KubeCon CloudNativeCon India 2026

#KubeCon #CloudNativeCon

Feast as a Cloud Native Data System

Bridging Batch Processing and Online Serving on Kubernetes

Nikhil Kathole,
Red Hat

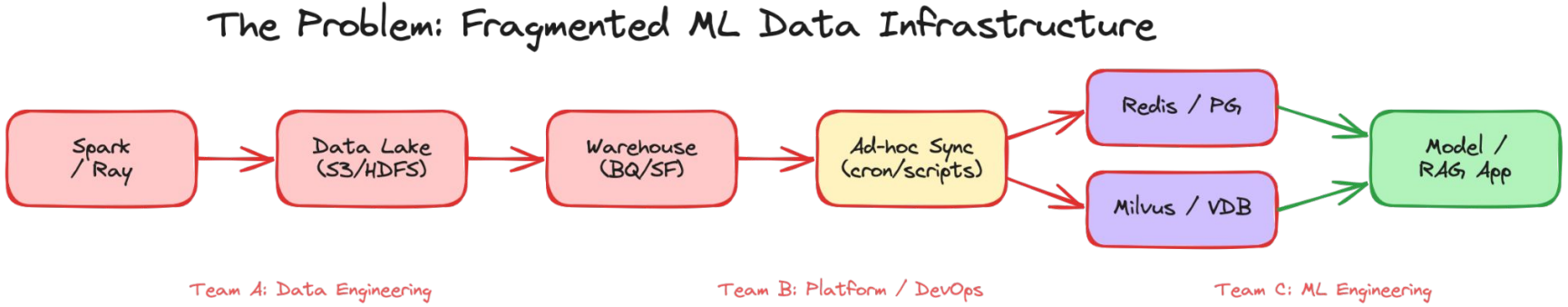
The Problem

Your Lake Has the Data.

Your Model Needs It in Milliseconds.

Who owns the path between?

Why Production ML breaks



Training-Serving Skew

Offline transforms ≠ online transforms.Silent correctness failures in production.

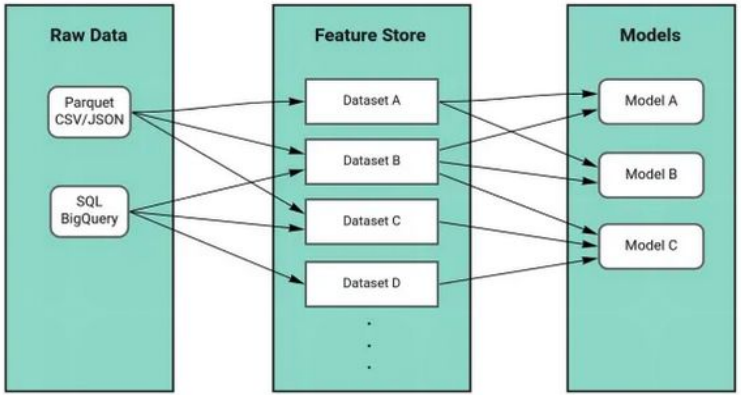
No Feature Reuse

Teams rebuild the same features for every new model. Zero discoverability.

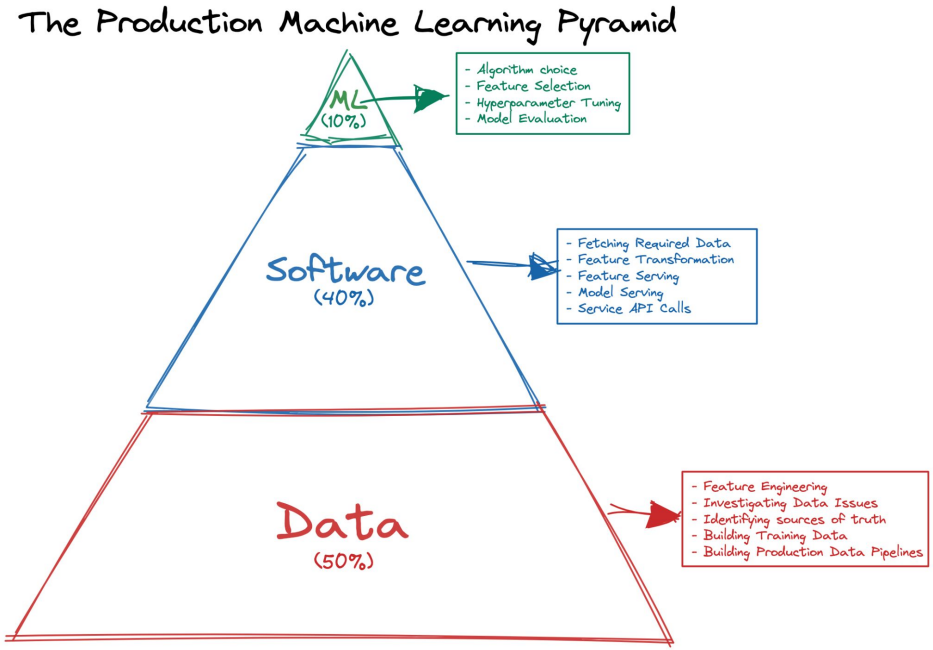
Duct-Tape Infrastructure

Data processing + lake + warehouse + cache + vector DB - stitched together by convention.

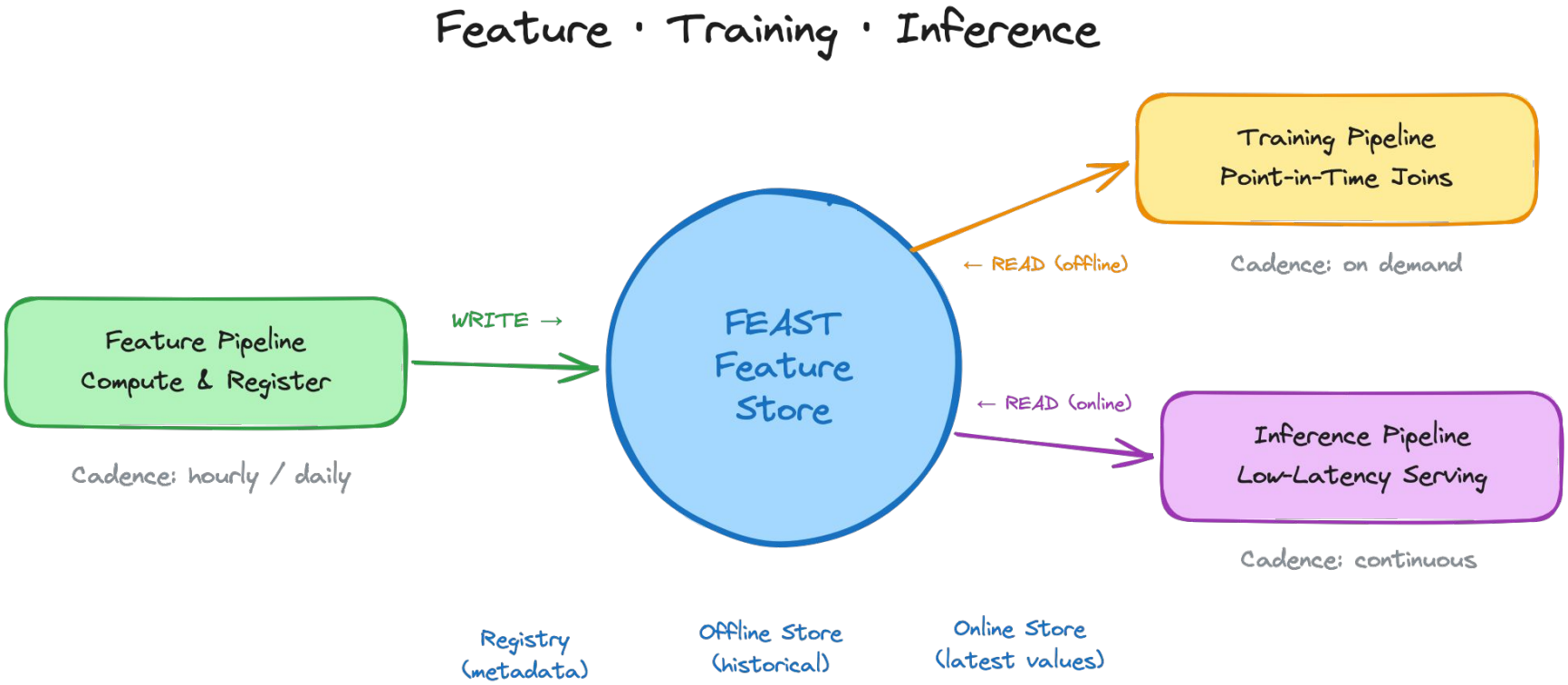
What is Feature Store ?



Feature Store as a platform layer serving features to the machine learning models



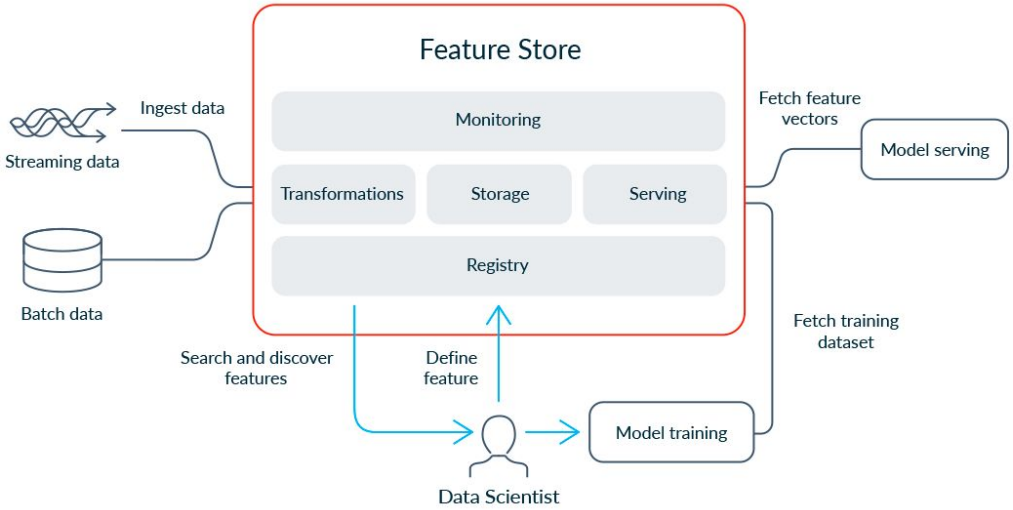
Define once. Train offline. Serve online.



Feast: The Missing Data Layer

KubeCon CloudNativeCon India 2026

- **Central Registry** - discover, version, and govern features
- **Dual storage** - offline for training, online for inference
- **Point-in-time correct joins** - no data leakage
- Materialization syncs batch compute to **low-latency serve**
- Build once, **reuse** across models and teams

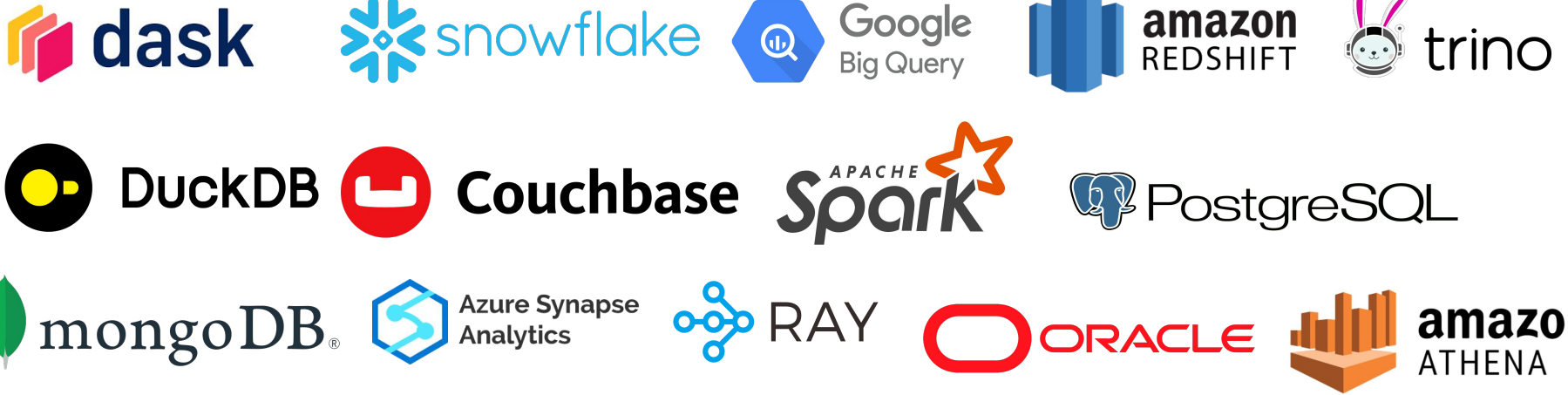


FEAST

Lakehouse/Warehouse Native

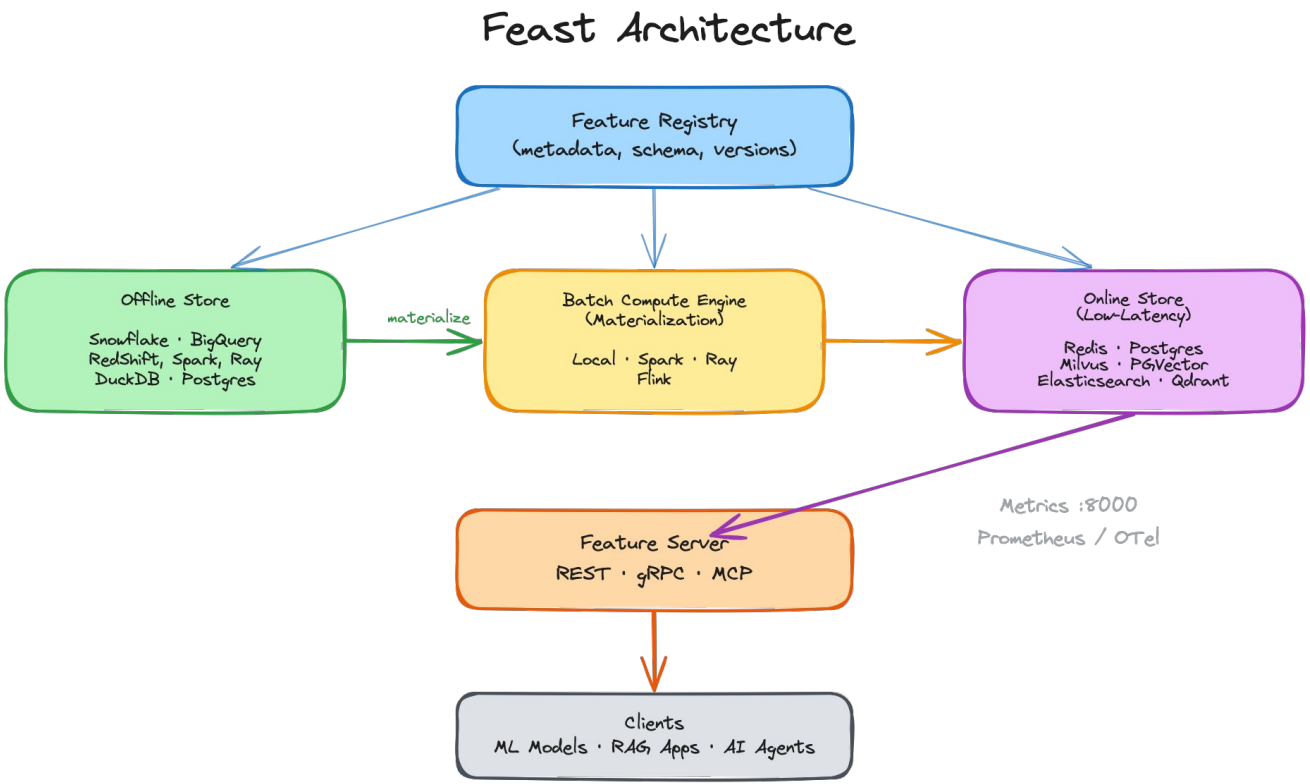
Do I need to migrate my data into Feast?

Feast integrates with what you already run.

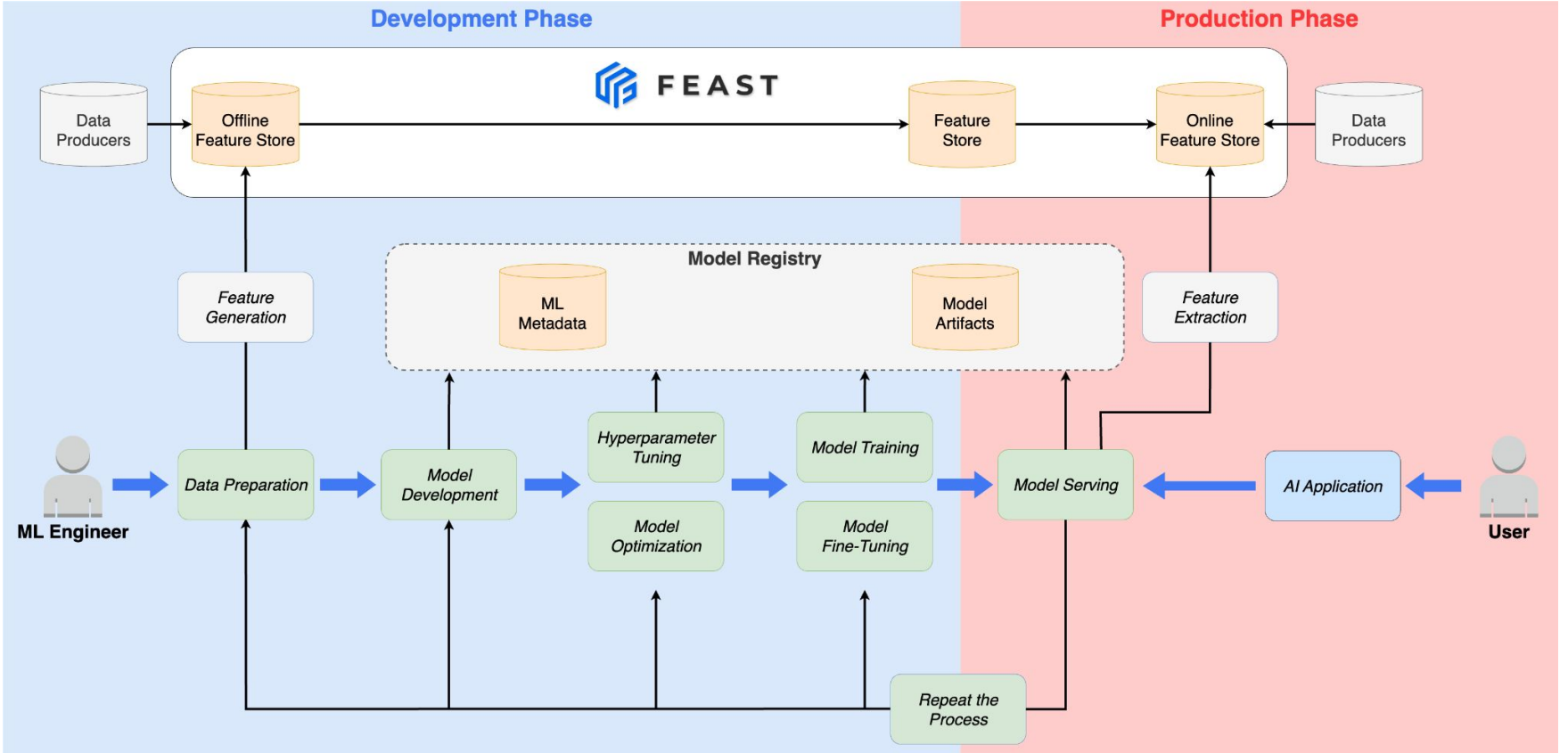


Feast Architecture

KubeCon CloudNativeCon India 2026



Feast in AI/ML Lifecycl



```python
from feast import FeatureStore

# Initialize the feature store
store = FeatureStore(repo_path="feature_repo")

# Get features for training
training_df = store.get_historical_features(
    entity_df=training_entities,
    features=[
        "customer_stats:daily_transactions",
        "customer_stats:lifetime_value",
        "product_features:price"
    ]
).to_df()

# Get online features for inference
features = store.get_online_features(
    features=[
        "customer_stats:daily_transactions",
        "customer_stats:lifetime_value",
        "product_features:price"
    ],
    entity_rows=[{"customer_id": "C123", "product_id": "P456"}]
).to_dict()

# Retrieve your documents using vector similarity search for RAG
features = store.retrieve_online_documents(
    features=[
        "corpus:document_id",
        "corpus:chunk_id",
        "corpus:chunk_text",
        "corpus:chunk_embedding",
    ],
    query="What is the biggest city in the USA?"
).to_dict()
```

KubeCon CloudNativeCon India 2026

Run it on Kubernetes

One CR | Full Stack | Day-2 Ready

Batch to Online on Kubernetes

KubeCon CloudNativeCon India 2026

Feast Operator: One CR → Full Kubernetes Stack

FeatureStore CR

```
apiVersion: feast.dev/v1

spec:
  feastProject: credit_scoring
  feastProjectDir:
    git:
      url: github.com/ … /features
      ref: main
  services:
    onlineStore:
      server: { metrics: true }
      persistence:
        store:
          type: postgres
    offlineStore:
      server: {}
    registry:
      local:
        server: {}
    ui: {}
  cronJob:
    schedule: "0 2 * * *"
  batchEngine:
    configMapRef: spark-engine
  openlineage:
    enabled: true
```

reconcile → **Feast Operator** (Go / Operator SDK) → creates

Kubernetes Cluster

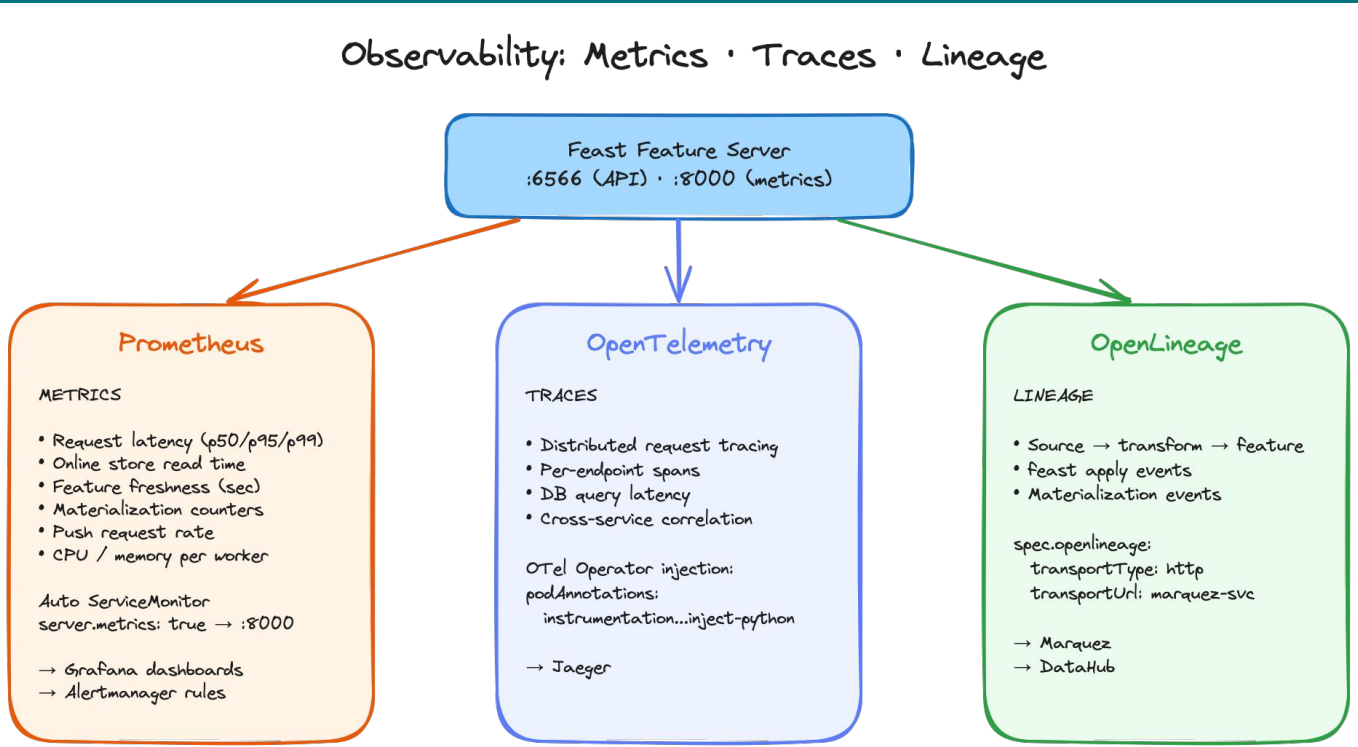
- Deployment: online-server
- Services (:6566, :8815)
- Deployment: offline-server
- ServiceMonitor (:8000)
- Deployment: registry
- CronJob (materialize)
- Deployment: feast-ui
- ConfigMap (client + YAML)
- RBAC
- Init Container (git clone)
- feature_store.yaml (operator-owned)

External (pre-provisioned): Postgres · Redis · Milvus

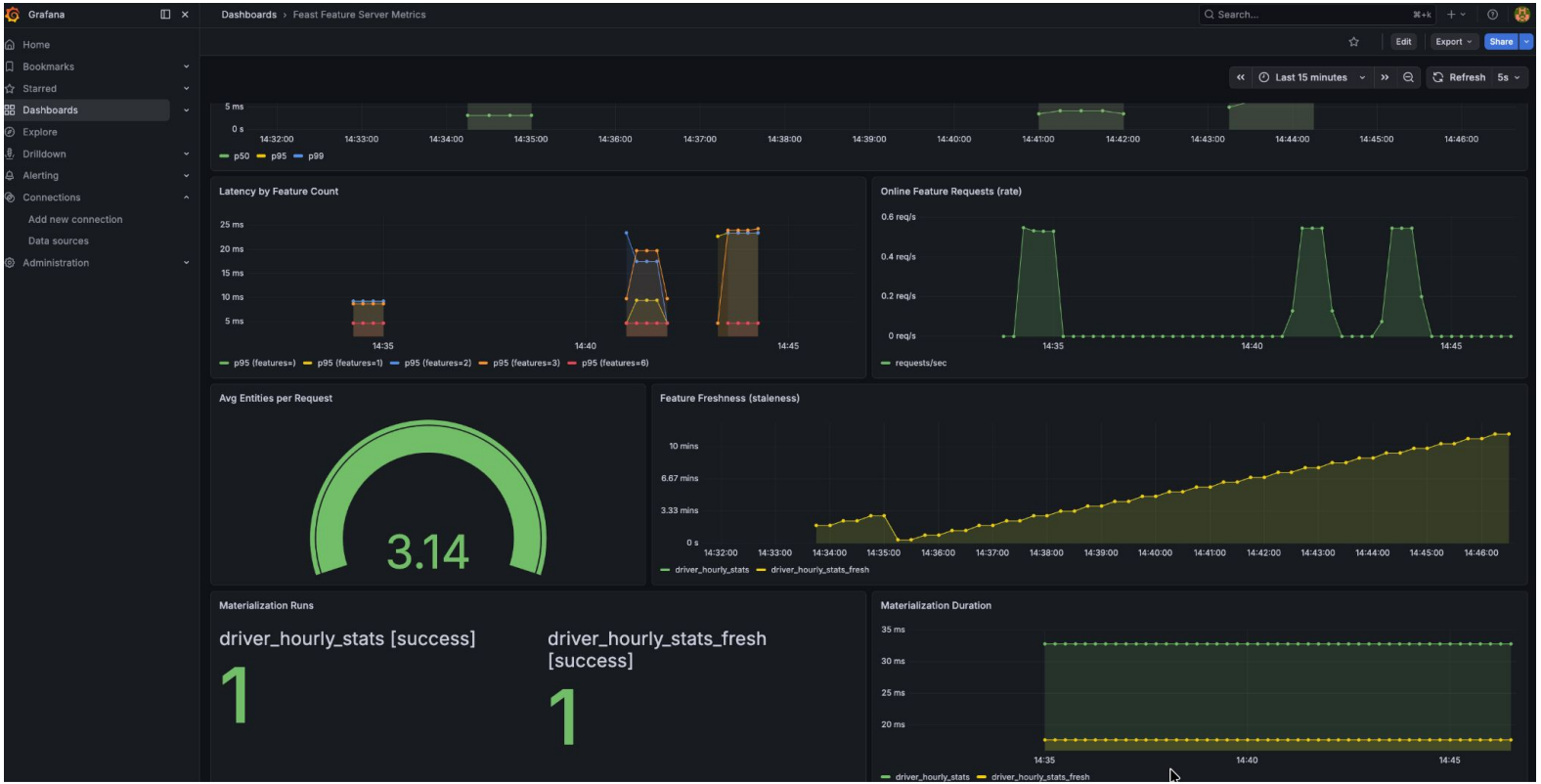
Operator wires Secrets → feature_store.yaml


KubeCon
CloudNativeCon
India 2026
Observe Everything
Metrics | Traces | Lineage

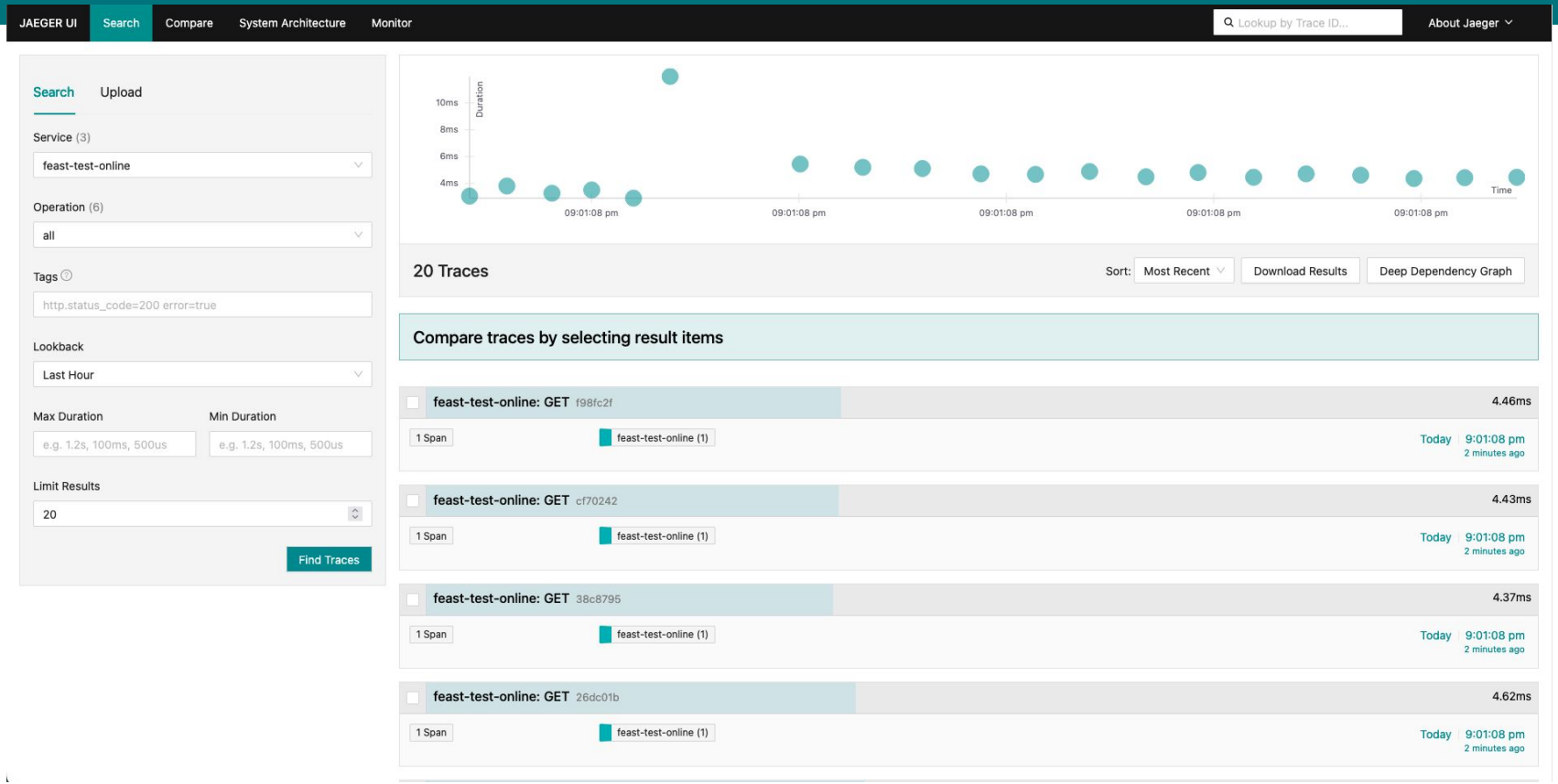
If You Cannot See It, You Cannot Scale It



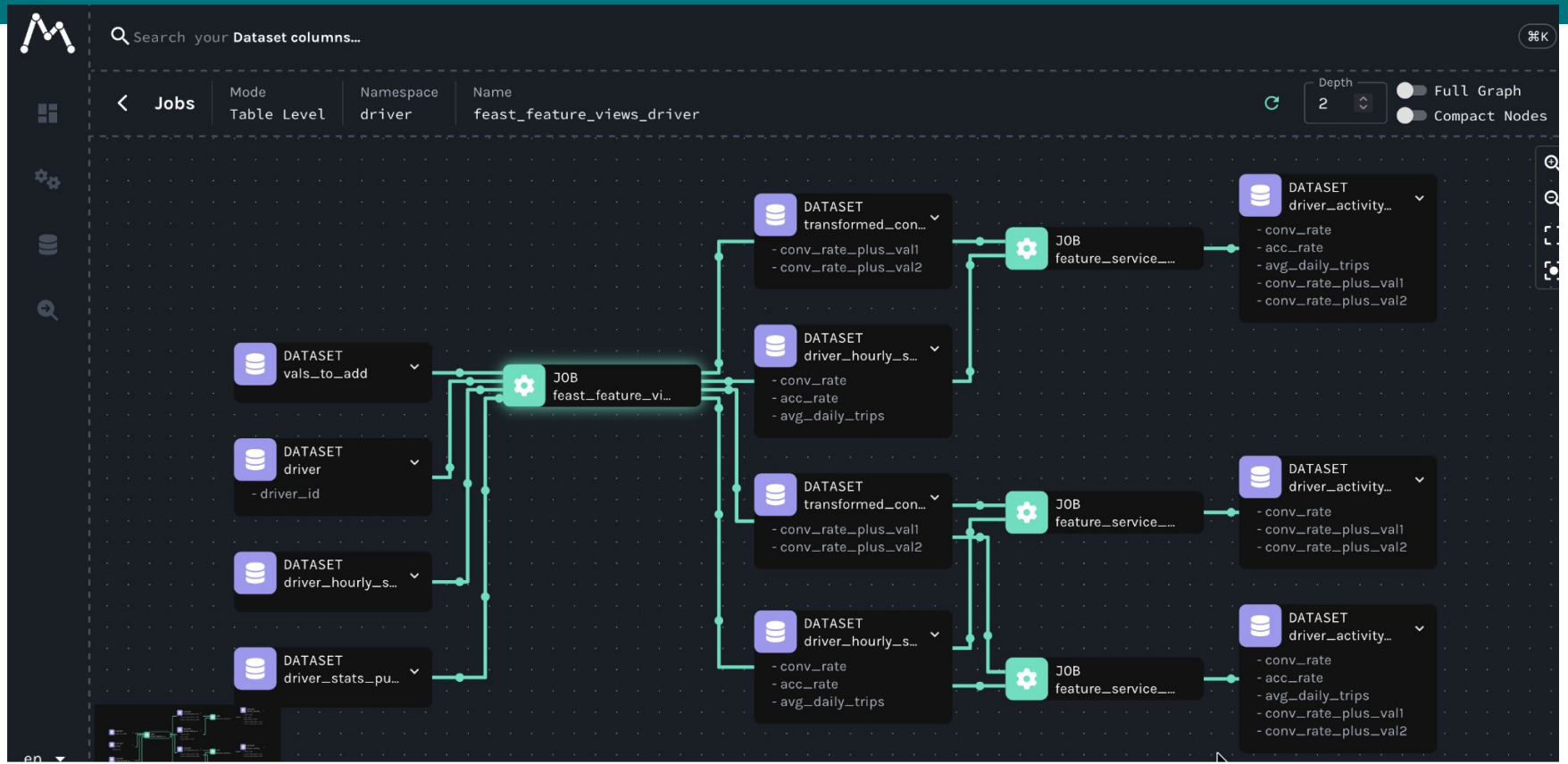
Metrics


Grafana
Dashboards › Feast Feature Server Metrics
Search...
Edit
Export
Share
Last 15 minutes
Refresh 5s
Home
Bookmarks
Starred
Dashboards
Explore
Drilldown
Alerting
Connections
Add new connection
Data sources
Administration
5 ms
0 s
14:32:00 14:33:00 14:34:00 14:35:00 14:36:00 14:37:00 14:38:00 14:39:00 14:40:00 14:41:00 14:42:00 14:43:00 14:44:00 14:45:00 14:46:00
p50 p95 p99
Latency by Feature Count
25 ms
20 ms
15 ms
10 ms
5 ms
14:35 14:40 14:45
p95 (features=) p95 (features=1) p95 (features=2) p95 (features=3) p95 (features=6)
Online Feature Requests (rate)
0.6 req/s
0.4 req/s
0.2 req/s
0 req/s
14:35 14:40 14:45
requests/sec
Avg Entities per Request
3.14
Feature Freshness (staleness)
10 mins
6.67 mins
3.33 mins
0 s
14:32:00 14:33:00 14:34:00 14:35:00 14:36:00 14:37:00 14:38:00 14:39:00 14:40:00 14:41:00 14:42:00 14:43:00 14:44:00 14:45:00 14:46:00
driver_hourly_stats driver_hourly_stats_fresh
Materialization Runs
driver_hourly_stats [success]
1
driver_hourly_stats_fresh [success]
1
Materialization Duration
35 ms
30 ms
25 ms
20 ms
14:35 14:40 14:45
driver_hourly_stats driver_hourly_stats_fresh

Traces


JAEGER UI
Search
Compare
System Architecture
Monitor
Lookup by Trace ID...
About Jaeger
Search
Upload
Service (3)
feast-test-online
Operation (6)
all
Tags
http.status_code=200 error=true
Lookback
Last Hour
Max Duration
e.g. 1.2s, 100ms, 500us
Min Duration
e.g. 1.2s, 100ms, 500us
Limit Results
20
Find Traces
Duration
10ms
8ms
6ms
4ms
09:01:08 pm
09:01:08 pm
09:01:08 pm
09:01:08 pm
09:01:08 pm
Time
20 Traces
Sort: Most Recent
Download Results
Deep Dependency Graph
Compare traces by selecting result items
feast-test-online: GET f98fc2f
4.46ms
1 Span
feast-test-online (1)
Today 9:01:08 pm
2 minutes ago
feast-test-online: GET cf70242
4.43ms
1 Span
feast-test-online (1)
Today 9:01:08 pm
2 minutes ago
feast-test-online: GET 38c8795
4.37ms
1 Span
feast-test-online (1)
Today 9:01:08 pm
2 minutes ago
feast-test-online: GET 26dc01b
4.62ms
1 Span
feast-test-online (1)
Today 9:01:08 pm
2 minutes ago

OpenLineage Integration



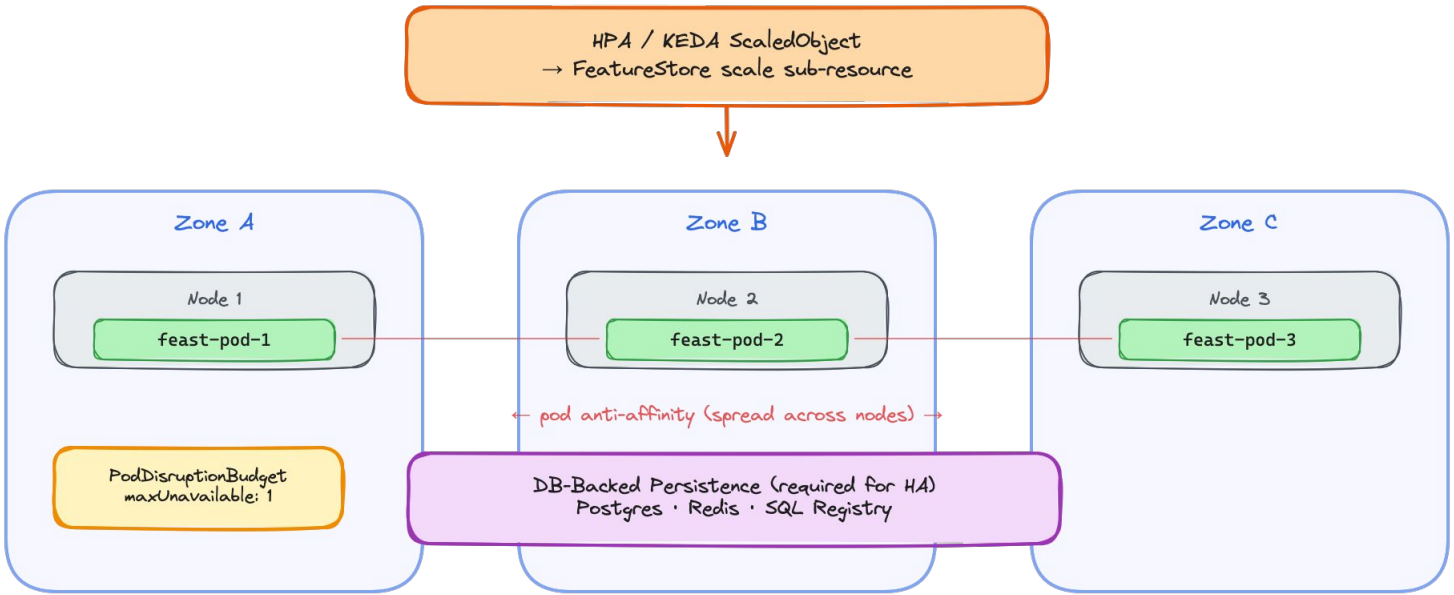
KubeCon CloudNativeCon India 2026

Scale Without Surprises

HPA | KEDA | PDB | HA

Scaling & High Availability

Scaling & High Availability

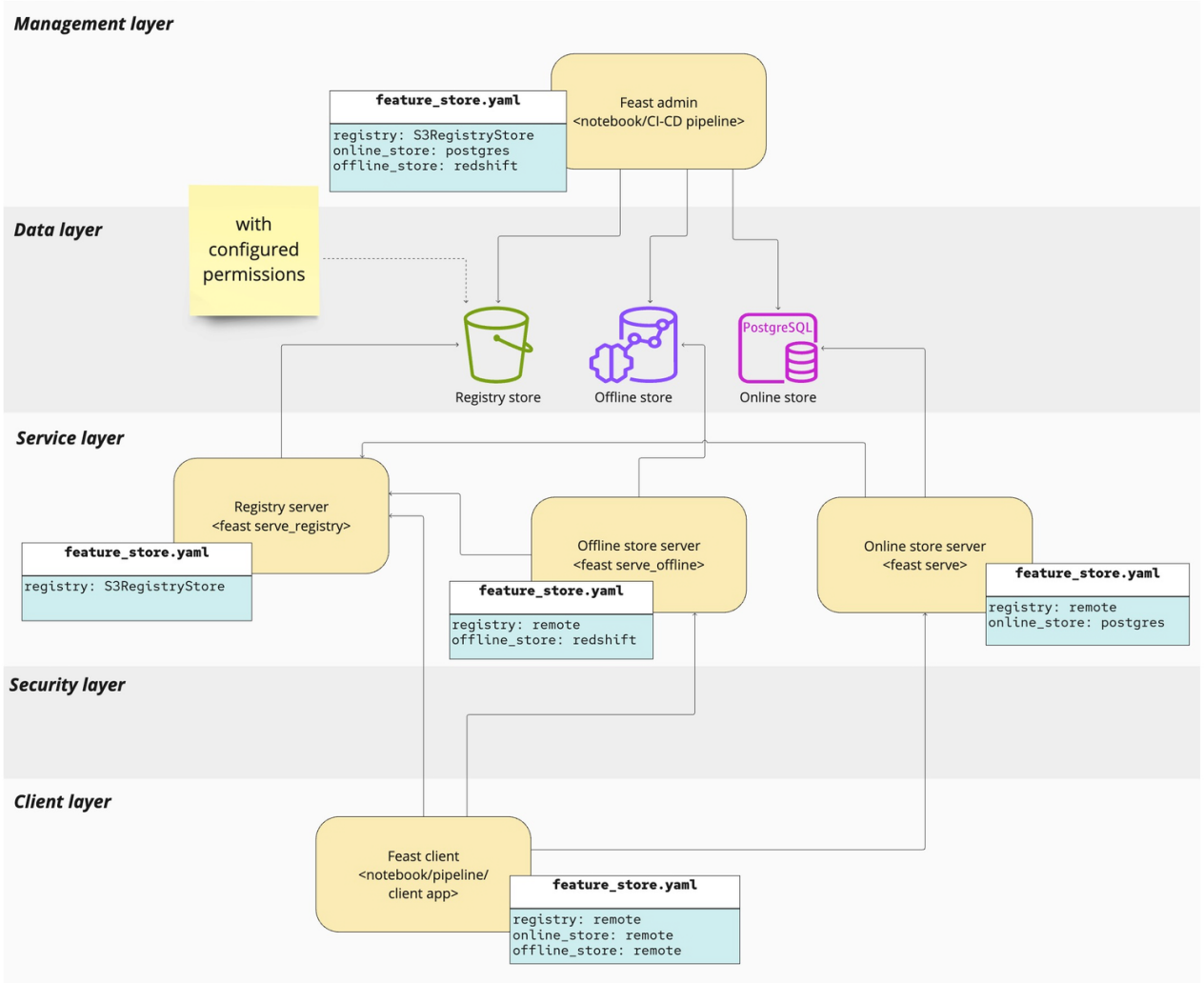


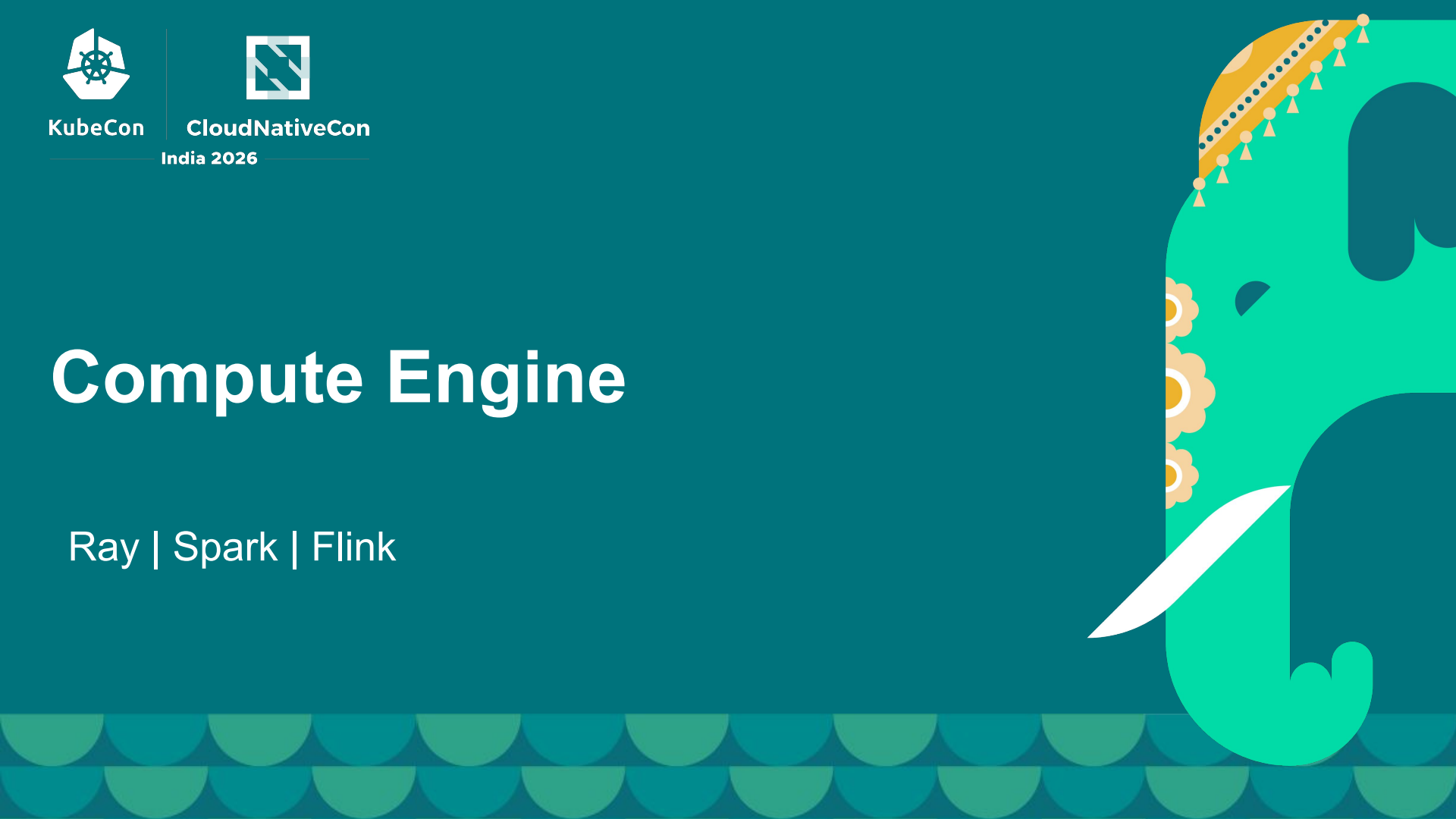
Auto-injected when replicas > 1:

- Pod anti-affinity (prefer spread across nodes)
- Zone topology spread (distribute across AZs)
- RollingUpdate strategy (zero-downtime)
- CEL validation rejects file-based stores

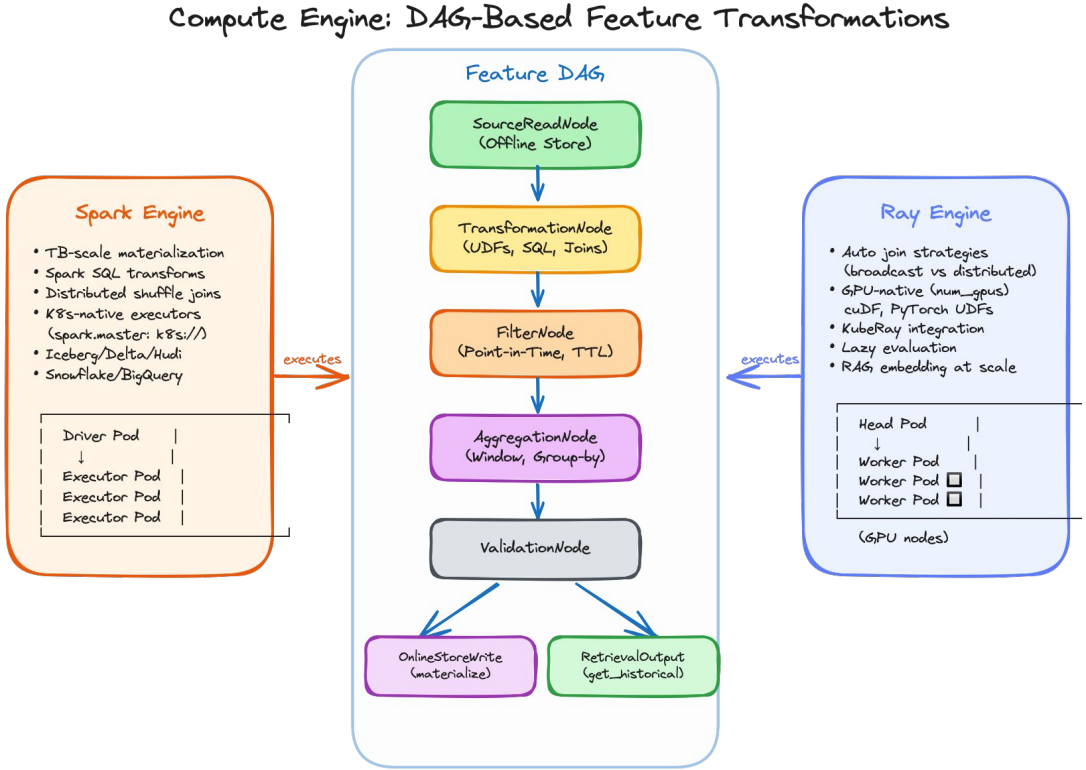
RBAC (Role based access control)

Governance | Multi-tenancy | Isolation

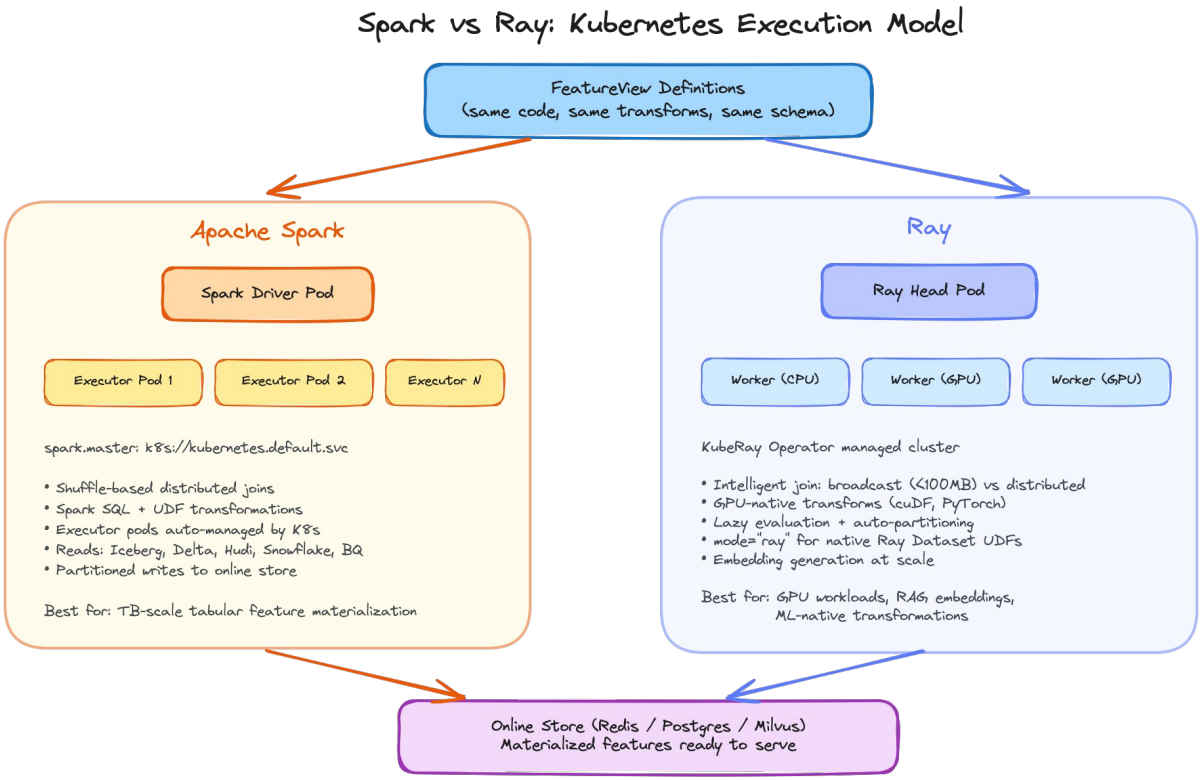

Management layer
feature_store.yaml
registry: S3RegistryStore
online_store: postgres
offline_store: redshift
Feast admin
<notebook/CI-CD pipeline>
Data layer
with configured permissions
Registry store
Offline store
PostgreSQL
Online store
Service layer
Registry server
<feast serve_registry>
feature_store.yaml
registry: S3RegistryStore
Offline store server
<feast serve_offline>
feature_store.yaml
registry: remote
offline_store: redshift
Online store server
<feast serve>
feature_store.yaml
registry: remote
online_store: postgres
Security layer
Client layer
Feast client
<notebook/pipeline/
client app>
feature_store.yaml
registry: remote
online_store: remote
offline_store: remote


KubeCon
CloudNativeCon
India 2026
Compute Engine
Ray | Spark | Flink

Compute Engines: DAG Execution



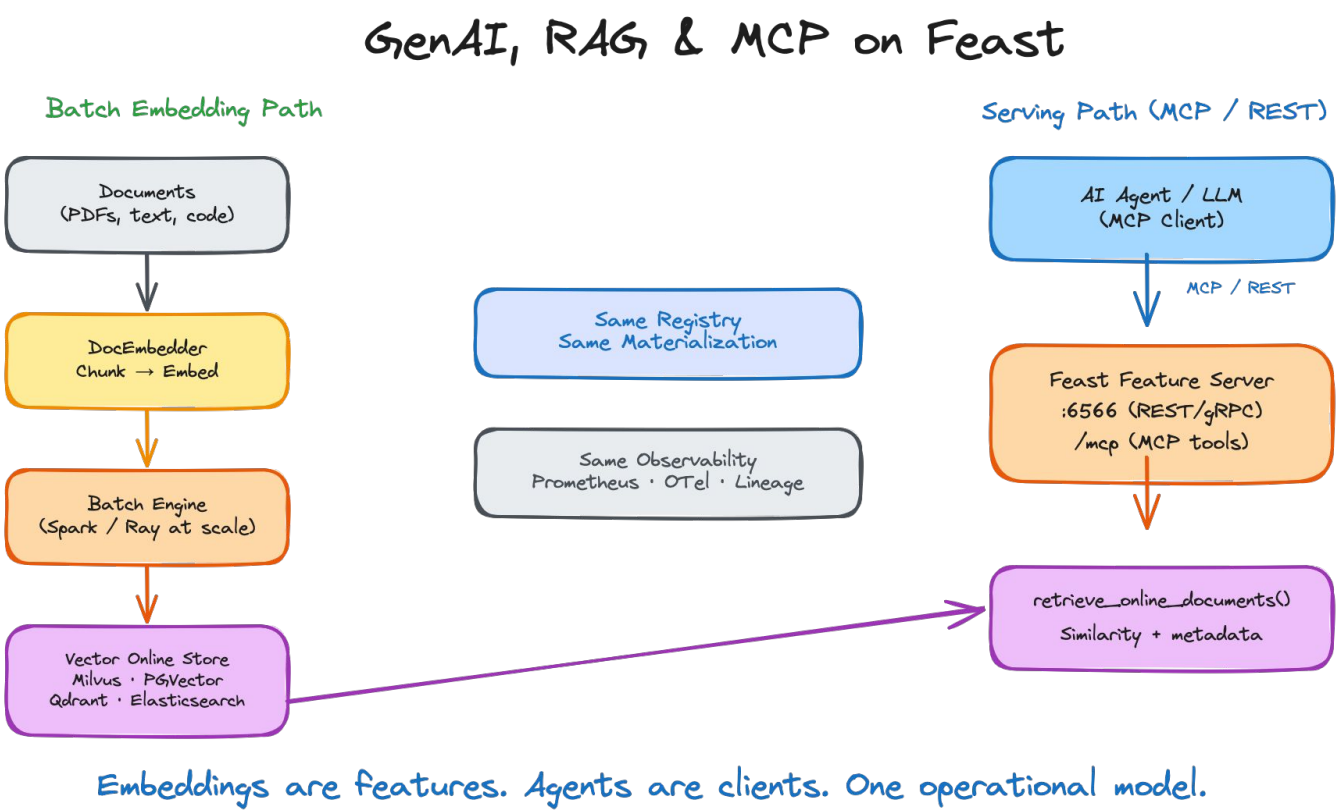
Compute Engines: Spark & Ray



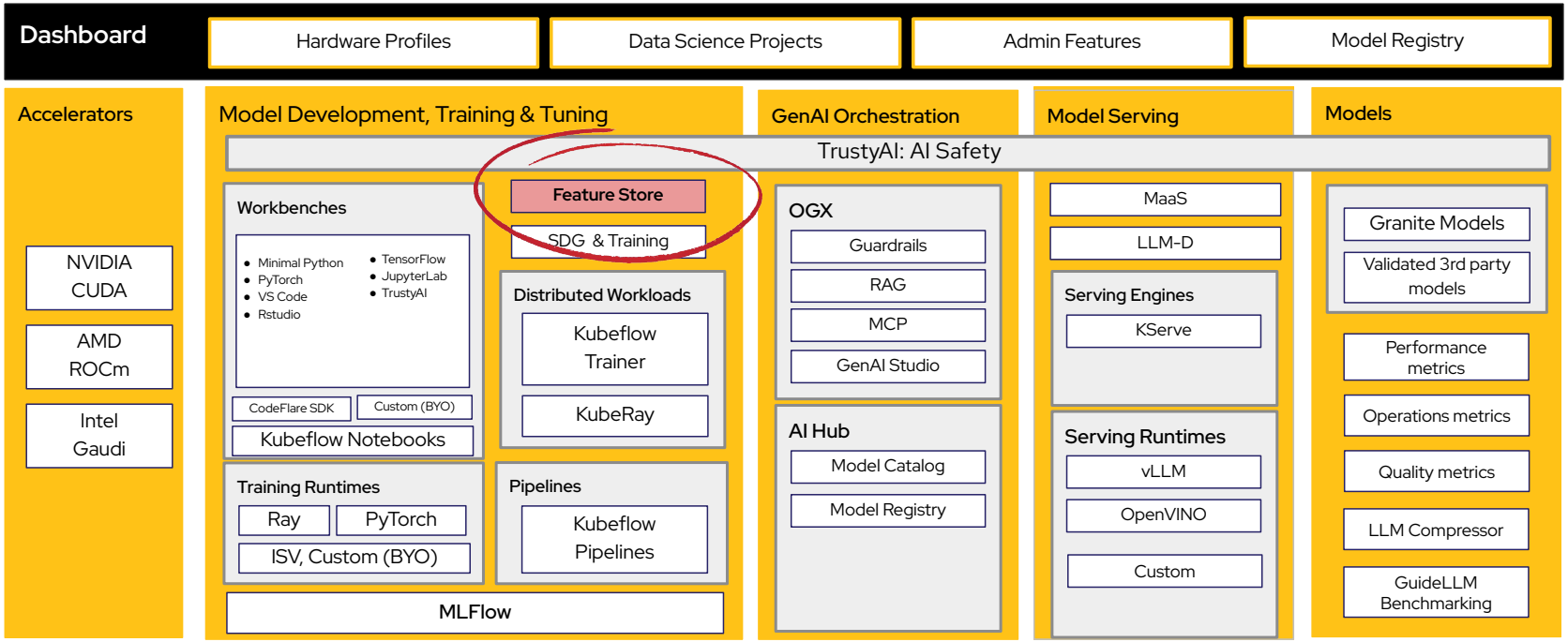
GenAI & RAG

Batch ML | Real-time ML | RAG | One stack

GenAI, RAG & MCP

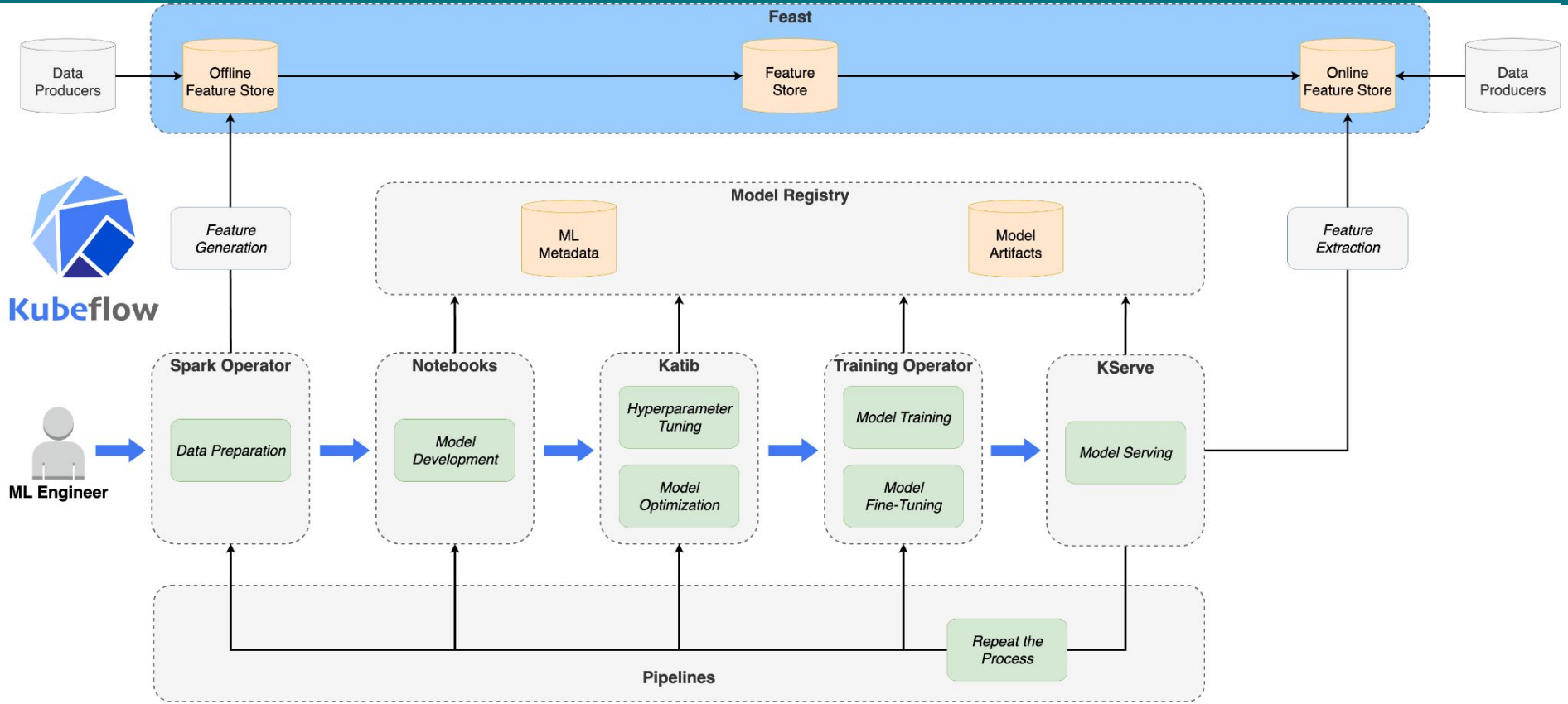


Feast in Open Data Hub



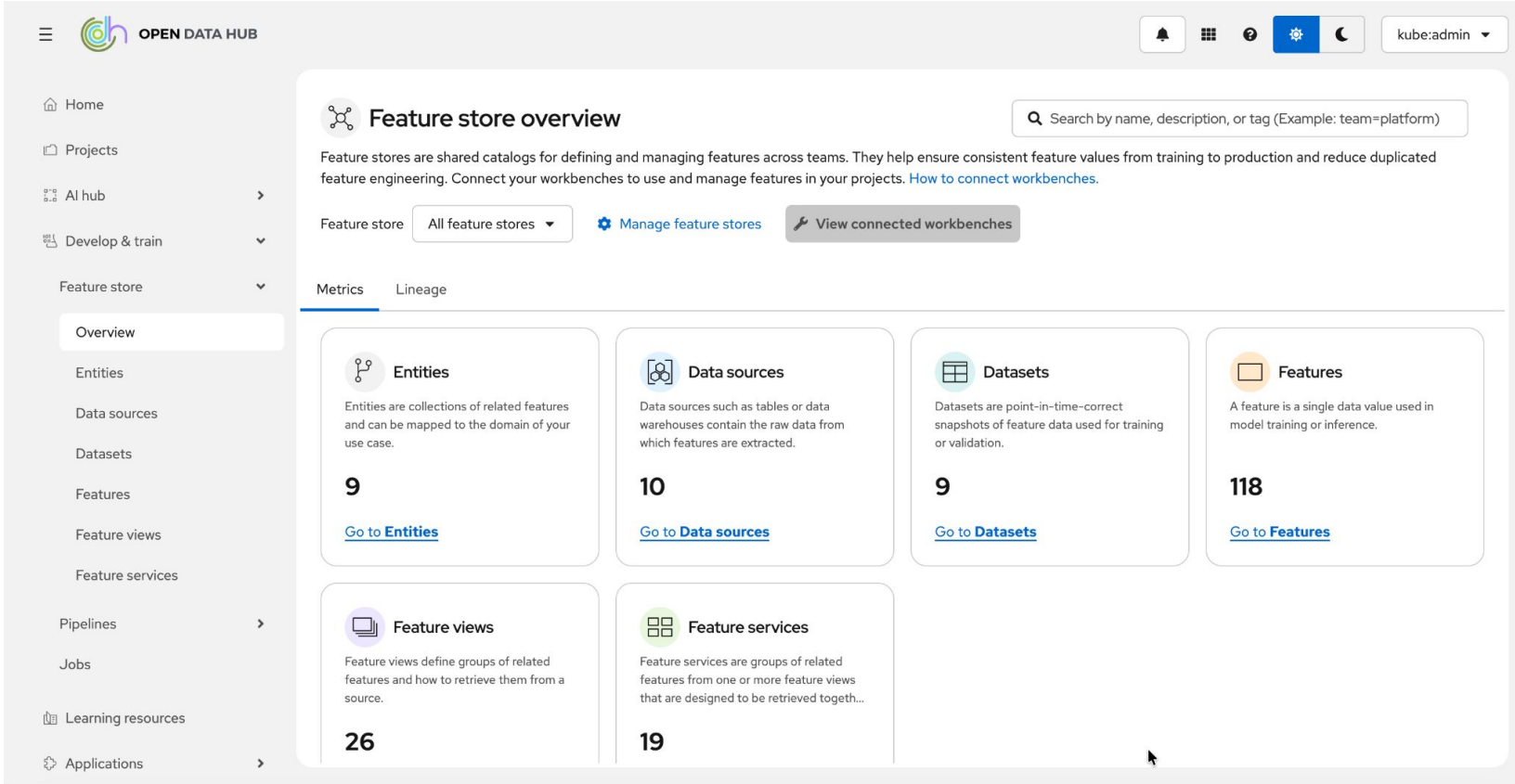
Feast in Open Data Hub

KubeCon CloudNativeCon India 2026

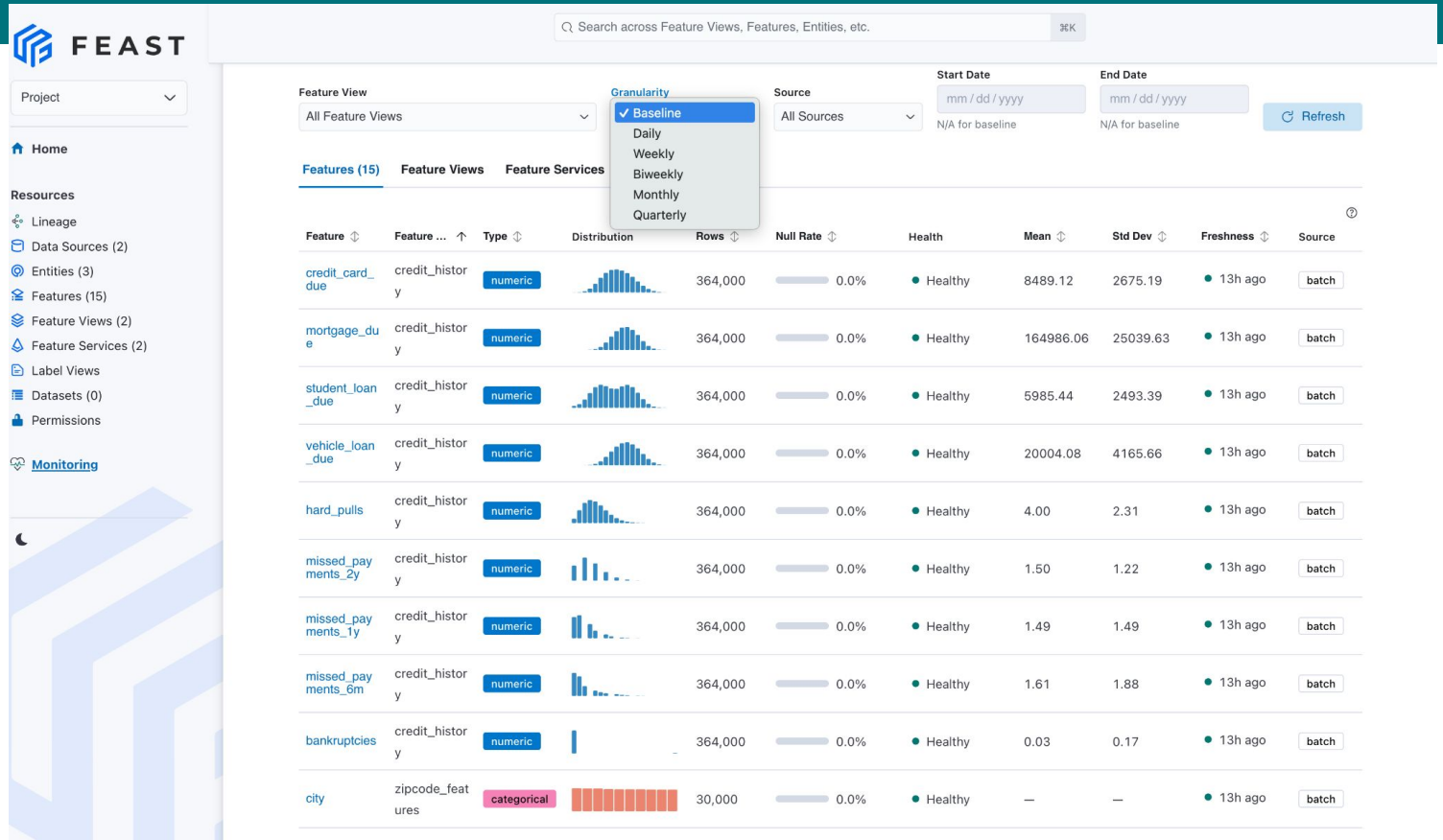


Feast in Open Data Hub

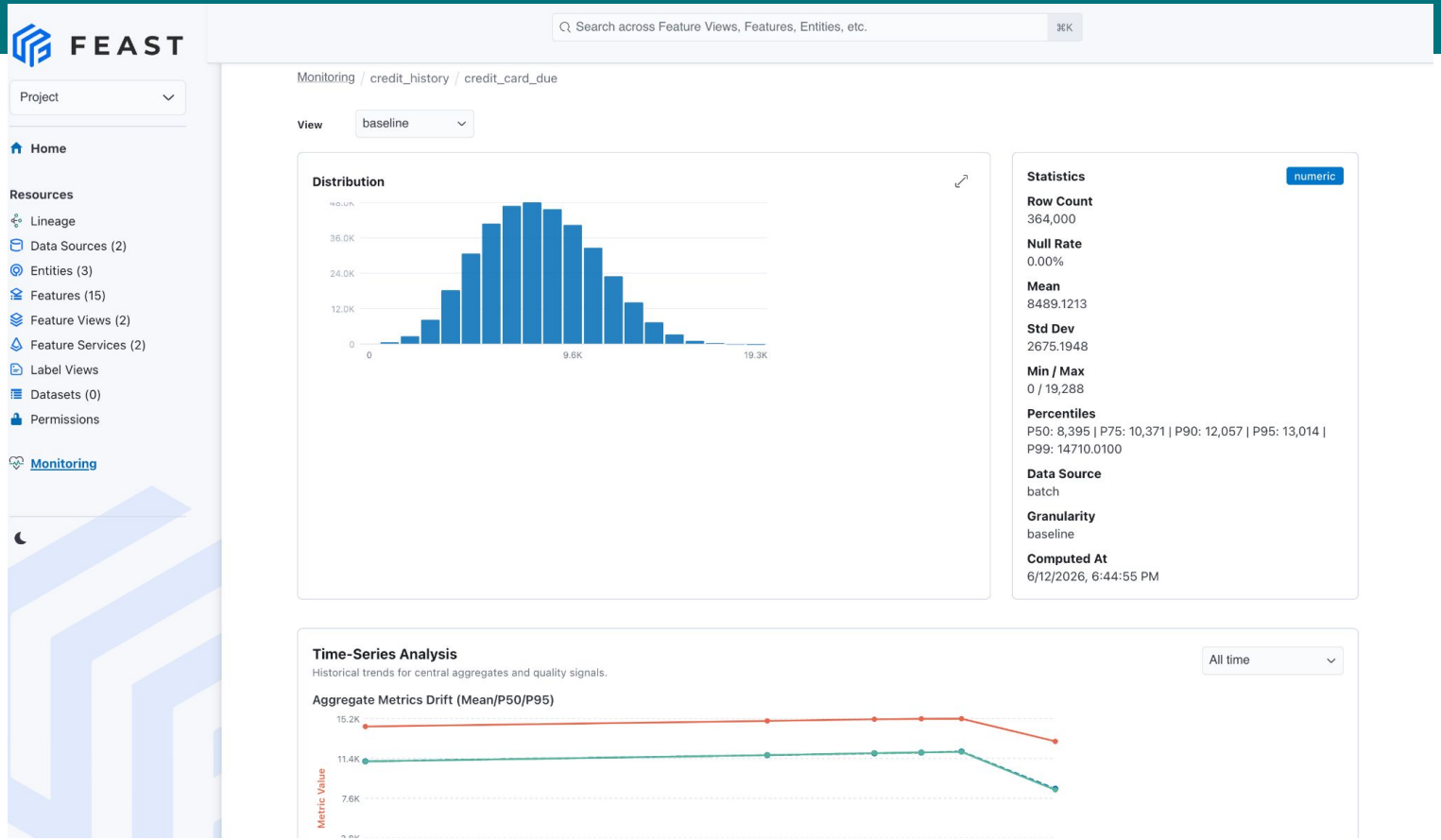
KubeCon CloudNativeCon India 2026



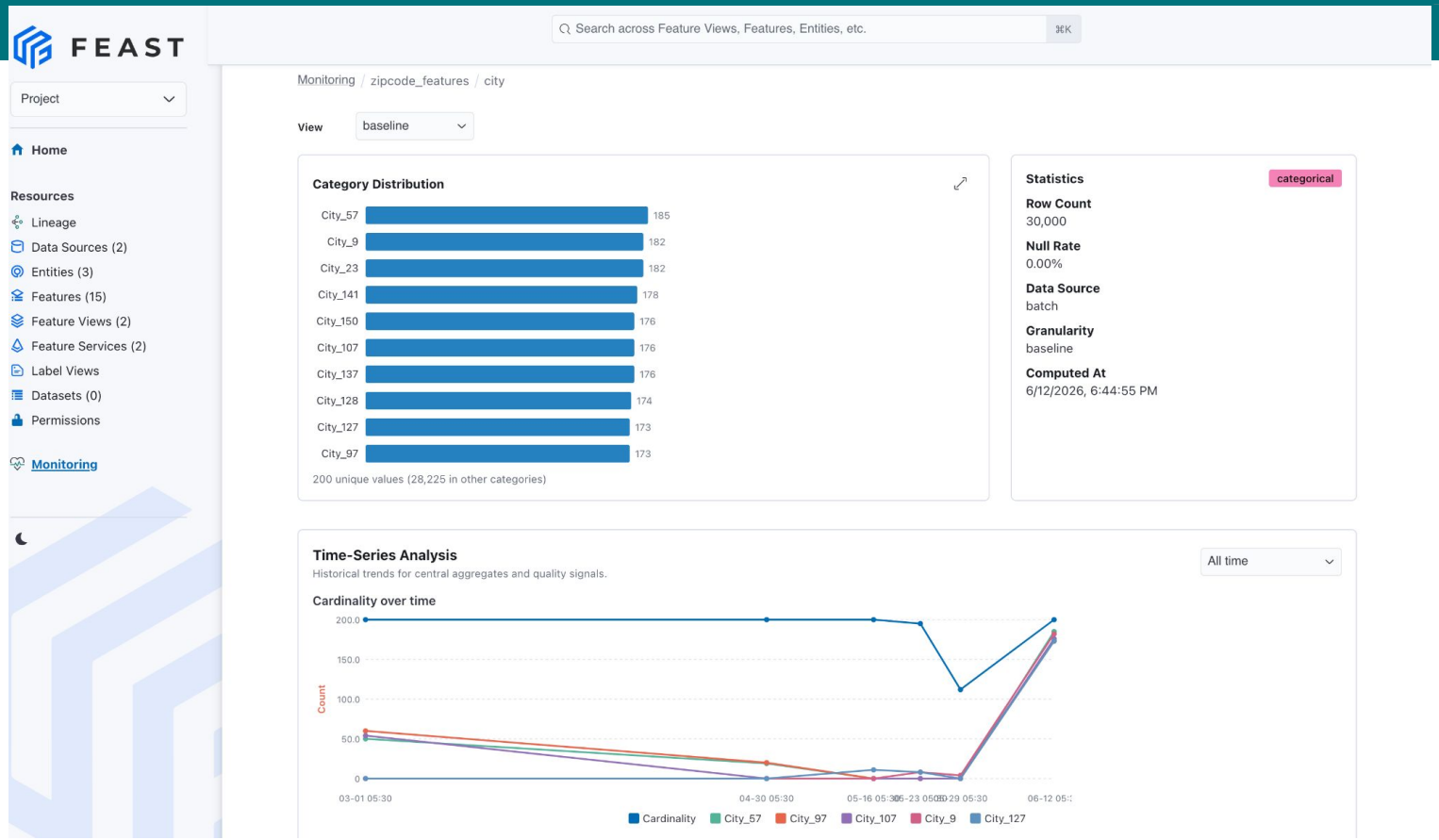
Monitoring


FEAST
Search across Feature Views, Features, Entities, etc.
⌘K
Project
Home
Resources
Lineage
Data Sources (2)
Entities (3)
Features (15)
Feature Views (2)
Feature Services (2)
Label Views
Datasets (0)
Permissions
Monitoring
Feature View
All Feature Views
Granularity
Baseline
Daily
Weekly
Biweekly
Monthly
Quarterly
Source
All Sources
Start Date
mm / dd / yyyy
N/A for baseline
End Date
mm / dd / yyyy
N/A for baseline
Refresh
Features (15)
Feature Views
Feature Services
Feature
Feature ...
Type
Distribution
Rows
Null Rate
Health
Mean
Std Dev
Freshness
Source
credit_card_due credit_history numeric 364,000 0.0% Healthy 8489.12 2675.19 13h ago batch
mortgage_due credit_history numeric 364,000 0.0% Healthy 164986.06 25039.63 13h ago batch
student_loan_due credit_history numeric 364,000 0.0% Healthy 5985.44 2493.39 13h ago batch
vehicle_loan_due credit_history numeric 364,000 0.0% Healthy 20004.08 4165.66 13h ago batch
hard_pulls credit_history numeric 364,000 0.0% Healthy 4.00 2.31 13h ago batch
missed_payments_2y credit_history numeric 364,000 0.0% Healthy 1.50 1.22 13h ago batch
missed_payments_1y credit_history numeric 364,000 0.0% Healthy 1.49 1.49 13h ago batch
missed_payments_6m credit_history numeric 364,000 0.0% Healthy 1.61 1.88 13h ago batch
bankruptcies credit_history numeric 364,000 0.0% Healthy 0.03 0.17 13h ago batch
city zipcode_features categorical 30,000 0.0% Healthy — — 13h ago batch

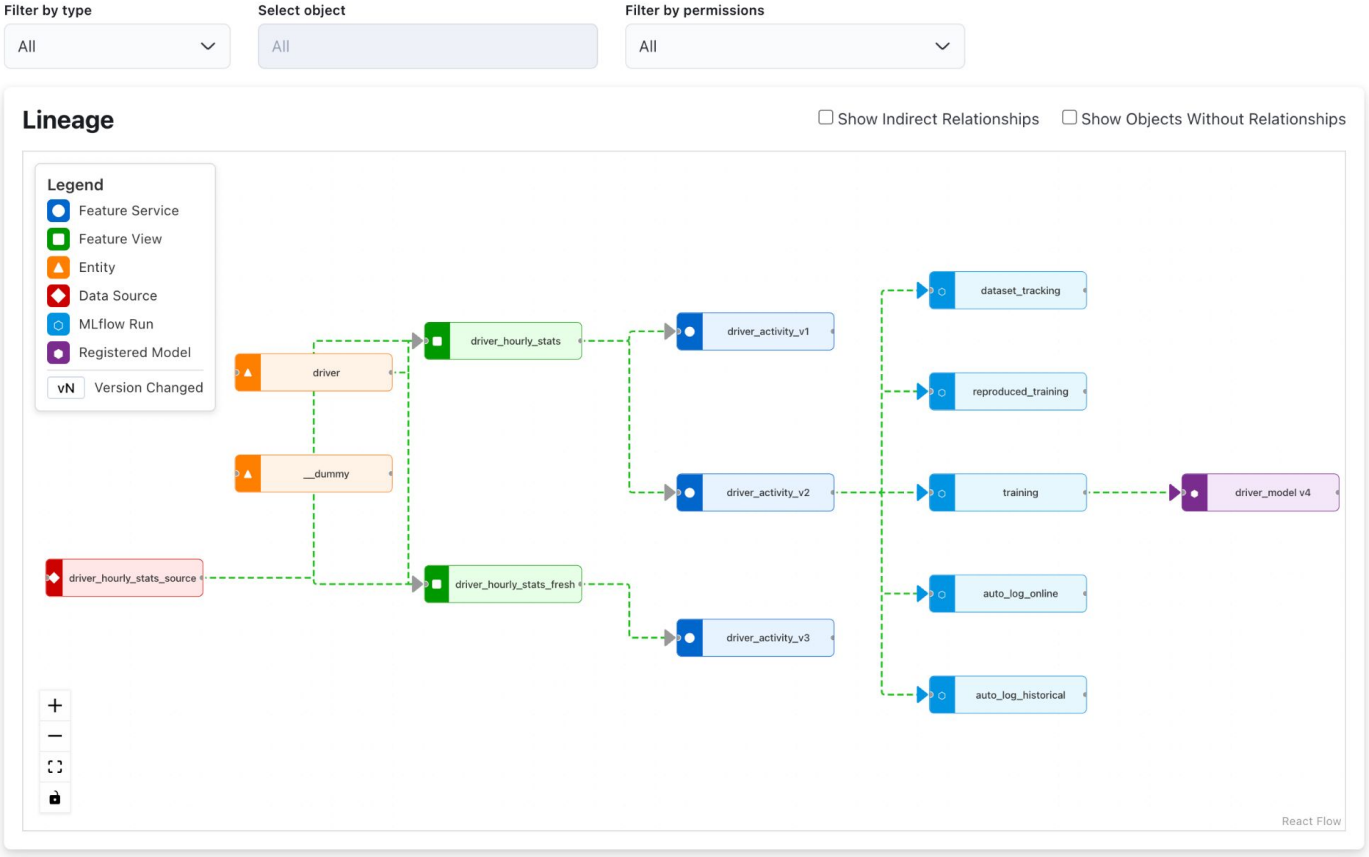
Monitoring


FEAST
Search across Feature Views, Features, Entities, etc.
⌘K
Project
Home
Resources
Lineage
Data Sources (2)
Entities (3)
Features (15)
Feature Views (2)
Feature Services (2)
Label Views
Datasets (0)
Permissions
Monitoring
Monitoring / credit_history / credit_card_due
View
baseline
Distribution
48.0K
36.0K
24.0K
12.0K
0
0
9.6K
19.3K
Statistics
numeric
Row Count
364,000
Null Rate
0.00%
Mean
8489.1213
Std Dev
2675.1948
Min / Max
0 / 19,288
Percentiles
P50: 8,395 | P75: 10,371 | P90: 12,057 | P95: 13,014 | P99: 14710.0100
Data Source
batch
Granularity
baseline
Computed At
6/12/2026, 6:44:55 PM
Time-Series Analysis
Historical trends for central aggregates and quality signals.
All time
Aggregate Metrics Drift (Mean/P50/P95)
15.2K
11.4K
7.6K
Metric Value

Monitoring


FEAST
Search across Feature Views, Features, Entities, etc.
⌘K
Project
Home
Resources
Lineage
Data Sources (2)
Entities (3)
Features (15)
Feature Views (2)
Feature Services (2)
Label Views
Datasets (0)
Permissions
Monitoring
Monitoring / zipcode_features / city
View
baseline
Category Distribution
City_57 185
City_9 182
City_23 182
City_141 178
City_150 176
City_107 176
City_137 176
City_128 174
City_127 173
City_97 173
200 unique values (28,225 in other categories)
Statistics
categorical
Row Count
30,000
Null Rate
0.00%
Data Source
batch
Granularity
baseline
Computed At
6/12/2026, 6:44:55 PM
Time-Series Analysis
Historical trends for central aggregates and quality signals.
All time
Cardinality over time
200.0
150.0
100.0
50.0
0
Count
03-01 05:30
04-30 05:30
05-16 05:30
05-23 05:30
05-29 05:30
06-12 05:
Cardinality
City_57
City_97
City_107
City_9
City_127

Native MLFlow integration



Resources

- https://feast.dev/
- https://github.com/feast-dev/feast
- https://opendatahub.io/
- https://docs.feast.dev/master/how-to-guides/production-deployment-topologies
- **Join us at** feastopensource.slack.com

Learn More:

https://feast.dev

FEAST

KubeCon CloudNativeCon India 2026

FEAST

Thank You!

Questions?

Nikhil Kathole

Principal Software Engineer, Red Hat