

SAURABH PANDEY

+91-7752989892 · pandey.saurabh4748@gmail.com · Chandigarh, India

SUMMARY

Data Engineer with 3.5+ years of experience designing scalable, cloud-native data platforms and backend data services across GCP and AWS. Specialized in CDC-based ingestion, warehouse optimization, CI/CD-enabled data workflows, and AI-integrated systems. Proven ability to improve performance, reduce infrastructure costs, and deliver reliable, low-latency data solutions supporting product analytics and enterprise applications.

TECHNICAL SKILLS

Cloud: GCP (BigQuery, Dataflow, Datastream, Cloud Run, Vertex AI, AlloyDB), AWS (S3, Glue, Lambda, Redshift)

Architecture & Engineering: Medallion (Bronze–Silver–Gold), CI/CD Pipelines, CDC Pipelines, Incremental Processing, Partitioning & Clustering, Metadata-Driven ETL, Data Governance, Query Optimization

Databases: BigQuery, Snowflake, PostgreSQL, MySQL, MS SQL Server

Programming: Python, Advanced SQL, FastAPI, Stored Procedures, Docker, Git

Generative AI: RAG Pipelines, ChromaDB, Embedding Workflows, LLM Integration, Semantic Search Optimization

Analytics & Monitoring: Looker, Splunk, Periscope

PROFESSIONAL EXPERIENCE

Harmony Data Integration Technology Pvt. Ltd. - Mohali, Punjab

AI & Data Engineer | August 2024 – Present

- Architected CDC-based Bronze–Silver–Gold data platform in BigQuery, improving reporting performance by 40% and reducing warehouse costs through partitioning and clustering strategies.
- Designed metadata-driven ETL framework enabling scalable schema evolution and accelerating onboarding of new data sources by 60%.
- Developed idempotent ingestion pipelines in AlloyDB and implemented CI/CD workflows for automated testing and deployment, improving release reliability.
- Optimized complex SQL workloads and warehouse schemas, reducing downstream API latency by 30% for product-facing analytics services.
- Led development of production-grade RAG system integrating vector embeddings and LLM APIs, reducing semantic retrieval latency by 50% and enabling intelligent search across structured and unstructured datasets.
- Collaborated with analytics, and product teams to design scalable data models aligned with application requirements.

Priority Technology Holdings, Inc. - Chandigarh

Data Engineer | July 2022 – August 2024

- Led database and warehouse performance optimization initiatives, improving API and query performance by 70–80% through indexing, execution plan analysis, and query refactoring.
- Owned MS SQL Server → MySQL & Snowflake migration strategy, ensuring full data reconciliation and zero critical downtime.
- Designed optimized reporting models and stored procedures for high-volume financial transaction systems handling sensitive payment data.
- Automated compliance and audit workflows using API-based integrations, improving traceability and reducing manual intervention.
- Performed peer reviews and established best practices for SQL and Python-based data pipelines.

Finxera - Chandigarh

Database Engineering Intern | January 2022 – June 2022

- Automated financial reporting pipelines using Python schedulers, eliminating repetitive manual reporting tasks.
- Optimized complex financial analytics queries, improving execution time and reporting accuracy.

EDUCATION

University Institute of Engineering & Technology, Panjab University - Chandigarh

Bachelor of Engineering in Computer Science | GPA: 8.4 / 10 | August 2018 – July 2022