

Guthikonda Ramteja – RESUME

Email : ramteja.bigdata@gmail.com

Mobile: +91 9492034307

Data Engineer Summary

- Having 4+ years of professional IT experience with over 4+ years of comprehensive experience in big data technologies.
- Expertise in Hadoop, **HDFS,Yarn, Spark, Sqoop, Hive, Scala**
- Strong understanding on Hadoop Architecture and MapReduce framework.
- Proficient at using Spark API to cleanse, aggregate, transform and store the datasets.
- Creating Sqoop jobs and Running **Sqoop** jobs daily for Incremental loading.
- Experience in importing and exporting data using Sqoop from HDFS to RDBMS and vice-versa
- Good experience in implementing business logic, optimizing queries using **HiveQL**.
- Good experience in using performance tuning techniques in Hive.
- Experience on working with RDDs, Data Frames and Datasets in spark.
- Knowledge on Kafka and Spark Streaming.
- Knowledge on AWS and basics of Python

Technical skills

- Hadoop Ecosystem : HDFS, Hive, Sqoop, spark
- Languages : Scala , Python
- Database : Oracle, Mysql , Teradata
- Query Languages : HiveQL ,SQL

Managerial skills

- Ability to provide leadership, motivation, and strategic direction to a multifaceted team
- Excellent interpersonal, collaboration, and problem-solving skills
- Proven ability to achieve results in very high-pressure environments and to keep teams motivated

Work Experience

- Worked as Hadoop Developer in MINDTREE from May 2018 to TILL NOW.

Qualifications

- Bachelor of Technology in Electrical & Electronics Engineering | JNTU KAKINADA.

Professional Experience

Project Name : Data Plat Form

Role : Data Engineer

Client : ABBVIE

Duration : Feb 2021 to Till Now

Project Description:

Data platform effort is intended to modernize and standardize data processing practices, infrastructure, and services to meet the needs of a growing organization. Its goals include. Democratizing access to data consumption and generation at scale. Developing and storing the technical aspects of compliance and data governance. Improving resilience of data processing and investing through developing robust tools, providing reliable infrastructure, and codifying best practices.

Roles & Responsibilities:

Data Acquisition:

- Involved in Implementation of Data Ingestion for boarding and processing input data.
- Responsible for loading the data into HDFS from different source systems which includes File systems and RDBMS and Kafka and Aws S3.
- Responsible for maintaining / organizing different data format of different layers in centralized data lake.

Compute and Transformation:

- Developed compute jobs based on business requirement for downstream applications.
- Communicating with BU to get the requirements and working towards completion of development activities.
- Implemented Spark - utilizing Spark core and Spark SQL for faster processing of data.
- Involved in developing Spark SQL to perform compute jobs on the hive tables.

Technologies used:

Spark Core, Spark SQL, Python, HDFS, Sqoop, Aws S3, Hive, Jenkins's pipeline, GIT Version Control, MYSQL, Autosis

Project Name : Finance and Risk Data Mart

Role : Spark & Hadoop Developer

Client : DBS

Domain : Banking & Finance Domain

Duration : April 2019– Dec 2020.

Project Description:

FRDM project deal with related to risk and financial department which consolidate data from different source system including loan accounting, equity, and other data into single unified system where these data will send to down streams application which are Risk Mart, QRM, Hedge accounts and this data model is optimized for risk management.

The main aim of the project is to maintain a centralized data lake which stores the data from all the sources to perform various analytics. In this project we have loaded the data from various external source systems to Hadoop HDFS data lake, then making it useful for the different downstream applications.

Roles & Responsibilities:

Data Acquisition:

- Involved in Implementation of Data Ingestion for boarding and processing input data.
- Responsible for loading the data into HDFS from different source systems which includes File systems and RDBMS.
- Responsible for maintaining / organizing different data format of different layers in centralized data lake.

Compute and Transformation:

- Developed compute jobs based on business requirement for downstream applications.
- Communicating with BU to get the requirements and working towards completion of development activities.
- Implemented Spark - utilizing Spark core and Spark SQL for faster processing of data.
- Involved in developing Spark SQL to perform compute jobs on the hive tables.

Technologies used:

- Spark Core, Spark SQL, Scala HDFS, Hive, Teradata, Jenkins's pipeline, GIT Version Control, Kafka


Project Title : Comprehensive Broadcast Reporting (CBR)

Client : Kantar Media

Technologies : HDFS, Spark, Sqoop, Hive, Scala

Duration : may 2018 to Mar 2019

Description:



In CBR all the broadcasters and Advertisers data are available in the existing structured databases (like Teradata and My SQL) hence we are bringing the data to OMNIA for the further analytics. After the analytics and the sliced records are moved again back to the Structured databases for the graphical representations. we have used Tableau and providing the report to client in a graphical view.

Roles and Responsibilities:

- Involved in the process of data acquisition, data preprocessing and data exploration of telecommunication project.
- Involved in creating Hive tables, loading with data, and writing hive queries which will run internally
- Data processing purpose using Spark (Spark 2.1)
- Used JIRA to get the requirement from the customer along with dependent metadata information
- Used INTELLIJ for code generation tool to generate a JAR
- Cloudera 5.3 as Distributed system for this project
- Preparing the filtered data ready for business analysis in text format and some data back to MySQL database in a separate table for Analysis as per the requirement using Sqoop.

Technologies used:

- Spark Core, Spark SQL, Scala HDFS, Hive, Teradata, Jenkins's pipeline, GIT Version Control, Kafka