

# Muhammad Usama Hameed

Principal Data Engineer | Data Architect | Azure & Microsoft Fabric Specialist

usamahameedx@gmail.com • linkedin.com/in/usamahameedx • usamahameedx.com

Principal Data Engineer and Data Architect with 7+ years of experience designing and delivering cloud-based data platforms across Azure. Specialist in Azure Data Engineering, Microsoft Fabric, Lakehouse architecture, ETL/ELT pipelines and analytics enablement.

## SKILLS

---

**Azure Cloud:** Microsoft Fabric, Azure Databricks, Azure Data Factory, Synapse Analytics, Delta Lake, ADLS Gen2 / Blob Storage, Microsoft Purview, Azure DevOps

**Data Engineering:** Lakehouse Architecture, Medallion Architecture, ETL / ELT Pipelines, Batch & Streaming Processing, Data Warehousing, Data Governance, Talend

**Big Data & Orchestration:** Apache Spark (PySpark), Apache Airflow, Apache Kafka, Apache NiFi

**AWS:** AWS Glue, Amazon S3, Amazon QuickSight

**Analytics & BI:** Power BI, Tableau, Apache Superset, Metabase

**Programming:** Python, SQL, Shell Scripting, Linux

**AI & Data Science:** Machine Learning, LLM Models, NLP & Text Analysis, EDA, Pandas, NumPy, Seaborn

**DevOps:** CI/CD, Docker, Git, GitHub Actions, Infrastructure as Code

## WORK EXPERIENCE

---

### Principal Data Engineer · NorthBay Solutions

Feb 2024 – Present

Remote (Client: USA)

- Architected and delivered scalable, cloud-native data platforms using Microsoft Fabric, Azure Data Factory, Azure Databricks, Delta Lake, and Azure Data Lake Storage Gen2 to support enterprise analytics and AI workloads.
- Designed and implemented modern Lakehouse and Medallion Architecture (Bronze, Silver, Gold) within Microsoft Fabric, improving data reliability, performance, and consumption across analytics teams.
- Built and optimized end-to-end ETL/ELT pipelines using Azure Data Factory, Fabric Data Pipelines, Spark (PySpark), and SQL, ensuring high throughput and low latency data processing.
- Led the adoption of Fabric OneLake as a unified storage layer, enabling seamless integration across data engineering, BI, and data science workloads.
- Implemented data governance, security, and monitoring frameworks using Microsoft Purview, Azure Monitor, and Log Analytics, ensuring compliance, lineage, and data quality standards.
- Collaborated with stakeholders, product owners, and analytics teams to translate business requirements into scalable Azure-based data solutions.
- Optimized performance and cost through partitioning, Delta Lake optimizations (Z-Order, OPTIMIZE), Spark tuning, and capacity management in Fabric.

- Enabled advanced analytics and BI by delivering curated datasets for Power BI, Fabric Warehouses, and Semantic Models.
- Led end-to-end project delivery, including architecture design, estimation, implementation, and deployment.
- Mentored and guided 5 junior data engineers, conducted code and architecture reviews, and established engineering best practices across teams.
- Worked with DevOps teams to implement CI/CD pipelines (Azure DevOps, GitHub Actions), Infrastructure as Code, and automated deployments for data platforms.

## **Associate Data Architect · Socrate Datai Ltd**

Jan 2023 – Feb 2024

Lahore, Pakistan

- Designed Azure data architecture by mapping source-to-target data flows, defining ingestion, storage, transformation, and consumption layers, which resulted in a more structured and scalable analytics platform.
- Improved ETL consistency by defining reusable ingestion and transformation patterns using Azure Data Factory and Talend, reducing variability across data pipelines.
- Enabled reliable analytics by designing analytics-ready data models (fact and dimension structures) optimized for Power BI, improving report performance and usability.
- Accelerated data ingestion by architecting centralized storage on Azure Data Lake Storage Gen2 and defining folder structures, partitioning strategies, and access patterns.
- Reduced processing complexity by defining Spark-based transformation standards using Azure Databricks (PySpark), Python, and SQL for batch processing workloads.
- Supported large-scale e-commerce analytics by designing scalable Azure architectures capable of handling high user and transaction volumes.
- Improved data quality by introducing validation rules and consistency checks within Azure pipelines and transformation layers.
- Simplified onboarding of new data sources by creating standardized integration templates and documented ingestion patterns.
- Improved deployment reliability by supporting container-based runtime patterns using Linux and Docker.
- Ensured architectural compliance by reviewing pipeline and data model designs against approved Azure architecture standards.
- Reduced long-term maintenance overhead by documenting data models, data flows, and architecture diagrams for engineering and BI teams.
- Strengthened business-engineering alignment by translating business requirements into clear Azure architecture designs and technical specifications.

## **Senior Data Engineer · Analytics Private Limited**

Nov 2020 – Dec 2022

Lahore, Pakistan

- Designed scalable AWS-based data architecture by defining ingestion, processing, and analytics layers capable of handling very large volumes of financial transaction data for banking clients.

- Built and maintained distributed data ingestion pipelines using Apache NiFi, Apache Kafka, AWS Glue, and Amazon S3, enabling reliable and continuous data flow.
- Implemented large-scale data processing using PySpark on AWS, optimizing batch transformations for downstream analytics and reporting.
- Improved data accuracy and consistency by translating stakeholder requirements into transformation logic implemented within AWS Glue and Spark jobs.
- Increased platform reliability by identifying and resolving pipeline failures, performance bottlenecks, and data quality issues, reducing operational disruptions.
- Enabled business reporting and operational transparency by delivering analytical datasets and dashboards using Amazon QuickSight.
- Enhanced analytics capabilities by writing optimized SQL queries and Python-based transformations, supporting deeper financial and operational insights.
- Reduced ETL errors and reprocessing effort by implementing validation, logging, and monitoring mechanisms across AWS data pipelines.
- Collaborated with stakeholders, analysts, and engineering teams to align AWS data solutions with financial reporting and regulatory needs.

## **Data Engineer · Omdena**

Sep 2018 – Oct 2020

Islamabad, Pakistan

- Designed and implemented end-to-end data pipelines using Python, PySpark, and SQL, enabling scalable ingestion, transformation, and analytics for social-impact projects.
- Orchestrated batch data workflows using Apache Airflow, improving pipeline reliability, scheduling, and observability.
- Built large-scale data processing pipelines with PySpark, enabling efficient handling of structured and unstructured datasets.
- Developed ETL workflows using Talend, standardizing data ingestion and transformation across multiple data sources.
- Implemented centralized data storage using open-source data formats (Parquet, CSV, JSON), improving data organization and accessibility.
- Automated data collection by developing Python-based scrapers and ingestion scripts, enabling timely ingestion of news and external datasets.
- Supported initiatives aligned with UN Sustainable Development Goal 16.2 by delivering structured datasets that improved access to justice and institutional transparency.
- Built data pipelines to detect and analyze patterns of online violence against children, enabling proactive identification of harmful activities.
- Supported the development of an educational chatbot for children by preparing clean, labeled datasets and feature-ready data for downstream models.
- Enabled analytics and reporting by publishing curated datasets to Apache Superset, supporting exploratory analysis and insights for stakeholders.

## CERTIFICATIONS

---

**Fabric Analytics Engineer Associate** — Microsoft

**Certified Data Engineer** — IBM

**Google Business Intelligence Certification** — Google

**Certified Apache Airflow Developer** — Astronomer

**Certified Matillion Practitioner** — Matillion

**Certified Python Programmer** — University of Michigan

## EDUCATION

---

**BS Software Engineering** — COMSATS University Islamabad

**Nanodegree in Artificial Intelligence & Data Science** — Udacity

**Skill & Career Fellowship (Stanford University Affiliated)** — Amal Academy

## VOLUNTEERING & LEADERSHIP

---

**Lead Data Educator · Data Seekho**

Jan 2025 – Present

- Founded and scaled a data education community impacting 1,000+ learners, delivering 50+ workshops, bootcamps, and webinars focused on data engineering, analytics, and AI.
- Designed industry-aligned curricula and mentored 200+ students, enabling hands-on project development and career readiness in data roles.

**Vice President · Roohnayiat Foundation**

- Led and coordinated community welfare initiatives benefiting 500+ individuals, overseeing volunteers, outreach programs, and resource planning.
- Introduced structured data collection and reporting to track program effectiveness and improve decision-making.