

Assignment-1 Individual Report

I'm Sobha Tejaswi Garapati, I'm in a team with Luke Ross, Rishabh Singh and Rishab Halthore and we are doing a project based on baseball. In assignment-1 we need to create 2 pipelines up to a data base and data warehouse with different technologies. My part in doing this Assignment-1 is I've done the Technical report for both the pipelines. I compared the performances, I briefed about the pros and cons of each stack and explained about the data sizes in both the pipelines in the technical report for assignment-1 and, I did some of the steps to see the loaded data.

From this Assignment-1 I've gained practical experience like how to set up and run services on a remote server (EC2), how to manage and run scripts for system processes. I've also acquired some knowledge on how to establish secure connections using SSH tunnels and also learned how to interact with a web application and understanding its user interface. I also learnt about how to configure connectors for data ingestion in Airbyte and also learnt about the testing data import using a specific dataset and understanding it's characteristics.

The resultant report of what I have learned is the practical experience in setting up a data pipeline, working with cloud-based VMs, data integration, and various tools and technologies commonly used in the field of data engineering and data science. I gained insights into best practices for secure access to remote servers, data source configuration, and data destination setup. I also have acquired a range of skills related to system administration, secure connections, and data integration using Airbyte.

