

**TEK011695 (Data Engineer - 5 + Year)**

■ MCA - Computer Science

Highlights

PROFILE SUMMARY

- Have 5y years of hands-on experience in migrating on premise ETLs to Google Cloud Platform (GCP) using cloud native tools such as BIG query, Cloud Data Proc, Google Cloud Storage, Composer. Having 1+ year of hands-on experience as trainee on SQL.
- Hands-on experience in building high-quality, scalable data products and pipelines using Pyspark, Python and Airflow. Experienced in SQL concepts, Python (Pandas) and Pyspark to cope up with the increasing volume of data.
- Have in depth experience in SQL for data manipulation, querying and database management. Hands on Bash scripting experience and building data pipelines on Unix/Linux systems.
- Developed and maintained complex data pipelines using Apache Airflow for workflow orchestration and task scheduling. Collaborated with cross-functional teams to gather requirements and design software solutions. Hands on experience in agile methodology

Skills

Primary Skills

- Apache Airflow
- Python
- AWS
- Google Cloud Dataflow

Secondary Skills

- ETL
- Unix
- GIT
- JQuery

Other Skills

TECHNICAL SKILLS

- Google Cloud Platform (GCP): GCP Cloud Storage, Big Query, Composer, Cloud Dataproc, Cloud SQL, Cloud Functions, Cloud Pub/Sub.
- Schedulers/Workflow: Airflow Composer
- Languages: SQL, Python Scripting
- Data Formats: CSV, JSON, Avro, Parquet
- Databases: MySQL, Cloud SQL, BigQuery
- Operating systems: Windows, Unix
- IDE's: Jupyter Notebook, Notepad++, Colab notebook
- Project Execution Tools: Agile methodology, JIRA, Confluence, GitHub

Projects

Project 1. Wirex - IT Industry

Role: Data Engineer

Technologies: BigQuery, GCS, Airflow/Composer, Cloud Pub/Sub and Dataflow.

Responsibilities:

- Responsible for analyzing requirements and designing the requirements sheet.
- Involving in building data pipeline in airflow in GCP for ETL related jobs using different airflow operators.
- Used cloud shell SDK in GCP to configure the services Data Proc, Storage, BigQuery.
- Coordinated with team and Developed framework to generate daily ad hoc reports and Extracts from enterprise data from BigQuery.
- Developing SQL scripts to Load data into BigQuery.
- Worked related to loading BigQuery data into pandas or Spark data frames for advanced ETL capabilities.
- Created BigQuery authorized views for row level security or exposing the data to other teams.
- Working on Creating DAG's and tasks to deploy and run PySpark jobs in the Airflow scheduling tool.
- Using Source code Management (SCM) as GITHUB.

Project 2. Hanes Australasia - IT Industry

Role: Data Engineer

Technologies: BigQuery, GCS, Airflow/Composer, Cloud Pub/Sub.

Responsibilities:

- Involved in requirements analysis and participated in implementing agile methodology for developing end to end gcp data pipeline.
- Created ETL Pipeline using Spark and Hive for ingest data from multiple sources.
- Designed and executed data schemas, achieving a 10% improvement in data validation.
- Conducted data cleaning processes, enhancing data accuracy.
- Carried out data transformation and cleansing using SQL queries, Python and Pyspark.
- Maintained and debugged pipeline issues, reducing failure rate.

Project 3. Tiki - IT Industry

Role - SQL Developer

Roles and Responsibilities:

- Assisted in managing and analyzing data sets using SQL.
- Contributed to data cleansing and enhanced system performance.
- Learned and assisted in implementing data validation checks.
- Worked under senior developers to optimize complex SQL queries resulting in the increase in performance.

- Assisted in implementing a row-level security algorithm on confidential data.
- Developed database schemas, tables.
- Involved in Designing, developing and maintenance of relational databases.
- Implemented SQL joins and window functions while writing the queries.