Mohamed Mahmoud Ali

Big Data Engineer

✉ mmahmoudbde@outlook.com   📞 +966596270667   📞 +201026180568   📍 Cairo, Egypt   in LinkedIn   ⌗ GitHub

PROFESSIONAL SUMMARY

Big Data & AI Engineer with 2+ years of experience architecting scalable data pipelines and deploying machine learning models across distributed systems. Specialized in building production-grade solutions using Hadoop ecosystem, PySpark, and cloud platforms. Proven expertise in processing terabyte-scale datasets, implementing real-time streaming architectures with Apache NiFi, and developing AI-driven analytics solutions. Currently pursuing Master's in Computer Science at Cairo University with focus on advanced machine learning and Data Engineering.

PROFESSIONAL EXPERIENCE

Big Data Engineer, Fingerprint Consultancy   05/2023 – Present
Cairo, Egypt

- Designed, automated, and optimized data pipelines processing **over 5TB of data daily** using Spark, Hadoop, Cloudera, NiFi, and Impala.
- Managed and configured multi-node Cloudera clusters supporting **50+ production jobs** with 99.9% uptime.
- Developed and maintained ETL/ELT workflows using Informatica and ODI, handling **100+ daily pipelines** across multiple business domains.
- Built full end-to-end data warehouse pipelines that improved reporting speed by **up to 40%** and reduced manual interventions by **70%**.

EDUCATION

Master of Science, Computer Science   08/2023 – Present
Cairo University, Faculty of Computers and Artificial Intelligence
Cairo, Egypt

- Focus: Advanced Machine Learning, Data Engineering, Big Data Analytics

Bachelor of Science, Bioinformatics and Computer Science   08/2018 – 08/2022
Cairo University, Faculty of Computers and Artificial Intelligence   GPA 3.4
Cairo, Egypt

- Graduation project: Predicting Protein Protein interactions using Deep Learning (A+)

PROJECTS

RAID Migration & Big Data Pipeline Project – Vodafone Egypt

- Migrated RAID data streams using PySpark 100+ Streams, processing terabytes of data every hour with high reliability and performance.
- Built 60+ automated ingestion pipelines to extract large-scale data from HDFS and deliver it to transformation workflows.
- Implementing PySpark transformations to handle massive distributed workloads with low latency 5 minutes per stream.
- Loaded transformed data into Kudu schema and partitioning for fast analytics.

Alinma Cross-Selling Project – Riyadh, KSA

- Developed and maintained end-to-end ETL workflows using Informatica for cross-selling analytics.
- Implemented data validation, error handling, and performance tuning to optimize large-scale ETL jobs 60+ for 3 business fields.

Naif College for National Security – Riyadh, KSA

- Built a full end-to-end data pipeline and data warehouse solution using ODI for ETL processes 150+ that runs in only 15 minutes.
- Integrated data from multiple sources (3 Sources) and ensured data quality, consistency, and reliability.
- Developed 15+ interactive dashboards and reports using Tableau to support decision-making and analytics.
- Delivered a scalable and maintainable solution that enables advanced reporting and insights for college operations.

Technical Skills

- **Big Data & Analytics:** Hadoop, Spark, PySpark, Hive, Impala, HBase, Kafka, NiFi
- **Cloud & Platforms:** Cloudera CDH, AWS (S3, EMR, Glue), Azure, GCP
- **Programming & Scripting:** Python, SQL, C++, Java, Shell,JS
- **Data Engineering:** ETL Development (Informatica, ODI), Data Warehousing, Data Modeling, Pipeline Orchestration
- **Databases:** PostgreSQL, MySQL, Oracle, Microsoft SQL Server
- **Visualization & BI:** Tableau, Power BI, Matplotlib, Seaborn, Plotly
- **AI/ML:** TensorFlow, PyTorch, Scikit-learn, Keras, XGBoost, Pandas, NumPy
- **DevOps & Tools:** Docker, Git, Airflow, Jupyter, Linux

LANGUAGES

- **Arabic:** Native
- **English:** Intermediate (read, write, and communicate effectively)