Md Ibrahim Khalil

+880 179 843 9302 — bsse1009@iit.du.ac.bd — linkedin.com/in/ibrahim-khalil — github.com/m-ibrahim-khalil

Summary — Software Engineer with 4+ years of experience architecting scalable ETL pipelines and MLOps infrastructure. Expert in building distributed data systems using Apache Airflow, Kafka, and Spark within Azure and Kubernetes environments. Proven track record of automating complex data workflows, reducing manual operational load, and deploying production-grade ML models.

Technical Skills

Data Engineering Airflow, Kafka, Mage, PySpark, InfluxDB, PostgreSQL, Playwright

Cloud & MLOps AWS, Azure ML Studio, MLflow, Kubernetes, Docker, CI/CD (GitHub Actions)

Core Languages Python, SQL, Bash, Go, TypeScript

Web & Backend FastAPI, Django, Node.js, React

Methodologies Agentic AI, TDD, Microservices, DataOps, LLM

Experience

Cefalo Bangladesh Ltd. **Jan 2023 – Present**
Software Engineer (Data & ML Engineer) *Dhaka, Bangladesh*

Kontali (Aquaculture Analytics) – *High Volume ETL*

- Architected scalable ETL/ELT pipelines using **Python, Airflow, and Kafka**, handling high-velocity data streams for a major Norwegian analytics provider.
- Optimized data ingestion and transformation workflows, significantly improving the latency and reliability of the data lake.
- Implemented CI/CD pipelines to automate testing and deployment of data DAGs, ensuring production stability.

Sensa (DataOps) – *MLOps Infrastructure*

- Designed the data architecture for a multi-tenant DataOps service, utilizing **InfluxDB** for time-series data and **PostgreSQL** for metadata.
- Built an end-to-end MLOps pipeline using **Azure ML Studio and MLflow**, automating model training, versioning, and deployment for anomaly detection.
- Engineered algorithms to detect ”frozen sensors” and data anomalies in real-time, improving data quality assurance.

Wirescan (Energy) – *Automation*

- Engineered an automated data ingestion system that extracts unstructured data from PDF, CSV, and Excel formats, reducing manual data entry load by **90%**.
- Developed visualization dashboards to render cable health indicators, enabling data-driven maintenance decisions.

Brain Station 23 **Jan 2022 – Jan 2023**
Associate Software Engineer *Dhaka, Bangladesh*

- **Unstructured Data ETL:** Developed a pipeline using Google Document AI to extract structured tabular data from medical diagnostic reports.
- Designed microservices using Python and PostgreSQL to serve NLP-derived insights to frontend applications.
- Collaborated with the R&D team to optimize ML model inference times for production use cases.

Education

University of Dhaka **Jan 2018 – Dec 2022**
Bachelor of Science in Software Engineering *CGPA: 3.63*

Technical Projects

LiDAR PointCloud Classification (3D Data Processing) **Code**

- Developed a processing pipeline to classify and segment 3D LiDAR point cloud data using **PyTorch** and **Open3D**.
- Implemented deep learning techniques to identify objects within interior spatial data.

Distributed System for Ride Sharing **Code**

- Designed a microservices architecture ensuring eventual consistency and high availability using **Docker, Nginx, and MongoDB**.

Deep Learning QA System (NLP) **Code**

- Built a context-aware Question Answering system using a BiDAF model trained on SQuAD v2.0, demonstrating proficiency in handling unstructured text data.