

Mohamed Taha Abo Heiba

Egypt | +20 114 971 1922 | Email: mohamed-aboheiba@outlook.com | Portfolio: motahaaboheiba.github.io

LinkedIn: linkedin.com/in/mohamed-taha-abo-heiba | GitHub: github.com/MoTahaAboHeiba

PROFESSIONAL SUMMARY

Data Engineer with hands-on experience designing and building **ETL pipelines**, **batch processing workflows**, and **real-time streaming data pipelines** on **Databricks**, **Hadoop**, and **AWS**. Processed 200K+ records through production-grade **Medallion Architecture data lakehouses** from raw ingestion to analytics-ready dimensional models. Skilled in **distributed computing** (Apache Spark, Kafka, Flink) and cloud-native data platform development on AWS. **Awarded best-performing trainee** in NTI's Huawei Big Data program. **HCIA-Big Data certified** and **AWS Certified Cloud Practitioner (CLF-C02)**.

EXPERIENCE & TRAINING

- **AWS Cloud Engineer Trainee | National Telecommunication Institute (NTI)** *Oct 2025 – Dec 2025*
 - Architected AWS web app from S3 static hosting to multi-tier EC2/RDS/VPC architecture, **applying Well-Architected Framework principles**
 - Configured **RDS Multi-AZ** and **S3 lifecycle policies** to ensure high availability and cut cold storage costs via **Glacier**
- **Big Data Engineer Trainee | National Telecommunication Institute (NTI)** *Jul 2025 – Sep 2025*
 - **Awarded best-performing trainee**; earned exam voucher and passed HCIA-Big Data certification on first attempt
 - Built GB-scale **Spark batch pipelines** on Hadoop (**HDFS/YARN/Hive**) and **Kafka+Flink streaming pipelines** for real-time log ingestion across 15+ labs each
- **Data Engineer Trainee | Digital Egypt Pioneers Initiative (DEPI)** *Apr 2024 – Oct 2024*
 - Built 4 **SSIS packages** ingesting 7 CSV sources into **SQL Server** via 17 stored procedures, automating **ETL workflows** that reduced manual integration effort and accelerated data delivery
 - Delivered validated dataset across 70K records powering churn model (97.4%), identifying at-risk customers for retention

PROJECTS

- **E-Commerce Lakehouse Databricks [GitHub](#)** | Databricks, PySpark, Delta Lake, Unity Catalog, Jobs
 - Built an end-to-end **cloud data lakehouse pipeline** processing 200K+ records from 6 data sources through a 3-task Databricks Jobs DAG (**Bronze** ingestion → 6 script **Silver** transformation → **Gold** star schema)
 - Delivered a governed data platform with 1 fact table and 2 dimension **Delta Tables** in **Unity Catalog** via **Medallion Architecture**, enabling auditable and analytics-ready data access for downstream consumers
- **SQL Data Warehouse Project [GitHub](#)** | SQL Server, T-SQL, Star Schema, Medallion Architecture
 - Built a fully scriptable **warehouse** triggered end-to-end via a single script; 2 master stored procedures handling Bronze→Silver ETL across 6 CRM/ERP sources, loading **60K+ records** into the fact layer
 - Delivered **star schema** as SQL views with **per-step execution-time logging**, cutting ETL debugging time by 25% and enabling bottleneck detection
- **Document Indexing & Retrieval Pipeline (EduMate RAG) [GitHub](#)** | FastAPI, LangChain, ChromaDB, Groq
 - Built a production RAG API (12 endpoints) ingesting and indexing 13K+ document chunks into ChromaDB via an automated PDF processing pipeline
 - Achieved sub-100ms **semantic search** with multi-turn context retention via **LangChain conversational memory**
- **Online Retail Analytics - Power BI & ETL Project [GitHub](#)** | Power BI, Power Query, DAX, SQL
 - Built **3-page Power BI Report** (30+ visuals, 5 DAX) over 9.73M revenue across 4k+ customers in 38 countries after cleaning 541K to 406K records and engineering 8 features; **surfaced top 25% of customers driving 65% of revenue**

CERTIFICATIONS

Huawei HCIA — Big Data Associate — Huawei Technologies Co., Ltd

Issued: Mar 2026 | Valid: Mar 2029

AWS Certified Cloud Practitioner (CLF-C02) — Amazon Web Services

Issued: May 2026 | Valid: May 2029

TECHNICAL SKILLS

Big Data & Distributed Computing: PySpark, Spark SQL, Databricks, Hadoop (HDFS, YARN, Hive), Kafka, Flink, Delta Lake
Data Engineering & ETL Pipelines: Medallion Architecture, Data Ingestion, (Batch, Stream Processing), ETL/ELT Workflows
Data Warehousing: SQL Server, Dimensional Modeling (Star / Snowflake Schema), T-SQL, Stored Procedures, Unity Catalog
Databases: PostgreSQL, MySQL, DynamoDB, HBase
Programming & DevOps: Python (Pandas, NumPy), Bash, Docker, Git / GitHub, Apache Airflow
Cloud & AWS Data Services: AWS (S3, EC2, RDS, VPC, IAM, Lambda, Glue, Redshift, Athena, EMR, ELB, Auto Scaling)
Analytics & Visualization: Power BI, DAX, Power Query, Matplotlib, Pipeline Monitoring & Logging

EDUCATION

B.Sc. Computer Science — Faculty of Computers & Artificial Intelligence, Fayoum University *Oct 2022 – Jun 2026*

GPA: 3.4 / 4.0 | **Core Coursework:** Big Data Engineering, Cloud Computing, Database Management Systems, Advanced SQL, Data Structures & Algorithms, Machine Learning, Operating Systems (Linux, Docker), Python Programming, Data Visualization (Power BI)