20. System Operations and Development Team

20.1. 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)
Mitsuo Iwamoto (Technical Staff)
Katsufumi Sugeta (Technical Staff)

20.2. 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 supercomptuers including the K computer from users'machines. HPCI has also provided large-scale storage systems which can be accessed from all over Japan.

AICS provided computing resources to some advanced users as an early access to a part of the K computer from the end of March 2011 to September, 2012. During this early access, the advanced users in five research fields of the Strategic Programs for Innovative Research (SPIRE) promoted by the MEXT implemented their application software. AICS could improve the system through their feedback. The K computer was completed in June 2012 and has been available for shared use since Septempber 28, 2012.

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 duiring the shared use, SODT has improved the system configuration, such as the job scheduling, the configuration of 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 around 80%.

SODT also helps users handle the K computer and utilize the K computer resources effectively by improving the compilers, MPI libraries and other system tools. This support has been conducted together with the Software Development Team.

20.3. Research Results and Achievements

20.3.1. Improvements of system software of the K computer

We have fixed and improved many points of the system software since the beginning of the shared use. Here, we describe the main improving points.

- Job Scheduling

The K computer has to manage various types of jobs simultaneously, so it is very difficult to achieve a high level efficiency of compute node usage. We have analyzed the operational statistics collected during the shared use. Investigating the analysis result, we have improved the system software (such as the job scheduler, the file staging system, and so on) and decided to change the job scheduling policy as follows:

1. To assign the designated compute nodes to a small resource group for small size jobs using less than 384 compute nodes for preventing the small size jobs from disturbing the large scale job execution.
2. To make an exclusive period for the large scale job execution using more than 36,864 compute nodes to execute the large scale jobs smoothly.

As a result, we could improve the system usage and reduce the waiting time for the job execution (Figure 1).

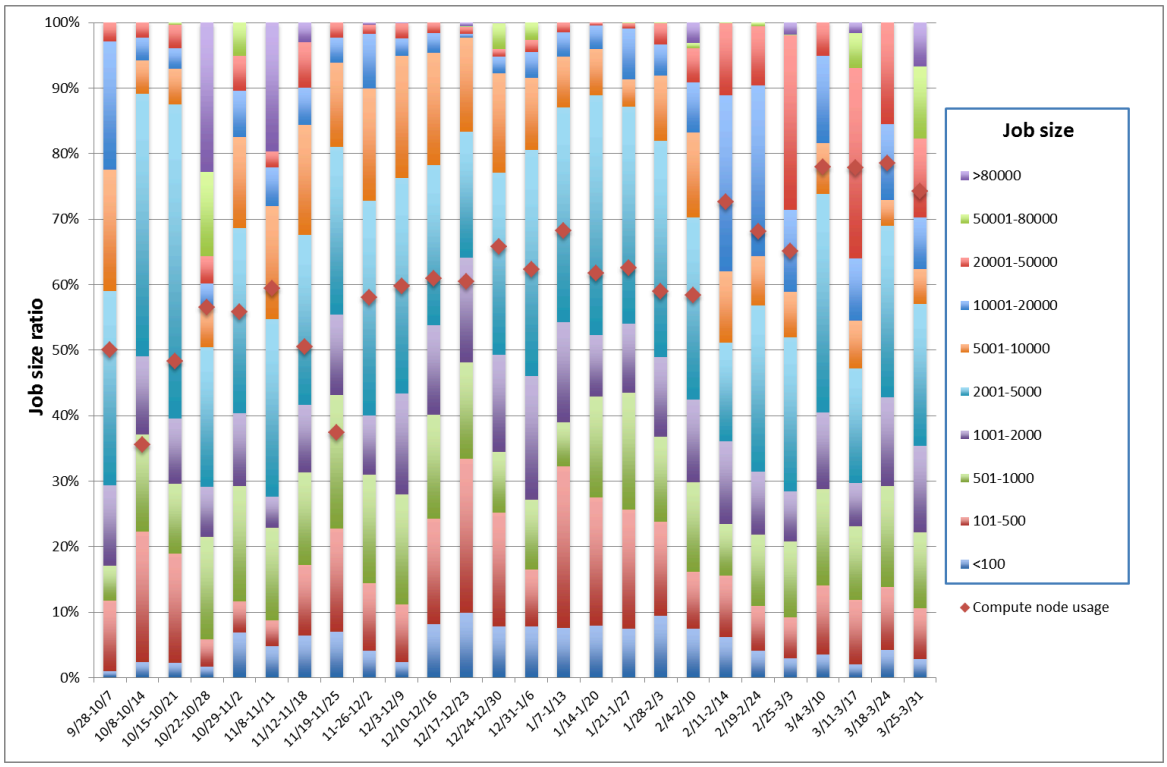

Figure 1 Details of resource usage during the shared use

Figure.1 indicates that the compute node usage has been improved and the number of executed large scale jobs has increased after changing the job scheduling policy on February, 2013.

- File system

The file system of the K computer is Fujitsu Exabyte File System (FEFS) based on Lustre file system and enhanced and tuned for the K computer. At the beginning of the shared use, we made 4 large volumes for user's home and data area. In this case, we required one month or more to check and repair the file system when an obstacle had occurred. This means that it is impossible to recover from the obstacles. So, we analyzed the users' usage and the setting of FEFS, and we decided to divide the file system to more smaller volumes and optimized the setting of FEFS. These changes enable to check and repair the files system within 48 hours without restrictions to users.

- Compilers, MPI libraries and System tools

Compilers and MPI libraries are customized and tuned for the K computer. But, through the shared use, we found the many improving points for them and improved them. These improvements have enabled the user to improve the performance of the user's application without any code changes. In figure 2, the performance of original Alltoallv and improved Alltoallv on the K computer are shown. This graph indicates that the improved Alltoallv is about 4 times as fast as original one using 10KB message size.

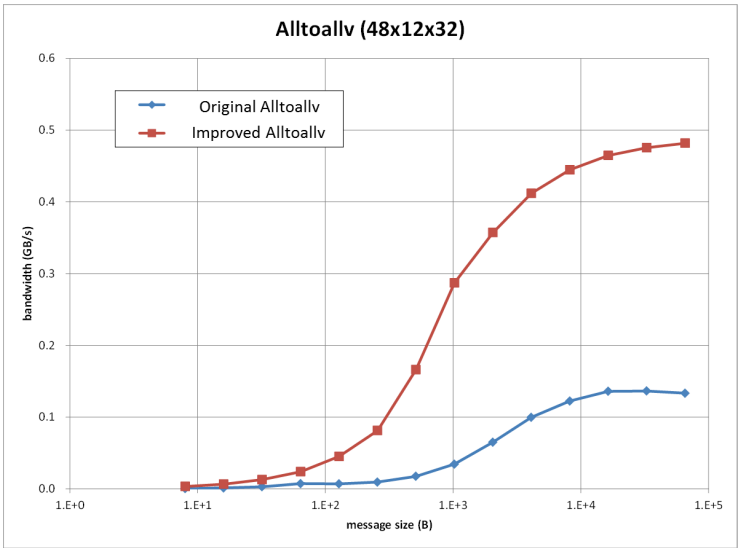

Figure 2 Performance of Alltoallv on the K computer

Many users request for tools that support the use of the K computer, so we are developing the tools as follows:

1. To provide the estimated waiting time for the job execution

2. To provide the available node information of the job execution
3. To provide the confusion information of the compute nodes

We consider these tools will be useful for the users.

20.3.2. User support

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

- User management

The K computer has 120 or more groups and 1,400 or more users at the end of March, 2013. The most of users are HPCI users and the members of AICS research divisions also study using the K computer. HPCI system performs the user management of HPCI users aside from us, so we have to adjust the difference between HPCI system and the K computer. We have developed the user managing system for this, and it enables to perform the user management of HPCI users and AICS users in the same way.

- 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. The consultation number in the period from the beginning of the shared use to the end of March, 2013 is approximately 1,400.

20.4. Schedule and Future Plan

We continue to improve the system software of the K computer and to provide the user support, and we will release the system tools that support the use of the K computer in FY2013.

20.5. Publication, Presentation and Deliverables

(1) Journal Papers

- None

(2) Conference Papers

1. Daisuke Takahashi, Atsuya Uno and Mitsuo Yokokawa: An Implementation of Parallel 1-D FFT on the K computer, HPCC2012, 25-27 Jun. 2012, Liverpool, UK.
2. T. Boku, K.-I. Ishikawa, Y. Kuramashi, K. Minami, Y. Nakamura, F. Shoji, D. Takahashi, M. Terai, A. Ukawa, T. Yoshie: Multi-block/multi-core SSOR preconditioner for the QCD quark solver for K computer, The 30th International Symposium on Lattice Field Theory, June 24-29, 2012, Cairns, Australia.

(3) Invited Talks

1. M.Kurokawa: The K computer and Expectations for Optical Devices, The Joint International Symposium on Optical Memory and Optical Data Storage 2012, 30 Sep.- 4 Oct. 2012.
2. M.Kurokawa: The K computer: 10 Peta-FLOPS supercomputer, 10th International Conference on Optical Internet (COIN), 29-31 May 2012.
3. F.Shoji: The K computer -System and Applications-, The 41st International Conference on Parallel Processing, Sep 10-13 2012, Pittsburgh, PA, USA.

(4) Posters and presentations

1. Atsuya Uno, Fumiyoshi Shoji and Mitsuo Yokokawa: The performance evaluation of the job scheduling with the file staging, IPSJ-SIGHPC 2012-HPC-136(22), 2012. (In Japanese)

(5) Patents and Deliverables

- None