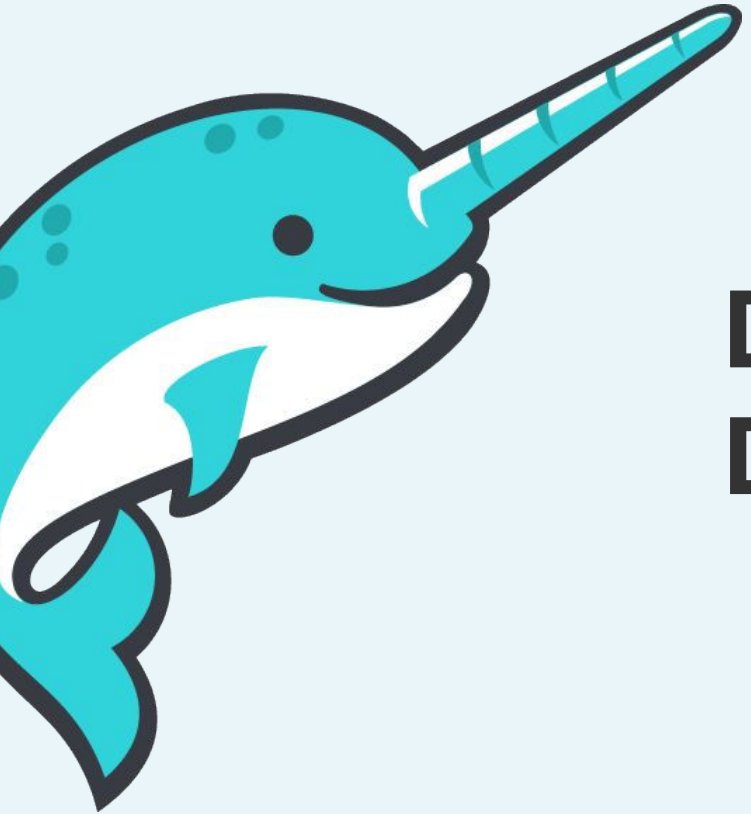


Data Mesh & Open Data Lakehouse

dremio
The Data Lake Engine

How orgs makes decisions today?

- move operational data to a centralized repo using ETL/ELT process

- needs specialized data engineers expertise; overload on centralized team

- wait time on data consumers

Centralized Approaches to Data Can Cause Friction

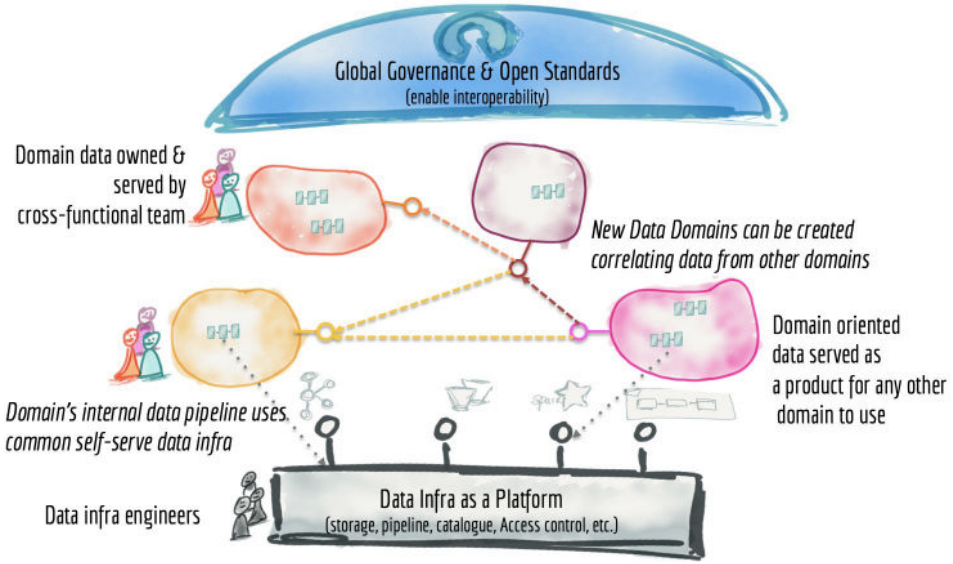
Data engineers
- Spend too much time managing ETL/ELT flows
- Constantly need to communicate with analysts to ensure they don't lose business context when preparing data

Data consumers
- Lack a self-service experience to analyze data
- Rely on data engineers for every task (experimenting, amending, sharing data)

But a lot of the problems are ORGANIZATIONAL!

Data Mesh: A Decentralized Solution

Data engineers
Now focused on developing a self-service infra platform



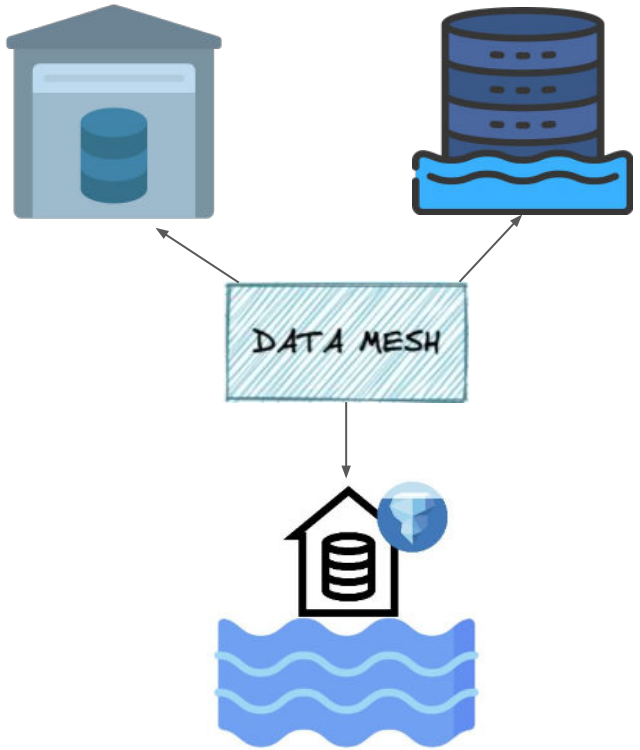
Business units/domains
Now responsible for creating/maintaining its datasets, treating data like a first-class product

Data consumers
Now have faster & easier access to data

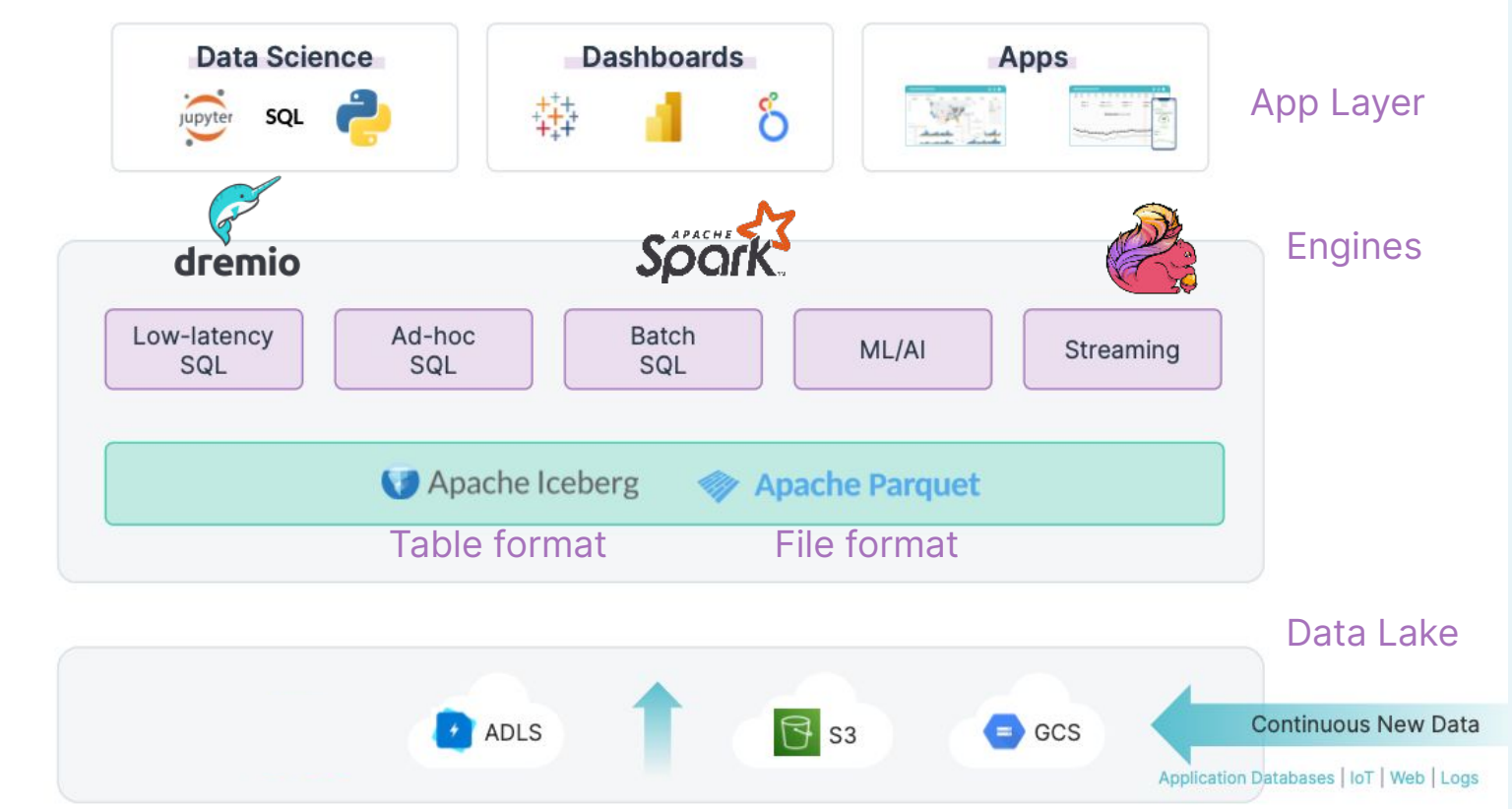
Domain Ownership	**Data as a Product**
Self-service Data Platform	**Federated Computational Governance**

Data Mesh for Data Architectures

- A data mesh strategy can technically be applied to any form of data architecture platform. E.g. warehouse, data lake, etc.
- Gives control to the team who knows that domain's data the best
- Brings a data product mindset (create, share data products with governance)
- Streamlines the central infrastructure team's effort
- Open Lakehouse - the **"right"** fit

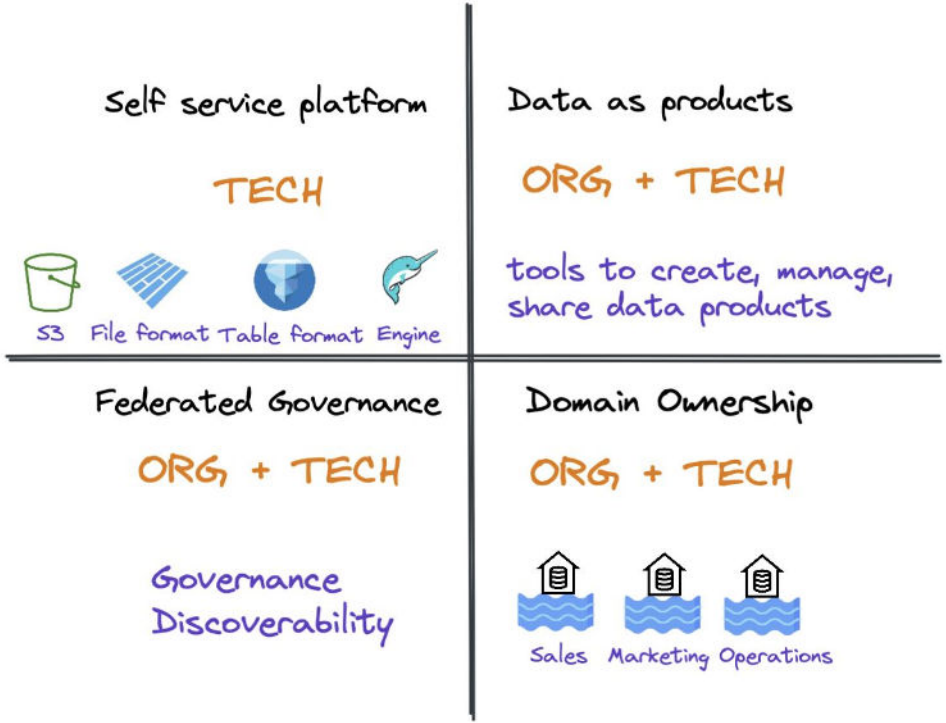


Introducing Open Data Lakehouse



Data Mesh, Open Lakehouse : complementing each other

- organizational problems needs the right technical approach & vice versa
- Both are designed to deal with the **scalability** problem from 2 different angles
- We need the right set of tools/infra to create, share & manage data products
- Aligning organizational & technological components with the 4 pillars

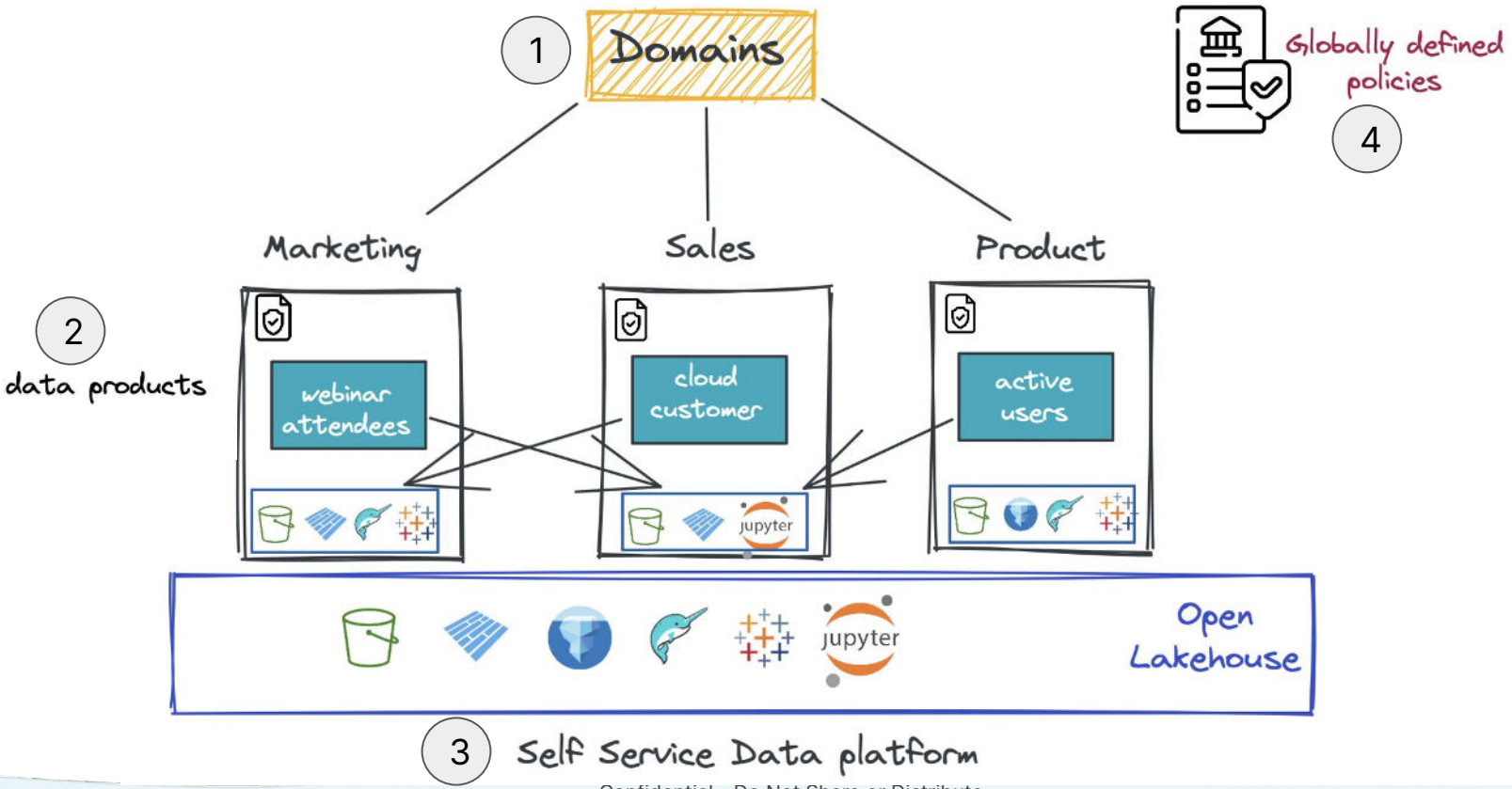


Data Mesh, Open Lakehouse : complementing each other

- organizational problems needs the right technical approach & vice versa
- Both are designed to deal with the **scalability** problem from 2 different angles
- We need the right set of tools/infra to create, share & manage data products
- Aligning organizational & technological components with the 4 pillars

Data Mesh Pillars	Organizational	Technological
Domain Ownership	✅	✅
Data as Product	✅	✅
Self service platform	❌	✅
Federated Governance	✅	✅

Open Lakehouse as the Platform to Support Data Mesh



Benefits of Open Lakehouse in a Data mesh

- Problem is scalability. An open lakehouse is best designed for this
- Data is *'open'* in an open lakehouse - sharing, consumption easy
- A lakehouse platform facilitates a standard set of tools for self-service that abstracts infrastructure for data consumers
- Makes data architecture future-proof (new use cases)
- Open table formats like Apache Iceberg helps managing & consuming data products

Open Data Lakehouse

+

Data Mesh