**Rekha**

**PROFESSIONAL SUMMARY:**

* Over 8+ years of software experience as a Developer, Designer with cross platform integration experience **with Hadoop, Hive,Pig,Sqoop, HBase,J2EE, JDBC, HTML, and JavaScript.**
* Experience in Design, Development, integration and implementation of web based applications, Client/server based applications, **Hadoop**, Mainframe and **SOA** based applications.
* Expertise in **HDFS architecture** and **Cluster Deployments**.
* Well experienced on development of Java **MapReduce**Programs and also on Various Hadoop Ecosystem Such as **Pig**, **Sqoop**, **Hive**,**Oozie**, **Flume,ZooKeeper** etc.
* Thorough knowledge on Development life cycle and usage of various Hadoop Eco systems **Hive, Pig, Sqoop, Map Reduce, Flume, Oozie, and Spark**.
* Extensively worked on **Pig Latin scripting** using Both Local and Map Reducing mode.
* Expertise in**Hive Query Language**as part of data processing in cloud data platform and debugging **Hive** issue.
* Expertise in Optimization of **Hive tables** using optimization techniques like **Partitioning** and **Bucketing.**
* Experience with Managing and Understanding of **Hadoop Log Files.**
* Worked on **NoSQL** database such as **Hbase, Cassandra** for storing high volume data using Sqoop.
* Acquired Skills in writing **Linux Scripts**, **Pig Scripts**, **Hive queries**and **Map Reduce** programs.
* Extensive knowledge on**Hadoop** and **Hive** Security.
* Hands-on experience with **Horton works & Cloudera Distributed Hadoop (CDH)**
* Experience with **Oozie** in scheduling jobs and workflow.
* Good knowledge of Hadoop Architecture and various components such as **HDFS**, **Name Node**, **Data Node** and **MapReduce** concepts.
* Hands on experience on **Amazon AWS** concepts like EMR & EC2 web services which provides fast and efficient processing of Big Data.
* Worked on the backend using **Scala and Spark** to perform several aggregation logics.
* Extensive experience working with different file formats, compression techniques, setting up batch jobs, ETL/ELT pipelines, DevOps, **automation** using Chef on **AWS**.
* Hands on experience working on**Java virtual machine** (JVM) and **multi-threaded processing**.
* Expertize on**Data Warehousing**and proficient in using MS Office 365.
* Experience in integration of various data sources like **Oracle, DB2, Sybase, SQL** server and MS access and non-relational sources like flat files into staging area.
* Expertise in creating Hadoop Clusters using **AWS** like Amazon EC2 & Amazon S3 bucket.
* Experienced as a Java Developer for 3+ years with hands-on experience on **J2EE** technologies such as **JDBC**, **JSP** which includes frameworks such as **spring**, **Hibernate** and **Struts**.
* Good knowledge in implementation of Spring Frameworks such as **Spring**, **Spring MVC**, **Spring Security** and **Spring JDBC**.
* Proficiency in writing **SQLQueries**, **Stored Procedures – PL/SQL.**
* Experience in Production Job-scheduling, Implementation and job-monitoring using **Control-m tool.**
* Excellent problem solving skills, high analytical skills, good communication and interpersonal skills.
* Used the **Agile** and **Waterfall methodology** to develop the applications.
* Efficiency to handle emergency and high-pressure circumstances with demanding production support activity.

**Technical skills:**

|  |  |
| --- | --- |
| **Operating Systems** | **DOS, Windows 98/XP/2000/7, Unix (Sun Solaris),Linux, Unix** |
| **Hadoop/BigData Technologies** | **Pig, Hive, HiveQL, Sqoop, Kafka, Spark, Flume, Oozie, Impala, Zoo Keeper, Java Map Reduce, Redshift.** |
| **Languages** | **Java, Java EE 1.5, Java Script, SQL, PL/SQL,C.** |
| **Web/Distributed Technologies** | **J2EE, WSDL, XML, JDBC, UML, MVC, STRUTS, Spring, XML, XSL, Hibernate.** |
| **Scripting** | **Java Script, HTML, CSS, Unix Shell Scripting.** |
| **Open Frame works** | **Hibernate, Spring, JSON, Prototype, Hadoop** |
| **RDBMS** | **ORACLE 8i/9i/10g, MS-Access 2000, SQL-Server, Hbase, MySQL** |
| **Web/Application Servers** | **WebSphere, Apache Tomcat 4.1/5.1/6.0/7.0.** |
| **IDE / Tools** | **Eclipse 3.1 with Eclipse 4.1.1, NetBeans.** |
| **Methodologies** | **Unified Modeling Language (UML), Rational Unified Process (RUP), Agile** |
| **Distributions** | **Cloudera, MapR, Hortonworks** |

**Client: Wells Fargo, Charlotte, NC.**

**Duration: Jun 2015 – Present.**

**Role: Sr.Hadoop/BigData Developer.**

**RESPONSIBILITIES:**

* Installed and configured Hadoop **MapReduce, HDFS**, Developed multiple MapReduce jobs in **java** and **python** for data cleaning and preprocessing.
* Expertise in designing and deployment of **Hadoop cluster** and different Big Data analytic tools including **Pig**, **Hive,HBase, Oozie, Sqoop, Kafka, Spark**with Cloudera distribution.
* Participated in development/implementation of **Clouderaenvironment**.
* Responsible for running **Hadoop streaming jobs** to process terabytes of xml format data.
* Handled **importing of data** from various **data sources**, performed transformations using **Hive Map reduce**, loaded data into **HDFS** and extracted data from **MYSQL** into HDFS vice-versa using **Sqoop**.
* Loaded data from **Linux** file system to **HDFS**.
* Experience in Importing and exporting data into **HDFS** and Hive using **Sqoop** and **Kafka**.
* Defined**job flows,**managing and reviewing **Hadoop log files**.
* Installed and configured Hive and also written **Hive UDFs** in **Java**.
* Optimized **Hive joins** for large tables and developed map reduce code for full outer join of two large tables.
* Experienced in using **Hive-vectorization** for faster query process to utilize the CPU usage time.
* **Created Hive** tables, loading with data and writing hive queries which will run internally in map reduce way.
* Loaded the dataset into **Hive** for ELT Operation
* Creating Hive tables with periodic backups, writing **complex Hive/Impala** queries to run on **Impala.**
* Performed **Data Ingestion** from multiple internal clients using **Apache Kafka.**
* Written **Hive** queries for data analysis to meet the Business requirements.
* Design and develop **Pig Latin** scripts and pig command line transformations for data joins and custom processing of map reduce outputs.
* Designed and implemented cost-effective and performance-optimized configurations for **EMR** and **Redshift**.
* Experience with running Hadoop streaming jobs to process terabytes of xml format data using **Flume** and **Kafka**.
* Wrote **Pig** and **Hive** scripts to perform data manipulation operations that included data cleansing, joins and aggregations to process and load the data to **Redshift** cluster
* Created **HBase** tables for random read/writes by map reduce programs.
* **POC** conducted for different Use cases and documented on **AWS** Platform.
* Extensively Involved in loading data from **UNIX** file system to **HDFS**.
* Load the data into **Spark RDD** and performed in-memory **data computation** to generate the output response
* Installed and configured Apache Hadoop, **Hive** and **Pig** environment on **AWS**platform.
* Writing **Pig Latin** scripts to process the data and also written **UDF** in **java** and **python**.
* Wrote Map Reduce programs in Java to achieve the required Output.
* Involved in writing Java API's for interacting with **HBase,** building with **Maven.**
* Assisted **SQL** Server Database Developers in code review and optimizing **SQL** queries
* Developed job flows in **TWS** to automate the workflow for extraction of data from **Teradata** and **Oracle**.
* Involved in **NOSQL** databases like **Hbase, Cassandra** in implementing and integration.
* Used **Oozie** workflow engine to create the workflows and automate the **Map Reduce, Hive, Pig** jobs.
* Tested web services for **ARD** integration using **SOAPUI**.
* Responsible for creating different sessions and workflows to load the data to **Data Warehouse** using Informatics Workflow Manager.
* Supported in setting up **QA environment** and updating configurations for implementing scripts with **Pig** and **Sqoop**. Cluster co-ordination through **Zookeeper**.
* Experience in using**Jenkins** for **build and continuous integration** for software development.
* Cluster design, installation and configuration using HDP 2.1 stack on Azure and HDP 2.2 on premise setup.
* Implemented NameNode backup using NFS for High availability.
* Configured Hadoop environment with**Kerberos** authentication, Name nodes, and Data nodes.

**ENVIRONMENT**: Hadoop, MapReduce, HDFS, Pig, Hive, Hbase, Oozie, Impala, Cloudera CDH4.5, Cloudera CDH5.1.3, Kerberos security, Jenkins, Cassandra, Kafka, SQL, TWS, Maven, Python, Spark,Linux, Java, J2EE, Amazon Web services (AWS).

**Client: Adobe, San Jose, CA**

**Duration: Jan 2014 – May 2015.**

**Role: Big Data/ Hadoop Developer.**

* + - **Involved in the architectural decision with of Hadoop for application development and ecosystems like Hive,Pig, Zookeeper, Flume, Hbase, Impala and Sqoop.**
    - **Performed research &POC on Hive to analyze the partitioned and bucketed data and compute variousmetrics to determine the performance on Hadoop cluster.**
    - **Used Sqoop to import and export data from RDBMS to HDFS & vice versa.**
    - **Involved in installing and configuring Kerberos for the authentication of users and Hadoop daemons.**
    - **Installed and configured Pig for ELT jobs.**
    - **Created Hive tables and involved in data loading and writing Hive UDFs.**
    - **Created Hive External tables on the existing HDFS file systems**
    - **Performed data analysis, queries on hive, pig on AMBARI(Hortonworks)**
    - **Worked on importing and exporting data from SQL Server and Teradata into HDFS and HIVE using Sqoop.**
    - **Experience in real time streaming the data using Spark with Kafka.**
    - **Experience in using SQL Server Integration Services (SSIS) to extract, Transform and Load data**
    - **Performed development, deployment, job scheduling, testing, and validation& troubleshooting of data fromdevelopment to production environment.**
    - **Involved in writing shell scripts for Access Control List permissions for roles, active directories, UNIX directories& file systems.**
* Experience in writing **UNIX** Shell scripts for various purposes like **file validation**, **automation of ELT process** and **job scheduling** using **Crontab.**
  + - **Provided technical and analytical expertise to complex, specialized report requests requiring higher-leveldata analysis and data management services.**
    - **POC/analysis on no. of mappers & reducers required for fast processing of data for certain size of data &cluster.**
    - **Implemented POC's using Amazon Cloud Components S3, EC2, Elastic beanstalk and SimpleDB.**
    - **Involved in documentation of high level & detail level design of complete development phase.**
    - **Monitored production deployment for successful implementation of data.**
    - **Extensively worked on changed control process of the company required for Hadoop implementation &documented the same to help Hadoop team understand & follow the overall process.**
    - **Implementing ETL processes using Informatics to load data from Flat Files to target Data Warehouse database**
    - **Fix the code review comments and Building the code withJenkins and support for the code deployment into all the environments.**
    - **Worked on source target mapping of data to analyze the transformation of data for business analysis**
    - **Established healthy working relationship with team & management to ensure efficient & productive workspace& helped each other at any situation in order to avoid delay in project & met higher management expectations.**
    - **Learned new technology through company training & self to contribute in continuous innovation in technology.**

**ENVIRONMENT:Hadoop, MapReduce, Hive, HDFS, PIG, Sqoop, Flume, Kafka, Impala, HBase, Spark, ZooKeeperAMBARI(Hortonworks), AWS, Tidal, Teradata,Talend Administrator Console, MYSQL, PL/SQL, Python, UNIX, TortoiseGit,MS Visio,TFS.**

**Client: UHG, Minnetonka, MN**

**Duration: April 2012 – Nov 2013.**

**Role: Hadoop Developer/Admin.**

* Build 20 node **Hadoop cluster** from bare-metal hardware
* Configured hardware RAID on name node and **job tracker** nodes. RAID 1 mirror for OS and RAID 5 for data.
* Configured **IPTABLES** to allow required services and block unwanted ports.
* Installed and configured PXE-server, **DHCP, TFTP, HTTP and FTP** as a utility server for bare metal deployment.
* Upgraded the cluster from **CDH3U1 to CDH3U3**. The tasks were first performed on the staging platform before doing it on production cluster.
* Automated all the jobs starting from pulling the data from MySQL to pushing the result setdata to **Hadoop Distributed File System**.
* Implemented **Namenode** backup using **NFS.** This was done for High availability.
* **Used Ganglia** to monitor the cluster around the clock.
* Used **Log4J** for logging purposes.
* Wrote **shell scripts** for log-Rolling day to day processes and it is automated.
* Experience in developing strategies for Extraction, Transformation and Loading (ETL/ELT) mechanism and **UNIX** shell scripting.
* Implemented Capacity schedulers on the **Job tracker** to share the resources of the cluster for the map reduce jobs given by the users.
* Responsible for architecting **Hadoop clusters**.
* Experience in converting **Hive/SQL** queries into **Spark** transformations using Java.
* Develop **high-performance** cache, making the site stable and improving its performance.
* Created a java utility to sort and convert log files to **XML files** based on different parameters.
* Supported Data Analysts in **running Map Reduce P**rograms.
* Worked on importing and exporting data into HDFS and Hive using **Sqoop**.
* Worked on analyzing data with **Hive and Pig**.
* Experience in Implementing Rack Topology scripts to the **Hadoop Cluster**.
* Manage the day-to-day operations of the cluster for **backup and support**.
* Upgraded the **Hadoop Cluster to Cloudera Manager 3.7**.

**ENVIRONMENT**: Hive, Pig, HBase, Spark, Zookeeper, Sqoop, Impala, Java, Hibernate, and spring, Maven, SVN, SQL, Junit, Eclipse, XML, Log4j.

**Client: InfraSoft, Mumbai, India**

**Role: Java/J2EE Developer.**

**Duration: Apr 2010 – Mar 2012.**

**RESPONSIBILITIES**:

* Migrated the code from Mainframes to Java environment and migrating the database from **IBM DB2 to DB2 UDB**.
* Responsible for project initiation documents preparation, **WSR** preparation, and PSM documents preparation
* Application packaging and deploying to all environments in **WebSphere** AppServer.
* Responsible for requirements analysis and carried out code migration and modification.
* Worked extensively with business team to develop the screens in **Java**.
* Used software design specifications to develop and implement the **Spring-Hibernate** persistence solution.
* Involved in **NOSQL** databases like **Hbase, Cassandra** in implementing and integration.
* Participated on Integration testing with business to deliver the project on time.
* Created **Hibernate** OR mapping of the tables and integrated with **spring** (Transaction Management).
* Responsible for **code** review and **re-work**.
* Participated on Unit **testing** with business.
* Database development using **DB2 UDB** database.
* Created the database (**DB2 UDB** for Windows) and tables and import the data into tables.
* Involved on sending the daily status reports and every day offsite call with **lead / manager** to deliver the work status.
* Coding of new business requirements using **Java** and **JDBC**.
* Experience in Setting up and configuring Connect direct, **TWS**, **DSN,** testing load balancers etc.
* Developed Action classes using Struts **MVC** framework.
* Implementation of Java Batch Wrapper that interacts with **UNIX scripts** run by **TWS.**
* Wrote standalone batch programs using the scheduler class for running the application at a desired time period.
* Developed Struts validator **framework** for form validations.
* Designed and developed **JSP** pages.
* Handled multiple user information in the **JSP** session by using session tracking mechanism.
* Used **Javascript** routines in **JSP** pages and performed Client side validations.
* Participated in peer reviews and make sure all the code is migrated.
* Worked on **J2EE** and **Java** Beans and **XML** to migrate the code from Mainframes.

**ENVIRONMENT:** JDK 1.6, Core Java, Hibernate 3.6, Spring 3, Eclipse, tomcat 7.0, JBOSS 5, Unix, drools, Log4j 1.4.2, Oracle11g, Subversion, profiler, scrum.

**Client: Krvy Group, Hyderabad, India.**

**Duration: June 2008 – Mar 2010.**

**Role: Java Developer.**

**RESPONSIBILITIES**:

* Individually worked on all the stages of a Software Development Life Cycle (**SDLC**).
* Implemented the design patterns like **DAO** Factory, Singleton, Service Locator and Business Delegate
* Developed application service components and configured beans using (applicationContext.xml) **Spring IoC**.
* Integrating **Web services** and **working** with data in different servers.
* Designed **User Interface** and the **business logic** for customer registration and maintenance.
* Experience in Creating Tables, Views, Triggers, Indexes, Constraints and functions in **SQL** Server2005.
* Implemented the application using **SpringMVC** Framework which is based on **MVC** design pattern.
* Worked on **Data fixing** and **Bug fixing**.
* Understanding the requirements from **business users** and **end users**.
* Involved in designing and Development of **SOA** services using Web **Services**.
* Working with **XML/XSLT** files.
* Experience creating **UML** class and sequence diagram.
* Used JavaScript code, **HTML** and **CSS** style declarations to enrich websites.
* Worked in content management for versioning and notifications.
* Worked on Integrating the Customer Service software with the Website version.
* Worked on Integration of **UI** interface with **SQL** Server 2005 database.

**Environment: Java, J2EE, Web services, Spring MVC, CSS, HTML,XML, XSTL, SQL, PL SQL, SOAP UI, Eclipse IDE.**