Madhava Kondapalli
Big Data Engineer
LinkedIn

Contact: **+91-8328398752**
E-Mail**:** alwaysmadhav08@gmail.com

Career Objective

A computer science graduate who is seeking to find the opportunity to work in a fun and challenging working environment that will encourage me to improve and learn new and necessary skills as well as be motivated by the company to my best for the sake of helping myself and the company advancement in the software engineering industry.

PROFILE SUMMARY

- 3.7 + years of experience in application development in Big Data Ecosystem (**HDFS**, **YARN**, Sqoop, Hive)
- 2 + experience in development of applications in **Apache Spark**, **Scala,** Python and Kafka.
- Extensive knowledge in writing Spark jobs in Scala using Spark SQL & Data Frames for fast processing of data.
- Extensive knowledge of data ingestion using **Sqoop** from External data sources to HDFS and vice versa.
- Extensive knowledge of creating **Hive** tables with performance optimization using bucketing & partitioning.
- Creating workflows using big data orchestration tool Oozie.
- Experience in working with Spark for data manipulation and data aggregations using Scala
- Extensively worked on Hive External and managed tables with different file formats.
- Automated Multiple repetitive **UNIX** flows and in Python using Paramiko library.
- Extensively worked on validating and comparing the data flow in data pipelines using **Python**
- Developed python code to fetch and validate data from different data storages like DB2, Elastic search, EDW
- Experience with DevOps tools **Git-hub** and **Jenkins**.
- Good knowledge on core **java** and oops concepts.
- Good knowledge in **Waterfall** and **Agile** Methodologies.
- Passion to excel in any assignment and have good debugging, analytical and problem-solving skills.

PROFESSIONAL EXPERIENCE

1. Big Data Engineer — From: October 2020 to Present
   IBM India Pvt. Ltd (Contract Employee). - Bangalore, Karnataka

2. Software Engineer — From: August 2017 to October 2020
   Capgemini India Pvt. Ltd. - Bangalore, Karnataka

TECHNICAL SKILLS

- Big Data/Hadoop Ecosystem : HDFS, YARN, Spark, Hive, Sqoop, Oozie, Kafka, Nifi
- Programming languages : Scala, Python and Java
- Database : SQL, DB2 (Basic)
- NoSQL Database : Hbase
- DevOps/Build Tools : Git, Jenkins, Maven, SBT
- IDE : Eclipse, jupyter Notebook, PyCharm, IntelliJ
- Other Tools : WinSCP, Putty, DB Visualizer, Microsoft SQL Server
- Operating systems : Windows, UNIX

WORK EXPERIENCE

Project 3 : Central Components (**CC**)
Role : Big Data Engineer
Duration : Oct 2020 – Present
Client : MBRDI
Team size : 5
Description :

Central Components develops Big Data applications which satisfy the requirements of Mercedes-Benz Bank. We are developing common libraries, applications to handle data and infrastructure, create data pipelines to store historical data, move based on needs, and transform to analyze the data based on business requirements.

Responsibilities:

- Creating pipelines to load the Oracle data into Kafka using Hive tables and Sqoop using Oozie workflow.
- Create and Maintain Hive tables which deals with multiple file formats data.
- Building data pipelines between Hive, Spark and Hbase to perform transformations and aggregations.
- Importing tables having normal and complex data types like XML etc from Oracle to Hive using Sqoop.
- Creating smoke tests to check all big data components are working properly.
- Capturing the changing schema from the Oracle and store in Schema Registry and Kafka.
- Responsible for Unit testing and troubleshooting of big data applications.

Project 2 : Automated Fraud Marking (**AFM**)
Role : Big Data Engineer **Client** : Royal Bank of Canada (RBC)
Duration : Jan 2020 – July 2020 **Team size** : 6
Description :

AFM is designed to automate fraud marking process of TSYS personal/business Fraud Flagged transactions by analyzing the data received from different sources and compare against Fraud case data. Once all conditions are satisfied then the transaction will be marked as fraud and an incident will be created for the transaction.

Responsibilities:

- Involved in Requirement Analysis, Design, and Development of spark application.
- Developed spark code to read and process data from Hive and SQL tables.
- Create and Maintaining Hive tables (Manage & External tables).
- Used Partitioning and Bucketing techniques in hive to optimize performance.
- Building data pipelines between Hive, Spark and SQL to perform join transformations and aggregations.
- Loading the required data into hive tables using Nifi.
- Responsible for Unit testing and troubleshooting.

Project 1 : RBC Fraud Analysis (**RFA**)
Role : Hadoop & Spark Developer **Client** : Royal Bank of Canada (RBC)
Duration : Feb 2018 – Dec 2019 **Team size** : 5
Description :

This project is intended to detect illegal and malicious money transactions based on various FINTRAC proposed scenarios and store in different storages which will be used by RBC fraud applications like CAMPS, Retina.

Responsibilities:

- Analyzing the different scenarios and developing respective Scala code for Spark jobs.
- Creating Sqoop jobs with maximum optimizations to move the data between HDFS/ Hive tables and RDBMS.
- Integration Hbase and Hive tables and executed HQL queries perform data analysis.
- Developed spark applications to read data from Hive tables into data-frames for aggregations and insights.
- Created Hive external and managed tables with dynamic partition using ORC format for better performance.
- Created python scripts to automate regular UNIX flows and data validation in data pipelines.

EDUCATIONAL QUALIFICATION

- MCA (2017-20) under Sri Venkateswara University in Tirupati, Andhra Pradesh
- B- Sc (2014-17) in Electronics and Computer Science under Sri Venkateswara University in Tirupati, Andhra Pradesh.

TRAININGS

- 3 Months training Big Data Technologies (Hadoop, Spark with Scala and python) in Capgemini, Pune.

CERTIFICATIONS

- IBM Data Science Professional certificate (From Coursera)
- Oracle Cloud Infrastructure Foundations 2020 Certified Associate
- IBM Certified Database Associate DB2 9 fundamentals
- Scrum Foundation Professional Certificate (SFPC)