

MAHESH PAWAR

Pune, India

+91-9168712328

pawarmahesh1699@gmail.com

LinkedIn

HackeRrank

Professional Summary

GCP Data Engineer with more than 3 years of experience in designing and optimizing large-scale data pipelines. Recognized as a **Top Data Engineer Voice** on LinkedIn and a **four-time award-winner** for impactful engineering initiatives. Expertise in **GCP, BigQuery, PySpark, and Dataproc**, with a focus on automation, cloud migration, and data validation. Skilled at delivering scalable, high-performance solutions that enhance system efficiency and reduce costs.

WORK EXPERIENCE

Datametica Solutions Private Limited

03 2024 – Present

Engineer I

Pune, India

- Project: **Data warehouse migration of Teradata to GCP for the healthcare domain (POC project).**
- Responsibilities:
- **Led the Proof of Concept (POC)** for the migration project, collaborating with the **DevOps team** to establish essential **GCP project services**.
- **Converted and productionized DataStage ETL pipelines into GCP Dataform**, significantly improving data processing capabilities.
- **Established a Delta Lake on Google Cloud Storage**, migrating existing ETL processes to **Spark** using **Apache Iceberg** for enhanced performance and scalability.
- **Modernized legacy Perl scripts on GCP VM**, developing and testing robust **ETL pipelines** using **Apache Airflow** to ensure seamless data workflows.

Datametica Solutions Private Limited

03 2023 – 02 2024

Engineer I

Pune, India

- Project: **Cloud modernization and migration of a data warehouse for a telecommunications client.**
- Responsibilities:
- **Transformed DB2 and Oracle PL/SQL scripts into PySpark SQL**, boosting processing efficiency.
- **Built end-to-end data ingestion pipelines using Dataproc and Airflow** for job scheduling and orchestration.
- **Developed a Python utility** for auto-converting DB2/Oracle schemas to PySpark structTypes, streamlining data processing.
- **Integrated a data generator utility** for schema and dummy data generation, enhancing testing processes.
- **Designed reusable PySpark scripts** for common tasks, including reading various data formats, logging, and parsing schemas.
- **Created metadata-driven Python scripts** for efficient **Slowly Changing Dimension (SCD) Type 2** operations.

Datametica Solutions Private Limited

02 2022 – 03 2023

Associate Engineer I

Pune, India

- Project: **Migrating On-premise Data-warehouses(Netezza, Teradata) to Google Cloud Data-warehouse(BigQuery) For Merchandise And Retail Domain Client**
- Responsibilities: **Converting DataStage/Informatica ETL Jobs into BQ Stored Procedures and building end-to-end data processing pipeline using Airflow DAG.**
- Taken complete ownership of scripts jobs extensively used in Extract and Transfer Data from various sources to **BigQuery Data-Warehouse**. Sources are **Google Analytics API, Flat Files, BTEQ, FLoad, MLoad, Json Files, SFTP Servers..**
- Complete Ownership of Developing Batch and Streaming Data Ingestion pipeline using **Airflow-Dag, Dataflow, python and shell scripts.** And given Handover and support to the team.
- Optimized SQL queries and Airflow DAG with parallel execution which **lowered execution time by 35 percent** for Data Processing.

- Hands on in UC4,Control-M(**Scheduling and Orchestration**).
- Knowledge of **CDC,CI/CD** and Version Controlling (**Git,Bit-Bucket**).
- **Expertise in Historical Data Migration.**

Datametica Solutions Private Limited

06 2021 – 01 2022

Associate Analyst

Pune, India

- Requirement : **Migrating Historical Data From Netezza to Google Cloud BigQuery And Automating Manual Work.**
- Action : **Migrated 500 TB of data by developing fully automated Data Transfer framework by using Python and Shell Scripts.**
- Extract and Transfer data from On-premises to Cloud Storage (GCS Bucket, BigQuery).
- By Leading a team of 5 member migrated 10000 tables.
- Automated overall Historical Data Migration activity which has reduced manual intervention.
- Some of the Innovations I have did -
 - * Reconciliation of count and aggregation of numeric columns between NZ and BQ.
 - * Creating intermediate worktables to enhance historical data migration pipeline processing.
 - * Automating reporting on extracted table statuses via email, minimizing manual checks after script executions.
 - * Finding The historical data gaps in tables and generating config for missed data in BQ table, in the outcome time and 2 resources manual intervention reduced.
- Outcome : **Pipeline process got improved and awarded by Rising Star Award for my commitment and contribution to the Asset Building.**

TECHNICAL SKILLS

Languages: Python, SQL, Shell-Scripting, Java, C, C++, JavaScript, XML, JSON

Big Data: Airflow, Spark(PySpark), Hadoop, Map Reduce, Apache Beam, Apache Iceberg, Kafka

Google Cloud Services : BigQuery, GCS Bucket, Cloud Composer, Cloud Function, Compute Engine, Cloud-Shell, Dataflow, Dataproc, Dataform, PubSub.

Tools: DataStage, Informatica, Teradata, Netezza, DBaver, VMware, Git, Bit-Bucket, Putty, Control-M, UC4

EDUCATION

Solapur University

06 2017 – 03 2020

Degree Name - BCA(Computer Science) Percentage - 72.47%

Pandharpur, India

Pune University

02 2015 – 02 2017

Exam Name - HSC - Percentage - 70.00%

Venunagar, India

ACADEMIC PROJECTS

Gas Agency management system | C Sharp ,MySQL

01 2020

- This is my academic project used to manage Gas Agency.
- It allows users to Store,Display, Delete, Update Customer, Vendor and Employee record.
- Automatic bill report generate as a PDF based on purchase.
- Automatic stock maintain based on purchase and sell.

Quiz Website | HTML,PHP

04 2019

- This is mini project used for college level quiz competition.
- It will save all data to database which will help to generate scorecard.

CERTIFICATIONS

- Professional Data Engineer - Google
- Associate Cloud Engineer - Google
- Apache Airflow Fundamentals - Astronomer
- SQL Certificate - HackeRank
- Google BigQuery and PostgreSQL - Udemy
- Python Programming - Udemy/HackeRank

ACHIEVEMENT/EXTRACURRICULAR

- Rising Star Award -Datametica
- LinkdIn MySQL Skill Badge - LinkdIn
- Two Times SPOT Award - Datametica
- Build and Optimize DataWarehouses with BigQuery - Qwiklabs