**Vummadi Venkata Sita Armani Rao**

**Summary:**

* Around 9 years of experience in software development, deployment and maintenance of applications of various stages.
* 4+ years of experience on major components in Hadoop Ecosystem like Hadoop Map Reduce, HDFS, HIVE, PIG, Pentaho, HBase, Zookeeper, Sqoop, Oozie, Flume, Storm, Yarn, Spark, Scala and Avro.
* Extensively worked on build tools like Maven, Log4j, Junit and Ant.
* Experience in applying the latest development approaches including applications in Spark using Scala to compare the performance of Spark with Hive and SQL/Oracle.
* Thorough knowledge with the data extraction, transformation and load in Hive, Pig and HBase
* Hands on experience in coding Map Reduce/Yarn Programs using Java, Scala for analyzing Big data.
* Worked with Apache Spark which provides fast and general engine for large data processing integrated with functional programming language Scala.
* Good understanding knowledge in installing and maintaining Cassandra by configuring the Cassandra. Yaml file as per the requirement and performed reads and writes using Java JDBC connectivity.
* Hands on experience in writing Pig Latin scripts, working with grunt shells and job scheduling with Oozie.
* Experience in designing and implementing of secure Hadoop cluster using Kerberos.
* Processing this data using Spark Streaming API with Scala.
* Good exposure to MongoDB, its functionality and Cassandra implementation.
* Have a good experience working in Agile development environment including Scrum methodology.
* Good Knowledge on Spark framework on both batch and real-time data processing.
* Hands on experience in MLlib from Spark are used for predictive intelligence, customer segmentation and for smooth maintenance in Spark streaming.
* Expertise in Storm for reliable real-time data processing capabilities to Enterprise Hadoop.
* Hands on experience in scripting for automation, and monitoring using Shell, PHP, Python & Perl scripts.
* Experience data processing like collecting, aggregating, moving from various sources using Apache Flume and Kafka.
* Extensive Experience on importing and exporting data using Flume and Kafka.
* Experience in migrating the data using Sqoop from HDFS to Relational Database System and vice-versa according to client's requirement.
* Experienced in deployment of Hadoop Cluster using Puppet tool.
* Hands on experience in ETL, Data Integration and Migration and Extensively used ETL methodology for supporting Data Extraction, transformations and loading using Informatica.
* Good knowledge in Cluster coordination services through Zookeeper and Kafka.
* Excellent knowledge in existing Pig Latin script migrating into Java Spark code.
* Experience in transferring data between a Hadoop ecosystem and structured data storage in a RDBMS such as MY SQL, Oracle, Teradata and DB2 using Sqoop.
* Strong knowledge in Upgrading Mapr, CDH and HDP Cluster.
* Hands on Experience in setting up automated monitoring and escalation infrastructure for Hadoop Cluster using Ganglia and Nagios.
* Experience in developing and designing POCs using Scala and deployed on the Yarn cluster, compared the performance of Spark, with Hive and SQL/Teradata.
* Good understanding knowledge in in MPP databases such as HP Vertica and Impala.
* I have been experience with AWS, AZURE, EMR and S3.
* Involved in developing Impala scripts for extraction, transformation, loading of data into data warehouse.
* I have been experience with Chef, Puppet, and Ansible.
* Extending HIVE and PIG core functionality by using custom User Defined Function's (UDF), User Defined Table-Generating Functions (UDTF) and User Defined Aggregating Functions (UDAF) for Hive and Pig.
* Experience working on various Cloudera distributions like (CDH 4/CDH 5), Knowledge of working on Horton works and Amazon EMR Hadoop distributors.
* Experience in Apache NIFI which is a Hadoop technology and also Integrating Apache NIFI and Apache Kafka.
* Worked on version control tools like CVS, GIT, and SVN.
* Experience in Web Services using XML, HTML, and SOAP.
* Good experience in working with cloud environment like Amazon Web Services (AWS) EC2 and S3.
* Experience in developing web pages using Java, JSP, Servlets, JavaScript, JQuery, Angular JS, Node Js,Mobile JQuery, JBOSS 4.2.3, XML, Web Logic, SQL, PL/SQL, JUnit, and Apache-Tomcat, Linux.
* Worked on version control tools like CVS, GIT, and SVN.
* Experienced in collecting metrics for Hadoop clusters using Ambari & Cloudera Manager.
* Expertise in implementing and maintaining an Apache Tomcat /MySQL/PHP, LDAP, LAMP web service environment.
* Worked with BI (Business Intelligence) teams in generating the reports and designing ETL workflows on Tableau. Deployed data from various sources into HDFS and building reports using Tableau.
* Self-starter always inclined to learn new technologies and Team Player with very good communication, organizational and interpersonal skills.
* Experience in all phases of Software development life cycle (SDLC).

**Education:**

* Bachelor’s in computer Science from Osmania University, Hyd in 2008.

**Technical skills:**

**Big Data Ecosystem:** Hadoop, MapReduce, Pig, Hive, YARN, Kafka, Flume, Sqoop, Impala, Oozie, Zookeeper, Spark, Ambari, Mahout, MongoDB, Cassandra, Avro, Storm, Parquet and Snappy.

**Hadoop Distributions:** Cloudera (CDH3, CDH4, and CDH5), Hortonworks, MapR and Apache

**Languages:**  Java, Python, Jruby, SQL, HTML, DHTML, Scala, JavaScript, XML and C/C++

**No SQL Databases:** Cassandra, MongoDB and HBase

**Java Technologies:**  Servlets, JavaBeans, JSP, JDBC, JNDI, EJB and struts

**XML Technologies:** XML, XSD, DTD, JAXP (SAX, DOM), JAXB

**Methodology:** Agile, waterfall

**Web Design Tools:** HTML, DHTML, AJAX, JavaScript, JQuery and CSS, AngularJs, ExtJS and JSON, NodeJs.

**Development / Build Tools:** Eclipse, Ant, Maven, IntelliJ, JUNIT and log4J.

**Frameworks:** Struts, spring and hibernate

**App/Web servers:** WebSphere, WebLogic, JBoss and Tomcat

**DB Languages:** MySQL, PL/SQL, PostgreSQL and Oracle RDBMS Teradata, Oracle 9i, 10g, 11i, MS SQL Server, MySQL and DB2

**Operating systems:** UNIX, LINUX, Mac os and Windows Variants

**Data analytical tools:** R and MATLAB

**ETL Tools:** Talend, Informatica, Pentaho

**Professional Experience:**

**Miracle software systems, Michigan July 2016 – Till Date**

**Sr. Hadoop/Spark Developer**

**Responsibilities:**

* Experienced in designing and deployment of Hadoop cluster and different Big Data analytic tools including Pig, Hive, Cassandra, Oozie, Sqoop, Kafka, Spark, Impala with Horton works distribution
* Performed source data transformations using Hive.
* Supporting infrastructure environment comprising of RHEL and Solaris.
* Involved in developing a Map Reduce framework that filters bad and unnecessary records.
* Developed Spark scripts by using Scala shell commands as per the requirement.
* Used Kafka to transfer data from different data systems to HDFS.
* Created Spark jobs to see trends in data usage by users.
* Responsible for generating actionable insights from complex data to drive real business results for various application teams.
* Designed the Column families in Cassandra.
* Ingested data from RDBMS and performed data transformations, and then export the transformed data to Cassandra as per the business requirement.
* Developed Spark code to using Scala and Spark-SQL for faster processing and testing.
* Experience in NoSQL Column-Oriented Databases like Cassandra and its Integration with Hadoop cluster.
* Used Spark API over Hadoop YARN as execution engine for data analytics using Hive.
* Exported the analyzed data to the relational databases using Sqoop to further visualize and generate reports for the BI team.
* Good experience with Talend open studio for designing ETL Jobs for Processing of data.
* Experience in processing large volume of data and skills in parallel execution of process using Talend functionality
* Worked on different file formats like Text files and Avro.
* Created various kinds of reports using Power BI and Tableau based on the client's needs.
* Worked on Agile Methodology projects extensively.
* NIFI is designed to pull data from various sources and push it in HDFS and Cassandra.
* Experience designing and executing time driven and data driven Oozie workflows.
* Setting up Kerberos principals and testing HDFS, Hive, Pig, and MapReduce access for the new users.
* Experienced in working with Spark eco system using SCALA and HIVE Queries on different data formats like Text file and parquet.
* Log4j framework has been used for logging debug, info & error data.
* Collected the logs data from web servers and integrated in to HDFS using Flume
* Developed Hive scripts in Hive QL to de-normalize and aggregate the data.
* Implemented map-reduce counters to gather metrics of good records and bad records.
* Work experience with cloud infrastructure like Amazon Web Services (AWS).
* Developed customized UDF's in java to extend Hive and Pig functionality.
* Worked with SCRUM team in delivering agreed user stories on time for every sprint.
* Experience in importing data from various data sources like Mainframes, Teradata, Oracle and Netezza using Sqoop, SFTP, performed transformations using Hive, Pig and Spark and loaded data into HDFS.
* Extracted the data from Teradata into HDFS/Databases/Dashboards using SPARK STREAMING.
* Implemented best income logic using Pig scripts.
* Worked on different file formats (ORCFILE, Parquet, Avro) and different Compression Codecs (GZIP, SNAPPY, LZO).
* Created applications using Kafka, which monitors consumer lag within Apache Kafka clusters. Used in production by multiple companies.
* Using Spark-Streaming APIs to perform transformations and actions on the fly for building the common learner data model which gets the data from Kafka in near real time and Persists into Cassandra.
* Design and document REST/HTTP, SOAP APIs, including JSON data formats and API versioning strategy.
* Experience in using Apache Kafka for collecting, aggregating, and moving large amounts of data from application servers.
* Used Hibernate ORM framework with Spring framework for data persistence and transaction management.
* Performance analysis of Spark streaming and batch jobs by using Spark tuning parameters.
* Used React Bindings for embracing Redux.
* Worked towards creating real time data streaming solutions using Apache Spark/Spark Streaming, Kafka.
* Worked on Express view engine which renders React components on server.
* Worked along with the Hadoop Operations team in Hadoop cluster planning, installation, maintenance, monitoring and upgrades.
* Used File System check (FSCK) to check the health of files in HDFS.
* Worked in Agile development environment in sprint cycles of two weeks by dividing and organizing tasks. Participated in daily scrum and other design related meetings.

**Environment:** Hadoop, Hive, Map Reduce, Sqoop, Kafka, Spark, Yarn, Pig, Cassandra, Oozie, shell Scripting, Scala, Maven, Java, React JS, JUnit, agile methodologies, Horton works, Soap, NIFI, Teradata, MySQL.

**Onset Computer Corporation, MA Oct 2014 - June 2016**

**Spark/Hadoop Developer**

**Responsibilities:**

* Experience in developing customized UDF's in java to extend Hive and Pig Latin functionality.
* Responsible for installing, configuring, supporting, and managing of Hadoop Clusters.
* Importing and exporting data into HDFS from Oracle 10.2 database and vice versa using SQOOP.
* Installed and configured Pig and written Pig Latin scripts.
* Designed and implemented HIVE queries and functions for evaluation, filtering, loading and storing of data.
* Created HBase tables and column families to store the user event data.
* Written automated HBase test cases for data quality checks using HBase command line tools.
* Developed a data pipeline using HBase, Spark and Hive to ingest, transform and analyzing customer behavioral data.
* Experience in collecting the log data from different sources like (webservers and social media) using Flume and storing on HDFS to perform MapReduce jobs.
* Handled importing of data from machine logs using Flume.
* Created Hive Tables, loaded data from Teradata using Sqoop.
* Worked extensively with importing metadata into Hive and migrated existing tables and applications to work on Hive and AWS cloud.
* Configured, monitored, and optimized Flume agent to capture web logs from the VPN server to be put into Hadoop Data Lake.
* Responsible for loading data from UNIX file systems to HDFS. Installed and configured Hive and written Pig/Hive UDFs.
* Wrote, tested and implemented Teradata Fast load, Multiload and Bteq scripts, DML and DDL.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDD, Scala and Python.
* Ec2 Collecting and aggregating large amounts of log data using Flume and staging data in HDFS for further analysis
* Exported the analyzed data to the relational databases using Sqoop to further visualize and generate reports for the BI team.
* Develop ETL Process using SPARK, SCALA, HIVE and HBASE.
* Collecting and aggregating large amounts of log data using Flume and staging data in HDFS for further analysis.
* Wrote Java code to format XML documents; upload them to Solr server for indexing.
* Used with NoSQL technology (Amazon Dynodb) to gather and track event-based metric.
* Maintenance of all the services in Hadoop ecosystem using ZOOKEPER.
* Worked on implementing Spark frame work.
* Designed and implemented Spark jobs to support distributed data processing.
* Expertise in Extraction, Transformation, loading data from Oracle, DB2, SQL Server, MS Access, Excel, Flat Files and XML using Talend.
* Experienced on loading and transforming of large sets of structured, semi and unstructured data.
* Help design of scalable Big Data clusters and solutions.
* Followed agile methodology for the entire project.
* Experience in working with Hadoop clusters using Cloudera distributions.
* Involved in Hadoop cluster task like Adding and Removing Nodes without any effect to running jobs and data.
* Developed workflows using Oozie to automate the tasks of loading the data into HDFS and pre-processing with Pig.
* Developed interactive shell scripts for scheduling various data cleansing and data loading process.
* Converting the existing relational database model to Hadoop ecosystem.

**Environment:** Hadoop, HDFS, Pig, Hive, Flume, Sqoop, Oozie, Python, Shell Scripting, SQL Talend, Spark, HBase, Elastic search, Linux- Ubuntu, Flume, Cloudera.

**Tyro labs, San Francisco Jan 2013 - Sept 2014**

**Hadoop Developer**

**Responsibilities:**

* Extensively involved in Installation and configuration of Cloudera distribution Hadoop, Name Node, Secondary Name Node, Job Tracker, Task Trackers, and Data Nodes.
* Developed MapReduce programs in Java and Sqoop the data from ORACLE database.
* Responsible for building scalable distributed data solutions using Hadoop. Written various Hive and Pig scripts.
* Moved data from HDFS to Cassandra using Map Reduce and Bulk Output Format class.
* Experienced with different scripting language like Python and shell scripts.
* Developed various Python scripts to find vulnerabilities with SQL Queries by doing SQL injection, permission checks and performance analysis.
* Installed Oozie workflow engine to run multiple Hive and Pig jobs which run independently with time and data availability.
* Experienced with handling administration activations using Cloudera manager.
* Expertise in understanding Partitions, Bucketing concepts in Hive.
* Used Oozie Scheduler system to automate the pipeline workflow and orchestrate the Map Reduces jobs that extract the data on a timely manner. Responsible for loading data from UNIX file system to HDFS.
* Developed suit of Unit Test Cases for Mapper, Reducer and Driver classes using MR Testing library.
* Analyzed the weblog data using the HiveQL, integrated Oozie with the rest of the Hadoop stack
* Utilized cluster co-ordination services through Zookeeper.
* Worked on the Ingestion of Files into HDFS from remote systems using MFT.
* Got good experience with various NoSQL databases and Comprehensive knowledge in process improvement, normalization/de-normalization, data extraction, data cleansing, and data manipulation.
* Experience with creating script for data modeling and data import and export. Extensive experience in deploying, managing and developing MongoDB clusters.
* Developed Pig scripts to convert the data from Text file to Avro format.
* Created Partitioned Hive tables and worked on them using HiveQL.
* Developed Shell scripts to automate routine DBA tasks.
* Used Maven extensively for building jar files of MapReduce programs and deployed to Cluster.
* Responsible for cluster maintenance, adding and removing cluster nodes, cluster monitoring and troubleshooting, managing and reviewing data backups and Hadoop log files.

**Environment:** HDFS, Python, Map Reduce, Pig, Hive, Oozie, Sqoop, Flume, HBase, Java, Maven, Avro, Cloudera, Eclipse and Shell Scripting.

**Shivam med iSOFT software – Hyderabad, India Oct 2010 - Nov 2012**

**Java Developer**

**Responsibilities:**

* Interact and coordinate with team members to develop detailed software requirements that will drive the design, implementation, and testing of the Consolidated Software application.
* Implemented the object-oriented programming concepts for validating the columns of the import file.
* Integrated Spring Dependency Injection (IOC) among different layers of an application.
* Designed the Database, written triggers, and stored procedures.
* Developed PL/SQL View function in Oracle 9i database for get available date module.
* Used Quartz schedulers to run the jobs in a sequential with in the given time
* Used JSP and JSTL Tag Libraries for developing User Interface components
* Implemented the online application using Core Java, JDBC, JSP, Servlets, spring, Hibernate, Web Services, SOAP, and WSD.
* Responsible for Checking in the code using the Rational Rose clear case explorer.
* Used Core Java concepts such as multi-threading, collections, garbage collection and other JEE technologies during development phase and used different design patterns.
* Created continuous integration builds using Maven and SVN control.
* Used Eclipse Integrated Development Environment (IDE) in entire project development.
* Responsible for Effort estimation and timely production deliveries.
* Written deployment scripts to deploy application at client site.
* Involved in design, analysis, and architectural meetings.
* Created the stored procedures using Oracle database and accessed through Java JDBC
* Configured log4j to log the warning and error messages.
* Implemented the reports module applications using jasper reports for business intelligence
* Supported Testing Teams and involved in defect meetings.
* Deployed web, presentation, and business components on Apache Tomcat Application Server.

**Environment:** HTML, Java Script, Ajax, Servlets, JSP, JavaScript, CSS, XML, ANT, Tomcat Server, Soap, Jasper Reports.

**Magna InfoTech, Hyderabad, India Nov 2008 - Sept 2010**

**Java/Etl Developer**

**Responsibilities:**

* Understanding business objectives and implementation of business logic
* Designed front end using JSP and business logic in Servlets.
* Used JSPs, HTML and CSS to develop user interface.
* Responsible for design and build data mart as per the requirements.
* Created complex mappings in Power Center Designer using Aggregate, Expression, Filter, and Sequence Generator, Update Strategy, Union, Lookup, Joiner, XML Source Qualifier, and Stored procedure transformations.
* Co-coordinating the UAT and Production migration for Informatica objects with Users and other business stake holders.
* Extensively used Oracle ETL process for address data cleansing.
* Automated new FNI Blades source file loads by creating functional design documents, Informatica mappings, sessions and workflows.
* Involved in technical design, logical data modeling, data validation, verification, data cleansing, data scrubbing.
* Created Rulesets for data quality index reports.
* Extensively worked on Views, Stored Procedures, Triggers and SQL queries and for loading the data (staging) to enhance and maintain the existing functionality.
* Involved in creating error logs and increase the performance of the jobs.
* Written queries to test the functionality of the code during testing.
* Develop Logical and Physical data models that capture current state/future state data elements and data flows using Erwin / Star Schema.
* Worked on production support tickets and the resolutions for high, medium and low priority incidents through Remedy incident system.
* Used debugger to debug mappings to gain troubleshooting information about data and error conditions.

**Environment:** Java, J2EE, JDBC, Servlets, EJB, JSP, Struts, HTML, CSS, Java Script, UML, Jboss Application Server 4.2, MySQL, Linux, and CVS. Informatica 8.6.1 Oracle 11g (TOAD and SQL Developer), Cognos & Tableau, UNIX, MS ACCESS, MS EXCEL 2007, Autosys.