

Umit Samanta

+91-6290896297 | hit18bt.umitsamanta@gmail.com | linkedin.com/in/umitsamanta103 | github.com/umitsam14

PROFILE

Results-driven **Data Engineer** with expertise in building and optimizing scalable data infrastructure across hybrid cloud architectures. Skilled in **ETL/ELT pipelines**, **data quality assurance**, **observability**, and **AI/ML integration**. Adept at ensuring **data governance**, driving visualization adoption, and implementing innovative solutions for business transformation.

TECHNICAL SKILLS

| | |
|-------------------------|--|
| Languages: | Python, SQL |
| Big Data & Streaming: | Spark, PySpark, Hadoop, Kafka |
| Cloud Platforms: | AWS (S3, Lambda, Glue, EMR, Redshift, DynamoDB, Kinesis), Azure (Data Factory, Blob Storage, Synapse, Event Hubs, Azure Functions), GCP (Basics) |
| Databases: | MySQL, PostgreSQL, MongoDB |
| Data Engineering: | ETL/ELT, Data Modeling, Delta Lake |
| DevOps & Observability: | Docker, Kubernetes, Jenkins, GitHub Actions, Terraform, Prometheus, Grafana |
| Orchestration: | Apache Airflow |
| Tools: | FastAPI, Streamlit, Power BI |

EXPERIENCE

BT Group

Mar 2023 – Present

Associate – Service Operations — Kolkata, India

- Automated **ETL workflows** using AWS Glue and Lambda integrated with Redshift.
- Reduced **report generation time by 40%** through pipeline optimization.
- Designed dashboards using **Power BI**, enabling data-driven business decisions.
- Collaborated on **observability improvements** to optimize performance and data reliability.

Rinex Technology Pvt. Ltd.

Jul 2022 – Sep 2022

Data Engineering Intern — Mangalore, India

- Optimized AWS Glue ETL jobs and Redshift queries, reducing ingestion latency by **50%**.
- Enhanced batch ETL processes for large-scale analytics workflows.

Flipkart

Sep 2021 – Oct 2021

Data Analytics Intern — Howrah, India

- Built analytics pipelines with SQL and Python for real-time processing using AWS Kinesis.
- Reduced data stream latency by **45%** through pipeline optimization.

PROJECTS

AWS Incremental ETL Pipeline (Serverless)

[GitHub]

May 2025 – June 2025

- Built a fully serverless ETL pipeline using **AWS Lambda**, **S3**, **Glue Crawler**, and **Athena**.
- Parsed nested JSON to Parquet using Pandas and automated schema detection with Glue.
- Use case:** Real-time ingestion of incremental e-commerce order data with schema evolution support.

Drug Recommendation System (ML + Data Engineering)

[GitHub]

Jan 2025 – Apr 2025

- Built ML pipeline using **Pandas**, **Scikit-learn**, and **MySQL** for structured healthcare data.
- Developed backend API using **FastAPI** and desktop UI using **Tkinter GUI**.
- Use case:** Predicts optimal drugs for patient conditions using input symptom data.

Real-Time Retail Analytics Pipeline on Azure

[GitHub]

Mar 2025 – May 2025

- Built a real-time data pipeline using **Azure Event Hubs**, **Azure Stream Analytics**, and **Blob Storage**.
- Integrated with **Azure Data Factory** for batch transformations and loaded data into **Azure Synapse Analytics**.
- Created Power BI dashboards to visualize key retail KPIs from streaming and batch datasets.
- Use case:** End-to-end Azure solution for e-commerce clickstream + sales analytics.

CERTIFICATIONS

AWS Certified Cloud Practitioner (CLF-C02)

[\[View Certificate\]](#)

Issued by Udemy — June 2024

ITIL V4 Foundation Certification

[\[View Certificate\]](#)

Issued by PeopleCert — May 2023

EDUCATION

M.Tech in Biotechnology – Haldia Institute of Technology

Aug 2023 – June 2025

B.Tech in Biotechnology – Haldia Institute of Technology

Aug 2018 – Jul 2022

ACHIEVEMENTS

- Best Shift Lead Award – BT Group (Oct 2024)
- “Just To Say” Award – BT Group
- **HackerRank Gold Badge** in Python and SQL
- Solved 100+ LeetCode problems on DSA and SQL
- Recognized for excellence in painting & performing arts