ROHIT KUMAR | SENIOR DATA ENGINEER | 4+ YRS

Bengaluru, India | +91-7906616602 | r.kumar01@hotmail.com | LinkedIn | Data Portfolio | GitHub

Work Experience

Moody's Ratings – Data Engineer

Apr'25 – Present | Bengaluru

- Built and led deployments of the AIDE framework, ingesting **80+ feeds from 10+ heterogeneous sources** (JDBC, APIs, S3) with a scalable design, **cutting infrastructure costs by 60%** and **reducing deployment time by 75%**
- Engineered and operationalized data quality, governance using Great Expectations, **covering 95%+ of feeds across 100+ TB data**; enabled stakeholders with AWS CloudWatch dashboards for unified visibility and faster debugging & issue resolution
- Led development of an Apache Iceberg Data Lakehouse with ACID transactions and data lineage, **reducing data prep time by 50% for 30+ data scientists**; **built an MCP server** for fast data access and exploration
- **Improved DB indexing for** unstructured data on pgvector, enabling fast similarity search - **cut index build time by ~95%** for Agentic AI workloads
- Architected and shipped 2 Spring Boot/MongoDB APIs serving relevance-scored news & sentiment trends to an analyst-workbench & Agentic AI workflows, cutting manual triage 90% at 30K+ requests / day, p95 <20ms

Axis Bank – Data Engineer

Jul'22 – Apr'25 | Mumbai

- **Designed and implemented** batch data pipelines and a self-service analytics dashboard for 16 product lines, **partnering cross-functionally** to automate repeat complaint detection, saving 72 hours/month and reducing repeats by 40% by Q1 FY'25
- Led the Missing Income Nudge Project to improve data quality for RBI compliance, **resulting in a 52% reduction in missing income customers** through **targeted communication strategies for 14 personas** in collaboration with business teams

Projects

AURUM: Stock Research Data Platform (AWS, DBT, ECS Fargate, Terraform) [GitHub] [Demo]

- Developed a **configuration driven ingestion framework** loading Yahoo OHLCV and SEC EDGAR XBRL for S&P 500 companies into RDS Postgres; chunked writes and unique keys supported **26-year memory-safe** historical loads and **idempotent** reruns
- Modelled a **DBT medallion warehouse** (bronze → silver → gold; 20 models) powering a **2.9M-row ML feature mart + training pipeline**, deployed via Terraform on ECS Fargate and orchestrated by **Step Functions**

TransitFlow: Real-Time Event Stream (using Kafka & Spark) [GitHub]

- **Architected a real-time data streaming solution** using Kafka and Spark, processing transit data (vehicle, weather) and writing to AWS S3 for real-time analytics and reporting
- Enhanced data retrieval capabilities through streamlined ETL workflows using AWS Glue; integrated diverse datasets into Redshift, **improving report generation speeds by approximately 30%**
- Deployed Docker containers for both Spark and Kafka, enhancing the development environment by accelerating real-time data ingestion processes; **reduced initial setup time for new projects by 50%** through streamlined workflows

Achievements

Moody's IM'PACT Award – Lead with Curiosity (2x): *Sep'26*

- Recognized for leading delivery of 2 application-level APIs that enabled **CLEANews** integration with internal systems, beyond core Data Engineering scope
- Recognized for optimizing DB indexing for Agentic AI, reducing documents insert time from 2 days to ~90 mins

BIU – Platinum Learner Award (2x): *Mar'25*

- Acknowledged for inclusion in the list of top learners in FY'24 & FY'25 within the Business Intelligence Unit (BIU)

BIU – Star Award: *Nov'23*

- Awarded for automating repeat complaint detection, reducing bank-wide repeats as of Q1 FY'25

Education

National Institute of Technology — *Tiruchirappalli*

M. Tech - Energy Engineering — *Jun'22*

Skills

Data Engineering: Apache Spark (PySpark, Structured Streaming), Apache Kafka, Apache Airflow, dbt, JDBC, REST APIs

Lakehouse & Storage: Apache Iceberg, Delta Lake, Snowflake, Databricks, Amazon S3, Athena, PostgreSQL, MongoDB

Data Quality & AI: Great Expectations, CloudWatch, MCP servers, pgvector, RAG, Langchain, Langgraph

Cloud & DevOps: AWS (S3, Glue, EMR, ECR, Redshift, Athena, DynamoDB), GCP (BigQuery, GCS), Terraform, Docker, Kubernetes, GitHub Actions

Languages: Python, SQL, Java (Spring Boot), C++, SAS, Bash

Courses and Certifications

- Data Engineering Professional Certificate – DeepLearning.AI
- Build Data Lakes and Data Warehouses on Google Cloud – Google Cloud
- IBM Data Engineering Professional Certificate, IBM Data Warehousing Certificate – IBM, Coursera
- Machine Learning Specialization by Andrew Ng – DeepLearning.AI
- AWS Cloud Practitioner Essentials – Amazon Web Services