

Mohamed Ali

Data Engineer

✉ Mohamed.Ali.c2021@Gmail.Com ☎ +201200104652 📍 5th Settlement, Cairo 11835, Egypt

in LinkedIn 🐙 Github 📁 Portfolio 🛡 Military Service Completed 🇪🇬 Egyptian 🌐 Single

Experience

Data Engineering, Information Technology Institute (ITI)

Jul 2023 – Dec 2023
Alexandria, Egypt

Relevant Projects: Azure Cloud Services: ECDC Covid-19 Reporting; Gravity Books Project; HealthCare Project; Adventure Work DataMart; Dockerized Airflow Data Pipeline; PySpark Batch & Streaming Data Analysis; Real-Time Sales Streaming Data Pipeline

Data Scientist & Business Analytics Intern, The Sparks Foundation

May 2023 – Jun 2023

Relevant Projects: Sentiment Analysis and Prediction for Stock Market Prices; Facial Recognition and Emotion Detection.

Education

Bachelor of Science, Major in Physics and Computer Science, Faculty of Science, Helwan University

Sep 2016 – Jul 2021
Cairo, Egypt

Graduation Project: Machine learning Model Predicting Critical Temperature of Superconductor.

Projects

Azure Cloud Services: ECDC Covid-19 Reporting

- Ingested data from diverse sources (HTTP, Azure Blob Storage) into Azure Data Lake Gen2 using Azure Data Factory (ADF).
- Utilized SQL for data querying, Data Lake, Dataflow, Data Bricks and HDInsight to store and process, CI/CD methodology, SQL OLTP Database and Azure Synapse OLAP DWH reporting using Power BI.

AWS: COVID-19 Full Stack Data Project for Analysis

- Utilized AWS Amazon S3 - Storing the data ,Crawler - Used to extract all the schema and information straight from S3, Amazon Athena - Running ad hoc sql queries on the available data in S3.
- AWS Glue - Data transformation, Amazon Redshift - Storing the transformed dimensional model in data warehouse.

Data Warehouse and BI: Gravity Books Project

- Developed a comprehensive data warehouse for Gravity Books.
- Integrated data using SSIS, leveraged SSAS for analytical layers.
- Enabled dynamic reports and dashboards in Power Pivot and Power BI.

Data Warehouse and BI: Healthcare Project

- Established a Healthcare Data Warehouse.
- Ensured integrated, quality, and secure healthcare analytics.
- Improved decision-making and patient care through streamlined data processes.

Big Data: Real-Time Stock Streaming Data Pipeline

- The Real-Time Stock Streaming Data Pipeline is a comprehensive project designed for my master's thesis, focusing on stream processing.
- It leverages Finnhub.io API/websocket for real-time trading data, providing a showcase of key aspects in streaming pipeline development and architecture.
- The goal is to achieve low latency, scalability, and availability.

Big Data: Real-Time Sales Streaming Data Pipeline

- This project is a comprehensive sales streaming data pipeline designed to handle real-time data processing, storage, and visualization.

- It utilizes Docker containers for easy deployment and scalability, Apache Airflow for data ingestion, Kafka as a message broker, Spark for stream processing, and Cassandra and ClickHouse as serving databases.
- The pipeline culminates in a Superset-powered dashboard for insightful data visualization.

Dockerized Airflow Data Pipeline

- Dockerized Apache Airflow data pipeline that fetches user data from two different sources and inserts it into a PostgreSQL database.
- The first data source is the RandomUser API, and the second is an existing MySQL database.
- The pipeline consists of two separate DAGs, one for each data source, and utilizes PythonOperator and SqlOperator tasks to execute the data transfer operations.

PySpark Batch & Streaming Data Analysis

- Conducted practical sessions on Spark batch and streaming processing using PySpark.

Machine Learning: Model Predicting Critical Temperature of Superconductors

- Analyzed and predicted superconductor critical temperatures using Pandas, NumPy, and Scikit-learn.
- Contributed to advancements in superconducting materials.

Machine Learning: Sentiment Analysis and Prediction for Stock Market Prices

- Used Python, Pandas, Scikit-learn, NLTK, and Vader Sentiment Analysis for data preprocessing.
- Trained and evaluated nine ML models for sentiment analysis, achieving impressive accuracy.

Machine Learning: Facial Recognition and Emotion Detection

- Developed a Face and Emotion Detection system using TensorFlow and Keras, achieving high accuracy in real-time emotion classification.
- Integrated pre-trained models for face detection and utilized OpenCV for efficient video processing.
- Led end-to-end project development, showcasing expertise in deep learning, computer vision, and collaborative problem-solving

Exploratory Data Analysis: Prosper Loan

- Utilized Python, Pandas, Matplotlib, Seaborn, and NumPy for loan data analysis.
- Conducted exploratory data analysis on loan characteristics.



Skills

Data Base (SQL Server, PostgreSQL, MongoDB, Cassandra, Relational and Non-Relational, RDBMS)

Data Architectures & Modeling (Bill Inmon, Ralph Kimball, Data Lake, Delta Lake, Data Lakehouse)

Big Data (Apache Spark, Apache Kafka, Hadoop Ecosystem (Hive, Impala, HBase, Yarn), Databricks)

Cloud (Azure, AWS) | **ETL & Workflow Management** (SSIS, Apache Airflow, Apache Ni-Fi, Talend)

Programming Languages (Python, OOP, SQL) | **Machine Learning** (TensorFlow, Scikit-Learn)

Data Analysis (NumPy, Pandas, SSAS) | **Visualization & Dashboard** (Matplotlib, Grafana, Power BI)

CI/CD & Virtualization (Kubernetes, Git, Jenkins, Docker, VMware)

Operating Systems (RedHat, UNIX/Linux Administration, Bash, and Shell Scripting)

Soft (Excellent Communication, Team Player, Problems solving, Passionate Learner)



Certificates

- Information Technology Institute (ITI) Data Engineering
- IBM Data Engineering Professional Certificate
- IBM Data Warehouse Engineer Professional Certificate
- Microsoft Azure Data Engineering Associate (DP-203) Professional Certificate
- Ministry of Communications and Information Technology Nanodegree Program Advanced Data Analysis
- Stanford University & DeepLearnig.AI Machine Learning Specialization
- Data Bricks Fundamentals of the Databricks Lakehouse Platform Accreditation (V2)