**SUJITHA ATTALURI**

|  |  |
| --- | --- |
| **Status** | **H1B** |
| **Contact Number** | **314-384-9963** |
| **Alternate Contact Number** | **314-384-9963** |
| **Current location of the candidate (City/State)** | **New York city** |
| **Email Address** | [**sujitha.hadoop@gmail.com**](mailto:sujitha.hadoop@gmail.com) |
| **Skype ID** | [**sujitha.hadoop@gmail.com**](mailto:sujitha.hadoop@gmail.com) |

**PROFESSIONAL SUMMARY**

**8+Years of experience in the field of Information Technology which includes a major concentration on Big Data Tools and Technologies, various Relational Databases and NoSQL Databases, Java Programming language and J2EE technologies with highly recommended software practices.**

* Having **4+** years of experience in the field of software development creating solutions using Enterprise Applications and Web based Applications using **JAVA & J2EE Technologies.**
* Having **4** years of experience as a **Big Data Engineer** with good understanding of Hadoop framework, **Big Data Tools** and Technologies for implementing Data analytics.
* **Hadoopdeveloper**: Excellent hands on experience using Hadoop tools like **HDFS, Hive, Pig, Apache Spark, Apache Sqoop, Flume, Oozie, Apache Kafka, Apache storm, Yarn, Impala, Zookeeper, Hue.** Experience in analyzing data using **HiveQL**, **Pig Latin**, and **MapReduce** Programs.
* Experienced in ingesting data into **HDFS** from various Relational databases like **MYSQL**, **Oracle**, **DB2**, **Teradata** using **sqoop**.
* Experienced in importing real time streaming logs and aggregating the data to **HDFS** using **Kafka** and **Flume**.
* Excellent knowledge on creating real-time data streaming solutions using **Apachestorm**, **sparkstreaming** and building **sparkapplications** using **scala**.
* Well versed with various Hadoop distributions which includes **apache Hadoop**, **cloudera**, **Hortonworks** and knowledge on **MAPR** distribution.
* Experienced in creating various tables in **Hive** which include Managed Tables and External tables and loading data into **Hive** from **HDFS**.
* Implemented **Bucketing** and **partitioning** concepts in Hive and also wrote **UDF’s** in order to create user defined functionalities.
* Implemented **Pig scripts** for analyzing large data sets in the **HDFS** by performing various **transformations**.
* Experience in analyzing data using **HiveQL**, **PigLatin**, **HBase**.
* Capable of processing large sets of structured, semi-structured and unstructured data and supporting system application architecture.
* Experience working on **NoSQL** Databases like **HBase, Cassandra and MongoDB.**
* Experience in **Python**, **Scala, shell scripting.**
* Experience in Creating various **Oozie** jobs to manage processing **workflows**.
* Experience in using **AWS Cloud components S3, EC2, Elastic beanstalk and DynamoDB.**
* Experience in using various file formats including **XML, JSON, CSV** and other file formats like **text**, **sequence files**, **avro**, **ORC** and **Parquette** using various compression techniques like **snappy,LZO**.
* Experience with Testing Map Reduce programs using **MRUnit, Junit and EasyMock**.
* Knowledge on Machine Learning algorithms and Predictive Analysis using **sparkMLLib**,**Mahout**.
* Worked on data visualization tools like **Tableu** and also integrated the data using ETL tool**Talend.**
* Worked on various Relational Databases like **Postgres, MySQL, Oracle 10g, DB2.**
* Created java applications which are used to connect to database using **JDBC**, **JSP**,**Spring** and **Hibernate**.
* Knowledge on design, development of web based applications using **HTML, DHTML, CSS, JavaScript, JQuery, JSP and Servlets.**
* Experience on various build tools like **ANT**, **MAVEN**, **Graddle**, **SBT**.
* Knowledge on creating dashboards/reports using reporting tools like **Tableu, Qlickview.**
* Development experience with **IDE’s** Eclipse, NetBeans, IntelliJ and repositories SVN, GIT and CVS.
* Having good experience in different software methodologies like **waterfall** and **agile** approach.
* Knowledge on writing **YARN** applications.
* Passionate about working on the most cutting-edge Big Data technologies.
* Ability to adapt to evolving technology, strong sense of responsibility and accomplishment.
* Willing to update my knowledge and learn new skills according to business requirement.

Educational Details : Bachelors in Computer science from Andhra University 2004 - 2008

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| Hadoop Technologies | HDFS, MapReduce, Hive, Impala, Pig, Sqoop, Flume, Oozie, Zookeeper, Ambari, Hue, Spark, Strom, Kafka, Yarn, Ganglia, TEZ |
| Operating System | Windows, Unix, Linux |
| Languages | Java, J2EE, SQL, PL/SQL, Shell Script, Python, scala |
| Testing tools | Junit, MRunit, EasyMock |
| Front – End | HTML, JSTL, DHTML, JavaScript, CSS, XML, XSL, XSLT |
| SQL Databases | MySQL, Oracle 11g/10g/9i, SQL Server |
| NoSQL Databases | HBase, Cassandra, MongoDB, Neo4j |
| File System | HDFS |
| Reporting Tools | Tableau, Qlickview |
| IDE Tools | Eclipse, NetBeans, Spring Tool Suite, IntelliJ |
| Application Server | IBM WebSphere, Web Logic, JBoss |
| Version control | SVN, GIT and CVS |
| Build Tools | Maven, Graddle, ANT,SBT. |
| ETL Tools | Talend, Datastage, Informatica. |
| Messaging & Web Services Technology | SOAP, WSDL, REST, UDDI, XML, SOA, JAX-RPC,  IBM WebSphere MQ v5.3, JMS. |

**Work experience:**

**Client: Brown Brothers Harriman & Co-NYFeb 2015 –Till date**

**Role: Big Data Developer**

Brown Brothers Harriman & Co. is the oldest and one of the largest private banks in the united states which provides commercial banking, investment management and trust services to private companies and individuals.

**Responsibilities:**

* Responsible for building scalable distributed data solutions using**Hadoop.**
* Importing different log files using **Apache Kafka** into **HDFS** and performed data analytics using **spark.**
* Involved in importing the data from various data sources into **HDFS** using **Sqoop** and applying various transformations using **Hive**, **Spark** and then loading data into **Hive** tables.
* Involved in requirement and design phase to implement Streaming **Lambda Architecture** to use real time streaming using **Spark** and **Kafka.**
* Collected the logs from the physical machines and the OpenStack controller and integrated into **HDFS** using **kafka.**
* Involved in converting **Hive/SQL** queries into spark transformations using **spark RDDs** and **python.**
* Experience in developing various **Spark applications** using **python**. (**pyspark**).
* Developing **spark code** to applying various **transformations and actions** for faster data processing.
* Used **Spark Stream** processing to get data into in-memory, implemented **RDDtransformations**, and performed **actions**.
* Developed various **Kafka** Producers and consumers for importing various transaction logs.
* Used **Zookeeper** to store offsets of messages consumed for a specific topic and partition by a specific Consumer Group in **Kafka**.
* Involved in integrating **HBase** with **Spark** to import data into **HBase** and also performed some CRUD operations on **Hbase**.
* Used various **HBase** commands and generated different Datasets as per requirements and provided access to the data when required using **grant** and **Revoke**
* Performed performance tuning and troubleshooting of **Map Reduce** jobs by analyzing and reviewing **Hadoop** log files.
* **Sqoop** jobs, **PIG** and **Hive** scripts were created for data ingestion from relational databases to compare with historical data.
* Experience in working with Elastic MapReduce(**EMR**) and setting up environments on amazon AWS EC2 instances.
* Experienced in migrating **HiveQL** into **Impala** to minimize query response time.
* Knowledge on handling **Hive** queries using Spark SQL that integrates with Spark environment.
* Executed **Hadoop/Spark** jobs on **AWS EMR** using programs, data stored in **S3 Buckets**.
* Loaded and performed some transform data into Hadoop cluster from large set of structured data using **TalendBig data studio**.
* Worked with different File Formats like **textfile**, **avro**, **orc** for HIVE querying and processing based on business logic.
* Experience in pulling the data from **Amazon S3** bucket to data lake and built **Hive** tables on top of it and created data frames in **Spark** to perform further analysis.
* Involved in writing Custom **Talend** jobs to ingest, enrich and distribute data in Hadoop ecosystem.
* Worked on Sequence files, RC files, Map side joins, bucketing, partitioning for **Hive** performance enhancement and storage improvement.
* Developed efficient ETL processes, including workflows and jobs scheduling, to move data from source to target according to requirements using **Talend**.
* Implemented **Hive**, **Pig UDF's** to implement business logic and Responsible for performing extensive data validation using **Hive**.
* Implemented Daily **Cron jobs** that automate parallel tasks of loading the data into **HDFS** using **Oozie** coordinator jobs.
* Involved in loading the structured and semi structured data into spark clusters using **Spark SQL** and **Data Frames API.**
* Used **Pig** as ETL tool to do transformations, event joins, filter and some pre-aggregations
* Used Visualization tools such as Power view for excel, Tableau for visualizing and generating reports.
* Written multiple **Map Reduce** programs for data extraction, transformation and aggregation from multiple file formats including **XML**, **JSON**, **CSV**& other compressed file formats.
* Used **Talend** Open Studio for data integration and for data migration from various location across the business.
* Integrated data quality plans as a part of ETL processes using **Talend.**
* Experience in build scripts using **Maven** and did continuous system integrations like **Jenkins**.
* Used **JIRA** for bug tracking and **GIT** for version control.
* Involved in story-driven **agile** development methodology and actively participated in daily scrum meetings.

**Environment**: **Cloudera, Map Reduce, HDFS, Pig, Hive, Sqoop, Spark, Kafka, Oozie, Java, Linux, Maven, HBase, Zookeeper, Tableau, TalendOpen studio,AWS.**

**Client: HighPoint Solutions-Philadelphia, PAMay 2014 – Nov 2015  
Role: Big Data Engineer.**

HighPoint solutions deliver practical IT strategies & solutions to various health care organizations. HighPoint Solutions also introduces various Big Data Services to its clients which results in improvements.

**Responsibilities:**

* Developed various data loading strategies and performed various transformations for analyzing the datasets by using **Hortonworks** Distribution for Hadoop ecosystem.
* Worked in Loading and transforming large sets of structured, semi structured and unstructured data.
* Involved in collecting, aggregating and moving data from servers to **HDFS** using **Flume**.
* Collecting data from various Flume agents that are imported on various servers using **Multi-hop Flow**.
* Knowledge on various **flume** sources, channels and sink by which data is ingested into **HDFS**
* Responsible for performing various transformations like sort, join, aggregations, filter in-order to retrieve various datasets using **spark.**
* Experience in extracting appropriate features from datasets in-order to handle bad, null, partial records using **spark SQL.**
* Experience in developing various **spark** application using **Spark-shell**(Scala).
* Involved in creating **Hive Tables**, loading with data and writing **Hive queries** which will invoke and run **MapReduce** jobs in the backend.
* Designed and implemented **Incremental Imports** into **Hive** tables and writing Hive queries to run on **TEZ**.
* Written **Hive** jobs to parse the logs and structure them in tabular format to facilitate effective querying on the log data.
* Migrated ETL jobs to **Pig** scripts do Transformations, even joins and some pre-aggregations before storing the data onto **HDFS**.
* Involved in writing optimized **Pig Script** along with developing and testing Pig Latin Scripts
* Implemented the workflows using Apache **Oozie** framework to automate tasks
* Worked on different file formats like Sequence files, **XML** files and Map files using **MapReduce** Programs. Worked with **Avro** Data Serialization system to work with JSON data formats.
* Exported data to **Cassandra**(NoSQL) database from **HDFS** using **sqoop** and performed various **CQL** commands on Cassandra to obtain various datasets as required.
* After performing all the transformations data is stored in **MongoDB**(**NOSQL**)using **Sqoop.**
* Created and imported various collections, documents into **MongoDB** and performed various actions like query, project, aggregation, sort, limit.
* Involved in Unit testing and delivered Unit test plans and results documents using **Junit** and **MRUnit**.
* Automated all the jobs to pull the data and load into Hive tables, using **Oozie** workflows
* Analyzed the data by performing **Hive queries** and running **Pig scripts** to study customer behavior.
* Created and maintained Technical documentation for launching Hadoop Clusters and for executing **Pig** Scripts.

**Environment**: **Hadoop, HDFS, Map Reduce, spark, Sqoop, Oozie, Pig, Hive, Flume, TEZ, LINUX, Java, Eclipse, Cassandra,MongoDB.**

**Client: VWR International-PA June 2013 - March 2014   
Role: Hadoop Developer**

VWR International is basically involved in distribution of research laboratory products. It has various suppliers from different sectors like biotechnology, life sciences and pharmaceutical.

**Responsibilities:**

* Responsible for loading the customer's data and event logs from MSMQ into **HBase** using Java API.
* Created **HBase** tables to store variable data formats of input data coming from different portfolios.
* Involved in adding huge volumes of data in columns to store data in **HBase**.
* Used **Sqoop** for transferring data from **HBase** to **HDFS** and vice versa.
* Responsible for architecting **Hadoop** clusters with **CDH4** on CentOS, managing with Cloudera Manager.
* Involved in initiating and successfully completing Proof of Concept on **FLUME** for Pre-Processing, Increased Reliability and Ease of Scalability over traditional MSMQ.
* Used **Flume** to collect the log data from different resources and transfer the data type to hive tables using different **SerDe’s** to store in **JSON**, **XML** and Sequence file formats.
* Managing and scheduling Jobs to remove the duplicate log data files in **HDFS** using **Oozie** workflows.
* Involved in creating **Hive** tables, loading data and running hive queries in those data.
* Extensive Working knowledge of **partitioned table, UDFs, performance tuning, compression-related properties, thrift server** in **Hive**.
* Used **Hive** to find correlations between customer's browser logs in different sites and analyzed them to build risk profile for such sites.
* End-to-end performance tuning of Hadoop clusters and **MapReduce** routines against very large data sets.
* Developed the **PigUDF'S** to pre-process the data for analysis.
* Monitored Hadoop cluster job performance and performed capacity planning and managed nodes on Hadoop cluster.
* Proficient in using **ClouderaManager**, an end to end tool to manage **Hadoop** operations.

**Environment: Hadoop (CDH4), BigData, HDFS, Pig, Hive, MapReduce, Sqoop, Cloudera manager, LINUX, FLUME, HBase, Pig, Hive**

**Client: Hexaware Technologies, Atlanta GA Nov 2012– April 2013**

**Role: Java Developer**

**Responsibilities:**

* Used **Hibernate ORM** tool as persistence Layer - using the database and configuration data to provide persistence services (and persistent objects) to the application.
* Implemented **Oracle Advanced Queuing** using **JMS** and Message driven beans.
* Responsible for developing **DAO** layer using Spring MVC and configuration XML's for Hibernate and to also manage **CRUD** operations (insert, update, and delete).
* Implemented Dependency injection of spring frame work.
* Developed and implemented the **DAO** and service classes.
* Developed reusable services using BPEL to transfer data.
* Participated in Analysis, interface design and development of **JSP**.
* Configured **log4j** to enable/disable logging in application.
* Developed Rich user interface using **HTML**, **JSP**, **AJAX**, **JSTL**, **JavaScript**, **JQuery** and **CSS**.
* Developed data mapping to create a communication bridge between various application interfaces using **XML**, and **XSL**.
* Responsible for deploying the application using **WebSphere Server** and worked with **SOAP**, **XML** messaging.
* Implemented **PL/SQL** queries, Procedures to perform data base operations.
* Wrote **UNIX** Shell scripts and used **UNIX** environment to deploy the EAR and read the logs.
* Used **JUnit** to develop Test cases for performing Unit Testing.
* Used Building tools like **Maven** to build, package, test and deploy application in the application server.
* Actively involved in code review and bug fixing for improving the performance.
* Involved in code deployment activities for different environments.
* Agile Scrum Methodology been followed for the development process.

**Environment: Java, Spring, Hibernate, JMS, EJB, Web logic Server, SQL Developer, Maven,XML, CSS, JavaScript, JSON.**

**Client:GLOBAL ENERGY CONSULTING ENGINEERS(GECE) Jan 2011 –Oct 2012 Role: JAVA/J2EE Developer**

**Responsibilities:**

* Responsible and active in the analysis, design, implementation and deployment of full software development lifecycle(**SDLC**) of the project.
* Designed and developed user interface using **JSP, HTML** and **JavaScript.**
* Developed **Struts** action classes, action forms and performed action mapping using **Struts** framework and performed data validation in form beans and action classes.
* Extensively used **Struts framework** as the controller to handle subsequent client requests and invoke the model based upon user requests.
* Defined the search criteria and pulled out the record of the customer from the database. Make the required changes and save the updated record back to the database.
* Validated the fields of user registration screen and login screen by writing **JavaScript** validations.
* Developed build and deployed scripts using **Apache ANT** to customize WAR and EAR files.
* Used **DAO** and **JDBC** for database access.
* Developed stored procedures and triggers using **PL/SQL** in order to calculate and update the tables to implement business logic.
* Design and develop **XML** processing components for dynamic menus on the application.
* Involved in post-production support and maintenance of the application.

**Environment: Oracle, Java , Struts, Servlets, HTML, XML, SQL, J2EE, JUnit, Tomcat .**

**Client: NCR Corporation – India Oct 2008- Dec 2010**

**Role: Jr. Java / Web Developer**

**Responsibilities:**

* Implemented the project according to the Software Development Life Cycle (**SDLC**).
* Analyzing and Preparing the requirement Analysis Document.
* Involved in developing Web Services using **SOAP** for sending and getting data from external interface.
* Involved in requirement gathering, requirement analysis, defining scope, and design.
* Worked with various **J2EE** components like **Servlets**, **JSPs**, **JNDI,JDBC** using Web Logic Application server.
* Involved in developing and coding the **Interfaces** and **classes** required for the application and created appropriate relationships between the **system classes** and the **interfaces** provided.
* Assisting project managers with drafting use case scenarios during the planning stages.
* Developing the **Use Cases**, **Class Diagram**s and **Sequence Diagrams**.
* Used **Java Script** for client-side Validation.
* Used **HTML, CSS, JavaScript** for create web pages.
* Involved in Database design and developing SQL Queries, stored procedures on **MySQL.**

**Environment:Java, J2EE, JDBC, HTML, CSS, JavaScript, Servlets, JSP, JDBC, Oracle, Eclipse, Web Logic, MySQL.**