**Varun Reddy**

**Ph: 832-794-7482**

**Email:** [**varunrred@gmail.com**](mailto:varunrred@gmail.com)

**Professional summary:**

* 5+ years of professional experience in IT, including work experience in Big Data Analytics, Distributed Computing using Hadoop Ecosystem tools and Java.
* String hands on experience using major components in Hadoop Ecosystem like Spark, Map Reduce, HIVE, PIG, HBase, Sqoop, Kafka, YARN.
* Excellent knowledge and understanding of Distributed Computing and Parallel processing frameworks.
* Developed transformations on high volume data in HDFS using JAVA and Spark framework
* Strong experience with developing end-to-end spark applications in scala.
* Worked extensively on troubleshooting issues related to memory management, resource management, with in spark applications.
* Strong knowledge on fine-tuning spark applications and hive scripts.
* Written complex mapreduce jobs to perform various data transformations on large scale datasets.
* Used Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive
* Possess good Linux and Hadoop System Administration skills, networking, shell scripting and familiarity with open source configuration management and deployment tools such as Chef.
* Integrated Kafka-Spark streaming for high-efficiency throughput and reliability
* Experience in installation, configuration, and monitoring Hadoop clusters both in house and on the cloud (AWS).
* Strong knowledge on creating and monitoring Hadoop clusters on Amazon EC2, VM, Hortonworks Data Platform 2.1 & 2.2
* Extending Hive and Pig core functionality by writing custom UDF's for Data Analysis.
* Handling importing of data from various data source, performed transformation, and hands on developing and debugging MR2 jobs to process large data sets.
* Experienced in working with Hadoop clusters using Horton works distributions
* Experience in using Apache Flume for collecting, aggregation, moving large amount of data from application server.
* Integrated Apache Storm with Kafka to perform web analytics and to perform clickstream data from Kafka to HDFS
* Real time predictive analytics capabilities using Spark Streaming, Spark SQL.

AWS Stack

* Explored Spark for improving the performance and optimization of the existing jobs in Hadoop using Spark context
* Hands on experience using tableau to perform operations like report generation.
* Used Sqoop to take data from relational database into HDFS for processing
* Experience in creating tables on top of data on AWS S3 obtained from different data sources and providing them to analytics team building reports using Tableau.
* Knowledge of working on Hortonworks and Amazon EMR Hadoop distributors
* Used sqoop extensively for ingesting data from relational data sources.
* Strong knowledge of entire SDLCE - Requirement Gathering & Analysis, Planning, Design, Development, Testing and Implementation.
* Developed multiple POCs using Spark Scala and deployed on the Yarn cluster, compared the performance of Spark, with Hive and SQL.
* Developed efficient MapReduce programs for filtering out the unstructured data and developed multiple MapReduce jobs to perform data cleaning and preprocessing on Hortonworks
* Used NoSql technologies like Hbase, Mongo dB for data extraction and storing huge volume of data. Knowledge in job/workflow scheduling and monitoring tools like Oozie, Zookeeper.
* Experience with Oozie Workflow Engine in running workflow jobs with actions that run Hadoop Map/Reduce and Pig jobs.
* Gathered data from twitter and stored them in HDFS and putting them in Hive
* Expertise in writing Map Reduce jobs using Java native code, Pig, Hive for data Processing
* Hands on experience in installation, configuration, management and development of big data solutions using Hortonworks distributions
* Used SVN repository for version control of the developed code
* Experience working on NoSQL databases including Cassandra and Hbase
* Major strengths are familiarity with multiple software systems, ability to learn quickly new technologies, adapt to new environments, self-motivated, team player, focused adaptive and quick learner with excellent interpersonal, technical and communication skills.
* Strong oral and written communication, initiation, interpersonal learning and organizing skills matched with the ability to manage time and people effectively.

**Skills**

Hadoop Stack HDFS , Map Reduce , Apache Pig , Hbase, Hive, ZooKeeper , Chef, Kafka, YARN, Sqoop, MRUnit.   
Database: Oracle 10g/11g, Sql Server […] R2 , My SQL, DB2 , HBase , Hortonworks, MongoDB  
Framework Structs2, Spring 2.5/2.0, Hibernate 3.0.   
Operating Systems: Windows 2008, 2003, 2000 Server, Windows […] DOS, Red Hat Linux    
Database Tools: SQL Enterprise Manager, SQL Profiler, Query Analyser, SQL Server Setup, Security Manager, Service manager, SQL Server Reporting Services(SSRS)   
Programming Languages: Java , J2EE

Script Languages: JavaScript , JQuery , Shell Script(BASH)    
Methodologies: Agile/Scrum

**Work Experience:**

**Vantiv, Cincinnati May 2017 – May 2018**

**Sr. Hadoop developer**

**Responsibilities:**

* Uploaded and processed more than 40 terabytes of data from various structured and unstructured sources into HDFS using Sqoop and Custom Input Adapters.
* Developed spark applications for performing large scale transformations and denormalization of relational datasets.
* Utilized Spark-SQL and Data frame API in spark for writing custom transformations and data aggregations.
* Used Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive
* Generated various reports using tableau.
* Created data visualization to view reports with tableau.
* Written Java code by using collections, maps, file streams, threads for enhancements
* Writing the queries using Impala query engine to get the faster results
* Developed shell scripts for running Hive scripts in Hive and Impala.
* Involved in identifying job dependencies to design workflow for Oozie and resource management for YARN.
* Modeled various hive tables and optimized the access by designing partitions and bucketing.
* Developed Kafka consumer's API in Scala for consuming data from Kafka topics
* Developed efficient MapReduce programs for filtering out the unstructured data and developed multiple MapReduce jobs to perform data cleaning and preprocessing on Hortonworks
* Proactively monitored systems and services, architecture design and implementation of Hadoop deployment, configuration management, backup, and disaster recovery systems and procedures.
* Designed HBase tables for time series data. Designed row key to avoid region server hot spotting.
* Involved in Analyzing system failures, identifying root causes, and recommended course of actions.
* Documented the systems processes and procedures for future references.
* Worked with systems engineering team to plan and deploy new Hadoop environments and expand existing Hadoop clusters.
* Involved in setting up a 80 node Hadoop cluster with the help of devops team.
* Used HBase API's to get and scan events data stored in HBase.
* Used Java extensively to write custom input adapters and rest services utilized by down stream applications teams.

**Environment**: HDP 2.2 - Hadoop, Java, Hortonworks, Impala, Spark, YARN, HBase, HDFS, ZooKeeper, Java, Kafka, Spark, Shell Scripting, JSON, JUnit MySQL, JIRA, Confluence, Putty.

**Paccar, Seattle July 2015 – April 2017**

**Hadoop Developer**

**Responsibilities:**

* Responsible for building scalable distributed data solutions using Hadoop.
* Responsible for fine tuning long running hive queries.
* Responsible for Cluster maintenance, adding and removing cluster nodes, Cluster Monitoring and Troubleshooting, Manage and review data backups and log files.
* Analyzed data using Hive and Pig and written custom UDFs in Hive.
* Worked hands on with ETL process.
* Worked with application teams to install operating system, Hadoop updates, patches, version upgrades as needed.
* Integrated Kafka-Spark streaming for high-efficiency throughput and reliability
* Responsible for running Hadoop MR jobs to process terabytes of xml's data.
* Load and transform large sets of structured, semi structured and unstructured data using Hadoop /Big Data concepts.
* Populated HDFS and HBase with huge amounts of data using Apache Kafka
* Worked on installation and configuration Horton works cluster ground up
* Responsible for creating Hive tables, loading data and writing Hive queries.
* Handled importing data from various data sources, performed transformations using Hive, Map Reduce, and loaded data into HDFS.
* Implemented Data Interface to get information of customers using RestAPI and Pre-process data using MapReduce 2.0 and store into HDFS (Hortonworks)
* Used Spark API over Hortonworks Hadoop YARN to perform analytics on data in Hive
* Extracted the data from Teradata into HDFS using the Sqoop.
* Used Kafka to ingest data into Spark engine
* Exported the patterns analyzed back to Teradata using Sqoop.
* Installed Oozie workflow engine to run multiple Hive and MR jobs which run independently with time and data availability.
* Involved in running Hadoop jobs for processing millions of records of text data.

**Environment:** CDH 5.4, Apache Solr, MapReduce, Hortonworks, Kafka, Hive, Pig, YARN, Sqoop, Java, Eclipse Kepler, SVN repository, Linux, Putty, WinSCP.

**National Payments Corp, Mumbai Feb 2014 – March 2015**

**Hadoop developer**

**Responsibilities:**

* Developed Java MapReduce programs for the analysis of sample log file stored in cluster
* Load data from various data sources into HDFS using Flume.
* Developed and maintained Hive QL, Pig Latin Scripts, Scala and Map Reduce.
* Experience in administration, installing, upgrading and managing CDH4, Pig, Hive & Hbase
* Used WebSphere as an application server and used Apache Maven to deploy and build the application in WebSphere, performed unit testing using JUnit, Developed JAX-WS client and JAX-WS web services to coordinate with outer systems, involved in design of data migration strategy to migrate the data from legacy system to Kenan FX 2.0 billing system.
* Involved in designing and developing Hadoop MapReduce jobs Using JAVA Runtime Environment for the batch processing to search and match the scores.
* Worked on Hortonworks Data Platform (HDP)
* Responsible for working with Message broker system such as Kafka Extracted data from mainframes and feed to KAFKA and ingested to HBase to perform Analytics
* Datacenter infrastructure design and implementation for cluster, hardware (rack and servers), network (switches) and other (power and cooling) infrastructure related requirements
* Involved in the design of staging database as part of migration strategy
* Complete involvement in Requirement Analysis and documentation on Requirement Specification.
* Involved in design of the core implementation logic using MVC architecture
* Used SVN repository for version control of the developed code
* Mentored junior developers in their development activities
* Developed design documents for the code developed.
* Performed unit testing for the developed code using Junit

**Environment:** CDH4, Mapreduce, Oracle, Hortonworks, Kafka, Hive, Pig, HBase, Sqoop, Java, Hibernate, and spring, Maven, SVN, SQL, Junit, Eclipse, XML, Log4j.

**Easylink, Chennai Jan 2013 – Feb 2014**

**Hadoop Developer**

 **Responsibilities:**

* Involved in the Design, Coding, Testing and Implementation of the web application.
* Developed JSP Java Server Pages starting from HTMLs and detailed technical design specification documents. Pages included HTML, CSS, JavaScript, Hibernate and JSTL.
* Developed SOAP based requests for communicating with Web Services.
* Used agile systems and strategies to provide quick and feasible solutions, based on agile system, to the organization.
* Extensively used JDBC concepts to connect to, retrieve and update information in the database.
* Implemented HTTP Modules for different applications in Struts Framework that uses Servlets, JSP, Action Form, Action Class and Action Mapping.
* Developing web applications using MVC Framework, Spring, Struts, Hibernate.
* Involved in the creation of custom interceptors for Validation purposes.
* Analyzed and fixed defects in the Login application.
* Involved in configuration and deployment of application on the JBoss Application.
* Involved in dynamic creation of error elements on demand when there is an error.
* Involved in Ajax - based Rich Browser User Interfaces.
* Ensured design consistency with client's development standards and guidelines.
* Improved user experience by designing and creating new web components and features.

**Environment:** Hadoop, MapReduce, Hive, Pig, HBase, Sqoop, Flume, Oracle 11g, Core Java.

**Education: Master’s in Computer Science, Lamar university**