James Li

- Senior Backend Engineer with Strong Infrastructure/SRE Capability.
- email: jamesli@csie.io
- Summary
  - High-performance systems architect with **5+ years** designing cloud-native SaaS and data center automation. Scaled systems to **100k+ QPS** with **99.5%+ SLA**. Optimized data retrieval and deployment latency **from 30+ min to <2 sec**. Expert in Python, Azure, K8s, with proven experience in **zero data loss and near-zero RTO**. Leverages modern AI-assisted tooling to accelerate development while maintaining code quality and system reliability.

Senior Software Engineer, Wiwynn, Oct 2022 - Present

Python/PostgreSQL/DevOps/K8s/Kafka/Azure

- Immersion Cooling Infrastructure Test Infra
  - Co-designed and developed a high-concurrency immersion cooling test infrastructure for global data centers, scaling the system to support **100k+** peak concurrent requests. Optimized system performance and database efficiency through **K8s Ingress/Load Balancers**, **Gunicorn worker tuning**, and database layer enhancements using **PgBouncer** and complex **query optimization**, ensuring **99.5%+** SLA.

- Data Center Data Management System
  - Architected a high-concurrency data platform supporting **150k+** simultaneous tasks with **zero data loss**. Implemented a state-machine driven lifecycle management system that enables **near-zero RTO** (Recovery Time Objective), allowing automated state synchronization and seamless task resumption immediately after power or network failures. Optimized data retrieval latency from ~30 minutes to 1-2 seconds —a **99% reduction** in access time, while maintaining **99.5%+ SLA** and high availability across global data center operations.

- Cloud Infrastructure & DevOps
  - Reduced K8s provisioning time **from hours to minutes** by architecting an automated package delivery system using Azure Front Door (CDN) and Blob Storage, effectively bypassing regional network constraints for internal apt repositories.

    - **Eliminated Single Point of Failure** (SPOF) and optimized the global deployment pipeline by migrating from self-hosted registries to Azure Container Registry (ACR). This strategic migration, combined with optimized caching and CI/CD workflows, reduced deployment cycles **from tens of minutes to mere seconds**, significantly accelerating the global delivery of critical services.

    - Standardized cross-region data workflows by designing an integrated Azure Blob Storage tool, **reducing data transfer latency from several minutes to just seconds**. Optimized the equilibrium between high-speed performance and cloud egress costs for international data centers, resulting in significantly improved operational efficiency and reduced infrastructure overhead.

- High-Performance Architecture & System Optimization
    - Achieved a **99% improvement** in API response time by leading cross-team diagnostic initiatives; utilized database profiling to redesign index architectures and optimize complex query logic.
    - Eliminated API bottlenecks by implementing Kafka as an asynchronous messaging layer, transforming response times from 3 minutes to near real-time (**99%+ latency reduction**).

- Infrastructure Automation & Operational Efficiency
    - Minimized MTTR (Mean Time to Recovery) **from hours to minutes** by developing a unified abstraction layer for multi-vendor switch commands. This interface empowered on-site non-engineering personnel to execute complex network troubleshooting, reducing the dependency on senior engineering intervention and significantly improving SLA.
    - Empowered global operations through structured technical training and a unified command abstraction layer, achieving a **100%** autonomous resolution rate for routine infrastructure incidents. This initiative successfully eliminated the need for senior engineering escalation for common tasks, while simultaneously slashing MTTR **from hours to minutes**.

Backend Engineer, PalUp, Jun 2022 - Sep 2022

Python/PyTorch/Flask

- Developed image recognition system for automated end-to-end testing
    - Developed image recognition system for automated end-to-end testing simulating user behavior — established data collection and annotation system with standardized workflow to ensure training data quality, trained YOLO model using

PyTorch for UI component detection and classification, deployed via Flask API achieving **95% accuracy** across multi-platform UI elements.

R&D Substitute Military Service, AUO, Aug 2020 - May 2022

Python/Flask/JS/DevOps

- Chiller Power Optimization System
  - Developed chiller power optimization system — analyzed motor load patterns during cooling cycles, optimized motor frequency control strategy through machine learning to reduce energy consumption costs, developed web interface for multi-site deployment.
- Infrastructure Management
  - Managed Linux server infrastructure and PostgreSQL/MySQL databases ensuring stable operation of ML SaaS platform.
- Established a DevOps and agile development culture
  - Led agile transformation initiative implementing Kanban workflow and GitLab CI/CD pipelines to improve team development efficiency.
  - Conducted hands-on Git workflow training

- Education
  - M.S., Computer Science and Information Engineering, National Chung Cheng University, 2018 - 2020
  - B.S., Computer Science and Information Engineering, National Chung Cheng University, 2014 - 2018
