Mudit Nautiyal

nautiyal.mudit.21@gmail.com | linkedin.com/in/mudit-nautiyal | github.com/MuditNautiyal-21 | Portfolio

Summary

Data Engineer with 6+ years building high-volume ETL pipelines, enterprise data integration systems, and ML-driven monitoring at HCL Technologies. MS in Data Science from University at Buffalo. Experienced in AWS, Spark, Docker, SQL optimization, and deploying production data applications with containerized architectures.

Technical Skills

Data Engineering & Cloud: AWS (S3, Glue, Redshift), Apache Spark, Airflow, Docker, ETL/ELT Pipelines, Data Modeling, Informatica MDM, SAP, CI/CD, Git/GitHub
Databases: SQL Server, Oracle, PostgreSQL, MySQL, SQLite
Languages: Python (Pandas, NumPy), SQL (Advanced), Bash, R, C++
ML & Analytics: PyTorch, Scikit-learn, GenAI/LLMs (LangChain, RAG), Power BI, Tableau, A/B Testing

Professional Experience

HCL Technologies Noida, India
Senior Data Engineer *Jan. 2021 – Apr. 2024*

- Architected high-volume data ingestion pipelines between SAP and Informatica MDM using AWS S3 and SQL, processing **100,000+** daily transactions with high-availability SLA compliance.
- Refactored legacy SQL transformations and ETL synchronization logic, reducing data processing cycle time by **45%** (60 min → 35 min) and improving downstream data freshness for operational reporting.
- Led SQL Server and Oracle integration to build a unified "Golden Record" source of truth, collaborating cross-functionally with analysts, SAP teams, and business owners to reduce data inconsistencies by **60%** and streamline enterprise reporting.
- Developed ML-driven pipeline failure prediction model identifying **85%** of system failures 24+ hours in advance, reducing unplanned downtime by **30%**. Awarded **HCL Innovation Box (2023)**.

Uttarakhand Aaj (Digital News Platform) Dehradun, India
Data Analyst & Product Strategist *June 2018 – Dec. 2020*

- Analyzed user traffic and engagement patterns across a **13-state** deployment using Python and SQL, delivering data-driven recommendations that shaped the engineering roadmap and platform scaling strategy.
- Reduced query latency by **30%** during peak traffic by profiling database bottlenecks and implementing composite indexing, enabling faster report generation for editorial and product teams.
- Built the analytical case for migrating from monolithic to modular architecture by quantifying downtime patterns and performance bottlenecks, delivering executive presentations that accelerated leadership approval for the infrastructure overhaul.

Projects

SUDNAXI (Trading Intelligence Platform) | *Python, Docker, PostgreSQL, Streamlit, REST APIs*

- Architected a full-stack trading intelligence platform with modular Python codebase (core/, ml/, utils/, database/), Docker containerization, and PostgreSQL/SQLite dual-database support, processing real-time market data across **9** global exchanges.
- Built ML-driven adaptive strategy engine with 30-minute learning cycles and a backtesting pipeline generating **1,000+** simulated trades for risk-adjusted position sizing.
- Shipped production deployment supporting sustained real-time data ingestion across **9** exchanges by containerizing with Docker Compose, implementing environment-based config management, and adding API rate limiting. **68 commits, MIT licensed**.

Eagle Eyes | Industry-Sponsored (Client: Nissha Medical Technologies) Buffalo, NY
Team Lead, Vision AI (University at Buffalo) | Capstone *Aug. 2025 – Dec. 2025*

- Optimized end-to-end YOLOv8 inspection pipeline using CUDA + TensorRT (GTX 1660 Ti), reducing inference latency by **45%** (117ms → 62ms) and enabling real-time throughput of **∼30M tickets/day**.
- Engineered deterministic validation gates (count, visibility, density, layout) with IoU-based logic, reducing false positives by **15%** and improving automated quality audit traceability.
- Built a two-stage fallback pipeline (primary + recovery with TTA/low-confidence settings) ensuring robust performance on degraded input frames for edge-deployment readiness.

MediFriend (AI-Powered Medical Billing) | *Python, OCR, LangChain, Gemini, React/TypeScript*

- **Winner, AI Hackathon:** Built medical billing automation using OCR document ingestion and a RAG pipeline (LangChain + Gemini) for billing code extraction and discrepancy detection, reducing manual reconciliation **∼70%**.

Education

University at Buffalo, SUNY Buffalo, NY
Master of Professional Studies, Data Science & Applications *Aug. 2024 – Dec. 2025*

Guru Gobind Singh Indraprastha University Delhi, India
Bachelor of Technology, Electronics & Communication Engineering *Aug. 2014 – May 2018*

Certifications: Math for ML, Data Science & Business Analytics, BCG X Data for Decision Makers (Forage), Project E-Portfolio