

EDUCATION

- **BTech Electrical Engineering**, Malviya National Institute of Technology, Jaipur 2025.

EXPERIENCE

ICICI Bank (July 2025 – Present) | Data Engineer, Mumbai

- Worked with **IBM InfoSphere DataStage and Designer** to schedule and monitor automated ETL jobs, ensuring reliable batch processing and timely data availability across enterprise systems.
- Built and enhanced **Azure ETL pipelines** using **Azure Data Factory** and **Azure Databricks** to ingest and transform source-system data, delivering analytics-ready datasets to business and risk teams.
- Implemented **PII and PCI data redaction** using **regular expressions** to identify and obscure sensitive fields across **on-premise Oracle and MySQL databases**, strengthening data security and compliance controls.

PROJECTS

Real-Time Quick-Commerce Operations Lakehouse [🔗](#)

- Built a real-time lakehouse ingesting **Apache Kafka** streams via **Spark Structured Streaming** into **Delta Lake** medallion tables with **exactly-once** semantics, watermarked late-event handling, and schema evolution, powering a sub-5-minute stockout detector.
- Engineered batch ETL in both **PySpark** and **Scala** on **Databricks** and **GCP Dataproc**, loading gold aggregates into **BigQuery** for a portable multi-cloud pipeline with identical idempotent behaviour across engines.
- Implemented in-stream data quality (**Great Expectations**, dead-letter routing), **Airflow** orchestration, and CDC-style Delta MERGE upserts with time travel, validated by **86 automated tests** in CI.

UPI Payments Compliance Lakehouse (AWS) [🔗](#)

- Built an **AWS** payments lakehouse processing **5M+ events** through an **S3** medallion architecture with **Lambda** schema validation and quarantine routing, intercepting **2.0% bad events** before downstream contamination.
- Engineered idempotent **PySpark** ETL on **Databricks** with PII tokenization, proving kill-and-rerun reproducibility by removing 77K+ duplicates with byte-identical row counts.
- Codified the platform in **Terraform** (parameterised dev/prod) with least-privilege **IAM**, **VPC** isolation, encryption, and **CI/CD** (plan-on-PR, approval-gated apply).

Parcel Intelligence Lakehouse on Azure [🔗](#)

- Built an **Azure Data Factory**-orchestrated logistics lakehouse transforming **200K+ parcels** through a **Databricks Delta** medallion into three serving stores: **Cosmos DB** point-reads, an **Azure SQL** SLA mart, and lakehouse analytics.
- Engineered event-time ordering repair, PII masking, and a cross-store control total that blocks publication on mismatch, catching a **6.2% delivered-count undercount** from out-of-order scans.
- Delivered secret-free, **Key Vault**-referenced pipelines validated by **60 automated tests** (real-Spark, ADF checks) in **GitHub Actions** CI.

SaaS Revenue Metrics Platform (Snowflake + dbt) [🔗](#)

- Built a **Snowflake** and **dbt** platform with star-schema dimensional modelling, **RBAC** role hierarchy, and layered staging/mart transformations on **500K+ records**, delivering a governed source of truth for ARR and churn.
- Implemented a **160+ test** data-quality framework (not-null, referential integrity, row-count reconciliation, metric tolerance) orchestrated with **Airflow**, catching a 2x ARR fan-out double-count in CI.
- Authored expert **SQL** (window-function MRR waterfalls, recursive CTEs, GROUPING SETS) and integrated **Snowflake Cortex** to auto-generate governed executive summaries.

TECHNICAL SKILLS

- **Programming:** Python, SQL, PySpark, Scala, Bash
- **Data Engineering:** ETL/ELT, Medallion Lakehouse Architecture, Dimensional Modeling (Star Schema), Data Warehousing, Data Quality & Testing (dbt, Great Expectations), Batch & Streaming, CDC, Idempotent Pipelines
- **Big Data & Orchestration:** Apache Spark (Structured Streaming), Delta Lake, Databricks, Apache Kafka, dbt, Apache Airflow, Snowflake
- **Cloud & IaC:** AWS (S3, Lambda, RDS, DynamoDB, IAM, VPC), Azure (Data Factory, Databricks, Azure SQL, Cosmos DB), GCP (BigQuery, Dataproc), Terraform, Docker, CI/CD (GitHub Actions)
- **Databases & Governance:** Snowflake, Azure SQL, PostgreSQL, Cosmos DB, Oracle, MySQL, RBAC, PII/PCI Redaction, MDM
- **Analytics & ML:** Power BI, Tableau, A/B Testing, XGBoost, Regression, RAG/LLMs

LEADERSHIP SKILLS

- **Executive, Think India MNIT Club:** Led the planning and execution of large-scale events with **5,000+ attendees**, managing and coordinating a team of **20+ students** to ensure smooth operations and delivery.
- **Founder, Unicorn Club (2022):** Built and scaled a community of entrepreneurs, organizing and hosting **30+ talks** focused on practical business insights, collaboration, and knowledge sharing.