

**Aruna D**

**[arunacharvik@gmail.com](mailto:arunacharvik@gmail.com)**

**Contact:8722927334**

## **PROFESSIONAL SUMMARY**

An IT professional with 4.6 years of experience and strong technical skills in SQL, Python, Hadoop, Spark and Pyspark using Big Data & Hadoop development using HDFS, Hive, Sqoop, AWS (S3, Athena, Glue, Lambda etc.) Data Bricks.

- Worked extensively with strong emphasis on Design, Development, Implementation and Deployment of applications.
- Experience in Hadoop ecosystem & Big-Data Analytics.
- Experience with Cloud at AWS (S3, Athena, Glue, Lambda) and Databricks.
- Experienced in Spark ecosystem components like Spark batch, Spark SQL, and Spark- core.
- Experienced in using HDFS, Hive.
- Good knowledge of Hadoop Architecture and Yarn Architecture.
- Good knowledge and Understanding of Spark-standalone-Architecture and Spark-on-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 big-data. Understands the complex processing needs of big data and have experience in developing codes and modules to address those needs.
- Strong Programming skills Python, Object-oriented programming & design. Excellent communication, organizational skills, attention to detail & the ability to meet tight deadlines

## **WORK EXPERIENCE**

- Worked as Software Engineer in the Organization **GMS Solutions PVT LTD** from Apr-2018 to Oct 2022.

## **QUALIFICATION**

B.E from Nagarjuna College of Engineering and Technology, Chikkaballapur.

## TECHNICAL SKILLS:

Technologies	: Python
Databases	: MYSQL
IDEs	: PyCharm and IntelliJ API
API	: Spark
Operating Systems	: Linux and Windows
Ecosystem	: HDFS, Hive, Spark-SQL
Ingestion tools	: Sqoop
Build Tools	: Git Hub
Cloud Technologies	: Athena, S3, Glue, EMR, Lambda, Linux, Data bricks.

## PROJECTS: Feb - 2021 to Oct– 2022 Project 2: MTN

### Project 2: MTN

**Environment:** HDFS, Hive, Python, Sqoop, Oracle Spark, Spark-SQL, Linux, Autosys, Putty.

**Role:** Hadoop Developer

**Description:** Millions and Billions of users under the MTN, they have products like Telecom, the data analytics is a critical component of MTN-business. It collects huge amount of usage information about all their products, services and sites such as Complete Calls and Call Records, product and device specific information and other similar data and that will be consumed for Business Analytics and build Big Data Solutions for various Toronto-Dominion Business units. The collected data is aggregated and analyzed for adhere legal compliance, business and marketing to improve the services, measure and analyze the usage and to help to find services and advertising more relevant to the customers.

### Responsibilities:

- Injection data from Oracle to Big data layer at Hive partition tables.
- Loading data from Hive tables to Python Spark for Processing Data.
- Understand business story and analyze functional specifications & design feasible solutions.
- Based on business logic at applying cleaning the data, applying transformations and analytics with Python spark.
- Developed and implementing of code to write the logic data into Hive External Tables.
- Involved in configuring Bitbucket and creating Build using Jenkins.
- Involved in various log files to fix the issues during unit testing and performance tuning Involved in developing the python spark.

**Project 1: Viacom18**

**Environment:** HDFS, Hive, Python, Athena, Lambda, Databricks, Airflow, Shell Scripting, Linux

**Role:** Hadoop Developer

**Description:** With overall millions of customers at VOOT, customers and a finance, data analytics is a critical component of HPI business. It collects huge amount of usage information about all their products, services and sites such as Channel records, amount, network traffic 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 HPI Business units. The collected data is aggregated and analyzed for adhere legal compliance, business and marketing to improve the services, measure and analyze the usage and to help to find services and advertising more relevant to the customers.

**Responsibilities:**

- Injection data from Oracle to AWS S3 layer at Athena partition tables.
- Loading data from Athena tables to Spark processing framework.
- Understand business story and analyze functional specifications & design feasible solutions.
- Based on business logic at applying cleaning the data, applying transformations and analytics with Python Spark at Azure Databricks.
- Developed and implementing of code to write the logic data into Athena Partition Tables.
- Involved in configuring Bitbucket and creating Build using Jenkins.
- Involved in various log files to fix the issues during unit testing and performance tuning Involved in developing the python spark.

(Aruna D)