Data Engineer

Ahmed Moustafa

ahmedmustafa045@gmail.com

01555740019

Objective

Data Engineer/ETL Developer with good knowledge in the field of data including data Integration, Data Modeling, Data warehousing, SQL, with passion to learn and discover everything related to Data Engineering field.

Education

Bachelor Of Engineering – Mansoura University 2017

Skills

Concepts
Data Warehouse – Data Lake
Data integration - Data Modeling
Data Quality- Big Data

Big Data Frameworks
Apache Hadoop
Apache Spark (DataFrames, Spark-SQL)
Microsoft Azure – Databricks
AWS (CL2 – S3 - RDS)

Programing Languages
Python

ETL & ELT Tools
SSIS

Orchestration
Apache Airflow

Database
MSSQL - Snowflake

VERSIONS CONTROL
Git - GitHub

Software Experience
Android Applications using Android Studio

Containers
Docker

Projects

Orchestrate a parallel processing ETL pipeline on AWS

- Extract data from open API and ensure it is existed, make some transformation to then load to AWS RDS Postgres
- load csv file on data lake into Postgres
- orchestrate a parallel processing between the 2 pipelines and make join between two tables and then load in S3 storage layer

Technologies:
- AWS CL2, S3, RDS Postgres, Airflow, Python

Build and Automate loading data from S3 to Snowflake

-Create our database on snowflake then Staging layer that point to S3

-Orchestrate process from sensor data exist on S3, Create table on Database, Copy data from Staging to created table and finally send mail notification

Technologies:
- AWS CL2, S3, Airflow, Python, Snowflake

Data Pipelines on Azure

-First create storage layer(bronze-silver-gold), data factory, key vault, sql database and sql server

-Load Adventure Works data source into data lake with parquet format

-Then load data into Databricks then make snapshot of it in silver layer to make transformations on it using python

-load transformed data as Data Mart with dimensions and fact into gold layer and finally generate dbt docs for all process

Technologies:
- Microsoft Azure, Databricks, Python, dbt

Visa Issued Between 2006:2017

visa issued csv data source from Kaggle we import to python then use Apache Spark for data Cleaning, formatting, creating new column and finally make visualization using panda and plotly with the help of Microsoft Azure as we create a VM to make processing using master and 4 workers

Technologies:
- Python, Apache Spark, Ubuntu, Docker-compose for os, Microsoft Azure for VM, Visualization using panda and plotly

E-COMME RCE

- I worked with AdventureWorksDW 2022 (DWH) from Microsoft for learning purposes
  Build and Maintain Data Pipeline using SSIS, ETL from Multiple Sources to Staging, then Migrate Data from Staging to DW
  Create Report using Power BI for Total Sales Quantity

Technologies:
- SQL Server. SSIS

Slowly Changing Dimension

-Transform data from multiple sources of data Database, flat files like CSV, xlsx, txt by reject null values, add derived column for more information or some replacement validate time stamp then loaded data to Data Warehouse
-Deployment and make job automated

Technologies:

- SSIS, SQL Server.

-Incremental Data Load for Dimension and Fact Tables:

-preparing Staging tables
-Migrate Data using Lookup and Conditional Split to split data into 3 parts New Records, SCD Type1 and SCD Type2
-Migrate data to Staging then Update Dimension With SCD Type 1, prepare for SCD Type 2 then Migrate SCD Type 2

-Load Fact Table by:

-preparing Staging tables
-Migrate Data using Lookup and Conditional Split to split data into 2 parts New Records and updated Records
-Update Fact Table From Staging table

Technologies:

- SSIS, SQL Server.

Experience

- **SQL & SSIS for ELT Jobs**
- **Apache Spark for Data Processing Structure and semi-Structure**
- **Technical Support Engineer at Telecom Egypt**

Under Learning

- PL-SQL
- NoSQL

Courses and Certificates

- SQL (Hackerrank)
- SSIS(LearnSSIS)
- Modern DataEngineering(codewithyu)
- coder2j

Languages

- Arabic -Mother tongue
- English – Good