Canchen Li

835 Cotati Ter, Sunnyvale, CA 94085
canchen.lee@gmail.com · +1 (412) 214-2464 · github.com/Frost-Lee · linkedin.com/in/canchen-li

Education Background

Carnegie Mellon University Information Networking Institute — Pittsburgh, PA, USA | Aug. 2021 - Dec. 2022
M.S. in **Mobile and IoT Engineering**; GPA: 4.00 / 4.00
Courses: Introduction to Computer System, Storage Systems, Distributed Systems, Machine Learning with Large Datasets

Zhejiang University College of Computer Science and Technology — Hangzhou, Zhejiang, China | Sep. 2016 - Jul. 2021
B.Eng. in **Computer Science and Technology**, B.Sc. in **Statistics** (double degree); GPA: 3.63 / 4.00
Courses: Database Systems, Operating System, Computer Networks, Artificial Intelligence, Advanced Practices on Big Data Apps

Massachusetts Institute of Technology Langer Research Lab — Cambridge, MA, USA | Sep. 2019 - May. 2020
Visiting Student Researcher Program; 1 first author publication with 9.776 impact factor and 4 other publications.

Technical Skills

- **Languages**: Python, C, C++, Java, Go, Swift, Javascript, Perl, Shell script, SQL, HTML.
- **Frameworks**: Flask, React, SQLAlchemy, GraphQL, PySpark, unittest, TensorFlow, Keras, numpy, pthread, STL, UIKit, Spring.
- **Tools**: Docker, Kubernetes, AWS (EC2, Lambda, S3, DynamoDB, VPC), RabbitMQ, Jenkins, ElasticSearch, Postgres, MySQL, git, Linux.

Professional Experience

NVIDIA Deep Learning Infrastructure Engineer Intern — Santa Clara, CA, USA | May. 2022 - Aug. 2022

- Developed an EDA job scheduler for achieving compute farm load balancing and reducing peak EDA software license usage.
  - Designed and implemented the **system architecture**, including database schema, scheduled job run, and a **GraphQL** interface built upon **SQLAlchemy ORM** API for system management and job status query.
  - Embedded EDA license statistics query from **ElasticSearch** and Linux process level job lifecycle control to the scheduler.
  - Integrated the scheduler with the place and routing workflow in VLSI design and tested under multiple build tasks. Reflected the potential of reducing NVIDIA's EDA software license budget in **million level** in the long term.
- Participated in the research and deployment of a graph convolution model based congestion prediction solution.
  - Improved the clustering and searching algorithm for graph generation from hardware design data.
  - Collaborated with deep learning engineers and deployed the model as an asynchronous endpoint using **docker**, LSF, and **RabbitMQ**.

Apple IS&T Software Intern, IT Development Program (ITDP) — Shanghai, China | Feb. 2021 - Jul. 2021

- Conceptualized and implemented a proof-of-concept continuous evaluation and monitoring framework for machine learning models.
  - Developed the **database schema** and **REST API** with Postgres and Flask, with support for horizontal scalability on **Kubernetes**.
  - Created a two-step machine learning metric calculation mechanism with intermediate result storage as **time series data** in **InfluxDB** and the second step calculation with flux query language, empowering fast on-demand metric query and low storage cost.
  - Built the frontend with **React** for configuration management. Adopted **Grafana** for metrics visualization and real-time alerting.
  - Packaged the framework as a **helm** chart for easy installation. Setup the pipeline for **automated testing and deployment**.
  - Communicated and collaborated with **3 other teams** on integrating and testing the framework on existing deployed machine learning evaluation services, including use cases on Apple Trade In and Apple Store.
  - Presented the project to the **IS&T Management Team** (senior director level, **CEO -3**).
- Refactored and migrated business teams' offline supply chain logic to **AWS** using CloudFormation, EC2, Lambda, and RDS.

Amazon Software Development Engineer Intern — Beijing, China | Jun. 2020 - Sep. 2020

- Engineered and launched the shipping capacity hard constraint feature for direct fulfillment warehouses.
  - Conducted the table design in **DynamoDB** that supports constraint record edition history tracking.
  - Developed the backend service in **Spring** with full unit test coverage. Implemented the corresponding frontend interface in **jQuery**.
- Conducted ship method allocation analysis for direct fulfillment warehouse shipments.
  - Synthesized **terabytes** of data from **multiple data warehouses** for recalculating intermediate results of the business logic.
  - Analyzed the impact of fulfillment network capacity settings against the shipping costs and delays with **AWS Redshift** and **Jupyter Notebook**, and provided algorithm and operational optimization insights for the management team.

Selected Projects

- **Raft**: A **Go** implementation of the **Raft consensus algorithm** for distributed systems. Supports leader election, consensus-based log replication, and node recovery from failure. The solution behaves correctly under **high concurrency** conditions.
- **Cloud File System**: A **cloud file system** built on fuse API written in **C++**. With supportability for common file operations, block-level deduplication with Rabin fingerprinting, snapshots based recovery mechanism, and block-level LRU caching.
- **Malloc Lab**: A **memory allocator** with **C** implementation of a segregated list, measured 74.3% utilization and 8486 KOPS.
- **Reversi Zero**: A reversi AI player with the **AlphaGo Zero** machine learning algorithm, ranked top 5% in the course-wide tournament.