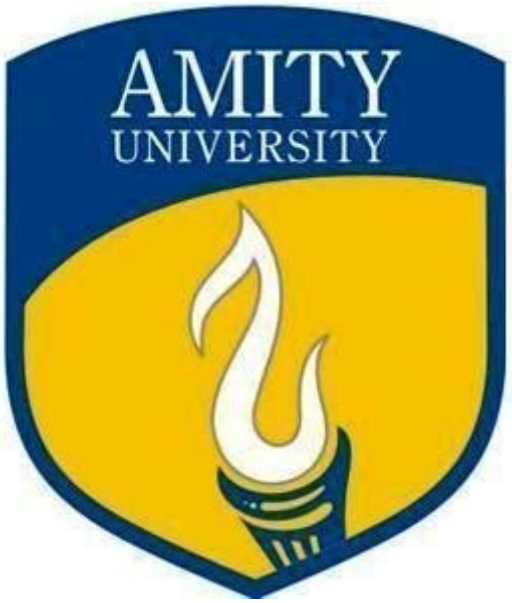
Amity University Haryana ASET 2023-26



Project File: Summer Internship Branch Year:

2025-26 Session: 2023-2026

Submitted by:

Rohit Yadav

Form no. A50504823044

BCA – 3rd Year

Submitted to:

Abstract for CCC

Introduction:

This internship project, conducted at Lirik Infotech as an Data Enginner, focused on developing an end-to-end Data Engineering pipeline using real-world tools and platforms. The program introduced students to essential industry skills such as Python scripting, PostgreSQL database operations, ETL (Extract, Transform, Load) pipelines, API integration, cloud storage handling with AWS S3, and data visualization with Power BI. The aim was to build a layered data architecture (Bronze, Silver, Gold) and load business-ready summaries into databases.

Methodology:

The internship involved hands-on training with tools such as Python (Pandas, Requests, psycopg2), PostgreSQL, AWS S3, and Power BI. The project followed a structured ETL methodology: extracting data from CSV and APIs, transforming and cleaning it, and finally loading it into a PostgreSQL database using Bronze-Silver-Gold layered architecture. Data was also visualized and analyzed in Power BI.

Analysis & Discussion:

The project demonstrated the practical use of data pipelines and cloud technologies in a business environment. It allowed exploration of various data layers and how they contribute to building accurate business summaries. It also helped improve performance by handling data incrementally and applying best practices for storage and transformation.

Learning Outcomes:

The internship highlighted the value of structured data flow using ETL pipelines, cloud integration, SQL database management, and visualization for

actionable insights. It also reinforced best practices in coding, error handling, and modular script design.

Conclusion:

This project underscores the importance of data engineering in modern business intelligence, showcasing how layered data processing and automation can help in delivering clean, reliable, and meaningful data to decision-makers.