NAVEEN SHIVANADULA

PLOT 74 VAIBAV COLONY, JODIMETLA, HYDERABAD · 8948058334
shivanadula.naveen14@gmail.com

CAREER OBJECTIVE

To obtain a challenging position in a dynamic organisation and be an integral part of growth-oriented company. To utilize and grow my analytical, research and technical skills that provides me the job satisfaction and makes me work with a whole-hearted commitment.

PROFESSIONAL SUMMARY

- Skilled developer with experience of 4 years 9 months in **Python, Spark, AWS services, Hadoop and Bigdata**.

EXPERIENCE

NOVEMBER 2020 – CURRENT
SENIOR DATA ENGINEER / ENQUERO GLOBAL LLP

MAY 2016 – OCTOBER 2020
BIGDATA/AWS DEVELOPER / INFOSYS LIMITED

EDUCATION

MAY 2016
BACHELOR OF TECHNOLOGY / NIT ALLAHABAD
Electronics & Communication Engineering

SKILLS

- AWS Skills: S3, EC2, EMR, Glue, Athena, Aurora, SNS, Crawler.
- Big-data Skills: Spark, Hive, HDFS, Hadoop, Sqoop.
- Programming Languages: Python, C#
- Databases: Aurora DB, Oracle, MySQL.
- Job Scheduler tools: Airflow.
- CICD: Bit bucket, Jenkins.
- Replicating Tools: Attunity.
- IDE: PyCharm, Eclipse, Visual Studio.

PROJECTS

Name: AMFAM – ADW CLOUD.
Client: American Family Insurance Company.
Role: Developer
Team Size: 5
Environment: Spark, Python, Sqoop, Hive, AWS EMR, S3, Glue, Athena, Aurora DB, Airflow, Attunity.

Description:

Client is a leading insurance company. To improve their customer service, client wants to periodically analyse their data to repair faults and problems. This project sets out to ingest data from different types of sources. We use PySpark to ingest high volume of data. Data is cleansed, transformed and then mapped to target data types and finally store in hive.

Responsibilities:

- Developed PySpark scripts using Data frames and RDD's in Spark for Data Ingestion and aggregation from scratch.
- Developed following internal frameworks using PySpark:

    i. Comet Framework:
        a. To process, filter and validate the data in AWS S3 using PySpark which is replicated by Attunity Replication tool from different sources to S3.
        b. To Process the CDC data in different mode like NRT (Near Real Time), BATCH, ACTIVE and finally store in hive.
    ii. File Framework:
        a. For uncompressing the different compressed files (zip, tar, tar.gz).
        b. To process delimited files like csv, tsv, dat, fixed width and pipe delimited files using python and spark data frames and finally saving into hive table.
    iii. Sqoop Framework:
        a. To import data from Oracle, DB2 and SQL server to s3 using Sqoop import and then processing the data from s3 and creating tables on it.
    iv. Airflow Framework:
        a. Created Dags in Airflow in Python which will run on AWS EC2.
        b. Scheduling the Dags and submitting the spark jobs from AWS EC2 to AWS EMR.

- Replicated data from Oracle Databases to S3 using Attunity Replication tool by creating FL and CDC tasks with source and destination connections.
- Experience in dealing JSON, CSV, TSV, DAT and PIPE delimiter files.
- Developed scripts using python boto3 to uncompress zipped files in S3.
- Closely worked with QA team to fix all the bugs and made the code more stabilized.

- Tracking the corrupted records for each table and updating those corrupted records separately in audit table.
- Experience in bit bucket code repository.
- Hands on deploying code using Jenkins.
- Developed SQL scripts for extracting metadata from on premise Oracle DB to Cloud Aurora DB.
- With the help of Object Tagging concept in s3, categorized the processed and unprocessed files in s3 using Python scripts.
- Created databases in AWS Glue catalogue using Python script.
- Developed Python script for converting fixed width delimited files to csv files.
- Done Validations checkouts for active and history tables.
- Used Amazon SNS in Framework to email the status of the job.
- Done exception handing in every UDF by raising exception wherever there is a chance of exception.
- Provided support for production release.

Name: DIGITAL MARKETING.
Client: VMware.
Role: Developer.
Team Size: 1
Environment: Python, Hadoop, Apache HAWQ.

Description:

VMware is a company developing compute, cloud, mobility, networking, and security products and services. Its marketing data is present in different platforms like Adobe Analytics, Brightedge etc. This project is to bring marketing data from different platform using Python REST API to Hadoop and Store the data in tables in Apache HAWQ. Later the data is used to create Tableau Reports.

Responsibilities:

- Developing Python scripts to pull the data from different API end points and extract different metrics like visits, entries, impression, bounces and cost from the pulled data.
- Created different Payload Jsons for different segments which is an input for Python Rest API PUT.
- Used Multithreading to run PUT and get responses parallelly.
- Used Python Pandas to create Dataframes on extracted data and storing the data in a csv file in Hadoop. External tables are created on it in HAWQ.
- An Internal Table is maintained for History data and Incremental data is loaded from external table to internal table daily.
- Created Stored procedures in HAWQ to load incremental data to internal tables from external tables.

ACHIEVEMENTS:

- Won **five** Insta Awards for my excellence in the project in Infosys.
- Won DNA Most Valuable Player Award for year 2019 in Infosys.

DECLARATION:

I hereby declare that the above-mentioned information is correct up to my knowledge and I bear the responsibility for the correctness of the above-mentioned.

Naveen Shivanadula