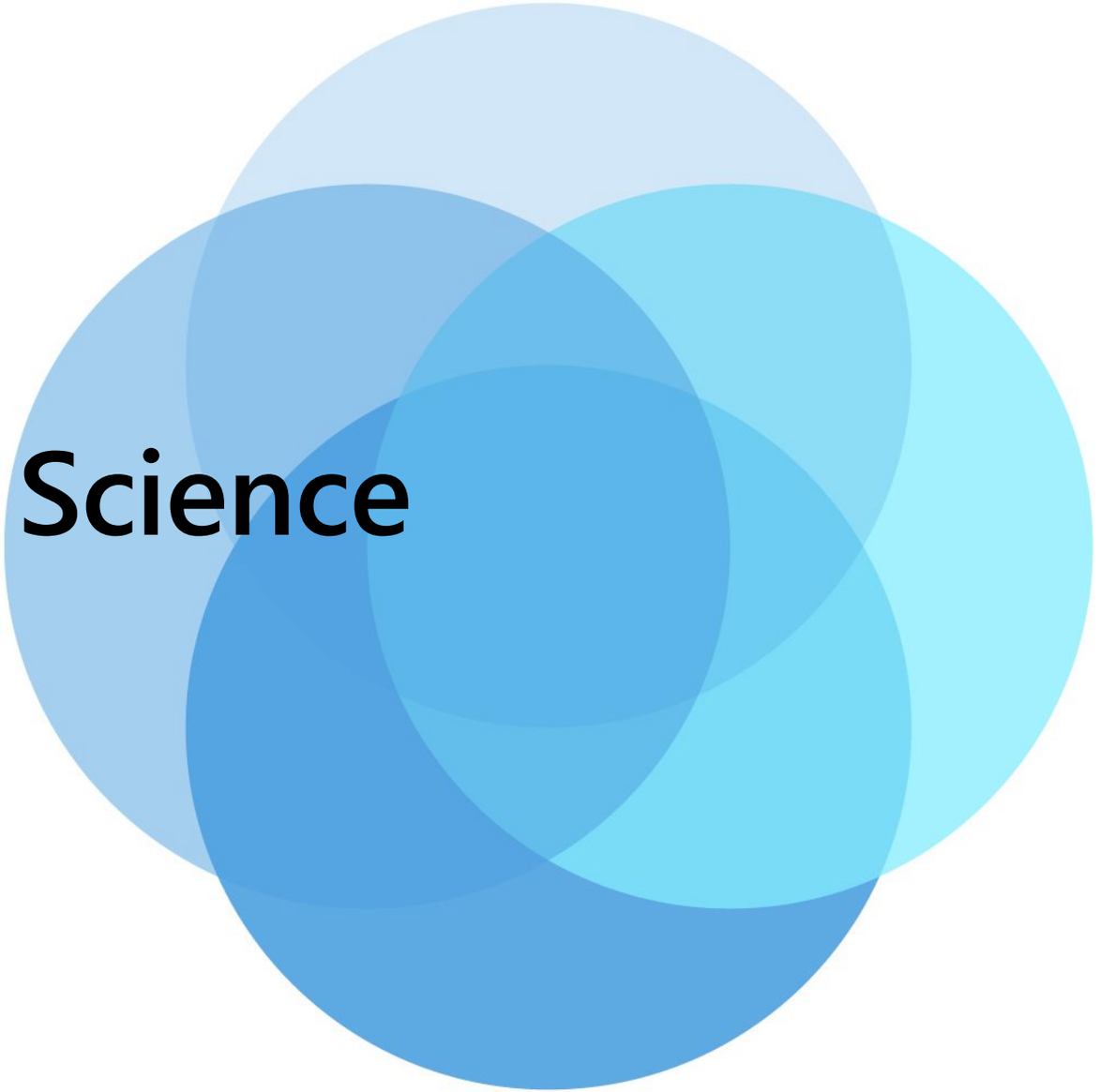
Microsoft Azure

Operationalizing Data Science at Microsoft

María Medina
Delivery Data Scientist @ Microsoft



Hello my name is

María Medina

@mariamedp

Delivery Data Scientist @ Microsoft

Ingeniera Informática y Matemática

1ª promoción MDSF Afi

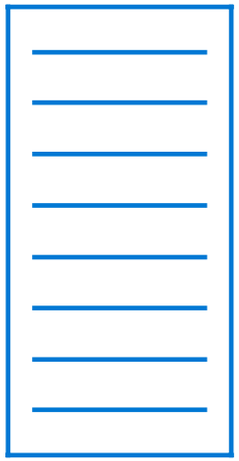
Agenda

1. Cloud & Azure
2. Artificial Intelligence in Azure
3. From predictions to production
4. Azure Machine Learning
5. AML Workshop

Before we start

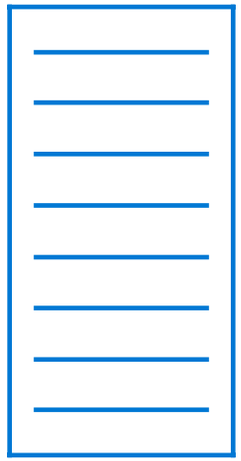
1. Cloud & Azure

On premises vs. cloud computing

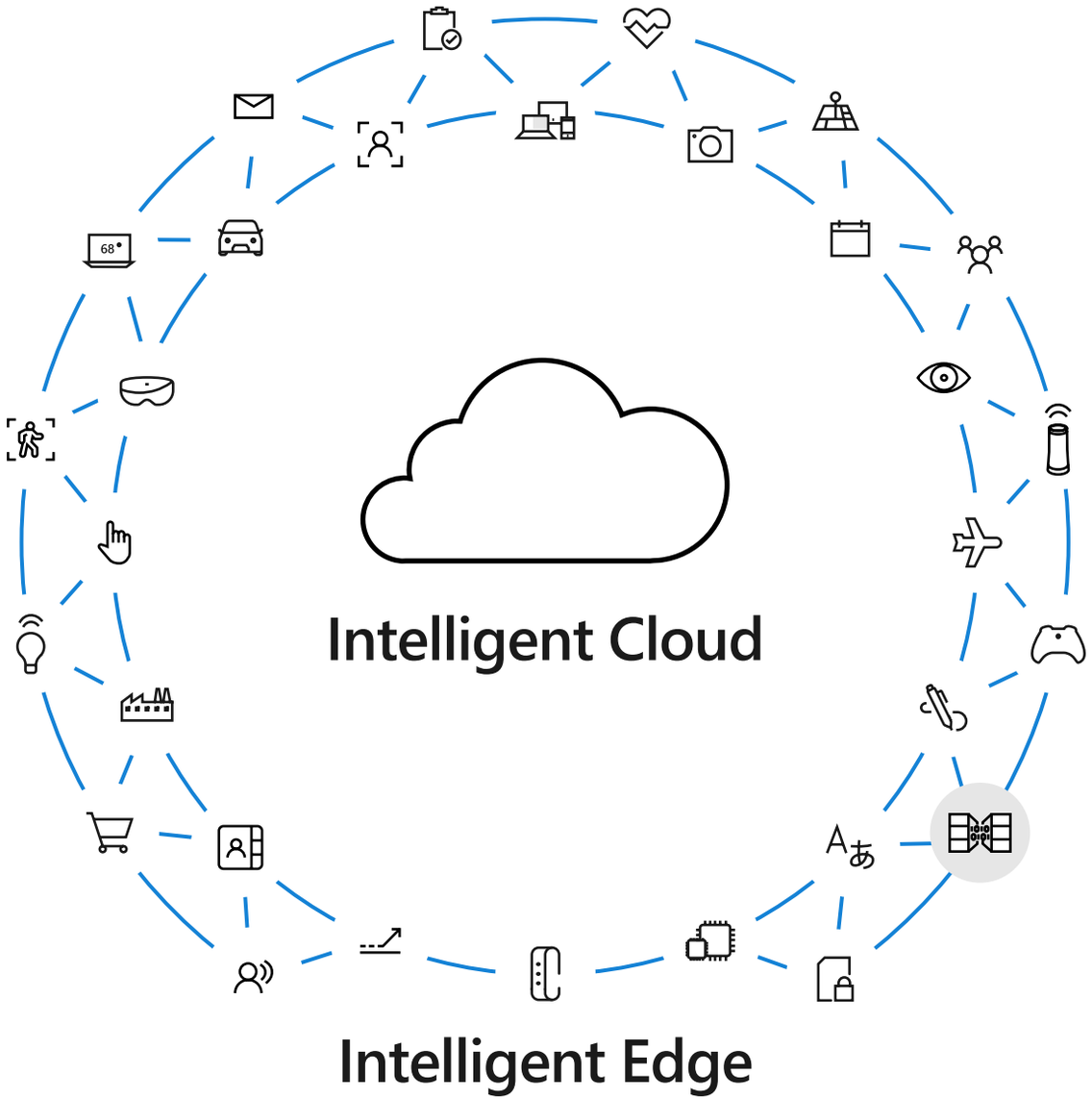


VS.

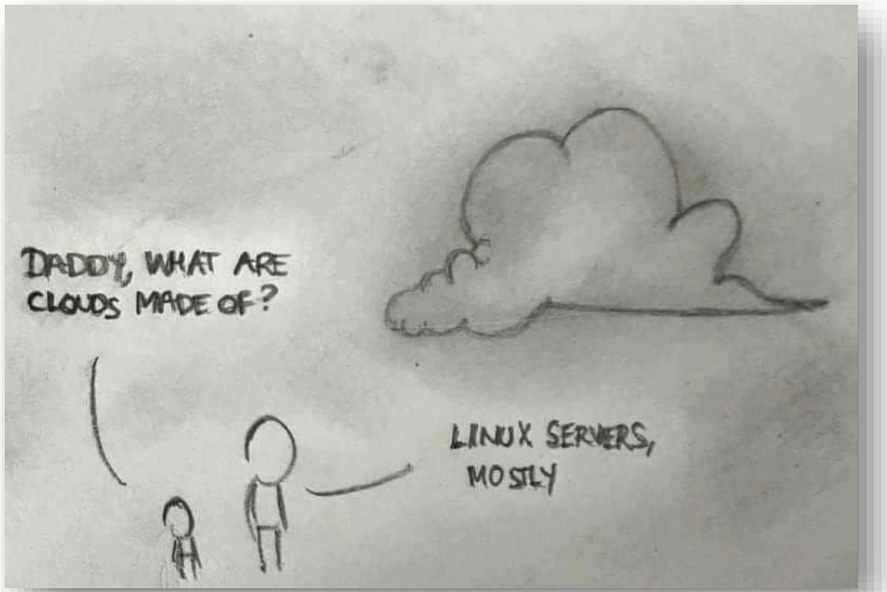
On premises vs. cloud computing

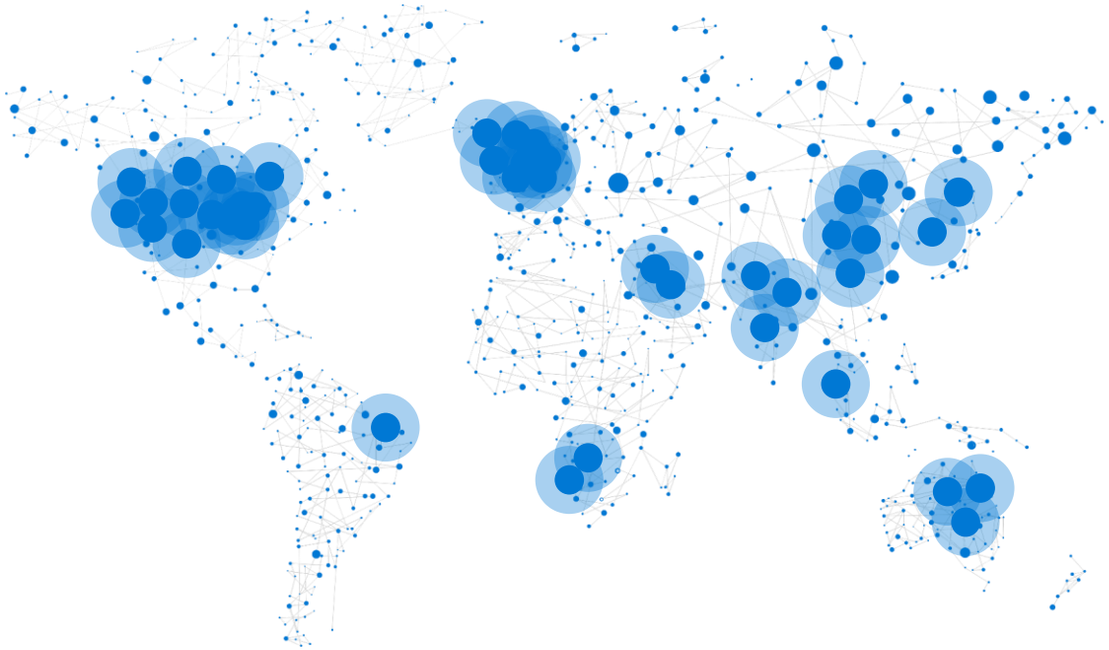


VS.



What is cloud?





60+ Azure regions

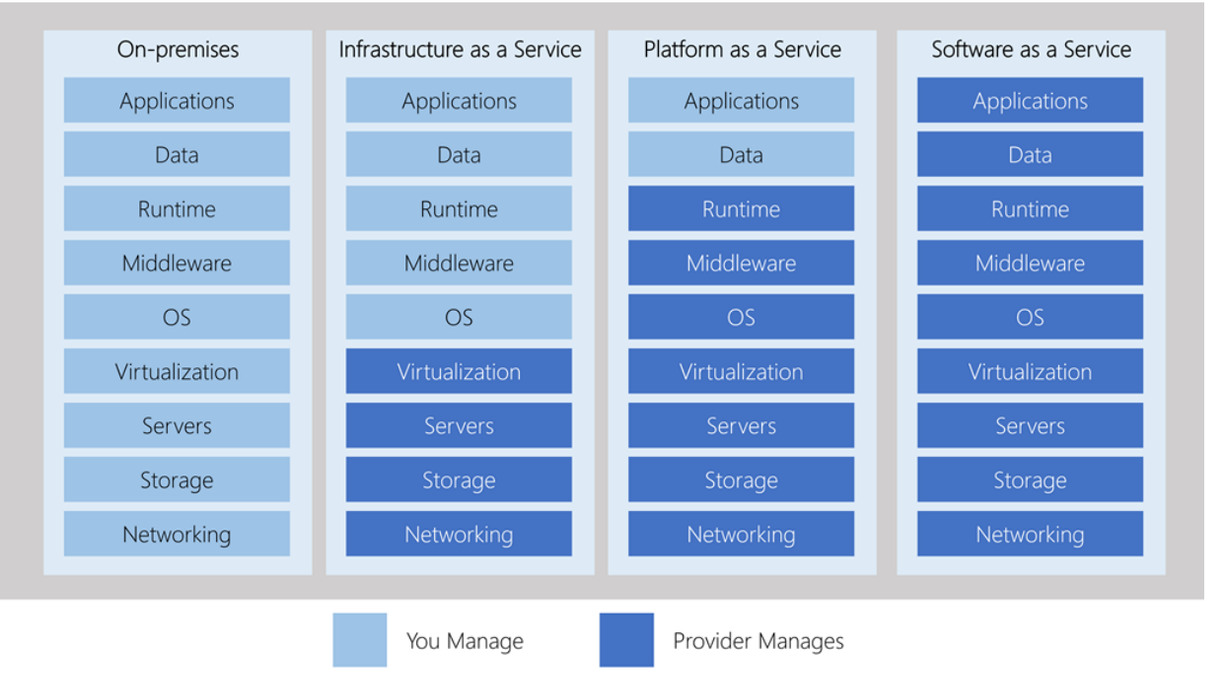
Azure Regions

Types of cloud computing

IaaS - Infrastructure as a Service

PaaS - Platform as a Service

SaaS - Software as a Service



What is Azure?

Edge Devices

Azure Stack Hub | Azure Stack Edge | Azure Sphere | Azure Kinect | HoloLens

Serverless

Web
Mobile
Mixed Reality
Containers
Events + Integration

Databases
Analytics
AI + Machine Learning
Internet of Things
Media

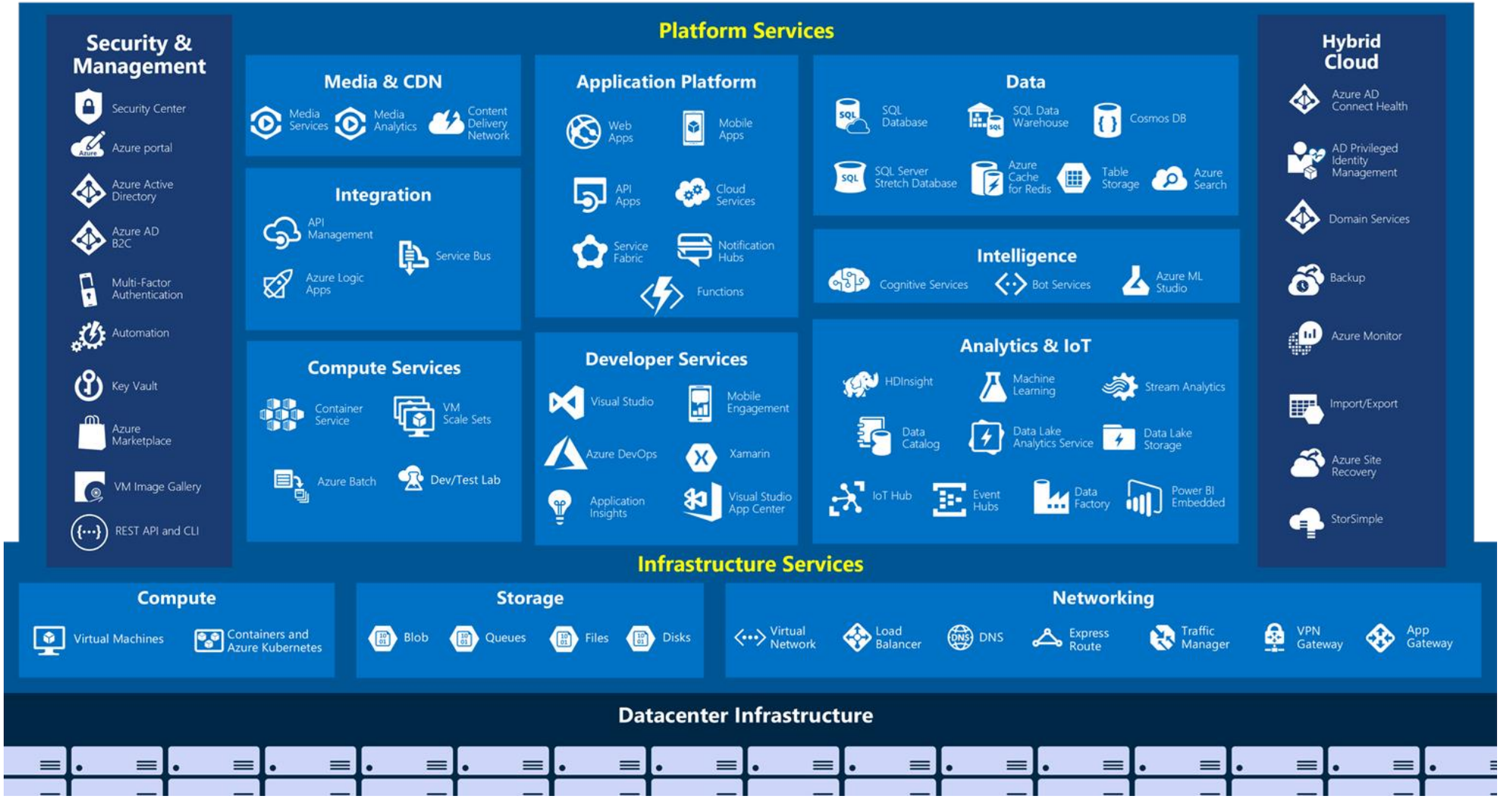
Infrastructure

Compute | Networking | Storage | Security | Identity

Tools

Visual Studio
GitHub
PowerApps
Power BI

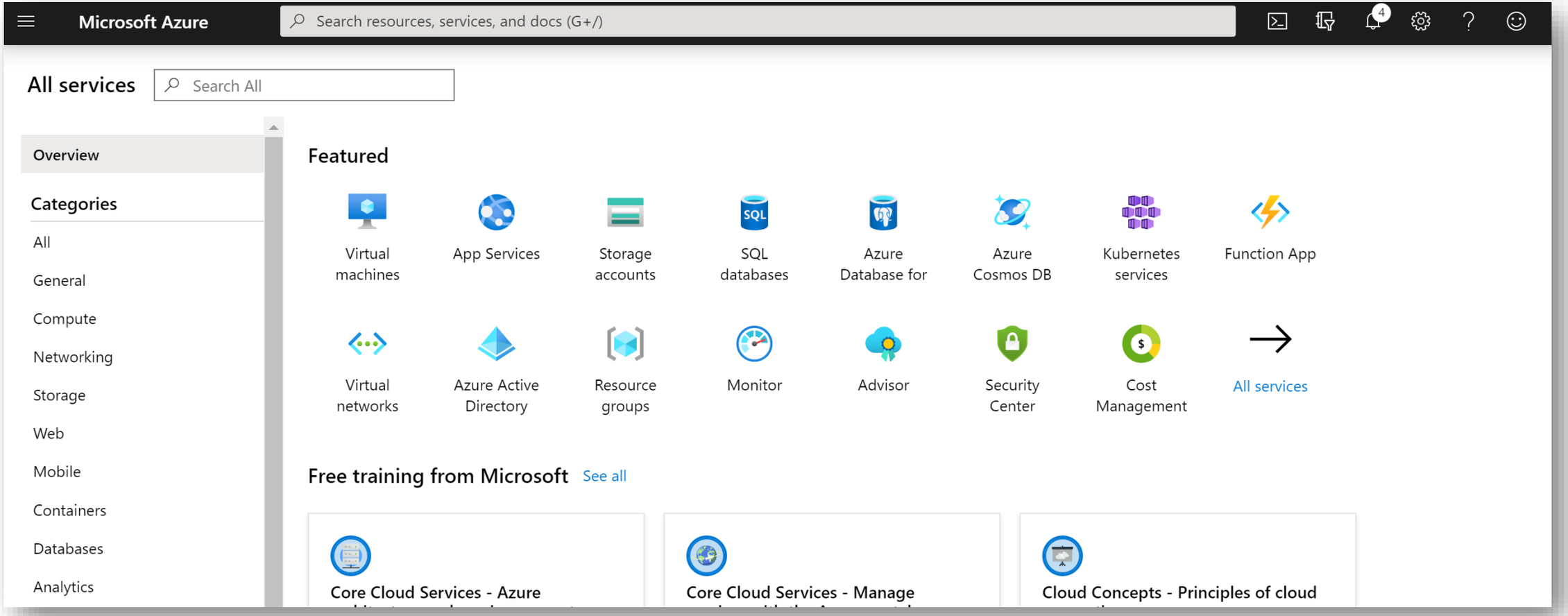
Azure Services



Tour of Azure services (2019)

Azure Portal

Azure free pass: [aka.ms/try-aml](aka.ms/try-aml)



portal.azure.com

2. Artificial Intelligence in Azure

AI in Azure

AI apps and agents

Azure Cognitive Services

Azure Bot Service

Knowledge mining

Cognitive Search

Machine learning

Azure Machine Learning

Azure Databricks

Azure Synapse

Cognitive Services

Vision

Language

Speech

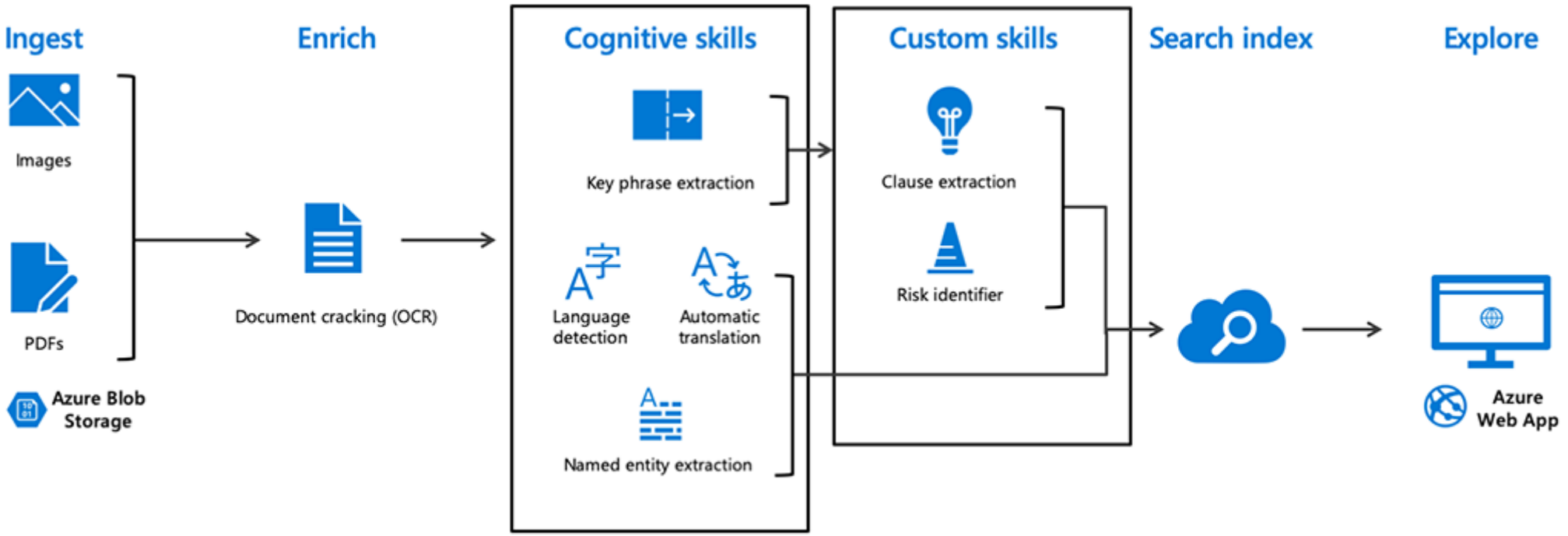
Search

Decision

Computer vision	Text analytics	Speaker recognition	Web search	Personalizer
Custom vision	Translator	Speech to text	Image search	Anomaly Detector
Face	Lang. understanding	Text to speech	Video search	Content Moderator
Form recognizer	Immersive reader	Speech translation	News search	Metrics Advisor
Ink recognizer	QnA maker		Autosuggest	
Video indexer				

Azure Cognitive Services

Azure Search



JFK Files Demo

Machine Learning

Familiar Data Science tools
To simplify model development


Visual Studio Code


Azure Notebooks

Jupyter


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

3. MLOps: from production to production

Why Azure Machine Learning?

¯_(ツ)_/¯

¿Qué hice para conseguir estos resultados?

¿Cuál es la versión buena del modelo?

Olvidarte de que estás entrenando y apagar el ordenador

Más habitual entrenar modelos
que hacer predicciones con ellos

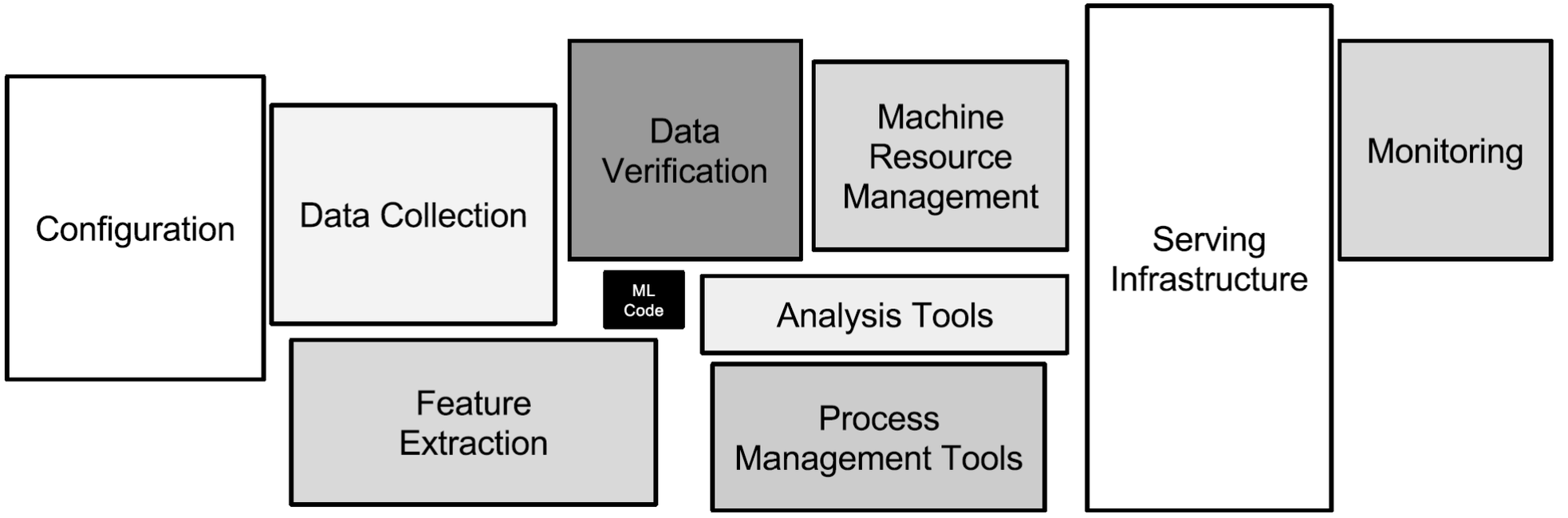
El modelo ejecuta en local, pero no en producción

Tenemos el modelo en producción, pero no sabemos cómo lo está haciendo

La precisión está bajando... ¿cómo reentrenamos?

Tenemos un nuevo modelo... ¿cómo lo cambiamos?

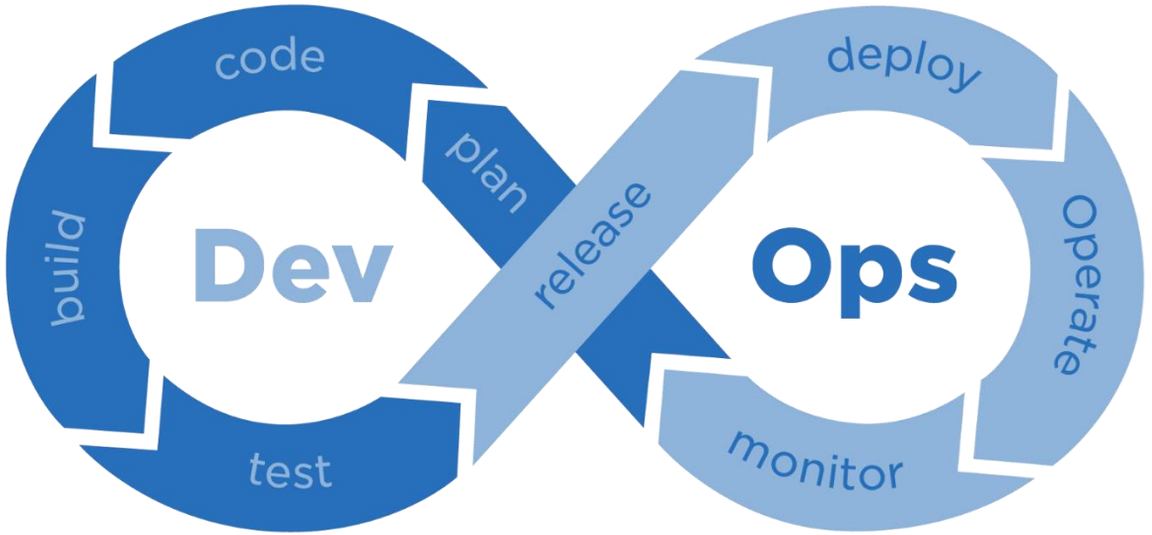
Machine Learning Systems



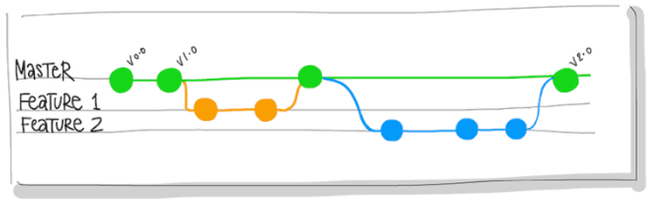
Sculley et al., "Hidden Technical Debt in Machine Learning Systems", 2015

What is DevOps?

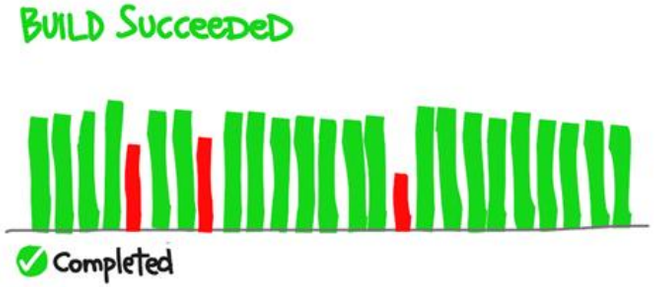
> DevOps is the union of **people**, **process**, and **products** to enable continuous delivery of value to your end users.
>
> *Donovan Brown*



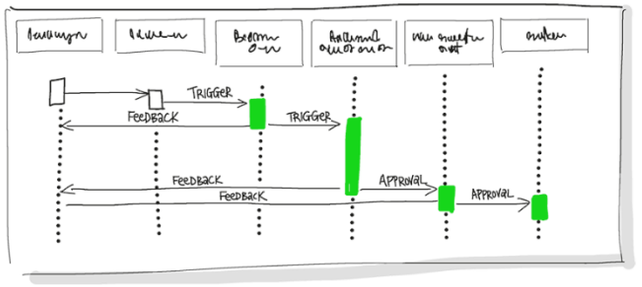
DevOps practices



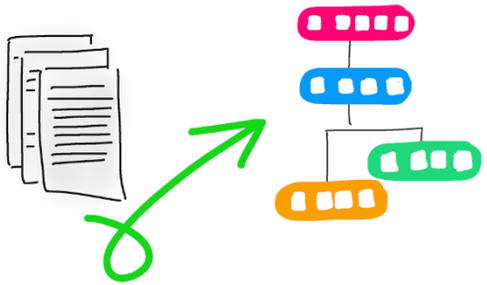
Version Control



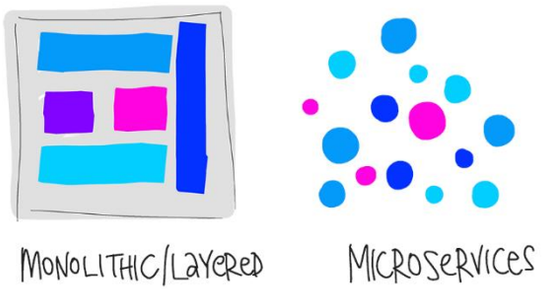
Continuous Integration



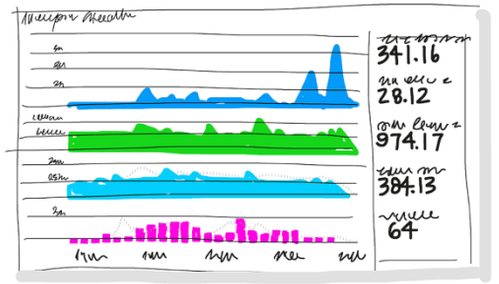
Continuous Delivery



Infrastructure as Code



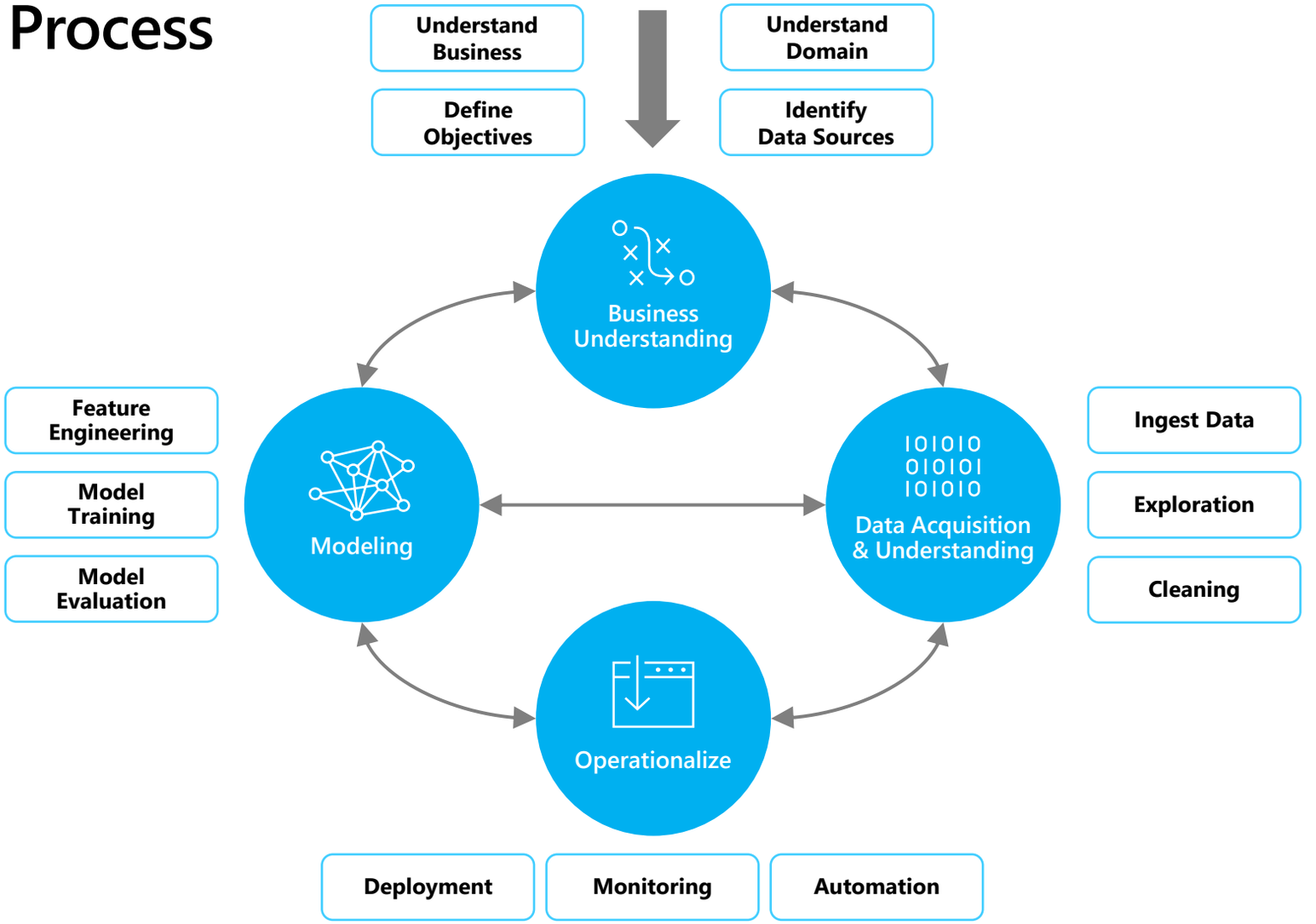
Microservices



Monitoring and logging

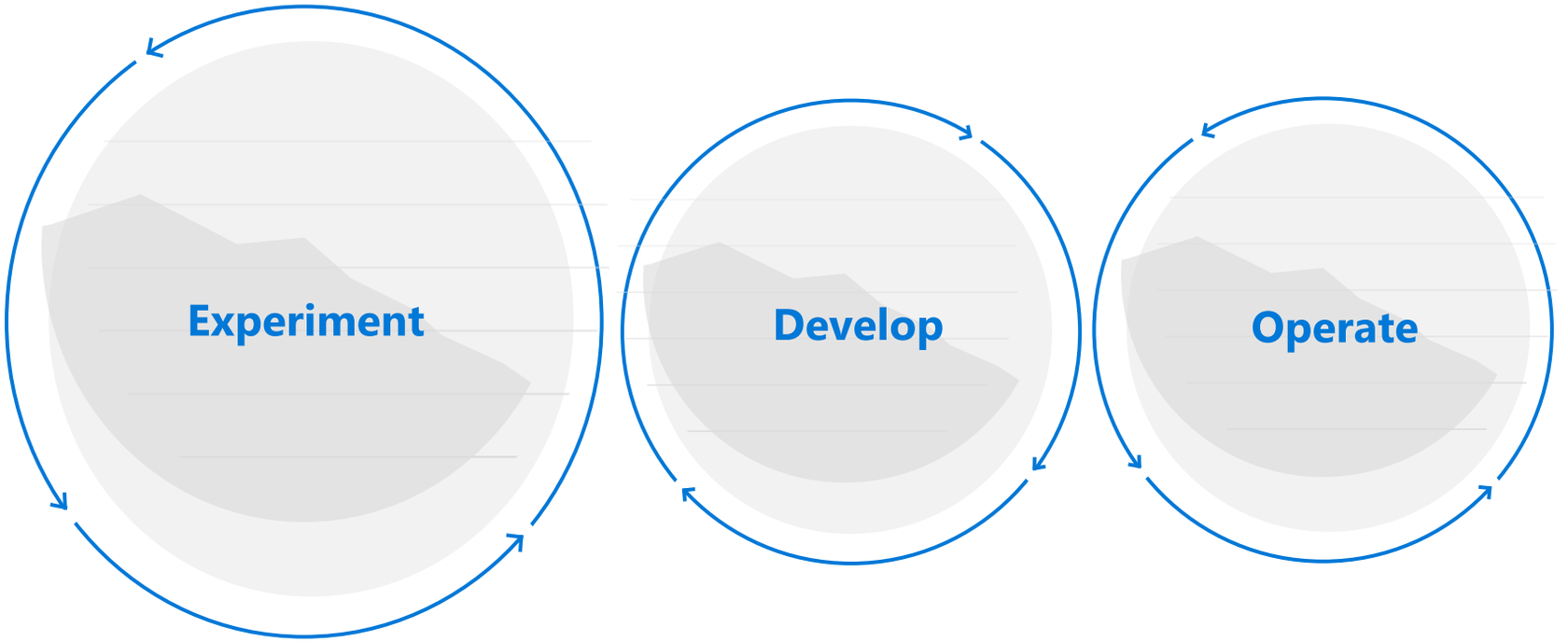
Sam Guckenheimer, "What is DevOps?", 2018

Data Science Process

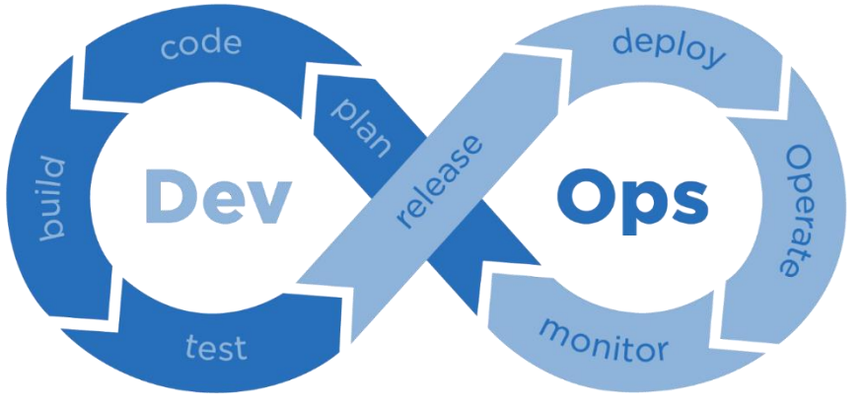


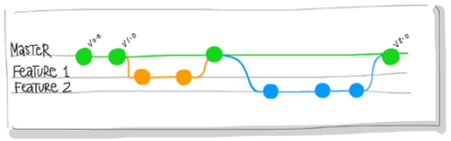
Team Data Science Process, Microsoft

Data Science Lifecycle

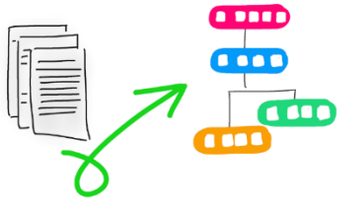


DevOps practices

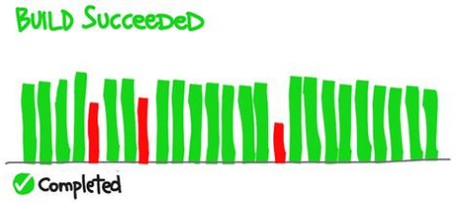




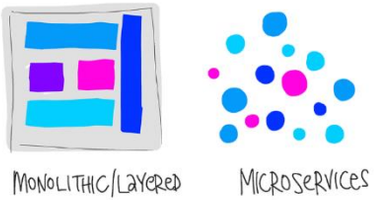
Version Control



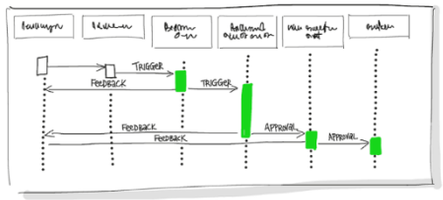
Infrastructure as Code



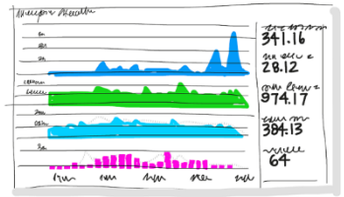
Continuous Integration



Microservices

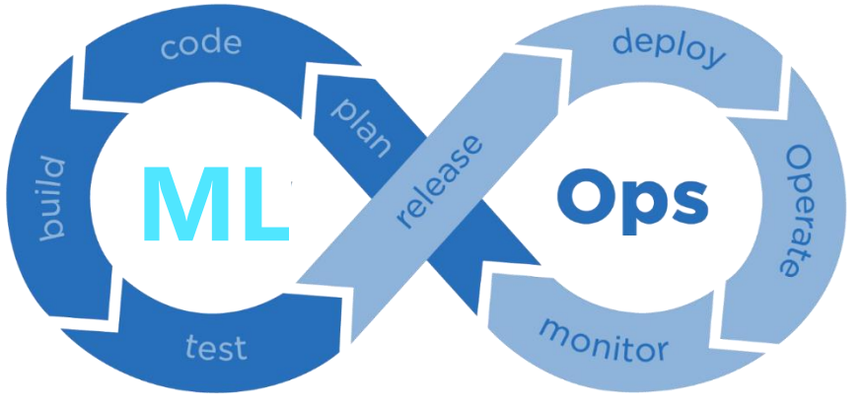


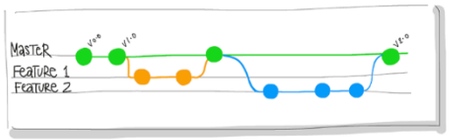
Continuous Delivery



Monitoring and logging

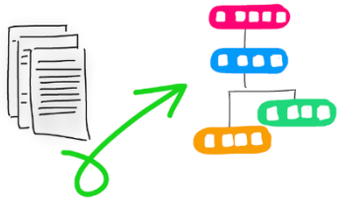
MLOps practices





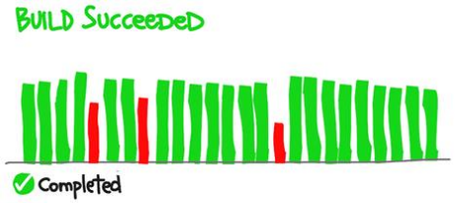
Version Control

code, data & models



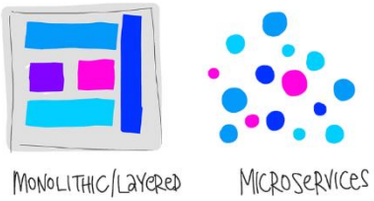
Infrastructure as Code

resources, compute & environments



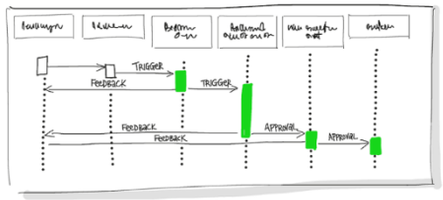
Continuous Integration

training



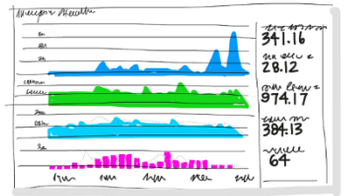
Microservices

Azure Machine Learning ecosystem



Continuous Delivery

model deployment



Monitoring and logging

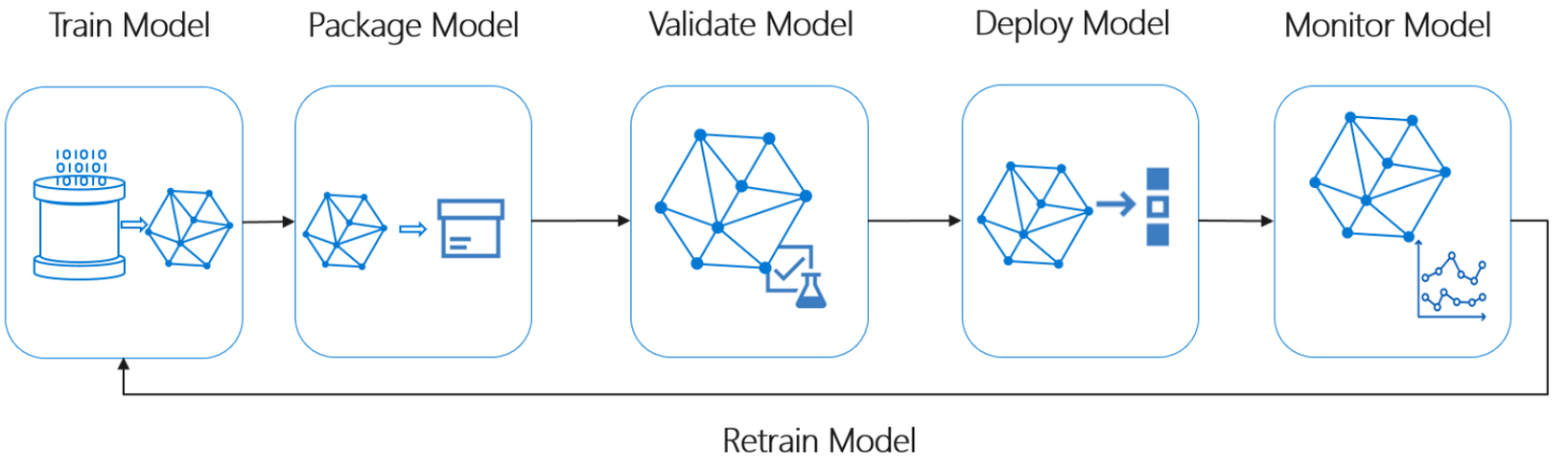
data & model monitoring

4. Azure Machine Learning

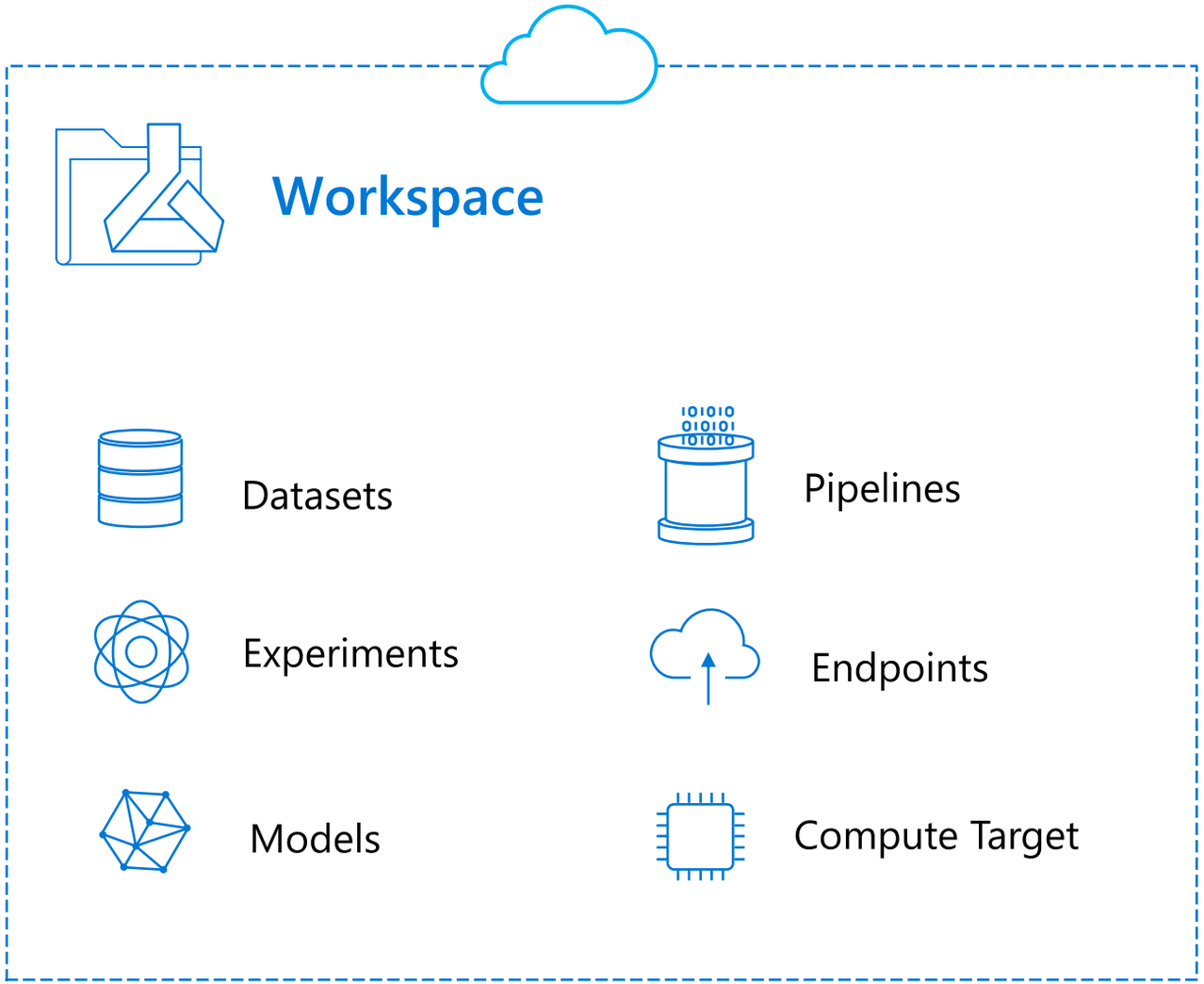
Azure Machine Learning (AML)

Asset management and orchestration services to assist in the lifecycle of model training and deployment workflows.

Azure Machine Learning



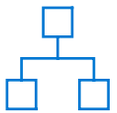
AML Workspace



Python SDK

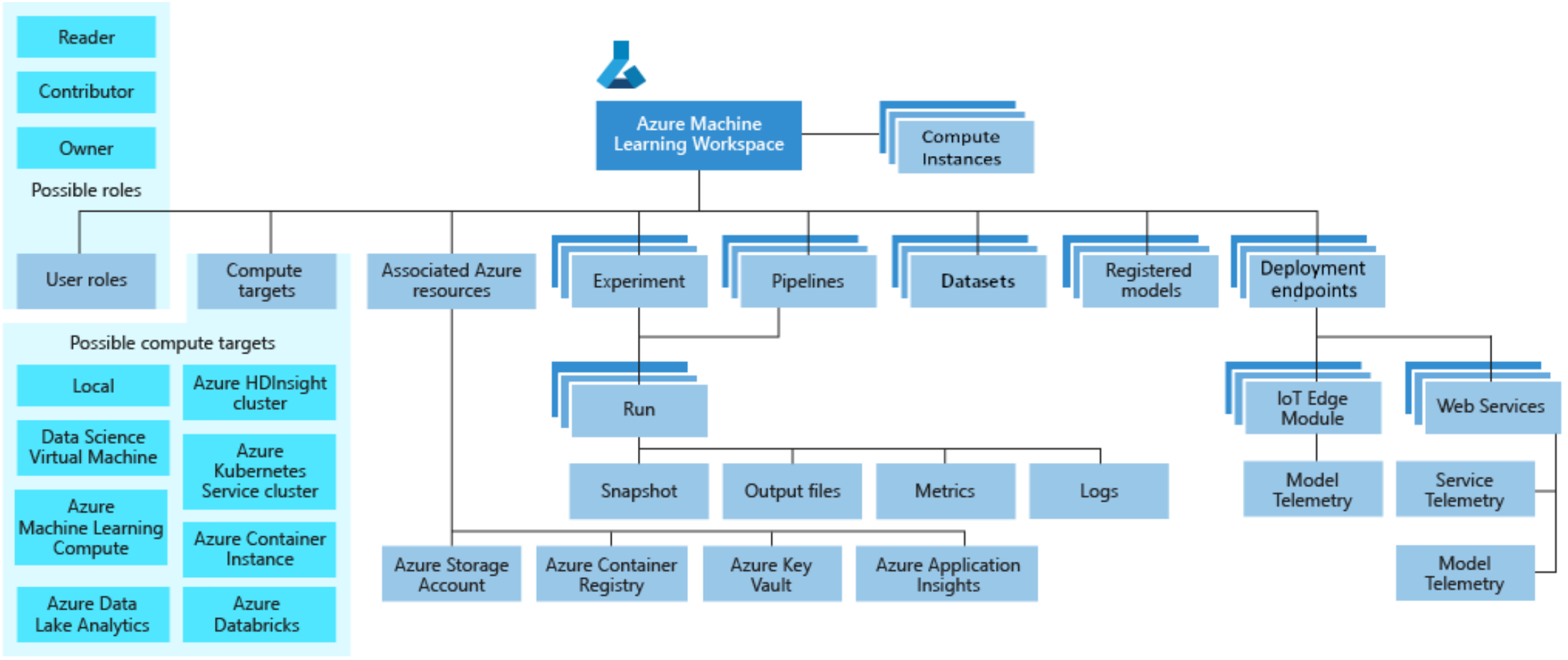
R SDK

Azure CLI

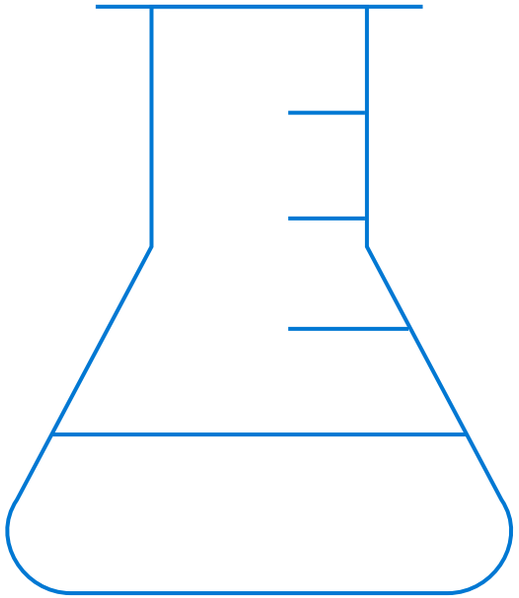


Designer

AML Workspace elements



AML Experiments and Runs



Experiment

Grouping of many runs from a given script.

- Always belongs to a workspace.
- Stores information about runs

Run

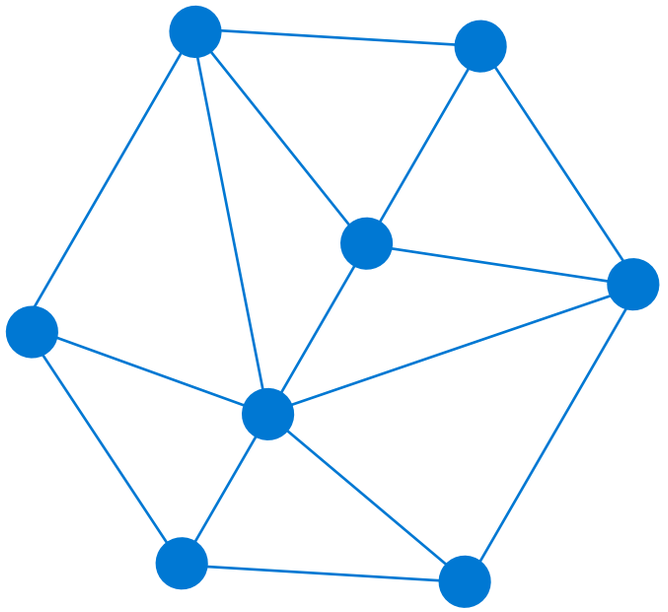
Produced when you submit a script to train a model. Contains:

- Metadata about the run (timestamp, duration etc.)
- Metrics logged by your script.
- Output files auto-collected by the experiment, or explicitly uploaded by you.
- A snapshot of the directory that contains your scripts, prior to the run.

Run configuration

A set of instructions that defines how a script should be run in a given compute target.

AML Models and Model Registry



Model

A machine learning model is an artifact that is created by your training process. You then use a model to get predictions on new data.

A model is produced by a run in Azure Machine Learning. You can also use a model trained outside of Azure Machine Learning.

Model Registry

Keeps track of all the models in your Azure Machine Learning service workspace.

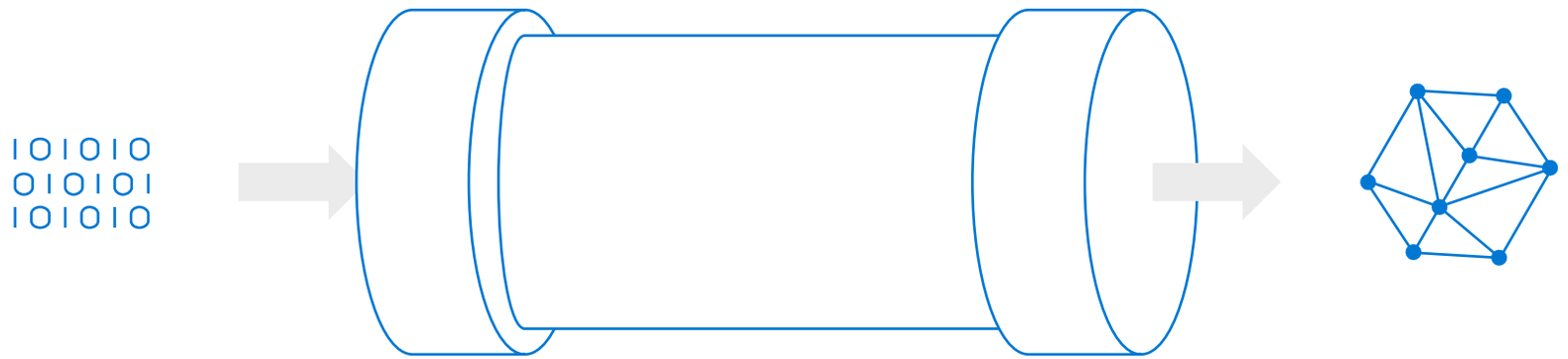
Models are identified by name and version.

You can provide additional metadata tags when you register the model, and then use these tags when searching for models.

You cannot delete models that are deployed.

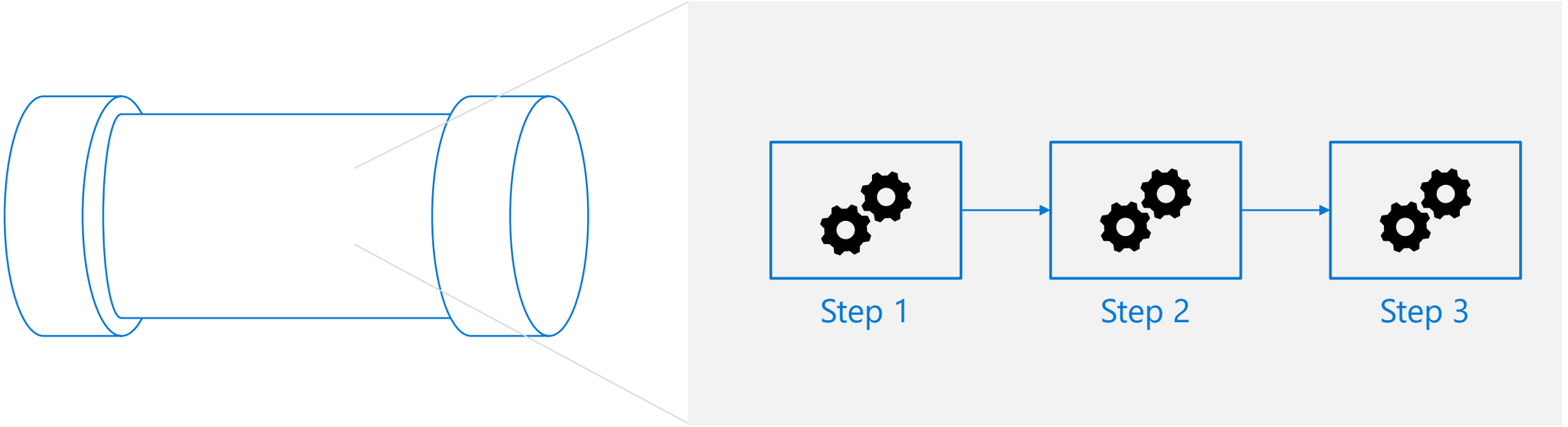
AML Pipelines

Used for creating workflows in your machine learning projects (training, batch inference)

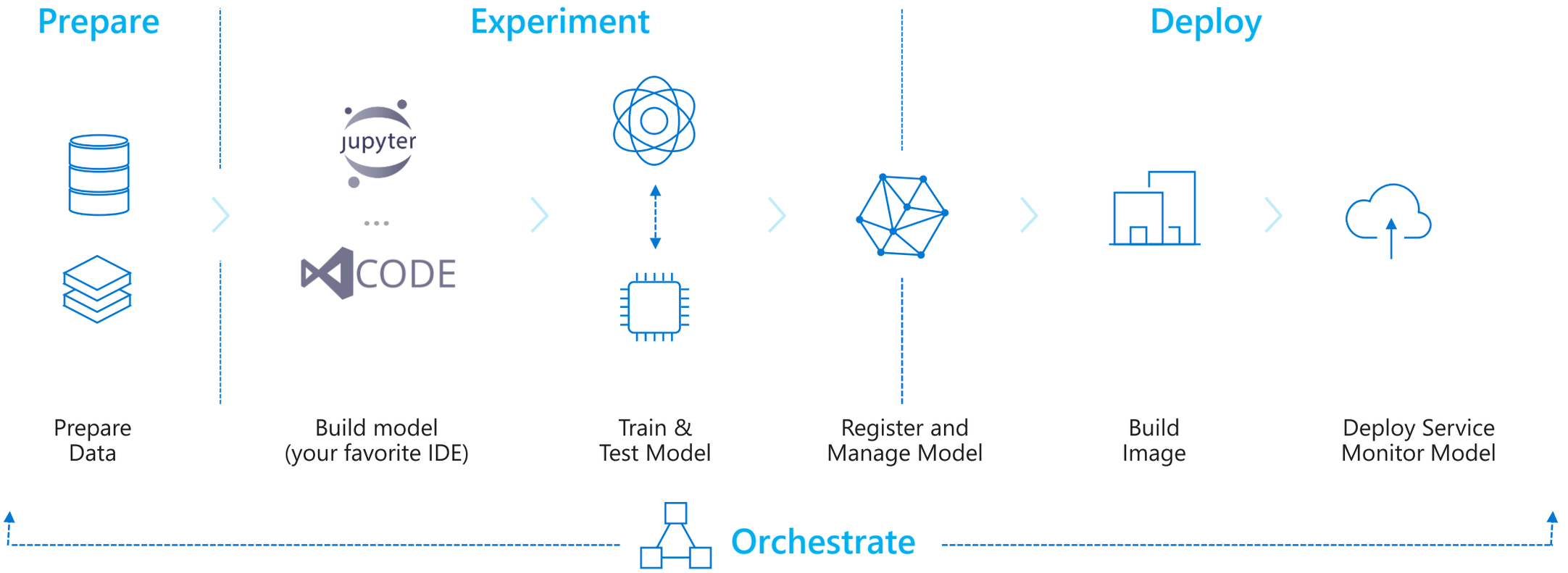


AML Pipelines

Used for creating workflows in your machine learning projects (training, batch inference)



AML end-to-end



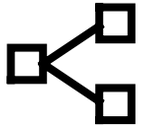
Training with AML

Datasets

+

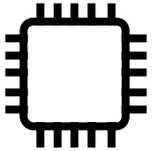
Training file
python/R

+



Dependencies

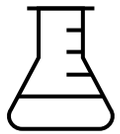
+



Compute Target

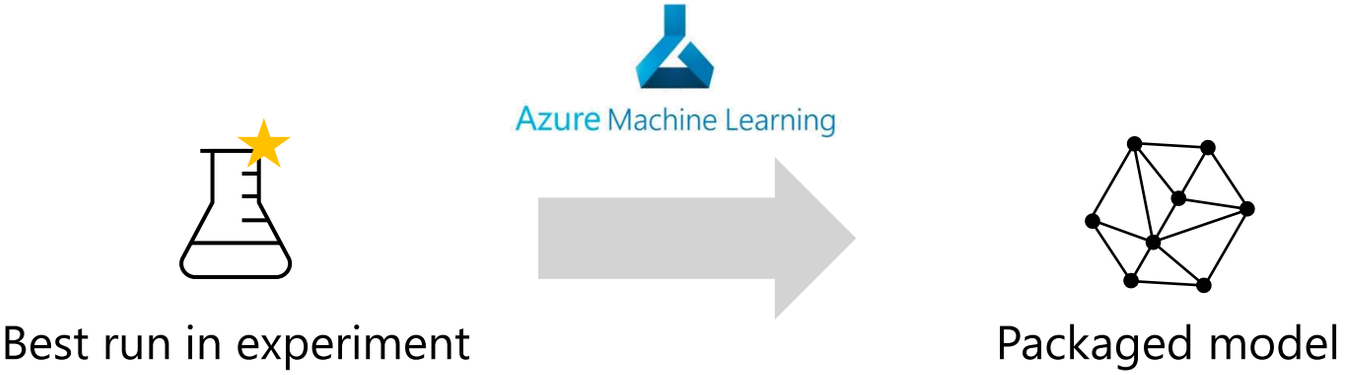
- Local
- Virtual Machine (CPU / GPU)
- Azure Databricks
- ...

Azure Machine Learning



Experiment

Register models from experiments



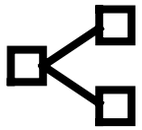
Deploy Model with AML

Model ID

+

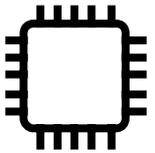
Scoring file
python/R

+



Dependencies

+



Compute Target

- Local
- Azure Container Instances
- Azure Kubernetes Service

Azure Machine Learning

Webservice

Inference with AML

Real-time inference

Webservices to serve real-time predictions upon request for a small number of records at a time.

Artifact: endpoint

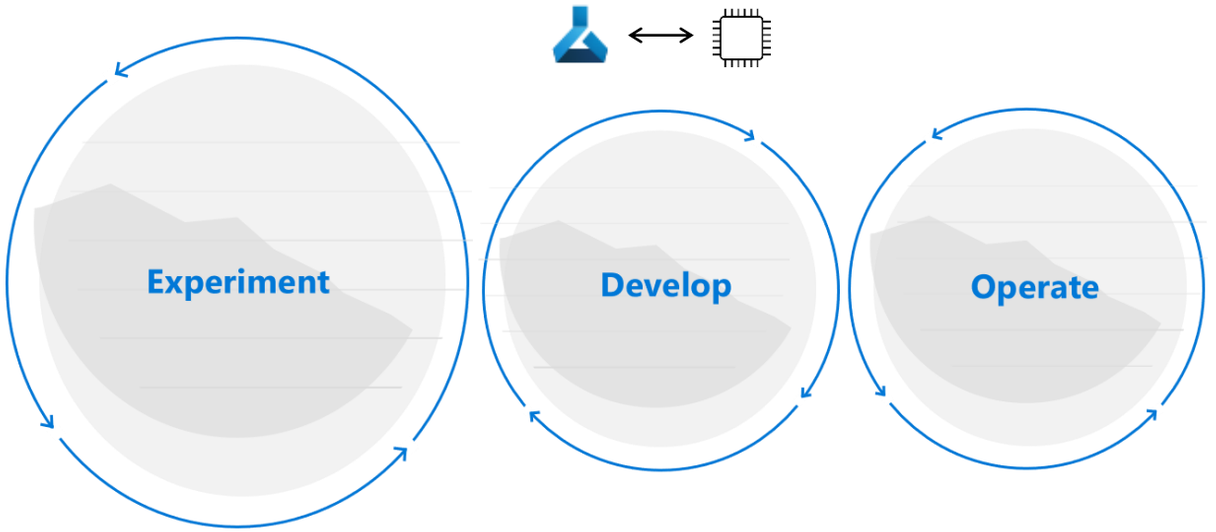
Batch inference

Pipeline to periodically generate asynchronous predictions for all records in the data

Artifact: pipeline

Data Science Lifecycle with AML

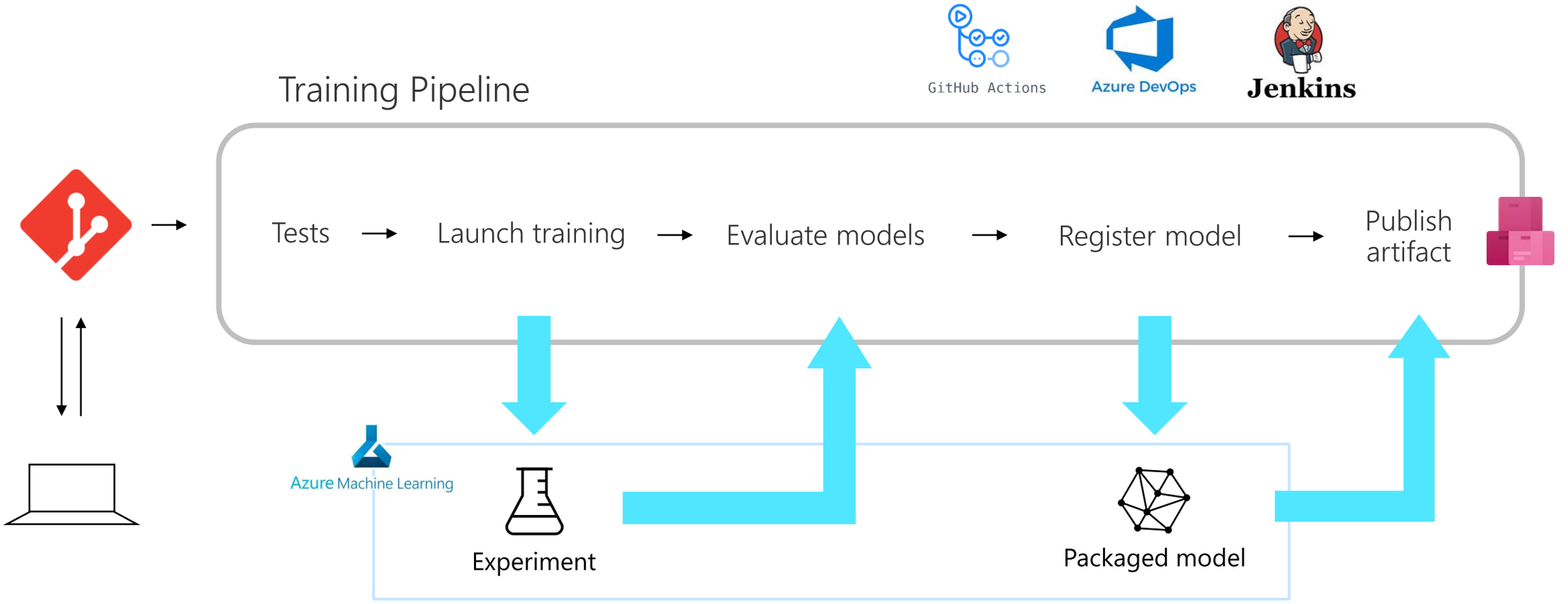
Move from experimental code to production-ready code.
Training process is defined with an AML pipeline and executed using a remote compute cluster.
User-level access to AML needed for low-level testing and review.



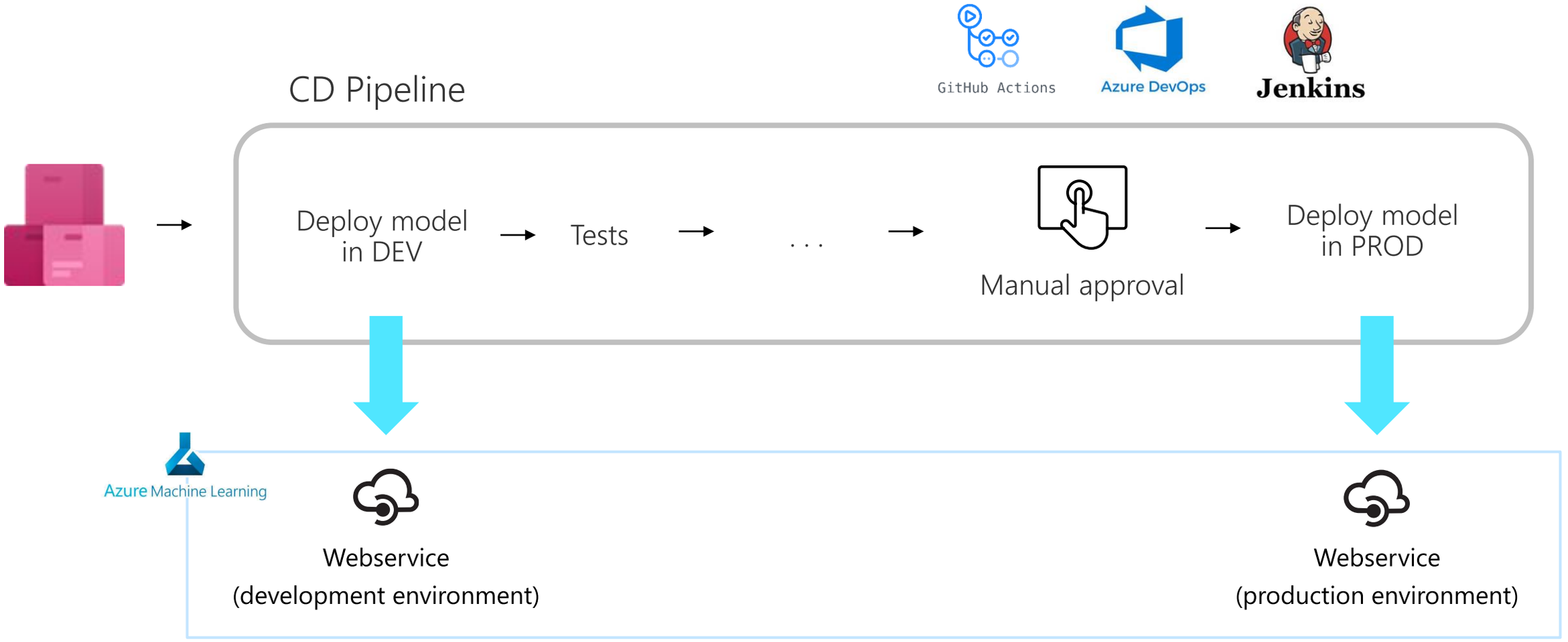
Experimentation done locally or from compute instance using notebooks for flexibility and speed.
Relies on AML workspace for tracking.
User-level access needed to create experiments, manage artifacts, etc.

Production code launched from CI/CD pipelines.
Uses AML Pipelines to manage ML flow.
No user-level access, only via CI/CD pipelines.
Reader access to AML might be needed.

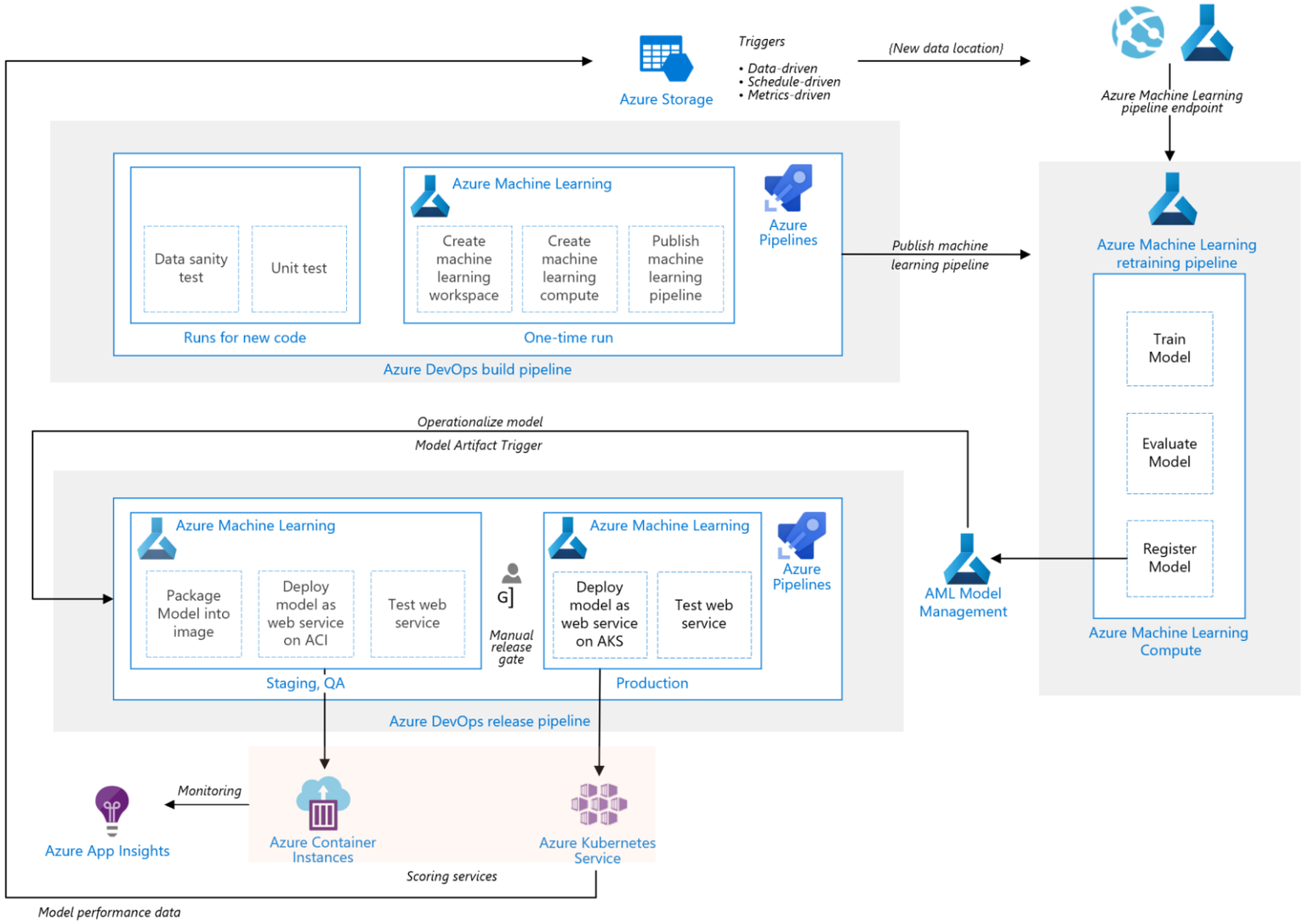
Continuous Integration in MLOps



Continuous Deployment in MLOps



MLOps Architecture



MLOps Reference Architecture, Microsoft

5. Demo time!

Setup:

1. Create Machine Learning workspace in Azure
2. Download *config.json* file
3. Install python SDK

```
pip install azureml-sdk
```

4. Download sample notebooks from Github
   - First experiment
   - Train on remote cluster
   - Deploy real-time webservice

Resources

Azure free pass: aka.ms/try-aml

Azure Machine Learning documentation: AML docs

Azure Machine Learning examples: github repo

Python SDK documentation: azureml docs

Azure Data Science Certification: learning path

Microsoft MLOps Accelerator: microsoft/dstoolkit-mlops-base

Thank you!