Robert Lewis

Email: contact@robertlewis.dev

 Portfolio   GitHub   LinkedIn

Cell: +1 (818) 730-0167

SUMMARY

Motivated Data Engineering intern candidate pursuing an M.S. in Computer Science at Georgia Tech. Proficient in **Python**, **SQL**, cloud platforms (**AWS, Azure, GCP**) and big data frameworks (**Spark, Hadoop, Kafka**). Demonstrated experience in building scalable pipelines through hands-on projects.

EDUCATION

Georgia Institute of Technology (Georgia Tech) **August 2024 - Present**
Master of Science in Computer Science *Atlanta, GA*
- **Coursework**: Advanced Database Concepts & Design, System Design for Cloud Computing, Data Analytics & Security, Computer Networks, Artificial Intelligence, Machine Learning for Trading

University of California, Los Angeles (UCLA) **September 2019 - June 2023**
Bachelor of Science in Statistics and Data Science *Los Angeles, CA*
- **Relevant Coursework**: Probability Theory, Deep Learning, Neural Networks, Bayesian Statistics, Computer Vision, Data Warehousing, Data Mining, Data Visualization, Data Structures & Algorithms

TECHNICAL SKILLS

Programming & Scripting: Python, Go, SQL, JavaScript, TypeScript, R, Bash

Cloud & MLOps: AWS, Azure, GCP, Docker, Kubernetes, Terraform, Airflow, CI/CD

Data Engineering & Pipelines: Kafka, Hadoop, Spark, Hive, dbt, Apache Beam

Databases & Storage: MySQL, PostgreSQL, MongoDB, Redis, S3, Cassandra

PROJECTS

Real-Time Audio Fingerprinting Engine — *Golang, MongoDB, FFmpeg, REST APIs*
- Created a **Golang** pipeline to extract audio fingerprints for real-time song IDs.
- Integrated **Spotify/YouTube APIs** and **FFmpeg** for metadata retrieval and audio preprocessing.
- Enabled scalable matching with **MongoDB** and **SQLite** for efficient fingerprint storage and retrieval.

AI-Driven Document Summarization Suite — *Python, LangChain, Pinecone, Docker, AWS*
- Streamlined retrieval of **millions** of documents (e.g., reports, articles) to speed up decision-making.
- Leveraged **LangChain** for LLM prompts and **Pinecone** for indexing; containerized with **Docker**.
- Deployed on **AWS Fargate** for serverless scaling, cutting latency by **30%**.

Automated MLOps Pipeline for Image Analytics — *Kubeflow, Kubernetes, Terraform, ETL*
- Developed an **MLOps** pipeline with **Kubeflow** for ETL in image classification tasks.
- Managed **Kubernetes clusters** via **Terraform** to enhance scalability and resource efficiency.
- Automated container orchestration with **ETL** workflows, minimizing manual oversight by **40%**.

CERTIFICATIONS

CompTIA Network+

CompTIA Security+