ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

A Survey of High-Performance Computing (HPC) Infrastructure in Thailand

Vara Varavithya[1] and Supakit Prueksaaroon[2]

ABSTRACT

In the last decade, government organizations and private companies in Thailand have invested considerably in computing resources. Research collaborations typically band together based on shared interests and propose projects to compete for funding. Thus, many institutions frequently use similar computing resources. A bird's eye view of high-performance computer infrastructure in Thailand is important in many ways. To the best of our knowledge and ability, we gathered information on government procurements of HPC resources in the past five years, which cover several organizations and ministries. We list the system specifications and tabulate the target application areas for each system. The aggregated number of cores and storage space for HPC in Thailand, commissioned during the past five years, is 54,838 cores and 21 PB, respectively. We also describe the large data transfers using UniNet for HPC applications. The survey results can be used by academics and decision-makers to build research agendas and national development strategies.

Article information:

Keywords: Digital Infrastructure, GPU Computing, High-performance Computing

Article history:
Received: January 23, 2023
Revised: April 1, 2022
Accepted: June 3, 2023
Published: June 17, 2023 (Online)

DOI: 10.37936/ecti-cit.2023172.251440

1. INTRODUCTION

Almost every research activity involves the processing of information, making digital infrastructure crucial for research. Not only do the theoretical and applied sciences heavily rely on computing services, but with the explosion of social data, the social sciences are also rapidly embracing the analytical capabilities of big-data services. These services comprise computing cycles, storage space, and network bandwidth. The amount of time required to complete a particular task strongly relates to the complexity and size of problems that can be processed [1]. High-performance computing (HPC) resources are, therefore, considered an essential part of the digital infrastructure for research.

Large-scale computing systems that are housed in a data center typically offer both computational and storage capabilities. Five decades of HPC computing have seen the evolution of mainframes, minicomputers, vector computers, supercomputers, massively parallel computers, and computer clusters. The network infrastructure is responsible for remote access and the movement of data. Most current-generation supercomputers are built as computer clusters with graphic processing units (GPUs) as co-processors interconnected with high-speed networks and storage systems [2].

The TOP500 list (www.top500.org) [3] compiles a list of the 500 highest-performing computing systems based on their ability to solve dense linear equations. With a performance of 1,102 peta floating point operations per second (PFlop/s), the Frontier System at the Department of Energy Oak Ridge National Laboratory in the United States came in first in the most recent ranking, which was completed in November 2022. The Fugaku supercomputer in Japan [4] came in second with a performance of 442 PFlop/s, as determined by the HPL benchmark [5]. Most systems in the TOP500 list have similar characteristics, adapted from the Beowulf approach [6].

In Thailand, the public and private sectors have made significant financial investments in HPC resources to meet their needs. The government's ongoing commitment to supporting HPC infrastructure has been demonstrated by the most recent establishment of the NSTDA Supercomputer Center (ThaiSC)

[1] The author is with Faculty of Digital Technology, Chitralada Technology Institute, E-mail: vara@cdti.ac.th
[2] The author is with Department of Electrical and Computer Engineering, Faculty of Engineering, Thammasat University, E-mail: psupakit@engr.tu.ac.th
[1] The corresponding author: vara@cdti.ac.th

[7]. Based on this effort, ThaiSC's LANTA, which at the time held the top spot in ASEAN, was placed 70th in the Top 500 in November 2022 with a Linpack performance of 8.15 PFlop/s (RMAX).

Research collaborations typically band together based on shared interests and propose projects to compete for funding. Thus, many institutions frequently use similar computing resources. A bird's eye view of high-performance computer infrastructure in Thailand is important in many ways. Researchers can better comprehend the complexity and scalability of problems that can be handled without employing cloud computing services by first assessing the total computer resources available at the national level. Second, a directory of HPC resources in the nation will promote knowledge sharing about the operation and maintenance of HPC systems as well as the sharing of computing resources.

For this paper, we gathered information on government procurements of HPC resources in the past five years. We present the characteristics of both the hardware and software systems and tabulate the target application areas for each system in this study. We further group the application areas and address large data transfers between HPC facilities. The structure of this paper is as follows. Section 2 provides an explanation of historical views. Thailand's current HPC resources are presented in Section 3. Section 4 addresses the challenges with data transport between HPC facilities. The opportunities and challenges of HPC in Thailand are discussed in Section 5. Finally, concluding remarks are drawn in Section 6.

2. HISTORICAL PERSPECTIVES

In the late 1980s, computational sciences emerged as a new field of research. In 1991, the National Electronics and Computer Technology Center (NECTEC) established a High Performance Computing Center (HPCC) to support high-performance computing research in Thailand. In August 1992, Thailand gained access to the Internet for the first time, which further advanced the research. In the early days, researchers used high-performance workstations as their workhorses to perform complex computations. These workstations included DEC Alpha, HP Apollo, IBM RS6000, and SUN SPARC.

Several small-scale HPC systems were put into operation in Thailand in the 1990s. The HPCC acquired CRAY EL98 and an 8-processor SGI Power Challenge XL. Kasetsart University (KU) and King Mongkut's University of Technology Thonburi (KMUTT) each had small-scale IBM SP2 systems installed. King Mongkut's Institute of Technology Ladkrabang (KMITL) acquired an 8-processor Convex system. In addition, the Thai Meteorological Department (TMD) procured a sizable 38-processor, 128 Mbytes/node IBM SP2 system for weather forecasting. During this time, computational chemistry, computational physics, and computer science and engineering were the dominant fields of study. The details of early efforts to address the usage of HPC in research can be found in [8]. The 1997 EECON paper [8] by Associate Professor Yuen Poovarawan and the late Assistant Professor Dr. Putchong Uthayopas concluded with the following statement:

"As Thailand moves toward the year 2000, high-performance computing will have an increasing role since it is one of the key enabling technologies that enable us to explore unknown technological territories that were not possible before."

The concept of the BEOWULF cluster [6] was proposed in 1995. In 1997, the SMILE project [9] started the boom of Beowulf Clusters in Thailand. In 1999, KU built the 72-node PIRUN Beowulf Cluster [10]. With this facility, the research group from KU pursued system software development, which included the cluster management SMILE Cluster Management System (SCMS) toolkit [11] and the batch scheduler Simple Queue Management System (SQMS) toolkit [12]. The subsequent versions of these software toolkits were integrated to be a part of the NPACI Rocks cluster distribution.

In the first decade of the new millennium, grid computing [13] became a new computing paradigm for creating virtual HPC infrastructure. The global research community showed ambitious efforts in sharing HPC resources, which streamlined processes of authentication, single sign-on, job scheduling and monitoring, data transfer, and security. Thailand entered the grid computing era with the establishment of the ThaiGrid project in 2001 between KU and King Mongkut's University of Technology North Bangkok (KMUTNB). In the later years, the members of the ThaiGrid project included the Asian Institute of Technology (AIT), Chulalongkorn University (CU), KU, KMIUNB, KMUTT, KMITL, Mahidol University (MU), Silpakorn University (SU), TMD, Suranaree University of Technology (SUT), Chiangmai University (CMU), Khon Kaen University (KKU), Walailak University (WU), and Prince of Songkla University (PSU).

Similar to grid computing initiatives such as EuroGrid, TeraGrid, CNGrid, and k*Grid, the Thai National Grid Project [14–17] was launched in 2006–2008 by the Ministry of Information and Communication Technology in Thailand. The project's objectives were to implement grid technology and develop a large-scale computing infrastructure. A national center called the Thai National Grid Center (TNGC) was established to supervise and manage the building of national grid infrastructure. Furthermore, a community of research and education institutes was created to pursue the human resources development and research components. The HPC infrastructure of the TNGC consisted of two parts, a tera-scale cluster and satellite clusters. The teraflops cluster [15] had 192 computing nodes in the system. Each computing node consisted of two dual-core Intel Xeon 3.0 GHz processors with 4 GB of main memory (32 nodes had 8 GB of main memory). The teraflops cluster could be accessed using four front-end nodes. Four high-end file servers provide storage of five terabytes. The satellite clusters included a set of small compute clusters distributed among 14 participating institutes. Each satellite cluster had five dual-core Intel Xeon 2.8 GHz servers with 4 GB of main memory. In the TNGC project, Globus [13] was adopted as a standard middleware. A virtual research laboratory initiative, Access Grid, was launched [18] to allow the members of the TNGC project to collaborate on their research via teleconferencing. A custom ThaiGrid portal [19] was developed to manage the TNGC infrastructure.

By the end of 2008, 21 clusters from 16 sites were connected on the ThaiGrid, with a total of more than 1,000 processor cores. Unfortunately, the TNGC project encountered budgeting challenges and was discontinued in 2010. However, the lessons learned from the TNGC project are valuable for designers and grid practitioners in experiencing large-scale computing at enterprise or campus levels. Additional contemporary HPC resources include the 284-core SILA cluster at Ramkhamhaeng University, the 256-core cluster at the National Center for Genetic Engineering and Biotechnology (BIOTEC), and the 240-core cluster at the KMUTNB. The National e-Science Infrastructure Consortium was established in 2011 to promote the collaborative development of research infrastructure. A group of universities and research institutes participated in the consortium. The goal was to provide the members with HPC services, storage systems for large datasets, scientific applications, and networking opportunities. The target areas included computational science and engineering, computer science and engineering, water resource management, energy and environment management, climate change, and high-energy physics. The aggregated resources of the consortium included 3,424 processor cores and 1,171 TB of storage.

After two decades, Thailand's HPC community has steadily developed with a wide range of applications, particularly in AI and the biological sciences. In the country, significant HPC systems have recently started operating. The following section contains a list of HPC systems acquired in Thailand over the past five years.

3. CURRENT STATUS OF HPC COMPUTING INFRASTRUCTURE IN THAILAND

We gathered public information about HPC resources that went through government procurement processes in Thailand. Table 1 lists the organizations using HPC resources in Thailand in 2022. Although by no means comprehensive, the list, to the best of our knowledge, includes the majority of HPC resources that are readily available. The HPC resources are scattered around the country in 13 universities, 4 research institutes, 3 government agencies, 1 state enterprise (the Electricity Generating Authority of Thailand (EGAT)), and 1 public company (the Siam Cement Group (SCG)). Out of these, the Ministry of Higher Education, Science, Research and Innovation (MHESI) and the Ministry of Digital Economy and Society (MDES) are responsible for operating the majority of the HPC resources. Table 2 presents the important characteristics of the HPC systems. The resources can be classified into two classes: HPC clusters with GPUs and NVIDIA DGX servers. The HPC clusters use high-end servers as their computing nodes. Intel Xeon and AMD processors are the most common in this class. On the other hand, the DGX servers use AMD processors with 64 processing cores. Some systems have additional "FAT nodes," with higher processing capability and memory size. Most larger systems in both classes use InfiniBand as an interconnection, while smaller systems employ Ethernet. The total number of cores and aggregated storage space are 54,838 and 21 PB, respectively. The NVIDIA V100 and A100 are the most common accelerators used, and their total number is close to one thousand.

ThaiSC's LANTA has the largest number of processing cores (33,024 cores), which can deliver performance at the 8.15 PFlop/s. The details of the commissioning date, the operating system installed on each system, and the application area are shown in Table 3. The LINUX OS family dominates the operating systems for high-performance computer systems. In most installations, the Slurm cluster management and job scheduling system has been used. The application areas can be categorized as follow:

- HPC services: parallel processing, Message Passing Interface (MPI), Open Multi-Processing (OpenMP), batch processing, and high-performance calculation libraries;
- Engineering designs: finite element calculation, ANSYS, CAD/CAM, computational fluid dynamics, mechanical modeling and simulation, Groningen MAchine for Chemical Simulations (GROMACS);
- Basic sciences: astronomy, high-energy physics, multi-physics, molecular dynamics, molecular modeling, Sybyl-X (molecular modeling and simulation suite), auto-dock, CERN root, computational chemistry, and Vienna Ab-initio Simulation Package (VASP);
- Weather forecast: Weather Research and Forecasting Model (WRF), WRF model coupled with Chemistry (WRF-Chem), and Regional Atmospheric Modeling System (RAMS);

Table 1: *A List of Organizations Operating HPC Resources in Thailand in 2022.*

Organization	Abbrivation	Organization	Abbrivation
Burapha University (under the Genomic Thailand Project, NBT)	BUU	National Astronomical Research Institute of Thailand	NARIT
Chitralada Technology Institute	CDTI	National Biobank of Thailand	NBT
Chulalongkorn University Technology	CU Consortium	National e-Science Infrastructure	KRYPTON Center
CMKL University (KMITL)	CMKL	NSTDA Supercomputer Center (ThaiSC)	LANTA, TARA
Electricity Generating Authority of Thailand	EGAT	Prince of Songkla University	PSU
Faculty of Medicine, Siriraj Hospital, Mahidol University	SIRIRACH	Siam Cement Group (SCG)	SCG
Faculty of Medicine, Chulalongkorn University	Med-CU	Suranaree University of Technology	SUT
Hydro-Informatics Institute	HII	Synchrotron Light Research Institute	SLRI
Kasetsart University	KU-WATA II	Thai Meteorological Department	TMD
Kasetsart University, Sriracha Campus	Poseidon AI	Thammasat University	TU
King Mongkut University of Technology Thonburi	KMUTT	Vidyasirimedhi Institute of Science and Technology	VISTEC
Khon Kaen University	KKU		
			Total of 23 organizations, 24 systems

Table 2: *Characteristics of HPC Resources in Thailand*

HPC Resource	♯ of nodes	♯ of cores	GPUs	Memory/ Node	Total Memory	Total Storage	Interconnections
BUU	1	224	N/A	3 TB	3 TB	1 PB	100 Gbps InfiniBand
CDTI	7	224	2 x T4 NVDIA	128 GB	896 GB	56 TB	10 Gbps Ethernet
CU	2	256	6 x T4, 16 x A100 NVIDIA	1 TB	2 TB	300 TB	200 Gbps InfiniBand
CMKL	6	768	48 x A100 NVIDIA, 1 x Phi	1 TB	6 TB	500 TB	200 Gbps InfiniBand
EGAT	13	1,008	1 x Xeon Phi	448 GB 704 GB	6,080 GB	200 TB	100 Gbps InfiniBand
SIRIRACH	2	256	16 x A100 NVIDIA	1 TB	2 TB	500 TB	200 Gbps InfiniBand
Med-CU	9	1,456	12 x A100 NVIDIA	256 GB 512 GB 1 TB	13 TB	2 PB	100 Gbps InfiniBand
HII	24	864	6 x V100, 3 x A100 NVIDIA	384 GB	9,216 GB	1.1 PB	100 Gbps InfiniBand
KMUTT	14	272	8 x V100	32GB	384GB	48TB	56 Gbps InfiniBand
KU-WATA II	12	240	N/A	256 GB	3,072 GB	46 TB	2x10 Gbps Ethernet
Poseidon AI	1	40	8 xV100 NVIDIA	512 GB	512 GB	120 TB	100 Gbps InfiniBand
KKU	7	396	4 x V100, 8 x A100 NVIDIA	1 TB	3,328 GB	50 TB	10 Gbps Ethernet, 200 Gbps InfiniBand
NARIT	38	1,600	12 x V100 NVIDIA	512GB	11 TB	90 TB	100 Gbps InfiniBand
NBT	4	640	16 x A100 NVIDIA	3 TB 1 TB	8 TB	3.3 PB	100 Gbps InfiniBand
KRYPTON	5	1,008	N/A	768 GB	3,840 GB	200 TB	100 Gbps InfiniBand
LANTA	160 176 10	20,480 11,264 1,280	Computing Nodes GPU 704 x A100 NVIDIA FAT Nodes	256 GB 512 GB 4 TB	104.4 TB	10 PB	200 Gbps InfiniBand
TARA	68	4,320	28 x V100 NVIDIA	192 GB 3 TB	36 TB	776 TB	100 Gbps InfiniBand
PSU	2	256	16 x A100 NVIDIA	1 TB	2 TB	550 TB	200 Gbps InfiniBand
SCG	26	1,144	1 x V100 NVIDIA	192 GB	5 TB	50 TB	56 Gbps InfiniBand
SUT	8	256	N/A	96 GB	768 GB	10 TB	56 Gbps InfiniBand
SLRI	18	168	N/A	48 GB	792 GB	90 TB	100 Gbps InfiniBand
TMD	172	5,504	16 x K80 NVIDIA, 8 x Xeon Phi	128 GB	22 TB	120 TB	100 Gbps Intel OPA
TU	6	96	6 x Geforce 1060Ti NVIDIA	64 GB	384 GB	10 TB	1 Gbps Ethernet
VISTEC	24	800	78 x V100, 2 x A100 NVIDIA	384 GB 768 GB 1 TB	6,144 GB 3,072 GB 2,048 GB	78 TB	100 Gbps InfiniBand
Total	**815**	**54,838**	**157 x V100, 825 x A100**		**251.64 TB**	**21,194 TB**	

Table 3: *Operating Systems and Applications used in the 2022 HPC Resources in Thailand.*

HPC Resource	Commission Date	Operating System	Applications
BUU	2021	CentOS	Genome processing
CDTI	2022	CentOS	AI, ML, TensorFlow, Big-Data
CU	2021	Ubuntu	AI, ML, TensorFlow
CMKL	2020	Ubuntu	AI, ML, TensorFlow, NLP, AI for health, AI for food, and AI for tourism
EGAT	2017	CentOS	Weather forecast, RAMS, calculated wind speeds for air pollution monitoring
SIRIRACH	2021	Ubuntu	AI, ML, TensorFlow, Clara Parabricks, genome sequencing
Med-CU	2018-2022	CentOS, Ubuntu	AI, ML, TensorFlow, OMICS, genome sequencing
HII	2020	CentOS	Weather forecast, WRF, rainfall forecasting, and water prediction for water management
KMUTT	2019	CentOS	AI, ML, Robotics
KU-WATA II	2019	CentOS	Computational fluid dynamics, molecular dynamics, mechanical modeling & simulation, OpenHPC 1.3, Open MPI, SGE, OpenOnDemand, CAD/CAM, ANSYS
Poseidon AI	2020	Ubuntu	AI, ML, TensorFlow, ANSYS, Slurm, Jupyter, GlusterFS
KKU	2021	Ubuntu	AI, ML, TensorFlow, GROMACS, VASP, Gaussian, Jupyter
NARIT	2019	CentOS	Astronomy
KRYPTON		CentOS	Life sciences, gaussian, Autodock, Sybyl-X
LANTA	2022	Cray OS (SuSe base)	Life sciences, chemistry, WRF-Chem
TARA	2019	CentOS	Life sciences, chemistry, WRF-Chem
PSU	2021	Ubuntu	AI, ML, TensorFlow, image/video detection, smart city
NBT	2019	CentOS, RHEL	Life sciences, genome processing, clara parabricks, variant annotations,
SCG	2019	CentOS	Multi-physics, ANSYS, finite element calculation
SUT	2011	CentOS	High-energy physics, CERN root
SLRI	2019	CentOS	Synchrotron applications, Vienna Ab initio simulation package
TMD	2017	CentOS	Weather forecast, WRF, rainfall forecasting
TU	2018	CentOS	Computational chemistry, molecular docking (drug discovery)
VISTEC	2018	CentOS	AI for health sciences, molecular dynamics, chemistry

- Artificial Intelligence (AI): Machine Learning (ML), Natural Language Processing (NLP), and TensorFlow;
- Life sciences: genome processing, Clara Parabricks, and molecular docking.

Figure 1 shows the number of processing cores and storage space, classified by application areas. The diameter of the dots reflects the relative number of cores to 54,838 cores. Not including HPC Services, the life sciences area has more systems deployed in different organizations. It is also observed that HPC resource allocation is dominated by resources from the LANTA system which spans over HPC services, basic sciences, life sciences, and weather forecast.

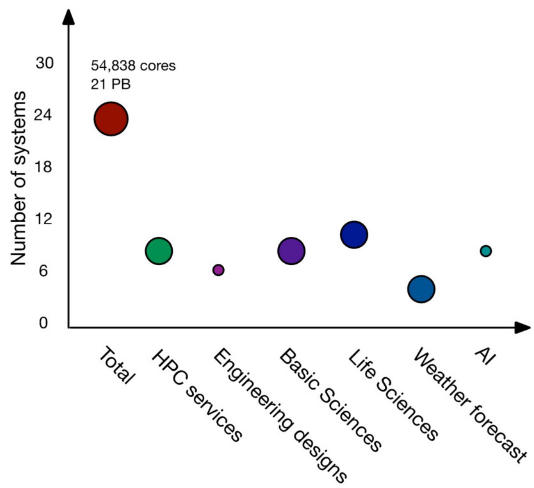

Fig.1: *The Number of Systems classified by the Application Areas.*

4. TRANSFER OF LARGE DATA FOR HPC APPLICATIONS

A certain class of applications involves transferring data between facilities. Large amounts of data come from science instruments, IoT devices, and social media. In recent years, new datasets have originated and expanded at a rapid rate, such as data from SLRI, TMD, and NBT. The processes of collecting, staging, processing, and visualizing these data require high-speed networks. At the national and international levels, REN programs are driven by a number of government initiatives, including Internet2, ESnet, DFN, JGN, MYREN, APAN, GE ANT, and TEIN. In Thailand, the Royal Thai Government has funded this endeavor under the UniNet project.

The UniNet network has started providing network services for higher education and research organizations since 1996. The backbone bandwidth has evolved from T1, OC3, OC12, OC48, 10 Gbps Ethernets to 100 Gbps Ethernet. Figure 2 shows the UniNet network topology that covers higher education and research institutes. The names of the provinces and the link bandwidth between them are shown. Multiple 10 Gbps links can be connected between HPC centers using UniNet services.

One reason that the high throughput requirements cannot be met is because of the best-effort service model of the Internet. A series of network and security devices along network paths creates bottlenecks that hamper fast data movement between two endpoints. Several techniques have been proposed to alleviate this problem, such as using overlay networks, dedicated links, L2-VPN, and QoS mechanisms.

The demilitarized zone (DMZ) concept has been adopted as the network pattern to provide a security isolation zone for Internet services and, later, for high-performance network services for researchers. In [21], six institutes participated in the UniNet DMZ project.

The software-defined networks could be used as overlay networks to link HPC facilities where operators have more control over performance and security for different requirements. An effective throughput of 700–900 Mbps between two research institutes was achieved, which is about one order of magnitude higher than network throughput in a campus environment. In [22], a transfer rate of up to 7.2 Gbps via the UniNet links using multi-stream parallel data transfer was demonstrated.

In the current production environment (April 2022), the NBT reported data transfer results between BUU and NSTDA (173 km), as shown in Figure 3. The parallel sessions of rsync were used to transfer human genome data from the sequencing center (BUU) to the genome data bank at NBT, NSTDA, for data processing. NBT is responsible for providing the Genomics Thailand ecosystem [23], which processes and sequences 200 Thai WGS per week. A total of 24 TB of data is transferred per week at a transfer rate of 4–6 Gbps. With this rate, it takes around 2.5 hours to transfer 5 TB of data. This result shows the potential for moving big data using Thailand's research and education network, the UniNet.

5. CHALLENGES AND OPPORTUNITIES

In this paper, not only did we collect data on HPC infrastructure, but we also conducted interviews and questionnaires with HPC experts in the country to obtain their perspectives on challenges and opportunities. We classify challenges and possibilities based on three factors: requirements, operations, and human resource development (HRD).

5.1 CREATING NEEDS

As observed in Section 3, HPC resources can be divided into three distinct categories: small clusters, application-specific resources, and large-scale HPC facilities. Small clusters in universities support the responsibilities of research and education. To scale for larger problems, researchers can either request ThaiSC resources or subscribe to public cloud services, which means that Thai researchers now have greater access to computational resources than they did in the past. To maintain a sustainable HPC infrastructure, however, a certain level of economy of scale must be achieved, with the community fostering new demands for HPC in various fields. Legacy HPC applications include scientific and engineering simulations, such as molecular dynamics, biology, chemistry, material sciences, finite element computation, and weather forecasting. Publicizing the value of HPC as a tool for decision-making requires an emphasis on applications that could help with problem-solving with a societal impact and enhancing the quality of life. To move forward, the community must collaborate on the development of requirements and use cases for new application areas, such as supporting digital services in clean air and water, smart cities, digital government, city planning, disaster planning and safety, digital twins, finance, logistics, medical sciences, and genomic medicine. Consequently, generating demand for HPC services is one of the main challenges that will help bring in new opportunities for sustaining HPC infrastructure.

The ThaiSc is the third significant HPC initiative in Thailand, following NECTEC's HPCC (1992) and ThaiGrid (2001). The results of the interviews show the HPC community's high regard for ThaiSc, particularly its ranking in the Top500, an important achievement in the development of HPC infrastructure. ThaiSc focuses on founding structures to provide a sustainable funding mechanism, which includes infrastructure ecosystem management models, close collaboration between funding agencies, resource providers, and users, and the promotion of HPC activities.

The recent projects supported by ThaiSC, such as those related to the COVID-19 pandemic and PM2.5 pollution forecasting, have demonstrated the significance of HPC as one of our national infrastructures, as well as the importance of HRDs, such as the provision of resources for the super AI engineer project and the ThaiSC pioneer program. The Genomics Thailand Project is another noteworthy HPC endeavor that demonstrates the potential for using genomic data in diagnosis. In the near future, healthcare providers may rely on genomic data, making HPC infrastructure an essential national infrastructure.

5.2 OPERATING INFRASTRUCTURE

The second difficulty involves infrastructure operations. Fluctuating HPC workload demands result in low utilization. In terms of return on investment, low-utilization systems will receive less budgetary support, while the average service time of computing resources is approximately seven years.

The cloud economy is far larger than the computer industry. In recent years, differences between on-

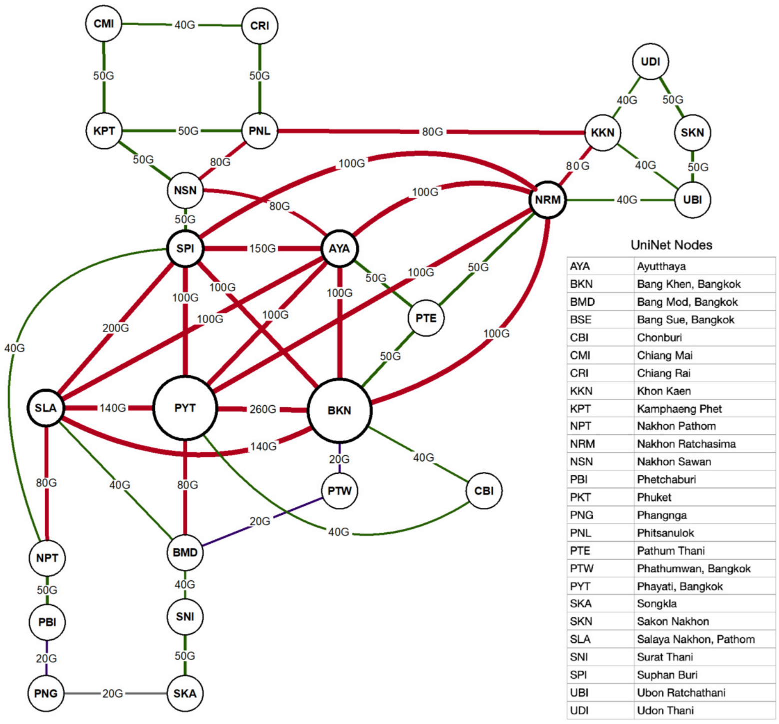

Fig.2: *The UniNet Backbone Network Operates at 100 Gbps. The Bandwidth of 50 Gbps Covers all Major Cities in the Country.*

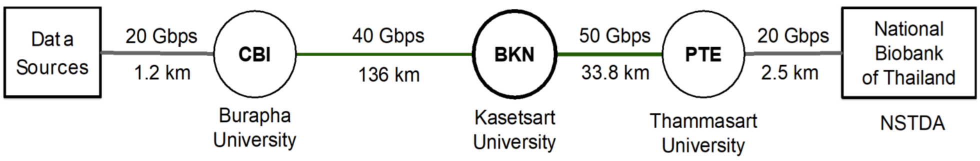

Fig.3: *The Network Path between Two Research Facilities to Demonstrate Large-Scale Data Transfer using the UniNet.*

premises HPC and cloud services have become narrower [28]. Cloud-based services have begun entering the HPC sector. In addition to on-demand access to a cloud provider's computing resources, HPC applications can benefit from a new form of application, namely cloud-native applications. In terms of scalability, service diversity, and cost, the cloud offers greater flexibility in creating HPC infrastructure computational environments. All main cloud providers, including Amazon Web Services, Microsoft Azure, and Google, now offer HPC services as standard cloud services. AWS provides a variety of services that can be utilized in HPC applications, such as EC2, Elastic Fabric Adaptor, AWS Parallel Cluster, Amazon FSx for Lustre, and AWS Batch. In native HPC cloud deployment, organizations may consider moving their existing HPC workload into the cloud environment. The hybrid HPC cloud environment provides the organization with augmented resources to accommodate the increase in HPC workload demands.

In the future, as the global trend [2] persists, the majority of HPC resources will adopt the cloud computing paradigm. To acquire next-generation HPC facilities, the Thailand HPC community must collaborate closely with cloud providers, even though it is becoming apparent that cloud services for HPC are not inexpensive.

Private sectors are more concerned with the low utilization of on-premise HPC resources, so subscription HPC cloud platforms (public cloud services with GPU, distributed file systems, and InfiniBand) have become attractive. Unfortunately, for government organizations, procurement directives have not yet taken cloud services into account in the same manner as utilities (such as water and electricity). Due to this, it is presently difficult to manage the pay-per-use model with government funding. However, government organizations have worked toward more flexible subscription procurement models.

Small HPC clusters installed on-site provide various benefits for educational and proof-of-concept applications. Research organizations that own such systems must devote resources—both financial and human—to the upkeep of a sophisticated computational environment. Some research groups have to maintain their HPC asset over concentrating on research as it consumes a large portion of their resources.

Different computational resources are needed for different applications. For instance, although bioinformatics applications need a lot of memory and scratch storage space, simulation programs operate effectively in a massively parallel environment (with many cores). Diverse requirements must be recognized, and infrastructure must be managed appropriately. The Personal Data Protection Act (PDPA) presents new regulations for operating HPC. As part of the Genomics Thailand project, a set of rules and processes necessary to abide by the PDPA law and handle sensitive data have been set up.

The HPC community must create an effective mechanism to transfer huge amounts of data if data mobility is unavoidable. Data Transfer Nodes (DTNs), SDWAN, SDN [24], and overlay networks are a few examples of functions that could be incorporated as part of the HPC infrastructure basic services. The method of application deployment in digital business has altered as a result of virtualization and container technologies [25–27]. These technologies are embraced by HPC communities as well. This strategy can be strengthened to utilize HPC resources better.

5.3 HUMAN RESOURCE DEVELOPMENT

Maintaining high availability of HPC resources require expertise and experiences that lead to opportunities in the third challenge, HRD. There are quite a few universities that offer courses on HPC, such as courses in parallel and distributed computing and big data. Some of the HPC topics are covered in other courses, such as computer organization, operating systems, and cloud computing. According to our questionnaires, there are approximately 50 researchers and faculties, 150 students who study big data and data sciences each year, 450 students who study HPC each year, and 300 students who participate in workshops, tutorials, and competitions each year in Thailand HPC community.

Human resources for maintaining and administering base HPC resources, applications, and interfaces fall into three broad categories. First, a position with job descriptions involving hardware and software systems, high-performance libraries, parallel file systems, storage, data center facilities, service offerings, performance tuning, bottleneck identification, job submission, workload scheduling, response to changes in requirements, and activities related to cyber security guidelines, experts for base HPC resources must have solid backgrounds in computer systems. Although these responsibilities are commonly found in IT departments, there is HPC-specific expertise that requires specialized knowledge and experience.

The demand for HPC administrative personnel is relatively small compared to the total number of IT personnel. Career opportunities in this field are limited. Although automation solutions can help to streamline some aspects of maintaining HPC systems, it is vital to identify critical knowledge areas in administering HPC systems and to manage community capabilities properly. The HPC administrators responsible for resource maintenance have shared their knowledge and experiences with a loosely affiliated group of interests. Efforts should be made to cultivate new generations of HPC administrators. Presently, at least five system integrators in the private sector in Thailand demonstrate high proficiency in implementing HPC systems, indicating the country potential for future growth.

The second category of human resources emphasizes the development of new HPC applications. This group is crucial as they will develop new applications and tools for high-performance computing. The third group covers the development of web application interfaces, APIs, and visualization. The WWW applications and API services will improve the utilization of HPC resources by connecting users, data, and services.

We need to concentrate more on developing the latter two skill sets to increase the use of high-performance computing in Thailand. There should be a balance between HRD and HPC initiatives because a steep learning curve requires a certain amount of time to master all the difficulties involved.

6. CONCLUSION AND FUTURE WORK

A high-performance computing landscape in Thailand aids in understanding the current status of scientific and technological development. In this paper, we present the developments in HPC in the past 25 years and provide information on the existing HPC resources in Thailand. We list the organizations that process these resources and tabulate a set of system specifications and applications.

We report approximately 54,838 processing cores and 21 PB of storage available for HPC applications in Thailand in 2022. We discuss the network for transferring big data and describe issues, directions, and trends for developing HPC infrastructure in the future. Researchers and policymakers can benefit from the presented survey results for planning research projects and national development plans.

The computation sciences and engineering fields in Thailand have grown over the past decades and expanded to several disciplines. In the future, social data will become an important input data source in new HPC applications, for example, health sciences. Therefore, the HPC community has to move forward and roll out procedures and standards to work with sensitive data.

ACKNOWLEDGMENT

We are appreciative of the HPC community in Thailand who gave us input and permitted us to use information about their systems. Information for the LANTA system is provided by the NSTDA Supercomputer Center (ThaiSC). We thank the the National Biobank of Thailand (NBT) and the National Science and Technology Development Agency (NSTDA) for sharing the data of their HPC utilization. The Genomics Thailand project, sponsored by the Health Systems Research Institute (HSRI), Ministry of Public Health.

Finally, we would like to acknowledge the outstanding contributions made to the advancement of computer systems in Thailand by the late Assistant Professor Dr. Putchong Uthayopas.

References

[1] J. Gustafson, "A paradigm for grand challenge performance evaluation," in *Toward Teraflop Computing and New Grand Challenge Applications*, Mardi Gras'94 Conference, pp. 279–28, 1994.

[2] D. Reed, D. Gannon, and J. Dongarra, "Reinventing high-performance computing: Challenges and opportunities," 2022. Available from `https://arxiv.org/abs/2203.02544. [AccessedDec.20,2022]`.

[3] E. Strohmaier, H. Meuer, J. Dongarra, and H. Simon, "The TOP500 list and progress in high-performance computing," *Computer*, vol. 48, no. 11, pp. 42–49, 2015.

[4] M. Sato et al., "Co-Design for A64FX Manycore Processor and "Fugaku"," *SC20: International Conference for High Performance Computing, Networking, Storage and Analysis*, Atlanta, GA, USA, pp. 1-15, 2020.

[5] J. Dongarra, P. Luszczek, and A. Petitet, "The LINPACK benchmark: Past, present, and future," *Concurrency and Computation: Practice and Experience*, vol. 15, no. 9, pp. 803–820, 2003. Available from `https://doi.org/10.1002/cpe.728`. [Accessed Dec. 15, 2022].

[6] D. Becker, T. Sterling, D. Savarese, J. Dorband, U. Ranawake, and C. Packer, "Beowulf: A Parallel Workstation for Scientific Computation," in *International Conference on Parallel Processing*, 1995.

[7] NSTDA Supercomputer Center (ThaiSC), `https://thaisc.io`, [Accessed April, 2023].

[8] Y. Poovarawan and P. Uthayopas, "High-performance computing: Needs and application to Thailand," in *20th Electrical Engineering Conference (EECON)*, 1997.

[9] P. Uthayopas, S. Katchamart, J. Fakcharoenpol, and P. Hemunnopjit, "SMILE: Toward the affordable high-performance computing platform using network-based multicomputer," in *1st Annual National Computational Science and Engineering Symposium (ANSCSE)*, 1997.

[10] P. Uthayopas, S. Sa-nguanpong, and Y. Poovarawan, "Building a large Beowulf cluster system: PIRUN experience," in *4th Annual National Symposium on Computational Science and Engineering (ANSCSE)*, 2000.

[11] P. Uthayopas, J. Maneesilp, and P. Ingongnamm, "SCMS: An integrated cluster management tool for Beowulf cluster system," in *International Conference on Parallel and Distributed Proceeding Techniques and Applications (PDPTA)*, 2000.

[12] C. Sriprayoonskul and P. Uthayopas, "SQMS: A batch scheduling system for large PC cluster," in *5th Annual National Symposium on Computational Science and Engineering (ANSCSE)*, 2001.

[13] I. Foster, C. Kesselman, and S. Tuecke, "The anatomy of the grid: Enabling scalable virtual organizations," in *International Journal of High-Performance Computing Applications*, vol. 15, no. 3, pp. 200–222, 2001.

[14] [V. Varavithya and P. Uthayopas, "ThaiGrid: Architecture and overview," *NECTEC Technical Journal*, vol. 2, no. 9, pp. 219–224, 2001.

[15] V. Varavithya and P. Uthayopas, "Design considerations in building terascale cluster computing systems," in *National Computer Science and Engineering Conference (NCSEC)*, 2005.

[16] P. Uthayopas and V. Varavithya, "Architecture and software designs for the Thai national grid infrastructure," in *National Computer Sciences and Engineering Conference (NCSEC)*, 2006.

[17] P. Uthayopas and V. Varavithya, "Building a national grid infrastructure: A ThaiGrid experience," *EIT Computer*, 2009.

[18] T. Achalakul, B. Sirinaovakul, and N. Nuttaworakul, "Virtual Laboratory: A distributed collaborative environment," *Computer Applications in Engineering Education*, vol. 12, no. 1, pp. 44–53, 2004.

[19] S. Phatanapherom and P. Uthayopas, "Building a scientific grid portal for the Thai national grid system," in *11th Annual National Symposium on Computational Science and Engineering (ANSCSE)*, 2007.

[20] S. Pumma and W. Feng, P. Phunchongharn, S. Chapeland, and T. Achalakul, "A runtime estimation framework for ALICE," *Future Generation Computer Systems*, vol. 72, pp. 65–77, 2017.

[21] K. Jutawongcharoen, V. Varavithya, K. Lekdee, A. Chaichit, and S. Sribuddee, "The implementation of the UniNet's research DMZ," in *20th International Computer Science and Engineering Conference (ICSEC)*, 2016.

[22] T. Szeto T and V. Varavithya, "Parallel file

transfers for HPC applications in the UniNet environment," in *16th Annual International Symposium on Computational Science and Engineering (ANSCSE)*, 2012.

[23] NSTDA Genomic Thailand Project, `https://genomicsthailand.com/`, [Accessed Jan, 2023]

[24] P. Morreale and J. Anderson, "Software-defined networking," *CRC Press*, 2014.

[25] T. Rosado T and J. Bernardino, "An overview of OpenStack architecture," in *18th International Database Engineering and Applications Symposium (IDEAS'14)*, Association for Computing Machinery, pp. 366–367, 2014.

[26] D. Merke, "Docker: Lightweight Linux containers for consistent development and deployment," *Linux Journal*, issue 239, no. 2, 2014.

[27] Production-grade container orchestration. `https://kubernetes.io/`. [Accessed Oct. 23, 2022].

[28] G. Guidi, M. Ellis, A. Buluc, K. Yelick, and D. Culler, "10 Years Later: Cloud computing is closing the performance gap," in *International Conference on Performance Engineering (ICPE'21)*, pp. 41–48, 2021.

Vara Varavithya is an associate professor at Chitralada Technology Institute, where he has served as dean in the Faculty of Digital Technology since 2010. From 1991 to 2019 he was a faculty member at the Faculty of Engineering, King Mongkut's University of Technology North Bangkok where he served as an Assistant Dean in Information Technology, an Associate Dean in Research and Information Technology, and the Head of Electrical Engineering Department in 1998-2002, 2002-2006, and 2006-2010, respectively. He received a B.Eng. in electrical engineering from King Mongkut's University of Technology North Bangkok. He received M.S. and Ph.D. in Computer Engineering form the Iowa State University in 1989. He has been involved in numerous stages of Thailand's research and educational networks' development. His research interests span in digital infrastructure, computer networking, performance evaluation, and computer systems.

Supakit Prueksaarooon is a lecturer in department of electrical and computer engineering at Thammasat university. He received his Ph.D. and master in computer engineering from King Mongkut's University of Technology North Bangkok. His expertise is in parallel file system, cluster computer, Cloud computing, vmware, hadoop and Cloudera. More specifically, his special expertise is aimed to providing large-scale level of data storage and computing infrastructure.