NEELA KONDA REDDY BEERAM

Senior Data Engineer | Data Platforms, Streaming & Lakehouse

Hyderabad, India | +91 8499951955 | bnkr.raven@gmail.com | linkedin.com/in/neela-konda-reddy-beeram-893298376 | github.com/bnkr-breakthrough

PROFESSIONAL SUMMARY

Data Engineer with 4+ years building reliable batch, CDC, and streaming pipelines for high-volume banking and payments data. Delivered production workflows processing 60M+ records per run with zero production failures since deployment. Hands-on across Python, PySpark, SQL, Apache Spark, Kafka, Airflow, dbt, AWS, Snowflake, Databricks, and enterprise ETL.

CORE TECHNICAL SKILLS

Data Engineering: Python, PySpark, SQL, Shell, ETL/ELT, batch processing, streaming, CDC, SCD-1, dimensional modeling, star schema, data marts

Streaming & Lakehouse: Apache Spark, Spark Structured Streaming, Apache Kafka, Debezium, Delta Lake, Delta Live Tables, Apache Iceberg, medallion architecture

Cloud & Warehousing: AWS Glue, Amazon S3, Amazon Athena, Snowflake, Snowpark, Databricks, Google BigQuery

Orchestration & Reliability: Apache Airflow, dbt, dbt tests, data-quality checks, monitoring/alerts, dashboards, Unity Catalog

Enterprise & Delivery: Ab Initio GDE, Control Center, SQL Server, PostgreSQL, Docker, Kubernetes, Git, GitHub Actions, CI/CD, code reviews, mentoring, technical design, stakeholder discussions

PROFESSIONAL EXPERIENCE

Cognizant Technology Solutions | Data Engineer

Jul 2022 – Present | Hyderabad, India

- Engineered a production Ab Initio recovery pipeline for a card-platform migration, consolidating 30 daily card-network files (~64M records per run) into monthly backfill extracts; achieved zero production failures since deployment.
- Designed a cross-platform extract-and-enrichment pipeline that moved ~25M records across 13 months from Google BigQuery to SQL Server, using a database-generated event ID as a stable reconciliation key.
- Architected parallel processing paths for ISO 8583 authorization and clearing messages, mapping raw attributes into standard IPM clearing layouts through substring parsing, lookups, and header generation.
- Built error, reject, and logging streams around table loads, isolating invalid records to quarantine files rather than failing jobs and improving reliability for large-scale backfills.
- Automated a two-stage card-reissuance eligibility workflow using multi-source joins over a 12-month activity window, generating customer letters and operational alerts.
- Delivered pipelines through sandbox, QA, and production with Control Center; contributed to technical design, code reviews, mentoring, and stakeholder discussions for client delivery.

SELECTED DATA PLATFORM PROJECTS

OmniLake — Designed an open, multi-engine lakehouse: one governed copy of data in open table formats (Apache Iceberg, Delta Lake) queried interchangeably by Spark, Trino, and Snowflake, eliminating data duplication and vendor lock-in.

FinShield — Real-time fraud detection platform: streamed transactions via Kafka and PySpark Structured Streaming into Snowflake, scored risk in-database with Snowpark (z-score anomaly detection), and surfaced alerts on a SOC-style Streamlit dashboard refreshing every 10 seconds.

LedgerFlow — CDC banking lakehouse: captured every PostgreSQL insert, update, and delete via Debezium and Kafka into a Databricks medallion pipeline (Auto Loader to Delta Live Tables), applying correct out-of-order SCD-1 merges with Unity Catalog row/column governance and lineage.

OrderLake — Serverless AWS data lake: S3 to Glue PySpark ETL with four data-quality gates and a quarantine zone (98.84% pass rate) landing partitioned Parquet queried in Athena; 92% cheaper scans, schema-drift resilient, and built entirely within the AWS free tier.

CERTIFICATIONS

- AWS Certified Data Engineer – Associate — Amazon Web Services
- Databricks Certified Data Engineer Associate — Databricks
- SnowPro Core Certification — Snowflake
- Claude 101 — Anthropic (applied GenAI / LLM foundations)

EDUCATION

B.Tech, Computer Science & Engineering | KL University, 2022 | CGPA 8.25