Diea Abdeltwab *Data Engineer*

diea.abdeltwab@gmail.com | diea.online | +201557050969 | Cairo, Egypt
github.com/diea-abdeltwab | linkedin.com/in/diea-abdeltwab

PROFILE

Data Engineer who designs and builds end-to-end **batch and real-time data pipelines**, from data ingestion and **ETL/ELT** to **dimensional modeling** (Star Schema, SCD Type 2) and data warehousing. Hands-on with **Python, SQL, Apache Spark, PySpark, Apache Kafka, and Apache Airflow**, with cloud experience across **Microsoft Azure, Databricks, Snowflake, and Microsoft Fabric**. Focused on building reliable, maintainable data pipelines. ITI Data Engineering graduate and on-demand **Apache Airflow instructor**.

SKILLS

Programming & Query Languages	**:** Python, SQL, Bash, Pandas, NumPy
Data Modeling & Warehousing	**:** Data Modeling, Dimensional Modeling, SCD Type 2, CDC, Data Warehousing
Data Architecture	**:** Data Lake, Lakehouse, Medallion Architecture
Databases & Storage	**:** PostgreSQL, SQL Server, MySQL, MongoDB, ClickHouse, MinIO
ETL/ELT & Orchestration	**:** ETL, ELT, Data Integration, Data Quality, Apache Airflow, dbt, SSIS
Data Streaming & Ingestion	**:** Apache Kafka, Apache Flink, Spark Streaming, REST APIs, Web Scraping
Big Data & Processing	**:** Apache Spark, PySpark, Apache Hadoop, Apache Hive, Delta Lake
Performance	**:** SQL Performance Tuning, Query Optimization, Indexing
Cloud & Cloud Data Platforms	**:** Microsoft Azure, Microsoft Fabric, Databricks, Snowflake
BI & Visualization	**:** Power BI, Grafana, Metabase, Excel
DevOps & Tools	**:** CI/CD Pipelines, Docker, Git, GitHub, Linux

EXPERIENCE

Data Engineering Instructor | ITI (On-Demand)
06/2026 – Present | Remote

- Designed and developed a complete **Apache Airflow** course for the ITI Data Engineering Track, from curriculum design to hands-on labs.
- Deliver **Airflow** sessions to new cohorts on request, covering **DAG** design, task dependencies, **scheduling**, **sensors**, and workflow **orchestration**.
- Built **ETL** labs where students orchestrate end-to-end data pipelines with **Airflow**, **Python**, and **SQL**.
- Coach students from diverse technical backgrounds through **debugging** and **troubleshooting** Airflow pipelines, adapting explanations to each learner's level.

Data Analytics Intern | Elevvo
08/2025 – 09/2025 | Remote

- Cleaned, transformed, and analyzed business datasets using **Python (Pandas)** and **SQL**.
- Built **Power BI dashboards** that tracked KPIs and trends, turning raw data into clear insights.
- Communicated findings through concise, easy-to-read reports.

PROJECTS

NYC MTA Bus Data Pipeline

- Ingested **Transitland** and **GTFS-Realtime API** data (batch and real-time) into **Azure Blob Storage**.
- Built a real-time streaming pipeline with **Apache Kafka** and **Spark Structured Streaming** for live bus tracking.
- Orchestrated batch **ETL** with **Apache Airflow**, deployed with **Docker** on an **Azure Virtual Machine**.
- Implemented **SCD Type 2** dimensions in **PostgreSQL** and served analytics from **ClickHouse**.
- Delivered a **Power BI** dashboard for daily batch reporting and a **Metabase** dashboard for real-time monitoring.

E-commerce Data Pipeline

- Ingested raw CSV files from **MinIO (S3-compatible object storage)** and loaded them into a MySQL staging layer.
- Transformed data with **PySpark** into a **Star Schema** data warehouse model (fact and dimension tables).
- Loaded data into **ClickHouse** for fast **OLAP analytics**, orchestrating **ETL** with **Apache Airflow**.
- Built **Power BI** dashboards on sales, customers, and delivery performance.

Real-Time Crypto Exchange Pipeline

- Streamed live cryptocurrency and fiat exchange prices from the **CoinGecko API** into **Apache Kafka**.
- Joined and aggregated real-time streams using **Apache Flink SQL** and stored results in **ClickHouse**.
- Built **Grafana** dashboards to monitor real-time price trends and exchange rates.

NYC Taxi Data Pipeline

- Ingested NYC Taxi trip **Parquet** files from **MinIO** object storage.
- Built a **PySpark ETL** pipeline modeling the data into a **Star Schema** in **PostgreSQL**.
- Visualized trip, revenue, and demand insights in **Power BI**.

EDUCATION

Data Engineering Track, Information Technology Institute (ITI) 03/2025 – 07/2025

- Five-month intensive program covering data engineering, big data architecture, ETL, data warehousing, Apache Spark, Apache Kafka, and Apache Airflow.

B.Sc. in Electronics and Communication Engineering, Fayoum University 2019 – 2024

- Grade: Very Good
- Graduation Project: V2X Communication – Embedded Software (Excellent)

CERTIFICATIONS

IBM Data Engineering Professional Certificate, IBM 09/2026

Microsoft Fabric Data Engineer DP-700 Professional Certificate, Microsoft/Coursera 11/2026

Professional Data Engineer in Python, DataCamp 09/2026

Associate Data Engineer in Databricks, DataCamp 10/2026

Microsoft Azure Fundamentals (AZ-900), DataCamp 09/2025

Data Engineer in Python, DataCamp 09/2025

Associate Data Engineer in Snowflake, DataCamp 06/2025

Associate Data Engineer in SQL, DataCamp 06/2025

Docker for Beginners with Hands-on labs, KodeKloud 06/2026

Introduction to Git and GitHub, Google 06/2026