JOYEB MULLA
Lead Data Engineer | Apache Spark | Azure | AWS | Databricks | Python | Fabric
Phone: +91 9890741250 / +44 7774448015 | Email: Mullajoyeb786@gmail.com
LinkedIn: https://www.linkedin.com/in/joyeb-m-04b94158
Location: India | UK On-site (Nov 2022–Oct 2024) | Seeking UK & Europe Opportunities

---

PROFILE

Lead Data Engineer (Senior Consultant level) with over 10 years of experience designing and delivering enterprise-scale analytics and data platforms within Financial Services and Banking domains. Expert in designing high-performance pipelines using Azure (Fabric, ADF, Synapse), AWS, and Databricks. Deeply skilled in Apache Spark optimisation, Python, and distributed computing. At KPMG UK, led engineering of the KPMG Insight Platform (KIP) and Fabric Analytics, building foundational data layers that support advanced analytics and financial modelling. Proven track record of collaborating with data scientists and product managers to optimise workflows for speed, security, and reliability. Passionate about mentoring engineers, driving best practices (CI/CD, Agile), and delivering value through large-scale data warehousing solutions.

CORE SKILLS

Cloud & Big Data: Azure (Fabric, ADF, Synapse, Data Lake Gen2, SQL DB), AWS (S3, Glue, Redshift, Lambda, Athena), Snowflake, Databricks, PySpark
Languages & Orchestration: Python (Strong), SQL, T-SQL, Airflow, ADF, Azure Logic Apps, YAML
Architecture & Optimisation: Distributed Computing, Parquet/Iceberg Optimisation, Data Modelling (Star/Snowflake), ML Pipeline Support, CI/CD, Git
Domain Expertise: Payments & Finance, P&L, General Ledger, Risk & Compliance, Retail, Industrial

---

PROFESSIONAL EXPERIENCE

KPMG India (KGS) – Lead Data Engineer, Advisory Tech UK | Dec 2020–Present
On-site: KPMG UK (Nov 2022–Oct 2024)
• Provided technical leadership to agile teams of 10+ engineers, designing high-performance data pipelines and tooling to support enterprise-scale financial analysis and reporting.
• Architected and optimised complex Apache Spark (PySpark/Spark SQL) workflows on Databricks, improving pipeline throughput and query performance by ~30% for large-scale financial datasets.
• Collaborated with cross-functional stakeholders, including Data Scientists and Product Managers, to build robust production data tooling that accelerates advanced analytics and modelling capabilities.
• Enforced strict engineering standards for security, reliability, and GDPR compliance across multi-region data systems, ensuring alignment with enterprise regulatory requirements.
• Designed scalable ETL/ELT frameworks using ADF, Databricks, and dbt, creating reusable patterns for ingesting and transforming high-volume transactional data.
• **Built AWS-based ETL pipelines implementing Slowly Changing Dimensions (SCD Type 1 and Type 2) using S3, AWS Glue, Redshift, and Spark to support historical data versioning and dimensional models.**
• Led the migration of critical financial workloads (P&L, Balance Sheet, GL) to Microsoft Fabric, modernising legacy infrastructure into a unified, secure cloud analytics platform.
• Mentored junior engineers and established software engineering best practices (Clean Code, Testing, CI/CD) to drive continuous improvement.

Nihilent Technologies – Senior Data Engineer | Dec 2018 – Dec 2020

• Designed and delivered Azure Data Factory and Databricks pipelines ingesting and transforming data into Azure Data Lake architectures.
• Built incremental and scalable ETL frameworks supporting finance and retail analytics use cases.
• Developed and optimised Power BI dashboards with Row-Level Security (RLS), incremental refresh, and performance-tuned semantic models.
• Improved data pipeline reliability, monitoring, and operational performance through proactive tuning and automation.

Wipro – Software Developer | Nov 2017 – Dec 2018

• Designed ADF and SSIS ingestion pipelines for industrial and energy datasets within enterprise environments.
• Developed Power BI reports using advanced DAX, secure workspace modelling, and governed dataset design.
• Supported deployment cycles, testing, and production enhancements in collaboration with UK-based teams.

KPIT Technologies – Software Developer | Nov 2014 – Oct 2017

• Managed SAP BusinessObjects and Power BI reporting estates across multiple business units.
• Designed governed dashboards and optimised semantic layers to support enterprise reporting needs.
• Supported production operations, release cycles, and quality improvements across BI platforms.

AWS PROJECT WORK (KPMG & Nihilent)

• Designed scalable S3 data lakes and orchestrated workflows using AWS Glue and Lambda.
• Optimised Athena and Redshift analytical queries using partitioned Parquet datasets to reduce cost and latency.
• Integrated hybrid cloud analytics pipelines connecting Azure and AWS ecosystems.

EDUCATION & CERTIFICATIONS

B.E. (Information Technology) – MIT College of Engineering, Pune University (2014)
Certifications: Microsoft Fabric Certified | Microsoft Azure DP-900 (Data Fundamentals)

ACHIEVEMENTS & LEADERSHIP

• Recognised by KPMG UK leadership for driving data modernisation and influencing engineering standards for the Insight Platform.
• Defined delivery standards and reusable PySpark/SQL frameworks adopted across multiple agile teams.
• Successfully optimised large-scale data processing jobs, significantly reducing runtime and operational costs.
• Created comprehensive monitoring and observability models, improving issue resolution speed by ~30%.