



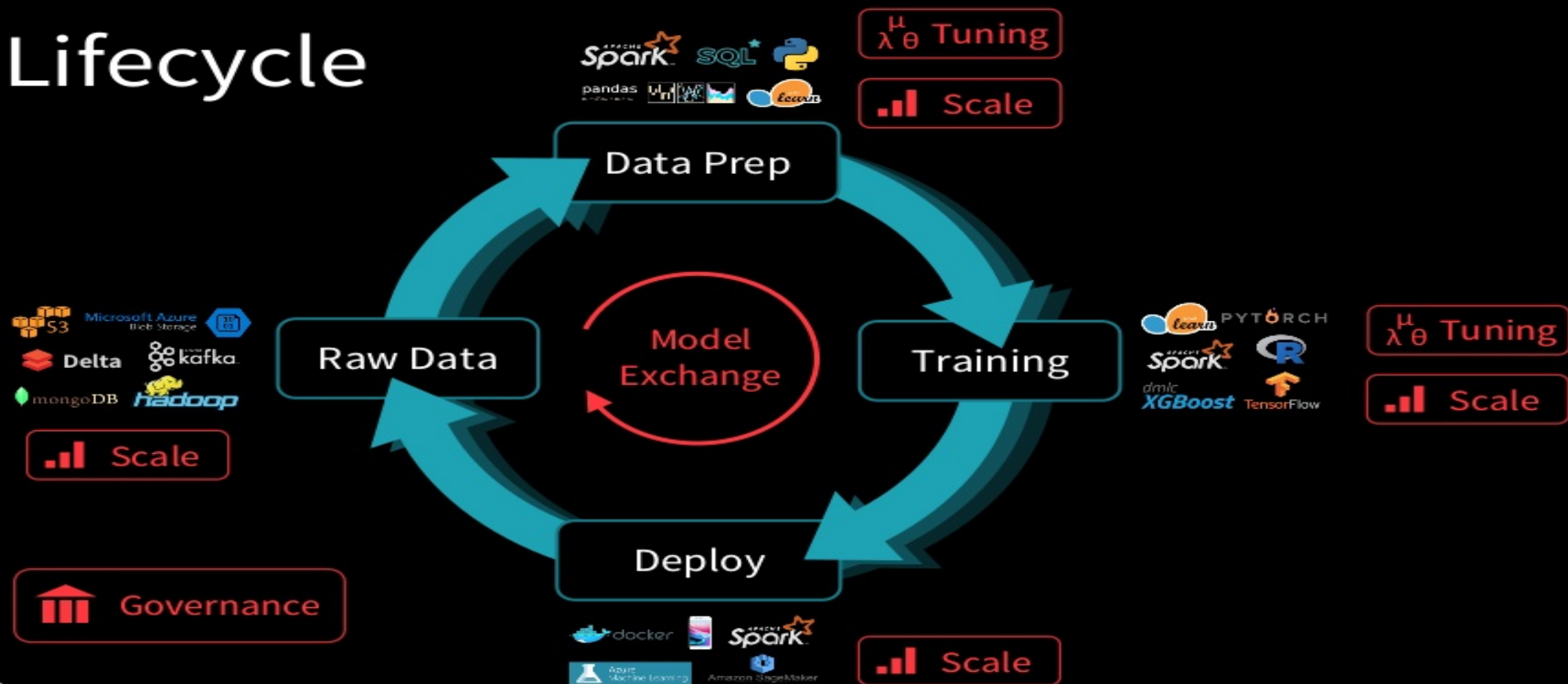
mlflow: Accelerating Production Machine Learning

Matei Zaharia
@matei_zaharia



Machine Learning Development is Complex

ML Lifecycle



Custom ML Platforms

Facebook FBLearner, Uber Michelangelo, Google TFX

**+ Standardize the data prep / training / deploy cycle:
if you work within the platform, you get these!**

– Limited to a few algorithms or frameworks

– Tied to one company's infrastructure

Can we provide similar benefits in an **open manner?**

Introducing *mlflow*

Open source machine learning platform

- Works with any ML library & language
- Runs the same way everywhere & cross-cloud
- Scales to big data with Apache Spark

Launched this June

- Already 48 contributors and many new features!

This Talk: MLflow Overview + New Announcements

How Does MLflow Work?

mlflow Tracking

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

mlflow Projects

Package workflows into reproducible and reusable steps

mlflow Models

Model packaging format for diverse deployment tools

Model Development without MLflow

```
data    = load_text(file)
ngrams  = extract_ngrams(data, N=n)
model   = train_model(ngrams,
                      learning_rate=lr)
score   = compute_accuracy(model)

print("For n=%d, lr=%f: accuracy=%f"
      % (n, lr, score))

pickle.dump(model, open("model.pkl"))
```

```
For n=2, lr=0.1: accuracy=0.71
For n=2, lr=0.2: accuracy=0.79
For n=2, lr=0.5: accuracy=0.83
For n=2, lr=0.9: accuracy=0.79
For n=3, lr=0.1: accuracy=0.83
For n=3, lr=0.2: accuracy=0.82
For n=4, lr=0.5: accuracy=0.75
...
```

What version of
my code was this
result from?

Model Deployment without MLflow

DATA
SCIENTIST



Code & Models



PRODUCTION
ENGINEER

Please deploy this
ArXiv paper!

...

Model Development *with* MLflow

```
data    = load_text(file)
ngrams  = extract_ngrams(data, N=n)
model   = train_model(ngrams,
                      learning_rate=lr)
score   = compute_accuracy(model)

print("For n=%d, lr=%f: accuracy=%f"
      % (n, lr, score))

pickle.dump(model, open("model.pkl"))
```

Model Development *with* MLflow

```
data    = load_text(file)
ngrams  = extract_ngrams(data, N=n)
model   = train_model(ngrams,
                      learning_rate=lr)
score   = compute_accuracy(model)
```

```
mlflow.log_param("data_file", file)
mlflow.log_param("n", n)
mlflow.log_param("learning_rate", lr)
mlflow.log_metric("score", score)
```

```
mlflow.sklearn.log_model(model)
```

mlflow

GitHub Docs

Language Model

Experiment ID: 0 Artifact Location: /Users/mlflow/mlruns/0

Search Runs: State:

Filter Params:


10 matching runs

					Parameters			Metrics	
<input type="checkbox"/>	Date	User	Name	Version	input_size	n	lr	accuracy	F1
<input type="checkbox"/>	2018-12-02 21:33:57	mlata	l2_lang_model.py	eebcb8	data_size	2.0	1	0.77	0.754
<input type="checkbox"/>	2018-12-02 21:33:59	mlata	l2_lang_model.py	eebcb8	data_size	2.0	4	0.88	0.859
<input type="checkbox"/>	2018-12-02 21:33:44	mlata	l2_lang_model.py	eebcb8	data_size	2.0	2	0.90	0.478
<input type="checkbox"/>	2018-12-02 21:33:53	mlata	l2_lang_model.py	eebcb8	data_size	1.0	1	0.98	0.469
<input type="checkbox"/>	2018-12-02 21:33:49	mlata	l2_lang_model.py	eebcb8	data_size	1.0	4	0.92	0.465

Track parameters, metrics,
output files & code version


Search using UI or API

MLflow UI: Inspecting Runs


[GitHub](#) [Docs](#)







Language Model

Experiment ID: 0 Artifact Location: /Users/matei/mlflow/mlruns/0

Search Runs:  State: Active ▾ Search


Filter Params: Filter Metrics: Clear

10 matching runs Compare Delete Download CSV 

<input type="checkbox"/>	Date ▾	User	Source	Version	Parameters			Metrics	
					input_file	lr	n	accuracy	f1
<input type="checkbox"/>	2018-10-02 21:53:57	matei	 lang_model.py	e55d56	data.txt	2.0	1	0.77	0.704
<input type="checkbox"/>	2018-10-02 21:53:56	matei	 lang_model.py	e55d56	data.txt	1.0	2	0.254	0.222
<input type="checkbox"/>	2018-10-02 21:53:55	matei	 lang_model.py	e55d56	data.txt	2.0	4	0.835	0.609
<input type="checkbox"/>	2018-10-02 21:53:53	matei	 lang_model.py	e55d56	data.txt	1.0	1	0.663	0.468
<input type="checkbox"/>	2018-10-02 21:53:52	matei	 lang_model.py	e55d56	data.txt	0.2	4	0.034	0.032
<input type="checkbox"/>	2018-10-02 21:53:51	matei	 lang_model.py	e55d56	data.txt	0.1	4	0.177	0.16




MLflow UI: Comparing Runs


[GitHub](#) [Docs](#)







Language Model

Experiment ID: 0 Artifact Location: /Users/matei/mlflow/mlruns/0

Search Runs:  State: Active ▾ Search

Filter Params: Filter Metrics: Clear

10 matching runs Compare Delete Download CSV 

<input type="checkbox"/>	Date ▾	User	Source	Version	Parameters			Metrics	
					input_file	lr	n	accuracy	f1
<input type="checkbox"/>	2018-10-02 21:53:57	matei	 lang_model.py	e55d56	data.txt	2.0	1	0.77	0.704
<input type="checkbox"/>	2018-10-02 21:53:56	matei	 lang_model.py	e55d56	data.txt	1.0	2	0.254	0.222
<input type="checkbox"/>	2018-10-02 21:53:55	matei	 lang_model.py	e55d56	data.txt	2.0	4	0.835	0.609
<input type="checkbox"/>	2018-10-02 21:53:53	matei	 lang_model.py	e55d56	data.txt	1.0	1	0.663	0.468
<input type="checkbox"/>	2018-10-02 21:53:52	matei	 lang_model.py	e55d56	data.txt	0.2	4	0.034	0.032
<input type="checkbox"/>	2018-10-02 21:53:51	matei	 lang_model.py	e55d56	data.txt	0.1	4	0.177	0.16



Packaging Code: MLflow Projects



Local Execution

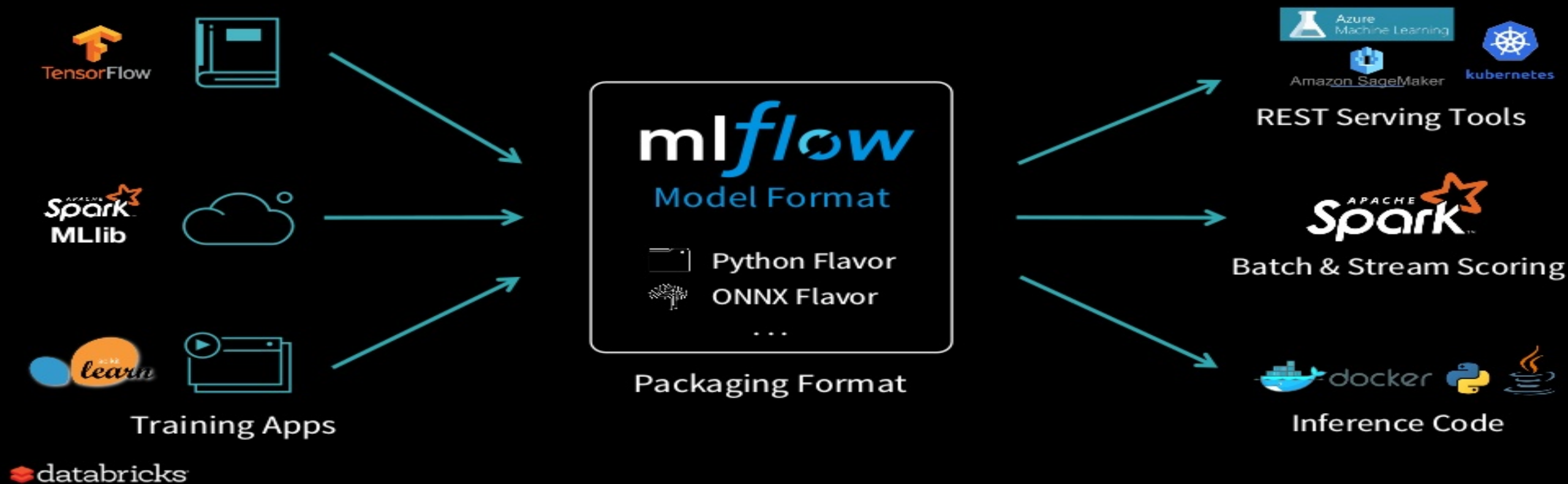


Remote Cluster



```
$ mlflow run git://...
```

Packaging Models: MLflow Models



Model Deployment *with* MLflow

DATA
SCIENTIST



PRODUCTION
ENGINEER

Please run this
[MLflow Project](#)
nightly for updates!

Don't even tell me
what ArXiv paper
that's from...

MLflow Development Status

Many new features since our release in June

- Model packaging for MLlib, H2O, TensorFlow, PyTorch, Keras
- Storage on Azure, AWS, Google, SFTP
- Java and Scala API
- New examples and UI features

Just released MLflow 0.7.0 today!

mlflow

Major Announcement in MLflow 0.7.0



RStudio partnered with Databricks to add an MLflow R API

See Kevin Kuo's talk on this at 14:40 today!

Conclusion

Workflow platforms can greatly simplify ML development

- Improve usability for **both** data scientists and engineers

Get started at mlflow.org, and come see our other talks:

- MLflow R API: 14:40 today
- ML Factories: 11:40 Thursday
- 1h Deep Dive: 14:00 Thursday

mlflow