

Sohag Ahammed Siyam

Certified Data Engineer - Databricks

Location: Espoo, Finland, Mobile: +358417216912

sohag35-292@diu.edu.bd | [linkedin.com/in/espsiyam/](https://www.linkedin.com/in/espsiyam/) | github.com/espSiyam

Experience

Data Engineer - Trainee, [KONE](#) – Finland

May 2024 – Present

- Automated query and alert using **Databricks Python SDK** and **Asset Bundle**.
- Data Timeliness** checks using **DBT SQL**, integrated with a CI/CD pipeline and **Databricks workflow**.
- Explored Databricks' **Data Monitoring**, internal mechanism and its integration.

Junior Backend Engineer - Data and AI, [Elelem](#) – Canada (Remote)

Dec 2022 – April 2024

- Implement custom **Vector Search** with Embedding and Clustering.
- Web Automation with **LLM (GPT models)**.
- Achieved a 10x reduction in **Bigquery** query cost through partitioning
- Developed, Deployed, and Managed different API (**FastAPI**),
- Configured and Modified **CI/CD pipeline**, Cloud Run, Logging, and Cloud Function in GCP.

Junior ML Engineer, [Fashol Dotcom Limited](#) – Dhaka, Bangladesh

April 2022 – Dec 2022

- Developed Machine Learning Models for **forecasting** Product Prices and Sales Quantity on **time series data**.
- Extracted data from various online sources (using Selenium, Scrapy).
- Established an automated pipeline with **ETL (Extract, Transform, Load)** for Extracting (PyMySQL), Cleaning (Python), Loading and finally Visualization (Data Studio).

Projects

On premises DB to Azure Analytical Pipeline

[Github Project Link](#)

- Linked and imported SQL database tables into **Azure Data Factory** using **Pipelines**.
- Managed **Medallion architecture (Bronze, Silver, Gold)** with Databricks to maintain the lakehouse architecture.
- Established a **Data Warehouse** in **Synapse Analytics**.
- Connected and visualized data in **PowerBI** for analytical purposes.

AI News Generator Pipeline

[Github Project Link](#)

- Developed ETL in a **dockerized Apache Airflow** to collect and store data in a GCP bucket, deployed on a GCP CE.
- Implemented a Cloud Function **trigger** for data insertion into **BigQuery** upon ETL saving data to the bucket.
- Created a pipeline to create news from the gathered data using the **OpenAI** API and creating the necessary schema.
- Done [Ad-hoc analysis](#) on **Apache Spark** (Databricks.)

Skills

Software Engineering: Python, Docker, API (Flask, Fast API), Micro-services, Unit-Test

Database: SQL, MongoDB, Data Modeling, Data Warehousing

DE tools: DBT, Apache Airflow, Apache Spark, Databricks, BigQuery

Cloud: GCP (Cloud Function, Cloud Build, BigQuery, Pub/Sub, Composer)

Azure (Synapse Analytics, Data Factory, Databricks, Logic App)

Learning AWS (Glue, Redshift, Elastic MapReduce)

Version Control: Git, Github Action.

Educations

[Jyväskylä University](#) – Master's in COIN (Artificial Intelligence)

Aug 2023 - Aug 2025

[Daffodil International University](#) – B.Sc. in Software Engineering

Jan 2018 -Jan 2022