**Maruthi**
**Sr. Big Data Developer | Ph: +91 8688784205, +1 669-226-8553 | Email:**
**maruthipydikondala@gmail.com**

## Professional Summary

- Around 5 years of experience in IT developing industry which includes 4+ years of comprehensive experience in Big Data processing using Hadoop and its ecosystem (MapReduce, Pig, Hive, Sqoop, Flume, Spark, Kafka and HBase) and 6 months in designing and developing various web applications using Java, Servlets, HTML, CSS and JavaScript.
- Solid understanding of the Hadoop Distributed File System.
- Excellent Experience in Hadoop architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node and MapReduce programming paradigm.
- Hands on experience in installing configuring and using Hadoop ecosystem components like Hadoop MapReduce HDFS HBase Hive Sqoop Pig Zookeeper Spark and Scala.
- Strong understanding of real time streaming technologies Spark and Kafka.
- Good Knowledge in writing **Spark Applications** in **PySpark and Scala using Dataframes.**
- Good Exposure on Apache Hadoop Map Reduce programming PIG Scripting and Distribute Application and HDFS
- Experience in importing and exporting data using Sqoop from HDFS to Relational Database Systems and vice-versa
- Cluster coordination services through Zookeeper.
- Knowledge of job workflow management and coordinating tools like Oozie.
- Experience in managing Hadoop clusters using Cloudera Manager tool.
- Worked on managing and automating Infrastructure Development and operations involving AMAZON WEB SERVICES(AWS).
- Install and configure chef server /workstation and nodes via CLI tools to AWS nodes.
- Created users and groups using IAM and assigned individual policies to each group.
- Configured Security group for EC2 window and Linux instances.
- Experience in working with Github private repositories and docker repositories.
- Experience with Docker to create, manage, deploy and run containerized applications
- Sound knowledge in various databases like MySQL & NoSQL(cassandra)
- Experience in working with various build tools like Maven.
- Working Experience in frameworks like Spring and Hibernate.
- Strong working experience using Agile methodologies including Scrum.
- Some Knowledge in some of unix/linux commands.
- Excellent ability to understand complex scenarios and business problems and transfer the knowledge to other team members in the most comprehensive manner.
- Strong communication skills, analytic skills, good team player and quick learner, organized and self-motivated.

## Technical Skills:

| BigData | Hadoop, Hbase, Hive, Sqoop, Oozie, Spark, Kafka, Flume, Mapreduce, Zookeeper |
|---|---|
| Operating Systems | UNIX, Mac, Linux, Windows 2000 / NT / XP / Vista. |

| | |
|---|---|
| **Programming Languages** | Java, Scala, Python, Go, C, VB, Objective C, PySpark. |
| **Databases** | DB2, MySQL, Oracle 9i, MongoDB, Cassandra |
| **IDE/Development Tools** | Eclipse 5.x, JBuilder 7.x and RAD 7.0 NetBeans, Intellij |
| **Web Technologies** | J2EE, Soap & REST Web Services, JSP, Servlets, EJB, JavaScript, Struts, Spring, Web works, Direct Web remoting, HTML, XML, JMS, JSF, Ajax. |
| **Frameworks** | Struts2.0, Spring Framework and Hibernate |
| **AWS Compute Services** | EC2, ECS, Lambda |
| **AWS Storage Services** | S3 |
| **AWS Database** | RDS |
| **Version Control Tools** | Git |
| **Web/Application Servers** | IBM Web sphere Application server, Jboss, Apache Tomcat, Nginx |
| **Automation/Build Tools** | Docker, Vagrant, Maven |

**Education:**

Bachelor of Technology from Jawaharlal Nehru Technological University Kakinada, India, 2013.
Masters from Silicon Valley University, California, 2016.

**Professional Experience:**

**Walmart, Bentonville, AR**                                                                    **Nov 2019 – Mar 2021**
**Sr. Data Engineer**
**Responsibilities:**
- Experience in monitoring and managing the Hadoop cluster using Cloudera Manager.
- Developed automic Workflows for daily incremental loads, which gets data from Teradata and then imported into hive tables.
- Very good understanding of Partitions, bucketing concepts in Hive and designed both Managed and External tables in Hive to optimize performance.
- Experience in working with different file formats like parquet, ORC, Avro, Json, csv. Based on the use case and environment will be able to decide the best option among the file formats.
- Solved performance issues in Hive with understanding of joins, Groups, and aggregation and how does it translate to MapReduce jobs. Created Hive tables and working on them using HiveQL.
- Experience in managing and reviewing Hadoop logs.
- Experience in dimensional data modeling, ETL development, and Data Warehousing
- Designed, developed, and tested all job automation workflow schedules via UC4.

- Worked with different teams and customers to maintain production stability, including testing, determine business requirements and change request process.
- Developing and building different spark projects using Maven.
- Worked on creating the RDD's, DF's for the required input data and performed the data transformations and actions using Spark Scala.
- Converted Hive jobs to spark, during migration process from on prem to GCP auto scaling cluster.
- Developed Spark/Scala, Python for regular expression (regex) project in the Hadoop/Hive environment with Linux for big data resources.
- Experience in creating tables, worksheets and pins/answers in ThoughtSpot.
- Importing and Exporting data into HDFS from MariaDB and vice versa using Sqoop.
- Monitor the complex application systems on daily basis and all application processes for system resource usage and performance using One Automation System monitor.
- Data sources are extracted, transformed and loaded to generate CSV data files with Python programming and SQL queries.
- Transferred the data from Teradata to AWS S3 based on the business requirements.
- Creating S3 buckets also managing policies for S3 buckets and Utilized S3 bucket for storage and backup on AWS.
- Defect/Bug fixes in non prod and in production environment tracked in Jira through tickets.
- Experience in creating documents in Atlassian Toolset such as Confluence, Stash.

**Environment:** Hadoop, CDH 5.x, AWS, GCP, Map Reduce, Hive, MariaDB, Sqoop, Scala, Python, Spark 2.x, Linux, UNIX. ThoughtSpot, Automic.

**Charter Communications, Charlotte, NC**　　　　　　　　　　　　　**Feb 2018 – Oct 2019**
**Spark/Hadoop Developer**
**Responsibilities:**
- Responsible for building scalable distributed data solutions using Hadoop Ecosystem.
- Responsible for troubleshooting issues in the execution of MapReduce jobs by inspecting and reviewing log files.
- Participate in the design and development of software using agile development practices.
- Develop Scala and SQL code to extract data from various databases, apply innovative ideas around the Data Science and Advanced Analytics Practices Creatively and present models to business customers and executives, utilizing a variety of formats and visualization methodologies.
- Implement POC for using Apache Impala for data processing on the top of Hive.
- Implemented POC to migrate map reduce jobs into Spark RDD transformations.
- Converts GoLang scripts into spark jobs which takes necessary fields from impala and populate them into HBase.
- Used Spark streaming to receive real time data from the Kafka and store the stream data to HDFS using Scala and databases such as HBase.
- Understanding of data storage and retrieval techniques, ETL and databases, to include graph stores, relational databases, tuple stores, NOSQL like HBASE. Hadoop, MySQL. Spark MLLIB libraries for designing recommendation Engines Analysis predicted by Statistical analysis using Spark
- Estimated the hardware requirements for Name Node and Data Nodes & planning the cluster.
- Experience in using Sqoop to import and export the data from Oracle DB into HDFS and HIVE.
- Developed workflow in Oozie to manage and schedule jobs on Hadoop cluster for generating reports on nightly, weekly and monthly basis and clean up jobs.
- Commissioning and decommissioning the nodes, manage cluster through performance tuning and enhancement.

- Involved in architecture and design of distributed time-series database platform using NOSQL technologies like Hadoop / Hbase, Zookeeper.
- Developed data pipeline using flume, Sqoop and pig to extract the data from weblogs and store in HDFS.
- Worked in Agile development environment in sprint cycles of two weeks by dividing and organizing tasks in Jira. Participated in daily scrum and other design related meetings.
- Transforming the data received from source systems using python and creating the files to load into hive.
- Perform live tests and maintain expert knowledge in areas of expertise.

**Environment:** Hadoop, Java, CDH 5.5, Map Reduce, Hive, Pig, Sqoop, Flume, HBase, Scala, Python, Go, Spark 1.6 & 2.x, Oozie, Linux, UNIX.

**Liberty Mutual, Westborough, MA**                                    **Feb 2017 - Jan 2018**
**Hadoop Developer**
**Responsibilities:**
- Scheduled and executed workflows in Oozie to run Hive and Spark jobs.
- Parse Json files through Spark core to extract schema for the production data using SparkSQL and Scala
- Experience in using DataStax Spark-Cassandra connectors to get data from Cassandra tables and process them using Apache Spark. Exploring with the Spark for improving the performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frame, Pair RDD's, YARN.
- Analyzed business requirements and cross-verified with functionality and features of NOSQL databases like HBase, Cassandra to determine the optimal DB.
- Developed Spark scripts to import large files from Amazon S3 buckets.
- Efficiently handled periodic exporting of SQL data into Elastic search.
- Worked on GitHub and Jenkins continuous integration tool for deployment of project packages.
- Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs and Scala.
- Uploaded and processed terabytes of data from various structured and unstructured sources into HDFS using Sqoop and Flume.
- Implemented various hive optimization techniques like Dynamic Partitions, Buckets, Map Joins, Parallel executions in Hive.
- Developed workflow in Oozie to automate the tasks of loading the data into HDFS and pre-processing with Pig.
- Experience in managing and reviewing Hadoop log files.
- Responsible to manage data coming from different sources.
- Involved in loading data from UNIX file system to HDFS.

**Environment:** Hadoop, HDFS, MapReduce, Spark, Pig, Hive, Impala, Sqoop, Flume, Kafka, HBase, Oozie, SQL scripting, Cassandra, Linux shell scripting, Eclipse and Cloudera, AWS.

**Sonos Inc, Santa Barbara, CA**                                    **Aug 2016 - Jan 2017**
**Role: Hadoop Developer**
**Responsibilities:**
- Responsible for building scalable distributed data solutions using Hadoop Ecosystem.
- Developed HIVE UDFs to perform data cleansing and transforming for ETL activities.
- Developed HIVE UDFs and UDAFs for Data analysis and Hive table loads.
- Wrote HIVE UDF's to get the Data from HBase and Put the Data on HBase
- Developed data pipeline using Flume and Sqoop to ingest data into HDFS for analysis
- Developed Sqoop scripts to import the data from SQL Server to HDFS

- Worked extensively on tuning Hive Jobs
- Successfully loaded files to Hive and HDFS from Cassandra. Processed the source data to structured data and store in NoSQL database Cassandra. Created alter, insert and delete queries involving lists, sets and maps in DataStax Cassandra. Worked in a language agnostic environment with exposure to multiple web platforms such as AWS, databases like Cassandra.
- Wrote UDF's for Apache Drill
- Configured Hive Storage plugin, HBase Storage Plugin and DFS plugin to create Views on Drill
- Implemented POC on Talend Open Studio for Big data.

**Environment:** Hadoop Framework, MapReduce, Hive, Sqoop, HBase, Cassandra, Flume, Oozie, Java (JDK1.6), SQL Server, Talend for Big Data, Apache Drill.

`