**VENKATESWARA REDDY**

**Email:** [**venky88.d@gmail.com**](mailto:venky88.d@gmail.com)*Sr. Hadoop / Spark Developer* **Cell: (571)-279-9966**

**Professional Summary:**

* 8+ years of professional experience in IT industry involved in Developing, Implementing and maintenance of various applications using **Java**, **J2EE** technologies, with 4+ years’ experience in **Hadoop** ecosystem's Developing, Implementing and maintenance.
* Excellent understanding of **Hadoop architecture** and underlying **framework** including storage management.
* Hands on experience on **Hadoop/Big Data** related technology experience in **Storage, Querying, Processing and analysis of data.**
* Experienced in using various Hadoop infrastructures such as **Map Reduce**, **Hive**, **Sqoop**, and **Oozie**.
* Experienced in Collected logs data from various sources and integrated in to HDFS using Flume and experience in developing custom **UDFs** for **Hive**.
  + - Experienced in testing data in **HDFS** and **Hive** for each transaction of data.
    - Experienced in importing and exporting data using **Sqoop** from **HDFS to Relational Database Systems** and vice-versa.
    - Experienced with the **Spark** improving the performance and optimization of the existing algorithms in **Hadoop** using **Spark Context, Spark-SQL, Data Frame, Pair RDD's, Spark YARN**.
    - Extensive knowledge in programming with Resilient Distributed Datasets (**RDDs**).
    - Experienced in using **Flume** to transfer log data files to **Hadoop Distributed File System** (HDFS)
    - Experienced with **Akka** building high performance and reliable distributed applications in **Java** and **Scala.**
* Worked **Akka**, toconfigure how actors will be created, destroyed, scheduled, and restarted upon failure.
* Knowledge and experience in job work-flow scheduling and monitoring tools like **Oozie and Zookeeper**.
* Good working knowledge of **Amazon Web Service** components like **EC2**, **EMR**, **S3**.
* Good experience in **Shell programming.**
* Experience in working Linux Based Operating systems like **Ubuntu** and **CentOS**
* Knowledge in configuration and managing - Cloudera’s Hadoop platform along with **CDH3&4 clusters.**
* Knowledge and experience of architecture and functionality of NOSQL DB like **Cassandra** and **Mongo DB**.
* Experienced in application development using **Java, J2EE, JDBC, spring, Junit.**
* Experienced in developing web based GUIs using **JavaScript, JSP**, **HTML**, **JQuery**, **XML** and **CSS.**
* Experienced to develop enterprise applications with J2EE/MVC architecture with application servers and Web servers such as, **JBoss**, and **Apache Tomcat** 6.0/7.0/8/0.
* Strong Experience in working with Databases like **Oracle 11g/10g/9i, DB2, SQL Server 2008** and **MySQL** and proficiency in writing complex **SQL** queries.
* Experienced in using database tools like **SQL** Navigator, **TOAD**.
  + - Experience in using **PL/SQL** to write **Stored Procedures**, **Functions** and **Triggers**.
* Demonstrated expertise utilizing **ETL** tools, including **SQL** Server Integration Services (SSIS), Data Transformation Services (DTS), and Data Stage and ETL package design, and RDBM systems like **SQL Servers, Oracle,** and **DB2**.
* Excellent technical and analytical skills with clear understanding of design goals of ER modeling for **OLTP** and dimension modeling for **OLAP**.
* Excellent communication, interpersonal and analytical skills and a highly motivated team player with the ability to work independently.
  + - Ability to learn and adapt quickly to the emerging new technologies and paradigms.

**Skill Set:**

|  |  |
| --- | --- |
| **Hadoop/Big Data** | MapReduce, HDFS, Hive, Pig, HBase, Zookeeper, Sqoop, Oozie, Flume, Spark, Scala, Akka, Kafka, Storm, MongoDB |
| **Languages** | Java, J2EE, PL/SQL, Pig Latin, HQL, Unix shell scripts |
| **Java/J2EE Technologies.** | JDBC, Java Script, JSP, Servlets, JQuery |
| **Databases** | Oracle 11g/10g/9i, Microsoft Access, MS SQL |
| **No SQL Databases** | Apache Cassandra, mongo DB |
| **Web Technologies** | HTML, DHTML, XML, XHTML, JavaScript, CSS, XSLT. |
| **Web/Application servers** | Apache Tomcat6.0/7.0/8.0, JBoss |
| **Frameworks** | MVC, Struts, Spring, Hibernate. |
| **Operating Systems** | UNIX, Ubuntu Linux and Windows 00/XP/Vista/7/8 |
| **Network protocols** | TCP/IP fundamentals, LAN and WAN. |

**Professional Experience:**

**Employer: Telma Solutions LLC**

**AARP, Baltimore, MD Nov 2015- Present**

**Sr. Hadoop / Spark Developer**

The AARP is an American holding company for various life insurance companies and investment firms doing business primarily in the United States, offering life and supplemental health insurance, investments, and retirement services. The Project is about collecting the Ads data from various sources into ware house for the further data analysis to make intelligent business decisions.

**Responsibilities:**

* Developed Spark code using **Scala** and **Spark-SQL**/Streaming for faster testing and processing of data.
* Exploring with the **Spark** improving the performance and optimization of the existing algorithms in Hadoop
* Load the data into **Spark RDD** and do in memory data Computation to generate the Output response.
* Import the data from different sources like **HDFS/HBase** into **Spark** **RDD**.
* Installed **Hadoop, Map Reduce, HDFS, AWS** and developed multiple **MapReduce** jobs in **PIG** and **Hive** for data cleaning and pre-processing.
* Integrating user data from **Cassandra** to data in **HDFS**. Integrating Cassandra with Storm for real time user attributes look up.
* Used **Hive** to analyze data ingested into **HBase** by using **Hive-HBase** integration and compute various metrics for reporting on the dashboard.
* Used **Spark API** over Hortonworks **Hadoop** **YARN** to perform analytics on data in **Hive**.
* Created **HBase** tables to store variable data formats **(Avro, JSON)** of data coming from different portfolios using **NOSQL.**
* Developed **PIG** **Latin** scripts for the analysis of semi structured data.
* Configured deployed and maintained multi-node Dev and **Test** **Kafka** Clusters.
* Involved in converting **Hive/SQL** queries into Spark transformations using **Spark RDDs, Python** and **Scala.**
* Performed transformations, cleaning and filtering on imported data using **Hive, Map Reduce,** and loaded final data into **HDFS**.
* Worked with **NoSQL** databases like **HBase** in creating HBase tables to load large sets of semi structured data coming from various sources.
* Successfully integrated **Hive** tables and **Mongo DB** collections and developed web service that queries **Mongo DB** collection and gives required data to web UI.
* Developed workflows using Oozie to automate the tasks of loading the data into **HDFS** and pre-processing with **Pig.**
* Worked on importing the unstructured data into the **HDFS** using **Flume**.
* Worked on **Oozie** workflow engine for job scheduling.
* Implemented **MapReduce** programs to handle semi/ unstructured data like **XML, JSON, Avro** data files and sequence files for log files.
* Experience in **Oozie** and **workflow** scheduler to manage **Hadoop** jobs by Direct Acyclic Graph (DAG) of actions with control flows.
* Involved in running real-time processing using **STORM.**
* To configure Hadoop environment in cloud through **Amazon Web Services (AWS**) and to provide a scalable distributed data solution.
* Worked on installation of **KAFKA** on Hadoop cluster and to use it for streaming & cleansing of raw data and have extracted useful information using Hive and stored the results in HBase and have enabled the clients to review the results using Tableau by connecting it through the IP address provided by **AWS**.
* Cluster co-ordination services through **ZooKeeper.**
* Used **Akka** as a framework to create reactive, distributed, parallel and resilient concurrent applications in **Scala**
* Expertise in different data Modeling and Data Warehouse design and development.
* Worked on Designing and Developing **ETL** Workflows using Java for processing data in **HDFS/HBase** using **Oozie.**

**Environment:** Apache Hadoop, HDFS, MapReduce, Pig, Hive, Sqoop, Kafka, Akka, Solr, HBase, Oozie, Flume, Spark, Scala, Apache Cassandra, Impala, Zookeeper, Java, SQL Scripting, Linux Shell Scripting, Python, and ETL

**Comcast, New Kensington, PA  Jan 2014 - Oct 2015**

**Big Data / Hadoop Developer**

Comcast Corporation, formerly registered as Comcast Holdings, is an American multinational mass media company and is the largest broadcasting and largest cable television company in the world by revenue. It is the second-largest pay-tv company after the AT&T-DirecTV acquisition, largest cable TV company and largest home Internet service provider in the United States, and the nation's third-largest home telephone service provider.

**Responsibilities**

* All the data was loaded from our relational **DBs** to **HIVE** using **Sqoop**. We were getting four flat files from different vendors. These were all in different formats e.g. text, EDI and XML formats
* Since our cluster was over 100 nodes, we were using resource manager (application master)
* Writing **Hive** join query to fetch info from multiple tables, writing multiple M**ap Reduce** jobs to collect output from **Hive**
* Used Hive to analyze the partitioned and bucketed data and compute various metrics for reporting on the dashboard.
* Used Hive to analyze data ingested into **HBase** by using **Hive**-**HBase** integration and compute various metrics for reporting on the dashboard
* Involved in developing **Map-reduce framework**, writing queries scheduling map-reduce
* Developed the code for Importing and exporting data into **HDFS** and Hive using **Sqoop**
* Installed and configured **Hadoop** and responsible for maintaining cluster and managing and reviewing Hadoop log files.
* Developed Shell, Perl and **Python** scripts to automate and provide Control flow to **Pig scripts**.
* Continuous monitoring and managing the **Hadoop** cluster through **Cloudera Manager**.
* Worked on configuring and managing disaster recovery and backup on **Cassandra** Data.
* Performed Filesystem management and monitoring on Hadoop log files.
* Utilized **Oozie** workflow to run **Pig** and **Hive** Jobs Extracted files from MongoDB through Sqoop and placed in **HDFS** and processed.
* Used **Flume** to collect, aggregate, and store the web log data from different sources like web servers, mobile and network devices and pushed to **HDFS**.
* Implemented partitioning, dynamic partitions and buckets in **HIVE**.
* Developed customized classes for serialization and Deserialization in Hadoop
* Analyzed large amounts of data sets to determine optimal way to aggregate and report on it.
* Involved in migration of data from existing **RDBMS** (oracle and SQL server) to Hadoop using **Sqoop** for processing data.
* Implemented Installation and configuration of multi-node cluster on Cloud using **Amazon Web Services(AWS)** on **EC2**.

**Environment:** Apache Hadoop, HDFS, Hive, Map Reduce, Cloudera**,** Pig, Sqoop, Kafka, Apache Cassandra, Oozie, Impala, Cloudera, Flume, Zookeeper, Java, MySQL, Eclipse, PL/SQL and Python.

**East West Bank, Pasadena, CA Nov 2012- Dec 2013**

**Big Data/Hadoop Developer**

East West Bank is a bank based in the state of California in the United States, with assets of $29.9 billion and a market capitalization of over $5.8 billion. It now has over 130 branch locations in the U.S.

**Responsibilities:**

* Involved with the application teams to install Hadoop updates, patches and version upgrades as required.
* Worked on analyzing, writing **Hadoop MapReduce** jobs using **JavaAPI, Pig** and **Hive**.
* Created **HBase** tables to store variable data formats of data coming from different portfolios.
* Involved in **Configuring core-site.xml** and mapred-site.xml according to the multi node cluster environment.
* Used **Flume** to collect, aggregate and store the web log data onto **HDFS**.
* Having experience on Hadoop eco system components **HDFS, MapReduce, Hive, Pig, Sqoop** and **HBase.**
* Used **Flume** to collect, aggregate, and store the web log data from different sources like web servers, mobile and network devices and pushed to **HDFS**.
* Load and transform large sets of structured, semi structured and unstructured data onto **RDBMS** from Hadoop environment using **Sqoop** for reporting on the dashboard.
* Responsible for writing **Hive** Queries for analyzing terabytes of customer data from **HBase** and put the results in output file.
* Involved in creating **Hive** tables and working on them using **Hive QL.**
* Configured periodic incremental imports of data from **MySQL** into **HDFS** using Sqoop.
* Responsible for Load, aggregate and move large amounts of log data using **Flume**.
* Involved in loading data from **UNIX** file system to **HDFS**.
* Worked on loading and transformation of large datasets of structured, semi structured and unstructured data into **Hadoop ecosystem**.
* Responsible to manage data coming from different data sources.
* Analyzed the data by performing **Hive** queries and running **Pig** **scripts** to study customer behavior
* Responsible for migrating tables from traditional **RDBMS** into Hive tables using **Sqoop** and later generate required visualizations and dashboards using **Tableau**.
* Responsible for architecting Hadoop clusters with **CDH4** on **CentOS**, managing with **Cloudera Manager.**
* Development experience in **UNIX, LINUX** and Windows (Vista, XP, NT, 2000, 95) and Cloud based virtual platforms.
* Creating Server Side of application for project management using **Mongo DB**.
* Designed Business classes and used Design Patterns like Data Access Object, **MVC** etc.
* Responsible for the overall layout design, color scheme of the web site using **HTML, bootstrap** and **CSS3.**

**Environment:**  Apache Hadoop, HDFS, Hive, HBase, Map Reduce, Pig, Sqoop, Kafka, Mongo DB, Apache Web server, Tableau**,** Flume, Storm, Oozie,Cloudera Java, HTML, CSS**,** JavaScript, jQuery, JDBC, UNIX, XML, Java Script, MVC and Jenkins (build tool).

**Zyno Medical - Boston, MA Dec 2011 – Oct 2012**

**JAVA/Hadoop Developer**

Providing corporate information and hospital statistics, which can be viewed by Directors and management or by those who are authorized to see it. The information includes all user’s registration, payments, consultation statistics, pharmacy, surgery, transplants, equipment's, patients and doctors, healthcare bill, multi valued diagnosis, preventive health care, corporate counseling, welfare programmers, claims administration, patient care co-ordination, doctor achievements, staff details and human resource.

**Responsibilities:**

* Involved in requirement analysis, design, coding and implementation.
* Worked in **Agile Methodology** and used **JIRA** for maintain the stories about project.
* Analysed large data sets by running **Hive queries.**
* Involved in Design, develop **Hive** Data model, loading with data and writing Java UDF for **Hive**
* Handled importing and exporting data into **HDFS** by developing solutions, analyzed the data using **Map** **Reduce**, **Hive** and produce summary results from Hadoop to downstream systems.
* Used **Sqoop** to import and export the data from Hadoop Distributed File System (**HDFS**) to RDBMS.
* Created **Hive tables** and loaded data from **HDFS** to **hive** tables as per the requirement.
* Established custom **Map Reduces** programs in order to analyze data and used **HQL queries** to clean unwanted data.
* Created components like **Hive** UDFs for missing functionality in **Hive** to analyze and process the large volumes of data.
* Worked on various performance optimizations like using distributed cache for small datasets, **Partition**, **Bucketing** in **hive** and Map Side joins.
* Involved in writing complex queries to perform **join** operations between multiple tables.
* Involved actively verifying and testing data in **HDFS** and **Hive** tables while Sqoopingdata from **Hive to RDBMS** tables.
* Developed job flows to automate the workflow for **pig** and **hive** jobs.
* Used cloud computing on multi-node cluster and deployed hadoop application on cloud S3 and used **Elastic Map Reduce**(EMR) to run a **Map-reduce**.
* Extracted the data onto **HDFS** using **Flume**
* Designed and Maintained **Oozie** workflows to manage the flow of jobs in the cluster.
* Created and maintained Technical documentation for launching Hadoop Clusters and for executing Hive queries and Pig Scripts
* Used **IMPALA** to pull the data from Hive tables.
* Work involved extensive usage of **HTML, CSS, JavaScript and Ajax** for client side development and validations.
* Involved in designing user screens and validations using **HTML, jQuery, Ext JS and JSP** as per user requirements
* Used parsers for the conversion of **XML** files to java objects and vice versa.
* Migration of **ETL** processes from **MySQL** to **Hive** to test the easy data manipulation.
* Used Apache **Maven** 2.x to build and deploy application to various environments
* Installed **Oozie workflow** engine to run multiple **Hive** jobs which run independently with time and data availability
* Worked with **Avro** **Data Serialization** system to work with **JSON** data formats.
* Actively updated the upper management with daily updates on the progress of project that include the classification levels that were achieved on the data.

**Environment**: Java API, Apache Hadoop, MapReduce, Pig, HDFS, HBase, HiveQL,Cloudera Oozie, Sqoop, Kafka, Storm, MySQL, Linux, AWS, Netbeans, Hibernate, XML, HTML, JavaScript, JDBC, JSP, PL/SQL, CSS, SOAP Web services.

**Virima soft solutions, India. Sep 2009 – Nov 2011**

**Java/UI Developer**

Virima Soft is IT services company based in Hyderabad, India. We are specialized in outsourced software application and product development using world-class IT development methodologies. We have rich expertise in developing robust, scalable and high-end web, collaborative and client-server based applications using Java, .NET, PHP, HTML 5, Web 2.0, Android, Symbian technologies.

**Responsibilities:**

* Participated in requirement gathering and converting the requirements into technical specifications.
* Developed UI using **HTML**, **JavaScript**, and **JSP**, and developed Business Logic and Interfacing components using Business Objects, **XML**, and **JDBC**.
* Created rapid prototypes of interfaces to be used as blueprints for technical development.
* Responsible for design, development, test and maintenance of applications designed on Java technologies.
* Used UML diagrams Use Cases, Object, Class, State, Sequence and Collaboration to design the application using Object Oriented analysis and design
* Managed connectivity using **JDBC** for querying/inserting & data management including triggers and stored procedures.
* Developed **JavaScript** behavior code for user interaction.
* Created database program in **SQL** server to manipulate data accumulated by internet transactions.
* Wrote **Servlets** class to generate dynamic **HTML** pages.
* Developed **SQL** queries and Stored Procedures using **PL/SQL** to retrieve and insert into multiple database schemas.
* Developed the **XML** Schema and **Web services** for the data maintenance and structures Wrote test cases in **JUnit** for unit testing of classes.
* Used DOM and **DOM Functions** using Firefox and IE Developer Tool bar for IE.
* Debugged the application using Firebug to traverse the documents.
* Involved in developing web pages using **HTML** and **JSP**.
* Provided Technical support for production environments resolving the issues, analysing the defects, providing and implementing the solution defects.
* Involved in writing **SQL Queries**, Stored Procedures and used JDBC for database connectivity with **MySQL** Server.
* Developed the presentation layer using **CSS** and **HTML** taken from bootstrap to develop for browsers.

**Environment:** Java**,** XML, HTML, JavaScript, JDBC, CSS, SQL, PL/SQL, XML, Web MVC, Eclipse, Ajax, JQuery, spring with Hibernate, Active MQ, Jasper Reports, Ant as build tool and My SQL and Apache Tomcat

**Gati Ltd, India Jun 2008 - Aug 2009**

**Junior Java Developer**

GATI is a cargo services company with many departments like HR, Finance, Accounting, Customer Service, Information Technology, Logistics, Operations, and Marketing.

**Responsibilities:**

* Involved in both maintenance and new enhancements of the application.
* Developed **Servlets** and **JDBC** were used in retrieving data.
* Designed and developed dynamic Web pages using **HTML** and **JSP**.
* Implemented **Object Relational** mapping in the persistence layer using Hibernate Framework in conjunction with Spring Functionality.
* Involved in the design and development of the entire application. Created **UML** diagrams (use case, class, sequence, and collaboration) based on the business requirements. Involved in Analysis, Design, Coding and Development of custom Interfaces.
* Used Spring IOC for injecting the **Hibernate** and used Hibernate annotations to design the modelling part of the applications.
* Involved in writing **PL/SQL, SQL** queries.
* Used **Oracle DB** for writing **SQL** scripts, **PL/SQL** code for procedures and functions.
* Wrote JUnit test cases to test the functionality of each method in the DAO layer. Configured and deployed the WebSphere application Server.
* Prepared technical reports and documentation manuals for efficient program development.

**Environment:** JSP, HTML, Servlets, Hibernate, Spring Framework, XML, JDBC, PL/SQL, WebSphere, Eclipse, Junit, J2EE, struts 1.2.9, Spring 2.5, JavaScript and Apache Tomcat.

**Education:**

* **Bachelors of Engineering in Computer Science from JNT University, Hyderabad**