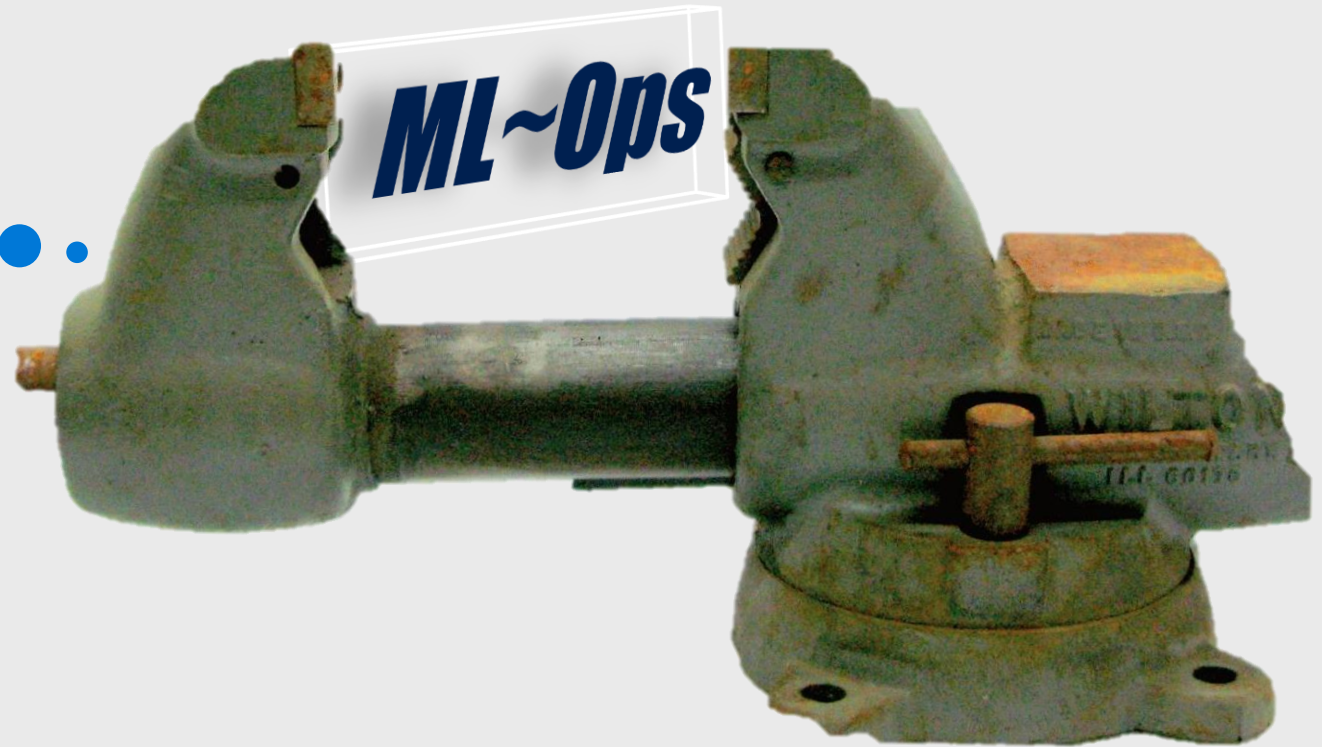
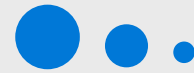


Machine Learning + DevOps *using* Azure ML Services

*ML + DevOps
Together at
Last!*



Rolf Tesmer

Microsoft Cloud Architect | Azure | Data | Analytics | AI

Mr. Fox SQL Blog - <https://mrfoxsql.wordpress.com/>

Linked In - <https://www.linkedin.com/in/rolftesmer/>

Today's Agenda

1. What is DevOps
2. *The zero!* Pain Points in Data Science projects
3. *The hero!* Azure Machine Learning Service
4. Walkthrough Demo
5. Different ways to build ML DevOps pipelines
6. Best Practices... *What should I do now?*



A couple of *assumptions* about you...

1. Have awareness of public cloud and Azure
2. Have a working background with data
3. Are curious about Devops
4. Love solving problems



Microsoft AI

Our strategy is to build best-in-class **platforms** and productivity services for an **intelligent cloud and an intelligent edge** infused with **artificial intelligence** ("AI").

With Azure we want to **democratize AI** so every company can be an AI company, and every partner can build AI solutions, while **helping customers build AI capability to transform.**

What exactly is DevOps? *And Why Should I Care?*

DevOps is a software engineering practice that aims at unifying software development and software operation. The main characteristic of the DevOps movement is to strongly advocate automation and monitoring at all steps of software construction, from integration, testing, releasing to deployment and management.

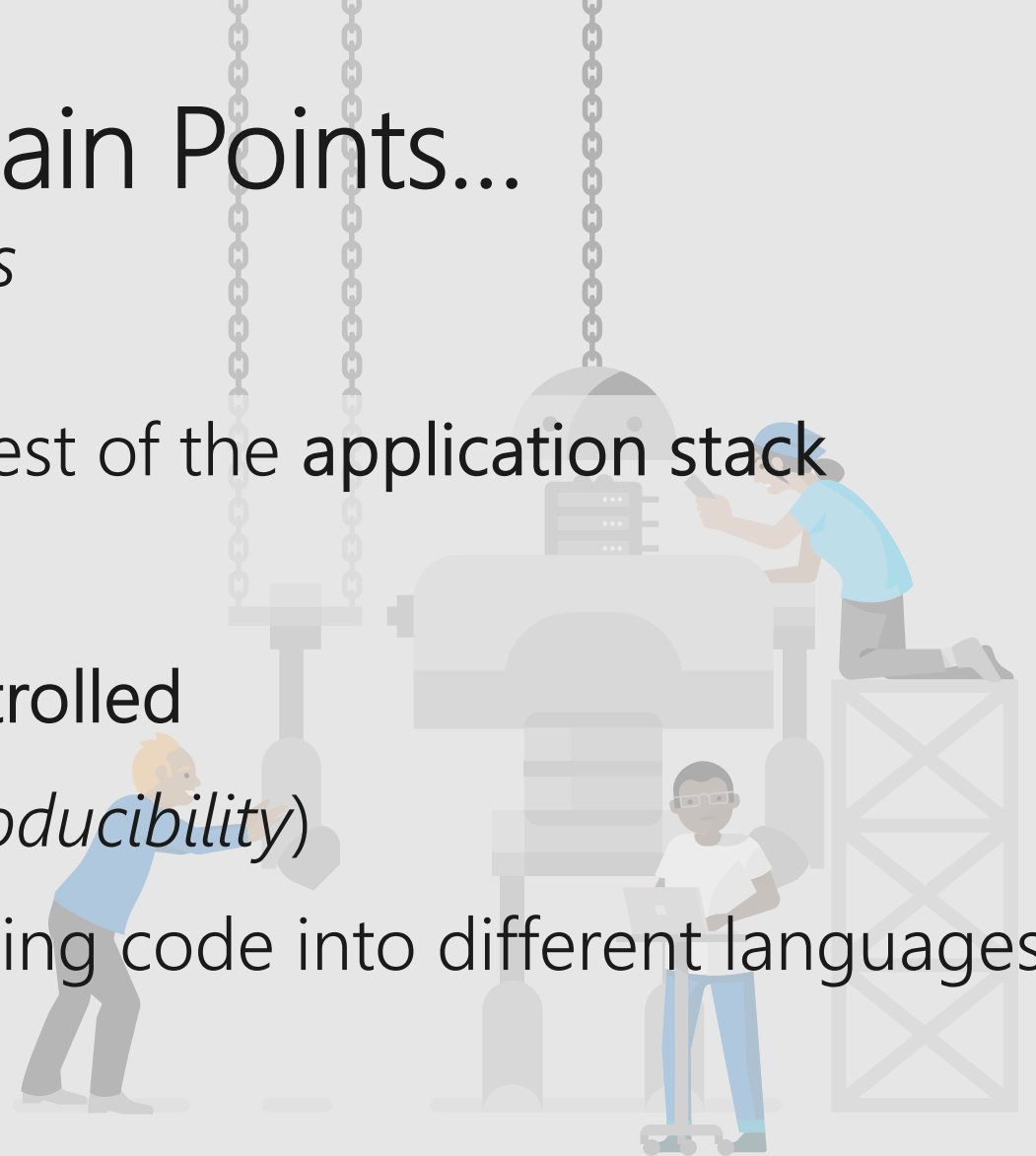
GOAL: *DevOps enables faster time to market, lower failure rate, shortened lead times, automated compliance, release consistency.*

method of development → **Agile != DevOps** ← *method of deployment*

7x Common Data Science Pain Points...

...when Applying DevOps Methods / Practices

1. ML stack might be different from rest of the application stack
2. Testing accuracy of ML model
3. ML code is not always version controlled
4. Hard to reproduce models (*ie reproducibility*)
5. Need to re-write featurizing + scoring code into different languages
6. Hard to track breaking changes
7. Difficult to monitor models & determine when to retrain



Machine Learning Capabilities on Azure

Domain specific pretrained models

To reduce time to market - [Azure Cognitive Services](#)



Vision



Speech



Language



Search

Familiar Data Science tools

To simplify model development



PyCharm



Jupyter



Visual Studio Code



Command line

Popular frameworks

To build advanced deep learning solutions



Pytorch



TensorFlow



Scikit-Learn



Onnx

Productive services

To empower data science and development teams



[Azure Databricks](#)



[Azure Machine Learning](#)



[Machine Learning VMs](#)

Powerful infrastructure

To accelerate deep learning



CPU



[GPU](#)



[FPGA](#)



From the Intelligent Cloud to the Intelligent Edge



Azure ML Concept

Model Management

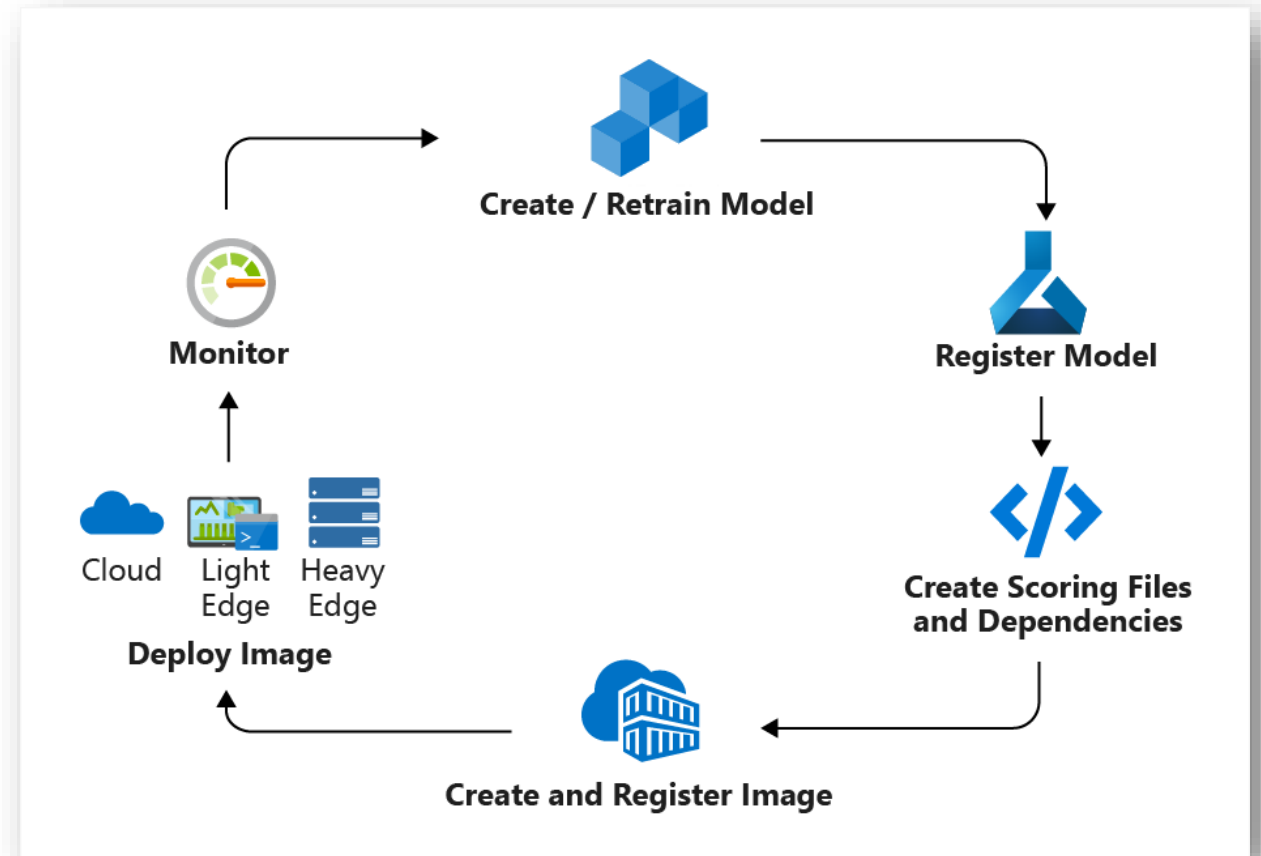
Model Management in Azure ML usually involves these four steps

Step 1: Register Model using the Model Registry

Step 2: Register Image using the Image Registry (the Azure Container Registry)

Step 3: Deploy the Image to cloud or to edge devices

Step 4: Monitor models—you can monitor input, output, and other relevant data from your model.



Walkthrough

Dashboard > AML2-rg > AMLService2

AMLService2

Machine Learning service workspace

Search (Ctrl+/)

- Overview
- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems

Settings

- Locks
- Export template
- Properties

Application

- Experiments
- Pipelines
- Compute
- Models
- Images
- Deployments
- Activities

Support + troubleshooting

- Usage + quotas
- New support request

Resource group: AML2-rg
Location: Australia East
Subscription: MrFoxSQL
Subscription ID: b4bc0fa0-4e1b-49a4-af89-572f07ea0dd9

Getting Started

train (Python)

Attached: CLUST_HC_RTL_P... File View: Code Permission

xvars_multiselect: hours_per_week, sex, education_num

Cmd 1

train.py

```
1 #####
2 # train.py
3 #####
4 # v0.1 rotesmer 17 Feb 2019 Created
5 #####
6 #
7 # Description
8 # This training code will extract Adult Census Data and build
9 # of the individual sample/observation exceeds USD$50K per ye
10 # The Dependant Variable (label) is Income. Possible values
11 # The dataset contains many independant Variables (features). No features are engineered.
12 # Model prediction = 1 --> the income was predicted to be greater than $50K
13 # Model prediction = 0 --> the income was predicted to be less than $50K
14 # Ref Data: https://www.kaggle.com/uciml/adult-census-income
15 #
```

Command took 0.13 seconds -- by rotesmer@microsoft.com at 17/03/2019, 20:37:48 on CLUST_HC_RTL_PY3_W1-4

Cmd 2

Variables, Imports, Check Azure ML SDK Version

```
1 import sys
2 import azureml.core
3 import os
4 import urllib
5 import pprint
6 import numpy as np
7 import shutil
8
9 from azureml.core.run import Run
10 from azureml.core.experiment import Experiment
11 from pyspark.ml import Pipeline, PipelineModel
12 from pyspark.ml.feature import OneHotEncoder, StringIndexer, VectorAssembler
```

Azure DevOps

rotesmer / AML / Pipelines

adult_census CI - CD > adult_census_model-25

Pipeline Variables History + Deploy Cancel Refresh Edit ...

Release

Continuous deployment
for Rolf Tesmer
17/03/2019, 17:46

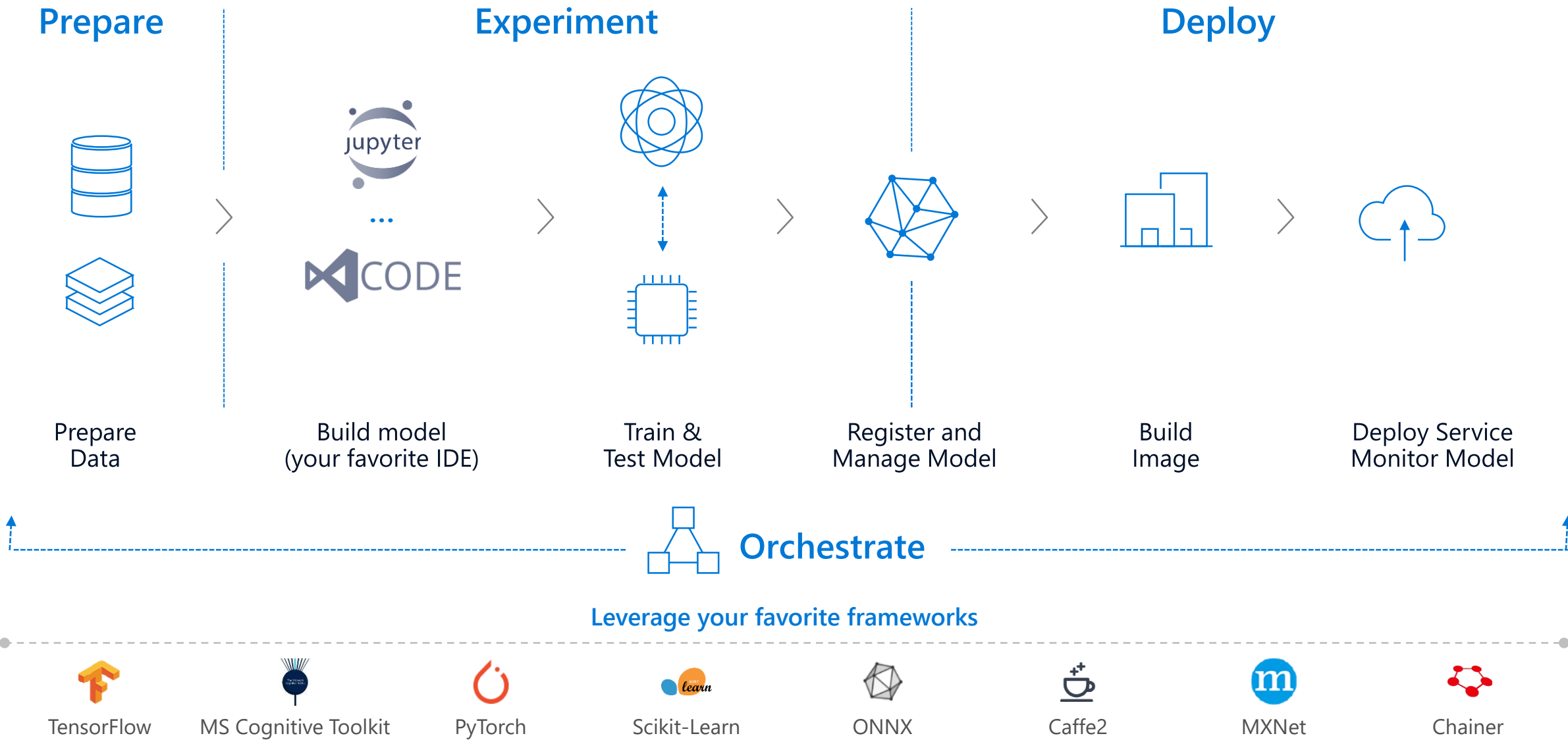
Artifacts
adult_census CI
34
master

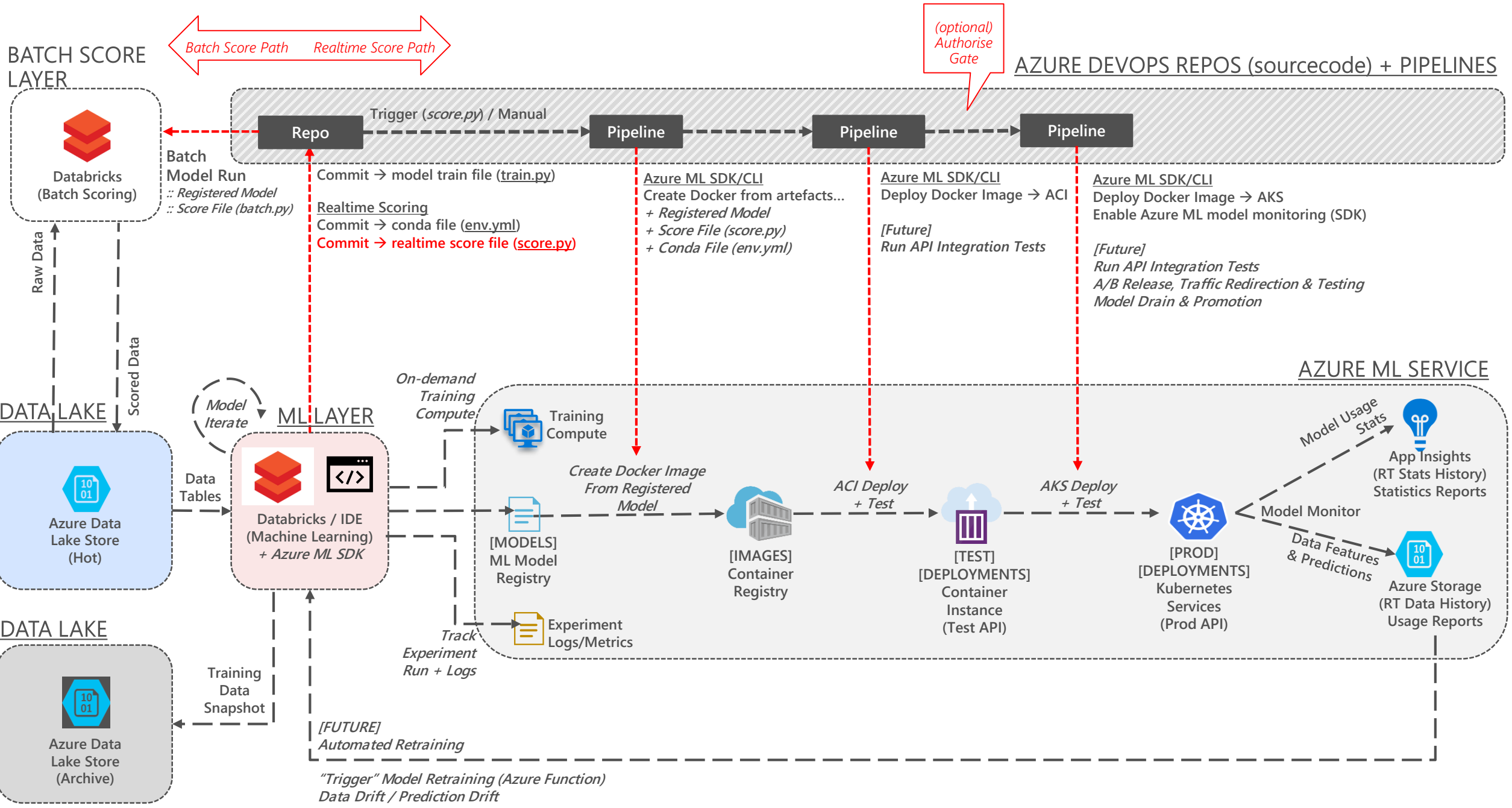
Stages

- Build Test Container
Succeeded
on 17/03/2019, 19:44
- Release
Succeeded
on 17/03/2019, 19:48

Navigation: Overview, Boards, Repos, Pipelines, Builds, Releases, Library, Task groups, Deployment groups, Test Plans, Artifacts, Compliance

Typical Azure Machine Learning E2E Process





5x Reasons to Bother with Model Containerisation

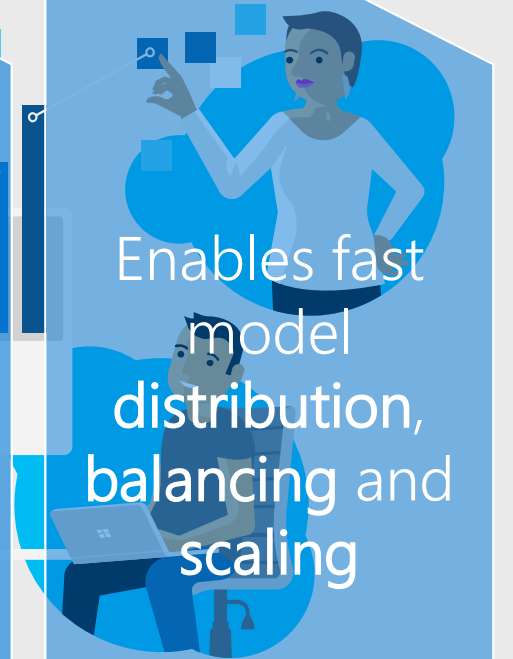
Combines everything needed for the model; libraries, dependencies, etc

Creates a form of physical model versioning and history

Can enable an external model REST API for interaction; batch, req-resp

Offers flexible deployment locations; ie AKS (Kubernetes), edge, etc

Enables fast model distribution, balancing and scaling



How Do I Know When I Should Retrain My Model?

1. Manually – gut feel
2. Time – some schedule
3. Event – trigger based
4. Drift – inputs/outputs exceed some threshold

Data Drift; change in model input data that leads to model performance degradation over time.

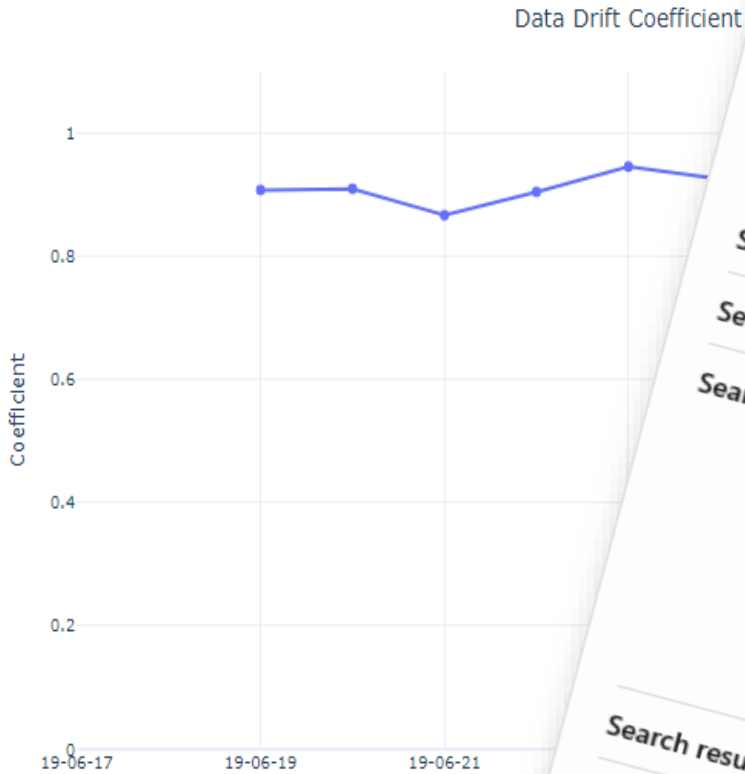
Measure; Monitor deployed model inputs and measure inference data to original train data.

- Measure magnitude drift → drift coefficient
- Measure drift contribution by feature
- Measure distance metrics. Wasserstein & Energy Distance
- Measure feature distribution. Kernel Density Estimation & Histograms

Action; Send drift email alerts. Kick-off Devops Pipeline

How Do I Know

When My Model



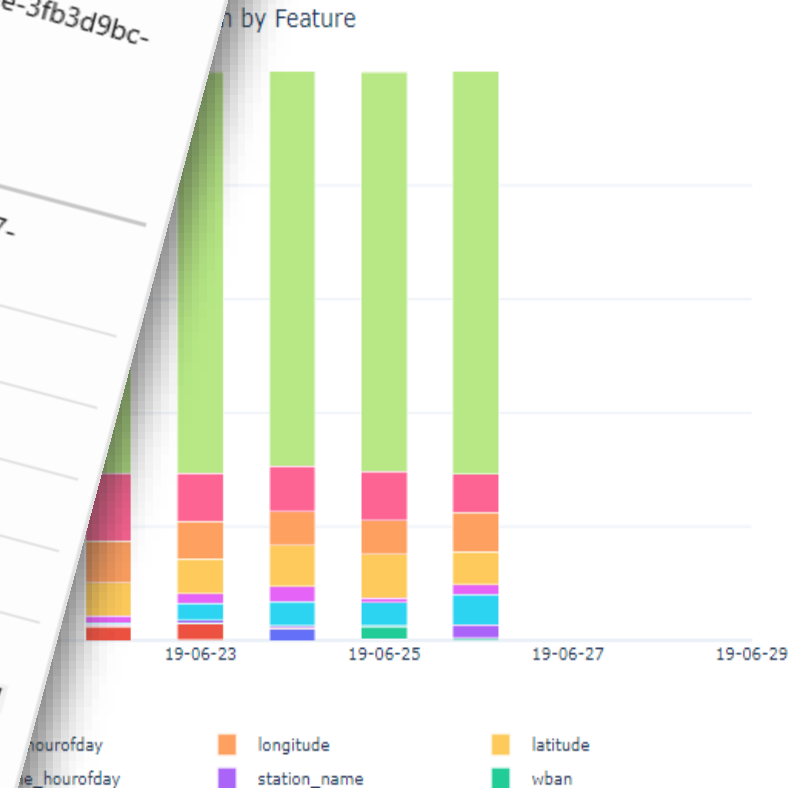
Microsoft Azure

Your Azure Monitor alert was triggered

We are notifying you because there are 1 counts of "DataDrift-Alert-Rule-3fb3d9bc-24f9-4675-b1b7-3e94c11352ac".

Essentials

Name	DataDrift-Alert-Rule-3fb3d9bc-24f9-4675-b1b7-3e94c11352ac
Severity	3
Resource	publicdeinsights6e499d8d
Search interval start time	June 11, 2019 16:59:39 UTC
Search interval duration	5 min
Search query	<pre>customMetrics where (name=='datadrift_coefficient' or name==' ds_mcc_test') and value > 0.2 and customDimension s.datadrift_id == '3fb3d9bc-24f9-4675-b1b7-3e94c1 1352ac' summarize AggregatedValue=count() by bin_at(tim estamp, 1m, datetime(2019-06-11T17:04:39.0000000))</pre>
Search results	1 result(s)
Description	Alert Rule is triggered every time data drift is detected



5x Best Practices for Build/Release Model Pipelines

Different DevOps pipelines **per model**

Pipelines should match **business requirements** for **consumption**

Put model training into its **own DevOps build pipeline**

Retraining **triggers** could be configured **based on ANY metric**

Don't forget about **reproducibility** – **build this into DevOps**

In Summary: How MLOps Supports Data Scientists

continued model code
management + versioning

ie. commit

trained model artefacts
management + versioning

ie. register

automate DevOps build +
release processes

ie. no click ops

training **data retention**

ie. keep data used to train

scoring + prediction **data retention**

ie. model monitoring

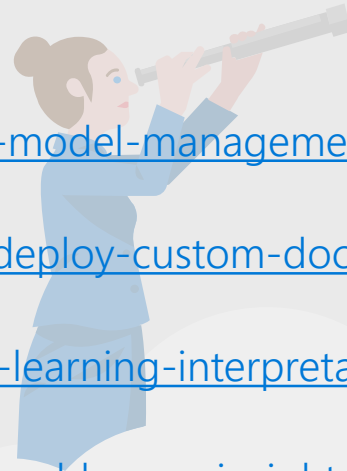
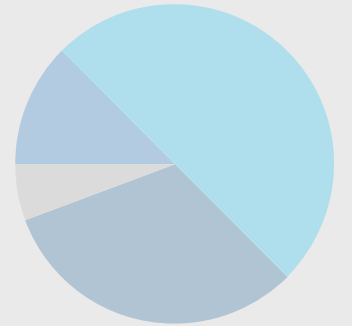




Appendix – References

References

1. <https://github.com/microsoft/MLOps>
2. <https://mrfoxsql.wordpress.com/2019/06/11/machine-learning-devops-ml-devops-together-at-last/>
3. <https://azure.microsoft.com/en-us/services/machine-learning-service/>
4. <https://docs.microsoft.com/en-au/azure/machine-learning/service/concept-model-management-and-deployment>
5. <https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-deploy-custom-docker-image>
6. <https://docs.microsoft.com/en-us/azure/machine-learning/service/machine-learning-interpretability-explainability>
7. <https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-enable-app-insights>
8. <https://docs.microsoft.com/en-au/azure/machine-learning/service/how-to-monitor-data-drift>
9. <https://www.youtube.com/watch?v=nst3UAGpiBA>
10. <https://en.wikipedia.org/wiki/DevOps>



Azure Databricks

Enable collaboration between data scientists and data engineers with an interactive productive workspace

Prepare and clean data at massive scale with the language of your choice

Build and train models with pre-configured machine learning and deep learning optimized clusters



Azure Machine Learning

Track experiments for reproducibility and auditing needs.

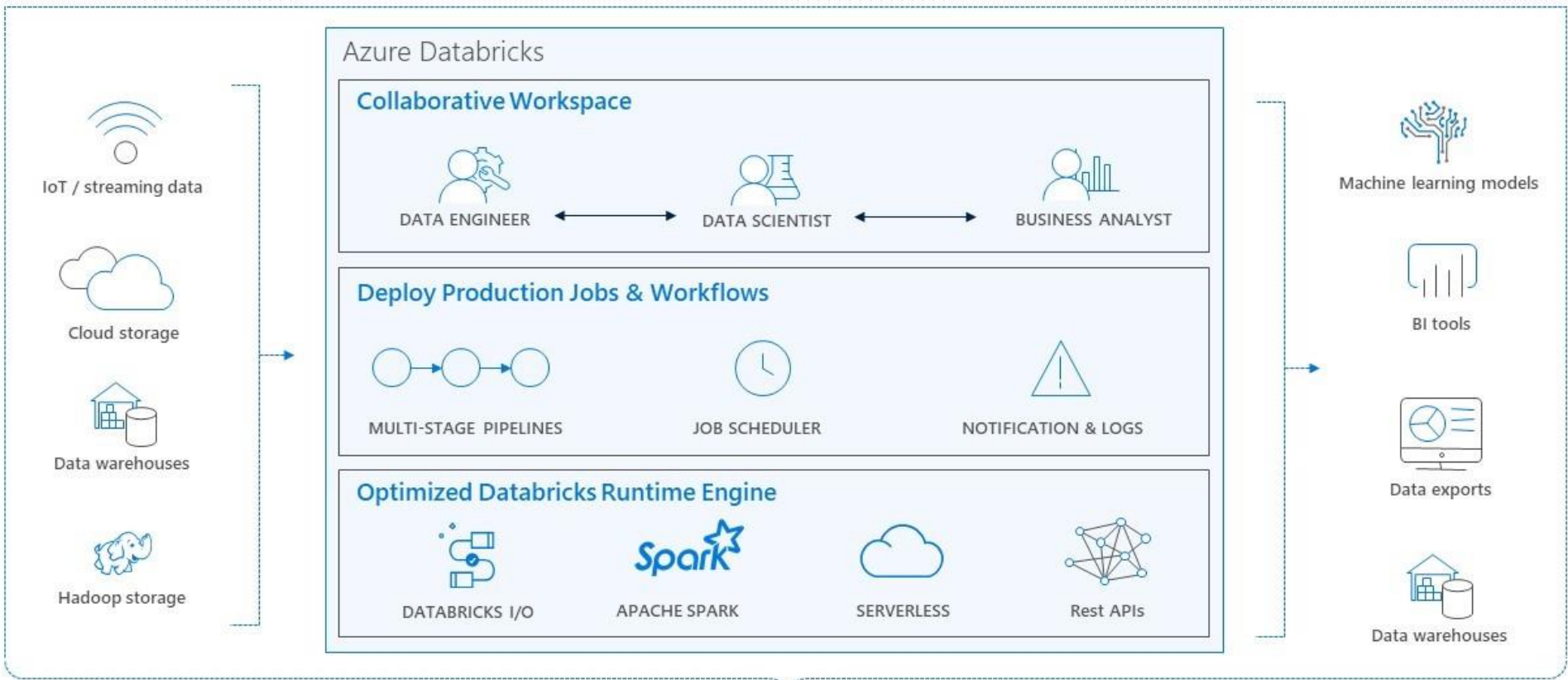
Identify and promote best performing models into production

Deploy and manage your models using containers to run them anywhere

Azure Databricks

<https://azure.microsoft.com/en-us/services/databricks/>

Azure Databricks



Enhance Productivity

Build on secure & trusted cloud

Scale without limits

Azure DevOps

<https://azure.microsoft.com/en-us/solutions/devops/>

Introducing Azure DevOps

*Azure Dev Ops...
...or use your own preferred
DevOps Automation suite
of tooling - !*



Azure
Boards

Plan, track, and discuss work across teams, deliver value to your users faster.



Azure
Repos

Unlimited cloud-hosted private Git repos. Collaborative pull requests, advanced file management, and more.



Azure
Pipelines

CI/CD that works with any language, platform, and cloud. Connect to GitHub or any Git provider and deploy continuously to any cloud.



Azure
Test Plans

The test management and exploratory testing toolkit that lets you ship with confidence.



Azure
Artifacts

Create, host, and share packages. Easily add artifacts to CI/CD pipelines.