Sumário

[Módulo 1 – Fundamentos 1](#_Toc84050106)

[Por quê Cloud? 1](#_Toc84050107)

# Módulo 1 – Fundamentos

## Por quê Cloud?

Iaas -> Paas -> Saas

## Well Architected Framework

Guia/Melhores práticas para projetar infraestrutura boas que são:

* Excelência Operacional -> Agregar valor operacional, automatizando operações. Executar Operações como código. Produzir Documentação. Prever Falhas.
* Confiabilidade -> Recuperar rapidamente (Automaticamente) de falhas. Tornar o sistema Confiável. Tornar a solução escalável automaticamente.
* Segura -> Proteger Informações. Avaliar Riscos. Gerenciar Acessos. Controle para detectar falhas de Segurança. Proteger Serviços. Confidencialidade e integridade dos dados (Importante) no desenho da solução.

Habilitar Rastreabilidade. Aplicar Segurança em todas as camadas.

* Eficiente de Desempenho -> Atender os requisitos de sistema e manter essa eficiência à medida que as mudanças na demanda e as tecnologias evoluem. Tecnologias avançadas. Usar arquitetura Serveless (Arquitetura gerenciado). Testar com mais frequência novas tecnologias.
* Otimização de Custos -> Agregar valor comercial pelo menor preço. Controlar o quanto de dinheiro está sendo gasto.

“Tudo falha, o tempo todo.” -> Diretor da Amazon.com

## Modelos de Arquitetura de Dados na Cloud

### Data Base as a Service (Banco de dados como serviço)

Não configuramos uma máquina Linux ou Windows com o banco. Contratamos apenas o sérvio de banco.

AWS -> RDS (Relation Database Service) -> Oracle, Mysql, MariaDb. Amazon Aurora ()

Soluções de DW -> Amazon Redshift (Cluster de análise de Dados). Amazon Redshift Spectrum.

Não-Relacional -> Chave-Valor(DynamoDB) . Documento(DocumentDB). Grafo (Amazon Neptune). Em memórias (Amazon Elasticache -> Memcached e Redis). Pesquisa(Amazon ElasticSearch Service)

Azure -> Banco de Dados SQL do Azure. SQL Server em Máquinas Virtuais. POstgree, MariaDB e Mysql.

Soluções de DW -> Azure SQL Data WareHouse. Synapse Analytics.

Não-Relacional -> Chave-Valor(CosmoDB) . Documento(CosmoDB). Grafo (CosmoDB). Em memórias (Azure Redis). Pesquisa(Elastic Azure)

Google Cloud -> Bare Metal (Migração para a base Oracle)

Cloud SQL (Mysqo, Sql Server)

Cloud Spanner (Oracle e DynamoDB)

Soluções de DW -> Big Query

Não-Relacional -> Chave-Valor(Cloud BigTable) . Documento(FireStore, Firebase Realtime DataBase). Em memórias (MemoryStore -> Memcached e Redis).

## Data Lake -> Desenho de Arquitetura

### Lambda ->

Duas Camadas rodando simultaneamente. São as camadas de Batch e de Speed (Rápida).

Speed Layer -> Ferramenta de Stream de Dados. Kafka, exemplo.

Diagrama

Descrição gerada automaticamente

### Kappa

Uma única camada de processamento. Será a Real Time Processamento rápido.

Diagrama

Descrição gerada automaticamente

### Unifield

Igual ao Lambda, porém contém uma camada de Machine Learn (Predições)

Diagrama

Descrição gerada automaticamente

### LakeHouse

Armazear objetos de vários tipos de dados.

Fonte única de verdade.

### Amazon S3. -> Para Data Lake

Armazenamento de Objetos.

Bucket -> Replica o banco em todas as regiões automaticamente.

Modelo Pay as you go:

Pague por GBs por mês; Tráfego para fora da AWS e outras regiões; Solicitações PUT, COPY, POST, LIST e GET

Não Paga por:

Tráfego para dentro do S3; Tráfego para fora do S3 para Amazon Cloud Front ou Amazon EC2 na mesma região.

DICA IMPORTANTE: Sempre colocar as instancias de computação na mesma região do Data Lake.

## Ferramenta de Ingestão de Dados

Batch, uma única vez.

Batch, por substituição programada.

Batch incremental

Near-Real-Time incremental.

BATCH -> SPARK; PYTHON; APACHE NIFI

Near-Real-Time -> KAFKA.

AWS -> DMS (Data Base Migration Service), para Data Lake

Eventos -> Amazon Kinesis, Google PubSub, Azure Event Hub.

## Processamento de BIG DATA

OPEN SOURCE:

Apache Spark (Batch e Real Time)

ksqlDB (Kafka – Real Time).

CLOUD:

AWS -> EMR (Batch e Real Time). Glue Job (Batch e Real Time)

Azure -> HD Insight (Batch e Real Time) Azure Stream Analytics (Real Time)

Google Cloud -> DataPRoc (Cluster Spark). Pub/Sub (Kafka). BigQuery (Processamento Batch).

DataBricks -> Processamento própria de BigData. Spark, Delta Lake.

## Data Lake – Consumo de Dados – DW e Engines – Consumir dados do Data Lake

AWS -> Amazon Athena. Em conjunto com o Glue Data Catalog.

Azure -> Azure Data Lake Analytics.

Google Cloud -> Big Query

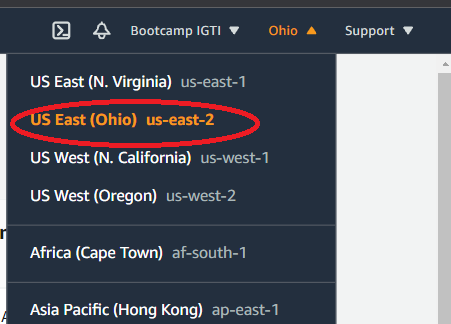
Presto (Motor do Athena utiliza ele), Trino, Apache Drill, Dremio.

# Módulo 1 – Prática – Data Bases Relacionais

*delete me*

RDS da AWS -> Implantar uma base PostGreSQL na AWS.

Link para acessar o Console da AWS: [https://us-east-2.console.aws.amazon.com/console/home?region=us-east-2#](https://us-east-2.console.aws.amazon.com/console/home?region=us-east-2)



Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Diagrama

Descrição gerada automaticamente

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Na Amazon, por default, tudo está inválido. Temos que habilitar o que queremos.

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

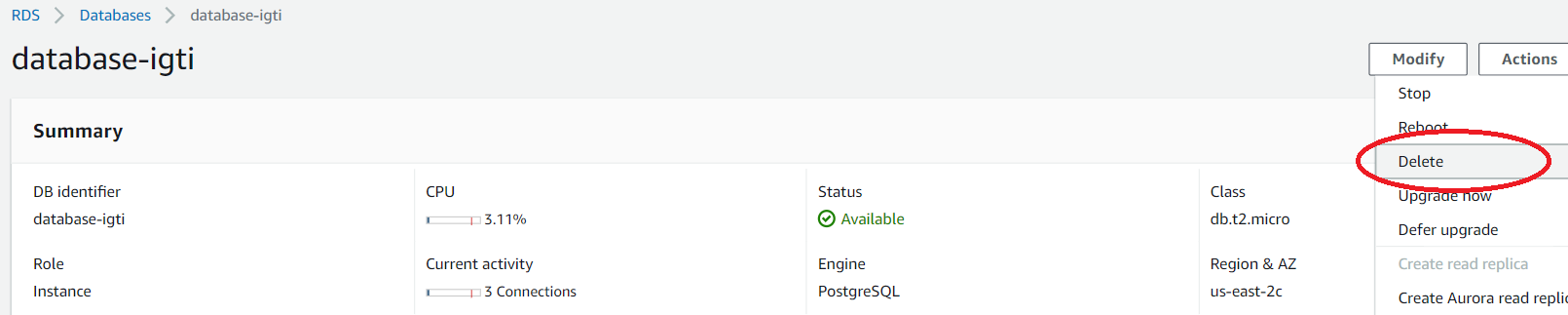
Só o Source (IP) tem acesso a este Serviço.

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo

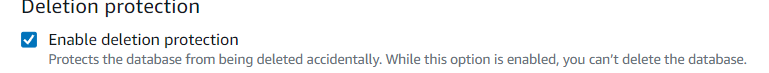
Descrição gerada automaticamente

v

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Clicar em Modify



Desabilitar para poder deletar

# Módulo 1 – Prática – Data Lake com AWS S3

Bucket um Storage de objetos.

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Quando interagimos com o Bucket S3, não precisa selecionar a zona que ficará os dados.

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Instalar o AWS CLI

Configurar o AWS CLI

Texto

Descrição gerada automaticamente

Mudar as Credenciais de acesso no IAM para poder usar o AWS externamente, fora do Site do Console da AWS.

Tela de celular com aplicativo aberto

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Conecto com o S3, crio um cliente. Faço Download de um arquivo. E faço um upload do arquivo.

Texto

Descrição gerada automaticamente

# Módulo 1 – Prática - **Extração de Dados com AWS DMS**

Serviço para migração de dados

Tela de celular com aplicativo aberto

Descrição gerada automaticamente

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente