Manan Mehta

London, United Kingdom
+44 7448889282 | manan161200@gmail.com | LinkedIn

Profile

Data Engineer/Data Scientist with 3+ years of experience designing scalable data platforms and AI-driven document intelligence systems. Strong expertise in Python, SQL, AWS, and MLOps, with hands-on experience in ingestion pipelines, Change Data Capture, LLM-based extraction, and native cloud deployment. Currently pursuing an **M.Sc. in Data Science at King's College London**.

Experience

ZS Associates **Pune, India**
Senior Engineer *Jan 2025 – Aug 2025*

- Architected production-grade **GitLab CI/CD pipelines** integrating SonarQube, PyTest (90%+ coverage), and TruffleHog, reducing release cycle time from 5 days to 3 days (40% improvement).
- Streamlined deployment of ML extraction models and MWAA (Airflow) DAGs across 3 environments (dev, test, prod), decreasing rollback incidents by 30% and improving deployment stability.
- Designed a **multi-agent RAG-based NLP chatbot** indexing 50K+ regulatory documents, reducing manual query resolution time from 2 hours to under 45 minutes (60% reduction).

ZS Associates **Pune, India**
Business Technology Solutions Associate Consultant *Jul 2024 – Dec 2024*

- Led development of a distributed **AWS Batch on EKS digitization pipeline** orchestrated via MWAA, processing 50K+ documents and handling peak loads of 3K documents/day.
- Integrated Veeva Vault APIs and AWS Textract, storing structured outputs in Amazon RDS and OpenSearch Serverless, enabling sub-second metadata retrieval across 1M+ extracted attributes.
- Engineered LLM-based document standardization and validation workflows, improving structured data extraction accuracy by 25% and reducing manual review effort by 40%.
- Optimized infrastructure configuration (compute sizing, indexing, storage strategy), reducing cloud costs by 15% % .

ZS Associates **Pune, India**
Business Technology Solutions Associate / Intern *Jan 2022 – Jun 2024*

- Designed scalable ingestion pipelines that integrate Snowflake, Oracle DB and SFTP, consolidating 10+ data sources and ingesting daily into Amazon S3.
- Architected **Change Data Capture (CDC)** mechanisms, reducing redundant data processing by 35% and cutting ingestion runtime by 25%.
- Implemented transformation workflows converting Parquet datasets to SAS XPORT format, ensuring 100% schema consistency across downstream reporting systems.
- Delivered secure Flask APIs with PingFederate SSO authentication, supporting internal users across multiple business units.
- Increased automated test coverage to 90%+, reducing post-release production defects by 30%.

Key Projects

Prostate Cancer Detection: Developed an MRI-based cancer classification pipeline involving image denoising, segmentation, and temporal progression modeling to distinguish benign vs malignant tumors.

Text Summarization: Built an NLP application for summarizing web articles using RNN models and TextRank. Integrated BeautifulSoup-based scraping and configurable summary length.

Patent Analytics Dashboard: Developed a patent analytics platform using ELK Stack (Elasticsearch, Logstash, Kibana) with web parsing for metadata extraction and interactive trend visualization.

Education

Sept 2025 – Ongoing: **M.Sc. Data Science**, King's College London, UK
Relevant Coursework: Big Data (Hadoop/Spark/NoSQL), Statistics, Neural Networks, Computer Vision

Jul 2018 – Jun 2022: **B.E. Computer Science and Engineering**, Panjab University CGPA: 8.49/10
Relevant Coursework: Data Structures, Databases, Image Processing

Technical Skills

Languages: Python, SQL, C++, Bash

Machine Learning: scikit-learn, NLP, Embeddings, Model Evaluation, Deep Learning, Computer Vision

Data Engineering: PySpark, Airflow (MWAA), PostgreSQL, Snowflake, CDC, ETL Design

Cloud & DevOps: AWS (S3, EC2, Lambda, Batch, Textract, RDS, OpenSearch), Docker, Kubernetes (EKS), GitLab CI/CD

Visualization: CloudWatch, ELK Stack

Certifications & Languages

SQL for Data Science (Coursera)

Image Denoising Using AutoEncoders in Keras and Python(Coursera)

English (Fluent), Hindi (Fluent)