Construir uma arquitetura de alta disponibilidade e recuperação de desastres com métodos nativos e híbridos para bancos de dados Microsoft SQL Server no Amazon EC2

AWS Orientação prescritiva

AWS Orientação prescritiva: Construir uma arquitetura de alta disponibilidade e recuperação de desastres com métodos nativos e híbridos para bancos de dados Microsoft SQL Server no Amazon EC2

Table of Contents

Construir uma arquitetura de alta disponibilidade e recuperação de desastres com métodos nativos e híbridos para bancos de dados Microsoft SQL Server no Amazon EC2

Ram Yellapragada e Alysia Tran, Amazon Web Services (AWS)

Fevereiro de 2022 (histórico do documento)

O Microsoft SQL Server tem muitas opções nativas para oferecer suporte à alta disponibilidade (HA) e à recuperação de desastres (DR), para ajudar a garantir a continuidade dos negócios de suas workloads de banco de dados. Este guia descreve uma configuração ideal para o SQL Server na Amazon Elastic Compute Cloud (Amazon EC2) na Nuvem do Amazon Web Services (AWS). A nova hospedagem do SQL Server no Amazon EC2 fornece um sistema autogerenciado onde você pode reter o controle total sobre as operações e a configuração do banco de dados.

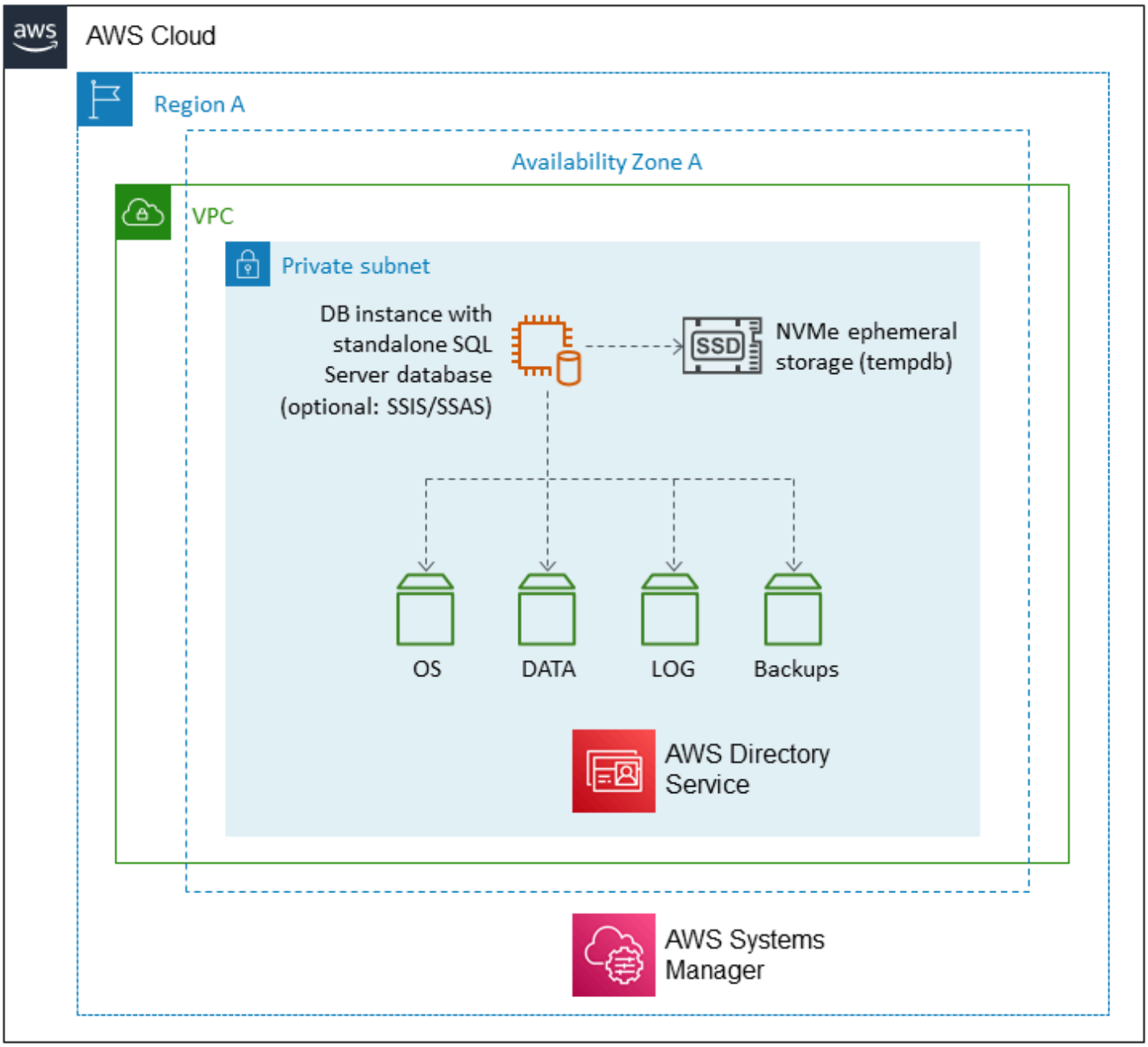
O guia discute as opções híbridas de HA/DR do SQL Server que incluem vários serviços e infraestrutura da AWS e fornece orientação sobre componentes e configurações de infraestrutura, incluindo classes de instância, opções de armazenamento e configuração de HA/DR. Este documento também explica como uma determinada estratégia de HA/DR pode se encaixar em um exemplo de caso de uso que tenha requisitos específicos de objetivo de tempo de recuperação (RTO) e objetivo de ponto de recuperação (RPO) e abrange alguns cenários de recuperação, incluindo diagramas de arquitetura relevantes. Este guia não fornece soluções projetadas para aplicações ou requisitos específicos. Ele apresenta algumas opções de HA/DR com base em RTO e RPO, para que você possa escolher uma arquitetura que atenda às suas necessidades.

Além disso, como um exercício de dimensionamento, o guia define as opções de HA/DR para uma workload típica de processamento de transações on-line (OLTP) do SQL Server e fornece uma comparação lado a lado dessas opções. Para uma discussão sobre como redefinir a hospedagem do SQL Server na AWS, consulte a seção Amazon EC2 para SQL Server no guia Migrar bancos de dados do Microsoft SQL Server para a Nuvem AWS. Para obter informações sobre outras opções de migração, consulte a seção Estratégias de migração de banco de dados do SQL Server nesse guia. Para leitura adicional, consulte a seção Próximas etapas e recursos.

SQL Server na arquitetura de nó único do Amazon EC2

O diagrama a seguir ilustra uma arquitetura recomendada para um SQL Server de nó único no Amazon Elastic Compute Cloud (Amazon EC2) antes de adicionar suporte para alta disponibilidade (HA) e recuperação de desastres (DR).

Nessa arquitetura, o banco de dados SQL Server é implantado em uma instância do EC2, usando uma imagem de máquina da Amazon (AMIs) para SQL Server e volumes separados para OS, DATA, LOG e backups. O armazenamento expresso de memória não volátil (NVMe) é associado diretamente à instância do EC2 e usado para o banco de dados tempdb do SQL Server. AWS Directory Service é usado para configurar a autenticação do Windows para o banco de dados do SQL Server. Você também pode usar AWS Systems Manager para detectar e instalar patches e atualizações do SQL Server.



A tabela a seguir resume as recomendações para configurar essa arquitetura. Essas recomendações são discutidas em detalhes nas seções a seguir.

AMI/Tipo de instância	• Tipo de instância otimizada para Amazon Elastic Block Store (Amazon EBS) para desempenho

	• NVMe para armazenamento de instâncias (temporário) • AMIs do Amazon EC2 para SQL Server
Edição do SQL Server	• Edição do SQL Server Developer (não produção) • Edições Standard e Enterprise do SQL Server (produção)
Tipo de armazenamento	• Amazon EBS • NVMe (tempdb) (gp2/io1/io2)
Volumes	• SO • DATA • LOG • tempdb • Espaço de rascunho para armazenar e baixar backups
Opções de DR	• Amazon EC2 • Snapshots do Amazon EBS • Backups nativos do SQL Server

Tipos de instância

A AWS oferece uma seleção de classes de instância para suas workloads do SQL Server. Você pode escolher entre otimizado para computação, otimizado para memória, otimizado para armazenamento, uso geral e outros tipos, dependendo do workload esperado no servidor de banco de dados, da versão, das opções de HA/DR, dos núcleos necessários e das considerações de licenciamento. Recomendamos que você escolha os tipos de instância otimizados para Amazon EBS para o SQL Server. Eles oferecem o melhor throughput com volumes do EBS anexados em uma rede dedicada, o que é essencial para workloads do SQL Server que podem ter requisitos pesados de acesso a dados. Para workloads de banco de dados padrão, você pode executar classes de instância otimizada para memória, como R5, R5b, R5d e R5n. Você também pode incluir

armazenamento de instância ou armazenamento NVMe. Ambos são ideais para tempdb e oferecem desempenho equilibrado para workloads de banco de dados.

Para workloads críticas, a instância z1d de alto desempenho é otimizada para workloads que acarretam altos custos de licenciamento, como o SQL Server. A instância z1d é criada com um processador escalável Intel Xeon personalizado que fornece uma frequência turbo sustentada de todos os núcleos de até 4,0 GHz, o que é significativamente mais rápido do que outras instâncias. Para workloads que precisam de processamento sequencial mais rápido, você pode executar menos núcleos com uma instância z1d e obter o mesmo desempenho ou melhor do que outras instâncias com mais núcleos.

A Amazon também fornece AMIs dedicadas para o SQL Server no Microsoft Windows Server para ajudá-lo a hospedar as edições mais recentes do SQL Server no Amazon EC2.

Armazenamento

Alguns tipos de instância oferecem volumes de armazenamento de instância NVMe. O NVMe é uma opção de armazenamento temporário (efêmero). Esse armazenamento é anexado diretamente à instância do EC2. Embora o armazenamento NVMe seja temporário e os dados sejam perdidos na reinicialização, ele oferece o melhor desempenho. Portanto, é adequado para o banco de dados tempdb do SQL Server, que tem altos padrões de E/S e acesso a dados aleatório. Não há cobrança adicional pelo uso do em uma instância de um armazenamento de instância NVMe para tempdb. Para obter orientação adicional, consulte a seção Adicionar tempdb em um armazenamento de instâncias no guia Práticas recomendadas para implantação do SQL Server no Amazon EC2.

O Amazon EBS é uma solução de armazenamento durável que atende aos requisitos do SQL Server para armazenamento rápido e disponível. A Microsoft recomenda manter os volumes de dados e logs separados para um desempenho ideal. Os motivos para essa separação incluem:

- Diferentes métodos de acesso aos dados. Os volumes de dados usam acesso aleatório aos dados de processamento de transações on-line (OLTP), enquanto os volumes de log usam acesso serial.
- Opções de recuperação de desastres. A perda de um volume não afeta o outro volume e ajuda na recuperação dos dados.
- Diferentes tipos de workloads. Os volumes de dados são para workloads OLTP, enquanto os volumes de log se destinam aos workloads de processamento analítico on-line (OLAP).
- Diferentes requisitos de desempenho. Os volumes de dados e logs têm diferentes requisitos de IOPS e latência, taxas mínimas de throughput e benchmarks de desempenho semelhantes.

Para selecionar o tipo de volume do Amazon EBS certo, você deve analisar seus métodos de acesso ao banco de dados, IOPS e throughput. Colete métricas durante o horário de trabalho padrão e durante o pico de uso. O SQL Server usa extensões para armazenar dados. A unidade atômica de armazenamento no SQL Server é uma página com tamanho de 8 KB. Oito páginas fisicamente contíguas formam uma extensão de 64 KB. Portanto, em uma máquina SQL Server, o tamanho da unidade de alocação NTFS para hospedar arquivos de banco de dados SQL (incluindo tempdb) deve ser 64 KB. Para obter informações sobre como verificar o tamanho da alocação de NTFS de suas unidades, consulte o guia Práticas recomendadas para implantar o SQL Server no Amazon EC2.

A escolha do volume do EBS depende do workload, ou seja, se o banco de dados exige muita leitura ou gravação, exige alto IOPS, armazenamento de arquivos e considerações semelhantes. A tabela a seguir mostra uma amostra de configuração.

Recurso do Amazon EBS	Tipo	Descrição
Disco OS	`gp3`	Armazenamento de uso geral.
Disco DATA	`io1`/`io2`	Armazenamento intensivo de gravação.
Disco LOG	`gp3` ou `io2`	Armazenamento de uso geral para workloads intensivos.
Disco de backup	`st1`	Armazenamento de arquivos mais barato. Para melhorar o desempenho, os backups também podem ser armazenados em um disco mais rápido se forem copiados regularmente para o Amazon Simple Storage Service (Amazon S3).

Considerações sobre o Amazon EBS e o Amazon S3

A tabela a seguir mostra uma comparação entre o Amazon EBS e o Amazon S3 para armazenamento. Use essas informações para entender as diferenças entre os dois serviços e escolher a melhor abordagem para seu caso de uso.

Serviço	Disponibilidade	Durabilidade	Observações
Amazon EBS	• Todos os tipos de volume EBS oferecem capacidades de snapshots duráveis e foram projetados para disponibilidade de 99,999%. • Você pode usar snapshots para provisionar novas instâncias em diferentes regiões da AWS em caso de desastre.	• Os dados dos volumes do EBS são replicados em vários servidores em uma única zona de disponibilidade para evitar perdas de dados causadas por falha em qualquer componente único. • Os volumes do EBS são projetados para uma taxa de falha anual (AFR) entre 0,1 e 0,2%, em que a falha se refere a uma perda total ou parcial do volume, dependendo do tamanho e do desempenho do volume.	• Uma instância otimizada para o Amazon EBS usa uma pilha de configuração otimizada e fornece largura de banda adicional dedicada para E/S do Amazon EBS. Essa otimização proporciona a melhor performance para seus volumes do EBS ao minimizar a contenção entre a E/S do Amazon EBS e outro tráfego de sua instância. • As restaurações rápidas de snapshots são suportadas para até 50 snapshots ao mesmo tempo. Você deve habilitar esse atributo

Serviço	Disponibilidade	Durabilidade	Observações
			explicitamente por snapshot. • Uma instância otimizada para Amazon EBS oferece desempenho provisionado completo na inicialização, portanto, nenhum tempo de aquecimento está envolvido.

Serviço	Disponibilidade	Durabilidade	Observações

Amazon S3	• Altamente disponível.
	• Projetado para 99,99% de disponibilidade em um determinado ano.
	• Várias classes de armazenamento estão disponíveis, como S3 Standard e S3 Standard — Infrequent Access ((S3 Standard — IA)). Você pode mover arquivos de backup para uma classe de armazenamento com base em um período de retenção.
		• O Amazon S3, o Amazon Glacier e o S3 Glacier Deep Archive foram projetados para oferecer 99,999999999% (11 noves) de durabilidade. As plataformas Amazon S3 e Amazon Glacier oferecem backup de dados confiável, com replicação de objetos em pelo menos três zonas de disponibilidade geograficamente dispersas.	• Você pode usar o Amazon S3 para backups de longo prazo em nível de arquivo do SQL Server (incluindo backups completos e logs de transações).
			• Suporte ao Amazon S3:
			  • Controle do tempo de replicação (RTC)
			  • Replicação entre regiões por meio do gerenciamento do ciclo de vida do S3 e AWS Backup
			  • Classificação inteligente
			• O Amazon S3 fornece o armazenamento mais barato. Aplica-se taxas de transferência de dados entre regiões.

SQL Server no Amazon FSx para Windows File Server

O Amazon FSx para Windows File Server fornece desempenho rápido com taxa de throughput básica de até 2 GB/segundo por sistema de arquivos, centenas de milhares de IOPS e latências consistentes de menos de um milissegundo. Para fornecer o desempenho certo para suas instâncias do SQL Server, você pode escolher um nível de taxa de throughput que seja independente do tamanho do seu sistema de arquivos. Níveis mais altos de capacidade de throughput também vêm com níveis mais altos de IOPS que o servidor de arquivos pode fornecer às instâncias do SQL Server que o acessam. A capacidade de armazenamento determina não apenas a quantidade de dados que você pode armazenar, mas também quantas operações de E/S por segunto (IOPS) você pode realizar no armazenamento - cada GB de armazenamento fornece 3 IOPS. Você pode provisionar cada sistema de arquivos para ter até 64 TiB de tamanho (em comparação com 16 TiB para o Amazon EBS). Você também pode usar os sistemas Amazon FSx como testemunha de compartilhamento de arquivos para implantações do Windows Server Failover Cluster (WSFC).

Opções e considerações de HA/DR

Embora a possibilidade de uma região ou zona de disponibilidade da AWS ficar completamente off-line seja extremamente rara, recomendamos uma abordagem multifacetada para backup e recuperação em caso de desastre para redundância e para minimizar a perda de dados. Os processos de backup e recuperação devem incluir o nível adequado de granularidade para atender ao objetivo de tempo de recuperação (RTO) e ao objetivo de ponto de recuperação (RPO) para o workload e para dar suporte a seus processos de negócios, e muitas vezes dependem no aplicativo. No caso de bancos de dados, a AWS também oferece suporte a todas as recomendações da Microsoft para instalação e configuração do SQL Server para alta disponibilidade e recuperação de desastres (HA/DR). Diferentes edições do SQL Server oferecem suporte a várias opções de HA/DR, e você deve considerar casos especiais, como bancos de dados muito grandes (VLDBs), caso a caso. Como em qualquer configuração de DR, o teste é essencial para garantir que cada aplicativo atenda aos seus Acordos de serviço (SLAs) para HA/DR. Para seu ambiente de teste/desenvolvimento, considere usar a edição SQL Server Developer, que é gratuita, mas vem com limitações.

Para um caso de uso que exige um RPO de 15 minutos e um RTO de 4 horas, você pode considerar uma combinação das seguintes opções de HA/DR:

- Opções nativas de HA/DR do SQL Server com standby quente (nível de banco de dados) — Para obter ilustrações de algumas dessas arquiteturas, consulte a seção Diagramas de arquitetura do SQL Server no Amazon EC2 mais adiante neste guia.
  - Dois nós, Multi-AZ em uma única região (modo de confirmação síncrona) ou em várias regiões (modo de confirmação assíncrona, grupo de disponibilidade básica)
  - Três nós (ou mais), Multi-AZ em várias regiões (modos de confirmação síncrona e confirmação assíncrona)
  - Envio de dois nós, Multi-AZ e logs em várias regiões (com backups de logs a cada 5 minutos)
- Backups nativos do SQL Server no Amazon S3 (somente em nível de banco de dados, DR) — Backups completos (uma vez por dia)
  - Backups diferenciais (a cada 2-4 horas).
  - Backups de logs (a cada 5-10 minutos).

  - Os backups precisam ser feitos e copiados para o Amazon Simple Storage Service (Amazon S3) usando scripts personalizados ou uma opção como um gateway de arquivos para backup e transferência eficientes.
  - Se você tiver centenas de bancos de dados, poderá continuar usando suas ferramentas de backup existentes (como Commvault ou Litespeed) para gerenciar backups com eficiência e armazená-los diretamente no Amazon S3.
  - Use o Amazon S3 Cross-Region Replication (CRR) com o S3 Replication Time Control (RTC) para controlar e monitorar a replicação de objetos dentro de um SLA de 15 minutos.
  - Para fins de conformidade e economia de custos, você também pode usar o gerenciamento do ciclo de vida do S3 para mover e armazenar backups antigos para armazenamento de longo prazo.
  - Se você fizer backups nativos do SQL Server e movê-los para o Amazon S3 regularmente, no caso de um desastre, os backups estarão prontamente disponíveis na região de destino. Isso elimina a necessidade de transferir backups ou restaurar snapshots.
  - Recomendamos usar o SQL Native Backup Compression para reduzir o tamanho dos arquivos.
- Snapshots da AWS (nível de instância e volume, somente DR)
  - Backups de imagem de máquina da Amazon (AMI) do Amazon Elastic Compute Cloud (Amazon EC2) para reconstruir bancos de dados do zero
  - Snapshots de volumes do Amazon Elastic Block Store (Amazon EBS) para anexar volumes do EBS ao Amazon EC2

Gerenciando recursos de HA/DR no AWS Backup

O AWS Backup é um serviço totalmente gerenciado que oferece a capacidade de criar planos e programações de backup e atribuir recursos da AWS envolvidos na configuração de HA/DR, como volumes do Amazon EBS para criar snapshots e AMIs do Amazon EC2, a esses planos de backup. Você também pode usar o AWS Backup para programar cópias multirregionais desses snapshots do EBS. Para um uso ideal, o AWS Backup exige um mecanismo de marcação eficiente para que os recursos estejam disponíveis. O AWS Backup também oferece suporte para backups consistentes com aplicativos por meio do Windows Volume Shadow Copy Service (VSS), que você pode usar para o SQL Server. Para proteção a nível de armazenamento, recomendamos o uso de snapshots do EBS. Os snapshots iniciais do EBS são completos, e os snapshots subsequentes são incrementais. Embora os snapshots do EBS ofereçam proteção a nível de armazenamento, eles não

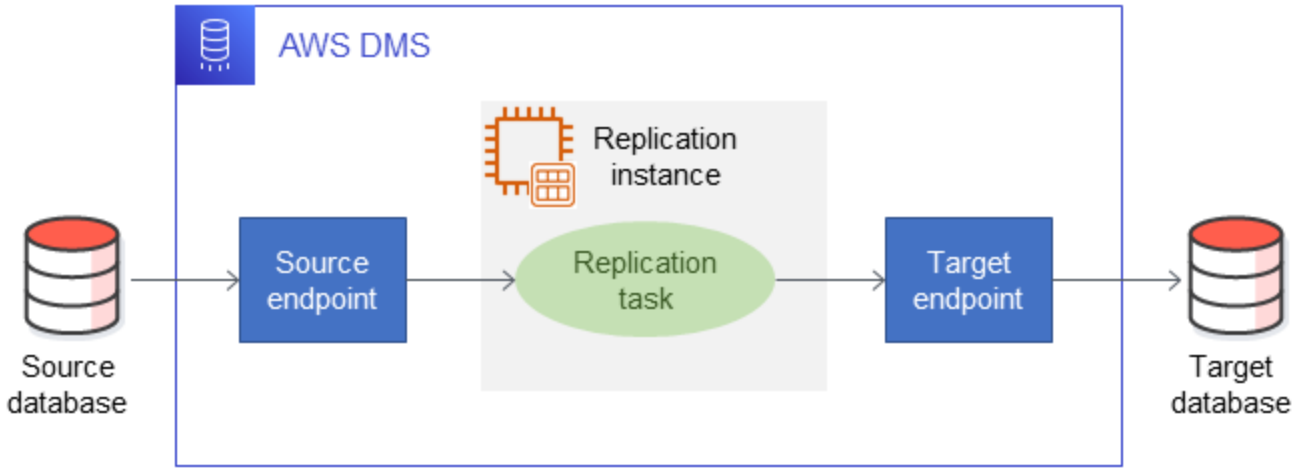
substituem os backups nativos baseados em arquivos do SQL Server que oferecem recuperação pontual.

Usando o AWS DMS para HA/DR

Se você estiver procurando uma alternativa às opções Always On do SQL Server para replicação ou se tiver bancos de dados heterogêneos de origem e destino, em uma configuração híbrida ou na AWS, você pode usar o AWS Database Migration Service (AWS DMS) das seguintes maneiras.

Se você usa o AWS DMS com o SQL Server em um contexto autogerenciado (hospedado no Amazon EC2 ou on-premises), ele suporta replicação única e contínua em dois modos: usando MS-REPLICATION (para capturar alterações em tabelas que têm chaves primárias) e MS-CDC (para capturar alterações em tabelas que não têm chaves primárias). No entanto, se você usar o Amazon Relational Database Service (Amazon RDS) como fonte para o AWS DMS, somente o MS-CDC é suportado. O AWS DMS oferece uma variedade de endpoints de origem e destino, oferece suporte a mecanismos de banco de dados heterogêneos e oferece controle refinado sobre o processo de replicação. Você também pode usar o AWS Schema Conversion Tool (AWS SCT) com AWS DMS para migrações heterogêneas de bancos de dados. O AWS SCT automatiza as mudanças no nível do esquema e também produz relatórios para preparação e planejamento da migração.

Você adiciona bancos de dados de origem e destino como endpoints no AWS DMS, conforme ilustrado no diagrama a seguir. Esse serviço implementa um processo de replicação lógica usando o MS-REPLICATION ou o MS-CDC. Se você tiver uma configuração híbrida, poderá configurar o AWS DMS para replicação contínua entre on-premises e a AWS. Durante a substituição, a tarefa de migração do AWS DMS pode ser interrompida e o aplicativo poderá se conectar ao banco de dados que já está sincronizado com o banco de dados on-premises sem mais demoras. O uso do AWS DMS do SQL Server como fonte tem algumas limitações, descritas na documentação do AWS DMS.

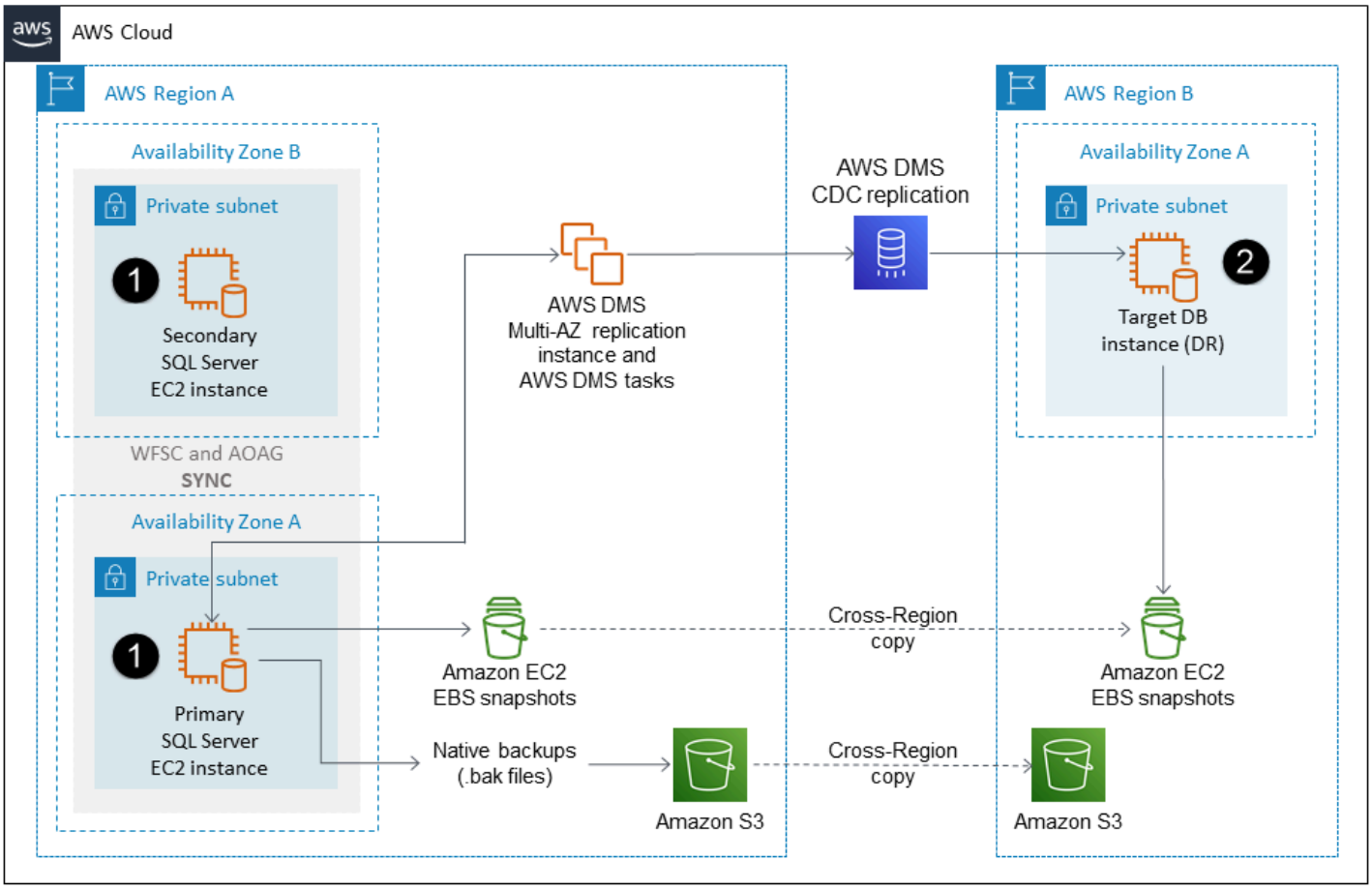


Considere usar o AWS DMS em vez de métodos nativos de HA/DR nos seguintes cenários:

- Quando você quiser economizar nos custos de licenciamento. Por exemplo, se você estiver usando uma versão avançada, como a edição SQL Server Enterprise somente para as opções Always On, considere configurar o AWS DMS, pois pode fornecer uma opção de replicação lógica sem o custo de uma licença da edição Enterprise.
- Quando você tem origens e destinos heterogêneos. As versões do SQL Server nos nós primários e de recuperação de desastres não precisam ser iguais (dentro das limitações do AWS DMS), o que oferece flexibilidade significativa.
- Para evitar a sobrecarga do Windows, do cluster do SQL Server e da configuração e gerenciamento de grupos de disponibilidade distribuídos. O AWS DMS oferece uma configuração simples e um gerenciamento fácil das tarefas de replicação.
- Para casos de uso comercial, como transferência quase em tempo real (dependendo da instância de replicação, da configuração da rede e do volume de dados), mascaramento de dados, filtragem seletiva, mapeamento de esquema/tabela (homogêneo e heterogêneo), avaliações de pré-migração e suporte a JSON.
- Para duplicar, interromper e iniciar tarefas com facilidade, conforme necessário, com base em números de sequência de log (LSNs), timestamps e opções semelhantes.

O diagrama a seguir mostra uma abordagem alternativa de como o AWS DMS pode fornecer suporte à replicação. Nessa configuração, a origem é um cluster de grupos de disponibilidade Always On do SQL Server, e o AWS DMS usa a opção de captura de dados de alteração (CDC) para replicar dados continuamente em um destino em uma região da AWS diferente. Para obter o

melhor desempenho, é fundamental garantir que a instância de replicação tenha o tamanho certo e permaneça na região de origem.



Os mecanismos de origem e de destino não precisam ser iguais. No diagrama, os nós primário e secundário marcados como (1) podem ser um cluster do SQL Server em uma configuração Single-AZ ou Multi-AZ. Ou a origem pode ser um único nó do SQL Server que ofereça suporte ao MS-CDC ou ao MS-REPLICATION.

A instância de banco de dados de destino, marcada como (2) no diagrama, pode ser qualquer versão do SQL Server no Amazon RDS, no Amazon EC2 ou em qualquer outro destino heterogêneo. Ele não precisa corresponder às instâncias primária e secundária nem oferecer suporte aos grupos de disponibilidade do Always On. Por exemplo, a origem pode ser um cluster de grupos de disponibilidade do SQL Server Always On e o destino pode ser uma edição compatível do Amazon Aurora PostgreSQL.

Usando o AWS Application Migration Service para DR

Recomendamos usar o AWS Application Migration Service para migrações mover sem alterações (lift-and-shift) para a AWS. O Application Migration Service replica continuamente suas máquinas (incluindo o sistema operacional, a configuração do estado do sistema, bancos de dados, aplicativos e arquivos) em uma área de armazenamento de baixo custo em sua conta da AWS de destino e região preferida. No caso de um desastre, você pode usar o Application Migration Service para iniciar automaticamente milhares de suas máquinas em seu estado totalmente provisionado em minutos.

Considerações adicionais

A lista a seguir identifica os possíveis gargalos que você deve considerar ao projetar uma estratégia de HA/DR.

- Largura de banda, latência, complexidade da rede e conectividade em uma configuração de nó multirregional.
- Tamanho dos snapshots do Amazon EBS ou do Amazon EC2 e o tempo necessário para copiá-los usando o AWS Backup.
  - Os snapshots do Amazon EBS e do Amazon EC2 são armazenados no Amazon S3 usando o AWS Backup.
  - Um snapshot do EBS não é replicado para a região de destino no Amazon S3 até que o snapshot atual seja concluído. A duração da replicação também depende do tamanho do volume.
  - Quando o snapshot estiver concluído, o tempo de cópia dos snapshots pode ser de apenas 15 minutos para 99,99% dos objetos. No entanto, testes completos são necessários para casos de uso específicos e grandes volumes críticos.
- Tempo necessário para restaurar os volumes do EBS na zona e região de disponibilidade de destino.
- Tempo necessário para restaurar as imagens do Amazon EC2 na zona e região de disponibilidade de destino.
- Se estiver construindo do zero, será necessário tempo para provisionar a infraestrutura para a imagem do Amazon EC2 ou restaurar os snapshots do EBS na zona de disponibilidade e região de destino.
- Se estiver restaurando do zero, será necessário tempo para restaurar backups completos, diferenciais e de log nativos do SQL Server na zona e região de disponibilidade de destino.

- Dependências externas e aplicativos que precisam estar disponíveis em todas as regiões.
- Limitações nos tamanhos dos arquivos para volumes e para upload para o Amazon S3.

Cenários de recuperação de desastres

Esta seção fornece exemplos de uma única falha na zona de disponibilidade ou região da AWS e discute as opções para recuperação de desastres (DR, disaster recovery). Os exemplos pressupõem um objetivo de ponto de recuperação (RPO) de 15 minutos e um objetivo de tempo de recuperação (RTO) de 4 horas.

Falha na zona de disponibilidade

Você pode usar uma das seguintes opções para se recuperar de uma única falha na zona de disponibilidade dentro dos parâmetros fornecidos (RPO de 15 minutos, RTO de 4 horas).

- Provisione a recuperação do aplicativo usando o backup de imagem mais recente do Amazon Elastic Compute Cloud (Amazon EC2) e conecte-se à instância de banco de dados standby quente existente por meio de uma implantação de grupo de disponibilidade Always On ou envio de logs.
  - Uma configuração de grupo de disponibilidade do SQL Server Always On para DR com dois ou mais nós fornece failover automático para o nó secundário por meio do modo de confirmação síncrona ou assíncrona, para que o banco de dados esteja disponível imediatamente. Para uma configuração de HA, os dois nós estão disponíveis para operações de leitura. Essa opção atende confortavelmente aos requisitos de RTO e RPO. Na edição SQL Server Standard, usar grupos de disponibilidade básicos também é uma opção, mas está limitado a dois nós, porque um grupo de disponibilidade pode incluir somente um banco de dados. No entanto, você pode configurar vários grupos de disponibilidade em uma região ou entre regiões. Essa configuração proporciona economia de custos, pois não há custo adicional para o nó secundário, que não é acessível para operações de leitura. A edição SQL Server Enterprise fornece funcionalidade completa e failover para todos os bancos de dados em um único grupo de disponibilidade. Para obter exemplos dessa opção, consulte os seguintes diagramas de arquitetura:
    - Arquitetura de HA/DR de dois nós com cluster de grupos de disponibilidade Always On (região única, multi-AZ)
    - Arquitetura de HA/DR de três nós (região única, Multi-AZ)
    - Arquitetura de HA/DR de quatro nós com cluster de grupos de disponibilidade Always On distribuída (múltiplas regiões, multi-AZ)
    - Arquitetura de HA/DR de três nós com um único grupo de disponibilidade (múltiplas reigões)

  - O envio de registros do SQL Server como uma solução de DR requer um failover manual para um servidor em espera e depende da frequência dos backups de log. Essa é uma das opções de DR mais baratas. As edições do SQL Server para o local de DR primário e enviado por log não precisam corresponder. Essa opção atende ao RPO (usando backups de logs de transações a cada 5 minutos) e ao RTO, mas requer manutenção por meio de scripts manuais e personalizados. Para obter um exemplo dessa opção, consulte os seguintes diagramas de arquitetura:
    - Arquitetura de HA/DR de três nós com envio de log (múltiplas regiões)
- Se você tiver um aplicativo como o SQL Server Reporting Services (SSRS) que tenha uma implantação em escala horizontal, o balanceador de carga poderá redirecionar todo o tráfego para o nó secundário.
- Você pode usar AMIs baseadas no Amazon EC2 para o servidor de aplicativos e bancos de dados para provisionar a infraestrutura. Os bancos de dados podem ser restaurados em uma nova zona de disponibilidade, dependendo do tamanho e da frequência de backup, a partir dos backups nativos mais recentes (backup completo, backup diferencial ou backups de logs de transações a cada 5 minutos) ou usando snapshots do EBS. Essa opção atende aos requisitos de RPO e RTO, mas requer scripts personalizados. Você também deve considerar o tempo necessário para provisionar a infraestrutura, e atender aos requisitos de RPO e RTO pode ser um desafio.
- As imagens do Amazon EC2 (incluindo volumes do EBS) para aplicativos e para o servidor de banco de dados podem ser restauradas em uma nova zona de disponibilidade. O RPO pode ser desafiador, dependendo do backup mais recente, mas essa opção pode ser combinada com os logs de transações mais recentes para atender aos requisitos. Essa opção oferece suporte a snapshots do Windows Volume Shadow Copy Service (VSS).

Falha de região

Você pode usar uma das seguintes opções para se recuperar de uma única falha na Região da AWS dentro dos parâmetros fornecidos (RPO de 15 minutos, RTO de 4 horas).

- Você pode usar imagens de máquina da Amazon (AMIs) baseadas no Amazon EC2 para o servidor de aplicativos e bancos de dados para provisionar a infraestrutura. Os bancos de dados podem ser restaurados em uma nova Região, dependendo do tamanho e da frequência de backup, a partir dos backups nativos mais recentes (backup completo, backup diferencial ou backups de logs de transações a cada 5 minutos). Essa opção atende aos requisitos de RPO e RTO, mas requer scripts personalizados.

  - O envio de registros do SQL Server como uma solução de DR requer um failover manual para um servidor em espera e depende da frequência dos backups de log. Essa é uma das opções de DR mais baratas. As edições do SQL Server para o local de DR primário e enviado por log não precisam corresponder. Essa opção atende ao RPO (usando backups de logs de transações a cada 5 minutos) e ao RTO, mas requer manutenção por meio de scripts manuais e personalizados. Bancos de dados grandes exigem tempos de restauração longos.
- Você pode usar uma AMI do Amazon EC2 para o aplicativo e o servidor de banco de dados e restaurá-la para um destino em uma nova região. O RPO depende do tamanho e da frequência dos backups.
  - As imagens mais recentes do aplicativo podem ser restauradas usando uma AMI. Você pode usar backups diferenciais nativos recentes ou de logs de transações a cada 5 minutos para atualizar o banco de dados para atender ao RPO.
  - O RTO depende do tamanho e do tempo para transferir e restaurar os snapshots para a nova região, se a origem ainda não estiver sincronizada com o destino.
- A solução com o menor tempo de inatividade é restaurar a imagem de backup do aplicativo e ter um nó SQL Server em standby quente em uma região remota usando uma configuração de grupo de disponibilidade de dois, três ou quatro nós (básica, clássica ou distribuída) e conectar-se ao servidor de banco de dados em espera após um failover. A réplica do modo de confirmação síncrona atende aos requisitos de RPO, enquanto a réplica do modo de confirmação assíncrona pode ser adiada dependendo do volume de transações. Você pode usar uma configuração de grupo de disponibilidade distribuído para aumentar a escala horizontalmente os nós do banco de dados em uma nova região, se necessário. Essa configuração também reduz a complexidade porque usa dois grupos de disponibilidade independentes em vez de um único grupo de disponibilidade distribuído pelas regiões no modo de confirmação síncrona ou assíncrona e atende confortavelmente aos requisitos de RTO e RPO. Como alternativa, usar grupos de disponibilidade básica do SQL Server na edição Standard também é uma opção. No entanto, ele tem limitações, porque suporta somente até dois nós, e somente um banco de dados pode estar em um único grupo de disponibilidade, embora vários grupos de disponibilidade sejam suportados. Você pode configurar a edição SQL Server Standard em uma região ou entre regiões. Esta edição oferece economia de custos porque não cobra pelo nó secundário, que não é acessível para operações de leitura. A edição SQL Server Enterprise fornece funcionalidade completa e é compatível com o failover de todos os bancos de dados em um único failover de grupo de disponibilidade.

Casos de uso comuns

Como um exercício de dimensionamento, 80% dos aplicativos do SQL Server executados no Amazon EC2 que têm uma workload normal de processamento de transações on-line (OLTP) podem ser agrupados em uma das três categorias com base em sua importância:

- SQL Server HA/DR com backups do SQL Server, usando duas réplicas de confirmação síncrona e uma réplica em modo de confirmação assíncrona
- AWS Backup HA/DR com backups do SQL Server, usando uma AMI do Amazon EC2 para o aplicativo e o banco de dados, e armazenamento Amazon EBS
- AWS Backup HA/DR com backups do SQL Server, usando uma AMI do Amazon EC2 para o servidor do banco de dados, uma imagem do Amazon EC2 para o aplicativo e snapshots do Amazon EBS

A tabela a seguir fornece detalhes sobre cada categoria.

	HA/DR do SQL Server com backups do SQL Server	AWS Backup HA/DR com AMIs, armazenamento EBS e backups do SQL Server	AWS Backup HA/DR com AMIs, snapshots EBS e backups do SQL Server
Processo de restauração em caso de desastre	• Restaure a AMI base do Amazon EC2 para o aplicativo a partir do AWS Backup • Failover para a instância em espera na região (no caso de falha na zona de disponibilidade) ou para a instância entre regiões (no	• Restaure imagens do Amazon EC2 a partir de backups do aplicativo e do banco de dados • Fornece suporte em regiões e entre regiões • Aplique os backups diferenciais e de log de transações mais recentes do SQL Server (a	• Restaure a imagem do Amazon EC2 a partir do backup para o aplicativo • Restaure a AMI base do Amazon EC2 para o servidor do banco de dados • Restaure snapshots do EBS (se houver) • O cluster precisa ser reconstruído

	HA/DR do SQL Server com backups do SQL Server	AWS Backup HA/DR com AMIs, armazenamento EBS e backups do SQL Server	AWS Backup HA/DR com AMIs, snapshots EBS e backups do SQL Server
	caso de falha na região) • Atende aos requisitos de RPO e RTO	cada 15 minutos) para atender aos requisitos de RPO e RTO do banco de dados	• Fornece suporte em regiões e entre regiões • Aplique os backups diferenciais e de log de transações mais recentes ao banco de dados para atender aos requisitos de RPO, mas pode não atender o RTO

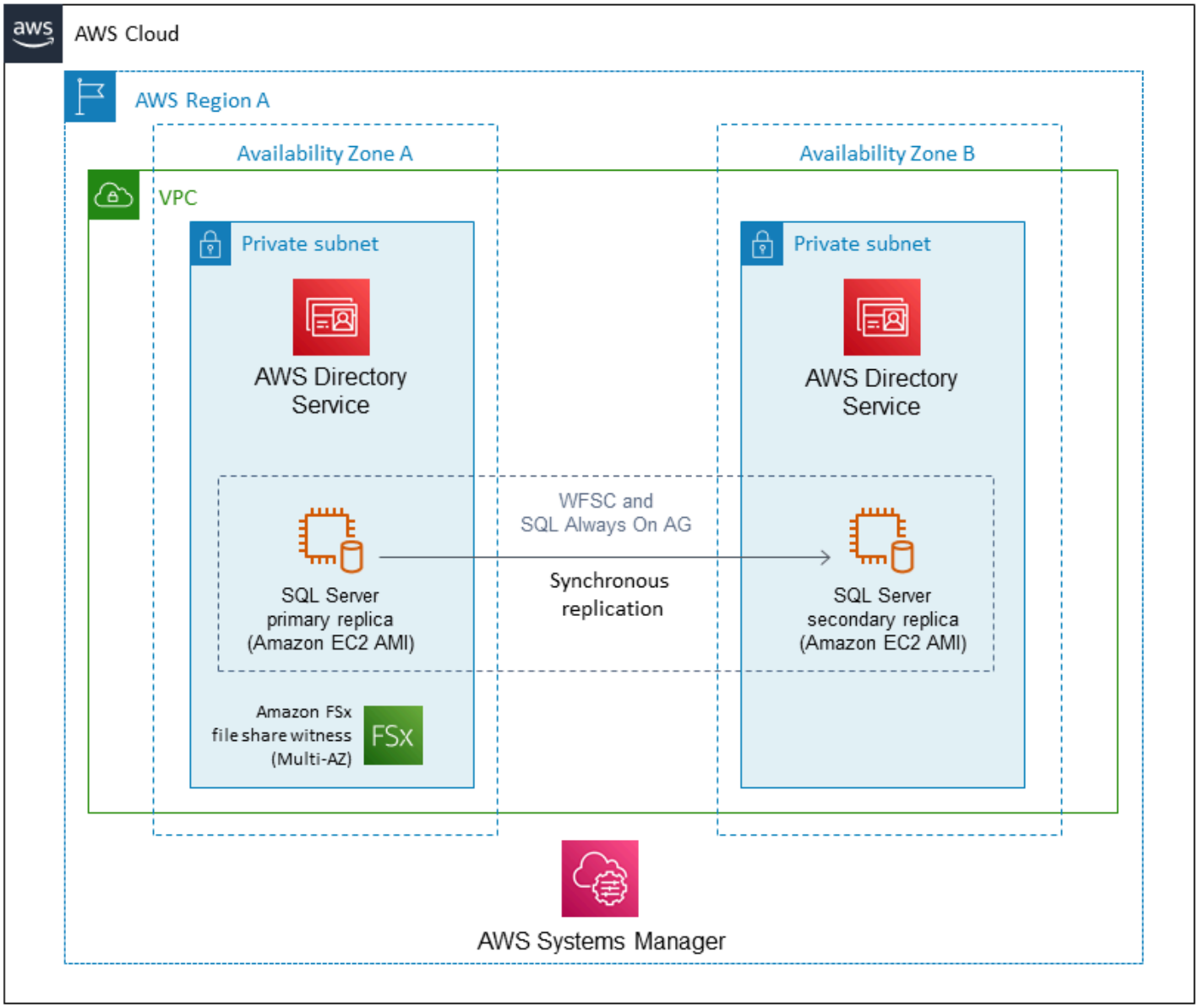
	HA/DR do SQL Server com backups do SQL Server	AWS Backup HA/DR com AMIs, armazenamento EBS e backups do SQL Server	AWS Backup HA/DR com AMIs, snapshots EBS e backups do SQL Server
Recursos primários	• Três licenças da edição SQL Server Enterprise (a licença passiva de nós HA e DR é gratuita se você tiver um contrato de licenciamento de Software Assurance existente com a Microsoft; veja o anúncio) • Espaço de backup do Amazon EC2 no Amazon Simple Storage Service (Amazon S3) • Transferência de dados entre regiões	• Uma licença do SQL Server (qualquer edição). • Espaço de backup do Amazon EC2 no Amazon S3 • Backups do SQL Server (arquivos diferenciais e de log) no Amazon S3 • Transferência de dados entre regiões	• Uma licença do SQL Server (qualquer edição). • Espaço de backup do Amazon EC2 no Amazon S3 • Backups do SQL Server (arquivos diferenciais e de log) no Amazon S3 • Transferência de dados entre regiões
HA/DR	Oferece HA e DR	Oferece somente DR	Oferece somente DR
RPO	O failover é tratado pelo grupo de disponibilidade do SQL Server (o DR é manual)	Manual ou com script personalizado	Manual ou com script personalizado
RTO	Segundos para minutos	Minutos para horas	Várias horas

	HA/DR do SQL Server com backups do SQL Server	AWS Backup HA/DR com AMIs, armazenamento EBS e backups do SQL Server	AWS Backup HA/DR com AMIs, snapshots EBS e backups do SQL Server
Risco de perder SLAs	Baixo	Médio	Alto
Capacidade de gerenciamento	Simples	Médio	Médio
Escalabilidade	Simples	Médio	Médio
Limitações de tamanho de arquivo para uploads para o Amazon S3 ou transferência entre regiões	N/A — Manipulado no modo de confirmação síncrona ou no modo de confirmação assíncrona para um standby quente	Sim	Sim
Perda de dados	Quase zero (depende da workload e da infraestrutura provisionada)	Depende da frequência das imagens de backup do Amazon EC2 e dos backups do SQL Server	Depende da frequência das imagens de backup do Amazon EC2 ou dos snapshots de EBS e dos backups do SQL Server
Custo	Médio	Baixo - médio	Baixo - médio

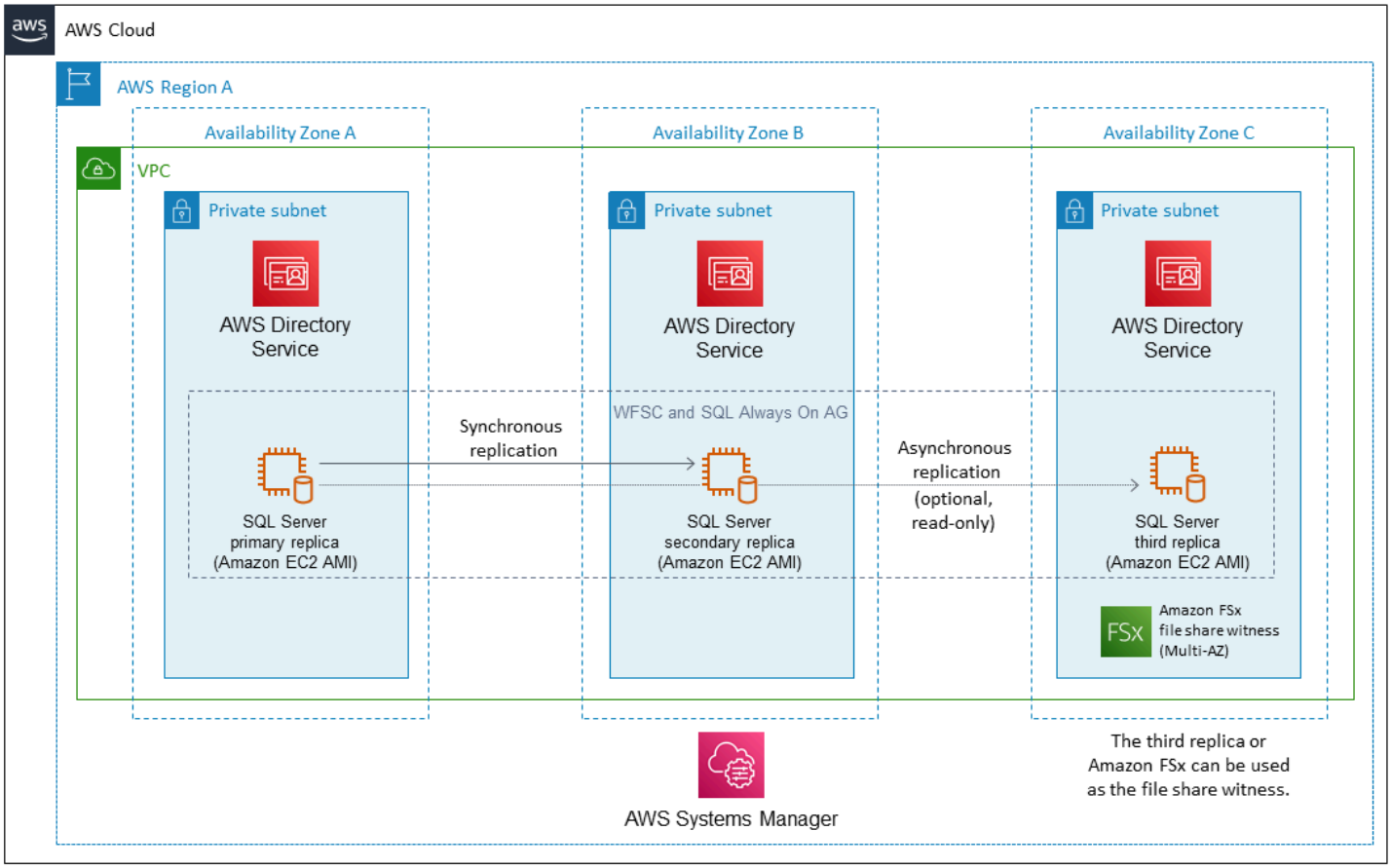
Diagramas de arquitetura do SQL Server no Amazon EC2

Esta seção fornece diagramas de arquitetura que ilustram as estratégias de HA/DR descritas nas seções anteriores.

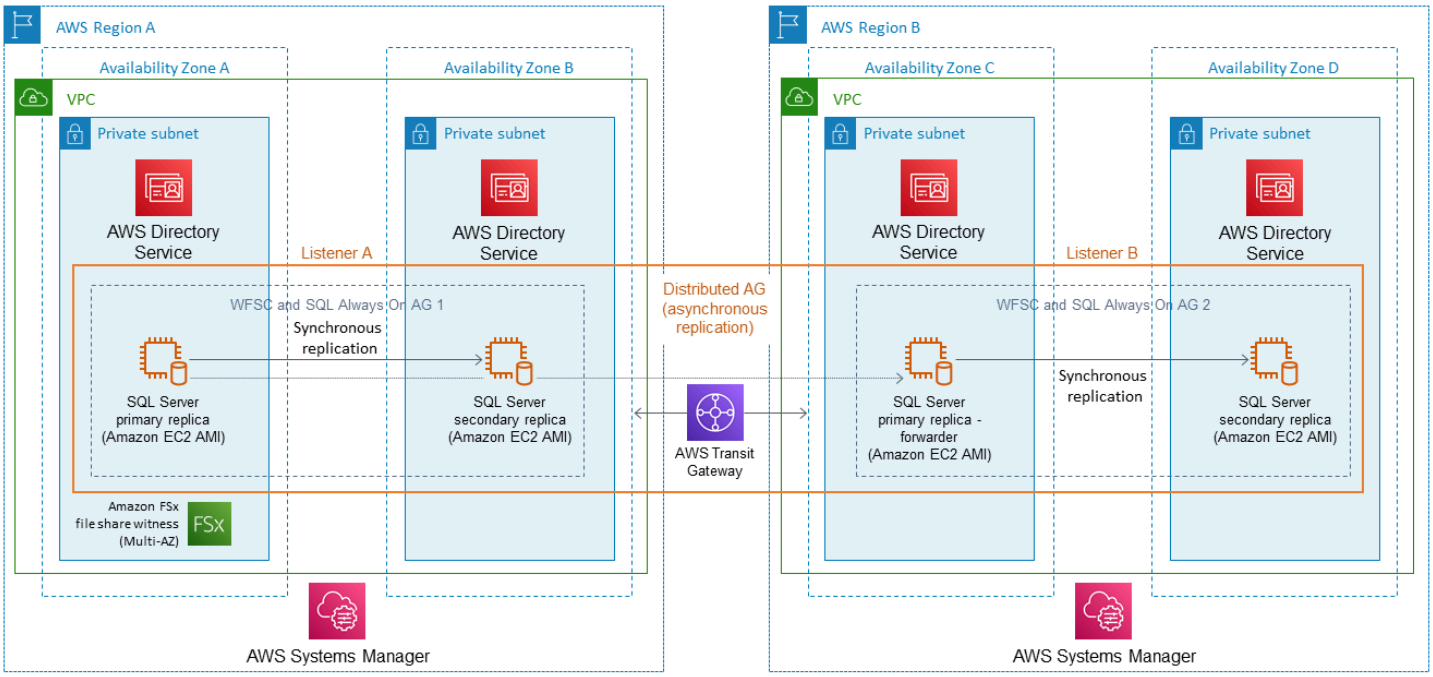
Arquitetura de HA/DR de dois nós com cluster de grupos de disponibilidade Always On (região única, multi-AZ)



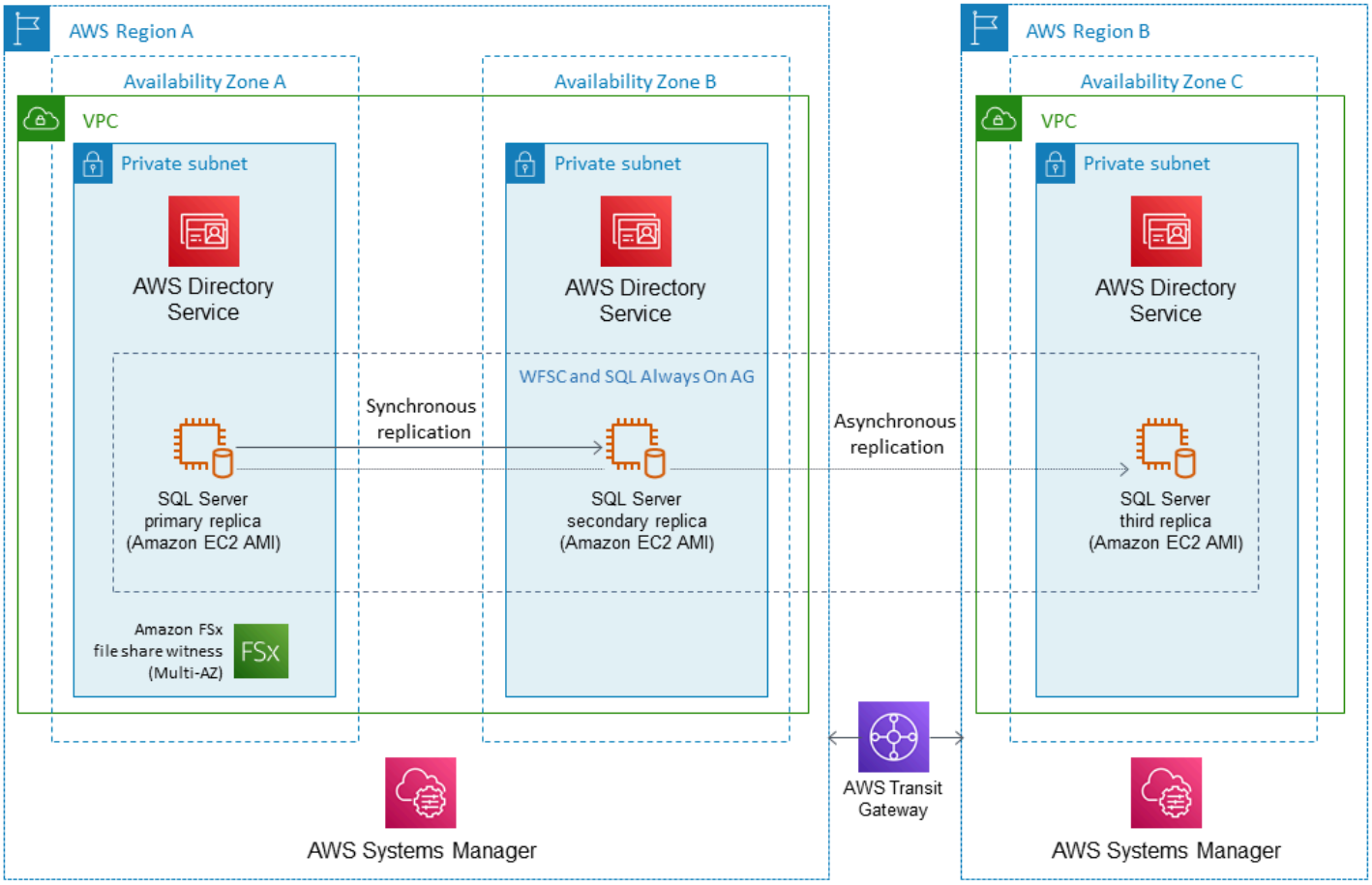
Arquitetura de três nós HA/DR (região única, Multi-AZ)



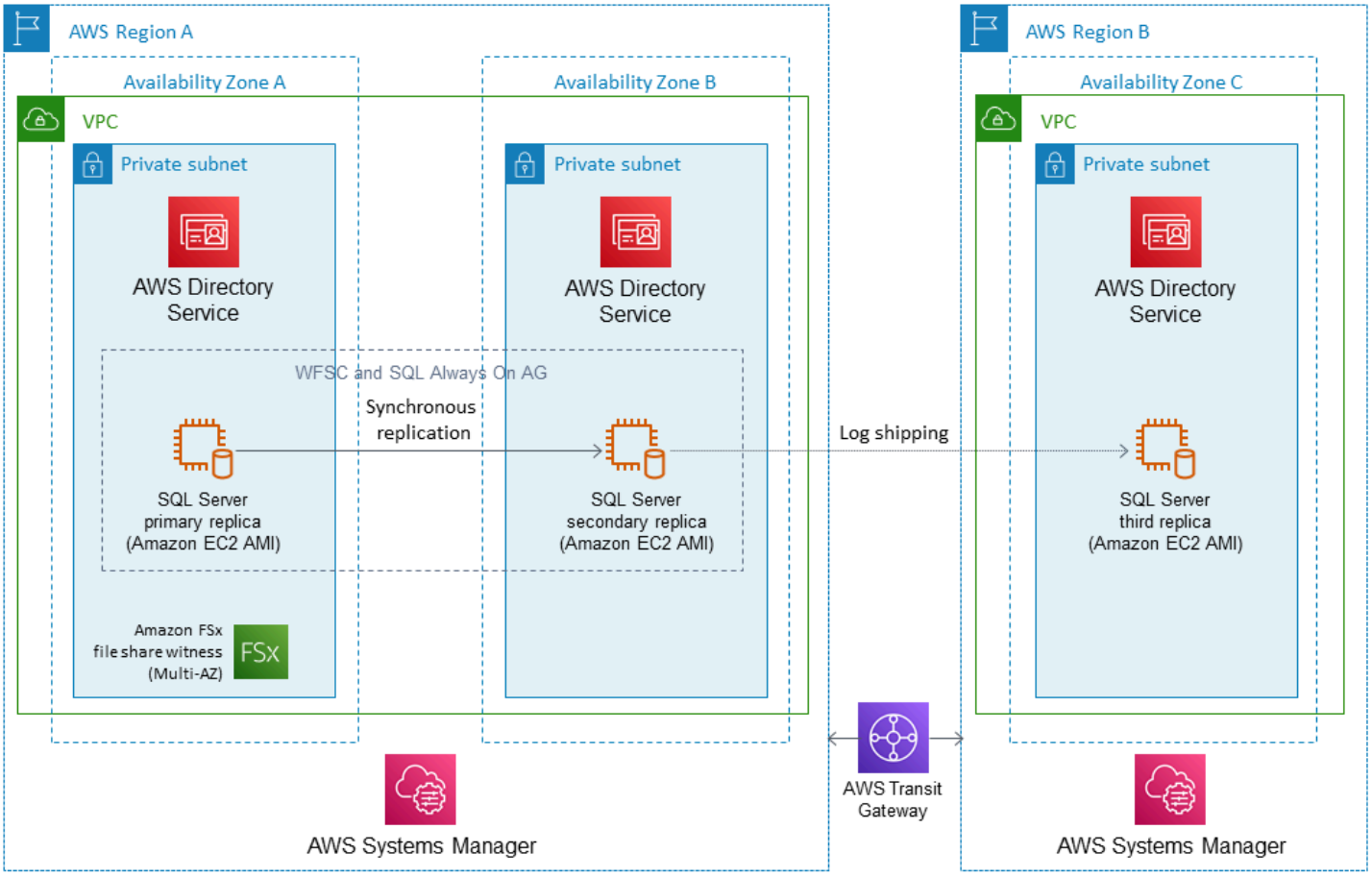
Arquitetura de HA/DR de quatro nós com cluster de grupos de disponibilidade Always On (multirregião, multi-AZ)



Arquitetura de HA/DR de três nós com grupo único de disponibilidade Always On (multirregião)



Arquitetura HA/DR de três nós com log de envio (multirregião)

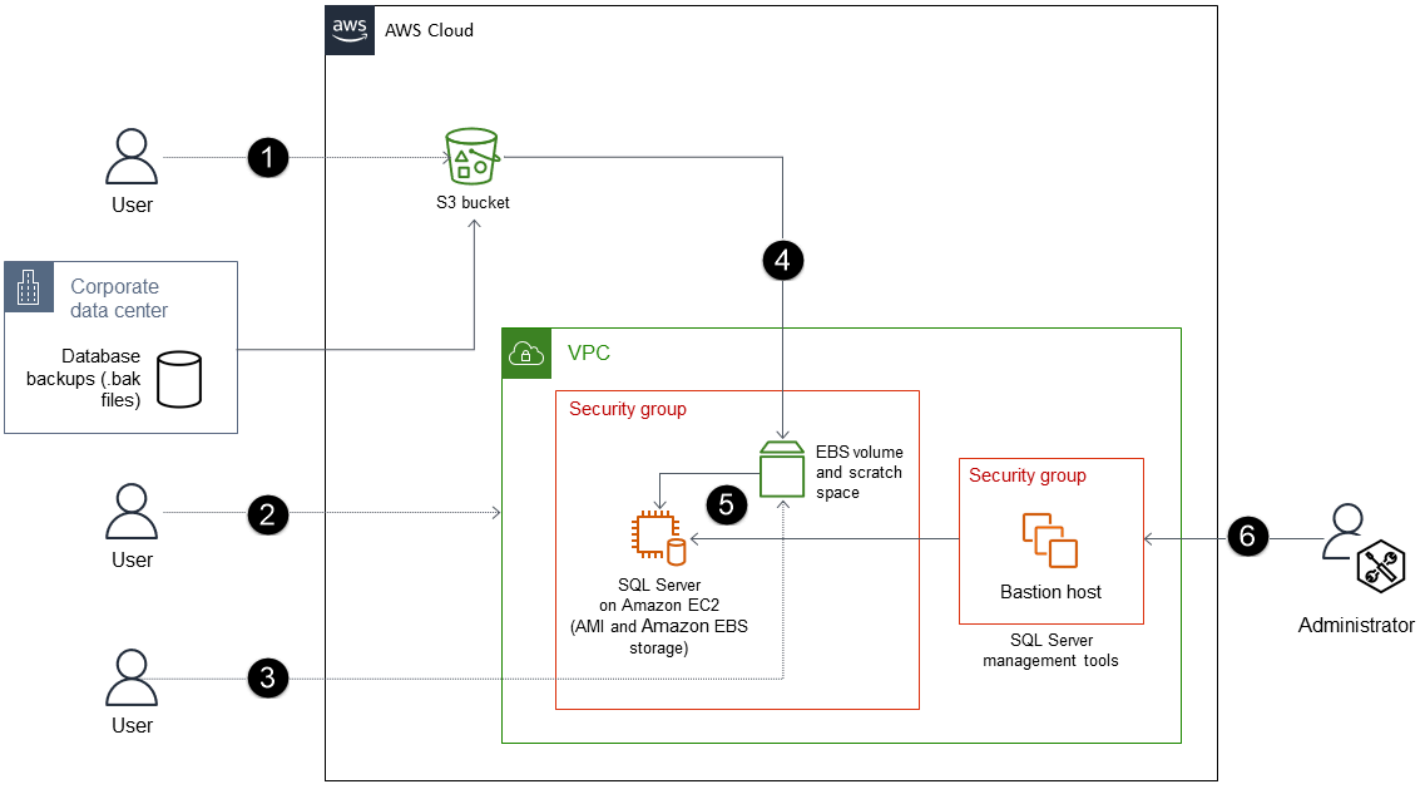


Opções de restauração

As seções a seguir fornecem duas opções de restauração de banco de dados para o SQL Server no Amazon Elastic Compute Cloud (Amazon EC2), quando seus backups estão on-premises.

Usar o Amazon S3

Essa abordagem de restauração de banco de dados do SQL Server usa os comandos do Amazon Simple Storage Service (Amazon S3) para o AWS Command Line Interface (AWS CLI) ou a API do Amazon S3 para fazer upload dos arquivos de backup diretamente para um bucket do S3.



O processo consiste nestas etapas:

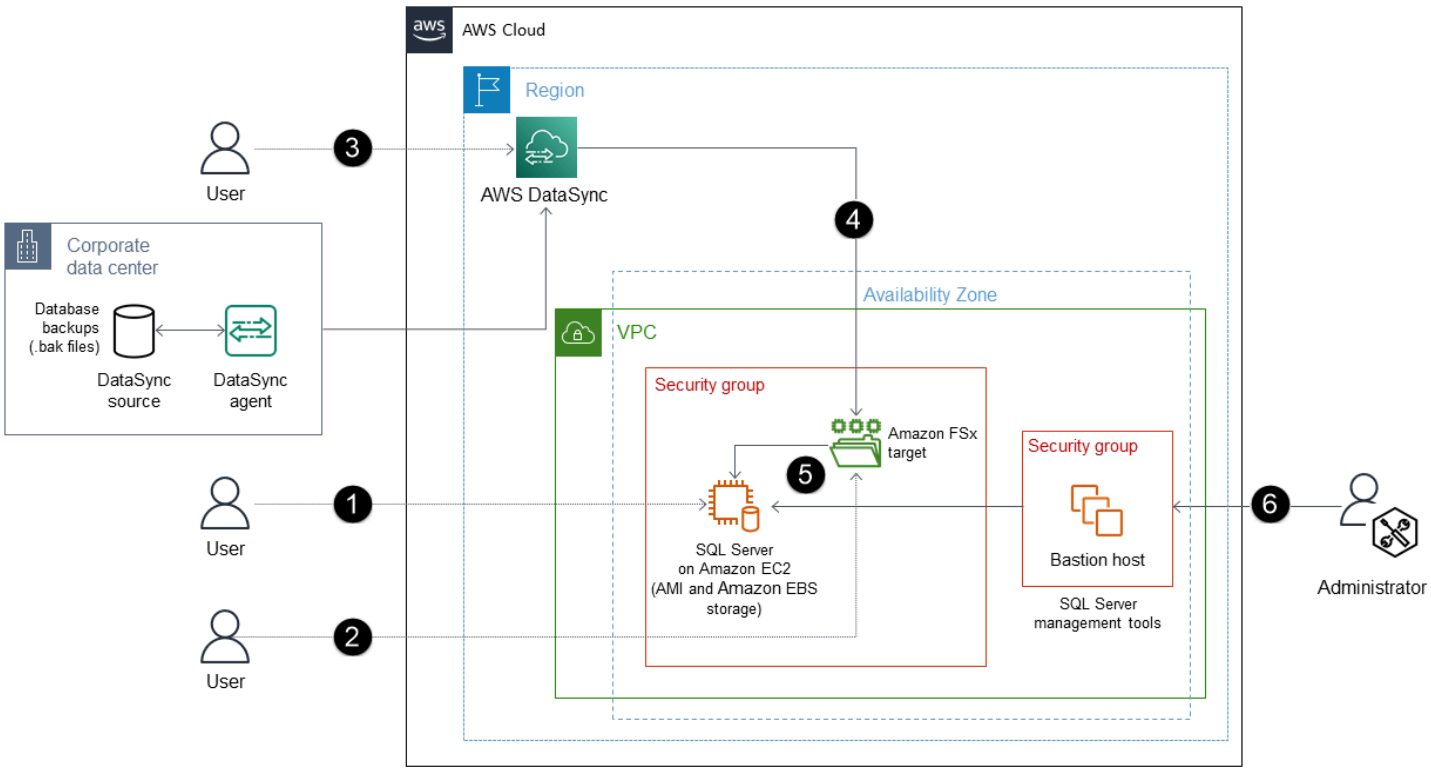
1. Crie um bucket do S3 (ou use um bucket existente) para armazenar os arquivos de backup e transfira os arquivos de backup (.bak) do seu banco de dados on-premises para o bucket do S3 usando a CLI da AWS ou a API do Amazon S3.
2. Implante o SQL Server em uma instância EC2 otimizada para EBS, usando uma imagem de máquina da Amazon (AMI) do SQL Server. Essa AMI deve conter volumes do EBS configurados

com uma partição OS, uma partição DATA, uma partição LOG, armazenamento tempdb (NVMe) e espaço de trabalho.

3. (Opcional) Anexe um volume do EBS não raiz à instância do EC2.
4. Copie os arquivos de backup para o volume do EBS não raiz.
5. Restaure os arquivos de backup do volume do EBS para o SQL Server na instância EC2.
6. Use as ferramentas de gerenciamento do SQL Server para gerenciar seu banco de dados.

Usar AWS DataSync e Amazon FSx

Essa abordagem de restauração de banco de dados do SQL Server usa AWS DataSync para transferir os arquivos de backup para o Amazon FSx para Windows File Server.



O processo consiste nestas etapas:

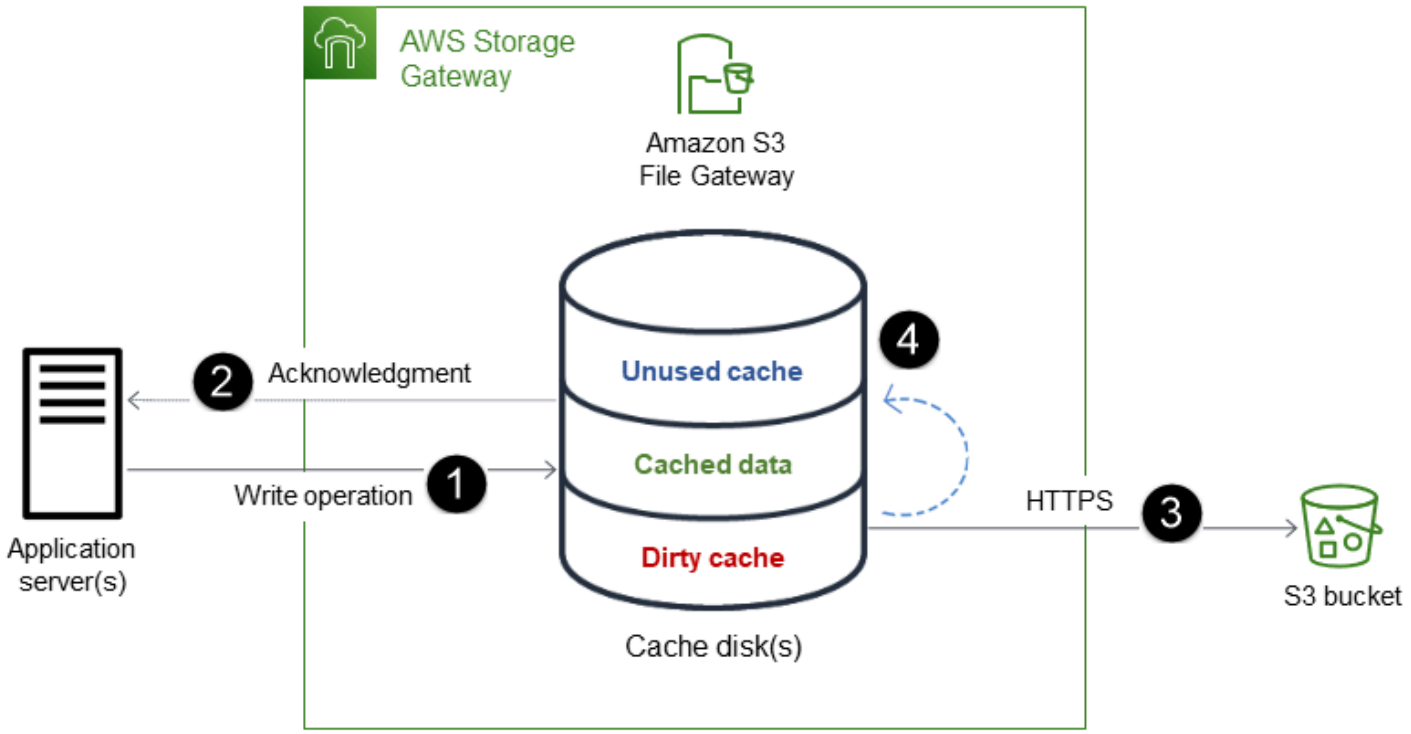
1. Implante o SQL Server em uma instância EC2 otimizada para EBS com NVMe anexado, usando uma AMI que contém volumes do EBS configurados com OS, DATA, LOG e tempdb. (Por exemplo, você pode usar a classe de instância `r5d.large` otimizada para memória.)

2. Usar o FSx for Windows File Server para criar um servidor de arquivos. Isso pode ser usado como um local de armazenamento temporário para baixar arquivos de backup do SQL Server (.bak) do seu ambiente on-premises.
3. Crie um endpoint e um agente do DataSync para o servidor de arquivos Amazon FSx.
4. O DataSync automatiza a sincronização de dados entre seu armazenamento on-premises e o servidor de arquivos Amazon FSx sem exigir o Amazon S3.
5. Restaure os arquivos de backup do servidor de arquivos Amazon FSx para o SQL Server na instância EC2.
6. Use as ferramentas de gerenciamento do SQL Server para gerenciar seu banco de dados.

> **Note**
>
> O Amazon EC2 oferece o Microsoft SQL Server nas AMIs do Microsoft Windows para várias edições de servidores de SQL.

Usar o Gateway de Arquivos do Amazon S3

Você pode usar o Gateway de Arquivos do Amazon S3 para armazenar backups nativos do SQL Server no Amazon S3, conforme ilustrado no diagrama a seguir. Como alternativa, existem ferramentas como Commvault e LiteSpeed que ajudam você a gerenciar backups a nível do arquivo em escala e armazená-los diretamente no Amazon S3. Você também pode usar uma ferramenta como o SIOS DataKeeper para backup/recuperação e configuração de DR.



O processo consiste nestas etapas:

1. Os dados são gravados no disco de cache local do gateway de arquivos.
2. Depois que os dados persistirem com segurança no cache local, o gateway de arquivos confirma a conclusão da operação de gravação no aplicativo cliente.
3. O gateway de arquivos transfere dados para o bucket do S3 de forma assíncrona. Ele otimiza a transferência de dados e usa o HTTPS para criptografar dados em trânsito.
4. Depois que os dados são carregados no bucket do S3, eles permanecem no cache local do gateway de arquivos até serem removidos.

Próximas etapas e recursos

Este guia abordou as práticas recomendadas para rápida recuperação de desastres de bancos de dados do SQL Server. As recomendações incluem o uso de imagens para restaurar a instância do aplicativo e o uso de métodos SQL nativos para restaurar o banco de dados ou, preferencialmente, fazer o failover do banco de dados. Ao contrário das grandes restaurações de bancos de dados que podem levar horas, o uso dos backups do Amazon Elastic Compute Cloud (EC2Amazon) Amazon Machine Image (AMI) em combinação com os registros de transações mais recentes ajuda você a atender aos requisitos de objetivo de ponto de recuperação (RPO) e objetivo de tempo de recuperação (RTO), mantendo seus custos gerais baixos. A abordagem ideal depende do tamanho do banco de dados, do número e da natureza dos backups e da frequência dos backups dos logs de transações para os quais uma estratégia de recuperação de desastres precisa ser projetada. Consulte os links a seguir para obter mais informações, melhores práticas, guias de início rápido e orientações prescritivas sobre migração e hospedagem do SQL Server na Amazon. EC2

Documentação

- Melhores práticas e recomendações para agrupamento do SQL Server na Amazon (documentação da EC2 Amazon EC2)
- Loja de EC2 instâncias da Amazon (EC2 documentação da Amazon)
- Replicação de objetos (documentação do Amazon S3)
- Restauração rápida de snapshots do Amazon EBS (documentação da Amazon EC2)
- SQL Server com replicação Always On na Nuvem AWS (implantação de referência do Quick Start)
- Tipos de volume do Amazon EBS (EC2 documentação da Amazon)
- Usando FSx o Windows File Server com o Microsoft SQL Server (FSx documentação da Amazon)
- O que AWS Backupé (AWS Backup documentação)
- AWS Windows AMIs (EC2 documentação da Amazon)

AWS Orientação prescritiva

- Melhores práticas para implantar o Microsoft SQL Server na Amazon EC2
- EC2 Backup e recuperação da Amazon com instantâneos e AMIs
- Coloque o tempdb em um armazenamento de instância

Publicações e notícias no blog

- Armazene facilmente seus backups do SQL Server no Amazon S3 usando o Gateway de Arquivos
- Monitore os custos de transferência de dados relacionados à replicação do Amazon S3
- Implantação multirregional do SQL Server usando grupos de disponibilidade distribuídos
- Notas de campo: Criando uma arquitetura multirregional para o SQL Server usando FCI e grupos de disponibilidade distribuídos
- A Amazon EC2 agora oferece o Microsoft SQL Server no Microsoft Windows Server 2022 AMIs

Documentação do SQL Server

- Edições e atributos compatíveis do SQL Server

Apêndice: Tipos de armazenamento SSD do Amazon EBS

O Amazon Elastic Block Store (Amazon EBS) oferece os seguintes volumes com suporte a unidades de estado sólido (SSD). Para obter as informações mais recentes, consulte Tipos de volume do Amazon EBS na documentação do Amazon EC2.

	General Purpose SSD		Provisioned IOPS SSD		
Tipo de volume	`gp3`	`gp2`	`io2 Block Express`[1]	`io2`	`io1`
Durabilidade	99,8% a 99,9% de durabilidade (taxa anual de falhas de 0,1% a 0,2%)	99,8% a 99,9% de durabilidade (taxa anual de falhas de 0,1% a 0,2%)	Durabilidade de 99,999% (taxa anual de falhas de 0,001%)	Durabilidade de 99,999% (taxa anual de falhas de 0,001%)	99,8% a 99,9% de durabilidade (taxa anual de falhas de 0,1% a 0,2%)
Casos de uso	• Aplicações interativas de baixa latência • Ambientes de teste e desenvolvimento	• Aplicações interativas de baixa latência • Ambientes de teste e desenvolvimento	Workloads que exigem: • Latência média abaixo de um milissegundo • Performance estável de IOPS • Mais de 64.000 IOPS ou 1.000	• Workloads que exigem performance de IOPS sustentado ou mais do que 16.000 IOPS • Workloads de banco de dados com alto consumo de E/S	• Workloads que exigem performance de IOPS sustentado ou mais do que 16.000 IOPS • Workloads de banco de dados com alto consumo de E/S

	General Purpose SSD		Provisioned IOPS SSD		
			MiB/s de throughput		
Tamanho do volume	1 GiB – 16 TiB	1 GiB – 16 TiB	4 GiB – 64 TiB	4 GiB – 16 TiB	4 GiB – 16 TiB
Máximo de IOPS por volume (16 KiB E/S)	16.000	16,000	256.000	64,000[2]	64,000[2]
Throughput máximo por volume	1.000 MiB/s	250 MiB/s[3]	4.000 MiB/s	1.000 MiB/s[2]	1.000 MiB/s[2]
Multi-attach do Amazon EBS	Não suportado	Não suportado	Compatível	Compatível	Compatível
Volume de inicialização	Compatível	Compatível	Compatível	Compatível	Compatível

[1] Os volumes do `io2 Block Express` são compatíveis apenas com instâncias R5b. Os volumes `io2` anexados a uma instância `R5b` durante ou após a inicialização são executados automaticamente no Block Express. Para obter mais informações, consulte Volumes do io2 Block Express na documentação do Amazon EC2.

[2] O número máximo de throughput são garantidos somente em instâncias baseadas no Nitro System provisionadas com mais de 32.000 IOPS. Outras instâncias garantem até 32,000 IOPS e 500 MiB/s. Volumes `io1` criados antes de 6 de dezembro de 2017 e que não foram modificados desde a criação podem não atingir a performance total, a menos que você modifique o volume.

[3] O limite de throughput é entre 128 MiB/s e 250 MiB/s, dependendo do tamanho do volume. Volumes menores ou iguais a 170 GiB fornecem uma throughput máxima de 128 MiB/s. Os volumes maiores que 170 GiB e menores que 334 GiB fornecerão uma throughput máxima de 250 MiB/s se houver créditos de expansão disponíveis. Volumes maiores ou iguais a 334 GiB fornecem 250 MiB/

s independentemente dos créditos de expansão. Volumes gp2 criados antes de 3 de dezembro de 2018 e que não foram modificados desde a criação podem não atingir a performance total, a menos que você modifique o volume.

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
Publicação inicial	—	28 de fevereiro de 2022

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:

- Refatorar/rearquitetar: mova uma aplicação e modifique sua arquitetura aproveitando ao máximo os recursos nativos de nuvem para melhorar a agilidade, a performance e a escalabilidade. Isso normalmente envolve a portabilidade do sistema operacional e do banco de dados. Exemplo: migrar seu banco de dados Oracle on-premises para o Amazon Aurora Edição Compatível com PostgreSQL.
- 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: migrar 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

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.

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 AIOps é usado na estratégia de AWS migração, 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 descriptografia. É possível compartilhar a chave pública porque ela não é usada na descriptografia, 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 disrupção 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 deployment (implantação azul/verde)

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 disrupção 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 Implement break-glass procedures nas orientações do AWS Well-Architected.

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 conteinerizados 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 estressam 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.

criptografia no lado do cliente

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

Centro de excelência em nuvem (CCoE)

Uma equipe multidisciplinar que impulsiona os esforços de adoção da nuvem em toda a organização, incluindo o desenvolvimento de práticas recomendas 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 publicações CCo E 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 — Fazer investimentos fundamentais para escalar sua adoção da nuvem (por exemplo, criar uma landing zone, definir um CCo E, estabelecer um modelo de operações)
- Migração: migrar aplicações individuais
- Reinvenção: otimizar produtos e serviços e inovar na nuvem

Esses estágios foram definidos por Stephen Orban na postagem do blog The Journey Toward Cloud-First & the Stages of Adoption no blog de estratégia Nuvem AWS empresarial. 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 pipeline de CI/CD 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

Um conjunto 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 no AWS Well-Architected Framework. 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.

defense-in-depth

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 defense-in-depth abordagem 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 disastre. Para obter mais informações, consulte Recuperação de desastres de cargas de trabalho em AWS: Recuperação na nuvem no AWS Well-Architected Framework.

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, Design orientado por domínio: lidando com a complexidade no coração do software (Boston: Addison-Wesley Professional, 2003). Para obter informações sobre como usar o design orientado por domínio com o padrão strangler fig, consulte Modernizar incrementalmente os serviços web herdados do Microsoft ASP.NET (ASMX) 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. Os sistemas big-endian armazenam o byte mais significativo antes. Os sistemas little-endian armazenam o byte menos significativo antes.

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 em contexto, em que os modelos aprendem com exemplos (shots) incorporados aos prompts. Prompts few-shot podem ser eficazes 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 vem treinando em grandes conjuntos de dados generalizados e não rotulados. 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?.

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 governar recursos, políticas e conformidade em todas as unidades organizacionais ()OUs. 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 .

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 predições do modelo com os dados de retenção.

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 normal de DevOps lançamento.

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

IaC

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 a prática recomendada Implantar usando infraestrutura imutável no AWS Well-Architected Framework.

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 AWS Segurança recomenda configurar sua conta de rede com entrada, saída e inspeção VPCs para proteger a interface bidirecional entre seu aplicativo 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 manufatura por meio de avanços em conectividade, dados em tempo real, automação, analytics e IA/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 industrial das coisas (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 Criando uma estratégia de transformação digital industrial da Internet das Coisas (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 AWS Segurança recomenda configurar sua conta de rede com entrada, saída e inspeção VPCs para proteger a interface bidirecional entre seu aplicativo 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 são LLMs.

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, vazar 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.

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 Construindo mecanismos no AWS Well-Architected Framework.

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 machine-to-machine (M2M), baseado no padrão de publicação/assinatura, para dispositivos de IoT com recursos limitados.

microsserviço

Um serviço pequeno e independente que se comunica de forma bem definida APIs e normalmente é de propriedade de equipes pequenas e independentes. 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 leveza. APIs 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

Equipes multifuncionais que simplificam a migração de workloads 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: rehospede 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, o AWS Well-Architected Framework 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.

Open Process Communications - Unified Architecture (OPC-UA)

Um protocolo de comunicação machine-to-machine (M2M) para automação industrial. O 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) no AWS Well-Architected Framework.

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 exigida 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 S3 Regiões da AWS, criptografia do lado do servidor com AWS KMS (SSE-KMS) e solicitações dinâmicas ao bucket S3. `PUT` `DELETE`

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 AWS Segurança recomenda configurar sua conta de rede com entrada, saída e inspeção VPCs para proteger a interface bidirecional entre seu aplicativo 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 políta 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 contém informações sobre como você deseja que o Amazon Route 53 responda às consultas de DNS para um domínio e seus subdomínios em um 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 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 em AWS Organizations. SCPs defina barreiras ou estabeleça limites nas ações que um administrador pode delegar a usuários ou funções. Você pode usar SCPs como listas de permissão ou listas 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.

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 cacordo de serviço.

SLI

Veja indicador de nível de serviço.

SLO

Veja objetivo de nível de serviço.

split-and-seed modelo

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 Modernizar incrementalmente os serviços Web herdados do Microsoft ASP.NET (ASMX) 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 supervisório 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

Pares de valores-chave que atuam como metadados para organizar seus recursos. AWS 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.

gateway de trânsito

Um hub de trânsito de rede que você pode usar para interconectar sua rede com VPCs a rede local. 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. Para obter mais informações, consulte o guia Como quantificar a incerteza em sistemas de aprendizado profundo.

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 da VPC

Uma conexão entre duas VPCs que permite rotear o 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.