 **Praneeth Kallepu**

*Data Engineer*

*Google Cloud Platform| Bigdata| PySpark*

+91 9550056076

Speaker PhoneMarkerHyderabad | INDIA

[Praneeth.kallepu123@gmail.com](mailto:%20Praneeth.kallepu123@gmail.com)

**D A T A E N G I N E E R I N G | C L O U D | B I G D A T A | P y S p a r k | I N G E S T I O N | S Q L**

# Professional Summary

* Having 4 years of experience into IT with 3+ years of experience implementing Bigdata Technologies in GCP Cloud
* Have 3 years Extensive Experience in IT data analytics projects, Hands on experience in migrating on premise ETLs to Google Cloud Platform (GCP) using cloud-native tools such as BIG query, Cloud Data Proc, Google Cloud Storage, Data flow, Composer – Airflow.
* Strong understanding of Big Data Analytics platforms and ETL in the context of Big Data.
* Hands on GCP cloud native data resources such as storage, ingestion, processing and analytical tools
* Hands-on different transformations in spark which the data getting from different data filesystems from sources Experience in implementing various business requirements using RDDs, Data Frames and Spark SQL - PySpark.
* Good knowledge in handling messaging services using Apache Kafka, Apache Airflow and Azure logic apps.
* Extensive knowledge in Data Warehousing concepts, dimension modelling and ETL process designing to load the data from OLTP (RDBMS), Oracle, SQL Server into OLAP (BAW). Practical understanding of the Data modelling (Dimensional & Relational) concepts like Star – Schema Modelling, Snowflake Schema Modelling, Fact and Dimension tables.
* Diverse experience in all phases of software development life cycle (SDLC) especially in Analysis, Design, Development, Testing and Deploying of applications.
* Strong SQL reporting skills with the ability to create SQL views and write SQL queries to build custom datasets for reporting or analysis
* Excellent communication and interpersonal skills and capable of learning new technologies very quickly.
* Ability to learn quickly and apply new skills in addressing business and technical problems
* Good in Managing & Coordinating among Team Self-Managed, Result-oriented and producing quality work on time.

# Education

* Bachelors in Electronics and Communication Engineer from Glocal University April *201**8*

# Work History

* Data Engineer – Harman Connected Services , Bangalore from *July 2021 to Sep13 - 2022.*
* Associate Engineer – Wipro Technologies from *July 2018 to Jun 2021. (SEGMAP ICT CONSULTING AND SOLUTIONS PRIVATE LIMITED)*

 **Technical Tools**

* Cloud Computing: GCP, MS- Azure.
* Bigdata frameworks: Python, Jupiter.
* Big data Technologies: Hdfs, Hive, Spark.
* ETL: Logic Apps, Informatica, ODBC
* DBMS/RDBMS: Oracle, SQL Server
* Programming: SQL, Python , Scala
* Operating systems: Windows, Linux, UNIX

# Work Experience

## Assignment -1:

**Client: NFU Mutual**

## Company: Harman connected Services

**Description**: The national Farmers Union Mutual Insurance Society Limited, trading as NFU Mutual, is a UK mutual insurance composite.

## Role: Data Engineer Responsibilities:

* + Coordinated with team and Developed framework to generate Daily ad-hoc reports and extracts large sets of structured, semi structured and unstructured data from enterprise data from BigQuery.
  + Extensively Used Dataflow as ETL to create Data pipelines, Mappings, Copy Activities, Transformations based on Import/export data between GCS, BigQuery and on-premise data storages by batch, incremental, stream processing.
  + Mapping and designing ingestion and transformation of data from multiple sources, creating a cohesive data asset.
  + Proficient in creating Data pipelines for on-cloud ETL processing; copy activity, custom Azure develop etc.
  + Carried out data transformation and cleansing using SQL queries, Python and PySpark.
  + Design and develop ETL integration patterns using Python on PySpark.
  + Worked on reading and writing multiple data formats like JSON, ORC, Parquet on storage using PySpark.
  + Build data pipelines in airflow in GCP for ET related jobs using different airflow operators.
  + Used cloud shell SDK in GCP to configure the services Data Proc, Storage, BigQuery
  + Participated and Coordinated with Data Science team in implementing Advanced Analytical Models in Spark.
  + Work related to downloading BigQuery data into pandas or Spark data frames for advanced ETL capabilities.
  + Worked with google cloud data catalogue and other google cloud APIs for monitoring, query and billing related analysis for BigQuery.
  + Created BigQuery authorized views for row level security or exposing the data to other teams.
  + Implement CICD (Continuous Integration and Continuous Development) pipeline for Code Deployment
  + Monitoring BigQuery, Dataproc and cloud Data flow jobs via Stack driver for all the environment.

**Environment:** Google cloud platform, Spark, PySpark, Apache Airflow, Scala , Data Bricks, Azure Data Factory, GCS, Big Query, Data Flow, Data Proc, Data bricks, Composer, Stack driver.



## Assignment -2:

**Client: ASC Consulting Pvt Ltd *Jun’2018– June’2021***

## Company: Wipro Technologies]

**Description:** ASC Group is the Chartered Accountant & legal consultancy company in India offering financial management, tax, legal advisory, outsourcing services to industries,

## Role: Associate Engineer Responsibilities:

* + - Designed and Developed SSIS packages using SQL SERVER 2005/2008 Integration Services.
    - Developed SSIS Packages using xml, OLEDB, file system task in SQL SERVER Integration Services to upload and download the ETL DATA.
    - Developed Stored Procedures and Functions.
    - Creating Reports with MS Reporting Services 2005.
    - Importing data for Data Warehouse and Reports with Integration Services using SQL SERVER 2005 Integration Services.

**Environment:** MSBI, SQL Server 2005, Windows 2003 Server.

