

# **Introdução Big Data e engenharia de dados**





# Sobre mim





# Sobre mim



- 20+ anos de experiência trabalhando em grandes organizações;
- Atuação como arquiteto de soluções, coordenador de equipes de desenvolvimento e infraestrutura além de Líder técnico de times e projetos;
- Experiência com projetos e tecnologias distintas e aprendendo constantemente;

# O que é Big Data?





# Quando o dado vira Big Data?

Coletar

Armazenar

Organizar

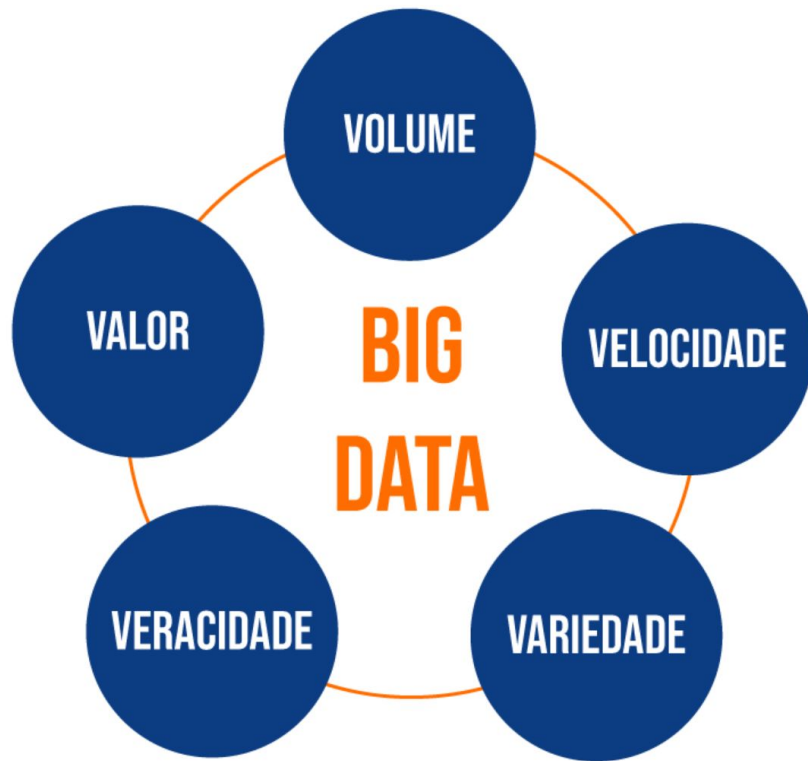
Analisar

Mover

Compartilhar



# Os 5 V's do Big Data





# Organizações orientadas a dados



# Por que agora?

- Dados e ciência de dados não são uma coisa nova. Sempre existiu por exemplo o censo, então por que agora?

## Poder computacional

Cloud;  
Acesso facilitado;  
Custo reduzido;  
Facilidade no gerenciamento;

## Acesso aos dados

Maior disponibilidade de dados;  
IoT;  
Redes sociais;  
Etc.

## Os algoritmos

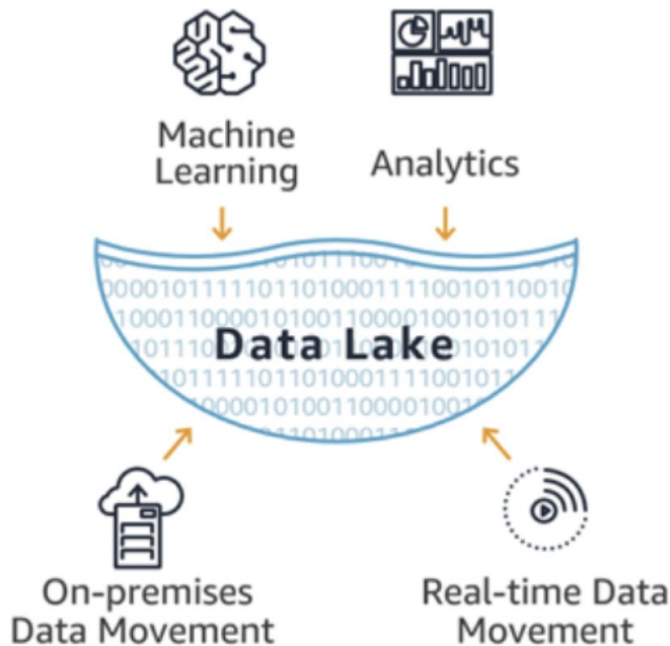
Inteligência de como tratar os dados;  
Avanços recentes;  
Paridade humana;





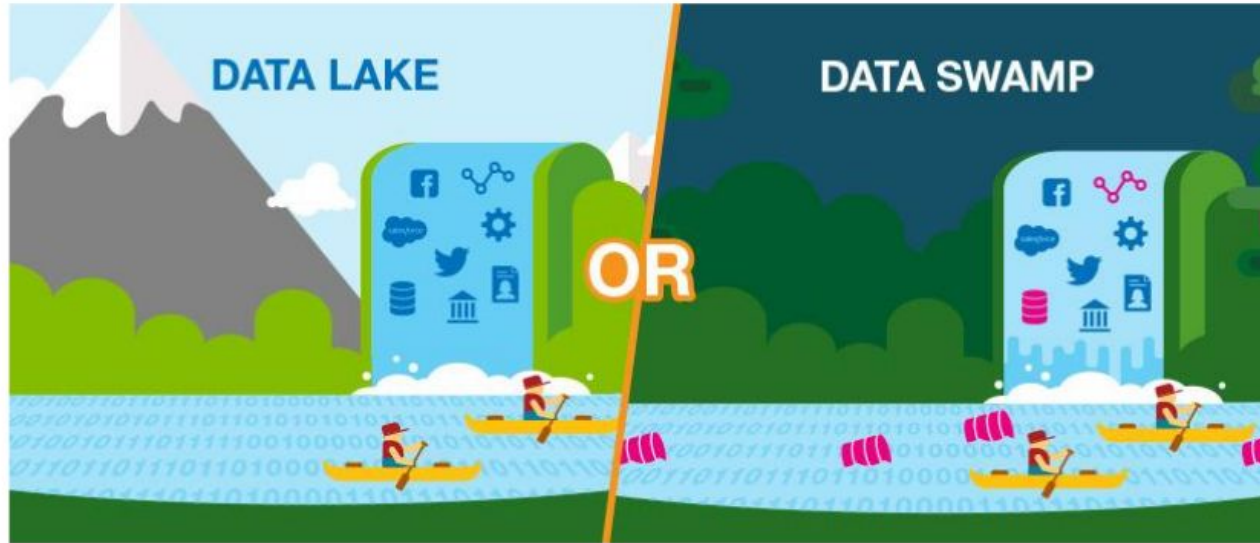


# O que é um Data Lake





# Desafios do Big Data



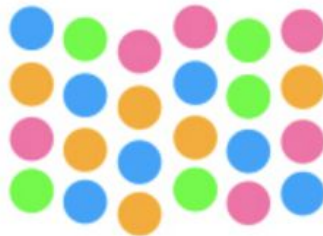


# Estruturas de dados

**Estruturados**



**Semiestruturados**



**Não estruturados**

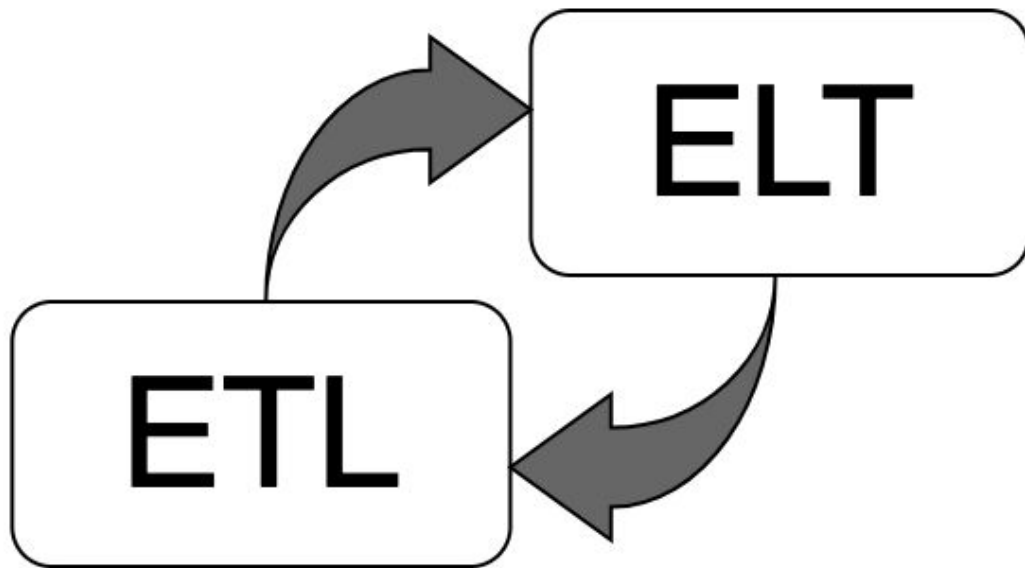


# Metadados





# ETL e ELT





# De onde vem os dados?



## 4 principais fontes: (pilares)

- Clientes
- Informação sobre sua operação
- Funcionários
- O produto em si



# Data lake na nuvem



- desempenho
- escalabilidade
- confiabilidade
- disponibilidade
- tempo de implantação mais rápido
- operação reduzida

# Pessoa engenheira de dados

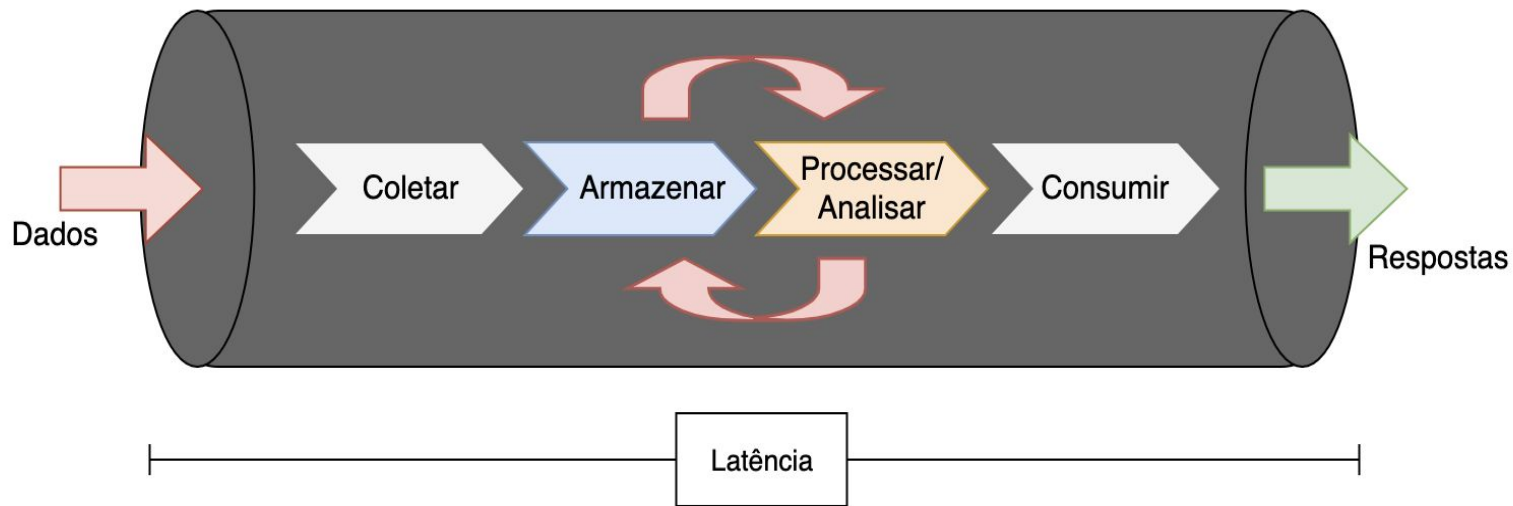
- Papel:
  - Decidir quais ferramentas são adequadas para o trabalho a ser feito;
  - Linguagens como Python, Java ou Scala.





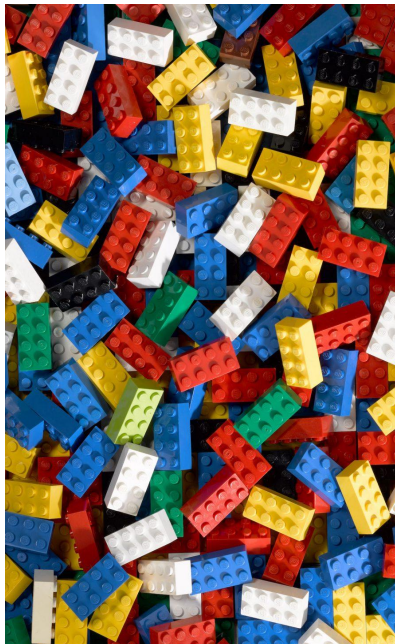


# Big Data "Pipeline"

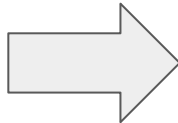




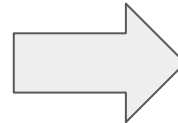
# Processamento e análise de dados



Dados de origem



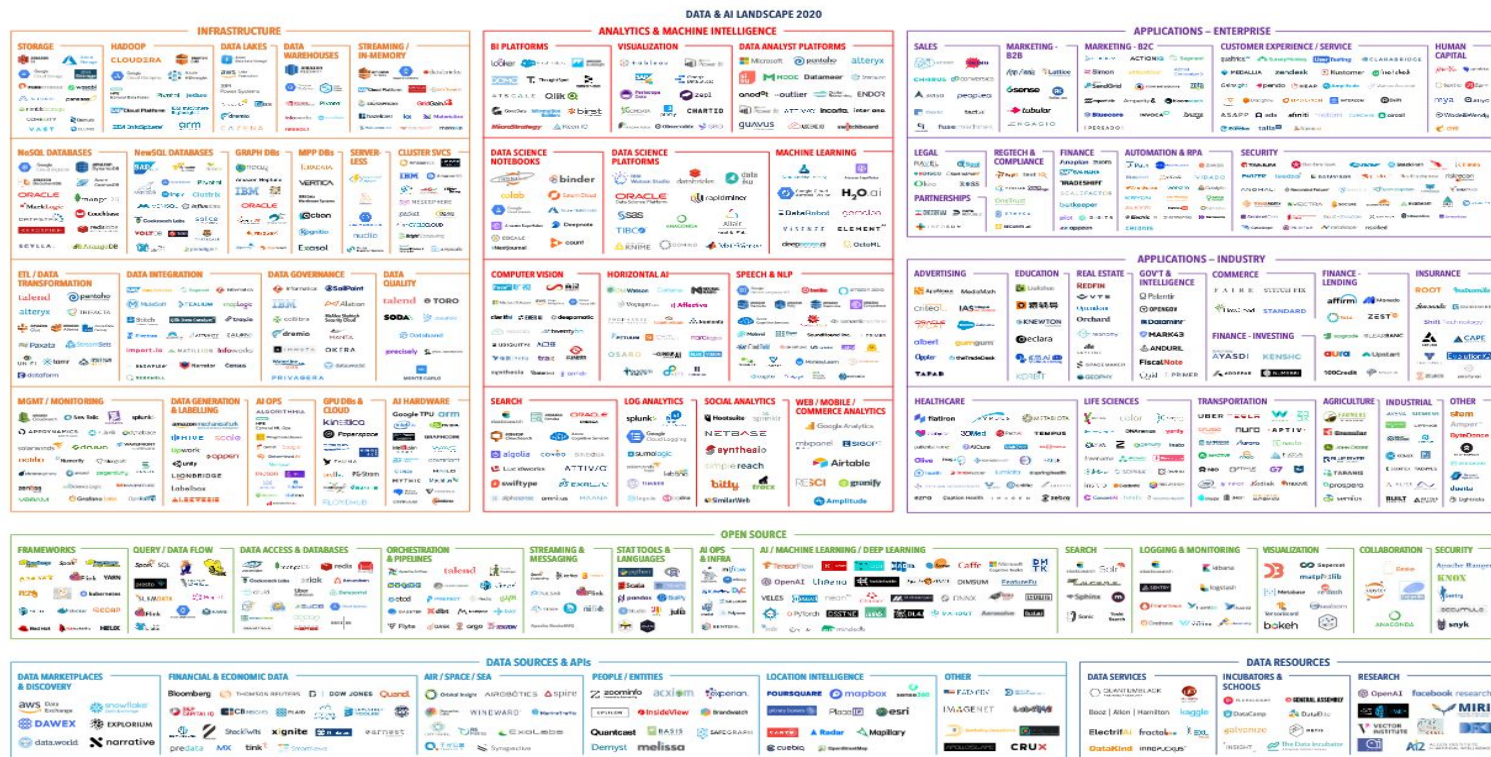
Inspecionados, organizados e armazenados



Utilizados para criar algo de valor

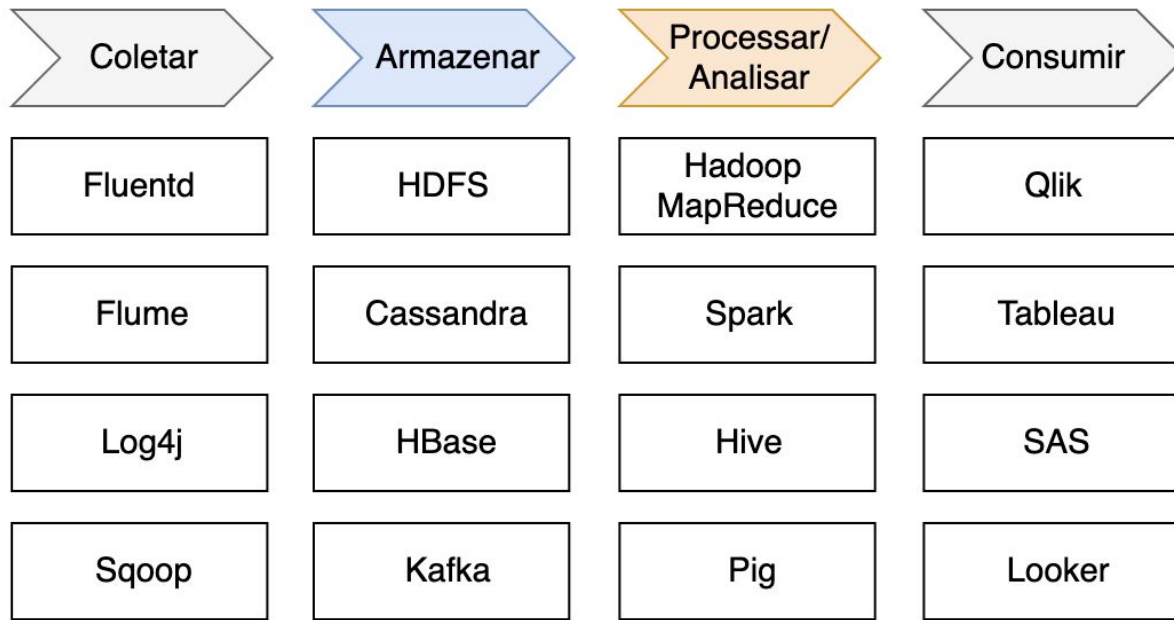


# Big Data landscape



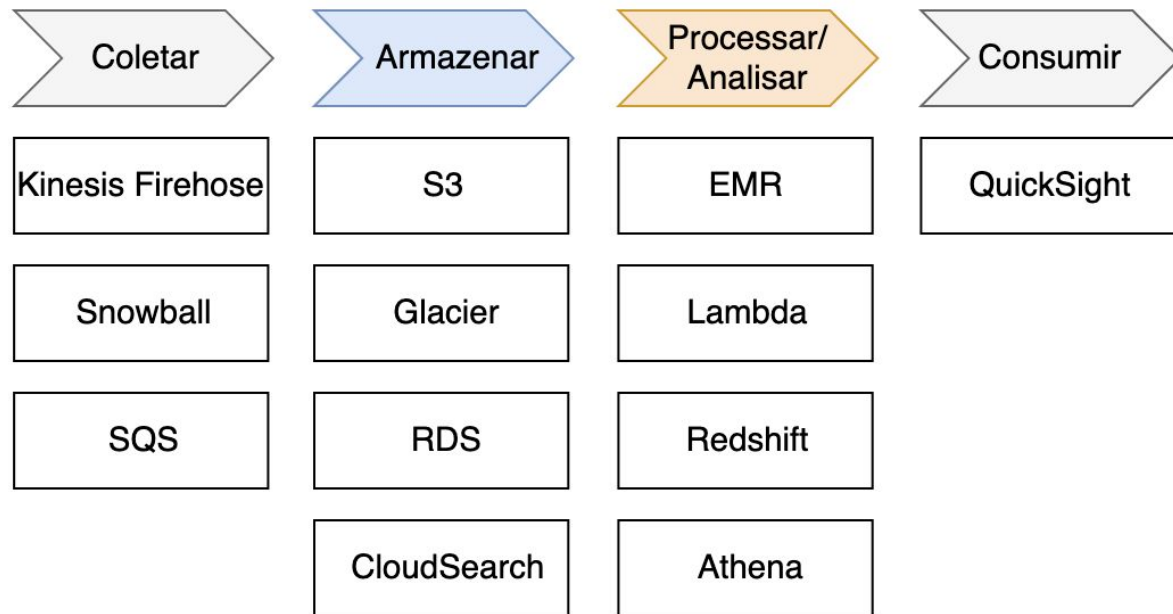


# Big Data "Pipeline" - Ferramentas



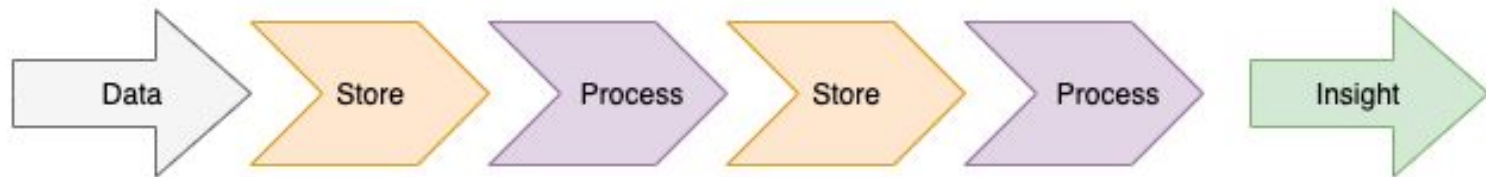


# Big Data "Pipeline" - Ferramentas





# Princípios de arquitetura de projetos de Big Data

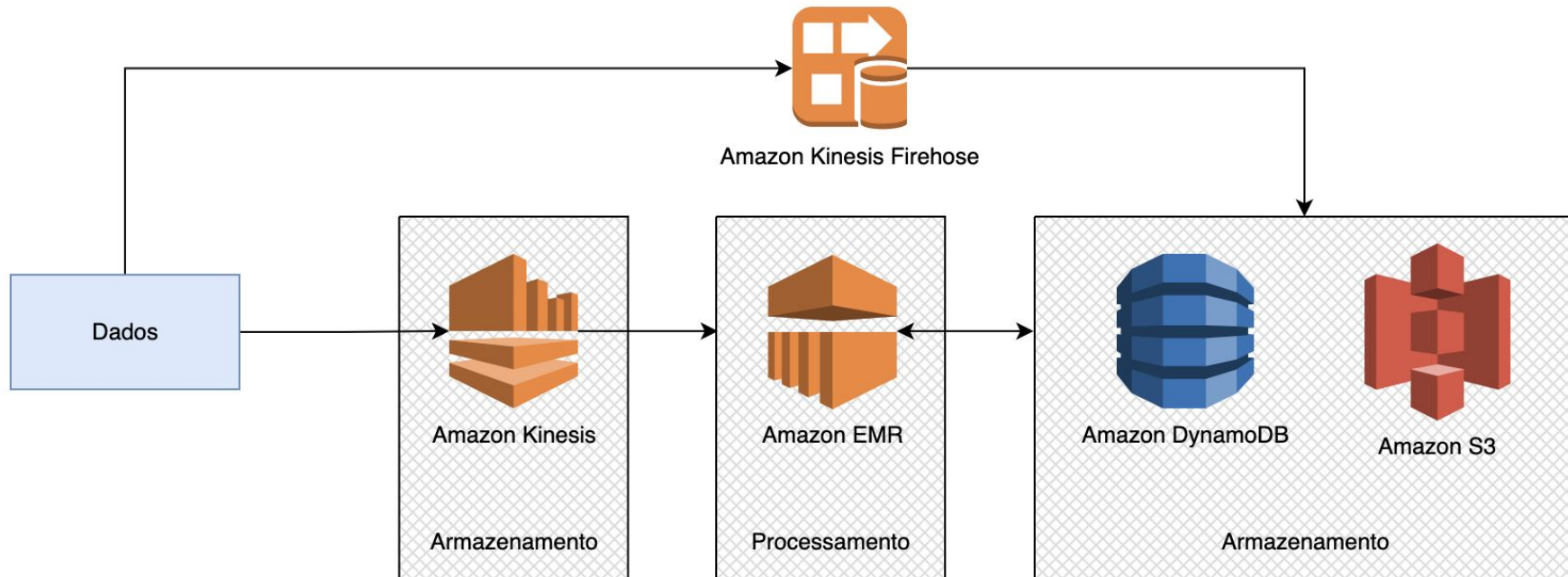


- Arquitetura descentralizada ou desacoplada
- Ferramentas específicas para cada tarefa
- Tolerância a falha
- Maior agilidade



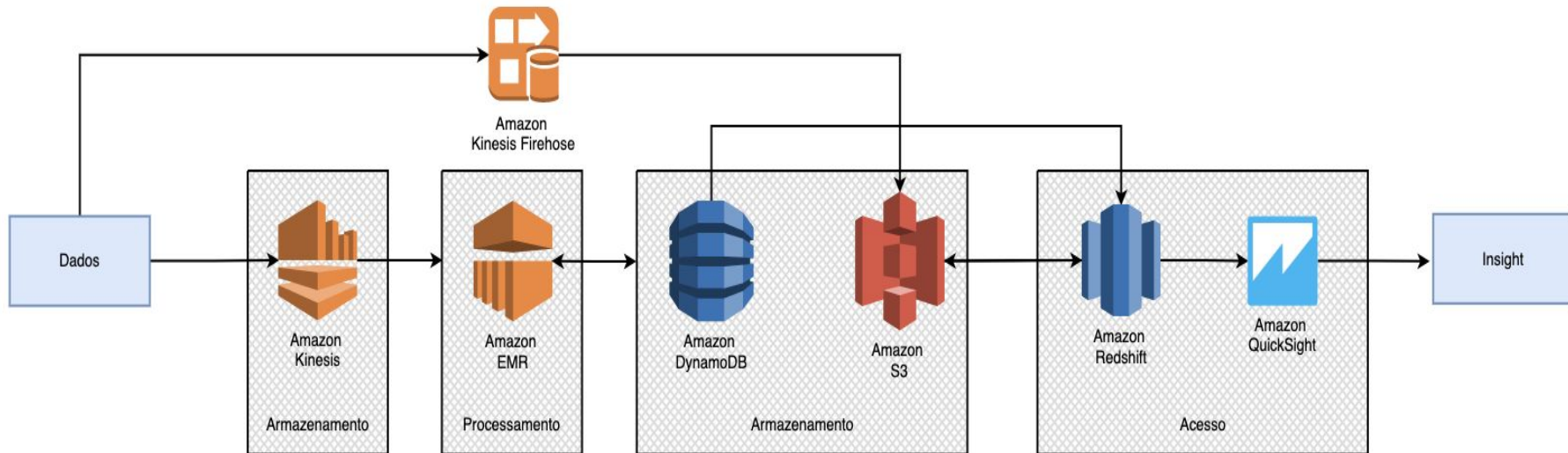


# Arquitetura desacoplada





# Arquitetura desacoplada - flexibilidade





# Projetos de Big Data





**Como os Seahawks estão usando um  
data lake para melhorar seu jogo**

SEATTLE  
SEAHAWKS



# Contextualização do case



- Time da NFL;
- Cidade de Seattle;
- Campeão em 2013.

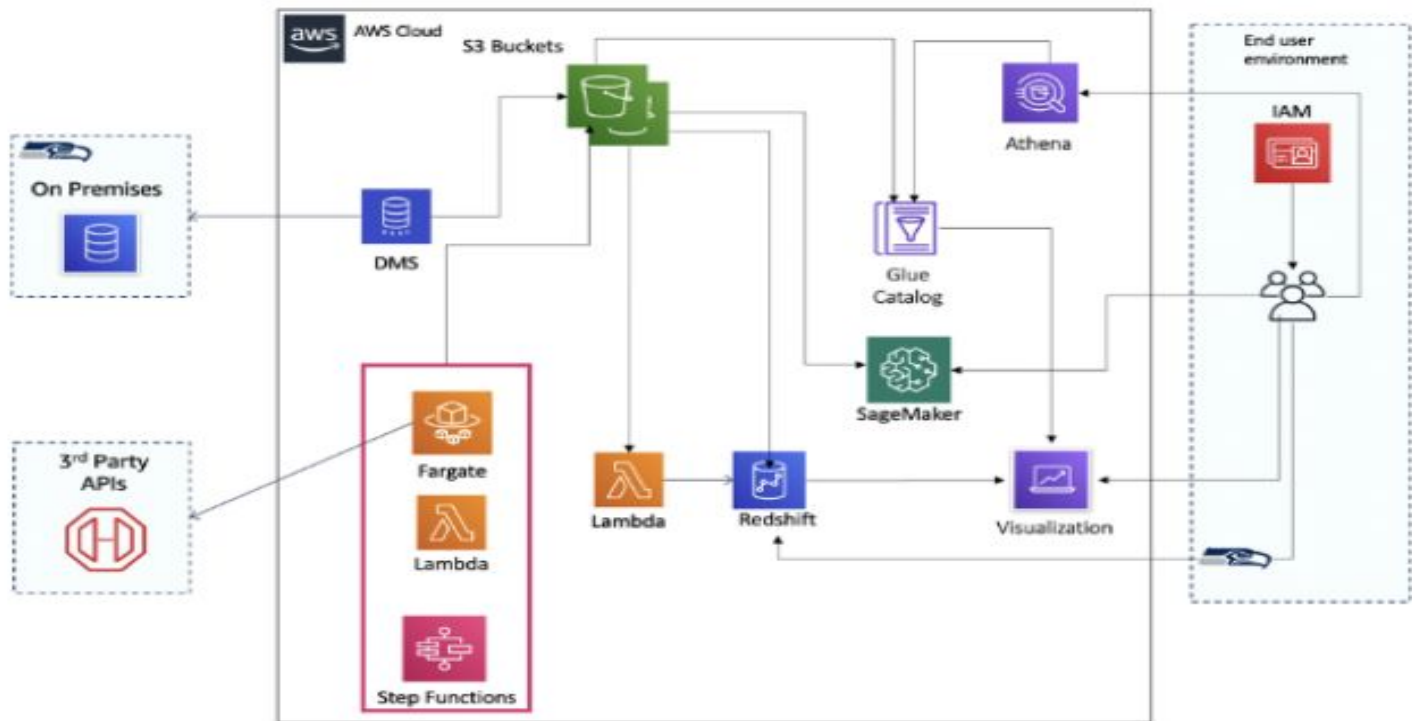
## Objetivo:

- Melhorar a avaliação e aquisição de talentos;
- Melhor observação sobre a saúde do jogador e seus tempos de recuperação;
- Planejamento do jogo.

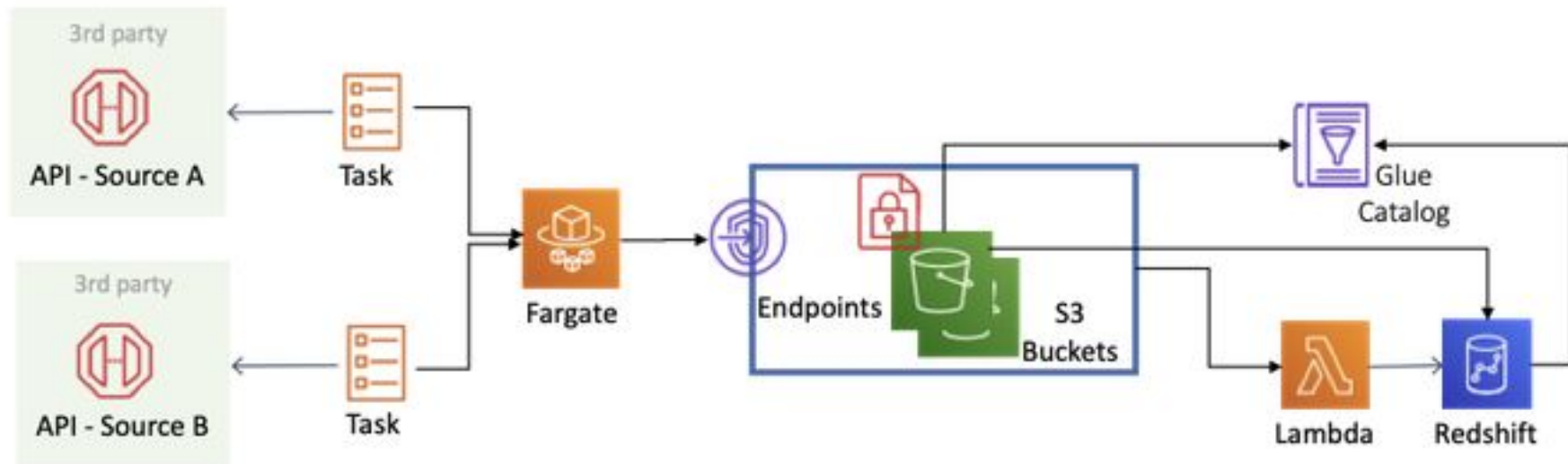
## Origem de dados:

- NFL Next Gen Stats;
- Pro Football Focus;
- Dados de telemetria dos jogadores de sensores IoT;
- Jogadas marcadas em vídeos dos treinos e jogos.

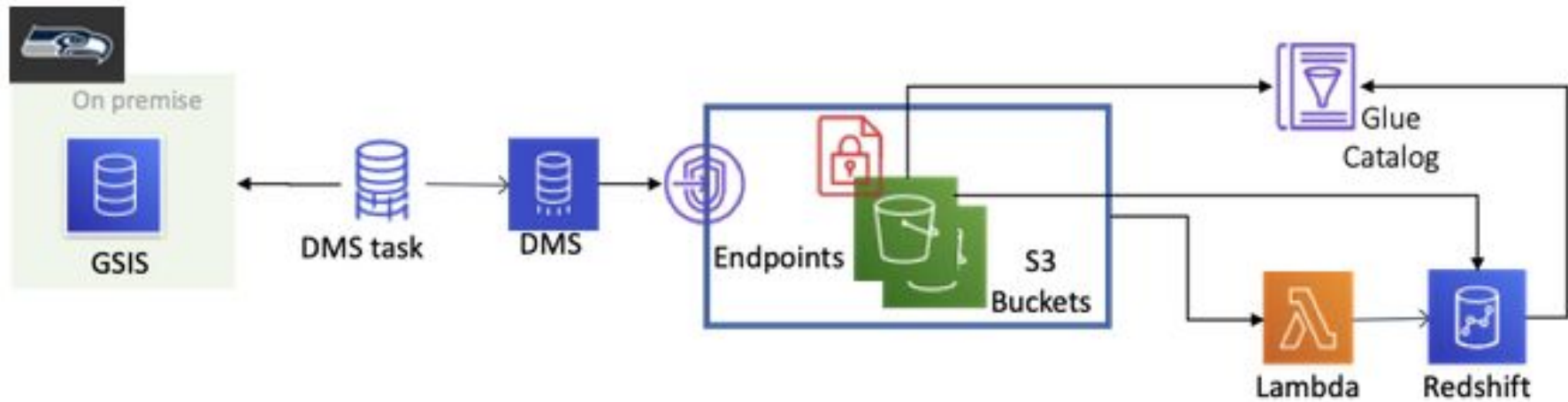
# Seahawks data lake



# Ingestão de dados de APIs



# Ingestão de dados do on-premises



# Dicas de carreira e estudos





# Principais certificações



## Professional

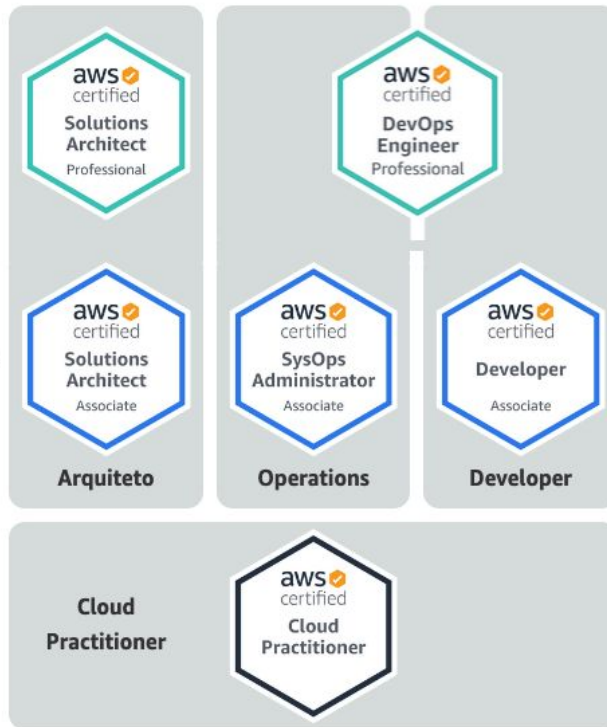
Dois anos de soluções experiência abrangente criando, operando e solucionando problemas usando a Nuvem AWS

## Associate

Um ano de experiência na solução de problemas e implementação de soluções usando a Nuvem AWS

## Foundational

Seis meses de Nuvem AWS fundamental e conhecimento do setor



## Specialty

Experiência técnica em Nuvem AWS no domínio Specialty conforme especificado no guia de exame







# Principais certificações



## Microsoft Certified: Azure Data Engineer Associate

**Requisitos:** Exame DP-203

Os Engenheiros de Dados da Azure integram, transformam e consolidam dados de vários sistemas de dados estruturados e não estruturados em estruturas adequadas para a construção de soluções de análise.

Um início opcional para quem é novo no Engenheiro de Dados



CERTIFICAÇÃO DE CONCEITOS BÁSICOS  
**Microsoft Certified: Azure Fundamentals**



CERTIFICAÇÃO DE ASSOCIADO  
**Microsoft Certified: Azure Data Engineer Associate**

Preencha uma certificação de associado



# Principais certificações



Google Cloud

## Foundational

Recommended experience:  
No hands-on experience with  
Google Cloud is required.

Cloud Digital Leader

## Associate

Recommended experience: 6+  
months building on Google Cloud.

Cloud Engineer

## Professional

Recommended experience: 3+ years  
industry experience, including 1+ years  
on Google Cloud.

Cloud Architect

Cloud Developer

Data Engineer

Cloud DevOps  
Engineer

Cloud Security  
Engineer

Cloud Network  
Engineer

Collaboration  
Engineer

Machine Learning  
Engineer

# Dúvidas?





**Fabio Saito**

Sr. Customer Solutions Manager at  
Amazon Web Services (AWS)

