**Hari Rajesh**

**Phone:** (609) 451-7629

**Email:**[Harir9119@gmail.com](mailto:Harir9119@gmail.com)

**Professional Summary:**

* 7+ years of extensive IT experience in all phases of Software Development Life Cycle (SDLC) with skills in data analysis, design, development, testing and deployment of software systems.
* 3+ years of strong experience, working on Apache Hadoop ecosystem components like **MapReduce**, **HDFS**, **HBase**, **Hive**, **Sqoop**, **Pig**, **Oozie**, **Zookeeper**, **Flume** with **CDH4&5** distributions and **EC2** cloud computing with **AWS**.
* Worked with different flavors of Hadoop distributions which includes **Cloudera** and **Hortonworks**.
* Good experience in working with cloud environment like Amazon Web Services (**AWS**) **EMR**, **EC2**, and **S3**.
* In depth understanding/knowledge of Hadoop Architecture and various components such as **HDFS, JobTracker, TaskTracker, NameNode, DataNode.**
* Hands on experience in developing in both Web based and Standalone applications using Spring framework modules like **Spring annotations**, **Spring Web MVC**, **Spring DAO** and **IOC**.
* Experience in working on SOAP based web services using **SOAP UI**, **WSDL**. Worked on **RESTFUL** web services.
* Expert in developing web applications using **Struts**, **Hibernate** and **Spring Frameworks**.
* Worked in designing the User Interface using **HTML**, **CSS**, **JavaScript**, **Bootstrap**, **JQuery** according to the client requirement.
* Experience in applying the latest development approaches including applications in Spark using **Scala** to compare the performance of **Spark** with **Hive** and **SQL/Oracle**.
* Experience working with **Cloudera** &**Hortonworks** Distribution of Hadoop.
* Expertise in using SparkSQL with various data sources like **JSON, Parquet and Hive**.
* Experienced in Developing Spark programs using **Scala** and **Java API's.**
* Expertise in using **Kafka** as a messaging system to implement real-time streaming solutions.
* Implemented **Sqoop** for large data transfers from **RDBMS** to **HDFS/HBase/Hive** and vice-versa.
* Experienced in using **Zookeeper** to coordinate the servers in clusters and to maintain the data consistency.
* Experienced with Cloudera Manager to monitor health and performance of the **Hadoop cluster**.
* Expertise in using **Flume** in Collecting, aggregating and loading log data from multiple sources into **HDFS.**
* Migrated an existing on-premises application to **AWS**. Used AWS services like EC2 and S3 for small data sets processing and storage, Experienced in Maintaining the Hadoop cluster on **AWSEMR**.
* Experience in Amazon **AWS** services such as **EMR**, **EC2**, **S3**, **CloudFormation**, **RedShift**, **DynamoDB** which provides fast and efficient processing of Big Data.
* Hands on experience with **AWS** (Amazon Web Services), Elastic Map Reduce (**EMR**), **StorageS3**, **EC2** instances and Data Warehousing.
* Scheduled various **ETL** process and **Hive scripts** by developing **Oozie workflows**.
* Experienced in working with structured data using **HiveQL**, **join operations, Hive UDF’s, partitions,** bucketingandinternal/externaltables**.**
* Experience in handling various file formats like **AVRO, Sequential, Parquet** etc. .
* Experience in migrating ETL process into Hadoop, Designing Hive data model and wrote **PigLatin scripts** to load data into **Hadoop**.
* Proficient in Various **NoSQL** Databases like **Cassandra, MongoDB, HBase** etc
* Good understanding and experience with Software Development methodologies like **Agile** and **Waterfall**.
* Worked coherently with multiple teams, conducted peer reviews, organized and participated in knowledge transfer(technical and domain) sessions.

**Technical Skills:**

|  |  |
| --- | --- |
| **Hadoop/Big Data** | MapReduce, Hive, AWS, Pig, Impala, Sqoop, Flume, HDFS, Oozie, Hue, HBase, Zookeeper, Spark, Kafka, Avro. |
| **Java & J2EE Technologies** | Java, Servlets, Struts, Spring, JSP, JDBC. |
| **Programming Languages** | C/C++, Unix shell scripting, SQL, PL/SQL, Scala, Python. |
| **Frameworks** | Hibernate, Struts, Spring Frameworks. |
| **Databases** | Oracle, MySQL, DB2, SQL Server, HQL. |
| **NoSQL Databases** | Cassandra, HBase, MongoDB. |
| **Web Technologies** | JavaScript, HTML, XML, REST, SOAP, CSS. |
| **Operating Systems** | Windows, Ubuntu, centos, Linux, Unix. |
| **IDE’s** | Eclipse, Net beans, Visual Studio, Intelli-J, WebStorm. |

**PROFESSIONAL EXPERIENCE:**

**Client: Coach - Atlantic City, NJ Jan 2016 - Present**

**Role: Hadoop Developer**

**Description:**

Maintaining the customer member details and rewards points transaction are very difficult in terms of storage and processing. Member loyalty management system is replacing the existing reward management system which is developed as a web service provider with the help of database shading. Aim of this system is to reduce the response time of web service. Used Hive and Pig to generate BI reports.

**Responsibilities:**

* Responsible for building scalable distributed data solutions using **Hadoop.**
* Involved in building all domain pipelines using Spark Data Frames, **Spark Steaming** and **Spark batch**   
  **processing**.
* Involved in file movements between HDFS and **AWSS3** and extensively worked with **S3** bucket in **AWS**.
* Developed spark application for filtering Json source data in AWS S3 location and store it into HDFs with partitions and used spark to extract schema of Json files.
* Improving the performance and optimization of existing algorithms in Hadoop using **Spark Context**, **Spark**-**SQL** and **SparkYarn** using **Scala**.
* Experience in ingesting real-time data using **Flume**, **Kafka**.
* Worked on importing and exporting data from **Oracle** and **DB2** into **HDFS** and **HIVE** using **Sqoop.**
* Responsible for trouble shooting issues in the execution of **MapReduce** jobs by inspecting and reviewing log files.
* Developed simple to complex **MapReduce** programs to analyze the datasets as per the requirement.
* Configured periodic incremental imports of data from **MySQL** into **HDFS** using **Sqoop**.
* Responsible for migrating tables from traditional **RDBMS** into **Hive** tables using **Sqoop** and later generate required visualizations and dashboards using **Tableau**.
* Responsible for Load, aggregate and move large amount of log data using **Flume**.
* Developed the Sqoop scripts in-order to make the interaction between **HDFS** and **RDBMS** (Oracle, MySQL).
* Worked extensively with importing metadata into Hive and migrated existing tables and applications to work on Hive and **AWS** cloud and making the data available in Athena.
* Imported the data from different sources like **AWSS3**, Local file system into Spark RDD.
* Scheduled an Event -Driven **AWSLambda** functions to trigger when data is uploaded into the **S3** buckets.
* 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.
* Involved in identifying job dependencies to design workflow for **Oozie** and resource management for **YARN**.
* Used **Cassandra** to store the analyzed and processed data for scalability.
* Involved in gathering requirements and participating in the **Agile** planning meetings in-order to finalize the scope of each development.

**Environment:** Hadoop, AWS, Java, JavaScript, MapReduce, HDFS, Hive, Pig, Impala, Kafka, Cassandra, Clouera Manager, ETL, Sqoop, Flume, Oozie, Zookeeper, MySql, Eclipse, Tableau and Unix.

**Client: Autodesk - San Francisco, CA June 2014 – Jan 2016**

**Role: Hadoop Developer**

**Description:**

Autodesk makes software for people who make things. At Autodesk, the current marketing system does not enable us to engage with our customers based on where they are in their journey, through their preferred communication channels, and with activities and content relevant to their needs and recent actions. Hence, we are trying to implement technology and process to enable a scalable, repeatable nurture marketing practice as well as optimize campaign effectiveness based on this data.

**Responsibilities:**

* Responsible for Cluster maintenance, commissioning and decommissioning Data nodes, Cluster Monitoring, TroubleShooting, Manage and review **data backups**, Manage & review **Hadoop**log files.
* Load log data into **HDFS** using **Flume, Kafka** and performing **ETL** integrations.
* Involved in gathering the business requirements, designing (Low-level and High-level), development and testing.
* Written **Map Reduce** jobs in java to process the log data.
* Created **Hive tables** to store the processed results in a tabular format.
* Involved in transforming data from Mainframe tables to **HDFS** and **HBASE** tables using **Sqoop** and worked on **Impala** to analyze stored data.
* Execution of Hadoop ecosystem and Applications through **Apache HUE**.
* Developing Hive User Defined Functions (UDF’s) in **java**, compiling them into jar’s and adding them to the HDFS and executing them with **Hive Queries**.
* Have deep and thorough understanding of **ETL** tools and how they can be applied in a Big Data environment, supporting and managing **Hadoop** **Clusters** using **Cloudera.**
* Involved in loading data from **UNIX** file system to **HDFS** and Created custom Solr Query components to enable optimum search matching.
* Used Partitioning, Dynamic-Partitioning and bucketing concepts for performance optimization in **hive**.
* Written the **Apache PIG Scripts** for loading, filtering and storing the data in **HDFS**, which later will be used by **Hive** or **MRjobs**.
* Implemented Pub/Sub model using **Apache Kafka** for real-time transactions to load into**HDFS**.
* Made use of Distributed cache to improve the performance of the jobs and reduce the load on cluster.
* Developed **Spark-SQL** scripts and involved in converting hive UDF’s to Spark-SQL UDF’s.
* Implemented POC to migrate MapReduce programs into Spark transformations using **Spark and Scala.**
* Automated workflows using **Oozie** jobs.
* Used flume to collect the entire web log from the online ad-servers and push into **HDFS**.

**Environment**: Hadoop, Java, Spring, Hibernate, HDFS, Hive, Impala, Sqoop, Flume, Pig, MapReduce, Oozie, Apache Kafka, Oracle, MySQL, UNIX, ETL, Spark and Scala.

**Client: Comcast - Los Angeles, CA Feb 2013 - June 2014**

**Role: Java Developer**

**Description:**

Comcast is a multinational mass media company and is the largest broadcasting and Cable Company. I worked as a Java developer in a project which is about developing and maintaining business process, information data and web enterprise architecture to ensure alignment with new business strategies.

**Responsibilities:**

* Development of Action Classes, DAO and Hibernate DAO Support classes.
* Worked in developing **Struts** related backend services (to access the database).
* Used **ORM** to integrate **Struts Framework** with **Hibernate**.
* Used **JSP** and **Servlets**, **EJB’s** on server side.
* Installed, configured, deployed, Monitored and troubleshoot **WebLogic**.
* Studied the implementation of **Struts**.
* Experience in developing **UML usecases**, **class**, **sequence** and **deployment** diagrams using Rational Rose and Microsoft Visio.
* Good Knowledge of using **IDE** Tools like **Eclipse**, **NetBeans**, **Rational Application Developer**(RAD) for **Java/J2EE** application development.
* Developed the Action Classes, Action Form Classes, created JSP’s using **Struts** tag libraries and configured in Struts-config.xml, Web.xml files.
* Involved in writing packages, stored procedures to retrieve/modify data in database.
* Performed extensive query optimization by analyzing **DB2** Explain Plans for faster report generation.
* Worked extensively with and implemented the use of collections.
* Took advantage of some tools such as **SQL Plus, Oracle 10g**, and **Toad** in-order to work with databases.
* Involved in **Apache**-**Tomcat server** for maintenance and deployment.
* Participated in **design reviews**, **testcase reviews**, **code reviews** and **testing**.
* Extensively involved in database activities like designing tables, writing stored procedures, functions, queries in the application.

**Environment**: Java/J2EE, Struts, Hibernate, spring, JSP, Servlets, EJB,Oracle 10g Application Server, UML, Toad, DB2, SQL Plus, Eclipse, WebLogic, RAD and NetBeans.

**Client: Hierarchy Technologies - Hyderabad, India Apr 2012 - Nov 2012**

**Role: Java/J2EE Developer**

**Description:**

Hierarchy Technologies has been designed to facilitate the business process of tracking a Bill of Materials Order through its life cycle. This encompasses the initial forecasting of the project through to the required reports being available to the relevant Finance and Billing teams for verification. The solution is designed to automate as much of the business process and eliminate as many manual steps as possible.

**Responsibilities:**

* Involved in analyzing the client requirements and convert them into technical specifications.
* Worked in Analysis, Design and Coding for client development using **J2EE** stack and**Eclipse** platform.
* Involved in analyzing the requirements, modeling & designing the features.
* Coding and maintaining **Oracle packages**, stored procedures and functions.
* Prepared technical design documents that include **UseCase** &**Sequence** diagrams using Rational Rose.
* Involved in creating web-based java components like **client Applets** and client-side UI using JFC in **Eclipse**.
* Set up Data sources and queues set up in **JBoss Application Server** that is responsible for transferring data to other queue managers.
* Used Spring **MVC** to develop the application & control the operation flows.
* Developed **PL/SQL** stored procedures to perform complex database operations.
* Designed and developed **SQL** queries in the application.
* Developed Design documents for various components identified in the system.
* Generated the **Hibernate, XML** and **Java Mappings** for the schemas.
* Extensively used **J2EE, Servlets, JSP and XML**.
* Generated the **Hibernate**, **XML** and Java Mappings for the schemas.
* Collaborated with offshore team to ensure the project **HTML** tasks are completed on time with accuracy.
* Used **Subversion** as the version control system.
* Transactional Data Analysis in **Matlab**.

**Environment:** Java/J2EE, JSP, EJB, Servlets, MVC, Hibernate, Applets, HTML, Oracle, JavaScript, JBOSS, XML, SQL, PL/SQL, Eclipse and MatLab.

**Client: Robotics Inc - Hyderabad, India June 2010 - Apr 2012**

**Role: Jr. Java Developer**

**Description:**

This project attempts to automate the income tax and the sales tax procedures. This project is not specific to any country, but to display the working of the model, the guidelines and business rules specific to the Indian government's tax procedures have been considered.

**Responsibilities:**

* Training in gathering user requirements and followed by **analysis** and **design**.
* Assisting in Developing **JSP's**, action classes form **beans**, response beans, **EJB's**.
* Designed customer maintenance module, client interactive pages using web **technologies** like **HTML**, **CSS** and **JavaScript**.
* Implemented innovative techniques in web services that helped the architecture to avoid the hassle of multiple pages for single screen.
* Developed analysis level documentation such as UseCase, Business Domain Model, Activity, Sequence and Class Diagrams.
* Coded **Servlets** for the Transactional Model to handle many requests.
* Participated in developing business objects helpers which interact with middleware tubs.
* Involved in implementing business delegate pattern to separate view from business process.
* Improved product browser compatibility, increasing accessibility.
* Collaborated with wide variety of roles to provide high-quality, robust solutions.
* Conducted Design reviews and Technical reviews with other project statehood Implemented services using Java.
* Extensively used **XML** to code configuration files.
* Developed **PL/SQL** stored procedures, **triggers**.
* Performed **functionalintegration**, **system** and **validation testing**.
* Technical Support to client and defect management.

**Environment:** Java/J2EE, Apache Tomcat Server, Servlets, XML, HTML, CSS, JavaScript, EJB, PL/SQL, Java Beans and JSP.