**VenkatRoyalu**

[**Venkat.royal1812@gmail.com**](mailto:Venkat.royal1812@gmail.com)

**224-788-0898**

Hadoop Developer

**SUMMARY:**

* **7+ years** of strong experience in software development.
* **5+ year**s of experience using **BigData Ecosystems & Java.**
* Experience in **Apache Spark with Scala and Apache Solr**.
* Solid Mathematics, Probability and Statistics foundation and broad practical statistical and data mining techniques cultivated through various industry work and academic programs
* Involved in the Software Development Life Cycle **(SDLC)**phaseswhich include **Analysis, Design, Implementation, Testing** and **Maintenance.**
* Strong **technical**, **administration**, &**mentoring** knowledge in **Linux** and **BigData/Hadoop** technologies.
* Have sound knowledge on **In-Memory MEMSQL**.
* Hands on experience on major components in **Hadoop Ecosystem** like **Hadoop Map Reduce**, **HDFS, HIVE, PIG, Pentaho, Hbase, Zookeeper, Sqoop, Oozie, Cassandra, Flume and Avro**.
* Experienced the deployment of **Hadoop Cluster using Puppet tool**
* Work experience with cloud infrastructure like **Amazon Web Services (AWS).**
* Experience in importing and exporting the data using **SQOOP** from **HDFS** to **Relational Database systems/mainframe** and vice-versa
* Experienced in defining data quality, data modeling, data warehouse, **Star/Snowflake Schema Designs** with requirements analysis - definition, database design, testing and implementation, and Quality process.
* In-depth knowledge and understanding of **dimensional modeling (Star schema and Snowflake schema, SCD types- 1, 2, 3), and data modeling (star schema)** at logical and physical level
* Installing, configuring and managing of **Hadoop Clusters** and Data Science tools.
* Managing the **Hadoop** distribution with **Cloudera Manager**, **Cloudera Navigator**, **Hue**.
* Experience with modern data processing technology stack including Spark, EMR, Airflow, Python and Snowflake within AWS.
* Having good knowledge in front end technologies like HTML4, CSS3, Android, AJAX, JSON, DOM, TypeScript, Bootstrap, XHTML, DHTML, JavaScript, jQuery, node JS, Tag libraries, Angular JS 2/4.
* Knowledge on various datawarehouse BI tools such as Pentaho, Looker, Tableau.
* Setting up the High-Availability for **Hadoop Clusters** components and **Edge nodes.**
* Experience in developing **Shell scripts** and **Python Scripts** for system management.
* Well versed in using Software development methodologies like **Rapid Application Development (RAD), Agile Methodology** and **Scrum** software development processes.
* Expert in understanding the data and designing/Implementing the enterprise platforms like **Hadoop Data lake** and **Huge Data warehouses**
* Experience in building and publishing interactive reports and dashboards with design customizations based on the client requirements in Tableau, Looker, PowerBI and SSRS.
* Used PIG Latin Scripts, join operations, custom user defined functions (UDF) to perform ETL operations
* Experience with **Object Oriented Analysis**and**Design (OOAD)**methodologies.
* Experience in installations of software, writing test cases, debugging, and testing of batch and online systems.
* Experience in Production, **quality assurance (QA), SIT (System Integration testing) and user acceptance (UA)** testing.
* Expertise in J2EEtechnologies like **JSP, Servlets, EJBs 2.0, JDBC, JNDI and AJAX.**
* Extensively worked on implementing **SOA (Service Oriented Architecture)**using**XMLWeb services** (**SOAP**, **WSDL**, **UDDI** and **XML Parsers**).
* Worked with XML parserslike**JAXP (SAX and DOM)**and**JAXB.**
* Expertise in applying **Java Messaging Service (JMS)**for reliable information exchange across Java applications.
* Proficient with **Core JAVA , AWT** and also with the markup languages like**HTML 5.0,XHTML,**

**DHTML, CSS, XML 1.1, XSL, XSLT, XPath, XQuery, Angular.js, Node.js**

* Expert in understanding the data and designing/Implementing the enterprise platforms like Hadoop Data lake and Huge Data warehouses.
* Worked with **version control systems** like **Subversion, Perforce, and GIT**for providing common platform for all the developers.
* Articulate in written and verbal communication along with strong interpersonal, analytical, and organizational skills.
* Hands on experience with Microsoft Azure Cloud services, Storage Accounts and Virtual Networks.
* Good at manage hosting plans for Azure Infrastructure, implementing and deploying workloads on Azure virtual machines (VMs).
* Highly motivated team player with the ability to work independently and adapt quickly to new and emerging technologies.
* Creatively communicate and present models to business customers and executives, utilizing a variety of formats and visualization methodologies.

**EDUCATION:**Bachelor Degree in Computer Science.

**Proof of Concepts:** Demonstrated real-time analytics using Apache Kafka and Apache Nifi

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| Languages | C,C++,Java 6,Java 7, Java 8, Scala. |
| Big Data Skills | Map reduce, Hadoop,Spark,Kafka,Storm |
| Web Technologies | HTML5 JavaScript, Ajax, CSS, JQuery, XML,BootStrap. |
| Servers | WebSphere, Tomcat 6.x,MIIS (Microsoft Internet Information Server) |
| Case Tools and IDE | Eclipse,NetBeans,RAD,IntelliJ,Netezza. |
| Frameworks in Hadoop | Spark, Kafka, Storm |
| Databases | DB2, Oracle and MySQL Server |
| Version Tools | SVN, CVS, ClearCases |
| Web Services | SOAP, REST |
| Certifications | Sun Certified Java Programmer. |

**WORK EXPERIENCE:**

**Capital One, Richmond, VA March 2017 – Till Date**

**Sr. Hadoop Developer**

**Responsibilities:**

* Migration of Oracle tables to the HDFS using SQOOP.
* Designed and developed rich front end screens using JSF (Ice faces), JSP, Docker, CSS, HTML ,Angular JS and JQuery.
* Developed Managed beans and defined Navigation rules for the application using JSF.
* Developed Angular JS 2.0 code and migrated pre-existing code to updated Angular JS 2.0 framework. Written a custom Sqoop class for the XMLTYPE datatype in the Oracle Databse
* Used Java Messaging Services (JMS) for reliable and asynchronous exchange of important information such as payment status report to MQ Server using MQ Series.
* Designed and implemented Apache Spark job which takes the Sequence File from th HDFS and migrate to the Hbase.
* Designing & Creating ETL Jobs through Talend to load huge volumes of data into Cassandra, Hadoop Ecosystem and relational databases.
* Built NodeJSnpm packages/libraries to support current and future Node Development.
* Created API using nodeJS for internal company use.
* Implemented modules into Node.js tointegrate with designs and requirements.
* Design and Develop Hadoop ETL solutions to move data to the data lake using big data tools like Sqoop, Hive, Spark, HDFS, Talend etc.
* Involved and guide the team for the preparing the technical specification
* Involved in Development,Build and Deployment Application
* Implemented micro services in order to separate the tasks and not to have dependency on other Parallel ongoing tasks of same Application.
* Converted existing Snowflake schema data into Star schema in hive for building OLAP cubes in Kylin
* Used Impala to read, write and query the Hadoop data in HDFS from HBase or Cassandra and configured Kafka to read and write messages from external programs.
* Used Pig as ETL tool to do transformations, event joins, filter both traffic and some pre-aggregations before storing the data onto HDFS
* Created Apache Spark RDD transformations on Data sets in the Hadoop data lake.
* Identifying opportunities to improve infrastructure that effectively and efficiently utilizes the Microsoft Azure Windows server 2008/2012/R2, Microsoft SQL Server, Microsoft Visual Studio, Windows PowerShell, Cloud infrastructure.
* Deployed Azure IaaS virtual machines (VMs) and Cloud services (PaaS role instances) into secure VNets and subnets.
* Developed Web service using Restful with Jersey, and implemented JAX-RS and also provided security-using SSL.
* Creation Of**HBase** Tables and implemented Salting on the Hbase.
* Migration of tables from oracle to Hbase on the Tenant basis.
* Used Pig as ETL tool to do Transformations, even joins and some pre-aggregations before storing the data onto HDFS
* One time migraion of 300 Billions of records using apache Spark BulkLoading.
* Involved in Delta Migration using Sqoop Incremental Updates.
* Worked in designing starschema, Snowflakeschema for Data Warehouse, ODSarchitecture.
* Identify the Snowflake aggregate tables to reverse load the data from Snowflake to Teradata environment.
* Data modelling into Snowflake environment for creation of new objects.
* Maintained, structured, and surveyed documents within the NoSQL MongoDB database; ensuring data integrity, correcting anomalies, and increasing the overall maintainability of the database.
* Reviewing ETL application use cases before on boarding to Hadoop
* Perform ETL on data from Dev and QA box for staging it for micro batching to Spark
* Involved in architecturing the Data Pipe-Line for the Add/Update flow of real-time analysis flow for the IMS Application
* Achitectured the whole design for the data migration process and also for the real time analysis.
* Written Apache Spark Jobs using **Scala** API.
* Involved in administrating and configuring the mapr distribution.
* Implemented **Mapr Streams** (KAFKA 0.9 API) to the Spark Streaming using Java API.

**Environment:** Apache Spark, Spark Streaming, Spark SQL, ETL, data Lake, Hadoop Security,Mapr Streams, Node.js, Mapr 5.1,Open-Shift,Scala,Java,Hbase, Snowflake, Eclipse,MVN,Sequence Files, Tableu, Looker

**BBVA Compass, Brimingham, AL July 2015 –Feb 2017**

**Hadoop Developer**

**Responsibilities:**

* Performed hands-on data manipulation, transformation, hypothesis testing and predictive modelingDeveloped robust set of codes that are tested, automated, structured and efficient
* Defined service layer using EJB3.0 and also defined remote and local services.
* Accessed remote and local EJB services from controller.
* Developed application using JSP, Tag libraries, JSF and Struts (MVC) Framework.
* Exposed web services to client developing WSDL also involved in developing web client for application interactions.
* Evaluate, refine, and continuously improve the efficiency and accuracy of existing Predictive Models using Netezza.
* Involved in developing several Data Pipelines using Apache Kafka.
* Developed Framework API for Tax calculations in Yoda using server-side components using J2EE and spring framework.
* Designed, developed and implemented a messaging module using Java Messaging Service (JMS) point-to-point messaging and Message Driven Beans to listen to the messages in the queue for interactions with client ordering data. Collaborated on insights with other Data Scientists, Business Analysts, and Partners.
* Evaluate, refine, and continuously improve the efficiency and accuracy of existing Predictive Models.
* Utilized various data analysis and data visualization tools to accomplish data analysis, report design and report delivery.
* Used Java Messaging Services (JMS) for reliable and asynchronous exchange of important information such as payment status report to MQ Server using MQ Series.
* Used NodeJS architectural concepts like event loop, call backs and event emitters
* Used Pig as ETL tool to do transformations, event joins and some pre-aggregations before storing the data onto HDFS, Tera data
* Importing and exporting the data from HDFS to RDBMS using Sqoop and Kafka.
* Worked on creating Custom Azure Templates for quick deployments and advanced PowerShell scripting.
* Contribute to the support forums (specific to Azure Networking, Azure Virtual Machines, Azure Active Directory, Azure Storage) for Microsoft Developers Network.
* Populated HDFS and Cassandra with huge amounts of data using ApacheKafka.
* Developed Scala and SQL code to extract data from various databases
* Champion new innovative ideas around the Data Science and Advanced Analytics practices
* Creatively communicated and presented models to business customers and executives, utilizing a variety of formats and visualization methodologies.
* Designed, developed and implemented a messaging module using Java Messaging Service (JMS) point-to-point messaging and Message Driven Beans to listen to the messages in the queue for interactions with client ordering data.
* Uploaded data to Hadoop Hive and combined new tables with existing databases.
* Developed statistical models to forecast inventory and procurement cycles.8
* Developed Python code to provide data analysis and generate complex data report.
* Deployed the Cassandra cluster in cloud (Amazon AWS) environment with scalable nodes as per the business requirement.
* Implemented the data backup strategies for the data in the Cassandra cluster.
* Generated the data cubes using hive, Pig, JAVA Map-Reducing on provisioning Hadoop cluster in AWS.
* Implemented the ETL design to dump the Map-Reduce data cubes to Cassandra cluster.
* Imported the data from relational databases into HDFS using SQOOP.
* Implemented POC for using APACHE IMPALA for data processing on top of HIVE.
* Utilized Python Panda Frame to provide data analysis.
* Worked on Hortonworks 2.3distribution
* Utilized Python regular expressions operation (NLP) to analysis customer review.
* Understanding of data storage and retrieval techniques, ETL, and databases, to include graph stores, relational databases, tuple stores, NOSQL, Hadoop, PIG,MySQLandOracle databases

**Environment**: Apache Spark, Pyspark, Spark Streaming, Spark SQL,Apache Kafka, Apache Flume, Python Pandas, ETL, data Lake, Cassandra, Hortonworks (HDP) 2.3,AWS,Scala,AKKA,Hive,PIG,Big Data.

**VColkswagan, Ann Arbor MI October 2014 - June 2015**

**Hadoop Developer-Java**

**Responsibilities:**

* Involved in architecture design, development and implementation of **Hadoop** deployment, backup and recovery systems.
* Experience in working on **multi-Petabyte** clusters both administration and development.
* Developed **Chef**modules to automate the installation, configuration and deployment of ecosystem tools, OS's and network infrastructure at a cluster level.
* Performed advanced procedures like text analytics and processing, using the in-memory computing capabilities of **Spark** using **Scala**.
* Performed cluster co-ordination and assisted with data capacity planning and node forecasting using **Zookeeper**.
* Developed Fault-tolerant data warehouse cluster by using **Amazon S3** and monitoring we nodes and automatic replication was developed by using **Amazon Redshift**
* Involved in performance Tuning of Hadoop clusters
* Implemented **Hadoop** framework to capture user navigation across the application to validate the user interface and provide analytic feedback/result to the UI team.
* Executed custom interceptors for **Flume** to filter data and defined channel selectors to multiplex the data into different sinks.
* Data warehouse, Business Intelligence architecture design and develop. Designed the ETL process from various sources in to Hadoop/HDFS for analysis and further processing
* Extracted data from **Oracle SQL server** and **MySQL** databases to **HDFS** using **Sqoop**.
* Optimized **MapReduce** jobs to use **HDFS** efficiently by using **Gzip**, **LZO**, **Snappy** compression techniques.
* Experience in writing **Pig** scripts to transform raw data from several data sources into forming baseline data.
* Created **Hive** tables to store the processed results in a tabular format and written **Hive** scripts to transform and aggregate the disparate data.
* Improve search results using SOLR and customize **lucence/Solr code**.
* Involved on **Lucence, Solr** and lead the index and search related development work.
* Query huge data quickly near real time while feeding log file into **solr**.
* Worked with the team to improve and ranking of search results using **SOLR**
* Experience in using **Avro, Parquet, RCFile and JSON** file formats and developed **UDFs** using **Hive** and **Pig**.
* Responsible for cluster maintenance, rebalancing blocks, commissioning and decommissioning of nodes, monitoring and troubleshooting, manage and review data backups and log files.
* Driving the application from development phase to production phase using Continuous Integration and Continuous Deployment (CICD) model using **Chef**, **Maven** and **Jenkins**.
* Develop **Pentaho Kettle** Graphs to cleanse and transform the raw data into useful information and load it to a **Kafka** Queue (further loaded to **HDFS**) and **Neo4j database** for UI team to display it using the Web application.
* Automated the process for extraction of data from warehouses and weblogs into **HIVE** tables by developing workflows and coordinator jobs in **Oozie**.
* Scheduled snapshots of volumes for backup to find root cause analysis of failures and document bugs and fixes for downtimes and maintenance of cluster.
* Tune/Modify SQL for batch and online processes.
* Commissioning and decommissioning the nodes.
* Manage cluster through performance tuning and enhancement.

**Environment**: Hortonworks (HDP 2.2), HDFS, Batch-Processing,MapReduce, Apache Cassandra,ApacheSolr, YARN, Spark, Scala,Hive, Pig, Flume, Sqoop, Chef,Puppet,Python, Oozie, ZooKeeper, Ambari, Oracle Database, MySQL, HBase, SparkSQL, AWS Redshift, Avro, Parquet, RCFile, JSON, UDF, Java (jdk1.7), Multi-Threading, Performance Tuning, CentOS

**Vanguard Group, Malvern, PA December 2012 – September 2014**

**Big Data Engineer**

**Responsibilities:**

* Worked with the business users to gather, define business requirements and analyze the possible technical solutions.
* **Hadoop** installation, configuration of multiple nodes in **Amazon EMR** platform.
* Setup and optimize Standalone-System/Pseudo-Distributed/Distributed Clusters.
* Developed Simple to complex **Map/reduce** streaming jobs
* Analyzing data with **Hive**, **Pig** and **Hadoop** Streaming.
* Build/Tune/Maintain **Hive QL** and **Pig** Scripts for reporting purpose.
* Handled importing of data from various data sources, performed transformations using **Hive**,**Map/Reduce**, loaded data into **HDFS** and Extracted the data from **MySQL** into **HDFS** using **Sqoop**
* **Processed** Data across a Hadoop cluster of virtual servers on the Amazon Elastic Computer Cloud(EC2) by using **Amazon EMR**
* Analyzed the data by performing Hive queries (**HiveQL**) and running **Pig** scripts (**Pig Latin**) to study customer behavior.
* Stored the data in an Apache Cassandra Cluster
* Used Impala to query the Hadoop data stored in **HDFS**.
* Manage and review **Hadoop** log files.
* Support/Troubleshoot **Map/Reduce** programs running on the cluster
* Load data from **Linux** file system into **HDFS**.
* Install and configure **Hive** and write **Hive** UDFs.
* Create tables, load data, and write queries in **Hive**.
* Develop scripts to automate routine DBA tasks using **Linux Shell Scripts, Python**

**Environment: Hadoop 0.20.2 - PIG, Hive, JAVA,AWS, AWS EMR,Cloudera manager, 30 Node cluster with Linux-Ubuntu**

**Mutual of Omaha, NE June 2012 – November 2012**

**Jr. Java Developer**

**Responsibilities:**

* Worked with the business users to gather, define business requirements and analyze the possible technical solutions.
* Developed the Controllers, Service Layer, Dao layer using Spring MVC, Spring JDBC.
* Performed unit testing Using JUnit and JUnit annotations.
* Implemented Restful Web service.
* Configured the Transaction Management for the project using Spring Container Managed Transactions.
* Developed web interface to display the customer information from the database tables
* Created HTML, CSS, JavaScript, DHTML pages for Presentation Layer.
* Experience in doing validation on the UI from one screen to other using JavaScript.
* Used Annotations for JSR 303, spring. The complete application was developed using annotations.
* Used UNIX and Linux commands for debugging.
* Involved in resolving SSO login issue.
* Involved in Daily Stand Up Meetings, Sprint Planning and Backlog Grooming for Agile Scrum Process.

**Oracle Financial Services Software, Hyderabad,India July 2010– May2012**

**Jr. Java/J2EE Developer**.

**Responsibilities:**

* Developed User Interfaces module using JSP, Java Script, DHTML and form beans for presentation layer.
* Developed Servlets and Java Server Pages (JSP).
* Developed PL/SQLqueries, and wrote stored procedures and JDBC routines to generate reports based on client requirements.
* Enhancement of the System according to the customer requirements.
* Involved in the customization of the available functionalities of the software for an NBFC (Non-Banking Financial Company).
* Involved in putting proper review processes and documentation for functionality development.
* Providing support and guidance for Production and Implementation Issues.
* Used Java Script validation inJSP.
* Used Hibernate framework to access the data from back-end SQL Server database.
* Used AJAX (Asynchronous JavaScript and XML) to implement user friendly and efficient client interface.
* Used MDB for consuming messages from JMS queue/topic.
* Designed and developed Web Application using Struts Framework.
* ANT to compile and generate EAR, WAR, and JAR files.
* Created tes1t case scenarios for Functional Testing and wrote Unit test cases with JUnit.
* Responsible for Integration, unit testing, systemtesting and stress testing for all the phases of project.

**Environment**: Java, J2EE, JSP 1.2, Performance Tuning,Spring1.2,Hibernate 2.0, JSF1.2,EJB 1.2 , IBM WebSphere6.0, Servlets , JDBC, XML, XSLT, DOM, CSS, HTML, DHTML, SQL, Java Script, Log4J, ANT1.6, WSAD6.0, Oracle 9i, Windows 2000.