ANURAG SRIVASTAVA

Data Engineer | GCP | anuanurag2410@gmail.com | +91-7XXXXXXX | LinkedIn | Github

Objective

Around 3 years of Experience in Data Engineering Field, worked on GCP Bigquery, Teradata and was part of Core Data Team responsible for Data Migration Activities. Looking for a long-term association with a growth-oriented organization with some added responsibilities.

Professional Work Experience

Working as Data Engineer with GCP BigQuery, GCS Buckets and Spark to Transform and Load the Data From On-Premise to Cloud Platform(GCP).

Professional Summary

- 3 Years of hands-on-experience using Google Cloud Platform(GCP) services and knowledge of Data Warehousing Concepts.
- Knowledge of Different Database Sources- Teradata, MYSQL and Google BigQuery
- Good Communication Skills and can adjust quickly in a team environment.
- While Contributing to the project worked over weekends and different shifts to meet the business requirements before SLA.
- Followed Agile Methodology during the project and JIRA to Track all the Issues and Bugs.
- Expertise in finding the issues and Bugs in the code and discrepancy in the Data after loading.
- In absence of Team Lead acted as Shift Lead to guide the team of 6 people to complete the Daily Business Requirements.

Skill Set

- **Primary Skill -** SQL(DDL, DML), GCP BigQuery,Python(Pandas, Numpy, os module), Pyspark(Basics), Data Analysis
- **Secondary Skills-** Linux, Bitbucket, JIRA, Bamboo Builds
- **Database Used-** Teradata(IDW),GCP BigQuery
- **Tools Used-** Putty,Teradata v15.0,Control-M, JIRA,Bitbucket, Win SCP

Project Experience

Cloud Data Warehouse Migration *Feb 2021-Present*

Client: *Charles Schwab*
Domain: *Banking and Multinational Financial Services*
Role: *Data Migration Developer, Data Engineer*

Project Description:

The Project on high-level scope was to migrate Integrated Data Warehouse Architecture(IDW) to Cloud Data Warehouse(CDW), which involves assessing and developing a data migration plan for ETL Extraction and transforming the data from a total of 1**26 data sources.** We dealt with migrating over **500 million** customers' data records from Teradata to Google Cloud that was also the first of its kind worldwide in terms of volume of data was **975 PetaBytes.**

Roles and Responsibilities:

- I worked for about 3 Years in the CDW Project where I had an experience working with 2 Different Teams (DDL Team and Data Migration Team).

- Part of Team where we designed Data Migration Utility that is an ETL Tool to Migrate data from Teradata to GCP, that gave End-to-End solution for loading the data from teradata to GCP by extracting data using TDCH/SQOOP from teradata then transforming the data using PySpark and loading it to GCP BigQuery.
    - *Technologies Used: Python,Teradata, GCP,Unix, Pyspark*
- Worked with cross functional teams to ensure smooth loading of data using ETLs and helped ETL teams in case of data discrepancy of the data.
- Worked on Conversion of **40,000 Objects**(Tables,Views and Procedures) and worked with cross functional teams to get DDL compatible as per data accepted in GCP.
- Developed Python code to find dependency order of the views and their dependent objects to deploy them smoothly through bamboo build.
- Also developed Python code to minimize the manual effort like changing the file name of the DDL objects as create table statements,Validating the column names as an exact match in Teradata and BigQuery, moving and bifurcating the objects DDLs to Bitbucket.
- Loaded **100PBs** of data from TD and BQ which includes history data loading and file migration.
- Worked on code fixed for incremental data load tables for **100TB** data.
- Worked on monitoring the daily data load jobs scheduled for control-m and getting instant fix for the job running in Pre-Prod and Prod Environments in BQ.
- Worked on Change Requests for the DDL deployment to BigQuery production and data loading in BQ Prod with help of external team.
- Found an easy approach to load huge volumes of data by using stage approaches and parallel loading of data to save time.

Personal Project Experience

1- Ola Data Engineering Project | Github

Project Description:

On High Level Scope the project was to take a raw taxi Dataset from Kaggle and Used pandas library to transform the code, also made **Data Model** using **LucidChart** then used **Mage(Data Pipeline) tool** for the Data Exporting to Google Bigquery, then made a Analytical Table in GCP BQ to make Dashboard in **Looker Studio.**

Technologies Used:*Python Pandas, LucidChart,Mage Data Pipeline Tool, GCP Bigquery, GCP GCS, LookerStudio*

Achievements

- Got Personal recognition from Vice president for my extraordinary contribution with a lump sum amount (Specially Rewarded) ***December 2022***
- Received Certificate of Appreciation: KUDOS AWARD from Project Manager for my phenomenal contribution and support to achieve our milestone *March 2023*
- Received Certificate of Appreciation: KUDOS AWARD from Project Manager for my urge to learn and apply my learning in project *September 2023*
- Got 4/5 and 5/5 yearly performance ratings *2022,2023*

Academic

- Bachelor of Technology in Computer Science Engineering *2016-2020*
  Maharaja Ranjit Singh Punjab Technical University ***CGPA 7.2/10***

- HSC(ICSE Board) with 67% at Little Flower School, Gorakhpur Uttar Pradesh *2013*
- SSC(CBSE Board) with 72% at RAMPUS,Gorakhpur Uttar Pradesh *2015*