# Jagadeesh S E-mail:jagadeeshbdf@gmail.com Mobile: +91-8074853682

**Summary**

* Having around 4+ years of extensive IT experience Development and implementation of software applications which includes experience in Big Data using **Hadoop**, **Hive**, **Sqoop**, **Spark**,**PySpark**,**AWS,S3,EMR**and **Scala** **Programming**.
* Well understanding of Hadoop architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, and Yarn.
* Working Experience on Hadoop Distributions like Cloudera.
* Working experience on Cloudera Hadoop Distribution and HUE web interface.
* Working knowledge of performing Batch processing using Spark & Scala.
* In-depth knowledge and hands-on experience in dealing with Hadoop components like HDFS, MapReduce, HiveQL, Hive, and Sqoop.
* Designing and creating Hive external tables using a shared meta-store instead of derby with partitioning, dynamic partitioning, and buckets.
* Experience in importing and exporting data using Sqoop from Relational Database Systems to HDFS and vice-versa.
* Proficient at using Spark API to cleanse, aggregate, transform and store the datasets.
* Experience in writing HiveQL queries to store processed data into Hive tables for analysis.
* Good interpersonal communication skills, and analytical and logical abilities with demonstrated ability to work as a team member as well as independently.
* Willingness and ability to easily adapt to learn any newer technology or software.

# Professional Summary:

* Working as a **Big Data Developer** in Novitas Technologies from December 2020 to till date.

# Education Details

# B Tech from JNTU Kakinada, Andhra Pradesh.

# Technical Skills

|  |  |
| --- | --- |
| **BigData/Hadoop**  **Components & Ecosystems** | HDFS, Yarn, MapReduce, Hive, Sqoop |
| **Programming Languages** | Scala, Unix |
| **Spark Ecosystem’s** | Spark Core, Spark SQL,Akka |
| **Hadoop Distribution** | Cloudera Manager(CDH) |

|  |  |
| --- | --- |
| **Databases** | Oracle, MySQL |
| **IDE's** | Eclipse, Scala IDE |
| **Build Tool** | Maven,CI/CD, Jenkins, Git |

**Professional Experience: Project # 1:**

|  |  |
| --- | --- |
| **Project Name** : | **Sales Analytics** |
| **Duration :** | **April 2022 - till date** |
| **Client :** | **Unilever Data IQ Team** |
| **Role :** | **Hadoop Developer** |
| **Environment :** | **HDFS, Hive, Sqoop, Spark, Scala** |

**Description:**

Unilever needed to improve its on-shelf availability and product assortments in order to meet its ambitious growth targets for India and South Asia markets. This required an understanding of changing buying patterns based on market insights and analytics and undertaking course corrections in the shortest possible time. As a result, Unilever was able to reduce the lead time required for business insight generation from 20 days to less than 10 hours for 4-6 TB of data, thereby significantly increasing business agility. This project provides solutions for Batch & real-time requirements, Data acquired using Sqoop methods, ETL using HIVE, and reported back to external tables.

# Responsibilities:

* + Interaction with client in gathering and understanding the business requirements.
  + Importing data from RDBMS to HIVE, HDFS using SQOOP.
  + Created and worked Sqoop jobs with incremental load to populate Hive External tables.
  + Load and transform large sets of structured, semi-structured, and unstructured data using Hive.
  + Creating a Hive table with partitions and Buckets and loading data into Buckets.
  + Access HBase data in Hive by integrating Hive with HBase.
  + Implemented Spark and utilizing RDD, Data frames, and Spark SQL for data cleansing, transformation, and referential integrity based on the requirement.
  + Process the code changes in schema validation and transformation as per the new enhancements or the business.

**Project # 2:**

**Project Name: Commercial Date Lake**

**Role: Spark & Hadoop Developer**

**Client: AbbVie**

**Domain: Life Sciences**

**Technologies: HDFS, Spark, Sqoop, Hive, Scala, Linux**

**Duration: Nov 2020 to Mar 2022**

Project Description: CDL houses data for commercial organization. It is one central repository for data related to Marketing, Sales, Patient Services, Health Care Professionals, and Market Access. Data is sourced from a variety of channels and is grouped as per the data classification of Internal, Regulated, and Secret. Data is across the brands in the therapeutic areas of Immunology, Oncology, and Speciality like Women's Health, and HCV.

The main aim of the project is to maintain a centralized data lake that stores the data from all the sources to perform various analytics. In this project, we have loaded the data from various external source systems to the Hadoop HDFS data lake and then made it useful for the different downstream applications.

**Roles & Responsibilities**:

● Responsible for loading of the data into the tables at different levels.

● Writing Spark with Scala for data quality management.

● Understand the business requirements and curate the data as per requirements.

● Writing Complex hive queries as per the business rules.

● Debug the issues in loading the data into Tables and make the data available for business.

● Work on ad hoc requests to create/modify existing tables based on business requirements.

● Bringing the data from Teradata, Oracle, and Informatica into HDFS using Sqoop and making the data available downstream.

● Monitoring of real-time jobs and fixing the issues.

**Technologies** **used**:

Spark Core, Spark SQL, Scala, Python, HDFS, Hive, Teradata, Informatica, GIT Version Control .