
RECIFE DATA
SATURDAY

APOIO

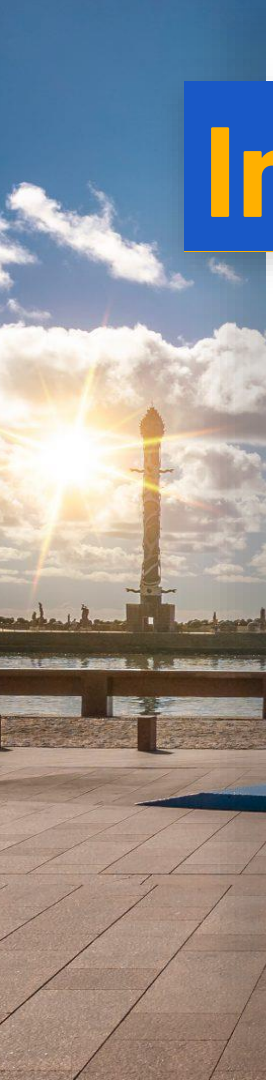
- LS Treinamentos
- Consulta BD
- Big Data Navigators
- UPE

QUEM EU?

- + 12 anos na área de TI/Dados.
- Sócio na Empresa 4BI
- Bacharelado em Engenharia de produção
- + 5 anos como Consultor
- Especialista ERP Totvs RM
- Data Viz Power BI
- Hobby – Cervejeiro

Ícaro Macedo

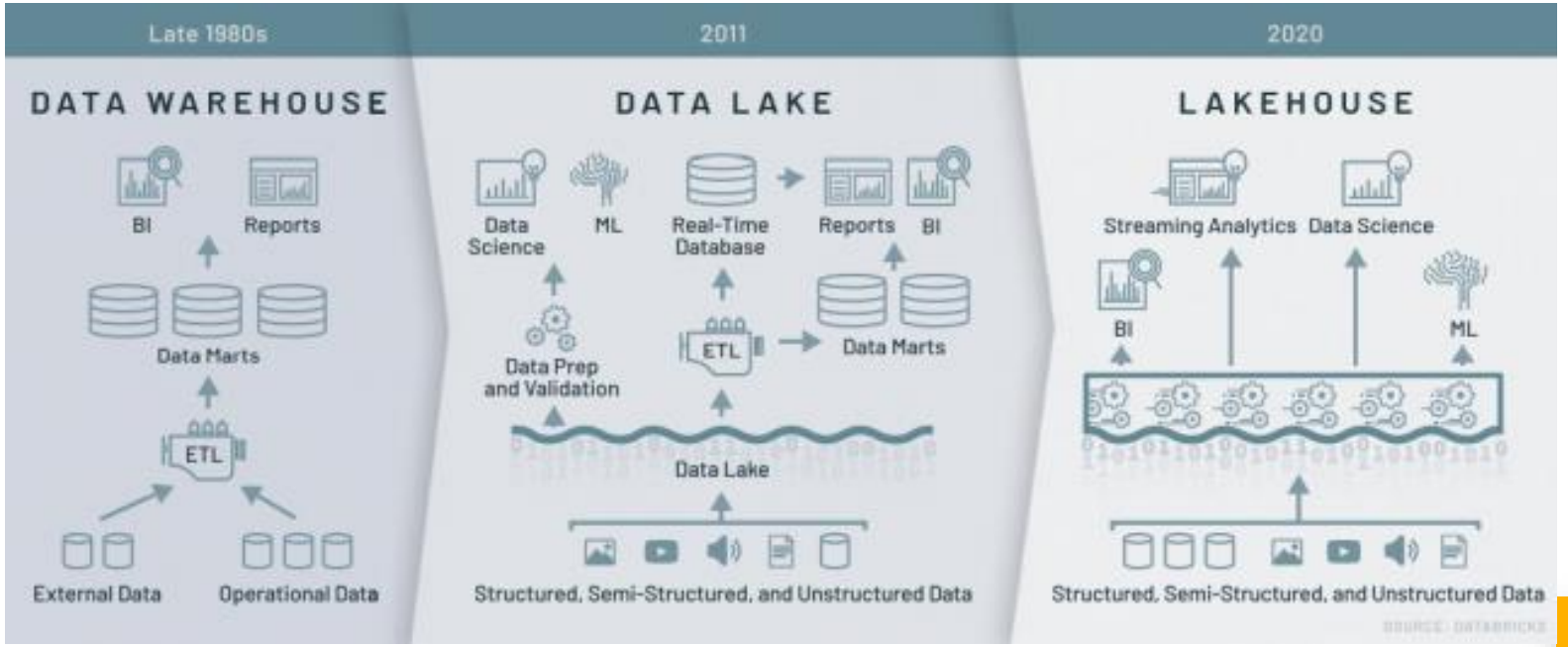
Evolução da Engenharia de dados e seus desafios.

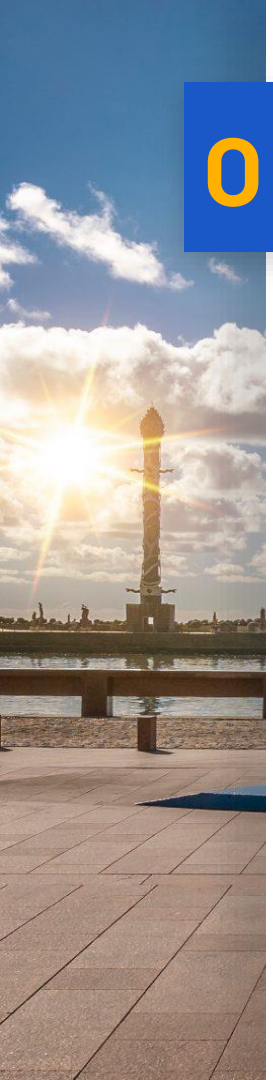


Introdução à palestra

Nesta palestra, vamos explorar a evolução da Engenharia de Dados e a cronologia dos conceitos de datawarehouse, datalake, lakehouse, data vault e data mesh.

Evolução: Arquitetura de dados





O que é um Data warehouse

Um data warehouse é um sistema de armazenamento de dados projetado para armazenar grandes volumes de dados de várias fontes, de modo que possam ser facilmente analisados e utilizados para tomada de decisões. Ele é uma parte essencial da estratégia de business intelligence (BI) de uma organização, pois permite a análise de dados históricos e atuais para identificar tendências, padrões e insights que podem impulsionar o sucesso dos negócios.

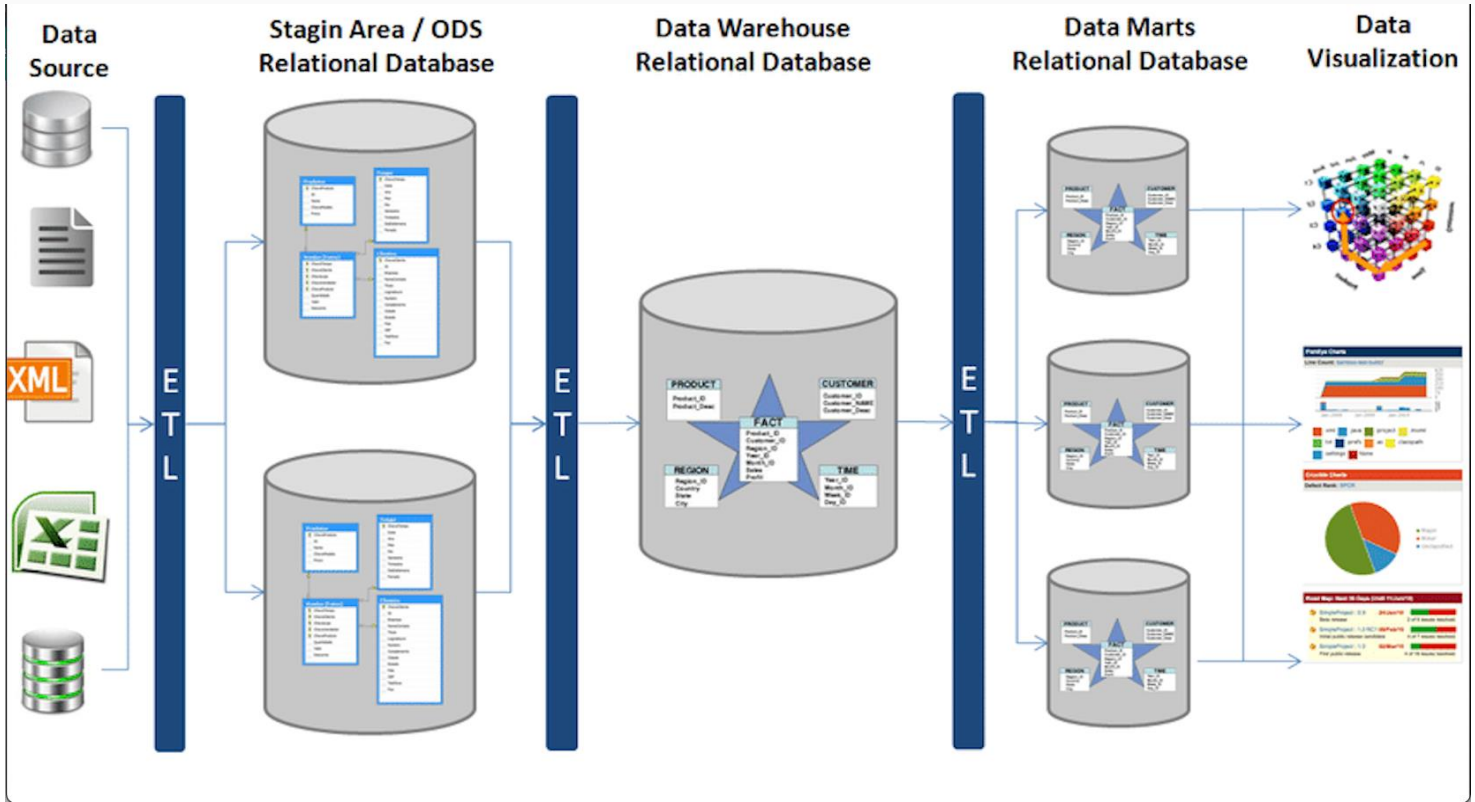
Ganhos:

1. **Consolidação de dados:** consolidação de dados de várias fontes diferentes **em um único local**, o que facilita a análise e a geração de relatórios.
2. **Análise de dados:** analisar grandes volumes de dados para identificar padrões, tendências e insights que podem ajudar a orientar as decisões de negócios.
3. **Suporte à decisão:** tomada decisões informadas e estratégicas com base em dados concretos.
4. **Melhoria da eficiência:** ajuda a melhorar a eficiência das operações de negócios e a reduzir o tempo gasto na análise manual de dados.

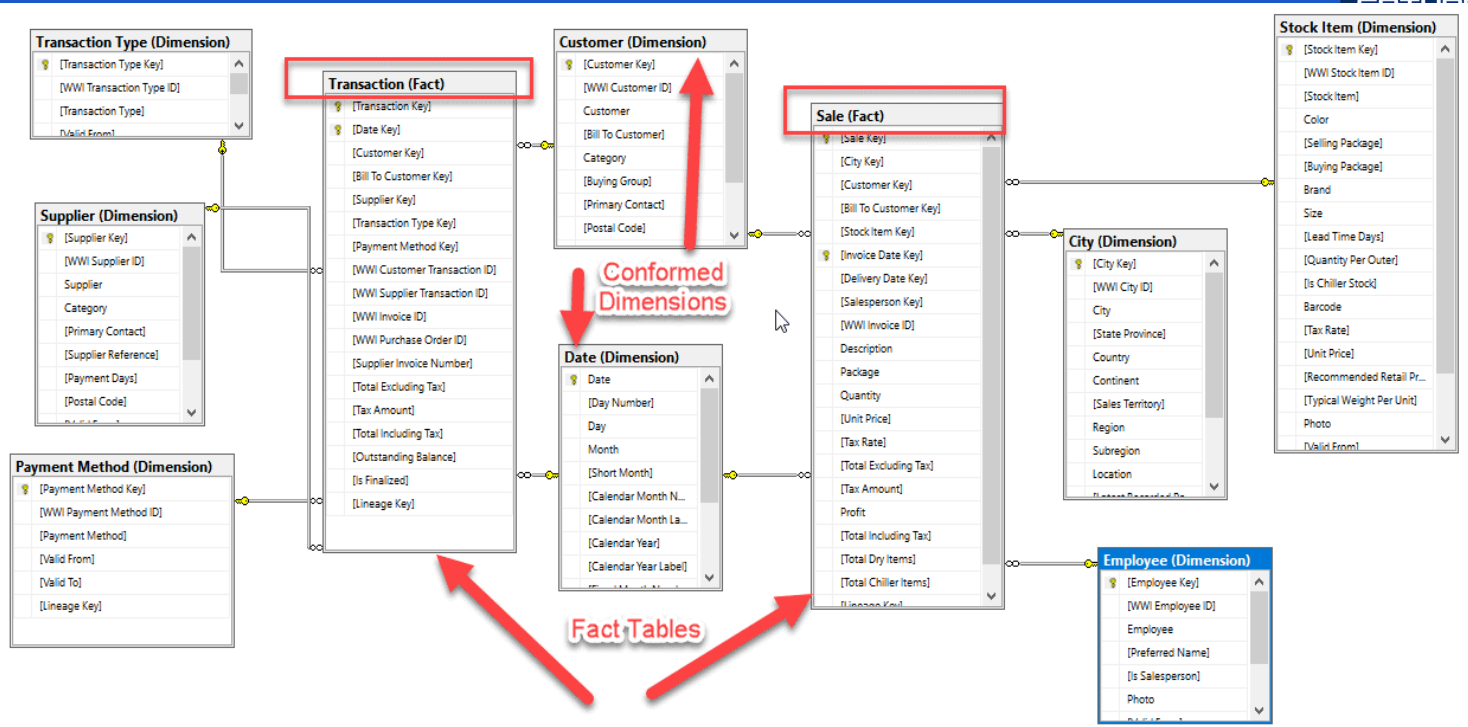
Perdas:

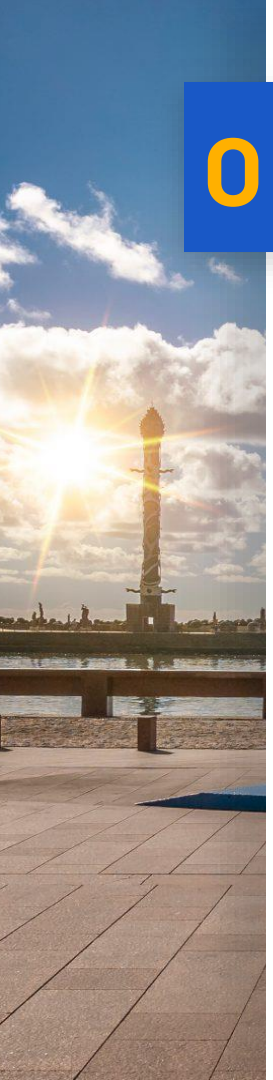
1. **Custo inicial:** pode ser cara, pois envolve a aquisição de hardware, software e a contratação de pessoal qualificado.
2. **Complexidade:** O design e a implementação de um data warehouse podem ser complexos, exigindo conhecimento técnico especializado.
3. **Integração de dados:** Integrar dados de várias fontes diferentes em um data warehouse pode ser desafiador e exigir esforço adicional para garantir a qualidade e a consistência dos dados.
4. **Manutenção contínua:** requer manutenção contínua para garantir que os dados sejam atualizados e precisos, o que pode exigir recursos adicionais.

Arquitetura Data warehouse



Star Schema


Transaction Type (Dimension)
[Transaction Type Key]
[WWI Transaction Type ID]
[Transaction Type]
Supplier (Dimension)
[Supplier Key]
[WWI Supplier ID]
Supplier
Category
[Primary Contact]
[Supplier Reference]
[Payment Days]
[Postal Code]
Payment Method (Dimension)
[Payment Method Key]
[WWI Payment Method ID]
[Payment Method]
[Valid From]
[Valid To]
[Lineage Key]
Transaction (Fact)
[Transaction Key]
[Date Key]
[Customer Key]
[Bill To Customer Key]
[Supplier Key]
[Transaction Type Key]
[Payment Method Key]
[WWI Customer Transaction ID]
[WWI Supplier Transaction ID]
[WWI Invoice ID]
[WWI Purchase Order ID]
[Supplier Invoice Number]
[Total Excluding Tax]
[Tax Amount]
[Total Including Tax]
[Outstanding Balance]
[Is Finalized]
[Lineage Key]
Customer (Dimension)
[Customer Key]
[WWI Customer ID]
Customer
[Bill To Customer]
Category
[Buying Group]
[Primary Contact]
[Postal Code]
Conformed Dimensions
Date (Dimension)
Date
[Day Number]
Day
Month
[Short Month]
[Calendar Month N...
[Calendar Month La...
[Calendar Year]
[Calendar Year Label]
Fact Tables
Sale (Fact)
[Sale Key]
[City Key]
[Customer Key]
[Bill To Customer Key]
[Stock Item Key]
[Invoice Date Key]
[Delivery Date Key]
[Salesperson Key]
[WWI Invoice ID]
Description
Package
Quantity
[Unit Price]
[Tax Rate]
[Total Excluding Tax]
[Tax Amount]
Profit
[Total Including Tax]
[Total Dry Items]
[Total Chiller Items]
City (Dimension)
[City Key]
[WWI City ID]
City
[State Province]
Country
Continent
[Sales Territory]
Region
Subregion
Location
Stock Item (Dimension)
[Stock Item Key]
[WWI Stock Item ID]
[Stock Item]
Color
[Selling Package]
[Buying Package]
Brand
Size
[Lead Time Days]
[Quantity Per Outer]
[Is Chiller Stock]
Barcode
[Tax Rate]
[Unit Price]
[Recommended Retail Pr...
[Typical Weight Per Unit]
Photo
[Valid From]
Employee (Dimension)
[Employee Key]
[WWI Employee ID]
Employee
[Preferred Name]
[Is Salesperson]
Photo



O que é Datalake

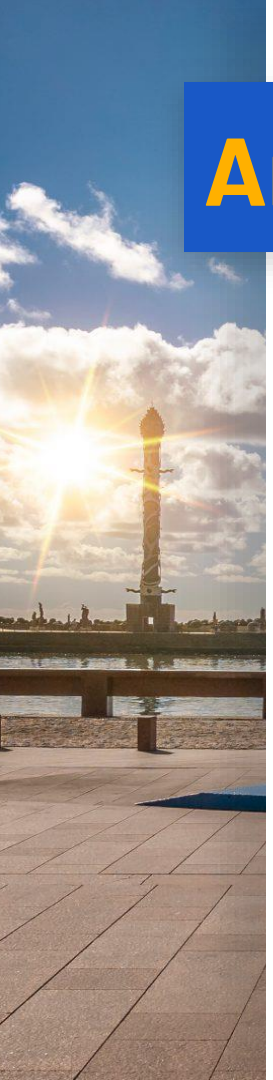
Um Data Lake é um repositório de armazenamento que detém uma vasta quantidade de dados brutos e não processados, em sua maioria em formatos como CSV, JSON, Parquet, entre outros. Ele difere dos bancos de dados tradicionais por não impor uma estrutura fixa aos dados, permitindo que sejam armazenados dados de diferentes formatos e origens sem a necessidade de pré-processamento.

Ganhos:

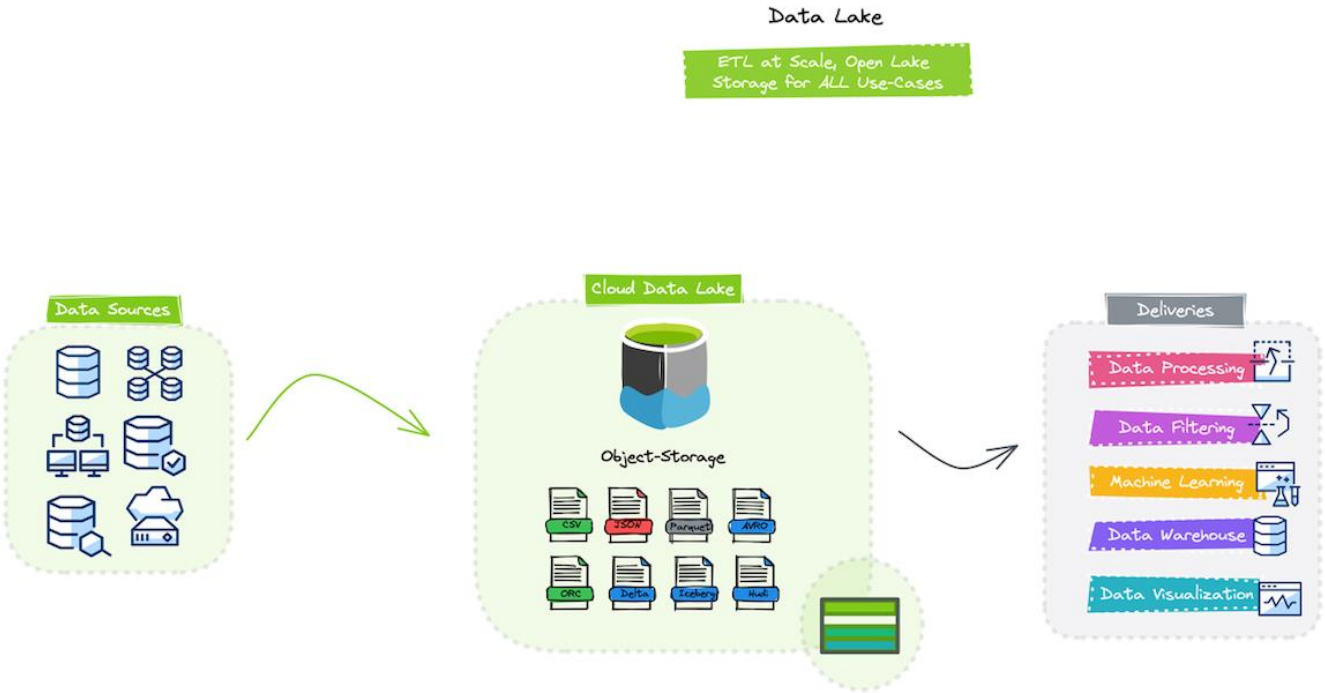
- **Armazenamento de Dados Semiestruturados e Não Estruturados:** O Data Lake permite armazenar uma grande variedade de dados em sua forma original, sem a necessidade de convertê-los para um formato específico.
- **Escalabilidade:** É possível aumentar a capacidade de armazenamento do Data Lake conforme a necessidade, sem interromper as operações.
- **Análise Avançada de Dados:** Com o uso de ferramentas adequadas, é possível realizar análises avançadas, como análise preditiva e machine learning, nos dados armazenados no Data Lake.

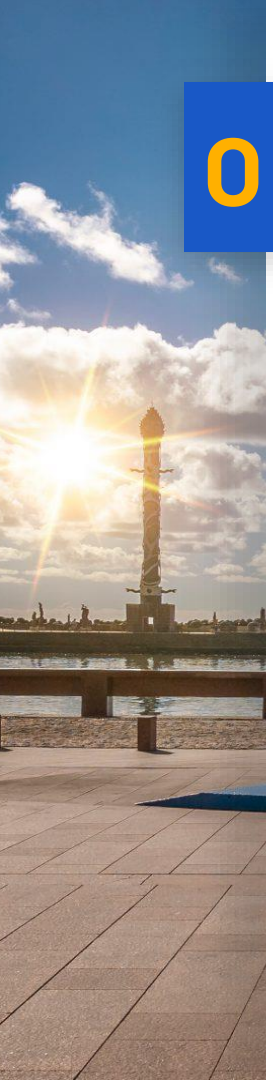
Perdas:

- **Complexidade:** O gerenciamento de um Data Lake pode ser complexo devido à diversidade de dados e à necessidade de garantir a qualidade e a segurança dos mesmos.
- **Custo:** A manutenção de um Data Lake pode ser custosa, especialmente ao lidar com grandes volumes de dados.
- **Segurança:** Como o Data Lake armazena uma grande quantidade de dados sensíveis, é necessário implementar medidas rigorosas de segurança para proteger esses dados contra acessos não autorizados.



Arquitetura Datalake





O que é Lakehouse

Lakehouse é um conceito que combina características de datalake e data warehouses, proporcionando um ambiente unificado para armazenamento e análise de dados. Aqui estão algumas características principais, bem como ganhos e perdas associados ao conceito:

Características do Lakehouse:

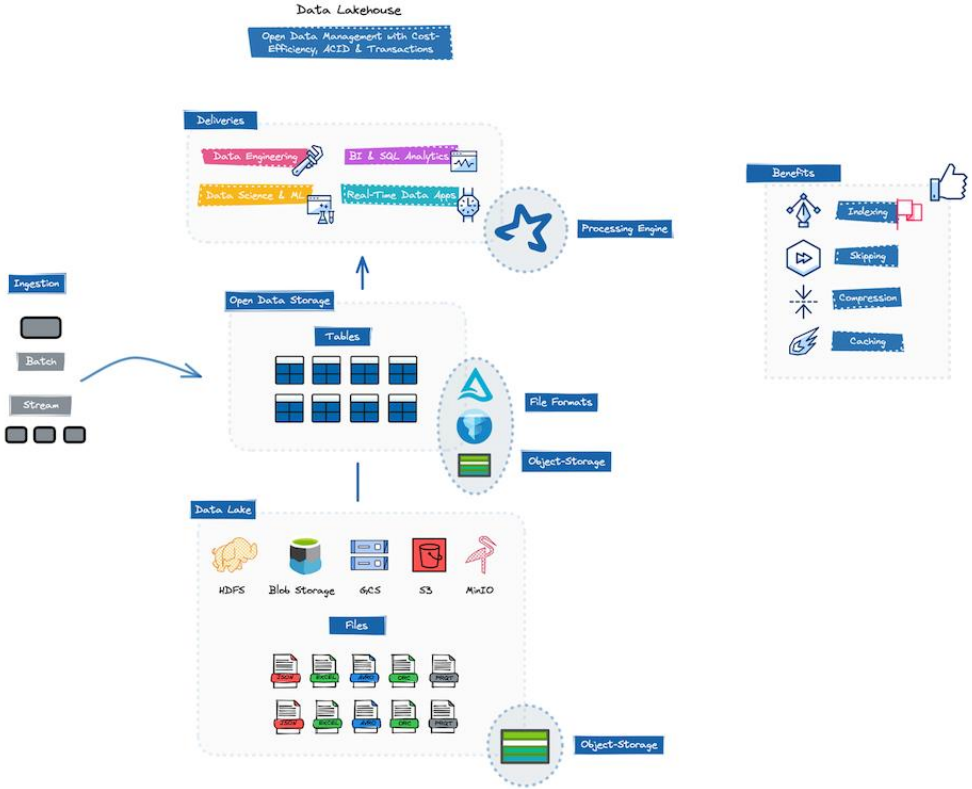
1. **Armazenamento de Dados:** O Lakehouse permite armazenar dados em seu formato original (como em um data lake), o que facilita a ingestão de dados brutos e semiestruturados.

2. **Processamento Analítico:** Oferece capacidades de processamento analítico avançado, semelhante a um data warehouse, permitindo consultas complexas e otimizadas.

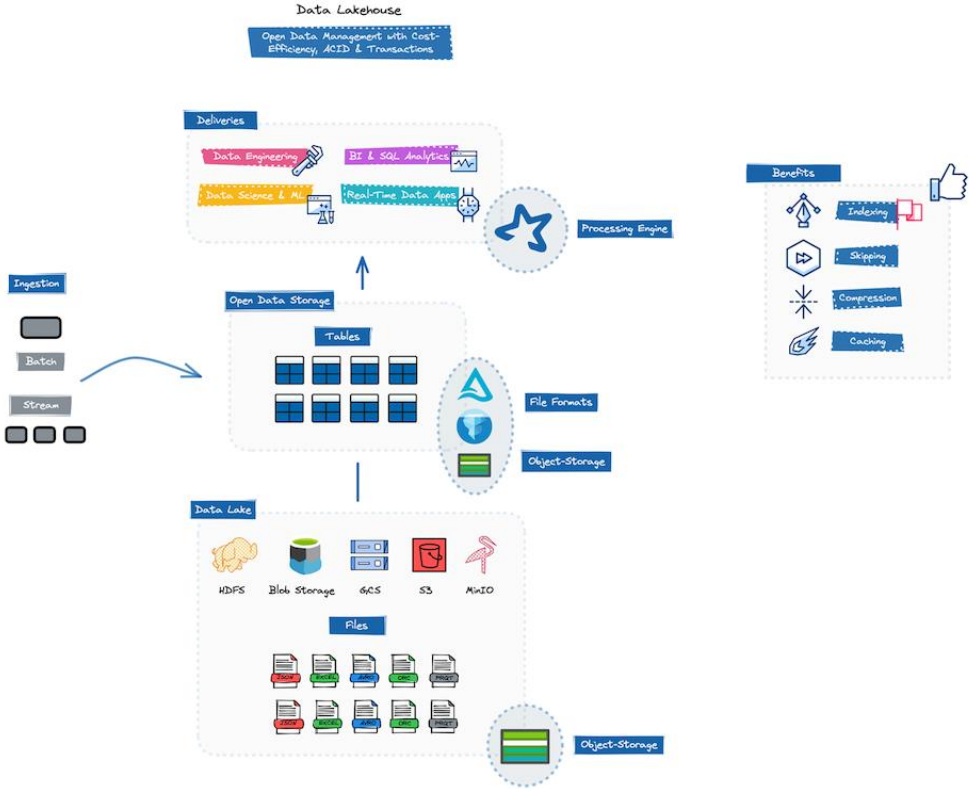
3. **Metadados e Controle de Dados:** Oferece metadados integrados e controle de dados para gerenciar a qualidade, segurança e conformidade dos dados.

4. **Unified Analytics:** Oferece suporte para diferentes tipos de análises, incluindo SQL, machine learning e processamento de gráficos, em um único ambiente.

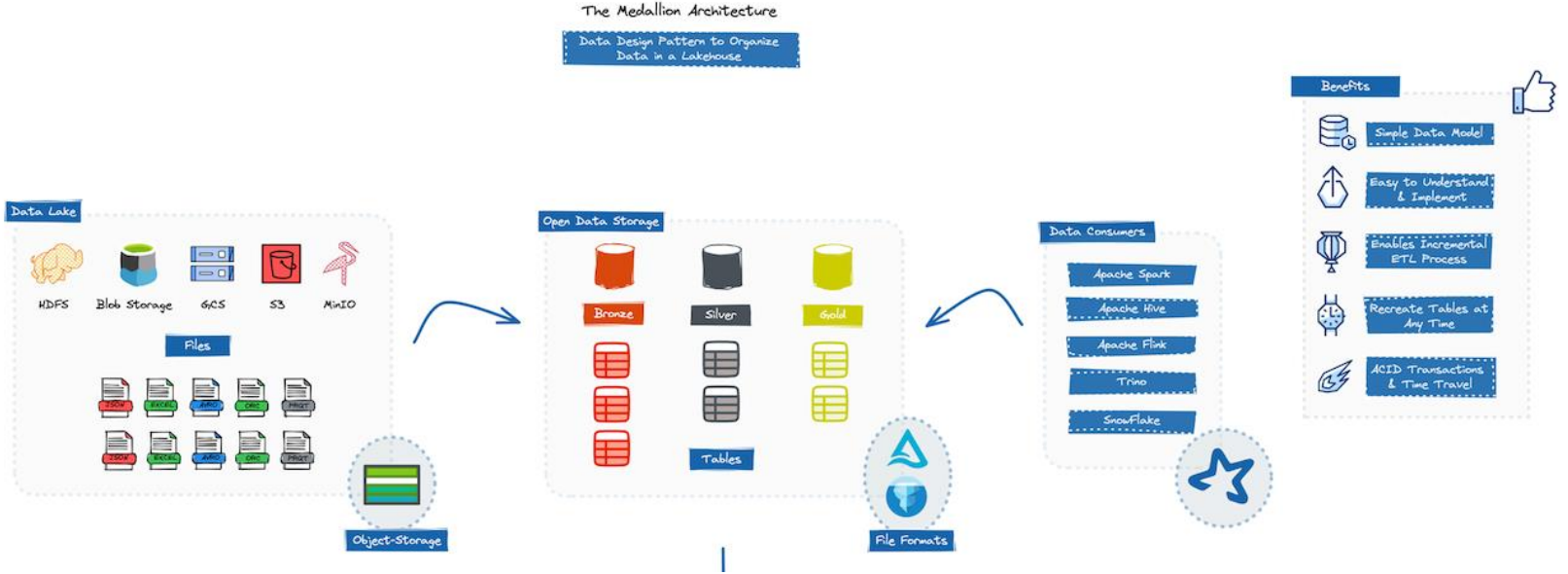
Arquitetura Lakehouse



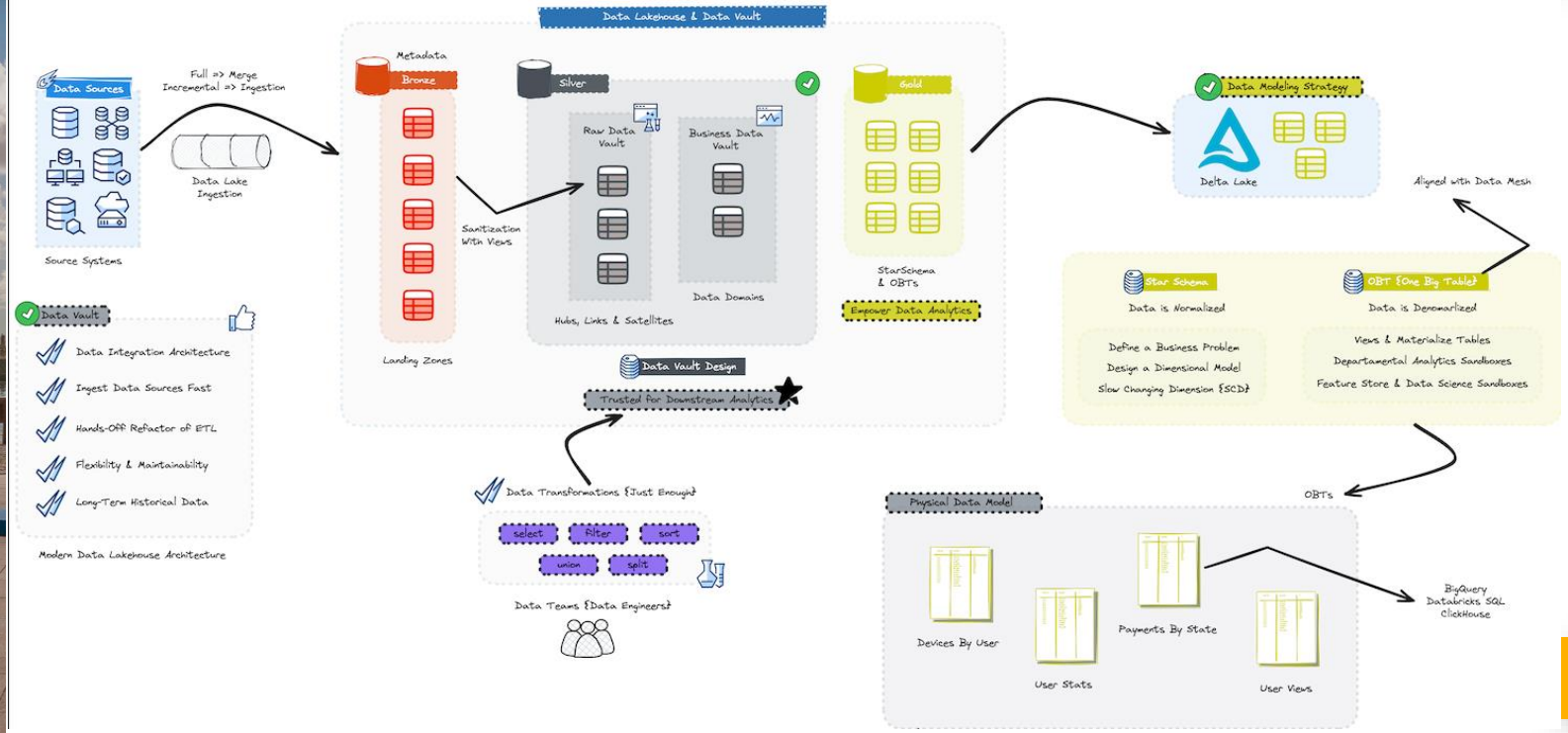
Arquitetura Lakehouse

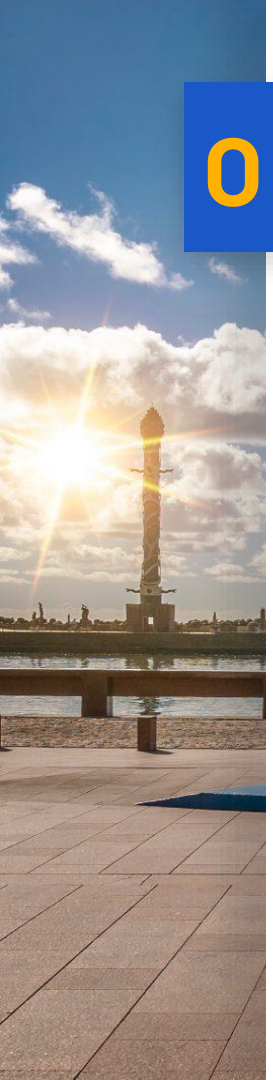


Arquitetura Medalhão



Arquitetura Data Vault





O que é Data Vault

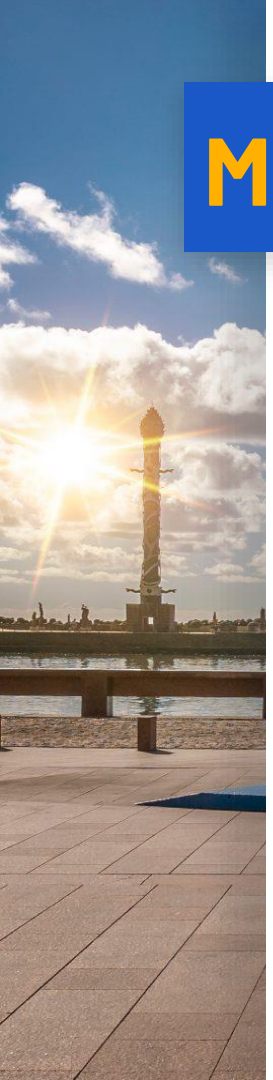
Data Vault é uma metodologia de modelagem de dados usada principalmente em data warehousing. Foi desenvolvida por Dan Linstedt e é projetada para fornecer uma abordagem flexível, escalável e resiliente para armazenar dados brutos de uma variedade de fontes, tornando-os disponíveis para análise de negócios.

A ideia fundamental por trás do Data Vault é dividir os dados em três tipos principais de tabelas:

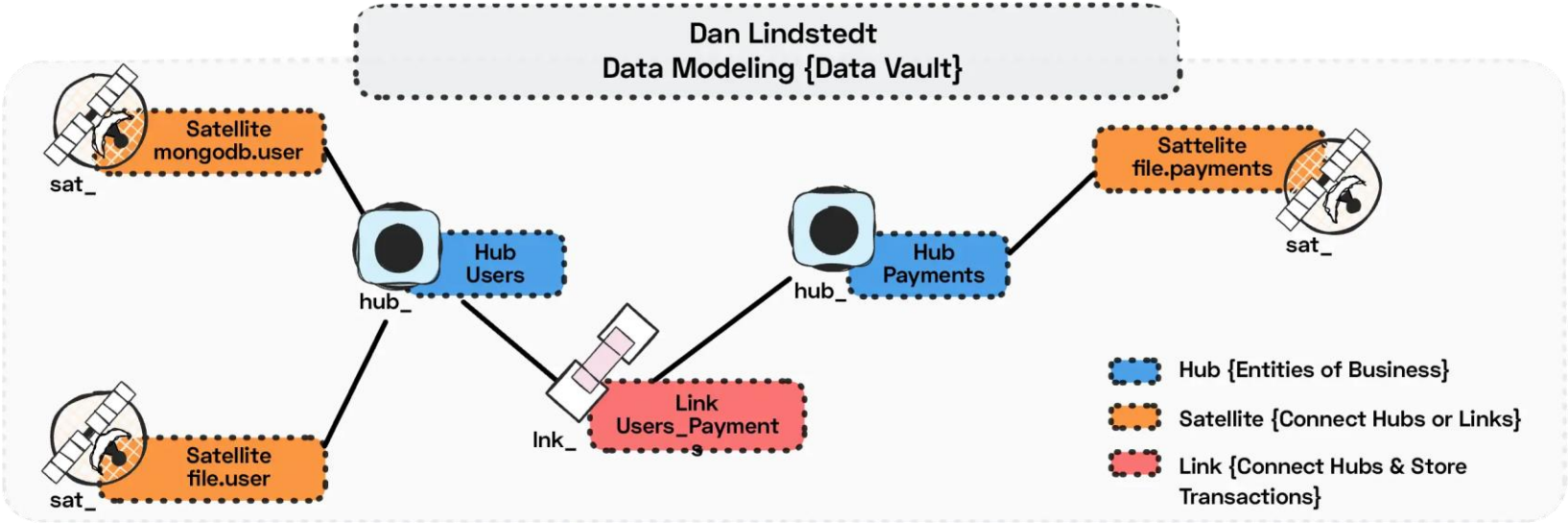
1.Hubs: representam entidades de negócios centrais, como clientes, produtos ou transações. Cada hub é uma tabela que contém uma chave de negócios única para cada registro.

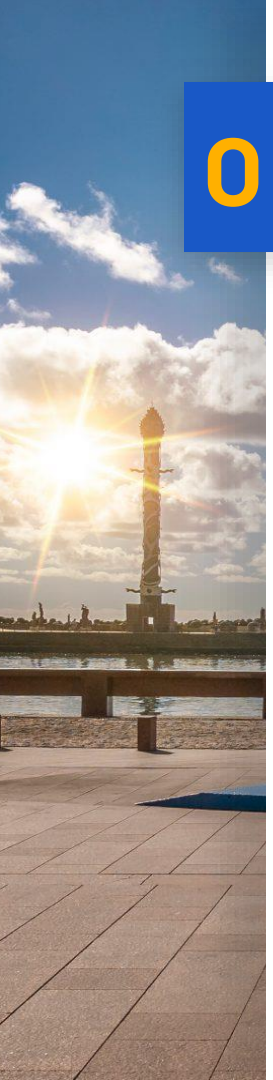
2.Links: são tabelas que conectam hubs e descrevem relacionamentos entre eles.

3.Satélites: contêm atributos de hub, como datas, status, descrições, etc. Cada satélite está associado a um hub ou link e contém informações adicionais sobre o hub/link.



Modelo Data Vault





O que é Data Mesh

Data Mesh é uma abordagem arquitetônica para gerenciar dados em organizações distribuídas e complexas. A ideia é tratar os dados como um produto e aplicar princípios de arquitetura de microsserviços para criar uma infraestrutura de dados distribuída e escalável.

Ganhos:

1.Descentralização: Ao adotar o Data Mesh, as organizações podem **descentralizar** a propriedade e a governança dos dados. Isso significa que os times que produzem os dados são responsáveis por sua qualidade e governança, facilitando a escalabilidade.

2.Agilidade: Com a abordagem de microsserviços, os times podem trabalhar de forma independente, o que aumenta a agilidade na entrega de soluções e permite uma resposta mais rápida às mudanças de requisitos.

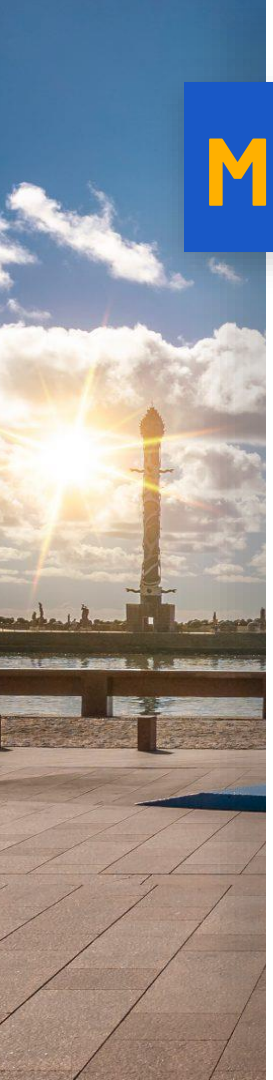
3.Escalabilidade: A arquitetura distribuída do Data Mesh permite escalar de forma mais eficiente, uma vez que cada componente pode ser escalado independentemente.

Perdas:

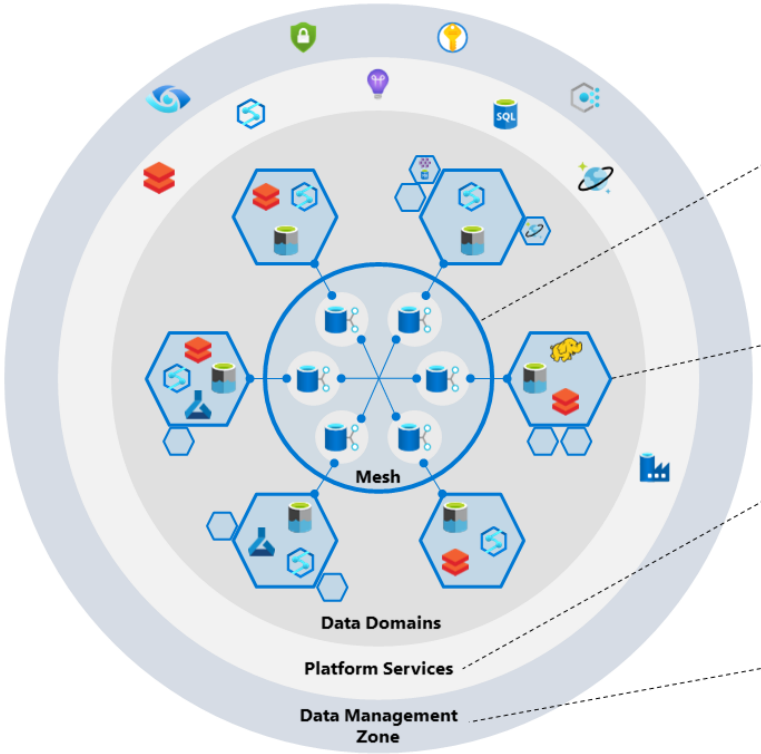
1.Complexidade inicial: Implementar o Data Mesh pode ser complexo no início, pois exige mudanças na cultura organizacional, na forma como os times trabalham e na infraestrutura de dados existente.

2.Governança distribuída: Distribuir a governança dos dados pode ser desafiador e requer uma coordenação eficaz entre os times para garantir a consistência e a qualidade dos dados.

3.Custo inicial: A transição para o Data Mesh pode exigir investimentos significativos em termos de tempo e recursos para redesenhar a arquitetura de dados e capacitar os times.



Modelo Data Mesh



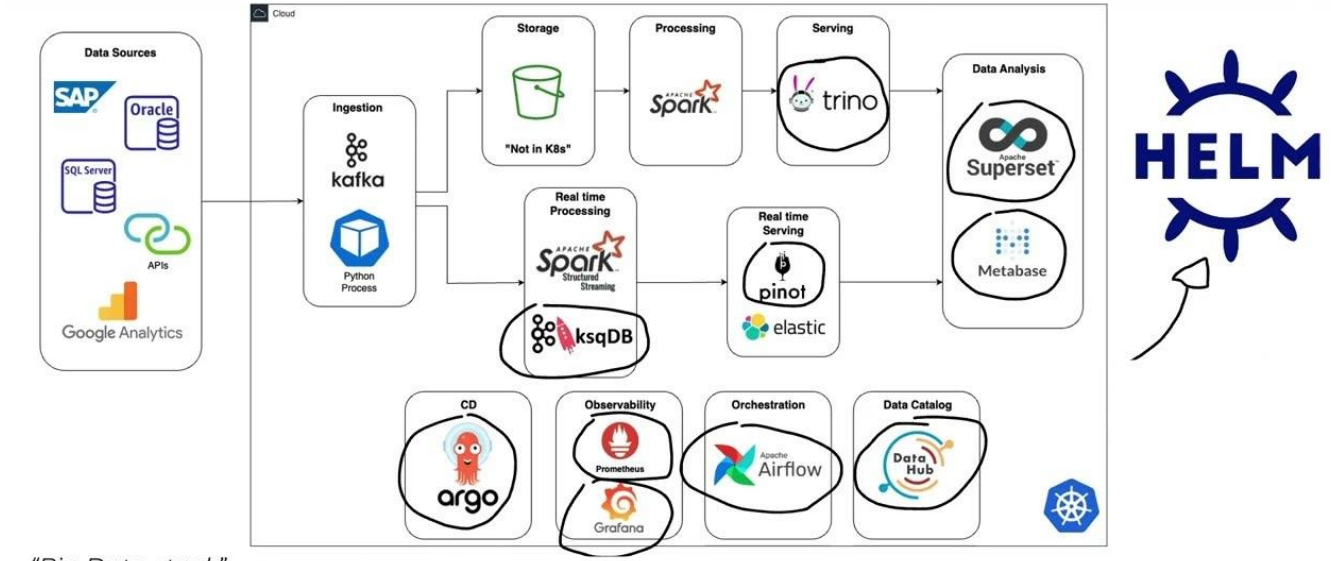
The **data mesh** intelligently distributes data products between data domains. Read data stores share compute resources. This reduces costs, solves interoperability concerns, and better addresses time-variant and non-volatile concerns of large data consumers.

Data domains operate their own applications or analytics platforms, whilst adhering to common policies and standards.

The central **platform services** defines blueprints that encompass baseline security, policies, capabilities, and standards.

A key concept for every enterprise-scale analytics and AI implementation is having one **data management zone**. This subscription, which is required for data management, contains resources that'll be shared across all landing zones.

Futuro da Engenharia de Dados



"Big Data stack"

Futuro da Engenharia de Dados

Mídias

icaromacedo

icaromcosta

http://www.4bi.com

OBRIGADO PELA PRESENÇA!