Talk'y digital by **xpr_campus**

Deploy and scale your machine learning models

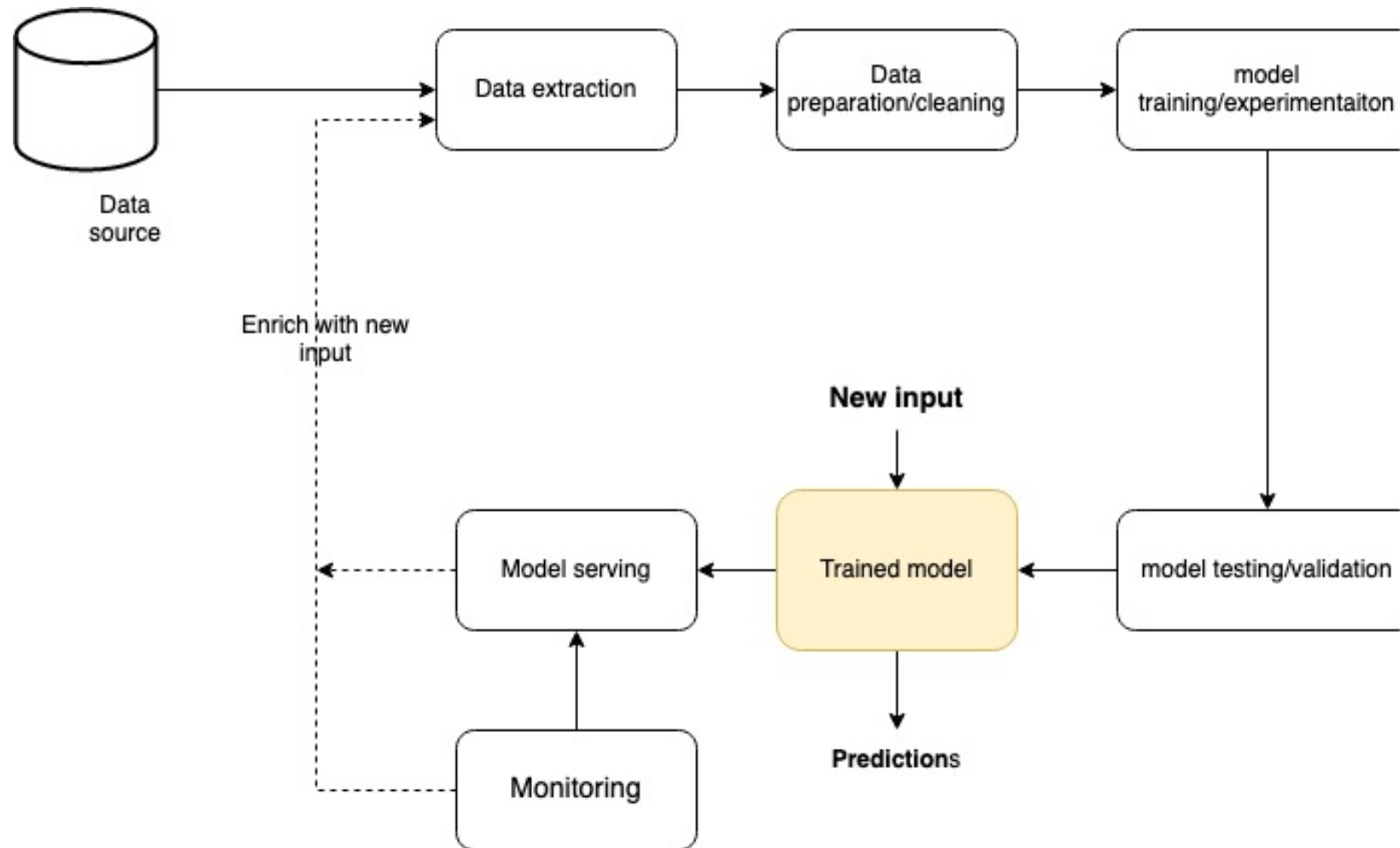Zakaria EL BAZI

06/02/2022

Zakaria EL BAZI

https://elbazi.co
Data/Cloud/Devops engineer @ Micropole Group

machine learning in Lab

Machine Learning Lab :

Machine Learning Lab :

Data scientist :

- Explore the data in local environments (a lab or even local machine with Jupyter notebooks).

- Experiment and test different approaches to understand the data (data and statistical analysis).

- "Train multiple" models on small datasets and compare them.

- Finetune the selected model and optimize it.

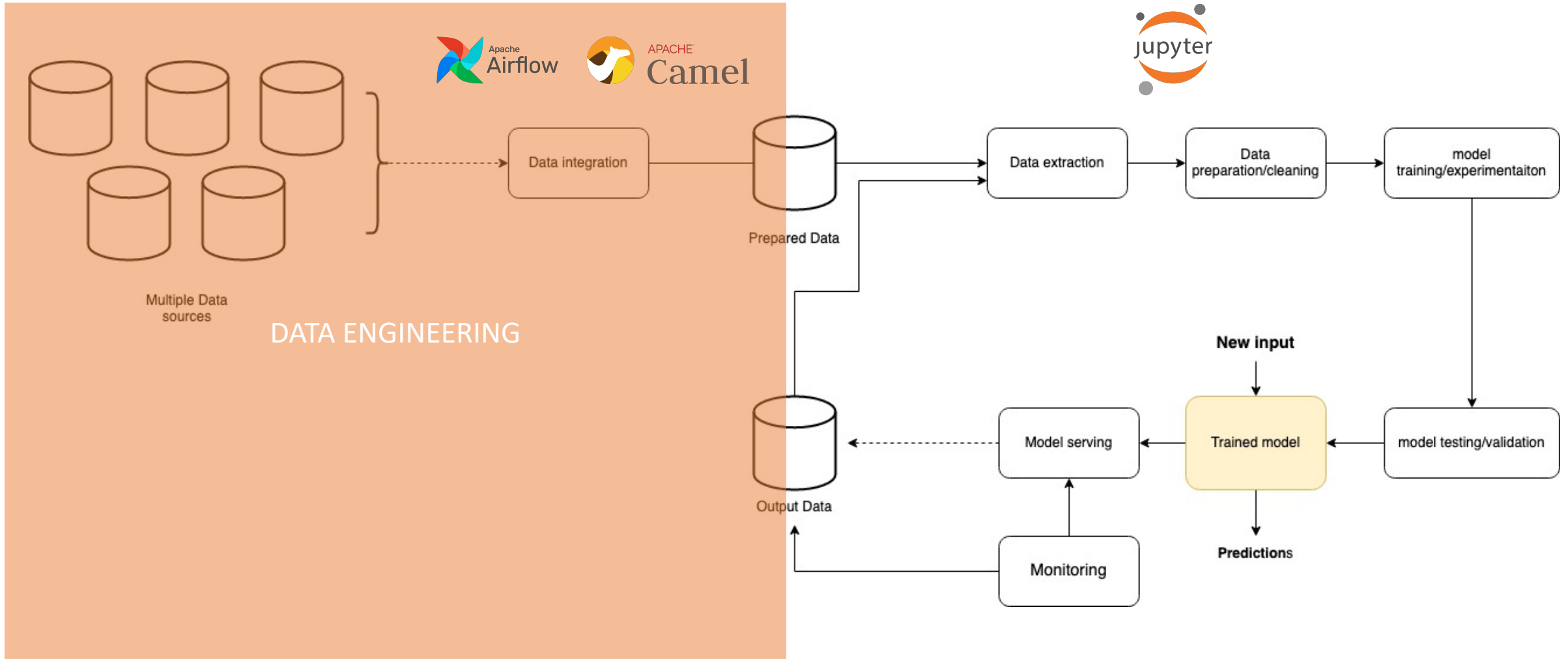- Export the trained model to a model registry (parameters).

Machine Learning Lab :

Assumptions:

- The business problem is well defined.
- **The data is prepared** and **perfectly documented**.
- The required work environment is prepared (labs, data access, tools configurations, etc).

DATA ENIGEERS

Machine Learning Lab :

DATA ENGINEERING

Multiple Data sources

Data integration

Prepared Data

Data extraction

Data preparation/cleaning

model training/experimentaiton

model testing/validation

Trained model

New input

Predictions

Model serving

Output Data

Monitoring

Machine Learning Lab :

Data engineer :

- Collect data from various sources.

- Create data integration pipelines and ensure data quality.

- Store the data correctly in proper data storage services (Sql, NoSql).

- Maintain and improve databases continually.

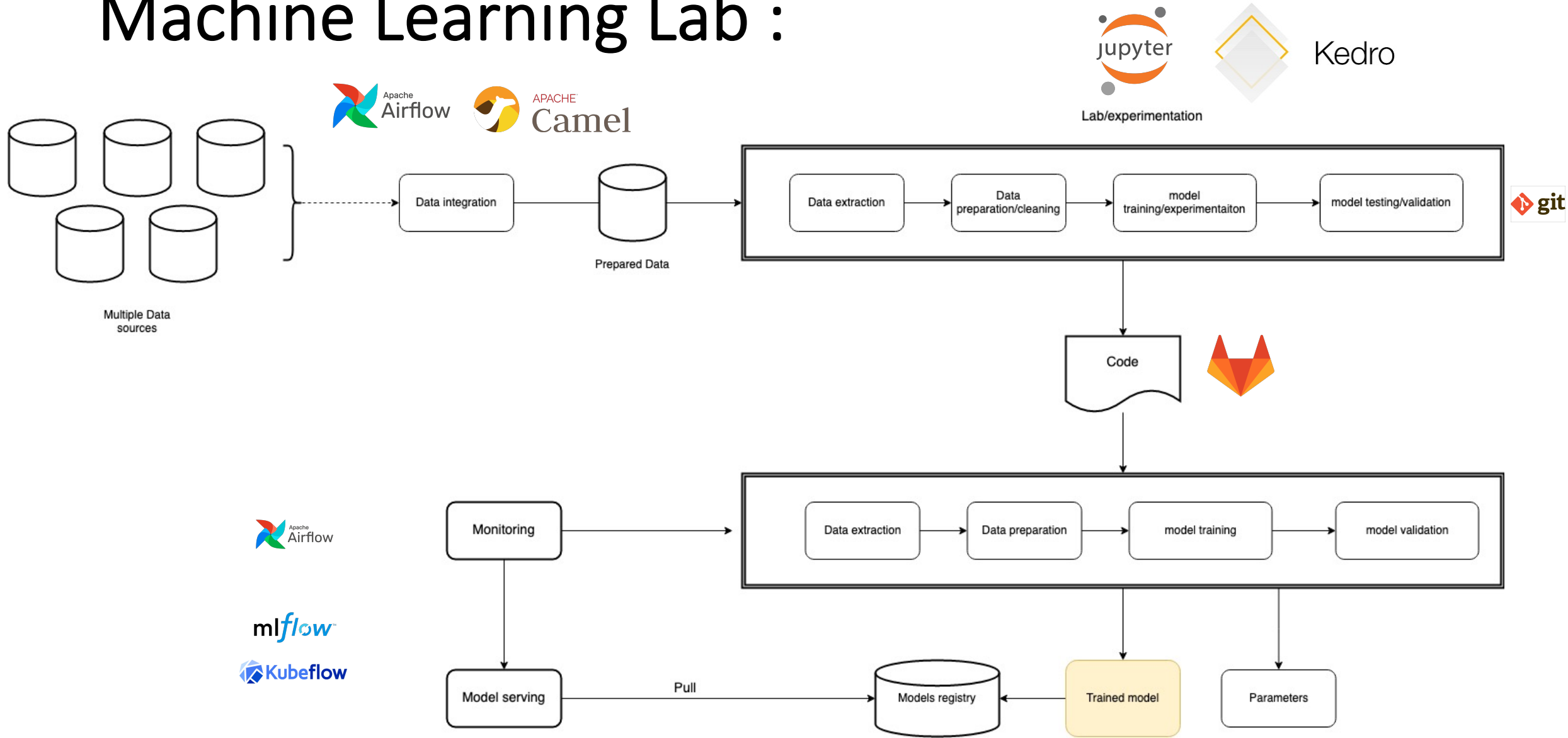- Work with data scientists to improve data pipelines.

machine learning in Prod & MLOPS

Machine Learning in prod :

Good production means :

- Model is reproducible.

- Code and models are version controlled.

- Continuous deployment.

- Collaborative (includes best practices).

Machine Learning Lab :

Machine Learning in prod :

Kedro:

- "Kedro is a python framework for creating reproducible, maintainable and modular data science code by applying **software engineering best practices** to machine-learning code "

- It offers a UI to visualize the data pipeline.

- Integrates perfectly with existing tools like Jupyter, PySpark, etc.

- Kedro pipelines are easy to deploy as packages or containers:
    - As Airflow DAGs
    - With AWS batch
    - To Databricks, Kubeflow and many more

Machine Learning in prod :

Mlflow:

- Mlflow is an end-to-end machine learning lifecycle that allows to :
  - Track experiments and record/compare parameters
  - Packaging ML code in a reusable, reproducible form in order to **share with other data scientists** or **transfer to production**.
  - **Managing and deploying models** from a variety of ML libraries to a variety of model serving and inference platforms.
  - **Model regitry ,model versioning, stage transitions, and annotations**.
  - **Model monitoring.**

Machine Learning Lab :