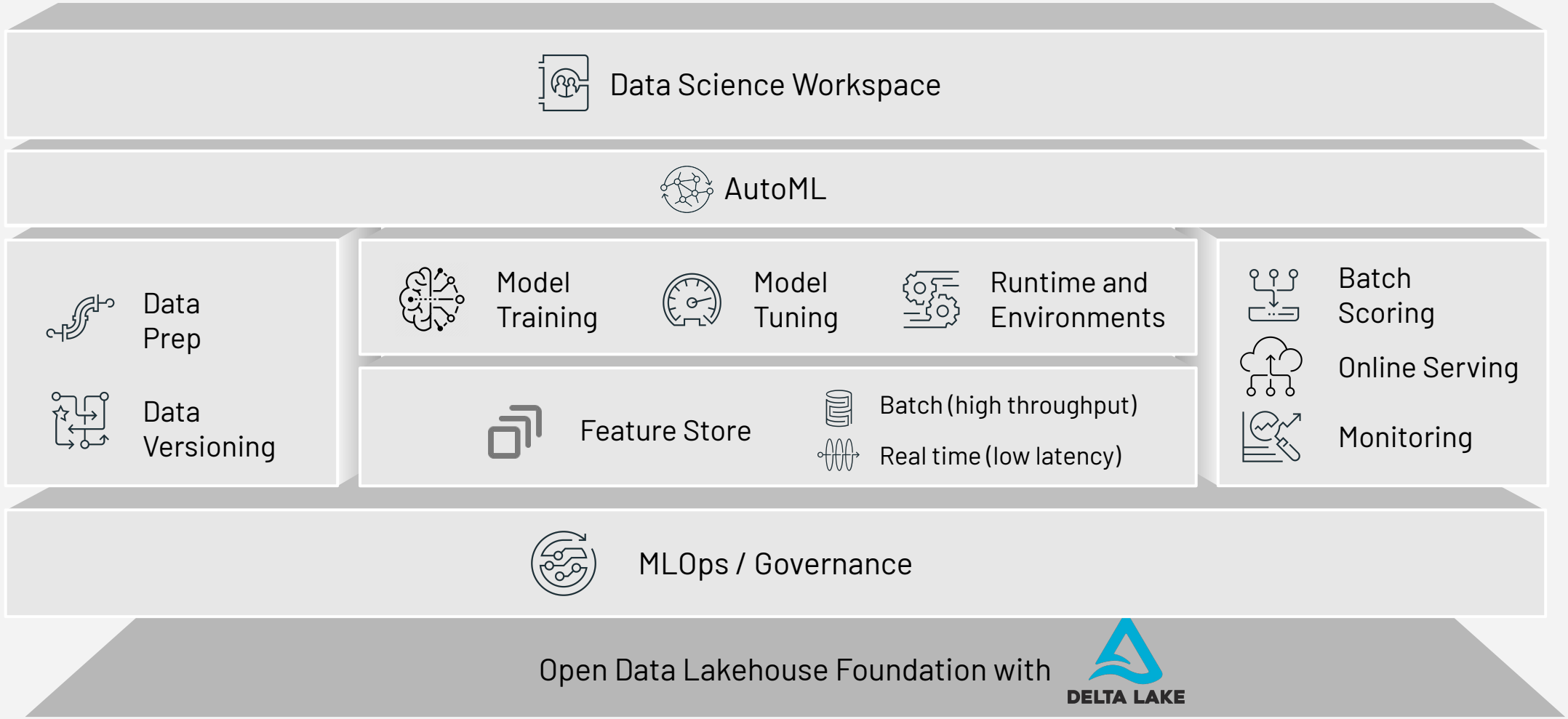
Databricks Feature Store

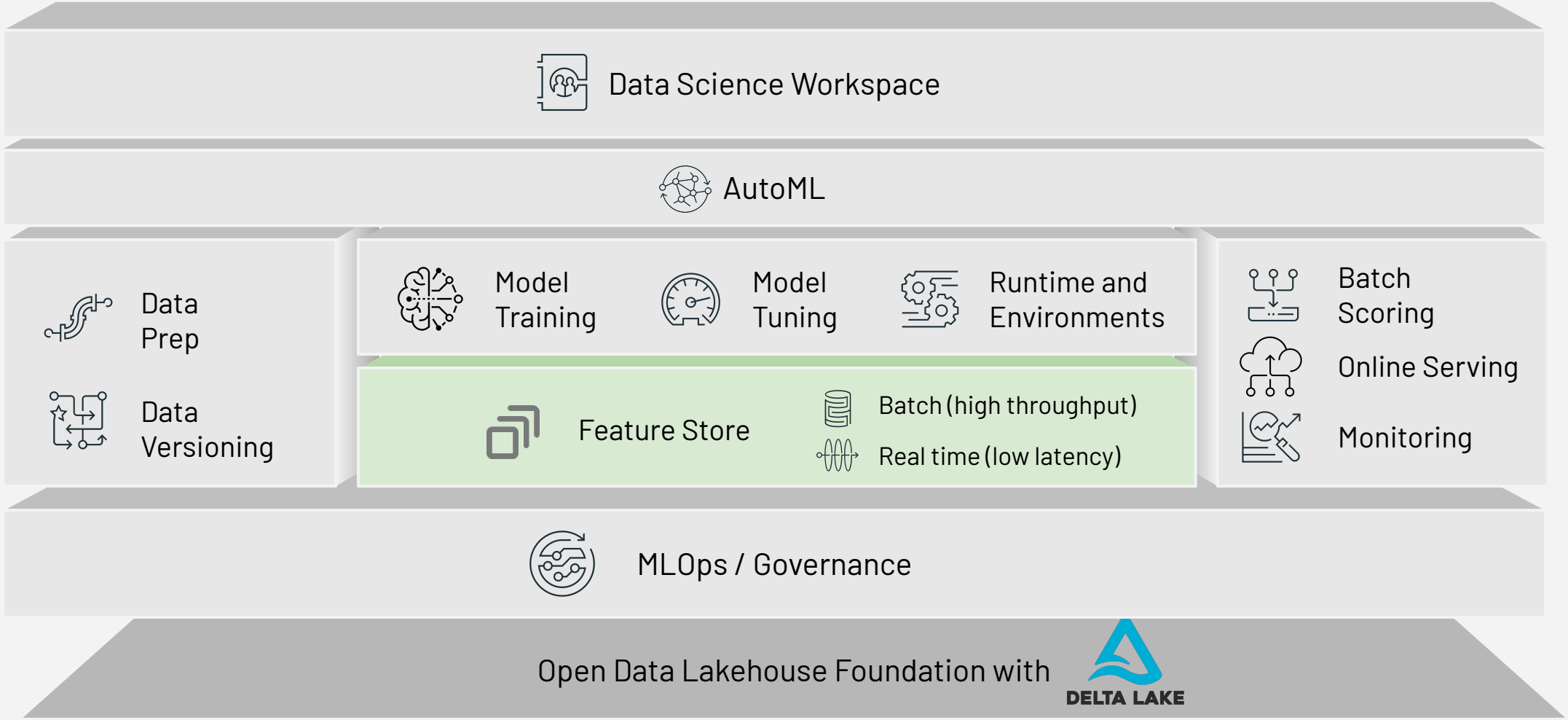
Lakehouse

A data-native and collaborative solution for the full ML lifecycle



Feature Store

The first Feature Store codesigned with a Data and MLOps Platform



What is a Feature?

A recommender model example

Raw - Data Origin

Users table
Zip code, Payment methods, etc.

Items table
Description, Category, etc.

Purchases
User ID, Item ID, Date, Quantity, Price

Types of Features

Transformations

e.g. Category Encoding

Context Features

e.g. Weekday

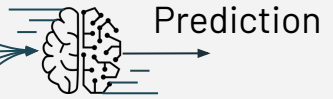
Feature Augmentation

e.g. Weather

Pre-computed Features

e.g. Purchases last 7, 14, 21 days

ML Model



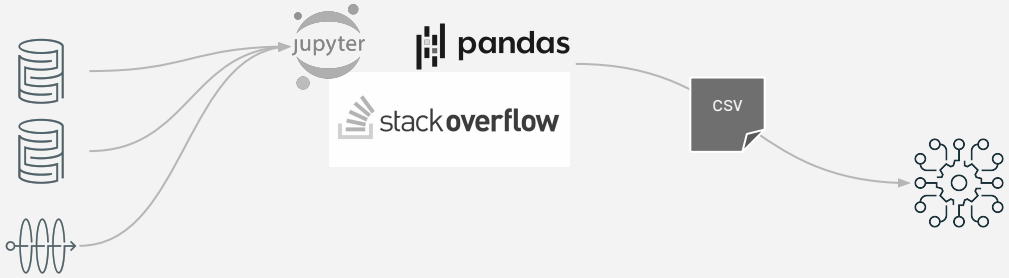
Results

Item	P(purchase|user)
	0.58
	0.13
	0.12
	0.01

databricks

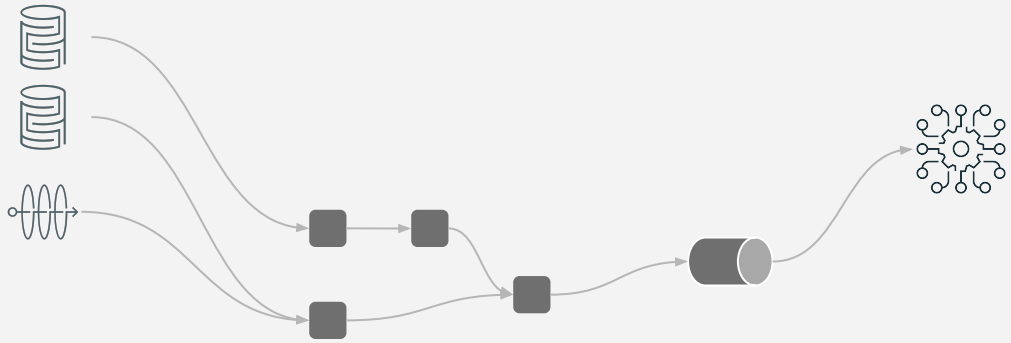
A day (or 6 months) in the life of an ML model

Raw Data	Featurization Joins, Aggregates, Transforms, etc.	Training	Serving	Client



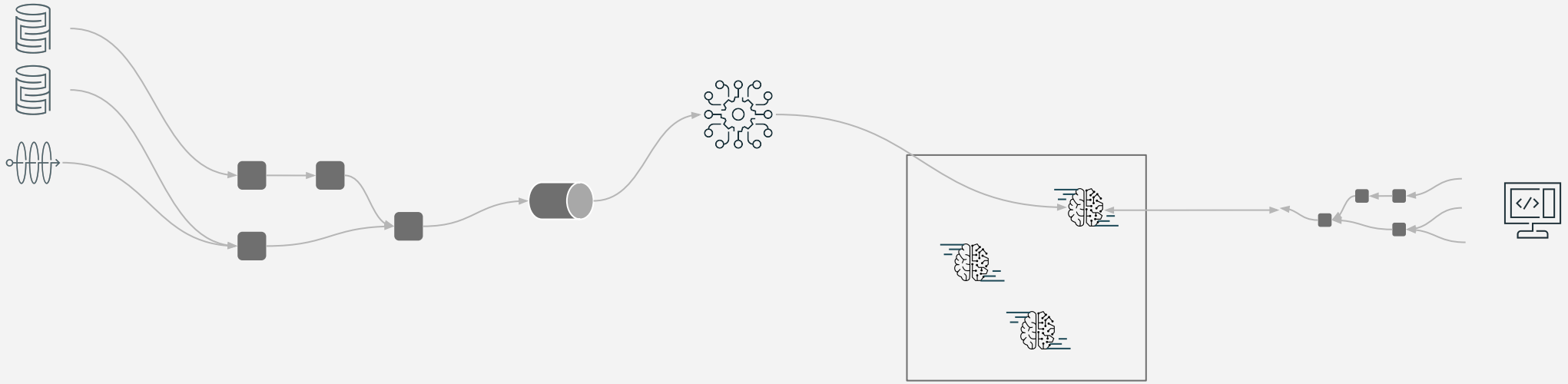
A day (or 6 months) in the life of an ML model

Raw Data	Featurization Joins, Aggregates, Transforms, etc.	Training	Serving	Client

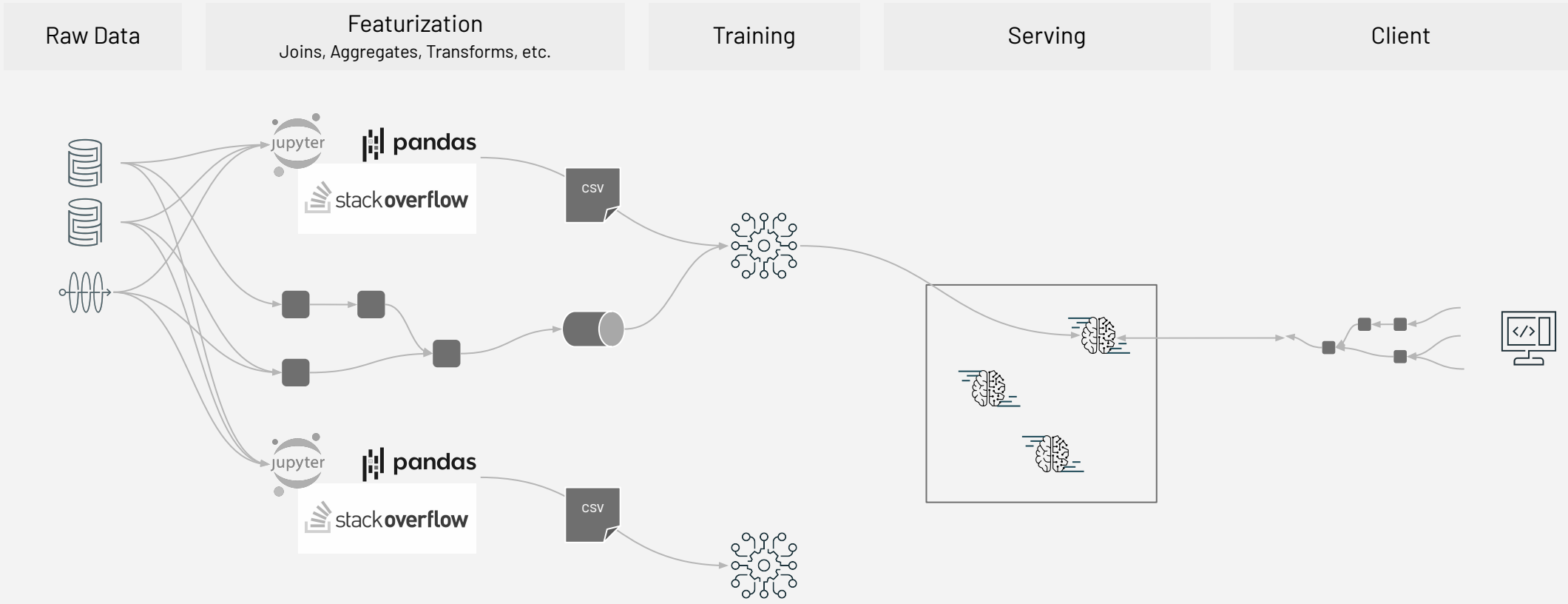


A day (or 6 months) in the life of an ML model

Raw Data	Featurization Joins, Aggregates, Transforms, etc.	Training	Serving	Client

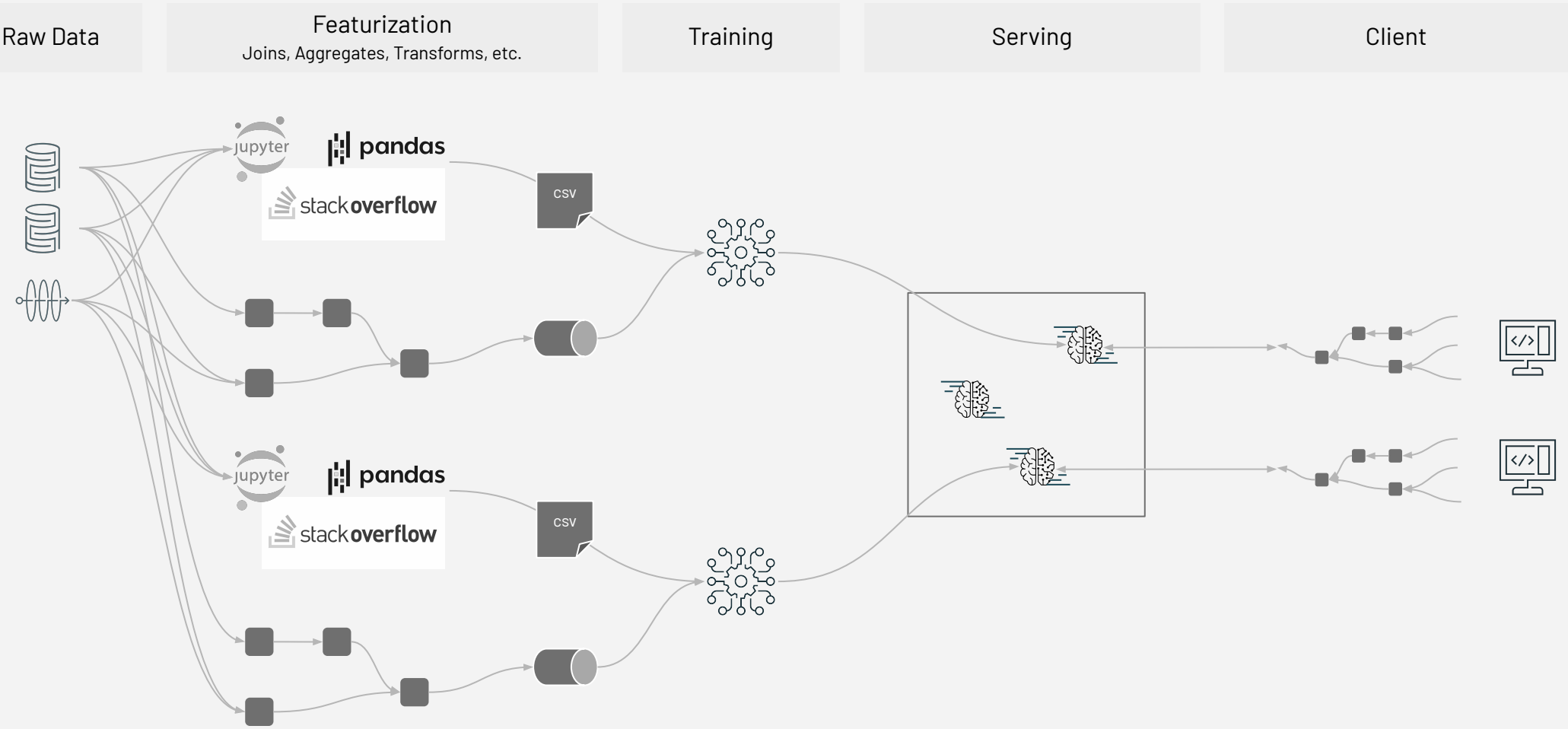


A day (or 6 months) in the life of an ML model



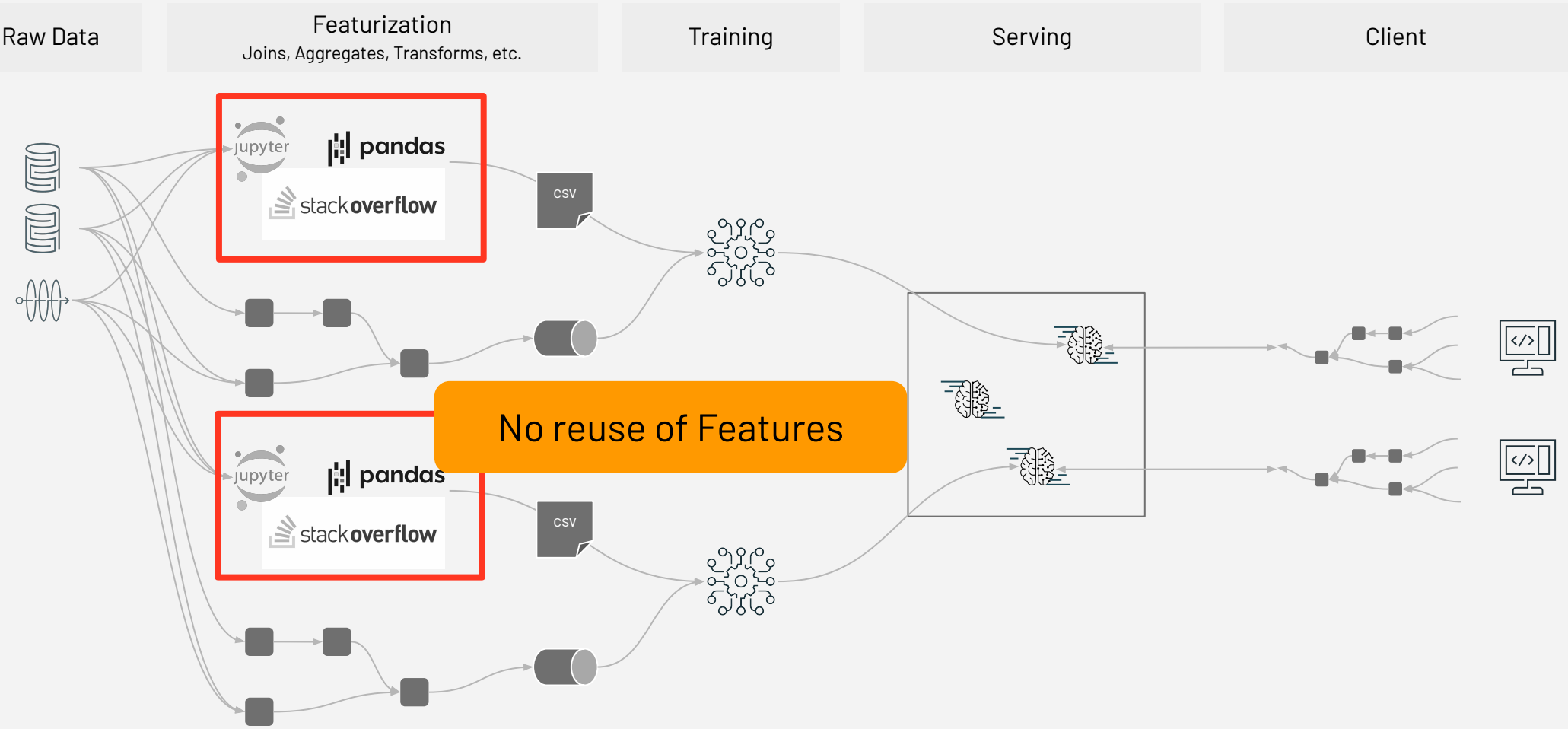
databricks

A day (or 6 months) in the life of an ML model



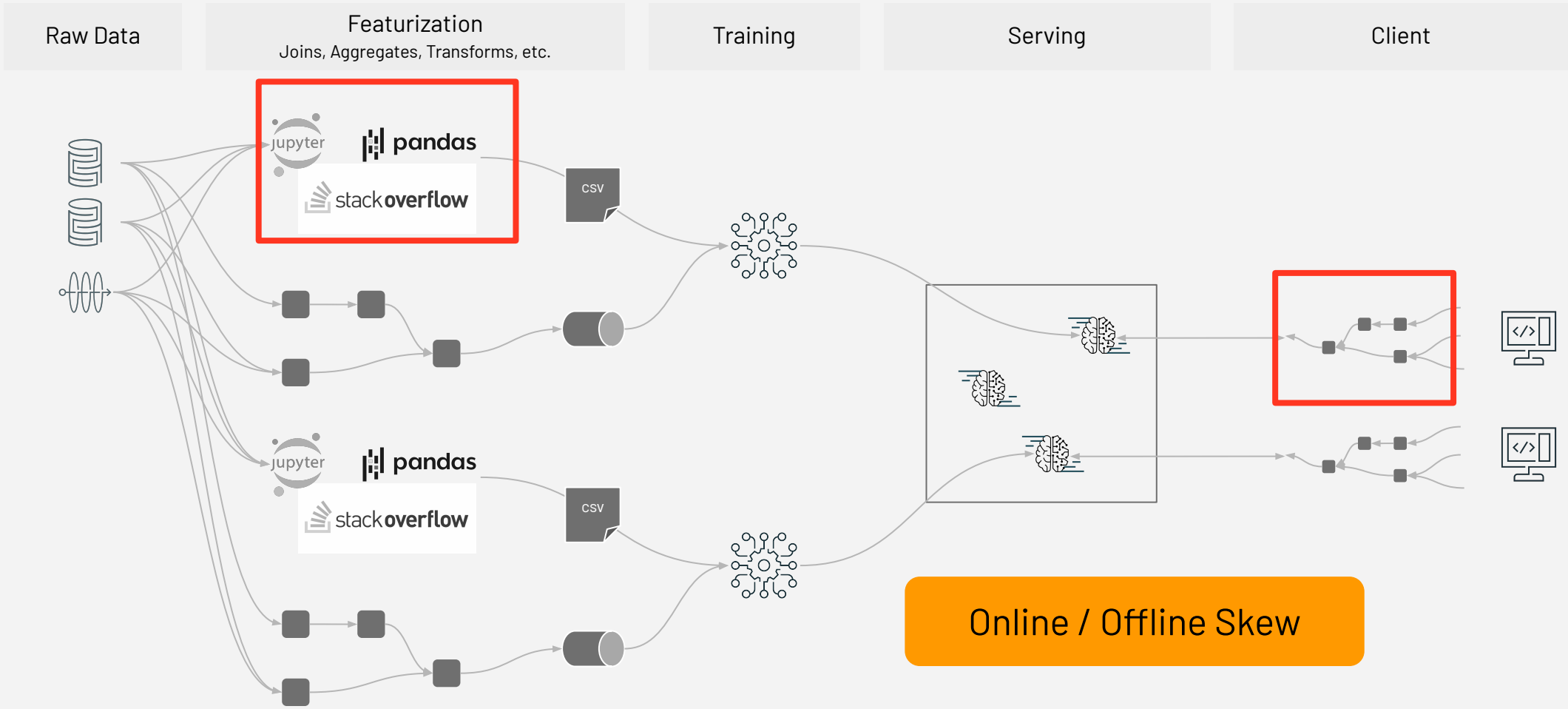
databricks

A day (or 6 months) in the life of an ML model



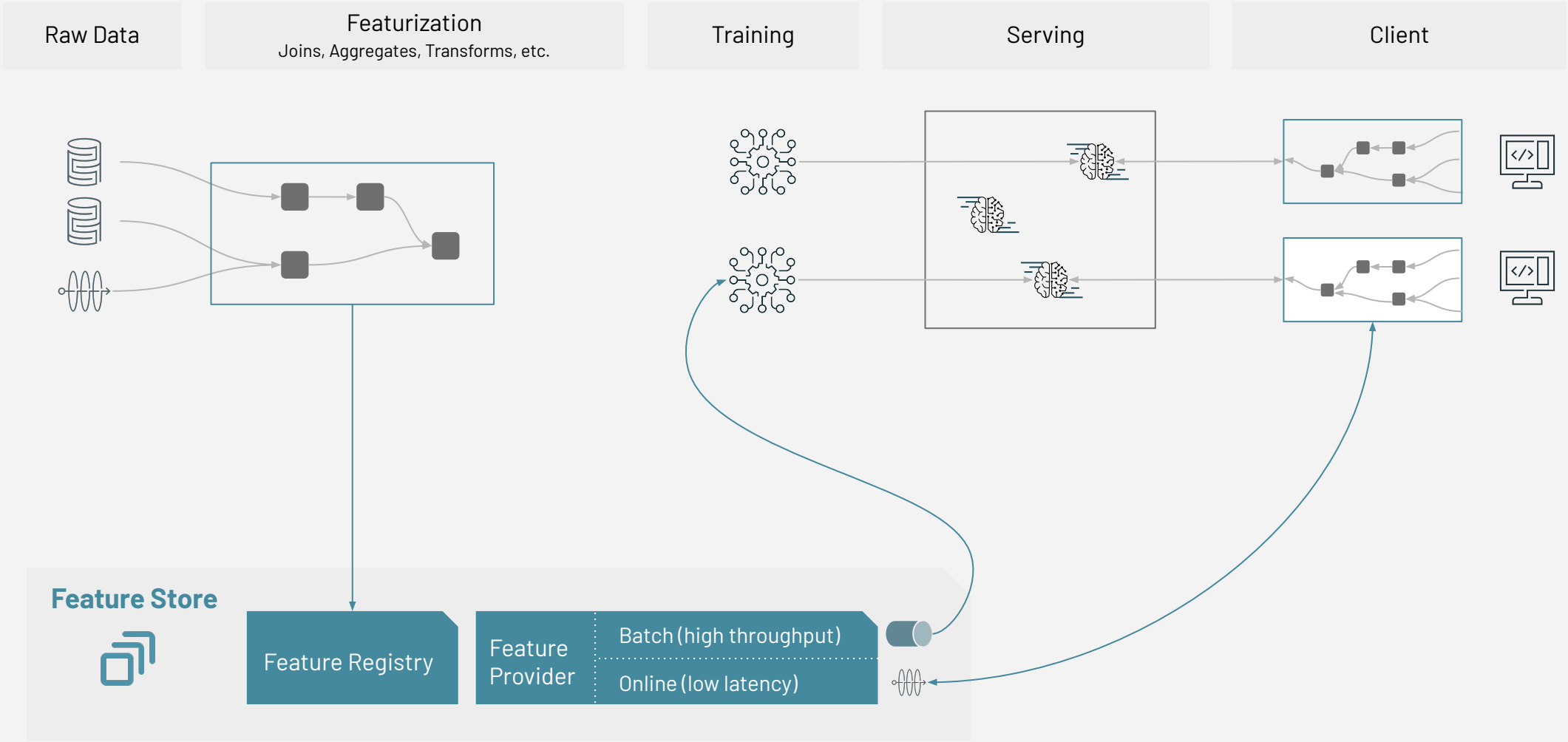
databricks

A day (or 6 months) in the life of an ML model

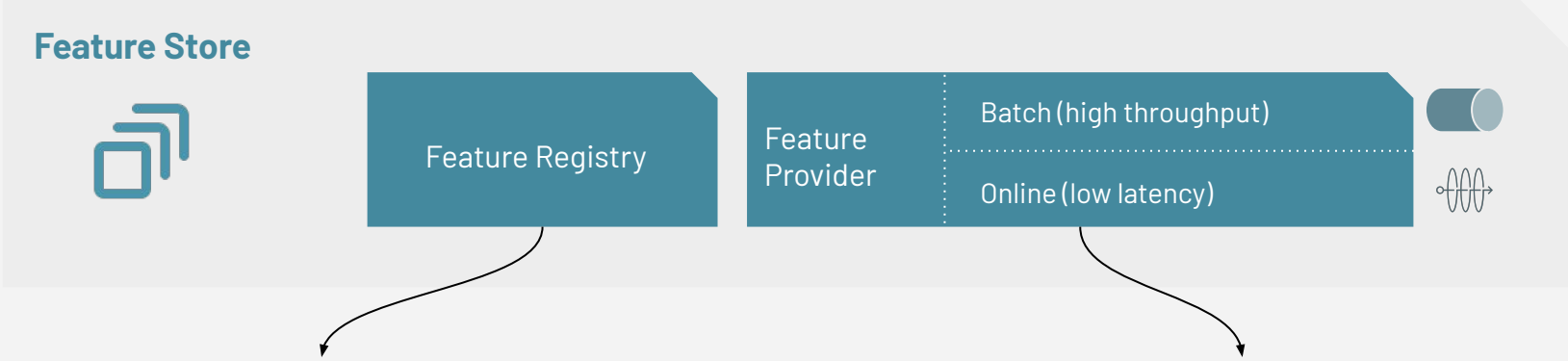


databricks

Solving the Feature Store Problem



Solving the Feature Store Problem



Feature Registry

- Discoverability and Reusability
- Versioning
- Upstream and downstream Lineage

Feature Provider

- Batch and online access to Features
- Feature lookup packaged with Models
- Simplified deployment process

Co-designed with DELTA LAKE

- Open format
- Built-in data versioning and governance
- Native access through PySpark, SQL, etc.

Co-designed with mlflow

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

ABN·AMRO

"The Databricks Feature Store is the missing piece to our unified ML platform. It creates a marketplace for features, enabling us to quickly develop and deploy new models from existing features."

— Jorg Klein, ABN Amro

Feature Store - Architecture Overview

Feature Definition

- Flexibility in creating features
- Attributes can be created in the preferred language

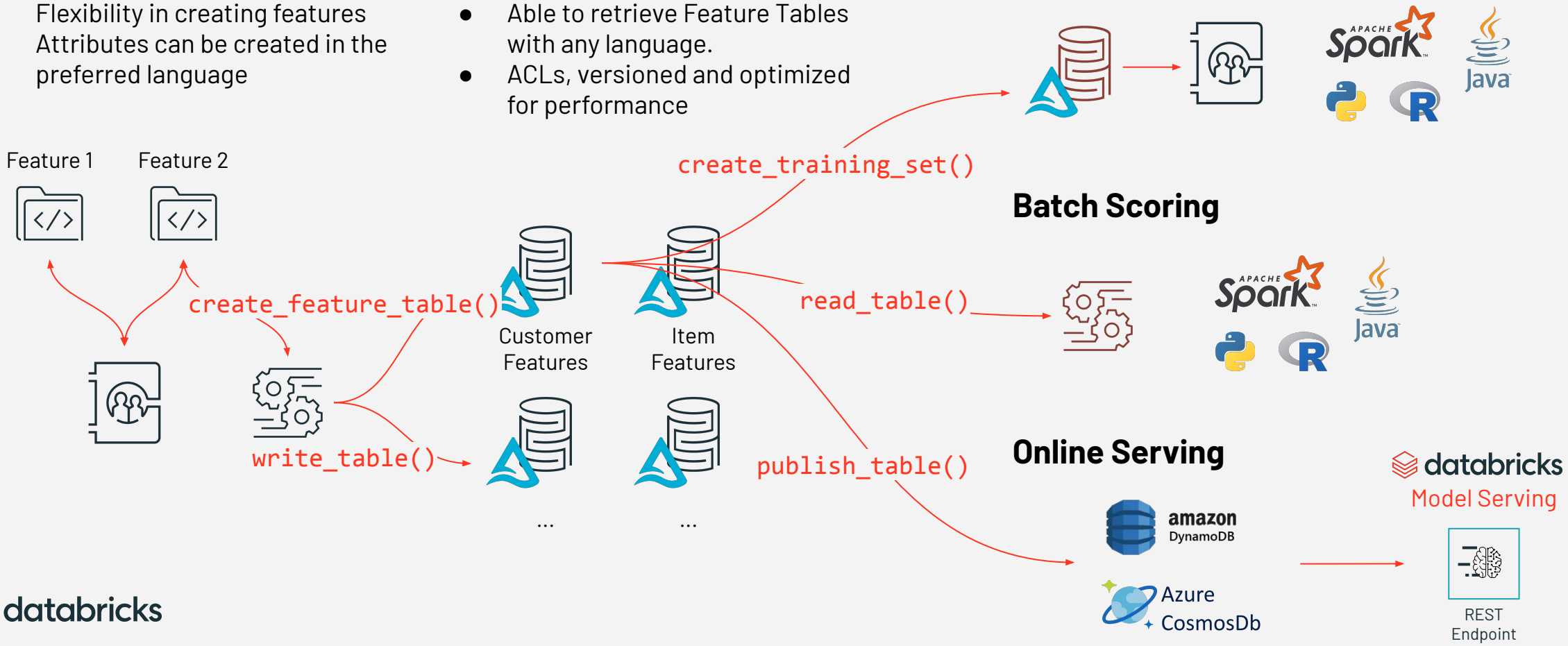
Feature Tables

- Able to retrieve Feature Tables with any language.
- ACLs, versioned and optimized for performance

Training Set Creation

Batch Scoring

Online Serving



Resources

- Documentation (AWS, Azure, GCP), including an example notebook
- Blog post introducing the Feature Store
- Data+AI Summit 2021 keynote introducing the Feature Store
- Data+AI Summit 2021 end-to-end demo including the Feature Store