**SudheerR. Bodabanda** 

**Mobile: 614-558-4423**

**Email:** [**sudheerrdy13@gmail.com**](mailto:sudheerrdy13@gmail.com)

|  |
| --- |
| **PROFESSIONAL SUMMARY** |

* Over 7+ years of expertise in various domains and maintenance of the software development lifecycle using Agile and Scrum methodology.
* Worked on systems programming, requirements gathering, technical documentation writing, extensive designing and developing Big Data solutions for different enterprise application systems across the domains.
* Experience in Hadoop 1.x/ 2.x ecosystem components like **HDFS**, **MapReduce**, **ClouderaCHD4/ CHD5, HBase, Oozie, Hive, Sqoop 1.4.5/1.4.6, Pig 0.5.x/0.7.x, Flume**.
* Experience in analyzing data using **HIVEQL**, **PIG** Latin and custom **MapReduce** programs in JAVA. Extending Hive and PIG core Functionality by using custom User Defined Functions.
* Excellent understanding / knowledge of Hadoop architecture and various components such as **HDFS**, **JobTracker**, **Task Tracker, Name Node, Data Node** and **MapReduce**programming paradigm.
* Application Development using **Java**, **Hadoop**, **RDBMS** and **shell scripting** with performance tuning.
* Experience in managing Hadoop clusters and services using **Cloudera Manager**. Proficient in using **Cloudera** Manager, an end-to-end tool to manage Hadoop operations.
* Experience in building and maintaining multiple Hadoop clusters of different sizes and configuration and setting up the rack topology for large clusters.
* Experienced in importing and exporting data into Hadoop distributed file systems (HDFS) and Hive using **Sqoop**, and in loading data to Hive partitions and creating buckets in **Hive**.
* Experienced in building **Pig scripts** to extract, transform and load data onto HDFS for processing.
* Excellent understanding of HDFS, Map Reduce, **YARN**, and tools including Pig and Hive for data analysis, **Sqoop** for data migration, **Flume** for data ingestion, **Oozie** for scheduling and **Zookeeper** 3.5.x for coordinating cluster resources.
* Worked with relational databases like **MySQL** and NoSQL databases like **HBase**.
* Having good knowledge in **Apache Spark**.
* Programmed datasets with Transformations and Actions using **RDDs.**
* Used **Dataframes** to query and join datasets.
* Created **Hive Tables** using Spark **Hive Context**.
* Good Knowledge on **Scalaconcepts** like creating **Arrays**, **Lists**, **Collection Objects, Inheritance** etc.
* Performed **Spark Streaming** on the real time data from various sources.
* Good Knowledge in **AWS EC2, EBS, Trusted Advisor, S3, Cloud Watch, Cloud Front, IAM**, Security Groups, Auto-Scaling.
* Setup and build **AWS infrastructure** various resources, VPC EC2, S3, IAM, EBS, Security Group, Auto Scaling, and RDS in Cloud Formation JSON templates.
* Redesigned infrastructure for high availability using multiple **AWS** availability zones.
* Development, Acceptance, Integration and Production **AWS** Endpoints.
* Good Understanding on **XML** and **JSON** files.
* Experience in Working with Build Tools like **MAVEN** and Continuous Integration Tools like **Jenkins**
* Good working knowledge data visualization using Tableau.
* Experienced in handling **Avro** data files in **MapReduce**programs using Avro data serialization system.
* Experience with **Oozie**Workflow Engine in running workflow jobs with actions that run Hadoop **MapReduce** and Pig jobs.
* Experienced in developing custom UDFs for Pig and Hive using Java by extending core functionality
* Experienced in loading unstructured data into HDFS using **Flume/Kafka**.
* Good hands on experience in NoSQL databases such as **Hbase, Cassandra** and **MongoDB.**

|  |
| --- |
| **TECHNICAL EXPERTISE** |

|  |  |
| --- | --- |
| **Big Data Eco System** | HDFS, Map Reduce, Hive, Pig 0.5/0.7, HBase, Cloudera CDH4, CDH5, Hadoop Streaming, ZooKeeper 3.5.x, Oozie, Sqoop 1.4.5/6, Flume 1.5.x, Kafka, Apache Spark and Scala |
| **No SQL** | HBase,Cassandra |
| **Languages** | Core Java, My SQL 5.6, Shell Scripting, Perl Scripting,Ruby, C/C++, Python |
| **Web Technologies** | HTML 4/5, JavaScript, CSS 2/3, XML, Servlets, Amazon AWS, Google App Engine |
| **Operating system** | Windows, Linux and Unix |
| **DBMS / RDBMS:** | Oracle 11g/10g/9i, SQL Server 2012/2008, MySQL |
| **IDE:** | Eclipse 3/4, Microsoft Visual Studio (2008,2012), NetBeans, Spring Tool Suits |
| **Tools:** | QTP, QC, PUTTY, MySQL Workbench, JUnit, SQL Oracle Developer, WinScp, Tableau, Datameer, Splunk, Tahiti Viewer, Cygwin, Jenkins, MAVEN, REST |
| **Network Protocols** | TCP/IP, UDP, HTTP, DNS, DHCP |

**Client: IRi, Chicago, IL June 2017– Till Date**

**Role: Big Data Developer**

**Description:** IRi is a maketing analytics company where in this project we deals with IRi has more than 150 retailer clients where we will handle the transactional data for those retailers. Here we will get flat files from Data lake and we will store the data into hadoop from linux server.

**Responsibilities:**

* Analyze and define researcher's strategy and determine system architecture and requirement to achieve goals.
* Developed multiple Kafka Producers and Consumers from as per the software requirement specifications.
* Used **Kafka** for log accumulation like gathering physical log documents off servers and places them in a focal spot like HDFS for handling.
* Configured Spark Streaming to get ongoing information from the Kafka and store the stream information to HDFS.
* Experienced in loading and transforming of large sets of structured, semi structured and unstructured data.
* Developed **Spark jobs** and **HiveJobs** to summarize and transform data.
* Involved in converting Hive/SQL queries into Spark transformations using Sparkdataframes, **Scala** and **Python**.
* Expertise in implementing **SparkScala application** using higher order functions for both batch and interactive analysis requirement.
* Experienced in developing Spark scripts for data analysis in both python and scala.
* Wrote Scala scripts to make **spark streaming** work with Kafka as part of sparkKafka integration efforts.
* Built on-premise data pipelines using kafka and spark for real time data analysis.
* Created reports in **TABLEAU** for visualization of the data sets created and tested native Drill, Impala and Spark connectors.
* Developed **shell scripts** to generate the Hive create statements from the data and load the data into the external table.
* Wrote **MapReduce** jobs using Java API and Pig Latin. Optimized Hive QL/Pig scripts by using execution engine like Tez, Spark.
* Involved in writing custom MapReduce programs using java API for data processing.
* Integrated Maven build and designed workflows to automate the build and deploy process.
* Involved in developing a linear regression model to predict a continuous measurement for improving the observation on wind turbine data developed using Spark with Scala API.
* Worked extensively on **Spark, MLlib** to develop a **Logical regression model** on operational Data.
* The Hive tables are created as per requirement were Internal or External tables defined with appropriate static, dynamic partitions and bucketing, intended for efficiency.
* Worked in **AWS** environment for development and deployment of custom Hadoop applications.
* Strong experience in working with **Elastic Map Reduce (EMR)** and setting up environments on **Amazon AWS EC2** instances.
* Load and transform large sets of structured, semi structured data using **Hive**.
* Extract real time feed using **Kafka** and **Spark Streaming** and convert it to **RDD** and process data in the form of Data Frame and save the data as Parquet format in HDFS.
* Used **Spark** and **Spark-SQL** to read the Parquet data and create the tables in Hive using the Scala API.
* Very good understanding on **Cassandra** cluster mechanism that includes replication strategies, snitch, gossip, consistent hashing and consistency levels.
* Imported data from various resources to the Cassandra cluster using Java APIs.
* Configured Performance Tuning and Monitoring for Cassandra Read and Write processes for fast I/O operations and low latency time.
* Used Java API and **Sqoop** to export data into **DataStax**Cassandra cluster from RDBS.
* Strong working experience on Cassandra for retrieving data from Cassandra clusters to run queries.
* Experience in **Data modeling** using Cassandra.
* Experienced in using the **Spark application master** to monitor the Spark jobs and capture the logs for the spark jobs.
* Involved in performing the **analytics and visualization** for the data from the logs and estimate the error rate and study the probability of future errors using regressing models.
* Used **WEB REST API** to make the HTTP GET, PUT, POST and DELETE requests from the webserver to perform analytics on the **Data lake**.
* Worked on High performance computing (HPC) to simulate tools required for the genomics pipeline.
* Used **Mingle** and later moved to**JIRA** for task/bug tracking.
* Used **GIT** for version control

**Environment:** HDP 2.3.4 ,Hadoop, Hive, HDFS, HPC, WEBHDFS, WEBHCAT, Cassandra, Spark, Spark-SQL, spark streaming, KAFKA, Java, Scala, python, Web Server's, JIRA, Maven Build and SBT build.

**Client: Synchrony Financial, Chicago, IL June 2016– May 2017**

**Role: Big Data Developer**

**Description:**This Project deals with ingestion of data into Hadoop ecosystem from various sources that include external and internal sources like Teradata and DB2 databases. Main aim of the project is to centralize the source of data that can be consumed for analytical purpose by the end user or application teams or for reporting purposes.

**Responsibilities**:

* Installed and configured **Hadoop** developed multiple **MapReduce** jobs in Java for data cleaning and preprocessing.
* Installed and configured **MapReduce**, **HIVE**, **HDFS** and implemented CDH4 Hadoop cluster on **CentOS**.
* Imported data using **Sqoop** from Teradata using Teradata connector. Load and transform large sets of structured, semi structured and unstructured data.
* Creating **Hive tables** and working on them for data analysis in order to meet the business requirements.
* Executed Hadoop commands on Hive Shell and implemented Hive DDls like **partitions** and **Buckets**.
* Created **Hive** User defined funtions (**UDfs**) in Hive as per the client’s requirements.
* Designed and implemented MapReduce-based large-scale parallel relation-learning system.
* Installed and benchmarked **Hadoop/HBase clusters** for internal use written HBASE Client program in Java and web services.
* Created Dynamic columns and **hfiles** to store them in the hadoop cluster.
* Participated in design and development of scalable and custom Hadoop solutions as per dynamic data needs.
* Worked on large sets of **structured**, **semi-structured** and **unstructured** data.
* Programing **MapReduce Jobs** on Data Aggregation, Indexing, searching, clustering and Text analysis with appropriate business logics.
* Installed and configured **Pig** for ETL jobs and used **Pig Scripts** for implementing MapReduce Jobs on a large scale.
* Done Data ingestion using **Flume** and imported large data sets to the Hadoop environment.
* Utilized **Oozie** for the Job workflow scheduling and troubleshooting the encountered issues.
* Utilization of hadoop admin commands and utilities like **Balancer**, **fsck**, **dfsadmin** and **distcp** and maintaining the clusters.
* Taking **Metadata Backups** and maintaining them in a company based repository at regular intervals and troubleshooting the cluster issues.
* Maintaining Directories and space quota for multi users across the systems using dfsadmin commands.
* Troubleshooting the cluster by reviewing Hadoop Log files.
* Created tables, stored procedures in SQL for data manipulation and retrieval, Databas
* Performed Transformations and Actions on various datasets using **RDDs**.
* Programming **spark streaming** on the real time data and processed the data by applying the business logic with accurate results.
* Hands-on experience on **Hue** Distribution and performed various actions on the real time applications datasets.
* Processing various datasets with **scala** programs and implementing business logic on them.
* Debugging various scala codes and doing code walkthrough to achieve the company formats and standards.
* Extensively involved in Business Analysis and Requirements Gathering.
* Involved in Developing Test Plans and Developed Test Cases/Test Strategy with input from the assigned Business Analysts.
* Analyzed data sources, prepared data conversion approach to load the data.
* Worked on profiling the data for understanding the data and mapping document.
* Designed the table structures and indexed them to pulled out the data fast.
* Worked in importing and cleansing of data from various sources like Oracle, flat files onto SQL Server with high volume data.
* Managed **Amazon web services** like EC2, S3, RDS, EBS, ELB, Auto-Scaling, AMI, IAM, SQS, SNS, Dynamo DB, Cloud watch through AWS.
* Complete understand of configuring EC2, VPC, automated stack configuration and Monitoring software for the cloud environment.
* **Build servers using AWS**: Importing volumes, launching EC2, RDS, creating security groups, auto-scaling, load balancers (ELBs) in the defined virtual private connection.
* Creating **alarms** in Cloud watch service for monitoring the servers performance, CPU Utilization, disk usage etc.

**Environment**: Hadoop 1.x, MapReduce, HDFS, Hive, J2EE 7, ClouderaCHD4, Pig, HBase, Flume, Oozie, Linux, XML, SQL, J2EE 7, Eclipse 3.7, PL/SQL, SQL\*PLUS, AWS, Apache Spark.

**Client: World Financial Group, Milwaukee, WI Nov 2014– May 2016**

**Role: Hadoop developer**

**Description:**As a Hadoop Developer, am responsible for building scalable distributed data solutions using Hadoop. Collection and Downloading of data generated from different sources to HDFS. Performed necessary transformations and aggregation to build the common learner data model in NoSQL store (HBase). Used Pig, Hive and MapReduce for analyzing the various kinds of data and information.

**Responsibilities:**

* Installed and configured **Hadoop** Developed multiple **Map Reduce** jobs in Java for data cleaning and preprocessing.
* Worked on data Importing and exporting into **HDFS** and **Hive** using **Sqoop**.
* Participated in development/implementation of **Cloudera** Hadoop environment.
* Load and transform large sets of structured, semi structured and unstructured data.
* Experienced in managing and reviewing Hadoop log files.
* Worked on automation of delta feeds from, Teradata using **Sqoop**, also from FTP Servers to Hive.
* Exported the analyzed data to the relational databases using **Sqoop** for visualization and to generate reports for the BI team.
* Developed **Map Reduce** programs to cleanse the data in **HDFS** obtained from heterogeneous data sources to make it suitable for ingestion into **Hive schema** for analysis
* Used **Sqoop** to import the data from RDBMS to Hadoop Distributed File System (HDFS) and later analyzed the imported data using Hadoop Components
* Established custom **Map Reduce** programs in order to analyze data and used **Pig Latin** to clean unwanted data.
* Did various Performance tuning like using distributed cache for small datasets, Partition, Bucketing in hive and Map Side joins
* Involved in loading and transforming large sets of Structured, Semi-Structured and Unstructured data and analyzed them by running **Hive** queries and **Pig** scripts
* Worked on **Storm** and **Kafka** tools to stream Real time data. And also, performed operations on Statistical analysis.
* Participated in requirement gathering from the Experts and Business Partners and converting the requirements into technical specifications
* Implemented daily workflow for extraction, processing and analysis of data with **Oozie**.
* Involved in loading data from **LINUX** file system to **HDFS.**
* Migrated applications to the **AWS** cloud.
* Implemented **AWS** solutions using EC2, S3, RDS, EBS, Elastic Load Balancer, Auto scaling groups, Worked on DMZ and non-DMZ (Local) zones of servers.
* Installed operating system on multiple machines using **Kick Start**. Work with different team members for automation of Release components.
* Resolved system issues and inconsistencies in coordination with quality assurance and engineering teams.
* Troubleshoot the build issue during the Jenkins build process.
* Actively participate in various internal meet / sessions with other project team members to understand the requirements and suggest suitable Hadoop Stack Technologies along with training sessions.

**Environment**: Cloudera HDFS, Hadoop, Map Reduce, Hive, Pig Latin, Java, SQL, Sqoop, Centos, NOSQL

database, Storm, Kafka, AWS.

**Client: Landis + Gyr, Alpharetta, GA          May 2013 – Oct 2014**

**Role: Hadoop Developer**

**Description:** The project was intended to convert already existing Risk Manager process to use Hadoop Eco System components. The goal of the project was to achieve scalability, Flexibility, faster processing speeds, fault tolerance and cost effective system to replace already existing high maintenance semi-automated process.

**Responsibilities:**

* Gathered business requirements from the Business Partners and Subject Matter Experts.
* Loaded and transformed large sets of structured, semi structured and unstructured data in various formats like text, zip, XML and JSON.
* Wrote **MapReduce** jobs to perform big data analytics on ingested data using Java API.
* Installed and configured Pig and also written **Pig Latin** scripts.
* Imported data using **Sqoop** to load data from Oracle to HDFS on regular basis or from Oracle server to HBase depending on requirements.
* Implemented advanced procedures like text analytics and processing using the in-memory computing capabilities like **Spark.**
* Enhanced and optimized product **Spark** code to aggregate, group and run data mining tasks using the Spark framework.
* Wrote MapReduce in **Ruby** using Hadoop Streaming to implement various functionalities.
* Created Hive tables and worked on them using **Hive QL**.
* Loaded the data into **Cassandra** using **CQL** and carefully generated the rowkey.
* Built web portal using JavaScript, which makes a **REST API** call to the elastic search and gets the rowkey.
* Built Hive table on top of Cassandra so that it can point to the reporting tools like **Tableau, Datameer, Splunk**.
* Used **netezza** for low latency queries by loading the data into netezza from big data cluster using **nzload**.
* Responsible for the Plugin Management, User Management, Build/Deploy Pipeline Setup and End-End Job Setup of all the projects.
* Version control through **GIT**. Writing new plugins in **Nagios** to monitor resources. Working in implementation team to build and engineer servers on Ubuntu and RHEL Linux. Provisioning virtual servers on VMware and ESX servers using **Vcloud.**
* Connected continuous integration system with **GIT** version control repository and continually build as the check-in's come from the developer.
* **Jenkins** is used as a continuous integration tool for automation of daily process.

**Environment:** Cassandra, Netezza, Pig, Hive, Map Reduce, Sqoop, Ruby, JavaScript, Apache Spark, Tableau, Talend, Datameer, Splunk, Pentaho, SFTP, Jenkins, GIT, AWS.

**Client:Hewlett Packard- Columbus,OH Jan 2012 to Apr 2013**

**Role:Java Developer**

**Description:**The project is to maintain manual details of blood camps, donor details, blood stock, blood types, instruments stock, bill generation etc. 'Red Cross Online' is an intranet system application to be used by doctors and technicians to reduce daily activities time, i.e. Register donor, blood camp details in the Red Cross Blood Bank. The system provides cross matching details, grouping details, bill generation, certificates for Blood Donors and organizations and information on blood stock, furniture and equipment's stocks

**Responsibilities:**

* Responsible for gathering **Business Requirements** and **User Specifications** from Business Analyst.
* Involved in developing, testing and deployment of the application using **Web Services**.
* Worked on Load Builder Module for Region services using **SOAP Web services**.
* Worked with **Servlets, JSP and Ajax** to design the user interface.
* Used JSP, Java Script, HTML5, and CSS for validating and customizing error messages to the User Interface.
* Used Eclipse 3.5 IDE for code development and deployed in Web Logic Server.
* Worked on **MVC framework, STRUTS 2.0** and used **Spring** dependency injection for application customization and upgrade.
* **UML diagrams** were used to create use cases and Sequence diagram.
* Implemented **Hibernate** in the data access object layer to access and update information in the Database.
* Wrote **PL/SQL queries**, stored procedures, and triggers to perform back-end database operations.
* Used multiple Action Controllers to control the page flow.
* Used Interceptors for client validations.
* Used **Subversion** for version control and log4j for logging errors.
* Developed **ANT scripts** to build and deploy the application.
* Wrote test cases in **JUnit** for unit testing.

**Environment:** Java 6, J2EE 5, Struts 2.0, Hibernate 3.0, MVC , WebLogic Application Server 10.3, UML, JSP, Servlets, Java Script, HTML5, CSS, Ajax, Web Services , Oracle 10g, Eclipse 3.5 IDE, PL/SQL, ANT, Junit, XML/XSL, log 4j 1.2.15.

**Ecolite Technologies May 2010 to Nov 2011**

**Java Developer**

**Description:**Ecolite Technologies is a certified company with inhouse design, manufacturing & testing facilities located in Gurgaon, India providing eco-friendly and energy saving green lighting solutions using LED based lighting products. Energy saving is a key focus of the current social, economic and political policy across both the developed and developing world.

**Responsibilities:**

* Prepared the Installation, Customer guide and Configuration document which were delivered to the customer along with the product.
* Worked on **JSP, Servlets and JDBC** in creating web components.
* Responsible for creating work model using **HTML and JavaScript** to understand the flow of the web application and created class diagrams.
* Participated in the daily stand up **SCRUM** agile meetings as part of **AGILE** process for reporting the day to day developments of the work done.
* J2EE is used to develop the application based on MVC architecture.
* Used HTML, XHTML, JavaScript, JQuery, DHTML and Ajax to improve the interactive front end.
* Used **EJB entity** and **session beans** to implement business logic and session handling and transactions.
* Designed, Implemented, Tested and Deployed Enterprise **Java Beans using Web**Logic as Application Server.
* Designed the database tables and indexes used for the project.
* Developed **stored procedures, packages and database triggers** to enforce data integrity.
* JDBC API was used with Query Statements and Prepared Statements to interact with the database using SQL.
* Used **SAX and DOM XML parsers** for data retrieval.
* Performed data analysis and created reports for user requirements.

**Environment:** Windows NT 2000/2003, XP, and Windows 7/ 8 C, Java, JSP, Servlets, JDBC, EJB, DOM, XML, SAX