

Armazenamento e Banco de Dados

Willyan Guimarães - @willyancaetanodev

Objetivo Geral

Conhecer os principais serviços de Armazenamento e Banco de Dados da AWS e compreender seus casos de uso alvo.

Percurso

Etapa 1

Armazenamento de dados em nuvem

Etapa 2

Amazon Elastic Block Store - EBS

Etapa 3

Amazon S3

Percurso

Etapa 4

EFS – Amazon Elastic File System

Etapa 5

Amazon Relational Database Service

Etapa 6

DynamoDB

Percurso

Etapa 7

Outros serviços de banco de dados

Etapa 8

Big Data com Amazon Redshift

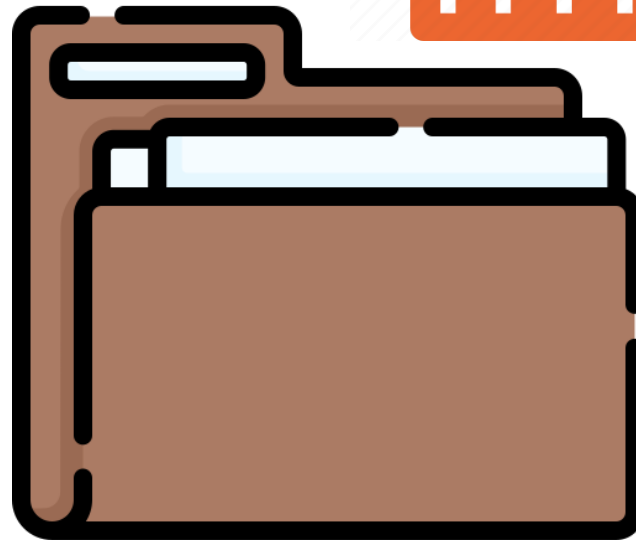
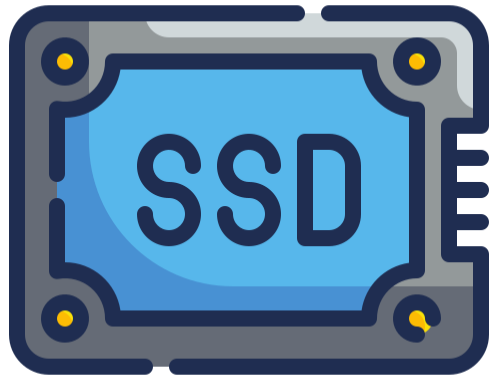
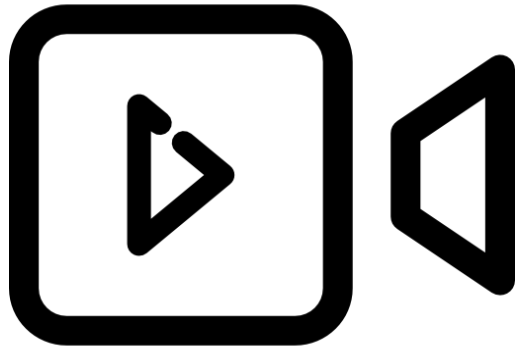
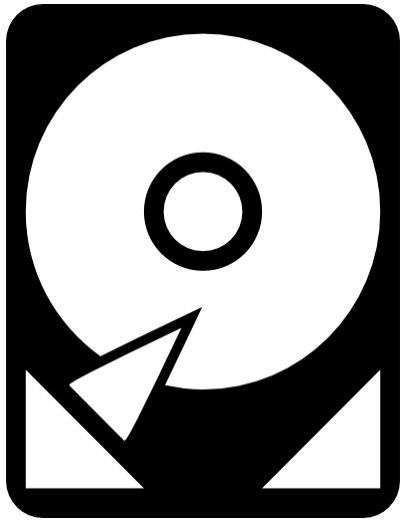
Pré-requisitos

- Nenhum pré-requisito específico

Etapa 1

Armazenamento de Dados em Nuvem

E os dados? O quê, onde e como?



Tipos de Armazenamento

- Armazenamento de Objetos (Object Storage)
- Armazenamento de Arquivos (File Storage)
- Armazenamento de Blocos (Block Storage)

Armazenamento de Objetos

- Object Storage
- Dados como objetos (arquivos e metadados)
- Dados não estruturados
- Casos de uso: Data lakes, Mídias, Backup e recuperação

Armazenamento de Arquivos

- File Storage
- Sistemas de arquivos compartilhados
- Permite acesso por meio de servidores, aplicações e usuários
- Analogia com pastas compartilhadas em uma rede
- Casos de uso: Ferramentas de desenvolvimento, Diretórios pessoais

Armazenamento de Blocos

- Block Storage
- Armazenamento de blocos: HDD, SSD
- Dispositivo com diferentes configurações de Leitura e Escrita
- Casos de uso: Máquinas virtuais, contêiners, banco de dados

Para saber mais

- <https://aws.amazon.com/pt/what-is/cloud-storage/>
- <https://aws.amazon.com/pt/what-is/cloud-file-storage/>
- <https://aws.amazon.com/pt/what-is/object-storage/>
- <https://aws.amazon.com/pt/what-is/block-storage/>

Dúvidas?

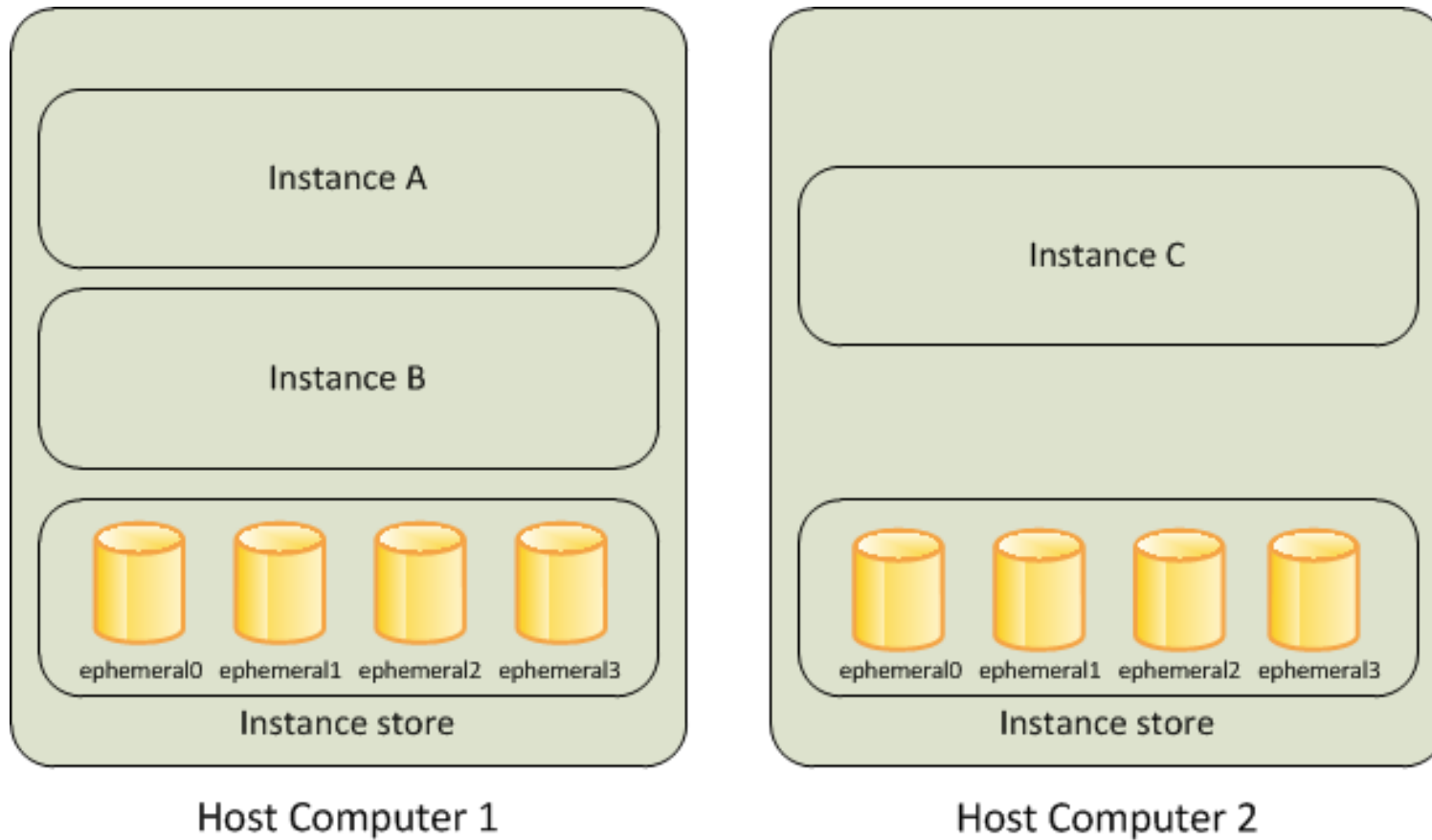
- > Fórum/Artigos
- > Comunidade Online (Discord)



Etapa 2

Amazon Elastic Block Store - EBS

Usando EC2



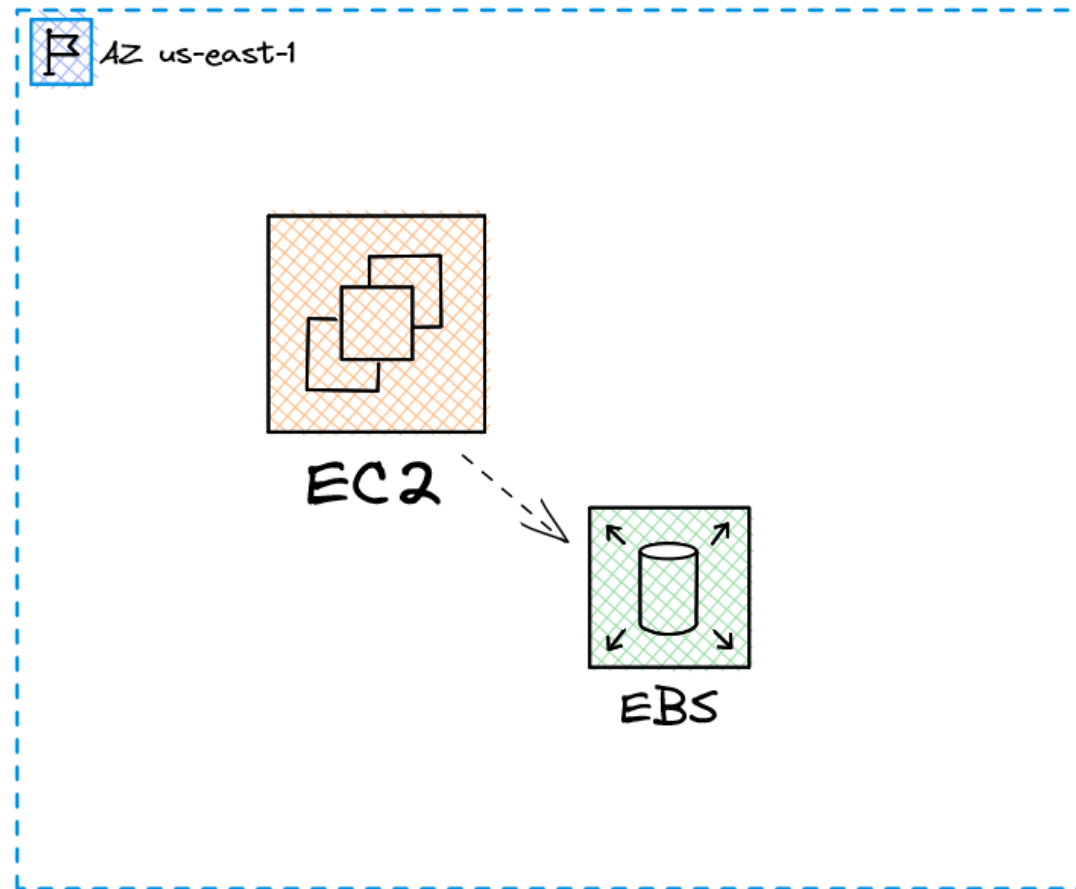
Volume Instance Store

- Armazenamento de Blocos
- Discos anexados fisicamente ao computador host
- Ideal para dados de armazenamento temporário como buffers, caches, dados de rascunho

Dados serão perdidos se

- Falha de disco de uma unidade
- Instância parada
- Instância hiberna
- Instância encerrada

Amazon Elastic Block Store - EBS



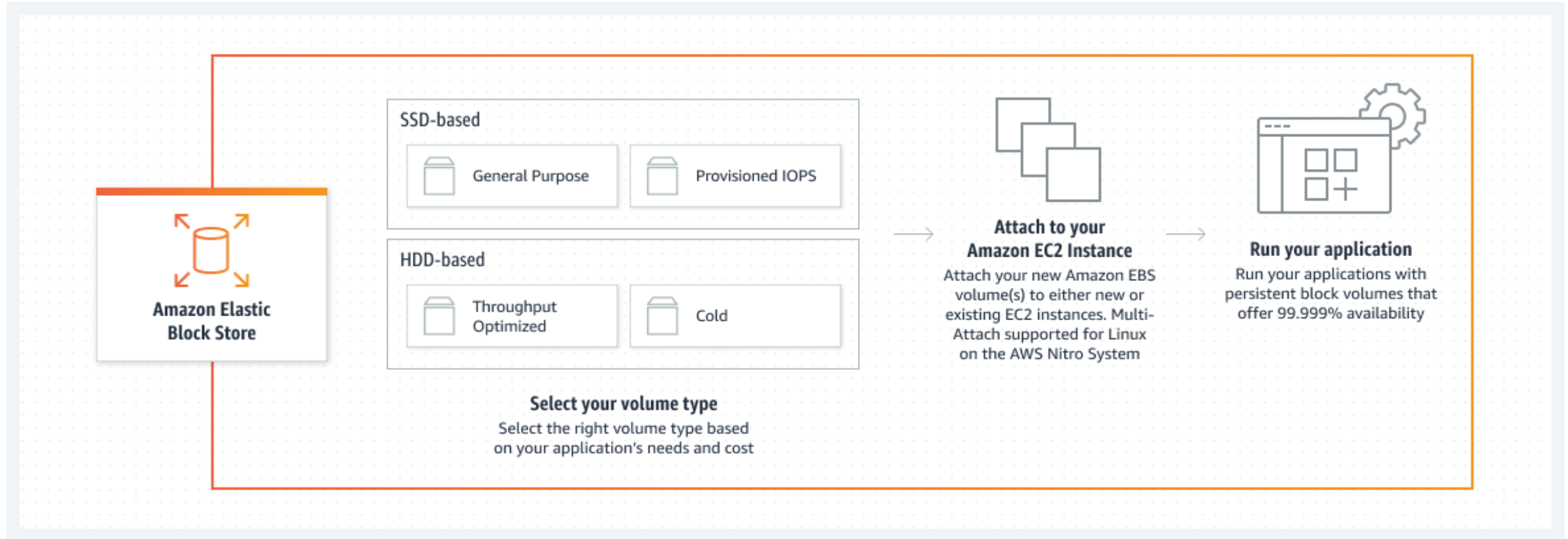
EBS – Elastic Block Store

- Armazenamento em Blocos
- Block, blocos = HD, físico
- Projetado para Amazon Elastic Compute Cloud (EC2)
- HDs são chamados "volumes"

Como funciona

1. Defina o tipo do volume
2. Escolha tamanho e configurações
3. Anexe o volume a uma instância EC2

Mais ou menos assim



HDD

- Mais lento
- Mais barato
- Dois tipos: Disco rígido frio e otimizado para throughput

SSD

- Mais rápido
- Mais caro
- Dois tipos: Volumes SSD de uso geral, IOPS provisionados

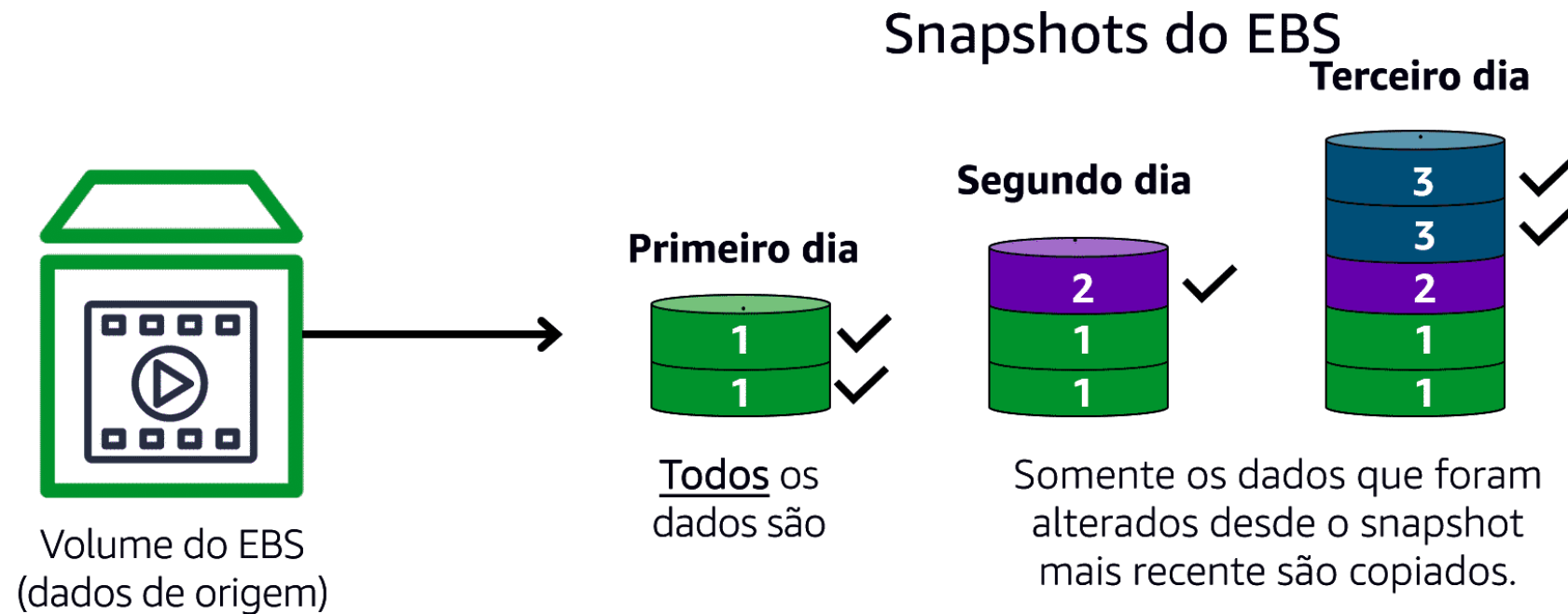
Lembre-se

SSD	HDD
Maior custo	Menor custo
Mais rápido	Mais lento
Tipos: Uso geral, IOPS provisionados	Tipos: Disco rígido frio e otimizado para throughput

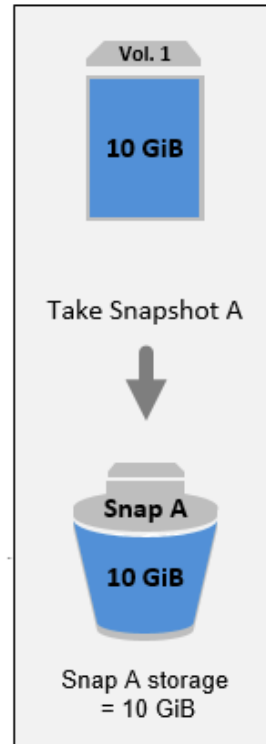
Como funcionam os backups?

- Snapshots
- Backup Incremental

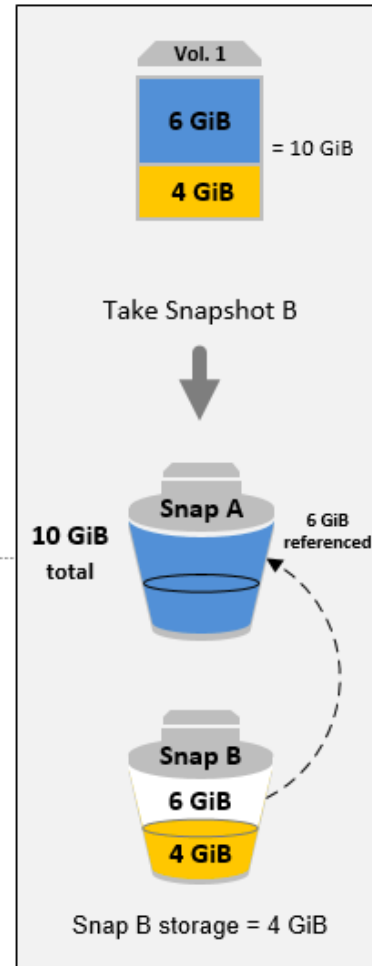
Como funcionam os backups?



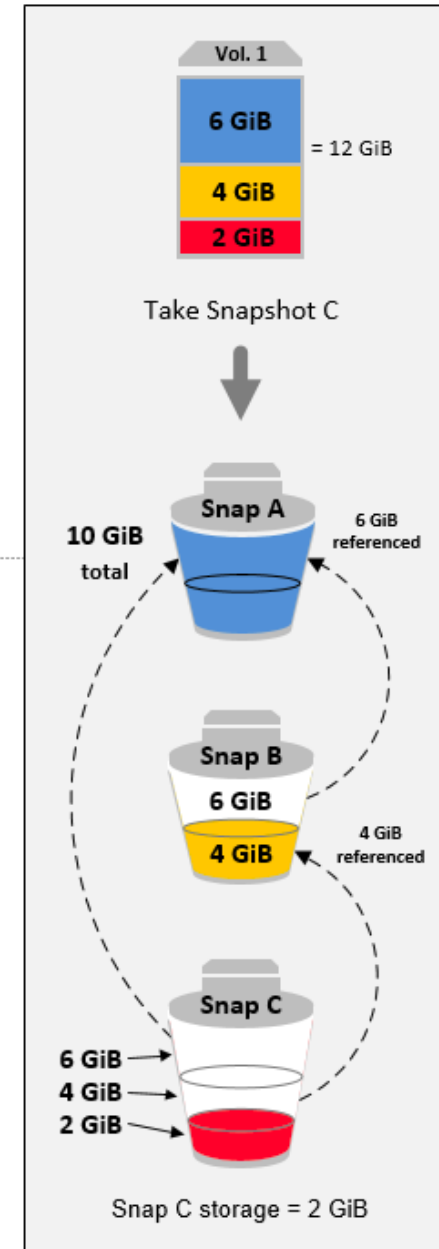
State 1: 10 GiB

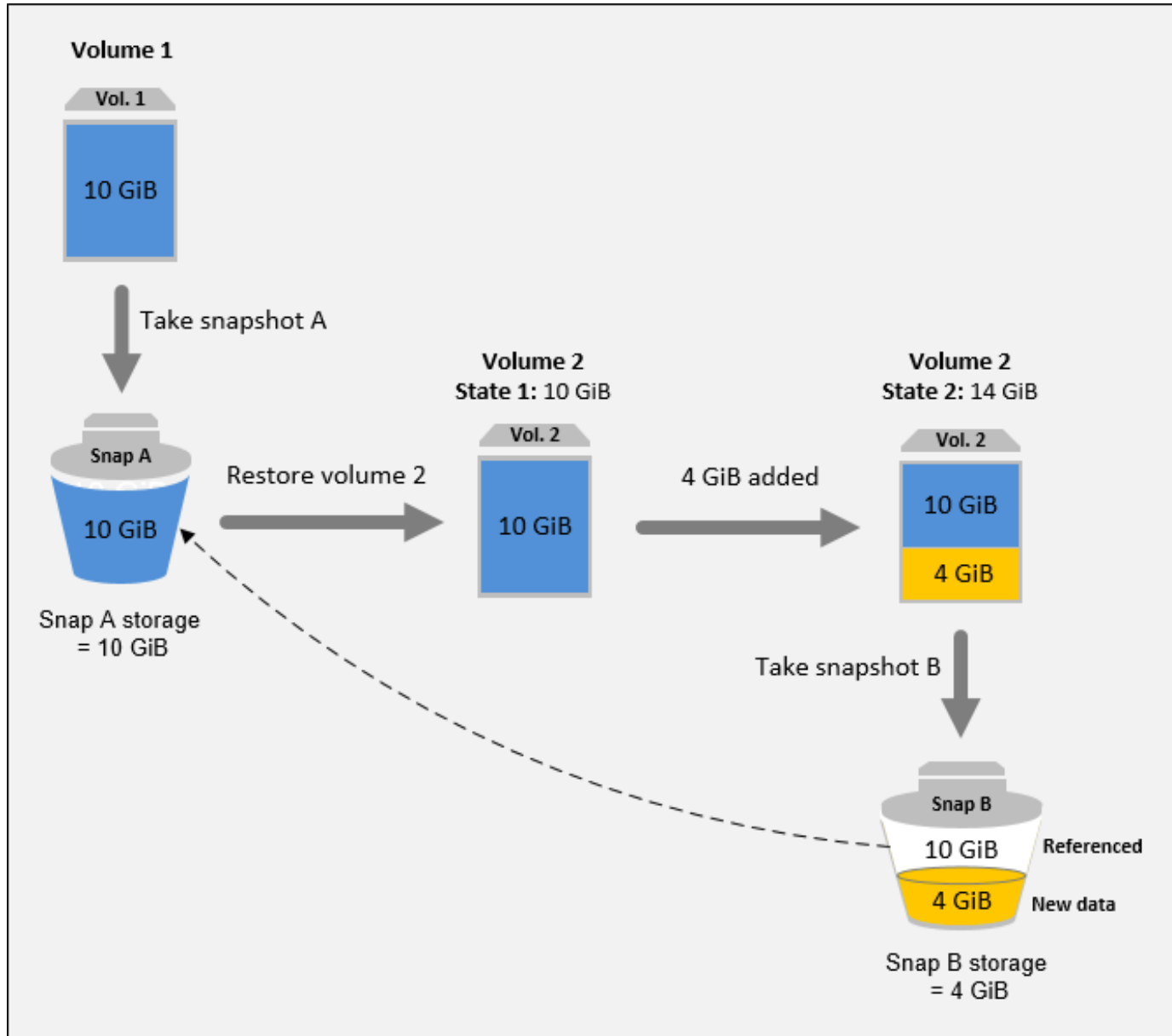


State 2: 4 GiB changed



State 3: 2 GiB added





Para saber mais

- https://docs.aws.amazon.com/pt_br/AWSEC2/latest/UserGuide/InstanceStorage.html
- https://docs.aws.amazon.com/pt_br/AWSEC2/latest/UserGuide/AmazonEBS.html
- https://docs.aws.amazon.com/pt_br/AWSEC2/latest/UserGuide/ebs-volume-types.html
- https://docs.aws.amazon.com/pt_br/AWSEC2/latest/UserGuide/EBSSnapshots.html

Dúvidas?

- > Fórum/Artigos
- > Comunidade Online (Discord)



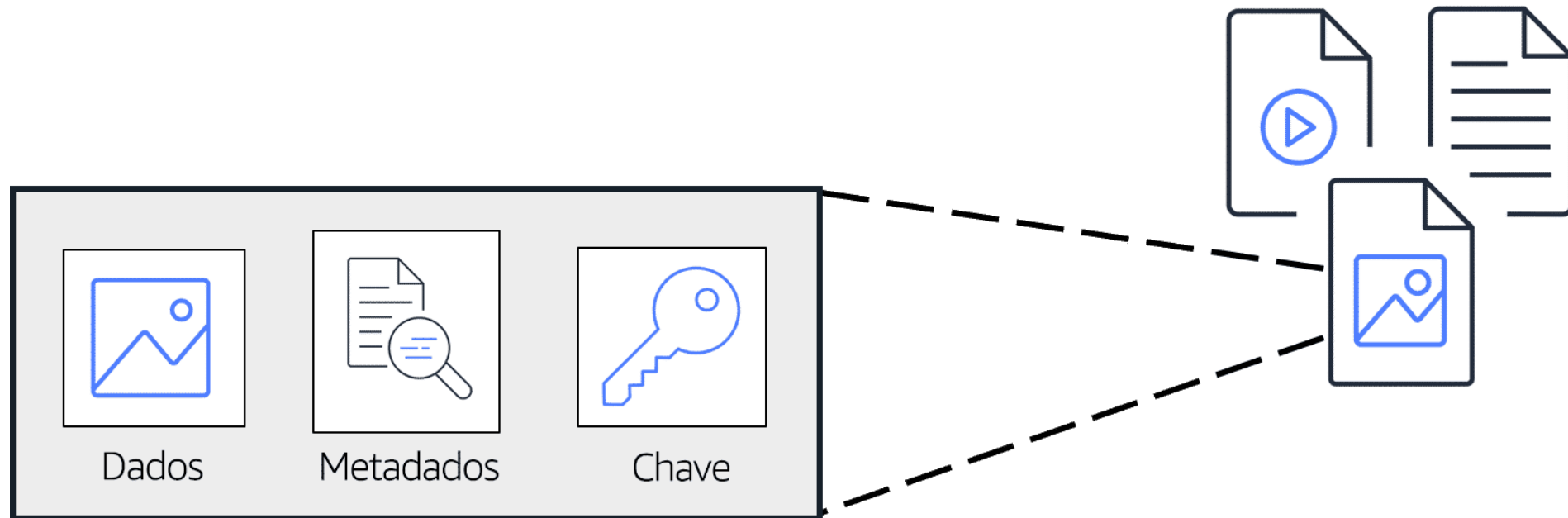
Etapa 3

Amazon S3

O que é Amazon S3 ?

- Serviço de armazenamento de objetos
- S3 – Simple Storage Service

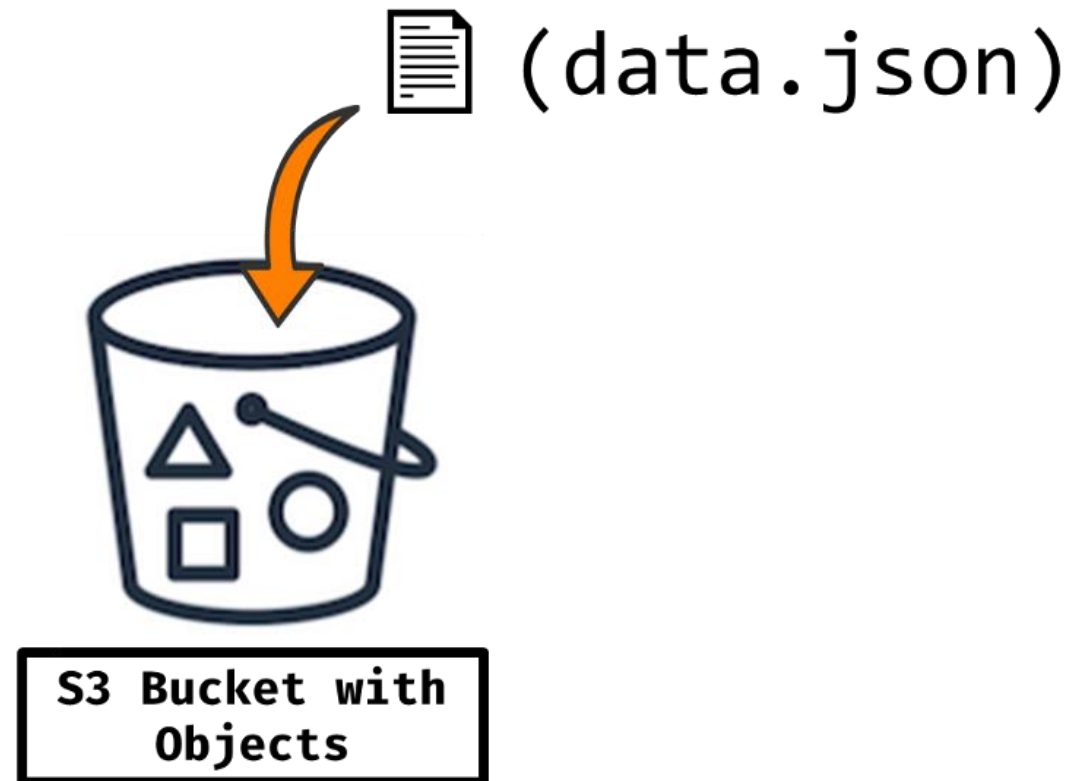
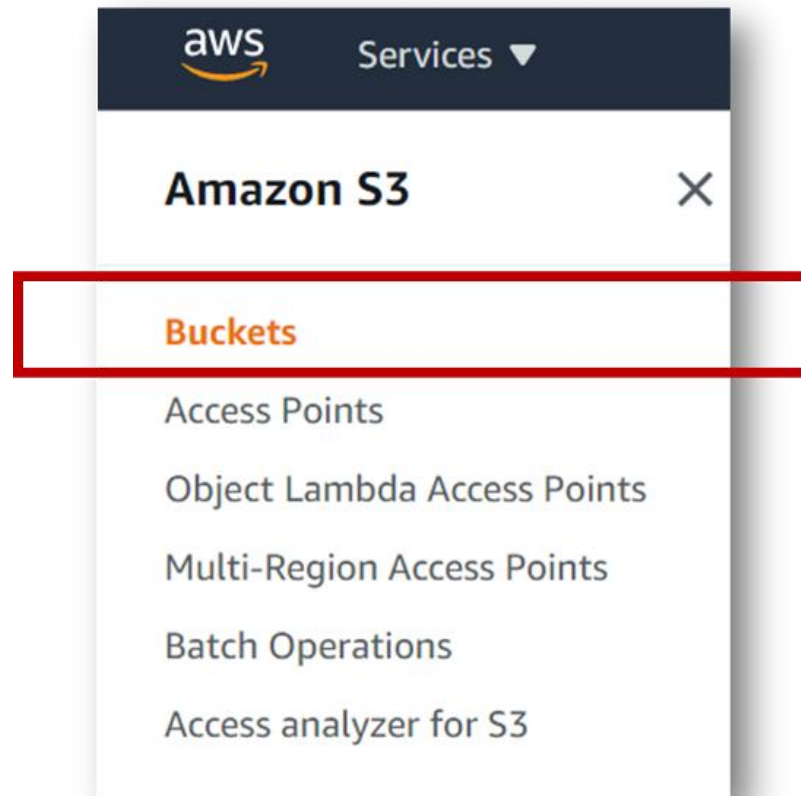
O que é um objeto no S3?



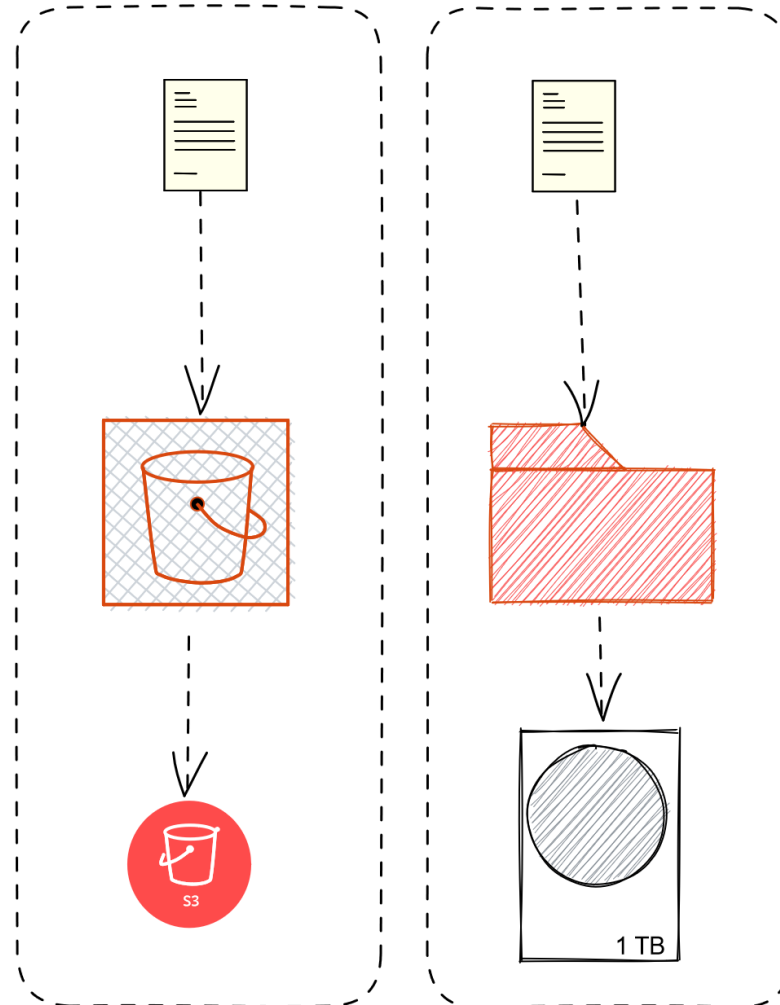
Composição de um objeto

- Chave: Nome que você atribui ao objeto. Usado para recuperar o objeto.
- Valor: O conteúdo que você está armazenando.
- Metadados: Um conjunto de pares de nome-valor com o qual é possível armazenar informações relacionadas ao objeto

Onde estão os objetos ?



Analogia



Buckets S3

- Antes de um upload do seus objetos, você precisa criar um
- É um contêiner para objetos armazenados no Amazon S3.
- Você pode armazenar qualquer número de objetos em um bucket
- Objetos podem ter de 0 até 5TB de tamanho
- Você pode ter até 100 buckets na sua conta

Você ainda pode

- Controlar acesso por objeto
- Utilizar versionamento de objetos

Casos de uso

- Data lakes
- Arquivamento de dados
- Hospedagem de sites estáticos

Classes de armazenamento

- Categorias para adequar melhor as necessidades de negócio e custo
- Fatores importantes na seleção de uma categoria:
- Com que frequência você planeja recuperar seus dados ?
- Seus dados precisam estar muito ou pouco disponíveis ?

Classes de armazenamento

- S3 Standard
- S3 Standard-Infrequent Access (S3 Standard-IA)
- S3 One Zone-Infrequent Access (S3 One Zone – IA)
- S3 Intelligent-Tiering
- S3 Glacier Instant Retrieval

Classes de armazenamento

- S3 Glacier Flexible Retrieval
- S3 Glacier Deep Archive

S3 Standard

- Projetado para dados acessados com frequência
- Armazena dados em um mínimo de três Zonas de Disponibilidade
- Boa escolha para diversos casos de uso como sites, distribuição de conteúdo e análise de dados
- Custo mais alto

S3 Standard-Infrequent Access

- Semelhante ao S3 Standard
- Armazena dados em um mínimo de três Zonas de Disponibilidade
- Ideal para dados acessados com pouca frequência
- Taxa por GB de armazenamento e recuperação mais baixo

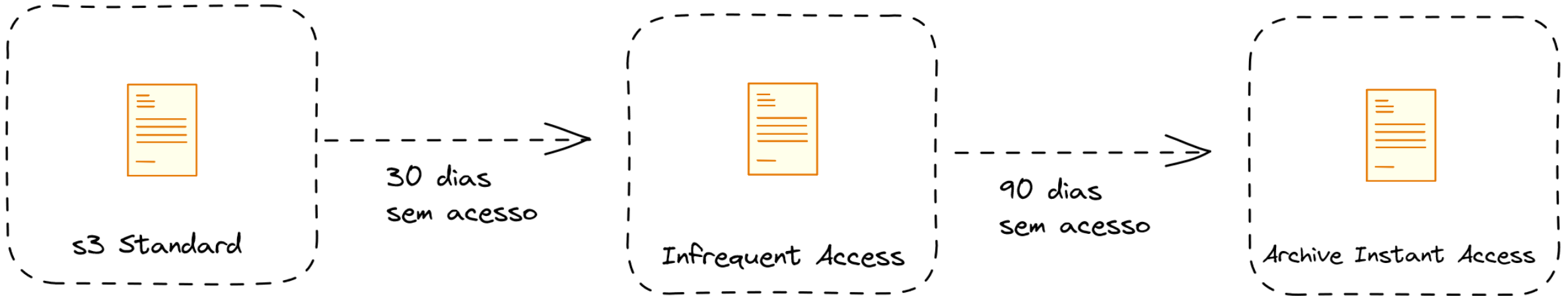
S3 One Zone-Infrequent Access

- Tem um preço de armazenamento menor do que o S3 Standard – IA
- Armazena dados em uma única Zona de Disponibilidade
- Cenários: Você quer economizar custos com armazenamento e você pode reproduzir facilmente seus dados em caso de falha na Zona de Disponibilidade

S3 Intelligent-Tiering

- Ideal para dados com padrões de acesso desconhecidos ou em alteração
- Gerencia automaticamente o ciclo de vida dos objetos armazenados otimizando custos
- Requer uma pequena taxa mensal de monitoramento e automação por objeto

Como funciona



S3 Glacier Instant Retrieval

- Ideal para dados de longa duração, raramente acessados mas que exigem recuperação rápida (milissegundos)
- Oferece acesso tão rápido quanto Standard e Standard-IA
- Ideal para dados acessados uma vez por trimestre

S3 Glacier Flexible Retrieval

- Para dados que não requerem acesso imediato
- Ideal para casos de uso de backups não urgentes, recuperação de desastres
- Usuário pode escolher qual velocidade de recuperação
- Ideal para dados acessados 1 ou 2 vezes por ano

S3 Glacier Deep Archive

- Suporte a retenção e preservação digital de longo prazo para dados que podem ser acessados 1 ou 2 vezes por ano
- Ideal para empresas que precisam manter dados por conformidades legais por 7 a 10 anos
- Recuperação de dados em até 12 horas

Para saber mais

- O que é Amazon S3 -
https://docs.aws.amazon.com/pt_br/AmazonS3/latest/userguide/Welcome.html
- Visão geral de objetos Amazon S3 -
https://docs.aws.amazon.com/pt_br/AmazonS3/latest/userguide/UsingObjects.html
- Trabalhar com metadados de objeto -
https://docs.aws.amazon.com/pt_br/AmazonS3/latest/userguide/UsingMetadata.html

Para saber mais

- Visão geral dos buckets - https://docs.aws.amazon.com/pt_br/AmazonS3/latest/userguide/UsingBucket.html
- Usando o versionamento em buckets do S3 - https://docs.aws.amazon.com/pt_br/AmazonS3/latest/userguide/Versioning.html
- Classes de armazenamento - <https://aws.amazon.com/pt/s3/storage-classes/>

Para saber mais

- Como o S3 Intelligent-Tiering funciona - https://docs.aws.amazon.com/pt_br/AmazonS3/latest/userguide/intelligent-tiering-overview.html

Dúvidas?

- > Fórum/Artigos
- > Comunidade Online (Discord)



Etapa 4

EFS – Amazon Elastic File System

O que é EFS

- Elastic File System
- Fornece um sistema de arquivos
- Servless e totalmente elástico
- Escala até Petabytes
- Aumente e diminuir conforme adição e remoção de arquivos

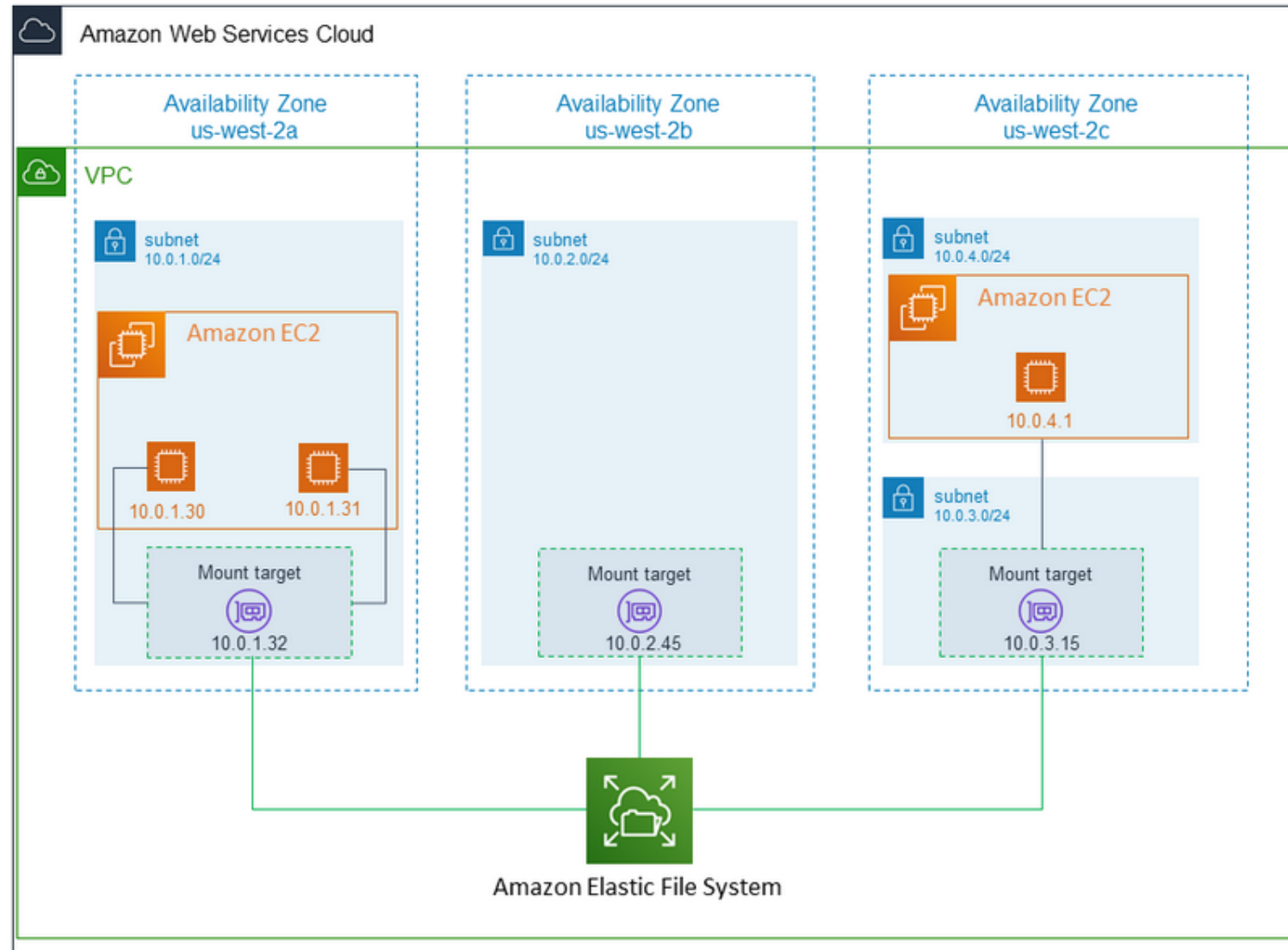
O que é EFS

- Compatível com protocolo NFS (Network file System)
- Pode ser acessado por EC2, Lambda, ECS
- Acesso simultâneo ao mesmos dados sem problemas de performance

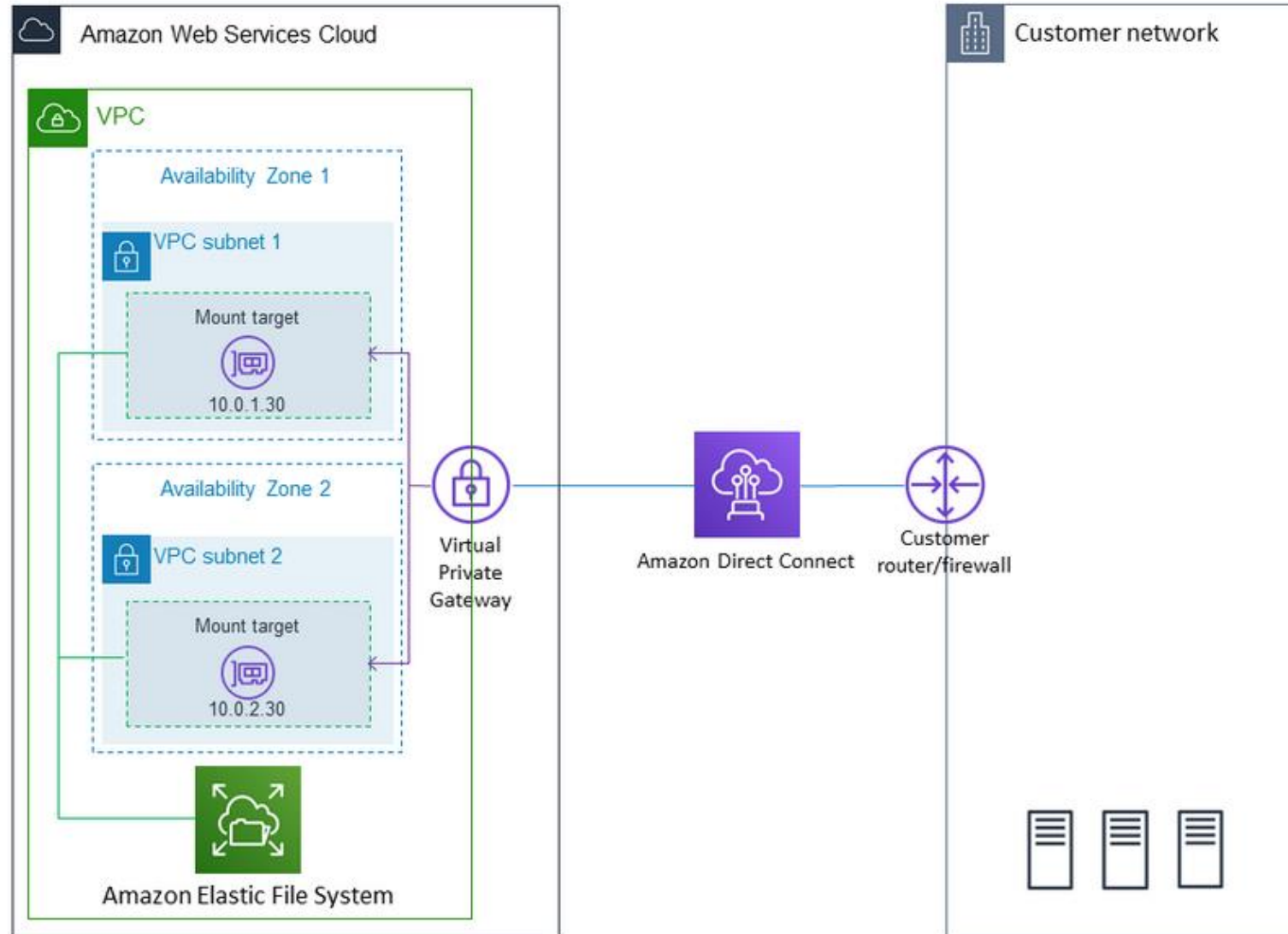
Classes de armazenamento

- Padrão (Instância regional): Standard e Standard – IA (Infrequent Access)
- Uma AZ: One Zone e One Zone – IA (Infrequent Access)

Casos de uso



Casos de uso



Para saber mais

- Página do produto - <https://aws.amazon.com/pt/efs/>
- O que é EFS ? - https://docs.aws.amazon.com/pt_br/efs/latest/ug/whatisefs.html

Dúvidas?

- > Fórum/Artigos
- > Comunidade Online (Discord)

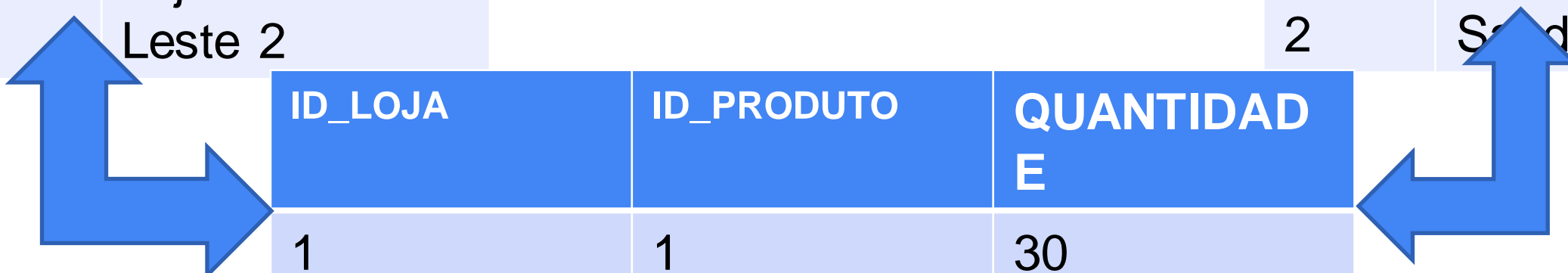


Etapa 5

Amazon Relational Database Service

Banco de dados relacional

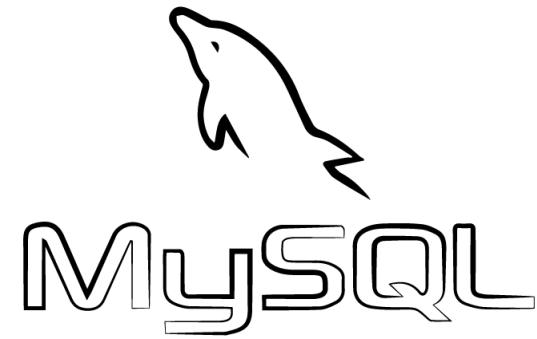
ID_LOJA	NOME_LOJA	ID_LOJA	ID_PRODUTO	QUANTIDADE	ID	PRODUTO
1	Loja Centro 1	1	1	30	1	Tênis
2	Loja Zona Leste 2	1	2	20	2	Sandália
		2	1	10		
		2	2	0		



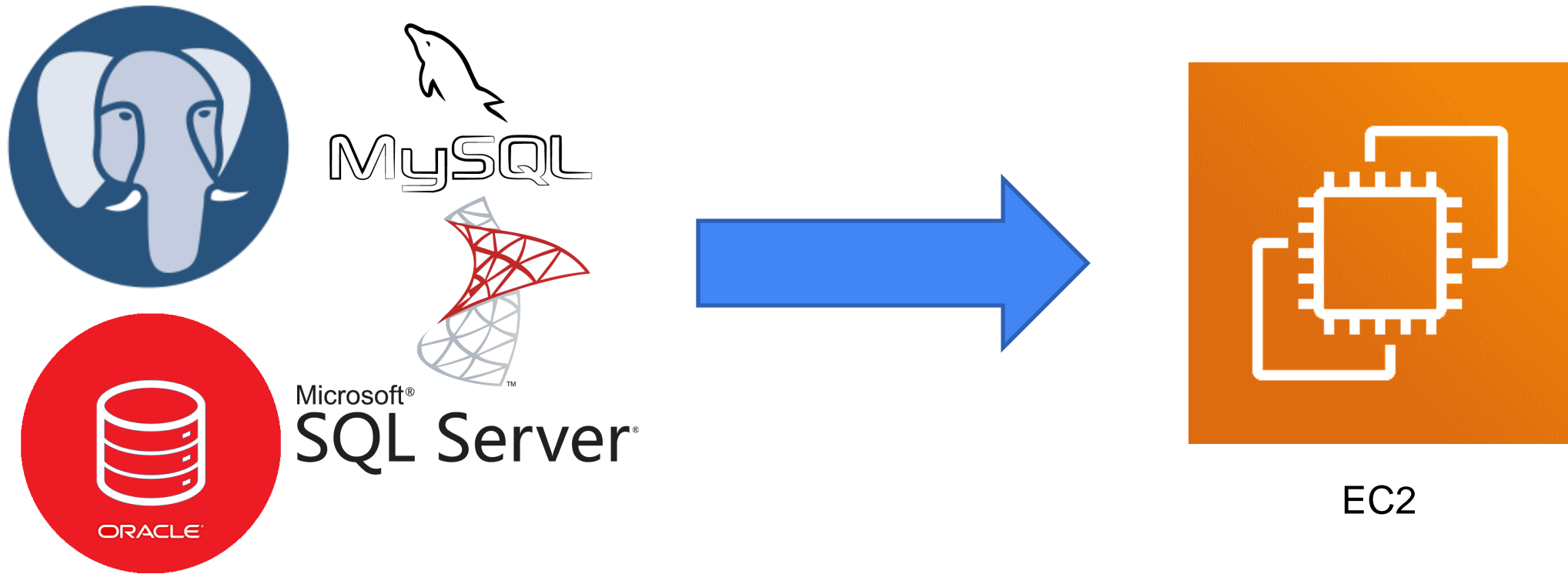
Requisitos

- Relação de dados
- Facilita a compreensão das informações
- SQL como linguagem de consulta
- RDBMS (Sistema de gerenciamento de banco de dados)

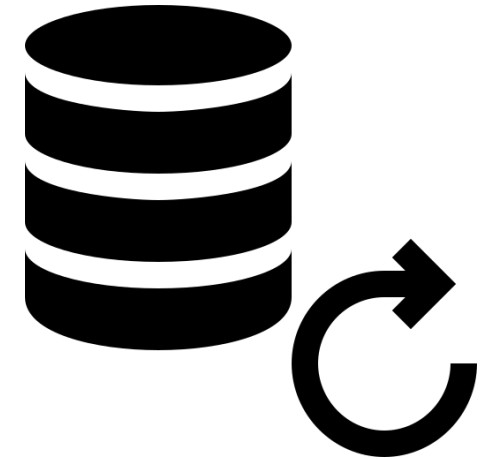
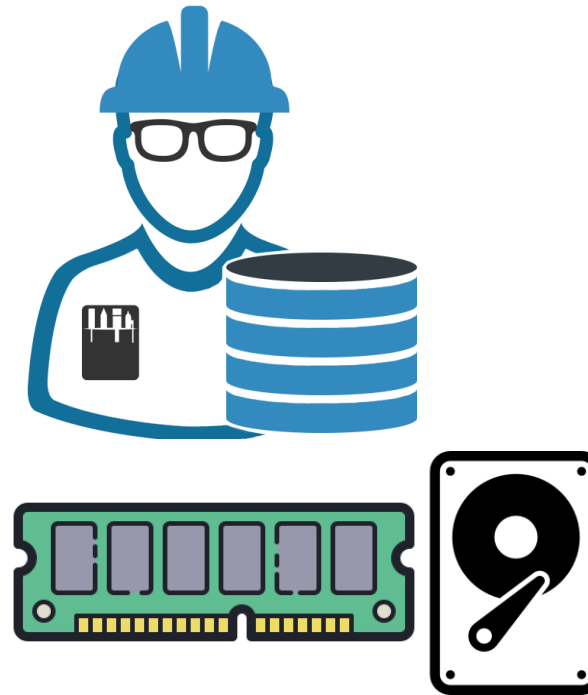
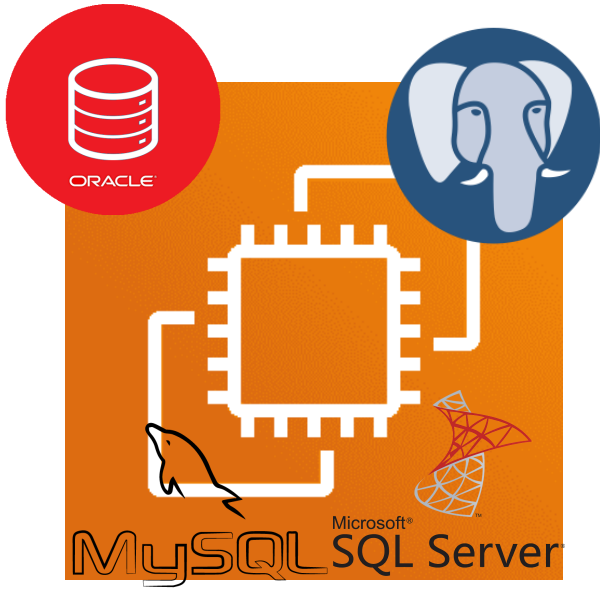
Vendors



Como usar na nuvem ?



Você usa, você gerencia



Relational Database Service

- Facilita configuração e provisionamento de hardware
- Patches automatizados
- Backups
- Redundância
- Failover e Recuperação de Desastres

Mecanismos compatíveis

- MySQL
- PostgreSQL
- MariaDB
- Oracle
- SQL Server

Amazon Aurora

- Servless
- Mecanismos compatíveis: PostgreSQL e MySQL
- Preço 1/10 de outros vendors
- Replicação multi-regional
- Até 15 réplicas de leituras
- Backup contínuo via S3

Para saber mais

- Página do RDS - <https://aws.amazon.com/pt/rds/?p=ft&c=db&z=3>
- O que é RDS? - https://docs.aws.amazon.com/pt_br/AmazonRDS/latest/UserGuide/Welcome.html
- Página Amazon Aurora - <https://aws.amazon.com/pt/rds/aurora/>
- Documentação Amazon Aurora - https://docs.aws.amazon.com/pt_br/AmazonRDS/latest/AuroraUserGuide/CHAP_AuroraOverview.html

Dúvidas?

- > Fórum/Artigos
- > Comunidade Online (Discord)



Etapa 6

DynamoDB

DynamoDB

- Banco de dados não relacional (NoSQL)
- Gerenciado (Servless)
- Performance abaixo de 10 milissegundos
- Escala automaticamente
- Replicação de dados regional
- Caso de uso: Muitos dados, baixa latência



Estrutura dos dados



```
[{
  "PersonID": 101,
  "LastName": "Smith",
  "FirstName": "Fred",
  "Phone": "555-4321"
},
{
  "PersonID": 102,
  "LastName": "Jones",
  "FirstName": "Mary",
  "Address": {
    "Street": "123 Main",
    "City": "Anytown",
    "State": "OH",
    "ZIPCode": 12345
  }
},
{
  "PersonID": 103,
  "LastName": "Stephens",
  "FirstName": "Howard",
  "Address": {
    "Street": "123 Main",
    "City": "London",
    "PostalCode": "E8 5K8"
  },
  "FavoriteColor": "Blue"
}]
```

Tabelas



```
[{
  "PersonID": 101,
  "LastName": "Smith",
  "FirstName": "Fred",
  "Phone": "555-4321"
},
{
  "PersonID": 102,
  "LastName": "Jones",
  "FirstName": "Mary",
  "Address": {
    "Street": "123 Main",
    "City": "Anytown",
    "State": "OH",
    "ZIPCode": 12345
  }
},
{
  "PersonID": 103,
  "LastName": "Stephens",
  "FirstName": "Howard",
  "Address": {
    "Street": "123 Main",
    "City": "London",
    "PostalCode": "E8 5K8"
  },
  "FavoriteColor": "Blue"
}]
```

Tabela "People"

Itens



```
[{
  "PersonID": 101,
  "LastName": "Smith",
  "FirstName": "Fred",
  "Phone": "555-4321"
},
{
  "PersonID": 102,
  "LastName": "Jones",
  "FirstName": "Mary",
  "Address": {
    "Street": "123 Main",
    "City": "Anytown",
    "State": "OH",
    "ZIPCode": 12345
  }
},
{
  "PersonID": 103,
  "LastName": "Stephens",
  "FirstName": "Howard",
  "Address": {
    "Street": "123 Main",
    "City": "London",
    "PostalCode": "ER3 5K8"
  },
  "FavoriteColor": "Blue"
}]
```

Itens da tabela

Atributos

```
[{
  "PersonID": 101,
  "LastName": "Smith",
  "FirstName": "Fred",
  "Phone": "555-4321"
},
{
  "PersonID": 102,
  "LastName": "Jones",
  "FirstName": "Mary",
  "Address": {
    "Street": "123 Main",
    "City": "Anytown",
    "State": "OH",
    "ZIPCode": 12345
  }
},
{
  "PersonID": 103,
  "LastName": "Stephens",
  "FirstName": "Howard",
  "Address": {
    "Street": "123 Main",
    "City": "London",
    "PostalCode": "ER3 5K8"
  },
  "FavoriteColor": "Blue"
}]
```



Resumindo

- Tabelas: Coleção de dados
- Item: Grupo de atributos identificável
- Atributo: Dados existentes dentro de cada item

Para saber mais

- Página do produto - <https://aws.amazon.com/pt/dynamodb/>
- Documentação - https://docs.aws.amazon.com/pt_br/amazondynamodb/latest/developerguide/Introduction.html
- DynamoDB – O quê, por que e quando usar - <https://dev.to/oieduardorabelo/amazon-dynamodb-o-que-por-que-e-quando-usar-o-design-de-tabela-unica-com-dynamodb-ao9>

Dúvidas?

- > Fórum/Artigos
- > Comunidade Online (Discord)



Etapa 7

Outros serviços de banco de dados

A escolha do banco de dados correto

A necessidade de negócio escolhe o tipo de banco de dados

Amazon DocumentDB

- Banco de dados de documentos
- Gerenciamento de conteúdo
- Catálogos, perfis de usuário
- Compatível com cargas de trabalho MongoDB

Amazon Neptune

- Redes sociais, mecanismos de recomendação, detecção de fraude e gráficos de conhecimento
- Banco de dados de grafos

Amazon QLDB

- Quantum Ledger Database
- Banco de dados serviço ledger
- Imutabilidade
- Indicado para históricos, registros digitais, transações financeiras

Amazon DynamoDB Accelerator

- Chamado também de DAX
- Camada de cache nativa pra otimizar tempo de leitura de dados

Amazon ElastiCache

- Camada de cache sobre bancos de dados
- Compatível com Redis e Memcached

Para saber mais

- DAX - <https://aws.amazon.com/pt/dynamodb/dax/>
- Elasticache <https://aws.amazon.com/pt/elasticache/>
- DocumentDB – <https://aws.amazon.com/pt/documentdb/>
- Neptune - <https://aws.amazon.com/pt/neptune/>
- QLDB - <https://aws.amazon.com/pt/qldb/>

Dúvidas?

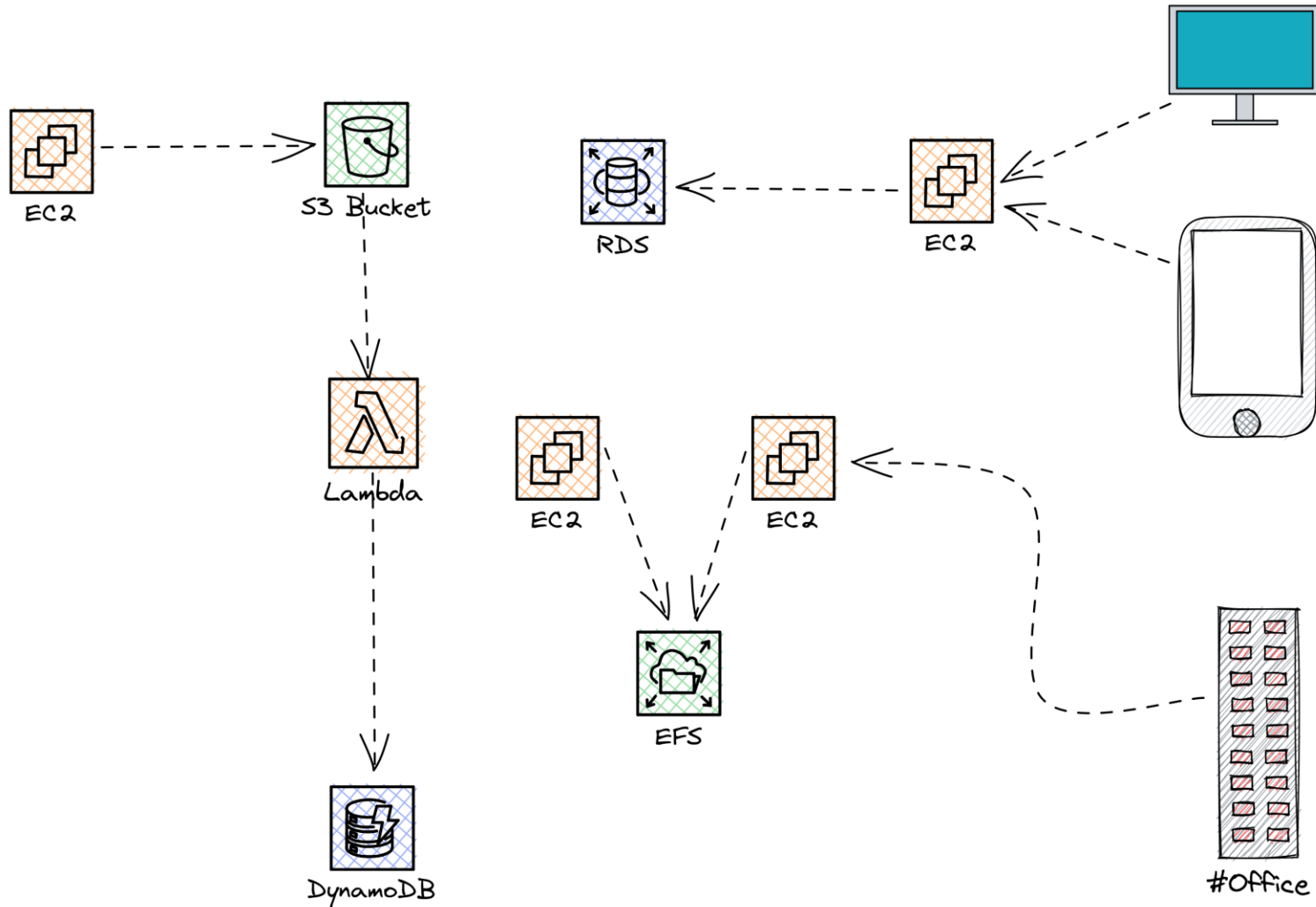
- > Fórum/Artigos
- > Comunidade Online (Discord)



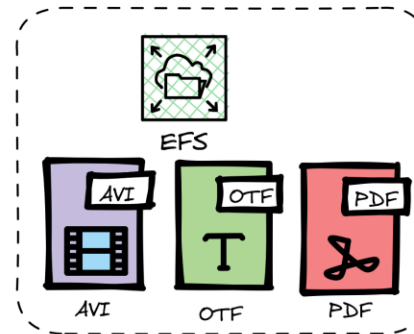
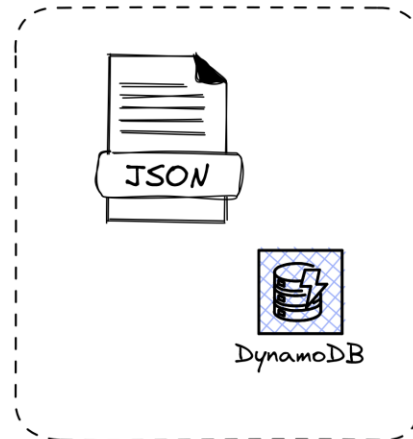
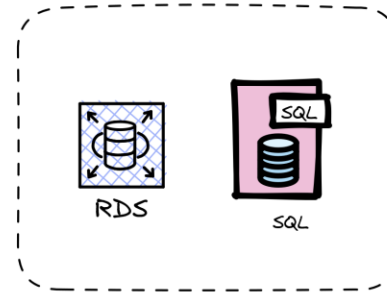
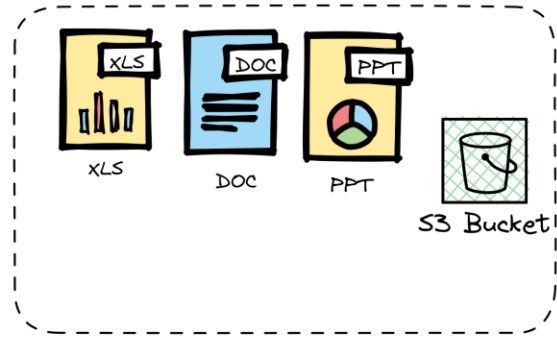
Etapa 8

Big Data com Amazon Redshift

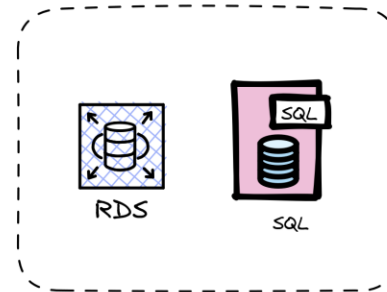
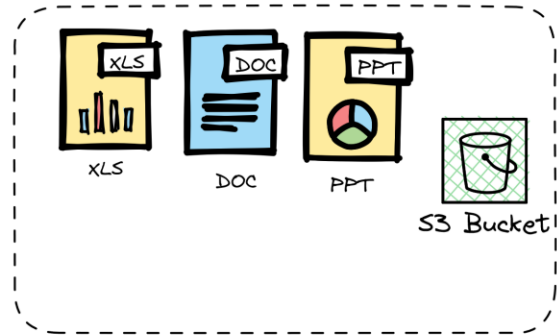
Cada vez mais dados



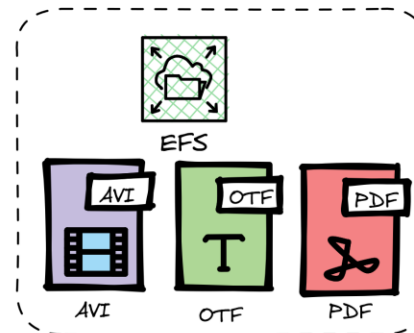
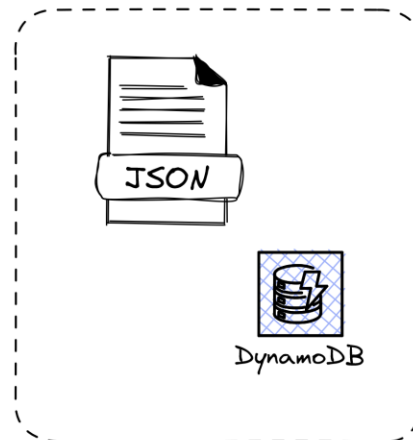
Nossas fontes de dados



Perguntas



Quantas vendas
o aplicativo fez
desde o início do
lançamento?



Quanto usuários fizeram
cadastro na última hora ?

Em um contexto de

- Velocidade de geração de dados
- Variedade de fontes de dados
- E quando os dados precisam responder com inteligência de negócio

Amazon Redshift

- Serviço de Data warehouse para análise de Big Data
- Oferece coletar informações de muitas fontes de dados
- Projeta relações e tendências de dados
- Usando Redshift Spectrum é possível rodar comandos SQL em cima de todas as fontes de dados agrupadas

Para saber mais

- Página do produto - <https://aws.amazon.com/pt/redshift/>
- Documentação - https://docs.aws.amazon.com/pt_br/redshift/latest/gsg/getting-started.html

Dúvidas?

- > Fórum/Artigos
- > Comunidade Online (Discord)

