Umit Samanta

+91-6290896297 | hit18bt.umitsamanta@gmail.com

https://www.linkedin.com/in/umitsamanta103

PROFILE

Results-driven Data Engineer with expertise in building, optimizing, and scaling data infrastructure. Skilled in cloud platforms, distributed systems, ETL pipelines, and integrating machine learning models. Passionate about solving complex data challenges and ensuring data reliability.

TECHNICAL SKILLS

Programming Languages:	Python, SQL
Big Data Technologies:	Spark, PySpark, YARN, Kubernetes, Hadoop
Streaming Technology:	Kafka, Spark structured streaming
Cloud computing:	AWS (S3, Lambda, EC2, Redshift, DynamoDB, Glue, RDS, SQS, Kinesis, EMR), Snowflake
Databases:	PostgreSQL, MySQL, Redshift, DynamoDB, MongoDB
Data Engineering Tools:	Data Modelling, ETL/ELT data Pipeline
DevOps:	Git, Terraform, Docker, Kubernetes, CI/CD (Jenkins, GitHub Actions)
Orchestration:	Airflow, Kubernetes
Familiar:	FastAPI, Elastic search

EXPERIENCE

BT Group — March 2023 – Present
Associate – Service Operations — *Kolkata, India*

- Developed and automated data pipelines using AWS services such as Glue, Lambda, and Redshift, reducing data processing time by 40% and improving data availability across systems.
- Built scalable and efficient ETL processes to transform raw customer data into structured formats, leading to 30% faster decision-making through real-time reporting and dashboards.
- Collaborated with cross-functional teams to gather business requirements and developed custom data solutions that resulted in a 20% increase in operational efficiency.

Rinex Technology Pvt. Ltd. — Jul 2022 – Sep 2022
Data Engineering Intern — *Mangalore, India*

- Developed ETL workflows with AWS Glue, reducing batch processing times by 50%.
- Optimized Redshift query execution, improving data retrieval speed by 40%.

Flipkart — Sep 2021 – Oct 2021
Data Analytics Intern — *Howrah, India*

- Built data analytics pipelines using SQL, Pandas, and AWS Lambda.
- Developed real-time ingestion workflows with AWS Kinesis, reducing latency by 45%.

EDUCATION

Haldia Institute of Technology — Haldia
Masters of Technology in Biotechnology — *Aug 2023 – Present*

Haldia Institute of Technology — Haldia
Bachelor of Technology in Biotechnology — *Aug 2018 – Jul 2022*

PROJECTS

Scalable Drug Discovery Data Pipeline — **Jan 2025 – Present**
Apache Airflow, AWS, Python, FastAPI, SQL

- **Issue:** Traditional drug discovery processes are time-consuming and computationally expensive.

- **Solution:** Designed an automated data pipeline using Apache Airflow to streamline data ingestion, transformation, and model execution.

- Integrated ML models with FastAPI for real-time prediction of drug-target interactions.

- Optimized data storage using AWS Redshift, DynamoDB, and S3, reducing query latency by 40%.

Distributed Batch Processing with Apache Spark　　　　　　　　　**Feb 2024 – Mar 2024**

- **Issue:** Inefficient batch processing led to delays in data availability.

- **Solution:** Developed scalable batch processing workflows with Apache Spark on AWS EMR to handle large-scale data processing efficiently.

- Leveraged Delta Lake for ACID-compliant ETL workflows, ensuring data integrity.

CERTIFICATIONS

- AWS Cloud Practitioner (CLF-02) - Udemy
  Certification Link

- ITIL V4 Foundation – Peoplecert
  Certification Link

ACHIEVEMENTS

- Best Shift Lead Award – Oct 2024

- "Just To Say" Award – BT Group

- Recognized for excellence in painting

- 5th star level in Singing and Painting