**Name :-** M.S. Adithya

**Number** :- 8801118152

**E-Mail** :- [santoshadithya](mailto:santoshadithyamathukumalli@gmail.com)30@gmail.com

**Objective:**

I can contribute to the development of the organization, which is growth oriented, by applying my skills and knowledge. Ability to learn quickly and desire to define and create the best solutions possible to become and invaluable asset to the company.

**Profile**

* **Developer- Big Data** with around 7+ years of professional experience in IT industry, involving in Software development and implementation
* Around 4 plus years of experience as Hadoop Developer with good knowledge in **Hadoop** Ecosystem **Data Warehouse**, and **SQL** technologies.
* Having working experience in **Databricks** and **Cloudera**.
* Worked on Databricks UI based distribution for **Spark,Azure** on **Delta tables,DBFS,Spark**.
* Working knowledge in Hadoop Core-Components like **HDFS, HIVE,PySpark,SQOOP ,HBASE, Spark, Scala,Spark SQL, Spark Structured Streaming,Nifi,postgres,elasticsearch**
* Maintaining Hive tables and executing hive queries
* Having Good working knowledge on **Spark** and **Scala.**
* Worked Knowledge on **Spark RDDs** (Resilient distributed datasets) **DataFrames, Datasets.**
* Having working knowledge on **Pyspark**.
* Having Good Working knowledge on **Datalakes, ADLS,Azure,Databricks.**
* Having Good Working knowledge on **Azure Data lake Gen 2, Data Factory,Data Flows**
* Worked on **Scala** **Functions, Collections and Traits**.
* Working Knowledge on  **Akka, Akka Streams, Akka Http** etc.
* Working Knowledge on **Kafka.**
* Working Knowledge on **Core Java.**
* Having Knowledge on **Zookeeper, Yarn,Pig,Flume.**
* Having Working Knowledge on **github and bitbucket** for code repository**.**
* **SQOOP** and**TDCH** and its integration with **Hive and Spark.**
* Importing and exporting data from **Teradata to HIVE** and **HIVE to Teradata**.
* Importing data from **Spark to sql server and Delta or Hive tables** in **Databricks.**
* Good experience on importing and exporting data from different systems to Hadoop file system using **SQOOP** and **TDCH**.
* Worked on Various distributed systems like **Cloudera, Databricks.**
* Working Knowledge in **HBase.**
* Working Knowledge on schedulers **Control M.**
* Working experience on client side testing.
* Working knowledge on **Power BI.**
* Involved in Managing **Power BI dashboard .**
* Ability to learn new things and take up challenges, team player with good communication skills.

Working experience:

|  |  |  |
| --- | --- | --- |
| **Company** | **Technology worked** | **Till Date** |
| Valuelabs  Larsen and Toubro  Infosys Ltd.  Select sys India Pvt Ltd. | Hadoop,Pyspark,Hive,Nifi,postgres  Hadoop, Spark, Scala,kafka, Spark Streaming, Core Java, Spark Sql.  Hadoop, Hive,Sqoop | June 2022 to till date  Dec 2020 to June 2022  Apr 2018 to Oct 2020  Oct 2014 to Jan 2017 |

**Project -1**

**Client : Value Labs**

**Designation : Hadoop Developer**

**Domain : Retail**

**Team Size : 1**

**Period : June 2022**

**Environment : Cloudera,AWS**

**Project Description:** DFS is retail store industry to reduce the time line in loading updated retail store prices and promos in all the stores across the globe build a pipeline it can overcome the timeline and timezone issues for the implementing above promos and prices.

**Responsibilities:**

* Involved in building pipelines end to end for DSL.
* Involved in writing spakr scripts for for batch process.
* Involved in processing semi structured data like xml and json to proper structure and transforming based on business
* Involved in creating and maintaining hive partitions tables.
* Involved in writing spark and scala scripts for data processing using dataframes and RDD’s .
* Performing transforming ,aggregations and cleansing of data.
* Involved in sending the data to AWS elastic search.

**Project -2**

**Client : Larsen and Toubro**

**Designation : Hadoop Developer**

**Domain : Dashboard Product for smart cities**

**Team Size : 1**

**Period : Dec 2020 – June 2022 Project -1**

**Client : Larsen and Toubro**

**Designation : Hadoop Developer**

**Domain : Dashboard Product for smart cities**

**Team Size : 1**

**Period : Dec 2020 – Till Date**

**Environment : Cloudera**

**Project Description:** Lnt building there own product for the smart and safe cities domain. In our product we are building ML models to capture crime pattern,gender,Licence plate,Mask detections etc. in the cities

**Responsibilities:**

* Involved in building pipelines end to end for Hyderabad,Nagpur smart cities project
* Involved in writing python scripts for api calling for batch process.
* Involved in maintaining and debugging the cluser and Nifi issues on server.
* Involved in increasing the space and creating mount points incase of space congestion issues in cluster.
* Involved in creating and maintaining hive partitions tables.
* Involved in writing pyspark scripts for data processing.
* Performing transforming ,aggregations and cleansing of data.
* Involved in extracting data from data dump anfd load into hive and postgres.

**Environment : Cloudera**

**Project Description:** Lnt building there own product for the smart and safe cities domain. In our product we are building ML models to capture crime pattern,gender,Licence plate,Mask detections etc. in the cities

**Responsibilities:**

* Involved in building pipelines end to end for Hyderabad,Nagpur smart cities project
* Involved in writing python scripts for api calling for batch process.
* Involved in maintaining and debugging the cluser and Nifi issues on server.
* Involved in increasing the space and creating mount points incase of space congestion issues in cluster.
* Involved in creating and maintaining hive partitions tables.
* Involved in writing pyspark scripts for data processing.
* Performing transforming ,aggregations and cleansing of data.
* Involved in extracting data from data dump anfd load into hive and postgres.

**Project-2**

**Client : 3M**

**Designation : Hadoop Developer**

**Domain : IOT filter devices**

**Team Size : 11**

**Period : Jan 2020 – Till Date**

**Environment : Databricks**

**Project Description:** 3M is mobile based company in USA. They have various IOT devices apart from mobiles like air filter, air purifiers, smart air filters etc.

**Responsibilities:**

* Involved in writing Scala and Python code for Spark Structured Streaming to Stream the data from IOT hubs and IBM outdoor.
* Involved in writing python scripts for api calling for streaming jobs.
* Performing transforming ,aggregations and cleansing of data.
* Involved in scheduling and automating jobs .
* Creating ,managing Delta tables and ADLS paths.
* Exporting the results from SQL servers and Spark into Power BI.

**Project-3**

**Client : VISA**

**Designation : Hadoop Developer**

**Domain : Credit cards**

**Team Size : 5**

**Period : March 2019 – Dec 2019**

**Environment : Cloudera**

**Project Description:** Small business is a module in VISA application of credit cards. All the backend changes of the UI will be involved in this module.

**Responsibilities:**

* Involved in creating java based API for importing and exporting of data from SQL server to Spark.
* Importing of data from SQL server to HBASE using java API.
* Exporting data from HBase to SQL server using Java API

**Project -4** :

**Client : Infosys**

**Designation : Hadoop Developer**

**Domain : Internal Project**

**Team Size : 1**

**Period : March 2019 – Aug 2019**

**Environment : Cloudera**

**Project Description:** Load the streaming data into Kafka cluster**.**Process and perform aggregation in spark streaming.

Responsibilities:

* Implemented Kafka producer and consumer applications on Kafka cluster setup with help of Zookeeper.
* Implemented to reprocess the failure messages in Kafka.
* Used Spark API to generate PairRDD using Scala programming.
* Have knowledge on partition of Kafka messages and setting up the replication factors in Kafka Cluster

**Project-5**

**Client : Anthem**

**Designation : Hadoop Developer**

**Domain : Health Care**

**Team Size : 6**

**Period : April 2018 – March 2019**

**Environment : Cloudera**

**Project Description:** This is a migration project from Teradata to Hadoop. Storing the increasing data day to day in Teradata is costly, so client interested in much reliable source Hadoop.

**Responsibilities:**

* Involved in importing the data from Teradata to HDFS using TDCH which is an internal tool used for importing and exporting the data from Teradata.
* Creating Hive tables and loading the data from HDFS into them.
* For processing the data we used Spark framework using spark-sql and scala as per client requirement.
* Involved in exporting the data back to Teradata using TDCH.
* For scheduling purpose we have used Control-m.
* Involved in client side testing for migrated data.

**Project-6**

**Client : CITIBANK**

**Designation : Hadoop Developer**

**Domain : Banking**

**Team Size : 9**

**Period : June 2016–Jan 2017**

**Environment : Cloudera.**

**Project Description-** Customer Insight & Retail Analytics is the process of evaluating the customer feedback collecting from different vendors about banking products like Accounts, loans, Credit cards, Mortgage process etc. Create survey request based on customer performance based on transactional history and send the Survey request to the third party vendors and once survey completed they send back data to the Bank database. Resulted data extracted and loaded into Hive in incremental basis and full refresh after applying ETL transformations

**Responsibilities:**

* Involved in loading data from file system to HDFS
* Loading files to HDFS and writing hive queries to process required data.
* Loading data to hive tables from HDFS and writing queries to process.
* Develop, validate and maintain HiveQL queries.
* Load and transform large sets of structured data.
* Involved in creating Hive tables, loading with data and writing hive queries that will run internally in map reduce way.
* Responsible to manage data coming from different sources.
* Importing and exporting data into HDFS and Hive using Sqoop.
* Written Hive queries for data analysis to meet the business requirements

**Project-7**

**Client : Take a Lot**

**Designation : Hadoop Developer**

**Domain : Ecommerce (Online Shopping)**

**Team Size : 6**

**Period : Oct 2014–May2016**

**Environment : Cloudera**

**Project Description:** Takealot is one of the Leading Online shopping Portal from South Africa. The purpose of the project is to store Huge Amount of data generated by the ecommerce website and extract meaningful information out of it. The solution is based on the open source Big Data s/w Hadoop .The data will be stored in Hadoop file system and processed using Hadoop which includes getting the raw data from the Server, process the obtain product and pricing information, extract various reports out of the product, pricing information and export the information for further processing.

**Responsibilities:**

• Executed queries using Hive jobs to analyze data.   
• Importing and exporting data in HDFS and Hive using Sqoop to Teradata and Hbase.

• Involved in loading data from LINUX and UNIX file system to HDFS.   
• Utilized Apache Hadoop environment by Cloudera 5.2.2 in UNIX putty.

• Creating .lst files for bulk import and export of data into HDFS and also for creating Hive tables.

**Education-**

|  |  |  |
| --- | --- | --- |
| **Degree** | **University** | **Percentage** |
| Btech (IT) | Jawaharlal Technological University Hyderabad | 62 |

**Personal Details:**

Nationality:Indian

LanguagesKnown:English, Hindi and Telugu

Place:Hyderabad

DOB: 30/05/1991

Fathers Name: Late M. Nageshwar Rao

Regards

MS. Aditya