



# COMMUNITY DAY

DACH

19th OCTOBER 2022

# Data Architectures for Cloud

(*& AWS services to build them*)

Anurag Kale  
anuraganil.kale@polestar.com  
Cloud Software Architect  
Architecture & Standards  
Polestar Digital

# What's in this session?

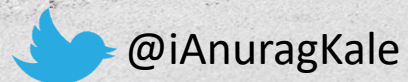
- Quick look at “Why” should we care?
- Abstract view of cloud data architectures
  - Data Lakes
  - Data Lake House
  - Data Mesh
- Meta-questions
  - What are these?
  - When to consider them?
  - What is required to build them?
- Walkthrough of some example architectures

# \$whoami

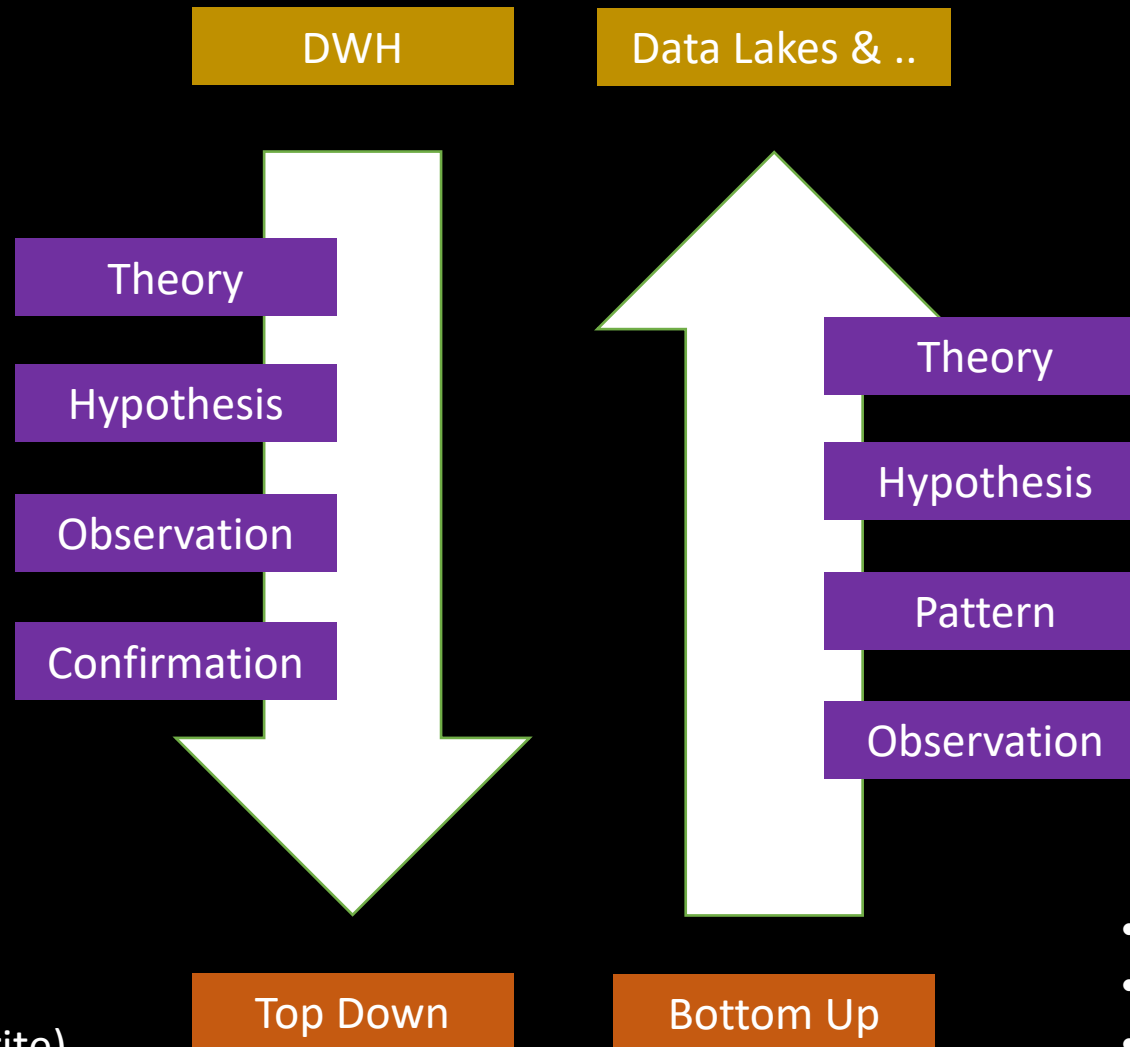
Anurag Kale

Cloud Software Architect at Polestar Digital

AWS Data Hero



# Two approach to get value from data



- Know the questions to ask
- Do a lot of upfront work
- Model First (Schema on Write)

- Don't know the questions to ask
- Little upfront work
- Model later (Schema on Read)

# DWH is awesome if ..

- You aim to reduce stress on production system
- Want to do descriptive analytics (what happened, why it happened)
- Your data structure is predictable
- The data is a representation of an entity
- Optimise read access and sequential reads
- Need one version of truth
- Use for Master Data Needs
- Build BI and Reports for Business Users

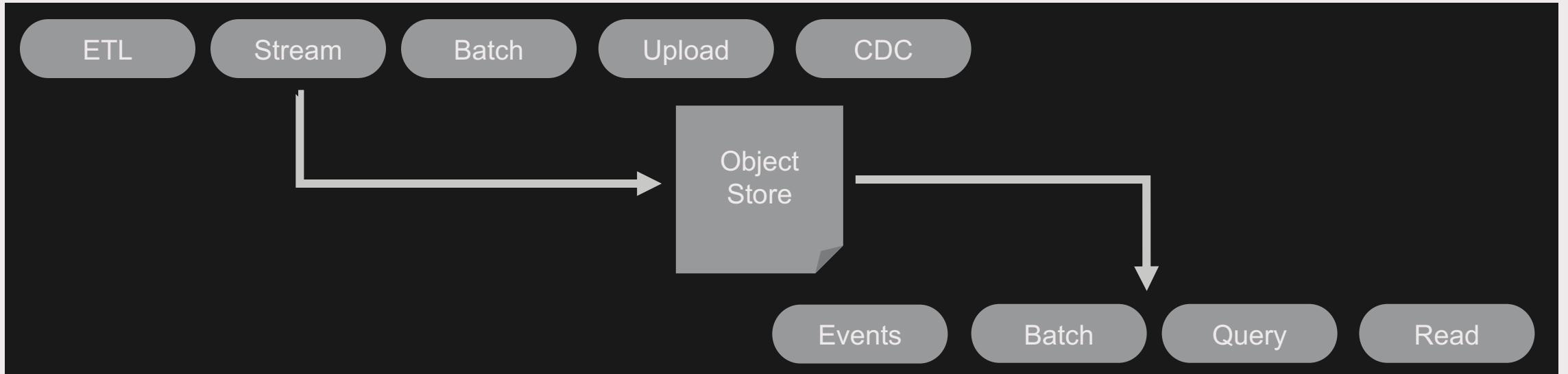
# Data Lakes

# Bottom Up Approach

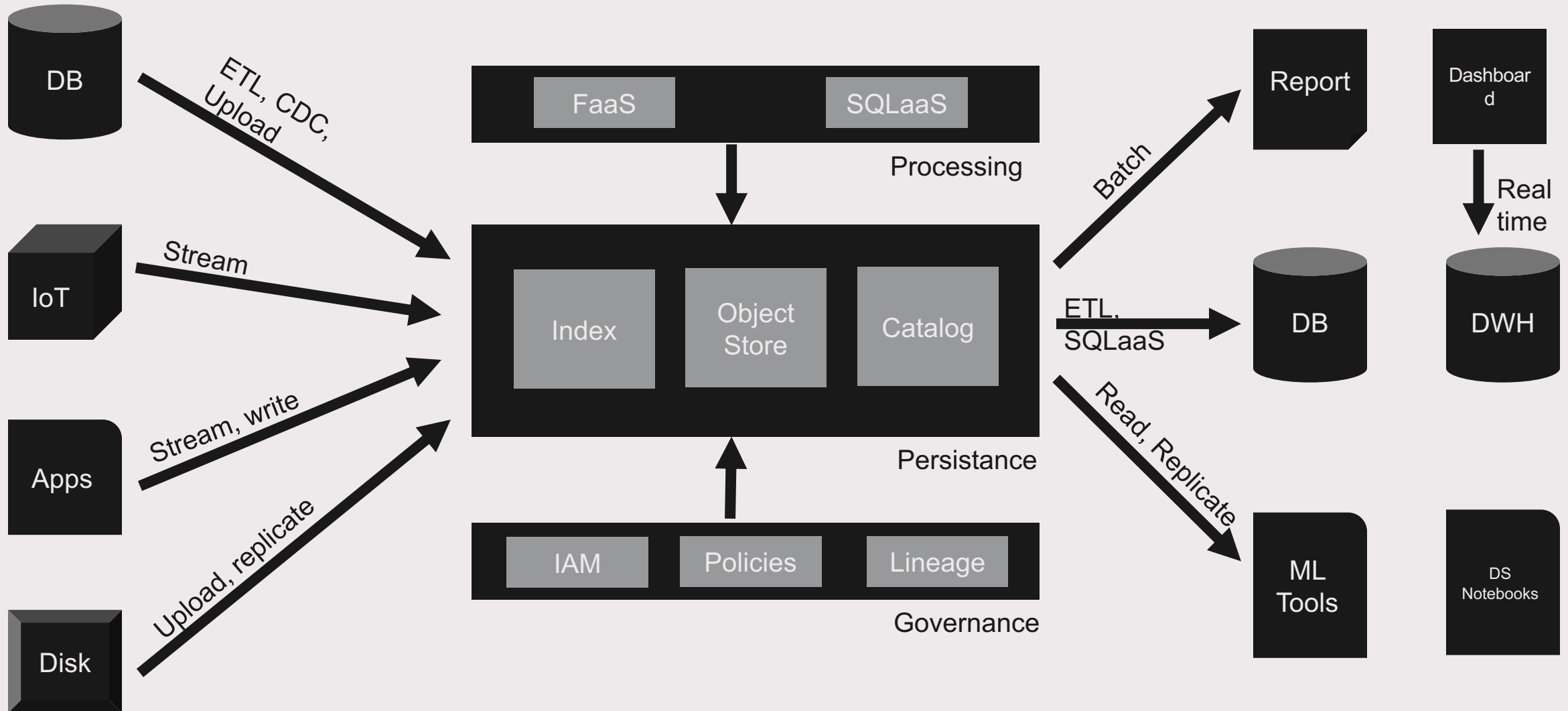




New kid on the block  
          
Data Lake in Cloud



# Data Lake as Platform



# Data Lakes is awesome if ..

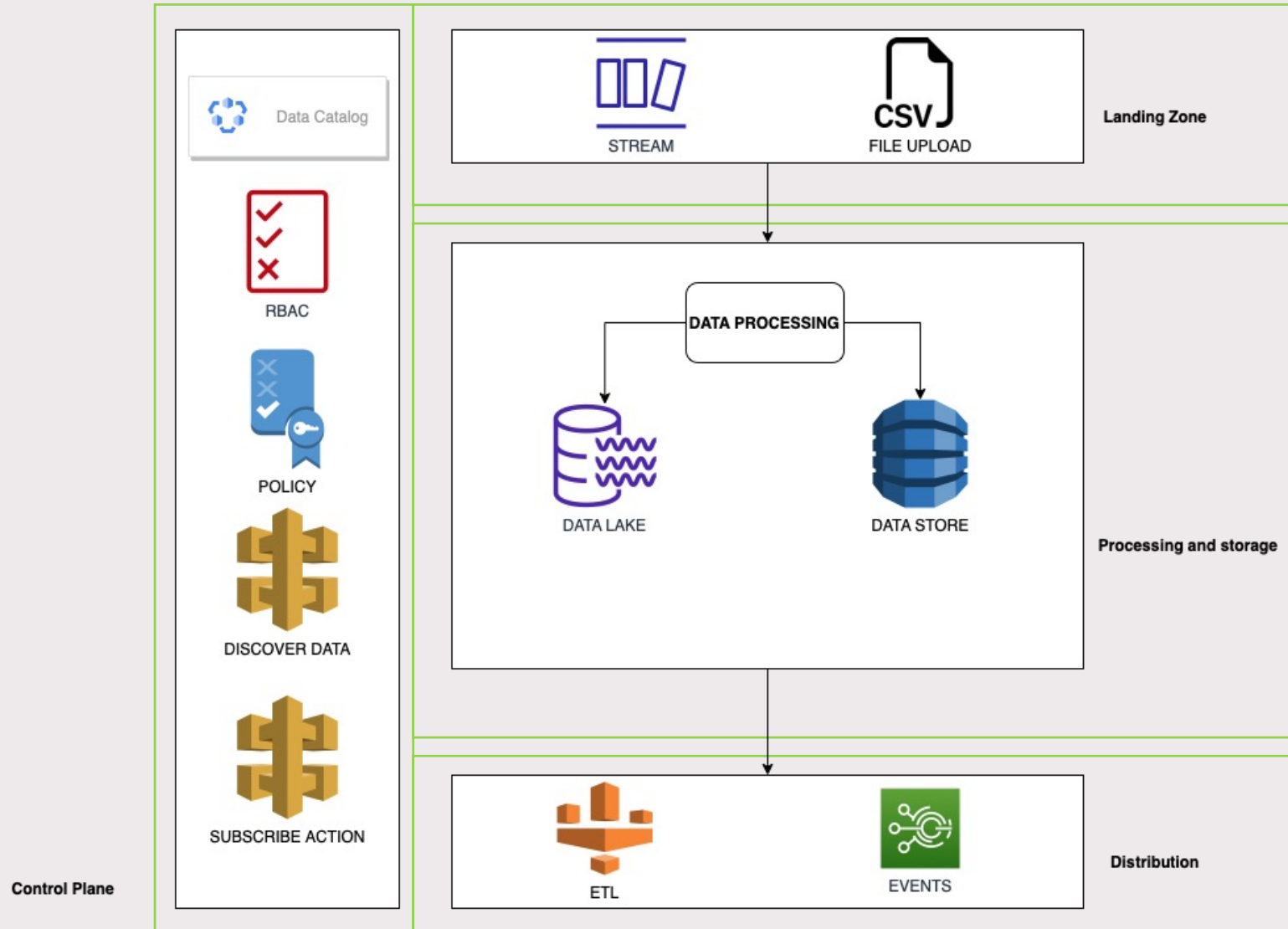
- You want to store data inexpensively
- Need a central place to source data
- Treat it as a staging area / repository
- Technically equipped to work with Schema-on-Read
- Lookin to empower your data enginners and data scientists
- Support multiple data formats
- Enable predictive analytics

# Sample implementation

Logical View

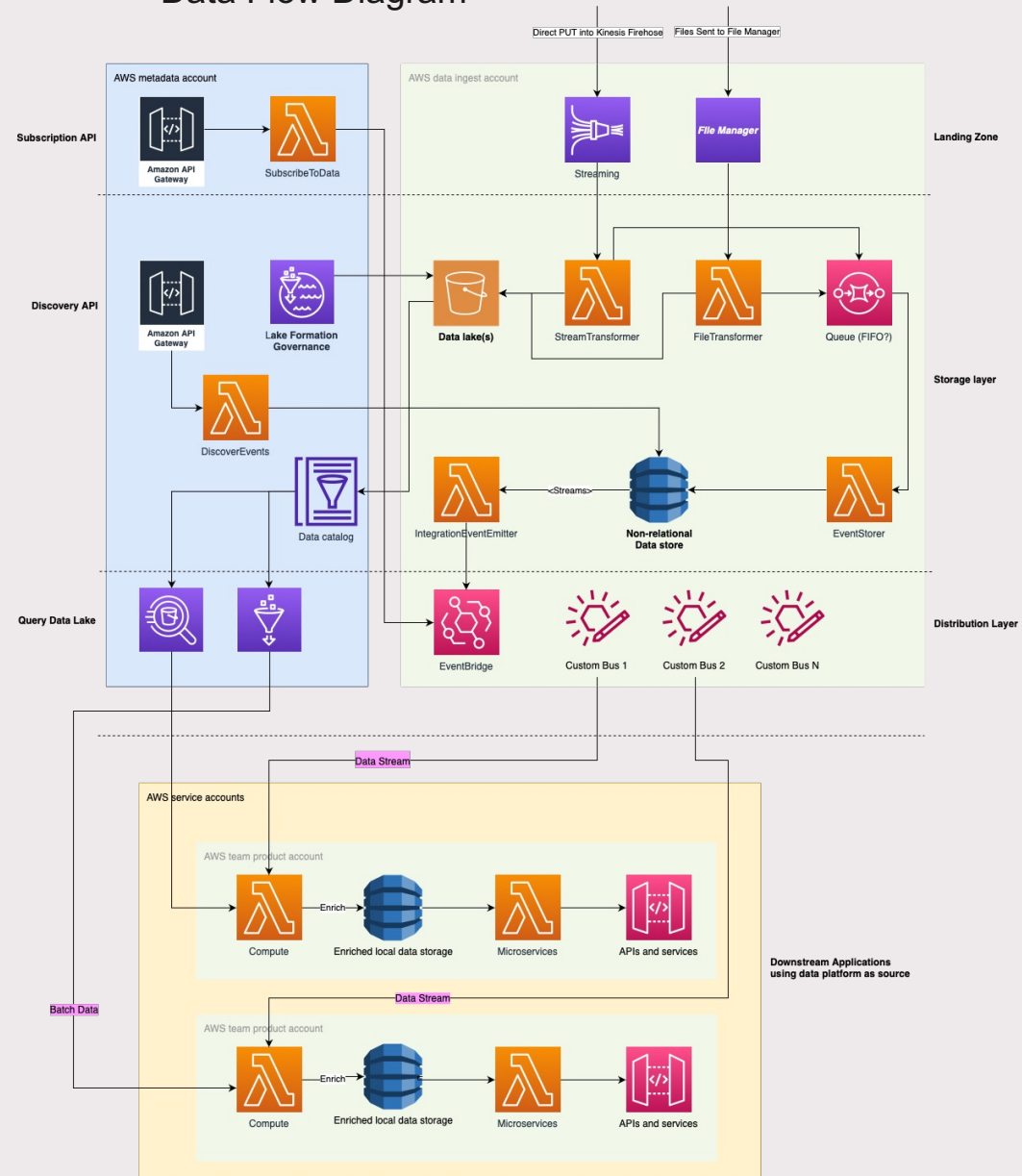
# Polestar Data Platform

## Logical componets



# Polestar Data Platform

## Data Flow Diagram



# Data Lake House

Goodness of both DWH and Data Lake

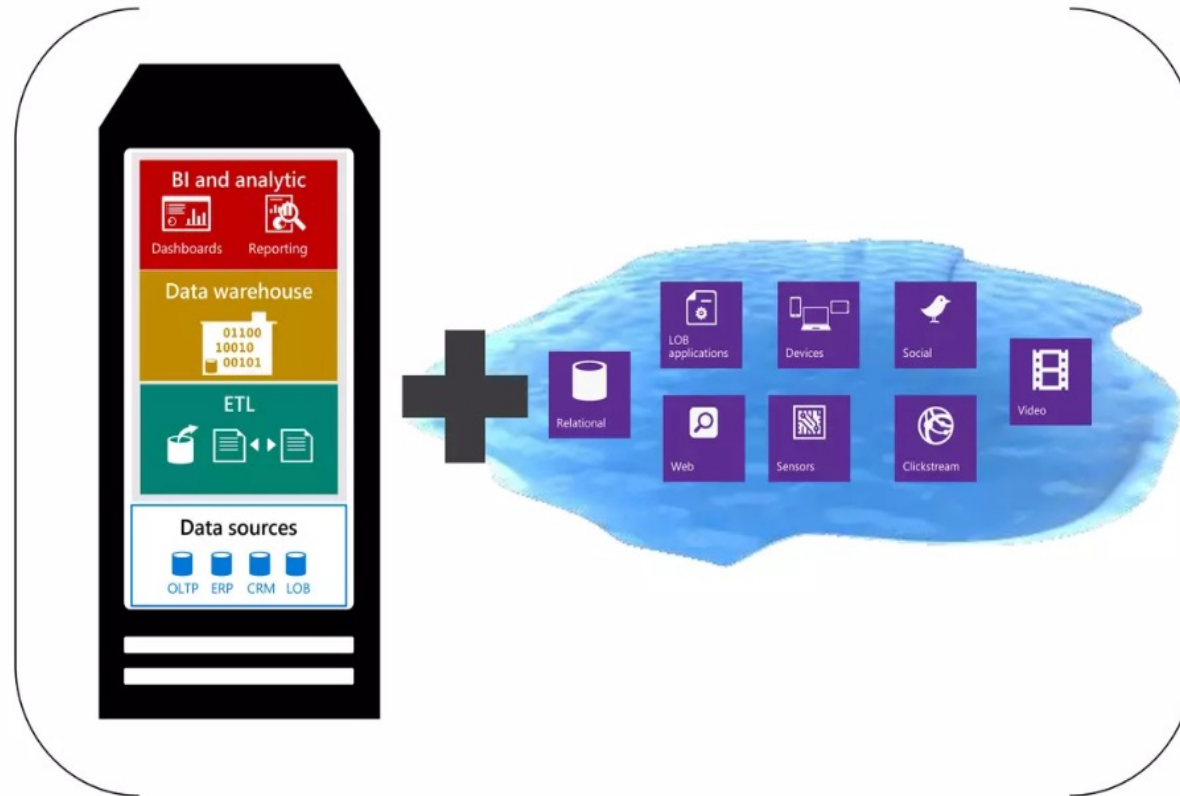
# Data Lake House

What happened?

Descriptive Analytics

Why did it happen?

Diagnostic Analytics



What will happen?

Predictive Analytics

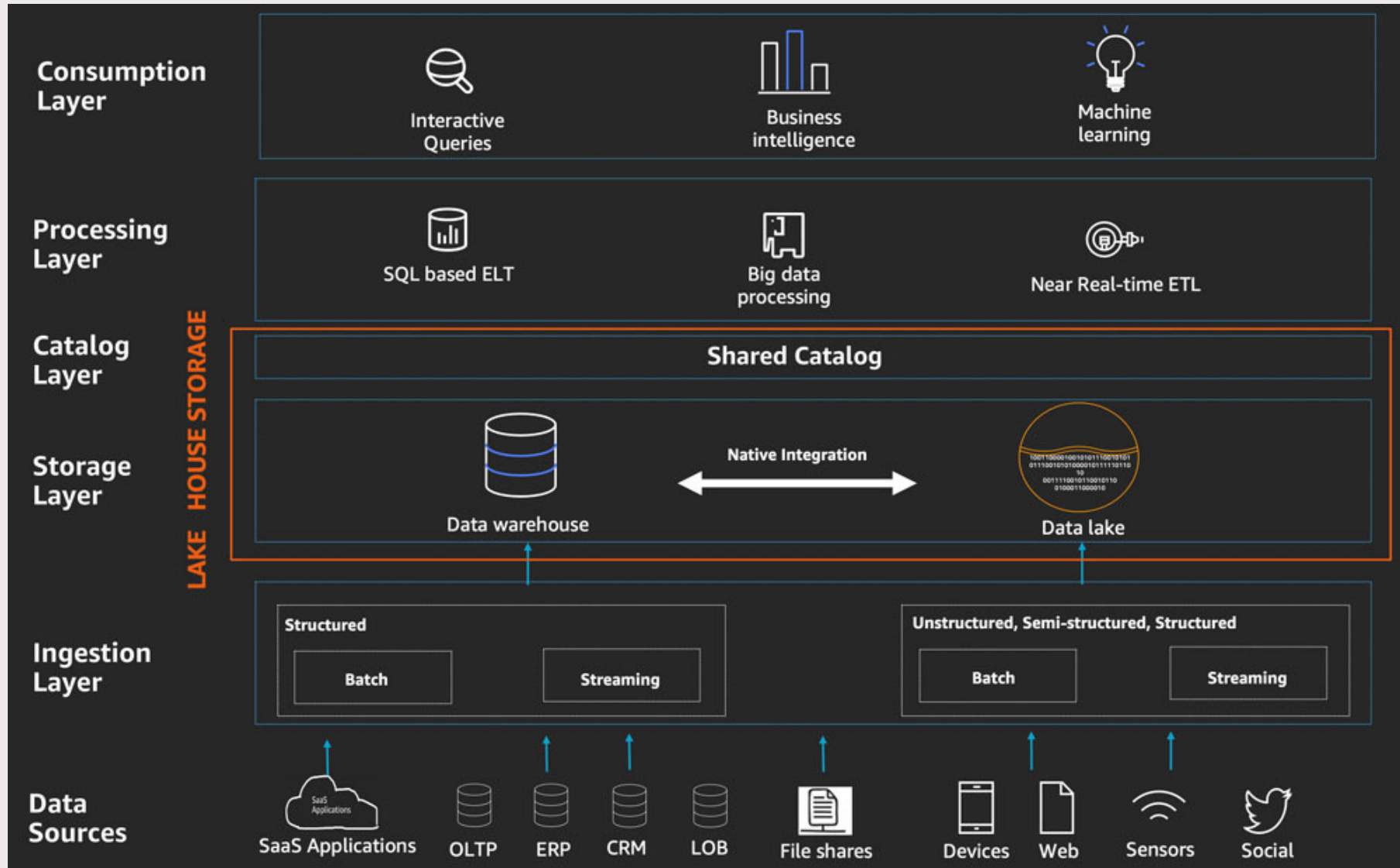
How can we make it happen?

Prescriptive Analytics



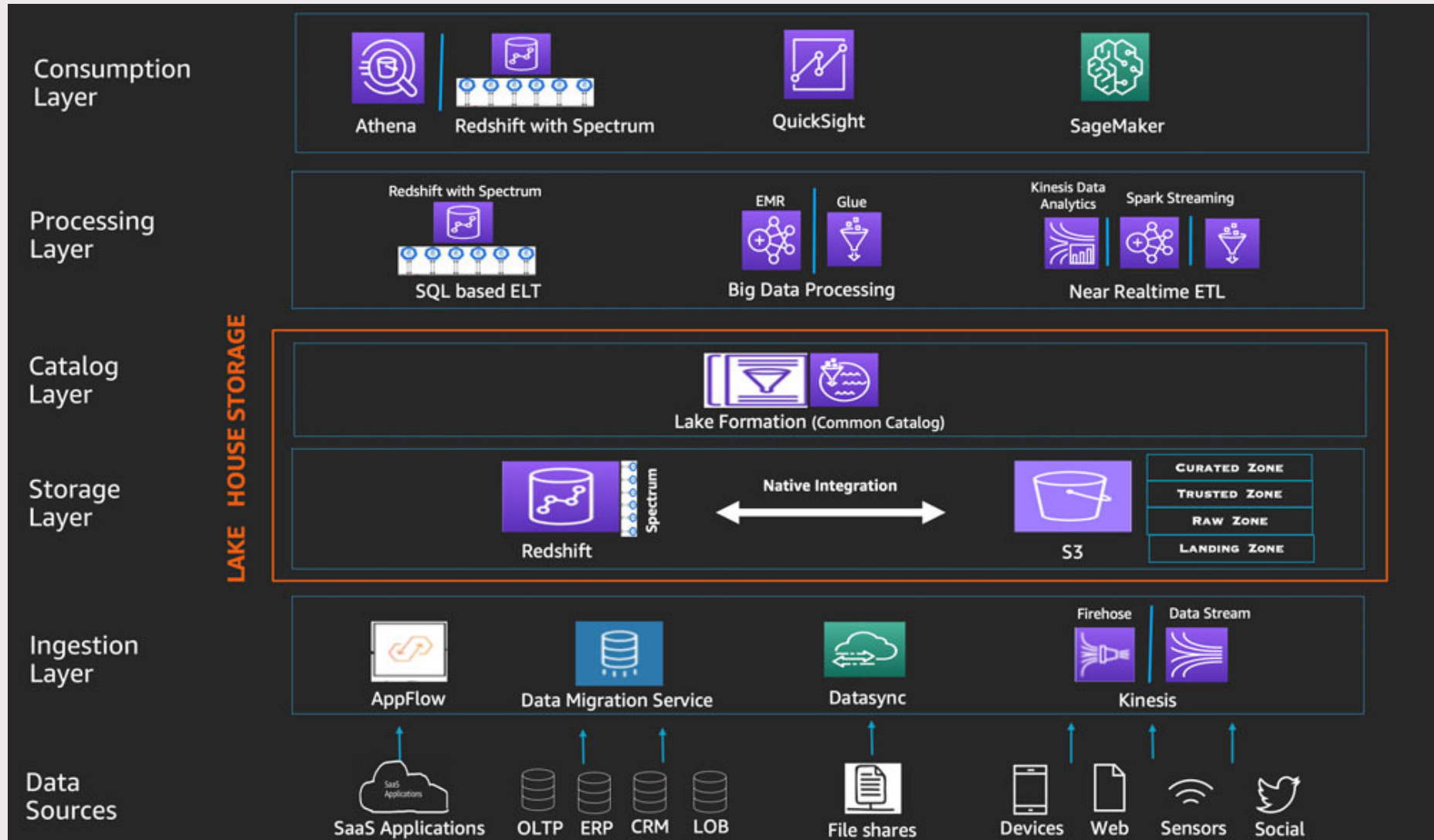
# Data Lake House

## Logical View



# Data Lake House

## Logical View

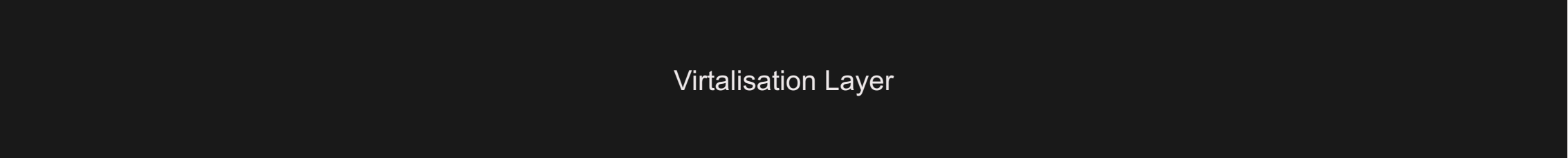


# Data Mesh

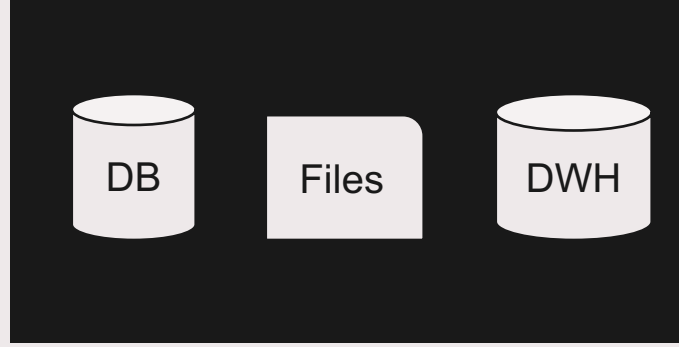
Or Data Fabric

Other Data Architectures

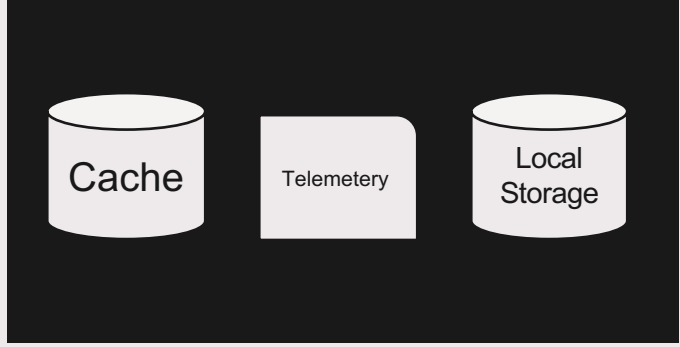
          
Data Mesh / Data Fabric



Cloud



On Prem

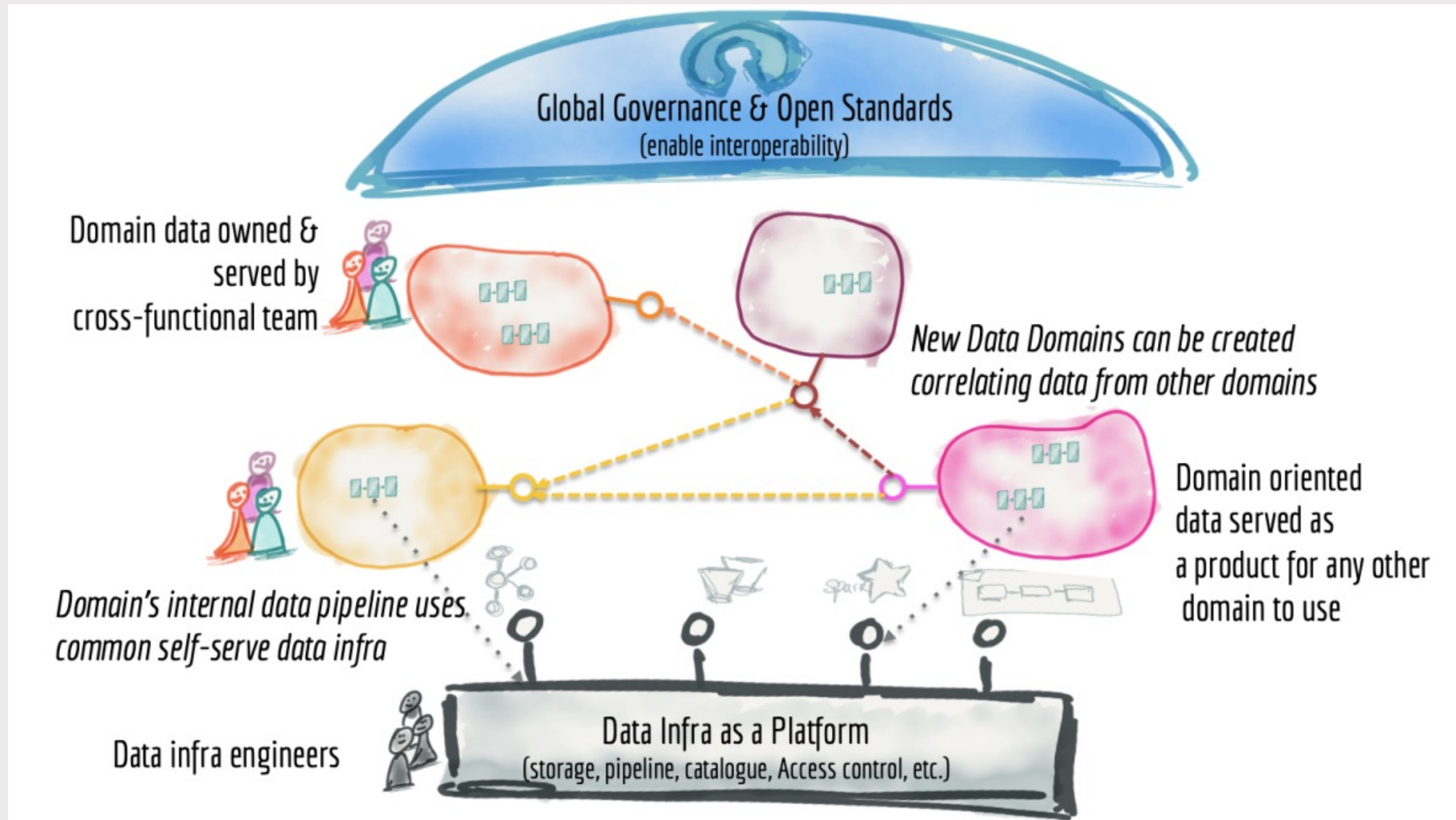


Edge

# So what's Data Mesh trying to do?

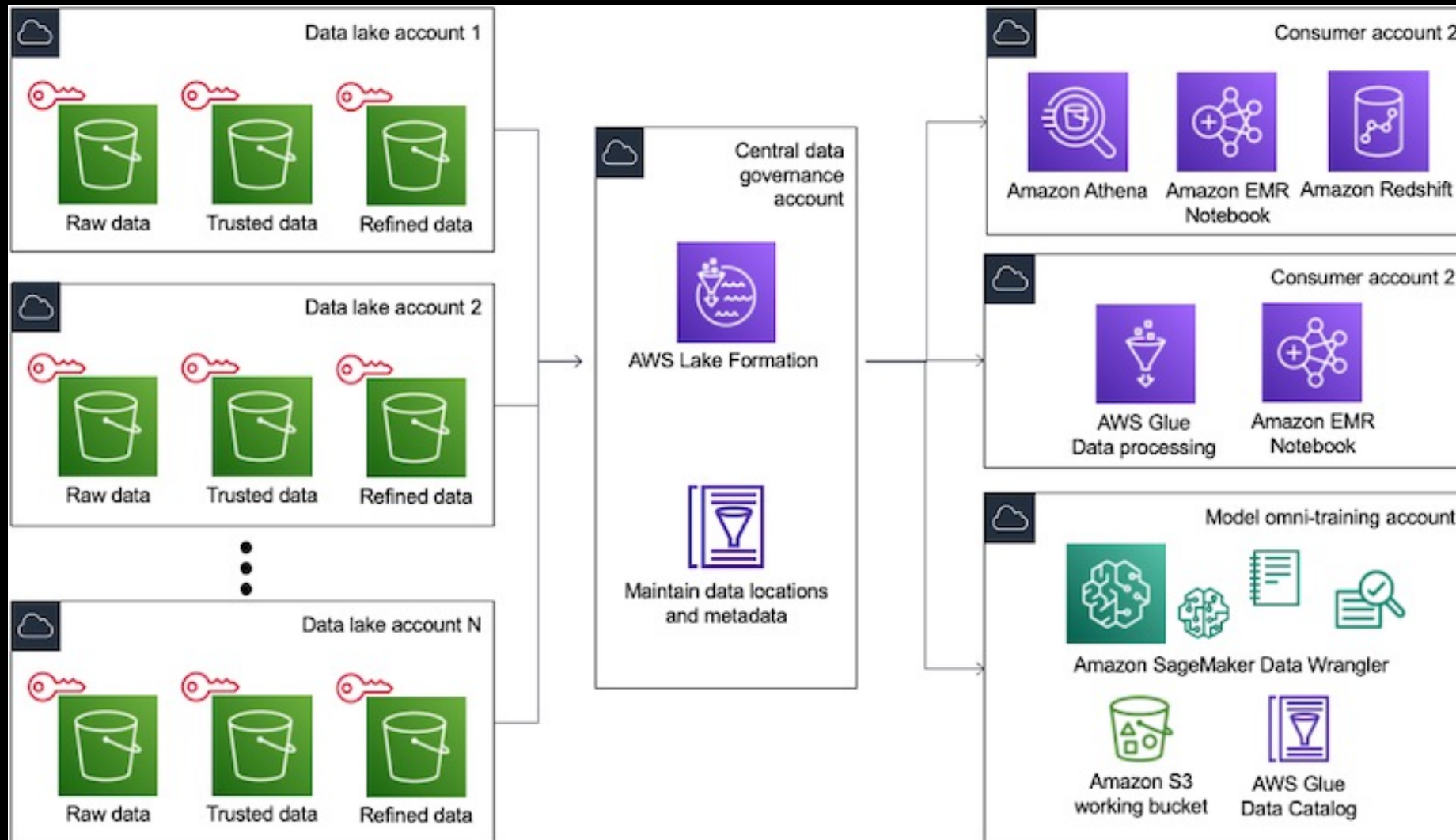
- Applied Domain Driven Design to Data
- Tries to NOT have multiple copies of data
- Access data where it is
- Provide generic data pipelines to address org wide concerns
- Utilise domain knowledge of the data owners
- Reduce egress cost's by copying minimal data

# Data Mesh Visualised



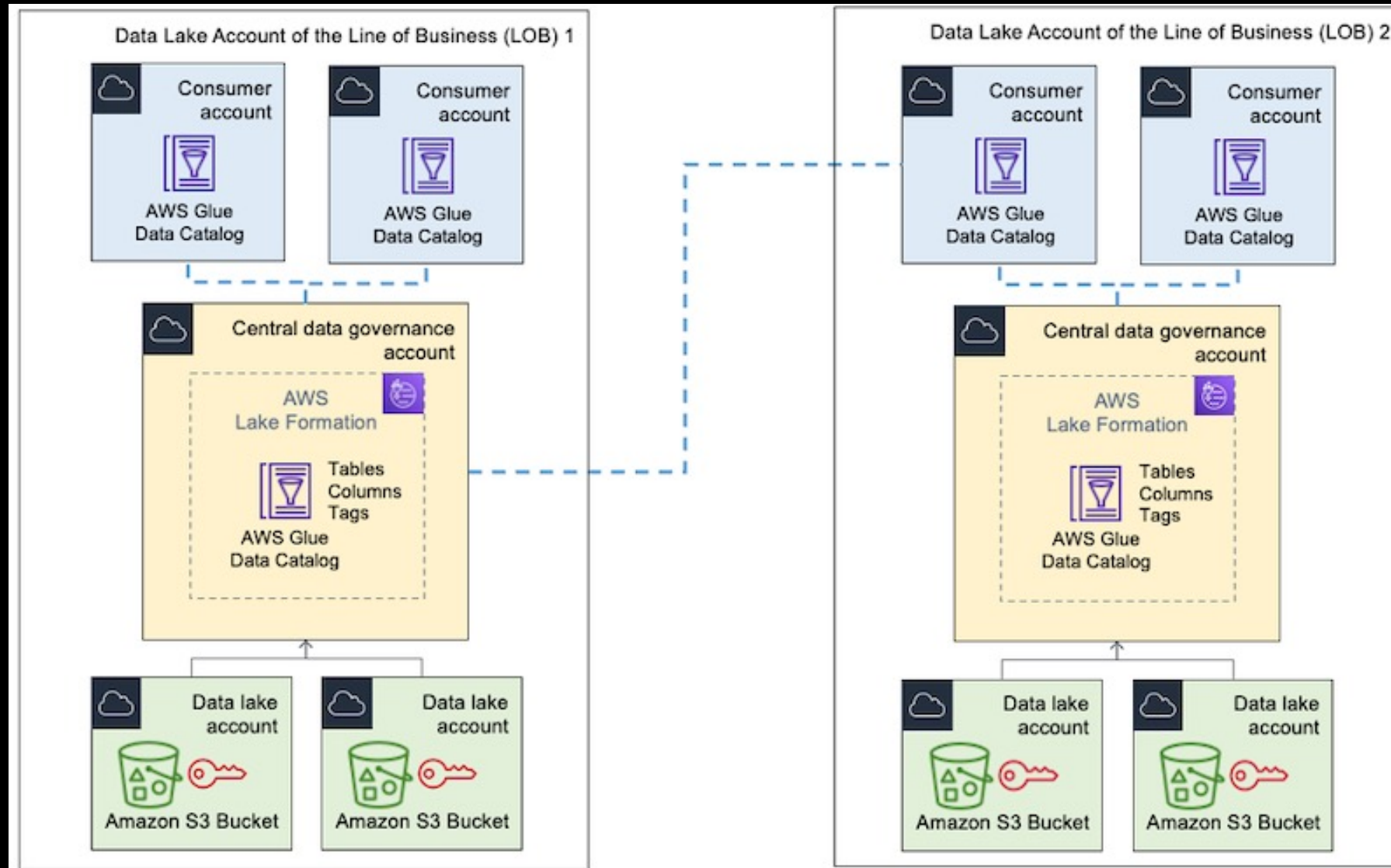
# Sample Implementations

# Collect Meta-Data – Centralised Approach





# Collect Meta-Data – LOB Approach



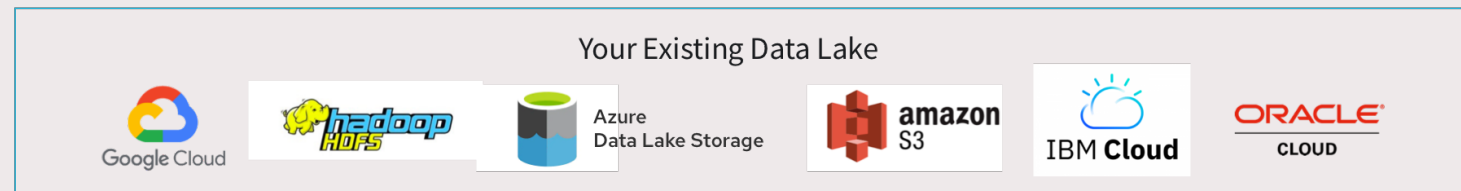
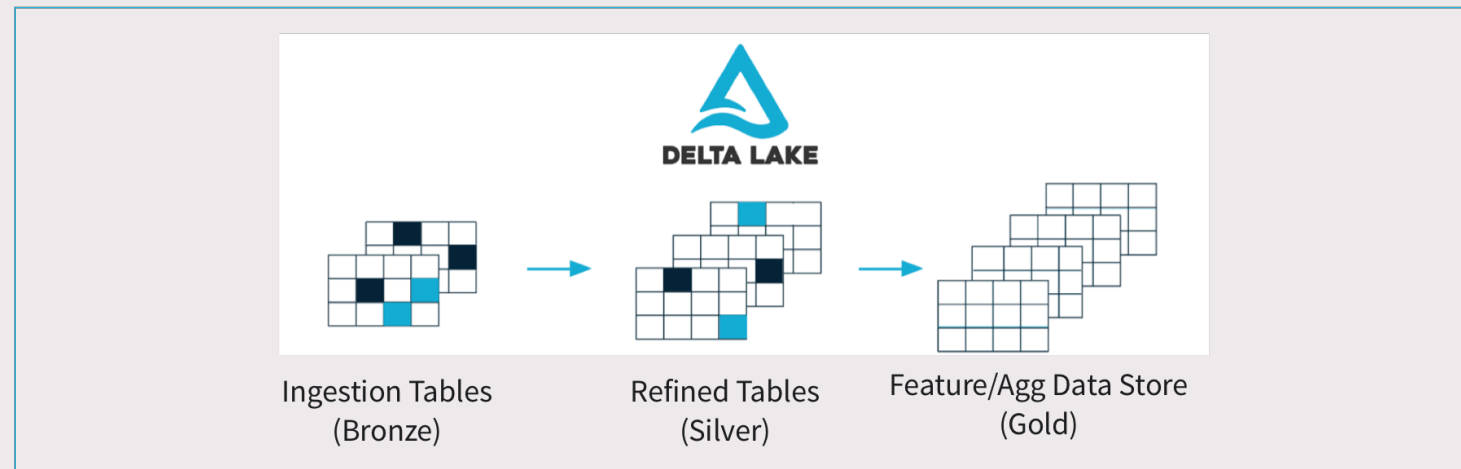
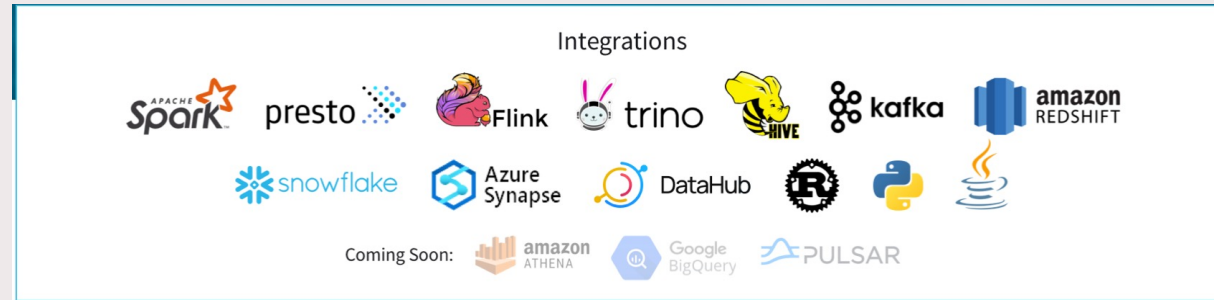
# Final notes on Data Mesh

- (CON) Lack of industry definition on
  - What exactly is it?
  - How should the interface look like?
  - Any standard approach to storage?
- (PRO/CON?) Data Expertise
  - Each team needs a data expert
  - Distributed ownership, needs trust based environment
- (PRO) Access of data
  - Better understanding of data
  - Easy sourcing

Honorable Mention

# Other Data Architectures

## Delta Lakes



Thank You! Questions?



@iAnuragKale