

SANDEEP KUMAR KAPARTHI

GCP - DATA ENGINEER



CONTACT

- +91 7330357797
- sandeepkaparthi27@gmail.com
- Kamareddy, Telangana

EDUCATION

- Osmania University**
Master of Computer Applications (MCA)
2019-2022
- Telangana University**
Bachelor of Science (B.Sc)
2016-2019

SKILLS

- Programming Languages: Python, SQL, Linux, Scala (Basic)
- Big Data & GCP: BigQuery, Dataflow, Pub/Sub, Composer, DataProc
- ETL & Workflow Tools: Airflow, Control-M, DBT
- Streaming: Spark (Basic)
- Visualization: Superset
- Storage/Tools: Hive, Trino, S3
- Databases: Teradata, NoSQL (MongoDB, Bigtable), PostgreSQL
- Version Control & DevOps: Git, Jenkins, GitHub
- Cloud Concepts: Data Lake, Data Warehouse
- Others: Jira, Confluence, ServiceNow

CERTIFICATIONS

- Associate Cloud Engineer (Google) - 2022
- Python Basic Certification (Hackerrank) - 2022
- SQL Advanced Certification (Hackerrank) - 2024
- Problem Solving Basic Certification (Hackerrank) - 2024

AWARDS

- Star of the Month
- Best Team

WEBSITE&LINKS

- [GitHub](#)
- [LinkedIn](#)
- [Portfolio](#)



PROFILE

Proactive Data Engineer with 3+ years of experience in designing scalable data pipelines on GCP using BigQuery, Python, Airflow, and Linux. Hands-on expertise in migration & implementing incremental loads, and optimizing SQL performance. Skilled in time series processing, interest rate projections, and automating ETL workflows with robust validation and error handling. Strong team player with a focus on data quality and business impact.



WORK EXPERIENCE

Data Engineer - GCP 2022 - PRESENT
Tata Consultancy Services Limited

Interest Analytics (Using Time Series Data) DEC 2024 - PRESENT

- Developed end-to-end data pipelines to process 40+ TB of historical data and resigned pipelines for raw ingestion to curated layers using BigQuery, Airflow and Control-M.
- Implemented incremental load strategies and time series partitioning to enhance query performance and pipeline efficiency.
- Automated ETL workflows using Airflow with built-in data validation, logging, and exception handling for fault-tolerant processing.
- Collaborated with the DNA analytics team to generate interest rate projections leveraging curated datasets and time series models.
- Ensured data quality and governance through schema validation, reconciliation checks, and standardized transformation logic across pipeline stages.

Data Migration - Teradata to GCP MAR 2023 - NOV 2024

- Designed and implemented scalable ETL workflows to migrate 3+ TB of structured data from Teradata to BigQuery, ensuring end-to-end data integrity and security.
- Utilized Google Cloud Storage and Compute Engine (VMs) for staging files and running migration scripts, improving operational efficiency.
- Tuned complex SQL queries using partitioning, clustering, and materialized views to optimize BigQuery performance and reduce query latency.
- Automated extraction, transformation, and validation tasks using Python scripts and Linux cron jobs, minimizing manual intervention.
- Developed and orchestrated Control-M jobs to schedule and monitor daily batch loads and incremental data updates.
- Conducted detailed data reconciliation and anomaly detection to handle UTC timestamp shifts and rounding logic mismatches between Teradata and BigQuery, ensuring consistent migration accuracy.