erro, a chave ou o valor da tag especificado não atende às restrições da tag. Para obter mais informações, consulte Compreender as restrições de tags.
7. Escolha Salvar alterações.

Para remover uma tag de um fluxo

1. Abra o console do Kinesis Data Streams.
2. No painel de navegação à esquerda, selecione Fluxos de dados.
3. Na página Fluxos de dados, selecione o fluxo do qual você deseja remover as tags.
4. Na página Detalhes do fluxo, selecione Configuração.
5. Na seção Tags, escolha Gerenciar tags.
6. Localize o par de tags Chave e Valor que você deseja remover. Em seguida, escolha Remover.
7. Escolha Salvar alterações.

Marque fluxos usando o AWS CLI

É possível adicionar, listar e remover tags em seus fluxos usando a AWS CLI. Para obter exemplos, consulte a seguinte documentação.

create-stream

Cria um fluxo com tags.

add-tags-to-stream

Adiciona ou atualiza as tags para o fluxo especificado.

list-tags-for-stream

Lista as tags para o fluxo especificado.

remove-tags-from-stream

Remove as tags do fluxo especificado.

Marque streams usando o Kinesis Data Streams APIs

Você pode adicionar, listar e remover tags em seus streams usando o Kinesis Data Streams. APIs Para obter exemplos, consulte a seguinte documentação:

CreateStream

Cria um fluxo com tags.

AddTagsToStream

Adiciona ou atualiza as tags para o fluxo especificado.

ListTagsForStream

Lista as tags para o fluxo especificado.

RemoveTagsFromStream

Remove as tags do fluxo especificado.

Marque os consumidores usando o AWS CLI

É possível adicionar, listar e remover tags em seus consumidores usando a AWS CLI. Para obter exemplos, consulte a seguinte documentação:

register-stream-consumer

Registra um consumidor para um fluxo de dados do Kinesis com tags.

tag-resource

Adiciona ou atualiza tags para o recurso do Kinesis especificado.

list-tags-for-resource

Lista as tags do recurso do Kinesis especificado.

untag-resource

Remove as tags do recurso do Kinesis especificado.

Marque consumidores usando o Kinesis Data Streams APIs

Você pode adicionar, listar e remover tags em seus consumidores usando o Kinesis APIs Data Streams. Para obter exemplos, consulte a seguinte documentação:

RegisterStreamConsumer

Registra um consumidor para um fluxo de dados do Kinesis com tags.

TagResource

Adiciona ou atualiza tags para o recurso do Kinesis especificado.

ListTagsForResource

Lista as tags do recurso do Kinesis especificado.

UntagResource

Remove as tags do recurso do Kinesis especificado.

Gerenciar registros grandes

O Amazon Kinesis Data Streams suporta registros de até 10 MiBs mebibytes (). Esse recurso é recomendado para processar cargas úteis intermitentes de dados que excedam o limite padrão de tamanho de registro de 1 MiB. O tamanho de registro máximo padrão para fluxos existentes e recém-criados é definido como 1 MiB.

Esse atributo beneficia aplicações de Internet das Coisas (IoT), pipelines de captura de dados de alteração (CDC) e fluxos de trabalho de machine learning que exigem o processamento ocasional de cargas úteis de dados de maior tamanho. Para começar a usar registros grandes em seu fluxo, atualize o limite máximo de tamanho de registro do seu fluxo.

> **Important**
>
> O limite individual de taxa de transferência de fragmentos de 1 MB/s para gravações e 2 MB/s para leituras permanece inalterado com suporte para tamanhos de registro maiores. O Kinesis Data Streams foi projetado para acomodar grandes registros intermitentes junto com um tráfego básico de registros menores ou iguais a 1 MiB. Ele não foi projetado para acomodar a ingestão sustentada de registros grandes e de alto volume.

Atualizar seu fluxo para usar registros grandes

Para processar registros maiores com o Kinesis Data Streams

1. Vá para o console do Kinesis Data Streams.
2. Selecione seu fluxo e acesse a guia Configuração.
3. Clique em Editar, localizado ao lado de Tamanho máximo do registro.
4. Defina o tamanho máximo do registro (até 10 MiB).
5. Salve as alterações.

Essa configuração ajusta somente o tamanho máximo do registro para esse fluxo de dados do Kinesis. Antes de aumentar esse limite, confirme se todas as aplicações downstream podem gerenciar registros maiores.

Você também pode atualizar essa configuração usando a AWS CLI:

```
aws kinesis update-max-record-size \ --stream-arn  \
        --max-record-size-in-ki-b 5000
```

Otimizar o desempenho do seu fluxo com registros grandes

É recomendável manter registros grandes em menos de 2% do seu tráfego geral. Em um fluxo, cada fragmento tem uma capacidade de throughput de 1 MiB por segundo. Para acomodar registros grandes, os streams do Kinesis Data aumentam até 10 MiBs, com uma média de 1 MiB por segundo. Essa capacidade de ser compatível com registros grandes é continuamente recarregada no fluxo. A taxa de recarga depende do tamanho dos registros grandes e do tamanho do registro básico. Para obter melhores resultados, use uma chave de partição distribuída uniformemente. Para obter mais informações sobre como o Kinesis sob demanda escala, consulte Atributos e casos de uso do modo sob demanda.

Reduzir o controle de utilização com registros grandes

Para mitigar o controle de utilização

1. Implemente a lógica de repetição com espera exponencial em sua aplicação produtora.
2. Use chaves de partição aleatórias para distribuir registros grandes pelos fragmentos disponíveis.
3. Armazene cargas úteis no Amazon S3 e envie somente referências de metadados ao fluxo para fluxos contínuos de registros grandes. Para obter mais informações, consulte Processamento de registros grandes com o Amazon Kinesis Data Streams.

Gerencie registros grandes usando o Kinesis Data Streams APIs

O suporte a registros grandes introduz uma nova API e atualiza dois planos de controle existentes APIs para lidar com registros de até 10 MiBs.

API para modificar o tamanho do registro:

- `UpdateMaxRecordSize`: configura o limite máximo de tamanho de registro para streams existentes de até 10. MiBs

Atualizações para os existentes APIs:

- `CreateStream`: adiciona o parâmetro `MaxRecordSizeInKiB` opcional para definir limites de tamanho de registro durante a criação do fluxo.
- `DescribeStreamSummary`: retorna o campo `MaxRecordSizeInKiB` para mostrar a configuração atual do fluxo.

Todos os APIs listados mantêm a compatibilidade com versões anteriores dos fluxos existentes. Para obter a documentação completa da API, consulte a Referência da API do serviço do Amazon Kinesis Data Streams.

AWS componentes compatíveis com registros grandes

Os AWS componentes a seguir são compatíveis com registros grandes:

Componente	Description
AWS SDK	AWS O SDK suporta o gerenciamento de registros grandes. Você pode atualizar o tamanho máximo do registro do seu stream em até 10 MiB usando os métodos disponíveis no. AWS SDKs Para obter mais informações, consulte Como usar esse serviço com um AWS SDK.
Kinesis Consumer Library (KCL)	A KCL é compatível com o gerencimento de registros grandes a partir da versão 2.x. Para usar o suporte a registros grandes, atualize o `maxRecordSize` do seu fluxo e use a KCL. Para obter mais informações, consulte Usar a Kinesis Client Library.
Kinesis Producer Library (KPL)	A partir da versão 1.0.5, a KPL é compatíve l com o gerenciamento de registros grandes. Para usar o suporte a registros grandes, atualize o `maxRecordSize` de seu fluxo e use a KPL. Para mais informações, consulte Desenvolver produtores usando a Amazon Kinesis Producer Library (KPL).

Componente	Description
Amazon EMR	O Amazon EMR com Apache Spark oferece suporte ao tratamento de registros grandes até o limite do Kinesis Data Streams (10). MiBs Para usar o suporte a registros grandes, use a função `readStream` . Para mais informaçõ es, consulte Integração do Amazon EMR com o Amazon Kinesis.

Componente	Description
Amazon Data Firehose	Quando usado com o Kinesis Data Streams, o comportamento do Amazon Data Firehose com registros grandes depende do destino da entrega: • Amazon S3: a entrega de registros grandes é compatível sem nenhuma configuração adicional. Quando você usa a conversão de formato de dados, a entrega de registros grandes é compatível com Firehose. Quando você usa particionamento dinâmico, a entrega de registros grandes não é compatível com Firehose. • Lambda: não recomendamos o uso de registros grandes com o Firehose quando ele aciona funções do Lambda downstream. Isso pode levar a falhas intermitentes. • HTTP: a entrega de registros grandes não é compatível com Firehose. • Snowflake: a entrega de registros grandes não é compatível com Firehose. • Amazon Redshift: a entrega de registros grandes não é compatível com o Firehose. Para aplicativos que exigem entrega para o Snowflake ou Redshift com registros grandes, entregue primeiro os dados para o Amazon S3. Depois disso, use os processos de Extração, Transformação e Carregamento (ETL) para carregar os dados. Para todos os outros destinos, teste o comportamento com registros grandes em um proof-of-concept ambiente antes de escalar para o uso em produção. O

Componente	Description
	gerenciamento de registros grandes varia de acordo com o destino.
AWS Lambda	AWS Lambda suporta cargas úteis de até 6 MiBs. Esse limite inclui a carga útil do Kinesis convertida em codificação de base 64 e os metadados associados ao mapeamento da origem do evento (ESM). Para registros menores que 6 MiBs, o Lambda os processa usando o ESM sem necessidade de configuração adicional. Para registros maiores que 6 MiBs, o Lambda os processa usando um destino em caso de falha. É necessário configurar um destino em caso de falha usando o ESM para gerenciar registros que excedam os limites de processamento do Lambda. Cada evento enviado ao destino em caso de falha é um documento JSON que contém metadados sobre a invocação que falhou. É recomendável criar um destino em caso de falha no ESM, independentemente do tamanho do registro. Isso garante que nenhum registro seja descartado. Para obter mais informações, consulte Configurar destinos para invocações com falha.

Componente	Description
banco de dados de origem	O Amazon Redshift só oferece suporte a tamanhos de registro menores que 1 MiB no fluxo de dados a partir do Kinesis Data Streams. Os registros que excederem esse limite não serão processados. Os registros não processados são registrados em log como `sys_stream_scan_errors` . Para obter mais informações, consulte SYS_STREAM_SCAN_ERRORS.
Conector Flink para Kinesis Data Streams	Há duas abordagens para consumir dados do Kinesis Data Streams: o conector de origem do Kinesis e o conector coletor do Kinesis. O conector de origem suporta o tratamento de registros menores que 1 MiB e até 10. MiBs Não use o conector do coletor para registros maiores que 1 MiB. Para obter mais informaçõ es, consulte Usar conectores para mover dados no Amazon Managed Service para Apache Flink com a API. DataStream

Regiões em que grandes registros são compatíveis

Esse recurso do Amazon Kinesis Data Streams está disponível somente nas AWS seguintes regiões:

AWS Região	Nome da região
eu-north-1	Europa (Estocolmo)
me-south-1	Oriente Médio (Bahrein)
ap-south-1	Ásia-Pacífico (Mumbai)
eu-west-3	Europa (Paris)

AWS Região	Nome da região
ap-southeast-3	Ásia-Pacífico (Jacarta)
us-east-2	Leste dos EUA (Ohio)
af-south-1	África (Cidade do Cabo)
eu-west-1	Europa (Irlanda)
me-central-1	Oriente Médio (Emirados Árabes Unidos)
eu-central-1	Europa (Frankfurt)
sa-east-1	América do Sul (São Paulo)
ap-east-1	Ásia-Pacífico (Hong Kong)
ap-south-2	Ásia-Pacífico (Hyderabad)
us-east-1	Leste dos EUA (Norte da Virgínia)
ap-northeast-2	Ásia-Pacífico (Seul)
ap-northeast-3	Ásia-Pacífico (Osaka)
eu-west-2	Europa (Londres)
ap-southeast-4	Ásia-Pacífico (Melbourne)
ap-northeast-1	Ásia-Pacífico (Tóquio)
us-west-2	Oeste dos EUA (Oregon)
us-west-1	Oeste dos EUA (N. da Califórnia)
ap-southeast-1	Ásia-Pacífico (Singapura)
ap-southeast-2	Ásia-Pacífico (Sydney)
il-central-1	Israel (Tel Aviv)

AWS Região	Nome da região
ca-central-1	Canadá (Central)
ca-west-1	Oeste do Canadá (Calgary)
eu-south-2	Europa (Espanha)
cn-northwest-1	China (Ningxia)
eu-central-2	Europa (Zurique)
us-gov-east-1	AWS GovCloud (Leste dos EUA)
us-gov-west-1	AWS GovCloud (Oeste dos EUA)

Realize testes de resiliência com AWS Fault Injection Service

AWS Fault Injection Service é um serviço totalmente gerenciado que ajuda você a realizar experimentos de injeção de falhas em suas AWS cargas de trabalho. AWS FIS a integração com o Amazon Kinesis Data Streams permite que você teste a resiliência do seu aplicativo contra erros comuns da API do Amazon Kinesis Data Streams em um ambiente controlado. Esse recurso permite validar o tratamento de erros, a lógica de novas tentativas e também monitorar os sistemas antes de encontrar falhas. Para obter mais informações, consulte O que é AWS Fault Injection Service? .

Ações

- Erro interno da API: injeta erros internos nas solicitações feitas pelo perfil do IAM de destino. A resposta específica depende de cada serviço e API. A ação `aws:fis:inject-api-internal-error` cria erros `InternalFailure` (HTTP 500).
- Erro de controle de utilização da API: injeta erros internos nas solicitações feitas pelo perfil do IAM de destino. A resposta específica depende de cada serviço e API. A ação `aws:fis:inject-api-throttle-error` cria erros `ThrottlingException` (HTTP 400).
- Erro de API indisponível: injeta erros internos nas solicitações feitas pelo perfil do IAM de destino. A resposta específica depende de cada serviço e API. A ação `aws:fis:inject-api-unavailable-error` cria erros `ServiceUnavailable` (HTTP 503).
- Exceção de throughput provisionada: injeta erros internos nas solicitações feitas pelo perfil do IAM de destino. A resposta específica depende de cada serviço e API. A ação

`aws:kinesis:inject-api-provisioned-throughput-exception` cria erros `ProvisionedThroughputExceededException` (HTTP 400).

- Exceção de iterador expirado da API: injeta erros internos nas solicitações feitas pelo perfil do IAM de destino. A resposta específica depende de cada serviço e API. A ação `aws:kinesis:inject-api-expired-iterator-exception` cria erros `ExpiredIteratorException` (HTTP 400).

Para obter mais informações, consulte Ações do Amazon Kinesis Data Streams.

Considerações

- Você pode usar as ações acima com ofertas provisionadas e sob demanda para o Amazon Kinesis Data Streams.
- Seu fluxo é retomado quando o experimento é concluído, com base na duração selecionada. Também é possível interromper um experimento em andamento antes que ele seja concluído. Como alternativa, você pode definir uma condição de parada para interromper o experimento com base em alarmes que definem a integridade do aplicativo em um Amazon CloudWatch Application Insights.
- Você pode testar até 280 fluxos.

Para obter mais informações sobre suporte regional, consulte Endpoints e quotas do AWS Fault Injection Service.

Erros de exceção de throughput provisionada

Os erros de exceção de throughput provisionado excedido (HTTP 400) ocorrem quando a taxa de solicitação de um fluxo do Kinesis ultrapassa os limites de throughput de um ou mais fragmentos. Cada fragmento tem limites específicos de capacidade de leitura e gravação e exceder esses limites aciona essa exceção. Os cenários que levam a essa exceção incluem: picos repentinos na ingestão ou no consumo de dados, capacidade insuficiente de fragmentos para o volume de dados processado ou distribuição desigual das chaves de partição.

Recomendações para lidar com exceções

- Implemente mecanismos exponenciais de espera e de novas tentativas.
- Aumente o número de fragmentos para acomodar uma maior throughput.
- Assegure que haja uma distribuição adequada das chaves de partição.

- Monitore as métricas do fluxo.

Além disso, o uso do modo de capacidade sob demanda do Kinesis ajuda a ajustar automaticamente workloads e minimizar a ocorrência dessa exceção. Para ter mais informações, consulte O que é o AWS Fault Injection Service?

> **Note**
>
> Os problemas de distribuição imprópria estão fora da capacidade de escalonamento automático no modo sob demanda.

Para realizar um experimento básico

1. Use métricas básicas: registre os padrões de throughput normais antes do teste.
2. Crie um experimento: use a ação `aws:kinesis:inject-api-provisioned-throughput-exception`.
3. Configure a intensidade: comece com um controle de utilização de solicitações de 25%.
4. Monitore as respostas: verifique a lógica de novas tentativas com espera exponencial.
5. Valide o ajuste de escala: confirme se o ajuste de escala automático aciona a ativação.
6. Verifique os alarmes: confira se os alarmes `CloudWatch` estão funcionando conforme o esperado.

As aplicações devem implementar estratégias de espera adequadas, monitorar as métricas `WriteProvisionedThroughputExceeded` e `ReadProvisionedThroughputExceeded` e acionar o ajuste de escala de fragmentos quando apropriado.

Detalhes da ação

- Tipo de recurso: ARN do perfil do IAM
- Operações de destino: `PutRecord`, `PutRecords`, `GetRecords`
- Código de erro: `ProvisionedThroughputExceededException` (HTTP 400)
- Descrição: simula cenários em que a taxa de solicitação excede os limites de capacidade de fragmento, testando o controle de utilização de aplicações e ajustando a escala das respostas.

Parâmetros

- ARN do perfil do IAM: perfil usado por sua aplicação para as operações do Kinesis Data Streams.
- Operações: operações de destino: `PutRecord`, `PutRecords`, `GetRecords`.
- Lista de recursos: nomes específicos do fluxo ou identificadores de fragmentos.
- Duração: duração do experimento, que varia de um minuto a 12 horas. Na AWS FIS API, o valor é uma string no formato ISO 8601. Por exemplo, PT1 M representa um minuto. No AWS FIS console, você insere o número de segundos, minutos ou horas.
- Intensidade: a porcentagem de solicitações a serem submetidas a controle de utilização.

Permissões obrigatórias

- `kinesis:InjectApiError`

Exemplo de modelo de experimento

O exemplo a seguir mostra uma exceção de taxa de transferência provisionada para todas as solicitações de até 5 streams do Kinesis Data com a tag especificada. AWS FIS seleciona os fluxos a serem afetados aleatoriamente. Após 5 minutos, a falha é removida.

```
{
    "description": "Kinesis stream experiment",
    "targets": {
        "KinesisStreams-Target-1": {
            "resourceType": "aws:kinesis:stream",
            "resourceTags": {
                   "tag-key": "tag-value"
            },
            "selectionMode": "COUNT(5)"
        }
    },
    "actions": {
         "kinesis": {
              "actionId": "aws:kinesis:stream-provisioned-throughput-exception",
              "description": "my-stream",
              "parameters": {
                   "duration": "PT5M",
                   "percentage": "100",
                   "service": "kinesis"
              },
```

```
            "targets": {
                  "KinesisStreams": "KinesisStreams-Target-1"
            }
      }
   },
   "stopConditions": [
      {
            "source": "none"
      }
   ],
   "roleArn": "arn:aws:iam::111122223333:role/role-name",
   "tags": {},
   "experimentOptions": {
       "accountTargeting": "single-account",
       "emptyTargetResolutionMode": "fail"
   }
}
```

Exemplo de permissões do perfil do experimento

A permissão a seguir permite que você execute as ações `aws:kinesis:stream-provisioned-throughput-exception` e `aws:kinesis:stream-expired-iterator-exception` em um fluxo específico que afeta 50% das solicitações.

Erros de exceção do iterador expirado

Os erros de exceção do iterador expirado (HTTP 400) ocorrem quando o iterador de fragmento expira e não é mais usado para recuperar os registros do fluxo durante a chamada de `GetRecords`. Isso ocorre quando há atrasos entre as operações de leitura, causados por tarefas de processamento de dados de longa duração, problemas de rede ou tempo de inatividade da aplicação.

> **Note**
>
> Um iterador de fragmento é válido por 5 minutos após o momento em que é emitido.

Recomendações para lidar com exceções

- Atualizar os iteradores de fragmento antes que eles expirem.
- Incorporar o tratamento de erros para obter novos iteradores.

- Utilizar a Kinesis Kinesis Client Library (KCL), que gerencia automaticamente a expiração do iterador de fragmento.

Para obter mais informações, consulte O que é AWS Fault Injection Service?

Para realizar um experimento básico

1. Crie um modelo de experimento: use o AWS FIS console.
2. Selecione a ação: use a ação `aws:kinesis:inject-api-expired-iterator-exception`.
3. Configure os destinos: especifique o perfil do IAM e as operações do Kinesis Data Streams.
4. Defina a duração: comece com 5 a 10 minutos para o teste inicial.
5. Adicionar condições de parada: condições de parada para AWS FIS.
6. Execute o experimento: monitore o comportamento da aplicação.

Detalhes da ação

- Tipo de recurso: ARN do perfil do IAM
- Operações de destino: `GetRecords`
- Código de erro: `ExpiredIteratorException` (HTTP 400)
- Descrição: o iterador fornecido excede o período máximo permitido, simulando cenários em que o processamento de registros é muito lento ou a lógica de verificação falha.

Parâmetros

- ARN do perfil do IAM: perfil usado por sua aplicação para as operações do Kinesis Data Streams.
- Operações: operações de destino: `GetRecords`
- Lista de recursos: os nomes específicos do fluxo ou ARNs.
- Duração: a duração do experimento. Isso é configurável.
- Intensidade: a porcentagem de solicitações a serem submetidas a controle de utilização.

Permissões obrigatórias

- `kinesis:InjectApiError`

Gravar dados no Amazon Kinesis Data Streams

Um produtor é uma aplicação que grava dados no Amazon Kinesis Data Streams. Você pode criar produtores para o Kinesis Data Streams AWS SDK para Java usando a e a Kinesis Producer Library (KPL).

Se este é o primeiro contato com o Kinesis Data Streams, comece familiarizando-se com os conceitos e a terminologia apresentados em O que é o Amazon Kinesis Data Streams? e Use o AWS CLI para realizar operações do Amazon Kinesis Data Streams.

> **Important**
>
> O Kinesis Data Streams oferece suporte a alterações do período de retenção do registro de dados no fluxo de dados. Para obter mais informações, consulte Alterar o período de retenção de dados.

Para colocar dados no fluxo, é necessário especificar o nome do fluxo, uma chave de partição e o blob de dados que serão adicionados ao fluxo. A chave de partição é usada para determinar em que fragmento do fluxo o registro de dados será adicionado.

Todos os dados no fragmento são enviados para o mesmo operador que está processando o fragmento. A chave de partição usada usa depende da lógica do aplicativo. O número de chaves de partição normalmente deve ser muito maior que o número de fragmentos. Isso ocorre porque a chave de partição é usada para determinar como mapear um registro de dados para um determinado fragmento. Se houver um número suficiente de chaves de partição, os dados podem ser distribuídos uniformemente pelos fragmentos de um fluxo.

Tópicos

- Desenvolver produtores usando a Amazon Kinesis Producer Library (KPL)
- Desenvolva produtores usando a API Amazon Kinesis Data Streams com o AWS SDK para Java
- Gravar no Amazon Kinesis Data Streams usando o Kinesis Agent
- Grave no Kinesis Data Streams usando outros serviços AWS
- Gravar no Kinesis Data Streams usando integrações de terceiros
- Solução de problemas de produtores do Amazon Kinesis Data Streams
- Otimizar os produtores do Kinesis Data Streams

Desenvolver produtores usando a Amazon Kinesis Producer Library (KPL)

Um produtor do Amazon Kinesis Data Streams é uma aplicação que coloca registros de dados de usuários em um fluxo de dados do Kinesis (o que também é chamado de ingestão de dados). A Amazon Kinesis Producer Library (KPL) simplifica o desenvolvimento de aplicações de produção, permitindo que os desenvolvedores atinjam alto throughput de gravação para um fluxo de dados do Kinesis.

Você pode monitorar o KPL com a Amazon CloudWatch. Para obter mais informações, consulte Monitore a Kinesis Producer Library com a Amazon CloudWatch.

Tópicos

- Analisar a função da KPL
- Perceber as vantagens de usar a KPL
- Entender quando não usar a KPL
- Instalar a KPL
- Migrar da KPL 0.x para a KPL 1.x
- Transição para certificados Amazon Trust Services (ATS) para a KPL
- Plataformas compatíveis com a KPL
- Principais conceitos da KPL
- Integrar a KPL com o código de produtor
- Gravar no fluxo de dados do Kinesis usando a KPL
- Configurar a Amazon Kinesis Producer Library
- Implementar a desagregação de consumidores
- Usar a KPL com o Amazon Data Firehose
- Use o KPL com o Registro do AWS Glue Esquema
- Definir a configuração do proxy da KPL
- Política de ciclo de vida da versão da KPL

> **Note**
>
> Recomendamos a atualização para a versão mais recente da KPL. A KPL é atualizada regularmente em versões que incluem os patches de dependência e segurança e as correções de bugs mais recentes, além de novos recursos compatíveis com versões anteriores. Para obter mais informações, consulte https://github.com/awslabs/amazon-kinesis-producer/releases/.

Analisar a função da KPL

A KPL é uma easy-to-use biblioteca altamente configurável que ajuda você a gravar em um stream de dados do Kinesis. Ela atua como um intermediário entre o código da aplicação de produção e as ações das APIs do Kinesis Data Streams. A KPL executa as seguintes tarefas principais:

- Grava em um ou mais fluxos de dados do Kinesis com um mecanismo automático e configurável de novas tentativas
- Coleta registros e usa `PutRecords` para gravar vários registros em vários fragmentos por solicitação
- Agrega registros de usuário para aumentar o tamanho da carga útil e melhorar o throughput
- Integra-se perfeitamente à Kinesis Client Library (KCL) para desagregar registros em lote no consumidor
- Envia CloudWatch métricas da Amazon em seu nome para fornecer visibilidade sobre o desempenho do produtor

Observe que o KPL é diferente da API Kinesis Data Streams que está disponível no. AWS SDKs A API do Kinesis Data Streams ajuda a gerenciar vários aspectos do Kinesis Data Streams (incluindo a criação de fluxos, a refragmentação e a colocação e obtenção de registros), enquanto a KPL fornece uma camada de abstração especificamente para a ingestão de dados. Para obter informações sobre a API do Kinesis Data Streams, consulte a Referência de APIs do Amazon Kinesis.

Perceber as vantagens de usar a KPL

A lista a seguir apresenta algumas das principais vantagens no uso da KPL para desenvolver produtores do Kinesis Data Streams.

A KPL pode ser usada em casos de uso síncronos ou assíncronos. Recomenda-se a interface assíncrona, que possui maior desempenho, a menos que haja um motivo específico para usar o comportamento síncrono. Para obter mais informações sobre esses dois casos de uso e o código de exemplo, consulte Gravar no fluxo de dados do Kinesis usando a KPL.

Benefícios de desempenho

A KPL pode ajudar a criar produtores de alta performance. Considere uma situação em que suas instâncias do Amazon EC2 atuam como proxy para coletar eventos de 100 bytes de centenas ou milhares de dispositivos de baixa potência e gravar registros em um fluxo de dados do Kinesis. Cada instância do EC2 precisa gravar milhares de eventos por segundo no fluxo de dados. Para alcançar o throughput necessário, os produtores precisam implementar uma lógica complexa, como agrupamento em lotes ou multithreading, lógica de retentativa e desagregação de registros no lado do consumidor. A KPL executa todas essas tarefas.

Facilidade de uso do lado do consumidor

No caso de desenvolvedores no lado do consumidor usando a KCL em Java, a integração da KPL não requer esforço adicional. Ao recuperar um registro agregado do Kinesis Data Streams que consiste em vários registros de usuário da KPL, a KCL chama automaticamente a KPL para extrair registros do usuário individual e retorná-los ao mesmo usuário.

Desenvolvedores no lado do conusmidor que não usam a KCL, mas usam a operação `GetRecords` da API diretamente, contam com uma biblioteca de Java da KPL disponível para extrair registros do usuário individual e retorná-los ao mesmo usuário.

Monitoramento de produtor

Você pode coletar, monitorar e analisar seus produtores do Kinesis Data Streams usando a CloudWatch Amazon e a KPL. O KPL emite taxa de transferência, erro e outras métricas CloudWatch em seu nome e é configurável para ser monitorado no nível de stream, fragmento ou produtor.

Arquitetura assíncrona

Como pode armazenar registros em buffer antes de enviá-los ao Kinesis Data Streams, a KPL não força a aplicação chamadora a bloquear e aguardar confirmação de que o registro chegou ao servidor antes de continuar o runtime. Uma chamada para colocar um registro na KPL sempre retorna imediatamente, sem esperar o registro ser enviado ou uma resposta ser recebida do servidor. Em vez disso, é criado um objeto `Future` que posteriormente recebe o resultado do envio do registro ao Kinesis Data Streams. Esse é o mesmo comportamento dos clientes assíncronos no SDK. AWS

Entender quando não usar a KPL

A KPL pode ter um atraso por processamento adicional de até `RecordMaxBufferedTime` dentro da biblioteca (configurável pelo usuário). Valores maiores que `RecordMaxBufferedTime` geram maiores eficiências de empacotamento e melhor desempenho. Aplicações que não toleram esse atraso adicional podem precisar usar o SDK da AWS diretamente. Para obter mais informações sobre como usar o AWS SDK com o Kinesis Data Streams, consulte. Desenvolva produtores usando a API Amazon Kinesis Data Streams com o AWS SDK para Java Para obter mais informações sobre `RecordMaxBufferedTime` e outras propriedades configuráveis pelo usuário da KPL, consulte Configurar a Amazon Kinesis Producer Library.

Instalar a KPL

A Amazon fornece binários pré-compilados da Amazon Kinesis Producer Library (KPL) de C++ para macOS, Windows e distribuições recentes do Linux (para obter detalhes das plataformas com suporte, consulte a próxima seção). Esses binários, empacotados como parte dos arquivos .jar do Java, são invocados e usados automaticamente ao usar o Maven para instalar o pacote. Para localizar as versões mais recentes da KPL e da KCL, use os seguintes links de pesquisa do Maven:

- KPL
- KCL

O binários do Linux foram compilados com GNU Compiler Collection (GCC) e vinculados estaticamente à libstdc++ no Linux. Espera-se que eles funcionem em qualquer distribuição do Linux de 64 bits que inclua um glibc versão 2.5 ou superior.

Usuários de distribuições Linux anteriores podem criar o KPL usando as instruções de compilação fornecidas junto com o código-fonte ativado. GitHub Para baixar o KPL de GitHub, consulte a Amazon Kinesis Producer Library.

> **Important**
>
> A Amazon Kinesis Producer Library (KPL) 0.x chegará em 30 de end-of-support janeiro de 2026. É altamente recomendável que você migre suas aplicações da KPL que usam a versão 0.x para a versão mais recente da KPL antes de 30 de janeiro de 2026. Para saber a versão mais recente da KPL, consulte a página da KPL no Github. Para ter mais informações sobre como migrar da KPL 0.x para a KPL 1.x, consulte Migrar da KPL 0.x para a KPL 1.x.

Migrar da KPL 0.x para a KPL 1.x

Este tópico fornece step-by-step instruções para migrar seu consumidor do KPL 0.x para o KPL 1.x. O KPL 1.x introduz suporte para o AWS SDK para Java 2.x, mantendo a compatibilidade da interface com as versões anteriores. Não é necessário atualizar sua lógica de processamento dos dados principais para migrar para a KPL 1.x.

1. É necessário atender aos seguintes pré-requisitos:

   - Java Development Kit (JDK) 8 ou posterior
   - AWS SDK para Java 2. x
   - Maven ou Gradle para gerenciamento de dependências

2. Adicionar dependências

   Se estiver usando Maven, adicione a dependência a seguir ao arquivo pom.xml. Atualize groupid de `com.amazonaws` para `software.amazon.kinesis` e a versão `1.x.x` para a versão mais recente da KPL.

   ```
   <dependency>
       <groupId>software.amazon.kinesis</groupId>
       <artifactId>amazon-kinesis-producer</artifactId>
       <version>1.x.x</version> <!-- Use the latest version -->
   </dependency>
   ```

   Se você estiver usando Gradle, adicione o seguinte ao seu arquivo `build.gradle`. Substitua `1.x.x` pela versão mais recente da KPL.

   ```
   implementation 'software.amazon.kinesis:amazon-kinesis-producer:1.x.x'
   ```

   A versão mais recente da KPL pode ser obtida no Repositório central do Maven.

3. Atualizar as instruções de importação para a KPL

   O KPL 1.x usa o AWS SDK para Java 2.x e usa um nome de pacote atualizado que começa com`software.amazon.kinesis`, em comparação com o nome do pacote no KPL anterior que começa com. `com.amazonaws.services.kinesis`

Substitua a importação de `com.amazonaws.services.kinesis` por `software.amazon.kinesis`. A tabela a seguir relaciona as importações que precisam ser substituídas.

Substituições de importações

Substitua:	Por:
importar com.amazonaws.services.kinesis.producer.Attempt;	importar software.amazon.kinesis.producer.Attempt;
importe com.amazonaws.services.kinesis.producer. BinaryToHexConverter;	importar software.amazon.kinesis.producer.BinaryToHexConverter;
importe com.amazonaws.services.kinesis.producer. CertificateExtractor;	importar software.amazon.kinesis.producer.CertificateExtractor;
importar com.amazonaws.services.kinesis.producer.Daemon;	importar software.amazon.kinesis.producer.Daemon;
importe com.amazonaws.services.kinesis.producer. DaemonException;	importar software.amazon.kinesis.producer.DaemonException;
importe com.amazonaws.services.kinesis.producer. FileAgeManager;	importar software.amazon.kinesis.producer.FileAgeManager;
importe com.amazonaws.services.kinesis.producer. FutureTimedOutException;	importar software.amazon.kinesis.producer.FutureTimedOutException;
importe com.amazonaws.services.kinesis.producer. GlueSchemaRegistrySerializerInstance;	importar software.amazon.kinesis.producer.GlueSchemaRegistrySerializerInstance;
importe com.amazonaws.services.kinesis.producer. HashedFileCopier;	importar software.amazon.kinesis.producer.HashedFileCopier;
importe com.amazonaws.services.kinesis.producer. IKinesisProdutor;	importar software.amazon.kinesis.producer.IKinesisProdutor;

Substitua:	Por:
importe com.amazonaws.services.kinesis.producer. IrrecoverableError;	importar software.amazon.kinesis.producer. IrrecoverableError;
importe com.amazonaws.services.kinesis.producer. KinesisProducer;	importar software.amazon.kinesis.producer. KinesisProducer;
importe com.amazonaws.services.kinesis.producer. KinesisProducerConfiguration;	importar software.amazon.kinesis.producer. KinesisProducerConfiguration;
importe com.amazonaws.services.kinesis.producer. LogInputStreamReader;	importar software.amazon.kinesis.producer. LogInputStreamReader;
importar com.amazonaws.services.kinesis.producer.Metric;	importar software.amazon.kinesis.producer.Metric;
importe com.amazonaws.services.kinesis.producer. ProcessFailureBehavior;	importar software.amazon.kinesis.producer. ProcessFailureBehavior;
importe com.amazonaws.services.kinesis.producer. UnexpectedMessageException;	importar software.amazon.kinesis.producer. UnexpectedMessageException;
importe com.amazonaws.services.kinesis.producer. UserRecord;	importar software.amazon.kinesis.producer. UserRecord;
importe com.amazonaws.services.kinesis.producer. UserRecordFailedException;	importar software.amazon.kinesis.producer. UserRecordFailedException;
importe com.amazonaws.services.kinesis.producer. UserRecordResult;	importar software.amazon.kinesis.producer. UserRecordResult;
importar com.amazonaws.services.kinesis.producer.protobuf.Messages;	importar software.amazon.kinesis.producer.protobuf.Messages;
importar com.amazonaws.services.kinesis.producer.protobuf.Config;	importar software.amazon.kinesis.producer.protobuf.Config;

4. Atualizar declarações de importação para classes de provedores AWS de credenciais

Ao migrar para o KPL 1.x, você deve atualizar pacotes e classes em suas importações no código do aplicativo KPL que são baseados no AWS SDK para Java 1.x para os correspondentes com base no 2.x. AWS SDK para Java As importações comuns na aplicação KPL são classes de provedores de credenciais. Consulte Alterações no provedor de credenciais na documentação do guia de migração AWS SDK para Java 2.x para ver a lista completa de alterações no provedor de credenciais. Aqui está a alteração comum de importação que talvez você precise fazer em suas aplicações KPL.

Importação na KPL 0.x

```
import com.amazonaws.auth.DefaultAWSCredentialsProviderChain;
```

Importação na KPL 1.x

```
import software.amazon.awssdk.auth.credentials.DefaultCredentialsProvider;
```

Se você importar outros provedores de credenciais com base no AWS SDK para Java 1.x, deverá atualizá-los para os equivalentes ao AWS SDK para Java 2.x. Se você não importou nada classes/packages do AWS SDK para Java 1.x, pode ignorar essa etapa.

5. Atualize a configuração do provedor de credenciais na configuração da KPL

   A configuração do provedor de credenciais na KPL 1.x requer os provedores de credenciais do AWS SDK para Java 2.x. Se você estiver passando provedores de credenciais para o AWS SDK para Java 1.x no `KinesisProducerConfiguration` substituindo o provedor de credenciais padrão, você deverá atualizá-lo com os AWS SDK para Java provedores de credenciais 2.x. Consulte Alterações no provedor de credenciais na documentação do guia de migração AWS SDK para Java 2.x para ver a lista completa de alterações no provedor de credenciais. Se não substituiu o provedor de credenciais padrão na configuração da KPL, você pode ignorar essa etapa.

   Por exemplo, se você estiver substituindo o provedor de credenciais padrão da KPL com o seguinte código:

```
KinesisProducerConfiguration config = new KinesisProducerConfiguration();
// SDK v1 default credentials provider
config.setCredentialsProvider(new DefaultAWSCredentialsProviderChain());
```

É necessário atualizá-lo com o código a seguir para usar o provedor de credenciais do AWS SDK para Java 2.x:

```
KinesisProducerConfiguration config = new KinesisProducerConfiguration();
// New SDK v2 default credentials provider
config.setCredentialsProvider(DefaultCredentialsProvider.create());
```

Transição para certificados Amazon Trust Services (ATS) para a KPL

Em 9 de fevereiro de 2018, às 9:00, horário do Pacífico, o Amazon Kinesis Data Streams instalou os certificados do ATS. Para poder continuar a gravar registros no Kinesis Data Streams usando a Amazon Kinesis Producer Library (KPL), é necessário atualizar a instalação da KPL para a versão 0.12.6 ou posterior. Essa mudança afeta todas as AWS regiões.

Para obter informações sobre a mudança para o ATS, consulte Como se preparar para AWS a mudança para sua própria autoridade de certificação.

Se encontrar problemas e precisar de suporte técnico, abra um caso na central de suporte da AWS .

Plataformas compatíveis com a KPL

A Amazon Kinesis Producer Library (KPL) é escrita em C++ e executada como um processo secundário do processo principal do usuário. Binários nativos pré-compilados de 64 bits são fornecidos com a versão do Java e gerenciados pelo wrapper Java.

O pacote Java é executado sem a necessidade de instalar qualquer biblioteca adicional nos seguintes sistemas operacionais:

- Distribuições do Linux com kernel 2.6.18 (setembro de 2006) e posterior
- Apple iOS X 10.9 e posterior
- Windows Server 2008 e posterior

> **Important**
>
> O Windows Server 2008 e posterior é compatível com todas as versões do KPL até a versão 0.14.0.
> A plataforma Windows NÃO é compatível a partir da versão KPL 0.14.0 ou superior.

Observe que a KPL é de 64 bits apenas.

Código-fonte

Se os binários fornecidos na instalação da KPL não forem suficientes para seu ambiente, o núcleo da KPL é escrito como um módulo C++. O código-fonte do módulo C++ e da interface Java é lançado sob a Amazon Public License e está disponível GitHub na Amazon Kinesis Producer Library. Embora seja possível usar a KPL em qualquer plataforma para a qual haja disponibilidade de JRE e um compilador C++ compatível com padrões recentes, a Amazon não oferece suporte oficial a nenhuma plataforma que não esteja na lista de plataformas compatíveis.

Principais conceitos da KPL

As seções a seguir contêm os conceitos e a terminologia necessários para entender e aproveitar a Amazon Kinesis Producer Library (KPL).

Tópicos

- Registros
- Agrupamento em lotes
- Agregação
- Coleta

Registros

Neste guia, há uma distinção entre registros de usuário da KPL e registros do Kinesis Data Streams. O termo registro sem um qualificador refere-se a um registro de usuário da KPL. Um registro do Kinesis Data Streams será explicitamente chamado de registro do Kinesis Data Streams.

Um registro de usuário da KPL é um blob de dados com um significado específico para o usuário. Os exemplos incluem um blob JSON que representa um evento de interface do usuário em um site ou uma entrada de log proveniente de um servidor da web.

Um registro do Kinesis Data Streams é uma instância da estrutura de dados de `Record` definida pela API do serviço Kinesis Data Streams. Ele contém uma chave de partição, um número sequencial e um blob de dados.

Agrupamento em lotes

Envio em lotes refere-se à execução de uma única ação em vários itens, em vez de executar a ação repetidamente em cada item.

Nesse contexto, o “item” é um registro, e a ação é seu envio ao Kinesis Data Streams. Em uma situação sem envio em lotes, cada registro é colocado em um registro separado do Kinesis Data Streams e faz uma única solicitação HTTP para enviá-lo ao Kinesis Data Streams. Com o agrupamento em lotes, cada solicitação HTTP pode carregar vários registros, em vez de apenas um.

A KPL aceita dois tipos de agrupamento em lotes:

- Agregação: armazenamento de vários registros em um único registro do Kinesis Data Streams.
- Coleta: uso da operação `PutRecords` da API para enviar vários registros do Kinesis Data Streams a um ou mais fragmentos no fluxo de dados do Kinesis.

Os dois tipos de agrupamento em lotes da KPL são projetados para coexistir e podem ser ativados ou desativados de forma independente. Por padrão, ambos são ativados.

Agregação

A agregação refere-se ao armazenamento de vários registros em um registro do Kinesis Data Streams. A agregação permite que os clientes aumentem o número de registros enviados por chamada de API, o que aumenta efetivamente o throughput do produtor.

Os fragmentos do Kinesis Data Streams oferecem suporte a até 1.000 registros do Kinesis Data Streams por segundo, ou um throughput de 1 MB. O limite de registros por segundo do Kinesis Data Streams restringe clientes com registros menores que 1 KB. A agregação de registros permite aos clientes combinar vários registros em um único registro do Kinesis Data Streams. Isso permite que os clientes melhorem o próprio throughput por fragmento.

Considere o caso de um fragmento na região us-east-1 que está atualmente em execução à taxa constante de 1.000 registros por segundo, com registros de 512 bytes cada. Com a agregação da KPL, é possível empacotar 1.000 registros em apenas 10 registros do Kinesis Data Streams reduzindo o número de registros por segundo para 10 (a 50 KB cada).

Coleta

Coleta refere-se ao agrupamento de vários registros do Kinesis Data Streams em um lote, que é enviado em uma única solicitação HTTP com uma chamada à operação `PutRecords` da API, em vez de enviar cada registro do Kinesis Data Streams em uma solicitação HTTP própria.

Isso aumenta o throughput em comparação com a não utilização de coleções, pois reduz a sobrecarga de fazer muitas solicitações HTTP separadas. Na verdade, `PutRecords`, por si só, foi projetado especificamente para essa finalidade.

A coleta difere da agregação porque opera com grupos de registros do Kinesis Data Streams. Os registros do Kinesis Data Streams coletados ainda podem conter vários registros do usuário. O relacionamento pode ser visualizado da seguinte maneira:

Integrar a KPL com o código de produtor

A Amazon Kinesis Producer Library (KPL) é executada em um processo separado e se comunica com o processo principal do usuário usando IPC. Essa arquitetura, às vezes chamada de microsserviço, é escolhida por dois motivos principais:

1) O processo do usuário não falhará mesmo que a KPL falhe

O processo pode ter tarefas não relacionadas ao Kinesis Data Streams e continuar em operação mesmo que a KPL falhe. Além disso, o processo de usuário pai pode reiniciar a KPL e recuperar um estado totalmente funcional (essa funcionalidade está nos wrappers oficiais).

Um exemplo é um servidor Web que envia métricas ao Kinesis Data Streams. O servidor pode continuar entregando páginas mesmo que a parte do Kinesis Data Streams tenha parado de funcionar. Portanto, causar uma falha em todo o servidor devido a um erro na KPL seria uma interrupção desnecessária.

2) Clientes arbitrários podem ser aceitos

Há sempre clientes que usam linguagens diferentes das oficialmente aceitas. Esses clientes também devem ser capazes de usar a KPL facilmente.

Matriz de uso recomendado

A matriz de uso a seguir lista as configurações recomendadas para diferentes usuários e indica se e como a KPL deve ser usada. Lembre-se de que, se a agregação estiver habilitada, será preciso usar a desagregação para extrair seus registros no lado do consumidor.

Linguagem do lado do produtor	Linguagem do lado do consumidor	Versão da KCL	Lógica do ponto de verificação	A KPL pode ser usada?	Advertências
Tudo menos Java	*	*	*	Não	N/D
Java	Java	Usa o Java SDK diretamente	N/D	Sim	Se a agregação for usada, será necessári o usar a biblioteca de desagrega ção fornecida após as chamadas a

Linguagem do lado do produtor	Linguagem do lado do consumidor	Versão da KCL	Lógica do ponto de verificação	A KPL pode ser usada?	Advertências
					`GetRecords`.
Java	Tudo menos Java	Usa o SDK diretamente	N/D	Sim	É preciso desabilitar a agregação.
Java	Java	1.3.x	N/D	Sim	É preciso desabilitar a agregação.
Java	Java	1.4.x	Chama o ponto de verificação sem nenhum argumento	Sim	Nenhum
Java	Java	1.4.x	Chama o ponto de verificação com um número sequencial explícito	Sim	Desative a agregação ou altere o código para usar números sequenciais estendidos para definir pontos de verificação.

Linguagem do lado do produtor	Linguagem do lado do consumidor	Versão da KCL	Lógica do ponto de verificação	A KPL pode ser usada?	Advertências
Java	Tudo menos Java	1.3.x + daemon de várias linguagens + wrapper específico de linguagem	N/D	Sim	É preciso desabilitar a agregação.

Gravar no fluxo de dados do Kinesis usando a KPL

As seções a seguir mostram o código de exemplo em uma progressão desde o produtor mais básico até o código totalmente assíncrono.

Código barebone de produtor

O código a seguir é todo o necessário para escrever um produtor minimamente funcional. Os registros de usuário da Amazon Kinesis Producer Library (KPL) são processados em segundo plano.

```
// KinesisProducer gets credentials automatically like
// DefaultAWSCredentialsProviderChain.
// It also gets region automatically from the EC2 metadata service.
KinesisProducer kinesis = new KinesisProducer();
// Put some records
for (int i = 0; i < 100; ++i) {
    ByteBuffer data = ByteBuffer.wrap("myData".getBytes("UTF-8"));
    // doesn't block
    kinesis.addUserRecord("myStream", "myPartitionKey", data);
}
// Do other stuff ...
```

Responder aos resultados de forma síncrona

No exemplo anterior, o código não verificou se os registros do usuário da KPL foram bem-sucedidos. A KPL faz todas as novas tentativas necessárias para dar conta das falhas. No entanto, para verificar

os resultados, é possível utilizar os objetos `Future` que são retornados de `addUserRecord`, como no exemplo a seguir (exemplo anterior mostrado para fins de contexto):

```
KinesisProducer kinesis = new KinesisProducer();

// Put some records and save the Futures
List<Future<UserRecordResult>> putFutures = new
 LinkedList<Future<UserRecordResult>>();
for (int i = 0; i < 100; i++) {
    ByteBuffer data = ByteBuffer.wrap("myData".getBytes("UTF-8"));
    // doesn't block
    putFutures.add(
        kinesis.addUserRecord("myStream", "myPartitionKey", data));
}

// Wait for puts to finish and check the results
for (Future<UserRecordResult> f : putFutures) {
    UserRecordResult result = f.get(); // this does block
    if (result.isSuccessful()) {
        System.out.println("Put record into shard " +
                           result.getShardId());
    } else {
        for (Attempt attempt : result.getAttempts()) {
            // Analyze and respond to the failure
        }
    }
}
```

Responder aos resultados de forma assíncrona

O exemplo anterior está chamando `get()` em um objeto `Future`, o que bloqueia o runtime. Se não quiser bloquear o runtime, você pode usar callback assíncrono, como mostrado no exemplo a seguir:

```
KinesisProducer kinesis = new KinesisProducer();

FutureCallback<UserRecordResult> myCallback = new FutureCallback<UserRecordResult>() {

    @Override public void onFailure(Throwable t) {
        /* Analyze and respond to the failure  */
    };
    @Override public void onSuccess(UserRecordResult result) {
        /* Respond to the success */
    };
```

```
};

for (int i = 0; i < 100; ++i) {
    ByteBuffer data = ByteBuffer.wrap("myData".getBytes("UTF-8"));
    ListenableFuture<UserRecordResult> f = kinesis.addUserRecord("myStream",
 "myPartitionKey", data);
    // If the Future is complete by the time we call addCallback, the callback will be
 invoked immediately.
    Futures.addCallback(f, myCallback);
}
```

Configurar a Amazon Kinesis Producer Library

Embora as configurações padrão devam funcionar bem para a maioria dos casos de uso, talvez convenha alterar algumas configurações padrão para ajustar o comportamento do `KinesisProducer` às suas necessidades. Para isso, uma instância da classe `KinesisProducerConfiguration` pode ser passada ao construtor `KinesisProducer`, por exemplo:

```
KinesisProducerConfiguration config = new KinesisProducerConfiguration()
        .setRecordMaxBufferedTime(3000)
        .setMaxConnections(1)
        .setRequestTimeout(60000)
        .setRegion("us-west-1");

final KinesisProducer kinesisProducer = new KinesisProducer(config);
```

Também é possível pode carregar uma configuração de um arquivo de propriedades:

```
KinesisProducerConfiguration config =
 KinesisProducerConfiguration.fromPropertiesFile("default_config.properties");
```

É possível substituir qualquer caminho e nome de arquivo a que o processo de usuário tem acesso. Também é possível chamar métodos definidos para a instância de `KinesisProducerConfiguration` criada dessa forma para personalizar a configuração.

O arquivo de propriedades deve especificar parâmetros usando seus nomes em PascalCase. Os nomes correspondem aos usados nos métodos definidos na classe `KinesisProducerConfiguration`. Por exemplo:

```
RecordMaxBufferedTime = 100
```

```
MaxConnections = 4
RequestTimeout = 6000
Region = us-west-1
```

Para obter mais informações sobre regras de uso de parâmetros de configuração e limites de valor, consulte o arquivo de exemplo de propriedades de configuração em GitHub.

Observe que, depois que o `KinesisProducer` é inicializado, alterar a instância de `KinesisProducerConfiguration` que foi usada não tem mais efeito. No momento, o `KinesisProducer` não oferece suporte à reconfiguração dinâmica.

Implementar a desagregação de consumidores

A partir da versão 1.4.0, a KCL oferece suporte à desagregação automática dos registros de usuário da KPL. O código da aplicação de consumo escrito com versões anteriores da KCL será compilado sem qualquer modificação após a atualziação da KCL. No entanto, se a agregação da KPL estiver sendo usada no lado do produtor, haverá uma sutileza envolvendo a definição de pontos de verificação: como todos os sub-registros dentro de um registro agregado têm o mesmo número de sequência, os dados adicionais precisarão ser armazenados com o ponto de verificação, se for preciso distinguir os sub-registros. Esses dados adicionais são chamados de números de subsequência.

Opções

- Migrar de versões anteriores da KCL
- Usar extensões da KCL para desagregação da KPL
- Use GetRecords diretamente

Migrar de versões anteriores da KCL

Não é necessário alterar as chamadas existentes para que definam pontos de verificação com a agregação. A recuperação bem-sucedida de todos os registros armazenados no Kinesis Data Streams ainda é garantida. A KCL agora oferece duas novas operações de ponto de verificação para suporte a casos de uso específicos, descritos abaixo.

Caso seu código existente tenha sido escrito para uma versão da KCL anterior ao suporte à KPL e a operação de ponto de verificação seja chamada sem argumentos, isso será equivalente a definir o ponto de verificação do número de sequência do último registro de usuário da KPL no lote. Se a operação de ponto de verificação é chamada com uma string de número sequencial, isso equivale a

definir o ponto de verificação do número sequencial do lote conhecido junto com o número 0 (zero) da subsequência implícita.

Chamar a nova operação de ponto de verificação `checkpoint()` da KCL sem argumentos é semanticamente equivalente a definir o ponto de verificação do número de sequência da última chamada a `Record` no lote com o número 0 (zero) de subsequência implícito.

Chamar a nova operação de ponto de verificação `checkpoint(Record record)` da KCL é semanticamente equivalente a definir o ponto de verificação do número de sequência do `Record` conhecido com o número 0 (zero) de subsequência implícito. Se a chamada a `Record` é na verdade um `UserRecord`, o número de sequencial e subsequencial de `UserRecord` será definido.

Chamar a nova operação de ponto de verificação `checkpoint(String sequenceNumber, long subSequenceNumber)` da KCL explicitamente define o ponto de verificação do número de sequência conhecido com o número de subsequência conhecido.

Em qualquer desses casos, depois que o ponto de verificação é armazenado na tabela de pontos de verificação do Amazon DynamoDB, a KCL pode retomar corretamente a recuperação de registros, mesmo quando a aplicação falha e reinicia. Se houver mais registros contidos na sequência, a recuperação ocorrerá a partir do próximo registro de número subsequencial dentro do registro com o número sequencial cujo ponto de verificação foi definido mais recentemente. Se o ponto de verificação mais recente inclui o último número subsequencial do registro de número sequencial anterior, a recuperação ocorrerá a partir do registro com o próximo número sequencial.

A próxima seção discute detalhes de definição do ponto de verificação de sequência e subsequência para consumidores que precisam evitar a omissão e a duplicação de registros. Se a omissão (ou duplicação) de registros ao parar e reiniciar o processamento de registros do consumidor não é importante, é possível executar seu código existente sem modificação.

Usar extensões da KCL para desagregação da KPL

A desagregação da KPL pode envolver a definição de um ponto de verificação de subsequência. Para facilitar a definição do ponto de verificação de subsequência, uma classe `UserRecord` foi adicionada à KCL:

```
public class UserRecord extends Record {
    public long getSubSequenceNumber() {
    /* ... */
    }
    @Override
    public int hashCode() {
```

```
    /* contract-satisfying implementation */
    }
    @Override
    public boolean equals(Object obj) {
    /* contract-satisfying implementation */
    }
}
```

Essa classe agora é usada em vez de `Record`. Ela não interrompe o código existente por ser uma subclasse de `Record`. A classe `UserRecord` representa os sub-registros reais e os registros padrão não agregados. Os registros não agregados podem ser considerados como registros agregadas com exatamente um sub-registro.

Além disso, duas operações novas são adicionadas a `IRecordProcessorCheckpointer`:

```
public void checkpoint(Record record);
public void checkpoint(String sequenceNumber, long subSequenceNumber);
```

Para começar a usar a definição de ponto de verificação de número subsequencial, é possível executar a seguinte conversão. Altere o seguinte código de formulário:

```
checkpointer.checkpoint(record.getSequenceNumber());
```

Novo código de formulário:

```
checkpointer.checkpoint(record);
```

Recomendamos usar o formulário `checkpoint(Record record)` para definição de ponto de verificação de subsequência. No entanto, se `sequenceNumbers` já estiverem sendo armazenados em strings para usar na definição de pontos de verificação, agora deve-se também armazenar `subSequenceNumber`, como mostrado no exemplo a seguir:

```
String sequenceNumber = record.getSequenceNumber();
long subSequenceNumber = ((UserRecord) record).getSubSequenceNumber();  // ... do other
 processing
checkpointer.checkpoint(sequenceNumber, subSequenceNumber);
```

A transmissão de `Record` para `UserRecord` sempre é bem-sucedida porque a implementação sempre usa `UserRecord`. A menos que haja uma necessidade de executar aritmética nos números sequenciais, essa abordagem não é recomendada.

Durante o processamento de registros de usuário da KPL, a KCL grava o número de subsequência no Amazon DynamoDB como um campo extra para cada linha. As versões anteriores da KCL usavam `AFTER_SEQUENCE_NUMBER` para obter registros ao retomar os pontos de verificação. Em vez disso, a KCL atual com suporte à KPL, usa `AT_SEQUENCE_NUMBER`. Quando é recuperado o registro no número sequencial para o qual foi definido o ponto de verificação, esse número é verificado e os sub-registros são descartados conforme apropriado (que podem ser todos eles, se o último sub-registro é aquele verificado). Novamente, os registros não agregados podem ser considerados como registros agregados com um único sub-registro. Portanto, o mesmo algoritmo funciona para os registros agregados e não agregados.

Use GetRecords diretamente

Em vez de usar a KCL, pode-se optar por invocar a operação de API `GetRecords` diretamente para recuperar os registros do Kinesis Data Streams. Para desempacotar os registros recuperados nos registros de usuário originais da KPL, chame uma das seguintes operações estáticas em `UserRecord.java`:

```
public static List<Record> deaggregate(List<Record> records)

public static List<UserRecord> deaggregate(List<UserRecord> records, BigInteger
 startingHashKey, BigInteger endingHashKey)
```

A primeira operação usa o valor `0` (zero) padrão para `startingHashKey` e o valor `2^128 -1` padrão para `endingHashKey`.

Cada uma dessas operações desagrega a lista de registros conhecidos do Kinesis Data Streams em uma lista de registros de usuário da KPL. Qualquer registro de usuário da KPL com uma chave de hash explícita ou chave de partição fora do intervalo de `startingHashKey` (inclusive) e de `endingHashKey` (inclusive) é descartado da lista de registros retornados.

Usar a KPL com o Amazon Data Firehose

Ao usar a Kinesis Producer Library (KPL) para gravar dados em um fluxo de dados do Kinesis, é possível usar agregação para combinar os registros gravados. Ao usar esse fluxo de dados como origem para o fluxo de entrega do Firehose, O Firehose desagregará os registros antes de entregá-los ao destino. Quando o fluxo de entrega é configurado para transformar os dados, o Firehose desagregará os registros antes de entregá-los ao AWS Lambda. Para obter mais informações, consulte Gravando no Amazon Firehose Usando o Kinesis Data Streams.

Use o KPL com o Registro do AWS Glue Esquema

Você pode integrar seus streams de dados do Kinesis com o AWS Glue Schema Registry. O registro de esquemas do AWS Glue permite detectar, controlar e evoluir esquemas centralmente, ao mesmo tempo que garante que os dados produzidos sejam validados continuamente por um esquema registrado. O esquema define a estrutura e o formato de um registro de dados. Um esquema é uma especificação versionada para publicação, consumo ou datastore confiáveis. O AWS Glue Schema Registry permite que você melhore a qualidade end-to-end dos dados e a governança de dados em seus aplicativos de streaming. Para obter mais informações, consulte Registro de esquemas do AWS Glue. Uma das formas de configurar essa integração é usar as bibliotecas KPL e Kinesis Client Library (KCL) em Java.

> **Important**
>
> Atualmente, a integração do Kinesis Data AWS Glue Streams e do registro de esquemas só é compatível com os streams de dados do Kinesis que usam produtores de KPL implementados em Java. Não há suporte para múltiplas linguagens.

Para obter instruções detalhadas sobre como configurar a integração do Kinesis Data Streams com o Schema Registry usando o KPL, consulte a seção “Interagindo com dados KPL/KCL usando as bibliotecas” em Caso de uso: integração do Amazon Kinesis Data Streams com o Glue Schema Registry. AWS

Definir a configuração do proxy da KPL

Para aplicativos que não podem se conectar diretamente à Internet, todos os clientes do AWS SDK oferecem suporte ao uso de proxies HTTP ou HTTPS. Em um ambiente empresarial típico, todo o tráfego de saída da rede precisa passar por servidores proxy. Se seu aplicativo usa a Kinesis Producer Library (KPL) para coletar e enviar dados AWS em um ambiente que usa servidores proxy, seu aplicativo exigirá a configuração do proxy KPL. A KPL é uma biblioteca de alto nível criada com base no SDK do AWS Kinesis. Ele é dividido em um processo nativo e um wrapper. O processo nativo executa todas as tarefas de processamento e envio de registros, enquanto o wrapper gerencia o processo nativo e se comunica com ele. Para obter mais informações, consulte Implementar Produtores Eficientes e Confiáveis com a Amazon Kinesis Producer Library.

O wrapper é escrito em Java e o processo nativo é escrito em C++ com o uso do SDK do Kinesis. A KPL versão 0.14.7 ou superior é compatível com a configuração de proxy no wrapper Java, que

pode passar todas as configurações de proxy para o processo nativo. Para obter mais informações, consulte https://github.com/awslabs/amazon-kinesis-producer/releases/tag/v0.14.7.

É possível usar o código a seguir para adicionar configurações de proxy às aplicações da KPL.

```
KinesisProducerConfiguration configuration = new KinesisProducerConfiguration();
// Next 4 lines used to configure proxy
configuration.setProxyHost("10.0.0.0"); // required
configuration.setProxyPort(3128); // default port is set to 443
configuration.setProxyUserName("username"); // no default
configuration.setProxyPassword("password"); // no default

KinesisProducer kinesisProducer = new KinesisProducer(configuration);
```

Política de ciclo de vida da versão da KPL

Este tópico descreve a política de ciclo de vida da versão para a Amazon Kinesis Producer Library (KPL). AWS fornece regularmente novos lançamentos para as versões do KPL para oferecer suporte a novos recursos e aprimoramentos, correções de erros, patches de segurança e atualizações de dependências. Recomendamos que você continue up-to-date com as versões do KPL para acompanhar os recursos, as atualizações de segurança e as dependências subjacentes mais recentes. Não recomendamos o uso contínuo de uma versão não compatível da KPL.

O ciclo de vida das principais versões da KPL consiste nas três fases a seguir:

- Disponibilidade geral (GA) — Durante essa fase, a versão principal é totalmente suportada. AWS fornece lançamentos regulares de versões secundárias e de patches que incluem suporte para novos recursos ou atualizações de API para o Kinesis Data Streams, bem como correções de bugs e segurança.
- Modo de manutenção — AWS limita os lançamentos da versão do patch para tratar apenas de correções críticas de bugs e problemas de segurança. A versão principal não receberá atualizações dos novos recursos ou APIs do Kinesis Data Streams.
- E nd-of-support — A versão principal não receberá mais atualizações ou lançamentos. As versões publicadas anteriormente continuarão disponíveis por meio de gerenciadores de pacotes públicos e o código permanecerá ativado GitHub. O uso de uma versão que chegou end-of-support é feito

a critério do usuário. Recomendamos que você faça a atualização para a versão principal mais recente.

Versão principal	Fase atual	Data de lançamento	Data do modo de manutenção	End-of-support encontro
KPL 0.x	Modo de manutenção	2015-06-02	2025-04-17	2026-01-30
KPL 1.x	Disponibilidade geral	2024-12-15	--	--

Desenvolva produtores usando a API Amazon Kinesis Data Streams com o AWS SDK para Java

Você pode desenvolver produtores usando a API Amazon Kinesis Data Streams AWS com o SDK for Java. Se você nunca usou o Kinesis Data Streams, comece familiarizando-se com os conceitos e a terminologia apresentados em O que é o Amazon Kinesis Data Streams? e Use o AWS CLI para realizar operações do Amazon Kinesis Data Streams.

Esses exemplos discutem a API do Kinesis Data Streams e usam o AWS SDK para Java para adicionar (colocar) dados a um fluxo. Contudo, na maioria dos casos de uso, é melhor usar a biblioteca KPL do Kinesis Data Streams. Para obter mais informações, consulte Desenvolver produtores usando a Amazon Kinesis Producer Library (KPL).

O código Java de exemplo neste capítulo demonstra como executar operações básicas da API do Kinesis Data Streams e está dividido logicamente por tipo de operação. Esses exemplos não representam um código pronto para produção, pois não verificam todas as exceções possíveis nem abrangem todas as considerações de segurança ou de performance possíveis. Também é possível chamar a API do Kinesis Data Streams usando outras linguagens de programação. Para obter mais informações sobre todas as opções disponíveis AWS SDKs, consulte Comece a desenvolver com a Amazon Web Services.

Cada tarefa tem pré-requisitos. Por exemplo, não é possível adicionar dados a um fluxo enquanto o fluxo não é criado, o que requer a criação de um cliente. Para obter mais informações, consulte Criar e gerenciar fluxos de dados do Kinesis.

Tópicos

- Adicionar dados a um fluxo
- Interaja com os dados usando o AWS Glue Schema Registry

Adicionar dados a um fluxo

Quando um fluxo é criado, pode-se adicionar dados a ele na forma de registros. Um registro é uma estrutura de dados que contém os dados a serem processados na forma de um blob de dados. Depois de armazenar os dados no registro, o Kinesis Data Streams não inspeciona, interpreta nem altera dados de forma alguma. Cada registro também tem um número sequencial e uma chave de partição associados.

Há duas operações diferentes na API do Kinesis Data Streams que adicionam dados a um fluxo, `PutRecords` e `PutRecord`. A operação `PutRecords` envia vários registros ao fluxo por solicitação HTTP e a operação singular `PutRecord` envia registros ao fluxo um por vez (uma solicitação HTTP separada é necessária para cada registro). Talvez convenha usar `PutRecords` para a maioria dos aplicativos, pois ele atingirá um throughput mais alto por produtor de dados. Para obter mais informações sobre cada uma dessas operações, consulte as subseções abaixo.

Tópicos

- Adicione vários registros com PutRecords
- Adicione um único registro com PutRecord

Sempre tenha em mente que, como a aplicação de origem está adicionando dados ao fluxo usando a API do Kinesis Data Streams, provavelmente há uma ou mais aplicações de consumo processando dados fora do fluxo simultaneamente. Para obter informações sobre como os consumidores obtêm dados usando a API do Kinesis Data Streams, consulte Como obter dados de um fluxo.

> **Important**
>
> Alterar o período de retenção de dados

Adicione vários registros com PutRecords

A operação `PutRecords` envia vários registros ao Kinesis Data Streams em uma única solicitação. Ao usar `PutRecords`, os produtores podem obter um throughput mais alto ao enviar dados para

o fluxo de dados do Kinesis. Cada solicitação `PutRecords` pode oferecer suporte a até 500 registros. Cada registro na solicitação pode ter no máximo 1 MB, até um limite de 5 MB para toda a solicitação, incluindo chaves de partição. Assim como a operação única `PutRecord` descrita abaixo, `PutRecords` usa números de sequência e chaves de partição. No entanto, o parâmetro `PutRecord` de `SequenceNumberForOrdering` não é incluído em uma chamada a `PutRecords`. A operação `PutRecords` tenta processar todos os registros na ordem natural da solicitação.

Cada registro de dados tem um número sequencial exclusivo. O número de sequência é atribuído pelo Kinesis Data Streams depois que `client.putRecords` é chamada para adicionar os registros de dados ao fluxo. Os números sequenciais da mesma chave de partição geralmente aumentam com o tempo: quanto maior o período entre as solicitações `PutRecords`, maiores ficam os números sequenciais.

> **Note**
>
> Os números de sequência não podem ser usados como índices para conjuntos de dados dentro do mesmo fluxo. Para separar logicamente conjuntos de dados, use chaves de partição ou crie um fluxo separado para cada conjunto de dados.

Uma solicitação `PutRecords` pode incluir registros com diferentes chaves de partição. O escopo da solicitação é um fluxo; cada solicitação pode incluir qualquer combinação de chaves de partição e registros, dentro dos limites da solicitação. As solicitações feitas com diferentes chaves de partição a streams com muitos fragmentos diferentes costumam ser mais rápidas do que as solicitações com um pequeno número de chaves de partição para um pequeno número de fragmentos. O número de chaves de partição deve ser muito maior do que o número de fragmentos para reduzir a latência e maximizar o throughput.

PutRecords exemplo

O código a seguir cria 100 registros de dados com chaves de partição sequenciais e os coloca em um fluxo denominado `DataStream`.

```
        AmazonKinesisClientBuilder clientBuilder =
AmazonKinesisClientBuilder.standard();

        clientBuilder.setRegion(regionName);
        clientBuilder.setCredentials(credentialsProvider);
        clientBuilder.setClientConfiguration(config);
```

```
        AmazonKinesis kinesisClient = clientBuilder.build();

        PutRecordsRequest putRecordsRequest  = new PutRecordsRequest();
        putRecordsRequest.setStreamName(streamName);
        List <PutRecordsRequestEntry> putRecordsRequestEntryList  = new ArrayList<>();
        for (int i = 0; i < 100; i++) {
            PutRecordsRequestEntry putRecordsRequestEntry  = new
PutRecordsRequestEntry();

putRecordsRequestEntry.setData(ByteBuffer.wrap(String.valueOf(i).getBytes()));
            putRecordsRequestEntry.setPartitionKey(String.format("partitionKey-%d",
i));
            putRecordsRequestEntryList.add(putRecordsRequestEntry);
        }

        putRecordsRequest.setRecords(putRecordsRequestEntryList);
        PutRecordsResult putRecordsResult  =
kinesisClient.putRecords(putRecordsRequest);
        System.out.println("Put Result" + putRecordsResult);
```

A resposta a `PutRecords` inclui uma matriz de resposta `Records`. Cada registro na matriz de resposta se correlaciona diretamente com um registro na matriz de solicitação por ordenação natural, do início ao fim da solicitação e da resposta. A matriz de resposta de `Records` sempre inclui o mesmo número de registros da matriz de solicitação.

Lidar com falhas ao usar PutRecords

Por padrão, a falha de registros individuais em uma solicitação não interrompe o processamento de registros subsequentes em uma solicitação `PutRecords`. Isso significa que uma matriz de resposta `Records` inclui os registros com processamento bem-sucedido e malsucedido. É preciso detectar os registros com processamento malsucedido e incluí-los em uma chamada subsequente.

Os registros bem-sucedidos incluem os valores `SequenceNumber` e `ShardID`, e os registros malsucedidos incluem os valores `ErrorCode` e `ErrorMessage`. O parâmetro `ErrorCode` reflete o tipo de erro e pode ter um dos seguintes valores: `ProvisionedThroughputExceededException` ou `InternalFailure`. `ErrorMessage` fornece informações mais detalhadas sobre a exceção `ProvisionedThroughputExceededException`, incluindo o ID da conta, o nome do fluxo e o ID do fragmento do registro que foi limitado. O exemplo abaixo tem três registros em uma solicitação `PutRecords`. O segundo registro falha e isso é refletido na resposta.

Example PutRecords Sintaxe da solicitação

```
{
    "Records": [
        {
     "Data": "XzxkYXRhPl8w",
     "PartitionKey": "partitionKey1"
        },
        {
     "Data": "AbceddeRFfg12asd",
     "PartitionKey": "partitionKey1"
        },
        {
     "Data": "KFpcd98*7nd1",
     "PartitionKey": "partitionKey3"
        }
    ],
    "StreamName": "myStream"
}
```

Example PutRecords Sintaxe de resposta

```
{
    "FailedRecordCount”: 1,
    "Records": [
        {
     "SequenceNumber": "21269319989900637946712965403778482371",
     "ShardId": "shardId-000000000001"

        },
        {
     “ErrorCode":”ProvisionedThroughputExceededException”,
     “ErrorMessage": "Rate exceeded for shard shardId-000000000001 in stream
 exampleStreamName under account 111111111111."

        },
        {
     "SequenceNumber": "21269319989999637946712965403778482985",
     "ShardId": "shardId-000000000002"
        }
    ]
}
```

Os registros com processamento malsucedido podem ser incluídos nas solicitações `PutRecords` subsequentes. Primeiro, verifique o parâmetro `FailedRecordCount` no `putRecordsResult` para confirmar se há registros com falha na solicitação. Assim sendo, cada `putRecordsEntry` com um `ErrorCode` que não seja `null` deve ser adicionado a uma solicitação subsequente. Para obter um exemplo desse tipo de handler, consulte o seguinte código.

Example PutRecords manipulador de falhas

```
PutRecordsRequest putRecordsRequest = new PutRecordsRequest();
putRecordsRequest.setStreamName(myStreamName);
List<PutRecordsRequestEntry> putRecordsRequestEntryList = new ArrayList<>();
for (int j = 0; j < 100; j++) {
    PutRecordsRequestEntry putRecordsRequestEntry = new PutRecordsRequestEntry();
    putRecordsRequestEntry.setData(ByteBuffer.wrap(String.valueOf(j).getBytes()));
    putRecordsRequestEntry.setPartitionKey(String.format("partitionKey-%d", j));
    putRecordsRequestEntryList.add(putRecordsRequestEntry);
}

putRecordsRequest.setRecords(putRecordsRequestEntryList);
PutRecordsResult putRecordsResult = amazonKinesisClient.putRecords(putRecordsRequest);

while (putRecordsResult.getFailedRecordCount() > 0) {
    final List<PutRecordsRequestEntry> failedRecordsList = new ArrayList<>();
    final List<PutRecordsResultEntry> putRecordsResultEntryList =
 putRecordsResult.getRecords();
    for (int i = 0; i < putRecordsResultEntryList.size(); i++) {
        final PutRecordsRequestEntry putRecordRequestEntry =
 putRecordsRequestEntryList.get(i);
        final PutRecordsResultEntry putRecordsResultEntry =
 putRecordsResultEntryList.get(i);
        if (putRecordsResultEntry.getErrorCode() != null) {
            failedRecordsList.add(putRecordRequestEntry);
        }
    }
    putRecordsRequestEntryList = failedRecordsList;
    putRecordsRequest.setRecords(putRecordsRequestEntryList);
    putRecordsResult = amazonKinesisClient.putRecords(putRecordsRequest);
}
```

Adicione um único registro com PutRecord

Cada chamada para `PutRecord` opera em um único registro. Prefira a operação `PutRecords` descrita em Adicione vários registros com PutRecords, a menos que seu aplicativo precise especificamente enviar sempre registos únicos por solicitação ou algum outro motivo para o não uso de `PutRecords`.

Cada registro de dados tem um número sequencial exclusivo. O número de sequência é atribuído pelo Kinesis Data Streams depois que `client.putRecord` é chamada para adicionar o registro de dados ao fluxo. Os números sequenciais da mesma chave de partição geralmente aumentam com o tempo: quanto maior o período entre as solicitações `PutRecord`, maiores ficam os números sequenciais.

Quando ocorrem colocações em rápida sucessão, não há garantia de que os números de sequência retornados aumentem, porque as operações put aparentam ser essencialmente simultâneas para o Kinesis Data Streams. Para garantir estritamente o aumento de números sequenciais para a mesma chave de partição, use o parâmetro `SequenceNumberForOrdering`, como mostrado no código de exemplo em PutRecord exemplo.

Usando ou não `SequenceNumberForOrdering`, os registros que o Kinesis Data Streams recebe por meio de uma chamada a `GetRecords` são estritamente ordenados por número de sequência.

> **Note**
>
> Os números de sequência não podem ser usados como índices para conjuntos de dados dentro do mesmo fluxo. Para separar logicamente conjuntos de dados, use chaves de partição ou crie um fluxo separado para cada conjunto de dados.

Uma chave de partição é usada para agrupar os dados dentro de um fluxo. Um registro de dados é atribuído a um fragmento dentro do fluxo com base em sua chave de partição. Especificamente, o Kinesis Data Streams usa a chave de partição como entrada para uma função de hash que mapeia essa chave (e os dados associados) a um determinado fragmento.

Como resultado desse mecanismo de hashing, todos os registros de dados com a mesma chave de partição são mapeados para o mesmo fragmento no fluxo. No entanto, se o número de chaves de partição ultrapassar o número de fragmentos, alguns fragmentos conterão necessariamente registros com chaves de partição diferentes. Do ponto de vista do design, para garantir que todos

os seus fragmentos sejam bem utilizados, o número de fragmentos (especificado pelo método `setShardCount` de `CreateStreamRequest`) deve ser substancialmente menor que o número de chaves de partição exclusivas, e o volume de dados que flui para uma única chave de partição deve ser substancialmente menor que a capacidade do fragmento.

PutRecord exemplo

O código a seguir cria dez registros de dados, distribuídos entre duas chaves de partição, e os coloca em um fluxo denominado `myStreamName`.

```
for (int j = 0; j < 10; j++)
{
  PutRecordRequest putRecordRequest = new PutRecordRequest();
  putRecordRequest.setStreamName( myStreamName );
  putRecordRequest.setData(ByteBuffer.wrap( String.format( "testData-%d",
 j ).getBytes() ));
  putRecordRequest.setPartitionKey( String.format( "partitionKey-%d", j/5 ));
  putRecordRequest.setSequenceNumberForOrdering( sequenceNumberOfPreviousRecord );
  PutRecordResult putRecordResult = client.putRecord( putRecordRequest );
  sequenceNumberOfPreviousRecord = putRecordResult.getSequenceNumber();
}
```

O código de exemplo anterior usa `setSequenceNumberForOrdering` para garantir estritamente o aumento da ordenação dentro de cada chave de partição. Para usar esse parâmetro de forma eficaz, defina o `SequenceNumberForOrdering` do registro atual (registro n) como o número de sequência do registro anterior (registro n-1). Para obter o número sequencial de um registro que foi adicionado ao fluxo, chame `getSequenceNumber` para o resultado de `putRecord`.

O parâmetro `SequenceNumberForOrdering` garante estritamente o aumento de números de sequência para a mesma chave de partição. `SequenceNumberForOrdering` não fornece a ordenação de registros em várias chaves de partição.

Interaja com os dados usando o AWS Glue Schema Registry

Você pode integrar seus streams de dados do Kinesis com o AWS Glue Schema Registry. O registro de esquemas do AWS Glue permite detectar, controlar e evoluir esquemas centralmente, ao mesmo tempo que garante que os dados produzidos sejam validados continuamente por um esquema registrado. O esquema define a estrutura e o formato de um registro de dados. Um esquema é uma especificação versionada para publicação, consumo ou datastore confiáveis. O AWS Glue Schema Registry permite que você melhore a qualidade end-to-end dos dados e a governança de dados

em seus aplicativos de streaming. Para obter mais informações, consulte Registro de esquemas do AWS Glue. Uma das formas de configurar essa integração é por meio do `PutRecords` `PutRecord` Kinesis APIs Data Streams disponível AWS no Java SDK.

Para obter instruções detalhadas sobre como configurar a integração do Kinesis Data Streams com o registro do esquema usando o e o PutRecord Kinesis Data Streams, consulte a seção “Interagindo com dados PutRecords usando o APIs Kinesis Data Streams” em Caso de uso: Integração do Amazon Kinesis Data APIs Streams com o Glue Schema Registry. AWS

Gravar no Amazon Kinesis Data Streams usando o Kinesis Agent

O Kinesis Agent é uma aplicação de software Java independente que oferece uma maneira fácil de coletar e enviar dados ao Kinesis Data Streams. O agente monitora continuamente um conjunto de arquivos e envia novos dados ao fluxo. Ele manipula o rodízio de arquivos, os pontos de verificação e as novas tentativas após falhas. Os dados são entregues de maneira confiável, imediata e simples. Ele também emite CloudWatch métricas da Amazon para ajudar você a monitorar e solucionar melhor o processo de streaming.

Por padrão, os registros são analisados em cada arquivo com base no caractere de nova linha (`'\n'`). No entanto, o agente também pode ser configurado para analisar registros de várias linhas (consulte Especificar as definições da configuração do agente).

É possível instalar o agente em ambientes de servidor baseados no Linux, como servidores web, servidores de log e servidores de banco de dados. Após instalar o agente, configure-o especificando os arquivos a serem monitorados e o fluxo dos dados. Depois que o agente é configurado, ele coleta dados dos arquivos de forma durável e os envia confiavelmente ao fluxo.

Tópicos

- Concluir os pré-requisitos do Kinesis Agent
- Fazer download e instalar o agente
- Configuração e inicialização do agente
- Especificar as definições da configuração do agente
- Monitorar vários diretórios de arquivos e gravação em vários fluxos
- Uso do agente para pré-processar dados
- Usar comandos da CLI do agente
- Perguntas frequentes

Concluir os pré-requisitos do Kinesis Agent

- O sistema operacional deve ser a AMI do Amazon Linux versão 2015.09 ou posterior ou o Red Hat Enterprise Linux versão 7 ou posterior.
- Se estiver usando o Amazon EC2 para executar o agente, inicie a instância do EC2.
- Gerencie suas AWS credenciais usando um dos seguintes métodos:
  - Especifique um perfil do IAM ao executar a instância do EC2.
  - Especifique AWS as credenciais ao configurar o agente (consulte awsAccessKeyID e awsSecretAccesschave).
  - Edite `/etc/sysconfig/aws-kinesis-agent` para especificar sua região e suas chaves de AWS acesso.
  - Se sua instância do EC2 estiver em uma AWS conta diferente, crie uma função do IAM para fornecer acesso ao serviço Kinesis Data Streams e especifique essa função ao configurar o agente (consulte AssumeroLearn e Id). assumeRoleExternal Use um dos métodos anteriores para especificar AWS as credenciais de um usuário na outra conta que tenha permissão para assumir essa função.
- A função ou AWS as credenciais do IAM que você especificar devem ter permissão para realizar a operação do Kinesis Data PutRecordsStreams para que o agente envie dados para seu stream. Se você ativar o CloudWatch monitoramento para o agente, a permissão para realizar a CloudWatch PutMetricDataoperação também será necessária. Para obter mais informações, consulte Controlar o acesso aos recursos do Amazon Kinesis Data Streams usando o IAMMonitore a integridade do Kinesis Data Streams Agent com a Amazon CloudWatch, e Controle de CloudWatch acesso.

Fazer download e instalar o agente

Primeiro, conecte-se à instância. Para obter mais informações, consulte Conectar-se à sua instância no Guia do usuário do Amazon EC2. Se tiver problemas para se conectar, consulte Solução de problemas para conectar-se à sua instância no Guia do usuário do Amazon EC2.

Como configurar o agente usando o Amazon Linux AMI

Use o comando a seguir para fazer download do agente e instalá-lo:

```
sudo yum install -y aws-kinesis-agent
```

Como configurar o agente usando o Red Hat Enterprise Linux

Use o comando a seguir para fazer download do agente e instalá-lo:

```
sudo yum install -y https://s3.amazonaws.com/streaming-data-agent/aws-kinesis-agent-latest.amzn2.noarch.rpm
```

Para configurar o agente usando GitHub

1. Baixe o agente em amazon-kinesis-agentawlabs/.
2. Instale o agente navegando até o diretório de download e executando o comando a seguir:

   ```
   sudo ./setup --install
   ```

Como configurar o agente em um contêiner do Docker

O Kinesis Agent também pode ser executado em um contêiner por meio da base de contêineres amazonlinux. Use o Dockerfile a seguir e depois execute o `docker build`.

```
FROM amazonlinux

RUN yum install -y aws-kinesis-agent which findutils
COPY agent.json /etc/aws-kinesis/agent.json

CMD ["start-aws-kinesis-agent"]
```

Configuração e inicialização do agente

Como configurar e iniciar o agente

1. Abra e edite o arquivo de configuração (como superusuário, se as permissões padrão de acesso a arquivos estiverem sendo usadas): `/etc/aws-kinesis/agent.json`

   Nesse arquivo de configuração, especifique os arquivos (`"filePattern"`) nos quais o agente coleta dados e o nome do fluxo (`"kinesisStream"`) ao qual o agente envia dados. Observe que o nome do arquivo é um padrão, e o agente reconhece os rodízios de arquivos. Só é possível fazer o rodízio de arquivos ou criar novos arquivos uma vez por segundo, no máximo. O agente usa o carimbo de data e hora de criação de arquivo para determinar quais arquivos serão rastreados e colocados no final do fluxo; a criação de novos arquivos ou o rodízio de

arquivos em uma frequência superior a uma vez por segundo não permite que o agente faça a distinção entre eles corretamente.

```
{
   "flows": [
        {
            "filePattern": "/tmp/app.log*",
            "kinesisStream": "yourkinesisstream"
        }
   ]
}
```

2. Inicie o agente manualmente:

```
sudo service aws-kinesis-agent start
```

3. (Opcional) Configure o agente para ser iniciado durante o startup do sistema:

```
sudo chkconfig aws-kinesis-agent on
```

Agora o agente está sendo executado como um serviço do sistema em segundo plano. Ele monitora continuamente os arquivos especificados e envia dados ao fluxo especificado. A atividade do agent é registrada em `/var/log/aws-kinesis-agent/aws-kinesis-agent.log`.

Especificar as definições da configuração do agente

O agente oferece suporte a duas configurações obrigatórias, `filePattern` e `kinesisStream`, além das configurações opcionais de recursos adicionais. É possível especificar configurações obrigatórias e opcionais em `/etc/aws-kinesis/agent.json`.

Sempre que o arquivo de configuração for alterado, o agente deverá ser interrompido e iniciado, usando os seguintes comandos:

```
sudo service aws-kinesis-agent stop
sudo service aws-kinesis-agent start
```

Se desejar, é possível usar o comando a seguir:

```
sudo service aws-kinesis-agent restart
```

Estas são as configurações gerais.

Definição da configuração	Description
`assumeRoleARN`	O ARN da função a ser assumida pelo usuário. Para obter mais informações, consulte Delegar acesso entre AWS contas usando funções do IAM no Guia do usuário do IAM.
`assumeRoleExternalId`	Um identificador opcional que determina quem pode assumir o perfil. Para obter mais informações, consulte Como usar um ID externo no Guia do usuário do IAM.
`awsAccessKeyId`	AWS ID da chave de acesso que substitui as credenciais padrão. Essa configuração tem precedência sobre todos os outros provedores de credenciais.
`awsSecretAccessKey`	AWS chave secreta que substitui as credenciais padrão. Essa configuração tem precedência sobre todos os outros provedores de credenciais.
`cloudwatch.emitMetrics`	Permite que o agente emita métricas para, CloudWatch se definidas (verdadeiras). Padrão: verdadeiro
`cloudwatch.endpoint`	O endpoint regional para CloudWatch. Padrão: `monitoring.us-east-1.amazonaws.com`
`kinesis.endpoint`	O endpoint regional do Kinesis Data Streams. Padrão: `kinesis.us-east-1.amazonaws.com`

Estas são as configurações de fluxo.

Definição da configuração	Description
`dataProcessingOptions`	A lista das opções de processamento aplicadas a cada registro analisado antes que ele seja enviado ao fluxo. As opções de processamento são executadas na ordem especificada. Para obter mais informações, consulte Uso do agente para pré-processar dados.
`kinesisStream`	[Obrigatório] O nome do fluxo.
`filePattern`	[Obrigatório] O diretório e o padrão de arquivo que devem ser combinados para serem coletados pelo agente. Para todos os arquivos correspondentes a esse padrão, deve ser concedida uma permissão de leitura a `aws-kinesis-agent-user` . Para o diretório que contém os arquivos, devem ser concedidas permissões de leitura e execução a `aws-kinesis-agent-user` .
`initialPosition`	A posição em que o arquivo começou a ser analisado. Os valores válidos são `START_OF_FILE` e `END_OF_FILE` . Padrão: `END_OF_FILE`
`maxBufferAgeMillis`	O tempo máximo, em milissegundos, durante o qual o agente armazena os dados em buffer antes de enviá-los ao fluxo. Intervalo de valores: 1.000 a 900.000 (1 segundo a 15 minutos) Padrão: 60.000 (1 minuto)
`maxBufferSizeBytes`	O tamanho máximo, em bytes, durante o qual o agente armazena os dados em buffer antes de enviá-los ao fluxo. Intervalo de valores: 1 a 4.194.304 (4 MB) Padrão: 4.194.304 (4 MB)
`maxBufferSizeRecords`	O número máximo de registros para os quais o agente armazena os dados em buffer antes de enviá-los ao fluxo. Intervalo de valores: 1 a 500

Definição da configuração	Description
	Padrão: 500
`minTimeBetweenFilePollsMillis`	O intervalo de tempo, em milissegundos, em que o agente consulta e analisa os arquivos monitorados em busca de novos dados. Intervalo de valores: 1 ou mais Padrão: 100
`multiLineStartPattern`	O padrão de identificação do início de um registro. Um registro é composto por uma linha que corresponde ao padrão e pelas linhas subsequentes que não correspondem ao padrão. Os valores válidos são expressões regulares. Por padrão, cada nova linha nos arquivos de log é analisada como um único registro.
`partitionKeyOption`	O método para gerar a chave de partição. Os valores válidos são `RANDOM` (inteiro gerado aleatoriamente) e `DETERMINISTIC` (um valor de hash calculado a partir dos dados). Padrão: `RANDOM`
`skipHeaderLines`	O número de linhas em que o agente ignorará a análise no início dos arquivos monitorados. Intervalo de valores: 0 ou mais Padrão: 0 (zero)
`truncatedRecordTerminator`	A string que o agente usa para truncar um registro analisado que excede o limite de tamanho de registro do Kinesis Data Streams. (1,000 KB) Padrão: `'\n'` (nova linha)

Monitorar vários diretórios de arquivos e gravação em vários fluxos

Ao especificar vários fluxos de configurações, é possível configurar o agente para monitorar vários diretórios de arquivos e enviar dados a vários streams. No exemplo de configuração a seguir, o

agente monitora dois diretórios de arquivos e envia dados para um fluxo do Kinesis e para um fluxo de entrega do Firehose, respectivamente. Observe que, como é possível especificar endpoints diferentes para o Kinesis Data Streams e o Firehose, os fluxos dos dois serviços não precisam estar na mesma região.

```
{
    "cloudwatch.emitMetrics": true,
    "kinesis.endpoint": "https://your/kinesis/endpoint",
    "firehose.endpoint": "https://your/firehose/endpoint",
    "flows": [
        {
            "filePattern": "/tmp/app1.log*",
            "kinesisStream": "yourkinesisstream"
        },
        {
            "filePattern": "/tmp/app2.log*",
            "deliveryStream": "yourfirehosedeliverystream"
        }
    ]
}
```

Para obter informações mais detalhadas sobre o uso do agente com o Firehose, consulte Gravar no Amazon Kinesis Data Firehose com o Kinesis Agent.

Uso do agente para pré-processar dados

O agente pode pré-processar os registros analisados a partir dos arquivos monitorados antes de enviá-los ao fluxo. É possível habilitar esse recurso adicionando a configuração `dataProcessingOptions` ao fluxo de arquivos. Um ou mais opções de processamento podem ser adicionadas e serão executadas na ordem especificada.

O agente oferece suporte às seguintes opções de processamento. Como o agente é de código aberto, é possível desenvolver e estender ainda mais suas opções de processamento. O download do agente pode ser feito em Kinesis Agent.

Opções de processamento

`SINGLELINE`

> Converte um registro de várias linhas em um registro de única linha removendo caracteres de nova linha, e espaços à esquerda e à direita.

```
{
    "optionName": "SINGLELINE"
}
```

CSVTOJSON

Converte um registro com formato separado por delimitador em um registro com formato JSON.

```
{
    "optionName": "CSVTOJSON",
    "customFieldNames": [ "field1", "field2", ... ],
    "delimiter": "yourdelimiter"
}
```

customFieldNames

[Obrigatório] Os nomes de campo usados como chaves em cada par de valores de chave JSON. Por exemplo, ao especificar `["f1", "f2"]`, o registro "v1, v2" será convertido em `{"f1":"v1","f2":"v2"}`.

delimiter

A string usada como delimitador no registro. O padrão é uma vírgula (,).

LOGTOJSON

Converte um registro com formato de log em um registro com formato JSON. Os formatos de log com suporte são Apache Common Log, Apache Combined Log, Apache Error Log e RFC3164 Syslog.

```
{
    "optionName": "LOGTOJSON",
    "logFormat": "logformat",
    "matchPattern": "yourregexpattern",
    "customFieldNames": [ "field1", "field2", … ]
}
```

logFormat

[Obrigatório] O formato da entrada de log. Os valores possíveis são:

- `COMMONAPACHELOG`: o formato do Apache Common Log. Cada entrada de log tem o seguinte padrão: `"%{host} %{ident} %{authuser} [%{datetime}] \"%{request}\" %{response} %{bytes}"`.
- `COMBINEDAPACHELOG`: o formato do Apache Combined Log. Cada entrada de log tem o seguinte padrão: `"%{host} %{ident} %{authuser} [%{datetime}] \"%{request}\" %{response} %{bytes} %{referrer} %{agent}"`.
- `APACHEERRORLOG`: o formato do Apache Error Log. Cada entrada de log tem o seguinte padrão: `"[%{timestamp}] [%{module}:%{severity}] [pid %{processid}:tid %{threadid}] [client: %{client}] %{message}"`.
- `SYSLOG`— O formato RFC3164 Syslog. Cada entrada de log tem o seguinte padrão: `"%{timestamp} %{hostname} %{program}[%{processid}]: %{message}"`.

`matchPattern`

O padrão da expressão regular usada para extrair valores de entradas de log. Essa configuração é usada se a entrada de log não estiver em um dos formatos de log predefinidos. Se essa configuração for usada, também é necessário especificar `customFieldNames`.

`customFieldNames`

Os nomes de campo personalizados usados como chaves em cada par de valores de chave JSON. É possível usar essa configuração para definir nomes de campo para valores extraídos de `matchPattern` ou substituir os nomes de campo padrão de formatos de log predefinidos.

Example: Configuração LOGTOJSON

Este é um exemplo de uma configuração `LOGTOJSON` para uma entrada Apache Common Log convertida em formato JSON:

```
{
    "optionName": "LOGTOJSON",
    "logFormat": "COMMONAPACHELOG"
}
```

Antes da conversão:

```
64.242.88.10 - - [07/Mar/2004:16:10:02 -0800] "GET /mailman/listinfo/hsdivision
 HTTP/1.1" 200 6291
```

Depois da conversão:

```
{"host":"64.242.88.10","ident":null,"authuser":null,"datetime":"07/
Mar/2004:16:10:02 -0800","request":"GET /mailman/listinfo/hsdivision
 HTTP/1.1","response":"200","bytes":"6291"}
```

Example: Configuração LOGTOJSON com campos personalizados

Este é outro exemplo de configuração LOGTOJSON:

```
{
    "optionName": "LOGTOJSON",
    "logFormat": "COMMONAPACHELOG",
    "customFieldNames": ["f1", "f2", "f3", "f4", "f5", "f6", "f7"]
}
```

Com essa configuração, a mesma entrada Apache Common Log do exemplo anterior é convertida em formato JSON, da seguinte forma:

```
{"f1":"64.242.88.10","f2":null,"f3":null,"f4":"07/Mar/2004:16:10:02 -0800","f5":"GET /
mailman/listinfo/hsdivision HTTP/1.1","f6":"200","f7":"6291"}
```

Example: Conversão da entrada Apache Common Log

A configuração de fluxo a seguir converte uma entrada Apache Common Log em um registro de linha única no formato JSON:

```
{
    "flows": [
        {
            "filePattern": "/tmp/app.log*",
            "kinesisStream": "my-stream",
            "dataProcessingOptions": [
                {
                    "optionName": "LOGTOJSON",
                    "logFormat": "COMMONAPACHELOG"
                }
            ]
        }
    ]
}
```

Example: Conversão de registros de várias linhas

A configuração de fluxo a seguir analisa registros de várias linha cuja primeira linha começa com "[SEQUENCE=". Cada registro é convertido primeiro em um registro de única linha. Em seguida, os valores são extraídos do registro com base em um delimitador por tabulações. Os valores extraídos são mapeados para os valores `customFieldNames` especificados, a fim de formar um registro de linha única no formato JSON.

```
{
    "flows": [
        {
            "filePattern": "/tmp/app.log*",
            "kinesisStream": "my-stream",
            "multiLineStartPattern": "\\[SEQUENCE=",
            "dataProcessingOptions": [
                {
                    "optionName": "SINGLELINE"
                },
                {
                    "optionName": "CSVTOJSON",
                    "customFieldNames": [ "field1", "field2", "field3" ],
                    "delimiter": "\\t"
                }
            ]
        }
    ]
}
```

Example: Configuração LOGTOJSON com padrão de correspondência

este é um exemplo de configuração `LOGTOJSON` referente a uma entrada Apache Common Log convertida em formato JSON, com o último campo (bytes) omitido:

```
{
    "optionName": "LOGTOJSON",
    "logFormat": "COMMONAPACHELOG",
    "matchPattern": "^([\\d.]+) (\\S+) (\\S+) \\[([\\w:/]+\\s[+\\-]\\d{4})\\] \"(.
+?)\" (\\d{3})",
    "customFieldNames": ["host", "ident", "authuser", "datetime", "request",
 "response"]
}
```

Antes da conversão:

```
123.45.67.89 - - [27/Oct/2000:09:27:09 -0400] "GET /java/javaResources.html HTTP/1.0"
 200
```

Depois da conversão:

```
{"host":"123.45.67.89","ident":null,"authuser":null,"datetime":"27/Oct/2000:09:27:09
 -0400","request":"GET /java/javaResources.html HTTP/1.0","response":"200"}
```

Usar comandos da CLI do agente

Inicie automaticamente o agente durante o startup do sistema:

```
sudo chkconfig aws-kinesis-agent on
```

Verifique o status do agente:

```
sudo service aws-kinesis-agent status
```

Interrompa o agente:

```
sudo service aws-kinesis-agent stop
```

Leia o arquivo de log do agente a partir deste local:

```
/var/log/aws-kinesis-agent/aws-kinesis-agent.log
```

Desinstale o agente:

```
sudo yum remove aws-kinesis-agent
```

Perguntas frequentes

Existe um Kinesis Agent para Windows?

O Kinesis Agent para Windows é um software diferente das plataformas do Kinesis Agent para Linux.

Por que a desaceleração do Kinesis Agent está aumentando? and/or **`RecordSendErrors`**

Isso geralmente ocorre devido ao controle de utilização do Kinesis. Verifique a métrica `WriteProvisionedThroughputExceeded` do Kinesis Data Streams ou a métrica `ThrottledRecords` dos fluxos de entrega do Firehose. Qualquer aumento de 0 nessas métricas indica que os limites do fluxo precisam ser aumentados. Para obter mais informações, consulte Kinesis Data Stream limits e Amazon Firehose Delivery Streams.

Depois de descartar o controle de utilização como causa, verifique se o Kinesis Agent está configurado para seguir um número grande de arquivos pequenos. Há um atraso quando o Kinesis Agent exibe os dados do final de um arquivo novo, portanto, o Kinesis Agent deveria estar exibindo os dados do final de um pequeno número de arquivos maiores. Tente consolidar os arquivos de log em arquivos maiores.

Por que estou recebendo exceções **`java.lang.OutOfMemoryError`**?

O Kinesis Agent não tem memória suficiente para lidar com a workload atual. Tente aumentar `JAVA_START_HEAP` e `JAVA_MAX_HEAP` no `/usr/bin/start-aws-kinesis-agent` e reiniciar o agente.

Por que estou recebendo exceções **`IllegalStateException : connection pool shut down`**?

O Kinesis Agent não tem conexões suficientes para lidar com a workload atual. Tente aumentar `maxConnections` e `maxSendingThreads` nas configurações gerais do agente em `/etc/aws-kinesis/agent.json`. O valor padrão para esses campos é 12 vezes o número de processadores de runtime disponíveis. Consulte AgentConfiguration.java para saber mais sobre as configurações avançadas do agente.

Como posso depurar outro problema com o Kinesis Agent?

Os logs do nível `DEBUG` podem ser habilitados em `/etc/aws-kinesis/log4j.xml`.

Como devo configurar o Kinesis Agent?

Quanto menor o `maxBufferSizeBytes`, mais frequentemente o Kinesis Agent enviará dados. Isso pode ser bom, pois diminui o tempo de entrega dos registros, mas também aumenta as solicitações por segundo feitas ao Kinesis.

Por que o Kinesis Agent está enviando registros duplicados?

Isso ocorre devido a uma configuração incorreta da exibição dos dados do final dos arquivos. Certifique-se de que cada `fileFlow's filePattern` corresponda a apenas um arquivo. Isso também pode ocorrer se o modo `logrotate` que está sendo usado estiver no modo `copytruncate`. Tente mudar o modo para o modo padrão ou criar para evitar duplicações. Para obter mais informações sobre como lidar com registros duplicados, consulte Handling Duplicate Records.

Grave no Kinesis Data Streams usando outros serviços AWS

Os AWS serviços a seguir podem se integrar diretamente ao Amazon Kinesis Data Streams para gravar dados nos streams de dados do Kinesis. Leia as informações sobre cada serviço que seja de seu interesse e consulte as referências fornecidas.

Tópicos

- Grave no Kinesis Data Streams usando AWS Amplify
- Gravar no Kinesis Data Streams usando o Amazon Aurora
- Escreva para o Kinesis Data Streams usando a Amazon CloudFront
- Grave no Kinesis Data Streams CloudWatch usando o Amazon Logs
- Gravar no Kinesis Data Streams usando o Amazon Connect
- Grave no Kinesis Data Streams usando AWS Database Migration Service
- Gravar no Kinesis Data Streams usando o Amazon DynamoDB
- Escreva para o Kinesis Data Streams usando a Amazon EventBridge
- Grave no Kinesis Data Streams usando AWS IoT Core
- Gravar no Kinesis Data Streams usando o Amazon Relational Database Service
- Gravar no Kinesis Data Streams usando o Amazon Pinpoint
- Gravar no Kinesis Data Streams usando o Amazon Quantum Ledger Database (Amazon QLDB)

Grave no Kinesis Data Streams usando AWS Amplify

Você pode usar o Amazon Kinesis Data Streams para transmitir dados de aplicações móveis criadas com o AWS Amplify para processamento em tempo real. Isso permite a criação de painéis em tempo real, capture exceções, gere alertas, estimule recomendações e tome outras decisões comerciais

ou operacionais oportunas. Você também pode enviar dados para outros serviços, como o Amazon Simple Storage Service, o Amazon DynamoDB e o Amazon Redshift.

Para obter mais informações, consulte Using Amazon Kinesis no AWS Amplify Developer Center.

Gravar no Kinesis Data Streams usando o Amazon Aurora

É possível usar o Amazon Kinesis Data Streams para monitorar atividades em clusters de banco de dados do Amazon Aurora. Usando o Database Activity Streams, seu cluster de banco de dados do Aurora envia atividades para um fluxo de dados do Amazon Kinesis em tempo real. Em seguida, pode-se criar aplicações de gerenciamento de conformidade para consumir essas atividades, auditá-las e gerar alertas. Também é possível usar o Amazon Firehose para armazenar dados.

Para obter mais informações, consulte Database Activity Streams no Guia do desenvolvedor do Amazon Aurora.

Escreva para o Kinesis Data Streams usando a Amazon CloudFront

Você pode usar o Amazon Kinesis Data CloudFront Streams com registros em tempo real e obter informações sobre solicitações feitas para uma distribuição em tempo real. Em seguida, você pode criar seu próprio consumidor de stream de dados do Kinesis ou usar o Amazon Data Firehose para enviar os dados de log para o Amazon S3, Amazon Redshift, Amazon Service ou um serviço de processamento de log de OpenSearch terceiros.

Para obter mais informações, consulte Registros em tempo real no Amazon CloudFront Developer Guide.

Grave no Kinesis Data Streams CloudWatch usando o Amazon Logs

Você pode usar CloudWatch assinaturas para ter acesso a um feed em tempo real de eventos de log do Amazon CloudWatch Logs e entregá-lo a um stream de dados do Kinesis para processamento, análise e carregamento em outros sistemas.

Para obter mais informações, consulte Processamento em tempo real de dados de log com assinaturas no Guia do usuário do Amazon CloudWatch Logs.

Gravar no Kinesis Data Streams usando o Amazon Connect

É possível usar o Kinesis Data Streams para exportar registros de contatos e eventos de agentes em tempo real da sua instância do Amazon Connect. Você também pode habilitar o fluxo de dados

do Amazon Connect Customer Profiles para receber automaticamente atualizações em um fluxo de dados do Kinesis sobre a criação de novos perfis ou alterações nos perfis existentes.

Em seguida, você pode criar aplicações de consumo para processar e analisar os dados em tempo real. Por exemplo, usando registros de contato e dados do perfil do cliente, você pode manter os dados do sistema de origem, como CRMs ferramentas de automação de marketing, up-to-date com as informações mais recentes. Usando os dados de eventos do agente, é possível criar painéis que exibem informações e eventos, além de acionar notificações personalizadas de atividades específicas do agente.

Para obter mais informações, consulte Fluxo de dados para sua instância, Configurar uma exportação em tempo real e Fluxos de eventos de agente no Guia do administrador do Amazon Connect.

Grave no Kinesis Data Streams usando AWS Database Migration Service

Você pode usar AWS Database Migration Service para migrar dados para um stream de dados do Kinesis. Em seguida, é possível criar aplicações de consumo para processar os registros de dados em tempo real. Você também pode enviar facilmente dados downstream para outros serviços, como o Amazon Simple Storage Service, o Amazon DynamoDB e o Amazon Redshift.

Para obter mais informações, consulte Using Kinesis Data Streams no Guia do usuário do AWS Database Migration Service .

Gravar no Kinesis Data Streams usando o Amazon DynamoDB

É possível usar o Amazon Kinesis Data Streams para capturar alterações no Amazon DynamoDB. O Kinesis Data Streams capta alterações no nível de item em qualquer DynamoDB e as replica em um fluxo de dados do Kinesis. Suas aplicações de consumo podem acessar esse fluxo para visualizar as alterações nos itens em tempo real e entregá-las downstream ou realizar ações com base no seu conteúdo.

Para obter mais informações, consulte Como o Kinesis Data Streams funciona com o DynamoDB no Guia do desenvolvedor do Amazon DynamoDB.

Escreva para o Kinesis Data Streams usando a Amazon EventBridge

Usando o Kinesis Data Streams, você AWS pode enviar EventBridge eventos de chamada de API para um stream, criar aplicativos de consumo e processar grandes quantidades de dados. Você

também pode usar o Kinesis Data Streams como EventBridge destino no Pipes e entregar registros de um stream de uma das fontes disponíveis após filtragem e enriquecimento opcionais.

Para obter mais informações, consulte Enviar eventos para um stream do Amazon Kinesis e EventBridge Pipes no Guia EventBridge do usuário da Amazon.

Grave no Kinesis Data Streams usando AWS IoT Core

Você pode gravar dados em tempo real a partir de mensagens MQTT no AWS IoT Core usando ações de regras de AWS IoT. Em seguida, pode-se criar aplicações para processar os dados, analisar seu conteúdo, gerar alertas e entregar os dados a aplicações de analytics ou outros serviços da AWS .

Para obter mais informações, consulte Kinesis Data Streams no Guia do desenvolvedor do AWS IoT Core.

Gravar no Kinesis Data Streams usando o Amazon Relational Database Service

É possível usar o Amazon Kinesis Data Streams para monitorar atividades em instâncias do Amazon RDS. Usando o Database Activity Streams, o Amazon RDS envia atividades para um fluxo de dados do Kinesis em tempo real. Em seguida, pode-se criar aplicações de gerenciamento de conformidade para consumir essas atividades, auditá-las e gerar alertas. Também é possível usar o Amazon Data Firehose para armazenar dados.

Para obter mais informações, consulte Database Activity Streams no Guia do desenvolvedor do Amazon RDS.

Gravar no Kinesis Data Streams usando o Amazon Pinpoint

É possível configurar o Amazon Pinpoint para enviar dados de eventos ao Amazon Kinesis Data Streams. O Amazon Pinpoint pode enviar dados de eventos para campanhas, viagens e mensagens de e-mail e SMS transacionais. Em seguida, é possível ingerir os dados em aplicações de analytics ou criar as próprias aplicações de consumo para realizar ações com base no conteúdo dos eventos.

Para obter mais informações, consulte Streaming Events no Guia do desenvolvedor do Amazon Pinpoint.

Gravar no Kinesis Data Streams usando o Amazon Quantum Ledger Database (Amazon QLDB)

Você pode criar um fluxo no Amazon QLDB para capturar todas as revisões de documentos confirmadas no diário e entregar esses dados ao Amazon Kinesis Data Streams em tempo real. Assim, o QLDB mantém um fluxo contínuo de dados do diário do livro-razão para um recurso de fluxo de dados do Kinesis. Em seguida, é possível usar a plataforma de fluxo Kinesis ou a Kinesis Client Library para consumir o fluxo, processar os registros de dados e analisar o conteúdo dos dados. Um fluxo do QLDB grava dados no Kinesis Data Streams em três tipos de registros `control`, `block summary` e `revision details`.

Para obter mais informações, consulte Streams no Guia do desenvolvedor do Amazon QLDB.

Gravar no Kinesis Data Streams usando integrações de terceiros

Você pode gravar dados no Kinesis Data Streams usando uma das opções de terceiros a seguir que se integram ao Kinesis Data Streams: Selecione uma opção para saber mais sobre ela e encontre recursos e links para a documentação relevante.

Tópicos

- Apache Flink
- Fluentd
- Debezium
- Oráculo GoldenGate
- Kafka Connect
- Adobe Experience
- Striim

Apache Flink

O Apache Flink é um framework de código aberto distribuído para computações com estado em fluxos de dados delimitados e não delimitados. Para obter mais informações sobre como gravar no Kinesis Data Streams do Apache Flink, consulte Amazon Kinesis Data Streams Connector.

Fluentd

O Fluentd é um coletor de dados de código aberto para uma camada de registro unificada. Para obter mais informações sobre como gravar no Kinesis Data Streams do Fluentd, consulte Stream processing with Kinesis.

Debezium

O Debezium é uma plataforma distribuída de código aberto para captura de mudanças nos dados. Para obter mais informações sobre como gravar no Kinesis Data Streams do Debezium, consulte Streaming MySQL Data Changes to Amazon Kinesis.

Oráculo GoldenGate

GoldenGate O Oracle é um produto de software que permite replicar, filtrar e transformar dados de um banco de dados para outro. Para obter mais informações sobre como gravar no Kinesis Data Streams GoldenGate da Oracle, consulte Replicação de dados para o Kinesis Data Stream usando Oracle. GoldenGate

Kafka Connect

O Kafka Connect é uma ferramenta para transmitir dados de forma escalável e confiável entre o Apache Kafka e outros sistemas. Para obter mais informações sobre como gravar no Kinesis Data Streams do Apache Kafka, consulte Kinesis kafka connector.

Adobe Experience

A Adobe Experience Platform permite que as organizações centralizem e padronizem dados dos clientes em qualquer sistema. Em seguida, a plataforma aplica ciência de dados e machine learning para melhorar significativamente o design e a entrega de experiências ricas e personalizadas. Para obter mais informações sobre como gravar dados no Kinesis Data Streams da Adobe Experience Platform, aprenda como criar uma conexão com o Amazon Kinesis.

Striim

O Striim é uma plataforma completa em memória para coletar, filtrar, transformar, enriquecer, agregar, analisar e fornecer dados em tempo real. end-to-end Para obter mais informações sobre como gravar dados no Kinesis Data Streams do Striim, consulte Kinesis Writer.

Solução de problemas de produtores do Amazon Kinesis Data Streams

Os tópicos a seguir oferecem soluções para problemas comuns com produtores do Amazon Kinesis Data Streams:

- Meu aplicativo produtor está gravando a uma taxa menor que a esperada
- Eu recebo um erro de permissão de chave mestra do KMS não autorizada
- Solucionar outros problemas comuns para produtores

Meu aplicativo produtor está gravando a uma taxa menor que a esperada

Os motivos mais comuns para o throughput de gravação ser mais lento do que o esperado são:

- Limites de serviço excedidos
- Quero otimizar meu produtor
- Uso indevido das operações flushSync()

Limites de serviço excedidos

Para descobrir se os limites de serviço estão sendo excedidos, verifique se o produtor está lançando exceções de throughput a partir do serviço e valide quais operações da API estão sendo aceleradas. Lembre-se de que há limites diferentes de acordo com a chamada, consulte Cotas e limites. Por exemplo, além dos limites de nível de fragmento para gravações e leituras que são mais comumente conhecidas, há limites de nível de fluxo a seguir:

- CreateStream
- DeleteStream
- ListStreams
- GetShardIterator
- MergeShards
- DescribeStream
- DescribeStreamSummary

As operações `CreateStream`, `DeleteStream`, `ListStreams`, `GetShardIterator` e `MergeShards` são limitadas a 5 chamadas por segundo. A operação `DescribeStream` é limitada a 10 chamadas por segundo. A operação `DescribeStreamSummary` é limitada a 20 chamadas por segundo.

Se essas chamadas não forem o problema, selecione uma chave de partição que permita distribuir operações put uniformemente em todos os fragmentos e não tenha uma determinada chave de partição que esteja colidindo com os limites de serviço quando as restantes não estão. Isso requer a medição do throughput de pico e que seja levado em conta o número de fragmentos no seu fluxo. Para obter mais informações sobre o gerenciamento de streams, consulte Criar e gerenciar fluxos de dados do Kinesis.

> **Tip**
>
> Lembre-se de arredondar para o kilobyte mais próximo para cálculos de limitação de taxa de transferência ao usar a operação de um único registro PutRecord, enquanto a operação de vários registros PutRecords é arredondada na soma cumulativa dos registros em cada chamada. Por exemplo, uma solicitação `PutRecords` com 600 registros com tamanho de 1,1 KB não serão aceleradas.

Quero otimizar meu produtor

Antes de começar a otimizar o produtor, conclua as seguintes tarefas importantes. Primeiro, identifique seu throughput de pico desejado em termos de tamanho do registro e registros por segundo. Em seguida, descarte a capacidade de fluxo conforme o fator de limitação (Limites de serviço excedidos). Se a capacidade de fluxo foi excluída, use as seguintes dicas de solução de problemas e diretrizes de otimização para os dois tipos comuns de aplicações de produção.

Produtor grande

Um grande produtor normalmente é executado em um servidor on-premises ou em uma instância do Amazon EC2. Os clientes que precisam de um throughput mais alto de um grande produtor normalmente se preocupam com a latência por registro. As estratégias para lidar com a latência incluem o seguinte: Se o cliente puder armazenar registros em microlote/buffer, use a Amazon Kinesis Producer Library (que tem lógica de agregação avançada), a operação de vários registros ou agregar registros em um arquivo maior antes de usar a operação PutRecordsde registro único. PutRecord Se não for possível criar microlotes ou armazenar registros em buffer, use vários threads

para gravar no serviço Kinesis Data Streams ao mesmo tempo. Os AWS SDK para Java e outros SDKs incluem clientes assíncronos que podem fazer isso com muito pouco código.

Produtor pequeno

Um pequeno produtor geralmente é um aplicativo móvel, dispositivo IoT ou cliente web. Se for um aplicativo móvel, recomendamos usar a `PutRecords` operação ou o Kinesis Recorder no celular. AWS SDKs Para obter mais informações, consulte AWS Mobile SDK para Android Guia de introdução e AWS Mobile SDK for iOS Guia de introdução. Aplicativos móveis devem lidar com conexões intermitentes inerentemente e precisam de algum tipo de alocação em lote, como `PutRecords`. Se não for possível alocar em lote por algum motivo, consulte as informações sobre Grande produtor acima. Se o seu produtor é um navegador, a quantidade de dados que está sendo gerada geralmente é muito pequena. No entanto, as operações put estão sendo colocadas no caminho crítico do aplicativo, o que não é recomendável.

Uso indevido das operações **`flushSync()`**

O uso incorreto do `flushSync()` pode afetar significativamente o desempenho de gravação. A operação do `flushSync()` foi projetada para cenários de desligamento, a fim de garantir que todos os registros armazenados em buffer sejam enviados antes que a aplicação KPL seja encerrada. Se você implementou essa operação após cada operação de gravação, ela pode adicionar uma latência extra substancial de cerca de 500 ms por gravação. Assegure que `flushSync()` tenha sido implementado somente para o desligamento da aplicação para evitar atrasos extras desnecessários no desempenho de gravação.

Eu recebo um erro de permissão de chave mestra do KMS não autorizada

Esse erro ocorre quando um aplicativo produtor grava em um fluxo criptografado sem permissões na chave mestra do KMS. Para atribuir permissões a uma aplicação para que acesse uma chave do KMS, consulte Using Key Policies in AWS KMS e Using IAM Policies with AWS KMS.

Solucionar outros problemas comuns para produtores

- Por que meu fluxo de dados Kinesis retorna um Erro de Servidor Interno 500?
- How do I troubleshoot timeout errors when writing from Flink to Kinesis Data Streams?
- How do I troubleshoot throttling errors in Kinesis Data Streams?
- Por que ocorre controle de utilização em meu fluxo de dados Kinesis?
- Como posso colocar registros de dados em um fluxo de dados usando a KPL?

Otimizar os produtores do Kinesis Data Streams

Você pode otimizar ainda mais seus produtores do Amazon Kinesis Data Streams dependendo do comportamento específico observado. Leia os seguintes tópicos para identificar soluções.

Tópicos

- Personalize o comportamento de limites de taxas e novas tentativas da KPL
- Aplique as práticas recomendadas à agregação da KPL

Personalize o comportamento de limites de taxas e novas tentativas da KPL

Ao adicionar registros de usuário do Amazon Kinesis Producer Library (KPL) usando a operação `addUserRecord()` da KPL, o registro recebe um carimbo de data e hora e é adicionado a um buffer com um prazo definido pelo parâmetro de configuração `RecordMaxBufferedTime`. Essa stamp/deadline combinação de tempo define a prioridade do buffer. Os registros são descarregados do buffer com base nos seguintes critérios:

- Prioridade do buffer
- Configuração de agregação
- Configuração de coleta

Os parâmetros de configuração de coleta e agregação que afetam o comportamento do buffer são os seguintes:

- `AggregationMaxCount`
- `AggregationMaxSize`
- `CollectionMaxCount`
- `CollectionMaxSize`

Em seguida, os registros descarregados são enviados para o fluxo de dados do Kinesis como registros do Amazon Kinesis Data Streams usando uma chamada para a operação `PutRecords` de API do Kinesis Data Streams. A operação `PutRecords` envia solicitações ao fluxo que ocasionalmente apresenta falhas parciais ou totais. Os registros com falha são automaticamente adicionados de volta ao buffer da KPL. O novo prazo é definido com base no mínimo destes dois valores:

- Metade da configuração de `RecordMaxBufferedTime` atual
- O time-to-live valor do registro

Essa estratégia permite que registros repetidos de usuários do KPL sejam incluídos nas chamadas subsequentes da API do Kinesis Data Streams, para melhorar a taxa de transferência e reduzir a complexidade, ao mesmo tempo em que reforça o valor do registro do Kinesis Data Streams. time-to-live Não há um algoritmo de espera, tornando essa uma estratégia de tentativa relativamente agressiva. O spam devido ao excesso de tentativas é evitado pela limitação de taxas, discutida na próxima seção.

Limitação de intervalo

A KPL inclui um recurso de limitação de taxas, que limita o throughput por fragmento enviado de um único produtor. A limitação de taxas é implementada usando um algoritmo de bucket de token com buckets separados para bytes e registros do Kinesis Data Streams. Cada gravação bem-sucedida em um fluxo de dados do Kinesis adiciona um token (ou vários tokens) a cada bucket, até um determinado limite. Esse limite é configurável, mas por padrão é definido como 50% maior que o limite de fragmento real, para permitir a saturação de fragmentos a partir de um único produtor.

É possível reduzir esse limite para diminuir o spam devido ao excesso de tentativas. No entanto, a melhor prática é que cada produtor tente novamente para um throughput máximo de maneira agressiva e lide com qualquer limitação resultante determinada como excessiva por meio da expansão da capacidade do fluxo e da implementação de uma estratégia de chave de partição apropriada.

Aplique as práticas recomendadas à agregação da KPL

Embora o esquema do número de sequência dos registros resultantes do Amazon Kinesis Data Streams permaneça o mesmo, a agregação faz com que a indexação dos registros de usuários do Amazon Kinesis Producer Library (KPL) contidos em um registro agregado do Kinesis Data Streams comece de 0 (zero). No entanto, desde que não haja dependência dos números de sequência para identificar exclusivamente os registros de usuários da KPL, o código poderá ignorar isso, já que a agregação (de registros de usuários da KPL em um registro do Kinesis Data Streams) e a desagregação subsequente (de um registro do Kinesis Data Streams em registros de usuários da KPL) cuidam disso automaticamente para você. Isso se aplica se seu consumidor estiver usando o KCL ou o AWS SDK. Para usar essa funcionalidade de agregação, você precisará inserir a parte Java do KPL em sua compilação se seu consumidor for escrito usando a API fornecida no AWS SDK.

Se quiser usar números de sequência como identificadores exclusivos dos registros de usuários da KPL, recomendamos que use as operações `public int hashCode()` e `public boolean equals(Object obj)`, que respeitam contratos, fornecidas em `Record` e `UserRecord` para habilitar a comparação desses registros. Além disso, para examinar o número subsequente do registro de usuários da KPL, é possível convertê-lo em uma instância de `UserRecord` e recuperar o número de subsequência.

Para obter mais informações, consulte Implementar a desagregação de consumidores.

Leitura de dados do Amazon Kinesis Data Streams

Um consumidor é uma aplicação que processa todos os dados de um fluxo de dados do Kinesis. Quando um consumidor usa o fan-out aprimorado, ele obtém sua própria cota de taxa MB/sec de transferência de leitura, permitindo que vários consumidores leiam dados do mesmo fluxo em paralelo, sem competir pela taxa de transferência de leitura com outros consumidores. Para usar o recurso de distribuição avançada de fragmentos, consulte Desenvolver consumidores de distribuição avançada com throughput dedicado.

Você pode criar consumidores para o Kinesis Data Streams usando a Kinesis Client Library (KCL) ou AWS SDK para Java. Você também pode desenvolver consumidores usando outros AWS serviços AWS Lambda, como o Amazon Managed Service for Apache Flink e o Amazon Data Firehose. O Kinesis Data Streams oferece suporte a integrações AWS com outros serviços, como Amazon EMR, Amazon e Amazon Redshift. Ele também oferece suporte a integrações de terceiros EventBridge AWS Glue, incluindo Apache Flink, Adobe Experience Platform, Apache Druid, Apache Spark, Databricks, Confluent Platform, Kinesumer e Talend.

Tópicos

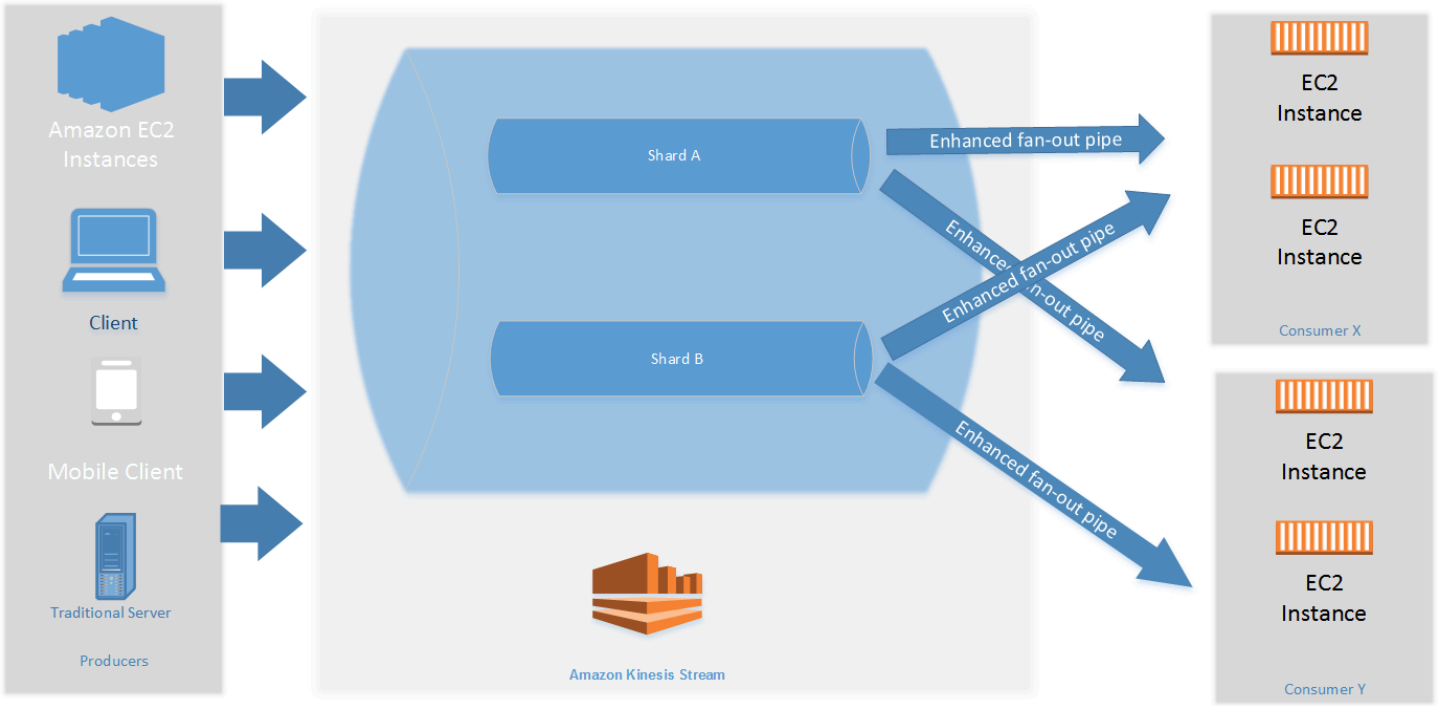
- Desenvolver consumidores de distribuição avançada com throughput dedicado
- Usar o visualizador de dados no console do Kinesis
- Consultar seus fluxos de dados no console do Kinesis
- Usar a Kinesis Client Library
- Desenvolva consumidores com o AWS SDK para Java
- Desenvolva consumidores usando AWS Lambda
- Desenvolver consumidores usando o Amazon Managed Service for Apache Flink
- Desenvolver consumidores usando o Amazon Data Firehose
- Leia dados do Kinesis Data Streams AWS usando outros serviços
- Leitura a partir do Kinesis Data Streams usando integrações de terceiros
- Solução de problemas de consumidores do Kinesis Data Streams
- Otimizar consumidores do Amazon Kinesis Data Streams

Desenvolver consumidores de distribuição avançada com throughput dedicado

No Amazon Kinesis Data Streams, é possível criar consumidores que usem um recurso chamado distribuição avançada. Esse atributo permite que consumidores recebam registros de um fluxo com throughput de até 2 MB de dados por segundo por fragmento. Esse throughput é dedicada, o que significa que consumidores que usam a divisão avançada não precisam lidar com outros consumidores que estejam recebendo dados do fluxo. O Kinesis Data Streams envia registros de dados do fluxo para consumidores que usam a distribuição avançada. Portanto, esses consumidores não precisam sondar dados.

> **Important**
>
> Com o modo On-demand Advantage, você pode registrar até 50 consumidores por stream para usar o fan-out aprimorado. Com streams padrão e provisionados sob demanda, você pode registrar até 20 consumidores por stream para usar o fan-out aprimorado.

O diagrama a seguir mostra a arquitetura de distribuição avançada. Ao usar a versão 2.0 ou posterior da Amazon Kinesis Client Library (KCL) para criar um consumidor, a KCL configurará o consumidor para usar a distribuição avançada e receber dados de todos os fragmentos do fluxo. Se usar a API para criar um consumidor que use distribuição avançada, é possível se inscrever em fragmentos individuais.



O diagrama mostra o seguinte:

- Um fluxo com dois fragmentos.
- Dois consumidores que estão usando a distribuição avançada para receber dados do fluxo: consumidores X e Y. Os dois consumidores estão inscritos em todos os fragmentos e em todos os registros do fluxo. Ao usar a versão 2.0 ou posterior da KCL para criar um consumidor, a KCL inscreverá automaticamente esse consumidor em todos os fragmentos do fluxo. Por outro lado, ao usar a API para criar um consumidor, é possível se inscrever em fragmentos individuais.
- Setas que representam as distribuições avançadas usadas pelos consumidores para receber dados do fluxo. Um tubo de ventilação aprimorado fornece até 2% MB/sec de dados por fragmento, independentemente de qualquer outro tubo ou do número total de consumidores.

Tópicos

- Diferenças entre o consumidor de throughput compartilhado e o consumidor de distribuição avançada
- Regiões com suporte para até 50 consumidores de fan-out aprimorados (somente sob demanda Advantage)
- Gerencie consumidores de distribuição aprimorados com o ou AWS CLI APIs

Diferenças entre o consumidor de throughput compartilhado e o consumidor de distribuição avançada

A tabela a seguir compara consumidores padrão de throughput compartilhada com consumidores de distribuição avançada. O atraso na propagação da mensagem é definido como o tempo gasto em milissegundos para que uma carga útil enviada usando o despacho de carga útil (como e) chegue ao aplicativo consumidor por meio da carga útil consumidora APIs (como `PutRecord` e`PutRecords`). APIs `GetRecords` `SubscribeToShard`

Essa tabela compara consumidores de throughput compartilhada com consumidores de distribuição avançada

Características	Consumidores de throughput compartilhado sem distribuição avançada	Consumidores com distribuição avançada
Throughput de leitura	Fixado em um total de 2 MB/sec por fragmento. Se houver vários consumidores lendo a partir do mesmo fragmento, todos eles compartilham esse throughput. A soma das taxas de transferência que eles recebem do fragmento não excede 2 MB/s.	Dimensionada de acordo com o registro dos consumidores para usar a distribuição avançada. Cada consumidor registrado para usar a distribuição avançada recebe seu próprio throughput de leitura por fragmento, de até 2 MB/s, independentemente de outros consumidores.
Atraso de propagação da mensagem	Uma média de cerca de 200 ms se houver um consumidor lendo no fluxo. Essa média chega até cerca de 1000 ms se houver cinco consumidores.	Normalmente, uma média de 70 ms se houver um ou cinco consumidores.
Custo	Não aplicável	Há um custo de recuperação de dados e um custo de hora de fragmento por consumidor. Para obter mais informaçõ

Características	Consumidores de throughput compartilhado sem distribuição avançada	Consumidores com distribuição avançada
		es, consulte Definição de preço do Amazon Kinesis Data Streams.
Registro de modelo de entrega	Extraia o modelo por meio de HTTP usando GetRecords.	O Kinesis Data Streams envia os registros para você por meio de HTTP/2 usando. SubscribeToShard

Regiões com suporte para até 50 consumidores de fan-out aprimorados (somente sob demanda Advantage)

Support para até 50 consumidores de fan-out aprimorados no modo On-Demand Advantage está disponível somente nas seguintes regiões: AWS

AWS Região	Nome da região
eu-north-1	Europa (Estocolmo)
me-south-1	Oriente Médio (Bahrein)
ap-south-1	Ásia-Pacífico (Mumbai)
eu-west-3	Europa (Paris)
ap-southeast-3	Ásia-Pacífico (Jacarta)
us-east-2	Leste dos EUA (Ohio)
af-south-1	África (Cidade do Cabo)
eu-west-1	Europa (Irlanda)
me-central-1	Oriente Médio (Emirados Árabes Unidos)

AWS Região	Nome da região
eu-central-1	Europa (Frankfurt)
sa-east-1	América do Sul (São Paulo)
ap-east-1	Ásia-Pacífico (Hong Kong)
ap-south-2	Ásia-Pacífico (Hyderabad)
us-east-1	Leste dos EUA (Norte da Virgínia)
ap-northeast-2	Ásia-Pacífico (Seul)
ap-northeast-3	Ásia-Pacífico (Osaka)
eu-west-2	Europa (Londres)
ap-southeast-4	Ásia-Pacífico (Melbourne)
ap-northeast-1	Ásia-Pacífico (Tóquio)
us-west-2	Oeste dos EUA (Oregon)
us-west-1	Oeste dos EUA (N. da Califórnia)
ap-southeast-1	Ásia-Pacífico (Singapura)
ap-southeast-2	Ásia-Pacífico (Sydney)
il-central-1	Israel (Tel Aviv)
ca-central-1	Canadá (Central)
ca-west-1	Oeste do Canadá (Calgary)
eu-south-2	Europa (Espanha)
cn-northwest-1	China (Ningxia)
eu-central-2	Europa (Zurique)

AWS Região	Nome da região
us-gov-east-1	AWS GovCloud (Leste dos EUA)
us-gov-west-1	AWS GovCloud (Oeste dos EUA)

Gerencie consumidores de distribuição aprimorados com o ou AWS CLI APIs

Os consumidores que usam a distribuição avançada no Amazon Kinesis Data Streams podem receber registros de um fluxo de dados com throughput dedicada de até 2 MB de dados por segundo por fragmento. Para obter mais informações, consulte Desenvolver consumidores de distribuição avançada com throughput dedicado.

Você pode usar o AWS CLI Kinesis APIs Data Streams para registrar, descrever, listar e cancelar o registro de um consumidor que usa fan-out aprimorado no Kinesis Data Streams.

Gerencie consumidores usando o AWS CLI

Você pode registrar, descrever, listar e cancelar o registro de consumidores de fan-out aprimorados usando o. AWS CLI Para obter exemplos, consulte a seguinte documentação.

register-stream-consumer

Registra um consumidor para um fluxo de dados do Kinesis. Você pode aplicar tags ao registrar o consumidor.

describe-stream-consumer

Obtém a descrição de um consumidor registrado com ARN do consumidor ou a combinação de nome e fluxo de ARN do consumidor.

list-stream-consumers

Lista os consumidores registrados para receber dados de um fluxo usando a distribuição avançada.

deregister-stream-consumer

Cancela o registro de um consumidor com ARN de consumidor ou a combinação de nome e fluxo de ARN do consumidor.

Gerencie consumidores usando o Kinesis Data Streams APIs

Você pode registrar, descrever, listar e cancelar o registro de consumidores de fan-out aprimorados usando o Kinesis Data Streams. APIs Para obter exemplos, consulte a seguinte documentação.

RegisterStreamConsumer

Registra um consumidor para um fluxo de dados do Kinesis com tags. Você pode aplicar tags ao registrar o consumidor.

DescribeStreamConsumer

Obtém a descrição de um consumidor registrado com ARN do consumidor ou a combinação de nome e fluxo de ARN do consumidor.

ListStreamConsumers

Lista os consumidores registrados para receber dados de um fluxo usando a distribuição avançada.

DeregisterStreamConsumer

Cancela o registro de um consumidor com ARN de consumidor ou a combinação de nome e fluxo de ARN do consumidor.

Atribuir tags a consumidores

É possível atribuir os próprios metadados aos fluxos e aos consumidores de distribuição avançada criados no Kinesis Data Streams na forma de tags. É possível usar tags para categorizar e monitorar os custos dos consumidores. Você também pode controlar o acesso aos consumidores usando tags com controle de acesso por atributo (ABAC). Para obter mais informações, consulte Marcar com tag seus recursos do Amazon Kinesis Data Streams.

Usar o visualizador de dados no console do Kinesis

O visualizador de dados no console de gerenciamento do Kinesis permite a visualização de registros de dados dentro do fragmento especificado do fluxo de dados sem precisar desenvolver uma aplicação de consumo. Para usar o visualizador de dados, siga estas etapas:

1. Faça login no Console de gerenciamento da AWS e abra o console do Kinesis em https://console.aws.amazon.com /kinesis.

2. Escolha o fluxo de dados ativo cujos registros serão visualizados e, em seguida, escolha a guia Visualizador de dados.
3. Na guia Visualizador de dados do fluxo de dados ativo selecionado, escolha o fragmento cujos registros serão visualizados, escolha a Posição inicial e clique em Obter registros. É possível definir a posição inicial como um dos seguintes valores:

   - No número de sequência: mostra os registros da posição indicada pelo número de sequência especificado no campo de número de sequência.
   - Depois do número de sequência: mostra os registros imediatamente após a posição indicada pelo número de sequência especificado no campo de número de sequência.
   - No carimbo de data/hora: mostra os registros da posição indicada pelo timestamp especificado no campo de timestamp.
   - Horizonte de corte: mostra os registros no último registro não cortado no fragmento, que é o registro de dados mais antigo no fragmento.
   - Mais recentes: mostra os registros imediatamente após o registro mais recente no fragmento, para que os dados mais recentes no fragmento sempr esejam lidos.

   Os registros de dados gerados que correspondem ao ID do fragmento e à posição inicial especificados são exibidos em uma tabela de registros no console. São exibidos no máximo 50 registros de cada vez. Para visualizar o próximo conjunto de registros, clique no botão Próximo.
4. Clique em qualquer registro individual para visualizar a carga útil do registro em dados brutos ou no formato JSON em uma janela separada.

Observe que quando você clica em Obter registros ou nos botões Avançar no Visualizador de dados, isso invoca a GetRecordsAPI e isso se aplica ao limite da GetRecordsAPI de 5 transações por segundo.

Consultar seus fluxos de dados no console do Kinesis

A guia Data analytics no console do Kinesis Data Streams permite a consulta dos fluxos de dados usando SQL. Para usar esse recurso, siga estas etapas:

1. Faça login no Console de gerenciamento da AWS e abra o console do Kinesis em https://console.aws.amazon.com /kinesis.

2. Escolha o fluxo de dados ativo que deseja consultar e, em seguida, escolha a guia Data analytics.
3. Na guia Data Analytics, é possível realizar a inspeção e a visualização do fluxo com um notebook gerenciado do Apache Flink Studio. É possível realizar consultas SQL ad-hoc para inspecionar seu fluxo de dados e visualizar os resultados em segundos usando o Apache Zeppelin. Na guia Data analytics, escolha Eu concordo e, em seguida, escolha Criar notebook para criar um notebook.
4. Depois que o notebook for criado, escolha Abrir no Apache Zeppelin. Isso abrirá seu notebook em uma nova guia. Um notebook é uma interface interativa na qual é possível enviar suas consultas SQL. Escolha a nota que contém o nome do seu fluxo.
5. Será exibida uma nota com um exemplo de consulta `SELECT` para gerar os dados no fluxo já em execução. Isso permite a visualização do esquema do seu fluxo de dados.
6. Para testar outras consultas, como janelas em cascata ou deslizantes, escolha Exibir exemplos de consultas na guia Data analytics. Copie a consulta, modifique-a de acordo com seu esquema de fluxo de dados e, em seguida, execute-a em um novo parágrafo em sua nota do Zeppelin.

Usar a Kinesis Client Library

O que é a Kinesis Client Library?

A Kinesis Client Library (KCL) é uma biblioteca de software Java independente projetada para simplificar o processo de consumo e processamento de dados do Amazon Kinesis Data Streams. A KCL lida com muitas das tarefas complexas associadas à computação distribuída, o que permite que os desenvolvedores se concentrem na implementação da lógica de negócios para o processamento dos dados. Ela gerencia atividades como balanceamento de carga entre vários operadores, resposta a falhas de operadores, verificação dos registros processados e resposta a mudanças no número de fragmentos no fluxo.

A KCL é atualizada com frequência para incorporar as versões mais recentes das bibliotecas subjacentes, promover melhorias de segurança e realizar correções de bugs. É recomendável que você use a versão mais recente da KCL para evitar os problemas já conhecidos e se beneficiar de todas as melhorias mais recentes. Para encontrar a versão mais recente da KCL, consulte Github da KCL.

Important

- É recomendável que você use a versão mais recente da KCL para evitar os bugs e problemas já conhecidos. Se você estiver usando a KCL 2.6.0 ou versão anterior, atualize para a KCL 2.6.1 ou versão posterior para evitar uma condição rara que pode bloquear o processamento de fragmentos quando a capacidade do fluxo for alterada.
- A KCL é uma biblioteca Java. Support para linguagens diferentes de Java é fornecido usando um daemon baseado em Java chamado. MultiLangDaemon MultiLangDaemoninterage com o aplicativo KCL por meio de STDIN e STDOUT. Para obter mais informações sobre o MultiLangDaemon on GitHub, consulteDesenvolva consumidores com a KCL em linguagens não Java.
- Não use as AWS SDK para Java versões 2.27.19 a 2.27.23 com KCL 3.x. Essas versões têm um problema que causa um erro de exceção relacionado ao uso do DynamoDB da KCL. Recomendamos que você use a AWS SDK para Java versão 2.28.0 ou posterior para evitar esse problema.

Principais atributos e benefícios da KCL

A seguir, alguns dos principais atributos e benefícios correspondentes da KCL:

- Escalabilidade: a KCL permite que as aplicações sejam escaladas dinamicamente distribuindo a carga de processamento entre vários operadores. Você pode escalar sua aplicação, removendo ou expandindo manualmente ou com ajuste de escala automático, sem se preocupar com a redistribuição de carga.
- Balanceamento de carga: a KCL equilibra automaticamente a carga de processamento entre os operadores disponíveis, resultando em uma distribuição uniforme do trabalho entre os operadores.
- Ponto de verificação: a KCL gerencia o ponto de verificação dos registros processados, permitindo que as aplicações retomem o processamento a partir da última posição processada com êxito.
- Tolerância a falhas: a KCL fornece mecanismos integrados de tolerância a falhas, garantindo que o processamento de dados continue, mesmo que operadores individuais falhem. A KCL também fornece at-least-once entrega.
- Tratamento de mudanças no fluxo: a KCL se adapta às divisões e mesclagens de fragmentos que podem ocorrer devido a alterações no volume de dados. Ela mantém o ordenamento ao

assegurar que os fragmentos secundários sejam processados somente após o fragmento principal ser concluído e verificado.

- Monitoramento: a KCL se integra à Amazon CloudWatch para monitoramento em nível de consumidor.
- Suporte a vários idiomas: o KCL oferece suporte nativo a Java e habilita várias linguagens de programação não Java. MultiLangDaemon

Conceitos da KCL

Essa seção explica os principais conceitos e interações da Kinesis Client Library (KCL). Esses conceitos são fundamentais para desenvolver e gerenciar aplicações consumidoras da KCL.

- Aplicação de consumo da KCL: uma aplicação personalizada projetada para ler e processar os registros dos fluxos de dados do Kinesis usando a Kinesis Client Library.
- Operador: as aplicações consumidoras da KCL são distribuídas de modo típico, com um ou mais operadores em execução simultânea. A KCL coordena os operadores para que consumam os dados do fluxo de forma distribuída e equilibra a carga uniformemente entre vários operadores.
- Agendador: uma classe de alto nível que um operador da KCL usa para começar a processar os dados. Cada operador da KCL tem um agendador. O agendador inicializa e supervisiona várias tarefas, incluindo a sincronização de informações de fragmentos dos fluxos de dados do Kinesis, o monitoramento de atribuições de fragmentos entre os operadores e o processamento dos dados do fluxo com base nos fragmentos atribuídos ao operador. O agendador pode usar várias configurações que afetam seu comportamento, como o nome do fluxo a ser processado e as credenciais da AWS . O agendador inicia a entrega dos registros de dados do fluxo para os processadores de registros.
- Processador de registros: define a lógica pela qual a aplicação de consumo da KCL processa os dados obtidos dos fluxos de dados. É necessário implementar sua lógica de processamento de dados personalizada no processador de registros. Um operador da KCL instancia um agendador. Em seguida, o agendador instancia um processador de registros para cada fragmento para o qual tenha concessão. Um operador pode executar vários processadores de registros.
- Concessão: define a atribuição entre um operador e um fragmento. As aplicações consumidoras da KCL usam concessões para distribuir o processamento de registros de dados por vários operadores. Cada fragmento está vinculado a apenas um operador por meio de uma concessão em um determinado momento e cada operador pode manter um ou mais concessões simultaneamente. Quando um operador deixa de manter uma concessão devido a interrupção

ou falha, a KCL atribui outro operador para assumir a concessão. Para mais detalhes sobre a concessão, consulte Documentação do Github: ciclo de vida da concessão.

- Tabela de concessão: é uma tabela exclusiva do Amazon DynamoDB usada para rastrear todas as concessões da aplicação de consumo da KCL. Cada aplicação de consumo da KCL cria sua própria tabela de concessões. A tabela de concessões é usada para manter o estado de todos os operadores para coordenar o processamento de dados. Para obter mais informações, consulte Tabelas de metadados do DynamoDB e balanceamento de carga na KCL.
- Ponto de verificação: é o processo de armazenar persistentemente a posição do último registro processado com sucesso em um fragmento. A KCL gerencia o ponto de verificação para garantir que o processamento possa ser retomado a partir da última posição desse ponto se um operador falhar ou a aplicação for reiniciada. Os pontos de verificação são armazenados na tabela de concessão do DynamoDB como parte dos metadados da concessão. Isso permite que os operadores continuem o processamento a partir do ponto em que o operador anterior parou.

Tabelas de metadados do DynamoDB e balanceamento de carga na KCL

A KCL gerencia metadados, como concessões e métricas de utilização da CPU dos operadores. A KCL rastreia esses metadados usando as tabelas do DynamoDB. Para cada aplicação do Amazon Kinesis Data Streams, a KCL cria três tabelas do DynamoDB para gerenciar os metadados: tabela de concessão, tabela de métricas do operador e tabela de estado do coordenador.

> **Note**
>
> A KCL 3.x introduziu duas novas tabelas de metadados: métricas do operador e tabelas de estado do coordenador.

> **Important**
>
> Adicione as permissões adequadas às aplicações da KCL para criar e gerenciar tabelas de metadados no DynamoDB. Para obter detalhes, consulte Permissões do IAM necessárias para aplicativos de consumo da KCL.
> O aplicação de consumo da KCL não remove automaticamente essas três tabelas de metadados do DynamoDB. Remova as tabelas de metadados do DynamoDB criadas pela aplicação de consumo da KCL ao descomissionar sua aplicação de consumo para evitar custos desnecessários.

Tabela de concessões

A tabela de concessões é uma tabela exclusiva do Amazon DynamoDB usada para monitorar os fragmentos cedidos e processados pelos agendadores da aplicação de consumo da KCL. Cada aplicação de consumo da KCL cria sua própria tabela de concessões. A KCL usa o nome da aplicação de consumo como nome da tabela de concessões por padrão. Você pode definir um nome de tabela personalizado usando a configuração. A KCL também cria um índice secundário global na tabela de concessões com a chave de partição do leaseOwner para uma descoberta eficiente da concessão. O índice secundário global reflete o atributo leaseKey da tabela básica de concessões. Se a tabela de concessões da aplicação de consumo da KCL não existir quando a aplicação for inicializada, um dos operadores a criará.

É possível visualizar a tabela usando o console do Amazon DynamoDB enquanto a aplicação de consumo está em execução.

> **Important**
>
> - Cada nome de aplicação de consumo da KCL deve ser exclusivo para evitar a duplicação do nome da tabela de concessões.
> - Sua conta é cobrada pelos custos associados à tabela do DynamoDB, além dos custos associados ao próprio Kinesis Data Streams.

Cada linha na tabela de concessões representa um fragmento que está sendo processado pelos agendadores da aplicação de consumo. Os campos principais incluem o seguinte:

- leaseKey: para processamento de fluxo único, essa é a ID do fragmento. Para o processamento de vários fluxos com a KCL, ele é estruturado como `account-id:StreamName:streamCreationTimestamp:ShardId`. A leaseKey é a chave de partição da tabela de concessões. Para obter mais informações sobre o processamento de vários fluxos, consulte Processamento de vários fluxos com a KCL.
- checkpoint: número de sequência do ponto de verificação mais recente do fragmento.
- checkpointSubSequenceNúmero: ao usar o recurso de agregação da Kinesis Producer Library, essa é uma extensão do ponto de verificação que rastreia registros individuais de usuários dentro do registro do Kinesis.

- leaseCounter: usado para verificar se um operador está processando atualmente a concessão de modo ativo. O leaseCounter aumenta se a propriedade da concessão for transferida para outro operador.
- leaseOwner: é o operador atual que detém essa concessão.
- ownerSwitchesSincePonto de controle: Quantas vezes esse contrato mudou de trabalhadores desde o último posto de controle.
- parentShardId: ID do pai desse fragmento. Garante que o fragmento principal seja totalmente processado antes do início do processamento dos fragmentos secundários, mantendo a ordem correta de processamento do registro.
- childShardId: Lista de fragmentos secundários IDs resultantes da divisão ou mesclagem desse fragmento. Usado para rastrear a linhagem de fragmentos e gerenciar a ordem de processamento durante as operações de refragmentação.
- startingHashKey: o limite inferior do intervalo de chaves de hash desse fragmento.
- endingHashKey: o limite superior do intervalo de chaves de hash desse fragmento.

Se adotar o processamento de vários fluxos com a KCL, você visualizará os dois campos adicionais a seguir na tabela de concessões. Para obter mais informações, consulte Processamento de vários fluxos com a KCL.

- shardID: o ID do fragmento.
- streamName: é o identificador do fluxo de dados no formato: `account-id:StreamName:streamCreationTimestamp`.

Tabela de métricas do operador

A tabela de métricas do operador é uma tabela exclusiva do Amazon DynamoDB para cada aplicação da KCL e é usada para registrar as métricas de utilização da CPU de cada operador. Essas métricas serão usadas pela KCL para atribuir concessões de modo eficiente, resultando no uso equilibrado dos recursos entre os operadores. A KCL usa `KCLApplicationName-WorkerMetricStats` como nome da tabela de métricas do operador por padrão.

Tabela de estados do coordenador

A tabela de estados do coordenador é uma tabela exclusiva do Amazon DynamoDB para cada aplicação da KCL e é usada para armazenar informações de estado internas dos operadores.

Por exemplo, a tabela de estados do coordenador armazena dados sobre a eleição do líder ou os metadados associados à migração local da KCL 2.x para a KCL 3.x. A KCL usa `KCLApplicationName-CoordinatorState` como nome da tabela de estados do coordenador por padrão.

Modo de capacidade do DynamoDB para tabelas de metadados criadas pela KCL

Por padrão, a Kinesis Client Library (KCL) cria tabelas de metadados do DynamoDB, como tabela de concessões, tabela de métricas de operadores e tabela de estados do coordenador usando o modo de capacidade sob demanda. Esse modo dimensiona automaticamente a capacidade de leitura e gravação para acomodar o tráfego sem exigir planejamento de capacidade. É altamente recomendável que você mantenha o modo de capacidade como modo sob demanda para uma operação mais eficiente dessas tabelas de metadados.

Se você decidir mudar a tabela de concessão para o modo de capacidade provisionada, siga estas práticas recomendadas:

- Analise os padrões de uso:
  - Monitore os padrões e usos de leitura e gravação do seu aplicativo (RCU, WCU) usando métricas da Amazon. CloudWatch
  - Entenda os requisitos de pico e médios de throughput.
- Calcule a capacidade necessária:
  - Estime as unidades de capacidade de leitura (RCUs) e as unidades de capacidade de gravação (WCUs) com base em sua análise.
  - Considere fatores como o número de fragmentos, a frequência dos pontos de verificação e o número de operadores.
- Implemente o ajuste de escala automático:
  - use o ajuste de escala automático do DynamoDB para ajustar automaticamente a capacidade provisionada e definir os limites de capacidade mínima e máxima apropriados.
  - O ajuste de escala automático do DynamoDB ajudará a evitar que a tabela de metadados da KCL atinja o limite de capacidade e a utilização passe a ser controlada.
- Monitoramento e otimização regulares:
  - Monitore continuamente CloudWatch as métricas de`ThrottledRequests`.
  - Ajuste a capacidade à medida que sua workload for alterada ao longo do tempo.

Se você tiver `ProvisionedThroughputExceededException` nas tabelas de metadados do DynamoDB para sua aplicação de consumo da KCL, você precisará aumentar a capacidade de throughput provisionada da tabela do DynamoDB. Se você definir um determinado nível de unidades de capacidade de leitura (RCU) e de unidades de capacidade de gravação (WCU) na primeira vez que criar sua aplicação de consumo, esse nível pode não ser suficiente à medida que seu uso aumentar. Por exemplo, se uma aplicação de consumo da KCL realiza pontos de verificação com frequência ou opera em um fluxo com vários fragmentos, talvez sejam necessárias unidades com capacidade maior. Para obter informações sobre throughput provisionada no DynamoDB, consulte a capacidade de throughput do DynamoDB e atualização de uma tabela no Guia do desenvolvedor do Amazon DynamoDB.

Como a KCL atribui concessões aos operadores e equilibra a carga

A KCL coleta e monitora continuamente as métricas de utilização da CPU dos hosts de computação que executam os operadores para garantir uma distribuição uniforme de workload. Essas métricas de utilização da CPU são armazenadas na tabela de métricas do operador no DynamoDB. Se a KCL detectar que alguns operadores apresentam taxas de utilização da CPU mais altas em comparação com outros, ela reatribuirá as concessões entre os operadores para reduzir a carga dos operadores mais usados. O objetivo é equilibrar a workload de forma mais uniforme em toda a frota de aplicações de consumo, evitando que algum dos operadores fique sobrecarregado. À medida que a KCL distribui a utilização da CPU em toda a frota de aplicações de consumo, você pode dimensionar corretamente a capacidade da frota da aplicação de consumo escolhendo o número certo de operadores ou usar o ajuste de escala automático para gerenciar com eficiência a capacidade computacional para obter custos mais baixos.

> **Important**
>
> A KCL pode coletar métricas de utilização da CPU dos operadores somente se determinados pré-requisitos forem atendidos. Para obter detalhes, consulte Pré-requisitos. Se a KCL não conseguir coletar dos operadores as métricas de utilização da CPU, a KCL voltará a usar o throughput por operador para atribuir concessões e equilibrar a carga entre os operadores da frota. A KCL monitorará o throughput que cada operador recebe em um determinado momento e reatribuirá as concessões para garantir que cada operador receba um nível de throughput total semelhante das concessões atribuídas.

Desenvolver consumidores com a KCL

É possível usar a Kinesis Client Library (KCL) para criar aplicações de consumo que processam dados dos fluxos de dados do Kinesis.

A KCL está disponível em várias linguagens. Este tópico aborda como desenvolver consumidores da KCL em linguagens Java e não Java.

- Para conhecer a referência em Javadoc da Kinesis Client Library, consulte Javadoc da Amazon Kinesis Client Library.
- Para baixar o KCL para Java em GitHub, consulte a Biblioteca de cliente do Amazon Kinesis para Java.
- Para localizar a KCL para Java no Apache Maven, consulte o Repositório central do Maven da KCL.

Tópicos

- Desenvolver consumidores com a KCL em Java
- Desenvolva consumidores com a KCL em linguagens não Java

Desenvolver consumidores com a KCL em Java

Pré-requisitos

Antes de começar a usar a KCL 3.x, certifique-se de ter os pré-requisitos apresentados a seguir:

- Java Development Kit (JDK) 8 ou posterior
- AWS SDK para Java 2. x
- Maven ou Gradle para gerenciamento de dependências

A KCL coleta métricas de utilização da CPU, como a utilização da CPU, do host de computação executado pelos operadores para equilibrar a carga e alcançar um nível uniforme de utilização de recursos entre os operadores. Para permitir que a KCL colete métricas de utilização da CPU dos operadores, é necessário que os seguintes pré-requisitos sejam atendidos:

Amazon Elastic Compute Cloud(Amazon EC2)

- Seu sistema operacional deve ser Linux OS.

- Você deve habilitar IMDSv2em sua instância do EC2.

Amazon Elastic Container Service (Amazon ECS) no Amazon EC2

- Seu sistema operacional deve ser Linux OS.
- É necessário ativar o endpoint de metadados de tarefas do ECS versão 4.
- Sua versão do atendente do contêiner do Amazon ECS deve ser 1.39.0 ou posterior.

Amazon ECS em AWS Fargate

- É necessário habilitar o endpoint de metadados de tarefas do Fargate versão 4. Se você usar a versão da plataforma 1.4.0 ou posterior do Fargate, essa opção fica habilitada por padrão.
- Fargate versão da plataforma 1.4.0 ou posterior.

Amazon Elastic Kubernetes Service (Amazon EKS) no Amazon EC2

- Seu sistema operacional deve ser Linux OS.

Amazon EKS em AWS Fargate

- Plataforma 1.3.0 ou posterior do Fargate.

> **Important**
>
> Se a KCL não conseguir coletar dos operadores as métricas de utilização da CPU, a KCL voltará a usar o throughput por operador para atribuir concessões e equilibrar a carga entre os operadores da frota. Para obter mais informações, consulte Como a KCL atribui concessões aos operadores e equilibra a carga.

Instalar e adicionar dependências

Se estiver usando Maven, adicione a dependência a seguir ao seu arquivo `pom.xml`. Certifique-se de ter substituído 3.x.x pela versão mais recente da KCL.

```
<dependency>
    <groupId>software.amazon.kinesis</groupId>
```

```
    <artifactId>amazon-kinesis-client</artifactId>
    <version>3.x.x</version> <!-- Use the latest version -->
</dependency>
```

Se você estiver usando Gradle, adicione o seguinte ao seu arquivo `build.gradle`. Certifique-se de ter substituído 3.x.x pela versão mais recente da KCL.

```
implementation 'software.amazon.kinesis:amazon-kinesis-client:3.x.x'
```

A versão mais recente da KCL pode ser obtida no Repositório central do Maven.

Implementar o consumidor

Uma aplicação de consumo da KCL consiste nos seguintes componentes principais:

Componentes principais

- RecordProcessor
- RecordProcessorFactory
- Agendador
- Aplicação de consumo principal

RecordProcessor

RecordProcessor é o componente principal em que reside sua lógica de negócios para processar registros de stream de dados do Kinesis. Ele define como sua aplicação processa os dados que recebe do fluxo do Kinesis.

Principais responsabilidades:

- Inicializar o processamento de um fragmento
- Processar lotes de registros do fluxo do Kinesis
- Encerrar o processamento de um fragmento (por exemplo, quando o fragmento é dividido ou mesclado ou ainda quando a concessão é entregue a outro host)
- Tratar do ponto de verificação para acompanhar o progresso

A seguir, um exemplo de implementação:

```
import org.slf4j.Logger;
```

```
import org.slf4j.LoggerFactory;
import org.slf4j.MDC;
import software.amazon.kinesis.exceptions.InvalidStateException;
import software.amazon.kinesis.exceptions.ShutdownException;
import software.amazon.kinesis.lifecycle.events.*;
import software.amazon.kinesis.processor.ShardRecordProcessor;

public class SampleRecordProcessor implements ShardRecordProcessor {
    private static final String SHARD_ID_MDC_KEY = "ShardId";
    private static final Logger log =
 LoggerFactory.getLogger(SampleRecordProcessor.class);
    private String shardId;

    @Override
    public void initialize(InitializationInput initializationInput) {
        shardId = initializationInput.shardId();
        MDC.put(SHARD_ID_MDC_KEY, shardId);
        try {
            log.info("Initializing @ Sequence: {}",
 initializationInput.extendedSequenceNumber());
        } finally {
            MDC.remove(SHARD_ID_MDC_KEY);
        }
    }

    @Override
    public void processRecords(ProcessRecordsInput processRecordsInput) {
        MDC.put(SHARD_ID_MDC_KEY, shardId);
        try {
            log.info("Processing {} record(s)", processRecordsInput.records().size());
            processRecordsInput.records().forEach(r ->
                log.info("Processing record pk: {} -- Seq: {}", r.partitionKey(),
 r.sequenceNumber())
            );

            // Checkpoint periodically
            processRecordsInput.checkpointer().checkpoint();
        } catch (Throwable t) {
            log.error("Caught throwable while processing records. Aborting.", t);
        } finally {
            MDC.remove(SHARD_ID_MDC_KEY);
        }
    }
```

```
    @Override
    public void leaseLost(LeaseLostInput leaseLostInput) {
        MDC.put(SHARD_ID_MDC_KEY, shardId);
        try {
            log.info("Lost lease, so terminating.");
        } finally {
            MDC.remove(SHARD_ID_MDC_KEY);
        }
    }

    @Override
    public void shardEnded(ShardEndedInput shardEndedInput) {
        MDC.put(SHARD_ID_MDC_KEY, shardId);
        try {
            log.info("Reached shard end checkpointing.");
            shardEndedInput.checkpointer().checkpoint();
        } catch (ShutdownException | InvalidStateException e) {
            log.error("Exception while checkpointing at shard end. Giving up.", e);
        } finally {
            MDC.remove(SHARD_ID_MDC_KEY);
        }
    }

    @Override
    public void shutdownRequested(ShutdownRequestedInput shutdownRequestedInput) {
        MDC.put(SHARD_ID_MDC_KEY, shardId);
        try {
            log.info("Scheduler is shutting down, checkpointing.");
            shutdownRequestedInput.checkpointer().checkpoint();
        } catch (ShutdownException | InvalidStateException e) {
            log.error("Exception while checkpointing at requested shutdown. Giving
 up.", e);
        } finally {
            MDC.remove(SHARD_ID_MDC_KEY);
        }
    }
}
```

Veja a seguir uma explicação detalhada de cada método usado no exemplo:

inicializar (InitializationInputInitializationInput)

- Objetivo: configurar todos os recursos ou estados necessários para processar registros.

- Quando é chamado: uma vez, quando a KCL atribui um fragmento a esse processador de registros.
- Principais pontos:
  - `initializationInput.shardId()`: a ID do fragmento que esse processador manipulará.
  - `initializationInput.extendedSequenceNumber()`: o número sequencial a partir do qual o processamento será iniciado.

processRecords () ProcessRecordsInput processRecordsInput

- Objetivo: processar os registros recebidos e, opcionalmente, verificar o progresso.
- Quando é chamado: repetidamente, desde que o processador de registros mantenha a concessão do fragmento.
- Principais pontos:
  - `processRecordsInput.records()`: lista de registros a serem processados.
  - `processRecordsInput.checkpointer()`: usado para verificar o progresso.
  - Verifique se todas as exceções foram tratadas durante o processamento para evitar que a KCL falhe.
  - Esse método deve ser idempotente, pois o mesmo registro pode ser processado mais de uma vez em alguns cenários, como dados que não foram verificados antes de falhas ou reinicializações inesperadas do operador.
  - Sempre limpe todos os dados armazenados em buffer antes de verificar para garantir a consistência de dados.

Locação perdida () LeaseLostInput leaseLostInput

- Objetivo: limpar todos os recursos específicos para o processamento desse fragmento.
- Quando é chamado: quando outro Agendador assume a concessão desse fragmento.
- Principais pontos:
  - A verificação não é permitida neste método.

Encerrado () ShardEndedInput shardEndedInput

- Objetivo: concluir o processamento desse fragmento e verificar.

- Quando é chamado: quando o fragmento é dividido ou mesclado, indicando que todos os dados desse fragmento foram processados.
- Principais pontos:
  - `shardEndedInput.checkpointer()`: usado para realizar a verificação final.
  - A verificação nesse método é obrigatória para concluir o processamento.
  - Deixar de liberar os dados e fazer a verificação aqui pode resultar na perda de dados ou no processamento duplicado quando o fragmento for reaberto.

Desligamento solicitado () ShutdownRequestedInput shutdownRequestedInput

- Objetivo: verifique e limpe os recursos quando a KCL estiver desligada.
- Quando é chamado: quando a KCL está sendo encerrada. Por exemplo, quando a aplicação está sendo encerrada).
- Principais pontos:
  - `shutdownRequestedInput.checkpointer()`: usado para realizar o ponto de verificação antes do desligamento.
  - Implemente o ponto de verificação no método para que o andamento seja salvo antes que a aplicação pare.
  - A falha na liberação dos dados e na implementação do ponto de verificação aqui pode resultar na perda de dados ou no reprocessamento de registros quando a aplicação for reiniciada.

> **Important**
>
> A KCL 3.x garante menos reprocessamento de dados quando a concessão é passada de um operador para outro por meio de um ponto de verificação antes que o operador anterior seja desligado. Se não implementar a lógica do ponto de verificação no método `shutdownRequested()`, você não obterá esse benefício. Implemente uma lógica de ponto de verificação dentro do método `shutdownRequested()`.

RecordProcessorFactory

RecordProcessorFactory é responsável pela criação de novas RecordProcessor instâncias. A KCL usa essa fábrica para criar um novo RecordProcessor para cada fragmento que o aplicativo precisa processar.

Principais responsabilidades:

- Crie novas RecordProcessor instâncias sob demanda
- Certifique-se de que cada um RecordProcessor esteja inicializado corretamente

A seguir, um exemplo de implementação:

```
import software.amazon.kinesis.processor.ShardRecordProcessor;
import software.amazon.kinesis.processor.ShardRecordProcessorFactory;

public class SampleRecordProcessorFactory implements ShardRecordProcessorFactory {
    @Override
    public ShardRecordProcessor shardRecordProcessor() {
        return new SampleRecordProcessor();
    }
}
```

Neste exemplo, a fábrica cria um novo SampleRecordProcessor cada vez que shardRecordProcessor () é chamado. Isso pode ser estendido para incluir qualquer lógica de inicialização necessária.

Agendador

O Agendador é um componente de alto nível que coordena todas as atividades da aplicação KCL. Ele é responsável pela orquestração geral do processamento de dados.

Principais responsabilidades:

- Gerencie o ciclo de vida do RecordProcessors
- Gerenciar o gerenciamento de concessão para fragmentos
- Coordenar os pontos de verificação
- Equilibrar a carga de processamento de fragmentos entre vários operadores da sua aplicação
- Gerenciar os sinais de desligamento normal e encerramento da aplicação

Normalmente, o Agendador é criado e iniciado na aplicação principal. Você pode verificar o exemplo de implementação do Agendador na seção a seguir: Aplicação de consumo principal.

Aplicação de consumo principal

A aplicação de consumo principal une todos os componentes. Ela é responsável por configurar o consumidor da KCL, criar os clientes necessários, configurar o Agendador e gerenciar o ciclo de vida da aplicação.

Principais responsabilidades:

- Configurar clientes AWS de serviço (Kinesis, DynamoDB,) CloudWatch
- Configurar a aplicação KCL
- Criar e iniciar o Agendador
- Controlar o desligamento da aplicação

A seguir, um exemplo de implementação:

```
import software.amazon.awssdk.regions.Region;
import software.amazon.awssdk.services.cloudwatch.CloudWatchAsyncClient;
import software.amazon.awssdk.services.dynamodb.DynamoDbAsyncClient;
import software.amazon.awssdk.services.kinesis.KinesisAsyncClient;
import software.amazon.kinesis.common.ConfigsBuilder;
import software.amazon.kinesis.common.KinesisClientUtil;
import software.amazon.kinesis.coordinator.Scheduler;
import java.util.UUID;

public class SampleConsumer {
    private final String streamName;
    private final Region region;
    private final KinesisAsyncClient kinesisClient;

    public SampleConsumer(String streamName, Region region) {
        this.streamName = streamName;
        this.region = region;
        this.kinesisClient =
 KinesisClientUtil.createKinesisAsyncClient(KinesisAsyncClient.builder().region(this.region));
    }

    public void run() {
        DynamoDbAsyncClient dynamoDbAsyncClient =
 DynamoDbAsyncClient.builder().region(region).build();
        CloudWatchAsyncClient cloudWatchClient =
 CloudWatchAsyncClient.builder().region(region).build();
```

```
        ConfigsBuilder configsBuilder = new ConfigsBuilder(
            streamName,
            streamName,
            kinesisClient,
            dynamoDbAsyncClient,
            cloudWatchClient,
            UUID.randomUUID().toString(),
            new SampleRecordProcessorFactory()
        );

        Scheduler scheduler = new Scheduler(
            configsBuilder.checkpointConfig(),
            configsBuilder.coordinatorConfig(),
            configsBuilder.leaseManagementConfig(),
            configsBuilder.lifecycleConfig(),
            configsBuilder.metricsConfig(),
            configsBuilder.processorConfig(),
            configsBuilder.retrievalConfig()
        );

        Thread schedulerThread = new Thread(scheduler);
        schedulerThread.setDaemon(true);
        schedulerThread.start();
    }

    public static void main(String[] args) {
        String streamName = "your-stream-name"; // replace with your stream name
        Region region = Region.US_EAST_1; // replace with your region
        new SampleConsumer(streamName, region).run();
    }
}
```

A KCL cria um consumidor de distribuição avançada (Enhanced Fan-out, EFO) com throughput dedicada por padrão. Para obter mais informações sobre distribuição avançada, consulte Desenvolver consumidores de distribuição avançada com throughput dedicado. Se você tiver menos de 2 consumidores ou não precisar de atrasos de propagação de leitura abaixo de 200 ms, defina a seguinte configuração no objeto do agendador para usar consumidores de throughput compartilhada:

```
configsBuilder.retrievalConfig().retrievalSpecificConfig(new PollingConfig(streamName,
 kinesisClient))
```

O código a seguir é um exemplo de criação de um objeto do agendador que usa consumidores de throughput compartilhada:

Importações:

```
import software.amazon.kinesis.retrieval.polling.PollingConfig;
```

Código:

```
Scheduler scheduler = new Scheduler(
            configsBuilder.checkpointConfig(),
            configsBuilder.coordinatorConfig(),
            configsBuilder.leaseManagementConfig(),
            configsBuilder.lifecycleConfig(),
            configsBuilder.metricsConfig(),
            configsBuilder.processorConfig(),
            configsBuilder.retrievalConfig().retrievalSpecificConfig(new
 PollingConfig(streamName, kinesisClient))
        );/
```

Desenvolva consumidores com a KCL em linguagens não Java

Esta seção aborda a implementação de consumidores que usam a Kinesis Client Library (KCL) em Python, Node.js, .NET e Ruby.

A KCL é uma biblioteca Java. O suporte para linguagens diferentes de Java é fornecido usando uma interface multilíngue chamada `MultiLangDaemon`. Esse daemon baseado em Java é executado em segundo plano quando uma KCL com uma linguagem diferente de Java é utilizada. Portanto, se você instalar a KCL para linguagens não Java e criar a aplicação de consumo inteiramente em linguagens não Java, ainda assim você precisará ter Java instalado no sistema por causa do `MultiLangDaemon`. Além disso, o `MultiLangDaemon` tem algumas configurações padrão que você pode precisar para personalizar de acordo com seu caso de uso (por exemplo, a região da AWS à qual ele se conecta). Para obter mais informações sobre o `MultiLangDaemon` on GitHub, consulte o MultiLangDaemon projeto KCL.

Embora os conceitos principais permaneçam os mesmos em todas as linguagens, existem algumas considerações e implementações específicas de uma linguagem. Para obter os principais conceitos sobre o desenvolvimento de consumo da KCL, consulte Desenvolver consumidores com a KCL em Java. Para obter informações mais detalhadas sobre como desenvolver consumidores de KCL em

Python, Node.js, .NET e Ruby e as atualizações mais recentes, consulte os seguintes repositórios: GitHub

- Python: amazon-kinesis-client-python
- Node.js: amazon-kinesis-client-nodejs
- .NET: amazon-kinesis-client-net
- Ruby: amazon-kinesis-client-ruby

> **Important**
>
> Não use as versões não Java da KCL a seguir se você estiver usando o JDK 8. Essas versões contêm uma dependência (logback) que é incompatível com o JDK 8.
>
> - KCL Python 3.0.2 e 2.2.0
> - KCL Node.js 2.3.0
> - KCL .NET 3.1.0
> - KCL Ruby 2.2.0
>
> Recomendamos que você use as versões lançadas antes ou depois dessas versões afetadas ao trabalhar com o JDK 8.

Processamento de vários fluxos com a KCL

Esta seção descreve as alterações necessárias na KCL que permitem criar aplicações de consumo da KCL que podem processar mais de um fluxo de dados ao mesmo tempo.

> **Important**
>
> - O processamento de vários fluxos é compatível somente na KCL 2.3 ou versão posterior.
> - O processamento de vários fluxos não será compatível se houver consumidores da KCL criados em linguagens não Java e executados com `multilangdaemon`.
> - O processamento de vários fluxos não tem suporte em nenhuma versão da KCL 1.x.

- MultistreamTracker interface
  - Para criar um aplicativo de consumidor que possa processar vários fluxos ao mesmo tempo, você deve implementar uma nova interface chamada MultistreamTracker. Essa interface inclui o método `streamConfigList`, que retorna a lista de fluxos de dados, e suas configurações, a serem processados pela aplicação de consumo da KCL. Observe que os fluxos de dados processados podem ser alterados durante o runtime da aplicação de consumo. `streamConfigList` é chamado periodicamente pela KCL para obter informações das mudanças nos fluxos de dados a serem processados.
  - O `streamConfigList` preenche a StreamConfiglista.

```
package software.amazon.kinesis.common;

import lombok.Data;
import lombok.experimental.Accessors;

@Data
@Accessors(fluent = true)
public class StreamConfig {
    private final StreamIdentifier streamIdentifier;
    private final InitialPositionInStreamExtended initialPositionInStreamExtended;
    private String consumerArn;
}
```

  - Os campos `StreamIdentifier` e `InitialPositionInStreamExtended` são obrigatórios, enquanto `consumerArn` é opcional. Só é necessário fornecer `consumerArn` se você estiver usando a KCL para implementar uma aplicação de consumo de distribuição avançada.
  - Para obter mais informações sobre`StreamIdentifier`, consulte https://github.com/awslabs/amazon-kinesis-client/blob/v2.5.8/amazon-kinesis-client/src/main/java/software/amazon/kinesis/common/StreamIdentifier.java #L129. Para criar um `StreamIdentifier`, recomenda-se a criação de uma instância de vários fluxos a partir do `streamArn` e do `streamCreationEpoch` que esteja disponível na KCL 2.5.0 ou versões posteriores. Na KCL v2.3 e v2.4, que não oferecem suporte ao `streamArm`, crie uma instância multifluxo usando o formato `account-id:StreamName:streamCreationTimestamp`. Esse formato será descontinuado e não terá mais suporte a partir da próxima versão principal.
  - MultistreamTracker também inclui uma estratégia para excluir locações de fluxos antigos na tabela de locação (). formerStreamsLeases DeletionStrategy Observe que a estratégia NÃO PODE ser alterada durante o runtime da aplicação de consumo. Para obter mais informações, consulte https://github.com/awslabs/amazon-kinesis-client/

blob/0c5042dadf794fe988438436252a5a8fe70b6b0 b/amazon-kinesis-client/src/main/java/software/amazon/kinesis/processor/FormerStreamsLeasesDeletionStrategy .java.

•

Ou você pode inicializar ConfigsBuilder com `MultiStreamTracker` se quiser implementar um aplicativo consumidor KCL que processe vários fluxos ao mesmo tempo.

```
* Constructor to initialize ConfigsBuilder with MultiStreamTracker
     * @param multiStreamTracker
     * @param applicationName
     * @param kinesisClient
     * @param dynamoDBClient
     * @param cloudWatchClient
     * @param workerIdentifier
     * @param shardRecordProcessorFactory
     */
    public ConfigsBuilder(@NonNull MultiStreamTracker multiStreamTracker, @NonNull
 String applicationName,
            @NonNull KinesisAsyncClient kinesisClient, @NonNull DynamoDbAsyncClient
 dynamoDBClient,
            @NonNull CloudWatchAsyncClient cloudWatchClient, @NonNull String
 workerIdentifier,
            @NonNull ShardRecordProcessorFactory shardRecordProcessorFactory) {
        this.appStreamTracker = Either.left(multiStreamTracker);
        this.applicationName = applicationName;
        this.kinesisClient = kinesisClient;
        this.dynamoDBClient = dynamoDBClient;
        this.cloudWatchClient = cloudWatchClient;
        this.workerIdentifier = workerIdentifier;
        this.shardRecordProcessorFactory = shardRecordProcessorFactory;
    }
```

- Com o suporte para vários fluxos implementado para sua aplicação de consumo da KCL, cada linha da tabela de concessões da aplicação contém agora a ID do fragmento e o nome do fluxo dos vários fluxos de dados que esta aplicação processa.
- Quando o suporte para vários fluxos para sua aplicação de consumo da KCL é implementado, leaseKey assume a estrutura `account-`

`id:StreamName:streamCreationTimestamp:ShardId`. Por exemplo, `.111111111:multiStreamTest-1:12345:shardId-000000000336`

> **Important**
>
> Quando sua aplicação de consumo da KCL atual está configurada para processar somente um fluxo de dados, a `leaseKey` (que é a chave de partição da tabela de concessões) é a ID do fragmento. Se você reconfigurar uma aplicação de consumo da KCL atual para processar vários fluxos de dados, a tabela de concessões será quebrada, pois a estrutura da `leaseKey` deve ser `account-id:StreamName:StreamCreationTimestamp:ShardId` para oferecer suporte a vários fluxos.

Use o registro do AWS Glue esquema com o KCL

Você pode integrar o Kinesis Data Streams AWS Glue ao registro do esquema. O registro do AWS Glue esquema permite que você descubra, controle e desenvolva esquemas de forma centralizada, ao mesmo tempo em que garante que os dados produzidos sejam validados continuamente por um esquema registrado. O esquema define a estrutura e o formato de um registro de dados. Um esquema é uma especificação versionada para publicação, consumo ou datastore confiáveis. O registro do AWS Glue esquema permite que você melhore a qualidade end-to-end dos dados e a governança de dados em seus aplicativos de streaming. Para obter mais informações, consulte Registro de esquemas do AWS Glue. Uma das formas de configurar essa integração é usando a KCL para Java.

> **Important**
>
> - AWS Glue A integração do registro de esquemas para o Kinesis Data Streams só é suportada no KCL 2.3 ou posterior.
> - AWS Glue A integração do registro de esquemas para o Kinesis Data Streams não é compatível com consumidores de KCL escritos em linguagens não Java que são executadas com. `multilangdaemon`
> - AWS Glue A integração do registro de esquemas para o Kinesis Data Streams não é suportada em nenhuma versão do KCL 1.x.

Para obter instruções detalhadas sobre como configurar a integração do Kinesis Data Streams com o registro do AWS Glue esquema usando o KCL, consulte a seção “Interagindo com dados usando as KPL/KCL bibliotecas” em Caso de uso: integração do Amazon Kinesis Data Streams com o Schema Registry. AWS Glue

Permissões do IAM necessárias para aplicativos de consumo da KCL

É necessário adicionar as permissões a seguir para o perfil do IAM ou o usuário associado à aplicação de consumo da KCL.

Práticas recomendadas de segurança para AWS ditar o uso de permissões refinadas para controlar o acesso a diferentes recursos. AWS Identity and Access Management (IAM) permite gerenciar usuários e permissões de usuários no AWS. Uma Política do IAM lista explicitamente as ações permitidas e os recursos aos quais as ações são aplicáveis.

A tabela a seguir mostra as permissões mínimas do IAM geralmente necessárias para aplicações de consumo da KCL:

Permissões do IAM mínimas para aplicações de consumo da KCL

Serviço	Ações	Recursos (ARNs)	Finalidade

Amazon Kinesis Data Streams	`DescribeStream`
	
	`DescribeStreamSummary`
	
	`RegisterStreamConsumer`
		O fluxo de dados do Kinesis a partir do qual sua aplicação da KCL processará os dados.
		
		`arn:aws:kinesis:region:account:stream/StreamName`
			Antes de tentar ler registros, o consumidor verifica se o fluxo de dados existe, se está ativo e se os fragmentos estão contidos no fluxo de dados.
			
			Registra os consumidores em um fragmento.

Amazon Kinesis Data Streams	`GetRecords`
	
	`GetShardIterator`
		Fluxo de dados do Kinesis a partir do qual sua aplicação da KCL processará os dados.	Lê registros de um fragmento.

Serviço	Ações	Recursos (ARNs)	Finalidade
	`ListShards`	`arn:aws:kinesis:region:account:stream/StreamName`	

Amazon Kinesis Data Streams	`SubscribeToShard`
	
	`DescribeStreamConsumer`
		Fluxo de dados do Kinesis a partir do qual sua aplicação da KCL processará os dados. Adicione essa ação somente se você usar consumidores de distribuição avançada (EFO).
		
		`arn:aws:kinesis:region:account:stream/StreamName/consumer/*`
			Assina um fragmento para consumidores de distribuição avançada (EFO).

Amazon DynamoDB	`CreateTable`
	
	`DescribeTable`
	
	`UpdateTable`
	
	`Scan`
	
	`GetItem`
	
	`PutItem`
	
	`UpdateItem`
	
	`DeleteItem`
		Tabela de concessão (tabela de metadados no DynamoDB criada pela KCL).
		
		`arn:aws:dynamodb:region:account:table/KCLApplicationName`
			Essas ações são necessárias para que a KCL gerencie a tabela de concessões criada no DynamoDB.

Serviço	Ações	Recursos (ARNs)	Finalidade

Amazon DynamoDB	`CreateTable`
	
	`DescribeTable`
	
	`Scan`
	
	`GetItem`
	
	`PutItem`
	
	`UpdateItem`
	
	`DeleteItem`
		Métricas do operador e tabela de estados do coordenador (tabelas de metadados no DynamoDB) criada pela KCL.
		
		`arn:aws:dynamodb:region:account:table/KCLApplicationName-WorkerMetricStats`
		
		`arn:aws:dynamodb:region:account:table/KCLApplicationName-CoordinatorState`
			Essas ações são necessárias para que a KCL gerencie as métricas do operador e as tabelas de metadados do estado do coordenador no DynamoDB.

Serviço	Ações	Recursos (ARNs)	Finalidade
Amazon DynamoDB	`Query`	Índice secundário global na tabela de concessões. `arn:aws:dynamodb:region:account:table/KCLApplicationName/index/*`	Essa ação é necessária para que a KCL leia o índice secundário global da tabela de concessão criada no DynamoDB.
Amazon CloudWatch	`PutMetricData`	*	Faça upload de métricas CloudWatch que sejam úteis para monitorar o aplicativo. O asterisco (*) é usado porque não há nenhum recurso específico CloudWatch no qual a `PutMetricData` ação seja invocada.

Note

Substitua “região”, “conta”StreamName, "e" KCLApplication Nome” no por seu próprio Conta da AWS número Região da AWS, nome do ARNs stream de dados Kinesis e nome do aplicativo KCL, respectivamente. A KCL 3.x cria mais duas tabelas de metadados no DynamoDB. Para obter detalhes sobre as tabelas de metadados do DynamoDB criadas pela KCL, consulte Tabelas de metadados do DynamoDB e balanceamento de carga na KCL. Se você usar as configurações para personalizar os nomes das tabelas de metadados criadas pela KCL, use esses nomes de tabela especificados em vez do nome da aplicação da KCL.

Um exemplo de documento de política para um aplicação de consumo da KCL é mostrado a seguir.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "kinesis:DescribeStream",
                "kinesis:DescribeStreamSummary",
                "kinesis:RegisterStreamConsumer",
                "kinesis:GetRecords",
                "kinesis:GetShardIterator",
                "kinesis:ListShards"
            ],
            "Resource": "arn:aws:kinesis:us-
east-1:123456789012:stream/STREAM_NAME"
        },
        {
            "Effect": "Allow",
            "Action": [
                "kinesis:SubscribeToShard",
                "kinesis:DescribeStreamConsumer"
            ],
            "Resource": "arn:aws:kinesis:us-
east-1:123456789012:stream/STREAM_NAME/consumer/*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "dynamodb:CreateTable",
                "dynamodb:DescribeTable",
                "dynamodb:UpdateTable",
                "dynamodb:GetItem",
                "dynamodb:UpdateItem",
                "dynamodb:PutItem",
                "dynamodb:DeleteItem",
                "dynamodb:Scan"
            ],
            "Resource": [
            "arn:aws:dynamodb:us-east-1:123456789012:table/KCL_APPLICATION_NAME"
```

```
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "dynamodb:CreateTable",
                "dynamodb:DescribeTable",
                "dynamodb:GetItem",
                "dynamodb:UpdateItem",
                "dynamodb:PutItem",
                "dynamodb:DeleteItem",
                "dynamodb:Scan"
            ],
            "Resource": [
            "arn:aws:dynamodb:us-east-1:123456789012:table/KCL_APPLICATION_NAME-
WorkerMetricStats",
    "arn:aws:dynamodb:us-east-1:123456789012:table/KCL_APPLICATION_NAME-
CoordinatorState"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "dynamodb:Query"
            ],
            "Resource": [
            "arn:aws:dynamodb:us-east-1:123456789012:table/KCL_APPLICATION_NAME/
index/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "cloudwatch:PutMetricData"
            ],
            "Resource": "*"
        }
    ]
}
```

Antes de usar esta política de exemplo, verifique os seguintes itens:

- Substitua REGION pela sua Região da AWS (por exemplo, us-east-1).

- Substitua ACCOUNT_ID pelo seu Conta da AWS ID.
- Substitua STREAM_NAME pelo nome do seu fluxo de dados do Kinesis.
- Substitua CONSUMER_NAME pelo nome do seu consumidor, normalmente o nome da sua aplicação ao usar a KCL.
- Substitua KCL_APPLICATION_NAME pelo nome da aplicação da KCL.

Configurações da KCL

Você pode definir as propriedades de configuração para personalizar a funcionalidade da Kinesis Client Library para atender às suas necessidades específicas. A tabela a seguir descreve as classes e propriedades de configuração.

> **Important**
>
> Na KCL 3.x, o algoritmo de balanceamento de carga visa a alcançar uma utilização uniforme da CPU entre os operadores, e não um número igual de concessões por operador. Se a configuração `maxLeasesForWorker` for muito baixa, será possível limitar a capacidade da KCL de equilibrar a workload de forma eficaz. Se você usar a configuração `maxLeasesForWorker`, considere aumentar seu valor para permitir a melhor distribuição de carga possível.

Esta tabela mostra as propriedades de configuração da KCL

Propriedade de configuração	Classe de configuração	Description	Valor padrão
`applicationName`	ConfigsBuilder	O nome da aplicação da KCL. Usado como padrão para o `tableName` e o `consumerName` .	Não aplicável
`tableName`	ConfigsBuilder	Permite substituir o nome usado para a tabela de	Não aplicável

Propriedade de configuração	Classe de configuração	Description	Valor padrão
		concessão do Amazon DynamoDB.	
`streamName`	ConfigsBuilder	O nome do fluxo a partir do qual esse aplicativo processa registros.	Não aplicável
`workerIdentifier`	ConfigsBuilder	Um identificador exclusivo que representa a instanciação do processador do aplicativo. Isso deve ser exclusivo.	Não aplicável
`failoverTimeMillis`	LeaseManagementConfig	O número de milissegundos que devem passar antes que se considere uma falha do proprietário da concessão. No caso de aplicações que têm um grande número de fragmentos, um número maior pode ser definido para reduzir o número de IOPS do DynamoDB necessário para rastrear as concessões.	10.000 (10 segundos)

Propriedade de configuração	Classe de configuração	Description	Valor padrão
`shardSyncIntervalMillis`	LeaseManagementConfig	O tempo entre as chamadas de sincronização de fragmentos.	60.000 (60 segundos)
`cleanupLeasesUponShardCompletion`	LeaseManagementConfig	Quando definidas, as concessões são removidas assim que as concessões filho iniciam o processamento.	TRUE
`ignoreUnexpectedChildShards`	LeaseManagementConfig	Quando definidos, fragmentos filho que possuem um fragmento aberto são ignorados. Essa configuração destina-se principalmente a fluxos do DynamoDB.	FALSE

Propriedade de configuração	Classe de configuração	Description	Valor padrão
`maxLeasesForWorker`	LeaseManagementConfig	O número máximo de concessões que um único operador deve aceitar. Se esse número for definido muito baixo, isso pode causar perda de dados se os operadores não conseguirem processar todos os fragmentos, levando a uma atribuição de concessões abaixo do ideal entre os operadores. Considere a contagem total de fragmentos, o número de operadores e a capacidade de processamento do operador ao configurá-lo.	Ilimitado
`maxLeaseRenewalThreads`	LeaseManagementConfig	Controla o tamanho do grupo de threads de renovação de concessão. Quanto mais concessões seu aplicativo aceitar, maior esse grupo deve ser.	20

Propriedade de configuração	Classe de configuração	Description	Valor padrão
`billingMode`	LeaseManagementConfig	Determina o modo de capacidade e da tabela de concessões criada no DynamoDB. Há duas opções: modo sob demanda (PAY_PER_REQUEST) e modo provisionado. Recomendamos usar a configuração padrão do modo sob demanda, pois ela é escalada automaticamente para acomodar sua workload sem a necessidade de planejamento de capacidade.	PAY_PER_REQUEST (modo sob demanda)

Propriedade de configuração	Classe de configuração	Description	Valor padrão
`initialLeaseTableReadCapacity`	LeaseManagementConfig	A capacidade de leitura do DynamoDB que será usada se a Kinesis Client Library precisar criar uma nova tabela de concessões do DynamoDB com o modo de capacidade provisionada. Você pode ignorar essa configuração se estiver usando o modo de capacidade sob demanda padrão na configuração `billingMode` .	10
`initialLeaseTableWriteCapacity`	LeaseManagementConfig	A capacidade de leitura do DynamoDB que será usada se a Kinesis Client Library precisar criar uma nova tabela de concessões do DynamoDB. Você pode ignorar essa configuração se estiver usando o modo de capacidade sob demanda padrão na configuração `billingMode` .	10

Propriedade de configuração	Classe de configuração	Description	Valor padrão
`initialPositionInStreamExtended`	LeaseManagementConfig	A posição inicial do aplicativo no fluxo. Isso é usado somente durante a criação da concessão inicial.	InitialPositionInStream.TRIM_HORIZON
`reBalanceThresholdPercentage`	LeaseManagementConfig	Um valor percentual que define quando o algoritmo de balanceamento de carga deve considerar a reatribuição de fragmentos entre os operadores. Essa é uma nova configuração introduzida na KCL 3.x.	10
`dampeningPercentage`	LeaseManagementConfig	Um valor percentual usado para amortecer a quantidade de carga que será movida do operador sobrecarregado em uma única operação de rebalanceamento. Essa é uma nova configuração introduzida na KCL 3.x.	60

Propriedade de configuração	Classe de configuração	Description	Valor padrão
`allowThroughputOvershoot`	LeaseManagementConfig	Determina se a concessão adicional ainda precisa ser obtida do operador sobrecarregado, mesmo que isso faça com que a quantidade total de throughput da concessão exceda a quantidade de throughput desejada. Essa é uma nova configuração introduzida na KCL 3.x.	TRUE
`disableWorkerMetrics`	LeaseManagementConfig	Determina se a KCL deve ignorar as métricas de recursos dos operadores (como a utilização da CPU) ao reatribuir concessões e balancear a carga. Defina isso como TRUE se quiser evitar que a KCL balanceie a carga com base na utilização da CPU. Essa é uma nova configuração introduzida na KCL 3.x.	FALSE

Propriedade de configuração	Classe de configuração	Description	Valor padrão
`maxThroughputPerHostKBps`	LeaseManagementConfig	Quantidade de throughput máxima a ser atribuída a um operador durante a atribuição da concessão. Essa é uma nova configuração introduzida na KCL 3.x.	Ilimitado

Propriedade de configuração	Classe de configuração	Description	Valor padrão
`isGracefulLeaseHandoffEnabled`	LeaseManagementConfig	Controla o comportamento da transferência de concessões entre os operadores. Quando definido como verdadeiro, a KCL tentará transferir os arrendamentos normalmente, permitindo que o fragmento RecordProcessor tenha tempo suficiente para concluir o processamento antes de entregar o contrato a outro trabalhador. Isso pode ajudar a garantir integridade dos dados e transições suaves, mas pode aumentar o tempo de transferência. Quando definido como falso, o contrato será entregue imediatamente, sem esperar que o RecordProcessor contrato seja encerrado normalmente. Isso pode levar a transferências mais	TRUE

Propriedade de configuração	Classe de configuração	Description	Valor padrão
		rápidas, mas pode causar um processamento incompleto. Nota: O ponto de verificação deve ser implementado dentro do método shutdownRequested() do RecordProcessor para se beneficiar do recurso elegante de transferência de locação. Essa é uma nova configuração introduzida na KCL 3.x.	

Propriedade de configuração	Classe de configuração	Description	Valor padrão
`gracefulLeaseHandoffTimeoutMillis`	LeaseManagementConfig	Especifica o tempo mínimo (em milissegundos) para esperar que os fragmentos atuais sejam encerrados normalmente antes de transferir o contrato RecordProcessor à força para o próximo proprietário. Se seu método processRecords é executado comumente por mais tempo do que o valor padrão, considere aumentar essa configuração. Isso garante que RecordProcessor tenha tempo suficiente para concluir seu processamento antes que a transferência da locação ocorra. Essa é uma nova configuração introduzida na KCL 3.x.	30.000 (30 segundos)

Propriedade de configuração	Classe de configuração	Description	Valor padrão
`maxRecords`	PollingConfig	Permite definir o número máximo de registros que o Kinesis retorna.	10.000
`retryGetRecordsInSeconds`	PollingConfig	Configura o atraso entre as GetRecords tentativas de falhas.	Nenhum
`maxGetRecordsThreadPool`	PollingConfig	O tamanho do pool de fios usado para GetRecords.	Nenhum
`idleTimeBetweenReadsInMillis`	PollingConfig	Determina quanto tempo a KCL espera entre as GetRecords chamadas para pesquisar os dados dos fluxos de dados. A unidade é milissegundos.	1.500
`callProcessRecordsEvenForEmptyRecordList`	ProcessorConfig	Quando definido, o processador de registros é chamado mesmo quando o Kinesis não fornece nenhum registro.	FALSE

Propriedade de configuração	Classe de configuração	Description	Valor padrão
`parentShardPollIntervalMillis`	CoordinatorConfig	Com que frequência um processador de registros deve sondar a conclusão de fragmentos pai. A unidade é milissegundos.	10.000 (10 segundos)
`skipShardSyncAtWorkerInitializationIfLeaseExist`	CoordinatorConfig	Desative a sincronização de dados de fragmento se a tabela de concessão contiver concessões existentes.	FALSE
`shardPrioritization`	CoordinatorConfig	A priorização de fragmentos a ser usada.	NoOpShardPrioritization

Propriedade de configuração	Classe de configuração	Description	Valor padrão
`ClientVersionConfig`	CoordinatorConfig	Determina em qual modo de compatibilidade de versão da KCL a aplicação será executada. Essa configuração serve somente para a migração de versões anteriores da KCL. Ao migrar para 3.x, é preciso definir essa configuração como `CLIENT_VERSION_CONFIG_COMPATIBLE_WITH_2X`. Essa configuração pode ser removida quando a migração for concluída.	CLIENT_VERSION_CONFIG_3X
`taskBackoffTimeMillis`	LifecycleConfig	Tempo de espera para repetir tarefas da KCL com falha. A unidade é milissegundos.	500 (0,5 segundo)
`logWarningForTaskAfterMillis`	LifecycleConfig	Quanto tempo esperar antes de um aviso ser registrado caso uma tarefa não seja concluída.	Nenhum

Propriedade de configuração	Classe de configuração	Description	Valor padrão
`listShardsBackoffTimeInMillis`	RetrievalConfig	O número de milissegundos de espera entre as chamadas para `ListShards` em caso de falha. A unidade é milissegundos.	1.500 (1,5 segundo)
`maxListShardsRetryAttempts`	RetrievalConfig	O número máximo de novas tentativas de `ListShards` antes de desistir.	50
`metricsBufferTimeMillis`	MetricsConfig	Especifica a duração máxima (em milissegundos) para armazenar métricas em buffer antes de publicá-las em. CloudWatch	10.000 (10 segundos)
`metricsMaxQueueSize`	MetricsConfig	Especifica o número máximo de métricas a serem armazenadas em buffer antes da publicação. CloudWatch	10.000

Propriedade de configuração	Classe de configuração	Description	Valor padrão
`metricsLevel`	MetricsConfig	Especifica o nível de granularidade das CloudWatch métricas a serem ativadas e publicadas. Valores possíveis: NENHUM, RESUMO, DETALHADO.	MetricsLevel.DETALHADO
`metricsEnabledDimensions`	MetricsConfig	Controla as dimensões permitidas para CloudWatch métricas.	Todas as dimensões

Configurações descontinuadas na KCL 3.x

As seguintes propriedades de configuração foram descontinuadas na KCL 3.x:

A tabela mostra as propriedades de configuração descontinuadas da KCL 3.x

Propriedade de configuração	Classe de configuração	Description
`maxLeasesToStealAtOneTime`	LeaseManagementConfig	O número máximo de concessões que um aplicativo deve tentar roubar de uma só vez. A KCL 3.x ignorará essa configuração e reatribuirá as concessões com base na utilização de recursos dos operadores.
`enablePriorityLeaseAssignment`	LeaseManagementConfig	Controla se os operadores devem priorizar concessões muito expiradas (concessõ

Propriedade de configuração	Classe de configuração	Description
		es não renovadas por 3 vezes o tempo de failover) e novas concessões de fragmentos, independe ntemente do número de concessões pretendidas, mas ainda respeitando os limites máximos de concessõe s. A KCL 3.x ignorará essa configuração e sempre distribuirá as concessões expiradas entre os operadore s.

> **Important**
>
> Você precisa ter as propriedades de configuração descontinuadas durante a migração das versões anteriores da KCL para a KCL 3.x. Durante a migração, o operador da KCL iniciará primeiro com o modo compatível com KCL 2.x e passará para o modo de funcionalidade da KCL 3.x quando detectar que todos os operadores da KCL da aplicação estão prontos para executar a KCL 3.x. Essas configurações descontinuadas são necessárias enquanto os operadores da KCL estiverem executando o modo compatível com a KCL 2.x.

Política de ciclo de vida da versão da KCL

Este tópico descreve a política de ciclo de vida da versão para a Amazon Kinesis Client Library (KCL). AWS fornece regularmente novos lançamentos para as versões KCL para oferecer suporte a novos recursos e aprimoramentos, correções de erros, patches de segurança e atualizações de dependências. Recomendamos que você continue up-to-date com as versões do KCL para acompanhar os recursos, as atualizações de segurança e as dependências subjacentes mais recentes. Não recomendamos o uso contínuo de uma versão não compatível da KCL.

O ciclo de vida das principais versões da KCL consiste nas três fases a seguir:

- Disponibilidade geral (GA) — Durante essa fase, a versão principal é totalmente suportada. AWS fornece lançamentos regulares de versões secundárias e de patches que incluem suporte para novos recursos ou atualizações de API para o Kinesis Data Streams, bem como correções de bugs e segurança.
- Modo de manutenção — AWS limita os lançamentos da versão do patch para tratar apenas de correções críticas de bugs e problemas de segurança. A versão principal não receberá atualizações dos novos recursos ou APIs do Kinesis Data Streams.
- E nd-of-support — A versão principal não receberá mais atualizações ou lançamentos. As versões publicadas anteriormente continuarão disponíveis por meio de gerenciadores de pacotes públicos e o código permanecerá ativado GitHub. O uso de uma versão que chegou end-of-support é feito a critério do usuário. Recomendamos que você faça a atualização para a versão principal mais recente.

Versão principal	Fase atual	Data de lançamento	Data do modo de manutenção	End-of-support encontro
KCL 1.x	Modo de manutenção	19/12/2013	2025-04-17	2026-01-30
KCL 2.x	Disponibilidade geral	2018-08-02	--	--
KCL 3.x	Disponibilidade geral	2024-11-06	--	--

Migrar de versões anteriores da KCL

Este tópico explica como migrar das versões anteriores da Kinesis Client Library (KCL).

Quais são as novidades da KCL 3.0?

A Kinesis Client Library (KCL) 3.0 conta com vários aprimoramentos importantes em comparação com as versões anteriores:

- ela reduz os custos de computação para aplicações de consumo ao redistribuir automaticamente o trabalho de operadores com utilização acima da capacidade para os operadores subutilizados na

frota de aplicações de consumo. Esse novo algoritmo de balanceamento de carga garante que a utilização da CPU seja distribuída de modo uniforme entre os operadores e elimina a necessidade do provisionamento excessivo de operadores.

- Ele reduz o custo do DynamoDB associado à KCL ao otimizar as operações de leitura na tabela de concessões.
- Ele minimiza o reprocessamento de dados quando as concessões são transferidas para outro operador ao permitir que o operador atual conclua a verificação dos registros processados.
- Ele é usado AWS SDK for Java 2.x para melhorar o desempenho e os recursos de segurança, removendo totalmente a dependência do AWS SDK para Java 1.x.

Para mais informações, consulte a nota de lançamento da KCL 3.0.

Tópicos

- Migrar da KCL 2.x para a KCL 3.x
- Reverter para a versão anterior da KCL
- Avançar para a KCL 3.x após uma reversão
- Práticas recomendadas para a tabela de concessões com modo de capacidade provisionada
- Migrar da KCL 1.x para a KCL 3.x

Migrar da KCL 2.x para a KCL 3.x

Este tópico fornece step-by-step instruções para migrar seu consumidor do KCL 2.x para o KCL 3.x. A KCL 3.x oferece suporte à migração local de consumidores da KCL 2.x. Você pode continuar consumindo os dados do seu fluxo de dados do Kinesis enquanto migra seus operadores de forma contínua.

> **Important**
>
> A KCL 3.x mantém as mesmas interfaces e métodos da KCL 2.x. Portanto, você não precisa atualizar seu código de processamento de registros durante a migração. No entanto, é necessário definir a configuração adequada e verificar as etapas necessárias para a migração. É altamente recomendável que você siga as etapas de migração a seguir para uma experiência de migração perfeita.

Etapa 1: pré-requisitos

Antes de começar a usar a KCL 3.x, certifique-se de ter os pré-requisitos apresentados a seguir:

- Java Development Kit (JDK) 8 ou posterior
- AWS SDK para Java 2. x
- Maven ou Gradle para gerenciamento de dependências

> **Important**
>
> Não use as AWS SDK para Java versões 2.27.19 a 2.27.23 com KCL 3.x. Essas versões têm um problema que causa um erro de exceção relacionado ao uso do DynamoDB da KCL. Recomendamos que você use a AWS SDK para Java versão 2.28.0 ou posterior para evitar esse problema.

Etapa 2: adicionar dependências

Se estiver usando Maven, adicione a dependência a seguir ao seu arquivo `pom.xml`. Certifique-se de ter substituído 3.x.x pela versão mais recente da KCL.

```
<dependency>
    <groupId>software.amazon.kinesis</groupId>
    <artifactId>amazon-kinesis-client</artifactId>
    <version>3.x.x</version> <!-- Use the latest version -->
</dependency>
```

Se você estiver usando Gradle, adicione o seguinte ao seu arquivo `build.gradle`. Certifique-se de ter substituído 3.x.x pela versão mais recente da KCL.

```
implementation 'software.amazon.kinesis:amazon-kinesis-client:3.x.x'
```

A versão mais recente da KCL pode ser obtida no Repositório central do Maven.

Etapa 3: definir a configuração relacionada à migração

Para migrar da KCL 2.x para a KCL 3.x, é necessário definir o seguinte parâmetro de configuração:

- CoordinatorConfig. clientVersionConfig: essa configuração determina em qual modo de compatibilidade de versão do KCL o aplicativo será executado. Ao

migrar da KCL 2.x para a 3.x, você precisa definir essa configuração como `CLIENT_VERSION_CONFIG_COMPATIBLE_WITH_2X`. Para definir a configuração, adicione a seguinte linha ao criar seu objeto do agendador:

```
configsBuilder.coordiantorConfig().clientVersionConfig(ClientVersionConfig.CLIENT_VERSION_CONFI
```

Veja a seguir um exemplo de como configurar `CoordinatorConfig.clientVersionConfig` para migrar da KCL 2.x para a 3.x. Você pode ajustar outras configurações conforme necessário, com base em seus requisitos específicos:

```
Scheduler scheduler = new Scheduler(
    configsBuilder.checkpointConfig(),

 configsBuilder.coordiantorConfig().clientVersionConfig(ClientVersionConfig.CLIENT_VERSION_CONF
    configsBuilder.leaseManagementConfig(),
    configsBuilder.lifecycleConfig(),
    configsBuilder.metricsConfig(),
    configsBuilder.processorConfig(),
    configsBuilder.retrievalConfig()
);
```

É importante que todos os operadores em sua aplicação de consumo usem o mesmo algoritmo de balanceamento de carga em um determinado momento, uma vez que a KCL 2.x e a 3.x usam algoritmos diferentes para balancear a carga. A execução de operadores com diferentes algoritmos de balanceamento de carga pode causar uma distribuição de carga abaixo do ideal, pois os dois algoritmos operam de forma independente.

Essa configuração de compatibilidade da KCL 2.x permite que sua aplicação da KCL 3.x seja executada em um modo compatível com a KCL 2.x e use o algoritmo de balanceamento de carga para a KCL 2.x até que todos os operadores da sua aplicação de consumo tenham sido atualizados para a KCL 3.x. Quando a migração for concluída, a KCL mudará automaticamente para o modo de funcionalidade completa da KCL 3.x e começará a usar o novo algoritmo de balanceamento de carga da KCL 3.x para todos os operadores em execução.

> **Important**
>
> Se você não estiver usando `ConfigsBuilder`, mas criar um objeto `LeaseManagementConfig` para definir as configurações, será preciso adicionar

mais um parâmetro chamado `applicationName` na versão 3.x ou posterior da KCL. Para obter detalhes, consulte Erro de compilação com o LeaseManagementConfig construtor. Recomendamos usar `ConfigsBuilder` para definir as configurações da KCL. `ConfigsBuilder` fornece uma maneira mais flexível e sustentável de configurar sua aplicação da KCL.

Etapa 4: siga as melhores práticas para a implementação do método shutdownRequested()

A KCL 3.x introduz um atributo chamado transferência normal de concessões para minimizar o reprocessamento de dados quando uma concessão é transferida para outro operador como parte do processo de reatribuição da concessão. Isso é obtido mediante verificação do último número de sequência processado na tabela de concessões antes da transferência da concessão. Para garantir que a transferência normal da concessão funcione corretamente, o objeto `checkpointer` tem de ser invocado dentro do método `shutdownRequested` em sua classe `RecordProcessor`. Se você invocar o objeto `checkpointer` dentro do método `shutdownRequested`, poderá implementá-lo conforme ilustrado no exemplo a seguir.

Important

- O exemplo de implementação a seguir é um requisito mínimo para uma transferência de concessão normal. Você pode estendê-la para incluir lógica adicional relacionada à verificação, se necessário. Se você estiver executando processamento assíncrono, assegure que todos os registros entregues ao downstream tenham sido processados antes de invocar a verificação.
- Embora a transferência normal da concessão reduza significativamente a probabilidade de reprocessamento de dados durante as transferências de concessão, ela não elimina totalmente essa possibilidade. Para preservar a integridade e a consistência dos dados, projete suas aplicações de consumo downstream como idempotentes. Isso significa que elas devem ser capazes de lidar com o possível processamento de registros duplicados sem efeitos adversos no sistema geral.

```
/**
 * Invoked when either Scheduler has been requested to gracefully shutdown
 * or lease ownership is being transferred gracefully so the current owner
 * gets one last chance to checkpoint.
```

```
 *
 * Checkpoints and logs the data a final time.
 *
 * @param shutdownRequestedInput Provides access to a checkpointer, allowing a record
 processor to checkpoint
 *                               before the shutdown is completed.
 */
public void shutdownRequested(ShutdownRequestedInput shutdownRequestedInput) {
    try {
       // Ensure that all delivered records are processed
       // and has been successfully flushed to the downstream before calling
       // checkpoint
       // If you are performing any asynchronous processing or flushing to
       // downstream, you must wait for its completion before invoking
       // the below checkpoint method.
        log.info("Scheduler is shutting down, checkpointing.");
        shutdownRequestedInput.checkpointer().checkpoint();
    } catch (ShutdownException | InvalidStateException e) {
        log.error("Exception while checkpointing at requested shutdown. Giving up.",
 e);
    }
}
```

Etapa 5: verifique os pré-requisitos da KCL 3.x para coletar métricas de operadores

A KCL 3.x coleta métricas de utilização da CPU, como a utilização da CPU pelos operadores, para equilibrar uniformemente a carga entre os operadores. Os operadores de aplicações de consumo podem ser executados em Amazon EC2, Amazon ECS, Amazon EKS ou AWS Fargate. A KCL 3.x pode coletar métricas de utilização da CPU dos operadores somente quando os seguintes pré-requisitos forem atendidos:

Amazon Elastic Compute Cloud(Amazon EC2)

- Seu sistema operacional deve ser Linux OS.
- Você deve habilitar IMDSv2em sua instância do EC2.

Amazon Elastic Container Service (Amazon ECS) no Amazon EC2

- Seu sistema operacional deve ser Linux OS.
- É necessário ativar o endpoint de metadados de tarefas do ECS versão 4.
- Sua versão do atendente do contêiner do Amazon ECS deve ser 1.39.0 ou posterior.

Amazon ECS em AWS Fargate

- É necessário habilitar o endpoint de metadados de tarefas do Fargate versão 4. Se você usar a versão da plataforma 1.4.0 ou posterior do Fargate, essa opção fica habilitada por padrão.
- Fargate versão da plataforma 1.4.0 ou posterior.

Amazon Elastic Kubernetes Service (Amazon EKS) no Amazon EC2

- Seu sistema operacional deve ser Linux OS.

Amazon EKS em AWS Fargate

- Plataforma 1.3.0 ou posterior do Fargate.

> **Important**
>
> Se a KCL 3.x não puder coletar as métricas de utilização da CPU dos operadores porque os pré-requisitos não foram atendidos, ela reequilibrará a carga e o nível de throughput por concessão. Esse mecanismo de rebalanceamento alternativo garante que todos os operadores obtenham níveis de throughput total semelhantes das concessões atribuídas a cada operador. Para obter mais informações, consulte Como a KCL atribui concessões aos operadores e equilibra a carga.

Etapa 6: atualizar as permissões do IAM para a KCL 3.x

É necessário adicionar as permissões a seguir para o perfil do IAM ou a política associada à aplicação de consumo da KCL 3.x. Isso inclui a atualização da política do IAM atual, usada pela aplicação da KCL. Para obter mais informações, consulte Permissões do IAM necessárias para aplicativos de consumo da KCL.

> **Important**
>
> Seus aplicativos da KCL atuais podem não ter as seguintes ações e recursos do IAM adicionados à política do IAM porque não eram necessários na KCL 2.x. Verifique se você os adicionou antes de executar a aplicação da KCL 3.x.
>
> - Ações: `UpdateTable`

  - Recursos (ARNs): `arn:aws:dynamodb:region:account:table/KCLApplicationName`
- Ações: `Query`
  - Recursos (ARNs): `arn:aws:dynamodb:region:account:table/KCLApplicationName/index/*`
- Ações: `CreateTableDescribeTable,,Scan,GetItem,PutItem,UpdateItem,DeleteItem`
  - Recursos (ARNs):`arn:aws:dynamodb:region:account:table/KCLApplicationName-WorkerMetricStats`, `arn:aws:dynamodb:region:account:table/KCLApplicationName-CoordinatorState`

  Substitua “região”, “conta” e “KCLApplicationNome” no ARNs por seu próprio Região da AWS Conta da AWS número e nome do aplicativo KCL, respectivamente. Se você usar as configurações para personalizar os nomes das tabelas de metadados criadas pela KCL, use esses nomes de tabela especificados em vez do nome da aplicação da KCL.

Etapa 7: implantar o código da KCL 3.x em seus operadores

Depois de definir a configuração necessária para a migração e concluir todas as listas de verificação de migração anteriores, você pode criar e implantar seu código para seus operadores.

> **Note**
>
> Se você ver um erro de compilação com o `LeaseManagementConfig` construtor, consulte Erro de compilação com o LeaseManagementConfig construtor para obter informações sobre solução de problemas.

Etapa 8: concluir a migração

Durante a implantação do código da KCL 3.x, a KCL continua usando o algoritmo de atribuição de concessões da KCL 2.x. Após a implantação do código da KCL 3.x em todos os seus operadores, a KCL detectará isso automaticamente e passará a adotar o novo algoritmo de atribuição de concessões com base na utilização de recursos pelos operadores. Para obter mais detalhes sobre

o novo algoritmo de atribuição de concessões, consulte Como a KCL atribui concessões aos operadores e equilibra a carga.

Durante a implantação, você pode monitorar o processo de migração com as seguintes métricas emitidas para o. CloudWatch É possível monitorar métricas na operação de `Migration`. Todas as métricas são per-KCL-application métricas e são definidas para o nível `SUMMARY` métrico. Se a estatística `Sum` da métrica `CurrentState:3xWorker` corresponder ao número total de operadores em sua aplicação da KCL, a migração para KCL 3.x foi concluída com êxito.

> **Important**
>
> A KCL leva pelo menos 10 minutos para mudar para o novo algoritmo de atribuição de concessões depois que todos os operadores estiverem prontos para executá-lo.

CloudWatch métricas para o processo de migração da KCL

Metrics	Description
`CurrentState:3xWorker`	Número de operadores da KCL migrados para a KCL 3.x e que executam o novo algoritmo de atribuição de concessões. Se a contagem `Sum` dessa métrica corresponder ao número total de seus operadores, a migração para a KCL 3.x foi concluída com êxito. • Nível de métrica: resumo • Unidades: contagem • Estatísticas: A estatística mais útil é `Sum`
`CurrentState:2xCompatibleWorker`	Número de operadores da KCL sendo executados no modo compatível com KCL 2.x durante o processo de migração. Um valor diferente de zero para essa métrica indica que a migração ainda está em andamento. • Nível de métrica: resumo • Unidades: contagem

Metrics	Description
	• Estatísticas: A estatística mais útil é `Sum`
`Fault`	Número de exceções encontradas durante o processo de migração. A maioria dessas exceções é composta por erros transitórios e a KCL 3.x tentará automaticamente concluir a migração novamente. Caso haja um valor métrico `Fault` persistente, revise seus logs do período de migração para solucionar o problema. Se o problema persistir, entre em contato Suporte. • Nível de métrica: resumo • Unidades: contagem • Estatísticas: A estatística mais útil é `Sum`
`GsiStatusReady`	Status da criação do índice secundário global (GSI) na tabela de concessões. Essa métrica indica se o GSI na tabela de concessões foi criado, pois isso é um pré-requisito para executar a KCL 3.x. O valor é 0 ou 1, com 1 indicando que a criação foi realizada. Durante um estado de reversão, essa métrica não será emitida. Depois de avançar novamente, você pode continuar a monitorar essa métrica. • Nível de métrica: resumo • Unidades: contagem • Estatísticas: A estatística mais útil é `Sum`

Metrics	Description
`workerMetricsReady`	Status da emissão de métricas de todos os operadores. As métricas indicam se todos os operadores estão emitindo métricas como, por exemplo, a utilização da CPU. O valor é 0 ou 1, com 1 indicando que todos os operadores estão emitindo métricas e estão prontos para o novo algoritmo de atribuição de concessões. Durante um estado de reversão, essa métrica não será emitida. Depois de avançar novamente, você pode continuar a monitorar essa métrica. • Nível de métrica: resumo • Unidades: contagem • Estatísticas: A estatística mais útil é `Sum`

O KCL fornece capacidade de reversão para o modo compatível com 2.x durante a migração. Depois que a migração para a KCL 3.x for concluída, recomendamos que você remova a configuração `CoordinatorConfig.clientVersionConfig` de `CLIENT_VERSION_CONFIG_COMPATIBLE_WITH_2X` se a reversão não for mais necessária. A remoção dessa configuração interrompe a emissão de métricas relacionadas à migração da aplicação da KCL.

> **Note**
>
> Recomendamos que você monitore o desempenho e a estabilidade da sua aplicação por um período durante a migração e após a conclusão da migração. Se houver um problema, você poderá fazer a reversão para que os operadores usem a funcionalidade compatível com a KCL 2.x por meio da Ferramenta de migração da KCL.

Reverter para a versão anterior da KCL

Este tópico explica as etapas para a reversão de seu consumidor para a versão anterior. Quando for preciso reverter, há um processo de duas etapas:

1. Execute a Ferramenta de Migração da KCL.
2. Reimplantar o código da versão anterior da KCL (opcional).

Etapa 1: executar a Ferramenta de Migração da KCL

Quando precisar reverter para a versão anterior da KCL, você deve executar a Ferramenta de Migração da KCL. A Ferramenta de migração da KCL executa duas tarefas importantes:

- Ela remove uma tabela de metadados chamada tabela de métricas do operador e o índice secundário global na tabela de concessões no DynamoDB. Esses dois artefatos são criados pela KCL 3.x, mas não são necessários quando você reverte para a versão anterior.
- Isso faz com que todos os trabalhadores funcionem em um modo compatível com o KCL 2.x e comecem a usar o algoritmo de balanceamento de carga usado nas versões anteriores do KCL. Se você tiver problemas com o novo algoritmo de balanceamento de carga na KCL 3.x, isso mitigará o problema imediatamente.

> **Important**
>
> A tabela de estados do coordenador no DynamoDB deve existir e não deve ser excluída durante o processo de migração, reversão e avanço.

> **Note**
>
> É importante que todos os operadores em sua aplicação de consumidor usem o mesmo algoritmo de balanceamento de carga em um determinado momento. A Ferramenta de migração da KCL garante que todos os operadores em sua aplicação de consumo da KCL 3.x mudem para o modo compatível com a KCL 2.x para que todos os operadores executem o mesmo algoritmo de balanceamento de carga durante a reversão da aplicação para a versão anterior da KCL.

Você pode baixar a Ferramenta de Migração KCL no diretório de scripts do repositório KCL GitHub. O script pode ser executado em qualquer operador ou em qualquer host que tenha as permissões necessárias para gravar na tabela de estados do coordenador, excluir a tabela de métricas do operador e atualizar a tabela de concessões. Consulte Permissões do IAM necessárias para

aplicativos de consumo da KCL para obter a permissão necessária do IAM para executar o script. É necessário executar o script uma única vez para cada aplicação da KCL. É possível executar a Ferramenta de migração da KCL com o seguinte comando:

```
python3 ./KclMigrationTool.py --region <region> --mode rollback [--
application_name <applicationName>] [--lease_table_name <leaseTableName>] [--
coordinator_state_table_name <coordinatorStateTableName>] [--worker_metrics_table_name
 <workerMetricsTableName>]
```

Parâmetros

- --region: substitua `<region>` por sua. Região da AWS
- --application_name: esse parâmetro é obrigatório se você estiver usando nomes padrão para suas tabelas de metadados do DynamoDB (tabela de concessões, tabela de estados do coordenador e tabela de métricas do operador). Se você tiver especificado nomes personalizados para essas tabelas, poderá omitir esse parâmetro. Substitua `<applicationName>` pelo nome real da aplicação da KCL. A ferramenta usa esse nome para obter os nomes de tabela padrão se os nomes personalizados não forem fornecidos.
- --lease_table_name (opcional): esse parâmetro é necessário quando você define um nome personalizado para a tabela de concessões na configuração da KCL. Se você estiver usando o nome padrão da tabela, poderá omitir esse parâmetro. Substitua `leaseTableName` pelo nome da tabela personalizada que você especificou para a tabela de concessões.
- --coordinator_state_table_name (opcional): esse parâmetro é necessário quando você define um nome personalizado para a tabela de estados do coordenador na configuração da KCL. Se você estiver usando o nome padrão da tabela, poderá omitir esse parâmetro. Substitua `<coordinatorStateTableName>` pelo nome da tabela personalizada que você especificou para a tabela de estados do coordenador.
- --worker_metrics_table_name (opcional): esse parâmetro é necessário quando você define um nome personalizado para a tabela de métricas do operador na configuração da KCL. Se você estiver usando o nome padrão da tabela, poderá omitir esse parâmetro. Substitua `<workerMetricsTableName>` pelo nome da tabela personalizada que você especificou para a tabela de métricas do operador.

Etapa 2: reimplantar o código com a versão anterior da KCL (opcional)

Depois de executar a Ferramenta de Migração da KCL para uma reversão, você verá uma destas mensagens:

- Mensagem 1: “Reversão concluída. Sua aplicação da KCL estava executando o modo compatível com KCL 2.x. Se não houver mitigação de nenhuma regressão, reverta para os binários anteriores da aplicação implantando o código com sua versão anterior da KCL.”
  - Ação necessária: Isso significa que seus trabalhadores estavam executando no modo compatível com KCL 2.x. Se o problema persistir, reimplante o código com a versão anterior da KCL para seus operadores.
- Mensagem 2: “Reversão concluída. Sua aplicação da KCL estava executando o modo de funcionalidade da KCL 3.x. A reversão para os binários anteriores da aplicação não é necessária, a menos que você não veja nenhuma mitigação do problema em 5 minutos. Se o problema persistir, reverta para os binários anteriores da aplicação implantando o código com sua versão anterior da KCL.”
  - Ação necessária: Isso significa que seus trabalhadores estavam executando no modo KCL 3.x e a Ferramenta de Migração KCL mudou todos os trabalhadores para o modo compatível com KCL 2.x. Se o problema for resolvido, você não precisa reimplantar o código com a versão anterior da KCL. Se o problema persistir, reimplante o código com a versão anterior da KCL para seus operadores.

Avançar para a KCL 3.x após uma reversão

Este tópico explica como voltar seu consumidor para a KCL 3.x após uma reversão. Quando precisar avançar, você deve passar por um processo de duas etapas:

1. Execute a Ferramenta de Migração da KCL.
2. Implantar o código com a KCL 3.x.

Etapa 1: executar a Ferramenta de Migração da KCL

Execute a Ferramenta de Migração da KCL. Ferramenta de migração da KCL com o seguinte comando para avançar para a KCL 3.x:

```
python3 ./KclMigrationTool.py --region <region> --mode rollforward [--application_name
 <applicationName>] [--coordinator_state_table_name <coordinatorStateTableName>]
```

Parâmetros

- --region: substitua `<region>` por sua. Região da AWS

- --application_name: esse parâmetro é obrigatório se você estiver usando nomes padrão para a tabela de estados do coordenador. Se você tiver especificado nomes personalizados para a tabela de estados do coordenador, poderá omitir esse parâmetro. Substitua `<applicationName>` pelo nome real da aplicação da KCL. A ferramenta usa esse nome para obter os nomes de tabela padrão se os nomes personalizados não forem fornecidos.
- --coordinator_state_table_name (opcional): esse parâmetro é necessário quando você define um nome personalizado para a tabela de estados do coordenador na configuração da KCL. Se você estiver usando o nome padrão da tabela, poderá omitir esse parâmetro. Substitua `<coordinatorStateTableName>` pelo nome da tabela personalizada que você especificou para a tabela de estados do coordenador.

Após a execução da Ferramenta de Migração no modo de avanço, o KCL cria os seguintes recursos do DynamoDB necessários para a KCL 3.x:

- Um índice secundário global na tabela de concessões
- Uma tabela de métricas do operador

Etapa 2: implantar o código com a KCL 3.x

Depois de executar a Ferramenta de Migração da KCL para um avanço, implante seu código com a KCL 3.x nos operadores. Siga Etapa 8: concluir a migração para concluir sua migração.

Práticas recomendadas para a tabela de concessões com modo de capacidade provisionada

Se a tabela de concessões da sua aplicação da KCL foi alterada para o modo de capacidade provisionada, a KCL 3.x cria um índice secundário global na tabela de concessões com o modo de cobrança provisionada e as mesmas unidades de capacidade de leitura (RCU) e de gravação (WCU) como tabela básica de concessões. Quando o índice secundário global for criado, recomendamos que você monitore o uso real no índice secundário global, no console do DynamoDB, e ajuste as unidades de capacidade, se necessário. Para obter um guia mais detalhado sobre como alternar o modo de capacidade das tabelas de metadados do DynamoDB criadas pela KCL, consulte Modo de capacidade do DynamoDB para tabelas de metadados criadas pela KCL.

> **Note**
>
> Por padrão, a KCL cria tabelas de metadados, como a tabela de concessões, a tabela de métricas do operador e a tabela de estados do coordenador, além do índice secundário global na tabela de concessões usando o modo de capacidade sob demanda. Recomendamos que você use o modo de capacidade sob demanda para ajustar automaticamente a capacidade com base nas alterações de uso.

Migrar da KCL 1.x para a KCL 3.x

Este tópico explica como migrar seu consumidor da KCL 1.x para a KCL 3.x. A KCL 1.x usa classes e interfaces diferentes em comparação com a KCL 2.x e a KCL 3.x. É necessário primeiro migrar o processador de registros, a fábrica do processador de registros e as classes de operador para o formato compatível com a KCL 2.x/3.x e depois seguir as etapas de migração da KCL 2.x para a KCL 3.x. É possível atualizar diretamente da KCL 1.x para a KCL 3.x.

- Etapa 1: migrar o processador de registros

  Siga a seção Migrar o processador de registros na página Migrar consumidores da KCL 1.x para a KCL 2.x.
- Etapa 2: migrar a fábrica do processador de discos

  Siga a seção Migrar a fábrica do processador de registros na página Migrar consumidores da KCL 1.x para a KCL 2.x.
- Etapa 3: migrar o trabalhador

  Siga a seção Migrar o operador na página Migrar consumidores da KCL 1.x para a KCL 2.x.
- Etapa 4: migrar a configuração da KCL 1.x

  Siga a seção Configurar o cliente Amazon Kinesis na página Migrar consumidores da KCL 1.x para a KCL 2.x.
- Etapa 5: verifique a remoção do tempo de inatividade e as remoções da configuração do cliente

  Siga as seções Remoção do tempo de inatividade e Remoções da configuração do cliente na página Migrar consumidores da KCL 1.x para a KCL 2.x.
- Etapa 6: Siga as step-by-step instruções no guia de migração do KCL 2.x para o KCL 3.x

Siga as instruções na página Migrar da KCL 2.x para a KCL 3.x para concluir a migração. Se precisar reverter para a versão anterior da KCL ou passar para a KCL 3.x após uma reversão, consulte Reverter para a versão anterior da KCL e Avançar para a KCL 3.x após uma reversão.

> **Important**
>
> Não use as AWS SDK para Java versões 2.27.19 a 2.27.23 com KCL 3.x. Essas versões têm um problema que causa um erro de exceção relacionado ao uso do DynamoDB da KCL. Recomendamos que você use a AWS SDK para Java versão 2.28.0 ou posterior para evitar esse problema.

Documentação da versão anterior da KCL

Os tópicos a seguir foram arquivados. Para ver a documentação atual da Kinesis Client Library, consulte Usar a Kinesis Client Library.

> **Important**
>
> As versões 1.x e 2.x da Amazon Kinesis Client Library (KCL) estão desatualizadas. O KCL 1.x chegará end-of-support em 30 de janeiro de 2026. É altamente recomendável que você migre suas aplicações da KCL que usam a versão 1.x para a versão mais recente da KCL antes de 30 de janeiro de 2026. Para encontrar a versão mais recente da KCL, consulte a página da biblioteca de cliente do Amazon Kinesis em. GitHub Para obter informações sobre as versões mais recentes da KCL, consulte Usar a Kinesis Client Library. Para ter mais informações sobre como migrar da KCL 1.x para a KCL 3.x, consulte Migrar da KCL 1.x para a KCL 3.x.

Documentação retirada

- Informações da KCL 1.x e 2.x
- Desenvolver consumidores personalizados com throughput compartilhada
- Migrar consumidores da KCL 1.x para a KCL 2.x

Informações da KCL 1.x e 2.x

> **Important**
>
> As versões 1.x e 2.x da Amazon Kinesis Client Library (KCL) estão desatualizadas. O KCL 1.x chegará end-of-support em 30 de janeiro de 2026. É altamente recomendável que você migre suas aplicações da KCL que usam a versão 1.x para a versão mais recente da KCL antes de 30 de janeiro de 2026. Para encontrar a versão mais recente da KCL, consulte a página da biblioteca de cliente do Amazon Kinesis em. GitHub Para obter informações sobre as versões mais recentes da KCL, consulte Usar a Kinesis Client Library. Para ter mais informações sobre como migrar da KCL 1.x para a KCL 3.x, consulte Migrar da KCL 1.x para a KCL 3.x.

Um dos métodos de desenvolvimento de aplicações de consumo personalizadas que podem processar dados de fluxos de dados do KDS é usar a Kinesis Client Library (KCL).

Tópicos

- Sobre a KCL (versões anteriores)
- Versões anteriores da KCL
- Conceitos da KCL (versões anteriores)
- Usar uma tabela de concessões para monitorar os fragmentos processados pela aplicação de consumo da KCL
- Processar vários fluxos de dados com a mesma aplicação de consumo da KCL 2.x para Java
- Use o KCL com o Registro do AWS Glue Esquema

> **Note**
>
> Recomenda-se o uso da versão mais recente da KCL 1.x ou da KCL 2.x, dependendo do cenário de uso. Ambas as versões da KCL, tanto 1.x como a 2.x, são atualizadas regularmente para incluir os patches de dependência e segurança e as correções de bugs mais recentes, além de novos recursos compatíveis com versões anteriores. Para obter mais informações, consulte https://github.com/awslabs/amazon-kinesis-client/releases.

Sobre a KCL (versões anteriores)

A KCL ajuda você a consumir e processar dados de um fluxo de dados do Kinesis lidando com muitas das tarefas complexas associadas à computação distribuída. Isso inclui balanceamento de carga em várias instâncias de aplicações de consumo, resposta a falhas nas instâncias de aplicações de clientes, verificação de registros processados e reação à refragmentação. A KCL cuida de todas essas subtarefas para possibilitar a concetração de esforços na escrita de uma lógica personalizada de processamento de registros.

O KCL é diferente dos Kinesis Data APIs Streams que estão disponíveis no. AWS SDKs O Kinesis APIs Data Streams ajuda você a gerenciar muitos aspectos do Kinesis Data Streams, incluindo a criação de streams, a refragmentação e a colocação e obtenção de registros. A KCL fornece uma camada de abstração em torno de todas essas subtarefas, especificamente para possibilitar a concentração na lógica de processamento de dados personalizada da aplicação de consumo. Para obter informações sobre a API do Kinesis Data Streams, consulte a Referência de APIs do Amazon Kinesis.

> **Important**
>
> A KCL é uma biblioteca Java. O suporte para linguagens diferentes de Java é fornecido usando uma interface multilíngue chamada MultiLangDaemon. Esse daemon baseado em Java é executado em segundo plano quando uma linguagem de KCL diferente de Java é utilizada. Por exemplo, se você instalar o KCL para Python e escrever seu aplicativo de consumidor inteiramente em Python, você ainda precisará do Java instalado em seu sistema por causa do. MultiLangDaemon Além disso, MultiLangDaemon tem algumas configurações padrão que talvez você precise personalizar para seu caso de uso, por exemplo, a AWS região à qual ele se conecta. Para obter mais informações sobre o MultiLangDaemon on GitHub, consulte o MultiLangDaemon projeto KCL.

A KCL atua como um intermediário entre a lógica de processamento de registros e o Kinesis Data Streams.

Versões anteriores da KCL

Atualmente, é possível usar qualquer uma destas versões compatíveis da KCL para criar aplicações de consumo personalizadas:

- KCL 1.x

Para obter mais informações, consulte Desenvolver aplicações de consumo da KCL 1.x.

- KCL 2.x

  Para obter mais informações, consulte Desenvolver aplicações de consumo da KCL 2.x.

Pode-se usar a KCL 1.x ou a KCL 2.x para criar aplicações de consumo que usam throughput compartilhada. Para obter mais informações, consulte Desenvolver aplicações de consumo personalizadas com throughput compartilhada usando a KCL.

Para criar aplicações de consumo que usam throughput dedicada (consumidores de distribuição avançada), só é possível usar a KCL 2.x. Para obter mais informações, consulte Desenvolver consumidores de distribuição avançada com throughput dedicado.

Para obter informações sobre as diferenças entre a KCL 1.x e a KCL 2.x e instruções sobre como migrar da KCL 1.x para a KCL 2.x, consulte Migrar consumidores da KCL 1.x para a KCL 2.x.

Conceitos da KCL (versões anteriores)

- Aplicação de consumo da KCL: uma aplicação personalizada que usa a KCL e é projetada para ler e processar registros de fluxos de dados.
- Instância de aplicação de consumo: as aplicações de cliente da KCL normalmente são distribuídas, com uma ou mais instâncias executadas simultaneamente para coordenar falhas e balancear dinamicamente a carga de processamento dos registros de dados.
- Operador: uma classe de alto nível que uma instância de aplicação de consumo da KCL usa para começar a processar dados.

  > **Important**
  >
  > Cada instância da aplicação de consumo da KCL tem um operador.

  O operador inicializa e supervisiona várias tarefas, incluindo a sincronização de informações de fragmentos e concessões, o monitoramento de atribuições de fragmentos e o processamento dos dados dos fragmentos. Um trabalhador fornece à KCL as informações de configuração do aplicativo consumidor, como o nome do fluxo de dados cujos registros de dados esse aplicativo consumidor KCL processará e AWS as credenciais necessárias para acessar esse fluxo de dados.

O operador também inicia a instância específica da aplicação de consumo da KCL para entregar registros de dados do fluxo de dados aos processadores de registros.

> **Important**
>
> Na KCL 1.x, essa classe é chamada de operador. Para obter mais informações (esses são os repositórios Java KCL), consulte https://github.com/awslabs/amazon-kinesis-client/blob/v1.x/src/main/java/com/amazonaws/services/kinesis/clientlibrary/lib/worker/Worker.java. Na KCL 2.x, essa classe é chamada de Agendador. A finalidade do Agendador na KCL 2.x é idêntica à finalidade do operador na KCL 1.x. Para obter mais informações sobre a classe Scheduler no KCL 2.x, consulte https://github.com/awslabs/amazon-kinesis-client/.java.blob/master/amazon-kinesis-client/src/main/java/software/amazon/kinesis/coordinator/Scheduler

- Concessão: dado que define a ligação entre um operador e um fragmento. As aplicações de cliente distribuídas da KCL usam concessões para particionar o processamento de registros de dados em uma frota de operadores. A qualquer momento, cada fragmento de registros de dados é associado a um determinado operador por uma concessão identificada pela variável leaseKey.

  Por padrão, um trabalhador pode manter um ou mais arrendamentos (sujeito ao valor da variável maxLeasesForWorker) ao mesmo tempo.

  > **Important**
  >
  > Os operadores competem para manter todas as concessões disponíveis para todos os fragmentos disponíveis em um fluxo de dados. Mas apenas um operador consegue manter uma concessão de cada vez.

  Por exemplo, se houver uma instância da aplicação de consumo A com o operador A que está processando um fluxo de dados com quatro fragmentos, o operador A poderá reter as concessões aos fragmentos 1, 2, 3 e 4 ao mesmo tempo. Mas, se houver duas instâncias de aplicações de consumo A e B com os operadores A e B, e essas instâncias estiverem processando um fluxo de dados com quatro fragmentos, o operador A e o operador B não poderão reter a concessão ao fragmento 1 ao mesmo tempo. Um operador retém a concessão a um fragmento específico até estar pronto para parar de processar os registros de dados do fragmento ou até que uma falha ocorra. Quando um operador libera a concessão, outro operador a assume e a retém.

Para obter mais informações (esses são os repositórios Java KCL), consulte https://github.com/awslabs/amazon-kinesis-client/blob/v1.x/src/main/java/com/amazonaws/services/kinesis/leases/impl/Lease.java para KCL 1.x e https://github.com/awslabs/amazon-kinesis-client/blob/master/amazon-kinesis-client/src/main/java/software/amazon/kinesis/leases/Lease.java para KCL 2.x.

- Tabela de concessões: uma tabela exclusiva do Amazon DynamoDB usada para monitorar os fragmentos em um fluxo de dados do KDS vinculados a uma concessão e sendo processados pelos operadores da aplicação de consumo da KCL. A tabela de concessões precisa permanecer sincronizada (em um operador e entre todos os operadores) com as informações mais recentes do fragmento do fluxo de dados enquanto a aplicação de consumo da KCL está em execução. Para obter mais informações, consulte Usar uma tabela de concessões para monitorar os fragmentos processados pela aplicação de consumo da KCL.
- Processador de registros: a lógica que define como a aplicação de consumo da KCL processa os dados obtidos dos fluxos de dados. Em runtime, uma instância da aplicação de consumo da KCL inicia um operador, que, por sua vez, inicia um processador de registros para cada fragmento cuja concessão retém.

Usar uma tabela de concessões para monitorar os fragmentos processados pela aplicação de consumo da KCL

Tópicos

- O que é uma tabela de concessões
- Throughput
- Como a tabela de concessões é sincronizada com fragmentos em um fluxo de dados do KDS

O que é uma tabela de concessões

Em cada aplicação do Amazon Kinesis Data Streams, a KCL usa uma tabela de concessões exclusiva (armazenada em uma tabela do Amazon DynamoDB) para monitorar os fragmentos em um fluxo de dados do KDS vinculados a uma concessão e sendo processados pelos operadores da aplicação de consumo da KCL.

> **Important**
>
> Como a KCL usa o nome da aplicação de consumo para criar o nome da tabela de concessões que a aplicação usa, cada aplicação de consumo deve ter um nome exclusivo.

É possível visualizar a tabela usando o console do Amazon DynamoDB enquanto a aplicação de consumo está em execução.

Se a tabela de concessões da aplicação de consumo da KCL não existir quando a aplicação for inicializada, um dos operadores a criará.

> **Important**
>
> Sua conta é cobrada pelos custos associados à tabela do DynamoDB, além dos custos associados ao próprio Kinesis Data Streams.

Cada linha na tabela de concessões representa um fragmento que está sendo processado pelos operadores da aplicação de consumo. Se a aplicação de consumo da KCL processar somente um fluxo de dados, a chave de hash da tabela de concessões, `leaseKey`, será o ID do fragmento. Em caso de Processar vários fluxos de dados com a mesma aplicação de consumo da KCL 2.x para Java, a estrutura da leaseKey será semelhante a `account-id:StreamName:streamCreationTimestamp:ShardId`. Por exemplo, `.111111111:multiStreamTest-1:12345:shardId-000000000336`

Além do ID do fragmento, cada linha também inclui os seguintes dados:

- checkpoint: número de sequência do ponto de verificação mais recente do fragmento. Esse valor é exclusivo entre todos os fragmentos no fluxo de dados.
- checkpointSubSequenceNúmero: ao usar o recurso de agregação da Kinesis Producer Library, essa é uma extensão do ponto de verificação que rastreia registros individuais de usuários dentro do registro do Kinesis.
- leaseCounter: usado para versionamento de concessão, para que os operadores possam detectar se a própria concessão foi assumida por outro operador.
- leaseKey: um identificador exclusivo para uma concessão. Cada concessão é específica a um fragmento no fluxo de dados e é retida por um operador por vez.
- leaseOwner: o operador que está retendo essa concessão.
- ownerSwitchesSincePonto de verificação: Quantas vezes esse contrato de arrendamento mudou de trabalhador desde a última vez que um posto de controle foi escrito.
- parentShardId: usado para garantir que o fragmento principal seja totalmente processado antes do início do processamento nos fragmentos secundários. Isso garante que os registros sejam processados na mesma ordem em que foram colocados no fluxo.

- hashrange: usado pelo `PeriodicShardSyncManager` para executar sincronizações periódicas a fim de encontrar fragmentos ausentes na tabela de concessões e criar concessões para eles, se necessário.

  > **Note**
  >
  > A partir da KCL 1.14 e da KCL 2.3, esse dado está presente na tabela de concessões de cada fragmento. Para obter mais informações sobre `PeriodicShardSyncManager` e a sincronização periódica entre concessões e fragmentos, consulte Como a tabela de concessões é sincronizada com fragmentos em um fluxo de dados do KDS.

- childshards: usado por `LeaseCleanupManager` para revisar o status de processamento do fragmento filho e decidir se o fragmento pai pode ser excluído da tabela de concessões.

  > **Note**
  >
  > A partir da KCL 1.14 e da KCL 2.3, esse dado está presente na tabela de concessões de cada fragmento.

- shardID: o ID do fragmento.

  > **Note**
  >
  > Esse dado só estará presente na tabela de concessões se você estiver Processar vários fluxos de dados com a mesma aplicação de consumo da KCL 2.x para Java. Isso só tem suporte na KCL 2.x para Java, a partir da KCL 2.3 para Java e versões posteriores.

- stream name o identificador do fluxo de dados no formato `account-id:StreamName:streamCreationTimestamp`.

  > **Note**
  >
  > Esse dado só estará presente na tabela de concessões se você estiver Processar vários fluxos de dados com a mesma aplicação de consumo da KCL 2.x para Java. Isso só tem suporte na KCL 2.x para Java, a partir da KCL 2.3 para Java e versões posteriores.

Throughput

Se sua aplicação do Amazon Kinesis Data Streams receber exceções de throughput provisionado, é necessário aumentar o throughput provisionado para a tabela do DynamoDB. A KCL cria a tabela com um throughput provisionado de 10 leituras por segundo e 10 gravações por segundo, mas isso pode não ser suficiente para a aplicação. Por exemplo, se uma aplicação do Amazon Kinesis Data Streams definir pontos de verificação ou usar operadores com frequência em um fluxo de dados composto por vários fragmentos, talvez seja necessário um throughput maior.

Para obter informações sobre o throughput provisionado no DynamoDB, consulte Modo de capacidade de leitura/gravação e Trabalhar com tabelas e dados no DynamoDB no Guia do desenvolvedor do Amazon DynamoDB.

Como a tabela de concessões é sincronizada com fragmentos em um fluxo de dados do KDS

Os operadores das aplicações de consumo da KCL usam concessões para processar fragmentos de um determinado fluxo de dados. As informações sobre qual operador usa a concessão a um fragmento em um momento determinado são armazenadas em uma tabela de concessões. A tabela de concessões precisa permanecer sincronizada com as informações mais recentes do fragmento do fluxo de dados enquanto a aplicação de consumo da KCL está em execução. A KCL sincroniza a tabela de concessões com as informações de fragmentos adquiridas do serviço Kinesis Data Streams durante a inicialização ou o reinício da aplicação de consumo e sempre que um fragmento sendo processado chega ao fim (refragmentação). Em outras palavras, os operadores ou uma aplicação de consumo da KCL são sincronizados com o fluxo de dados que estão processando durante a inicialização da aplicação e sempre que a aplicação encontra um evento de refragmentação do fluxo de dados.

Tópicos

- Sincronização na KCL 1.0 - 1.13 e na KCL 2.0 - 2.2
- Sincronização na KCL 2.x a partir da KCL 2.3 e em versões posteriores
- Sincronização na KCL 1.x a partir da KCL 1.14 e em versões posteriores

Sincronização na KCL 1.0 - 1.13 e na KCL 2.0 - 2.2

No KCL 1.0 - 1.13 e no KCL 2.0 - 2.2, durante a inicialização do aplicativo consumidor e também durante cada evento de refragmentação do fluxo de dados, o KCL sincroniza a tabela de leasing com as informações de fragmentos adquiridas do serviço Kinesis Data Streams invocando a ou a descoberta. `ListShards` `DescribeStream` APIs Em todas as versões do KCL listadas acima,

cada trabalhador de um aplicativo consumidor do KCL conclui as seguintes etapas para realizar o processo de lease/shard sincronização durante a inicialização do aplicativo consumidor e em cada evento de refragmentação do stream:

- Busca todos os fragmentos de dados do fluxo sendo processado
- Busca todas as concessões do fragmento da tabela de concessões
- Filtra cada fragmento aberto sem uma concessão na tabela de concessões
- Itera em todos os fragmentos abertos encontrados e, para cada fragmento aberto sem pai aberto:
  - Percorre a árvore hierárquica no caminho dos ancestrais para determinar se o fragmento é um descendente. Um fragmento será considerado descendente se um fragmento ancestral estiver sendo processado (a entrada de concessão do fragmento ancestral existe na tabela de concessões) ou se houver um fragmento ancestral que deve ser processado (por exemplo, a posição inicial é `TRIM_HORIZON` ou `AT_TIMESTAMP`).
  - Se o fragmento aberto for descendente, a KCL verificará sua posição inicial e criará concessões para seus pais, se necessário.

Sincronização na KCL 2.x a partir da KCL 2.3 e em versões posteriores

A partir das versões mais recentes compatíveis da KCL 2.x (KCL 2.3) e posteriores, a biblioteca oferece suporte às alterações no processo de sincronização listadas a seguir. Essas alterações de lease/shard sincronização reduzem significativamente o número de chamadas de API feitas pelos aplicativos consumidores da KCL para o serviço Kinesis Data Streams e otimizam o gerenciamento de leasing em seu aplicativo consumidor da KCL.

- Na inicialização da aplicação, se a tabela de concessões estiver vazia, a KCL utilizará a opção de filtragem da API `ListShard` (o parâmetro de solicitação `ShardFilter` opcional) para recuperar e criar concessões somente para um instantâneo dos fragmentos abertos no momento especificado pelo parâmetro `ShardFilter`. O parâmetro `ShardFilter` permite filtrar a resposta da API `ListShards`. A única propriedade obrigatória do parâmetro `ShardFilter` é `Type`. A KCL usa a propriedade de filtro `Type` e os seguintes valores válidos para identificar e retornar um instantâneo dos fragmentos abertos que podem exigir novas concessões:
  - `AT_TRIM_HORIZON`: a resposta inclui todos os fragmentos abertos em`TRIM_HORIZON`.
  - `AT_LATEST`: a resposta inclui somente os fragmentos do fluxo de dados abertos no momento.
  - `AT_TIMESTAMP`: a resposta inclui todos os fragmentos com timestamp inicial menor ou igual ao timestamp fornecido e timestamp final maior ou igual ao timestamp fornecido ou ainda abertos.

`ShardFilter` é usado ao criar uma concessão em uma tabela de concessões vazia para inicializar concessões para um instantâneo dos fragmentos especificados em `RetrievalConfig#initialPositionInStreamExtended`.

Para saber mais sobre o `ShardFilter`, consulte https://docs.aws.amazon.com/kinesis/latest/APIReference/API_ShardFilter.html.

- Em vez de todos os trabalhadores realizarem a lease/shard sincronização para manter a tabela de leasing atualizada com os fragmentos mais recentes no fluxo de dados, um único líder de trabalhadores eleito executa a sincronização de arrendamento/fragmento.
- O KCL 2.3 usa o parâmetro de `ChildShards` retorno do `GetRecords` e do `SubscribeToShard` APIs para realizar a lease/shard sincronização que acontece em `SHARD_END` para fragmentos fechados, permitindo que um trabalhador do KCL crie concessões somente para os fragmentos secundários do fragmento que ele concluiu o processamento. Para aplicativos compartilhados em todos os consumidores, essa otimização da lease/shard sincronização usa o `ChildShards` parâmetro da `GetRecords` API. Para aplicativos de consumo de taxa de transferência dedicada (fan-out aprimorado), essa otimização da lease/shard sincronização usa o `ChildShards` parâmetro da API. `SubscribeToShard` Para obter mais informações, consulte GetRecords, SubscribeToShards e ChildShard.
- Com as mudanças acima, o comportamento da KCL está passando de um modelo no qual todos os operadores obtêm informações de todos os fragmentos existentes para um modelo em que os operadores só obtêm informações dos filhos do fragmento que possui. Portanto, além da sincronização que ocorre durante a inicialização do aplicativo do consumidor e os eventos de refragmentação, a KCL agora também realiza shard/lease varreduras periódicas adicionais para identificar possíveis falhas na tabela de leasing (em outras palavras, para conhecer todos os novos fragmentos) para garantir que o intervalo de hash completo do fluxo de dados esteja sendo processado e criar concessões para eles, se necessário. `PeriodicShardSyncManager`é o componente responsável pela execução de lease/shard varreduras periódicas.

  Para obter mais informações sobre o `PeriodicShardSyncManager` KCL 2.3, consulte https://github.com/awslabs/amazon-kinesis-client/blob/master/amazon-kinesis-client/src/main/java/software/amazon/kinesis/leases/LeaseManagementConfig.java #L201 -L213.

  A KCL 2.3 tem novas opções disponíveis para configuração de `PeriodicShardSyncManager` em `LeaseManagementConfig`:

Nome	Valor padrão	Description
leasesRecoveryAuditorExecutionFrequencyMillis	120.000 (2 minutos)	Frequência (em milissegundos) do trabalho do auditor para verificar concessões parciais na tabela de concessões. Se detectar alguma falha nas concessões de um fluxo, o auditor acionará a sincronização de fragmentos com base em `leasesRecoveryAuditorInconsistencyConfidenceThreshold` .

Nome	Valor padrão	Description
leasesRecoveryAuditorInconsistencyConfidenceThreshold	3	Limite de confiança no trabalho periódico do auditor para determinar se as concessões de um fluxo de dados na tabela de concessões são inconsistentes. Se encontrar consecutivamente o mesmo conjunto de inconsistências em um fluxo de dados pelo número de vezes definido, o auditor acionará uma sincronização de fragmentos.

Agora, novas CloudWatch métricas também são emitidas para monitorar a integridade do`PeriodicShardSyncManager`. Para obter mais informações, consulte PeriodicShardSyncManager.

- Inclui uma otimização de `HierarchicalShardSyncer` para criar apenas concessões em uma camada de fragmentos.

Sincronização na KCL 1.x a partir da KCL 1.14 e em versões posteriores

A partir das versões mais recentes compatíveis da KCL 1.x (KCL 1.14) e posteriores, a biblioteca oferece suporte às alterações no processo de sincronização listadas a seguir. Essas alterações de lease/shard sincronização reduzem significativamente o número de chamadas de API feitas pelos aplicativos consumidores da KCL para o serviço Kinesis Data Streams e otimizam o gerenciamento de leasing em seu aplicativo consumidor da KCL.

- Na inicialização da aplicação, se a tabela de concessões estiver vazia, a KCL utilizará a opção de filtragem da API `ListShard` (o parâmetro de solicitação `ShardFilter` opcional) para recuperar e criar concessões somente para um instantâneo dos fragmentos abertos no momento especificado pelo parâmetro `ShardFilter`. O parâmetro `ShardFilter` permite filtrar a resposta da API `ListShards`. A única propriedade obrigatória do parâmetro `ShardFilter` é `Type`. A KCL usa a propriedade de filtro `Type` e os seguintes valores válidos para identificar e retornar um instantâneo dos fragmentos abertos que podem exigir novas concessões:
  - `AT_TRIM_HORIZON`: a resposta inclui todos os fragmentos abertos em`TRIM_HORIZON`.
  - `AT_LATEST`: a resposta inclui somente os fragmentos do fluxo de dados abertos no momento.
  - `AT_TIMESTAMP`: a resposta inclui todos os fragmentos com timestamp inicial menor ou igual ao timestamp fornecido e timestamp final maior ou igual ao timestamp fornecido ou ainda abertos.

  `ShardFilter` é usado ao criar uma concessão em uma tabela de concessões vazia para inicializar concessões para um instantâneo dos fragmentos especificados em `KinesisClientLibConfiguration#initialPositionInStreamExtended`.

  Para saber mais sobre o `ShardFilter`, consulte https://docs.aws.amazon.com/kinesis/latest/APIReference/API_ShardFilter.html.
- Em vez de todos os trabalhadores realizarem a lease/shard sincronização para manter a tabela de leasing atualizada com os fragmentos mais recentes no fluxo de dados, um único líder de trabalhadores eleito executa a sincronização de arrendamento/fragmento.
- O KCL 1.14 usa o parâmetro de `ChildShards` retorno do `GetRecords` e do `SubscribeToShard` APIs para realizar a lease/shard sincronização que acontece em `SHARD_END` para fragmentos fechados, permitindo que um trabalhador do KCL crie concessões somente para os fragmentos secundários do fragmento que ele concluiu o processamento. Para obter mais informações, consulte GetRecords e ChildShard.
- Com as mudanças acima, o comportamento da KCL está passando de um modelo no qual todos os operadores obtêm informações de todos os fragmentos existentes para um modelo em que os operadores só obtêm informações dos filhos do fragmento que possui. Portanto, além da

sincronização que ocorre durante a inicialização do aplicativo do consumidor e os eventos de refragmentação, a KCL agora também realiza shard/lease varreduras periódicas adicionais para identificar possíveis falhas na tabela de leasing (em outras palavras, para conhecer todos os novos fragmentos) para garantir que o intervalo de hash completo do fluxo de dados esteja sendo processado e criar concessões para eles, se necessário. `PeriodicShardSyncManager`é o componente responsável pela execução de lease/shard varreduras periódicas.

Quando `KinesisClientLibConfiguration#shardSyncStrategyType` é definido como `ShardSyncStrategyType.SHARD_END`, `PeriodicShardSync` `leasesRecoveryAuditorInconsistencyConfidenceThreshold` é usado para determinar o limite do número de varreduras consecutivas contendo lacunas na tabela de concessões após o qual é necessário impor uma sincronização de fragmentos. Quando `KinesisClientLibConfiguration#shardSyncStrategyType` é definido como `ShardSyncStrategyType.PERIODIC`, `leasesRecoveryAuditorInconsistencyConfidenceThreshold` é ignorado.

Para obter mais informações sobre o `PeriodicShardSyncManager` KCL 1.14, consulte https://github.com/awslabs/amazon-kinesis-client/blob/v1.x/src/main/java/com/amazonaws/services/kinesis/clientlibrary/lib/worker/KinesisClientLibConfiguration.java #L987 -L999.

A KCL 1.14 tem uma nova opção disponível para configuração de `PeriodicShardSyncManager` em `LeaseManagementConfig`:

Nome	Valor padrão	Description
leasesRecoveryAuditorInconsistencyConfidenceThreshold	3	Limite de confiança no trabalho periódico do auditor para determinar se as concessões de um fluxo de dados na tabela de concessões são inconsistentes. Se encontrar consecutivamente o mesmo conjunto de inconsistências em um fluxo de dados pelo número de vezes definido, o auditor acionará uma sincronização de fragmentos.

Agora, novas CloudWatch métricas também são emitidas para monitorar a integridade do`PeriodicShardSyncManager`. Para obter mais informações, consulte PeriodicShardSyncManager.

- A KCL 1.14 agora também oferece suporte à limpeza adiada de concessões. As concessões são excluídas de forma assíncrona por `LeaseCleanupManager` ao chegar ao `SHARD_END` quando

um fragmento ultrapassar o período de retenção do fluxo de dados ou quando for fechado por uma operação de refragmentação.

Novas opções disponíveis para configuração de `LeaseCleanupManager`:

Nome	Valor padrão	Description
leaseCleanupIntervalMillis	1 minuto	Intervalo de execução do thread de limpeza de concessões.
completedLeaseCleanupIntervalMillis	5 minutos	Intervalo de verificação de conclusão da concessão.
garbageLeaseCleanupIntervalMillis	30 minutos	Intervalo de verificação do estado de lixo de uma concessão (ou seja, reduzida após o período de retenção do fluxo de dados).

- Inclui uma otimização de `KinesisShardSyncer` para criar apenas concessões em uma camada de fragmentos.

Processar vários fluxos de dados com a mesma aplicação de consumo da KCL 2.x para Java

Esta seção descreve as seguintes alterações na KCL 2.x para Java que permitem criar aplicações de consumo da KCL que podem processar mais de um fluxo de dados ao mesmo tempo.

> **Important**
>
> O processamento multifluxo só tem suporte na KCL 2.x para Java, a partir da KCL 2.3 para Java e versões posteriores.
> O processamento multifluxo NÃO tem suporte em nenhuma outra linguagem na qual a KCL 2.x possa ser implementada.
> O processamento multifluxo NÃO tem suporte em nenhuma versão da KCL 1.x.

- MultistreamTracker interface

  Para criar um aplicativo de consumidor que possa processar vários fluxos ao mesmo tempo, você deve implementar uma nova interface chamada MultistreamTracker. Essa interface inclui o método `streamConfigList`, que retorna a lista de fluxos de dados, e suas configurações, a serem processados pela aplicação de consumo da KCL. Observe que os fluxos de dados sendo processados podem ser alterados durante o runtime da aplicação de consumo. `streamConfigList` é chamado periodicamente pela KCL para obter informações das mudanças nos fluxos de dados a serem processados.

  O `streamConfigList` método preenche a StreamConfiglista.

```
package software.amazon.kinesis.common;

import lombok.Data;
import lombok.experimental.Accessors;

@Data
@Accessors(fluent = true)
public class StreamConfig {
    private final StreamIdentifier streamIdentifier;
    private final InitialPositionInStreamExtended initialPositionInStreamExtended;
    private String consumerArn;
}
```

  Observe que os campos `StreamIdentifier` e `InitialPositionInStreamExtended` são obrigatórios, enquanto `consumerArn` é opcional. Só é necessário fornecer `consumerArn` se

a KCL 2.x estiver sendo usada para implementar uma aplicação de consumo de distribuição avançada.

Para obter mais informações sobre`StreamIdentifier`, consulte https://github.com/awslabs/amazon-kinesis-client/blob/v2.5.8/amazon-kinesis-client/src/main/java/software/amazon/kinesis/common/StreamIdentifier.java #L129. Para criar um `StreamIdentifier`, recomenda-se a criação de uma instância multifluxo a partir do `streamArn` e do `streamCreationEpoch` que esteja disponível na v2.5.0 e versões posteriores. Na KCL v2.3 e v2.4, que não oferecem suporte ao `streamArm`, crie uma instância multifluxo usando o formato `account-id:StreamName:streamCreationTimestamp`. Esse formato será descontinuado e não terá mais suporte a partir da próxima versão principal.

`MultistreamTracker` também inclui uma estratégia para excluir concessões de fluxos antigos na tabela de concessões (`formerStreamsLeasesDeletionStrategy`). Observe que a estratégia NÃO PODE ser alterada durante o runtime da aplicação de consumo. Para obter mais informações, consulte https://github.com/awslabs/amazon-kinesis-clientb/amazon-kinesis-client/src/main/java/software/amazon/kinesis/processor/FormerStreamsLeasesDeletionStrategy/blob/0c5042dadf794fe988438436252a5a8fe70b6b0 .java

- ConfigsBuilderé uma classe de todo o aplicativo que você pode usar para especificar todas as configurações do KCL 2.x a serem usadas ao criar seu aplicativo consumidor KCL. `ConfigsBuilder`a classe agora tem suporte para a `MultistreamTracker` interface. Você pode inicializar ConfigsBuilder com o nome do único fluxo de dados do qual consumir registros:

```
/**
    * Constructor to initialize ConfigsBuilder with StreamName
    * @param streamName
    * @param applicationName
    * @param kinesisClient
    * @param dynamoDBClient
    * @param cloudWatchClient
    * @param workerIdentifier
    * @param shardRecordProcessorFactory
    */
   public ConfigsBuilder(@NonNull String streamName, @NonNull String
applicationName,
           @NonNull KinesisAsyncClient kinesisClient, @NonNull DynamoDbAsyncClient
dynamoDBClient,
           @NonNull CloudWatchAsyncClient cloudWatchClient, @NonNull String
workerIdentifier,
```

```
            @NonNull ShardRecordProcessorFactory shardRecordProcessorFactory) {
        this.appStreamTracker = Either.right(streamName);
        this.applicationName = applicationName;
        this.kinesisClient = kinesisClient;
        this.dynamoDBClient = dynamoDBClient;
        this.cloudWatchClient = cloudWatchClient;
        this.workerIdentifier = workerIdentifier;
        this.shardRecordProcessorFactory = shardRecordProcessorFactory;
    }
```

Ou você pode inicializar ConfigsBuilder com `MultiStreamTracker` se quiser implementar um aplicativo consumidor KCL que processe vários fluxos ao mesmo tempo.

```
* Constructor to initialize ConfigsBuilder with MultiStreamTracker
     * @param multiStreamTracker
     * @param applicationName
     * @param kinesisClient
     * @param dynamoDBClient
     * @param cloudWatchClient
     * @param workerIdentifier
     * @param shardRecordProcessorFactory
     */
    public ConfigsBuilder(@NonNull MultiStreamTracker multiStreamTracker, @NonNull
 String applicationName,
            @NonNull KinesisAsyncClient kinesisClient, @NonNull DynamoDbAsyncClient
 dynamoDBClient,
            @NonNull CloudWatchAsyncClient cloudWatchClient, @NonNull String
 workerIdentifier,
            @NonNull ShardRecordProcessorFactory shardRecordProcessorFactory) {
        this.appStreamTracker = Either.left(multiStreamTracker);
        this.applicationName = applicationName;
        this.kinesisClient = kinesisClient;
        this.dynamoDBClient = dynamoDBClient;
        this.cloudWatchClient = cloudWatchClient;
        this.workerIdentifier = workerIdentifier;
        this.shardRecordProcessorFactory = shardRecordProcessorFactory;
    }
```

- Com o suporte multifluxo implementado na aplicação de consumo da KCL, cada linha da tabela de concessões da aplicação contém o ID do fragmento e o nome do fluxo que a aplicação processa.

- Quando o suporte multifluxo para sua aplicação de consumo da KCL é implementado, leaseKey assume a estrutura `account-id:StreamName:streamCreationTimestamp:ShardId`. Por exemplo, .`111111111:multiStreamTest-1:12345:shardId-000000000336`

> **Important**
>
> Quando sua aplicação de consumo da KCL está configurada para processar somente um fluxo de dados, leaseKey (a chave de hash da tabela de concessões) é o ID do fragmento. Ao reconfigurar a aplicação de consumo da KCL existente para processar vários fluxos de dados, a tabela de concessões será quebrada, pois no suporte multifluxo, a estrutura leaseKey deve ser a `account-id:StreamName:StreamCreationTimestamp:ShardId`.

Use o KCL com o Registro do AWS Glue Esquema

Você pode integrar seus streams de dados do Kinesis com o AWS Glue Schema Registry. O registro de esquemas do AWS Glue permite detectar, controlar e evoluir esquemas centralmente, ao mesmo tempo que garante que os dados produzidos sejam validados continuamente por um esquema registrado. O esquema define a estrutura e o formato de um registro de dados. Um esquema é uma especificação versionada para publicação, consumo ou datastore confiáveis. O AWS Glue Schema Registry permite que você melhore a qualidade end-to-end dos dados e a governança de dados em seus aplicativos de streaming. Para obter mais informações, consulte Registro de esquemas do AWS Glue. Uma das formas de configurar essa integração é usando a KCL em Java.

> **Important**
>
> Atualmente, a integração do Kinesis Data AWS Glue Streams e do Schema Registry só é compatível com os streams de dados do Kinesis que usam consumidores do KCL 2.3 implementados em Java. Não há suporte para múltiplas linguagens. Os clientes da KCL 1.0 não são compatíveis. Os clientes da KCL 2.x anteriores à KCL 2.3 não são compatíveis.

Para obter instruções detalhadas sobre como configurar a integração do Kinesis Data Streams com o Schema Registry usando o KCL, consulte a seção “Interagindo com dados KPL/KCL usando as bibliotecas” em Caso de uso: integração do Amazon Kinesis Data Streams com o Glue Schema Registry. AWS

Desenvolver consumidores personalizados com throughput compartilhada

> **Important**
>
> As versões 1.x e 2.x da Amazon Kinesis Client Library (KCL) estão desatualizadas. O KCL 1.x chegará end-of-support em 30 de janeiro de 2026. É altamente recomendável que você migre suas aplicações da KCL que usam a versão 1.x para a versão mais recente da KCL antes de 30 de janeiro de 2026. Para encontrar a versão mais recente da KCL, consulte a página da biblioteca de cliente do Amazon Kinesis em. GitHub Para obter informações sobre as versões mais recentes da KCL, consulte Usar a Kinesis Client Library. Para ter mais informações sobre como migrar da KCL 1.x para a KCL 3.x, consulte Migrar da KCL 1.x para a KCL 3.x.

Caso não seja necessário um throughput específico ao receber dados do Kinesis Data Streams, nem atrasos de propagação de leitura de até 200 ms, pode-se criar aplicações de consumo seguindo as etapas descritas nos tópicos a seguir. É possível usar a Kinesis Client Library (KCL) ou o AWS SDK para Java.

Tópicos

- Desenvolver aplicações de consumo personalizadas com throughput compartilhada usando a KCL

Para obter informações sobre a criação de consumidores que podem receber registros de fluxos de dados do Kinesis com throughput dedicada, consulte Desenvolver consumidores de distribuição avançada com throughput dedicado.

Desenvolver aplicações de consumo personalizadas com throughput compartilhada usando a KCL

> **Important**
>
> As versões 1.x e 2.x da Amazon Kinesis Client Library (KCL) estão desatualizadas. O KCL 1.x chegará end-of-support em 30 de janeiro de 2026. É altamente recomendável que você migre suas aplicações da KCL que usam a versão 1.x para a versão mais recente da KCL antes de 30 de janeiro de 2026. Para encontrar a versão mais recente da KCL, consulte a página da biblioteca de cliente do Amazon Kinesis em. GitHub Para obter informações sobre as versões mais recentes da KCL, consulte Usar a Kinesis Client Library. Para ter mais

informações sobre como migrar da KCL 1.x para a KCL 3.x, consulte Migrar da KCL 1.x para a KCL 3.x.

Um dos métodos de desenvolvimento de aplicações de consumo personalizadas com throughput compartilhada envolve o uso da Kinesis Client Library (KCL).

Escolha um dos tópicos a seguir para a versão KCL que esteja sendo usada.

Tópicos

- Desenvolver aplicações de consumo da KCL 1.x
- Desenvolver aplicações de consumo da KCL 2.x

Desenvolver aplicações de consumo da KCL 1.x

Important

As versões 1.x e 2.x da Amazon Kinesis Client Library (KCL) estão desatualizadas. O KCL 1.x chegará end-of-support em 30 de janeiro de 2026. É altamente recomendável que você migre suas aplicações da KCL que usam a versão 1.x para a versão mais recente da KCL antes de 30 de janeiro de 2026. Para encontrar a versão mais recente da KCL, consulte a página da biblioteca de cliente do Amazon Kinesis em. GitHub Para obter informações sobre as versões mais recentes da KCL, consulte Usar a Kinesis Client Library. Para ter mais informações sobre como migrar da KCL 1.x para a KCL 3.x, consulte Migrar da KCL 1.x para a KCL 3.x.

É possível criar uma aplicação de consumo para o Amazon Kinesis Data Streams usando a Kinesis Client Library (KCL).

Para obter mais informações sobre o KCL, consulte Sobre a KCL (versões anteriores).

Escolha um dos seguintes tópicos, dependendo do que deseja usar.

Conteúdo

- Desenvolver uma aplicação de consumo da Kinesis Client Library em Java
- Desenvolver uma aplicação de consumo da Kinesis Client Library em Node.js
- Desenvolver uma aplicação de consumo da Kinesis Client Library em .NET

- Desenvolver uma aplicação de consumo da Kinesis Client Library em Python
- Desenvolver uma aplicação de consumo da Kinesis Client Library em Ruby

Desenvolver uma aplicação de consumo da Kinesis Client Library em Java

> **Important**
>
> As versões 1.x e 2.x da Amazon Kinesis Client Library (KCL) estão desatualizadas. O KCL 1.x chegará end-of-support em 30 de janeiro de 2026. É altamente recomendável que você migre suas aplicações da KCL que usam a versão 1.x para a versão mais recente da KCL antes de 30 de janeiro de 2026. Para encontrar a versão mais recente da KCL, consulte a página da biblioteca de cliente do Amazon Kinesis em. GitHub Para obter informações sobre as versões mais recentes da KCL, consulte Usar a Kinesis Client Library. Para ter mais informações sobre como migrar da KCL 1.x para a KCL 3.x, consulte Migrar da KCL 1.x para a KCL 3.x.

É possível usar a Kinesis Client Library (KCL) para criar aplicações que processam dados dos fluxos de dados do Kinesis. A Kinesis Client Library está disponível em várias linguagens. Este tópico discute Java. Para ver a referência de Javadoc, consulte o tópico AWS Javadoc para Classe. AmazonKinesisClient

Para baixar o Java KCL de GitHub, acesse a Kinesis Client Library (Java). Para localizar a KCL Java no Apache Maven, acesse a página de resultados da pesquisa de KCL. Para baixar o código de amostra para um aplicativo consumidor Java KCL em GitHub, acesse a página do projeto de amostra KCL for Java em. GitHub

O aplicativo de exemplo usa Apache Commons Logging. É possível alterar a configuração do registro em log no método `configure` estático definido no arquivo `AmazonKinesisApplicationSample.java`. Para obter mais informações sobre como usar o Apache Commons Logging com aplicativos Log4j e AWS Java, consulte Logging with Log4j no Guia do desenvolvedor.AWS SDK para Java

É necessário concluir as seguintes tarefas ao implementar uma aplicação de consumo da KCL em Java:

Tarefas

- Implemente os métodos IRecord do processador

- Implemente uma fábrica de classes para a interface IRecord do processador
- Criar um operador
- Modificar as propriedades de configuração
- Migrar para a versão 2 da interface do processador de registros

Implemente os métodos IRecord do processador

Atualmente, a KCL oferece suporte a duas versões da interface do `IRecordProcessor`: a interface original está disponível com a primeira versão da KCL e a versão 2 está disponível desde a versão 1.5.0. As duas interfaces são totalmente compatíveis. A escolha depende dos requisitos de cenário específicos. Consulte os Javadocs criados localmente ou o código-fonte para ver todas as diferenças. As seções a seguir descrevem a implementação mínima para os conceitos básicos.

IRecordVersões do processador

- Interface original (versão 1)
- Interface atualizada (versão 2)

Interface original (versão 1)

A interface `IRecordProcessor` original (`package com.amazonaws.services.kinesis.clientlibrary.interfaces`) expõe os seguintes métodos de processador de registros que o consumidor precisa implementar. O exemplo fornece implementações que podem ser usadas como ponto de partida (consulte `AmazonKinesisApplicationSampleRecordProcessor.java`).

```
public void initialize(String shardId)
public void processRecords(List<Record> records, IRecordProcessorCheckpointer
 checkpointer)
public void shutdown(IRecordProcessorCheckpointer checkpointer, ShutdownReason reason)
```

inicializar

A KCL chama o método `initialize` quando o processador de registros é instanciado, passando um ID de fragmento específico como um parâmetro. Esse processador de registros processa apenas esse fragmento e, normalmente, o inverso também é verdadeiro (esse fragmento é processado somente por esse processador de registro). No entanto, a aplicação de consumo deve considerar a possibilidade de que um registro de dados pode ser processado mais de uma vez. A semântica

do Kinesis Data Streams é do tipo pelo menos uma vez, o que significa que cada registro de dados de um fragmento é processado pelo menos uma vez por um operador na aplicação de consumo. Para obter mais informações sobre casos em que um fragmento específico pode ser processado por mais de um operador, consulte Use refragmentação, escalonamento e processamento paralelo para alterar o número de fragmentos.

```
public void initialize(String shardId)
```

processRecords

A KCL chama o método `processRecords` passando uma lista de registros de dados do fragmento especificado pelo método `initialize(shardId)`. O processador de registros processa os dados nesses registros de acordo com a semântica da aplicação de consumo. Por exemplo, o operador pode executar uma transformação nos dados e, em seguida, armazenar o resultado em um bucket do Amazon Simple Storage Service (Amazon S3).

```
public void processRecords(List<Record> records, IRecordProcessorCheckpointer
 checkpointer)
```

Além dos dados em si, o registro também contém um número de sequência e uma chave de partição. O operador pode usar esses valores ao processar os dados. Por exemplo, o operador pode escolher o bucket do S3 no qual armazenar os dados com base no valor da chave de partição. A classe `Record` expõe os seguintes métodos que oferecem acesso aos dados do registro, número de sequência e chave de partição.

```
record.getData()
record.getSequenceNumber()
record.getPartitionKey()
```

No exemplo, o método privado `processRecordsWithRetries` tem código que mostra como um operador pode acessar os dados do registro, o número de sequência e a chave de partição.

O Kinesis Data Streams requer que o processador de registros rastreie os registros que já foram processados em um fragmento. A KCL faz esse rastreamento passando um checkpointer (`IRecordProcessorCheckpointer`) para o `processRecords`. O processador de registros chama o método `checkpoint` nesta interface para informar a KCL sobre o progresso do processamento dos registros no fragmento. Se o operador falhar, a KCL usará essas informações para reiniciar o processamento do fragmento no último registro processado conhecido.

Em uma operação de divisão ou mesclagem, a KCL só começará a processar os novos fragmentos quando os processadores dos fragmentos originais chamarem `checkpoint` para indicar que o processamento dos fragmentos originais foi concluído.

Se nenhum parâmetro for fornecido, a KCL presumirá que a chamada para `checkpoint` significa que todos os registros foram processados até o último registro passado para o processador de registros. Portanto, o processador de registros deve chamar `checkpoint` somente após ter processado todos os registros na lista que foi passada a ele. Os processadores de registros não precisam chamar `checkpoint` em cada chamada para `processRecords`. Um processador pode, por exemplo, chamar `checkpoint` a cada terceira chamada para `processRecords`. É possível, opcionalmente, especificar o número de sequência exato de um registro como um parâmetro para `checkpoint`. Nesse caso, a KCL presume que todos os registros foram processados somente até o registro especificado.

No exemplo, o método privado `checkpoint` mostra como chamar `IRecordProcessorCheckpointer.checkpoint` usando a lógica de novas tentativas e o tratamento de exceções apropriados.

A KCL depende do `processRecords` para lidar com qualquer exceção ocorrida no processamento dos registros de dados. Se ocorrer uma exceção em `processRecords`, a KCL ignorará os registros de dados passados antes da exceção. Ou seja, esses registros não serão reenviados para o processador de registros que lançou a exceção ou para qualquer outro processador de registros na aplicação de consumo.

shutdown

A KCL chama o método `shutdown` quando o processamento termina (o motivo do desligamento é `TERMINATE`) ou quando o operador não está mais respondendo (o motivo do desligamento é `ZOMBIE`).

```
public void shutdown(IRecordProcessorCheckpointer checkpointer, ShutdownReason reason)
```

O processamento termina quando o processador de registros não recebe mais registros do fragmento porque ele foi dividido ou intercalado, ou o fluxo foi excluído.

A KCL também passa uma interface do `IRecordProcessorCheckpointer` para `shutdown`. Se o motivo do desligamento é `TERMINATE`, o processador de registros deve terminar o processamento de todos os registros de dados e, em seguida, chamar o método `checkpoint` nesta interface.

Interface atualizada (versão 2)

A interface `IRecordProcessor` atualizada (`package com.amazonaws.services.kinesis.clientlibrary.interfaces.v2`) expõe os seguintes métodos de processador de registros que o consumidor precisa implementar:

```
void initialize(InitializationInput initializationInput)
void processRecords(ProcessRecordsInput processRecordsInput)
void shutdown(ShutdownInput shutdownInput)
```

Todos os argumentos da versão original da interface podem ser acessados por meio de métodos get nos objetos de contêiner. Por exemplo, para recuperar a lista de registros em `processRecords()`, pode-se usar `processRecordsInput.getRecords()`.

Além das entradas fornecidas pela interface original, estas novas entradas estão disponíveis a partir da versão 2 da interface (KCL 1.5.0 e posterior):

número de sequência inicial

: No objeto `InitializationInput` passado para a operação `initialize()`, o número de sequência inicial a partir do qual os registros seriam fornecidos à instância do processador de registros. Esse é o número de sequência que foi verificado pela última vez pela instância do processador de registros que processou anteriormente o mesmo fragmento. Isso será fornecido no caso de o aplicativo precisar de informações.

número de sequência do ponto de verificação pendente

: No objeto `InitializationInput` passado para a operação `initialize()`, o número de sequência de verificação pendente (se houver) que não pôde ser confirmado antes que a instância do processador de registros anterior parasse.

Implemente uma fábrica de classes para a interface IRecord do processador

Também será necessário implementar uma fábrica para a classe que implementa os métodos do processador de registros. Quando a aplicação de consumo instancia o operador, ela passa uma referência a essa fábrica.

O exemplo implementa a classe de fábrica no arquivo `AmazonKinesisApplicationSampleRecordProcessorFactory.java` usando a interface de processador de registros original. Para que a fábrica da classe

crie a versão 2 dos processadores de registros, deve-se usar o nome do pacote `com.amazonaws.services.kinesis.clientlibrary.interfaces.v2`.

```
public class SampleRecordProcessorFactory implements IRecordProcessorFactory {
    /**
    * Constructor.
    */
    public SampleRecordProcessorFactory() {
        super();
    }
    /**
    * {@inheritDoc}
    */
    @Override
    public IRecordProcessor createProcessor() {
        return new SampleRecordProcessor();
    }
}
```

Criar um operador

Como discutido em Implemente os métodos IRecord do processador, há duas versões da interface do processador de registros da KCL para escolha, o que afeta a forma de criar um operador. A interface do processador de registros original usa a seguinte estrutura de código para criar um operador:

```
final KinesisClientLibConfiguration config = new KinesisClientLibConfiguration(...)
final IRecordProcessorFactory recordProcessorFactory = new RecordProcessorFactory();
final Worker worker = new Worker(recordProcessorFactory, config);
```

Com a versão 2 da interface do processador de registros, é possível usar `Worker.Builder` para criar um operador sem a necessidade de se preocupar com qual construtor usar e a ordem dos argumentos. A interface do processador de registros atualizada usa a seguinte estrutura de código para criar um operador:

```
final KinesisClientLibConfiguration config = new KinesisClientLibConfiguration(...)
final IRecordProcessorFactory recordProcessorFactory = new RecordProcessorFactory();
final Worker worker = new Worker.Builder()
    .recordProcessorFactory(recordProcessorFactory)
    .config(config)
```

```
    .build();
```

Modificar as propriedades de configuração

O exemplo fornece valores padrão para propriedades de configuração. Esses dados de configuração para o operador são então consolidados em um objeto `KinesisClientLibConfiguration`. Esse objeto e uma referência à fábrica de classe para `IRecordProcessor` são passados na chamada que instancia o operador. É possível substituir qualquer uma dessas propriedades por seus próprios valores usando um arquivo de propriedades do Java (consulte `AmazonKinesisApplicationSample.java`).

Nome da aplicação

A KCL exige um nome de aplicação exclusivo entre as aplicações e as tabelas do Amazon DynamoDB na mesma região. Ela usa o valor de configuração de nome de aplicativo das seguintes formas:

- Presume-se que todos os operadores associados com esse nome de aplicativo estejam trabalhando juntos no mesmo fluxo. Esses operadores podem ser distribuídos em várias instâncias. Ao executar uma instância adicional do mesmo código da aplicação, mas com um nome diferente, a KCL tratará a segunda instância como uma aplicação totalmente independente operando no mesmo fluxo.
- A KCL cria uma tabela do DynamoDB com o nome da aplicação e usa essa tabela para manter informações de estado (como pontos de verificação e mapeamento de operador-fragmento) da aplicação. Cada aplicação tem sua própria tabela do DynamoDB. Para obter mais informações, consulte Usar uma tabela de concessões para monitorar os fragmentos processados pela aplicação de consumo da KCL.

Configurar credenciais

Você deve disponibilizar suas AWS credenciais para um dos provedores de credenciais na cadeia de provedores de credenciais padrão. Por exemplo, ao executar a aplicação de consumo em uma instância do EC2, recomenda-se que a instância seja executada com um perfil do IAM. As credenciais da AWS que refletem as permissões associadas a esse perfil do IAM são disponibilizadas às aplicações na instância por meio dos metadados da instância. Essa é a maneira mais segura de gerenciar credenciais para uma aplicação de consumo em execução em uma instância do EC2.

A aplicação de exemplo primeiro tenta recuperar as credenciais do IAM nos metadados da instância:

```
credentialsProvider = new InstanceProfileCredentialsProvider();
```

Se a aplicação de exemplo não consegue obter credenciais dos metadados da instância, ele tenta recuperar as credenciais de um arquivo de propriedades:

```
credentialsProvider = new ClasspathPropertiesFileCredentialsProvider();
```

Para obter mais informações sobre os metadados da instância, consulte Metadados da instância no Guia do usuário do Amazon EC2.

Usar o ID do operador para várias instâncias

O código de inicialização de exemplo cria um ID para o operador, `workerId`, usando o nome do computador local e anexando um identificador exclusivo globalmente, conforme mostrado no seguinte trecho de código. Essa abordagem é compatível com o cenário de várias instâncias da aplicação de consumo em execução em um único computador.

```
String workerId = InetAddress.getLocalHost().getCanonicalHostName() + ":" +
 UUID.randomUUID();
```

Migrar para a versão 2 da interface do processador de registros

Para migrar o código que usa a interface original, além das etapas descritas anteriormente, as seguintes etapas serão necessárias:

1. Altere a classe do processador de registros para importar a versão 2 da interface do processador de registros:

   ```
   import com.amazonaws.services.kinesis.clientlibrary.interfaces.v2.IRecordProcessor;
   ```

2. Altere as referências para as entradas para usar métodos `get` nos objetos de contêiner. Por exemplo, na operação `shutdown()`, altere "`checkpointer`" para "`shutdownInput.getCheckpointer()`".
3. Altere a classe da fábrica do processador de registros para importar a versão 2 da interface da fábrica do processador de registros:

   ```
   import
    com.amazonaws.services.kinesis.clientlibrary.interfaces.v2.IRecordProcessorFactory;
   ```

4. Altere a construção do operador para usar `Worker.Builder`. Por exemplo:

```
final Worker worker = new Worker.Builder()
    .recordProcessorFactory(recordProcessorFactory)
    .config(config)
    .build();
```

Desenvolver uma aplicação de consumo da Kinesis Client Library em Node.js

> **Important**
>
> As versões 1.x e 2.x da Amazon Kinesis Client Library (KCL) estão desatualizadas. O KCL 1.x chegará end-of-support em 30 de janeiro de 2026. É altamente recomendável que você migre suas aplicações da KCL que usam a versão 1.x para a versão mais recente da KCL antes de 30 de janeiro de 2026. Para encontrar a versão mais recente da KCL, consulte a página da biblioteca de cliente do Amazon Kinesis em. GitHub Para obter informações sobre as versões mais recentes da KCL, consulte Usar a Kinesis Client Library. Para ter mais informações sobre como migrar da KCL 1.x para a KCL 3.x, consulte Migrar da KCL 1.x para a KCL 3.x.

É possível usar a Kinesis Client Library (KCL) para criar aplicações que processam dados dos fluxos de dados do Kinesis. A Kinesis Client Library está disponível em várias linguagens. Este tópico discute Node.js.

A KCL é uma biblioteca Java; o suporte para linguagens diferentes de Java é fornecido usando uma interface multilíngue chamada de. MultiLangDaemon Esse daemon baseado em Java é executado em segundo plano quando uma linguagem de KCL diferente de Java é utilizada. Portanto, se você instalar o KCL para Node.js e escrever seu aplicativo de consumidor inteiramente em Node.js, ainda precisará do Java instalado em seu sistema por causa do MultiLangDaemon. Além disso, MultiLangDaemon tem algumas configurações padrão que você pode precisar personalizar para seu caso de uso, por exemplo, a AWS região à qual ele se conecta. Para obter mais informações sobre o MultiLangDaemon on GitHub, acesse a página do MultiLangDaemon projeto KCL.

Para baixar a KCL do Node.js GitHub, acesse a Biblioteca de Cliente Kinesis (Node.js).

Downloads de códigos de exemplo

Há dois exemplos de código disponíveis para KCL em Node.js:

- basic-sample

  Usado nas seções a seguir para ilustrar os conceitos básicos de criação de uma aplicação de consumo da KCL em Node.js.

- click-stream-sample

  Levemente mais avançado e usa um cenário real, para depois que houver familiaridade com o código de exemplo básico. Esse exemplo não é discutido aqui, mas há um arquivo README com mais informações.

É necessário concluir as seguintes tarefas ao implementar uma aplicação de consumo da KCL em Node.js:

Tarefas

- Implementar o processador de registros
- Modificar as propriedades de configuração

Implementar o processador de registros

A aplicação de consumo mais simples possível usando a KCL para Node.js deve implementar uma função `recordProcessor`, que, por sua vez, contém as funções `initialize`, `processRecords` e `shutdown`. O exemplo fornece uma implementação que pode ser usada como ponto de partida (consulte `sample_kcl_app.js`).

```
function recordProcessor() {
  // return an object that implements initialize, processRecords and shutdown
 functions.}
```

inicializar

A KCL chama a função `initialize` quando o processador de registros é iniciado.
Esse processador de registros processa apenas o ID do fragmento passado como `initializeInput.shardId` e, normalmente, o inverso também é verdadeiro (esse fragmento é processado somente por esse processador de registro). No entanto, a aplicação de consumo deve considerar a possibilidade de que um registro de dados pode ser processado mais de uma vez. Isso acontece porque a semântica do Kinesis Data Streams é do tipo pelo menos uma vez, o que significa que cada registro de dados de um fragmento é processado pelo menos uma vez por um operador na

aplicação de consumo. Para obter mais informações sobre casos em que um fragmento específico pode ser processado por mais de um operador, consulte Use refragmentação, escalonamento e processamento paralelo para alterar o número de fragmentos.

```
initialize: function(initializeInput, completeCallback)
```

processRecords

A KCL chama essa função com uma entrada contendo uma lista de registros de dados do fragmento especificado para a função `initialize`. O processador de registros implementado processará os dados nesses registros de acordo com a semântica da aplicação de consumo. Por exemplo, o operador pode executar uma transformação nos dados e, em seguida, armazenar o resultado em um bucket do Amazon Simple Storage Service (Amazon S3).

```
processRecords: function(processRecordsInput, completeCallback)
```

Além dos dados em si, o registro também contém um número de sequência e uma chave de partição, que o operador pode usar ao processar os dados. Por exemplo, o operador pode escolher o bucket do S3 no qual armazenar os dados com base no valor da chave de partição. O dicionário de `record` expõe os seguintes pares de chave/valor para acessar os dados do registro, o número de sequência e a chave de partição:

```
record.data
record.sequenceNumber
record.partitionKey
```

Observe que os dados são codificados em Base64.

No exemplo básico, a função `processRecords` tem código que mostra como um operador pode acessar os dados do registro, o número de sequência e a chave de partição.

O Kinesis Data Streams requer que o processador de registros rastreie os registros que já foram processados em um fragmento. A KCL faz esse rastreamento com um objeto `checkpointer` passado como `processRecordsInput.checkpointer`. O processador de registros chama a função `checkpointer.checkpoint` para informar a KCL sobre o progresso do processamento dos registros no fragmento. Se o operador falhar, a KCL usará essas informações ao reiniciar o processamento do fragmento para continuar a partir do último registro processado conhecido.

Em uma operação de divisão ou mesclagem, a KCL só começará a processar os novos fragmentos quando os processadores dos fragmentos originais chamarem `checkpoint` para indicar que o processamento dos fragmentos originais foi concluído.

Se u mnúmero de sequência não for passado para a função `checkpoint`, a KCL presumirá que a chamada para `checkpoint` significa que todos os registros foram processados até o último registro passado para o processador de registros. Portanto, o processador de registros deve chamar `checkpoint` somente após ter processado todos os registros na lista que foi passada para ele. Os processadores de registros não precisam chamar `checkpoint` em cada chamada para `processRecords`. Um processador pode, por exemplo, chamar `checkpoint` a cada terceira chamada ou algum evento externo ao seu processador de gravação, como um verification/validation serviço personalizado que você implementou.

É possível, opcionalmente, especificar o número de sequência exato de um registro como um parâmetro para `checkpoint`. Nesse caso, a KCL presume que todos os registros foram processados somente até o registro especificado.

O aplicativo de exemplo básico mostra a chamada mais simples possível para a função `checkpointer.checkpoint`. É possível adicionar outra lógica de verificação que precisar para o consumidor neste ponto da função.

shutdown

A KCL chama a função `shutdown` quando o processamento termina (`shutdownInput.reason` é `TERMINATE`) ou quando o operador não está mais respondendo (`shutdownInput.reason` é `ZOMBIE`).

```
shutdown: function(shutdownInput, completeCallback)
```

O processamento termina quando o processador de registros não recebe mais registros do fragmento porque ele foi dividido ou intercalado, ou o fluxo foi excluído.

A KCL também passa um objeto `shutdownInput.checkpointer` para `shutdown`. Se o motivo do desligamento for `TERMINATE`, é necessário verificar se o processador de registros terminou o processamento de todos os registros de dados e, em seguida, chamar a função `checkpoint` nessa interface.

Modificar as propriedades de configuração

O exemplo fornece valores padrão para as propriedades de configuração. É possível substituir qualquer uma dessas propriedades por seus próprios valores (consulte `sample.properties` no exemplo básico).

Nome da aplicação

A KCL exige uma aplicação exclusiva entre as aplicações e as tabelas do Amazon DynamoDB na mesma região. Ela usa o valor de configuração de nome de aplicativo das seguintes formas:

- Presume-se que todos os operadores associados com esse nome de aplicativo estejam trabalhando juntos no mesmo fluxo. Esses operadores podem ser distribuídos em várias instâncias. Ao executar uma instância adicional do mesmo código da aplicação, mas com um nome diferente, a KCL tratará a segunda instância como uma aplicação totalmente independente operando no mesmo fluxo.
- A KCL cria uma tabela do DynamoDB com o nome da aplicação e usa essa tabela para manter informações de estado (como pontos de verificação e mapeamento de operador-fragmento) da aplicação. Cada aplicação tem sua própria tabela do DynamoDB. Para obter mais informações, consulte Usar uma tabela de concessões para monitorar os fragmentos processados pela aplicação de consumo da KCL.

Configurar credenciais

Você deve disponibilizar suas AWS credenciais para um dos provedores de credenciais na cadeia de provedores de credenciais padrão. A propriedade `AWSCredentialsProvider` pode ser usada para definir um provedor de credenciais. O arquivo `sample.properties` precisa disponibilizar as credenciais para um dos provedores de credenciais na cadeia de provedores de credenciais padrão. Se você estiver executando seu consumidor em uma instância do Amazon EC2, recomendamos que você configure a instância com uma função do IAM. AWS as credenciais que refletem as permissões associadas a essa função do IAM são disponibilizadas aos aplicativos na instância por meio dos metadados da instância. Essa é a maneira mais segura de gerenciar credenciais para uma aplicação de consumo em execução em uma instância do EC2.

O exemplo a seguir configura a KCL para processar um fluxo de dados do Kinesis chamado `kclnodejssample` usando o processador de registros fornecido em `sample_kcl_app.js`:

```
# The Node.js executable script
executableName = node sample_kcl_app.js
```

```
# The name of an Amazon Kinesis stream to process
streamName = kclnodejssample
# Unique KCL application name
applicationName = kclnodejssample
# Use default AWS credentials provider chain
AWSCredentialsProvider = DefaultAWSCredentialsProviderChain
# Read from the beginning of the stream
initialPositionInStream = TRIM_HORIZON
```

Desenvolver uma aplicação de consumo da Kinesis Client Library em .NET

> **Important**
>
> As versões 1.x e 2.x da Amazon Kinesis Client Library (KCL) estão desatualizadas. O KCL 1.x chegará end-of-support em 30 de janeiro de 2026. É altamente recomendável que você migre suas aplicações da KCL que usam a versão 1.x para a versão mais recente da KCL antes de 30 de janeiro de 2026. Para encontrar a versão mais recente da KCL, consulte a página da biblioteca de cliente do Amazon Kinesis em. GitHub Para obter informações sobre as versões mais recentes da KCL, consulte Usar a Kinesis Client Library. Para ter mais informações sobre como migrar da KCL 1.x para a KCL 3.x, consulte Migrar da KCL 1.x para a KCL 3.x.

É possível usar a Kinesis Client Library (KCL) para criar aplicações que processam dados dos fluxos de dados do Kinesis. A Kinesis Client Library está disponível em várias linguagens. Este tópico discute .NET.

A KCL é uma biblioteca Java; o suporte para linguagens diferentes de Java é fornecido usando uma interface multilíngue chamada de. MultiLangDaemon Esse daemon baseado em Java é executado em segundo plano quando uma linguagem de KCL diferente de Java é utilizada. Portanto, se você instalar o KCL para.NET e escrever seu aplicativo de consumidor inteiramente no.NET, ainda precisará do Java instalado em seu sistema por causa do MultiLangDaemon. Além disso, MultiLangDaemon tem algumas configurações padrão que você pode precisar personalizar para seu caso de uso, por exemplo, a AWS região à qual ele se conecta. Para obter mais informações sobre o MultiLangDaemon on GitHub, acesse a página do MultiLangDaemon projeto KCL.

Para baixar o.NET KCL de GitHub, acesse Kinesis Client Library (.NET). Para baixar o código de amostra para um aplicativo de consumidor do.NET KCL, acesse a página do projeto de amostra de consumidor KCL para.NET em. GitHub

É necessário concluir as seguintes tarefas ao implementar uma aplicação de consumo da KCL em .NET:

Tarefas

- Implemente os métodos da classe IRecord Processor
- Modificar as propriedades de configuração

Implemente os métodos da classe IRecord Processor

A aplicação de consumo precisa implementar os seguintes métodos para `IRecordProcessor`. A aplicação de consumo de exemplo fornece implementações que podem ser usadas como ponto de partida (consulte a classe `SampleRecordProcessor` em `SampleConsumer/AmazonKinesisSampleConsumer.cs`).

```
public void Initialize(InitializationInput input)
public void ProcessRecords(ProcessRecordsInput input)
public void Shutdown(ShutdownInput input)
```

Inicializar

A KCL chama este método quando o processador de registros é instanciado, passando um ID de fragmento específico no parâmetro `input` (`input.ShardId`). Esse processador de registros processa apenas esse fragmento e, normalmente, o inverso também é verdadeiro (esse fragmento é processado somente por esse processador de registro). No entanto, a aplicação de consumo deve considerar a possibilidade de que um registro de dados pode ser processado mais de uma vez. Isso acontece porque a semântica do Kinesis Data Streams é do tipo pelo menos uma vez, o que significa que cada registro de dados de um fragmento é processado pelo menos uma vez por um operador na aplicação de consumo. Para obter mais informações sobre casos em que um fragmento específico pode ser processado por mais de um operador, consulte Use refragmentação, escalonamento e processamento paralelo para alterar o número de fragmentos.

```
public void Initialize(InitializationInput input)
```

ProcessRecords

A KCL chama este método passando uma lista de registros de dados no parâmetro `input` (`input.Records`) do fragmento especificado pelo método `Initialize`. O processador de

registros implementado processará os dados nesses registros de acordo com a semântica da aplicação de consumo. Por exemplo, o operador pode executar uma transformação nos dados e, em seguida, armazenar o resultado em um bucket do Amazon Simple Storage Service (Amazon S3).

```
public void ProcessRecords(ProcessRecordsInput input)
```

Além dos dados em si, o registro também contém um número de sequência e uma chave de partição. O operador pode usar esses valores ao processar os dados. Por exemplo, o operador pode escolher o bucket do S3 no qual armazenar os dados com base no valor da chave de partição. A classe `Record` expõe os seguintes itens para acessar os dados do registro, o número de sequência e a chave de partição:

```
byte[] Record.Data
string Record.SequenceNumber
string Record.PartitionKey
```

No exemplo, o método `ProcessRecordsWithRetries` tem código que mostra como um operador pode acessar os dados do registro, o número de sequência e a chave de partição.

O Kinesis Data Streams requer que o processador de registros rastreie os registros que já foram processados em um fragmento. A KCL faz esse rastreamento, passando um objeto `Checkpointer` para `ProcessRecords` (`input.Checkpointer`). O processador de registros chama o método `Checkpointer.Checkpoint` para informar a KCL sobre o progresso do processamento dos registros no fragmento. Se o operador falhar, a KCL usará essas informações para reiniciar o processamento do fragmento no último registro processado conhecido.

Em uma operação de divisão ou mesclagem, a KCL só começará a processar os novos fragmentos quando os processadores dos fragmentos originais chamarem `Checkpointer.Checkpoint` para indicar que o processamento dos fragmentos originais foi concluído.

Se nenhum parâmetro for fornecido, a KCL presumirá que a chamada para `Checkpointer.Checkpoint` significa que todos os registros foram processados até o último registro passado para o processador de registros. Portanto, o processador de registros deve chamar `Checkpointer.Checkpoint` somente após ter processado todos os registros na lista que foi passada a ele. Os processadores de registros não precisam chamar `Checkpointer.Checkpoint` em cada chamada para `ProcessRecords`. Um processador pode, por exemplo, chamar `Checkpointer.Checkpoint` a cada terceira ou quarta chamada. É possível, opcionalmente, especificar o número de sequência exato de um registro como um parâmetro para

`Checkpointer.Checkpoint`. Nesse caso, a KCL presume que os registros foram processados somente até o registro especificado.

No exemplo, o método privado `Checkpoint(Checkpointer checkpointer)` mostra como chamar o método `Checkpointer.Checkpoint` usando a lógica de novas tentativas e o tratamento de exceções apropriados.

De maneira diferente das bibliotecas KCL em outras linguagens, a KCL para .NET não lida com nenhuma exceção ocorrida no processamento dos registros de dados. As exceções não detectadas do código do usuário causam uma falha no programa.

Shutdown

A KCL chama o método `Shutdown` quando o processamento termina (o motivo do desligamento é `TERMINATE`) ou quando o operador não está mais respondendo (o valor `input.Reason` do desligamento é `ZOMBIE`).

```
public void Shutdown(ShutdownInput input)
```

O processamento termina quando o processador de registros não recebe mais registros do fragmento porque ele foi dividido ou intercalado, ou o fluxo foi excluído.

A KCL também passa um objeto `Checkpointer` para `shutdown`. Se o motivo do desligamento é `TERMINATE`, o processador de registros deve terminar o processamento de todos os registros de dados e, em seguida, chamar o método `checkpoint` nesta interface.

Modificar as propriedades de configuração

A aplicação de consumo de exemplo fornece valores padrão para as propriedades de configuração. É possível substituir qualquer uma dessas propriedades por seus próprios valores (consulte `SampleConsumer/kcl.properties`).

Nome da aplicação

A KCL exige uma aplicação exclusiva entre as aplicações e as tabelas do Amazon DynamoDB na mesma região. Ela usa o valor de configuração de nome de aplicativo das seguintes formas:

- Presume-se que todos os operadores associados com esse nome de aplicativo estejam trabalhando juntos no mesmo fluxo. Esses operadores podem ser distribuídos em várias

instâncias. Ao executar uma instância adicional do mesmo código da aplicação, mas com um nome diferente, a KCL tratará a segunda instância como uma aplicação totalmente independente operando no mesmo fluxo.

- A KCL cria uma tabela do DynamoDB com o nome da aplicação e usa essa tabela para manter informações de estado (como pontos de verificação e mapeamento de operador-fragmento) da aplicação. Cada aplicação tem sua própria tabela do DynamoDB. Para obter mais informações, consulte Usar uma tabela de concessões para monitorar os fragmentos processados pela aplicação de consumo da KCL.

Configurar credenciais

Você deve disponibilizar suas AWS credenciais para um dos provedores de credenciais na cadeia de provedores de credenciais padrão. A propriedade `AWSCredentialsProvider` pode ser usada para definir um provedor de credenciais. As sample.properties precisam disponibilizar as credenciais para um dos provedores de credenciais na cadeia de provedores de credenciais padrão. Ao executar a aplicação de consumo em uma instância do EC2, recomenda-se configurar a instância com um perfil do IAM. As credenciais da AWS que refletem as permissões associadas a esse perfil do IAM são disponibilizadas às aplicações na instância por meio dos metadados da instância. Essa é a maneira mais segura de gerenciar credenciais para uma aplicação de consumo em execução em uma instância do EC2.

O arquivo de propriedades do exemplo configura a KCL para processar um fluxo de dados do Kinesis chamado “words” usando o processador de registros fornecido em `AmazonKinesisSampleConsumer.cs`.

Desenvolver uma aplicação de consumo da Kinesis Client Library em Python

> **Important**
>
> As versões 1.x e 2.x da Amazon Kinesis Client Library (KCL) estão desatualizadas. O KCL 1.x chegará end-of-support em 30 de janeiro de 2026. É altamente recomendável que você migre suas aplicações da KCL que usam a versão 1.x para a versão mais recente da KCL antes de 30 de janeiro de 2026. Para encontrar a versão mais recente da KCL, consulte a página da biblioteca de cliente do Amazon Kinesis em. GitHub Para obter informações sobre as versões mais recentes da KCL, consulte Usar a Kinesis Client Library. Para ter mais informações sobre como migrar da KCL 1.x para a KCL 3.x, consulte Migrar da KCL 1.x para a KCL 3.x.

É possível usar a Kinesis Client Library (KCL) para criar aplicações que processam dados dos fluxos de dados do Kinesis. A Kinesis Client Library está disponível em várias linguagens. Este tópico discute Python.

A KCL é uma biblioteca Java; o suporte para linguagens diferentes de Java é fornecido usando uma interface multilíngue chamada de. MultiLangDaemon Esse daemon baseado em Java é executado em segundo plano quando uma linguagem de KCL diferente de Java é utilizada. Portanto, se você instalar o KCL para Python e escrever seu aplicativo de consumidor inteiramente em Python, ainda precisará do Java instalado em seu sistema por causa do. MultiLangDaemon Além disso, MultiLangDaemon tem algumas configurações padrão que você pode precisar personalizar para seu caso de uso, por exemplo, a AWS região à qual ele se conecta. Para obter mais informações sobre o MultiLangDaemon on GitHub, acesse a página do MultiLangDaemon projeto KCL.

Para baixar o Python KCL em GitHub, acesse a Kinesis Client Library (Python). Para baixar o código de amostra para um aplicativo de consumidor do Python KCL, acesse a página do projeto de amostra do KCL for Python em. GitHub

É necessário concluir as seguintes tarefas ao implementar uma aplicação de consumo da KCL em Python:

Tarefas

- Implemente os métodos RecordProcessor de classe
- Modificar as propriedades de configuração

Implemente os métodos RecordProcessor de classe

A classe `RecordProcess` precisa estender o `RecordProcessorBase` para implementar os métodos a seguir. O exemplo fornece implementações que podem ser usadas como ponto de partida (consulte `sample_kclpy_app.py`).

```
def initialize(self, shard_id)
def process_records(self, records, checkpointer)
def shutdown(self, checkpointer, reason)
```

inicializar

A KCL chama o método `initialize` quando o processador de registros é instanciado, passando um ID de fragmento específico como um parâmetro. Esse processador de registros processa apenas

esse fragmento e, normalmente, o inverso também é verdadeiro (esse fragmento é processado somente por esse processador de registro). No entanto, a aplicação de consumo deve considerar a possibilidade de que um registro de dados pode ser processado mais de uma vez. Isso acontece porque a semântica do Kinesis Data Streams é do tipo pelo menos uma vez, o que significa que cada registro de dados de um fragmento é processado pelo menos uma vez por um operador no consumidor. Para obter mais informações sobre casos em que um fragmento específico pode ser processado por mais de um operador, consulte Use refragmentação, escalonamento e processamento paralelo para alterar o número de fragmentos.

```
def initialize(self, shard_id)
```

process_records

A KCL chama este método passando uma lista de registros de dados do fragmento especificado pelo método `initialize`. O processador de registros implementado processará os dados nesses registros de acordo com a semântica da aplicação de consumo. Por exemplo, o operador pode executar uma transformação nos dados e, em seguida, armazenar o resultado em um bucket do Amazon Simple Storage Service (Amazon S3).

```
def process_records(self, records, checkpointer)
```

Além dos dados em si, o registro também contém um número de sequência e uma chave de partição. O operador pode usar esses valores ao processar os dados. Por exemplo, o operador pode escolher o bucket do S3 no qual armazenar os dados com base no valor da chave de partição. O dicionário de `record` expõe os seguintes pares de chave/valor para acessar os dados do registro, o número de sequência e a chave de partição:

```
record.get('data')
record.get('sequenceNumber')
record.get('partitionKey')
```

Observe que os dados são codificados em Base64.

No exemplo, o método `process_records` tem código que mostra como um operador pode acessar os dados do registro, o número de sequência e a chave de partição.

O Kinesis Data Streams requer que o processador de registros rastreie os registros que já foram processados em um fragmento. A KCL faz esse rastreamento, passando um objeto `Checkpointer`

para `process_records`. O processador de registros chama o método `checkpoint` neste objeto para informar a KCL sobre o progresso do processamento dos registros no fragmento. Se o operador falhar, a KCL usará essas informações para reiniciar o processamento do fragmento no último registro processado conhecido.

Em uma operação de divisão ou mesclagem, a KCL só começará a processar os novos fragmentos quando os processadores dos fragmentos originais chamarem `checkpoint` para indicar que o processamento dos fragmentos originais foi concluído.

Se você não passar um parâmetro, a KCL presumirá que a chamada para `checkpoint` significa que todos os registros foram processados até o último registro passado para o processador de registros. Portanto, o processador de registros deve chamar `checkpoint` somente após ter processado todos os registros na lista que foi passada a ele. Os processadores de registros não precisam chamar `checkpoint` em cada chamada para `process_records`. Um processador pode, por exemplo, chamar `checkpoint` a cada terceira chamada. É possível, opcionalmente, especificar o número de sequência exato de um registro como um parâmetro para `checkpoint`. Nesse caso, a KCL presume que todos os registros foram processados somente até o registro especificado.

No exemplo, o método privado `checkpoint` mostra como chamar o método `Checkpointer.checkpoint` usando a lógica de novas tentativas e o tratamento de exceções apropriados.

A KCL depende do `process_records` para lidar com qualquer exceção ocorrida no processamento dos registros de dados. Se ocorrer uma exceção em `process_records`, a KCL ignorará os registros de dados passados para `process_records` antes da exceção. Ou seja, esses registros não serão reenviados para o processador de registros que lançou a exceção ou para qualquer outro processador de registros na aplicação de consumo.

shutdown

A KCL chama o método `shutdown` quando o processamento termina (o motivo do desligamento é `TERMINATE`) ou quando o operador não está mais respondendo (o `reason` do desligamento é `ZOMBIE`).

```
def shutdown(self, checkpointer, reason)
```

O processamento termina quando o processador de registros não recebe mais registros do fragmento porque ele foi dividido ou intercalado, ou o fluxo foi excluído.

A KCL também passa um objeto `Checkpointer` para `shutdown`. Se o `reason` do desligamento é `TERMINATE`, o processador de registros deve terminar o processamento de todos os registros de dados e, em seguida, chamar o método `checkpoint` nesta interface.

Modificar as propriedades de configuração

O exemplo fornece valores padrão para as propriedades de configuração. É possível substituir qualquer uma dessas propriedades por seus próprios valores (consulte `sample.properties`).

Nome da aplicação

A KCL exige um nome de aplicação exclusivo entre as aplicações e as tabelas do Amazon DynamoDB na mesma região. Ela usa o valor de configuração de nome de aplicativo das seguintes formas:

- Presume-se que todos os operadores associados a esse nome de aplicativo estejam trabalhando juntos no mesmo fluxo. Esses operadores podem ser distribuídos em várias instâncias. Se você executar uma instância adicional do mesmo código da aplicação, mas com um nome diferente, a KCL tratará a segunda instância como uma aplicação totalmente independente operando no mesmo fluxo.
- A KCL cria uma tabela do DynamoDB com o nome da aplicação e usa essa tabela para manter informações de estado (como pontos de verificação e mapeamento de operador-fragmento) da aplicação. Cada aplicação tem sua própria tabela do DynamoDB. Para obter mais informações, consulte Usar uma tabela de concessões para monitorar os fragmentos processados pela aplicação de consumo da KCL.

Configurar credenciais

Você deve disponibilizar suas AWS credenciais para um dos provedores de credenciais na cadeia de provedores de credenciais padrão. A propriedade `AWSCredentialsProvider` pode ser usada para definir um provedor de credenciais. As sample.properties precisam disponibilizar as credenciais para um dos provedores de credenciais na cadeia de provedores de credenciais padrão. Ao executar a aplicação de consumo em uma instância do Amazon EC2, recomenda-se que a instância seja configurada com um perfil do IAM. As credenciais da AWS que refletem as permissões associadas a esse perfil do IAM são disponibilizadas às aplicações na instância por meio dos metadados da instância. Essa é a maneira mais segura de gerenciar credenciais para uma aplicação de consumo em execução em uma instância do EC2.

O arquivo de propriedades do exemplo configura a KCL para processar um fluxo de dados do Kinesis chamado “words” usando o processador de registros fornecido em `sample_kclpy_app.py`.

Desenvolver uma aplicação de consumo da Kinesis Client Library em Ruby

> **Important**
>
> As versões 1.x e 2.x da Amazon Kinesis Client Library (KCL) estão desatualizadas. O KCL 1.x chegará end-of-support em 30 de janeiro de 2026. É altamente recomendável que você migre suas aplicações da KCL que usam a versão 1.x para a versão mais recente da KCL antes de 30 de janeiro de 2026. Para encontrar a versão mais recente da KCL, consulte a página da biblioteca de cliente do Amazon Kinesis em. GitHub Para obter informações sobre as versões mais recentes da KCL, consulte Usar a Kinesis Client Library. Para ter mais informações sobre como migrar da KCL 1.x para a KCL 3.x, consulte Migrar da KCL 1.x para a KCL 3.x.

É possível usar a Kinesis Client Library (KCL) para criar aplicações que processam dados dos fluxos de dados do Kinesis. A Kinesis Client Library está disponível em várias linguagens. Este tópico discute Ruby.

A KCL é uma biblioteca Java; o suporte para linguagens diferentes de Java é fornecido usando uma interface multilíngue chamada de. MultiLangDaemon Esse daemon baseado em Java é executado em segundo plano quando uma linguagem de KCL diferente de Java é utilizada. Portanto, se você instalar o KCL para Ruby e escrever seu aplicativo de consumo inteiramente em Ruby, ainda precisará do Java instalado em seu sistema por causa do. MultiLangDaemon Além disso, MultiLangDaemon tem algumas configurações padrão que você pode precisar personalizar para seu caso de uso, por exemplo, a AWS região à qual ele se conecta. Para obter mais informações sobre o MultiLangDaemon on GitHub, acesse a página do MultiLangDaemon projeto KCL.

Para baixar a KCL do Ruby GitHub, acesse a Kinesis Client Library (Ruby). Para baixar o código de amostra para um aplicativo de consumidor Ruby KCL, acesse a página do projeto de amostra KCL for Ruby em. GitHub

Para obter mais informações sobre a biblioteca de suporte da KCL Ruby, consulte a documentação da KCL para gems da Ruby.

Desenvolver aplicações de consumo da KCL 2.x

> **Important**
>
> As versões 1.x e 2.x da Amazon Kinesis Client Library (KCL) estão desatualizadas. O KCL 1.x chegará end-of-support em 30 de janeiro de 2026. É altamente recomendável que você migre suas aplicações da KCL que usam a versão 1.x para a versão mais recente da KCL antes de 30 de janeiro de 2026. Para encontrar a versão mais recente da KCL, consulte a página da biblioteca de cliente do Amazon Kinesis em. GitHub Para obter informações sobre as versões mais recentes da KCL, consulte Usar a Kinesis Client Library. Para ter mais informações sobre como migrar da KCL 1.x para a KCL 3.x, consulte Migrar da KCL 1.x para a KCL 3.x.

Este tópico mostra como usar a versão 2.0 da Kinesis Client Library (KCL).

Para obter mais informações sobre a KCL, consulte a visão geral fornecida em Developing Consumers Using the Kinesis Client Library 1.x.

Escolha um dos seguintes tópicos, dependendo do que deseja usar.

Tópicos

- Desenvolver uma aplicação de consumo da Kinesis Client Library em Java
- Desenvolver uma aplicação de consumo da Kinesis Client Library em Python
- Desenvolver consumidores de distribuição avançada com o KCL 2.x

Desenvolver uma aplicação de consumo da Kinesis Client Library em Java

> **Important**
>
> As versões 1.x e 2.x da Amazon Kinesis Client Library (KCL) estão desatualizadas. O KCL 1.x chegará end-of-support em 30 de janeiro de 2026. É altamente recomendável que você migre suas aplicações da KCL que usam a versão 1.x para a versão mais recente da KCL antes de 30 de janeiro de 2026. Para encontrar a versão mais recente da KCL, consulte a página da biblioteca de cliente do Amazon Kinesis em. GitHub Para obter informações sobre as versões mais recentes da KCL, consulte Usar a Kinesis Client Library. Para ter mais

informações sobre como migrar da KCL 1.x para a KCL 3.x, consulte Migrar da KCL 1.x para a KCL 3.x.

O código a seguir mostra uma implementação de exemplo em Java de `ProcessorFactory` e `RecordProcessor`. Para aproveitar o recurso de distribuição avançada, consulte Usar consumidores com distribuição avançada.

```
/*
 *  Copyright 2019 Amazon.com, Inc. or its affiliates. All Rights Reserved.
 *
 *  Licensed under the Amazon Software License (the "License").
 *  You may not use this file except in compliance with the License.
 *  A copy of the License is located at
 *
 *  http://aws.amazon.com/asl/
 *
 *  or in the "license" file accompanying this file. This file is distributed
 *  on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either
 *  express or implied. See the License for the specific language governing
 *  permissions and limitations under the License.
 */


/*
 * Copyright 2019 Amazon.com, Inc. or its affiliates. All Rights Reserved.
 *
 * Licensed under the Apache License, Version 2.0 (the "License").
 * You may not use this file except in compliance with the License.
 * A copy of the License is located at
 *
 *     http://www.apache.org/licenses/LICENSE-2.0
 *
 * or in the "license" file accompanying this file. This file is distributed
 * on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either
 * express or implied. See the License for the specific language governing
 * permissions and limitations under the License.
 */

import java.io.BufferedReader;
import java.io.IOException;
```

```
import java.io.InputStreamReader;
import java.util.UUID;
import java.util.concurrent.ExecutionException;
import java.util.concurrent.Executors;
import java.util.concurrent.Future;
import java.util.concurrent.ScheduledExecutorService;
import java.util.concurrent.ScheduledFuture;
import java.util.concurrent.TimeUnit;
import java.util.concurrent.TimeoutException;

import org.apache.commons.lang3.ObjectUtils;
import org.apache.commons.lang3.RandomStringUtils;
import org.apache.commons.lang3.RandomUtils;
import org.slf4j.Logger;
import org.slf4j.LoggerFactory;
import org.slf4j.MDC;

import software.amazon.awssdk.core.SdkBytes;
import software.amazon.awssdk.regions.Region;
import software.amazon.awssdk.services.cloudwatch.CloudWatchAsyncClient;
import software.amazon.awssdk.services.dynamodb.DynamoDbAsyncClient;
import software.amazon.awssdk.services.kinesis.KinesisAsyncClient;
import software.amazon.awssdk.services.kinesis.model.PutRecordRequest;
import software.amazon.kinesis.common.ConfigsBuilder;
import software.amazon.kinesis.common.KinesisClientUtil;
import software.amazon.kinesis.coordinator.Scheduler;
import software.amazon.kinesis.exceptions.InvalidStateException;
import software.amazon.kinesis.exceptions.ShutdownException;
import software.amazon.kinesis.lifecycle.events.InitializationInput;
import software.amazon.kinesis.lifecycle.events.LeaseLostInput;
import software.amazon.kinesis.lifecycle.events.ProcessRecordsInput;
import software.amazon.kinesis.lifecycle.events.ShardEndedInput;
import software.amazon.kinesis.lifecycle.events.ShutdownRequestedInput;

import software.amazon.kinesis.processor.ShardRecordProcessor;
import software.amazon.kinesis.processor.ShardRecordProcessorFactory;
import software.amazon.kinesis.retrieval.polling.PollingConfig;

/**
 * This class will run a simple app that uses the KCL to read data and uses the AWS SDK
 to publish data.
 * Before running this program you must first create a Kinesis stream through the AWS
 console or AWS SDK.
 */
```

```
public class SampleSingle {

    private static final Logger log = LoggerFactory.getLogger(SampleSingle.class);

    /**
     * Invoke the main method with 2 args: the stream name and (optionally) the region.
     * Verifies valid inputs and then starts running the app.
     */
    public static void main(String... args) {
        if (args.length < 1) {
            log.error("At a minimum, the stream name is required as the first argument.
 The Region may be specified as the second argument.");
            System.exit(1);
        }

        String streamName = args[0];
        String region = null;
        if (args.length > 1) {
            region = args[1];
        }

        new SampleSingle(streamName, region).run();
    }

    private final String streamName;
    private final Region region;
    private final KinesisAsyncClient kinesisClient;

    /**
     * Constructor sets streamName and region. It also creates a KinesisClient object
 to send data to Kinesis.
     * This KinesisClient is used to send dummy data so that the consumer has something
 to read; it is also used
     * indirectly by the KCL to handle the consumption of the data.
     */
    private SampleSingle(String streamName, String region) {
        this.streamName = streamName;
        this.region = Region.of(ObjectUtils.firstNonNull(region, "us-east-2"));
        this.kinesisClient =
 KinesisClientUtil.createKinesisAsyncClient(KinesisAsyncClient.builder().region(this.region));
    }

    private void run() {
```

```
        /**
         * Sends dummy data to Kinesis. Not relevant to consuming the data with the KCL
         */
        ScheduledExecutorService producerExecutor =
Executors.newSingleThreadScheduledExecutor();
        ScheduledFuture<?> producerFuture =
producerExecutor.scheduleAtFixedRate(this::publishRecord, 10, 1, TimeUnit.SECONDS);

        /**
         * Sets up configuration for the KCL, including DynamoDB and CloudWatch
dependencies. The final argument, a
         * ShardRecordProcessorFactory, is where the logic for record processing lives,
and is located in a private
         * class below.
         */
        DynamoDbAsyncClient dynamoClient =
DynamoDbAsyncClient.builder().region(region).build();
        CloudWatchAsyncClient cloudWatchClient =
CloudWatchAsyncClient.builder().region(region).build();
        ConfigsBuilder configsBuilder = new ConfigsBuilder(streamName, streamName,
kinesisClient, dynamoClient, cloudWatchClient, UUID.randomUUID().toString(), new
SampleRecordProcessorFactory());

        /**
         * The Scheduler (also called Worker in earlier versions of the KCL) is the
entry point to the KCL. This
         * instance is configured with defaults provided by the ConfigsBuilder.
         */
        Scheduler scheduler = new Scheduler(
                configsBuilder.checkpointConfig(),
                configsBuilder.coordinatorConfig(),
                configsBuilder.leaseManagementConfig(),
                configsBuilder.lifecycleConfig(),
                configsBuilder.metricsConfig(),
                configsBuilder.processorConfig(),
                configsBuilder.retrievalConfig().retrievalSpecificConfig(new
PollingConfig(streamName, kinesisClient))
        );

        /**
         * Kickoff the Scheduler. Record processing of the stream of dummy data will
continue indefinitely
         * until an exit is triggered.
         */
```

```
        Thread schedulerThread = new Thread(scheduler);
        schedulerThread.setDaemon(true);
        schedulerThread.start();

        /**
         * Allows termination of app by pressing Enter.
         */
        System.out.println("Press enter to shutdown");
        BufferedReader reader = new BufferedReader(new InputStreamReader(System.in));
        try {
            reader.readLine();
        } catch (IOException ioex) {
            log.error("Caught exception while waiting for confirm. Shutting down.",
ioex);
        }

        /**
         * Stops sending dummy data.
         */
        log.info("Cancelling producer and shutting down executor.");
        producerFuture.cancel(true);
        producerExecutor.shutdownNow();

        /**
         * Stops consuming data. Finishes processing the current batch of data already
received from Kinesis
         * before shutting down.
         */
        Future<Boolean> gracefulShutdownFuture = scheduler.startGracefulShutdown();
        log.info("Waiting up to 20 seconds for shutdown to complete.");
        try {
            gracefulShutdownFuture.get(20, TimeUnit.SECONDS);
        } catch (InterruptedException e) {
            log.info("Interrupted while waiting for graceful shutdown. Continuing.");
        } catch (ExecutionException e) {
            log.error("Exception while executing graceful shutdown.", e);
        } catch (TimeoutException e) {
            log.error("Timeout while waiting for shutdown.  Scheduler may not have
exited.");
        }
        log.info("Completed, shutting down now.");
    }

    /**
```

```
     * Sends a single record of dummy data to Kinesis.
     */
    private void publishRecord() {
        PutRecordRequest request = PutRecordRequest.builder()
                .partitionKey(RandomStringUtils.randomAlphabetic(5, 20))
                .streamName(streamName)
                .data(SdkBytes.fromByteArray(RandomUtils.nextBytes(10)))
                .build();
        try {
            kinesisClient.putRecord(request).get();
        } catch (InterruptedException e) {
            log.info("Interrupted, assuming shutdown.");
        } catch (ExecutionException e) {
            log.error("Exception while sending data to Kinesis. Will try again next
cycle.", e);
        }
    }

    private static class SampleRecordProcessorFactory implements
ShardRecordProcessorFactory {
        public ShardRecordProcessor shardRecordProcessor() {
            return new SampleRecordProcessor();
        }
    }

    /**
     * The implementation of the ShardRecordProcessor interface is where the heart of
the record processing logic lives.
     * In this example all we do to 'process' is log info about the records.
     */
    private static class SampleRecordProcessor implements ShardRecordProcessor {

        private static final String SHARD_ID_MDC_KEY = "ShardId";

        private static final Logger log =
LoggerFactory.getLogger(SampleRecordProcessor.class);

        private String shardId;

        /**
         * Invoked by the KCL before data records are delivered to the
ShardRecordProcessor instance (via
         * processRecords). In this example we do nothing except some logging.
         *
```

```
         * @param initializationInput Provides information related to initialization.
         */
        public void initialize(InitializationInput initializationInput) {
            shardId = initializationInput.shardId();
            MDC.put(SHARD_ID_MDC_KEY, shardId);
            try {
                log.info("Initializing @ Sequence: {}",
initializationInput.extendedSequenceNumber());
            } finally {
                MDC.remove(SHARD_ID_MDC_KEY);
            }
        }

        /**
         * Handles record processing logic. The Amazon Kinesis Client Library will
invoke this method to deliver
         * data records to the application. In this example we simply log our records.
         *
         * @param processRecordsInput Provides the records to be processed as well as
information and capabilities
         *                            related to them (e.g. checkpointing).
         */
        public void processRecords(ProcessRecordsInput processRecordsInput) {
            MDC.put(SHARD_ID_MDC_KEY, shardId);
            try {
                log.info("Processing {} record(s)",
processRecordsInput.records().size());
                processRecordsInput.records().forEach(r -> log.info("Processing record
pk: {} -- Seq: {}", r.partitionKey(), r.sequenceNumber()));
            } catch (Throwable t) {
                log.error("Caught throwable while processing records. Aborting.");
                Runtime.getRuntime().halt(1);
            } finally {
                MDC.remove(SHARD_ID_MDC_KEY);
            }
        }

        /** Called when the lease tied to this record processor has been lost. Once the
lease has been lost,
         * the record processor can no longer checkpoint.
         *
         * @param leaseLostInput Provides access to functions and data related to the
loss of the lease.
         */
```

```
        public void leaseLost(LeaseLostInput leaseLostInput) {
            MDC.put(SHARD_ID_MDC_KEY, shardId);
            try {
                log.info("Lost lease, so terminating.");
            } finally {
                MDC.remove(SHARD_ID_MDC_KEY);
            }
        }

        /**
         * Called when all data on this shard has been processed. Checkpointing must
occur in the method for record
         * processing to be considered complete; an exception will be thrown otherwise.
         *
         * @param shardEndedInput Provides access to a checkpointer method for
completing processing of the shard.
         */
        public void shardEnded(ShardEndedInput shardEndedInput) {
            MDC.put(SHARD_ID_MDC_KEY, shardId);
            try {
                log.info("Reached shard end checkpointing.");
                shardEndedInput.checkpointer().checkpoint();
            } catch (ShutdownException | InvalidStateException e) {
                log.error("Exception while checkpointing at shard end. Giving up.", e);
            } finally {
                MDC.remove(SHARD_ID_MDC_KEY);
            }
        }

        /**
         * Invoked when Scheduler has been requested to shut down (i.e. we decide to
stop running the app by pressing
         * Enter). Checkpoints and logs the data a final time.
         *
         * @param shutdownRequestedInput Provides access to a checkpointer, allowing a
record processor to checkpoint
         *                               before the shutdown is completed.
         */
        public void shutdownRequested(ShutdownRequestedInput shutdownRequestedInput) {
            MDC.put(SHARD_ID_MDC_KEY, shardId);
            try {
                log.info("Scheduler is shutting down, checkpointing.");
                shutdownRequestedInput.checkpointer().checkpoint();
            } catch (ShutdownException | InvalidStateException e) {
```

```
                log.error("Exception while checkpointing at requested shutdown. Giving
 up.", e);
            } finally {
                MDC.remove(SHARD_ID_MDC_KEY);
            }
        }
    }

}
```

Desenvolver uma aplicação de consumo da Kinesis Client Library em Python

> **Important**
>
> As versões 1.x e 2.x da Amazon Kinesis Client Library (KCL) estão desatualizadas. O KCL 1.x chegará end-of-support em 30 de janeiro de 2026. É altamente recomendável que você migre suas aplicações da KCL que usam a versão 1.x para a versão mais recente da KCL antes de 30 de janeiro de 2026. Para encontrar a versão mais recente da KCL, consulte a página da biblioteca de cliente do Amazon Kinesis em. GitHub Para obter informações sobre as versões mais recentes da KCL, consulte Usar a Kinesis Client Library. Para ter mais informações sobre como migrar da KCL 1.x para a KCL 3.x, consulte Migrar da KCL 1.x para a KCL 3.x.

É possível usar a Kinesis Client Library (KCL) para criar aplicações que processam dados dos fluxos de dados do Kinesis. A Kinesis Client Library está disponível em várias linguagens. Este tópico discute Python.

A KCL é uma biblioteca Java; o suporte para linguagens diferentes de Java é fornecido usando uma interface multilíngue chamada de. MultiLangDaemon Esse daemon baseado em Java é executado em segundo plano quando uma linguagem de KCL diferente de Java é utilizada. Portanto, se você instalar o KCL para Python e escrever seu aplicativo de consumidor inteiramente em Python, ainda precisará do Java instalado em seu sistema por causa do. MultiLangDaemon Além disso, MultiLangDaemon tem algumas configurações padrão que você pode precisar personalizar para seu caso de uso, por exemplo, a AWS região à qual ele se conecta. Para obter mais informações sobre o MultiLangDaemon on GitHub, acesse a página do MultiLangDaemon projeto KCL.

Para baixar o Python KCL em GitHub, acesse a Kinesis Client Library (Python). Para baixar o código de amostra para um aplicativo de consumidor do Python KCL, acesse a página do projeto de amostra do KCL for Python em. GitHub

É necessário concluir as seguintes tarefas ao implementar uma aplicação de consumo da KCL em Python:

Tarefas

- Implemente os métodos RecordProcessor de classe
- Modificar as propriedades de configuração

Implemente os métodos RecordProcessor de classe

A classe `RecordProcess` precisa estender a classe `RecordProcessorBase` para implementar os seguintes métodos:

```
initialize
process_records
shutdown_requested
```

Este exemplo fornece implementações que podem ser usadas como ponto de partida.

```
#!/usr/bin/env python

# Copyright 2014-2015 Amazon.com, Inc. or its affiliates. All Rights Reserved.
#
# Licensed under the Amazon Software License (the "License").
# You may not use this file except in compliance with the License.
# A copy of the License is located at
#
# http://aws.amazon.com/asl/
#
# or in the "license" file accompanying this file. This file is distributed
# on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either
# express or implied. See the License for the specific language governing
# permissions and limitations under the License.

from __future__ import print_function

import sys
import time
```

```
from amazon_kclpy import kcl
from amazon_kclpy.v3 import processor


class RecordProcessor(processor.RecordProcessorBase):
    """
    A RecordProcessor processes data from a shard in a stream. Its methods will be
 called with this pattern:

    * initialize will be called once
    * process_records will be called zero or more times
    * shutdown will be called if this MultiLangDaemon instance loses the lease to this
 shard, or the shard ends due
        a scaling change.
    """
    def __init__(self):
        self._SLEEP_SECONDS = 5
        self._CHECKPOINT_RETRIES = 5
        self._CHECKPOINT_FREQ_SECONDS = 60
        self._largest_seq = (None, None)
        self._largest_sub_seq = None
        self._last_checkpoint_time = None

    def log(self, message):
        sys.stderr.write(message)

    def initialize(self, initialize_input):
        """
        Called once by a KCLProcess before any calls to process_records

        :param amazon_kclpy.messages.InitializeInput initialize_input: Information
 about the lease that this record
            processor has been assigned.
        """
        self._largest_seq = (None, None)
        self._last_checkpoint_time = time.time()

    def checkpoint(self, checkpointer, sequence_number=None, sub_sequence_number=None):
        """
        Checkpoints with retries on retryable exceptions.

        :param amazon_kclpy.kcl.Checkpointer checkpointer: the checkpointer provided to
 either process_records
```

```
            or shutdown
        :param str or None sequence_number: the sequence number to checkpoint at.
        :param int or None sub_sequence_number: the sub sequence number to checkpoint
at.
        """
        for n in range(0, self._CHECKPOINT_RETRIES):
            try:
                checkpointer.checkpoint(sequence_number, sub_sequence_number)
                return
            except kcl.CheckpointError as e:
                if 'ShutdownException' == e.value:
                    #
                    # A ShutdownException indicates that this record processor should
be shutdown. This is due to
                    # some failover event, e.g. another MultiLangDaemon has taken the
lease for this shard.
                    #
                    print('Encountered shutdown exception, skipping checkpoint')
                    return
                elif 'ThrottlingException' == e.value:
                    #
                    # A ThrottlingException indicates that one of our dependencies is
is over burdened, e.g. too many
                    # dynamo writes. We will sleep temporarily to let it recover.
                    #
                    if self._CHECKPOINT_RETRIES - 1 == n:
                        sys.stderr.write('Failed to checkpoint after {n} attempts,
giving up.\n'.format(n=n))
                        return
                    else:
                        print('Was throttled while checkpointing, will attempt again in
{s} seconds'
                              .format(s=self._SLEEP_SECONDS))
                elif 'InvalidStateException' == e.value:
                    sys.stderr.write('MultiLangDaemon reported an invalid state while
checkpointing.\n')
                else:  # Some other error
                    sys.stderr.write('Encountered an error while checkpointing, error
was {e}.\n'.format(e=e))
            time.sleep(self._SLEEP_SECONDS)

    def process_record(self, data, partition_key, sequence_number,
sub_sequence_number):
        """
```

```
        Called for each record that is passed to process_records.

        :param str data: The blob of data that was contained in the record.
        :param str partition_key: The key associated with this recod.
        :param int sequence_number: The sequence number associated with this record.
        :param int sub_sequence_number: the sub sequence number associated with this
record.
        """
        ####################################
        # Insert your processing logic here
        ####################################
        self.log("Record (Partition Key: {pk}, Sequence Number: {seq}, Subsequence
Number: {sseq}, Data Size: {ds}"
                 .format(pk=partition_key, seq=sequence_number,
sseq=sub_sequence_number, ds=len(data)))

    def should_update_sequence(self, sequence_number, sub_sequence_number):
        """
        Determines whether a new larger sequence number is available

        :param int sequence_number: the sequence number from the current record
        :param int sub_sequence_number: the sub sequence number from the current record
        :return boolean: true if the largest sequence should be updated, false
otherwise
        """
        return self._largest_seq == (None, None) or sequence_number >
self._largest_seq[0] or \
            (sequence_number == self._largest_seq[0] and sub_sequence_number >
self._largest_seq[1])

    def process_records(self, process_records_input):
        """
        Called by a KCLProcess with a list of records to be processed and a
checkpointer which accepts sequence numbers
        from the records to indicate where in the stream to checkpoint.

        :param amazon_kclpy.messages.ProcessRecordsInput process_records_input: the
records, and metadata about the
            records.
        """
        try:
            for record in process_records_input.records:
                data = record.binary_data
                seq = int(record.sequence_number)
```

```
                sub_seq = record.sub_sequence_number
                key = record.partition_key
                self.process_record(data, key, seq, sub_seq)
                if self.should_update_sequence(seq, sub_seq):
                    self._largest_seq = (seq, sub_seq)

            #
            # Checkpoints every self._CHECKPOINT_FREQ_SECONDS seconds
            #
            if time.time() - self._last_checkpoint_time >
 self._CHECKPOINT_FREQ_SECONDS:
                self.checkpoint(process_records_input.checkpointer,
 str(self._largest_seq[0]), self._largest_seq[1])
                self._last_checkpoint_time = time.time()

        except Exception as e:
            self.log("Encountered an exception while processing records. Exception was
 {e}\n".format(e=e))

    def lease_lost(self, lease_lost_input):
        self.log("Lease has been lost")

    def shard_ended(self, shard_ended_input):
        self.log("Shard has ended checkpointing")
        shard_ended_input.checkpointer.checkpoint()

    def shutdown_requested(self, shutdown_requested_input):
        self.log("Shutdown has been requested, checkpointing.")
        shutdown_requested_input.checkpointer.checkpoint()


if __name__ == "__main__":
    kcl_process = kcl.KCLProcess(RecordProcessor())
    kcl_process.run()
```

Modificar as propriedades de configuração

O exemplo fornece valores padrão para as propriedades de configuração, conforme mostrado no script a seguir. É possível substituir qualquer uma dessas propriedades por seus próprios valores.

```
# The script that abides by the multi-language protocol. This script will
# be executed by the MultiLangDaemon, which will communicate with this script
# over STDIN and STDOUT according to the multi-language protocol.
```

```
executableName = sample_kclpy_app.py

# The name of an Amazon Kinesis stream to process.
streamName = words

# Used by the KCL as the name of this application. Will be used as the name
# of an Amazon DynamoDB table which will store the lease and checkpoint
# information for workers with this application name
applicationName = PythonKCLSample

# Users can change the credentials provider the KCL will use to retrieve credentials.
# The DefaultAWSCredentialsProviderChain checks several other providers, which is
# described here:
# http://docs.aws.amazon.com/AWSJavaSDK/latest/javadoc/com/amazonaws/auth/
DefaultAWSCredentialsProviderChain.html
AWSCredentialsProvider = DefaultAWSCredentialsProviderChain

# Appended to the user agent of the KCL. Does not impact the functionality of the
# KCL in any other way.
processingLanguage = python/2.7

# Valid options at TRIM_HORIZON or LATEST.
# See http://docs.aws.amazon.com/kinesis/latest/APIReference/
API_GetShardIterator.html#API_GetShardIterator_RequestSyntax
initialPositionInStream = TRIM_HORIZON

# The following properties are also available for configuring the KCL Worker that is
 created
# by the MultiLangDaemon.

# The KCL defaults to us-east-1
#regionName = us-east-1

# Fail over time in milliseconds. A worker which does not renew it's lease within this
 time interval
# will be regarded as having problems and it's shards will be assigned to other
 workers.
# For applications that have a large number of shards, this msy be set to a higher
 number to reduce
# the number of DynamoDB IOPS required for tracking leases
#failoverTimeMillis = 10000

# A worker id that uniquely identifies this worker among all workers using the same
 applicationName
```

```
# If this isn't provided a MultiLangDaemon instance will assign a unique workerId to
 itself.
#workerId =

# Shard sync interval in milliseconds - e.g. wait for this long between shard sync
 tasks.
#shardSyncIntervalMillis = 60000

# Max records to fetch from Kinesis in a single GetRecords call.
#maxRecords = 10000

# Idle time between record reads in milliseconds.
#idleTimeBetweenReadsInMillis = 1000

# Enables applications flush/checkpoint (if they have some data "in progress", but
 don't get new data for while)
#callProcessRecordsEvenForEmptyRecordList = false

# Interval in milliseconds between polling to check for parent shard completion.
# Polling frequently will take up more DynamoDB IOPS (when there are leases for shards
 waiting on
# completion of parent shards).
#parentShardPollIntervalMillis = 10000

# Cleanup leases upon shards completion (don't wait until they expire in Kinesis).
# Keeping leases takes some tracking/resources (e.g. they need to be renewed,
 assigned), so by default we try
# to delete the ones we don't need any longer.
#cleanupLeasesUponShardCompletion = true

# Backoff time in milliseconds for Amazon Kinesis Client Library tasks (in the event of
 failures).
#taskBackoffTimeMillis = 500

# Buffer metrics for at most this long before publishing to CloudWatch.
#metricsBufferTimeMillis = 10000

# Buffer at most this many metrics before publishing to CloudWatch.
#metricsMaxQueueSize = 10000

# KCL will validate client provided sequence numbers with a call to Amazon Kinesis
 before checkpointing for calls
# to RecordProcessorCheckpointer#checkpoint(String) by default.
#validateSequenceNumberBeforeCheckpointing = true
```

```
# The maximum number of active threads for the MultiLangDaemon to permit.
# If a value is provided then a FixedThreadPool is used with the maximum
# active threads set to the provided value. If a non-positive integer or no
# value is provided a CachedThreadPool is used.
#maxActiveThreads = 0
```

Nome da aplicação

A KCL exige um nome de aplicação exclusivo entre as aplicações e as tabelas do Amazon DynamoDB na mesma região. Ela usa o valor de configuração de nome de aplicativo das seguintes formas:

- Presume-se que todos os operadores associados a esse nome de aplicativo estejam trabalhando juntos no mesmo fluxo. Esses operadores podem ser distribuídos entre várias instâncias. Se você executar uma instância adicional do mesmo código da aplicação, mas com um nome diferente, a KCL tratará a segunda instância como uma aplicação totalmente independente operando no mesmo fluxo.
- A KCL cria uma tabela do DynamoDB com o nome da aplicação e usa essa tabela para manter informações de estado (como pontos de verificação e mapeamento de operador-fragmento) da aplicação. Cada aplicação tem sua própria tabela do DynamoDB. Para obter mais informações, consulte Usar uma tabela de concessões para monitorar os fragmentos processados pela aplicação de consumo da KCL.

Credenciais

Você deve disponibilizar suas AWS credenciais para um dos provedores de credenciais na cadeia de provedores de credenciais padrão. A propriedade `AWSCredentialsProvider` pode ser usada para definir um provedor de credenciais. Se você executar seu aplicativo de consumidor em uma instância do Amazon EC2, recomendamos que você configure a instância com uma função do IAM. AWS as credenciais que refletem as permissões associadas a essa função do IAM são disponibilizadas aos aplicativos na instância por meio dos metadados da instância. Essa é a maneira mais segura de gerenciar credenciais para uma aplicação de consumo em execução em uma instância do EC2.

Desenvolver consumidores de distribuição avançada com o KCL 2.x

> **Important**
>
> As versões 1.x e 2.x da Amazon Kinesis Client Library (KCL) estão desatualizadas. O KCL 1.x chegará end-of-support em 30 de janeiro de 2026. É altamente recomendável que você migre suas aplicações da KCL que usam a versão 1.x para a versão mais recente da KCL antes de 30 de janeiro de 2026. Para encontrar a versão mais recente da KCL, consulte a página da biblioteca de cliente do Amazon Kinesis em. GitHub Para obter informações sobre as versões mais recentes da KCL, consulte Usar a Kinesis Client Library. Para ter mais informações sobre como migrar da KCL 1.x para a KCL 3.x, consulte Migrar da KCL 1.x para a KCL 3.x.

Os consumidores que usam a distribuição avançada no Amazon Kinesis Data Streams podem receber registros de um fluxo de dados com throughput dedicada de até 2 MB de dados por segundo por fragmento. Esse tipo de consumidor não precisa lidar com outros consumidores que estejam recebendo dados do fluxo. Para obter mais informações, consulte Desenvolver consumidores de distribuição avançada com throughput dedicado.

É possível usar a versão 2.0 ou posterior da Kinesis Client Library (KCL) para desenvolver aplicações que usam a distribuição avançada para receber dados de fluxos. A KCL inscreve automaticamente seu aplicativo em todos os fragmentos de um stream e garante que seu aplicativo consumidor possa ler com um valor de taxa de transferência de 2 por fragmento. MB/sec Para usar a KCL sem ativar a distribuição avançada, consulte Desenvolvendo aplicações de consumo usando a Kinesis Client Library 2.0.

Tópicos

- Desenvolver consumidores de distribuição avançada usando o KCL 2.x em Java

Desenvolver consumidores de distribuição avançada usando o KCL 2.x em Java

> **Important**
>
> As versões 1.x e 2.x da Amazon Kinesis Client Library (KCL) estão desatualizadas. O KCL 1.x chegará end-of-support em 30 de janeiro de 2026. É altamente recomendável que você migre suas aplicações da KCL que usam a versão 1.x para a versão mais recente da KCL antes de 30 de janeiro de 2026. Para encontrar a versão mais recente da KCL, consulte

a página da biblioteca de cliente do Amazon Kinesis em. GitHub Para obter informações sobre as versões mais recentes da KCL, consulte Usar a Kinesis Client Library. Para ter mais informações sobre como migrar da KCL 1.x para a KCL 3.x, consulte Migrar da KCL 1.x para a KCL 3.x.

É possível usar a versão 2.0 ou posterior da Kinesis Client Library (KCL) para desenvolver aplicações no Amazon Kinesis Data Streams que recebem dados de fluxos usando a distribuição avançada. O código a seguir mostra uma implementação de exemplo em Java de `ProcessorFactory` e `RecordProcessor`.

É recomendável o uso de `KinesisClientUtil` para criar `KinesisAsyncClient` e configurar `maxConcurrency` no `KinesisAsyncClient`.

Important

O Amazon Kinesis Client pode ter latência significativamente maior, a menos que se configure `KinesisAsyncClient` para ter um `maxConcurrency` alto o suficiente para permitir todas as concessões e usos adicionais do `KinesisAsyncClient`.

```
/*
 *  Copyright 2019 Amazon.com, Inc. or its affiliates. All Rights Reserved.
 *
 *  Licensed under the Amazon Software License (the "License").
 *  You may not use this file except in compliance with the License.
 *  A copy of the License is located at
 *
 *  http://aws.amazon.com/asl/
 *
 *  or in the "license" file accompanying this file. This file is distributed
 *  on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either
 *  express or implied. See the License for the specific language governing
 *  permissions and limitations under the License.
 */

/*
 * Copyright 2019 Amazon.com, Inc. or its affiliates. All Rights Reserved.
 *
 * Licensed under the Apache License, Version 2.0 (the "License").
```

```
 * You may not use this file except in compliance with the License.
 * A copy of the License is located at
 *
 *     http://www.apache.org/licenses/LICENSE-2.0
 *
 * or in the "license" file accompanying this file. This file is distributed
 * on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either
 * express or implied. See the License for the specific language governing
 * permissions and limitations under the License.
 */

import java.io.BufferedReader;
import java.io.IOException;
import java.io.InputStreamReader;
import java.util.UUID;
import java.util.concurrent.ExecutionException;
import java.util.concurrent.Executors;
import java.util.concurrent.Future;
import java.util.concurrent.ScheduledExecutorService;
import java.util.concurrent.ScheduledFuture;
import java.util.concurrent.TimeUnit;
import java.util.concurrent.TimeoutException;

import org.apache.commons.lang3.ObjectUtils;
import org.apache.commons.lang3.RandomStringUtils;
import org.apache.commons.lang3.RandomUtils;
import org.slf4j.Logger;
import org.slf4j.LoggerFactory;
import org.slf4j.MDC;

import software.amazon.awssdk.core.SdkBytes;
import software.amazon.awssdk.regions.Region;
import software.amazon.awssdk.services.cloudwatch.CloudWatchAsyncClient;
import software.amazon.awssdk.services.dynamodb.DynamoDbAsyncClient;
import software.amazon.awssdk.services.kinesis.KinesisAsyncClient;
import software.amazon.awssdk.services.kinesis.model.PutRecordRequest;
import software.amazon.kinesis.common.ConfigsBuilder;
import software.amazon.kinesis.common.KinesisClientUtil;
import software.amazon.kinesis.coordinator.Scheduler;
import software.amazon.kinesis.exceptions.InvalidStateException;
import software.amazon.kinesis.exceptions.ShutdownException;
import software.amazon.kinesis.lifecycle.events.InitializationInput;
import software.amazon.kinesis.lifecycle.events.LeaseLostInput;
import software.amazon.kinesis.lifecycle.events.ProcessRecordsInput;
```

```
import software.amazon.kinesis.lifecycle.events.ShardEndedInput;
import software.amazon.kinesis.lifecycle.events.ShutdownRequestedInput;
import software.amazon.kinesis.processor.ShardRecordProcessor;
import software.amazon.kinesis.processor.ShardRecordProcessorFactory;

public class SampleSingle {

    private static final Logger log = LoggerFactory.getLogger(SampleSingle.class);

    public static void main(String... args) {
        if (args.length < 1) {
            log.error("At a minimum, the stream name is required as the first argument.
 The Region may be specified as the second argument.");
            System.exit(1);
        }

        String streamName = args[0];
        String region = null;
        if (args.length > 1) {
            region = args[1];
        }

        new SampleSingle(streamName, region).run();
    }

    private final String streamName;
    private final Region region;
    private final KinesisAsyncClient kinesisClient;

    private SampleSingle(String streamName, String region) {
        this.streamName = streamName;
        this.region = Region.of(ObjectUtils.firstNonNull(region, "us-east-2"));
        this.kinesisClient =
 KinesisClientUtil.createKinesisAsyncClient(KinesisAsyncClient.builder().region(this.region));
    }

    private void run() {
        ScheduledExecutorService producerExecutor =
 Executors.newSingleThreadScheduledExecutor();
        ScheduledFuture<?> producerFuture =
 producerExecutor.scheduleAtFixedRate(this::publishRecord, 10, 1, TimeUnit.SECONDS);

        DynamoDbAsyncClient dynamoClient =
 DynamoDbAsyncClient.builder().region(region).build();
```

```
        CloudWatchAsyncClient cloudWatchClient =
CloudWatchAsyncClient.builder().region(region).build();
        ConfigsBuilder configsBuilder = new ConfigsBuilder(streamName, streamName,
kinesisClient, dynamoClient, cloudWatchClient, UUID.randomUUID().toString(), new
SampleRecordProcessorFactory());

        Scheduler scheduler = new Scheduler(
                configsBuilder.checkpointConfig(),
                configsBuilder.coordinatorConfig(),
                configsBuilder.leaseManagementConfig(),
                configsBuilder.lifecycleConfig(),
                configsBuilder.metricsConfig(),
                configsBuilder.processorConfig(),
                configsBuilder.retrievalConfig()
        );

        Thread schedulerThread = new Thread(scheduler);
        schedulerThread.setDaemon(true);
        schedulerThread.start();

        System.out.println("Press enter to shutdown");
        BufferedReader reader = new BufferedReader(new InputStreamReader(System.in));
        try {
            reader.readLine();
        } catch (IOException ioex) {
            log.error("Caught exception while waiting for confirm. Shutting down.",
ioex);
        }

        log.info("Cancelling producer, and shutting down executor.");
        producerFuture.cancel(true);
        producerExecutor.shutdownNow();

        Future<Boolean> gracefulShutdownFuture = scheduler.startGracefulShutdown();
        log.info("Waiting up to 20 seconds for shutdown to complete.");
        try {
            gracefulShutdownFuture.get(20, TimeUnit.SECONDS);
        } catch (InterruptedException e) {
            log.info("Interrupted while waiting for graceful shutdown. Continuing.");
        } catch (ExecutionException e) {
            log.error("Exception while executing graceful shutdown.", e);
        } catch (TimeoutException e) {
            log.error("Timeout while waiting for shutdown. Scheduler may not have
exited.");
```

```
        }
        log.info("Completed, shutting down now.");
    }

    private void publishRecord() {
        PutRecordRequest request = PutRecordRequest.builder()
                .partitionKey(RandomStringUtils.randomAlphabetic(5, 20))
                .streamName(streamName)
                .data(SdkBytes.fromByteArray(RandomUtils.nextBytes(10)))
                .build();
        try {
            kinesisClient.putRecord(request).get();
        } catch (InterruptedException e) {
            log.info("Interrupted, assuming shutdown.");
        } catch (ExecutionException e) {
            log.error("Exception while sending data to Kinesis. Will try again next
cycle.", e);
        }
    }

    private static class SampleRecordProcessorFactory implements
ShardRecordProcessorFactory {
        public ShardRecordProcessor shardRecordProcessor() {
            return new SampleRecordProcessor();
        }
    }


    private static class SampleRecordProcessor implements ShardRecordProcessor {

        private static final String SHARD_ID_MDC_KEY = "ShardId";

        private static final Logger log =
LoggerFactory.getLogger(SampleRecordProcessor.class);

        private String shardId;

        public void initialize(InitializationInput initializationInput) {
            shardId = initializationInput.shardId();
            MDC.put(SHARD_ID_MDC_KEY, shardId);
            try {
                log.info("Initializing @ Sequence: {}",
initializationInput.extendedSequenceNumber());
            } finally {
```

```
                MDC.remove(SHARD_ID_MDC_KEY);
            }
        }

        public void processRecords(ProcessRecordsInput processRecordsInput) {
            MDC.put(SHARD_ID_MDC_KEY, shardId);
            try {
                log.info("Processing {} record(s)",
processRecordsInput.records().size());
                processRecordsInput.records().forEach(r -> log.info("Processing record
pk: {} -- Seq: {}", r.partitionKey(), r.sequenceNumber()));
            } catch (Throwable t) {
                log.error("Caught throwable while processing records. Aborting.");
                Runtime.getRuntime().halt(1);
            } finally {
                MDC.remove(SHARD_ID_MDC_KEY);
            }
        }

        public void leaseLost(LeaseLostInput leaseLostInput) {
            MDC.put(SHARD_ID_MDC_KEY, shardId);
            try {
                log.info("Lost lease, so terminating.");
            } finally {
                MDC.remove(SHARD_ID_MDC_KEY);
            }
        }

        public void shardEnded(ShardEndedInput shardEndedInput) {
            MDC.put(SHARD_ID_MDC_KEY, shardId);
            try {
                log.info("Reached shard end checkpointing.");
                shardEndedInput.checkpointer().checkpoint();
            } catch (ShutdownException | InvalidStateException e) {
                log.error("Exception while checkpointing at shard end. Giving up.", e);
            } finally {
                MDC.remove(SHARD_ID_MDC_KEY);
            }
        }

        public void shutdownRequested(ShutdownRequestedInput shutdownRequestedInput) {
            MDC.put(SHARD_ID_MDC_KEY, shardId);
            try {
                log.info("Scheduler is shutting down, checkpointing.");
```

```
                shutdownRequestedInput.checkpointer().checkpoint();
            } catch (ShutdownException | InvalidStateException e) {
                log.error("Exception while checkpointing at requested shutdown. Giving
 up.", e);
            } finally {
                MDC.remove(SHARD_ID_MDC_KEY);
            }
        }
    }

}
```

Migrar consumidores da KCL 1.x para a KCL 2.x

> **Important**
>
> As versões 1.x e 2.x da Amazon Kinesis Client Library (KCL) estão desatualizadas. O KCL 1.x chegará end-of-support em 30 de janeiro de 2026. É altamente recomendável que você migre suas aplicações da KCL que usam a versão 1.x para a versão mais recente da KCL antes de 30 de janeiro de 2026. Para encontrar a versão mais recente da KCL, consulte a página da biblioteca de cliente do Amazon Kinesis em. GitHub Para obter informações sobre as versões mais recentes da KCL, consulte Usar a Kinesis Client Library. Para ter mais informações sobre como migrar da KCL 1.x para a KCL 3.x, consulte Migrar da KCL 1.x para a KCL 3.x.

Este tópico explica as diferenças entre as versões 1.x e 2.x da Kinesis Client Library (KCL). Ele também mostra como migrar o consumidor da versão 1.x para a versão 2.x da KCL. Depois de migrar o cliente, ele iniciará o processamento de registros a partir do local verificado pela última vez.

A versão 2.0 da KCL apresenta as seguintes alterações de interface:

Alterações de interface da KCL

Interface KCL 1.x	Interface KCL 2.0
`com.amazonaws.services.kinesis.clientlibrary.interfaces.v2.IRecordProcessor`	`software.amazon.kinesis.processor.ShardRecordProcessor`

Interface KCL 1.x	Interface KCL 2.0
`com.amazonaws.services.kinesis.clientlibrary.interfaces.v2.IRecordProcessorFactory`	`software.amazon.kinesis.processor.ShardRecordProcessorFactory`
`com.amazonaws.services.kinesis.clientlibrary.interfaces.v2.IShutdownNotificationAware`	Compactada em `software.amazon.kinesis.processor.ShardRecordProcessor`

Tópicos

- Migrar o processador de registros
- Migrar a fábrica do processador de registros
- Migração do operador
- Configurar o cliente do Amazon Kinesis
- Remoção do tempo ocioso
- Remoções de configuração de cliente

Migrar o processador de registros

Este exemplo mostra um processador de registros implementado para a KCL 1.x:

```
package com.amazonaws.kcl;

import com.amazonaws.services.kinesis.clientlibrary.exceptions.InvalidStateException;
import com.amazonaws.services.kinesis.clientlibrary.exceptions.ShutdownException;
import
 com.amazonaws.services.kinesis.clientlibrary.interfaces.IRecordProcessorCheckpointer;
import com.amazonaws.services.kinesis.clientlibrary.interfaces.v2.IRecordProcessor;
import
 com.amazonaws.services.kinesis.clientlibrary.interfaces.v2.IShutdownNotificationAware;
import com.amazonaws.services.kinesis.clientlibrary.lib.worker.ShutdownReason;
import com.amazonaws.services.kinesis.clientlibrary.types.InitializationInput;
import com.amazonaws.services.kinesis.clientlibrary.types.ProcessRecordsInput;
import com.amazonaws.services.kinesis.clientlibrary.types.ShutdownInput;

public class TestRecordProcessor implements IRecordProcessor,
 IShutdownNotificationAware {
```

```
    @Override
    public void initialize(InitializationInput initializationInput) {
        //
        // Setup record processor
        //
    }

    @Override
    public void processRecords(ProcessRecordsInput processRecordsInput) {
        //
        // Process records, and possibly checkpoint
        //
    }

    @Override
    public void shutdown(ShutdownInput shutdownInput) {
        if (shutdownInput.getShutdownReason() == ShutdownReason.TERMINATE) {
            try {
                shutdownInput.getCheckpointer().checkpoint();
            } catch (ShutdownException | InvalidStateException e) {
                throw new RuntimeException(e);
            }
        }
    }

    @Override
    public void shutdownRequested(IRecordProcessorCheckpointer checkpointer) {
        try {
            checkpointer.checkpoint();
        } catch (ShutdownException | InvalidStateException e) {
            //
            // Swallow exception
            //
            e.printStackTrace();
        }
    }
}
```

Como migrar a classe de processador de registro

1. Altere as interfaces de `com.amazonaws.services.kinesis.clientlibrary.interfaces.v2.IRecordProcessor` e

`com.amazonaws.services.kinesis.clientlibrary.interfaces.v2.IShutdownNotificat` para `software.amazon.kinesis.processor.ShardRecordProcessor`, da seguinte forma:

```
// import
 com.amazonaws.services.kinesis.clientlibrary.interfaces.v2.IRecordProcessor;
// import
 com.amazonaws.services.kinesis.clientlibrary.interfaces.v2.IShutdownNotificationAware;
import software.amazon.kinesis.processor.ShardRecordProcessor;

// public class TestRecordProcessor implements IRecordProcessor,
 IShutdownNotificationAware {
public class TestRecordProcessor implements ShardRecordProcessor {
```

2. Atualize as instruções `import` para os métodos `initialize` e `processRecords`.

```
// import com.amazonaws.services.kinesis.clientlibrary.types.InitializationInput;
import software.amazon.kinesis.lifecycle.events.InitializationInput;

//import com.amazonaws.services.kinesis.clientlibrary.types.ProcessRecordsInput;
import software.amazon.kinesis.lifecycle.events.ProcessRecordsInput;
```

3. Substitua o método `shutdown` pelos novos métodos a seguir: `leaseLost`, `shardEnded`, e `shutdownRequested`.

```
//    @Override
//    public void shutdownRequested(IRecordProcessorCheckpointer checkpointer) {
//        //
//        // This is moved to shardEnded(...)
//        //
//        try {
//            checkpointer.checkpoint();
//        } catch (ShutdownException | InvalidStateException e) {
//            //
//            // Swallow exception
//            //
//            e.printStackTrace();
//        }
//    }

    @Override
    public void leaseLost(LeaseLostInput leaseLostInput) {
```

```
    }

    @Override
    public void shardEnded(ShardEndedInput shardEndedInput) {
        try {
            shardEndedInput.checkpointer().checkpoint();
        } catch (ShutdownException | InvalidStateException e) {
            //
            // Swallow the exception
            //
            e.printStackTrace();
        }
    }

//    @Override
//    public void shutdownRequested(IRecordProcessorCheckpointer checkpointer) {
//        //
//        // This is moved to shutdownRequested(ShutdownReauestedInput)
//        //
//        try {
//            checkpointer.checkpoint();
//        } catch (ShutdownException | InvalidStateException e) {
//            //
//            // Swallow exception
//            //
//            e.printStackTrace();
//        }
//    }

    @Override
    public void shutdownRequested(ShutdownRequestedInput shutdownRequestedInput) {
        try {
            shutdownRequestedInput.checkpointer().checkpoint();
        } catch (ShutdownException | InvalidStateException e) {
            //
            // Swallow the exception
            //
            e.printStackTrace();
        }
    }
```

Veja a seguir a versão atualizada da classe de processador de registro.

```
package com.amazonaws.kcl;

import software.amazon.kinesis.exceptions.InvalidStateException;
import software.amazon.kinesis.exceptions.ShutdownException;
import software.amazon.kinesis.lifecycle.events.InitializationInput;
import software.amazon.kinesis.lifecycle.events.LeaseLostInput;
import software.amazon.kinesis.lifecycle.events.ProcessRecordsInput;
import software.amazon.kinesis.lifecycle.events.ShardEndedInput;
import software.amazon.kinesis.lifecycle.events.ShutdownRequestedInput;
import software.amazon.kinesis.processor.ShardRecordProcessor;

public class TestRecordProcessor implements ShardRecordProcessor {
    @Override
    public void initialize(InitializationInput initializationInput) {

    }

    @Override
    public void processRecords(ProcessRecordsInput processRecordsInput) {

    }

    @Override
    public void leaseLost(LeaseLostInput leaseLostInput) {

    }

    @Override
    public void shardEnded(ShardEndedInput shardEndedInput) {
        try {
            shardEndedInput.checkpointer().checkpoint();
        } catch (ShutdownException | InvalidStateException e) {
            //
            // Swallow the exception
            //
            e.printStackTrace();
        }
    }

    @Override
    public void shutdownRequested(ShutdownRequestedInput shutdownRequestedInput) {
        try {
            shutdownRequestedInput.checkpointer().checkpoint();
```

```
        } catch (ShutdownException | InvalidStateException e) {
            //
            // Swallow the exception
            //
            e.printStackTrace();
        }
    }
}
```

Migrar a fábrica do processador de registros

A fábrica do processador de registros é responsável por criar processadores de registro quando uma concessão é realizada. Veja a seguir um exemplo de uma fábrica da KCL 1.x.

```
package com.amazonaws.kcl;

import com.amazonaws.services.kinesis.clientlibrary.interfaces.v2.IRecordProcessor;
import
 com.amazonaws.services.kinesis.clientlibrary.interfaces.v2.IRecordProcessorFactory;

public class TestRecordProcessorFactory implements IRecordProcessorFactory {
    @Override
    public IRecordProcessor createProcessor() {
        return new TestRecordProcessor();
    }
}
```

Migrar a fábrica do processador de registros

1. Altere a interface implementada de `com.amazonaws.services.kinesis.clientlibrary.interfaces.v2.IRecordProcessorFa` para `software.amazon.kinesis.processor.ShardRecordProcessorFactory`, da seguinte forma:

   ```
   // import
    com.amazonaws.services.kinesis.clientlibrary.interfaces.v2.IRecordProcessor;
   import software.amazon.kinesis.processor.ShardRecordProcessor;

   // import
    com.amazonaws.services.kinesis.clientlibrary.interfaces.v2.IRecordProcessorFactory;
   import software.amazon.kinesis.processor.ShardRecordProcessorFactory;
   ```

```
// public class TestRecordProcessorFactory implements IRecordProcessorFactory {
public class TestRecordProcessorFactory implements ShardRecordProcessorFactory {
```

2. Alterar a assinatura de retorno para `createProcessor`.

```
// public IRecordProcessor createProcessor() {
public ShardRecordProcessor shardRecordProcessor() {
```

Veja a seguir um exemplo da fábrica do processador de registros em 2.0:

```
package com.amazonaws.kcl;

import software.amazon.kinesis.processor.ShardRecordProcessor;
import software.amazon.kinesis.processor.ShardRecordProcessorFactory;

public class TestRecordProcessorFactory implements ShardRecordProcessorFactory {
    @Override
    public ShardRecordProcessor shardRecordProcessor() {
        return new TestRecordProcessor();
    }
}
```

Migração do operador

Na versão 2.0 da KCL, uma nova classe, chamada `Scheduler`, substitui a classe `Worker`. Veja a seguir um exemplo de um operador da KCL 1.x.

```
final KinesisClientLibConfiguration config = new KinesisClientLibConfiguration(...)
final IRecordProcessorFactory recordProcessorFactory = new RecordProcessorFactory();
final Worker worker = new Worker.Builder()
    .recordProcessorFactory(recordProcessorFactory)
    .config(config)
    .build();
```

Para migrar o operador

1. Altere a instrução `import` para a classe `Worker` para as instruções de importação para as classes `Scheduler` e `ConfigsBuilder`.

```
// import com.amazonaws.services.kinesis.clientlibrary.lib.worker.Worker;
import software.amazon.kinesis.coordinator.Scheduler;
```

```
import software.amazon.kinesis.common.ConfigsBuilder;
```

2. Crie o `ConfigsBuilder` e um `Scheduler` conforme mostrado no exemplo a seguir.

   É recomendável o uso de `KinesisClientUtil` para criar `KinesisAsyncClient` e configurar `maxConcurrency` no `KinesisAsyncClient`.

> **Important**
>
> O Amazon Kinesis Client pode ter latência significativamente maior, a menos que se configure `KinesisAsyncClient` para ter um `maxConcurrency` alto o suficiente para permitir todas as concessões e usos adicionais do `KinesisAsyncClient`.

```
import java.util.UUID;

import software.amazon.awssdk.regions.Region;
import software.amazon.awssdk.services.dynamodb.DynamoDbAsyncClient;
import software.amazon.awssdk.services.cloudwatch.CloudWatchAsyncClient;
import software.amazon.awssdk.services.kinesis.KinesisAsyncClient;
import software.amazon.kinesis.common.ConfigsBuilder;
import software.amazon.kinesis.common.KinesisClientUtil;
import software.amazon.kinesis.coordinator.Scheduler;

...

Region region = Region.AP_NORTHEAST_2;
KinesisAsyncClient kinesisClient =
 KinesisClientUtil.createKinesisAsyncClient(KinesisAsyncClient.builder().region(region));
DynamoDbAsyncClient dynamoClient =
 DynamoDbAsyncClient.builder().region(region).build();
CloudWatchAsyncClient cloudWatchClient =
 CloudWatchAsyncClient.builder().region(region).build();

ConfigsBuilder configsBuilder = new ConfigsBuilder(streamName, applicationName,
 kinesisClient, dynamoClient, cloudWatchClient, UUID.randomUUID().toString(), new
 SampleRecordProcessorFactory());

Scheduler scheduler = new Scheduler(
    configsBuilder.checkpointConfig(),
    configsBuilder.coordinatorConfig(),
```

```
    configsBuilder.leaseManagementConfig(),
    configsBuilder.lifecycleConfig(),
    configsBuilder.metricsConfig(),
    configsBuilder.processorConfig(),
    configsBuilder.retrievalConfig()
    );
```

Configurar o cliente do Amazon Kinesis

Com a versão 2.0 da Kinesis Client Library, a configuração do cliente passou de uma única classe de configuração (`KinesisClientLibConfiguration`) para seis classes. A tabela a seguir descreve a migração.

Campos de Configuração e suas Novas Classes

Campo Original	Nova classe de configuração	Description
`applicationName`	`ConfigsBuilder`	O nome da aplicação da KCL. Usado como padrão para o `tableName` e o `consumerName` .
`tableName`	`ConfigsBuilder`	Permite substituir o nome usado para a tabela de concessão do Amazon DynamoDB.
`streamName`	`ConfigsBuilder`	O nome do fluxo a partir do qual esse aplicativo processa registros.
`kinesisEndpoint`	`ConfigsBuilder`	Essa opção não existe mais. Consulte remoções de configuração de cliente.
`dynamoDBEndpoint`	`ConfigsBuilder`	Essa opção não existe mais. Consulte remoções de configuração de cliente.
`initialPositionInStreamExtended`	`RetrievalConfig`	A localização no fragmento a partir da qual a KCL começa a obter registros, começando com a execução inicial do aplicativo.

Campo Original	Nova classe de configuração	Description
`kinesisCredentialsProvider`	`ConfigsBuilder`	Essa opção não existe mais. Consulte remoções de configuração de cliente.
`dynamoDBCredentialsProvider`	`ConfigsBuilder`	Essa opção não existe mais. Consulte remoções de configuração de cliente.
`cloudWatchCredentialsProvider`	`ConfigsBuilder`	Essa opção não existe mais. Consulte remoções de configuração de cliente.
`failoverTimeMillis`	LeaseManagementConfig	O número de milissegundos que devem passar antes que se considere uma falha do proprietário da concessão.
`workerIdentifier`	`ConfigsBuilder`	Um identificador exclusivo que representa a instanciação do processador do aplicativo. Isso deve ser exclusivo.
`shardSyncIntervalMillis`	LeaseManagementConfig	O tempo entre as chamadas de sincronização de fragmentos.
`maxRecords`	`PollingConfig`	Permite definir o número máximo de registros que o Kinesis retorna.
`idleTimeBetweenReadsInMillis`	`CoordinatorConfig`	Essa opção não existe mais. Consulte a remoção do tempo ocioso.
`callProcessRecordsEvenForEmptyRecordList`	`ProcessorConfig`	Quando definido, o processador de registros é chamado mesmo quando o Kinesis não fornece nenhum registro.

Campo Original	Nova classe de configuração	Description
`parentShardPollIntervalMillis`	`CoordinatorConfig`	Com que frequência um processador de registros deve sondar a conclusão de fragmentos pai.
`cleanupLeasesUponShardCompletion`	`LeaseManagementConfig`	Quando definidas, as concessões são removidas assim que as concessões filho iniciam o processamento.
`ignoreUnexpectedChildShards`	`LeaseManagementConfig`	Quando definidos, fragmentos filho que possuem um fragmento aberto são ignorados. Essa configuração destina-se principalmente a fluxos do DynamoDB.
`kinesisClientConfig`	`ConfigsBuilder`	Essa opção não existe mais. Consulte remoções de configuração de cliente.
`dynamoDBClientConfig`	`ConfigsBuilder`	Essa opção não existe mais. Consulte remoções de configuração de cliente.
`cloudWatchClientConfig`	`ConfigsBuilder`	Essa opção não existe mais. Consulte remoções de configuração de cliente.
`taskBackoffTimeMillis`	`LifecycleConfig`	O tempo de espera para repetir tarefas com falha.
`metricsBufferTimeMillis`	`MetricsConfig`	Controla a publicação de CloudWatch métricas.
`metricsMaxQueueSize`	`MetricsConfig`	Controla a publicação de CloudWatch métricas.
`metricsLevel`	`MetricsConfig`	Controla a publicação de CloudWatch métricas.

Campo Original	Nova classe de configuração	Description
`metricsEnabledDimensions`	`MetricsConfig`	Controla a publicação de CloudWatch métricas.
`validateSequenceNumberBeforeCheckpointing`	`CheckpointConfig`	Essa opção não existe mais. Consulte a validação do número de sequência do ponto de verificação.
`regionName`	`ConfigsBuilder`	Essa opção não existe mais. Consulte remoção de configuração de cliente.
`maxLeasesForWorker`	`LeaseManagementConfig`	O número máximo de concessões que uma única instância do aplicativo deve aceitar.
`maxLeasesToStealAtOneTime`	`LeaseManagementConfig`	O número máximo de concessões que um aplicativo deve tentar roubar de uma só vez.
`initialLeaseTableReadCapacity`	`LeaseManagementConfig`	A IOPs leitura do DynamoDB que é usada se a bibliotec a cliente do Kinesis precisar criar uma nova tabela de lease do DynamoDB.
`initialLeaseTableWriteCapacity`	`LeaseManagementConfig`	A IOPs leitura do DynamoDB que é usada se a bibliotec a cliente do Kinesis precisar criar uma nova tabela de lease do DynamoDB.
`initialPositionInStreamExtended`	LeaseManagementConfig	A posição inicial do aplicativo no fluxo. Isso é usado somente durante a criação da concessão inicial.

Campo Original	Nova classe de configuração	Description
`skipShardSyncAtWorkerInitializationIfLeasesExist`	`CoordinatorConfig`	Desative a sincronização de dados de fragmento se a tabela de concessão contiver concessões existentes. TODO: KinesisEco -438
`shardPrioritization`	`CoordinatorConfig`	A priorização de fragmentos a ser usada.
`shutdownGraceMillis`	N/D	Essa opção não existe mais. Consulte MultiLang Remoções.
`timeoutInSeconds`	N/D	Essa opção não existe mais. Consulte MultiLang Remoções.
`retryGetRecordsInSeconds`	`PollingConfig`	Configura o atraso entre as GetRecords tentativas de falhas.
`maxGetRecordsThreadPool`	`PollingConfig`	O tamanho do pool de fios usado para GetRecords.
`maxLeaseRenewalThreads`	`LeaseManagementConfig`	Controla o tamanho do grupo de threads de renovação de concessão. Quanto mais concessões seu aplicativo aceitar, maior esse grupo deve ser.
`recordsFetcherFactory`	`PollingConfig`	Permite substituir a fábrica usada para criar extratores que recuperam dados dos streams.
`logWarningForTaskAfterMillis`	`LifecycleConfig`	Quanto tempo esperar antes de um aviso ser registrado caso uma tarefa não seja concluída.

Campo Original	Nova classe de configuração	Description
`listShardsBackoffTimeInMillis`	`RetrievalConfig`	O número de milissegundos de espera entre as chamadas para `ListShards` em caso de falha.
`maxListShardsRetryAttempts`	`RetrievalConfig`	O número máximo de novas tentativas de `ListShards` antes de desistir.

Remoção do tempo ocioso

Na versão 1.x da KCL, `idleTimeBetweenReadsInMillis` corresponde a duas quantidades:

- A quantidade de tempo entre as verificações de envio de tarefas. Agora é possível configurar esse tempo entre tarefas, definindo `CoordinatorConfig#shardConsumerDispatchPollIntervalMillis`.
- A quantidade de tempo inativo quando nenhum registro é retornado do Kinesis Data Streams. Na versão 2.0, em distribuição avançada registros são enviados a partir de sua respectiva recuperação. Atividade no do consumidor fragmento só ocorre quando uma solicitação é enviada.

Remoções de configuração de cliente

Na versão 2.0, a KCL não cria mais clientes. Ela depende do fornecimento de um cliente válido pelo usuário. Com essa alteração, todos os parâmetros de configuração que controlavam a criação do cliente foram removidos. Se esses parâmetros forem necessários, pode-se configurá-los nos clientes antes de fornecê-los ao `ConfigsBuilder`.

Campo removido	Configuração equivalente
`kinesisEndpoint`	Configure o SDK `KinesisAsyncClient` com o endpoint de sua preferência: `KinesisAsyncClient.builder().endpointOverride(URI.create("https://<kinesis endpoint>")).build()` .

Campo removido	Configuração equivalente
`dynamoDBEndpoint`	Configure o SDK `DynamoDbAsyncClient` com o endpoint de sua preferência: `DynamoDbAsyncClient.builder().endpointOverride(URI.create("https://<dynamodb endpoint>")).build()` .
`kinesisClientConfig`	Configure o SDK `KinesisAsyncClient` com a configuração necessária: `KinesisAsyncClient.builder().overrideConfiguration(<your configuration>).build()` .
`dynamoDBClientConfig`	Configure o SDK `DynamoDbAsyncClient` com a configuração necessária: `DynamoDbAsyncClient.builder().overrideConfiguration(<your configuration>).build()` .
`cloudWatchClientConfig`	Configure o SDK `CloudWatchAsyncClient` com a configuração necessária: `CloudWatchAsyncClient.builder().overrideConfiguration(<your configuration>).build()` .
`regionName`	Configure o SDK com a Região de sua preferência. Ela é a mesma para todos os clientes do SDK. Por exemplo, `.KinesisAsyncClient.builder().region(Region.US_WEST_2).build()`

Desenvolva consumidores com o AWS SDK para Java

Você pode desenvolver consumidores personalizados usando o Amazon Kinesis APIs Data Streams. Esta seção descreve o uso do Kinesis APIs Data Streams AWS SDK para Java com o.

> **Important**
>
> O método recomendado para desenvolvimento de consumidores personalizados do Kinesis Data Streams com throughput compartilhada envolve o uso da Kinesis Client Library (KCL). A KCL ajuda a consumir e processar dados de um fluxo de dados do Kinesis lidando com muitas das tarefas complexas associadas à computação distribuída. Para obter mais informações, consulte Desenvolver consumidores com a KCL em Java.

Tópicos

- Desenvolva consumidores com produtividade compartilhada com o AWS SDK para Java
- Desenvolva consumidores expandidos aprimorados com o AWS SDK para Java
- Interaja com os dados usando o AWS Glue Schema Registry

Desenvolva consumidores com produtividade compartilhada com o AWS SDK para Java

Um dos métodos para desenvolver consumidores personalizados do Kinesis Data Streams com compartilhamento total é usar o Amazon APIs Kinesis Data Streams com o. AWS SDK para Java Esta seção descreve o uso do Kinesis APIs Data Streams AWS SDK para Java com o. Você pode chamar o Kinesis APIs Data Streams usando outras linguagens de programação diferentes. Para obter mais informações sobre todas as opções disponíveis AWS SDKs, consulte Comece a desenvolver com a Amazon Web Services.

O código de exemplo Java nesta seção demonstra como executar operações básicas da API do Kinesis Data Streams e está dividido logicamente por tipo de operação. Esses exemplos não representam um código pronto para produção. Eles não verificam todas as exceções possíveis nem levam em conta todas as considerações de segurança ou desempenho possíveis.

Tópicos

- Como obter dados de um fluxo
- Como usar iteradores de fragmentos
- Use GetRecords
- Adaptar para uma nova fragmentação

Como obter dados de um fluxo

Os Kinesis APIs Data Streams `getShardIterator` incluem `getRecords` os métodos e que você pode invocar para recuperar registros de um stream de dados. Esse é o modelo de pull, em que o código extrai registros de dados diretamente dos fragmentos do fluxo de dados.

> **Important**
>
> Recomenda-se usar o suporte do processador de registros fornecido pela KCL para recuperar registros dos fluxos de dados. Esse é o modelo push, no qual é implementado o

código que processa os dados. A KCL recupera registros de dados do fluxo de dados e os entrega ao código da aplicação. Além disso, a KCL fornece as funcionalidades de failover, recuperação e balanceamento de carga. Para obter mais informações, consulte Desenvolver consumidores personalizados com throughput compartilhada usando a KCL.

No entanto, em alguns casos, talvez você prefira usar o Kinesis APIs Data Streams. Um exemplo disso é a implementação de ferramentas personalizadas de monitoramento ou depuração dos fluxos de dados.

Important

O Kinesis Data Streams oferece suporte a alterações do período de retenção do registro de dados no fluxo de dados. Para obter mais informações, consulte Alterar o período de retenção de dados.

Como usar iteradores de fragmentos

Os registros do fluxo são recuperados por fragmentos. Para cada fragmento e para cada lote de registros que recupera desse fragmento, é necessário obter um iterador de fragmentos. O iterador de fragmentos é usado no objeto `getRecordsRequest` para especificar o fragmento a partir do qual os registros devem ser recuperados. O tipo associado ao iterador de fragmentos determina o ponto no fragmento a partir do qual os registros devem ser recuperados (veja mais à frente nesta seção para obter mais detalhes). Para trabalhar com o iterador de fragmentos, é necessário recuperar o fragmento. Para obter mais informações, consulte Listar fragmentos.

Obtenha o iterador de fragmentos inicial usando o método `getShardIterator`. Obtenha iteradores de fragmentos para obter mais lotes de registros usando o método `getNextShardIterator` do objeto`getRecordsResult` retornado pelo método `getRecords`. Um iterador de fragmentos é válido por 5 minutos. Se um iterador de fragmentos for usado durante sua validade, um novo será fornecido. Cada iterador de fragmentos permanecerá válido por 5 minutos, mesmo depois de ser usado.

Para obter o iterador de fragmentos inicial, instancie `GetShardIteratorRequest` e passe-o ao método `getShardIterator`. Para configurar a solicitação, especifique o fluxo e o ID do fragmento. Para obter informações sobre como obter os streams em sua AWS conta, consulteListar fluxos. Para obter informações sobre como obter os fragmentos em um fluxo, consulte Listar fragmentos.

```
String shardIterator;
GetShardIteratorRequest getShardIteratorRequest = new GetShardIteratorRequest();
getShardIteratorRequest.setStreamName(myStreamName);
getShardIteratorRequest.setShardId(shard.getShardId());
getShardIteratorRequest.setShardIteratorType("TRIM_HORIZON");

GetShardIteratorResult getShardIteratorResult =
 client.getShardIterator(getShardIteratorRequest);
shardIterator = getShardIteratorResult.getShardIterator();
```

Esse código de exemplo especifica `TRIM_HORIZON` como o tipo de iterador ao obter o iterador de fragmentos inicial. Esse tipo de iterador significa que o retorno dos registros deve começar a partir do primeiro registro adicionado ao fragmento, em vez de começar pelo registro adicionado mais recentemente, também conhecido como extremidade. Estes são tipos de iterador possíveis:

- `AT_SEQUENCE_NUMBER`
- `AFTER_SEQUENCE_NUMBER`
- `AT_TIMESTAMP`
- `TRIM_HORIZON`
- `LATEST`

Para obter mais informações, consulte ShardIteratorType.

Alguns tipos de iterador exigem que um número de sequência seja especificado, além do tipo. Por exemplo:

```
getShardIteratorRequest.setShardIteratorType("AT_SEQUENCE_NUMBER");
getShardIteratorRequest.setStartingSequenceNumber(specialSequenceNumber);
```

Depois de obter um registro usando `getRecords`, pode-se conseguir o número de sequência do registro chamando o método `getSequenceNumber` do registro.

```
record.getSequenceNumber()
```

Além disso, o código que adiciona registros ao fluxo de dados pode obter o número de sequência de um registro adicional chamando `getSequenceNumber` no resultado de `putRecord`.

```
lastSequenceNumber = putRecordResult.getSequenceNumber();
```

É possível usar números sequenciais para garantir estritamente o ordenamento crescente do registros. Para obter mais informações, consulte o exemplo de código em PutRecord exemplo.

Use GetRecords

Após obter o iterador de fragmentos, instancie um objeto `GetRecordsRequest`. Especifique o iterador para a solicitação usando o método `setShardIterator`.

Opcionalmente, também é possível definir o número de registros a recuperar usando o método `setLimit`. O número de registros retornados por `getRecords` sempre é igual ou menor que esse limite. Se esse limite não for especificado, `getRecords` retornará 10 MB de registros recuperados. O código de exemplo a seguir define esse limite para 25 registros.

Se nenhum registro for retornado, isso significa que não há registros de dados disponíveis atualmente nesse fragmento no número de sequência referenciado pelo iterador de fragmentos. Nessa situação, a aplicação deve aguardar o tempo adequado de acordo com as fontes de dados do fluxo. Em seguida, tente obter os dados do fragmento novamente usando o iterador de fragmentos retornado pela chamada anterior a `getRecords`.

Passe o `getRecordsRequest` para o método `getRecords` e capture o valor retornado como um objeto `getRecordsResult`. Para obter os registros de dados, chame o método `getRecords` no objeto `getRecordsResult`.

```
GetRecordsRequest getRecordsRequest = new GetRecordsRequest();
getRecordsRequest.setShardIterator(shardIterator);
getRecordsRequest.setLimit(25);

GetRecordsResult getRecordsResult = client.getRecords(getRecordsRequest);
List<Record> records = getRecordsResult.getRecords();
```

Para preparar-se para outra chamada para `getRecords`, obtenha o próximo iterador de fragmentos de `getRecordsResult`.

```
shardIterator = getRecordsResult.getNextShardIterator();
```

Para obter os melhores resultados, aguarde pelo menos 1 segundo (1.000 milissegundos) entre as chamadas para `getRecords` a fim de evitar exceder o limite na frequência de `getRecords`.

```
try {
  Thread.sleep(1000);
}
catch (InterruptedException e) {}
```

Normalmente, deve-se chamar `getRecords` em um loop, mesmo se estiver recuperando um único registro em um cenário de teste. Uma única chamada para `getRecords` pode retornar uma lista de registros vazia, mesmo quando o fragmento contém mais registros em números de sequência subsequentes. Quando isso ocorre, o `NextShardIterator` retornado com a lista de registros vazia faz referência a um número de sequência subsequente no fragmento, e as chamadas posteriores a `getRecords` acabam retornando os registros. O exemplo a seguir demonstra o uso de um loop.

Exemplo: getRecords

O código de exemplo a seguir reflete as dicas de `getRecords` nesta seção, inclusive chamadas em um loop.

```
// Continuously read data records from a shard
List<Record> records;

while (true) {

  // Create a new getRecordsRequest with an existing shardIterator
  // Set the maximum records to return to 25

  GetRecordsRequest getRecordsRequest = new GetRecordsRequest();
  getRecordsRequest.setShardIterator(shardIterator);
  getRecordsRequest.setLimit(25);

  GetRecordsResult result = client.getRecords(getRecordsRequest);

  // Put the result into record list. The result can be empty.
  records = result.getRecords();

  try {
    Thread.sleep(1000);
  }
  catch (InterruptedException exception) {
    throw new RuntimeException(exception);
```

```
  }

  shardIterator = result.getNextShardIterator();
}
```

Se a Kinesis Client Library estiver sendo usada, ela poderá fazer várias chamadas antes de retornar dados. Esse comportamento é projetado e não indica um problema com a KCL ou com seus dados.

Adaptar para uma nova fragmentação

`getRecordsResult.getNextShardIterator` retorna `null` para indicar que o fragmento passou por uma divisão ou uma mesclagem. O fragmento agora está em estado de `CLOSED`, e todos os registros de dados disponíveis nele foram lidos.

Nesse cenário, pode-se usar `getRecordsResult.childShards` para conhecer os fragmentos filho que foram criados pela divisão ou mesclagem do fragmento sendo processado. Para obter mais informações, consulte ChildShard.

No caso de uma divisão, os dois novos fragmentos têm `parentShardId` igual ao ID do fragmento que estavam sendo processados anteriormente. O valor de `adjacentParentShardId` dos dois fragmentos é `null`.

No caso de uma fusão, o fragmento unificado criado terá `parentShardId` igual ao ID de um dos fragmentos pai e `adjacentParentShardId` igual ao ID do outro fragmento. Seu aplicativo já leu todos os dados de um desses fragmentos. Esse é o fragmento para o qual `getRecordsResult.getNextShardIterator` retornou `null`. Se a ordem dos dados for importante para o aplicativo, deve-se garantir que ele também leia todos os dados de outro fragmento pai antes de ler qualquer dado novo de fragmento filho criado pela fusão.

Se vários processadores forem usados para recuperar dados do fluxo (digamos, um processador por fragmento) e ocorrer uma divisão ou fusão de fragmentos, deve-se aumentar ou diminuir o número de processadores para se adaptar à alteração no número de fragmentos.

Para obter mais informações sobre a refragmentação, incluindo uma discussão sobre estados de fragmentos (como `CLOSED`), consulte Refragmentar um fluxo.

Desenvolva consumidores expandidos aprimorados com o AWS SDK para Java

A distribuição avançada é um recurso do Amazon Kinesis Data Streams que permite que os consumidores recebam registros de um fluxo de dados com throughput dedicada de até 2 MB de dados por segundo por fragmento. Um consumidor que usa distribuição avançada não precisa lidar com outros consumidores que estejam recebendo dados do fluxo. Para obter mais informações, consulte Desenvolver consumidores de distribuição avançada com throughput dedicado.

É possível usar operações de API para criar um consumidor que usa a distribuição avançada no Kinesis Data Streams.

Para registrar um consumidor com distribuição avançada usando a API do Kinesis Data Streams

1. Ligue RegisterStreamConsumerpara registrar seu aplicativo como um consumidor que usa fan-out aprimorado. O Kinesis Data Streams gera um nome do recurso da Amazon (ARN) para o consumidor e o retorna na resposta.
2. Para começar a ouvir um fragmento específico, passe o ARN do consumidor em uma chamada para. SubscribeToShard Em seguida, o Kinesis Data Streams começa a enviar os registros desse fragmento para você, na forma de eventos SubscribeToShardEventdo tipo em uma conexão HTTP/2. A conexão permanece aberta por até 5 minutos. Ligue SubscribeToShardnovamente se quiser continuar recebendo registros do fragmento após o `future` retorno da chamada para ser SubscribeToShardconcluído normalmente ou excepcionalmente.

   > **Note**
   >
   > `SubscribeToShard` A API também retorna a lista dos fragmentos filho do fragmento atual quando chega ao final do fragmento.

3. Para cancelar o registro de um consumidor que está usando o fan-out aprimorado, ligue. DeregisterStreamConsumer

O de código a seguir é um exemplo de como é possível inscrever o consumidor em um fragmento, renovar a assinatura periodicamente e manipular os eventos.

```
    import software.amazon.awssdk.services.kinesis.KinesisAsyncClient;
    import software.amazon.awssdk.services.kinesis.model.ShardIteratorType;
```

```
    import software.amazon.awssdk.services.kinesis.model.SubscribeToShardEvent;
    import software.amazon.awssdk.services.kinesis.model.SubscribeToShardRequest;
    import
 software.amazon.awssdk.services.kinesis.model.SubscribeToShardResponseHandler;

    import java.util.concurrent.CompletableFuture;

    /**
     * See https://github.com/awsdocs/aws-doc-sdk-examples/blob/master/javav2/
example_code/kinesis/src/main/java/com/example/kinesis/KinesisStreamEx.java
     * for complete code and more examples.
     */
    public class SubscribeToShardSimpleImpl {

        private static final String CONSUMER_ARN = "arn:aws:kinesis:us-
east-1:123456789123:stream/foobar/consumer/test-consumer:1525898737";
        private static final String SHARD_ID = "shardId-000000000000";

        public static void main(String[] args) {

            KinesisAsyncClient client = KinesisAsyncClient.create();

            SubscribeToShardRequest request = SubscribeToShardRequest.builder()
                    .consumerARN(CONSUMER_ARN)
                    .shardId(SHARD_ID)
                    .startingPosition(s -> s.type(ShardIteratorType.LATEST)).build();

            // Call SubscribeToShard iteratively to renew the subscription
 periodically.
            while(true) {
                // Wait for the CompletableFuture to complete normally or
 exceptionally.
                callSubscribeToShardWithVisitor(client, request).join();
            }

            // Close the connection before exiting.
            // client.close();
        }


        /**
         * Subscribes to the stream of events by implementing the
 SubscribeToShardResponseHandler.Visitor interface.
         */
```

```
        private static CompletableFuture<Void>
callSubscribeToShardWithVisitor(KinesisAsyncClient client, SubscribeToShardRequest
request) {
            SubscribeToShardResponseHandler.Visitor visitor = new
SubscribeToShardResponseHandler.Visitor() {
                @Override
                public void visit(SubscribeToShardEvent event) {
                    System.out.println("Received subscribe to shard event " + event);
                }
            };
            SubscribeToShardResponseHandler responseHandler =
SubscribeToShardResponseHandler
                    .builder()
                    .onError(t -> System.err.println("Error during stream - " +
t.getMessage()))
                    .subscriber(visitor)
                    .build();
            return client.subscribeToShard(request, responseHandler);
        }
    }
```

`event.ContinuationSequenceNumber` retorna `null` para indicar que o fragmento passou por uma divisão ou uma mesclagem. O fragmento agora está em estado de `CLOSED`, todos os registros de dados disponíveis nele foram lidos. Nesse cenário, como mostrado no exemplo acima, é possível usar `event.childShards` para conhecer os fragmentos filho que foram criados pela divisão ou mesclagem do fragmento sendo processado. Para obter mais informações, consulte ChildShard.

Interaja com os dados usando o AWS Glue Schema Registry

Você pode integrar seus streams de dados do Kinesis com o AWS Glue Schema Registry. O registro de esquemas do AWS Glue permite detectar, controlar e evoluir esquemas centralmente, ao mesmo tempo que garante que os dados produzidos sejam validados continuamente por um esquema registrado. O esquema define a estrutura e o formato de um registro de dados. Um esquema é uma especificação versionada para publicação, consumo ou datastore confiáveis. O AWS Glue Schema Registry permite que você melhore a qualidade end-to-end dos dados e a governança de dados em seus aplicativos de streaming. Para obter mais informações, consulte Registro de esquemas do AWS Glue. Uma das formas de configurar essa integração é por meio da API `GetRecords` Kinesis Data Streams, disponível AWS no SDK Java.

Para obter instruções detalhadas sobre como configurar a integração do Kinesis Data Streams com o Schema Registry `GetRecords` usando o Kinesis Data Streams, consulte a seção “Interagindo com

dados usando o APIs Kinesis Data Streams" em Caso de uso: Integração do Amazon Kinesis Data APIs Streams com o Glue Schema Registry. AWS

Desenvolva consumidores usando AWS Lambda

Você pode usar uma AWS Lambda função para processar registros em um fluxo de dados. AWS Lambda é um serviço de computação que permite executar código sem provisionar ou gerenciar servidores. Ele executa seu código somente quando necessário e escala automaticamente, de algumas solicitações por dia a milhares por segundo. É cobrado apenas o tempo de computação consumido. Não haverá cobranças quando seu código não estiver em execução. Com AWS Lambda, você pode executar código para praticamente qualquer tipo de aplicativo ou serviço de back-end, tudo sem nenhuma administração. Ele executa seu código em uma infraestrutura de computação de alta disponibilidade e executa toda a administração dos recursos computacionais, inclusive manutenção de servidor e sistema operacional, provisionamento de capacidade e escalabilidade automática, monitoramento do código e registro em log. Para obter mais informações, consulte Como usar AWS Lambda com o Amazon Kinesis.

Para obter informações sobre solução de problemas, consulte Por que o gatilho do Kinesis Data Streams não consegue invocar minha função do Lambda?

Desenvolver consumidores usando o Amazon Managed Service for Apache Flink

É possível usar uma aplicação do Amazon Managed Service for Apache Flink para processar e analisar dados em um fluxo do Kinesis usando SQL, Java ou Scala. As aplicações do Managed Service for Apache Flink podem enriquecer os dados usando fontes de referência, agregar dados ao longo do tempo ou usar machine learning para encontrar anomalias nos dados. Em seguida, é possível gravar os resultados da análise em outro fluxo do Kinesis, em um fluxo de entrega do Firehose ou em uma função do Lambda. Para obter mais informações, consulte o Managed Service for Apache Flink Developer Guide for SQL Applications ou o Managed Service for Apache Flink Developer Guide for Flink Applications.

Desenvolver consumidores usando o Amazon Data Firehose

É possível usar o Firehose para ler e processar registros de um fluxo do Kinesis. O Firehose é um serviço totalmente gerenciado para fornecer dados de streaming em tempo real para destinos como Amazon S3, Amazon Redshift, Amazon OpenSearch Service e Splunk. O Firehose oferece suporte

a qualquer endpoint HTTP personalizado ou de propriedade de provedores de serviços de terceiros compatíveis, incluindo Datadog, MongoDB e New Relic. Também é possível configurar o Firehose para transformar seus registros de dados e para converter o formato do registro antes de entregar os dados para seu destino. Para obter mais informações, consulte Gravar no Firehose usando o Kinesis Data Streams.

Leia dados do Kinesis Data Streams AWS usando outros serviços

Os AWS serviços a seguir podem se integrar diretamente ao Amazon Kinesis Data Streams para ler dados dos fluxos de dados do Kinesis. Leia as informações sobre cada serviço que seja de seu interesse e consulte as referências fornecidas.

Tópicos

- Leitura de dados do Kinesis Data Streams usando Amazon EMR
- Leia dados do Kinesis Data Streams EventBridge usando o Amazon Pipes
- Leia dados do Kinesis Data Streams usando AWS Glue
- Leitura de dados do Kinesis Data Streams usando Amazon Redshift

Leitura de dados do Kinesis Data Streams usando Amazon EMR

Os clusters do Amazon EMR podem ler e processar streams do Kinesis diretamente, usando ferramentas conhecidas no ecossistema do Hadoop, como Hive, Pig, a API de streaming do Hadoop e o Cascading. MapReduce Você também pode unir dados em tempo real do Kinesis Data Streams a dados existentes no Amazon S3, no Amazon DynamoDB e no HDFS em um cluster em execução. É possível carregar diretamente os dados do Amazon EMR no Amazon S3 ou no DynamoDB para atividades de pós-processamento.

Para obter mais informações, consulte Amazon Kinesis no Guia de apresentação do Amazon EMR.

Leia dados do Kinesis Data Streams EventBridge usando o Amazon Pipes

O Amazon EventBridge Pipes oferece suporte aos streams de dados do Kinesis como fonte. O Amazon EventBridge Pipes ajuda você a criar point-to-point integrações entre produtores e consumidores de eventos com etapas opcionais de transformação, filtragem e enriquecimento. Você pode usar o EventBridge Pipes para receber registros em um stream de dados do Kinesis e, opcionalmente, filtrar ou aprimorar esses registros antes de enviá-los para um dos destinos disponíveis para processamento, incluindo o Kinesis Data Streams.

Para obter mais informações, consulte Amazon Kinesis Stream como fonte no Amazon EventBridge Release Guide.

Leia dados do Kinesis Data Streams usando AWS Glue

Usando o ETL de AWS Glue streaming, você pode criar trabalhos de extração, transformação e carregamento (ETL) de streaming que são executados continuamente e consomem dados do Amazon Kinesis Data Streams. Os trabalhos limpam e transformam os dados e, em seguida, carregam os resultados em data lakes do Amazon S3 ou armazenamentos de dados JDBC.

Para obter mais informações, consulte Trabalhos de transmissão de ETL no AWS Glue no Guia de apresentação do AWS Glue .

Leitura de dados do Kinesis Data Streams usando Amazon Redshift

A ingestão de fluxo do Amazon Redshift oferece suporte ao Amazon Kinesis Data Streams. O recurso de ingestão de fluxo do Amazon Redshift fornece ingestão de dados de baixa latência e alta velocidade do Amazon Kinesis Data Streams em uma visão materializada do Amazon Redshift. A ingestão de dados do Amazon Redshift elimina a necessidade de preparar dados no Amazon S3 antes da ingestão no Amazon Redshift.

Para obter mais informações, consulte Ingestão de fluxo no Guia de lançamento do Amazon Redshift.

Leitura a partir do Kinesis Data Streams usando integrações de terceiros

Você pode ler os dados dos fluxos de dados do Amazon Kinesis Data Streams usando uma das opções de terceiros a seguir que se integrem ao Kinesis Data Streams: Selecione a opção desejada e encontre recursos e links para a documentação relevante.

Tópicos

- Apache Flink
- Adobe Experience Platform
- Apache Druid
- Apache Spark
- Databricks

- Kafka Confluent Platform
- Kinesumer
- Talend

Apache Flink

O Apache Flink é um framework de código aberto distribuído para computações com estado em fluxos de dados delimitados e não delimitados. Para obter mais informações sobre o consumo do Kinesis Data Streams usando o Apache Flink, consulte Amazon Kinesis Data Streams Connector.

Adobe Experience Platform

A Adobe Experience Platform permite que as organizações centralizem e padronizem dados dos clientes em qualquer sistema. Em seguida, a plataforma aplica ciência de dados e machine learning para melhorar significativamente o design e a entrega de experiências ricas e personalizadas. Para obter mais informações sobre o consumo de fluxo de dados do Kinesis usando a Adobe Experience Platform, consulte Conector do Amazon Kinesis.

Apache Druid

O Druid é um banco de dados de analytics em tempo real de alta performance que entrega consultas em menos de um segundo feitas em dados de fluxo e lote em escala e sob carga. Para obter mais informações sobre a ingestão de fluxos de dados do Kinesis usando o Apache Druid, consulte Ingestão do Amazon Kinesis.

Apache Spark

O Apache Spark é um mecanismo de analytics unificado para processamento de dados em grande escala. Ele fornece alto nível APIs em Java, Scala, Python e R e um mecanismo otimizado que suporta gráficos de execução geral. Você pode usar o Apache Spark para criar aplicações de processamento de fluxos que consumam os dados nos fluxos de dados do Kinesis.

Para consumir os fluxos de dados do Kinesis usando o Apache Spark Structured Streaming, use o conectordo Amazon Kinesis Data Streams. Esse conector é compatível com o consumo com a distribuição avançada, o que fornece à sua aplicação throughput de leitura dedicada de até 2 MB de dados por segundo por fragmento. Para ter mais informações, consulte Desenvolver consumidores personalizados com throughput dedicado (distribuição avançada).

Para o consumo de fluxos de dados do Kinesis usando o Spark Streaming, consulte Spark Streaming + integração com Kinesis.

Databricks

O Databricks é uma plataforma baseada em nuvem que fornece um ambiente colaborativo para engenharia de dados, ciência de dados e machine learning. Para obter mais informações sobre o consumo de fluxos de dados do Kinesis usando Databricks, consulte Conectar ao Amazon Kinesis.

Kafka Confluent Platform

A Confluent Platform, construída com base no Kafka, fornece recursos e funcionalidades adicionais que ajudam as empresas a criar e gerenciar pipelines de dados e aplicações de fluxo em tempo real. Para obter mais informações sobre o consumo de fluxos de dados do Kinesis usando a Confluent Platform, consulte Amazon Kinesis Source Connector for Confluent Platform.

Kinesumer

O Kinesumer é um cliente para Go que implementa um cliente do grupo de consumidores distribuídos no lado do cliente para os fluxos de dados do Kinesis. Para obter mais informações, consulte o repositório do Kinesumer no GitHub.

Talend

O Talend é um software de integração e gerenciamento de dados que permite que os usuários coletem, transformem e conectem dados de várias fontes de maneira escalável e eficiente. Para obter mais informações sobre o consumo de fluxos de dados Kinesis usando o Talend, consulte Conectar o Talend a um fluxo Amazon Kinesis.

Solução de problemas de consumidores do Kinesis Data Streams

Os tópicos a seguir oferecem soluções para problemas comuns com consumidores do Amazon Kinesis Data Streams:

- Erro de compilação com o construtor LeaseManagementConfig
- Alguns registros do Kinesis Data Streams são ignorados quando a Kinesis Client Library é usada
- Registros pertencentes ao mesmo fragmento são processados por processadores de registros diferentes ao mesmo tempo
- O aplicativo consumidor está lendo a uma taxa menor que a esperada

- GetRecords retorna uma matriz de registros vazia mesmo quando há dados no fluxo
- Iterador do fragmento expira inesperadamente
- Processamento de registros de consumidores ficando atrasados
- Erro de permissão de chave do KMS não autorizada
- DynamoDbException: o caminho do documento fornecido na expressão de atualização é inválido para atualização
- Solucionar outros problemas comuns para consumidores

Erro de compilação com o construtor LeaseManagementConfig

Ao fazer o upgrade para a Kinesis Client Library (KCL) versão 3.x ou posterior, pode haver um erro de compilação relacionado ao construtor `LeaseManagementConfig`. Se você estiver criando diretamente um objeto `LeaseManagementConfig` para definir as configurações em vez de usar `ConfigsBuilder` nas versões 3.x ou posteriores da KCL, talvez a seguinte mensagem de erro seja exibida ao compilar o código da aplicação da KCL.

```
Cannot resolve constructor 'LeaseManagementConfig(String, DynamoDbAsyncClient,
 KinesisAsyncClient, String)'
```

A KCL com versões 3.x ou posteriores exige que você adicione mais um parâmetro, applicationName (type: String), após o parâmetro tableName.

- Antes: leaseManagementConfig = new LeaseManagementConfig (tableName, DBClient dynamo, kinesisClient, streamName, WorkerIdentifier)
- Depois de: leaseManagementConfig = new LeaseManagementConfig (tableName, applicationName, dynamo, kinesisClient, streamNameDBClient, WorkerIdentifier)

Em vez de criar diretamente um LeaseManagementConfig objeto, recomendamos usá-lo `ConfigsBuilder` para definir configurações no KCL 3.x e versões posteriores. `ConfigsBuilder`fornece uma maneira mais flexível e sustentável de configurar seu aplicativo KCL.

Veja a seguir um exemplo de uso do `ConfigsBuilder` para definir as configurações da KCL.

```
ConfigsBuilder configsBuilder = new ConfigsBuilder(
    streamName,
    applicationName,
```

```
    kinesisClient,
    dynamoClient,
    cloudWatchClient,
    UUID.randomUUID().toString(),
    new SampleRecordProcessorFactory()
);

Scheduler scheduler = new Scheduler(
    configsBuilder.checkpointConfig(),
    configsBuilder.coordinatorConfig(),
    configsBuilder.leaseManagementConfig()
    .failoverTimeMillis(60000), // this is an example
    configsBuilder.lifecycleConfig(),
    configsBuilder.metricsConfig(),
    configsBuilder.processorConfig(),
    configsBuilder.retrievalConfig()
);
```

Alguns registros do Kinesis Data Streams são ignorados quando a Kinesis Client Library é usada

A causa mais comum de registros ignorados é uma exceção não processada lançada de `processRecords`. A Kinesis Client Library (KCL) usa o `processRecords` para lidar com todas as exceções que ocorrem no processamento de registros de dados. Qualquer exceção lançada por `processRecords` é absorvida pela KCL. Para evitar infinitas novas tentativas no caso de uma falha recorrente, a KCL não reenvia o lote de registros processados no momento da exceção. A KCL chama `processRecords` para o próximo lote de registros de dados sem reiniciar o processador de registros. Isso resulta efetivamente em aplicativos consumidores observando registros ignorados. Para impedir registros ignorados, processe todas as exceções em `processRecords` da forma apropriada.

Registros pertencentes ao mesmo fragmento são processados por processadores de registros diferentes ao mesmo tempo

Para qualquer aplicação da Kinesis Client Library (KCL) em execução, cada fragmento tem apenas um proprietário. No entanto, vários processadores de registro pode temporariamente processar o mesmo fragmento. Se uma instância de operador perder a conectividade com a rede, a KCL pressupõe, após a expiração do tempo de failover, que o operador inacessível não está mais processando registros e instrui outras instâncias de operador a assumir o controle. Por um breve

período, novos processadores de registros e processadores de registros provenientes do operador inacessível podem processar dados do mesmo fragmento.

Defina um tempo de failover que seja apropriado para sua aplicação. Para aplicativos de baixa latência, o padrão de 10 segundos pode representar o tempo máximo que deseja esperar. No entanto, nos casos em que se esperam problemas de conectividade, como fazer chamadas em áreas geográficas em que a conectividade pode ser perdida com mais frequência, esse número pode ser muito baixo.

O aplicativo deve prever e lidar com esse cenário, especialmente porque a conectividade de rede normalmente é restaurada para o operador anteriormente inacessível. Se um processador de registros tem seus fragmentos executados por outro processador de registros, ele deve lidar com os dois casos seguintes para executar o encerramento normal:

1. Após a conclusão da chamada atual para `processRecords`, a KCL invoca o método de desligamento no processador de registros com o motivo de desligamento “ZOMBIE”. Espera-se que seus processadores de registros limpem quaisquer recursos conforme apropriado e, em seguida, saiam.
2. Ao tentar criar um ponto de verificação a partir de um operador “zombie”, a KCL lança `ShutdownException`. Depois de receber essa exceção, seu código deve sair do método atual de forma limpa.

Para obter mais informações, consulte Tratamento de registros duplicados.

O aplicativo consumidor está lendo a uma taxa menor que a esperada

Os motivos mais comuns para o throughput de leitura ser mais lento do que o esperado são os seguintes:

1. Vários aplicativos consumidores com um total de leituras que excedem os limites por fragmento. Para obter mais informações, consulte Cotas e limites. Nesse caso, aumente o número de fragmentos no fluxo de dados do Kinesis.
2. O limite que especifica o número máximo de GetRecords por chamada pode ter sido configurado com um valor baixo. Usando a KCL, é possível configurar o operador com um valor baixo para a propriedade `maxRecords`. Em geral, recomendamos o uso dos padrões do sistema para essa propriedade.
3. A lógica de sua `processRecords` chamada pode estar demorando mais do que o esperado por vários motivos possíveis; a lógica pode estar consumindo muita CPU, I/O bloqueando

ou congestionando a sincronização. Para testar se isso é verdadeiro, teste a execução de processadores de registros vazios e comparar o throughput de leitura. Para obter informações sobre como acompanhar os dados de entrada, consulte Use refragmentação, escalonamento e processamento paralelo para alterar o número de fragmentos.

Se houver apenas uma aplicação de consumo, sempre é possível ler pelo menos duas vezes mais rápido do que a taxa de colocação. É possível gravar até 1.000 registros por segundo, até uma taxa máxima total de gravação de dados de 1 MB por segundo (incluindo chaves de partição). Cada fragmento aberto oferece suporte a até cinco transações por segundo para leituras, até uma taxa máxima total de leitura de dados de 2 MB por segundo. Observe que cada leitura (chamada a GetRecords) obtém um lote de registros. O tamanho dos dados retornados pelo GetRecords varia de acordo com a utilização do fragmento. O tamanho máximo de dados que GetRecords pode retornar é 10 MB. Se uma chamada retorna esse limite, as chamadas subsequentes feitas nos próximos 5 segundos lançam `ProvisionedThroughputExceededException`.

GetRecords retorna uma matriz de registros vazia mesmo quando há dados no fluxo

O consumo, ou a obtenção, de registros é um modelo de envio. Espera-se que os desenvolvedores GetRecordsliguem em um loop contínuo, sem atrasos. Cada chamada a GetRecords também retorna um valor de `ShardIterator` que deve ser usado na próxima iteração do loop.

A operação GetRecords não bloqueia. Em vez disso, ela retorna imediatamente; com registros de dados relevantes ou com um elemento `Records` vazio. Um elemento `Records` vazio é retornado em duas condições:

1. Atualmente, não há mais dados no fragmento.
2. Não há dados perto da parte do fragmento indicada pelo `ShardIterator`.

A última condição é sutil, mas é uma característica de compensação necessária para evitar tempo de busca ilimitado (latência) ao recuperar registros. Desse modo, a aplicação de consumo de fluxo deve fazer um loop e chamar GetRecords tratando os registros vazios como de costume.

Em um cenário de produção, o único momento em que o loop contínuo deve ser encerrado é quando o valor de `NextShardIterator` é `NULL`. Quando `NextShardIterator` é `NULL`, significa que o fragmento foi fechado e o valor de `ShardIterator` apontaria para além do último registro. Se o

aplicativo consumidor nunca chamar SplitShard ou MergeShards, o fragmento permanecerá aberto e as chamadas a GetRecords nunca retornarão um valor de `NextShardIterator` que seja `NULL`.

Ao usar a Kinesis Client Library (KCL), o padrão de consumo anterior será abstraído. Isso inclui a manipulação automática de um conjunto de fragmentos que mudam dinamicamente. Com a KCL, o desenvolvedor fornece apenas a lógica para processar registros de entrada. Isso é possível porque a biblioteca faz chamadas contínuas a GetRecords.

Iterador do fragmento expira inesperadamente

Um novo iterador de fragmento é retornado por toda solicitação a GetRecords (como `NextShardIterator`), que é usado na próxima solicitação GetRecords (como `ShardIterator`). Normalmente, esse iterador do fragmento não expira antes de ser usado. No entanto, pode-se encontrar iteradores de fragmento que expiram por não chamar GetRecords por mais de cinco minutos, ou porque executar uma reinicialização da aplicação de consumo.

Se o iterador do fragmento expirar imediatamente, antes que seja possível usá-lo, isso poderá indicar que a tabela do DynamoDB usada pelo Kinesis não tem capacidade suficiente para armazenar os dados de concessões. Essa situação tem maior probabilidade de ocorrer se houver um grande número de fragmentos. Para solucionar esse problema, aumente a capacidade de gravação atribuída à tabela do fragmento. Para obter mais informações, consulte Usar uma tabela de concessões para monitorar os fragmentos processados pela aplicação de consumo da KCL.

Processamento de registros de consumidores ficando atrasados

Para a maioria dos casos de uso, as aplicações de consumo estão lendo os últimos dados do fluxo. Em determinadas circunstâncias, as leituras dos consumidores podem ficar atrasadas, o que pode não ser desejado. Depois de identificar a dimensão do atraso da leitura dos consumidores, veja os motivos mais comuns disso ocorrer.

Comece com a métrica `GetRecords.IteratorAgeMilliseconds`, que rastreia a posição de leitura em todos os fragmentos e aplicações de consumo no fluxo. Observe que, se a idade de um iterador passar de 50% do período de retenção (24 horas por padrão, configurável até 365 dias), haverá risco de perda de dados devido à expiração de registro. Uma solução provisória rápida é aumentar o período de retenção. Isso interrompe a perda de dados importantes enquanto o problema é solucionado. Para obter mais informações, consulte Monitore o serviço Amazon Kinesis Data Streams com a Amazon CloudWatch. Em seguida, identifique o quão atrasado seu aplicativo consumidor está lendo cada fragmento usando uma CloudWatch métrica personalizada emitida

pela Kinesis Client Library (KCL),. `MillisBehindLatest` Para obter mais informações, consulte Monitore a biblioteca de cliente do Kinesis com a Amazon CloudWatch.

Veja os motivos mais comuns para consumidores ficarem atrasados:

- Grandes aumentos súbitos em `GetRecords.IteratorAgeMilliseconds` ou `MillisBehindLatest` geralmente indicam um problema transitório, como falhas de operação da API para um aplicativo downstream. Investigue esses aumentos repentinos se uma das métricas exibir esse comportamento constantemente.
- Um aumento gradual nessas métricas indica que um consumidor não está acompanhando o fluxo porque não está processando registros com a rapidez necessária. As causas raiz mais comuns para esse comportamento são recursos físicos insuficientes ou lógica de processamento de registros que não escalou com o aumento no throughput do fluxo. Você pode verificar esse comportamento observando as outras CloudWatch métricas personalizadas que o KCL emite associadas à `processTask` operação, incluindo `RecordProcessor.processRecords.TimeSuccess`, e. `RecordsProcessed`
  - Em caso de aumento na métrica `processRecords.Time` correlacionada ao aumento no throughput, deve-se analisar a lógica do processamento de registros para identificar por que ela não está se dimensionando de acordo com o maior throughput.
  - Caso veja um aumento nos valores de `processRecords.Time` que não esteja correlacionado com aumento no throughput, verifique se estão sendo feitas chamadas de bloqueio no caminho crítico, que muitas vezes são a causa de lentidão no processamento de registros. Uma abordagem alternativa é aumentar o paralelismo, aumentando o número de fragmentos. Finalmente, confirme se há uma quantidade adequada de recursos físicos (memória, utilização da CPU, etc.) nos nós de processamento subjacentes durante o pico de demanda.

Erro de permissão de chave do KMS não autorizada

Esse erro ocorre quando um aplicativo consumidor lê um fluxo criptografado sem permissões na AWS KMS chave. Para atribuir permissões a uma aplicação para que acesse uma chave do KMS, consulte Using Key Policies in AWS KMS e Using IAM Policies with AWS KMS.

DynamoDbException: o caminho do documento fornecido na expressão de atualização é inválido para atualização

Ao usar o KCL 3.x com AWS SDK para Java as versões 2.27.19 a 2.27.23, você pode encontrar a seguinte exceção do DynamoDB:

“software.amazon.awssdk.services.dynamodb.model. DynamoDbException: o caminho do documento fornecido na expressão de atualização é inválido para atualização (Serviço: DynamoDb, Código de status: 400, ID da solicitação: xxx)”

Esse erro ocorre devido a um problema conhecido AWS SDK para Java que afeta a tabela de metadados do DynamoDB gerenciada pelo KCL 3.x. O problema foi introduzido na versão 2.27.19 e afeta todas as versões até a 2.27.23. O problema foi resolvido no AWS SDK para Java versão 2.27.24. Para obter desempenho e estabilidade ideais, recomendamos a atualização para a versão 2.28.0 ou posterior.

Solucionar outros problemas comuns para consumidores

- Por que o gatilho do Kinesis Data Streams não consegue invocar minha função do Lambda?
- Como detecto e soluciono problemas de ReadProvisionedThroughputExceeded exceções no Kinesis Data Streams?
- Why am I experiencing high latency issues with Kinesis Data Streams?
- Why is my Kinesis data stream returning a 500 Internal Server Error?
- How do I troubleshoot a blocked or stuck KCL application for Kinesis Data Streams?
- Can I use different Amazon Kinesis Client Library applications with the same Amazon DynamoDB table?

Otimizar consumidores do Amazon Kinesis Data Streams

Você pode otimizar ainda mais seu consumidor do Amazon Kinesis Data Streams com base no comportamento específico observado.

Leia os seguintes tópicos para identificar soluções.

Tópicos

- Melhorar o processamento de baixa latência
- Processe dados serializados usando AWS Lambda com a Amazon Kinesis Producer Library
- Use refragmentação, escalonamento e processamento paralelo para alterar o número de fragmentos
- Tratamento de registros duplicados
- Tratamento do startup, do desligamento e do controle de utilização

Melhorar o processamento de baixa latência

O atraso de propagação é definido como a end-to-end latência do momento em que um registro é gravado no stream até ser lido por um aplicativo consumidor. Esse atraso varia dependendo de uma série de fatores, mas é afetado principalmente pelo intervalo de sondagem de aplicativos de consumidor.

Para a maioria dos aplicativos, recomendamos a sondagem de cada fragmento uma vez por segundo por aplicativo. Isso permite várias aplicações de consumo processando um fluxo simultaneamente sem atingir os limites do Amazon Kinesis Data Streams de cinco chamadas de `GetRecords` por segundo. Além disso, o processamento de lotes de dados maiores tende a ser mais eficiente na redução da latência de rede e outras latências de downstream no seu aplicativo.

Os padrões da KCL estão definidos para seguir a prática recomendada de sondagem a cada segundo. Esse padrão gera atrasos de propagação médios que costumam ser menores que 1 segundo.

Os registros do Kinesis Data Streams ficam disponíveis para serem lidos imediatamente após serem gravados. Há alguns casos de uso que precisam aproveitar isso e exigir o consumo de dados do fluxo assim que ele estiver disponível. É possível reduzir significativamente o atraso de propagação sobrepondo as configurações padrão da KCL para sondar com mais frequência, como mostrado nos exemplos a seguir.

Código de configuração Java da KCL:

```
kinesisClientLibConfiguration = new
        KinesisClientLibConfiguration(applicationName,
        streamName,
        credentialsProvider,

 workerId).withInitialPositionInStream(initialPositionInStream).withIdleTimeBetweenReadsInMilli
```

Configuração de arquivo de propriedades da KCL para Python e Ruby:

```
idleTimeBetweenReadsInMillis = 250
```

> **Note**
>
> Como o Kinesis Data Streams tem um limite de cinco chamadas de `GetRecords` por segundo por fragmento, configurar a propriedade `idleTimeBetweenReadsInMillis`

abaixo de 200 ms pode fazer com que a aplicação receba a exceção `ProvisionedThroughputExceededException`. Muitas dessas exceções podem gerar esperas exponenciais significativas e, portanto, causar latências inesperadas significativas no processamento. Ao definir essa propriedade em 200 ms ou acima e houver mais de um aplicativo de processamento, ocorrerá um controle de utilização semelhante.

Processe dados serializados usando AWS Lambda com a Amazon Kinesis Producer Library

A Amazon Kinesis Producer Library (KPL) agrega pequenos registros formatados pelo usuário em registros maiores de até 1 MB para usar melhor o throughput do Amazon Kinesis Data Streams. Embora o KCL para Java ofereça suporte à desagregação desses registros, você precisa usar um módulo especial para desagregar registros ao usá-lo AWS Lambda como consumidor de seus streams. Você pode obter o código e as instruções do projeto necessários GitHub nos Módulos de desagregação da biblioteca de produtores do Amazon Kinesis para Lambda. AWS Os componentes deste projeto oferecem a capacidade de processar dados serializados KPL em Java AWS Lambda, Node.js e Python. Esses componentes também podem ser usados como parte de um aplicativo de várias linguagens da KCL.

Use refragmentação, escalonamento e processamento paralelo para alterar o número de fragmentos

A refragmentação permite aumentar ou diminuir o número de fragmentos em um fluxo para se adaptar às alterações na taxa de dados que fluem pelo fluxo. A refragmentação costuma ser realizado por um aplicativo administrativo que monitora as métricas de tratamento de dados de fragmento. Embora não inicie operações de refragmentação, a KCL é projetada para se adaptar às alterações no número de fragmentos resultantes desse processo.

Conforme observado em Usar uma tabela de concessões para monitorar os fragmentos processados pela aplicação de consumo da KCL, a KCL rastreia os fragmentos no fluxo usando uma tabela do Amazon DynamoDB. Quando novos fragmentos são criados em consequência da refragmentação, a KCL descobre os novos fragmentos e preenche novas linhas na tabela. Os operadores descobrem automaticamente os novos fragmentos e criam processadores para tratar os dados provenientes deles. A KCL também distribui os fragmentos no fluxo entre todos os operadores e processadores de registros disponíveis.

A KCL garante que os dados existentes nos fragmentos antes do refragmentação sejam processados primeiro. Depois do processamento dos dados, os dados dos novos fragmentos são enviados para processadores de registros. Dessa forma, a KCL preserva a ordem em que os registros de dados foram adicionados ao fluxo para uma determinada chave de partição.

Exemplo: refragmentação, escalonamento e processamento paralelo

O exemplo a seguir ilustra como a KCL ajuda a lidar com escalabilidade e refragmentação:

- Por exemplo, a aplicação está sendo executada em uma instância do EC2 e processa um fluxo de dados do Kineses com quatro fragmentos. Essa instância tem um operador da KCL e quatro processadores de registros (um processador de registros para cada fragmento). Esses quatro processadores de registros são executados em paralelo no mesmo processo.
- Em seguida, ao escalar o aplicativo para usar outra instância, duas instâncias processarão um único fluxo que tem quatro fragmentos. Ao ser iniciado na segunda instância, o operador da KCL faz balanceamento de carga com a primeira instância, de modo que cada instância passa a processar dois fragmentos.
- Se decidir dividir os quatro fragmentos em cinco, a KCL coordenará novamente o processamento entre as instâncias: uma instância processará três fragmentos e a outra processará dois fragmentos. Uma coordenação semelhante ocorre ao mesclar fragmentos.

Normalmente, ao usar a KCL, é necessário garantir que o número de instâncias não exceda o número de fragmentos (exceto para fins de espera de falha). Cada fragmento é processado por exatamente um operador da KCL e tem exatamente um processador de registros correspondente, para nunca necessitar de várias instâncias para processar apenas um fragmento. Mas como um operador pode processar qualquer número de fragmentos, tudo bem se o número de fragmentos ultrapassar o número de instâncias.

Para expandir o processamento do aplicativo, deve-se testar uma combinação destas abordagens:

- Aumentar o tamanho da instância (porque todos os processadores de registros são executados em paralelo em um processo)
- Aumentar o número de instâncias até o número máximo de fragmentos abertos (porque os fragmentos podem ser processados de forma independente)
- Aumentar o número de fragmentos (o que aumenta o nível de paralelismo)

Observe que é possível usar o ajuste de escala automático para escalar automaticamente as instâncias com base em métricas apropriadas. Para obter mais informações, consulte o Guia do usuário do Amazon EC2 Auto Scaling.

Quando a refragmentação aumenta o número de fragmentos no fluxo, o aumento correspondente no número de processadores de registros aumenta a carga nas instâncias do EC2 que os hospedam. Se as instâncias fazem parte de um grupo de Auto Scaling e a carga aumenta suficientemente, o grupo de Auto Scaling adiciona mais instâncias para lidar com o aumento de carga. É necessário configurar as instâncias para iniciar a aplicação do Amazon Kinesis Data Streams no startup, para que os operadores e processadores de registros adicionais se tornem imediatamente ativos na nova instância.

Para obter mais informações sobre a refragmentação, consulte Refragmentar um fluxo.

Tratamento de registros duplicados

Há dois principais motivos pelos quais os registros podem ser entregues mais de uma vez à aplicação do Amazon Kinesis Data Streams: novas tentativas de produtor e novas tentativas de consumidor. Seu aplicativo precisa prever e tratar devidamente o processamento de registros individuais várias vezes.

Novas tentativas do produtor

Considere um produtor que tem um tempo limite esgotado de rede depois de fazer uma chamada a `PutRecord`, mas antes de poder receber uma confirmação do Amazon Kinesis Data Streams. O produtor não tem certeza de que o registro foi entregue ao Kinesis Data Streams. Supondo que cada registro é importante para o aplicativo, o produtor teria sido concebido de modo a tentar novamente a chamada com os mesmos dados. Se as duas chamadas a `PutRecord` para os mesmos dados foram confirmadas com sucesso no Kinesis Data Streams, haverá dois registros do Kinesis Data Streams. Embora os dois registros tenham dados idênticos, eles também têm números sequenciais exclusivos. Os aplicativos que precisam de rigorosas garantias devem incorporar uma chave primária ao registro para remover duplicatas mais tarde ao processar. Observe que o número de duplicatas resultante das retentativas de produtor costuma ser baixo em comparação com o número de duplicatas resultante das retentativas de consumidor.

> **Note**
>
> Se você usa o AWS SDK`PutRecord`, saiba mais sobre o comportamento de repetição do SDK no guia do usuário AWS SDKs and Tools.

Novas tentativas do consumidor

As retentativas de consumidor (aplicativo de processamento de dados) acontecem quando os processadores de registros são reiniciados. Os processadores de registros do mesmo fragmento reiniciam nos seguintes casos:

1. Um operador é encerrado inesperadamente
2. Instâncias de operador são adicionadas ou removidas
3. Os fragmentos são mesclados ou divididos
4. O aplicativo é implantado

Em todos esses casos, o mapeamento shards-to-worker-to -record-processor é atualizado continuamente para balancear a carga do processamento. Processadores de fragmentos que foram migrados para outras instâncias reiniciam o processamento de registros a partir do último ponto de verificação. Isso acarreta processamento de registros duplicado, conforme mostrado no exemplo abaixo. Para obter mais informações sobre balanceamento de carga, consulte Use refragmentação, escalonamento e processamento paralelo para alterar o número de fragmentos.

Exemplo: novas tentativas de consumidor gerando nova entrega de registros

Neste exemplo, há uma aplicação que lê continuamente registros de um fluxo, agrega registros em um arquivo local e faz upload do arquivo para o Amazon S3. Para simplificar, suponha que haja apenas 1 fragmento e 1 operador processando o fragmento. Considere a sequência de eventos de exemplo a seguir, supondo que o último ponto de verificação ocorreu no registro de número 10.000:

1. Um operador lê o próximo lote de registros a partir do fragmento, registros de 10.001 a 20.000.
2. O operador passa o lote de registros para o processador de registros associado.
3. O processador de registros agrega os dados, cria um arquivo do Amazon S3 e faz upload do arquivo para o Amazon S3 com êxito.
4. O operador é encerrado inesperadamente antes que ocorra um novo ponto de verificação.
5. O aplicativo, o operador e o processador de registros são reiniciados.
6. O operador agora começa a ler a partir do último ponto de verificação bem-sucedido, que neste caso é 10.001.

Desse modo, os registros de 10.001 a 20.000 são consumidos mais de uma vez.

Ser resiliente a novas tentativas de consumidor

Mesmo que os registros possam ser processados mais de uma vez, seu aplicativo pode apresentar efeitos colaterais como se os registros tivessem sido processados apenas uma vez (processamento idempotente). As soluções para esse problema variam em termos de complexidade e precisão. Se o destino dos dados finais puder tratar bem das duplicatas, recomendamos confiar no destino final para obter o processamento idempotente. Por exemplo, com o Opensearch, você pode usar uma combinação de controle de versão e exclusividade IDs para evitar o processamento duplicado.

A aplicação de exemplo da seção anterior lê continuamente os registros de um fluxo, agrega os registros em um arquivo local e faz upload do arquivo para o Amazon S3. Conforme ilustrado, os registros de 10.001 a 20.000 são consumidos mais de uma vez, resultando em vários arquivos do Amazon S3 com os mesmos dados. Uma forma de reduzir as duplicatas desse exemplo é garantir que a etapa 3 use o seguinte esquema:

1. O processador de registros usa um número fixo de registros por arquivo do Amazon S3, como 5.000.
2. O nome do arquivo usa este esquema: prefixo do Amazon S3, shard-id e `First-Sequence-Num`. Neste caso, pode ser algo como `sample-shard000001-10001`.
3. Depois de fazer upload do arquivo do Amazon S3, defina o ponto de verificação especificando `Last-Sequence-Num`. Neste caso, o ponto de verificação seria definido no registro de número 15.000.

Com esse esquema, mesmo que os registros sejam processados mais de uma vez, o arquivo do Amazon S3 resultante terá o mesmo nome e os mesmos dados. As retentativas geram apenas a gravação dos mesmos dados no mesmo arquivo mais de uma vez.

No caso de uma operação de refragmentação, o número de registros deixados no fragmento pode ser menor que o número fixo necessário. Nesse caso, o método `shutdown()` precisa descarregar o arquivo no Amazon S3 e definir o ponto de verificação no último número de sequência. O esquema acima também é compatível com as operações de refragmentação.

Tratamento do startup, do desligamento e do controle de utilização

Leia aqui algumas considerações adicionais para incorporar ao projeto da aplicação do Amazon Kinesis Data Streams.

Tópicos

- Inicializar produtores de dados e consumidores de dados
- Encerre um aplicativo Amazon Kinesis Data Streams
- Controle de utilização de leitura

Inicializar produtores de dados e consumidores de dados

Por padrão, a KCL começa a ler registros pela extremidade do fluxo, que é o registro adicionado mais recentemente. Nessa configuração, se um aplicativo de produção de dados adicionar registros ao fluxo antes da execução de qualquer processador de registros de recebimento, os registros não serão lidos pelos processadores de registros após a inicialização.

Para alterar o comportamento dos processadores de registros para que eles sempre leiam dados a partir do início do fluxo, defina o seguinte valor no arquivo de propriedades da aplicação do Amazon Kinesis Data Streams:

```
initialPositionInStream = TRIM_HORIZON
```

Por padrão, o Amazon Kinesis Data Streams armazena todos os dados por 24 horas. Ele também oferece suporte à retenção prolongada de até 7 dias e a retenção de longo prazo de até 365 dias. Esse período é chamado de período de retenção. Definir a posição de início como `TRIM_HORIZON` inicia o processador de registros com os dados mais antigos no fluxo, conforme definido pelo período de retenção. Mesmo com a definição de `TRIM_HORIZON`, se um processador de registros iniciar após decorrido um tempo maior que o período de retenção, alguns dos registros no fluxo não estarão mais disponíveis. Por esse motivo, você deve sempre ter aplicativos de consumo lendo o stream e usando a CloudWatch métrica `GetRecords.IteratorAgeMilliseconds` para monitorar se os aplicativos estão acompanhando os dados recebidos.

Em alguns cenários, pode ser adequado aos processadores de registros perder os primeiros registros no fluxo. Por exemplo, você pode executar alguns registros iniciais no stream para testar se o stream está funcionando end-to-end conforme o esperado. Depois de fazer essa verificação inicial, inicie seus operadores e comece a colocar os dados de produção no fluxo.

Para obter mais informações sobre a configuração de `TRIM_HORIZON`, consulte Como usar iteradores de fragmentos.

Encerre um aplicativo Amazon Kinesis Data Streams

Quando a aplicação do Amazon Kinesis Data Streams tiver concluído a tarefa pretendida, é necessário desligá-la encerrando as instâncias do EC2 nas quais ela está em execução. É possível encerrar as instâncias usando o Console de gerenciamento da AWS ou a AWS CLI.

Depois de desligar a aplicação do Amazon Kinesis Data Streams, deve-se excluir a tabela do Amazon DynamoDB que a KCL usou para rastrear o estado da aplicação.

Controle de utilização de leitura

O throughput de um fluxo é provisionado no nível do fragmento. Cada fragmento tem um throughput de leitura de até cinco transações por segundo para leituras, até uma taxa máxima total de leitura de dados de 2 MB por segundo. Se uma aplicação (ou grupo de aplicações operando no mesmo fluxo) tentar obter dados de um fragmento a uma taxa mais rápida, o Kinesis Data Streams limitará as operações Get correspondentes.

Em uma aplicação do Amazon Kinesis Data Streams, se um processador de registros processar dados mais rapidamente do que o limite, como no caso de um failover, o controle de utilização ocorrerá. Como a KCL gerencia as interações entre a aplicação e o Kinesis Data Streams, as exceções de controle de utilização ocorrem no código da KCL, e não no código da aplicação. No entanto, como a KCL registra essas exceções, elas podem ser vistas nos logs.

Se acreditar que o aplicativo fica limitado de forma consistente, considere aumentar o número de fragmentos do fluxo.

Monitorar o Kinesis Data Streams

É possível monitorar os fluxos de dados no Amazon Kinesis Data Streams usando os seguintes recursos:

- CloudWatch métricas — O Kinesis Data Streams CloudWatch envia métricas personalizadas à Amazon com monitoramento detalhado para cada stream.
- Kinesis Agent — O Kinesis Agent publica CloudWatch métricas personalizadas para ajudar a avaliar se o agente está funcionando conforme o esperado.
- Registro em log da API: o Kinesis Data Streams usa o AWS CloudTrail para registrar chamadas à API em log e armazenar os dados em um bucket do Amazon S3.
- Kinesis Client Library: a Kinesis Client Library (KCL) fornece métricas por fragmento, operador e aplicação da KCL.
- Kinesis Producer Library: a Amazon Kinesis Producer Library (KPL) fornece métricas por fragmento, operador e aplicação da KPL.

Para obter mais informações sobre problemas comuns de monitoramento, perguntas e solução de problemas, consulte o seguinte:

- Quais métricas devem ser usadas para monitorar e solucionar problemas com o Kinesis Data Streams?
- Por que o IteratorAgeMilliseconds valor do Kinesis Data Streams continua aumentando?

Monitore o serviço Amazon Kinesis Data Streams com a Amazon CloudWatch

O Amazon Kinesis Data Streams CloudWatch e o Amazon estão integrados para que você possa coletar CloudWatch, visualizar e analisar métricas para seus streams de dados do Kinesis. Por exemplo, para rastrear o uso de fragmentos, é possível monitorar as `IncomingBytes` e as métricas de `OutgoingBytes` e compará-las com o número de fragmentos no fluxo.

As métricas de stream e as métricas em nível de fragmento que você configura são coletadas automaticamente e enviadas a CloudWatch cada minuto. As métricas são arquivadas por duas semanas. Depois desse período, os dados serão descartados.

A tabela a seguir descreve o monitoramento básico no nível do fluxo e o monitoramento avançado no nível do fragmento em fluxos de dados do Kinesis.

Tipo	Description
Básico (nível de fluxo)	Dados de nível de fluxo são enviados automaticamente a cada minuto, sem custo adicional.
Avançado (nível de fragmento)	Dados de nível de fragmento são enviados a cada minuto por um custo adicional. Para obter esse nível de dados, você precisa habilitá-lo especificamente para o stream usando a operação EnableEnhancedMonitoring. Para obter informações sobre preços, consulte a página CloudWatch do produto Amazon.

Métricas e dimensões do Amazon Kinesis Data Streams

O Kinesis Data Streams envia CloudWatch métricas para dois níveis: o nível do stream e, opcionalmente, o nível do fragmento. As métricas no nível do fluxo se destinam aos casos de uso de monitoramento mais comuns em condições normais. As métricas em nível de estilhaço são para tarefas de monitoramento específicas, geralmente relacionadas à solução de problemas, e são habilitadas com o uso da operação EnableEnhancedMonitoring.

Para obter uma explicação das estatísticas coletadas a partir das CloudWatch métricas, consulte CloudWatch Estatísticas no Guia CloudWatch do Usuário da Amazon.

Tópicos

- Métricas no nível do fluxo básicas
- Métricas do nível de fragmento aprimoradas
- Dimensões de métricas do Amazon Kinesis Data Streams
- Métricas recomendadas do Amazon Kinesis Data Streams

Métricas no nível do fluxo básicas

O namespace `AWS/Kinesis` inclui métricas de nível do fluxo a seguir.

O Kinesis Data Streams envia essas métricas em nível de stream CloudWatch a cada minuto. Essas métricas estão sempre disponíveis.

Métrica	Description
`GetRecords.Bytes`	O número de bytes recuperados do fluxo do Kinesis, medido no período especificado. As estatísticas mínima, máxima e média representam os bytes em uma única operação `GetRecords` para o fluxo no período especificado. Nome da métrica do nível de fragmento: `OutgoingBytes` Dimensões: StreamName Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: bytes
`GetRecords.IteratorAge`	Essa métrica não é mais usada. Use `GetRecords.IteratorAgeMilliseconds` .
`GetRecords.IteratorAgeMilliseconds`	A idade do último registro em todas as chamadas `GetRecords` feitas com relação a um fluxo do Kinesis, medida no período especificado. Idade é a diferença entre a hora atual e quando o último registro da chamada `GetRecords` foi gravado no fluxo. As estatísticas de mínimo e máximo podem ser usadas para acompanhar o progresso das aplicações de consumo do Kinesis. Um valor zero indica que os registros lidos estão em completa sincronização com o fluxo. Nome da métrica do nível de fragmento: `IteratorAgeMilliseconds`

Métrica	Description
	Dimensões: StreamName Estatísticas válidas: mínima, máxima, média, amostras Unidade: milissegundos
`GetRecords.Latency`	O tempo gasto por operação `GetRecords` , medido ao longo do período de tempo especificado. Dimensões: StreamName Estatísticas: mínimo, máximo, média Unidade: milissegundos
`GetRecords.Records`	O número de registros recuperados do fragmento, medido ao longo do período de tempo especificado. As estatísticas mínima, máxima e média representam os registros em uma única operação `GetRecords` para o fluxo no período especificado. Nome da métrica do nível de fragmento: `OutgoingRecords` Dimensões: StreamName Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: contagem

Métrica	Description
`GetRecords.Success`	O número de operações `GetRecords` bem-sucedidas por fluxo, medido ao longo do período de tempo especificado. Dimensões: StreamName Estatísticas válidas: média, soma, amostras Unidades: contagem
`IncomingBytes`	O número de bytes colocados com êxito no fluxo do Kinesis no período especificado. Essa métrica inclui bytes das operações `PutRecord` e `PutRecords`. As estatísticas mínima, máxima e média representam os bytes em uma única operação put para o fluxo no período especificado. Nome da métrica do nível de fragmento: `IncomingBytes` Dimensões: StreamName Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: bytes

Métrica	Description
IncomingRecords	O número de registros colocados com êxito no fluxo do Kinesis no período especificado. Essa métrica inclui registros das operações PutRecord e PutRecords . As estatísticas mínima, máxima e média representam os registros em uma única operação put para o fluxo no período especificado. Nome da métrica do nível de fragmento: IncomingRecords Dimensões: StreamName Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: contagem
PutRecord.Bytes	O número de bytes colocados no fluxo do Kinesis usando a operação PutRecord no período especificado. Dimensões: StreamName Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: bytes
PutRecord.Latency	O tempo gasto por operação PutRecord , medido ao longo do período de tempo especificado. Dimensões: StreamName Estatísticas: mínimo, máximo, média Unidade: milissegundos

Métrica	Description
PutRecord.Success	O número de operações PutRecord bem-sucedidas por fluxo do Kinesis, medido no período especificado. A média reflete a porcentagem de gravações bem-suced idas em um fluxo. Dimensões: StreamName Estatísticas válidas: média, soma, amostras Unidades: contagem
PutRecords.Bytes	O número de bytes colocados no fluxo do Kinesis usando a operação PutRecords no período especificado. Dimensões: StreamName Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: bytes
PutRecords.Latency	O tempo gasto por operação PutRecords , medido ao longo do período de tempo especificado. Dimensões: StreamName Estatísticas: mínimo, máximo, média Unidade: milissegundos
PutRecords.Records	Essa métrica foi substituída. Use PutRecord s.SuccessfulRecords . Dimensões: StreamName Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: contagem

Métrica	Description
PutRecords.Success	O número de operações PutRecords com pelo menos um registro bem-sucedido por fluxo do Kinesis, medido no período especificado. Dimensões: StreamName Estatísticas válidas: média, soma, amostras Unidades: contagem
PutRecords.TotalRecords	O número total de registros enviados em uma operação PutRecords por fluxo de dados do Kinesis, medido no período especificado. Dimensões: StreamName Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: contagem
PutRecords.SuccessfulRecords	O número de registros bem-sucedidos em uma operação PutRecords por fluxo de dados do Kinesis, medido no período especificado. Dimensões: StreamName Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: contagem

Métrica	Description
PutRecords.FailedRecords	O número de registros rejeitados devido a falhas internas em uma operação PutRecords por fluxo de dados do Kinesis, medido no período especificado. Falhas internas ocasionais são esperadas e a operação deve ser repetida. Dimensões: StreamName Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: contagem
PutRecords.ThrottledRecords	O número de registros rejeitados devido ao controle de utilização em uma operação PutRecords por fluxo de dados do Kinesis, medido no período especificado. Dimensões: StreamName Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: contagem

Métrica	Description
`ReadProvisionedThroughputExceeded`	O número de chamadas `GetRecords` limitadas para o fluxo ao longo do período especificado. A estatística mais usada para essa métrica é Média. Quando a estatística mínima tem um valor de 1, todos os registros foram limitados ao fluxo durante o período especificado. Quando a estatística máxima tem um valor de 0 (zero), nenhum registro foi limitado ao fluxo durante o período especificado. Nome da métrica do nível de fragmento: `ReadProvisionedThroughputExceeded` Dimensões: StreamName Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: contagem
`SubscribeToShard.RateExceeded`	Essa métrica é emitida quando uma tentativa de nova assinatura apresenta falha porque já existe uma assinatura ativa com o mesmo consumidor ou se o número de chamadas por segundo permitido para essa operação for excedido. Dimensões: StreamName, ConsumerName
`SubscribeToShard.Success`	Essa métrica registra se a SubscribeToShard assinatura foi estabelecida com sucesso. A assinatura se mantém ativa por no máximo 5 minutos. Portanto, essa métrica é emitida pelo menos uma vez a cada 5 minutos. Dimensões: StreamName, ConsumerName

Métrica	Description
`SubscribeToShardEvent.Bytes`	O número de bytes recebidos do fragmento, medidos no período especificado. As estatísticas mínima, máxima e média representam os bytes publicados em um único evento no período especificado. Nome da métrica do nível de fragmento: `OutgoingBytes` Dimensões: StreamName, ConsumerName Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: bytes
`SubscribeToShardEvent.MillisBehindLatest`	O número de milissegundos em que os registros de leitura estão na extremidade do fluxo de dados, indicando o quão atrasado o consumidor está em relação ao horário atual. Dimensões: StreamName, ConsumerName Estatísticas válidas: mínima, máxima, média, amostras Unidade: milissegundos

Métrica	Description
`SubscribeToShardEvent.Records`	O número de registros recebidos do fragmento, medidos no período especificado. As estatísticas mínima, máxima e média representam os registros em um único evento no período especificado. Nome da métrica do nível de fragmento: `OutgoingRecords` Dimensões: StreamName, ConsumerName Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: contagem
`SubscribeToShardEvent.Success`	Essa métrica é emitida sempre que um evento é publicado com êxito. Ela será emitida somente quando houver uma assinatura ativa. Dimensões: StreamName, ConsumerName Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: contagem

Métrica	Description
`WriteProvisionedThroughputExceeded`	O número de registros rejeitados por causa da limitação para o fluxo ao longo do período especificado. Essa métrica inclui a limitação das operações `PutRecord` e `PutRecords` . A estatística mais usada para essa métrica é Média. Quando a estatística mínima tem um valor diferente de zero, nenhum registro é limitado ao fluxo durante o período especificado. Quando a estatística máxima tem um valor de 0 (zero), nenhum registro foi limitado ao fluxo durante o período especificado. Nome da métrica do nível de fragmento: `WriteProvisionedThroughputExceeded` Dimensões: StreamName Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: contagem

Métricas do nível de fragmento aprimoradas

O namespace `AWS/Kinesis` inclui métricas de nível do fragmento a seguir.

O Kinesis envia as seguintes métricas em nível de fragmento a cada minuto. CloudWatch Cada dimensão métrica cria 1 CloudWatch métrica e faz aproximadamente 43.200 chamadas de `PutMetricData` API por mês. Essas métricas não são permitidas por padrão. Há uma cobrança para as métricas aprimoradas emitidas pelo Kinesis. Para obter mais informações, consulte Amazon CloudWatch Pricing sob o título Amazon CloudWatch Custom Metrics. As cobranças são indicadas por fragmento pela métrica por mês.

Métrica	Description
IncomingBytes	O número de bytes colocados com sucesso no fragmento ao longo do período especificado. Essa métrica inclui bytes das operações PutRecord e PutRecords . As estatísticas mínima, máxima e média representam os bytes em uma única operação put para o fragmento no período especificado. Nome da métrica no nível do fluxo: IncomingBytes Dimensões: StreamName, ShardId Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: bytes
IncomingRecords	O número de registros colocados com sucesso no fragmento ao longo do período especificado. Essa métrica inclui registros das operações PutRecord e PutRecords . As estatísticas mínima, máxima e média representam os registros em uma única operação put para o fragmento no período especificado. Nome da métrica no nível do fluxo: IncomingRecords Dimensões: StreamName, ShardId Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: contagem
IteratorAgeMilliseconds	A idade do último registro em todas as chamadas GetRecords feitas com relação a um fragmento, medida ao longo do período de tempo especificado. Idade é a diferença entre a hora atual e quando o último registro da chamada GetRecords foi gravado no

Métrica	Description
	fluxo. As estatísticas de mínimo e máximo podem ser usadas para acompanhar o progresso das aplicações de consumo do Kinesis. Um valor de 0 (zero) indica que os registros lidos estão em completa sincronização com o fluxo. Nome da métrica no nível do fluxo: `GetRecords.IteratorAgeMilliseconds` Dimensões: StreamName, ShardId Estatísticas válidas: mínima, máxima, média, amostras Unidade: milissegundos
`OutgoingBytes`	O número de bytes recuperados do fragmento, medido ao longo do período de tempo especificado. As estatísticas mínima, máxima e média representam os bytes retornados em uma única operação `GetRecords` ou publicados em um único evento `SubscribeToShard` para o fragmento no período especificado. Nome da métrica no nível do fluxo: `GetRecords.Bytes` Dimensões: StreamName, ShardId Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: bytes

Métrica	Description
`OutgoingRecords`	O número de registros recuperados do fragmento, medido ao longo do período de tempo especificado. As estatísticas mínima, máxima e média representam os registros retornados em uma única operação `GetRecords` ou publicados em um único evento `SubscribeToShard` para o fragmento no período especificado. Nome da métrica no nível do fluxo: `GetRecords.Records` Dimensões: StreamName, ShardId Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: contagem

Métrica	Description
`ReadProvisionedThroughputExceeded`	O número de chamadas `GetRecords` limitadas para o fragmento ao longo do período especificado. Esta contagem de exceções abrange todas as dimensões dos seguintes limites: 5 leituras por fragmento por segundo ou 2 MB por segundo por fragmento. A estatística mais usada para essa métrica é Média. Quando a estatística mínima tem um valor de 1, todos os registros foram limitados ao fragmento durante o período especificado. Quando a estatística máxima tem um valor de 0 (zero), nenhum registro foi limitado ao fragmento durante o período especificado. Nome da métrica no nível do fluxo: `ReadProvisionedThroughputExceeded` Dimensões: StreamName, ShardId Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: contagem

Métrica	Description
`WriteProvisionedThroughputExceeded`	O número de registros rejeitados por causa da limitação para o fragmento ao longo do período especificado. Esta métrica inclui limitação das operações `PutRecord` e `PutRecords` e abrange todas as dimensões os seguintes limites: 1.000 registros por segundo por fragmento ou 1 MB por segundo por fragmento. A estatística mais usada para essa métrica é Média. Quando a estatística mínima tem um valor diferente de zero, nenhum registro é limitado ao fragmento durante o período especificado. Quando a estatística máxima tem um valor de 0 (zero), nenhum registro foi limitado ao fragmento durante o período especificado. Nome da métrica no nível do fluxo: `WriteProvisionedThroughputExceeded` Dimensões: StreamName, ShardId Estatísticas válidas: mínima, máxima, média, soma, amostras Unidades: contagem

Dimensões de métricas do Amazon Kinesis Data Streams

Dimensão	Description
`StreamName`	O nome do fluxo do Kinesis. Todas as estatísticas disponíveis são filtradas por `StreamName` .

Métricas recomendadas do Amazon Kinesis Data Streams

Várias métricas do Amazon Kinesis Data Streams podem ser particularmente interessantes para os clientes do Kinesis Data Streams. A lista a seguir oferece métricas recomendadas e suas utilizações.

Métrica	Observações sobre o uso
`GetRecords.IteratorAgeMilliseconds`	Rastreia a posição de leitura em todos os fragmentos e consumidores no fluxo. Se a idade de um iterador passa de 50% do período de retenção (por padrão 24 horas, configurável até 7 dias), há risco de perda de dados devido à expiração de registro. Recomendamos que você use CloudWatch alarmes na estatística Máximo para alertá-lo antes que essa perda seja um risco. Para ver um exemplo de cenário que usa essa métrica, consulte Processamento de registros de consumidores ficando atrasados.
`ReadProvisionedThroughputExceeded`	Quando o processamento do registro do lado do consumidor está ficando para trás, às vezes é difícil saber onde está o gargalo. Use essa métrica para determinar se as leituras estão sendo limitadas por terem ultrapassado os limites de throughput de leitura. A estatística mais usada para essa métrica é Média.
`WriteProvisionedThroughputExceeded`	Ela tem a mesma finalidade da métrica `ReadProvisionedThroughputExceeded` , mas para o lado do produtor (put) do fluxo. A estatística mais usada para essa métrica é Média.
`PutRecord.Success` , `PutRecords.Success`	Recomendamos o uso de CloudWatch alarmes na estatística média para indicar quando os registros estão falhando na transmissão. Escolha um ou ambos os tipos put, dependendo do que o produtor usa. Se estiver usando a Amazon Kinesis Producer Library (KPL), use `PutRecords.Success` .
`GetRecords.Success`	Recomendamos o uso de CloudWatch alarmes na estatística média para indicar quando os registros estão falhando na transmissão.

Acesse CloudWatch as métricas da Amazon para Kinesis Data Streams

Você pode monitorar as métricas do Kinesis Data Streams CloudWatch usando o console, a linha de comando ou a API. CloudWatch Os procedimentos a seguir mostram como acessar as métricas usando os seguintes métodos:

Para acessar métricas usando o CloudWatch console

1. Abra o CloudWatch console em https://console.aws.amazon.com/cloudwatch/.
2. Na barra de navegação, escolha uma Região.
3. No painel de navegação, selecione Métricas.
4. No painel CloudWatch Métricas por categoria, escolha Kinesis Metrics.
5. Clique na linha relevante para ver as estatísticas do MetricNameStreamNamee.

   Observação: a maioria dos nomes de estatísticas do console corresponde aos nomes de CloudWatch métricas correspondentes listados anteriormente, exceto para taxa de transferência de leitura e taxa de transferência de gravação. Essas estatísticas são calculadas em intervalos de 5 minutos: a taxa de transferência de gravação monitora a `IncomingBytes` CloudWatch métrica e a taxa de transferência de leitura monitora. `GetRecords.Bytes`
6. (Opcional) No painel gráfico, selecione uma estatística e um período de tempo e, em seguida, crie um CloudWatch alarme usando essas configurações.

Para acessar métricas usando o AWS CLI

Use as métricas e get-metric-statisticsos comandos da lista.

Para acessar métricas usando a CloudWatch CLI

Use os comandos mon-list-metrics e mon-get-stats.

Para acessar métricas usando a CloudWatch API

Use as operações ListMetrics e GetMetricStatistics.

Monitore a integridade do Kinesis Data Streams Agent com a Amazon CloudWatch

O agente publica CloudWatch métricas personalizadas com um namespace de. AWS KinesisAgent Essas métricas ajudam a avaliar se o agente está enviando dados ao Kinesis Data Streams conforme especificado, e se está íntegro e consumindo a quantidade apropriada de recursos de CPU e memória no produtor de dados. As métricas, como número de registros e bytes enviados, são úteis para compreender a taxa em que o atendente está enviando dados ao fluxo. Quando essas métricas ficarem abaixo dos limites esperados em alguns percentuais ou caírem para zero, isso poderá indicar problemas de configuração, erros de rede ou problemas de integridade do agente. As métricas como consumo de CPU e memória no host e contadores de erros do agente indicam uso de recurso por parte do produtor de dados e fornece informações sobre erros potenciais de configuração ou de host. Por fim, o agente também registra exceções de serviço para ajudar a investigar problemas do agente. Essas métricas são relatadas na região especificada na configuração do agente`cloudwatch.endpoint`. CloudWatch as métricas publicadas de vários agentes do Kinesis são agregadas ou combinadas. Para obter mais informações sobre a configuração do atendente, consulte Especificar as definições da configuração do agente.

Monitor com CloudWatch

O agente do Kinesis Data Streams envia as seguintes métricas para o. CloudWatch

Métrica	Description
`BytesSent`	O número de bytes enviados para o Kinesis Data Streams no período especificado. Unidades: bytes
`RecordSendAttempts`	O número de tentativas de registro (primeira vez ou como nova tentativa) em uma chamada para `PutRecords` no período especificado. Unidades: contagem
`RecordSendErrors`	O número de registros que retornaram status de falha em uma chamada para `PutRecords` , incluindo novas tentativas, no período especificado. Unidades: contagem

Métrica	Description
`ServiceErrors`	O número de chamadas para `PutRecords` que resultaram em erro de serviço (diferente de um erro de controle de utilização) no período especificado. Unidades: contagem

Registre as chamadas de API do Amazon Kinesis Data Streams com AWS CloudTrail

O Amazon Kinesis Data Streams está AWS CloudTrail integrado com, um serviço que fornece um registro das ações realizadas por um usuário, função AWS ou serviço no Kinesis Data Streams. CloudTrail captura todas as chamadas de API para o Kinesis Data Streams como eventos. São capturadas as chamadas do console do Kinesis Data Streams e as chamadas de código às operações de API do Kinesis Data Streams. Se você criar uma trilha, poderá habilitar a entrega contínua de CloudTrail eventos para um bucket do Amazon S3, incluindo eventos para o Kinesis Data Streams. Se você não configurar uma trilha, ainda poderá ver os eventos mais recentes no CloudTrail console no Histórico de eventos. Usando as informações coletadas por CloudTrail, você pode determinar a solicitação que foi feita ao Kinesis Data Streams, o endereço IP a partir do qual a solicitação foi feita, quem fez a solicitação, quando ela foi feita e detalhes adicionais.

Para saber mais CloudTrail, inclusive como configurá-lo e ativá-lo, consulte o Guia AWS CloudTrail do usuário.

Informações do Kinesis Data Streams em CloudTrail

CloudTrail é ativado em sua AWS conta quando você cria a conta. Quando uma atividade de evento suportada ocorre no Kinesis Data Streams, essa atividade é registrada CloudTrail em um evento junto AWS com outros eventos de serviço no histórico de eventos. Você pode visualizar, pesquisar e baixar eventos recentes em sua AWS conta. Para obter mais informações, consulte Visualizar eventos com o histórico de eventos do CloudTrail.

Para um registro contínuo dos eventos em sua AWS conta, incluindo eventos do Kinesis Data Streams, crie uma trilha. Uma trilha permite CloudTrail entregar arquivos de log para um bucket do Amazon S3. Por padrão, quando você cria uma trilha no console, a trilha se aplica a todas as AWS regiões. A trilha registra eventos de todas as regiões na AWS partição e entrega os arquivos de

log ao bucket do Amazon S3 que você especificar. Além disso, você pode configurar outros AWS serviços para analisar e agir com base nos dados de eventos coletados nos CloudTrail registros. Para saber mais, consulte:

- Visão Geral para Criar uma Trilha
- CloudTrail Serviços e integrações compatíveis
- Configurando notificações do Amazon SNS para CloudTrail
- Recebendo arquivos de CloudTrail log de várias regiões e recebendo arquivos de CloudTrail log de várias contas

O Kinesis Data Streams oferece suporte ao registro das seguintes ações como CloudTrail eventos em arquivos de log:

- AddTagsToStream
- CreateStream
- DecreaseStreamRetentionPeriod
- DeleteStream
- DeregisterStreamConsumer
- DescribeStream
- DescribeStreamConsumer
- DisableEnhancedMonitoring
- EnableEnhancedMonitoring
- GetRecords
- GetShardIterator
- IncreaseStreamRetentionPeriod
- ListStreamConsumers
- ListStreams
- ListTagsForStream
- MergeShards
- PutRecord
- PutRecords
- RegisterStreamConsumer

- RemoveTagsFromStream
- SplitShard
- StartStreamEncryption
- StopStreamEncryption
- SubscribeToShard
- UpdateShardCount
- UpdateStreamMode

Cada entrada de log ou evento contém informações sobre quem gerou a solicitação. As informações de identidade ajudam a determinar o seguinte:

- Se a solicitação foi feita com credenciais de usuário root ou AWS Identity and Access Management (IAM).
- Se a solicitação foi feita com credenciais de segurança temporárias de um perfil ou de um usuário federado.
- Se a solicitação foi feita por outro AWS serviço.

Para obter mais informações, consulte Elemento userIdentity do CloudTrail.

Exemplo: entradas do arquivo de log do Kinesis Data Streams

Uma trilha é uma configuração que permite a entrega de eventos como arquivos de log para um bucket do Amazon S3 que você especificar. CloudTrail os arquivos de log contêm uma ou mais entradas de log. Um evento representa uma única solicitação de qualquer fonte e inclui informações sobre a ação solicitada, a data e hora da ação, parâmetros de solicitação, e assim por diante. arquivos de log do CloudTrail não são um rastreamento de pilha ordenada das chamadas da API pública. Assim, elas não são exibidas em nenhuma ordem específica.

O exemplo a seguir mostra uma entrada de CloudTrail registro que demonstra as `MergeShards` ações `CreateStream` `DescribeStreamListStreams`,`DeleteStream`,`SplitShard`,, e.

```
{
    "Records": [
        {
            "eventVersion": "1.01",
            "userIdentity": {
                "type": "IAMUser",
```

```
                "principalId": "EX_PRINCIPAL_ID",
                "arn": "arn:aws:iam::012345678910:user/Alice",
                "accountId": "012345678910",
                "accessKeyId": "EXAMPLE_KEY_ID",
                "userName": "Alice"
            },
            "eventTime": "2014-04-19T00:16:31Z",
            "eventSource": "kinesis.amazonaws.com",
            "eventName": "CreateStream",
            "awsRegion": "us-east-1",
            "sourceIPAddress": "127.0.0.1",
            "userAgent": "aws-sdk-java/unknown-version Linux/x.xx",
            "requestParameters": {
                "shardCount": 1,
                "streamName": "GoodStream"
            },
            "responseElements": null,
            "requestID": "db6c59f8-c757-11e3-bc3b-57923b443c1c",
            "eventID": "b7acfcd0-6ca9-4ee1-a3d7-c4e8d420d99b"
        },
        {
            "eventVersion": "1.01",
            "userIdentity": {
                "type": "IAMUser",
                "principalId": "EX_PRINCIPAL_ID",
                "arn": "arn:aws:iam::012345678910:user/Alice",
                "accountId": "012345678910",
                "accessKeyId": "EXAMPLE_KEY_ID",
                "userName": "Alice"
            },
            "eventTime": "2014-04-19T00:17:06Z",
            "eventSource": "kinesis.amazonaws.com",
            "eventName": "DescribeStream",
            "awsRegion": "us-east-1",
            "sourceIPAddress": "127.0.0.1",
            "userAgent": "aws-sdk-java/unknown-version Linux/x.xx",
            "requestParameters": {
                "streamName": "GoodStream"
            },
            "responseElements": null,
            "requestID": "f0944d86-c757-11e3-b4ae-25654b1d3136",
            "eventID": "0b2f1396-88af-4561-b16f-398f8eaea596"
        },
        {
```

```
            "eventVersion": "1.01",
            "userIdentity": {
                "type": "IAMUser",
                "principalId": "EX_PRINCIPAL_ID",
                "arn": "arn:aws:iam::012345678910:user/Alice",
                "accountId": "012345678910",
                "accessKeyId": "EXAMPLE_KEY_ID",
                "userName": "Alice"
            },
            "eventTime": "2014-04-19T00:15:02Z",
            "eventSource": "kinesis.amazonaws.com",
            "eventName": "ListStreams",
            "awsRegion": "us-east-1",
            "sourceIPAddress": "127.0.0.1",
            "userAgent": "aws-sdk-java/unknown-version Linux/x.xx",
            "requestParameters": {
                "limit": 10
            },
            "responseElements": null,
            "requestID": "a68541ca-c757-11e3-901b-cbcfe5b3677a",
            "eventID": "22a5fb8f-4e61-4bee-a8ad-3b72046b4c4d"
        },
        {
            "eventVersion": "1.01",
            "userIdentity": {
                "type": "IAMUser",
                "principalId": "EX_PRINCIPAL_ID",
                "arn": "arn:aws:iam::012345678910:user/Alice",
                "accountId": "012345678910",
                "accessKeyId": "EXAMPLE_KEY_ID",
                "userName": "Alice"
            },
            "eventTime": "2014-04-19T00:17:07Z",
            "eventSource": "kinesis.amazonaws.com",
            "eventName": "DeleteStream",
            "awsRegion": "us-east-1",
            "sourceIPAddress": "127.0.0.1",
            "userAgent": "aws-sdk-java/unknown-version Linux/x.xx",
            "requestParameters": {
                "streamName": "GoodStream"
            },
            "responseElements": null,
            "requestID": "f10cd97c-c757-11e3-901b-cbcfe5b3677a",
            "eventID": "607e7217-311a-4a08-a904-ec02944596dd"
```

```
        },
        {
            "eventVersion": "1.01",
            "userIdentity": {
                "type": "IAMUser",
                "principalId": "EX_PRINCIPAL_ID",
                "arn": "arn:aws:iam::012345678910:user/Alice",
                "accountId": "012345678910",
                "accessKeyId": "EXAMPLE_KEY_ID",
                "userName": "Alice"
            },
            "eventTime": "2014-04-19T00:15:03Z",
            "eventSource": "kinesis.amazonaws.com",
            "eventName": "SplitShard",
            "awsRegion": "us-east-1",
            "sourceIPAddress": "127.0.0.1",
            "userAgent": "aws-sdk-java/unknown-version Linux/x.xx",
            "requestParameters": {
                "shardToSplit": "shardId-000000000000",
                "streamName": "GoodStream",
                "newStartingHashKey": "11111111"
            },
            "responseElements": null,
            "requestID": "a6e6e9cd-c757-11e3-901b-cbcfe5b3677a",
            "eventID": "dcd2126f-c8d2-4186-b32a-192dd48d7e33"
        },
        {
            "eventVersion": "1.01",
            "userIdentity": {
                "type": "IAMUser",
                "principalId": "EX_PRINCIPAL_ID",
                "arn": "arn:aws:iam::012345678910:user/Alice",
                "accountId": "012345678910",
                "accessKeyId": "EXAMPLE_KEY_ID",
                "userName": "Alice"
            },
            "eventTime": "2014-04-19T00:16:56Z",
            "eventSource": "kinesis.amazonaws.com",
            "eventName": "MergeShards",
            "awsRegion": "us-east-1",
            "sourceIPAddress": "127.0.0.1",
            "userAgent": "aws-sdk-java/unknown-version Linux/x.xx",
            "requestParameters": {
                "streamName": "GoodStream",
```

```
                "adjacentShardToMerge": "shardId-000000000002",
                "shardToMerge": "shardId-000000000001"
            },
            "responseElements": null,
            "requestID": "e9f9c8eb-c757-11e3-bf1d-6948db3cd570",
            "eventID": "77cf0d06-ce90-42da-9576-71986fec411f"
        }
    ]
}
```

Monitore a biblioteca de cliente do Kinesis com a Amazon CloudWatch

A Kinesis Client Library (KCL) para Amazon Kinesis Data Streams publica métricas personalizadas da CloudWatch Amazon em seu nome, usando o nome do seu aplicativo KCL como namespace. Você pode visualizar essas métricas navegando até o CloudWatch console e escolhendo Métricas personalizadas. Para obter mais informações sobre métricas personalizadas, consulte Publicar métricas personalizadas no Guia CloudWatch do usuário da Amazon.

Há uma cobrança nominal pelas métricas enviadas CloudWatch pela KCL; especificamente, as cobranças de Amazon CloudWatch Custom Metrics e Amazon CloudWatch API Requests se aplicam. Para obter mais informações, consulte Amazon CloudWatch Pricing.

Tópicos

- Métricas e namespace
- Dimensões e níveis de métricas
- Configuração da métrica
- Lista de métricas

Métricas e namespace

O namespace usado para o carregamento de métricas é o nome da aplicação especificado ao executar o KCL.

Dimensões e níveis de métricas

Há duas opções para controlar quais métricas são carregadas para o CloudWatch:

níveis de métrica

Cada métrica é atribuída a um nível individual. Quando você define um nível de relatório de métricas, métricas com um nível individual abaixo do nível de relatório não são enviadas para CloudWatch. Os níveis são: `NONE`, `SUMMARY` e `DETAILED`. A configuração padrão é`DETAILED`; ou seja, todas as métricas são enviadas para CloudWatch. Um nível de relatório `NONE` significa que nenhuma métrica é enviada. Para obter informações sobre quais níveis são atribuídos a quais métricas, consulte Lista de métricas.

dimensões habilitadas

Cada métrica KCL tem dimensões associadas que também são enviadas para CloudWatch. Na KCL 2.x, se a KCL estiver configurada para processar um único fluxo de dados, todas as dimensões de métricas (`Operation`, `ShardId` e `WorkerIdentifier`) serão habilitadas por padrão. Além disso, na KCL 2.x, se a KCL estiver configurada para processar um único fluxo de dados, a dimensão `Operation` não poderá ser desabilitada. Na KCL 2.x, se a KCL estiver configurada para processar vários fluxos de dados, todas as dimensões de métricas (`Operation`, `ShardId`, `StreamId` e `WorkerIdentifier`) serão habilitadas por padrão. Além disso, na KCL 2.x, se a KCL estiver configurada para processar vários fluxos de dados, as dimensões `Operation` e `StreamId` não poderão ser desabilitadas. A dimensão `StreamId` só está disponível para métricas por fragmento.

Na KCL 1.x, apenas as dimensões `Operation` e `ShardId` estão habilitadas por padrão, e a dimensão `WorkerIdentifier` está desabilitada. Na KCL 1.x, a dimensão `Operation` não pode ser desabilitada.

Para obter mais informações sobre dimensões CloudWatch métricas, consulte a seção Dimensões no tópico Amazon CloudWatch Concepts, no Guia CloudWatch do usuário da Amazon.

Quando a `WorkerIdentifier` dimensão é ativada, se um valor diferente for usado para a propriedade de ID do trabalhador toda vez que um determinado trabalhador da KCL for reiniciado, novos conjuntos de métricas com novos valores de `WorkerIdentifier` dimensão serão enviados para. CloudWatch Se houver necessidade de que o valor da dimensão `WorkerIdentifier` seja o mesmo nas reinicializações de um operador da KCL específico, será necessário especificar explicitamente o mesmo valor do ID do operador durante a inicialização para cada operador. Observe que o valor do ID do operador para cada operador da KCL ativo precisa ser único entre todos os operadores da KCL.

Configuração da métrica

Os níveis métricos e as dimensões habilitadas podem ser configurados usando a KinesisClientLibConfiguration instância, que é passada para o Worker ao iniciar o aplicativo KCL. MultiLangDaemon Nesse caso, as `metricsEnabledDimensions` propriedades `metricsLevel` e podem ser especificadas no arquivo.properties usado para iniciar o aplicativo MultiLangDaemon KCL.

Os níveis de métrica podem ser atribuídos a um dos três valores: NONE, SUMMARY ou DETAILED. Os valores das dimensões ativadas devem ser sequências de caracteres separadas por vírgula com a lista de dimensões permitidas para as métricas. CloudWatch As dimensões usadas pela aplicação da KCL são `Operation`, `ShardId` e `WorkerIdentifier`.

Lista de métricas

As tabelas a seguir listam as métricas da KCL agrupadas por escopo e operação.

Tópicos

- Per-KCL-application métricas
- Métricas por operador
- Métricas por fragmento

Per-KCL-application métricas

Essas métricas são agregadas em todos os trabalhadores da KCL dentro do escopo do aplicativo, conforme definido pelo namespace da Amazon CloudWatch .

Tópicos

- LeaseAssignmentManager
- InitializeTask
- ShutdownTask
- ShardSyncTask
- BlockOnParentTask
- PeriodicShardSyncManager
- MultistreamTracker

LeaseAssignmentManager

A operação `LeaseAssignmentManager` é responsável por atribuir concessões aos operadores e reequilibrar as concessões entre os operadores para obter uma utilização uniforme dos recursos dos operadores. A lógica dessa operação inclui a leitura dos metadados da concessão presentes na tabela da concessão e as métricas da tabela de métricas do operador e a execução de atribuições da concessão.

Métrica	Description
LeaseAndWorkerMetricsLoad.Hora	Tempo gasto para carregar todas as concessões e informações de métricas dos operadores no gerenciador de atribuição de concessão (LAM), o novo algoritmo de atribuição de concessão e balanceamento de carga introduzido na KCL 3.x. Nível de métrica: detalhado Unidade: milissegundos
TotalLeases	Número total de concessões para a aplicação da KCL atual. Nível de métrica: resumo Unidades: contagem
NumWorkers	O número total de operadores na aplicação atual da KCL. Nível de métrica: resumo Unidades: contagem
AssignExpiredOrUnassignedLeases.Hora	Tempo para a realização da atribuição em memória de concessões expiradas. Nível de métrica: detalhado Unidade: milissegundos
LeaseSpillover	Número de concessões que não foram atribuídas por atingirem o limite do número máximo de concessões ou de throughput máxima por operador.

Métrica	Description
	Nível de métrica: resumo Unidades: contagem
BalanceWorkerVariance.Hora	Tempo para a realização do balanceamento na memória das concessões entre operadores. Nível de métrica: detalhado Unidade: milissegundos
NumOfLeasesReassignment	Número total de reatribuições de concessões feitas na iteração de reatribuição atual. Nível de métrica: resumo Unidades: contagem
FailedAssignmentCount	Número de falhas nas AssignLease chamadas para a tabela de lease do DynamoDB. Nível de métrica: detalhado Unidades: contagem
ParallelyAssignLeases.Hora	Tempo para a transferência de novas atribuições para a tabela de concessões do DynamoDB. Nível de métrica: detalhado Unidade: milissegundos
ParallelyAssignLeases.Sucesso	Número de transferências de novas atribuições com êxito. Nível de métrica: detalhado Unidades: contagem

Métrica	Description
TotalStaleWorkerMetricsEntry	Número total de informações de métricas do operador que devem ser eliminadas. Nível de métrica: detalhado Unidades: contagem
StaleWorkerMetricsCleanup.Hora	Tempo para a realização da exclusão das informações das métricas do operador da tabela de métricas de operadores do DynamoDB. Nível de métrica: detalhado Unidade: milissegundos
Hora	Tempo gasto pela operação `LeaseAssignmentManager` . Nível de métrica: resumo Unidade: milissegundos
Bem-sucedida	Número de vezes que a operação `LeaseAssignmentManager` foi concluída com sucesso. Nível de métrica: resumo Unidades: contagem
ForceLeaderRelease	Indica que o gerenciador de atribuição de concessões falhou 3 vezes consecutivas e que o operador líder está liberando a liderança. Nível de métrica: resumo Unidades: contagem
NumWorkersWithInvalidEntry	Número de informações de métricas do operador que são consideradas inválidas. Nível de métrica: resumo Unidades: contagem

Métrica	Description
NumWorkersWithFailingWorkerMetric	Número de informações de métricas do operador que têm -1 (representando que o valor da métrica do operador não está disponível) como um dos valores das métricas do operador. Nível de métrica: resumo Unidades: contagem
LeaseDeserializationFailureCount	Informações da concessão a partir da tabela de concessões que falhou em desserializar. Nível de métrica: resumo Unidades: contagem

InitializeTask

A operação `InitializeTask` é responsável por inicializar o processador de registros para a aplicação da KCL. A lógica dessa operação inclui a obtenção de um iterador de fragmentos do Kinesis Data Streams e a inicialização do processador de registros.

Métrica	Description
KinesisDataFetcher.getIterator.Success	Número de operações `GetShardIterator` bem-sucedidas por aplicação da KCL. Nível de métrica: detalhado Unidades: contagem
KinesisDataFetcher.getIterator.time	Tempo decorrido por operação `GetShardIterator` para a aplicação da KCL determinada. Nível de métrica: detalhado Unidade: milissegundos

Métrica	Description
RecordProcessor.Initialize.time	Tempo percorrido pelo método de inicialização do processador de registros. Nível de métrica: resumo Unidade: milissegundos
Bem-sucedida	Número de inicializações bem-sucedidas do processador de registros. Nível de métrica: resumo Unidades: contagem
Hora	Tempo decorrido pelo operador da KCL para inicialização do processador de registros. Nível de métrica: resumo Unidade: milissegundos

ShutdownTask

A operação `ShutdownTask` inicia a sequência de desligamento para o processamento de fragmento. Isso pode ocorrer porque um fragmento é dividido ou mesclado, ou quando a concessão do fragmento é perdida no operador. Em ambos os casos, a função `shutdown()` do processador de registros é chamada. Novos fragmentos também são descobertos no caso em que um fragmento é dividido ou mesclado, resultando na criação de um ou dois novos fragmentos.

Métrica	Description
CreateLease.Sucesso	O número de vezes que novos fragmentos filho são adicionados com êxito à tabela do DynamoDB da aplicação da KCL após o desligamento do fragmento pai. Nível de métrica: detalhado Unidades: contagem

Métrica	Description
CreateLease.Hora	Tempo decorrido para adicionar informações de novos fragmentos filho à tabela do DynamoDB da aplicação da KCL. Nível de métrica: detalhado Unidade: milissegundos
UpdateLease.Sucesso	Número de pontos de verificação finais bem-sucedidos durante o desligamento do processador de registros. Nível de métrica: detalhado Unidades: contagem
UpdateLease.Hora	Tempo necessário para a operação de pontos de verificação durante o desligamento do processador de registros. Nível de métrica: detalhado Unidade: milissegundos
RecordProcessor.Horário de desligamento	Tempo percorrido pelo método de desligamento do processador de registros. Nível de métrica: resumo Unidade: milissegundos
Bem-sucedida	Número de tarefas de desligamento bem-sucedidas. Nível de métrica: resumo Unidades: contagem
Hora	Tempo decorrido pelo operador da KCL para a tarefa de desligamento. Nível de métrica: resumo Unidade: milissegundos

ShardSyncTask

A operação `ShardSyncTask` detecta alterações nas informações de fragmentos do fluxo de dados do Kinesis para que novos fragmentos possam ser processados pela aplicação da KCL.

Métrica	Description
CreateLease.Sucesso	Número de tentativas bem-sucedidas para adicionar novas informações de fragmentos à tabela do DynamoDB da aplicação da KCL. Nível de métrica: detalhado Unidades: contagem
CreateLease.Hora	Tempo decorrido para adicionar informações de novos fragmentos à tabela do DynamoDB da aplicação da KCL. Nível de métrica: detalhado Unidade: milissegundos
Bem-sucedida	Número de operações bem-sucedidas de sincronização de fragmentos. Nível de métrica: resumo Unidades: contagem
Hora	Tempo percorrido para a operação de sincronização de fragmentos. Nível de métrica: resumo Unidade: milissegundos

BlockOnParentTask

Se o fragmento é dividido ou mesclado com outros, novos fragmentos filhos são criados. A operação `BlockOnParentTask` garante que o processamento de registros de novos fragmentos não será iniciado até que os fragmentos pai sejam completamente processados pela KCL.

Métrica	Description
Bem-sucedida	Número de verificações bem-sucedidas para a conclusão de fragmentos pai. Nível de métrica: resumo Unidades: contagem
Hora	Tempo percorrido para a conclusão de fragmentos pai. Nível de métrica: resumo Unidade: milissegundos

PeriodicShardSyncManager

A `PeriodicShardSyncManager` é responsável por examinar os fluxos de dados sendo processados pela aplicação de consumo da KCL, identificando fluxos de dados com concessões parciais e entregando-os para sincronização.

As métricas a seguir estão disponíveis quando a KCL está configurada para processar um único fluxo de dados (em seguida, o valor de NumStreamsToSync e NumStreamsWithPartialLeases é definido como 1) e também quando a KCL está configurada para processar vários fluxos de dados.

Métrica	Description
NumStreamsToSync	O número de fluxos de dados (por AWS conta) processados pelo aplicativo consumidor que contém concessões parciais e que devem ser entregues para sincronização. Nível de métrica: resumo Unidades: contagem
NumStreamsWithPartialLeases	O número de fluxos de dados (por AWS conta) que o aplicativo do consumidor está processando e que contêm concessões parciais. Nível de métrica: resumo

Métrica	Description
	Unidades: contagem
Bem-sucedida	O número de vezes que `PeriodicShardSyncManager` conseguiu identificar com êxito as concessões parciais nos fluxos de dados que a aplicação de consumo está processando. Nível de métrica: resumo Unidades: contagem
Hora	O tempo (em milissegundos) que `PeriodicShardSyncManager` leva para examinar os fluxos de dados processados pela aplicação de consumo, a fim de determinar quais dos fluxos de dados exigem sincronização de fragmentos. Nível de métrica: resumo Unidade: milissegundos

MultistreamTracker

A interface `MultistreamTracker` permite a criação de aplicações de consumo da KCL que podem processar vários fluxos de dados ao mesmo tempo.

Métrica	Description
DeletedStreams.Contagem	O número de fluxos de dados excluídos no período. Nível de métrica: resumo Unidades: contagem
ActiveStreams.Contagem	O número de fluxos de dados ativos sendo processados. Nível de métrica: resumo Unidades: contagem

Métrica	Description
StreamsPendingDeletion.Contagem	O número de fluxos de dados com exclusão pendente com base em `FormerStreamsLeasesDeletionStrategy` . Nível de métrica: resumo Unidades: contagem

Métricas por operador

Essas métricas são agregadas em todos os processadores de registros que consomem dados de um fluxo de dados do Kinesis, como uma instância do Amazon EC2.

Tópicos

- WorkerMetricStatsReporter
- LeaseDiscovery
- RenewAllLeases
- TakeLeases

WorkerMetricStatsReporter

A operação `WorkerMetricStatReporter` é responsável por publicar periodicamente as métricas do operador atual na tabela de métricas do operador. Essas métricas são usadas pela operação `LeaseAssignmentManager` para realizar atribuições de concessão.

Métrica	Description
InMemoryMetricStatsReporterFailure	Número de falhas na captura do valor da métrica do operador na memória, devido à falha de algumas métricas do operador. Nível de métrica: resumo Unidades: contagem
WorkerMetricStatsReporter.Hora	Tempo gasto pela operação `WorkerMetricsStats` . Nível de métrica: resumo

Métrica	Description
	Unidade: milissegundos
WorkerMetricStatsReporter.Sucesso	Número de vezes que a operação `WorkerMetricsStats` foi concluída com sucesso. Nível de métrica: resumo Unidades: contagem

LeaseDiscovery

A operação `LeaseDiscovery` é responsável por identificar as novas concessões atribuídas ao operador atual pela operação `LeaseAssignmentManager`. A lógica dessa operação envolve a identificação das concessões atribuídas ao operador atual por meio da leitura do índice secundário global da tabela de concessões.

Métrica	Description
ListLeaseKeysForWorker.Hora	O tempo para chamar o índice secundário global na tabela de concessões e obter as chaves de concessão atribuídas ao operador atual. Nível de métrica: detalhado Unidade: milissegundos
FetchNewLeases.Hora	O tempo para buscar todas as novas concessões na tabela de concessões. Nível de métrica: detalhado Unidade: milissegundos
NewLeasesDiscovered	O número total de novas concessões atribuídas aos operadores. Nível de métrica: detalhado Unidades: contagem
Hora	Tempo gasto pela operação `LeaseDiscovery` .

Métrica	Description
	Nível de métrica: resumo Unidade: milissegundos
Bem-sucedida	Número de vezes que a operação `LeaseDiscovery` foi concluída com sucesso. Nível de métrica: resumo Unidades: contagem
OwnerMismatch	O número de incompatibilidades do proprietário em relação à resposta do GSI e à leitura consistente da tabela de concessão. Nível de métrica: detalhado Unidades: contagem

RenewAllLeases

A operação `RenewAllLeases` renova periodicamente concessões de fragmentos de propriedade de uma determinada instância de operador.

Métrica	Description
RenewLease.Sucesso	Número de renovações bem-sucedidas de concessões por parte do operador. Nível de métrica: detalhado Unidades: contagem
RenewLease.Hora	O tempo necessário para a operação de renovação de concessões. Nível de métrica: detalhado Unidade: milissegundos

Métrica	Description
CurrentLeases	Número de concessões de fragmentos pertencentes ao operador depois que todas as concessões foram renovadas. Nível de métrica: resumo Unidades: contagem
LostLeases	Número de concessões de fragmentos perdidas após uma tentativa de renovar todas as concessões pertencentes ao operador. Nível de métrica: resumo Unidades: contagem
Bem-sucedida	Número de vezes que a operação de renovação da concessão foi bem-sucedida para o operador. Nível de métrica: resumo Unidades: contagem
Hora	Tempo percorrido para renovar todas as concessões para o operador. Nível de métrica: resumo Unidade: milissegundos

TakeLeases

A operação `TakeLeases` equilibra o processamento de registros entre todos os operadores da KCL. Se tiver menos concessões de fragmentos que o necessário, o operador atual da KCL usará concessões de fragmentos de outro operador que esteja sobrecarregado.

Métrica	Description
ListLeases.Sucesso	Número de vezes que todas as concessões de fragmentos foram recuperadas com êxito da tabela do DynamoDB da aplicação da KCL.

Métrica	Description
	Nível de métrica: detalhado Unidades: contagem
ListLeases.Hora	Tempo decorrido para recuperar todas as concessões de fragmentos da tabela do DynamoDB da aplicação da KCL. Nível de métrica: detalhado Unidade: milissegundos
TakeLease.Sucesso	Número de vezes que o operador usou com êxito concessões de fragmentos de outros operadores da KCL. Nível de métrica: detalhado Unidades: contagem
TakeLease.Hora	Tempo decorrido para atualizar a tabela de concessão com concessões executadas pelo operador. Nível de métrica: detalhado Unidade: milissegundos
NumWorkers	Número total de operadores, conforme identificado por um operador específico. Nível de métrica: resumo Unidades: contagem
NeededLeases	Número de concessões de fragmentos que o operador atual precisa para uma carga de processamento de fragmentos equilibrada. Nível de métrica: detalhado Unidades: contagem

Métrica	Description
LeasesToTake	O número de concessões que o operador tentará executar. Nível de métrica: detalhado Unidades: contagem
TakenLeases	Número de concessões realizadas com sucesso pelo operador. Nível de métrica: resumo Unidades: contagem
TotalLeases	Número total de fragmentos que a aplicação da KCL está processando. Nível de métrica: detalhado Unidades: contagem
ExpiredLeases	Número total de fragmentos que não estão sendo processados por nenhum operador, conforme identificado pelo operador específico. Nível de métrica: resumo Unidades: contagem
Bem-sucedida	Número de vezes que a operação `TakeLeases` foi concluída com sucesso. Nível de métrica: resumo Unidades: contagem
Hora	Tempo percorrido pela operação `TakeLeases` para um operador. Nível de métrica: resumo Unidade: milissegundos

Métricas por fragmento

Essas métricas são agregadas em um único processador de registros.

ProcessTask

A operação `ProcessTask` chama GetRecords com a posição do iterador atual para recuperar registros do streaming e chama a função `processRecords` do processador de registros.

Métrica	Description
KinesisDataFetcher .getRecords.Sucesso	Número de operações `GetRecords` bem-sucedidas por fragmento do fluxo de dados do Kinesis. Nível de métrica: detalhado Unidades: contagem
KinesisDataFetcher .Obter registros. Hora	Tempo decorrido por operação `GetRecords` para o fragmento do fluxo de dados do Kinesis. Nível de métrica: detalhado Unidade: milissegundos
UpdateLease.Sucesso	Número de pontos de verificação bem-sucedidos feitos pelo processador de registros para o determinado fragmento. Nível de métrica: detalhado Unidades: contagem
UpdateLease.Hora	Tempo percorrido para cada operação de ponto de verificação para o determinado fragmento. Nível de métrica: detalhado Unidade: milissegundos
DataBytesProcessed	Tamanho total de registros processados em bytes em cada chamada de `ProcessTask` .

Métrica	Description
	Nível de métrica: resumo Unidades: byte
RecordsProcessed	Número de registros processados em cada chamada de `ProcessTask` . Nível de métrica: resumo Unidades: contagem
ExpiredIterator	Número de ExpiredIteratorException recebidos ao ligar`GetRecords` . Nível de métrica: resumo Unidades: contagem
MillisBehindLatest	Tempo em que o iterador atual está atrás do registro mais recente (ponta) no fragmento. Esse valor é menor ou igual à diferença da hora entre o registro mais recente em uma resposta e a hora atual. Esse é um reflexo mais preciso da distância em que um fragmento está da ponta do que a comparação de carimbos de data/hora no último registro de resposta. Esse valor se aplica ao último lote de registros, não a uma média de todos os carimbos de data/hora em cada registro. Nível de métrica: resumo Unidade: milissegundos
RecordProcessor.Registros do processo. Tempo	Tempo percorrido pelo método `processRecords` do processador de registros. Nível de métrica: resumo Unidade: milissegundos

Métrica	Description
Bem-sucedida	Número de operações bem-sucedidas de tarefas do processo. Nível de métrica: resumo Unidades: contagem
Hora	Tempo percorrido para a operação de tarefas do processo. Nível de métrica: resumo Unidade: milissegundos

Monitore a Kinesis Producer Library com a Amazon CloudWatch

A Amazon Kinesis Producer Library (KPL) para Amazon Kinesis Data Streams publica métricas personalizadas da Amazon em seu nome. CloudWatch Você pode visualizar essas métricas navegando até o CloudWatch console e escolhendo Métricas personalizadas. Para obter mais informações sobre métricas personalizadas, consulte Publicar métricas personalizadas no Guia CloudWatch do usuário da Amazon.

Há uma cobrança nominal pelas métricas enviadas CloudWatch pela KPL; especificamente, as taxas de métricas CloudWatch personalizadas da Amazon e das solicitações de CloudWatch API da Amazon se aplicam. Para obter mais informações, consulte Amazon CloudWatch Pricing. A coleta de métricas locais não gera cobranças do CloudWatch.

Tópicos

- Métricas, dimensões e namespaces
- Granularidade e nível de métrica
- Acesso local e CloudWatch upload da Amazon
- Lista de métricas

Métricas, dimensões e namespaces

É possível especificar um nome de aplicação ao ativar a KPL, que será usado como parte do namespace durante o carregamento de métricas. Esse passo é opcional. A KPL atribuirá um valor padrão se o nome da aplicação não for definido.

Também é possível configurar a KPL para adicionar dimensões arbitrárias às métricas. Isso é útil se você quiser dados mais detalhados em suas métricas. CloudWatch Por exemplo, pode-se adicionar o nome de host como uma dimensão, o que permite a identificação de distribuições de carga irregulares na frota. Como todas as definições de configuração da KPL são imutáveis, não será possível alterar essas dimensões adicionais depois que a instância da KPL for inicializada.

Granularidade e nível de métrica

Há duas opções para controlar o número de métricas carregadas para o CloudWatch:

nível de métrica

Este é um indicador aproximado da importância de uma métrica. Cada métrica é atribuída a um nível. Quando você define um nível, as métricas com níveis abaixo desses não são enviadas para CloudWatch. Os níveis são `NONE`, `SUMMARY` e `DETAILED`. A configuração padrão é `DETAILED`. Ou seja, todas as métricas. `NONE` significa que não há métricas, de modo que nenhuma métrica é atribuída a esse nível.

granularidade

Controla se a mesma métrica é emitida em níveis adicionais de granularidade. Os níveis são `GLOBAL`, `STREAM` e `SHARD`. A configuração padrão é `SHARD`, que contém as métricas mais granulares.

Quando `SHARD` é escolhida, as métricas são emitidas com o nome do fluxo e o ID do fragmento como dimensões. Além disso, a mesma métrica também é emitida com somente a dimensão do nome do fluxo, e a métrica sem o nome do fluxo. Isso significa que, para uma métrica específica, dois fluxos com dois fragmentos cada produzirão sete CloudWatch métricas: uma para cada fragmento, uma para cada fluxo e uma geral; todas descrevendo as mesmas estatísticas, mas em diferentes níveis de granularidade. Para ter um esclarecimento, consulte o diagrama abaixo.

Os diferentes níveis de granularidade formam uma hierarquia, e todas as métricas no sistema formam árvores, enraizadas nos nomes da métrica:

```
MetricName (GLOBAL):          Metric X                    Metric Y
```

```
                              |                       |
                     -----------------           ------------
                     |               |           |          |
StreamName (STREAM):    Stream A        Stream B     Stream A   Stream B
                        |               |
                     --------        ---------
                     |      |        |       |
ShardID (SHARD):     Shard 0 Shard 1  Shard 0 Shard 1
```

Nem todas as métricas estão disponíveis no nível de fragmento; algumas são de nível de fluxo ou globais por natureza. Elas não são produzidas no nível de fragmento, mesmo quee as métricas estejam habilitadas nesse nível (`Metric Y` no diagrama acima).

Ao especificar uma dimensão adicional, é necessário fornecer valores para `tuple:<DimensionName, DimensionValue, Granularity>`. A granularidade é usada para determinar onde a dimensão personalizada é inserida na hierarquia: `GLOBAL` significa que a dimensão adicional é inserida após o nome da métrica, `STREAM` significa que ela é inserida após o nome do fluxo e `SHARD` significa que ela é inserida depois do ID de fragmento. Se várias dimensões adicionais são fornecidas por nível de granularidade, elas são inseridas na ordem determinada.

Acesso local e CloudWatch upload da Amazon

As métricas para a instância atual da KPL estão disponíveis localmente em tempo real, e é possível consultar a KPL a qualquer momento para obtê-las. O KPL calcula localmente a soma, média, mínimo, máximo e contagem de cada métrica, como em. CloudWatch

Pode-se obter estatísticas cumulativas desde o início do programa até o presente momento ou usar uma janela contínua ao longo dos últimos N segundos, em que N é um número inteiro entre 1 e 60.

Todas as métricas estão disponíveis para upload em CloudWatch. Isso é especialmente útil para agregar dados em vários hosts, monitoramentos e alarmes. Essa funcionalidade não está disponível localmente.

Conforme descrito anteriormente, é possível selecionar de quais métricas fazer upload com as configurações de nível de métrica e granularidade. As métricas que não são carregadas estão disponíveis localmente.

Carregar pontos de dados individualmente é insustentável, porque isso pode produzir milhões de carregamentos por segundo se o tráfego for alto. Por esse motivo, o KPL agrega métricas localmente

em intervalos de 1 minuto e carrega um objeto de estatísticas uma vez por minuto, por métrica ativada. CloudWatch

Lista de métricas

Métrica	Description
`UserRecordsReceived`	Contagem de registros de usuário lógicos recebidos pelo núcleo da KPL para operações put. Não disponível no nível de fragmento. Nível de métrica: detalhado Unidade: Contagem
`UserRecordsPending`	Exemplo periódico de quantos registros de usuário estão pendentes atualmente. Um registro está pendente se está atualmente em buffer e esperando para ser enviado ou se foi enviado e está em trânsito para o serviço de backend. Não disponível no nível de fragmento. A KPL oferece um método dedicado para recuperar essa métrica global para permitir que os clientes gerenciem sua taxa de colocação. Nível de métrica: detalhado Unidade: Contagem
`UserRecordsPut`	Contagem de quantos registros de usuário lógicos foram colocados com êxito. A KPL retorna zero para registros com falha. Isso permite que a média ofereça a taxa de sucesso, que a contagem ofereça o total de tentativas e que a diferença entre a contagem e a soma ofereça o número de falhas. Nível de métrica: resumo Unidade: Contagem
`UserRecordsDataPut`	Bytes nos registros de usuário lógicos colocados com êxito.

Métrica	Description
	Nível de métrica: detalhado Unidade: bytes
`KinesisRecordsPut`	Contagem de registros do Kinesis Data Streams colocados com êxito (cada registro do Kinesis Data Streams pode conter vários registros de usuário). A KPL retorna zero para registros com falha. Isso permite que a média ofereça a taxa de sucesso, que a contagem ofereça o total de tentativas e que a diferença entre a contagem e a soma ofereça o número de falhas. Nível de métrica: resumo Unidade: Contagem
`KinesisRecordsDataPut`	Bytes em registros do Kinesis Data Streams. Nível de métrica: detalhado Unidade: bytes
`ErrorsByCode`	Contagem de cada tipo de código de erro. Apresenta uma dimensão adicional de `ErrorCode` , além de dimensões normais, como `StreamName` e `ShardId`. Nem todo erro pode ser rastreado até um fragmento. Os erros que não podem ser rastreados são apenas emitidos nos níveis globais ou de fluxo. Essa métrica captura informações sobre fatores como limitações, alterações no mapa de fragmentos, falhas internas, serviço indisponível, limites de tempo e assim por diante. Os erros de API do Kinesis Data Streams são contados uma vez por registro do Kinesis Data Streams. Vários registros de usuário em um registro do Kinesis Data Streams não geram várias contagens. Nível de métrica: resumo Unidade: Contagem

Métrica	Description
`AllErrors`	Isso é acionado pelos mesmos erros de Erros por código, mas não faz distinção entre tipos. Isso é útil como um monitor geral da taxa de erros sem exigir uma soma manual das contagens de todos os tipos diferentes de erros. Nível de métrica: resumo Unidade: Contagem
`RetriesPerRecord`	Número de tentativas realizadas por registro de usuário. Zero é emitido para registros que são bem-sucedidos em uma tentativa. Os dados são emitidos no momento em que um usuário termina (quando o registro é bem-sucedido ou não pode mais ser repetido). Se o registro time-to-live for um valor grande, essa métrica poderá ser significa tivamente atrasada. Nível de métrica: detalhado Unidade: Contagem
`BufferingTime`	O tempo entre a chegada de um registro de usuário na KPL e sua saída para o backend. Essas informações são transmitidas de volta ao usuário por registro, mas também estão disponíveis como estatística agregada. Nível de métrica: resumo Unidade: milissegundos
`Request Time`	O tempo necessário para executar `PutRecordsRequests` . Nível de métrica: detalhado Unidade: milissegundos

Métrica	Description
`User Records per Kinesis Record`	O número de registros de usuário lógicos agregados em um único registro do Kinesis Data Streams. Nível de métrica: detalhado Unidade: Contagem
`Amazon Kinesis Records per PutRecord sRequest`	O número de registros do Kinesis Data Streams agregados em um único `PutRecordsRequest` . Não disponível no nível de fragmento. Nível de métrica: detalhado Unidade: Contagem
`User Records per PutRecord sRequest`	O número total de registros de usuário contidos em um `PutRecord sRequest` . Isso é equivalente aproximadamente ao produto das últimas duas métricas. Não disponível no nível de fragmento. Nível de métrica: detalhado Unidade: Contagem

Segurança no Amazon Kinesis Data Streams

A segurança na nuvem AWS é a maior prioridade. Como AWS cliente, você se beneficiará de uma arquitetura de data center e rede criada para atender aos requisitos das organizações mais sensíveis à segurança.

A segurança é uma responsabilidade compartilhada entre você AWS e você. O modelo de responsabilidade compartilhada descreve isto como segurança da nuvem e segurança na nuvem.

- Segurança da nuvem — AWS é responsável por proteger a infraestrutura que executa AWS os serviços na AWS nuvem. AWS também fornece serviços que você pode usar com segurança. A eficácia da nossa segurança é regularmente testada e verificada por auditores de terceiros como parte dos Programas de conformidade da AWS. Para saber mais sobre os programas de conformidade que se aplicam ao Kinesis Data Streams, consulte Serviços da AWS no escopo por programa de conformidade.
- Segurança na nuvem — Sua responsabilidade é determinada pelo AWS serviço que você usa. Também existe a responsabilidade por outros fatores, incluindo a confidencialidade de dados, os requisitos da organização e as leis e regulamentos aplicáveis.

Esta documentação ajuda a entender como aplicar o modelo de responsabilidade compartilhada ao usar o Kinesis Data Streams. Os tópicos a seguir mostram como configurar o Kinesis Data Streams para atender aos seus objetivos de segurança e conformidade. Você também aprenderá a usar outros AWS serviços que podem ajudá-lo a monitorar e proteger seus recursos do Kinesis Data Streams.

Tópicos

- Proteção de dados no Amazon Kinesis Data Streams
- Controlar o acesso aos recursos do Amazon Kinesis Data Streams usando o IAM
- Validação de conformidade para o Amazon Kinesis Data Streams
- Resiliência no Amazon Kinesis Data Streams
- Segurança de infraestrutura no Kinesis Data Streams
- Práticas recomendadas de segurança para o Kinesis Data Streams

Proteção de dados no Amazon Kinesis Data Streams

A criptografia do lado do servidor usando chaves AWS Key Management Service (AWS KMS) facilita o cumprimento de requisitos rígidos de gerenciamento de dados, criptografando seus dados em repouso no Amazon Kinesis Data Streams.

> **Note**
>
> Se você precisar de módulos criptográficos validados pelo FIPS 140-2 ao acessar AWS por meio de uma interface de linha de comando ou de uma API, use um endpoint FIPS. Para ter mais informações sobre endpoints do FIPS disponíveis, consulte Federal Information Processing Standard (FIPS) 140-2.

Tópicos

- O que é criptografia no lado do servidor para o Kinesis Data Streams?
- Custos, regiões e considerações sobre desempenho
- Como começo a usar a criptografia no lado do servidor?
- Criar e usar chaves do KMS geradas pelo usuário
- Permissões para usar as chaves do KMS geradas pelo usuário
- Verificar e solucionar problemas de permissões de chaves do KMS
- Usar o Amazon Kinesis Data Streams com endpoints da VPC de interface

O que é criptografia no lado do servidor para o Kinesis Data Streams?

A criptografia do lado do servidor é um recurso do Amazon Kinesis Data Streams que criptografa automaticamente os dados antes que estejam em repouso AWS KMS usando uma chave mestra de cliente (CMK) especificada por você. Os dados são criptografados antes de serem gravados na camada de armazenamento do fluxo do Kinesis e descriptografados depois de recuperados do armazenamento. Como resultado, os dados são criptografados em repouso no serviço Kinesis Data Streams. Isso permite atender a requisitos normativos rígidos e aprimorar a segurança dos dados.

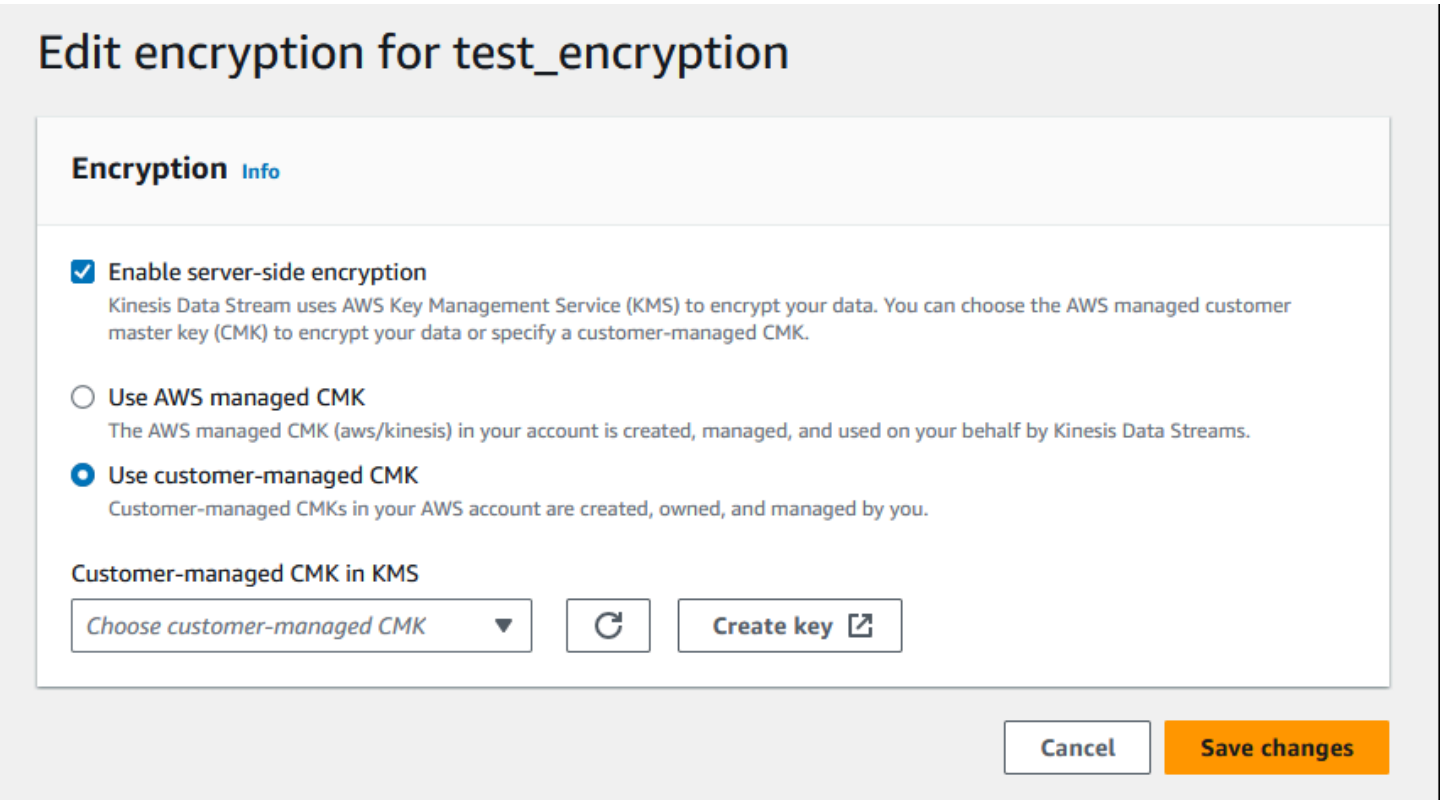
Com a criptografia no lado do servidor, seus produtores e consumidores de fluxos do Kinesis não precisam gerenciar chaves mestras ou operações de criptografia. Seus dados são criptografados automaticamente quando entram e saem do serviço Kinesis Data Streams, então seus dados em

repouso são criptografados. AWS KMS fornece todas as chaves mestras usadas pelo recurso de criptografia do lado do servidor. AWS KMS facilita o uso de uma CMK for Kinesis gerenciada AWS por, uma CMK AWS KMS especificada pelo usuário ou uma chave mestra importada para o serviço. AWS KMS

> **Note**
>
> Criptografia no lado do servidor criptografa os dados de entrada somente após a criptografia ser ativada. Os dados preexistentes em um fluxo não criptografado não são criptografados após a ativação da criptografia no lado do servidor.

Ao criptografar seus fluxos de dados e compartilhar o acesso a outros diretores, você deve conceder permissão na política de chaves da AWS KMS chave e nas políticas do IAM na conta externa. Para ter mais informações, consulte o tópico sobre como Permitir que usuários de outras contas utilizem uma chave do KMS.

Se você tiver habilitado a criptografia do lado do servidor para um fluxo de dados com chave KMS AWS gerenciada e quiser compartilhar o acesso por meio de uma política de recursos, deverá passar a usar a chave gerenciada pelo cliente (CMK), conforme mostrado a seguir:



Além disso, deve-se permitir que as entidades principais de compartilhamento tenham acesso à sua CMK, usando os recursos de compartilhamento entre contas do KMS. Certifique-se também de fazer a alteração nas políticas do IAM para as entidades principais de compartilhamento. Para ter mais informações, consulte o tópico sobre como Permitir que usuários de outras contas utilizem uma chave do KMS.

Custos, regiões e considerações sobre desempenho

Ao aplicar a criptografia do lado do servidor, você está sujeito ao uso AWS KMS da API e aos custos principais. Ao contrário das chaves mestras do KMS personalizadas, a chave mestra do cliente (CMK) do `(Default) aws/kinesis` é oferecida gratuitamente. No entanto, os custos de uso de API que o Amazon Kinesis Data Streams gera em seu nome ainda devem ser pagos.

Os custos de uso da API aplicam-se a cada CMK, incluindo as personalizadas. O Kinesis Data Streams chama o AWS KMS aproximadamente a cada cinco minutos quando está mudando a chave de dados. Em um mês de 30 dias, o custo total das chamadas de AWS KMS API iniciadas por um stream do Kinesis deve ser inferior a alguns dólares. Esse custo aumenta com o número de credenciais de usuário que você usa em seus produtores e consumidores de dados, pois cada credencial de usuário exige uma chamada de API exclusiva para. AWS KMS Ao usar um perfil do IAM para autenticação, cada chamada para assumir um perfil resulta em credenciais exclusivas ao usuário. Para economizar gastos com o KMS, armazene em cache as credenciais de usuário que são retornadas pela chamada de função assumida.

Veja a seguir a descrição dos custos por recurso:

Chaves

- A CMK para Kinesis gerenciada AWS por (alias `aws/kinesis` =) é gratuita.
- As chaves do KMS geradas pelo usuário estão sujeitas aos custos de chave do KMS. Para obter mais informações, consulte AWS Key Management Service Pricing.

Os custos de uso da API aplicam-se a cada CMK, incluindo as personalizadas. O Kinesis Data Streams chama o KMS aproximadamente a cada cinco minutos quando está mudando a chave de dados. Em um mês de 30 dias, o custo total das chamadas à API do KMS que são iniciadas por um fluxo de dados do Kinesis deve ser apenas alguns dólares. Observe que esse custo aumenta de acordo com o número de credenciais de usuário que você usa em seus produtores e consumidores de dados, pois cada credencial de usuário exige uma chamada de API exclusiva para AWS o KMS. Quando você usa a função do IAM para autenticação, cada uma assume-role-call resultará em

credenciais de usuário exclusivas e talvez você queira armazenar em cache as credenciais do usuário retornadas pela assume-role-call para economizar custos de KMS.

Uso da API do KMS

Para cada stream criptografado, ao ler o TIP e usar uma única chave de account/user acesso IAM entre leitores e gravadores, o serviço Kinesis liga para o AWS KMS serviço aproximadamente 12 vezes a cada 5 minutos. Não ler o TIP pode levar a um aumento no número de chamadas para o AWS KMS serviço. As solicitações de API para gerar novas chaves de criptografia de dados estão sujeitas aos custos AWS KMS de uso. Para obter mais informações, consulte AWS Key Management Service Pricing: Usage.

Disponibilidade da criptografia no lado do servidor por região

Atualmente, a criptografia do lado do servidor dos Kinesis Streams está disponível em todas as regiões compatíveis com o Kinesis Data Streams, incluindo (Oeste dos EUA) e as regiões da China. AWS GovCloud Para obter mais informações sobre regiões compatíveis com o Kinesis Data https://docs.aws.amazon.com/general/latest/gr/akStreams, consulte .html.

Considerações sobre desempenho

Devido à sobrecarga de serviço da aplicação de criptografia, a aplicação de criptografia do lado do servidor aumenta a latência típica de `PutRecord`, `PutRecords` e `GetRecords` em menos de 100 µs.

Como começo a usar a criptografia no lado do servidor?

A maneira mais fácil de começar a usar a criptografia do lado do servidor é usar a Console de gerenciamento da AWS e a chave de serviço do Amazon Kinesis KMS,. `aws/kinesis`

O procedimento a seguir demonstra como habilitar a criptografia no lado do servidor para um fluxo do Kinesis.

Habilitar a criptografia no lado do servidor para um fluxo do Kinesis

1. Faça login no Console de gerenciamento da AWS e abra o console do Amazon Kinesis Data Streams.
2. Crie ou selecione um fluxo do Kinesis no Console de gerenciamento da AWS.
3. Selecione a guia Detalhes.

4. Em Criptografia no lado do servidor, selecione Editar.
5. A não ser que seja preferível usar uma chave mestra do KMS gerada pelo usuário, selecione a chave mestra do KMS (Padrão) aws/kinesis. Essa é a chave mestra do KMS gerada pelo serviço Kinesis. Selecione Habilitado e, em seguida, selecione Salvar.

   > **Note**
   >
   > A chave mestra padrão do serviço Kinesis é gratuita, no entanto, as chamadas de API feitas pelo Kinesis para o AWS KMS serviço estão sujeitas aos custos de uso do KMS.

6. O fluxo passa por um estado pendente. Assim que o fluxo voltar a um estado ativo com a criptografia habilitada, todos os dados recebidos gravados no fluxo serão criptografados usando a chave mestra selecionada do KMS.
7. Para desativar a criptografia do lado do servidor, escolha Desativado na criptografia do lado do servidor no e, em seguida, escolha Salvar Console de gerenciamento da AWS.

Criar e usar chaves do KMS geradas pelo usuário

Esta seção descreve como criar e usar as próprias chaves do KMS em vez de usar a chave mestra administrada pelo Amazon Kinesis.

Criar chaves do KMS geradas pelo usuário

Para obter instruções sobre como criar as próprias chaves, consulte Criar chaves no Guia do desenvolvedor do AWS Key Management Service . Depois de criar chaves para sua conta, o serviço Kinesis Data Streams as retorna na lista de chaves mestras do KMS.

Usar chaves do KMS geradas pelo usuário

Depois que as permissões corretas forem aplicadas aos seus consumidores, produtores e administradores, você poderá usar chaves KMS personalizadas em sua própria AWS conta ou em outra AWS conta. Todas as chaves mestras do KMS em sua conta aparecem na lista Chave mestra do KMS no Console de gerenciamento da AWS.

Para usar chaves mestras do KMS personalizadas localizadas em outra conta, são necessárias permissões para usar essas chaves. Também deve-se especificar o ARN da chave mestra do KMS na caixa de entrada ARN no Console de gerenciamento da AWS.

Permissões para usar as chaves do KMS geradas pelo usuário

Antes de usar a criptografia do lado do servidor com uma chave KMS gerada pelo usuário, você deve configurar AWS KMS políticas de chaves para permitir a criptografia de fluxos e a criptografia e descriptografia de registros de fluxo. Para obter exemplos e mais informações sobre AWS KMS permissões, consulte Permissões da API AWS KMS: referência de ações e recursos.

> **Note**
>
> O uso da chave padrão do serviço para criptografia não requer a aplicação de permissões personalizadas do IAM.

Antes de usar chaves mestras do KMS geradas pelo usuário, verifique se seus produtores e consumidores de fluxos do Kinesis (entidades principais do IAM) são usuários na política de chaves mestras do KMS. Caso contrário, as gravações e as leituras de um fluxo falharão, o que pode resultar, em última análise, em perda de dados, processamento atrasado, ou travamento de aplicativos. É possível gerenciar permissões para chaves do KMS usando políticas do IAM. Para obter mais informações, consulte Como usar políticas do IAM com o AWS KMS.

Exemplo de permissões de produtor

Seus produtores de fluxos do Kinesis devem ter a permissão `kms:GenerateDataKey`.

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
        "Effect": "Allow",
        "Action": [
            "kms:GenerateDataKey"
        ],
        "Resource": "arn:aws:kms:us-
west-2:123456789012:key/1234abcd-12ab-34cd-56ef-1234567890ab"
    },
    {
        "Effect": "Allow",
```

```
        "Action": [
            "kinesis:PutRecord",
            "kinesis:PutRecords"
        ],
        "Resource": "arn:aws:kinesis:*:123456789012:MyStream"
    }
  ]
}
```

Exemplo de permissões de consumidor

Seus consumidores de fluxos do Kinesis devem ter a permissão `kms:Decrypt`.

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
        "Effect": "Allow",
        "Action": [
            "kms:Decrypt"
        ],
        "Resource": "arn:aws:kms:us-
west-2:123456789012:key/1234abcd-12ab-34cd-56ef-1234567890ab"
    },
    {
        "Effect": "Allow",
        "Action": [
            "kinesis:GetRecords",
            "kinesis:DescribeStream"
        ],
        "Resource": "arn:aws:kinesis:*:123456789012:MyStream"
    }
  ]
}
```

Amazon Managed Service para Apache Flink e AWS Lambda use funções para consumir streams do Kinesis. Adicione a permissão `kms:Decrypt` às funções que esses consumidores usam.

Permissões de administrador de fluxo

Os administradores de fluxos do Kinesis precisam ter autorização para chamar `kms:List*` e `kms:DescribeKey*`.

Verificar e solucionar problemas de permissões de chaves do KMS

Depois de habilitar a criptografia em um stream do Kinesis, recomendamos que você monitore o sucesso de suas `getRecords` chamadas`putRecord`,`putRecords`, e usando as seguintes métricas da Amazon CloudWatch :

- `PutRecord.Success`
- `PutRecords.Success`
- `GetRecords.Success`

Para obter mais informações, consulte Monitorar o Kinesis Data Streams.

Usar o Amazon Kinesis Data Streams com endpoints da VPC de interface

É possível usar um endpoint da VPC de interface para impedir que o tráfego entre sua Amazon VPC e o Kinesis Data Streams saia da rede da Amazon. Os endpoints VPC de interface não exigem um gateway de internet, dispositivo NAT, conexão VPN ou conexão. Direct Connect Os endpoints de VPC de interface são alimentados por AWS PrivateLink uma AWS tecnologia que permite a comunicação privada entre AWS serviços usando uma interface de rede elástica com privacidade em IPs sua Amazon VPC. Para obter mais informações, consulte Amazon Virtual Private Cloud e Interface VPC Endpoints ().AWS PrivateLink

Tópicos

- Usar os endpoints da VPC de interface para o Kinesis Data Streams
- Controlar o acesso a endpoints da VPC para o Kinesis Data Streams
- Disponibilidade de políticas de endpoints da VPC para o Kinesis Data Streams

Usar os endpoints da VPC de interface para o Kinesis Data Streams

Para começar, você não precisa alterar as configurações para os fluxos, produtores ou consumidores. Crie um endpoint da VPC de interface para que o Kinesis Data Streams comece o fluxo de tráfego de e para os recursos da Amazon VPC pelo endpoint da VPC de interface. Os

endpoints da VPC de interface habilitada para FIPS estão disponíveis para as regiões dos EUA. Para obter mais informações, consulte Criação de um endpoint de interface.

A Amazon Kinesis Producer Library (KPL) e a Kinesis Consumer Library (KCL) chamam serviços como AWS Amazon e Amazon CloudWatch DynamoDB usando endpoints públicos ou endpoints VPC de interface privada, os que estiverem em uso. Por exemplo, se a aplicação da KCL estiver sendo executada em uma VPC com interface do DynamoDB com endpoints da VPC habilitados, as chamadas entre o DynamoDB e a aplicação da KCL passarão pelo endpoint da VPC de interface.

Controlar o acesso a endpoints da VPC para o Kinesis Data Streams

As políticas de endpoint da VPC permitem que você controle o acesso anexando uma política a um endpoint da VPC ou usando campos adicionais em uma política anexada a um usuário, perfil ou grupo do IAM para restringir o acesso somente por meio do endpoint da VPC especificado. Use essas políticas para restringir o acesso a fluxos específicos a um endpoint da VPC especificado quando tais políticas forem usadas em conjunto com as políticas do IAM para conceder acesso somente a ações de fluxo de dados do Kinesis por meio do endpoint da VPC especificado.

Veja a seguir exemplos de políticas de endpoint para acessar fluxos de dados do Kinesis.

- Exemplo de política de VPC: acesso somente leitura — esse exemplo de política pode ser anexado a um VPC endpoint. (Para obter mais informações, consulte Como controlar o acesso aos recursos da Amazon VPC). Ele restringe as ações a somente listar e descrever um fluxo de dados do Kinesis por meio do VPC endpoint ao qual está anexado.

```
{
  "Statement": [
    {
      "Sid": "ReadOnly",
      "Principal": "*",
      "Action": [
        "kinesis:List*",
        "kinesis:Describe*"
      ],
      "Effect": "Allow",
      "Resource": "*"
    }
  ]
}
```

- Exemplo de política de VPC: restringir o acesso a um fluxo de dados específico do Kinesis — esse exemplo de política pode ser anexado a um VPC endpoint. Ele restringe o acesso a um fluxo de dados específico por meio do VPC endpoint ao qual está anexado.

```
{
  "Statement": [
    {
      "Sid": "AccessToSpecificDataStream",
      "Principal": "*",
      "Action": "kinesis:*",
      "Effect": "Allow",
      "Resource": "arn:aws:kinesis:us-east-1:123456789012:stream/MyStream"
    }
  ]
}
```

- Exemplo de política do IAM: restringir o acesso a um fluxo específico apenas de determinado endpoint da VPC. Esse exemplo de política pode ser anexado a um usuário, um perfil ou um grupo do IAM. Ele restringe o acesso a um fluxo de dados especificado do Kinesis para ocorrer somente em determinado VPC endpoint.

  JSON

```
{
   "Version":"2012-10-17",
   "Statement": [
      {
         "Sid": "AccessFromSpecificEndpoint",
         "Action": "kinesis:*",
         "Effect": "Deny",
         "Resource": "arn:aws:kinesis:us-east-1:123456789012:stream/MyStream",
         "Condition": { "StringNotEquals" : { "aws:sourceVpce":
 "vpce-11aa22bb" } }
      }
   ]
}
```

Disponibilidade de políticas de endpoints da VPC para o Kinesis Data Streams

Endpoints da VPC de interface do Kinesis Data Streams com políticas são compatíveis com as seguintes regiões:

- Europa (Paris)
- Europa (Irlanda)
- Leste dos EUA (Norte da Virgínia)
- Europa (Estocolmo)
- Leste dos EUA (Ohio)
- Europa (Frankfurt)
- América do Sul (São Paulo)
- Europa (Londres)
- Ásia-Pacífico (Tóquio)
- Oeste dos EUA (N. da Califórnia)
- Ásia-Pacífico (Singapura)
- Ásia-Pacífico (Sydney)
- China (Pequim)
- China (Ningxia)
- Ásia-Pacífico (Hong Kong)
- Oriente Médio (Bahrein)
- Oriente Médio (Emirados Árabes Unidos)
- Europa (Milão)
- África (Cidade do Cabo)
- Ásia-Pacífico (Mumbai)
- Ásia-Pacífico (Seul)
- Canadá (Central)
- Oeste dos EUA (Oregon), exceto usw2-az4
- AWS GovCloud (Leste dos EUA)
- AWS GovCloud (Oeste dos EUA)
- Ásia-Pacífico (Osaka)

- Europa (Zurique)
- Ásia-Pacífico (Hyderabad)

Controlar o acesso aos recursos do Amazon Kinesis Data Streams usando o IAM

AWS Identity and Access Management (IAM) permite que você faça o seguinte:

- Crie usuários e grupos em sua AWS conta
- Atribua credenciais de segurança exclusivas a cada usuário em sua conta AWS
- Controle as permissões de cada usuário para realizar tarefas usando AWS recursos
- Permita que os usuários de outra AWS conta compartilhem seus AWS recursos
- Crie funções para sua AWS conta e defina os usuários ou serviços que podem assumi-las
- Use identidades existentes para sua empresa para conceder permissões para realizar tarefas usando recursos AWS

Ao usar o IAM com o Kinesis Data Streams, é possível controlar se os usuários de sua organização podem executar uma tarefa usando ações específicas da API do Kinesis Data Streams e se podem usar recursos específicos da AWS .

Se você estiver desenvolvendo um aplicativo usando a Kinesis Client Library (KCL), sua política deve incluir permissões para o Amazon DynamoDB e a Amazon; CloudWatch o KCL usa o DynamoDB para rastrear informações de estado do aplicativo e enviar métricas de KCL para você. CloudWatch CloudWatch Para obter mais informações sobre o recurso KCL, consulte Desenvolver aplicações de consumo da KCL 1.x.

Para obter mais informações sobre IAM, consulte o seguinte:

- AWS Identity and Access Management (IAM)
- Introdução ao IAM
- Guia do usuário do IAM

Para obter mais informações sobre o IAM e o DynamoDB, consulte Usar o IAM para controlar o acesso a recursos do Amazon DynamoDB no Guia do desenvolvedor do Amazon DynamoDB.

Para obter mais informações sobre o IAM e a Amazon CloudWatch, consulte Controlando o acesso do usuário à sua AWS conta no Guia CloudWatch do usuário da Amazon.

Conteúdo

- Sintaxe da política
- Ações para o Kinesis Data Streams
- Nomes de recursos da Amazon (ARNs) para Kinesis Data Streams
- Exemplo de políticas para o Kinesis Data Streams
- Compartilhamento do fluxo de dados com outra conta
- Configurar uma AWS Lambda função para ler do Kinesis Data Streams em outra conta
- Compartilhamento de acesso usando políticas baseadas em recursos

Sintaxe da política

A política do IAM é um documento JSON que consiste em uma ou mais declarações. Cada instrução é estruturada da seguinte maneira:

```
{
  "Statement":[{
    "Effect":"effect",
    "Action":"action",
    "Resource":"arn",
    "Condition":{
      "condition":{
        "key":"value"
        }
      }
    }
  ]
}
```

Existem vários elementos que compõem uma instrução:

- Effect: o efeito pode ser `Allow` ou `Deny`. Por padrão, os usuários do IAM não têm permissão para usar recursos e ações da API. Por isso, todas as solicitações são negadas. Uma permissão explícita substitui o padrão. Uma negação explícita substitui todas as permissões.
- Ação: é a ação de API específica para a qual a permissão esteja sendo concedida ou negada.

- Recurso: o recurso afetado pela ação. Para especificar um recurso na declaração, é necessário usar o nome do recurso da Amazon (ARN).
- Condição: condições são opcionais. Elas podem ser usadas para controlar quando as políticas entrarão em vigor.

Ao criar e gerenciar políticas do IAM, pode ser conveniente usar o gerador de políticas do IAM e o simulador de políticas do IAM.

Ações para o Kinesis Data Streams

Em uma declaração de política do IAM, é possível especificar qualquer ação de API de qualquer serviço que dê suporte ao IAM. Para o Kinesis Data Streams, use o seguinte prefixo com o nome da ação da API: `kinesis:`. Por exemplo: `kinesis:CreateStream`, `kinesis:ListStreams` e `kinesis:DescribeStreamSummary`.

Para especificar várias ações em uma única instrução, separe-as com vírgulas, como segue:

```
"Action": ["kinesis:action1", "kinesis:action2"]
```

Também é possível especificar várias ações usando asteriscos. Por exemplo, é possível especificar todas as ações cujo nome começa com a palavra "Obter", conforme o seguinte:

```
"Action": "kinesis:Get*"
```

Para especificar todas as operações do Kinesis Data Streams, use o curinga *, como a seguir:

```
"Action": "kinesis:*"
```

Para obter a lista completa das ações da API do Kinesis Data Streams, consulte a Referência de API do Amazon Kinesis.

Nomes de recursos da Amazon (ARNs) para Kinesis Data Streams

Cada declaração de política do IAM se aplica aos recursos que você especifica usando seus ARNs.

Use o seguinte formato de recursos do ARN para os fluxos de dados do Kinesis:

```
arn:aws:kinesis:region:account-id:stream/stream-name
```

Por exemplo:

```
"Resource": arn:aws:kinesis:*:111122223333:stream/my-stream
```

Exemplo de políticas para o Kinesis Data Streams

As políticas de exemplo a seguir demonstram como é possível controlar o acesso do usuário aos fluxos de dados do Kinesis.

Example 1: Allow users to get data from a stream

Example

Esta política permite que um usuário ou grupo execute as operações `DescribeStreamSummary`, `GetShardIterator` e `GetRecords` no fluxo especificado e `ListStreams` em qualquer stream. Esta política pode ser aplicada a usuários que devem conseguir obter dados de um fluxo específico.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "kinesis:Get*",
                "kinesis:DescribeStreamSummary"
            ],
            "Resource": [
            "arn:aws:kinesis:us-east-1:111122223333:stream/stream1"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "kinesis:ListStreams"
            ],
            "Resource": [
                "*"
            ]
        }
```

```
    ]
}
```

Example 2: Allow users to add data to any stream in the account

Example

Esta política permite que um usuário ou grupo use a operação `PutRecord` com qualquer um dos streams da conta. Esta política pode ser aplicada a usuários que devem conseguir adicionar registros de dados a todos os streams em uma conta.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "kinesis:PutRecord"
            ],
            "Resource": [
                "arn:aws:kinesis:us-east-1:111122223333:stream/*"
            ]
        }
    ]
}
```

Example 3: Allow any Kinesis Data Streams action on a specific stream

Example

Esta política permite que um usuário ou grupo use qualquer operação do Kinesis Data Streams no fluxo especificado. Esta política poderia ser aplicada a usuários que devem ter controle administrativo por meio de um fluxo específico.

JSON

```
{
```

```
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "kinesis:*",
            "Resource": [
            "arn:aws:kinesis:us-east-1:111122223333:stream/stream1"
            ]
        }
    ]
}
```

Example 4: Allow any Kinesis Data Streams action on any stream

Example

Esta política permite que um usuário ou grupo use qualquer operação do Kinesis Data Streams em qualquer fluxo em uma conta. Como esta política concede acesso total a todos os fluxos, deve-se restringi-la somente aos administradores.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "kinesis:*",
            "Resource": [
                "arn:aws:kinesis:*:111122223333:stream/*"
            ]
        }
    ]
}
```

Compartilhamento do fluxo de dados com outra conta

> **Note**
>
> Atualmente, a Kinesis Producer Library não é compatível com a a especificação de um ARN de fluxo ao gravar em um fluxo de dados. Use o AWS SDK se quiser gravar em um stream de dados entre contas.

Anexe uma política baseada em recursos ao fluxo de dados para conceder acesso a outra conta, a um usuário do IAM ou a um perfil do IAM. As políticas baseadas em recursos são documentos de política JSON anexadas a um recurso, como um fluxo de dados. Essas políticas concedem permissão para a entidade principal especificada executar ações específicas nesse recurso e definem sob quais condições isso se aplica. Uma política pode ter várias declarações. É necessário especificar um principal em uma política baseada em recursos. Os diretores podem incluir contas, usuários, funções, usuários federados ou AWS serviços. É possível configurar políticas no console do Kinesis Data Streams, na API ou no SDK.

Observe que compartilhar o acesso a consumidores registrados, como o Enhanced Fan Out, exige uma política tanto no ARN do fluxo de dados quanto no ARN do consumidor.

Ativação do acesso entre contas

Para permitir o acesso entre contas, é possível especificar uma conta inteira ou as entidades do IAM em outra conta como a entidade principal em uma política baseada em atributo. Adicionar uma entidade principal entre contas à política baseada em recurso é apenas metade da tarefa de estabelecimento da relação de confiança. Quando o principal e o recurso estão em AWS contas separadas, você também deve usar uma política baseada em identidade para conceder ao principal acesso ao recurso. No entanto, se uma política baseada em recurso conceder acesso a uma entidade principal na mesma conta, nenhuma política baseada em identidade adicional será necessária.

Para obter mais informações sobre como usar políticas baseadas em recursos para acesso entre contas, consulte Acesso a recursos entre contas no IAM.

Os administradores de fluxo de dados podem usar AWS Identity and Access Management políticas para especificar quem tem acesso ao quê. Ou seja, qual entidade principal pode executar ações em quais recursos e em que condições. O elemento `Action` de uma política JSON descreve as

ações que podem ser usadas para permitir ou negar acesso em uma política. As ações de política geralmente têm o mesmo nome da operação de AWS API associada.

Ações do Kinesis Data Streams que podem ser compartilhadas:

Ação	Nível de acesso
DescribeStreamConsumer	Consumidor
DescribeStreamSummary	Fluxo de dados
GetRecords	Fluxo de dados
GetShardIterator	Fluxo de dados
ListShards	Fluxo de dados
PutRecord	Fluxo de dados
PutRecords	Fluxo de dados
SubscribeToShard	Consumidor

Veja a seguir exemplos do uso de uma política baseada em recursos para conceder acesso entre contas ao fluxo de dados ou ao consumidor registrado.

Para realizar uma ação entre contas, é necessário especificar o ARN do fluxo para acesso ao fluxo de dados e o ARN do consumidor para o acesso do consumidor registrado.

Exemplo de políticas baseadas em recursos para fluxo de dados do Kinesis

Compartilhar um consumidor registrado envolve uma política de fluxo de dados e uma política de consumidor devido às ações necessárias.

> **Note**
>
> Veja os seguintes exemplos de valores válidos para `Principal`:
>
> - `{"AWS": "123456789012"}`

- Usuário do IAM: `{"AWS": "arn:aws:iam::123456789012:user/user-name"}`
- Perfil do IAM: `{"AWS":["arn:aws:iam::123456789012:role/role-name"]}`
- Várias entidades principais (pode ser uma combinação de contas, usuários, perfis): `{"AWS":["123456789012", "123456789013", "arn:aws:iam::123456789012:user/user-name"]}`

Example 1: Write access to the data stream

Example

JSON

```
{
    "Version":"2012-10-17",
    "Id": "__default_write_policy_ID",
    "Statement": [
        {
            "Sid": "writestatement",
            "Effect": "Allow",
            "Principal": {
                "AWS": "Account12345"
            },
            "Action": [
                "kinesis:DescribeStreamSummary",
                "kinesis:ListShards",
                "kinesis:PutRecord",
                "kinesis:PutRecords"
            ],
            "Resource": "arn:aws:kinesis:us-east-2:123456789012:stream/
datastreamABC"
        }
    ]
}
```

Example 2: Read access to the data stream

Example

JSON

```
{
    "Version":"2012-10-17",
    "Id": "__default_sharedthroughput_read_policy_ID",
    "Statement": [
        {
            "Sid": "sharedthroughputreadstatement",
            "Effect": "Allow",
            "Principal": {
                "AWS": "Account12345"
            },
            "Action": [
                "kinesis:DescribeStreamSummary",
                "kinesis:ListShards",
                "kinesis:GetRecords",
                "kinesis:GetShardIterator"
            ],
            "Resource": "arn:aws:kinesis:us-east-2:123456789012:stream/
datastreamABC"
        }
    ]
}
```

Example 3: Share enhanced fan-out read access to a registered consumer

Example

Declaração de política de fluxo de dados:

JSON

```
{
    "Version":"2012-10-17",
    "Id": "__default_sharedthroughput_read_policy_ID",
    "Statement": [
```

```
        {
            "Sid": "consumerreadstatement",
            "Effect": "Allow",
            "Principal": {
                "AWS": "arn:aws:iam::111122223333:role/role-name"
            },
            "Action": [
                "kinesis:DescribeStreamSummary",
                "kinesis:ListShards"
            ],
            "Resource": "arn:aws:kinesis:us-east-2:123456789012:stream/
datastreamABC"
        }
    ]
}
```

Declaração de política de consumidor:

JSON

```
{
    "Version":"2012-10-17",
    "Id": "__default_efo_read_policy_ID",
    "Statement": [
        {
            "Sid": "eforeadstatement",
            "Effect": "Allow",
            "Principal": {
                "AWS": "arn:aws:iam::111122223333:role/role-name"
            },
            "Action": [
                "kinesis:DescribeStreamConsumer",
                "kinesis:SubscribeToShard"
            ],
            "Resource": "arn:aws:kinesis:us-east-2:123456789012:stream/
datastreamABC/consumer/consumerDEF:1674696300"
        }
    ]
}
```

Não há suporte para o caractere curinga (*) no campo de ações ou entidade principal, a fim de manter o princípio do privilégio mínimo.

Gerenciamento da política para o fluxo de dados de forma programática

Além do Console de gerenciamento da AWS, o Kinesis Data Streams tem três APIS para gerenciar sua política de stream de dados:

- PutResourcePolicy
- GetResourcePolicy
- DeleteResourcePolicy

Use `PutResourePolicy` para anexar ou substituir uma política de um fluxo de dados ou consumidor. Use `GetResourcePolicy` para verificar e visualizar uma política do fluxo de dados ou consumidor especificado. Use `DeleteResourcePolicy` para excluir uma política do fluxo de dados ou consumidor especificado.

Limites de políticas

As políticas de recursos do Kinesis Data Streams têm as restrições a seguir.

- Curingas (*) não são aceitos para ajudar a impedir que o acesso amplo seja concedido por meio de políticas de recursos diretamente vinculadas a um fluxo de dados ou consumidor registrado. Além disso inspecione com cuidado as seguintes políticas para garantir que elas não concedam acesso amplo:
  - Políticas baseadas em identidade vinculadas aos AWS diretores associados (por exemplo, funções do IAM)
  - Políticas baseadas em recursos anexadas aos AWS recursos associados (por exemplo, chaves AWS Key Management Service KMS)
- AWS Os diretores de serviço não recebem suporte para diretores para evitar possíveis deputados confusos.
- Não há suporte para entidades principais federadas.
- Usuários canônicos não IDs são suportados.
- O tamanho da política não pode exceder 20 kB.

Compartilhamento de acesso a dados criptografados

Se você habilitou a criptografia do lado do servidor para um fluxo de dados com chave KMS AWS gerenciada e deseja compartilhar o acesso por meio de uma política de recursos, deve passar a usar a chave gerenciada pelo cliente (CMK). Para obter mais informações, consulte O que é criptografia no lado do servidor para o Kinesis Data Streams?. Além disso, deve-se permitir que as entidades principais de compartilhamento tenham acesso à sua CMK, usando os recursos de compartilhamento entre contas do KMS. Certifique-se também de fazer a alteração nas políticas do IAM para as entidades principais de compartilhamento. Para ter mais informações, consulte o tópico sobre como Permitir que usuários de outras contas utilizem uma chave do KMS.

Configurar uma AWS Lambda função para ler do Kinesis Data Streams em outra conta

Para ver um exemplo de como configurar uma função do Lambda para ler do Kinesis Data Streams em outra conta, consulte Compartilhe o acesso com funções entre contas AWS Lambda.

Compartilhamento de acesso usando políticas baseadas em recursos

> **Note**
>
> Atualizar uma política existente baseada em recursos significa substituir a atual, portanto, certifique-se de incluir todas as informações necessárias em sua nova política.

Compartilhe o acesso com funções entre contas AWS Lambda

Operador do Lambda

1. Acesse o console do IAM para criar uma função do IAM que será usada como a função de execução do Lambda para sua AWS Lambda função. Inclua a política do IAM gerenciada `AWSLambdaKinesisExecutionRole` que tem as permissões de invocação necessárias do Kinesis Data Streams e do Lambda. Essa política também concede acesso a todos os possíveis recursos do Kinesis Data Streams aos quais você possa ter acesso.
2. No AWS Lambda console, crie uma AWS Lambda função para processar registros em um stream de dados do Kinesis Data Streams e, durante a configuração da função de execução, escolha a função que você criou na etapa anterior.

3. Forneça o perfil de execução ao proprietário do recurso do Kinesis Data Streams para configurar a política de recursos.
4. Conclua a configuração da função do Lambda.

Proprietário do recurso do Kinesis Data Streams

1. Obtenha o perfil de execução entre contas do Lambda que invocará a função do Lambda.
2. No console do Amazon Kinesis Data Streams, escolha o fluxo de dados. Escolha a guia Compartilhamento de fluxo de dados e clique no botão Criar política de compartilhamento para iniciar o editor visual de políticas. Para compartilhar um consumidor registrado em um fluxo de dados, escolha o consumidor e, em seguida, escolha Criar política de compartilhamento. Também é possível escrever a política JSON diretamente.
3. Especifique o perfil de execução entre contas do Lambda como a entidade principal e as ações exatas do Kinesis Data Streams com as quais o acesso seja compartilhado. Certifique-se de incluir a ação `kinesis:DescribeStream`. Para obter mais informações sobre exemplos de políticas de recursos do Kinesis Data Streams, consulte Exemplo de políticas baseadas em recursos para fluxo de dados do Kinesis.
4. Escolha Criar política ou use o PutResourcePolicypara anexar a política ao seu recurso.

Compartilhamento de acesso com consumidores de várias contas da KCL

- Se estiver usando a KCL 1.x, verifique se é a versão KCL 1.15.0 ou superior.
- Se estiver usando a KCL 2.x, verifique se é a versão KCL 2.5.3 ou superior.

Operador da KCL

1. Forneça ao proprietário do recurso seu usuário do IAM ou o perfil do IAM que executará aplicação KCL.
2. Peça ao proprietário do recurso o fluxo de dados ou o ARN do consumidor.
3. Certifique-se de especificar o ARN do fluxo fornecido como parte da configuração da KCL.
    - Para KCL 1.x: use o KinesisClientLibConfigurationconstrutor e forneça o ARN do fluxo.
    - Para o KCL 2.x: você pode fornecer apenas o ARN do stream ou para a biblioteca de cliente StreamTrackerdo Kinesis. ConfigsBuilder Para StreamTracker, forneça o ARN do stream e o Epoch de criação a partir da tabela de lease do DynamoDB que é gerada pela biblioteca. Se

você quiser ler de um consumidor registrado compartilhado, como o Enhanced Fan-Out, use StreamTracker e também forneça o ARN do consumidor.

Proprietário do recurso do Kinesis Data Streams

1. Obtenha o usuário do IAM ou o perfil do IAM de várias contas que executará a aplicação KCL.
2. No console do Amazon Kinesis Data Streams, escolha o fluxo de dados. Escolha a guia Compartilhamento de fluxo de dados e clique no botão Criar política de compartilhamento para iniciar o editor visual de políticas. Para compartilhar um consumidor registrado em um fluxo de dados, escolha o consumidor e, em seguida, escolha Criar política de compartilhamento. Você também pode escrever a política JSON diretamente.
3. Especifique o usuário do IAM ou o perfil do IAM da aplicação KCL de várias contas como entidade principal e as ações exatas do Kinesis Data Streams às quais o acesso seja compartilhado. Para obter mais informações sobre exemplos de políticas de recursos do Kinesis Data Streams, consulte Exemplo de políticas baseadas em recursos para fluxo de dados do Kinesis.
4. Escolha Criar política ou use o PutResourcePolicypara anexar a política ao seu recurso.

Compartilhamento de acesso a dados criptografados

Se você habilitou a criptografia do lado do servidor para um fluxo de dados com chave KMS AWS gerenciada e deseja compartilhar o acesso por meio de uma política de recursos, deve passar a usar a chave gerenciada pelo cliente (CMK). Para obter mais informações, consulte O que é criptografia no lado do servidor para o Kinesis Data Streams?. Além disso, deve-se permitir que as entidades principais de compartilhamento tenham acesso à sua CMK, usando os recursos de compartilhamento entre contas do KMS. Certifique-se também de fazer a alteração nas políticas do IAM para as entidades principais de compartilhamento. Para ter mais informações, consulte o tópico sobre como Permitir que usuários de outras contas utilizem uma chave do KMS.

Validação de conformidade para o Amazon Kinesis Data Streams

Auditores terceirizados avaliam a segurança e a conformidade do Amazon Kinesis Data Streams como parte de vários programas de conformidade. AWS Isso inclui SOC, PCI, FedRAMP, HIPAA e outros.

Para obter uma lista de AWS serviços no escopo de programas de conformidade específicos, consulte AWS Serviços no escopo por programa de conformidade. Para obter informações gerais, consulte Programas de conformidade da AWS.

Você pode baixar relatórios de auditoria de terceiros usando AWS Artifact. Para obter mais informações, consulte Baixando relatórios no AWS Artifact.

Sua responsabilidade de conformidade ao usar o Kinesis Data Streams é determinada pela confidencialidade dos dados, pelos objetivos de conformidade da empresa e pelos regulamentos e leis aplicáveis. Se seu uso do Kinesis Data Streams estiver sujeito à conformidade com padrões como HIPAA, PCI ou AWS FedRAMP, fornece recursos para ajudar a:

- Guias de início rápido sobre segurança e conformidade — Esses guias de implantação discutem considerações arquitetônicas e fornecem etapas para a implantação de ambientes básicos focados em segurança e conformidade em. AWS
- Documento técnico sobre arquitetura para segurança e conformidade com a HIPAA — Este whitepaper descreve como as empresas podem usar para criar aplicativos compatíveis com a HIPAA. AWS
- AWS Recursos de conformidade — esta coleção de pastas de trabalho e guias que podem ser aplicados ao seu setor e localização
- AWS Config— Esse AWS serviço que avalia o quão bem suas configurações de recursos estão em conformidade com as práticas internas, as diretrizes e os regulamentos do setor.
- AWS Security Hub CSPM— Esse AWS serviço fornece uma visão abrangente do seu estado de segurança interno, AWS que ajuda você a verificar sua conformidade com os padrões e as melhores práticas do setor de segurança.

Resiliência no Amazon Kinesis Data Streams

A infraestrutura AWS global é construída em torno de AWS regiões e zonas de disponibilidade. AWS As regiões fornecem várias zonas de disponibilidade fisicamente separadas e isoladas, conectadas a redes de baixa latência, alta taxa de transferência e alta redundância. Com as zonas de disponibilidade, é possível projetar e operar aplicações e bancos de dados que executam o failover automaticamente entre as zonas de disponibilidade sem interrupção. As zonas de disponibilidade são mais altamente disponíveis, tolerantes a falhas e escaláveis que uma ou várias infraestruturas de data center tradicionais.

Para obter mais informações sobre AWS regiões e zonas de disponibilidade, consulte Infraestrutura AWS global.

Além da infraestrutura AWS global, o Kinesis Data Streams oferece vários recursos para ajudar a suportar suas necessidades de resiliência e backup de dados.

Recuperação de desastres no Amazon Kinesis Data Streams

Ao usar uma aplicação do Amazon Kinesis Data Streams para processar dados de um fluxo, podem ocorrer falhas nos seguintes níveis:

- Um processador de registros pode falhar
- Um operador pode falhar ou a instância do aplicativo que instanciou o operador pode falhar
- Uma instância do EC2 que hospeda uma ou mais instâncias do aplicativo pode falhar

Falha do processador de registros

O trabalhador invoca métodos do processador de registros usando tarefas Java ExecutorService. Se uma tarefa falhar, o operador manterá o controle do fragmento que o processador de registros estava processando. O operador inicia uma nova tarefa de processador de registros para processar esse fragmento. Para obter mais informações, consulte Controle de utilização de leitura.

Falha de operador ou aplicação

Se um operador (ou uma instância) da aplicação do Amazon Kinesis Data Streams falhar, é necessário detectar e resolver a situação. Por exemplo, se o método `Worker.run` lançar uma exceção, é necessário identificá-la e tratá-la.

Se o próprio aplicativo falhar, é necessário detectar isso e reiniciá-lo. Quando o aplicativo é iniciado, ele instancia um novo operador, que, por sua vez, instancia novos processadores de registros aos quais são atribuídos fragmentos automaticamente para processamento. Podem ser os mesmos fragmentos que esses processadores de registros estavam processando antes da falha ou fragmentos novos para esses processadores.

Em uma situação em que o operador ou o aplicativo falha, a falha não é detectada e há outras instâncias do aplicativo sendo executadas em outras instâncias do EC2, os operadores nessas outras instâncias lidam com a falha. Eles criam processadores de registro adicionais para processar os fragmentos que não estão mais sendo processados pelo operador com falha. A carga nessas outras instâncias do EC2 aumenta de forma correspondente.

O cenário descrito aqui pressupõe que, embora o operador ou a aplicação tenha falhado, a instância do EC2 que hospeda o recurso ainda está em execução e, portanto, não é reiniciada por um grupo do Auto Scaling.

Falha na instância do Amazon EC2

Recomenda-se a execução das instâncias do EC2 referentes a sua aplicação em um grupo do Auto Scaling. Dessa forma, se uma das instâncias do EC2 falhar, o grupo do Auto Scaling iniciará automaticamente uma nova instância para substituí-la. É necessário configurar as instâncias para iniciar a aplicação do Amazon Kinesis Data Streams no startup.

Segurança de infraestrutura no Kinesis Data Streams

Como um serviço gerenciado, o Amazon Kinesis Data Streams é protegido AWS pelos procedimentos globais de segurança de rede descritos no whitepaper Amazon Web Services: Visão geral dos processos de segurança.

Você usa chamadas de API AWS publicadas para acessar o Kinesis Data Streams pela rede. Os clientes devem oferecer suporte a Transport Layer Security (TLS) 1.2 ou posterior. Os clientes também devem ter compatibilidade com conjuntos de criptografia com perfect forward secrecy (PFS) como Ephemeral Diffie-Hellman (DHE) ou Ephemeral Elliptic Curve Diffie-Hellman (ECDHE). A maioria dos sistemas modernos como Java 7 e versões posteriores oferece compatibilidade com esses modos.

Além disso, as solicitações devem ser assinadas usando um ID da chave de acesso e uma chave de acesso secreta associada a uma entidade principal do IAM. Ou pode-se usar o AWS Security Token Service (AWS STS) para gerar credenciais de segurança temporárias para assinar solicitações.

Práticas recomendadas de segurança para o Kinesis Data Streams

O Amazon Kinesis Data Streams fornece uma série de recursos de segurança a serem considerados no desenvolvimento e na implementação das suas próprias políticas de segurança. As práticas recomendadas a seguir são diretrizes gerais e não representam uma solução completa de segurança. Como essas práticas recomendadas podem não ser adequadas ou suficientes para o seu ambiente, trate-as como considerações úteis em vez de prescrições.

Implemente o acesso de privilégio mínimo

Ao conceder permissões, é necessário decidir quem receberá quais permissões para quais recursos do Kinesis Data Streams. Habilite ações específicas que quer permitir nesses recursos. Portanto, é necessário conceder somente as permissões necessárias para executar uma tarefa. A implementação do privilégio de acesso mínimo é fundamental para reduzir o risco de segurança e o impacto que pode resultar de erros ou usuários mal-intencionados.

Usar funções do IAM

Aplicações de clientes e produtores precisam ter credenciais válidas para acessar fluxos de dados do Kinesis. Você não deve armazenar AWS credenciais diretamente em um aplicativo cliente ou em um bucket do Amazon S3. Essas são credenciais de longo prazo que não são automaticamente alternadas e podem ter um impacto comercial significativo se forem comprometidas.

Em vez disso, use um perfil do IAM para gerenciar credenciais temporárias que suas aplicações de clientes e produtores usarão para acessar fluxos de dados do Kinesis. Ao usar uma função, não é necessário usar credenciais de longo prazo (como um nome de usuário e uma senha ou chaves de acesso) para acessar outros recursos.

Para obter mais informações, consulte os seguintes tópicos no Manual do usuário do IAM:

- Perfis do IAM
- Cenários comuns para perfis: usuários, aplicações e serviços

Implementação da criptografia do lado do servidor em recursos dependentes

É possível criptografar dados em repouso e dados em trânsito no Kinesis Data Streams. Para obter mais informações, consulte Proteção de dados no Amazon Kinesis Data Streams.

Use CloudTrail para monitorar chamadas de API

O Kinesis Data Streams é AWS CloudTrail integrado com, um serviço que fornece um registro das ações realizadas por um usuário, função AWS ou serviço no Kinesis Data Streams.

Usando as informações coletadas por CloudTrail, você pode determinar a solicitação que foi feita ao Kinesis Data Streams, o endereço IP a partir do qual a solicitação foi feita, quem fez a solicitação, quando ela foi feita e detalhes adicionais.

Para obter mais informações, consulte the section called “Registre as chamadas de API do Amazon Kinesis Data Streams com AWS CloudTrail”.

Usando esse serviço com um AWS SDK

AWS kits de desenvolvimento de software (SDKs) estão disponíveis para muitas linguagens de programação populares. Cada SDK fornece uma API, exemplos de código e documentação que permitem que os desenvolvedores criem facilmente aplicações em seu idioma de preferência.

Documentação do SDK	Exemplos de código
AWS SDK para C++	AWS SDK para C++ exemplos de código
AWS CLI	AWS CLI exemplos de código
AWS SDK para Go	AWS SDK para Go exemplos de código
AWS SDK para Java	AWS SDK para Java exemplos de código
AWS SDK para JavaScript	AWS SDK para JavaScript exemplos de código
AWS SDK para Kotlin	AWS SDK para Kotlin exemplos de código
AWS SDK para .NET	AWS SDK para .NET exemplos de código
AWS SDK para PHP	AWS SDK para PHP exemplos de código
Ferramentas da AWS para PowerShell	Ferramentas da AWS para PowerShell exemplos de código
AWS SDK para Python (Boto3)	AWS SDK para Python (Boto3) exemplos de código
AWS SDK para Ruby	AWS SDK para Ruby exemplos de código
AWS SDK para Rust	AWS SDK para Rust exemplos de código
AWS SDK para SAP ABAP	AWS SDK para SAP ABAP exemplos de código
AWS SDK para Swift	AWS SDK para Swift exemplos de código

Exemplo de disponibilidade

Não consegue encontrar o que precisa? Solicite um exemplo de código usando o link Fornecer feedback na parte inferior desta página.

Exemplos de código para Kinesis usando AWS SDKs

Os exemplos de código a seguir mostram como usar o Kinesis com um kit de desenvolvimento AWS de software (SDK).

As noções básicas são exemplos de código que mostram como realizar as operações essenciais em um serviço.

Ações são trechos de código de programas maiores e devem ser executadas em contexto. Embora as ações mostrem como chamar perfis de serviço individuais, você pode ver as ações no contexto em seus cenários relacionados.

Para obter uma lista completa dos guias do desenvolvedor do AWS SDK e exemplos de código, consulteUsando esse serviço com um AWS SDK. Este tópico também inclui informações sobre como começar e detalhes sobre versões anteriores do SDK.

Exemplos de código

- Exemplos básicos de uso do Kinesis AWS SDKs
  - Aprenda o básico do Kinesis com um SDK AWS
  - Ações para o Kinesis usando AWS SDKs
    - Use AddTagsToStream com um AWS SDK ou CLI
    - Use CreateStream com um AWS SDK ou CLI
    - Use DeleteStream com um AWS SDK ou CLI
    - Use DeregisterStreamConsumer com um AWS SDK ou CLI
    - Use DescribeStream com um AWS SDK ou CLI
    - Use GetRecords com um AWS SDK ou CLI
    - Usar GetShardIterator com uma CLI
    - Use ListStreamConsumers com um AWS SDK
    - Use ListStreams com um AWS SDK ou CLI
    - Use ListTagsForStream com um AWS SDK ou CLI
    - Use PutRecord com um AWS SDK ou CLI
    - Use PutRecords com um AWS SDK ou CLI
    - Use RegisterStreamConsumer com um AWS SDK ou CLI
- Exemplos de tecnologia sem servidor para o Kinesis

- Invocar uma função do Lambda em um trigger do Kinesis
- Relatando falhas de itens em lote para funções do Lambda com um trigger do Kinesis

Exemplos básicos de uso do Kinesis AWS SDKs

Os exemplos de código a seguir mostram como usar os conceitos básicos do Amazon AWS SDKs Kinesis com.

Exemplos

- Aprenda o básico do Kinesis com um SDK AWS
- Ações para o Kinesis usando AWS SDKs
  - Use AddTagsToStream com um AWS SDK ou CLI
  - Use CreateStream com um AWS SDK ou CLI
  - Use DeleteStream com um AWS SDK ou CLI
  - Use DeregisterStreamConsumer com um AWS SDK ou CLI
  - Use DescribeStream com um AWS SDK ou CLI
  - Use GetRecords com um AWS SDK ou CLI
  - Usar GetShardIterator com uma CLI
  - Use ListStreamConsumers com um AWS SDK
  - Use ListStreams com um AWS SDK ou CLI
  - Use ListTagsForStream com um AWS SDK ou CLI
  - Use PutRecord com um AWS SDK ou CLI
  - Use PutRecords com um AWS SDK ou CLI
  - Use RegisterStreamConsumer com um AWS SDK ou CLI

Aprenda o básico do Kinesis com um SDK AWS

O exemplo de código a seguir mostra como:

- Criar um fluxo e inserir um registro nele.
- Criar um iterador de fragmento.
- Ler o registro e, em seguida, limpar os recursos.

SAP ABAP

SDK para SAP ABAP

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
   DATA lo_stream_describe_result TYPE REF TO /aws1/cl_knsdescrstreamoutput.
   DATA lo_stream_description TYPE REF TO /aws1/cl_knsstreamdescription.
   DATA lo_sharditerator TYPE REF TO /aws1/cl_knsgetsharditerator01.
   DATA lo_record_result TYPE REF TO /aws1/cl_knsputrecordoutput.

   "Create stream."
   TRY.
       lo_kns->createstream(
           iv_streamname = iv_stream_name
           iv_shardcount = iv_shard_count ).
       MESSAGE 'Stream created.' TYPE 'I'.
     CATCH /aws1/cx_knsinvalidargumentex.
       MESSAGE 'The specified argument was not valid.' TYPE 'E'.
     CATCH /aws1/cx_knslimitexceededex.
       MESSAGE 'The request processing has failed because of a limit exceeded
exception.' TYPE 'E'.
     CATCH /aws1/cx_knsresourceinuseex.
       MESSAGE 'The request processing has failed because the resource is in
use.' TYPE 'E'.
   ENDTRY.

   "Wait for stream to becomes active."
   lo_stream_describe_result = lo_kns->describestream( iv_streamname =
iv_stream_name ).
   lo_stream_description = lo_stream_describe_result->get_streamdescription( ).
   WHILE lo_stream_description->get_streamstatus( ) <> 'ACTIVE'.
     IF sy-index = 30.
       EXIT.               "maximum 5 minutes"
     ENDIF.
     WAIT UP TO 10 SECONDS.
```

```
      lo_stream_describe_result = lo_kns->describestream( iv_streamname =
 iv_stream_name ).
      lo_stream_description = lo_stream_describe_result-
>get_streamdescription( ).
    ENDWHILE.

    "Create record."
    TRY.
        lo_record_result = lo_kns->putrecord(
            iv_streamname = iv_stream_name
            iv_data       = iv_data
            iv_partitionkey = iv_partition_key ).
        MESSAGE 'Record created.' TYPE 'I'.
      CATCH /aws1/cx_knsinvalidargumentex.
        MESSAGE 'The specified argument was not valid.' TYPE 'E'.
      CATCH /aws1/cx_knskmsaccessdeniedex.
        MESSAGE 'You do not have permission to perform this AWS KMS action.' TYPE
 'E'.
      CATCH /aws1/cx_knskmsdisabledex.
        MESSAGE 'KMS key used is disabled.' TYPE 'E'.
      CATCH /aws1/cx_knskmsinvalidstateex.
        MESSAGE 'KMS key used is in an invalid state. ' TYPE 'E'.
      CATCH /aws1/cx_knskmsnotfoundex.
        MESSAGE 'KMS key used is not found.' TYPE 'E'.
      CATCH /aws1/cx_knskmsoptinrequired.
        MESSAGE 'KMS key option is required.' TYPE 'E'.
      CATCH /aws1/cx_knskmsthrottlingex.
        MESSAGE 'The rate of requests to AWS KMS is exceeding the request
 quotas.' TYPE 'E'.
      CATCH /aws1/cx_knsprovthruputexcdex.
        MESSAGE 'The request rate for the stream is too high, or the requested
 data is too large for the available throughput.' TYPE 'E'.
      CATCH /aws1/cx_knsresourcenotfoundex.
        MESSAGE 'Resource being accessed is not found.' TYPE 'E'.
    ENDTRY.

    "Create a shard iterator in order to read the record."
    TRY.
        lo_sharditerator = lo_kns->getsharditerator(
          iv_shardid = lo_record_result->get_shardid( )
          iv_sharditeratortype = iv_sharditeratortype
          iv_streamname = iv_stream_name ).
        MESSAGE 'Shard iterator created.' TYPE 'I'.
      CATCH /aws1/cx_knsinvalidargumentex.
```

```
        MESSAGE 'The specified argument was not valid.' TYPE 'E'.
      CATCH /aws1/cx_knsprovthruputexcdex.
        MESSAGE 'The request rate for the stream is too high, or the requested
data is too large for the available throughput.' TYPE 'E'.
      CATCH /aws1/cx_sgmresourcenotfound.
        MESSAGE 'Resource being accessed is not found.' TYPE 'E'.
    ENDTRY.

    "Read the record."
    TRY.
        oo_result = lo_kns->getrecords(                    " oo_result is
returned for testing purposes. "
            iv_sharditerator   = lo_sharditerator->get_sharditerator( ) ).
        MESSAGE 'Shard iterator created.' TYPE 'I'.
      CATCH /aws1/cx_knsexpirediteratorex.
        MESSAGE 'Iterator expired.' TYPE 'E'.
      CATCH /aws1/cx_knsinvalidargumentex.
        MESSAGE 'The specified argument was not valid.' TYPE 'E'.
      CATCH /aws1/cx_knskmsaccessdeniedex.
        MESSAGE 'You do not have permission to perform this AWS KMS action.' TYPE
'E'.
      CATCH /aws1/cx_knskmsdisabledex.
        MESSAGE 'KMS key used is disabled.' TYPE 'E'.
      CATCH /aws1/cx_knskmsinvalidstateex.
        MESSAGE 'KMS key used is in an invalid state. ' TYPE 'E'.
      CATCH /aws1/cx_knskmsnotfoundex.
        MESSAGE 'KMS key used is not found.' TYPE 'E'.
      CATCH /aws1/cx_knskmsoptinrequired.
        MESSAGE 'KMS key option is required.' TYPE 'E'.
      CATCH /aws1/cx_knskmsthrottlingex.
        MESSAGE 'The rate of requests to AWS KMS is exceeding the request
quotas.' TYPE 'E'.
      CATCH /aws1/cx_knsprovthruputexcdex.
        MESSAGE 'The request rate for the stream is too high, or the requested
data is too large for the available throughput.' TYPE 'E'.
      CATCH /aws1/cx_knsresourcenotfoundex.
        MESSAGE 'Resource being accessed is not found.' TYPE 'E'.
    ENDTRY.

    "Delete stream."
    TRY.
        lo_kns->deletestream(
            iv_streamname = iv_stream_name ).
        MESSAGE 'Stream deleted.' TYPE 'I'.
```

```
      CATCH /aws1/cx_knslimitexceededex.
        MESSAGE 'The request processing has failed because of a limit exceeded
exception.' TYPE 'E'.
      CATCH /aws1/cx_knsresourceinuseex.
        MESSAGE 'The request processing has failed because the resource is in
use.' TYPE 'E'.
    ENDTRY.
```

- Consulte detalhes da API nos tópicos a seguir na Referência de API do AWS SDK para SAP ABAP.
  - CreateStream
  - DeleteStream
  - GetRecords
  - GetShardIterator
  - PutRecord

Para obter uma lista completa dos guias do desenvolvedor do AWS SDK e exemplos de código, consulteUsando esse serviço com um AWS SDK. Este tópico também inclui informações sobre como começar e detalhes sobre versões anteriores do SDK.

Ações para o Kinesis usando AWS SDKs

Os exemplos de código a seguir demonstram como realizar ações individuais do Kinesis com. AWS SDKs Cada exemplo inclui um link para GitHub, onde você pode encontrar instruções para configurar e executar o código.

Os exemplos a seguir incluem apenas as ações mais utilizadas. Consulte uma lista completa na Referência de API do Amazon Kinesis.

Exemplos

- Use AddTagsToStream com um AWS SDK ou CLI
- Use CreateStream com um AWS SDK ou CLI
- Use DeleteStream com um AWS SDK ou CLI
- Use DeregisterStreamConsumer com um AWS SDK ou CLI
- Use DescribeStream com um AWS SDK ou CLI
- Use GetRecords com um AWS SDK ou CLI

- Usar GetShardIterator com uma CLI
- Use ListStreamConsumers com um AWS SDK
- Use ListStreams com um AWS SDK ou CLI
- Use ListTagsForStream com um AWS SDK ou CLI
- Use PutRecord com um AWS SDK ou CLI
- Use PutRecords com um AWS SDK ou CLI
- Use RegisterStreamConsumer com um AWS SDK ou CLI

Use **`AddTagsToStream`** com um AWS SDK ou CLI

Os exemplos de código a seguir mostram como usar o `AddTagsToStream`.

.NET

SDK para .NET

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
using System;
using System.Collections.Generic;
using System.Threading.Tasks;
using Amazon.Kinesis;
using Amazon.Kinesis.Model;

/// <summary>
/// This example shows how to apply key/value pairs to an Amazon Kinesis
/// stream.
/// </summary>
public class TagStream
{
    public static async Task Main()
    {
        IAmazonKinesis client = new AmazonKinesisClient();

        string streamName = "AmazonKinesisStream";
```

```
            var tags = new Dictionary<string, string>
            {
                { "Project", "Sample Kinesis Project" },
                { "Application", "Sample Kinesis App" },
            };

            var success = await ApplyTagsToStreamAsync(client, streamName, tags);

            if (success)
            {
                Console.WriteLine($"Taggs successfully added to {streamName}.");
            }
            else
            {
                Console.WriteLine("Tags were not added to the stream.");
            }
        }

        /// <summary>
        /// Applies the set of tags to the named Kinesis stream.
        /// </summary>
        /// <param name="client">The initialized Kinesis client.</param>
        /// <param name="streamName">The name of the Kinesis stream to which
        /// the tags will be attached.</param>
        /// <param name="tags">A sictionary containing key/value pairs which
        /// will be used to create the Kinesis tags.</param>
        /// <returns>A Boolean value which represents the success or failure
        /// of AddTagsToStreamAsync.</returns>
        public static async Task<bool> ApplyTagsToStreamAsync(
            IAmazonKinesis client,
            string streamName,
            Dictionary<string, string> tags)
        {
            var request = new AddTagsToStreamRequest
            {
                StreamName = streamName,
                Tags = tags,
            };

            var response = await client.AddTagsToStreamAsync(request);

            return response.HttpStatusCode == System.Net.HttpStatusCode.OK;
        }
    }
```

- Para obter detalhes da API, consulte AddTagsToStreama Referência AWS SDK para .NET da API.

CLI

AWS CLI

Para adicionar tags a um fluxo de dados

O exemplo `add-tags-to-stream` a seguir atribui uma tag com a chave `samplekey` e o valor `example` ao fluxo especificado.

```
aws kinesis add-tags-to-stream \
    --stream-name samplestream \
    --tags samplekey=example
```

Este comando não produz saída.

Para obter mais informações, consulte Adicionar tags a fluxos no Guia do desenvolvedor do Amazon Kinesis Data Streams.

- Para obter detalhes da API, consulte AddTagsToStreamem Referência de AWS CLI Comandos.

Para obter uma lista completa dos guias do desenvolvedor do AWS SDK e exemplos de código, consulteUsando esse serviço com um AWS SDK. Este tópico também inclui informações sobre como começar e detalhes sobre versões anteriores do SDK.

Use **`CreateStream`** com um AWS SDK ou CLI

Os exemplos de código a seguir mostram como usar o `CreateStream`.

Exemplos de ações são trechos de código de programas maiores e devem ser executados em contexto. É possível ver essa ação em contexto no seguinte exemplo de código:

- Conheça os conceitos básicos

.NET

SDK para .NET

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
using System;
using System.Threading.Tasks;
using Amazon.Kinesis;
using Amazon.Kinesis.Model;

/// <summary>
/// This example shows how to create a new Amazon Kinesis stream.
/// </summary>
public class CreateStream
{
    public static async Task Main()
    {
        IAmazonKinesis client = new AmazonKinesisClient();

        string streamName = "AmazonKinesisStream";
        int shardCount = 1;

        var success = await CreateNewStreamAsync(client, streamName,
shardCount);
        if (success)
        {
            Console.WriteLine($"The stream, {streamName} successfully
created.");
        }
    }

    /// <summary>
    /// Creates a new Kinesis stream.
    /// </summary>
    /// <param name="client">An initialized Kinesis client.</param>
    /// <param name="streamName">The name for the new stream.</param>
    /// <param name="shardCount">The number of shards the new stream will
```

```
        /// use. The throughput of the stream is a function of the number of
        /// shards; more shards are required for greater provisioned
        /// throughput.</param>
        /// <returns>A Boolean value indicating whether the stream was created.</
returns>
        public static async Task<bool> CreateNewStreamAsync(IAmazonKinesis
 client, string streamName, int shardCount)
        {
            var request = new CreateStreamRequest
            {
                StreamName = streamName,
                ShardCount = shardCount,
            };

            var response = await client.CreateStreamAsync(request);

            return response.HttpStatusCode == System.Net.HttpStatusCode.OK;
        }
    }
```

- Para obter detalhes da API, consulte CreateStreama Referência AWS SDK para .NET da API.

CLI

AWS CLI

Como criar um fluxo de dados

O exemplo de `create-stream` a seguir cria um fluxo de dados chamado samplestream com três fragmentos.

```
aws kinesis create-stream \
    --stream-name samplestream \
    --shard-count 3
```

Este comando não produz saída.

Para obter mais informações, consulte Criar um fluxo no Guia do desenvolvedor do Amazon Kinesis Data Streams.

- Para obter detalhes da API, consulte CreateStreamem Referência de AWS CLI Comandos.

Java

SDK para Java 2.x

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
import software.amazon.awssdk.regions.Region;
import software.amazon.awssdk.services.kinesis.KinesisClient;
import software.amazon.awssdk.services.kinesis.model.CreateStreamRequest;
import software.amazon.awssdk.services.kinesis.model.KinesisException;

/**
 * Before running this Java V2 code example, set up your development
 * environment, including your credentials.
 *
 * For more information, see the following documentation topic:
 *
 * https://docs.aws.amazon.com/sdk-for-java/latest/developer-guide/get-
started.html
 */
public class CreateDataStream {
    public static void main(String[] args) {

        final String usage = """

                Usage:
                    <streamName>

                Where:
                    streamName - The Amazon Kinesis data stream (for example,
 StockTradeStream).
                """;

        if (args.length != 1) {
            System.out.println(usage);
            System.exit(1);
```

```
        }

        String streamName = args[0];
        Region region = Region.US_EAST_1;
        KinesisClient kinesisClient = KinesisClient.builder()
                .region(region)
                .build();
        createStream(kinesisClient, streamName);
        System.out.println("Done");
        kinesisClient.close();
    }

    public static void createStream(KinesisClient kinesisClient, String
 streamName) {
        try {
            CreateStreamRequest streamReq = CreateStreamRequest.builder()
                    .streamName(streamName)
                    .shardCount(1)
                    .build();

            kinesisClient.createStream(streamReq);

        } catch (KinesisException e) {
            System.err.println(e.getMessage());
            System.exit(1);
        }
    }
}
```

- Para obter detalhes da API, consulte CreateStreama Referência AWS SDK for Java 2.x da API.

PowerShell

Ferramentas para PowerShell V4

Exemplo 1: cria um fluxo. Por padrão, esse cmdlet não retorna nenhuma saída, então a PassThru opção - é adicionada para retornar o valor fornecido ao StreamName parâmetro - para uso posterior.

```
$streamName = New-KINStream -StreamName "mystream" -ShardCount 1 -PassThru
```

- Para obter detalhes da API, consulte CreateStreamem Referência de Ferramentas da AWS para PowerShell cmdlet (V4).

Ferramentas para PowerShell V5

Exemplo 1: cria um fluxo.

```
New-KINStream -StreamName "mystream" -ShardCount 1
```

- Para obter detalhes da API, consulte CreateStreamem Referência de Ferramentas da AWS para PowerShell cmdlet (V5).

Python

SDK para Python (Boto3)

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
class KinesisStream:
    """Encapsulates a Kinesis stream."""

    def __init__(self, kinesis_client):
        """
        :param kinesis_client: A Boto3 Kinesis client.
        """
        self.kinesis_client = kinesis_client
        self.name = None
        self.details = None
        self.stream_exists_waiter = kinesis_client.get_waiter("stream_exists")


    def create(self, name, wait_until_exists=True):
        """
        Creates a stream.

        :param name: The name of the stream.
        :param wait_until_exists: When True, waits until the service reports that
```

```
                                the stream exists, then queries for its
metadata.
        """
        try:
            self.kinesis_client.create_stream(StreamName=name, ShardCount=1)
            self.name = name
            logger.info("Created stream %s.", name)
            if wait_until_exists:
                logger.info("Waiting until exists.")
                self.stream_exists_waiter.wait(StreamName=name)
                self.describe(name)
        except ClientError:
            logger.exception("Couldn't create stream %s.", name)
            raise
```

- Para obter detalhes da API, consulte a CreateStreamReferência da API AWS SDK for Python (Boto3).

Rust

SDK para Rust

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
async fn make_stream(client: &Client, stream: &str) -> Result<(), Error> {
    client
        .create_stream()
        .stream_name(stream)
        .shard_count(4)
        .send()
        .await?;

    println!("Created stream");

    Ok(())
```

```
}
```

- Para obter detalhes da API, consulte a CreateStreamreferência da API AWS SDK for Rust.

SAP ABAP

SDK para SAP ABAP

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
   TRY.
       lo_kns->createstream(
           iv_streamname = iv_stream_name
           iv_shardcount = iv_shard_count ).
       MESSAGE 'Stream created.' TYPE 'I'.
     CATCH /aws1/cx_knsinvalidargumentex.
       MESSAGE 'The specified argument was not valid.' TYPE 'E'.
     CATCH /aws1/cx_knslimitexceededex.
       MESSAGE 'The request processing has failed because of a limit exceed
exception.' TYPE 'E'.
     CATCH /aws1/cx_knsresourceinuseex.
       MESSAGE 'The request processing has failed because the resource is in
use.' TYPE 'E'.
   ENDTRY.
```

- Para obter detalhes da API, consulte a CreateStreamreferência da API AWS SDK for SAP ABAP.

Para obter uma lista completa dos guias do desenvolvedor do AWS SDK e exemplos de código, consulteUsando esse serviço com um AWS SDK. Este tópico também inclui informações sobre como começar e detalhes sobre versões anteriores do SDK.

Use **`DeleteStream`** com um AWS SDK ou CLI

Os exemplos de código a seguir mostram como usar o `DeleteStream`.

Exemplos de ações são trechos de código de programas maiores e devem ser executados em contexto. É possível ver essa ação em contexto no seguinte exemplo de código:

- Conheça os conceitos básicos

.NET

SDK para .NET

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
using System;
using System.Threading.Tasks;
using Amazon.Kinesis;
using Amazon.Kinesis.Model;

/// <summary>
/// Shows how to delete an Amazon Kinesis stream.
/// </summary>
public class DeleteStream
{
    public static async Task Main()
    {
        IAmazonKinesis client = new AmazonKinesisClient();
        string streamName = "AmazonKinesisStream";

        var success = await DeleteStreamAsync(client, streamName);

        if (success)
        {
            Console.WriteLine($"Stream, {streamName} successfully deleted.");
        }
        else
```

```
            {
                Console.WriteLine("Stream not deleted.");
            }
        }

        /// <summary>
        /// Deletes a Kinesis stream.
        /// </summary>
        /// <param name="client">An initialized Kinesis client object.</param>
        /// <param name="streamName">The name of the string to delete.</param>
        /// <returns>A Boolean value representing the success of the operation.</
returns>
        public static async Task<bool> DeleteStreamAsync(IAmazonKinesis client,
 string streamName)
        {
            // If EnforceConsumerDeletion is true, any consumers
            // of this stream will also be deleted. If it is set
            // to false and this stream has any consumers, the
            // call will fail with a ResourceInUseException.
            var request = new DeleteStreamRequest
            {
                StreamName = streamName,
                EnforceConsumerDeletion = true,
            };

            var response = await client.DeleteStreamAsync(request);

            return response.HttpStatusCode == System.Net.HttpStatusCode.OK;
        }
    }
```

- Para obter detalhes da API, consulte DeleteStreama Referência AWS SDK para .NET da API.

CLI

AWS CLI

Como excluir um fluxo de dados

O exemplo de `delete-stream` a seguir exclui o fluxo de dados especificado.

```
aws kinesis delete-stream \
    --stream-name samplestream
```

Este comando não produz saída.

Para obter mais informações, consulte Excluir um fluxo no Guia do desenvolvedor do Amazon Kinesis Data Streams.

- Para obter detalhes da API, consulte DeleteStreamem Referência de AWS CLI Comandos.

Java

SDK para Java 2.x

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
import software.amazon.awssdk.regions.Region;
import software.amazon.awssdk.services.kinesis.KinesisClient;
import software.amazon.awssdk.services.kinesis.model.DeleteStreamRequest;
import software.amazon.awssdk.services.kinesis.model.KinesisException;

/**
 * Before running this Java V2 code example, set up your development
 * environment, including your credentials.
 *
 * For more information, see the following documentation topic:
 *
 * https://docs.aws.amazon.com/sdk-for-java/latest/developer-guide/get-
started.html
 */
public class DeleteDataStream {

    public static void main(String[] args) {
        final String usage = """

                Usage:
                    <streamName>
```

```
                Where:
                    streamName - The Amazon Kinesis data stream (for example,
 StockTradeStream)
                """;

        if (args.length != 1) {
            System.out.println(usage);
            System.exit(1);
        }

        String streamName = args[0];
        Region region = Region.US_EAST_1;
        KinesisClient kinesisClient = KinesisClient.builder()
                .region(region)
                .build();

        deleteStream(kinesisClient, streamName);
        kinesisClient.close();
        System.out.println("Done");
    }

    public static void deleteStream(KinesisClient kinesisClient, String
 streamName) {
        try {
            DeleteStreamRequest delStream = DeleteStreamRequest.builder()
                    .streamName(streamName)
                    .build();

            kinesisClient.deleteStream(delStream);

        } catch (KinesisException e) {
            System.err.println(e.getMessage());
            System.exit(1);
        }
    }
}
```

- Para obter detalhes da API, consulte DeleteStreama Referência AWS SDK for Java 2.x da API.

PowerShell

Ferramentas para PowerShell V4

Exemplo 1: exclui o fluxo especificado. A confirmação será solicitada antes que a operação continue. Para ignorar a solicitação de confirmação, use a opção -Force.

```
Remove-KINStream -StreamName "mystream"
```

- Para obter detalhes da API, consulte DeleteStreamem Referência de Ferramentas da AWS para PowerShell cmdlet (V4).

Ferramentas para PowerShell V5

Exemplo 1: exclui o fluxo especificado. A confirmação será solicitada antes que a operação continue. Para ignorar a solicitação de confirmação, use a opção -Force.

```
Remove-KINStream -StreamName "mystream"
```

- Para obter detalhes da API, consulte DeleteStreamem Referência de Ferramentas da AWS para PowerShell cmdlet (V5).

Python

SDK para Python (Boto3)

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
class KinesisStream:
    """Encapsulates a Kinesis stream."""

    def __init__(self, kinesis_client):
        """
        :param kinesis_client: A Boto3 Kinesis client.
        """
```

```
        self.kinesis_client = kinesis_client
        self.name = None
        self.details = None
        self.stream_exists_waiter = kinesis_client.get_waiter("stream_exists")


    def delete(self):
        """
        Deletes a stream.
        """
        try:
            self.kinesis_client.delete_stream(StreamName=self.name)
            self._clear()
            logger.info("Deleted stream %s.", self.name)
        except ClientError:
            logger.exception("Couldn't delete stream %s.", self.name)
            raise
```

- Para obter detalhes da API, consulte a DeleteStreamReferência da API AWS SDK for Python (Boto3).

Rust

SDK para Rust

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
async fn remove_stream(client: &Client, stream: &str) -> Result<(), Error> {
    client.delete_stream().stream_name(stream).send().await?;

    println!("Deleted stream.");

    Ok(())
}
```

- Para obter detalhes da API, consulte a DeleteStreamreferência da API AWS SDK for Rust.

SAP ABAP

SDK para SAP ABAP

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
   TRY.
       lo_kns->deletestream(
           iv_streamname = iv_stream_name ).
       MESSAGE 'Stream deleted.' TYPE 'I'.
     CATCH /aws1/cx_knslimitexceededex.
       MESSAGE 'The request processing has failed because of a limit exceed
exception.' TYPE 'E'.
     CATCH /aws1/cx_knsresourceinuseex.
       MESSAGE 'The request processing has failed because the resource is in
use.' TYPE 'E'.
   ENDTRY.
```

- Para obter detalhes da API, consulte a DeleteStreamreferência da API AWS SDK for SAP ABAP.

Para obter uma lista completa dos guias do desenvolvedor do AWS SDK e exemplos de código, consulteUsando esse serviço com um AWS SDK. Este tópico também inclui informações sobre como começar e detalhes sobre versões anteriores do SDK.

Use **`DeregisterStreamConsumer`** com um AWS SDK ou CLI

Os exemplos de código a seguir mostram como usar o `DeregisterStreamConsumer`.

.NET

SDK para .NET

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
    using System;
    using System.Threading.Tasks;
    using Amazon.Kinesis;
    using Amazon.Kinesis.Model;

    /// <summary>
    /// Shows how to deregister a consumer from an Amazon Kinesis stream.
    /// </summary>
    public class DeregisterConsumer
    {
        public static async Task Main(string[] args)
        {
            IAmazonKinesis client = new AmazonKinesisClient();

            string streamARN = "arn:aws:kinesis:us-west-2:000000000000:stream/
AmazonKinesisStream";
            string consumerName = "CONSUMER_NAME";
            string consumerARN = "arn:aws:kinesis:us-west-2:000000000000:stream/
AmazonKinesisStream/consumer/CONSUMER_NAME:000000000000";

            var success = await DeregisterConsumerAsync(client, streamARN,
 consumerARN, consumerName);

            if (success)
            {
                Console.WriteLine($"{consumerName} successfully deregistered.");
            }
            else
            {
                Console.WriteLine($"{consumerName} was not successfully
 deregistered.");
            }
```

```
        }

        /// <summary>
        /// Deregisters a consumer from a Kinesis stream.
        /// </summary>
        /// <param name="client">An initialized Kinesis client object.</param>
        /// <param name="streamARN">The ARN of a Kinesis stream.</param>
        /// <param name="consumerARN">The ARN of the consumer.</param>
        /// <param name="consumerName">The name of the consumer.</param>
        /// <returns>A Boolean value representing the success of the operation.</
returns>
        public static async Task<bool> DeregisterConsumerAsync(
            IAmazonKinesis client,
            string streamARN,
            string consumerARN,
            string consumerName)
        {
            var request = new DeregisterStreamConsumerRequest
            {
                StreamARN = streamARN,
                ConsumerARN = consumerARN,
                ConsumerName = consumerName,
            };

            var response = await client.DeregisterStreamConsumerAsync(request);

            return response.HttpStatusCode == System.Net.HttpStatusCode.OK;
        }
    }
```

- Para obter detalhes da API, consulte DeregisterStreamConsumera Referência AWS SDK para .NET da API.

CLI

AWS CLI

Como cancelar o registro de um consumidor de fluxo de dados

O exemplo `deregister-stream-consumer` a seguir cancela o registro do consumidor especificado do fluxo de dados especificado.

```
aws kinesis deregister-stream-consumer \
    --stream-arn arn:aws:kinesis:us-west-2:123456789012:stream/samplestream \
    --consumer-name KinesisConsumerApplication
```

Este comando não produz saída.

Para obter mais informações, consulte Desenvolver consumidores com Fan-Out aprimorado usando a API Kinesis Data Streams no Guia do desenvolvedor do Amazon Kinesis Data Streams.

- Para obter detalhes da API, consulte DeregisterStreamConsumerem Referência de AWS CLI Comandos.

Para obter uma lista completa dos guias do desenvolvedor do AWS SDK e exemplos de código, consulteUsando esse serviço com um AWS SDK. Este tópico também inclui informações sobre como começar e detalhes sobre versões anteriores do SDK.

Use **`DescribeStream`** com um AWS SDK ou CLI

Os exemplos de código a seguir mostram como usar o `DescribeStream`.

CLI

AWS CLI

Como descrever um fluxo de dados

O exemplo de `describe-stream` a seguir retorna detalhes sobre o fluxo de dados especificado.

```
aws kinesis describe-stream \
    --stream-name samplestream
```

Saída:

```
{
    "StreamDescription": {
        "Shards": [
            {
                "ShardId": "shardId-000000000000",
                "HashKeyRange": {
```

```
                    "StartingHashKey": "0",
                    "EndingHashKey": "113427455640312821154458202477256070484"
                },
                "SequenceNumberRange": {
                    "StartingSequenceNumber":
 "49600871682957036442365024926191073437251060580128653314"
                }
            },
            {
                "ShardId": "shardId-000000000001",
                "HashKeyRange": {
                    "StartingHashKey": "113427455640312821154458202477256070485",
                    "EndingHashKey": "226854911280625642308916404954512140969"
                },
                "SequenceNumberRange": {
                    "StartingSequenceNumber":
 "49600871682979337187563555549332609155523708941634633746"
                }
            },
            {
                "ShardId": "shardId-000000000002",
                "HashKeyRange": {
                    "StartingHashKey": "226854911280625642308916404954512140970",
                    "EndingHashKey": "340282366920938463463374607431768211455"
                },
                "SequenceNumberRange": {
                    "StartingSequenceNumber":
 "49600871683001637932762086172474144873796357303140614178"
                }
            }
        ],
        "StreamARN": "arn:aws:kinesis:us-west-2:123456789012:stream/
samplestream",
        "StreamName": "samplestream",
        "StreamStatus": "ACTIVE",
        "RetentionPeriodHours": 24,
        "EnhancedMonitoring": [
            {
                "ShardLevelMetrics": []
            }
        ],
        "EncryptionType": "NONE",
        "KeyId": null,
        "StreamCreationTimestamp": 1572297168.0
```

```
    }
}
```

Para obter mais informações, consulte Criar e gerenciar fluxos no Guia do desenvolvedor do Amazon Kinesis Data Streams.

- Para obter detalhes da API, consulte DescribeStreamem Referência de AWS CLI Comandos.

PowerShell

Ferramentas para PowerShell V4

Exemplo 1: retorna detalhes do fluxo especificado.

```
Get-KINStream -StreamName "mystream"
```

Saída:

```
HasMoreShards        : False
RetentionPeriodHours : 24
Shards               : {}
StreamARN            : arn:aws:kinesis:us-west-2:123456789012:stream/mystream
StreamName           : mystream
StreamStatus         : ACTIVE
```

- Para obter detalhes da API, consulte DescribeStreamem Referência de Ferramentas da AWS para PowerShell cmdlet (V4).

Ferramentas para PowerShell V5

Exemplo 1: retorna detalhes do fluxo especificado.

```
Get-KINStream -StreamName "mystream"
```

Saída:

```
HasMoreShards        : False
RetentionPeriodHours : 24
Shards               : {}
StreamARN            : arn:aws:kinesis:us-west-2:123456789012:stream/mystream
StreamName           : mystream
```

```
StreamStatus         : ACTIVE
```

- Para obter detalhes da API, consulte DescribeStreamem Referência de Ferramentas da AWS para PowerShell cmdlet (V5).

Python

SDK para Python (Boto3)

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
class KinesisStream:
    """Encapsulates a Kinesis stream."""

    def __init__(self, kinesis_client):
        """
        :param kinesis_client: A Boto3 Kinesis client.
        """
        self.kinesis_client = kinesis_client
        self.name = None
        self.details = None
        self.stream_exists_waiter = kinesis_client.get_waiter("stream_exists")


    def describe(self, name):
        """
        Gets metadata about a stream.

        :param name: The name of the stream.
        :return: Metadata about the stream.
        """
        try:
            response = self.kinesis_client.describe_stream(StreamName=name)
            self.name = name
            self.details = response["StreamDescription"]
            logger.info("Got stream %s.", name)
        except ClientError:
```

```
            logger.exception("Couldn't get %s.", name)
            raise
        else:
            return self.details
```

- Para obter detalhes da API, consulte a DescribeStreamReferência da API AWS SDK for Python (Boto3).

Rust

SDK para Rust

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
async fn show_stream(client: &Client, stream: &str) -> Result<(), Error> {
    let resp = client.describe_stream().stream_name(stream).send().await?;

    let desc = resp.stream_description.unwrap();

    println!("Stream description:");
    println!("  Name:              {}:", desc.stream_name());
    println!("  Status:            {:?}", desc.stream_status());
    println!("  Open shards:       {:?}", desc.shards.len());
    println!("  Retention (hours): {}", desc.retention_period_hours());
    println!("  Encryption:        {:?}", desc.encryption_type.unwrap());

    Ok(())
}
```

- Para obter detalhes da API, consulte a DescribeStreamreferência da API AWS SDK for Rust.

SAP ABAP

SDK para SAP ABAP

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
    TRY.
        oo_result = lo_kns->describestream(
            iv_streamname = iv_stream_name ).
        DATA(lt_stream_description) = oo_result->get_streamdescription( ).
        MESSAGE 'Streams retrieved.' TYPE 'I'.
      CATCH /aws1/cx_knslimitexceededex.
        MESSAGE 'The request processing has failed because of a limit exceed
exception.' TYPE 'E'.
      CATCH /aws1/cx_knsresourcenotfoundex.
        MESSAGE 'Resource being accessed is not found.' TYPE 'E'.
    ENDTRY.
```

- Para obter detalhes da API, consulte a DescribeStreamreferência da API AWS SDK for SAP ABAP.

Para obter uma lista completa dos guias do desenvolvedor do AWS SDK e exemplos de código, consulteUsando esse serviço com um AWS SDK. Este tópico também inclui informações sobre como começar e detalhes sobre versões anteriores do SDK.

Use **`GetRecords`** com um AWS SDK ou CLI

Os exemplos de código a seguir mostram como usar o `GetRecords`.

Exemplos de ações são trechos de código de programas maiores e devem ser executados em contexto. É possível ver essa ação em contexto no seguinte exemplo de código:

- Conheça os conceitos básicos

CLI

AWS CLI

Como obter registros de um fragmento

O exemplo de `get-records` a seguir obtém registros de dados do fragmento de um fluxo de dados do Kinesis usando o iterador de fragmento especificado.

```
aws kinesis get-records \
    --shard-iterator AAAAAAAAAAF7/0mWD7IuHj1yGv/
TKuNgx2ukD5xipCY4cy4gU96orWwZwcSXh3K9tAmGYeOZyLZrvzzeOFVf9iN99hUPw/w/
b0YWYeehfNvnf1DYt5XpDJghLKr3DzgznkTmMymDP3R+3wRKeuEw6/kdxY2yKJH0veaiekaVc4N2VwK/
GvaGP2Hh9Fg7N++q0Adg6fIDQPt4p8RpavDbk+A4sL9SWGE1
```

Saída:

```
{
    "Records": [],
    "MillisBehindLatest": 80742000
}
```

Para obter mais informações, consulte Desenvolvimento de consumidores usando a API Kinesis Data Streams AWS com o SDK for Java no Guia do desenvolvedor do Amazon Kinesis Data Streams.

- Para obter detalhes da API, consulte GetRecordsem Referência de AWS CLI Comandos.

Java

SDK para Java 2.x

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
import software.amazon.awssdk.core.SdkBytes;
import software.amazon.awssdk.regions.Region;
import software.amazon.awssdk.services.kinesis.KinesisClient;
```

```
import software.amazon.awssdk.services.kinesis.model.DescribeStreamResponse;
import software.amazon.awssdk.services.kinesis.model.DescribeStreamRequest;
import software.amazon.awssdk.services.kinesis.model.Shard;
import software.amazon.awssdk.services.kinesis.model.GetShardIteratorRequest;
import software.amazon.awssdk.services.kinesis.model.GetShardIteratorResponse;
import software.amazon.awssdk.services.kinesis.model.Record;
import software.amazon.awssdk.services.kinesis.model.GetRecordsRequest;
import software.amazon.awssdk.services.kinesis.model.GetRecordsResponse;
import java.util.ArrayList;
import java.util.List;

/**
 * Before running this Java V2 code example, set up your development
 * environment, including your credentials.
 *
 * For more information, see the following documentation topic:
 *
 * https://docs.aws.amazon.com/sdk-for-java/latest/developer-guide/get-
started.html
 */
public class GetRecords {
    public static void main(String[] args) {
        final String usage = """

                Usage:
                    <streamName>

                Where:
                    streamName - The Amazon Kinesis data stream to read from (for
 example, StockTradeStream).
                """;

        if (args.length != 1) {
            System.out.println(usage);
            System.exit(1);
        }

        String streamName = args[0];
        Region region = Region.US_EAST_1;
        KinesisClient kinesisClient = KinesisClient.builder()
                .region(region)
                .build();

        getStockTrades(kinesisClient, streamName);
```

```
        kinesisClient.close();
    }

    public static void getStockTrades(KinesisClient kinesisClient, String
streamName) {
        String shardIterator;
        String lastShardId = null;
        DescribeStreamRequest describeStreamRequest =
DescribeStreamRequest.builder()
                .streamName(streamName)
                .build();

        List<Shard> shards = new ArrayList<>();
        DescribeStreamResponse streamRes;
        do {
            streamRes = kinesisClient.describeStream(describeStreamRequest);
            shards.addAll(streamRes.streamDescription().shards());

            if (shards.size() > 0) {
                lastShardId = shards.get(shards.size() - 1).shardId();
            }
        } while (streamRes.streamDescription().hasMoreShards());

        GetShardIteratorRequest itReq = GetShardIteratorRequest.builder()
                .streamName(streamName)
                .shardIteratorType("TRIM_HORIZON")
                .shardId(lastShardId)
                .build();

        GetShardIteratorResponse shardIteratorResult =
kinesisClient.getShardIterator(itReq);
        shardIterator = shardIteratorResult.shardIterator();

        // Continuously read data records from shard.
        List<Record> records;

        // Create new GetRecordsRequest with existing shardIterator.
        // Set maximum records to return to 1000.
        GetRecordsRequest recordsRequest = GetRecordsRequest.builder()
                .shardIterator(shardIterator)
                .limit(1000)
                .build();

        GetRecordsResponse result = kinesisClient.getRecords(recordsRequest);
```

```
        // Put result into record list. Result may be empty.
        records = result.records();

        // Print records
        for (Record record : records) {
            SdkBytes byteBuffer = record.data();
            System.out.printf("Seq No: %s - %s%n", record.sequenceNumber(), new
 String(byteBuffer.asByteArray()));
        }
    }
}
```

- Para obter detalhes da API, consulte GetRecordsa Referência AWS SDK for Java 2.x da API.

PowerShell

Ferramentas para PowerShell V4

Exemplo 1: esse exemplo mostra como retornar e extrair dados de uma série de um ou mais registros. O iterador fornecido Get-KINRecord determina a posição inicial dos registros a serem retornados, os quais, neste exemplo, são capturados em uma variável, $records. Cada registro individual pode então ser acessado indexando a coleção $records. Supondo que os dados no registro sejam texto codificado em UTF-8, o comando final mostra como você pode extrair os dados do MemoryStream objeto e retorná-los como texto para o console.

```
$records
$records = Get-KINRecord -ShardIterator "AAAAAAAAAAGIc....9VnbiRNaP"
```

Saída:

```
MillisBehindLatest NextShardIterator           Records
------------------ -----------------           -------
0                  AAAAAAAAAAERNIq...uDn11HuUs {Key1, Key2}
```

```
$records.Records[0]
```

Saída:

```
ApproximateArrivalTimestamp Data                    PartitionKey SequenceNumber
--------------------------- ----                    ------------ --------------
3/7/2016 5:14:33 PM         System.IO.MemoryStream Key1
 4955986459776...931586
```

```
[Text.Encoding]::UTF8.GetString($records.Records[0].Data.ToArray())
```

Saída:

```
test data from string
```

- Para obter detalhes da API, consulte GetRecordsem Referência de Ferramentas da AWS para PowerShell cmdlet (V4).

Ferramentas para PowerShell V5

Exemplo 1: esse exemplo mostra como retornar e extrair dados de uma série de um ou mais registros. O iterador fornecido Get-KINRecord determina a posição inicial dos registros a serem retornados, os quais, neste exemplo, são capturados em uma variável, $records. Cada registro individual pode então ser acessado indexando a coleção $records. Supondo que os dados no registro sejam texto codificado em UTF-8, o comando final mostra como você pode extrair os dados do MemoryStream objeto e retorná-los como texto para o console.

```
$records
$records = Get-KINRecord -ShardIterator "AAAAAAAAAAGIc....9VnbiRNaP"
```

Saída:

```
MillisBehindLatest NextShardIterator           Records
------------------ -----------------           -------
0                  AAAAAAAAAAERNIq...uDn11HuUs {Key1, Key2}
```

```
$records.Records[0]
```

Saída:

```
ApproximateArrivalTimestamp Data                    PartitionKey SequenceNumber
--------------------------- ----                    ------------ --------------
```

```
3/7/2016 5:14:33 PM         System.IO.MemoryStream Key1
 4955986459776...931586
```

```
[Text.Encoding]::UTF8.GetString($records.Records[0].Data.ToArray())
```

Saída:

```
test data from string
```

- Para obter detalhes da API, consulte GetRecordsem Referência de Ferramentas da AWS para PowerShell cmdlet (V5).

Python

SDK para Python (Boto3)

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
class KinesisStream:
    """Encapsulates a Kinesis stream."""

    def __init__(self, kinesis_client):
        """
        :param kinesis_client: A Boto3 Kinesis client.
        """
        self.kinesis_client = kinesis_client
        self.name = None
        self.details = None
        self.stream_exists_waiter = kinesis_client.get_waiter("stream_exists")


    def get_records(self, max_records):
        """
        Gets records from the stream. This function is a generator that first
 gets
```

```
        a shard iterator for the stream, then uses the shard iterator to get
records
        in batches from the stream. The shard iterator can be accessed through
the
        'details' property, which is populated using the 'describe' function of
this class.
        Each batch of records is yielded back to the caller until the specified
        maximum number of records has been retrieved.

        :param max_records: The maximum number of records to retrieve.
        :return: Yields the current batch of retrieved records.
        """
        try:
            response = self.kinesis_client.get_shard_iterator(
                StreamName=self.name,
                ShardId=self.details["Shards"][0]["ShardId"],
                ShardIteratorType="LATEST",
            )
            shard_iter = response["ShardIterator"]
            record_count = 0
            while record_count < max_records:
                response = self.kinesis_client.get_records(
                    ShardIterator=shard_iter, Limit=10
                )
                shard_iter = response["NextShardIterator"]
                records = response["Records"]
                logger.info("Got %s records.", len(records))
                record_count += len(records)
                yield records
        except ClientError:
            logger.exception("Couldn't get records from stream %s.", self.name)
            raise



    def describe(self, name):
        """
        Gets metadata about a stream.

        :param name: The name of the stream.
        :return: Metadata about the stream.
        """
        try:
            response = self.kinesis_client.describe_stream(StreamName=name)
```

```
            self.name = name
            self.details = response["StreamDescription"]
            logger.info("Got stream %s.", name)
        except ClientError:
            logger.exception("Couldn't get %s.", name)
            raise
        else:
            return self.details
```

- Para obter detalhes da API, consulte a GetRecordsReferência da API AWS SDK for Python (Boto3).

SAP ABAP

SDK para SAP ABAP

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
   TRY.
       oo_result = lo_kns->getrecords(             " oo_result is returned for
testing purposes. "
           iv_sharditerator = iv_shard_iterator ).
       DATA(lt_records) = oo_result->get_records( ).
       MESSAGE 'Record retrieved.' TYPE 'I'.
     CATCH /aws1/cx_knsexpirediteratorex.
       MESSAGE 'Iterator expired.' TYPE 'E'.
     CATCH /aws1/cx_knsinvalidargumentex.
       MESSAGE 'The specified argument was not valid.' TYPE 'E'.
     CATCH /aws1/cx_knskmsaccessdeniedex.
       MESSAGE 'You do not have permission to perform this AWS KMS action.' TYPE
'E'.
     CATCH /aws1/cx_knskmsdisabledex.
       MESSAGE 'KMS key used is disabled.' TYPE 'E'.
     CATCH /aws1/cx_knskmsinvalidstateex.
       MESSAGE 'KMS key used is in an invalid state. ' TYPE 'E'.
```

```
      CATCH /aws1/cx_knskmsnotfoundex.
        MESSAGE 'KMS key used is not found.' TYPE 'E'.
      CATCH /aws1/cx_knskmsoptinrequired.
        MESSAGE 'KMS key option is required.' TYPE 'E'.
      CATCH /aws1/cx_knskmsthrottlingex.
        MESSAGE 'The rate of requests to AWS KMS is exceeding the request
quotas.' TYPE 'E'.
      CATCH /aws1/cx_knsprovthruputexcdex.
        MESSAGE 'The request rate for the stream is too high, or the requested
data is too large for the available throughput.' TYPE 'E'.
      CATCH /aws1/cx_knsresourcenotfoundex.
        MESSAGE 'Resource being accessed is not found.' TYPE 'E'.
    ENDTRY.
```

- Para obter detalhes da API, consulte a GetRecordsreferência da API AWS SDK for SAP ABAP.

Para obter uma lista completa dos guias do desenvolvedor do AWS SDK e exemplos de código, consulteUsando esse serviço com um AWS SDK. Este tópico também inclui informações sobre como começar e detalhes sobre versões anteriores do SDK.

Usar **`GetShardIterator`** com uma CLI

Os exemplos de código a seguir mostram como usar o `GetShardIterator`.

Exemplos de ações são trechos de código de programas maiores e devem ser executados em contexto. É possível ver essa ação em contexto no seguinte exemplo de código:

- Conheça os conceitos básicos

CLI

AWS CLI

Para obter um iterador de fragmentos

O exemplo `get-shard-iterator` a seguir usa o tipo iterador de fragmento `AT_SEQUENCE_NUMBER` e gera um iterador de fragmento para começar a ler os registros de dados exatamente da posição indicada pelo número de sequência especificado.

```
aws kinesis get-shard-iterator \
    --stream-name samplestream \
    --shard-id shardId-000000000001 \
    --shard-iterator-type LATEST
```

Saída:

```
{
    "ShardIterator": "AAAAAAAAAAFEvJjIYI+3jw/4aqgH9FifJ+n48XWTh/
IFIsbILP6o5eDueD39NXNBfpZ10WL5K6ADXk8w+5H+Qhd9cFA9k268CPXCz/kebq1TGYI7Vy
+lUkA9BuN3xvATxMBGxRY3zYK05gqgvaIRn94O8SqeEqwhigwZxNWxID3Ej7YYYcxQi8Q/fIrCjGAy/
n2r5Z9G864YpWDfN9upNNQAR/iiOWKs"
}
```

Para obter mais informações, consulte Desenvolvimento de consumidores usando a API Kinesis Data Streams AWS com o SDK for Java no Guia do desenvolvedor do Amazon Kinesis Data Streams.

- Para obter detalhes da API, consulte GetShardIteratorem Referência de AWS CLI Comandos.

PowerShell

Ferramentas para PowerShell V4

Exemplo 1: retorna um iterador de fragmentos para o fragmento e a posição inicial especificados. Detalhes dos identificadores de fragmentos e dos números de sequência podem ser obtidos na saída do Get-KINStream cmdlet, fazendo referência à coleção Shards do objeto de stream retornado. O iterador retornado pode ser usado com o Get-KINRecord cmdlet para extrair registros de dados no fragmento.

```
Get-KINShardIterator -StreamName "mystream" -ShardId "shardId-000000000000" -
ShardIteratorType AT_SEQUENCE_NUMBER -StartingSequenceNumber "495598645..."
```

Saída:

```
AAAAAAAAAAGIc....9VnbiRNaP
```

- Para obter detalhes da API, consulte GetShardIteratorem Referência de Ferramentas da AWS para PowerShell cmdlet (V4).

Ferramentas para PowerShell V5

Exemplo 1: retorna um iterador de fragmentos para o fragmento e a posição inicial especificados. Detalhes dos identificadores de fragmentos e dos números de sequência podem ser obtidos na saída do Get-KINStream cmdlet, fazendo referência à coleção Shards do objeto de stream retornado. O iterador retornado pode ser usado com o Get-KINRecord cmdlet para extrair registros de dados no fragmento.

```
Get-KINShardIterator -StreamName "mystream" -ShardId "shardId-000000000000" -
ShardIteratorType AT_SEQUENCE_NUMBER -StartingSequenceNumber "495598645..."
```

Saída:

```
AAAAAAAAAAGIc....9VnbiRNaP
```

- Para obter detalhes da API, consulte GetShardIteratorem Referência de Ferramentas da AWS para PowerShell cmdlet (V5).

Para obter uma lista completa dos guias do desenvolvedor do AWS SDK e exemplos de código, consulteUsando esse serviço com um AWS SDK. Este tópico também inclui informações sobre como começar e detalhes sobre versões anteriores do SDK.

Use **`ListStreamConsumers`** com um AWS SDK

O código de exemplo a seguir mostra como usar `ListStreamConsumers`.

.NET

SDK para .NET

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
using System;
using System.Collections.Generic;
using System.Threading.Tasks;
using Amazon.Kinesis;
```

```
    using Amazon.Kinesis.Model;

    /// <summary>
    /// List the consumers of an Amazon Kinesis stream.
    /// </summary>
    public class ListConsumers
    {
        public static async Task Main()
        {
            IAmazonKinesis client = new AmazonKinesisClient();

            string streamARN = "arn:aws:kinesis:us-east-2:000000000000:stream/
AmazonKinesisStream";
            int maxResults = 10;

            var consumers = await ListConsumersAsync(client, streamARN,
 maxResults);

            if (consumers.Count > 0)
            {
                consumers
                    .ForEach(c => Console.WriteLine($"Name: {c.ConsumerName} ARN:
 {c.ConsumerARN}"));
            }
            else
            {
                Console.WriteLine("No consumers found.");
            }
        }

        /// <summary>
        /// Retrieve a list of the consumers for a Kinesis stream.
        /// </summary>
        /// <param name="client">An initialized Kinesis client object.</param>
        /// <param name="streamARN">The ARN of the stream for which we want to
        /// retrieve a list of clients.</param>
        /// <param name="maxResults">The maximum number of results to return.</
param>
        /// <returns>A list of Consumer objects.</returns>
        public static async Task<List<Consumer>>
 ListConsumersAsync(IAmazonKinesis client, string streamARN, int maxResults)
        {
            var request = new ListStreamConsumersRequest
            {
```

```
                StreamARN = streamARN,
                MaxResults = maxResults,
            };

            var response = await client.ListStreamConsumersAsync(request);

            return response.Consumers;
        }
    }
```

- Para obter detalhes da API, consulte ListStreamConsumersa Referência AWS SDK para .NET da API.

Para obter uma lista completa dos guias do desenvolvedor do AWS SDK e exemplos de código, consulteUsando esse serviço com um AWS SDK. Este tópico também inclui informações sobre como começar e detalhes sobre versões anteriores do SDK.

Use **`ListStreams`** com um AWS SDK ou CLI

Os exemplos de código a seguir mostram como usar o `ListStreams`.

.NET

SDK para .NET

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
using System;
using System.Collections.Generic;
using System.Threading.Tasks;
using Amazon.Kinesis;
using Amazon.Kinesis.Model;

/// <summary>
```

```
    /// Retrieves and displays a list of existing Amazon Kinesis streams.
    /// </summary>
    public class ListStreams
    {
        public static async Task Main(string[] args)
        {
            IAmazonKinesis client = new AmazonKinesisClient();
            var response = await client.ListStreamsAsync(new
 ListStreamsRequest());

            List<string> streamNames = response.StreamNames;

            if (streamNames.Count > 0)
            {
                streamNames
                    .ForEach(s => Console.WriteLine($"Stream name: {s}"));
            }
            else
            {
                Console.WriteLine("No streams were found.");
            }
        }
    }
```

- Para obter detalhes da API, consulte ListStreamsa Referência AWS SDK para .NET da API.

CLI

AWS CLI

Para listar fluxos de dados

O exemplo de `list-streams` a seguir lista todos os fluxos de dados ativos na conta e região atuais.

```
aws kinesis list-streams
```

Saída:

```
{
```

```
    "StreamNames": [
        "samplestream",
        "samplestream1"
    ]
}
```

Para obter mais informações, consulte Listar fluxos no Guia do desenvolvedor do Amazon Kinesis Data Streams.

- Para obter detalhes da API, consulte ListStreamsem Referência de AWS CLI Comandos.

Rust

SDK para Rust

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
async fn show_streams(client: &Client) -> Result<(), Error> {
    let resp = client.list_streams().send().await?;

    println!("Stream names:");

    let streams = resp.stream_names;
    for stream in &streams {
        println!("  {}", stream);
    }

    println!("Found {} stream(s)", streams.len());

    Ok(())
}
```

- Para obter detalhes da API, consulte a ListStreamsreferência da API AWS SDK for Rust.

SAP ABAP

SDK para SAP ABAP

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
   TRY.
       oo_result = lo_kns->liststreams(        " oo_result is returned for
testing purposes. "
           "Set Limit to specify that a maximum of streams should be returned."
           iv_limit = iv_limit ).
       DATA(lt_streams) = oo_result->get_streamnames( ).
       MESSAGE 'Streams listed.' TYPE 'I'.
     CATCH /aws1/cx_knslimitexceededex.
       MESSAGE 'The request processing has failed because of a limit exceed
exception.' TYPE 'E'.
   ENDTRY.
```

- Para obter detalhes da API, consulte a ListStreamsreferência da API AWS SDK for SAP ABAP.

Para obter uma lista completa dos guias do desenvolvedor do AWS SDK e exemplos de código, consulteUsando esse serviço com um AWS SDK. Este tópico também inclui informações sobre como começar e detalhes sobre versões anteriores do SDK.

Use **`ListTagsForStream`** com um AWS SDK ou CLI

Os exemplos de código a seguir mostram como usar o `ListTagsForStream`.

.NET

SDK para .NET

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
    using System;
    using System.Collections.Generic;
    using System.Threading.Tasks;
    using Amazon.Kinesis;
    using Amazon.Kinesis.Model;

    /// <summary>
    /// Shows how to list the tags that have been attached to an Amazon Kinesis
    /// stream.
    /// </summary>
    public class ListTags
    {
        public static async Task Main()
        {
            IAmazonKinesis client = new AmazonKinesisClient();
            string streamName = "AmazonKinesisStream";

            await ListTagsAsync(client, streamName);
        }

        /// <summary>
        /// List the tags attached to a Kinesis stream.
        /// </summary>
        /// <param name="client">An initialized Kinesis client object.</param>
        /// <param name="streamName">The name of the Kinesis stream for which you
        /// wish to display tags.</param>
        public static async Task ListTagsAsync(IAmazonKinesis client, string
streamName)
        {
            var request = new ListTagsForStreamRequest
            {
                StreamName = streamName,
```

```
                Limit = 10,
            };

            var response = await client.ListTagsForStreamAsync(request);
            DisplayTags(response.Tags);

            while (response.HasMoreTags)
            {
                request.ExclusiveStartTagKey = response.Tags[response.Tags.Count
- 1].Key;
                response = await client.ListTagsForStreamAsync(request);
            }
        }

        /// <summary>
        /// Displays the items in a list of Kinesis tags.
        /// </summary>
        /// <param name="tags">A list of the Tag objects to be displayed.</param>
        public static void DisplayTags(List<Tag> tags)
        {
            tags
                .ForEach(t => Console.WriteLine($"Key: {t.Key} Value:
{t.Value}"));
        }
    }
```

- Para obter detalhes da API, consulte ListTagsForStreama Referência AWS SDK para .NET da API.

CLI

AWS CLI

Para listar tags para um fluxo de dados

O exemplo `list-tags-for-stream` a seguir lista as tags anexadas ao fluxo de dados especificado.

```
aws kinesis list-tags-for-stream \
    --stream-name samplestream
```

Saída:

```
{
    "Tags": [
        {
            "Key": "samplekey",
            "Value": "example"
        }
    ],
    "HasMoreTags": false
}
```

Para obter mais informações, consulte Adicionar tags a fluxos no Guia do desenvolvedor do Amazon Kinesis Data Streams.

- Para obter detalhes da API, consulte ListTagsForStreamem Referência de AWS CLI Comandos.

Para obter uma lista completa dos guias do desenvolvedor do AWS SDK e exemplos de código, consulteUsando esse serviço com um AWS SDK. Este tópico também inclui informações sobre como começar e detalhes sobre versões anteriores do SDK.

Use **`PutRecord`** com um AWS SDK ou CLI

Os exemplos de código a seguir mostram como usar o `PutRecord`.

Exemplos de ações são trechos de código de programas maiores e devem ser executados em contexto. É possível ver essa ação em contexto no seguinte exemplo de código:

- Conheça os conceitos básicos

CLI

AWS CLI

Como gravar um registro em um fluxo de dados

O exemplo de `put-record` a seguir grava um único registro de dados no fluxo de dados especificado usando a chave de partição especificada.

```
aws kinesis put-record \
```

```
    --stream-name samplestream \
    --data sampledatarecord \
    --partition-key samplepartitionkey
```

Saída:

```
{
    "ShardId": "shardId-000000000009",
    "SequenceNumber": "49600902273357540915989931256901506243878407835297513618",
    "EncryptionType": "KMS"
}
```

Para obter mais informações, consulte Desenvolvimento de produtores usando a API Amazon Kinesis Data Streams AWS com o SDK for Java no Guia do desenvolvedor do Amazon Kinesis Data Streams.

- Para obter detalhes da API, consulte PutRecordem Referência de AWS CLI Comandos.

Java

SDK para Java 2.x

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
import software.amazon.awssdk.core.SdkBytes;
import software.amazon.awssdk.regions.Region;
import software.amazon.awssdk.services.kinesis.KinesisClient;
import software.amazon.awssdk.services.kinesis.model.PutRecordRequest;
import software.amazon.awssdk.services.kinesis.model.KinesisException;
import software.amazon.awssdk.services.kinesis.model.DescribeStreamRequest;
import software.amazon.awssdk.services.kinesis.model.DescribeStreamResponse;

/**
 * Before running this Java V2 code example, set up your development
 * environment, including your credentials.
 *
 * For more information, see the following documentation topic:
```

```
 *
 * https://docs.aws.amazon.com/sdk-for-java/latest/developer-guide/get-
started.html
 */
public class StockTradesWriter {
    public static void main(String[] args) {
        final String usage = """

                Usage:
                    <streamName>

                Where:
                    streamName - The Amazon Kinesis data stream to which records
 are written (for example, StockTradeStream)
                """;

        if (args.length != 1) {
            System.out.println(usage);
            System.exit(1);
        }

        String streamName = args[0];
        Region region = Region.US_EAST_1;
        KinesisClient kinesisClient = KinesisClient.builder()
                .region(region)
                .build();

        // Ensure that the Kinesis Stream is valid.
        validateStream(kinesisClient, streamName);
        setStockData(kinesisClient, streamName);
        kinesisClient.close();
    }

    public static void setStockData(KinesisClient kinesisClient, String
 streamName) {
        try {
            // Repeatedly send stock trades with a 100 milliseconds wait in
 between.
            StockTradeGenerator stockTradeGenerator = new StockTradeGenerator();

            // Put in 50 Records for this example.
            int index = 50;
            for (int x = 0; x < index; x++) {
                StockTrade trade = stockTradeGenerator.getRandomTrade();
```

```
                sendStockTrade(trade, kinesisClient, streamName);
                Thread.sleep(100);
            }

        } catch (KinesisException | InterruptedException e) {
            System.err.println(e.getMessage());
            System.exit(1);
        }
        System.out.println("Done");
    }

    private static void sendStockTrade(StockTrade trade, KinesisClient
kinesisClient,
            String streamName) {
        byte[] bytes = trade.toJsonAsBytes();

        // The bytes could be null if there is an issue with the JSON
serialization by
        // the Jackson JSON library.
        if (bytes == null) {
            System.out.println("Could not get JSON bytes for stock trade");
            return;
        }

        System.out.println("Putting trade: " + trade);
        PutRecordRequest request = PutRecordRequest.builder()
                .partitionKey(trade.getTickerSymbol()) // We use the ticker
symbol as the partition key, explained in
                                                       // the Supplemental
Information section below.
                .streamName(streamName)
                .data(SdkBytes.fromByteArray(bytes))
                .build();

        try {
            kinesisClient.putRecord(request);
        } catch (KinesisException e) {
            System.err.println(e.getMessage());
        }
    }

    private static void validateStream(KinesisClient kinesisClient, String
streamName) {
        try {
```

```
            DescribeStreamRequest describeStreamRequest =
 DescribeStreamRequest.builder()
                    .streamName(streamName)
                    .build();

            DescribeStreamResponse describeStreamResponse =
 kinesisClient.describeStream(describeStreamRequest);

            if (!
describeStreamResponse.streamDescription().streamStatus().toString().equals("ACTIVE"))
 {
                System.err.println("Stream " + streamName + " is not active.
 Please wait a few moments and try again.");
                System.exit(1);
            }

        } catch (KinesisException e) {
            System.err.println("Error found while describing the stream " +
 streamName);
            System.err.println(e);
            System.exit(1);
        }
    }
}
```

- Para obter detalhes da API, consulte PutRecorda Referência AWS SDK for Java 2.x da API.

PowerShell

Ferramentas para PowerShell V4

Exemplo 1: grava um registro contendo a string fornecida ao parâmetro -Text.

```
Write-KINRecord -Text "test data from string" -StreamName "mystream" -
PartitionKey "Key1"
```

Exemplo 2: grava um registro contendo os dados contidos no arquivo especificado. O arquivo é tratado como uma sequência de bytes, portanto, se contiver texto, ele deverá ser gravado com qualquer codificação necessária antes de ser usado com esse cmdlet.

```
Write-KINRecord -FilePath "C:\TestData.txt" -StreamName "mystream" -PartitionKey
 "Key2"
```

- Para obter detalhes da API, consulte PutRecordem Referência de Ferramentas da AWS para PowerShell cmdlet (V4).

Ferramentas para PowerShell V5

Exemplo 1: grava um registro contendo a string fornecida ao parâmetro -Text.

```
Write-KINRecord -Text "test data from string" -StreamName "mystream" -
PartitionKey "Key1"
```

Exemplo 2: grava um registro contendo os dados contidos no arquivo especificado. O arquivo é tratado como uma sequência de bytes, portanto, se contiver texto, ele deverá ser gravado com qualquer codificação necessária antes de ser usado com esse cmdlet.

```
Write-KINRecord -FilePath "C:\TestData.txt" -StreamName "mystream" -PartitionKey
 "Key2"
```

- Para obter detalhes da API, consulte PutRecordem Referência de Ferramentas da AWS para PowerShell cmdlet (V5).

Python

SDK para Python (Boto3)

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
class KinesisStream:
    """Encapsulates a Kinesis stream."""

    def __init__(self, kinesis_client):
        """
        :param kinesis_client: A Boto3 Kinesis client.
```

```
        """
        self.kinesis_client = kinesis_client
        self.name = None
        self.details = None
        self.stream_exists_waiter = kinesis_client.get_waiter("stream_exists")


    def put_record(self, data, partition_key):
        """
        Puts data into the stream. The data is formatted as JSON before it is
passed
        to the stream.

        :param data: The data to put in the stream.
        :param partition_key: The partition key to use for the data.
        :return: Metadata about the record, including its shard ID and sequence
number.
        """
        try:
            response = self.kinesis_client.put_record(
                StreamName=self.name, Data=json.dumps(data),
PartitionKey=partition_key
            )
            logger.info("Put record in stream %s.", self.name)
        except ClientError:
            logger.exception("Couldn't put record in stream %s.", self.name)
            raise
        else:
            return response
```

- Para obter detalhes da API, consulte a PutRecordReferência da API AWS SDK for Python (Boto3).

Rust

SDK para Rust

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
async fn add_record(client: &Client, stream: &str, key: &str, data: &str) ->
 Result<(), Error> {
    let blob = Blob::new(data);

    client
        .put_record()
        .data(blob)
        .partition_key(key)
        .stream_name(stream)
        .send()
        .await?;

    println!("Put data into stream.");

    Ok(())
}
```

- Para obter detalhes da API, consulte a PutRecordreferência da API AWS SDK for Rust.

SAP ABAP

SDK para SAP ABAP

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
   TRY.
       oo_result = lo_kns->putrecord(            " oo_result is returned for
testing purposes. "
           iv_streamname = iv_stream_name
           iv_data       = iv_data
           iv_partitionkey = iv_partition_key ).
       MESSAGE 'Record created.' TYPE 'I'.
     CATCH /aws1/cx_knsinvalidargumentex.
       MESSAGE 'The specified argument was not valid.' TYPE 'E'.
     CATCH /aws1/cx_knskmsaccessdeniedex.
       MESSAGE 'You do not have permission to perform this AWS KMS action.' TYPE
'E'.
     CATCH /aws1/cx_knskmsdisabledex.
       MESSAGE 'KMS key used is disabled.' TYPE 'E'.
     CATCH /aws1/cx_knskmsinvalidstateex.
       MESSAGE 'KMS key used is in an invalid state. ' TYPE 'E'.
     CATCH /aws1/cx_knskmsnotfoundex.
       MESSAGE 'KMS key used is not found.' TYPE 'E'.
     CATCH /aws1/cx_knskmsoptinrequired.
       MESSAGE 'KMS key option is required.' TYPE 'E'.
     CATCH /aws1/cx_knskmsthrottlingex.
       MESSAGE 'The rate of requests to AWS KMS is exceeding the request
quotas.' TYPE 'E'.
     CATCH /aws1/cx_knsprovthruputexcdex.
       MESSAGE 'The request rate for the stream is too high, or the requested
data is too large for the available throughput.' TYPE 'E'.
     CATCH /aws1/cx_knsresourcenotfoundex.
       MESSAGE 'Resource being accessed is not found.' TYPE 'E'.
   ENDTRY.
```

- Para obter detalhes da API, consulte a PutRecordreferência da API AWS SDK for SAP ABAP.

Para obter uma lista completa dos guias do desenvolvedor do AWS SDK e exemplos de código, consulteUsando esse serviço com um AWS SDK. Este tópico também inclui informações sobre como começar e detalhes sobre versões anteriores do SDK.

Use **`PutRecords`** com um AWS SDK ou CLI

Os exemplos de código a seguir mostram como usar o `PutRecords`.

CLI

AWS CLI

Para gravar vários registros em um fluxo de dados

O exemplo `put-records` a seguir grava um registro de dados usando a chave de partição especificada e outro registro de dados usando uma chave de partição diferente em uma única chamada.

```
aws kinesis put-records \
    --stream-name samplestream \
    --
records Data=blob1,PartitionKey=partitionkey1 Data=blob2,PartitionKey=partitionkey2
```

Saída:

```
{
    "FailedRecordCount": 0,
    "Records": [
        {
            "SequenceNumber":
 "49600883331171471519674795588238531498465399900093808706",
            "ShardId": "shardId-000000000004"
        },
        {
            "SequenceNumber":
 "49600902273357540915989931256902715169698037101720764562",
            "ShardId": "shardId-000000000009"
        }
    ],
    "EncryptionType": "KMS"
}
```

Para obter mais informações, consulte Desenvolvimento de produtores usando a API Amazon Kinesis Data Streams AWS com o SDK for Java no Guia do desenvolvedor do Amazon Kinesis Data Streams.

- Para obter detalhes da API, consulte PutRecordsem Referência de AWS CLI Comandos.

JavaScript

SDK para JavaScript (v3)

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
import { PutRecordsCommand, KinesisClient } from "@aws-sdk/client-kinesis";

/**
 * Put multiple records into a Kinesis stream.
 * @param {{ streamArn: string }} config
 */
export const main = async ({ streamArn }) => {
  const client = new KinesisClient({});
  try {
    await client.send(
      new PutRecordsCommand({
        StreamARN: streamArn,
        Records: [
          {
            Data: new Uint8Array(),
            /**
             * Determines which shard in the stream the data record is assigned
 to.
             * Partition keys are Unicode strings with a maximum length limit of
 256
             * characters for each key. Amazon Kinesis Data Streams uses the
 partition
             * key as input to a hash function that maps the partition key and
             * associated data to a specific shard.
             */
            PartitionKey: "TEST_KEY",
          },
          {
            Data: new Uint8Array(),
            PartitionKey: "TEST_KEY",
          },
```

```
        ],
      }),
    );
  } catch (caught) {
    if (caught instanceof Error) {
      //
    } else {
      throw caught;
    }
  }
};

// Call function if run directly.
import { fileURLToPath } from "node:url";
import { parseArgs } from "node:util";

if (process.argv[1] === fileURLToPath(import.meta.url)) {
  const options = {
    streamArn: {
      type: "string",
      description: "The ARN of the stream.",
    },
  };

  const { values } = parseArgs({ options });
  main(values);
}
```

- Para obter detalhes da API, consulte PutRecordsa Referência AWS SDK para JavaScript da API.

Para obter uma lista completa dos guias do desenvolvedor do AWS SDK e exemplos de código, consulteUsando esse serviço com um AWS SDK. Este tópico também inclui informações sobre como começar e detalhes sobre versões anteriores do SDK.

Use **`RegisterStreamConsumer`** com um AWS SDK ou CLI

Os exemplos de código a seguir mostram como usar o `RegisterStreamConsumer`.

.NET

SDK para .NET

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
    using System;
    using System.Threading.Tasks;
    using Amazon.Kinesis;
    using Amazon.Kinesis.Model;

    /// <summary>
    /// This example shows how to register a consumer to an Amazon Kinesis
    /// stream.
    /// </summary>
    public class RegisterConsumer
    {
        public static async Task Main()
        {
            IAmazonKinesis client = new AmazonKinesisClient();
            string consumerName = "NEW_CONSUMER_NAME";
            string streamARN = "arn:aws:kinesis:us-east-2:000000000000:stream/
AmazonKinesisStream";

            var consumer = await RegisterConsumerAsync(client, consumerName,
 streamARN);

            if (consumer is not null)
            {
                Console.WriteLine($"{consumer.ConsumerName}");
            }
        }

        /// <summary>
        /// Registers the consumer to a Kinesis stream.
        /// </summary>
        /// <param name="client">The initialized Kinesis client object.</param>
```

```
        /// <param name="consumerName">A string representing the consumer.</
param>
        /// <param name="streamARN">The ARN of the stream.</param>
        /// <returns>A Consumer object that contains information about the
 consumer.</returns>
        public static async Task<Consumer> RegisterConsumerAsync(IAmazonKinesis
 client, string consumerName, string streamARN)
        {
            var request = new RegisterStreamConsumerRequest
            {
                ConsumerName = consumerName,
                StreamARN = streamARN,
            };

            var response = await client.RegisterStreamConsumerAsync(request);
            return response.Consumer;
        }
    }
```

- Para obter detalhes da API, consulte RegisterStreamConsumera Referência AWS SDK para .NET da API.

CLI

AWS CLI

Para registrar um consumidor de fluxo de dados

O exemplo `register-stream-consumer` a seguir registra um consumidor chamado `KinesisConsumerApplication` com o fluxo de dados especificado.

```
aws kinesis register-stream-consumer \
    --stream-arn arn:aws:kinesis:us-west-2:012345678912:stream/samplestream \
    --consumer-name KinesisConsumerApplication
```

Saída:

```
{
    "Consumer": {
        "ConsumerName": "KinesisConsumerApplication",
```

```
        "ConsumerARN": "arn:aws:kinesis:us-west-2: 123456789012:stream/
samplestream/consumer/KinesisConsumerApplication:1572383852",
        "ConsumerStatus": "CREATING",
        "ConsumerCreationTimestamp": 1572383852.0
    }
}
```

Para obter mais informações, consulte Desenvolver consumidores com Fan-Out aprimorado usando a API Kinesis Data Streams no Guia do desenvolvedor do Amazon Kinesis Data Streams.

- Para obter detalhes da API, consulte RegisterStreamConsumerem Referência de AWS CLI Comandos.

SAP ABAP

SDK para SAP ABAP

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no AWS Code Examples Repository.

```
   TRY.
       oo_result = lo_kns->registerstreamconsumer(       " oo_result is returned
for testing purposes. "
           iv_streamarn = iv_stream_arn
           iv_consumername = iv_consumer_name ).
       MESSAGE 'Stream consumer registered.' TYPE 'I'.
     CATCH /aws1/cx_knsinvalidargumentex.
       MESSAGE 'The specified argument was not valid.' TYPE 'E'.
     CATCH /aws1/cx_sgmresourcelimitexcd.
       MESSAGE 'You have reached the limit on the number of resources.' TYPE
'E'.
     CATCH /aws1/cx_sgmresourceinuse.
       MESSAGE 'Resource being accessed is in use.' TYPE 'E'.
     CATCH /aws1/cx_sgmresourcenotfound.
       MESSAGE 'Resource being accessed is not found.' TYPE 'E'.
   ENDTRY.
```

- Para obter detalhes da API, consulte a RegisterStreamConsumerreferência da API AWS SDK for SAP ABAP.

Para obter uma lista completa dos guias do desenvolvedor do AWS SDK e exemplos de código, consulteUsando esse serviço com um AWS SDK. Este tópico também inclui informações sobre como começar e detalhes sobre versões anteriores do SDK.

Exemplos de tecnologia sem servidor para o Kinesis

Os exemplos de código a seguir mostram como usar o Kinesis com o. AWS SDKs

Exemplos

- Invocar uma função do Lambda em um trigger do Kinesis
- Relatando falhas de itens em lote para funções do Lambda com um trigger do Kinesis

Invocar uma função do Lambda em um trigger do Kinesis

Os exemplos de código a seguir mostram como implementar uma função do Lambda que recebe um evento acionado pelo recebimento de mensagens de stream do Kinesis. A função recupera a carga útil do Kinesis, decodifica do Base64 e registra o conteúdo do registro em log.

.NET

SDK para .NET

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no repositório dos Exemplos sem servidor.

Consumir um evento do Kinesis com o Lambda usando .NET.

```
// Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
// SPDX-License-Identifier: Apache-2.0
using System.Text;
using Amazon.Lambda.Core;
using Amazon.Lambda.KinesisEvents;
```

```
using AWS.Lambda.Powertools.Logging;

// Assembly attribute to enable the Lambda function's JSON input to be converted
 into a .NET class.
[assembly:
 LambdaSerializer(typeof(Amazon.Lambda.Serialization.SystemTextJson.DefaultLambdaJsonSeri

namespace KinesisIntegrationSampleCode;

public class Function
{
    // Powertools Logger requires an environment variables against your function
    // POWERTOOLS_SERVICE_NAME
    [Logging(LogEvent = true)]
    public async Task FunctionHandler(KinesisEvent evnt, ILambdaContext context)
    {
        if (evnt.Records.Count == 0)
        {
            Logger.LogInformation("Empty Kinesis Event received");
            return;
        }

        foreach (var record in evnt.Records)
        {
            try
            {
                Logger.LogInformation($"Processed Event with EventId:
 {record.EventId}");
                string data = await GetRecordDataAsync(record.Kinesis, context);
                Logger.LogInformation($"Data: {data}");
                // TODO: Do interesting work based on the new data
            }
            catch (Exception ex)
            {
                Logger.LogError($"An error occurred {ex.Message}");
                throw;
            }
        }
        Logger.LogInformation($"Successfully processed {evnt.Records.Count}
 records.");
    }

    private async Task<string> GetRecordDataAsync(KinesisEvent.Record record,
 ILambdaContext context)
```

```
    {
        byte[] bytes = record.Data.ToArray();
        string data = Encoding.UTF8.GetString(bytes);
        await Task.CompletedTask; //Placeholder for actual async work
        return data;
    }
}
```

Go

SDK para Go V2

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no repositório dos Exemplos sem servidor.

Consumir um evento do Kinesis com o Lambda usando Go.

```
// Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
// SPDX-License-Identifier: Apache-2.0
package main

import (
 "context"
 "log"

 "github.com/aws/aws-lambda-go/events"
 "github.com/aws/aws-lambda-go/lambda"
)

func handler(ctx context.Context, kinesisEvent events.KinesisEvent) error {
 if len(kinesisEvent.Records) == 0 {
  log.Printf("empty Kinesis event received")
  return nil
 }

 for _, record := range kinesisEvent.Records {
  log.Printf("processed Kinesis event with EventId: %v", record.EventID)
  recordDataBytes := record.Kinesis.Data
  recordDataText := string(recordDataBytes)
```

```
  log.Printf("record data: %v", recordDataText)
  // TODO: Do interesting work based on the new data
 }
 log.Printf("successfully processed %v records", len(kinesisEvent.Records))
 return nil
}

func main() {
 lambda.Start(handler)
}
```

Java

SDK para Java 2.x

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no repositório dos Exemplos sem servidor.

Consumir um evento do Kinesis com o Lambda usando Java.

```
// Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
// SPDX-License-Identifier: Apache-2.0
package example;

import com.amazonaws.services.lambda.runtime.Context;
import com.amazonaws.services.lambda.runtime.LambdaLogger;
import com.amazonaws.services.lambda.runtime.RequestHandler;
import com.amazonaws.services.lambda.runtime.events.KinesisEvent;

public class Handler implements RequestHandler<KinesisEvent, Void> {
    @Override
    public Void handleRequest(final KinesisEvent event, final Context context) {
        LambdaLogger logger = context.getLogger();
        if (event.getRecords().isEmpty()) {
            logger.log("Empty Kinesis Event received");
            return null;
        }
        for (KinesisEvent.KinesisEventRecord record : event.getRecords()) {
```

```
            try {
                logger.log("Processed Event with EventId: "+record.getEventID());
                String data = new String(record.getKinesis().getData().array());
                logger.log("Data:"+ data);
                // TODO: Do interesting work based on the new data
            }
            catch (Exception ex) {
                logger.log("An error occurred:"+ex.getMessage());
                throw ex;
            }
        }
        logger.log("Successfully processed:"+event.getRecords().size()+"
 records");
        return null;
    }

}
```

JavaScript

SDK para JavaScript (v3)

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no repositório dos Exemplos sem servidor.

Consumindo um evento do Kinesis com o uso do Lambda. JavaScript

```
// Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
// SPDX-License-Identifier: Apache-2.0
exports.handler = async (event, context) => {
  for (const record of event.Records) {
    try {
      console.log(`Processed Kinesis Event - EventID: ${record.eventID}`);
      const recordData = await getRecordDataAsync(record.kinesis);
      console.log(`Record Data: ${recordData}`);
      // TODO: Do interesting work based on the new data
    } catch (err) {
      console.error(`An error occurred ${err}`);
```

```
      throw err;
    }
  }
  console.log(`Successfully processed ${event.Records.length} records.`);
};

async function getRecordDataAsync(payload) {
  var data = Buffer.from(payload.data, "base64").toString("utf-8");
  await Promise.resolve(1); //Placeholder for actual async work
  return data;
}
```

Consumindo um evento do Kinesis com o uso do Lambda. TypeScript

```
// Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
// SPDX-License-Identifier: Apache-2.0
import {
  KinesisStreamEvent,
  Context,
  KinesisStreamHandler,
  KinesisStreamRecordPayload,
} from "aws-lambda";
import { Buffer } from "buffer";
import { Logger } from "@aws-lambda-powertools/logger";

const logger = new Logger({
  logLevel: "INFO",
  serviceName: "kinesis-stream-handler-sample",
});

export const functionHandler: KinesisStreamHandler = async (
  event: KinesisStreamEvent,
  context: Context
): Promise<void> => {
  for (const record of event.Records) {
    try {
      logger.info(`Processed Kinesis Event - EventID: ${record.eventID}`);
      const recordData = await getRecordDataAsync(record.kinesis);
      logger.info(`Record Data: ${recordData}`);
      // TODO: Do interesting work based on the new data
    } catch (err) {
      logger.error(`An error occurred ${err}`);
```

```
      throw err;
    }
    logger.info(`Successfully processed ${event.Records.length} records.`);
  }
};

async function getRecordDataAsync(
  payload: KinesisStreamRecordPayload
): Promise<string> {
  var data = Buffer.from(payload.data, "base64").toString("utf-8");
  await Promise.resolve(1); //Placeholder for actual async work
  return data;
}
```

PHP

SDK para PHP

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no repositório dos Exemplos sem servidor.

Consumir um evento do Kinesis com o Lambda usando PHP.

```
// Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
// SPDX-License-Identifier: Apache-2.0
<?php

# using bref/bref and bref/logger for simplicity

use Bref\Context\Context;
use Bref\Event\Kinesis\KinesisEvent;
use Bref\Event\Kinesis\KinesisHandler;
use Bref\Logger\StderrLogger;

require __DIR__ . '/vendor/autoload.php';

class Handler extends KinesisHandler
{
```

```
    private StderrLogger $logger;
    public function __construct(StderrLogger $logger)
    {
        $this->logger = $logger;
    }

    /**
     * @throws JsonException
     * @throws \Bref\Event\InvalidLambdaEvent
     */
    public function handleKinesis(KinesisEvent $event, Context $context): void
    {
        $this->logger->info("Processing records");
        $records = $event->getRecords();
        foreach ($records as $record) {
            $data = $record->getData();
            $this->logger->info(json_encode($data));
            // TODO: Do interesting work based on the new data

            // Any exception thrown will be logged and the invocation will be
 marked as failed
        }
        $totalRecords = count($records);
        $this->logger->info("Successfully processed $totalRecords records");
    }
}

$logger = new StderrLogger();
return new Handler($logger);
```

Python

SDK para Python (Boto3)

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no repositório dos Exemplos sem servidor.

Consumir um evento do Kinesis com o Lambda usando Python.

```
# Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
# SPDX-License-Identifier: Apache-2.0
import base64
def lambda_handler(event, context):

    for record in event['Records']:
        try:
            print(f"Processed Kinesis Event - EventID: {record['eventID']}")
            record_data = base64.b64decode(record['kinesis']
['data']).decode('utf-8')
            print(f"Record Data: {record_data}")
            # TODO: Do interesting work based on the new data
        except Exception as e:
            print(f"An error occurred {e}")
            raise e
    print(f"Successfully processed {len(event['Records'])} records.")
```

Ruby

SDK para Ruby

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no repositório dos Exemplos sem servidor.

Consumir um evento do Kinesis com o Lambda usando Ruby.

```
# Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
# SPDX-License-Identifier: Apache-2.0
require 'aws-sdk'

def lambda_handler(event:, context:)
  event['Records'].each do |record|
    begin
      puts "Processed Kinesis Event - EventID: #{record['eventID']}"
      record_data = get_record_data_async(record['kinesis'])
      puts "Record Data: #{record_data}"
      # TODO: Do interesting work based on the new data
```

```
    rescue => err
      $stderr.puts "An error occurred #{err}"
      raise err
    end
  end
  puts "Successfully processed #{event['Records'].length} records."
end

def get_record_data_async(payload)
  data = Base64.decode64(payload['data']).force_encoding('UTF-8')
  # Placeholder for actual async work
  # You can use Ruby's asynchronous programming tools like async/await or fibers
 here.
  return data
end
```

Rust

SDK para Rust

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no repositório dos Exemplos sem servidor.

Consuma um evento do Kinesis com o Lambda usando Rust.

```
// Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
// SPDX-License-Identifier: Apache-2.0
use aws_lambda_events::event::kinesis::KinesisEvent;
use lambda_runtime::{run, service_fn, Error, LambdaEvent};

async fn function_handler(event: LambdaEvent<KinesisEvent>) -> Result<(), Error>
 {
    if event.payload.records.is_empty() {
        tracing::info!("No records found. Exiting.");
        return Ok(());
    }

    event.payload.records.iter().for_each(|record| {
```

```
        tracing::info!("EventId:
 {}",record.event_id.as_deref().unwrap_or_default());

        let record_data = std::str::from_utf8(&record.kinesis.data);

        match record_data {
            Ok(data) => {
                // log the record data
                tracing::info!("Data: {}", data);
            }
            Err(e) => {
                tracing::error!("Error: {}", e);
            }
        }
    });

    tracing::info!(
        "Successfully processed {} records",
        event.payload.records.len()
    );

    Ok(())
}

#[tokio::main]
async fn main() -> Result<(), Error> {
    tracing_subscriber::fmt()
        .with_max_level(tracing::Level::INFO)
        // disable printing the name of the module in every log line.
        .with_target(false)
        // disabling time is handy because CloudWatch will add the ingestion
 time.
        .without_time()
        .init();

    run(service_fn(function_handler)).await
}
```

Para obter uma lista completa dos guias do desenvolvedor do AWS SDK e exemplos de código, consulteUsando esse serviço com um AWS SDK. Este tópico também inclui informações sobre como começar e detalhes sobre versões anteriores do SDK.

Relatando falhas de itens em lote para funções do Lambda com um trigger do Kinesis

Os exemplos de código a seguir mostram como implementar uma resposta parcial em lote para funções do Lambda que recebem eventos de um fluxo do Kinesis. A função relata as falhas do item em lote na resposta, sinalizando para o Lambda tentar novamente essas mensagens posteriormente.

.NET

SDK para .NET

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no repositório dos Exemplos sem servidor.

Relatar falhas de itens em lote do Kinesis com o Lambda usando o .NET.

```
// Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
// SPDX-License-Identifier: Apache-2.0
using System.Text;
using System.Text.Json.Serialization;
using Amazon.Lambda.Core;
using Amazon.Lambda.KinesisEvents;
using AWS.Lambda.Powertools.Logging;

// Assembly attribute to enable the Lambda function's JSON input to be converted
 into a .NET class.
[assembly:
 LambdaSerializer(typeof(Amazon.Lambda.Serialization.SystemTextJson.DefaultLambdaJsonSeri

namespace KinesisIntegration;

public class Function
{
    // Powertools Logger requires an environment variables against your function
    // POWERTOOLS_SERVICE_NAME
    [Logging(LogEvent = true)]
    public async Task<StreamsEventResponse> FunctionHandler(KinesisEvent evnt,
 ILambdaContext context)
```

```
    {
        if (evnt.Records.Count == 0)
        {
            Logger.LogInformation("Empty Kinesis Event received");
            return new StreamsEventResponse();
        }

        foreach (var record in evnt.Records)
        {
            try
            {
                Logger.LogInformation($"Processed Event with EventId:
{record.EventId}");
                string data = await GetRecordDataAsync(record.Kinesis, context);
                Logger.LogInformation($"Data: {data}");
                // TODO: Do interesting work based on the new data
            }
            catch (Exception ex)
            {
                Logger.LogError($"An error occurred {ex.Message}");
                /* Since we are working with streams, we can return the failed
item immediately.
                   Lambda will immediately begin to retry processing from this
failed item onwards. */
                return new StreamsEventResponse
                {
                    BatchItemFailures = new
List<StreamsEventResponse.BatchItemFailure>
                    {
                        new StreamsEventResponse.BatchItemFailure
{ ItemIdentifier = record.Kinesis.SequenceNumber }
                    }
                };
            }
        }
        Logger.LogInformation($"Successfully processed {evnt.Records.Count}
records.");
        return new StreamsEventResponse();
    }

    private async Task<string> GetRecordDataAsync(KinesisEvent.Record record,
ILambdaContext context)
    {
        byte[] bytes = record.Data.ToArray();
```

```
        string data = Encoding.UTF8.GetString(bytes);
        await Task.CompletedTask; //Placeholder for actual async work
        return data;
    }
}

public class StreamsEventResponse
{
    [JsonPropertyName("batchItemFailures")]
    public IList<BatchItemFailure> BatchItemFailures { get; set; }
    public class BatchItemFailure
    {
        [JsonPropertyName("itemIdentifier")]
        public string ItemIdentifier { get; set; }
    }
}
```

Go

SDK para Go V2

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no repositório dos Exemplos sem servidor.

Relatar falhas de itens em lote do Kinesis com o Lambda usando Go.

```
// Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
// SPDX-License-Identifier: Apache-2.0
package main

import (
 "context"
 "fmt"
 "github.com/aws/aws-lambda-go/events"
 "github.com/aws/aws-lambda-go/lambda"
)

func handler(ctx context.Context, kinesisEvent events.KinesisEvent)
 (map[string]interface{}, error) {
```

```
 batchItemFailures := []map[string]interface{}{}

 for _, record := range kinesisEvent.Records {
  curRecordSequenceNumber := ""

  // Process your record
  if /* Your record processing condition here */ {
   curRecordSequenceNumber = record.Kinesis.SequenceNumber
  }

  // Add a condition to check if the record processing failed
  if curRecordSequenceNumber != "" {
   batchItemFailures = append(batchItemFailures, map[string]interface{}
{"itemIdentifier": curRecordSequenceNumber})
  }
 }

 kinesisBatchResponse := map[string]interface{}{
  "batchItemFailures": batchItemFailures,
 }
 return kinesisBatchResponse, nil
}

func main() {
 lambda.Start(handler)
}
```

Java

SDK para Java 2.x

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no repositório dos Exemplos sem servidor.

Relatar falhas de itens em lote do Kinesis com o Lambda usando Java.

```
// Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
// SPDX-License-Identifier: Apache-2.0
```

```
import com.amazonaws.services.lambda.runtime.Context;
import com.amazonaws.services.lambda.runtime.RequestHandler;
import com.amazonaws.services.lambda.runtime.events.KinesisEvent;
import com.amazonaws.services.lambda.runtime.events.StreamsEventResponse;

import java.io.Serializable;
import java.util.ArrayList;
import java.util.List;

public class ProcessKinesisRecords implements RequestHandler<KinesisEvent,
 StreamsEventResponse> {

    @Override
    public StreamsEventResponse handleRequest(KinesisEvent input, Context
 context) {

        List<StreamsEventResponse.BatchItemFailure> batchItemFailures = new
 ArrayList<>();
        String curRecordSequenceNumber = "";

        for (KinesisEvent.KinesisEventRecord kinesisEventRecord :
 input.getRecords()) {
            try {
                //Process your record
                KinesisEvent.Record kinesisRecord =
 kinesisEventRecord.getKinesis();
                curRecordSequenceNumber = kinesisRecord.getSequenceNumber();

            } catch (Exception e) {
                /* Since we are working with streams, we can return the failed
 item immediately.
                   Lambda will immediately begin to retry processing from this
 failed item onwards. */
                batchItemFailures.add(new
 StreamsEventResponse.BatchItemFailure(curRecordSequenceNumber));
                return new StreamsEventResponse(batchItemFailures);
            }
        }

       return new StreamsEventResponse(batchItemFailures);
    }
}
```

JavaScript

SDK para JavaScript (v3)

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no repositório dos Exemplos sem servidor.

Relatar falhas de itens em lote do Kinesis com o Lambda usando Javascript.

```
// Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
// SPDX-License-Identifier: Apache-2.0
exports.handler = async (event, context) => {
  for (const record of event.Records) {
    try {
      console.log(`Processed Kinesis Event - EventID: ${record.eventID}`);
      const recordData = await getRecordDataAsync(record.kinesis);
      console.log(`Record Data: ${recordData}`);
      // TODO: Do interesting work based on the new data
    } catch (err) {
      console.error(`An error occurred ${err}`);
      /* Since we are working with streams, we can return the failed item
 immediately.
           Lambda will immediately begin to retry processing from this failed
 item onwards. */
      return {
        batchItemFailures: [{ itemIdentifier: record.kinesis.sequenceNumber }],
      };
    }
  }
  console.log(`Successfully processed ${event.Records.length} records.`);
  return { batchItemFailures: [] };
};

async function getRecordDataAsync(payload) {
  var data = Buffer.from(payload.data, "base64").toString("utf-8");
  await Promise.resolve(1); //Placeholder for actual async work
  return data;
}
```

Relatando falhas de itens em lote do Kinesis com o uso do Lambda. TypeScript

```
// Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
// SPDX-License-Identifier: Apache-2.0
import {
  KinesisStreamEvent,
  Context,
  KinesisStreamHandler,
  KinesisStreamRecordPayload,
  KinesisStreamBatchResponse,
} from "aws-lambda";
import { Buffer } from "buffer";
import { Logger } from "@aws-lambda-powertools/logger";

const logger = new Logger({
  logLevel: "INFO",
  serviceName: "kinesis-stream-handler-sample",
});

export const functionHandler: KinesisStreamHandler = async (
  event: KinesisStreamEvent,
  context: Context
): Promise<KinesisStreamBatchResponse> => {
  for (const record of event.Records) {
    try {
      logger.info(`Processed Kinesis Event - EventID: ${record.eventID}`);
      const recordData = await getRecordDataAsync(record.kinesis);
      logger.info(`Record Data: ${recordData}`);
      // TODO: Do interesting work based on the new data
    } catch (err) {
      logger.error(`An error occurred ${err}`);
      /* Since we are working with streams, we can return the failed item
 immediately.
           Lambda will immediately begin to retry processing from this failed
 item onwards. */
      return {
        batchItemFailures: [{ itemIdentifier: record.kinesis.sequenceNumber }],
      };
    }
  }
  logger.info(`Successfully processed ${event.Records.length} records.`);
  return { batchItemFailures: [] };
};
```

```
async function getRecordDataAsync(
  payload: KinesisStreamRecordPayload
): Promise<string> {
  var data = Buffer.from(payload.data, "base64").toString("utf-8");
  await Promise.resolve(1); //Placeholder for actual async work
  return data;
}
```

PHP

SDK para PHP

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no repositório dos Exemplos sem servidor.

Relatar falhas de itens em lote do Kinesis com o Lambda usando PHP.

```
// Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
// SPDX-License-Identifier: Apache-2.0
<?php

# using bref/bref and bref/logger for simplicity

use Bref\Context\Context;
use Bref\Event\Kinesis\KinesisEvent;
use Bref\Event\Handler as StdHandler;
use Bref\Logger\StderrLogger;

require __DIR__ . '/vendor/autoload.php';

class Handler implements StdHandler
{
    private StderrLogger $logger;
    public function __construct(StderrLogger $logger)
    {
        $this->logger = $logger;
    }
```

```
    /**
     * @throws JsonException
     * @throws \Bref\Event\InvalidLambdaEvent
     */
    public function handle(mixed $event, Context $context): array
    {
        $kinesisEvent = new KinesisEvent($event);
        $this->logger->info("Processing records");
        $records = $kinesisEvent->getRecords();

        $failedRecords = [];
        foreach ($records as $record) {
            try {
                $data = $record->getData();
                $this->logger->info(json_encode($data));
                // TODO: Do interesting work based on the new data
            } catch (Exception $e) {
                $this->logger->error($e->getMessage());
                // failed processing the record
                $failedRecords[] = $record->getSequenceNumber();
            }
        }
        $totalRecords = count($records);
        $this->logger->info("Successfully processed $totalRecords records");

        // change format for the response
        $failures = array_map(
            fn(string $sequenceNumber) => ['itemIdentifier' => $sequenceNumber],
            $failedRecords
        );

        return [
            'batchItemFailures' => $failures
        ];
    }
}

$logger = new StderrLogger();
return new Handler($logger);
```

Python

SDK para Python (Boto3)

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no repositório dos Exemplos sem servidor.

Relatar falhas de itens em lote do Kinesis com o Lambda usando Python.

```
# Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
# SPDX-License-Identifier: Apache-2.0
def handler(event, context):
    records = event.get("Records")
    curRecordSequenceNumber = ""

    for record in records:
        try:
            # Process your record
            curRecordSequenceNumber = record["kinesis"]["sequenceNumber"]
        except Exception as e:
            # Return failed record's sequence number
            return {"batchItemFailures":[{"itemIdentifier":
 curRecordSequenceNumber}]}

    return {"batchItemFailures":[]}
```

Ruby

SDK para Ruby

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no repositório dos Exemplos sem servidor.

Relatar falhas de item em lote do Kinesis com o Lambda usando Ruby.

```
# Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
# SPDX-License-Identifier: Apache-2.0
require 'aws-sdk'

def lambda_handler(event:, context:)
  batch_item_failures = []

  event['Records'].each do |record|
    begin
      puts "Processed Kinesis Event - EventID: #{record['eventID']}"
      record_data = get_record_data_async(record['kinesis'])
      puts "Record Data: #{record_data}"
      # TODO: Do interesting work based on the new data
    rescue StandardError => err
      puts "An error occurred #{err}"
      # Since we are working with streams, we can return the failed item
 immediately.
      # Lambda will immediately begin to retry processing from this failed item
 onwards.
      return { batchItemFailures: [{ itemIdentifier: record['kinesis']
['sequenceNumber'] }] }
    end
  end

  puts "Successfully processed #{event['Records'].length} records."
  { batchItemFailures: batch_item_failures }
end

def get_record_data_async(payload)
  data = Base64.decode64(payload['data']).force_encoding('utf-8')
  # Placeholder for actual async work
  sleep(1)
  data
end
```

Rust

SDK para Rust

> **Note**
>
> Tem mais sobre GitHub. Encontre o exemplo completo e saiba como configurar e executar no repositório dos Exemplos sem servidor.

Relate falhas de itens em lote do Kinesis com o Lambda usando Rust.

```
// Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
// SPDX-License-Identifier: Apache-2.0
use aws_lambda_events::{
    event::kinesis::KinesisEvent,
    kinesis::KinesisEventRecord,
    streams::{KinesisBatchItemFailure, KinesisEventResponse},
};
use lambda_runtime::{run, service_fn, Error, LambdaEvent};

async fn function_handler(event: LambdaEvent<KinesisEvent>) ->
 Result<KinesisEventResponse, Error> {
    let mut response = KinesisEventResponse {
        batch_item_failures: vec![],
    };

    if event.payload.records.is_empty() {
        tracing::info!("No records found. Exiting.");
        return Ok(response);
    }

    for record in &event.payload.records {
        tracing::info!(
            "EventId: {}",
            record.event_id.as_deref().unwrap_or_default()
        );

        let record_processing_result = process_record(record);

        if record_processing_result.is_err() {
            response.batch_item_failures.push(KinesisBatchItemFailure {
                item_identifier: record.kinesis.sequence_number.clone(),
```

```
            });
            /* Since we are working with streams, we can return the failed item
 immediately.
            Lambda will immediately begin to retry processing from this failed
 item onwards. */
            return Ok(response);
        }
    }

    tracing::info!(
        "Successfully processed {} records",
        event.payload.records.len()
    );

    Ok(response)
}

fn process_record(record: &KinesisEventRecord) -> Result<(), Error> {
    let record_data = std::str::from_utf8(record.kinesis.data.as_slice());

    if let Some(err) = record_data.err() {
        tracing::error!("Error: {}", err);
        return Err(Error::from(err));
    }

    let record_data = record_data.unwrap_or_default();

    // do something interesting with the data
    tracing::info!("Data: {}", record_data);

    Ok(())
}

#[tokio::main]
async fn main() -> Result<(), Error> {
    tracing_subscriber::fmt()
        .with_max_level(tracing::Level::INFO)
        // disable printing the name of the module in every log line.
        .with_target(false)
        // disabling time is handy because CloudWatch will add the ingestion
 time.
        .without_time()
        .init();
```

```
    run(service_fn(function_handler)).await
}
```

Para obter uma lista completa dos guias do desenvolvedor do AWS SDK e exemplos de código, consulteUsando esse serviço com um AWS SDK. Este tópico também inclui informações sobre como começar e detalhes sobre versões anteriores do SDK.

Histórico do documento

A tabela a seguir descreve as mudanças importantes na documentação do Amazon Kinesis Data Streams.

Alteração	Descrição	Alterado em
Foi adicionado o suporte para testes de resiliência com AWS Fault Injection Service.	Adição do Realize testes de resiliência com AWS Fault Injection Service.	15 de outubro de 2025
Foi adicionado o suporte para tamanhos de registro grandes até 10 MiB.	Adição do Gerenciar registros grandes.	27 de outubro de 2025
End-of-support datas e política de ciclo de vida de versões para KPL e KCL.	Foram adicionadas informações sobre end-of-support datas e políticas de ciclo de vida de versões para a Amazon Kinesis Producer Library (KPL) e a Amazon Kinesis Client Library (KCL). Para obter mais informações, consulte Política de ciclo de vida da versão da KPL e Política de ciclo de vida da versão da KCL.	13 de março de 2025
Foi adicionado o suporte para compartilhar fluxos de dados entre contas.	Adição do Compartilhamento do fluxo de dados com outra conta.	22 de novembro de 2023
Adição de suporte para os modos de capacidade de fluxo de dados	Adição do Escolha o modo certo para transmitir.	29 de novembro de 2021

Alteração	Descrição	Alterado em
provisionado e sob demanda.		
Novo conteúdo de criptografia do lado do servidor	Adição do Proteção de dados no Amazon Kinesis Data Streams.	7 de julho de 2017
Novo conteúdo para CloudWatch métricas aprimoradas.	Atualizado Monitorar o Kinesis Data Streams.	19 de abril de 2016
Novo conteúdo para agente aprimorado do Kinesis.	Atualizado Gravar no Amazon Kinesis Data Streams usando o Kinesis Agent.	11 de abril de 2016
Novo conteúdo para o uso de agentes do Kinesis.	Adição do Gravar no Amazon Kinesis Data Streams usando o Kinesis Agent.	2 de outubro de 2015
Atualização do conteúdo da KPL para a versão 0.10.0.	Adição do Desenvolver produtores usando a Amazon Kinesis Producer Library (KPL).	15 de julho de 2015
Atualização do tópico sobre métricas da KCL com métricas configuráveis.	Adição do Monitore a biblioteca de cliente do Kinesis com a Amazon CloudWatch.	9 de julho de 2015
Reorganização do conteúdo.	Reorganização significativa dos tópicos do conteúdo para obter visualização em árvore mais concisa e agrupamento mais lógico.	01 de julho de 2015

Alteração	Descrição	Alterado em
Novo tópico do guia do desenvolvedor da KPL.	Adição do Desenvolver produtores usando a Amazon Kinesis Producer Library (KPL).	02 de junho de 2015
Novo tópico sobre métricas da KCL.	Adição do Monitore a biblioteca de cliente do Kinesis com a Amazon CloudWatch.	19 de maio de 2015
Compatibilidade com KCL .NET	Adição do Desenvolver uma aplicação de consumo da Kinesis Client Library em .NET.	1 de maio de 2015
Compatibilidade com KCL Node.js	Adição do Desenvolver uma aplicação de consumo da Kinesis Client Library em Node.js.	26 de março de 2015
Compatibilidade com KCL Ruby	Adição de links para a biblioteca KCL Ruby.	12 de janeiro de 2015
Nova API PutRecords	Foram adicionadas informações sobre a nova PutRecords API aothe section called “Adicione vários registros com PutRecords”.	15 de dezembro de 2014
Compatibilidade com atribuição de tags	Adição do Marcar com tag seus recursos do Amazon Kinesis Data Streams.	11 de setembro de 2014
Nova CloudWatch métrica	Adição da métrica `GetRecords.IteratorAgeMilliseconds` a Métricas e dimensões do Amazon Kinesis Data Streams.	3 de setembro de 2014
Novo capítulo de monitoramento	Adicionadas Monitorar o Kinesis Data Streams e Monitore o serviço Amazon Kinesis Data Streams com a Amazon CloudWatch.	30 de julho de 2014
Limite padrão de fragmento	Atualização do Cotas e limites: o limite padrão de fragmento foi elevado de 5 para 10.	25 de fevereiro de 2014
Limite padrão de fragmento	Atualização do Cotas e limites: o limite padrão de fragmento foi elevado de 2 para 5.	28 de janeiro de 2014

Alteração	Descrição	Alterado em
Atualizações de versão de API	Atualizações para a versão 2013-12-02 da API do Kinesis Data Streams.	12 de dezembro de 2013
Lançamento inicial	Versão inicial do Guia do desenvolvedor do Amazon Kinesis.	14 de novembro de 2013

As traduções são geradas por tradução automática. Em caso de conflito entre o conteúdo da tradução e da versão original em inglês, a versão em inglês prevalecerá.