

## **Chandrakant Bidve**

**GCP Data Engineer**

---

### **Professional Snapshot:**

- Having 4+ years of experience in Data Engineering using **Google Cloud platform, Apache Spark, Python, Hive and SQL** .
- Having experience in various Google Cloud Products like **Cloud Composer, Data Proc, Big Query, Cloud Storage, Cloud Composer**.
- Hands-on experience of Google Cloud Platform components and **Big Data technologies**.
- Experience Cloud computing – Development, Implementation.
- Experience in dealing with various file formats like **CSV, JSON, Parquet, ORC, Avro** .
- Experienced on cloud Platform services **SaaS, IaaS and PaaS**.
- Working experience on GCP partitioning Experience with **RDD, Spark Data frames, Spark SQL**.
- Capable of processing large datasets structured and semi structured by writing Hive and SQL Queries as per business requirements and supporting system application architecture.
- Strong experience in migrating on-premise **Data warehouse applications to Cloud** .
- Hands on experience on Apache Spark. Strong experience in Spark code to **perform ETL**.
- Involved in writing **data transformations, data cleansing using Apache Spark**.
- Good Knowledge About **Cloud Dataflow**.
- Team player with interpersonal and good communication skills.
- Worked on importing and exporting data from different databases like Oracle **SQL , MySQL into GCP** .
- Strong experience in **HIVE and SQL** scripts as per the requirement.

### **Technical Skills:**

- **Programming Language** : Python, PySpark.
- **Big Data (GCP Tools and Services)** : Big Query, Data Proc, Google Cloud Storage, Google Cloud SDK, Shell, Cloud Composer, SQL.
- **Scheduling Tools** : Airflow.
- **IDE** : PyCharm, Jupyter notebook.
- **Operating Systems** : Linux, Window.

---

### **Organizational Experience:**

- **Company Name** – Sparkynode Analytics.
- **Duration** – Nov 2018 to till now.

### **Work Experience:**

#### **PROJECT-3**

**TITLE : CIGNA**

### **ROLES AND RESPONSIBILITIES:**

- Analyze and organize raw data.
- Combine raw information from different sources.
- Implementation of Python scripts for various Google cloud API clients like BigQuery, Configuration, Dataproc and many more.
- Developed Python utilities which will be used to connect source and target systems for the data processes.
- Refactoring of existing Pyspark codebase to make it generic across the platform.
- Pre-processing and cleaning on data using Spark.
- Design and develop Spark transformation in Dataproc.
- Write complex SQL Queries for data analysis.
- Development of custom frameworks/components for better reusability.
- Storing the extracted data in Oracle database for further querying and analysis.
- Develop data pipeline using cloud composer to process large volume of data.
- Analyse logging information and start or resume failed tasks.
- Responsible for debugging and troubleshooting during data processing.
- Creating database schemas for newly generated data.
- Debugging and fixing the issues by sampling the data.
- Coordinating with the team to meet the company requirements.
- Attending daily stand-up meetings (Scrum), and Requirement review meetings to analyse requirements for each user story in a sprint.

## **PROJECT-2**

**TITLE : THALES**

### **ROLES AND RESPONSIBILITIES:**

- Created Python utilities which will be used to connect source and target systems For the data process we have used PySpark.
- Combine raw information from different sources.
- Updated existing pyspark code and tried to make them more generic across the platform.
- Develop SQL Queries for fetching the data from database.
- Oracle database for further querying and analysis.
- Responsible for debugging and troubleshooting during data processing.

## **PROJECT-1**

**TITLE : MENZIES DISTRIBUTION**

### **ROLES AND RESPONSIBILITIES:**

- Updated existing pyspark code and tried to make them more generic across the platform.
- There are various frameworks with their specific use and I am part of developing most of their code.
- Created Python DataFlow and dataflow templates to process large volume of data.
- Performing pre-processing and cleaning on data using Spark.
- Develop SQL Queries for fetching the data from database.

---

### **Educational Qualification:**

- Completed PG from Pune University