**Assignment 3**: Essay on MLOps Lifecycle, including the integration of DevOps, DataOps, and ModelOps

**------ SANTOSH BR**

Preface: This document explains the need of MLOps Lifecycle including integration of DevOps, DataOps and Model Ops.

## Why Lifecycle required for MLOps:

In a large enterprise or in any organization if we need to rollout products based on AI/ML techniques we need to have a sustainable MLOps Lifecycle. This helps in smooth roll-out of the products / features based on AI/ML technology.

## What is MLOps Lifecycle:

MLOps involves many divisions of **Data Science**. In large scale organisation we will have individual teams or stakeholders to manage each one of them. In smaller organization few people will be taking care of the multiple roles.

Stakeholders: Business decision makers, data engineers, data scientists, ML engineer, Devops engineer, Customers etc.,

### Lifecycle:

A diagram of a diagram of a project

Description automatically generated with medium confidence

The **Business management** had to make a careful study of the problem statement as the ML systems are expensive, time consuming to setup and roll-out and maintain.

Once the problem statement is identified the management had to identify the right stake holders, the **data scientists** who are subject matter experts should be involved in preparation of the data and continuously involved in development of the ML System.

The **data engineers** have roll to run through all the DataOps related tasks. It can involve data munging, cleaning or preparing new set of data as the business requirement changes. The data engineers will be maintaining the continuous DataOps pipelines.

Then comes the **ML engineers** who will work on the given data. They will work on the model to be selected and apply the model. The predictions, accuracy of the solution are developed and automated by the ML engineers using the required ML model.

Then comes the **Devops engineers** who have the key role to develop the complete infrastructure and manage/enhance as the business need. The Devops Engineer generally plays the roll of developing the monitoring solutions for the ML pipelines. Developing smooth deployment pipelines and rollout of the product/features needs a Devops pipelines.

Product/Feature analysis of the deployments is done either by **Business analysts** along with the management.

As, mentioned above the lifecycle goes through many divisions of and individuals, developing a stable lifecycle flow is very important for long-term standing for the AI/ML products/features. Each stakeholder must closely work with each other for a sustainable ML System.

## DataOps for ML Lifecycle:

Data is critical for any ML system, having a good DataOps will help in sustainable and predictable ML system and output.

### Some key points about the DataOps:

What type of data to use for given ML System?

Are we reading/writing the data in **JSON / CSV / Databases**. Developing solution of converting the given set of data to the appropriate format is very important. This helps in easy expansion of the system for future needs. Automating the data **CRUD** activity and having unit tests for the sample of the data are key tasks of the DataOps. Cleaning of the data / error handling is another task for DataOps engineer for a robust ML System.

What is the expected latency/iops of the large set of data?

Since we are taking of big data here, we need to keep a close watch on the latency which can degrade the ML System. Having low latency is key for many ML systems. Making decision of on-prem infra or Cloud infra usage should be analysed before rolling out solution. Since we read / write data frequently for the ML System having a clear understanding on the latency and the iops of the storage systems will help in manging the performance of the ML System.

## DevOps for ML Lifecycle:

DevOps plays a major role in developing a sustainable ML System. M/L Devops engineers have a larger role than the regular software devops engineer.

Devops teams should develop CI/CD pipelines for the ML System. A stable CI/CD pipelines helps in identifying the product/feature deficiencies at very early stage. Deployment can be faster with a fully automated CI/CD pipeline. The CI/CD pipelines will also help business management/analysts (including customers) to do requirement analysis of the business need.

Along with the deployment responsibilities the Devops teams has a key role of developing the ML System infrastructure. When it comes to ML Systems its expensive solution so effective use of infrastructure is a key aspect. The compute / storage / RAM are expensive requirement of the ML System. The ML models which use heavy compute / storage needs a close understanding of the resource requirement and usage. The Devops teams should work with the ML Engineers for a efficient usage models and systems.

With effective usage of resource comes the monitoring and alerting. The Devops teams had to develop a robust monitoring / alerting solutions so that the breakdowns can be cut down and the resources are used efficiently.

Also, the Devops must work with dataops in developing scalable ML System.

## ModelOps for ML Lifecyle:

The key system which works on the data to develop a AI/ML based product. The ModelOps where the ML Engineers will work on developing tools and automations for analysing data and coming up with the ML models as per the business management requirement.

The modelops needs a good understanding of the problem statement and had to work closely with all other stakeholder particularly the data scientist (subject matter experts) to understand and leverage their experience and develop the required ML model.

The modelops will deicide the required features/hyper parameters for finding a ML solution for the given problem. The modelops will work on improving the accuracy of the ML model by using different ML models and applying them. They will work closely with dataops to see the required data is in place and also work with devops for the resource usage and rollouts.

The modelops are the brain behind a successful AI/ML based product, so they must keep improving the solution / accuracy of the predictions. The modelops should develop the pipelines such that they are sustainable and easily expandable for the future business needs.

## Conclusion:

As mentioned earlier ML Systems and solutions are expensive both in terms of resource and time. Having a good ML System where the workflow is well defined from…

**problem statement -> data gathering -> ML Model development -> Deployment -> monitoring -> Business analysis is key for a stable AI/ML based solution**.

Develop and automate each stage and identify the stakeholders for a successful and sustainable AI/ML System.