



GPS Data/AI Strategy FY23

Delivered by CSA Team



Franck Gaillard
Cloud Solution Architect
Data AI
frgail@microsoft.com



Narjes Majdoub
Cloud Solution Architect
Data AI
nmajdoub@microsoft.com



Ali Bouhaddou
Cloud Solution Architect
Data Analytics
albouhad@microsoft.com



Frederic Gisbert
Cloud Solution Architect
Data Analytics
frgisber@microsoft.com

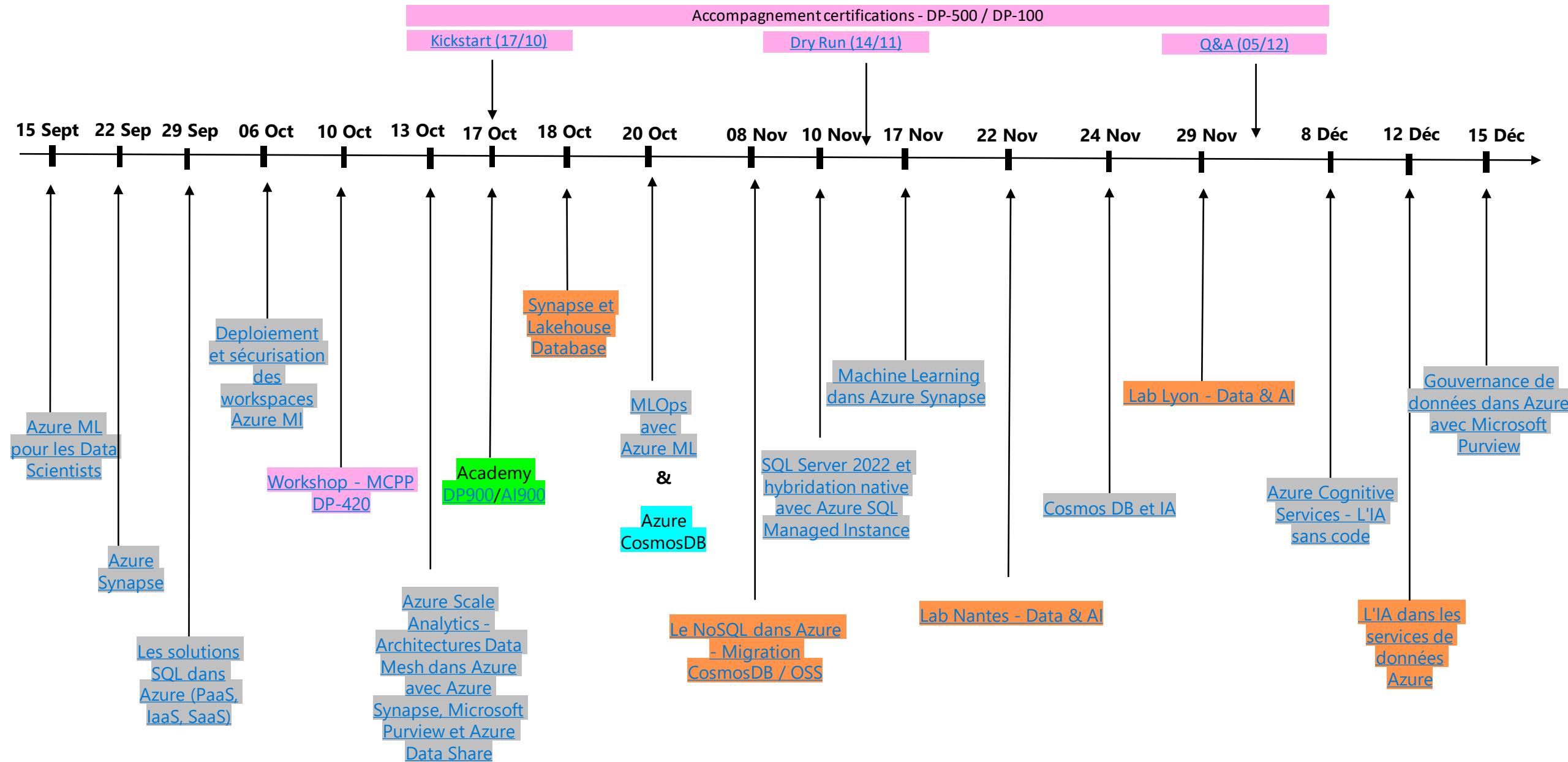


Azure Data & AI technical intensity plan

- From June 2022 to June 2023
- Focus on "Azure Data & AI" tech intensity
- Many content, from L100 Beginner to L400 Expert level:
 - Academy L100
 - Webinar L200/L300
 - Workshop L300/L400
 - Certification kickstart L300/L400
 - Openhack / Microhack L400

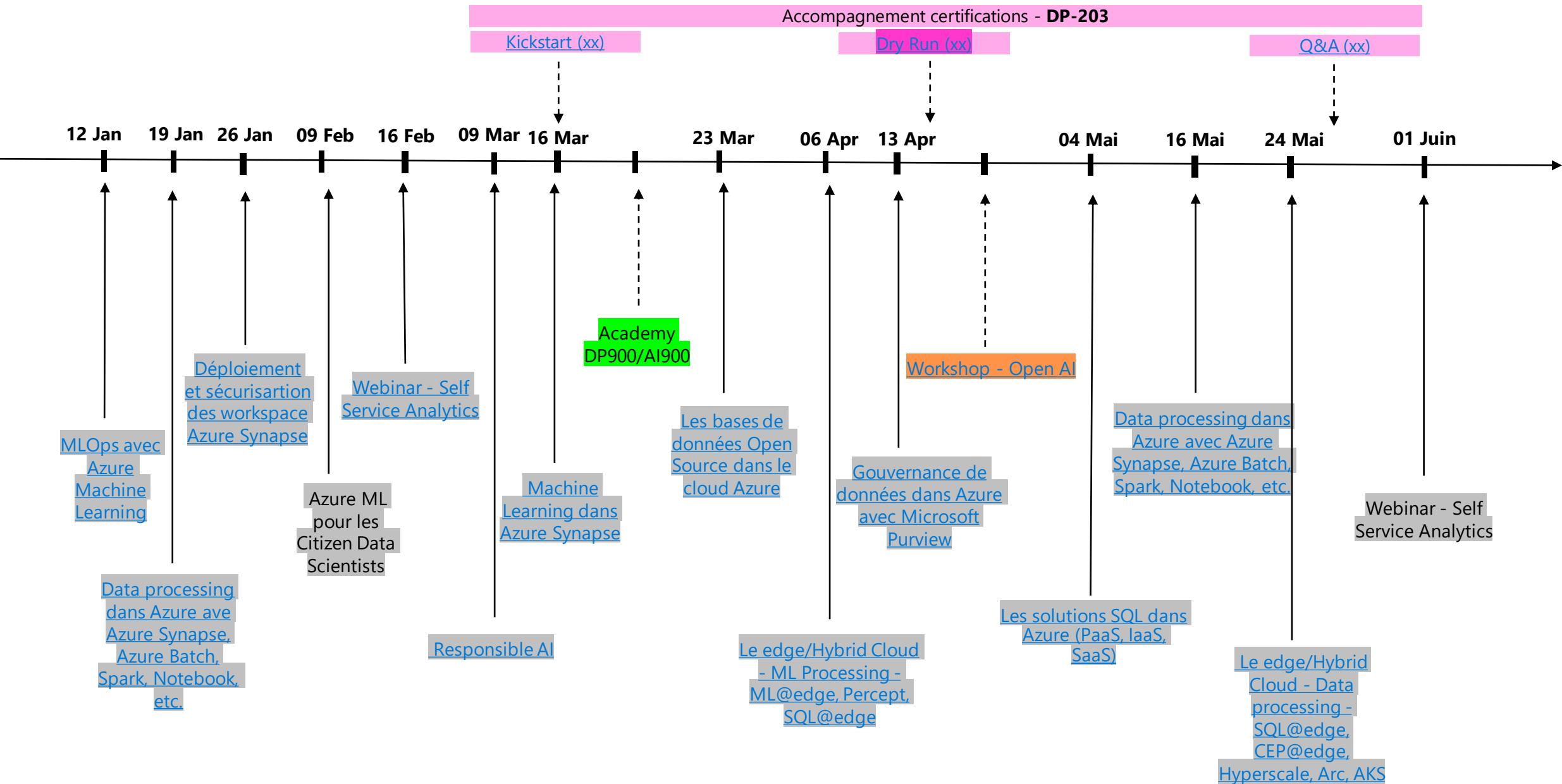
Data & AI events timeline – H1

Webinar/Academy - L 200/300
Workshop/Openhack/Certifications - L 300/400



Data & AI events timeline – H2

Webinar/Academy - L 200/300
Workshop/Openhack/Certifications - L 300/400



Liste des évènements de type Webinar 2H

Event Webinar (Les jeudis de la Data & AI) - L200/300	Date	Duration (min)	Link
Azure Machine Learning pour les Data Scientists	15/09/2022	120	https://msevents.microsoft.com/event?id=2454281594
Azure Synapse	22/09/2022	120	https://msevents.microsoft.com/event?id=857781749
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	29/09/2022	120	https://msevents.microsoft.com/event?id=502366997
Déploiement et sécurisation des workspaces Azure Machine learning	06/10/2022	120	https://msevents.microsoft.com/event?id=1505714138
Azure Scale Analytics - Architectures Data Mesh dans Azure avec Azure Synapse, Microsoft Purview et Azure Data Share	13/10/2022	120	https://msevents.microsoft.com/event?id=139685175
MLOps avec Azure Machine Learning	20/10/2022	120	https://msevents.microsoft.com/event?id=1245885767
SQL Server 2022 et hybridation native avec Azure SQL Managed Instance	10/11/2022	120	https://msevents.microsoft.com/event?id=145826476
Machine Learning dans Azure Synapse Analytics	17/11/2022	120	https://msevents.microsoft.com/event?id=3637723312
Azure Cosmos DB et IA	24/11/2022	120	https://msevents.microsoft.com/event?id=2646013445
Azure et les Services Cognitifs	08/12/2022	120	https://msevents.microsoft.com/event?id=3772037220
La gouvernance de données dans Azure avec Microsoft Purview	15/12/2022	120	https://msevents.microsoft.com/event?id=1499560981
MLOps avec Azure Machine Learning	12/01/2023	120	https://msevents.microsoft.com/event?id=4115194515
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	19/01/2023	120	https://msevents.microsoft.com/event?id=1537241181
Déploiement et sécurisation des workspace Azure Synapse	26/01/2023	120	https://msevents.microsoft.com/event?id=1806467748
Azure Machine Learning pour les Citizen Data Scientists	09/02/2023	120	En cours
PowerBI - Self Service Analytics	16/02/2023	120	https://msevents.microsoft.com/event?id=1401519679
L'IA responsable avec Azure machine learning	09/03/2023	120	https://msevents.microsoft.com/event?id=2072953112
Machine Learning dans Azure Synapse Analytics	16/03/2023	120	https://msevents.microsoft.com/event?id=3413014857
Les bases de données Open Source dans le cloud Azure	23/03/2023	120	https://msevents.microsoft.com/event?id=2727487131
Hybridation des services de Machine Learning Azure	06/04/2023	120	https://msevents.microsoft.com/event?id=1624914222
La gouvernance de données dans Azure avec Microsoft Purview	13/04/2023	120	https://msevents.microsoft.com/event?id=3909342839
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	04/05/2023	120	https://msevents.microsoft.com/event?id=1162207895
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	16/05/2023	120	https://msevents.microsoft.com/event?id=3517068442
Hybridation des services de données Azure	24/05/2023	120	https://msevents.microsoft.com/event?id=2996507398
Self Service Analytics	01/06/2023	120	En cours

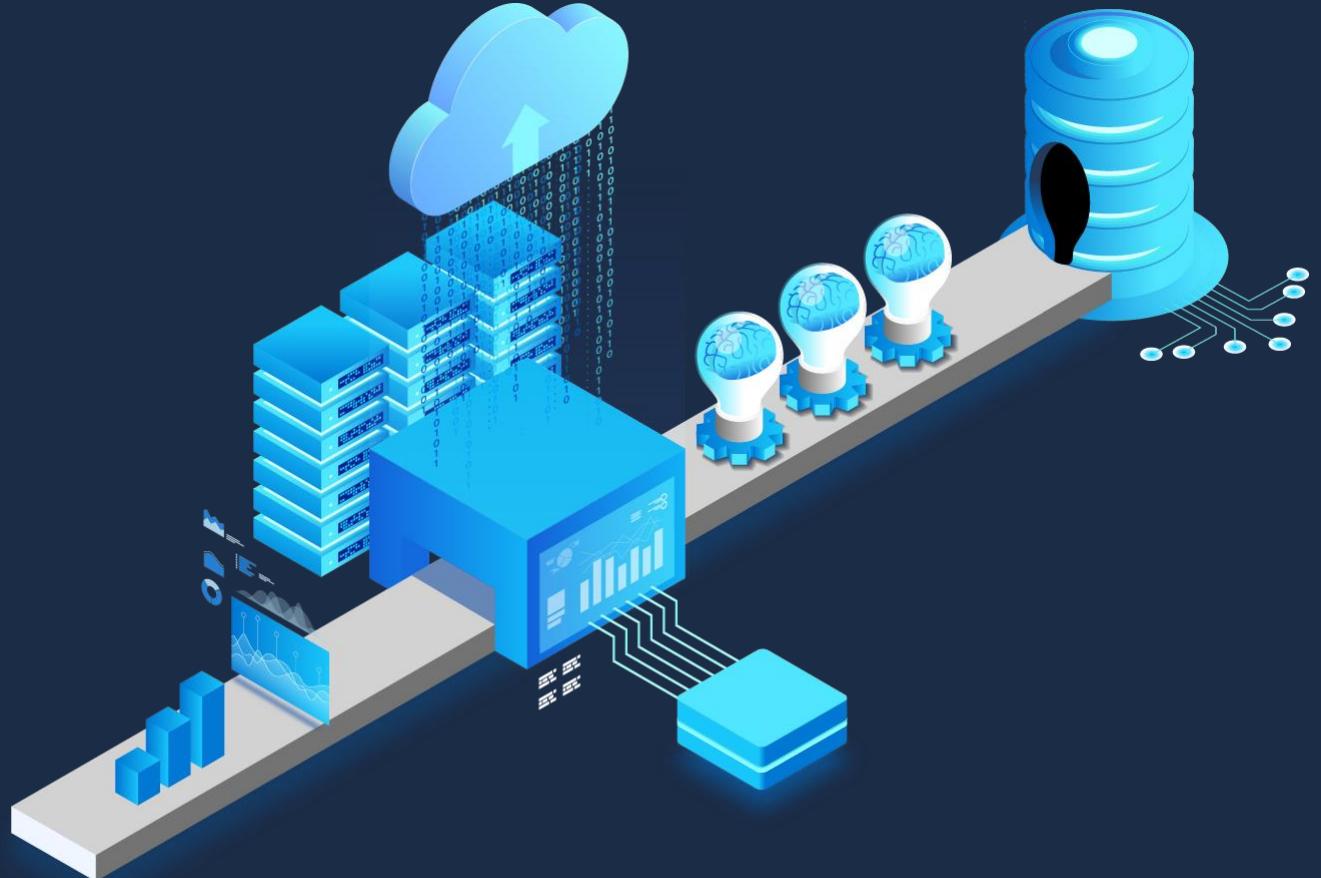
Liste des évènements de type Workshop/Prepa Cert/Academy

Event Workshop L300/400	Date	Duration (min)	Link
Synapse et Lakehouse Database	18/10/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cy9kwWS4u
Le NoSQL dans Azure - Migration CosmosDB / OSS	08/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cy9kwWS4u
Lab Lyon - Data & AI	22/11/2022	240	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUMIZZOURETORSWjcyTERYRkJGTTFFUjaUi4u
Lab Nantes - Data & AI	29/11/2022	240	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUMIZZOURETORSWjcyTERYRkJGTTFFUjaUi4u
L'IA dans les services de données Azure	12/12/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cy9kwWS4u
Open AI	H2	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFYSDVLR0cy9kwWS4u

Event Academy, kickstart certifications, workshop certifications	Date	Duration (min)	Link
MCPP - DP-420	10/10/2022	420	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUMkJSIRKSU1RRFA0OVgzSFdTSTY0E9WQy4u
Micro Hack CosmosDB	20/10/2022	420	H1 - Inscriptions PTA
Academy DP900	17-21/10/2022	300	https://msevents.microsoft.com/event?id=3250818161
Academy AI900	17-21/10/2022	300	https://msevents.microsoft.com/event?id=2717528090
Kickstart DP-500	17/10/2022	60	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNEk3WFQ1TEdNNTQ2Uk85V0cxQzM3E9ZRS4u
Dry Run DP-500	14/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNEk3WFQ1TEdNNTQ2Uk85V0cxQzM3E9ZRS4u
Q&A DP-500	05/12/2022	90	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNEk3WFQ1TEdNNTQ2Uk85V0cxQzM3E9ZRS4u
Kickstart DP-100	17/10/2022	60	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNDAxV0hSN0FHM1YzUzI3OUNMFYx\\$RIMi4u
Dry Run DP-100	14/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNDAxV0hSN0FHM1YzUzI3OUNMFYx\\$RIMi4u
Q&A DP-100	05/12/2022	90	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNDAxV0hSN0FHM1YzUzI3OUNMFYx\\$RIMi4u
Kickstart DP-203	17/10/2022	60	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUOVFWOUVCNFcyQk5SVjFBUFczNktCLFpLMi4u
Dry Run DP-203	14/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUOVFWOUVCNFcyQk5SVjFBUFczNktCLFpLMi4u
Q&A DP-203	05/12/2022	90	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUOVFWOUVCNFcyQk5SVjFBUFczNktCLFpLMi4u

MLOps with Azure ML

12/01/2023



Speaker info



Narjes Majdoub

Cloud Solutions Architect
Data & AI @ Microsoft

nmajdoub@microsoft.com



Nicolas Clerc

Cloud Solution Architect
Application Innovation & Dev

Nicolas.Clerc@microsoft.com
<https://github.com/flyingoverclouds>
<https://twitter.com/nclerc>

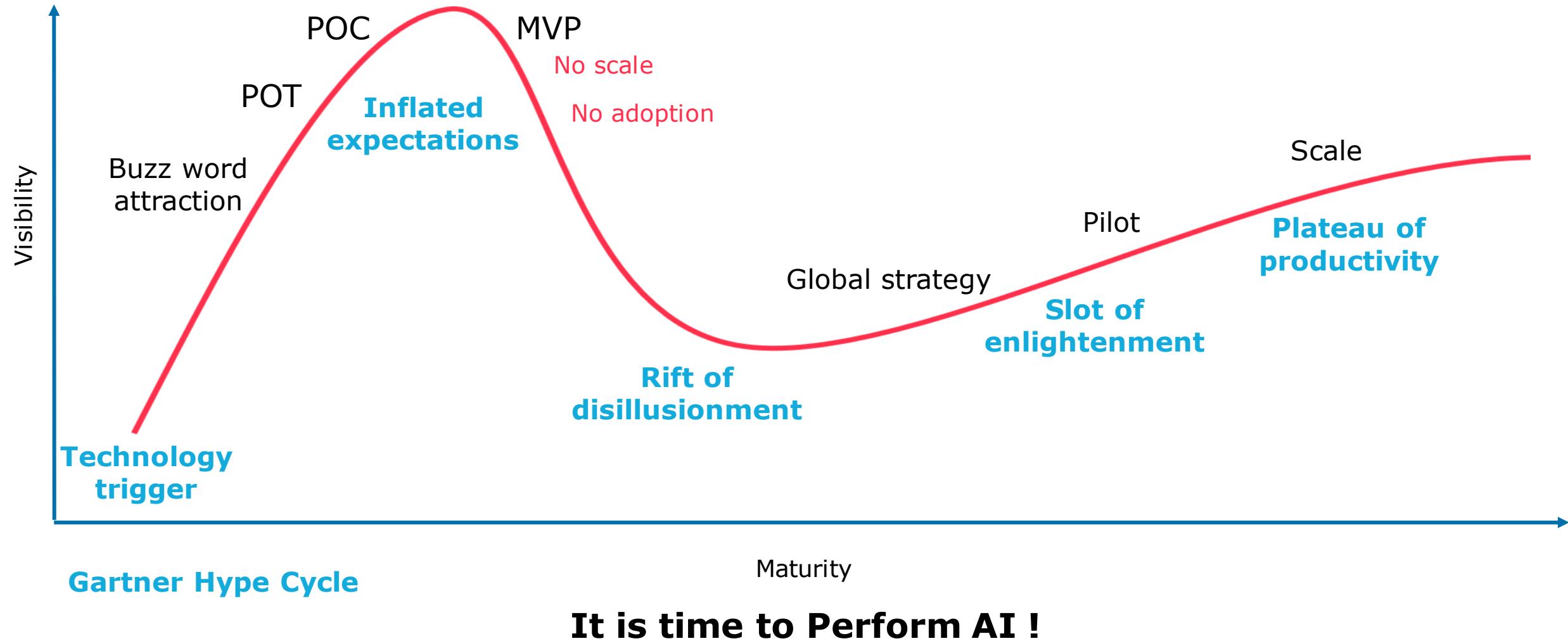
Agenda

-
1. Introduction to MLOps
 2. Build pipeline
 3. Release pipeline
 4. Monitoring and retraining

Machine Learning Challenges (origin)



AI adoption curve in a company



Barriers to AI at scale

More than 3 projects out of 4 don't make it to production

Cycle time is
too high

Manual

Model Decay

Training
Serving Skew

Lack of
reusability

Lack of
reproducibility

Model Bias

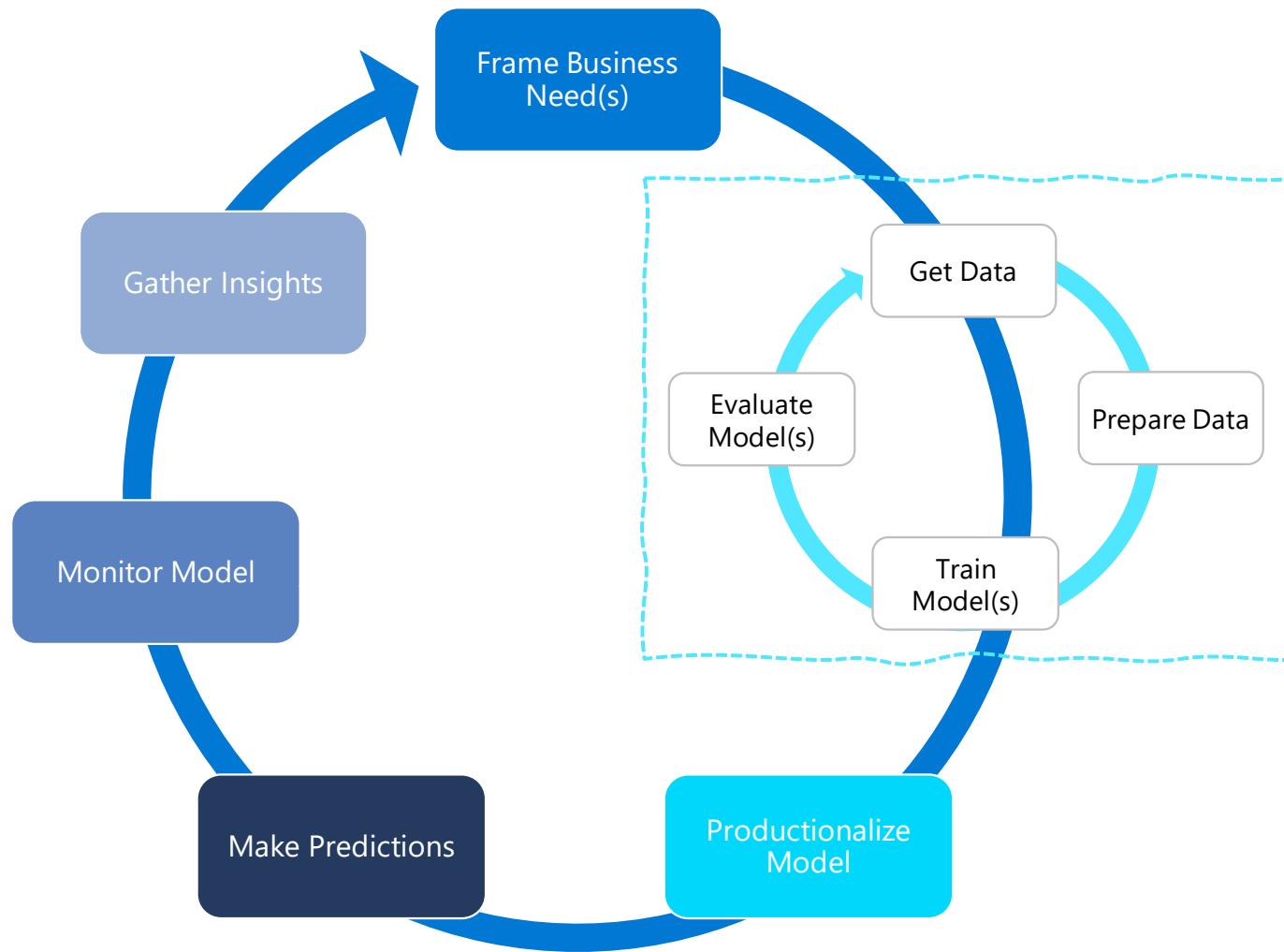
Model
Transparency

Duplicated
efforts

Limited
collaboration

Scalability

Classical Machine learning lifecycle



Most of time :

- Manual Experiment Steps
- Occurrence on fixed basis
- Decision to go into production based on a human reading
- Repetitive action
- Manual action to deploy (+ dedicated to the IT team)
- Lack of track & versioning over the time

Why is not so simple ?!



Communication & implementation
between roles & BU



Budget desynchronize with cost-benefit
analysis



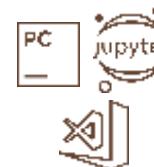
Realistic metrics (shared with technical
& business team)



Degree of scripts automation & code
refactoring



Triggering strategy to retrain the
model



Large Tools & template environment

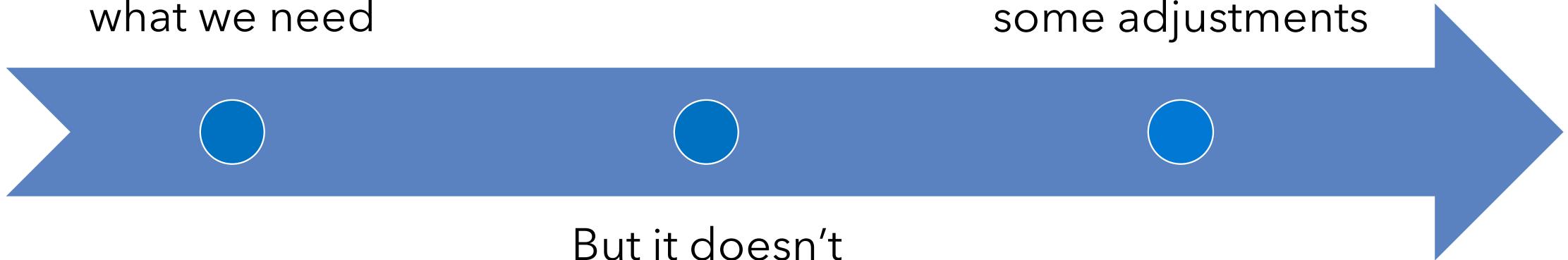
MLOps concepts



So why not just use DevOps?

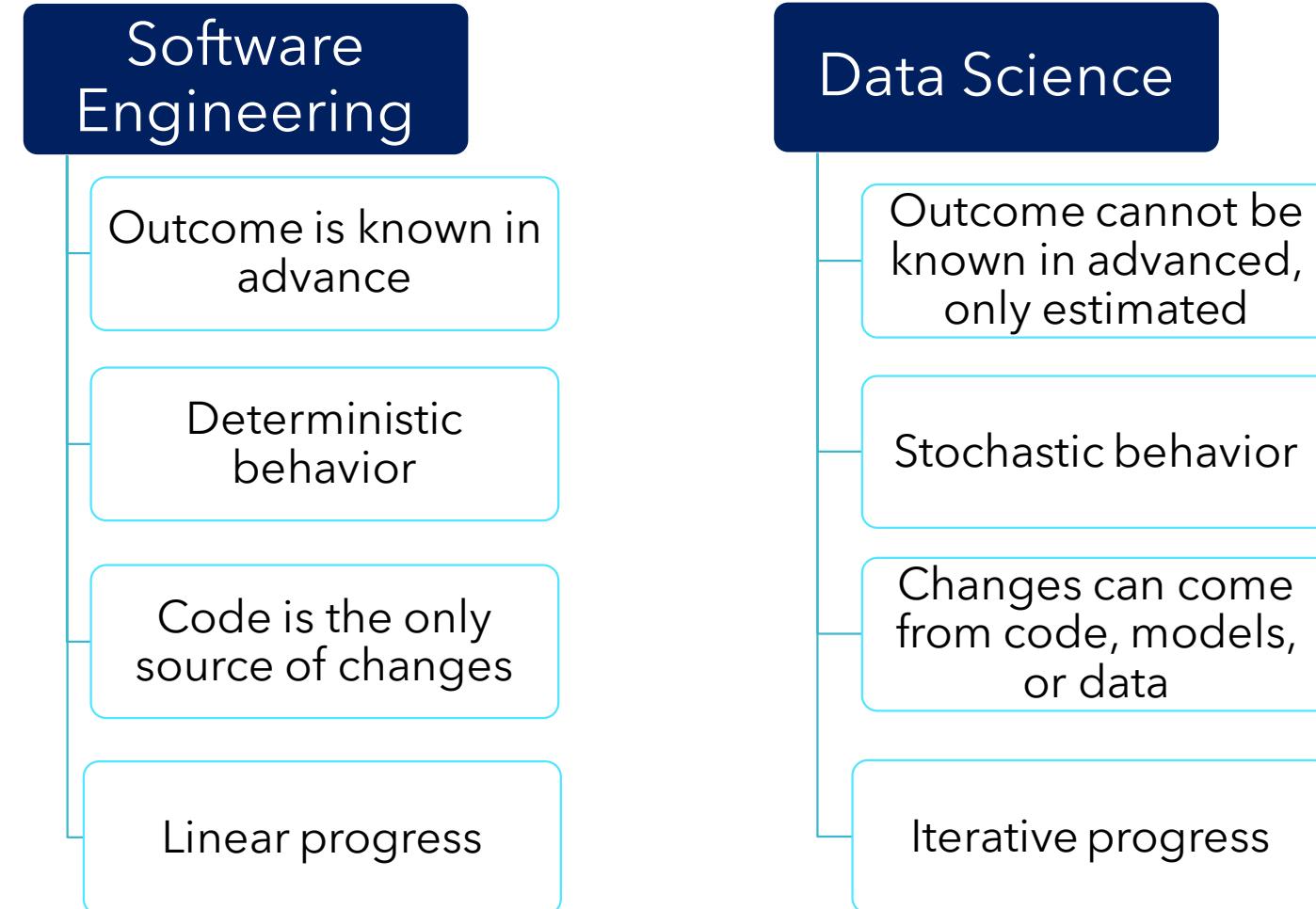
DevOps has many ingredients for what we need

We need to make some adjustments

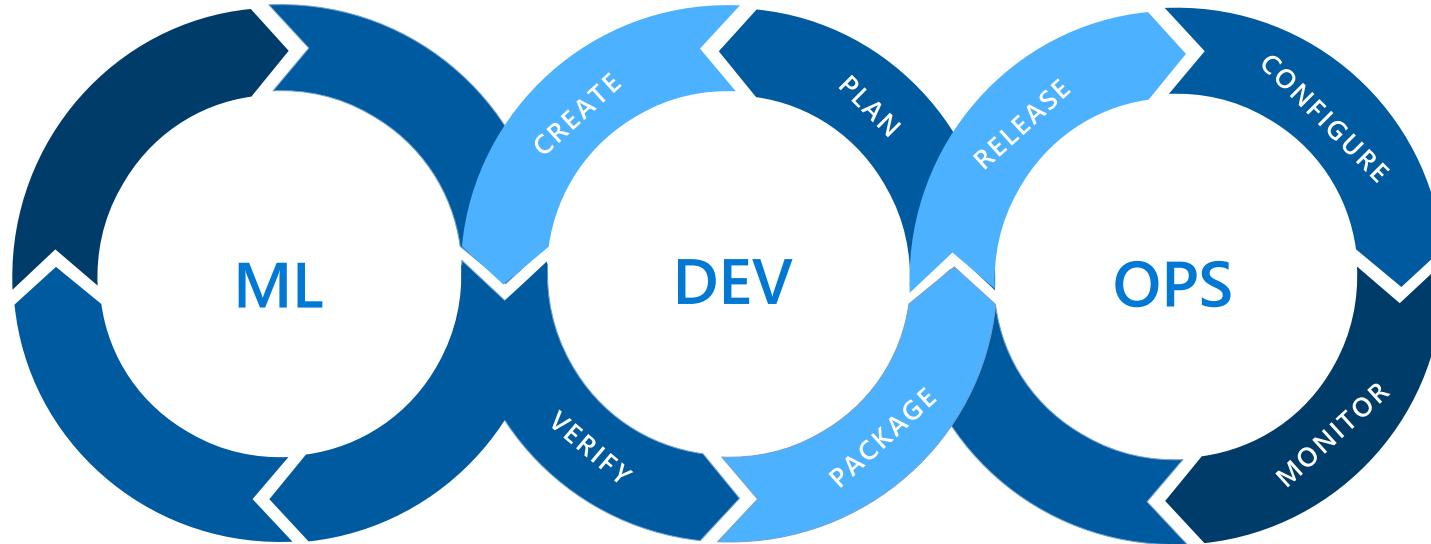


But it doesn't account for some key challenges in Data Science

Key Differences Between Software and Data Science



MLOps = ML + DEV + OPS



Experiment

Business Understanding
Data Acquisition
Exploratory data analysis
Modeling

Develop

Code Refactoring
Versioning
Testing & Validation

Operate

Model Deployment
CI/CD pipeline
Monitoring & Triggering



MLOps – Focus Areas

Challenge	How to solve – Best Practices
Reproducibility & versioning	Track, snapshot & manage assets used to create the model Enable collaboration and sharing of ML pipelines
Auditability & explainability	Maintain asset integrity & persist access control logs Certify model behavior meets regulatory & adversarial standards
Packaging & validation	Support model portability across a variety of platforms Certify model performance meets functional requirements
Deployment & monitoring	Release models with confidence Monitor & know when to retrain



What is Azure MLOps?

Asset management & services to manage and automate the ML lifecycle.

Asset Management



Environments



Code



Experiment Runs



Datasets



Models



Endpoints

Azure ML Services



ML Pipelines



Model Interpretability



Model Packaging



Model Monitoring



Model Profiling



ML Data Drift



Model Deployment



Eventing & Notifications



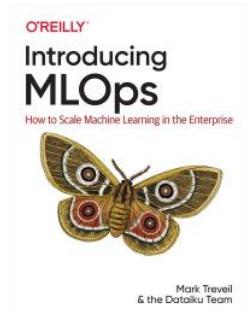
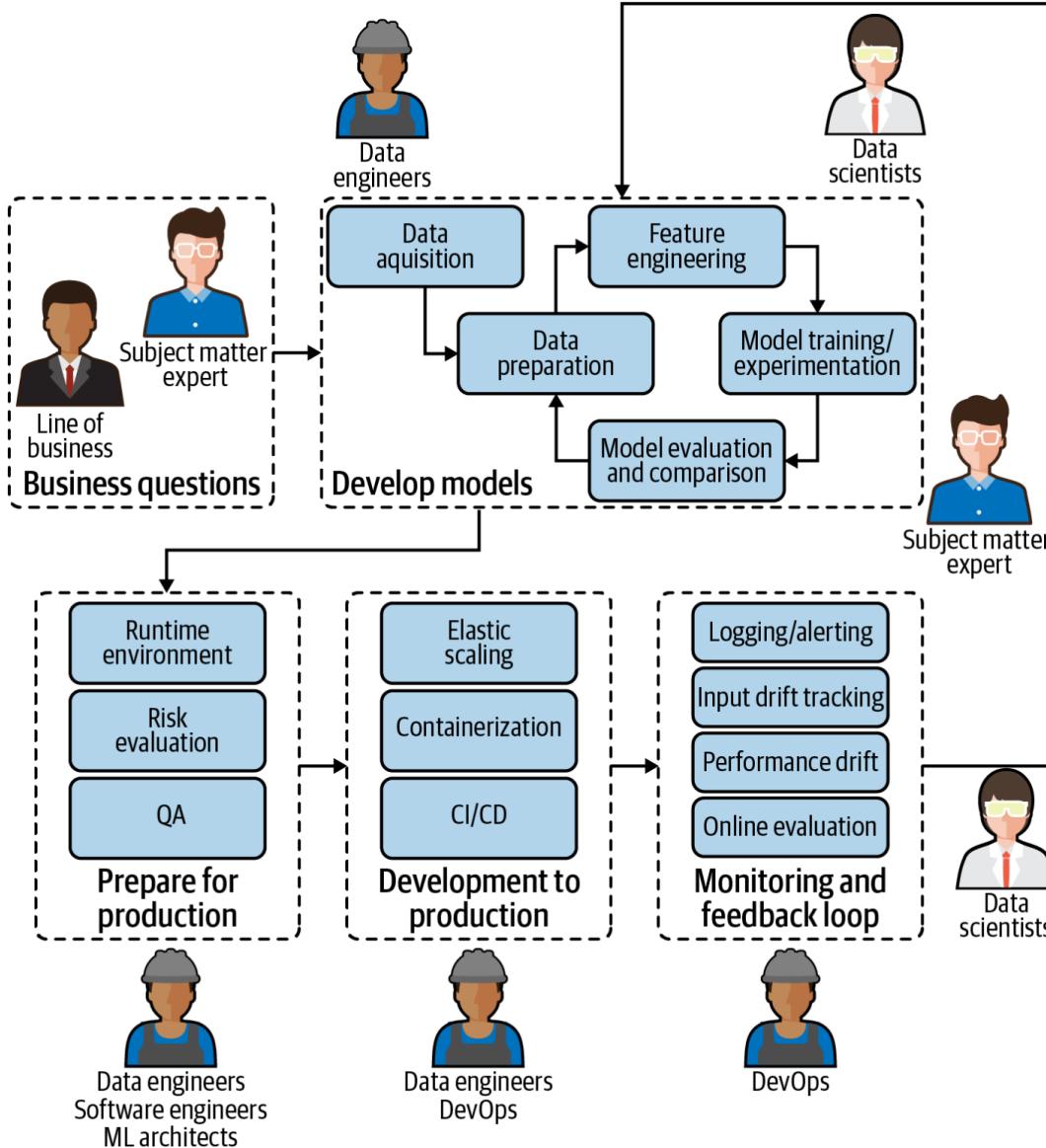
ML Audit Trail



GitHub / DevOps Extension

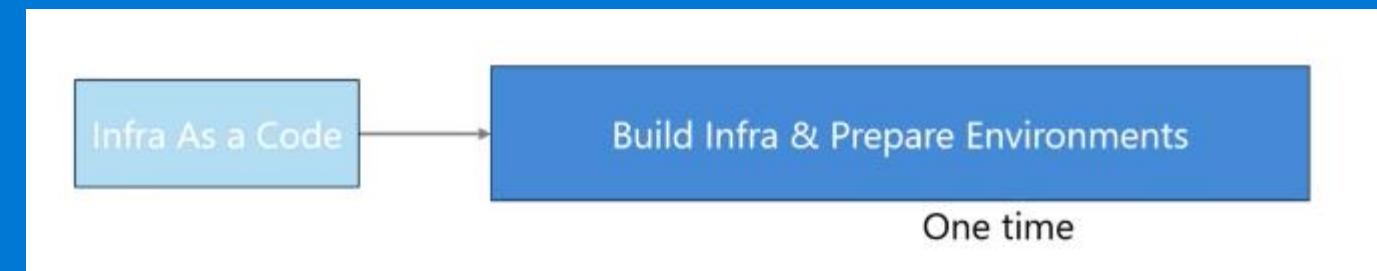


Roles & responsibilities



Demo

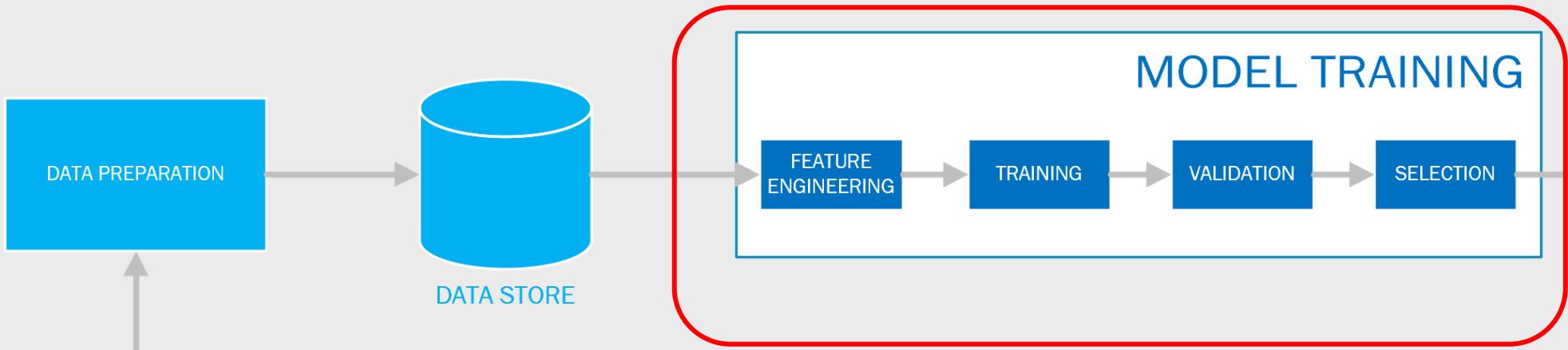
Create resources



Build/Modeling pipeline



MODELING PIPELINE



MODEL TRAINING

DATA PREPARATION

DATA STORE

FEATURE
ENGINEERING

TRAINING

VALIDATION

SELECTION

Data Scientist

RELEASE PIPELINE

MODEL DEPLOYMENT



DATA SOURCE

DEPLOY

APPROVE

MONITOR

VALIDATE

CREATE
CONTAINER
IMAGE

BUILD
ENVIRONMENT

PACKAGE

MODEL REGISTRY

MLOps Engineer



Azure ML – The Workspace

Top-level resource for the Azure Machine Learning service.

Centralized place to **work with all the artifacts** you create when using Azure Machine Learning service.

Models are registered with the workspace.

Each workspace can be shared by multiple people.

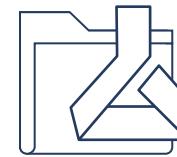
Required resources:

[Azure Container Registry](#)

[Azure Storage](#)

[Azure Application Insights](#)

[Azure Key Vault](#)



What's in the workspace?



Data



ML Pipelines



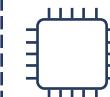
Jobs



Models



Endpoints



Computes



Datastores



Azure ML - Asset Versioning

Facilitated with Azure Machine Learning features:

Code Versioning

- Track changes in code (and configuration) over time, integrate work, reproducibility and collaboration.

Dataset Versioning

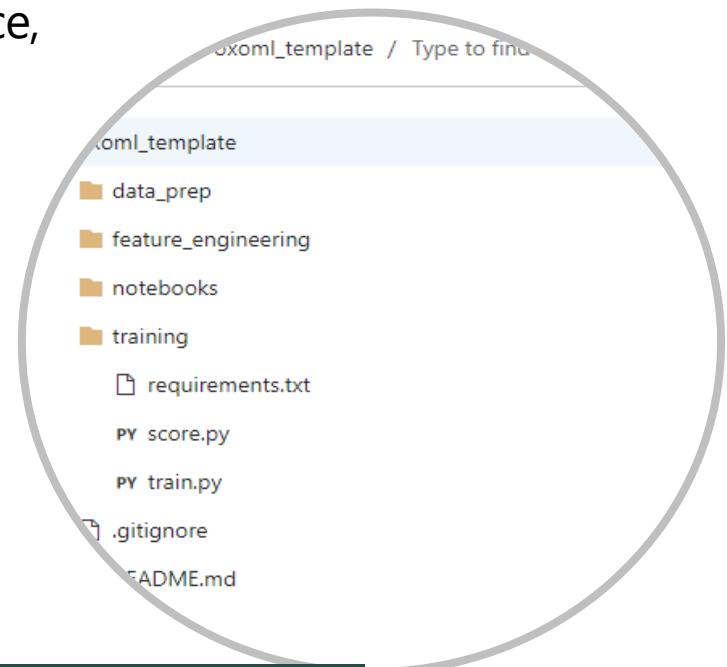
- Training data plays an important role in the quality of the software build. Hence, versioning of data is required for reproducibility.

Model Versioning

- Version trained models in relation to code and training data for traceability.

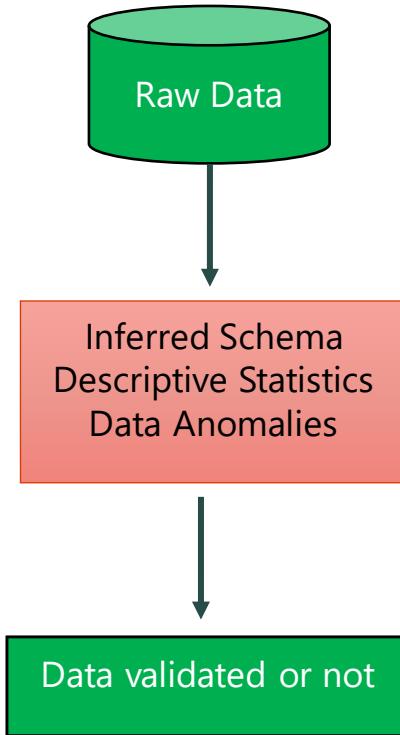
Experiment Tracking

- Version model experiment runs to understand which code, data and e.g. selected features led to what output and performance and allow for reproducibility.



→ Make it easy to revert to a previous version or know exactly where a bug is fixed when a problem

Data validation



```
# Compare stats of eval data with training data.  
tfdv.visualize_statistics(lhs_statistics=day2_stats, rhs_statistics=day1_stats,  
    lhs_name='DAY2', rhs_name='DAY1')
```

```
Sort by  
# Infer a schema from the training data stats.  
schema = tfdv.infer_schema(statistics=train_stats, infer_feature_shape=False)  
tfdv.display_schema(schema=schema)
```

```
tfdv.display_anomalies(anomalies)
```

Feature name	Anomaly short description		Anomaly long description
payment_type	Unexpected string values	Examples contain values missing from the schema: Pcard (<1%).	

```
tfdv.display_anomalies(serving_anomalies)
```

Feature name	Anomaly short description		Anomaly long description
	'tips'	Column dropped	Column is completely missing
'pickup_community_area'	INT	required	single
'payment_type'	STRING	required	single 'payment_type'
'trip_seconds'	FLOAT	optional	single
'trip_start_day'	INT	required	single

➔ Validate the data to ensure the result is aligned with the expected and decide to launch the remaining pipeline

Model Validation – Common Issues

Edge cases

- The **model response** on a given record is not the expected one.
- Investigate the trainset and detect **potential bias**.
- Ensure that the preprocessing is **not clipping any values** etc.
- Document these corner cases & add them to validation process

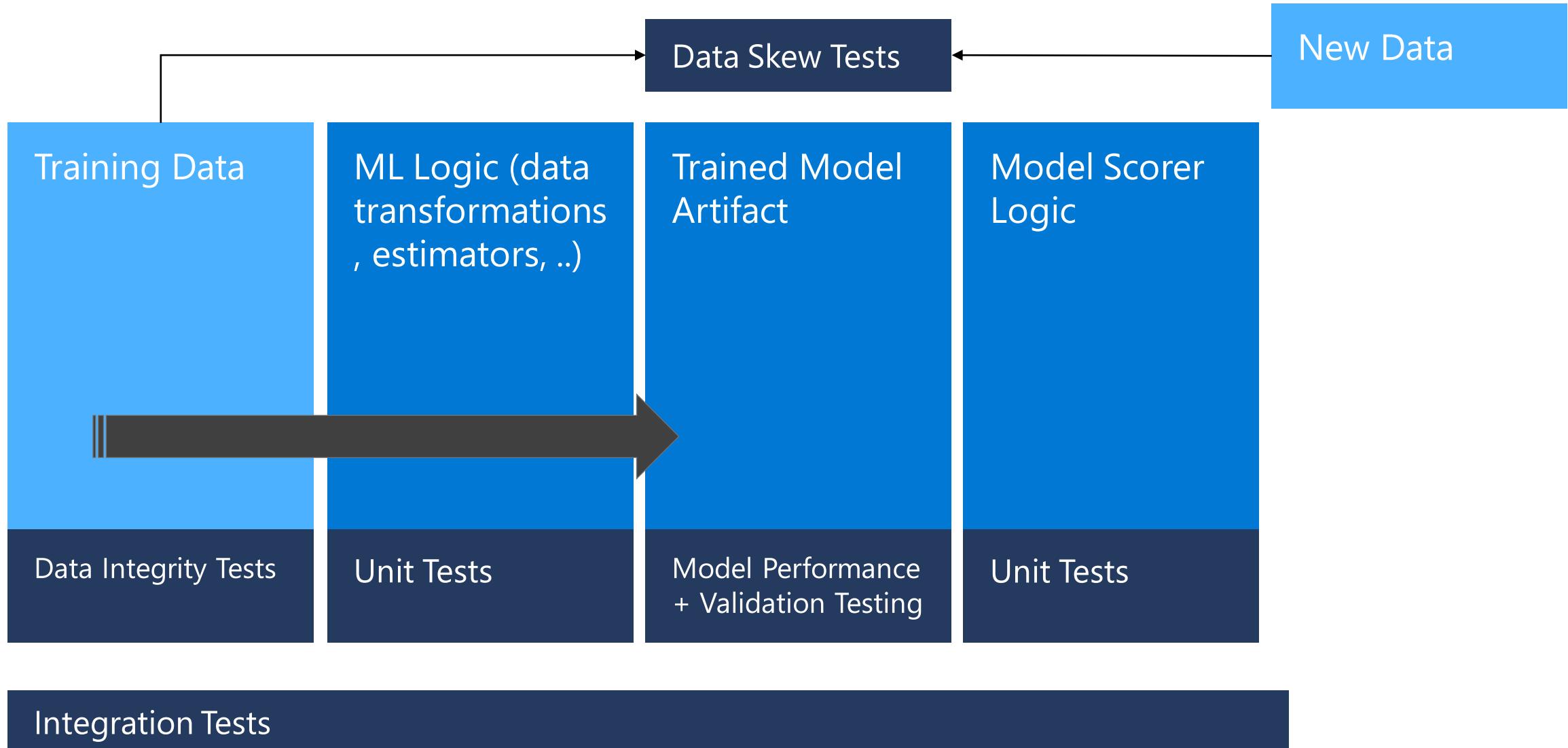
Null values / unknown categories

- This type of bugs refers to the resiliency of the model in case of missing **values and how well can it handle unseen categorical values**.

Input issues

- An input stream may stop producing **data causing unexpected responses** by the model.

ML Validation – Test Flow

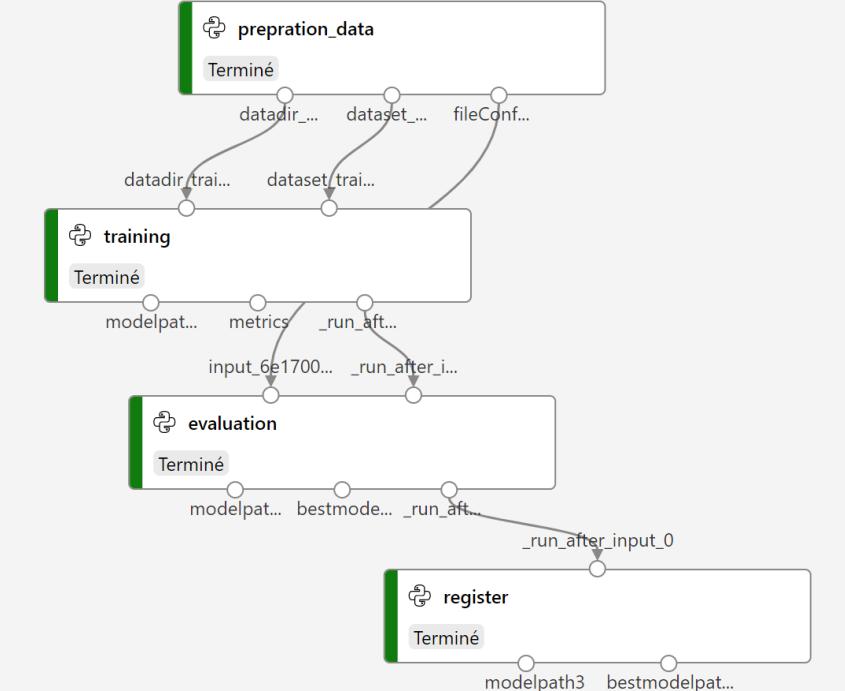


Reproducibility

“When the model is passed the same input data, the reproduced model would return the same output.”

Main points :

- Profile models + package requirements (AML environment)
- Optimize & scale (AML compute clusters)
- Central assets (model, code, data and documentation)
- Reproducible ML pipelines



➔ Reducing the randomness of machine learning to ensure reproducibility



Demo 1

Create Azure ML Pipeline for Model Build

Automate ML lifecycle – What pipeline to use?

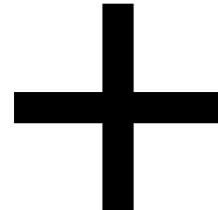
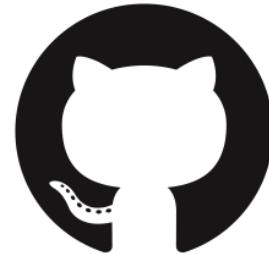
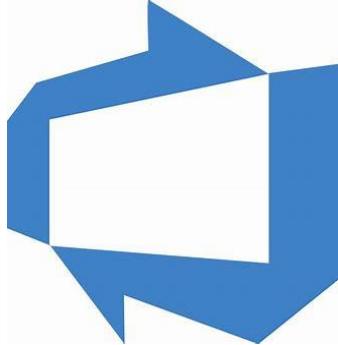
Scenario	Primary persona	Azure offering	OSS offering	Canonical pipe
Model orchestration (Machine learning)	Data scientist	Azure Machine Learning Pipelines	Kubeflow Pipelines	Data -> Model
Data orchestration (Data prep)	Data engineer	Azure Data Factory pipelines	Apache Airflow	Data -> Data
Code & app orchestration (CI/CD)	App Developer / Ops	Azure Pipelines / Github Actions	Jenkins	Code + Model -> App/Service

Branch Strategy

- ✓ **Feature Branch** – Data Scientists and ML Engineers work in their own branches and open pull requests to the integration branch
- ✓ **Integration Branch** – When a PR is opened to this branch, the CI pipeline will begin. If the run is successful, the merge will complete. If a higher performing model is produced, a PR to the main branch is open
- ✓ **Main Branch** – When a PR is opened to this branch, the CD pipeline will begin. Once the CD pipeline is successfully run, the merge will complete

GitHub / Azure DevOps and AzureML Better Together

Ideas, Collaboration, and Code



**Machine Learning Lifecycle
and MLOps**



Azure DevOps Pipelines

Cloud-hosted pipelines for Linux, Windows and macOS.



Any language, any platform, any cloud

Build, test, and deploy Node.js, Python, Java, PHP, Ruby, C/C++, .NET, Android, and iOS apps. Run in parallel on Linux, macOS, and Windows. Deploy to Azure, AWS, GCP or on-premises



Extensible

Explore and implement a wide range of community-built build, test, and deployment tasks, along with hundreds of extensions from Slack to SonarCloud. Support for YAML, reporting and more



Containers and Kubernetes

Easily build and push images to container registries like Docker Hub and Azure Container Registry. Deploy containers to individual hosts or Kubernetes.

The screenshot shows the Azure DevOps Pipelines interface. At the top, it displays the project path: Contoso / AdventureWorks Mobile / Pipelines / Builds / 10382. Below this, there's a sidebar with links: AdventureWorks, Overview, Pipelines (which is selected), Builds, Releases, Library, and Deployment groups. The main area shows a pipeline run titled "Enabling feature flags for Preview Attachment and Grid Views". It includes tabs for Summary, Logs (which is selected), Tests, and YAML. The summary shows three parallel jobs: "Windows Job" (Running, 1m 53s), "Linux Job" (Running, 3m 29s), and "macOS Job" (Running, 3m 07s). On the right, there's a detailed view of the "Linux Job" with a list of steps: Prepare job, Initialize job, Get sources, Cmdline, Nodetool, and Install dependencies. Below this is a log window showing command-line output:

```
yarn install v1.7.0
$ node build/npm/preinstall.js
[1/4] Resolving packages...
[2/4] Fetching packages...
[3/4] Linking dependencies...
[4/4] Building fresh packages...
$ npm run compile
#####
> code-oss-dev-build@1.0.0 compile ./adventureworks/build
> tsc -p tsconfig.build.json

⚡ Done in 4.89s.
$ node ./postinstall
[##] 2/2 removed './adventureworks/extensions/node_modules/typescript/lib/tsc.js'
removed './adventureworks/extensions/node_modules/typescript/lib/tsserverlibrary.d.ts'
removed './adventureworks/extensions/node_modules/typescript/lib/tsserverlibrary.js'
removed './adventureworks/extensions/node_modules/typescript/lib/typescriptServices.d.ts'
removed './adventureworks/extensions/node_modules/typescript/lib/typescriptServices.js'
```

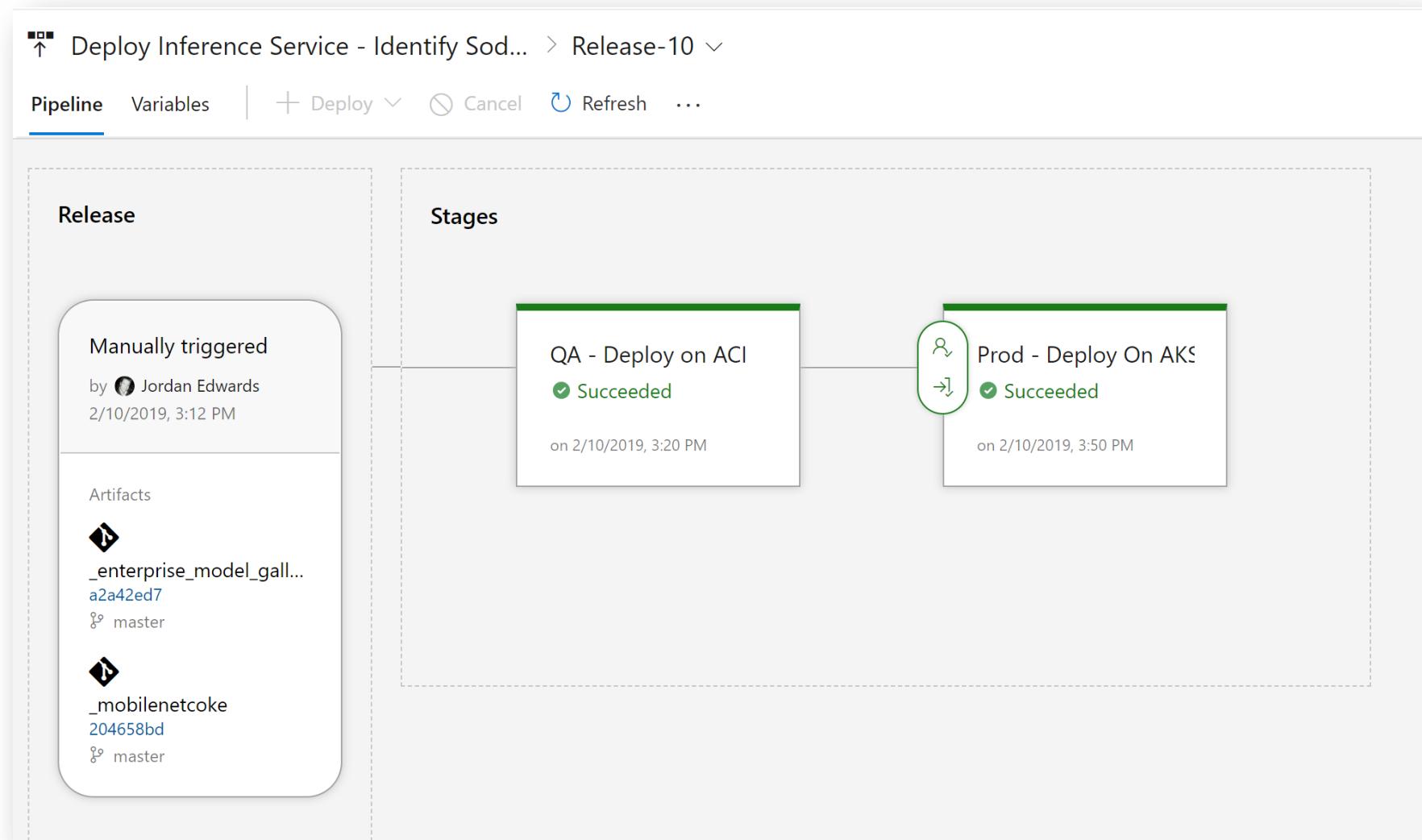
Azure DevOps + Azure ML

Trigger release pipelines
on model registration

Use **Azure DevOps +
Azure ML CLI** to manage
E2E release flow

Leverage Azure DevOps
approvals for controlling
application rollout

**Announcing: The AML
Extension for MLOps**



Azure Pipeline Machine Learning Extension

New MLOps capabilities in Azure Machine Learning brings the sophistication of DevOps to data science, with orchestration and management capabilities to enable effective ML Lifecycle management.

Key Features:

- Using Azure Resource Manager an AzureML Service Connection type can be created to access your artifacts in an AzureML workspace.
- By registering a new version of a model into an AzureML service workspace, a trigger can be configured to kick off a release pipeline.
- Azure DevOps server task for running published ML pipelines
- Azure DevOps tasks for AzureML specific actions such as model profiling and model deployment.

First Class Model Training Tasks

CI pipeline captures:

1. Create sandbox
2. Run unit tests and code quality checks
3. Attach to compute
4. Run training pipeline
5. Evaluate model
6. Register model

The screenshot shows the Azure DevOps interface for a pipeline named "DevOps for AI - Build Model - AzureML". The pipeline is currently set to "Tasks". The tasks listed are:

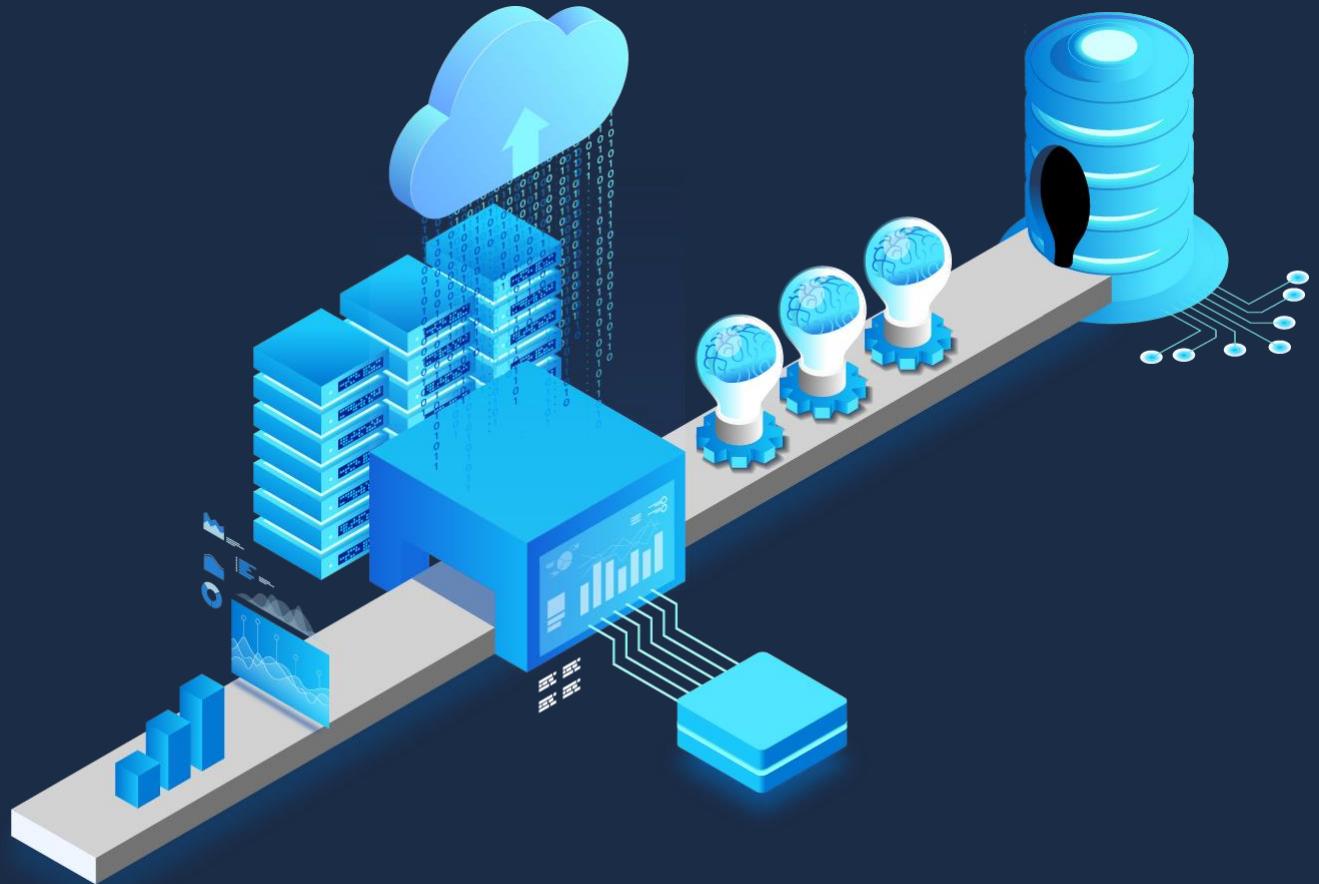
- Create Conda Environment (Conda Environment)
- Prepare Conda Environment (using yml) (Command Line)
- Install Azure CLI ML Extension (Preview) (Azure CLI)
- Unit tests (model code) (Command Line)
- Code Quality (flake8) (Command Line)
- Publish Unit Test Results (Publish Test Results)
- AzureML: Attach to an experiment (Azure CLI)
- AzureML: create dsvm run config (Azure CLI)
- AzureML: train model (Azure CLI)
- AzureML: download trained model (Azure CLI)
- Integration tests (Command Line)
- Publish Test Results (Publish Test Results)
- AzureML: register model (Azure CLI)
- Copy Model Requirements (Copy Files)

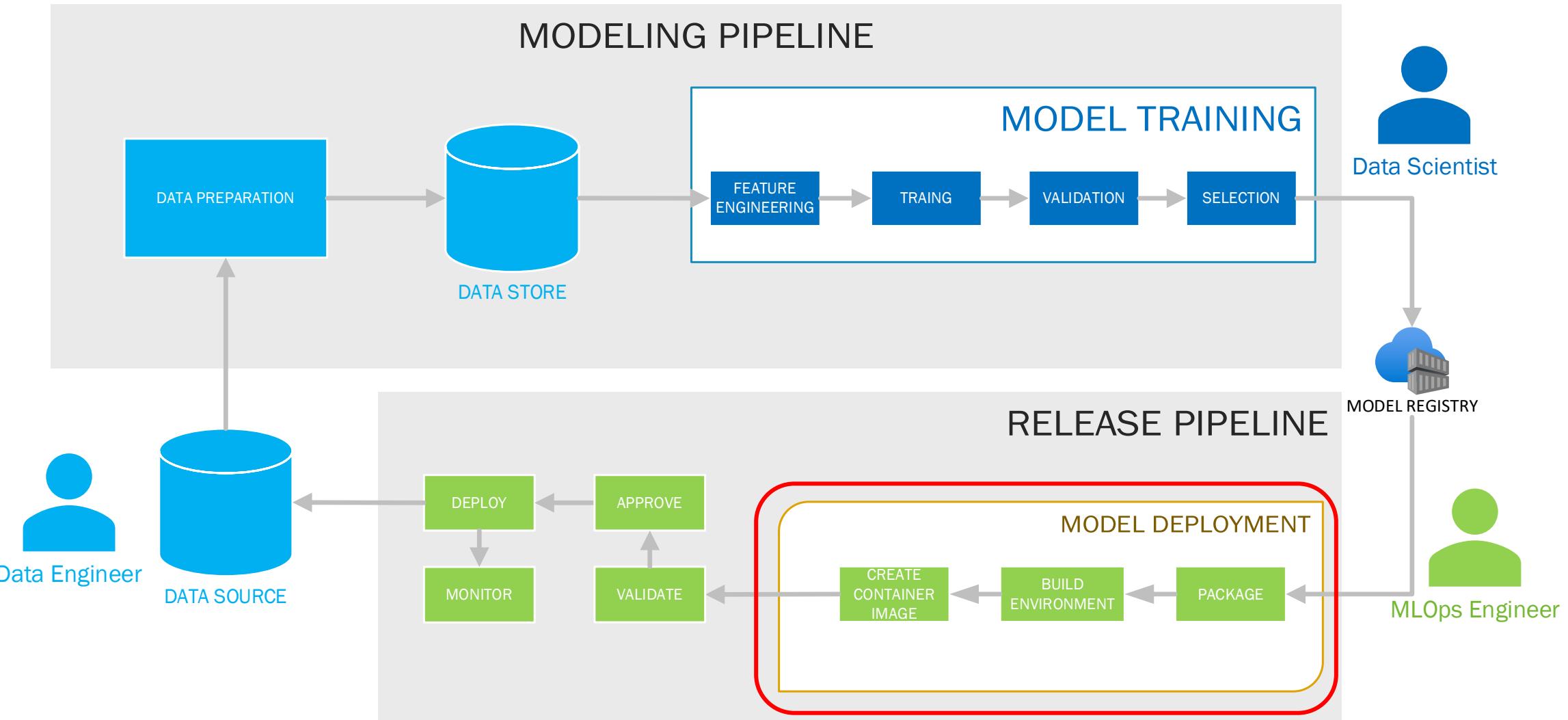


Demo 2

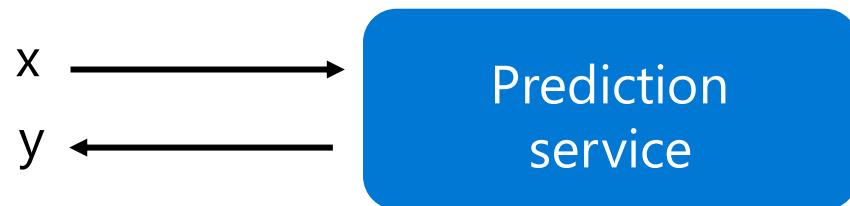
Build pipeline driven by Azure DevOps

Release pipeline





Deployment – Check list of questions



Realtime vs Batch



Cloud vs Edge/Browser



Compute resources (CPU/GPU/memory)



Latency, throughput (QPS)



Logging



Security and privacy

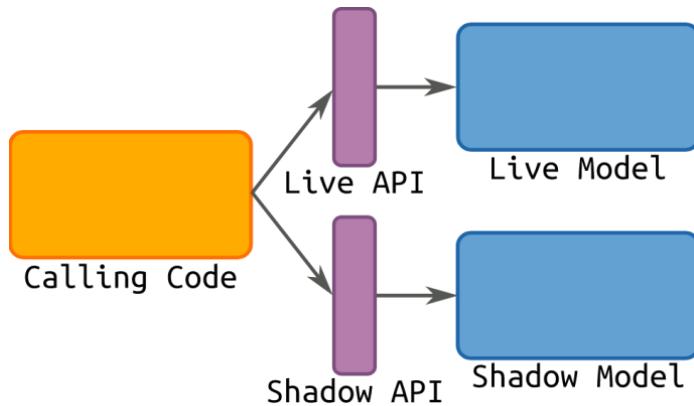
Deployment Patterns – Common use cases

1. New product/capability
2. Automate/assist with manual task
3. Replace previous ML system

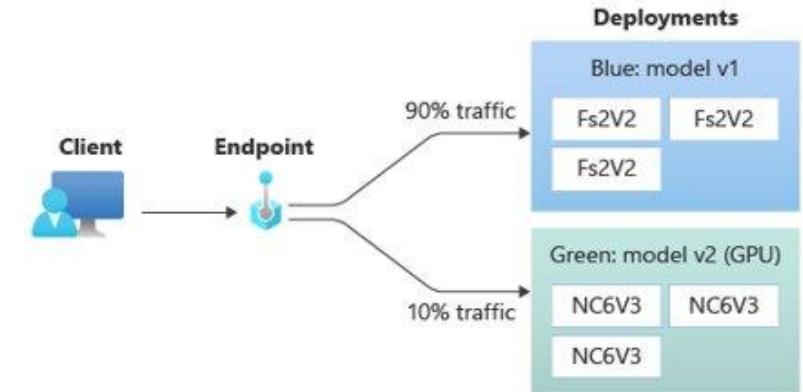
Key ideas:

- Gradual ramp up with monitoring
- Rollback

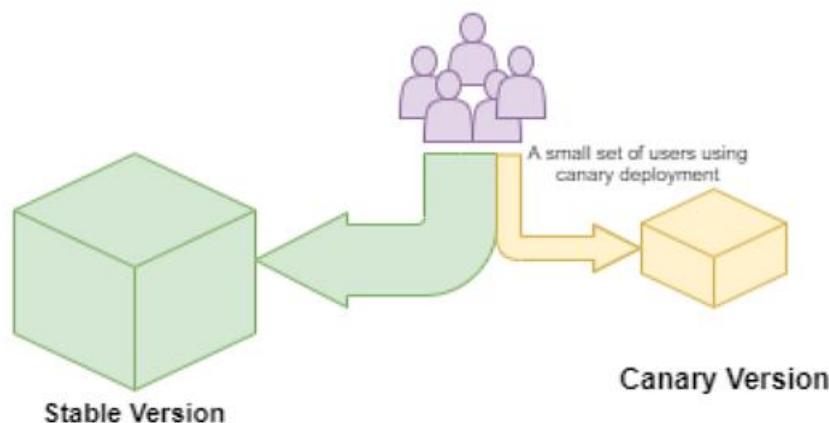
Some model deployment strategies



Shadow deployment

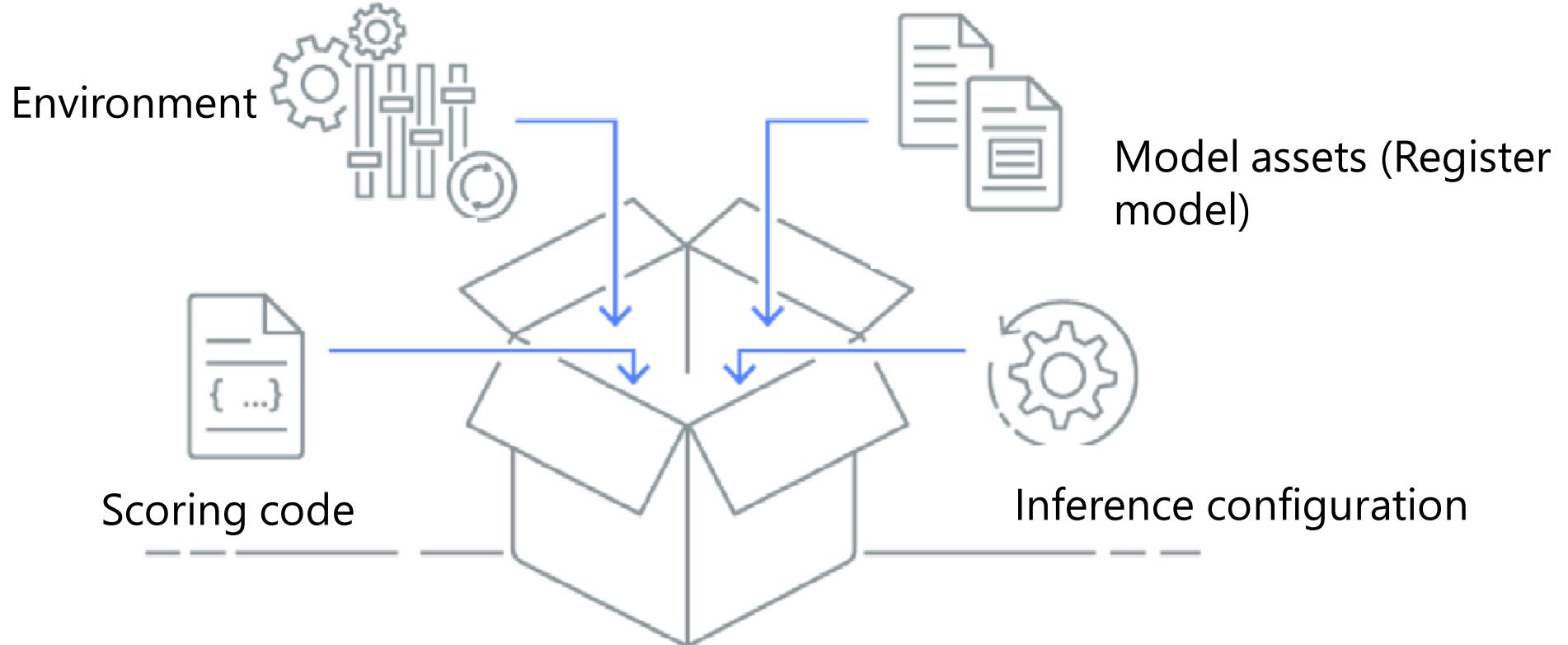


Blue-green deployment

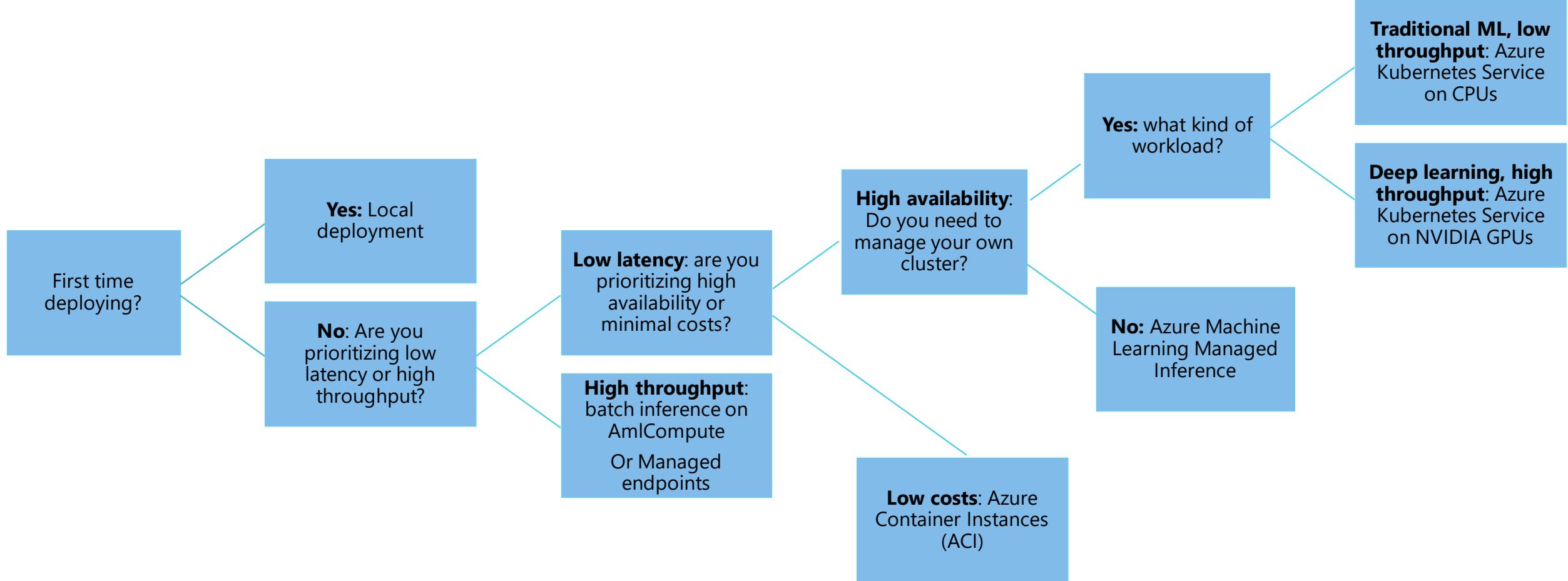


Canary deployment

AML model deployment- requirements



AML model deployment – inferencing target



Azure Pipeline - Release

Release pipelines in Azure Pipelines help your team **continuously deliver** models at a faster pace and with lower risk. You can **fully automate** the testing and delivery of your models in multiple stages all the way to production. Or, set up semi-automated processes with **approvals** and **on-demand deployments**.

Operations can be automated:

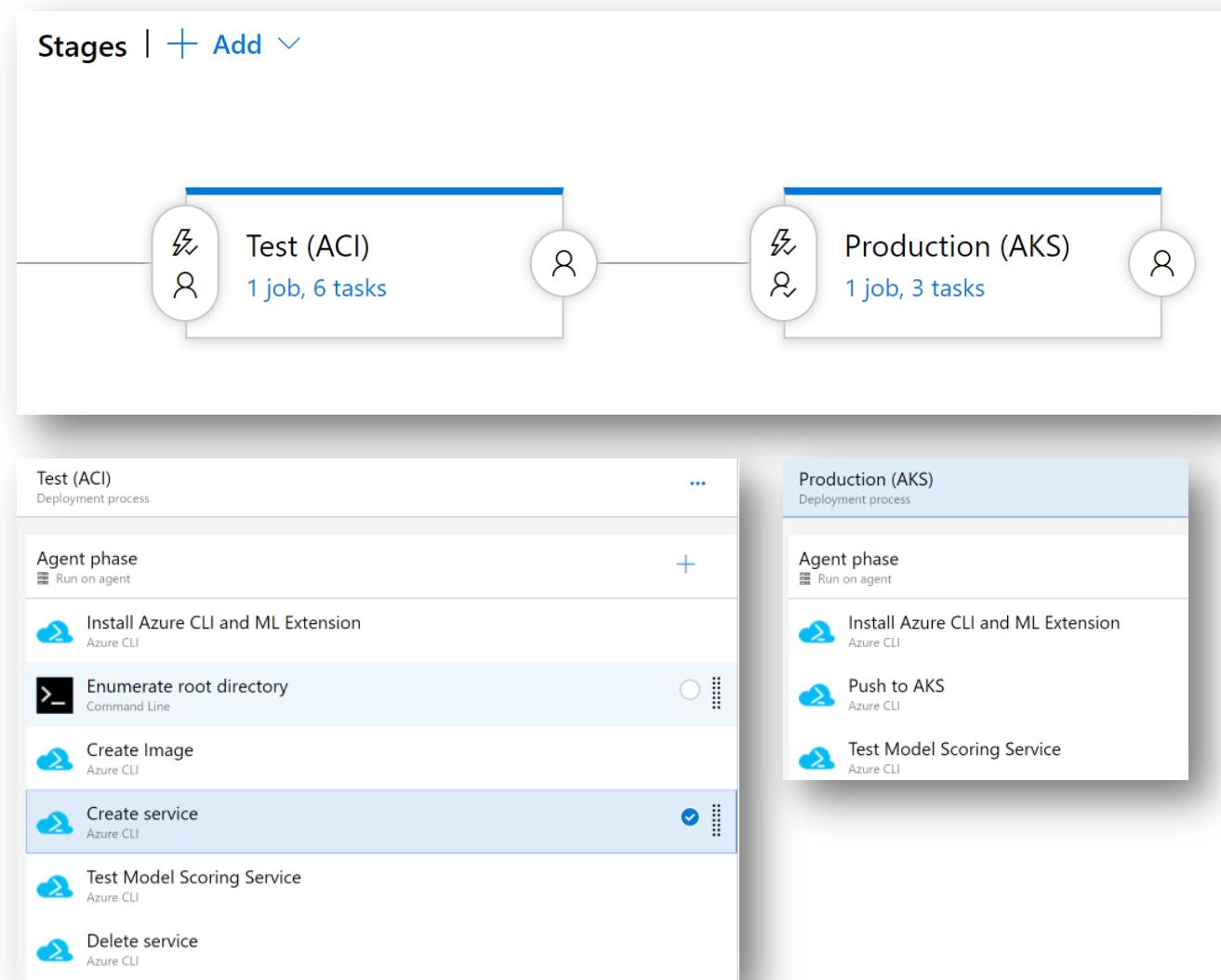
- Data preparation (extract, transform, load operations)
- Training machine learning models with on-demand scale-out and scale-up
- Deployment of machine learning models as public or private web services
- Monitoring deployed machine learning models (such as for performance or data-drift analysis)

Automated Deployment

CD pipeline captures:

1. Package model into container image
2. Validate and profile model
3. Deploy model to Dev/Test (ACI)
4. If all is well, proceed to rollout to AKS

Everything is done via the CLI





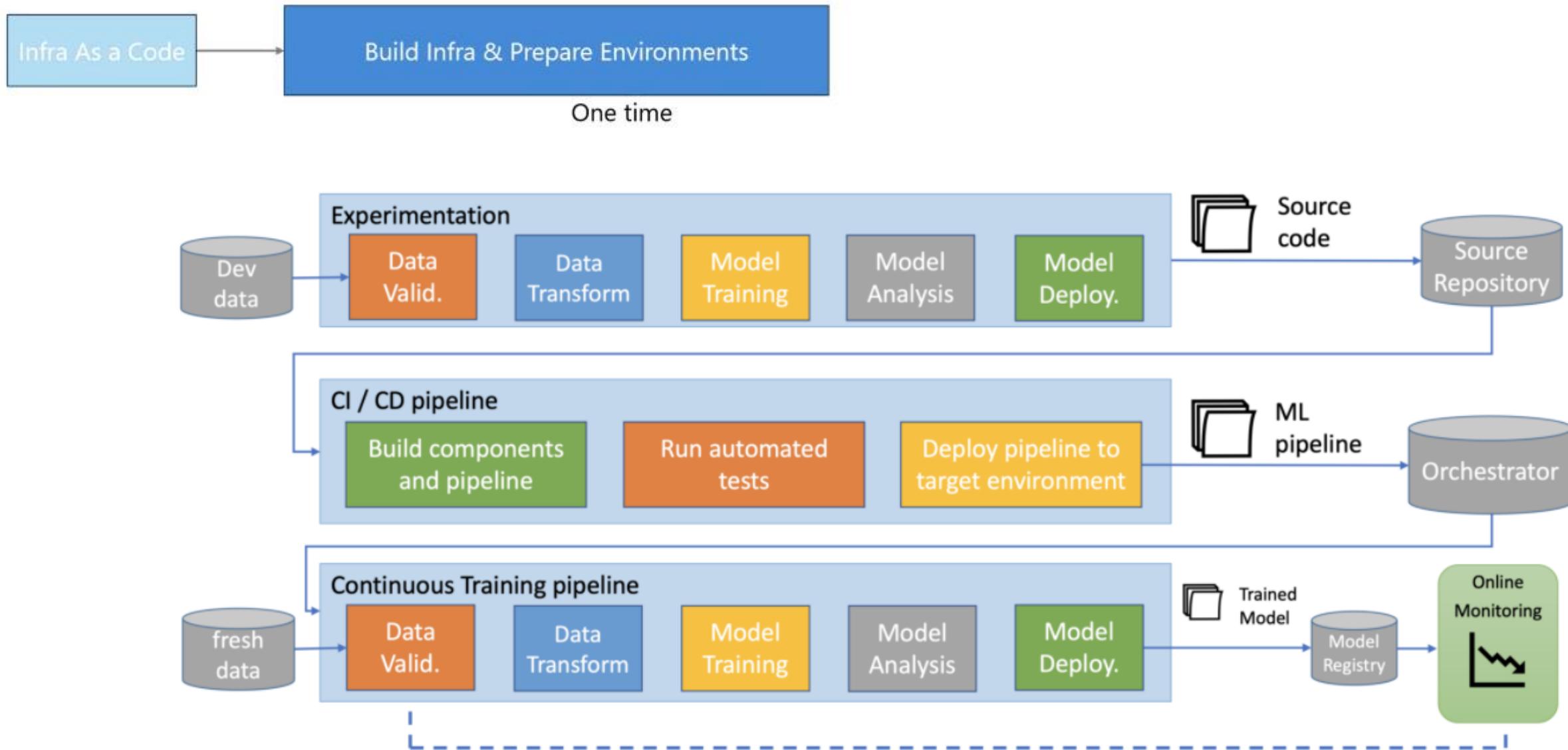
Demo

Create model release pipeline Azure DevOps

Monitoring/ Retraining



Complete MLOps chain



Monitoring ML systems – Why?

FORTUNE

This is not a drill: The coronavirus pandemic is testing A.I.'s ability to handle extreme events.

Model Drift

Tik Tok

We believe all companies should disclose their algorithms, moderation policies, and data flows to regulators

Model Transparency

BUSINESS INSIDER

Amazon built an AI tool to hire people but had to shut it down because it was discriminating against women

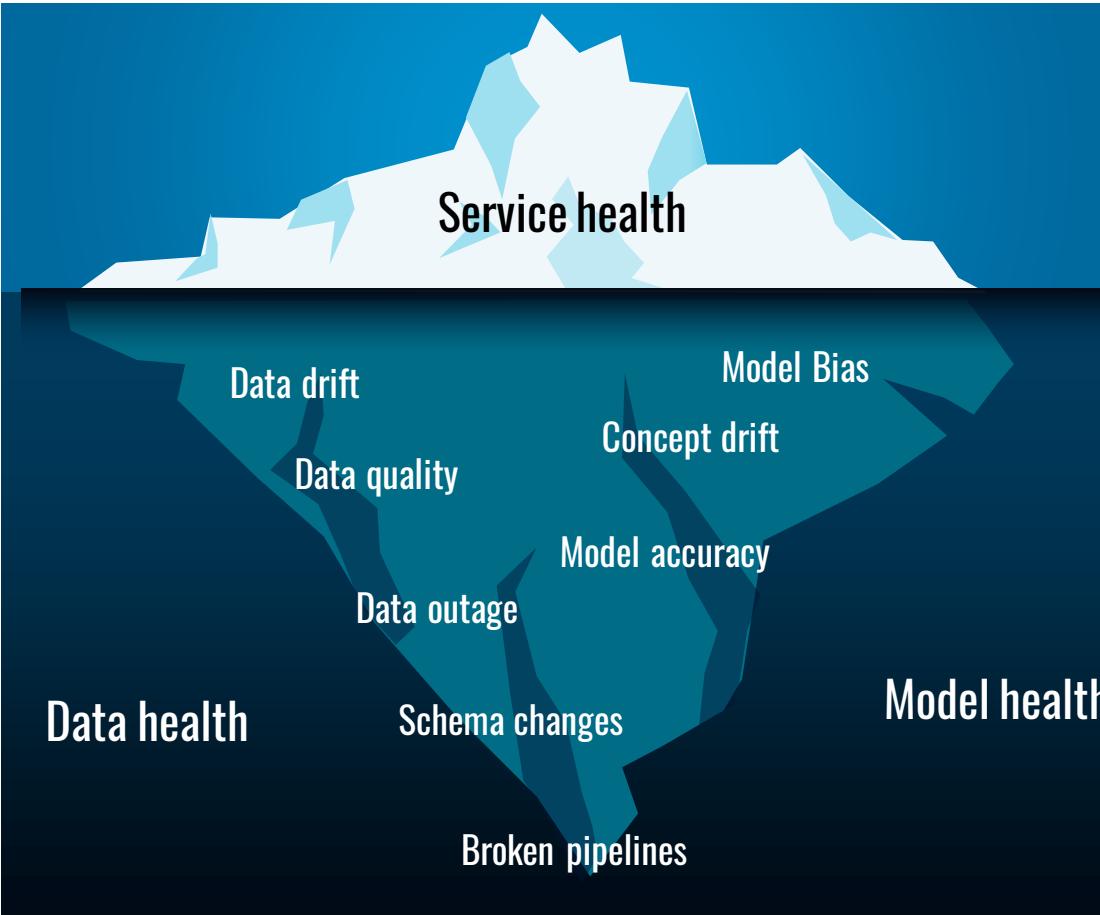
Model Bias

THE WASHINGTON POST

JPMorgan settles federal mortgage discrimination suit for \$55 million

Model Compliance

Monitoring ML systems – What?



- ✓ Brainstorm what could go wrong
- ✓ Brainstorm a few statistics that will detect the problem
- ✓ Define thresholds to trigger alerts

[source](#)

Monitor for Operational Issues – Application Insights

- Azure Application Insights
 - Output data
 - Responses
 - Request rates, response time, and failure rates
 - Exceptions

Microsoft Azure Machine Learning

my-ws > Endpoints > aks-service-appinsights

aks-service-appinsights

--

CPU
0.1

Memory
0.5 GB

Application Insights enabled
true

Application Insights url
<https://portal.azure.com/#@microsoft.onmicrosoft.com/resource/subscriptions/xxxxxx...>

Event Hubs enabled
false

Storage enabled
false

Autoscale enabled
true

Min replicas
1

New

Home

Author

Notebooks

Automated ML (preview)

Designer (preview)

Assets

Datasets

Experiments

Pipelines

Models

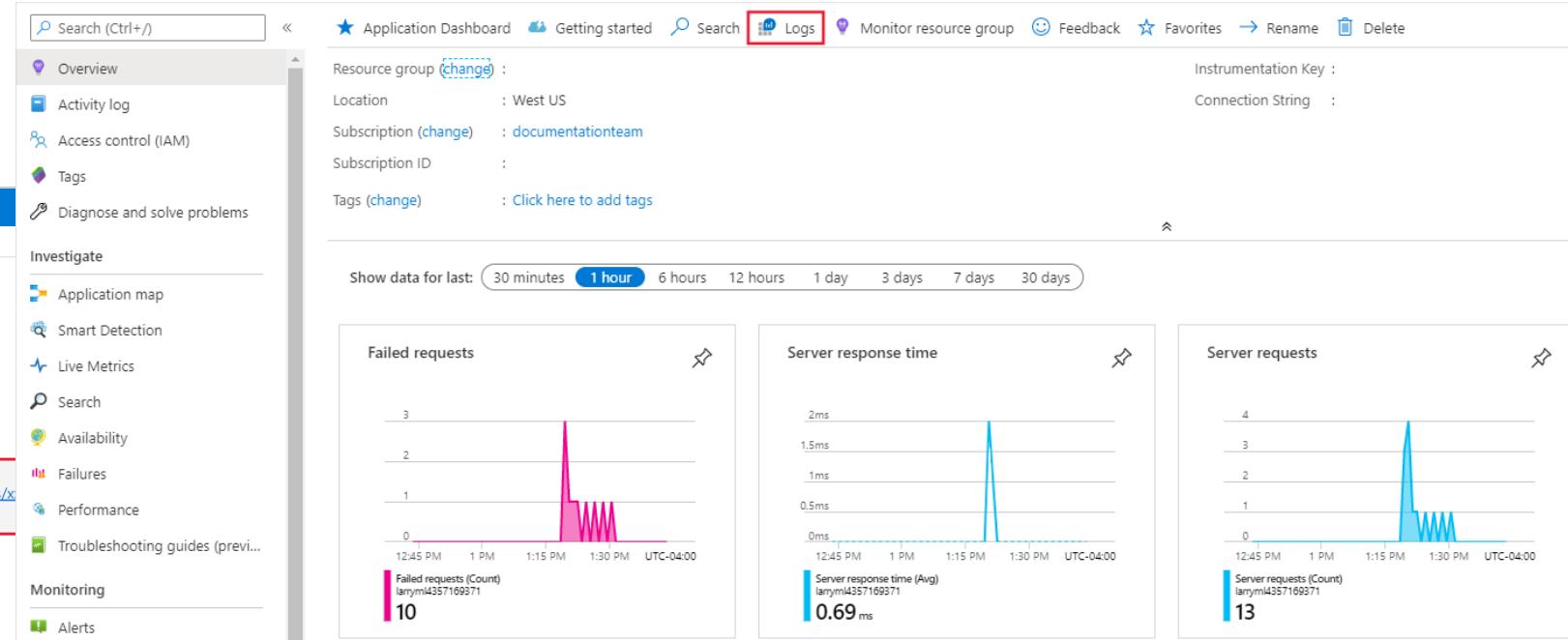
Endpoints

Manage

Compute

Datastores

Data Labeling



Monitor Azure Machine Learning – Azure monitor

ContosoRetailWebAppDb (consqlhqytc4j6j6uibc2/ContosoRetailWebAppDb) - Activity log

Activity log

Search (Ctrl+)

Overview

Activity log

Tags

Diagnose and solve problems

Quick start

Query editor (preview)

Settings

Configure

Geo-Replication

Connection strings

Sync to other databases

Add Azure Search

Properties

Locks

Export template

Security

Advanced Data Security

Auditing

Dynamic Data Masking

Transparent data encryption

Search (Ctrl+)

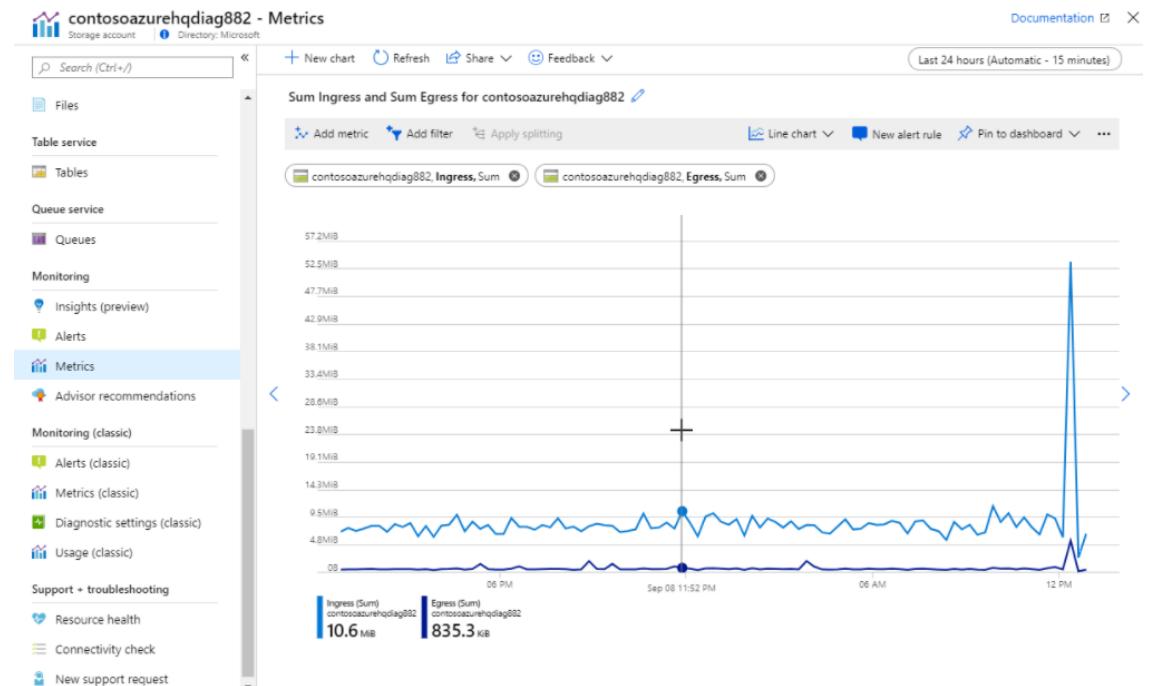
Quick Insights

Management Group : None Subscription : Contoso IT - demo Timespan : Last week Event severity : All

Resource group : ContosoAzureHQ Resource : ContosoRetailWebAppDb Add filter

13 items.

OPERATION NAME	STATUS	TIME	TIME STAMP	SUBSCRIPTION	EVENT INITIATED BY
DeployIfNotExists	Succeeded	3 h ago	Mon Sep 09...	Contoso IT - demo	Microsoft Azure Policy Insights
DeployIfNotExists	Succeeded	1 d ago	Sun Sep 08 ...	Contoso IT - demo	Microsoft Azure Policy Insights
DeployIfNotExists	Succeeded	2 d ago	Sat Sep 07 ...	Contoso IT - demo	Microsoft Azure Policy Insights
DeployIfNotExists	Succeeded	3 d ago	Fri Sep 06 ...	Contoso IT - demo	Microsoft Azure Policy Insights
Get Database Top Queries query	Succeeded	3 d ago	Fri Sep 06 ...	Contoso IT - demo	rosmithj@microsoft.com
Audit	Succeeded	4 d ago	Thu Sep 05 ...	Contoso IT - demo	Microsoft Azure Policy Insights
Audit	Succeeded	5 d ago	Wed Sep 04 ...	Contoso IT - demo	Microsoft Azure Policy Insights
Get Database Top Queries query	Succeeded	6 d ago	Tue Sep 03 ...	Contoso IT - demo	phnakorn@microsoft.com
Get Database Top Queries query	Succeeded	6 d ago	Tue Sep 03 ...	Contoso IT - demo	phnakorn@microsoft.com
Get Database Top Queries query	Succeeded	6 d ago	Tue Sep 03 ...	Contoso IT - demo	phnakorn@microsoft.com
Get Database Top Queries query	Succeeded	6 d ago	Tue Sep 03 ...	Contoso IT - demo	phnakorn@microsoft.com
Audit	Succeeded	6 d ago	Tue Sep 03 ...	Contoso IT - demo	Microsoft Azure Policy Insights
Get Database Top Queries query	Succeeded	7 d ago	Tue Sep 03 ...	Contoso IT - demo	andersbe@microsoft.com



Activity Log & metrics

<https://learn.microsoft.com/en-us/azure/machine-learning/monitor-azure-machine-learning>

Resource Log Events

Compute

- AmlComputeClusterEvent
- AmlComputeCpuGpuUtilization
- AmlComputeJobEvent
- ComputeInstanceEvent

MLOps

- ModelsChangeEvent / ModelsReadEvent / ModelsActionEvent
- DeploymentReadEvent / DeploymentEventACI / DeploymentEventAKS
- InferencingOperationAKS / InferencingOperationACI
- EnvironmentChangeEvent / EnvironmentReadEvent
- PipelineChangeEvent / PipelineReadEvent
- AmlRunStatusChangedEvent / RunEvent / RunReadEvent

Data Management

- DataLabelChangeEvent / DataLabelReadEvent
- DataStoreChangeEvent / DataStoreReadEvent
- DataSetChangeEvent / DataSetReadEvent

Recommended alerts

Alert type	Condition	Description
Model Deploy Failed	Aggregation type: Total, Operator: Greater than, Threshold value: 0	When one or more model deployments have failed
Quota Utilization Percentage	Aggregation type: Average, Operator: Greater than, Threshold value: 90	When the quota utilization percentage is greater than 90%
Unusable Nodes	Aggregation type: Total, Operator: Greater than, Threshold value: 0	When there are one or more unusable nodes

Data evolves with its environment

Issue

- Production inferences could result bad predictions

Cause

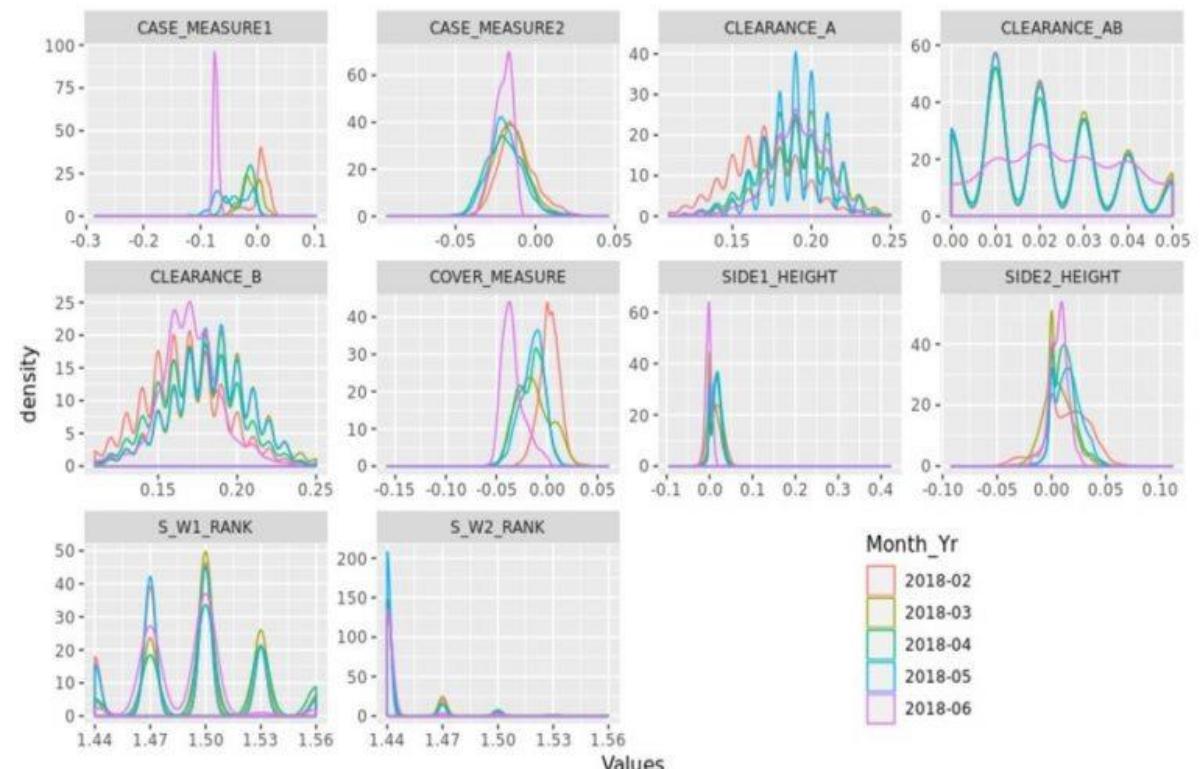
- Upstream process changes
- Data quality issues, such as a broken sensor always reading 0.
- Natural drift in the data.
- Change in relation between features, or covariate shift.

Solution

- Trigger alert
- Trigger training pipeline

(Drift service from azure machine learning)

Monitoring : Data Drift



Monitoring : Model Drift

Data evolves with its environment

Issue

Feature drift (features importance changed over the time)

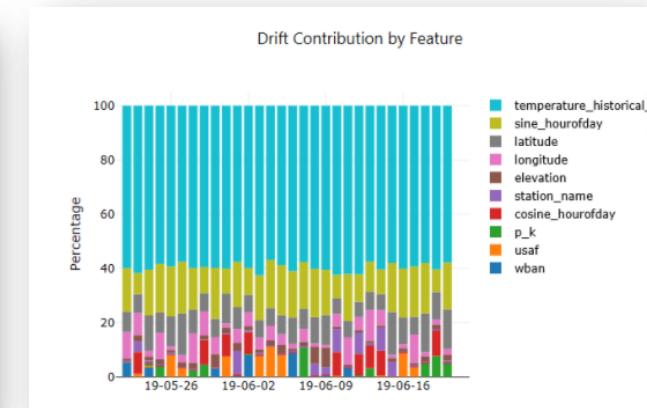
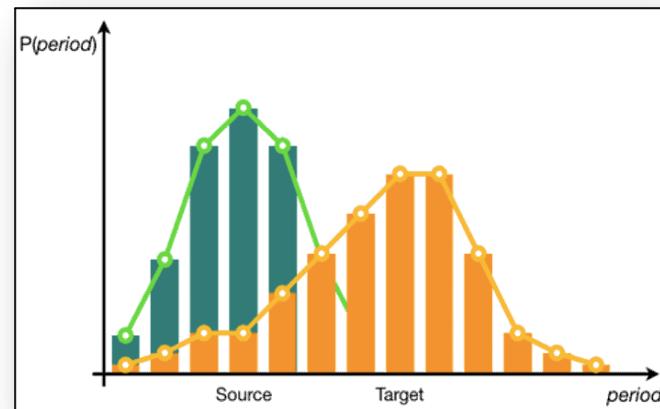
Label drift (distribution of the target changed)

Solution

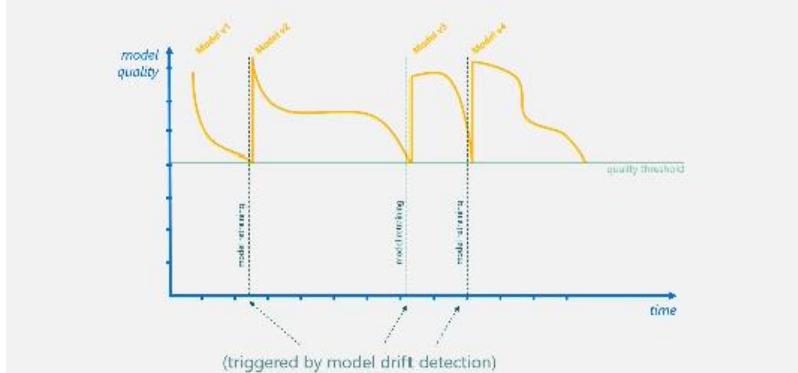
- Trigger alert
- Trigger training pipeline

(Monitoring dashboard – powerbi connected with your artifact store)

>> Anticipate performance issues in production



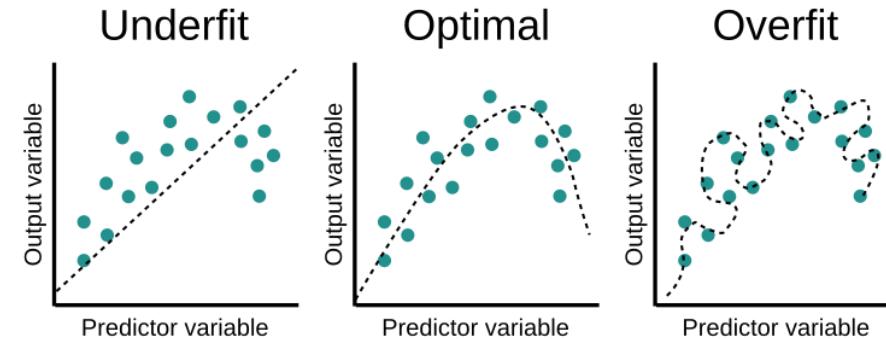
Model decay and retraining



Monitoring : Model Bias

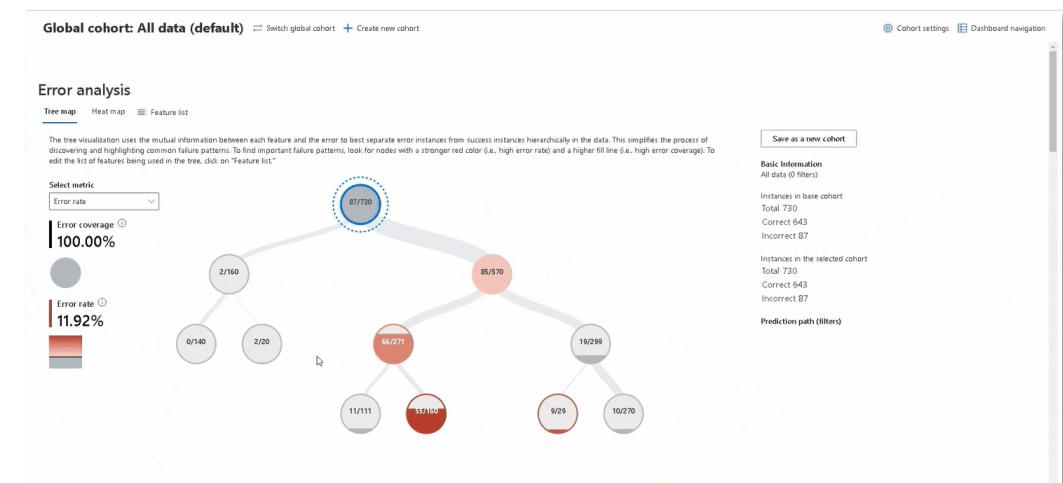
Issue

The model was wrongly trained, or the data were simply insufficient to represent reality which led to a *data drift*.



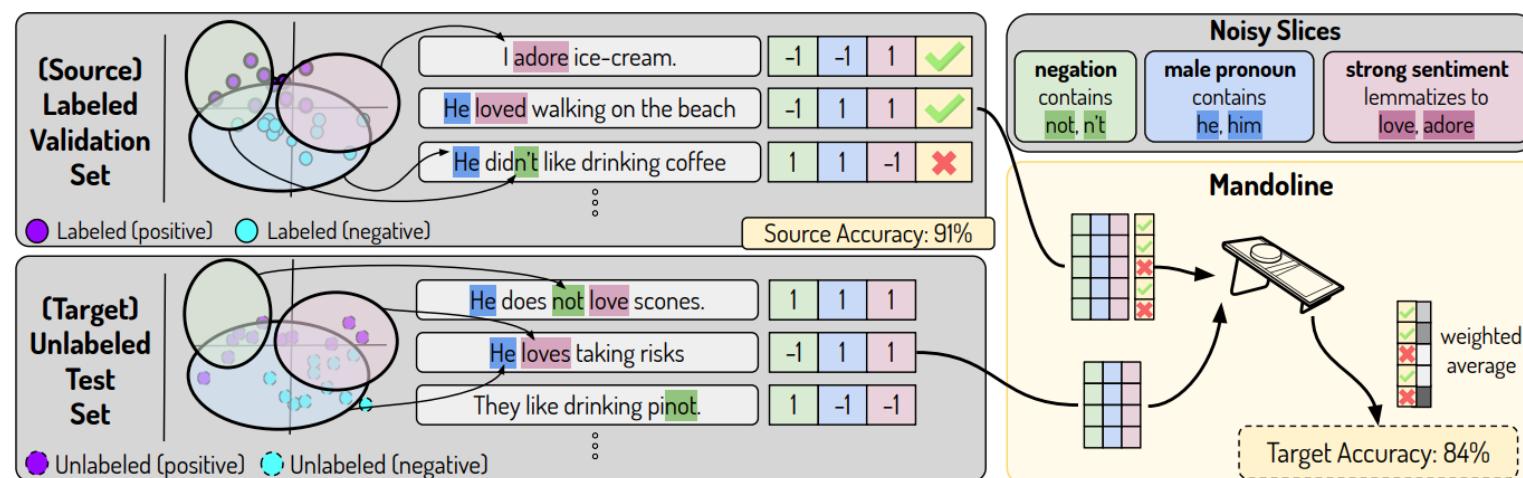
Solution

- AML Responsible AI dashboards → Error analysis/feature importance/ fairness...
- Rethink about data sampling/feature engineering
- Retraining



Monitoring with delayed outcomes

- ✓ Devise an **approximate signal** that can help us *estimate* the model's performance
- ✓ Label a small subset of our live dataset to estimate performance
- ✓ Rely on input data – importance weighting

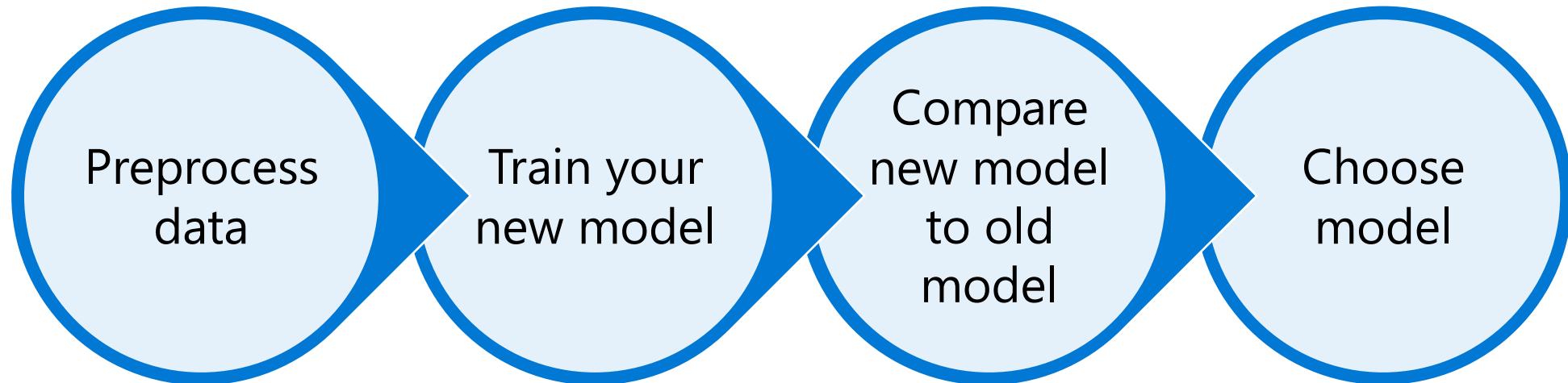


[Mandoline: Model Evaluation under Distribution Shift](#)

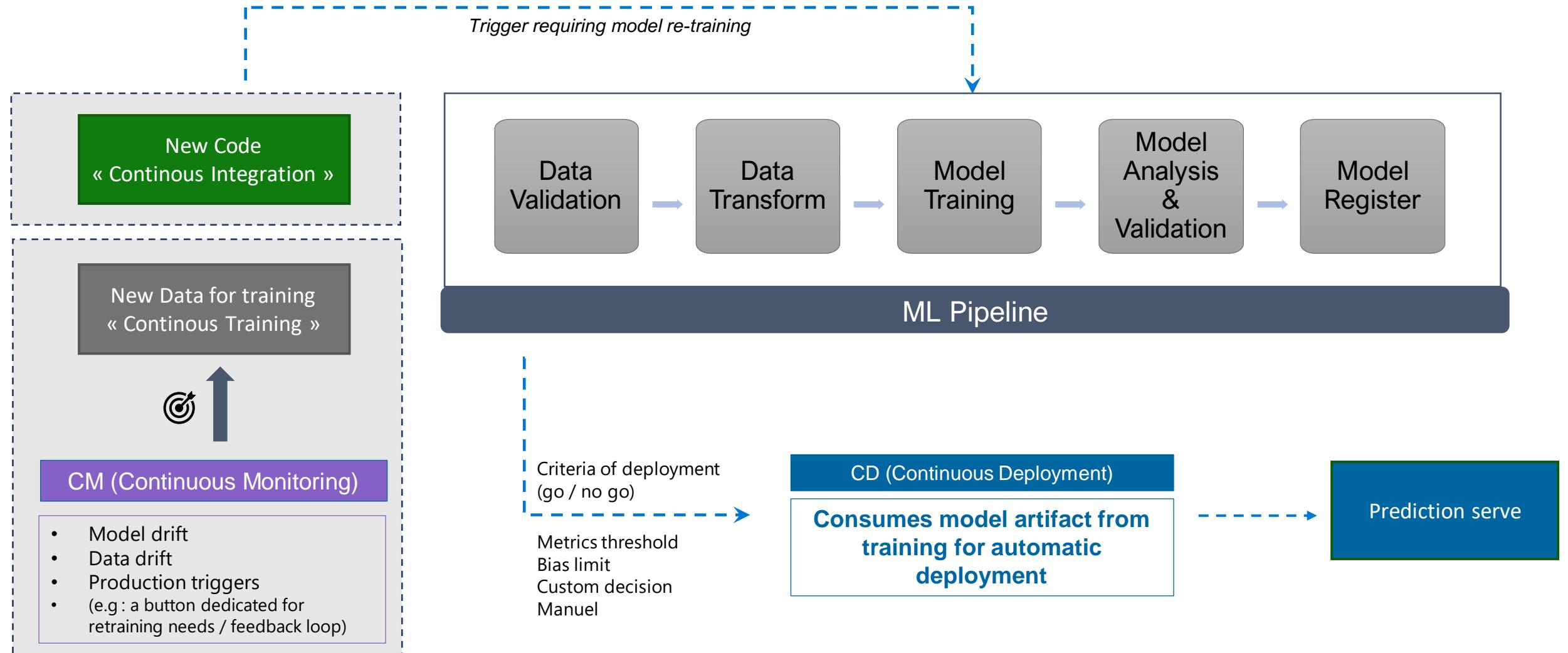


Retrain the model

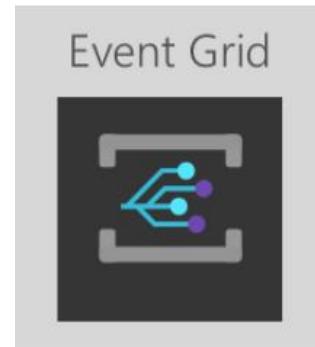
- ✓ On schedule
- ✓ When there is new data
- ✓ When an alert was triggered (drift, model decay, ...)



« Simplified flow »



Notify, Automate, Alert with Azure EventGrid (Preview)



<https://learn.microsoft.com/en-us/azure/machine-learning/how-to-use-event-grid>

Notify, Automate, Alert with Azure EventGrid (Preview)

Microsoft.MachineLearningServices.**RunCompleted**

- Raised when a machine learning experiment run is completed

Microsoft.MachineLearningServices.**ModelRegistered**

- Raised when a machine learning model is registered in the workspace

Microsoft.MachineLearningServices.**ModelDeployed**

- Raised when a deployment of inference service with one or more models is completed

Microsoft.MachineLearningServices.**DatasetDriftDetected**

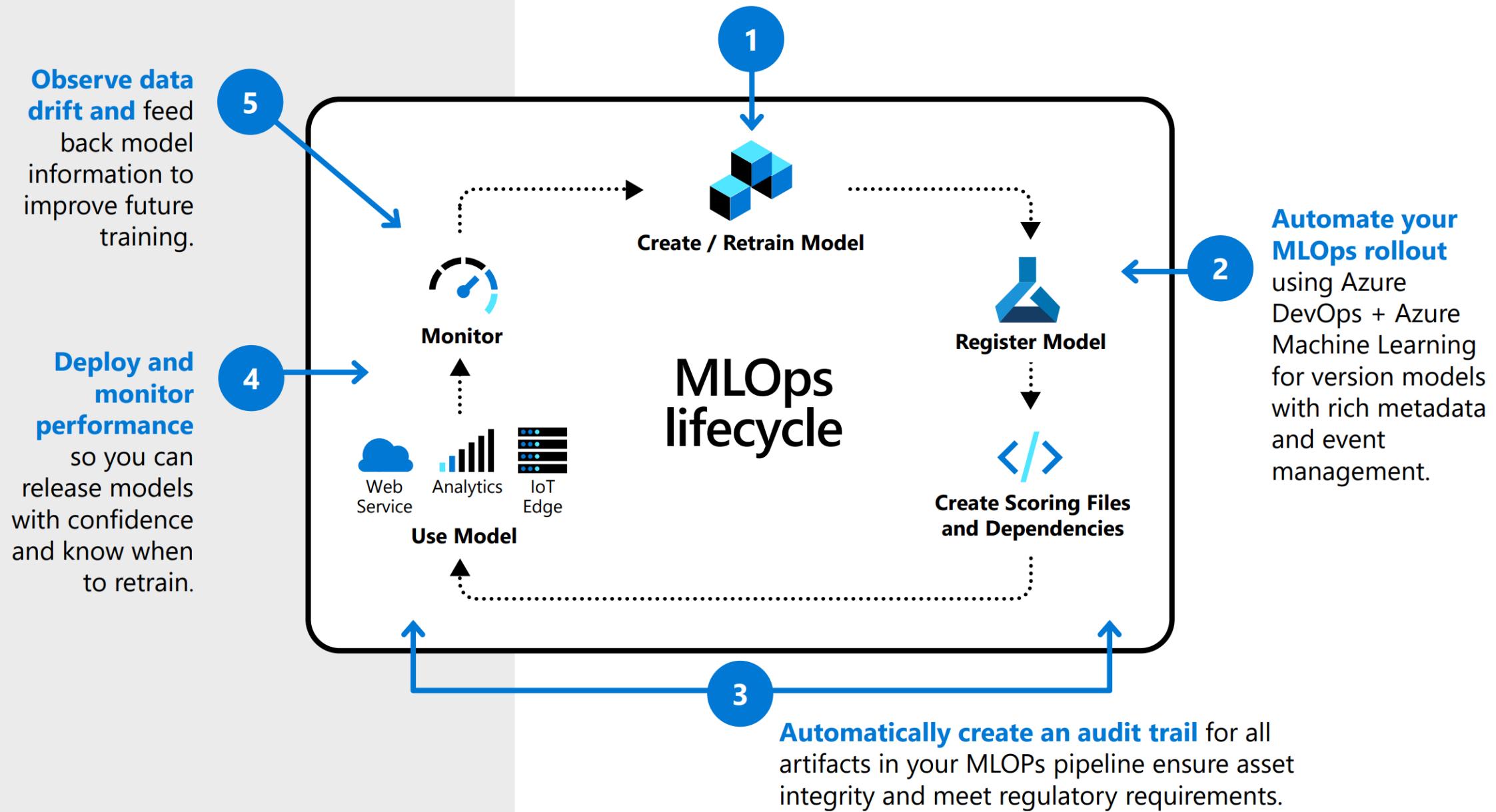
- Raised when a data drift detection job for two datasets is completed

Microsoft.MachineLearningServices.**RunStatusChanged**

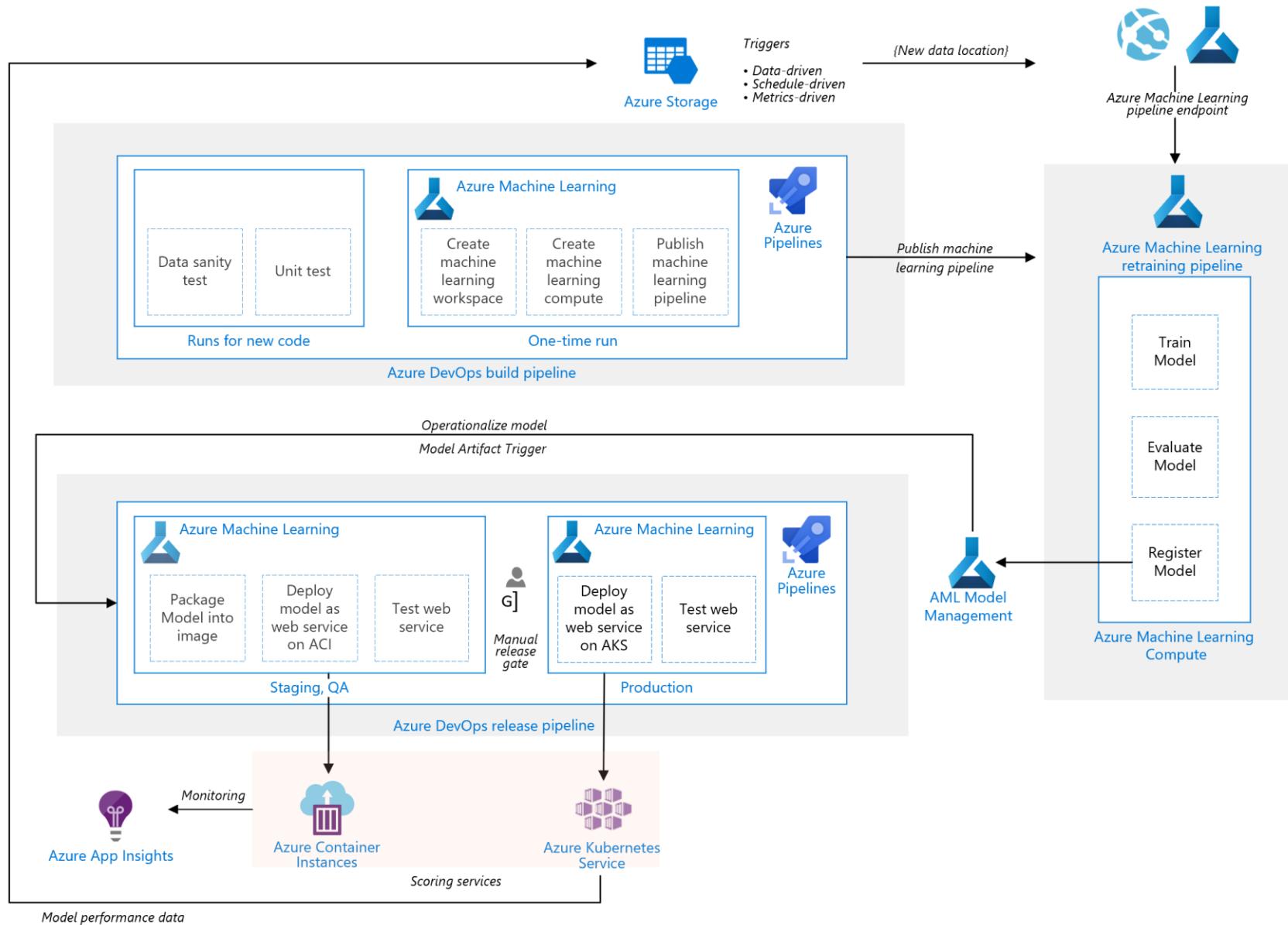
- Raised when a run status changed, currently only raised when a run status is 'failed'

5 Best Practices to optimize your MLOps lifecycle on Azure:

Create models with reusable ML pipelines using the Azure Machine Learning extension for Azure DevOps. Store your code in GitHub so it automatically integrates into your MLOps pipeline.



End to End Demo Architecture



Useful Links

- Architecture center :

<https://learn.microsoft.com/en-us/azure/architecture/reference-architectures/ai/mlops-python>

- MS learn MLOps learning path:

<https://learn.Microsoft.com/en-us/training/paths/introduction-machine-learn-operations>

- MLOps framework:

<https://learn.microsoft.com/en-us/azure/architecture/example-scenario/mlops/mlops-technical-paper>

- Try it out!

<https://github.com/microsoft/MLOpsPython>

Q&A

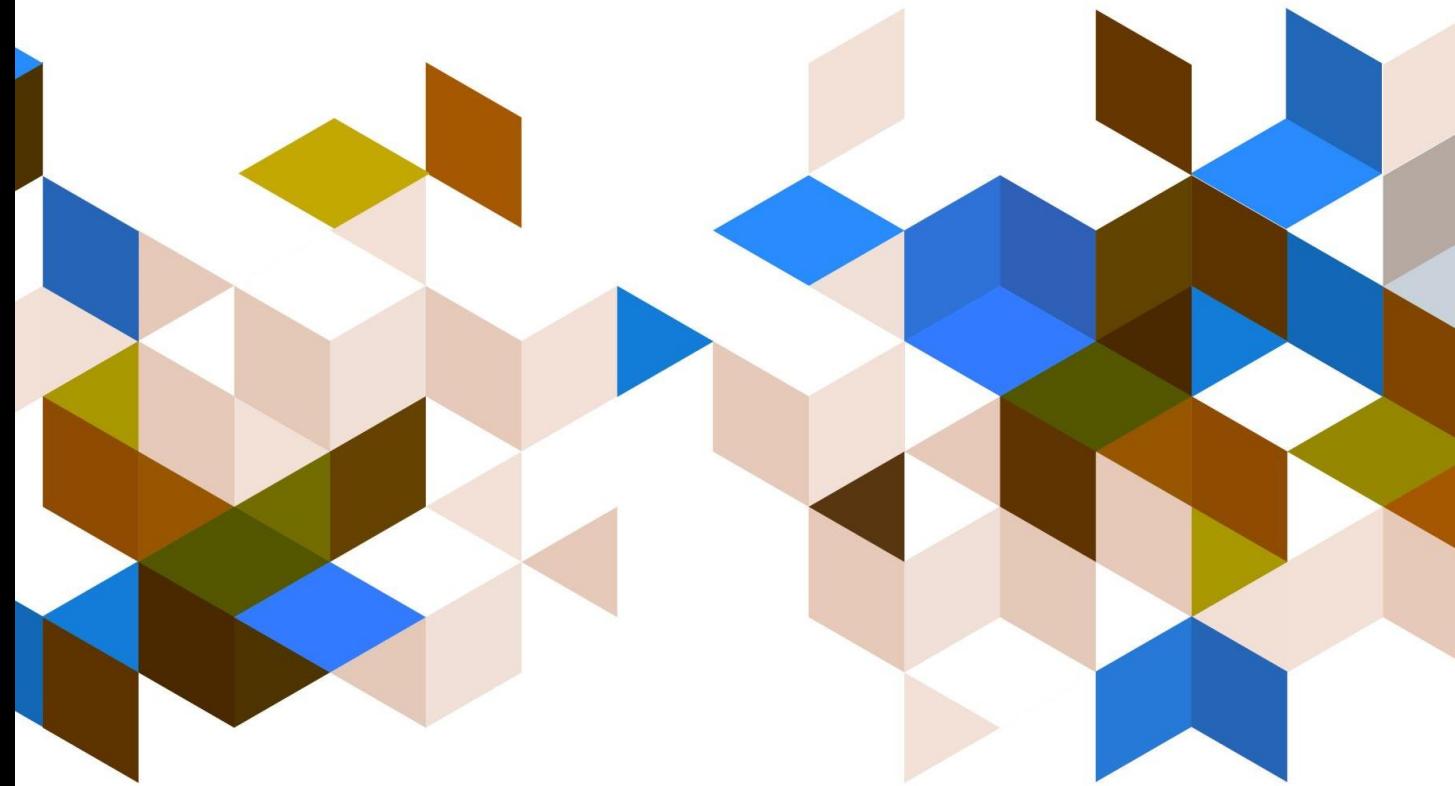




Enregistrez vous dès maintenant au prochain Webinars Data AI

Event Webinar (Les jeudis de la Data & AI) - L200/300	Date	Duration (min)	Link
Azure Synapse	22/09/2022	120	https://msevents.microsoft.com/event?id=857781749
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	29/09/2022	120	https://msevents.microsoft.com/event?id=502366997
Déploiement et sécurisation des workspaces Azure Machine learning	06/10/2022	120	https://msevents.microsoft.com/event?id=1505714138
Azure Scale Analytics - Architectures Data Mesh dans Azure avec Azure Synapse, Microsoft Purview et Azure Data Share	13/10/2022	120	https://msevents.microsoft.com/event?id=139685175
MLOps avec Azure Machine Learning	20/10/2022	120	https://msevents.microsoft.com/event?id=1245885767
SQL Server 2022 et hybridation native avec Azure SQL Managed Instance	10/11/2022	120	https://msevents.microsoft.com/event?id=145826476
Machine Learning dans Azure Synapse Analytics	17/11/2022	120	https://msevents.microsoft.com/event?id=3637723312
Azure Cosmos DB et IA	24/11/2022	120	https://msevents.microsoft.com/event?id=2646013445
Azure et les Services Cognitifs	08/12/2022	120	https://msevents.microsoft.com/event?id=3772037220
La gouvernance de données dans Azure avec Microsoft Purview	15/12/2022	120	https://msevents.microsoft.com/event?id=1499560981
MLOps avec Azure Machine Learning	12/01/2023	120	https://msevents.microsoft.com/event?id=4115194515
	19/01/2023	120	https://msevents.microsoft.com/event?id=1537241181
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	26/01/2023	120	https://msevents.microsoft.com/event?id=1806467748
Déploiement et sécurisation des workspace Azure Synapse	09/02/2023	120	En cours
Azure Machine Learning pour les Citizen Data Scientists	16/02/2023	120	https://msevents.microsoft.com/event?id=1401519679
L'IA responsable avec Azure machine learning	09/03/2023	120	https://msevents.microsoft.com/event?id=2072953112
Machine Learning dans Azure Synapse Analytics	16/03/2023	120	https://msevents.microsoft.com/event?id=3413014857
Les bases de données Open Source dans le cloud Azure	23/03/2023	120	https://msevents.microsoft.com/event?id=2727487131
Hybridation des services de Machine Learning Azure	06/04/2023	120	https://msevents.microsoft.com/event?id=1624914222
La gouvernance de données dans Azure avec Microsoft Purview	13/04/2023	120	https://msevents.microsoft.com/event?id=3909342839
Les solutions SQL dans Azure (PaaS, IaaS, SaaS)	04/05/2023	120	https://msevents.microsoft.com/event?id=1162207895
	16/05/2023	120	https://msevents.microsoft.com/event?id=3517068442
Data processing dans Azure ave Azure Synapse, Azure Batch, Spark, Notebook, etc.	24/05/2023	120	https://msevents.microsoft.com/event?id=2996507398
Self Service Analytics	01/06/2023	120	En cours

Annexes



Model Versioning & Storage

Provide a consistent way to discover, store, track & share models

Provide a consistent model metadata format

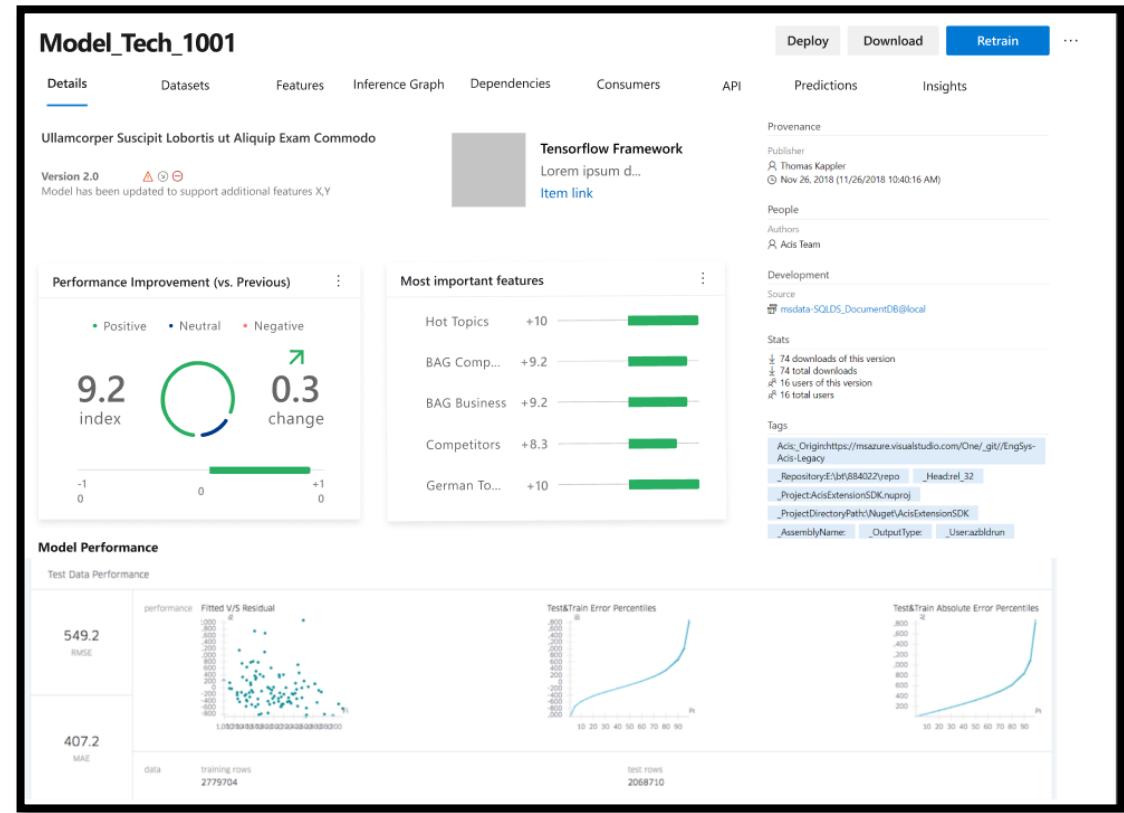
Track where a model came from

- which data,
- which experiment / previous model(s),
- where's the code / notebook)
- Was it converted / quantized?

Track where model is running & model health

Control who has access to what models

- Private / compliant data

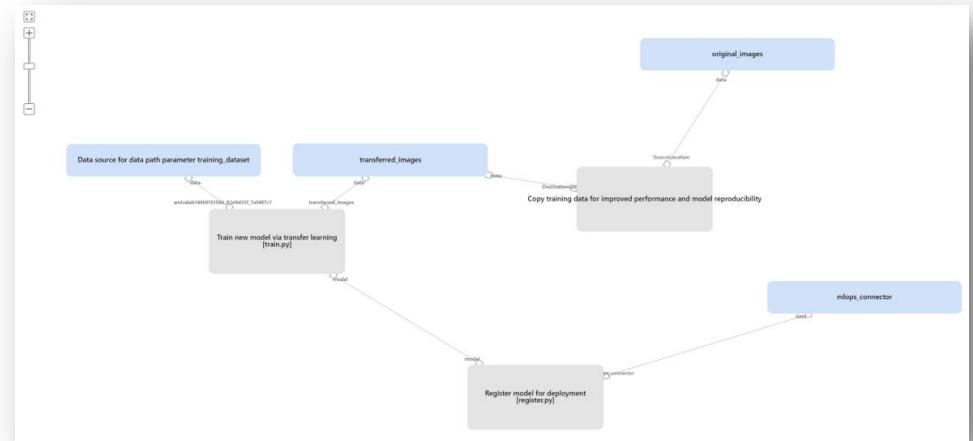


Azure ML – ML Pipelines

Manage & scale up your multi-step workflows

Increase experiment **velocity, reliability, repeatability**
Use the **technology of your choice** for each step

Create & manage ML workflows concurrently



Key advantage	Description
Unattended runs	Schedule a few steps to run in parallel or in sequence in a reliable and unattended manner. Since data prep and modeling can last days or weeks, you can now focus on other tasks while your pipeline is running.
Mixed and diverse compute	Use multiple pipelines that are reliably coordinated across heterogeneous and scalable computes and storages. Individual pipeline steps can be run on different compute targets, such as HDInsight, GPU Data Science VMs, and Databricks, to make efficient use of available compute options.
Reusability	Pipelines can be templated for specific scenarios such as retraining and batch scoring. They can be triggered from external systems via simple REST calls.
Tracking and versioning	Instead of manually tracking data and result paths as you iterate, use the pipelines SDK to explicitly name and version your data sources, inputs, and outputs as well as manage scripts and data separately for increased productivity.

Azure ML – Data Drift

Know when you should retrain your models

What is Data Drift

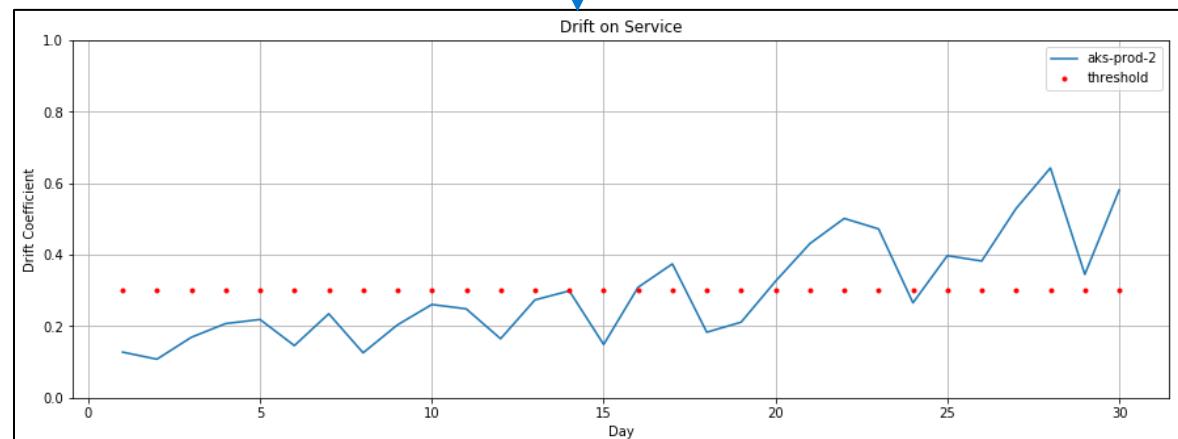
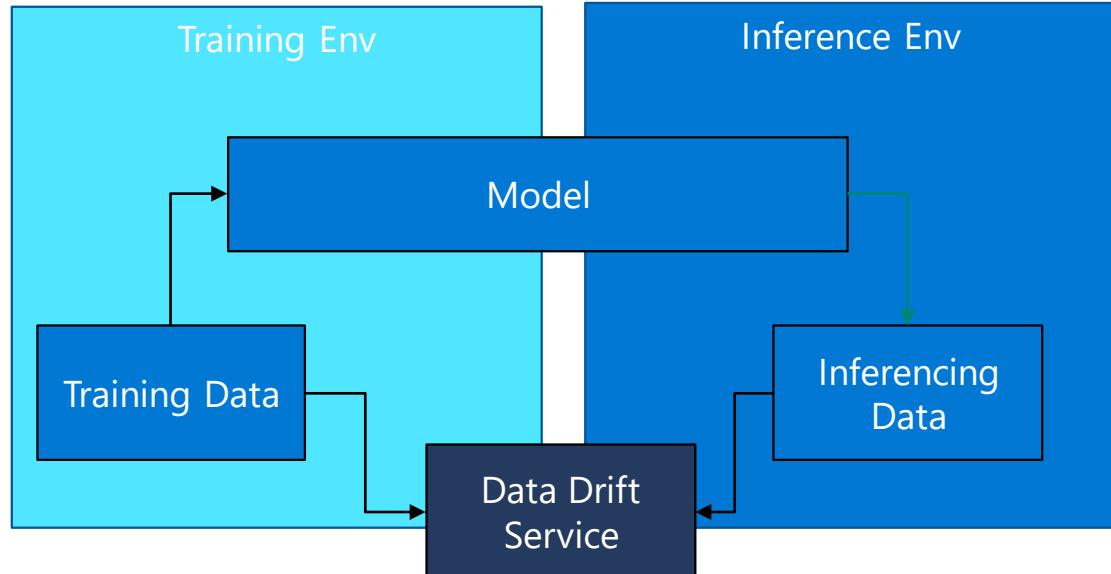
- When the inference data varies enough from the training data to cause model degradation, we consider it data drift

Why is Data Drift important

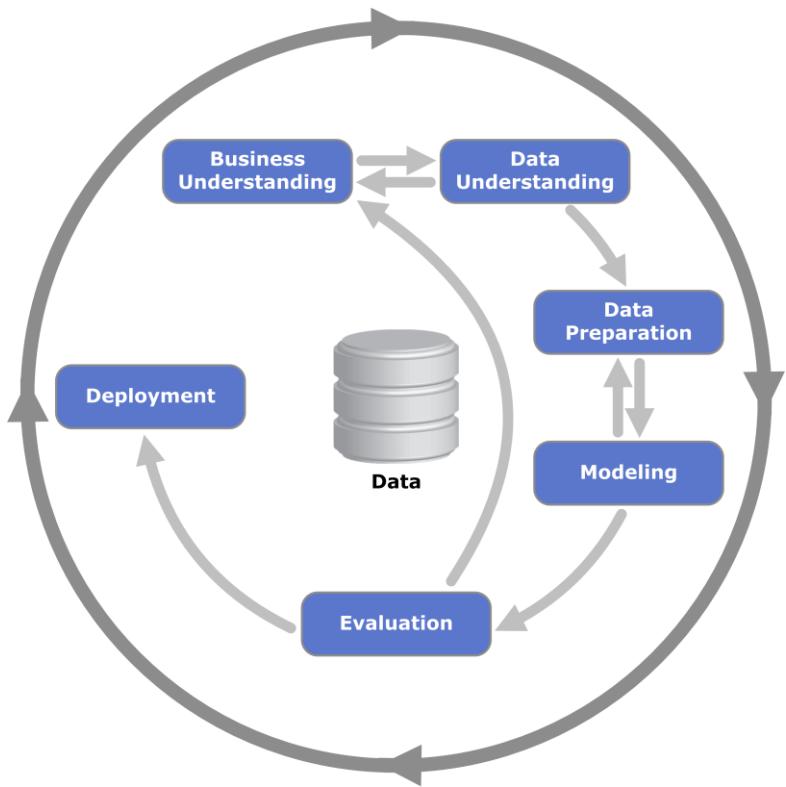
- in absence of labels, data drift can be used as a proxy for model performance metrics

What can I expect from data drift service

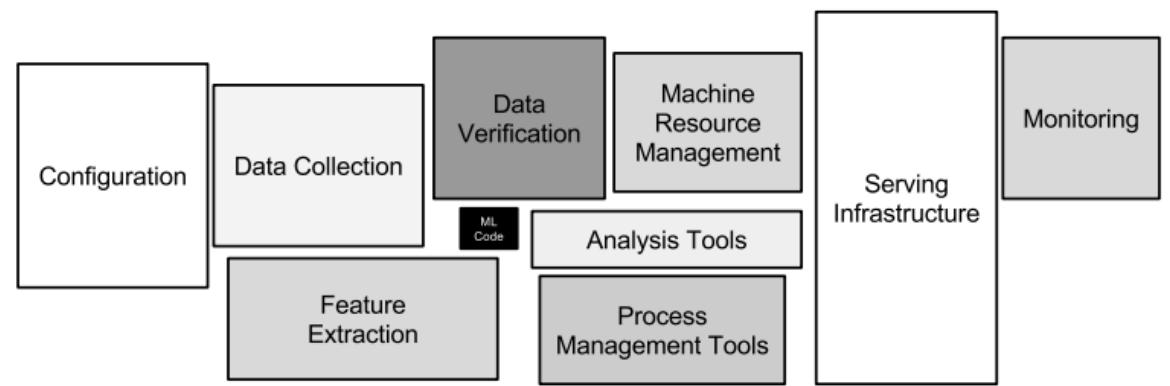
- Alerting and Insights** in Previews
- Drift Coefficient** - measures magnitude of drift
- Drift Contribution by Feature** - measures feature that caused data drift
- Distance Metrics by Feature** - distance metrics on features
- Distribution Visualization** - capture and visualize change in distributions over time



Traditional ML modeling



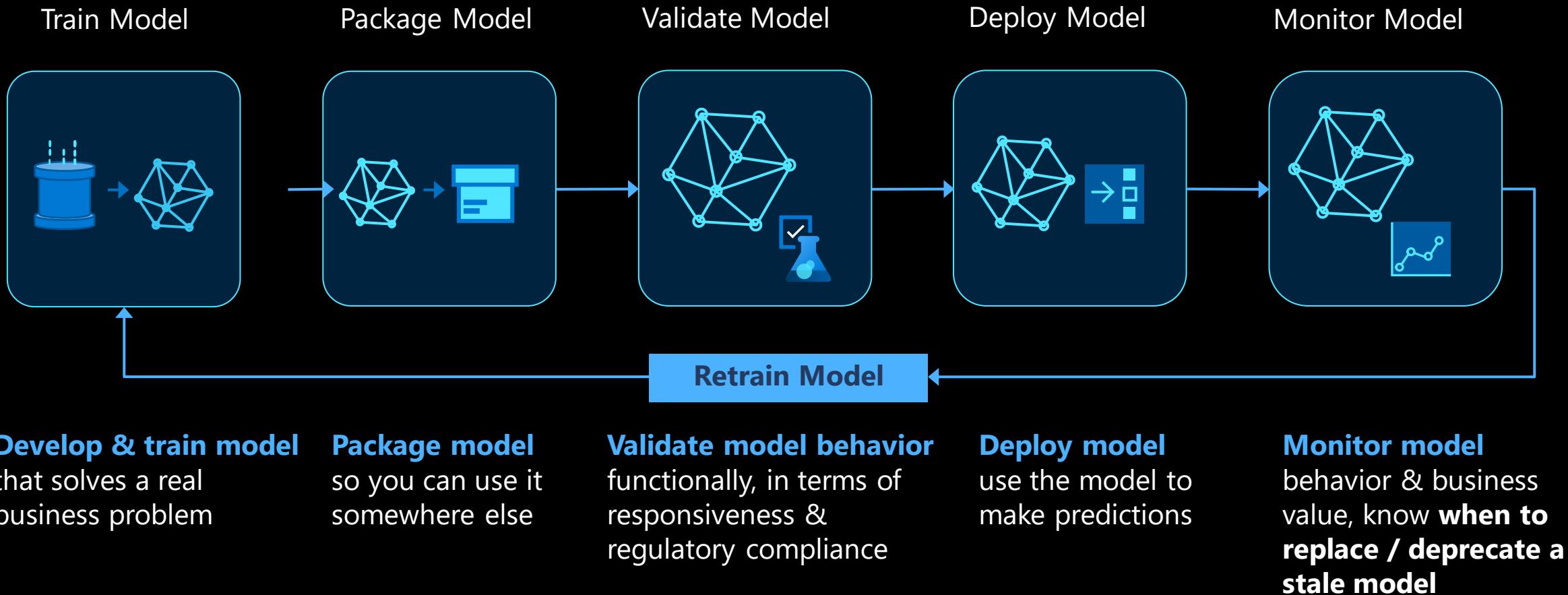
Production ML requires much more



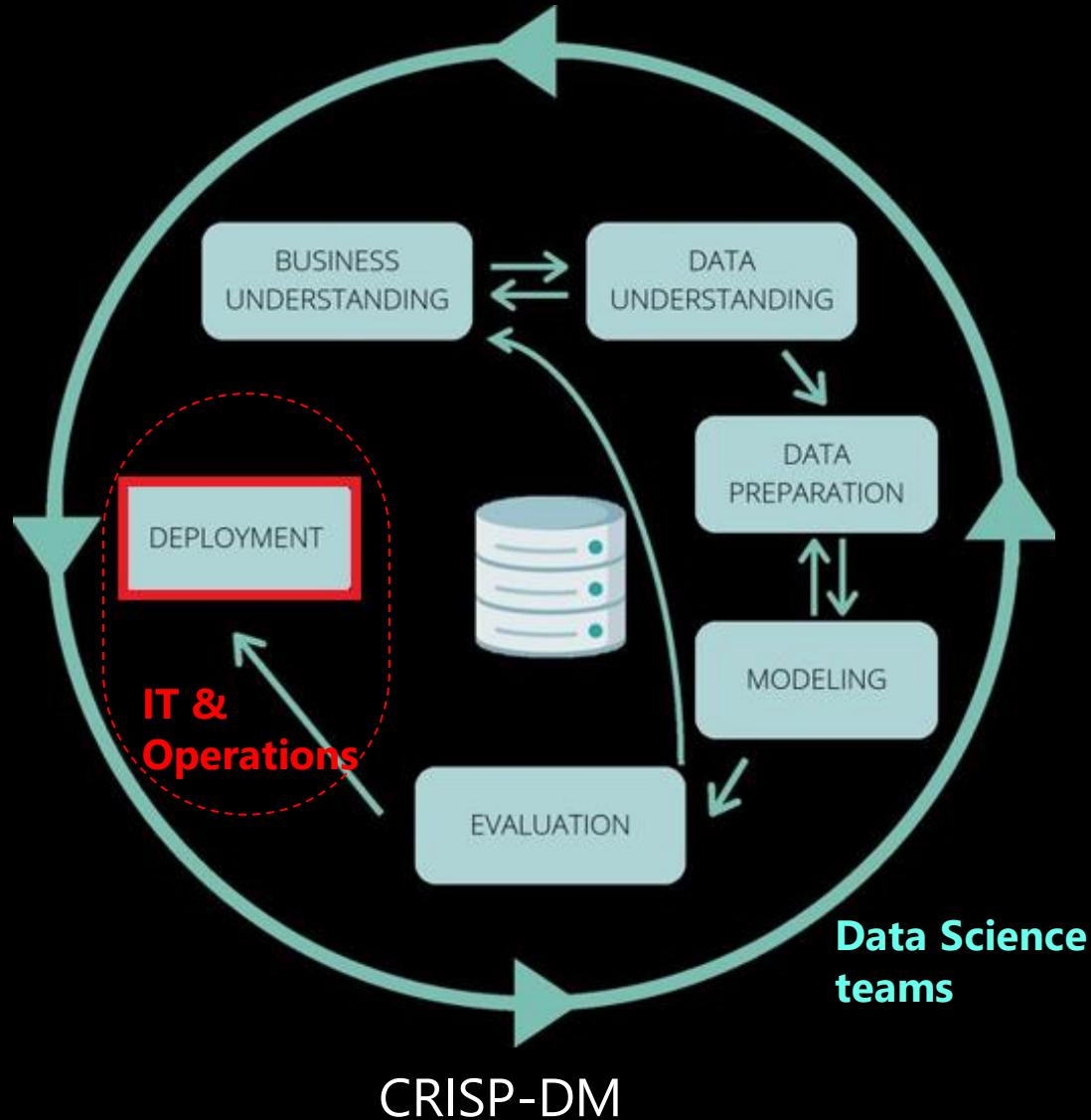
Source: Hidden technical debt in ML systems

Prod ML = ML dev + Modern software dev

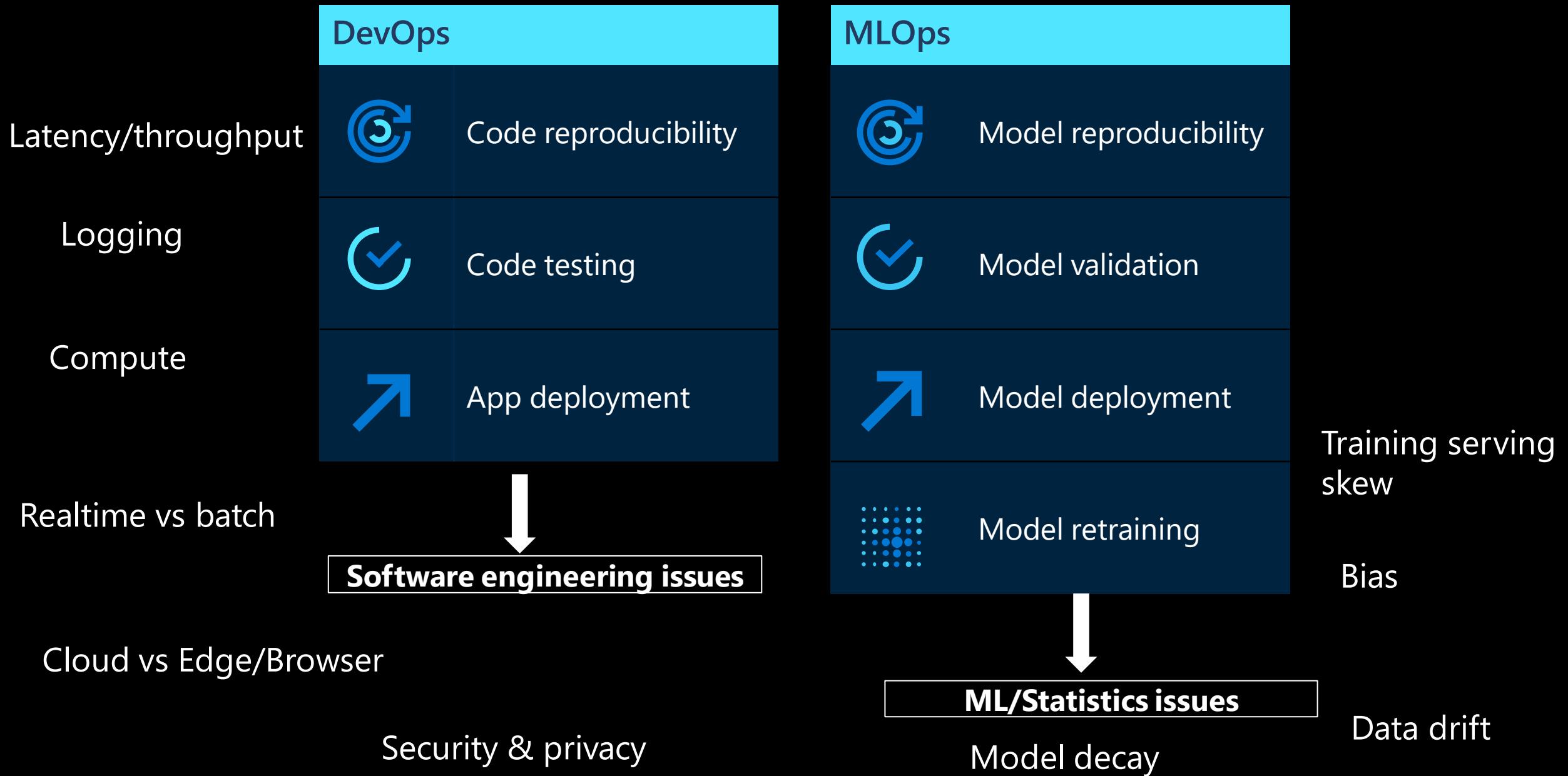
What does MLOps look like?



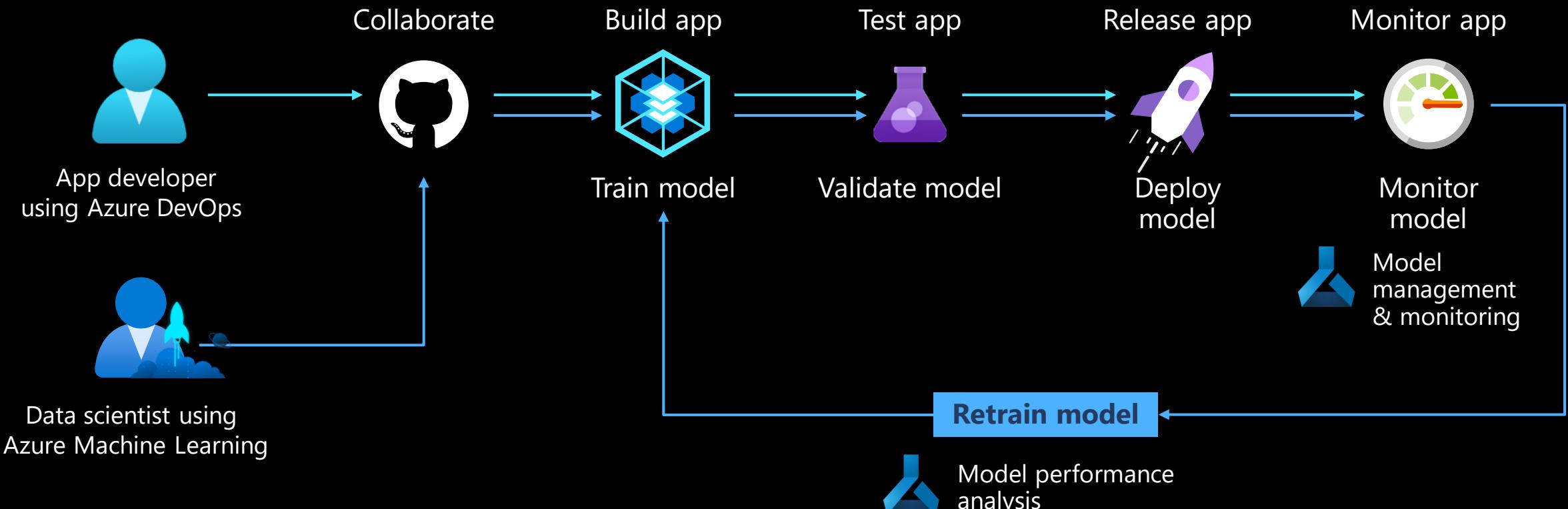
Traditional Data Mining Process



- Need to align business metrics and ML metrics
- Need to define clear processes
- Shorten 1st iteration duration – quick baseline then iterate to improve



MLOps Workflow



✓ Model reproducibility

✓ Model validation

✓ Model deployment

✓ Model retraining

Model Validation - Test Types

Responsibilities across roles in the context of *model development*.

Test Type	Data Scientist	App Dev / Ops
Unit Tests	X	
Data Integrity Tests		X
Model Performance	X	
Model Validation		X
Integration Tests	X	X
Load Tests		X
Data Monitoring		X
Skew Monitoring		X
Model Monitoring	X	X

*Note the additional testing responsibilities for the DE in the context of data pipeline development.

Model Validation - Summary

Profile and validate

- Data (changes to shape / profile)
- Model in isolation (offline A/B)
- Model + app (functional testing)

Control model rollout on validation

- Only deploy after initial validation passes
- Ramp up traffic to new model using A/B experimentations

Assess performance

- Functional behavior
- Performance characteristics

The screenshot shows a validation summary for a build. At the top, under 'Policies', four items are listed as required: '2 reviewers approved' (green checkmark), 'Required reviewers have approved' (green checkmark), 'Work items linked' (green checkmark), and 'Gated Build - AEther Modules succeeded' (green checkmark). Below this, the 'Run History' tab is selected, showing a single completed run. The run details include: Status: Completed, Start Time: May 1, 2018, 11:18:45 AM, Duration: 5s, Target: sdk, Run Id: 9ee0e521-d33b-45f1-8938-0edd9acd62f, Run Number: 40, Script Name: run_client.py, and Arguments. The 'Attributes' section also lists these same details.

The screenshot shows a metrics dashboard for a build. At the top, there's a search bar with 'Build filter: a83c0947|125b6e2b|b238a' and a 'Filter' button. Below the search bar, the 'Metrics' section displays tracked metrics: accuracy (0.995253164556962), precision (0.8888888888888888), recall (0.9583333333333334), and feature_importance ({"Importance": {"min1": 0.016218768, "num..."}). A chart titled 'feature_importance va...' shows a sharp increase from near zero to approximately 0.45 at index 12. The bottom section, 'Build summary', shows performance metrics for 'Hippogriff.Server performance MSIT': .NET CLR Memory(Hippogriff.Server)\% Time in GC (85th %) at 4.35, Process(Hippogriff.Server)\% Processor Time (85th %) at 29.4, Process(Hippogriff.Server)\Private Bytes (85th %) at 490Mb, and Process(Hippogriff.Server)\Working Set - Private (85th %) at 459Mb. All values are shown with small green arrows indicating slight increases.

The screenshot continues the metrics dashboard from the previous section. The 'Build summary' table is identical to the one above. Below it, a more detailed table provides specific performance metrics for 'Hippogriff.Server performance MSIT': .NET CLR Memory(Hippogriff.Server)\% Time in GC (85th %) at 4.35, Process(Hippogriff.Server)\% Processor Time (85th %) at 29.4, Process(Hippogriff.Server)\Private Bytes (85th %) at 490Mb, and Process(Hippogriff.Server)\Working Set - Private (85th %) at 459Mb. The last two rows show 'Show hidden metrics V' and 'PR#6821 b238a3d' respectively.