Venkataramana T [venkat195.job@gmail.com](mailto:venkat195.job@gmail.com) 9686329153

# EXPERIENCE SUMMARY

I have 4+years of experience in total IT and relevant into bigdata 3.11 years, strong technical skills in **SQL, Python, Spark, PySpark,Databricks,Gremlin,GraphQL,** and developing applications using **Hadoop Ecosystem components**.

* Good experience in Hadoop ecosystem& Big-Data Analytics.
* Experienced in Spark ecosystem components like Spark batch, Spark streaming, Spark SQL, Spark-core.
* Experienced in using HDFS, Map Reduce, Hive, Cloudera Manager.
* Good knowledge of Hadoop Architecture and Yarn Architecture.
* Good knowledge of SQL, Spark Architecture
* Good knowledge and Understanding of Spark-standalone-Architecture and Spark-Yarn Architecture.
* Experienced in working Hive data warehouse system, developing data pipeline, implementing.
* Complex business logic and optimizing business logic.
* Experienced Hadoop Developer, have a strong background with file distribution systems in a Bigdata.
* Understands the complex processing needs of big data.

# SKILLS SUMMARY

Domain : Media and Telecommunications

Programming Languages : Python

Operating System/ ERP Version : Windows 10,Ubuntu/(Linux)Tools / DB / Packages Framework / ERP Components : SQL, Spark, PySpark, Python, Hive, Sqoop,HDFS and Linux

# PROFESSIONAL SUMMARY

* Working as Data Engineer at **Comcast india Engineering LLP** India from Oct 2022 to Aug 2023.

# Work Experience

**Project 2**

Project Name : ROCI Client : Comcat

Team Size 14

Start Date : Oct2022 Date : Aug 2023

# PROJECT DESCRIPTION:

Modelling the complete footprint of Comcast's Access network as a graph, right from headend till cable modems, using physical and logical entities, relationships and algorithms . This can be used to visualize the complete access network topology of Comcast (~50 million Cable modems, 300K+ fiber nodes, 30K+ RPD's, 4000+ CMTS, RF plant elements from the headend till CPE, by dynamically updating the data non & near real-time . This platform will be the source of truth for world's largest cable network upgrade worth $4Billion - to find the nodes, cable drops and modems that can be updated, and for Proactive Network Maintenance of RF plant.

# ROLE & CONTRIBUTION:

* Played a pivotal role in migrating from AWS NeptuneDB to Neo4j DB, including initial data modeling for optimized database structures.
* Successfully converted Gremlin queries to Cypher queries, facilitating the migration and ensuring compatibility with Neo4j.
* Created and managed data ETL pipelines in Databricks using PySpark, ensuring efficient data processing and transformation.
* Utilized GraphQL APIs to publish data to consumers, providing a streamlined and flexible data access solution.
* Designed and implemented data quality monitoring pipelines, ensuring data accuracy and reliability.
* Developed interactive Grafana dashboards to visualize and report on data quality metrics, enabling informed decision-making.
* Wrote comprehensive unit tests for all GraphQL APIs, ensuring the robustness and reliability of data access and manipulation.

# TECHNOLOGY & TOOLS:

Python,Pyspark, AWS ,SQL, AWSNeptuneDB, Neo4jDB ,databricks, ETL, Grafana, dataAnalytics, Concourse, CI/CD, Kubernetes, git ,Linux

# Inceptez Technologies Pvt Ltd Project 1

Project Name : LB DataLake

Client : Llyods Bank

Team Size 08

Start Date :Jan 2019

End Date : Aug 2022

# PROJECT DESCRIPTION:

With overall million of customers like rats(ora-std) MTG, PCL,CRD, customers and a finance, data Optima. analytics is a critical, component of Llyods Bank business. Bank collects huge amount of usage information about all their products, services and sites such as business data, Financial Data, product and device specific information and other similar data and that will be consumed for Business Analytics and build Big Data Solutions for various Llyods Bank Business units.compliance, business and marketing to improve the services, measureThe collected data is aggregated and analysed for adhere legal and analyse the usage and to help.

# ROLE & CONTRIBUTION:

* Data is pulling from Oracle using Spark to HDFS as Hive tables.
* Reading data from source (hive) LZ.
* Understand business story and analyses functional specifications & design feasible solutions. Using PySpark For processing and Applying Analytics on data as per client Requirement.
* Developed code to write the logic data into Hive External Tables. Involved in configuring and creating Autosy schedule.
* Involved in various log files to fix the issues during unit testing and performance tuning & involved in developing the PySpark.

# TECHNOLOGY & TOOLS:

SQL,Python, Hive, Spark, Yarn, PySpark, Windows, Linux and Hadoop Ecosystems.

# EDUCATIONAL QUALIFICATION:

B.Tech from JNTUA 2018

Q1. How conversant are you with Python/Scala on a scale 1-5

----> PYTHON(4/5),

Q2. Would you be interested to take a data structures and Algorithm test in any programming language of your choice

--->Yes

Q3. What Development Techniques do you adopt for your data project?

---> agile development

Q4. Explain in brief what measures were taken to assure data quality.

--> unit test,

Q5. What data Modelling approach did u adopt for enabling Business Intelligence If an

--->LOGICAL

Q6. What technical gaps do you want to work on once you join our organisation?

---->DEVELOPEMENT