**Name: Haindavi**

**Mobile Number: 512-221-2402**

* Over **10+ years** of IT experience in designing and developing big data applications using **Apache Hadoop, HDFS, MapReduce, HBase, Hive, Oozie, Tez, Yarn, Sqoop and Spark.**
* Extensive experience with analysis, design, development, customizations and implementation of Big Data and Data warehousing applications.
* Proficient in analyzing and translating business requirements to technical requirements and architecture.
* Good exposure to Proof of Concept (POC) design and solution designing and architecture design.
* Good leadership abilities, multi-tasking, excellent interpersonal, analytical skills, self-motivated, quick learner, team player, and vendor management
* Actively involved in all phases of software development life cycle including functional specifications, prototypes and documentation. Worked effectively in agile software development environment.
* Experience in writing system specifications, translating user requirements to technical specifications.
* Expertise in developing Spark SQL and batch applications for big data pipelines using HBase, Hive, Avro, parquet.
* Work experience in solving BIG DATA problems using Apache Hadoop (YARN, MapReduce, HDFS) and Ecosystems (Apache HBase, Hive, PIG Latin, Sqoop, Flume, Oozie, Spark, Avro, Zookeeper).
* Hands on experience in writing **Hive Query** Language and optimizing hive queries.
* Experience working on NoSQL databases like **HBase**.
* Developed MapReduce programs in Java, Queries in Hive (HQL) and UDF’s in Core Java.
* Experience in creating spark applications in both Scala and Python Context.
* Actively Worked in Sqoop to move Structured Data from multiple databases to HDFS.
* Expertise in writing MapReduce programs in Java, PIG Latin, HQL, Shell scripting, SQL, PL/SQL, Core Java.
* Working knowledge in setting up Hadoop (pseudo-distributed and Multi Node) Cluster and configuring cluster properties.
* Experience in implementing ETL/ELT processes with MapReduce, PIG, Hive.
* Experience in creation of Database objects like Stored Procedures, Functions and Triggers in database languages like PL/SQL, PostgreSQL.
* Experience in performance tuning, SQL Tuning, partitioning the data and creation of indexes for faster database access and better query performance.
* Experience in **Oozie** and workflow scheduler to manage Hadoop jobs by Direct Acyclic Graph (DAG) of actions with control flows.
* Experience in Importing and exporting data from different databases like MySQL, RDBMS into HDFS and HBASE using **Sqoop**.
* Experience in Performance Tuning, Backup and Recovery process, and product support on various platforms. SQL Tuning and creation of indexes for faster database access and better query performance.
* Working knowledge on Crontab for Batch scheduling.
* Experience in optimizing queries for maximum throughput, benchmarking, providing Proof of concept for Enterprise architects.
* Working experience in Map Reduce programming model and Hadoop Distributed File System.
* Strong technical and architectural knowledge in solution development.
* Effective in working independently and collaboratively in teams.
* Good analytical, communication, problem solving and interpersonal skills.
* Flexible and ready to take on new challenges.
* Self-starter and team player, capable of working independently and motivating a team of professionals.

**TECHNICAL SUMARY:**

|  |  |
| --- | --- |
| Big Data Stack | Spark, HBase, Hive, Sqoop, Flume, Hadoop, HDFS, Mapreduce, YARN, PIG Latin, Oozie, Avro, Parquet, Zookeeper, Hue. |
| Programming Languages | Core Java, Unix shell scripting, SQL, PL/SQL, HQL |
| RDBMS | Oracle 10g/9i/8i, MySQL, SQL server, Teradata |
| Automation | Crontab |
| Server Tools | WINSCP, SSH, Putty |
| Operating System | Windows 95/98/2000/XP, Windows NT, MS-DOS, UNIX, LINUX |
| Others | Eclipse, JDK 1.7/1.8, Oracle SQL Developer, TOAD, ODBC, HTML, Service NOW |

**EDUCATION:**

* **MCA** from JNTU in 2011
* **BSC** from OU in 2008

**PROFESSIONAL EXPERIENCE**

**Client: CBRE May’2022 – March’2024**

**Role: Senior Software Engineer**

**Responsibilities:**

* Gather functional requirements by analyzing the use cases to ingest data from variety of sources in streaming as well as in batch mode to create Hadoop Data Lake.
* Build the Modern Data Architecture (MDA) Pipeline using appropriate Hadoop technologies that best fits the use case ensuring performance.
* Build code to create the workflows, instrumentation and audit as per business requirements to avoid Data Loss / Data Duplication and to improve Data Accuracy.
* Build Connectors API to import and export bulk data from databases and other systems.
* Integrate data from different sources to provide unified view of the combined data to users.
* Create file Watcher to continuously look for any new legacy files from Members in Local File system and move to MDA SFTP structure.
* Actively worked in Hive for data warehousing, cleansing data, generating reports and adhoc historic reports.
* Created Managed/External tables in Hive and tuned performance of the HQLs by creating partitions and buckets on Hive tables.
* Optimizing Hive Queries to improve performance.
* Load the data into HBase tables for UI web application.
* Designed and created Hive external tables using shared meta-store instead of derby with partitioning, dynamic partitioning, and buckets.
* Building scripts to Monitor System health and logs and respond accordingly to any warning or failure conditions.
* Develop Parser code to Parse Member files into unified format (Parquet) based on the metadata stored in SQL Server and egress parsed data (Parquet) to HDFS.
* Build Service to load Historical data to MDA Data Lake.
* Build Sqoop Scripts to copy data from RDBMS to HDFS and vice versa.
* Build Hive User Defined Functions (UDF), User Defined Aggregated Functions (UDAF) and User Defined Tabular Functions (UDTF).
* Implement Tool to generate reports by comparing Legacy Output Data with MDA Output Data.
* Build big data batch data processing pipelines to Extract, Load, Validate and Transform data with high processing speed by taking advantage of Spark in-memory Computation.

**Environment: MapR, Hadoop 2.1, HDP 3.x, CDP 7.x, HDFS, Spark, Hive, Sqoop, Shell Scripts.**

**Client: OPTUM (United Health Group) Aug’2016 – May’2022**

**Role: Senior Software Engineer**

**Responsibilities:**

* Importing and exporting data from MySQL into HDFS using Sqoop as flat files into HDFS.
* Building HiveQL scripts to create tables, load data and query tables in a Hive.
* Used Spark SQL to create Data frame from Hive data, applied transformations based on business logic, loaded data back to Hive tables/Views. Also, Used Spark SQL to process JSON files from ingestion, register them as temp tables and loaded data to Hive tables/views.
* Created Hive external tables to read the persisted HBase tables using HBase SerDe properties.
* Created Hive views on top of HBase table (single column family, different qualifier) which has different data model versions.
* Expertise in developing spark batch applications to transform data from Hive/HBase and write it back to Hive tables for reporting layer consumption.
* Worked in Analyzing/solving BIG DATA problems using Apache Hadoop (MapReduce, HDFS) and Ecosystems (Hive, PIG Latin, Sqoop, Flume, Oozie, Spark, Avro, Zookeeper).
* Actively Worked in Sqoop to Import/Export RDBMS Data from multiple databases (Oracle and MySQL) to HDFS/Hive. Used Sqoop's Import-all-tables for the initial table creation and import from RDBMS to HDFS/Hive.
* Created and scheduled Sqoop batch jobs to incrementally import/export on daily basis.
* Tuned performance of Sqoop imports by using option boundry-query and by increasing number of mappers.
* Actively worked in Hive/PIG for data warehousing, cleansing data, generating reports and adhoc historic reports.
* Created Managed/External tables in Hive and tuned performance of the HQLs by creating partitions and buckets on Hive tables.
* Optimizing Hive Queries to improve performance.
* Load the data into HBase tables for UI web application.
* Designed and created Hive external tables using shared meta-store instead of derby with partitioning, dynamic partitioning, and buckets.
* Building scripts to Monitor System health and logs and respond accordingly to any warning or failure conditions.
* Worked on Big Data Integration and Analytics based on Hadoop and Elasticsearch.
* Run various Hive queries on the data dumps and generate aggregated datasets for downstream systems for further analysis.
* Worked on writing code for validating the data exported from the traditional databases and the Hive using Sqoop.
* Worked on moving the data from HDFS to Object Storage (Clever Safe).

**Environment: MapR, Hadoop 2.1, HDFS, Spark, Hive, Sqoop, Shell Scripts.**

**Client: INTERAKT Digital Solutions Pvt Ltd June’2014 – Aug’2016**

**Role: Hadoop Developer**

**Responsibilities:**

* Involved in loading data from UNIX file system to HDFS.
* Involved in creating Hive tables, loading with data and writing hive queries which will run internally in map reduce way.
* Worked on writing Hive Queries.
* Worked on Hive Query optimization by setting different queue.
* Run various **Hive** queries on the data dumps and generate aggregated datasets for downstream systems for further analysis.
* Migrating the needed data from MySQL into HDFS using **Sqoop** and importing various formats of flat files into **HDFS.**
* Load the data into HBase tables for UI web application.
* Maintain System integrity of all sub-components related to Hadoop.
* Designed and created Hive external tables using shared meta-store instead of derby with partitioning, dynamic partitioning and buckets.
* **HiveQL** scripts to create, load, and query tables in a Hive.
* Monitored System health and logs and respond accordingly to any warning or failure conditions.
* Worked on Big Data Integration and Analytics based on **Hadoop and Solr.**
* Importing and exporting data from different databases like MySQL, RDBMS into HDFS and HBASE using Sqoop.
* Worked on indexing the HBase tables using Solr and indexing the Json data and Nested data.
* Worked on taking Snapshot backups for HBase tables.
* Worked on fixing the cluster issues.
* Involved in Region split and major compaction manually in HBase.
* Importing and exporting data from different databases like MySQL, RDBMS into HDFS and HBASE using Sqoop.
* Involved in writing Hive queries for Modules (Algorithms).

**Environment: Hortonworks Hadoop 2.1 and HDP 2.3, HDFS, Hive, Map Reduce, HBase, Pig, Sqoop, Shell Scripts, Oozie Co-coordinator, Solr.**

**Client: iGrid Technologies Jan’2013 - June ‘2014**

**Role: Hadoop Developer**

**Responsibilities:**

* Importing T-Mobile telco data from oracle database using sqoop.
* Building **HiveQL** scripts to process the data based on the client requirement
* Maintained System integrity of all sub-components (primarily HDFS, Hive).
* Integrated the hive warehouse with HBase
* Load the data into HBase tables for UI web application.
* Designed and created Hive external tables using shared meta-store instead of derby with partitioning, dynamic partitioning, and buckets.
* Monitored System health and logs and respond accordingly to any warning or failure conditions.
* Involved to load the data into Hadoop distributed file system (HDFS).
* Involved to create tables in HIVE and writing hive queries on the data.
* Involved load the output data into HBase database.
* Involved in load data to HDFS and Hbase using Sqoop.
* Involved in configuring Hive queries in Oozie scheduler.

**Environment:** **Hadoop, MapReduce, HDFS, Hive, Java, SQL, Pig, Sqoop, Oozie, ZooKeeper, MySQL HBase**

**Client: iGrid Technologies Jan’2012 – Dec ‘2012**

**Role: Software Developer**

**Responsibilities:**

* Designed Entegrate Screens with Java Swings for displaying the transactions.
* Involved in the development of code for connecting to database using JDBC with the help of Oracle JDevelper 9i.
* Involved in the development of database coding including Procedures, Triggers in Oracle.
* Worked as Research Assistant and a Development Team Member
* Coordinated with Business Analysts to gather the requirement and prepare data flow diagrams and technical documents.
* Identified Use Cases and generated Class, Sequence and State diagrams using UML.
* Used JMS for the asynchronous exchange of critical business data and events among J2EE components and legacy system.
* Worked in Designing, coding and maintaining of Entity Beans and Session Beans using EJB 2.1 Specification.
* Worked in the development of Web Interface using MVC Struts Framework.
* User Interface was developed using JSP and tags, HTML and Java Script.
* Database connection was made using properties files.
* Used Session Filter for implementing timeout for ideal users.
* Used Stored Procedure to interact with database.
* Development of Persistence was done using Hibernate Framework.
* Used Log4j for logging.

**Environment: Java EE 6, Eclipse 4.2, Oracle 11g/SQL**