Samiksha Raut

1 Years as Data Engineer at IBM

CONTACT

8010694451

rautsamikshaa22@gmail.com

India, Pune, 900 Sadashiv peth

EDUCATION

2021 - 2025

Bachelor of Science (BS)

Yashoda Technical Campus,Satara, India, Satara

LANGUAGES

- English
- Hindi
- Marathi

PERSONAL INFO

- Date of birth: 22 March 2004
- Place of birth: Baramati
- Nationality: indian

PROFESSIONAL SUMMARY

1 Years+ as Data Engineer at IBM from Dec 2024 to Till Date

Data Engineer with 8+ Month of experience in Data Engineering, Big Data technologies and Data Analysis, . Proficient in Python, PySpark, Databricks SQL, Hadoop, Hive , Azure Cloud with expertise in building and managing scalable data pipelines using Azure Data Factory. Skilled in Data Mining, Data Preparation, Data modeling and ETL processes, ensuring high-quality data flow for analytical and business needs. Experienced in handling large datasets, implementing Machine Learning algorithms, and working with cloud platforms for data storage and processing. Strong knowledge of Proof of Concepts (PoC) and gap analysis to drive data-driven solutions for enterprise applications.

EXPERIENCE

Data Engineer 2024 - Now

Expeditors, United Kingdom

Shipping Data Pipeline – Cargo Capacity Planning

Domain: Shipping & Logistics

Tech Stack: Azure Data Factory, Azure Databricks (PySpark), Azure Logic Apps, Azure Blob Storage, Azure Data Lake Gen2, Unity Catalog, Python, SQL, ETL

Summary:

Designed and implemented a scalable, automated data pipeline on Azure for cargo capacity planning. The pipeline ingested shipment forecast and bulk files from multiple external sources, applied data validation, schema alignment, and CBM (Cubic Meter) calculations, enriched records with master references (port codes, stacking rules, product mappings), and delivered curated datasets for planning and reporting. The solution provided end-to-end automation, governance, and performance optimization, enabling logistics teams to make faster, more accurate capacity planning decisions.

Key Achievements:

- Reduced processing time by ~75%, cutting daily effort from 4–5 hours to less than 1 hour.
- Improved forecast accuracy by 20–25% through standardized schema and CBM calculations.
- Enabled incremental processing, reducing redundant runs and saving 30–40% compute costs.
- Built an audit and error-handling framework with logging for 100+ daily files, reducing troubleshooting time by ~60%.
- Optimized storage with partitioned, organized data, improving query response times by 40%.
- Enforced governance and role-based access with Unity Catalog, ensuring 100% compliance.
-

Achieved scalability by onboarding 5–10 new clients monthly without major redesign.

- Reduced cloud costs by ~25% using autoscaling clusters and efficient file formats (Parquet/Delta).
- Improved cross-team collaboration by documenting business rules, reducing requirement clarification cycles by ~50%.
- Designed the pipeline to handle over 1 TB of daily data load with 99.9% uptime using Spark tuning and partitioning.
- Reduced Power BI/Synapse report refresh times by ~30%, enabling faster and more reliable decision-making.

SKILLS

Python
SQL
Big Data
Data Warehousing
Data Mining
ETL
Apache Hadoop
Apache Spark
NoSQL
Data Modeling
Cloud Computing
Data Visualization
Machine Learning
Data Analysis
Business Intelligence
AWS
Azure Data Factory
Pyspark
Data bricks