NITESH RANJAN SINGH

University of Glasgow | Glasgow, UK | +44 7900 945064 | 3090808S@student.gla.ac.uk | LinkedIn | GitHub | Portfolio

PROFESSIONAL SUMMARY

MSc Data Science student at the University of Glasgow with 3+ years of experience building scalable data solutions using Azure, Databricks, and PySpark. Certified Azure and Databricks Data Engineer skilled in ETL pipeline design, cloud data integration, and analytics across healthcare and enterprise domains. Passionate about turning complex data into actionable insights and now seeking a Graduate Data Engineer, Analyst, or Data Scientist role in the UK.

KEY SKILLS & TOOLS

Programming: Python, SQL, PySpark, Scala, DAX
Data Engineering: Azure (ADF, Databricks, ADLS Gen2, Synapse, Fabric), Snowflake, Delta Lake, Apache Spark, Airflow, Kafka
Cloud & DevOps: Azure Data & AI Stack, AWS (S3, Glue, Lambda), Docker, CI/CD (Azure DevOps, GitHub Actions)
Databases: Azure SQL, PostgreSQL, MySQL, Vector DB
Data Science & ML: Pandas, NumPy, Scikit-learn, MLflow, TensorFlow, PyTorch, Feature Engineering
Visualization & BI: Power BI, Tableau, Streamlit, Plotly
Certifications: Azure Data Engineer (DP-203), Databricks Data Engineer Associate, IBM Data Science Professional Certified

EDUCATION

University of Glasgow – MSc Data Science | Sep 2025 – Sep 2026

Key Modules: Big Data Systems, ML & AI, Deep Learning, Information Retrieval, Information Visualization, Statistics & Probability
Scholarship: £10,000 Excellence Award
Achievement: ***Runner-Up – GUTS 2025 Hackathon (Morgan Stanley Challenge***): Built Market Wars Simulator: an AI-driven investment strategy simulator with market regimes, agent-based trading models, and genetic algorithm optimization. Integrated risk metrics (Sharpe, Drawdown) and adaptive optimization for wealth-generation strategies; recognized for financial realism and technical design. **M & PI**

UPES, India – B. Tech in Computer Science Engineering with Spz. in BFSI | Jun 2018 – May 2022

Final Year Project: Digi-Attendance: Designed and developed an IoT-based facial recognition system using Python (OpenCV, face recognition, PyQt5) to automate employee attendance tracking in real time. Implemented live video feed processing and facial encoding to ensure 99% recognition accuracy, eliminating manual entry errors and improving operational efficiency.

PROFESSIONAL EXPERIENCE

Data Engineer | IQVIA, Gurgaon, India | Dec 2022 – Aug 2025
Client: GSK | Omnichannel Data Platform (Global Brand Analytics) & GSK BaaS Project

- Architected and delivered a large-scale Azure-based data platform for GSK, integrating multi-source healthcare, CRM, and digital engagement data across UK, Germany, and Italy, enabling real-time omnichannel analytics and personalized marketing insights.
- Designed and implemented the Change History & Data Quality Layer (L2) in Azure Databricks (PySpark), developing delta detection and schema validation frameworks that improved data accuracy to 99.8% and reduced processing time by 35%.
- Automated metadata-driven ingestion and validation pipelines using Azure Data Factory and SQL-based quality logging, cutting manual QA effort by 65% and accelerating new-market onboarding by 60%.
- Delivered analytical and model-ready datasets via Azure Synapse and Power BI, empowering marketing and data science teams with actionable KPIs for campaign performance, channel effectiveness, and ROI optimization.
- *Tools:* Azure Data Factory, Azure Databricks (SQL & PySpark), ADLS Gen2, Delta Lake, Azure Synapse Analytics, Power BI
- *Awards:* IQVIA Impact Award (Ovation FY23-24) and Spotlight Award (FY24-25)

Data Analyst | Axtria, Noida, India | Jan 2022 – Dec 2022
Client: AZ & J&J | J&J Omnichannel MDM & AZ O2 Project

- Designed and implemented a multi-layered data processing framework within the Azure ecosystem (ADF, Databricks, ADLS Gen2, Synapse) to ingest and integrate large-scale omnichannel healthcare datasets for Johnson & Johnson MDM and AstraZeneca Omnichannel.
- Engineered the L2 Change History / Delta Detection layer in Azure Databricks (PySpark) to identify incremental changes, enforce schema consistency, and maintain referential integrity across 10M+ records per batch.
- Developed metadata-driven data quality automation, including null validation, duplicate elimination, volume comparison, and schema drift detection — improving data reliability by 40% and reducing manual QA time by 60%.
- *Tools:* Azure Data Factory, Azure Databricks (SQL & PySpark), ADLS Gen2, Delta Lake, Azure Synapse Analytics, Power BI
- *Award:* Axtria Bravo Award (Aug – 2022)

PROJECTS

AutoPipelineAI
Designed and built an autonomous ETL and Data-Ops framework to eliminate manual data preparation and visualization bottlenecks. Developed using Streamlit, Databricks, and Llama3 (via Ollama), it automates data ingestion, transformation, and insight generation with integrated dashboards and NLP capabilities, reducing analysis time by ~70% and enabling real-time, self-serve analytics for end users.

Explainable & Fair Recommender System (Neuro-Symbolic AI Framework) *Ongoing Research Project — 2025*
Building a neuro-symbolic recommender system that combines Graph Neural Networks (GNNs) and Knowledge Graphs (KGs) to ensure fair, diverse, and user-centric recommendations. Integrating Large Language Models (GPT-4, LLaMA 3) to produce transparent, stakeholder-specific explanations grounded in model reasoning. Designing a multi-objective optimization framework to balance accuracy, fairness, and platform diversity across domains like music, recruitment, and education.

ADDITIONAL INFORMATION

Work Authorization: Currently on a UK Student Visa; eligible for full-time employment under the Graduate Route Visa upon course completion (no sponsorship required)
Languages: English (Fluent), Hindi (Native)