**Rajashekhar M**

**Phone: 704-449-1881**

**Email: Raj.neni123@gmail.com**

**Professional Summary:**

* **Proficient** **around 8+ years** of professional experience as a Java/Big Data Developer.
* Expertise in using various Bigdata technologies such as MapReduce, Pig, Hive, Zookeeper, Sqoop, Oozie, Spark, Scala, NiFi.
* Proficient in writing Hadoop jobs for analyzing data using Map Reduce, Hive and Pig
* Developed Kafka consumer to retrieve information from producers. And, Converting retrieved data into multiple files.
* Developed Java Mapper and Reducer programs for complex business requirements
* Developed customized spark internal framework for applying business logic
* Optimized Hive tables using techniques such as partitions and bucketing to provide better performance with HiveQL queries.
* Developed shell script to run spark application on cloud.
* Good knowledge in SQL database - Oracle and MySQL.
* Proficient in job workflow scheduling and monitoring tools like Oozie, Zookeeper and NiFi.
* Experienced in developing applications using Java/J2EE technologies like Servlet, JSP, Hibernate, Spring, JDBC, JNDI, JMS, RESTful Web Services etc.
* Expertise in developing applications using Spring and Hibernate (Object/Relational mapping framework).
* Proficient in writing build scripts using Ant & Maven.
* Experience in Requirements gathering, Design & Development.
* Experience in writing database objects like Stored Procedures, Triggers, SQL, PL/SQL packages and Cursors for Oracle.
* Knowledge of various design patterns.
* Experienced in using SVN, Clearcase for version control.
* Proficient in unit testing the application using JUnit, MRUnit and logging the application using Log4j.
* Experienced in Waterfall & Agile development methodology.
* A Good team player with excellent communication and inter-personal skills with a positive attitude towards work, loyalty towards time schedules for worthful delivery.

**Technical Expertise:**

**Big Data Ecosystems** : Hadoop, HDFS, MapReduce, HBase, Hive, Scala, Sqoop, Flume, Oozie,

Spark, Zookeeper

**Databases** : MySQL, HBase, MS\_SQL

**Operating Systems** : Linux, Windows

**Tools**  : Eclipse IDE, Putty

**Programming Languages**: Scala, SQL, NoSQL, PL/SQL, T-SQL, HiveQL

**Packages** : VMWare, Oracle VM Virtual Box, SSIS(ETL), MS Office

**Professional Experience:**

**Sammons Financial Group, Sioux Falls, SD March '2017 - Till Date**

**Title: Big Data Developer**

**Responsibilities:**

* Involved in initial requirement gathering. Played an active role in analyzing, documenting and designing the application.
* Played an active role in creating business document, functional document, Technical document.
* Experience with Hortonworks distribution.
* Experienced in loading data from different relational databases to HDFS using Sqoop.
* Created Hive internal/external tables with proper static and dynamic partitions and working on them using HQL.
* Deployed data from various sources into HDFS and building reports using Tableau.
* Written several Map reduce Jobs using Java API, also used Jenkins for Continuous integration.
* Collected the logs data from Web Servers and integrated in to HDFS using Flume.
* Used the Spark - Cassandra Connector to load data to and from Cassandra.
* Building, managing and scheduling Oozie workflows for end to end job processing.
* Experienced in extending Hive and Pig core functionality by writing custom UDFs using Java.
* Analyzing of large volumes of structured data using SparkSQL.
* Migrated HiveQL queries on structured into SparkSQL to improve performance.
* Created Hive tables, worked on loading data into hive tables and Experienced in loading and transforming heavy sets of structured, semi structured and unstructured data.
* Extracted Real time feed using Spark streaming and convert it to RDD and process data into Data Frame and load the data into Cassandra.
* Used Spark streaming api for performing transformations and actions on the fly for building common learner data model which gets data from Kafka in real time.
* Has developed custom components and multi-threaded configurations with a flat file by writing JAVA code in Talend.
* Written Spark Streaming jobs in Scala Programming language.
* Extensive experience in creating Lambda functions manually using Java and Python.
* Implemented S3 by creating a bucket and passing the event information to Lambda.
* Experienced in using Automation Scheduling tools like Autosys and Control-M.
* Knowledge on using spring and Rest services and connecting it to the Kubernetes cluster.
* Extensively used Microservices and Postman for hitting the Kubernetes DEV and Hadoop clusters.
* Deployed various Microservices like Spark, MongoDB, Cassandra in Kubernetes and Hadoop clusters using Docker.
* Launched Kinesis Streams services in new AWS regions and granting access to new users through IAM.
* Setting up and worked on Kerberos authentication principals to establish secure network communication on cluster and testing of HDFS, Hive, Pig and MapReduce to access cluster for new users.
* Used Amazon Kinesis as a messaging system for getting all the weblogs, data from rest end points into the Amazon Redshift.
* Used Hortonworks AMBARI for job browser, file browser, running hive and impala queries.
* Configured Spark streaming to receive real time data from the Kafka and store the stream data to HDFS and store in databases such as HBase.
* Worked on RapidMiner, a Data scientist tool and created several Hadoop operators like Hive Operator, Spark Operator, Mongo Operator.
* Designed ETL process using Talend Tool to load from Sources to Targets through data Transformations.
* Experienced in ETL methodology for performing Data Migration, Data Profiling, Extraction, Transformation and Loading using Talend and designed data conversions from large variety of source systems including Oracle DB2, SQL server, Teradata, Hive and non-relational sources like flat files, XML files.
* Implemented Security Side car and AAF as security authentications for Hadoop cluster.
* Used Python for pattern matching in build logs to format errors and warnings.
* Involved in creating UI using Node.js and called different microservices to setup the frontend.
* Worked on CodeCloud, a git repo for continuous code checkins.
* Responsible for developing, support and maintenance for the ETL (Extract, Transform and Load) processes using Talend.
* Developed the Talend jobs and make sure to load the data into HIVE tables & HDFS files and develop the Talend jobs to integrate with Teradata system from HIVE tables.
* Assisted with testing and monitoring Control- M upgrade and active schedule functions, ensuring effective issue resolution and continuous operations.
* Extensive experience in using Microservices, Kubernetes (prod, test, dev) environments and Docker.
* Good experience in developing several seed templates like Scala, MongoDB, Zeppelin and Spark.
* Experienced in configuring the yaml files for Spark, MongoDB, and Cassandra and deployed in Docker for connecting to the several Microservices.

**Environment**: Hadoop 2.7.0, YARN, HDFS, Spark 2.1, Sqoop 1.99.7, Hive 2.1.1, Flume 1.7.0, Oozie 4.2.0 HDP 2.5, Pig 0.16.0, Kafka 0.9.0, Hbase 1.1.2, Zookeeper 3.4.8, Talend Studio, Jenkins 2.0 MySQL 5.6.33, Java 8, Superputty, Scala IDE, Bit Bucket.

**Intermountain Health Care, Salt Lake City, UT March '2016 - Feb ‘2017**

**Title: Big Data Developer**

**Responsibilities:**

* Followed the standard for **Software Development Life** Cycle Processes (Selection of a Life Cycle Model, Project Initiation, Project Monitoring and Control, Requirements, Design, Implementation, Installation, Operation and Support.
* Participated in periodic technical reviews of various modules and developed schema to improve them.
* Involved in architectural discussions, and Installed and configured Hadoop MapReduce, HDFS, Developed multiple MapReduce jobs for data cleaning and preprocessing.
* Batch-processed data using PIG and Hive.
* Used various built-in functions and regex operations to extract data
* Developed Pig UDF's in java to derive business metrics.
* Created Hive Tables, loaded data from Oracle using Sqoop; and loaded the processed data into Netezza database.
* Optimized processing by using specialized joins in Pig and bucketing in Hive.
* Performed data analytics in Hive and then exported this metrics back to Oracle Database using Sqoop.
* Developed Map reduce programs for data extraction, data manipulation.
* Worked on generating reports using SPARK SQL for fast processing of data.
* Worked on data frames in SparkSQL to get data from Hive tables and applied transformations on data frames.
* Worked with different file formats and compression techniques in Hadoop.
* Created unix shell scripting to handle day to day jobs.
* Used Oozie workflows to automate various jobs i.e., Pig, Hql and shell scripts.
* Continuously monitored and managed the Hadoop cluster through Cloudera Manager.
* Worked on developing different kinds of reports using different loaders and delimiters.
* Indirectly managed offshore team.
* Worked along with Datawarehouse team and business analysts to reduce the gaps between the requirement gathering.
* Reviewed code to ensure that all the Processes and Standards are being ensured that correct implementation of requirements, coding standard, project architecture, efficient code writing techniques.

**Environment**: HDFS, Sqoop, Pig, Hive, MapReduce, Oozie, Spark SQL, MySQL, Eclipse, RedHat Linux, RDBMS.

**Harley Davidson, Milwaukee, WI Jan '2015 - Feb ‘2016**

**Title: Hadoop/Bigdata Developer**

**Responsibilities:**

* Went through the total project development life cycle and have a clear understanding of various tiers as the project was a multi-tier application.
* Coordinated with various groups of the client to gather business and system requirements.
* Experience on working with Hortonworks distribution.
* Involved in review of functional and non-functional requirements (NFR's).
* Responsible to manage data coming from different sources.
* Responsible for coding MapReduce program, Hive queries, testing and debugging the Map Reduce programs.
* Worked on migrating MapReduce programs into Spark transformations using Spark and Scala, initially done using python (PySpark).
* Designed, developed and maintained data integration programs in Hadoop and RDBMS environment with both RDBMS and NoSQL data stores for data access and analysis.
* Implemented Map Reduce programs to handle semi/unstructured data like xml, Json, Avro data files and sequence files for log files.
* Developed Simple to complex Map/reduce Jobs using Hive, Pig and Python.
* Used Spark-SQL to Load data into Hive tables and written queries to fetch data from these tables.
* Experienced in querying data using SparkSQLon top of Spark engine for faster data sets processing.
* Imported semi-structured data from Avro files using Pig to make serialization faster.
* Managing and scheduling Jobs on a Hadoop cluster using Ooziework flows and java schedulers.
* Indexed documents using Elastic Search.
* Developed various Python scripts to find vulnerabilities with SQL Queries by doing SQL injection, permission checks and performance analysis.
* Experienced with Kerberos authentication to establish a more secure network communication on the cluster.
* Handled importing of data from various data sources, performed transformations using Hive, MapReduce, and loaded data into HDFS.
* Extraction, transformation and loading of data from various file formats like .csv, .xls, .txt and various delimited formats using Talend Open Studio.
* Experienced in writing Spark Applications in Scala and Python (Pyspark).
* Imported data from AWS S3 and into spark RDD and performed transformations and actions on RDD's.
* Processed the web server logs by developing Multi-hop flume agents by using Avro Sink and loaded into MongoDB for further analysis.
* Experienced in converting Hive/SQL queries into Spark transformations using Spark RDD, Scala and Python.
* Imported weblogs & unstructured data using the Apache Flume and stores the data in Flume channel.
* Exported event weblogs to HDFS by creating a HDFS sink which directly deposits the weblogs in HDFS
* Experienced in connecting Avro Sink ports directly to Spark Streaming for analyzation of weblogs.
* Implementing data quality checks and transformations using Flume Interceptor.
* Implemented the business logic in Flume Interceptor in Java.
* Worked on MongoDB for distributed storage and processing.
* Responsible for using Flume sink to remove the data from Flume channel and to deposit in No-SQL database like MongoDB.

**Environment**: Hortonworks HDP, Hadoop, Spark, Flume, Elastic Search, AWS, EC2, S3, Pig, Hive, Python, MapReduce, HDFS, Tableau, Informatica, VPC.

**KMPG Group, Mumbai, India Jan '2010 - Sep '2014**

**Title: Java Developer**

**Responsibilities:**

* Involved in Design and Code Construction with **OOPs** Implementation.
* Involved in Maintenance / Enhancements of the application which includes developing the new reports.
* Developed **JSP**, Tiles, Validator, Action classes and **EJBs** for various user transactions.
* Designed **Java Servlets** and Objects using **J2EE** standards.
* Implemented the project by using **MVC** framework.
* Used Struts Validator framework to validate user input.
* Used Struts **MVC** framework along with **JST** for developing **J2EE** based web application.
* Implemented Hibernate to map all the tables from different data sources to make database updating.
* Designed Implementation of **GUI** using **HTML**, **JSP** and **Java Script** for front-end validations.
* Implemented Singleton Design Pattern by sharing the single service locator object across different Business Delegates.
* Developed the **XML** Schema for the data maintenance and structures.
* Implemented EJB session beans, messaging technologies like Message Driven Bean, **JMS**.
* Involved in designing Database Connections using **JDBC**.
* Involved in design and Development of **UI** using **HTML**, JavaScript and **CSS**.
* Used **JSP** for presentation layer, entity beans as the model to represent the persistent data and interacted with Oracle database as per **MVC** Architecture.
* Used **JDBC** calls in the Enterprise  to access Oracle Database.
* Involved in Integration testing and defect fixes.
* Developed **Unit Test** Scenarios and tested Test Scenarios in Web methods using **N-Unit**.

**Environment:** Java, J2EE, Junit, JDK 1.5, Spring core, Spring MVC, Hibernate, SOAP/Rest Web services, Ajax, XML, jQuery, HTML, CSS, TC Server, SVN, Maven, Jenkins, Splunk, SFX, MAX, SEED. Oracle