10 Ideas for MLOps Projects



**Anish Mahapatra**

Contents

[List of Figures 3](#_Toc82419635)

[Introduction 4](#_Toc82419636)

[Designing a MLOps Solution 5](#_Toc82419637)

# List of Figures

[Fig 1: The complete MLOps process 4](#_Toc82419626)

[Fig 2: The MLOps Stack Template (Source) 5](#_Toc82419627)

# Introduction

Data Science is losing it’s title of the “Sexiest job of the 21st Century”.

*87% of Data Science Projects never make it to production -* [*VentureBeat*](https://venturebeat.com/2019/07/19/why-do-87-of-data-science-projects-never-make-it-into-production/)

Jobs over the next decade are going to leverage the fundamentals of Data Science and build on top of that for production. Data Science has flourished over the decade on the promise that companies will make money using analytics. Data Science can make money only when AI/ML projects are in production. **MLOps roles will replace Data Science jobs** in the next decade. MLOps seems to be the new buzz word floating around in the Data Science world and it is critical to understand what it really means.

MLOps aims to provide an end-to-end machine learning development process to design, build and manage reproducible, testable and evolvable ML-powered software. The cloud strategy of organizations plays a key role when it comes to production as custom solutions have to be built depending on the needs of the company. The complexity associated with MLOps is one of the main reasons we have not seen a one-size-fits-all approach. In this article, we will go through a few end-to-end projects that will explain how MLOps can be leveraged to make a Data Science project production ready.

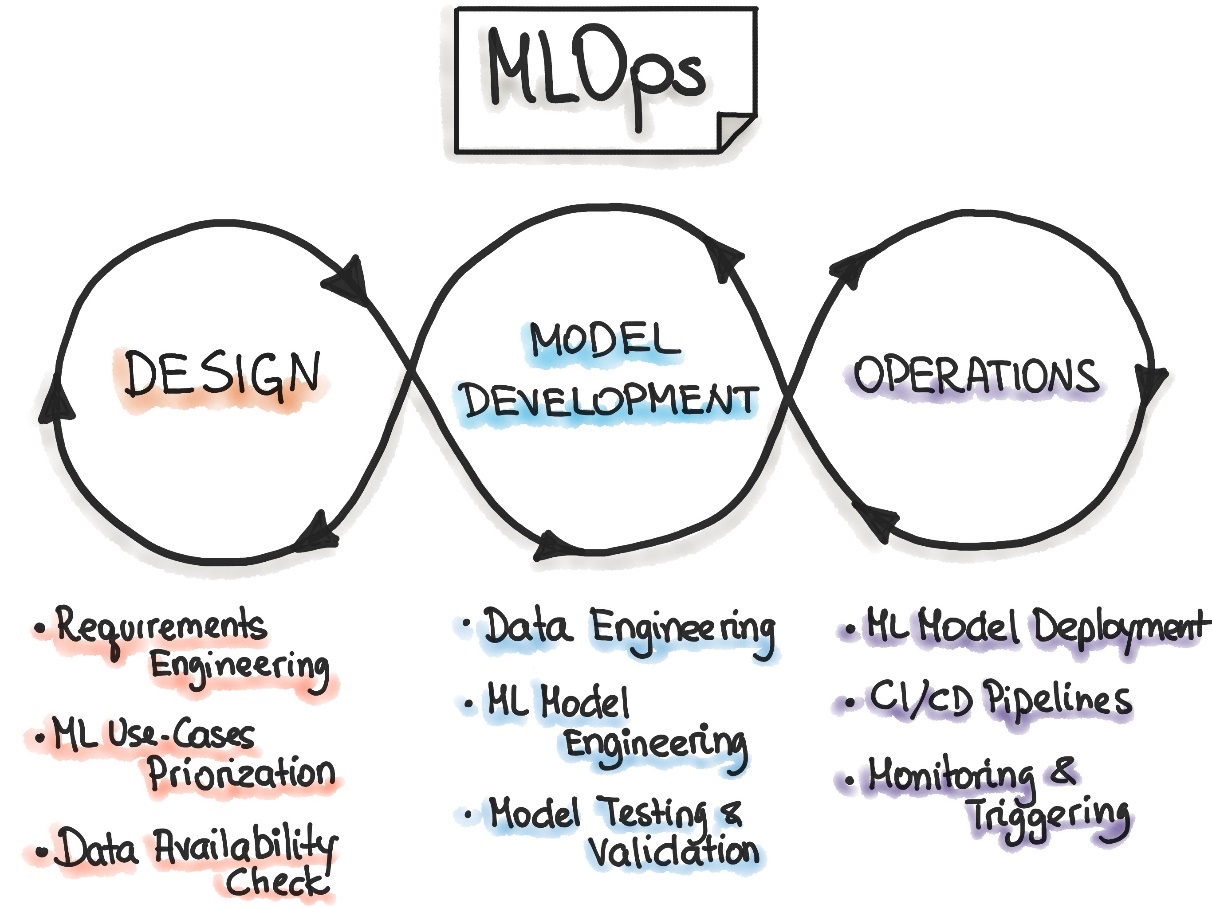


Fig 1: The complete MLOps process

# Designing a MLOps Solution

At the moment, the MLOps landscape is frequently evolving. The number of tools available for a single task are many – this leads to multiple options, which is alwas a good thing, but, it also means that we have to decide what will work best in out use-case. A good way to make projects more tangible is to focus on the stack. This can be complicated in deifferent settings and we have to make MLOps real for it to work. Where DevOps has a CI/CD approach to software solutioning, MLOps has a Continuous Integration (CI), Continuous Delivery (CD), Continuous Training (CT), Continuous Monitoring (CM) approach.

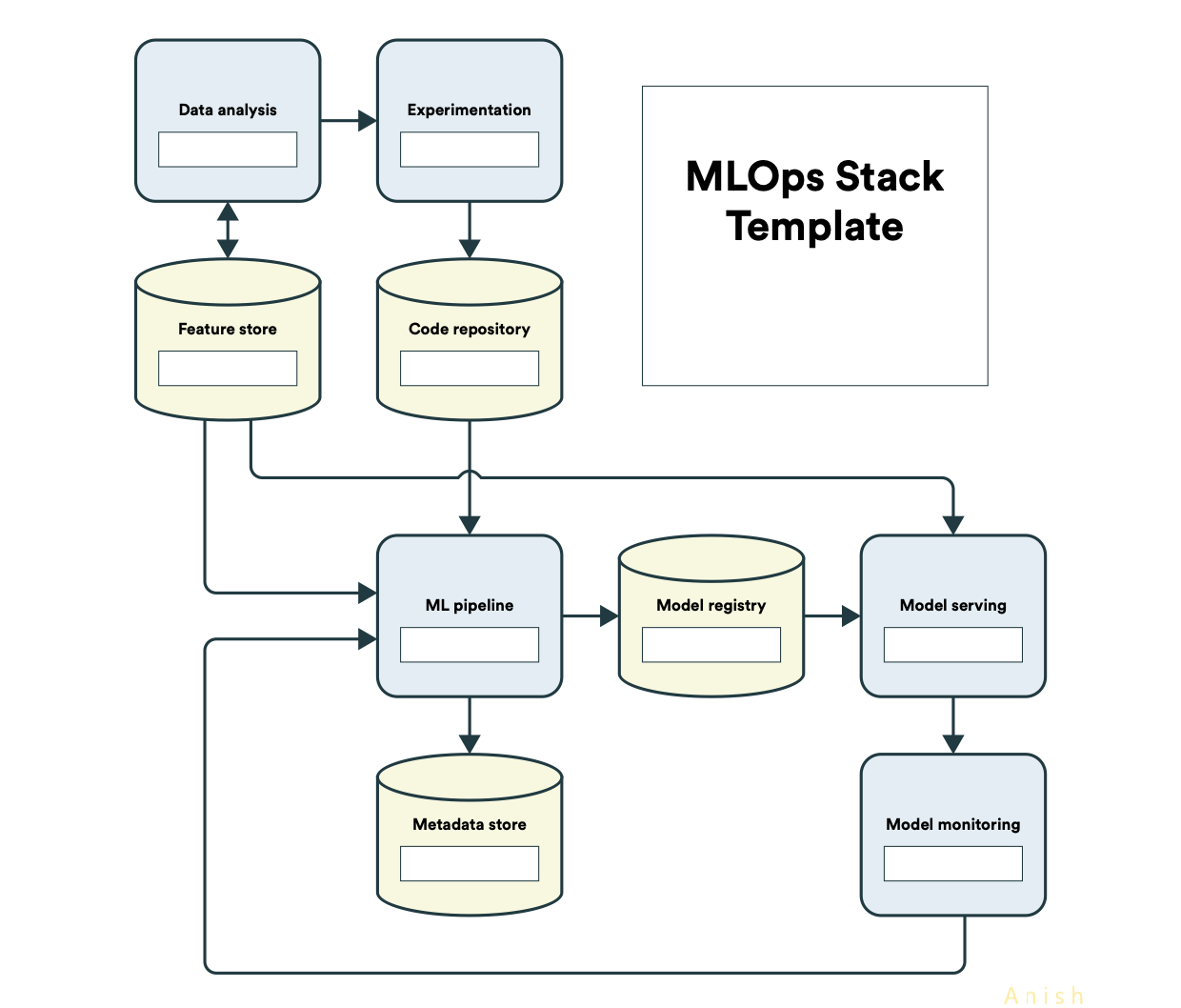


Fig 2: The MLOps Stack Template ([Source](https://ml-ops.org/content/state-of-mlops))

The MLOps Stack template is a good start on how we can think about MLOps solution design when implementing at scale for various organizations. The components are as follows:

* **Data Analysis**: The analysis component of the MLOps flow can be done with the help of various tools
* **Experimentation**: Output-focused experimentation along with domain knowledge can help select the relevant toolset
* **Feature Store**: Feature stores are used to store variations on the feature set leveraged for machine learning models that multiple teams can access
* **Code Repository**: The data and code repository has to be selected such that it fits into the MLOps stack being used, especially if it is on the cloud
* **ML Pipeline**: Similar to data pipeline, ML pipelines help carry the state of the machine learning project from data to ML output
* **Metadata Store**: Metadata for larger and evolving datasets can be housed in metadata stores
* **Model Registry**: Logging models are done in the model registry, this setup helps reflect back on multiple iterations
* **Model Serving**: Model serving is critical to production, it is the interface of machine learning with the real world
* **Model Monitoring**: Monitoring the model parameters in the real-world post production is critical to adjust various components to ensure that the feedback loop is met and the expected output is delivered in production consistently

Now that we have got a clear understanding on the components required to design for solution, we are on track to look at project ideas for MLOps. One nuance of the current state of MLOps is that it is in the very nascent stages of development – all major cloud platforms and various open-source applications are all trying to solve for production-ready machine learning. So, in the project ideas below, we will look at the project ideas where we can implement components of MLOps to make data science projects more MLOps ready.

# 1 Perfect the project Structure – Cookiecutter & readme.so

A

# 2 Automate your EDA to minutes – Pandas Profiling, SweetViz

A

# 3 Track Data Science Projects with CI, CD, CT, CM – DVC

A

# 4 Explainable AI / XAI – SHAP, LIME, SHAPASH

A

# 5 Deploy projects in minutes – Docker, FastAPI

A

# 6 End to End Machine Learning – mlflow, Seldon.ai

A

# 7 Model Registry – MLStore, FeatureStore

A

# 8 Big Data using Python, instead of PySpark – DASK

A

# 9 Build a Chatbot and deploy it (open-source)

A

# 10 FaaS framework implementation – Apache OpenWhisk, OpenFaas

A