



# FEATURE STORE SUMMIT

**12-13 OCTOBER | 08:30 AM - 4:00 PM PT**

ORGANIZED BY HOPSWORKS

# Databricks Feature Store

co-designed with a Data and ML-ops platform



Mani Parkhe  
ML Tech Lead  
Databricks



Avesh Singh  
Sr. Software Engineer  
Databricks

# databricks Lakehouse Platform

SIMPLE OPEN COLLABORATIVE

Data Engineering

BI & SQL  
Analytics

Real-time Data  
Applications

Data Science  
& Machine Learning

Data Management & Governance



Open Data Lake



Structured



Semi-structured

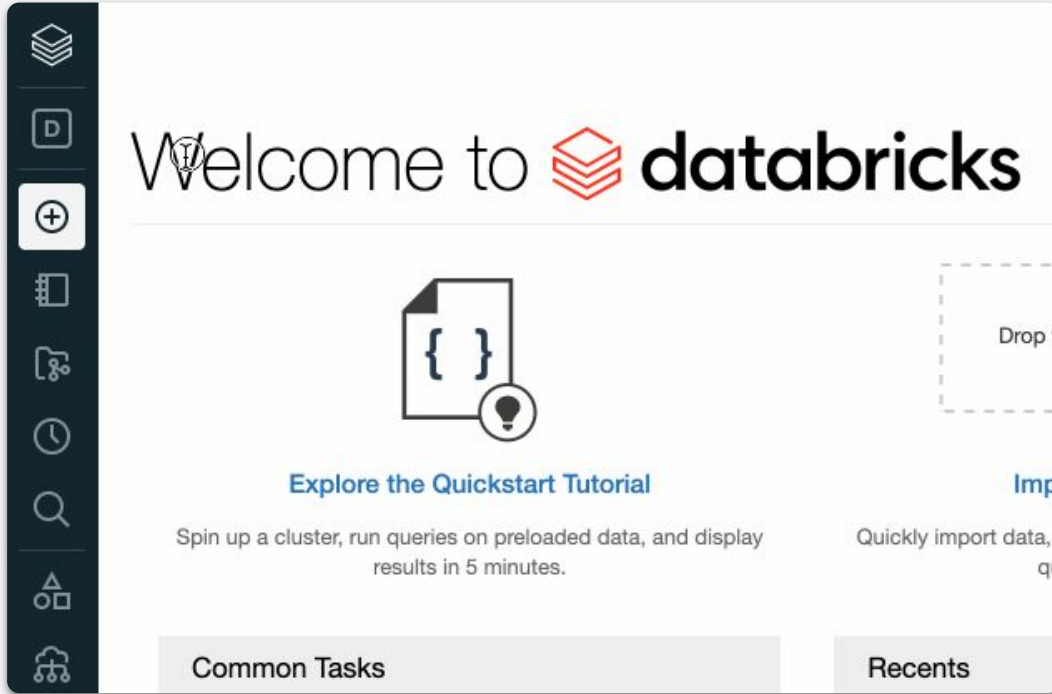


Unstructured



Streaming

# Databricks Machine Learning



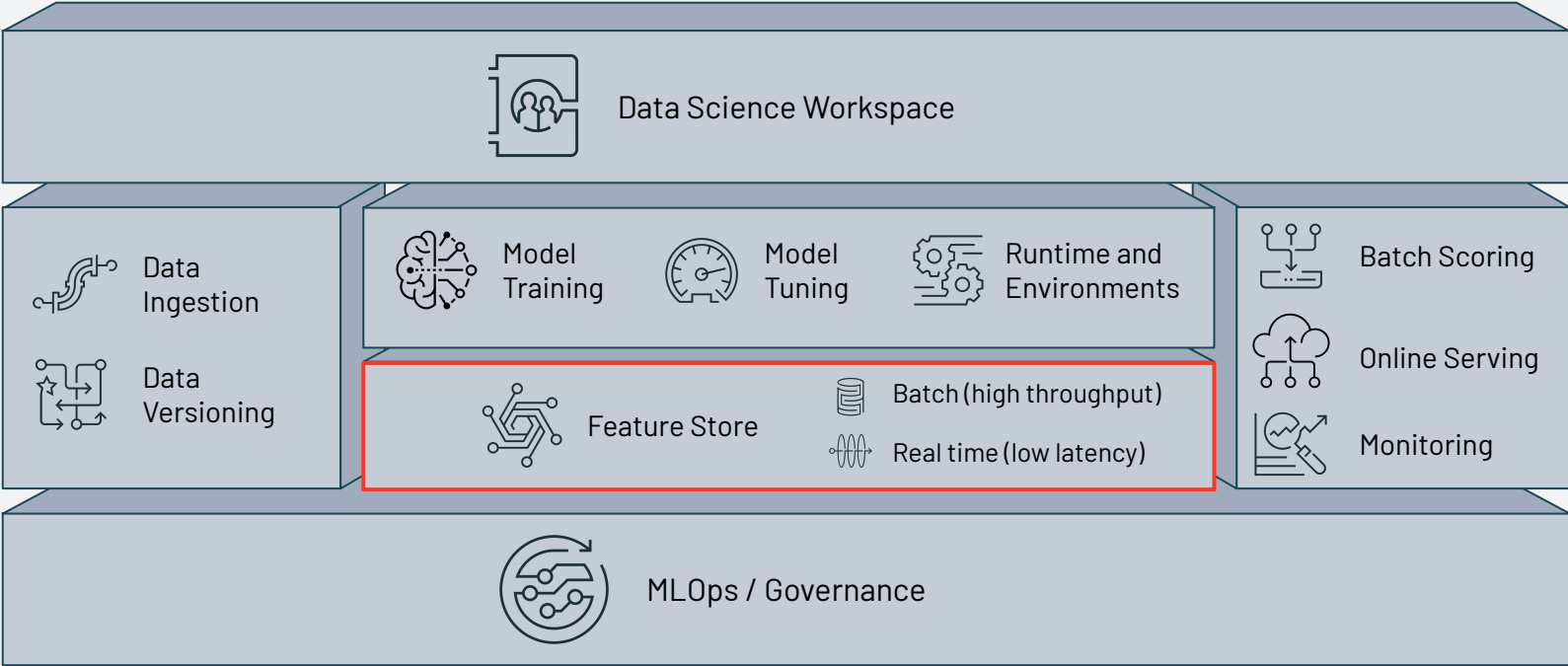
All the power of Databricks, plus:

- + ML ready environments (e.g. for TensorFlow)
- + Experiment Tracking + Reproducibility
- + Model Interpretability/Understanding
- + Model Registry (Model Lifecycle Management)
- + Model Serving
- + AutoML
- + Feature Store + Feature Serving

# The data problem in AI

1. **Complexity** of data models, data types, custom computation logic
2. **Online/offline skew** affecting model performance
3. **Reusability** and **Discoverability** of existing features
4. **Proprietary data formats, tooling, and vendor lock in**

# Databricks Feature Store



# Databricks Feature Store

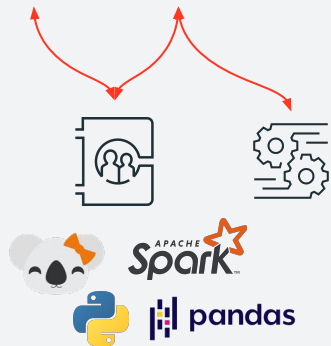
## 1. Open data and model formats

### Feature Definitions

- Flexibility in feature definition
- Performance attributes based on language choice

Feature 1

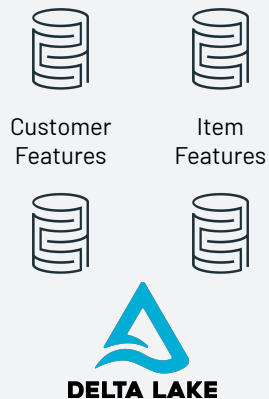
Feature 2



write

### Feature Tables

- SQL tables that can be queried from any language
- ACL'd, versioned, and optimized for performance



training set

read

publish

### Model Training



### Batch Scoring



mlflow

### Online Serving



databricks  
Model Serving



REST  
Endpoint

# Databricks Feature Store

## 2. Backward and Forward Lineage Tracking

Scheduled jobs and notebooks that compute the features

Table and file sources used for feature computation

Consumers like model, endpoints, jobs, .. etc that use these features.

The screenshot displays the Databricks Feature Store interface for a feature named 'prod.customer\_features'. The interface includes a sidebar with navigation options like 'Create', 'Workspace', 'Repos', 'Recents', 'Search', 'Data', 'Compute', 'Jobs', 'Experiments', 'Feature Store', and 'Models'. The main content area shows the feature's details, including its creation and modification dates, data sources ('customers.page\_view\_events' and 'customers.receipts'), and a description: 'Customer features using purchase receipts and app interactions data.' Below this, there are two tables: 'Producers (2)' and 'Features (7)'. The 'Producers' table lists 'feature\_creator job' (Every hour (US/Pacific), Success) and 'feature\_creator' (No schedule). The 'Features' table lists various features like 'customer\_id', 'page\_visits\_7d', 'page\_visits\_6mly', 'total\_purchases\_30d', 'total\_purchases\_6m', and 'total\_purchases\_7d', each with its data type and consumers (Models, Endpoints, Jobs, Notebooks). Red arrows indicate lineage from the 'feature\_creator' job to the 'customer\_id' feature and from the 'feature\_creator' job to the 'page\_visits\_7d' feature. Another red arrow points from the 'feature\_creator' job to the 'total\_purchases\_30d' feature. A red arrow also points from the 'feature\_creator' job to the 'total\_purchases\_6m' feature. A red arrow points from the 'feature\_creator' job to the 'total\_purchases\_7d' feature. A red arrow points from the 'feature\_creator' job to the 'total\_purchases\_30d' feature. A red arrow points from the 'feature\_creator' job to the 'total\_purchases\_6m' feature. A red arrow points from the 'feature\_creator' job to the 'total\_purchases\_7d' feature. A red arrow points from the 'feature\_creator' job to the 'total\_purchases\_30d' feature. A red arrow points from the 'feature\_creator' job to the 'total\_purchases\_6m' feature. A red arrow points from the 'feature\_creator' job to the 'total\_purchases\_7d' feature.

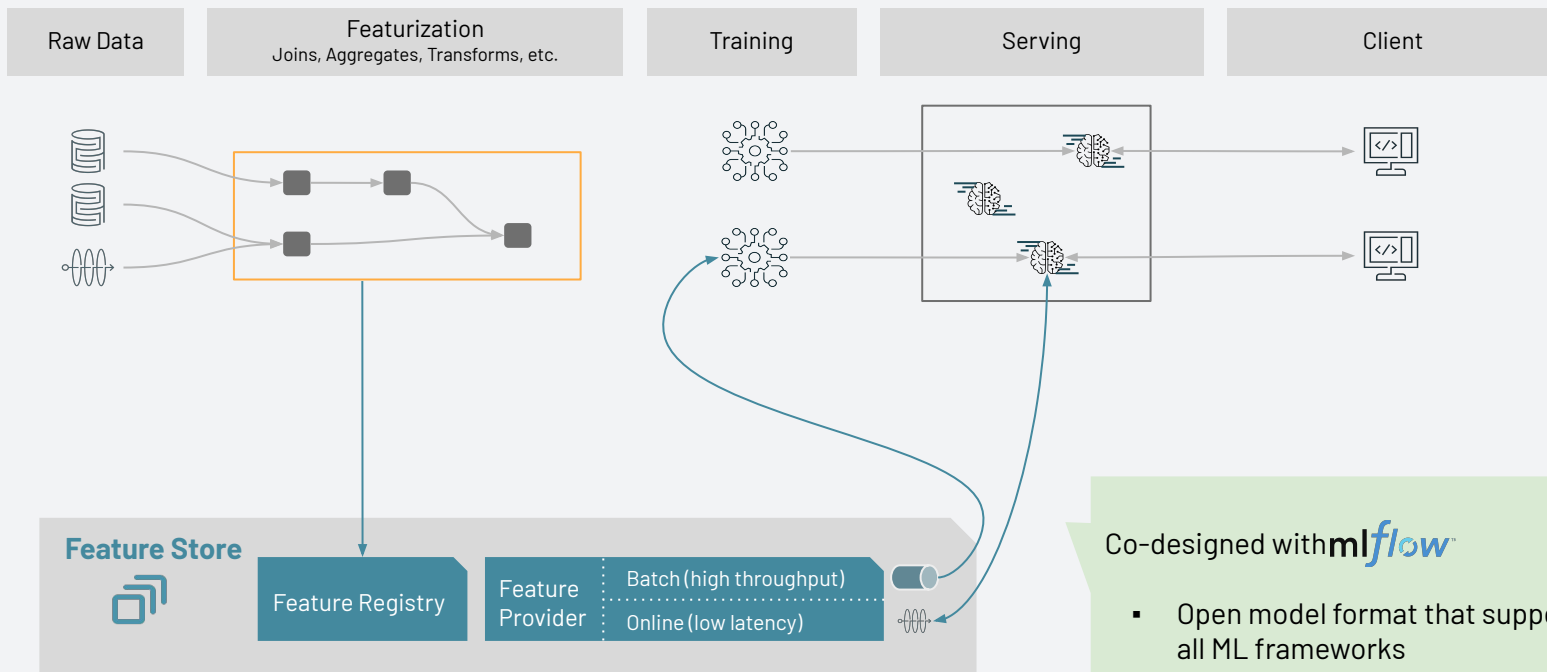
Name	Schedule	Status	Last run	Last written
feature_creator job	Every hour (US/Pacific)	Success	2021-10-08 22:38:08	2021-10-08 22:38:29
feature_creator	No schedule	-	-	2021-10-08 22:35:30

Feature	Data Type	Consumers			
		Models	Endpoints	Jobs	Notebooks
customer_id	INTEGER	-	-	-	-
page_visits_7d	FLOAT	product_recommendation_model/1	product_recommendation_model/1	feature_creator job +1 more	score_product_recommendations +1 more
page_visits_6mly	FLOAT	product_recommendation_model/1	product_recommendation_model/1	feature_creator job +1 more	score_product_recommendations +1 more
total_purchases_30d	FLOAT	product_recommendation_model/1	product_recommendation_model/1	purchasing_power +2 more	score_product_recommendations +1 more
total_purchases_6m	FLOAT	product_recommendation_model/1	product_recommendation_model/1	purchasing_power +2 more	score_product_recommendations +1 more
total_purchases_7d	FLOAT	product_recommendation_model/1	product_recommendation_model/1	purchasing_power +2 more	score_product_recommendations +1 more



# Databricks Feature Store

## 3. Seamless integration with MLflow and in-built feature lookup



Co-designed with **mlflow**

- Open model format that supports all ML frameworks
- Feature version and lookup logic hermetically logged with Model



# Thank you!

**Do you have any questions?**

**Mani Parkhe**  
**[mani@databricks.com](mailto:mani@databricks.com)**