Min-Chi Chiang

Berkeley, CA 94704
mencher@berkeley.edu | louis5499.github.io | linkedin.com/in/mencher-chiang | (707)727-5028

Education

University of California, Berkeley — Berkeley, CA
Master of Engineering in EECS — *Aug. 2021 – May 2022 (Expected)*
Concentration: Data Science & Systems
Award: UC Berkeley College of Engineering Fung Excellence Scholarship

National Tsing Hua University — Hsinchu, Taiwan
Bachelor of Computer Science — *Sept. 2017 – Jan. 2021 (Early graduated)*
Overall GPA: 3.95/4.30
Major GPA: 4.05/4.30
Award: Outstanding Academic Achievement Award (10 out of 138 students)
Key Coursework: Operating System, Design and Analysis of Algorithms, Computer Architecture, Data Structures, Database Systems, Software Engineering, Massive Data Analysis, Embedded Systems, Parallel Programming

Work Experience

9dynamics — Jan. 2020 – Dec. 2020
Backend Tech Lead — *Taipei, Taiwan*
- Led 5 people team to develop 60k concurrent instant messaging systems by employing NATS Streaming, Websocket, gRPC, MongoDB and Redis with Kubernetes
- Supervised microservice system design, service mesh networking, SQL & non-SQL DB integration on Social App
- Built 100% automated DevOps mechanism such as CI/CD, monitoring on E-Commerce, Social App projects
- Integrated Scrum with our DevOps process from scratch to well manage with 10+ people developers team & PM

giloo Inc. — July 2017 – Jan. 2020
System Architect & Software Engineer — *Taipei, Taiwan*
- Oversaw backend infrastructure reconstruction from monolithic system to microservice with Kubernetes & Istio
- Reached 99.99 % availability by integrating monitoring systems & fault tolerance mechanism with Kubernetes
- Designed payment systems supporting Google & Apple Pay by integrating with Google, Apple and Stripe APIs and PostgreSQL & Redis Database
- Implemented authorization systems supporting Google, Apple, Facebook OAuth by using JWT and Redis
- Collaborated with KKStream & Bitmovin to develop online streaming platform with Typescript

Research Experience

DynamoML: Dynamic Resource Management Operators for Machine Learning Workloads
Min-Chi Chiang and Jerry Chou
- Speedup 3.44 times of Resnet-50 model training time by employing gang scheduling with locality awareness
- Keep 74ms of inference response time by designing inference auto-scaling controller for priority scheduling
- Fully-utilize GPU resources by utilizing CUDA library hook & API blocking to provide fractional GPU allocation and time-share fairness scheduling
- Reduce 60% overall ML workload execution time by integrating a set of custom runtime scheduling operators
- Awarded the **best paper** in the 11th International Conference on Cloud Computing and Services Science

Leadership Experience

7th Mei-Chu Hackathon | *President* — Nov. 2018 – Oct. 2019
TED at MingdaoHS 2020 Summit | *Speaker* — Sept. 2020

Skills

Programming: C/C++, Python, Verilog, Java, Golang, Javascript, SQL
Framework: PySpark, Pthread, MPI, OpenMP, CUDA, Nuxt.js, Egg.js, Moleculer, Go-Micro
Deployment Tool: Kubernetes, Docker, Istio, Amazon Web Service, Google Cloud Platform, Gitlab CI/CD, Jaeger
Database Management: MySQL, PostgreSQL, MongoDB, Redis, Elasticsearch, Neo4j, Dgraph, Prometheus