



# SUMMIT ONLINE

AMERICAS

A N T 2 0 1

# Transforme Dados em Insights

Hugo Rozestraten

Arquiteto de soluções especialista em analíticos

AWS

# Por que escolher a AWS para data lakes e análises?

1



---

Mais fácil de  
construir data  
lakes e análises

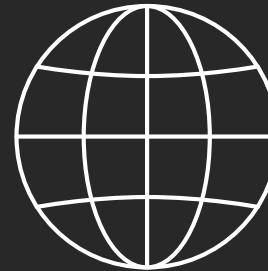
2



---

Infraestrutura  
mais segura  
para análise

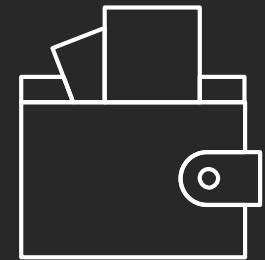
3



---

Mais abrangente  
e aberto

4



---

Mais escalável  
e econômico

# O portfólio de análises da AWS

## Dados, visualização, engajamento e aprendizado de máquina



## Análise



## Infraestrutura e gerenciamento de data lake



## Movimentação de dados

AWS Database Migration Service | AWS Snowball | AWS Snowmobile | Amazon Kinesis Data Firehose | Amazon Kinesis Data Streams |  
Amazon Managed Streaming for Apache Kafka



# Serviços de movimentação de dados

## **Movimentação de dados**

Serviços de Migração de Dados e Streaming

# Principais maneiras de mover dados para o data lake



## Movimentação de dados de data centers locais

Conexão de rede dedicada

Aparelhos seguros

Recipientes de transporte robusto

Migração de banco de dados

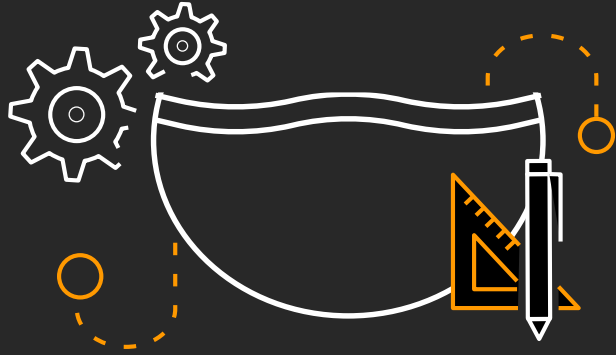
Gateway que permite que os aplicativos gravem na nuvem

## Movimentação de dados de fontes em tempo real

Conecte dispositivos à AWS

Fluxos de dados em tempo real

Transmissões de vídeo em tempo real



# Infraestrutura de data lake & serviços de gestão

## Infraestrutura e gerenciamento de data lake



Amazon S3/  
Amazon S3 Glacier



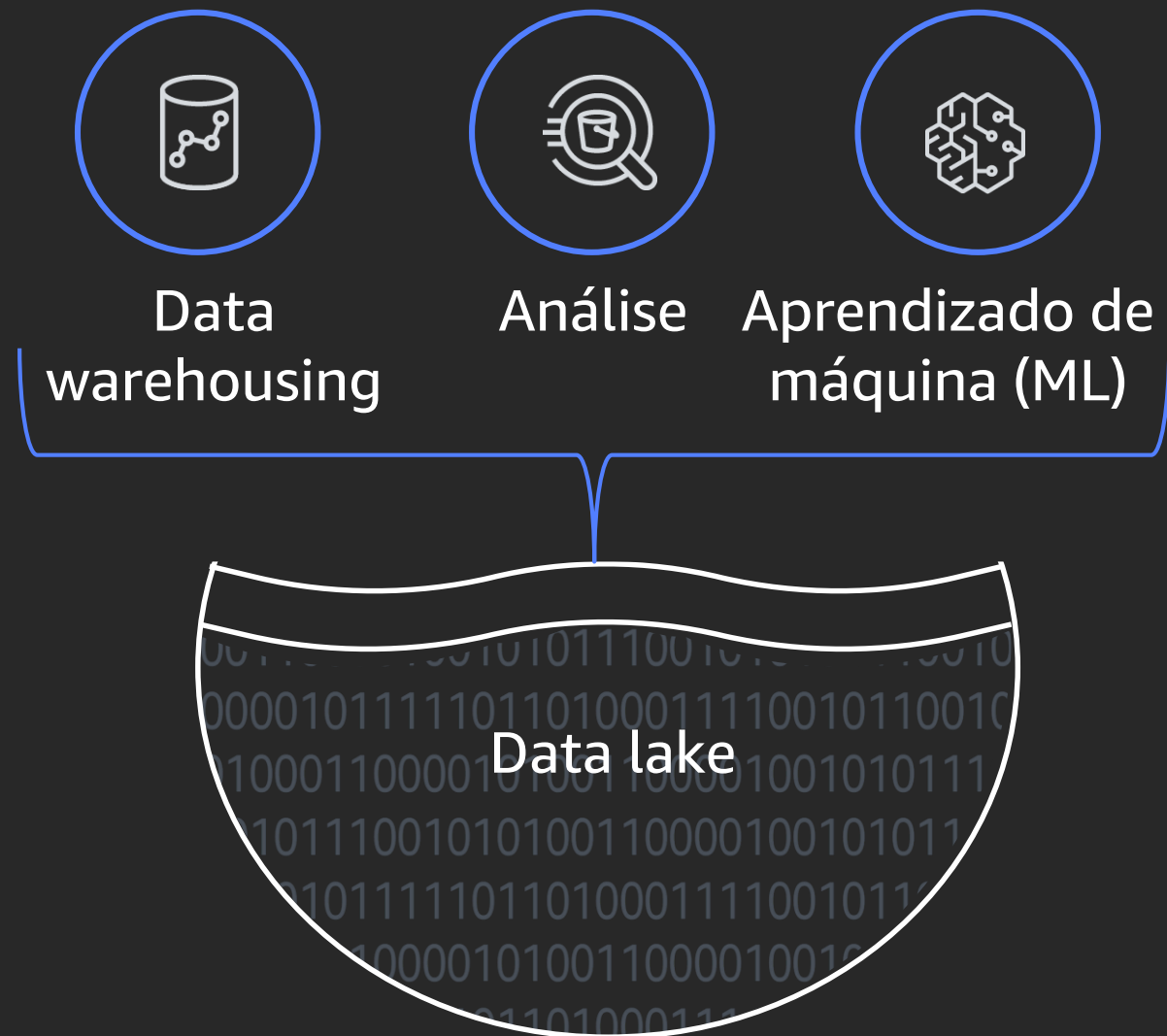
AWS Lake  
Formation



AWS Glue

# Cientes que mudam para data lakes

Reunindo o melhor dos dois mundos



Amplia ou evolui arquiteturas de data warehouse

---

Armazenar quaisquer dados em qualquer formato

---

Durável, disponível e em escala exabyte

---

Seguro, compatível e auditável

---

Execute qualquer tipo de análise a do data warehouse à predição



# Crie uma infraestrutura robusta de data lake com o Amazon S3

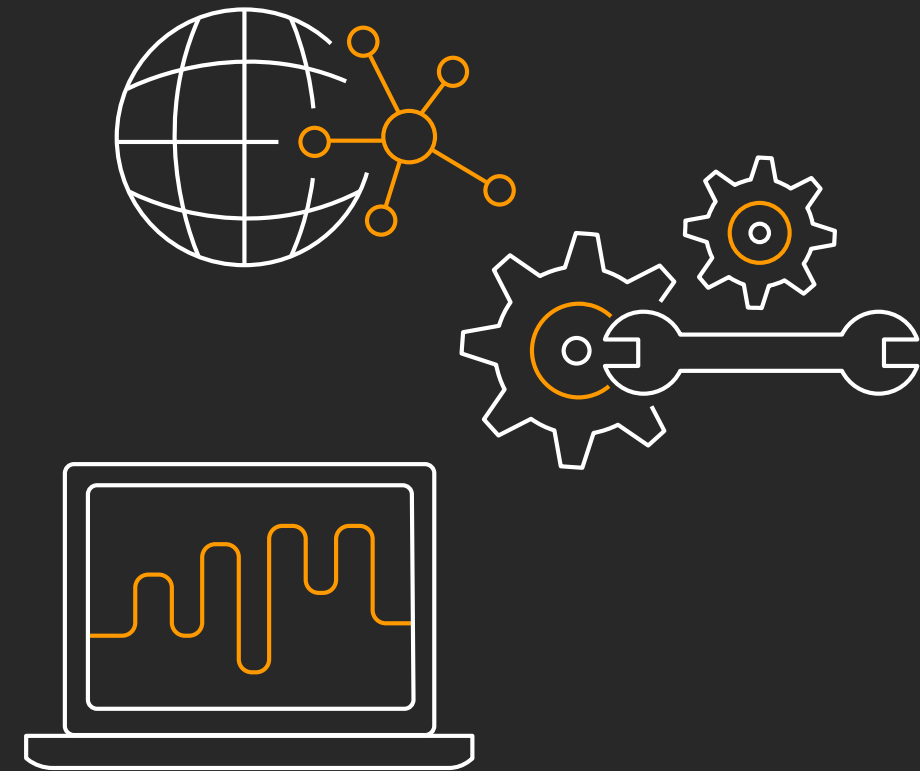
99.9999999999% de durabilidade

Recursos globais de replicação

Recursos de gerenciamento

Classes de armazenamento econômicas

Integrações com parceiros



# ETL Serverless e integração de dados com o AWS Glue

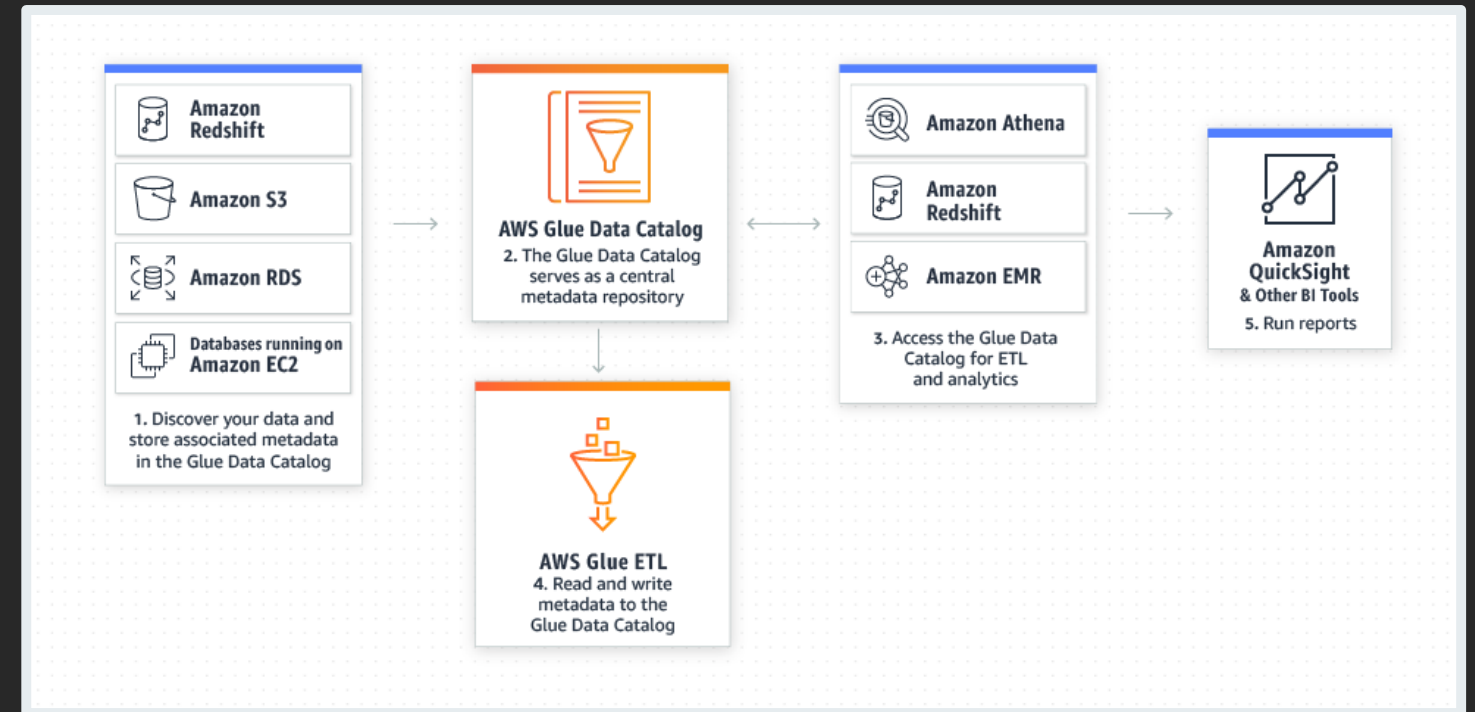
Provisionamento, configuração e dimensionamento Serverless para executar seus trabalhos ETL no Apache Spark e Python

Pague somente pelos recursos usados para trabalhos

Rastreie suas fontes de dados, identifique formatos de dados e sugira esquemas e transformações

Automatiza o esforço na criação, manutenção e execução de trabalhos ETL

Em breve: tempos de início de trabalho mais rápidos (menos de 2 minutos)



# AWS Glue

Catálogo de dados e ETL simples, flexível e econômico

## Menos trabalho



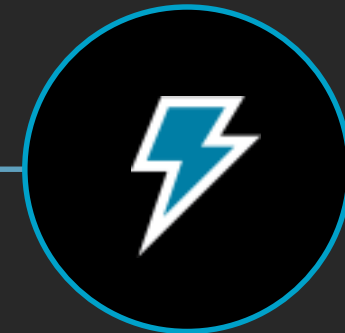
Integrado com serviços AWS – suporta Amazon Aurora, Amazon RDS, Amazon Redshift, Amazon S3, e bancos de dados na sua VPC rodando em Amazon EC2

## Serverless



Serverless – Sem infraestrutura para provisionar

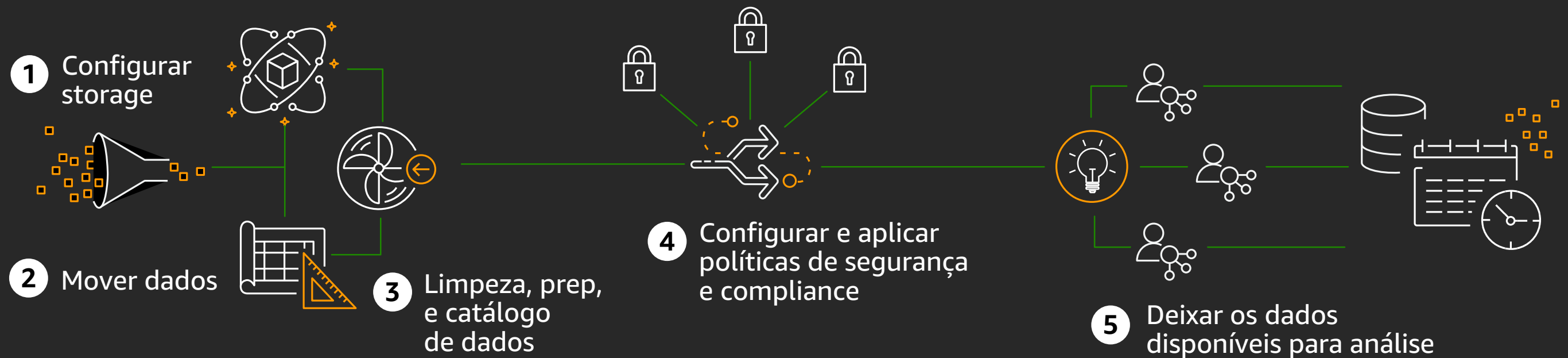
## Mais produtividade



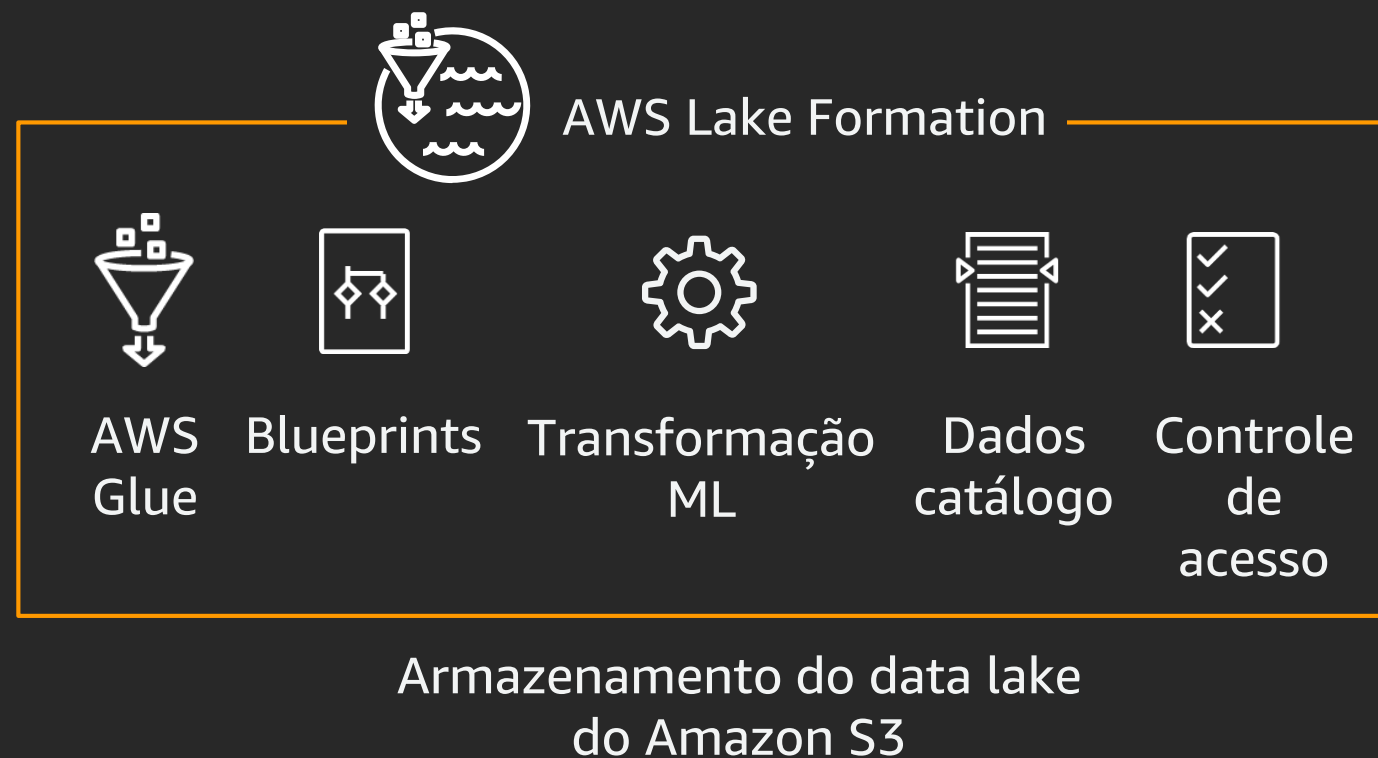
Gera códigos automaticamente para realizar transformações e cargas de

# Desafios para criar um data lake seguro

## Etapas típicas da construção de um data lake



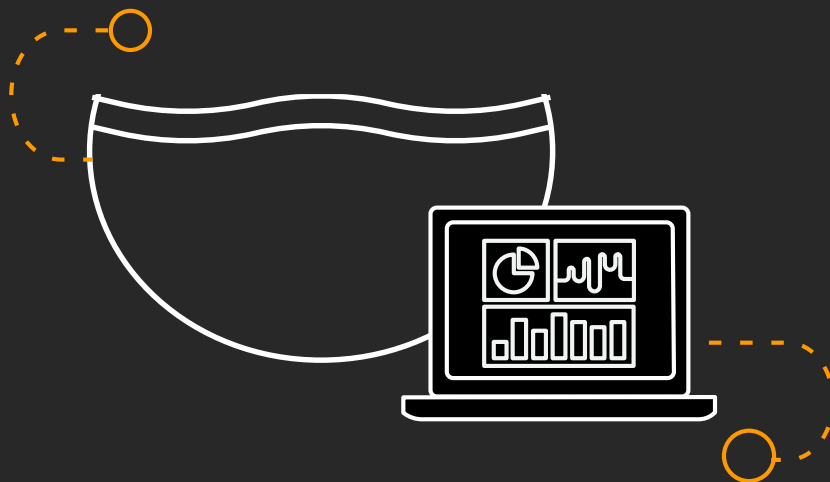
# Crie um data lake seguro em dias com AWS Lake Formation



Conjunto abrangente de ferramentas integradas permite o acesso consistente do usuário

Gerenciamento centralizado de permissões refinadas capacita agentes de segurança

Simplifica funções de ingestão e limpeza, permite que os engenheiros de dados criem mais rapidamente



# Serviços de análise



**Big Data  
processamento**



**Dados  
armazenagem**



**Análise tempo  
real**



**Análise  
Operacional**



**Consulta  
Interativo**



**Processamento  
de dados  
Serverless**

# Amazon EMR

Execute facilmente Apache Spark, Hadoop, Hive, Presto, HBase e mais aplicativos de big data na AWS

**Versões mais novas**



Atualizado com os frameworks open-source dentro de 30 dias

**Baixo custo**



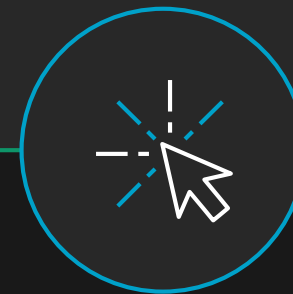
50%–80% redução de com Amazon EC2 Spot e Instâncias reservadas  
Cobrança por segundo

**Usa Amazon S3 storage**



Processe dados no Amazon S3 de forma segura e performática com EMRFS

**Fácil**



Totalmente gerenciado

# Recursos para modernização para o Amazon EMR

Guia detalhado de migração

Workshop gratuito de 2 dias

Desconstrução de cargas de trabalho

Arquitetura do estado futuro

Plano de migração e próximas etapas recomendadas

Serviços profissionais da AWS e parceiros disponíveis para implementação

Visite: [aws.amazon.com/emr/emr-migration/](https://aws.amazon.com/emr/emr-migration/)

E-mail: [emr-migration-help@amazon.com](mailto:emr-migration-help@amazon.com)

## Amazon EMR Migration Guide

How to Move Apache Spark and Apache Hadoop  
From On-Premises to AWS

August 2019





# Armazenamento de dados: Amazon Redshift

Primeiro e mais popular data warehouse na nuvem

## Data lake e AWS



Analise exabytes de dados em data warehouses, data lakes, e bancos de dados operacionais

Explore dados com vários serviços analíticos

## Melhor performance, escalável



3x mais rápido com RA3\*  
10x mais rápido com AQUA\*

*\*vs. Outras nuvens*

Adiciona capacidade **ilimitada** on demand para atender mais usuários

## Mais seguro & compliant



Segurança AWS-grade (e.g., VPC, criptografia com AWS KMS, AWS CloudTrail)

Certificações como SOC, PCI DSS, ISO, FedRAMP, and HIPAA

## Baixo custo



Otimize custos pagando processamento e storage separadamente

1/10 do custo de um data warehouse tradicional at \$1,000/TB/year

Até 75% menos do que outros data warehouses com custo previsíveis

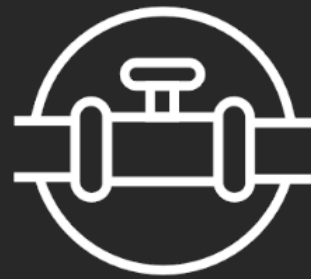
# Em tempo real: Amazon Kinesis

Colete, processe e analise facilmente fluxos de vídeo e dados em tempo real



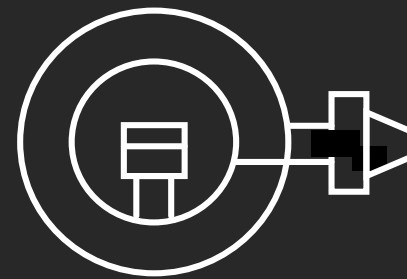
## **Amazon Kinesis Video Streams**

Capture, processe e armazene fluxos de vídeo para análise



## **Amazon Kinesis Data Streams**

Crie aplicativos personalizados que analisem fluxos de dados



## **Amazon Kinesis Data Firehose**

Carregar fluxos de dados em armazenamentos de dados da AWS

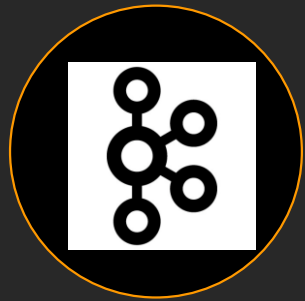


## **Amazon Kinesis Data Analytics**

Analisar dados fluxos com SQL

# Amazon Managed Streaming for Apache Kafka (Amazon MSK)

Um serviço Apache Kafka totalmente gerenciado, altamente disponível e seguro



Apache Kafka  
gerenciado  
para você



Totalmente  
gerenciado



Alta  
Disponibilidade



Múltiplos níveis  
de segurança

# Análise operacional

Amazon Elasticsearch Service totalmente gerenciado, escalável e seguro

Open-source  
Elasticsearch APIs,  
Kibana, and Logstash



Open-source  
Elasticsearch APIs  
Kibana gerenciado  
Integração com Logstash

Totalmente  
gerenciado



Elasticsearch clusters em minutos – simplifica provisionamento hardware, instalação e patching, recuperação de falhas, backups, e monitoramento

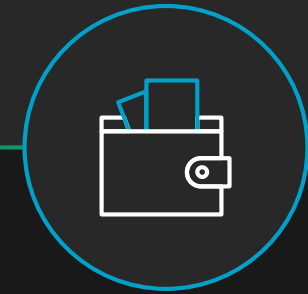
Escalável, Seguro  
e compliant



Escale clusters para cima e para baixo via API call ou poucos cliques  
VPC, dados criptografados no storage e em trânsito

Compliance – HIPAA,  
PCI DSS, and ISO

Pague  
pelo uso



Cargas otimizadas,  
sem gasto antecipado

# Open Distro para Elasticsearch

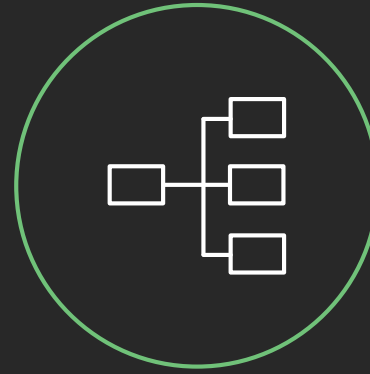


Uma distribuição licenciada do Apache 2.0 para o Elasticsearch,  
e com segurança empresarial, alertas, SQL e muito mais



## 100% open source

Liberdade para ver, usar,  
mudar e distribuir o código



## Nível Enterprise

Entrega segurança e  
configurações avançadas  
como alertas, SQL  
e diagnósticos



## Community-driven

Provê Liberdade e facilidade  
para contribuir em  
mudanças distro

# Comece a usar opções flexíveis de implantação



Visite o [site](#) para Open Distro para Elasticsearch



O Docker



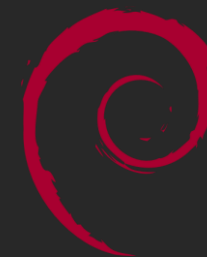
Baixe o Elasticsearch e pacotes Kibana



RPM



Carregar e consultar dados



Debian

# Amazon Athena

Serviço de consulta interativa Serverless

Query  
instantânea



Setup de custo zero  
Aponte para o S3 e  
faça queries

Pague por query



Pague somente pelas  
queries que executa  
Economize 30%–90% com  
compressão Snappy/gzip  
(formatos Orc e Parquet)

Use o storage S3



ANSI SQL  
JDBC/ODBC drivers  
Múltiplos formatos,  
de compressão, joins  
complexos e tipos de dados

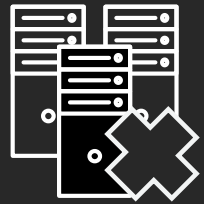
Fácil



Serverless.  
zero infraestrutura,  
zero administração  
Integrado com  
Amazon QuickSight

# Análise Serverless

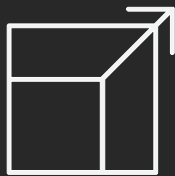
Forneça análises sob demanda no data lake



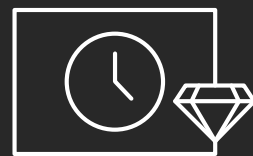
Sem servidor,  
administração zero



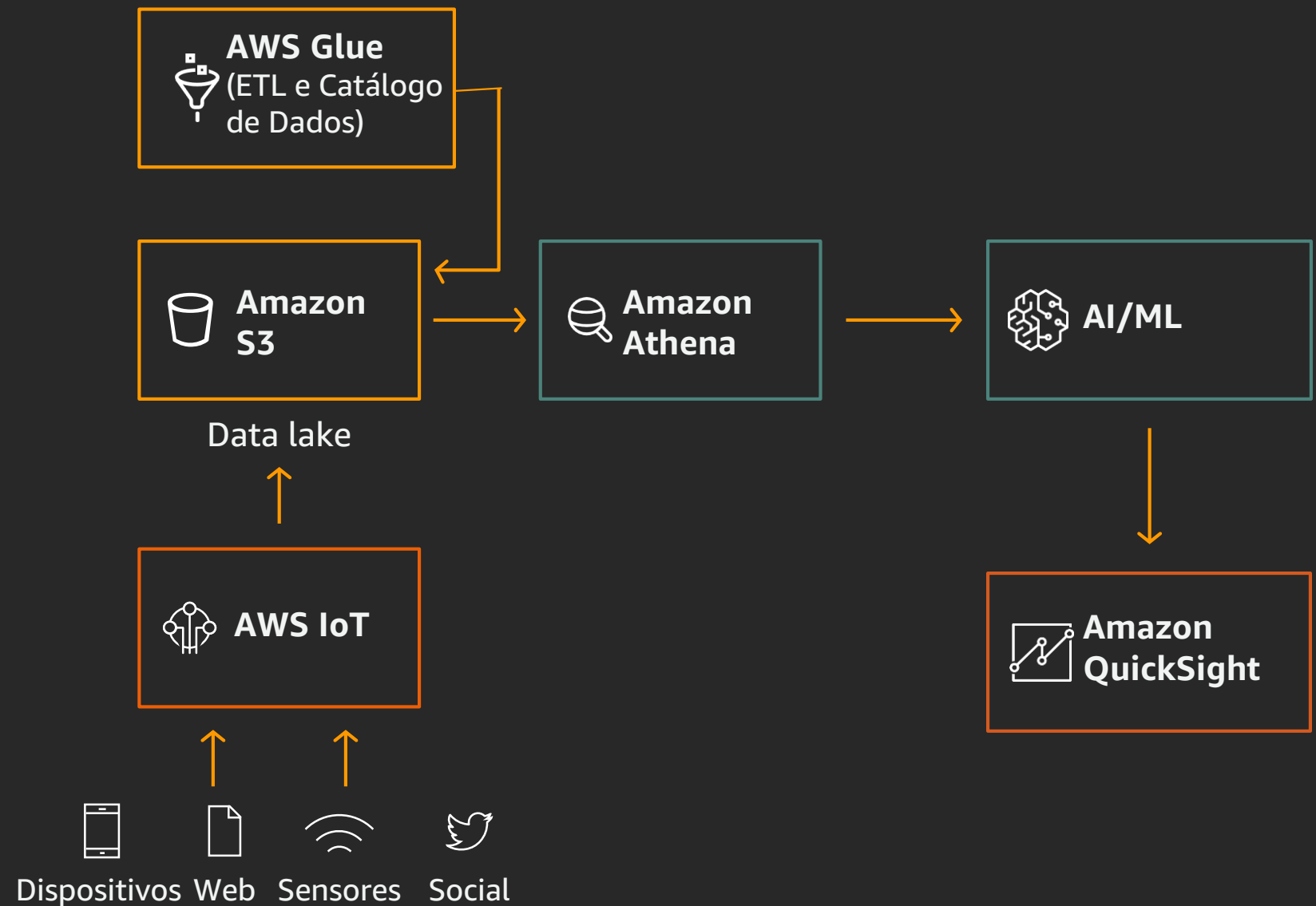
Nunca pague  
por recursos  
ociosos



Dimensiona recursos  
automaticamente com uso



Disponibilidade e  
tolerância a falhas  
integradas







# Dados, visualização, engajamento Serviços de aprendizado de máquina

## Dados, visualização, engajamento e aprendizado de máquina



**Dados**



**Painéis**



**Amazon Pinpoint**



**Análise preditiva**

# AWS Data Exchange

Encontre e subscreva facilmente dados de terceiros na nuvem

**Encontre facilmente  
diversos dados em  
um só lugar**



>1,000 produtos de dados  
>80 provedores de dados incluindo  
Dow Jones, Change Healthcare,  
Foursquare, Dun & Bradstreet,  
Thomson Reuters, Pitney Bowes,  
LexisNexis, e Deloitte

**Analise dados  
facilmente**



Faça download ou copie para  
Amazon S3  
Combine, analise, e modele com  
dados existentes  
Analise dados com Amazon EMR,  
Amazon Redshift, Amazon  
Athena, e AWS Glue

**Acesse dados de terceiros  
de maneira eficiente**



Simplique o acesso ao dado –  
sem necessidade de receber  
medias físicas, FTP gerenciado,  
ou integre com APIs  
Minimize revisões legais  
e negociações

# Amazon QuickSight

Primeiro serviço de BI criado para a nuvem com preços de pagamento por sessão e insights de ML

## Escalabilidade



Escale de 1 a 10,000+ usuários em minutos

Pay as you go

## Serverless



Crie Dashboards em minutos

Sem servidores na sua conta

## Integrado com Serviços AWS



Seguro, acesso privado

Integrado com data lake S3  
permissões com AWS IAM e  
AWS Lake Formation

## Suporte API



Crie conteúdos e faça onboard com código

Fácil embed nas suas aplicações

# Cientes referência Latino Americanos



mercado  
libre

YoUSE

 webmotors

Arezzo & Co

 reachcore

mobile  
we think mobile

UFM  UNIVERSIDAD  
FRANCISCO  
MARROQUÍN  
VERITAS • LIBERTAS • JUSTITIA

 despegar

Sua  úsica

 BPP

 VivaReal

**Unimed**   
Belo Horizonte

 **EMBRAER**

**GFG** GLOBAL  
FASHION  
GROUP

 GuiaBolso  
Você no controle da sua vida

 **kueski**

  
CALEIDOS

**embraco**  
COMPRESORES

# Mais data lakes e análises do que em qualquer outro lugar

Dezenas de milhares de data lakes são executados na AWS em todos os setores

3M Science.  
Applied to Life.™

3VICTORS  
TRAVEL BI

Acquia

A+E

amazon

Adobe

AMGEN

ATLASSIAN

Agero

airbnb

BARCLAYS

Capital One

Chick-fil-A

CIVIS

CIVITAS  
LEARNING

CoffeeMeetsBagel

dataxu

CROWDSTRIKE

DOMO

EA

EPIC  
GAMES

FAST RETAILING

FINANCIAL TIMES

FOURSQUARE

FUNNY  
& DIE

GP Georgia-Pacific

gogo

HEARST corporation

hudl

HYATT™

instacart

JustGiving

lyft

MONSTER

mparticle

Nasdaq

NETFLIX

docomo

NU  
SKIN

Olive

OLX GROUP

Panasonic

Pfizer

Pinterest

pitney bowes

real networks

REDFIN

ROVIO

SCHOLASTIC

slack

SONOS

reddit

T-Mobile

Sysco

THOMSON REUTERS

ticketmaster

TUNE

verizon

yelp

Zillow

ZUMBA

# Demo

# Aprenda com o treinamento da AWS e Certificação

Recursos criados pelos especialistas da AWS para ajudá-lo a criar e validar habilidades de análise de dados



Novo curso digital gratuito – **Fundamentos da Análise de Dados**



Ofertas em sala de aula, apresentam instrutores especializados da AWS e laboratórios práticos



Valide a experiência com **AWS Certified Data Analytics – Specialty**

Visite [aws.amazon.com/training/paths-specialty/](https://aws.amazon.com/training/paths-specialty/)



# Obrigado!

Hugo Rozestraten  
hhug@amazon.com