Avinash Jadhav

Add: Flat No 8 Swaraj Park,
Tal/Dist. - Aurangabad -431008.

Mob No: +91–8329858176
Email:avijadhav20061996@gmail.com

CAREER OBJECTIVE

Seeking a lasting career within a technical role, commencing with the role of a Data Engineer. I possess a versatile skill set without a specific technology preference but boast substantial exposure to cutting-edge technologies such as Big Data. This exposure positions me to make a significant impact early in my career and contribute effectively to the organizational objectives of my future workplace.

PROFESSIONAL

- Having **2.6 years**' experience which includes **Hadoop Eco System** and **Big DataEngineering.**
- Good knowledge of Hadoop Ecosystems such as **HDFS, Job Tracker, Task Tracker, Name Node, Data Node and Map Reduce.**
- Uploaded and processed terabytes of data from various **structured** and **unstructured** sources into **HDFS.**
- Collecting large amount of log data using **Apache Sqoop** and aggregating using **HIVE** in **HDFS.**
- Very Good understanding of **Partitioning, bucketing** concepts in Hive and designed both **Internal**, **External** tables in Hive to optimize performance.
- Good exposure to **agile** software development process.
- Experience in rebalancing the data in the cluster.
- Involving in Importing/Exporting data to/from **RDBMS** using Sqoop.
- Configuring various **XMLs** file such as **Core-site.xml, Hdfs-site.xml, Mapred-site.xml, and Yarn-site.xml.**
- Manage and review Hadoop log files.
- Strong problem solving & technical skills coupled with clear decision-making.
- Good exposures in Requirement Analysis, Customization, and user acceptance.
- Participating in all **Sprint** activities including **daily stand ups,** grooming sessions etc.
- Proficient to Develop **ETL Pipelines** and **Data integration Solution.**
- Experience in Spark Architecture including **Spark-Core, Spark-SQL**.
- Good Understanding of **RDD, Data frame,** and optimization techniques in spark like, **Cache, Persist, Broadcast joins, repartition, and coalesce.**
- Have Good experience in Spark **Data Cleaning, Data integration, Data Transformation for compatibility.**
- Hands on experience of collecting large amount of data using **Apache Sqoop** and aggregating in **Hive.**

TECHNOLOGY PROFICIENCY

- **Platforms** : Windows, Linux.
- **Programming language** : Python, SQL
- **Scripting Language** : Shell Script.
- **Big Data Ecosystem** : HDFS, Sqoop, Hive, Spark, Pyspark
- **Distribution Package** : Cloudera
- **IDE** : PyCharm
- **Cloud Technologies** : AWS S3, EC2, EMR
- **Databases** : MySQL , Oracle
- **Other tools** : GitLab , Putty , Bit Bucket , jira

ORGANIZATIONAL SCAN

- **Data Engineer at MITGO IT Solution's Pvt. Ltd, Mumbai (Nov 2021 – Till Present)**

WORK EXPERIENCE

1. Data Migration   **Duration: Nov-2021–Apr-2022**
Role: Data Engineer

The Purpose of the Project is to ingest that data which is present on cloud storage to Data Hub which was modeled as per LDM. It's Cloud to hive ingestion Project in Real Time Data Migration Utility to migrate existing system into Data hub which was built on top of Hadoop.

Description &Responsibilities:

- Import Data into the Hive from various Relational Data bases **(Oracle, MySQL)** using Sqoop.
- Write Hive **DDL** to Create Hive Table for **Optimize Query Performance.**
- Ingest Flat Files like Delimited, Fixed Length, etc. into Hive Warehouse.
- Involved in importing data from various **RDBMS** into HDFS using **Sqoop** which includes Incremental Load to populate Hive External tables.
- Involved in importing data from various RDBMS into Hive Tables which includes Queries using Sqoop.
- Designed both **Managed and External** Hive Tables and Defined **static and dynamic** partitions as per requirement for optimized performance on production assets.
- Worked with various File Formats like Text File, Sequence File**, ORC Files, Avro Files** and various Compression Formats like **Snappy, bzip2.**
- Written Hive Queries for Data Analysis to meet the business requirements **KPI's.**
- Create **Sqoop jobs** for extracting data from RDBMS and file sources and loading it into **data warehouse.**

Environment: Hadoop, Hive, HDFS, Sqoop, SQL

2. Health and Insurance Data Analysis   **Duration:May 2022 present**
                                                                                Role: Data Engineer

Health and Insurance Data Analysis is a Data Hub where analysis for Medicine and Insurance Policies data is done on Big Data platform. At least 10% of the ethereal insurance payments are attributed to fraudulent claims. Worldwide this is estimated to be a multi-million-dollar problem. Fraudulent claims are not a novel problem, but the complexity of the insurance frauds seem to be increasing exponentially making it difficult for the Ethereal insurance companies to deal with them.

Roles and Responsibilities:

- Import Data into the **Hive** from various Relational Databases (**Oracle, MYSQL Server**) into HDFS using Sqoop.
- Write Hive DDL to Create Hive Table for **Optimize Query Performance**.
- Ingest Files like **CSV, TSV, and JSON** etc. into Hive Warehouse.
- Raise **JIRA** for Infrastructure and Platform issues.
- Engage with BA to understand the requirements clearly.
- Define all the possible Test Cases along with the Test Data and stored data into **S3, HDFS**.
- Designed both **Managed and External** Hive Tables and Defined static and **dynamic partitions** as per requirement for optimized performance on production assets.
- Perform optimization techniques to enhance performance of **Sqoop job**.
- **Written Hive Queries** for Data Analysis to meet the business requirements.
- Must monitor **Sqoop job**.

Environment: HDFS, Yarn, Hive, Sqoop, Spark , Pyspark , spark-SQL , Python, Big query

EDUCATIONAL QUALIFICATION
- **B.Tech. Engineering from DR. BATU University During (2019 – 2022)**

PERSONAL DETAILS
- Date of Birth   : 20 June 1996
- Languages known : English, Hindi, and Marathi

I hereby declare that the above information given is true and correct to the best of my knowledge and belief.

Place: Aurangabad   **(Avinash Jadhav)**