

Subhajit Chowdhury

Cloud Data Engineer

Kolkata, India | +91 7029860941 | er.subhajitchowdhury@gmail.com

LinkedIn | GitHub | Portfolio | Medium

SUMMARY

Cloud Data Engineer with **nearly 4 years** of experience designing **production-grade ETL/ELT pipelines** on AWS and Snowflake. Proven track record of distributed system optimization, reducing infrastructure costs by **~£487K annually** while engineering **TB-scale Data Lakehouse** architectures using **Delta Lake** and PySpark.

SKILLS

Cloud & Data Platforms: AWS (S3, Glue, EMR, Redshift, Athena, Lambda, Step Functions, DynamoDB, DMS), Snowflake (Snowpipe), Delta Lake

Programming & Querying: Python (Pandas, NumPy), PySpark, SQL (PostgreSQL, MySQL, SQL Server, Oracle), Shell Scripting

Data Engineering: ETL/ELT, Dimensional modeling, Data lakes, Medallion architecture, Incremental processing, CDC, SCD

Orchestration & DevOps: AWS Step Functions, Jenkins, Git, Airflow (basic exposure)

Streaming & Messaging: EventBridge, Apache Kafka (working knowledge)

Tools: GitHub, Jira, Confluence, ServiceNow, Amazon QuickSight

EXPERIENCE

Tata Consultancy Services

Jul 2022 – Present

Data Engineer

- Reduced compute and storage costs by **~£487K annually** by optimizing S3 storage patterns and Glue/Athena query execution, migrating from legacy on-prem vendor systems to AWS.
- Engineered and maintained **150+ production-grade ETL/ELT pipelines** processing **hundreds of GBs daily with periodic TB-scale full refreshes** for downstream analytics.
- Decreased downstream data latency by **~30%** by implementing a metadata-driven ingestion framework using AWS Glue Crawlers for automated schema discovery.
- Designed and deployed **Delta Lake-based lakehouse architecture** providing ACID guarantees, schema enforcement, and upserts for reliable incremental data processing.
- Developed **near real-time event-driven pipelines** using EventBridge, Lambda, DynamoDB Streams, and Snowflake Snowpipe to handle continuous semi-structured JSON ingestion.
- Optimized SQL query performance by **20–25%** via partition pruning, predicate pushdown, and efficient joins across Redshift and Athena MPP environments.
- Built resilient orchestration workflows via **AWS Step Functions** and established CI/CD deployment pipelines using Jenkins and Git.

PROJECTS

Serverless Batch Analytics ETL Pipeline (AWS, PySpark, Athena)

- Built serverless ETL pipeline using S3, Glue, and Athena to ingest and structure analytical datasets.
- Implemented Medallion architecture with automated schema enforcement and incremental batch processing.

Near Real-Time ELT Pipeline (AWS, Snowflake)

- Constructed event-driven ELT pipeline utilizing EventBridge, Lambda, and Snowflake Snowpipe.
- Enabled near real-time continuous ingestion of semi-structured API data, replacing scheduled batch loads.

EDUCATION

B.Tech in Computer Science and Engineering, MAKAUT

2018 – 2022

DGPA: 8.90

CERTIFICATIONS

- AWS Certified Cloud Practitioner (Validation: a6447414ae25419cba12941481cd10a2) July 2025 – July 2028
- AWS Certified Data Engineer – Associate (DEA-C01) *In Progress*
- Databricks Fundamentals & Generative AI Fundamentals (Accreditations)