

Feature Store

COE.Efficiency

1. **WHY** - Ideal FS
2. **WHAT** - FSs in market
3. **HOW** - Levis FS v1.0
4. Demo

WHY

Ideal FS

Ideal FS - DS headache

1. **Data**

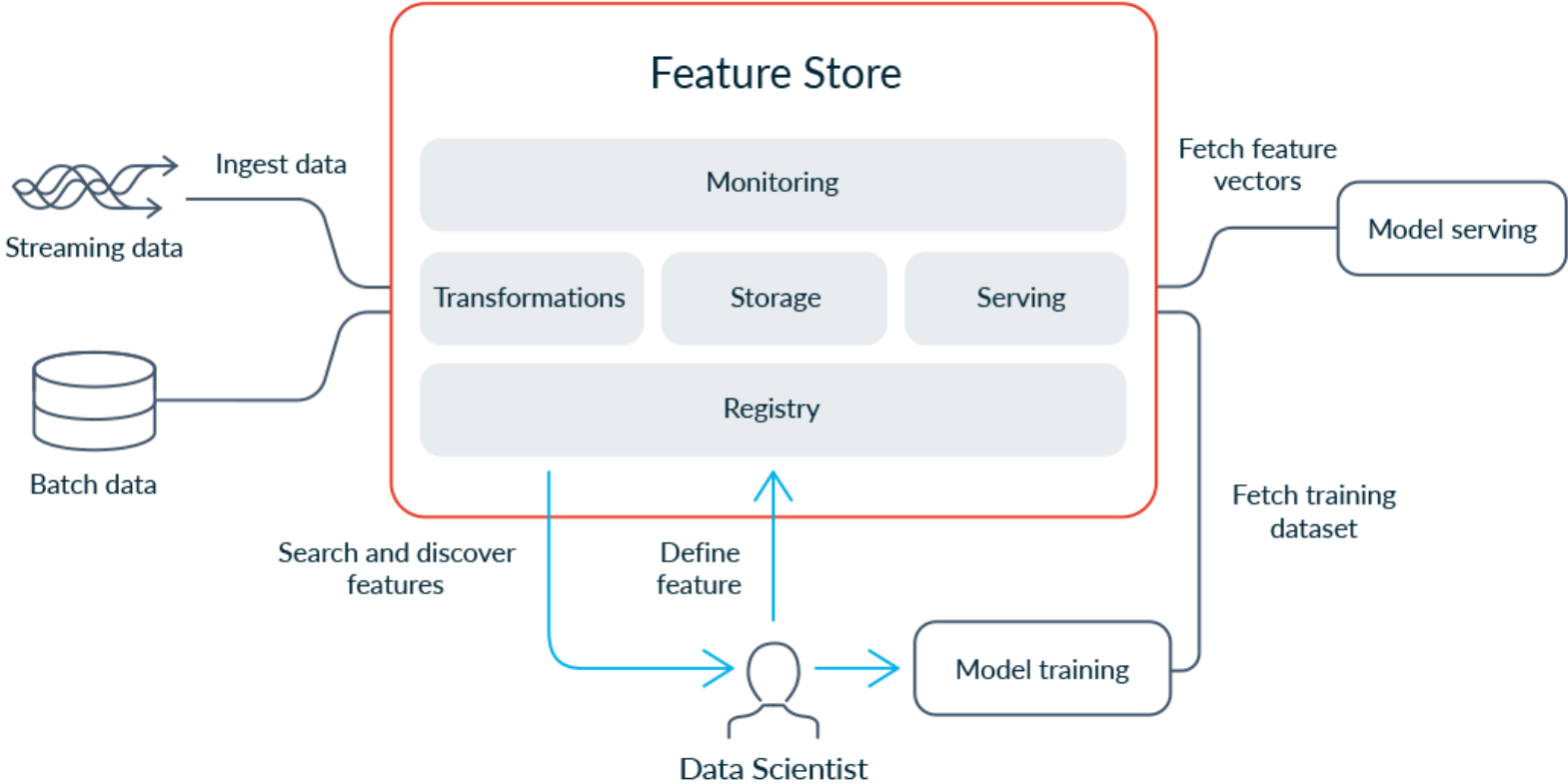
- Search -> Access -> Understand -> DQ

2. **Online/Offline Skew**

- POC vs Prod

3. **Data freshness**

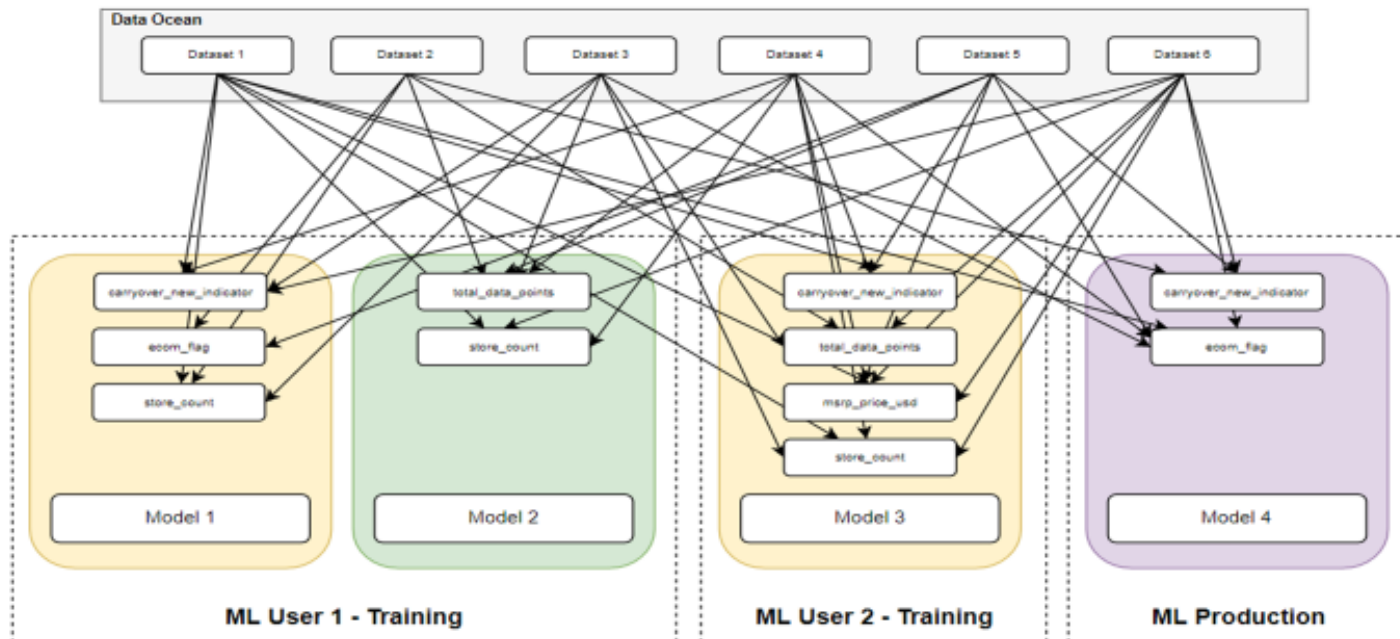
Ideal FS



Ideal FS - Bottleneck

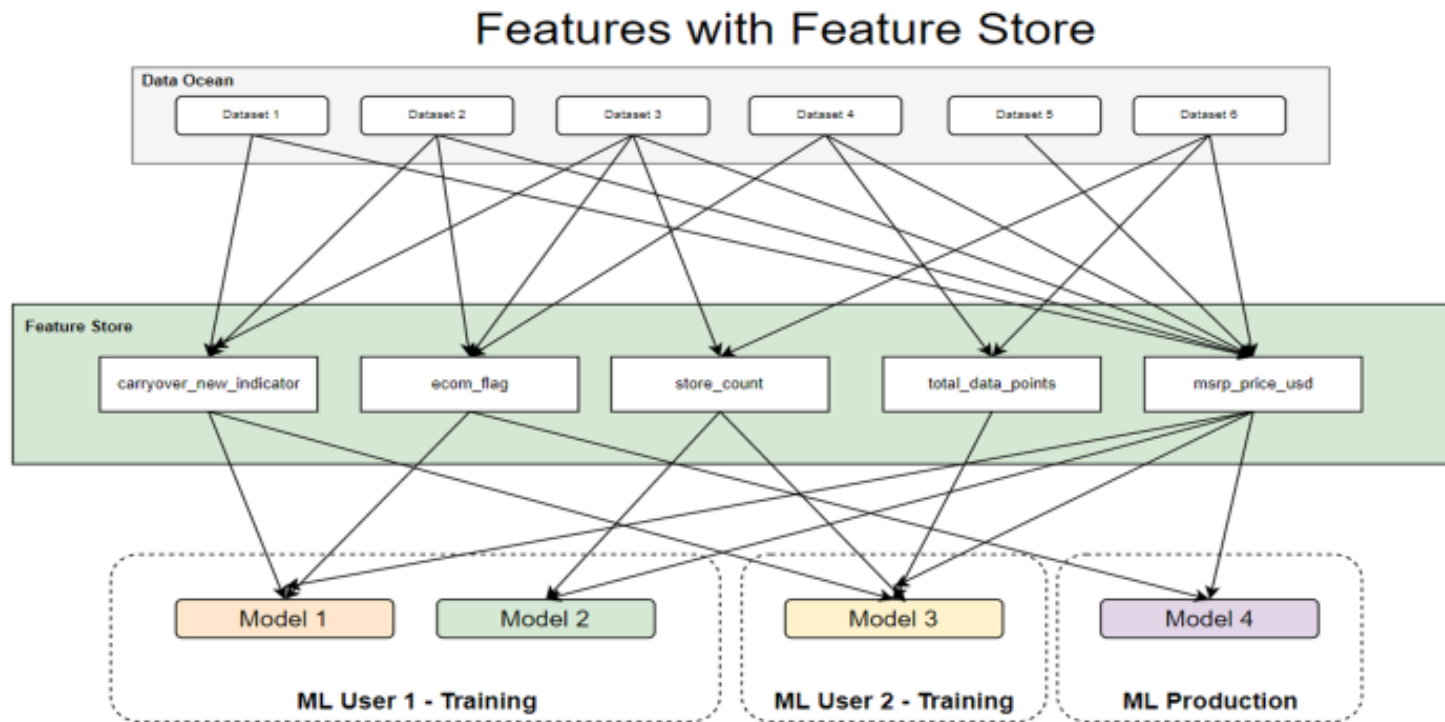
Why a Feature Store ?

Features **without** Feature Store



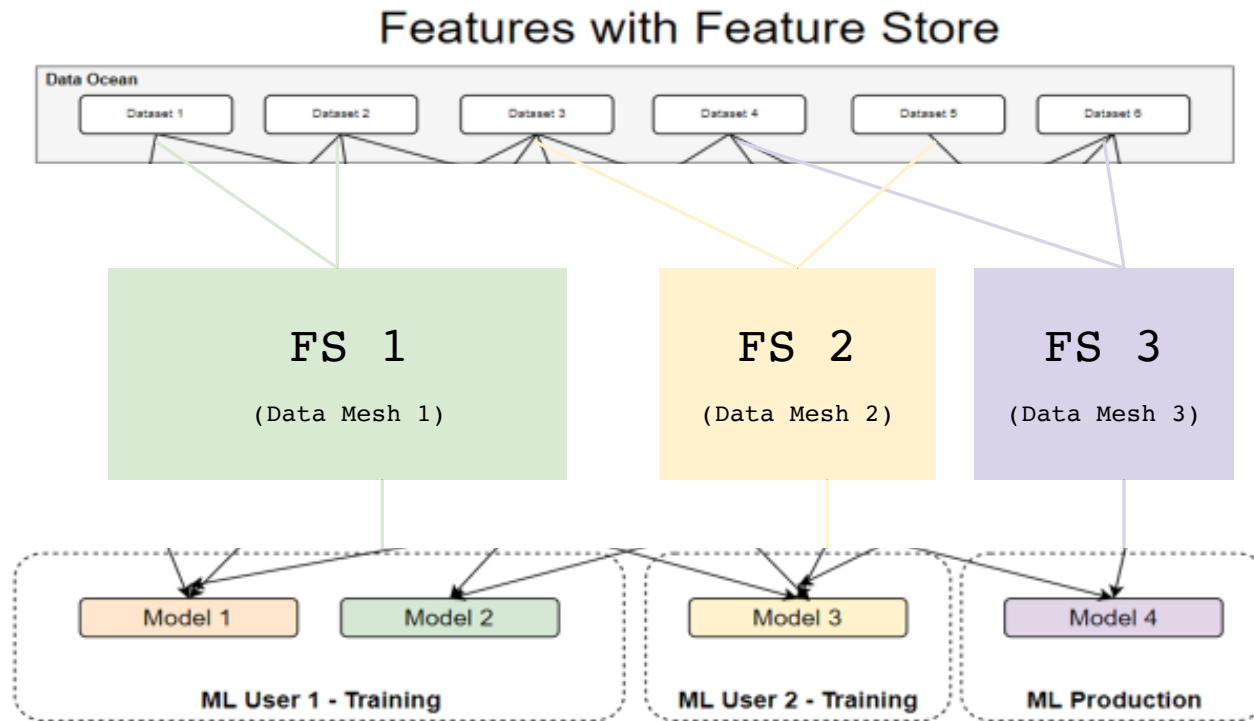
Ideal FS - Lighter, Faster, Stronger

Why a Feature Store ?



Ideal FS - Data Mesh

Why a Feature Store ?

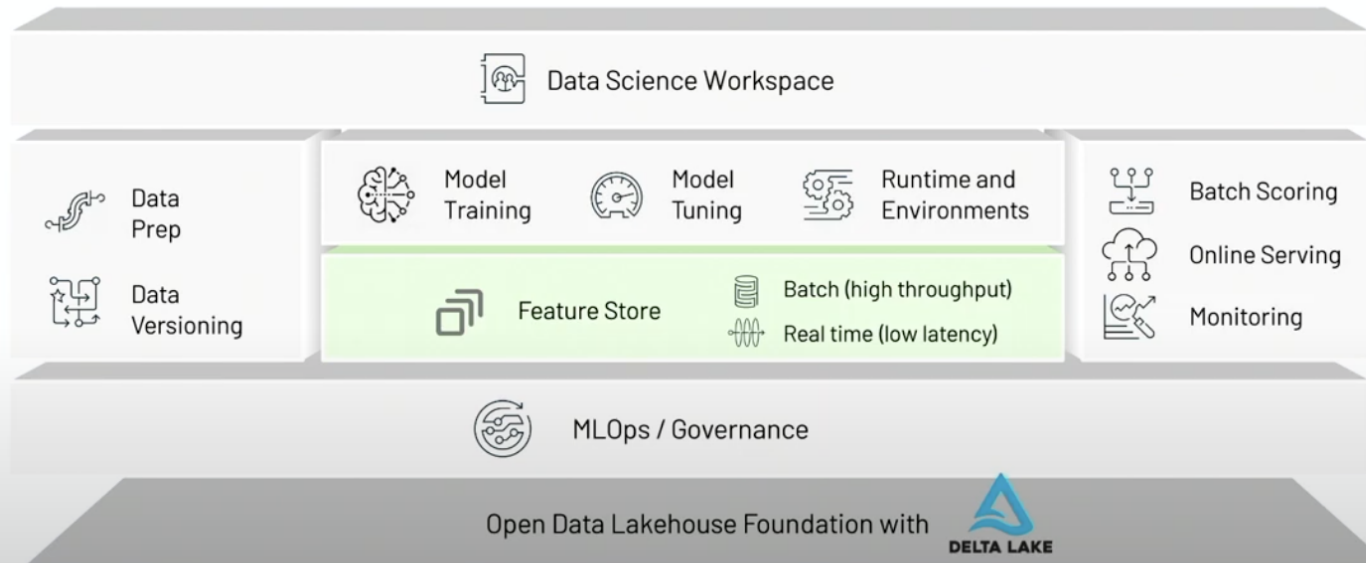


WHAT FSs in market

Ideal FS - Databricks FS

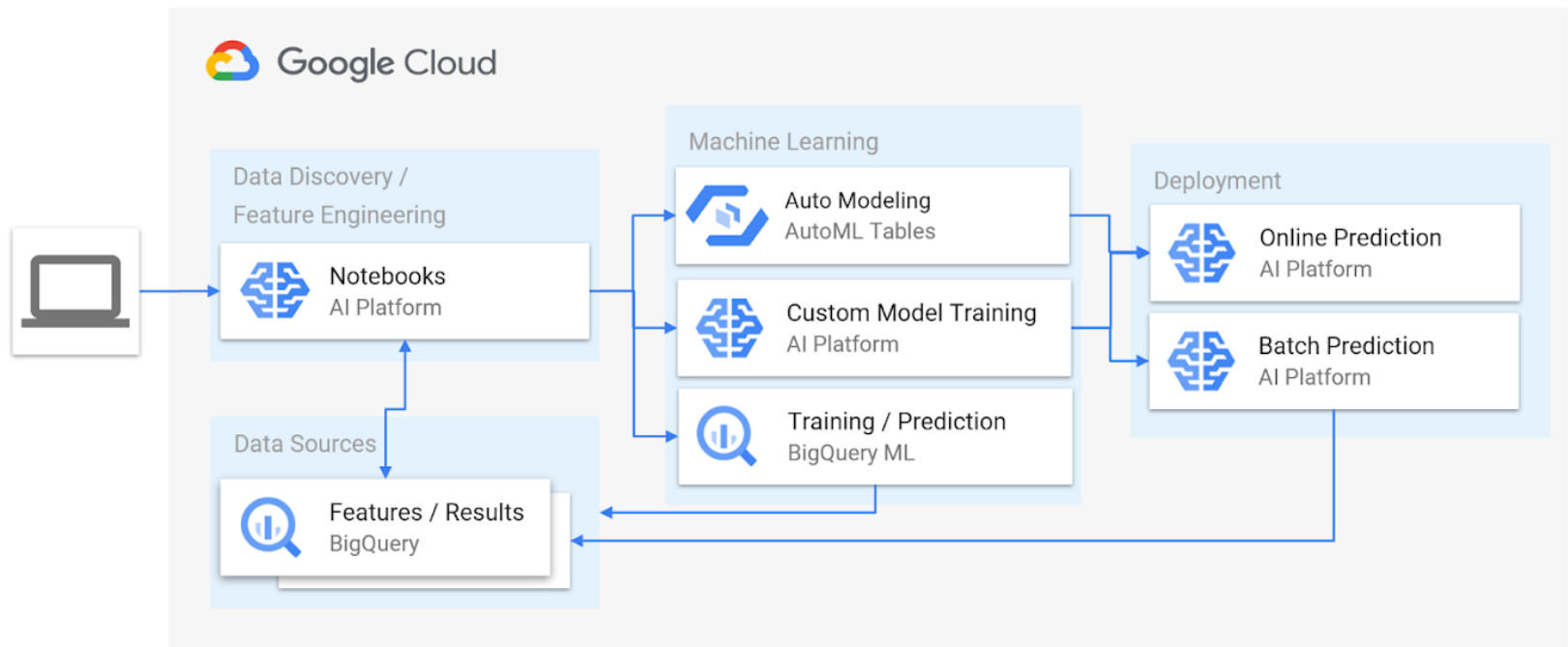
Databricks Feature Store

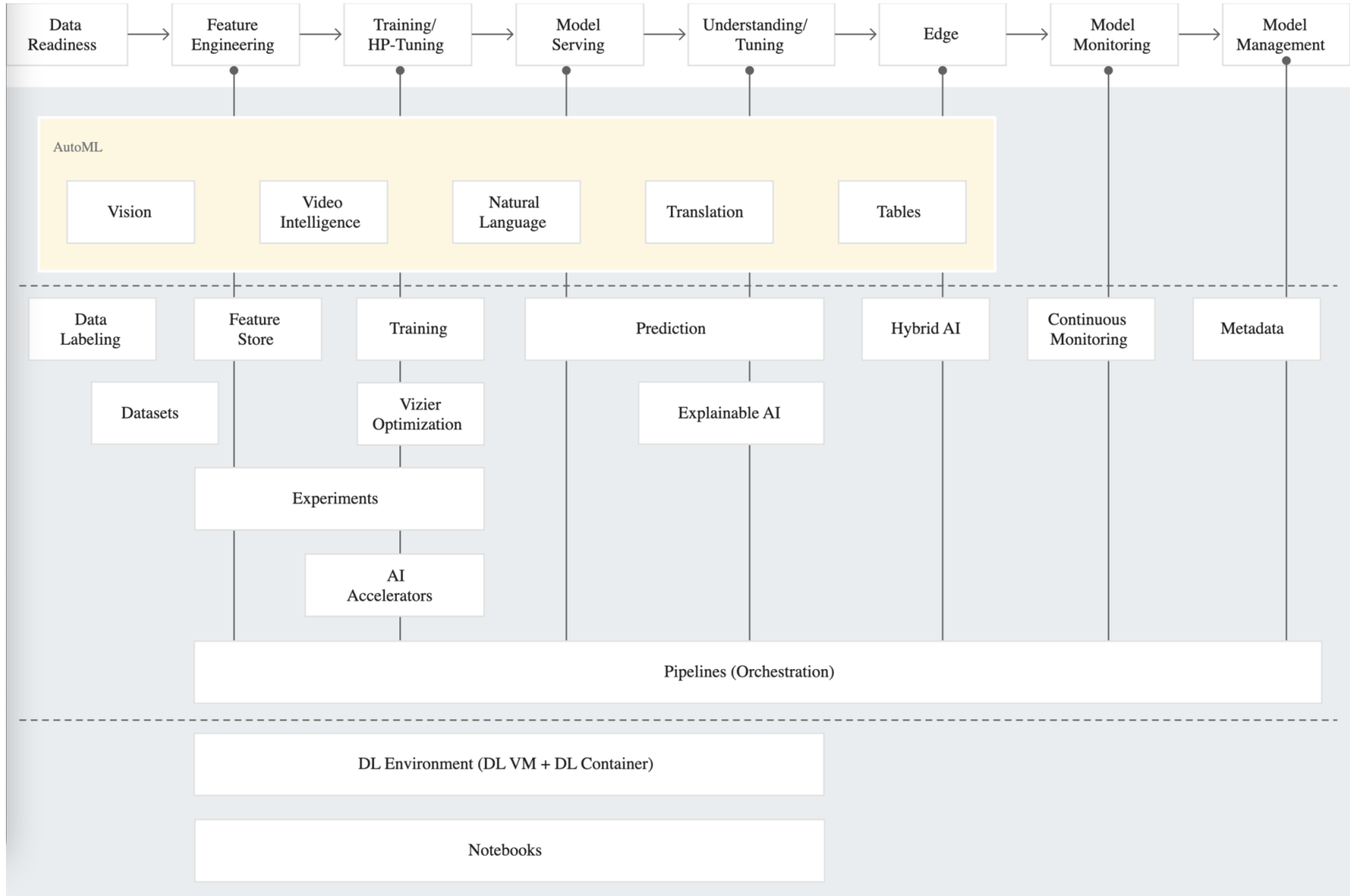
A feature store built on top of the data lakehouse.



Ideal FS - GCP Vertex AI

Effective Machine Learning iteration Architecture



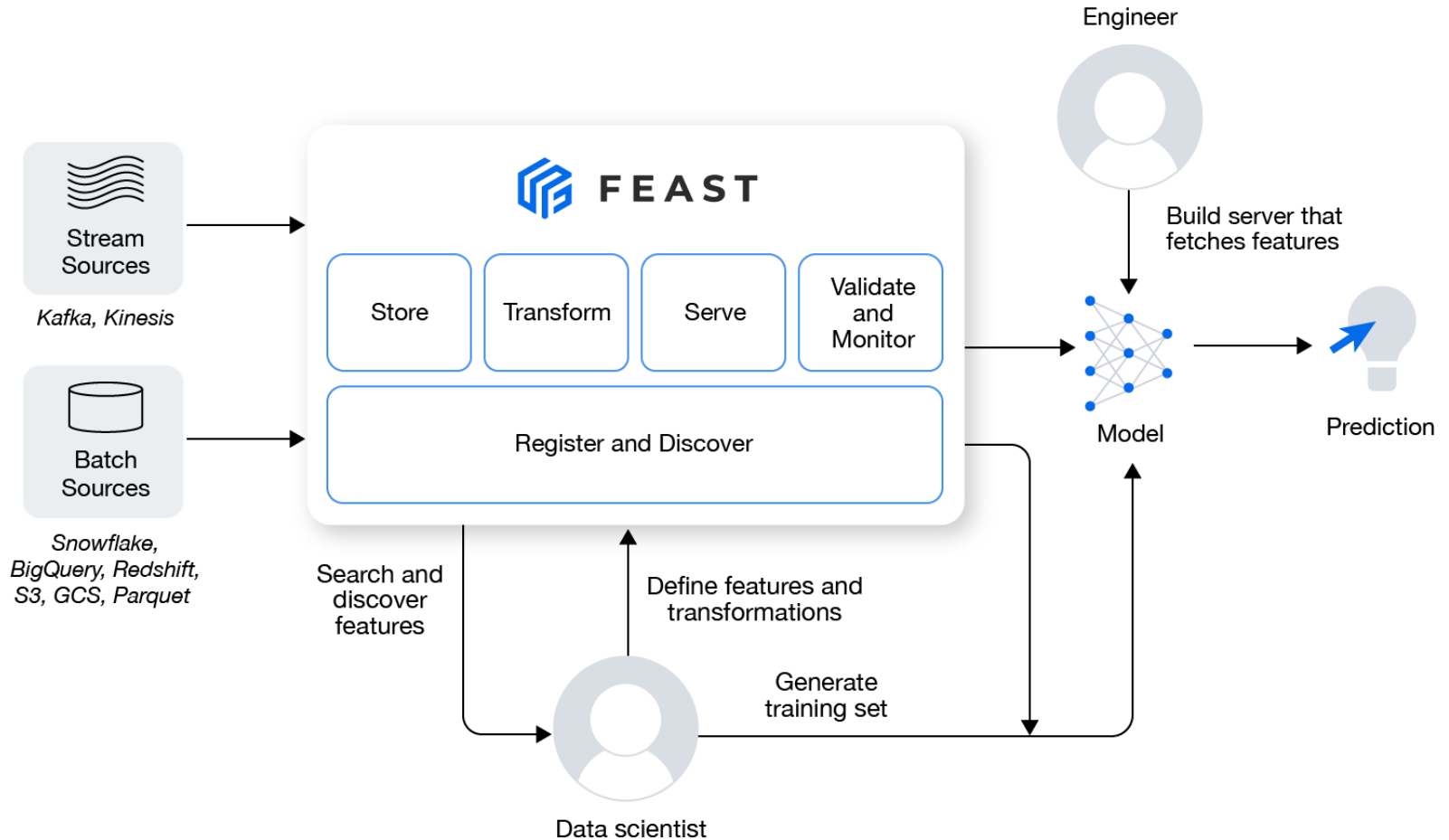


Ideal FS - GCP Vertex AI

The screenshot displays the Google Cloud Platform interface for Vertex AI. The top navigation bar includes the Google Cloud Platform logo, the user account 'levi-dataeng-d', and a search bar for products and resources. The left sidebar lists various Vertex AI features: Dashboard, Datasets (highlighted), Features, Labeling tasks, Workbench, Pipelines, Training, Experiments, Models, Endpoints, Batch predictions, and Metadata. The main content area shows the 'Isa_sellout_weekly' dataset in the 'ANALYZE' tab. It provides dataset information such as creation time (Nov 08, 2021 11:03 PM), format (CSV), and location. A summary section indicates 58 total columns and 2,773,282 total rows. Below this, a table of general statistics is shown, generated on Nov 08, 2021 at 11:22 PM. The table has columns for 'Column name', 'Missing % (count)', and 'Distinct values'. The data rows are as follows:

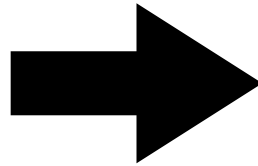
Column name	Missing % (count)	Distinct values
ad_count_of_sizes	-	84
ad_number_of_doors	-	629
affiliate_code	-	1
affiliate_name	-	1
aur_usd	-	1156464

Ideal FS - FEAST



Ideal FS

- FS Databricks
- Vertex AI - GCP
- FEAST
- etc.

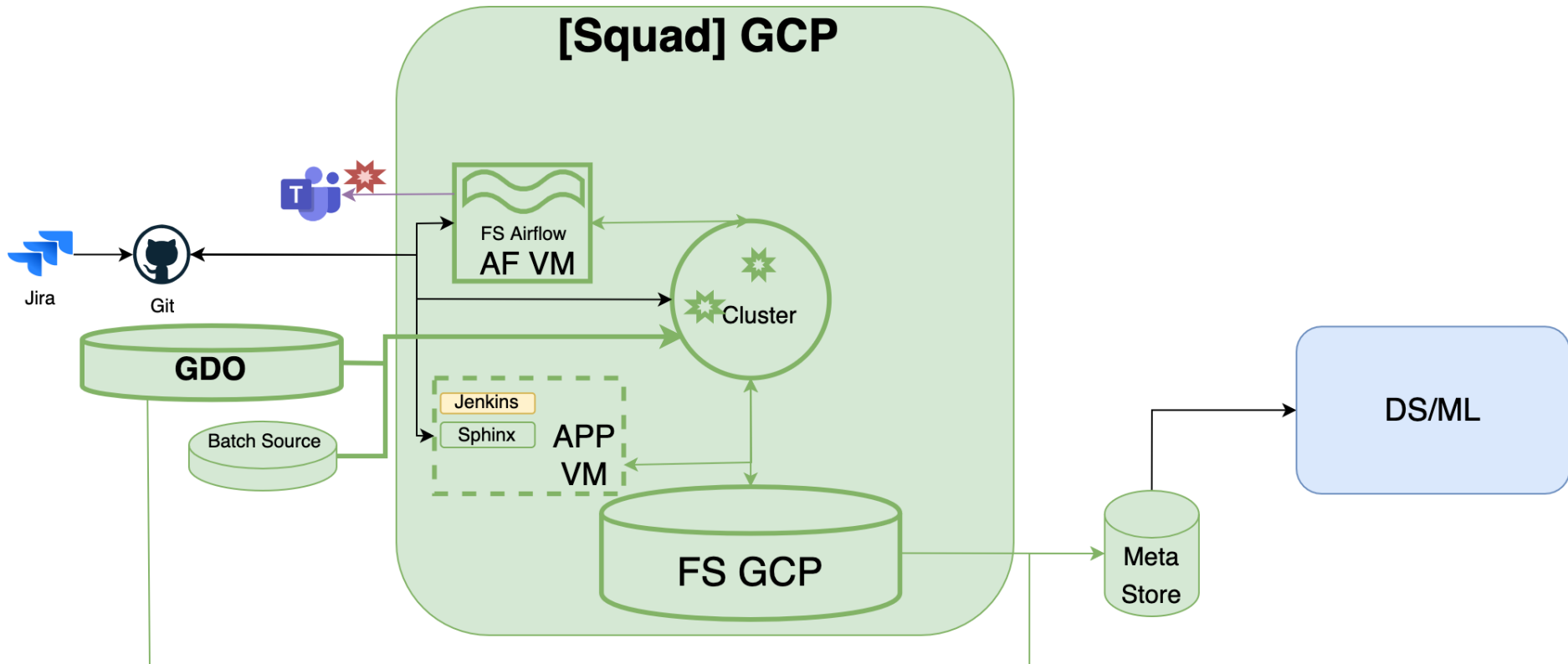


- Cost
- Dependency
- Customize
- Data Silos

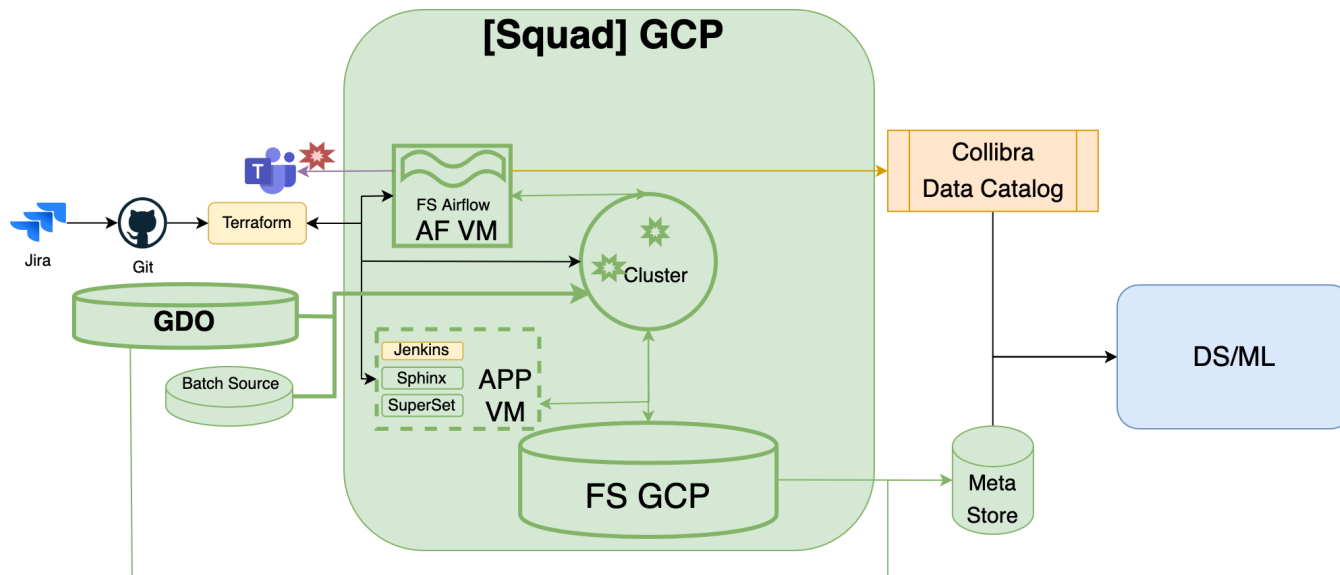


HOW FS v1.0

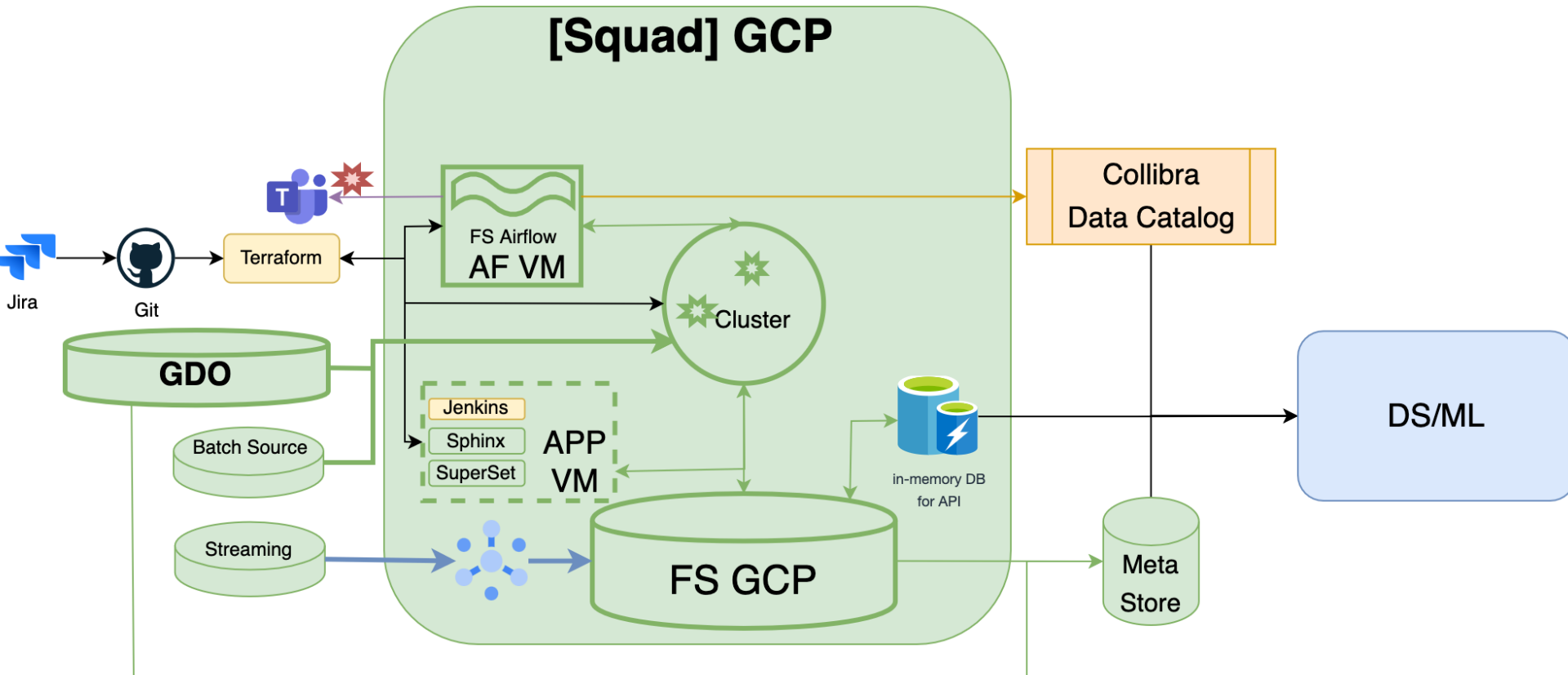
FS Ecosystem V1.0



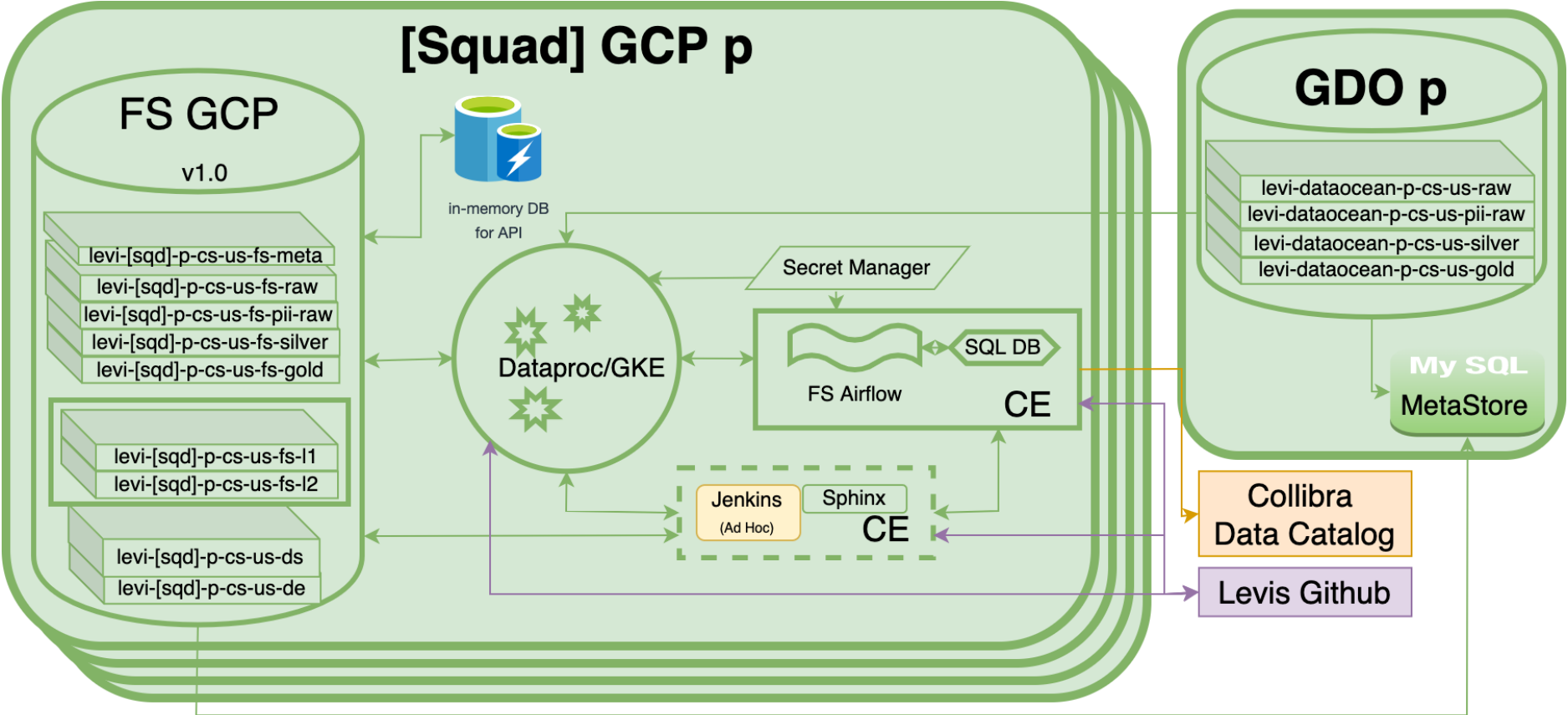
FS Ecosystem V2.0



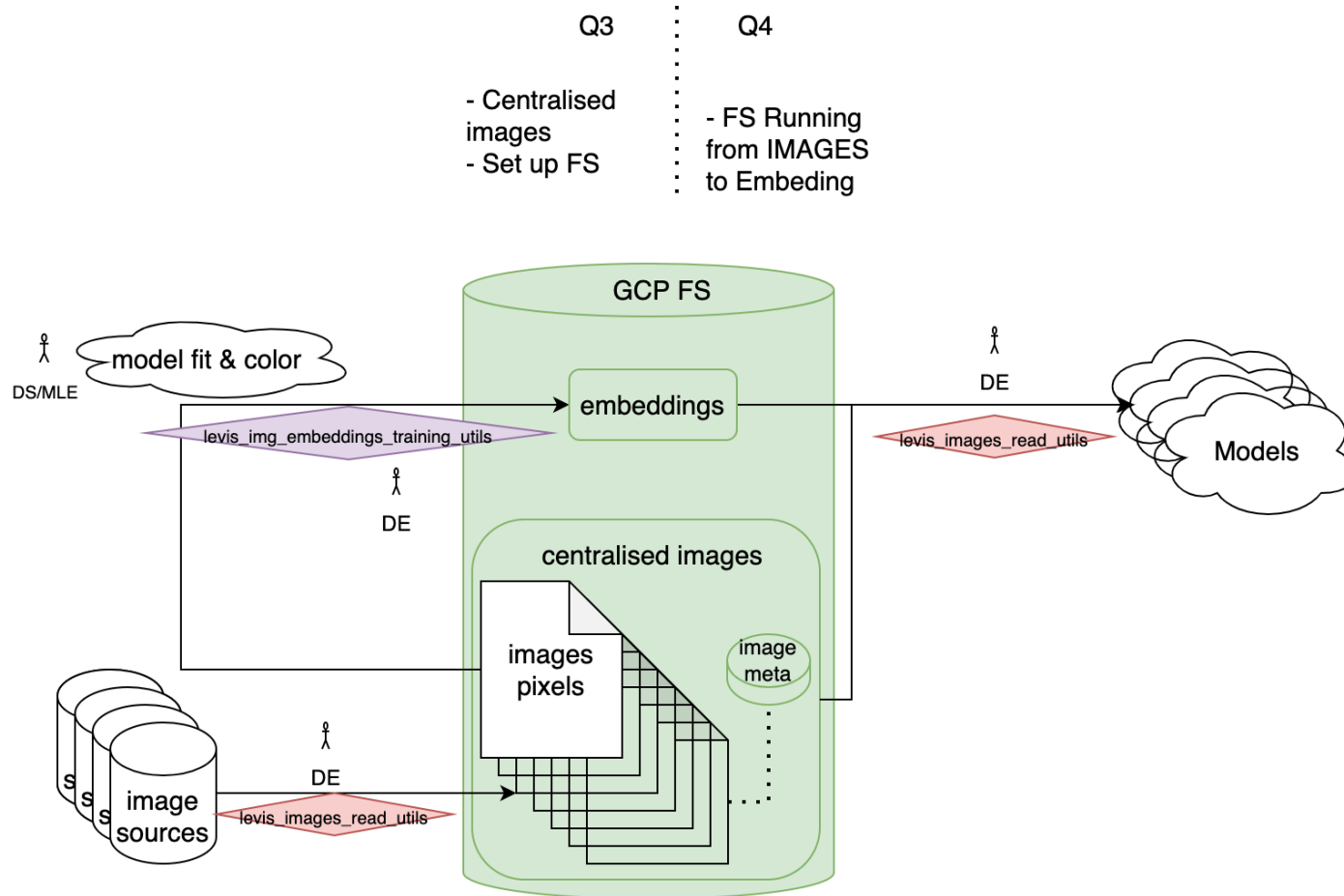
FS Ecosystem 3.0



Cloud Archi

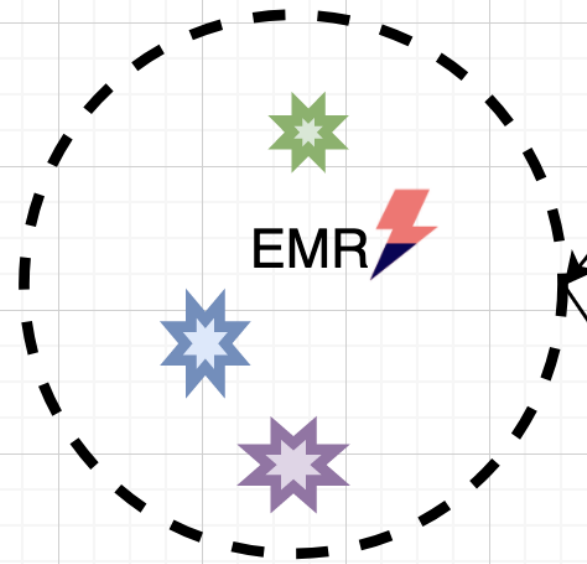
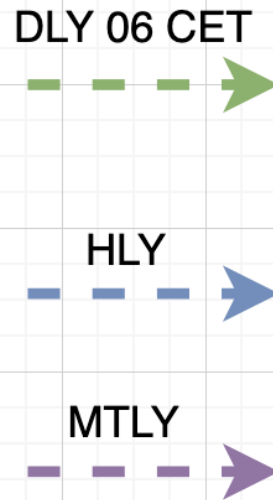
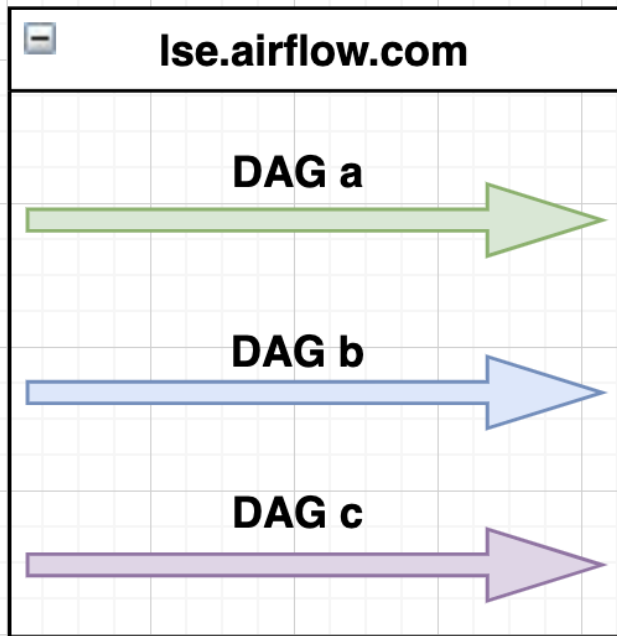


Cloud Archi (Common Lib)



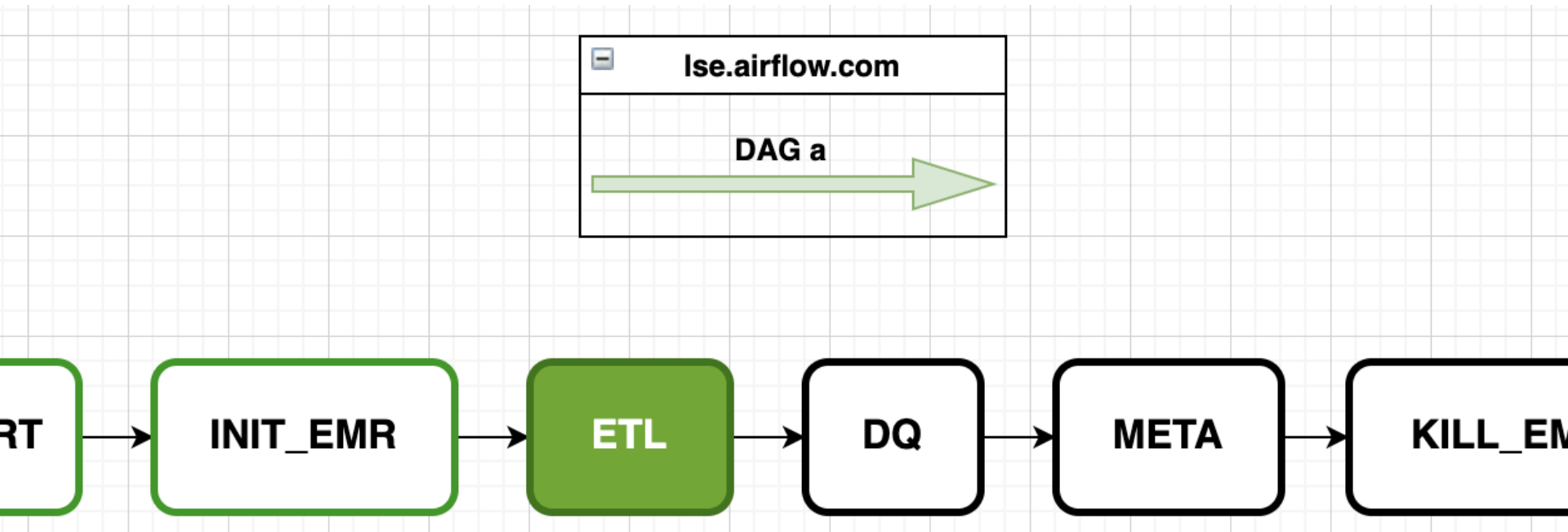
Airflow

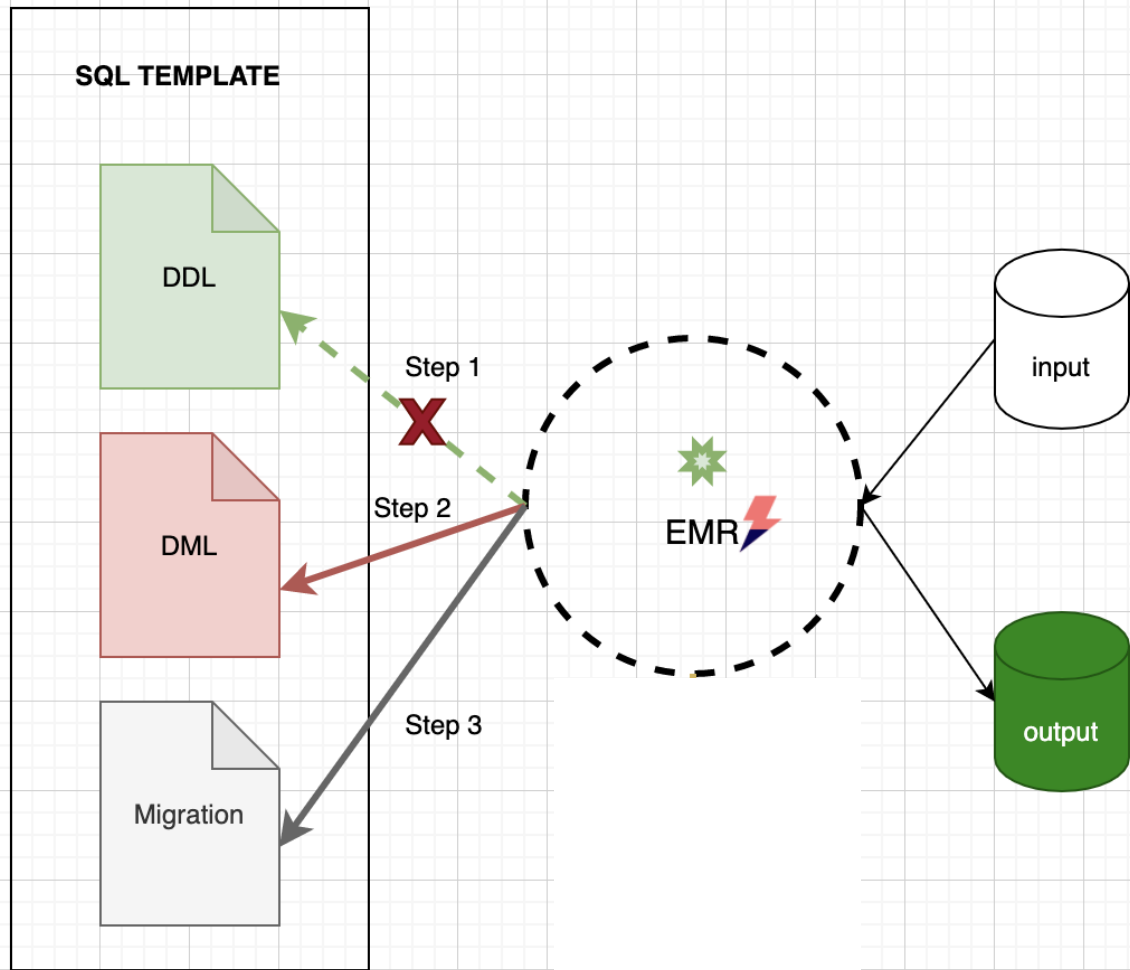
Orchestration & Automation



Airflow

Orchestration & Automation





DEMO