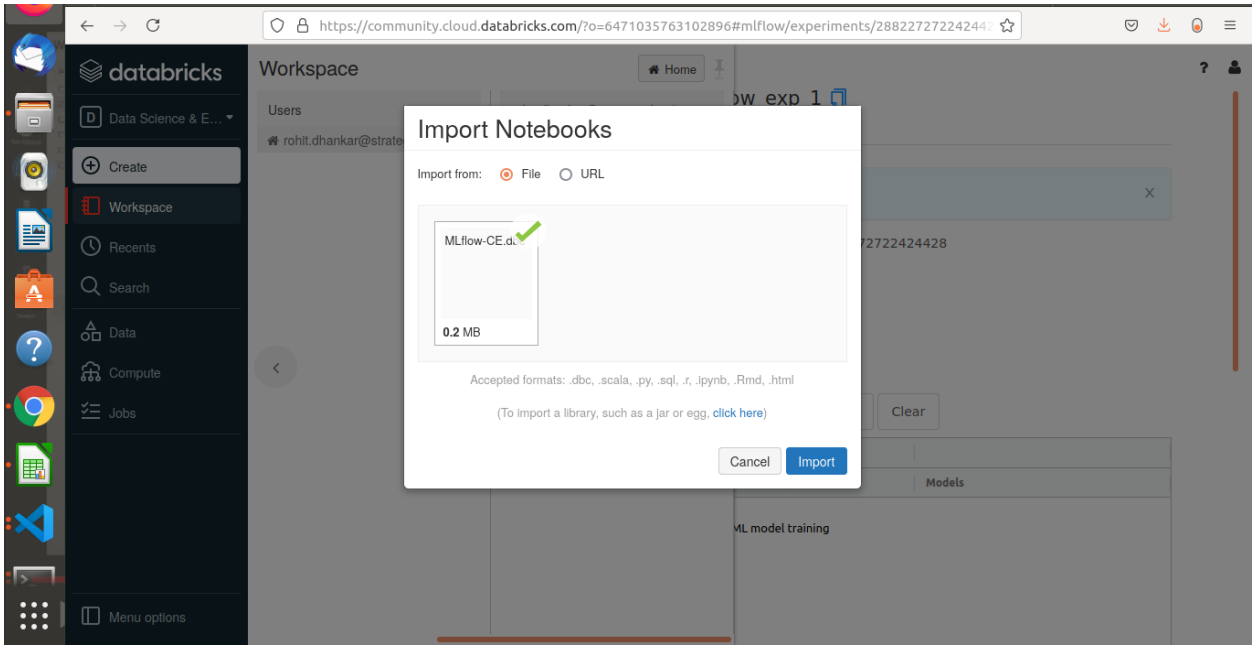
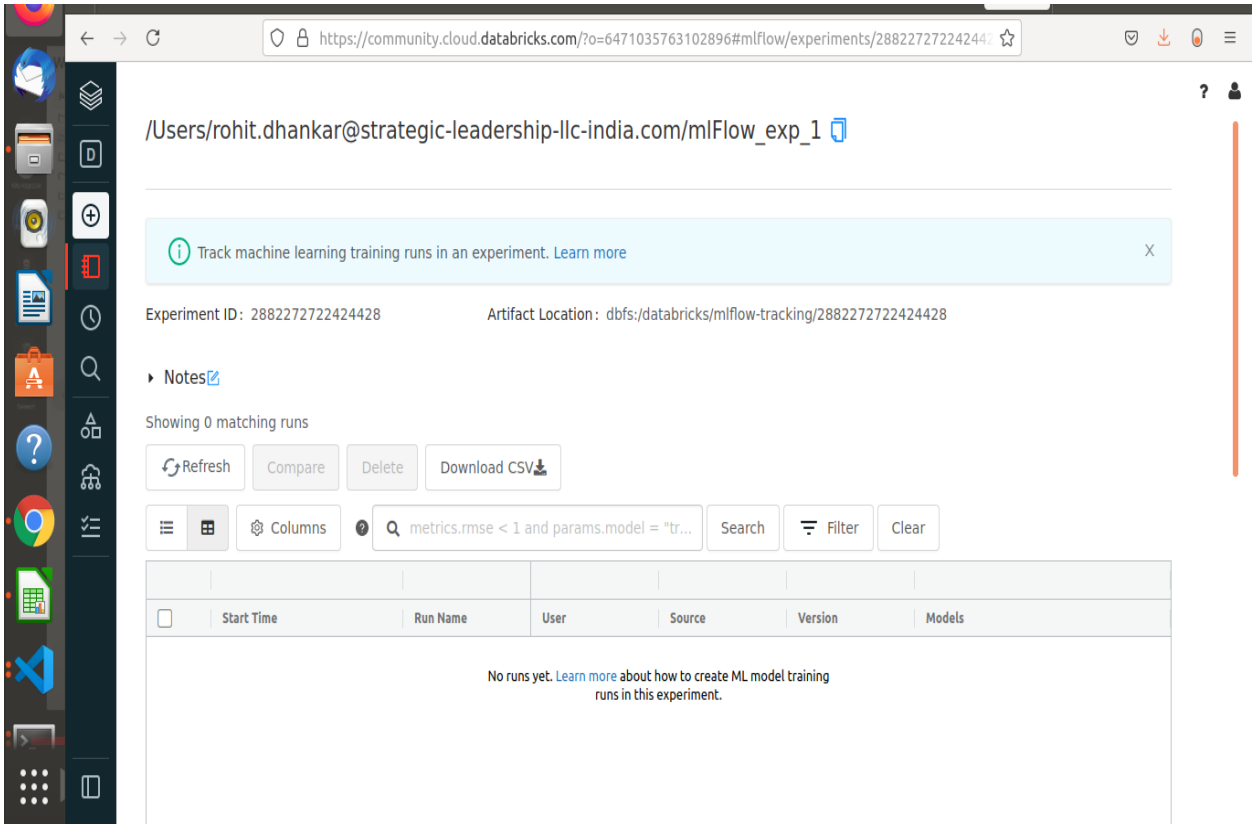
Table of Contents -

GIT → https://github.com/mlflow/mlflow
PyPi → https://pypi.org/project/mlflow/
SO → https://stackoverflow.com/questions/tagged/mlflow

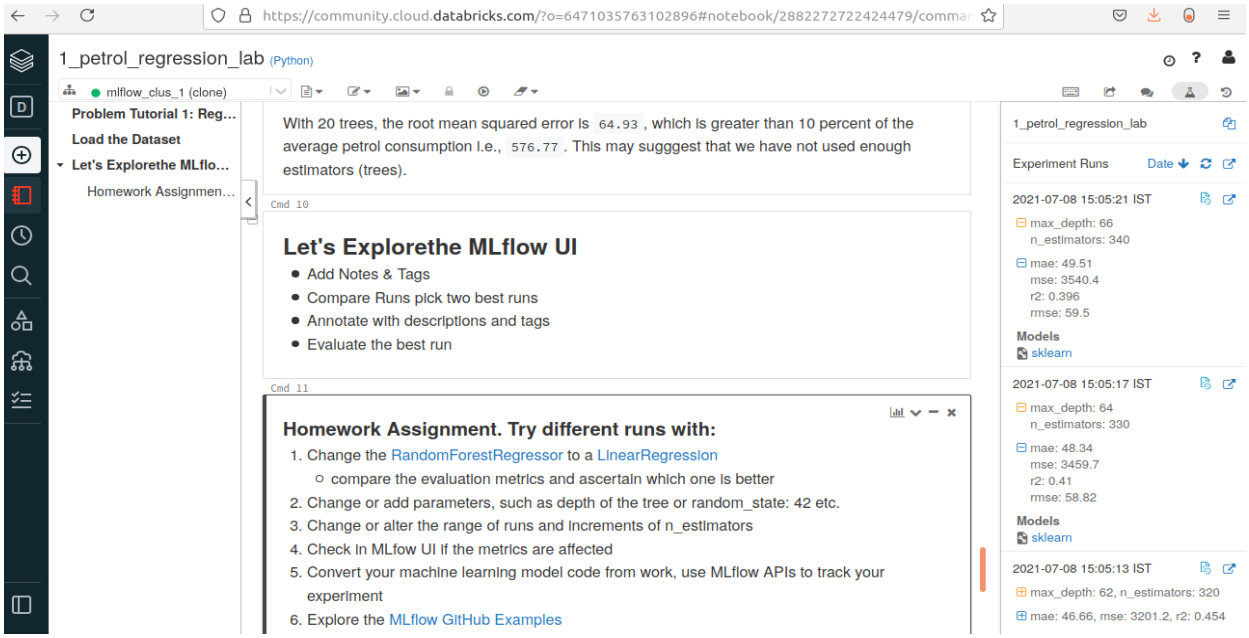
$ pip install mlflow
$ pip3 freeze | grep mlflow
mlflow==1.18.0

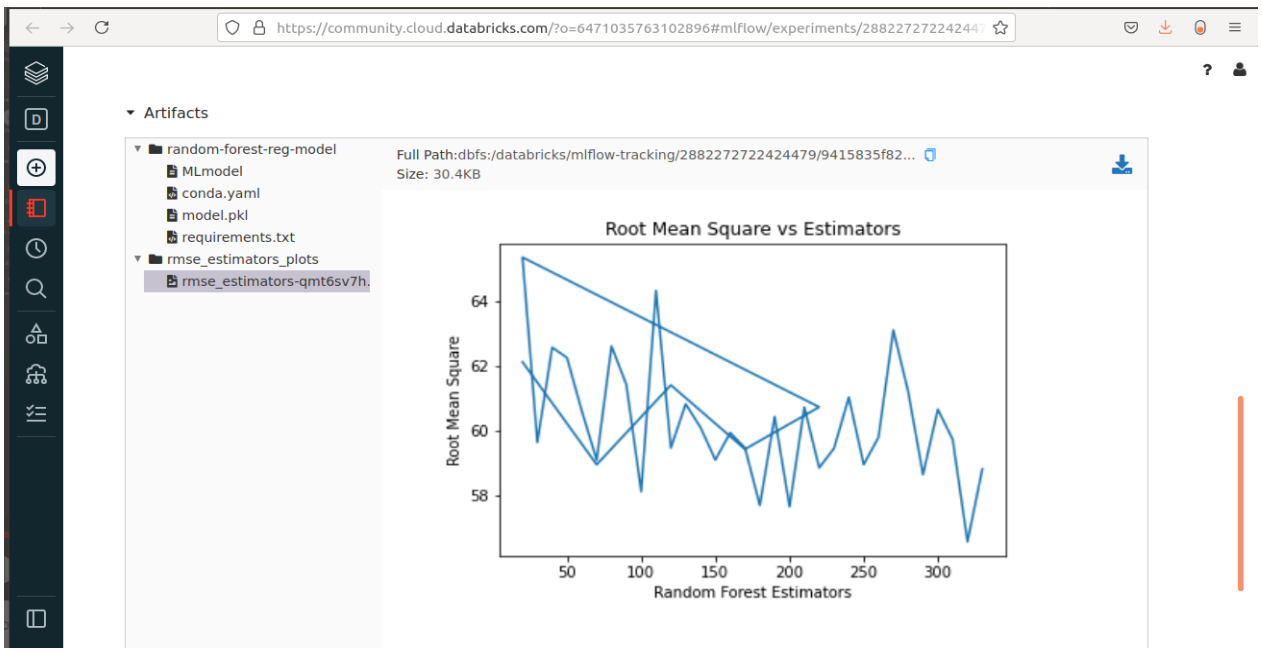
Spark MLFlow on Azure Data Bricks **Community Edition**





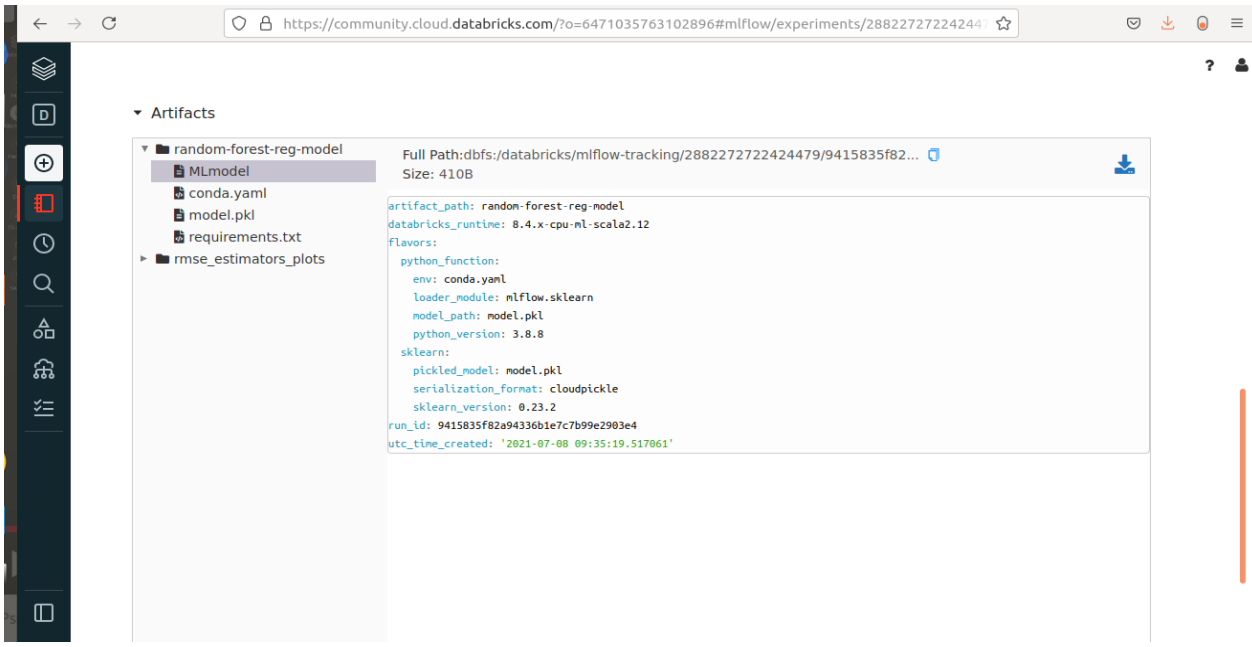
Following along with the official tutorials - we run the code on the Community Instance .
Screen captures as below - Code files attached
Notice on the top right side we have the **MIFLow - Tracking API (Widgets)**
/Users/<your_EmailID_here>/MLflow-CE/1_petrol_regression_lab





The RMSE vs. ESTIMATORS are diff from the tutorials as we are using Diff values for input params

YAML File with Model Flavours



random-forest-reg-model

- **MLmodel**
- conda.yaml
- model.pkl
- requirements.txt
- rmse_estimators_plots

Full Path: dbfs:/databricks/mlflow-tracking/2882272722424479/9415835f82...
Size: 410B

```yaml
artifact_path: random-forest-reg-model
databricks_runtime: 8.4.x-cpu-ml-scala2.12
flavors:
  python_function:
    env: conda.yaml
    loader_module: mlflow.sklearn
    model_path: model.pkl
    python_version: 3.8.8
  sklearn:
    pickled_model: model.pkl
    serialization_format: cloudpickle
    sklearn_version: 0.23.2
run_id: 9415835f82a94336b1e7c7b99e2903e4
utc_time_created: '2021-07-08 09:35:19.517061'
```

random-forest-reg-model

- MLmodel
- **conda.yaml**
- model.pkl
- requirements.txt
- rmse_estimators_plots

```
Full Path: dbfs:/databricks/mlflow-tracking/2882272722424479/9415835f82...
Size: 142B
channels:
- conda-forge
dependencies:
- python=3.8.8
- pip
- pip:
  - mlflow
  - scikit-learn==0.23.2
  - cloudpickle==1.6.0
name: mlflow-env
```

The model.pkl File

Example MLProjects File --

https://github.com/mlflow/mlflow-example

https://raw.githubusercontent.com/mlflow/mlflow/master/examples/sklearn_elasticnet_wine/MLproject

```
name: tutorial

conda_env: conda.yaml

entry_points:
  main:
    parameters:
      alpha: {type: float, default: 0.5}
      l1_ratio: {type: float, default: 0.1}
    command: "python train.py {alpha} {l1_ratio}"
```

Example Conda.yaml File

https://github.com/mlflow/mlflow/blob/master/examples/sklearn_elasticnet_wine/conda.yaml

```
name: tutorial
channels:
  - conda-forge
dependencies:
  - python=3.6
  - pip
  - pip:
    - scikit-learn==0.23.2
    - mlflow>=1.0
```

Local system - Running MLFlow from a URI

Project is the sample project from official MlFlow GITHUB

mlflow run https://github.com/mlflow/mlflow-example.git -P alpha=0.4

```
(dbfs_env) dhankar@dhankar-1:~/.../0521$ pip3 freeze | grep mlflow
mlflow==1.18.0
(dbfs_env) dhankar@dhankar-1:~/.../0521$ mkdir test_mlflow_run
(dbfs_env) dhankar@dhankar-1:~/.../0521$ cd test_mlflow_run
(dbfs_env) dhankar@dhankar-1:~/.../test_mlflow_run$

#

(dbfs_env) dhankar@dhankar-1:~/.../test_mlflow_run$ mlflow run
https://github.com/mlflow/mlflow-example.git -P alpha=0.4
2021/07/09 12:04:01 INFO mlflow.projects.utils: === Fetching project from
https://github.com/mlflow/mlflow-example.git into /tmp/tmprucs0le3 ===
2021/07/09 12:04:07 INFO mlflow.utils.conda: === Creating conda
environment mlflow-1abc00771765dd9dd15731cbda4938c765fbb90b ===
Collecting package metadata (repodata.json): done
Solving environment: done
```

```
==> WARNING: A newer version of conda exists. <==
 current version: 4.8.3
 latest version: 4.10.3

Please update conda by running

   $ conda update -n base -c defaults conda



Preparing transaction: done
Verifying transaction: done
Executing transaction: done
Ran pip subprocess with arguments:
['/home/dhankar/anaconda3/envs/mlflow-1abc00771765dd9dd15731cbda4938c765fb
b90b/bin/python', '-m', 'pip', 'install', '-U', '-r',
'/tmp/tmprucs0le3/condaenv.kuw5jjip.requirements.txt']
Pip subprocess output:
Collecting mlflow
 Using cached mlflow-1.18.0-py3-none-any.whl (14.2 MB)
Collecting sqlalchemy
 Downloading
SQLAlchemy-1.4.20-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.manyli
nux_2_17_x86_64.manylinux2014_x86_64.whl (1.5 MB)
Requirement already satisfied: click>=7.0 in
/home/dhankar/.local/lib/python3.7/site-packages (from mlflow->-r
/tmp/tmprucs0le3/condaenv.kuw5jjip.requirements.txt (line 1)) (7.1.1)
……….
Successfully built alembic databricks-cli prometheus-flask-exporter
Installing collected packages: zipp, typing-extensions, MarkupSafe,
importlib-metadata, Werkzeug, urllib3, smmap, Jinja2, itsdangerous, idna,
greenlet, click, chardet, websocket-client, tabulate, sqlalchemy,
requests, python-editor, pyparsing, prometheus-client, Mako, gitdb, Flask,
sqlparse, querystring-parser, pyyaml, protobuf, prometheus-flask-exporter,
packaging, gunicorn, gitpython, entrypoints, docker, databricks-cli,
cloudpickle, alembic, mlflow
 Attempting uninstall: click
   Found existing installation: click 7.1.1
   Uninstalling click-7.1.1:
```

```
     Successfully uninstalled click-7.1.1
Successfully installed Flask-2.0.1 Jinja2-3.0.1 Mako-1.1.4
MarkupSafe-2.0.1 Werkzeug-2.0.1 alembic-1.4.1 chardet-4.0.0 click-8.0.1
cloudpickle-1.6.0 databricks-cli-0.14.3 docker-5.0.0 entrypoints-0.3
gitdb-4.0.7 gitpython-3.1.18 greenlet-1.1.0 gunicorn-20.1.0 idna-2.10
importlib-metadata-4.6.1 itsdangerous-2.0.1 mlflow-1.18.0 packaging-21.0
prometheus-client-0.11.0 prometheus-flask-exporter-0.18.2 protobuf-3.17.3
pyparsing-2.4.7 python-editor-1.0.4 pyyaml-5.4.1 querystring-parser-1.2.4
requests-2.25.1 smmap-4.0.0 sqlalchemy-1.4.20 sqlparse-0.4.1
tabulate-0.8.9 typing-extensions-3.10.0.0 urllib3-1.26.6
websocket-client-1.1.0 zipp-3.5.0

#
# To activate this environment, use
#
#     $ conda activate mlflow-1abc00771765dd9dd15731cbda4938c765fbb90b
#
# To deactivate an active environment, use
#
#     $ conda deactivate

2021/07/09 12:05:48 INFO mlflow.projects.utils: === Created directory
/tmp/tmplc7eo1z6 for downloading remote URIs passed to arguments of type
'path' ===
2021/07/09 12:05:48 INFO mlflow.projects.backend.local: === Running
command 'source /home/dhankar/anaconda3/bin/../etc/profile.d/conda.sh &&
conda activate mlflow-1abc00771765dd9dd15731cbda4938c765fbb90b 1>&2 &&
python train.py 0.4 0.1' in run with ID 'b69e1632c6764308a799adc175dc8e12'
===
/home/dhankar/anaconda3/envs/mlflow-1abc00771765dd9dd15731cbda4938c765fbb9
0b/lib/python3.7/site-packages/sklearn/utils/__init__.py:4:
DeprecationWarning: Using or importing the ABCs from 'collections' instead
of from 'collections.abc' is deprecated since Python 3.3,and in 3.9 it
will stop working
 from collections import Sequence
/home/dhankar/anaconda3/envs/mlflow-1abc00771765dd9dd15731cbda4938c765fbb9
0b/lib/python3.7/site-packages/sklearn/model_selection/_split.py:18:
DeprecationWarning: Using or importing the ABCs from 'collections' instead
of from 'collections.abc' is deprecated since Python 3.3,and in 3.9 it
will stop working
```

```
 from collections import Iterable
/home/dhankar/anaconda3/envs/mlflow-1abc00771765dd9dd15731cbda4938c765fbb90b/lib/python3.7/site-packages/sklearn/model_selection/_search.py:16:
DeprecationWarning: Using or importing the ABCs from 'collections' instead of from 'collections.abc' is deprecated since Python 3.3,and in 3.9 it will stop working
 from collections import Mapping, namedtuple, defaultdict, Sequence
Elasticnet model (alpha=0.400000, l1_ratio=0.100000):
 RMSE: 0.7909069124367867
 MAE: 0.6174288492244517
 R2: 0.19207580388574486
2021/07/09 12:05:49 INFO mlflow.projects: === Run (ID 'b69e1632c6764308a799adc175dc8e12') succeeded ===
(dbfs_env) dhankar@dhankar-1:~/.../test_mlflow_run$ ls -ltr
total 4
drwxr-xr-x 4 dhankar dhankar 4096 Jul  9 12:04 mlruns
(dbfs_env) dhankar@dhankar-1:~/.../test_mlflow_run$
```

As seen above - we have results from the SkLearn Elasticnet Model

```
Elasticnet model (alpha=0.400000, l1_ratio=0.100000):
 RMSE: 0.7909069124367867
 MAE: 0.6174288492244517
 R2: 0.19207580388574486
```

Model signature

from mlflow.models.signature import infer_signature
from mlflow.models.signature import ModelSignature

MLFlow PROJECTS

MultiStep projects -- mlflow/examples/multistep_workflow/MLproject

https://github.com/mlflow/mlflow/blob/c635d1aa12e1749ab1321128ac61c0f3e6309c1d/examples/multistep_workflow/MLproject

```
name: multistep_example

conda_env: conda.yaml

entry_points:
 load_raw_data:
   command: "python load_raw_data.py"

 etl_data:
   parameters:
     ratings_csv: path
     max_row_limit: {type: int, default: 100000}
   command: "python etl_data.py --ratings-csv {ratings_csv}
--max-row-limit {max_row_limit}"

 als:
   parameters:
     ratings_data: path
     max_iter: {type: int, default: 10}
     reg_param: {type: float, default: 0.1}
     rank: {type: int, default: 12}
   command: "python als.py --ratings-data {ratings_data} --max-iter
{max_iter} --reg-param {reg_param} --rank {rank}"

 train_keras:
   parameters:
     ratings_data: path
```

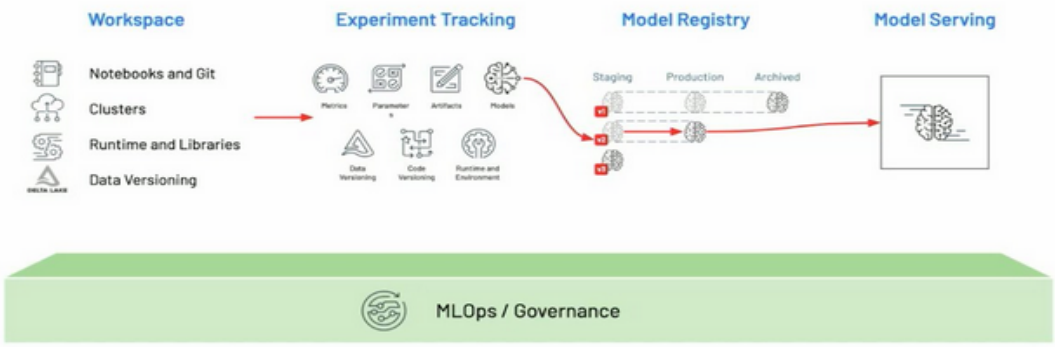
```
    als_model_uri: string
    hidden_units: {type: int, default: 20}
  command: "python train_keras.py --ratings-data {ratings_data}
--als-model-uri {als_model_uri} --hidden-units {hidden_units}"

main:
  parameters:
    als_max_iter: {type: int, default: 10}
    keras_hidden_units: {type: int, default: 20}
    max_row_limit: {type: int, default: 100000}
  command: "python main.py --als-max-iter {als_max_iter}
--keras-hidden-units {keras_hidden_units}
                            --max-row-limit {max_row_limit}"
```

MLFlow MODELS REGISTRY

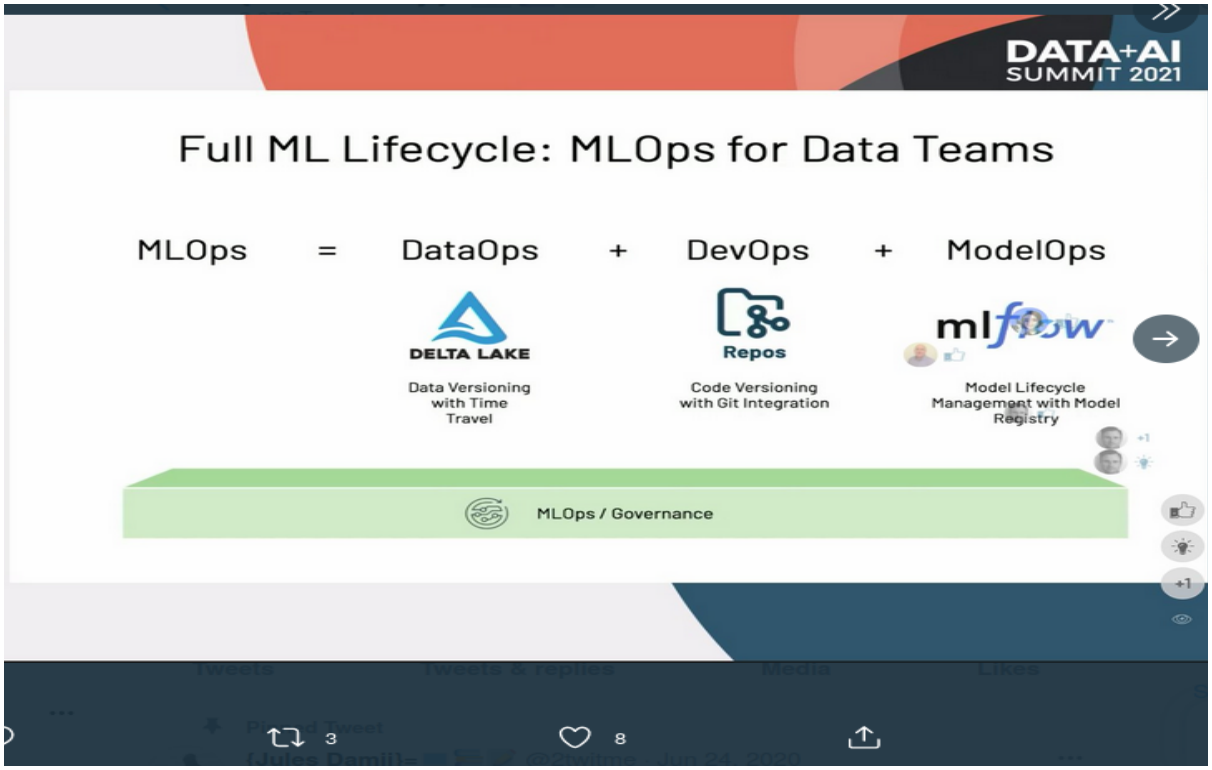
REGISTRY centralized model store, set of APIs, and UI, to collaboratively manage the full lifecycle of MLflow Models.

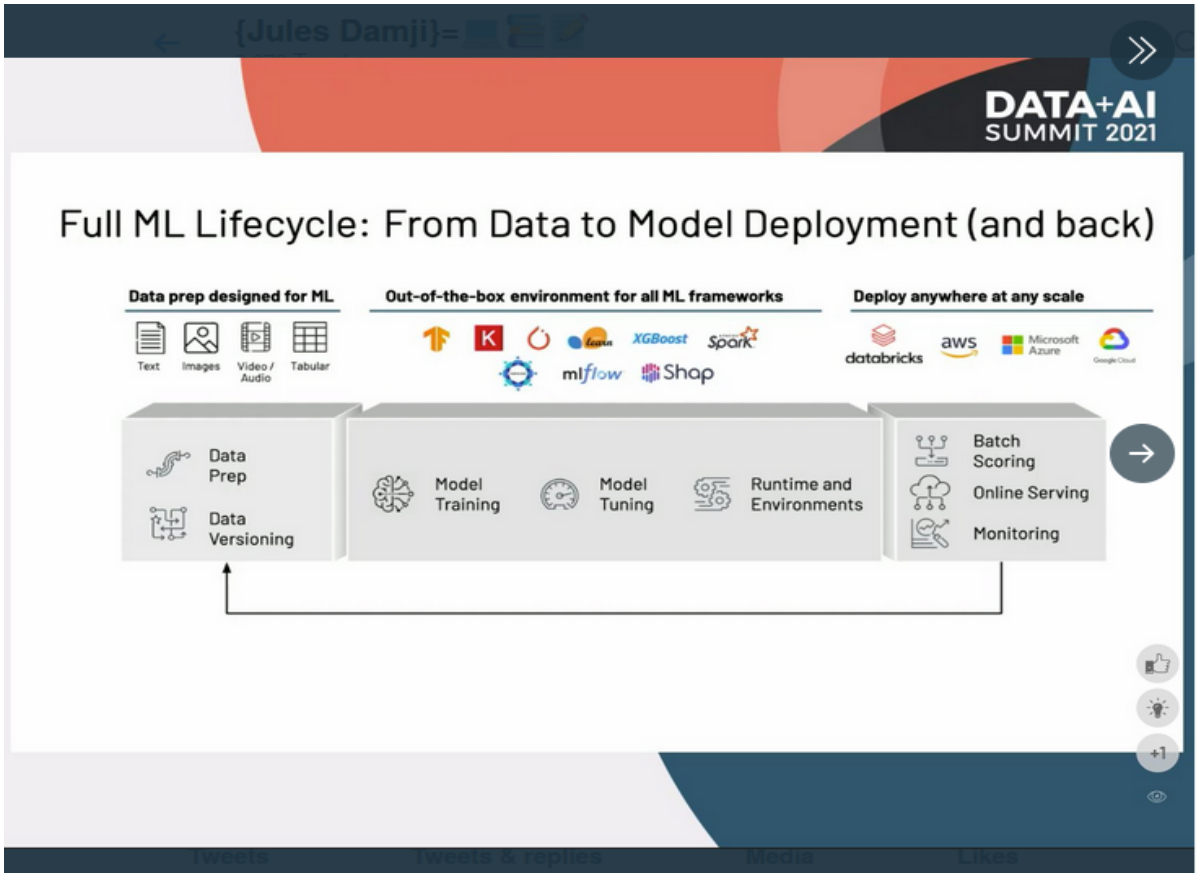
{Jules Damji}=

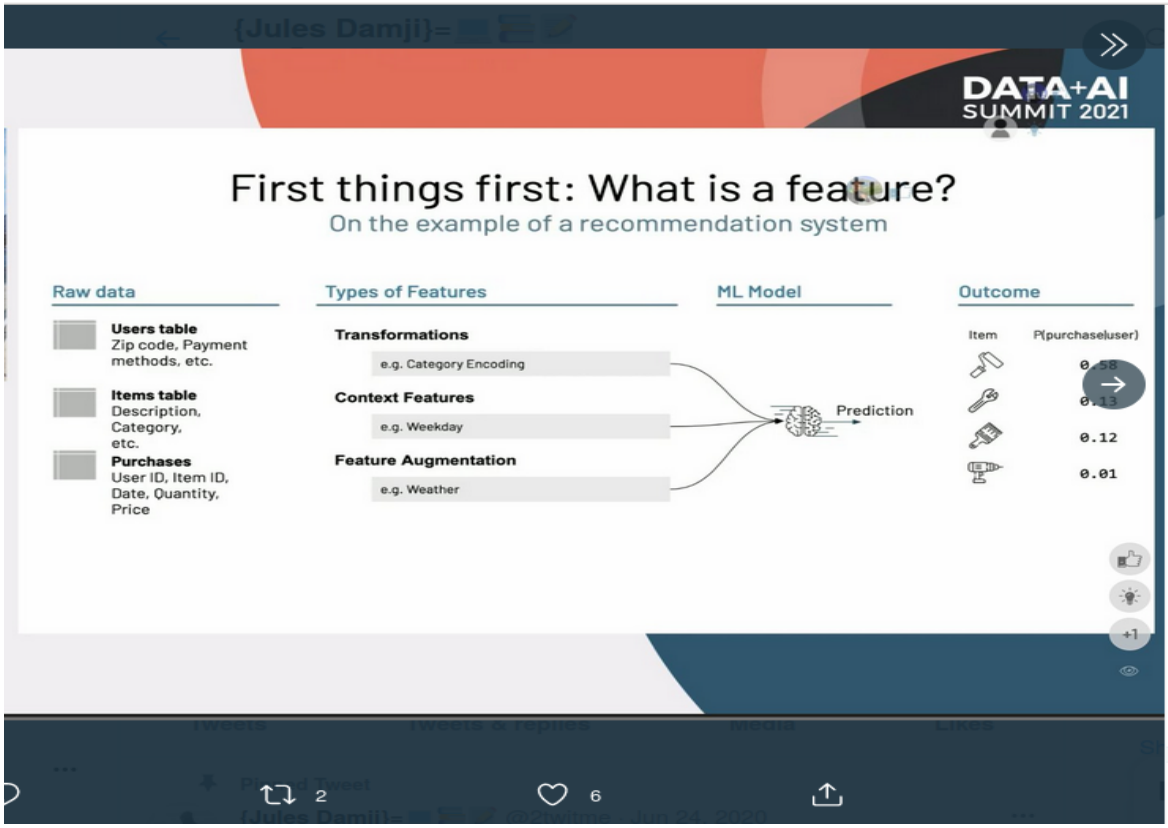
Full ML Lifecycle: How you know you did it right

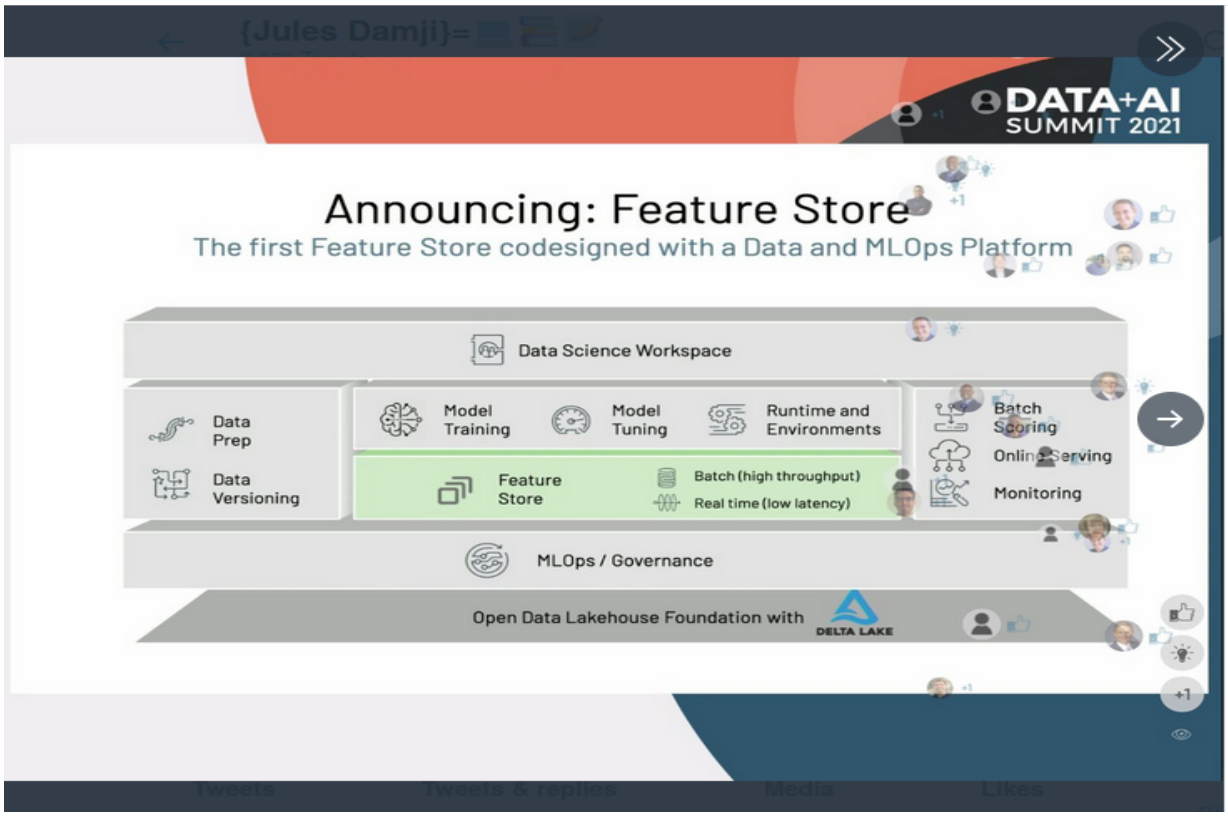


Tweets
Tweets & replies
Media
Likes


DATA+AI
SUMMIT 2021
Full ML Lifecycle: MLOps for Data Teams
MLOps = DataOps + DevOps + ModelOps
DELTA LAKE
Data Versioning with Time Travel
Repos
Code Versioning with Git Integration
mlflow
Model Lifecycle Management with Model Registry
MLOps / Governance


DATA+AI
SUMMIT 2021
Full ML Lifecycle: From Data to Model Deployment (and back)
Data prep designed for ML
Text
Images
Video / Audio
Tabular
Out-of-the-box environment for all ML frameworks
XGBoost
Spark
mlflow
Shap
Deploy anywhere at any scale
databricks
aws
Microsoft Azure
Data Prep
Data Versioning
Model Training
Model Tuning
Runtime and Environments
Batch Scoring
Online Serving
Monitoring


DATA+AI
SUMMIT 2021
First things first: What is a feature?
On the example of a recommendation system
Raw data
Users table
Zip code, Payment methods, etc.
Items table
Description, Category, etc.
Purchases
User ID, Item ID, Date, Quantity, Price
Types of Features
Transformations
e.g. Category Encoding
Context Features
e.g. Weekday
Feature Augmentation
e.g. Weather
ML Model
Prediction
Outcome
Item
P(purchase|user)
0.58
0.13
0.12
0.01


DATA+AI
SUMMIT 2021
Announcing: Feature Store
The first Feature Store codesigned with a Data and MLOps Platform
Data Science Workspace
Data Prep
Data Versioning
Model Training
Model Tuning
Runtime and Environments
Feature Store
Batch (high throughput)
Real time (low latency)
Batch Scoring
Online Serving
Monitoring
MLOps / Governance
Open Data Lakehouse Foundation with
DELTA LAKE