Bulletin of The Nippon Dental University **48**：13-16. March 2019
http://doi.org/10.14983/00000854/
CODEN:NSDKDD ISSN 0385-1605

Construction of distributed computing site DIRAC.NDU.jp for Belle II experiment

Hiroaki ONO [1], Minori WATANABE [1], Kiyoshi HAYASAKA [2]

[1] Department of Physics, The Nippon Dental University School of Life Dentistry at Niigata, 1-8 Hamaura-Cho, Chuo-ku, Niigata 951-8580, Japan

[2] Department of Physics, Faculty of Science, Niigata University, 8580 2-no-cho, Ikarashi, Nishi-ku, Niigata 950-2181, Japan

Abstract: We have constructed a new distributed computing site, called DIRAC.NDU.jp, in The Nippon Dental University School of Life Dentistry at Niigata (NDU) aiming for the Monte Carlo (MC) simulation data production for Belle II experiment [1]. In this article, we introduce the details of DIRAC.NDU.jp and the situation of the performances of our site.

Key words: Belle II experiment, Distributed computing, DIRAC

（Accepted 8 March 2019）

1. Introduction

Belle II experiment has started taking beam collision data since 2018 with SuperKEKB accelerator, which had been upgraded from the Belle experiment with KEKB accelerator at High Energy Accelerator Research Organization (KEK) in Tsukuba, Japan.

Monte Carlo (MC) simulation data is used to understand the detailed physics processes including its intermediate state since only the final state information can be obtained from the experimental data. Therefore, a large statistics of MC simulation data production is required well before the data taking begins. To ensure the requirement of huge number of CPU power and disk spaces for MC data production, Belle II experiment adopts the DIRAC (Distributed Infrastructure with Remote Agent Control) [2] as distributed computing management platform using worldwide collaboration resources.

NDU-Niigata physics group has joined the Belle experiment from the beginning and started to join the Belle II collaboration since 2017 as a satellite member of Niigata University group. NDU group contributes to Belle II experiment with belonging to the distributed computing (DC) operation group as an expert of DC operation. Concurrently, we have started to construct new computing facilities to process MC production jobs as a part of Belle II DC sites.

2. Distributed computing structure in Belle II

DIRAC provides the software framework of distributed computing infrastructure and interface for management, which was initially produced for the LHCb experiment [3-5] but now commonly used for several high energy physics (HEP) experiments with implementing particular extension like BelleDIRAC for Belle II experiment and ILCDirac for ILC experiment.

Belle II DC sites consist of heterogeneous architectures; Grid base distributed computing sites, commercial cloud base sites, and non-grid type local clusters. Many universities, in particular Japanese universities, provide their relatively small local computing resources to Belle II for the contribution of MC production, where Secure Shell (SSH) base

job submission system is prepared to manage such resources without hosting their own Grid base servers. This job submission scheme is managed by Nagoya University DIRAC server and much easily involved those resources to distributed computing world. We called these sites as "DIRAC" sites and NDU-Niigata has also joined the distributed computing system as "DIRAC.NDU.jp" site.

3. Setup of computing facilities

Thanks to the research promotion budgets from our university, we purchased multicore, high density servers to contribute the Belle II MC production. Table 1 shows specifications of head node server, called Computing Element (CE) in DIRAC, and worker nodes (WNs). We gradually increase our WNs year by year since 2017 and 400 job slots can be utilized from July 2019.

High density multicore nodes are prepared for Belle II MC production dedicated usage in order to ensure the requirement for Belle II DC sites [6]; 2 GB memory and 10 GB disk space for scratch directory (/tmp) per job with Scientific Linux 6 (SL6) OS. Specifications of six computer nodes we prepare are Intel Xeon E5-2660 v4 2.0 GHz, 2×14 cores/28 threads; 56 job slots/node in total, 128 GB memory, and 4 TB disk space for CE or 500 GB to 1 TB disk space for other WNs used for scratch directory. In addition, 22 TB of RAID disk is equipped to store the beam background files (local SE), which are used to overlay the reaction to the detector by the real beam background to the one by the simulated event. To keep the enough bandwidth to transfer the files to destination storage element (SE), internal network is connected with 10 Gbps backbone L2 switch and connected to 1 Gbps network to outside. Figure 1 shows the latest setup of DIRAC.NDU.jp computer nodes. Switches and most of nodes are connected to uninterruptable power supply to prevent unexpected site down.

Figure 1 Setup of new computer nodes. Six nodes are installed in total at 2019.

Nodes	CPU	Memory	Disk	# of nodes
Computing element (CE)	Intel Xeon E5-2660 v4 2.0 GHz 28 cores/56 threads	128 GB	4 TB RAID 22 TB RAID for localSE	1
Worker node (WN)	Intel Xeon E5-2660 v4 2.0 GHz 28 cores/56 threads	128 GB	500 GB to 1 TB SSD	5
	Intel Xeon E31220 3.1 GHz 2 cores/4 threads	8 GB	128 GB SSD	3
	Intel Core i5-2500 3.3 GHz 2 cores/4 threads	8 GB	128 GB SSD	15

Table 1 Components of DIRAC.NDU.jp nodes. All WNs disk for /tmp is replaces as SSD to cope with heavy IO access.

We have also reused 15 of retired desktop computers which were used at IT center computer room for students and replaced at 2018. Retired computers are set up as WNs, whose specification is Intel Core i5 3.3 GHz 2 core/4 threads with hyper threading; 4 job slots in total, and increase to 8 GB memory with 128 GB disk. Figure 2 shows the array of reused desktop computers involved into site nodes.

On CE and all WNs, SL6 has been installed and CE home directory is shared by WNs via mounting Network File System (NFS) service. Job queue scheduling is managed by the Sun Grid Engine (SGE) and Belle II dedicated MC production software components are provided through the CernVM File System (CVMFS) service, which is commonly used for HEP community developed by CERN and easily distributes software environment to worldwide DC sites. A web cache proxy service is also prepared on CE using the Squid service to prevent the WNs duplicating file access to outside.

Figure 2 Reused computers for WNs replaced at IT center computer room.

4. Performance of DIRAC.NDU.jp

From May 2016, DIRAC.NDU.jp site had launched with a small desktop server and started to process the Belle II MC production jobs, then computing facilities are gradually increased year by year. Figure 3 and 4 show the growing of the number of processed jobs on our site from 2016 April to 2019 February. A large number of jobs have been submitted in the series of production period, called "MC production campaign", and busy period or less active period have been observed in Fig. 3 which is correlated with the period of MC campaign. In the MC9 campaign period, our site faced heavy disk Input/Output (IO) access by multi-file extraction in jobs and disks remained in high load state with low responsibility. In order to solve this situation, whole WNs disks were replaced with solid-state disk (SSD) to cope with the heavy disk IO access. In 2019, six of high-density servers and many reused desktops are hosted utilizing 400 job slots and around 200k of MC production jobs have been executed as shown in Fig. 4.

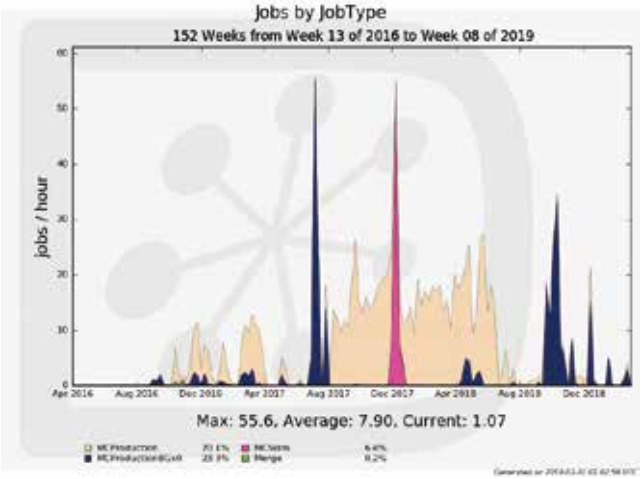

Figure 3 Processed jobs at DIRAC.NDU.jp from Apr. 2016 to Feb. 2019.

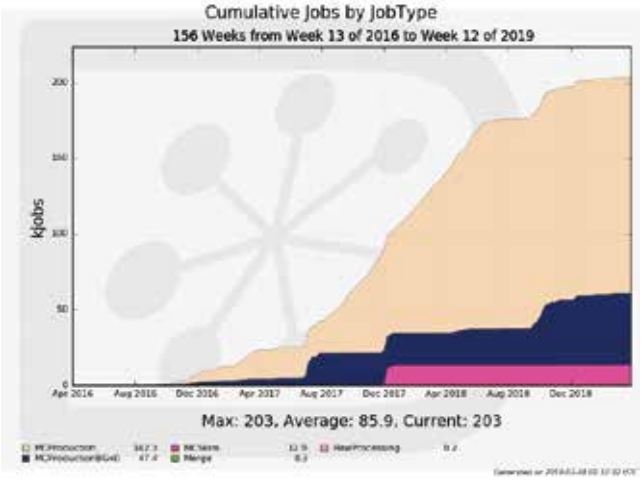

Figure 4 Cumulative number of processed jobs at DIRAC.NDU.jp with different job types from Apr. 2016 to Feb. 2019.

Figure 5 shows the summary of CPU usage among the sites in Japanese institutes from Apr. 2016 to Feb. 2019. DIRAC.NDU.jp has already been fourth largest site in the Japanese sites or third largest site among the Japanese university sites.

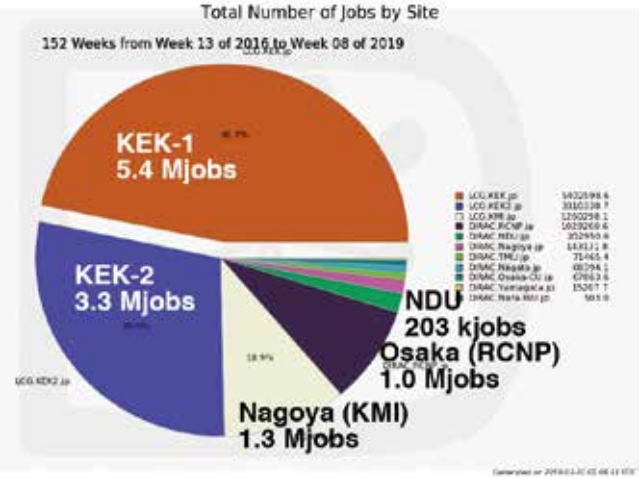

Figure 5 CPU usages summary among the Japanese sites.

5. Conclusion and Acknowledgement

We have constructed new distributed computing site, called DIRAC.NDU.jp, for the Belle II MC production. Our site started as small test site from 2016 but currently increased to 400 job slots for Belle II dedicated MC production jobs. At the present, all nodes use SSD and additional 22 TB disk space is prepared for the smooth and efficient operation in order to allocate enough background overlay files. We are planning to increase our job slots greater than 500 job slots and host new storage element (NDU-SE) in our site for further contribution to Belle II experiment.

We would like to thank to the members of Belle II Distributed Computing group and Data Production group for the stable operation of DIRAC servers and efficient MC data production.

This work is supported by NDU research promotion grant number NDU Grants N-16001, N-17001, N-18001.

References

1) T. Abe et al (Belle-II Collaboration) Belle II Technical Design Report KEKB, Nov 2010. 480 pp.
2) F. Stagni et al., DIRAC in Large Particle Physics Experiments, J. Phys. Conf. Ser. 898 (2017), no. 9 092020
3) S. Amato et al., LHCb Collaboration, Technical Proposal, CERN-LHCC/98-4.
4) R. Antunes Nobrega et al., LHCb Collaboration, LHCb Reoptimized Detector Design and Performance TDR, CERN-LHCC/2003-030
5) R. Antunes Nobrega et al., LHCb Collaboration, LHCb Trigger TDR, CERN- LHCC/2003-031
6) https://operations-portal.egi.eu/vo/view/voname/belle