Aniket Majety

+1 404-457-2790 | aniketmaj21@gmail.com | linkedin.com/amajety | github.com/amajety1

EXPERIENCE

Software Engineer May 2025 – August 2025
HomeReelz *Remote*
- Created a mobile app frontend using **React Native** and **Typescript** to deliver a real estate content creation platform for real estate agents.
- Integrated **Firestore** for live data sync and used **React Query** to reduce database lookups through cache-first patterns.
- Deployed backend to **Docker** with **CI/CD pipelines**, ensuring consistency across development and production environments
- Used **SQL (Postgres)** database to ensure consistent data relations

Data Engineer Intern January 2025 – May 2025
InpharmD *Remote*
- Engineered scalable data ingestion workflows using **Python**, **Snowflake**, and **SQL Server Integration Services (SSIS)** to extract and transform clinical literature from PubMed and Google Scholar.
- Designed multi-threaded ETL pipelines incorporating **TF-IDF ranking**, **Apache Airflow**, **Databricks**, and **SSIS** orchestration with **PostgreSQL** storage to preprocess biomedical text data.
- Developed and deployed fine-tuned **LLM pipelines** using **Transformers**, **PyTorch**, and structured biomedical datasets stored in **AWS S3**.
- Implemented data partitioning and indexing strategies on biomedical datasets, reducing query scan times by 60%.
- Designed a reproducible **CI/CD workflow** with **GitHub Actions** and **Docker** for seamless deployment of data pipelines.

Data Engineer Co-op January 2024 – December 2024
Oakwin
- Designed ETL pipelines using **Apache Airflow**, **Apache Spark** on **Databricks**, **Python**, and **Docker** to orchestrate and automate daily data ingestion workflows.
- Built data pipelines integrating **REST APIs**, **pandas**/**Apache Spark** for data transformation, **AWS S3** for storage, and **PostgreSQL/Redis** for workflow orchestration.
- Automated ETL preprocessing of raw data from diverse formats (**databases**, **flat files**, **CSV**, **Excel**), enabling ingestion into distributed data pipelines.
- Implemented a cloud-native data analytics platform using **AWS Glue** with **Apache Spark** for ETL transformations, **Amazon Athena** for ad-hoc SQL queries, and **Amazon Redshift**/**Snowflake** for scalable analytics.
- Developed automation scripts with **Shell** and **AWS CLI** to validate ETL job outputs, monitor logs, and integrate alerts with **CloudWatch** and **Grafana** dashboards to ensure SLA compliance.
- Optimized complex **SQL queries** and large-scale **Spark SQL** transformations using **CTEs** and execution plan tuning.
- Designed **data modeling schemas** (star/snowflake) to support OLAP analytics and downstream BI dashboards.
- Worked within Agile teams to manage deliverables, deploying containerized applications with **Docker Compose** for local development and production environments.

TECHNICAL SKILLS

Languages: Python, SQL
Data Engineering: ETL Pipelines, Data Modeling, Apache Airflow, Apache Spark, Databricks, AWS Glue, pandas, numpy, SSIS
Cloud Data Services: Amazon S3, Amazon Athena, Amazon Redshift, Snowflake, AWS EC2
Orchestration & Workflow: Apache Airflow, Docker, Docker Compose
Databases: PostgreSQL, Redis, SQL Server
APIs & Integration: REST APIs, AWS SDK, curl
DevOps Tools: Docker, Git, AWS CLI

CERTIFICATIONS

AWS Certified Cloud Practitioner (CCP) View Certificate

EDUCATION

Georgia State University Atlanta, GA
Bachelor of Science in Computer Science *Expected May 2025*
- GPA: 3.7/4.0
- President's List (Spring 2024, Summer 2024), Dean's List (Fall 2024)