

Rohit Jadhav

Data Engineer

Redmond, WA · rhjadhav2795@gmail.com · (979) 250-2973 · linkedin.com/in/rohitjadhavmis · Portfolio & AI assistant ↗

SUMMARY

Data Engineer with ~5 years at Amazon designing and modernizing large-scale data platforms on AWS — distributed batch and near-real-time pipelines, data modeling, and data-quality and governance. Comfortable owning systems end to end, from infrastructure-as-code through production operations and cross-team migrations. AWS-certified, CFA Level I, M.S. in MIS.

EXPERIENCE

Amazon — Data Engineer

Jul 2021 – Present · Seattle / Redmond, WA

AWS Analytics Engineering · Feb 2025 – Present

- Re-architected a business-critical cloud usage & revenue pipeline (~360 TB/day, ~451B rows) from a Redshift-based processing engine to **Spark + Apache Iceberg** — removing a single point of failure and shared-cluster contention, and enabling fast rollback of bad data via time travel.
- Improved data availability by analyzing downstream usage and re-partitioning so the critical data used by **80%+ of consumers lands ~2 hours earlier**; added data-quality alarms and a source-team SLA + runbook to catch issues before they reach reporting.
- Safely migrated and deprecated **53 tasks across 3 legacy ETL workflows with zero disruption to 50+ downstream reports**, authoring a backfill SOP reused by later migrations.
- Prototyped **knowledge-graph memory for AI agents** (Amazon Neptune + Bedrock) for deterministic multi-hop reasoning over operational relationships where RAG fell short, and proposed a federated framework for org-wide adoption.

Intelligent Talent Acquisition · Apr 2023 – Jan 2025

- Built a **self-service audience-segmentation engine**: a recursive translator compiling a UI's nested JSON rule tree into Spark SQL predicates, with SCD2 versioning on Iceberg, scoring membership across **10M+ candidate records** — replacing a manual, ticket-driven CSV workflow.
- Rebuilt the pipelines behind a **media-mix model** to be globally configurable (config-only country onboarding) with a 3-strategy geo-cluster resolver and multi-grain weekly reconciliation; integrated a new paid-social channel for first-ever site-level ROI.
- Built an **observability data platform from scratch** — a dimensional model over 3 heterogeneous sources (NoSQL CDC, multi-schema S3 files, unstructured logs) into 12 queryable datasets, on event-driven serverless ETL with reusable CDK constructs.

Seller Partner Support · Jul 2021 – Mar 2023

- Built a **near-real-time, serverless (Lambda) ETL** replacing fragile EC2 polling — 60s OpenSearch scrolls with high-water marks, nested-JSON→relational transforms, and merge-upsert into partitioned Aurora PostgreSQL across 3 regions — keeping ops dashboards unchanged through a routing-platform migration.
- Revamped a legacy seller-support analytics ETL from script-based **PostgreSQL to AWS Glue + S3, cutting query time from several hours to under a minute**.
- Led an org-wide **infrastructure cost-efficiency initiative** — decomposed a composite KPI, built a reusable prioritization framework, and drove the org to a green efficiency score.

Texas A&M Health — Database Administrator & AWS Support Associate

Mar 2020 – May 2021 · College Station, TX

- Delivered a relational data model for a health-research initiative, migrating from manual data management to a maintainable research-lifecycle data store.
- Implemented HIPAA-compliant patient surveys and performed data-security analysis for sensitive research data.

Mays Innovation Research Center, Texas A&M — Lead Math Analyst

Sep 2019 – Mar 2020 · College Station, TX

- Built MATLAB models, optimizations, and visualizations supporting the center's immigration-research and performance-targets studies, evaluating data under varying constraints to inform the directors' analysis.

Tata Consultancy Services — SAP Process Integration & ABAP Developer

Oct 2017 – May 2019 · Mumbai, India

- Developed a SAP PI data interface automating purchase-order and shipment transfer to a Truck Access Control System, and built ABAP reports and batch data-migration programs (legacy → SAP R/3).
- Sustained a 95%+ ticket-resolution rate across cross-functional support; awarded Star of the Month.

SKILLS

Technical: Python · SQL · Java · Apache Spark · Apache Iceberg · Apache Kafka · Apache Airflow · Amazon EMR · AWS Glue · Amazon S3 · Amazon Redshift · AWS Lambda · Amazon Kinesis · Amazon DynamoDB · Aurora / RDS PostgreSQL · Amazon OpenSearch · Amazon Neptune · Amazon Bedrock · Amazon SageMaker · AWS Lake Formation · Amazon CloudWatch · AWS CDK (IaC) · Dimensional data modeling · Data warehousing & lakehouse · ETL / ELT · Change data capture (CDC) · Slowly-changing dimensions (SCD2) · Data partitioning & performance tuning · Data quality & observability · SAP (PI, ABAP)

Non-technical: Cross-functional collaboration · Stakeholder management · Technical leadership & mentoring · Agile / Scrum · End-to-end system ownership

EDUCATION

Texas A&M University — Mays Business School

College Station, TX · 2021

M.S., Management Information Systems · GPA 3.83 / 4.0

University of Mumbai — D. J. Sanghvi College of Engineering

Mumbai, India · 2017

B.E., Electronics & Telecommunication Engineering · GPA 3.51 / 4.0

CERTIFICATIONS

AWS Certified Solutions Architect – Associate · AWS Certified Cloud Practitioner · CFA Program — Level I (Passed)

PUBLICATION

Hardware Implementation of a Convolution Encoder — DJ Spark 2017, D. J. Sanghvi College of Engineering