# Muhammad Abubakar

📍 Lahore, Pakistan ✉ abubakar.dataengineer@gmail.com 📞 +92 335 4631993
💼 linkedin.com/in/muhammadabubakar12 🌐 dataengineerabubakar.vercel.app

*Open to relocation and remote opportunities*

## Professional Summary

Innovative and results-driven **Data Engineer** with over **4 years of experience** in designing, building, and optimizing **data pipelines, ELT frameworks, and analytics platforms** on **AWS, Databricks, and PySpark**. Proven expertise in developing scalable, cost-effective, and compliant cloud-based data ecosystems supporting real-time analytics and business intelligence.

Proficient in **Python, SQL, PL/SQL, and Shell scripting**, with a strong foundation in **data modeling, automation, and governance**. Experienced in transforming complex business data into actionable insights across **finance, healthcare, and enterprise** domains. Recognized for driving efficiency through automation, cloud cost optimization, and robust data architectures.

**Currently seeking Data Engineering or Cloud Data roles in Dubai or remote settings**, where I can leverage my technical depth and cloud expertise to deliver data-driven innovation and business impact.

## Core Competencies

**Cloud:** AWS (EMR, Glue, Athena, Step Functions, S3, Redshift, RDS, DynamoDB, CloudWatch, CLI)
**Big Data:** PySpark, Databricks, Delta Lake, Medallion Architecture, Oracle Data Integrator (ODI 12c)
**Databases:** PostgreSQL, Oracle, MySQL, DynamoDB, MongoDB
**Programming:** Python, SQL, PL/SQL, Shell, Java, JavaScript
**Data Governance:** Security, Encryption, Compliance (FCCM, TBAML, KYC, HIPAA)
**Tools:** Stonebranch, Postman, Toad, WebLogic, WebSphere, Visual Studio Code

## Professional Experience

**Optrix Technologies** *May 2025 – Present*
*Senior Software Engineer – Data Engineering*

- Architected and deployed scalable **ETL pipelines** using **AWS S3, Databricks, and PySpark**, cutting Databricks runtime costs by **25%** through Delta Lake optimization.
- Built and maintained a **multi-layered AWS data warehouse** integrating structured (PostgreSQL) and unstructured (DynamoDB) datasets for advanced analytics.
- Automated ingestion, transformation, and workflow orchestration using **Shell scripting** and **Stonebranch**, increasing operational throughput by **30%**.
- Implemented **data quality monitoring**, error handling, and lineage documentation to ensure compliance and traceability.
- Collaborated with analytics teams to enable real-time dashboards and predictive insights for business stakeholders.

**Techlogix** *Dec 2021 – May 2025*
*Senior Data Engineer – Cloud & Big Data*

- Designed and optimized **petabyte-scale data pipelines** leveraging AWS EMR, Glue, and Athena, improving performance by **40%** and reducing compute cost by **25%**.
- Implemented **Medallion Architecture** (Bronze–Silver–Gold) on AWS S3 using Delta Lake and Glue Catalog for self-service analytics.
- Automated 30+ data workflows with **Step Functions, Lambda, and Stonebranch**, improving delivery timelines by **40%**.
- Ensured **HIPAA and AML compliance** via IAM policies, KMS encryption, and secure access management.
- Designed ODI 12c mappings for **FCCM TBAML and KYC frameworks**, ensuring accurate ingestion and regulatory compliance.

*Key Projects:*

- **AWS Data Pipeline Optimization:** Enhanced PySpark-based pipelines for financial data, achieving 40% faster processing and 25% cost savings.
- **Cloud Data Warehouse for BCBS:** Built Delta Lake data lakehouse on AWS S3 for 50M+ daily records with Glue and Step Functions.

- **FCCM Customization (AML/TBML/KYC):** Automated ODI-based ingestion workflows, improving data quality and audit readiness.

**Hypoters (Remote)** *Oct 2020 – Nov 2021*

*Software Engineer – Backend Development*

- Developed and deployed RESTful APIs using Node.js and MongoDB, supporting high-availability web applications.

- Tuned and optimized MySQL schemas for scalable, low-latency database performance in production.

## Education

**Bachelor of Science in Computer Science (BSCS)**

National University of Sciences and Technology (NUST), Pakistan *2017 – 2021*

## Achievements

- **Achiever of the Month – Techlogix** (Sep 2022, Jun 2023, Jul 2024) for excellence in cloud automation and data optimization.

- Reduced AWS Databricks compute expenses by 25% through Delta Lake and PySpark performance tuning.

- Automated 30+ production workflows, improving operational efficiency and data reliability by 40%.

## Professional Interests

Data Engineering | Cloud Data Platforms | Big Data Analytics | ETL Automation | DataOps | AWS Solutions | Python | PySpark | Machine Learning Integration