SUBHOJEET RUDRA

Email ID: subhojeetr@gmail.com
Contact: 9845199391

SUMMARY:

- Working with Accenture as a Senior Software Engineer
- Overall 5.6 years of experience in IT industry.
- Expertise in Python, Pandas, Pyspark, Hive, Sqoop, Big Data, Hadoop, SQL, SAS, AWS
- Efficient in writing codes.
- Having good analytical skills and fast learner.

Technical Skills:

- Programming Language: Python, Pandas
- BigData Technology: Pyspark, Hive, Sqoop, HDFS
- Package: pandas, numpy, pyspark, psycopg2, pyodbc, sklearn, requests, re, sys, logging, xml, json
- Tools: SAS EG, Ab Initio
- Cloud Computing: AWS EC2, S3, IAM Role, CLI

- Databases: Redshift, Snowflake, PostgreSQL, Teradata
- Scheduling: Control-m, Autosys, cron
- Version Control: GitHub
- Ticketing Tools: HPSM
- Operating System: Windows, Linux

Professional Experience:

Accenture Solutions Pvt. Ltd. [Feb 2018 to till date]
Designation: Senior Software Engineer

Roles & Responsibilities:
- Expertise in migrating complex SAS processes (2000 - 4500 lines) to pandas and pyspark process using **Spark APIs** and **SparkSQLs**.
- Experience in establishing ETL pipelines by importing data into hive using **Sqoop**, Loading it as a pyspark dataframes, manipulating it by using **UDFs** and other column based inbuilt API and finally loaded into **dynamically partitioned** hive tables as ORC and parquet file format.
- Expertise in working with different file formats (**fixed width, csv, xml, json, avro, parquet, orc**).
- Automate the process to load the archival files to **S3** bucket and reconciliation.
- Expertise in writing **Unix Shell Scripting**.
- Experience in data analysis using **SAS EG** and pandas.
- Experience in calculation of statistical values and data cleaning and preprocessing.
- Experience in migrating 500 tables from Teradata to Hive.
- Migrated **Ab Initio** graphs to pyspark process.
- Developed complex HQLs to load data into **partitions** of **external tables** from stage to final target table.
- As part of POC, worked in virtual agent using **rasa**.
- Scheduled jobs using **Autosys**.

Wipro Technologies Pvt. Ltd. [Aug 2014 to Feb 2018]
Designation: Project Engineer

Roles & Responsibilities:
- Experience in automating the extraction of schema and table name from Teradata SQLs using Python regular expression and retrieving the counts of extracted table names to ensure **Count Validation**.
- Expertise in developing framework from **AWS S3** bucket to Data Lake using Shell Scripting.

- Experience in migrating 800 tables from Teradata to Redshift.
- Expertise in working with large **SQLs** (around 3000-6500 lines).
- Built historical load framework to load historical data from **AWS Redshift** to **Snowflake**.
- Experience in developing **HQLs** to load the target Hive Table.
- Automatic Report Generation and Automatic Mails using Python.
- Migrated tables from Teradata to **AWS Redshift** and **Snowflake** using Shell Script, Python, SQL and Ab Initio.
- Experience in demos to different Teams.
- Scheduled jobs with **Control-M**.

Academic Profile:

- B.Tech, KIIT University 2014 -84%
- Higher secondary from RSVM, 2009 (C.B.S.E) -83%
- Matriculation from RSVM, 2007 (C.B.S.E) -80%

Awards & Achievements:

- Received Trophy in Wipro PRP Program.
- Accenture ACE Award Winner.
- Conducted training in Pyspark.

Personal Profile:

Name   : Subhojeet Rudra
Date of Birth   : 11th March, 1991
Languages Known   : Hindi, English, and Bengali
Nationality   : Indian