

# AHMED HASSAN

## Big Data Engineer

Address: Nasr city – Cairo | Mobile: 01016445683

[Email](#) | [LinkedIn](#) | [GitHub](#)

---

### SUMMARY

I am a highly accurate and experienced Data Engineer with a strong proficiency in collecting, analyzing, and interpreting large datasets. I possess extensive analytical skills that enable me to effectively derive valuable insights from complex data. My expertise in data engineering tools and technologies allows me to develop efficient data pipelines and optimize data workflows. I am committed to delivering high-quality results and collaborating effectively with cross-functional teams to drive data-driven decision-making.

### EDUCATION

**B.Sc.:** Zagazig university [2015-2020]

**Major:** Computer Engineering

**Grade:** Very Good

**Graduation project:** Autonomous Car with Excellent Grade

**Description:** Build Autonomous System in Car Has Ability to Track Eyes for driver to warn him during driving.

### WORK EXPERIENCE

**FEB 2023 – current**

**Giza Systems, Egypt — senior Data Engineer**

#### **Ministry of Interior in KSA**

- Working on migrating and integrating Database from SQL server into graph database using GraphQL API.
- understand the data requirements of the organization and design appropriate data models. This includes creating logical and physical data models, defining data structures, and ensuring data integrity, consistency, and performance. Using This tool like ER/Studio tools.
- Design data modeling to graph database to ingest data from SQL server database.
- Working on Nifi tool to migrate database from SQL server to graph database (Arrangodb).
- Integrate Multi data source from different systems on Kafka and ingest data into Arrangodb.
- implement data quality checks and data governance processes to ensure that data is accurate, consistent, and compliant with relevant regulations. This includes profiling data, identifying data quality issues, and implementing data cleansing or enrichment processes using Python scripts.
- scheduling and monitoring data integration workflows, ensuring data is processed in a timely manner, and troubleshooting any issues that arise working with Apache Airflow tools.
- Dockerize all services in containers to facilitate on devops team in production.

#### **Advanced Analytics**

- Start by understanding the specific requirements of the big data infrastructure, including the volume and variety of data, the desired data retrieval and reporting capabilities, and any performance or scalability considerations.
- Set up Kudu and Impala on This infrastructure. Kudu is a columnar storage engine, while Impala provides SQL-like querying capabilities. Configure them to work together seamlessly.
- Write SQL scripts to create the necessary tables, columns, and indexes on Kudu. This will enable efficient querying and retrieval of data.
- Develop endpoints using the Django framework and GraphQL to interact with Kudu. This will allow you to retrieve and manipulate data through a unified interface.
- Create REST endpoints on top of Kudu to retrieve data from the data warehouse schema.
- Define a data warehouse schema on Kudu that organizes the data for reporting purposes.

MAY 2022 – FEB 2023

Etisalat, Egypt — *Data Engineer*

#### **Cloudera Platform**

- Translate complex technical and functional requirements into detailed designs to start Development phase.
- Develop and implement data pipelines that extracts, transforms, and loads data from Staging area into Datawarehouse (Hive) Using spark tool on Cloudera Platform.
- Developing ETL (Extract, Transform, Load) processes to transform and cleanse the ingested data using Apache Spark, Apache Hive and Apache Pig to perform data manipulation, aggregation, and enrichment tasks.
- Working with distributed file systems (HDFS) to store and manage large volumes of data. This includes designing and implementing data partitioning, compression, and indexing strategies for efficient data storage and retrieval.
- working with the Nifi tool to handle the ingestion, storage, processing, and analysis of large datasets. This involves setting up data flows, configuring data processors, and ensuring efficient data movement and processing within the Nifi environment.
- developing visually appealing and intuitive dashboards using Tableau. These dashboards are designed to monitor the quality of data, providing meaningful insights and visualizations to stakeholders.
- Organizing the work alongside the team using version control.

AUG 2020 – JAN 2022

Corelia, France — *Data Engineer*

- Collected, analyzed, and interpreted raw data from various websites. Collaborated with the Operations and Technology Department on the development of new automated data management/analysis software which increased the overall productivity and cut unnecessary costs.
- Implemented Data pipeline to help data science teams build prediction model surge in customer orders to lower average.
- Conduct data mining, data modeling, statistical analysis and business intelligence Data analytics supports decisions for high-priority, enterprise initiatives involving IT/product development, customer service improvement, organizational realignment, and process reengineering.
- Built out the data and reporting infrastructure from the ground up using Data Studio and SQL to provide insights into the product and marketing funnels. Designed and implemented A/B experiments for products to improve the conversion rate.
- Filtering and Wrangling millions of textual data to deliver neat and meaningful Reports.
- Working on social media real data for many platforms (YouTube, Instagram, Change, Reddit, and Twi

## SKILLS

### Big Data:

- Apache Hadoop
- NIFI
- Apache Hive
- Apache Spark
- Apache HBase
- Kafka
- Airflow
- Cloudera

### BI Tools

- Tableau
- Data Analysis Using Excel
- Concepts and Analytical Tools
- Business Statistics
- Data Mining methods.

### Data Warehousing and ETL

- Data Warehousing
- Snowflake Tool

### Databases

- Oracle SQL and PL/SQL
- SQL (Access, Server, and My
- Mongo
- Graph Database
- Teradata

### Administration

- Red Hat System Admin I
- Kubernetes
- Shell scripting

### Programming Languages

- Python - Java SE - C- Scala - React, Nodejs (Express)

## CERTIFICATIONS

- Feb 2019 – Apr 2020 “Data Architect Nanodegree from Udacity”
- Oct 2020 – Dec 2020 “advanced data analytics nanodegree from egfwd”
- Jan 2021 – May 2021 “AWS machine learning specialist from MCIT”
- Aug 2021 – Dec 2021 “Data Engineer Nanodegree from Udacity”
- Apr 2021 – Jul 2021 “Udacity Data Streaming Nanodegree”

## language

- Arabic: Mother tongue
- English: Very Good