

Mushirul Aman

mushirul.aman@gmail.com / www.linkedin.com/in/mushirul-aman / +91 9162178254



Deloitte



PEPSICO

verizon



Summary

Experienced Lead Data Engineer (7+ years) specializing in **end-to-end data lifecycle management, real-time streaming, and cloud-native architectures** across **AWS, Azure, and GCP**. Proven track record in designing **fault-tolerant, high-throughput pipelines** using **Snowflake, Airflow, PySpark, DBT, Kafka, and Databricks**, ensuring **25% reduction in downtime** and faster business decisions. Delivered **enterprise-scale solutions** for **Chevron, Verizon, PepsiCo and Google**, including **20% improvement in demand forecasting accuracy, 60% reduction in manual reporting, and real-time anomaly detection** in FinTech data streams.

EDUCATION

VIT University, Vellore (2015 – 2019)

B.Tech, Computer Science (8 CGPA)

Vellore, India

2015 - 2019

WORK EXPERIENCE

Lead Data Engineer – Rsystems International, Bengaluru

Client -> Lendistry | Multinational Lending Business for Grants & Loans

06/2025 – Present

Data Vault 2.0, Data Discovery , Optimization, Streaming Events and Dimensional Modelling:

- Architected and implemented an enterprise-wide Data Vault 2.0 framework with DBT-driven transformations, dimensional modelling, and data discovery layers, increasing reporting accuracy by 30% and consolidating 120+ KPIs across finance, sales, and customer operations into a unified single source of truth, improving C-level decision confidence.

Senior Associate 2 – PWC AC, Bengaluru

Client -> Google | Multinational Corporation and Technology Company

09/2024 – 06/2025

Real-time Data Processing, Rollouts, Implementation, Streaming Events and Automation:

- Built real-time location intelligence pipelines for Google Maps using AWS Kinesis, Lambda, and Glue, handling 50M+ GPS events/day to power ETA accuracy and ranking model enrichment.
- Designed an AI-ready data lake on S3 + Redshift Spectrum for search query logs, enabling batch and streaming ML model training, reducing offline training time by 40%.
- Provisioned secure, scalable ML infrastructure on AWS SageMaker and EC2, supporting deep learning experimentation for natural language processing and map query understanding.
- Enabled fine-grained access control using IAM and Lake Formation, ensuring privacy-compliant data pipelines for LLM-based search enhancements.

Senior Associate 2 – PWC AC, Bengaluru | DOJ : 26th March 2024

Client -> PepsiCo Phase 2 and Phase 3 | Global Leader in Convenient Foods and Beverages

03/2024 – 09/2024

Data Pipeline Development, Cloud Integration, Analytics Automation:

- Delivered demand forecasting pipelines on AWS Glue and Redshift, integrating POS data and IoT signals from 5 global factories, improving forecast accuracy by 22%.
- Built ML feature pipelines using PySpark on Databricks for real-time shelf availability insights, contributing to dynamic pricing and replenishment models across 1M+ SKUs.
- Implemented CI/CD for ELT pipelines using GitHub Actions + Terraform, reducing deployment time by 70% and increasing pipeline reliability.
- Established S3-based zoned architecture (raw, staging, curated) with role-based access and lineage tracking, supporting both business reporting and AI model consumption.

Data Migration Specialist – Deloitte USI, Bengaluru

Client -> Chevron Release 2 | Oil & Gas American multinational energy corporation

08/2021 - 02/2024

Enhanced Data Delivery, ERP Data Integration, Streamlined Business Processes:

- Ingested IoT sensor data from 5000+ rigs into AWS Redshift using Kinesis and Glue, enabling predictive maintenance dashboards and reducing equipment failure risk by 35%.
- Implemented multi-region data governance using Lake Formation + IAM roles to ensure SOX and internal audit

compliance, with zero security escalations during audits.

- Optimized complex joins and transformations in PySpark on EMR, cutting batch processing time from 7 hours to under 2 hours for safety event data analytics.
- Established S3-based zoned architecture (raw, staging, curated) with role-based access and lineage tracking, supporting both business reporting and AI model consumption.

Client -> Chevron Release 1 | Oil & Gas American multinational energy corporation

DBT Transformation Frameworks & CI/CD Automation, Data Management and Data Lineage:

- Built scalable, cloud-native data platform to unify field operations data from Chevron's global oil fields for analytics and predictive maintenance using modular DBT models on top of Snowflake to enable layered, version-controlled transformations with CI/CD integration via Azure DevOps. This Boosted model accuracy by 30% through improved data consistency, lineage, and quality validations within DBT.
- Engineered parameterized ELT pipelines using Azure Data Factory for ingestion, Snowflake for storage, DBT for transformations, (snapshots, tests, incremental models) to ensure auditable, versioned, and traceable data models. This achieved 100% regulatory compliance traceability with time-travel-ready datasets and versioned DBT models.
- Enhanced retrieval and ranking models (BM25, learning-to-rank, semantic embeddings) to improve top-10 precision by 20% and CTR by 15%. Reduced zero-result queries by 30% through advanced analyzer/tokenizer configurations and multilingual synonym handling in Elasticsearch.

Business Technology Analyst – Deloitte USI, Hyderabad

08/2019 - 08/2021

Client -> Verizon Release 1 | Telecommunications Mass media Global Leader

Data Analysis, Data Visualization, ETL/ELT pipelines, Scalable Data Lake Engineering:

- Developed Real-Time Network Analytics Platform with real-time data ingestion pipelines using Apache Kafka, PySpark, and AWS Kinesis to process network logs and error events. Also Built streaming applications to detect anomalies and trigger alerts using sliding window aggregations resulting in a 30% reduction in downtime and enhanced operational visibility.
- Partnered with stakeholders to gather business requirements related to NPS, call center, and billing data to build Power BI dashboards for customer churn insights, integrating multiple data sources with row-level security enhancing data governance across business units.
- Designed and deployed distributed search pipelines using Apache Kafka, Flink, and Elasticsearch/OpenSearch, enabling near real-time indexing of 1B+ documents/day with <5s freshness SLA. Improved throughput from 10K QPS to 18K QPS and reduced query latency by 35% through sharding strategies, caching, and index tuning.

STRENGTHS

Data Engineering, ETL, ELT, Data Pipelines, Data Warehousing, Data Lake, Batch Processing, Stream Processing, Data Modeling, Data Integration, Data Transformation, Snowflake, BigQuery, Redshift, Azure Synapse, SQL Server, Cassandra, Apache Airflow, AWS Glue, Apache Kafka, Apache Spark, Python, PySpark, SQL, Data Governance, Data Quality, Cloud Computing, Cloud Data Platforms, Azure Data Factory, Google Cloud Dataflow, Databricks, DBT, Power BI, Tableau, Real-time Analytics, Data Security, CI/CD, Performance Tuning, Data Democratization, Schema Evolution, Data Lake , DLT , Flink , ML , DeepLearning, NLP , AI

TECHNICAL SKILLS

Data Management and Integration: Data Warehousing, Data Integration, Data Governance, Data Transformation, ETL/ELT
Data Analytics and Modelling: Data Modelling, Data Analytics, Data Profiling, Data Quality, Cloud-to-Cloud & On-prem Migration
Data Technical Proficiency: SAP BODS, Data Cleansing, Performance Optimization, Information Steward, Big Data
Project Management and Methodologies: Agile/Scrum, Planning, Integration, Risk Analysis, DevOps, JIRA

ACCOMPLISHMENTS

- **Spot Award for Excellence:** Recognized for exceptional performance by partnering with data scientists at Google and Verizon to create ML-ready datasets for LLMs, personalization engines, and fraud detection models; delivered a unified training-ready data zone using S3 + Glue + IAM.
- **Outstanding Award for Financial Data Pipelines Design:** Awarded for leading complex data loads in financial transactional data, conducting rigorous performance testing on ETL/ELT pipelines, and managing data volumes ranging from ~10M records.
- **Recognition for Handling Fresher training on Data Engineering Technologies:** Received multiple recognition letters from Client leadership for contributions to Verizon and Chevron projects, including the design and implementation of data pipelines, data integration, and training on data technologies over 5 years.

CERTIFICATION



Databricks Certified Data Engineer Professional, AWS Certified Solutions Architect – Associate