Ganyu Xu

Software engineer

626-327-8890
xuganyu96@gmail.com
/xuganyu96
/in/ganyu-bruce-xu

Skills

ETL pipeline

- **Python 3**
- **Apache Airflow**
- **pandas**
- Data warehousing with **S3 + Parquet + Athena**

Cloud infrastructure

- Amazon web services:
  Elastic Container Service
  Fargate
  Lambda
  Simple Queuing Service
- Docker container
- DataDog

Web development

- **Flask**
- **MySQL, PostgreSQL**
- SQLAlchemy
- HTML, CSS, JavaScript

Miscellaneous

- Jupyter Notebook
- matplotlib
- C, Go, R
- Shell scripting
- vim, tmux, git

Education

UC Berkeley

Aug, 2015 - May, 2019

B.A. Mathematics

B.A. Statistics

GPA 3.464

Professional Experience

Software engineer, data platform

July, 2019 - present

LeanTaaS Inc. — Santa Clara, CA

- **Converted $1.1M+ contract into annual revenue** by implementing ETL pipeline from EHR backend database to internal analytics warehouse
- **Scaled raw data ingestion throughput** by migrating data storage from PostgreSQL to Apache Parquet on AWS S3 and AWS Athena
- **Shortened ETL implementation process from 12 to 4 weeks** by building scalable backend microservice that automates data schema validation
- **Expanded regression testing infrastructure** by migrating testing environment from a single EC2 instance to container clusters on AWS ECS
- **Automated unit testing** by integrating pytest into deployment pipeline
- **Increased ETL pipeline uptime** by implementing service availability monitoring via DataDog and engineer on-call alarms on PagerDuty
- **Reverse-engineered proprietary analytics/operational features** from EHR vendors and industry incumbents such as Epic, Cerner, Meditech, and Picis
- Led a team of 8+ data engineers/analysts, recruited and mentored junior teammates, carried out product vision from business and engineering leaders

Software engineering projects

AWS ECS/Fargate Executor for Apache Airflow

- Reduced server maintenance overhead and monetary cost from idle worker instances by leveraging serverless container deployment on AWS Fargate
- Achieved superior parallelism for bursty workload compared to conventional EC2 auto-scaling solution, such as with the popular choice of Celery Executor

Chives exchange

- Simulated order matching algorithms that pair stock orders and execute trade
- Leveraged relational constraints in data model to resolve race conditions and achieve horizontal scalability for the order matching algorithm
- Implemented web client (Flask) featuring simple user authentication, simulated stock trading, and market data visualization (ChartJS)
- Published to AWS on EC2 with self-hosted MySQL, NginX, and RabbitMQ; traffic routed under personal domain (GoDaddy) through AWS Route53

Simple resume generator

- A fun learning experience for HTML/CSS Flexbox, DOM manipulation in plain JavaScript, and a single-page application design
- This resume you are reading was created using this project, with a single HTML page that loads the rendering app and content from GitHub Pages