



FRANK LÜ

Data Infrastructure Engineer • Software Engineer

[in](#) frankcholula [G](#) frankcholula [G](#) frankcholula.notion.site
[@](#) tsufanglu@email.com [📍](#) Taiwan | USA | Remote

DATA TOOLS

Data Science and Analytics

Snowflake     
Numpy|Pandas     
DBT     
PySpark     

Data Engineering

Dagster     
Airflow     
Debezium     
Kafka     

Data Infrastructure

K8S|Helm     
AWS     
Datadog     

PROGRAMMING

Frontend

React.js     

Backend

Python     
Node.js     
Java     

Database | Infrastructure

SQL     
Terraform     

LANGUAGES

English: Native



Chinese: Native

ABOUT ME

Hello! I empower data scientists and data analysts by building scalable data infrastructures and data tools. My experience spans real estate, healthcare, and logistics. I'm interested in green tech/climate change-related jobs, preferably fully remote or hybrid.

EDUCATION



Electrical Engineering and Computer Science | UC Berkeley

 Jun 2012 – Jun 2016  CA, USA

- Minor in Mechanical Engineering



EXPERIENCE

Senior Data Platform Engineer • Infra Engineer | Flexport

 Jun 2019 – Jan 2023  CA, USA

- Implemented Kafka on Kubernetes with Strimzi and cluster re-balancing using Cruise Control, migrated Airflow DAGs to Dagster for data pipeline orchestration, and established a self-service software catalog for microservices using Backstage with full-fledged CI/CD and infrastructure provisioning in Github Actions.
- Deployed 3 automated shipping assignment and consolidation models for ocean, air, and trucking using the FICO Xpress Optimization tool and turned the respective models into services.
- Designed and implemented Flexport's Data Mesh vision by leading a team of 3 engineers to modernize the analytics stack with Snowflake, DBT, and Looker.

Software Engineer II • Data Scientist I | Virta Health

 May 2018 – Apr 2019  CA, USA

- Standardized data science model-deployment workflow and hyper-parameter tuning with Amazon SageMaker.
- Created and optimized two ML models. One model tracks patient retention for health coaches and the other one predicts a patient's weight given a patient's A1C level.
- Worked with the clinician experience team to design and create a supervised learning DKA(Diabetic Ketoacidosis) model, with a 75 percent accuracy rate of predicting patient with the complication.