**Sethuvardhan Reddy Undela**

**860-603-0189  
teja@itarks.com**

Professional Summary

* Over 9 years of experience in Data Warehousing, ETL, Design and Development, Implementation of Business Intelligence solutions and Cloud environments/platforms like Google Cloud platform and Azure.
* Designed a complete architecture for data migration from Oracle (on-prem) to google cloud platform (Google Big query) and Azure Synapse.
* Developed both ELT and ETL solutions with Composer Airflow and Cloud dataflow(Apache Bean), ADF and python.
* Have Very good experience in google big query. Implemented partitioning and clustering which helped us to reduce 80% of query cost.
* Developed a python function to create high-end virtual machines on fly and delete the instance immediately after the work done automatically in GCP computer engine environment which helped us to reduce the billing cost for computer engine services.
* Developed scripts to integrate multiple software components.
* Developed an alerting system for Failure, Success, Delay of the ETL jobs which helped us to avoid continuous monitoring of the ETL Jobs and saved nearly 4 hours per day.
* Involved in Data cleansing and data remediation for historical data in production environment and lower environments.
* Good in understanding the business processes/requirements and translating them into technical requirement.
* Experience in PostgreSQL, Cloud SQL, oracle, and Big Query development and involved in analysis and verification of data loaded.
* Good communication and interpersonal skills, quick learner and good performer in both team and individual job environment.

**Technical Expertise:**

**Programming Languages** :Advanced SQL, Python, shell scripting and PySpark

**Databases** : Google Big query, Cloud SQL, Postgres and Snowflake

**Operating systems :** Windows and Linux

**Cloud Technologies (GCP) :** Big-query, composer, cloud SQL, Compute engine and GCS.

**Cloud Technologies (Azure)** **:** ADF, Azure Devops, Blob, Databricks and Azure SQL.

**Project management tools :** Bit bucket, Git, JIRA, Confluence and CA Rally.

Professional Project # 1

**Name** : Diverts and Triggers.

**Role** :GCP Data engineer

**Client** : **CVS Health.**

**Portfolio** : Health Care Firm in USA.

**Duration** : July’24 to till date.

**Technologies** :Cloud SQL | Postgres | Advanced SQL | Apache Airflow | Data processing | ADF | Big query | JIRA | Data cleansing | ETL/ELT | Google cloud platform | Cloud Dataflow (Apache Beam) | Python (pandas) | Google cloud Storage | Google compute engine | Composer (Airflow)

**Description**

Engineered a solution to Diverts and Triggers process which facilitates reimbursement for Med-D plan members at the benefit level, aligning payments with what is covered by the plan and CMS (Centers for Medicare & Medicaid Services) guidelines.

**Responsibilities**

* Developing robust data pipelines using BigQuery to enable seamless extraction, transformation, and loading (ETL) processes.
* Created detailed mapping documents for seamless Cloud SQL claim data migration.
* Implemented ETL processes on the Google Cloud Platform, leveraging native tools like Big-Query for effective data management.
* Developed Apache Beam transformations and I/O connectors for data ingestion and processing.
* Optimized Big-Query processes with techniques like partitioning, clustering, time travel, and cache management to ensure cost-efficiency.
* Developed and scheduled data pipelines using Cloud Composer (Airflow) to ingest RX claim data from GCS to Big-Query.
* Implemented robust Airflow pipelines using configuration variables and XCom variables to parameterize and streamline the Diverts and Triggers process in GCP Big-Query.
* Developed Python scripts using Pandas for data manipulation, ensuring efficient data transformation and analysis.
* Enhanced the reimbursement process for Med-D plan members by transitioning from Cloud SQL to GCP Big-Query, streamlining operations and ensuring compliance with CMS standards.
* Conducted monthly maintenance activities, including housekeeping on GCP Big-Query tables, GCS Buckets, and VM refresh operations.

Professional Project # 2

**Name** : Deductions, Fees, and Fines.

**Role** :Consultant

**Client** : **Kraft Heinz.**

**Portfolio** : CPG Firm in USA.

**Duration** : Dec’23 to till date.

**Technologies** :Azure SQL | Postgres | Advanced SQL | Apache Airflow | Data processing | ADF | Snowflake | Azure Devops | Data cleansing | ETL/ELT | Google cloud platform | Snowflake | Synapse | ADF Managed Apache Airflow | Python (pandas) | Google cloud Storage | Google compute engine.

**Description**

Engineered a solution to identify and mitigate financial losses within a CPG firm, focusing on fees, fines, and customer claims. Developed data pipelines in Snowflake and Big query for seamless ETL processes and employed advanced data modeling techniques. Constructed analytical frameworks to pinpoint areas of financial leakage and propose recovery strategies. This solution facilitated the recoupment of historical losses and the implementation of proactive risk mitigation measures.

**Responsibilities**

* Designing and engineering a comprehensive solution to detect and address financial losses within a CPG firm, with a focus on fees, fines, and customer claims.
* Developing robust data pipelines using BigQuery to enable seamless extraction, transformation, and loading (ETL) processes.
* Developed and managed complex data pipelines using Python, Apache Beam (Cloud Dataflow), and Apache Airflow (Cloud Composer) for efficient data processing and workflow automation.
* Built scalable ETL pipelines for real-time and batch processing.
* Used Dataflow templates for reusable and scalable pipeline deployments
* Leveraging advanced data modeling techniques to structure and optimize data within Snowflake and BigQuery, ensuring accuracy and efficiency in financial analysis.
* Creating analytical frameworks to identify areas of financial leakage and propose proactive recovery strategies, leveraging insights from data analysis.
* Collaborating closely with stakeholders to understand business requirements and translate them into technical solutions.
* Implementing measures to facilitate the recoupment of historical losses and the implementation of proactive risk mitigation measures.
* Continuously monitoring and refining the solution to adapt to evolving business needs and changing market conditions.

Professional Project # 3

**Name** : Universal Profile.

**Role** :Consultant

**Client** : **Blue Nile.**

**Portfolio** : Jewelry retailer.

**Duration** : Nov`19 to Dec’23.

**Technologies** :Oracle | SQL | Advanced SQL | Apache Airflow | Data processing | ADF | Big-Query | Azure Devops | Data cleansing | ETL/ELT | Google cloud platform | Google Big query | Synapse | Google cloud Composer | Python (pandas) | Google cloud Storage | Google compute engine.

**Description**

Provide a single 360 view of customer and Activate customer-centric marketing through Purchase Journeys and rapid experimentation by Pulling data from on-prem Oracle database and prepared Airflow pipelines to push all the data into Google Big-Query or Azure Synapse after and before data transformations. And building a universal profile (Unique ID across the globe) for customers using the PII information collected from different kind of sources like online, fresh desk and historical data.

**Responsibilities**

* Designed a complete architecture for data migration from Oracle (on-prem) to google cloud platform (Google Big query).
* Understanding the business processes/requirements and translating them into technical requirement.
* Developing ADF Pipelines that will migrate the data from GCS to blob to Azure SQL.
* Developing Airflow ETL pipelines to transform the data from oracle to Big query.
* Used Cloud SQL for managed relational databases with automatic backups and scaling.
* Involved in creating the unique ID for customers using the advanced SQL in Big query.
* Creating an analytical data set where data science team proactively use it for recommendation systems.
* Developing advanced SQL scripts to clean the phone number and email address and other PIII information to increase the matching probability while creating a universal profile.
* Utilized Firestore for flexible, scalable data storage in real-time applications.
* Involved in Sprint planning, daily stand-ups & Review meetings.
* Coordinating Releases & Hotfixes
* Coordinating weekly status call with Client & sending MOM
* Organizing the QC plan and test cases to make sure that every item or task is passing the expected outputs.
* Contributing to the overall data strategy and architecture to support long-term business objectives and growth initiatives.

Professional Project # 4

**Name** : Trade Experience.

**Role** :AssociateConsultant

**Portfolio** : **Global Health care.**

**Duration** :Dec`18 to Nov`19

**Technologies** :Informatica Power center (ETL) | PostgreSQL (Green plum) | Teradata (Database) | Chronicle(scheduler) | Excel as analytical tool | Python(pandas) | Azure |

**Description:**

The Trade in Experience project is for tracking the healthcare trade in business. Basically, whenever a customer wants to buy back Install base products from customer, then a Quote will be created, and asset value will be provided. And these products are segregated based on their usage and kind of modality. Categories like GSC and SP will have more value than others. This trade in information we will be receiving as files and the same will be processed with Informatica. Target database is PostgreSQL (Green plum).

# **Responsibilities**

* Worked as a DB and ETL developer, building Data Base objects, pipelines, performing data analysis, setting up QA unit/SIT test cases, code & data migration, providing documentation.
* Created Mappings to bring the initial (Raw) data from Teradata to PostgreSQL (green plum) database by using different kind of Transformations in informatica mapping designer.
* Created SQL scripts which will convert Raw data (Excel) to informative (Normalized) data which will reside in relational tables.
* Created triggers, Functions in PostgreSQL which will maintain the AUDIT tables automatically.
* Automated the process of validating the Data base objects like tables, views, columns, datatypes from one environment (stage server) to another environment (production server).
* Build reports with Excel which will help to understand the business Flow.
* Involved in query optimization and reduced query running time and increased query performance.
* Providing guidance and support to the broader team on data engineering best practices and methodologies.
* Documenting processes, workflows, and solutions to ensure knowledge sharing and maintain system integrity.

Professional Project # 5

**Name** : **Colombia broadcasting system**

**Role** : Sr. Associate.

**Portfolio**  : Telecom and media

**Duration** : May`16 to Dec`18

**Technologies :** Informatica Power center (ETL) | Matillion (ETL/ELT) | Oracle 11g | Teradata (Database) | Google Big query as (Database) | Excel as analytical tool

**Description:**

Our client is an American English language Commercial Broadcasting Network that is flagship property of private Corporation. This project's main objective is to help business users analyze booking time sales in a multi-dimensional way to improve their advertisement sales. Client wanted to migrate their entire data from existing oracle database to different environment GCP platform in Google Big-query database using informatica PowerCenter and Matillion as an ETL tool.

# **Responsibilities**

* Understanding the Requirements and modifying/enhancing the existing code/logic to implement the requirements.
* Extracted data through complex SQL queries gathering data stored across multiple tables in the database, while data is available at different granularities across tables.
* Build the jobs in the Matillion to transform the data from the On-Premises to the Big Query tables.
* Developed the Matillion job to perform data cleansing, Transformation and loading to Dimension and Fact tables.
* Written Scripts to copy files from Unix server to GCP environment (Storage Bucket).
* Built ETL jobs to load the data from Storage Bucket to Google Big query database.
* Migrated data from Teradata to Google cloud environment for analytics.

# **Educational Qualification:**

B. Tech (Electronics and communication engineering) from Sri Vidyanikethan Engineering college, C/O-2015

# **Certifications:**

Google Cloud Professional Data Engineer Certified (**Series ID:** 21525 and **Certification ID:** 9KtGrg)

Azure Data Engineer DP-203 (In Progress)