

Overview and Highlights of the Belle II Computing

Yuji Kato^{*a} and Kiyoshi Hayasaka^a on behalf of the Belle II computing group^b

^a *Kobayashi-Maskawa Institute for the Origin of Particles and the Universe, Nagoya University, Chikusa-ku Furo-cho, Nagoya, Japan*

^b <https://belle2.cc.kek.jp/wiki/pub/Public/ComputingPublic/AuthorList4Belle2Computing.tex>

E-mail: kato@hepl.phys.nagoya-u.ac.jp

The Belle II is a next-generation B-factory experiment which aims to find the physics beyond the Standard Model with a data sample of 50 ab^{-1} integrated luminosity. In order to provide computing resources which can process and store such a huge data sample, a distributed computing model is adopted. Nagoya team leads monitoring group in Belle II computing to maximize the active computing resources. Also, we are contacting many Asian universities which have local computing resources to make them join Belle II computing using the method which does not need GRID middleware. In this paper, overview of the Belle II computing and activities of the Nagoya team are described.

Flavor Physics & CP Violation 2015

May 25-29, 2015

Nagoya, Japan

*Speaker.

1. Introduction

Belle II is a next-generation B-factory experiment at the SuperKEKB accelerator in Japan, which aims to find the physics beyond the Standard Model from the precise measurement of the decay of B-meson, τ lepton, charm meson and so on. The final target for the peak luminosity of the SuperKEKB is $8 \times 10^{35} \text{ cm}^2/\text{s}$, which corresponds to about 40 times higher than that of the KEKB, the predecessor of the SuperKEKB. The detail of the Belle II experiment can be found in [1]. The physics run without vertex detector will started in 2017 and the one with full Belle II detector will start in 2018. A data sample of 50 ab^{-1} will be accumulated at the end of the data taking. The computing resources required to process and store such a large data sample are: 1MHS06 CPU resources, 200 PB tape for raw data and 100 PB disk for processed and simulated data. The Belle II experiment has adopted a distributed computing model to provide the computing resources. In this paper, overview of the Belle II computing and activities of Nagoya team are described.

2. Overview of the Belle II distributed computing system.

The computing resources are provided by collaborative institutes around the world. The role of a computing site depends on the scale of the site as follows:

- ‘*Raw Data Center*’, which stores raw data and process it to produce ‘mDST’, which is a root-based format file containing all necessary information for physics analysis. KEK, where the experiment is performed, plays this role. In addition, big computing sites in each region i.e., North America, Europe and Asia store one copy of the raw data. In the first 3 years of the experiment, PNNL (Pacific Northwest National Laboratory) will keep the whole copy. From the 4th year, the copy will be distributed in other big computing sites.
- ‘*Regional Data Center*’, where mDST files for data are distributed.
- ‘*MC production site*’, which is relatively small site, performs the Monte-Carlo (MC) data production and user skim to produce Ntuple.

The schematic view of the Belle II computing model is shown in Fig. 1.

We have been developing the computing system based on existing technologies. We choose DIRAC[2] as a workload and data management system. It provides a common interface to a number of heterogeneous resources such as grids with different middlewares, cloud and local computing clusters. A file catalog is provided by LCG File Catalog[3]. A meta data catalog is provided by ARDA Metadata Catalog (AMGA)[4]. The Belle II software is distributed by CernVM-FS (CVMFS)[5]. Distribution of data is performed by FTS3 (File Transfer Service, version 3.0)[6]. In order to test and understand the bottleneck of the computing system, Belle II computing group has performed MC mass production campaigns four times. Computing resources have been increased gradually. We have solved scalability issue by an improvement of in the AMGA[7, 8] or load valancing of DIRAC servers. Now, 18k concurrent jobs on 31 sites over 15 countries can be handled.

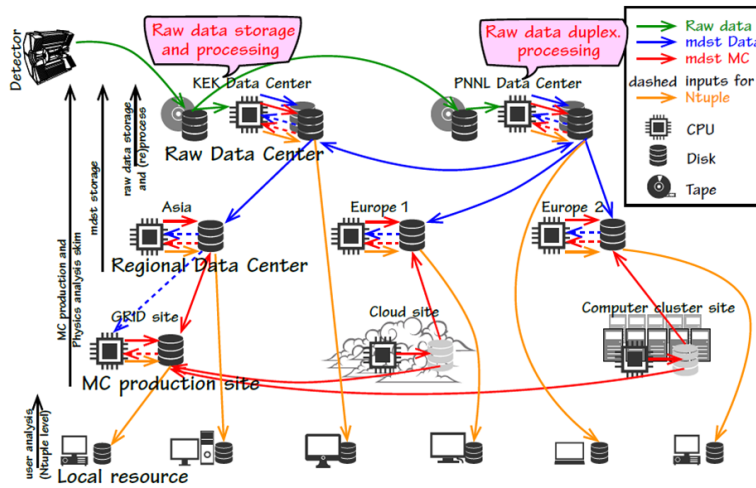


Figure 1: Schematic view of the Belle II computing model.

3. Monitoring system

Because the required computing resource is huge and the number of interfaces to execute jobs or transfer data is large in the distributed computing system, the monitoring system is important to utilize the computing resource effectively. We are developing the monitoring system in two ways. One is active way, where test jobs to check the status of each site are submitted or the test file transfer is performed to storage and so on. Another is the passive way, where the information on jobs stored in DIRAC databases is processed and visualized. In both cases, results are visualized in the web as shown in Fig. 2. More details are described in [9, 10].

Site status summary

site	worker node	CPU	#core	memory	OS	Kernel	rpm	cvmsfs	releases	CPU Norm.	last updated
CLOUD.AWS.Singapore.sg	ip172-31-12-162	Intel(R) Xeon(R) CPU E5-2670 v2 @ 2.50GHz	x1	3768MB	None	3.14.25-28.38.amzn2.x86_64	7 problem found	Rev. 52	OK (build-2015-02-09)	6.6 H506	2015/04/02 17:07:26
CLOUD.AWS.Sydney.au	ip172-31-5-30	Intel(R) Xeon(R) CPU E5-2670 v2 @ 2.50GHz	x1	3768MB	None	3.14.25-28.38.amzn2.x86_64	7 problem found	Rev. 52	OK (build-2015-02-09)	6.3 H506	2015/04/02 20:44:11
CLOUD.AWS.Tokyo.jp	ip172-31-5-18	Intel(R) Xeon(R) CPU E5-2670 v2 @ 2.50GHz	x1	3768MB	None	3.14.24-27.48.amzn2.x86_64	7 problem found	Rev. 52	OK (build-2015-02-09)	6.5 H506	2015/04/02 16:25:16
DIRAC.Bethesda.co	hepf02.buaa.edu.cn	Intel(R) Xeon(R) CPU E5-2650 v2 @ 2.60GHz	x32	1004MB/cores	Scientific Linux release 5.8 (Eaton)	2.6.18-0081.1.e5	OK	Rev. 52	OK (build-2015-02-09)	12.8 H506	2015/04/01 09:05:42
DIRAC.BNP.ru	sec-138-2.belle2	GEMU Virtual CPU version 0.8.1	x2	1508MB/cores	Scientific Linux release 5.11 (Eaton)	2.6.18-400.e5	OK	NL	OK (build-2015-02-09)	1.06 H506	2015/04/02 20:34:16

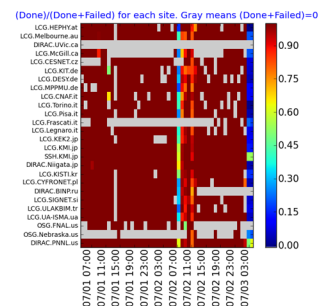
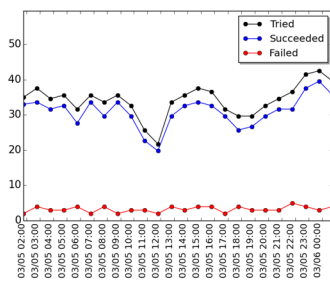


Figure 2: Outputs of the monitoring system. Top table shows the result of the test job to check the status of sites. Information on CPU, memory, number of cores, OS and status of the CVMFS etc are shown. Bottom left plot shows the history of the job submission status. Right plot shows the time profile of the efficiency of the job execution for each site.

4. Participation of local clusters to Belle II computing

Many universities have their own clusters on which grid middleware is not installed. DIRAC can submit jobs on traditional cluster sites via the SSH protocol. Nagoya team has contacted many Asian universities to configure their systems to accept job submission from DIRAC central servers. We have obtained many experiences to handle traditional cluster sites, and have supported their operation. Currently, more than 10 local cluster sites join the Belle II computing and executing MC production jobs stably.

5. Summary

We have been developing Belle II distributed computing system towards the start of the physics run in 2017. We adopt DIRAC as the workload and data management system and utilize many other existing technologies. Current system can handle 18 K concurrent jobs. However, this number is still smaller than finally required one. We need more tests for the scalability and improvements for the system. Nagoya team contributes by developing monitoring system and working as 'local cluster center'.

6. Acknowledgments

We are grateful for the support and the provision of computing resources by CoEPP in Australia, HEPHY in Austria, McGill HPC in Canada, CESNET in the Czech Republic, DESY, GridKa, LRZ/RZG in Germany, INFN-CNAF, INFN-LFN, INFN-LNL, INFN Pisa, INFN Torino, ReCaS (Univ. & INFN) Napoli in Italy, KEK-CRC, KMI in Japan, KISTI GSDC in Korea, Cyfronet, CC1 in Poland, NUSC, SSCC in Russia, SiGNET in Slovenia, ULAKBIM in Turkey, UA-ISMA in Ukraine, and OSG, PNNL in USA. We acknowledge the service provided by CANARIE, Dante, ESnet, GARR, GEANT, and NII. We thank the DIRAC and AMGA teams for their assistance and CERN for the operation of a CVMFS server for Belle II.

References

- [1] T. Abe *et al.* [Belle-II Collaboration], arXiv:1011.0352 [physics.ins-det].
- [2] <http://diracgrid.org/>.
- [3] <https://twiki.cern.ch/twiki/bin/view/LCG/LfcGeneralDescription>.
- [4] S. Ahn *et al.* Journal of the Korean Physical Society 57 issue 4 715, 2010.
- [5] <http://cernvm.cern.ch/portal/>.
- [6] “FTS3: New Data Movement Service For WLCG”, proceedings of the CHEP2013 conference.
- [7] “Improvement of AMGA Python Client Library for the Belle II Experiment”, J. H. Kwak, proceedings of the CHEP2015 conference.
- [8] “Directory Search Performance Optimization of AMGA for the Belle II Experiment”, G. Park, proceedings of the CHEP2015 conference.
- [9] “Monitoring system for the Belle II distributed computing”, K. Hayasaka, proceedings of the CHEP2015 conference.
- [10] “Job monitoring on DIRAC for Belle II distributed computing”, Y. Kato, proceedings of the CHEP2015 conference.