

ROHIT KUMAR

+91-7906616602 | r.kumar01@hotmail.com | [LinkedIn](#) | [GitHub](#)

Skills

Data Engineering & ML – PySpark, Airflow, Kafka, DBT, Apache Iceberg, Spark MLlib, LangChain, LangGraph, MCP

Big Data & Databases – Apache Spark, Hadoop, MySQL, PostgreSQL, Cassandra, MongoDB

DevOps & Cloud – Docker, Kubernetes, Terraform, AWS Cloud Services (Athena, Redshift, S3, Glue, CloudFormation, SageMaker AI, EMR, ECR)

Programming & Analytics – Python, SQL, C, SAS, Power BI, Tableau

Experience

Moody's Ratings – Data Engineer

Apr'25 – Present | Bengaluru

- Built and led deployments of the AIDE framework, ingesting **50+ feeds from 8+ heterogeneous sources** (JDBC, APIs, S3) with a scalable design, **cutting infrastructure costs by 60%** and **reducing deployment time by 75%**.
- Engineered and operationalized data governance using Great Expectations, **covering 90%+ of feeds across 10+ TB data**; enabled stakeholders with AWS CloudWatch dashboards for unified visibility and faster 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.

Axis Bank – Deputy Manager

Jul'22 – Apr'25 | Mumbai

- Designed and implemented** ETL pipelines and a self-service 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.
- Designed data infrastructure and pipelines for detractor close looping on the CRM platform, **resulting in a 17% NPS (Net promoter Score) increase** by providing actionable insights to sales teams.

Projects

TransitFlow: Real-Time Event Stream (using Kafka & Spark) [\[LINK\]](#)

- 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; enabled seamless integration of diverse datasets into Redshift, thus **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.

Reddit Sentiment Analysis Pipeline (using Airflow & Spark) [\[LINK\]](#)

- Engineered a comprehensive ETL pipeline utilizing Apache Airflow, PySpark, and MySQL for sentiment analysis of Reddit posts; **achieved 30% increase in data processing efficiency** while enhancing analytics capabilities for trend identification.
- Developed tailored Python scripts that automated extraction and loading workflows, resulting in a **time savings of 15 hours weekly** on routine database administration tasks for enhanced operational efficiency.

Achievements

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

- Recognition for the insights on Repeat Complaints led to a reduction in bank wide repeats by 40% as of Q1 FY'25.

Education

National Institute of Technology

Tiruchirappalli

M. Tech - Energy Engineering

Jun'22

Galgotias College of Engineering & Technology

Greater Noida

B. Tech - Mechanical Engineering

Jun'18

Courses & Certifications

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