



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

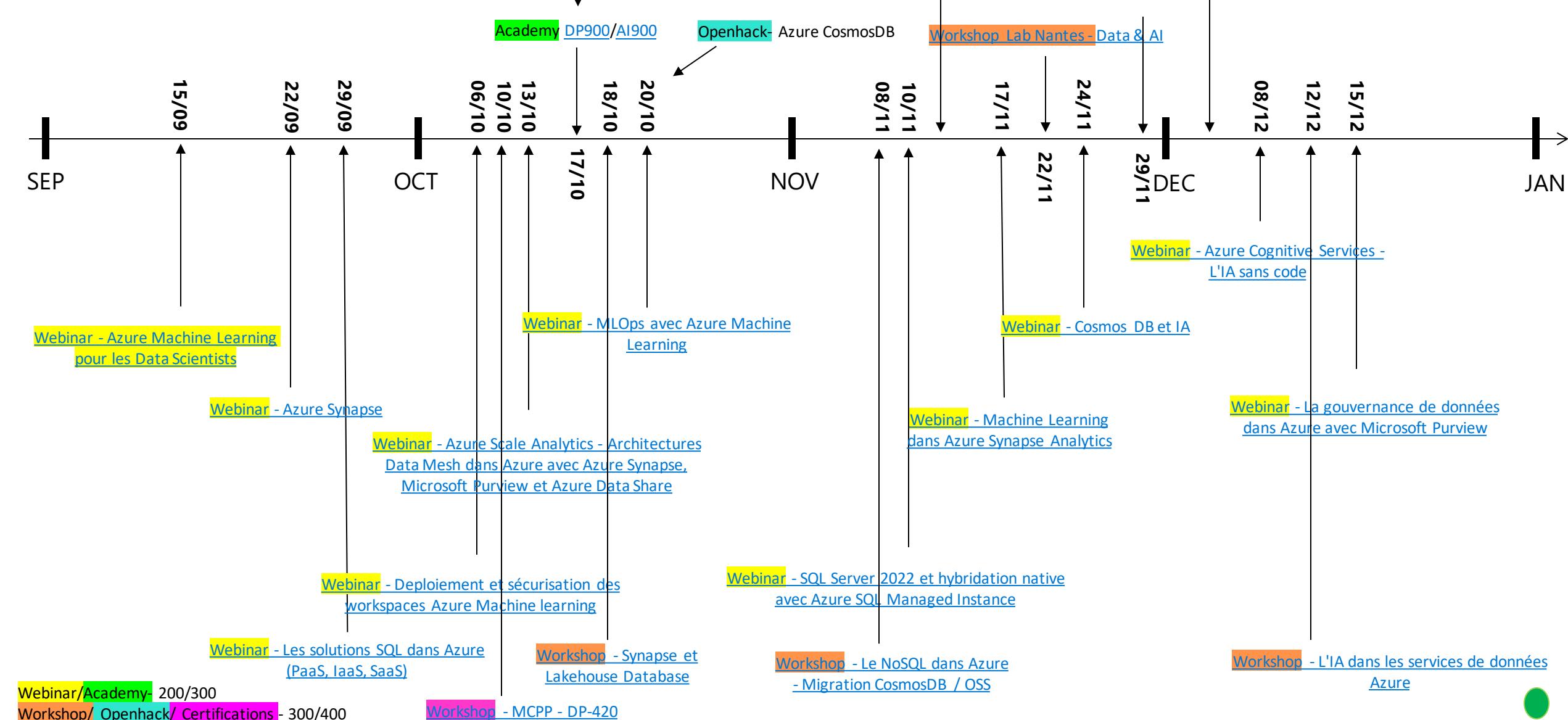


Kickstart (17/10)

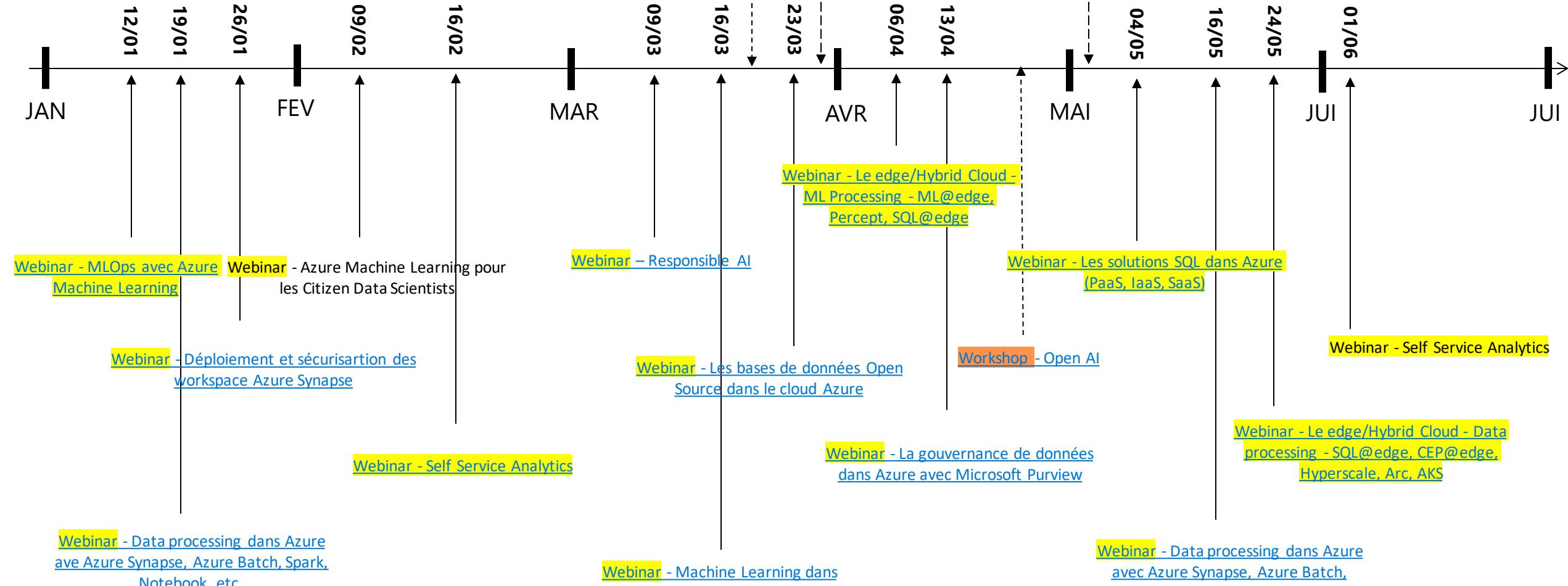
Dry Run (14/11)

Q&A (05/12)

Plan GPS Data/AI global FY23 (H1)



Plan GPS Data/AI global FY23 (H2)



Webinar - Data processing dans Azure avec Azure Synapse, Azure Batch, Spark, Notebook, etc.

Webinar/Academy- 200/300

Workshop/ Openhack/ Certifications - 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=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFSDVLR0cyS1kwWS4u
Le NoSQL dans Azure - Migration CosmosDB / OSS	08/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFSDVLR0cyS1kwWS4u
Lab Lyon - Data & AI	22/11/2022	240	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUMIZZOURET0RSWjcyTERYRkJGTIFFUjaUi4u
Lab Nantes - Data & AI	29/11/2022	240	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUMIZZOURET0RSWjcyTERYRkJGTIFFUjaUi4u
L'IA dans les services de données Azure	12/12/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFSDVLR0cyS1kwWS4u
Open AI	H2	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdURE1RMVgwTDNISTE1TDFSDVLR0cyS1kwWS4u

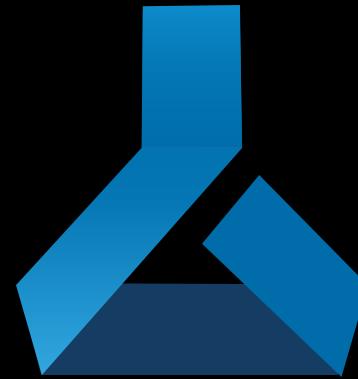
Event Academy, kickstart certifications, workshop certifications	Date	Duration (min)	Link
MCPP - DP-420	10/10/2022	420	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUMkJSIRKSU1RRFA0OVgzSFdtSTY0RE9WQy4u
Micro Hack CosmosDB	20/10/2022	420	H1 - Inscriptions 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=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNEk3WFQ1TEdNNTO2Uk85V0cxQzM3TE9ZRS4u
Dry Run DP-500	14/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNEk3WFQ1TEdNNTO2Uk85V0cxQzM3TE9ZRS4u
Q&A DP-500	05/12/2022	90	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNEk3WFQ1TEdNNTO2Uk85V0cxQzM3TE9ZRS4u
Kickstart DP-100	17/10/2022	60	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNDAxV0hSN0FHM1YzUzI30UNMFYxSkRIMi4u
Dry Run DP-100	14/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNDAxV0hSN0FHM1YzUzI30UNMFYxSkRIMi4u
Q&A DP-100	05/12/2022	90	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUNDAxV0hSN0FHM1YzUzI30UNMFYxSkRIMi4u
Kickstart DP-203	17/10/2022	60	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUOVFWOUVCNFcyOk5SVjFBUFczNktCUFpLMi4u
Dry Run DP-203	14/11/2022	120	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUOVFWOUVCNFcyOk5SVjFBUFczNktCUFpLMi4u
Q&A DP-203	05/12/2022	90	https://forms.office.com/Pages/ResponsePage.aspx?id=v4j5cvGGr0GRqy180BHB3zwJTO3s11AuaqpNnBbrwdUOVFWOUVCNFcyOk5SVjFBUFczNktCUFpLMi4u





Azure Machine Learning Service for data scientists

15/09/2022



Speaker info



Narjes Majdoub

Cloud Solutions Architect
Data & AI @ Microsoft

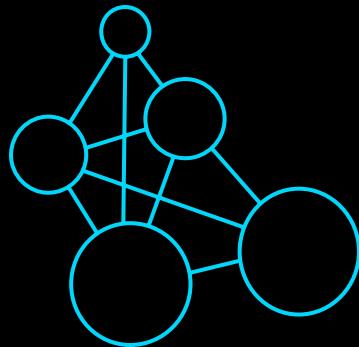


Franck Gaillard

Cloud Solutions Architect
Data & AI @ Microsoft

Agenda

-
1. Introduction to AML
 2. Technical details of AML
 3. Automated ML
 4. Introduction to Azure ML and MLOps

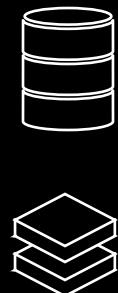


Requirements of an advanced ML Platform

Machine Learning

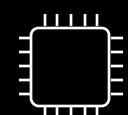
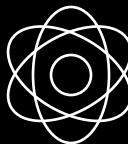
Typical E2E Process

Prepare

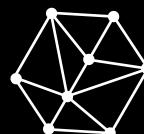


Prepare Data

Experiment

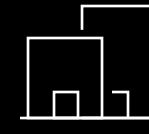


Build model
(your favorite IDE)



Train &
Test Model

Deploy

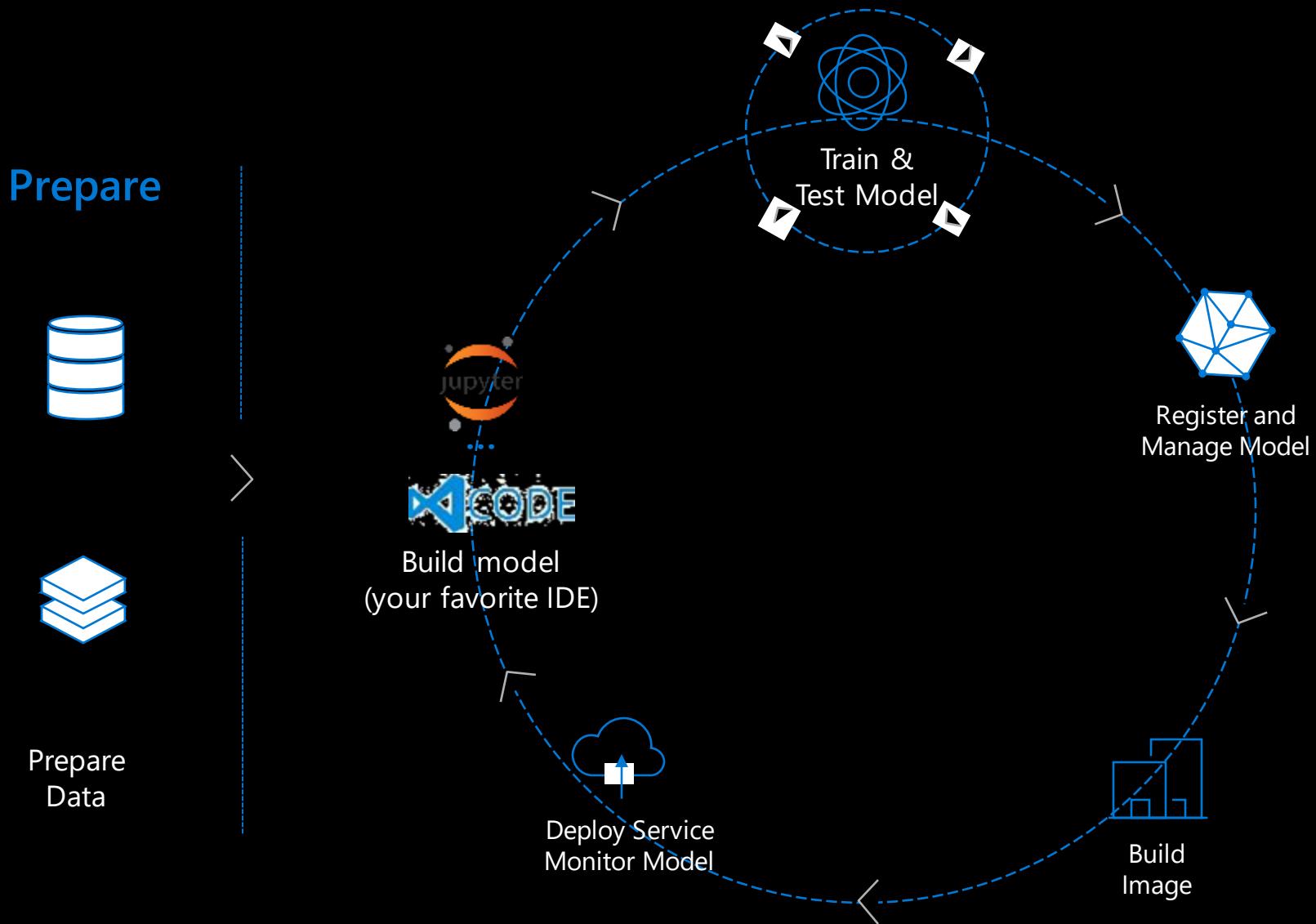


Deploy Service
Monitor Model

Orchestrate

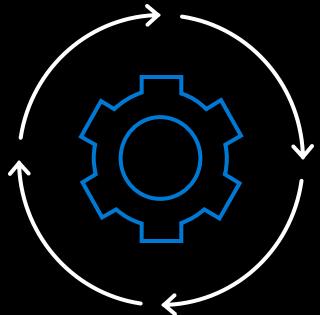


DevOps loop for data science



Key Trends

Data Science and ML platforms



Automation

Automated workflows for deployment and management 1000s of models

Composite AI and transfer learning techniques with models like GPT3

Configurable and repeatable recipes like NLP, Recommenders, many-models

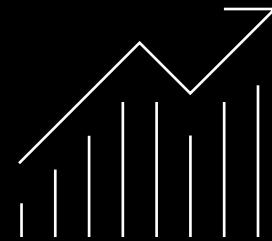


Collaboration

Collaborative tools and processes as multiple roles contributing to ML practice

Robust Responsible AI approach to ensure ethical use due to multiple stakeholders

Enterprise grade data and model governance/security

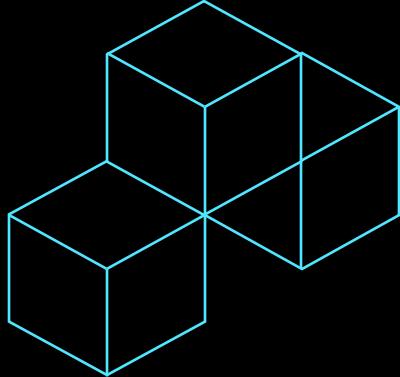


Acceleration

Rise of MLOps to accelerate operationalization of models

Advanced techniques like reinforcement learning, GANs and synthetic data





Azure offers a comprehensive
AI/ML platform that meets—and
exceeds—requirements



Intended Audience

BUILD MODELS



Citizen
Data Scientist



Professional
Data Scientist



Developers

OPERATIONALIZE MODELS



ML Engineers



IT



Azure Machine Learning

The one central hub for your data science team

Boosted collaboration

Integration with other Azure services

The screenshot shows the Microsoft Azure Machine Learning studio interface. On the left is a sidebar menu with the following items:

- New
- Home
- Author
- Notebooks
- Automated ML
- Designer
- Assets
- Datasets
- Experiments
- Pipelines
- Models
- Endpoints
- Manage
- Compute
- Datastores
- Data Labeling

The main area is titled "Azure Machine Learning studio" and contains four cards:

- Notebooks**: "Code with Python SDK and run sample experiments." with a "Start now" button.
- Automated ML**: "Automatically train and tune a model using a target metric." with a "Start now" button.
- Designer**: "Drag-and-drop interface from prepping data to deploying models." with a "Start now" button.
- A large blue "+" button labeled "Create new" with a dropdown arrow.

Below these cards is a section titled "My recent resources" with a table titled "Runs".

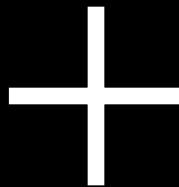
Run	Run ID	Experiment	Status	Submitted time	Submitted by	Run type
Run 74	AutoML_133595d2-2485...	ntFlightD...	Completed	Oct 29, 2020 4:42 PM	Nishant Thac...	Automated...
Run 630	69fd25f-882e-4845-aa7e...	manymo...	Completed	Oct 29, 2020 2:05 PM	Service Princi...	Pipeline
Run 31	AutoML_e5431f1-663d-4...	ntFlightD...	Completed	Oct 29, 2020 1:51 PM	Nishant Thac...	Automated...
Run 613	0f0ebe29-d3d7-4083-9fe...	manymo...	Completed	Oct 29, 2020 12:49 PM	Service Princi...	Pipeline
Run 1	AutoML_f7583e85-bb3f-4...	ntFlightD...	Completed	Oct 28, 2020 11:34 PM	Nishant Thac...	Automated...

Azure Machine Learning Service

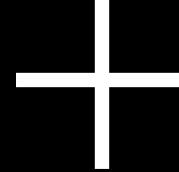
Make data scientists to be more productive

Enable your organization to manage the ML lifecycle through MLOps

Azure Cloud
Services



Python
SDK



Cross-Platform
CLI

That enables you to:

- ✓ Prepare Data
- ✓ Build Models
- ✓ Train Models

- ✓ Manage Models
- ✓ Track Experiments
- ✓ Deploy Models



Machine Learning on Azure

Domain specific pretrained models

To reduce time to market



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

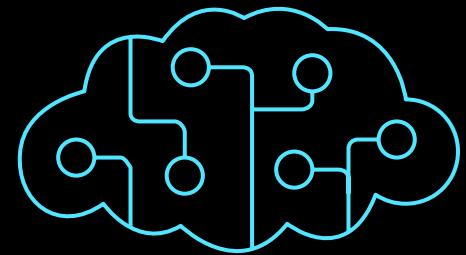


FPGA



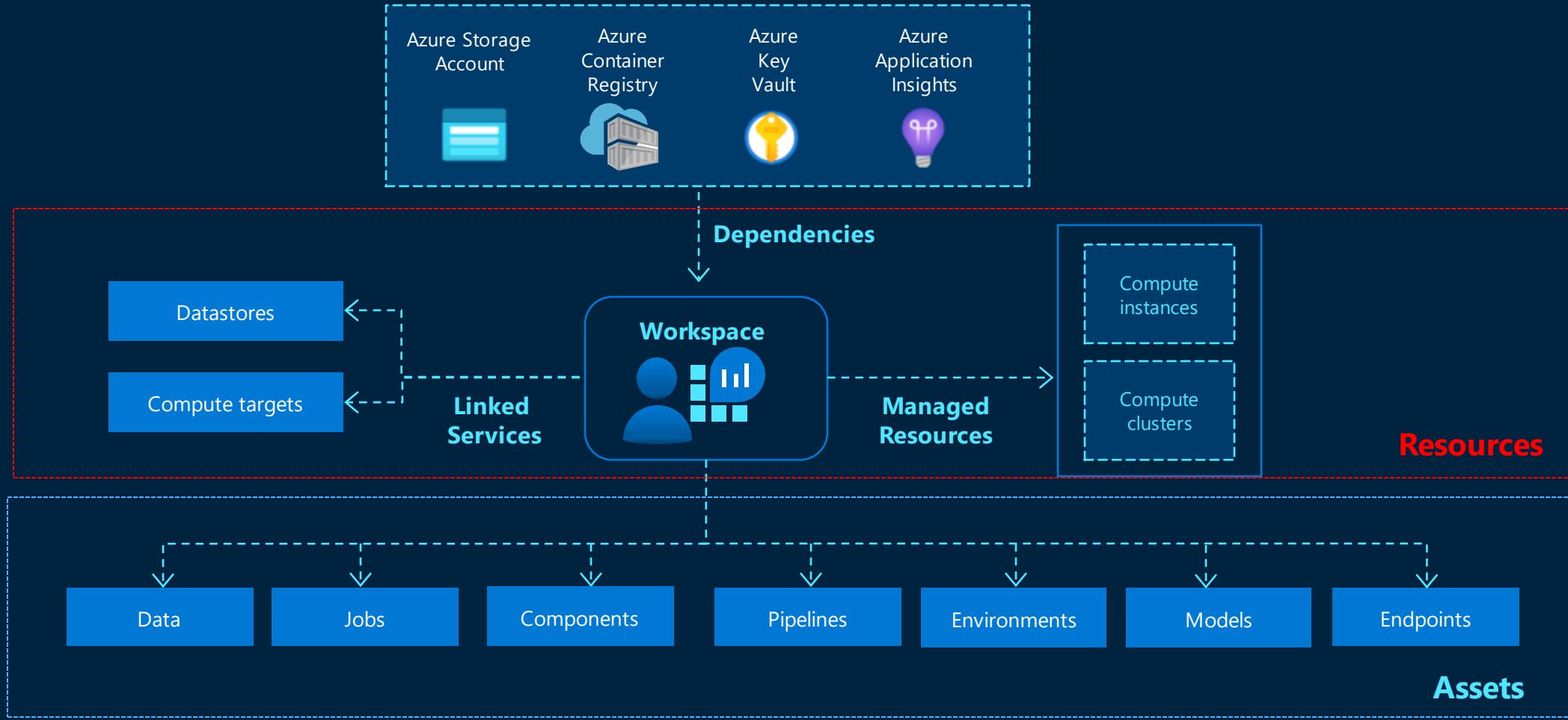
From the Intelligent Cloud to the Intelligent Edge





Azure Machine Learning: Technical Details

Key Elements of Azure Machine Learning



Azure ML service

Key Artifacts



Workspace

Central resource used to register all other resources, and to track the history of your training job executions.



Models



Model registry



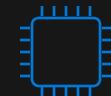
Jobs



Pipelines



Compute Instance



Compute Target



Images



Images registry



Deployment / Endpoints



Data / Datastores

Azure ML service

Key Artifacts



Workspace



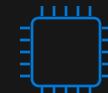
Models



Models registry



Jobs



Compute Target



Images



Images registry



Deployment / Endpoints



Dataset / Datastores

Managed VMs for data scientists optimized for data science environments.



Compute Instance



Different types of compute

The screenshot shows the 'Compute' blade in the Azure portal. It lists four main categories of compute instances:

- Compute instances**: VM set up for running dev ML code.
- Compute clusters**: Set of VMs that can auto scale up based on traffic.
- Inference clusters**:
 - Create/attach an AKS cluster
 - ACI replaced by managed endpoints for v2
- Attached computes**: Any compute target that you manage yourself outside of AML for train/inference.

A sidebar on the right shows a list of additional compute targets:

- + New
- Azure Databricks
- Data Lake Analytics
- HDInsight
- Kubernetes (preview)
- Synapse Spark pool (preview)
- Virtual machine

Environments

Environments: Runtime, packages,
environment variables, specific settings



Prepare Data



Build & Train Models



Deploy & Predict

1. **Curated environments:** pre-created for faster deployment + provided by AML by default
2. **Custom environments:** user is responsible for set up. Can be created from:
 - A docker image
 - A base docker image with a conda YAML to customize further
 - A docker build context



Azure ML service

Key Artifacts



Workspace



Models



Model registry



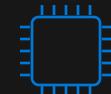
Jobs



Pipelines



Compute Instance



Compute Targets



Images



Images registry



Deployment and endpoints



Data / Datastores

Data: **Reference** to the data source location + copy of its **metadata**.

Datastores: used for storing connection information to Azure storage services

Datastores



Microsoft Azure Machine Learning Studio

Microsoft > nm_data_ai > Datastores

Datastores

+ New Microsoft

Search

Showing 1-6 of 6 datastores

Name	Type	Storage name	Created on
azureml	Azure Blob Storage	nmdataai4280352326	Mar 19, 2022
sample_datastore	Azure Blob Storage	nmdataai4280352326	Mar 15, 2022
workspaceworkingdirectory	Azure file share	nmdataai4280352326	Mar 11, 2022
workspaceartifactstore	Azure Blob Storage	nmdataai4280352326	Mar 11, 2022
workspacefilestore	Azure file share		Mar 11, 2022
workspaceblobstore (Default)	Azure Blob Storage		Mar 11, 2022

New datastore Refresh Unregister Set as default datastore Edit columns Reset view

All filters Clear all

workspaceblobstore (Default) ☆

Overview Browse (preview)

Create dataset Refresh Update authentication Set as default datastore

General

Datastore name workspaceblobstore
Datastore type Azure Blob Storage
Created by Service Principal
Subscription ID 77c6f1ae-da9d-4379-b3cd-5a41778a26c5
Resource group name nm_data_ai
Protocol https
Endpoint core.windows.net
Account name nmdataai4280352326
Blob container azureml-blobstore-cca1b977-2339-414a-994f-9dcf0635d6cc
Data URL https://nmdataai4280352326.blob.core.windows.net/azureml-blobstore-cca1b977-2339-414a-994f-9dcf0635d6cc
Created on Mar 11, 2022 12:50 PM

Datastore type *

- Azure Blob Storage
- Azure Blob Storage
- Azure file share
- Azure Data Lake Storage Gen1
- Azure Data Lake Storage Gen2
- Azure SQL database
- Azure PostgreSQL database
- Azure MySQL database

Authentications Allowed workspace managed identity access Yes Authentication type Account key

Data

Microsoft Azure Machine Learning Studio

Microsoft > nm_data_ai > Data

Data

Registered data assets Dataset monitors (preview)

+ Create Refresh Unregister Edit columns Reset view

Search All filters Clear all

Showing 1-12 of 12 data assets

Page size: 25

Name	Version	Data source	Created on	Modified on	Data type (new)	Properties	Created by	Tags
heart-data	1	workspaceblobstore	Jul 20, 2022 5:20 PM	Jul 20, 2022 5:20 PM	mltable	Tabular	Narjes Majdoub	
diabetes target	1	workspaceblobstore	Apr 25, 2022 4:51 PM	Apr 25, 2022 4:51 PM	mltable	Tabular, Ti...	Narjes Majdoub	format: CSV
diabetes baseline	1	workspaceblobstore	Apr 25, 2022 4:51 PM	Apr 25, 2022 4:51 PM	mltable	Tabular	Narjes Majdoub	format: CSV
batch-data	1	workspaceblobstore	Apr 24, 2022 3:12 PM	Apr 24, 2022 3:12 PM	mltable	File	Narjes Majdoub	
diabetes file dataset	1	workspaceblobstore	Apr 21, 2022 1:28 PM	Apr 21, 2022 1:28 PM	mltable	File	Narjes Majdoub	format: CSV
TD-Auto_Price_Training-Clean...	1	workspaceblobstore	Mar 19, 2022 1:13 PM	Mar 19, 2022 1:13 PM	mltable	File	Narjes Majdoub	azureml.Designer : true
TD-Auto_Price_Training-Normal...	1	workspaceblobstore	Mar 19, 2022 12:54 PM	Mar 19, 2022 12:54 PM	mltable			
MD-Auto_Price_Training-Tran...	1	workspaceblobstore						
bike-rentals	1	URI						
test_werknemers_huish...	1	workspaceblobstore						

Create dataset

New version Refresh Generate profile Unregister

Preview Profile

Number of columns: 8 Number of rows: 50 (of 10000)

ID	storeID	productCode	quantity	logQuantity	advertising	price	weekStarting	ID
1	2	surface.go	105	9.265	1	159	2017-06-15 00:00:00	d6bd47a7-2ad6-4f0a-b8de-1c1386cae5ea
2	2	surface.go	80	8.987	0	269	2017-07-27 00:00:00	64cc74c2-c7da-4e12-af64-c95bd429934
3	2	surface.go	68	8.832	1	209	2017-08-03 00:00:00	9a2d164b-5e44-44d7-9837-cf9ae566c99
4	2	surface.go	28	7.966	0	209	2017-08-10 00:00:00	b6cd9987-1d5a-4f4f-9346-719d73b1f7f0
5	2	surface.go	16	7.378	0	209	2017-08-24 00:00:00	ac0e099-e102-4bf7-9775-983b151dc0d3
6	2	surface.go	253	10.14	1	189	2017-08-31 00:00:00	3d22c002-b04c-4092-abc-b3f0f01c9da8f
7	2	surface.go	107	9.283	0	189	2017-09-07 00:00:00	b6e19699-d684-449e-9c98-c19d5288bc7b
8	2	surface.go	66	8.803	0	189	2017-09-14 00:00:00	e89a5838-fb8f-413a-a23d-4d000c4b5282

From local files

From datastore

From web files

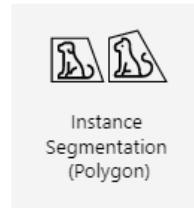
From Open Datasets



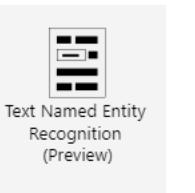
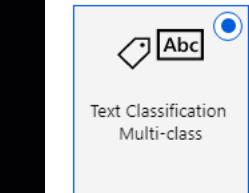
Media type *

Image Text

Labeling task type *



Labeling task type *



Apply only a single label from a set of classes to a piece of text

Azure ML service

Key Artifacts



Workspace



Models



Models registry



Jobs



Pipelines



Compute Instance

Grouping of executions of the same script. Stores metadata for an execution. There are currently four types of jobs:

Command Job

Pipeline Job

Sweep Job

AutoML Job



Compute Target



Images



Images registry



Deployment / Endpoints



Dataset / Datastores

Execution :
Running a script to train a
model.

Contains:
Execution metadata: (date,
time, duration, etc.)
The metrics logged by the
script

The output files

A snapshot of the directory
containing the scripts, before
execution

Azure ML service

Key Artifacts



Workspace



Models



Models registry



Jobs

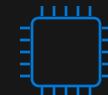


Pipelines



Compute Instance

Binary file(s) that represent a machine learning model and any corresponding metadata



Compute Target



Images



Images registry



Deployment / Endpoints



Dataset / Datastores

Azure ML service

Key Artifacts



Workspace

Grouping of all the models, their different versions and the tags added by the user (useful for research) of a workspace.



Models



Models registry



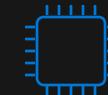
Jobs



Pipelines



Compute Instance



Compute Target



Images



Images registry



Deployment / Endpoints



Dataset / Datastores

Azure ML service

Key Artifacts

A package including a model, its scoring script dan its dependencies.



Workspace



Models



Models registry



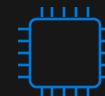
Experiments



Pipelines



Compute Instance



Computer



Images



Images registry



Deployment / Endpoints



Dataset / Datastores

Azure ML service

Key Artifacts

Grouping of images and tags added by the user (useful for research).



Workspace



Models



Models registry



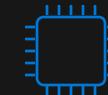
Experiments



Pipelines



Compute Instance



Compute 1



Images



Images registry



Deployment / Endpoints



Dataset / Datastores

Model Management

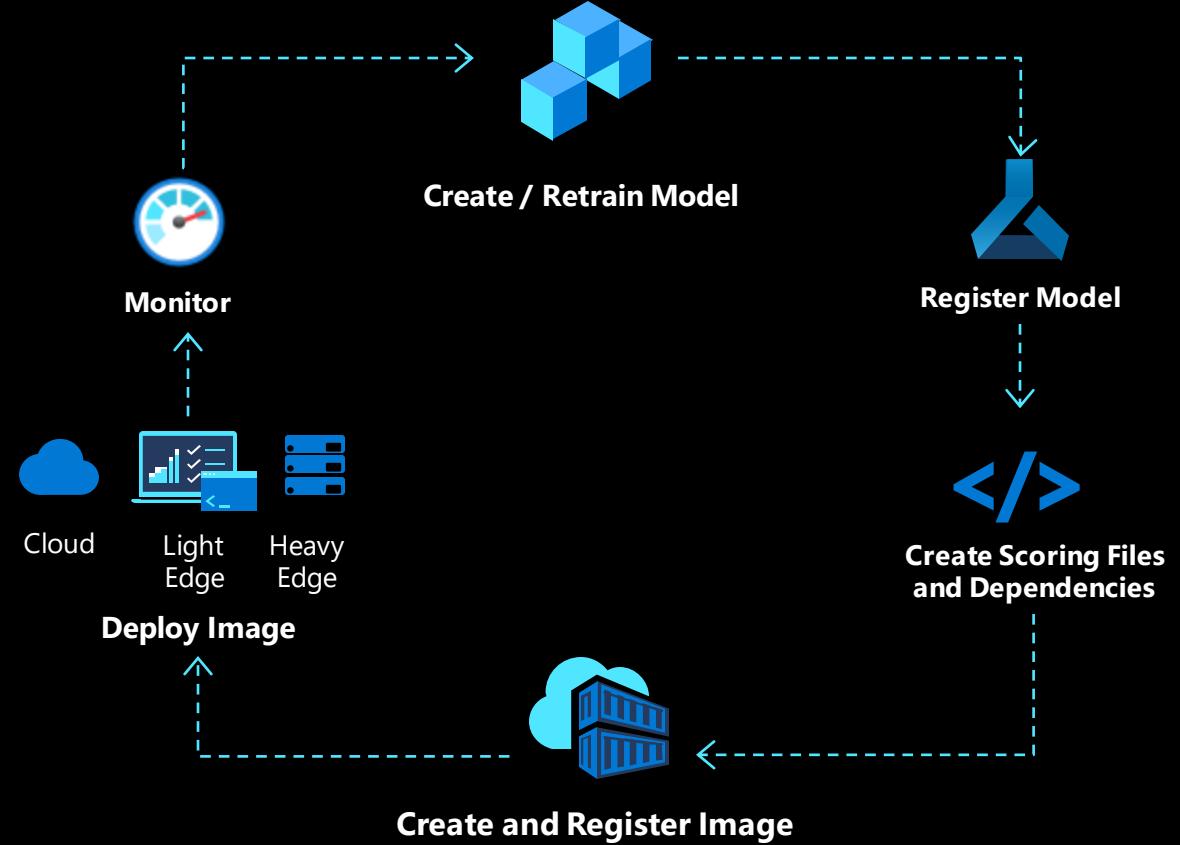
Model Management in Azure ML usually involves these four steps

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.



Azure ML service

Key Artifacts



Workspace



Models



Models registry



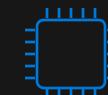
Experiments



Pipelines



Compute Instance



Compute Ta



Images



Images registry



Deployment / Endpoints

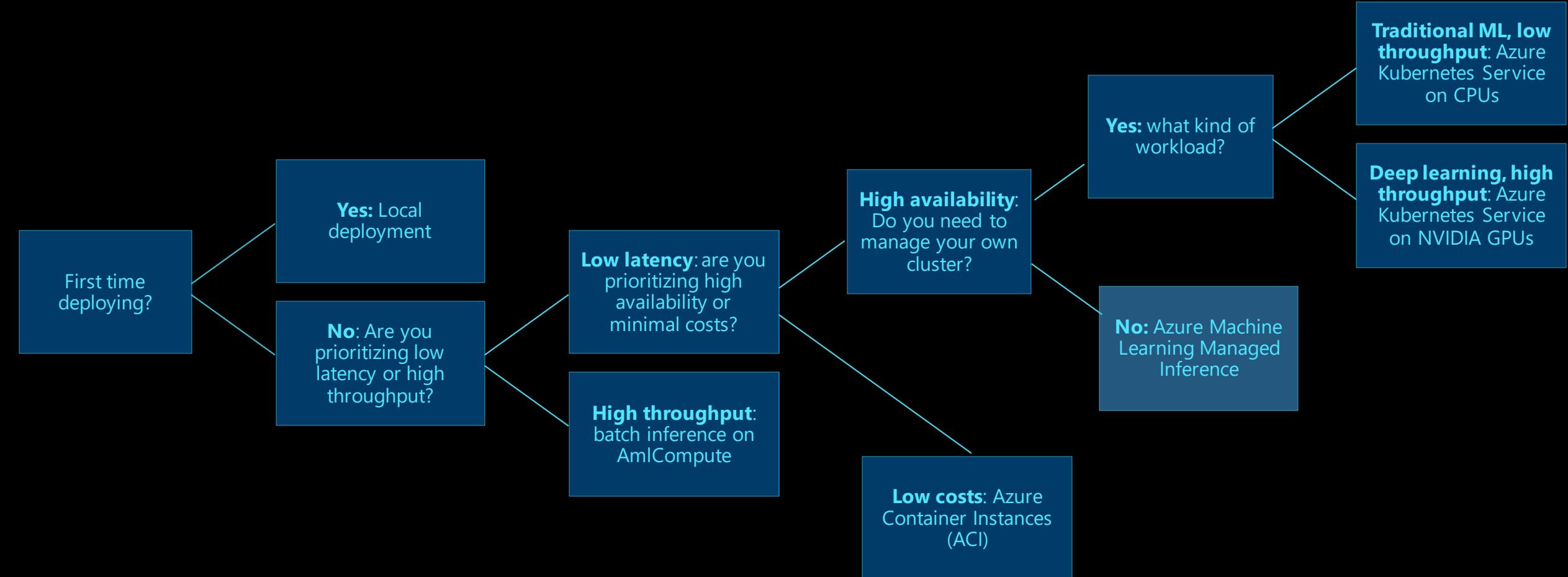


Dataset / Datastores

Instantiation of an image either in Web service (ACI, AKS or FPGA) or via an IoT Module (Docker container).

We need to specify:
• Model assets
• Scoring script
• Environment
• Compute size + scale settings

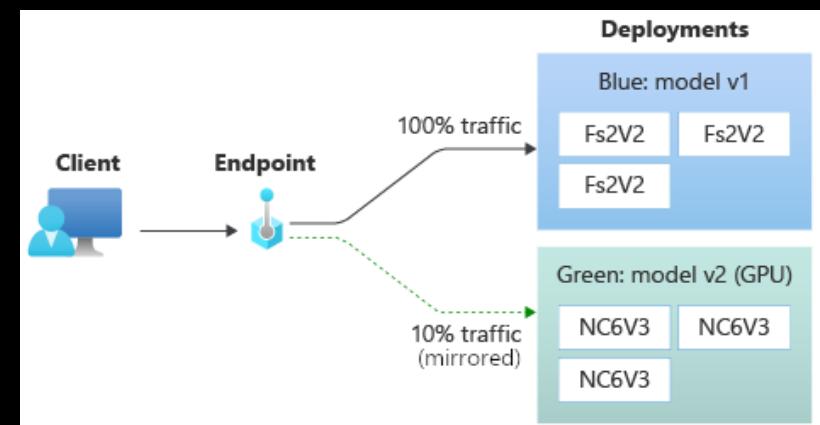
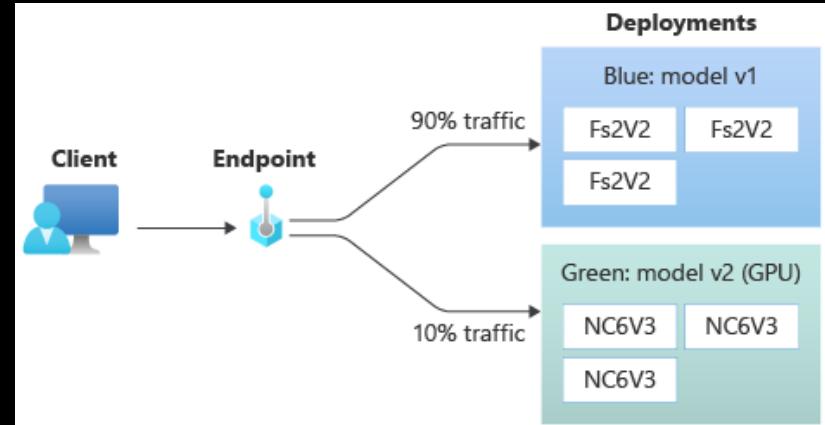
Choosing an inferencing target



Managed Endpoints - GA

- Azure Machine Learning **managed endpoints**, now generally available, help developers and data scientists more easily deploy large-scale machine learning models for both real-time and batch inferencing.

	Managed online endpoints	Kubernetes online endpoints
Recommended users	Users who want a managed model deployment and enhanced MLOps experience	Users who prefer Kubernetes and can self-manage infrastructure requirements
Infrastructure management	Managed compute provisioning, scaling, host OS image updates, and security hardening	User responsibility
Compute type	Managed (AmlCompute)	Kubernetes cluster (Kubernetes)
Out-of-box monitoring	Azure Monitoring (includes key metrics like latency and throughput)	Supported
Out-of-box logging	Azure Logs and Log Analytics at endpoint level	Unsupported
Application Insights	Supported	Supported
Managed identity	Supported	Supported
Virtual Network (VNET)	Supported (preview)	Supported
View costs	Endpoint and deployment level	Cluster level
Mirrored traffic	Supported	Unsupported



Azure ML service

Key Artifacts



Workspace



Models



Models registry



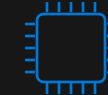
Jobs



Pipelines



Compute Instance



Compute Target



Images



Images registry



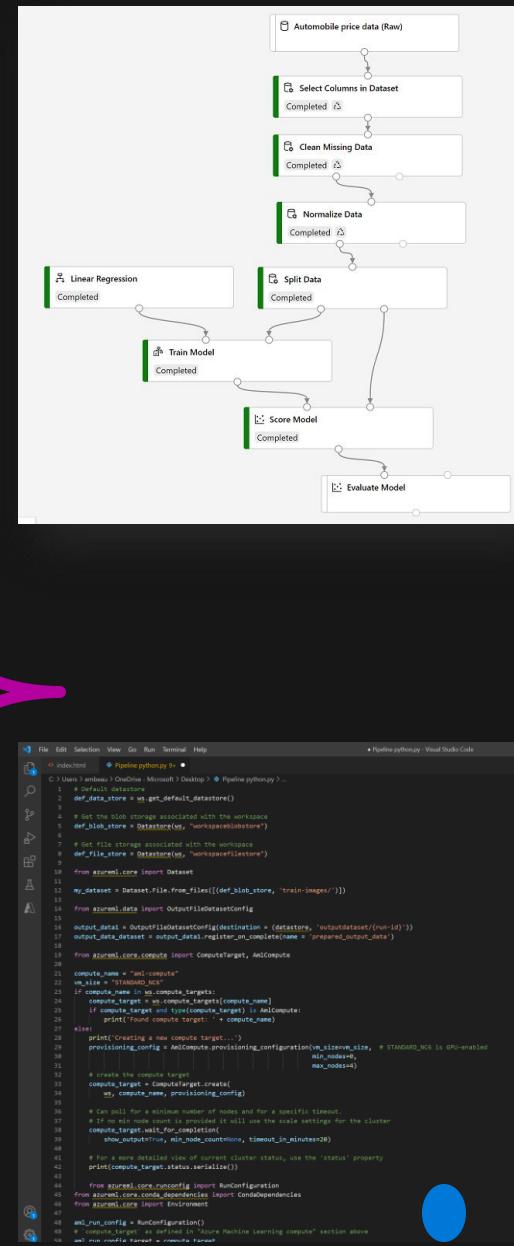
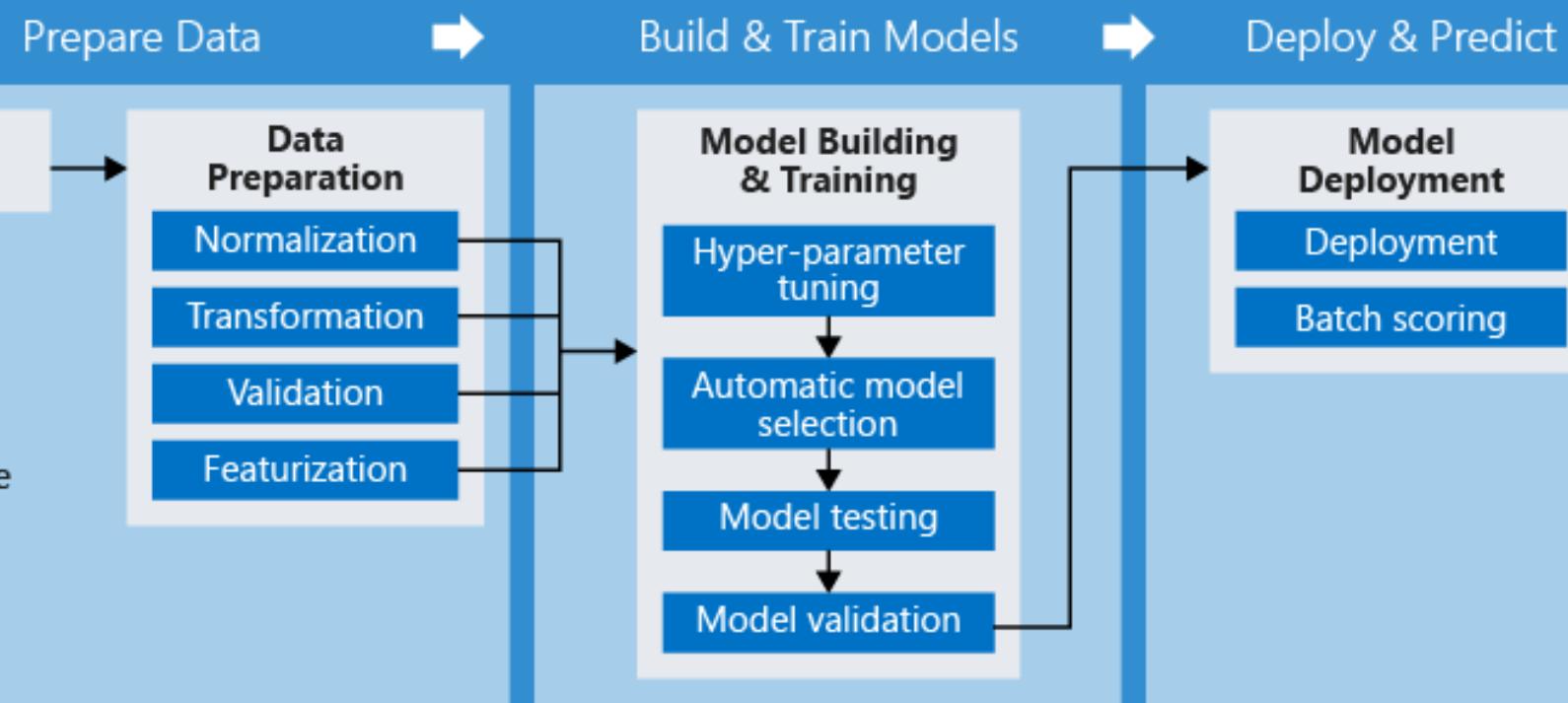
Deployment / Endpoints



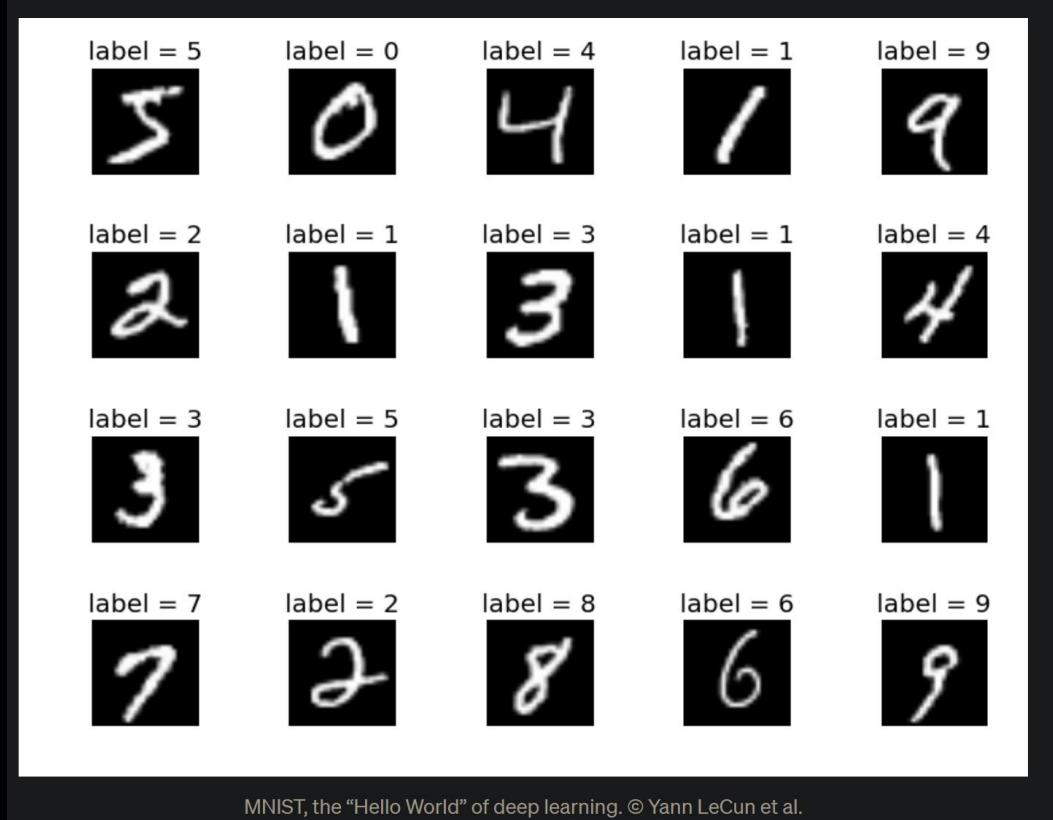
Dataset / Datastores

A workflow composed of independent steps linked together.
Each step is an event that can be executed (data prep, training
model, model evaluation, deployment etc...)

Azure ML Artifact Pipeline



[Demo] Train and deploy a model with AML



- 70.000 handwritten gray scale images of 28*28 pixels
- Goal: Create a multi-class classifier to identify the digit a given image represents.

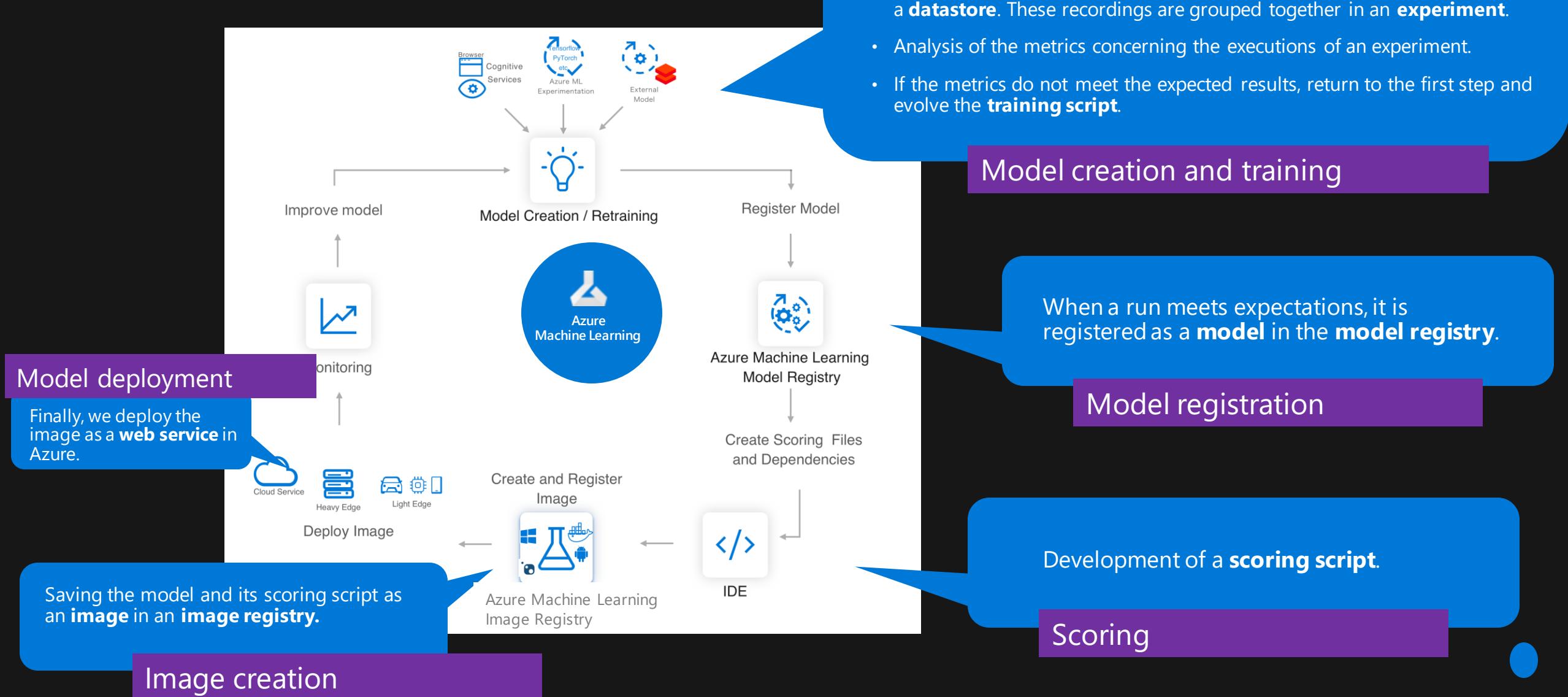


TensorFlow

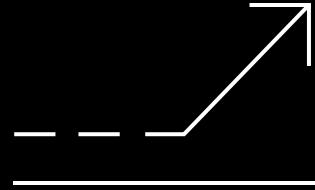


Azure ML service

AI/ML Lifecycle

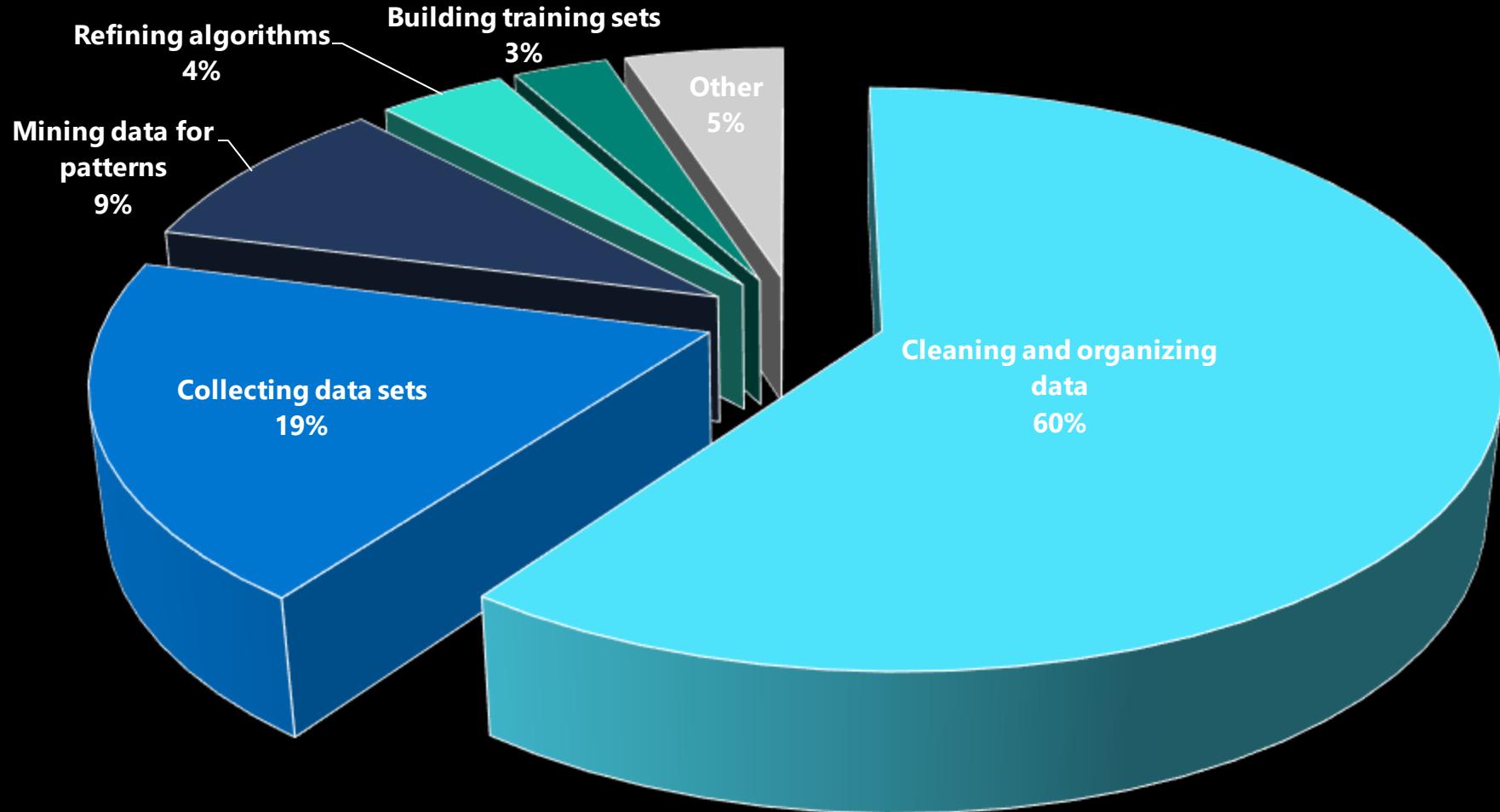


Automated Machine learning



Automated Machine Learning
'simplifies' the creation and selection
of the optimal model

What data scientist spend the most time doing?



Source : Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says (forbes.com)

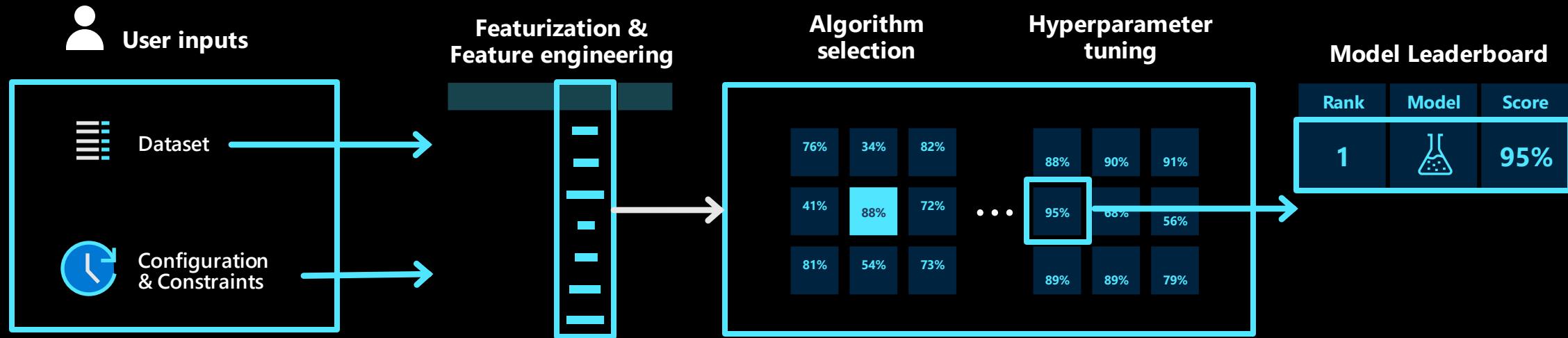


What is Automated Machine Learning?

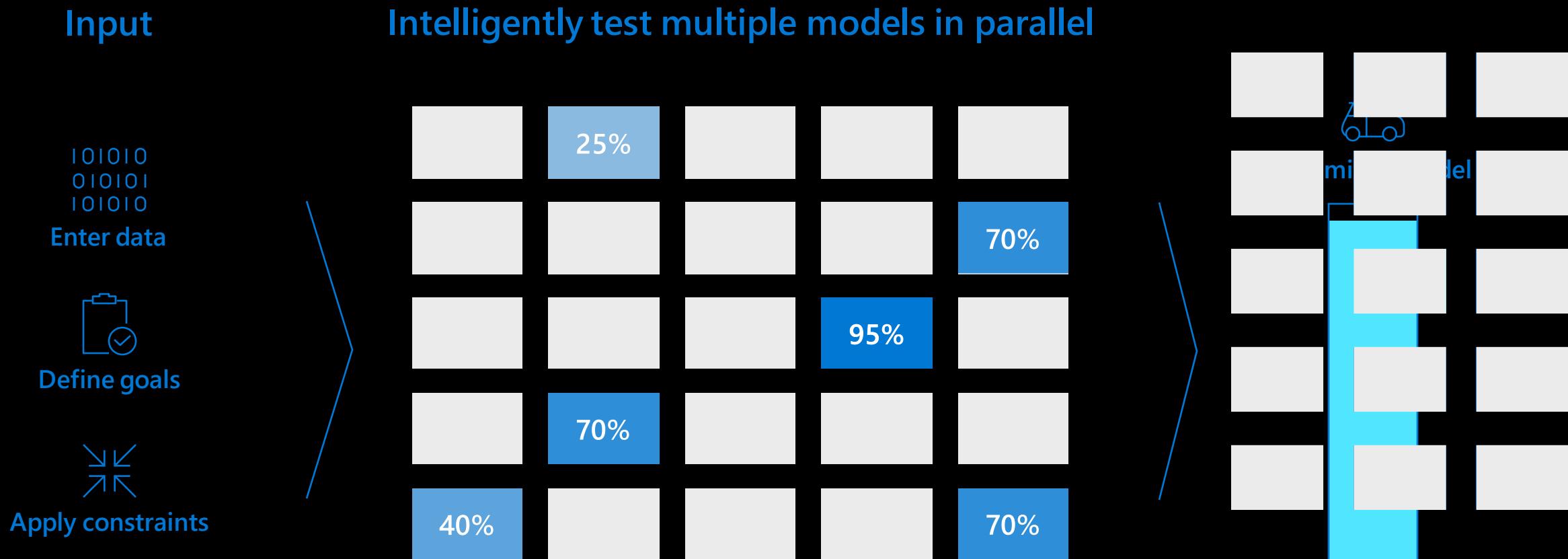
Automated machine learning (automated ML) automates feature engineering,

algorithm and hyperparameter selection to find the 'best model' for your data.

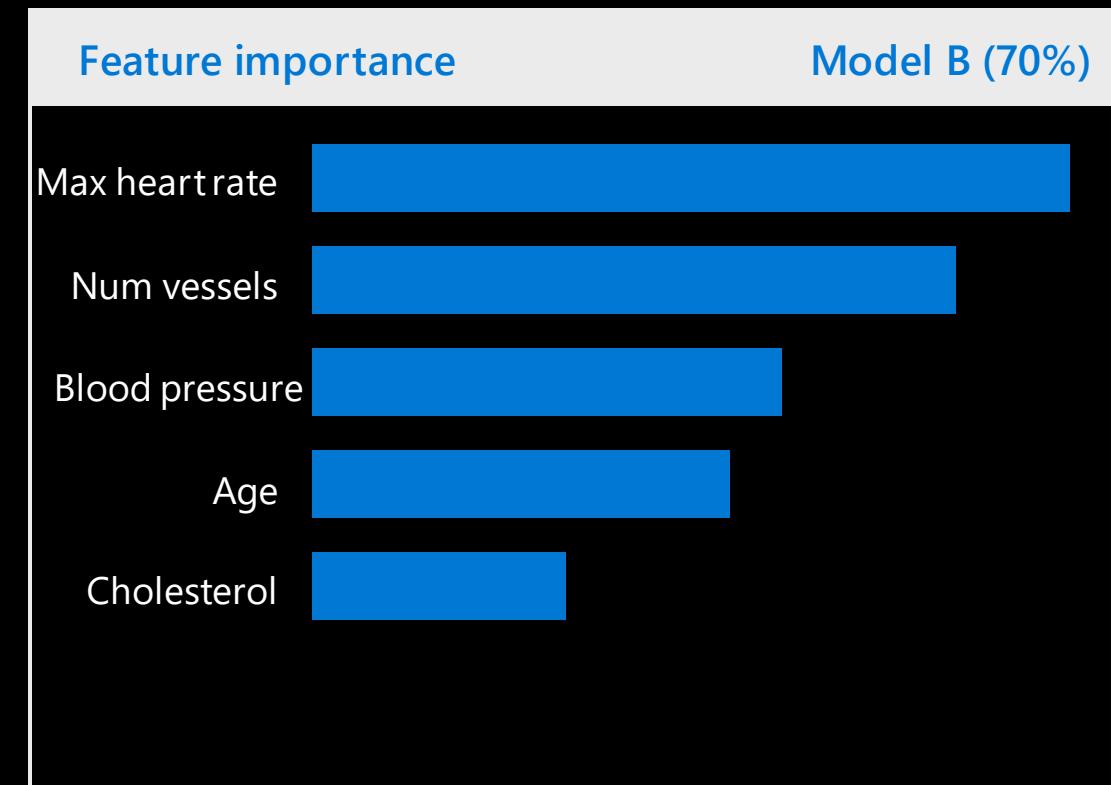
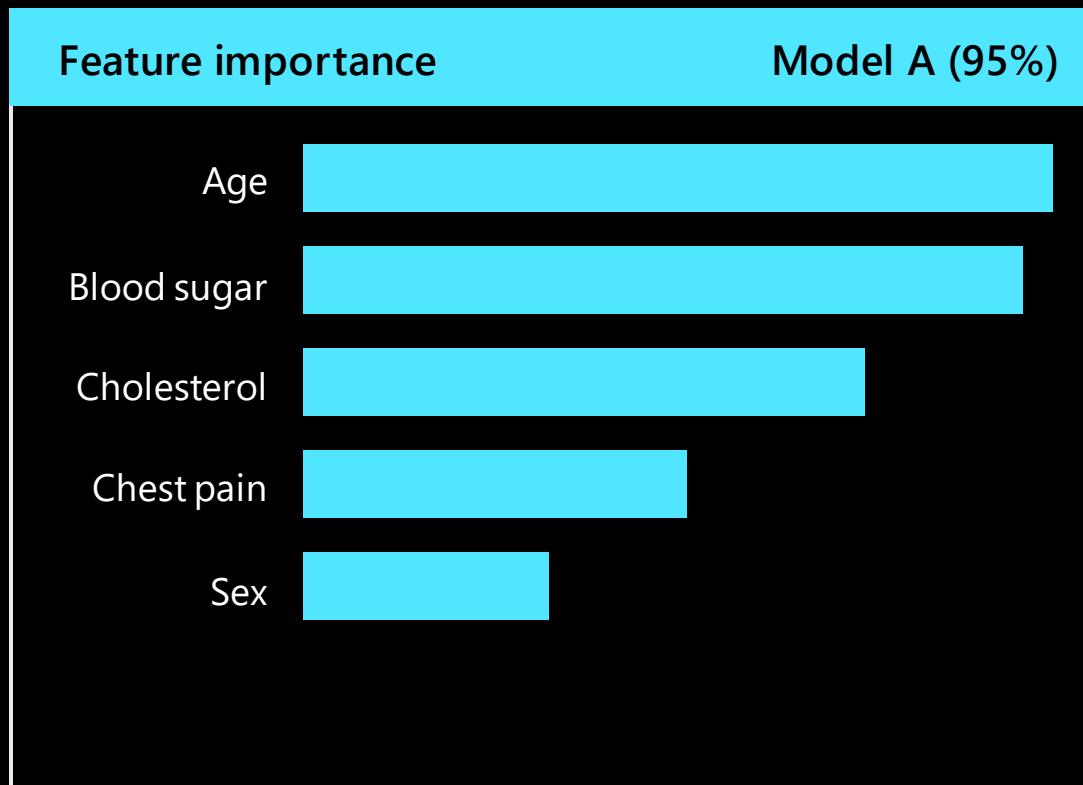
Loop until reaching **exit criteria**



Azure Machine Learning accelerates model development with automated machine learning



Azure Machine Learning accelerates model selection with model explainability



Azure Machine Learning accelerates model selection with code generation [Public Preview]

AutoML Model's training code generation (i.e. Scikit-Learn code, Transformers and Algorithms)

AutoML UI
(Experiment's model leader board)

Microsoft > cesardi-automl-centraluseuap-ws > Experiments > SDK_Codgen_remote_porto_seguro

CodeGen_Internal_SDK_remote_porto_seguro_1 🎒 ⭐

Refresh Edit and submit Cancel Delete

Details Data guardrails Models Outputs + logs Child runs Snapshot

Refresh Deploy Download Explain model # View generated code (preview) 2)

Search

Showing 1-10 of 10 models

Algorithm name	Explained	AUC weighted ↓
VotingEnsemble	View explanation	0.63816
MaxAbsScaler, XGBoostClassifier		0.63739
MaxAbsScaler, LightGBM		0.63594
TruncatedSVDWrapper, XGBoostClassifier		0.62862
MaxAbsScaler, LightGBM		0.62337
StandardScalerWrapper, XGBoostClassifier		0.62047

script.py with model's training generated code (Scikit-Learn, etc.)

```
def generate_algorithm_config():
    from xgboost.sklearn import XGBClassifier

    algorithm = XGBClassifier(
        base_score=0.5,
        booster='gbtree',
        colsample_bylevel=1,
        colsample_bynode=1,
        colsample_bytree=1,
        gamma=0,
        learning_rate=0.1,
        max_delta_step=0,
        max_depth=3,
        min_child_weight=1,
        missing=None,
        n_estimators=100)
    return algorithm

#...
def build_model_pipeline():
    from sklearn.pipeline import Pipeline
    pipeline = Pipeline(
        steps=[
            ('featurization', generate_data_transformation_config()),
            ('preproc', generate_preprocessor_config()),
            ('model', generate_algorithm_config()),])
    return pipeline
#...

model_pipeline = build_model_pipeline()

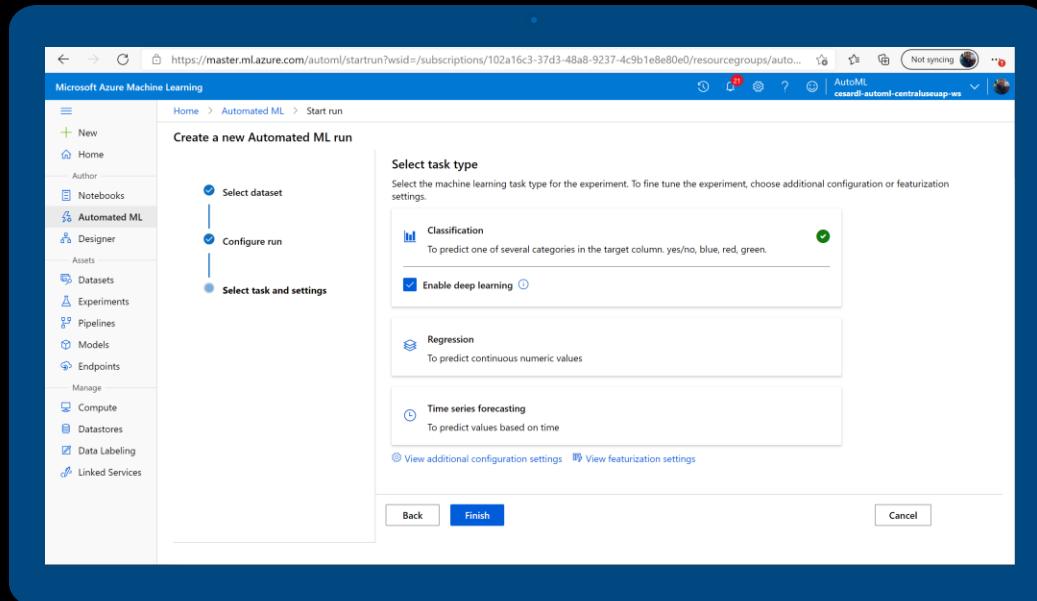
model = model_pipeline.fit(X, y)
```

- Data preprocessing
- Algorithm selection
- Featurization
- Hyperparameters

AutoML

UI and Python SDK

Automated ML UI in Azure portal



**End-to-end no-code experience for non-data scientists
to train ML models**

Easy to get started

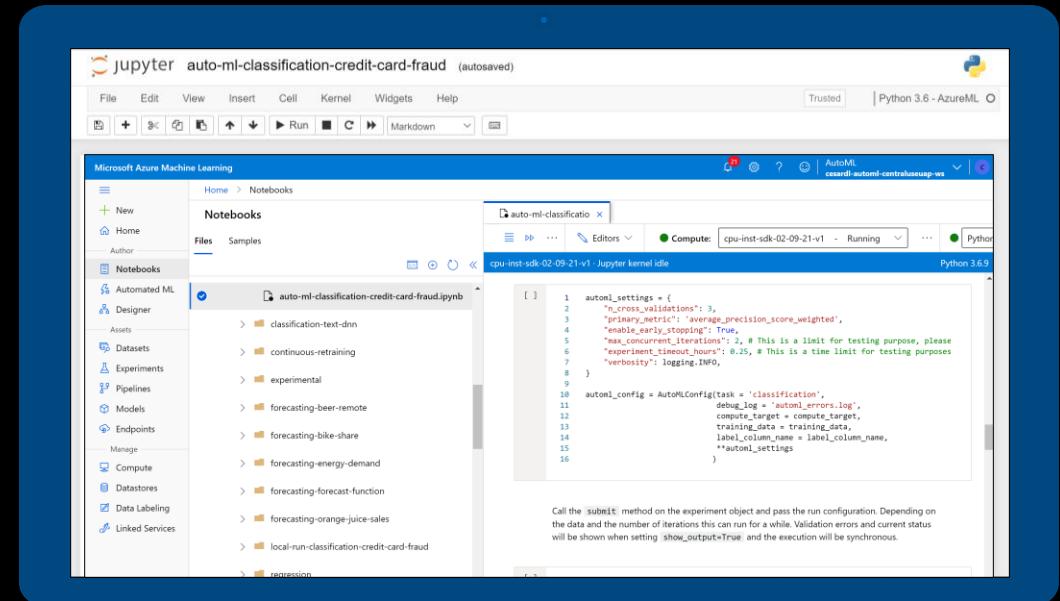
Deploy models easily and quickly

Most common scenarios supported

UI also covers results from SDK runs:

- Model metrics
- Model explainability

Python SDK (Notebooks)



Full control

Maximum flexibility based on Python
Supports all scenarios and parameters
Sample notebooks for SDK 1.x:

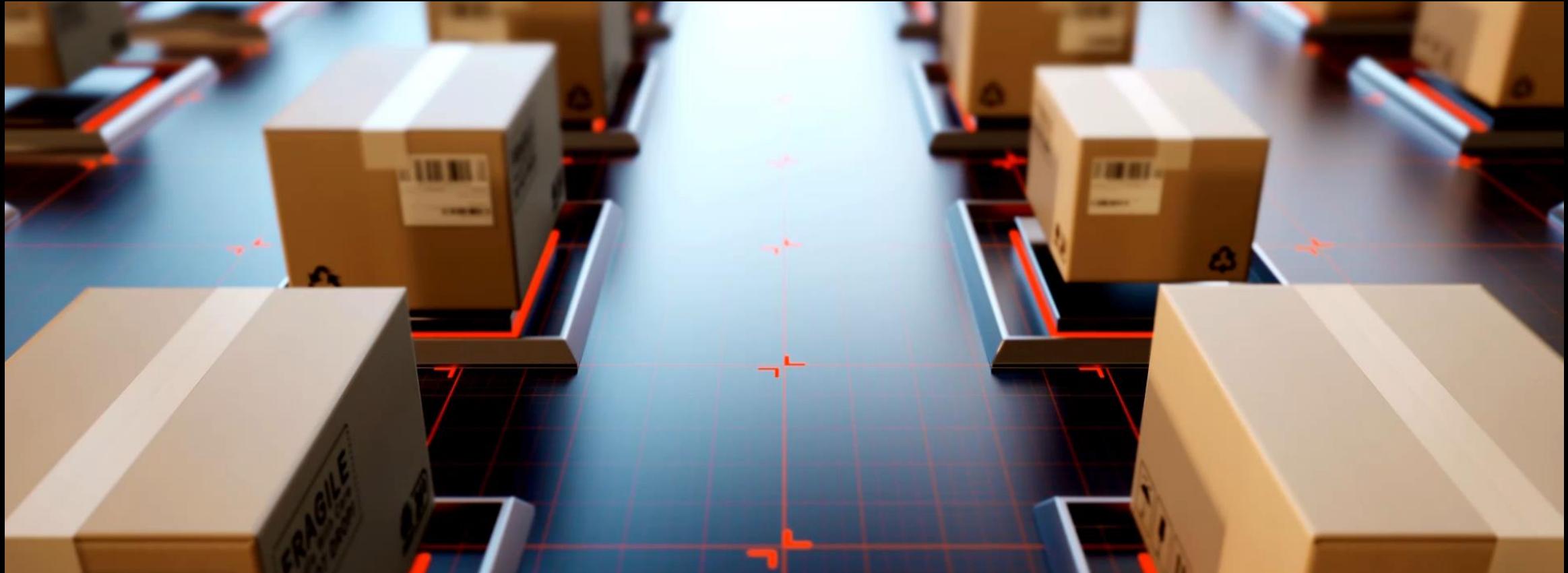
<https://github.com/Azure/MachineLearningNotebooks/tree/master/how-to-use-azureml/automated-machine-learning>

Automated ML

Current Capabilities

Category	Value
ML Problem Spaces	Classification Regression Forecasting Computer Vision (Preview) NLP (Preview)
Frameworks	Scikit Learn
Languages	Python
Data Type and Data Formats	Numerical Text Scikit-learn supported data formats (Numpy, Pandas)
Data sources	Local Files, Azure Blob Storage
Compute Target	Automated Hyperparameter Tuning Azure ML Compute (Batch AI), Azure Databricks Automated Model Selection Local Compute, Azure ML Compute (Batch AI), Azure Databricks

Automated ML Demo



Use case



10M deaths by 2030
[WHO]



Counseling for smoking
cessation ineffective and
time-consuming.

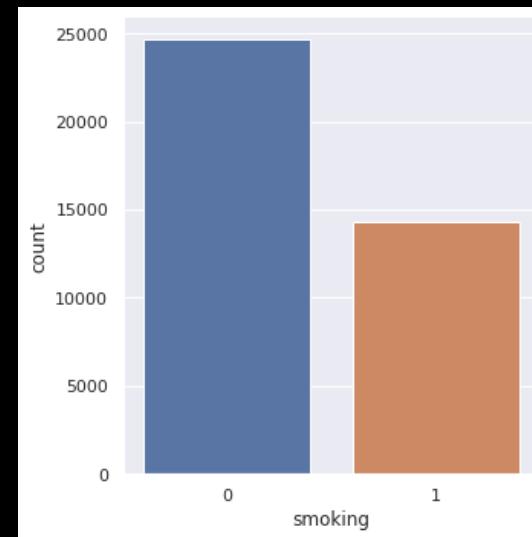
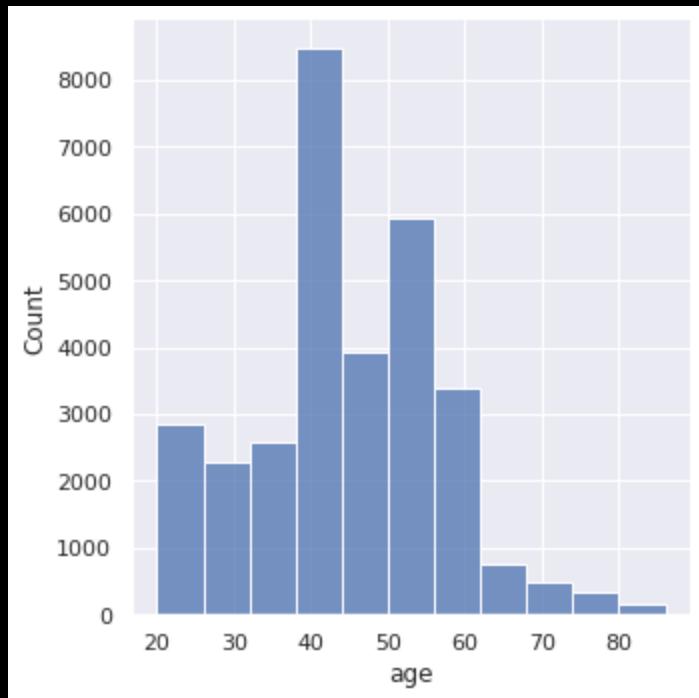


Only 1/3 quit smoking



Use case

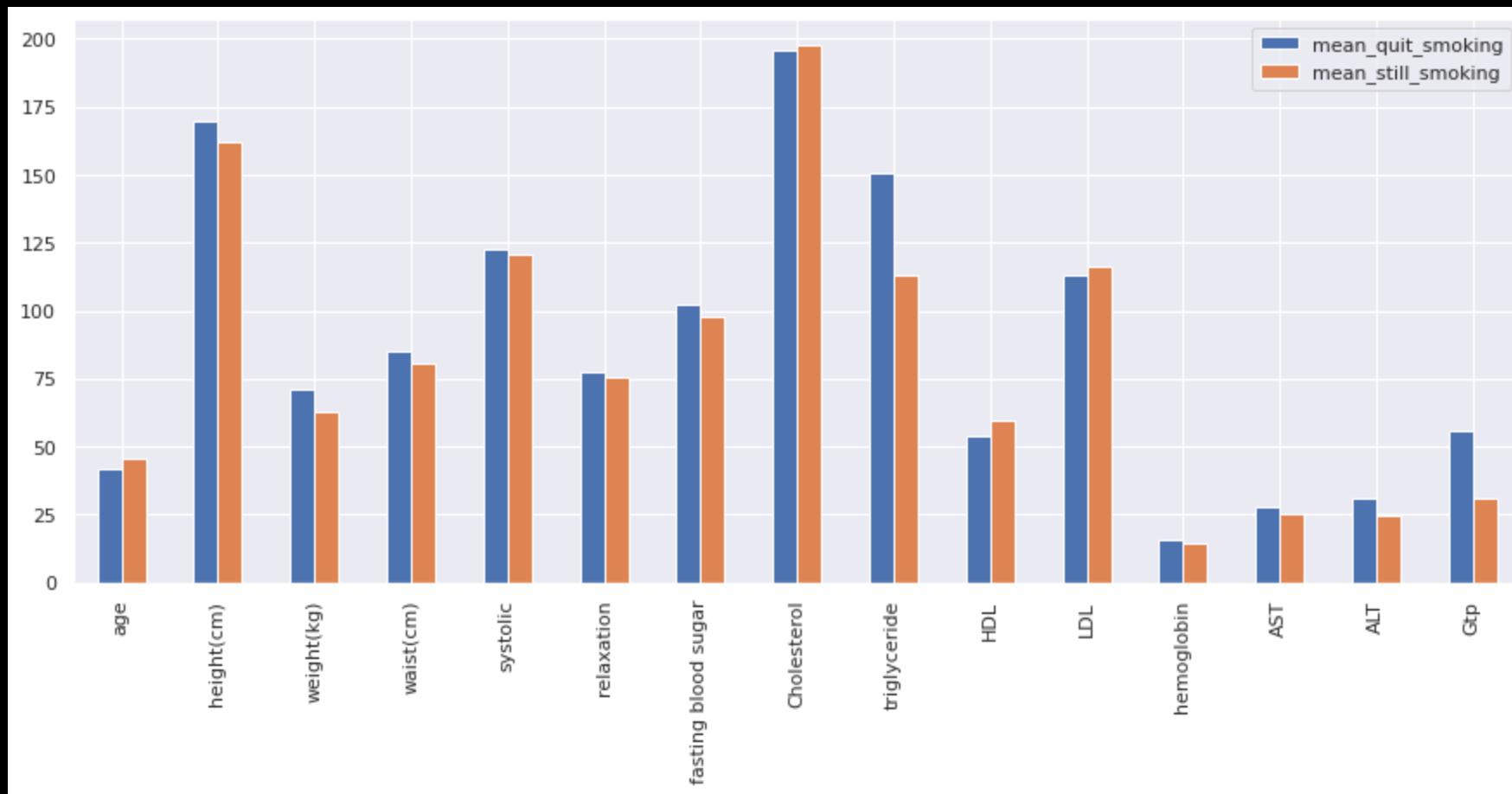
- Data for 40k patients + 23 columns bio-signals



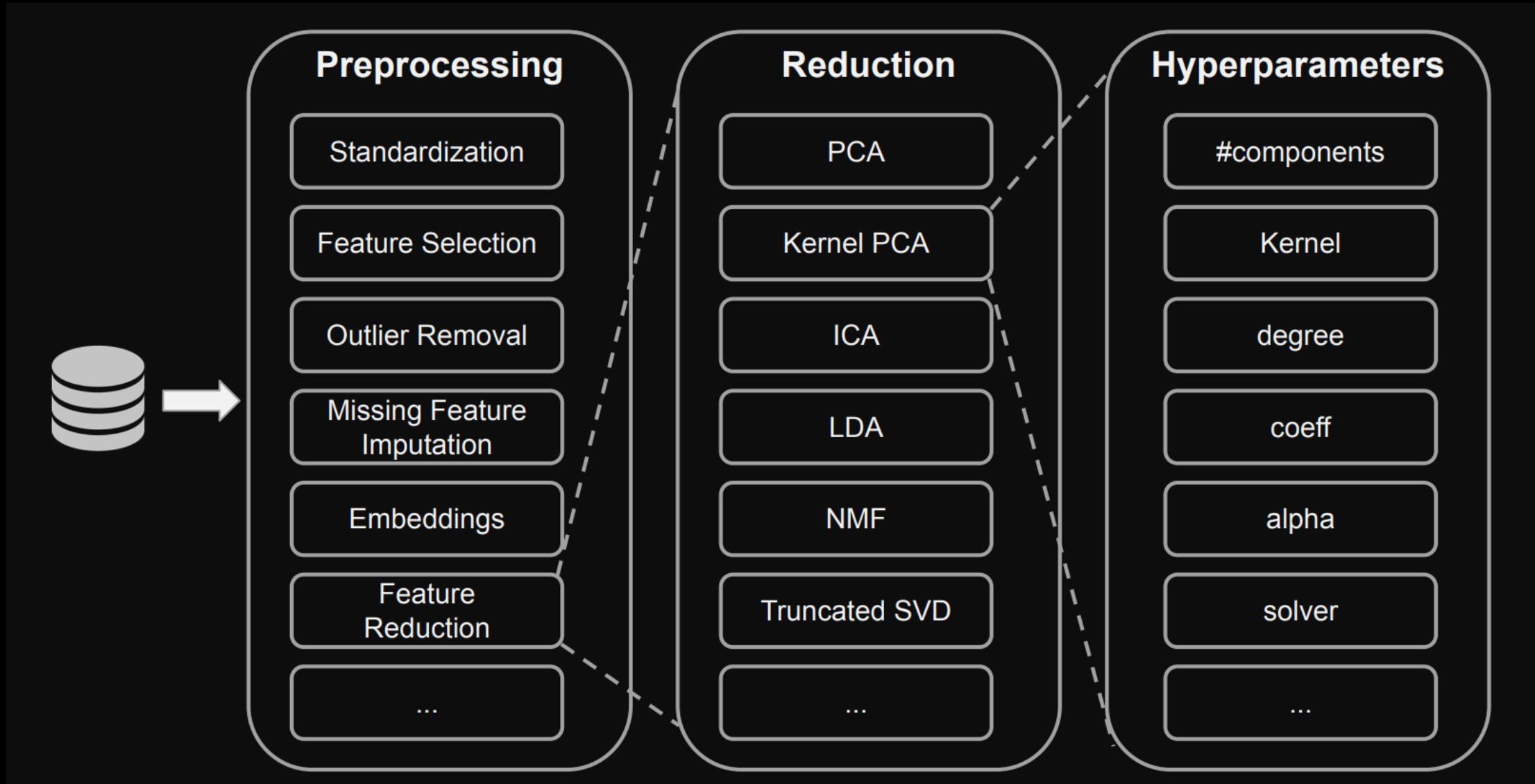
Goal: Identify the smoking status of individuals using bio-signals

Use case

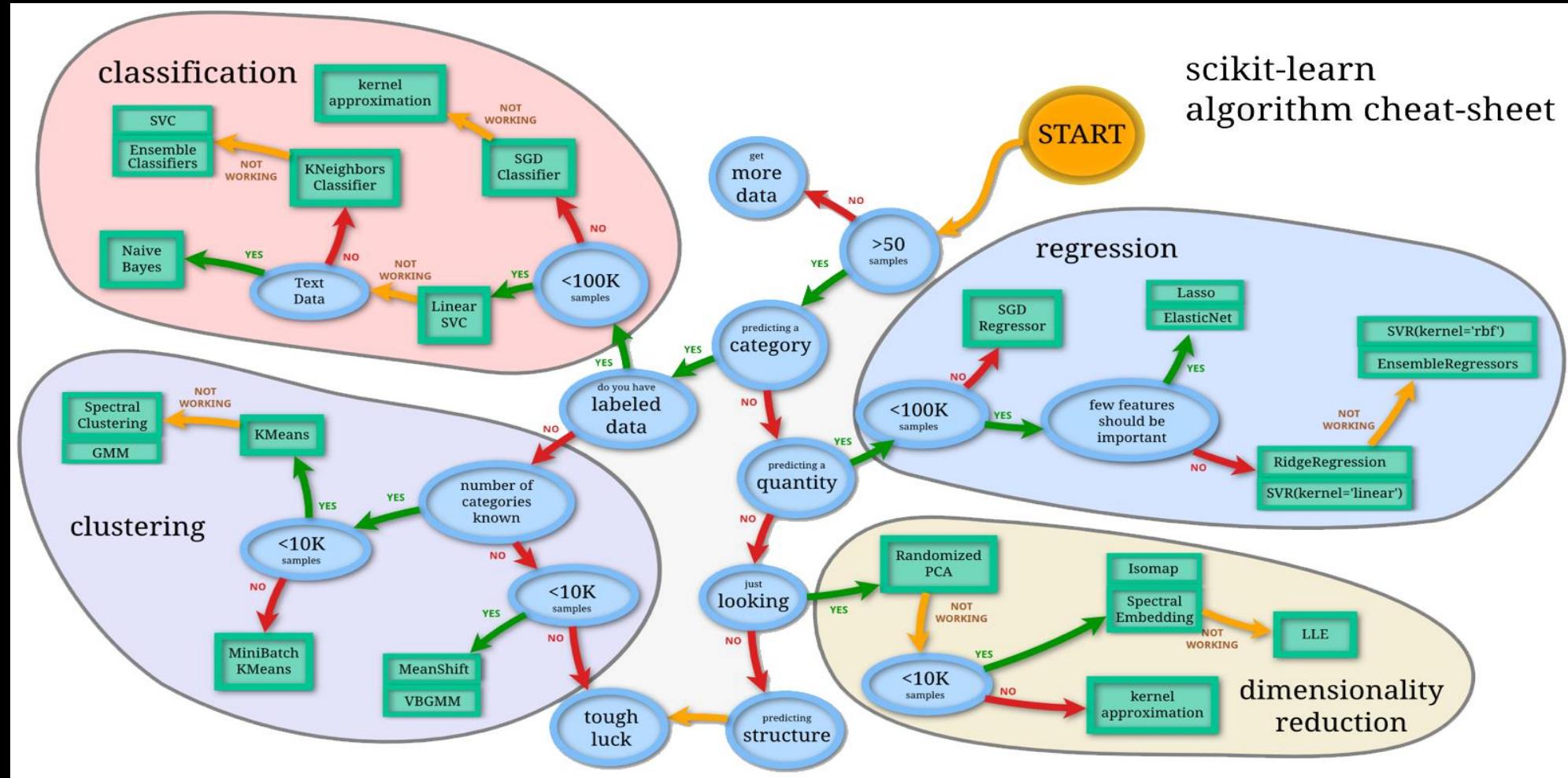
- Data for 40k patients + 23 columns bio-signals
- 36% quit – 63% didn't



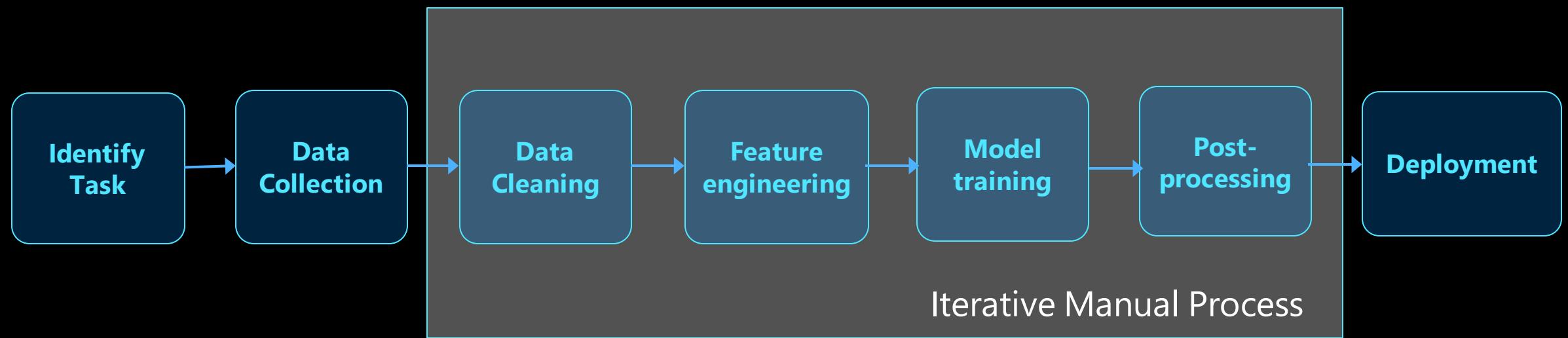
Preprocessing?



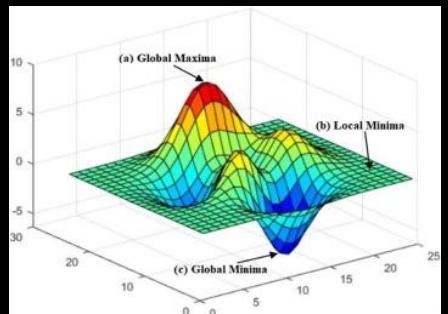
Choice of algorithm?



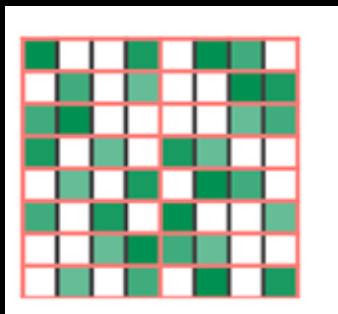
Challenges in Designing ML pipelines



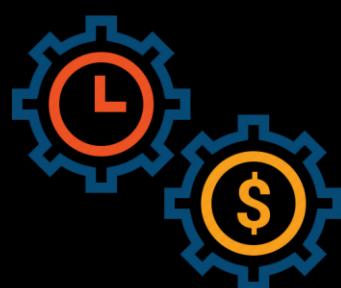
Complex search space



Sparsity of good configurations



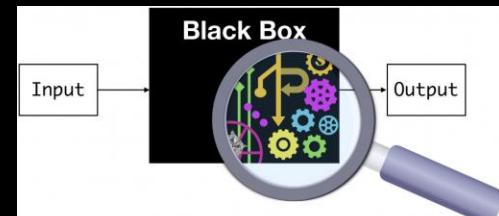
Expensive evaluations



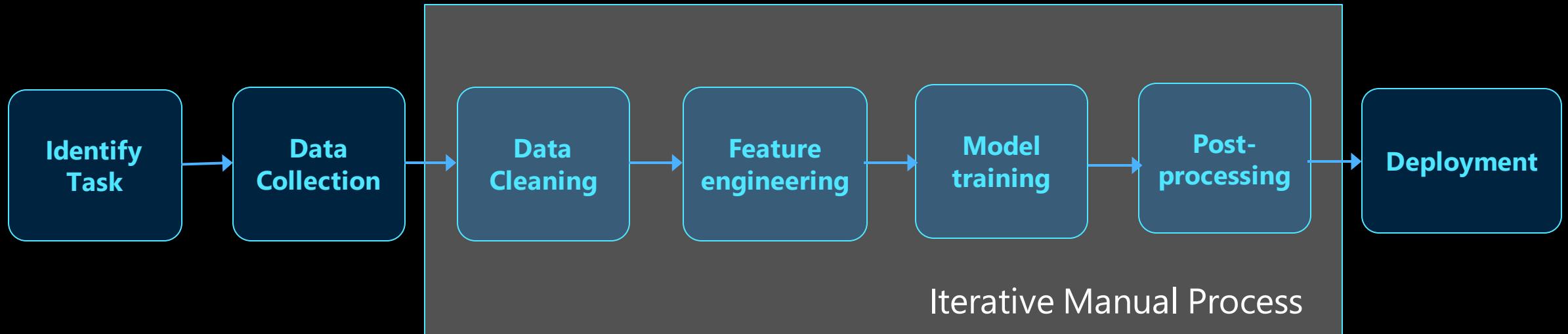
Noise on observations



Black-Box Problem



From Manual to Automated ML



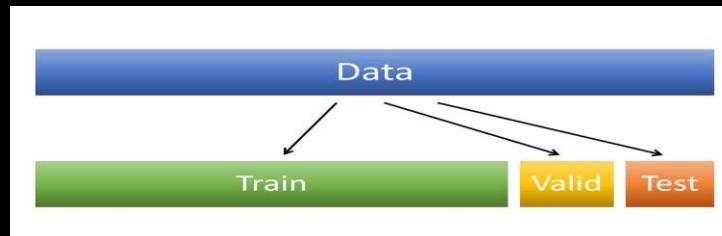
AutoML Demo

1- Create compute resources

Remote ML compute cluster/ Databricks/ local compute



2- Create data assets



- [MLTable](#) format :

Use case = Data schema changes & Data size increases

Creator defines schema in 1 file

```
type: mltable

paths:
  - pattern: ./*_use_this.txt

traits:
  - index_columns: store_location

transformations:
  - read_delimited:
    | encoding: ascii
    | header: all_files_same_headers
    | delimiter: " "
  - keep_columns: ["store_location", "zip_code", "date", "amount", "x", "y", "z"]
  - convert_column_types:
    | columns: ["x", "y", "z"]
    | to_type: boolean
    | columns: "date"
    | to_type: datetime
```

Consumer materialize into a dataframe

```
import mltable

tbl = mltable.load("./my_data")
df = tbl.to_pandas_dataframe()
```

AutoML Demo

1- Create compute resources

2- Create data assets

3- Create configuration file

4- Submit & track experiment

```
! cli-automl-smoker-status-prediction.yml •
cli > jobs > automl-standalone-jobs > cli-automl-smoker-status-prediction > ! cli-automl-smoker-status-prediction.yml > {} training_data > type
    AutoMLJob specification (autoMLJob.schema.json)
    $schema: https://azurerm-sdk2.blob.core.windows.net/preview/0.0.1/autoMLJob.schema.json
    type: automl
    ...
    experiment_name: dpv2-cli-automl-classifier-experiment
    description: A Classification job to predict smoker status
    ...
    compute: azureml:cpu-cluster
    ...
    task: classification
    log_verbose: debug
    primary_metric: average_precision_score_weighted
    ...
    target_column_name: "smoking"
    ...
    validation_data_size: 0.20
    n_cross_validations: 5
    test_data_size: 0.1
    ...
    training_data:
        path: "./training-mhtable-folder"
        type: mhtable
    ...
    limits:
        timeout_minutes: 15
        max_trials: 10
        max_concurrent_trials: 7
        trial_timeout_minutes: 5
        enable_early_termination: true
        exit_score: 0.9
    ...
    featurization:
        mode: auto
    ...
    training:
        enable_model_explainability: true
        allowed_training_algorithms:
            - gradient_boosting
            - logistic_regression
    ...

```

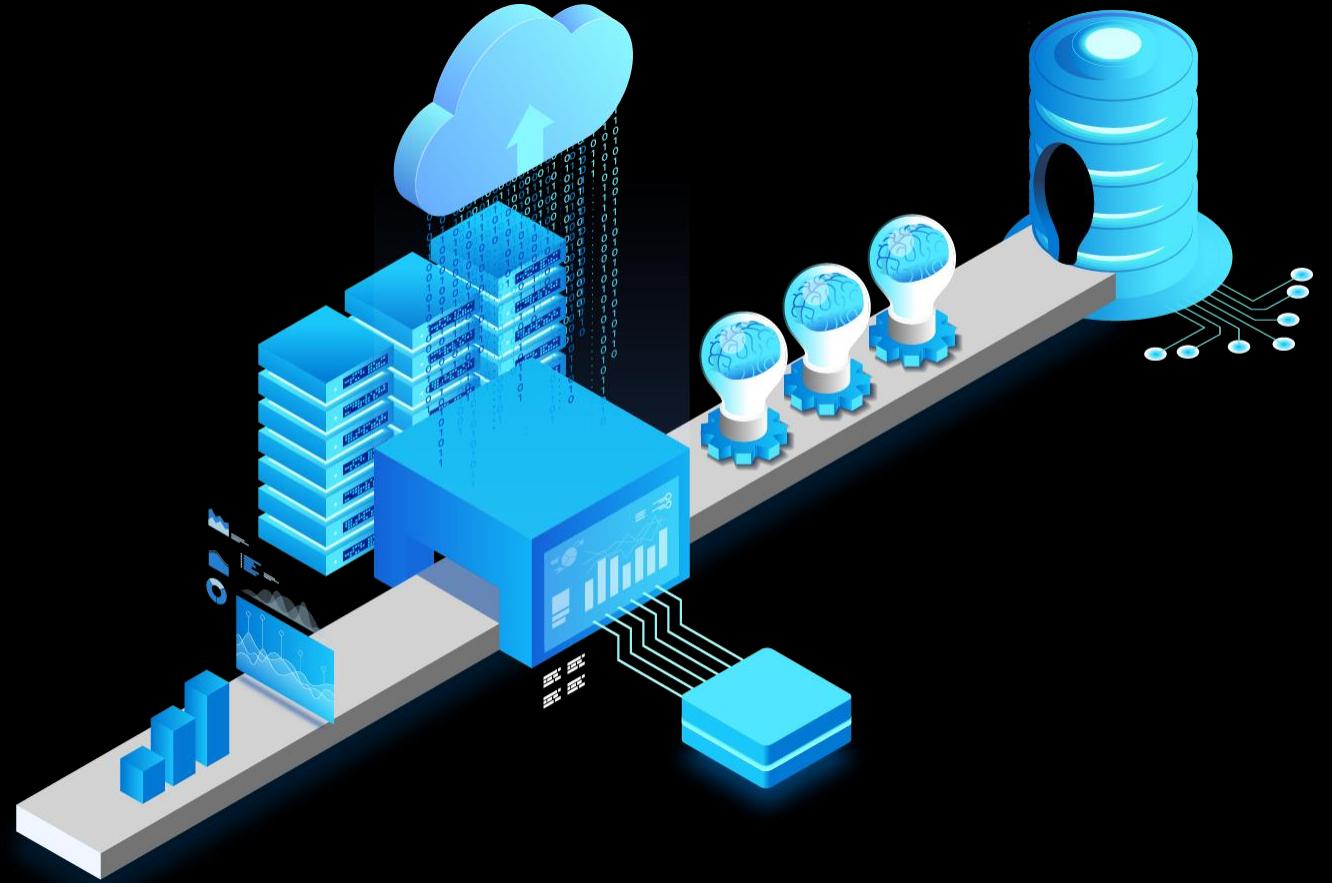
Compute target

Task / primary metric specification

Data set path

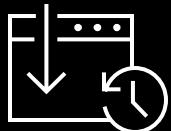
**18 mins
0.89 AUC weighted**

Azure ML and MLOps



MLOps == How to bring ML to production

Bring together **people**, **process**, and **platform** to automate ML-infused software delivery & provide continuous value to our users.



People

- Blend together the work of individual engineers in a repository.
- Each time you commit, your work is automatically built and tested, and bugs are detected faster.
- Code, data, models and training pipelines are shared to accelerate innovation.

101010
010101
101010

Process

- Provide templates to bootstrap your infrastructure and model development environment, expressed as code.
- Automate the entire process from code commit to production.

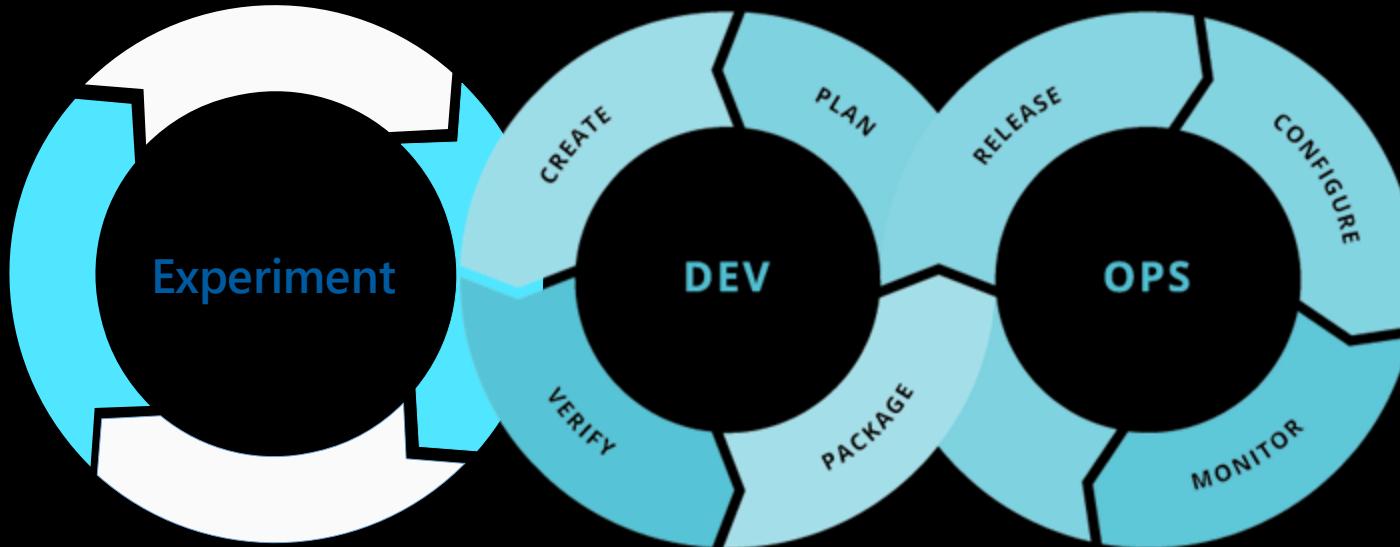


Platform

- Safely deliver features to your customers as soon as they're ready.
- Monitor your pipelines, infrastructure and products in production and know when they aren't behaving as expected



MLOps lifecycle



Experiment

Data Acquisition
Business Understanding
Initial Modeling

Develop

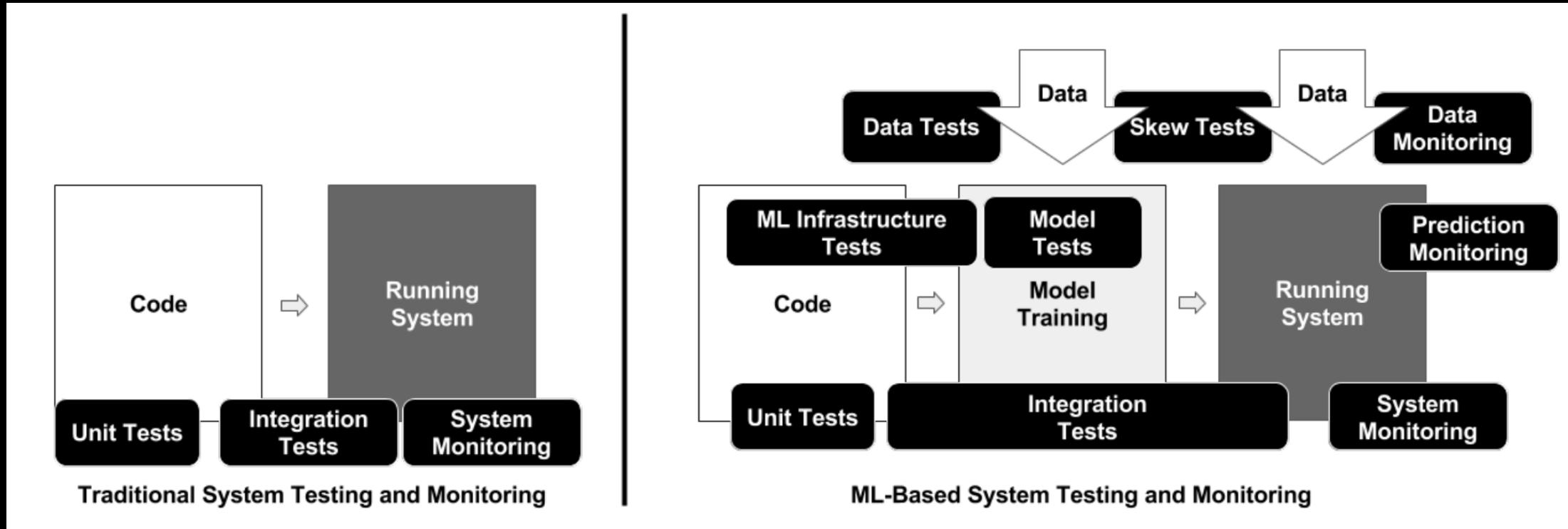
Modeling + Testing
Continuous Integration
Continuous Deployment

Operate

Continuous Delivery
Data Feedback Loop
System + Model Monitoring



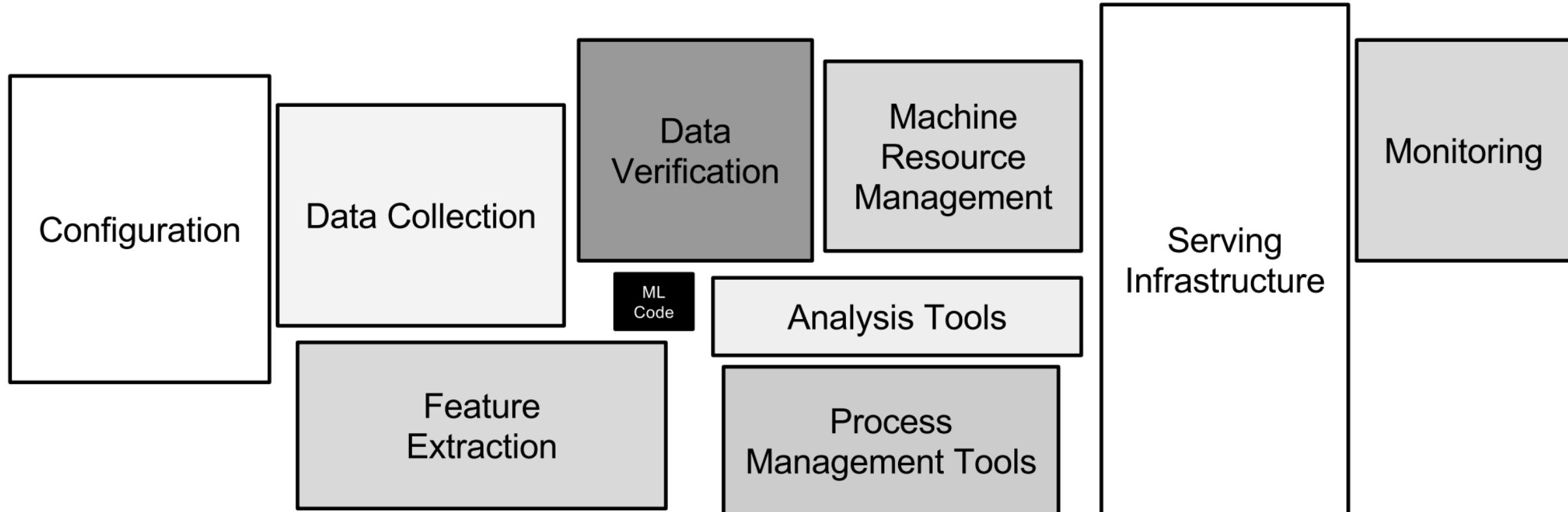
Traditional vs. ML infused systems



ML introduces two new assets into the software development lifecycle – **data** and **models**.



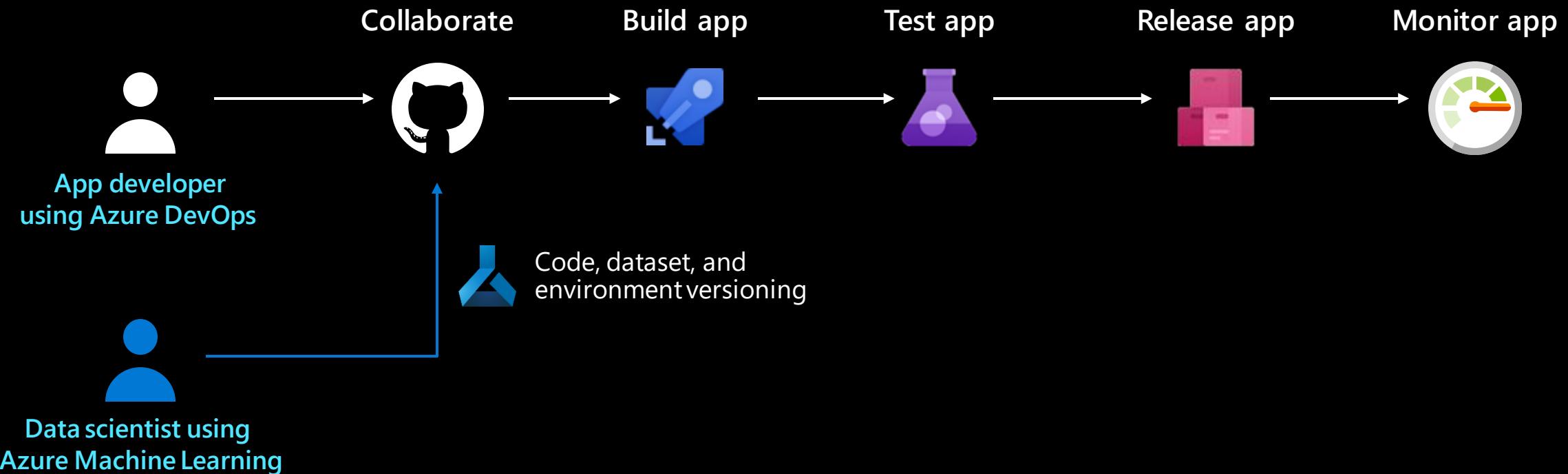
More assets & process to manage



Sculley, D.; Holt, Gary; Golovin, Daniel; Davydov, Eugene; Phillips, Todd; Ebner, Dietmar; Chaudhary, Vinay; Young, Michael; Crespo, Jean-Francois; Dennison, Dan (7 December 2015). ["Hidden Technical Debt in Machine Learning Systems"](#)



MLOps with Azure Machine Learning



Model reproducibility

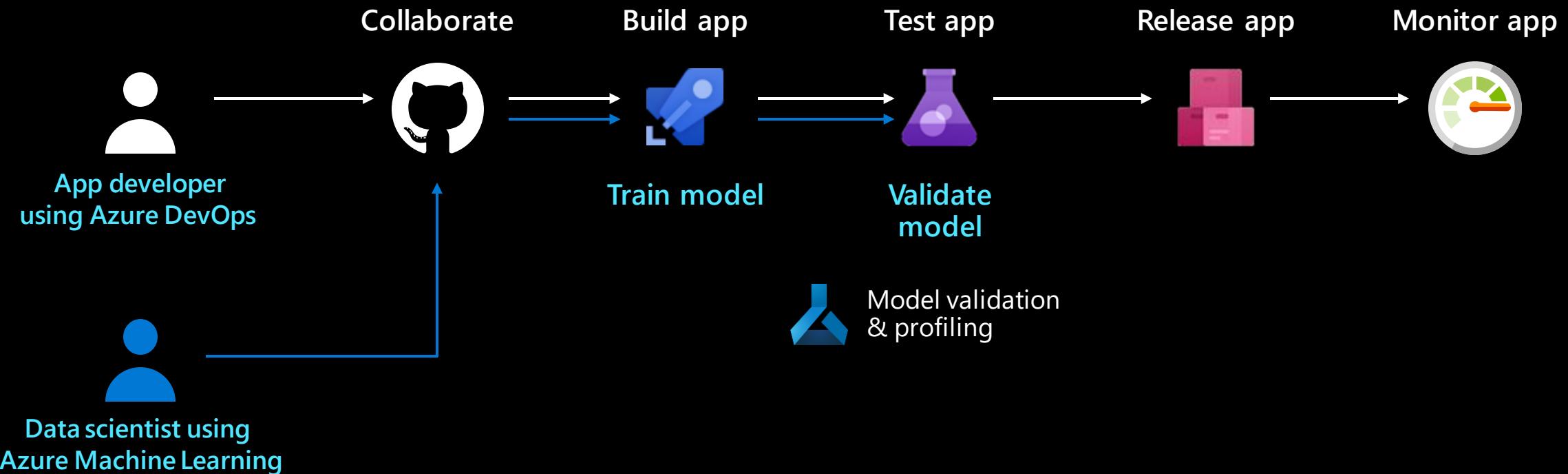
Model validation

Model deployment

Model retraining



MLOps with Azure Machine Learning



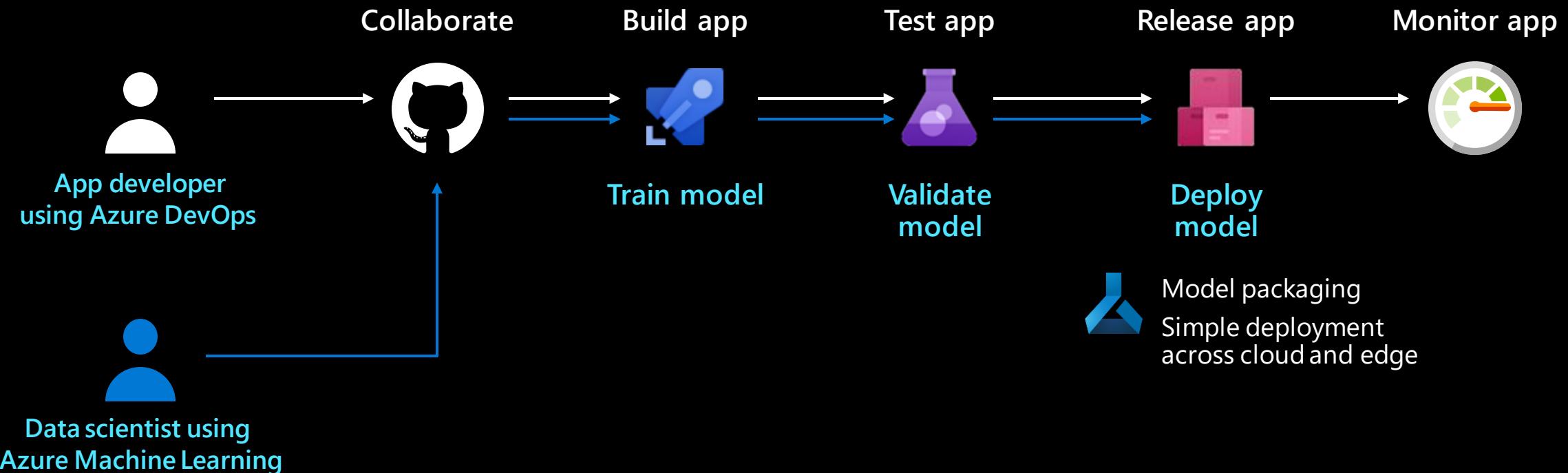
Model reproducibility

Model validation

Model deployment

Model retraining

MLOps with Azure Machine Learning



Model reproducibility

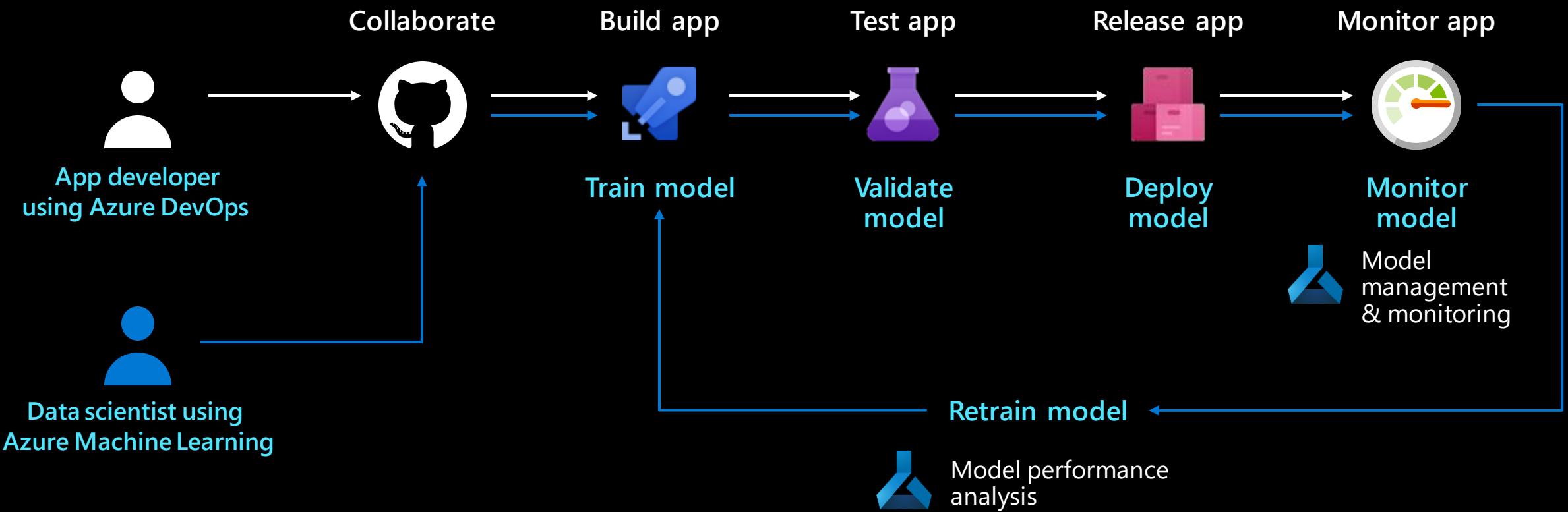
Model validation

Model deployment

Model retraining



MLOps with Azure Machine Learning



Model reproducibility

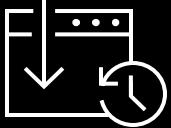
Model validation

Model deployment

Model retraining



MLOps: Supporting Technologies



Infrastructure as Code

- **Azure Resource Manager** Templates
- **Azure ML Python SDK & CLI**
- Azure SDK's

101010
010101
101010

CI/CD

- **Azure DevOps Pipelines**
- **Azure ML Pipelines**
- Azure Repos / GitHub
- Azure Boards



Testing / Release / Monitoring

- **Azure DevOps** for automated testing
- **R** - Runit and testthat
- **Python** - PyUnit, pytest, nose, ...
- **Azure ML & MLFlow Tracking**
- **Azure Data Prep SDK** (analyse/profile)
- **Azure ML Model Management**
(Instrumentation, Telemetry)
- **Azure ML – Data Drift Detection**
- **Azure Monitor** for app telemetry

Azure DevOps + Azure ML

Fast, easy to use, well-controlled Mlops lifecycle



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



Which Azure pipeline technology should I use?

The Azure cloud provides several types of pipeline, each with a different purpose. The following table lists the different pipelines and what they're used for:

Scenario	Primary persona	Azure offering	OSS offering	Canonical pipe	Strengths
Model orchestration (Machine learning)	Data scientist	Azure Machine Learning Pipelines	Kubeflow Pipelines	Data -> Model	Distribution, caching, code-first, reuse
Data orchestration (Data prep)	Data engineer	Azure Data Factory pipelines	Apache Airflow	Data -> Data	Strongly typed movement, data-centric activities
Code & app orchestration (CI/CD)	App Developer / Ops	Azure Pipelines ↗	Jenkins	Code + Model -> App/Service	Most open and flexible activity support, approval queues, phases with gating

How to create AML resources?

3- AML CLI v2 [GA] 4- AML SDK v2 [Preview]

Language of your choice with CLI v2

Automation and Deployment

New Features

Improved Python SDK

CLI v2 supports python, R, Java, Julia or C#

Command lines enable easier automation and deployment

Managed inferencing and Reusable components for pipelines available in v2

Python functions to build any workflow
Go from simple to complex tasks incrementally

Azure ML CLI v2

Format:

```
az ml <noun> <verb> <options>
```

For the CLI (v2), assets and workflows are authored in YAML format : What to run, How to run it, Where to run it, ...

- Assets
 - Component
 - Data
 - Environment
 - Model
- Endpoints
 - Batch
 - Online
- Jobs
 - basics
 - single-step
 - Pipelines
 - Automl
- Resources
 - Compute
 - Connections
 - Datastore
 - workspace

```
az ml model create --file my_model.yml
```

```
az ml online-endpoint create --name $ENDPOINT_NAME -f endpoints/online/managed/sample/endpoint.yml
```

```
az ml job create --file pipeline.yml
```

```
az ml compute --file my_compute.yml
```

<https://github.com/Azure/azureml-examples/tree/main/cli>

SDK v2 [Preview]

- The Python SDK v2, in preview, simplifies the developer experience

- Use Python functions to build a single step or a complex workflow

SDK v2 allows you to build a single command or a chain of commands like python functions - the command has a name, parameters, expects input, and returns output.

- Move from simple to complex concepts incrementally

SDK v2 allows you to:

- Construct a single command.
- Add a hyperparameter sweep on top of that command,
- Add the command with various others into a pipeline one after the other. This construction is useful, given the iterative nature of machine learning.

- Reusable components in pipelines

Azure ML introduces components for managing and reusing common logic across pipelines. This functionality is available only via CLI v2 and SDK v2.

- Managed inferencing

Azure ML offers endpoints to streamline model deployments for both real-time and batch inference deployments. This functionality is available only via CLI v2 and SDK v2.

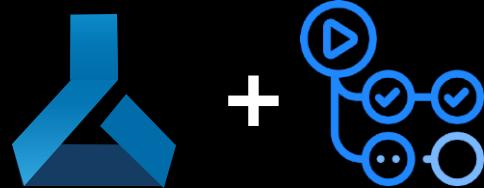
▼ azure.ai.ml

Overview

- azure.ai.ml.automl
- azure.ai.ml.constants
- azure.ai.ml.dsl
- azure.ai.ml.entities
- azure.ai.ml.identity
- azure.ai.ml.sweep
- azure.ai.ml.AmlToken
- azure.ai.ml.Input
- azure.ai.ml.MLClient
- azure.ai.ml.ManagedIdentity
- azure.ai.ml.MpiDistribution
- azure.ai.ml.Output
- azure.ai.ml.PyTorchDistribution
- azure.ai.ml.TensorFlowDistribution
- azure.ai.ml.UserIdentity

<https://github.com/Azure/azureml-examples/tree/main/sdk>

<https://docs.microsoft.com/en-us/python/api/overview/azure/ml/installv2?view=azure-ml-py>



GitHub Actions Integration

- Create AzureML Workspace
- Manage Azure Compute
- Run Jobs in AzureML
- Register Models in AzureML
- Deploy Models in AzureML

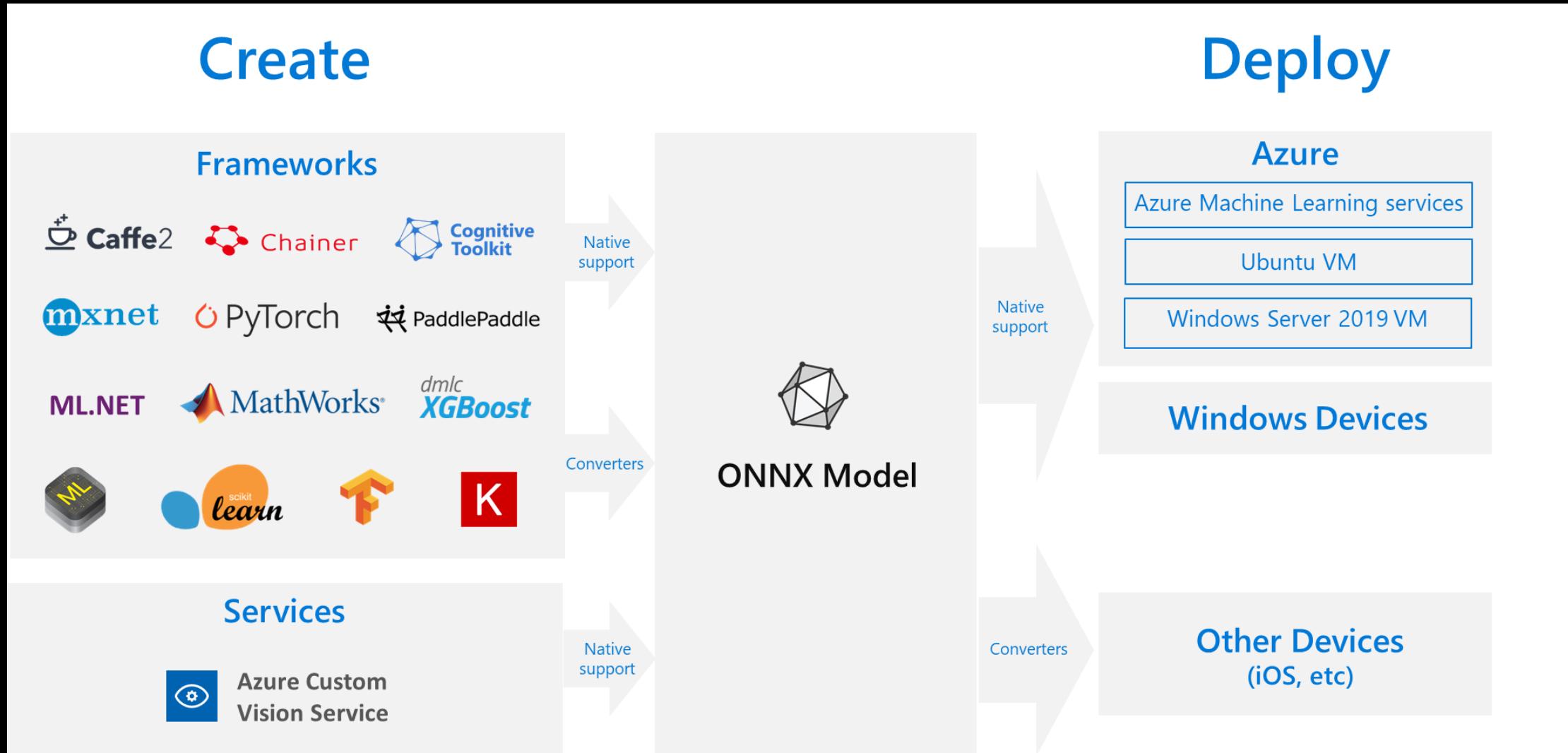
<http://mlops-github.com/actions>

The screenshot shows a GitHub Pull Request interface with several annotations:

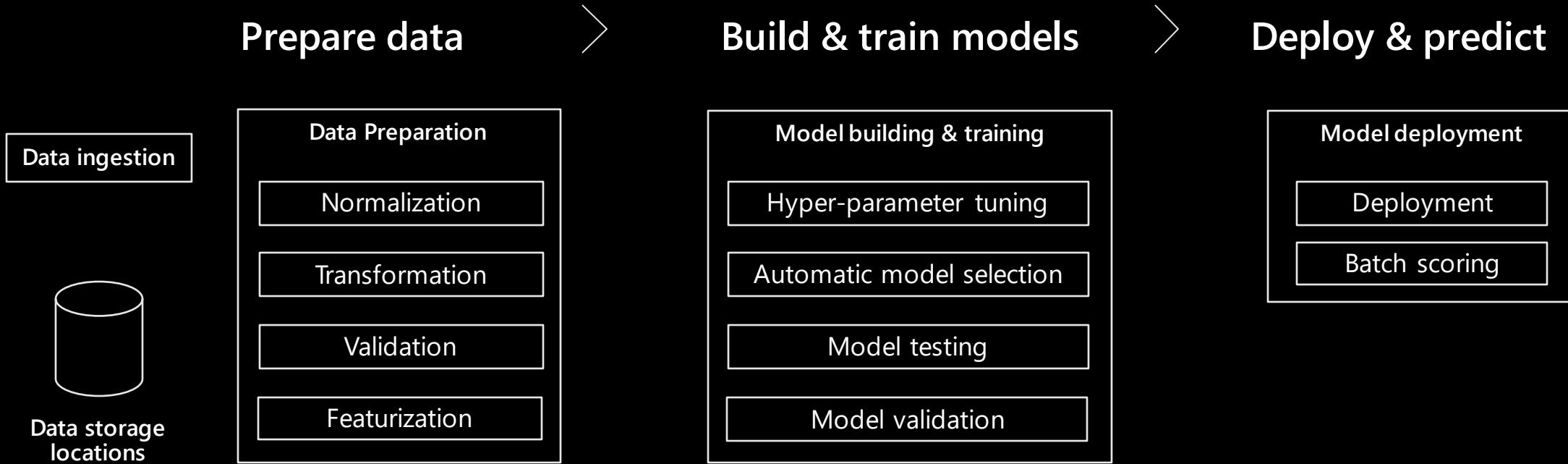
- A red box highlights the GitHub Actions status bar at the top, which includes links for Conversation (4), Commits (1), Checks (2), and Files changed (1).
- An annotation on the right side says "Issue a chat command to GitHub Actions "/run-full-test"" and points to a comment by the user "hamelmsmu" containing the command "/run-full-test".
- A red box highlights the "github-actions" bot comment below, which states "ML Workflow For SHA 5287272 has been instantiated." and lists the built Docker images: "hamelmsu/ml-cicd" and "hamelmsu/ml-cicd-gpu".
- An annotation on the right side says "A Link to my ML Pipeline running the test is dropped into the PR" and points to the workflow instantiation message.
- A red box highlights the "github-actions" bot comment below, which displays "Model Evaluation Results" and provides a table of statistics.
- An annotation on the right side says "Model statistics are dropped into the PR with links back to the experiment tracking system." and points to the evaluation results table.

Category	Run ID	SHA	Train Loss	Val Loss	Acc	Val Acc	Runtime
candidate	ddscgocn	5287272	0.366	0.534	0.862	0.796	542.478
baseline	d2mg9r7l	0fbe4ae	0.392	0.527	0.851	0.798	577.173

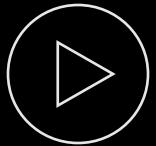
ONNX format



Azure Machine Learning pipelines

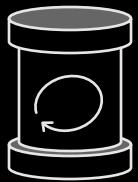


Advantages of Azure ML Pipelines



Unattended runs

Schedule a few steps to run in parallel or in sequence to focus on other tasks while your pipeline runs



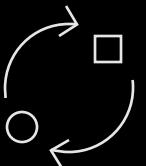
Reusability

Create templates of pipelines for specific scenarios such as retraining and batch scoring



Tracking and versioning

Name and version your data sources, inputs and outputs with the pipelines SDK



Mixed and diverse compute

Use multiple pipelines that are reliably coordinated across heterogeneous and scalable computes and storages



Components (preview) – only SDK v2 and CLI v2

In Azure Machine Learning, a component represents one reusable step in a pipeline, designed to help improve the productivity of pipeline building.

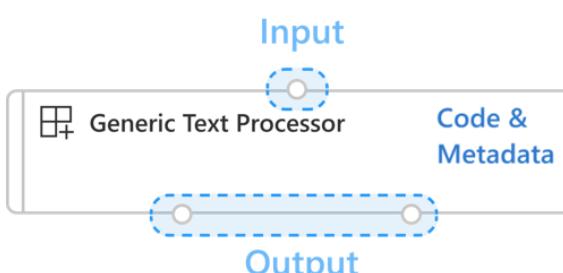
Pipeline



```
graph TD; A[HuggingFace Transformer Model] --> B[Generic Text Processor]; B --> C[GPT2 Finetuning Base]; C --> D[Register GPT2 Model]
```

Components

- Metadata
name, display_name, version, type, etc.
- Interface
input/output specifications (name, type, description, default value, etc).
- Command, Code & Environment
command, code and environment required to run the component



```
graph LR; Input((Input)) --> GP[Generic Text Processor]; GP --> CM[Code & Metadata]; CM --> Output((Output))
```

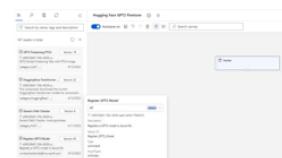
CLI

```
name: "generic_text_processor"
type: "script"
description: "A generic text processor component that takes input text and performs various processing steps like tokenization, stemming, and lemmatization before outputting the processed text."
```

SDK

```
name: "generic_text_processor"
type: "script"
description: "A generic text processor component that takes input text and performs various processing steps like tokenization, stemming, and lemmatization before outputting the processed text."
```

UI



- The interface allows the user to build steps and connect steps easily
- Components can be easily shared and reused across pipelines, workspaces, and subscriptions
- Components are versioned
- Unit testable