

SAI SUMEETH PAGADALA

Data Engineer

Hyderabad | +91 6281751011 | pagadalas1906@gmail.com

SUMMARY

Entry-level Data Engineer with hands-on experience in building batch ETL pipelines and cloud-native data solutions on Google Cloud Platform (GCP). Skilled in Python, SQL, BigQuery, GCS, Dataproc, Dataflow, and Airflow for data ingestion, transformation, and orchestration. Experienced in designing layered data models, performing data quality checks, and preparing analytics-ready datasets. Strong foundation in data engineering concepts with a focus on reliability, scalability, and clean data delivery.

SKILLS

Programming: Python, SQL, PySpark

Data Engineering: ETL Pipelines, Data Modeling, Data Warehousing

Big Data: Apache Spark, Dataproc, Dataflow (Batch)

Orchestration: Apache Airflow

Cloud (GCP): BigQuery, GCS

Databases & Tools: MySQL, Git, VS Code, Jira

Data Quality: Validation, Cleaning, Reconciliation

PROJECTS

Project 1: On-prem to Google Cloud Platform (GCP) Data Migration Project *(Jan 2025 – Jun 2025)*

Tech Stack: MySQL, Google Cloud Storage (GCS), SQL, Apache Airflow

- Set up the project environment and configured Apache Airflow to orchestrate data migration workflows.
- Created and managed Airflow DAGs to automate extraction of data from on-prem MySQL systems.
- Moved data through an SFTP staging layer before loading into Google Cloud Storage (GCS) for secure transfer and validation.
- Performed data validation and reconciliation checks to ensure data completeness and consistency after migration.

Project 2: GCS to BigQuery – Batch Data Ingestion Pipeline *(Jul 2025 – Dec 2025)*

Tech Stack: Google Cloud Storage (GCS), BigQuery, SQL

- Ingested batch data from Google Cloud Storage (GCS) into BigQuery for analytical processing.
 - Designed a three-layer BigQuery structure (Raw, Refined, View) to organize data for reporting and analysis.
 - Applied SQL-based transformations such as trimming, null handling, and basic standardization in the refined layer.
 - Created analytics-ready BigQuery views to support downstream dashboards and reporting use cases.
-

EDUCATION

B.Tech – Computer Science Engineering

Teegala Krishna Reddy Engineering College, Hyderabad (2020–2024) | CGPA: 6.93

CERTIFICATION

Data Engineering – Cirrus Technologies, Hyderabad

Hands-on training in Python for Data Engineering, SQL, ETL development, Data Pipelines, GCS, and BigQuery.