
ISC'23 PRELIMINARY CONTEST PROPOSAL

November 12, 2022

GeekPie_HPC Team
ShanghaiTech University

Contents

| | | |
|----------|---|-----------|
| 1 | SIST of ShanghaiTech | 3 |
| 1.1 | HPC Hardware & Software Platforms | 4 |
| 1.2 | HPC-related Courses, Training, and Groups | 4 |
| 1.2.1 | Supercomputing-related Courses | 5 |
| 1.2.2 | Supercomputing-related Training | 5 |
| 1.2.3 | Supercomputing-related Interest Groups | 5 |
| 1.3 | HPC Research and Applications | 6 |
| 1.3.1 | HPC-related Research | 6 |
| 1.3.2 | HPC Applications | 6 |
| 1.4 | Key Achievements in Supercomputing | 7 |
| 2 | GeekPie_HPC Team Introduction | 9 |
| 2.1 | History of GeekPie_HPC | 9 |
| 2.2 | Team Members | 10 |
| 2.2.1 | Team Captain Information | 12 |
| 2.2.2 | Team Centripetal Force | 12 |
| 2.2.3 | Team Demonstration | 13 |
| 2.2.3.1 | Team Initiative | 13 |
| 2.2.3.2 | Team Diversity | 13 |
| 2.2.3.3 | Team Curiosity | 13 |
| 2.2.3.4 | Team execution | 14 |
| 3 | HPC System Design | 15 |
| 3.1 | Essential Online Competition Components | 15 |
| 3.1.1 | Bridges-2 supercomputer | 15 |
| 3.1.1.1 | Normal usage similar to our chassis | 15 |
| 3.1.2 | Niagara supercomputer | 16 |

| | | |
|---------|---|----|
| 3.1.2.1 | Higher usage for memory bound application | 16 |
| 3.1.3 | HPCAC-AI Cluster Center, Thor cluster | 18 |
| 3.1.3.1 | The Migration for DPU | 18 |
| 3.2 | Essential Components | 19 |
| 3.3 | Trade-offs between Models | 20 |
| 3.3.1 | Hardware Topological Structure | 20 |
| 3.3.1.1 | AMD Zen2 Architecture | 21 |
| 3.3.1.2 | The (dis)advantages of our choice | 21 |
| 3.3.2 | Software Configuration | 22 |
| 3.3.2.1 | Stack of Software | 22 |
| 3.3.2.2 | Management of environment variables | 23 |

Chapter 1

SIST of ShanghaiTech

The School of Information Science and Technology (SIST) is one of the four departments of ShanghaiTech University. As its name reveals, SIST's mission is to educate, cultivate, nurture, and train future innovators, entrepreneurs, researchers, and leaders in the field of information science and technology. SIST integrates a world-leading scientific research environment with a life-enriching and innovative educational experience for students.



Figure 1.1: Logo of SIST, ShanghaiTech University

The school's cutting-edge research activities are designed to address global and national strategic challenges and needs of information science and technology. It promotes both fundamental and pragmatic research of the highest level by recruiting world-class scholars and researchers with an international reputation as faculty members, by collaborating broadly and closely with world-leading institutions and companies, and by taking advantages of Yangtze Delta's vibrant regional economic environment. SIST is committed to undertaking and supporting national key R&D projects with a strong aspiration for scientific and technological breakthroughs. SIST has the clear objective of becoming a regional, national, and international development workhorse by directly contributing to global technological advances with creative ideas and disruptive innovations.

1.1 HPC HARDWARE & SOFTWARE PLATFORMS

SIST provides three separate HPC platforms:

- General-purpose HPC platform aiming at teaching and research
- AI clusters for machine learning and computer vision related research
- Novel cluster for medicine finding using AlphaFold 2.

The general-purpose HPC platform consists of 84 nodes: four *high performance GPU nodes* equipped with NVIDIA Tesla V100, three *normal GPU nodes* equipped with NVIDIA M40 and NVIDIA TITAN X respectively, two *GPU fat nodes* equipped with NVIDIA Tesla K40M, eight *CPU nodes* with enhanced main memory, and eight *general computing nodes*. In total, this cluster system provides 3194 cores, 24TB memory, and 800TB storage carrying *Lustre* parallel file system. This platform also provides support for courses like Operating Systems, Parallel Computer Architecture, Machine Learning, Computer Graphics, etc.

The AI-specific platform at SIST is dedicated to machine learning and computer vision related researches. It contains various CPU / GPU nodes equipped with different levels of Intel Xeon CPUs and NVIDIA GPUs, optimized for different application scenarios encountered in machine learning and computer vision research.

As for the newly set up cluster unique for providing computation power for the Department of Biomedical Engineering (BME) to find drugs faster ever before. It requires new forms of data structure and distributed filesystem to resolve the storage of huge data. Also, since proteins consist of chains of amino acids which spontaneously fold, in a process called protein folding, to form the three dimensional structures of the proteins. To enable the AI algorithm identify parts of a larger problem, much A100s are configured for their cluster.

1.2 HPC-RELATED COURSES, TRAINING, AND GROUPS

SIST provides comprehensive courses within the field of computer systems and high-performance computing. Supercomputing-related training at ShanghaiTech University is led by the GeekPie_HPC student association, with vigorous assistance from SIST and the Library and Information Center. Other interest

groups include the iHuman Institute, the School of Life Science and Technology (SLST), and the School of Physical Science and Technology (SPST), who consecutively conduct inter-disciplinary researches across Biology, Chemistry, Physics, and Computer Science.

1.2.1 Supercomputing-related Courses

Various supercomputing-related courses are provided by the professors at SIST. The courses cover topics of computer system architectures (e.g. CS 130: Operating Systems, CS 110/210/211: Computer Architecture I/II/III), implementations and algorithms (e.g. CS 121: Parallel Computing, CS 240: Advanced Algorithms Design and Analysis), and applications (e.g. CS 181/281: Artificial Intelligence I/II, CS 171/271: Computer Vision I/II, and CS280: Deep Learning). More than 40% of the final score of the aforementioned courses come from the course projects, which require the design and implementation of real computer systems and applications. Thanks to the well established curricular system, students at ShanghaiTech have more opportunities to get hands dirty and gain precious experiences from system optimization to application developments.

1.2.2 Supercomputing-related Training

Most of the HPC-related student activities are organized by GeekPie_HPC student association. The GeekPie_HPC team provides periodic tutorials in system setup and parallel coding techniques for students who show great interest in HPC. Besides, serving as seminar co-organizers, GeekPie_HPC frequently invites experienced engineers and researchers from top-ranked HPC institutes to provide talks around up-to-date HPC achievements and the state-of-art supercomputing technologies.

1.2.3 Supercomputing-related Interest Groups

The close collaborations amongst GeekPie_HPC association, SIST, the iHuman Institute, and SLST at ShanghaiTech provides a promising method of HPC practice along with interdisciplinary research. SIST serves as a platform provider and the technical supporter, which covers most of the aspects of HPC

usage and optimization, while the iHuman Institute / SLST serves as HPC users who provide real-world demands. The excellence of research and sophisticated HPC applications from iHuman, SLST, and SPST, along with the open-minded teaching mechanisms and advanced educational program from SIST, provide students with good opportunities to apply practical supercomputer applications into their research projects.

1.3 HPC RESEARCH AND APPLICATIONS

Research topics over parallel / distributed computer systems and parallel / distributed algorithms (which eventually aims at creating more efficient and reliable high-performance computing systems) are both covered in SIST. Meanwhile, a wide range of different applications across different fields is utilizing the current HPC resources we provide.

1.3.1 HPC-related Research

On the system platform side, the Laboratory of I/O Systems and Data Science (LION), led by Prof. Shu Yin, is promoting various research projects on large-scale parallel / distributed storage systems. On the application and algorithms perspective, Prof. Rui Fan is leading a series of research projects on efficient parallel / distributed algorithms. With the combination of theory and practice, along with in-depth communication with the Library and Information Center, HPC-related research topics are paid more and more attention to and are generating better and better results. Prof. Chundong Wang, as the newly enrolled Computer Architecture P.I., is devoting his knowledge into new forms of memory media called 3DXPoint and new filesystem over this non-volatile memory. In scenario of HPC, we expect the Memory Mode, so that more application could be fit within the eADR domain [1].

1.3.2 HPC Applications

The HPC platform at SIST is the core equipment supporting research labs, whose research interest focuses on cutting-edge computational techniques

spanning across deep learning, natural language processing, large-scale images capturing and rendering, and computer vision recognition.

Besides the above artificial intelligence specific clusters, the iHuman Institute, SLST, and SPST share other general-purpose HPC clusters to receive bio-informatics support. Those research projects cover genomics, proteomics, structural biology, and theoretical biology.

1.4 KEY ACHIEVEMENTS IN SUPERCOMPUTING

Two key achievements in supercomputing research are listed as follows:

- **GPCR Project.** It was initiated in the year 2014 to help to coordinate and manage the generation of high-resolution structure-function studies of medically important proteins known as G-Protein Coupled Receptors (GPCRs) while making all data publicly available. More than 800 distinct GPCRs are found in the human body, where they carry out important cellular functions, including communication between cells and their environment. With the help of complex Molecular Dynamics simulation on a sophisticated HPC parallel simulation platform, these new findings provide clues to understand why some drugs that interact with this receptor have had unexpectedly complex and sometimes harmful effects.
- **Hard X-ray Free-Electron Lasers.** The Hard X-ray Free-Electron Lasers project (SHINE) was started on April 27, 2018, which also marks the beginning of the rapid development of the photonic science discipline. By using advanced photonic science devices represented by free-electron lasers, scientists can recognize material changes on atomic space scales and femtosecond time scales through the rich interaction of light and matter. During a hard x-ray laser experiment, up to 500 GigaBytes of data (mostly snapshots of the interact scene) can be generated within a single second. This puts huge pressure on the backend supercomputing platform, especially on the I/O system. With the help of cutting-edge I/O technologies, Prof. Huaidong Jiang and Prof. Yuhai Jiang's research groups have both made important progress in DNA single-particle imaging and atto-second quantum coherent detection, respectively. The relevant

results have been published on the well-known journals *ACS Nano* and *Physical Review Letters*.

- **AlphaFold Project.** It was initiated in the year 2020 to help to find the latest possible 3-D structure of protein. Other than a peek data usage pulse on filesystem, this project requires the persistency and scalability of the storage. Simultaneously, new forms of data storage form modified from the HDF5 is also hard to tackle.

Chapter 2

GeekPie_HPC Team Introduction

GeekPie_ is a student association founded by the first batch of undergraduate students of ShanghaiTech University. It collects all of the computer science enthusiasts at our university. As a newly established team of GeekPie_, GeekPie_HPC is less than three years old. However, inspired by the success in ASC'18 and ISC'18 student cluster competitions, many new members who are interested in supercomputing technologies have joined the team, making GeekPie_HPC a rapidly flourishing force in student supercomputing contests.

2.1 HISTORY OF GEEKPIE_HPC

The GeekPie_HPC Team is the youngest team among all the sub-groups of GeekPie_ association. It was originally founded by Zhiqiang Xie, a senior student who spent efforts on machine learning projects and realizes the importance of the HPC system supports advanced analysis applications. Thanks to the comprehensive HPC resource support from SIST and the Library and Information Center, it did not take too long to attract the attention of a decent number of students who are interested in HPC development and applications. GeekPie_HPC has participated in ASC'18, ISC'18, SC'19, SC'20, ISC'21, SC'21, ISC'22 ASC'22 and IndySCC'22. Fresh blood is injected into this vigorous team this year, making GeekPie_HPC a stronger student supercomputing organization.



Figure 2.1: Logo of GeekPie_ (left) and GeekPie_HPC (right)

2.2 TEAM MEMBERS

The core part of GeekPie_HPC is comprised of around 15 undergraduate students. Each of them is 100% enthusiastic about HPC systems, development, and applications. Six of them will participate in the ISC'23 competition: Aibo Hu, Zecheng Li, Zongze Li, Yichi Zhang, Lei Huang and Xuanjun Wen, under the supervision of Prof. Shu Yin whose research focuses on parallel and distributed computer systems, especially I/O systems. We are connected by our mutual interests over academic fields and future plan of being a practitioner in the Computer System. ShanghaiTech University is a four-year school which means senior students were enrolled in Sept. 2019, junior students were enrolled in Sept. 2020, sophomore students are enrolled in Sept. 2021 and freshman students are enrolled in Sept. 2022.

Aibo, HU is the team captain, sophomore majoring in Computer Science and Technology. He is interested in distributed systems. He currently works as a full-stack developer maintaining the ShanghaiTech mirror. In addition, he is also interested in algorithms and has won several awards in ICPC.

Zecheng, LI is a senior majoring in Computer Science and Technology at SIST, ShanghaiTech University. He has participated in former SC and ISC competitions and is experienced in HPC software tuning. He is also an intern at an HFT company working on low-latency trading systems. His research interests lie in performance analysis, parallel computer architecture, and HPC systems.

Zongze, LI is a junior majoring in Computer Science and Technology. His interests vary from Database to Operating System. He enjoys figure out every details in architecture of computer basics.

Lei, HUANG is a sophomore majoring in Computer Science and Technology in ShanghaiTech. He is interested in algorithm analysis and design. He is



Figure 2.2: Team picture of GeekPie_HPC in Nov. 2021

now learning theory of computation like formal languages and computational complexity. In addition he is an ICPC medalist in Asia Regional.

Xuanjun, WEN, sophomore, majoring in Computer Science and Technology, is interested in Linux Operations and Maintenance.

Yichi, ZHANG is a sophomore computer science major interests in operating systems, compilers, and algorithms. He has also won some awards in ICPC and worked in GeekPie_ as a backend developer.

Our First advisor, **Shu, YIN** is an Assistant Professor in the School of Information Science and Technology at ShanghaiTech University. He received the B.S. and M.S. degrees in Electronic Engineering from Wuhan University of Technology, China, in 2006 and 2008, respectively. He received a Ph.D. in Computer Science from Auburn University, USA in 2012. Before joining ShanghaiTech University in 2016, he had been serving as an assistant professor and an associate professor at Hunan University from 2012 to 2016. He worked as a post-doctoral research at the State Key Lab of HPC (HPCL) from 2015 to 2017. His research interests include parallel and distributed systems, storage systems, high-performance computing, energy-efficiency, fault tolerance, and reliability analysis. His research is supported by the National Natural Sci-

ence Foundation of China, China's Ministry of Education, and ShanghaiTech University. He has been on the program committees of various international conferences, including IEEE ICPP, IEEE IPDPS, IEEE Cluster, and IEEE NAS. Our second advisor, **Yindong, ZHANG** is an HPC engineer working for Library and Information Centre, ShanghaiTech University. He is excel at deploying scientific applications onto High Performance Computers and gain huge experience from daily maintenance of the schools' devices. **Jiajun, CHENG** is the previous captain for the team, this year, he will be the advisor for the team.

2.2.1 Team Captain Information

Name: Aibo Hu
Email: huab@shanghaitech.edu.cn
Address: 393 Middle Huaxia Road, Pudong, Shanghai, 201210
Phone: +86 182 5874 3132
Name(Chinese): 胡艾博
Address(Chinese): 上海市浦东新区华夏中路393号201210

2.2.2 Team Centripetal Force

Since their success in various supercomputing competitions, the students have completed a self-contained curriculum based on optimization of algorithms, architecture and maintenance. At the same time, the students involved in the course were taught the tool chain, such as Vim, tmux, CMake, CUDA, Grafana and so on. Among the participants, there are many students with future aspirations, and our team also extensively helps them to make contacts with industry and academia, providing them with internship and lab research opportunities. These opportunities also help our team members to grow faster and be able to compete in the future.

We have weekly meetings of the club to discuss past scientific computing topics on our unique machine. We practice things like profiling with vtune, arm forge and μ prof programs in progress, and modifying code to make it faster. We provided a wiki [2] for all the practitioners to view and a manual to quick look up during the competition, in which all pitfalls of previous members are recorded. Our team member is always ready to help those who do not come from this field and want to have some inter-disciplinary try.

2.2.3 Team Demonstration

2.2.3.1 Team Initiative

The initiative of GeekPie_HPC is to embrace a technology-neutral geek spirit. While we value winning and losing, we focus more on developing people's practical (proficient application of various tools) and communicative (good at drawing pies, taking pots and re-blooding) skills. Our team member is squeezed by a majority of their college time to maintain a GPAs and since almost all of the competition falls around mid-term and final exams, we also requires the spirit of YOLO.

It's all about how to be a good technical sharing officer to be a teammate, whether it's academic or scientific, it's all about generating value while brain storming with others. The work we do will only value us if it has value for others, just being a technical deal is useless. We hope ourselves will treasure the opportunity to work with great people, see what others are doing at slack, and then contribute what we can.

2.2.3.2 Team Diversity

Our team have diversity in aspects like focusing on different personal research interest in academic fields like Distributed IO system, Formal Execution, Cyber security and Programming Language, future plans of working or researching and extracurricular activities like reinventing wheels, designing drones, or tearing down Supercomputer and assemble them again which are good manifestations of geek's daily surroundings. The diversity of focus help us learn more from the competition.

2.2.3.3 Team Curiosity

As our team name GeekPie_HPC indicates, we are a team with geeky mind, to explore the unknown and the limit of system performance. We equipped our minds with full curiosity, which has already promoted us to mastered knowledge in varies fields such as System, Graphics and Vision. Curiosity is our ultimate motivation to put great effort in this competition.

2.2.3.4 Team execution

Though our team members are busy all the time, when there's new tasks for us, we will start working on it immediately. Whenever there's difficult problems, the seniors of the team will reply immediately and help us. With the help of our advisor, our team can realize those blueprint in our mind without procrastinating or giving up.

Chapter 3

HPC System Design

The ultimate objective of the ISC student cluster contest is to build a supercomputing cluster, which should be as efficient as possible and solves the given problems as good as possible. With that in mind, we first give a brief analysis of the performance and power estimation on all the essential components we might use to form the HPC cluster. In the following part, we are going to describe our configuration for training machine.

3.1 ESSENTIAL ONLINE COMPETITION COMPONENTS

In 2022, ISC goes virtual, we also have to care about the choice for remote performance diagnose. Last year, we've tried to run profiling tools such as Intel VTune and Arm Forge to diagnose performance issues. We also maintains our own fork of IPM. This year, we may make better use of these profiling tools to improve the performance for HPC applications.

3.1.1 Bridges-2 supercomputer

3.1.1.1 Normal usage similar to our chassis

The setup and the environment from the Pittsburgh Supercomputing Center (PSC) is pretty similar to our daily used machine. All we need is to watch the time of estimated queue. Also, since the cluster is set up in March 2021, the kernel is relatively new, so that we could deploy newer scheduler like Kubernetes, newer compilers like NVHPC 22.9 and newer tools to monitor the

job like eBPF. The scheduler is unique to newer slurm. Also, we could overclock using zenstate tool [3] on EPYC 7742.

| Category | Class | Component | Number Est. |
|-----------|------------------|--|-------------|
| CPU Nodes | Bridges-2 | CPU: AMD EPYC Rome 7742 × 2, 2.25-3.40GHz, 64 cores/ Intel Cascade Lake Xeon Gold 6248 CPUs | |
| | | Memory: 256GB, 512GB, 4TB | 488, 16, 16 |
| | | Storage: NVMe SSD 3.84TiB | |
| GPU Nodes | HCA Card | Mellanox IB ConnectX® -6 HCA card HDR Adapter | |
| | Add-ons | GPU: NVIDIA Tesla V100 (Volta) Accelerator | 24 * 8 |
| | | Mellanox IB ConnectX® QM8700 HDR Switch | |

*Key Idea: Converged HPC + AI + Data

Table 3.1: Detailed Parameters of PSC's Components

The last column of the table lies the number estimated queue line for the slurm. We could utilize the information to better balance our job queue. Custom topology optimized for data-centric HPC, AI and HPDA should be able to predefined by software so that we may be able to compose hypercube like [?] or more dynamically to avoid the fat tree communication over bands. Also, it can do community data collections and take Big Data as a Service to tailor to the latest software stack.

3.1.2 Niagara supercomputer

3.1.2.1 Higher usage for memory bound application

We found that last year, if we assign the job to higher number on the scheduler, we may possibly gain newly deployed Cascadelake. So our choice is to schedule on these jobs with this nodes as much as possible.

| Category | Class | Component | Number Est. |
|-----------|----------|--|-------------|
| CPU Nodes | Niagara | CPU: Intel Cascade Lake Xeon Gold 6148/6248 CPUs 2.4/5GHz | |
| | | Memory: 188GB | 2,016 |
| | | Storage: Lustre | |
| | HCA Card | Mellanox IB ConnexX® -5 HCA card EDR Adapter | |
| | Switch | Mellanox IB ConnexX® QSFP HDR Switch | |

*Key Idea: Energy efficiency, and Network and Storage performance and capacity.

Table 3.2: Detailed Parameters of Niagara's Components

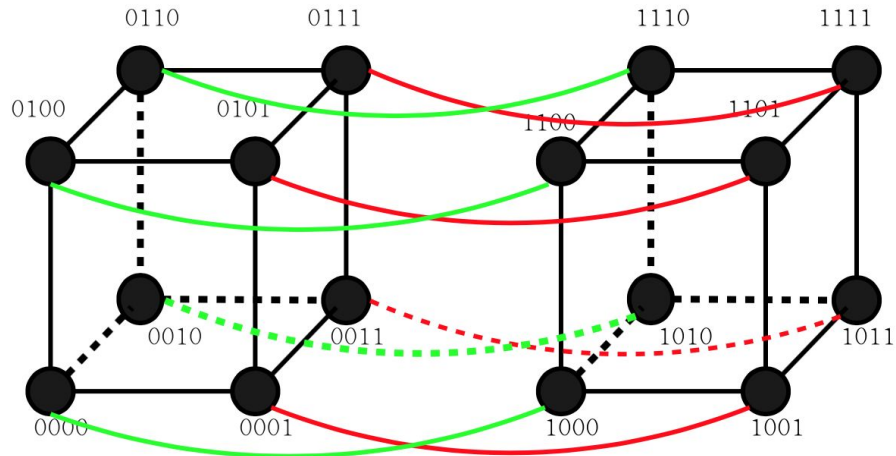


Figure 3.1: Topology of proposed cluster

3.1.3 HPCAC-AI Cluster Center, Thor cluster

3.1.3.1 The Migration for DPU

| Category | Class | Component | Number Est. |
|-----------|----------|--|-------------|
| CPU Nodes | Thor | CPU: 16-core CPUs E5-2697A V4 2.60GHz | |
| | | Memory: 256GB | 36 |
| | | Storage: 1TB 7.2K RPM SATA | |
| | HCA Card | Mellanox IB ConnexX® -6 HCA card HDR Adapter | |
| | | Nidia DPU, BlueField-2 SoC | |
| | Switch | Mellanox IB ConnexX® QM7800 HDR Switch | |

*Key Idea: Novel hardware, novel choice.

Table 3.3: Detailed Parameters of HPCAC-AI's Components

DPU is a new type of programmable processor that combines three key elements. DPU is a SOC (System On Chip) that combines. An industry standard, high performance and software programmable multi-core CPU, often based on the already widely used Arm architecture, closely matched to its SOC components. High performance network interfaces that can parse, process and efficiently transfer data to the GPU and CPU at wire speed or at the speed available in the network. A variety of flexible and programmable acceleration engines that can offload applications such as AI, machine learning, security, telecoms and storage, and boost performance. All of these DPU capabilities are critical to enabling secure, bare-bones performance, native cloud computing for the next generation of large-scale computing on the cloud. We are going to migrate some application tailoring to this device.

3.2 ESSENTIAL COMPONENTS

This year, our training system for the contest will be based on the SUPER-MICRO server [4], plus NVIDIA GPU add-ons without NV-Link [5]. Detailed parameters and power consumption estimations about all the essential components are summarized in Table 3.4 below. All the components in this table, along with the estimation of their power consumption, will be the factors to consider when designing the power-sensitive HPC cluster for ASC'20 competition.

| Category | Class | Component | Power Est. |
|----------|---------------------------|---|---------------|
| Server | SUPER-MICRO 4124GS-TNR | CPU: AMD EPYC Rome 7742 × 2, 2.25 GHz, 64 cores | 225W TDP* × 2 |
| | | Memory: 32GiB × 32, DDR4, 3200 MHz | 9 W |
| | | Storage: SATA SSD 1.9TiB | 10 W |
| | HCA Card | Mellanox IB ConnectX® -5 HCA card, single port QSFP HDR | 9 W |
| | Add-ons | GPU: NVIDIA Tesla V100 (Volta) Accelerator | 250W TDP |
| | | Storage: PCIe SSD 256GB | 10 W |
| Switch | GbE | 10/100/1000 Mbps, 24 ports Ethernet | 30 W |
| | FDR-IB | SwitichX™, 36 QSFP HDR Infini-Band | 130 W |

**Thermal Designed Power* (TDP): the maximum amount of heat generated by a computer chip or component (often a CPU, GPU or system on a chip) that the cooling system within is designed to dissipate under any workload. [6]

Table 3.4: Detailed Parameters of Essential Components

As the real source of computation power, the performance of CPUs and GPUs in this system should also be considered. Theoretical performance of

the given AMD EPYC Rome 7742 CPUs [7] and NVIDIA Tesla V100 GPUs are summarized in Table 3.5 below.

| Item | Estimation |
|---|--------------------------|
| CPU: AMD EPYC Rome 7742, 2.25GHz, 64 cores | 2.3 TFLOPs / <i>chip</i> |
| GPU: NVIDIA Tesla V100 (<i>Volta</i>) | 7.8 TFLOPs / <i>card</i> |

Table 3.5: Theoretical Performance of Given CPUs and GPUs

3.3 TRADE-OFFS BETWEEN MODELS

The above three possible configurations both have their own advantages and disadvantages considering efficiency, CPU / GPU balance, the danger of power exceeding, etc.

3.3.1 Hardware Topological Structure

We finally fixed our decision on the second option in Section 3.3, which is comprised of 2 nodes, each of which equips with a dual-socket AMD EPYC Rome 7742 and one NVIDIA Tesla V100 GPU card. Figure 3.2 demonstrate our system's configuration.

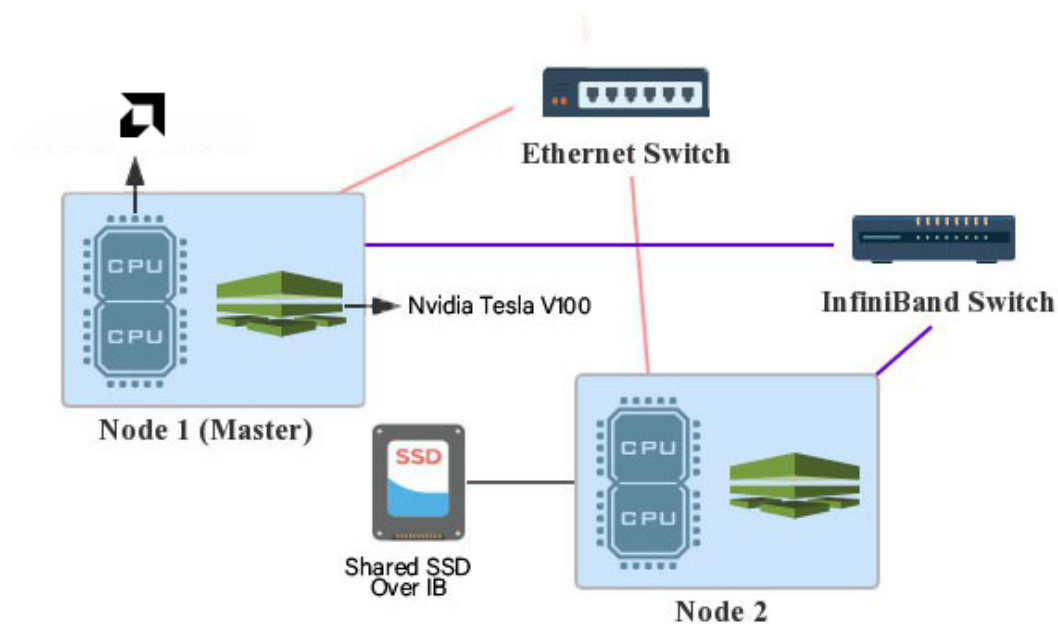


Figure 3.2: The Configuration of Our System

Both nodes are connected via both Gigabit Ethernet and InfiniBand. Each node extends one Ethernet cable to the Ethernet switch as well as one InfiniBand cable to the InfiniBand switch (called *Star-shape Topology*). The Ethernet switch functions as an auxiliary route of inter-connection, for communications that do not need to go through IB. Applications use InfiniBand for minimum communication latency on computations.

3.3.1.1 AMD Zen2 Architecture

Since years before, we have witnessed the leap-frog process of Zen. The Zen2 Architecture makes a huge jump in IPC, CPU-to-Memory Bandwidth, NUMA performance, and PCIe4.0 to outer devices. The "Zeppelin" Die applies the modular design approach, each die is connected by DDR+IFOP and DDR+IFIS, in total, 4 layers package routing. All the die has muxing IO access to the PCIe lane and voltage variation between MCM cores. For dual-socket AMD system, it has a similar data routing topology as inside one.

3.3.1.2 The (dis)advantages of our choice

This architecture configuration has the following pros (+) & cons (-):

- + Relatively high (and sufficient) theoretical performance (68.55 TFLOPs)
- + Mild power consumption (2823 W), neither too high causing the danger of breaking the 3,000 W restriction, nor too low leaving wasted computation power
- + Balances CPU and GPU computation power, therefore able to serve CPU intensive task and GPU intensive task.
- Not as energy-saving and energy-efficient as the dual GPU configuration
- Possibly slightly lower performance for CPU-intensive applications

Nevertheless, this will be the most suitable choice for us.

3.3.2 Software Configuration

The software environment built upon the system will include at least the following elements:

3.3.2.1 Stack of Software

- **Operating System:** Ubuntu 22.04.1 LTS.
- **Libraries:**
 - OpenMPI, MPICH, IntelMPI, MVAICH2
 - OpenBLAS, cuBlas, IntelMKL, AOCL-BLIS
- **Build tools:**
 - GNU Compiler Collection (GCC 11.3), Intel ICC (from *OneAPI 2022, cluster edition), AMD AOCC 3.2
 - NVIDIA CUDA Toolkit 11.8
 - AMDNetLib, AMDlibFlames, MKL
- **Deep Learning:** Anaconda, PyTorch Framework, etc.

3.3.2.2 Management of environment variables

Management of environment variables are achieved by using *Spack* [8], a useful tool based on *Environment Modules* and *Python*. These software environment tools will support the successful and efficient execution of all the tasks required.

Bibliography

- [1] Intel, “Intel optane memory technology.” [Online]. Available: <https://www.intel.com/content/www/us/en/developer/articles/technical/eadr-new-opportunities-for-persistent-memory-applications.html>
- [2] GeekPie_HPC, “Geekpie_hpc wiki.” [Online]. Available: <https://hpc.geekpie.club/wiki/>
- [3] Github, “Zenstate overclocking tool.” [Online]. Available: <https://github.com/geekpiehpc/ZenStates-Linux>
- [4] “Supermicro 4124gs-tnr datasheet.” [Online]. Available: <https://www.supermicro.com/en/Aplus/system/4U/4124/AS-4124GS-TNR.cfm>
- [5] “Nvlink and nvswith.” [Online]. Available: <https://www.nvidia.com/en-gb/data-center/nvlink/>
- [6] Wikipedia, “Thermal design power.” [Online]. Available: https://en.wikipedia.org/wiki/Thermal_design_power
- [7] “Epyc-rome.” [Online]. Available: <https://www.hpcwire.com/2019/08/08/amd-launches-epyc-rome-first-7nm-cpu/>
- [8] “package manager.” [Online]. Available: <https://spack.io>