

About Sun Yat-sen University



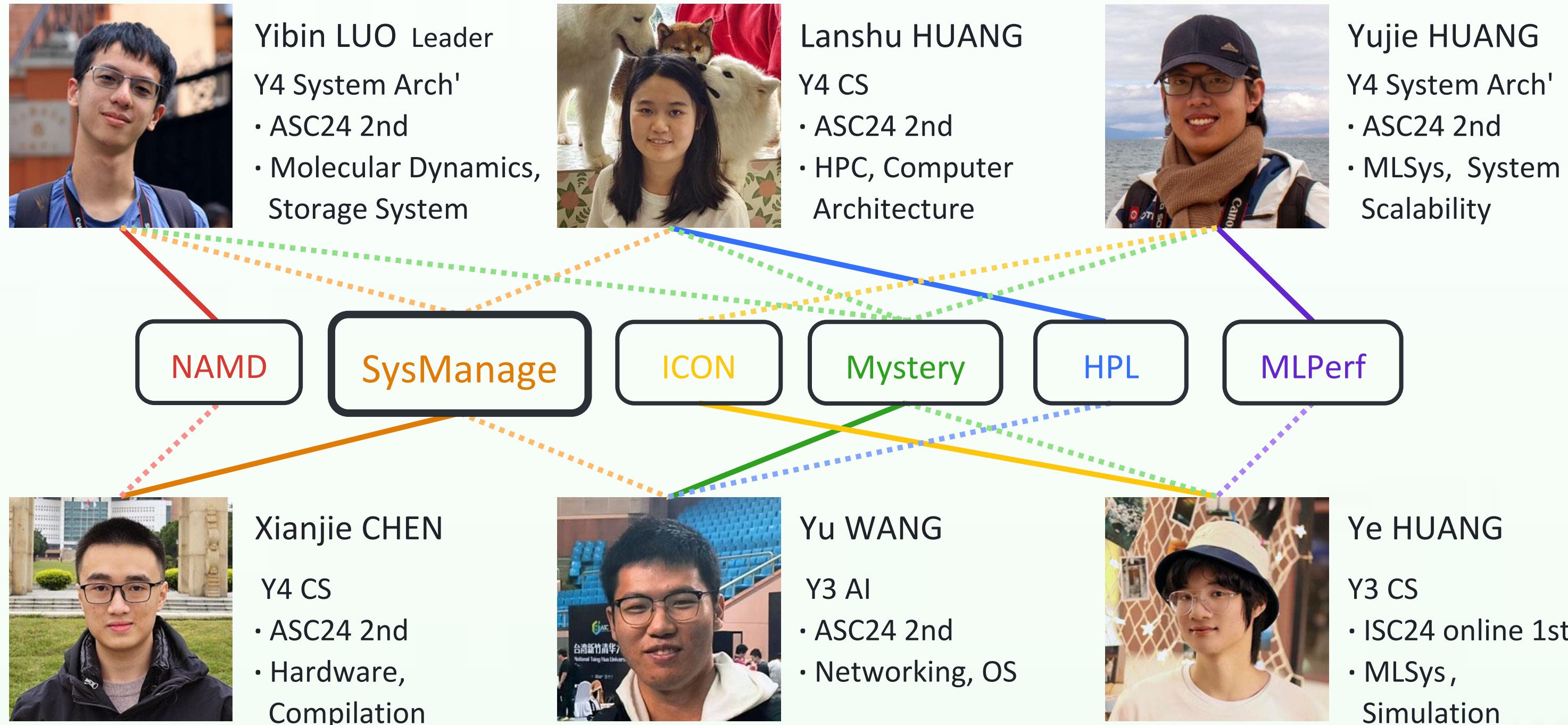
Established in 1924 by the great pioneer of China's democratic revolution, Dr. Sun Yat-sen, SYSU has since grown into one of the top national universities in China. With more than **68,000** students, **70** academic departments, and **five** campuses located in Guangzhou, Shenzhen and Zhuhai, SYSU stands at the forefront of research and education.



As SYSU celebrates its **100th** anniversary in 2024, we are honored to represent our university at the IndySCC24 competition, continu-ing our legacy of excellence.

About Our Team

The SYSU HPC club has a strong tradition in supercomputing competitions. We won the **Silver Price (2nd)** at **ASC24** with five members, while one member claimed the **Online Championship at ISC24**. Our team has also excelled in many other performance optimization contests. Now, we're preparing for our next challenge at SC24.



Unity in Diversity drives innovation and success

Diversity

♀ Diversity of Genders

Team consists of **ONE** female and **FIVE** male members.

📖 Diversity of Majors and Interests

We are from **THREE** different majors (CS, AI, Arch'), as well as different personal interests and skills.

📍 Diversity of Birthplace and Culture

Our members come from over **FOUR** different provinces across China, speaking four distinct accents / dialects.

Unity

👥 Teamwork

All of us have participated in HPC competitions, with the best experience in teamwork.

✅ Task Assignment

Tasks are assigned based on individual interests and strengths, with each task having at least one leader and another assistant.

🎯 Shared Goal & Passion

We are united by a common passion for HPC and a commitment to overcoming the boundaries of performance & efficiency.

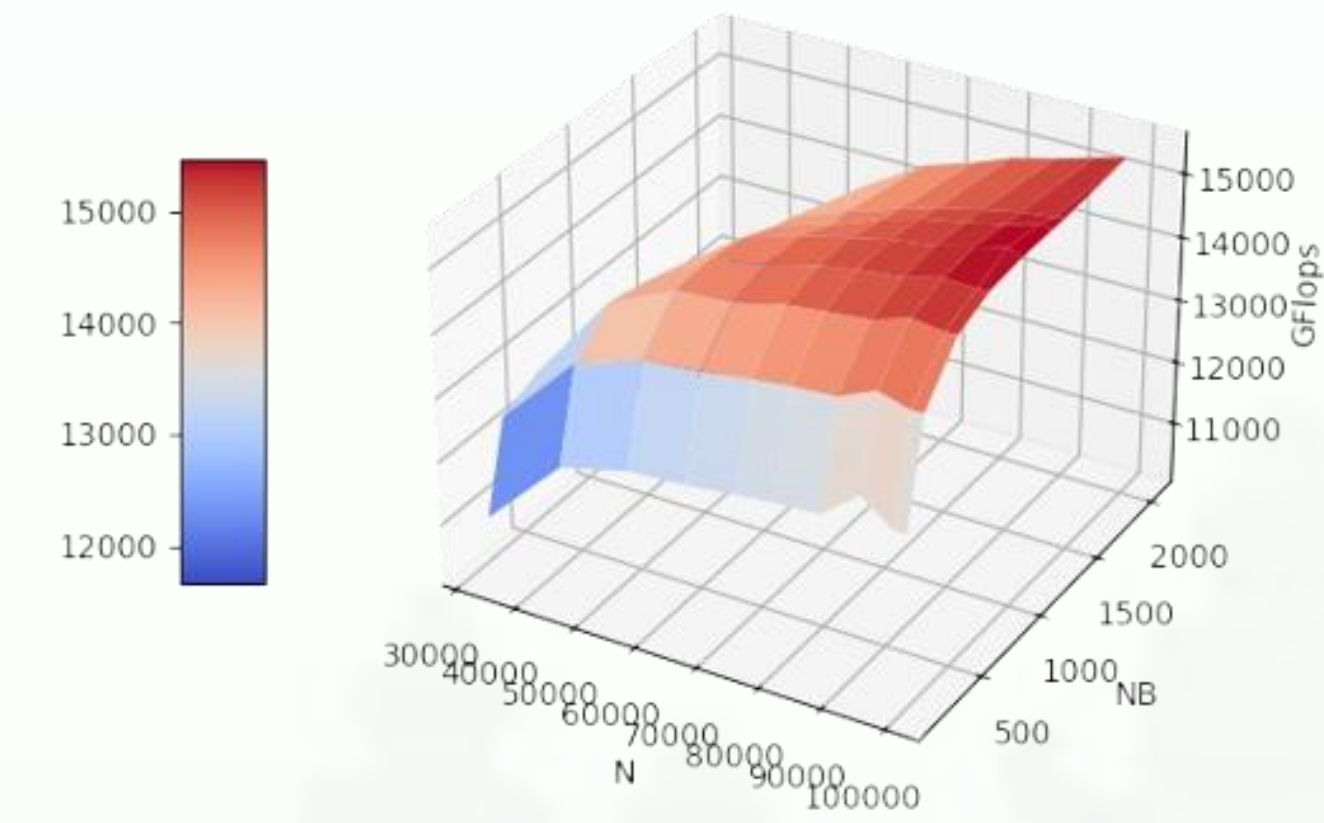
General Optimization Strategies

- Improving code by performing **vectorization** and compute-comm-unication **overlapping** on hotspots, port it to **GPU** if feasible.
- Tuning the **file system** and **MPI libraries** on the provided cloud-based platform before running applications.
- Exploit various **profilers** to identify hotspots and performance bottlenecks.
- Deliberately bind processes and threads to CPUs and NUMA nodes.
- Endeavor to different compilers, compile options and libraries.

Applications Optimizations

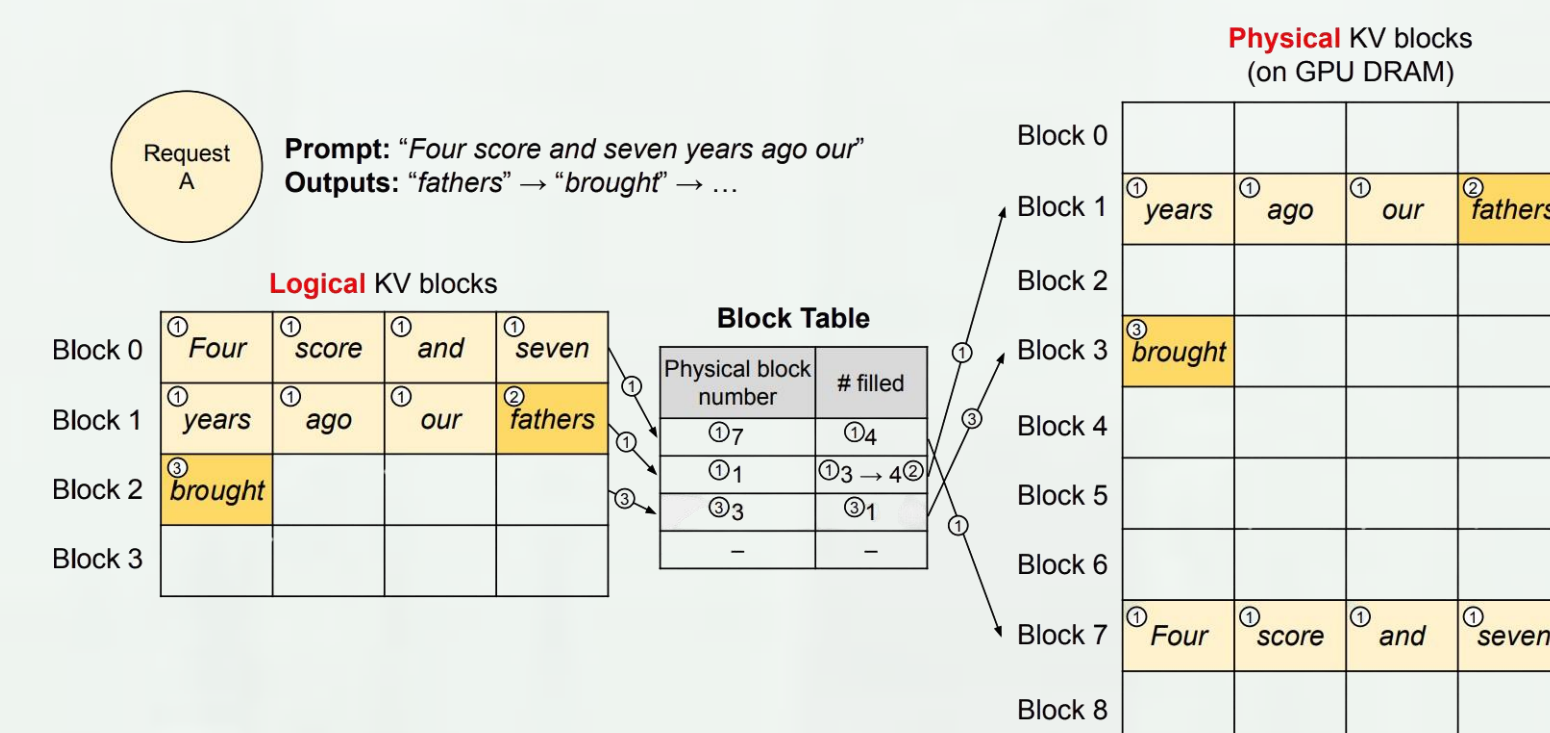
LINPACK

- Use the latest **AMD binary** for optimal performance.
- Adjust **input parameters** to understand their impact on performance and explore the underlying rationales.



MLPerf

- Offline LLM inference with **BERT-large** from MLPerf-Inference-DataCenter, aiming to maximize system **throughput** while maintaining accuracy.
- Apply quantization, specialized kernels, and paged KV-Cache to enhance performance.
- Design heterogeneous data parallelism across both CPU and GPU nodes.

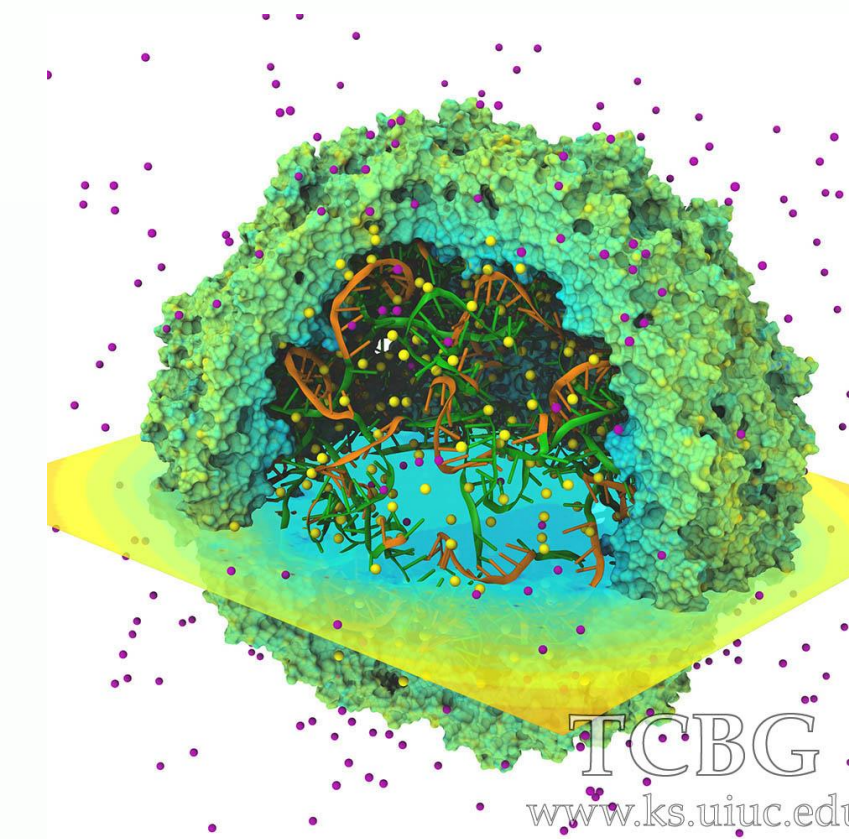


Time Scheduling

- During the 48-hour competition, we will run **ICON** and **NAMD** first, then the **mystery application**.
 - Get familiar with the mystery application on the login node when running ICON and NAMD.
- Estimate** roughly running time from try-run output, and prioritize the tasks with the higher scoring efficiency.
- For weak scalability applications, run multiple tasks simultaneously.
- We have a hardware system "**Quarterback**" who is fully in charge of all cluster resources to schedule tasks wisely and prevent job conflicts.

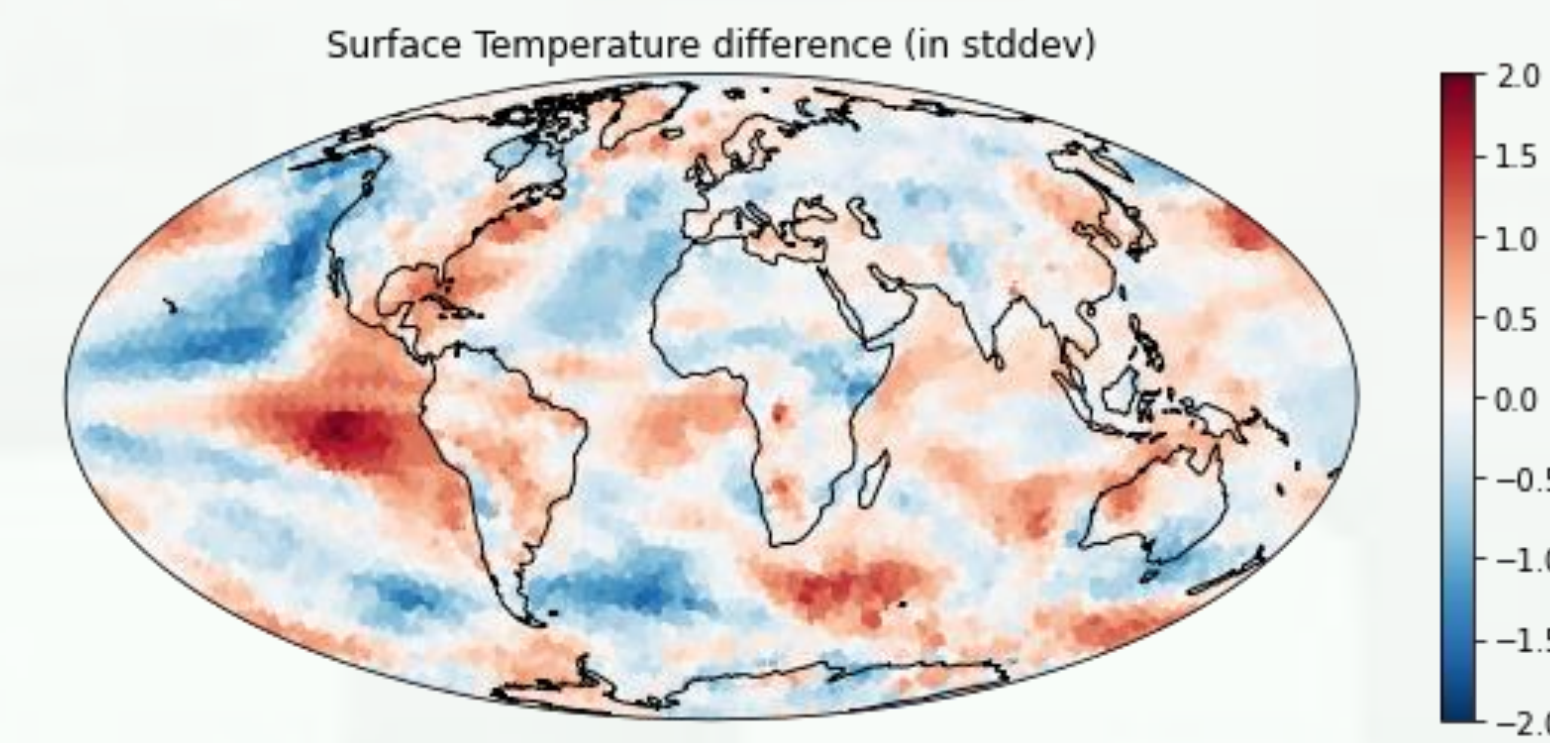
NAMD

- Compared to **GPU offload** mode, **GPU resident** mode and CPU mode for performance differences.
- Profile the system to identify hotspots and analyze communication patterns.



ICON

- Focus on analyzing **I/O behavior**.
- Parallel I/O: use parallel Filesystem
- Aim to achieve asynchronous I/O and minimize redundant I/O operations.



Mystery Application

- Explore different compilation methods, such as GNU, OneAPI, and NVCC.
- Analyze scalability and determine whether the application is more suitable for CPU or GPU.
- Estimate execution time and co-schedule with other applications to optimize resource utilization and performance.

Software



- Rocky Linux** release 9.4
- with customized kernel settings



- BeeGFS**
 - a parallel file system designed for HPC applications
 - to speed up the I/O processing of applications
- Data Plane Development Kit**
 - a robust framework that boosts packet processing rates
 - therefore accelerating network performance



- NVIDIA CUDA Toolkit**
 - provide Highly optimized GPU-accelerated libraries
- Intel OneAPI**
 - Intel compilers produce highly-optimized programs, even on AMD platforms
 - Intel HPC libraries (MKL, MPI, etc.) are widely-used for their high performance



- Spack**
 - offer convenient environment control & fast compilation

Preparation

Preparation fuels Performance

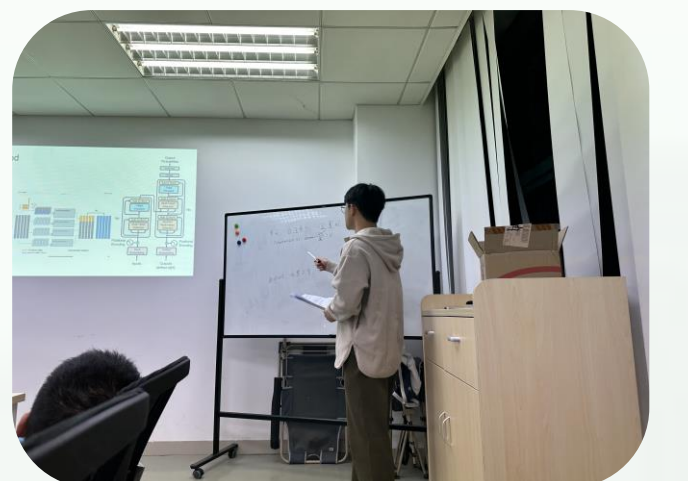
✓ Wide Breath of Practical Experience

Our **coursework curriculum** covers various HPC topics, including Parallel Programming Models, SIMD, OpenMP, MPI, CUDA, and HIP.

We have investigated parallel programming frameworks and system-related **courses** like supercomputers, computer architecture, and distributed systems.

✓ Adequate guidance

- Our HPC club benefits from **alumni** with sophisticated expertise in storage, networks, compilers, and CUDA optimization, offering valuable guidance.
- Our institution provides ample **computational resources** for us to use.



✓ Collaboration

- We have a comprehensive preparation plan with **regular courses**, **daily training**, and **weekly meetings** to prepare for the competition.
- We divide our team into **hardware/system management** and **software optimization** groups specialized in building clusters and optimizing application performance.