**BRAHMAIAH B**

**Mobile: +91-9121471233**

**Email Id: brahmaiah.bollam2@gmail.com**

|  |
| --- |
| **Professional Summary** |

4+ years of experience in developing, delivering of software applicationsusing avariety of technologies within the Big Data frameworks. Experienced in all phases of the software development life cycle. Proven ability in project-based team work.

**Skill Set:**

* Hands-on development and implementation experience in Big Data Management Platform (BMP) using Databricks, S3, AuroraDb, DynamoDB, HDFS, Hive, pyspark, Oozie, Sqoop, HBase, Linux, Apache airflow and HUE.
* Experience in Data transfer from structured data stores to HDFS using Sqoop, pyspark JDBC connection.
* Developed Batch Processing jobs using pyspark, SparkSQL and Hive.
* Thorough Understanding of HIVE, Hive Query language, Hive tables, Partitioning, Bucketing and optimization techniques.
* Depth working experience on different files formats like parquet, JSON, Orc..etc.
* Experience in extracting and loading data to databases like Oracle, MS SQL Server, and Teradata, SAPBW.
* Experience in understanding and managing Hadoop Log Files, cloudWatch logs.
* Experience in using Flume to load logs files into HDFS.
* Experience on AWS Analytical services like Aws Glue, Lambda, S3,Secrete Manager, Athena, and Cloud Watch.
* Implemented Pyspark scripts in Glue pyshell and Azure Databricks.
* Implemented separate ETL-Framework to automate the spark jobs
* Implemented the cloud-formation templates to create the dynamodb tables.
* Implemented the code to trigger the Databricks job from Lambda on event bases.
* Experience Azure Databricks cluster and also implemented the job cluster to trigger the airflow jobs parallelly.
* Hands on experience integrating Hive with Spark to perform HQL in Spark.
* Experience with Airflow Workflow Engine in running workflow jobs with actions that run Hadoop, Sqoop, hive and pyspark, linux.
* Experience on Sub versioning and continues integration and development with Jenkins.
* Having knowledge on Kafka architecture and also have hands-on knowledge on Kafka.
* Experience on GitLab and GitaBash for check-in and checkout.

|  |
| --- |
| **Experience** |

**1**. Worked as Big data Developer with **Novartis** from Feb 2020 to Till Date

**Client : Novartis**

**Project : F1 SpotOn**

**Role : Bigdata Engineer**

**Organization : Altimetrik India Pvt Ltd.**

**Roles and Responsibilities:**

* As per the project requirements we are fetching the data from different data sources to **s3** using databricks with pyspark jdbc connection.
* Implemented pyspark scripts to fetch the incremental data from sql server and SAP BW.
* Implemented the logic to load the tables from Aurora to Databricks to DynamoDB tables and global secondary indexes
* Implemented the logic to write dynamo-DB tables from s3 file system
* Implemented the lambda code to trigger the Databricks job cluster once the glue job has completed using cloud-watch Rules.
* Experience on cloud-formation stacks and templates.
* Worked on Airflow for job scheduling.
* Implemented the python scripts to create the scopes and secrets in databricks.
* Experience in Git-lab and Git-bash for code check-in and checkouts.
* Experience on AWS Glue, lambda and Cloud Watch etc.

**Environment**: Databricks, pyspark, S3, Apache Airflow, DBeaver, RDS (Aurora DB), dynamodb, lambda, cloudwatch.

--------------------------------------------------------------------------------------------------------------------------------------------------------

**2**. Worked as Big data Developer with **Travelers Insurance** from Oct 2018 to 31st Jan 2020

**Client : Travelers**

**Project : Business Insurance**

**Role : Big Data Developer**

**Organization : Infosys**

**Description:**

Traveler is one of the biggest insurance companies in fortune 500 companies. Currently project is migrating from Teradata to Bigdata as per their data limitations. As per their requirement they are maintaining different businesses like RDF, dun& Bradstreet and Pre-issuance. For all these businesses they use TIF as a tool to import/export the data from Teradata to HDFS vice versa. And also on daily bases we used to get the client events from abinitio to Hadoop file system by using rabbit MQ. Once data placed in hdfs, we wrote generic pyspark scripts to process the data and load the data frames into hive partition tables and physical zed views

**Roles and Responsibilities:**

* Implemented the TIF Tool (SQOOP) to import and export the data from Teradata to HDFS.
* On daily bases we used get the RDF events from Abinitio to temporary NAS drive by using Rabbit MQ, once data in NAs drive.
* Implemented the python and pyspark scripts to parse the nested xml’s and convert them into Json format.
* Developed spark coding using python scripting to analyse the data we are getting from different sources.
* Created data frames in a Spark SQL from data in HDFS and did transformations analyse the data and stores back into the HDFS and used the msck repair to merge the partition files with hive tables on daily bases.
* Worked on broadcast variable to put the data on single variable and used it for entire process
* Developed UDFs for different logics based on requirement.
* Closely worked with testing team to fix all the bugs and made the code more stabilized.
* Experience in SVN code check-in and checkouts.
* As per the AWS POC, worked on AWS Glue, Athena and Cloud Watch etc.

**Environment**: HDP 2.6, pyspark 2.3.0, HDFS, Ambari, Hive, Sqoop, AWS Glue, Athena, SVN, Jenkins.

**---------------------------------------------------------------------------------------------------------------------------------------------**

**3.** Worked as Big data Developer with **Prime Therapeutics** from Dec 2017 to Oct 2018.

**Project : Claims ODS**

**Organization : Infosys**

**Client : Prime Therapeutics**

**Role : Big Data Developer**

**Description:**

Its Health Care project where we process the files on daily basis as well as history data as per the client requirement. Files from landing areas are stored in RAW layer i.e. HDFS and from RAW the files are moved to stage layer i.e. HIVE by performing several data quality rule i.e. schema mismatching, deduplications and performing hard reject by capturing the error records in error hive tables. In conform stage Data Frame are created from hive stage table and required transformations are done as per the STT documents.

**Roles and Responsibilities:**

• Created Hive External tables with partitioning and bucketing logic.

• Tracking the corrupted records for each table and updating those corrupted records separately in error table.

• Experience in dealing JSON, CSV, TSV, ORC and pipe delimiter files.

• Experience in loading data from landing area to RAW layer (HDFS) and from Raw to Stage and Conformed layers (HIVE).

• Generated hash codes for each record to remove duplicate records with in the data.

• Developed UDFs for different logics based on requirement.

• Experience in manual execution of DAG’s using Airflow to load data to Raw, Stage and Conform layers

• Closely worked with testing team to fix all the bugs and made the code more stabilized.

* Experience in SVN code check-in and checkouts.

• Monitoring and Debugging Hadoop jobs/Applications running in production.

**Environment**: CDH 5, pyspark, HDFS, Apache Airflow, Hue, MariaDB.

**------------------------------------------------------------------------------------------------------------------------------------------**

**4**. Worked as Big data Specialist with **FINO Payments Bank Mumbai** from May 2017 to Nov 2017.

**Organization : TransOrg Analytics**

**Client / Project : FINO Payments Bank.**

**Description:**

Joint Liability Group (JLG) is a lending model that enables a group of individuals to take loans for income generating activity by forming a group, wherein group members guarantee each other loans. Fino is innovative online payment solutions and extensive reach offers customers a safe, reliable and quick means to transfer money back home, while giving banks the opportunity to tap the huge 70% money transfer market that is currently served by informal players.

**Responsibilities**

* Developed data pipeline using Sqoop, Pyspark and Hive to ingest data into HDFS.
* Extensively used cron schedulers for doing manipulations of the flat files, and to run the spark scripts.
* Developed Sqoop scripts to import and export data from and to relational sources.
* Worked on persist and cache to store the required RDDs on the memory.
* Developed spark coding using python scripting to analyze the data we are getting from different sources.
* Worked on broadcast variable to put the data on single variable and used it for entire process.
* Involved in loading and transforming large sets of structured, semi structured data from databases into HDFS using Sqoop imports.
* Importing and exporting different kinds of data like incremental, updated, and column base data from RDBMS to HIVE.
* Export the data from HDFS to SAS cluster for visualization

**Environment**: CDH 5, Pyspark, SPARK SQL, Hadoop, SAS, Sqoop, Hue, Oozie, SQL Server.

**-------------------------------------------------------------------------------------------------------------------------------------------------**

|  |
| --- |
| **Academic Credentials** |

* Bachelor of Technology on Mechanical Engineering, JNTU Kakinada – 2014.

|  |
| --- |
| **Personal Details** |

Name : Brahmaiah B

Sex : Male

Current city : Hyderabad

**(BRAHMAIAH)**