



The Floow

Data Products and Data Meshing

Date: 16 March 2022

Prepared for: General use

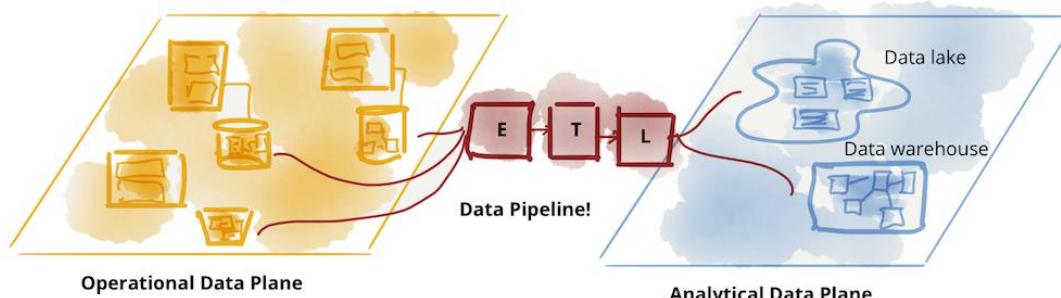


Data Mesh is a new architecture paradigm shift.

- It's a new way of thinking how to do analytics within organizations
- It's not a new piece of technology
- It's not a new software you going to run out and buy
- It's not a technology centric approach
- no single technology is the perfect fit to build a Data Mesh
- It's a total rethink of who in charge of what, when it comes in who maintains the data within an company

A definition:

An intentionally designed distributed data architecture, under centralized governance and standardization for interoperability, enabled by a shared and harmonized self-serve data infrastructure

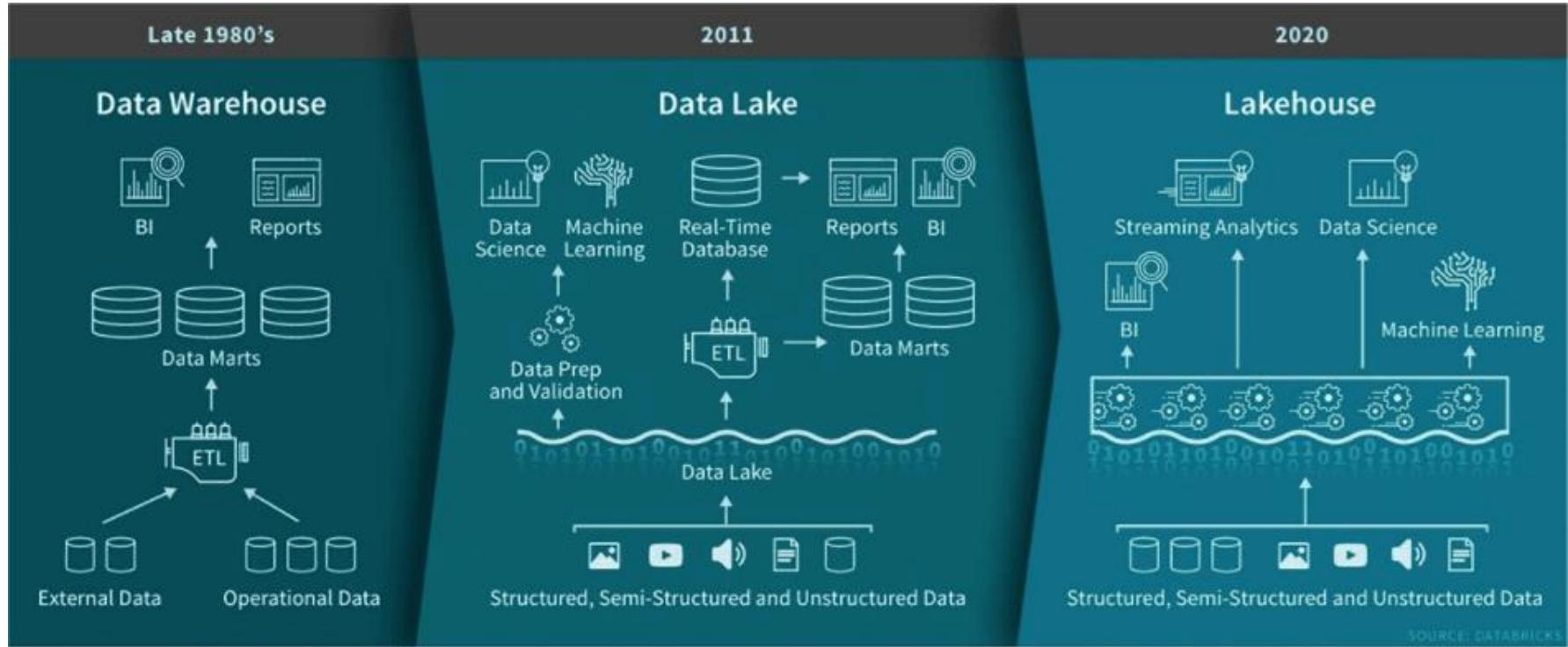


- Running the Business
- Capabilities & Services
- Operation Data
- Products & Applications
- Communication
- Infrastructure

Running the business
Serving the users

Optimizing the business
Augmenting the user experience
with intelligence

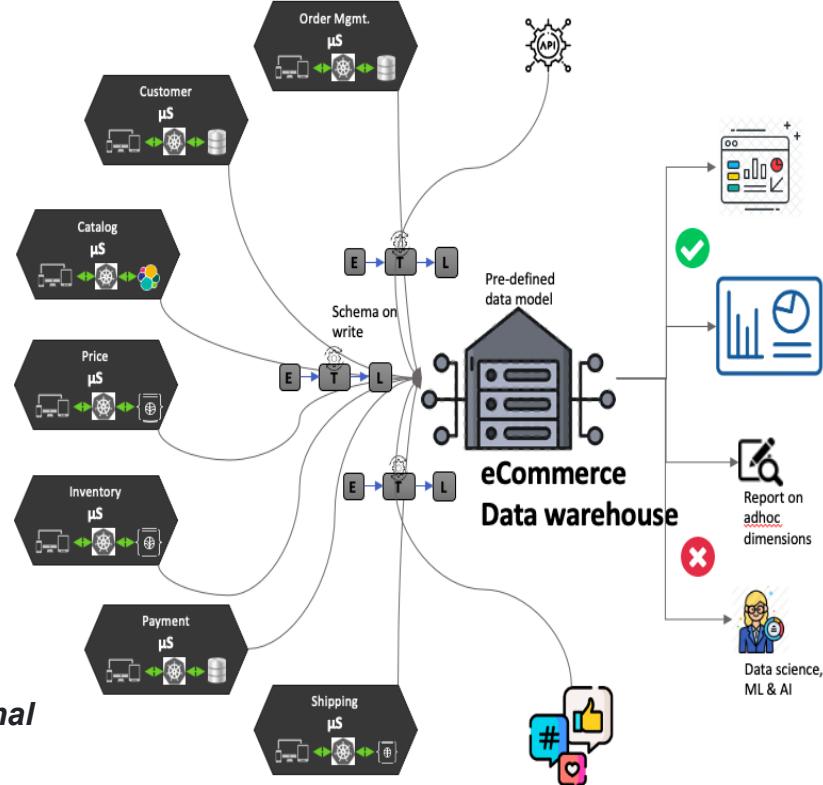
- Optimising the Business
- Insights
- Analytical Data
- Report | ML Models
- Data Transfer & Copying
- Data Pipelines & Storage Infra



Data Warehouse Limitations



- Data is aggregated so visibility into the lowest levels is lost
- Only pre-determined questions can be answered
- Modeling a single unified data model is impractical for complex domains.
- Siloed data engineers with no domain expertise
- A complex labyrinth of data pipelines with constant operational challenges for centralized data teams

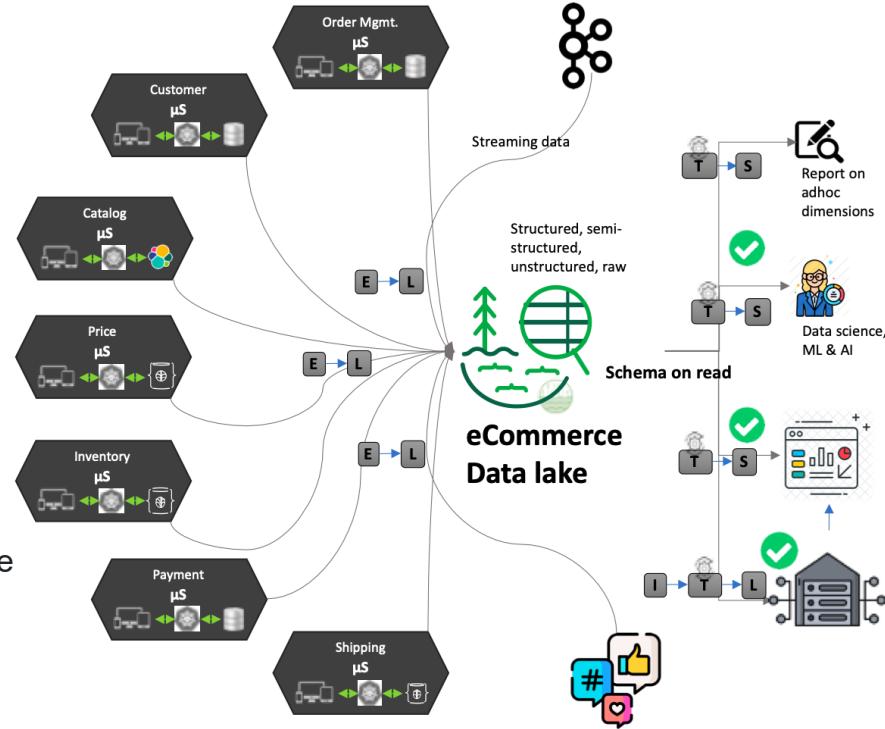


* **Limitations for using relational database in general on operational plane**

Data Lakes and Lakehouse Limitations



- Complex data pipelines
- Data quality: quickly morph into data swamps
- Continuous data management
- Limited data governance and security
- Query performance deteriorates
- Metadata management
- Above all, data is centralized leading to a monolithic data lake

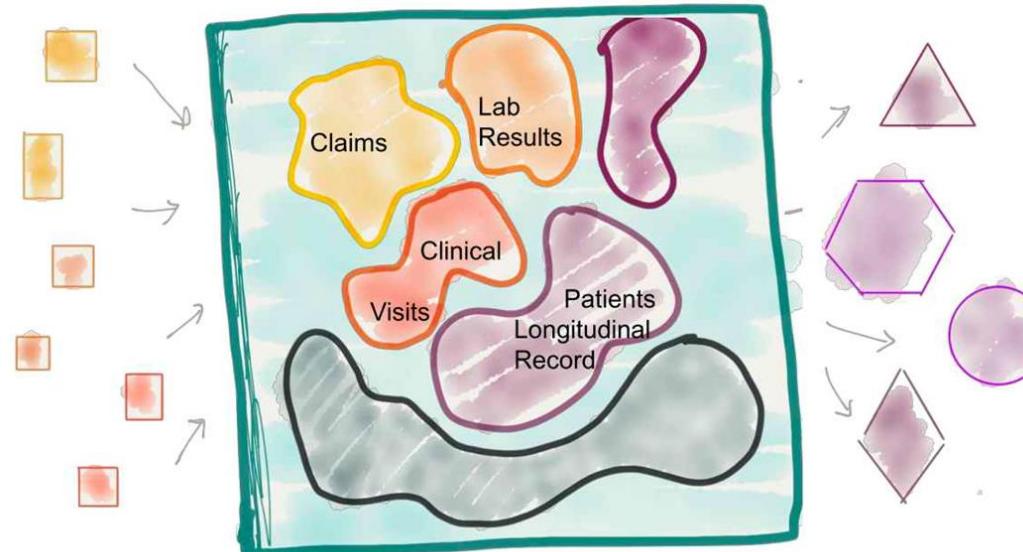


*Lakehouse does solve some of the problems but not centralized data platform

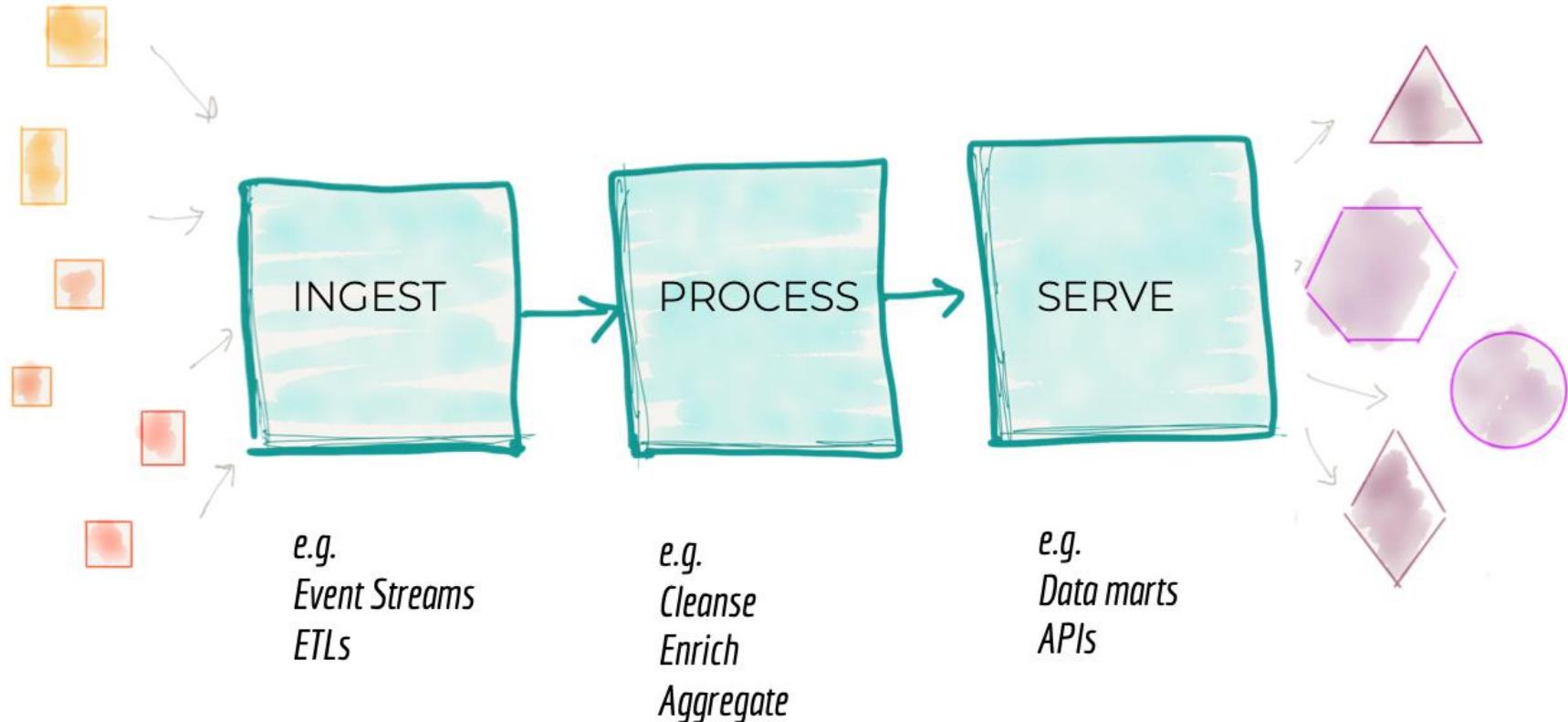




MONOLITHIC No Domain Boundaries

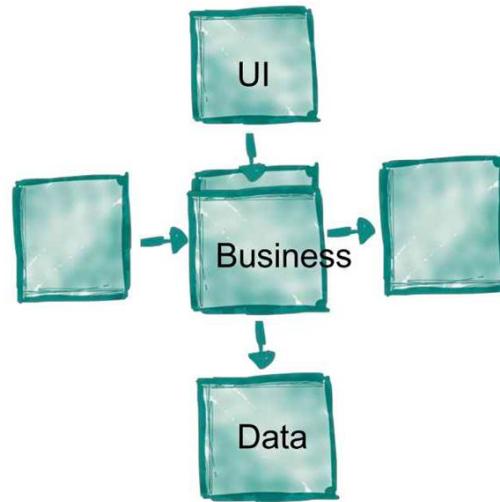


Big Data Platform Pipeline



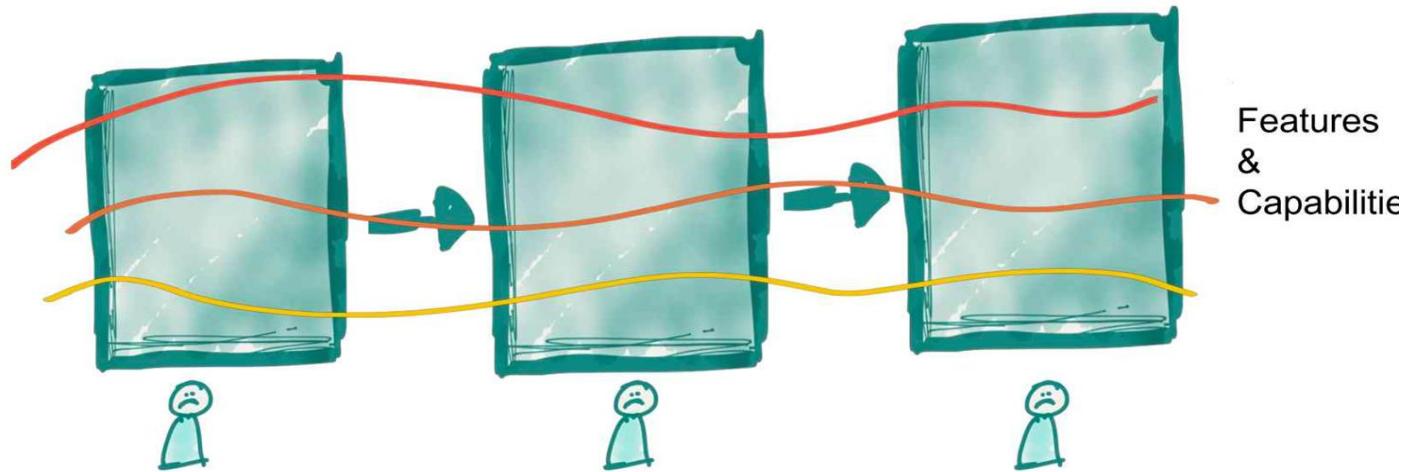


LAYERED Top Level Technical Partitioning



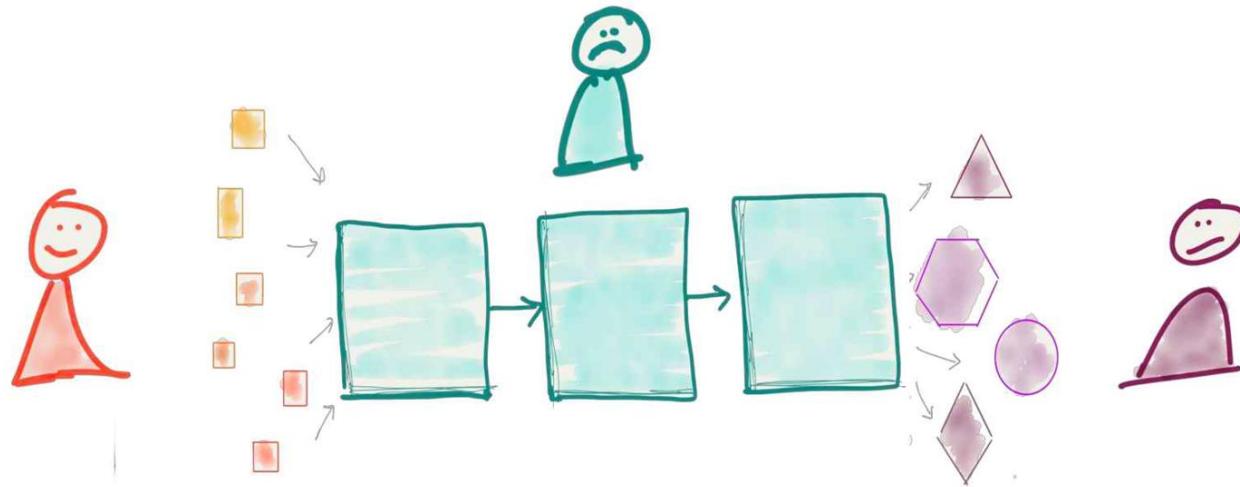


Orthogonal to the Axis of Change!





Data Platform Engineers



Domains' Operational systems Teams

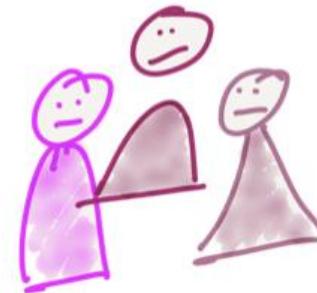
Data Scientists, BI teams, ...



Cross-functional
Domain oriented source teams



Hyper-specialized
Data & ML Platform Engineers

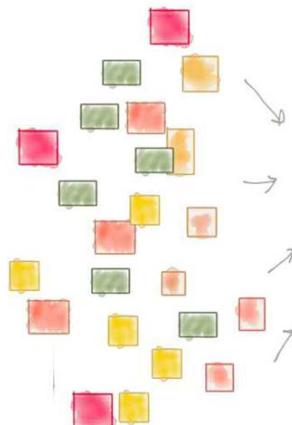


Cross-functional
Domain oriented consumer teams



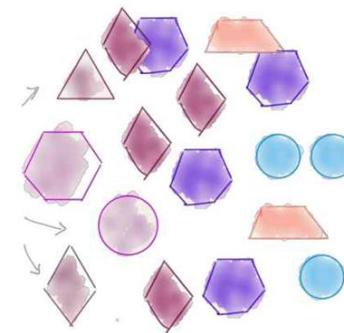


UBIQUITOUS DATA



SOURCE PROLIFERATION

INNOVATION AGENDA



CONSUMER PROLIFERATION



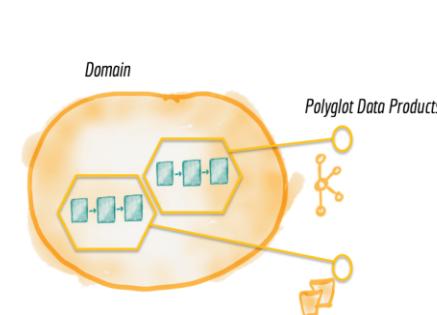
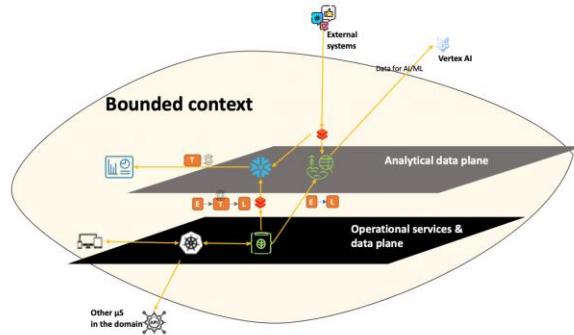
A data mesh tries to solve three challenges that come with using a centralized data platform:

- **Organizational scaling:** The central team or the one that runs the data lake or warehouse becomes the bottleneck.
- **Lack of ownership:** Lets users know who owns the data—the source team or the infrastructure team.
- **Lack of quality:** While the infrastructure team is responsible for maintaining the data's quality, it may not know the data as well as the source team.

What is Data Mesh? – Paradigm Shift



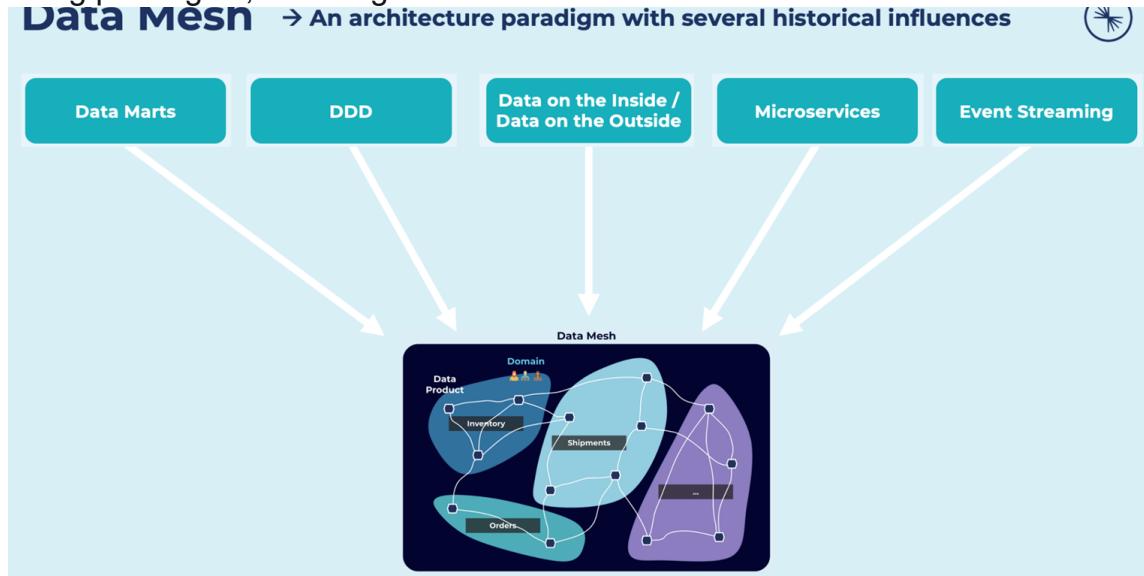
- Focuses more on business domains, rather than technical implementation
- Results in domain-oriented, decentralized data ownership and architecture
- Allows each bounded context to own and make available its domain data as “data product”
- Leaves the choice of storing raw or aggregated data to be determined by individual bounded contexts based on users
- Pushes data processing pipelines into each bounded context who own and understand the data



Data Mesh - An Architecture Paradigm



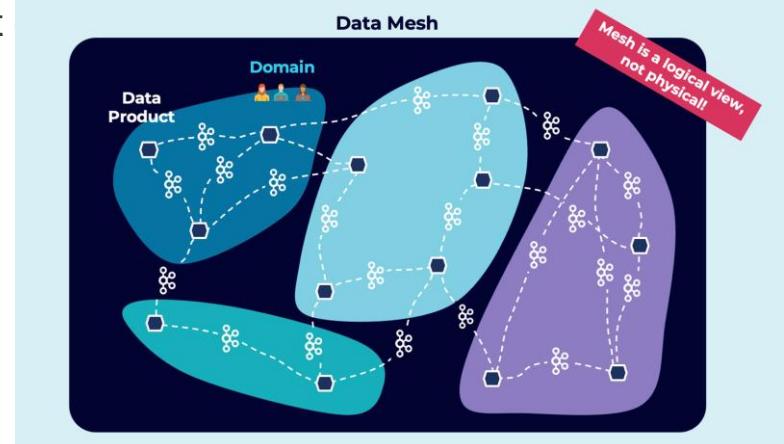
- Is an implementation pattern (not unlike microservices or domain-driven design) but applied to data.
- Combines existing paradigms, including **Domain-driven + Microservices + Event Streaming**



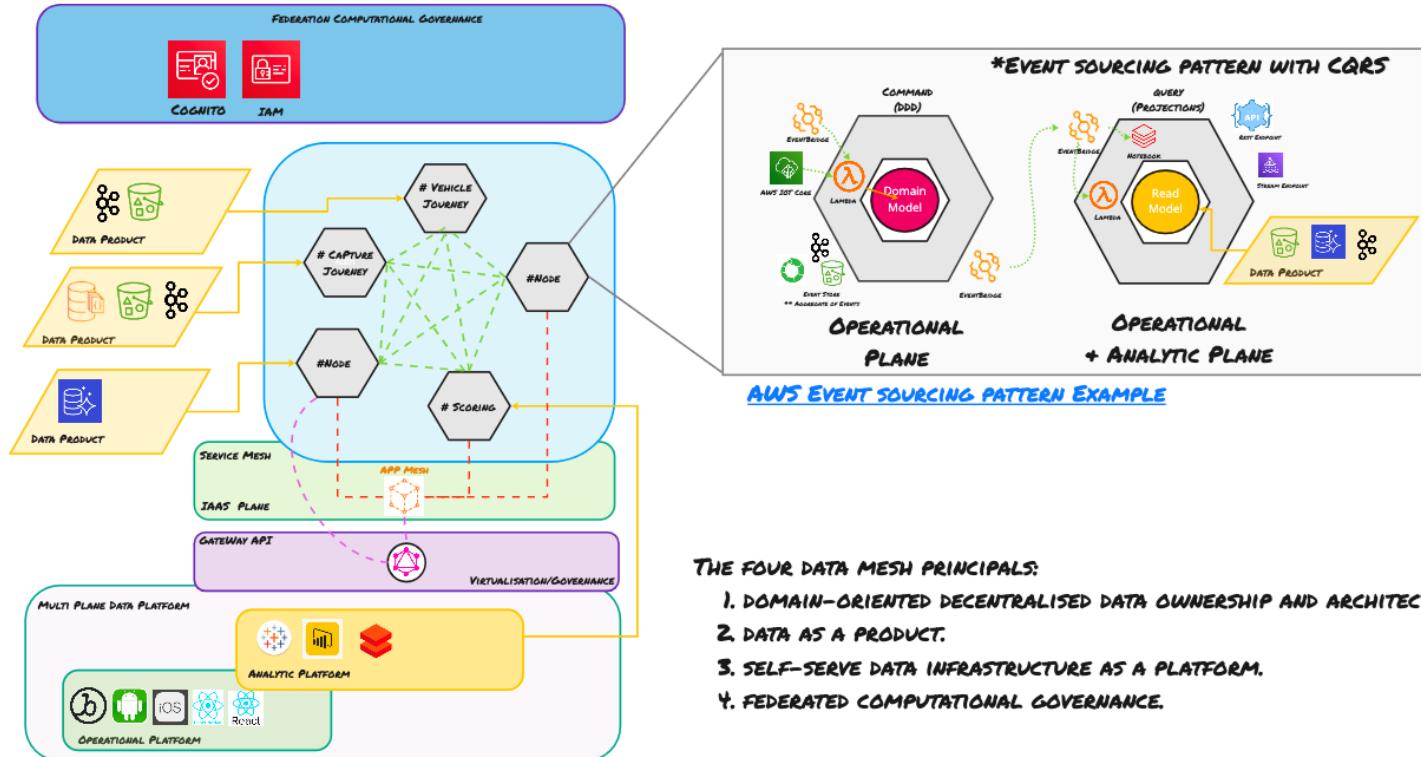
Key features of the data mesh architecture?



- All data is a domain that can be accessed by anyone who has access to it.
- Decentralized ownership/ data management.
- **Data as a product**, not as a by-product.
- Each domain is responsible for its data, its data quality, and its security.
- Domains don't influence each other, each domain has its own resources.
- Domains are owned by those who know data best. It ownership.
- Mesh is a logical view, NOT Physical



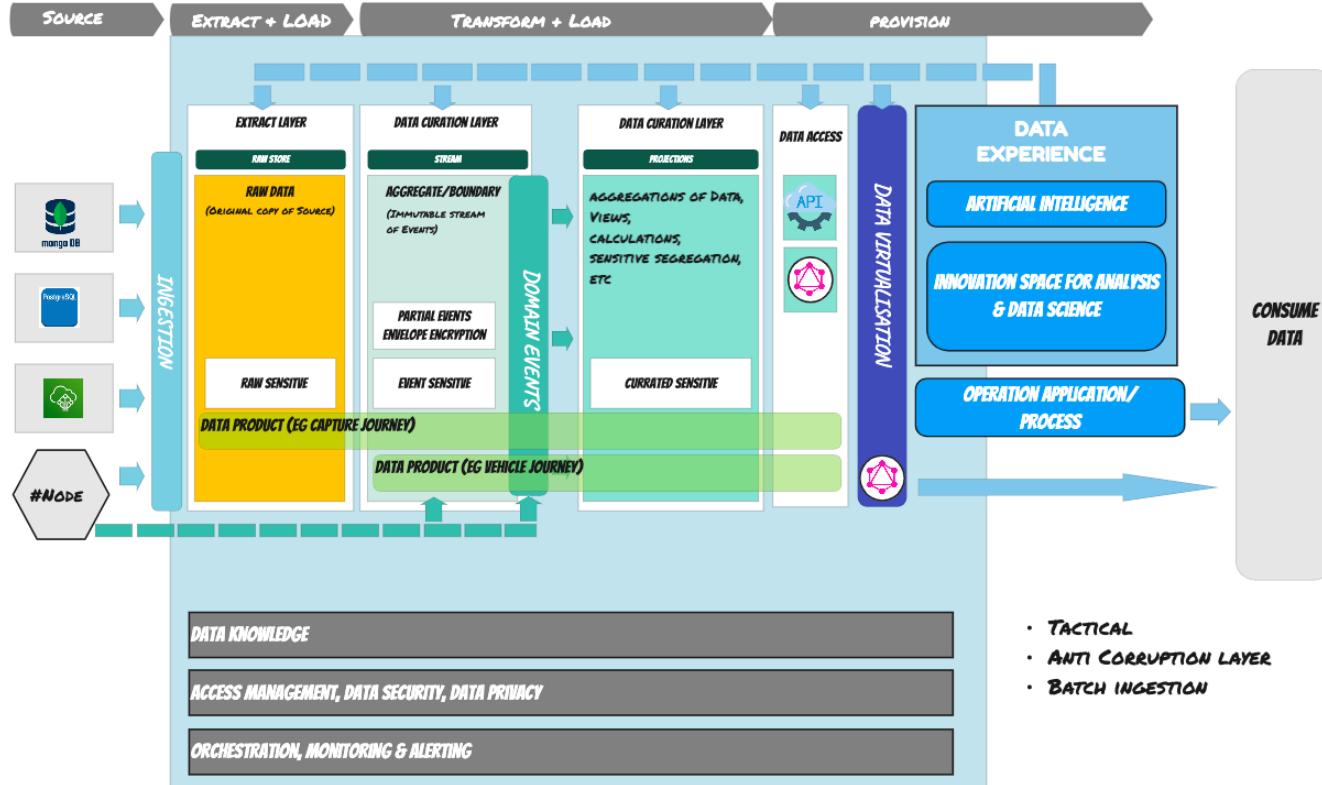
Data Mesh Holistic Architecture



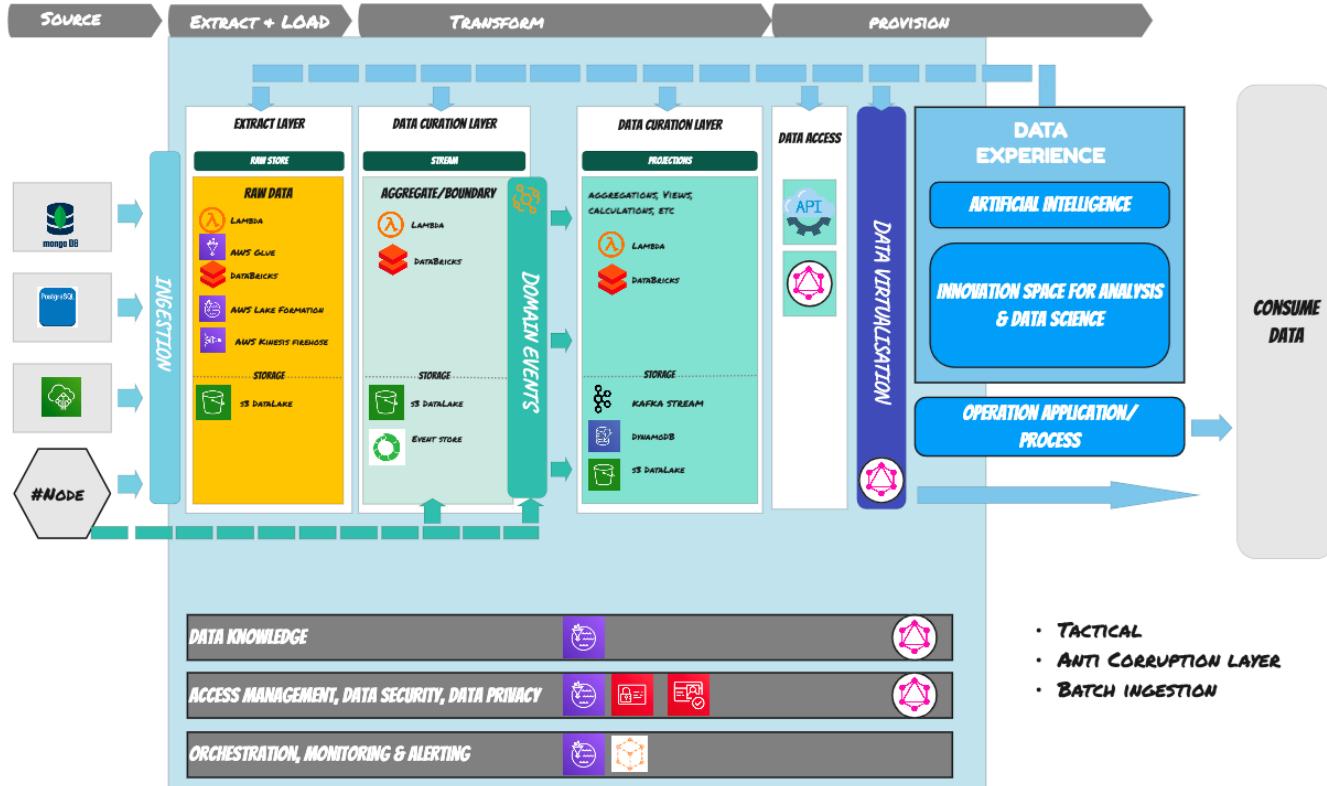
THE FOUR DATA MESH PRINCIPALS:

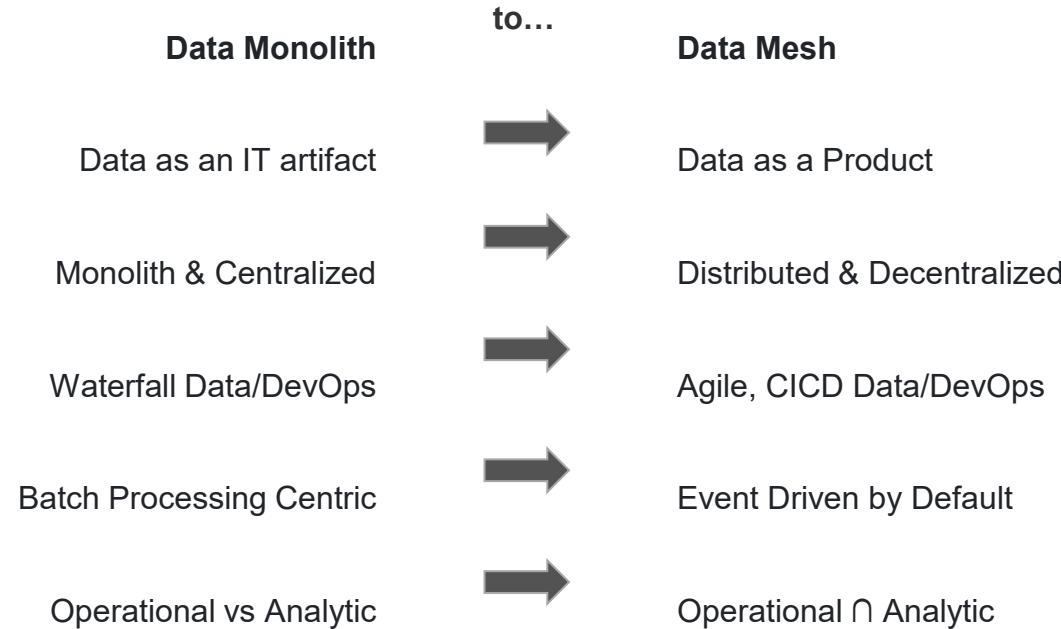
1. DOMAIN-ORIENTED DECENTRALISED DATA OWNERSHIP AND ARCHITECTURE.
2. DATA AS A PRODUCT.
3. SELF-SERVE DATA INFRASTRUCTURE AS A PLATFORM.
4. FEDERATED COMPUTATIONAL GOVERNANCE.

Data Mesh - ELT-CQRS



Data Mesh - ELT-CQRS



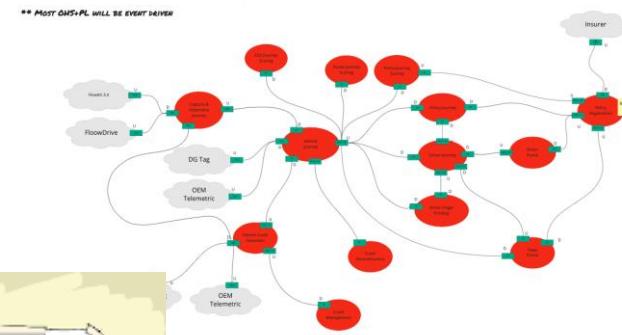
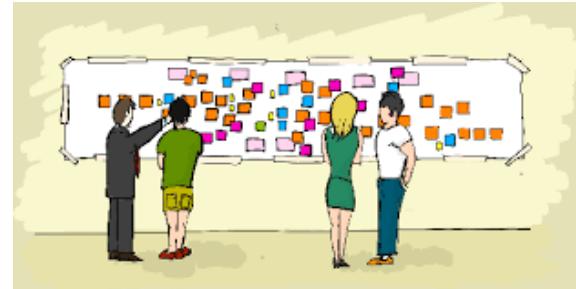
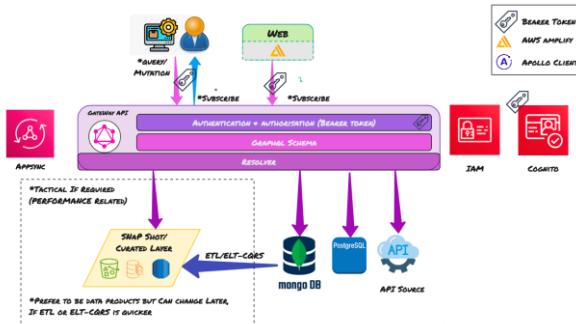




- Teams should be aligned against data products and not tech
 - Establish data products, by identity bounded context and their domains and subdomains
 - Event Storming Workshops
 - Context Map, DDD Models
 - Migrate from Monolith to Microservice
 - Using “The Strangler Pattern”
 - Look at moving towards an Event Driven Architecture

** MOST GMFS+PL WILL BE EVENT DRIVEN







The Data Mesh concept could be relevant for global deployments and go across **data centres** and **multiple clouds**.

An example spanning a streaming Data Mesh across multiple cloud providers like AWS, Azure, GCP, or Alibaba, and on-premise / edge sites:

