Adithya Shankaran

Data Engineer
+91 8780135961 | adikshan11@gmail.com | Gurugram, Haryana
LinkedIn: linkedin.com/in/adithya-shankaran | GitHub: github.com/adikshan11 | Portfolio: adithyashankaran.vercel.app

Summary

Data Engineer with **2 years** at **Xebia**, working directly with a global apparel retail brand's product owners and **BI analysts** to build the **ETL/ELT pipelines**, **CDC**, **data models** and **data quality** checks on **GCP** behind their **Looker Studio** sales and demand reporting, plus **RAG** and **LangGraph** AI agents.

Experience

Xebia | *Full-time, Gurugram, India* **Aug 2024 – Present**

Junior Consultant – Data Engineer | Global apparel retail brand (client) Jul 2026 – Present

- Unified duplicated **Dataproc** pipeline templates into one reusable framework component powering **524 production schedules** with zero output regressions, plus time limits on **600+ DAGs** that expose silent 4-hour failures.
- Designed and released per-task time limits in the shared **Vertex AI** pipeline framework, sized from measured production run times with alerts naming the failed task; enforced on **59 tasks** across **25** e-commerce pipelines.
- Ended repeated 4-hour timeouts on a high-cost Hybris **Spark** pipeline (0 of 9 runs finished before, 11 of 16 after) and cut its daily cost **19.7%** by pruning joins and scans; right-sized clusters for **29 pipelines** (**17.6%** lower cost).
- Led preprod validation of **7 Hybris** pipelines on the new **SQL Server Hyperscale** source; all 7 matched.
- Triaged **production incidents**, reviewed **102 PRs** in **7 repos** and wrote **7 design and analysis docs**.
- Restored a missing campaign attribute for the site marketing team's twice-weekly reporting with a chained **DBT** backfill across **6 months** of events, reconciling **4,200** hourly windows and repairing **1,455 missing records**.

Technical Trainee | Centre of Excellence department | Client-focused project Aug 2024 – Jun 2026

- Led the migration of **13 e-commerce data sources** from Databricks to **Dataproc** and **BigQuery**, each validated.
- Designed **CDC** for the **OMS order tables** with **Datastream** and BigQuery snapshot and change views as the **single source of truth** for order reporting, to replace **Qlik**.
- Built **Dataplex** data-quality scans for **2 data products** with thresholds tested on years of production data, and a **CI gate** that blocks DBT models without a scan.
- Centre of Excellence: built **RAG** extraction with **LlamaIndex** chunking into **Weaviate** and **ChromaDB** vector stores, and a **Claude** (**AWS Bedrock**) Terraform linter for all **6 Well-Architected pillars**.

Projects

imaarat.ai | *FastAPI, DBT, DuckDB, LangGraph, Langfuse, PostgreSQL*

- Open-source **AI underwriting platform** for Indian commercial property: a deterministic rule engine scores each proposal against official seismic, flood and cyclone data for all **19,312 PIN codes**, AI explains, a reviewer signs off.
- Built the data layer: **Shapely** joins for PIN hazards, nightly **ELT** from PostgreSQL to **Parquet** with **9 DBT models** and **26 tests** on DuckDB; **LangGraph** agents read **26 languages** over **MCP/A2A**.

NYC Taxi Lakehouse | *PySpark, Delta Lake, DBT, DuckDB, Airflow*

- **Delta Lake** lakehouse over **9.5M trips**: quarantines **3.9%** bad records via 5 reject rules, blocks bad loads with a 7-check quality gate, and serves DBT marts with **63 tests**.

Technical Skills

Languages: Python, SQL
Frameworks: PySpark, DBT, Airflow, Databricks, Delta Lake, LangGraph, LlamaIndex, MCP, A2A
Cloud (GCP): BigQuery, Managed Service for Apache Spark (Dataproc), Knowledge Catalog (Dataplex), Vertex AI Pipelines, Datastream, Dataflow, Cloud Storage, Pub/Sub
Cloud (AWS): S3, Bedrock, SageMaker, Redshift
Databases: Hive Metastore, PostgreSQL, SAP HANA, Weaviate, ChromaDB, MongoDB, Snowflake
Tools: Git, GitHub Actions, Docker, Looker Studio, Tableau, Claude Code, Copilot, Amazon Q, Jira, Confluence

Certifications & Achievements

Certified: FDE-1 Level 2, GCP (Oct 2026); GitHub Copilot X-AI Practitioner+ (2026); Salesforce Developer (2023)
Activities: Mentor, Xebia Tech-AI-Thon; NSS Community Lead; Grade 8 Keyboard, Trinity College London

Education

Birla Institute of Technology and Science (BITS), Pilani **Oct 2026 – Present**
Master of Technology in Data Science & Engineering *India*

Kalinga Institute of Industrial Technology **Sep 2020 – Jun 2024**
Bachelor of Technology in Computer Science & Communication Engineering (8.32 CGPA) *Odisha, India*