YOUSEF MAHMOUD
Data Engineer

Cairo, Egypt   01004389030   Yousef.soliman.de@gmail.com

linkedin.com/in/y0usefma7m0ud   github.com/YOU5F   y0u5f.github.io

Professional Summary

Junior Data Engineer with a strong foundation in Computer Science, AI, and Cloud Computing (AWS). Currently a Data Engineering Trainee at ITI, with hands-on experience in building scalable ETL pipelines (Airflow, Spark), Data Warehousing (Snowflake, dbt), and deploying ML models. Proficient in DevOps tools (Docker) and Agile methodologies. Passionate about transforming raw data into actionable insights using modern data stacks.

Education

Information Technology Institute (ITI)   **Minya, Egypt**
Data Engineering Trainee   *June 2025 – Nov 2025*
- Developing end-to-end data pipelines using Apache Airflow, Spark, and Python in an intensive 4-month program.
- Designing and implementing Data Warehouses using Snowflake and dbt, reducing query latency by 30%.
- Collaborating with cross-functional teams to integrate AI models into data workflows using Docker containers.

Beni-Suef University   **Beni-Suef, Egypt**
Bachelor of Computer Science   *2020 – 2024*

Experience

Instant Software Solutions   **Hybrid**
AI Engineer Intern   *May 2025*
- Implemented 6 AI projects including recommendation systems, Instagram scraping, and Twitter sentiment analysis.
- Developed chatbots and real-time emotion recognition systems leveraging NLP and Computer Vision techniques.
- Optimized ML models using advanced algorithms (ALS, SVD) and processed diverse datasets with Pandas/NumPy.

Projects

Hollywood in the Cloud | *AWS (Lambda, Glue, Athena, S3), Power BI* [GitHub]
- Built a Serverless Data Pipeline to ingest 100+ years of movie data from TMDB API using AWS Lambda.
- Automated ETL processes with AWS Glue and enabled SQL-based analysis on S3 data via Amazon Athena.
- Visualized insights identifying Q4 as the peak release window and market trends using Power BI.

Olist E-commerce ELT Pipeline | *Snowflake, Airflow, dbt, Docker* [GitHub]
- Architected a scalable ELT pipeline processing 100K+ records using Medallion Architecture (Bronze, Silver, Gold).
- Orchestrated daily workflows using Apache Airflow and containerized the environment with Docker.
- Implemented dbt for transformation with 26+ automated tests and modeled a Star Schema for analytics.

ITI Examination System | *SQL Server, T-SQL, SSMS, Docker* [GitHub]
- Engineered a normalized relational database with complex Stored Procedures for automated exam generation.
- Established a secure Role-Based Access Control (RBAC) system and created Audit Triggers for data integrity.

Skin Cancer Detection (AI Platform) | *Python, TensorFlow, Flutter*
- Spearheaded an AI healthcare platform for skin cancer detection using CNNs, achieving **96%** classification accuracy.
- Built a patient ecosystem with drug reminders and chat modules; awarded "Excellent" for technical impact.

Technical Skills

Languages: Python (Advanced), SQL, Bash Scripting, Java, C++
Data Engineering: Apache Airflow, Apache Spark, Hadoop, Kafka, dbt, MongoDB
Cloud & DevOps: AWS (S3, EC2, Glue, EMR, Redshift), Docker, Linux, Git/GitHub, Jenkins
AI & Visualization: Machine Learning, NLP, Power BI, Matplotlib
Soft Skills: Agile Project Management, Problem Solving, Team Leadership, Critical Thinking
Languages: Arabic (Native), English (Advanced)

Certifications & Training

AWS Academy Graduate - Data Engineering & Cloud Foundations *Amazon Web Services (AWS)*   **2025**

Artificial Intelligence Diploma (170 Hours) *Instant Software Solutions*   **2024**

Fundamentals of Agile Project Management *Project Management Institute (PMI)*   **2022**

Technical Courses: Apache Spark (Frank Kane), Docker for DevOps (Mumshad Mannambeth), MongoDB (Maximilian), Modern Data Warehouse (Baraa Salkini)