System Operations and Development Team

1. Team members

Fumiyoshi Shoji (Team Head)
Atsuya Uno (Research & Development Scientist)
Hitoshi Murai (Research & Development Scientist)
Motoyoshi Kurokawa (Research & Development Scientist)
Keiji Yamamoto (Postdoctoral Researcher)
Toshiyuki Tsukamoto (Research & Development Scientist)
Fumio Inoue (Research & Development Scientist)
Mitsuo Iwamoto (Technical Staff)
Katsufumi Sugeta (Technical Staff)

2. Research Activities

The K computer is a distributed-memory parallel computer system consisting of 82,944 compute nodes and has played a central role of the High Performance Computing Infrastructure (HPCI) initiative granted by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). HPCI has achieved an integrated operation of the K computer and other supercomputer centers in Japan and has enabled seamless accesses of a cluster of supercomputers including the K computer from users' machines. HPCI has also provided large-scale storage systems which can be accessed from all over Japan.

System Operations and Development Team (SODT) has conducted the research and development on advanced management and operations of the K computer. While analyzing the operational statistics collected during the shared use, SODT has improved the system configuration, such as the job scheduling, the file system, and users' environments. For example, it is very difficult to achieve higher system utilization because the K computer has to process various sizes and types of jobs simultaneously. SODT has responded flexibly to the user's requests and made analysis of the operational status, and then has realized high level utilization approximately 76% in FY2013. Furthermore, SODT has developed tools that support the use of the K computer.

SODT has conducted research in cooperation with the research team of RIKEN Spring-8 Center and RIKEN AICS for big data processing on the K computer. The K computer is used to analyze the huge data transmitted from RIKEN Harima where SACLA XFEL facility is located.

SODT also helps users handle the K computer and utilize the K computer resources effectively by improving the system software. This support has been conducted together with the Software Development Team.

3. Research Results and Achievements

3.1 Improvements of system software of the K computer

We have fixed and improved many points of the system software through the shared use. Here, we describe the main activities in FY2013.

- **Analyzing Operation Statistics and Job Scheduling**

  We have analyzed logs of jobs executed on the K computer and have performed job scheduling simulation using the simulator which can simulate the job scheduling same as real job scheduler on the K computer. Referring these results, we have tuned some parameters of the scheduler (such as scheduling map, file staging timing and so on).

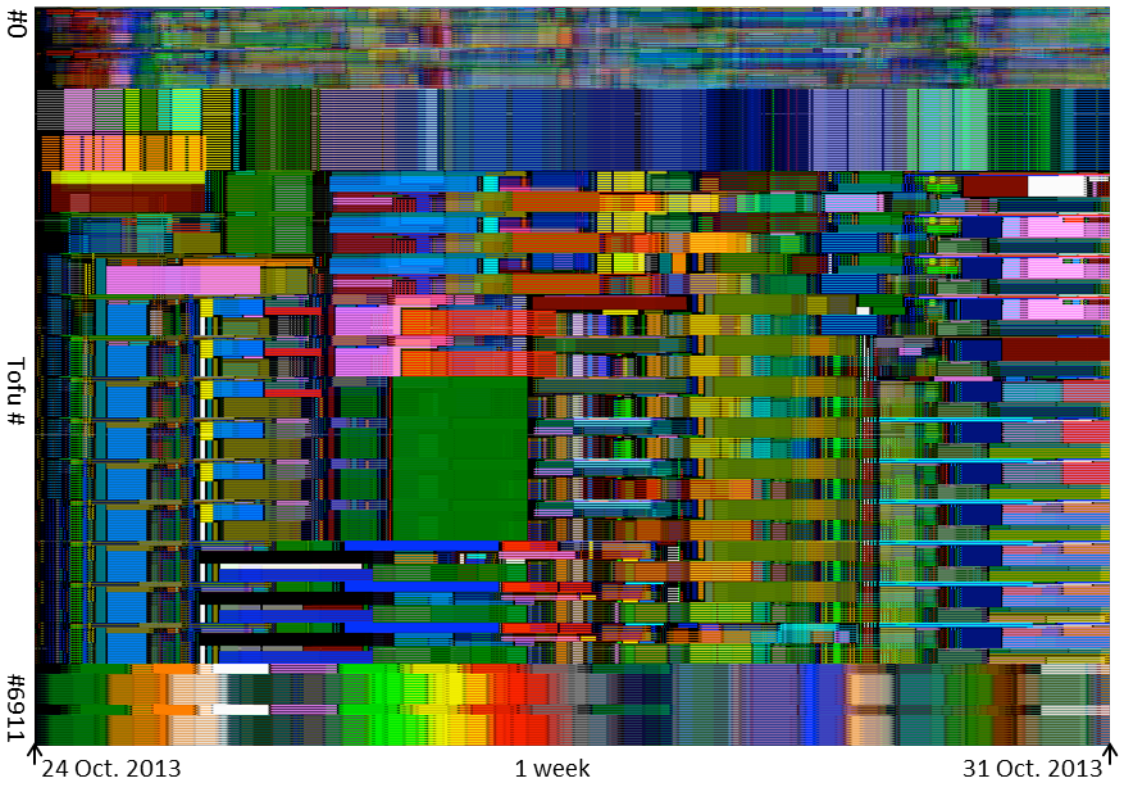

Figure 1. Job execution situation on the K computer.

In order to evaluate the results of the job scheduling, we have visualized the scheduling map. Figure 1 shows the number and size of jobs executed on the K computer during one week from October 24 to October 31, 2013. Since the scheduler allocates a job to nodes shaped a rectangle

in logical 3dimensional coordinate of the K computer, the nodes allocated to one job could be scattered on this figure. The boxes with the same color on the same vertical axis correspond to the same job. The width of a rectangle denotes the elapsed time from the start time to the termination time. The black area shows the nodes either waiting for job execution or for maintenance. The one-ninth of the K computer that corresponds to upper area of this figure is assigned for a resource group "**Small**" which is a limited resource only for small-sized job. On the other side, eight-ninth one is assigned for a resource group "**Large**." As we can see this figure, the K computer can process many jobs in various sizes simultaneously which contributes for high ratio of the node utilization. During this week, approximately 3,600 and 3,000 jobs were processed in **Small** and **Large** respectively, and the total node utilization was approximately 86%.

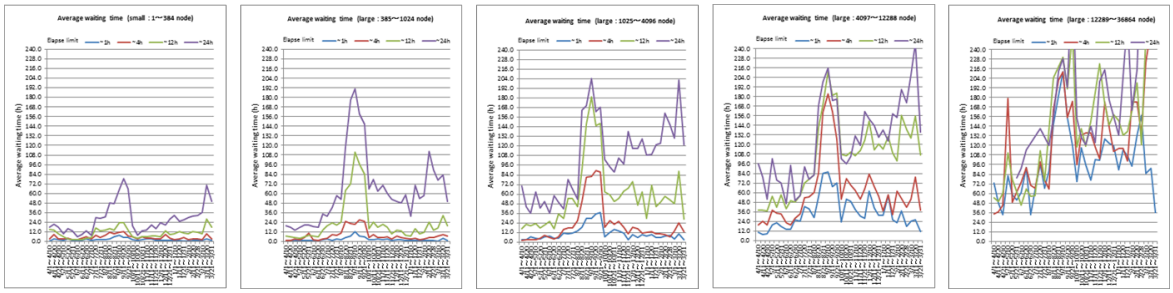

Figure 2. Average waiting time.

Figure 2 shows average waiting times with respect to both job sizes and elapsed times in FY 2013. The average waiting times are directly proportional to both elapsed time and number of nodes. It means that a fairness of opportunity for job execution has been kept.

The average waiting time in September 2013 was very longer than that in other months. Each group had appropriate compute resources for a year, and the resources were divided into two periods: from April to September and from October to March of the next year. In order to consume the remaining compute resources before the resources will be expired many jobs are submitted at the end of the periods. At the job congestion in September 2013 we adjusted system parameters, such as the maximum number of jobs executed simultaneously for each user group and the waiting time has been reduced. But job congestion has occurred for large scale jobs in March 2014. We have to analyze the operational statistics more and need to tune the parameters.

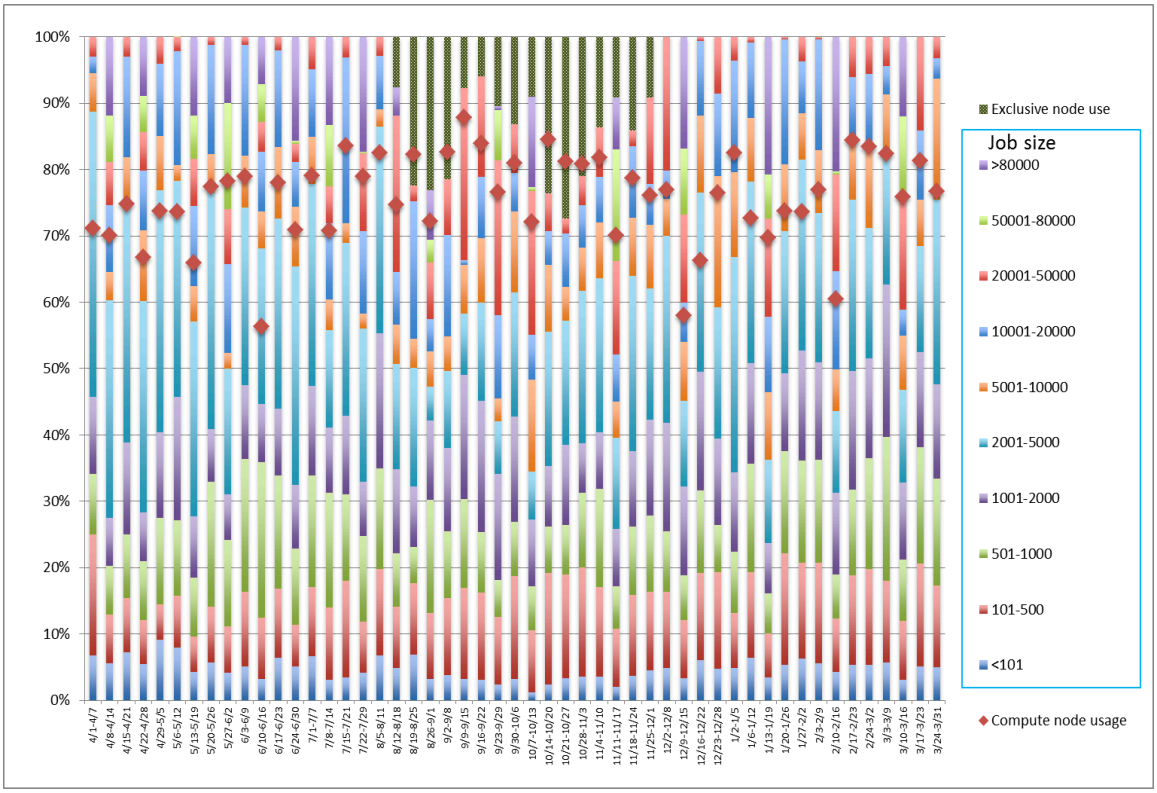

Figure 3. Details of resource usage in FY2013.

Figure 3 shows the compute node usage in FY 2013. We have achieved approximately 76% node usage in FY2013. This is more than 15% higher than last year's.

- **MPI libraries and System tools**

MPI libraries for the K computer are optimized for Tofu interconnect. But most of them are the most effective when 3D torus is specified. Since many jobs on the K computer are executed with 1D or 2D torus we extended some MPI libraries to be able to perform effectively on 1D or 2D torus.

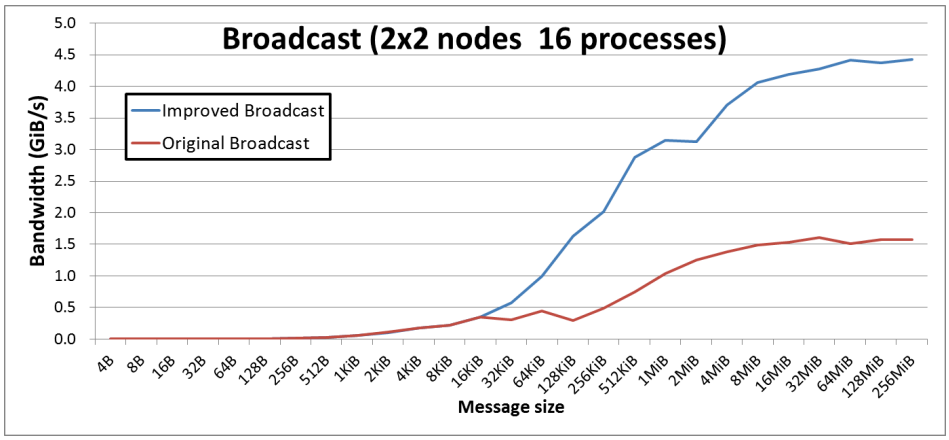

Figure 4. Performance of Broadcast on 2D torus.

Figure 4 shows the performance of original Broadcast and improved Broadcast on 2D torus. This graph indicates that the improved Broadcast is about 5 times as fast as original one using 128KiB message size.

We have developed tools that support the use of the K computer. "Waiting for the K" is a tool that enables users to know the estimated waiting time for the job execution and the available node information of the job execution. We have also been developing another system tools that can enhance the job management. Using these tools, users will be able to submit many jobs at once.

3.2 Cooperative research

- **Big data processing on the K computer**

This research is conducted by collaboration between the RIKEN Spring-8 Center and the RIKEN AICS. The goal of this project is to establish the path to discover the 3D structure of a molecule from a number of XFEL snapshots. Each snapshot size is around 20 Mbytes, but may vary depending on the resolution of image sensor. However, the number of images required to develop a 3D structure of a molecule is millions, resulting 20 PBytes of data size in total. The K computer is used to analyze the huge data. SODT cooperates with the big data transfer from SACLA to the K computer. Further, each image is classified into thousands of images to have every possible snapshot orientations and to reduce the quantum noise.

3.3 User support

We have conducted the user support, such as the user management and the consulting services.

- **User management**

The K computer has 170 or more groups and 2,000 or more users at the end of March, 2014. The number of HPCI users and AICS researchers are approximately 1,750 and 250 respectively. The number of daily active users is approximately 120.

- **Consulting services**

We support users through "the K support desk" and provide users the technical information on the K computer including system environments, system tools, and software libraries. The consulting services have been conducted together with the Software Development Team. Figure 5 indicates the consultation number in FY2013 and the consultation number in FY2013 is approximately 230.

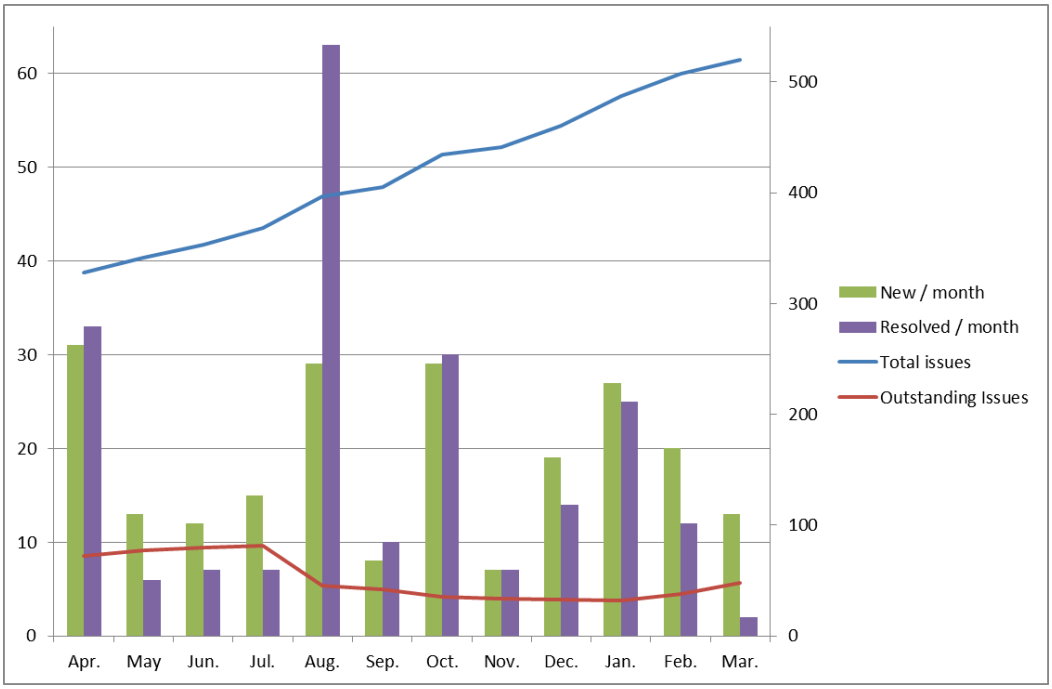

Figure 5. Consultation number in FY2013.

4. Schedule and Future Plan

We continue to improve the system software of the K computer and to provide the user support. The improved system software that supports the use of the K computer will be released in FY2014.

5. Publication, Presentation and Deliverables

(1) Journal Papers

[1] Tomoya Adachi, Naoyuki Shida, Kenichi Miura, Shinji Sumimoto, Atsuya Uno, Motoyoshi Kurokawa, Fumiyoshi Shoji, and Mitsuo Yokokawa," The Design of Ultra Scalable MPI Collective Communication on the K Computer", Comput. Sci., 28(2-3):147–155, May 2013.

[2] Fumiyoshi Shoji, "Trend of supercomputer", J.HTSJ, Vol. 52, No220, (2013) 15-20 (In Japanese).

[3] Masaaki Terai, Ken-ichi Ishikawa, Yoshinori Sugisaki, Kazuo Minami, Fumiyoshi Shoji, Yoshifumi Nakamura, Yoshinobu Kuramashi, Mitsuo Yokokawa,"Performance Tuning of a Lattice QCD Code on a Node of the K computer", Transactions of Information Processing Society of Japan, Vol.6, No.3, 43-57 (Sep. 2013) (In Japanese).

[4] Atsushi Tokuhisa, Junya Arai, Yasumasa Joti, Yoshiyuki Ohno, Toyohisa Kameyama, Keiji Yamamoto, Masayuki Hatanaka, Balazs Gerofi, Akio Shimada, Motoyoshi Kurokawa, Fumiyoshi Shoji, Kensuke Okada, Takashi Sugimoto, Mitsuhiro Yamaga, Ryotaro Tanaka, Mitsuo Yokokawa, Atsushi Hori, Yutaka Ishikawa, Takaki Hatsui, Nobuhiro Go, "High-speed

classification of coherent X-ray diffraction patterns on the K computer for high-resolution single biomolecule imaging", In Journal of Synchrotron Radiation, volume 20, 2013.

[5] SHIMIZU Toshiyuki, AJIMA Yuichiro, YOSHIDA Toshio, ASATO Akira, SHIDA Naoyuki, MIURA Kenichi, SUMIMOTO Shinji, NAGAYA Tadao, MIYOSHI Ikuo, AOKI Masaki, HARAGUCHI Masatoshi, YAMANAKA Eiji, MIYAZAKI Hiroyuki, KUSANO Yoshihiro, SHINJO Naoki, OINAGA Yuji, UNO Atsuya, KUROKAWA Motoyoshi, TSUKAMOTO Toshiyuki, MURAI Hitoshi, SHOJI Fumiyoshi, INOUE Shunsuke, KURODA Akiyoshi, TERAI Masaaki, HASEGAWA Yukihiro, MINAMI Kazuo, YOKOKAWA Mitsuo, "Design and Evaluation of K Computer", IEICE TRANSACTIONS on Information and Systems:D, Vol.J96-D No.10 pp.2118-2129,Oct. 2013 (In Japanese)

(2) Conference Papers

[6] Keiji Yamamoto, Atsuya Uno, Hitoshi Murai, Toshiyuki Tsukamoto, Fumiyoshi Shoji, Shuji, Matsui, Ryuichi Sekizawa, Fumichika Sueyasu, Hiroshi Uchiyama, Mitsuo Okamoto, Nobuo Ohgushi, Katsutoshi Takashina, Daisuke Wakabayashi, Yuki Taguchi, Mitsuo Yokokawa, "The *K computer Operations: Experiences and Statistics", International Conference on* Computational Science (ICCS), 2014, Australia. (to appear)

(3) Invited Talks

[7] Fumiyoshi.Shoji, “Introduction to the K computer”, The K day in MQM 2013.