



Integrando microsserviços usando serviços sem servidor AWS

AWS Orientação prescritiva



AWS Orientação prescritiva: Integrando microsserviços usando serviços sem servidor AWS

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 desprestige 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

Introdução	1
Público-alvo	1
Objetivos	2
Segurança	2
Padrões de comunicação	3
Comunicação síncrona	3
Comunicação assíncrona	5
Acionar e esquecer	5
Verificação de declaração	6
Retorno de chamada	7
Comunicação bidirecional	8
Opções de coordenação	11
Orquestração	11
Exemplo: Step Functions	11
Exemplo: Amazon MWAA	14
Principais diferenças entre o Step Functions e o Amazon MWAA	16
Coreografia	16
Como escolher sua abordagem de coordenação	17
Gerenciando APIs	18
Amazon API Gateway	18
Autenticação e autorização	18
Chaves de API e limites de taxas	19
Público e privado APIs	19
Quando utilizar o API Gateway	20
Sistema de mensagens	21
Amazon SQS	21
Sondagem	22
Orientação	22
Amazon SNS	22
Orientação	24
Amazon EventBridge	24
Orientação	25
AWS AppSync Events e API Gateway	26
Orientação	26

Perguntas frequentes	27
Como posso combinar diferentes padrões de integração?	27
Qual é o principal benefício de usar uma arquitetura de microsserviços?	27
Como posso implementar o tratamento de erros?	27
Quais são os benefícios do padrão de verificação de declaração?	27
Quais são os benefícios do padrão de retorno de chamada?	28
Posso implementar a comunicação bidirecional?	28
Como posso otimizar o uso de funções do Lambda?	28
Quais são as principais diferenças entre usar o Amazon SNS e? EventBridge	28
Recursos	29
Documentação da AWS service (Serviço da AWS)	29
Leitura adicional	29
Histórico do documento	30
Glossário	31
#	31
A	32
B	35
C	37
D	41
E	45
F	47
G	49
H	50
eu	52
L	54
M	56
O	60
P	63
Q	66
R	66
S	69
T	73
U	75
V	75
W	76
Z	77

..... lxxviii

Integrando microsserviços usando serviços sem servidor AWS

Tabby Ward, Abhishek Agawane e Matt Kahn, Amazon Web Services

Setembro de 2025 ([histórico do documento](#))

Um aspecto importante da modernização dos softwares da sua organização é selecionar os padrões de arquitetura corretos que permitem agilidade e capacidade de resposta às mudanças nas necessidades de negócios. Em algumas aplicações, uma arquitetura monolítica é uma escolha comum. No entanto, para muitas organizações, a estratégia de modernização de [refatorar monólitos em microsserviços](#) pode ser eficaz quando o caso de uso se alinha aos benefícios dos microsserviços.

Microsserviços e monólitos não são mutuamente exclusivos: muitas organizações de sucesso empregam os dois padrões juntos, em que monólitos modulares atendem a alguns domínios e microsserviços lidam com outros.

Quando os microsserviços fazem parte da sua arquitetura, vários serviços podem ser chamados para obter dados para uma transação comercial. A implementação dessas integrações é necessária, pois requer um projeto cuidadoso para lidar com possíveis desafios, como consistência de dados, latência e complexidade operacional. Quando os microsserviços são integrados adequadamente, eles podem oferecer benefícios como escalabilidade independente, maior velocidade de desenvolvimento e otimização potencial de custos.

O guia faz parte de uma série de conteúdo que aborda a abordagem de modernização de aplicativos recomendada pela AWS. A série também inclui:

- [Estratégia para modernizar aplicativos no Nuvem AWS](#)
- [Abordagem em fases para modernizar aplicativos no Nuvem AWS](#)
- [Avaliação da prontidão da modernização para aplicações na Nuvem AWS](#)
- [Decompor monólitos em microsserviços](#)

Público-alvo

Este guia está destinado a proprietários de aplicações, empresários, líderes técnicos e gerentes de projeto que determinaram que microsserviços são apropriados para seu caso de uso específico.

O guia apresenta vários padrões de comunicação síncrona e assíncrona entre microsserviços usando sistemas sem servidor, como o Amazon AWS Lambda API Gateway, para Serviços da AWS autonomia e escalabilidade.

Objetivos

Ao usar este guia para integrar seus novos microsserviços, você pode transformar com eficiência a arquitetura da sua organização em uma arquitetura de microsserviços. Isso ajuda a fornecer um ajuste rápido às necessidades comerciais flutuantes por meio de alta escalabilidade, resiliência aprimorada, entrega contínua e isolamento de falhas. Uma arquitetura de microsserviços também acelera a inovação, pois cada microsserviço pode ser implantado e testado individualmente.

Uma arquitetura de microsserviços também pode ajudar a reduzir o tempo de comercialização de seus produtos ou serviços, porque cada microsserviço tem uma base de código independente que torna mais fácil e rápido adicionar novos recursos e iterá-los.

Segurança

Você deve proteger seus microsserviços adequadamente para proteger a integridade dos seus serviços e dados, mas garantir que a segurança não afete negativamente a performance da sua aplicação.

Em um ambiente de microsserviços, você deve considerar como cada serviço autenticará e autorizará as solicitações recebidas de um cliente externo ou outro microsserviço. Considere também como cada serviço acessará outros Serviços da AWS com segurança.

O acesso a Serviços da AWS deve ser concedido por meio de funções de escopo restrito [AWS Identity and Access Management \(IAM\)](#). Assumir um perfil do IAM fornece ao microsserviço credenciais do IAM de curto prazo na forma de uma chave de acesso, um segredo de acesso e um token de sessão. Eles são usados pelos vários kits de desenvolvimento de software (SDKs) para assinar solicitações Serviços da AWS usando o [AWS Signature Version 4 \(SigV4\)](#).

Padrões de comunicação

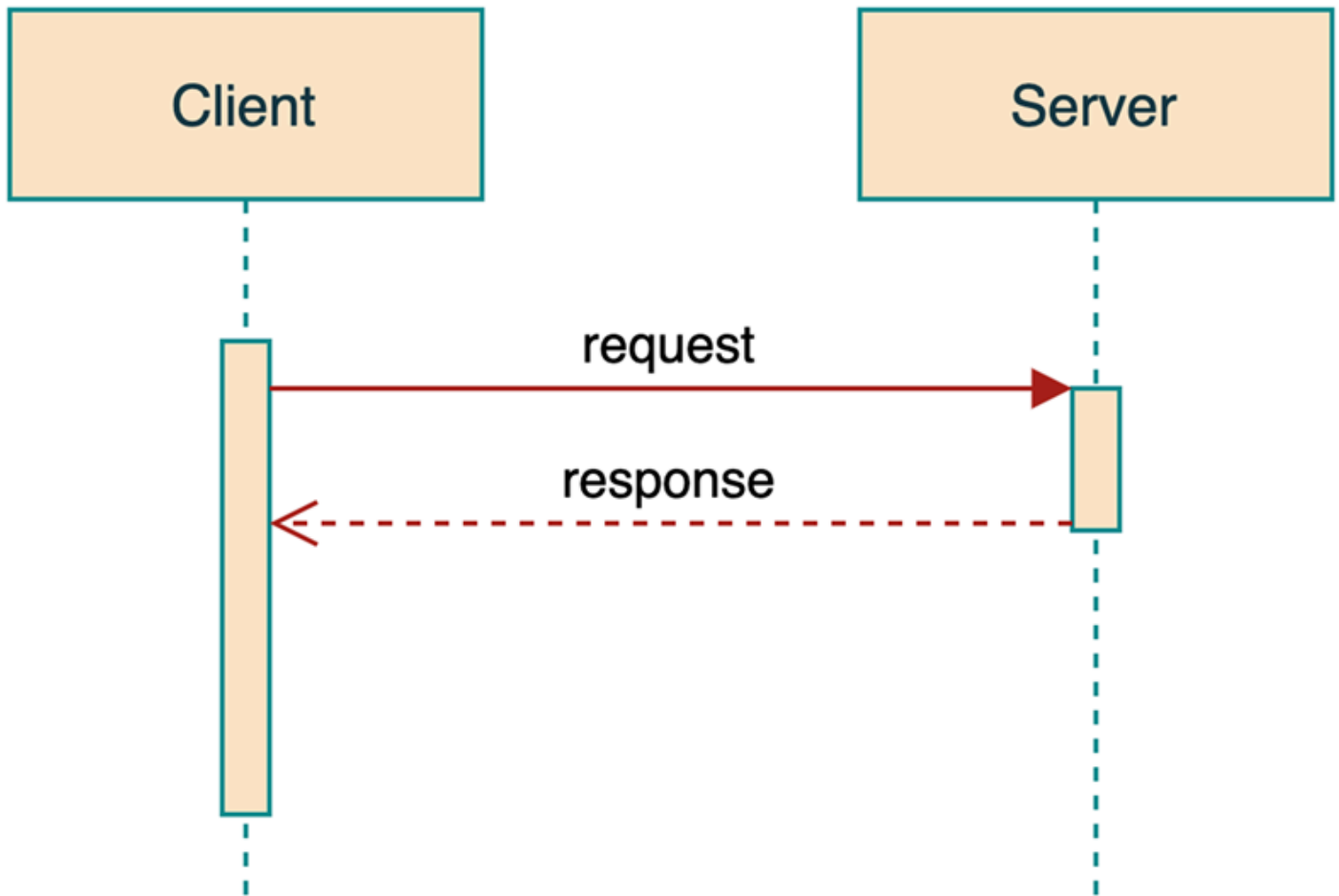
Em uma arquitetura de microsserviços, a comunicação ocorre em dois padrões principais: comunicação síncrona e comunicação assíncrona. Na comunicação síncrona, o chamador espera uma resposta antes de continuar, semelhante a uma chamada da API HTTP REST em tempo real. A comunicação assíncrona segue um padrão baseado em mensagens em que o chamador continua processando sem esperar uma resposta, como ao usar filas de mensagens. As próximas seções examinam detalhadamente a implementação, os benefícios e os casos de uso de cada padrão.

Tópicos

- [Comunicação síncrona](#)
- [Comunicação assíncrona](#)

Comunicação síncrona

Na comunicação síncrona, um cliente inicia uma solicitação para um serviço, conforme ilustrado no diagrama a seguir. Exemplos incluem uma solicitação para obter informações, como uma solicitação HTTP GET, ou uma solicitação para alterar dados, como uma solicitação HTTP PUT. Em ambos os casos, o cliente espera que o servidor responda antes de continuar. Chamadas síncronas são conhecidas da maioria dos desenvolvedores, são fáceis de implementar e solucionar problemas e, em muitos casos, são o padrão de comunicação amplamente aceito.



Os benefícios da comunicação síncrona incluem:

- Controle de fluxo previsível: execução determinística e ciclos claros de solicitação/resposta que são mais fáceis de entender em comparação com a comunicação assíncrona.
- Forte consistência: confirmação imediata de alterações de dados e atualizações de estado.
- Tratamento simples de erros: propagação direta de erros e exceções.
- Fácil depuração: rastreamento e monitoramento simples de solicitações.
- Suporte para protocolos: protocolos bem estabelecidos, como HTTP e REST, que simplificam a implementação.

A comunicação síncrona tem algumas desvantagens:

- Acoplamento forte: dependência direta dos serviços da disponibilidade uns dos outros.
- Impacto na rede: aumento da carga da rede devido às conexões abertas constantes.

- Utilização de recursos: maior uso de memória devido à manutenção dos estados de conexão.
- Falhas em cascata: capacidade de os problemas em um serviço se propagarem rapidamente pelo sistema.

Comunicação assíncrona

Por outro lado, na comunicação assíncrona, o cliente emite uma solicitação para um serviço, mas não recebe uma resposta imediata. Nesse caso, ele geralmente recebe apenas uma confirmação de que a solicitação foi aceita.

Os benefícios da comunicação assíncrona incluem:

- Suporte de arquitetura orientado por eventos: adequação natural para padrões de terceirização de eventos e segregação de responsabilidade por consultas de comando (CQRS).
- Melhor gerenciamento de recursos: capacidade dos serviços de processar solicitações com base em suas capacidades.
- Melhor isolamento de falhas: desacoplamento de serviços, o que evita falhas em cascata.
- Manipulação de picos de carga: melhor gerenciamento de picos de tráfego por meio do enfileiramento de mensagens.

As desvantagens incluem a complexidade. Por exemplo:

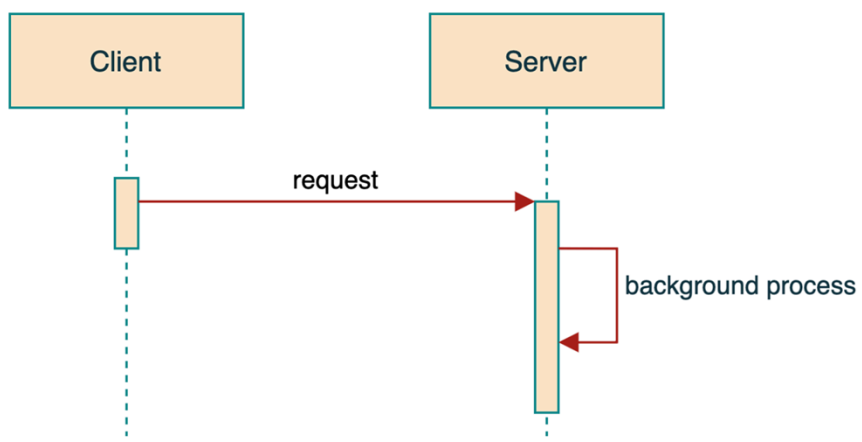
- Se o cliente exigir o resultado da operação assíncrona, a implementação de um mecanismo para obter ou receber esse resultado exigirá mais esforços.
- Pode ser mais difícil solucionar problemas em operações assíncronas, porque a solução de problemas exige análises de logs em vários sistemas.
- Pode ser mais difícil testar operações assíncronas, pois os testes exigem coordenação entre vários sistemas e serviços.

As abordagens de comunicação assíncrona incluem acionar e esquecer, verificar declarações, retorno de chamada e comunicação bidirecional.

Acionar e esquecer

No padrão “acionar e esquecer”, um cliente emite uma solicitação ao servidor e recebe de forma síncrona uma confirmação indicando que o servidor recebeu a mensagem e a processará. No

entanto, o processamento real ainda não ocorreu, e o cliente não tem visibilidade de quando ou como ele será feito. O diagrama a seguir ilustra esse padrão.



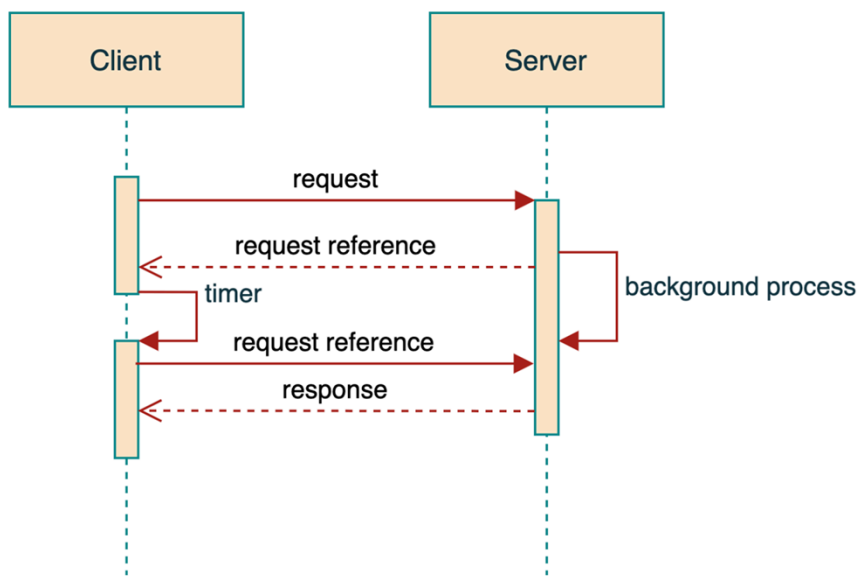
Nesse caso, o serviço não deve enviar a confirmação até que o objeto seja mantido de forma duradoura. Essa persistência pode ser implementada como uma operação de gravação no banco de dados ou colocando um item em uma fila.

Considerações adicionais:

- Implemente idempotência para lidar com mensagens duplicadas. Ou seja, cada mensagem deve ser processada apenas uma vez.
- Considere [filas de mensagens não entregues](#) em caso de falha no processamento.
- Monitore taxas de sucesso de processamento de mensagens.

Verificação de declaração

Se um cliente precisar do resultado de uma chamada de serviço, você poderá criar o serviço para emitir uma verificação de declaração ao receber uma solicitação. O diagrama a seguir ilustra esse padrão. A verificação da declaração é implementada como um identificador que o serviço retorna em sua confirmação. O cliente pode usar esse identificador posteriormente para verificar o status da solicitação e recuperar o resultado quando a solicitação estiver concluída.



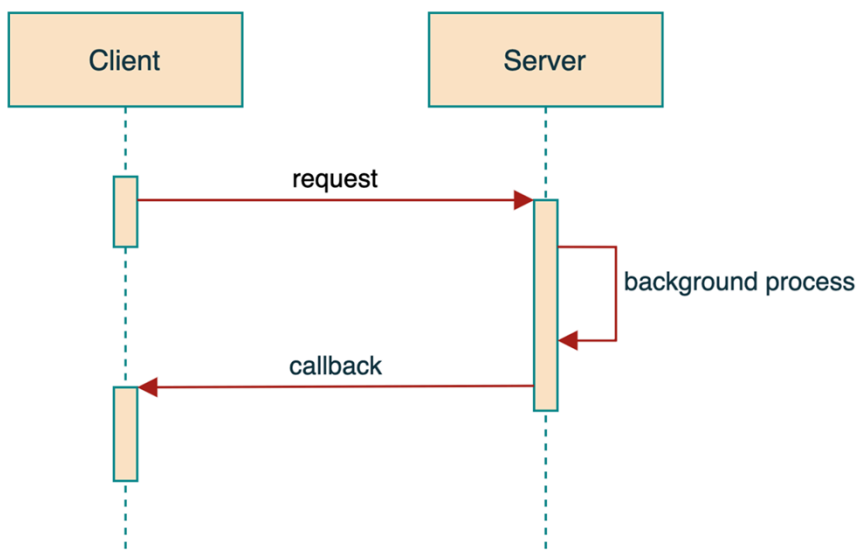
Os clientes devem implementar um mecanismo para pesquisar os resultados. Isso pode ser automatizado (por exemplo, uma verificação pode ser realizada a cada n minutos) ou implementado manualmente, em que a verificação é realizada em resposta a outro evento ou ação do usuário. Os serviços que implementam o padrão de verificação de declaração devem ser explícitos sobre o período de validade de uma verificação de declaração.

Práticas recomendadas:

- Implemente um recuo exponencial para sondagem.
- Defina um tempo de vida útil (TTL) para verificações de declarações.
- Forneça endpoints de status para acompanhamento do progresso.

Retorno de chamada

No padrão de retorno de chamada, um cliente emite uma solicitação para um serviço e fornece um local para o serviço entrar em contato quando o processamento estiver concluído. O cliente não espera um resultado, e o processamento continua. O serviço é responsável por entrar em contato com o local quando o processamento estiver concluído e fornecer o resultado. Os tipos comuns de locais para respostas são REST APIs ou filas. O diagrama a seguir ilustra o padrão de retorno de chamada.

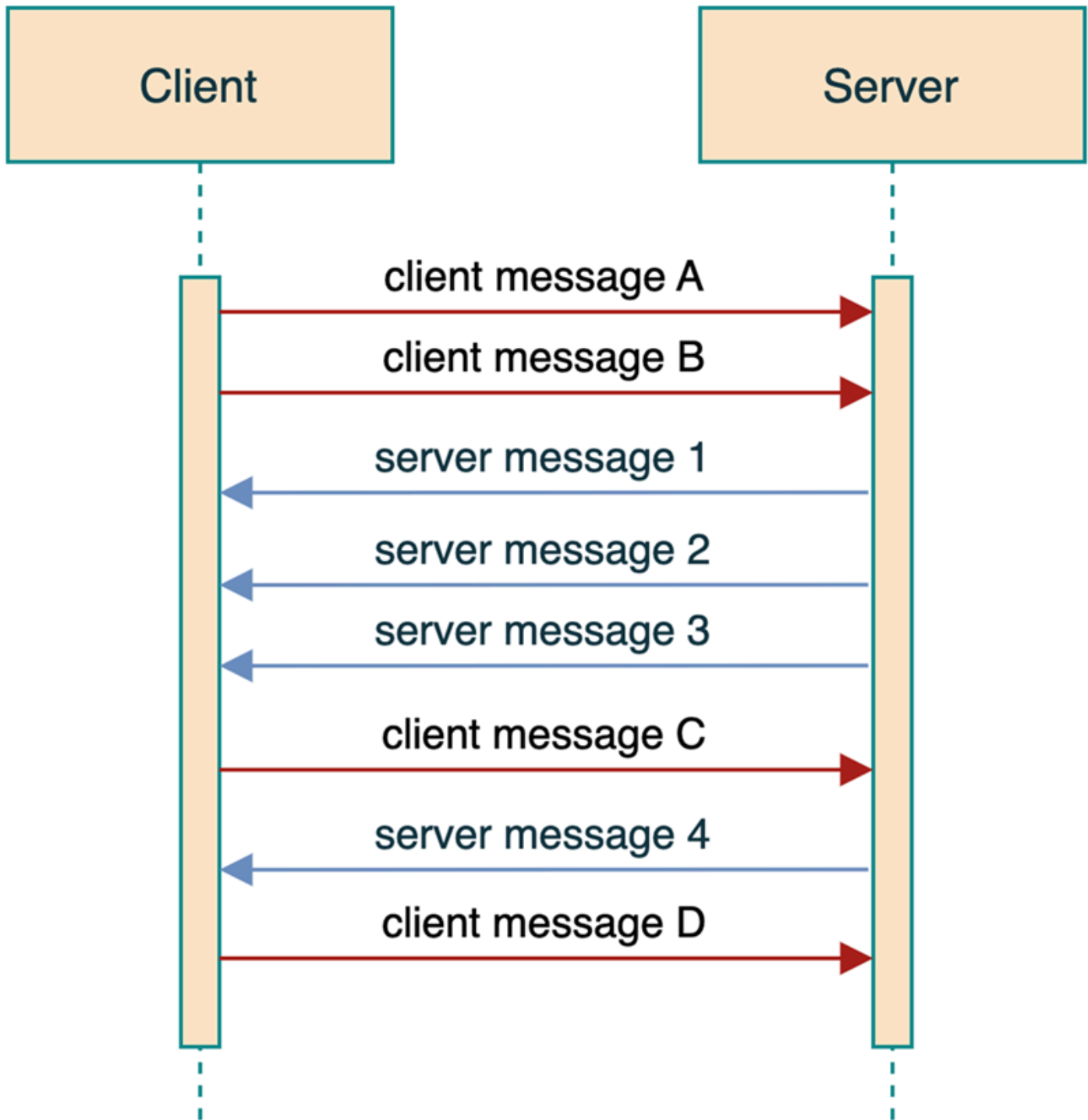


Implementação:

- Implemente mecanismos de repetição para retornos de chamada com falha.
- Proteja o local de retorno da chamada como faria com outros serviços.
- Gerencie os tempos limite de retorno de chamada.

Comunicação bidirecional

Para implementar a comunicação bidirecional, você deve criar uma conexão estável entre um cliente e um serviço, que permita que o cliente e o serviço enviem e processem mensagens. Isso é ilustrado no diagrama a seguir. Embora a comunicação seja assíncrona, o serviço deve ser capaz de sustentar uma conexão aberta para cada cliente.



Considerações de implementação:

- Ordenação de mensagens
- Números de sequência

- Estratégias de partição
- Ordenação de mensagens
- Gerenciamento de estados
 - Padrões de fornecimento de eventos
 - Conciliação de estado
 - Modelos de consistência
- Tratamento de erros
 - [Filas de mensagens não entregues](#)
 - Política de novas tentativas
 - [Disjuntores](#)
 - Estratégias de fallback
- Monitoramento e observabilidade
 - Correlação IDs
 - Rastrear mensagens
 - Métricas de performance
 - Indicadores de integridade do sistema

Opções de coordenação

A comunicação síncrona e assíncrona funciona bem para um cliente que chama um único serviço ou um pequeno número de serviços. No entanto, em um ambiente real, essa comunicação pode rapidamente se tornar complicada e com escalabilidade limitada. A realização de uma unidade de trabalho pode ser necessário que envolva vários microsserviços, que podem ter interdependências. Frequentemente, essas interações são modeladas como um fluxo de trabalho. Existem duas abordagens para projetar esses fluxos de trabalho: orquestração e coreografia.

Tópicos

- [Orquestração](#)
- [Coreografia](#)
- [Como escolher sua abordagem de coordenação](#)

Orquestração

Nessa abordagem, um único orquestrador é responsável por chamar cada microsserviço, determinar se as chamadas devem ser feitas em sequência ou em paralelo, manipular as respostas individuais dos serviços ao longo do processo e compilar o resultado final. Um orquestrador pode misturar invocações síncronas e assíncronas.

O [AWS Step Functions](#) e o [Amazon Managed Workflows for Apache Airflow \(Amazon MWAA\)](#) são ótimas opções para orquestradores de fluxo de trabalho.

A orquestração é uma boa escolha quando há ramificações lógicas em seu processo e você precisa de um único local para encapsular essa lógica. Também é útil quando você deseja implementar o padrão de verificação de reivindicação assíncrona. Por exemplo, os fluxos de trabalho padrão no Step Functions podem pausar um fluxo de trabalho e aguardar um retorno de chamada de outro serviço. O uso de um orquestrador também melhora o monitoramento e a observabilidade de um processo.

Exemplo: Step Functions

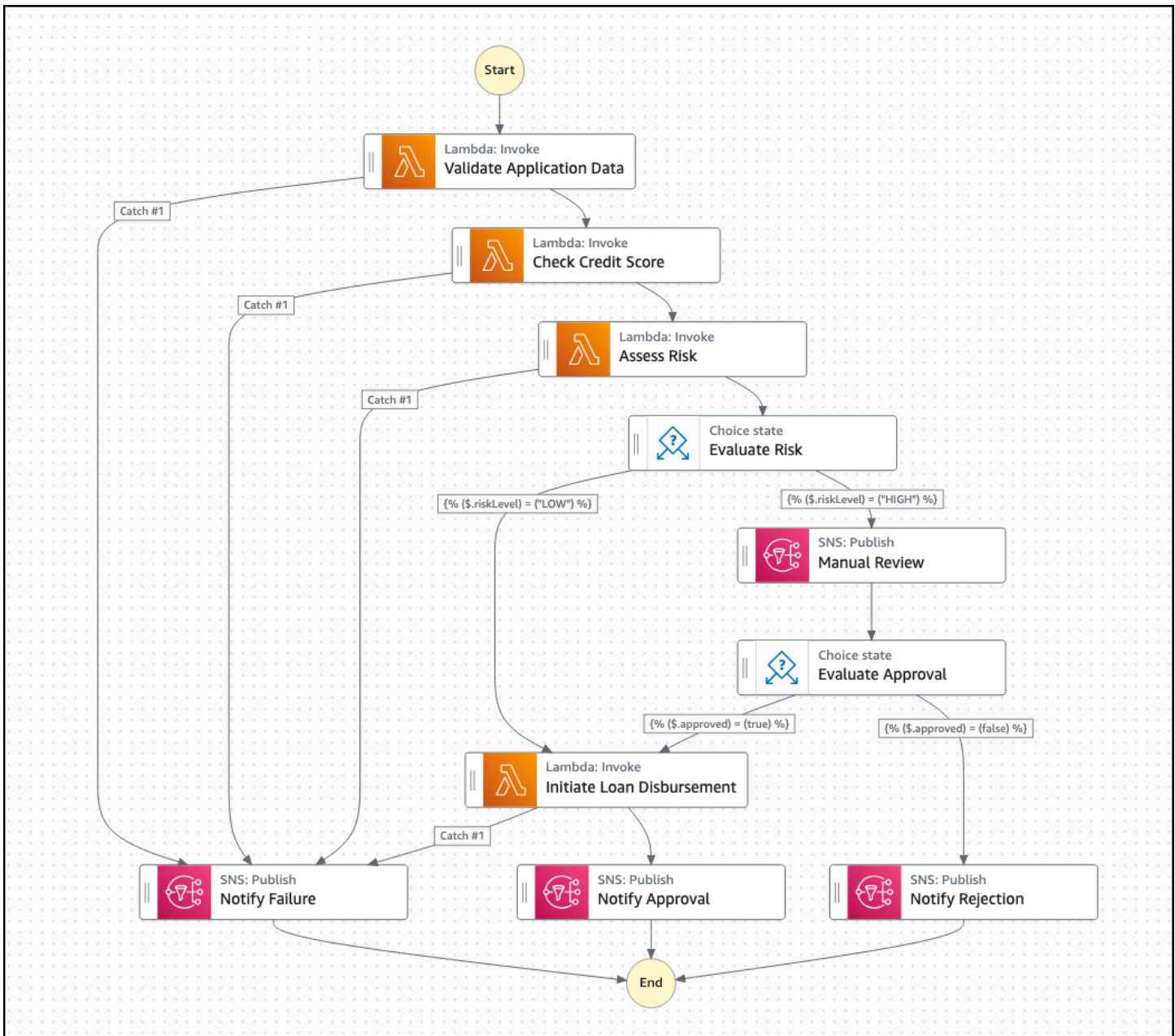
Você pode usar o Step Functions para coordenar várias funções do Lambda e Serviços da AWS outras para criar fluxos de trabalho complexos para integração de microsserviços. Essa opção é particularmente útil para processos longos e com várias etapas que envolvem vários microsserviços.

Considere o uso do Step Functions se:

- A integração do seu microsserviço envolve processos complexos e com várias etapas.
- Você precisa manter o estado durante operações de longa duração.
- Você deseja implementar o tratamento de erros e a lógica de repetição no nível do fluxo de trabalho.
- Você precisa coordenar operações síncronas e assíncronas.

O Step Functions oferece um editor visual para projetar fluxos de trabalho complexos, o que simplifica o processo de criação e gerenciamento de máquinas de estado. Ele fornece mecanismos integrados de tratamento de erros, incluindo lógica de repetição e gerenciamento de estado de erro, que aumentam a confiabilidade e a robustez das suas aplicações. Os fluxos de trabalho padrão oferecem suporte a processos de longa duração, de até um ano, o que é adequado para fluxos de trabalho que abrangem períodos prolongados. Essa opção separa a lógica de orquestração do código da aplicação, reduzindo significativamente a complexidade do código. Isso significa que os desenvolvedores podem se concentrar na lógica central do negócio, enquanto o Step Functions lida com o controle de fluxo e a coordenação dos componentes distribuídos.

Por exemplo, considere um processo de aprovação de empréstimo em uma aplicação de serviços financeiros, ilustrado no diagrama a seguir. O processo começa quando uma aplicação de empréstimo é apresentada.



Na máquina de estados ilustrada no diagrama anterior, o Step Functions coordena as seguintes etapas:

- Validar dados da aplicação (função do Lambda)
- Verificar pontuação de crédito (função do Lambda que chama uma API externa)
- Avaliar o risco (função do Lambda)
- Se o risco for alto, encaminhar para revisão manual (tarefa de aprovação humana)
- Se aprovado, inicie o desembolso do empréstimo (função do Lambda)

- Enviar notificação ao candidato (Amazon SNS)

Você pode usar essa abordagem para gerenciar um processo complexo e potencialmente demorado de maneira confiável, com tratamento de erros integrado e a capacidade de incluir etapas automatizadas e manuais.

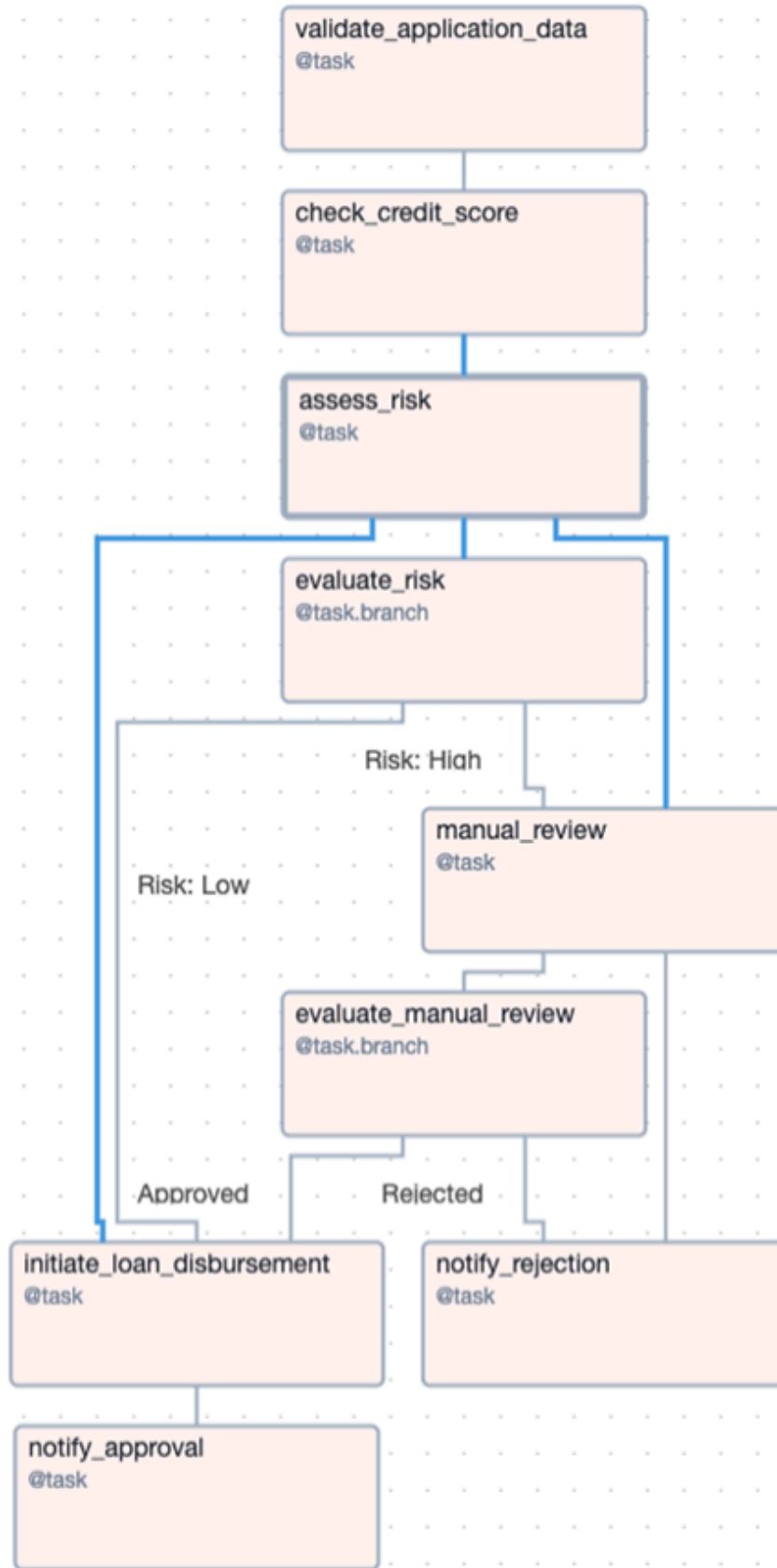
Considerações:

- Projete sua máquina de estados cuidadosamente para lidar com todos os cenários possíveis.
- Execute as etapas em paralelo, sempre que possível.
- Use os mecanismos integrados de tratamento de erros e repetição de tentativas do Step Functions para falhas permanentes e temporárias.
- Considere usar [fluxos de trabalho padrão ou expressos](#) com base no seu caso de uso. Os fluxos de trabalho expressos podem ser preferíveis para fluxos de trabalho de curta duração ou de alto volume.
- [Monitore as métricas de execução](#) para otimizar seu fluxo de trabalho.
- Use fluxos de trabalho aninhados para encapsular e reutilizar funcionalidades em várias máquinas de estado.
- Para fluxos de trabalho complexos, considere usar o [Amazon Bedrock Agents](#) como alternativa ao Step Functions.

Para obter mais informações, consulte a [documentação do Step Functions](#).

Exemplo: Amazon MWAA

Se sua organização já utiliza o Apache Airflow, o Amazon MWAA é uma escolha natural como orquestrador de fluxo de trabalho. No Apache Airflow, você cria seus fluxos de trabalho como grafos acíclicos direcionados (DAGs) usando Python. A representação DAG da máquina de estados ilustrada na seção Step Functions pode ser semelhante a esta:



Para obter informações sobre como trabalhar com DAGs, consulte a documentação do [Amazon MWAA](#).

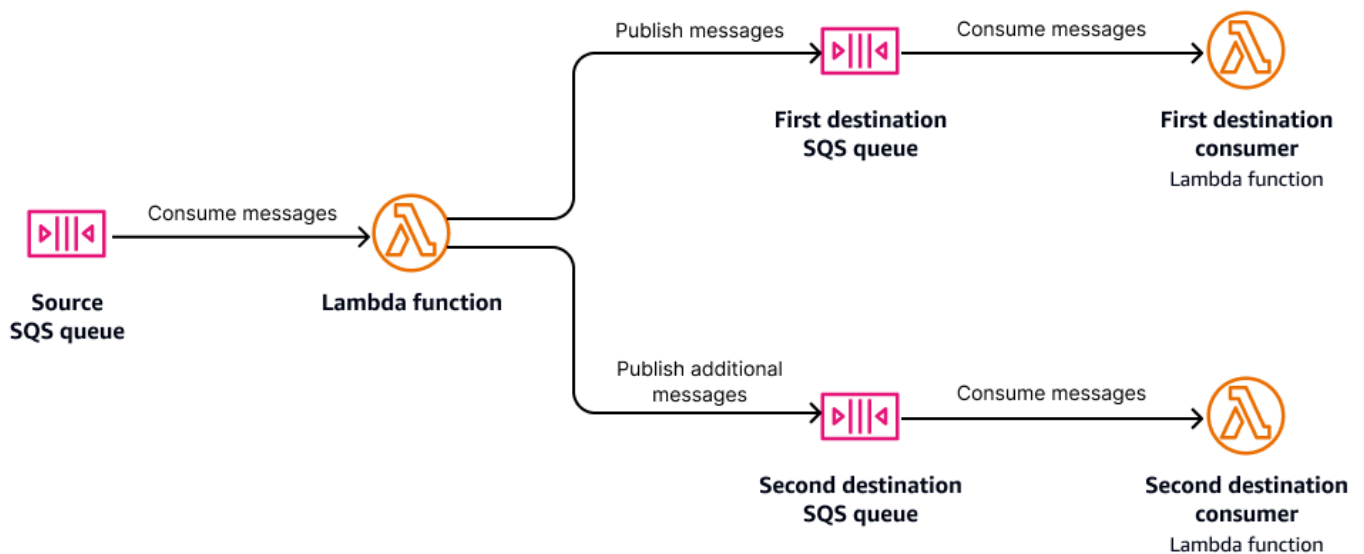
Principais diferenças entre o Step Functions e o Amazon MWAA

- O Step Functions é um serviço sem servidor totalmente gerenciado, portanto, não há infraestrutura para provisionar previamente e não é necessário programar uma janela de manutenção. O Amazon MWAA deve passar por sua implantação com antecedência, e você escolhe o tamanho e o número de nós em seu cluster.
- No Step Functions, você pode criar máquinas de estado de várias maneiras, incluindo o Workflow Studio, diretamente como JSON ou usando o AWS Cloud Development Kit (AWS CDK). O Apache DAGs Airflow é escrito em Python.
- Com o Step Functions, você não incorre em nenhum custo quando não há fluxos de trabalho em execução. Com o Amazon MWAA, você incorrerá em custos mesmo quando não DAGs estiver funcionando.

Coreografia

Em um sistema coreografado, os componentes individuais recebem uma tarefa, realizam algum trabalho e, possivelmente, emitem uma tarefa para que o trabalho subsequente seja realizado. Não existe um mecanismo central de coordenação. A coreografia facilita a escalabilidade independente dos serviços, pois cada serviço opera de forma relativamente isolada. Ele executa o trabalho quando o recebe, independentemente do throughput do serviço. A coreografia costuma ser uma parte central de uma [arquitetura orientada a eventos \(EDA\)](#).

No diagrama a seguir, não há coordenação entre as funções do Lambda. Cada função processa apenas mensagens na fila inscrita. Cada função é responsável pelo seu próprio tratamento de erros e pode controlar a simultaneidade; por exemplo, se uma dependência downstream tiver um limite de solicitações por segundo (RPS).



Uma EDA oferece vários benefícios, como acoplamento flexível de serviços e extensibilidade. Uma discussão completa sobre os princípios da EDA está além do escopo deste guia. Para obter mais informações, consulte:

- [AWS Well-Architected Framework — lente de aplicação sem servidor](#)
- [Introdução à arquitetura orientada a eventos](#) (Serverless Land)
- [Transição para uma arquitetura orientada a eventos](#) (Guia do desenvolvedor de tecnologias sem servidor)

Como escolher sua abordagem de coordenação

A coreografia e a orquestração têm suas utilidades na integração de microsserviços. Escolha a coreografia dentro dos limites de um único microsserviço, onde você tem controle total sobre as dependências. Escolha a orquestração quando trabalhar além dos limites dos microsserviços. Por exemplo, vários microsserviços que participam de uma transação distribuída se beneficiarão da orquestração para contabilizar a reversão de falhas. Os microsserviços que lidam com eventos que podem ser de interesse para outros microsserviços se beneficiarão da coreografia e de uma arquitetura orientada a eventos.

Um padrão comum para implementar a reversão quando vários sistemas estão envolvidos em uma única transação é o padrão saga.

Gerenciando APIs

O gerenciamento adequado de APIs torna seus microsserviços acessíveis tanto para consumidores internos quanto externos. AWS oferece uma variedade de serviços que você pode usar em conjunto para expor seu microsserviço com segurança. APIs Esses serviços permitem que você reforce a segurança APIs e implemente monitoramento e observabilidade a partir de um local central. Você também pode usar CloudFront a [Amazon](#) para melhorar o desempenho se seus usuários APIs estiverem geograficamente distantes do local em Região da AWS que os serviços estão hospedados.

Amazon API Gateway

[O Amazon API Gateway](#) é um serviço totalmente gerenciado que permite que os desenvolvedores criem, publiquem, mantenham, monitorem e protejam REST WebSocket APIs em qualquer escala. É possível usar o API Gateway para implementar muitos dos padrões descritos na seção [Padrões de comunicação](#) deste guia.

Há dois tipos principais de REST APIs: REST e HTTP. Ambos os tipos oferecem suporte RESTful APIs, mas oferecem recursos diferentes. Para determinar o que melhor atende às suas necessidades, consulte [Escolha entre REST APIs e HTTP APIs](#) na documentação do API Gateway. Esta seção do guia se concentra no API Gateway REST APIs.

Usar o API Gateway como ponto de entrada para você APIs fornece um único local para implementar preocupações comuns, como validação e segurança de solicitações. O API Gateway REST APIs oferece [validação de solicitações](#), o que permite definir o formato de suas solicitações usando o [esquema JSON](#). O API Gateway valida as solicitações recebidas em relação ao esquema definido e rejeita aquelas malformadas.

Autenticação e autorização

O API Gateway REST APIs oferece suporte aos seguintes mecanismos de autenticação (AuthN) e autorização (AuthZ):

- IAM: se você usa o IAM, as solicitações para sua API devem ser assinadas usando o [AWS Signature Version 4 \(SigV4\)](#).
- Amazon Cognito: o API Gateway validará um token portador apresentado como tendo sido emitido por um grupo de usuários do Amazon Cognito. Você também pode configurar o grupo de usuários

do Amazon Cognito para integração a um provedor de identidades (IdP) de terceiros, caso já esteja usando um. Você também pode usar um grupo de usuários do Amazon Cognito para autenticação machine-to-machine (M2M).

- AWS Lambda authorizer — O API Gateway invocará uma função Lambda que você especificar para realizar as verificações desejadas, a fim de determinar se uma solicitação deve ser autorizada.

Para obter mais informações, consulte [Controlar e gerenciar o acesso ao REST APIs](#) na documentação do API Gateway.

Chaves de API e limites de taxas

Você pode controlar quem tem permissão para ligar para você APIs e a que taxa usando chaves de API e planos de uso. Chaves de API não devem ser usadas para autenticação, mas podem ser usadas em conjunto com os esquemas mencionados anteriormente. Os usuários nem sempre precisam fornecer suas próprias chaves de API. Por exemplo, autorizadores do Lambda podem retornar uma chave de API para um usuário. O plano de uso permite especificar o throughput, o limite de intermitência e a cota mensal. Para obter mais informações, consulte [Planos de uso e chaves de API para REST APIs](#) na documentação do API Gateway.

Público e privado APIs

O API Gateway REST APIs que pode ser acessado pela Internet oferece suporte a dois tipos de endpoint:

- Otimizado para borda, o que significa que as solicitações dos chamadores são encaminhadas para um CloudFront ponto de presença (POP) próximo. Isso pode resultar em melhor performance para clientes geograficamente diferentes.
- Regional, o que significa que as solicitações são encaminhadas a um recurso dentro de uma determinada Região da AWS. Essa é uma boa opção quando todos os seus clientes estão perto da região em que sua API está implantada.

O API Gateway REST APIs também oferece suporte a endpoints de API privados, que podem ser acessados de uma nuvem privada virtual (VPC) usando uma interface VPC endpoint. Você também pode compartilhar com segurança o REST privado APIs criando endpoints de interface VPC em outros e até mesmo outros. VPCs Contas da AWS Para obter mais informações, consulte [Tipos de endpoint de API para REST APIs](#) na documentação do API Gateway.

Quando utilizar o API Gateway

O API Gateway é uma boa opção para serviços RESTful web e WebSocket conexões em tempo real. Ao usar WebSocket APIs no API Gateway, você pode adicionar comportamento para eventos de conexão e desconexão, como armazenar a conexão IDs em um armazenamento de dados externo associado aos atributos do cliente. Também pode encaminhar solicitações para comportamentos personalizados usando atributos de mensagem.

Tanto o REST quanto o WebSocket APIs podem se integrar diretamente a muitos Serviços da AWS sem exigir recursos computacionais separados, como funções Lambda. Isso pode melhorar a performance e reduzir os custos.

O REST APIs suporta roteamento baseado em caminhos e cabeçalhos, e você pode usá-los separadamente ou em conjunto. Um padrão comum é fornecer uma API REST como porta de entrada para várias, implementar preocupações compartilhadas APIs, conforme discutido anteriormente, e depois se comportar como um proxy reverso e rotear solicitações autorizadas para o endpoint correto da API.

Sistema de mensagens

Conforme discutido na seção [Padrões de comunicação](#), é possível usar mensagens para se comunicar de forma síncrona ou assíncrona entre serviços. Há muitos serviços sem servidor da AWS para escolher, e sua escolha deve ser baseada em suas necessidades de integração. Por exemplo, se você precisar de uma entrega ordenada de mensagens, deve escolher um serviço como o Amazon Simple Queue Service (Amazon SQS) ou o Amazon Simple Notification Service (Amazon SNS). Ambos os serviços oferecem suporte à entrega do tipo primeiro a entrar, primeiro a sair (FIFO), ao contrário do Amazon EventBridge, que não oferece esse recurso.

As seções a seguir discutem esses serviços em mais detalhes.

Tópicos

- [Amazon SQS](#)
- [Amazon SNS](#)
- [Amazon EventBridge](#)
- [AWS AppSync Events e API Gateway](#)

Amazon SQS

O [Amazon SQS](#) oferece suporte a filas padrão, que não garantem a ordem, e filas FIFO, que garantem a ordem dentro de um determinado grupo de mensagens.

As filas são um método comum para coreografar microsserviços e fornecem armazenamento durável para mensagens por até 14 dias. As filas são preenchidas por produtores e esvaziadas por consumidores. Ao usar o AWS Lambda como consumidor, é possível configurar uma fila do SQS como uma origem de eventos. Nesse caso, o mapeamento da origem do evento (ESM) do serviço Lambda pesquisa a fila para você e entrega as mensagens à sua função do Lambda quando elas ficam disponíveis. Os microsserviços executados em outros tipos de serviços de computação, como o Amazon Elastic Container Service (Amazon ECS) ou o Amazon Elastic Compute Cloud (Amazon EC2), precisam implementar seu próprio mecanismo de pesquisa para obter novas mensagens da fila, quando elas estiverem disponíveis.

O Lambda ESM para Amazon SQS também oferece suporte à filtragem de mensagens, o que permite processar apenas um subconjunto de mensagens em uma fila com base no conteúdo do corpo da mensagem.

Sondagem

O Amazon SQS oferece suporte a sondagens curtas e longas de mensagens. A pesquisa rápida consulta um subconjunto de servidores para encontrar mensagens disponíveis e as retorna imediatamente. No entanto, pode não retornar todas as mensagens disponíveis. Isso é útil quando sua aplicação precisa consumir mensagens o mais rápido possível ou não pode tolerar uma espera por um período mais longo.

A sondagem longa aguarda até que um período de tempo configurável tenha passado ou um número configurável de mensagens tenha sido recebido antes de retornar as mensagens. Isso pode reduzir o número de sondagens vazias, ou seja, o número de sondagens em que nenhuma mensagem é retornada, especialmente para filas que não recebem muitas mensagens. Reduzir o número de pesquisas vazias pode diminuir seus custos com o Amazon SQS, pois esse serviço cobra por cada solicitação, e cada operação de pesquisa é uma solicitação.

Orientação

As filas são uma boa opção quando:

- Você deseja desacoplar componentes e não precisa de comunicação síncrona entre eles.
- Você está se comunicando entre componentes que têm diferentes acordos de nível de serviço (SLAs) ou objetivos de nível de serviço (SLOs) de disponibilidade.
- Geralmente, você tem um único consumidor para um conjunto de mensagens.

Considere uma opção alternativa se:

- Você precisa de comunicação síncrona.
- Você precisa de uma lógica de roteamento complicada para enviar mensagens ao consumidor correto.

Amazon SNS

O [Amazon SNS](#) permite que você crie tópicos comuns e FIFO. Os tópicos são usados para implementar uma arquitetura de publicação/subscrição (pub/sub). O Amazon SNS oferece suporte a vários tipos de assinatura, incluindo e-mail, SMS (supondo que você tenha configurado uma identidade de origem, como um número gratuito ou um código longo de 10 dígitos), endpoint

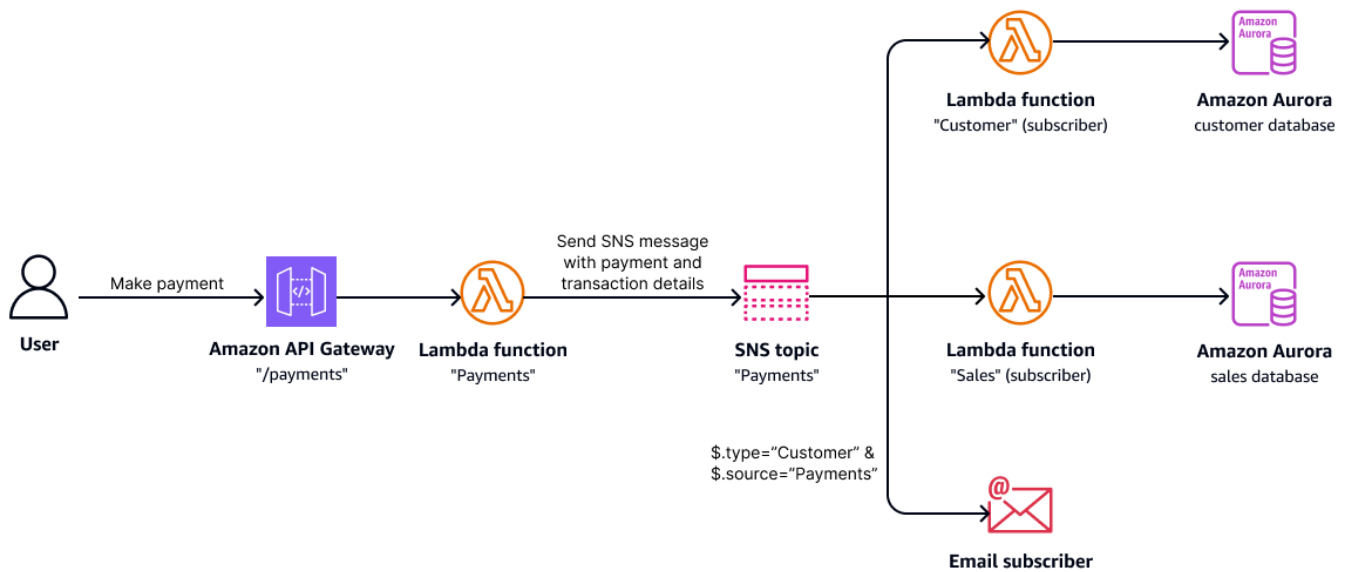
HTTP(S) e filas SQS. As assinaturas de usuários finais, como mensagens de e-mail e SMS, para um tópico do SNS devem ser confirmadas pelo assinante. O Amazon SNS permite que os serviços sejam amplamente disseminados, o que significa que uma única mensagem pode ser entregue a um número potencialmente grande de assinantes. Um tópico padrão do SNS tem um limite padrão de 12,5 milhões de assinaturas.

Em um ambiente de microsserviços, um tópico do SNS é útil para separar o roteamento de mensagens e a lógica de entrega de um editor. Isso pode ser implementado usando filtros de tópicos. Conceitualmente, os filtros de tópicos são semelhantes às regras do Amazon EventBridge, mas são configurados para cada assinante, em vez de estarem disponíveis em um local centralizado. Por exemplo, digamos que você tenha:

- Um serviço de pedidos, que processa pedidos.
- Um serviço de logística, que lida com a logística de pedidos.
- Um serviço de fidelidade, que concede pontos de fidelidade aos membros por pedidos.

Quando um pedido está pronto para ser atendido, ele publica uma mensagem em um tópico. O serviço de logística assina o tópico, mas não aplica um filtro, pois deseja saber sobre todos os pedidos. Imagine que você tem um serviço de fidelidade, responsável por conceder pontos aos membros quando eles fazem um pedido. No entanto, nem todos os pedidos são feitos por membros. O serviço de Fidelidade se inscreveria no tópico, mas implementaria um filtro de inscrição para verificar um atributo que indica se o pedido foi feito por um membro ou por um convidado.

Considere o caso em que um sistema recebe uma solicitação de um usuário final para efetuar um pagamento, conforme ilustrado no diagrama a seguir. Nesse caso, vários sistemas downstream precisam saber que a solicitação foi feita para que várias ações possam ser realizadas. Quando você usa o Amazon SNS, os pagamentos são publicados em um tópico do SNS, e as funções do Lambda se inscrevem no tópico para atualizar os bancos de dados de clientes e vendas. Além disso, uma inscrição de e-mail (que deve ser confirmada pelo cliente) envia um e-mail de confirmação ao cliente usando um filtro de inscrição.



Orientação

Algumas das funcionalidades descritas nesta seção para o Amazon SNS se sobrepõem às funcionalidades oferecidas por um barramento de eventos, como o EventBridge. Considere usar o Amazon SNS quando:

- Você terá um grande número de assinantes para um tópico.
- Você deseja usar tipos de inscrição (como e-mail ou SMS) que não são compatíveis nativamente com o EventBridge.
- Os assinantes devem poder determinar seu filtro de assinatura.
- É necessário solicitar a entrega ordenada aos assinantes (por grupo de mensagens).

Se você tiver muitos tópicos e assinaturas e filtros estiverem sendo usados para encaminhar mensagens entre microsserviços, o EventBridge provavelmente será a melhor opção.

Amazon EventBridge

O [Amazon EventBridge](#) é um serviço de barramento de eventos sem servidor e geralmente atua como base para uma arquitetura orientada a eventos (EDA). Você também pode usá-lo para rotear e entregar mensagens de forma assíncrona entre microsserviços. Usando o EventBridge, os produtores publicam eventos em um barramento. Você configura regras que correspondem a

eventos com base em seu conteúdo e escolhe um ou mais destinos para os quais os eventos que correspondem a essa regra serão entregues. O EventBridge oferece suporte a uma [ampla variedade de destinos de regras](#). O uso de um barramento de eventos permite separar os produtores dos consumidores e consolidar sua lógica de roteamento e entrega.

No EventBridge, você também pode criar regras programadas para realizar ações em momentos específicos. Você pode definir eventos usando expressões baseadas em cron e em taxas.

O [EventBridge Pipes](#) oferece a capacidade de canalizar mensagens de uma [origem](#) para um [destino](#) sem usar um serviço de computação, como o AWS Lambda. Por exemplo, digamos que você tenha uma fila do SQS que recebe mensagens, que devem servir como gatilho para uma máquina de estados do AWS Step Functions. Em vez de criar uma função do Lambda que tenha um mapeamento de fonte de evento para consumir mensagens da fila e escrever código usando um AWS SDK para invocar a máquina de estados, você pode usar o EventBridge Pipes para fazer isso sem escrever nenhum código personalizado.

O EventBridge é comumente usado com outros serviços de mensagens, como Amazon SQS e Amazon SNS. Por exemplo, entregar eventos a uma fila do SQS dá ao serviço receptor a flexibilidade de consumir mensagens quando for possível, a uma taxa independente da taxa na qual os eventos são produzidos. Da mesma forma, você pode enviar eventos que devem ser divulgados para um grande número de assinantes de um tópico do SNS.

Orientação

Use o EventBridge quando:

- Não é necessária comunicação síncrona entre os serviços.
- Você deseja separar a lógica de roteamento de mensagens dos seus microsserviços. Os microsserviços apenas produzem eventos e os publicam no barramento de eventos, e os serviços interessados criam regras para corresponder e entregar esses eventos.
- Você precisa entregar mensagens de um serviço compatível para outro.

Considere outros serviços em situações em que:

- É necessária uma ordem rígida dos eventos. Nesses casos, considere as filas FIFO do Amazon SQS ou os tópicos FIFO do Amazon SNS. Como alternativa, considere serviços de fluxo de eventos, como Amazon Kinesis Data Streams ou Amazon Managed Streaming for Apache Kafka (Amazon MSK).

AWS AppSync Events e API Gateway

O AWS AppSync Events e o Amazon API Gateway oferecem ambos uma experiência WebSocket gerenciada para os seus microsserviços.

O [AWS AppSync Events](#) oferece uma experiência simplificada para mensagens em tempo real usando um WebSocket. O AWS AppSync Events oferece suporte a mensagens unicast e multicast e um agrupamento flexível de canais em namespaces, com suporte para caracteres curinga. Os microsserviços podem se comunicar entre si de várias maneiras, usando o AWS AppSync Events. Por exemplo, um serviço que recebe dados em tempo real pode transformar e publicar os dados no canal apropriado, onde os assinantes os receberão em tempo real.

O [API Gateway](#) também é compatível com APIs WebSocket. Você pode definir integrações com Serviços da AWS, como o AWS Lambda e o Amazon DynamoDB, e configurar expressões de seleção de rotas que são mapeadas para essas integrações. O API Gateway possui rotas especiais que você pode usar para autorizar e gerenciar suas conexões WebSocket. Dependendo das suas necessidades, você pode armazenar informações de conexão WebSocket em um armazenamento de dados como o DynamoDB. Usando essas informações, as mensagens podem ser publicadas em conexões WebSocket específicas por meio de uma API REST, considerando um ID de conexão específico.

Orientação

Use o AWS AppSync Events quando:

- Você tem vários canais de mensagens agrupados em namespaces e deseja publicar e se inscrever em grupos de canais usando curingas.
- Suas comunicações ocorrem principalmente entre diferentes sistemas, em vez de entre Serviços da AWS.

Use APIs de WebSocket do API Gateway quando:

- Você deseja permitir que os clientes tenham conexões persistentes em tempo real com integrações de AWS service (Serviço da AWS).
- Você mesmo deseja gerenciar conexões WebSocket. Por exemplo, você pode querer permitir que outros sistemas enviem mensagens para um cliente específico após consultar seu ID de conexão.
- Você deseja usar recursos do API Gateway, como implantações em estágios ou integrações de proxy, ou deseja configurar seus próprios subprotocolos.

Perguntas frequentes

Como posso combinar diferentes padrões de integração?

Na maioria das situações, convém combinar padrões de integração. Por exemplo, você pode usar o AWS Step Functions para orquestrar um processo que chama um serviço remoto usando o padrão de verificação de declaração. Ou pode ter um processo orquestrado que coloca mensagens em filas, o que, por sua vez, aciona serviços coreografados.

Qual é o principal benefício de usar uma arquitetura de microsserviços?

As principais vantagens incluem escalabilidade independente de serviços, melhor isolamento de falhas, maior velocidade de desenvolvimento por meio do trabalho em equipe paralelo e a capacidade de entrega e implantação contínuas (CI/CD).

Como posso implementar o tratamento de erros nesses padrões?

Você pode implementar o tratamento de erros usando mecanismos integrados em Serviços da AWS. Por exemplo, AWS Lambda as funções podem ser configuradas com a lógica de repetição, e o Amazon SQS suporta filas de mensagens mortas para lidar com falhas persistentes. Além disso, o Step Functions fornece mecanismos de tratamento e repetição de erros em nível de fluxo de trabalho.

Quais são os benefícios do uso do padrão de verificação de declaração na comunicação assíncrona?

O padrão de verificação de declaração permite que os clientes recebam um identificador no momento do envio da solicitação. Esse identificador pode ser usado mais tarde para verificar o status e recuperar o resultado. Esse padrão beneficia os clientes por fornecer um mecanismo para pesquisar resultados sem esperar de maneira síncrona. Para obter mais informações, consulte a seção [Verificação de reivindicações](#) anterior deste guia.

Como o padrão de retorno de chamada melhora a comunicação assíncrona em microsserviços?

O padrão de retorno de chamada melhora a comunicação assíncrona, permitindo que o cliente forneça um local para o serviço entrar em contato após a conclusão do processamento. Isso dissocia o cliente da espera por uma resposta e permite que ele continue com outras tarefas. Para obter mais informações, consulte a seção [Retorno de chamada](#) anterior deste guia.

Posso implementar a comunicação bidirecional em microsserviços usando os padrões descritos?

Você pode implementar a comunicação bidirecional criando uma conexão estável entre um cliente e um serviço, para que eles possam enviar e processar mensagens de maneira assíncrona. Isso exige que o serviço dê suporte a uma conexão aberta para cada cliente. Para obter mais informações, consulte a seção [Comunicação bidirecional](#) anterior deste guia.

Como posso otimizar o uso de funções do Lambda em padrões de comunicação assíncrona?

É possível otimizar funções do Lambda garantindo que elas sejam idempotentes para lidar com possíveis duplicações de mensagens, usando recursos do Amazon SQS, como grupos de mensagens para ordenação, e implementando pesquisas longas para reduzir os custos. Além disso, você pode monitorar métricas de execução para identificar oportunidades de otimização.

Quais são as principais diferenças entre usar o Amazon SNS e EventBridge o pub/sub padrão?

O Amazon SNS envia uma única mensagem a todos os assinantes, o que pode incluir dados desnecessários para alguns assinantes. A Amazon EventBridge permite um controle mais granular, permitindo que você tenha várias regras que correspondam a um único evento, com cada regra acionando um serviço ou ação downstream diferente. Para obter mais informações, consulte o [Amazon SNS](#) e [EventBridge](#) as seções anteriores deste guia.

Recursos

Documentação da AWS service (Serviço da AWS)

- [Amazon API Gateway](#)
- [AWS AppSync Eventos do](#)
- [Amazon EventBridge](#)
- [Amazon MWAA](#)
- [Amazon SNS](#)
- [Amazon SQS](#)
- [AWS Step Functions](#)

Leitura adicional

- [Estratégia para modernizar aplicações na Nuvem AWS](#)
- [Abordagem em fases para modernizar aplicações na Nuvem AWS](#)
- [Avaliação da prontidão da modernização para aplicações na Nuvem AWS](#)
- [Decompor monólitos em microsserviços](#)
- [Como implementar padrões de integração corporativa com serviços AWS de mensagens: canais ponto a ponto](#)
- [Mensagens Pub/Sub: notificações assíncronas de eventos](#)

Histórico do documento

A tabela a seguir descreve alterações significativas feitas neste guia. Se desejar receber notificações sobre futuras atualizações, inscreva-se em um [feed RSS](#).

Alteração	Descrição	Data
Atualizações principais	Guia expandido, reorganizado e atualizado para refletir atualizações em AWS service (Serviço da AWS).	10 de setembro de 2025
Publicação inicial	—	11 de janeiro de 2021

AWS Glossário de orientação prescritiva

A seguir estão os termos comumente usados em estratégias, guias e padrões fornecidos pela Orientação AWS Prescritiva. Para sugerir entradas, use o link Fornecer feedback no final do glossário.

Números

7 Rs

Sete estratégias comuns de migração para mover aplicações para a nuvem. Essas estratégias baseiam-se nos 5 Rs identificados pela Gartner em 2011 e consistem em:

- **Refactor/re-architect** — mova um aplicativo e modifique sua arquitetura aproveitando ao máximo os recursos nativos da nuvem para melhorar a agilidade, o desempenho e a escalabilidade. Isso normalmente envolve a portabilidade do sistema operacional e do banco de dados. Exemplo: migre seu banco de dados Oracle local para a Amazon PostgreSQL-Compatible Aurora Edition.
- **Redefinir a plataforma (mover e redefinir [mover e redefinir (lift-and-reshape)]):** mova uma aplicação para a nuvem e introduza algum nível de otimização a fim de aproveitar os recursos da nuvem. Exemplo: migrar seu banco de dados Oracle on-premises para o Amazon Relational Database Service (Amazon RDS) para Oracle na Nuvem AWS.
- **Recomprar (drop and shop):** mude para um produto diferente, normalmente migrando de uma licença tradicional para um modelo SaaS. Exemplo: Migre seu sistema de gerenciamento de relacionamento com o cliente (CRM) para o Salesforce.com
- **Redefinir a hospedagem (mover sem alterações [lift-and-shift]):** mover uma aplicação para a nuvem sem fazer nenhuma alteração a fim de aproveitar os recursos da nuvem. Exemplo: migrar seu banco de dados Oracle on-premises para o Oracle em uma instância do EC2 na Nuvem AWS.
- **Realocar (mover o hipervisor sem alterações [hypervisor-level lift-and-shift]):** mover a infraestrutura para a nuvem sem comprar novo hardware, reescrever aplicações ou modificar suas operações existentes. Você migra servidores de uma plataforma on-premises para um serviço de nuvem para a mesma plataforma. Exemplo: Migrar um Microsoft Hyper-V aplicativo para o AWS
- **Reter (revisitar):** mantenha as aplicações em seu ambiente de origem. Isso pode incluir aplicações que exigem grande refatoração, e você deseja adiar esse trabalho para um

momento posterior, e aplicações antigas que você deseja manter porque não há justificativa comercial para migrá-las.

- Retirar: desative ou remova aplicações que não são mais necessárias em seu ambiente de origem.

A

A2A () Agent-to-Agent

Um protocolo com estado para colaboração entre agentes, apoiando a delegação de tarefas e a transferência de estados.

ABAC

Consulte [controle de acesso baseado em atributo](#).

serviços abstraídos

Veja [serviços gerenciados](#).

ACID

Veja [atomicidade, consistência, isolamento, durabilidade](#).

migração ativa-ativa

Um método de migração de banco de dados no qual os bancos de dados de origem e de destino são mantidos em sincronia (por meio de uma ferramenta de replicação bidirecional ou operações de gravação dupla), e ambos os bancos de dados lidam com transações de aplicações conectadas durante a migração. Esse método oferece suporte à migração em lotes pequenos e controlados, em vez de exigir uma substituição única. É mais flexível, mas exige mais trabalho do que a [migração ativa-passiva](#).

migração ativa-passiva

Um método de migração de banco de dados em que os bancos de dados de origem e de destino são mantidos em sincronia, mas somente o banco de dados de origem manipula as transações das aplicações conectadas, enquanto os dados são replicados no banco de dados de destino. O banco de dados de destino não aceita nenhuma transação durante a migração.

Agente

Um sistema de IA que pode raciocinar, planejar e realizar ações de forma autônoma usando ferramentas para atingir metas.

Agente Ops

Práticas operacionais para criar, testar, implantar e executar agentes de IA na produção em grande escala.

AGGREGATE FUNCTION

Uma função SQL que opera em um grupo de linhas e calcula um único valor de retorno para o grupo. Exemplos de funções agregadas incluem SUM e MAX.

AI

Veja [inteligência artificial](#).

AIOps

Veja [operações de inteligência artificial](#).

anonimização

O processo de excluir permanentemente informações pessoais em um conjunto de dados. A anonimização pode ajudar a proteger a privacidade pessoal. Dados anônimos não são mais considerados dados pessoais.

antipadrões

Uma solução frequentemente usada para um problema recorrente em que a solução é contraproducente, ineficaz ou menos eficaz do que uma alternativa.

controle de aplicações

Uma abordagem de segurança que permite o uso somente de aplicações aprovadas para ajudar a proteger um sistema contra malware.

portfólio de aplicações

Uma coleção de informações detalhadas sobre cada aplicação usada por uma organização, incluindo o custo para criar e manter a aplicação e seu valor comercial. Essas informações são fundamentais para [o processo de descoberta e análise de portfólio](#) e ajudam a identificar e priorizar as aplicações a serem migradas, modernizadas e otimizadas.

inteligência artificial (IA)

O campo da ciência da computação que se dedica ao uso de tecnologias de computação para desempenhar funções cognitivas normalmente associadas aos humanos, como aprender, resolver problemas e reconhecer padrões. Para obter mais informações, consulte [O que é inteligência artificial?](#)

operações de inteligência artificial (AIOps)

O processo de usar técnicas de machine learning para resolver problemas operacionais, reduzir incidentes operacionais e intervenção humana e aumentar a qualidade do serviço. Para obter mais informações sobre como as AIOps são usadas na estratégia de migração para a AWS , consulte o [guia de integração de operações](#).

criptografia assimétrica

Um algoritmo de criptografia que usa um par de chaves, uma chave pública para criptografia e uma chave privada para descryptografia. É possível compartilhar a chave pública porque ela não é usada na descryptografia, mas o acesso à chave privada deve ser altamente restrito.

atomicidade, consistência, isolamento, durabilidade (ACID)

Um conjunto de propriedades de software que garantem a validade dos dados e a confiabilidade operacional de um banco de dados, mesmo no caso de erros, falhas de energia ou outros problemas.

controle de acesso por atributo (ABAC)

A prática de criar permissões minuciosas com base nos atributos do usuário, como departamento, cargo e nome da equipe. Para obter mais informações, consulte [ABAC AWS](#) na documentação AWS Identity and Access Management (IAM).

fonte de dados autorizada

Um local onde você armazena a versão principal dos dados, que é considerada a fonte de informações mais confiável. Você pode copiar dados da fonte de dados autorizada para outros locais com o objetivo de processar ou modificar os dados, como anonimizá-los, redigi-los ou pseudonimizá-los.

Zona de disponibilidade

Um local distinto dentro de um Região da AWS que está isolado de falhas em outras zonas de disponibilidade e fornece conectividade de rede barata e de baixa latência a outras zonas de disponibilidade na mesma região.

AWS Estrutura de adoção da nuvem (AWS CAF)

Uma estrutura de diretrizes e melhores práticas AWS para ajudar as organizações a desenvolver um plano eficiente e eficaz para migrar com sucesso para a nuvem. AWS O CAF organiza a orientação em seis áreas de foco chamadas perspectivas: negócios, pessoas, governança, plataforma, segurança e operações. As perspectivas de negócios, pessoas e governança têm

como foco habilidades e processos de negócios; as perspectivas de plataforma, segurança e operações concentram-se em habilidades e processos técnicos. Por exemplo, a perspectiva das pessoas tem como alvo as partes interessadas que lidam com recursos humanos (RH), funções de pessoal e gerenciamento de pessoal. Nessa perspectiva, o AWS CAF fornece orientação para desenvolvimento, treinamento e comunicação de pessoas para ajudar a preparar a organização para a adoção bem-sucedida da nuvem. Para obter mais informações, consulte o [site da AWS CAF](#) e o [whitepaper da AWS CAF](#).

AWS Estrutura de qualificação da carga de trabalho (AWS WQF)

Uma ferramenta que avalia as cargas de trabalho de migração do banco de dados, recomenda estratégias de migração e fornece estimativas de trabalho. AWS O WQF está incluído com AWS Schema Conversion Tool (AWS SCT). Ela analisa esquemas de banco de dados e objetos de código, código de aplicações, dependências e características de performance, além de fornecer relatórios de avaliação.

B

bot malicioso

Um [bot](#) destinado a causar interrupção ou danos a indivíduos ou organizações.

BCP

Veja [planejamento de continuidade de negócios](#)

gráfico de comportamento

Uma visualização unificada e interativa do comportamento e das interações de recursos ao longo do tempo. É possível usar um gráfico de comportamento com o Amazon Detective para examinar tentativas de login malsucedidas, chamadas de API suspeitas e ações similares. Para obter mais informações, consulte [Dados em um gráfico de comportamento](#) na documentação do Detective.

sistema big-endian

Um sistema que armazena o byte mais significativo antes. Veja também [endianness](#).

classificação binária

Um processo que prevê um resultado binário (uma de duas classes possíveis). Por exemplo, seu modelo de ML pode precisar prever problemas como “Este e-mail é ou não é spam?” ou “Este produto é um livro ou um carro?”

filtro de bloom

Uma estrutura de dados probabilística e eficiente em termos de memória que é usada para testar se um elemento é membro de um conjunto.

blue/green implantação

Uma estratégia de implantação em que você cria dois ambientes separados, mas idênticos. Você executa a versão atual da aplicação em um ambiente (azul) e a nova versão da aplicação no outro ambiente (verde). Essa estratégia ajuda você a reverter rapidamente com o mínimo de impacto.

bot

Uma aplicação de software que executa tarefas automatizadas na internet e simula a atividade ou interação humana. Alguns bots são úteis ou benéficos, como crawlers da web que indexam informações na internet. Outros bots, conhecidos como bots maliciosos, têm como objetivo causar interrupção ou danos a indivíduos ou organizações.

botnet

Redes de [bots](#) infectadas por [malware](#) e sob o controle de uma única parte, conhecidas como bot herder ou operador de bots. Os botnets são o mecanismo mais conhecido para escalar bots e seu impacto.

ramo

Uma área contida de um repositório de código. A primeira ramificação criada em um repositório é a ramificação principal. Você pode criar uma nova ramificação a partir de uma ramificação existente e, em seguida, desenvolver recursos ou corrigir bugs na nova ramificação. Uma ramificação que você cria para gerar um recurso é comumente chamada de ramificação de recurso. Quando o recurso estiver pronto para lançamento, você mesclará a ramificação do recurso de volta com a ramificação principal. Para obter mais informações, consulte [Sobre filiais](#) (GitHub documentação).

Acesso de emergência

Em circunstâncias excepcionais e por meio de um processo aprovado, um meio rápido para um usuário obter acesso a um Conta da AWS que ele normalmente não tem permissão para acessar. Para obter mais informações, consulte o indicador [Implementar procedimentos de quebra de vidros](#) na AWS Well-Architected orientação.

estratégia brownfield

A infraestrutura existente em seu ambiente. Ao adotar uma estratégia brownfield para uma arquitetura de sistema, você desenvolve a arquitetura de acordo com as restrições dos sistemas e da infraestrutura atuais. Se estiver expandindo a infraestrutura existente, poderá combinar as estratégias brownfield e [greenfield](#).

cache do buffer

A área da memória em que os dados acessados com mais frequência são armazenados.

capacidade de negócios

O que uma empresa faz para gerar valor (por exemplo, vendas, atendimento ao cliente ou marketing). As arquiteturas de microsserviços e as decisões de desenvolvimento podem ser orientadas por recursos de negócios. Para obter mais informações, consulte a seção [Organizados de acordo com as capacidades de negócios](#) do whitepaper [Executar microsserviços containerizados na AWS](#).

planejamento de continuidade de negócios (BCP)

Um plano que aborda o impacto potencial de um evento disruptivo, como uma migração em grande escala, nas operações e permite que uma empresa retome as operações rapidamente.

C

CAF

Veja [AWS Cloud Adoption Framework](#).

implantação canário

O lançamento lento e incremental de uma versão para usuários finais. Quando estiver confiante, você implanta a nova versão e substitui a versão atual por completo.

CCoE

Veja [Centro de Excelência da Nuvem](#).

CDC

Veja [captura de dados de alteração](#).

captura de dados de alterações (CDC)

O processo de rastrear alterações em uma fonte de dados, como uma tabela de banco de dados, e registrar metadados sobre a alteração. É possível usar o CDC para várias finalidades, como auditar ou replicar alterações em um sistema de destino para manter a sincronização.

engenharia do caos

Introduzir intencionalmente falhas ou eventos disruptivos para testar a resiliência de um sistema. Você pode usar [AWS Fault Injection Service \(AWS FIS\)](#) para realizar experimentos que stressam suas AWS cargas de trabalho e avaliar sua resposta.

CI/CD

Veja [integração e entrega contínuas](#).

classificação

Um processo de categorização que ajuda a gerar previsões. Os modelos de ML para problemas de classificação predizem um valor discreto. Os valores discretos são sempre diferentes uns dos outros. Por exemplo, um modelo pode precisar avaliar se há ou não um carro em uma imagem.

Desenvolvedor cidadão

Um usuário corporativo que cria aplicativos de IA usando plataformas sem code/low código sem habilidades técnicas especializadas.

criptografia no lado do cliente

Criptografia de dados localmente, antes que o alvo os AWS service (Serviço da AWS) receba.

Centro de Excelência da Nuvem (CCoE)

Uma equipe multidisciplinar que impulsiona os esforços de adoção da nuvem em toda a organização, incluindo o desenvolvimento de práticas recomendadas de nuvem, a mobilização de recursos, o estabelecimento de cronogramas de migração e a liderança da organização em transformações em grande escala. Para obter mais informações, consulte as [postagens do CCoE no blog](#) de estratégia Nuvem AWS corporativa.

computação em nuvem

A tecnologia de nuvem normalmente usada para armazenamento de dados remoto e gerenciamento de dispositivos de IoT. A computação em nuvem é normalmente conectada à tecnologia de [computação de borda](#).

modelo operacional em nuvem

Em uma organização de TI, o modelo operacional usado para criar, amadurecer e otimizar um ou mais ambientes de nuvem. Para obter mais informações, consulte [Criar seu modelo operacional de nuvem](#).

estágios de adoção da nuvem

As quatro fases pelas quais as organizações normalmente passam ao migrar para a Nuvem AWS:

- Projeto: executar alguns projetos relacionados à nuvem para fins de prova de conceito e aprendizado
- Fundação: realizar investimentos fundamentais para escalar sua adoção da nuvem (por exemplo, criar uma zona de pouso, definir um CCoE, estabelecer um modelo de operações)
- Migração: migrar aplicações individuais
- Re-invention — Otimizando produtos e serviços e inovando na nuvem

Esses estágios foram definidos por Stephen Orban na postagem do blog The [Journey Toward Cloud-First & the Stages of Adoption](#) no blog Nuvem AWS Enterprise Strategy. Para obter informações sobre como eles se relacionam com a estratégia de AWS migração, consulte o [guia de preparação para migração](#).

CMDB

Veja [banco de dados de gerenciamento de configuração](#).

repositório de código

Um local onde o código-fonte e outros ativos, como documentação, amostras e scripts, são armazenados e atualizados por meio de processos de controle de versão. Os repositórios de nuvem comuns incluem o GitHub ou o Bitbucket Cloud. Cada versão do código é chamada de ramificação. Em uma estrutura de microsserviços, cada repositório é dedicado a uma única peça de funcionalidade. Um único CI/CD pipeline pode usar vários repositórios.

cache frio

Um cache de buffer que está vazio, não está bem preenchido ou contém dados obsoletos ou irrelevantes. Isso afeta a performance porque a instância do banco de dados deve ler da memória principal ou do disco, um processo que é mais lento do que a leitura do cache do buffer.

dados frios

Dados que raramente são acessados e geralmente são históricos. Ao consultar esse tipo de dados, consultas lentas geralmente são aceitáveis. Mover esses dados para níveis ou classes de armazenamento de baixo desempenho e menos caros pode reduzir os custos.

visão computacional (CV)

Um campo de [IA](#) que usa machine learning para analisar e extrair informações de formatos visuais, como vídeos e imagens digitais. Por exemplo, a Amazon SageMaker AI fornece algoritmos de processamento de imagem para CV.

desvio de configuração

Em uma workload, uma alteração de configuração em relação ao estado esperado. Isso pode fazer com que a workload se torne incompatível e, normalmente, é gradual e não intencional.

banco de dados de gerenciamento de configuração (CMDB)

Um repositório que armazena e gerencia informações sobre um banco de dados e seu ambiente de TI, incluindo componentes de hardware e software e suas configurações. Normalmente, os dados de um CMDB são usados no estágio de descoberta e análise do portfólio da migração.

pacote de conformidade

Uma coleção de AWS Config regras e ações de remediação que você pode montar para personalizar suas verificações de conformidade e segurança. Você pode implantar um pacote de conformidade como uma entidade única em uma Conta da AWS região ou em uma organização usando um modelo YAML. Para obter mais informações, consulte [Pacotes de conformidade na documentação](#). AWS Config

integração contínua e entrega contínua (CI/CD)

O processo de automatizar os estágios de origem, criação, teste, preparação e produção do processo de lançamento do software. CI/CD é comumente descrito como um pipeline. CI/CD pode ajudá-lo a automatizar processos, melhorar a produtividade, melhorar a qualidade do código e entregar com mais rapidez. Para obter mais informações, consulte [Benefícios da entrega contínua](#). CD também pode significar implantação contínua. Para obter mais informações, consulte [Entrega contínua versus implantação contínua](#).

CV

Veja [visão computacional](#).

D

dados em repouso

Dados estacionários em sua rede, por exemplo, dados que estão em um armazenamento.

classificação de dados

Um processo para identificar e categorizar os dados em sua rede com base em criticalidade e confidencialidade. É um componente crítico de qualquer estratégia de gerenciamento de riscos de segurança cibernética, pois ajuda a determinar os controles adequados de proteção e retenção para os dados. A classificação de dados é um componente do pilar de segurança na AWS Well-Architected Estrutura. Para obter mais informações, consulte [Classificação de dados](#).

desvio de dados

Uma variação significativa entre os dados de produção e os dados usados para treinar um modelo de ML ou uma alteração significativa nos dados de entrada ao longo do tempo. O desvio de dados pode reduzir a qualidade geral, a precisão e a imparcialidade das previsões do modelo de ML.

dados em trânsito

Dados que estão se movendo ativamente pela sua rede, como entre os recursos da rede.

data mesh

Um framework de arquitetura que fornece propriedade de dados distribuída e descentralizada com gerenciamento e governança centralizados.

minimização de dados

O princípio de coletar e processar apenas os dados estritamente necessários. Praticar a minimização de dados no Nuvem AWS pode reduzir os riscos de privacidade, os custos e a pegada de carbono de sua análise.

perímetro de dados

Um conjunto de proteções preventivas em seu AWS ambiente que ajudam a garantir que somente identidades confiáveis acessem recursos confiáveis das redes esperadas. Para obter mais informações, consulte [Construindo um perímetro de dados em AWS](#)

pré-processamento de dados

A transformação de dados brutos em um formato que seja facilmente analisado por seu modelo de ML. O pré-processamento de dados pode significar a remoção de determinadas colunas ou linhas e o tratamento de valores ausentes, inconsistentes ou duplicados.

proveniência dos dados

O processo de rastrear a origem e o histórico dos dados ao longo de seu ciclo de vida, por exemplo, como os dados foram gerados, transmitidos e armazenados.

titular dos dados

Um indivíduo cujos dados estão sendo coletados e processados.

data warehouse

Um sistema de gerenciamento de dados compatível com business intelligence, como analytics. Os data warehouses geralmente contêm grandes quantidades de dados históricos e geralmente são usados para consultas e análises.

linguagem de definição de dados (DDL)

Instruções ou comandos para criar ou modificar a estrutura de tabelas e objetos em um banco de dados.

linguagem de manipulação de dados (DML)

Instruções ou comandos para modificar (inserir, atualizar e excluir) informações em um banco de dados.

DDL

Veja [linguagem de definição de banco de dados](#).

deep ensemble

A combinação de vários modelos de aprendizado profundo para gerar previsões. Os deep ensembles podem ser usados para produzir uma previsão mais precisa ou para estimar a incerteza nas previsões.

Aprendizado profundo

Um subcampo do ML que usa várias camadas de redes neurais artificiais para identificar o mapeamento entre os dados de entrada e as variáveis-alvo de interesse.

defesa completa

Uma abordagem de segurança da informação na qual uma série de mecanismos e controles de segurança são cuidadosamente distribuídos por toda a rede de computadores para proteger a confidencialidade, a integridade e a disponibilidade da rede e dos dados nela contidos. Ao adotar essa estratégia AWS, você adiciona vários controles em diferentes camadas da AWS Organizations estrutura para ajudar a proteger os recursos. Por exemplo, uma abordagem de defesa aprofundada pode combinar autenticação multifatorial, segmentação de rede e criptografia.

administrador delegado

Em AWS Organizations, um serviço compatível pode registrar uma conta de AWS membro para administrar as contas da organização e gerenciar as permissões desse serviço. Essa conta é chamada de administrador delegado para esse serviço. Para obter mais informações e uma lista de serviços compatíveis, consulte [Serviços que funcionam com o AWS Organizations](#) na documentação do AWS Organizations .

implantação

O processo de criar uma aplicação, novos recursos ou correções de código disponíveis no ambiente de destino. A implantação envolve a implementação de mudanças em uma base de código e, em seguida, a criação e execução dessa base de código nos ambientes da aplicação

ambiente de desenvolvimento

Veja [ambiente](#).

controle detectivo

Um controle de segurança projetado para detectar, registrar e alertar após a ocorrência de um evento. Esses controles são uma segunda linha de defesa, alertando você sobre eventos de segurança que contornaram os controles preventivos em vigor. Para obter mais informações, consulte [Controles detectivos](#) em Como implementar controles de segurança na AWS.

mapeamento do fluxo de valor de desenvolvimento (DVSM)

Um processo usado para identificar e priorizar restrições que afetam negativamente a velocidade e a qualidade em um ciclo de vida de desenvolvimento de software. O DVSM estende o processo de mapeamento do fluxo de valor originalmente projetado para práticas de manufatura enxuta. Ele se concentra nas etapas e equipes necessárias para criar e movimentar valor por meio do processo de desenvolvimento de software.

gêmeo digital

Uma representação virtual de um sistema real, como um prédio, fábrica, equipamento industrial ou linha de produção. Os gêmeos digitais oferecem suporte à manutenção preditiva, ao monitoramento remoto e à otimização da produção.

tabela de dimensões

Em um [esquema em estrela](#), uma tabela menor que contém atributos de dados sobre dados quantitativos em uma tabela de fatos. Os atributos da tabela de dimensões geralmente são campos de texto ou números discretos que se comportam como texto. Esses atributos normalmente são usados para restringir consultas, filtrar e rotular conjuntos de resultados.

desastre

Um evento que impede que uma workload ou sistema cumpra seus objetivos de negócios em seu local principal de implantação. Esses eventos podem ser desastres naturais, falhas técnicas ou o resultado de ações humanas, como configuração incorreta não intencional ou ataque de malware.

Recuperação de desastres (RD)

A estratégia e o processo que você usa para minimizar o tempo de inatividade e a perda de dados causados por um [desastre](#). Para obter mais informações, consulte [Recuperação de desastres de cargas de trabalho em AWS: Recuperação na nuvem](#) na AWS Well-Architected estrutura.

DML

Veja [linguagem de manipulação de banco de dados](#).

design orientado por domínio

Uma abordagem ao desenvolvimento de um sistema de software complexo conectando seus componentes aos domínios em evolução, ou principais metas de negócios, atendidos por cada componente. Esse conceito foi introduzido por Eric Evans em seu livro Domain-Driven Design: Tackling Complexity in the Heart of Software (Boston: Addison-Wesley Professional, 2003). Para obter informações sobre como você pode usar o design orientado por domínio com o padrão strangler fig, consulte Modernizando os [serviços web legados da Microsoft ASP.NET \(ASMX\) de forma incremental usando](#) contêineres e o Amazon API Gateway.

DR

Veja [recuperação de desastres](#).

Detecção da oscilação

Rastreamento de desvios de uma configuração de linha de base. Por exemplo, você pode usar AWS CloudFormation para [detectar desvios nos recursos do sistema](#) ou AWS Control Tower para [detectar mudanças em seu landing zone](#) que possam afetar a conformidade com os requisitos de governança.

DVSM

Veja [mapeamento do fluxo de valor de desenvolvimento](#).

E

EDA

Veja [análise exploratória de dados](#).

EDI

Veja [intercâmbio eletrônico de dados](#).

computação de borda

A tecnologia que aumenta o poder computacional de dispositivos inteligentes nas bordas de uma rede de IoT. Quando comparada com a [computação em nuvem](#), a computação de borda pode reduzir a latência da comunicação e melhorar o tempo de resposta.

intercâmbio eletrônico de dados (EDI)

A troca automatizada de documentos comerciais entre organizações. Para obter mais informações, consulte [O que é EDI \(Intercâmbio eletrônico de dados\)?](#).

criptografia

Um processo de computação que transforma dados de texto simples, legíveis por humanos, em texto cifrado.

chave de criptografia

Uma sequência criptográfica de bits aleatórios que é gerada por um algoritmo de criptografia. As chaves podem variar em tamanho, e cada chave foi projetada para ser imprevisível e exclusiva.

endianismo

A ordem na qual os bytes são armazenados na memória do computador. Big-endian os sistemas armazenam primeiro o byte mais significativo. Little-endian os sistemas armazenam primeiro o byte menos significativo.

endpoint

Veja [endpoint de serviço](#).

serviço de endpoint

Um serviço que pode ser hospedado em uma nuvem privada virtual (VPC) para ser compartilhado com outros usuários. Você pode criar um serviço de endpoint com AWS PrivateLink e conceder permissões a outros diretores Contas da AWS ou a AWS Identity and Access Management (IAM). Essas contas ou entidades principais podem se conectar ao serviço de endpoint de maneira privada criando endpoints da VPC de interface. Para obter mais informações, consulte [Criar um serviço de endpoint](#) na documentação do Amazon Virtual Private Cloud (Amazon VPC).

planejamento de recursos empresariais (ERP)

Um sistema que automatiza e gerencia os principais processos de negócios (como contabilidade, [MES](#) e gerenciamento de projetos) para uma empresa.

criptografia envelopada

O processo de criptografar uma chave de criptografia com outra chave de criptografia. Para obter mais informações, consulte [Criptografia de envelope](#) na documentação AWS Key Management Service (AWS KMS).

ambiente

Uma instância de uma aplicação em execução. Estes são tipos comuns de ambientes na computação em nuvem:

- ambiente de desenvolvimento: uma instância de uma aplicação em execução que está disponível somente para a equipe principal responsável pela manutenção da aplicação. Ambientes de desenvolvimento são usados para testar mudanças antes de promovê-las para ambientes superiores. Esse tipo de ambiente às vezes é chamado de ambiente de teste.
- ambientes inferiores: todos os ambientes de desenvolvimento para uma aplicação, como aqueles usados para compilações e testes iniciais.
- ambiente de produção: uma instância de uma aplicação em execução que os usuários finais podem acessar. Em um CI/CD pipeline, o ambiente de produção é o último ambiente de implantação.

- ambientes superiores: todos os ambientes que podem ser acessados por usuários que não sejam a equipe principal de desenvolvimento. Isso pode incluir um ambiente de produção, ambientes de pré-produção e ambientes para testes de aceitação do usuário.

epic

Em metodologias ágeis, categorias funcionais que ajudam a organizar e priorizar seu trabalho. Os epics fornecem uma descrição de alto nível dos requisitos e das tarefas de implementação. Por exemplo, os épicos de segurança AWS da CAF incluem gerenciamento de identidade e acesso, controles de detetive, segurança de infraestrutura, proteção de dados e resposta a incidentes. Para obter mais informações sobre epics na estratégia de migração da AWS, consulte o [guia de implementação do programa](#).

ERP

Veja [planejamento de recursos empresariais](#).

análise exploratória de dados (EDA)

O processo de analisar um conjunto de dados para entender suas principais características. Você coleta ou agrega dados e, em seguida, realiza investigações iniciais para encontrar padrões, detectar anomalias e verificar suposições. O EDA é realizado por meio do cálculo de estatísticas resumidas e da criação de visualizações de dados.

F

tabela de fatos

A tabela central em um [esquema em estrela](#). Ela armazena dados quantitativos sobre as operações comerciais. Normalmente, uma tabela de fatos contém dois tipos de colunas: as que contêm medidas e as que contêm uma chave externa para uma tabela de dimensões.

Antecipar-se à falha

Uma filosofia que usa testes frequentes e incrementais para reduzir o ciclo de vida do desenvolvimento. É uma parte essencial de uma abordagem ágil.

delimitação de isolamento contra falhas

No Nuvem AWS, um limite, como uma zona de disponibilidade, Região da AWS um plano de controle ou um plano de dados, que limita o efeito de uma falha e ajuda a melhorar a resiliência das cargas de trabalho. Para obter mais informações, consulte [AWS Fault Isolation Boundaries](#).

ramificação de recursos

Veja [ramificação](#).

recursos

Os dados de entrada usados para fazer uma previsão. Por exemplo, em um contexto de manufatura, os recursos podem ser imagens capturadas periodicamente na linha de fabricação.

importância do recurso

O quanto um recurso é importante para as previsões de um modelo. Isso geralmente é expresso como uma pontuação numérica que pode ser calculada por meio de várias técnicas, como Shapley Additive Explanations (SHAP) e gradientes integrados. Para obter mais informações, consulte [Interpretabilidade do modelo de aprendizado de máquina com AWS](#).

transformação de recursos

O processo de otimizar dados para o processo de ML, incluindo enriquecer dados com fontes adicionais, escalar valores ou extrair vários conjuntos de informações de um único campo de dados. Isso permite que o modelo de ML se beneficie dos dados. Por exemplo, se a data “2021-05-27 00:15:37” for dividida em “2021”, “maio”, “quinta” e “15”, isso poderá ajudar o algoritmo de aprendizado a aprender padrões diferenciados associados a diferentes componentes de dados.

prompt few shot

Fornecer a um [LLM](#) um pequeno número de exemplos que demonstram a tarefa e o resultado desejado antes de solicitar que ele execute uma tarefa semelhante. Essa técnica é uma aplicação do aprendizado contextual, em que os modelos aprendem com exemplos (fotos) incorporados aos prompts. Few-shot a solicitação pode ser eficaz para tarefas que exigem formatação, raciocínio ou conhecimento de domínio específicos. Veja também [prompts zero-shot](#).

FGAC

Veja [controle de acesso refinado](#).

Controle de acesso refinado (FGAC)

O uso de várias condições para permitir ou negar uma solicitação de acesso.

migração flash-cut

Um método de migração de banco de dados que usa replicação contínua de dados via [captura de dados de alteração](#) para migrar os dados no menor tempo possível, em vez de usar uma abordagem em fases. O objetivo é reduzir ao mínimo o tempo de inatividade.

FM

Veja [modelo de base](#).

modelo de base (FM)

Uma grande rede neural de aprendizado profundo que treina em grandes conjuntos de dados generalizados e não rotulados. Os FMs são capazes de realizar uma ampla variedade de tarefas gerais, como entender a linguagem, gerar texto e imagens e conversar em linguagem natural. Para obter mais informações, consulte [O que são modelos de base?](#).

Gateway FM

[Um intermediário centralizado que controla e normaliza o acesso aos modelos de fundação.](#)

Também conhecido como gateway LLM.

G

IA generativa

Um subconjunto de modelos de [IA](#) que foram treinados em grandes quantidades de dados e que podem usar um simples prompt de texto para criar novos artefatos e conteúdo, como imagens, vídeos, texto e áudio. Para obter mais informações, consulte [O que é IA generativa?](#).

bloqueio geográfico

Veja [restrições geográficas](#).

restrições geográficas (bloqueio geográfico)

Na Amazon CloudFront, uma opção para impedir que usuários em países específicos acessem distribuições de conteúdo. É possível usar uma lista de permissões ou uma lista de bloqueios para especificar países aprovados e banidos. Para obter mais informações, consulte [Restringir a distribuição geográfica do seu conteúdo](#) na CloudFront documentação.

Fluxo de trabalho do GitFlow

Uma abordagem na qual ambientes inferiores e superiores usam ramificações diferentes em um repositório de código-fonte. O fluxo de trabalho do Gitflow é considerado legado, e o [fluxo de trabalho trunk-based](#) é a abordagem moderna e preferencial.

golden image

Um snapshot de um sistema ou software usado como modelo para implantar novas instâncias desse sistema ou software. Por exemplo, na manufatura, uma golden image pode ser usada para

provisionar software em vários dispositivos e ajudar a melhorar a velocidade, a escalabilidade e a produtividade nas operações de fabricação de dispositivos.

estratégia greenfield

A ausência de infraestrutura existente em um novo ambiente. Ao adotar uma estratégia greenfield para uma arquitetura de sistema, é possível selecionar todas as novas tecnologias sem a restrição da compatibilidade com a infraestrutura existente, também conhecida como [brownfield](#). Se estiver expandindo a infraestrutura existente, poderá combinar as estratégias brownfield e greenfield.

barreira de proteção

Uma regra de alto nível que ajuda a gerenciar recursos, políticas e conformidade em todas as unidades organizacionais (UOs). Barreiras de proteção preventivas impõem políticas para garantir o alinhamento a padrões de conformidade. Elas são implementadas usando políticas de controle de serviço e limites de permissões do IAM. Barreiras de proteção detectivas detectam violações de políticas e problemas de conformidade e geram alertas para remediação. Eles são implementados usando AWS Config, AWS Security Hub CSPM, Amazon GuardDuty AWS Trusted Advisor, Amazon Inspector e verificações personalizadas AWS Lambda .

grades de proteção (IA)

Mecanismos de segurança que filtram, validam e restringem as entradas e saídas dos [agentes](#) para ajudar a garantir um comportamento de IA responsável e seguro.

H

HA

Veja [alta disponibilidade](#).

migração heterogênea de bancos de dados

Migrar seu banco de dados de origem para um banco de dados de destino que usa um mecanismo de banco de dados diferente (por exemplo, Oracle para Amazon Aurora). A migração heterogênea geralmente faz parte de um esforço de redefinição da arquitetura, e converter o esquema pode ser uma tarefa complexa. [O AWS fornece o AWS SCT](#) para ajudar nas conversões de esquemas.

alta disponibilidade (HA)

A capacidade de uma workload operar continuamente, sem intervenção, em caso de desafios ou desastres. Os sistemas AH são projetados para realizar o failover automático, oferecer consistentemente desempenho de alta qualidade e lidar com diferentes cargas e falhas com impacto mínimo no desempenho.

modernização de historiador

Uma abordagem usada para modernizar e atualizar os sistemas de tecnologia operacional (OT) para melhor atender às necessidades do setor de manufatura. Um historiador é um tipo de banco de dados usado para coletar e armazenar dados de várias fontes em uma fábrica.

dados de hold-out

Uma parte dos dados históricos rotulados que são retidos de um conjunto de dados usado para treinar um modelo de [machine learning](#). Você pode usar dados de hold-out para avaliar a performance do modelo comparando as previsões do modelo com os dados de retenção.

humano no circuito (HiTL)

Um padrão de fluxo de trabalho em que a execução do [agente](#) é pausada para análise e aprovação humana em pontos críticos de decisão.

migração homogênea de bancos de dados

Migrar seu banco de dados de origem para um banco de dados de destino que compartilha o mesmo mecanismo de banco de dados (por exemplo, Microsoft SQL Server para Amazon RDS para SQL Server). A migração homogênea geralmente faz parte de um esforço de redefinição da hospedagem ou da plataforma. É possível usar utilitários de banco de dados nativos para migrar o esquema.

dados quentes

Dados acessados com frequência, como dados em tempo real ou dados translacionais recentes. Esses dados normalmente exigem uma camada ou classe de armazenamento de alto desempenho para fornecer respostas rápidas às consultas.

hotfix

Uma correção urgente para um problema crítico em um ambiente de produção. Devido à sua urgência, um hotfix geralmente é feito fora do fluxo de trabalho típico de uma DevOps versão.

período de hipercuidados

Imediatamente após a substituição, o período em que uma equipe de migração gerencia e monitora as aplicações migradas na nuvem para resolver quaisquer problemas. Normalmente, a duração desse período é de 1 a 4 dias. No final do período de hipercuidados, a equipe de migração normalmente transfere a responsabilidade pelas aplicações para a equipe de operações de nuvem.

eu

laC

Veja [infraestrutura como código](#).

Política baseada em identidade

Uma política anexada a um ou mais diretores do IAM que define suas permissões no Nuvem AWS ambiente.

aplicação ociosa

Uma aplicação que tem um uso médio de CPU e memória entre 5 e 20% em um período de 90 dias. Em um projeto de migração, é comum retirar essas aplicações ou retê-las on-premises.

IIoT

Veja [Internet das Coisas Industrial](#).

infraestrutura imutável

Um modelo que implanta uma nova infraestrutura para workloads de produção em vez de atualizar, aplicar patches ou modificar a infraestrutura existente. Infraestruturas imutáveis são inerentemente mais consistentes, confiáveis e preditivas do que [infraestruturas mutáveis](#). Para obter mais informações, consulte as melhores práticas de [implantação usando infraestrutura imutável](#) na AWS Well-Architected Estrutura.

VPC de entrada (admissão)

Em uma arquitetura de AWS várias contas, uma VPC que aceita, inspeciona e roteia conexões de rede de fora de um aplicativo. A [Arquitetura de referência de segurança da AWS](#) recomenda configurar sua conta de rede com VPCs de entrada, saída e inspeção para proteger a interface bidirecional entre a aplicação e a Internet em geral.

migração incremental

Uma estratégia de substituição na qual você migra a aplicação em pequenas partes, em vez de realizar uma única substituição completa. Por exemplo, é possível mover inicialmente apenas alguns microsserviços ou usuários para o novo sistema. Depois de verificar se tudo está funcionando corretamente, mova os microsserviços ou usuários adicionais de forma incremental até poder descomissionar seu sistema herdado. Essa estratégia reduz os riscos associados a migrações de grande porte.

Indústria 4.0

Um termo que foi introduzido por [Klaus Schwab](#) em 2016 para se referir à modernização dos processos de fabricação por meio de avanços na conectividade, dados em tempo real, automação, análise e. AI/ML

infraestrutura

Todos os recursos e ativos contidos no ambiente de uma aplicação.

Infraestrutura como código (IaC)

O processo de provisionamento e gerenciamento da infraestrutura de uma aplicação por meio de um conjunto de arquivos de configuração. A IaC foi projetada para ajudar você a centralizar o gerenciamento da infraestrutura, padronizar recursos e escalar rapidamente para que novos ambientes sejam reproduzíveis, confiáveis e consistentes.

Internet das Coisas Industrial (IIoT)

O uso de sensores e dispositivos conectados à Internet nos setores industriais, como manufatura, energia, automotivo, saúde, ciências biológicas e agricultura. Para obter mais informações, consulte [Construir uma estratégia de transformação digital para a Internet das Coisas Industrial \(IIoT\)](#).

VPC de inspeção

Em uma arquitetura de AWS várias contas, uma VPC centralizada que gerencia as inspeções do tráfego de rede entre VPCs (na mesma ou em diferentes Regiões da AWS), a Internet e as redes locais. A [Arquitetura de referência de segurança da AWS](#) recomenda configurar sua conta de rede com VPCs de entrada, saída e inspeção para proteger a interface bidirecional entre a aplicação e a Internet em geral.

Internet das coisas (IoT)

A rede de objetos físicos conectados com sensores ou processadores incorporados que se comunicam com outros dispositivos e sistemas pela Internet ou por uma rede de comunicação local. Para obter mais informações, consulte [O que é IoT?](#)

interpretabilidade

Uma característica de um modelo de machine learning que descreve o grau em que um ser humano pode entender como as previsões do modelo dependem de suas entradas. Para obter mais informações, consulte [Interpretabilidade do modelo de aprendizado de máquina com AWS](#).

IoT

Veja [Internet das Coisas](#).

Biblioteca de informações de TI (ITIL)

Um conjunto de práticas recomendadas para fornecer serviços de TI e alinhar esses serviços a requisitos de negócios. A ITIL fornece a base para o ITSM.

Gerenciamento de serviços de TI (ITSM)

Atividades associadas a design, implementação, gerenciamento e suporte de serviços de TI para uma organização. Para obter informações sobre a integração de operações em nuvem com ferramentas de ITSM, consulte o [guia de integração de operações](#).

ITIL

Veja [biblioteca de informações de TI](#).

ITSM

Veja [gerenciamento de serviços de TI](#).

L

controle de acesso baseado em etiqueta (LBAC)

Uma implementação do controle de acesso obrigatório (MAC) em que os usuários e os dados em si recebem explicitamente um valor de etiqueta de segurança. A interseção entre a etiqueta de segurança do usuário e a etiqueta de segurança dos dados determina quais linhas e colunas podem ser vistas pelo usuário.

zona de pouso

Uma landing zone é um AWS ambiente bem arquitetado, com várias contas, escalável e seguro. Um ponto a partir do qual suas organizações podem iniciar e implantar rapidamente workloads e aplicações com confiança em seu ambiente de segurança e infraestrutura. Para obter mais informações sobre zonas de pouso, consulte [Configurar um ambiente da AWS com várias contas seguro e escalável](#).

grande modelo de linguagem (LLM)

Um modelo de [IA](#) de aprendizado profundo pré-treinado em uma grande quantidade de dados. Um LLM pode realizar várias tarefas, como responder a perguntas, resumir documentos, traduzir texto para outros idiomas e completar frases. Para obter mais informações, consulte [O que é grande modelo de linguagem \(LLM\)?](#).

migração de grande porte

Uma migração de 300 servidores ou mais.

LBAC

Veja [controle de acesso baseado em rótulo](#).

privilégio mínimo

A prática recomendada de segurança de conceder as permissões mínimas necessárias para executar uma tarefa. Para obter mais informações, consulte [Aplicar permissões de privilégios mínimos](#) na documentação do IAM.

mover sem alterações (lift-and-shift)

Veja [7 Rs](#).

sistema little-endian

Um sistema que armazena o byte menos significativo antes. Veja também [endianness](#).

LLM

Veja [grande modelo de linguagem](#).

ambientes inferiores

Veja [ambiente](#).

M

machine learning (ML)

Um tipo de inteligência artificial que usa algoritmos e técnicas para reconhecimento e aprendizado de padrões. O ML analisa e aprende com dados gravados, por exemplo, dados da Internet das Coisas (IoT), para gerar um modelo estatístico baseado em padrões. Para obter mais informações, consulte [Machine learning](#).

ramificação principal

Veja [ramificação](#).

Malware

Software projetado para comprometer a segurança ou a privacidade do computador. O malware pode interromper os sistemas do computador, vaziar informações sensíveis ou obter acesso não autorizado. Exemplos de malware incluem vírus, worms, ransomware, cavalos de Troia, spyware e keyloggers.

Serviços gerenciados

Serviços da AWS para o qual AWS opera a camada de infraestrutura, o sistema operacional e as plataformas, e você acessa os endpoints para armazenar e recuperar dados. O Amazon Simple Storage Service (Amazon S3) e o Amazon DynamoDB são exemplos de serviços gerenciados. Eles também são conhecidos como serviços abstraídos.

sistema de execução de manufatura (MES)

Um sistema de software para rastrear, monitorar, documentar e controlar processos de produção que convertem matérias-primas em produtos acabados no chão de fábrica.

MAP

Veja [Programa de Aceleração da Migração](#).

MCP

Consulte [Protocolo de contexto do modelo](#).

Protocolo de contexto para modelos (MCP)

Um protocolo sem estado para comunicação entre [agentes](#) e [ferramentas](#).

Servidor MCP

Um serviço que expõe uma ou mais [ferramentas](#) por meio do [Model Context Protocol](#).

mecanismo

Um processo completo em que você cria uma ferramenta, impulsiona a adoção da ferramenta e, em seguida, inspeciona os resultados para fazer ajustes. Um mecanismo é um ciclo que se reforça e se aprimora à medida que opera. Para obter mais informações, consulte [Criação de mecanismos](#) na AWS Well-Architected estrutura.

conta de membro

Todos, Contas da AWS exceto a conta de gerenciamento, que fazem parte de uma organização em AWS Organizations. Uma conta só pode ser membro de uma organização de cada vez.

MES

Veja [sistema de execução de manufatura](#).

Transporte de Telemetria de Enfileiramento de Mensagens (MQTT)

[Um protocolo de comunicação leve, máquina a máquina \(M2M\), baseado no padrão, para dispositivos de IoT com recursos publish/subscribelimitados.](#)

microsserviço

Um serviço pequeno e independente que se comunica por meio de APIs bem definidas e normalmente pertence a equipes pequenas e autônomas. Por exemplo, um sistema de seguradora pode incluir microsserviços que mapeiam as capacidades comerciais, como vendas ou marketing, ou subdomínios, como compras, reclamações ou análises. Os benefícios dos microsserviços incluem agilidade, escalabilidade flexível, fácil implantação, código reutilizável e resiliência. Para obter mais informações, consulte [Integração de microsserviços usando serviços sem AWS servidor](#).

arquitetura de microsserviços

Uma abordagem à criação de aplicações com componentes independentes que executam cada processo de aplicação como um microsserviço. Esses microsserviços se comunicam por meio de uma interface bem definida usando APIs leves. Cada microsserviço nessa arquitetura pode ser atualizado, implantado e escalado para atender à demanda por funções específicas de uma aplicação. Para obter mais informações, consulte [Implementação de microsserviços em. AWS](#)

Programa de Aceleração da Migração (MAP)

Um AWS programa que fornece suporte de consultoria, treinamento e serviços para ajudar as organizações a criar uma base operacional sólida para migrar para a nuvem e ajudar a

compensar o custo inicial das migrações. O MAP inclui uma metodologia de migração para executar migrações legadas de forma metódica e um conjunto de ferramentas para automatizar e acelerar cenários comuns de migração.

migração em escala

O processo de mover a maior parte do portfólio de aplicações para a nuvem em ondas, com mais aplicações sendo movidas em um ritmo mais rápido a cada onda. Essa fase usa as práticas recomendadas e lições aprendidas nas fases anteriores para implementar uma fábrica de migração de equipes, ferramentas e processos para agilizar a migração de workloads por meio de automação e entrega ágeis. Esta é a terceira fase da [estratégia de migração para a AWS](#).

fábrica de migração

Cross-functional equipes que simplificam a migração de cargas de trabalho por meio de abordagens automatizadas e ágeis. As equipes da fábrica de migração geralmente incluem operações, analistas e proprietários de negócios, engenheiros de migração, desenvolvedores e DevOps profissionais que trabalham em sprints. Entre 20 e 50% de um portfólio de aplicações corporativas consiste em padrões repetidos que podem ser otimizados por meio de uma abordagem de fábrica. Para obter mais informações, consulte [discussão sobre fábricas de migração](#) e o [guia do Cloud Migration Factory](#) neste conjunto de conteúdo.

metadados de migração

As informações sobre a aplicação e o servidor necessárias para concluir a migração. Cada padrão de migração exige um conjunto de metadados de migração diferente. Exemplos de metadados de migração incluem a sub-rede, o grupo de segurança e AWS a conta de destino.

padrão de migração

Uma tarefa de migração repetível que detalha a estratégia de migração, o destino da migração e a aplicação ou o serviço de migração usado. Exemplo: rehoste a migração para o Amazon EC2 AWS com o Application Migration Service.

Avaliação de Portfólio para Migração (MPA)

Uma ferramenta on-line que fornece informações para validar o caso de negócios para migrar para a Nuvem AWS. O MPA fornece avaliação detalhada do portfólio (dimensionamento correto do servidor, preços, comparações de TCO, análise de custos de migração), bem como planejamento de migração (análise e coleta de dados de aplicações, agrupamento de aplicações, priorização de migração e planejamento de ondas). A [ferramenta MPA](#) (requer login) está disponível gratuitamente para todos os AWS consultores e consultores parceiros da APN.

Avaliação de Preparação para Migração (MRA)

O processo de obter insights sobre o status de prontidão de uma organização para a nuvem, identificar pontos fortes e fracos e criar um plano de ação para fechar as lacunas identificadas, usando o CAF. AWS Para mais informações, consulte o [guia de preparação para migração](#). A MRA é a primeira fase da [estratégia de migração para a AWS](#).

estratégia de migração

A abordagem usada para migrar uma workload para a Nuvem AWS. Para obter mais informações, veja a entrada [7 Rs](#) neste glossário e consulte [Mobilize sua organização para acelerar migrações em grande escala](#).

ML

Veja [machine learning](#).

modernização

Transformar uma aplicação desatualizada (herdada ou monolítica) e sua infraestrutura em um sistema ágil, elástico e altamente disponível na nuvem para reduzir custos, ganhar eficiência e aproveitar as inovações. Para obter mais informações, consulte [Strategy for modernizing applications in the Nuvem AWS](#).

avaliação de preparação para modernização

Uma avaliação que ajuda a determinar a preparação para modernização das aplicações de uma organização. Ela identifica benefícios, riscos e dependências e determina o quão bem a organização pode acomodar o estado futuro dessas aplicações. O resultado da avaliação é um esquema da arquitetura de destino, um roteiro que detalha as fases de desenvolvimento e os marcos do processo de modernização e um plano de ação para abordar as lacunas identificadas. Para obter mais informações, consulte [Evaluating modernization readiness for applications in the Nuvem AWS](#).

aplicações monolíticas (monólitos)

Aplicações que são executadas como um único serviço com processos fortemente acoplados. As aplicações monolíticas apresentam várias desvantagens. Se um recurso da aplicação apresentar um aumento na demanda, toda a arquitetura deverá ser escalada. Adicionar ou melhorar os recursos de uma aplicação monolítica também se torna mais complexo quando a base de código cresce. Para resolver esses problemas, é possível criar uma arquitetura de microsserviços. Para obter mais informações, consulte [Decompor monólitos em microsserviços](#).

MPA

Veja [Avaliação do Portfólio para Migração](#).

MQTT

Veja [Transporte de Telemetria de Enfileiramento de Mensagens](#).

classificação multiclasse

Um processo que ajuda a gerar previsões para várias classes (prevendo um ou mais de dois resultados). Por exemplo, um modelo de ML pode perguntar “Este produto é um livro, um carro ou um telefone?” ou “Qual categoria de produtos é mais interessante para este cliente?”

infraestrutura mutável

Um modelo que atualiza e modifica a infraestrutura existente para workloads de produção. Para melhorar a consistência, confiabilidade e previsibilidade, a AWS Well-Architected Estrutura recomenda o uso de [infraestrutura imutável](#) como uma prática recomendada.

O

OAC

Veja [controle de acesso de origem](#).

OAI

Veja [identidade de acesso de origem](#).

OCM

Veja [gerenciamento de alterações organizacionais](#).

migração offline

Um método de migração no qual a workload de origem é desativada durante o processo de migração. Esse método envolve tempo de inatividade prolongado e geralmente é usado para workloads pequenas e não críticas.

OI

Veja [integração de operações](#).

Ola

Veja [acordo de nível operacional](#).

migração online

Um método de migração no qual a workload de origem é copiada para o sistema de destino sem ser colocada offline. As aplicações conectadas à workload podem continuar funcionando durante a migração. Esse método envolve um tempo de inatividade nulo ou mínimo e normalmente é usado para workloads essenciais para a produção.

OPC-UA

Veja [Open Process Communications - Unified Architecture](#).

Comunicação de processo aberto - Arquitetura unificada (OPC-UA)

Um protocolo de comunicação máquina a máquina (M2M) para automação industrial. OPC-UA fornece um padrão de interoperabilidade com esquemas de criptografia, autenticação e autorização de dados.

acordo de nível operacional (OLA)

Um acordo que esclarece o que os grupos funcionais de TI prometem oferecer uns aos outros para apoiar um acordo de serviço (SLA).

análise de prontidão operacional (ORR)

Uma lista de verificação de perguntas e práticas recomendadas associadas que ajudam você a entender, avaliar, prevenir ou reduzir o escopo de incidentes e possíveis falhas. Para obter mais informações, consulte [Operational Readiness Reviews \(ORR\)](#) na AWS Well-Architected Estrutura.

tecnologia operacional (TO)

Sistemas de hardware e software que trabalham com o ambiente físico para controlar operações, equipamentos e infraestrutura industriais. Na manufatura, a integração dos sistemas de tecnologia da informação (TI) e tecnologia operacional (TO) é o foco principal das transformações da [Indústria 4.0](#).

integração de operações (OI)

O processo de modernização das operações na nuvem, que envolve planejamento de preparação, automação e integração. Para obter mais informações, consulte o [guia de integração de operações](#).

trilha organizacional

Uma trilha criada por ela AWS CloudTrail registra todos os eventos de todos Contas da AWS em uma organização em AWS Organizations. Essa trilha é criada em cada Conta da AWS que faz parte da organização e monitora a atividade em cada conta. Para obter mais informações, consulte [Criação de uma trilha para uma organização](#) na CloudTrail documentação.

gerenciamento de alterações organizacionais (OCM)

Uma estrutura para gerenciar grandes transformações de negócios disruptivas de uma perspectiva de pessoas, cultura e liderança. O OCM ajuda as organizações a se prepararem e fazerem a transição para novos sistemas e estratégias, acelerando a adoção de alterações, abordando questões de transição e promovendo mudanças culturais e organizacionais. Na estratégia de AWS migração, essa estrutura é chamada de aceleração de pessoas, devido à velocidade de mudança necessária nos projetos de adoção da nuvem. Para obter mais informações, consulte o [guia do OCM](#).

controle de acesso de origem (OAC)

Em CloudFront, uma opção aprimorada para restringir o acesso para proteger seu conteúdo do Amazon Simple Storage Service (Amazon S3). O OAC oferece suporte a todos os buckets do S3 Regiões da AWS, à criptografia do lado do servidor com AWS KMS (SSE-KMS) e à dinâmica PUT e DELETE às solicitações ao bucket do S3.

Identidade do acesso de origem (OAI)

Em CloudFront, uma opção para restringir o acesso para proteger seu conteúdo do Amazon S3. Quando você usa o OAI, CloudFront cria um principal com o qual o Amazon S3 pode se autenticar. Os diretores autenticados podem acessar o conteúdo em um bucket do S3 somente por meio de uma distribuição específica. CloudFront Veja também [OAC](#), que fornece um controle de acesso mais granular e aprimorado.

ORR

Veja [análise de prontidão operacional](#).

OT

Veja [tecnologia operacional](#).

VPC de saída (egresso)

Em uma arquitetura de AWS várias contas, uma VPC que gerencia conexões de rede que são iniciadas de dentro de um aplicativo. A [Arquitetura de referência de segurança da AWS](#)

recomenda configurar sua conta de rede com VPCs de entrada, saída e inspeção para proteger a interface bidirecional entre a aplicação e a Internet em geral.

P

limite de permissões

Uma política de gerenciamento do IAM anexada a entidades principais do IAM para definir as permissões máximas que o usuário ou perfil podem ter. Para obter mais informações, consulte [Limites de permissões](#) na documentação do IAM.

Informações de identificação pessoal (PII)

Informações que, quando visualizadas diretamente ou combinadas com outros dados relacionados, podem ser usadas para inferir razoavelmente a identidade de um indivíduo. Exemplos de PII incluem nomes, endereços e informações de contato.

PII

Veja [informações de identificação pessoal](#).

manual

Um conjunto de etapas predefinidas que capturam o trabalho associado às migrações, como a entrega das principais funções operacionais na nuvem. Um manual pode assumir a forma de scripts, runbooks automatizados ou um resumo dos processos ou etapas necessários para operar seu ambiente modernizado.

PLC

Veja [controlador lógico programável](#).

PLM

Veja [gerenciamento do ciclo de vida do produto](#).

política

Um objeto que pode definir permissões (veja [política baseada em identidade](#)), especificar condições de acesso (veja [política baseada em recurso](#)) ou definir as permissões máximas para todas as contas em uma organização no AWS Organizations (veja [política de controle de serviços](#)).

persistência poliglota

Escolher de forma independente a tecnologia de armazenamento de dados de um microsserviço com base em padrões de acesso a dados e outros requisitos. Se seus microsserviços tiverem a mesma tecnologia de armazenamento de dados, eles poderão enfrentar desafios de implementação ou apresentar baixa performance. Os microsserviços serão implementados com mais facilidade e alcançarão performance e escalabilidade melhores se usarem o armazenamento de dados mais bem adaptado às suas necessidades.

avaliação do portfólio

Um processo de descobrir, analisar e priorizar o portfólio de aplicações para planejar a migração. Para obter mais informações, consulte [Avaliar a preparação para a migração](#).

predicado

Uma condição de consulta que retorna `true` ou `false`, normalmente localizada em uma cláusula `WHERE`.

pushdown de predicados

Uma técnica de otimização de consultas de banco de dados que filtra os dados na consulta antes da transferência. Isso reduz a quantidade de dados que devem ser recuperados e processados do banco de dados relacional e melhora a performance das consultas.

controle preventivo

Um controle de segurança projetado para evitar que um evento ocorra. Esses controles são a primeira linha de defesa para ajudar a evitar acesso não autorizado ou alterações indesejadas em sua rede. Para obter mais informações, consulte [Controles preventivos](#) em Como implementar controles de segurança na AWS.

principal (entidade principal)

Uma entidade AWS que pode realizar ações e acessar recursos. Essa entidade geralmente é um usuário raiz para um Conta da AWS, uma função do IAM ou um usuário. Para obter mais informações, consulte Entidade principal em [Termos e conceitos de perfis](#) na documentação do IAM.

Privacidade por design

Uma abordagem em engenharia de sistemas que leva em consideração a privacidade em todo o processo de desenvolvimento.

zonas hospedadas privadas

Um contêiner que armazena informações sobre como você quer que o Amazon Route 53 responda a consultas ao DNS para um domínio e seus subdomínios dentro de uma ou mais VPCs. Para obter mais informações, consulte [Como trabalhar com zonas hospedadas privadas](#) na documentação do Route 53.

controle proativo

Um [controle de segurança](#) desenvolvido para evitar a implantação de recursos não conformes. Esses controles verificam os recursos antes de serem provisionados. Se o recurso não estiver em conformidade com o controle, ele não será provisionado. Para obter mais informações, consulte o [guia de referência de controles](#) na AWS Control Tower documentação e consulte [Controles proativos](#) em Implementação de controles de segurança em AWS.

gerenciamento do ciclo de vida do produto (PLM)

O gerenciamento de dados e processos de um produto em todo o seu ciclo de vida, desde a concepção, o desenvolvimento e o lançamento, passando pelo crescimento e maturidade, até o declínio e a remoção.

ambiente de produção

Veja [ambiente](#).

controlador lógico programável (PLC)

Na manufatura, um computador altamente confiável e adaptável que monitora as máquinas e automatiza os processos de fabricação.

encadeamento de prompts

Uso da saída de um prompt do [LLM](#) como entrada para o próximo prompt para gerar respostas melhores. Essa técnica é usada para dividir uma tarefa complexa em subtarefas, ou para refinar ou expandir iterativamente uma resposta preliminar. Isso ajuda a melhorar a precisão e a relevância das respostas de um modelo e permite resultados mais granulares e personalizados.

pseudonimização

O processo de substituir identificadores pessoais em um conjunto de dados por valores de espaço reservado. A pseudonimização pode ajudar a proteger a privacidade pessoal. Os dados pseudonimizados ainda são considerados dados pessoais.

publish/subscribe (pub/sub)

Um padrão que permite comunicações assíncronas entre microsserviços para melhorar a escalabilidade e a capacidade de resposta. Por exemplo, em um [MES](#) baseado em microsserviços, um microsserviço pode publicar mensagens de eventos em um canal em que outros microsserviços possam assinar. O sistema pode adicionar novos microsserviços sem alterar o serviço de publicação.

Q

plano de consulta

Uma série de etapas, como instruções, usadas para acessar os dados em um sistema de banco de dados relacional SQL.

regressão de planos de consultas

Quando um otimizador de serviço de banco de dados escolhe um plano menos adequado do que escolhia antes de uma determinada alteração no ambiente de banco de dados ocorrer. Isso pode ser causado por alterações em estatísticas, restrições, configurações do ambiente, associações de parâmetros de consulta e atualizações do mecanismo de banco de dados.

R

Matriz RACI

Veja [responsável, aprovador, consultado, informado \(RACI\)](#).

RAG

Veja [geração aumentada via recuperação](#).

ransomware

Um software mal-intencionado desenvolvido para bloquear o acesso a um sistema ou dados de computador até que um pagamento seja feito.

Matriz RASCI

Veja [responsável, aprovador, consultado, informado \(RACI\)](#).

RCAC

Veja [controle de acesso por linha e coluna](#).

réplica de leitura

Uma cópia de um banco de dados usada somente para leitura. É possível encaminhar consultas para a réplica de leitura e reduzir a carga no banco de dados principal.

Redefinir arquitetura

Veja [7 Rs](#).

objetivo de ponto de recuperação (RPO).

O máximo período de tempo aceitável desde o último ponto de recuperação de dados.

Isso determina o que é considerado uma perda aceitável de dados entre o último ponto de recuperação e a interrupção do serviço.

objetivo de tempo de recuperação (RTO)

O máximo atraso aceitável entre a interrupção e a restauração do serviço.

refatorar

Veja [7 Rs](#).

Região

Uma coleção de AWS recursos em uma área geográfica. Cada um Região da AWS é isolado e independente dos outros para fornecer tolerância a falhas, estabilidade e resiliência. Para obter informações, consulte [Specify which Regiões da AWS your account can use](#).

regressão

Uma técnica de ML que prevê um valor numérico. Por exemplo, para resolver o problema de “Por qual preço esta casa será vendida?” um modelo de ML pode usar um modelo de regressão linear para prever o preço de venda de uma casa com base em fatos conhecidos sobre a casa (por exemplo, a metragem quadrada).

redefinir a hospedagem

Veja [7 Rs](#).

versão

Em um processo de implantação, o ato de promover mudanças em um ambiente de produção.

realocar

Veja [7 Rs](#).

redefinir a plataforma

Veja [7 Rs](#).

recomprar

Veja [7 Rs](#).

resiliência

A capacidade de uma aplicação de resistir ou se recuperar de interrupções. [Alta disponibilidade](#) e [recuperação de desastres](#) são considerações comuns ao planejar a resiliência na Nuvem AWS. Para obter mais informações, consulte [Nuvem AWS Resilience](#).

política baseada em recurso

Uma política associada a um recurso, como um bucket do Amazon S3, um endpoint ou uma chave de criptografia. Esse tipo de política especifica quais entidades principais têm acesso permitido, ações válidas e quaisquer outras condições que devem ser atendidas.

matriz responsável, accountable, consultada, informada (RACI)

Uma matriz que define as funções e responsabilidades de todas as partes envolvidas nas atividades de migração e nas operações de nuvem. O nome da matriz é derivado dos tipos de responsabilidade definidos na matriz: responsável (R), responsabilizável (A), consultado (C) e informado (I). O tipo de suporte (S) é opcional. Se você incluir suporte, a matriz será chamada de matriz RASCI e, se excluir, será chamada de matriz RACI.

controle responsivo

Um controle de segurança desenvolvido para conduzir a remediação de eventos adversos ou desvios em relação à linha de base de segurança. Para obter mais informações, consulte [Controles responsivos](#) em Como implementar controles de segurança na AWS.

reter

Veja [7 Rs](#).

Retirada

Veja [7 Rs](#).

Geração Aumentada de Recuperação (RAG)

Uma tecnologia de [IA generativa](#) em que um [LLM](#) faz referência a uma fonte de dados autorizada que está fora de suas fontes de dados de treinamento antes de gerar uma resposta. Por exemplo, um modelo RAG pode realizar uma pesquisa semântica na base de conhecimento ou nos dados personalizados de uma organização. Para obter mais informações, consulte [O que é RAG \(geração aumentada via recuperação\)?](#).

alternância

O processo de atualizar periodicamente um [segredo](#) para dificultar o acesso de um invasor às credenciais.

controle de acesso por linha e coluna (RCAC)

O uso de expressões SQL básicas e flexíveis que tenham regras de acesso definidas. O RCAC consiste em permissões de linha e máscaras de coluna.

RPO

Veja [objetivo de ponto de recuperação](#).

RTO

Veja [objetivo de tempo de recuperação](#).

runbook

Um conjunto de procedimentos manuais ou automatizados necessários para realizar uma tarefa específica. Eles são normalmente criados para agilizar operações ou procedimentos repetitivos com altas taxas de erro.

S

SAML 2.0

Um padrão aberto que muitos provedores de identidade (IdPs) usam. Esse recurso permite o login único federado (SSO), para que os usuários possam fazer login no Console de gerenciamento da AWS ou chamar as operações da AWS API sem que você precise criar um usuário no IAM para todos em sua organização. Para obter mais informações sobre a federação baseada em SAML 2.0, consulte [Sobre a federação baseada em SAML 2.0](#) na documentação do IAM.

SCADA

Veja [controle de supervisão e aquisição de dados](#).

SCP

Veja [política de controle de serviço](#).

secret

Em AWS Secrets Manager, informações confidenciais ou restritas, como uma senha ou credenciais de usuário, que você armazena de forma criptografada. Consiste no valor secreto e em seus metadados. O valor secreto pode ser binário, uma única string ou várias strings. Para obter mais informações, consulte [What's in a Secrets Manager secret?](#) na documentação do Secrets Manager.

segurança desde a concepção

Uma abordagem em engenharia de sistemas que leva em consideração a segurança em todo o processo de desenvolvimento.

controle de segurança

Uma barreira de proteção técnica ou administrativa que impede, detecta ou reduz a capacidade de uma ameaça explorar uma vulnerabilidade de segurança. Existem quatro tipos primários de controles de segurança: [preventivos](#), [detectivos](#), [responsivos](#) e [proativos](#).

hardening da segurança

O processo de reduzir a superfície de ataque para torná-la mais resistente a ataques. Isso pode incluir ações como remover recursos que não são mais necessários, implementar a prática recomendada de segurança de conceder privilégios mínimos ou desativar recursos desnecessários em arquivos de configuração.

sistema de gerenciamento de eventos e informações de segurança (SIEM)

Ferramentas e serviços que combinam sistemas de gerenciamento de informações de segurança (SIM) e gerenciamento de eventos de segurança (SEM). Um sistema SIEM coleta, monitora e analisa dados de servidores, redes, dispositivos e outras fontes para detectar ameaças e violações de segurança e gerar alertas.

automação de resposta de segurança

Uma ação predefinida e programada projetada para responder ou remediar automaticamente um evento de segurança. Essas automações servem como controles de segurança [responsivos](#) ou [detectivos](#) que ajudam você a implementar as melhores práticas AWS de segurança. Exemplos de ações de resposta automatizada incluem a modificação de um grupo de segurança da VPC, a aplicação de patches em uma instância do Amazon EC2 ou a alternância de credenciais.

Criptografia do lado do servidor

Criptografia dos dados em seu destino, por AWS service (Serviço da AWS) quem os recebe.
política de controle de serviços (SCP)

Uma política que fornece controle centralizado sobre as permissões de todas as contas em uma organização no AWS Organizations. As SCPs definem barreiras de proteção ou estabelecem limites para as ações que um administrador pode delegar a usuários ou perfis. É possível usar SCPs como listas de permissão ou de negação para especificar quais serviços ou ações são permitidos ou proibidos. Para obter mais informações, consulte [Políticas de controle de serviço](#) na AWS Organizations documentação.

service endpoint (endpoint de serviço)

O URL do ponto de entrada para um AWS service (Serviço da AWS). Você pode usar o endpoint para se conectar programaticamente ao serviço de destino. Para obter mais informações, consulte [Endpoints do AWS service \(Serviço da AWS\)](#) na Referência geral da AWS.

acordo de serviço (SLA)

Um acordo que esclarece o que uma equipe de TI promete fornecer aos clientes, como tempo de atividade e performance do serviço.

indicador de nível de serviço (SLI)

Uma avaliação de um aspecto de performance de um serviço, como taxa de erro, disponibilidade ou throughput.

objetivo de nível de serviço (SLO)

Uma métrica alvo que representa a integridade de um serviço, conforme avaliado por um [indicador de nível de serviço](#).

modelo de responsabilidade compartilhada

Um modelo que descreve a responsabilidade com a qual você compartilha AWS pela segurança e conformidade na nuvem. AWS é responsável pela segurança da nuvem, enquanto você é responsável pela segurança na nuvem. Para obter mais informações, consulte o [Modelo de responsabilidade compartilhada](#).

Inteligência artificial sombria

Aplicativos de [IA](#) não autorizados criados ou usados fora dos canais controlados dentro de uma organização.

SIEM

Veja [sistema de gerenciamento de eventos e informações de segurança](#).

ponto único de falha (SPOF)

Uma falha em um único componente crítico de uma aplicação que pode interromper o sistema.

SLA

Veja [acordo de serviço](#).

SLI

Veja [indicador de nível de serviço](#).

SLO

Veja [objetivo de nível de serviço](#).

modelo dividir e semear

Um padrão para escalar e acelerar projetos de modernização. À medida que novos recursos e lançamentos de produtos são definidos, a equipe principal se divide para criar novas equipes de produtos. Isso ajuda a escalar os recursos e os serviços da sua organização, melhora a produtividade do desenvolvedor e possibilita inovações rápidas. Para obter mais informações, consulte [Phased approach to modernizing applications in the Nuvem AWS](#).

SPOF

Veja [ponto único de falha](#).

esquema em estrela

Uma estrutura organizacional de banco de dados que usa uma grande tabela de fatos para armazenar dados transacionais ou medidos e usa uma ou mais tabelas dimensionais menores para armazenar atributos de dados. Essa estrutura foi projetada para ser usada em um [data warehouse](#) ou para fins de inteligência comercial.

padrão strangler fig

Uma abordagem à modernização de sistemas monolíticos que consiste em reescrever e substituir incrementalmente a funcionalidade do sistema até que o sistema herdado possa ser desativado. Esse padrão usa a analogia de uma videira que cresce e se torna uma árvore estabelecida e, eventualmente, supera e substitui sua hospedeira. O padrão foi [apresentado por Martin Fowler](#)

como forma de gerenciar riscos ao reescrever sistemas monolíticos. Para ver um exemplo de como aplicar esse padrão, consulte [Modernizando os serviços web legados da Microsoft ASP.NET \(ASMX\) de forma incremental usando contêineres e o Amazon API Gateway](#).

sub-rede

Um intervalo de endereços IP na VPC. Cada sub-rede fica alocada em uma única zona de disponibilidade.

controle supervisorio e aquisição de dados (SCADA)

Na manufatura, um sistema que usa hardware e software para monitorar ativos físicos e operações de produção.

symmetric encryption (criptografia simétrica)

Um algoritmo de criptografia que usa a mesma chave para criptografar e descriptografar dados.

testes sintéticos

Testar um sistema de forma que simule as interações do usuário para detectar possíveis problemas ou monitorar a performance. Você pode usar o [Amazon CloudWatch Synthetics](#) para criar esses testes.

prompt do sistema

Uma técnica para fornecer contexto, instruções ou orientações a um [LLM](#) a fim de direcionar seu comportamento. Os prompts do sistema ajudam a definir o contexto e a estabelecer regras para interações com os usuários.

T

tags

Key-value pares que atuam como metadados para organizar seus AWS recursos. As tags podem ajudar você a gerenciar, identificar, organizar, pesquisar e filtrar recursos da . Para obter mais informações, consulte [Marcar seus recursos do AWS](#).

variável-alvo

O valor que você está tentando prever no ML supervisionado. Ela também é conhecida como variável de resultado. Por exemplo, em uma configuração de fabricação, a variável-alvo pode ser um defeito do produto.

lista de tarefas

Uma ferramenta usada para monitorar o progresso por meio de um runbook. Uma lista de tarefas contém uma visão geral do runbook e uma lista de tarefas gerais a serem concluídas. Para cada tarefa geral, ela inclui o tempo estimado necessário, o proprietário e o progresso.

ambiente de teste

Veja [ambiente](#).

treinamento

O processo de fornecer dados para que seu modelo de ML aprenda. Os dados de treinamento devem conter a resposta correta. O algoritmo de aprendizado descobre padrões nos dados de treinamento que mapeiam os atributos dos dados de entrada no destino (a resposta que você deseja prever). Ele gera um modelo de ML que captura esses padrões. Você pode usar o modelo de ML para obter previsões de novos dados cujo destino você não conhece.

ferramenta

Uma função ou API que um [agente](#) pode invocar para realizar operações em sistemas externos.

gateway de trânsito

Um hub de trânsito de rede que pode ser usado para interconectar as VPCs e as redes on-premises. Para obter mais informações, consulte [O que é um gateway de trânsito](#) na AWS Transit Gateway documentação.

fluxo de trabalho baseado em troncos

Uma abordagem na qual os desenvolvedores criam e testam recursos localmente em uma ramificação de recursos e, em seguida, mesclam essas alterações na ramificação principal. A ramificação principal é então criada para os ambientes de desenvolvimento, pré-produção e produção, sequencialmente.

Acesso confiável

Conceder permissões a um serviço que você especifica para realizar tarefas em sua organização AWS Organizations e em suas contas em seu nome. O serviço confiável cria um perfil vinculado ao serviço em cada conta, quando esse perfil é necessário, para realizar tarefas de gerenciamento para você. Para obter mais informações, consulte [Usando AWS Organizations com outros AWS serviços](#) na AWS Organizations documentação.

tuning (ajustar)

Alterar aspectos do processo de treinamento para melhorar a precisão do modelo de ML. Por exemplo, você pode treinar o modelo de ML gerando um conjunto de rótulos, adicionando rótulos e repetindo essas etapas várias vezes em configurações diferentes para otimizar o modelo.

equipe de duas pizzas

Uma pequena DevOps equipe que você pode alimentar com duas pizzas. Uma equipe de duas pizzas garante a melhor oportunidade possível de colaboração no desenvolvimento de software.

U

incerteza

Um conceito que se refere a informações imprecisas, incompletas ou desconhecidas que podem minar a confiabilidade dos modelos preditivos de ML. Há dois tipos de incertezas: a incerteza epistêmica é causada por dados limitados e incompletos, enquanto a incerteza aleatória é causada pelo ruído e pela aleatoriedade inerentes aos dados.

tarefas indiferenciadas

Também conhecido como trabalho pesado, trabalho necessário para criar e operar um aplicativo, mas que não fornece valor direto ao usuário final nem oferece vantagem competitiva. Exemplos de tarefas indiferenciadas incluem aquisição, manutenção e planejamento de capacidade.

ambientes superiores

Veja [ambiente](#).

V

aspiração

Uma operação de manutenção de banco de dados que envolve limpeza após atualizações incrementais para recuperar armazenamento e melhorar a performance.

controle de versões

Processos e ferramentas que rastreiam mudanças, como alterações no código-fonte em um repositório.

emparelhamento de VPC

Uma conexão entre duas VPCs que permite rotear tráfego usando endereços IP privados. Para ter mais informações, consulte [O que é emparelhamento de VPC?](#) na documentação da Amazon VPC.

Vulnerabilidade

Uma falha de software ou hardware que compromete a segurança do sistema.

W

cache quente

Um cache de buffer que contém dados atuais e relevantes que são acessados com frequência. A instância do banco de dados pode ler do cache do buffer, o que é mais rápido do que ler da memória principal ou do disco.

dados mornos

Dados acessados raramente. Ao consultar esse tipo de dados, consultas moderadamente lentas geralmente são aceitáveis.

função de janela

Uma função SQL que executa um cálculo em um grupo de linhas que se relacionam de alguma forma com o registro atual. As funções de janela são úteis para processar tarefas, como calcular uma média móvel ou acessar o valor das linhas com base na posição relativa da linha atual.

workload

Uma coleção de códigos e recursos que geram valor empresarial, como uma aplicação voltada para o cliente ou um processo de backend.

workstreams

Grupos funcionais em um projeto de migração que são responsáveis por um conjunto específico de tarefas. Cada workstream é independente, mas oferece suporte aos outros workstreams do projeto. Por exemplo, o workstream de portfólio é responsável por priorizar aplicações, planejar ondas e coletar metadados de migração. O workstream de portfólio entrega esses ativos ao workstream de migração, que então migra os servidores e as aplicações.

WORM

Veja [gravação única e várias leituras](#).

WQF

Veja [AWS Workload Qualification Framework](#).

gravação única e várias leituras (WORM)

Um modelo de armazenamento que grava dados uma única vez e evita que os dados sejam excluídos ou modificados. Os usuários autorizados podem ler os dados quantas vezes forem necessárias, mas não podem alterá-los. Essa infraestrutura de armazenamento de dados é considerada [imutável](#).

Z

exploração de dia zero

Um ataque, normalmente malware, que tira proveito de uma [vulnerabilidade zero-day](#).

vulnerabilidade de dia zero

Uma falha ou vulnerabilidade não mitigada em um sistema de produção. Os agentes de ameaças podem usar esse tipo de vulnerabilidade para atacar o sistema. Os desenvolvedores frequentemente ficam cientes da vulnerabilidade como resultado do ataque.

prompt zero shot

Fornecer a um [LLM](#) instruções para realizar uma tarefa, mas sem exemplos (shots) que possam ajudar a orientá-lo. O LLM deve usar seu conhecimento pré-treinado para lidar com a tarefa. A eficácia dos prompts zero-shot depende da complexidade da tarefa e da qualidade do prompt. Veja também [prompts few-shot](#).

aplicação zumbi

Uma aplicação que tem um uso médio de CPU e memória inferior a 5%. Em um projeto de migração, é comum retirar essas aplicações.

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.