

DIKSHA TOKKAR

Pune India | +91 9549477846 | tokkardiksha@gmail.com | [LinkedIn](#)

OBJECTIVE

Data engineer with 3+ years of experience in Banking and Pharmaceutical domain. Built ETL pipelines using Google Cloud Services. Proficient in writing optimized SQL queries for creating complex tables with performance metrics like **partitioning** and **clustering**. I am seeking an opportunity to leverage my expertise in cloud services to support building scalable data solutions.

SKILLS

Key Skills: Python, MySQL, PySpark, Apache Hadoop, Apache Hive, Apache Airflow, Shell Scripting
Cloud Technologies: GCP (BigQuery, Composer, Dataproc, Cloud Functions, Cloud Storage)
Project Management Tools: Gitlab, IntelliJ, DBeaver, Jira, Confluence

EXPERIENCE

GCP Data Engineer **June 2023 - Present**
Accenture **Pune, India**

- **Led** the migration of 100+ TB of data from on-premises Oracle Database to Google Cloud Platform (GCP) BigQuery for a major Canadian bank.
- **Orchestrated** data movement across three zones: Landing, Processing, and Curated, improving data processing efficiency by 30% and reducing time-to-insight for business teams.
- **Managed** the migration of 300+ database tables to the cloud, leveraging batch, delta, and file-based loading strategies, reducing migration time by 30%.
- **Automated** ETL workflows using dynamic Directed Acyclic Graphs (DAGs) in Apache Airflow, decreasing end-to-end data processing time by 25% and enhancing operational efficiency.
- **Engineered** Spark jobs for ETL processes using Dataproc to transfer data from Oracle databases to BigQuery, optimizing data load performance.
- **Automated** file arrival monitoring via DAGs, reducing manual intervention by 90% and ensuring timely data availability for downstream processes.
- Successfully validated end-to-end pipeline functionality through User Acceptance Testing (UAT), ensuring accuracy and reliability before production deployment.

Data Engineer Associate **July 2021 - May 2023**
Accenture **Pune, India**

- **Worked** on migration of data from on-premises Hadoop framework to GCP.
- **Optimized and refactored** the Hive scripts to BQ compatible SQL scripts, Shell scripts to equivalent Python scripts.
- **Transferred** file from GCS to shared drive using SFTP operator and worked on email clients like Sendgrid API, SMTP.
- **Utilized** PySpark and Spark SQL to streamline data pipelines, enhancing performance by 30%.
- **Used** Apache Airflow for orchestrating ETL processes using Airflow operators, BigQuery and Storage client libraries, and task groups.

EDUCATION

Bachelor of Technology (Information Technology), **2017 - 2021**
Medicaps University **(CGPA 8.31)**

CERTIFICATION

- GCP Associate Cloud Engineer
- GCP Professional Data Engineer
- 5 star on Hackerrank for SQL

INTERESTS

- Cooking
- Travelling