MLflow: Platform for Machine Learning Lifecycle

Machine Learning Architects Basel

Bassem Ben Hamed

December 2022

ML ARCHITECTS BASEL

Agenda

- Introduction to MLflow
- Mlflow Components
- MLflow Tracking
- MLflow Projects
- MLflow Models
- MLflow Model Registry

ML ARCHITECTS BASEL

Introduction to MLflow

Open machine learning platform

- Works with popular ML library and language
- Runs the same way anywhere (e.g. any cloud or locally)
- Designed to be useful for all persons on your organisation
- Simple. Modular. Easy-to-use
- Offers positive developer experience to get started!

MLflow Components

Record and query experiments: code, data, config, and results

Package data science code in a format that enables reproducible runs on many platform

Deploy machine learning models in diverse serving environments

Store, annotate and manage models in a central repository

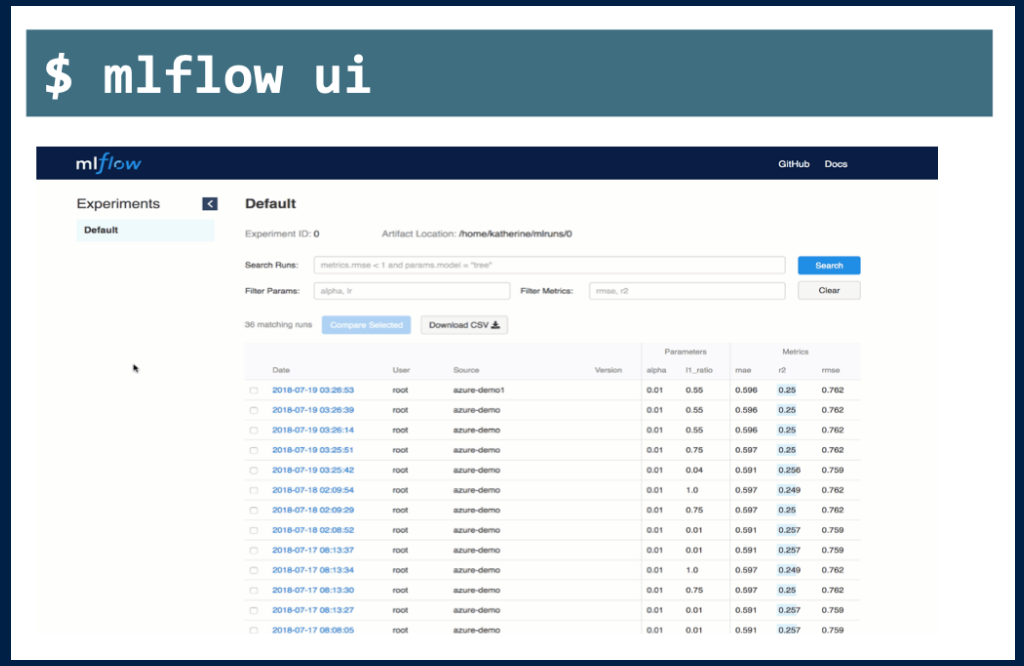
Key Concepts in MLflow Tracking

- Parameters: Key-value inputs to your code
- Metrics: numerics values (can update over time)
- Tags and Notes: information about a run
- Artifacts: files, data, and models
- Source: what code ran?
- Version: what of the code?
- Run: an instance of code that runs by Mlflow
- Experiment: {Run, …, Run}

Model Development with MLflow is Simple

```
import mlflow
data    = load_text(file)
ngrams  = extract_ngrams(data, N=n)
model   = train_model(ngrams,
                learning_rate=lr)
score   = compute_accuracy(model)
with mlflow.start_run():
  mlflow.log_param("data_file", file)
  mlflow.log_param("n", n)
  mlflow.log_param("learn_rate", lr)
  mlflow.log_metric("score", score)
  mlflow.sklearn.log_model(model)
```

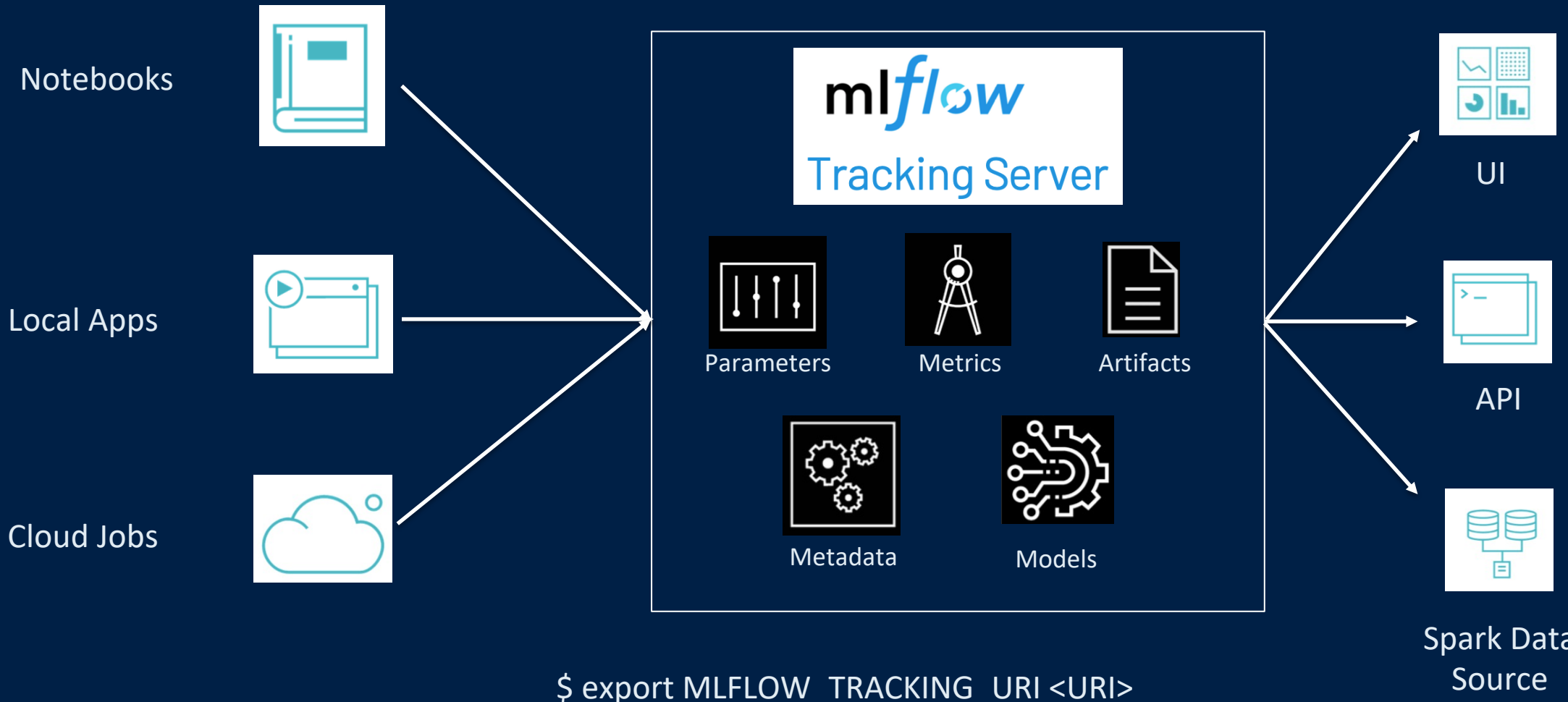
```
$ mlflow ui
```



Track parameters, metrics, output files and code version

ML ARCHITECTS BASEL

MLflow Tracking



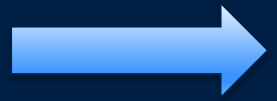

```
$ export MLFLOW_TRACKING_URI <URI>
  mlflow.set_tracking_uri(URI)
```

ML ARCHITECTS BASEL

MLflow Projects Motivation

Diverse set of tools

Diverse set of environments

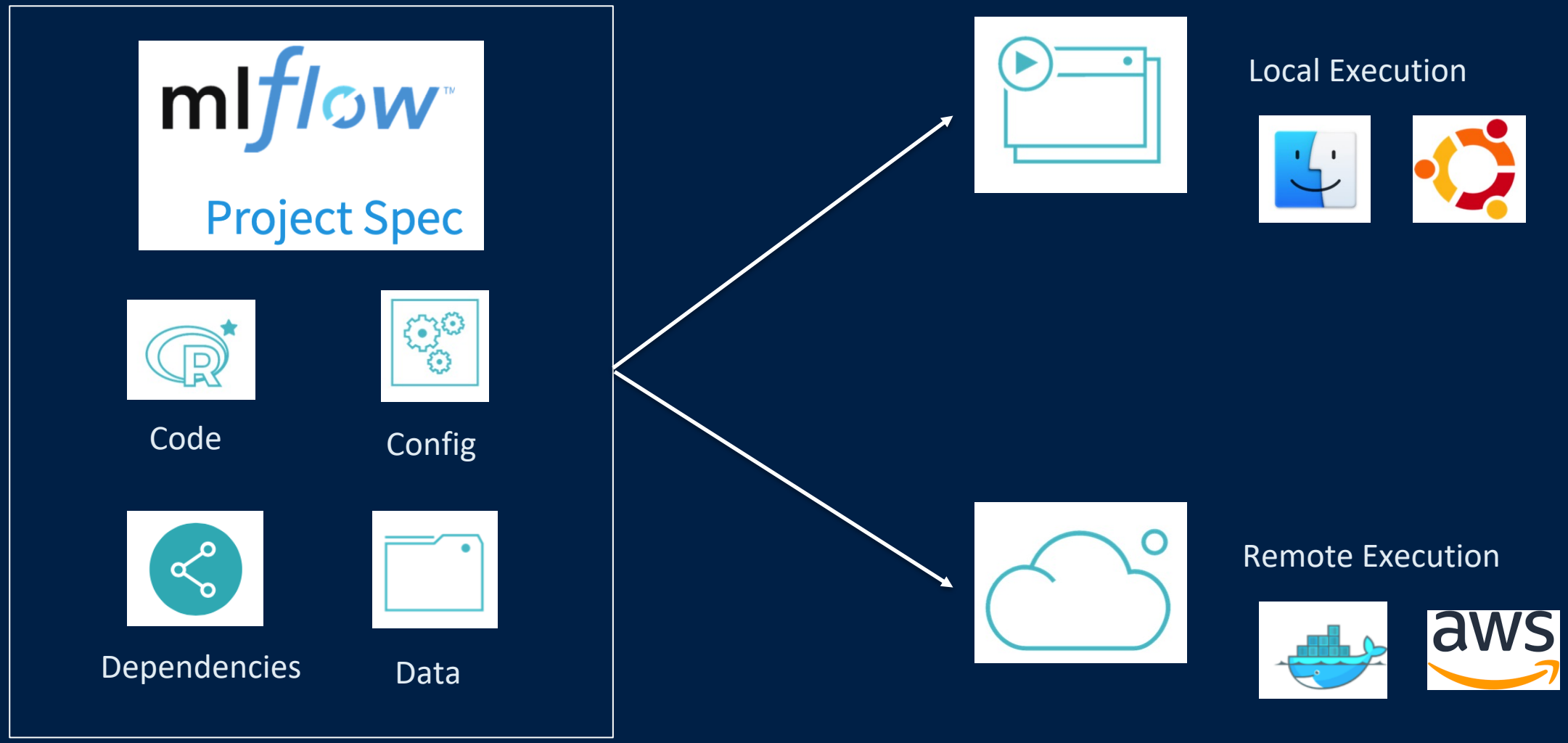


mlflow™ Projects

Package data science code in a format that enables reproducible runs on any plateform

Challenge:
ML results difficult to reproduce

MLflow Projects



MLflow Projects

Packaging format for reproducible ML runs

- Any code folder or Dagshub repository MLProject file with project configuration

Defines dependencies for reproducibility

- Conda dependencies can be specified in MLProject
- Reproducible in any environment

Execution API for running projects

- Python / R
- Supports local and remote execution
  - mlflow run -help
  - mlflow run https://dagshub.com/mlflow-project
  - mlflow run (<project_uri>, parameters = {})

MLflow Models

ML Frameworks

N x M Comination of Model support for all Serving tools

Inference Code

Batch and Stream Scoring

Serving Tools

MLflow Models

Packaging format for ML Models

- Any directory with MLmodel file

Defines dependencies for reproducibility

- Conda environment can be specified in MLmodel configuration

Model creation and loading utilities

- Mlflow.<model_flavor>.save_model(…) or log_model(…)
- mlflow .<model_flavor>.load_model(…)

Deployment APIs

- Python / R
- mlflow models [OPTIONS] COMMAND [ARGS] …
  - mlflow models serve [OPTIONS [ARGS] …
  - mlflow models predict [OPTIONS [ARGS] …

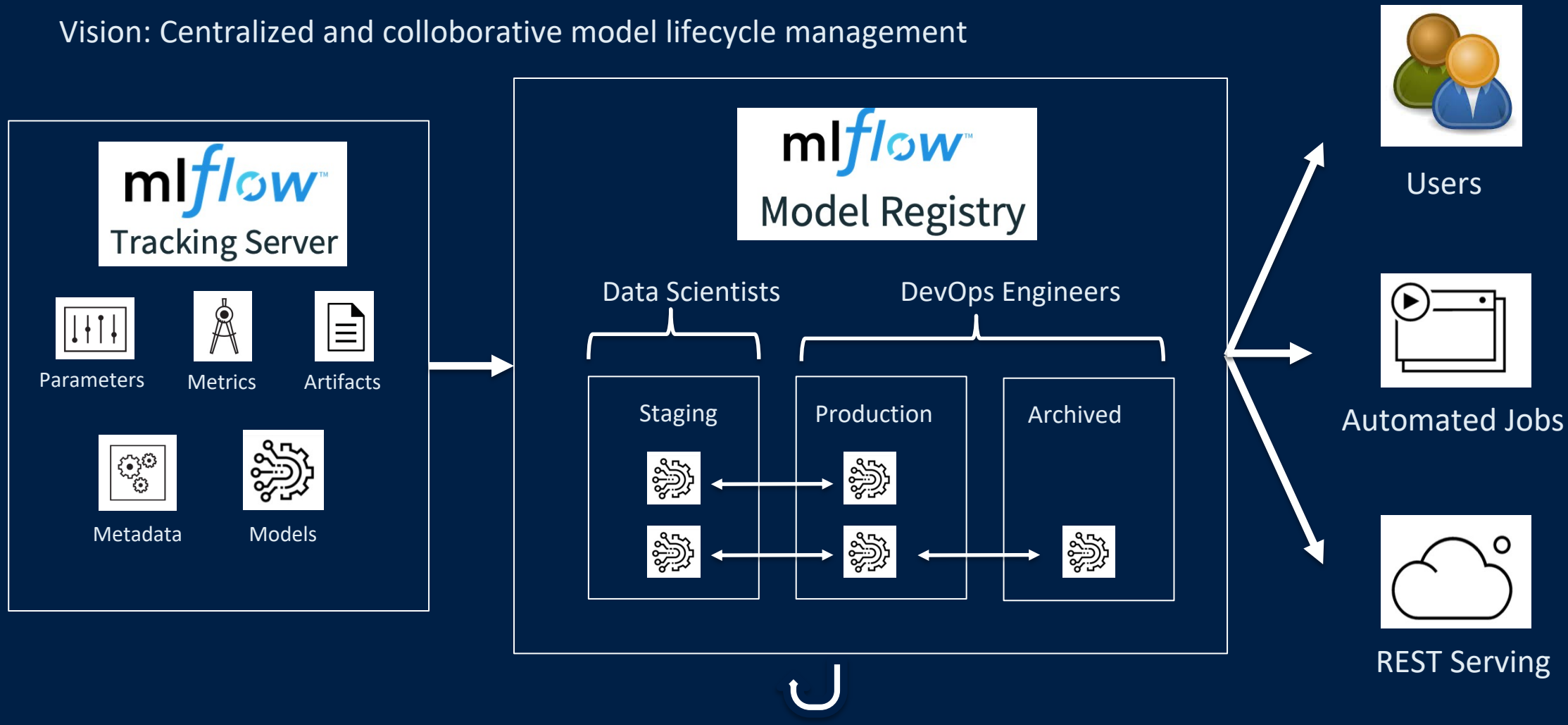
Model Management Problem

When you work in a large organization with many models, many data teams, management becomes a major challenge:

- Where can I find the best version of this model?
- How was this model trained?
- How can I track docs for each model?
- How can I review models?
- How can I integrate with CI/CD?

MLflow Model Registry

Vision: Centralized and colloborative model lifecycle management



MLflow Model Registry

- Repository of named, versioned models with comments and tags
- Track each model's stage: none, staging, production, or archived
- Easily inspect a specific version and its run info
- Easily load a specific version
- Provides model description, lineage and activities

Learning More About MLflow

- pip install mlflow – to get started
- Find documents and examples at mlflow.org
- Peruse code and contribute at github.com/mlflow/mlflow