

AKSHAY CHAUDHARY

Senior Data Engineer/ Data Platform Solutions Architect

(425) 287-8887 | akshay.96c@gmail.com | Seattle, WA | [linkedin.com/in/akshay96c](https://www.linkedin.com/in/akshay96c)

SUMMARY

Senior Data Engineer and data platform architect with 8+ years of experience designing and operating large-scale cloud data systems across streaming ingestion, distributed processing, data warehousing, governance, and analytics. Built high-throughput pipelines, self-service data platforms, and analytics-ready datasets supporting AI and business-critical workloads at Amazon across Selling Assistant, Prime Video, and Amazon Devices, with earlier 0-to-1 impact at Cars24, Rivigo, and OYO.

SKILLS

Data Engineering & Warehousing: Spark/PySpark, advanced SQL, batch and streaming pipelines, dimensional and lakehouse modeling, schema design and evolution, change-data handling, backfills/reprocessing, partitioning and performance tuning, data quality gates, monitoring and operational guardrails

Cloud & Data Platforms: AWS (S3, Bedrock, DynamoDB/Streams, Lambda, Kinesis, Firehose, Glue, Step Functions, Redshift, Athena, IAM/STS, CloudWatch), OpenSearch, Redis, Airflow, dbt, Kafka, Flink, Apache Iceberg, AWS CDK/CloudFormation.

AI/LLM Data Systems: RAG engineering, knowledge-base ingestion, document normalization, retrieval-context preparation, vector search/indexing, agentic workflows, MCP servers, LLM observability and evaluation loops, LangChain, Anthropic SDK.

Full-Stack Data Products: Python/FastAPI, TypeScript, Next.js/React, Streamlit, REST APIs, CI/CD, cloud deployment, internal tooling, system design.

Product Analytics: KPI design, experimentation/A/B testing, funnel/cohort and retention analysis, Tableau.

EXPERIENCE

Senior Data Engineer | Amazon · AI Selling Assistant

Seattle, WA · Apr 2025 – Present

- Founding data engineer for Selling Assistant; owned the LLM data platform for knowledge ingestion, retrieval context, product telemetry, and analytics across the product used by 1.3M+ sellers. ([public launch link](#))
- Architected a **multi-region, near-real-time telemetry ingestion platform processing hundreds of thousands of LLM trace signals per second**, with idempotent and replay-safe processing, automated data quality checks, failure recovery, and curated gold datasets powering PO metrics, benchmarking, evaluation, and self-service analytics.
- Architected the **analytical warehouse for LLM operations**, defining fact and dimension models for requests, retrievals, evaluations, costs, feedback, and downstream business outcomes. Implemented incremental processing, late-arriving data handling, historical backfills, and SCD Type 2 history, with aggregate tables supporting quality, latency, cost, adoption, and LLM observability.
- Designed and launched a **full-stack self-service LLM evaluation platform** that enabled PMs to configure sampling and audit rubrics, generated multilingual audit workflows, and transformed returned annotations into science- and analytics-ready datasets. Adopted by 100+ human and model-based auditors.

Data Engineer | Amazon · Amazon Devices

Bellevue, WA · Jun 2022 – Apr 2025

- Architected the **end-to-end pipeline powering ad targeting for Amazon Devices, processing terabytes of traffic data to publish curated query-ASIN candidate sets and feature tables** (impressions, clicks, CTR, organic rank, eligibility) across global marketplaces, consumed by modeling and real-time ads publishing systems. Improved feature quality and release safety (guardrails, alerting, feedback loops), driving **+27% attributed units and +23% attributed revenue at flat budget**.
- Developed an AWS-native governance platform that **scanned 300+ Redshift clusters across orgs daily to audit Seller Data Protection Policy** compliance and auto-generate violation tickets, **helping reduce exposure following Amazon's record \$887M GDPR fine**.

Business Intelligence Engineer | Amazon · Prime Video Search

Seattle, WA · Jun 2021 – Jun 2022

- Designed a last-impression attribution data model linking streams to final search impressions and built the analytical serving layer for a Tableau product used by 50+ stakeholders to analyze 1,000+ keywords across 8+ marketplaces and device surfaces, influencing franchise planning, campaign effectiveness, and content acquisition.

Product Data Scientist | Rivigo

Gurugram, India · Nov 2019 – Oct 2020

- Built data-driven incentive and payout systems for logistics operations, combining operational data, performance metrics, and scenario modeling to save \$574K annually and improve network efficiency during COVID-19.

Data Scientist | Cars24

Gurugram, India · Oct 2017 – Nov 2019

- Early data science hire; partnered with the CFO and led four analysts in building analytics infrastructure and decision systems for finance and product teams. Developed credit-risk and fraud-detection systems that **increased loan conversions by 37% and reduced delinquencies by 58%**.

EDUCATION

Master of Science in Business Analytics · University of California, San Diego · GPA 3.8/4

2020 – 2021

Bachelor of Technology in Computer Science · Guru Gobind Singh Indraprastha University · GPA 3.6/4

2014 – 2018