



mlflow

ML Pipelines with MLflow

Concepcion Diaz
ML Technical Trainer



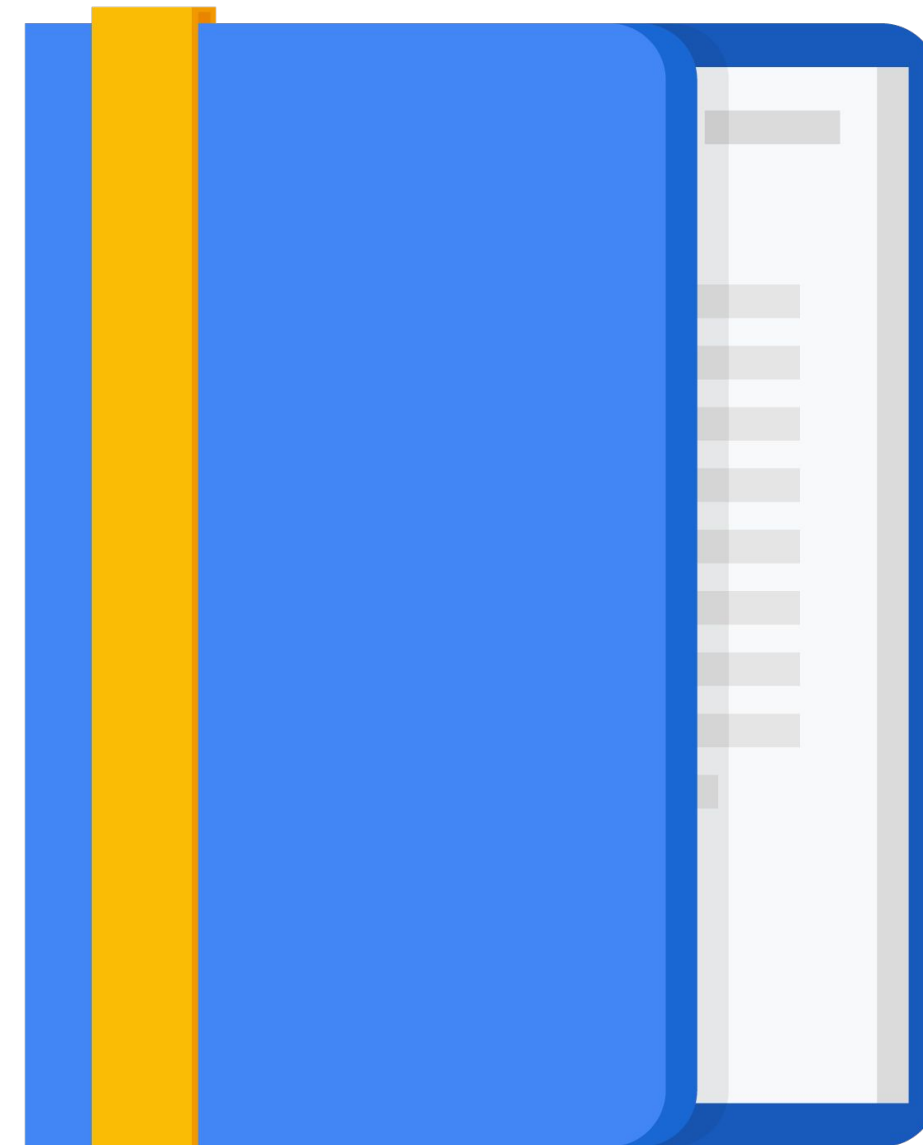
Agenda

Overview of ML development challenges

How MLflow tackles these challenges

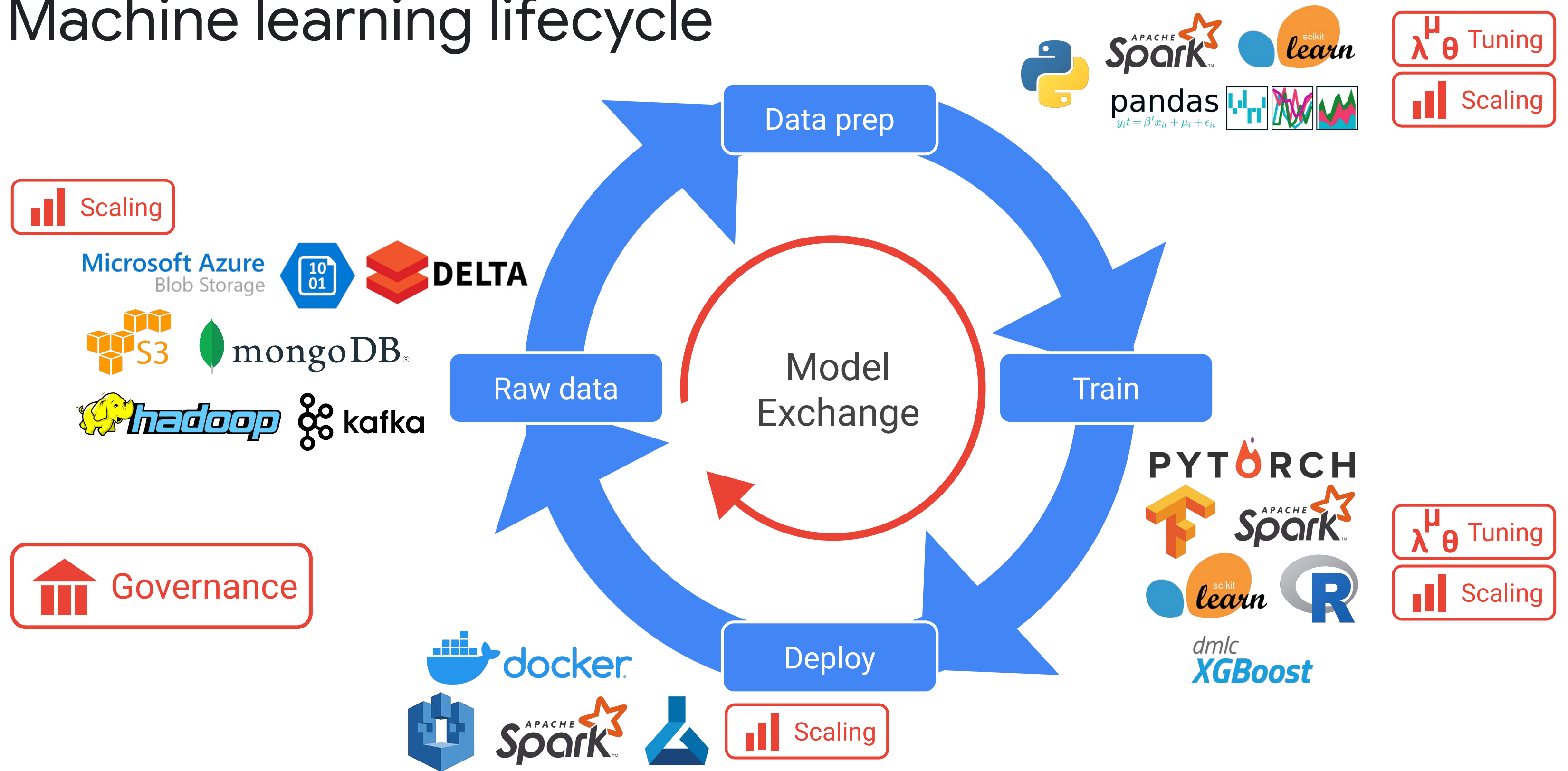
MLflow components

Demo



Machine learning
development is complex

Machine learning lifecycle

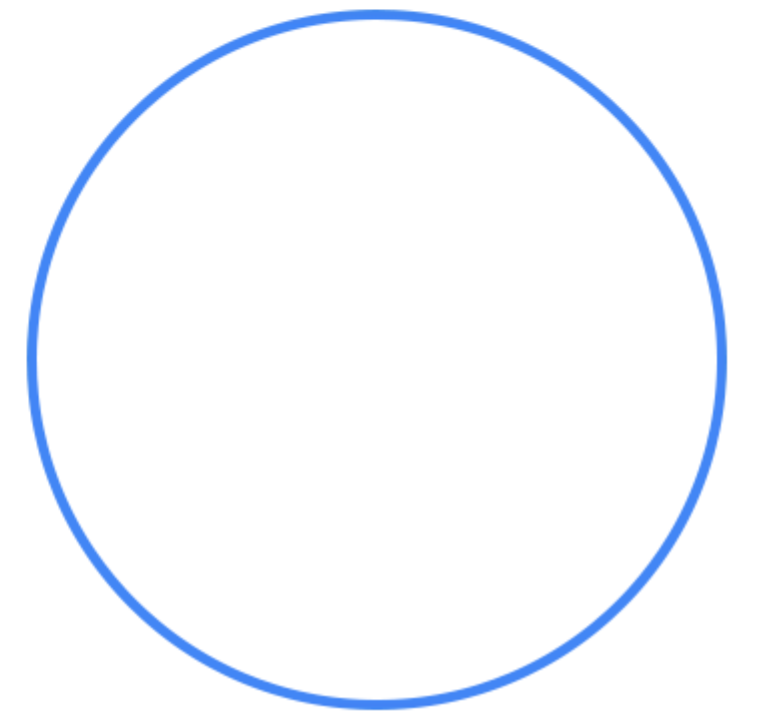


Custom machine learning platforms

Some big data companies
standardize the data prep/training/deploy loop.

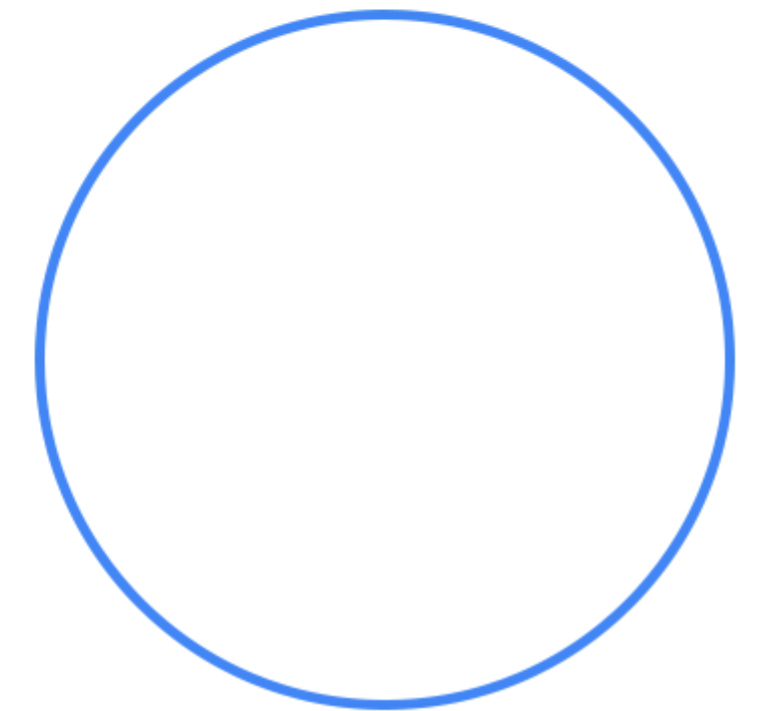
- Limited to a few algorithms or frameworks
- Tied to one company's infrastructure
- Out of luck if you leave the company...

Can we provide similar benefits in an open manner?



Introducing 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 orgs of 1 person or 1000+
- Simple. Modular. Easy-to-use
- Offers positive developer experience to get started!



1,363 commits

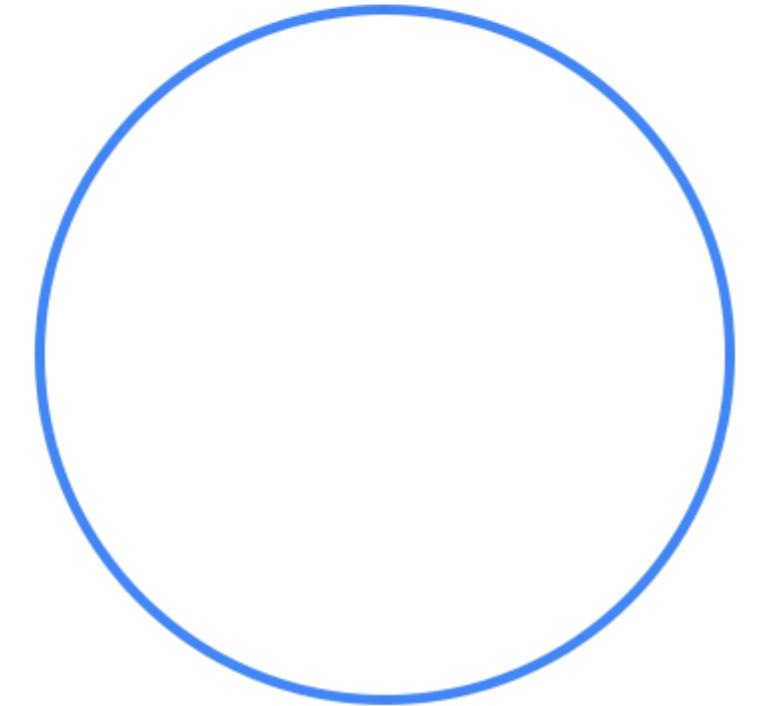
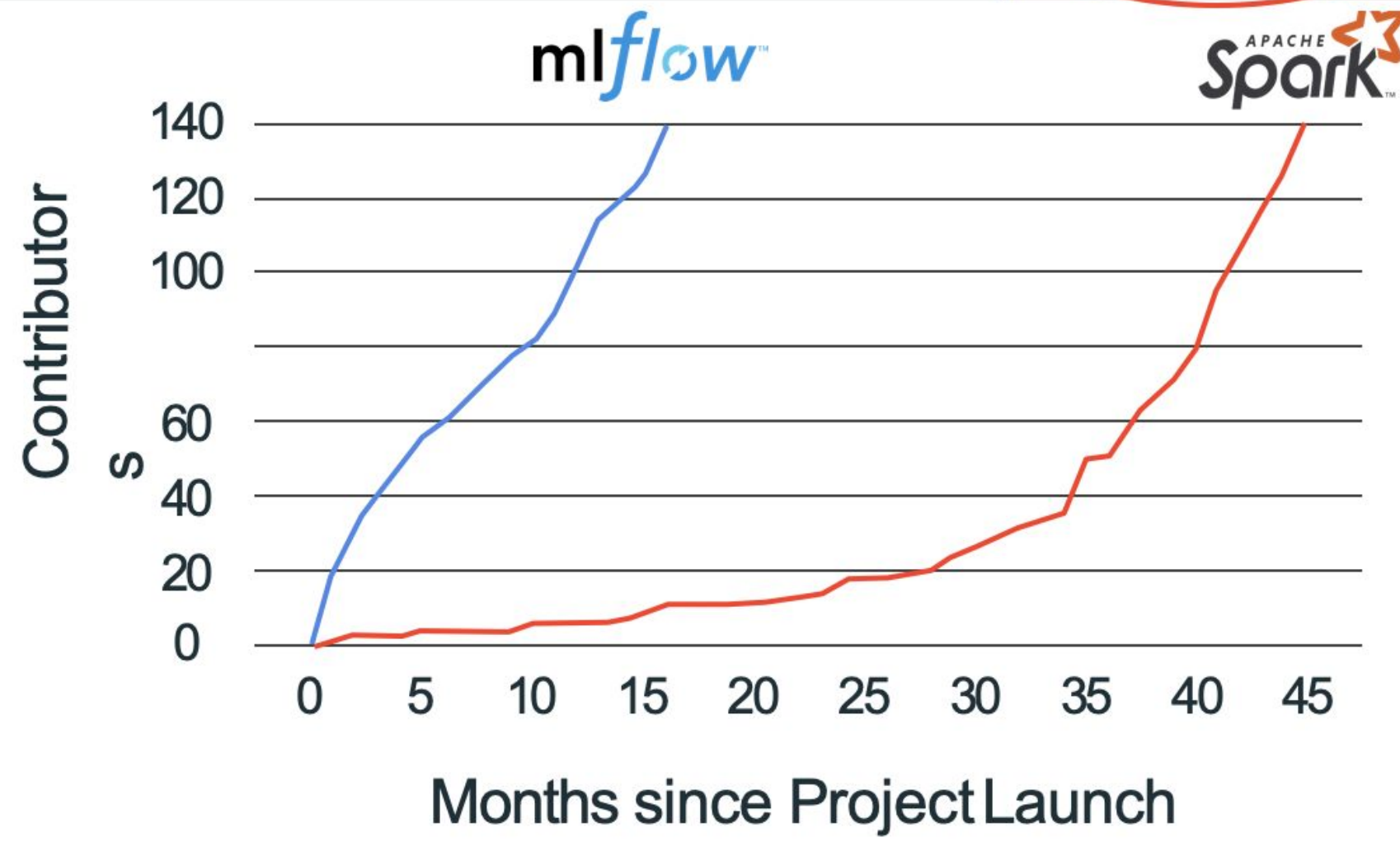
37 branches

0 packages

28 releases

200 contributors

Apache-2.0



MLflow components

mlflow

Tracking

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

databricks.com/mlflow

mlflow

Projects

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



mlflow.org

mlflow

Models

Deploy machine learning models in diverse serving environments.



github.com/mlflow

mlflow

Model Registry

Store, annotate, and manage models in a central repository.

twitter.com/MLflow

MLflow components

mlflow

Tracking

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

databricks.com/mlflow

mlflow

Projects

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



mlflow.org

mlflow

Models

Deploy machine learning models in diverse serving environments.



github.com/mlflow

mlflow

Model Registry

Store, annotate, and manage models in a central repository.

twitter.com/MLflow

MLflow components

mlflow

Tracking

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

databricks.com/mlflow

mlflow

Projects

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



mlflow.org

mlflow

Models

Deploy machine learning models in diverse serving environments.



github.com/mlflow

mlflow

Model Registry

Store, annotate, and manage models in a central repository.

twitter.com/MLflow

MLflow components

mlflow

Tracking

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

databricks.com/mlflow

mlflow

Projects

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



mlflow.org

mlflow

Models

Deploy machine learning models in diverse serving environments.



github.com/mlflow

mlflow

Model Registry

Store, annotate, and manage models in a central repository.

twitter.com/MLflow

MLflow components

mlflow

Tracking

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

databricks.com/mlflow

mlflow

Projects

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



mlflow.org

mlflow

Models

Deploy machine learning models in diverse serving environments.



github.com/mlflow

mlflow

Model Registry

Store, annotate, and manage models in a central repository.

twitter.com/MLflow

MLflow Tracking

The MLflow logo, consisting of the word "mlflow" in a blue, lowercase, sans-serif font. The "l" and "f" are connected, and the "o" has a small circular arrow around it.

Tracking

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

The MLflow logo, consisting of the word "mlflow" in a blue, lowercase, sans-serif font. The "l" and "f" are connected, and the "o" has a small circular arrow around it.

Projects

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

The MLflow logo, consisting of the word "mlflow" in a blue, lowercase, sans-serif font. The "l" and "f" are connected, and the "o" has a small circular arrow around it.

Models

Deploy machine learning models in diverse serving environments.

The MLflow logo, consisting of the word "mlflow" in a blue, lowercase, sans-serif font. The "l" and "f" are connected, and the "o" has a small circular arrow around it.

Model Registry

Store, annotate, and manage models in a central repository.

MLflow Tracking



Tracking

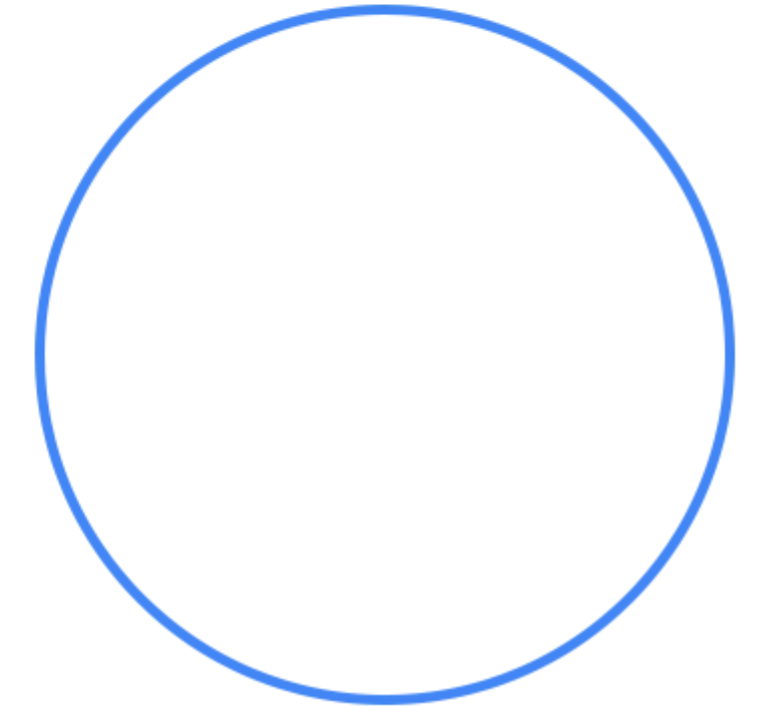
Record and query experiments: code, data, config, and results.

```
import mlflow

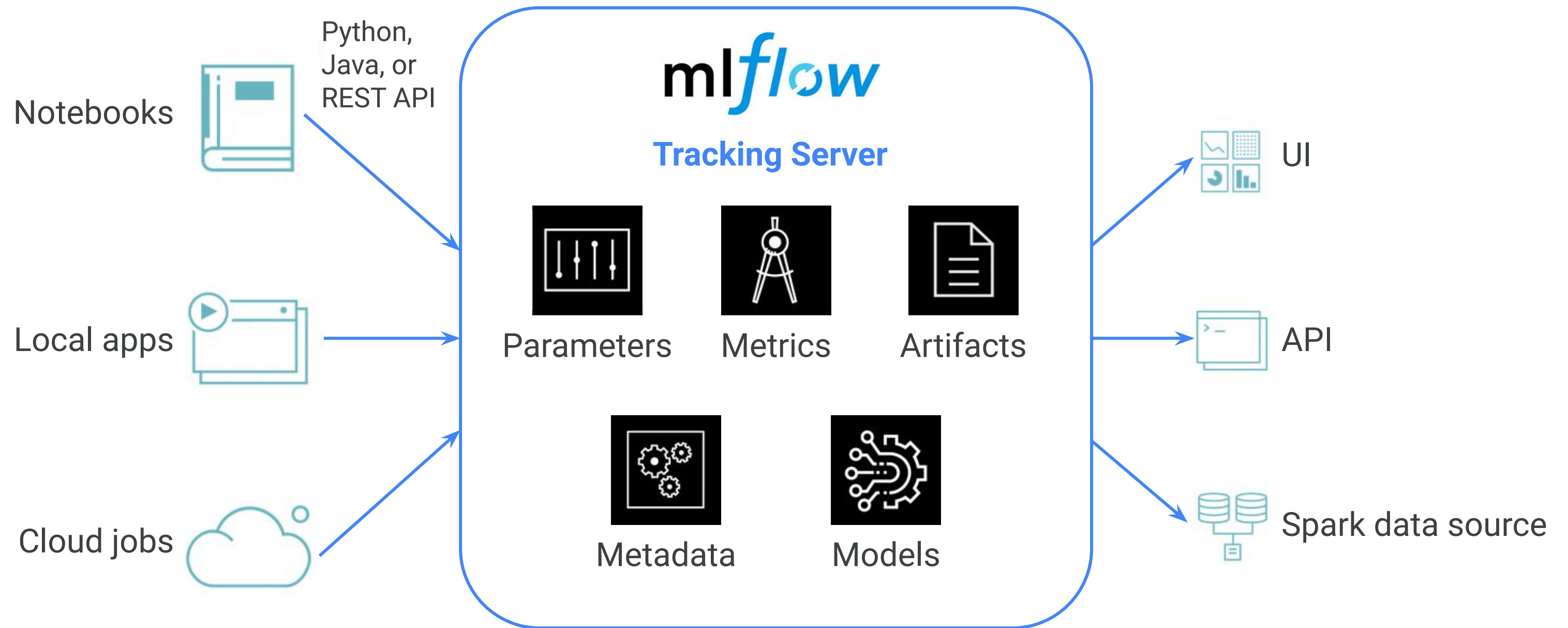
with mlflow.start_run():
    mlflow.log_param("layers", layers)
    mlflow.log_param("alpha", alpha)

    # train model

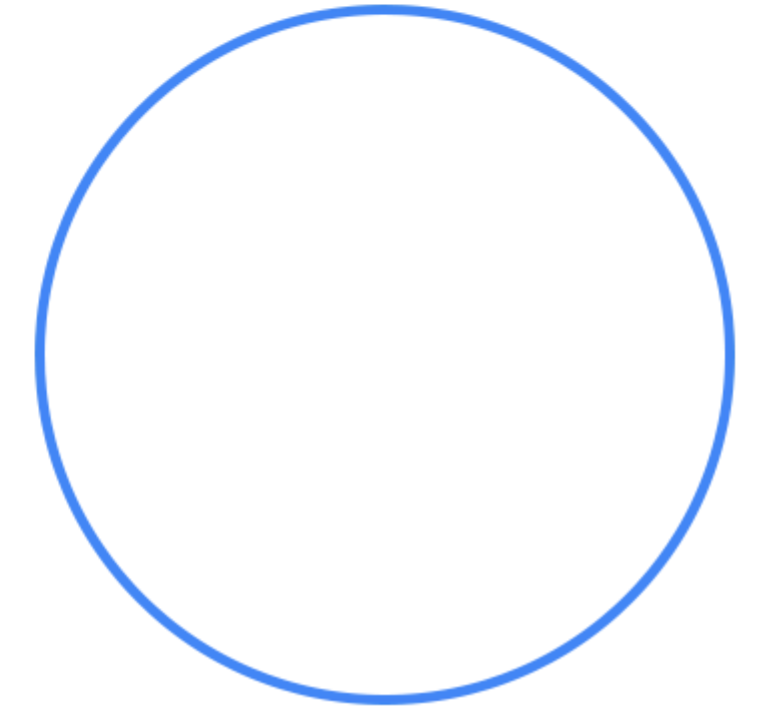
    mlflow.log_metric("mse", model.mse())
    mlflow.log_artifact("plot", model.plot(test_df))
    mlflow.tensorflow.log_model(model)
```



MLflow Tracking



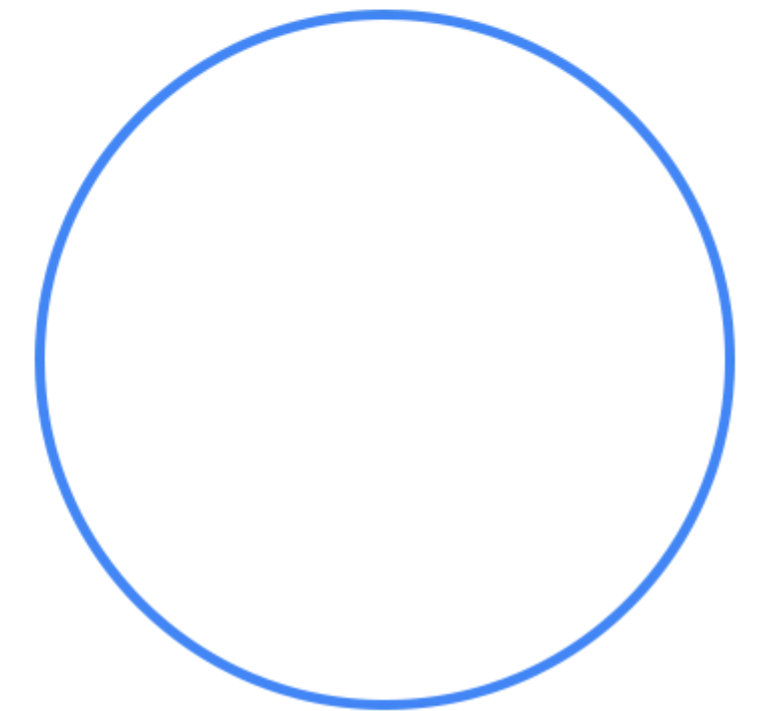
MLflow Tracking backend store



MLflow Tracking backend store

Entity (metadata) store

- FileStore (local file system)
mlruns directory by default
- SQLStore (via SQLAlchemy)
PostgreSQL, MySQL, SQLite
- MLflow plugins scheme
Customized entity metastore
- Managed MLflow on Databricks
MySQL on AWS and Azure



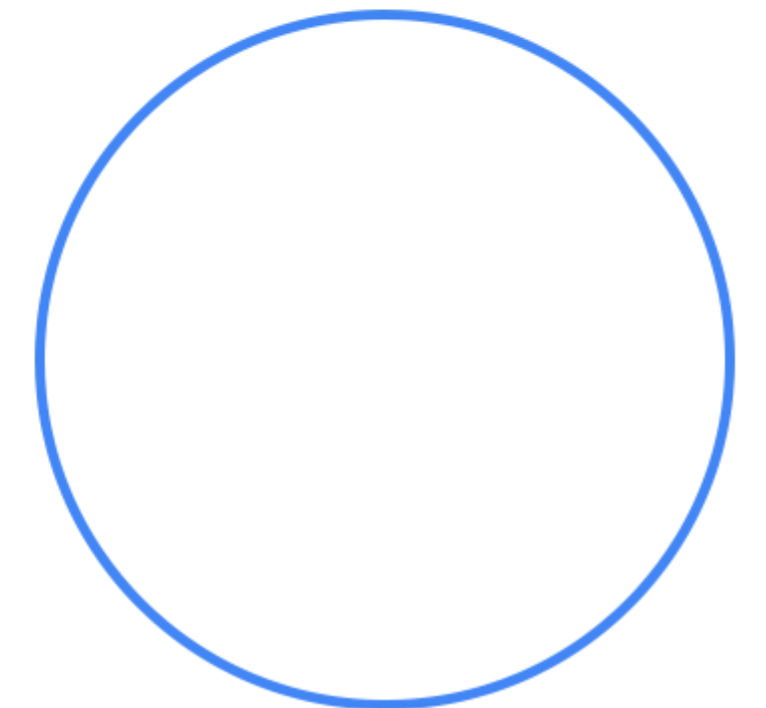
MLflow Tracking backend store

Entity (metadata) store

- FileStore (local file system)
 mlruns directory by default
- SQLStore (via SQLAlchemy)
 PostgreSQL, MySQL, SQLite
- MLflow plugins scheme
 Customized entity metastore
- Managed MLflow on Databricks
 MySQL on AWS and Azure

Artifact store

- Local file system
 mlruns directory
- S3-backed store
- Azure Blob storage
- Google Cloud Storage
- DBFS artifact repo



MLflow Projects

mlflow

Tracking

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

mlflow

Projects

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

mlflow

Models

Deploy machine learning models in diverse serving environments.

mlflow

Model Registry

Store, annotate, and manage models in a central repository.

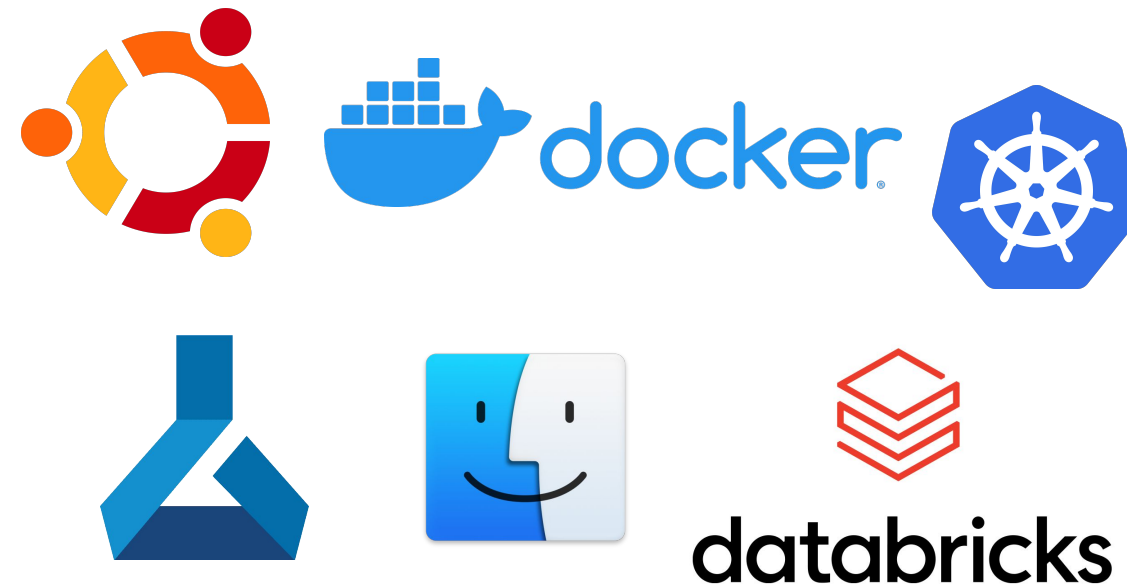
MLflow Projects motivation

mlflow

Projects

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

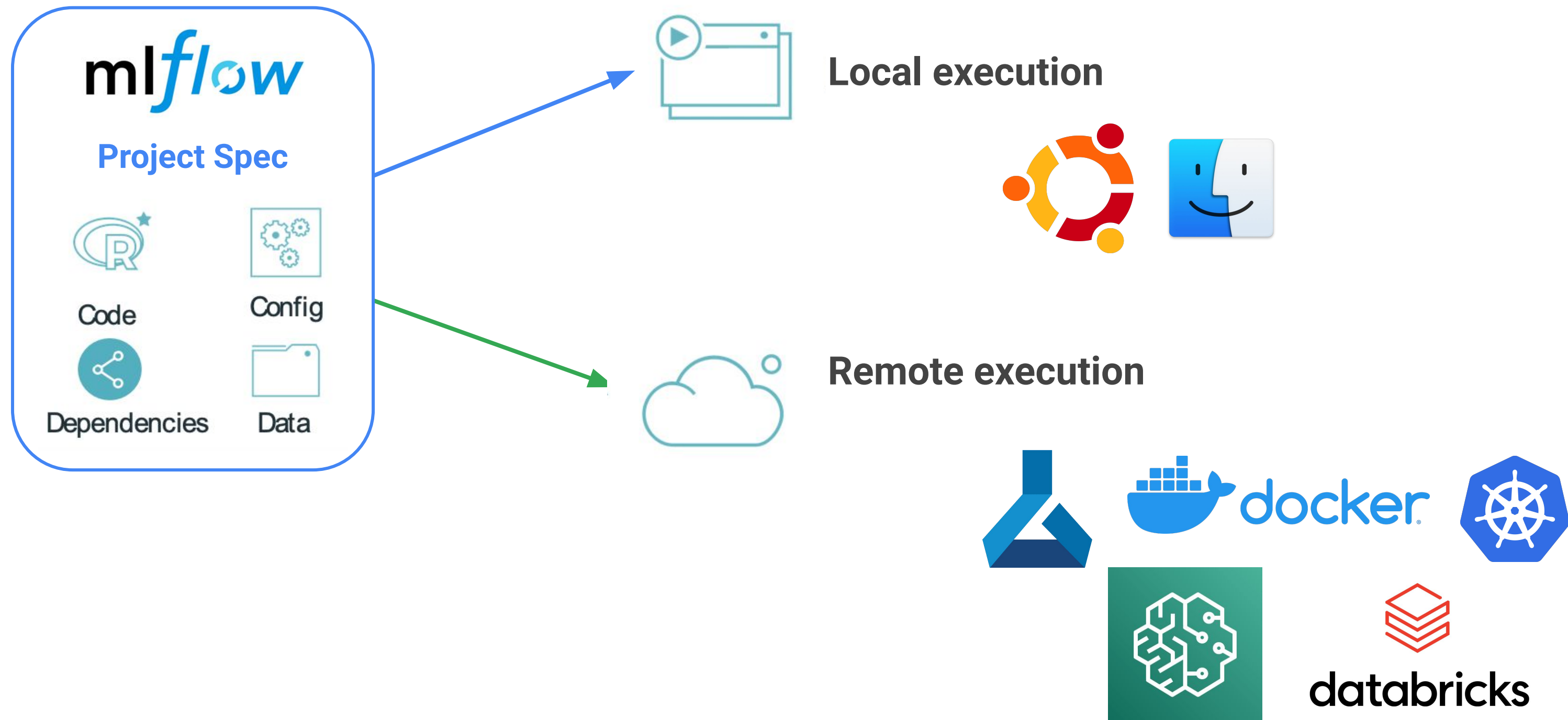
Diverse set of environments



Diverse set of tools



MLflow Projects



MLflow Projects

Packaging format for reproducible ML runs

Any code folder or GitHub repository

MLproject file with project configuration

Defines dependencies for reproducibility

Conda (+ R, Docker, ...) dependencies can be specified in MLproject

Reproducible in (almost) any environment

Execution API for running projects

CLI /Python/R/Java mml

- Supports local and remote execution



Example MLflow project file

my_project/

MLproject

conda.yaml

main.py

model.py

...

```
conda_env: conda.yaml

entry_points:
  main:
    parameters:
      training_data: path
      lambda: {type: float, default: 0.1}
    command: python main.py {training_data} {lambda}
```

```
$ mlflow run git://<my_project>.git -P lambda=0.3
mlflow.run("git://<my_project>", parameters={...})
mlflow run . -e main -P lambda=0.3
```

Example conda.yaml

my_project/

- MLproject
- conda.yaml
- main.py
- model.py

...

```
channels:
```

```
-defaults
```

```
dependencies:
```

```
- python=3.7.3
```

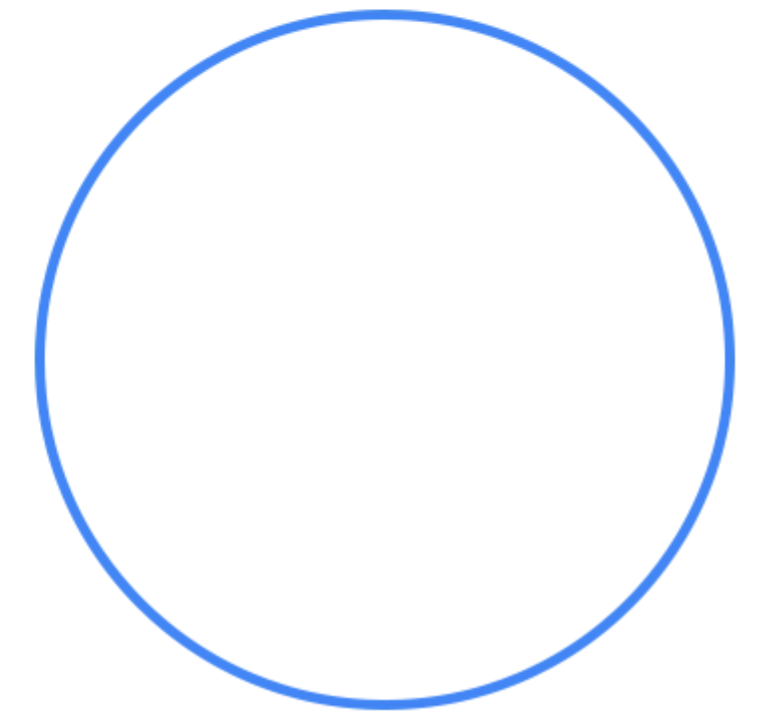
```
- scikit-learn=0.20.3
```

```
- pip:
```

```
- mlflow
```

```
- cloudpickle==0.8.0
```

```
name: mlflow-env
```



MLflow Models

mlflow

Tracking

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

mlflow

Projects

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

mlflow

Models

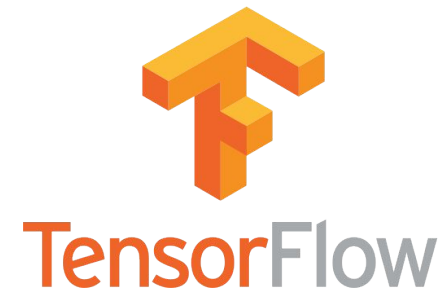
Deploy machine learning models in diverse serving environments.

mlflow

Model Registry

Store, annotate, and manage models in a central repository.

MLflow Models motivations

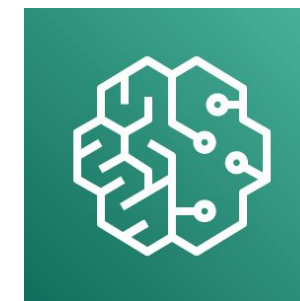


Inference code

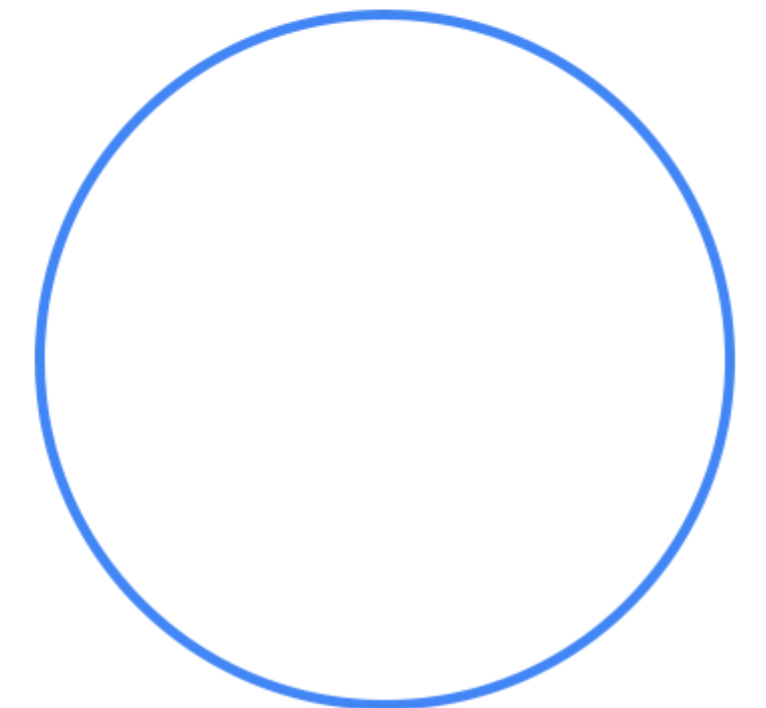


Batch and stream scoring

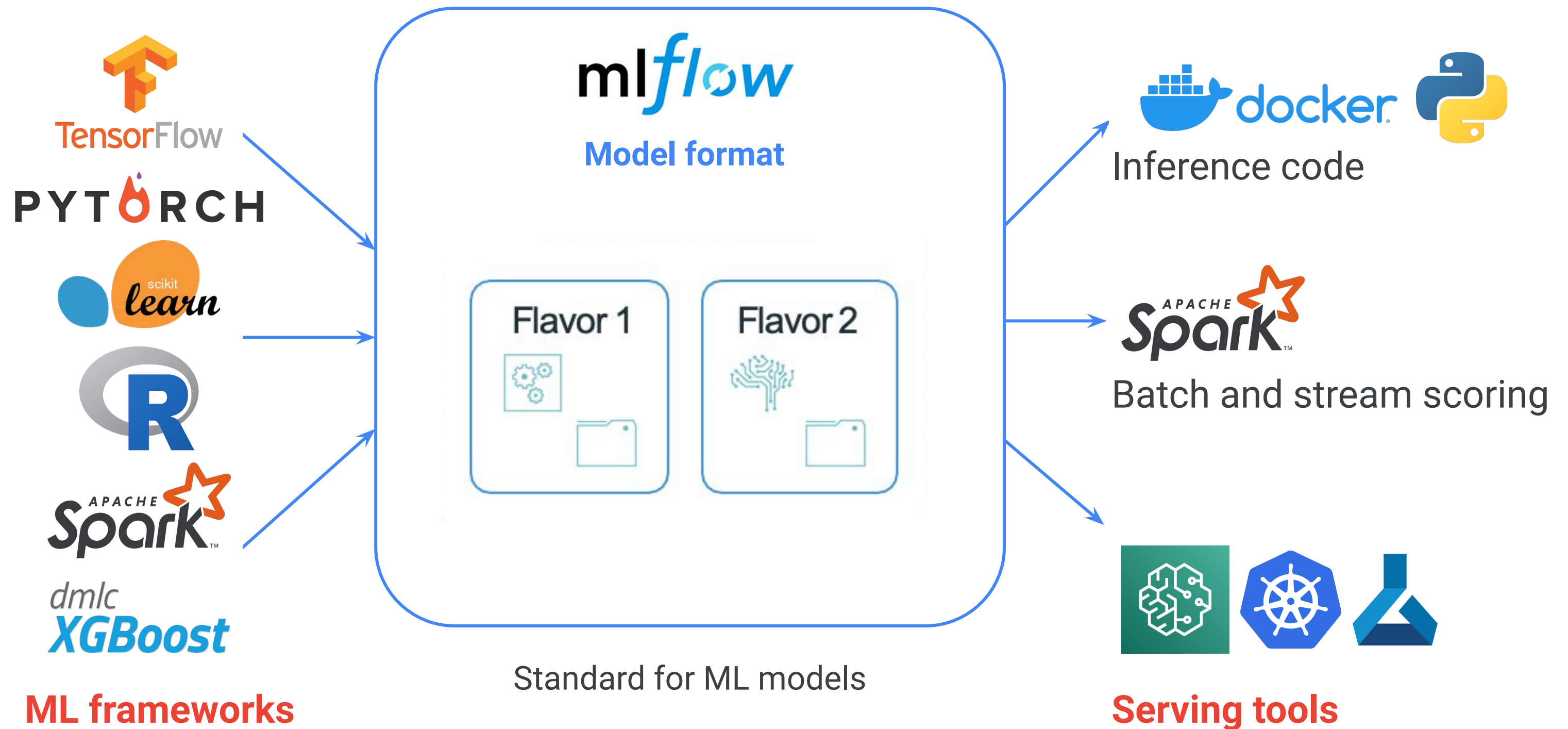
ML frameworks



Serving tools



MLflow Models



Example MLflow model

my_model/

MLmodel

estimator/

saved_model.pb

variables/

...

mlflow.tensorflow.log_model(...)

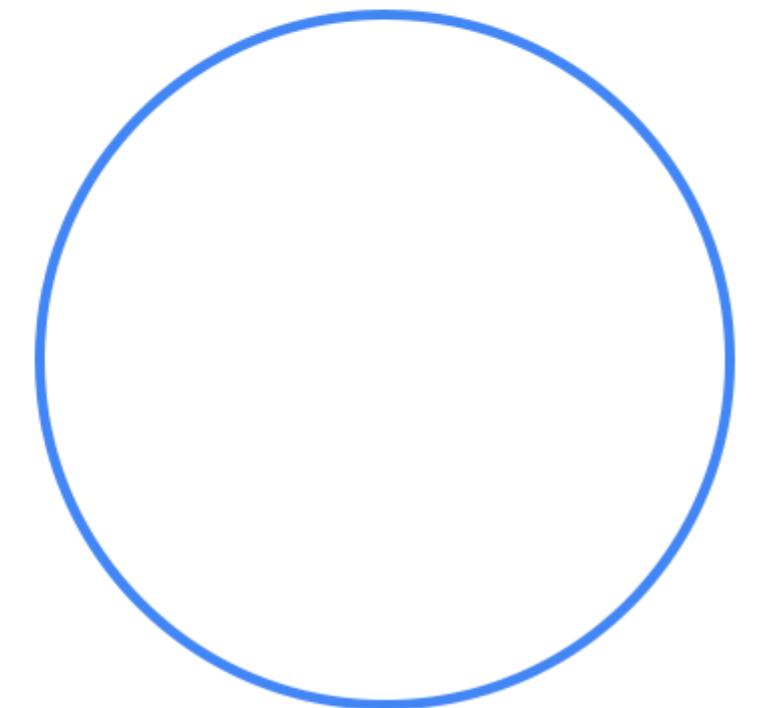
```
run_id: 769915006efd4c4bbd662461
time_created: 2018-06-28T12:34
flavors:
```

```
tensorflow:
  saved_model_dir: estimator
  signature_def_key: predict
```

```
python_function:
  loader_module: mlflow.tensorflow
```

} Usable by tools that understand TensorFlow model format

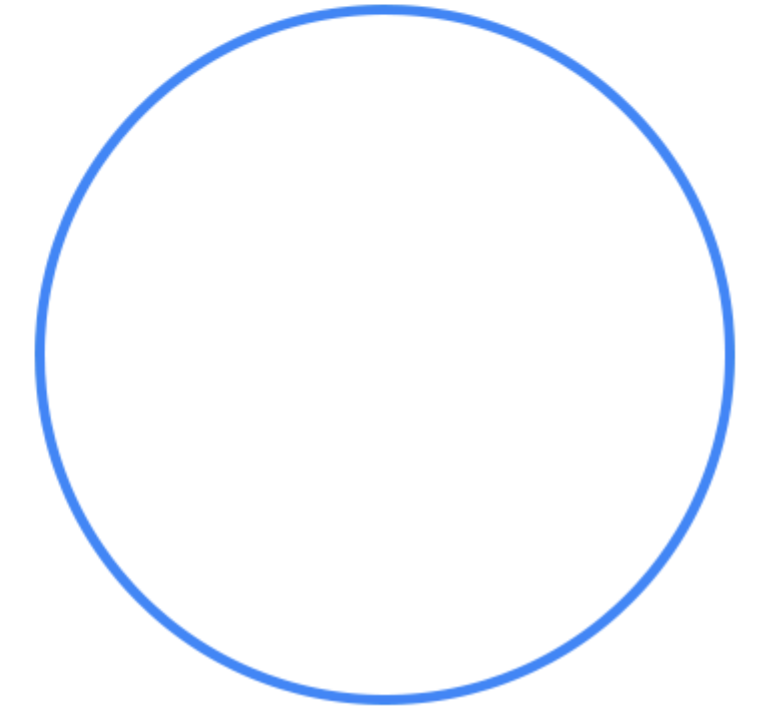
} Usable by any tool that can run Python (Docker, Spark, etc.)



Model Keras flavor example



Train a model



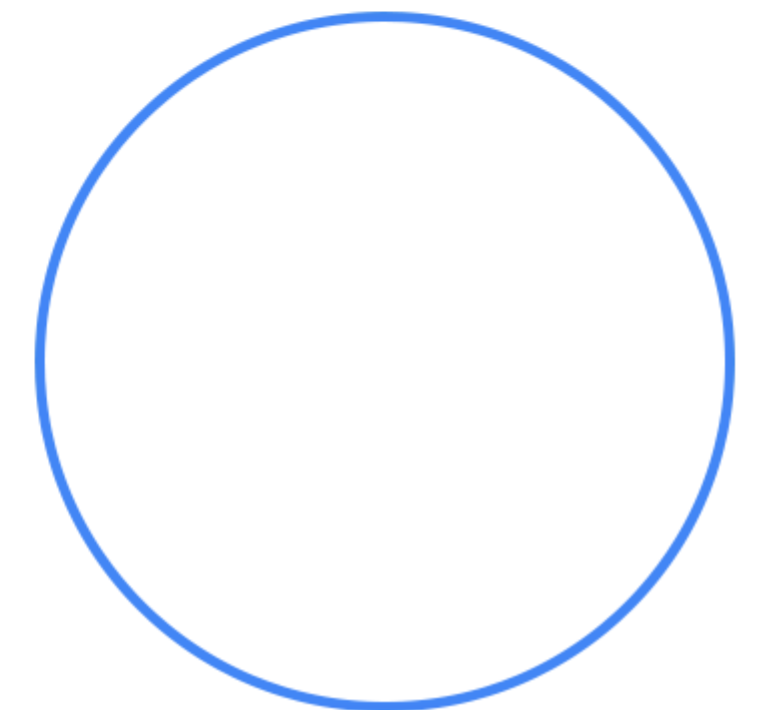
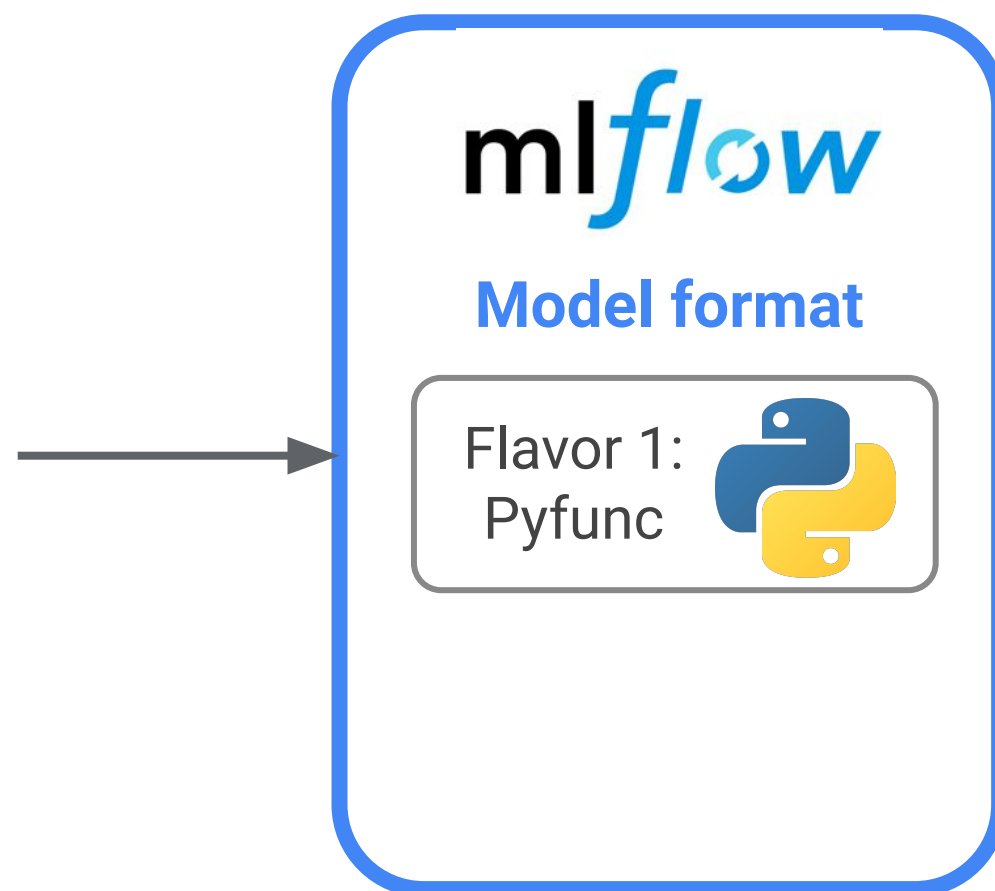
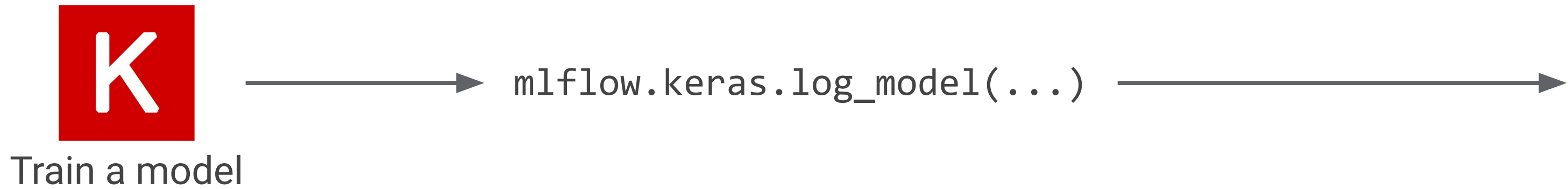
Model Keras flavor example



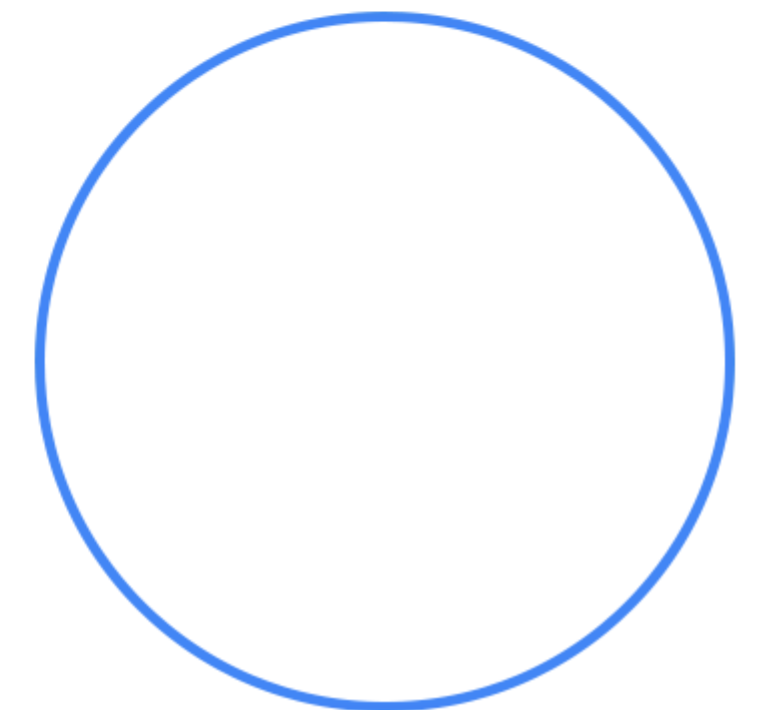
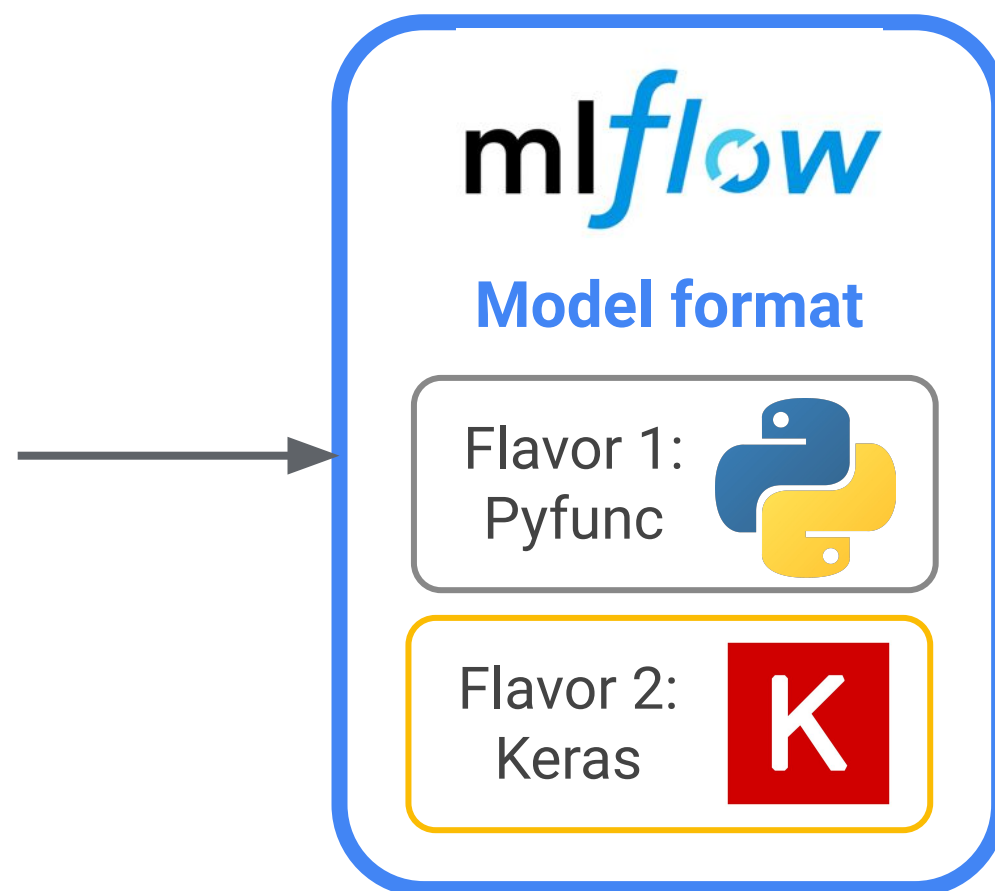
Model Keras flavor example



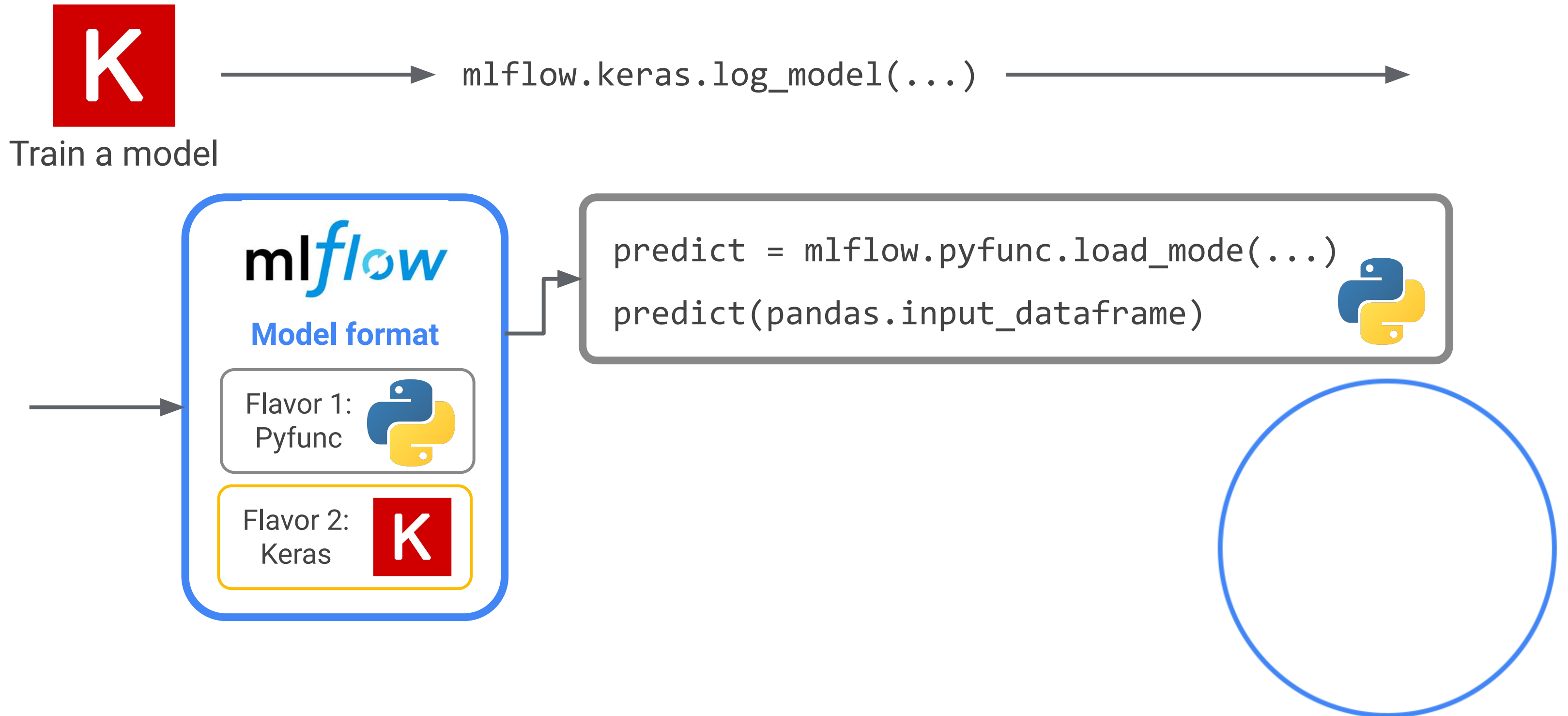
Model Keras flavor example



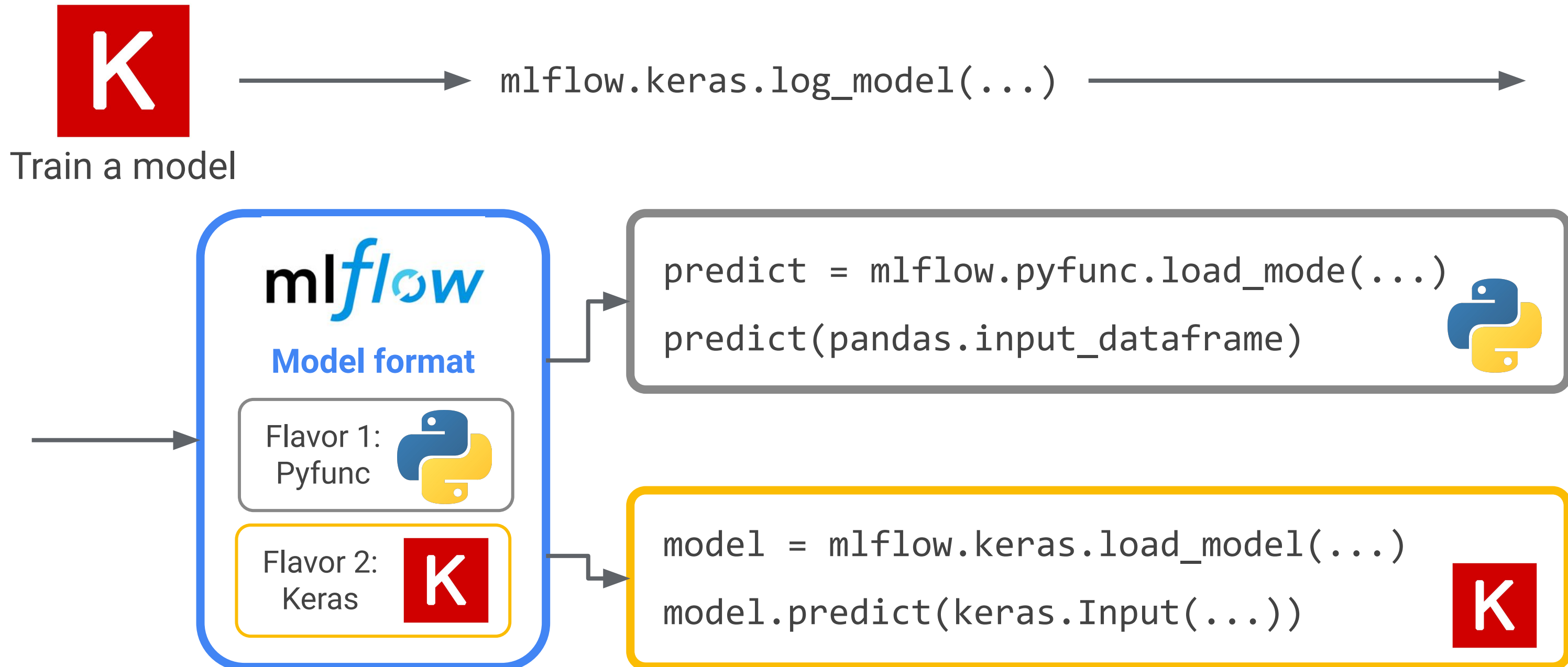
Model Keras flavor example



Model Keras flavor example

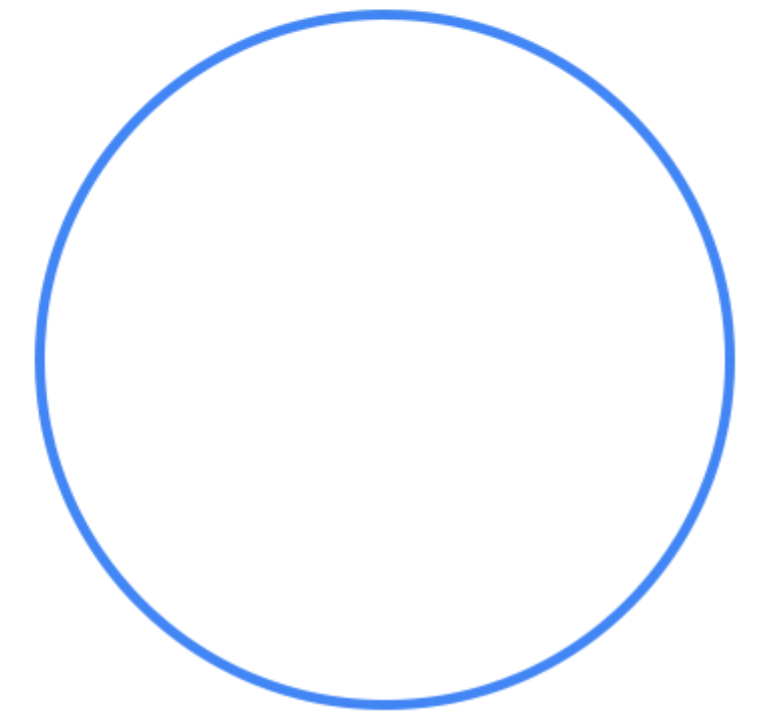
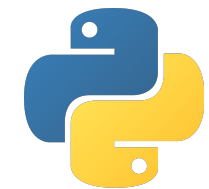


Model Keras flavor example

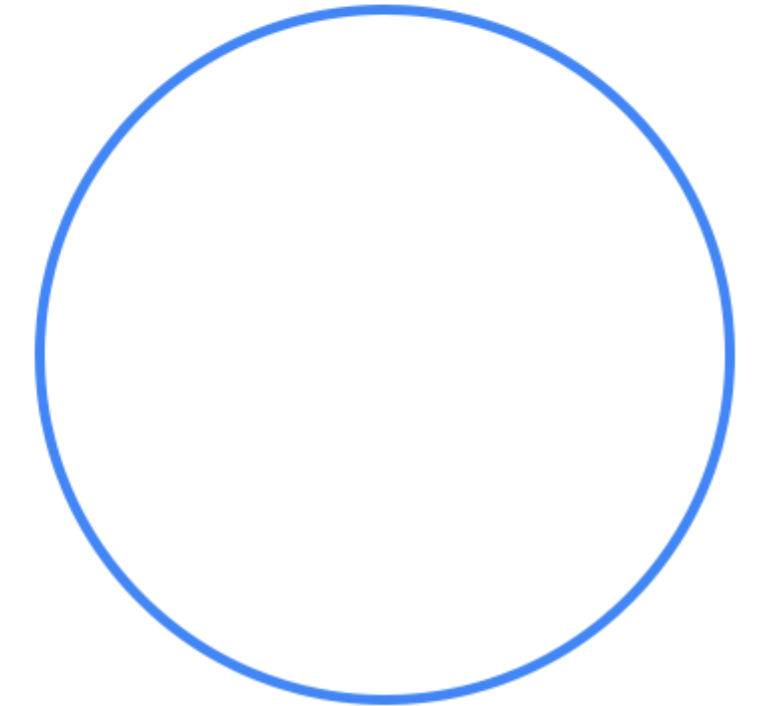
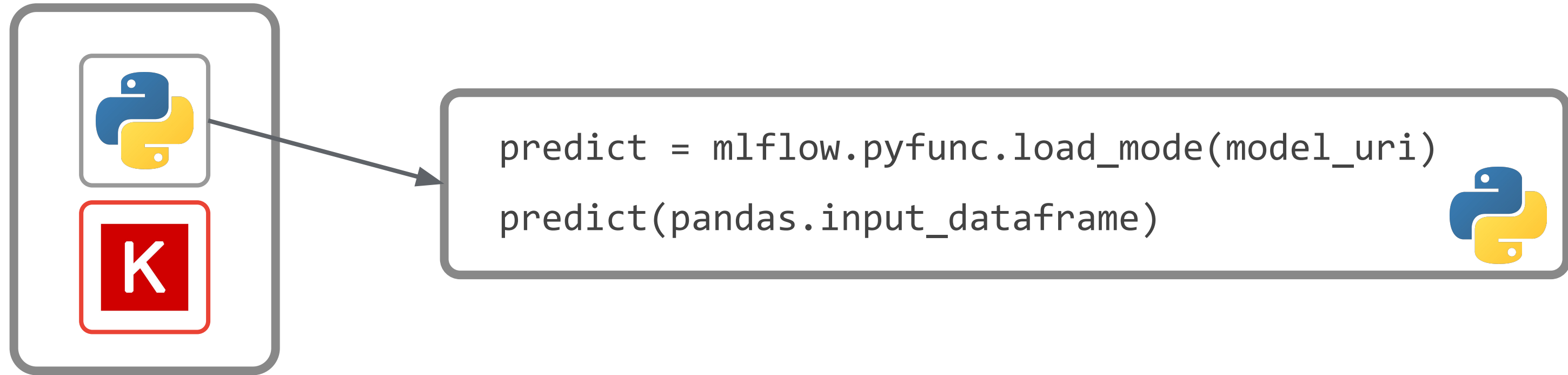


Model flavors example

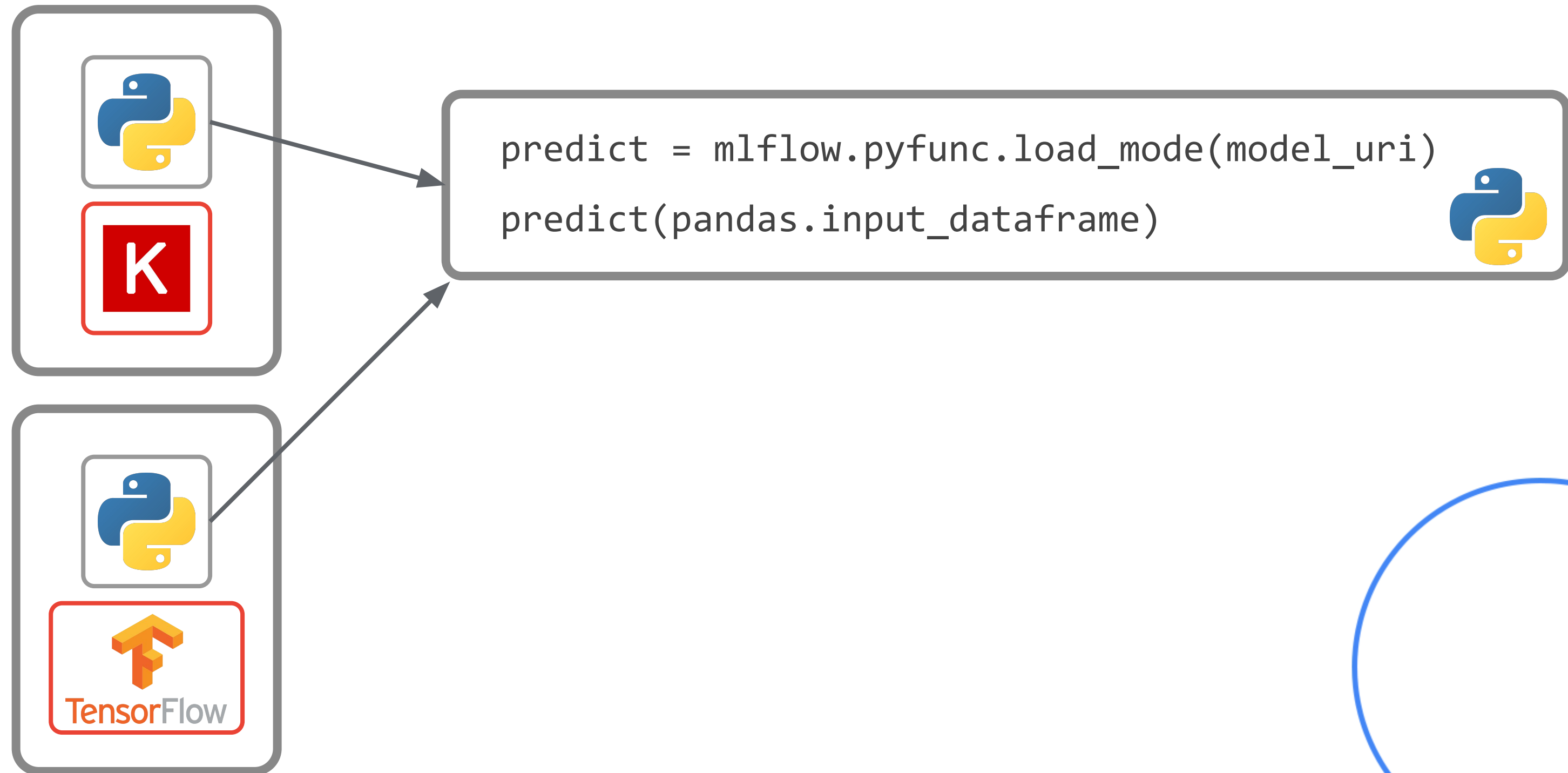
```
predict = mlflow.pyfunc.load_model(model_uri)  
predict(pandas.input_dataframe)
```



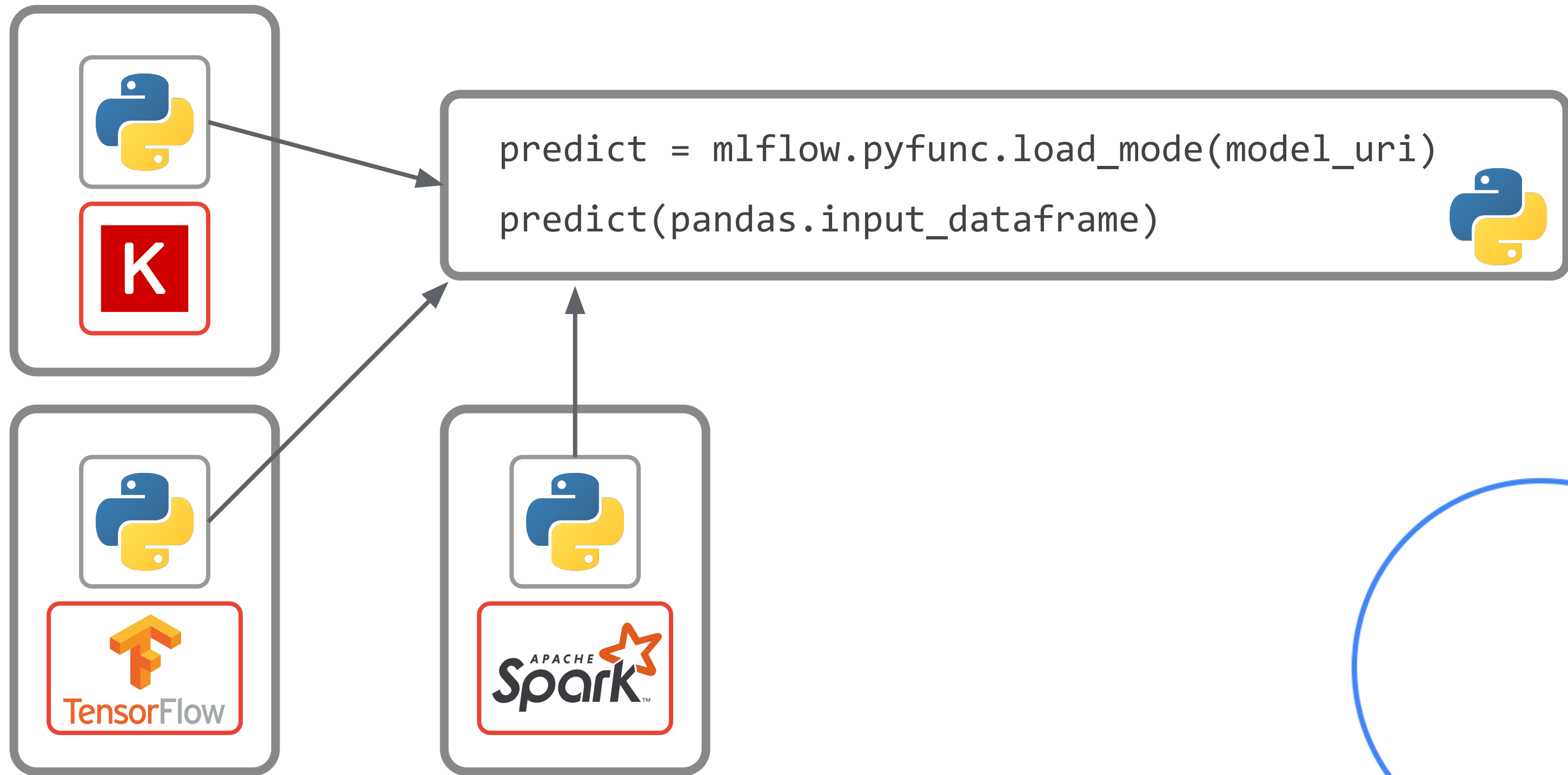
Model flavors example



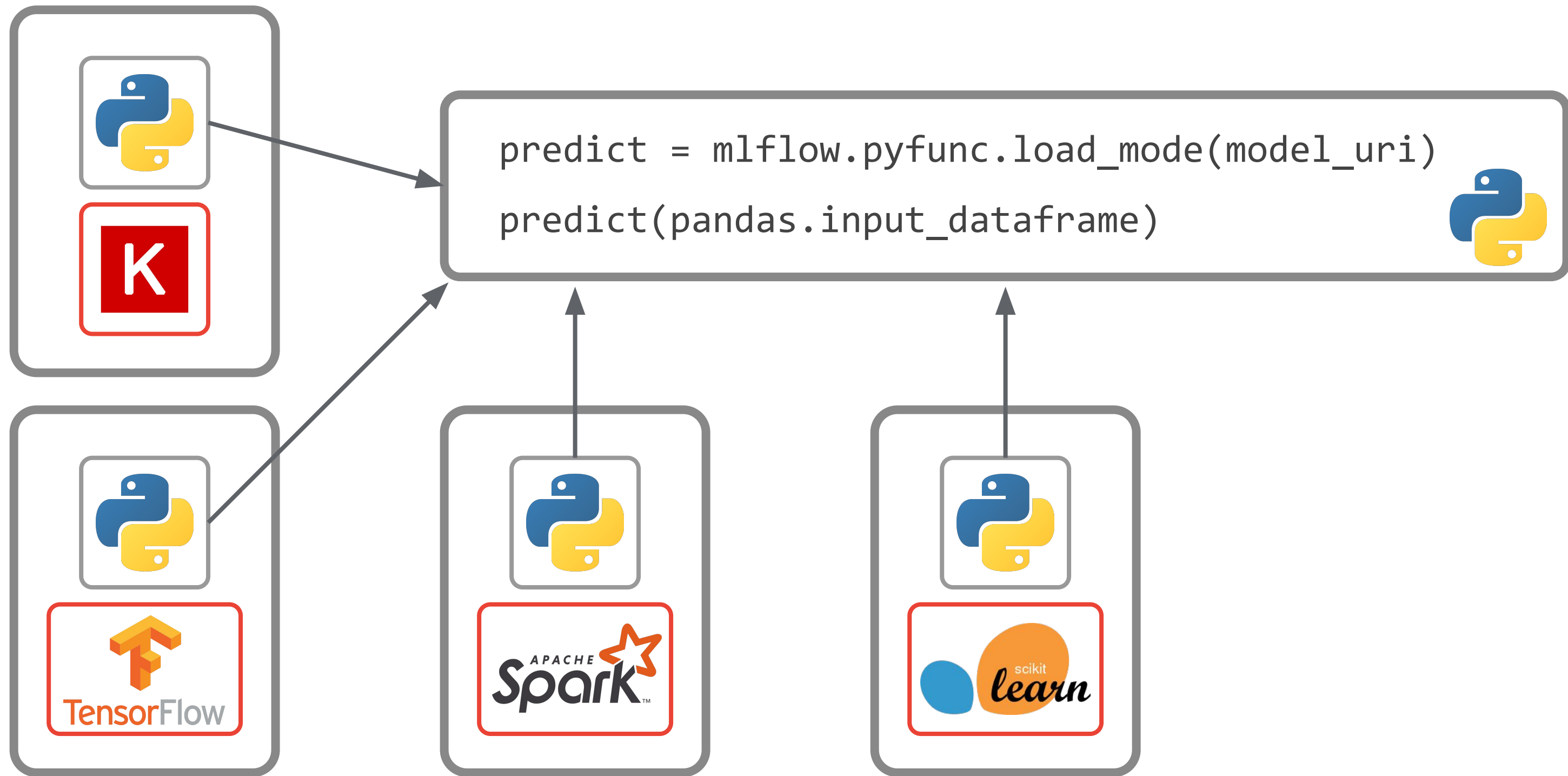
Model flavors example



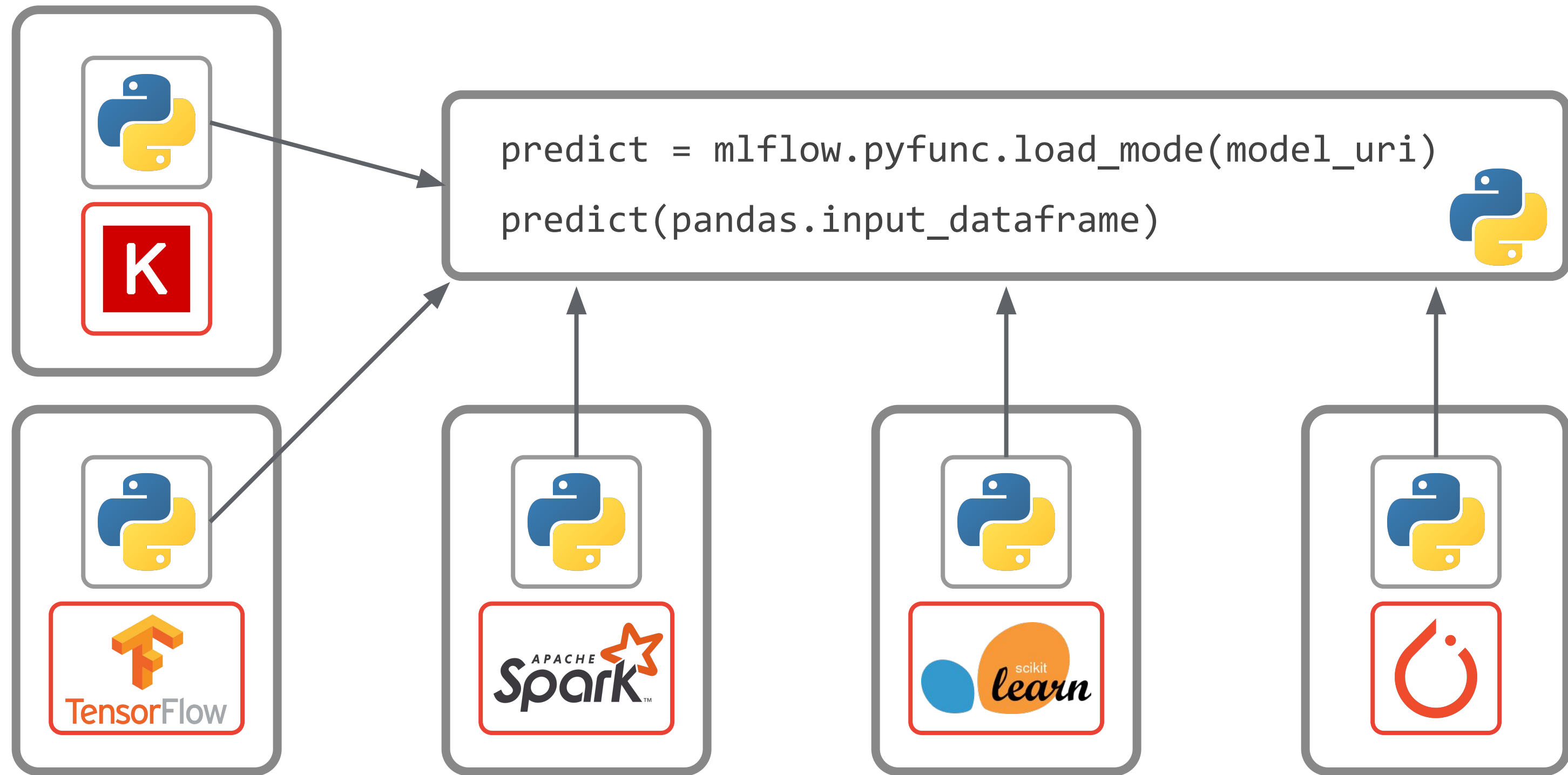
Model flavors example



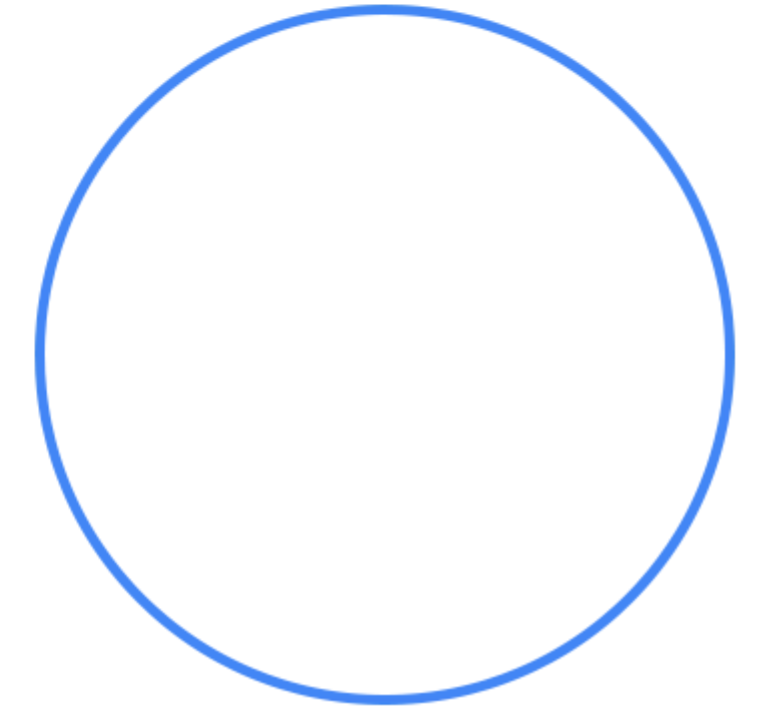
Model flavors example



Model flavors example



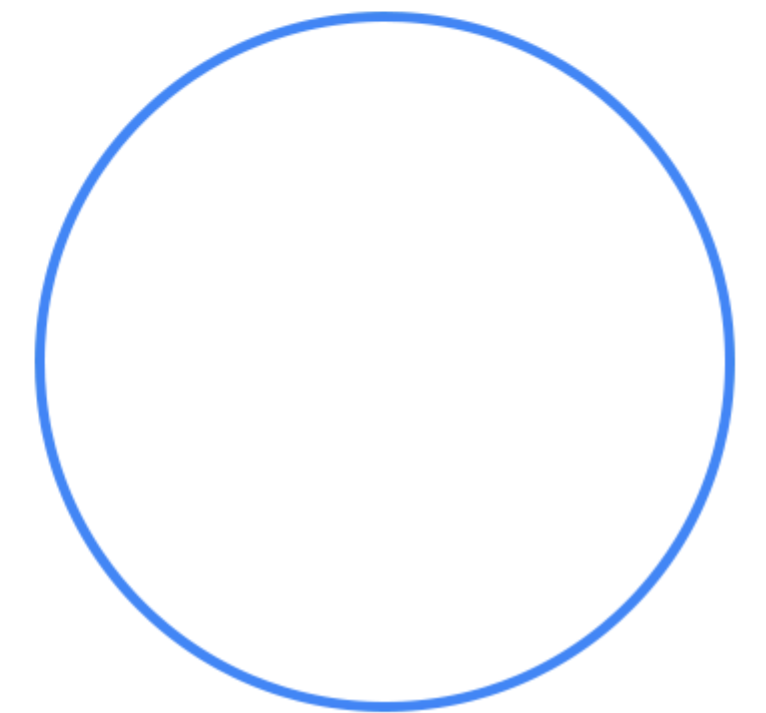
MLflow Models



MLflow Models

Packaging format for ML Models

Any directory with MLmodel file



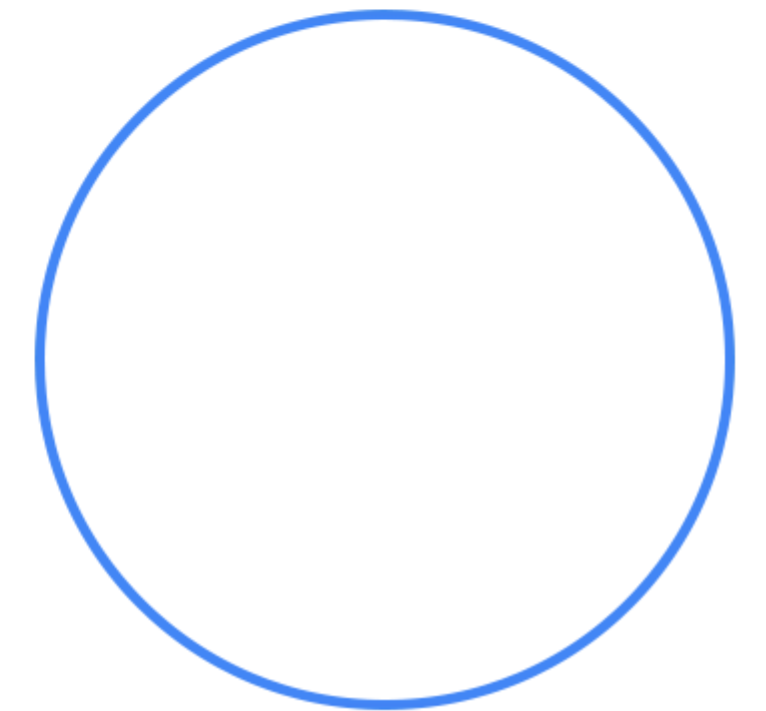
MLflow Models

Packaging format for ML Models

Any directory with MLmodel file

Defines dependencies for reproducibility

Conda environment can be specified in MLmodel configuration



MLflow Models

Packaging format for ML Models

Any directory with MLmodel file

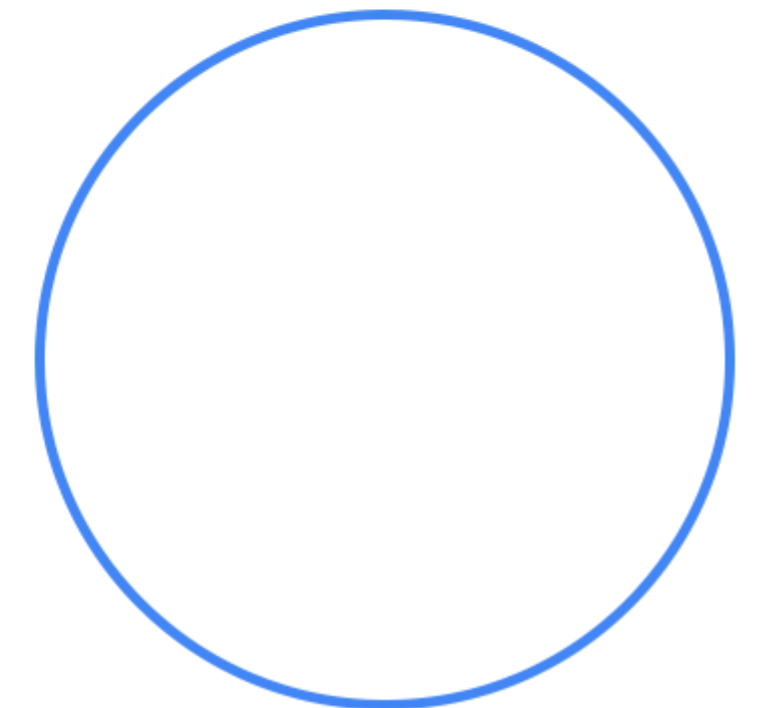
Defines dependencies for reproducibility

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(...)`



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

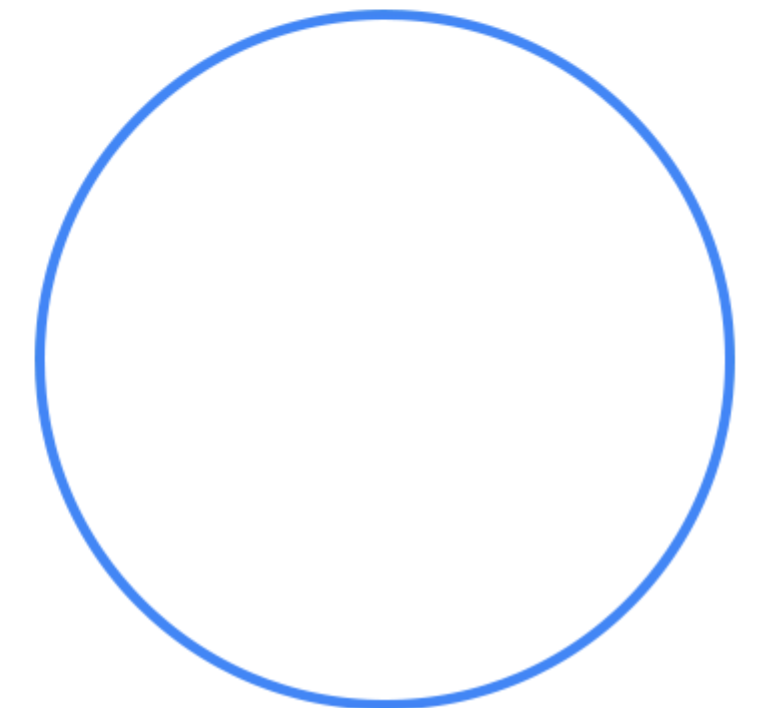
`mlflow.<model_flavor>.save_model(...)` or `log_model(...)`

`mlflow.<model_flavor>.load_model(...)`

Deployment APIs

CLI/Python/R/Java

`mlflow models [OPTIONS] COMMAND [ARGS]...`



MLflow Model Registry

mlflow

Tracking

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

mlflow

Projects

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

mlflow

Models

Deploy machine learning models in diverse serving environments

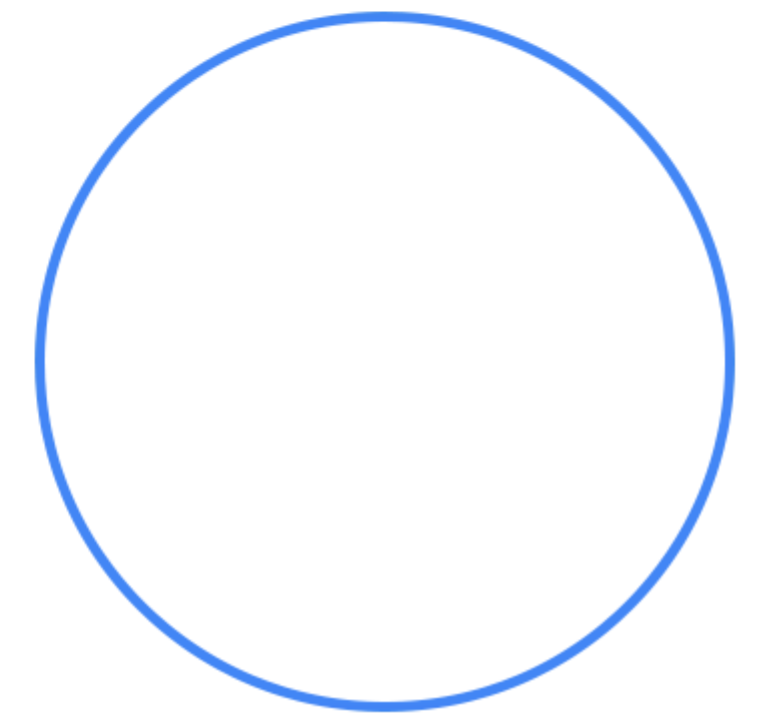
mlflow

Model Registry

Store, annotate, and manage models in a central repository

The model management problem

When you are working on one ML app alone, storing the models in files is manageable.

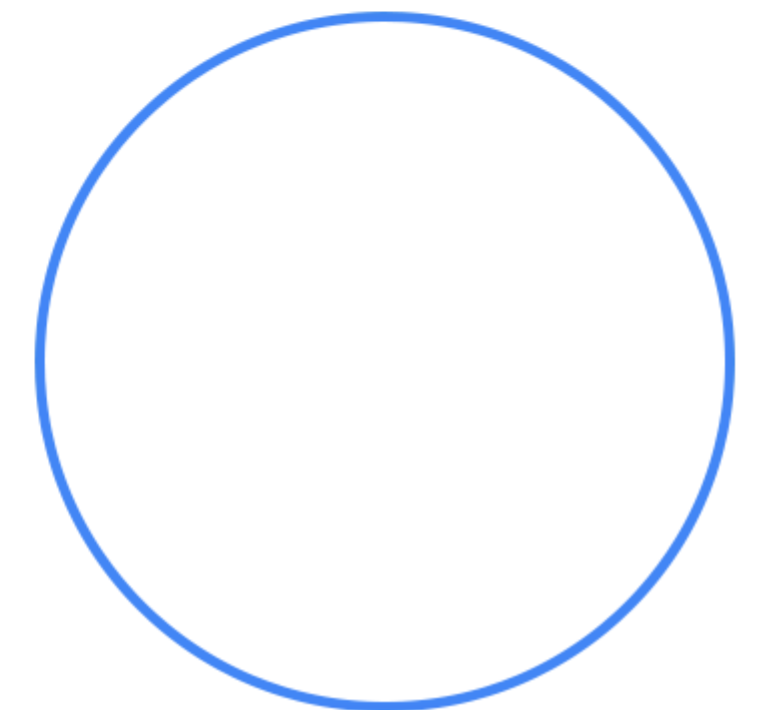
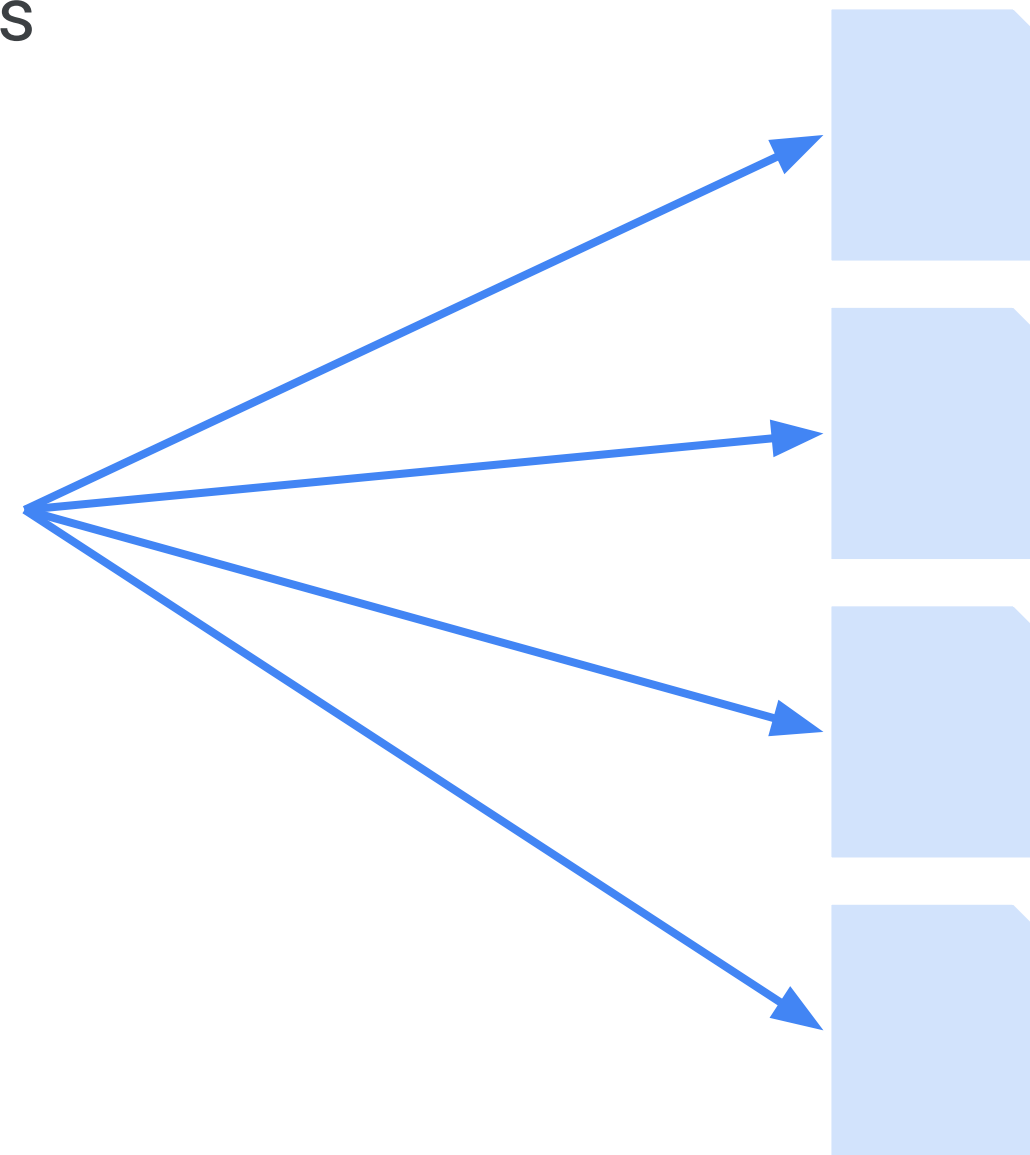


The model management problem

When you are working on one ML app alone, storing the models in files is manageable.



**MODEL
DEVELOPER**

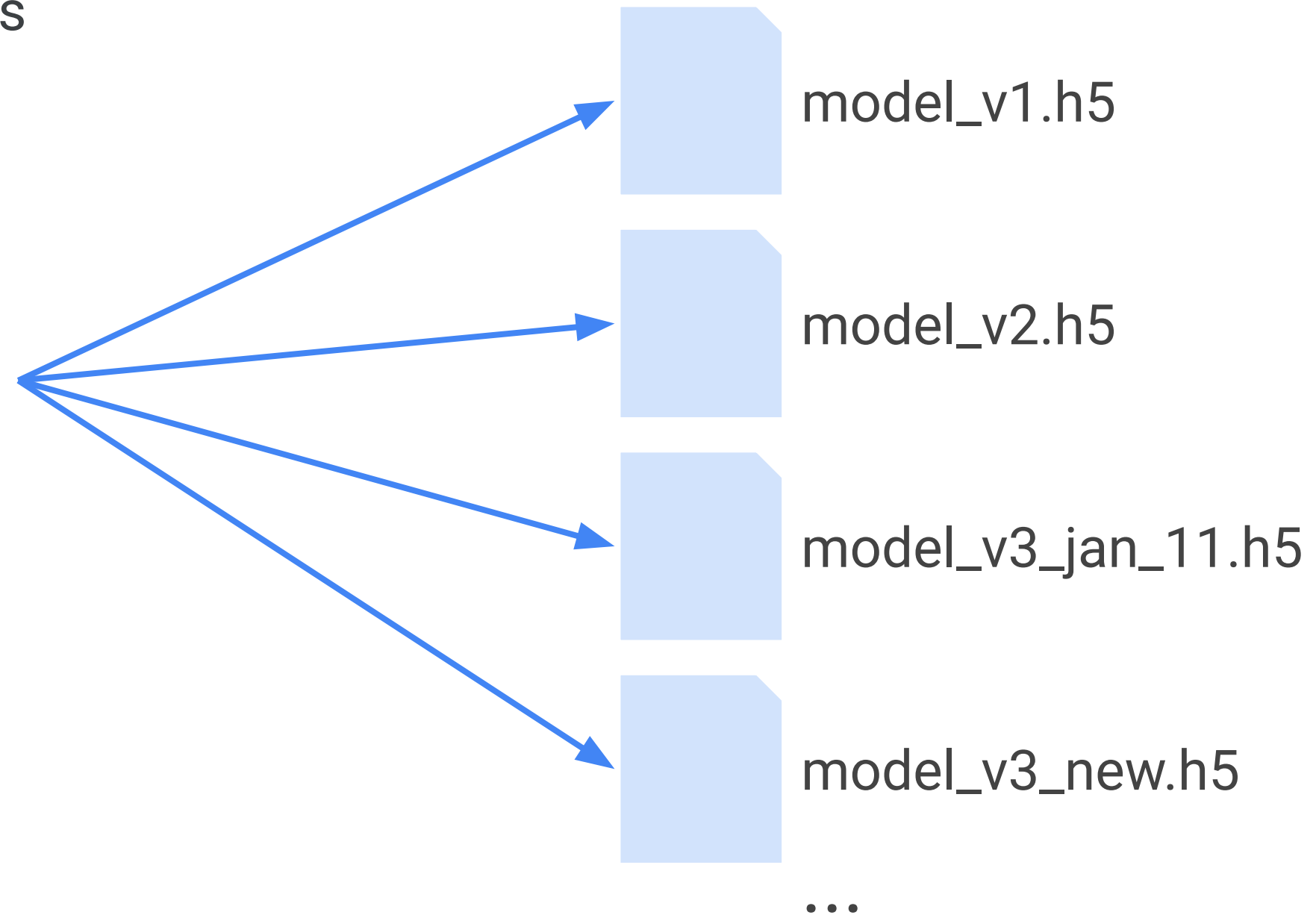


The model management problem

When you are working on one ML app alone, storing the models in files is manageable.



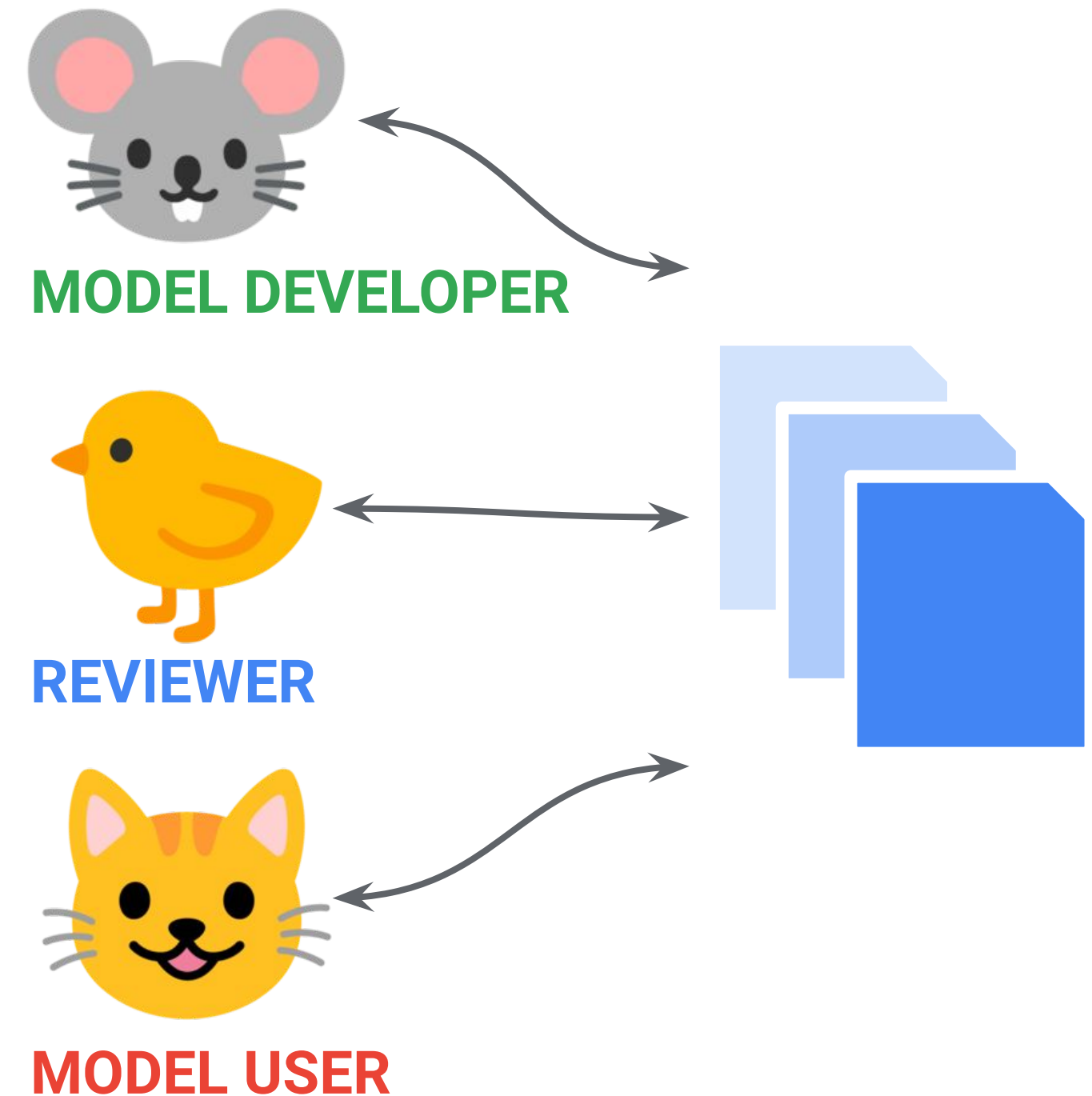
**MODEL
DEVELOPER**



The model management problem

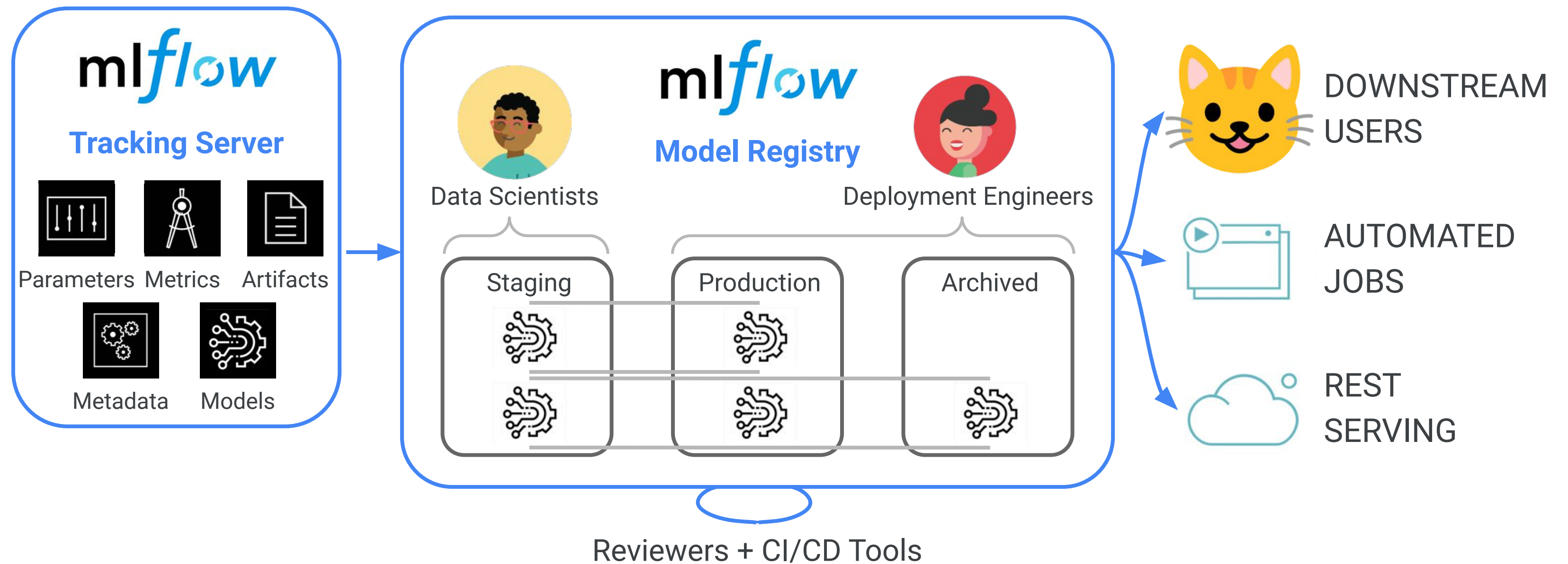
When you work in a large organization with many models and many data teams, management becomes a major challenge:

- 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?



Model Registry

VISION: Centralized and collaborative 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.

Registered Models > Airline_Delay_SparkML ▾






Created Time: 2019-10-10 15:20:29

Last Modified: 2019-10-14 12:17:04

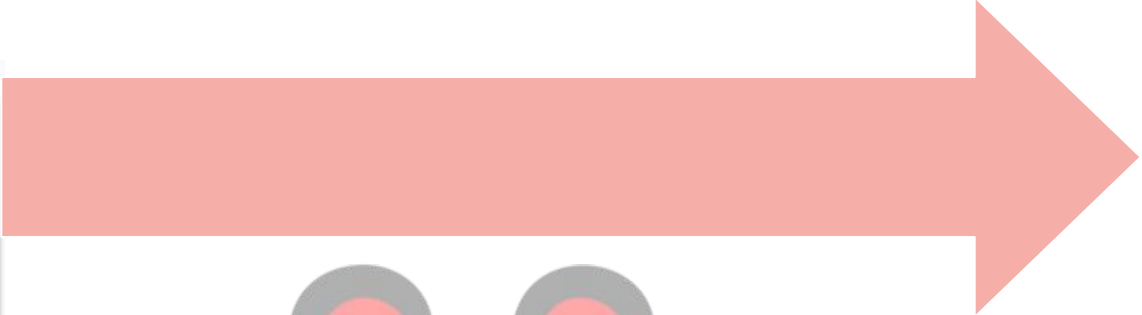
▾ Description [🔗](#)

Predicts airline delays (in minutes) using the best Spark RIF model from the AutoML Toolkit.

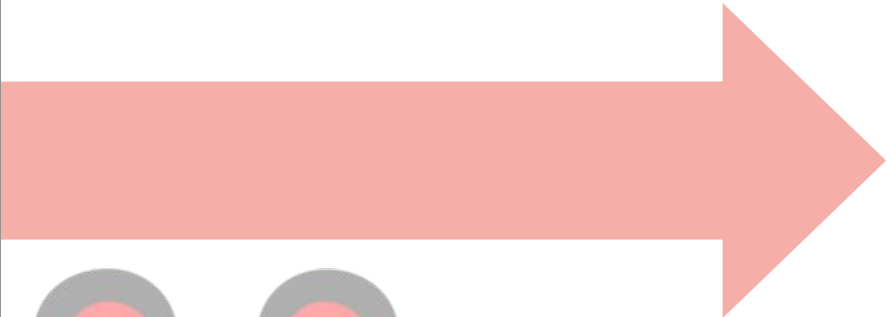
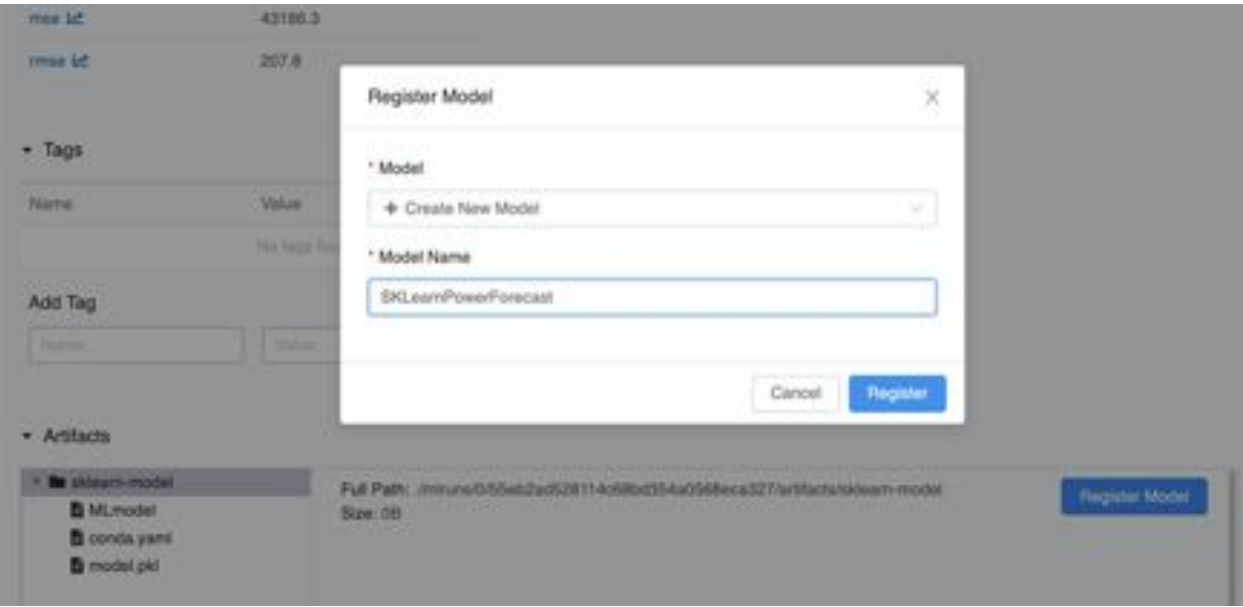
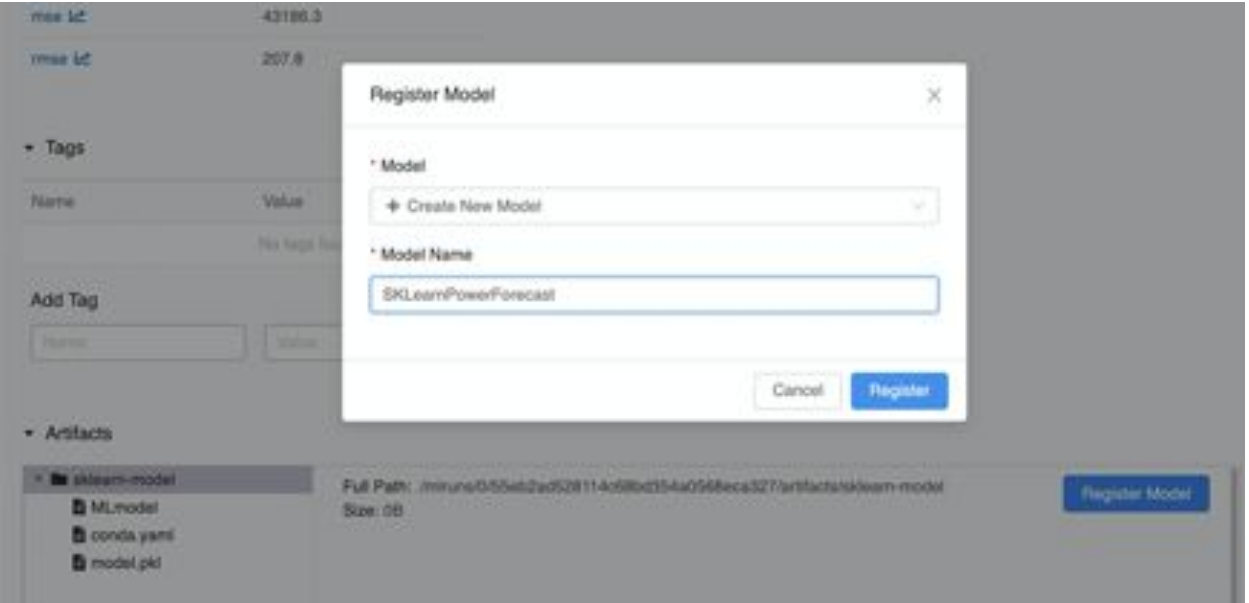
▾ Versions All Active(1)

Version	Registered at	Created by	Stage
 Version 1	2019-10-10 15:20:30	clemens@demo.com	Archived
 Version 2	2019-10-10 21:47:29	clemens@demo.com	Archived
 Version 3	2019-10-10 23:39:43	clemens@demo.com	Production
 Version 4	2019-10-11 09:55:29	clemens@demo.com	None
 Version 5	2019-10-11 12:44:44	matei@demo.com	Staging

Model Registry workflow UI



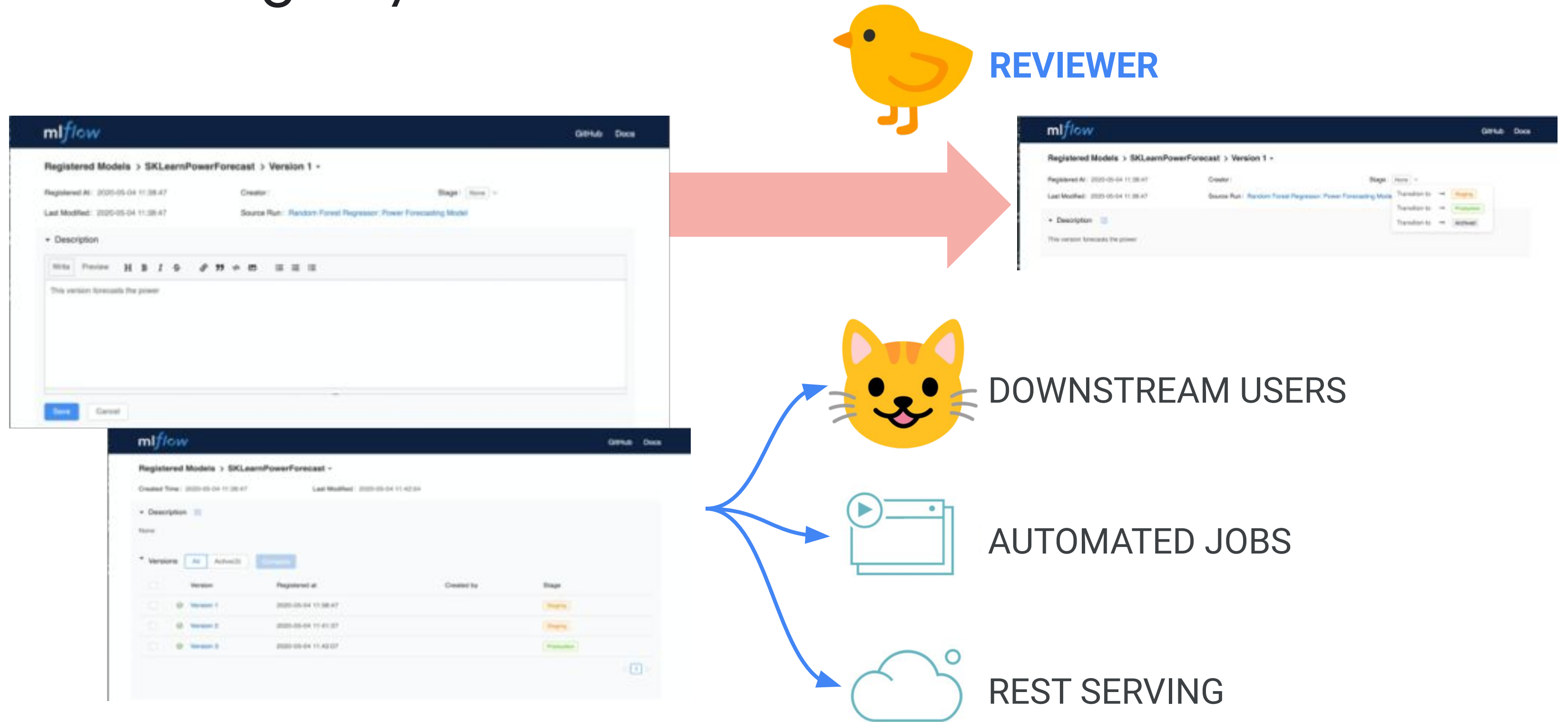
**MODEL
DEVELOPER**



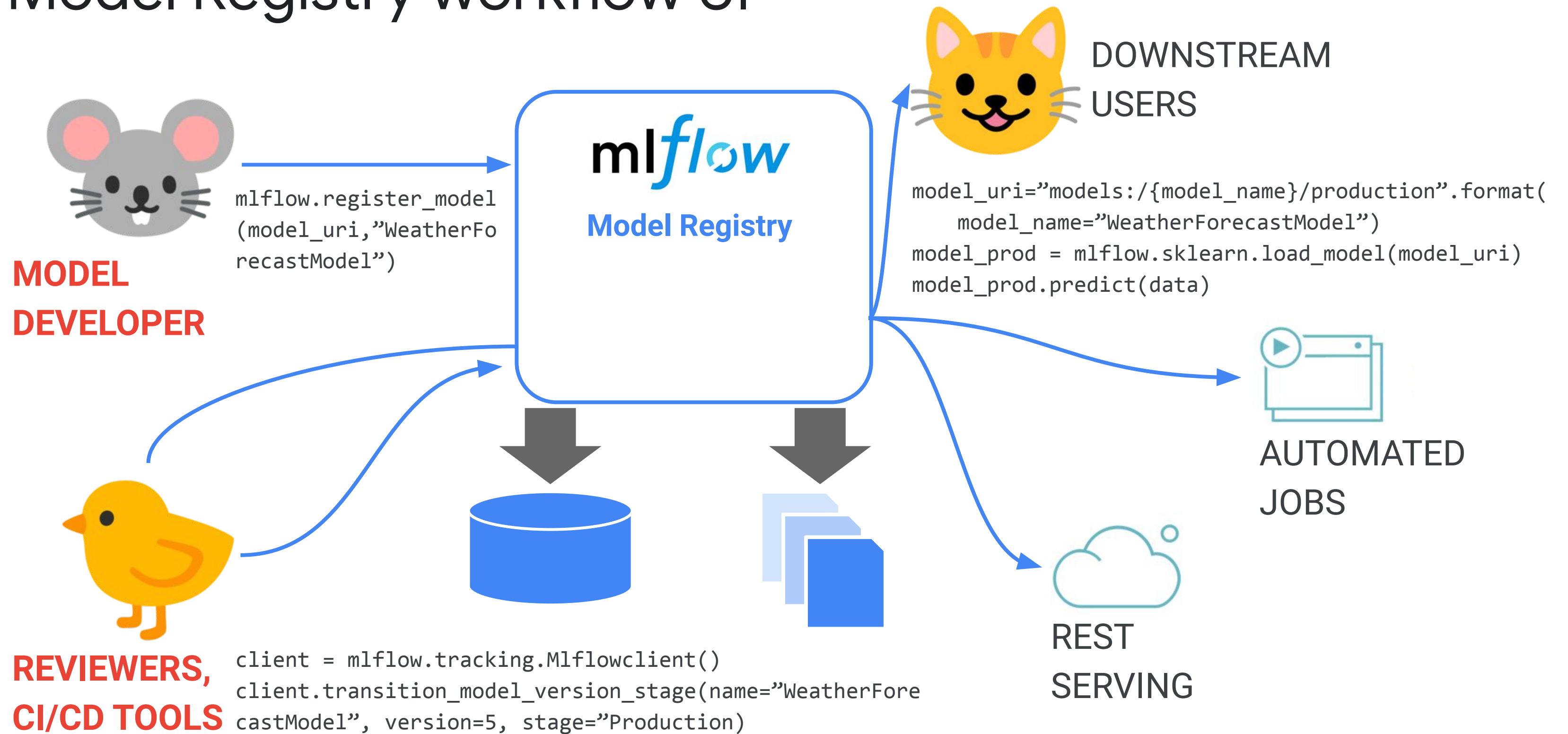
**MODEL
DEVELOPER**



Model Registry workflow UI



Model Registry workflow UI



Demo

Deploying MLflow Locally

Tracking Keras, TensorFlow, and Sckit-learn experiments



Demo

MLflow Using Databricks Community Edition



