

Shunyu Yao

shunyu@vt.edu | (+1) 434 466 6359 | shunyuyao515.github.io

EDUCATION

Virginia Tech, Blacksburg, VA, United States

Ph.D. Student, Computer Science

August 2020 - May 2025 (Expected)

University of Virginia, Charlottesville, VA, United States

Master of Computer Science (M.CS.), Computer Science

September 2018 - August 2020

BNU - HKBU United International College, Zhuhai, China

Bachelor of Science (B.S.), Computer Science and Technology

September 2014 - June 2018

PROJECT EXPERIENCE

Towards a Stateful Serverless Platform for HPC Applications

Distributed Systems & Storage Laboratory, Virginia Tech

September 2022 - Current

- Designing near-data scheduling for serverless workflows with in-memory caching intermediate states exchanged between memory intensive HPC serverless functions; Enabling native support for MPI on serverless platforms

Runtime-OS Co-Design for CXL Memory System

Distributed Systems & Storage Laboratory, Virginia Tech

February 2023 - Current

- Extracting performance characteristics of memory-intensive applications through DAMON; Studying CPython object access tracking

Automated FaaSification of Monolithic Workloads

Distributed Systems & Storage Laboratory, Virginia Tech

March 2021 - January 2023

- Created a tool that automates the process of dynamically decomposing Python workloads into serverless functions that are deployment-ready with fine-tuned performance

Cyberinfrastructure for Waterborne Antibiotic Resistance Risk Surveillance

Distributed Systems & Storage Laboratory, Virginia Tech

September 2020 - May 2021

- Enabled flexible pipeline by refactoring components into serverless microservices

Checkpoint/Restart Containerized GPU Tasks

University of Virginia (Advisor: Dr. Andrew Grimshaw)

September 2019 - May 2020

- Evaluated existing attempts to capture out-of-kernel GPU states during runtime

Partitioned Parallel Processing Spaces (PCubeS)

University of Virginia (Advisor: Dr. Andrew Grimshaw)

October 2018 - May 2019

- Translated the serial C version of LULESH hydrodynamic model into PCubeS parallel programming language to test the correctness of PCubeS type architecture

Final Year Project: JavaScript Interpreter

BNU - HKBU United International College (Advisor: Dr. Philippe Meunier)

September 2017 - May 2018

- Implemented a partial JavaScript interpreter using Antlr4

WORK EXPERIENCE

Graduate Research Assistant

Distributed Systems & Storage Laboratory, Virginia Tech

September 2021 - Current

- Projects: Automated FaaSification of Monolithic Workloads; Stateful serverless Platform for HPC

Graduate Teaching Assistant

Virginia Tech

September 2020 - Current

- Teaching Assistant for Computer Organization (CS2506) and Computer Systems (CS3214)

Research Aide Intern

Programming Models & Runtime Systems Group, Argonne National Lab

June 2021 - August 2021

- Participated in developing OSHMPI, an implementation of OpenSHMEM on MPICH

Distributed Systems Intern

Lancium

June 2019 - August 2019

- Automated HPC resource management and task containerization using Ansible, Slurm, Docker and Singularity

PUBLICATION

- Yuze Li, **Shunyu Yao**, Jaiaid Mobin, M. Mustafa Rafique, Dimitrios Nikolopoulos, Kirshanthan Sundararajah, Huaicheng Li, Ali R. Butt. 2024. Towards Efficient Python Interpreter for Tiered Memory Systems. In Proceedings of the 22nd USENIX Conference on File and Storage Technologies (FAST'24). February 27, 2024, Santa Clara, CA, USA. USENIX Association, USA, Work-in-Progress Report & Poster, 3 pages. (pdf) **To Appear**
- Yuze Li, **Shunyu (David) Yao**. 2023. Understanding and Optimizing Serverless Workloads in CXL-Enabled Tiered Memory. Project Report, 8 pages. (pdf)
- **Shunyu (David) Yao**, Muhammad Ali Gulzar, Liqing Zhang, Ali R. Butt. 2021. Towards a Serverless Bioinformatics Cyberinfrastructure Pipeline. In Proceedings of the 1st Workshop on High Performance Serverless Computing (HiPS '21), June 25, 2021, Virtual Event, Sweden. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3452413.3464787> (pdf)

SKILLS

Programming Languages: Python, C/C++, Bash, MPI, OpenMP, CUDA, Matlab

Frameworks: Docker, Kubernetes, AWS, Slurm, Openwhisk, Memcached, Redis, Ansible, Singularity, Flask, Antlr4

Languages: English, Mandarin (Native Speaker)