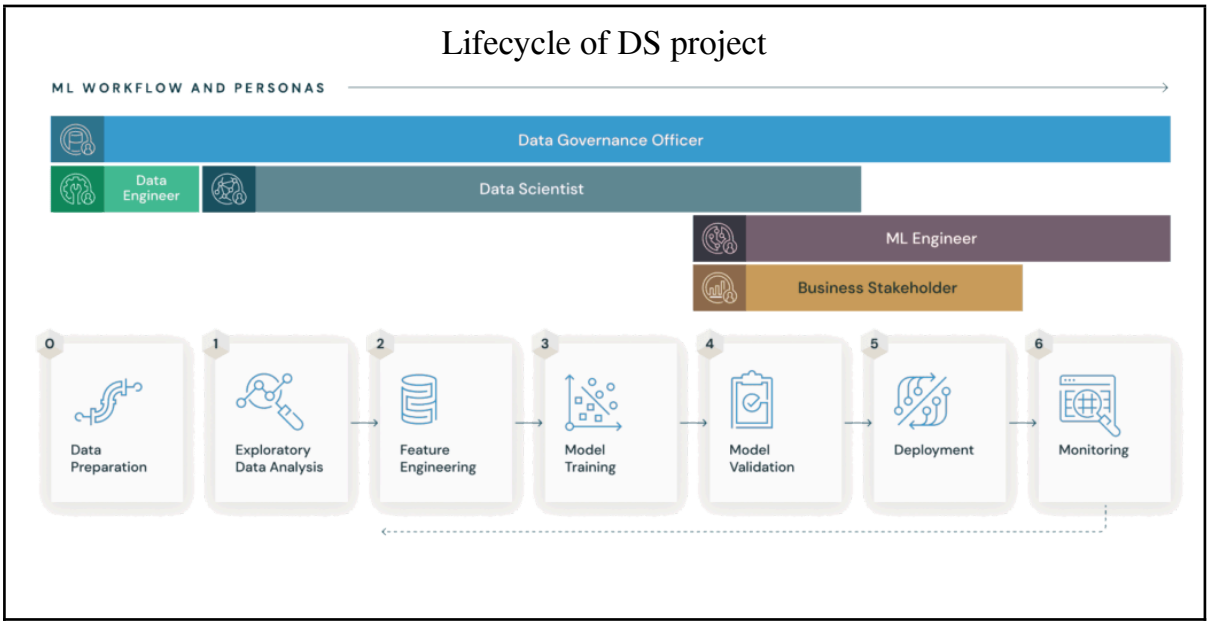
MLFlow

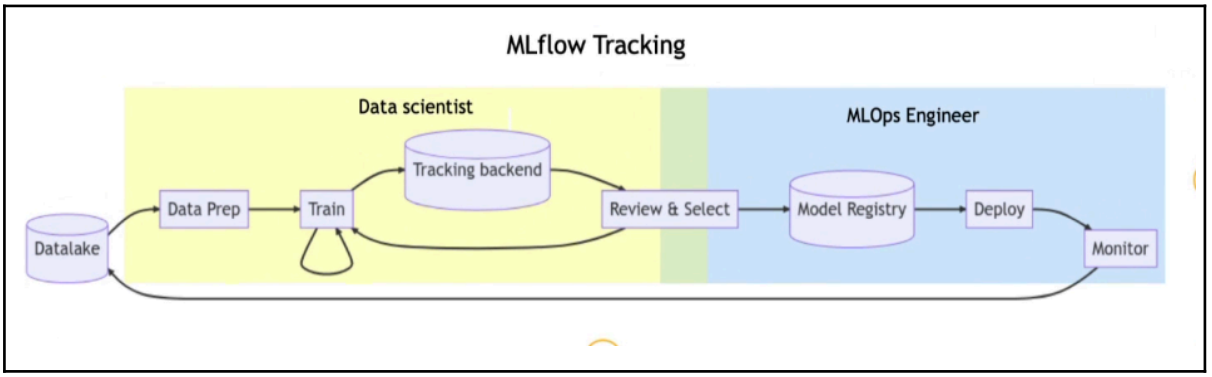
MLFlow is an open-source MLOps platform designed to manage the complete machine learning lifecycle from experimentation to deployment. It provides a comprehensive set of tools and frameworks to manage and track the end-to-end ML development process, including experimentation, reproducibility, deployment, and collaboration. MLflow enables data scientists and ML engineers to focus on building and deploying models while maintaining visibility, control, and reproducibility.

The core concept behind MLFlow is that you will be able to run ML and Generative AI projects that can solve complex, real-world challenges. It organizes work around experiments (grouping related runs) and runs (individual model executions with logged data). It supports deployment to REST APIs, cloud platforms, or edge devices, ensuring scalability.



- In data preparation, the data engineer is responsible for creating ETL Pipelines so that they can aggregate data from various sources (APIs, databases) and after aggregating that data, they will try to store it in some kind of source like MongoDB or any other database.
- One the ETL pipeline has been created by data engineers, then the data is available in the source. The data scientist will now take those entire data from the source and they'll start the life cycle of a data science project.

- The data scientist will perform EDA from where they are going to analyze the data, perform hypothesis testing, statistical analysis and more.
- After EDA they will go to feature engineering where they will handle missing values, check how many categorical variables are there, what kind of categorical encoding needs to be done, check if the dataset is imbalanced or not.
- Data scientists are extensively involved until the model validation. During the model validation both Data Scientist and ML Engineers will be taking part. However when we talk about deployment, ML Engineers will be extensively doing this particular part because the main aim is basically to deploy the model in an efficient way and to make sure that our entire application is scalable.
- ML Engineers play an important role in Model Validation, Deployment and Monitoring. In model monitoring, they need to implement concepts like data drift which will help to understand when their model needs to be re-trained.



What does a Data Scientist do with MLFlow?

A Data Scientist leverages MLFlow for:

1. Experiment Tracking & Hypothesis Testing.
2. Pipeline Creation & Code Structuring.
3. Model Packaging & Dependency Management.
4. Evaluating Hyperparameter Tuning.
5. Compare the results of Models retraining over time.

What does a MLOps Professional/ML Engineer do with MLFlow?

A Machine Learning Engineer leverages MLFlow for:

1. Manage the lifecycle of trained models, both pre and post deployment.
2. Deploy models securely to the production environment.
3. Manage Deployment dependencies.

How does MLFlow come into usage for prompt engineering users?

A prompt engineer leverages MLFlow for:

1. Evaluate & Experiment with different LLMs.
2. Create Custom Prompts & Experiments.
3. Deciding on the best base model suitable for project requirements.

MLFlow Overall Usecases

1. Experiment Tracking (Parameters & Metrics)
2. Model selection & Deployment.
3. Model Performance Monitoring.
4. Collaborative Project.