YUE (LUNA) JIANG

yuejian2@andrew.cmu.edu 4127731426 Yue Jiang Website

EDUCATION

Carnegie Mellon University, Pittsburgh, GPA: 3.9/4 *Jun 2022 - Dec 2023*
Master of Information Technology Strategy, School of Computer Science
Courses: Storage Systems, Database Systems, Distributed Systems, Advanced Cloud Computing, Computer Networks

Nankai University, Tianjin, Major GPA: 90.8/100 **(Top 5%)** *Aug 2017 - Jun 2021*
Bachelor of Engineering, Computer Science and Information Technology
Honors: National Scholarship (**Top 2%**) *2018 & 2019*
Courses: Operating System, Principles of Compilers, Principles of Computer Architechtures

TECHNICAL SKILLS

Programming: C/C++, Java, Python, TypeScript, Go, MATLAB, SQL, PHP, Assembly
Frameworks: Django, Spark, Kubernetes, Spring MVC, MongoDB, NumPy, CUDA, React
Tools: Linux, Git, AWS, Vim, Shell, Terraform, Clang, CMake, Docker, GDB, LaTeX

WORK EXPERIENCE

Splunk - *Backend Software Engineer Intern – Core Framework Team* *San Jose, United States*
⋆ Ingest Actions: a key feature for data masking, filtering, and routing | C++, Python *June 2023 - August 2023*
- Developed the routing of less frequently used data in **Parquet** format from Splunk to S3 to help customers reduce storage costs. The **Parquet** support is also required by over **50%** customers using Splunk Federated Search-S3.
- Provided **2** new license supports for Ingest Actions, leading to an estimated **40%** increase in its exposure to customers.
- Participated in **QA testing** process for Ingest Actions, identifying and fixing **2** critical bugs.

Higgs Asset Management - *C++ Backend Development Intern* *Hangzhou, China*
⋆ OpsAdapter: Market Data & Trading System | C++ *February 2022 - July 2022*
- Built the Market Data Retrieval System for **3** exchanges with efficient real-time market data processing.
- Built the Trading System for **3** exchanges with robust communication between exchanges and the Higgs, handling synchronous/asynchronous requests and callbacks partnered with traders.
- Enhanced the system using concurrency control, parallelism, and modern C++ features, reducing packet loss **from 10% to 3%**, thus enabling prompt trading responses.

Megvii - *Full-Stack Software Development Intern* *Beijing, China*
⋆ Hubble: Data Management System | Python *July 2020 - September 2020*
- Constructed Hubble's backend to manage data storage with **Django**, greatly optimizing researchers' data access time.
⋆ Sisyphus: Data Transfer System | Python, HTML, JS *July 2019 - September 2019*
- Accomplished data transmission scripts under various scenarios, leading to a **35%** reduction in data migration cost.
- Developed frontend of Sisyphus with HTML and Bootstrap; connected it to backend with **REST APIs**.

PROJECTS

⋆ CloudFS: A Cloud-backed Local File System on Linux | C++ *Storage Systems Project, CMU*
- Designed a **hybrid file system** with a local SSD and a cloud storage service akin to S3.
- Applied block-level deduplication based on Rabin Fingerprinting to cut down cloud storage cost by around **60%**.
- Proposed a snapshot mechanism to ensure data integrity and facilitate recovery in case of failures.
- Utilized spare SSD capacity as cache for cloud-stored data, improving performance and cutting cloud costs by **50%**.

⋆ Cloud-Computing Projects | Python *Advanced Cloud Computing Project, CMU*
- Applied **Terraform** to automate the provisioning of resources on AWS, reducing costs and improving RPS by **30%**.
- Developed an **ETL** Processing pipeline using **Spark** on **2.85 billion** web pages for large-scale topic modeling.
- Designed a **K8s** scheduler and a scheduling strategy for a certain set of workloads to maximize utility.

⋆ BusTub: A Relational Database Management System | C++ *Database Systems Project, CMU*
- Realized a buffer pool manager responsible for moving physical pages back and forth from main memory to disk. Optimized the QPS by **10x** under certain workloads by fine-grain locking and better replacement policy.
- Implemented a concurrent disk-backed **B+ Tree** with latch crabbing mechanism for fast data retrieval.