

Ritesh Nahavre

Data Engineer | Python, Airflow, Databricks, Kafka, Snowflake | HIPAA/PHI Data Platforms | Generative AI
+91 8527621700 | riteshnahavre2909@gmail.com | [LinkedIn](#) | [GitHub](#) | [LeetCode](#) | Gurugram / New Delhi, India
Serving notice period | Last Working Day: 27 November 2026

SUMMARY

Data Engineer with **2.5 years** owning HIPAA-compliant PHI data platforms for **50+** channel partners end to end. Architects identified and de-identified data layers; codes platforms and ETL pipelines in native **Python** on **Airflow** and on **Databricks**, with **Kafka** streaming; maintains data quality and operationalizes support with **LLM/RAG** assistants; and publishes analytics-ready data to **Snowflake**, powering **10** client dashboards.

WORK EXPERIENCE

Data Engineer Jul 2024 – Present
ZS Associates Gurugram, India

Architect: Future-State Data Platform | Identified & De-identified Layers

- Served as data architect, designing the future-state identified (PHI) and de-identified data layers for a big pharma client and integrating **6+** sources (Hub, Specialty Pharmacy, CRM, Survey, Website, Digital) through source-to-target mappings into a unified data lake.
- Standardized patient data via MDM (entity resolution, deduplication), cutting duplicate active records by **10%**.

Build: Code-First Platforms & ETL Pipelines | Native Python, Airflow, Databricks, Kafka

- Engineered the pipeline backend as production **Python** microservices deployed on **Airflow** (code-first, not low-code ETL), processing **10M+** records across ingestion, DQM, transformation, and export on **S3** and **Redshift**.
- Built **Databricks** pipelines with **Delta Live Tables** and **Delta Lake** and tuned Spark/SQL workloads, cutting runtime by **25%** and compute cost.
- Architected a **Kafka** streaming pipeline handling **50+** daily patient enrollments, giving Field Reimbursement Managers new patients within **30 minutes** to start Benefits Investigation and Prior Authorization.
- Packaged reusable Snowflake and Databricks accelerator modules for client pipelines, cutting delivery effort **50%**.

Operate: Data Quality & RAG-Assisted Support | Workstream Lead, LLMs, RAG, AI Agents

- Led the data quality workstream, rolling out **1000+** DQM checks to raise completeness and accuracy.
- Automated release validation with a Snowflake-based framework, cutting manual validation by **70%**.
- Created an **LLM/RAG** smart assist that reads Airflow failure logs and returns the next best action to fix failed jobs, so on-call engineers resolve incidents without manual log triage.
- Extended the assistant to propose and apply context-aware DQ checks during project setup.

Deliver: Analytics on Snowflake & Downstream Systems | Client Impact Award 2025

- Implemented HIPAA Expert Determination and published de-identified PHI to **Snowflake** and downstream systems as the analytics layer, powering **10** client dashboards in Tableau.
- Delivered the SFLC clinical-trial orchestration tool (slot management, benefit verification, apheresis, surgery scheduling) for a Parkinson's program across **100+** trial sites, earning the Client Impact Award 2025.

INTERNSHIP

Software Engineering Intern Jun 2023 – Aug 2023
DRDO – DESIDOC New Delhi, India

- Developed a full-stack web platform on relational databases managing **1000+** digital publications.

TECHNICAL SKILLS

Programming: Python, SQL, C++ (Data Structures & Algorithms)
Orchestration & Streaming: Apache Airflow (custom Python microservices), Apache Kafka
Lakehouse & Big Data: Databricks, Delta Live Tables, Delta Lake, PySpark, Spark SQL
Cloud & Warehousing: AWS (S3, Lambda, Redshift), Snowflake, MSSQL, Data Modeling, Data Warehousing
Quality & Compliance: DQM Frameworks, MDM, HIPAA Compliance, PHI De-identification, Expert Determination
Generative AI: LLMs, RAG, AI Agents, Prompt Engineering, Claude, GitHub Copilot
Tools: Git, Jira, Agile, Tableau, Salesforce Life Sciences Cloud

EDUCATION

Maharaja Agrasen Institute of Technology, GGSIPU New Delhi, India
Bachelor of Technology in Computer Science and Engineering — CGPA: 9.27/10 Nov 2020 – Jul 2024