Hazem Omar Mohamed **Atwa**

Cairo, Egypt. 01118559305
hazematwwa@gmai.com
linkedin.com/in/hazem-omar

My objective is to get involved in various large-scale projects that are related to data engineering including creating data pipelines, ETL/ELT, data-oriented system design, etc. Also, to be a team player and to interact with different mindsets and cultures to push myself and my company forward.

EXPERIENCE

Sr. Data Engineer & Cloud | Dotpy (FULL TIME)

JUN 2023 – PRESEN

As part of AI-Pro track at Information technology institute (ITI), I conduct cloud and big data fundamental's courses and implement them on AWS.

Big Data Topics:

- Basic storage and database systems (OLAP and OLTP, Row vs columnar storage, ACID transactions, CAP theorem, etc.)
- Hadoop ecosystem (HDFS, MapReduce, Hive, etc)
- Apache Hive and Apache Spark Architecture and use cases.
- NoSql-db like mongo-db and cassendra to how handle schema less with a lot of changeable of data

Cloud Topics:

- Mainly AWS, explaining basic AWS services like Networking, IAM, S3, EC2, etc.
- MS Azure to build a pipeline from end-to-end

Teaching Assist (TA) | AAST | CS Faculty (SESSIONAL)

APR 2022 – 2024 OCT

Teaching Assistant in Computer and Artificial Intelligence Faculty) .Specialist in AI & Big Data Department

Courses Related for Big Data:

- Database Theory (OLTP)
- DWH (OLAP)
- Data Mining
- Big Data & Streaming Analysis (Hadoop Echo System)
- Python Programming Language
- Linux Administration (Bash Script)

Data engineer II | Emuratic solution (FULL TIME)

MAY 2021 – MAR 2023

- Use Talend for big data to design, develop and automate data pipelines that perform a specific task -usually data transformation and integration-

- Develop Python Py/SQL code and sometimes integrate that code in Talend jobs to automate it.
- Write Linux bash scripts and use SQL*Loader on Linux to load very large files to the database.
- Deal with vendors from different companies and facilitate/maintain their applications.

1- **Create distributed system for how treat with big data:**
   a. Create distributed system with different machine and make a network between all this machine and store the

2- **Create a pipeline to handle a different velocity:**
   Use all tools to how make a pipeline for how treat with verity of data and process it, using a batch query processing and real time analysis with different services.

3- **DevOps and CRC:**
   Achieved required DevOps tasks that ease team daily operations:
   a. Implemented a CI/CD environment using Gitlab.
   b. Installed and configured services with containerization technologies using Docker

ETL Developer | Vezzeta (FULL TIME)
MAY 2018 – MAR 2020

In this role, I help clients design, deploy, maintain, modernize and migrate data lakes, data warehouses and data pipelines using cloud technologies, mainly using AWS. Projects:

1. **Major data lake migration and modernization:**
   My contributions involved redesigning data pipelines using AWS services, refactoring legacy codes, configuring CI/CD pipeline and monitoring dashboards. Technologies used:
   - Scala Spark on AWS EMR for processing.
   - Elasticsearch for logging and monitoring.
   - Apache Airflow for orchestration.
   - Write IaC using CDK to provision all resources.
2. **Migrate and Redesign SQL server Architecture**:
   That involved designing a highly available and disaster recoverable SQL Server instance installed on EC2 in a multi-AZ architecture. Technologies used: SQL Server on EC2 including Always on availability group and log shipping between 2 availability zones.

EDUCATION

BSC of Computer Science | Faculty of CS & AI | MTI
OCT 2014 – JUN 2018

Major: Computer Science, GPA: 3.4

Minor: Computer Science (CS)

Graduation Project: Build a full software cycle of development in health and blood donation system to extract the main knowledge help CEO to make a decision on it

Grade: A

Certification:

- AWS Certified Data Analytics – Specialty, 2023.
- Google Cloud Professional Data Engineer, 2023
- AWS Certified Solutions Architect – Associate, 2021.

MSC of Computer Science | Faculty of CS | AAST
JAN 2021 – APR 2024

Major: Computer Science, GPA:3.8

Minor: Computer Science (CS)

Research Topic: Especially Big Data Field in Hadoop echo system in job scheduling and how make a fair scheduling with very large amount of data and high traffic jobs without missing jobs

Related coursework:

- Cloud DevOps Engineer nanodegree – Udacity, 2020.
- Big Data & Data Lake Specialization – Coursera, 2021.
- Python and MySQL courses – Udemy, 2020.
- RDBMS for (OLTP) - IBM 2020.
- DWH for (OLAP) – Coursera 2021.
- LPIC-1 Linux system administration – LPI Egypt partner, 2022.
- Docker & Kubernetes for data engineer

SKILLS AND ACTIVITES

- Cloud skills specially AWS.
- Linux scripting and administration.
- Git and version control.
- Docker and Kubernetes on AWS.
- DevOps with CircleCI, and Jenkins with Blue Ocean.
- 3D design with Solidworks.
- Languages I speak:
    1- Arabic: Native.
    2- English: Fluent.
    3- French: B1.