Aditya Sharma

SENIOR DATA ENGINEER

Berlin, Germany · +49 174 529 9541 · adisharma2104@gmail.com · linkedin.com/in/aditya-sharma-b57b20169 · github.com/adisharma30

PROFESSIONAL SUMMARY

Senior Data Engineer with 7+ years building and operating production data platforms across telecom, payments, healthcare and finance. Designs distributed batch and streaming ETL/ELT pipelines in Python and PySpark, models data in dbt across Medallion (Bronze/Silver/Gold) layers in Snowflake and Databricks, orchestrates workloads with Apache Airflow, and manages real-time ingestion with Apache Kafka on AWS (S3, IAM, Lambda). Track record of replacing manual data handling with automated, tested, monitored workflows — cutting executive reporting effort by 42%, reducing reporting delays by 37%, and running platforms processing over 200 million records daily — alongside hands-on Spark performance tuning and containerised deployment on Kubernetes. Leads small engineering teams and translates stakeholder requirements into durable data models. English C2, German A2 (targeting B1 by Feb 2027).

TECHNICAL SKILLS

Languages	Python (Pandas, NumPy, PySpark), SQL
Data engineering	ETL/ELT design and development, dbt, batch and streaming processing, data modelling (relational and dimensional), data warehousing, pipeline automation
Streaming	Apache Kafka (producers, consumers, topic design, consumer groups), real-time ingestion, event-driven data flows
Warehousing	Snowflake, Medallion architecture (Bronze/Silver/Gold), Amazon Redshift, PostgreSQL, Teradata, MySQL, NoSQL
Distributed processing	PySpark (partitioning strategy, join optimisation, performance tuning), Databricks
Orchestration & CI/CD	Apache Airflow, Jenkins, Git
Cloud & infrastructure	AWS (S3, IAM, Lambda), Databricks, Kubernetes
Data quality	Validation and reconciliation logic, dbt tests, data quality management, pipeline monitoring
Serving & BI	Tableau, Power BI, Streamlit, Plotly Dash, FastAPI
AI & generative AI	LLM integration, RAG, LangChain, Hugging Face, prompt engineering
Ways of working	Agile delivery, stakeholder management, requirements gathering, team leadership, technical documentation

PROFESSIONAL EXPERIENCE

Senior Data Engineer 07/2023 – 06/2026

EXL

- Led a team of four engineers consolidating five fragmented payments source systems into a single standardised, modelled schema processing over 200 million records daily, using Python, Pandas and PySpark, delivering every milestone on schedule.
- Designed and deployed automated ETL/ELT workflows orchestrated in Apache Airflow, replacing recurring manual data-handling routines and recovering approximately 5 staff hours per week for higher-value engineering work.
- Built and managed Apache Kafka streaming pipelines for near real-time ingestion of transactional events, designing topics and consumer groups to deliver data into the warehouse layer with reliability and backpressure handling that batch processing could not provide.
- Structured the platform using a Medallion architecture across Snowflake and Databricks, defining Bronze, Silver and Gold layers so raw ingestion, cleansed modelling and business-ready datasets were cleanly separated and independently testable.

- Built a modular transformation layer in dbt with 80+ version-controlled models and automated tests that reduced data-quality incidents by approximately 45% and materially strengthened the reliability of downstream reporting.
- Optimised heavy PySpark transformation workloads through repartitioning and join optimisation, reducing job runtime by approximately 42% and improving pipeline throughput on the platform's largest datasets.
- Engineered a production generative AI capability — LLMs embedded into existing data and finance pipelines and deployed as containerised services on Kubernetes — automating insight generation, cutting manual analysis time by approximately 60%, and establishing a reusable pattern for enterprise AI adoption.
- Delivered executive KPI dashboards consumed directly by six C-suite leaders, reducing manual reporting effort by 42% and giving leadership a single trusted view of core operational metrics.

Data Engineer

01/2021 – 07/2023

ZS Associates

- Automated large-scale clinical data pipelines in Python and PostgreSQL, replacing manual extract-and-refresh routines and accelerating reporting cycles by approximately 30% for trial operations teams.
- Modelled patient- and site-level datasets to support operational analysis across clinical trials, contributing to a 13% improvement in trial efficiency.
- Delivered the data engineering and modelling foundation behind a clinical-trial operations optimisation programme that increased annual revenue by 32% and saved approximately USD 20M (approx. EUR 17.2M).
- Built an ML-driven automation application that mined historical trial data to auto-generate clinical study protocols, cutting protocol drafting from 8–9 months to 3–4 weeks (an approximately 90% reduction); initiated as a proof of concept, it was adopted by client leadership and funded as a full-scale project.
- Trained junior associates on advanced SQL and Tableau concepts through hands-on workshops, raising the team's self-sufficiency and reducing dependency on senior engineers for routine analysis.

Data Engineer — ETL & BI

01/2019 – 12/2020

Reliance Jio Platforms

- Designed and implemented ETL processes integrating heterogeneous source systems into a central data warehouse, reducing reporting delays by 37%.
- Authored and optimised complex SQL queries and data models across customer and network datasets, improving query performance by approximately 38% and the accuracy of business reporting.
- Built interactive Tableau dashboards on engineered network and customer data, increasing operational KPI visibility across 5+ business teams.
- Supported new product releases by building the data layer behind launch tracking — instrumenting adoption, active users and feature engagement so product and business teams could monitor uptake from day one.

EDUCATION

B.Tech, Information Technology
Ramdeobaba University (RCOEM), India
06/2015 – 05/2019

CERTIFICATIONS

- AWS Certified Data Engineer – Associate
- Tableau Desktop 1: Fundamentals (07/2020)

LANGUAGES

English (C2) · Hindi (Native) · German (A2, targeting B1 by Feb 2027)

SELECTED PROJECTS

- **Zomato E2E Data Engineering** — FastAPI to Kafka to MinIO to Snowflake, with dbt staging, dimensions and marts, orchestrated in Airflow.
- **Real-Time Trading Modern Data Stack** — streaming quote pipeline through Kafka and MinIO into Snowflake, with dbt bronze, silver and gold models.

RECOGNITION & ACHIEVEMENTS

- Recipient of the "Game Changer" award at ZS Associates.
- Designed and delivered internal training on advanced data-engineering methodologies to 50+ analysts across 7–8 sessions, recognised internally at EXL.
- Authored the team's data engineering standards — dbt model conventions, testing requirements and pipeline documentation — adopted across the payments transformation workstream.

- Mentored junior data engineers onto Spark and Airflow workloads, and interviewed candidates for data engineering roles.
- Finalist, HackM Hackathon (Government of Maharashtra); Top 10 nationally, Smart India Hackathon, 2018.