Yu-Chen CHENG
https://rudeigerc.dev

E-mail: rudeigerc@gmail.com

EDUCATION

Shanghai Jiao Tong University — Shanghai, China
M.Eng. in Software Engineering — *Sep. 2019 – Mar. 2023*
Thesis: *Distributed Data Processing Job Scheduling in Multi-Tenant Shared Clusters*

Shanghai Jiao Tong University — Shanghai, China
B.Eng. in Software Engineering — *Sep. 2015 – Jun. 2019*

EXPERIENCE

WizardQuant — Shanghai, China
AI Platform Development Engineer — *May. 2023 – Sep. 2024*

- **AI Platform.** Participared in the redesign and refactor of internal MLOps platform upon Kubernetes, including the refactor of Python SDK and CLI, the redesign of unified Job models, the implementation of microservices via Protobuf and gRPC, and the development of dashboard via React and Vue.
- **GitOps Pipeline.** Designed and built efficient Python and Go DevOps pipelines upon Tekton to accelarate the developing and testing processes.
- **Large Language Models.** Designed and conducted a series of experiments on large language models inference backends, focusing on metrics including Time to First Token (TTFT), Time Per Output Token (TPOT) and end-to-end throughput.
- **Developer Culture.** Introduced **a)** team-scoped modern Python project template with PDM **b)** Helm-based templates for hosting static sites on Kubernetes with integrations of Tekton pipelines and OAuth Proxy.
- **Mentopship.** Mentored **a)** the development and integration of the Dex identity provider with LDAP, providing an authentication and authorization middleware for internal systems. **b)** the benchmarking framework for Kubernetes schedulers, providing a unified interface for benchmarking and comparing different schedulers.

#KUBERNETES #GO #PYTHON #RAY #REACT #VUE

Ant Group — Shanghai, China
Site Reliability Engineer Intern — *Jun. 2021 – Sep. 2021*

- **Federation Observability Platform.** Participated in the design and development of the internal multi-clusters observability platform.
- **Identity Provider Connector.** Led the development of the Dex identity provider plugin connecting to the internal identity and access management system, providing OIDC services for internal systems such as Argo CD and Grafana.

#KUBERNETES #GO #REACT

The Linux Foundation — Remote
Mentee of LFX Mentorship Program, CNCF - Volcano — *Feb. 2021 – May. 2021*

- **System Stability Enhancement.** Built an efficient testing framework based on Kubernetes and imported unit tests and E2E tests to cover more classical scenarios for Volcano, a Kubernetes native batch scheduling system for compute-intensive workloads.

#KUBERNETES #GO

JD.com — Shanghai, China
Software Development Engineer Intern — *Jun. 2018 – Sep. 2018*

- **JingZao.** Refactored module JingZao with React Native and introduced several sections in Objective-C for JD.com mobile application.

#OBJECTIVE-C #REACT NATIVE

SELECTED PROJECTS

Moirai: Distributed Data Processing Job Scheduling in Multi-Tenant Shared Clusters — *Sep. 2021 – Dec. 2022*

- Built a scheduler aiming to improve the resource utilization of multi-tenant shared clusters for distributed data processing jobs with Kubernetes scheduling framework.
- *Unified Multi-Tenant Abstraction Model:* Designed a model providing abstraction and binding relationships between tenants and jobs.
- *Multi-Tenant-Oriented Scheduling System for Jobs:* Designed a mechanism of scheduling based on placement constraints and resource sharing.

PUBLICATIONS

[1] **Yuchen Cheng**, Chunghsuan Wu, Yanqiang Liu, Rui Ren, Hong Xu, Bin Yang, and Zhengwei Qi. 2020. OPS: Optimized Shuffle Management System for Apache Spark. In *49th International Conference on Parallel Processing - ICPP (ICPP '20)*.

Programming Skills

Languages: Go, Python, TypeScript, JavaScript, C++
Topics: Resource Management, Scheduling
Tools and Frameworks: Kubernetes, Ray, React, Vue

Licenses and Certifications

Japanese-Language Proficiency Test Level N1   *2021*

Honors and Awards

The Second Prize in Taiwan Student Scholarship, Ministry of Education, P.R.C.   *Dec. 2018, Mar. 2021 and Oct. 2021*
The Third Prize in Taiwan Student Scholarship, Ministry of Education, P.R.C.   *Dec. 2017*

Miscellaneous

Blog: https://rudeigerc.dev
GitHub: https://github.com/rudeigerc
LinkedIn: https://www.linkedin.com/in/rudeigerc/
Languages: Chinese (Native proficiency), English (Professional working proficiency), Japanese (Limited working proficiency)
Selected Open Source Contributions: Member of **Tekton** and **Tensorchord**. tektoncd/cli, tektoncd/triggers, tektoncd/dashboard, strimzi/strimzi-kafka-operator, volcano-sh/volcano, tensorchord/envd

Note

I am currently serving in military training service in Taiwan, expected to last from October 2024 to February 2025.

Last updated: October 6, 2024