

# ROHAN DUBEY

Lead Data Engineer | Distributed Systems | Real-Time Data Platforms | AWS Data Infrastructure

Hyderabad, India | +91-7093123447 | [rohankumardubey497@gmail.com](mailto:rohankumardubey497@gmail.com) | [GitHub](#) | [LinkedIn](#) | [Portfolio](#)

## SUMMARY

Lead Data Engineer with 8+ years of experience building production-grade distributed data platforms across real-time streaming, batch processing, and cloud-native lakehouse systems. Deep hands-on experience with Apache Flink, Kafka/MSK, Spark, Kinesis Data Analytics, DynamoDB, S3, and AWS data services. Strong record of reducing hot-path IO, improving reliability and recovery, tuning JVM/RocksDB workloads, and delivering cost-efficient data systems for analytics and AI/ML workflows.

## CORE TECHNICAL SKILLS

**Languages:** Java, Python, Go, Rust, Scala, SQL, C++, Shell

**Streaming & Batch:** Apache Flink, Kafka/MSK, Kinesis Data Analytics, Spark, Airflow, event-time processing, watermarks, exactly-once, checkpointing

**Cloud & Platform:** AWS S3, Glue, EMR, KDA, MSK, Redshift, Athena, Kubernetes, Docker, CI/CD

**Storage & Lakehouse:** DynamoDB, PostgreSQL, MySQL, Delta Lake, Apache Iceberg, Parquet, ORC, S3 data lake

**Specialties:** distributed systems, low-latency pipelines, state management, data modeling, schema evolution, observability, runtime tuning, cost optimization

## PROFESSIONAL EXPERIENCE

### Flutter Entertainment - Lead Data Engineer, Enterprise Tech

Jun 2026 - Present

Hyderabad, India

- Selected for Enterprise Tech to drive company-wide data and platform initiatives across Flutter's enterprise technology organization.
- Establishing scalable data delivery patterns for enterprise workflows, bringing production discipline from Streaming AI into reliability, operational clarity, and platform governance.

### PokerStars - Lead Big Data Engineer, Streaming AI

Jan 2026 - Jun 2026

Hyderabad, India

- Promoted to lead the Streaming AI data engineering track, guiding architecture and delivery for real-time AWS workloads supporting high-value production use cases.
- Improved runtime behavior and infrastructure efficiency by reducing hot-path IO, optimizing storage layout, and strengthening checkpoint/recovery strategies.
- Partnered with product and stakeholder teams to translate streaming and AI requirements into resilient data platform designs.

### PokerStars - Senior Big Data Engineer, Streaming AI

Aug 2023 - Jan 2026

Hyderabad, India

- Designed and operated real-time pipelines using Flink on Kinesis Data Analytics, MSK/Kafka, DynamoDB, and S3 for production workloads requiring low latency and high reliability.
- Built exactly-once sink patterns with checkpoint alignment, idempotent upserts, and failure-safe recovery; tuned RocksDB state, JVM memory, and backpressure behavior.
- Reduced cloud operating cost through DynamoDB access-pattern modeling, payload compaction, partition strategy, and S3 file/layout optimization.
- Improved observability and runbook quality across streaming jobs through metrics, alerting, and reproducible runtime configuration.

### Experian PLC - Senior Data Engineer

Mar 2023 - Aug 2023

Hyderabad, India

- Delivered secure, regulated data pipelines with ingestion controls, data quality checks, lineage, and monitoring to improve analytics readiness for client-facing use cases.
- Standardized batch and streaming job configuration to improve reproducibility, operational visibility, and handover quality.

### Model N - Member of Technical Staff 3

Jun 2021 - Mar 2023

Hyderabad, India

- Built event-driven analytics platform components with Kafka, Spark, Delta Lake, Dremio, and REST access patterns for downstream consumers.
- Improved query performance and cost efficiency using partitioning, Z-ordering, compaction, predicate pushdown, and lakehouse storage design.

### Carelon Global Solutions - Software Engineer

Aug 2020 - Jun 2021

Hyderabad, India

- Integrated Medicare and Medicaid datasets and SQL analytics workflows to support revenue operations, reconciliation, and reporting.

### Legato Health Technologies - Associate Software Engineer

Jul 2018 - Aug 2020

Hyderabad, India

- Built ETL workflows in Teradata and Informatica with SLA tracking, validation checks, and delivery standardization for healthcare data processing.

## SELECTED ENGINEERING PROJECTS

- **DataWizz** ([GitHub](#)): Production-grade lakehouse workspace inspired by Databricks, Snowflake, and ClickHouse; supports file ingestion, SQL querying, notebooks, pipeline orchestration, Delta-style publishing, and dashboards.
- **GoXStream** ([GitHub](#)): Flink-inspired stream processor in Go with operator graphs, connectors, checkpointing, and execution/runtime primitives.
- **FlowCore** ([GitHub](#)): Rust-based stream processor with event-time windows, watermarks, late-event handling, checkpointing, and a live dashboard.

## EDUCATION

Jawaharlal Nehru Technological University, Hyderabad - B.Tech in Electrical Engineering, GPA 4.0/4.0 (2014 - 2018)