MLflow Workshop

Spyros Stasis

Workshop Outline

MLflow Overview

- Project Lifecycle
- MLflow components

Hands on examples:

- Boston housing prices regression
- MNIST classification

Reproducible Research Tools

Machine Learning Project;

- Stages:

Raw Data → Data Preparation → Model Development → Deployment

- Considerations:
  1. Data
  2. Environment
  3. Parameters
  4. Deployment

Tools for tracking/reproducible machine learning pipelines:

- Uber Michelangelo, Facebook FBLearner, Tensorflow TFX
- Data Version Control (DVC)
- Pachyderm

What is MLflow?

- Open source project developed by Databricks
- Main principles:
  - Agnostic
  - Universal
  - Ease of use
  - Flexibility/Scalability

Open source machine learning platform

- Works with any ML library, algorithm, language, etc
- *Open interface* design (use with any code you already have)

mlflow Tracking
Record and query experiments: code, data, confs, results

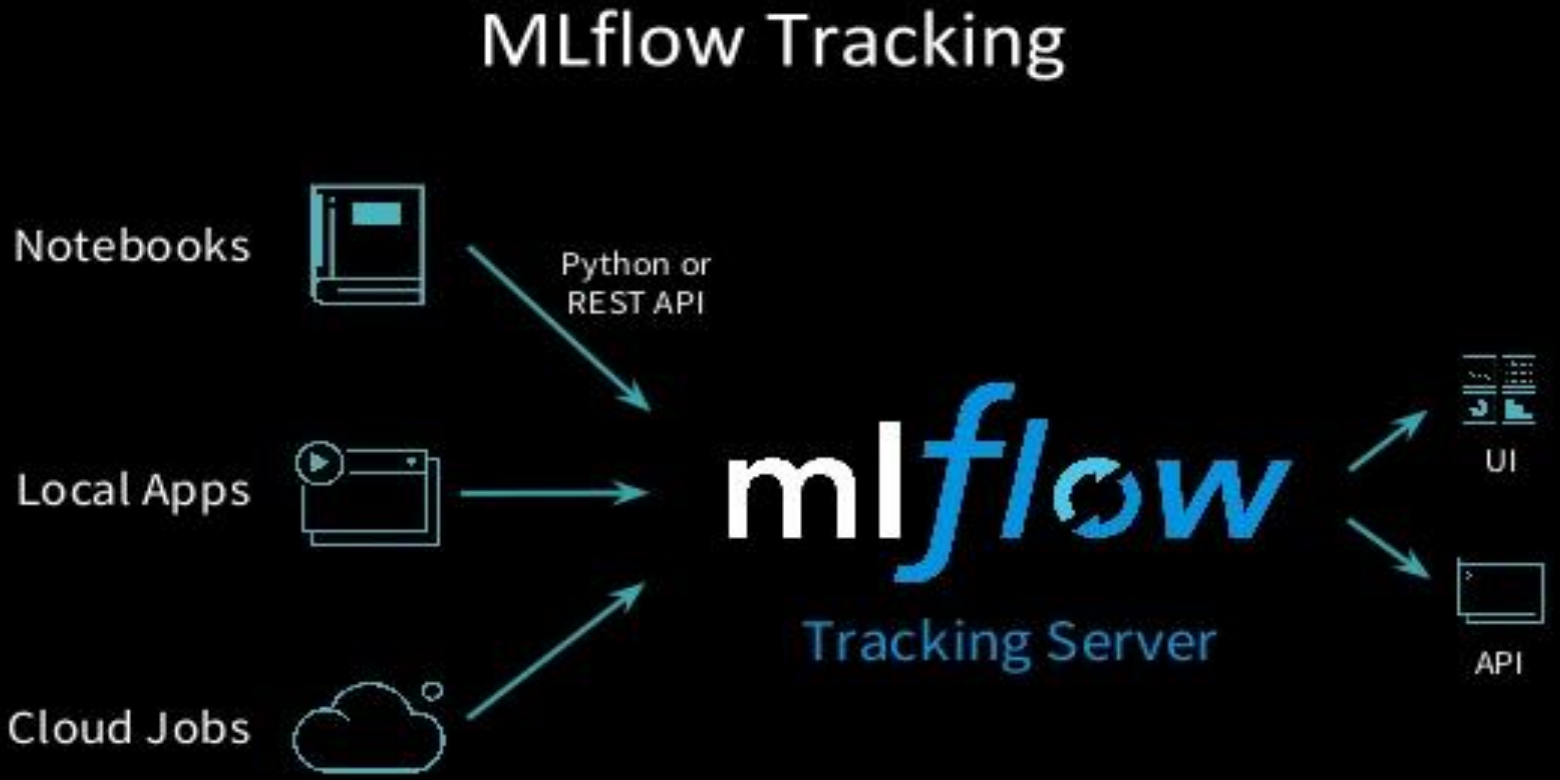
mlflow Projects
Packaging format for reproducible runs and workflows

mlflow Models
General format that standardizes deployment paths

new
mlflow Model Registry
Centralized model management, review & sharing

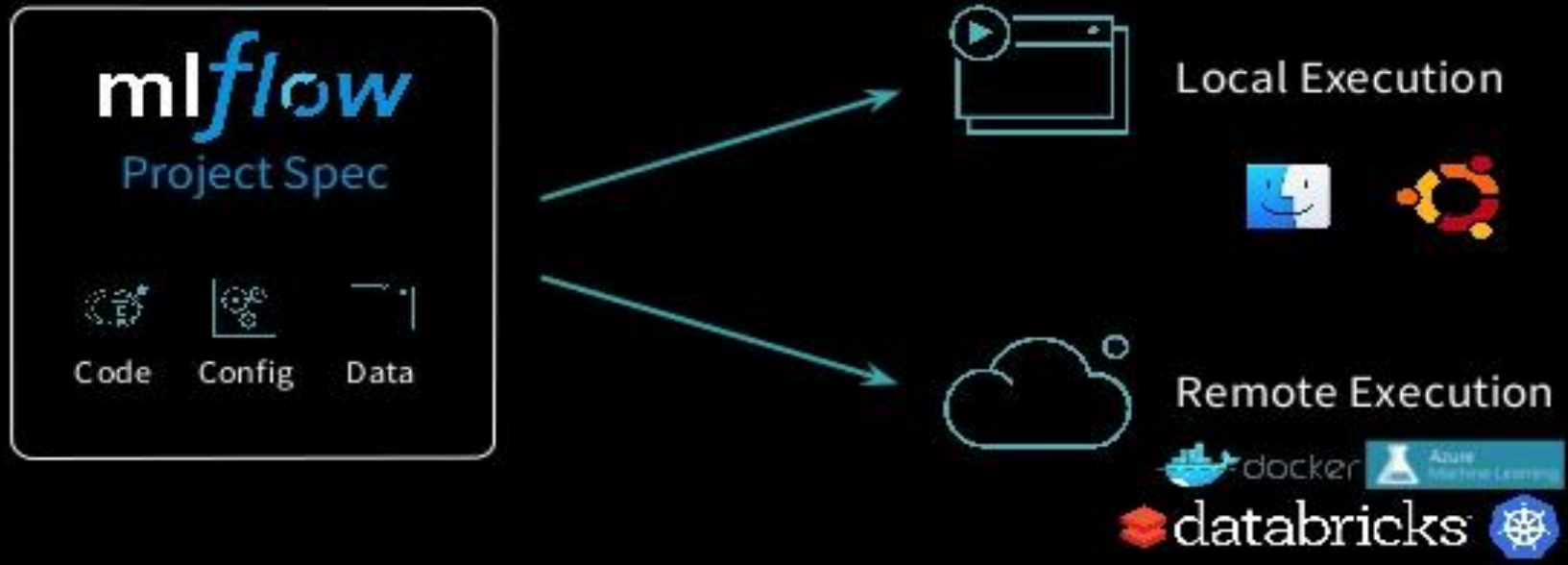
databricks

Slide taken from MLflow: A Platform for Production Machine Learning presentation by Matei Zaharia, (link)

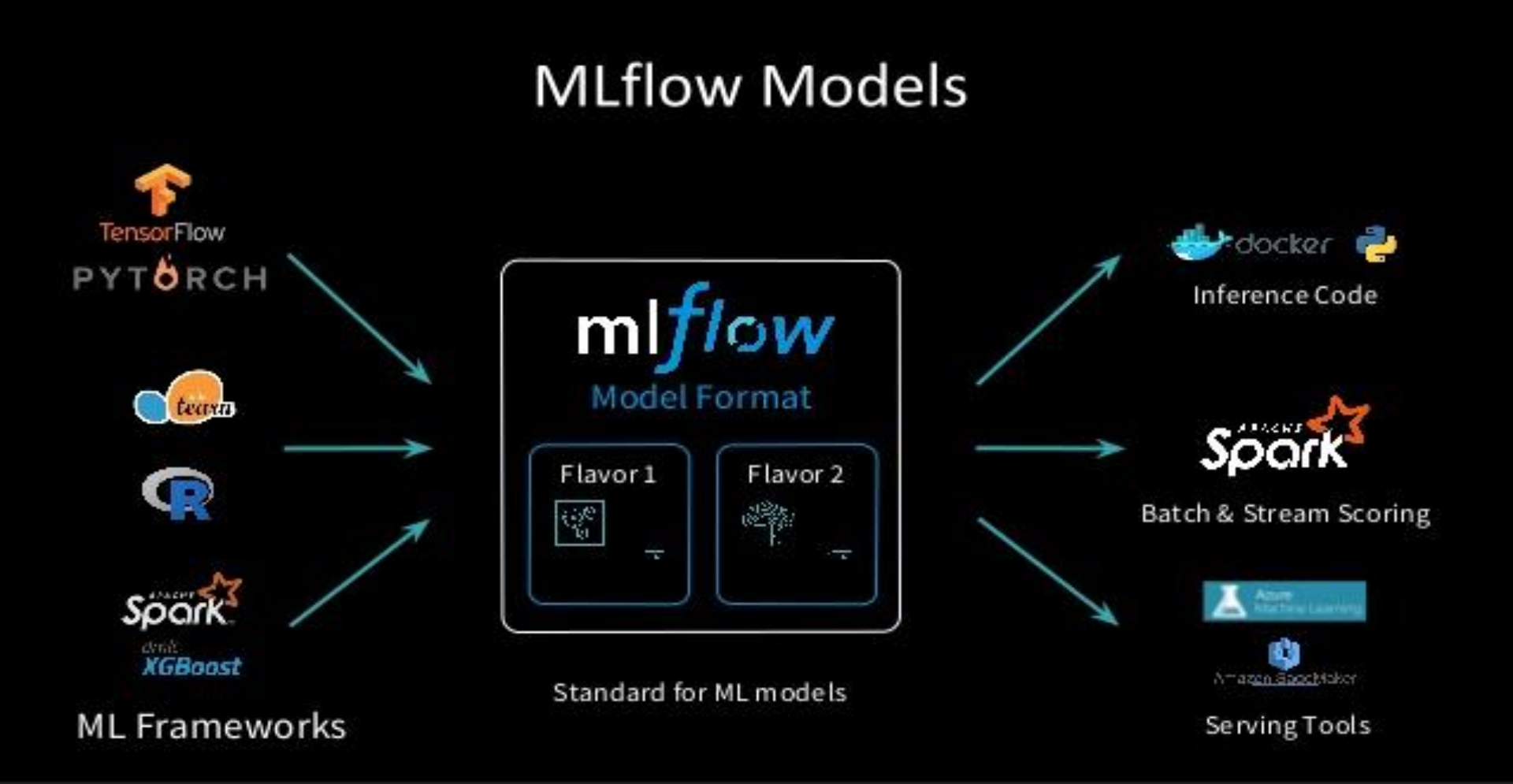


MLflow: Infrastructure for a Complete Machine Learning Life Cycle with Mani Parkhe and Tomas Nykodym (link)

MLflow Projects



MLflow: Infrastructure for a Complete Machine Learning Life Cycle with Mani Parkhe and Tomas Nykodym (link)



MLflow: Infrastructure for a Complete Machine Learning Life Cycle with Mani Parkhe and Tomas Nykodym (link)

MLflow Model Registry

GitHub-like environment for organizing & reviewing models

databricks

Slide taken from MLflow: A Platform for Production Machine Learning presentation by Matei Zaharia, (link)

https://github.com/Pyrsos/qmul_mlflow

Boston housing regression

- Try out:
  - Different regression techniques:
    - Linear Regression
    - Lasso
  - Additional metrics:
    - Max Error
    - Median absolute error
  - Logging artifacts
    - Perform PCA on the input data
    - Store the image in the MLflow logs

MNIST example

- Try out:
  - Different model architectures:
    - Add/remove dense layers
    - Change layer sizes/activation
    - Try and keep track of these changes
  - Custom metrics:
    - Define and log custom metrics
    - Track the metric over time (over epochs/batches)

Resources

- Mlflow Infrastructure for a complete Machine Learning Lifecycel
- Accelerating the Machine Learning Lifecycle with mlflow
- Simplifying model management with mlflow

Thank you! Any questions?