



# Open Source Feature Store Workshop

## Caixabank Developer Session

—

June 2021

# Oracle Cloud Environment for Feature DataStore

Users	Jupyter URL	login
holuser01 to holuser10	<a href="https://129.213.113.81/hopworks/">https://129.213.113.81/hopworks/</a>	<a href="mailto:holuser01@oracle.com">holuser01@oracle.com</a> to <a href="mailto:holuser10@oracle.com">holuser10@oracle.com</a>
holuser11 to holuser20	<a href="https://150.136.14.68/hopworks/">https://150.136.14.68/hopworks/</a>	<a href="mailto:holuser11@oracle.com">holuser11@oracle.com</a> to <a href="mailto:holuser20@oracle.com">holuser20@oracle.com</a>

## Instances *in* DataScienceHOL Compartmet

The [Compute service](#) helps you provision VMs and bare metal instances to meet your compute and application requirements. An [instance](#) is a compute host. Choose between virtual machines (VMs) and bare metal instances. The image that you use to launch an instance determines its operating system and other software.

Create Instance

Name	State	Public IP	Shape	OCPU Count	Memory (GB)	Availability Domain	Fault Domain	Created
<a href="#">featured11</a>	<span>● Running</span>	150.136.14.68	VM.Standard.E4.Flex	8	128	AD-2	FD-1	Sun, Jun 20, 2021, 19:50:42 UT
<a href="#">featured11</a>	<span>● Running</span>	129.213.113.81	VM.Standard.E4.Flex	8	128	AD-1	FD-2	Sun, Jun 20, 2021, 19:32:16 UT

# Environment to download Documentation / Scripts

---

GitHub Repository: <https://github.com/operard/featuredslab>

Documentation for Workshop: <https://github.com/operard/featuredslab/tree/main/doc>

Documentation of Hopsworks:

<https://github.com/logicalclocks/hops-examples>

<https://examples.hopsworks.ai/>

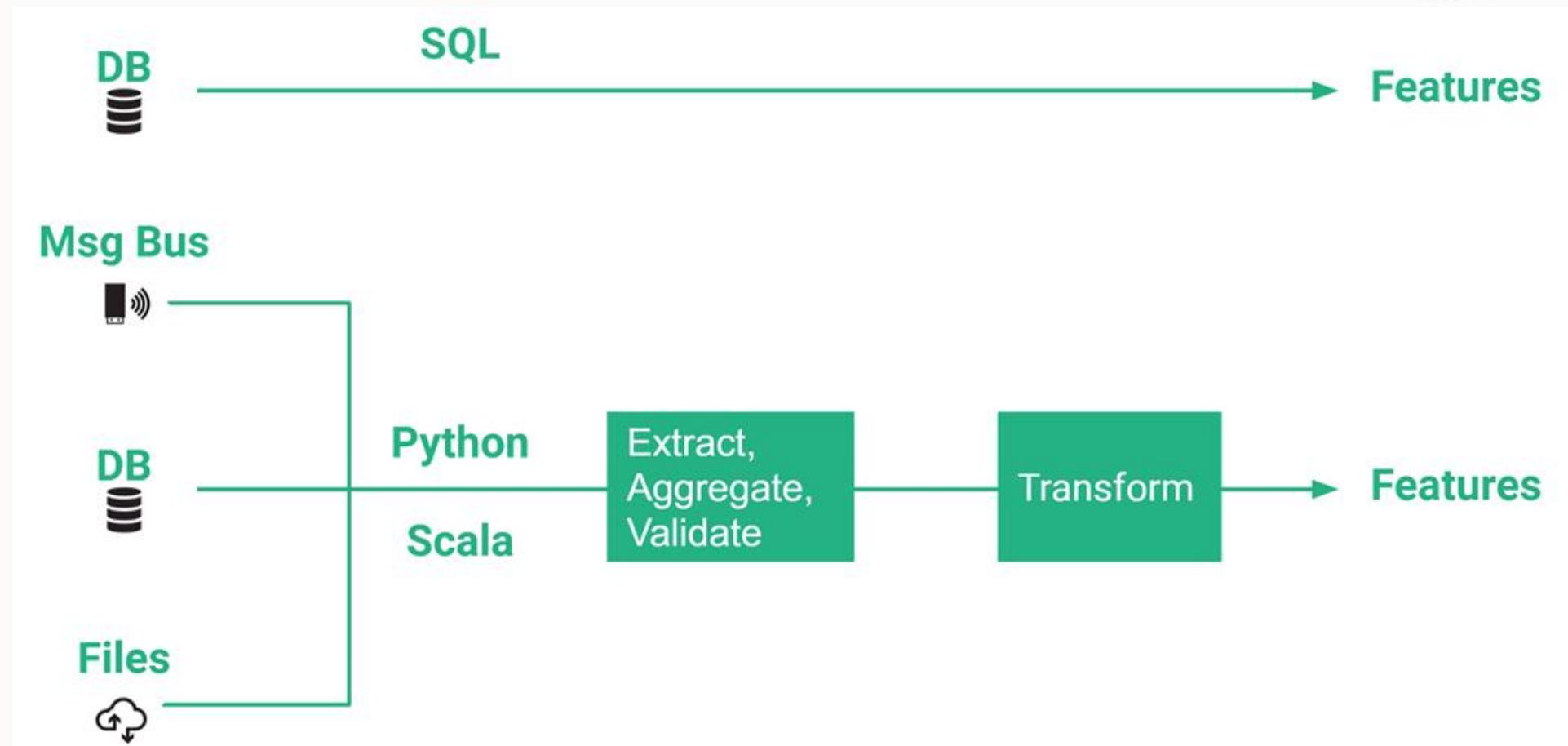
# Open Source Feature Store Concepts

---

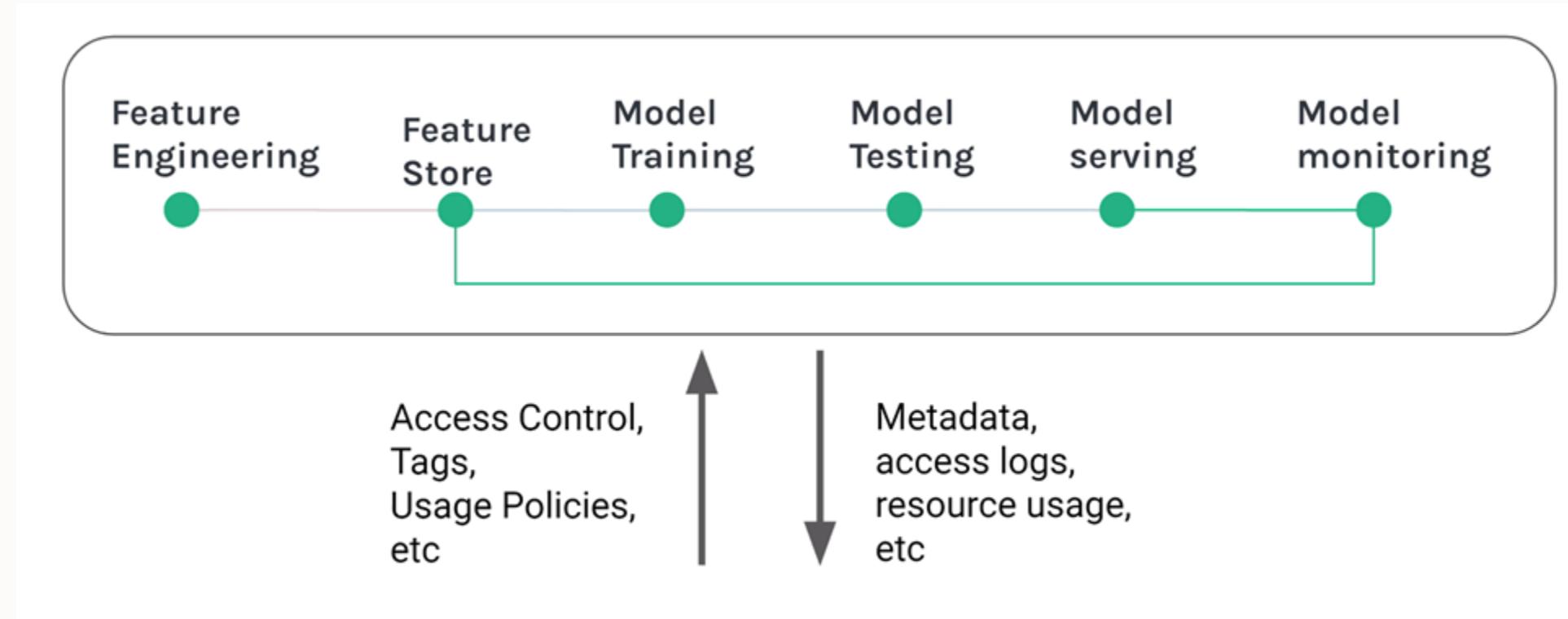
# From ML Pipelines to Production Models



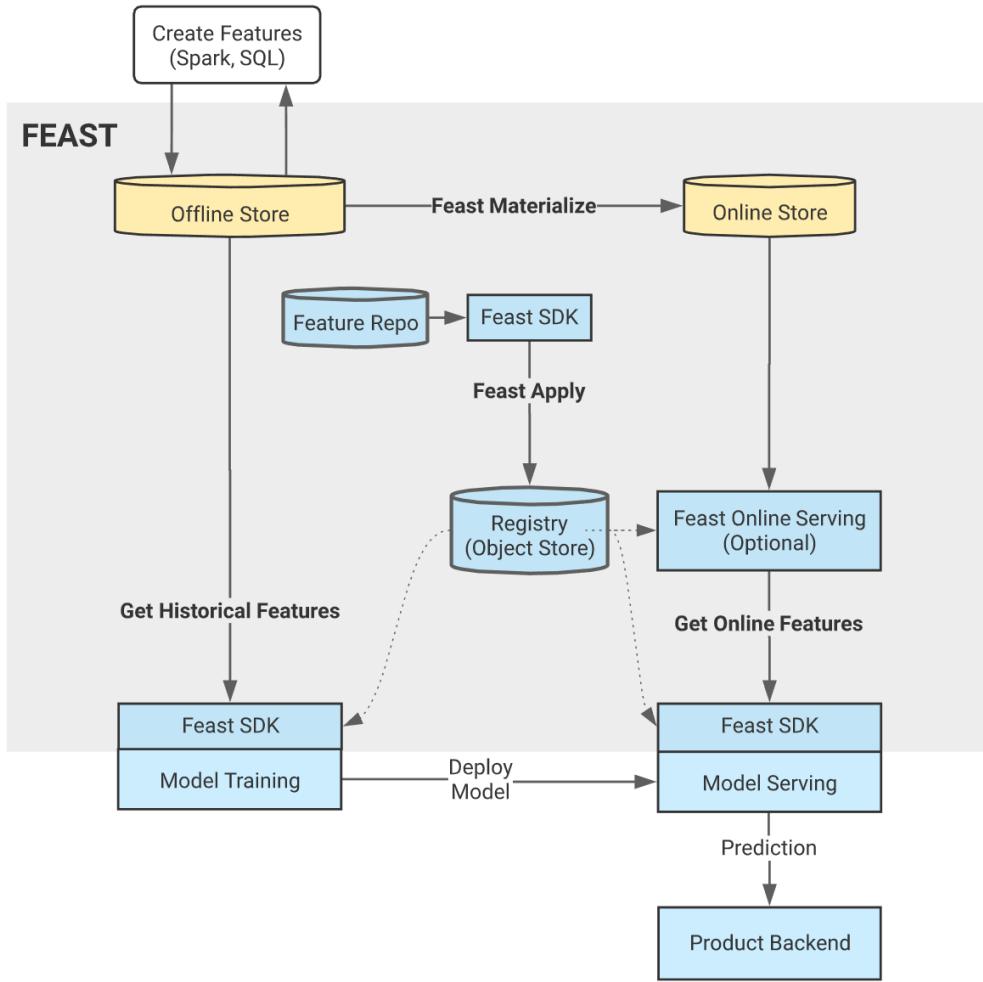
# Feature Engineering Concept



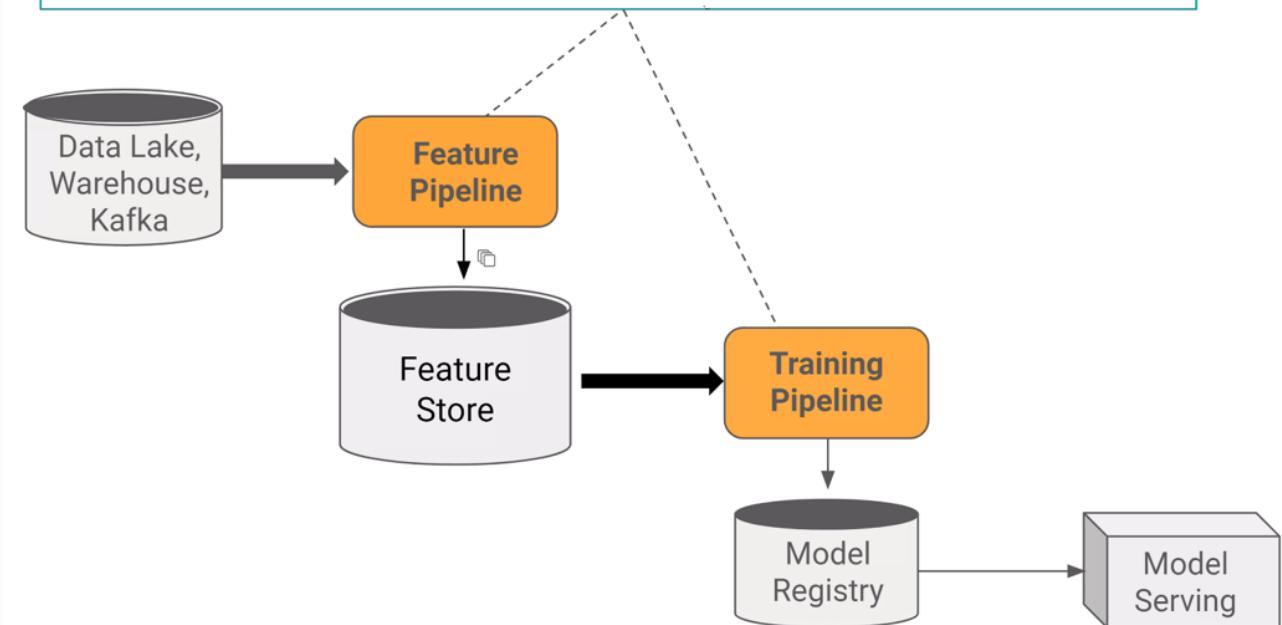
# MLOps for architect- Governance



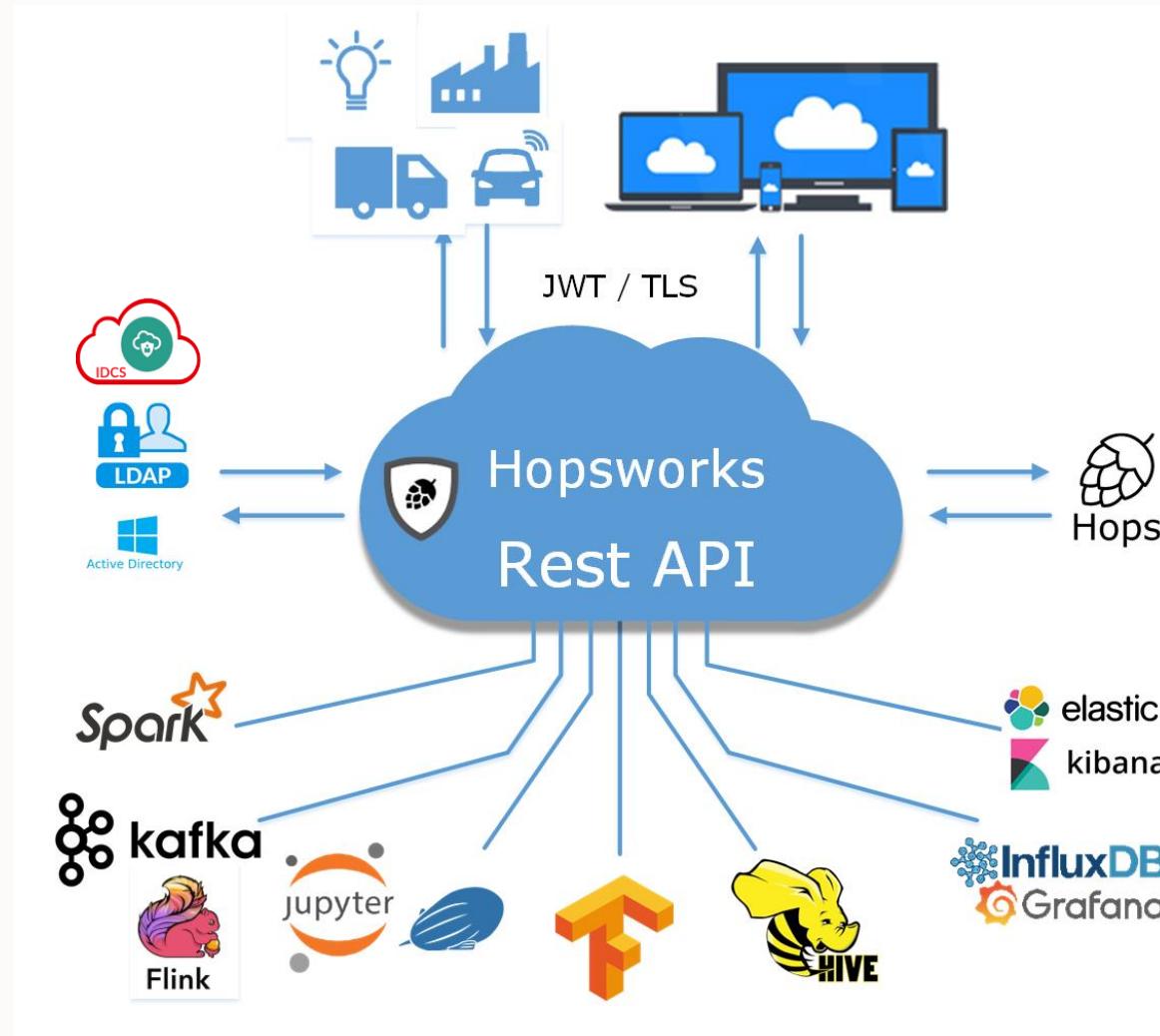
# Open Source Feature Store: FEAST vs HopsWorks



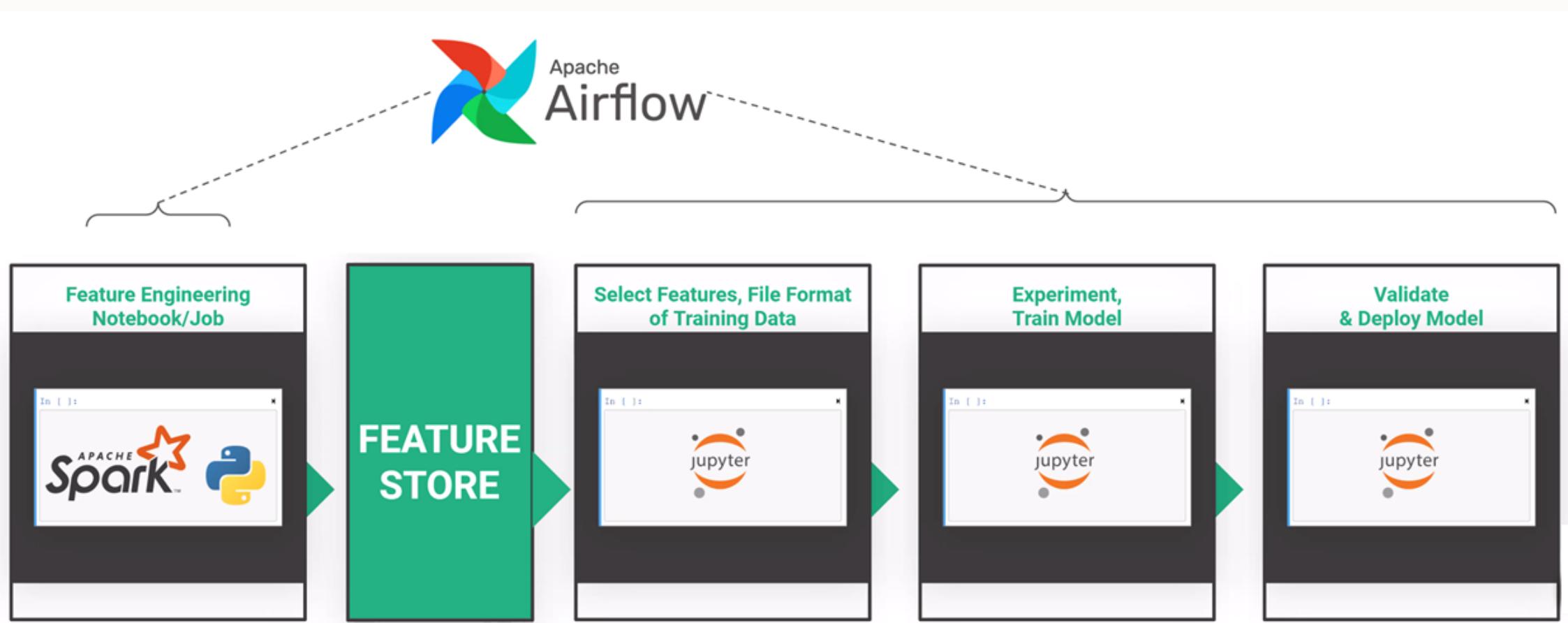
**Orchestrator:** Airflow, Github Actions, Jenkins  
**Actions:** Code commit, New data, time trigger (e.g., daily)



# Open Source HopsWorks

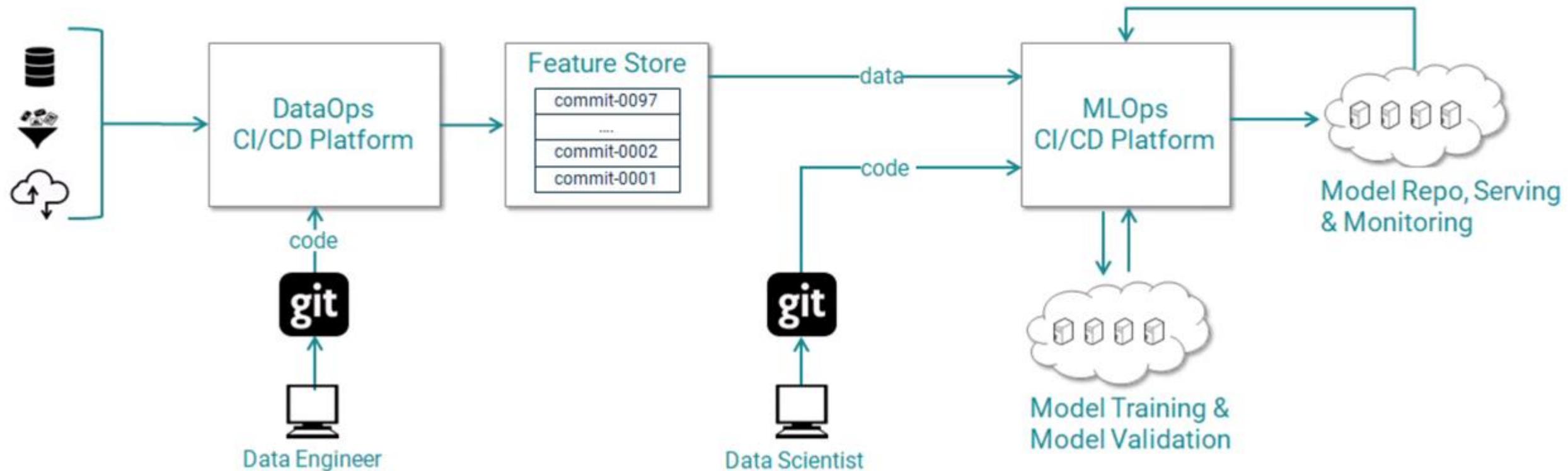


# Feature and Training Pipeline

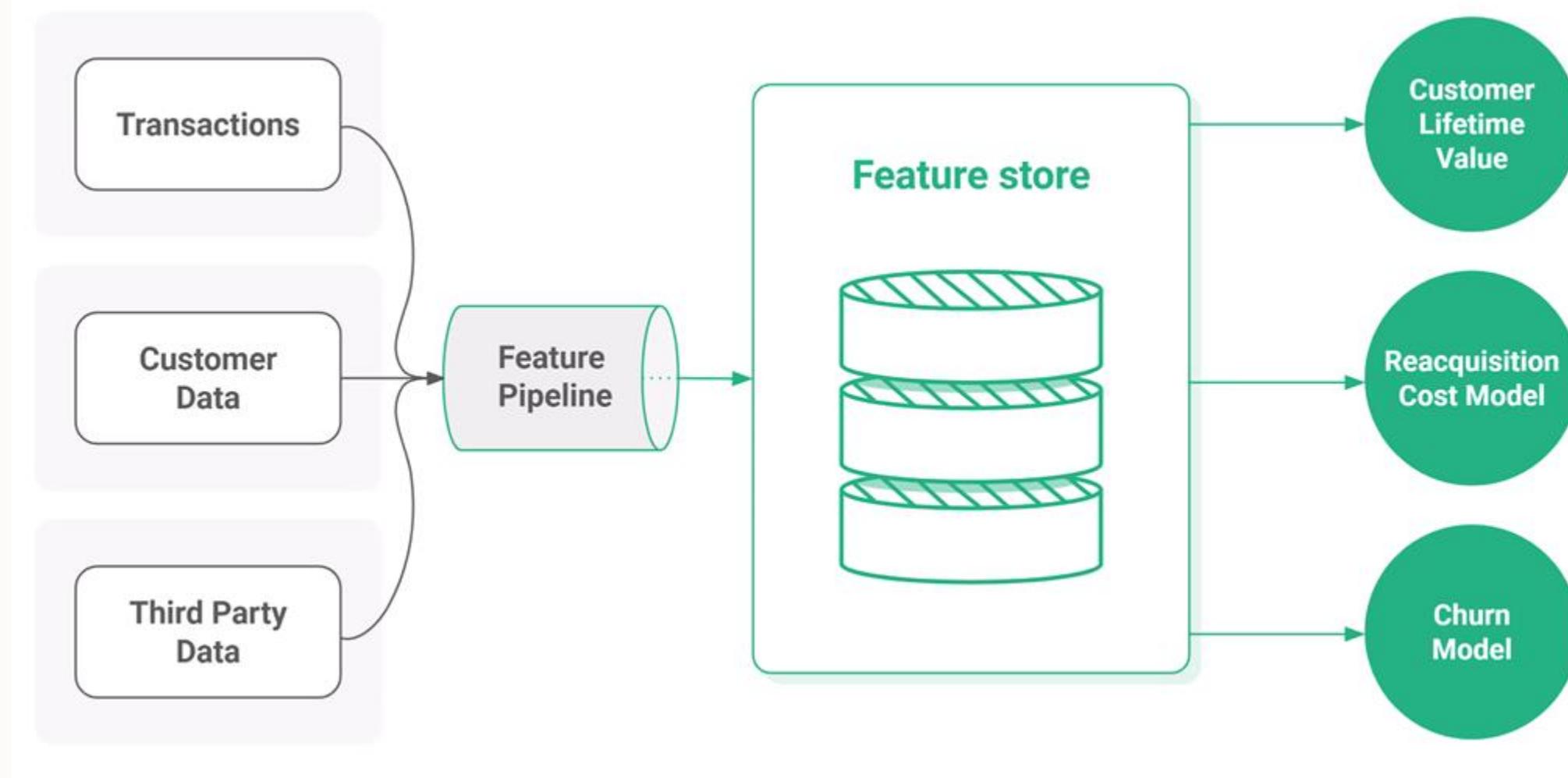


# Feature and Training Pipeline

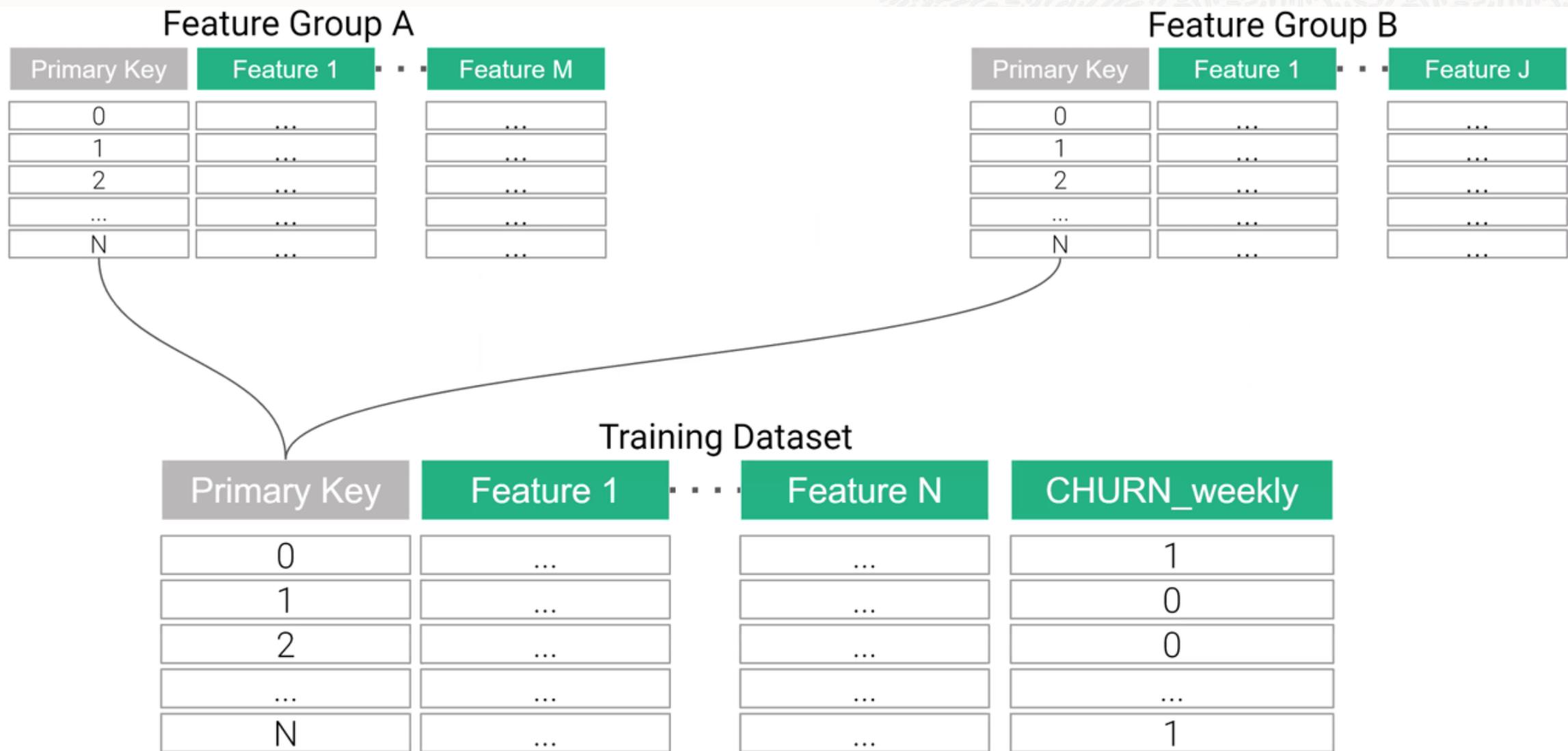
1. Data
2. Develop/Test Feature Pipelines
3. Develop Model
4. Train/Validate Model
5. Deploy/Monitor



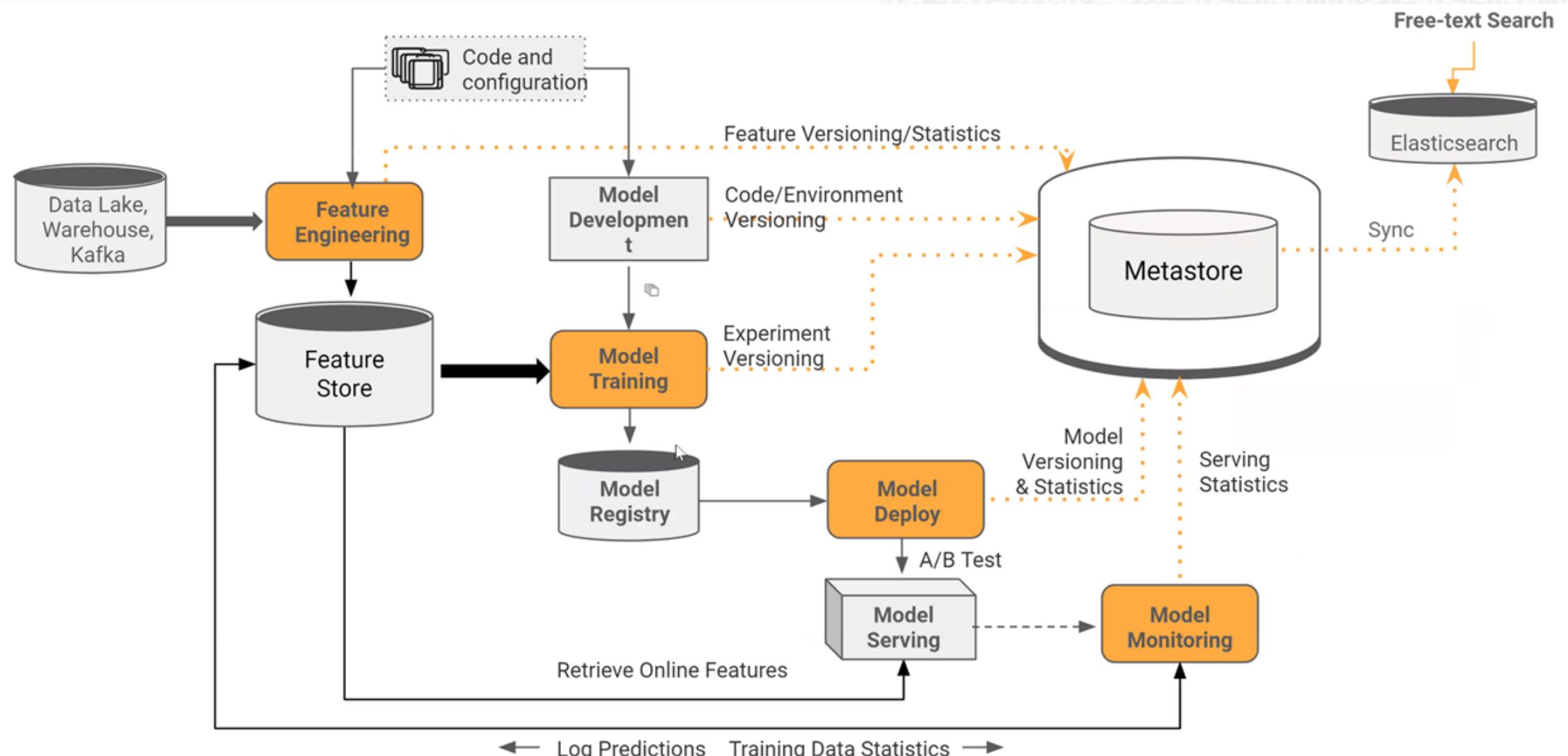
# Feature Store for ML



# Training Dataset

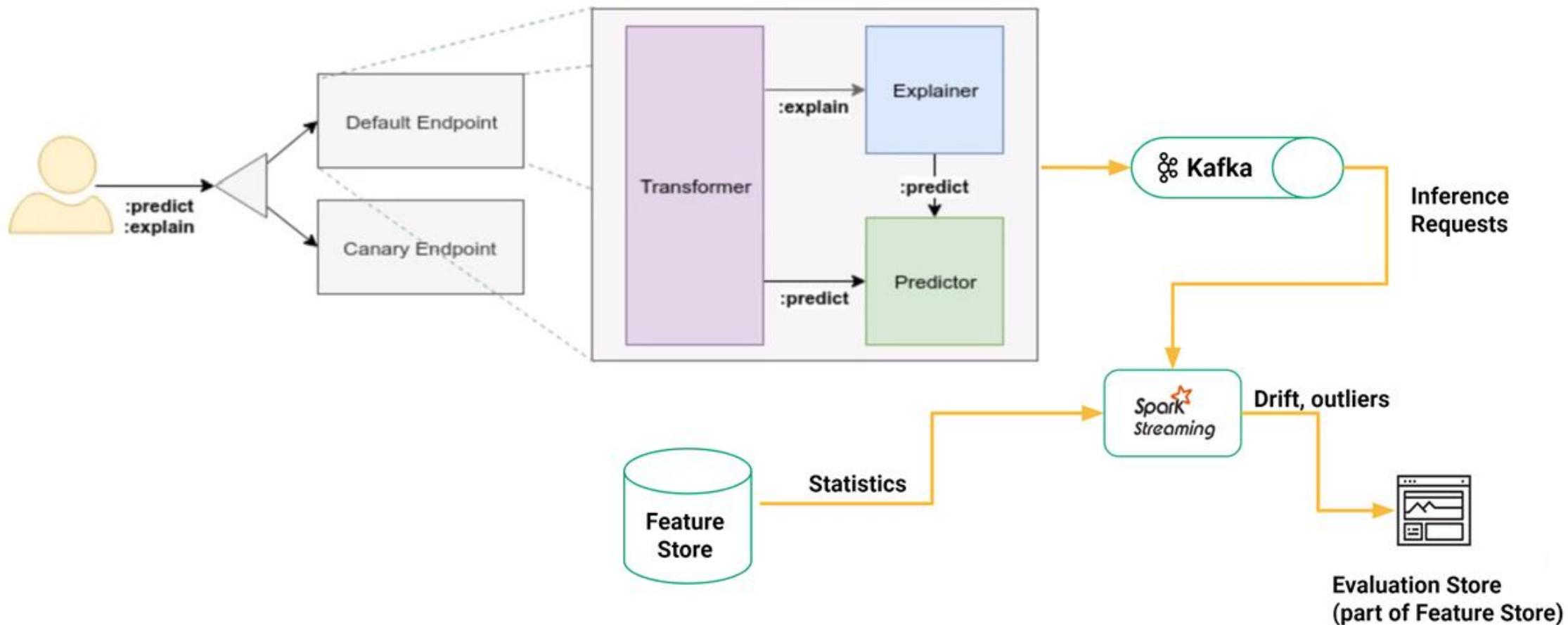


# MLOps with HopsWorks Feature Store



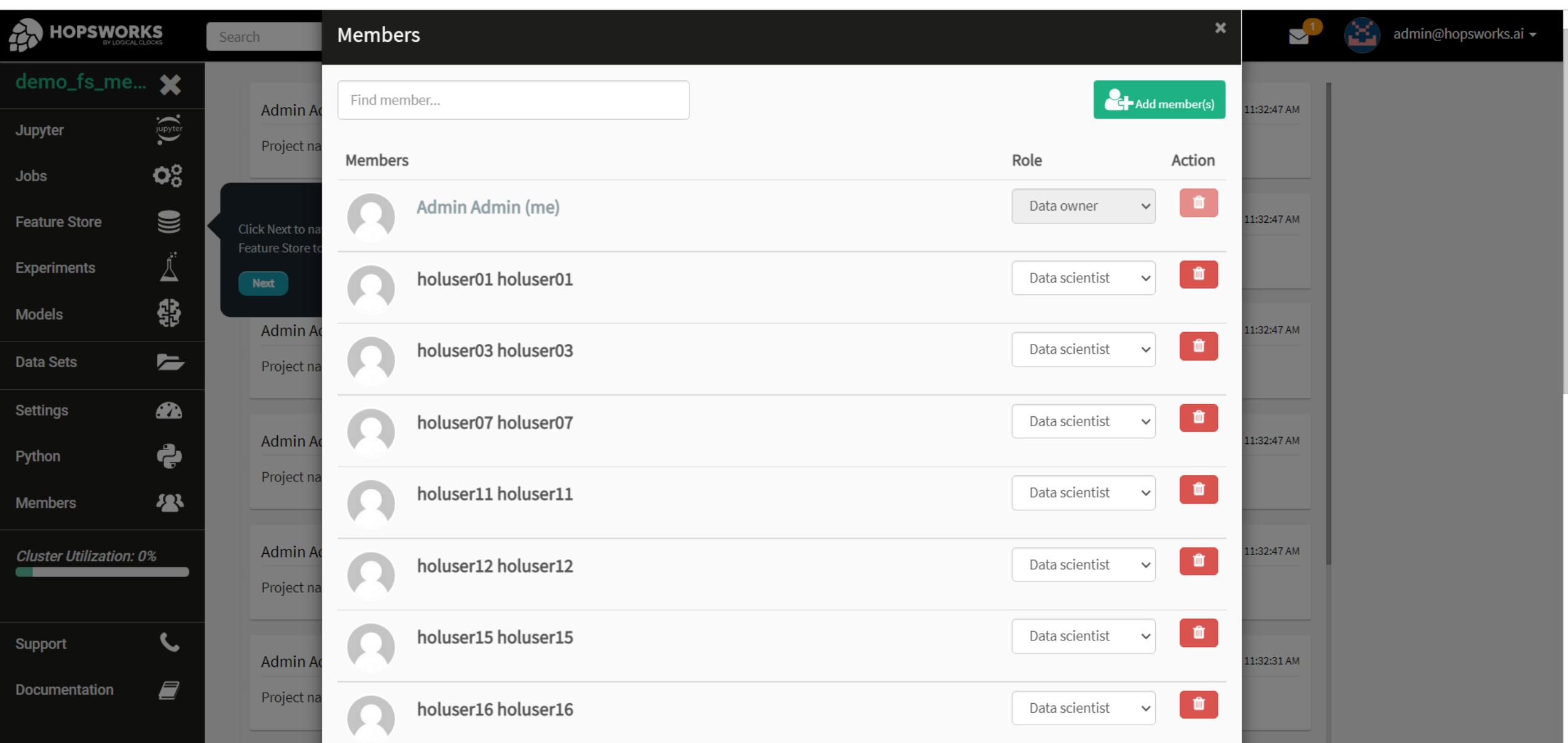
# Model Serving in Hopsworks

Source: <https://github.com/kubeflow/kfserving/tree/master/docs>



# Open Source Feature Store Example: demo\_fs\_meb10000

---







**HOPSWORKS** BY LOGICAL CLOCKS

Search 

demo\_fs\_me... 

Jupyter 

Jobs 

Feature Store 

Experiments 

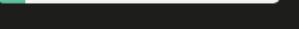
Models 

Data Sets 

Settings 

Python 

Members 

Cluster Utilization: 0% 

Support 

Documentation 

Search: By name type, owner (case sensitive)

Created on: 2/20/2021 to 6/23/2021 Jobs/page: 10

Name	Created on	Type	Owner	Ongoing	Actions
featurestore_tour_job	Jun 14, 2021 11:32:47 AM	SPARK	Admin Admin	0	     More 

Apps free computer ebo... microsoft openkm bigdata olivier dev amazon mcentric Bookmarks perso oracle teradata gartner » Other bookmarks Reading list

 **HOPSWORKS** BY LOGICAL CLOCKS  

**demo\_fs\_me...** 

**Jupyter** 

**Jobs** 

**Feature Store** 

**Experiments** 

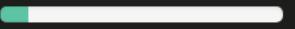
**Models** 

**Data Sets** 

**Settings** 

**Python** 

**Members** 

**Cluster Utilization: 0%** 

**Support** 

**Documentation** 

**Job definition** 

**Import job**

Job name - featurestore\_tour\_job

Job type - SPARK

App file - hops-examples-featurestore-tour-2.2.0.jar

Job details

Main class name  io.hops.examples.featurestore\_tour.Main

Default arguments  --input TestJob/data

Configure and create

Experiment  Parallel Experiments  Distributed Training  Spark (Static)  Spark (Dynamic)

Driver memory (MB): 1024

Executor memory (MB): 2048

Executor virtual cores: 1

Driver virtual cores: 1

The screenshot shows the HOPSWORKS interface. On the left is a sidebar with navigation links: Jupyter, Jobs, Feature Store, Experiments, Models, Data Sets (highlighted in green), Settings, Python, Members, Cluster Utilization (0%), Support, and Documentation. On the right is a main content area titled 'DataSets / Jupyter'. It features a search bar and a table with columns: Type, Name, Owner, Last modified, and File size. The table lists various datasets:

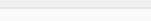
Type	Name	Owner	Last modified	File size
file	README.md	Admin Admin	Jun 14, 2021 11:32:26 AM	0.2KB
folder	aws	Admin Admin	Jun 14, 2021 11:32:35 AM	-
folder	azure	Admin Admin	Jun 14, 2021 11:32:35 AM	-
folder	databricks	Admin Admin	Jun 14, 2021 11:32:35 AM	-
folder	datasets	Admin Admin	Jun 14, 2021 1:03:52 PM	-
folder	delta	Admin Admin	Jun 14, 2021 11:32:35 AM	-
folder	feature-bias	Admin Admin	Jun 14, 2021 1:22:18 PM	-
folder	gcp	Admin Admin	Jun 14, 2021 11:32:35 AM	-
folder	hsfs	Admin Admin	Jun 14, 2021 11:32:39 AM	-
folder	images	Admin Admin	Jun 14, 2021 11:32:44 AM	-
folder	petastorm	Admin Admin	Jun 14, 2021 11:32:44 AM	-
folder	snowflake	Admin Admin	Jun 14, 2021 11:32:45 AM	-



Search



holuser20@oracle.com ▾

demo_fs_me... 		pypi	notebook	6.1.4	PIP	INSTALLED	
Jupyter		pypi	numpy	1.19.2	PIP	INSTALLED	
Jobs		pypi	nvidia-ml-py3	7.352.0	PIP	INSTALLED	
Feature Store		pypi	oauth2client	4.1.3	PIP	INSTALLED	
Experiments		pypi	oauthlib	3.1.0	PIP	INSTALLED	
Models		pypi	oci	2.40.0	PIP	INSTALLED	
Data Sets		conda	olefile	0.46	CONDA	INSTALLED	
Settings		pypi	opencv-python	4.2.0.32	PIP	INSTALLED	
Python		conda	openssl	1.1.1h	CONDA	INSTALLED	
Members		pypi	opt-einsum	3.3.0	PIP	INSTALLED	
<i>Cluster Utilization: 26%</i>		pypi	orderedmultidict	1.0.1	PIP	INSTALLED	
Support		pypi	oscrypto	1.2.1	PIP	INSTALLED	
Documentation		pypi	packaging	20.9	PIP	INSTALLED	

 5\_online\_serving....ipynb

 1\_create\_feature....ipynb

 stores data-set.csv

Show all



## User Administration



## New Users



## Spam Requests



## Register User



## Grafana



## Services



Projects Management



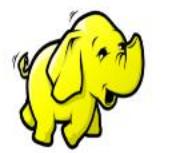
## Ndb Backups



## Master encryption key



### Edit variables



YARN(jobs) Admin



HDFS Admin



## Jupyter Notebook Servers



## Manage cluster nodes



## Manage Conda

# Thank You

---

**DevRel Team**





ORACLE

O