

**QAZI HAFEEZ**

**GCP Data Engineer
Ph: +1 (510)-854-1314**

**Summary:**

* 9+ years of experience in IT, which includes experience in Cloud Data Engineering, Hadoop, Big Data, Data Warehousing, SQL related technologies in Retail, Manufacturing, Financial and Communication sectors.
* Certified by Google like **Professional Data Engineer** since July of 2021.
* More than 3 years working with cloud environments, like Google Cloud: **Big Query, Cloud SQL, Cloud Storage, Cloud SDK, Compute Engine, App Engine, Cloud APIs, and other tools like Dataflow, Data proc, Pub/Sub, IAM Security, Cloud Function, Cloud Composer.**
* Worked on structured and semi structured data storage formats such as Parquet, ORC, CSV, JSON.
* Created DataMart’s using Pyspark, Big query job config and **Airflow** big query operators as per the reporting need.
* Created **Airflow dags** to schedule batch jobs, trigger **Big Query** or Pyspark operator to refresh the DataMart’s. Utilized Airflow variable to control the Airflow job tasks and the logic.
* Expert in working with cloud **PUB/SUB** to replicate data real-time from source system to **GCP Big Query.**
* Good knowledge on **GCP service accounts**, billing projects, **authorized views**, datasets, GCS buckets and **gsutil commands**.
* Experience with Unix/Linux systems with scripting experience and building **data pipelines**.
* Played a key role in migrating Tidal on GCP and defined different read/write strategies.
* Writing Cloud Formation Templates in **JSON** for Network and Content Delivery of the Google Cloud Environment.
* Expertise in managing full life cycle of Data Science project includes transforming business requirements into Data Collection, Data Cleaning, Data Preparation, Data Validation, Data Mining, and Data Visualization from structured and unstructured Data Sources.
* Strong SQL development skills including writing **Stored Procedures**, Triggers, Views, and User Defined functions.
* Hands on experience with different programming languages such as Python, SAS.
* Developed complex SQL queries and performance tuning of SQL queries.
* Hands on experience in setting up workflow using **Apache Airflow** engine for managing and scheduling Hadoop jobs.
* Experienced in working with Source Code management tools in **GIT**, **Bit Bucket** and **Source Tree.**

**Technical Skills:**

|  |  |
| --- | --- |
|  **GCP** | Google Cloud: Big Query, Cloud Run, Cloud Storage, Cloud SDK, Compute Engine, Cloud Functions, App Engine, Cloud APIs, and other tools like Dataflow, Dataproc, Data prep, Data Studio, Pub Sub, Spanner, IAM Security, Service Data Transfer, VPC Configuration. |
| **Programming languages** | Java, C#, PYTHON. |
| **Application/web servers** | Tomcat, IIS, Jenkins.  |
|  **Sql** | SQL, PL/SQL DB2, PL/SQL Oracle, Transact SQL. Views, Materialized Views, Store procedures, Windows queries, Triggers. |
|  **Data Modeling** | Star Schema Modelling, Snowflake Modeling |
|  **Methodologies** | Agile, Scrum, Waterfall. |
|  **Database servers** | Oracle 10g/11G/12c, SQL Server 2000/2005/2008/2012/2014/2017, MySQL, PostgreSQL, Big Query, Cloud SQL |
|  **Version control** | CVS, SVN, GIT |
|  **Platforms** | Windows 2000/98/95/NT4.0, UNIX, Windows 7,8,10, MacOS, Linux Debian/RedHat/Fedora/AIX/Ubuntu, Windows Server 2003,2008,2012,2016,2019. |

**Certification:**

 **Certified** in Google Cloud Platform **Professional Data Engineer**.

**Professional Experience:**

**Client:** Virgin Voyages, FL **Oct 23 – Till Date
Rol**e: GCPData Engineer

**Responsibilities:**

* Experience in building the multiple data pipelines using cloud services.
* Hands on experience in Google Cloud as Cloud Storage, Dataflow, Cloud Composer, Bigquery, Cloud Functions, Cloud Pub/sub and Dataproc.
* Experience in creating the GCP Projects and setting up an environment as per the requirements.
* Writing Python scripts to load the data from SFTP Server to BigQuery.
* Experience in creating pyspark scripts to load the data from different sources to BigQuery using Cloud Run and Cloud Function services.
* Experience in working with Airtable and loading the Airtable bases to BigQuery using pyspark scripts.
* Involving with different teams and back tracking the flows and experience in solving the critical issues.
* Identified and documented strategies, tools and phases in migration to Google Cloud Platform.
* Experience in loading the data from GCS to S3 bucket using private keys.
* Responsible for Google Production Support for environments and application related issues.
* Involving with business teams and different teams to solve the critical issues. Involving with business teams and different teams to solve the critical issues.

**Environment:** GCP, Cloud SQL, Big Query, Cloud DataProc, GCS, Cloud SQL, Cloud Composer, Hadoop, Hive, Teradata, SAS, Teradata, Spark, Python, SQL Server, Service Now, Confluence, Airtable, IBM Console.

**Client:** Verizon, Dallas, TX **Apr 22 – Oct 23
Rol**e: GCPData Engineer

**Responsibilities:**

* Experience in building the multiple data pipelines using cloud services.
* Hands on experience in Google Cloud as Cloud Storage, Dataflow, Cloud Composer, Bigquery, Cloud Functions, Cloud Pub/sub and Dataproc.
* Experience in IBM Console for monitoring of streaming jobs.
* Experience in testing the data through streaming jobs for Events and Outages.
* Open SSH tunnel for running the SPARK\_SQL Jobs.
* Writing Python scripts to load the data from Bigquery to Bigquery using Dataflow and Composer.
* Experience in data validations and analysis for Prod defects.
* Running the Cron Jobs using the Omega Data to GCP and checking the logs in Omega.
* Experienced in GCP Jobs and Egress jobs testing once the migration is done.
* Experience in writing and creating the Hive tables in Omega and data validation.
* Working experience with Support team and took the responsibility for the issues in production.
* Experience in solving the priority issues and involving in SOC calls while there is any production issues.
* Handling all priority incidents created by the end users and provide the solution on time via Service Now.
* Experience in creating Teradata scripts, pyspark scripts to load the data from Teradata tables to Hadoop.
* Responsible for L2 support for environment and application related issues.
* Handling Change Requests and Service Requests.
* Troubleshooting production issues under client defined SLA’s.
* Responsible for Google Production Support for environments and application related issues.
* Service Now applications implemented: Incidents, Change Requests, Service Requests, Configuration, Dashboards.
* Involving with business teams and different teams to solve the critical issues.

**Environment:** GCP, Cloud SQL, Big Query, Cloud DataProc, GCS, Cloud SQL, Cloud Composer, Hadoop, Hive, Teradata, SAS, Teradata, Spark, Python, SQL Server, Service Now, Confluence, IBM Console.

**Client:** ULTA BEAUTY, Bolingbrook, IL **Feb 21 – Mar 22**

**Rol**e: GCPData Engineer

**Responsibilities:**

* Experience in building multiple Data pipelines, end to end ETL and ELT process for Data ingestion and transformation in GCP and coordinate task among the team.
* Design and implement various layer of Data Lake, Design star schema in Big Query.
* Using g-cloud function with Python to load data into Big query for on arrival csv files in GCS bucket.
* Worked on GCP POC to migrate data and applications from On-Prem to Google Cloud.
* Exposure on IAM roles in GCP.
* Good knowledge on GCP service accounts, billing projects, authorized views, datasets, GCS buckets and gsutil commands.
* Developed and demonstrated the POC, to migrate on-prem workload to Google Cloud Platform using GCS, Big Query, Cloud SQL and Cloud DataProc.
* Follow Agile methodology process.
* Designed Pipelines with Apache Beam, Dataflow and orchestrated jobs into GCP.
* Process and load bound and unbound Data from Google Pub/Sub topic to Big query using Cloud Dataflow with Python.
* Experienced on ETL processes and used Python.
* Speech to text specialist for Speech Suites solution.
* Create the architecture to transactional data in cloud sql (MySQL).
* Involved in Development and Created SQL stored procedures, functions.
* Created Airflow Scheduling scripts in Python and DAG’s for SFTP Process.
* Create models with the data to solve problems.

**Environment:** GCP, Big Query, Cloud SQL, Cloud Storage, Airflow, DataStudio, MySQL, Natural Languages, Windows server, python, IAM Security, Service Data Transfer, VPC Configuration, Data Catalog.

**Client:** Nemours Health, Wilmington, DE  **Feb 20 – Jan 21**

**Role:** Data Engineer

**Responsibilities:**

* Used Deployment Manager of GCP to launch instances and copy configurations into additional regions to tackle issues related to region outage.
* Configured the Kubernetes cluster on GCP to establish communication between pods and MySQL instance.
* Participated in Querying RDBMS such as Oracle, MYSQL and SQL Server by using SQL for data integrity.
* Setup Alerting and monitoring using stack driver in GCP.
* Deployed and monitored scalable infrastructure on Google cloud & configuration management using Docker and Google Kubernetes Engine (GKE).
* Created and deployed Kubernetes pod definitions, tags, labels, multi-pod container replication. Managed multiple Kubernetes pod containers scaling, and auto-scaling.
* Deployed pods using Replication Controllers by interacting with Kubernetes API server defining through declarative YAML files.
* Working with Solution architects to set up a dedicated GCP cloud environment for our project.
* Leveraged cloud SDK and Google cloud CLI to package and deploy containerized web applications on to the Google cloud platform and using Google Kubernetes Engine (GKE) to manage clusters and implemented auto scaling conditions to accommodate traffic variance.
* ETL Pipelines that brings and transforms the huge volumes of data from different source systems Using Python Data Ingestion Pipelines.
* Developed ELT jobs using Apache beam to load data into Big Query tables.
* Designed Pipelines with Apache Beam, Dataflow and orchestrated jobs into GCP.
* Process and load bound and unbound Data from Google pub/sub topic to Big query using cloud dataflow with python.
* Devised simple and complex SQL scripts to check and validate Dataflow in various applications.
* Exporting the analyzed and processed data to RDBMS using Sqoop for visualization and for generation of reports for the BI team.
* Used ETL tool to do transformations, event joins, filters and some pre-aggregations before storing the data onto HDFS.
* Configuring and Setting up Airflow dags as per the requirement to run our pyspark commands in airflow parallel and sequential.
* Creating Job flow using Airflow in Python language and automating the jobs. Airflow will have separate stack for deploying dags on and will run jobs on GCP.
* Implemented GCP Code Pipeline and created Cloud formation JSON templates in Terraform for infrastructure as code.
* Leveraged cloud and GPU computing technologies for automated machine learning and analytics pipelines, such as GCP.
* Understand the data model document and analyze to identify the source systems and transformations required to ingest the data into Data.
* Prepared different documents which includes Data flow diagram, mapping doc, testing scenarios.

**Client:** Rcam Soft Tech Pvt Ltd, Hyderabad, India **Oct 17 – Jun 19**

**Role:** ETL Developer

**Responsibilities:**

* Design and Implement ETL for data load from Source to target databases and for Fact and Slowly Changing Dimensions (SCD) Type1, Type 2, Type 3 to capture the changes.
* Involved in writing SQL Queries and used Joins to access data from Oracle, and MySQL.
* Designing, developing, and deploying end-to-end Data Integration solution.
* Implemented custom error handling in Talend jobs and worked on different methods of logging.
* Developed mappings to load Fact and Dimension tables, SCD Type 1 and SCD Type 2 dimensions and Incremental loading and unit tested the mappings.
* Created context variables and groups to run Talend jobs against different environments.
* Created Data model, Data entities and view for the Master Data Management.
* Involved in creating roles and access control.
* Used the triggers to launch the process with given set of conditions.
* Worked on creating Data model, Campaigns for the Talend Data stewardship.
* Created Data entities in the Data model and defined roles at the entity level.

 **Environment:** Talend Platform for Data management 5.6, UNIX Scripting, Oracle 10g, SQL Server.

**Client:** (DRDL) – Hyderabad, India **Aug 15 – Sept 17 Role:** Data Acquisition Engineer

**Responsibilities:**

* Improve operations & impact growth in Research and Development Section for various projects like AGNI, PRITHVI, TRISHUL, NAG, DHANUSH, BRAHMOS, LRSAM, AKASH, ASTRA, PROGRAMME AD and PROGRAMME B-05.
* Worked as an Engineer with functional expertise in Real Time Data Acquisition and Data Analysis in Project Executions.
* Responsible for providing Ground Instrumentation Support during static testing of Missiles.
* Responsible for Programming in LabVIEW for Real time display.
* Performing Data Acquisition using WINDAQ software.
* Generated waveform reports for static tests of the Rocket Motors.
* Preparation of key project plans for new ventures including new channels.
* Reviewing & performing analysis of data using DADiSP to related projects.
* Analyzing the signal conditioners to generate dual outputs (Filtered and Unfiltered).
* Testing of fiber optic cables involving loss measurement and locating breaks in fiber.
* Trouble shooting of fiber optic cables involving connectorization of cables and splicing.
* Establishing Ethernet communication link over optical fibers.

**Education:**

* **Master’s Degree in Computer Science** / Governor’s State University, IL 2021.
* **Master’s Degree in Business Administration** /Karnataka State University, India 2015.
* **Bachelor’s Degree in Electronics & Instrumentation** / Osmania University, India 2012.