

ZAIN UL ABDEEN ALVI

Karachi, Pakistan | zainalvi552@gmail.com | +92 318 273 3998 | [LinkedIn](#) | [GitHub](#)

PROFESSIONAL SUMMARY

Data Engineering professional with hands-on experience architecting end-to-end streaming and batch pipelines using Apache Kafka, Databricks Delta Lake, AWS Kinesis, and Spark Structured Streaming. Formally trained in Machine Learning and Deep Learning through Samsung's AI Innovation Campus. Engineered production-grade ETL/ELT systems handling high-velocity data at scale. Seeking roles in Data Engineering, AI Engineering, or Data Analytics in Pakistan or remotely across Europe.

EXPERIENCE & WORK HISTORY

Samsung AI Innovation Campus × UIT Karachi | *AI Engineering Training* Jan 2026 – Apr 2026

- Designed and deployed 5+ end-to-end AI models using Python, scikit-learn, TensorFlow, and neural network architectures, achieving 92%+ accuracy on classification tasks.
- Applied statistical foundations including linear algebra, probability theory, and gradient-based optimization to solve real-world prediction problems.
- Delivered fully documented AI projects across design, implementation, evaluation, and results communication in competitive, nationally recognized program.

Core Data Engineers (Berlin, Germany) | *Data Engineering BootCamp* Aug 2025 – Dec 2025

- Completed rigorous production-grade bootcamp covering ETL/ELT architecture, data warehousing, and distributed cloud platforms aligned with European engineering standards.
- Architected and delivered 8+ end-to-end data projects using SQL, Python, Apache Airflow, and AWS within sprint-based agile cycles.
- Gained international engineering perspective by training alongside a Berlin-based cohort, replicating the workflow expectations of European data teams.

Saylani Mass IT Training | *Data Engineering Training* Nov 2024 – Jun 2025

- Built 3+ end-to-end ETL pipelines using Python and SQL, eliminating 40+ hours/month of manual data processing tasks.
- Implemented multi-step workflow orchestration using Apache Airflow and Prefect, deploying pipelines across AWS and GCP (BigQuery)
- Optimized schema design and query performance across Snowflake and Databricks data warehouses, achieving 35% faster analytical queries.

PROJECTS

Customer Complaint Management Pipeline:

- Engineered multi-layer data pipeline using medallion architecture (Bronze/Silver/Gold) to ingest, transform, and report customer complaint data at scale, reducing report generation time from 6 hours to 15 minutes.
- Implemented Delta Live Tables (DLT) on Databricks with AWS Kinesis as streaming source, reducing data availability latency from batch-level (8+ hours) to near real-time (< 2 minutes).

Technologies: AWS Kinesis · Databricks · Delta Lake · Delta Live Tables (DLT)

Telematics Data Lake Pipeline:

- Built real-time streaming pipeline ingesting 50+ MB/sec of high-velocity telematics data via AWS Kinesis into Databricks Delta Lakehouse with full medallion architecture, enabling real-time vehicle insurance tracking and analytics.
- Designed schema enforcement and data quality checks at each transformation layer, ensuring 99.8% data accuracy for downstream analytics consumers.

Technologies: AWS Kinesis · Databricks · Delta Live Tables · Delta Lake

Real-Time Data Engineering Pipeline

- Architected end-to-end streaming system capable of ingesting, transforming, and serving 100k+ events/second in real-time using Kafka producers/consumers and Spark Structured Streaming.
- Integrated MongoDB Atlas for flexible data storage and Elasticsearch + Kibana for real-time search indexing and interactive dashboards, enabling <500ms query latency.

Technologies: Apache Kafka · Spark Structured Streaming · MongoDB Atlas · Elasticsearch · Kibana

CORE SKILLS

Data Engineering & Pipelines: ETL/ELT · Apache Kafka · Apache Airflow · Prefect · Databricks · Delta Lake · Delta Live Tables · Spark Structured Streaming

Cloud Platforms: AWS (Kinesis, S3) · Google Cloud Platform (BigQuery) · Docker

Data Warehousing & Storage: Snowflake · Google BigQuery · Databricks Lakehouse · MongoDB Atlas · Elasticsearch

AI & Machine Learning: Supervised/Unsupervised Learning · Neural Networks · Deep Learning · scikit-learn · Python (pandas, NumPy)

Programming & Databases: Python · SQL (SQL Server, MySQL) · NoSQL · Git / GitHub

CERTIFICATIONS

- **Google Data Analytics Professional Certificate** — Google
- **Samsung AI Innovation Campus — Certificate in Artificial Intelligence** — Samsung Pakistan
- **Google Soft Skills Programme Certificate** — Google

EDUCATION

Sindh Madresatul Islam University

Feb 2023 – Dec 2026

Bachelor of Science in Computer Science

- Relevant Coursework: Data Engineering, Cloud Computing, Database Systems, Algorithms, Operating Systems