Modelagem de dados com o Amazon DynamoDB

AWS Orientação prescritiva

AWS Orientação prescritiva: Modelagem de dados com o Amazon DynamoDB

Table of Contents

Modelagem de dados com o Amazon DynamoDB

Processo, modelos e melhores práticas

Amazon Web Services (colaboradores)

Dezembro de 2023 (histórico do documento)

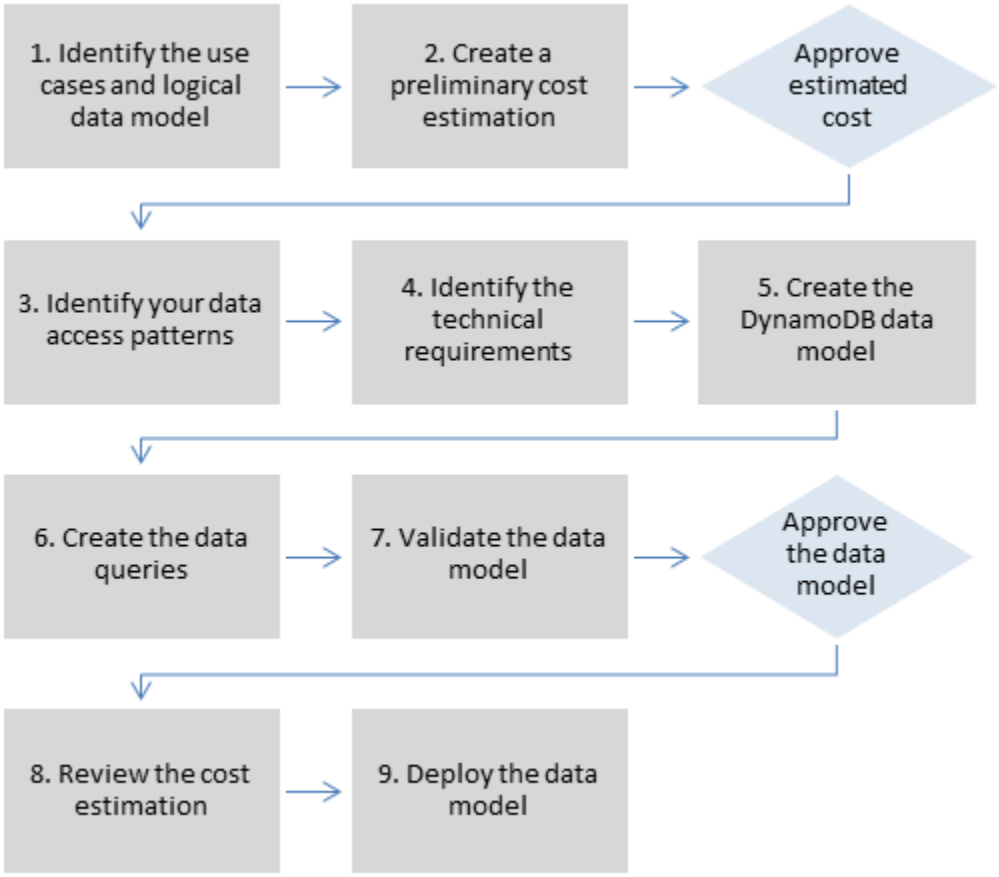
Os bancos de dados NoSQL fornecem esquemas flexíveis para a criação de aplicativos modernos. Eles são amplamente reconhecidos por sua facilidade de desenvolvimento, funcionalidade e desempenho em escala. O Amazon DynamoDB fornece performance rápida e previsível com escalabilidade integrada ao banco de dados NoSQL na Amazon Web Services (AWS) Cloud. Como um serviço de banco de dados totalmente gerenciado, o DynamoDB ajuda você a aliviar a carga administrativa de operar e escalar um banco de dados distribuído. Você não precisa se preocupar com provisionamento, instalação e configuração de hardware, replicação, aplicação de patches de software ou escalabilidade de clusters.

O design do esquema NoSQL requer uma abordagem diferente do design de sistema de gerenciamento de banco de dados relacional (RDBMS) tradicional. O modelo de dados RDBMS se concentra na estrutura dos dados e em seus relacionamentos com outros dados. A modelagem de dados NoSQL se concentra nos padrões de acesso ou em como o aplicativo vai consumir os dados, então ela armazena os dados de uma forma que ofereça suporte a operações de consulta diretas. Para um RDBMS como o Microsoft SQL Server ou o IBM Db2, você pode criar um modelo de dados normalizado sem pensar muito nos padrões de acesso. Você pode estender o modelo de dados para dar suporte aos seus padrões e consultas posteriormente.

Este guia apresenta um processo de modelagem de dados para usar o DynamoDB que fornece requisitos funcionais, desempenho e custos efetivos. O guia é para engenheiros de banco de dados que planejam usar o DynamoDB como banco de dados operacional para seus aplicativos em execução. AWS AWS A Professional Services usou o processo recomendado para ajudar empresas corporativas com a modelagem de dados do DynamoDB para diferentes casos de uso e cargas de trabalho.

Fluxo do processo de modelagem de dados

Recomendamos o seguinte processo ao modelar dados usando o Amazon DynamoDB. As estapas são discutidas em detalhes mais adiante neste guia.



Matriz RACI

Algumas organizações usam uma matriz de atribuição de responsabilidades (também conhecida como matriz RACI) para descrever as várias funções envolvidas em um projeto ou processo comercial específico. Este guia apresenta uma matriz RACI sugerida que pode ajudar sua organização a identificar as pessoas certas e as responsabilidades certas para o processo de modelagem de dados do DynamoDB. Para cada etapa no processo, ele lista as partes interessadas e seu envolvimento:

- R — responsável por concluir a etapa

- A — responsabilizável pela aprovação do e por assinar o trabalho
- C — consultado para fornecer informações para uma tarefa
- I — informado do progresso, mas não diretamente envolvido na tarefa

Dependendo da estrutura da organização e da equipe do projeto, as funções na matriz RACI a seguir podem ser desempenhadas pela mesma parte interessada. Em algumas situações, as partes interessadas são responsáveis por etapas específicas. Por exemplo, engenheiros de banco de dados podem ser responsáveis por criar e aprovar o modelo de dados, porque essa é sua área de domínio.

Etapa do processo	Usuário empresarial	Analista de negócios	Arquiteto de soluções	Engenheiro do banco de dados	Desenvolvedor de aplicações	DevOps engenheiro
1. Identificar os casos de uso e o modelo lógico de dados	C	R/A	eu	R		
2. Criar uma estimativa preliminar de custo	C	A	eu	R		
3. Identificar padrões de acesso a dados	C	A	eu	R		
4. Identificar os requisitos técnicos	C	C	A	R		

Etapa do processo	Usuário empresarial	Analista de negócios	Arquiteto de soluções	Engenheiro do banco de dados	Desenvolvedor de aplicações	DevOps engenheiro
5. Criar o modelo de dados do DynamoDB	eu	eu	eu	R/A		
6. Criar as consultas de dados	eu	eu	eu	R/A	R	
7. Validar o modelo de dados	A	R	eu	C		
8. Revisar a estimativa de custo	C	A	eu	R		
9. Implantar o modelo de dados do DynamoDB	eu	eu	C	C		R/A

Etapas do processo de modelagem de dados

Esta seção detalha cada etapa do processo de modelagem de dados recomendado para o Amazon DynamoDB.

Tópicos

Etapa 1. Identificar os casos de uso e o modelo lógico de dados

Objetivos

- Reunir as necessidades de negócios e os casos de uso que exigem um banco de dados NoSQL.
- Definir o modelo lógico de dados usando um diagrama de relacionamento entre entidades (ER).

Processo

- Analistas de negócios entrevistam usuários corporativos para identificar os casos de uso e os resultados esperados.
- O engenheiro de banco de dados cria o modelo conceitual de dados.
- O engenheiro de banco de dados cria o modelo lógico de dados.
- O engenheiro de banco de dados reúne informações sobre o tamanho do item, o volume de dados e o throughput esperado de leitura e gravação.

Ferramentas e recursos

- Avaliação de requisitos de negócios (consulte o modelo)
- Matriz de padrões de acesso (consulte o modelo)
- Sua ferramenta preferida para criar diagramas

RACI

Usuário empresarial	Analista de negócios	Arquiteto de soluções	Engenheiro do banco de dados	Desenvolvedor de aplicações	DevOps engenheiro
C	R/A	eu	R		

Saídas

- Casos de uso e requisitos comerciais documentados
- Modelo lógico de dados (diagrama ER)

Etapa 2. Criar uma estimativa preliminar de custo

Objetivo

- Desenvolver uma estimativa de custo preliminar para o DynamoDB.

Processo

- O engenheiro de banco de dados cria a análise de custo inicial usando as informações disponíveis e os exemplos apresentados na página de preços do DynamoDB.
  - Criar uma estimativa de custo para a capacidade sob demanda (consulte o exemplo).
  - Criar uma estimativa de custo para a capacidade provisionada (consulte o exemplo).
    - Para o modelo de capacidade provisionada, obtenha o custo estimado na calculadora e aplique o desconto pela capacidade reservada.

  - Compare os custos estimados dos dois modelos de capacidade.
  - Crie uma estimativa para todos os ambientes (Dev, Prod, QA).
- O analista de negócios analisa e aprova ou rejeita a estimativa preliminar de custo.

Ferramentas e recursos

- AWS Calculadora de preços

RACI

Usuário empresarial	Analista de negócios	Arquiteto de soluções	Engenheiro do banco de dados	Desenvolvedor de aplicações	DevOps engenheiro
C	A	eu	R		

Saídas

- Estimativa preliminar de custo

Etapa 3. Identificar padrões de acesso a dados

Os padrões de acesso ou padrões de consulta definem como os usuários e o sistema acessam os dados para satisfazer as necessidades comerciais.

Objetivo

- Documentar padrões de acesso a dados.

Processo

- O engenheiro de banco de dados e o analista de negócios entrevistam os usuários finais para identificar como os dados serão consultados usando o modelo de matriz de padrões de acesso a dados.

  - Para novos aplicativos, analise as histórias dos usuários referentes a atividades e objetivos. Documente os vários casos de uso identificados e analise os padrões de acesso que eles exigem.
  - Para aplicativos existentes, analise os logs de consulta para saber como as pessoas estão usando o sistema atualmente e quais são os principais padrões de acesso.
- O engenheiro de banco de dados identifica as seguintes propriedades dos padrões de acesso:
  - Tamanho de dados: saber o volume de dados que serão armazenados e solicitados ao mesmo tempo ajudará a determinar a maneira mais eficiente de particionar os dados (consulte o post do blog).
  - Forma dos dados: em vez de remodelar dados quando uma consulta é processada (como um sistema RDBMS faz), um banco de dados NoSQL organiza os dados para que sua forma no banco de dados corresponda ao que será consultado. Esse é um fator importante no aumento da velocidade e da escalabilidade.
  - Velocidade dos dados: o DynamoDB é escalado aumentando-se o número de partições físicas que estão disponíveis para processar consultas e distribuindo-se os dados com eficiência entre essas partições. Saber antecipadamente qual é o pico das cargas de consulta pode ajudar a determinar como particionar os dados para melhor utilização da capacidade de E/S.
- O usuário corporativo prioriza os padrões de acesso ou consulta.
  - As consultas prioritárias geralmente são as mais usadas ou mais relevantes. Também é importante identificar consultas que exijam menor latência de resposta.

Ferramentas e recursos

- Matriz de padrões de acesso (consulte o modelo)
- Escolher a chave de partição correta do DynamoDB (blog do banco de dados AWS)
- Design NoSQL para DynamoDB (documentação do DynamoDB)

RACI

Usuário empresarial	Analista de negócios	Arquiteto de soluções	Engenheiro do banco de dados	Desenvolvedor de aplicações	DevOps engenheiro
C	A	eu	R		

Saídas

- Matriz de padrões de acesso a dados

Exemplo

Padrão de acesso	Priority	Ler ou escrever	Descrição	Tipo (item único, vários itens ou todos)	Atributo chave	Filtros	Ordenação de resultados
Criar perfil de usuário	Alto	Escrever	O usuário cria um novo perfil	Item único	Nome de usuário	N/D	N/D
Atualizar perfil de usuário	Médio	Escrever	O usuário atualiza seu perfil	Item único	Nome de usuário	Nome de usuário = usuário atual	N/D

Etapa 4. Identificar os requisitos técnicos

Objetivo

- Reúna os requisitos técnicos do banco de dados do DynamoDB.

Processo

- Os analistas de negócios entrevistam o usuário comercial e a DevOps equipe para reunir os requisitos técnicos usando o questionário de avaliação.

Ferramentas e recursos

- Avaliação de requisitos técnicos (veja exemplo de questionário)

RACI

Usuário empresarial	Analista de negócios	Arquiteto de soluções	Engenheiro do banco de dados	Desenvolvedor de aplicações	DevOps engenheiro
C	C	A	R		

Saídas

- Documento de requisitos técnicos

Etapa 5. Criar o modelo de dados do DynamoDB

Objetivo

- Criar o modelo de dados do DynamoDB.

Processo

- O engenheiro de banco de dados identifica quantas tabelas serão necessárias para cada caso de uso. Você deve manter o mínimo de tabelas possível em uma aplicação do DynamoDB.
- Com base nos padrões de acesso mais comuns, identifique a chave primária que pode ser de dois tipos: uma chave primária com uma chave de partição que identifica dados ou uma chave primária com uma chave de partição e uma chave de classificação. Uma chave de classificação é uma chave secundária para agrupar e organizar dados para que possam ser consultados em uma partição de forma eficiente. Você pode usar chaves de classificação para definir relacionamentos hierárquicos em seus dados que podem ser consultados em qualquer nível da hierarquia (consulte o post do blog).
  - Design de chave de partição
    - Definir a chave de partição e avaliar sua distribuição.
    - Identificar a necessidade de fragmentação de gravação para distribuir uniformemente as workloads.
  - Design de chaves de classificação
    - Identificar a chave de classificação.
    - Identificar a necessidade de uma chave de classificação composta.
    - Identificar a necessidade de controle de versão.
- Com base nos padrões de acesso, identificar os índices secundários para atender aos requisitos de consulta.
  - Identifique a necessidade de índices secundários locais (LSIs). Esses são índices que possuem a mesma chave de partição da tabela base, mas uma chave de classificação diferente.
    - Para tabelas com LSIs, há um limite de tamanho de 10 GB por valor de chave de partição. Uma tabela com LSIs pode armazenar qualquer número de itens, desde que o tamanho total de qualquer valor de chave de partição não exceda 10 GB.
  - Identifique a necessidade de índices secundários globais (GSIs). Esses são índices com uma chave de partição e uma chave de classificação que podem ser diferentes das contidas na tabela base (consulte o post do blog).
  - Definir as projeções do índice. Considere projetar menos atributos para reduzir o tamanho dos itens gravados no índice. Nesta etapa, você deverá determinar se deseja usar o seguinte:
    - Índices esparsos
    - Consultas de agregação materializadas

    - Sobrecarga do GSI
    - Fragmentação do GSI
    - Uma réplica eventualmente consistente usando GSI
- O engenheiro de banco de dados determina se os dados incluirão itens grandes. Nesse caso, eles projetam a solução usando compressão ou armazenando dados no Amazon Simple Storage Service (Amazon S3).
- O engenheiro de banco de dados determina se dados de séries temporais serão necessários. Nesse caso, eles usam o padrão de design de séries temporais para modelar os dados.
- O engenheiro de banco de dados determina se o modelo ER inclui many-to-many relacionamentos. Nesse caso, ele usa um padrão de design de lista de adjacências para modelar os dados.

Ferramentas e recursos

- NoSQL Workbench para Amazon DynamoDB — Fornece modelagem de dados, visualização de dados e recursos de desenvolvimento e teste de consultas para ajudá-lo a projetar seu banco de dados do DynamoDB
- Design NoSQL para DynamoDB (documentação do DynamoDB)
- Escolher a chave de partição correta do DynamoDB (blog da base de dados AWS)
- Práticas recomendadas para usar índices secundários no DynamoDB (documentação do DynamoDB)
- Como criar índices secundários globais do Amazon DynamoDB (blog da base de dados AWS)

RACI

Usuário empresarial	Analista de negócios	Arquiteto de soluções	Engenheiro do banco de dados	Desenvolvedor de aplicações	DevOps engenheiro
eu	eu	eu	R/A		

Saídas

- Esquema de tabela do DynamoDB que satisfaz seus padrões e requisitos de acesso

Exemplo

A captura de tela a seguir mostra o NoSQL Workbench.

RetailDatabase | GSI: BundleProducts-VendorProducts-CustomerOrdersByOrderId | GSI: VendorOrdersByStatusDate-ProductInventory | GSI: ProductCatalog-CustomerOrdersByProduct

Primary Key		Attributes					
Partition Key: pk	Sort Key: sk						
P1	B1	GSI1-PK	GSI1-SK	name	desc		
		B1	P1	The Tiki Bundle	Everything you need for an island theme party.		
P4	B2	GSI1-PK	GSI1-SK	name	desc		
		B2	P4	Tiki Bar Set	Be the Mai Tai master with your very own Tiki Bar.		
P2	B1	name	desc	qty	GSI1-PK	GSI1-SK	location
		Tiki Torch	Bamboo tiki torch, 4 ft	6	B1	P2	W1-A9-S10-B52
	B2	name	desc	qty	GSI1-PK	GSI1-SK	location
		Tiki Torch	Bamboo tiki torch, 4 ft	2	B2	P2	W1-A9-S10-B52
	P2	name	desc	qty	location	reorderAt	GSI3-SK
		Tiki Torch	Bamboo tiki torch, 4 ft	656	W1-A9-S10-B52	100	/GardenOutdoor/OutdoorDecor/Lighting/LanternsT
	B1	name	desc	qty	GSI1-PK	GSI1-SK	location
		Tiki Statue - Pele	Tiki of the Hawaiian Fire Goddess Pele, 5 ft.	1	B1	P3	W1-A15-S6-B27

Etapa 6. Criar as consultas de dados

Objetivo

- Criar as consultas principais para validar o modelo de dados.

Processo

- O engenheiro de banco de dados cria manualmente uma tabela do DynamoDB na região ou em AWS seu computador (DynamoDB Local).
- O engenheiro de banco de dados adiciona dados de amostra à tabela do DynamoDB.

- O engenheiro de banco de dados cria facetas usando o NoSQL Workbench para Amazon DynamoDB ou o SDK AWS for Java ou Python para criar exemplos de consultas (veja a postagem do blog).

  As facetas são como uma visualização da tabela do DynamoDB.
- O engenheiro de banco de dados e o desenvolvedor de nuvem criam exemplos de consultas usando o AWS Command Line Interface (AWS CLI) ou o AWS SDK para o idioma preferido.

Ferramentas e recursos

- Uma AWS conta ativa, para obter acesso ao console do DynamoDB
- DynamoDB Local (opcional), se você quiser criar o banco de dados em seu computador sem acessar o serviço da web do DynamoDB
- NoSQL Workbench para Amazon DynamoDB (download e documentação)
- AWS SDK na linguagem de sua escolha (PythonJavaScript, PHP, .NET, Ruby, Java, Go, Node.js, C++ e SAP ABAP)

RACI

Usuário empresarial	Analista de negócios	Arquiteto de soluções	Engenheiro do banco de dados	Desenvolvedor de aplicações	DevOps engenheiro
eu	eu	eu	R/A	R	

Saídas

- Código para consultar a tabela do DynamoDB

Exemplos

- Exemplos do DynamoDB usando AWS o SDK for Java
- Exemplos de Python
- JavaScriptexemplos

Etapa 7. Validar o modelo de dados

Objetivo

- Certifique-se de que o modelo de dados satisfaça seus requisitos.

Processo

- O engenheiro de banco de dados preenche a tabela do DynamoDB com dados de exemplo.
- O engenheiro de banco de dados executa o código para consultar a tabela do DynamoDB.
- O engenheiro de banco de dados coleta os resultados da consulta.
- O engenheiro de banco de dados coleta as métricas de desempenho da consulta.
- O usuário corporativo valida que os resultados da consulta satisfazem as necessidades comerciais.
- Os analistas de negócios validam os requisitos técnicos.

Ferramentas e recursos

- Uma AWS conta ativa, para obter acesso ao console do DynamoDB
- DynamoDB Local (opcional), se você quiser criar o banco de dados em seu computador sem acessar o serviço da web do DynamoDB
- AWS SDK no idioma de sua escolha

RACI

Usuário empresarial	Analista de negócios	Arquiteto de soluções	Engenheiro do banco de dados	Desenvolvedor de aplicações	DevOps engenheiro
A	R	eu	C		

Saídas

- Modelo de dados aprovado

Etapa 8. Revisar a estimativa de custo

Objetivos

- Definir o modelo de capacidade e estimar os custos do DynamoDB para refinar a estimativa de custo a partir da etapa 2.
- Obtenha a aprovação financeira final do analista de negócios e das partes interessadas.

Processo

- O engenheiro de banco de dados identifica a estimativa do volume de dados.
- O engenheiro de banco de dados identifica os requisitos de transferência de dados.
- O engenheiro de banco de dados define as unidades de capacidade de leitura e gravação necessárias.
- O analista de negócios decide entre modelos de capacidade sob demanda e provisionados.
- O engenheiro de banco de dados identifica a necessidade de ajuste de escala automático do DynamoDB.
- O engenheiro de banco de dados insere os parâmetros na ferramenta de calculadora mensal simples.
- O engenheiro de banco de dados apresenta a estimativa final do preço às partes interessadas do negócio.
- O analista de negócios e as partes interessadas aprovam ou rejeitam a solução.

Ferramentas e recursos

- AWS Calculadora de preços

RACI

Usuário empresarial	Analista de negócios	Arquiteto de soluções	Engenheiro do banco de dados	Desenvolvedor de aplicações	DevOps engenheiro
C	A	eu	R		

Saídas

- Modos de capacidade
- Estimativa de custo revisada

Etapa 9. Implantar o modelo de dados

Objetivo

- Implante a tabela (ou tabelas) do DynamoDB no. Região da AWS

Processo

- DevOps O arquiteto cria um AWS CloudFormation modelo ou outra ferramenta de infraestrutura como código (IaC) para a tabela (ou tabelas) do DynamoDB. AWS CloudFormation fornece uma forma automatizada de provisionar e configurar suas tabelas e recursos associados.

Ferramentas e recursos

- AWS CloudFormation

RACI

Usuário empresarial	Analista de negócios	Arquiteto de soluções	Engenheiro do banco de dados	Desenvolvedor de aplicações	DevOps engenheiro
eu	eu	C	C		R/A

Saídas

- AWS CloudFormation modelo

Exemplo

```
mySecondDDBTable:
   Type: AWS::DynamoDB::
   Table DependsOn: "myFirstDDBTable"
   Properties:
   AttributeDefinitions:
      -  AttributeName: "ArtistId"
         AttributeType: "S"
      -  AttributeName: "Concert"
         AttributeType: "S"
      -  AttributeName: "TicketSales"
         AttributeType: "S"
   KeySchema:
      -  AttributeName: "ArtistId"
         KeyType: "HASH"
      -  AttributeName: "Concert"
         KeyType: "RANGE"
   ProvisionedThroughput:
      ReadCapacityUnits:
         Ref: "ReadCapacityUnits"
      WriteCapacityUnits:
         Ref: "WriteCapacityUnits"
   GlobalSecondaryIndexes:
      -  IndexName: "myGSI"
         KeySchema:
```

```
            - AttributeName: "TicketSales"
            KeyType: "HASH"
         Projection:
            ProjectionType: "KEYS_ONLY"
         ProvisionedThroughput:
         ReadCapacityUnits:
            Ref: "ReadCapacityUnits"
         WriteCapacityUnits:
            Ref: "WriteCapacityUnits"
   Tags:
      -  Key: mykey
         Value: myvalue
```

Modelos

Os modelos fornecidos nesta seção são baseados na modelagem de dados de jogadores de jogos com o Amazon DynamoDB no site. AWS

> **Note**
>
> As tabelas nesta seção usam MM como abreviatura de milhão e K como abreviatura de mil.

Tópicos

Modelo de avaliação de requisitos de negócios

Forneça uma descrição para o caso de uso:

> **Descrição**
>
> Imagine que você está criando um jogo multiplayer on-line. No seu jogo, grupos de 50 jogadores participam de uma sessão para jogar uma partida, que normalmente leva cerca de 30 minutos para ser jogada. Durante o jogo, você precisa atualizar o registro de um jogador específico para indicar a quantidade de tempo que o jogador está jogando, suas estatísticas ou se ele venceu o jogo. Os usuários querem ver os jogos anteriores que jogaram, seja para ver os vencedores dos jogos ou para assistir a uma repetição da ação de cada jogo.

Forneça informações sobre seus usuários:

Usuário	Descrição	Número esperado
Jogador	Jogador de jogos online.	1 MM

Equipe de desenvolvimento	Equipe interna que usará as estatísticas do jogo para melhorar o experiência de jogo.	100

Forneça informações sobre as fontes de dados e como os dados serão ingeridos:

Origem	Descrição	Usuário
Jogo online	Os jogadores criarão perfis e iniciarão novos jogos.	Jogador
Aplicativo de jogo	O aplicativo do jogo coletará automaticamente estatísticas sobre os jogos, como horário de início e término, número de jogadores, posição de cada jogador e mapa do jogo.	

Forneça informações sobre como os dados serão consumidos:

Consumidor	Descrição	Usuário
Jogo online	Os jogadores verão os perfis e analisarão as estatísticas do jogo.	Jogador
Análise de dados	A equipe de desenvolvimento do jogo extrairá as estatísticas do jogo para análise de dados e para melhorar a experiênc ia do usuário. Os dados serão exportados do armazenam ento de dados e importado s para o Amazon S3 para	Equipe de desenvolvimento

	apoiar a análise por meio de um aplicativo Spark.	

Forneça uma lista de entidades e como elas são identificadas:

Nome da entidade	Descrição	Identifier
Jogador de jogo	Armazena informações como identificação, endereço, dados demográficos, interesses de cada usuário (jogador).	Nome de usuário
Instância do jogo	Fornece informações sobre cada jogo jogado, incluindo criador, início, fim e mapa Yplayed.	ID do jogo
Mapeamento de usuários do jogo	Representa as many-to-m any relações entre usuários e jogos.	ID do jogo e nome de usuário

Criar um modelo ER para as entidades:

Fornecer estatísticas de alto nível sobre as entidades:

Entity Name	Nº estimado de registros	Tamanho do registro	Observações
Jogador de jogo	1 MM	< 1 KB	A plataforma de jogos tem cerca de 1 MM de usuários.
Instância do jogo	6 MM (100.000 K/dia * 60 dias)	< 1 KB	Em média, há 100 mil jogos todos os dias. Precisamos armazenar os últimos 60 dias.
Mapeamento de usuários do jogo	300 MM (6 jogos MM * 50 jogadores)	< 1 KB	Em média, cada jogo tem 50 jogadores sobre os quais precisamos armazenar informaçõ es.

Modelo de avaliação de requisitos técnicos

Fornecer informações sobre os tipos de ingestão de dados:

Tipo de ingestão de dados	S/N	Descrição	Frequência
Acesso ao aplicativo	Y		
Gateway de API	Y		
Streaming de dados	N		
Processo em lote	N		
ETL	N		
Importação de dados	N		
Séries temporais	N		

Fornecer informações sobre os tipos de consumo de dados:

Tipo de consumo de dados	S/N	Descrição	Frequência
Acesso ao aplicativo			
Gateway de API			
Exportação de dados			
Análise de dados			
Agregação de dados			
Relatórios			
Pesquisar			

Streaming de dados

ETL

Fornecer estimativas do volume de dados:

Nome da entidade	Nº estimado de registros	Tamanho do registro	Volume de dados
Jogador	1 MM	< 1 KB	~ 1 GB (1 MM * 1 KB)
Instância do jogo	6 MM (100K/dia * 60 dias)	< 1 KB	~ 6 GB (6 MILÍMETROS* 1 KB)
Mapeamento de usuários do jogo	300 MM (6 jogos MM * 50 jogadores)	< 1 KB	~ 300 GB (300 MILÍMETROS* 1 KB)

> **Note**
>
> O período de retenção de dados é 60 dias. Depois de 60 dias, os dados devem ser armazenados no Amazon S3 para análise, usando o DynamoDB Time to Live (TTL) para mover automaticamente os dados do DynamoDB para o Amazon S3.

Responda a estas perguntas sobre padrões de tempo:

- Em que período o aplicativo está disponível para o usuário (por exemplo, 24 horas por dia, 7 dias por semana ou das 9h às 17h nos dias de semana)?
- Há um pico de uso durante o dia? Quantas horas? Qual é o percentual de utilização do aplicativo?

Especificar os requisitos de throughput de gravação:

Nome da entidade	Escreve por dia	Horas/dia	Escreve/segundo
Jogador	10.000 atualizações	18	< 1
Instância do jogo	300.000	18	< 5
Mapeamento de usuários do jogo	1.800.000.000	18	~ 27.777

Observações

Operações de gravação dos jogadores do jogo: 1% dos usuários atualizam seus perfis todos os dias, então esperamos 10.000 atualizações para 1.000.000 de usuários.
Operações de gravação de instância do jogo: 100.000 jogos/dia. Para cada jogo, temos pelo menos três operações de gravação, na criação, no início e no final, então o total é de 300.000 operações de gravação.
Operações de gravação de mapeamento dos usuários do jogo: 100.000 jogos/dia para cada jogo com 50 jogadores. A duração média do jogo é de 30 minutos e a posição do jogador é atualizada a cada cinco segundos. Estimamos uma média de 360 atualizações por jogador, então o total é 100.000 X 50 X 360 = 1.800.000.000 operações de gravação.

Especificar os requisitos de throughput de leitura:

Nome da entidade	Leituras//dia	Horas/dia	Leituras/seg
Jogador	200.000	18	~ 3
Instância do jogo	5.000.000	18	~ 77
Mapeamento de usuários do jogo	1.800.000.000	18	~ 27.777

Observações

Operações de leitura do jogador do jogo: 20% dos usuários iniciam jogos, então 1 MM X 0,2 = 200.000.

Operações de leitura de instância do jogo: 100.000 jogos/dia. Para cada jogo, temos pelo menos uma operação de leitura por jogador e 50 jogadores por jogo, então o total é de 5.000.000 de operações de leitura.
Operações de leitura de mapeamento dos usuários do jogo: 100.000 jogos/dia para cada jogo com 50 jogadores. A duração média do jogo é de 30 minutos e a posição do jogador é atualizada a cada cinco segundos. Estimamos uma média de 360 atualizações por jogador, e cada atualizaão requer uma operação de leitura, então o total é 100.000 X 50 X 360 = 1.800.000.000 operações de leitura.

Especificar os requisitos de latência de acesso aos dados:

Operation	99 percentis	Latência máxima
Read	30 ms	100 ms
Write	10 ms	50 ms

Especificar os requisitos de disponibilidade de dados:

Requisito	S/N	Métrica	Observações
Alta disponibilidade	Y	99,9%	
RTO	Y	1 hora	Objetivo de tempo de recuperação
RPO	Y	1 hora	Objetivo do ponto de recuperação
Recuperação de desastres	N		
Replicação de dados na região	N		
Replicação de dados entre regiões	N	Latência de 3 segundos	Qual Regiões da AWS?

Especificar os requisitos de segurança:

Requisito	S/N	Observações
Armazenamento de dados confidenciais	N	Informações de saúde protegidas (PHI), informaçõ es do setor de cartões de pagamento (PCI), informações de identificação pessoal (PII)?
Criptografia em repouso	Y	
Criptografia em trânsito	Y	
Criptografia do lado do cliente	N	
Qualquer biblioteca de criptografia proprietária ou de terceiros	N	
Registro de acesso a dados	N	
Auditoria de acesso a dados	N	

Modelo de padrões de acesso

Coletar e documentar informações sobre os padrões de acesso para o caso de uso usando os seguintes campos:

Campo	Descrição
Padrão de acesso	Forneça um nome para o padrão de acesso.
Descrição	Forneça uma descrição mais detalhada do padrão de acesso.

Campo	Descrição
Priority	Defina uma prioridade para o padrão de acesso (alto, médio ou baixo). Isso define os padrões de acesso mais relevantes para o aplicativo.
Ler ou escrever	É um padrão de acesso de leitura ou gravação?
Tipo	O padrão acessa um único item, vários itens ou todos os itens?
Filtro	O padrão de acesso requer algum filtro?
Sort	O resultado requer alguma classificação?

Modelo

Padrão de acesso	Descrição	Priority	Ler ou escrever	Tipo (item único, múltiplo itens, ou todos)	Atributo chave	Filtros	Ordenação de resultados
Criar perfil de usuário	O usuário cria um novo perfil.	Alto	Write	Item único	Nome de usuário	N/D	N/D
Atualizar perfil de usuário	O usuário atualiza seu perfil.	Médio	Write	Item único	Nome de usuário	Nome de usuário = usuário atual	N/D

Obter perfil de usuário	O usuário analisa seu perfil.	Alto	Read	Item único	Nome de usuário	Nome de usuário = usuário atual	N/D
Crie um jogo	O usuário cria um novo jogo.	Alto	Write	Item único	ID do jogo	N/D	N/D
Encontre jogos abertos	O usuário pesquisa jogos abertos. Os resultados da pesquisa são classificados por data e hora de início em ordem decrescente.	Alto	Read	Vários itens		GameStatus = aberto	Iniciar o registro de data e hora descendente

Encontre jogos abertos por mapa	O usuário pesquisa jogos abertos usando um mapa específico classificado por data e hora de início em ordem decrescente ordem.	Médio	Read	Vários itens		GameStatus = abrir e mapear = XYZ	Iniciar o registro de data e hora descendente
Ver jogo	O usuário analisa os detalhes de um jogo.	Alto	Read	Item único	ID do jogo	N/D	N/D
Exibir usuários em um jogo	O usuário obtém uma lista de todos os usuários em um jogo.	Médio	Read	Vários itens		ID do jogo = XYZ	N/D

Junte um usuário a um jogo	O usuário entra em um jogo aberto.	Alto	Write	Item único	ID do jogo e nome de usuário	GameStatus = aberto	N/D
Comece um jogo	O usuário inicia um novo jogo.	Alto	Write	Item único	ID do jogo	N/D	N/D
Atualize o jogo para o usuário	Atualize a posição do usuário no jogo.	Médio	Write	Item único	ID do jogo e nome de usuário	N/D	N/D
Atualizar jogo	O jogo termina; atualize as estatísticas.	Médio	Write	Item único	ID do jogo	N/D	N/D
Encontre todos os jogos anteriores de um usuário	Liste todos os jogos que um usuário jogou ordenados pela data e hora de início do jogo.	Baixo	Read	Vários itens	Nome de usuário e ID do jogo	Nome de usuário = usuário atual	Carimbo de data e hora de início

Exportar dados para análise de dados	A equipe de desenvolvimento executará um trabalho em lotes para exportar dados para o Amazon S3.	Baixo	Read	Todos	N/D	N/D	N/D

Práticas recomendadas

Considere usar as seguintes práticas recomendadas de design do DynamoDB:

- Design de chave de partição: use uma chave de partição de alta cardinalidade para distribuir a carga uniformemente.
- Padrão de design da lista de adjacências — Use esse padrão de design para gerenciamento one-to-many e many-to-many relacionamentos.
- Índice esparso — Use índice esparso para seus índices secundários globais (). GSIs Ao criar um GSI, você especifica uma chave de partição e, opcionalmente, uma chave de classificação. Somente itens na tabela base que contêm uma chave de partição GSI correspondente aparecem no índice esparso. Isso ajuda a manter o GSIs tamanho menor.
- Sobrecarga de índice: use o mesmo GSI para indexar vários tipos de itens.
- Fragmentação de gravação de GSI: fragmente de forma inteligente para distribuir dados entre as partições para permitir consultas mais rápidas e eficientes.
- Itens grandes: armazene somente metadados dentro da tabela, salve o blob no Amazon S3 e mantenha a referência no DynamoDB. Divida itens grandes em vários itens e indexe com eficiência usando chaves de classificação.

Para obter mais práticas recomendadas de design, consulte a documentação do Amazon DynamoDB.

Exemplo de modelagem hierárquica de dados

As seções a seguir usam um exemplo de empresa automotiva para mostrar como você pode usar as etapas do processo de modelagem de dados para projetar um sistema de gerenciamento de componentes de vários níveis no DynamoDB.

Tópicos

Etapa 1: identificar os casos de uso e o modelo lógico de dados

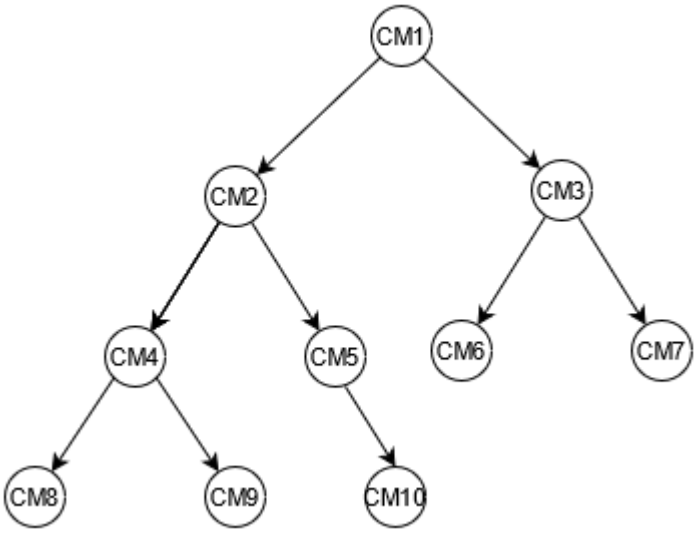
Uma empresa automotiva deseja criar um sistema de gerenciamento de componentes transacionais para armazenar e pesquisar todas as peças automotivas disponíveis e criar relacionamentos entre diferentes componentes e peças. Por exemplo, um carro contém várias baterias, cada bateria contém vários módulos de alto nível, cada módulo contém várias células e cada célula contém vários componentes de baixo nível.

Geralmente, para criar um modelo de relacionamento hierárquico, um banco de dados de grafos como o Amazon Neptune é uma escolha melhor. Em alguns casos, no entanto, o Amazon DynamoDB é uma alternativa melhor para a modelagem hierárquica de dados devido à sua flexibilidade, segurança, desempenho e escala.

Por exemplo, é possível criar um sistema em que 80 a 90% das consultas sejam transacionais no qual o DynamoDB se encaixe bem. Neste exemplo, os outros 10 a 20% das consultas são relacionais, onde um banco de dados gráfico como o Neptune se encaixa melhor. Nesse caso, incluir um banco de dados adicional na arquitetura para atender apenas 10 a 20% das consultas poderia aumentar o custo. Isso também aumenta a carga operacional de manter vários sistemas e

sincronizar os dados. Em vez disso, você pode modelar essas consultas relacionais de 10 a 20% no DynamoDB.

Criar o diagrama de uma árvore de exemplo para componentes automotivos pode ajudar a mapear a relação entre eles. O diagrama a seguir mostra um gráfico de dependência com quatro níveis. CM1 é o componente de nível superior do próprio carro de exemplo. Ele tem dois subcomponentes para dois exemplos de baterias CM2 e. CM3 Cada bateria tem dois subcomponentes, que são os módulos. CM2 tem módulos CM4 e CM5, e CM3 tem módulos CM6 CM7 e. Cada módulo tem vários subcomponentes, que são as células. O CM4 módulo tem duas células CM8 CM9 e. CM5 tem uma célula, CM1 0. CM6 e ainda CM7 não tem nenhuma célula associada.



Este guia usará essa árvore e seus identificadores de componentes como referência. Um componente superior será chamado de pai, enquanto um subcomponente será chamado de filho. Por exemplo, o componente superior CM1 é o pai de CM2 CM3 e. CM2 é pai de CM4 CM5 e. Isso representa graficamente os relacionamentos entre pais e filhos.

Na árvore, é possível ver o gráfico completo de dependências de um componente. Por exemplo, CM8 é dependente de CM4, que é dependente de CM2, que é dependente de CM1. A árvore define o gráfico de dependências completo como o caminho. Um caminho descreve duas coisas:

- O gráfico de dependências
- A posição na árvore

Preenchendo os modelos de acordo com os requisitos de negócios:

Forneça informações sobre seus usuários:

Usuário	Descrição
Funcionário	Funcionário interno da empresa automotiva que precisa de informações sobre carros e seus componentes

Forneça informações sobre as fontes de dados e como os dados serão ingeridos:

Origem	Descrição	Usuário
Sistema de gestão	Sistema que armazenará todos os dados relaciona dos às peças automotivas disponíveis e suas relações com outros componentes e peças.	Funcionário

Forneça informações sobre como os dados serão consumidos:

Consumidor	Descrição	Usuário
Sistema de gestão	Recupere todos os component es secundários imediatos para um ID de componente pai.	Funcionário
Sistema de gestão	Recupere uma lista recursiva de todos os componentes secundários para um ID de componente.	Funcionário
Sistema de gestão	Veja os ancestrais de um componente.	Funcionário

Etapa 2: criar uma estimativa preliminar de custo

É importante calcular uma estimativa do custo de todos os ambientes do seu aplicativo para que você possa verificar se a solução é financeiramente viável. A melhor prática é fazer uma estimativa de alto nível e obter a aprovação do analista de negócios antes de prosseguir com o desenvolvimento e a implantação.

- O engenheiro de banco de dados cria a análise de custo inicial usando as informações disponíveis e os exemplos apresentados na página de preços do DynamoDB.
  - Criar uma estimativa de custo para a capacidade sob demanda (consulte o exemplo).
  - Criar uma estimativa de custo para a capacidade provisionada (consulte o exemplo).
    - Para o modelo de capacidade provisionada, obtenha o custo estimado na calculadora e aplique o desconto pela capacidade reservada.
  - Compare os custos estimados dos dois modelos de capacidade.
  - Crie uma estimativa para todos os ambientes (Dev, Prod, QA).
- O analista de negócios analisa e aprova ou rejeita a estimativa preliminar de custo.

Usando esses valores de referência, você pode criar um preço estimado para enviar para aprovação. Para criar o orçamento, você pode usar a página de preços do DynamoDB e a Calculadora de preços da AWS.

Etapa 3: identificar seus padrões de acesso a dados

Este exemplo de caso de uso tem os seguintes padrões de acesso para gerenciar relacionamentos entre diferentes componentes do carro.

Padrão de acesso	Priority	Ler ou escrever	Descrição	Tipo	Filtros	Ordenação de resultados
Criança imediata	Alto	Leitura	Recupere todos os componentes secundários	Vários	`Component ID`	N/D

			imediatos para um ID de componente pai.			
Todos os componentes infantis	Alto	Leitura	Recupere uma lista recursiva de todos os componentes secundários para um ID de componente.	Vários	`Component ID`	N/D
Antepassados	Alto	Leitura	Recupere os ancestrais de um componente.	Vários	`Component ID`	N/D

Etapa 4: Identificar os requisitos técnicos

Este exemplo não tem requisitos técnicos específicos, que estão fora do escopo deste exemplo. Em casos reais, é uma prática recomendada concluir essa etapa e validar se todos os requisitos técnicos foram atendidos antes de prosseguir com o desenvolvimento e a implantação. Você pode usar o questionário de exemplo para concluir essa etapa em seu caso comercial. Além disso, recomendamos validar as cotas de serviço do DynamoDB para garantir que não haja limites rígidos na solução projetada.

Etapa 5: criar um modelo de dados do DynamoDB

Defina as chaves de partição para sua tabela base e índices secundários globais (GSIs):

- Seguindo as melhores práticas de design de chaves, use `ComponentId` como chave de partição para a tabela base neste exemplo. Por ser único, `ComponentId` pode oferecer granularidade. O DynamoDB usa o valor de hash da sua chave de partição para determinar a partição em que os dados são armazenados fisicamente. O ID exclusivo do componente gera um valor de hash diferente, o que pode facilitar a distribuição dos dados dentro da tabela. Você pode consultar a tabela base usando uma chave de `ComponentId` partição.
- Para encontrar filhos imediatos de um componente, crie um GSI em que `ParentId` esteja a chave de partição e `ComponentId` seja a chave de classificação. Você pode consultar esse GSI usando `ParentId` como chave de partição.
- Para todos os filhos recursivos de um componente, crie um GSI em que `GraphId` é a chave de partição, e `Path` é a chave de classificação. É possível consultar esse GSI usando `GraphId` como a chave de partição e o operador `BEGINS_WITH(Path, "$path")` na chave de classificação.

	Chave de partição	Chave de classific ação	Atributos de mapeamento
Tabela base	`ComponentId`		`ParentId`, `GraphId`, `Path`
GSI1	`ParentId`	`ComponentId`	
GSI2	`GraphId`	`Path`	`ComponentId`

Armazenar componentes na tabela

A próxima etapa é armazenar cada componente na tabela base do DynamoDB. Depois de inserir todos os componentes da árvore de exemplo, você obtém a tabela base a seguir.

ComponentId	ParentId	GraphId	Path
CM1		CM1#1	CM1

CM2	CM1	CM1#1	CM1|CM2
CM3	CM1	CM1#1	CM1|CM3
CM4	CM2	CM1#1	CM1|CM2|CM4
CM5	CM2	CM1#1	CM1|CM2|CM5
CM6	CM3	CM1#1	CM1|CM3|CM6
CM7	CM3	CM1#1	CM1|CM3|CM7
CM8	CM4	CM1#1	CM1|CM2|CM4|CM8
CM9	CM4	CM1#1	CM1|CM2|CM4|CM9
CM10	CM5	CM1#1	CM1| CM2 | CM5 | CM1 0

O GSI1 índice

Para verificar todos os filhos imediatos de um componente, você cria um índice que é usado `ParentId` como chave de partição e `ComponentId` como chave de classificação. A tabela dinâmica a seguir representa o GSI1 índice. Esse índice pode ser usado para recuperar todos os componentes filhos imediatos por meio de um ID de componente pai. Por exemplo, você pode descobrir quantas

baterias estão disponíveis em um carro (CM1) ou quais células estão disponíveis em um módulo (CM4).

ParentId	ComponentId
CM1	CM2
	CM3
CM2	CM4
	CM5
CM3	CM6
	CM7
CM4	CM8
	CM9
CM5	CM10

O GSI2 índice

A tabela dinâmica a seguir representa o GSI2 índice. Ela é configurada usando-se `GraphId` como uma chave de partição e`Path` como uma chave de classificação. Usando `GraphI` d e a `begins_with` operação na chave de classificação (`Path`), você pode encontrar a linhagem completa de um componente em uma árvore.

GraphId	Path	ComponentId
CM1#1	CM1	CM1
	CM1|CM2	CM2
	CM1|CM3	CM3
	CM1|CM2|CM4	CM4

CM1|CM2|CM5	CM5
CM1|CM2|CM4|CM8	CM8
CM1|CM2|CM4|CM9	CM9
CM1| CM2 | CM5 | CM1 0	CM10
CM1|CM3|CM6	CM6
CM1|CM3|CM7	CM7

Etapa 6: criar consultas de dados

Depois de definir seus padrões de acesso e projetar seu modelo de dados, você pode consultar dados hierárquicos no banco de dados do DynamoDB. Como prática recomendada para economizar custos e ajudar a garantir o desempenho, os exemplos a seguir usam somente a operação de consulta sem`Scan`.

- Encontre ancestrais de um componente.

  Para encontrar os ancestrais (pai, avô, bisavô etc.) do CM8 componente, consulte a tabela base usando. `ComponentId = "CM8"` A consulta retornará o registro a seguir.

  Para reduzir o tamanho dos dados resultantes, é possível usar uma expressão de projeção para retornar somente o atributo `Path`.

ComponentId	ParentId	GraphId	Path
CM8	CM4	CM1#1	CM1|CM2|CM4|CM8

  Path

  CM1|CM2|CM4|CM8

  Agora, divida o caminho usando o tubo (“|”) e pegue os primeiros componentes N-1 para obter ancestrais.

Resultado da consulta: Os ancestrais de CM8 são CM1, CM2, CM4.

- Encontre filhos imediatos de um componente.

  Para obter todos os componentes secundários imediatos ou de um nível a jusante do CM2 componente, consulte GSI1 usando. `ParentId = "CM2"` A consulta retornará o registro a seguir.

ParentId	ComponentId
CM2	CM4
	CM5

- Encontre todos os componentes secundários posteriores usando um componente de nível superior.

  Para obter todos os componentes secundários ou posteriores do componente de nível superior CM1, consulte GSI2 usando `GraphId = "CM1#1"` e e `begins_with("Path", "CM1|")` use uma expressão de projeção com. `ComponentId` Isso retornará todos os componentes relacionados a essa árvore.

  Este exemplo tem uma única árvore, com CM1 o componente superior. Na realidade, é possível ter milhões de componentes de nível superior na mesma tabela.

GraphId	ComponentId
	CM2
CM1#1	CM3
	CM4
	CM5
	CM8
	CM9
	CM10

CM6
CM7

- Encontre todos os componentes secundários posteriores usando um componente de nível médio.

  Para obter todos os componentes secundários ou posteriores recursivamente para componente CM2, você tem duas opções. Você pode consultar recursivamente nível por nível ou consultar o GSI2 índice.

  - Consulte GSI1, nível por nível, recursivamente, até atingir o último nível dos componentes secundários.

    1. Consulte GSI1 usando`ParentId = "CM2"`. Isso retornará o registro a seguir.

ParentId	ComponentId
CM2	CM4
	CM5

    2. Novamente, consulte GSI1 usando`ParentId = "CM4"`. Isso retornará o registro a seguir.

ParentId	ComponentId
CM4	CM8
	CM9

    3. Novamente, consulte GSI1 usando`ParentId = "CM5"`. Isso retornará o registro a seguir.

       Continue o loop: consulte para cada `ComponentId` até chegar ao último nível. Quando uma consulta usando `ParentId = "<ComponentId>"` não retorna nenhum resultado, o resultado anterior pertencia ao último nível da árvore.

ParentId	ComponentId
CM5	CM10

    4. Mescle todos os resultados.

resultado= [CM4, CM5] + [CM8, CM9] + [CM10]

= [CM4, CM5 CM8, CM9, CM1 0]

- Consulta GSI2, que armazena uma árvore hierárquica para um componente de nível superior (um carro ou). CM1

  1. Primeiro, encontre o componente de nível superior ou o ancestral superior e `Path` de. CM2 Para isso, consulte a tabela base usando `ComponentId = "CM2"` para encontrar o caminho desse componente na árvore hierárquica. Selecione os atributos `GraphId` e `Pat h`. A consulta retornará o registro a seguir.

GraphId	Path
CM1#1	CM1|CM2

  2. Consulte GSI2 usando`GraphId = "CM1#1" AND BEGINS_WITH("Path", "CM1|CM2|")`. A consulta retornará os resultados a seguir.

GraphId	Path	ComponentId
CM1#1	CM1|CM2|CM4	CM4
	CM1|CM2|CM5	CM5
	CM1|CM2|CM4|CM8	CM8
	CM1|CM2|CM4|CM9	CM9
	CM1| CM2 | CM5 | CM1 0	CM10

  3. Selecione o `ComponentId` atributo para o qual devolver todos os componentes secundários CM2.

Etapa 7: Validar o modelo de dados

Nessa etapa, o usuário corporativo valida os resultados da consulta e verifica se eles atendem às necessidades comerciais. Você pode usar a tabela a seguir para verificar os padrões de acesso em relação aos requisitos do usuário.

Pergunta	Tabela base//GSI	Consulta
Como usuário, quero recuperar todos os componentes filhos imediatos para um ID de componente pai.	GSI1	`ParentId = "<ComponentId>"` (Encontrar filhos imediatos de um componente.)
Como usuário, quero recuperar uma lista recursiva de todos os componentes filhos para um ID de componente.	GSI1 or GSI2	GSI1: `ParentId = "<ComponentId>"` or GSI2: `GraphId = "<TopLevelComponentId>#N" AND BEGINS_WITH("Path", "<PATH_OF_Component>")` (Encontrar todos os componentes filhos de nível inferior usando um componente de nível superior. Encontrar todos os componentes filhos de nível inferior usando um componente de nível intermediário.)
Como usuário, quero ver os ancestrais de um componente.	Tabela base	`ComponentId = "<ComponentId>"` e, em seguida, selecione o atributo Path.

(Encontrar ancestrais de um componente.)

Você também pode implementar um script (teste) em qualquer linguagem de programação para consultar o DynamoDB diretamente e comparar os resultados com os resultados esperados.

Etapa 8: Revise a estimativa de custo

Revise e refine a estimativa de custo novamente. Além disso, é uma boa prática validá-la com as partes interessadas da empresa e obter aprovação para passar para a próxima etapa.

Objetivos

- Defina o modelo de capacidade e estime os custos do DynamoDB para refinar a estimativa de custo a partir da etapa 2.
- Obtenha a aprovação financeira final do analista de negócios e das partes interessadas.

Processo

- O engenheiro de banco de dados identifica a estimativa do volume de dados.
- O engenheiro de banco de dados identifica os requisitos de transferência de dados.
- O engenheiro de banco de dados define as unidades de capacidade de leitura e gravação necessárias.
- O analista de negócios decide entre modelos de capacidade sob demanda e provisionados.
- O engenheiro de banco de dados identifica a necessidade de ajuste de escala automático do DynamoDB.
- O engenheiro de banco de dados insere os parâmetros no AWS Calculadora de Preços.
- O engenheiro de banco de dados apresenta a estimativa final do preço às partes interessadas do negócio.
- O analista de negócios e as partes interessadas aprovam ou rejeitam a solução.

Etapa 9: implantar o modelo de dados

Neste exemplo específico, a implantação do modelo foi feita usando o NoSQL Workbench, um aplicativo para desenvolvimento e operação de bancos de dados modernos. Usando essa ferramenta, você tem a opção de criar um modelo de dados, fazer upload de dados e implantá-los diretamente no seu. Conta da AWS Se quiser implementar este exemplo, você pode usar o AWS CloudFormation modelo a seguir, que foi gerado pelo NoSQL Workbench.

```
AWSTemplateFormatVersion: 2010-09-09
Resources:
  Components:
    Type: 'AWS::DynamoDB::Table'
    Properties:
      KeySchema:
        - AttributeName: ComponentId
          KeyType: HASH
      AttributeDefinitions:
        - AttributeName: ComponentId
          AttributeType: S
        - AttributeName: ParentId
          AttributeType: S
        - AttributeName: GraphId
          AttributeType: S
        - AttributeName: Path
          AttributeType: S
      GlobalSecondaryIndexes:
        - IndexName: GS1
          KeySchema:
            - AttributeName: ParentId
              KeyType: HASH
            - AttributeName: ComponentId
              KeyType: RANGE
          Projection:
            ProjectionType: KEYS_ONLY
        - IndexName: GSI2
          KeySchema:
            - AttributeName: GraphId
              KeyType: HASH
            - AttributeName: Path
              KeyType: RANGE
          Projection:
            ProjectionType: INCLUDE
```

```
        NonKeyAttributes:
          - ComponentId
    BillingMode: PAY_PER_REQUEST
    TableName: Components
```

Recursos adicionais

Mais informações sobre o DynamoDB

- Preços do DynamoDB
- Documentação do DynamoDB
- Design em NoSQL para DynamoDB
- Fragmentação de gravação
- Índices secundários locais () LSIs
- Índices secundários globais () GSIs
- Sobrecarga GSIs
- Fragmentação do GSI
- Usando GSIs para criar uma réplica eventualmente consistente
- Índices esparsos
- Consultas de agregação materializadas
- Padrão de design para séries temporais
- Padrão de design da lista de adjacências
- O modo de capacidade provisionado e sob demanda
- Ajuste de escala automático do DynamoDB
- Vida útil (TTL) do DynamoDB
- Modelagem de dados de jogadores de jogos com o DynamoDB (laboratório)

AWS serviços

- AWS CloudFormation
- Amazon S3

Ferramentas

- AWS Calculadora de Preços
- NoSQL Workbench para DynamoDB
- DynamoDB Local

- DynamoDB e AWS SDKs

Práticas recomendadas

- Práticas recomendadas de design e arquitetura com o DynamoDB (documentação do Dynamo DB)
- Práticas recomendadas para uso de índices secundários (documentação do DynamoDB)
- Práticas recomendadas para armazenar itens e atributos grandes (documentação do DynamoDB)
- Escolhendo a chave AWS de partição correta do DynamoDB (blog do banco de dados)
- Como criar índices DBglobal secundários do Amazon Dynamo (blog do AWS banco de dados)
- Quais são as facetas do NoSQL Workbench para Amazon DynamoDB (site Medium)

AWS recursos gerais

- AWS Site de orientação prescritiva
- AWS documentação
- AWS referência geral

Colaboradores

Os colaboradores deste guia incluem:

- Camilo Gonzalez, arquiteto de dados sênior, AWS
- Moinul Al-Mamun, arquiteto sênior de Big Data, AWS
- Santiago Segura, consultor de serviços profissionais, AWS
- Satheish Kumar Chandraprakasam, arquiteto de aplicativos em nuvem, AWS

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
Foi adicionada uma seção de melhores práticas e um exemplo para modelagem hierárquica de dados.	Adicionamos um resumo das melhores práticas do DynamoDB e step-by-step um exemplo de design e validação de um modelo hierárquico.	5 de dezembro de 2023
Publicação inicial	—	26 de outubro de 2020

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: migre seu banco de dados Oracle local para a edição compatível com o Amazon Aurora 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: Migre seu banco de dados Oracle local para o Amazon Relational Database Service (Amazon RDS) for Oracle no. Nuvem AWS
- Recomprar (drop and shop): mude para um produto diferente, normalmente migrando de uma licença tradicional para um modelo SaaS. Exemplo: migre seu sistema de gerenciamento de relacionamento com o cliente (CRM) para a 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: Migre seu banco de dados Oracle local para o Oracle em uma EC2 instância no. 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 local para um serviço em nuvem para a mesma plataforma. Exemplo: migrar um Microsoft Hyper-V aplicativo para 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 atributos.

serviços abstratos

Veja os 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 no qual 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.

função agregada

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 a inteligência artificial.

AIOps

Veja as 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 aplicativos

Uma abordagem de segurança que permite o uso somente de aplicativos aprovados 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 ruim

Um bot destinado a perturbar ou causar danos a indivíduos ou organizações.

BCP

Veja o 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 do aplicativo em um ambiente (azul) e a nova versão do aplicativo no outro ambiente (verde). Essa estratégia ajuda você a reverter rapidamente com o mínimo de impacto.

bot

Um aplicativo de software que executa tarefas automatizadas pela Internet e simula a atividade ou interação humana. Alguns bots são úteis ou benéficos, como rastreadores da Web que indexam informações na Internet. Alguns outros bots, conhecidos como bots ruins, têm como objetivo perturbar ou causar danos a indivíduos ou organizações.

botnet

Redes de bots infectadas por malware e sob o controle de uma única parte, conhecidas como pastor de bots ou operador de bots. As redes de bots 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 em vidro quebrado

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

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

Consulte Estrutura de adoção da AWS nuvem.

implantação canária

O lançamento lento e incremental de uma versão para usuários finais. Quando estiver confiante, você implanta a nova versão e substituirá a versão atual em sua totalidade.

CCoE

Veja o Centro de Excelência em Nuvem.

CDC

Veja a captura de dados de alterações.

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 a integração e a 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 Empresarial.

computação em nuvem

A tecnologia de nuvem normalmente usada para armazenamento de dados remoto e gerenciamento de dispositivos de IoT. A computação em nuvem geralmente está conectada à tecnologia de computação de ponta.

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 quando migram para o 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

Consulte o 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 GitHub or 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 da IA que usa aprendizado de máquina para analisar e extrair informações de formatos visuais, como imagens e vídeos digitais. Por exemplo, AWS Panorama oferece dispositivos que adicionam CV às redes de câmeras locais, e a Amazon SageMaker AI fornece algoritmos de processamento de imagem para CV.

desvio de configuração

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

banco de dados de gerenciamento de configuração (CMDB)

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

pacote de conformidade

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

integração contínua e entrega contínua (CI/CD)

O processo de automatizar os estágios de origem, criação, teste, preparação e produção do processo de lançamento do software. CI/CD is commonly described as a pipeline. CI/CDpode 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.

malha de dados

Uma estrutura arquitetônica 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 que oferece suporte à inteligência comercial, como análises. 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

Consulte a 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 o 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 são comumente usados para restringir consultas, filtrar e rotular conjuntos de resultados.

desastre

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

Recuperação de desastres (RD)

A estratégia e o processo que você usa para minimizar o tempo de inatividade e a perda de dados causados por um desastre. Para obter mais informações, consulte Recuperação de desastres de cargas de trabalho em AWS: Recuperação na nuvem no AWS Well-Architected Framework.

DML

Veja a 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 a recuperação de desastres.

detecção de deriva

Rastreando desvios de uma configuração básica. 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 o mapeamento do fluxo de valor do desenvolvimento.

E

EDA

Veja a 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 à computação em nuvem, a computação de ponta 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 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 o endpoint do 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 corporativos (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 pipeline de CI/CD, 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 o planejamento de recursos corporativos.

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. Ele armazena dados quantitativos sobre as operações comerciais. Normalmente, uma tabela de fatos contém dois tipos de colunas: aquelas que contêm medidas e aquelas que contêm uma chave externa para uma tabela de dimensões.

falham rapidamente

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

limite de isolamento de 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 Limites de isolamento de AWS falhas.

ramificação de recursos

Veja a filial.

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.

solicitação de alguns instantes

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

FGAC

Veja o 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 por meio da captura de dados alterados para migrar 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 o modelo da fundação.

modelo de fundação (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 básicos.

G

IA generativa

Um subconjunto de modelos de IA que foram treinados em grandes quantidades de dados e que podem usar uma simples solicitação de texto para criar novos conteúdos e artefatos, como imagens, vídeos, texto e áudio. Para obter mais informações, consulte O que é IA generativa.

bloqueio geográfico

Veja as 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 baseado em troncos é a abordagem moderna e preferida.

imagem dourada

Um instantâneo de um sistema ou software usado como modelo para implantar novas instâncias desse sistema ou software. Por exemplo, na manufatura, uma imagem dourada 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, Amazon GuardDuty AWS Trusted Advisor, Amazon Inspector e verificações personalizadas AWS Lambda .

H

HA

Veja a 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 retenção

Uma parte dos dados históricos rotulados que são retidos de um conjunto de dados usado para treinar um modelo de aprendizado de máquina. Você pode usar dados de retenção para avaliar o desempenho do modelo comparando as previsões 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 típico de uma DevOps versão.

período de hipercuidados

Imediatamente após a substituição, o período em que uma equipe de migração gerencia e monitora as aplicações migradas na nuvem para resolver quaisquer problemas. Normalmente,

a duração desse período é de 1 a 4 dias. No final do período de hipercuidados, a equipe de migração normalmente transfere a responsabilidade pelas aplicações para a equipe de operações de nuvem.

eu

IaC

Veja a 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 a Internet das Coisas industrial.

infraestrutura imutável

Um modelo que implanta uma nova infraestrutura para cargas de trabalho de produção em vez de atualizar, corrigir ou modificar a infraestrutura existente. Infraestruturas imutáveis são inerentemente mais consistentes, confiáveis e previsíveis do que infraestruturas mutáveis. Para obter mais informações, consulte as melhores práticas de implantação usando infraestrutura imutável no Well-Architected AWS 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 fabricação por meio de avanços em conectividade, dados em tempo real, automação, análise 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

Consulte 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

Consulte a biblioteca de informações de TI.

ITSM

Veja o 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.

modelo de linguagem grande (LLM)

Um modelo de IA de aprendizado profundo que é pré-treinado em uma grande quantidade de dados. Um LLM pode realizar várias tarefas, como responder 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 rótulos.

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 um modelo de linguagem grande.

ambientes inferiores

Veja o 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 a filial.

malware

Software projetado para comprometer a segurança ou a privacidade do computador. O malware pode interromper os sistemas do computador, vazar informações confidenciais ou obter acesso não autorizado. Exemplos de malware incluem vírus, worms, ransomware, cavalos de Tróia, 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 abstratos.

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

Consulte Migration Acceleration Program.

mecanismo

Um processo completo no qual 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 o 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 a Amazon EC2 com o AWS 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 o. 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 carga de trabalho para o. Nuvem AWS Para obter mais informações, consulte a entrada de 7 Rs neste glossário e consulte Mobilize sua organização para acelerar migrações em grande escala.

ML

Veja o aprendizado de máquina.

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 Estratégia para modernizar aplicativos no 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 Avaliação da prontidão para modernização de aplicativos no. 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.

MAPA

Consulte Avaliação do portfólio de migração.

MQTT

Consulte 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 cargas de trabalho 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 o controle de acesso de origem.

CARVALHO

Veja a identidade de acesso de origem.

OCM

Veja o gerenciamento de mudanças 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 a integração de operações.

OLA

Veja o contrato em 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

Consulte Comunicação de processo aberto — Arquitetura unificada.

Comunicação de processo aberto — Arquitetura unificada (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 melhores práticas 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 Well-Architected AWS Framework.

tecnologia operacional (OT)

Sistemas de hardware e software que funcionam com o ambiente físico para controlar operações, equipamentos e infraestrutura industriais. Na manufatura, a integração dos sistemas OT e de tecnologia da informação (TI) 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 a análise de prontidão operacional.

OT

Veja a 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 as 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

Consulte controlador lógico programável.

AMEIXA

Veja o gerenciamento do ciclo de vida do produto.

política

Um objeto que pode definir permissões (consulte a política baseada em identidade), especificar as condições de acesso (consulte a política baseada em recursos) ou definir as permissões máximas para todas as contas em uma organização em AWS Organizations (consulte a 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. Para obter mais informações, consulte Habilitar a persistência de dados em microsserviços.

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 `WHERE` cláusula.

pressão de predicados

Uma técnica de otimização de consulta 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 o desempenho 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 de engenharia de sistema 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 projetado para impedir a implantação de recursos não compatíveis. Esses controles examinam 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 o design, desenvolvimento e lançamento, passando pelo crescimento e maturidade, até o declínio e a remoção.

ambiente de produção

Veja o ambiente.

controlador lógico programável (PLC)

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

encadeamento imediato

Usando a 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 no qual outros microsserviços possam se inscrever. 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, responsável, consultado, informado (RACI).

RAG

Consulte Geração Aumentada de 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, responsável, consultado, informado (RACI).

RCAC

Veja o 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.

rearquiteta

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 mais informações, consulte Especificar o que Regiões da AWS sua conta pode usar.

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 um aplicativo de resistir ou se recuperar de interrupções. Alta disponibilidade e recuperação de desastres são considerações comuns ao planejar a resiliência no. Nuvem AWS Para obter mais informações, consulte Nuvem AWS Resiliência.

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.

aposentar-se

Veja 7 Rs.

Geração Aumentada de Recuperação (RAG)

Uma tecnologia de IA generativa na qual 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.

alternância

O processo de atualizar periodicamente um segredo para dificultar o acesso das credenciais por um invasor.

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 o objetivo do ponto de recuperação.

RTO

Veja o objetivo do 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 AWS Management Console 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 a política de controle de serviços.

secret

Em AWS Secrets Manager, informações confidenciais ou restritas, como uma senha ou credenciais de usuário, que você armazena de forma criptografada. Ele 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 O que há em um segredo do Secrets Manager? na documentação do Secrets Manager.

segurança por design

Uma abordagem de engenharia de sistema 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 principais de controles de segurança: preventivos, detectivos, responsivos e proativos.

fortalecimento 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 correção de uma instância EC2 da Amazon ou a rotação 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 medida de um aspecto de desempenho de um serviço, como taxa de erro, disponibilidade ou taxa de transferência.

objetivo de nível de serviço (SLO)

Uma métrica alvo que representa a integridade de um serviço, conforme medida 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 informações de segurança e sistema de gerenciamento de eventos.

ponto único de falha (SPOF)

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

SLA

Veja o contrato de nível de serviço.

ESGUIO

Veja o indicador de nível de serviço.

SLO

Veja o objetivo do 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 Abordagem em fases para modernizar aplicativos no. Nuvem AWS

CUSPE

Veja um único ponto de falha.

esquema de estrelas

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 uso 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 de supervisão 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 o desempenho. Você pode usar o Amazon CloudWatch Synthetics para criar esses testes.

prompt do sistema

Uma técnica para fornecer contexto, instruções ou diretrizes a um LLM para direcionar seu comportamento. Os prompts do sistema ajudam a definir o contexto e 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. 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 o 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 o 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 back-end.

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.

MINHOCA

Veja escrever uma vez, ler muitas.

WQF

Consulte Estrutura de qualificação AWS da carga de trabalho.

escreva uma vez, leia muitas (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, geralmente malware, que tira proveito de uma vulnerabilidade de dia zero.

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.

aviso de disparo zero

Fornecer a um LLM instruções para realizar uma tarefa, mas sem exemplos (fotos) que possam ajudar a orientá-la. O LLM deve usar seu conhecimento pré-treinado para lidar com a tarefa. A

eficácia da solicitação zero depende da complexidade da tarefa e da qualidade da solicitação. Veja também a solicitação de algumas fotos.

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.