

Anthony Lam

Senior Data Engineer @ Mastek

Email: so3-w94-vfg@mail.dice.com

Phone: 3176949941

Summary

Previous	Senior Data Engineer @ Olive
Preferred	Senior Data Engineer
Location	Indianapolis, IN, US
Desired Work Settings	Remote or Hybrid
Willing to Relocate	No
Work Authorization(s):	Authorized to work in the United States on a full-time basis.
Security Clearance Info:	US Citizen (for security purposes)
Employment Type	Full-time Contract - W2 Contract to Hire - Corp-to-Corp Contract to Hire - Independent Contract to Hire - W2
Total Experience	17 years
Education	Masters @ Georgia Institute of Technology Masters @ Miami University Bachelors @ University of Indianapolis
Profile Source	Dice
Profile Downloaded	Monday, December 9, 2024

Anthony Lam

[|| linkedin.com/in/lamah](https://www.linkedin.com/in/lamah)

Skills

Languages: Python, SQL, R, SAS

Tools: Databricks (PySpark, Delta Live Tables, Unity Catalog), Azure (Data Factory, Functions, Storage Container, DevOps), Apache Airflow, AWS (Glue, Lambdas, S3, Athena, Redshift), DBT (Core, Cloud), Docker, Kubernetes, GitHub, GitLab

Experience

Senior Data Engineer

Mastek, May 2023 - Nov 2024

- Data Pipeline Development: Designed and maintained ELT pipelines supporting production processes and machine learning models across bronze, silver, and gold layers using Azure Data Factory, Databricks (PySpark, Python, Delta Live Tables, Unity Catalog), DBT Cloud, and Azure Data Lake Storage.
- Advanced Feature Engineering: Extracted key insights from raw text blobs through named entity recognition (NER) by leveraging prompt engineering with ChatGPT, OpenAI, Regex, and Databricks.
- Automation & Monitoring: Created monitoring pipelines to auto-submit ServiceNow tickets via APIs using Azure Functions, Azure Data Factory, GitHub, and Python.
- Real-Time Data Integration: Extended pipelines to process and visualize data from Azure Service Bus for dashboards using Azure Functions, Python, and Scala.
- ETL for Transactional Data: Built pipelines to aggregate client SQL Server transactional data into AWS Redshift using AWS DMS, Apache Airflow, and S3.
- Scalable Architecture: Engineered dynamic Airflow tasks to manage incremental loads and automatic ingestion of over 200 tables.
- CI/CD Implementation: Led CI/CD pipeline development with GitHub Workflows to sync repositories to S3, ensuring version control for Airflow DAGs.
- Client Solutions: Developed product demos showcasing Airflow, AWS (S3, MWAA, RDS), and Snowflake capabilities to potential clients.

Senior Data Engineer

Olive, Apr 2019 - Mar 2022

- Dynamic Data Processing: Designed and implemented ELT pipelines to process 90M+ records and 100+ JSON variations using AWS Athena, Airflow, Python, DBT Core, and TypeScript, storing processed data in AWS Redshift.
- Custom Integrations: Developed Airflow operators and hooks to interact with Olive's internal APIs and external APIs such as Workfront, Databricks, Google Sheets, and Salesforce.
- Database Optimization: Integrated DBT Core with Airflow to automate table dependency tracking, testing, and documentation generation across the database.
- Bot Monitoring & Automation: Built a data pipeline to monitor 100+ bots using JavaScript and AWS (Redshift, Athena) and developed a web application for anomaly detection and alerting using Python and AWS Lambda.
- Operational Support: Monitored and maintained 20+ Airflow DAGs critical for updating a database used by multiple departments.

Market Research Statistician

Spencer Research Inc., Feb 2018 - Dec 2018

- Comprehensive Analytics Tool: Developed a robust analysis program in R to streamline and centralize the company's analytical processes, improving efficiency and consistency.
- Client-Focused Reporting: Produced detailed analysis reports for diverse product tests, delivering actionable insights tailored to client needs.
- Advanced Statistical Techniques: Designed and implemented a new procedure for applying factor analysis to survey data, enhancing the accuracy and depth of survey insights.

Education

Master of Science, Statistics

Miami University, 05/2016

Courses: Regression Analysis, Experimental Design Methods, General Linear Models, Bayesian Statistics, Analysis of Forecasting Systems, Statistical Learning, Multivariate Statistical Analysis

Projects

Automated Day Trader

Stock Market, Mar 2022 - May 2023

- Used Python to call IBKR API to stream live stock data from 50 symbols
- Used a local PostgreSQL
- Searched through over 400 different stocks and their option prices
- Found likely profitable opportunities to sell puts using time series analysis