

Data Engineer

Charles Obuseh, MSc, PE

charleside2001@yahoo.com

Cypress, TX 77429 • (972)904-3380

Self-driven individual, analytical, problem solver with over 3 years leveraging data engineering skills. I am proficient in technical software, languages, and tools with the ability to learn new tools, databases and systems to maintain/enhance strategic vision of the organization and able to provide cutting-edge data engineering skills. Seeking a reputable organization to contribute to with opportunities for personal growth.

TECHNICAL SKILLS

Languages

- Python, SQL

Others

- Git, Bash

Data Manipulation & Visualization

- Pandas, Tableau BI, ETL, Apache Spark, Hadoop HDFS, Snowflake Data Factory, SnowSQL and Snowpipe, Amazon Kinesis Firehose, Databricks, AWS EMR, Athena, Lambda, Amazon MW Apache Airflow and AWS EC2 QuickSight and SNS, Cloudera infrastructure – Ambari, Hortonworks Sandbox and Zeppelin Notebook

Databases and Storage

- AWS RDS, MySQL, MongoDB, AWS DynamoDB, SQLAlchemy, PostgreSQL, AWS S3 Buckets, SQLite, Azure, Oracle.

Career Experience

Data Engineer | TC Energy | Houston, TX | 09/2019– Present | Odfjell Terminal | Seabrook, TX | 02/2019– 09/2019 | Enterprise Products (Through CF) | Houston, TX | 10/2018 – 02/2019

- Design and implementation of process data pipeline for asset data migration into AWS cloud environment
- Built Snowflake Data Pipeline using Amazon Kinesis Firehose starting from the AWS EC2 logs to storage in Snowflake and AWS S3 bucket post-transformation and orchestrating through Amazon Managed Workflows for Apache Airflow (MWAA) DAGs.. Programming language used are Pyspark, Python, SparkSql
- Architecting data pipeline with Cloudera infrastructure – Ambari, Hortonworks Sandbox using Zeppelin Notebook. Programming language used are Python, SQL
- Created DataStream using AWS kinesis DataStream and kinesis firehose. Programming language used are Pyspark, Python, SparkSql
- Migration solution from SAP and Maximo to Data Lake.
- Solved data quality issues using AWS Databrew and Apache PySpark on AWS EMR and Databricks on AWS EC2 Clusters for data wrangling and transformations.
- AWS Project monitoring using AWS Lambda and Aurora
- Performed SQL data analysis using Oracle Database
- Regular meeting with stakeholders to gather user requirements and ascertain objectives.

Senior Data Analyst | Tesoro Logistics (Through IG) | San Antonio, TX | 04/2017 – 10/2018

...continued...

- Design and implementation of process data pipeline and loading of structured data for performing quantitative and qualitative asset risk modelling, predictive analysis for program budgeting and forecast.
- Created process safety risk reduction metrics

Project Engineer | Weatherford | Houston, TX | 09/2013– 03/2016

- Performed ETL of drilling well bottom hole pressure '.mbd' data from field and wireline instrumentations (bottom hole pressure, well depth and true vertical height of drilling bit).
- Cleaned and transformed data into a '.csv' data file. Load data into Weatherford microflux database for analysis.
- Project management of secure drilling operations of over 80 wells in Canada, Iraq, Nigeria, Angola, and Cameroon for clients such as Repsol, Shell, Exxon Mobile, Chevron and Total

Education & Certifications

University of North Texas | Master of Science | Mechanical and Energy Engineering| Denton, TX

Rice University | Certificate | Data Analytics | Houston, TX

License

Licensed Texas Professional Engineer | Texas Board of Professional Engineer and Land Surveyors

Additional Skills

Excellent project and time manager with outstanding excellent presentation skills, a great team player and ability to extend empathy.