Automated experiment tracking

FULLY AUTOMATED MLOPS

Arturo Opsetmoen Amador
Senior Consultant - Machine Learning

Machine learning has an experimental nature

Many "levers" to experiment with:

- Data transformations
  - Features used
- Training algorithms
  - Hyperparameter tuning
- Evaluation metrics
  - Precision, recall, etc.

A large space of possibilities!

Problems of Manual Machine Learning Workflows

Lack of automation = problems in manual ML workflows

- Difficult to track experiments and results
- Waste time and resources
- Hard to reproduce/share results

Automated logging in ML

Important things to log include:

- Code
- Environment
- Data
- Parameters
- Metrics

The importance of logging

Logging is essential for:

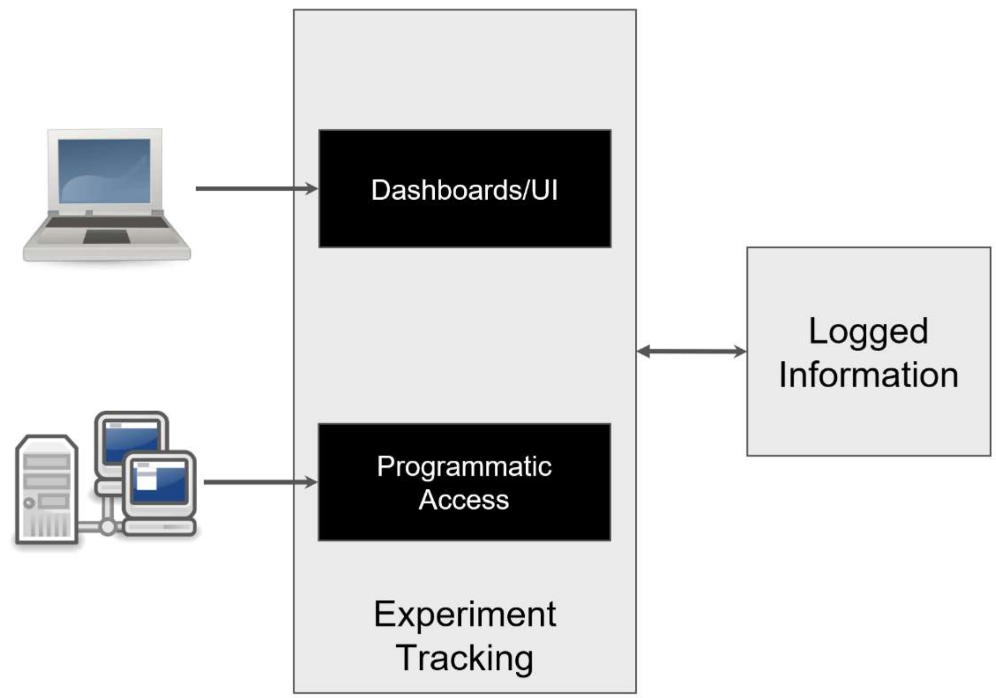
- The reproducibility of experiments in ML systems.
- Tracking system performance and making informed decisions.
- Identifying potential issues for improvement.

Reproducibility provides transparency and is crucial to make our systems trustworthy.

Automated experiment tracking system

Organize logs per run, or experiment to:

- See model training metadata
- Compare model training runs
- Reproduce model training runs



Automated experiment tracking - Today's market

Several tools that automate experiment tracking:

- TensorBoard
- MLFlow
- Weights & Biases
- Neptune

Let's practice!

FULLY AUTOMATED MLOPS

datacamp

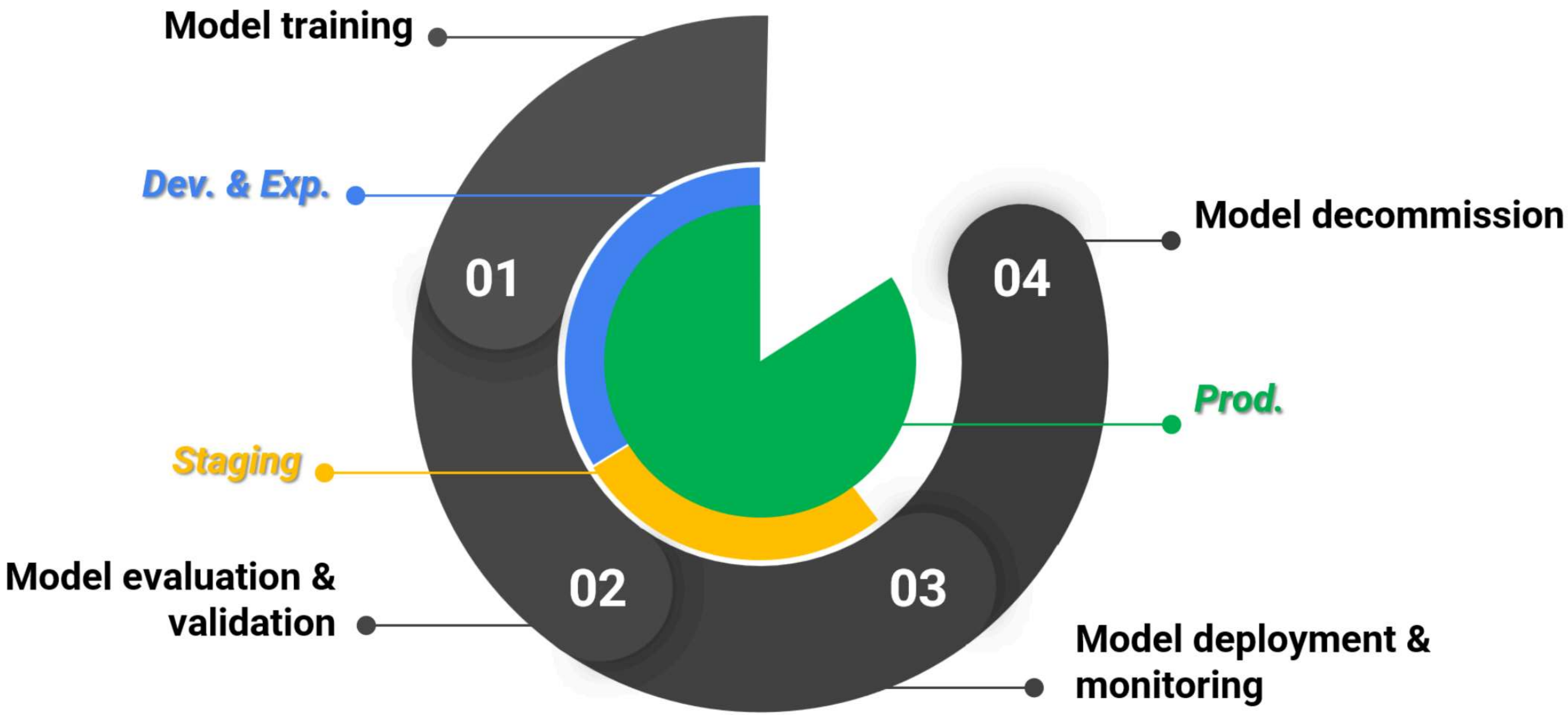
The model registry

FULLY AUTOMATED MLOPS

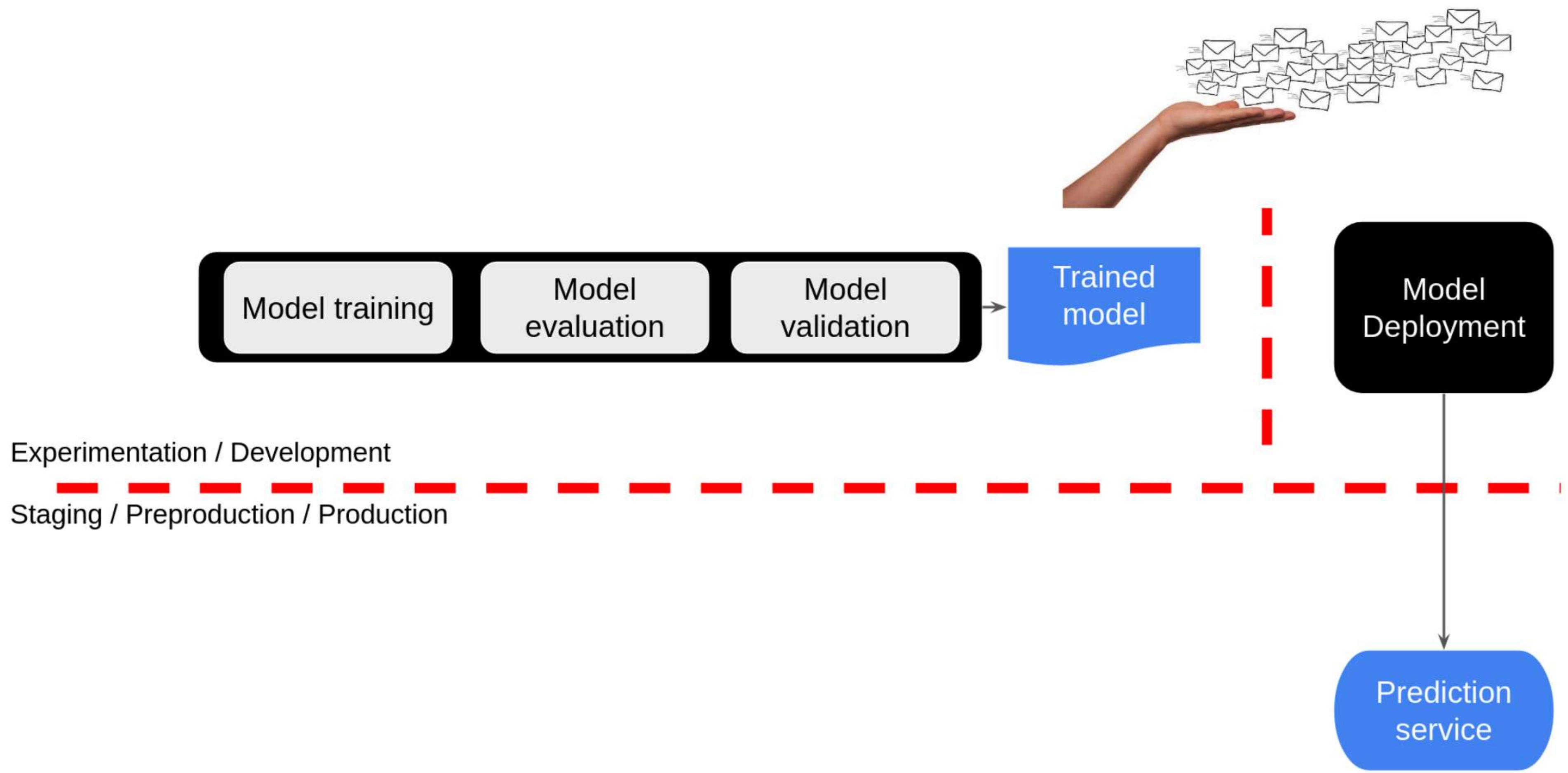
Arturo Opsetmoen Amador
Senior Consultant - Machine Learning

datacamp

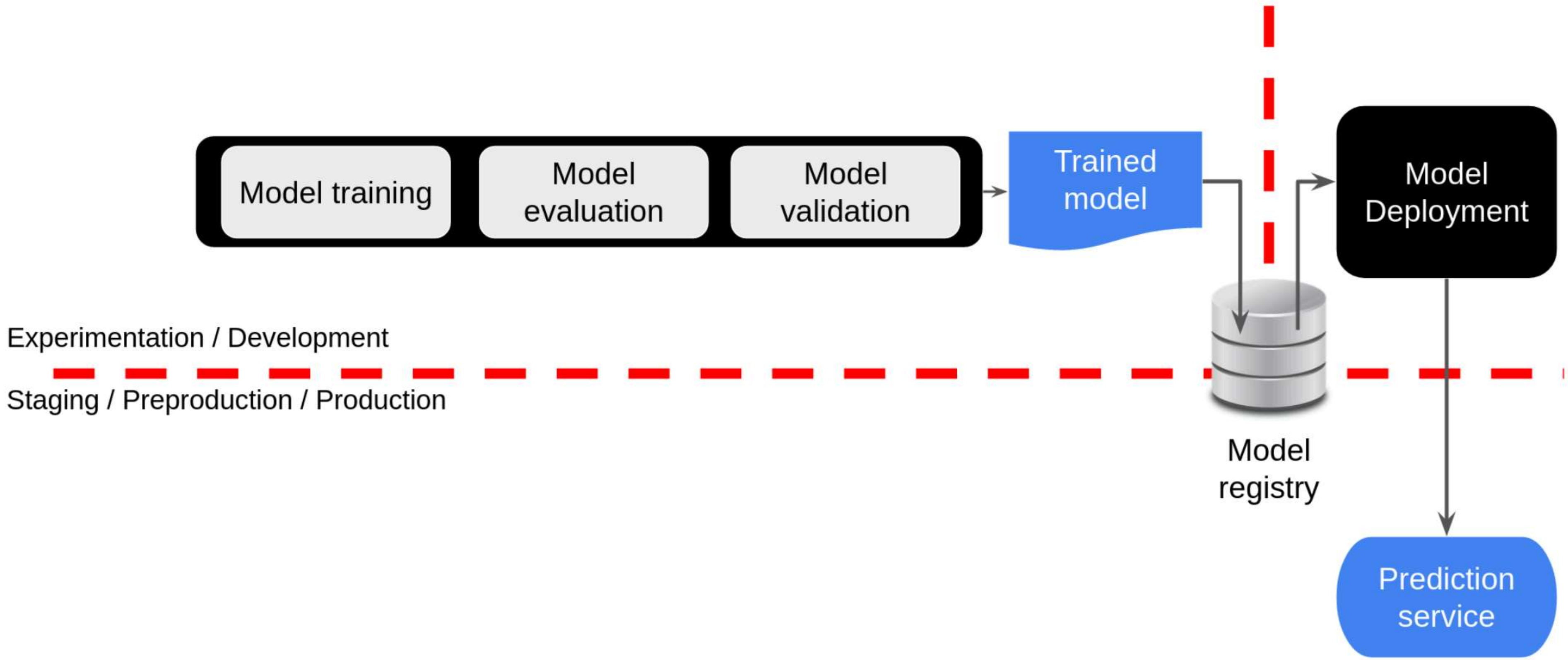
The ML model lifecycle



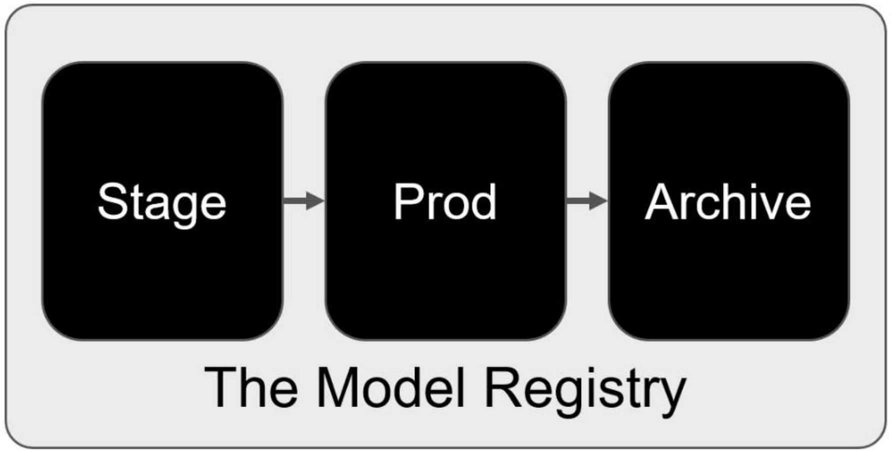
Throwing a model over the fence



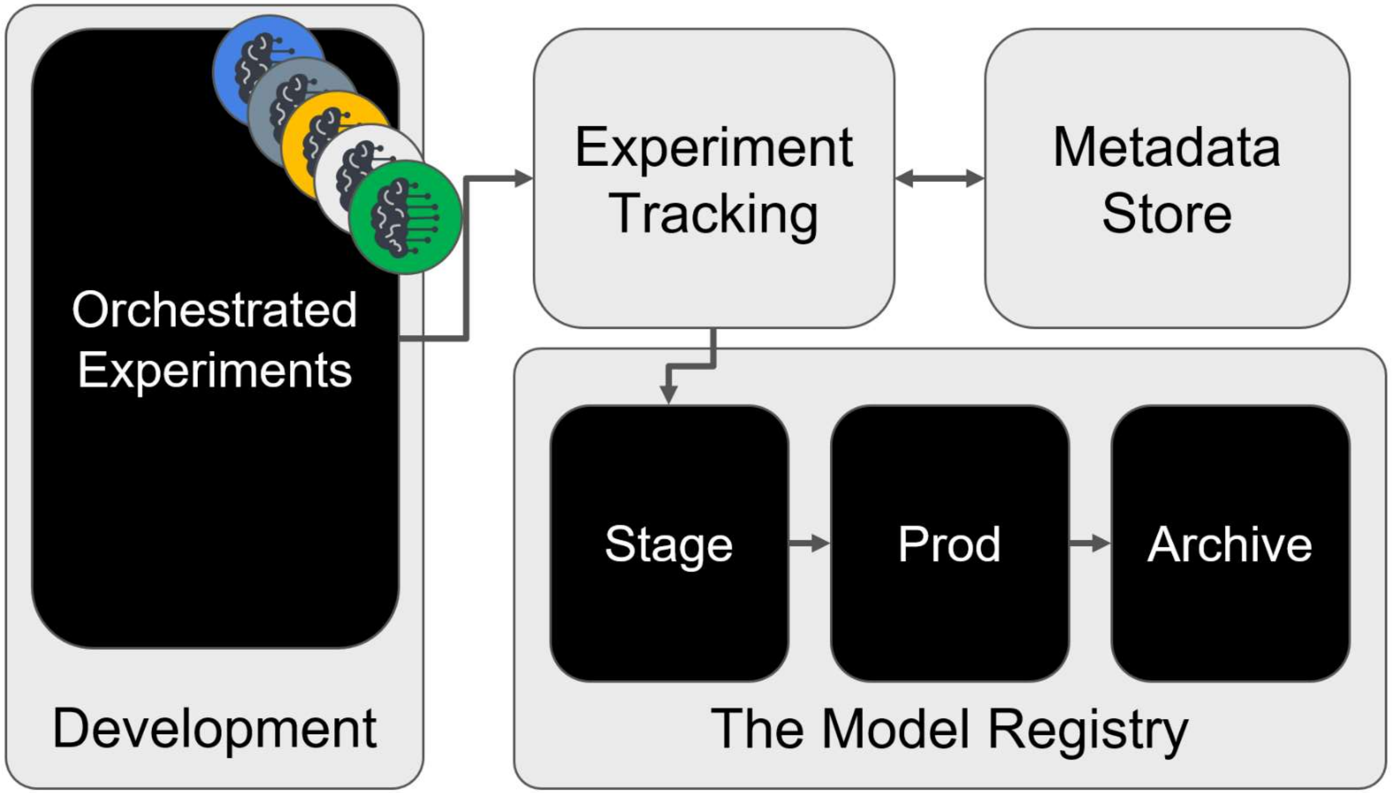
A first step towards automated MLOps



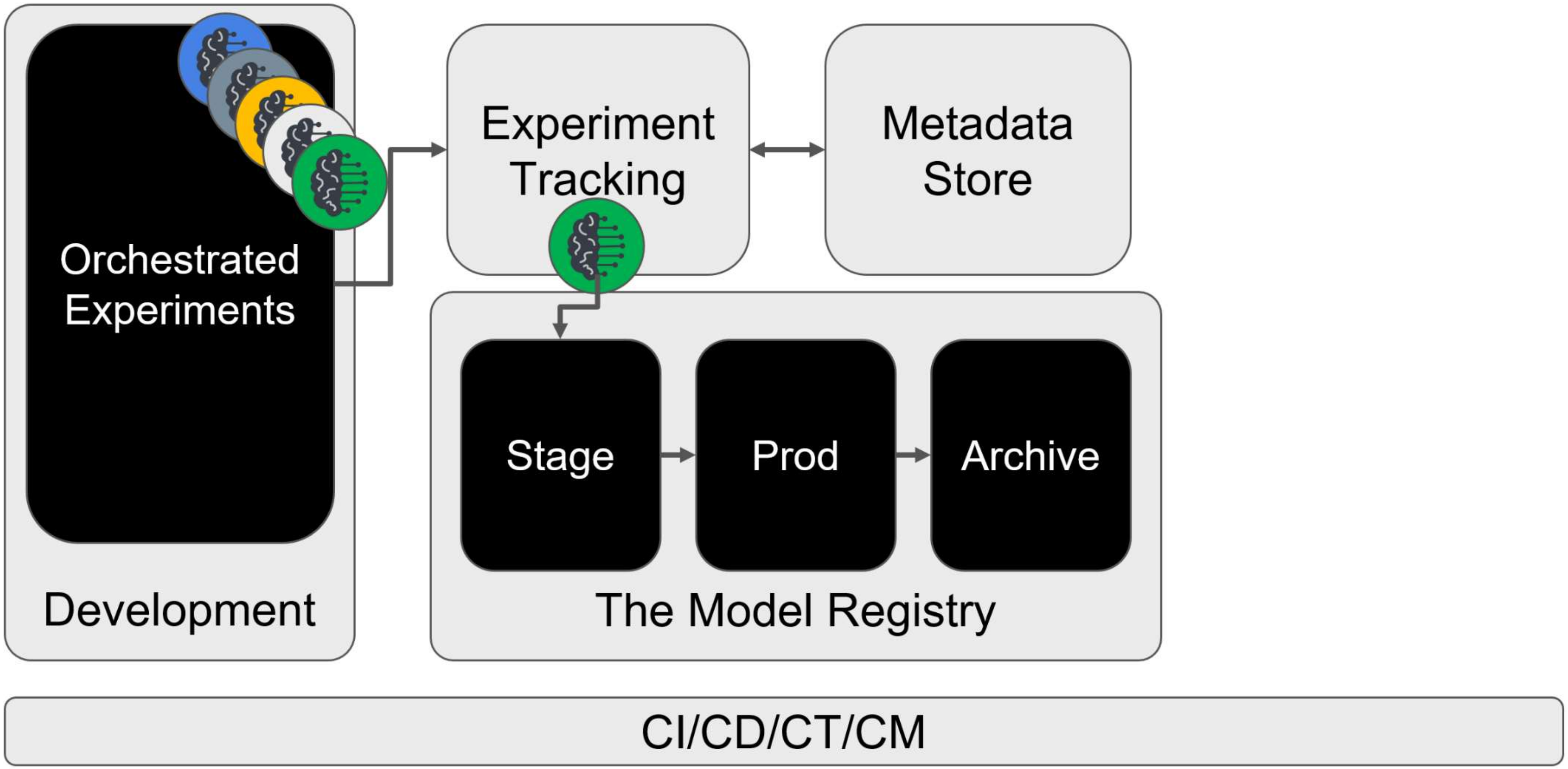
What is the model registry?



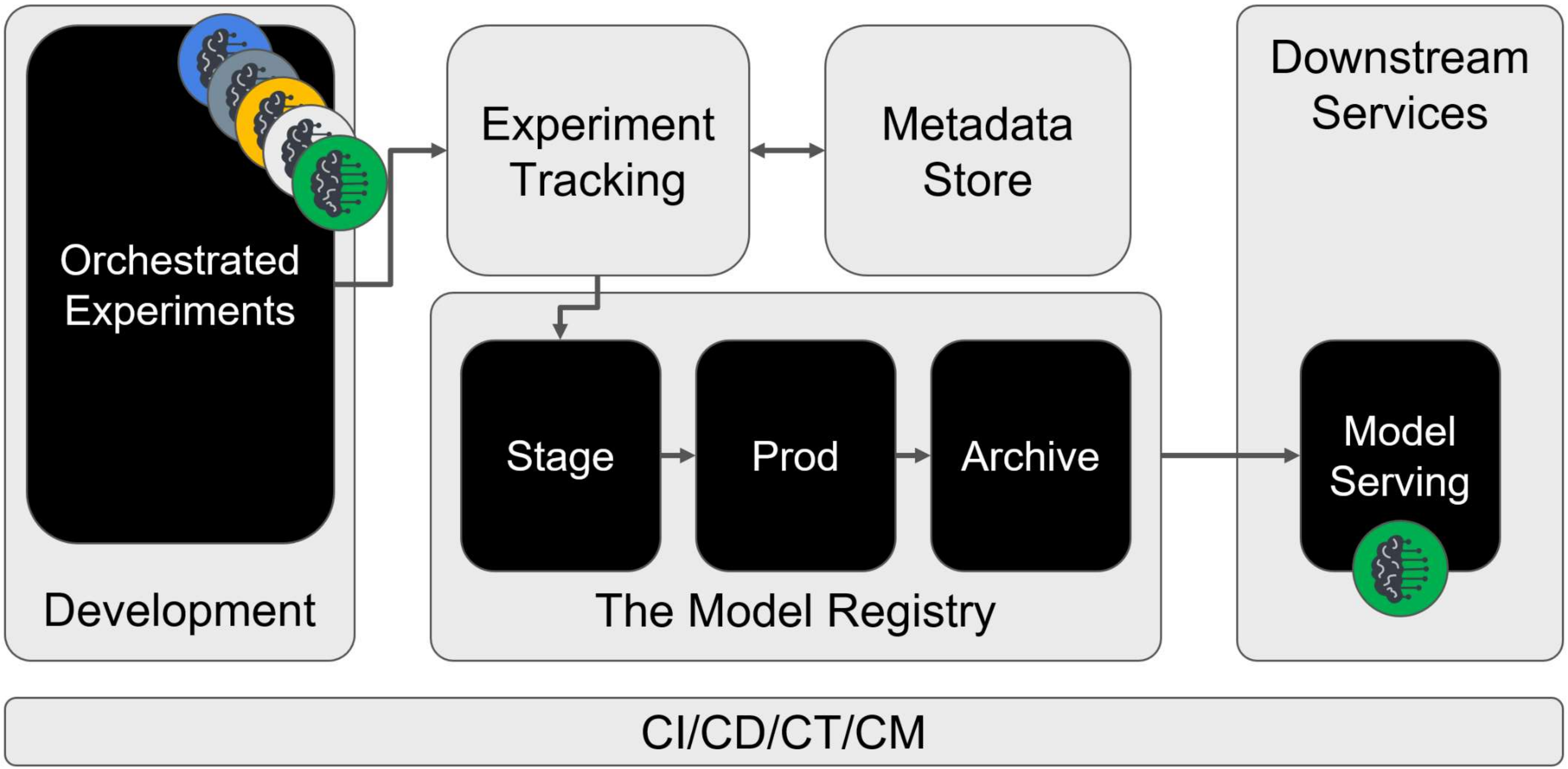
What is the model registry? - Experimentation



What is the model registry? - Registering a model

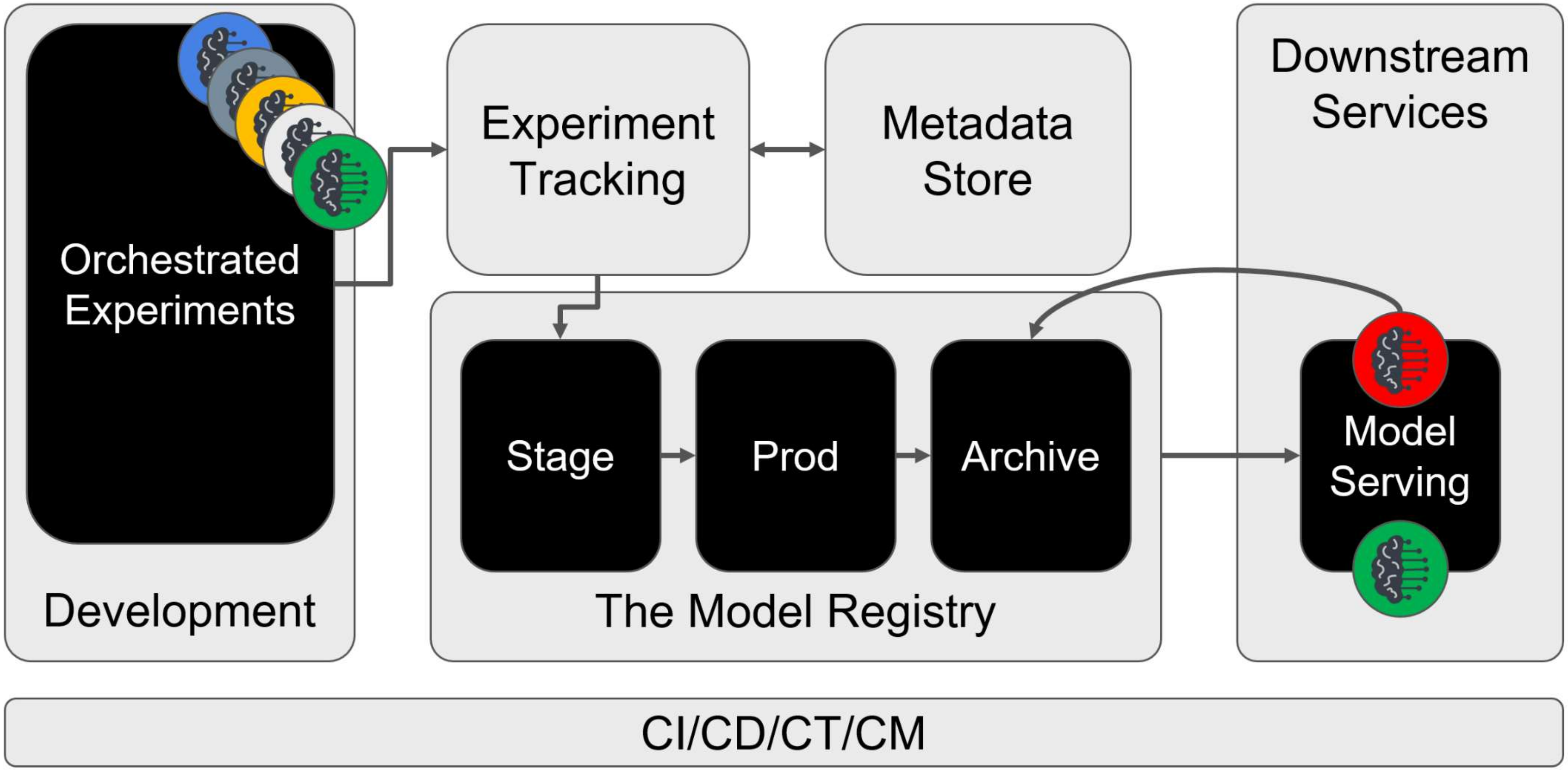


What is the model registry? - Updated deployment



FULLY AUTOMATED MLOPS

What is the model registry? - Model decommission



datacamp

FULLY AUTOMATED MLOPS

Let's practice!

FULLY AUTOMATED MLOPS

The feature store in an automated MLOps architecture

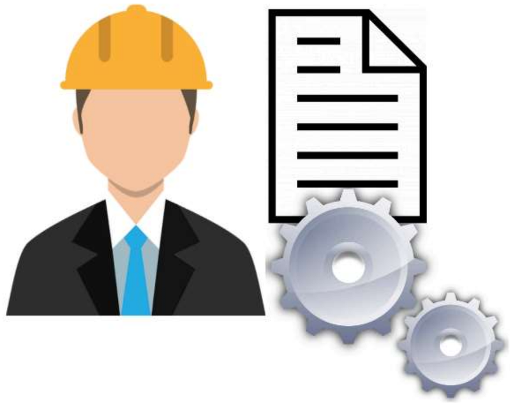
FULLY AUTOMATED MLOPS

Arturo Opsetmoen Amador
Senior Consultant - Machine Learning

Features in machine learning

Feature Engineering

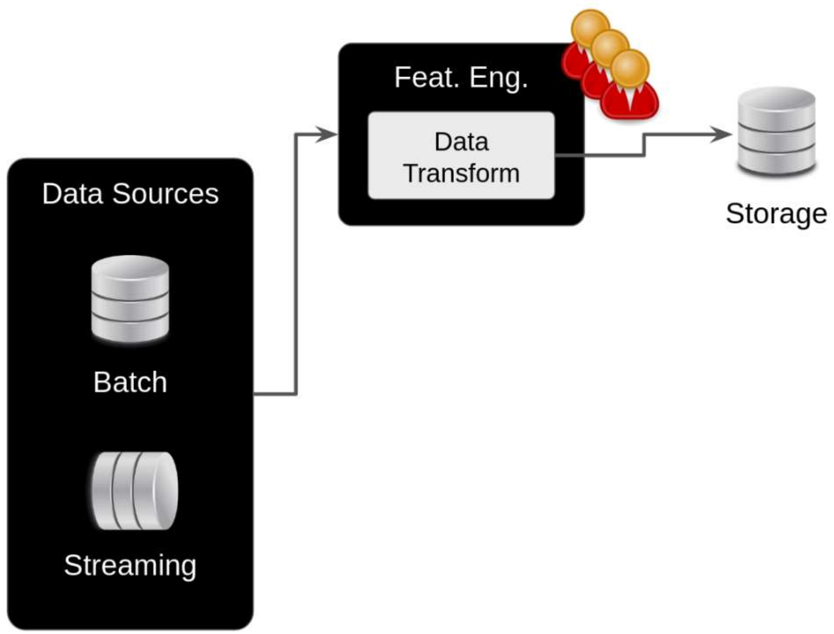
> Select, manipulate, and transform raw data sources to create features used as input for our ML algorithms



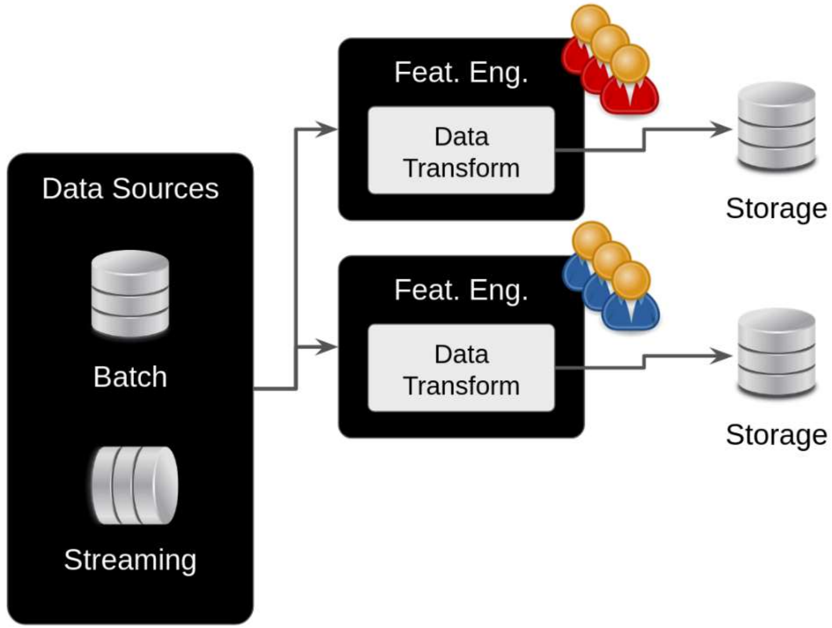
Examples:

- Numerical transformations
- Encoding of categories
- Grouping of values
- Constructing new features

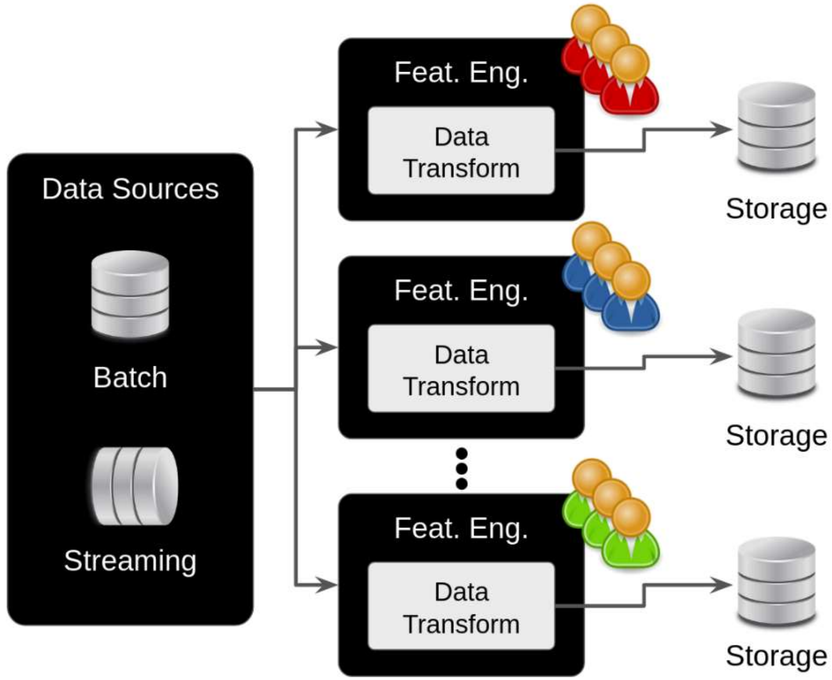
Feature engineering in the enterprise



Feature engineering in the enterprise

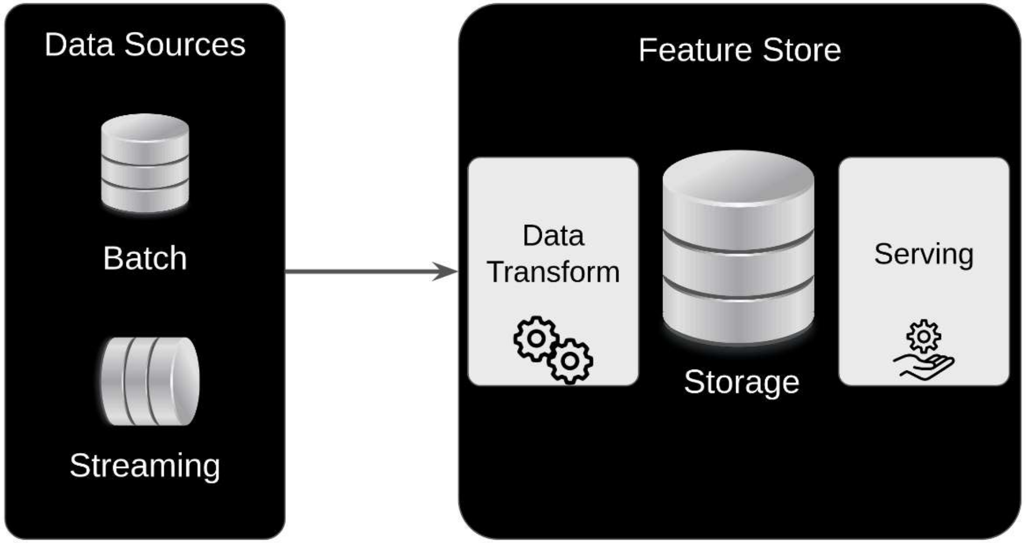


Feature engineering in the enterprise



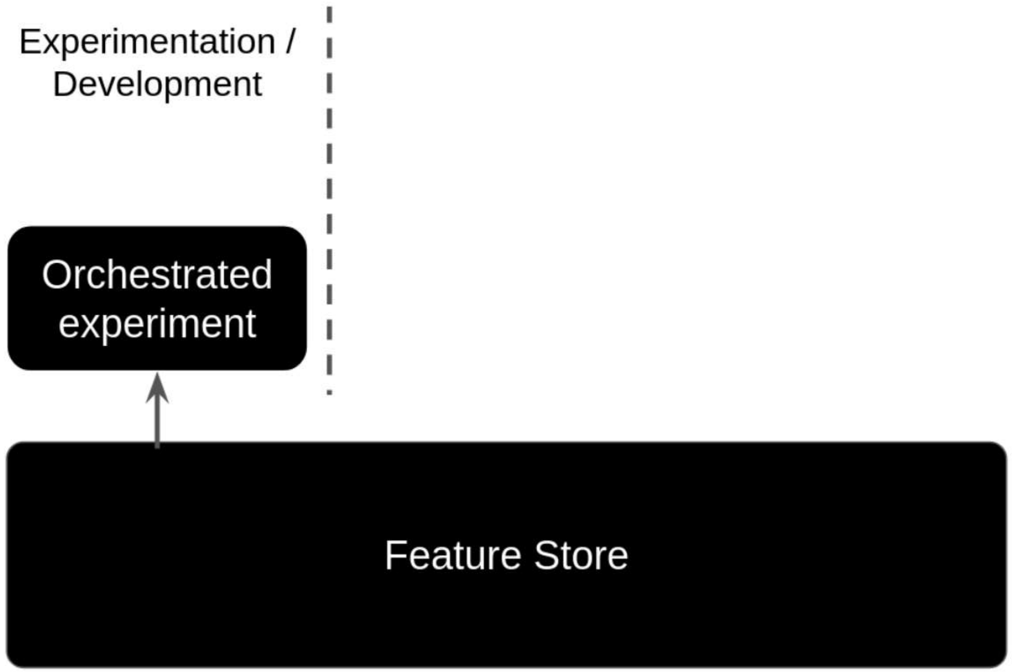
The feature store

- Centralized feature repository
- Avoid duplication of work with automation
- Transformation standardization
- Centralized storage
- Feature serving for batch and real-time



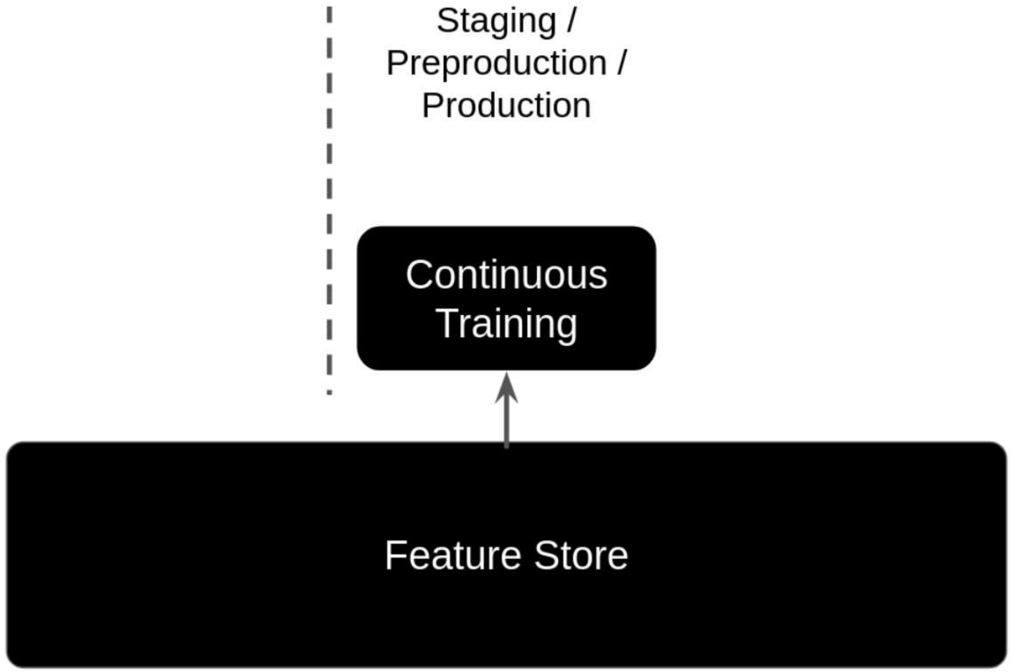
The feature store - Accelerated experimentation

- **Accelerated experimentation**
  - Data extracts for experiments
  - Feature discovery
  - Avoids multiple definitions for identical features



The feature store - Continuous training

- **Continuous Training (CT)**
  - Data extracts for automated pipelines in prod

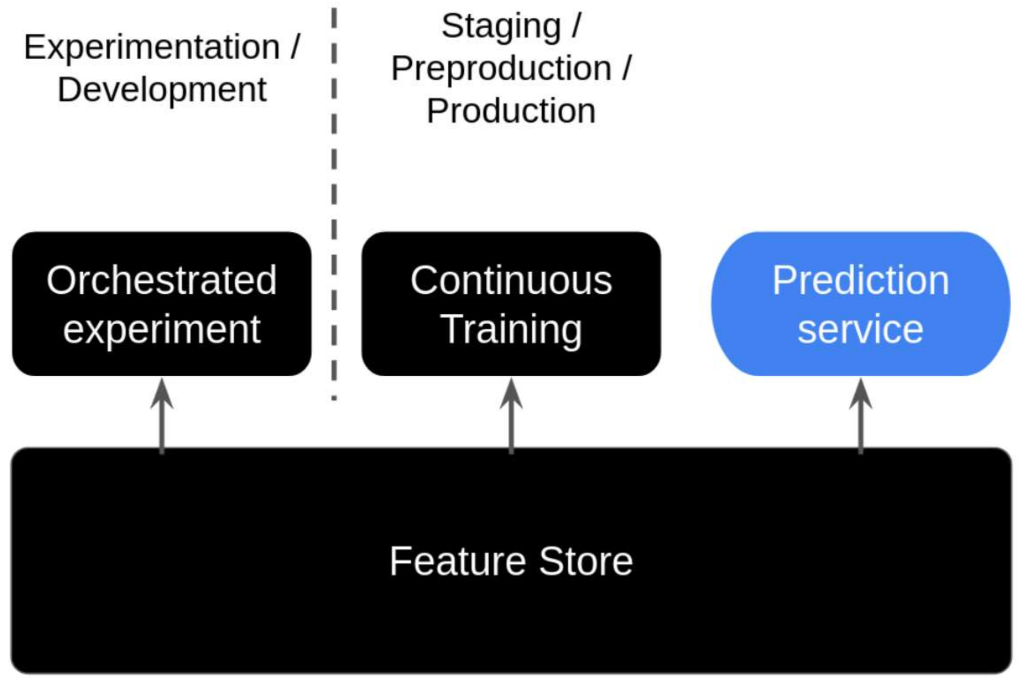


The feature store - Online predictions

- **Online predictions**
  - Use pre-defined features for prediction services

The feature store - Environment symmetry

- Avoids training-serving skew



Let's practice!

FULLY AUTOMATED MLOPS

datacamp

The metadata store

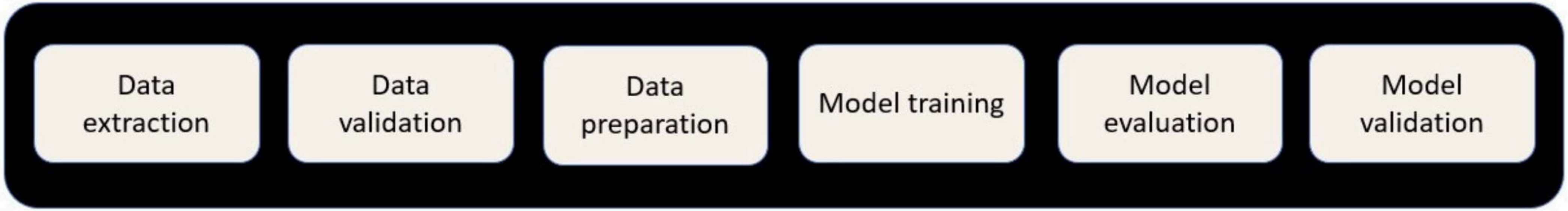
FULLY AUTOMATED MLOPS

Arturo Opsetmoen Amador
Senior Consultant - Machine Learning

datacamp

What is metadata in MLOps?

- **Metadata** is the information about the artifacts
  - created during the execution of different components of an ML pipeline



Metadata examples:

- Data versioning: Different versions of the same data are kept
- Metadata about training artifacts such as hyperparameters
- Pipeline execution logs

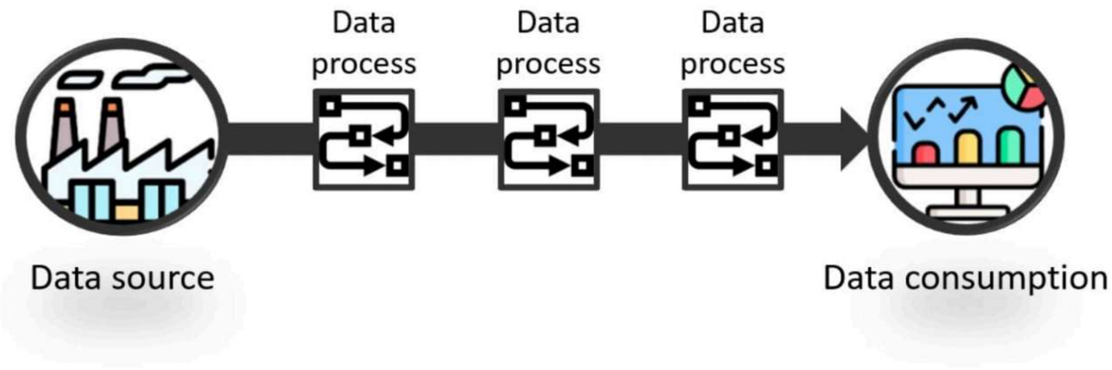
[1] https://datacentricai.org/

Important aspects of metadata in ML

- Data lineage
- Reproducibility
- Monitoring
- Regulatory

The importance of metadata - Data lineage

- Data lineage metadata tracks the information about data:
  - from its creation point
  - to the points of consumption



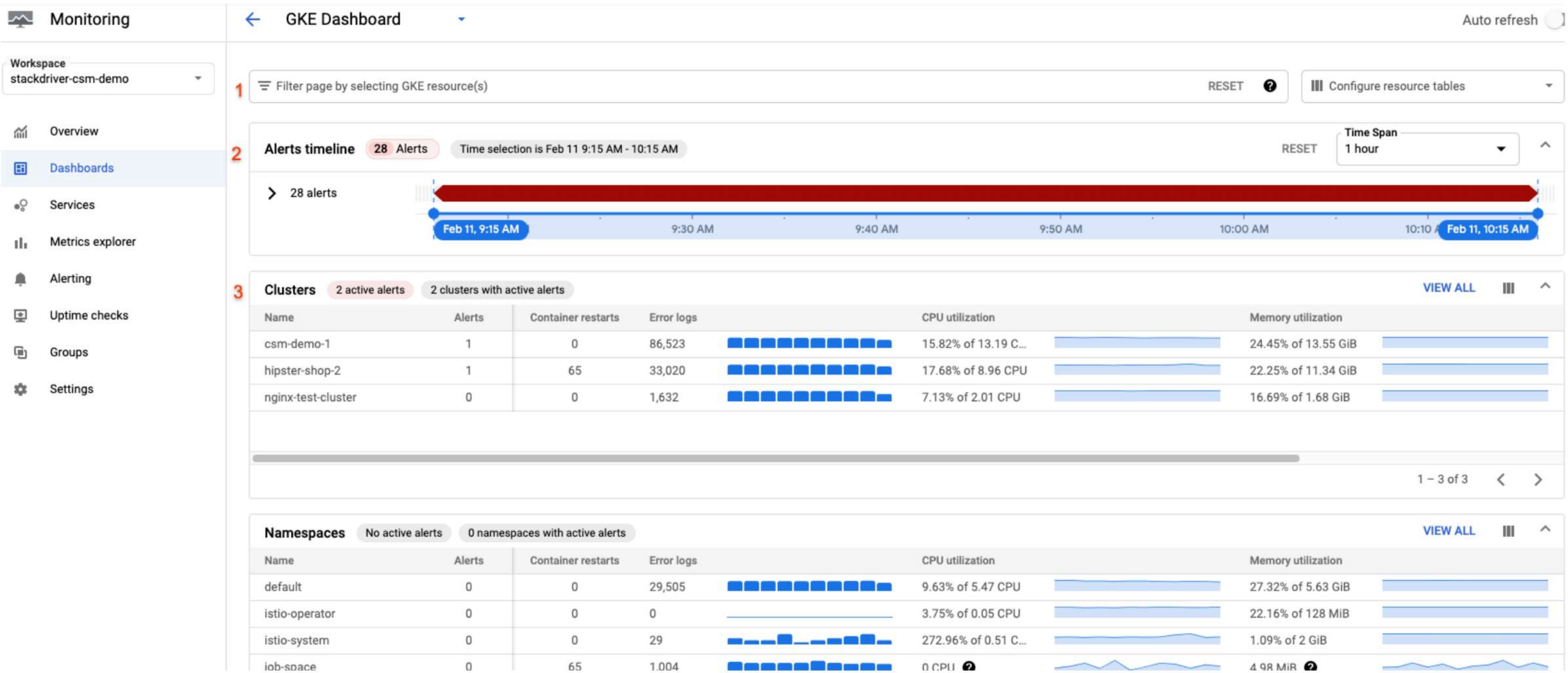
The importance of metadata - Reproducibility

- The metadata about our machine learning experiments:
  - Allows others to reproduce our results
  - Increases trust in our ML systems
  - Introduces scientific rigor to our ML process

The importance of metadata - Monitoring

- It allows machine learning engineers to
  - Follow the execution of the different parts of the MLOps pipeline
  - Check the status of the ML system at any time

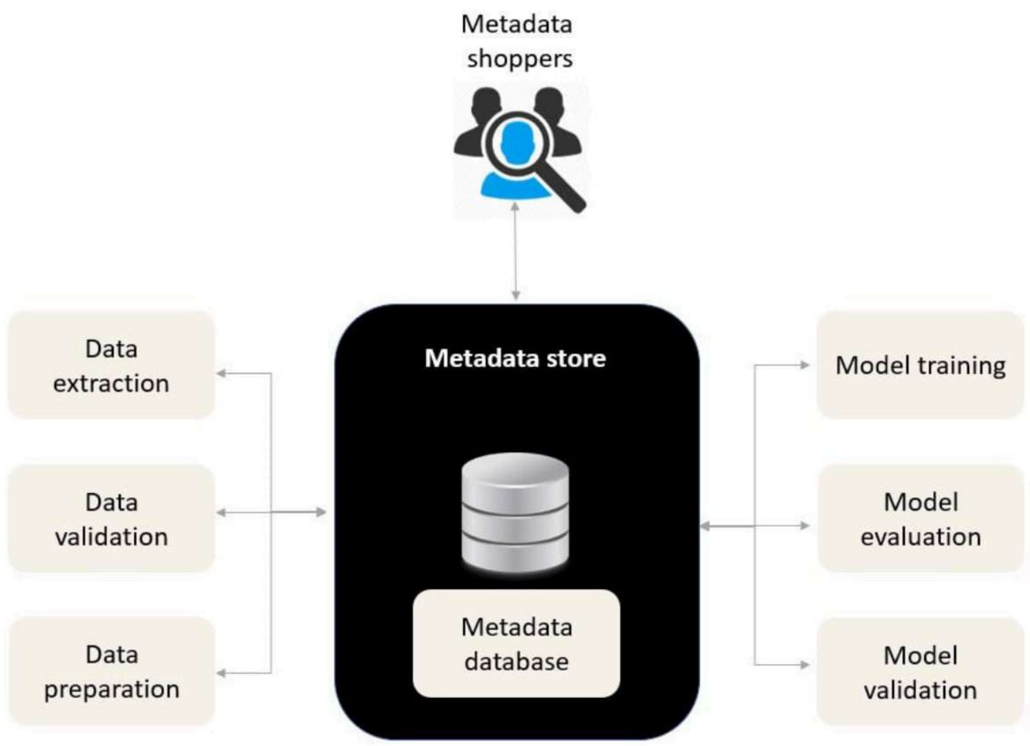
Example monitoring tool



[1] https://cloud.google.com/stackdriver/docs/solutions/gke/observing

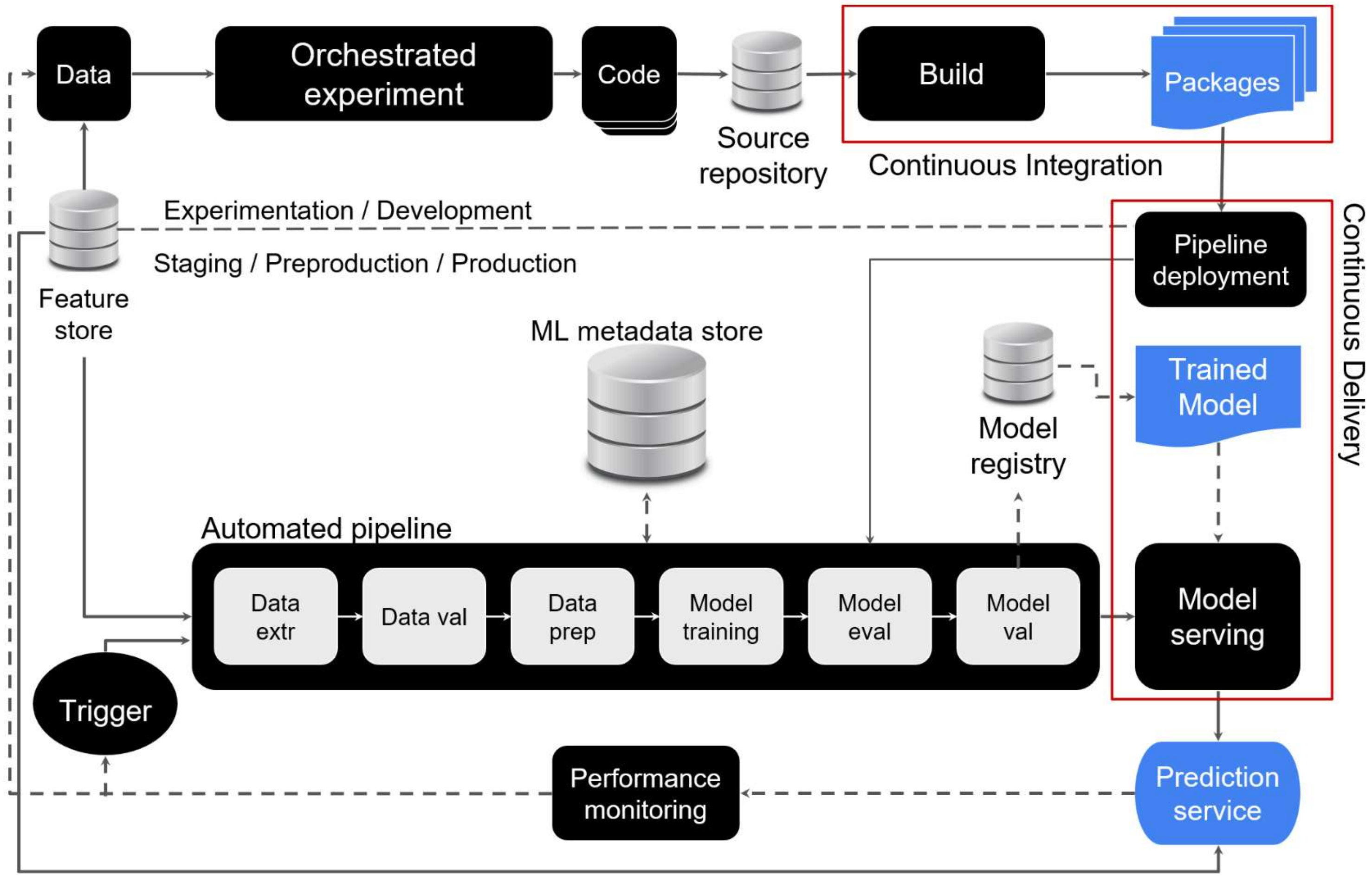
The metadata store

- Centralized place to manage all the MLOps metadata about:
  - experiments (logs)
  - artifacts
  - models
  - pipelines
- It has a user interface that allows us to:
  - read and write all model-related metadata



[1] https://cloud.google.com/architecture/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning

The metadata store in an MLOps architecture



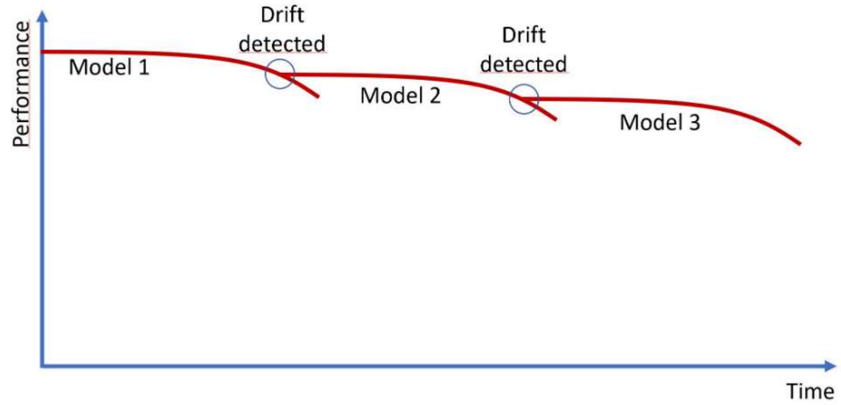
datacamp

FULLY AUTOMATED MLOPS

Metadata store in fully automated MLOps

It enables the **automatic monitoring** of the functioning of the fully automated MLOps pipeline

- Facilitating **automatic incident response.** For example,
  - Automatic model re-training
  - Automatic rollbacks



Let's practice!

FULLY AUTOMATED MLOPS

datacamp