Build CI/CD Pipeline for Machine Learning Projects using Jenkins

Business Overview

Delivering a new software version requires a sequence of actions known as a continuous integration and continuous deployment (CI/CD) pipeline. A "CI/CD pipelines" technique aims to automate software delivery to improve software quality across the software development life cycle. Organizations may produce higher-quality code more quickly by automating CI/CD across the software development lifecycle's development, testing, production, and monitoring stages. The real benefit of CI/CD pipelines is realized through automation, even though it is feasible to complete each step manually.

A pipeline is a procedure, often known as CI/CD, that directs software development along a route of developing, testing, and deploying code. Automating the procedure aims to reduce human error and keep the software release process consistent. The pipeline may contain tools for code compilation, unit testing, code analysis, security, and binaries production. This process would also involve packaging the code as a container image for deployment across a hybrid cloud in containerized settings.

A DevOps methodology's foundation is CI/CD, which unites teams from IT operations and development to deliver software. The speed at which code can be published has evolved into a competitive differentiator as specialized applications become essential to how businesses differentiate themselves.

We aim to build a CI/CD Pipeline for this search engine application using Jenkins, which is an open-source automation server.

Prerequisite:

We have already developed a search engine application using the Faiss similarity search algorithm deployed on streamlit. A search engine is an application that helps individuals find information online by using keywords or phrases. The Projectpro search bar app collects all project descriptions, organizes them into an efficient index, and returns the top results. Semantic search aims to increase search accuracy by comprehending the search query's content. This project uses SBERT, a variant of the conventional pre-trained BERT network that employs siamese and triplet networks to generate sentence embeddings for each sentence, which can then be compared using cosine-similarity, allowing a semantic search for a considerable number of sentences (only requiring a few seconds of training time). Facebook AI Similarity Search (Faiss), a library that allows us to search for similar multimedia documents quickly, is then used to create a query index from all the documents to return the top results.

Tech Stack
➔
Language: Python
➔
Services: AWS EC2, Docker, Streamlit, Jenkins, Github.

Jenkins:
Jenkins is a Java-based open-source automation tool with continuous integration plugins. Jenkins is used to continually create and test your software projects, making it easier for developers to integrate changes to the project and for users to acquire a new build. It also enables you to provide software continuously by interacting with a wide range of testing and deployment platforms.

Organizations may use Jenkins to automate and speed up the software development process. Jenkins unifies all development life-cycle operations, including build, documentation, testing, packaging, staging, deployment, static analysis, and much more.

Key Takeaways
- Understanding the streamlit application
- Creating an EC2 instance
- Setting up docker on the EC2 instance
- Understanding the different branches of the GitHub repository
- What is personal access token in GitHub?
- Generating new personal access token in GitHub
- Connecting to EC2 instance using SSH
- Deploying the streamlit application on the EC2 instance
- Cloning to the GitHub repository from the EC2 instance
- Setting up the Jenkins server on the EC2 instance
- Configuring Jenkins server
- Understanding the various components of Jenkins Dashboard
- Installing additional plugins in the Jenkins server
- Understanding the different jobs in the Jenkins server
- Creating a Freestyle job in the Jenkins server
- Creating a Pipeline job in the Jenkins server
- Understanding the configurations of the Jenkins job
- Building a Pipeline job in the Jenkins server

Project workflow: