MLOps Overview

Table of Contents:

What is MLOps (Machine Learning Operations)?

It's essentially a collection of Machine Learning, Software Development, and Operations, integrated in a way that allows for the reliable and effective deployment and maintenance of machine learning models in production.
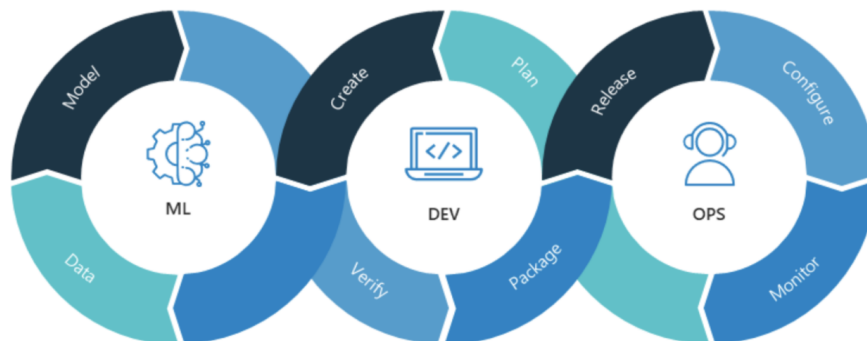

source: Nvidia blog

The complete MLOps process includes the following list of concepts:

1. Data Engineering
2. Version control of Data, ML models and Code
3. Continuous Integration and Continuous Delivery pipelines,
4. Automating deployments and experiments
5. Model Performance assessment and
6. Model Monitoring in production.

MLOps Stack Template:

Machine Learning Workflow is divided into multiple components. This template allows you to consider where you need tooling.
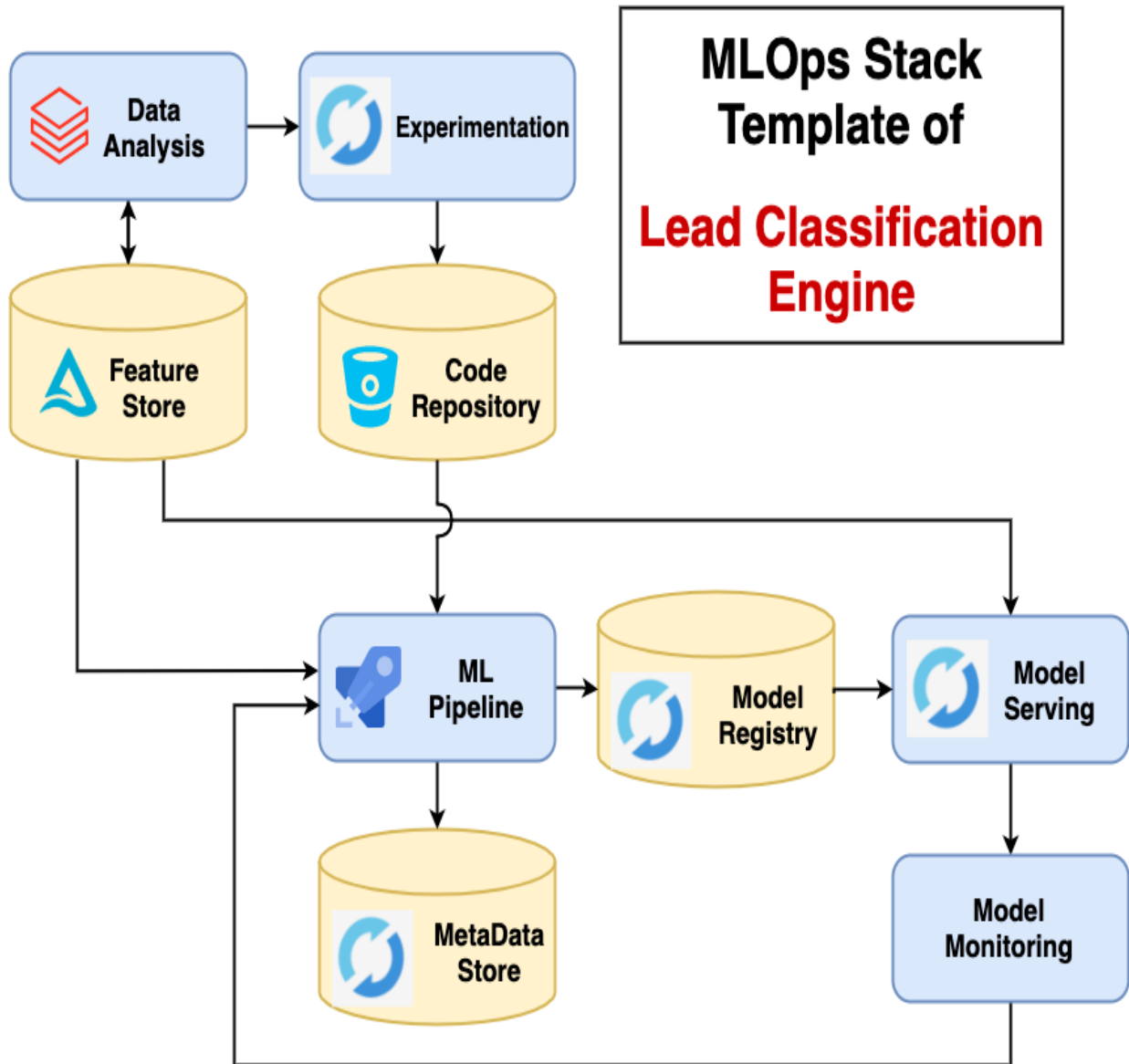
Template Source: Valohai

MLOps Production Model Monitoring Overview:

What is Model Monitoring and why it is important?

Tracking the performance of machine learning models in production. It helps the team to identify and eliminate a variety of issues, including bad quality predictions and poor technical performance. As a result, your machine learning models deliver the best performance.

What is Data Monitoring and Why?

Tracking model input data, which affects model performance.

If a deployed model receives an input sample that is significantly different from anything observed during training, model performance will decay. Monitoring data drift helps detect these model performance issues.

1. Feature Drift - Change in the distribution of model input data

2. Concept Drift - Shift in actual relation between model input and output

Methodology to Detect Drift:

Drift Detection

Detector	Tabular	Image	Time Series	Text	Categorical Features	Online	Feature Level
Kolmogorov-Smirnov	✓	✓		✓	✓		✓
Maximum Mean Discrepancy	✓	✓		✓	✓	✓	
Learned Kernel MMD	✓	✓		✓	✓		
Least-Squares Density Difference	✓	✓		✓	✓	✓	
Chi-Squared	✓				✓		✓
Mixed-type tabular data	✓				✓		✓
Classifier	✓	✓	✓	✓	✓		
Spot-the-diff	✓	✓	✓	✓	✓		✓
Classifier Uncertainty	✓	✓	✓	✓	✓		
Regressor Uncertainty	✓	✓	✓	✓	✓		

Source: SeldonIO

Refer to this link for the above methodologies

Top MLOps frameworks for Production Model Monitoring:

- **Evidently**
  - **Open-source** ML model monitoring system, which **generates interactive reports from pandas DataFrame**.
  - Ability to capture **Data Drift, Target Drift** (Numerical and Categorical targets)
  - Ease to integrate, platform-agnostic
  - **No Alert Mechanism available**
  - Can Log Metrics to MLFlow

Feature	Type	Reference Distribution	Current Distribution	Data drift	P-Value for Similarity Test ↑
worst concave points	num			Detected	0.000002

- **Whylogs**
  - **Open source**, and **high-performance statistical data logging** library
  - Ability to capture **model metrics and Monitor data quality**, Collects all possible **statistical metrics** about structured or unstructured data
  - Ease to integrate, platform-agnostic
  - **No Alert Mechanism available**

- **Fiddler**
  - **Managed** Model monitoring tool, **has very simple, clean, interactive dashboard**.
  - Visual tool to **monitor data drift, model metrics, data integrity and outliers.**
  - **Alert Mechanism available,** lets you set up alerts about production concerns.
  - Overall, it's a fantastic tool for monitoring machine learning models, with all of the capabilities you'll need.

o Pricing is not available on any source

● **Alibi Detect**

    o **Open source** python library focussed on outlier and drift detection,

    o Ability to capture drift for tabular data, text, images and time-series data

    o Ease to Integrate

    o Log metrics to MLFlow

    o **No Alert Mechanism**

References

● https://valohai.com/mlops-platforms-compared/

● https://ml-ops.org/content/mlops-principles

● https://neptune.ai/blog/end-to-end-mlops-platforms

● Get started with MLOps