Déployer une stack MLOps sous Azure : *gagnez progressivement en maturité !*

Meetup Data Cloud Paris

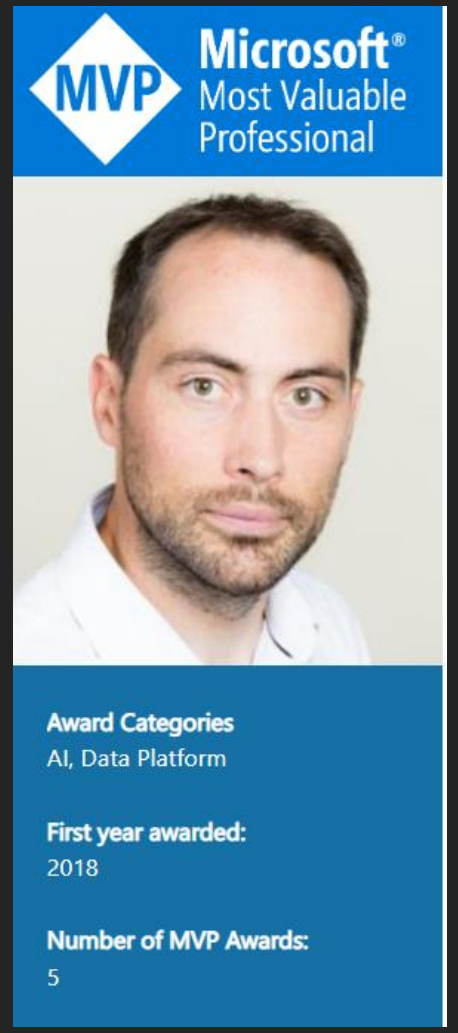
Mercredi 12 octobre 2022

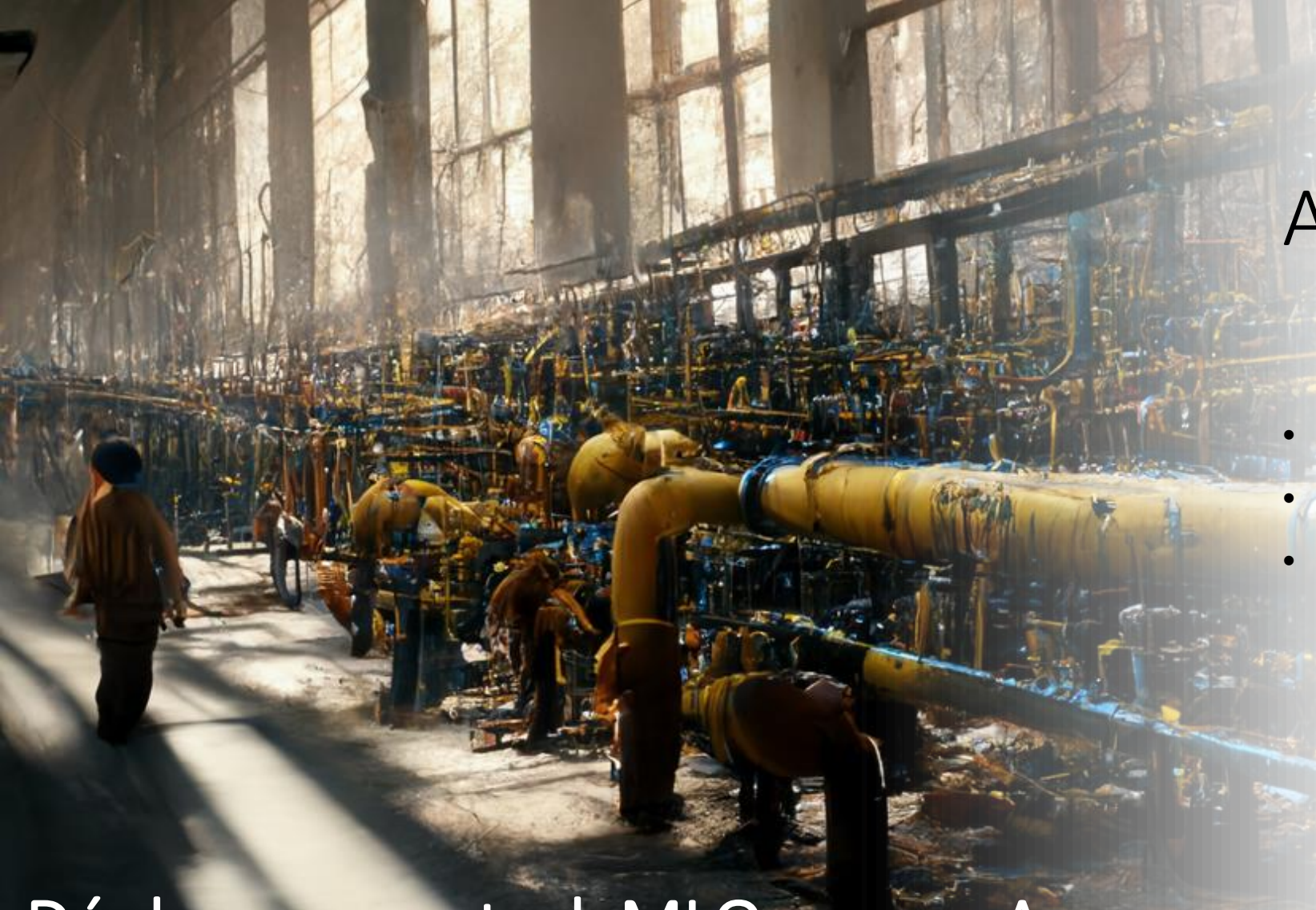
DALL-E : *gradually gain in maturity*

Speaker : Paul PETON

- Tech Lead Intelligence Artificielle
- Data Platform & AI MVP since 2018
- Co-animateur du podcast BigData Hebdo

- GitHub.com/methodidacte/meetups





Déployer une stack MLOps sous Azure :

gagnez progressivement en maturité

people dealing with process on a platform (DALL.E)

Agenda

- MLOps dans une coquille de noix
- Les niveaux de maturité du MLOps
- Une stack MLOps sur Azure
  - Par où commencer ?
  - L'architecture sous-jacente
  - Distinguer les familles de composants
  - Piocher dans la « boîte à outils »
  - Utiliser plusieurs environnements

MLOps dans une coquille de noix

- Pourquoi industrialiser ?
  - Pour **diminuer les coûts** (d'opérations manuelles, d'erreurs…)
  - Pour **augmenter la qualité** (par des opérations testées et automatisées)
  - Pour **réduire le temps de mise en production** (*time to market*)
- MLOps = ML + DevOps, mais pas que
  - ***Framework* complet** : design + model development + operations
  - Des niveaux de **maturité**
  - Une ***stack*** construite à partir d'un ***tooling*** vaste et mouvant
  - Sites de référence :
    - https://ml-ops.org/
    - https://mlops-guide.github.io/
- Une **complexité forte** liée à la nature de l'apprentissage automatique
  - Notion de **reproductibilité**
  - Notions de **dérive** (*drift*)

machine learning operations

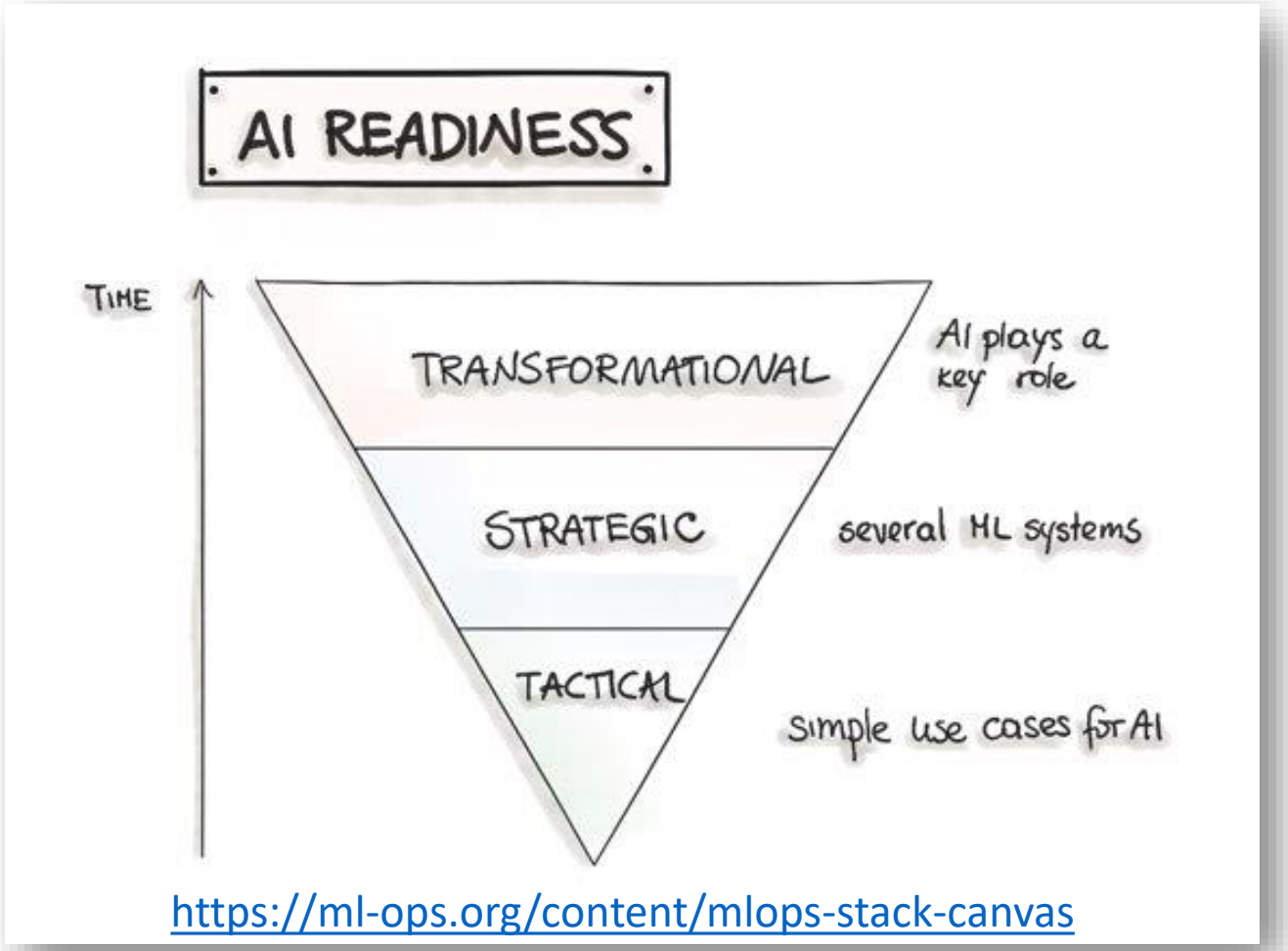

78:18
PAR 1 | 1 BEN
DIRECT RMC SPORT 1
TOP COACH
Il n'y a pas de certitude dans
dans le MLOps.

Les points de tension à résoudre

- La **collaboration** entre les Data Scientists
- Le temps passé et les efforts pour **déployer** les modèles
- La non **performance** des modèles lors de l'inférence
- La **gouvernance** des modèles et des données d'entrainement

tensions to be resolved

What is your AI readiness ?



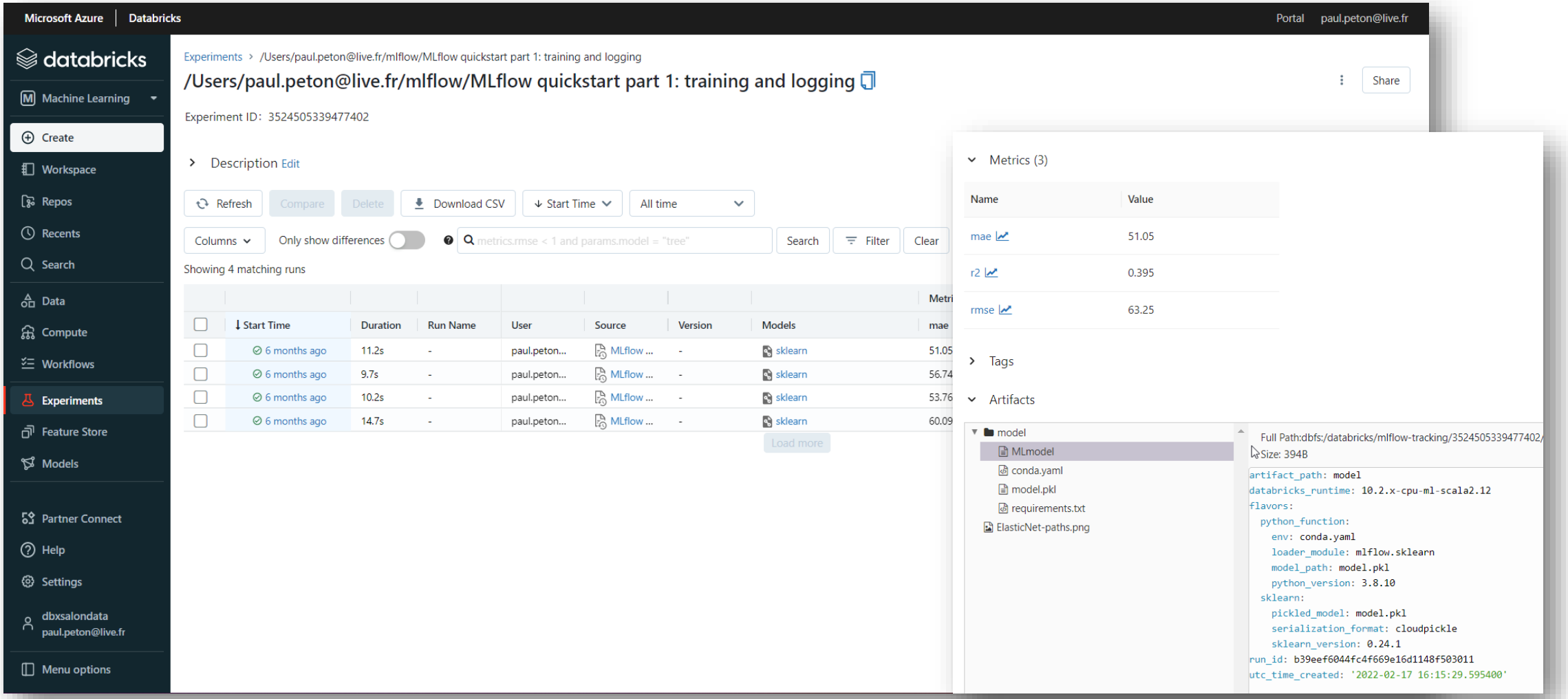
https://ml-ops.org/content/mlops-stack-canvas

Tracking

Maturité – niveau 0

- Suivre les **performances** des entrainements
- Assurer le **linéage** “code – data – model”

Tracking et versioning des modèles avec mlflow

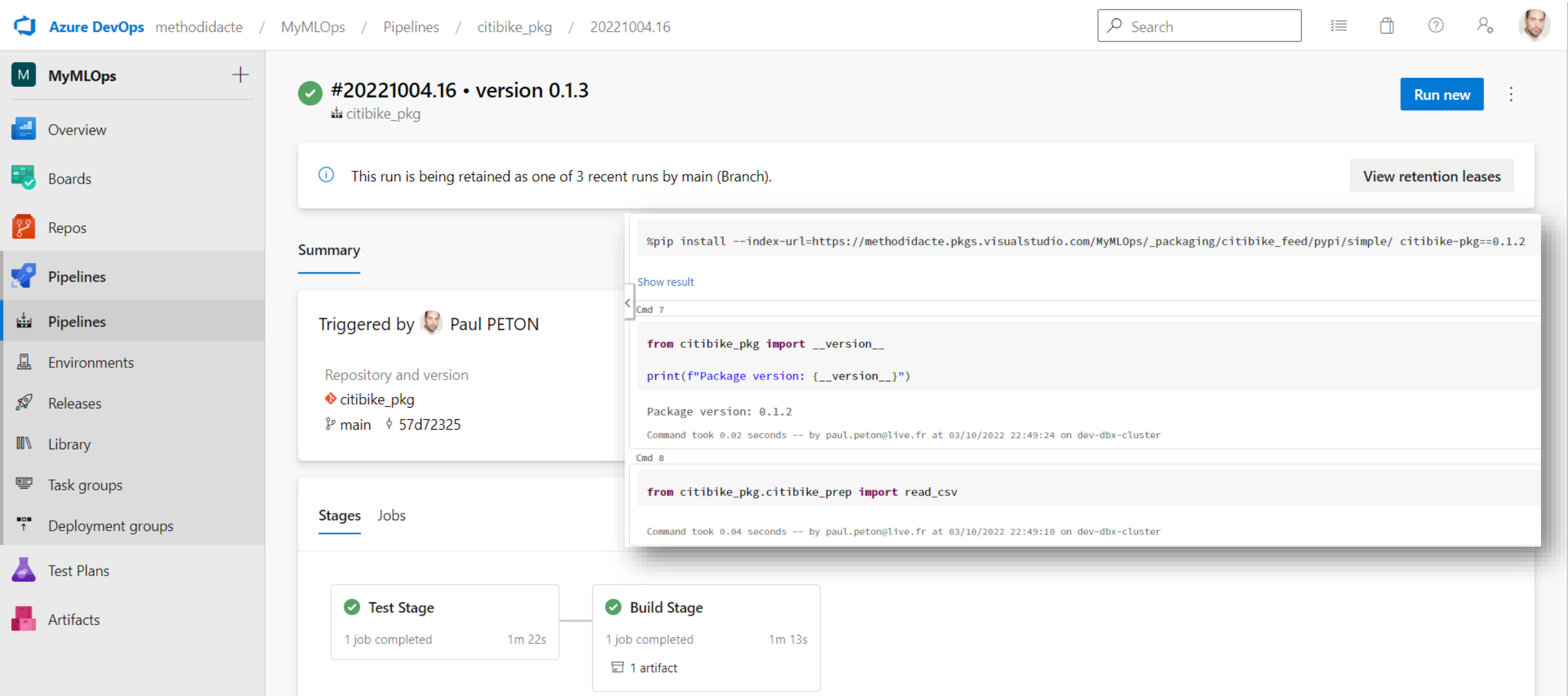


Automation

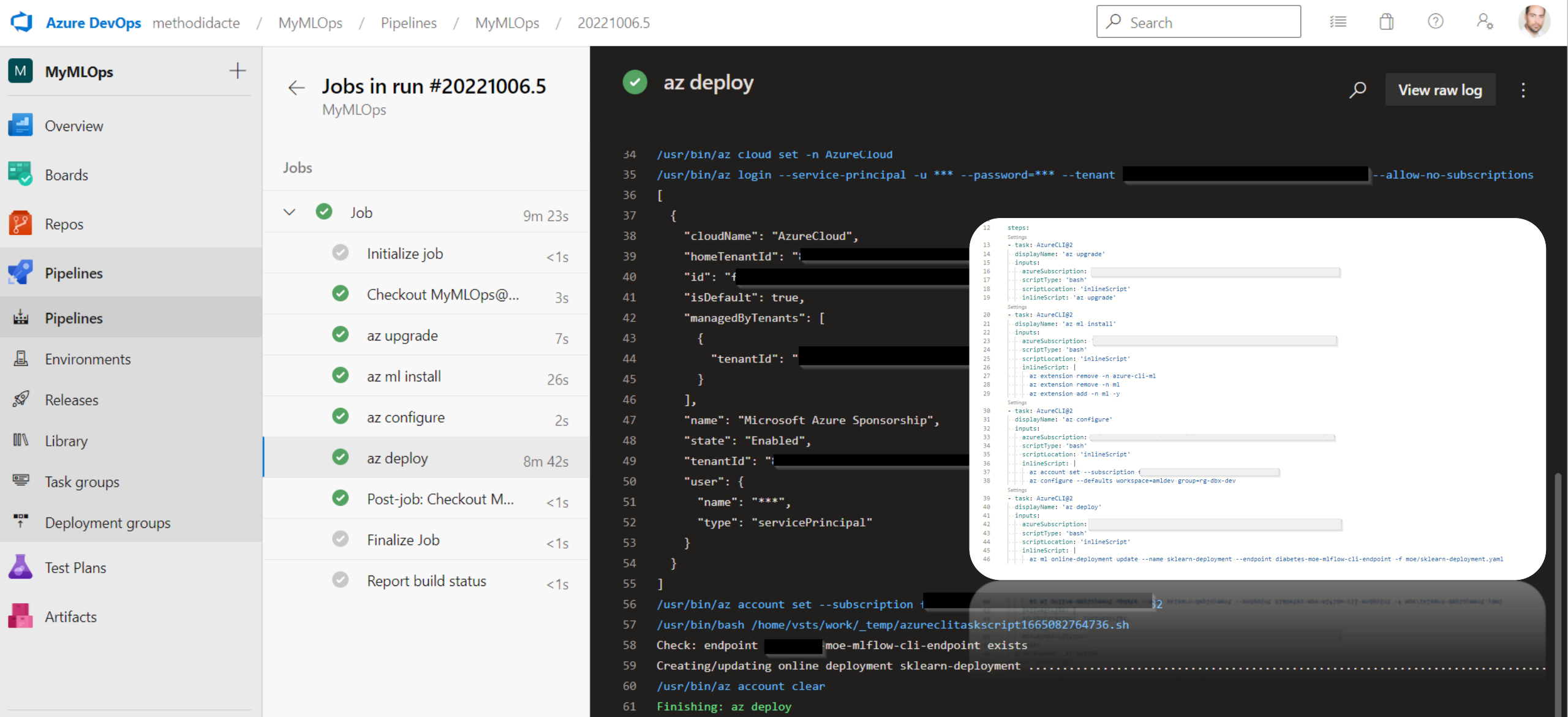
Maturité – niveau 1

- CI : **packager** les livrables
- CD : **déployer** le code et les modèles

CI : tests automatisés et packaging du code



CD : déploiement de modèle par CLI



Monitoring

Maturité – niveau 2

- Détecter les **dérives**
- Ajouter une ***feedback loop*** ?

Drift detection avec Azure Machine Learning

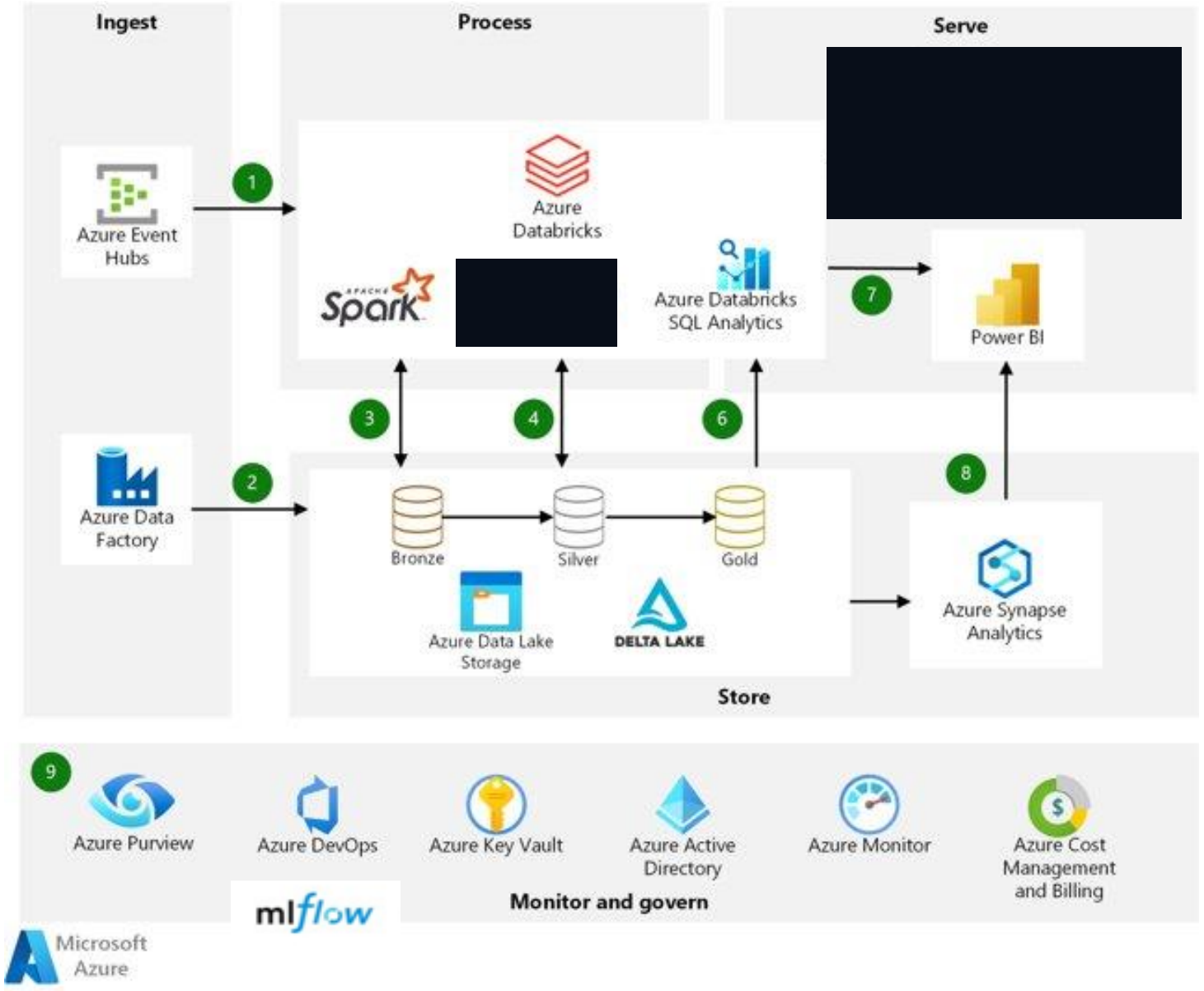
Stack MLOps dans Azure

Un exemple parmi tant d’autres

L’architecture sous-jacente : le *lake house*

- … sauf si vous êtes déjà passé au *Data Mesh* !

- Un lac de données qui **centralise** tout, découpé en couches :
  - Brute
  - Raffinée
  - Valeur métier
- Des outils transverses
  - Dont les **registres**



Lire des papiers de recherche ?

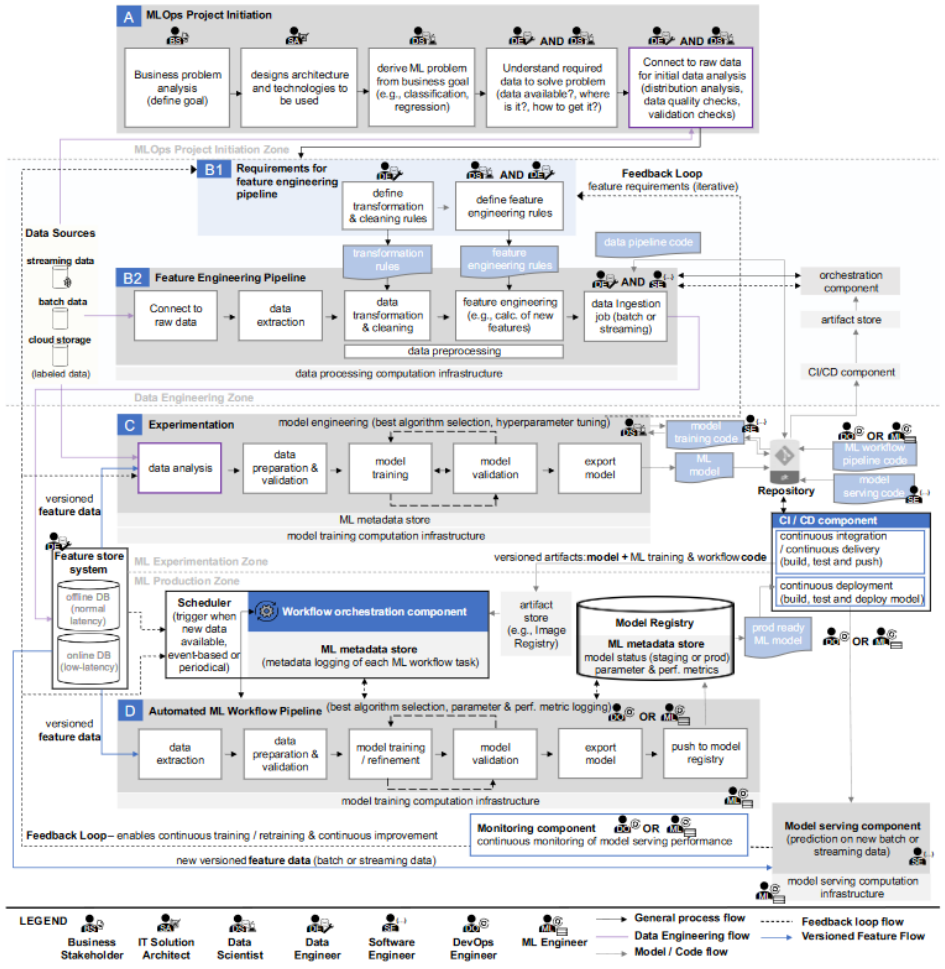


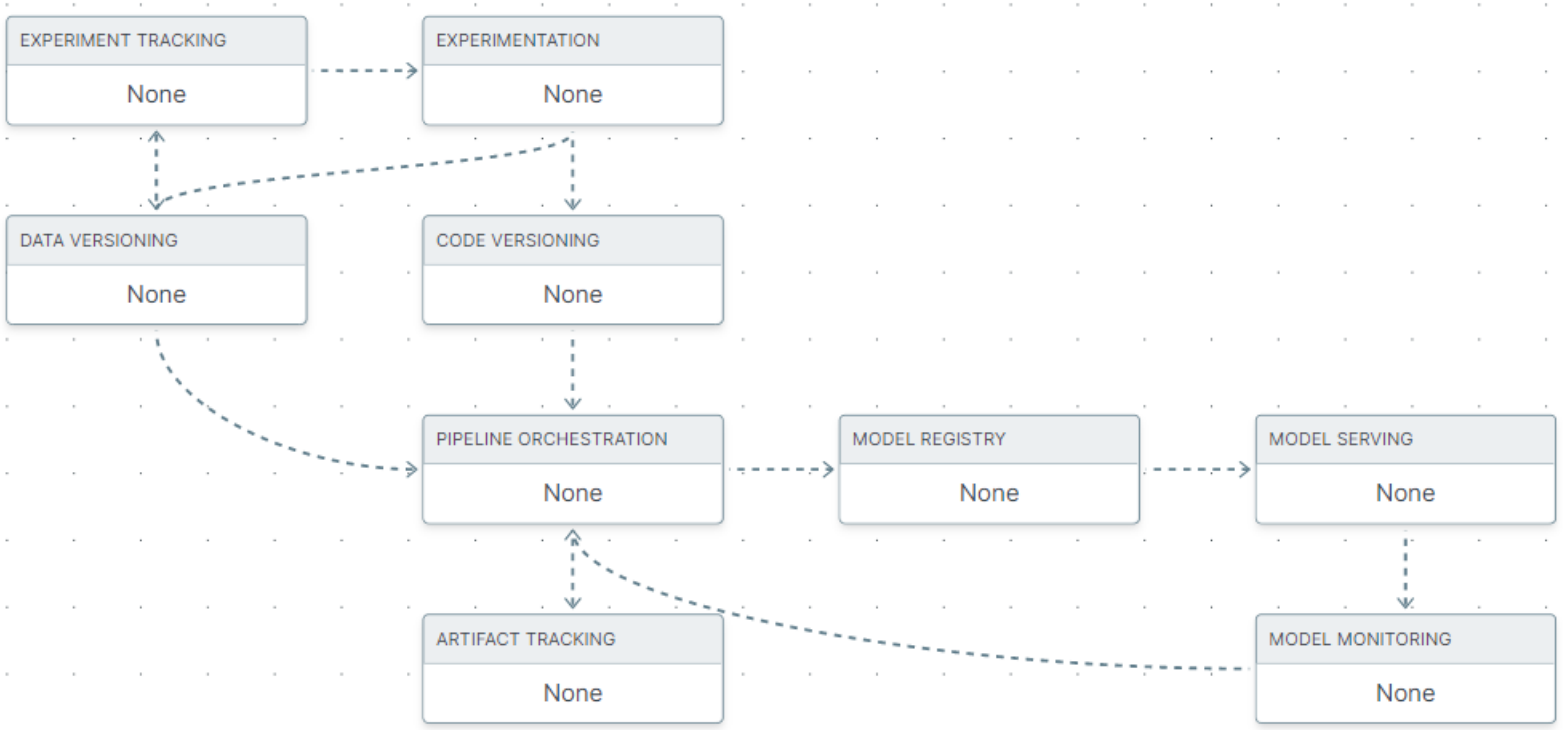
Figure 4. End-to-end MLOps architecture and workflow with functional components and roles

https://arxiv.org/ftp/arxiv/papers/2205/2205.02302.pdf

Build your open-source MLOps stack ?

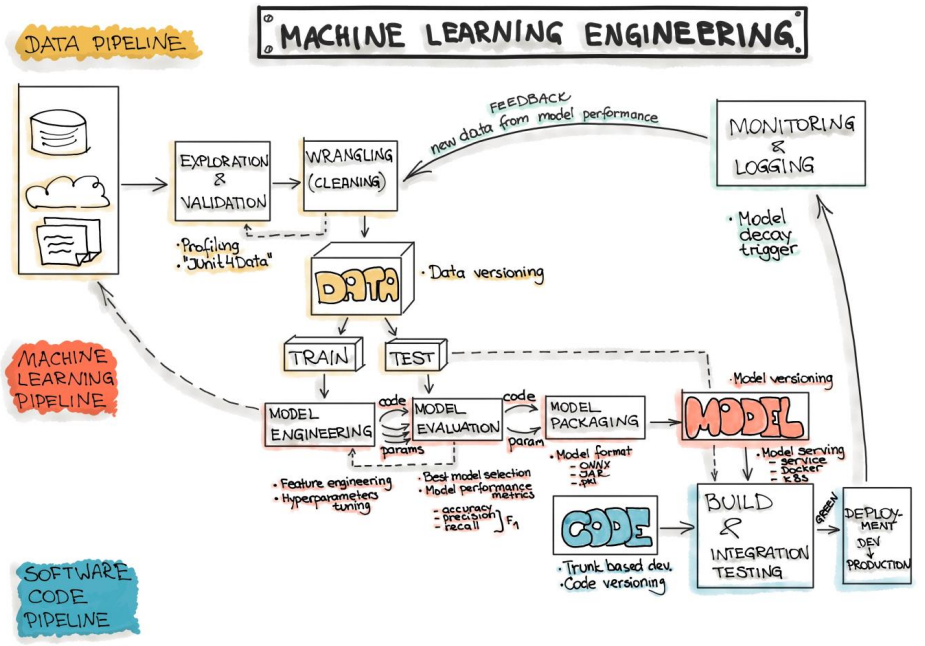
Explore tools and components of the MLOps stack. Select any tool to find more about it. Add it to build your own stack.

https://mymlops.com/



Les trois pipelines (de code) du Machine Learning

Data Engineering Pipelines	*Data Ingestion*
	Exploration and Validation
	Data Wrangling
	Data Splitting
ML Pipelines / workflows	*Model Training*
	Model Evaluation
	Model Testing
	Model Packaging
Model Serving	*Batch*
	On-demand (microservices, REST API)



https://ml-ops.org/content/end-to-end-ml-workflow

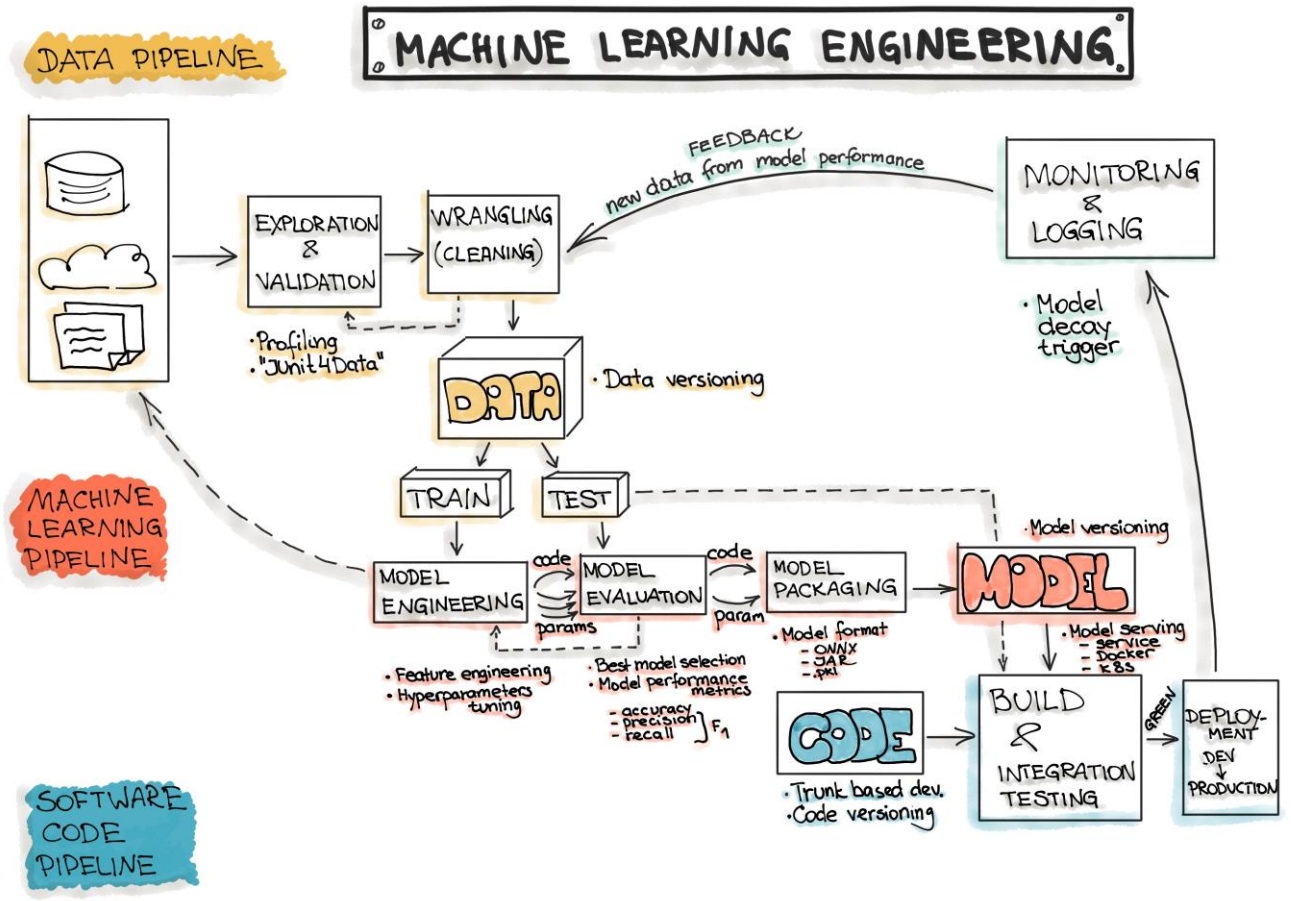
Distinguer les familles de composants

Code

Registry

Pipeline

Monitoring



https://ml-ops.org/content/end-to-end-ml-workflow

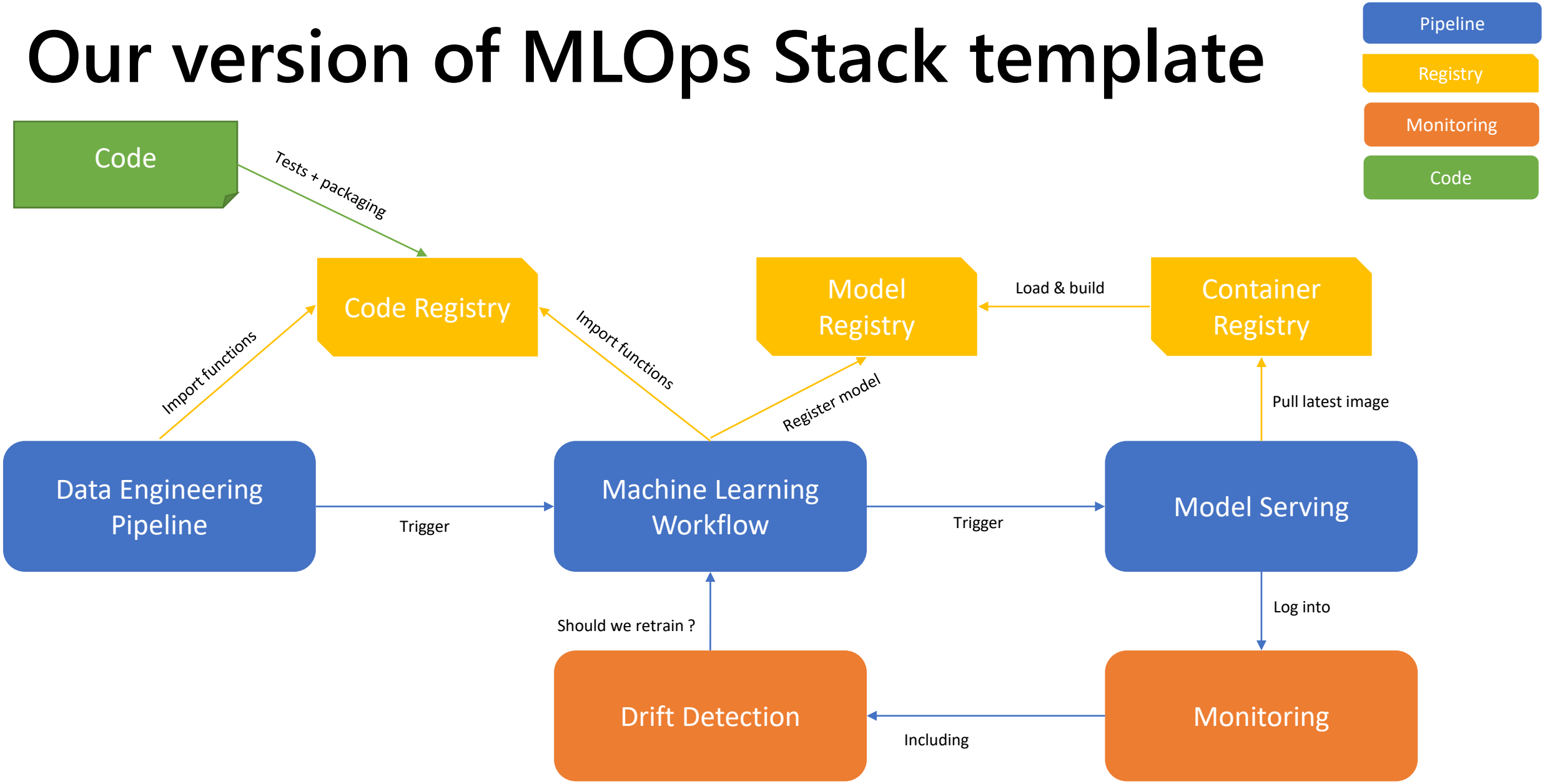
Code
Registry
Pipeline
Monitoring

Distinguer les familles de composants

- Pipeline
  - Data preparation
    - Code versionné dans un Repository Git
  - Model training
    - Artefact de modèle déposé dans un Model Registry
  - Model serving
    - Image (Docker) déposée dans un Container Registry
- Monitoring
  - Inclut la détection de la dérive

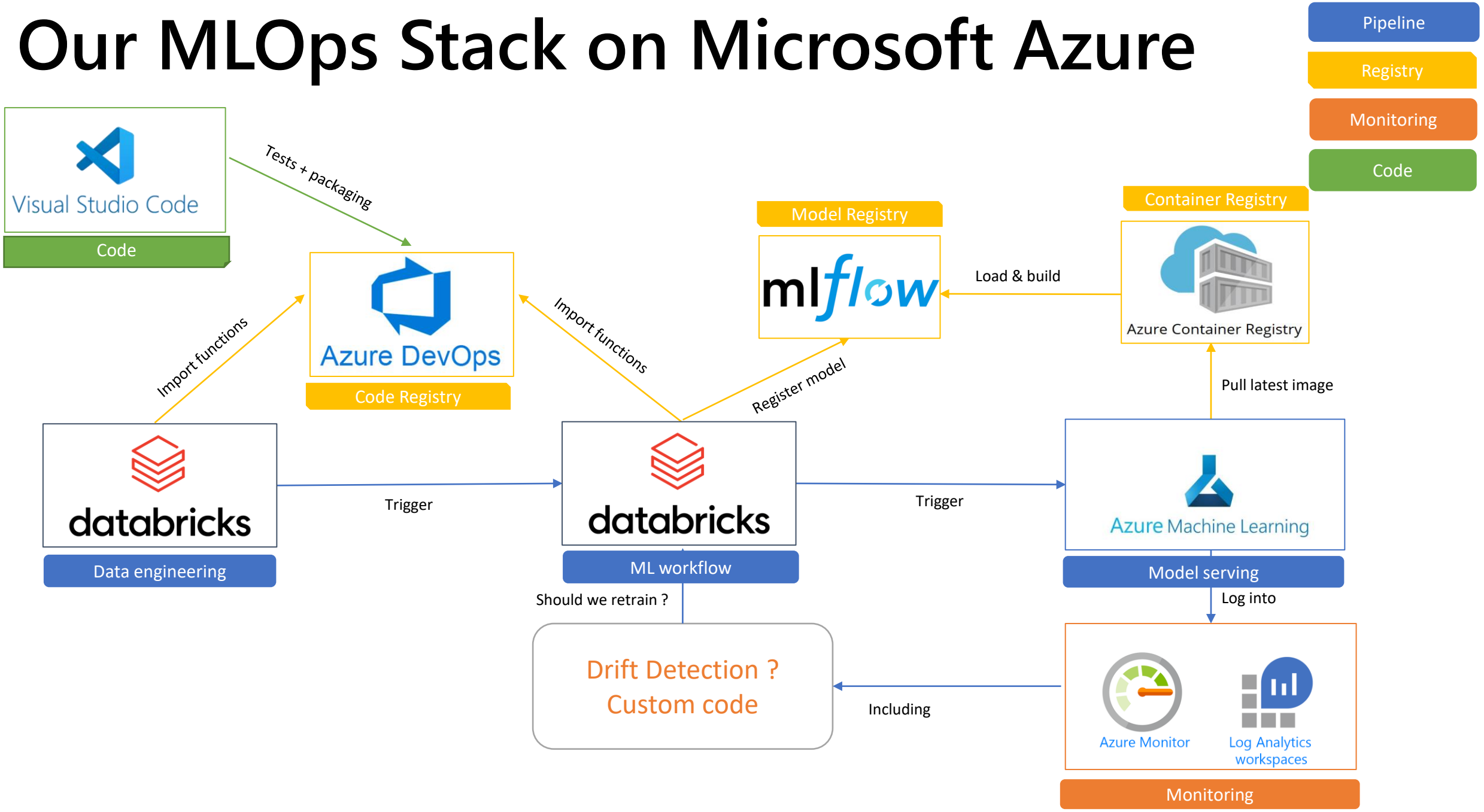
Code

Our version of MLOps Stack template

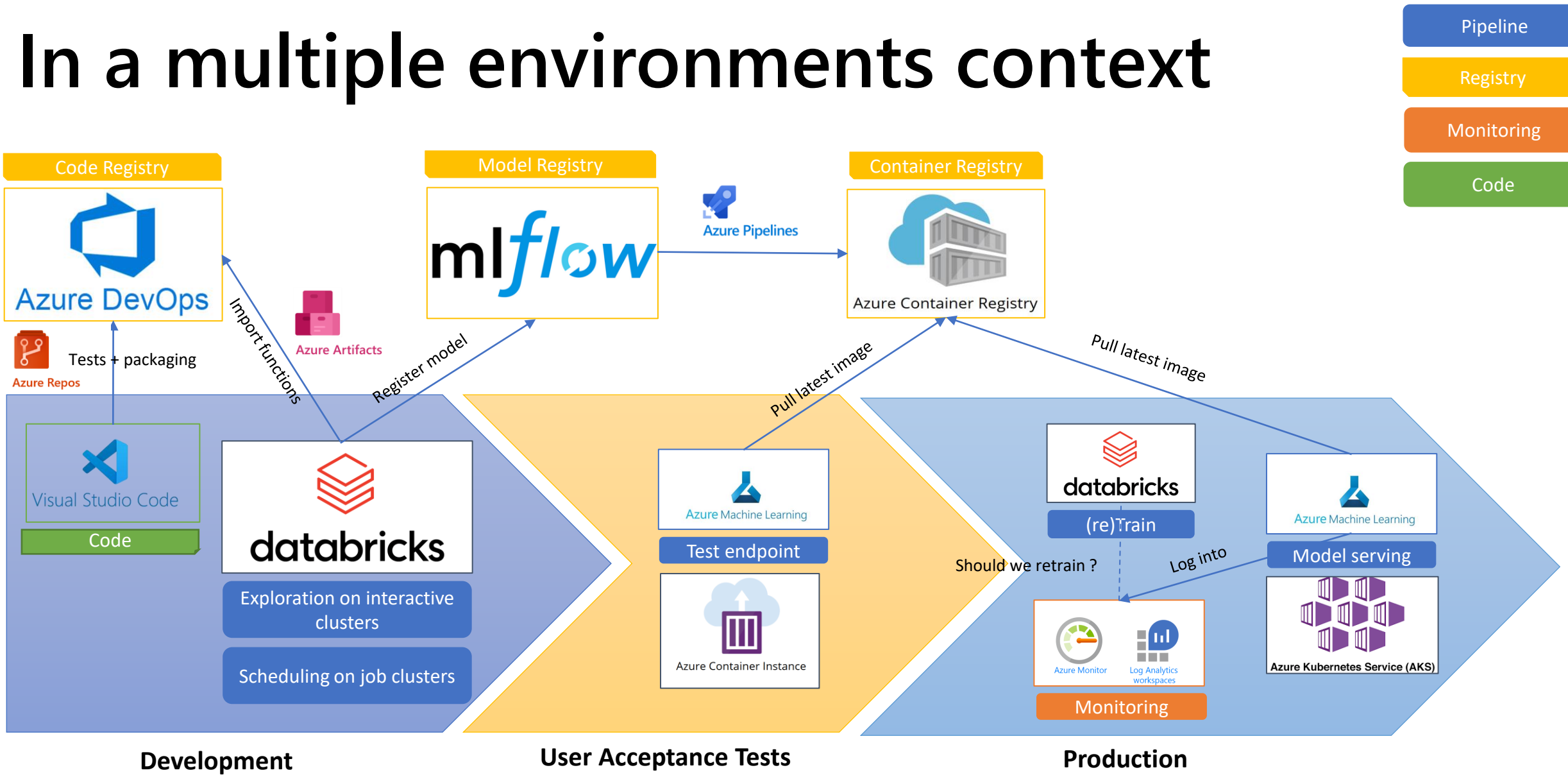



a data factory surrounded by five-legged unicorns

Our MLOps Stack on Microsoft Azure



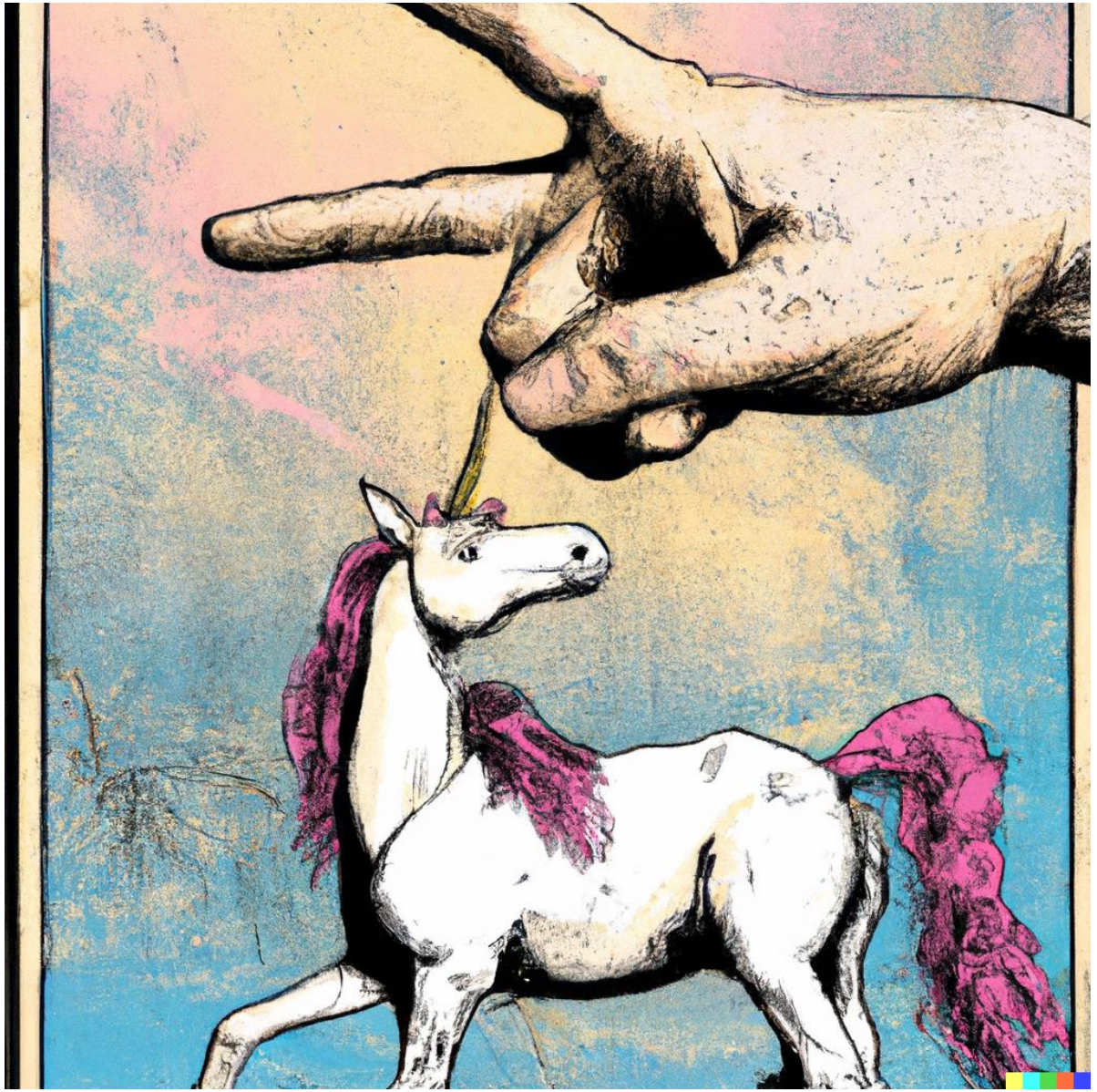
In a multiple environments context



Oh, que je préfère,

à un mouton à cinq pattes,

les doigts de la main



A five-legged unicorn and a giant hand with human fingers comicbook style by J.M.W. Turner