GLOBAL AZURE 2021

Torino, 15 aprile 2021

Grazie al prezioso contributo di:

"MLOps con Azure Machine Learning"

Gianni Rosa Gallina
R&D Senior Software Engineer at Deltatre

Clemente Giorio
R&D Senior Software Engineer at Deltatre

Vito Flavio Lorusso
Senior Program Manager at Microsoft

BY COMMUNITY • FOR COMMUNITY
#GLOBALAZURE
2021

TORINO

15 APRILE 2021

welol
beyond the digital

Azure Meetup
TORINO

Deltatre Innovation Lab

Microsoft

AGENDA

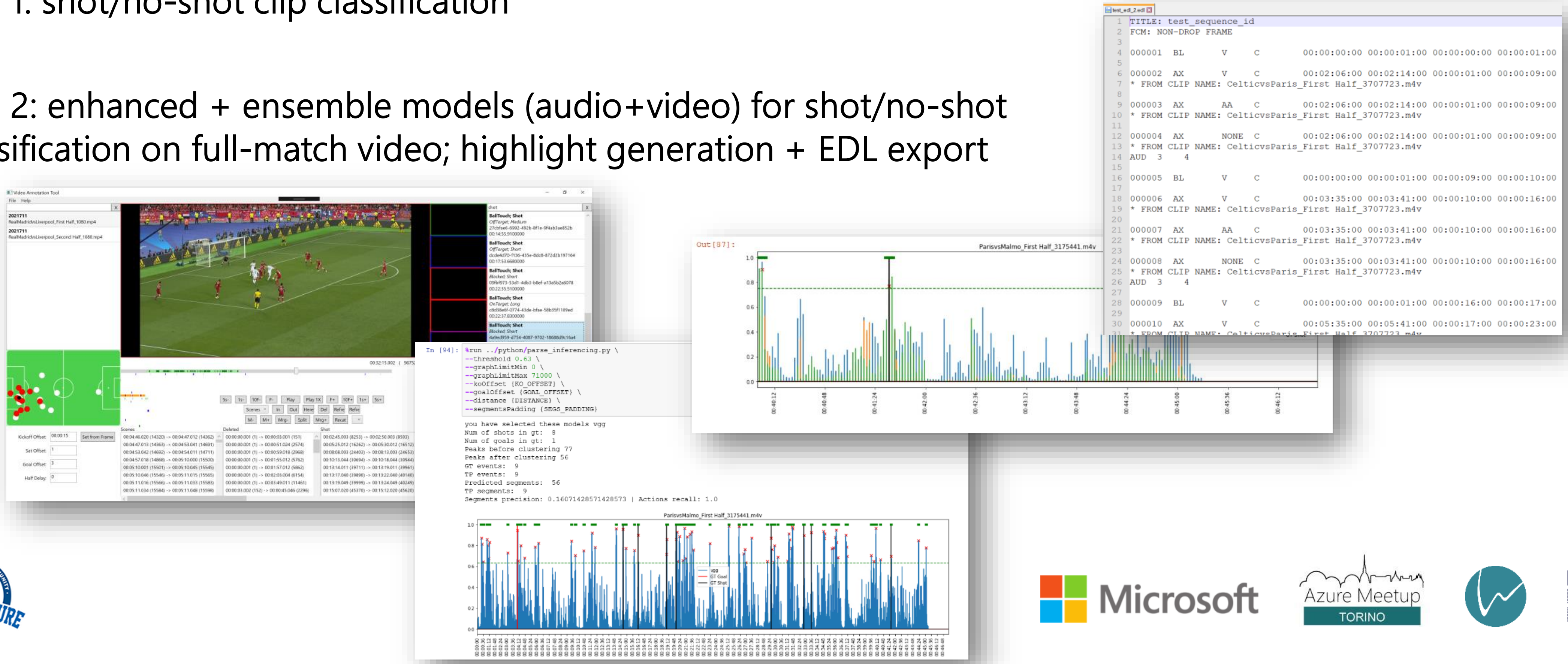
- Our journey
- AI & MLOps Overview
- ML/DL Workflows
- Pipelines and MLOps
- Tips

WHERE IT ALL STARTED

EXPLORATIONS AND PROTOTYPES

Data analysis, tooling and exploration for **action recognition in video**
PoC 1: shot/no-shot clip classification

PoC 2: enhanced + ensemble models (audio+video) for shot/no-shot classification on full-match video; highlight generation + EDL export



#GLOBALAZURE 2021

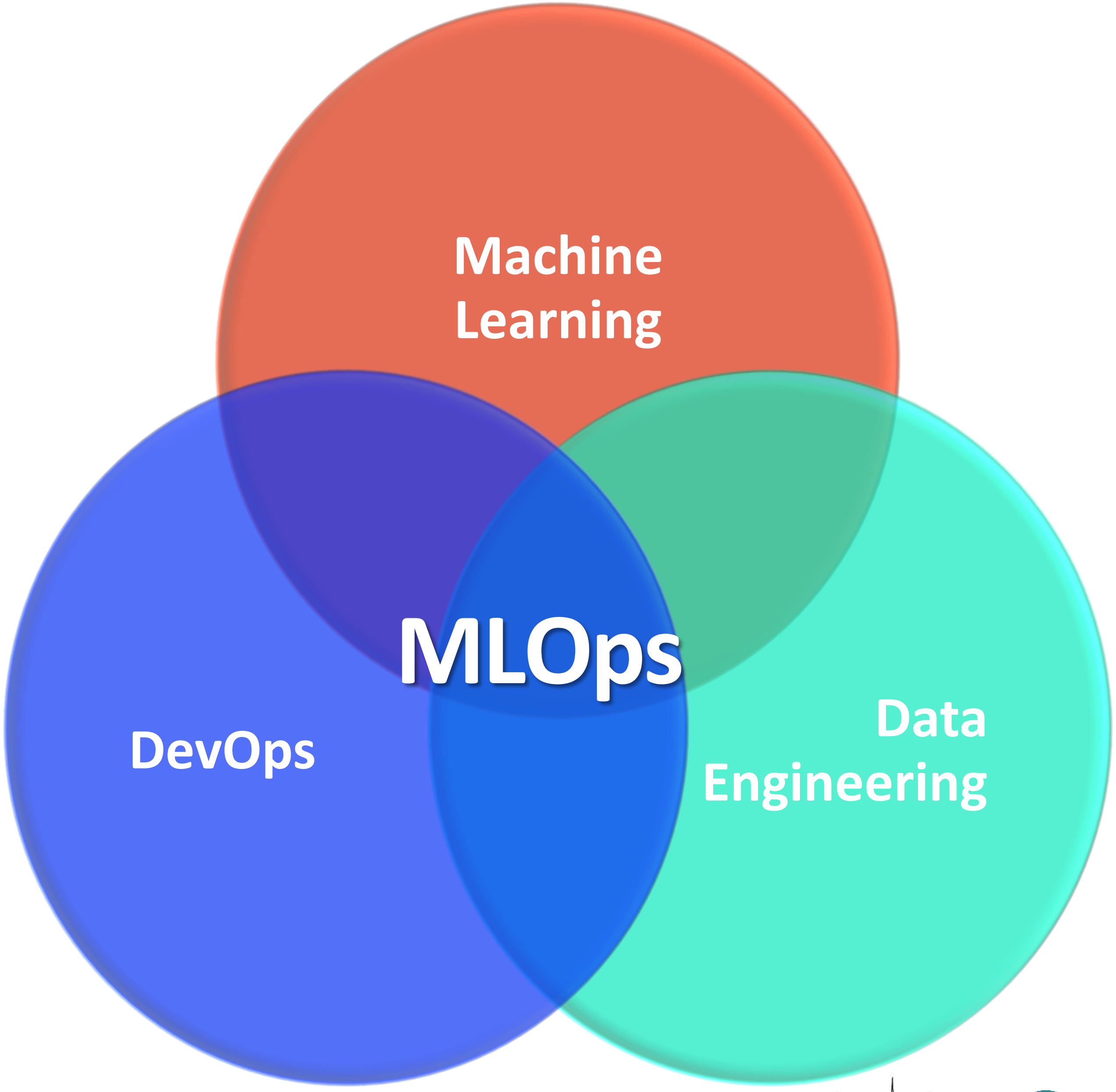
Microsoft

Azure Meetup TORINO

Deltatre Innovation Lab

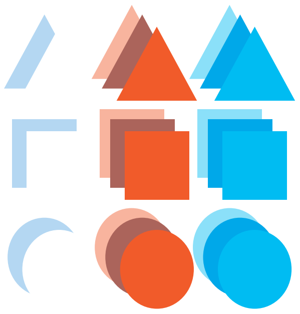

HOW TO GO TO
PRODUCTION?
MVP
Microsoft®
Most Valuable
Professional

MACHINE LEARNING AND MLOPS OVERVIEW

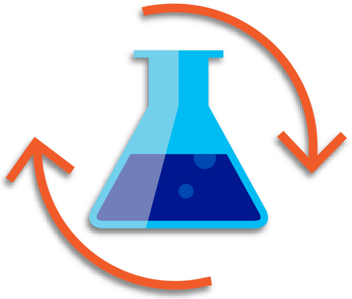


Microsoft Azure Meetup TORINO Deltatre Innovation Lab

TYPICAL ML WORKFLOW



Data preparation

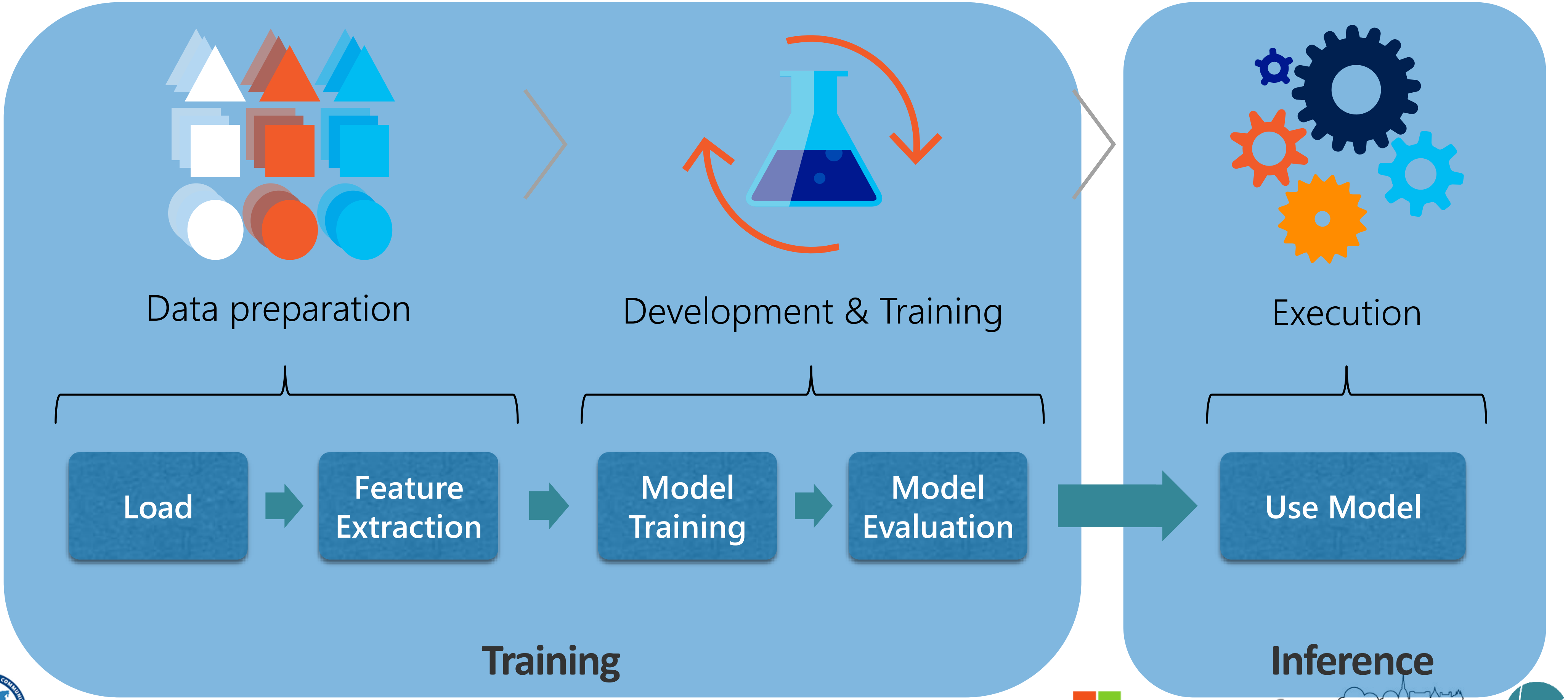


Development & Training

Execution

Load → Feature Extraction → Model Training → Model Evaluation → Use Model

TRAINING VS INFERENCE/SCORING



TRAINING

"Development" phase of a **Machine Learning project**

Usually **lead** and executed by **Data Scientists**
from start to end, **through experiments** and **trial & error**

Iterative process of **variable duration and results**
(until **specified target metrics are achieved**)

Typically, **resource & time intensive**
usually done on (lots of) CPUs and GPUs or AI-accelerators (TPU, FPGA)

Scalability ➔ reduce training time or **improve quality**
size of datasets and/or models + parallel processing = hardware/storage/bandwidth

Datasets

Training
+
Eval Loop

Trained
Model File

INFERENCE

"Production" phase of a **Machine Learning project**

Usually **lead** by **Software/AI Engineers** and **DevOps**

Data scientists' role is mainly to **monitor** model **behavior** in the field

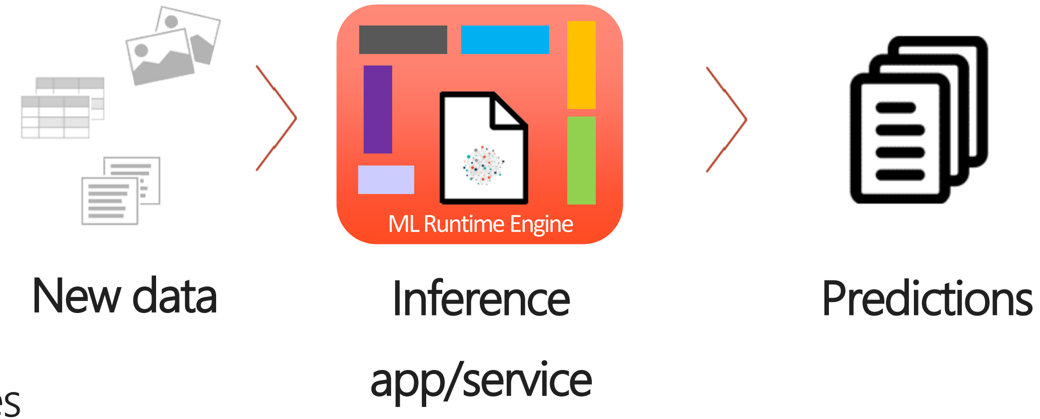
Different requirements, tools and frameworks

compared to training, more **similar** to **traditional development**

Deploy on **edge device** or **on-prem/cloud datacenter**

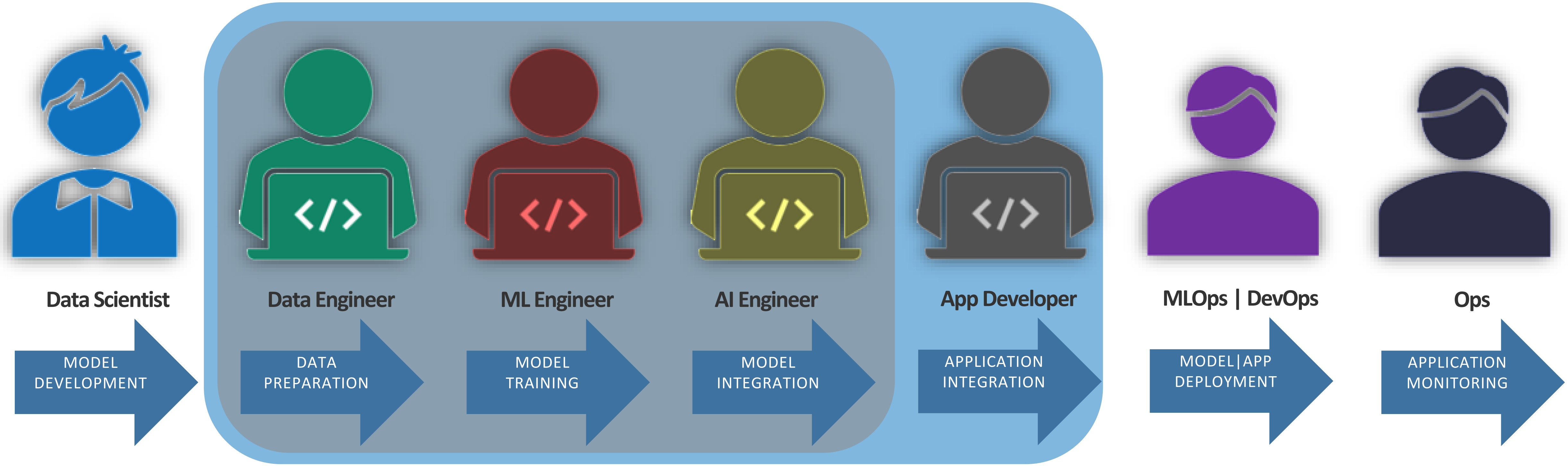
Scalability ➔ increase # of requests/sec

optimize models, parallel processing, increase/scale scoring instances



THE AI-TEAM

TYPICAL ML PROJECT TEAM



#GLOBALAZURE 2021

Microsoft

Azure Meetup TORINO

Deltatre Innovation Lab

MLOPS: DEVOPS ON ML COMPONENTS

Model reproducibility & versioning

Track, snapshot & manage assets used to create the model

Enable collaboration and sharing of ML pipelines

Model auditability & explainability

Maintain asset integrity & persist access control logs

Certify model behavior meets regulatory & adversarial standards

Model packaging & validation

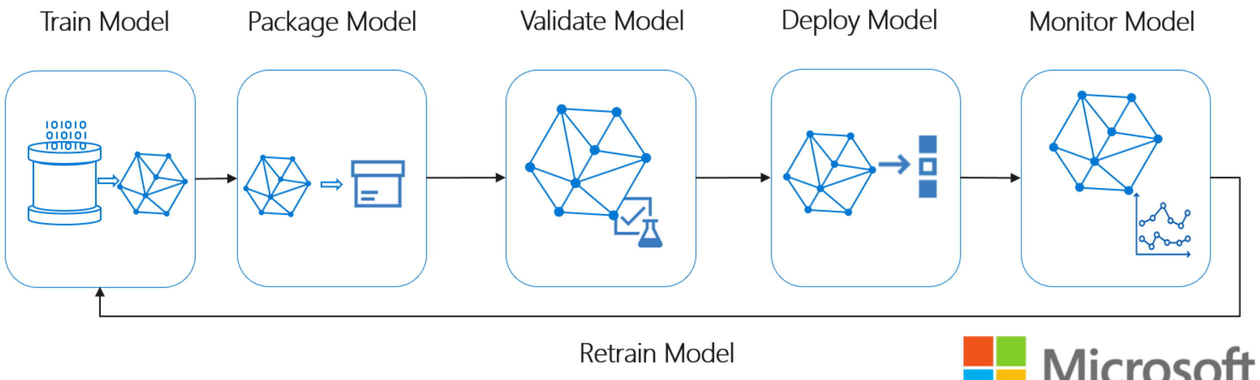
Support model portability across a variety of platforms

Certify performance meets functional and latency requirements

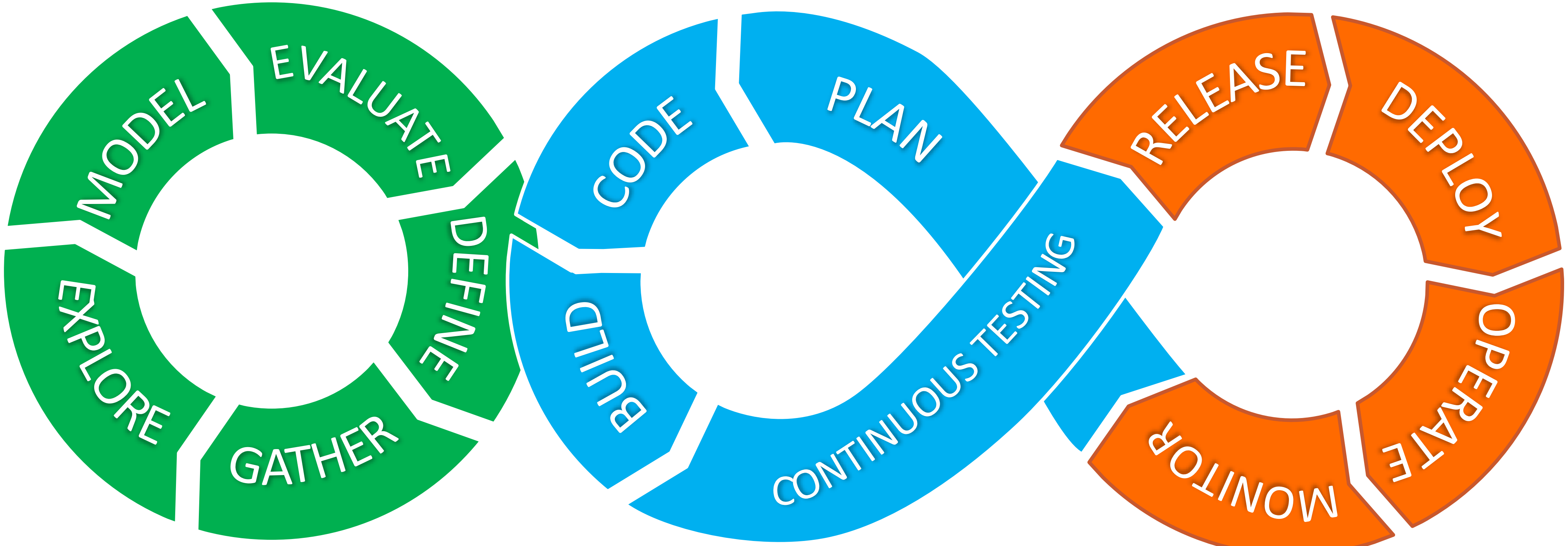
Model deployment & monitoring

Release models with confidence

Monitor & know when to retrain by analyzing signals such as data drift



THE FORGOTTEN EXPLORATION PHASE



ML PROJECT NEEDS AND REQUIREMENTS

Hybrid infrastructure → On-premise + Cloud
Data and/or computation on-premise (or other cloud providers)

Data security and IP protection

Use existing infrastructure capacity for AI/ML

Cognitive Services + Azure ML "enterprise-ready" features
(infrastructure, security, monitoring, etc.)

ML/DL WORKFLOWS

THE BUILDING BLOCKS OF “USABLE” MACHINE LEARNING

Trigger Endpoint

Monitoring

Scoring

Pipeline workflow

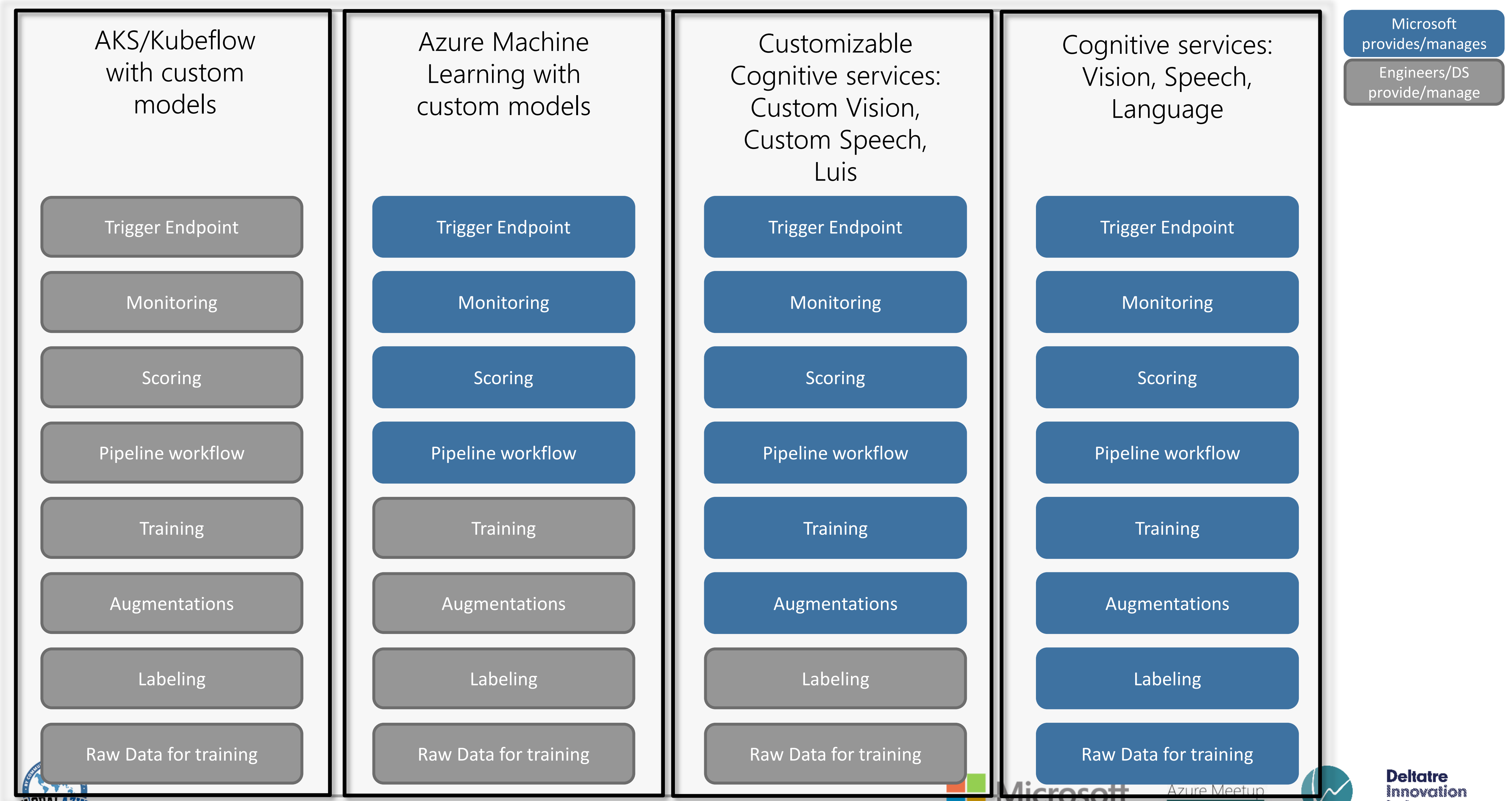
Training

Augmentations

Labeling

Raw Data for training

- **First of all,** you need **DATA** (possibly good data)
- Most ML problems can’t be solved without **Labeling**
- **Augmentations or processing** make data usable across different sources for our chosen algorithm
- **Most ML problems** can’t be solved with just one ML model. **Pipelines** and **Workflows** are key to chain **transformations** and **model evaluations and results**
- In order to use and maintain a ML pipeline proper **MLOps** and results **Monitoring** must be in place


AKS/Kubeflow with custom models
Trigger Endpoint
Monitoring
Scoring
Pipeline workflow
Training
Augmentations
Labeling
Raw Data for training
Azure Machine Learning with custom models
Trigger Endpoint
Monitoring
Scoring
Pipeline workflow
Training
Augmentations
Labeling
Raw Data for training
Customizable Cognitive services: Custom Vision, Custom Speech, Luis
Trigger Endpoint
Monitoring
Scoring
Pipeline workflow
Training
Augmentations
Labeling
Raw Data for training
Cognitive services: Vision, Speech, Language
Trigger Endpoint
Monitoring
Scoring
Pipeline workflow
Training
Augmentations
Labeling
Raw Data for training
Microsoft provides/manages
Engineers/DS provide/manage

AZURE COGNITIVE SERVICES

Vision
recognize, identify, caption, index, and understand what is in your pictures or videos.

Language
NLP tasks such as evaluate sentiment and learn how to recognize what users want.

Speech
convert speech into text vice versa, translate between languages, speaker verification and recognition

Decision
system that helps to give best recommendations for informed and efficient decision-making.

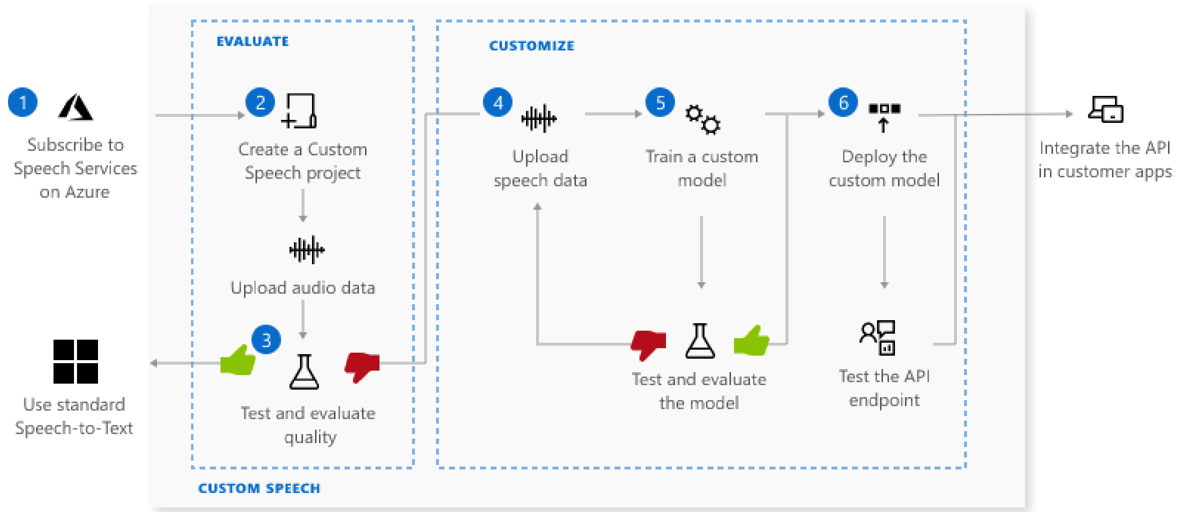
Search
search-APIs that gives you the ability to comb billions of contents with a single API call

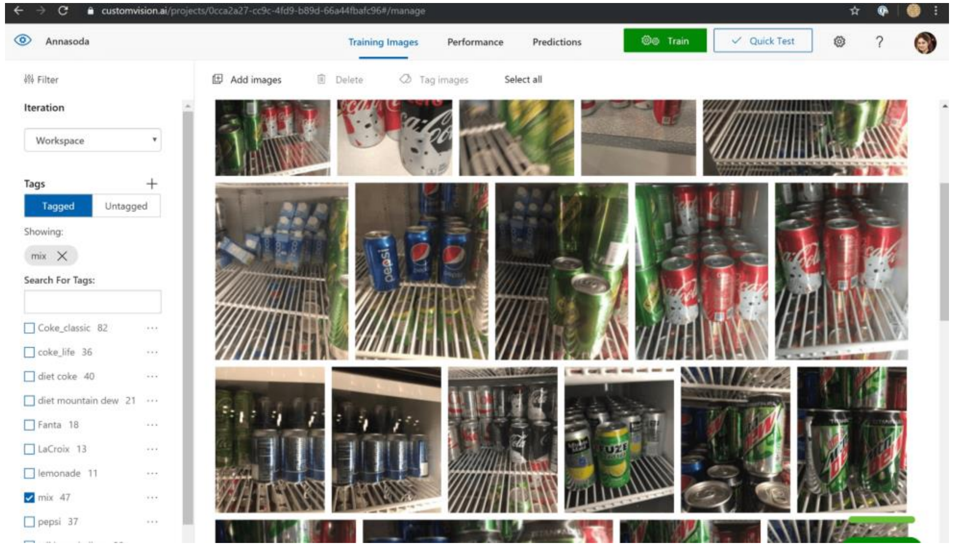
Cloud-based services to help you build cognitive intelligence into applications with very little knowledge of AI/ML or data science:

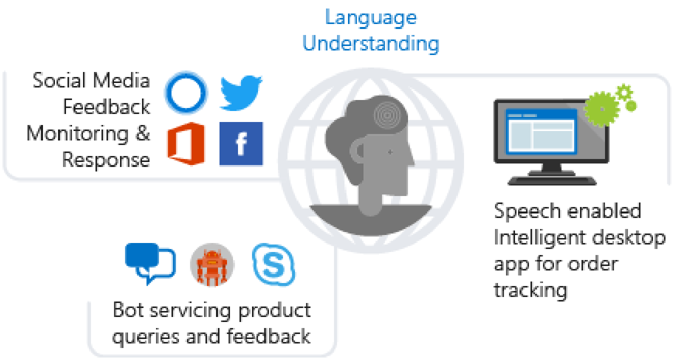
Available with REST APIs and client library SDKs available

They comprise various **AI services** such as computer vision, audio processing and speech processing and understanding

CUSTOMIZABLE VISION, SPEECH AND LANGUAGE







Still **developer** “friendly” ML services with ways to customize the training dataset to create *domain specific* ML models

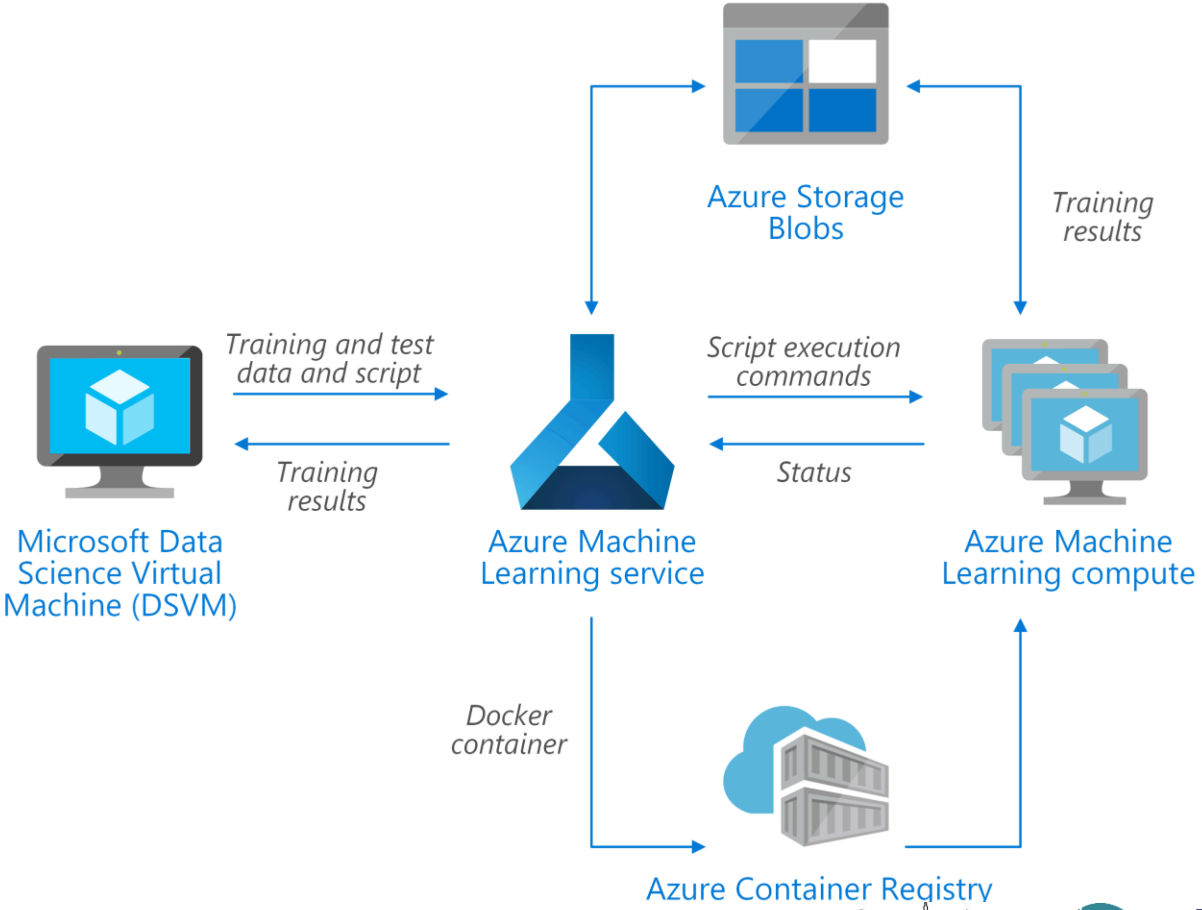
Microsoft Azure Meetup TORINO Deltatre Innovation Lab

AZURE MACHINE LEARNING
(MANAGED WORKFLOW, PIPELINE AND STREAMLINED MLOPS)

AML is a cloud-based environment that accelerates the **end-to-end machine learning lifecycle**.

You can use AML to:

- Train
- Deploy
- Automate
- Manage
- Track ML models



https://azure.microsoft.com/services/machine-learning/

DOING DATA SCIENCE WITH AML

Develop on your **local machine** using the AML **Python SDK** or R SDK

Open-source frameworks such as PyTorch, TensorFlow, scikit-learn, and others

Tools for ML workflows, including:

- **Jupyter notebooks**: widely used in the **exploratory phase** by data scientists because they allow to easily create, view, and share code
- **Azure Machine Learning CLI**: extension to automate ML activities in AML such as run experiments or deploy models

AML Studio: web portal combining no-code and code-first experiences for data science

Languages

Development tools

Frameworks

COGNITIVE SERVICES VS AML

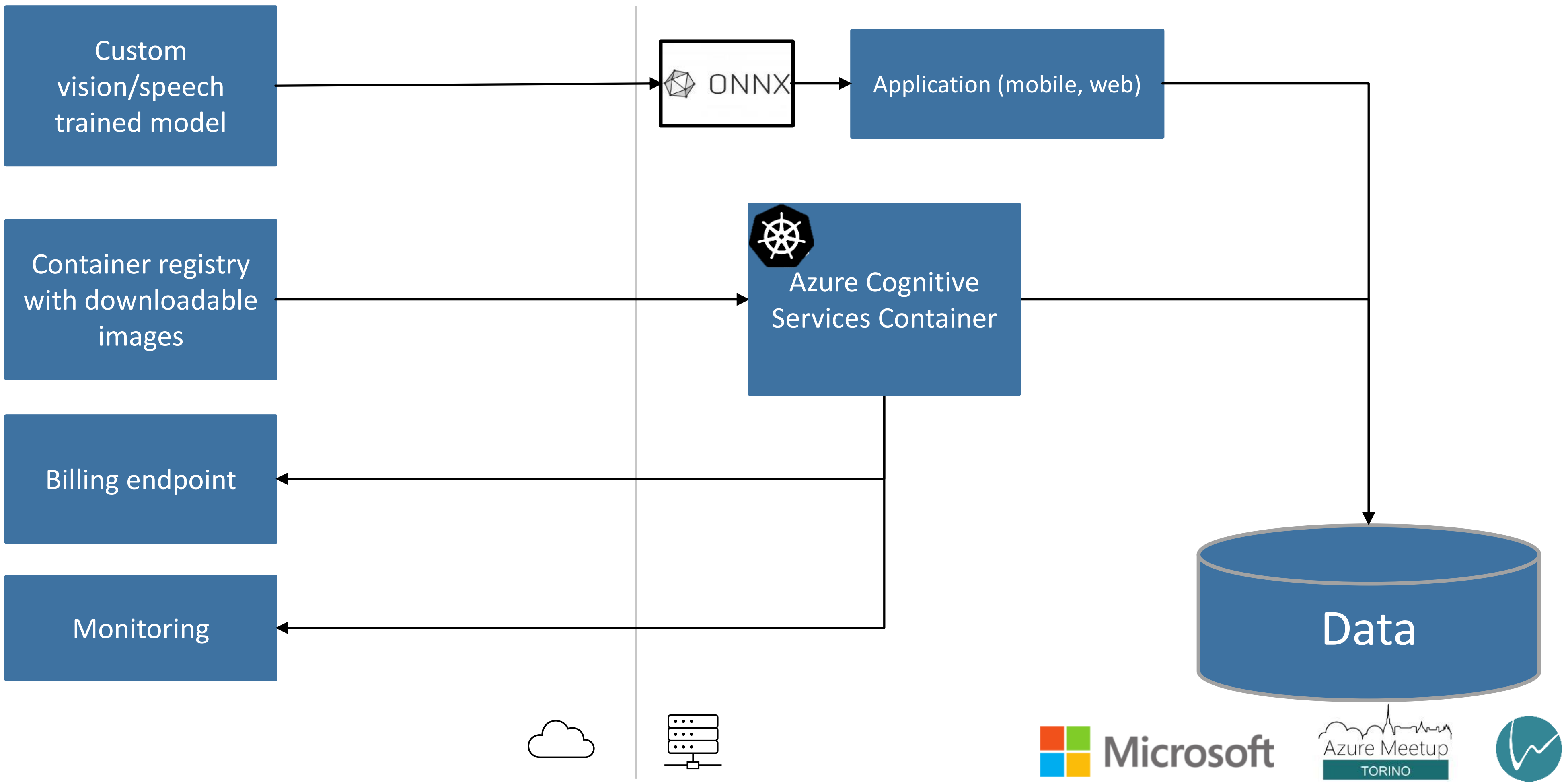
Cognitive Services are for developers without machine-learning experience

- A Cognitive Service provides a (general-purpose) trained model, made available using a SaaS REST API or an SDK.
- Services can be used and integrated within minutes, depending on the scenario.
- Cognitive Service readiness is ideal for who has **no AI knowledge and deals with general problems**.

Azure Machine Learning is tailored for data scientists

- AML works for highly specialized or specific problems, often requires suiting all ML operations (data collections, cleaning, training, evaluating, …).
- Model implementation may require weeks, if not months, and engineering, maintaining and serving them requires infrastructure + software engineering + data science skills. **Familiarity and expertise with data science are required**

DOES IT ALL NEED TO BE IN THE PUBLIC CLOUD?

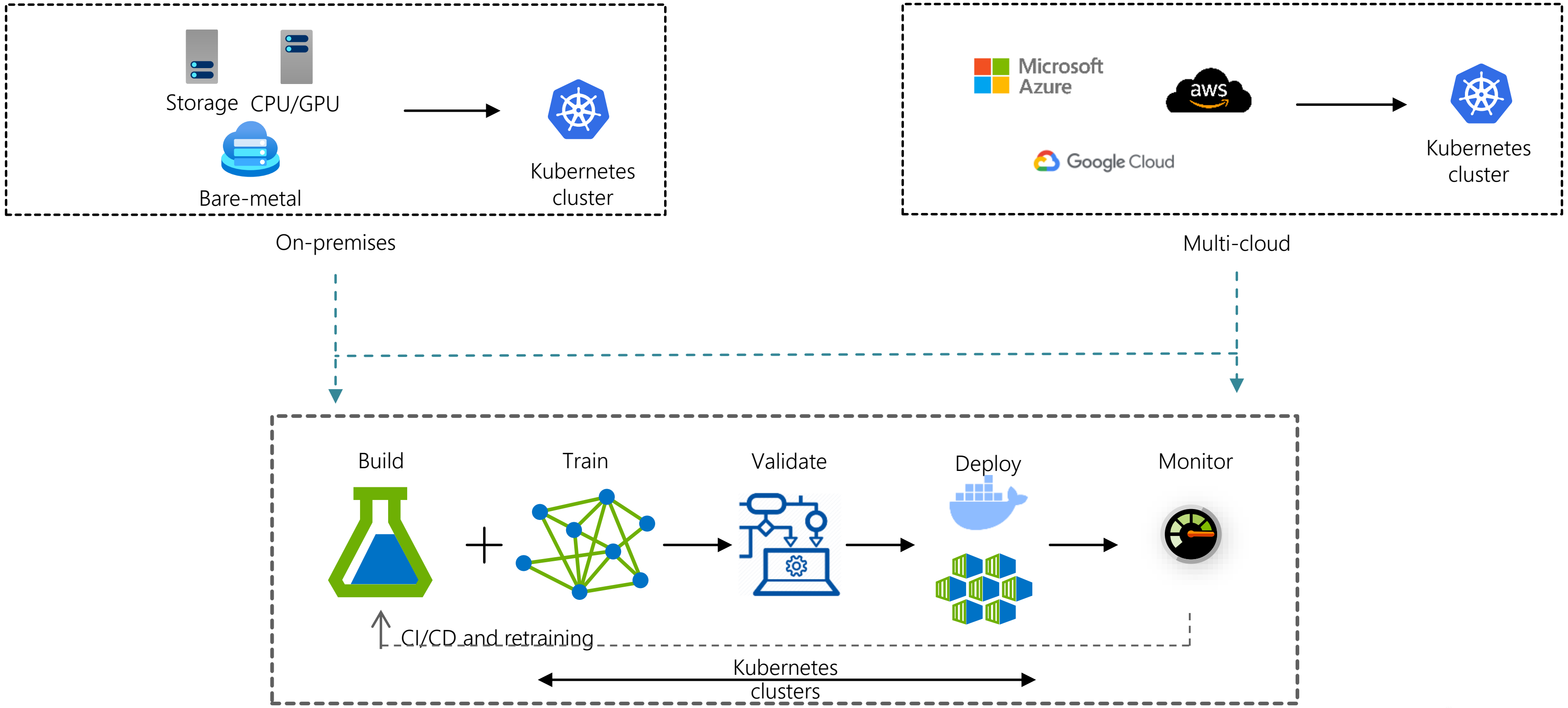


Microsoft

Azure Meetup TORINO

Deltatre Innovation Lab

HYBRID SCENARIO (ON-PREMISE + CLOUD)



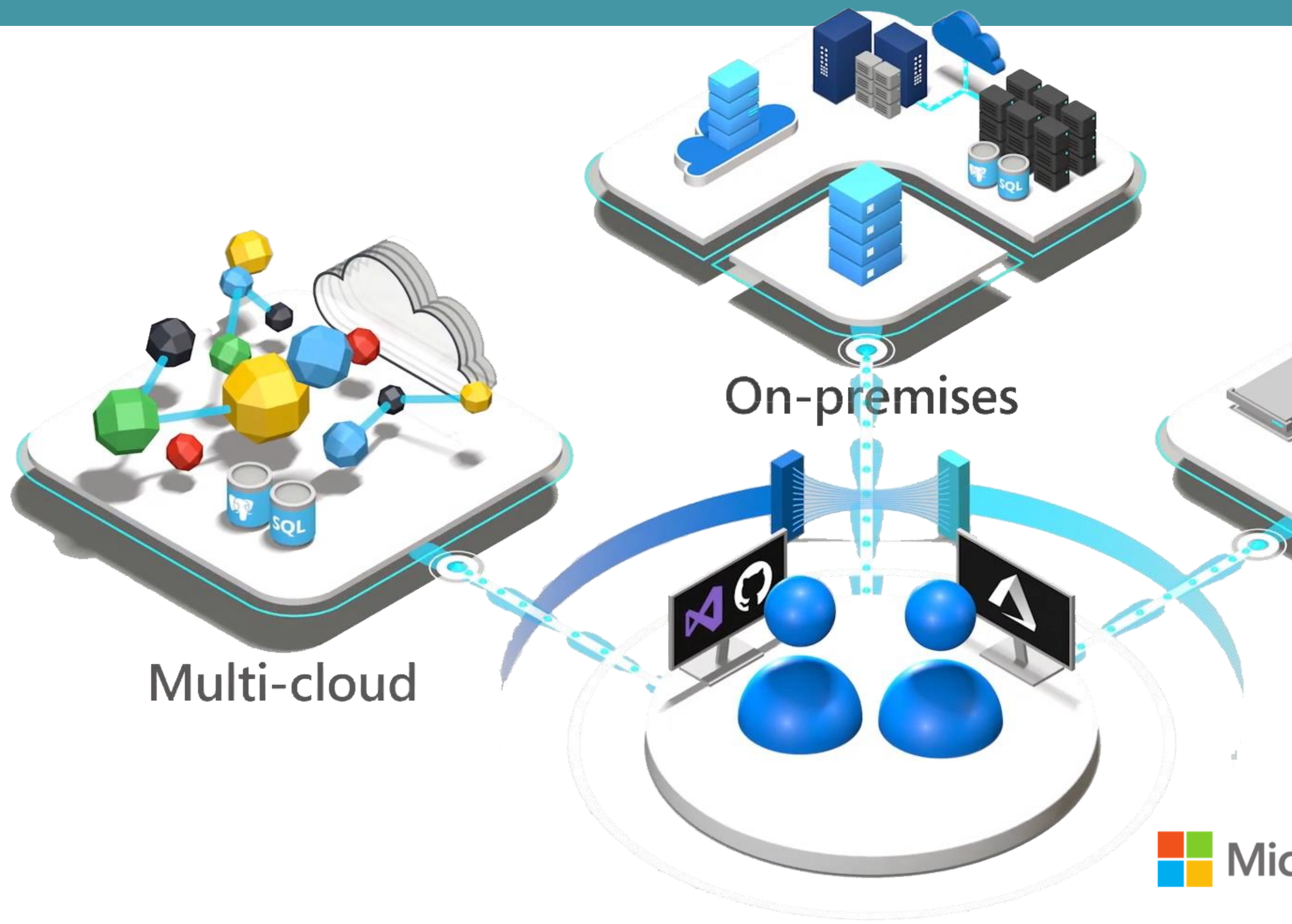
Azure Machine Learning

Microsoft

Azure Meetup TORINO

Deltatre Innovation Lab

AZURE ARC ENABLED ML



MACHINE LEARNING PIPELINES

What are AML pipelines?

- **AML Pipeline** is an independently executable workflow of a **complete ML task**. Subtasks are encapsulated as a **series of steps** within the pipeline.
- **Step**: discrete processing action designed for a specific task with its own resources and environment
- AML **automatically orchestrates** all the dependencies between pipeline steps. This may include:
  - Spinning up and down Docker images
  - Attaching and detaching compute resources
  - Moving data between steps consistently and automatically

Pipeline benefits

- Simplicity
- Speed
- Repeatability
- Flexibility
- Modularity
- Versioning and tracking
- Quality assurance
- Cost control
- Controlled Runtime Environment

WHAT TYPE OF PIPELINES?

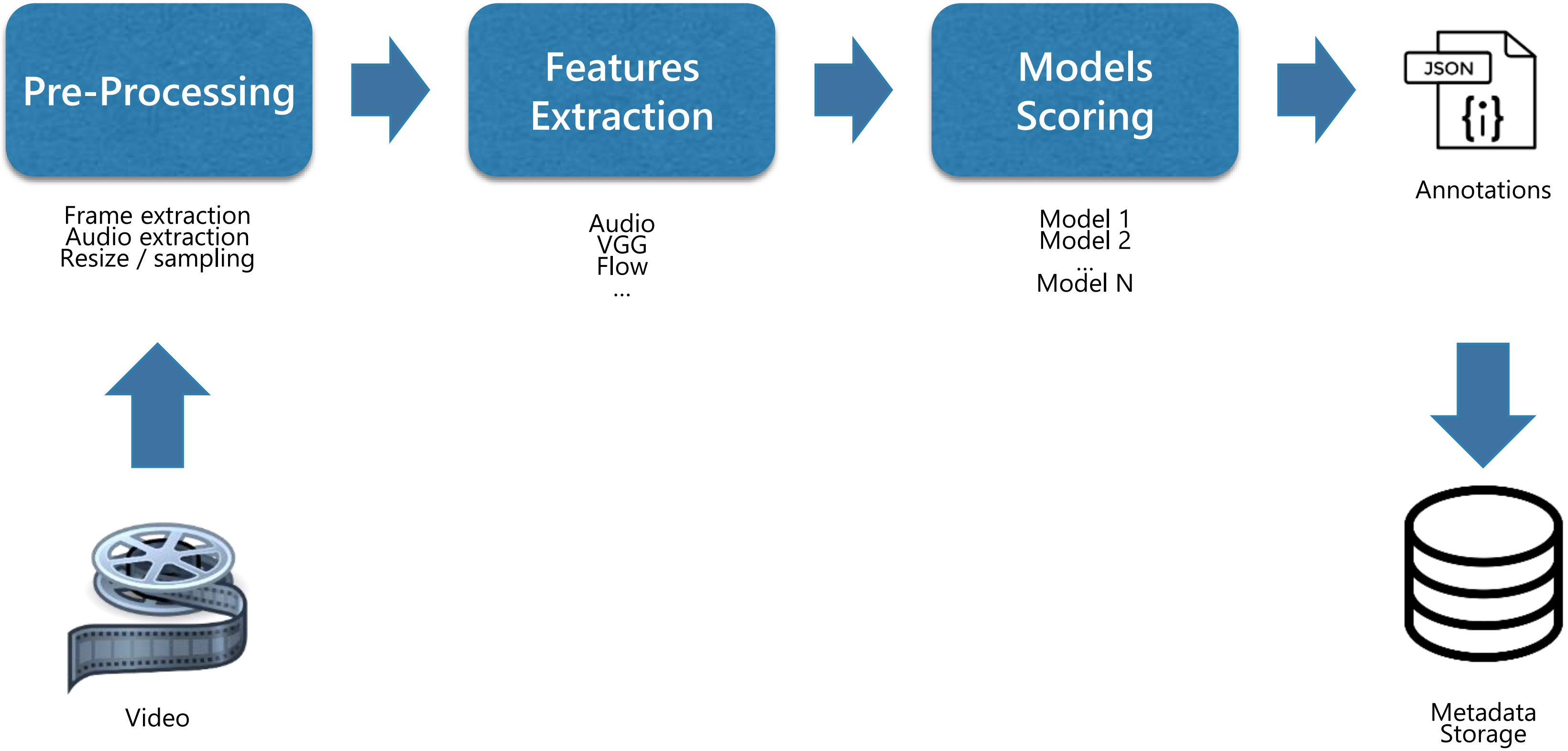
DevOps CI/CD Pipelines
Related to implement CI/CD processes. They check:

- Code quality
- Makes sure that any changes in the repository are deployed to the AML Service once they are committed/tested/approved, etc.

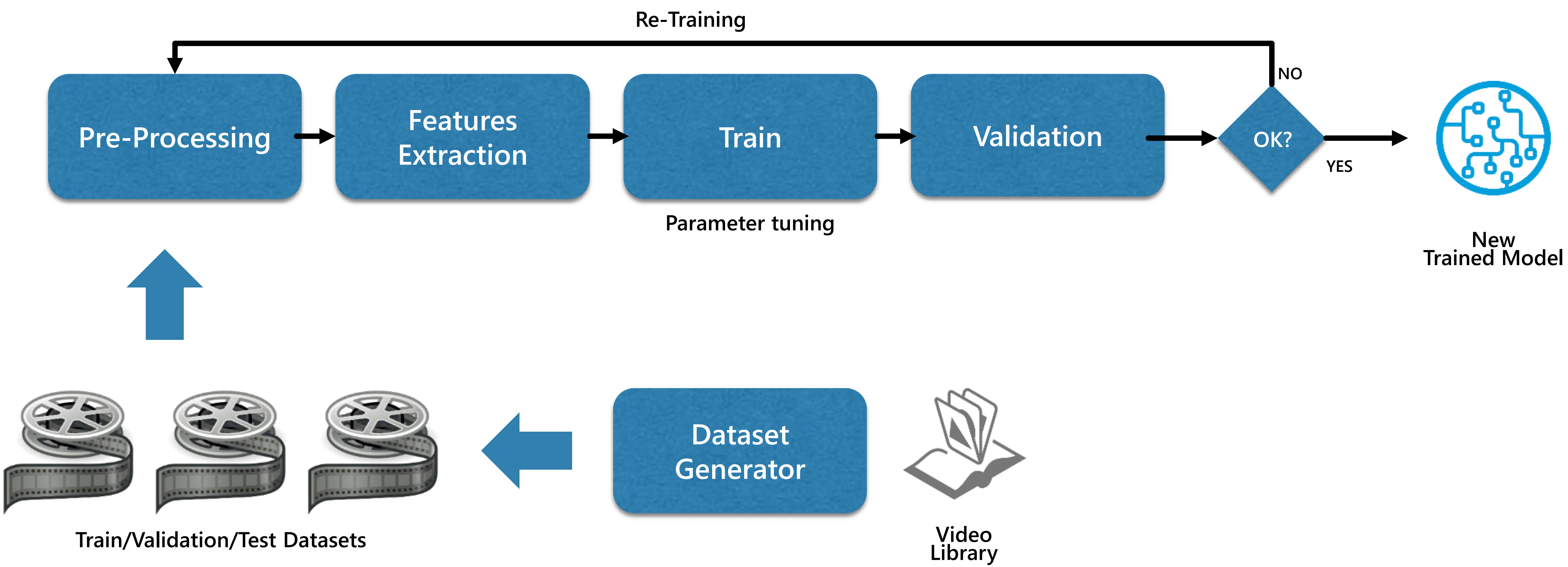
Azure Machine Learning Pipelines
Related to entities in the AML service such as

- Training models
- Scoring models
- Data augmentation pipelines

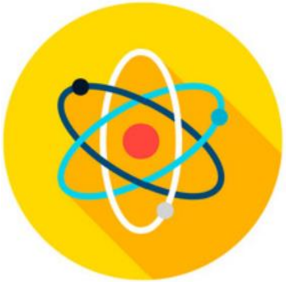
INFERENCE PIPELINE FOR VIDEO INDEXING



TRAINING PIPELINE FOR VIDEO/AUDIO MODELS



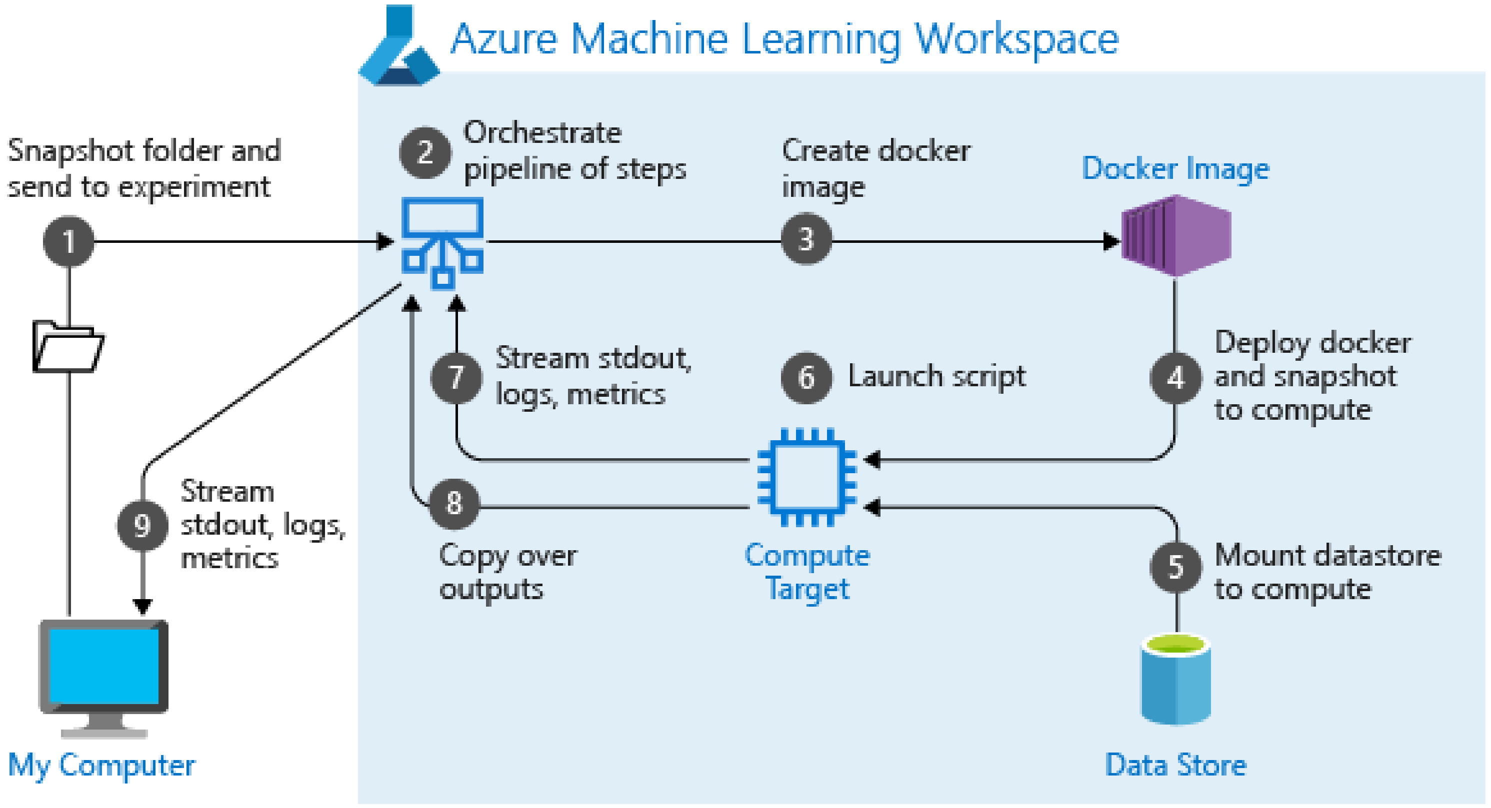
AML PIPELINES PHASES



Build

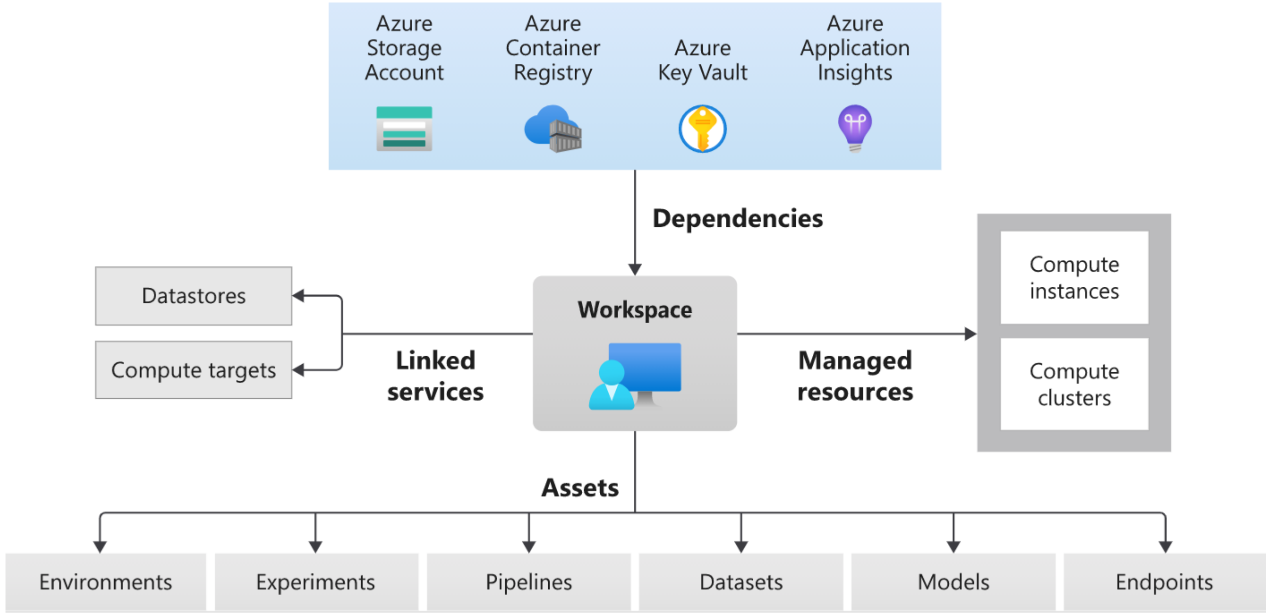
Publish

Run



AML PIPELINE KEY ELEMENTS

- **Workspace**: top-level resource for AML. Contains all the artifacts created when the user utilizes AML
- **Environment**: specify the Python packages, environment variables, and software settings (Python, Spark, Docker)
- **Compute targets**: machines or clusters that perform the computational steps in ML pipeline
- **Storage**: reference to datastore. Many service types (i.e., Azure blob container, Azure file share, …)



AML PIPELINE PHASES

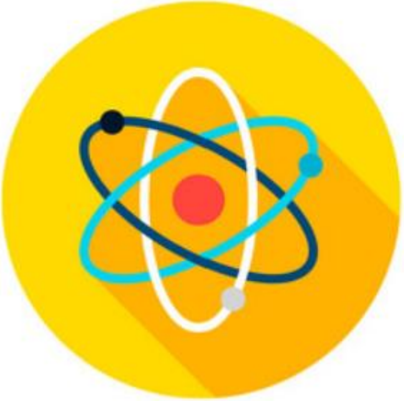
Building a pipeline

Build an AML pipeline implies to configure and set up the proper local environment. Therefore, you must:

- **Create and configure an Azure ML workspace** (WS)
- **Set up the compute target**: that means also to define the Python environment and package dependencies needed by your script.
- **Set up the machine learning resources**: where data is stored and how to access to it
- **Register a Model with the WS** to use it inside a step
- **Authenticate to Azure services** and call your environment for local developing
- **Define a pipeline** as sequence of PipelineStep objects and use DataReference to point the data and PipelineData to introduce a data dependency between steps and for creating an implicit execution order in the pipeline.

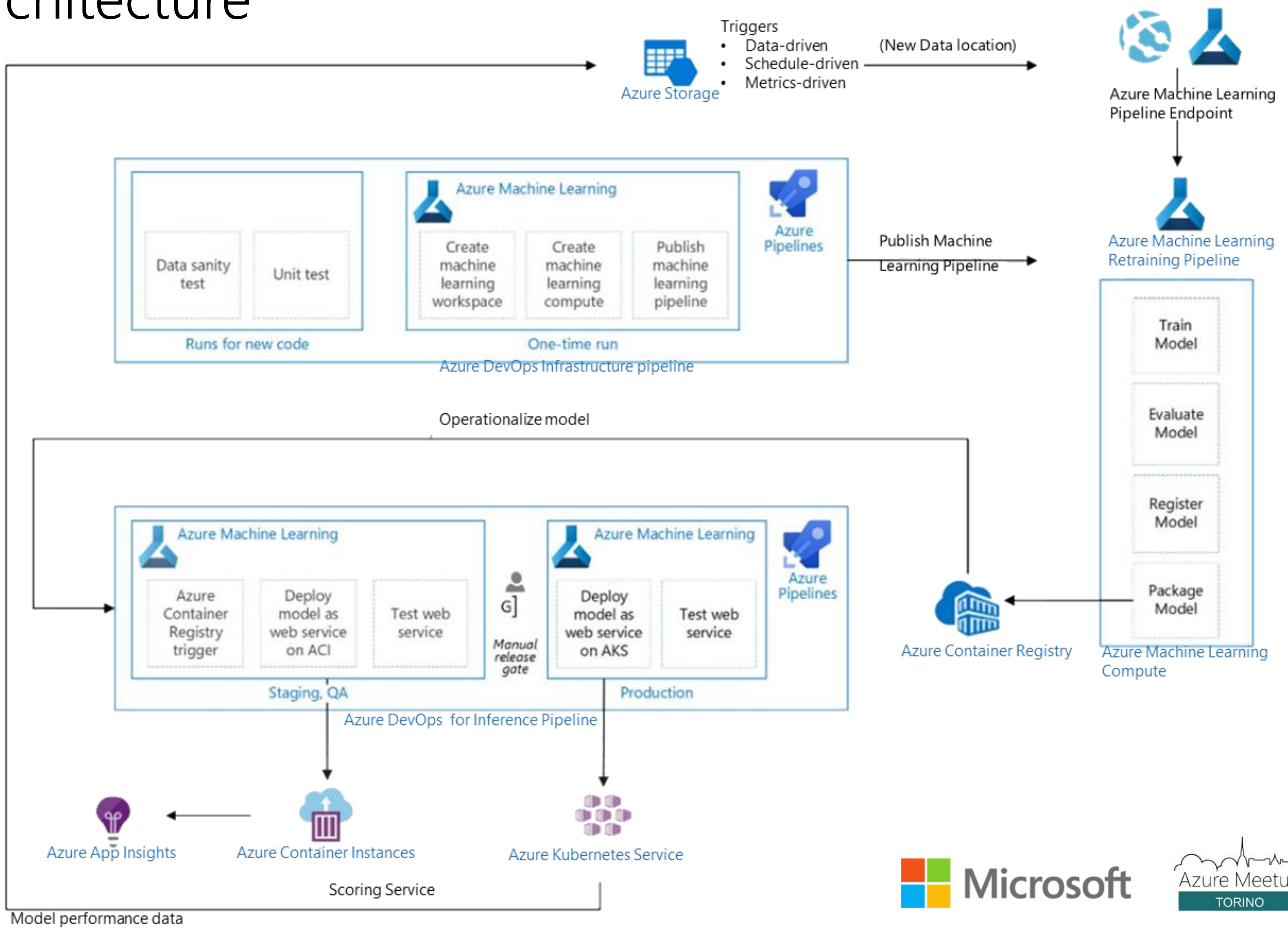
Publishing a pipeline

Publishing a pipeline at the end of the building script, allows you to **run the pipeline more times with different inputs**.



MLOPS & PIPELINES

DevOps architecture



Microsoft

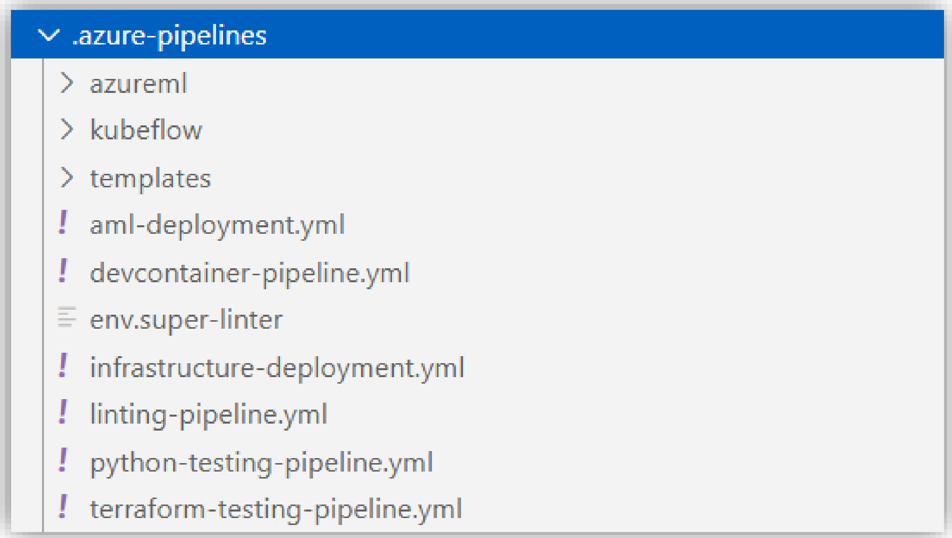
Azure Meetup TORINO

Deltatre Innovation Lab

DEVOPS, MLOPS, CI/CD

YAML Azure Pipelines + Variable groups + Multi-stage Docker build files

- **Build validators** (linting, Python tests, Terraform tests)
- **Azure ML Pipelines CI/CD** (build & publish & run)
- **Infrastructure deployment** (Terraform + scripts + testing)



USEFUL TIPS

DATASET MANAGEMENT

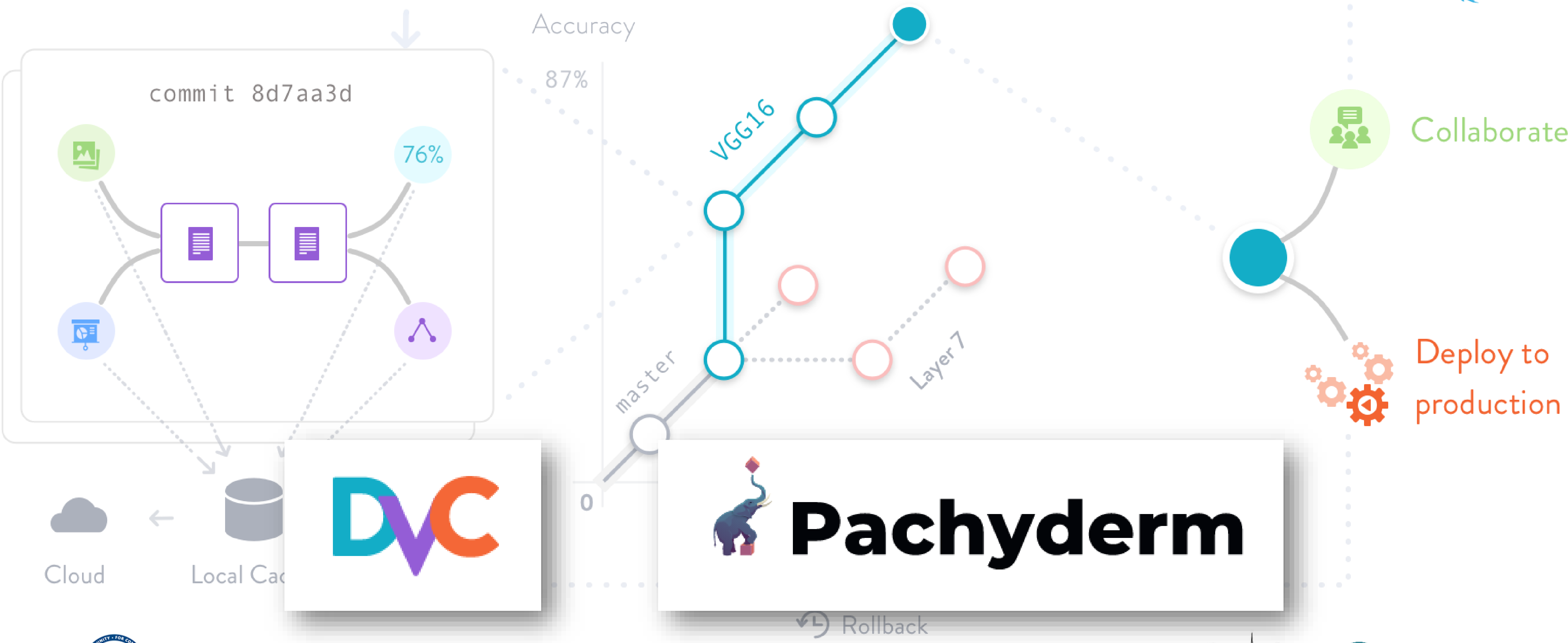


Image taken from DVC.org

PROJECT STRUCTURE

Cookie cutter
a cross-platform logical, simple, but flexible project structure to carry out and share work

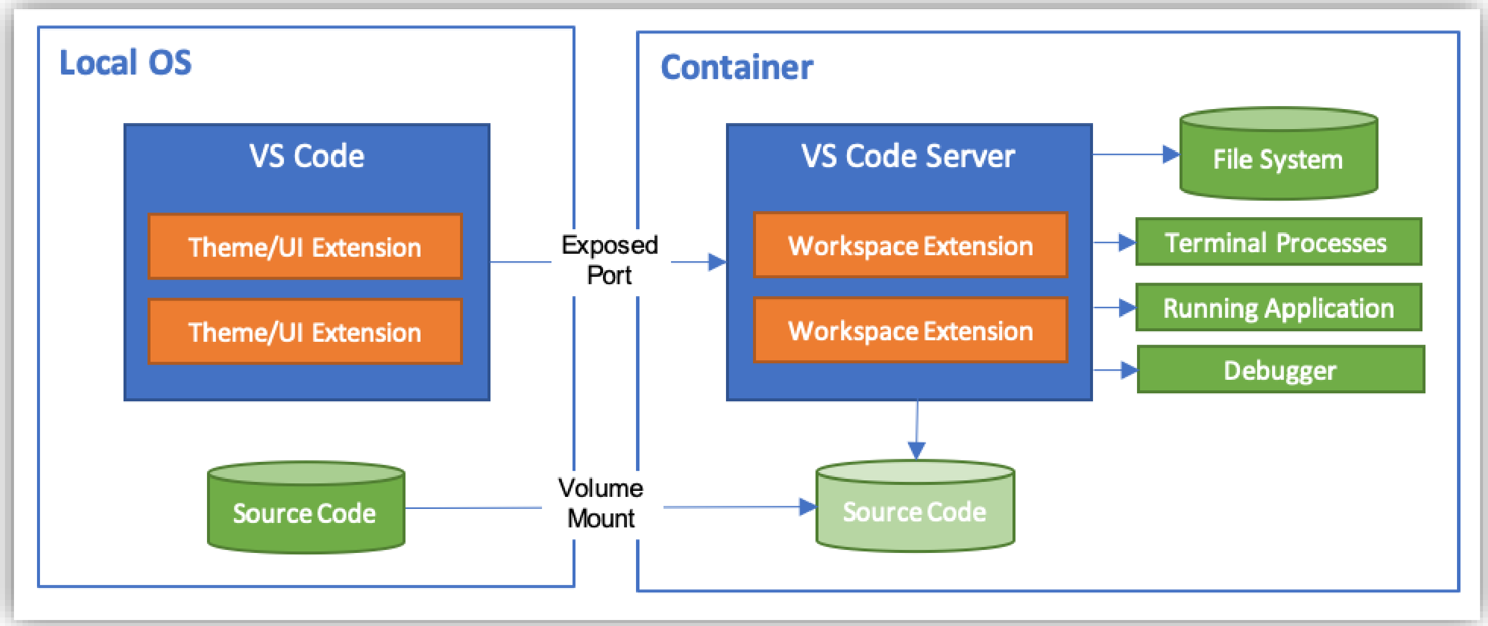
Documentation
strong, detailed and meaningful documentation for all the project phases

Testing
Unit and Integration tests are done using PyTest for Python. When possible, acceptance criteria should be verified with acceptance tests.

- .devcontainer
- .pipeline
- data
- docs
- envutils
- ml_service
- models
- notebooks
- reports
- src
- test
- terraform
- .env
- ReadMe.md
- requirements.txt

DEV ENVIRONMENT IN CONTAINERS

Remote Development extension pack for **Visual Studio Code**
→ Use a **Docker container** as a **full-featured development environment**
→ Use the **same Docker image** for all dev team (local machine, remote machine or cloud VMs)



Exploration

Data Preparation, Dataset Augmentation,
Jupyter Notebooks, Model Definition,
Toy Datasets, Hyperparameters, Model Tuning,
Output Metrics Evaluation, … a lot of coffee

Engineering & MLOps

Machine Learning Pipelines Definition,
Code Re-engineering, Automatic Testing,
CI/CD Pipelines, Performance Testing

Production Ready

Infrastructure as Code,
End to End Testing with real datasets,
Performance Tuning, Data Security,
Infrastructure Security, App/Service Integration,
Monitoring

Thank You!

ευχαριστώ Salamat Po متشکرم شكراً Grazie

благодаря ありがとうございます Kiitos Teşekkürler 谢谢

ขอบคุณครับ Obrigado شکریہ Terima Kasih Dziękuję

Hvala Köszönöm Tak Dank u wel дякую Tack

Mulţumesc спасибо Danke Cám ơn Gracias

多謝晒 Ďakujem תודה நன்றி Děkuji 감사합니다

REFERENCES (1/2)

Tools and IDE

https://www.python.org/
https://jupyter.org/
https://code.visualstudio.com/

ML/Deep Learning Services & Frameworks

https://azure.microsoft.com/services/machine-learning/
https://azure.microsoft.com/services/cognitive-services/
https://azure.microsoft.com/services/azure-arc/
https://www.tensorflow.org/
https://keras.io/
https://pytorch.org/
https://fast.ai/

REFERENCES (2/2)

MLOps

https://mlflow.org/
https://www.kubeflow.org
https://dvc.org/
https://www.pachyderm.com/
https://github.com/microsoft/MLOps
https://docs.microsoft.com/azure/machine-learning/team-data-science-process
https://docs.microsoft.com/azure/architecture/reference-architectures/ai/mlops-python
https://docs.microsoft.com/azure/architecture/reference-architectures/#ai-and-machine-learning

VSCode DevContainers

https://github.com/polatengin/project-standards/blob/master/DevContainers.md
https://code.visualstudio.com/docs/remote/containers
https://channel9.msdn.com/Series/Beginners-Series-to-Dev-Containers

About us

Ing. Gianni ROSA GALLINA

R&D Senior Software Engineer @ **Deltatre**

@giannirg

- AI, Machine Learning, Deep Learning on multimedia content
- Virtual/Augmented/Mixed Reality
- Immersive video streaming & 3D graphics for sport events
- Cloud solutions, web backends, serverless, video workflows
- Mobile apps dev (Windows / Android / Xamarin)
- End-to-end solutions with Microsoft Azure

MVP Microsoft Most Valuable Professional

Microsoft Specialist
Programming in C#
Programming in HTML5 with JavaScript & CSS3

Microsoft CERTIFIED Solutions Developer
Windows Store Apps Using C#
Web Applications

Microsoft Certified Azure AI Engineer Associate

Microsoft Certified Azure Data Scientist Associate

#GLOBALAZURE 2021

unity Certified Programmer

PLURALSIGHT Author

https://gianni.rosagallina.com

Microsoft

Azure Meetup TORINO

Deltatre Innovation Lab

About us

Vito Flavio Lorusso

Program Manager @ **Microsoft**

@vflorusso

- Former web developer
- Former data engineer and developer
- Former “doing Database cluster installations in datacenters”
- Former Solutions Architect
- Former “cloud evangelist”
- Former Distributed systems engineer
- Constantly looking for my place in the digital world to help work get done

#GLOBALAZURE

Microsoft Azure Meetup TORINO Deltatre Innovation Lab

About us

Clemente Giorio

R&D Senior Software Engineer @ **Deltatre**

@tinux80

- Augmented/Mixed/Virtual Reality
- Artificial Intelligence, Machine Learning, Deep Learning
- Internet of Things
- Embedded Apps
- Multimodal Tracking

Author