Mzamo Ntshaba

WORK EXPERIENCE

Data Engineer(Learnership) // Sandtech (Remote)

15 January - 28 February 2024 :

- In one of our capstone Projects we had to use Python together with Pandas Dataframes to develop Eskom key metrics to aid data scientists in generating actionable insights

1 March - 31 March 2024 :

- Normalized an unstructured database and optomised the queries using CTE's to improve performance and data integrity

1 April - 30 April 2024 :

- Deployed a static website on AWS Amplify then Utilized AWS API Gateway to trigger a AWS Lambda function for backend processing where user data are stored in a NoSql database AWS Dynamo DB to be integrated with a NLP tool AWS Comprehend for Sentiment Analysis then auto generates a smart email response using AWS SNS

1 May - 31 May 2024 :

- Configured AWS Cloud Formation using YAML templates and then used RDP client to connect to the EC2 instance inside the Virutal Private Cloud , for further security of the data migration had to SSH to a linux instance then used NFS protocol to transfer the data via the file gateway through to the storage gateway then finally into the S3 which served as the storage solution

- Part two involved creating ticker values in AWS DynamoDB then using AWS Lambda and Kinesis Firehose for real time streaming and monitoring

1 June - 30 June 2024 :

- In our specialisation capstone project i learnt how to create a robust ETL pipeline where I stored raw sensitive data in a private S3 bucket this setup allows for CI/CD Then mounted the bucket in an EC2 instance. launched a Docker container with an Airflow environment where i orchestrated the ETL process of transforming historical stock data and loading it into an RDS database, Automation was triggered by success file being dropped in a monitored s3 bucket

1 July - 31 july 2024 :

- Utilized PySpark for large scale data manipulation and analysis

Data Engineer(Internship) //Nedbank (Hybrid)

1 September 2024 - 31 August 2025 :

- Developing and automating Power BI dashboards to streamline reporting and deliver insights to key stakeholders

- Utilizing SSIS to extract , transform and load data from various sources into the data warehouse , implementing data modelling to optimize data structures and ensuring DQ standards are met through monitoring and cleaning

OBJECTIVE/SUMMARY

As a highly driven graduate data engineer , I am eager to apply the skills that i have acquired at Sandtech to drive innovation and make a meaningful impact , i am commited to continuous learning and using my expertise to transform complex datasets into actionable insights that support strategic decision making , I thrive in dynamic environments and I am very passionate about leveraging technology to deliver high quality solutions

SKILLS

- Python
- Pandas
- MySql
- AWS
- Docker
- Apache Airflow
- Apache Spark
- Team Collaboration
- Effective Communication

CERTIFICATIONS & LICENSES

Python(Basic)
HackerRank
03 October 2023

SQL(Intermediate)
HackerRank
23 April 2024

Google Big Data and Machine Learning Fundamentals
Coursera
07 April 2024

EDUCATION

NQF4 National Senior Certificate
Empangeni High School, South Africa
I January 2014 - 31 December 2018

NQF5 MICT SETA ACCREDITATI0N
Sand Tech Academy, South Africa
1 January 2024 - 31 December 2024

Bachelors Degree in Applied Artificial Intelligence (Scholarship)
IU International University of Applied Sciences, Germany
1 November 2024 - 1 November 2027