

# Accelerate Data for AI Models

**Victor Alberto Olmedo**

Principal Analytics and AI FSA  
Pure Storage

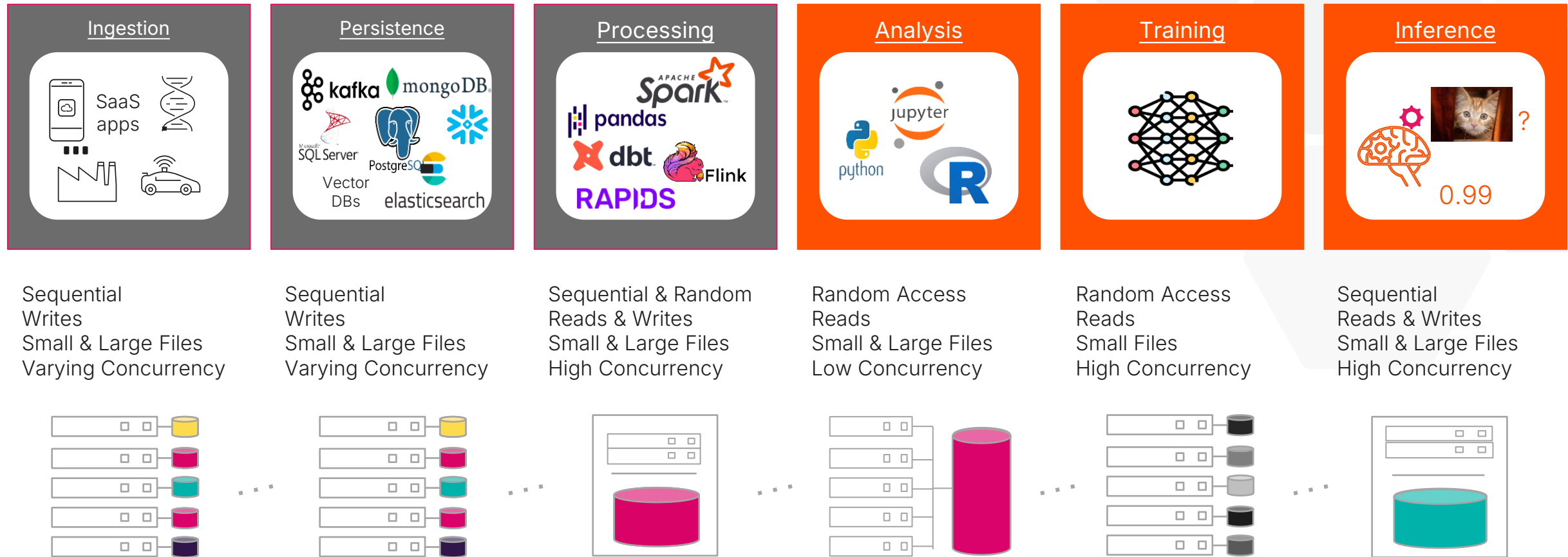


# Data Pipelines Source AI Pipelines

The view inside a typical AI pipeline

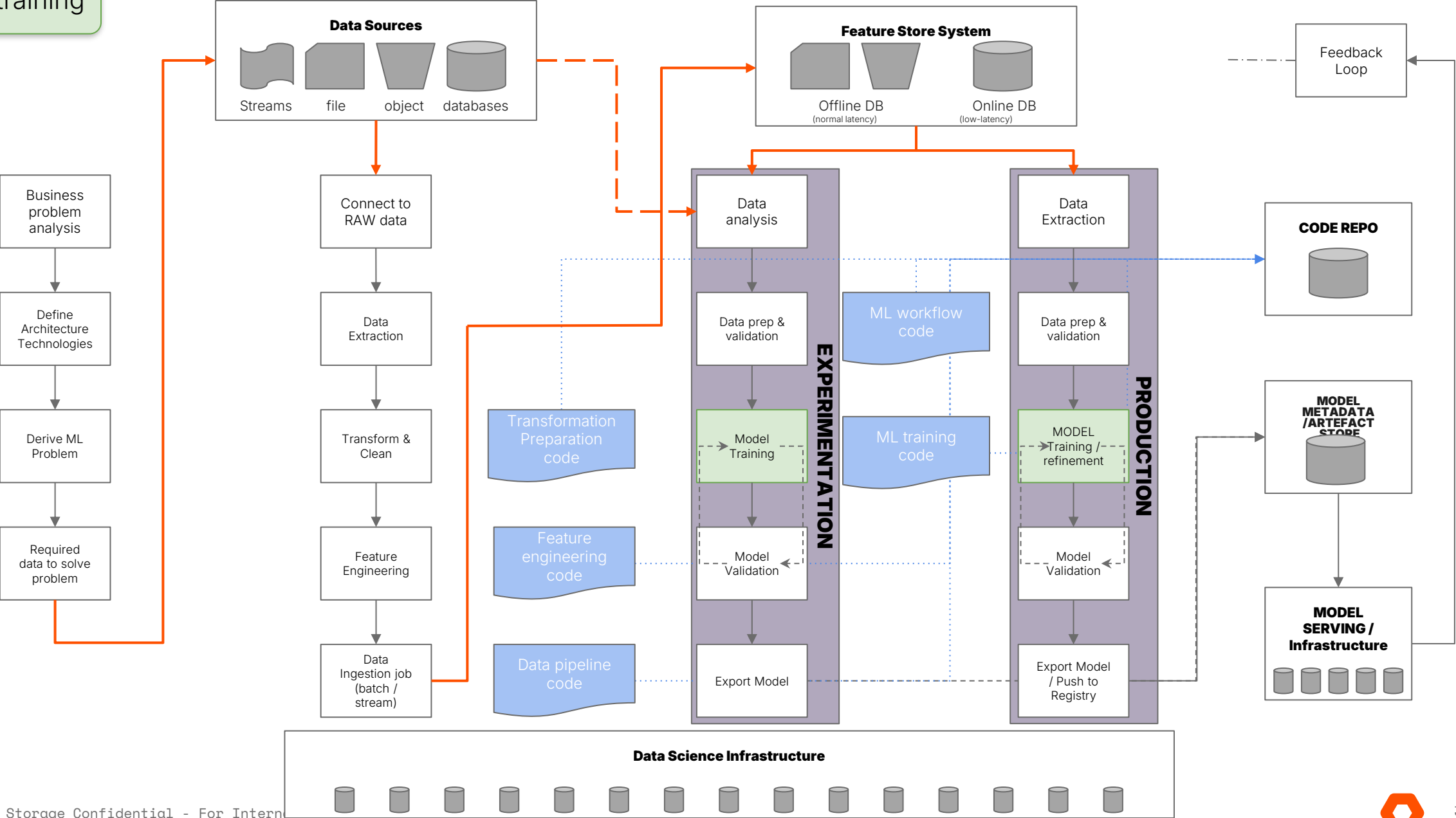
## Data curation

## AI training and inference

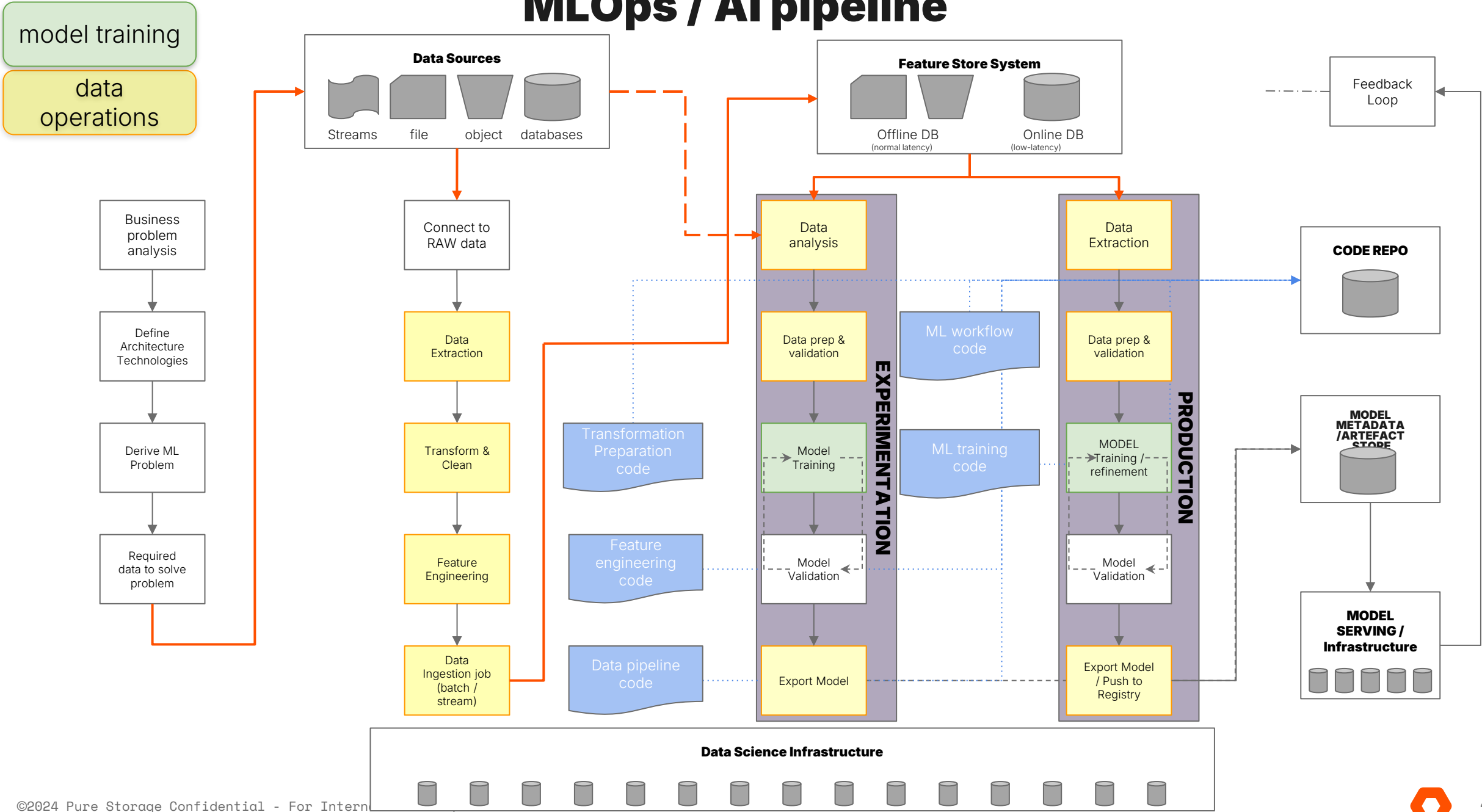


# MLOps / AI pipeline

model training



# MLOps / AI pipeline

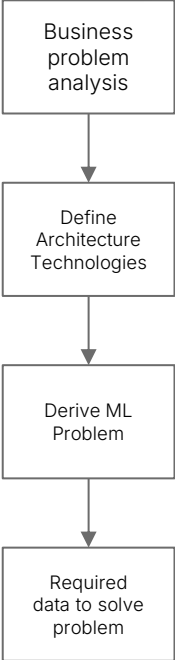


# MLOps / AI pipeline

model training

data operations

storage

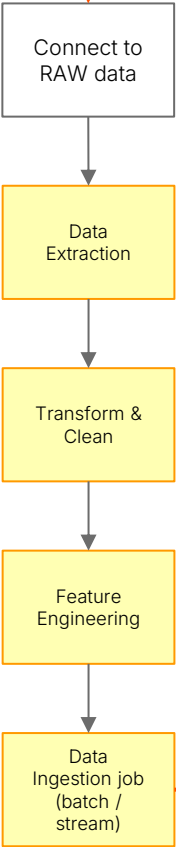


Data Sources

Streams file object databases

Feature Store System

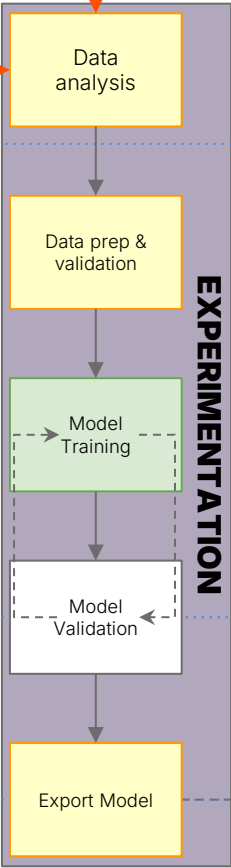
Offline DB (normal latency) Online DB (low-latency)



Transformation Preparation code

Feature engineering code

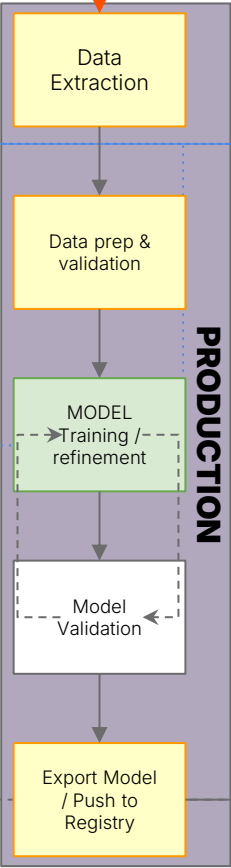
Data pipeline code



EXPERIMENTATION

ML workflow code

ML training code



PRODUCTION



CODE REPO

MODEL METADATA / ARTEFACT STORE

MODEL SERVING / Infrastructure

Data Science Infrastructure

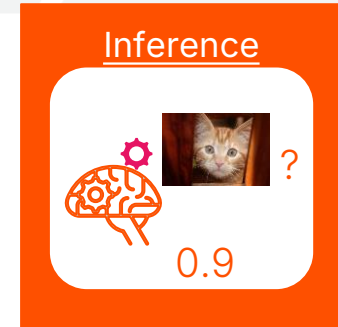
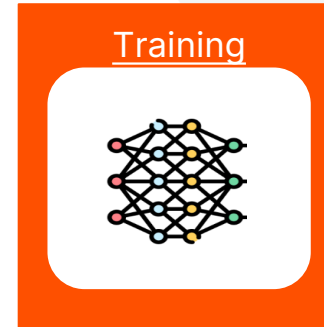
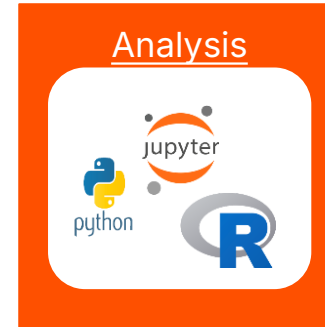
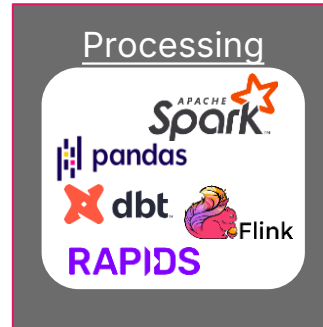
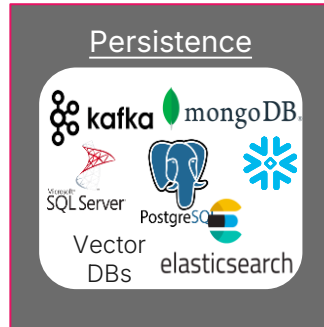
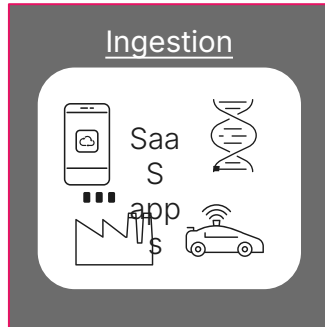
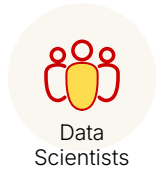


# Accelerate and Unify Data

Single Storage Platform for Unstructured Data

## Data curation

## AI training and inference



The portworx logo, featuring a stylized 'P' and the text 'portworx by Pure Storage'.

**DBaaS Platform**  
One Platform for Any Database

**Data Management**  
Run Kubernetes in Production

**Data Protection**  
Protect Apps Anywhere

**The #1 Kubernetes Data Platform**



