

AMIT PRASAD

Senior Data Engineer

aprasa02@gmail.com • +1 469-514-5855 • linkedin.com/in/amit-prasad-b030b616a/

SUMMARY

Results-driven **Senior Data Engineer** with 15+ years building cloud-scale data platforms and real-time/near-real-time pipelines. Deep expertise in **AWS (Glue, EMR, Lambda, Step Functions)**, **Snowflake**, **Apache Iceberg**, and **PySpark/Flink**. Strong in performance tuning, cost optimization, data modeling, and data quality/governance. Lead cross-functional delivery, mentor engineers, and ship reliable, observable systems end to end.

IMPACT HIGHLIGHTS

- Shipped a bureau-compliant **Metro2 single-line generator** handling **20M+ tradelines/run**, with parameterized reruns, auditability, and controls.
 - Improved Glue Spark throughput with **optimal partitioning, file sizing, and join strategies**; added **Iceberg compaction** to curb small files and cut S3 reads.
 - Built **event-driven** pipelines (Lambda → Step Functions → Glue) with **idempotency, retries, DLQs**, and end-to-end **tracing/alerts**.
 - Modernized analytics to **Databricks + Snowflake**, standardizing ELT and improving key workloads by **~40%**.
 - Operationalized **data governance** with DQ rules (Deequ-style), PII masking, lineage, and CI/CD for datasets and jobs.
-

CORE SKILLS

Programming & Big Data: Python, SQL, PySpark, Flink, Kafka

Cloud & Storage: AWS (Glue, EMR, S3, Lambda, Step Functions, CloudWatch), Snowflake, Apache Iceberg, Databricks

Pipelines & Orchestration: Airflow, Step Functions, DBT, Jenkins, Terraform, Git

DataOps & Governance: Deequ/GE-style DQ, masking, lineage, observability, cost/perf optimization

Ways of Working: Agile delivery, stakeholder alignment, mentoring

EXPERIENCE

Capital One — *Senior Data Engineer*

Apr 2024 – Present - Architected and delivered **Metro2** bureau reporting at scale (20M+ tradelines per

cycle) with parameterized runs, lineage, and audit trail.

- Built **serverless event flows** (API → Lambda → Step Functions → Glue) with DLQs, retries, and strong observability (metrics, logs, alarms).
- Tuned Spark jobs (skew handling, broadcast/hash joins, partition pruning, file sizing) to improve throughput and reliability.
- Implemented **data quality gates** and schema/version controls to protect downstream SLAs.

LTIMindtree — *Senior Specialist Data Engineer*

May 2023 – Jan 2024 - Led **Databricks + Snowflake** migration; standardized ELT patterns and tuned clustering/warehouses; key workloads improved **~40%**.

- Built **serverless ingestion** (Lambda + Python) for Qualtrics → curated models; added automated DQ and lineage.

Ascendion / Wipro — *AWS Data Engineer / Data Platform Lead*

Ascendion / Wipro — *AWS Data Engineer / Data Platform Lead*

May 2019 – Apr 2023 - Designed and implemented large-scale ETL on **Glue / EMR / PySpark**; led Snowflake migration and secure ingestion APIs.

- Established S3 data lake zones (raw → curated → consumer), schema evolution, and automated compaction jobs.

Wipro Ltd — *Senior Project Engineer – Data Engineering*

Mar 2015 – May 2019 - Engineered **Kafka-based** ingestion for financial events; stored raw feeds with registry and replay controls.

- Introduced partitioning/bucketing strategies; reduced shuffle and improved job reliability.

IG • Symphony Teleca — *Software Engineer / Developer*

2010 – 2015 - Delivered Java/Spring/JDBC modules, integration, and deployment; collaborated across QA and release teams.

SELECTED PROJECTS

Metro2 Credit Bureau Reporting (Capital One)

Event-driven architecture producing bureau-compliant single-line files for 20M+ tradelines with lineage and DQ checks.

Analytics Modernization (LTIMindtree)

Databricks + Snowflake migration; standardized ELT; improved performance ~40% and simplified governance.

Data Lake Foundation (Ascendion/Wipro)

Multi-zone S3 lake with schema registry, compaction, and incremental processing for BI/ML consumers.

EDUCATION & CERTIFICATIONS

B.Tech — Computer Science & Engineering, **Pandit Ravishankar Shukla University**
C-DAC India — Advanced Computing

TECH STACK (AT-A-GLANCE)

AWS (Glue, EMR, S3, Lambda, CloudWatch, API GW) • Snowflake • Iceberg • Databricks • Athena • Redshift • Flink • Kafka • Spark
Airflow • Step Functions • DBT • Terraform • Jenkins • Git • Tableau (basic)
Python • SQL • Java • Shell • Golang (basic)

Keywords for ATS

Senior Data Engineer • AWS • Glue • EMR • Lambda • Step Functions • Snowflake • Databricks • Iceberg • PySpark • Flink • Kafka • ETL/ELT • Airflow • Data Quality • Deequ • Lineage • CI/CD • Observability • Cost Optimization • Real-time Streaming

References available on request