aws
COMMUNITY DAY
WARSZAWA
14th OCTOBER 2022

Building a Serverless Event Store

Store, distribute and discover events

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

Polestar

There is no sense in talking about the solution before we agree on the problem, and no sense talking about the implementation steps before we agree on the solution.

-Efrat Goldratt-Ashlog [The buy in process according to theory of constraints]

What’s in this session?

- Explore business use case
- Blueprint the business solution
- Evaluate different technical candidates for the problem
- Logical Solution Architecture
- Data Flow Diagram
- What’s next?

$whoami

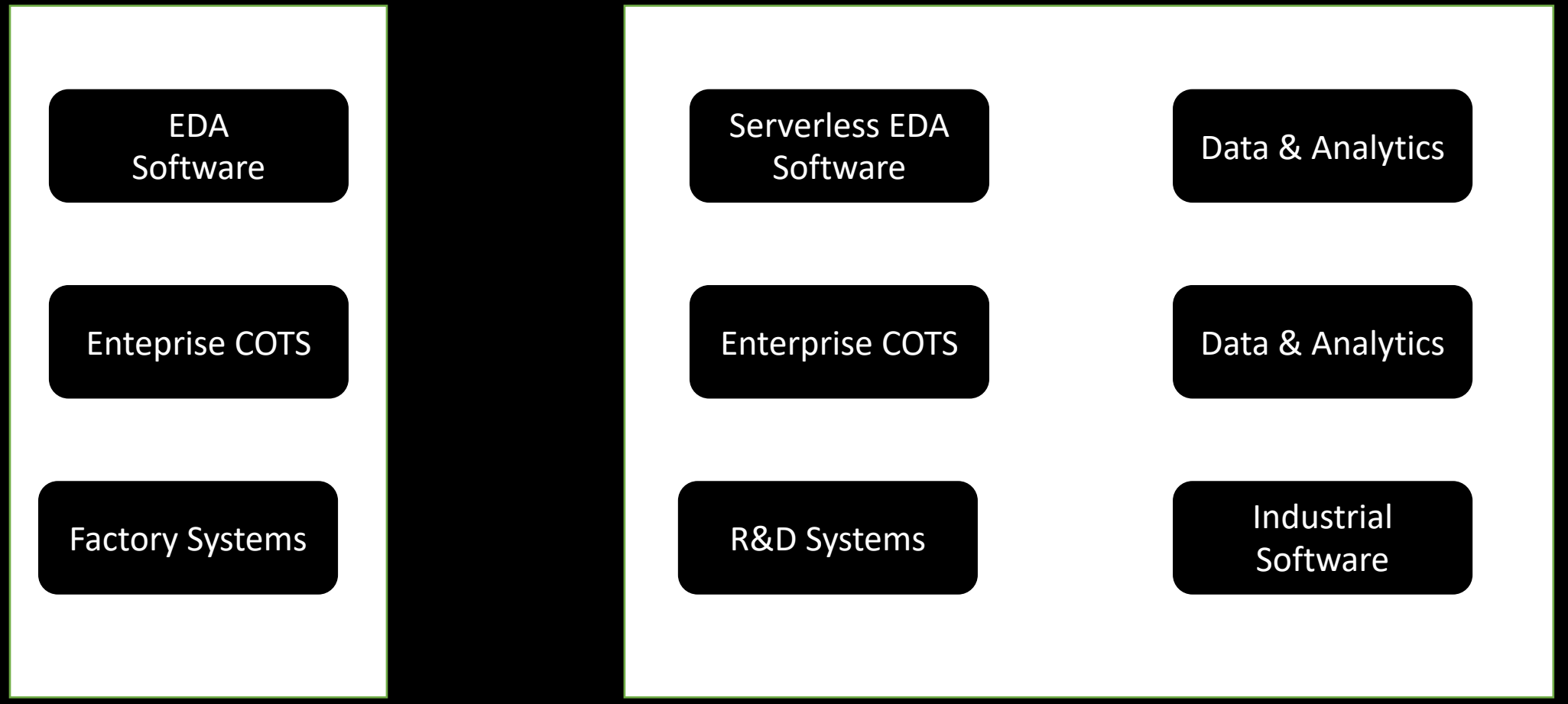
Anurag Kale

Cloud Software Architect at Polestar Digital

AWS Data Hero

@iAnuragKale

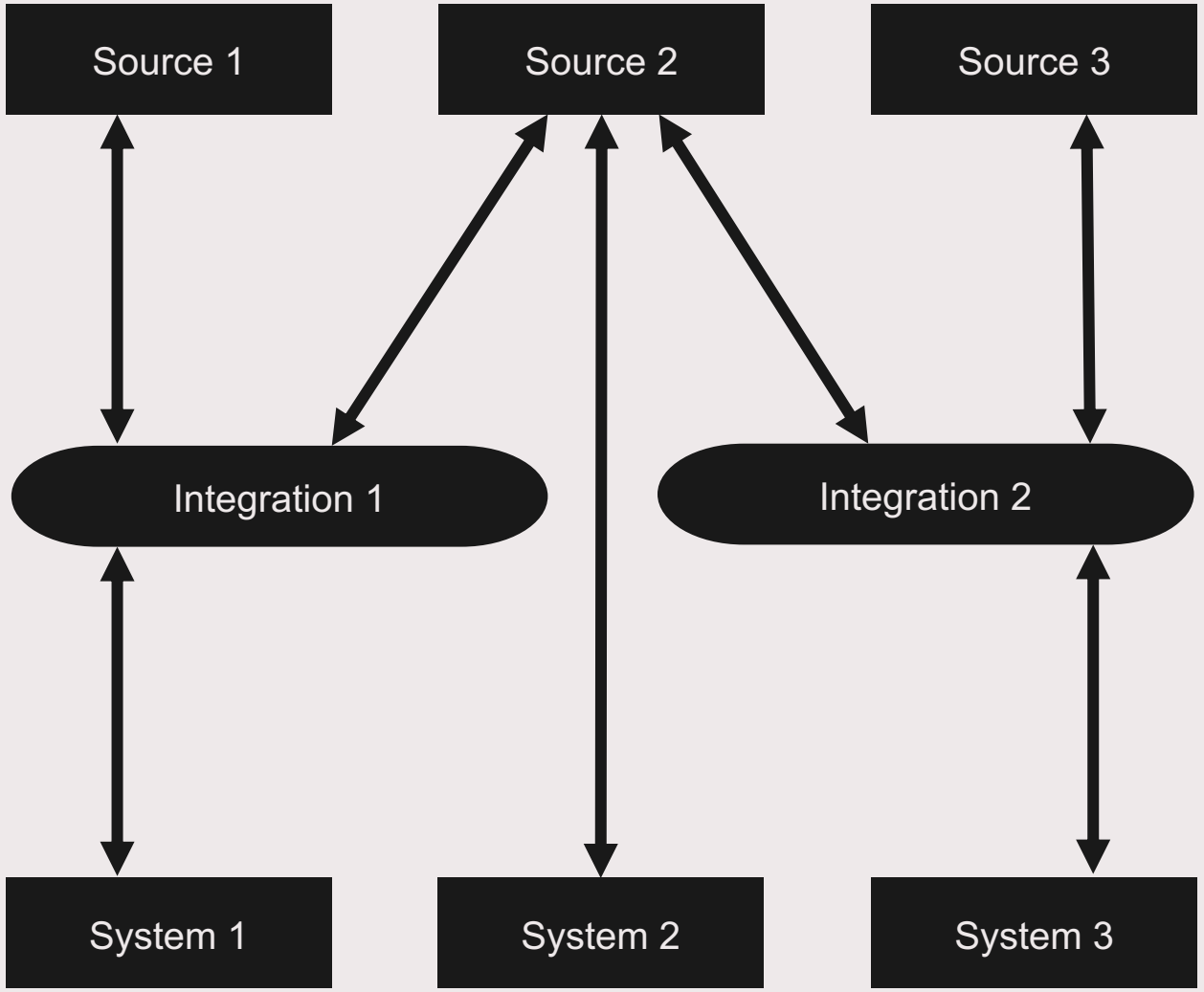
Current Landscape



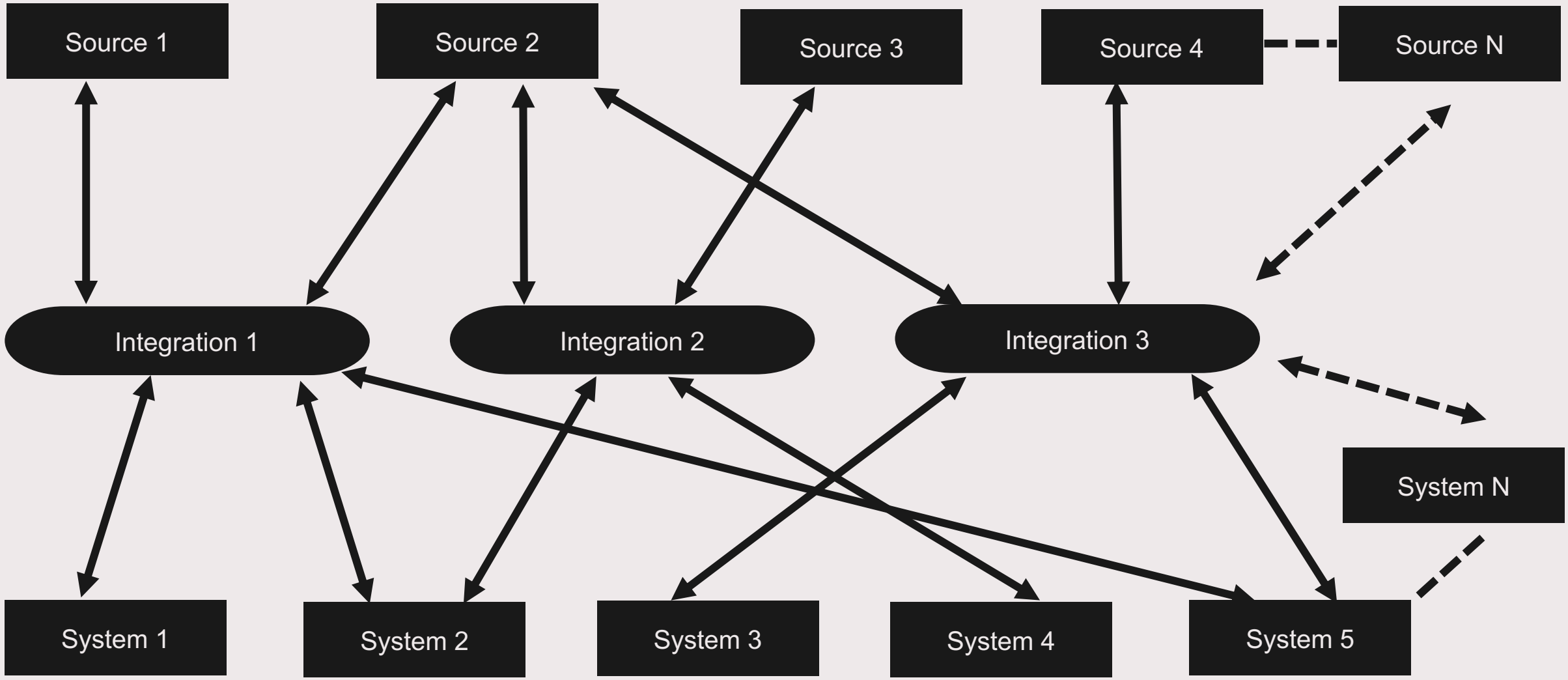
Polestar

The Problem

The problem - starts from here….



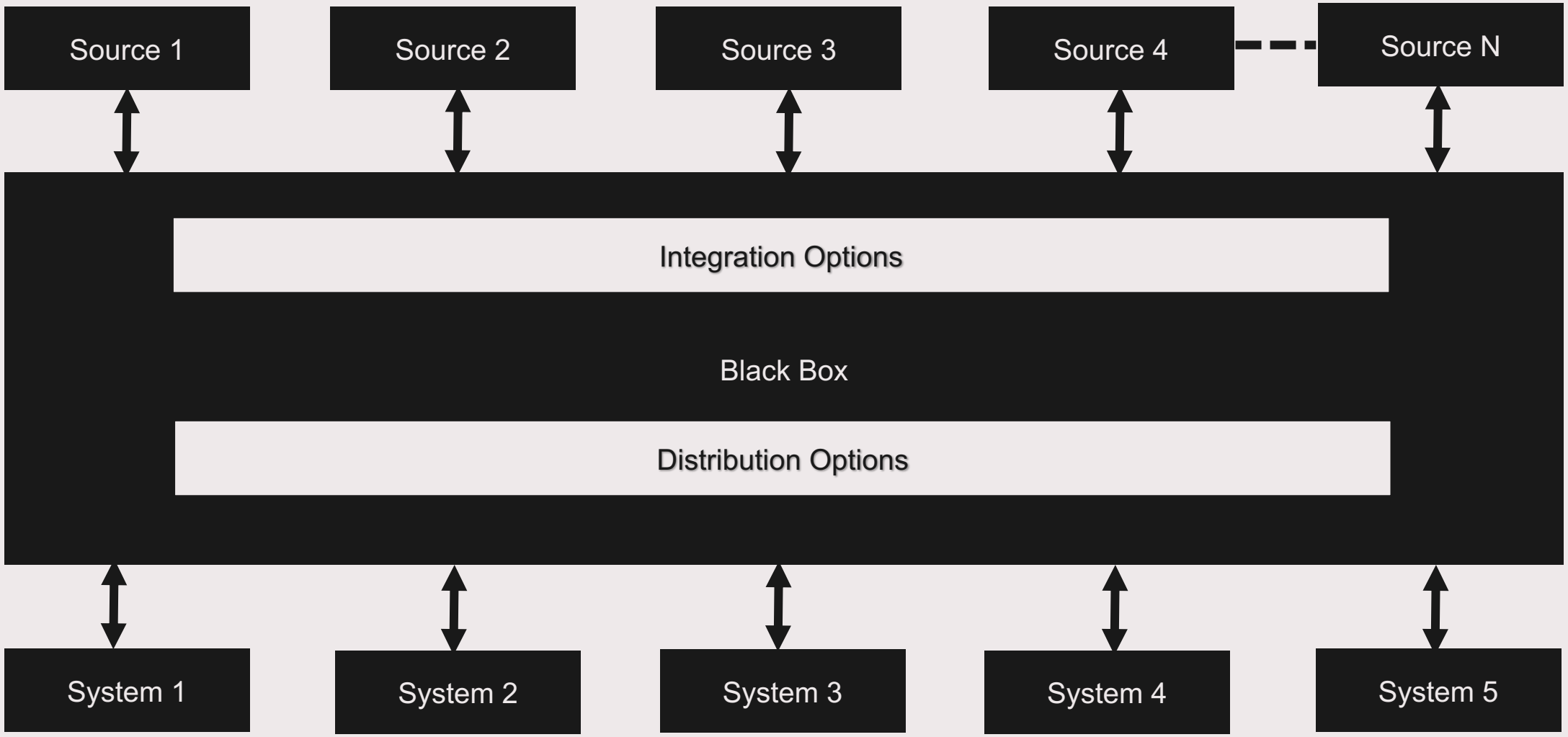
Soon this happens…..



What are we trying to address?

- Standardise collection of data from external sources
- Avoid point to point integrations
- Consolidate efforts for common used data sources
- Own Polestar data in unmutated format
- Enable current way of event driven architectures

Ideal Solution – Business Architecture



The proposed black box

Ideal properties

- Support multiple input methods
- Flexible and store multiple data formats
- Store data in original format i.e. unmutated data
- Store data for longer run in cost efficient way
- Good data discoverability
- Distibute data in events
- Adheres to IT Principles (Serverless / Cloud Native)
- Open to future expansions
- High sustanability impact (minimal CO2 emissions)

Polestar

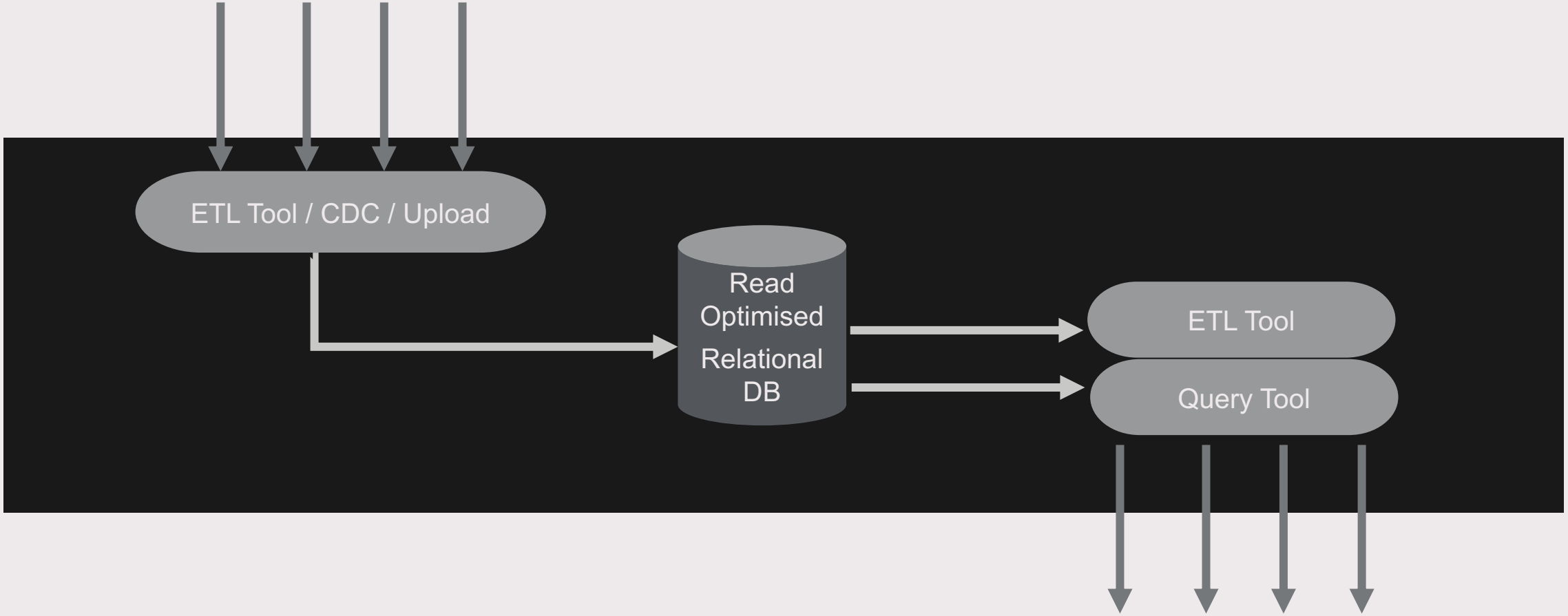
Candidates for Black Box

Data Architectures

- Enterprise Data Warehouse Approach
- Data Lake
- Data Lake House
- Delta Lakes
- Data Mesh / Data Fabric

The Usual Suspect

Enterprise Data Warehouse



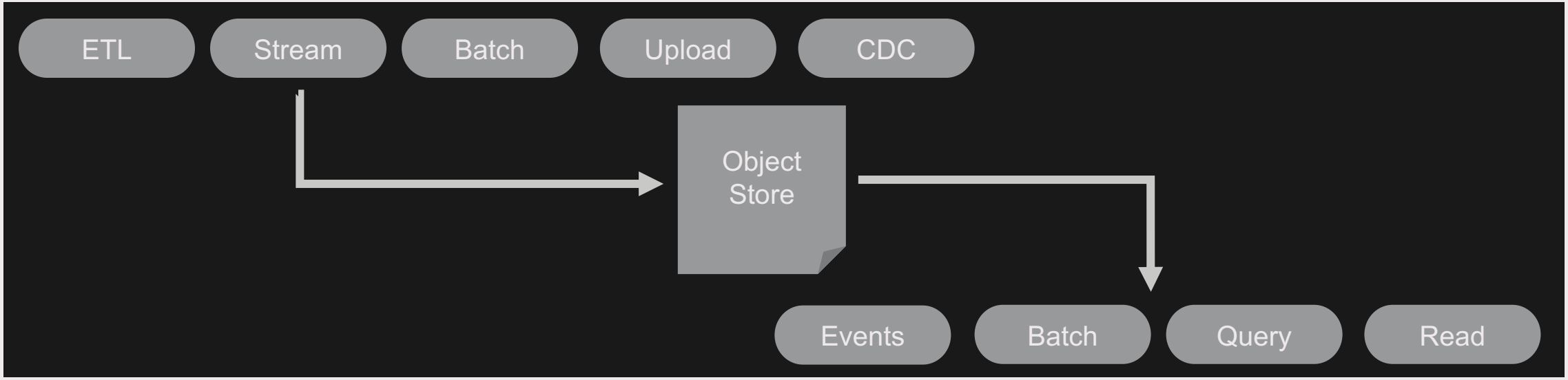
Evaluation of EDW Approach

- Proprietary tools, OS and expertise required
- Heavy licence costs
- Needs provisoned capacity for compute, network, storage
- Connection pooling issue (not suitable for AWS Lambda)
- Violates IT principles (requires down time, non-modular, does not work well with EDA etc)
- Long project effors for addition of every new data source
- 24x7 provisoning → Higher CO2 emissions

New kid on the block

—

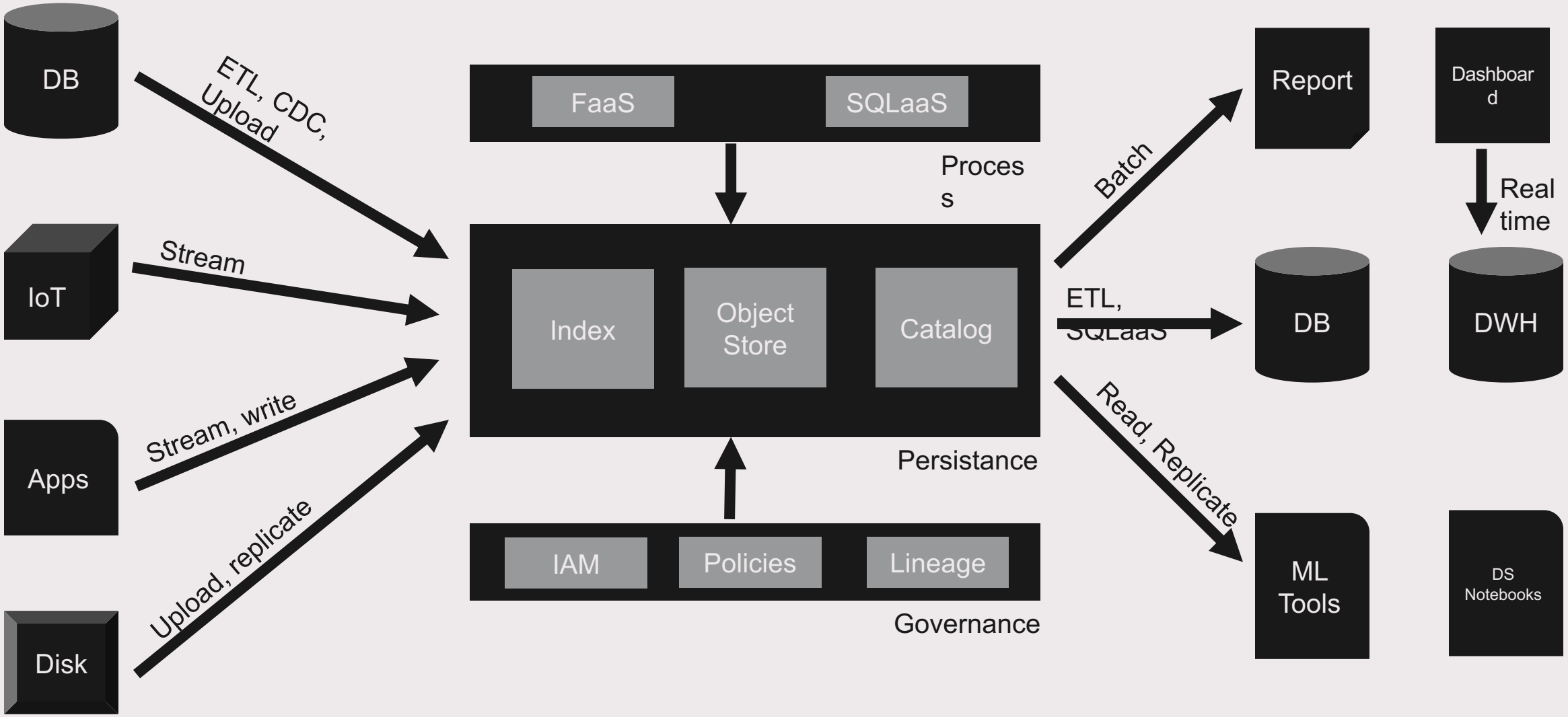
Data Lake in Cloud



Evaluation of Data Lake Approach

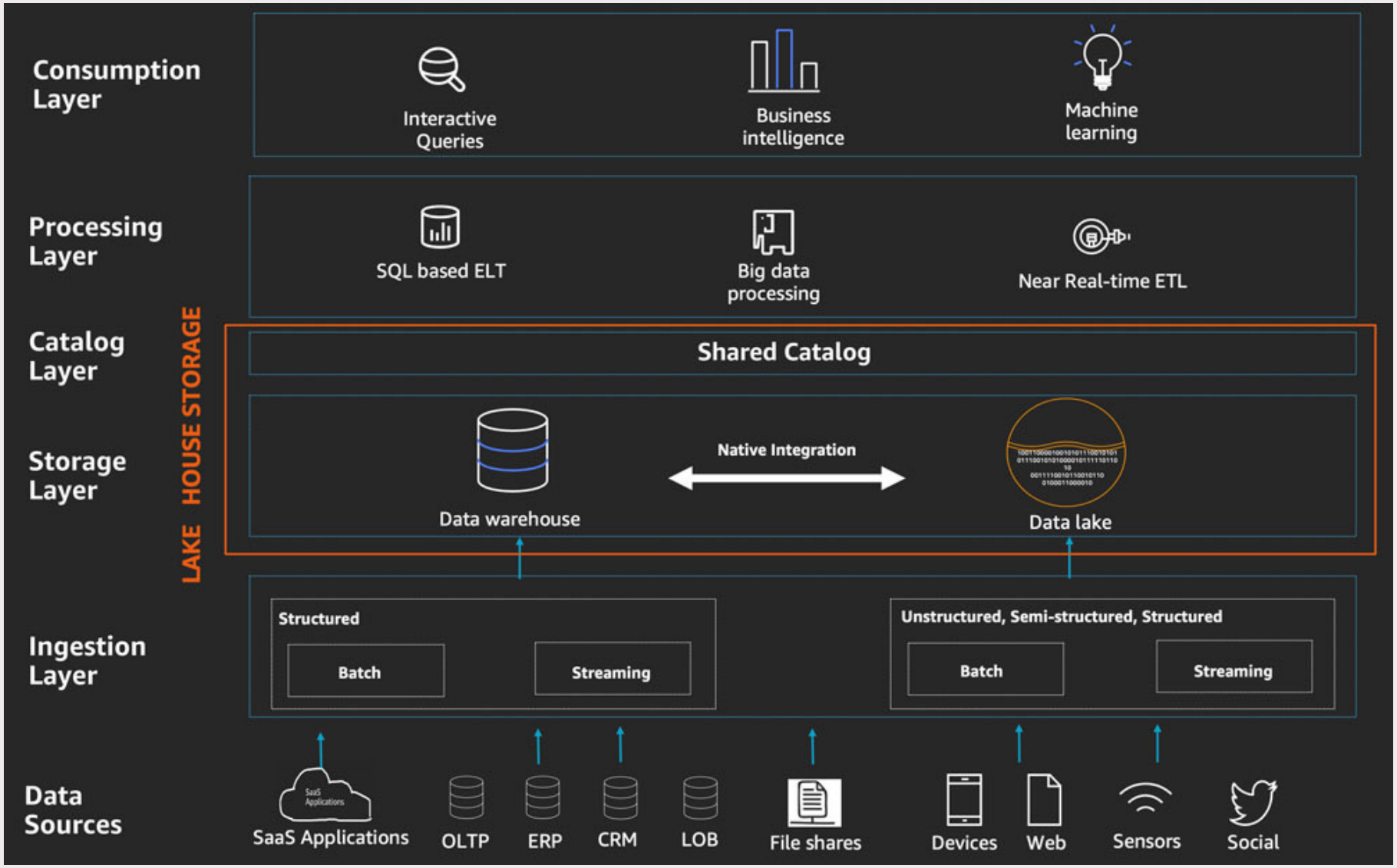
- Cloud Native, flat file storage
- No licence cost involved
- No provisioning of servers needed
- API based reads / writes
- Fits into all IT principles
- New source addition addition trivial task
- Significantly lower CO2 emissions
- Leaves room for expansion

Data Lake as Platfrom



Other Data Architectures

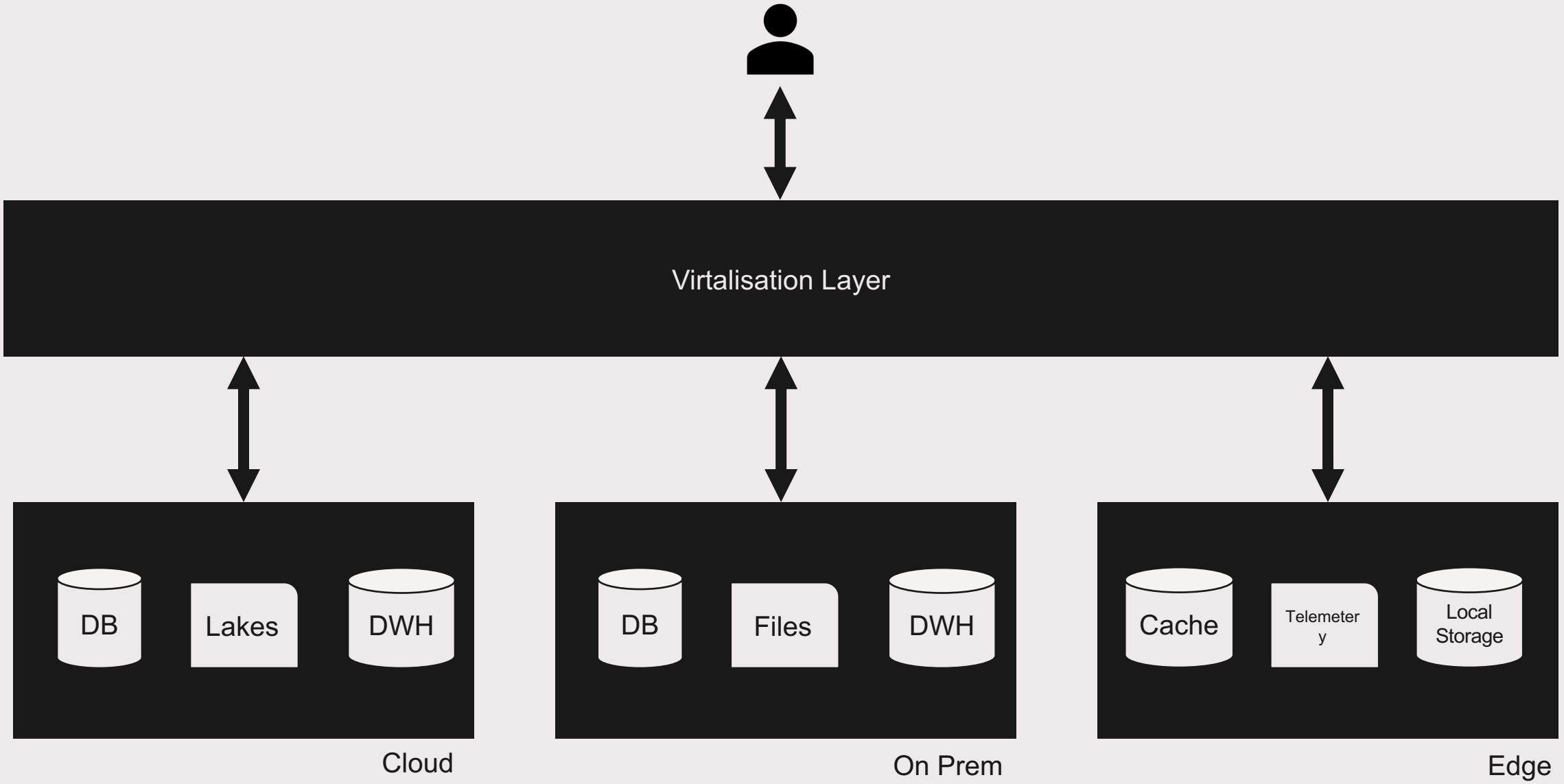
Data Lake House



Source - AWS

Other Data Architectures

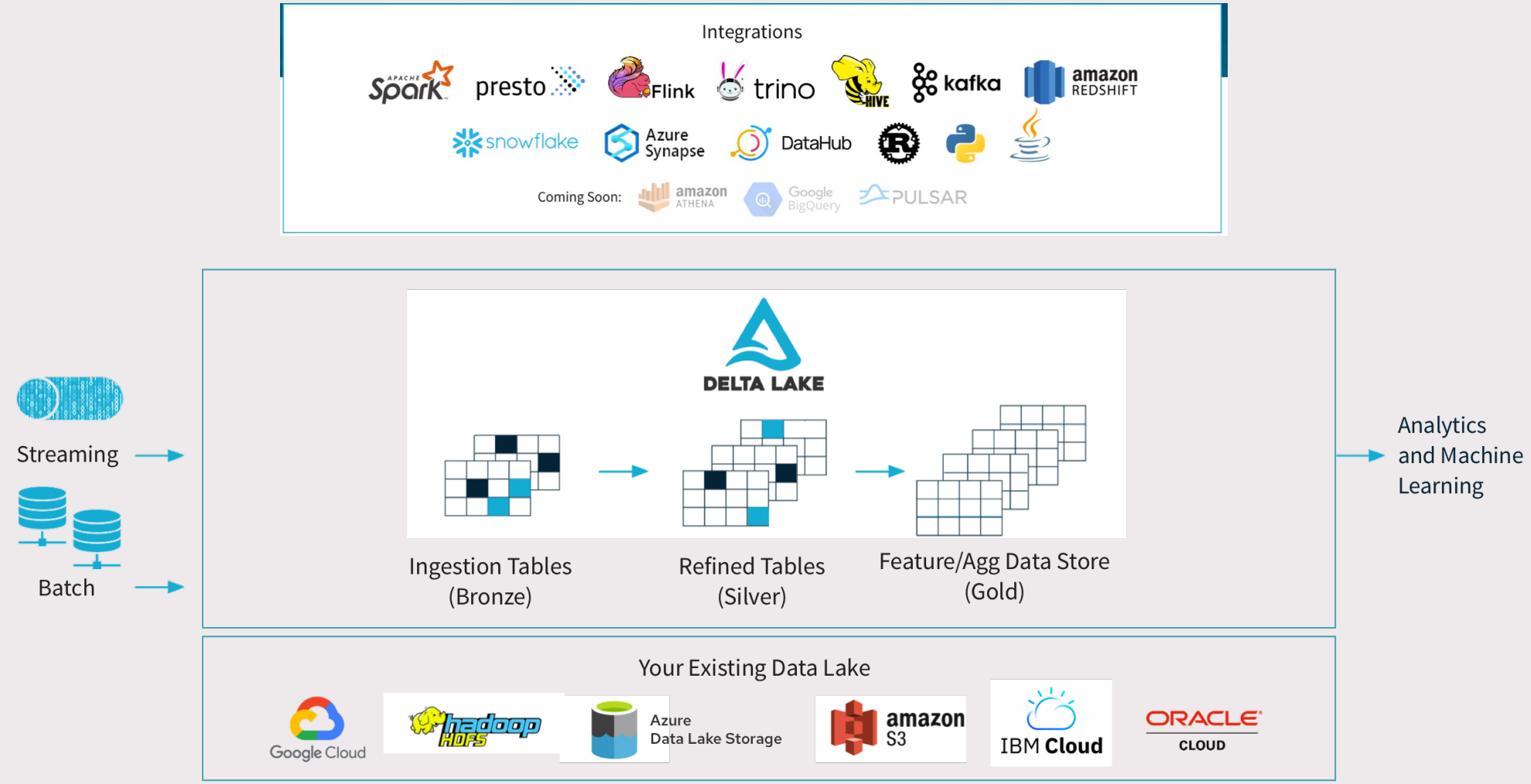
Data Mesh / Data Fabric



Source - AWS

Other Data Architectures

Delta Lakes

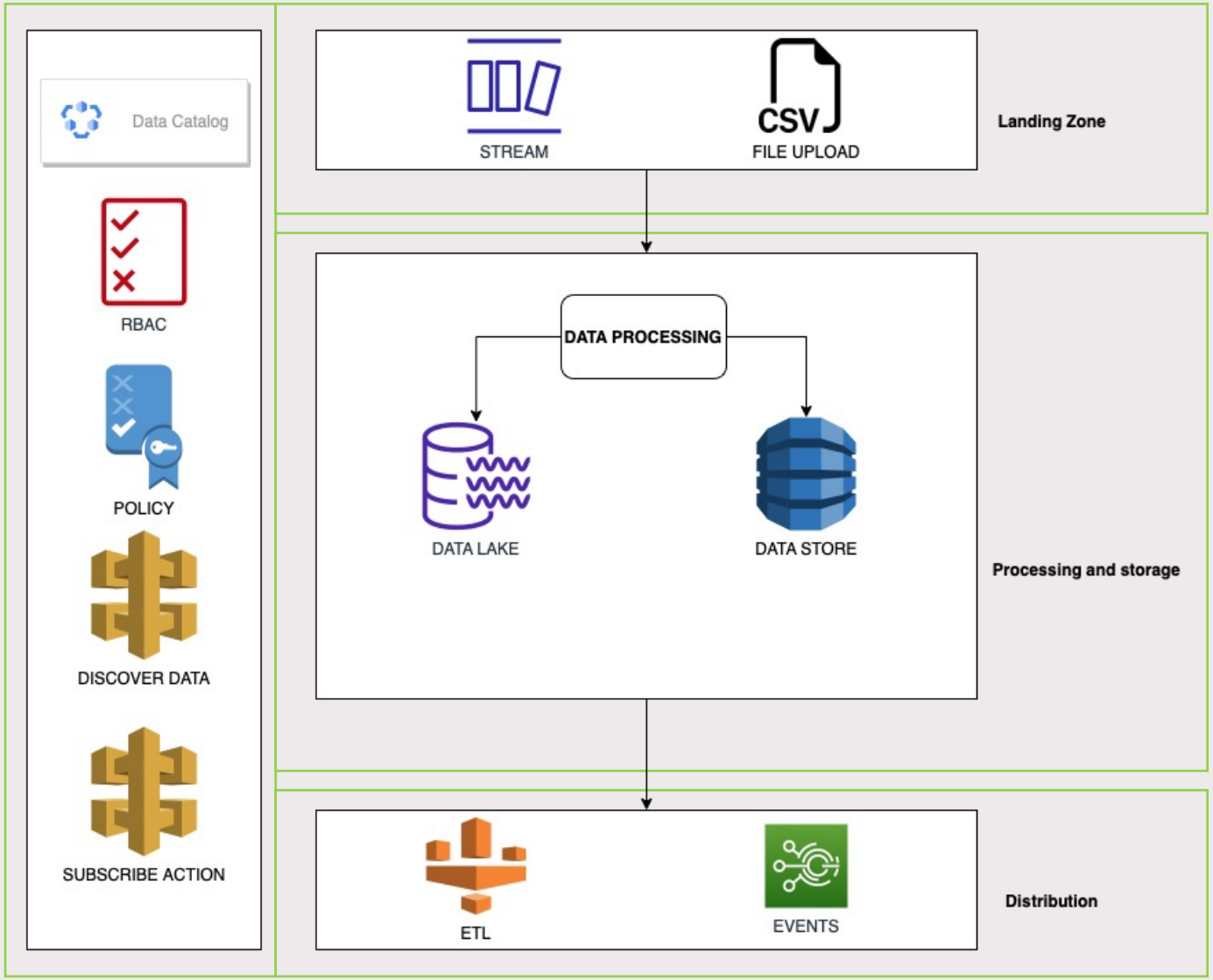


Polestar

Our Solution

Polestar Data Platfrom

Logical componets

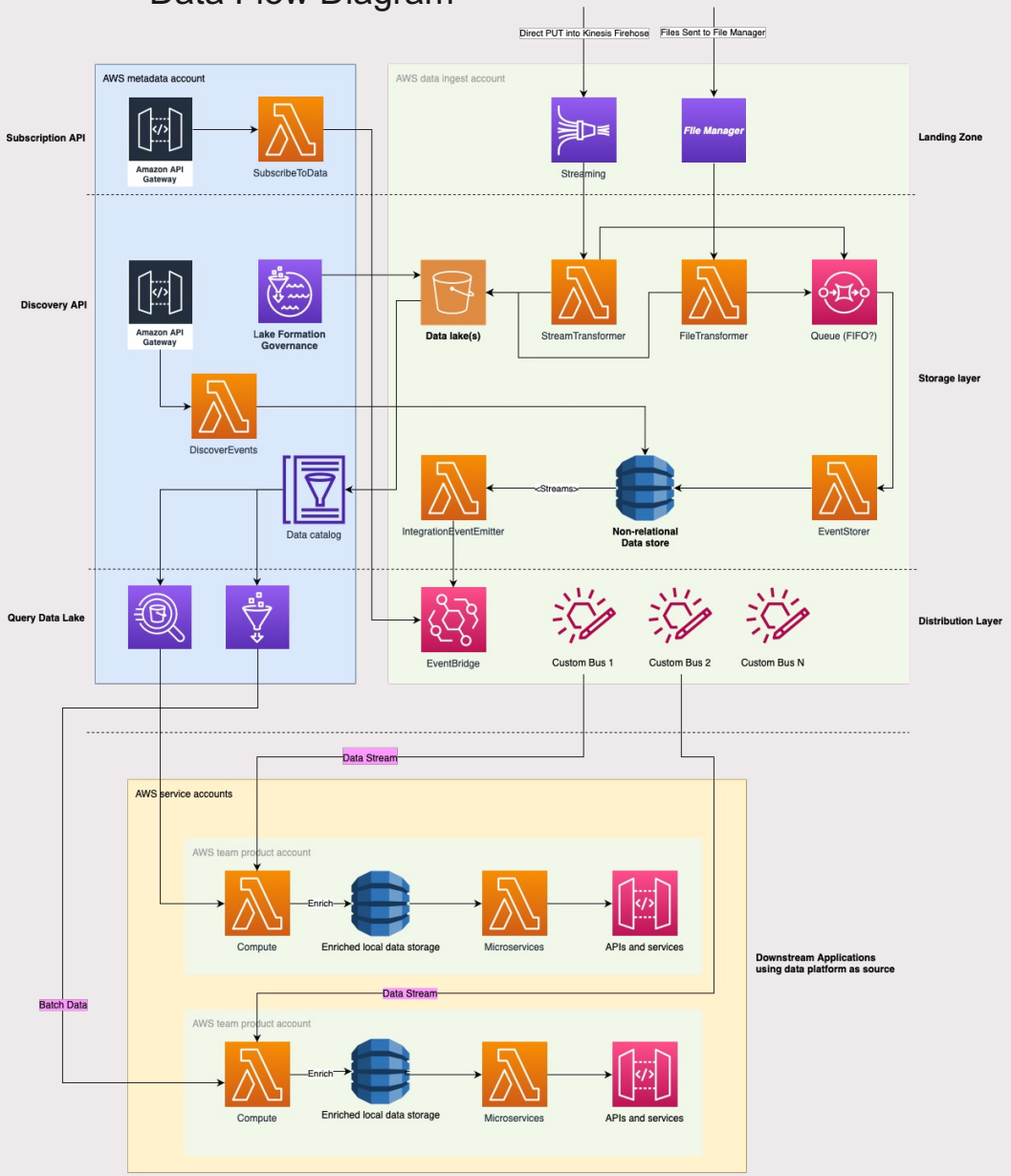


Envelope Format

```
{
  "metadata": {
    // Required
    "id": "abc123-f29f-278",
    "timestamp": "1566833470262",
    "timestampHuman": "2022-08-15T09:55:26Z",
    "requestTimeEpoch": "1566833470100",
    "user": "",
    "resource": "/createOrder",
    "source": "vista",
    "hostPlatform": "aws",
    "owner": "polestar",
    "eventType": "IntegrationEvent",
    // Optional
    "version": 1,
    "region": "eu-north-1",
    "jurisdiction": "eu",
    "tags": [""],
    "dataSensitivity": "public"
  },
  "data": {
    "YourData": "Goes here"
  }
}
```

Polestar Data Platform

Data Flow Diagram



What will we achieve?

- A staging area to store and operate variety of data
- Polestar will own the data in its landscape
- Highly suited platform for the current way of software design
- Highly decoupled, cloud native and modular software
- Light weight; minimal inertia to get off the ground
- Pay-as-you-go model
- No lock in and commitments
- Contains reusable pattens
- Self-service oriented and discoverable
- Sustainable solution

What more can be done?

- Standard support for multiple input sources
- Privacy module to scramble, encrypt or mask sensitive data
- Provide reusable patterns to consume data from the platfrom
- Enable data lineage capabilities
- Provide standard BI reports
- Standard access patterns for power users i.e. data scientists, data engineers etc.

Open source version- https://github.com/mikaelvesavuori/example-aws-stream-data-to-events

Thank You! Questions?

aws

COMMUNITY DAY

PARTNERZY

PLATINUM

N-iX

GOLD

REPLY

PHILIP MORRIS INTERNATIONAL IT HUB

SILVER

ASTEK Polska

TechData A TD SYNNEX Company

TAMESHI