



SÃO  
PAULO  
TECH  
SCHOOL

# **Arquitetura de soluções em nuvem**

## **Armazenamento em nuvem + arquitetura data lake**

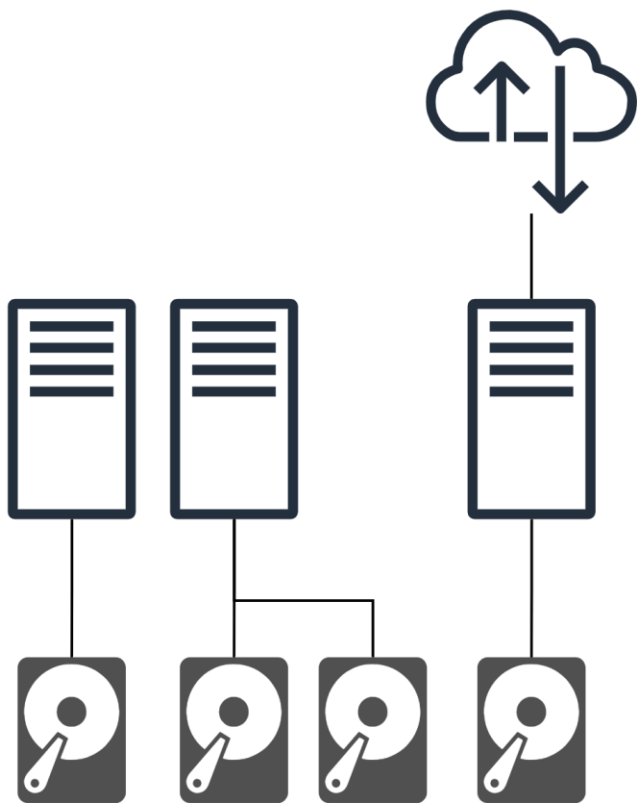
**Eduardo Verri**

[eduardo.verri@sptech.school](mailto:eduardo.verri@sptech.school)

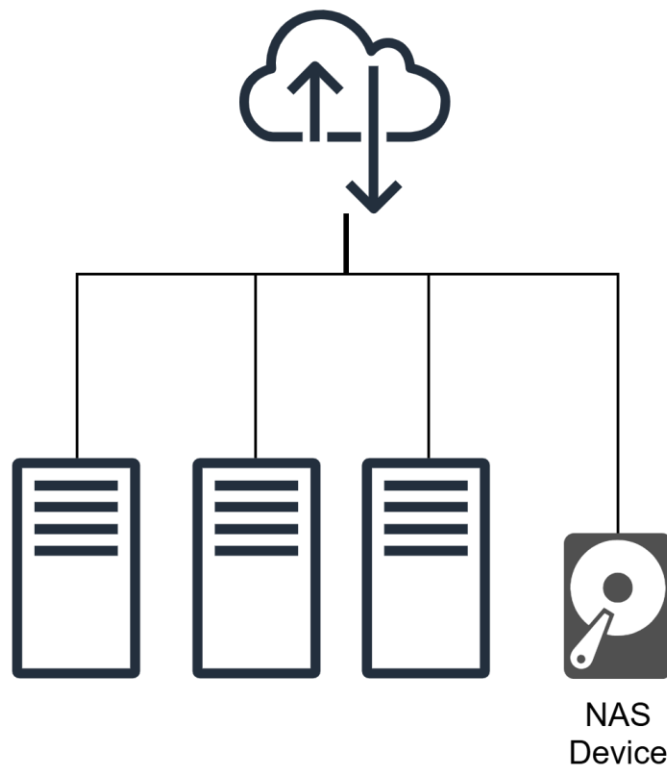
# Topologias de armazenamento

- ❑ **DAS – Direct Attached Storage:** dispositivos atachados diretamente ao computador como HD, SSD, HD externo e Pen Drives
- ❑ **NAS – Network Area Storage:** é um dispositivo (aplicação) com sistema operacional próprio e discos com ampla capacidade de armazenamento. São conectados à rede através de uma conexão Ethernet padrão.
- ❑ **SAN – Storage Area Network:** o acesso é através de uma estrutura de rede dedicada oferecendo flexibilidade e alto desempenho, porém maior complexidade de administração

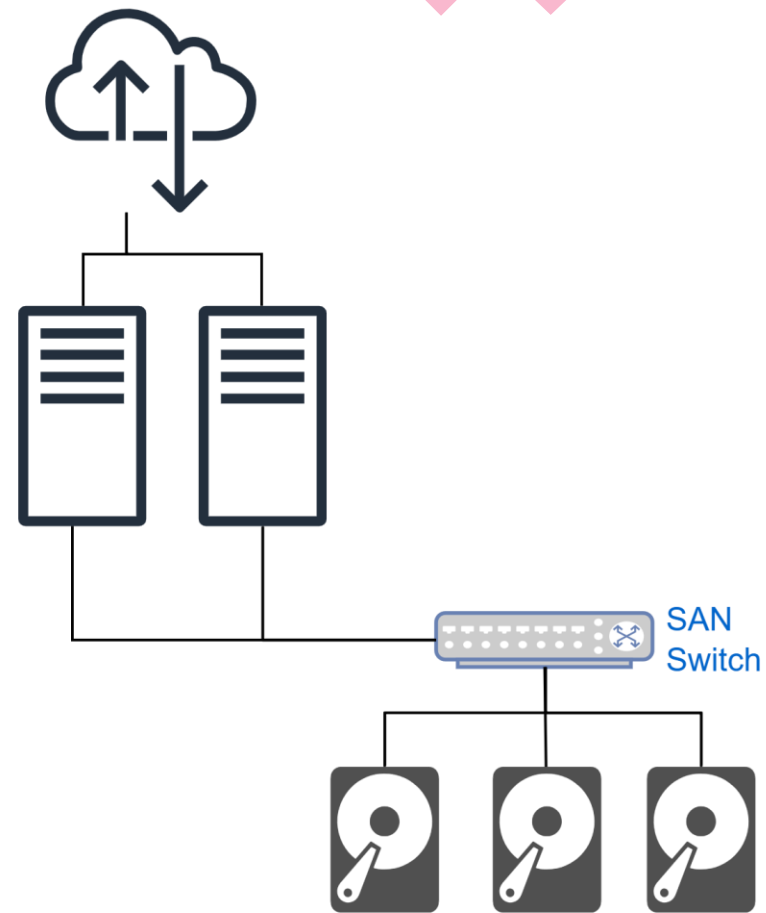
*Uma NAS aparece para um sistema operacional cliente como um servidor de arquivos, enquanto uma SAN aparece como um disco e existe como sua própria rede separada de dispositivos de armazenamento*



DAS



NAS



SAN

# Serviços de armazenamento da AWS



## Amazon Simple Storage Service (S3)

Armazenamento de objetos com escalabilidade, disponibilidade e segurança líder do setor para você armazenar e recuperar qualquer volume de dados de qualquer lugar.



## Amazon Elastic Block Store (EBS)

Serviço de armazenamento de blocos de alta performance, fácil de usar, tanto para workloads com alta throughput de dados quanto com intenso consumo de transações em qualquer escala.



## Amazon Elastic File System (EFS)

Um sistema de arquivos simples, sem servidor, elástico e com configuração única para você compartilhar dados de arquivos sem gerenciamento de armazenamento.



## Amazon File Cache

Cache de alta velocidade para conjuntos de dados armazenados em qualquer lugar, acelera workloads de estouro de nuvem.

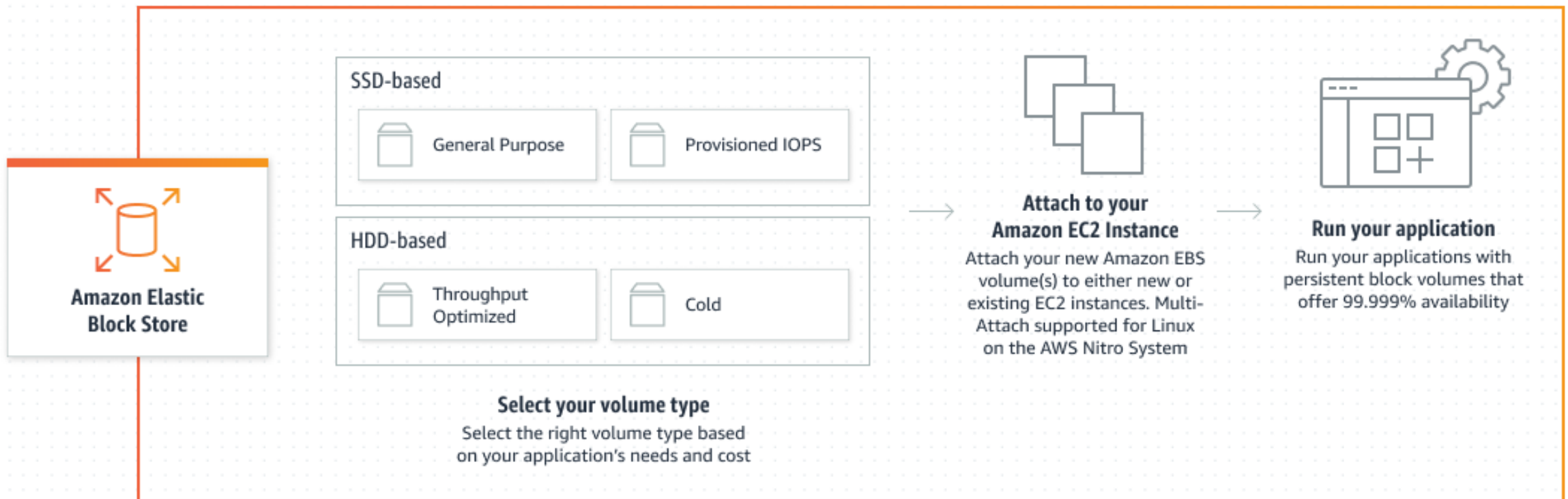
FSx

## Amazon FSx

Armazenamento de arquivos com economia de custo e totalmente gerenciado oferece as capacidades e a performance de sistemas de arquivos populares comerciais e de código aberto.

# Amazon EBS

Projetado para armazenar dados em um volume dedicado conectado a uma instância do Amazon EC2, assim como uma unidade de disco local em sua máquina física,



# Amazon EBS

## Principais Características:

1. **Persistência e Velocidade:** O EBS é adequado para armazenar dados de aplicativos que exigem alta velocidade de leitura e gravação, oferecendo persistência mesmo quando a instância é desligada.
2. **Conectividade Direta:** É conectado diretamente a instâncias EC2, permitindo que o armazenamento seja tratado como um disco físico.

## Cenários de Uso:

1. **Armazenamento de Dados de Aplicativos:** Dados que precisam ser persistentes e exigem alta taxa de transferência são ideais para serem armazenados no EBS.
2. **Execução de Bancos de Dados:** O EBS é frequentemente utilizado para armazenar dados de bancos de dados em instâncias EC2.

# Amazon S3

Armazena dados como objetos em um ambiente simples (sem hierarquia) que são associados a um identificador único (chave), para que possam ser acessados por meio de solicitações da web de qualquer lugar. Aceita GET e POST.





# Amazon S3

## Principais Características:

1. **Durabilidade e Confiabilidade:** O Amazon S3 oferece uma durabilidade excepcional, tornando-o ideal para armazenar backups, arquivos estáticos, e dados críticos que precisam ser preservados a longo prazo.
2. **Escalabilidade Automática:** Sua capacidade de escalar automaticamente faz do S3 uma escolha sólida para armazenamento de objetos em grande escala.

## Cenários de Uso:

1. **Armazenamento de Arquivos Estáticos:** Imagens, vídeos, e outros ativos estáticos encontram no S3 um ambiente propício para armazenamento.
2. **Backup e Recuperação de Dados:** A durabilidade do S3 o torna uma escolha popular para backups seguros e recuperação de desastres.

## O que é um data lake?

- ❑ Data lake é um repositório de armazenamento de dados que suporta grandes quantidades de dados, podendo ser estruturados e não estruturados.
- ❑ É um repositório centralizado que ingere e armazena grandes volumes de dados em sua forma original. Os dados podem ser processados e usados como base para uma variedade de necessidades analíticas.
- ❑ Pode acomodar todos os tipos de dados de qualquer fonte, desde dados estruturados (tabelas de banco de dados, planilhas do Excel) até semiestruturados (arquivos XML, páginas da Web) e não estruturados (imagens, arquivos de áudio, tweets)
- ❑ Os arquivos normalmente são armazenados em zonas preparadas – raw, trusted, curated – para que diferentes tipos de usuários possam usar os dados em suas várias formas para atender às suas necessidades.

## Elementos essenciais

- ❑ **Migração de dados:** Os data lakes permitem que você importe qualquer quantidade de dados que possa vir em tempo real. Os dados são coletados de várias fontes e movidos para o data lake em seu formato original. Esse processo permite escalar para dados de qualquer tamanho, economizando tempo na definição de estruturas de dados, esquemas e transformações.
- ❑ **Armazene e catalogue dados com segurança:** Os data lakes permitem que você armazene dados relacionais, como bancos de dados operacionais e dados de aplicações de linha de negócios, e dados não relacionais, como aplicações móveis, dispositivos IoT e mídias sociais. Eles também oferecem a capacidade de entender quais dados estão no lago por meio de crawling, catalogação e indexação de dados. Por fim, os dados devem ser protegidos para garantir que seus ativos de dados estejam protegidos.

## Elementos essenciais

- ❑ **Análise:** Os data lakes permitem que várias funções da organização, como cientistas de dados, desenvolvedores de dados e analistas de negócios, acessem dados com sua escolha de ferramentas e frameworks analíticos. Isso inclui frameworks de código aberto, como Apache Hadoop, Presto e Apache Spark, e ofertas comerciais de fornecedores de data warehouse e inteligência empresarial. Os data lakes permitem executar análises sem a necessidade de mover seus dados para um sistema de análise separado.
- ❑ **Machine Learning:** O data lakes permitem que as organizações gerem diferentes tipos de insights, incluindo relatórios sobre dados históricos e machine learning, onde os modelos são criados para prever resultados prováveis e sugerir uma série de ações prescritas para alcançar o resultado ideal.

## Data lake vs data warehouse

- ❑ Embora ambos armazenem dados, cada repositório tem seus próprios requisitos de armazenamento, o que os torna escolhas ideais para cenários diferentes.
- ❑ Os **data warehouses** exigem um esquema definido para atender a requisitos específicos de análise de dados para saídas de dados, como dashboards, visualizações de dados e outras tarefas de business intelligence. Esses requisitos são geralmente especificados pelos usuários corporativos que utilizarão os resultados do relatório regularmente. A **estrutura** de um data warehouse é normalmente organizada como um sistema relacional, obtendo dados de bancos de dados transacionais.
- ❑ Os **data lakes**, por outro lado, incorporam dados de sistemas relacionais e não relacionais, permitindo que os cientistas de dados incorporem dados estruturados e não estruturados em mais projetos de ciência de dados.

## Data lake vs data warehouse

- ❑ Cada sistema também tem seu próprio conjunto de vantagens e desvantagens. Por exemplo, os data warehouses tendem a ter melhor desempenho, mas têm um custo mais elevado.
- ❑ Os data lakes podem ser mais lentos no retorno dos resultados das consultas, mas têm custos de armazenamento mais baixos. Além disso, a capacidade de armazenamento dos data lakes os torna ideais para dados corporativos.

## Desafios de um Data lake

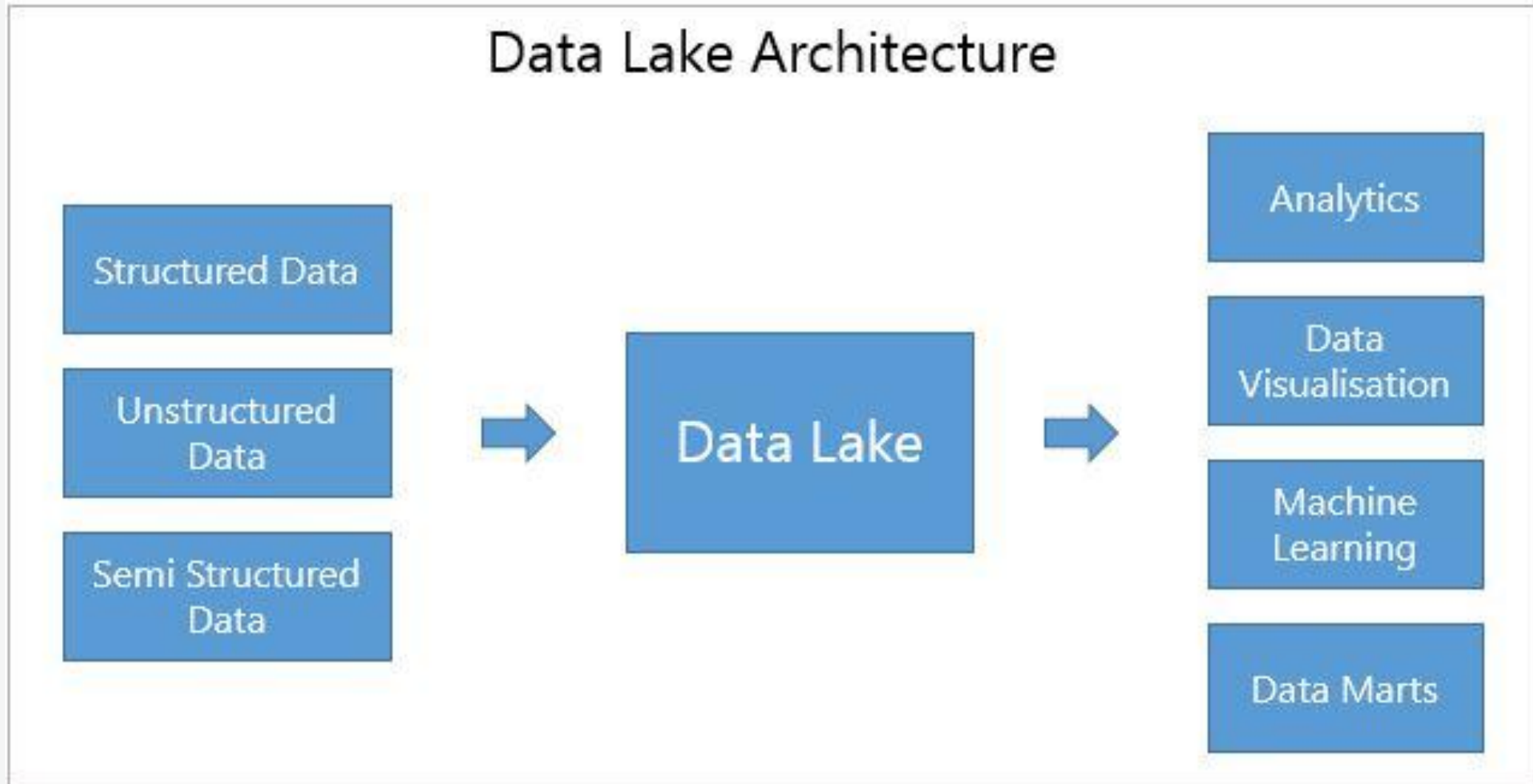
- ❑ **Desempenho:** Conforme a quantidade de dados inseridos em um data lake aumenta, isso afeta o desempenho, que já é mais lento do que outros sistemas alternativos de armazenamento de dados.
- ❑ **Governança:** Ele também requer uma forte governança para gerenciar adequadamente os dados. Os dados devem ser marcados e classificados com metadados relevantes para evitar **pântanos de dados**, e essas informações devem ser facilmente acessíveis por meio de um catálogo de dados, permitindo a funcionalidade de autoatendimento para uma equipe menos técnica, como analistas de negócios. Por fim, as proteções também devem ser implementadas para atender aos padrões regulatórios e de privacidade; isso pode incluir controles de acesso, criptografia de dados e muito mais.

# Arquitetura de um data lake

- ❑ Um Data Lake é um repositório centralizado de dados brutos e estruturados e não estruturados de várias fontes, incluindo dados de transações.
- ❑ Ele é projetado para armazenar dados em seu formato original e permitir que eles sejam acessados e processados por usuários e aplicativos, sem a necessidade de transformação prévia ou esquematização.
- ❑ O objetivo de um Data Lake é fornecer uma fonte única de confiança de dados para análises de negócios e tomada de decisões. Possibilitando a agilidade no sentido de conseguir receber dados de fontes sem esquematização prévia, separação da estrutura de processamento e de armazenamento, além de ser compatível com diversas ferramentas de processamento de dados e análise.

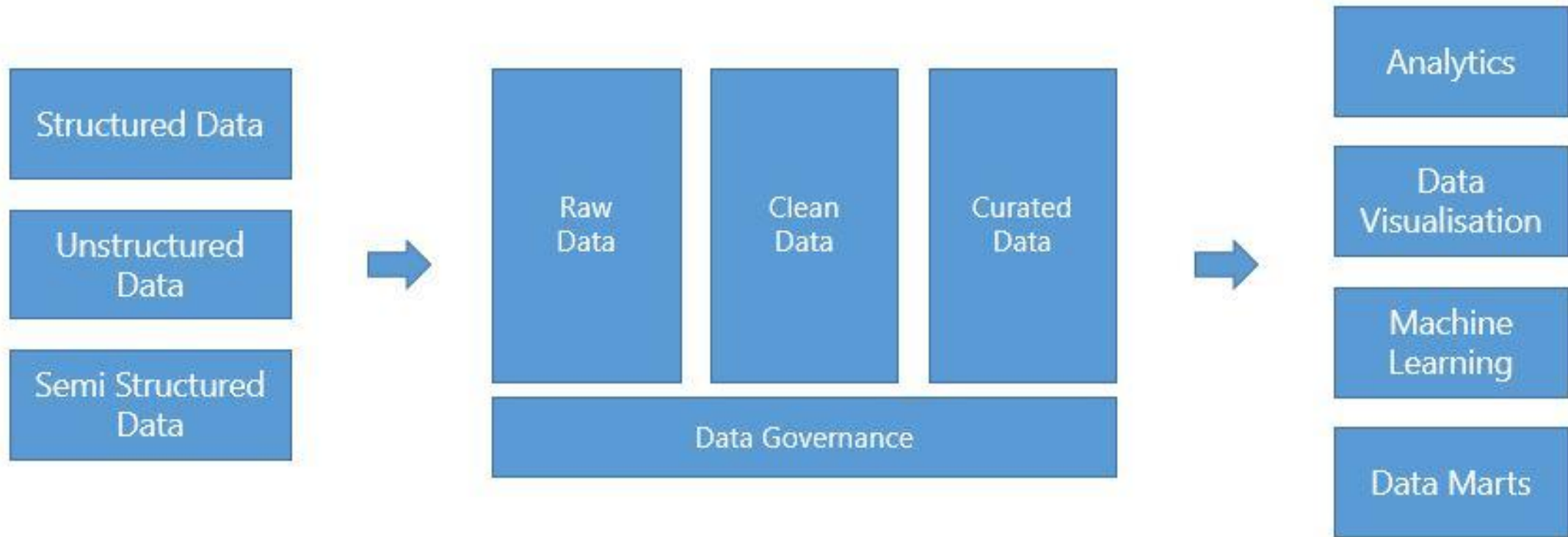


## Arquitetura de um data lake



# Arquitetura de um data lake

## Data Lake Architecture



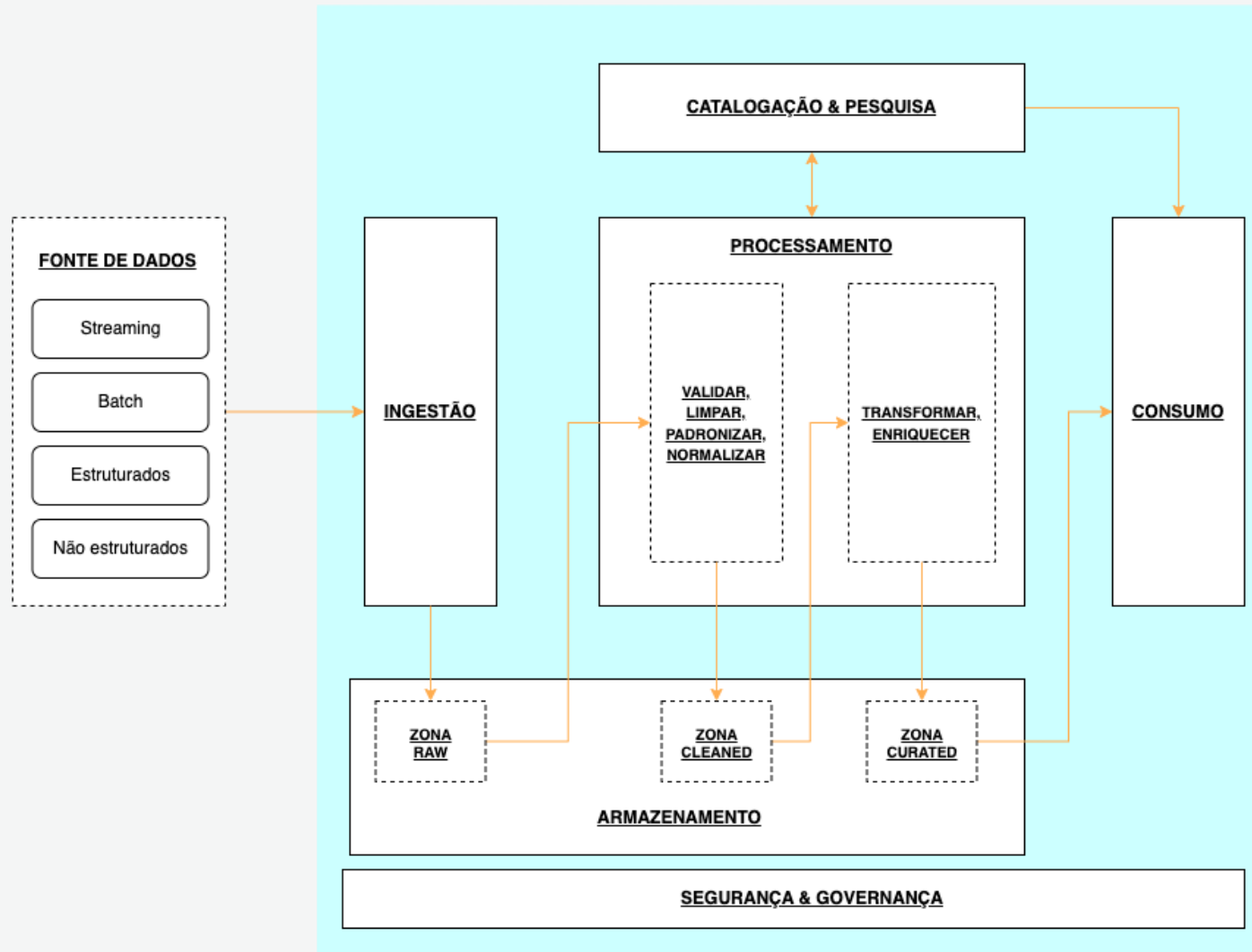
# Arquitetura de um data lake

- ❑ As três camadas clássicas do Data Lake tem diversos nomes, dependendo da literatura seguida, como Raw, Clean, Curated ou Bronze, Prata e Ouro, etc. Nesse caso, o importante não é muito o nome da camada escolhida e sim o que elas representam.
- ❑ **Uma primeira camada** é responsável por ser permissiva no sentido de suportar a escrita de fontes, para dar agilidade e também suporte. A primeira camada também costuma ter algumas características que não permitem a deleção, com objetivo de salvar fielmente todos os dados que em algum momento passaram por ela.
- ❑ **A segunda camada** de um Data Lake tem como principal propósito a organização e a preparação dos dados para serem utilizados em análises e tomadas de decisão. Nessa camada, é realizada a limpeza, transformação e integração dos dados provenientes de diferentes fontes, permitindo que os mesmos sejam consolidados e estruturados de forma padronizada e coerente

# Arquitetura de um data lake

- ❑ **A terceira camada** de um data lake tem como propósito a análise avançada e extração de valor dos dados. Nesta camada, são aplicadas técnicas de processamento de dados mais sofisticadas, como machine learning, inteligência artificial e modelagem estatística, com o objetivo de descobrir padrões, tendências e insights relevantes para o negócio.
- ❑ A terceira camada é utilizada para estruturar os dados de modo a serem analisados em um banco de dados específico, como Data Warehouse, Data Mart ou banco de dados para análise de grafos e séries temporais.
- ❑ Além disso, é possível utilizar a terceira camada para disponibilizar os dados para ferramentas de visualização de dados e dashboards interativos.

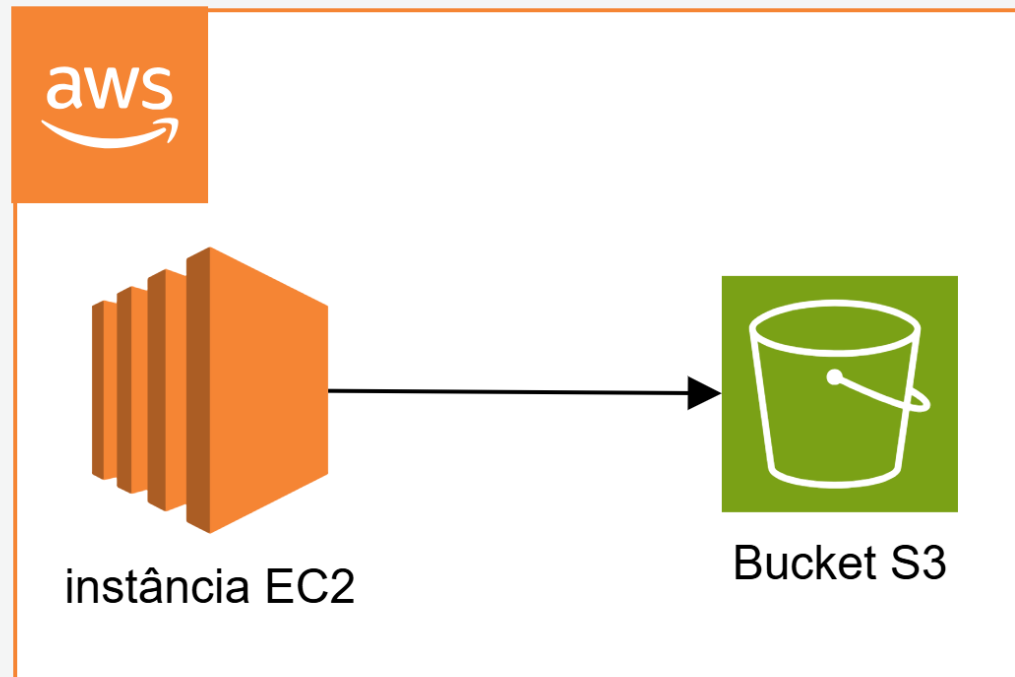
# Arquitetura de um data lake



**Lab S3**








## Objetivo




- ❑ Criar uma conexão entre uma instância EC2 e o serviço S3 de armazenamento.




**Criando um bucket S3**



 **Serviços**  [Alt+S]     Norte da Virgínia  voclabs/user3453499=Testar\_aluno @ 8321-5216-1309 

 EC2  VPC  RDS

## Amazon S3



- Buckets
- Access Grants
- Pontos de acesso
- Pontos de acesso Lambda de objeto
- Pontos de acesso de várias regiões
- Operações em lotes
- IAM Access Analyzer para S3

---

- Configurações de bloqueio do acesso público desta conta

Armazenamento

# Amazon S3

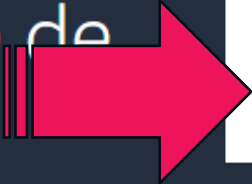
## Armazene e recupere qualquer quantidade de dados a partir de qualquer lugar

O Amazon S3 é um serviço de armazenamento de objetos que oferta escalabilidade, disponibilidade de dados, segurança e performance líderes do setor.

### Criar um bucket

Cada objeto no S3 é armazenado em um bucket. Para carregar arquivos e pastas no S3, será preciso criar um bucket no qual os objetos serão armazenados.

**Criar bucket**



## Configuração geral

Região da AWS

Leste dos EUA (Norte da Virgínia) us-east-1

Tipo de bucket | Informações



### Propósito geral

Recomendados para a maioria dos casos de uso e padrões de acesso. Os buckets de uso geral são do tipo original do S3. Eles permitem uma combinação de classes de armazenamento que armazenam objetos de maneira redundante em várias zonas de disponibilidade.



### Diretório

Recomendados para casos de uso de baixa latência. Esses buckets usam somente a classe de armazenamento do S3 Express One Zone, que fornece processamento mais rápido de dados em uma única zona de disponibilidade.

Nome do bucket | Informações

s3-raw-lab

O nome do bucket deve ser exclusivo no namespace global e seguir as regras de nomenclatura do bucket. [Veja as regras para nomenclatura de buckets](#)

Copiar configurações do bucket existente - *opcional*

Somente as configurações de bucket na configuração a seguir são copiadas.

**Escolher bucket**

Formato: s3://bucket/prefix

## Propriedade de objeto [Informações](#)

Controle a propriedade de objetos gravados nesse bucket a partir de outras contas da AWS e o uso de listas de controle de acesso (ACLs). A propriedade do objeto determina quem pode especificar o acesso aos objetos.

### ☒ ACLs desabilitadas (recomendado)

Todos os objetos nesse bucket são de propriedade dessa conta. O acesso a esse bucket e seus objetos é especificado usando apenas políticas.


### ☐ ACLs habilitadas

Os objetos nesse bucket podem ser de propriedade de outras contas da AWS. O acesso a esse bucket e seus objetos pode ser especificado usando ACLs.

Propriedade do objeto

Imposto pelo proprietário do bucket

## Configurações de bloqueio do acesso público deste bucket

O acesso público é concedido a buckets e objetos por meio de listas de controle de acesso (ACLs), políticas de bucket, políticas de ponto de acesso ou todas elas. Para garantir que o acesso público a este bucket e todos os seus objetos seja bloqueado, ative a opção de Bloquear todo o acesso público. Essas configurações serão aplicadas apenas a este bucket e aos respectivos pontos de acesso. A AWS recomenda ativar a opção Bloquear todo o acesso público. Porém, antes de aplicar qualquer uma dessas configurações, verifique se as aplicações funcionarão corretamente sem acesso público. Caso precise de algum nível de acesso público a este bucket ou aos objetos que ele contém, é possível personalizar as configurações individuais abaixo para que atendam aos seus casos de uso de armazenamento específicos. [Saiba mais](#) 

### ☒ Bloquear *todo* o acesso público

Ativar essa configuração é o mesmo que ativar todas as quatro configurações abaixo. Cada uma das configurações a seguir são independentes uma da outra.

#### ☒ Bloquear acesso público a buckets e objetos concedidos por meio de *novas* listas de controle de acesso (ACLs)

O S3 bloqueará as permissões de acesso público aplicadas a blocos ou objetos recém-adicionados e impedirá a criação de novas ACLs de acesso público para blocos e objetos existentes. Essa configuração não altera nenhuma permissão existente que permita o acesso público aos recursos do S3 usando ACLs.

#### ☒ Bloquear acesso público a buckets e objetos concedidos por meio de *qualquer* lista de controle de acesso (ACLs)

O S3 ignorará todas as ACLs que concedem acesso público a buckets e objetos.


#### ☒ Bloquear acesso público a buckets e objetos concedidos por meio de *novas* políticas de ponto de acesso e bucket público

O S3 bloqueará novas políticas de bucket e ponto de acesso que concedem acesso público a buckets e objetos. Essa configuração não altera nenhuma política existente que permita o acesso público aos recursos do S3.

#### ☒ Bloquear acesso público e entre contas a buckets e objetos por meio de *qualquer* política de bucket ou ponto de acesso público

O S3 ignorará o acesso público e entre contas para buckets ou pontos de acesso com políticas que concedem acesso público a buckets e objetos.

## Versionamento de bucket

O versionamento é um meio de manter múltiplas variantes de um objeto no mesmo bucket. Você pode usar o versionamento para preservar, recuperar e restaurar todas as versões de cada objeto armazenado no bucket do Amazon S3. Com o versionamento, você pode recuperar facilmente ações não intencionais do usuário e falhas da aplicação. [Saiba mais](#) 

### Versionamento de bucket

☒ Desativar

☐ Ativar

## Tags - *opcional* (0)

Você pode usar tags de bucket para rastrear custos de armazenamento e organizar buckets. [Saiba mais](#) 

Nenhuma tag associada a este bucket.

Adicionar tag

## Criptografia padrão [Informações](#)

A criptografia no lado do servidor é aplicada automaticamente a novos objetos armazenados nesse bucket.

### Tipo de criptografia | [Informações](#)

- ☒ Criptografia do lado do servidor com chaves gerenciadas do Amazon S3 (SSE-S3)
- ☐ Criptografia do lado do servidor com chaves do AWS Key Management Service (SSE-KMS)
- ☐ Criptografia de duas camadas no lado do servidor com chaves do AWS Key Management Service (DSSE-KMS)  
Proteja seus objetos com duas camadas separadas de criptografia. Para obter detalhes sobre a precificação, consulte os [preços do DSSE-KMS](#) na guia [Armazenamento](#) da [página de preços do Amazon S3](#). [↗](#)


### Chave do bucket

O uso de uma chave de bucket do S3 para SSE-KMS reduz os custos de criptografia ao diminuir as chamadas para o AWS KMS. As chaves de bucket do S3 não são compatíveis com o DSSE-KMS. [Saiba mais](#) [↗](#)

- ☐ Desativar
- ☒ Ativar

## ▼ Configurações avançadas


### Bloqueio de objeto


Armazene objetos usando um modelo write-once-read-many (WORM - uma gravação, várias leituras) para ajudar a evitar que objetos sejam excluídos ou substituídos por um período fixo ou indefinidamente. O Bloqueio de objetos funciona apenas em buckets versionados. [Saiba mais](#) 

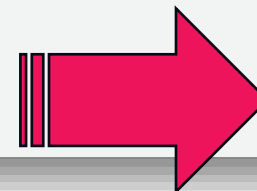
☒ Desativar

☐ Ativar

Permite sempre que os objetos neste bucket sejam bloqueados. Será necessário usar uma configuração adicional nos detalhes do bucket após sua criação para proteger os objetos contidos nele contra exclusão ou substituição.

 O bloqueio de objetos funciona somente em buckets versionados. Habilitar o bloqueio de objetos habilitará automaticamente o versionamento.

 Depois de criar o bucket, você pode fazer upload de arquivos e pastas para o bucket e definir configurações adicionais do bucket.



Criar bucket

Buckets de uso geral (1) [Informações](#) Todas as regiões da AWS

Buckets são contêineres para dados armazenados no S3.

Q

Encontrar buckets por nome

Copiar ARN

Vazio

Excluir

Criar bucket

|                       | Nome ▲                     | Região da AWS ▼                             | IAM Access Analyzer                              | Data de criação ▼           |
|-----------------------|----------------------------|---|--|-----------------------------|
| <input type="radio"/> | <a href="#">s3-raw-lab</a> | Leste dos EUA (Norte da Virgínia) us-east-1 | <a href="#">Exibir analisador para us-east-1</a> | 17 Sep 2024 10:59:40 AM -03 |



**Inserindo Função/Role na instância EC2**

aws

Serviços

Search

[Alt+S]

Norte da Virgínia

voclabs/user3453499=Testar\_aluno @ 8321-5216-1309

EC2

VPC

RDS

Instâncias (1/1) Informações

Última atualização less than a minute atrás

Conectar

Estado da instância

Ações

Executar instâncias

Localizar Instância por atributo ou tag (case-sensitive)

Todos os es...

|                                     | Name          | ID da instância     | Estado da instância | Tipo de inst... | Verificação de status | Status do a  |
|-------------------------------------|---------------|---------------------|---------------------|-----------------|-----------------------|--------------|
| <input checked="" type="checkbox"/> | public-server | i-0bbd62c721ef2770e | Executando          | t2.micro        | Inicializando         | Exibir alarm |

Alterar grupos de segurança

Obter senha do Windows

Modificar função do IAM

# Modificar função do IAM Informações

Associe uma função do IAM à sua instância.

ID da instância

 i-0bbd62c721ef2770e (public-server)


Função do IAM

Selecione uma função do IAM para associar à sua instância ou crie uma nova função caso ainda não tenha criado nenhuma. A função selecionada substitui todas as funções atualmente associadas à sua instância.



LabInstanceProfile ▼



[Criar nova função do IAM](#) 

Carregar



**Atualizar função do IAM**

**Acessando o Bucket S3 da instância**

# Configurando a EC2

1. Realize a conexão SSH com sua EC2
2. Atualize os pacotes do sistema operacional com

```
$ sudo apt update && sudo apt upgrade -y
```

3. Instale a ferramenta AWS CLI, que serve para gerenciar a AWS console via linha de comando ([doc](#))

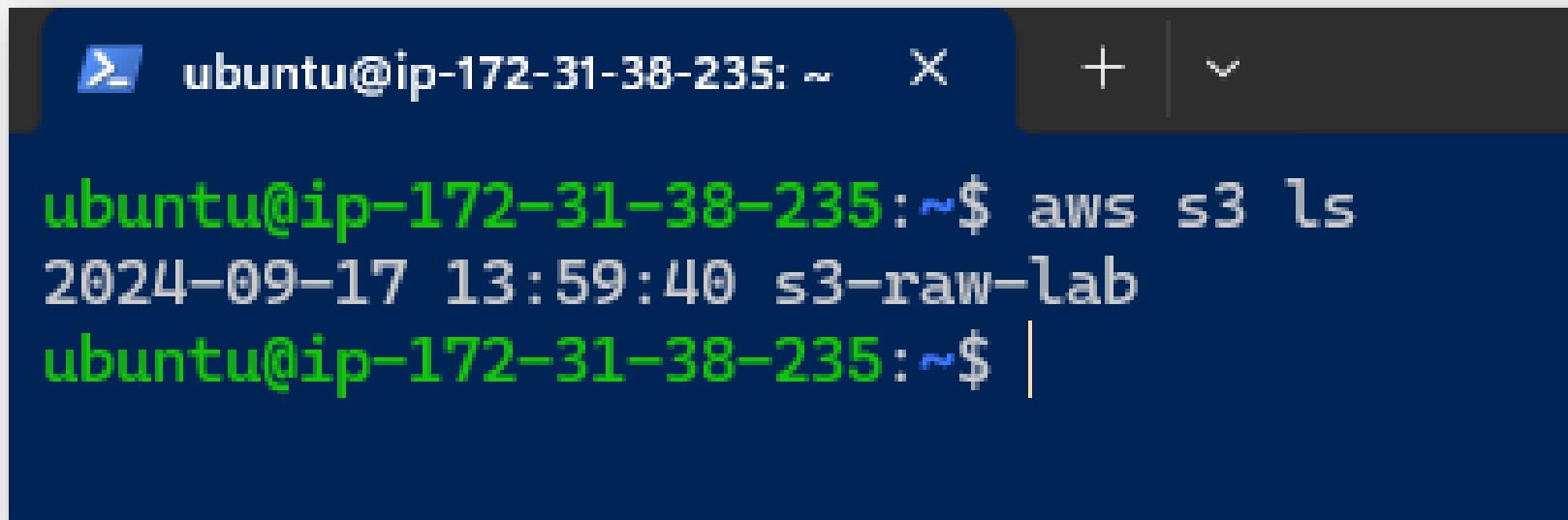
```
$ curl "https://awscli.amazonaws.com/awscli-exe-linux-x86_64.zip" -o "awscliv2.zip"  
$ unzip awscliv2.zip  
$ sudo ./aws/install
```

```
ubuntu@ip-172-31-38-235:~$ aws --version  
aws-cli/2.17.52 Python/3.12.6 Linux/6.8.0-1012-aws exe/x86_64.ubuntu.24
```

## Checando o acesso

1. Utilize a instância para listar os buckets da S3 que você possui. A função IAM está sendo utilizada nesse momento e podemos observar o bucket que criamos

```
$ aws s3 ls
```

A terminal window with a dark blue background. The title bar shows 'ubuntu@ip-172-31-38-235: ~' and window control buttons. The terminal text shows the command 'aws s3 ls' being executed, followed by the output '2024-09-17 13:59:40 s3-raw-lab'. The prompt 'ubuntu@ip-172-31-38-235:~\$' is visible at the end of the line.

```
ubuntu@ip-172-31-38-235: ~ X + v  
ubuntu@ip-172-31-38-235:~$ aws s3 ls  
2024-09-17 13:59:40 s3-raw-lab  
ubuntu@ip-172-31-38-235:~$ |
```

## Criando arquivo na EC2

```
ubuntu@ip-172-31-38-235: ~  
ubuntu@ip-172-31-38-235:~$ cat > arquivoS3.txt  
Hello world S3  
  
Enviando meu primeiro arquivo da EC2 para o bucket S3
```

Finalize com  
CTRL+C

## Enviando o Arquivo para o Bucket S3

Envie o arquivo para Bucket S3 com

```
$ aws s3 cp ./nome_do_arquivo s3://nome_do_bucket/nome_do_arquivo
```

```
ubuntu@ip-172-31-38-235:~$ aws s3 cp ./arquivoS3.txt s3://s3-raw-lab/arquivoS3.txt
upload: ./arquivoS3.txt to s3://s3-raw-lab/arquivoS3.txt
ubuntu@ip-172-31-38-235:~$ |
```

```
ubuntu@ip-172-31-38-235:~$ aws s3 ls s3://s3-raw-lab
2024-09-17 14:25:06          71 arquivoS3.txt
ubuntu@ip-172-31-38-235:~$ |
```



## O caminho inverso copia o arquivo do bucket para sua instância

```
ubuntu@ip-172-31-38-235:~$ mkdir bucket_download
ubuntu@ip-172-31-38-235:~$ aws s3 cp s3://s3-raw-lab/arquivoS3.txt ./bucket_download/download_file.txt
download: s3://s3-raw-lab/arquivoS3.txt to bucket_download/download_file.txt
ubuntu@ip-172-31-38-235:~$ cd bucket_download/
ubuntu@ip-172-31-38-235:~/bucket_download$ ll
total 12
drwxrwxr-x 2 ubuntu ubuntu 4096 Sep 17 14:30 ./
drwxr-x--- 6 ubuntu ubuntu 4096 Sep 17 14:29 ../
-rw-rw-r-- 1 ubuntu ubuntu   71 Sep 17 14:25 download_file.txt
ubuntu@ip-172-31-38-235:~/bucket_download$ cat download_file.txt
Hello world S3
```

Enviando meu primeiro arquivo da EC2 para o bucket S3

# s3-raw-lab

Informações

- Objetos
- Propriedades
- Permissões
- Métricas
- Gerenciamento
- Pontos de acesso

## Objetos (1) Informações

- 
- Copiar URI do S3
- Copiar URL
- Fazer download
- Abrir
- Excluir
- Ações
- Criar pasta
- Carregar

Os objetos são as entidades fundamentais armazenadas no Amazon S3. Você pode usar o [inventário do Amazon S3](#) para obter uma lista de todos os objetos em seu bucket. Para outras pessoas acessarem seus objetos, você precisará conceder permissões explicitamente a eles. [Saiba mais](#)

Localizar objetos por prefixo

| <input type="checkbox"/> | Nome                          | Tipo | Última modificação          | Tamanho | Classe de armazenamento |
|--------------------------|-------------------------------|------|-----------------------------|---------|-------------------------|
| <input type="checkbox"/> | <a href="#">arquivoS3.txt</a> | txt  | 17 Sep 2024 11:25:06 AM -03 | 71.0 B  | Padrão                  |

## Comandos úteis

\$ `aws s3 mb s3://bucket-name` Criar um bucket

\$ `aws s3 ls s3://bucket-name` Listar buckets e objetos

\$ `aws s3 rb s3://bucket-name --force` Excluir um bucket

\$ `aws s3 rm s3://bucket-name/example/filename.txt` Excluir um arquivo

\$ `aws s3 rm s3://bucket-name/example --recursive` Excluir um diretório

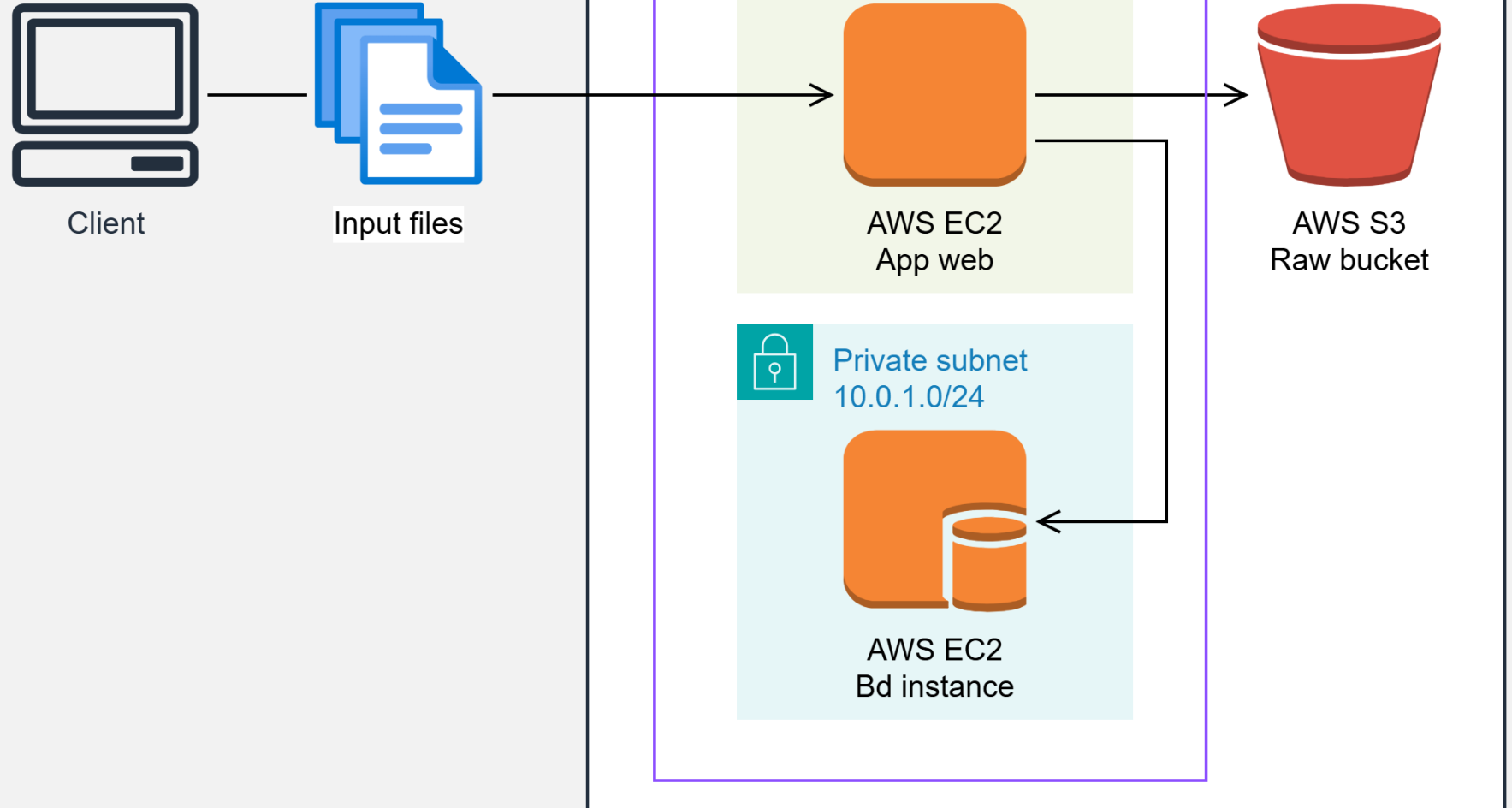
\$ `aws s3 cp s3://bucket-name/file1.txt s3://my-bucket/` Copia um arquivo de um bucket a outro

\$ `aws s3 sync <source> <target> [--options]` Sincroniza o conteúdo de um bucket e/ou diretório ou entre dois buckets

**Hands on projeto**

# Lembrando a arquitetura base

1. Configurar a arquitetura base, com pelo menos uma máquina e um bucket conectado
2. Ajustar a App web para capturar os dados e carregar um arquivo .csv no bucket **raw**.



# Links úteis

- [Arquitetura referência de Análise de Dados sem servidor na AWS | O blog da AWS \(amazon.com\)](#)
- [Life Sciences Data Collection, Storage, and Processing - Life Sciences Data Collection, Storage, and Processing \(amazon.com\)](#)
- [Customer Data Platform for Lodging \(awsstatic.com\)](#)
- [Reference Architecture Title \(awsstatic.com\)](#)
- [Centralized Log Analytics \(awsstatic.com\)](#)
- [Data Lake Architecture: 10 Critical Aspects | Learn - Hevo \(hevodata.com\)](#)
- [Data lakehouse architecture: Databricks well-architected framework - Azure Databricks | Microsoft Learn](#)
- [Guiding principles for the lakehouse - Azure Databricks | Microsoft Learn](#)
- [Download lakehouse reference architectures - Azure Databricks | Microsoft Learn](#)
- [Data lake architecture: Zones explained | by Capital One Tech | Capital One Tech | Medium](#)
- [Design patterns for an enterprise data lake using AWS Lake Formation cross-account access | AWS Big Data Blog \(amazon.com\)](#)
- [Design Patterns for Data Lakes. Data Lake is the heart of big data... | by Lackshu Balasubramaniam | Medium](#)
- [Databricks Reference Architectures for the documentation 2024-05-16 - Google Slides \(microsoft.com\)](#)
- [The scope of the lakehouse platform - Azure Databricks | Microsoft Learn](#)

**Agradeço**  
a sua atenção!

**Eduardo Verri**

eduardo.verri@sptech.school

SÃO  
PAULO  
TECH  
SCHOOL