Development of stable Grid service
at the next generation system of KEKCC

T. Nakamura, G. Iwai, H. Matsunaga, K. Murakami,
T. Sasaki, S. Suzuki, W. Takase

Computing Research Center,
KEK, High Energy Accelerator Research Organization,
1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan
E-mail: tomoaki.nakamura@kek.jp

Abstract.
A lot of experiments in the field of accelerator based science are actively running at High Energy Accelerator Research Organization (KEK) by using SuperKEKB and J-PARC accelerator in Japan. In these days at KEK, the computing demand from the various experiments for the data processing, analysis, and MC simulation is monotonically increasing. It is not only for the case with high-energy experiments, the computing requirement from the hadron and neutrino experiments and some projects of astro-particle physics is also rapidly increasing due to the very high precision measurement. Under this situation, several projects, Belle II, T2K, ILC and KAGRA experiments supported by KEK are going to utilize Grid computing infrastructure as the main computing resource. The Grid system and services in KEK, which is already in production, are upgraded for the further stable operation at the same time of whole scale hardware replacement of KEK Central Computer System (KEKCC). The next generation system of KEKCC starts the operation from the beginning of September 2016. The basic Grid services e.g. BDII, VOMS, LFC, CREAM computing element and StoRM storage element are made by the more robust hardware configuration. Since the raw data transfer is one of the most important tasks for the KEKCC, two redundant GridFTP servers are adapted to the StoRM service instances with 40 Gbps network bandwidth on the LHCONE routing. These are dedicated to the Belle II raw data transfer to the other sites apart from the servers for the data transfer usage of the other VOs. Additionally, we prepare the redundant configuration for the database oriented services like LFC and AMGA by using LifeKeeper. The LFC servers are made by two read/write servers and two read-only servers for the Belle II experiment, and all of them have an individual database for the purpose of load balancing. The FTS3 service is newly deployed as a service for the Belle II data distribution. The service of CVMFS stratum-0 is started for the Belle II software repository, and stratum-1 service is prepared for the other VOs. In this way, there are a lot of upgrade for the real production service of Grid infrastructure at KEK Computing Research Center. In this paper, we would like to introduce the detailed configuration of the hardware for Grid instance, and several mechanisms to construct the robust Grid system in the next generation system of KEKCC.
1. Introduction
KEK, the High Energy Accelerator Research Organization [1], is the largest accelerator science research laboratory using high-energy particle beams in Japan. Indeed, a lot of experiments and projects in the field of accelerator-based science are actively being carried out at KEK by using the SuperKEKB [2] and J-PARC [3] accelerator. Not only the accelerator, KEK provides various support based on the advanced technology to meet researcher’s demand as an Inter-University Research Institute Corporation across the countries. The Computing Research Center (KEK-CRC) located at Applied Research Laboratory in KEK provides IT infrastructures and computer resources for the simulation and analysis of the data collected by experiments.

In these days at KEK, the computing demand from the various experiments for the data processing, data analysis and production of Monte Carlo simulation (MC) data is monotonically increasing. It is not only for the case with high-energy experiments, the computing requirement from the hadron and neutrino experiments and some projects of astro-particle physics is also rapidly increasing due to the huge amount of data coming from the very high precision measurement based on the recent technology development.

Under this situation, several projects, Belle II [4], T2K [5], ILC [6] and KAGRA [7] experiments supported by KEK are going to utilize the Grid computing infrastructure as the main computing resource. The research and development of the Grid system and service were started at KEK-CRC in 2002, and the production services have already been deployed for the small virtual organizations (VO) since 2006. Figure 1 shows the fraction of CPU usage in KEK Central Computer System (KEKCC) from 2012 to 2015 [8] among the experiments supported by KEK. The fraction of Grid jobs are increasing and reached roughly 30% of the total CPUs.

The Grid system has been upgraded for the full-scale operation at the same time of hardware replacement of KEKCC. The next generation system of KEKCC has already been started the operation since at the beginning of September 2016. The new KEKCC has 10,000 CPU cores (236 kHS06), 13 PB disk storage, and 70 PB maximum capacity of tape archiving system [9]. The development of the Grid service at the new KEKCC is described in the following sections.

![Figure 1](image_url). Fraction of CPU usage at KEKCC, break down by groups from 2012 to 2015 [8]. Bars indicated by blue, red, and green correspond to the CPU usage from the local batch system. The fraction of Grid jobs is indicated as purple bars.
2. Construction of the robust Grid service

The Grid system and services at the new KEKCC have been developed by focusing on the robustness of each service instance to accommodate the requirement for the massive data production and MC production of the Belle II experiment. The basic Grid service instances, for example, information systems based on Berkeley Database Information Index (BDII) [10], two CREAM [11] computing elements with the aid of LSF [12] batch job scheduler, and some storage elements using Storage Resource Manager (StoRM) [13] are made by the redundant hardware configuration. Additionally, we applied high availability (HA) configurations to the database oriented services like Virtual Organization Membership Service (VOMS) [14], ARDA Metadata Grid Application (AMGA) [15], and LCG File Catalogue (LFC) [16] with the external RAID boxes as the central services by using SIOS LifeKeeper [17] commercial product. It enables us to make the HA configuration even if the running software itself does not support the HA configuration.

Figure 2 shows the schematic view on the configuration of the VOMS, AMGA, and LFC. The LFC servers are made by two read/write servers and two independent read-only servers to gain the access performance from the Belle II Grid. The read-only LFCs have replicated database in each SSD storage for the purpose of load balancing. One more read-write LFC is available for the other VOs to eliminate any interference with Belle II activities.

Although all of the service instances are connected to the UPS to avoid the instantaneous interruption of power supply, we have to take a few days downtime in every year due to the inspection and maintenance of the electric power supply facilities. We rearranged the power routing for the critical services referred from the other sites, for example, BDII, VOMS, AMGA, LFC, Argus [18] authorization service, and File Transfer Service (FTS3) [19], to achieve the uninterruptible operation against the scheduled power outage by introducing the external power source.
3. Deployment of the new central service

In addition to the basic Grid service instances, AMGA, FTS3, and CernVM File System (CVMFS) [20] stratum-0/1 servers are deployed as the new central services at KEKCC. Figure 3 shows the schematic view of the configuration for CVMFS stratum-0/1 and http-proxy, and repository update servers at KEKCC. The service of CVMFS stratum-0 is started for the Belle II software repository "belle.kek.jp". The CVMFS stratum-1 server retrieves the contents in other repositories stored in other stratum-1s. All of the servers are made by the redundant configuration.

We are taking care of security issue for the deployment of CVMFS stratum-0. The files created by the repository maintainers have been scanned automatically before the publication in the update cycle by the Sophos Anti-Virus software [21]. It is to avoid distributing any unwanted software and files to the other sites.

4. Reinforcement of the data transfer capability

Since the Belle II raw data transfer is one of the most important tasks for the KEKCC, two redundant GridFTP [22] servers dedicated to the Belle II raw data transfer are adapted to the special StoRM storage element. The Belle II raw data is expected to be sent by 3 GB/sec of the transfer rate from Belle II online storage to KEKCC storage. Each GridFTP server is connected by InfiniBand (4×FDR: 56 Gbps) to the KEKCC storage, and 4×10 Gbps Ethernet to the external network. The capability of total data throughput of HSM and disk storage in KEKCC corresponds to 50 GB/sec and 100 GB/sec, respectively. The network path is completely separated from the GridFTP servers for the data transfer usage of the other VOs and user analysis jobs via the Grid as shown in Figure 4.

The connectivity of the international network for the sites in Japan has been improved by the upgrade of SINET5 (Japanese National Research and Education Network) provided by NII [23] since April 2016. KEK has been connected to the SINET5 by 100 Gbps (+ 10 Gbps) [24]. We can utilize 100 Gbps + 10 Gbps of a transpacific network to the US and 20 Gbps of a Siberian network to Europe. The SINET5 has started the peering of LHCONE [25] route for
both ESnet [26] and GÉANT [27], and provided the route for KEK since September 2016 as shown in Figure 5.

KEKCC assigned 40 Gbps bandwidth for the Grid service instances including GridFTP servers, and the LHCONE traffic can bypass the KEKCC firewall. Therefore, there is no bottleneck for the Belle II data transfer from KEKCC to the sites in Belle II Grid.

The data transfer test has been actively ongoing by the collaboration of the Grid sites and Belle II computing group. One can find some initial results of the transfer test with new data transfer servers at the recent network environment elsewhere (see for example [28]).

![Figure 4. Schematic view of the configurations for the StoRM and GridFTP servers at KEKCC. The capability on the data transfer between the instances is indicated inside the figure.](image)

![Figure 5. The bandwidth of the international network connection, and the status of peering for LHCONE route at KEKCC [24].](image)
5. Summary

A lot of Grid services have been developed and deployed at KEK with the launch of the new KEK Central Computer System (KEKCC) at September 1st, 2016. The new Grid system is designed focusing on the system robustness to accommodate the requirement for the massive data production of the Belle II experiment. There are many kinds of improvements on the central Grid services for both service level and performance by introducing High Availability configuration. The critical central services, for example, BDII, VOMS, AMGA, LFC, and Argus are accomplished uninterruptible operation even at the maintenance period for the electric power supply facility. The data transfer performance is upgraded significantly by the high bandwidth international network provided by SINET with LHCONE routing and powerful GridFTP servers dedicated to the Belle II experiment. The smooth Belle II raw data transfer to the other sites can be expected without any interference originating from the other computing activities at KEK.

References

[1] KEK: https://www.kek.jp/en/
[2] SuperKEKB: http://www-superkekb.kek.jp/
[3] J-PARC: https://j-parc.jp/index-e.html
[4] Belle II: https://www.belle2.org/
[5] T2K: http://t2k-experiment.org/
[6] ILC: http://www.linearcollider.org/ILC
[7] KAGRA: http://gwcenter.icrr.u-tokyo.ac.jp/en/
[8] G. Iwai et al., ISGC2016, PoS (ISGC 2016) 037.
[9] K. Murakami et al., CHEP2016, J. Phys.: Conf. Ser. (to be published).
[10] BDII: http://gridinfo.web.cern.ch/
[11] CREAM: https://wiki.italiangrid.it/CREAM
[12] LSF: https://www-03.ibm.com/systems/spectrum-computing/products/lsf/
[13] StoRM: https://italiangrid.github.io/storm/
[14] VOMS: https://italiangrid.github.io/voms/index.html
[15] AMGA: http://amga.web.cern.ch/amga/
[16] LFC: http://lfcdm.web.cern.ch/lfc
[17] SIOS LifeKeeper: http://docs.us.sios.com/#LKSSP
[18] Argus: http://argus-documentation.readthedocs.io/en/latest/index.html
[19] FTs3: http://fts3-service.web.cern.ch/
[20] CVMFS: https://cvmfs.cern.ch/portal/filesystem
[21] Sophos Server Protection: https://www.sophos.com/en-us/products/server-security.aspx
[22] GridFTP: http://toolkit.globus.org/toolkit/docs/latest-stable/gridftp/
[23] SINET5: https://www.sinet.ad.jp/en/top-en
[24] S. Suzuki et al., HEPiX Fall 2017 Workshop, https://indico.cern.ch/event/531810/contributions/2298933/
[25] LHCONE: http://lhcone.web.cern.ch/
[26] ESnet: https://www.es.net/
[27] GÉANT: http://www.geant.org/
[28] S. Pardi, LHCOPN-LHCONE meeting, Helsinki, https://indico.cern.ch/event/527372/#3-belleii-update