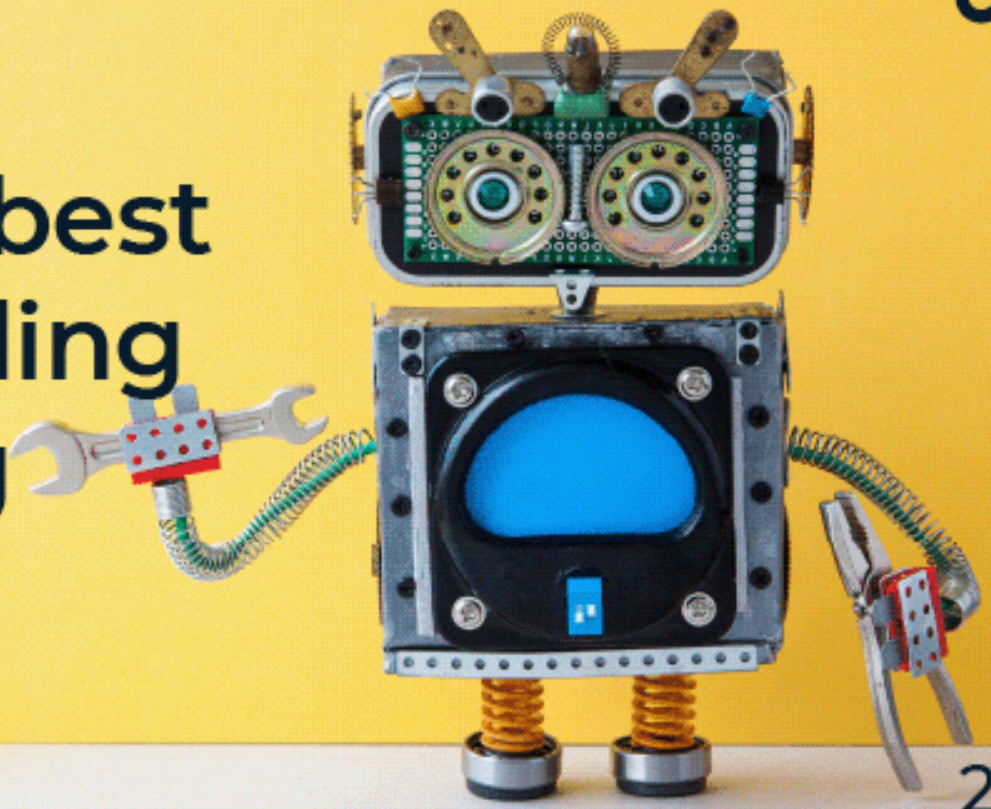
plain concepts

Virtual Coffee

MLOps. Software best practices for building Machine Learning solutions



Microsoft

22/4/2020
10:00 (GMT+02)
Length: 30 min

Rediscover
the meaning of technology

Founded in 2006 by 4 Microsoft MVPs, Plain Concepts was created to help companies adopt new technologies aimed at improving their productivity and processes.

Awarded in 2016 as **Microsoft Partner of the Year**, we currently have over **350 employees**, reaching a milestone in the technology sector by having **12 professionals** recognized as **Microsoft MVP** and over a dozen certifications at business level.

Present in **Spain, USA, UAE, United Kingdom, Germany and the Netherlands**, we have developed over 2,000 projects for companies across all industrial vertical.

plain concepts

Awards

Microsoft Gold Partner

Microsoft AI Partner

Microsoft Partner | Mixed Reality

MICROSOFT PARTNER OF THE YEAR 2016

12 MICROSOFT **MOST VALUE** PROFESSIONAL

Great Place To Work. Best Workplaces™ ESPAÑA 2019

MICROSOFT **GOLDEN PROVIDER**

- Application Development Cloud Platform
- Cloud Productivity Collaboration and Content Datacenter
- DevOps

MICROSOFT **SILVER PROVIDER**

- Application Integration Data Analytics
- Data Platform Messaging
- Small and Midmarket Cloud Solutions
- Windows and Devices

AGILE ALLIANCE CORPORATE MEMBER

XAMARIN PREMIUM CONSULTING PARTNER

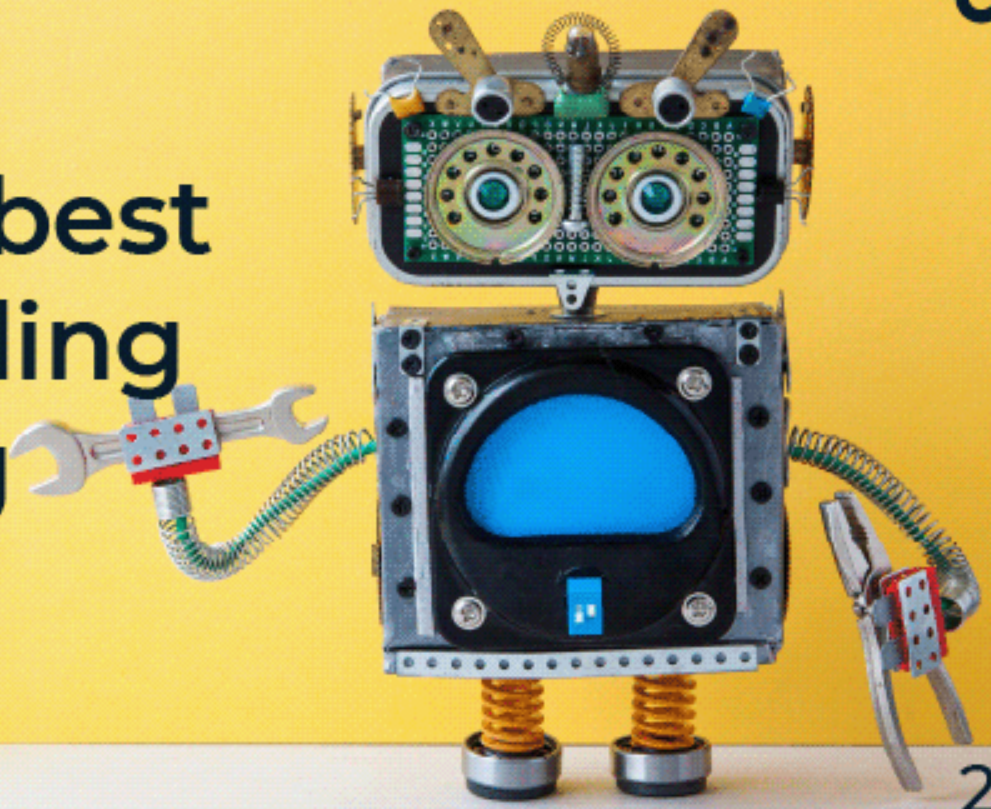
ALM PARTNER OF THE YEAR FOR 7 CONSECUTIVE YEARS

for more information,
visit http://plainconcepts.com

plain concepts

Virtual Coffee

MLOps. Software best practices for building Machine Learning solutions



Microsoft

22/4/2020
10:00 (GMT+02)
Length: 30 min

plain concepts

Agenda

1. Machine Learning is hard
2. MLOps **Components**
3. MLOps **Teams and Processes**
4. MLOps **Flows**

plain concepts

QUICK POLL

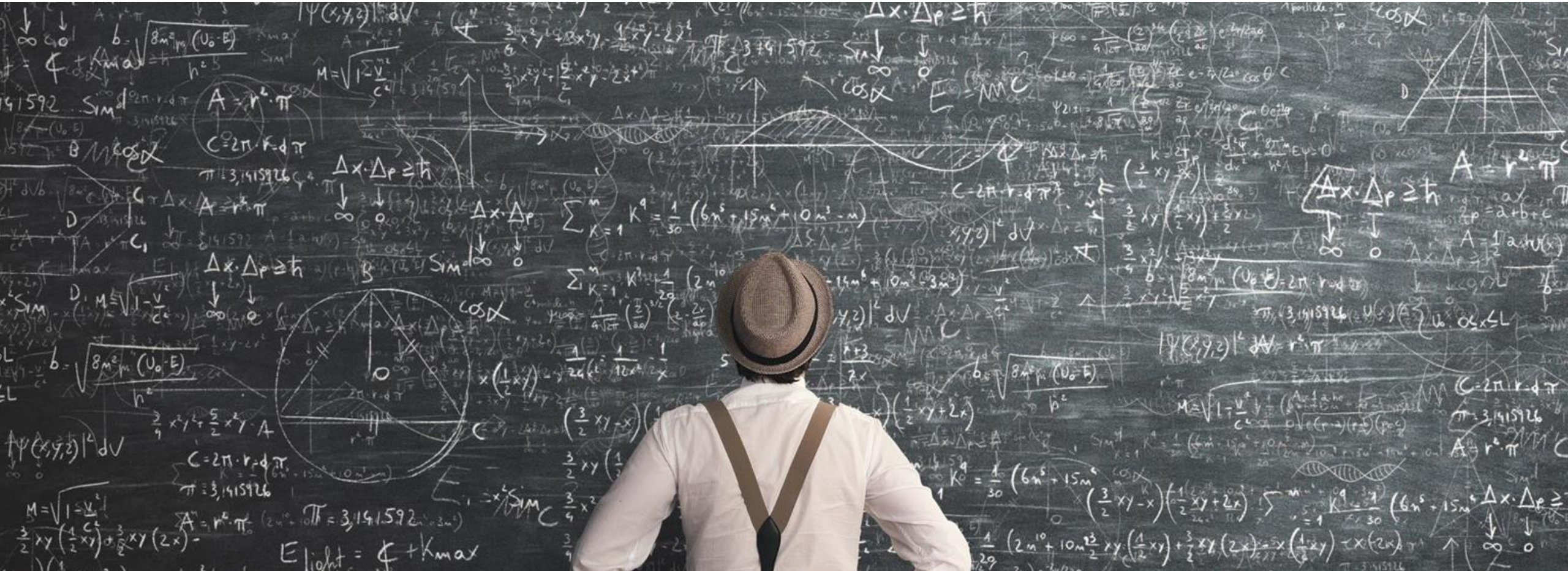
Did you have problems deploying your latest model to production?

YES NO

If YES - What problems?

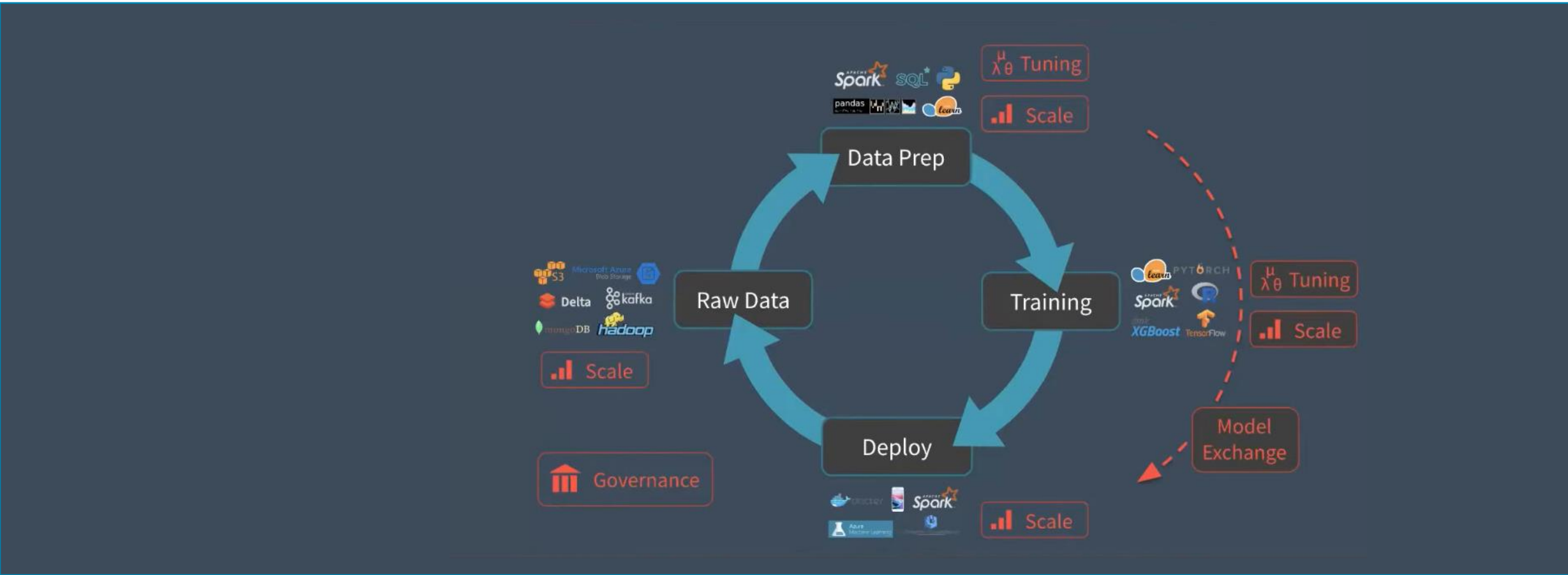
plain concepts

Machine Learning is hard



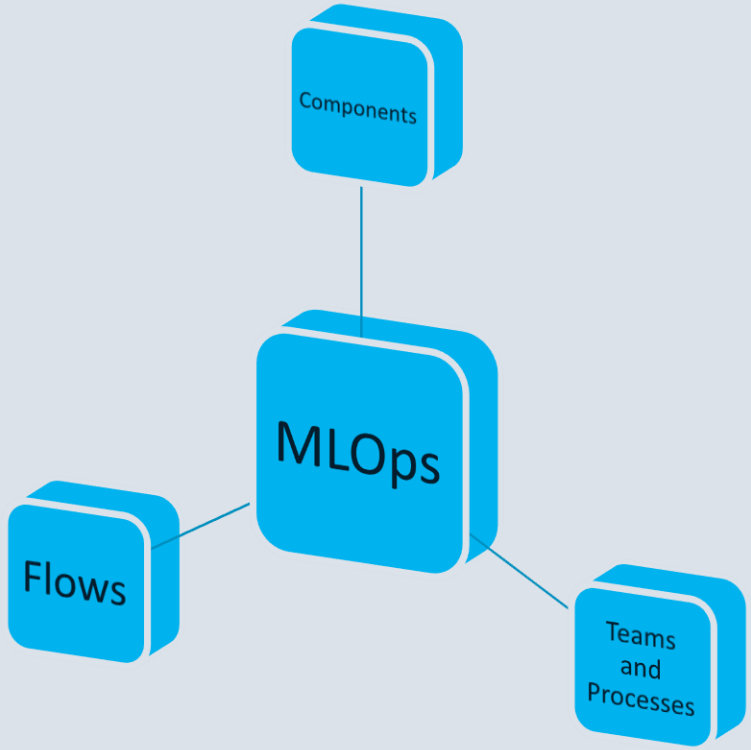
plain concepts

Machine Learning lifecycle



plain concepts

Machine Learning best practices

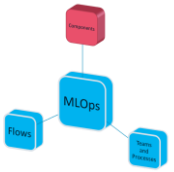


“

MLOps is a new practice for collaboration and communication between data scientists and IT professionals while **automating and productizing** machine learning algorithms

”

plain concepts



MLOps Components

SOURCE CONTROL

GitLab

GitHub

Bitbucket

ML LIFECYCLE

Kubeflow

mlflow

Azure ML

CI/CD PLATFORM

Azure DevOps

Jenkins

circleci

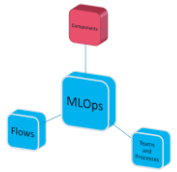
EXECUTION PLATFORM

Azure Data Factory

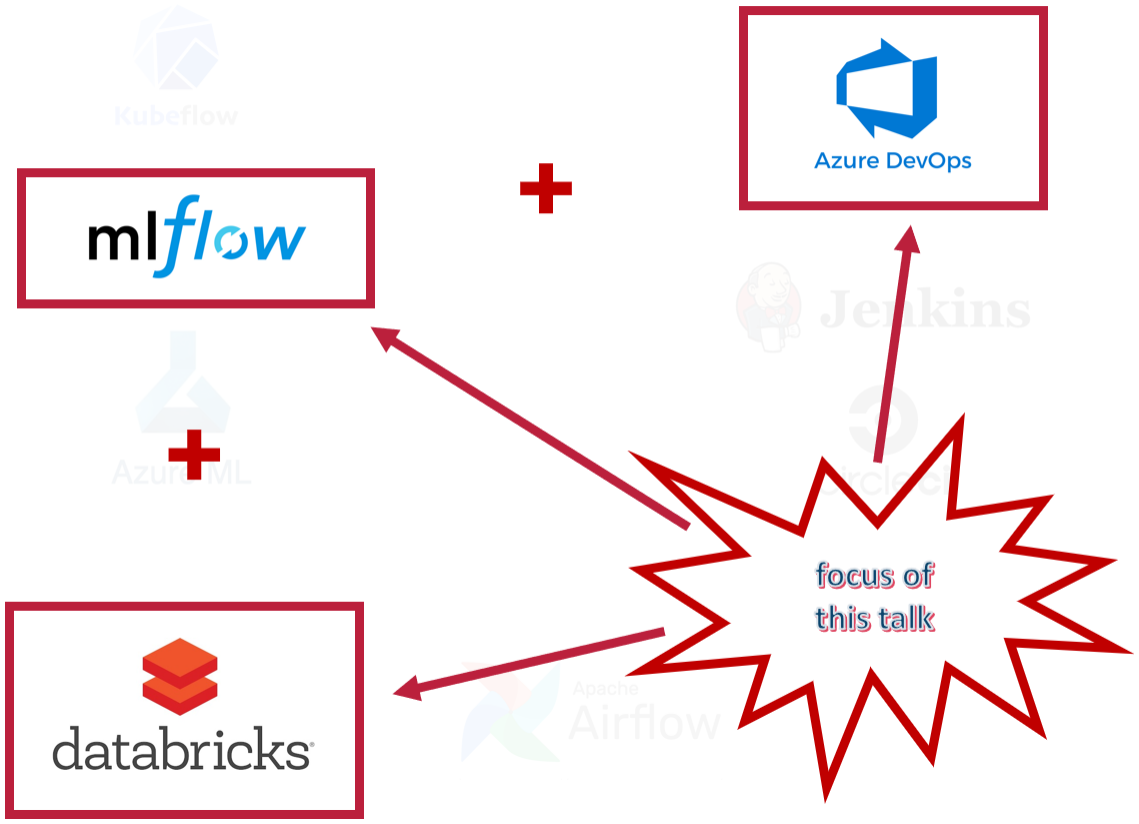
databricks

Apache Airflow

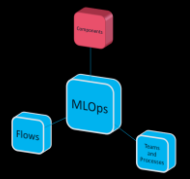
plain concepts



MLOps Components

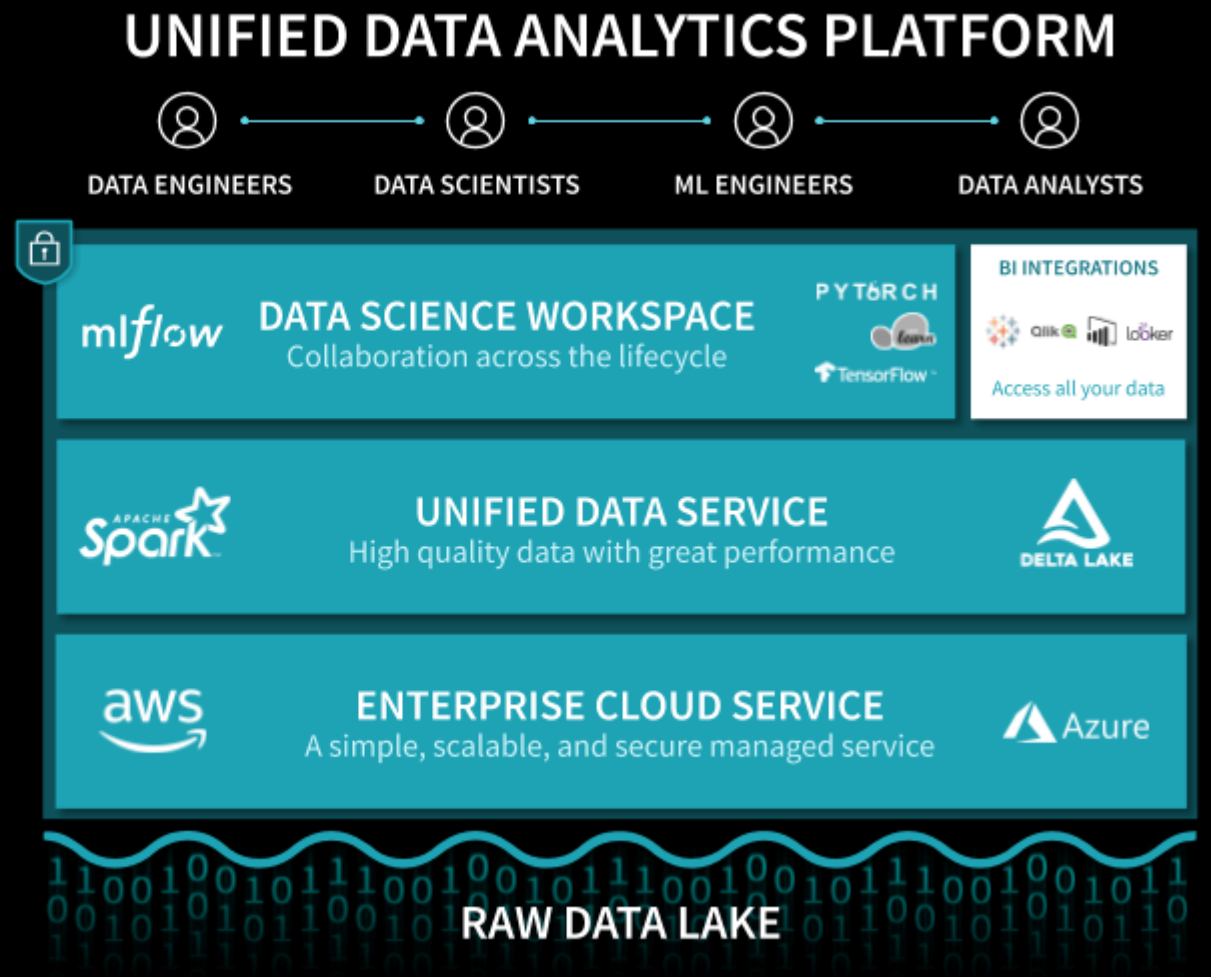


plain concepts



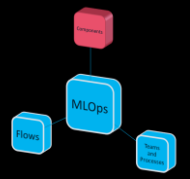
databricks

Accelerating data-driven innovation across data science, data engineering, and business analytics

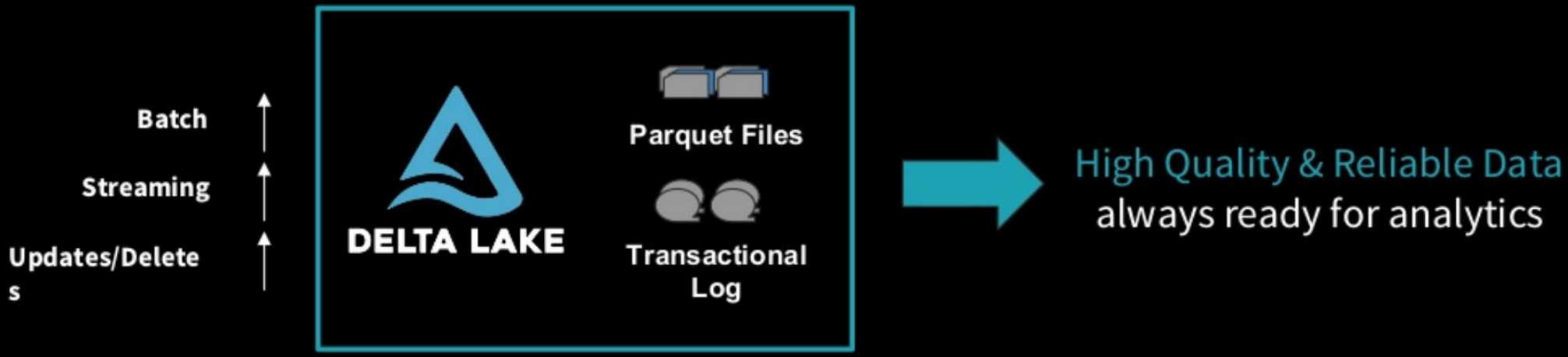


https://databricks.com/

plain concepts



Delta Lake ensures data reliability



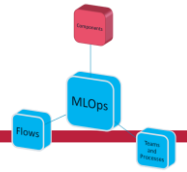
Key Features

- ACID Transactions
- Schema Enforcement
- Unified Batch & Streaming
- Time Travel/Data Snapshots

https://databricks.com/

plain concepts

Azure DevOps



Azure Pipelines

CI/CD for any platform

Build, test, and deploy in any language, to any cloud—or on-premises. Run in parallel on Linux, macOS, and Windows, and deploy containers to individual hosts or Kubernetes.

Azure Repos

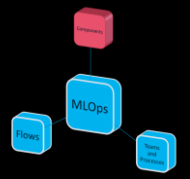
Unlimited free private repos

Get flexible, powerful Git hosting with effective code reviews and unlimited free repositories for all your ideas—from a one-person project to the world's largest repository.

focus of this talk

https://azure.com/devops

plain concepts



mlflow

Open source machine learning platform

- Works with any ML library, algorithm, language, etc
- *Open interface* design (use with any code you already have)

mlflow Tracking

Record and query experiments: code, data, confs, results

mlflow Projects

Packaging format for reproducible runs and workflows

mlflow Models

General format that standardizes deployment paths

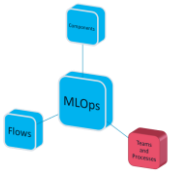
new

mlflow Model Registry

Centralized model management, review & sharing

https://mlflow.org/

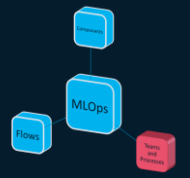
plain concepts



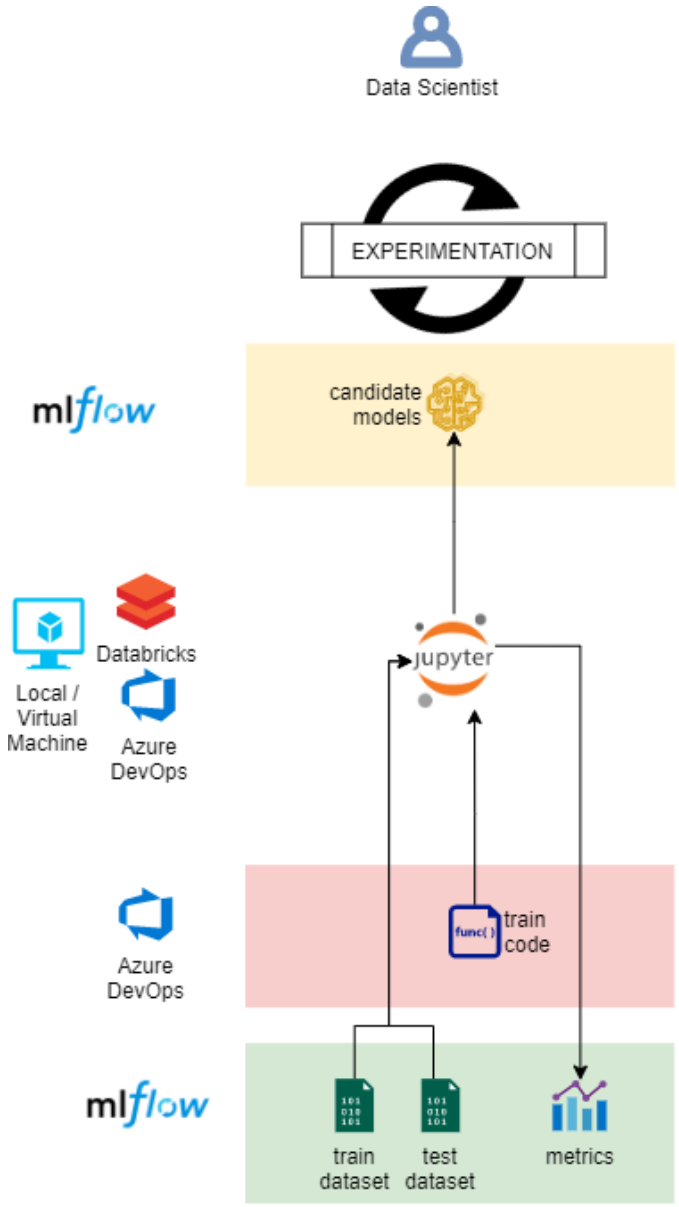
MLOps Teams and Processes

Area \ Stage	Data Scientist EXPERIMENTATION	Machine Learning Engineer OPERATIONALIZATION	Operations Engineer DEPLOYMENT
MODEL			
=			
CODE			
+			
DATA			

plain concepts



EXPERIMENTATION



Model Selection (mlFlow Models)

EDA (Exploratory Data Analysis)

Preprocessing
- Data Collection
- Feature Engineering

Train Model
- Hyperparameter tuning

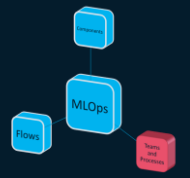
Evaluate Model

Feature branches
Notebooks (w/out Spaghetti code)

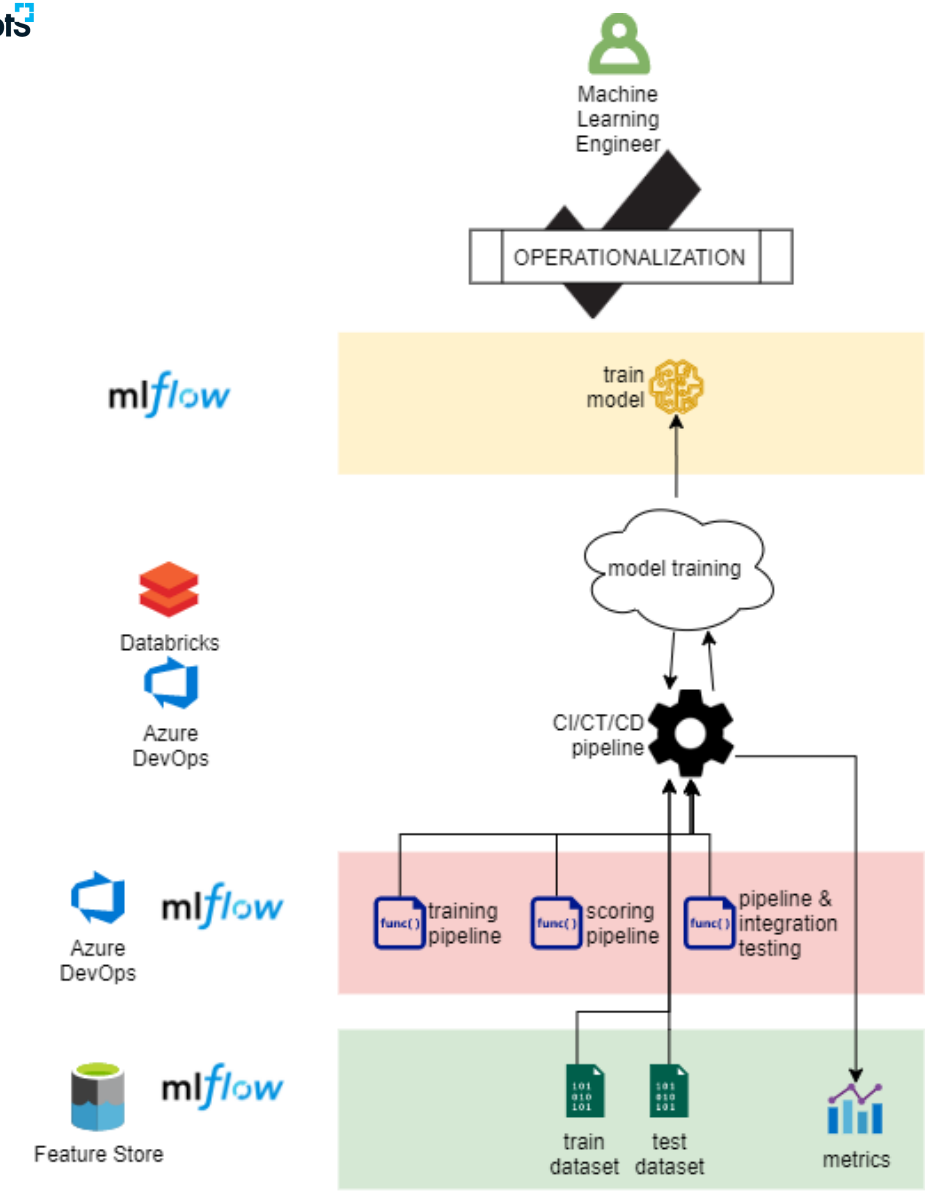
Dataset samples

Model Hyperparameters (mlFlow Tracking)

plain concepts



OPERATIONALIZATION



Model Store (mlFlow Models)

DevOps Pipelines:
- Build
- Release

Simple Orchestration:
- code
- gated releases

Infrastructure deployment

Code versioning
- Git flows
- mlFlow project

ML Pipelines (train and scoring)

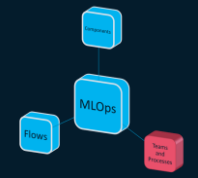
Unit and Integration testing

Train and Testing Dataset

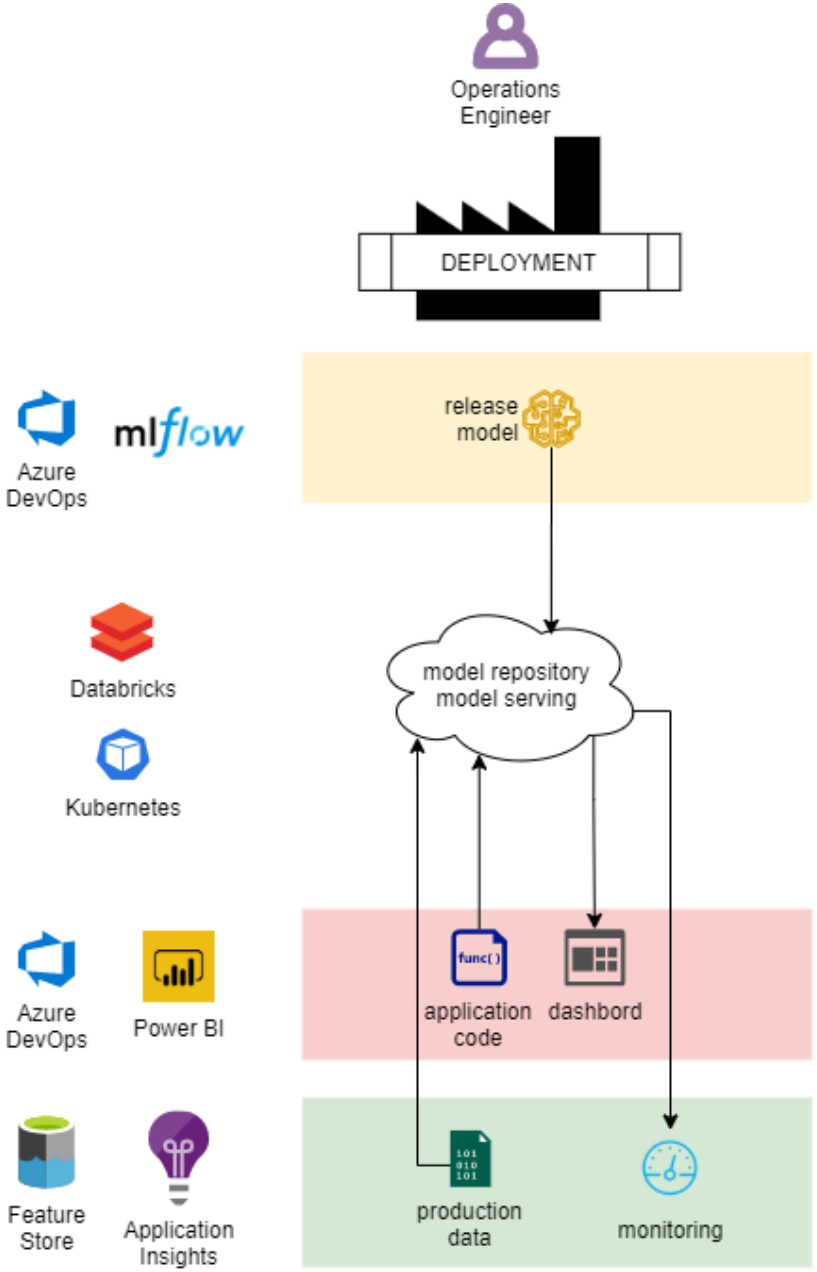
Feature Store (data lineage)

Data Pipelines integration

Model evaluation (mlFlow tracking)



DEPLOYMENT



Model Registry (mlFlow Model Registry)

Model deployment:
- Single / Multiple model (partitioned models)
- Shadow model (canary deployment)
- Competing models (A/B testing)

Model serving:
- Embedded model (model + code)
- Model as a Service (REST API)
- Model as artifact (deployment by label)

Scalability:
- Kubernetes
- Databricks

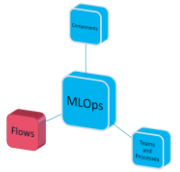
Scoring:
- Batch (offline)
- Single Request (online)

Visualization
API's
Analytics

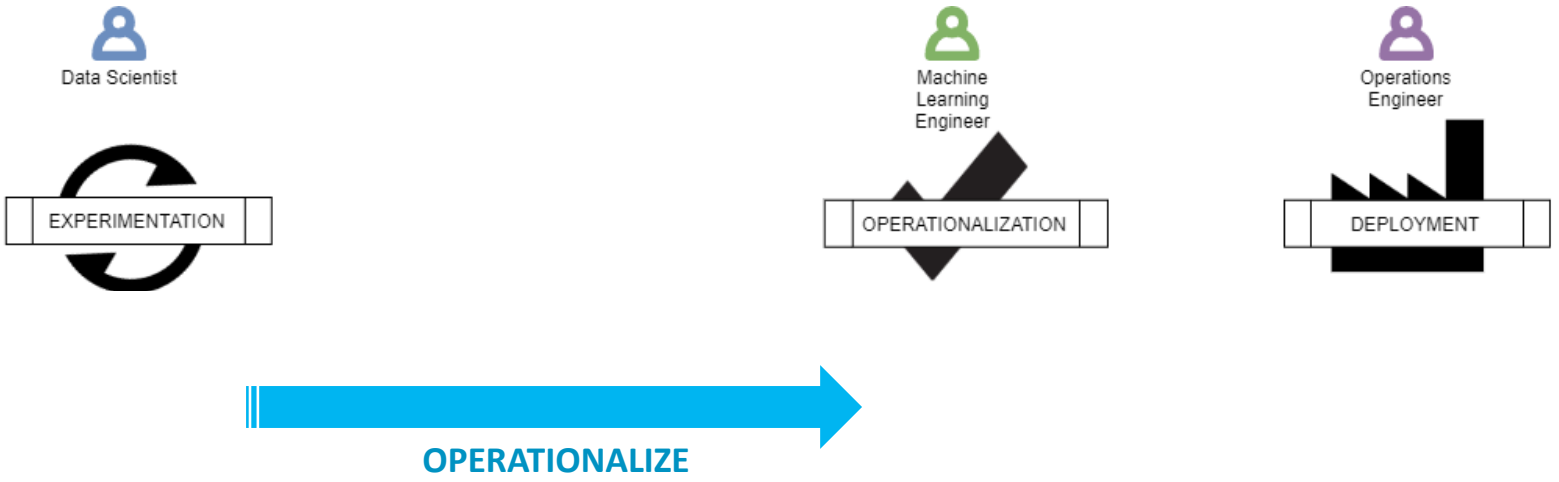
Production Data (w/out past data)

Monitoring:
- Concept drift
- Data drift

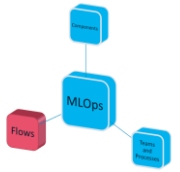
plain concepts



MLOps Flows

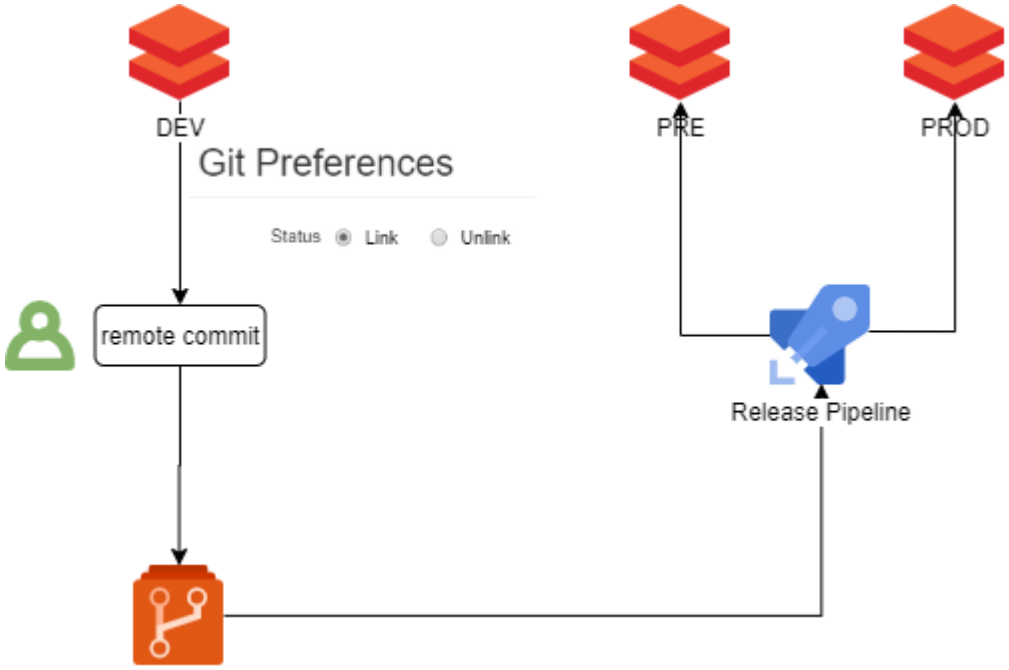


plain concepts

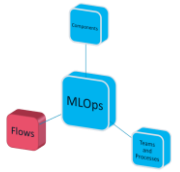


Databricks Development

Databricks based development

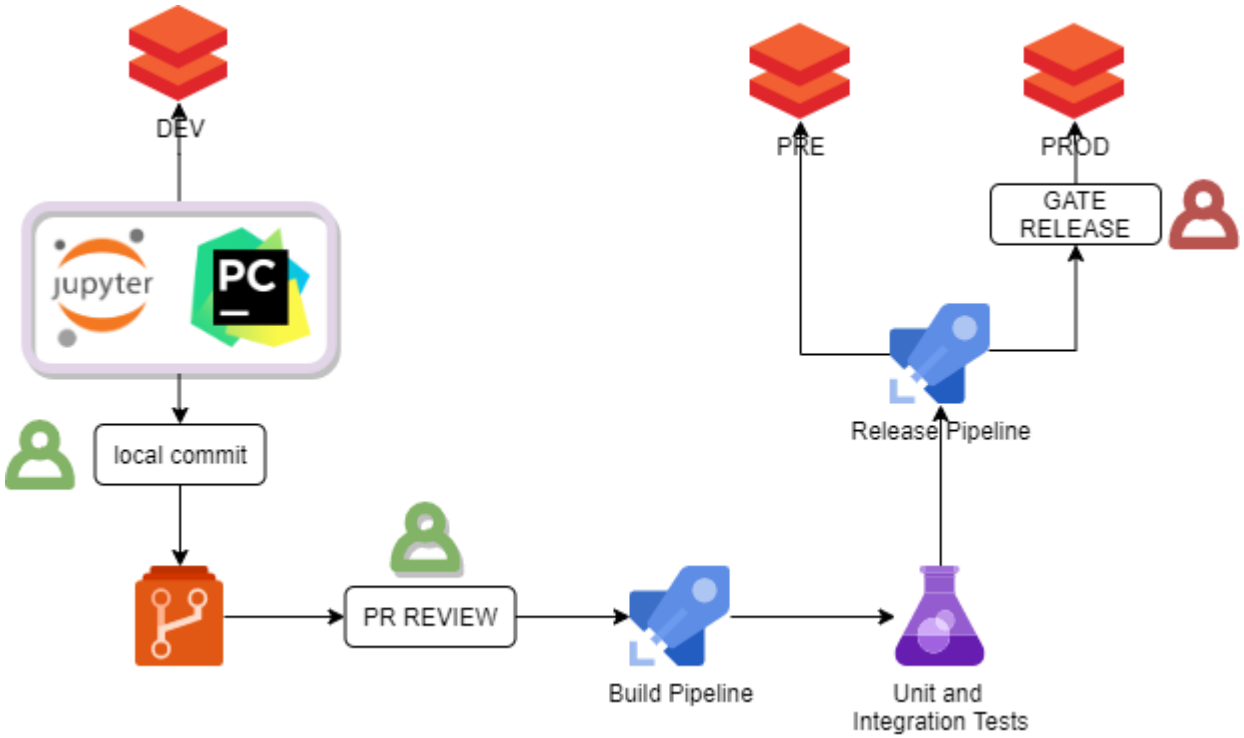


plain concepts

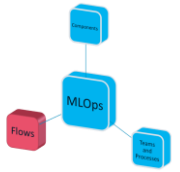


Databricks Development

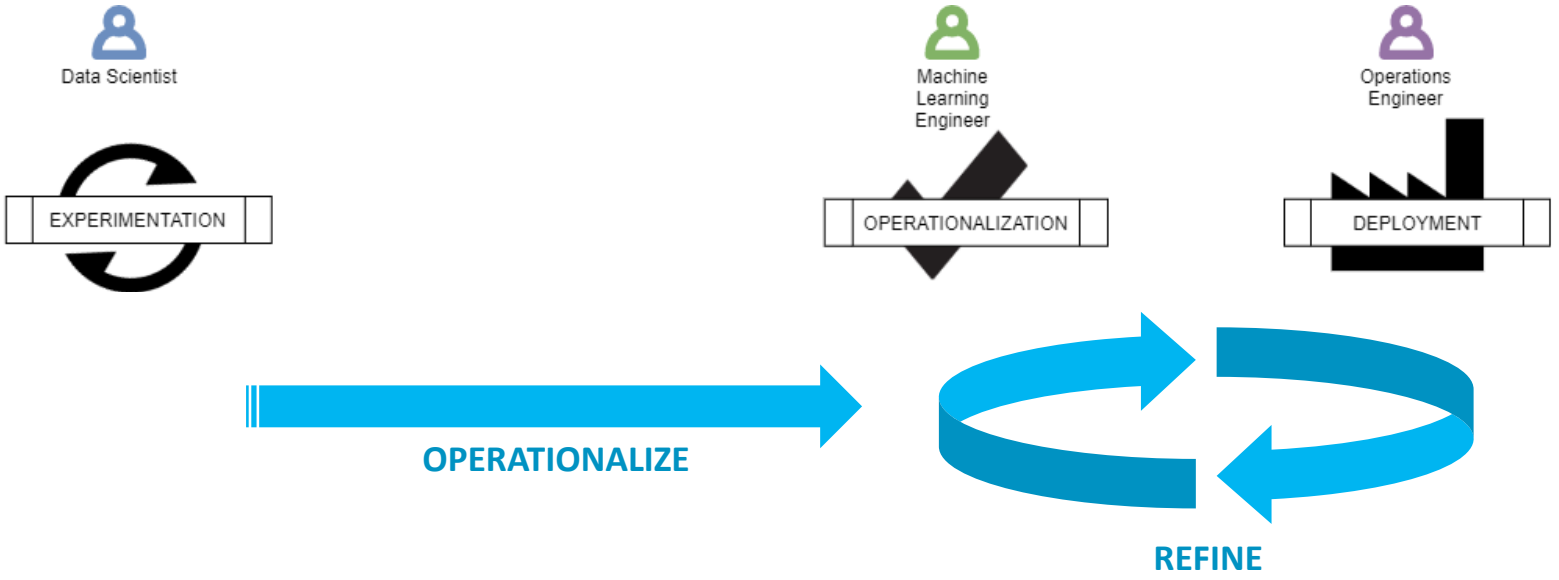
Mixed based development



plain concepts



MLOps Flows



plain concepts

Virtual Coffee

Optimización de modelos de Deep Learning: Mejora la eficiencia y el rendimiento de tus predicciones

Microsoft

22/04/2020
16:00 (GMT+02)
Duración: 30 min

plain concepts

Thank you

fplopez@plainconcepts.com

kalbes@plainconcepts.com

www.plainconcepts.com