

Muhammad Ramzan Khan

📍 Lahore

✉ rn45819@gmail.com

☎ +923063961138

🐙 ramzan1813

🌐 ramzan1813

Profile

Data Engineer around **4 years of experience** across **Data Engineering**, automation, and software development. Skilled in designing and building scalable **ETL/ELT pipelines**, orchestrating workflows, and delivering cloud-based data solutions on **AWS**. Strong expertise in **Python, SQL, PostgreSQL**, and AWS services including **Lambda, S3, RDS, Glue, Athena, Redshift, Step Functions, EventBridge**, and **MWAA (Airflow)**. Hands-on experience with **PySpark, Databricks**, and **distributed data processing**, along with exposure to **Snowflake Data Warehousing, Cloud Analytics**, and **ELT workflows**. Proficient in **Docker, Pentaho, Flask**, and **Selenium**, with a proven ability to troubleshoot complex systems, optimize data pipelines, mentor team members, and deliver reliable data solutions that support analytics, reporting, and business decision-making.

Core Competencies

- **Cloud Computing:** AWS (EC2, S3, RDS, Lambda, API Gateway, Step Functions, EventBridge, CloudWatch, ECR, Glue, Athena, Redshift, Cognito), Serverless Architecture, Workflow Orchestration
- **Data Engineering:** ETL Pipelines, Data Integration, Data Transformation, Data Modeling, Data Warehousing, Redshift, Snowflake, Query Optimization
- **Programming & Scripting:** Python, SQL, PostgreSQL, JavaScript, Node.js, C#, Java, C++, MongoDB, MariaDB, Oracle
- **Tools & Technologies:** Docker, Pentaho, Flask, PySpark, Databricks, Selenium, Airflow
- **Analytics & AI/ML:** Data Visualization, Business Intelligence, Predictive Analytics, ML Models, YOLOv4 (academic project), GPT APIs, Google Vision API
- **Data Warehousing & DB:** Athena, PostgreSQL, MongoDB, MariaDB, Redshift, Data Warehousing, Relational Databases, Query Optimization
- **Automation & Workflow Optimization:** Workflow Automation, Serverless Computing, API Integration
- **CI/CD & DevOps:** Git, GitHub, Bitbucket, CI/CD, SAM CLI, Confluence, Jira

Work Experience

SystemsLtd

Apr-2026 to Present

Data Engineer (Rak Bank)

- Built and maintained scalable ETL pipelines using PySpark on the Databricks platform for large-scale batch data processing, focusing on distributed computing, performance optimization, and efficient transformation logic.
- Worked extensively with Databricks notebooks, Delta Lake architecture, and Spark SQL for developing data engineering workflows across medallion architecture layers (Bronze, Silver, Gold).
- Designed and managed workflow orchestration using scheduling and pipeline automation frameworks, handling task dependencies, retries, and incremental data processing in production environments.

- Implemented cloud-based data lake solutions using S3-compatible object storage, applying partitioning strategies, Parquet optimization, and Delta format capabilities to improve query performance and data reliability.

SystemsLtd

Nov-2025 to Mar-2026

Data Engineer (Regeneron)

- Engineered and maintained scalable AWS-based data pipelines, ensuring consistent data availability and reliability across environments.
- Diagnosed and resolved complex data pipeline issues in production, including optimizing S3 data flows, enhancing Airflow (MWAA) DAG performance, and addressing AWS service-level constraints within defined access boundaries.
- Managed end-to-end pipeline lifecycle activities, including controlled promotion of workflows from development to production and implementing enhancements to improve pipeline stability and performance.
- Collaborated with cross-functional teams to design and implement data pipeline improvements, introducing fixes and incremental features aligned with evolving business requirements.
- Worked with analytical data platforms such as Amazon Redshift and Snowflake-style architectures, performing in-depth data validation and query optimization using SQL to ensure data accuracy and integrity.

Kavtech Solutions (Pvt.) Limited

Jan-2024 to Nov-2025

AWS Cloud Data Engineer (ylopo)

- Designed and built **event-driven, serverless data pipelines** on AWS using Lambda, Step Functions, EventBridge, and API Gateway to automate data ingestion, transformation, and integration from diverse sources (including APIs and RETS).
- Implemented **secure, scalable architectures** leveraging Cognito for authentication, Secrets Manager for credential management, and Amazon RDS/S3 for data storage, ensuring compliance and reliability in production workflows.
- Optimized ETL pipelines by **automating repetitive workflows**, reducing processing time and manual intervention while improving pipeline stability.
- Deployed and supported **new data feed implementations**, ensuring seamless integration and alignment with evolving client requirements.
- **Trained and mentored team members** on AWS ETL development, troubleshooting, and cloud best practices, enabling faster onboarding and improved team productivity.
- Partnered with cross-functional teams to **diagnose and resolve production challenges**, ensuring uninterrupted operation of mission-critical data processes.
- Continuously evaluated and integrated **AWS cloud-native services** to improve automation, scalability, and reliability of data workflows.

AlphaSol

Sep-2022 to Dec-2023

Python Engineer

- Developed automation scripts in Python (and Google Apps Script), streamlining repetitive workflows like file conversions and data transformations, saving 30+ hours/month.

- Designed and implemented web scraping pipelines (Selenium, Python) to extract business data from sources like Google Maps and Zillow, with outputs stored in MongoDB or exported to CSV for analytics.
- Built data analytics workflows, applying data cleaning and visualization (Matplotlib) to generate insights for client datasets.
- Integrated and fine-tuned GPT APIs for custom automation, including structured document generation from unstructured text, improving efficiency in client reporting processes.
- Experimented with machine learning models (scikit-learn), including a salary prediction model, and deployed prototypes using Flask and Streamlit.
- Designed a proof-of-concept automation integrating MongoDB and Selenium to track and sync YouTube view records with database systems.
- Contributed to AI/ML adoption within the company by combining automation, data pipelines, and GPT-based solutions into real-world workflows.

Achievements

- **Optimized Data Engineering Pipelines:** Spearheaded the redesign of ETL workflows, reducing data processing time and increasing reliability across multiple business-critical systems.
- **Cloud-Native Architecture:** Designed and deployed **serverless, event-driven solutions** using AWS Lambda, Step Functions, and EventBridge — enabling highly scalable and automated pipelines that minimize manual interventions.
- **Data Automation & Efficiency:** Automated repetitive data ingestion and transformation tasks, which freed up engineering hours and allowed the team to focus on higher-value strategic initiatives.
- **Production Reliability:** Diagnosed and resolved critical production bottlenecks in real-time data pipelines, ensuring **99.9% uptime** and uninterrupted delivery of business insights.
- **Advanced Analytics Enablement:** Integrated **AI/ML-driven models** into workflows (e.g., computer vision, object detection), contributing to innovative data solutions that improved business decision-making.
- **Data Quality & Governance:** Enhanced **data cleaning and transformation processes**, ensuring accuracy, consistency, and compliance for analytics and reporting teams.
- **Business Intelligence Impact:** Delivered actionable **BI dashboards and reporting solutions**, providing stakeholders with real-time, data-driven insights to guide strategic decisions.
- **Mentorship & Team Growth:** Trained and mentored junior engineers on ETL best practices, troubleshooting, and AWS data services, fostering a culture of knowledge-sharing and continuous improvement.
- **Recognition & Delivery:** Consistently delivered **high-impact results under tight deadlines**, receiving recognition from stakeholders for reliability, innovation, and execution excellence.

Education

PMAS-Arid Agriculture University

Oct-2018 to Aug-2022

BS Information Technology

- Built a strong technical foundation in **programming, data structures, algorithms, and database design**, with hands-on exposure to multiple languages including **C++, C#, Java, Python, and JavaScript**.
- Gained practical experience with **relational databases (Oracle, MariaDB)** through coursework and projects, learning schema design, normalization, query optimization.
- Projects: Student Record Management (C# + DB), AI-based Number Plate Recognition (YOLOv4 + Flask + MariaDB).
- Experimented with **Google Vision API** and other APIs to improve recognition accuracy and pipeline efficiency.

Certificate

- Data Science Foundations
- Data Visualization with Power BI
- AI & ML Fundamentals
- Microsoft Fabric (Dp-700)