

JIARU LIU

Phone: (314)-229-5314 | Email: liujiaru97@gmail.com | LinkedIn: [linkedin.com/in/claire-jiaru-liu](https://www.linkedin.com/in/claire-jiaru-liu)

PROFESSIONAL SKILLS

Programming Languages: Python, SQL, PySpark, Bash

Cloud Platforms: AWS, Google Cloud Platform, Databricks, Snowflake

Tools & Databases: Git, Docker, Terraform, Airflow, DBT, Tableau, Fivetran, MongoDB, AWS RDS, AWS Redshift

Technical Skills: Data Preparation, Data Analytics, Data Modeling, Data Visualization, CI/CD, MLOps

PROFESSIONAL EXPERIENCE

GOGUARDIAN, Los Angeles, CA

April 2021 - Present

Data Engineer II

February 2023 – Present

- **Cloud & Lakehouse Optimization:** Led migration of AWS ETL pipelines to Databricks, supporting stakeholder transition from Redshift to the Databricks Lakehouse architecture, resulting in ~\$400K annual cost savings.
- **Streaming Event Data:** Designed and implemented real-time application event ingestion pipelines using Kinesis Firehose, S3, Auto Loader, and Delta tables, processing 60M–800M records daily and reducing latency by 90–95%.
- **MongoDB CDC Ingestion:** Re-architected MongoDB CDC ingestion from Kafka-based streams to Spark Structured Streaming with the MongoDB connector, supporting reliable batch and streaming downstream processing.
- **Reverse ETL & APIs:** Built reverse ETL pipelines for Marketing and Sales teams, integrating with HubSpot REST APIs and Salesforce Bulk APIs to enable reliable, scalable data syncs and power automated business workflows.
- **Data Modeling & Quality:** Designed and optimized dbt-powered Medallion architecture data models and semantic layers for key Tableau dashboards, standardizing metrics, reducing duplicate reporting by 40%, improving data accuracy, and implementing data quality and freshness tests.
- **Data Observability:** Developed a stakeholder-facing observability layer by publishing dbt data catalog and Elementary reports through an AWS CloudFront website, and automated Tableau dashboard refreshes once data landed in the lakehouse.
- **MLOps:** Collaborated with the Data Science team to productionize 5+ ML/LLM pipelines, including model load testing, model deployment, and CI/CD integration, improving model reliability and release efficiency.
- **Data Privacy:** Implemented end-to-end PII data handling and governance, ensuring sensitive data was properly classified, tagged, and secured prior to ingestion into analytics environments using Unity Catalog for fine-grained access control.
- **Technical Leadership:** Led cross-team initiatives, mentored junior engineers, and drove best practices to improve data platform reliability, scalability, and team productivity.

Data Engineer I

April 2021 – February 2023

- **ETL Development:** Built ETL pipelines ingesting data from 30+ sources, developing AWS Batch and Glue jobs for transformation and loading data into S3 data lake and Redshift data warehouse to support analytics.
- **Task Orchestration:** Set up Airflow, built custom operators and hooks, and wrote DAGs to schedule and manage jobs across various tools, enhancing workflow automation and efficiency.
- **Infrastructure as Code:** Developed and maintained Terraform scripts for provisioning and managing AWS resources across multiple environments, ensuring consistency and scalability.
- **Analytics & Reporting:** Designed and built data model for customer usage report enabling CSMs to track engagement, identify upsell opportunities, support renewals, reducing query time by 75–80%.

KROLL BOND RATING AGENCY, New York, NY

October 2020 - January 2021

Data Engineer Intern

- **Data Preparation and Training for ML:** Built convolutional neural network models (Jupyterlab, Tensorflow, Pandas) to detect time series anomalies, processing over 50+ GiB of data for training and testing.
- **CI/CD Development & Testing:** Wrote installable Python packages with GitLab CI/CD, including running unit tests, doing static code check, and publishing the packages to package manager.
- **Containerization & Infrastructure as Code:** Created Docker images to containerize the Python packages using remote pipeline triggers and deployed the containers in multiple Terraformed environments.
- **Technical Presentation:** Delivered a presentation to the technology department, showcasing innovative data validation approaches, engaging an audience of over 50 professionals, and received recognition from the leadership team for

pioneering complex engineering solutions.

REGATTA CRAFT MIXERS, New York, NY

June 2020 - July 2020

Student Consultant

- **Industry Research and Analysis:** Conducted industry research and identified industry trends and 8 major competitors in the mixed drink market.
- **Data Wrangling and Cleaning:** Compiled and cleaned one-year Facebook social media data using Python.
- **Customer Behavior Analysis:** Analyzed customer behavior and preferences by visualizing data from an online survey and social media in Tableau.
- **Strategic Recommendations:** Summarized key takeaways based on analysis to enter the grocery store market as a new brand, including customers portfolio analysis, competitor analysis, and the tiered marketing and sales strategy.

EMERSON, Saint Louis, MO

January 2020 - May 2020

Student Consultant

- **Data Utilization Assessment:** Assessed Emerson's data utilization across different business units by conducting interviews with Marketing and IT leaders.
- **KPI Standardization:** Standardized and prioritized key marketing performance indicators (KPIs) across Emerson's Commercial and Residential Solution Business (\$6+ billion business unit).
- **Process Efficiency Optimization:** Reduced work process time on regular data check by an estimation of 20% by designing frameworks to identify data gaps and data inputs for KPIs.
- **Strategic Recommendations & Plans:** Designed an actionable solution implementation plan to quantify deliverables for Emerson's marketing team.

EDUCATION

WASHINGTON UNIVERSITY, OLIN BUSINESS SCHOOL, Saint Louis, MO

January 2021

Master of Science in Business Analytics

DALIAN UNIVERSITY OF TECHNOLOGY, Dalian, China

June 2019

Bachelor of Management in E-commerce