National Supercomputing Centre Singapore: The HPC Leader in the Southeast Asia Region

● Shun Utsui ● Motohiro Yamada

The National Supercomputing Centre Singapore (NSCC) is Singapore's national supercomputer center, and it has the objectives of supporting national R&D initiatives, attracting industrial research collaboration, and enhancing Singapore's research capabilities. In assisting this national-level research initiative for high performance computing (HPC) in Singapore, Fujitsu provided the system platform and is in charge of system operation. As an optimal system platform for NSCC, we selected the PRIMERGY Series for the computing nodes, and integrated storage, and a high-speed network from third-party vendors. For the system operation, Fujitsu developed an optimized support system that combines Fujitsu's proprietary know-how in HPC and the existing Managed Services of Fujitsu Singapore. This paper describes the challenges that were encountered and the solutions that were developed during system introduction and operation.

1. Introduction

On March 28, 2016, the system of the National Supercomputing Centre Singapore (NSCC) went into operation. Fujitsu was selected to supply the system platform and operate the system.[1)]

The NSCC system was established with the cooperation of the National University of Singapore (NUS), Nanyang Technological University (NTU), and the Singapore University of Technology and Design (SUTD), under the leadership of Singapore's Agency for Science, Technology and Research (A*STAR). The system boasts computing power of 1 PFLOPS (peta floating-point operations per second). In other words, it has the capacity to perform 1 quadrillion floating-point operations per second, which is approximately one tenth of the K computer[note 1)] processing power of Japan's second most powerful system. The K computer is currently ranked 7th in the world (according to the TOP500 list of November 2016). NSCC achieves this level of performance by linking approximately 1,300 servers in a high-speed network. This makes the NSCC system the 115th most powerful high-performance computing (HPC) system in the TOP500 list released in November 2016, and the highest ranked system in Southeast Asia. The HPC market has been led mainly by Japan, the U.S., and Europe, but Southeast Asian countries are developing rapidly. With the inauguration of NSCC, it is expected that the growth of the HPC community in this region will further accelerate. In this regard, one of the missions of NSCC is to promote the use of HPC in Singapore. Sharing Fujitsu's expertise in operating numerous large-scale HPC systems around the world with NSCC is also an important role of Fujitsu in this project.

This paper discusses the challenges faced in the NSCC project and the role that Fujitsu has played in meeting them.

2. Challenges

As mentioned at the beginning, Fujitsu's roles in this project are to provide the system platform and to operate the system. The challenges faced in fulfilling those missions were as follows.

First, there were three main technical challenges during the system implementation. The first challenge was the design of the data center. The NSCC data center is situated on the 17th floor of a mixed-use office building called Connexis North Tower, a location that presented a number of challenges. High power

note 1) A supercomputer jointly developed by RIKEN and Fujitsu.

efficiency was also required owing to Singapore's climate, which is characterized by warm temperatures year round.

The second challenge was the development of a high-speed network connecting the location of each stakeholder and NSCC. NSCC's main system is located in the aforementioned mixed-use office building, but some of its equipment is in different locations at NUS and NTU. The goal was to link all these locations with a high-speed network to make it easy for users to access the NSCC system and transfer data to it.

The third challenge was to meet hierarchical storage system and fast storage performance requirements, since storage performance is as important as computing capacity for the practical implementation of HPC. Storage performance may become a bottleneck for applications that generate a lot of I/O (or in layman terms, "do a lot of writing to and reading from files"). Taking this point into account for the NSCC system, the highest layer of the hierarchical storage was required to provide a high 500 GB/s performance. In addition, a hierarchical but transparent system to users was required to ensure ease of use.

Challenges were also faced in system operation. In general, it is harder in Singapore to find human resources experienced in HPC that are capable of operating a system of this scale compared with regions such as Japan, the U.S., and Europe, where HPC communities are very sophisticated. Notwithstanding this, the dispatch of a team capable of fully handling HPC operations to the customer's site was required.

Solutions provided by Fujitsu to meet the above challenges are described in the following sections.

3. System implementation

This section begins with an overview of the NSCC system. The system is roughly divided into three components: compute nodes, storage, and high-speed network (**Figure 1**).

The compute nodes consist of Fujitsu PRIMERGY Series servers. Out of the total of 1,298 nodes, 1,160 are FUJITSU Server PRIMERGY CX2550 M1 (standard nodes), 128 are FUJITSU Server PRIMERGY CX2570 M1 (graphics processing unit [GPU] nodes) with one GPU card, and 10 are FUJITSU Server RX4770 M2 (large memory nodes) with larger memory capacity than the other nodes. The CPUs used are Intel's Xeon E5-2690 v3 and Xeon E7-4850 v3 processors. This configuration achieves performance of 1.01 PFLOPS as measured by LINPACK of HPC Challenge, which is the standard benchmark test for the TOP500 list[note 2)].

The solutions provided by Fujitsu for each challenge are introduced below.

3.1 Challenge 1: Design of data center

The location of the data center in this project presented a number of challenges. As mentioned above, NSCC's data center was set up on the 17th floor of a mixed-use office building. Owing to rack height restrictions due to the dimensions of the elevators used to bring the equipment up to the 17th floor, as well as weight restrictions, the supplied hardware had to be designed so as not to conflict with those restrictions. Moreover, while average PUE (power usage effectiveness, an index value indicating the efficiency of a data center, the lower the value, the better the efficiency) in Singapore is 2.5 or higher (per NSCC survey), a more ambitious target PUE value of 1.4 was set for the project. Fujitsu's solution was the combination of three cooling methods, namely water cooling, air cooling, and water-cooled rear door heat exchanger (RDHx).

For the CX2550 M1 and the CX2570 M1 nodes, Fujitsu delivered the first water-cooled models of the PRIMERGY Series. The proposal of these water-cooled machines was essential for minimizing power consumption. Using warm water of approximately 40°C for direct water cooling of the CPU, memory, and GPU, a reduction in power consumption of approximately 140 W to 160 W can be achieved per compute node chassis (housing either four CX2550 M1 units or two GPU-equipped CX2570 M1 units). Applied to the entire system, this cooling method by itself yields an economy of 50 kW or more. RDHx were mounted on all units except the CX2550 M1 and CX 2570 M1 node racks. The remaining parts (such as network equipment equipped with compute node racks) are cooled by air cooling methods.

3.2 Challenge 2: Remote site linking

Understanding a high-speed network including

note 2) This test uses only the CPUs of the CX2550 M1 and CX2570 M1 nodes. It does not use the GPU of the CX2570 M1 nodes and the CPU of the RX 4770 M2 nodes.

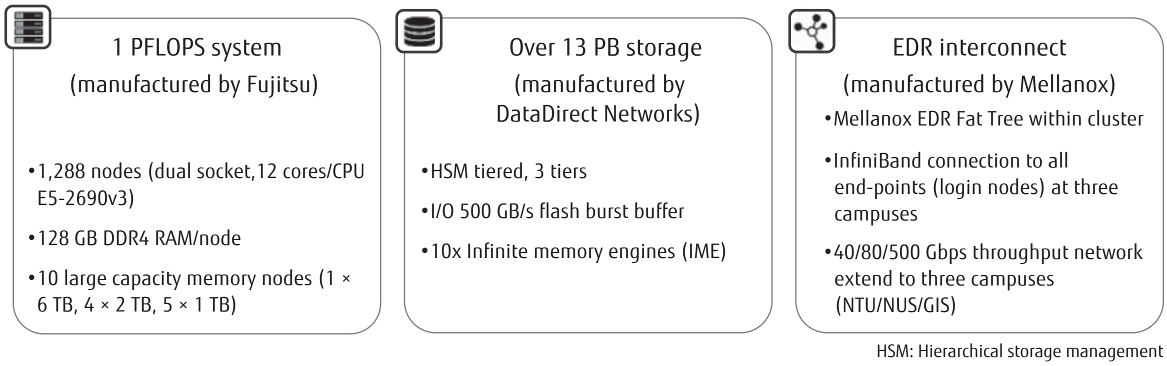

Figure 1
Three components of NSCC system.

its various locations requires an overview of the overall network design. The NSCC system comprises three networks: a user network, a management network, and a high-speed network. The user network and management network are Ethernet networks that communicate using the TCP/IP protocol. The high-speed network is used for communication between compute nodes and data access to the storage system. The InfiniBand technology of Mellanox Technologies was adopted for the high-speed network. InfiniBand EDR, which provides bandwidth of up to 100 Gbps per link and is the latest technology currently available, was used for the NSCC system. It is configured in Quasi Fat Tree non-blocking topology, meaning that whichever switch the traffic goes through, the per-link native performance of the switch is unaffected.

The challenge in this project was remote linking of NTU and NUS, as well as the Genome Institute of Singapore (GIS), an organization within A*STAR. Login nodes, which are gateways for users to access the system, were physically installed at multiple locations. At present, 4 units have been installed on the NUS campus and 4 units on the NTU campus, in addition to the 4 units at NSCC. One large-capacity memory node was also installed at GIS (**Figure 2**).

As a solution, it was decided to connect these nodes directly to NSCC's InfiniBand network with dark fiber cables via Mellanox Technologies' MetroX switches. The physical wiring distances from NSCC to GIS, NTU, and NUS are 2 km, 22.4 km, and 42 km, respectively. These connections ensure links with minimum capacity of 40 Gbps for system access from each remote site to NSCC. Instead of using TCP/IP, the remote direct memory access (RDMA) protocol is used for accessing data over this network. This design allows users to transfer data more comfortably than via the Internet, even from within their respective university campuses.

3.3 Challenge 3: Storage system

Another characteristic of the NSCC system is its requirements in terms of hierarchical storage system and high-speed storage performance. Fujitsu worked with DataDirect Networks (DDN) to fulfill those storage requirements. The installed storage system consists of four layers. The top layer consists of a burst buffer cache, followed by a Lustre file system storage, an IBM Spectrum Scale (previously known as General Parallel File System) storage, and lastly an archive area. The top layer performance requirement was especially high, reaching 500 GB/s. In response, Fujitsu proposed DDN's Infinite Memory Engine (IME). IME is a shared pool that utilizes non-volatile memory express (NVMe) solid-state drives (SSDs), as opposed to conventional hard disks, and acts like a cache,[2] which is beneficial to users running applications that require frequent I/O access on HPC systems. The second layer consists of a 4 PB Lustre file system (scratch area), the third layer consists of a 4 PB GPFS file system (home area), and the fourth layer is an archive file system that clears older files from the home area. The GPFS (home area) is where users mainly upload their files and binary data. The Lustre file system is an area where users mainly execute their jobs into the computing resources. Both the parallel file systems Lustre and GPFS are tuned for their respective intended use.

4. System operation

Apart from the system implementation, there

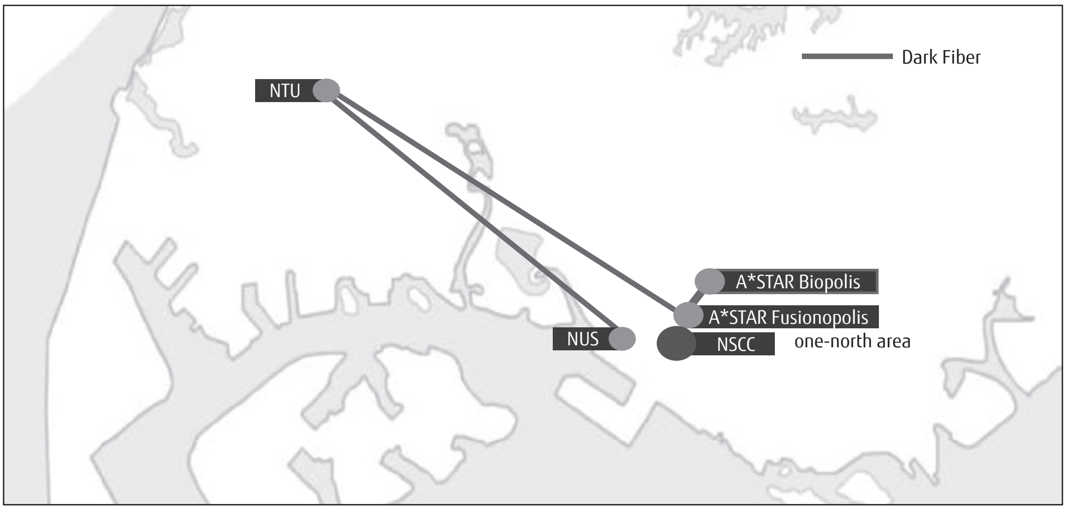

Figure 2
Remote linking of NSCC.

were also challenges in providing services for NSCC's daily system operation. For NSCC, Fujitsu developed a unique and optimal operational structure by introducing the large amount of HPC experience in Japan and other countries, and combining it with the existing system operation service (Managed Services) at Fujitsu Singapore (**Figure 3**). Fujitsu deployed an HPC system operation team to NSCC. The team is permanently seated in the customer's office to support their HPC system operation. At present, this team consists of six members, including two engineers from partner companies. These include HPC system engineers, a DDN storage engineer, an HPC code optimizer, a mechanical and engineering (M&E) engineer, and a service delivery manager. As such, this is a team made up of professionals in each required field of expertise. Moreover, it is a highly international team with Singaporean, Indian, British, and Japanese (the author) members. The team counts only one Singaporean member, a fact that may be less surprising considering the fact that there are only about 3.9 million Singaporeans and permanent residents in Singapore out of a population of about 5.54 million.[3] Finding the right people with HPC expertise in Singapore is much harder than in Japan, considering Singapore's small population. Therefore, Fujitsu gathered those professionals from all over the world.

Additionally, a Service Desk (SD) team that serves as the first point of contact for inquiries and requests from users, and a Remote Monitoring Center (RMC) team for monitoring system alerts round out the Managed Services team. The SD and RMC provide similar services to other customers in Singapore. They are located at Fujitsu's own operations center rather than at customer sites.

Here is an example of how Fujitsu provides services to NSCC users. First, let us assume that a user notices some system-related problems during job execution in the NSCC system. The first point of contact to report his/her issue is the SD team. After confirming the problem and its scope, SD escalates the problem to the HPC system operation team of NSCC if necessary. From this point on, on-site HPC system engineers aim to solve the problem while sharing information on an on-going basis with SD and the user. Once the problem is resolved, SD will follow up with the user to seek for closure on the problem. Once the user confirms that the issue has been solved satisfactorily, the case is closed.

Furthermore, for more complicated problems, a support team of third-party vendors and the HPC team of Fujitsu in Japan provide back-end support to the Managed Service team of Fujitsu Singapore.

At many HPC sites other than NSCC, there is an IT team on the customer side that will handle the first level support. It is not common for Fujitsu to handle the entire cycle from first response to closing. By contrast, at NSCC, Fujitsu is responsible for the entire support

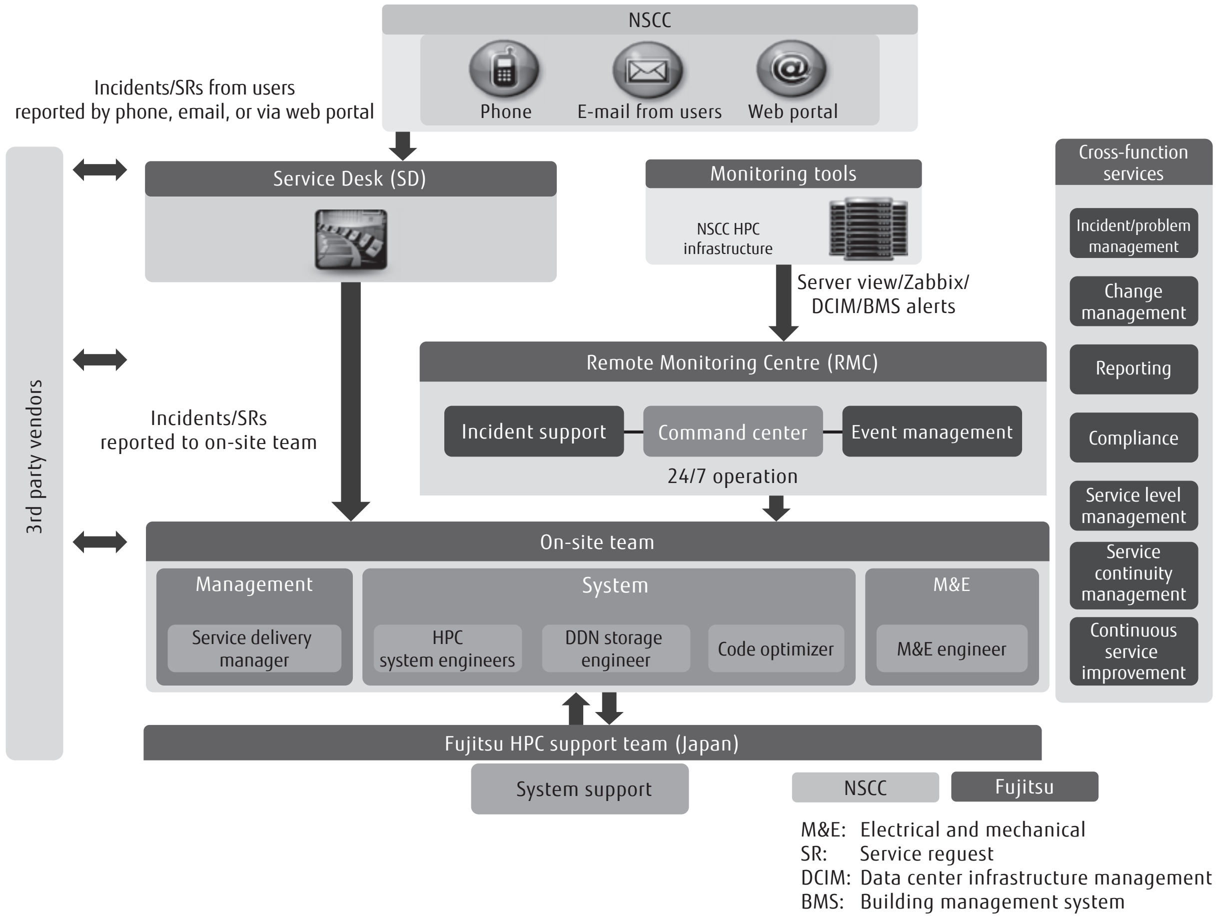

Figure 3
NSCC operation system.

cycle. What made this possible is the strong and structured Managed Services team of Fujitsu Singapore. The inauguration of the NSCC project brought two things together that might not have been combined otherwise: HPC operation know-how and a structured Managed Service. The result was the creation of an optimal HPC operation system exclusively for NSCC.

5. Future HPC trend as seen from NSCC

Operating the HPC system at NSCC, something interesting can be observed: whereas traditional HPC simulations such as fluid dynamics and molecular modelling still account for the majority of the jobs taking up resources, usage of Deep learning applications such as Caffe and Theano are also increasing. This is an obvious change to what we saw in many HPC centers a few years ago. A driving force behind this phenomenon might well be the fact that NSCC has built an open environment that is easily accessible even for HPC beginners.

Generally, in the world of Deep learning, the greater the amount of accumulated data, the more accurately a computer can make predictions. A factor that is involved here is a project called InfiniCortex that connects Singapore with countries all over the world with a high-speed network. Just as NUS and NTU are connected to NSCC via dark fiber, Singapore has laid undersea optical fiber cables that link it to various countries. For example, an optical fiber line directly connects Singapore to the Tokyo Institute of Technology in Japan. Singapore has many other such connections to the United States and Europe. According to a paper published by A*STAR, Computational Resource Centre (A*CRC), the leader of the InfiniCortex project, these high-speed links are "the world's first globally distributed and concurrent galaxy of supercomputers."[4)] Besides connecting the various HPC systems, these high-speed links are of critical importance in terms of facilitating data transfers. Singapore, being a small country with few resources, has no choice but to rely on foreign resources.[5] The InfiniCortex project is setting up an ecosystem for Singapore to easily import data as a precious resource.

6. Conclusion

A year has passed since NSCC started operating. Fujitsu's role is to continue supporting NSCC and its users utilizing its huge HPC knowledge base and expertise, and to be flexible in finding the right solution to any issue found. As mentioned above, the future HPC market is evolving into a new stage where the purpose of HPC usage will diversify. It will also be important to identify the potential needs of new HPC users, and promote the use of HPC in collaboration with NSCC.

Furthermore, starting with Singapore, it will also be important for Fujitsu to promote the use of HPC to other Southeast Asian countries. Solving problems through technology is what Fujitsu has been doing since its foundation. Just as Fujitsu was able to combine its HPC expertise with the existing Managed Services to form an optimum operational structure, our mission today is to work flexibly on issues both technical and non-technical to offer innovative solutions to each challenge.

References

1) NIKKEI ASIAN REVIEW: Singapore taps Fujitsu for region's most powerful machine.
http://asia.nikkei.com/Tech-Science/Tech/Singapore-taps-Fujitsu-for-region-s-most-powerful-machine
2) DDN storage website.
http://www.ddn.com/
3) Ministry of Foreign Affairs of Japan website, "Republic of Singapore" (in Japanese).
http://www.mofa.go.jp/mofaj/area/singapore/data.html
4) M. T. Michalewicz et al.: InfiniCortex: Present and Future. Proceedings of the ACM International Conference on Computing Frontiers, pp. 267-273, 2016.
http://dl.acm.org/ft_gateway.cfm?id=2912887&ftid=1734211&dwn=1&CFID=922850348&CFTOKEN=95619110
5) I. Iwasaki: The history of Singapore. Chuko Shinsho, Chuokoron-Shinsha, Inc., 2013 (in Japanese).

Shun Utsui
Fujitsu Asia Pte. Ltd.
Mr. Utsui is currently engaged in the HPC business in Singapore (NSCC project) and Southeast Asia.

Motohiro Yamada
Fujitsu Ltd.
Mr. Yamada is currently engaged in the global HPC business planning and promotion.