Jianshen **Liu**

Software Engineer · Computational Storage, SmartNIC, and Distributed System Expert

Cupertino, CA 95014, USA

+1 (415) 728-6512 | ljishen@gmail.com | www.jianshenliu.com | ljishen | jianshenliu

Skills

DevOps	AWS/GCP, Jupyter Notebook, Docker, Kubernetes, Vagrant, Jenkins, Git, Mercurial, Maven, Gradle, VSCode, IntelliJ IDEA, CI/CD
Back-end	Linux Kernel, eBPF, DPDK, Ceph, Hadoop, MySQL, PostgreSQL, RocksDB, Apache Arrow, Ansible, REST API, NVMe, Microservices, Machine Learning, Distributed Systems
Programming Languages	C++, Python, Java, Bash, SQL, PHP, C#, Javascript, LaTeX, Vimscript
Others	English, Cantonese, Mandarin
	Patience, Open-Mindedness, Active Listening, Versatility, Enthusiasm, Agile Methodologies, Best Practices

Work Experience

University of California, Santa Cruz — *Santa Cruz, USA*

Graduate Student Researcher — *Jan 2016 - Present*

- Developing a **SmartNICs**-based **distributed data-flow management platform** for HPC big data workloads at Sandia National Laboratories
- Implemented *Bitar*, a **C++** library based on **Apache Arrow** and **DPDK** to accelerate (>8x) data compression with hardware using the **BlueField-2 DPU**
- Evaluated the limits of network packet processing with ARM processors in SmartNICs using **DPDK** and an **in-kernel packet generator**; work highlighted by The Next Platform and Data Centre Dynamics
- Contributed to **Ceph** to fix performance evaluation issues, reduced the stddev of the result by 90%
- Proposed a novel **media-based methodology** to quantify the benefits of offloading data access functions to storage devices, prototyped with **YCSB** and **RocksDB** on **ROCKPro64**, and automated with **Ansible** and **Python**
- Developed *no_fscache*, a **Linux kernel module** to address page cache pollution before RWF_UNCACHED was introduced into the POSIX API

KIOXIA America Inc. — *San Jose, USA*

Strategic Marketing Engineer Intern — *Jun 2020 - Sep 2020*

- Quantified the performance speedup (>10x) of offloading SQL query processing to SSDs using **MariaDB**, **Docker**, **eBPF**, **FlameGraph**, **Linux perf**, **cgroups**, and the **TPC-H benchmark**
- Reported monthly to the senior vice president to drive internal strategic planning for **computational storage device** products
- Collaborated on cross-team projects to identify IO performance issues in NUMA systems

Samsung Semiconductor — *San Jose, USA*

Performance Analysis Engineer Intern — *Jul 2019 - Sep 2019*

- Established a **distributed testing framework** to automate the performance evaluation for Samsung Network KV devices
- Created **eBPF**- and **kprobe**-based tools to expose the **performance observability** of the device API
- Identified major system bottlenecks with the tools developed, resolving issues boosted the write performance by 60%

Wumii Technology — *Shenzhen, China*

Android Development Team Lead — *Dec 2012 - Jun 2014*

- Led design and code reviews of the first **reading recommendation app** (among the top 25 social apps) in China
- Implemented an energy-efficient and highly available Paho **MQTT**-based notification module in **Java** that outperformed commercial solutions in terms of latency by 35%

Senior Software Engineer — *Jun 2010 - Dec 2012*

- Worked as a core engineer to develop the first **news recommendation system** in China
- Drove the system architecture to handle distributed requests based on the **Spring** and **Hibernate** framework, and the code obfuscation system using Google's **Closure Compiler**
- Launched a **Java**-, **PHP**-, and **Javascript**-based platform to present picturized article recommendations across multiple platforms, achieved 200 million monthly active users

Education

University of California, Santa Cruz — *Santa Cruz, USA*

Doctor of Philosophy (Ph.D.), Computer Science and Engineering — *2016 - Present*

PhD Candidate advised by Prof. Carlos Maltzahn with a GPA of 3.95/4.00

Dissertation topic: **Eusocial Storage Devices**

South China Agricultural University — *Guangzhou, China*

Bachelor of Science (BS), Information and Computing Science — *2006 - 2010*

GPA: 3.45/4.00

Honors & Awards

Sep 2022 **Outstanding Student Paper**, IEEE High Performance Extreme Computing (HPEC) (top 5%) *Virtual Conference*
Oct 2009 **National Encouragement Scholarship**, South China Agricultural University *Guangzhou, China*
Oct 2009 **The Second Prize Scholarship**, South China Agricultural University (top 10%) *Guangzhou, China*
Oct 2008 **State Grants**, South China Agricultural University *Guangzhou, China*
Oct 2008 **The Second Prize Scholarship**, South China Agricultural University (top 10%) *Guangzhou, China*

Publications

Processing Particle Data Flows with SmartNICs
Jianshen Liu, Carlos Maltzahn, Matthew Leon Curry, Craig Ulmer
26th Annual 2022 IEEE High Performance Extreme Computing (IEEE-HPEC 2022), 2022, Virtual Conference

Performance Characteristics of the BlueField-2 SmartNIC
Jianshen Liu, Carlos Maltzahn, Craig Ulmer, Matthew Leon Curry
arXiv preprint arXiv:2105.06619 (May 2021). 2021

Implementing a Kernel Module to Eliminating External Caching Effects
Jianshen Liu
The Center for Research in Open Source Software (CROSS) (Jan. 2020). 2020

Scale-out Edge Storage Systems with Embedded Storage Nodes to Get Better Availability and {Cost-Efficiency} At the Same Time
Jianshen Liu, Matthew Leon Curry, Carlos Maltzahn, Philip Kufeldt
3rd USENIX Workshop on Hot Topics in Edge Computing (HotEdge'20), 2020, Virtual Conference

MBWU: Benefit Quantification for Data Access Function Offloading
Jianshen Liu, Philip Kufeldt, Carlos Maltzahn
International Conference on High Performance Computing (HPC-IODC 2019), 2019, Springer, Cham

Presentations & Posters

Experience of using the BlueField-2 DPU
Presented at Sandia National Laboratories (May 2022). 2022

Processing Particle Data Flows with SmartNICs
Presented at UC Santa Cruz Open Source Symposium (Sept. 2022). 2022

Eusocial Storage Devices
Presented at KIOXIA America Inc., Seagate Technology, and Fujitsu America Inc. (2020). 2020

MBWU (MibeeWu): Quantifying benefits of offloading data management to storage devices
Poster Session at 17th USENIX Conference on File and Storage Technologies (FAST'19) (Feb. 2019). 2019

Quantifying benefits of offloading data management to storage devices
Poster Session at 2019 OCP Global Summit Symposium (Mar. 2019). 2019

Certifications

Jan 2022 **Cruzhacks Mentor Certificate**, CruzHacks *Santa Cruz, USA*
Dec 2020 **Graduate Student Professional Communication Certificate**, University of California, Santa Cruz *Santa Cruz, USA*

Program Committees

CruzHacks *Santa Cruz, USA*
Tech Mentor *Jan 2020, Jan 2021*
Mentored hackers to build web applications to solve real-world problems

Center for Research in Open Source Software (CROSS) *Santa Cruz, USA*
Workshop Chair *Oct 2021, Oct 2020*
Chaired the SmartNICs and Eusocial Storage Devices workshop at the 5th and 6th CROSS Research Symposium at UC Santa Cruz