Data Engineer

Hari Prasad   hariprasadk0073@gmail.com
PH: 9986281613

Experience Summary:

I have 4.6 years of experience in total IT and strong technical skills in SQL, Hadoop (Bigdata) and Spark, PySpark, Python, AWS (EMR, Glue, Athena, S3, Lambda) Hive, MapReduce, Sqoop, HBase, data bricks etc. Include designing, developing applications using Hadoop Ecosystem components. Good knowledge in HTML, CSS.

- Overall, 3.6 years of relevant experience in Data engineer.
- Good experience in Hadoop ecosystem& Big-Data Analytics.
- Experienced in Spark ecosystem components like Spark batch, Spark streaming, Spark SQL.
- Good experience in Spark, Pyspark, AWS, data bricks.

- Experienced in using HDFS, MapReduce, Hive, Sqoop, HBase, Cloudera Manager. Good knowledge of Hadoop Architecture and Yarn Architecture.
- Good knowledge of Spark Architecture and AWS.
- Good knowledge and Understanding of Spark-standalone-Architecture and Spark-Yarn Architecture.
- Experienced in working Hive data warehouse system, developing data pipeline, implementing complex business logic and optimizing business logic.
- Good knowledge of AWS,S3, EMR.
- Experienced Hadoop Developer, have a strong background with file distribution systems in a bigdata.
- Understands the complex processing needs of big data and have experience in developing codes and modules to address those needs.
- Excellent communication, organizational skills, attention to detail & the ability to meet tight deadlines.

Educational Qualification:

MTech From Jawaharlal Nehru Technological University Anantapur (JNTUA).

Work Experience:

Worked as a Software engineer in StackNous Technologies.

Skills Summary:

Domain	: Telecom, Retail, Banking, Media
Programming Languages	: Python
Operating System/ ERP Version	: Windows10, Ubuntu Linux Tools / DB / Packages /
Framework / ERP Components	: Spark, PySpark, Python, Hive, HDFS, Sqoop, MapReduce, HBase, AWS, (Athena, Glue, EMR, S3).

Project 2:

Project Name	: AT&T
Team Size	: 8
Start Date	: Apr'2021
End Date	: Till Now

Environment:

HDFS, Hive, Python, Sqoop, Oracle, Spark, Spark-SQL, Linux, Autosys, Putty.

Description:

Millions and Billions of users under the AT&T, they have products like Telecom, the data analytics is a critical component of AT&T-business. It collects huge amount of usage information about all their products, services and sites such as Complete Calls and Call Records, product and device specific information and other similar data and that will be consumed for Business Analytics and build Big Data Solutions for various Toronto-Dominion Business units. The collected data is aggregated and analyzed for adhere legal compliance, business and marketing to improve the services, measure and analyze the usage and to help to find services and advertising more relevant to the customers.

Roles & Responsibilities:

- Injection data from Oracle to Big data layer at Hive partition tables.
- Loading data from Hive tables to Python Spark for Processing Data.
- Understand business story and analyze functional specifications & design feasible solutions.
- Based on business logic at applying cleaning the data, applying transformations and analytics with Python spark.
- Developed and implementing of code to write the logic data into Hive External Tables.
- Involved in configuring Bitbucket and creating Build using Jenkins.
- Involved in various log files to fix the issues during unit testing and performance tuning Involved in developing the spark.

Technologies & Tools:

Spark and Python, AutoSys, Arcadia Data, Hadoop Hive, Sqoop, Yarn, Bitbucket, PyCharm, Windows, Linux, JMS, Jenkins.

Project 1:

Project Name   : P&G
Team Size   : 08
Start Date   : July'2019
End Date   : Mar'2021

Project Description:

P&G Business such as P&G Products, enterprise customers and a huge global Products, data analytics is a critical component of P&G business. P&G collects huge amount of usage information about all their products, services and sites such as call records, wireless location, network traffic data, product and device specific information and other similar data and that will be consumed for Business Analytics and build Big Data Solutions for various P&G Business units. The collected data is aggregated and analyses for adhere legal compliance, business and marketing to we improve the services, measure and analyze the usage and to help to find services and advertising more relevant to the customers.

Roles & Responsibilities:

- Data Reading from Mount Path to Spark for Processing data.
- Understand business story and analyses functional specifications& Design feasible solutions.
- Based on business logic at applying cleaning the data, applying transformations and analytics with Spark.
- Developed code to write the logic data into Athena External.
- Involved in configuring and creatingAirflow schedule.
- Involved in setting up the properties for Apache Airflow.
- Involved in various log files to fix the issues during unit testing and performance tuning& involved in developing the spark.
- Tuning &Optimizing Involved in developing the PySpark.

Technology & Tools:

PySpark, Python, Spark, AWS, S3, Athena, data bricks,Airflow, CI/CD.S