Siddharth Nautiyal
Big Data Consultant

Harinagar, Ashram New Delhi, Delhi-110014

+91-8700179561, +91-8373977459

sidnautiyal2010@gmail.com

Big Data Engineer with more than 3 years of experience in building data intensive applications, tackling challenging architectural and scalability problems in Finance, Hospitality and Energy Sector. Currently helping Deloitte in building scalable data pipelines for one of their US based clients.

⚙ Skills

- Python
- Azure
- Apache Spark
- Scala
- Amazon Web Services
- SQL
- Apache NIFI
- Pandas
- Informatica Cloud

💼 Work History

2020-11 - Current

Big Data Consultant
Deloitte Consulting India Pvt Ltd, Gurgaon, Haryana

- Developed and setup an end to end scalable ETL datapipeline over **Azure.**
- Developed robust extraction flow using **Azure DataFactory** which will be handling incremental as well as full data load on the daily basis.
- Developed generic and reusable framework to implement **Slowly Changing Dimensions (1 &2)** and **Facts**, with the capability to handle the changed data capture using **Pyspark** and **SQL** in **Azure DataBricks**.
- Implemented **Azure delta lake** to improve the performance of the above mention framework, which reduces the time to half as compared to the basic spark-sql framework,which is being used previously.
- Implemented **Dynamic Data Masking** on **Azure Synapse**, which masks the PII Information present in the data.

2020-01 – 2020-11

Big Data Engineer
RedDoorz, Noida, Uttar Pradesh

- Developed and setup an end to end scalable **ETL datapipeline over AWS**.
- Developed robust **incremental/full extraction** flow using **NIFI**.
- Developed reusable **validation framework** using **python** for cleaning and massaging the data.
- Developed reusable framework to implement **SCD1/SCD2** and **incremental/full** data load to **Redshift** tables using python.
- Developed complex and optimized **SQL** queries to implement business logic.
- Created well optimized tables with **AWS Redshift** best practices,reducing querying and loading time.
- Created **AWS Step Function** to automate the whole datapipeline,which triggers at scheduled time, with the help of **AWS CloudWatch**.
- Developed daily automated **business summary report**, which contains daily, monthly and last month business related details and the comparison between them, using **python**.

2017-06-2019-12	**Technical Associate**

Genpact India Pvt. Ltd., Noida, Uttar Pradesh

☐ Created **data extraction** framework to extract flat files like csv, excel, xml, txt from **FTP** server using **Python**.

☐ Developed reusable data **validation** framework for cleaning and massaging the data using **Spark**.

☐ Developed reusable framework to implement business logic and perform **SCD1/SCD2** using **SPARK-SQL**.

☐ Developed complex **SQL queries** to implement business logic implementation.

☐ Developed framework to load **full/Incremental** data to **AWS Redshift** using **Spark-Sql**.

☐ Developed framework in **Scala** to process **Excel** files using **Apache POI library**.

☐ Developed **Spark** framework, which is improvement over Scala framework **for processing Excel** files using **Crealytics library**.

☐ Created tables in **AWS Redshift** using best practices, which leads to better query and data loading performance.

☐ Created **AWS Datapipeline** scripts for orchestration of all the jobs as per batches.

☐ Created **AWS lambda** to automate AWS EMR creation/termination and triggering of **AWS datapipeline**.

☐ Created **AWS Stepfunction** to orchestrate whole process.

☐ Developed **AWS Data Migration Service** pipeline to replicate 100 GBs of Historical data.

🎓Education

2017-06	**Bachelor of Technology**

Greater Noida Institute of Technology,Greater Noida, Uttar Pradesh
Acquired percentage- 65%

2013-05	**Intermediate**

G.L.T Saraswati Bal Mandir Sr. Sec. School,New Delhi, Delhi
Acquired Percentage- 67.7%

2011-03	**High School**

G.L.T Saraswati Bal Mandir Sr. Sec. School, New Delhi, Delhi
Acquired CGPA-8.8