

## MD SALEH

Bronx, New York 10458 • +1-347-493-7141 • [md.saleh8323@gmail.com](mailto:md.saleh8323@gmail.com)  
<https://www.linkedin.com/in/mdsaleh8323/>

### DATA ENGINEER

Highly motivated Data Engineer with nearly 3 years of experience building and maintaining data pipelines across the finance and healthcare sectors with a special focus on transactional data. Expertise in ETL, data modeling, data warehousing, and Big Data tools such as Apache Spark, Kafka, Airflow, and Hive and proficiency in SQL, Python, and AWS cloud services. Demonstrated leadership in data warehouse maintenance, ETL development, reporting, and the seamless execution of data migrations.

---

ETL Development • Data Warehousing • Data Modeling • Data Pipeline • Data Ingestion and Processing  
• Object Oriented Programming • Data Visualization and Reporting • Data Analysis

**Languages:** Python, SQL

**Databases:** MS SQL Server, MySQL

**Big Data:** Spark, PySpark, Hive, Kafka, NiFi, Airflow, Delta Lake, Amazon Redshift

**Web:** JavaScript, HTML, CSS

**Web Services:** REST, SOAP

**Frameworks:** Django, Django Rest Framework

**SDLC:** Agile / Scrum

**Tools:** SSMS, SSRS, Pervasive (ETL), Pentaho (ETL), Power BI, Tableau, Github, Jira, Postman, BeautifulSoup, Pycharm, Jupyter Notebook, Github Actions (CI/CD), Docker, Power Apps, Power Automate, SharePoint

**Platforms:** AWS (EC2, S3, RDS, DMS, Lambda, Redshift), Azure Cloud Services, Linux, Windows

**Application/Software:** FICO Debt Manager

---

### PROFESSIONAL EXPERIENCE

**CAREER NOTE:** Completed on-campus studies and currently taking distance education courses to complete a **Master's Degree in Computer Science** (Available for full-time, W-2 employment).

#### **SAVE THE CHILDREN IN BANGLADESH, Dhaka, Bangladesh • 11/2021 – 06/2023**

International non-profit organization focused on child welfare.

##### **Data Analyst (Data Engineer)**

Developed and maintained a data lakehouse using Apache NiFi, Spark, and Kafka, and crafted Power BI dashboards for improved data analysis.

- Developed and maintained a streamlined data lakehouse infrastructure utilizing Apache NiFi, Spark, Kafka, and Delta Lake, resulting in an improvement in data ingestion and storage efficiency.
- Implemented PySpark applications, ETL pipelines, and AWS cloud services to enable efficient data processing and storage in Delta Lake, thereby enhancing analysis capabilities and achieving 30% increase in productivity and data analysis efficiency.
- Designed Power BI dashboards to visualize data, thereby enhancing accessibility for data-driven decision-making, resulting in an increase in efficiency and informed choices.
- Spearheaded the development of Power Apps and Flow using Microsoft Power Platform to collect and automate surveys, reducing the complexity of manual data collection and migration by 100%.
- Leveraged Python for custom data cleaning and processing, resulting in a significant improvement in accuracy across healthcare, child protection, and education data within the central Data Warehouse, achieving a 20% increase in data integrity.
- Developed RESTful APIs using Python and Django Rest Framework, contributing to seamless data management, and achieving a 40% improvement in data processing efficiency.
- Collaborated with clients to translate their business needs into actionable data requirements, resulting in a 30% decrease in development rework and a 15% faster time-to-insight.

**Technologies Used:** Python, Apache NiFi, Spark, Kafka, Delta Lake, Airflow, Power BI, Power Apps, MS SQL Server, MySQL, Django, Django Rest Framework

**NASSPROTECH, Dhaka, Bangladesh • 12/2020 – 10/2021**

USA based management consulting company.

### **Data Engineer**

Developed ETL processes, designed and developed data flow and pipelines, and developed stored procedures for data processing and reporting.

- Orchestrated the design, automation, and support of intricate ETL processes, employing Pentaho, Pervasive ETL tools, SSRS, SQL Server, and Java, resulting in efficient data collection and storage from diverse sources.
- Constructed dashboards and reports using Power BI and SSRS, providing valuable insights for business intelligence and informed decision-making.
- Led the development of Extract, Transform, Load (ETL) processes and maintained a data warehouse, resulting in automated pipelines that facilitated seamless data movement from source systems.
- Maintained ETL jobs, pipelines, and batch processing using FICO Debt Manager, overseeing scheduling and monitoring tasks to ensure smooth operation.
- Addressed ticket-based production issues related to data warehouse and ETL, ensuring continuous operational efficiency by 100%.

**Technologies Used:** SQL Server, SSRS, Pervasive, Pentaho, Power BI, VBA, Fico Debt Manager, Java

---

## **ACADEMIC PROJECT**

**Maharishi International University (2024) [GitHub Public Repository Analyzer](#):** Built a real-time data processing pipeline as the final project in the Big Data Technology course during my master's program. Using Apache Kafka and Spark, this project efficiently processes data from GitHub public APIs, offering insights on open-source projects by different organizations. The result provides a holistic view of open-source GitHub repositories, organizational contributions, and trending programming languages. Utilized Java, Hadoop, Hive, Spark, Kafka, Airflow, Tableau.

**Maharishi International University (2024) [Attendance Management System](#):** Developed a web-based attendance management system using Java Spring Boot and MySQL as the as the final project of Enterprise Architecture course during my master's program. Also, I developed the CI/CD pipeline using GitHub action and docker to ensure the continuous deployment of the application to AWS EC2 instance.

**Maharishi International University (2023) [Migration of Auction Management Data to AWS](#):** Leveraged AWS cloud services (RDS, DMS, S3, Lambda, EC2, SQS, and SNS) to automate the migration of auction management data and generate reports on any data changes within the AWS RDS database. This final project for the Cloud Computing course streamlined data transfer and reporting processes, improving overall data accessibility.

## **PERSONAL PROJECT**

**[Multi-vendor E-Commerce 2021](#):** Developed a comprehensive E-commerce web application with a vendor portal, admin portal, and customer portal. Core features include product management, efficient order processing, secure payment gateways, user authentication, real-time inventory tracking, a review and rating system, dynamic discounts and promotions, dynamic content management, and responsive UI design.

Utilized Python, Django, DRF, JWT, PostgreSQL, HTML, JavaScript, jQuery, and CentOS Server.

## **EDUCATION**

**Master of Science in Computer Science** (Major in Data Science)

*(In progress via distance education; expected completion 04/2026)*

Maharishi International University, Fairfield, Iowa

**Key Courses:** Big Data Technology, Enterprise Architecture, Algorithms, Cloud Computing

**Bachelor of Science in Software Engineering**  
Daffodil International University, Dhaka, Bangladesh (2020)