

Ahmed Refat Abd Elmoaty

Data Engineer

Cairo, Egypt | ahmedrefat1412@gmail.com | +20 01020519935 [GitHub](#) | [LinkedIn](#) | [Portfolio](#)

Professional Summary

Data Engineer with strong experience in Python, SQL, PySpark, and ETL pipelines. Skilled in building batch and streaming data pipelines, data modeling, and working with structured and semi-structured data. Experienced with Airflow and cloud-based data platforms, and passionate about scalable big data solutions.

Education

Bachelor of Computer Science and Information Systems

Kafr El-Sheikh University, Faculty of Computers and Information – Information Systems Department

Graduation: May 2023

Experience & Training

Microsoft Data Engineer Trainee

Digital Egypt Pioneers Initiative (DEPI) | Nov 2025 – Present

- Gained hands-on experience building data pipelines using Python and SQL.
- Worked on data modeling, ETL concepts, and cloud-based data platforms

Cloud Services Management & Operations Trainee

National Telecommunication Institute (NTI) | Jan 2026 – Mar 2026

- Learned cloud fundamentals, networking basics, and cloud service operations.
- Practiced deploying and managing services on AWS and Linux environments.

Technical Skills

- **Programming & Data processing** : Python(Numpy,Pandas,Matplotlib,Seaborn) ,SQL
- **Databases & Data Warehouse** : Data Modeling , SQL Server , Advanced SQL , MongoDB, Postgres
- **BIG Data & Data Governance** : Hadoop ,HDFS ,Spark , Kafka
- **ETL & Orchestration** : Apache Airflow ,dbt
- **Cloud & DevOps Tools** :AWS, Snowflake , Docker , GitHub ,Linux
- **BI & Visualization** :Power BI (DAX)
- **Languages** :Arabic(native) , English

Projects

E-Commerce Pipeline Data Engineering | [GitHub](#)

Built a data engineering pipeline that ingests data from MongoDB and CSV files, processes and transforms it using PySpark, and loads analytics-ready datasets into PostgreSQL. Visualized insights with Power BI, orchestrated workflows with Airflow inside Docker for a scalable and automated solution

NYC Smart Traffic End-to-End Pipeline Project | [GitHub](#)

Building an end-to-end data pipeline for NYC traffic data, handling batch and streaming ingestion, distributed processing with PySpark, and delivering analytics-ready data to cloud storage and a data warehouse through automated transformations.

Startup Expansion Analytics Dashboard | [GitHub](#)

Developed an analytical dashboard to evaluate startup expansion performance, focusing on data cleaning, transformation, and visual storytelling to support business decision-making.

Sales Data Analysis Dashboard | [GitHub](#)

Analyzed sales data and built interactive dashboards to track performance metrics, trends, and insights through structured data modeling and visualization.