

KEVIN JOJI MATHEWS

Data Engineer

☎ 07453879484 ✉ kevinjmathewskj@gmail.com 🔗 linkedin.com/in/kevin-jmathews 🌐 github.com/KevinJM17

EDUCATION

MSc Big Data Science, Queen Mary University of London, London Sept 2023 - Sept 2024
Relevant modules: Data Mining, Cloud Computing, Machine Learning, Big Data Processing, Neural Networks and Deep Learning.

B.Tech in Electronics and Communication, National Institute of Technology, Calicut Jun 2018 - Sept 2022
Relevant modules: Artificial Intelligence, Data Structures using C++, ANN and Fuzzy Logic Systems.

SKILLS

Technical Skills	Python, Java, SQL, Postgres, Git, Bash, Docker, Terraform, Airflow
Data Engineering	ETL Pipelines, Data Warehousing, API Integration
Cloud	Azure (Azure Data Factory, Databricks, Data Lake), Google BigQuery
Data Analysis	Numpy, Pandas, Matplotlib, Seaborn, Tensorflow, PyTorch, PySpark
Soft Skills	Communication, Responsible, Team Collaboration, Problem-Solving, Analytical

EXPERIENCE

Network Data Engineer Intern Dec 2021 - Jan 2022
GBM Qatar *Doha, Qatar*

- Designed and optimized **ETL pipelines** to process and analyze network performance data from **50+** access points across the Qatar Metro Rail, leading to a **20%** improvement in Wi-Fi performance.
- Developed a user-friendly web application using **Python and Django** for project management, enhancing project tracking and collaboration efficiency by **40%**.
- Partnered with a cross-functional team to implement **data ingestion workflows** using **Google BigQuery**, automating Wi-Fi module performance analysis for the FIFA World Cup 2022 stadium.

PROJECTS

Reddit ETL Pipeline | Python, Airflow, Docker, PySpark, Databricks, Azure Data Factory

- Developed an end-to-end **ETL pipeline** to process Reddit API data using **Python, Docker, and Airflow**.
- Orchestrated data workflows with **Airflow**, automating data ingestion and transformation tasks.
- Implemented **Azure Blob Storage** for raw data storage and **Databricks** for data processing using **PySpark**.

Text Extractor | Python, Flask, GCP, Google BigQuery, Google Cloud Storage

- Designed a full-stack application with user authentication, image upload, and text extraction using **Flask and Google Cloud Vision API**.
- Integrated **Google Cloud Storage and Google BigQuery** for efficient data management, reducing manual processing time by **75%**.

Covid Pneumonia Detection from Chest X-Rays | Python, Matplotlib, Numpy, Tensorflow

- Developed and fine-tuned **image classification models** (VGG19 and custom models) to detect Covid-induced pneumonia, achieving an accuracy of **82%**.
- Implemented pooling methods to optimize model training and reduce computation time by **30%**.

PUBLICATION

Wang, H., Mathews, K.J., Golec, M., Gill, S.S. and Uhlig, S., 2025. "AmazonAICloud: proactive resource allocation using amazon chronos based time series model for sustainable cloud computing". Computing, 107(3), p.77.