CCJ operation in 2013

Y. Ikeda,[*1] H. En'yo,[*1] T. Ichihara,[*1] Y. Watanabe,[*1] and S. Yokkaichi[*1]

1 Overview

The RIKEN Computing Center in Japan (CCJ)[1] commenced operations in June 2000 as the largest offsite computing center for the PHENIX[2] experiment being conducted at the RHIC[3]. Since then, the CCJ has been providing numerous services as a regional computing center in Asia. We have transferred several hundred TBs of raw data files and nano data summary tape (nDST) files, which is a term for a type of summary data files at PHENIX, from the RHIC Computing Facility (RCF)[4] to the CCJ. The transferred data files are first stored in a High Performance Storage System (HPSS)[5] before performing the analysis. The CCJ maintains sufficient computing power for simulation and data analysis by operating a PC cluster running a PHENIX-compatible environment.

A joint operation with the RIKEN Integrated Cluster of Clusters (RICC)[6] was launched in July 2009. Twenty PC nodes have been assigned to us for dedicated use, sharing the PHENIX computing environment.

Many analysis and simulation projects are being carried out at the CCJ, and these projects are listed on the web page `http://ccjsun.riken.go.jp/ccj/proposals/`. As of December 2013, CCJ has been contributed 31 published papers and more than 33 doctoral theses.

2 Configuration

2.1 Calculation nodes

In our machine room 258/260 in the RIKEN main building, we have 28 PC nodes[a], and these nodes have been used for the analysis of the PHENIX nDST data. Table 1 lists the numbers of job slots, CPU threads, and CPU cores, and the number of nodes (there is no change in 2013). These nodes are operated using a data-oriented analysis scheme that carries out optimization using local disks[7][8]. The OS on the calculation nodes is Scientific Linux (SL) 5.3[9], and the same OS works on the 20 nodes at the RICC. As a batch-queuing system, LSF 8.0.0[10] and Condor 7.4.2[11] were run on the CCJ and RICC nodes, respectively, as of Dec 2013.

Table 2 lists the numbers of malfunctioned SATA or SAS disks in the HP servers (including NFS/AFS servers described in the next section).

Table 1. Limitation of number of job slots from LSF queue with cluster node.

	Nodes	Cores	Threads	Jobs
CCJ-hp1	18	144	144	180
CCJ-hp2	10	120	240	200
RICC	19	152	152	144
Total	47	416	536	524

Table 2. Malfunctioned HDDs in 2011, 2012, and 2013

			Malfunctioned		
Ype	Size	Total	2013	2012	2011
SATA	1 TB	192	16	20	9
	2 TB	120	2	5	4
SAS	146 GB	38	0	1	1
	300 GB	24	0	0	1

2.2 Data servers

Two data servers (HP ProLiant DL180 G6 with 20 TB SATA raw disks) are used to manage the RAID framework of the internal hard disks, which contain the user data and nDST files of PHENIX. The disks are not NFS-mounted on the calculation nodes to prevent performance degradation by process and network congestion. These disks can be accessed only using the "rcpx" command, which is the wrapper program of "rcp" developed at CCJ, and it has an adjustable limit for the number of processes on each server.

The Domain Name System, Network Information System, Network Time Protocol, and Network File System servers are operated on the server ccjnfs20[b] with a 10-TB FC-RAID, where the users' home and work spaces are located. The home and work spaces are formatted with VxFS 5.0[12]. The backup of home spaces on ccjnfs20 is saved to another disk server once a day and to HPSS once a week. The backups on HPSS are stored for three weeks.

2.3 HPSS

Since Dec 2008, the HPSS servers and the tape robot have been located in our machine room, although they are owned and operated by RICC. The specifications of the hardware used can be found in the literature[13]. The amount of data and the number of files archived in the HPSS were approximately 1.7 PB and 2.1 million files, respectively, as of Dec 2013. Table 3 lists the files and the current class of service (COS) in the HPSS. No new file has been added in 2013.

*1 RIKEN Nishina Center

a) HP ProLiant DL180 G5 with dual Xeon E5430 (2.66 GHz, 4 cores), 16 GB memory and 10 TB local SATA data disks for each node, and HP ProLiant DL180 G6 with dual Xeon X5650 (2.66 GHz, 6 cores), 24 GB/20 TB as above, for each node

b) SUN Enterprise M4000 with Solaris 10

Table 3. DST and raw data files in HPSS on Dec 31, 2013

	DST		Raw data	
Run	Size [TB]	COS	Size [TB]	COS
1	4	2,3,100	3	3,205
2	24	2,3,4,100	36	1,3,5,205
3	10	2,3,6	46	100,205
4	14	2,3	11	205
5	287	2,3,6,100	292	5,205
6	92	3,6,100	339	11,100
8	22	3	128	12
9	106	3,7	13	
10	32	3	0	
11	142	3	0	
12	3	3	0	
Total	736		854	

3 Data transfer from BNL and the PHENIX software environment

Data collected during the PHENIX experiment was transferred from the RCF to the CCJ using grid-FTP [16] through the science information nework (SINET) 4 (maintained by NII[17]) with a 10 Gbps bandwidth. The data which transfered from BNL is moved to local disks on the HP calculation nodes and the HPSS. The files are transferred using grid-FTP at a maximum speed of about 300 MB/s. Two PostgreSQL[14] server nodes are operated for the PHENIX database, whose data size was 285 GB as of Dec 2013. The data are copied from the RCF everyday and made accessible to the users. One AFS[15] server node is operated for the PHENIX AFS. The size of the libraries for the PHENIX analysis setup was 1.7 TB as of Dec 2013. The libraries are also copied from the RCF by AFS everyday.

3.1 Uninterruptible power-supply system (UPS)

The power consumption of the CCJ system, excluding the HPSS, is about 25 kW, and the power is supplied through five UPSs (10.5 kVA each) as of Dec 2013. For the HPSS, there is one 7.5-kVA UPS for 100 V and three 10.5-kVA UPSs for 200 V purchased by CCJ. The batteries of the three UPSs expire in 2014 (One have expired in 2013).

3.2 Login server upgrade

CCJ had two login servers, ccjsun (HP Proliant DL145) with SL 5.3 and ccjgw (Supermicro 5011E) with CentOS 3, in 2012. The ccjgw server was upgraded to the HP Proliant DL145 with SL 5.3 in 2013.

References

1) S. Yokkaichi et al.: RIKEN Accel. Prog. Rep. **44**, p228 (2011).
2) http://www.phenix.bnl.gov/
3) http://www.bnl.gov/rhic/
4) https://www.racf.bnl.gov/
5) http://www.hpss-collaboration.org/
6) http://accc.riken.jp/ricc/
7) T. Nakamura et al.: RIKEN Accel. Prog. Rep. **43**, p167 (2010)
8) J. Phys.: Conf. Ser. **331**, 072025 (2011).
9) http://www.scientificlinux.org/
10) http://www-03.ibm.com/systems/technicalcomputing/platformcomputing/products/lsf/index.html
11) http://www.cs.wisc.edu/condor/description.html
12) Veritas file system (Symantec Corporation).
13) S. Yokkaichi et al.: RIKEN Accel. Prog. Rep. **42**, p223 (2009).
14) http://www.postgresql.org/
15) http://www.openafs.org/
16) http://www.globus.org/toolkit/docs/latest-stable/gridftp/
17) http://www.nii.ac.jp/