**NAGAGOPI KARANKI**

**Mail Id: nagagopi2916@GMAIL.COM Mobile #: 7207360311**

**Profile Snapshot**

* Overall having 4.6 years of experience which includes 4 years on Big Data Eco Systems **Hadoop/Spark** in ingestion, storage, querying, processing.
* Worked on AWS Cloud with S3, EMR, Glue, Lambda, SNS Services.
* Expertise with the tools in Hadoop Ecosystem including **HDFS, Sqoop, Hive.**
* Hands-on experience in spark environments in **Spark** components like **Spark-Core, Spark-SQL**.
* Experience in **PYSPARK** environment.
* Experience in Manipulating/Analyzing Large datasets and finding patterns and insights within structured data.
* Expertise in using various Hadoop ecosystems such as HDFS, Hive, Sqoop for data storage and analysis.
* Experience in migrating the data using SQOOP from HDFS to Relational Database System and vice-versa according to client's requirement.
* Experience in writing HIVE queries.
* Load data from RDBMS to HDFS / Hive tables.
* Implement a solution around Hadoop to store and process the customer data.
* Having experience in Banking and Health Care domains.

**Education:**

M. Tech(CSE) K L University(KLU).

**Work Experience:**

**Accenture India Pvt Ltd** **:** Apr2018 to till date

**Technical Skills**

**Big Data :** HDFS, Sqoop, Hive, Python, Pyspark, Spark-Core, Spark-Sql, Yarn,Jenkins, Gitlab, SQL

**Operating Systems** : Linux, Windows

**Programming Languages** : SQL, Python, HiveQL

**Database skills** : MYSQL, DB2, Oracle

**Other tools** : Eclipse, Winscp, control-m, Autosys, Ops console, service now.

**Work Experience**

**Project Undertaken: 1**

**Client : Novartis**

**Environment : HDFS, Hive, Python, Spark-sql, Data bricks Scheduler, Airflow, AWS**

**Role : Big Data Engineer**

**Project Description:**

Novartis International is a Swiss multinational pharmaceutical company based in Basel, Switzerland. Novartis manufactures the drugs. I worked in Culture Sense Project where we collect data for the HR operations to find the cultures based on the employee learning KPI's.

The objective of the project is to collect, consolidate and analyse HR Operational data coming from multiple channels. The results from analytics are used to track our culture transformation and understand the factors of culture that drive performance.

**Roles & Responsibilities:**

* Gathering Requirement from the client for the new Business Requirements
* Collecting all the required details to onboard new data source(s) into the environment
* Preparing Metadata entries for the new data source(s)
* Preparing Design Document and Data Model Sheet for mapping the attributes
* Loading the data from sources into the Raw Layer by validating the Data Quality Rules
* Maintaining History Load and Incremental load based on the requests
* Creating parquet tables using spark sql on different layers
* Performing aggregations in published layer for various KPI’s
* Prod Data Validation for the History/Incremental on KPIs and providing the signoff to the production.
* Maintaining the code on Gitlab Repository
* Automatic Build and Deployment in Jenkins using Jenkins Pipeline

**Project Undertaken: 2**

**Client : CITI BANK**

**Environment : HDFS, Hive, Sqoop, Scala, Spark-sql.**

**Role : Hadoop Developer**

**Project Description:**

Citibank is the consumer division of financial services multinational Citigroup. Citibank provides credit cards, mortgages, personal loans, commercial loans. One area of Citi’s operations where Big Data analytics has been implemented successfully is in customer retention and acquisition. This project involves analysing data and targeting promotional spending using machine learning algorithms. Another is to scan transactional records to spot anomalies, which in the case of Citi’s customers, can mean incorrect, unusual or fraudulent charges. The costs resulting from these anomalies are far easier to manage if the problem is spotted quickly – or even before it happens, through predictive modelling. The platform is primarily built on Hadoop, and datasets are sourced from different applications that ingest multi-structured data streams from transactional stores, customer feedback, and business process data sources.

**Roles & Responsibilities:**

* Ingest Customer Transaction data into the Hdfs data lake using Sqoop
* Loaded the data into Spark RDD and did in-memory data Computation to generate the Output response.
* Create External Hive tables and load the data into the hive tables from the datasets in the data lake.
* Created Dataframes to read the data from JSON files using Pyspark and loaded into Hive tables
* Attending the Client meetings to gather the new requirements
* Involved in loading data from the UNIX file system to HDFS.
* Created Hive tables to store the processed results in a tabular format.
* Wrote complex SQL queries to verify target tables with business rules.
* Created Data frames using PYSPARK to load the data from local and HDFS system.
* Created hive tables to store the processed results in the hive table using Spark.
* Worked on Spark-SQL to fetch the data from Data Frames by creating temp tables to validate the data.
* Queries into Spark transformations using Spark Data frames

**Declaration:**

I hereby, solemnly declare that all information furnished above is genuine to the best of my knowledge and belief.

Date: Hyderabad **NagaGopi K**

Place: 18 OCT 2022 (Signature)