

Masood Salman Choudhury

📍 Manchester, United Kingdom ✉ MSalman5230@gmail.com ☎ +44 7552245629 📅 02/10/1996

🌐 [linkedin.com/in/MSalman5230](https://www.linkedin.com/in/MSalman5230) 🐙 github.com/MSalman5230 🔗 <https://msalman.de>

PERSONAL BIO

Senior Data Engineer & AI Solutions Architect with 5+ years' experience delivering **end-to-end data platforms** and **intelligent applications** across fintech, SaaS, and industrial analytics. Expert in designing **scalable data pipelines**, building **AI-powered systems with LLMs and machine learning models**, and deploying robust cloud-native solutions on **AWS, Azure, and GCP**. Proven track record in **leading cross-functional teams**, mentoring engineers, and driving **CI/CD best practices** to accelerate product delivery. Passionate about leveraging data and AI to solve complex business challenges and enable data-driven decision making.

WORK EXPERIENCE

Senior Data Engineer & AI Solutions Architect

04/2024 – present | Netherlands

Helixiora

Worked as a lead data engineer delivering **end-to-end solutions** for multiple **international clients** across fintech, enterprise SaaS, and industrial analytics domains. Responsible for **architecting** and **implementing data platforms, AI solutions, and cloud-based applications**. -

- Designed, built, and deployed **Azure** data pipelines in **Databricks (PySpark, SparkSQL)** to ingest large-scale structured and semi-structured datasets from **Blob Storage** and Azure **Cosmos DB**, execute complex transformations, and persist curated features in **Delta Lake** for downstream machine learning workflows.
- Led development of an enterprise-grade **Multi-Stage RAG system** for clients, combining **Python**-based data pipelines (Google Drive, Slack) with **LLMs, Langchain, Reranker**, and Pinecone **VectorDB** to deliver highly accurate, context-aware retrieval workflows.
- Designed, built, and deployed an **end-to-end Azure data pipeline** using App Functions and **CosmosDB**, applying **Kimball dimensional modeling** for **Power BI** datasets with **DAX**-based measures, enabling KPI reporting and forecasting.
- Led development of a **full-stack AI SaaS app (Android and iOS)** leveraging OpenAI Assistant API with a FastAPI backend, React Native (**Expo**) frontend, and **PostgreSQL** database, **including OAuth 2.0** authentication and integrated **Stripe payment** processing.
- Designed and implemented end-to-end **CI/CD pipelines** for fully automated deployment of containerized applications to **AWS, ECS** using **Docker**, Docker Compose, and infrastructure-as-code best practices (**Terraform**).
- Secured and deployed SaaS applications with **SSL**, reverse proxies, **Zero Trust** controls, and **firewall** rules, optimising load balancing for high availability.
- Set up **logging** and **monitoring** with **Prometheus** and **Grafana**, improving system observability, proactive issue detection, and **performance** tracking in production environments.
- Provided **technical mentorship** and conducted **code reviews** for junior and mid-level engineers, promoting best practices, improving code quality, and accelerating team growth.

Senior Data Engineer

02/2022 – 03/2024 | United Kingdom

Vaultoro

- Architected a **Kimball style star schema data warehouse** using **Elasticsearch** and **BigQuery**, enabling **real-time KPI dashboards in Kibana** and empowering data-driven decision-making for stakeholders.
- Led the design and implementation of multiple **scalable ETL pipelines** with Python, **GCP Dataflow**, Scrapy, and managed workflow orchestration with Apache **Airflow**, processing over **10 million financial data rows daily** for real-time analytics.
- Co-led the agile development of a secure, scalable Savings Platform using **FastAPI**, delivering the product in 3 months; **currently manages \$2M+ in monthly customer deposits**.

- Managed and optimized databases including **MongoDB, PostgreSQL, Elasticsearch, and BigQuery** by tuning queries, indexes, and partitions, resulting in significant performance improvements.
- Conducted deep data analysis with Pandas to detect anomalies and identify potential fraud patterns, enhancing platform security.
- Automated data validation workflows using Python scripts, ensuring pipeline integrity and achieving **100%** uptime for critical microservices.
- Deployed containerized applications with **Docker and Kubernetes**, ensuring high availability, scalability, and **streamlined CI/CD** operations.
- Developed and implemented a Random Forest classification model to identify high-value clients during signup, enabling personalized onboarding experiences.
- Configured and monitored **Google Analytics** and **Tag Manager dashboards** to track user behavior and support data-driven marketing strategies.

Data Analyst

06/2019 – 10/2020 | Guwahati, India

SoftCrop IT

- Delivered actionable insights via **Tableau dashboards** (**waterfall/cohort** analysis), improving stakeholder decision-making.
- Automated competitor **data scraping** (**Scrapy**) and ETL into **MySQL**, reducing manual effort by 50%.
- Analyzed sales and **geographic data** to guide strategic fibre network expansion.

PROJECTS

ETA Prediction for Container Ports using Machine Learning

09/2021 – 11/2021

Techniques at Port of Southampton

MSc Project at University Of Liverpool

- Predicted ship delays 10+ days in advance with 80% accuracy.
- Cleaned and engineered features using **Pandas, NumPy, and Pearson correlation**.
- Evaluated multiple ML models (SVM, DT, RF, NN) and selected Random Forest.
- Deployed the model with FastAPI for real-time predictions using docker.

Diabetes Prediction using XGboost

01/2022 – 01/2022

(Personal Project for self learning)

- **Exploratory Data Analysis (EDA)** with Seaborn, and **UMAP** for (Uniform Manifold Approximation and Projection) for visualizing high-dimensional data in a lower-dimensional space
- Tuned XGboost **hyperparameters** with **Optuna** and **ML-Flow**
- Deployed using Streamlit framework and docker

EDUCATION

MSc Business Analytics and Big Data

11/2020 – 02/2022

University of Liverpool

Liverpool, United Kingdom

Key Modules: •**Data Mining & Machine Learning** •Big Data Analytics •Enterprise Systems with SAP •Digital Business Technology and Management •Digital Strategy

Result: **Distinction**

Bachelor of Business Administration

08/2016 – 06/2019 | Guwahati, India

Asian Institute of Management and Technology

Key Modules: Statistics, Mathematics, Production & Operation Management.

Result: First-Class

CERTIFICATES

IBM Data Science Professional Certificate

Microsoft Certified: Azure Fundamentals

AZ-900

Microsoft Certified: Azure AI Fundamentals

AI-900