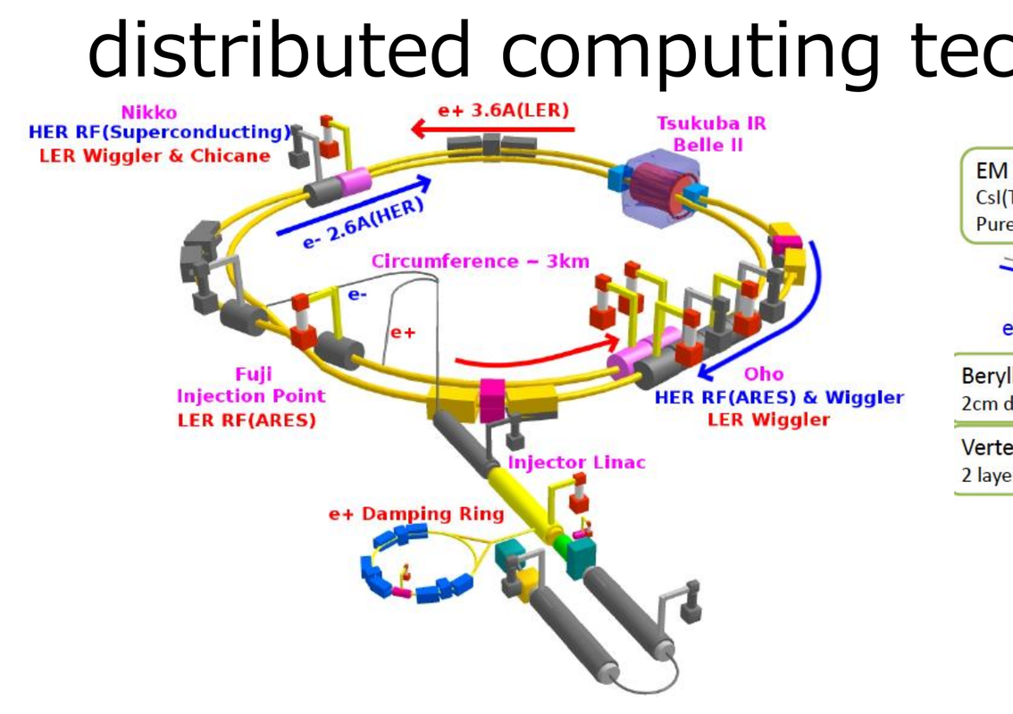
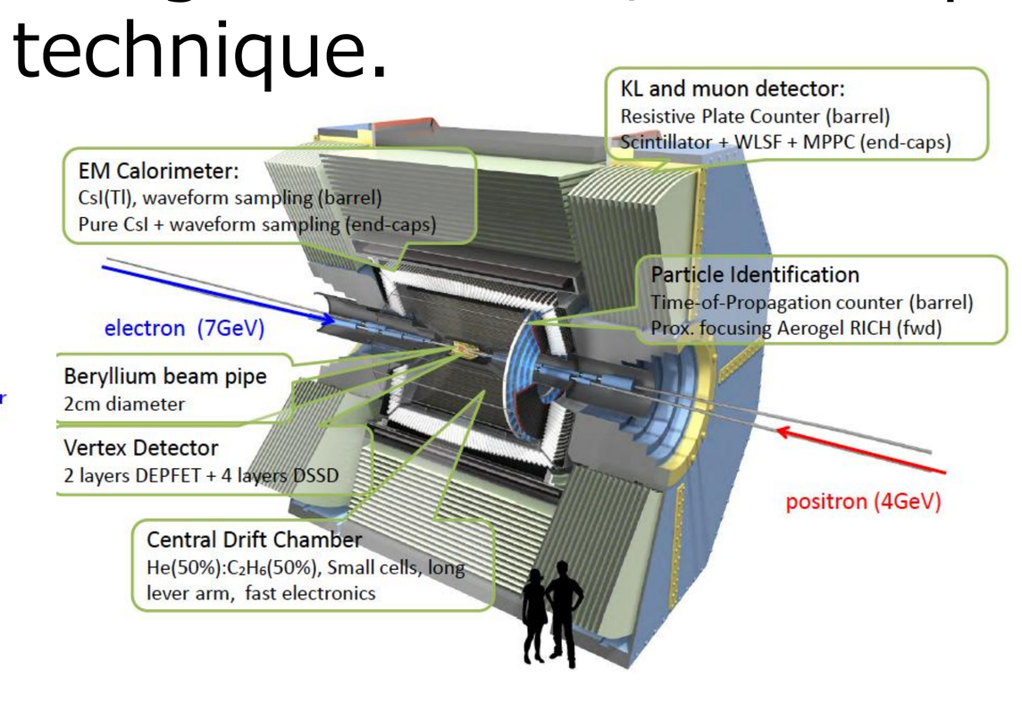
Monitoring system for the Belle II distributed computing (Poster ID: 314)

K.Hayasaka (KMI, Nagoya U.), Y.Kato (KMI, Nagoya U.), T.Hara (KEK), H.Miyake (KEK), I.Ueda (U. Tokyo/KEK) for Belle II computing group

Introduction

- Belle II experiment is a next-generation B-factory at KEK in Japan, which will start for physics run w/o (with) vertex detector in 2017 (2018), where $50ab^{-1}$ data sample will be collected for 10 years, which corresponds to about 5×10^{10} $B\bar{B}$-pair events.
- We roughly need to handle 1MHS06 CPU resources, 100PB storage for one set of raw data and 100PB one for MC/analysis data, finally.
- In order to utilize these huge resources, we adopt distributed computing technique.





KEKB superKEKB

Vertical β function: 5.9 mm → 0.27/0.30 mm (x20)

Beam current: 1.7/1.4 A → 3.6/2.6 A (x2)

→ $L = 2 \times 10^{34}\ cm^{-2}s^{-1} \rightarrow 8 \times 10^{35}\ cm^{-2}s^{-1}$ (x40)

SVD: 4 DSSD lyrs → 2 DEPFET lyrs + 4 DSSD lyrs

CDC: small cell, long lever arm

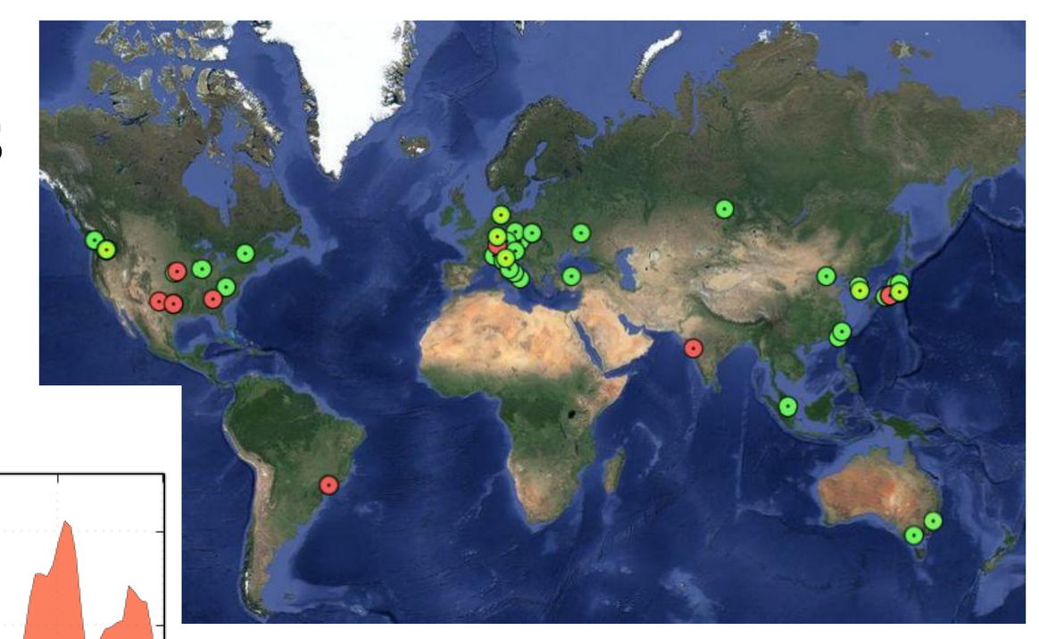
ACC+TOF → TOP+A-RICH

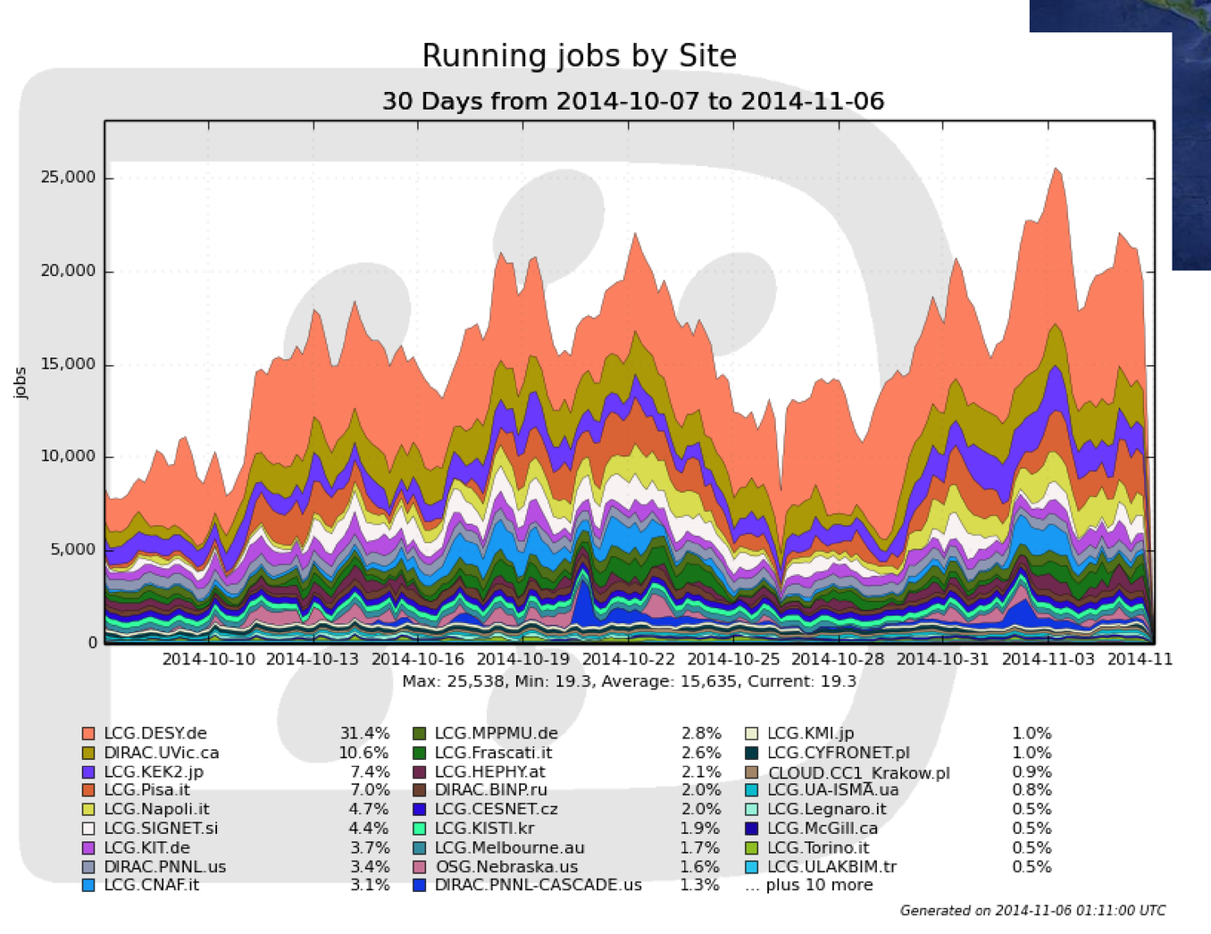
ECL: waveform sampling, pure CsI for end-caps

KLM: RPC → Scintillator +SiPM (end-caps)

Belle II computing

- Belle II has adopted DIRAC as the distributing computing software framework, which can handle grid, cloud and local cluster resources. (http://diracgrid.org/)
- CVMFS is used to provide Belle II software and libraries.
- At the present, around 40 sites participate (LCG, OSG, HPC, cloud and traditional cluster) and 25K concurrent jobs are handled at peak.





Troubles during operation

- We observed different types of troubles so far:
  - ➢ CVMFS
    - ✓ CVMFS is not running.
    - ✓ CVMFS is running with some error.
      - When some file is accessed, I/O error appears.
    - ✓ Files on CVMFS are not up-to-date.
  - ➢ Computing Element (CE)
    - ✓ CE is down.
    - ✓ CE is alive but batch job system is down.
    - ✓ Server Certification on CE is expired.
  - ➢ Worker Node (WN)
    - ✓ WN has hardware trouble, mostly, HDD failure.
    - ✓ WN does not have enough free space for HDD.
    - ✓ Required package is not installed.
    - ✓ Clock is not adjusted.
  - ➢ Central DIRAC servers
    - ✓ Due to the heavy load, server can not reply.

Monitoring

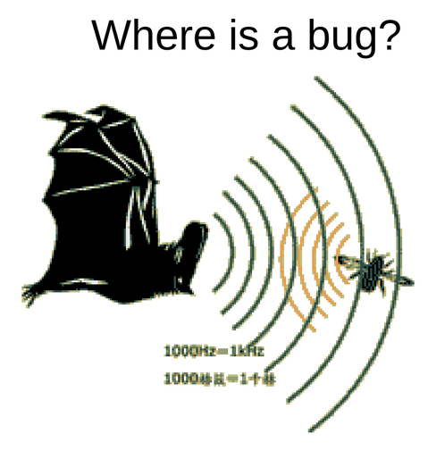
- To maximize the operation efficiency of the resources = increase of the resources
- To do so, it is important to find any problem faster.

→Effective monitoring system is necessary!

Two kinds of monitoring:

Active way

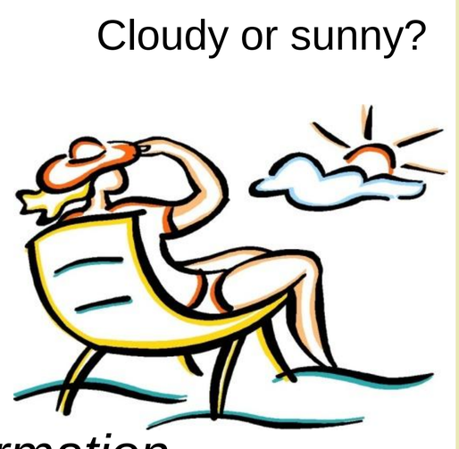
Where is a bug?



- *Make some action and check the result.*
  →Submit test job to check WN condition, or to check CE reaction.
  Try to open the DIRAC ports.

Passive way

Cloudy or sunny?



- *Check the result of the action for the operation.*
  →Pilot/Job behavior, log of pilot/job or information inside DIRAC.

Of course, DIRAC equips some passive-way monitoring systems such as job accounting, etc…

<u>We are developing the active-way monitor!</u>

to check…

1. Worker node status by submitting diagnosing job.
2. CE health by test submission.
3. DIRAC load by port check.

These results are stored in DB and summarized/visualized in the web pages.

<u>We are also developing the passive-way monitor.</u> → See Poster#337.

SiteCrawler

- We have developed diagnosing job submitting system to check CPU, size of memory, OS, kernel version, installed rpm packages, clock adjustment, free disk space, http proxy accessibility, cvmfs status, Belle II lib. files' md5sum and make a test execution for Belle II software on a WN.
- Every one hour, one job is submitted to each site.

Site status summary

site	worker node	CPU	#core	memory	OS	Kernel	rpm	cvmfs	releases	CPU Norm.	last updated
CLOUD.AWS.Singapore.sg	ip-172-31-12-162	Intel(R) Xeon(R) CPU E5-2670 v2 @ 2.50GHz	x1	3768MB	None	3.14.35-28.38.amzn1.x86_64	7 problems found	Rev. 52	OK (build-2015-02-09)	6.6 HS06	2015/04/02 17:07:36
CLOUD.AWS.Sydney.au	ip-172-31-5-30	Intel(R) Xeon(R) CPU E5-2670 v2 @ 2.50GHz	x1	3768MB	None	3.14.35-28.38.amzn1.x86_64	7 problems found	Rev. 52	OK (build-2015-02-09)	6.3 HS06	2015/04/02 20:44:11
CLOUD.AWS.Tokyo.jp	ip-172-31-5-18	Intel(R) Xeon(R) CPU E5-2670 v2 @ 2.50GHz	x1	3768MB	None	3.14.34-27.48.amzn1.x86_64	7 problems found	Rev. 52	OK (build-2015-02-09)	6.5 HS06	2015/04/02 16:25:16
DIRAC.Beihang.cn	hep102.buaa.edu.cn	Intel(R) Xeon(R) CPU E5-2650 v2 @ 2.60GHz	x32	1004MB/cores	Scientific Linux release 5.8 (Boron)	2.6.18-308.1.1.el5	OK	Rev. 52	OK (build-2015-02-09)	12.8 HS06	2015/04/01 09:05:42
DIRAC.BINP.ru	sscc-139-2.belle2	QEMU Virtual CPU version 0.9.1	x2	1508MB/cores	Scientific Linux release 5.11 (Boron)	2.6.18-400.el5	OK	N.I.	OK (build-2015-02-09)	10.6 HS06	2015/04/02 20:34:16
DIRAC.PNNL.us	cwn26	AMD Opteron 62xx class CPU	x32	1950MB/cores	Scientific Linux release 6.5 (Carbon)	2.6.32-431.17.1.el6.x86_64	OK	Rev. 52	OK (build-2015-02-09)	4.3 HS06	2015/04/02 19:14:38
DIRAC.TIFR.in	bellenet.ehep.tifr.res.in	Intel(R) Xeon(R) CPU E5-2640 0 @ 2.50GHz	x24	1336MB/cores	Scientific Linux release 6.6 (Carbon)	2.6.32-504.1.3.el6.x86_64	OK	Rev. 51	OK (build-2015-02-09)	11.6 HS06	2015/03/19 10:26:08
DIRAC.Tokyo.jp	belgdh.u-tokyo.jp.hep.net	Intel(R) Celeron(R) CPU G1610 @ 2.60GHz	x2	1901MB/cores	Scientific Linux CERN SLC release 6.6 (Carbon)	2.6.32-431.29.2.el6.x86_64	OK	Rev. 52	OK (build-2015-02-09)	11.3 HS06	2015/04/02 16:38:20
LCG.DESY.de	grid-wn0242.desy.de	AMD Opteron(TM) Processor 6234	x48	2688MB/cores	Scientific Linux release 6.6 (Carbon)	2.6.32-504.8.1.el6.x86_64	OK	Rev. 52	OK (build-2015-02-09)	4.9 HS06	2015/04/02 20:45:57
LCG.Frascati.it	atlaswn023.lnf.infn.it	Intel(R) Xeon(R) CPU E5520 @ 2.27GHz	x16	1501MB/cores	Scientific Linux release 6.5 (Carbon)	2.6.32-504.3.3.el6.x86_64	2 problems found	Rev. 52	OK (build-2015-02-09)	7.5 HS06	2015/04/02 20:52:39
LCG.HEPHY.at	r2b0h.grid.local	Intel(R) Xeon(R) CPU E5-2680 v2 @ 2.80GHz	x40	1609MB/cores	Scientific Linux release 6.6 (Carbon)	2.6.32-504.12.2.el6.x86_64	4 problems found	Rev. 52	OK (build-2015-02-09)	14.0 HS06	2015/04/02 20:13:51
LCG.KEK2.jp	ccb0153.cc.kek.jp	Intel(R) Xeon(R) CPU X5670 @ 2.93GHz	x12	4012MB/cores	Scientific Linux SL release 5.6 (Boron)	2.6.18-400.1.1.el5	one problem found	Rev. 52	OK (build-2015-02-09)	12.5 HS06	2015/04/02 20:52:12
LCG.NTU.tw	bgrid1.phys.ntu.edu.tw	AMD Opteron(tm) Processor 6134	x32	4032MB/cores	Scientific Linux CERN SLC release 6.6 (Carbon)	2.6.32-504.3.3.el6.x86_64	OK	Rev. 52	OK (build-2015-02-09)	7.3 HS06	2015/04/02 19:35:51
LCG.Pisa.it	ne1wn5	Quad-Core AMD Opteron(tm) Processor 8350 HE	x32	1009MB/cores	Scientific Linux release 6.6 (Carbon)	2.6.32.12-0.7-default	2 problems found	N.I.	OK (build-2015-02-09)	6.3 HS06	2015/04/02 21:07:52
LCG.SIGNET.si	f9nd128	Six-Core AMD Opteron(tm) Processor 2431	x12	2681MB/cores	Scientific Linux release 6.4 (Carbon)	3.14.27-gentoo	2 problems found	N.I.	OK (build-2015-02-09)	7.6 HS06	2015/04/02 21:01:49
LCG.Torino.it	t2-wn-018.to.infn.it	Intel(R) Xeon(R) CPU E5420 @ 2.50GHz	x8	1993MB/cores	CentOS release 6.6 (Final)	2.6.32-504.8.1.el6.x86_64	3 problems found	Rev. 52	OK (build-2015-02-09)	9.9 HS06	2015/04/02 19:43:22
LCG.UA-ISMA.ua	n106	Intel(R) Xeon(R) CPU E5345 @ 2.33GHz	x8	2009MB/cores	Scientific Linux CERN SLC release 6.6 (Carbon)	3.10.69-1.el6.elrepo.x86_64	one problem found	Rev. 52	OK (build-2015-02-09)	9.5 HS06	2015/04/02 19:29:38
LCG.ULAKBIM.tr	lufer117.ulakbim.gov.tr	AMD Opteron(tm) Processor 6174	x48	2688MB/cores	Scientific Linux release 6.3 (Carbon)	2.6.32-504.3.3.el6.x86_64	OK	Rev. 52	OK (build-2015-02-09)	7.1 HS06	2015/04/02 20:11:25

click! You can see information for the checked WNs.

WN status summary for LCG.ULAKBIM.tr

worker node	CPU	#core	memory	OS	Kernel	rpm	cvmfs	releases	CPU norm.	last updated
lufer115.ulakbim.gov.tr	AMD Opteron(tm) Processor 6174	x48	2688MB/cores	Scientific Linux release 6.3 (Carbon)	2.6.32-504.3.3.el6.x86_64	OK	Rev. 51	OK (build-2015-02-09)	7.2 HS06	2015/03/17 00:00:59
lufer116.ulakbim.gov.tr	AMD Opteron(tm) Processor 6174	x48	2688MB/cores	Scientific Linux release 6.3 (Carbon)	2.6.32-504.3.3.el6.x86_64	OK	Rev. 51	OK (build-2014-10-18)	7.1 HS06	2015/03/14 22:09:08
lufer117.ulakbim.gov.tr	AMD Opteron(tm) Processor 6174	x48	2688MB/cores	Scientific Linux release 6.3 (Carbon)	2.6.32-504.3.3.el6.x86_64	OK	Rev. 52	OK (build-2015-02-09)	7.1 HS06	2015/04/02 20:11:25
lufer118.ulakbim.gov.tr	AMD Opteron(tm) Processor 6174	x48	2688MB/cores	Scientific Linux release 6.3 (Carbon)	2.6.32-504.3.3.el6.x86_64	OK	Rev. 52	OK (build-2015-02-09)	7.2 HS06	2015/03/28 09:58:46
lufer119.ulakbim.gov.tr	AMD Opteron(tm) Processor 6174	x48	2688MB/cores	Scientific Linux release 6.3 (Carbon)	2.6.32-504.3.3.el6.x86_64	OK	Rev. 52	OK (build-2015-02-09)	7.1 HS06	2015/04/01 06:13:18
lufer120.ulakbim.gov.tr	AMD Opteron(tm) Processor 6174	x48	2688MB/cores	Scientific Linux release 6.3 (Carbon)	2.6.32-504.3.3.el6.x86_64	OK	Rev. 52	OK (build-2015-02-09)	7.2 HS06	2015/04/02 09:03:53

shaded row means that information is older than 2 days.
colored row means the latest one.
N.I.=not installed

Before submitting cordial jobs, we can check if Belle II software can be actually executed or not, from various viewpoints.
In particular, trouble relating cvmfs is often observed.

```
$ cvmfs_config stat -v belle.cern.ch
Thu Apr  2 11:11:10 UTC 2015
Version: 2.1.19.0
PID: 37692
Uptime: 14233 minutes
Memory Usage: 44620k
File Catalog Revision: 52 (expires in 2 minutes)
File Catalog ID: 19679a522a7641adc6d75b66ed3c656d022d99f3
No. Active File Catalogs: 5
Cache Usage: 12149781k / 20480001k
File Descriptor Usage: 6721 / 65024
No. Open Directories: 0
No. IO Errors: 5
Connection: http://cvmfs-stratum-one.cern.ch/opt/belle through proxy http://squid-frontier.ulakbim.gov.tr:3128 (online)
Usage: 41710585 open() calls (hitrate 99.9997%), 184130 opendir() calls
Transfer Statistics: 12684k read, avg. speed: 18k/s

$ cvmfs_readconfig belle.cern.ch
$ echo $CVMFS_CACHE_DIR
/var/cache/cvmfs2/shared
$ df /var/cache/cvmfs2/shared /cvmfs/belle.cern.ch/sl6
Filesystem                                  1K-blocks       Used  Available Use% Mounted on
/dev/sda3                                   117233580   69772736   41499040  63% /
tmpfs                                        66070488          0   66070488   0% /dev/shm
/dev/sda5                                   103078048   12256244   85578844  13% /var/cache/cvmfs2
torik1.ulakbim.gov.tr:/software/exp_soft   4326280208 1794269344 2484047096  42% /opt/exp_soft
cvmfs2                                       20480000   12149781    8330220  60% /cvmfs/belle.cern.ch
```

You can see the result for cvmfs checks!

CE test job submitter

- When DIRAC fails 70% of job submission to some site or SiteCrawler does not run for 6 hours, we submit a simple job to the site, outside DIRAC, i.e., using glite-ce-job-submit, by cron.
- The job executes just "cat /proc/cpuinfo".
- CE's reaction is recorded in DB.

CE Job Submission test result

sitename	CE	queue	status	last updated time
LCG.CNAF.it	ce08-lcg.cr.cnaf.infn.it	cream-lsf-belle	IDLE	2015/04/02 03:40:14 UTC
LCG.Cosenza.it	recas-ce-01.cs.infn.it	cream-pbs-belle	ce-job-status_failed	2015/04/02 03:40:16 UTC
LCG.DESY.de	grid-cr3.desy.de	cream-pbs-desy	DONE-OK	2015/04/02 02:00:14 UTC
LCG.Frascati.it	atlasce1.lnf.infn.it	cream-pbs-belle	IDLE	2015/04/02 03:40:19 UTC
LCG.HEPHY.at	creamce01.hephy.oeaw.ac.at	cream-pbs-belle	submission_failed	2015/04/02 03:25:02 UTC
LCG.KIT.de	cream-ge-3-kit.gridka.de	cream-sge-sl6	IDLE	2015/04/02 03:40:29 UTC
LCG.KMI.jp	ncream03.hepl.phys.nagoya-u.ac.jp	cream-pbs-belle	ce-job-status_failed	2015/04/02 03:40:30 UTC
LCG.MPPMU.de	grid-emicream2.rzg.mpg.de	cream-sge-cream2.q	submission_failed	2015/04/02 02:55:34 UTC
LCG.Napoli.it	t2-recas-ce01.na.infn.it	cream-pbs-belle		
LCG.Pisa.it	gridce1.pi.infn.it	cream-lsf-belle		
LCG.UA-ISMA.ua	gl-ce.isma.kharkov.ua	cream-pbs-belle		
LCG.ULAKBIM.tr	kalkan1.ulakbim.gov.tr	cream-pbs-belle		

CE Job Submission test result on LCG.Pisa.it

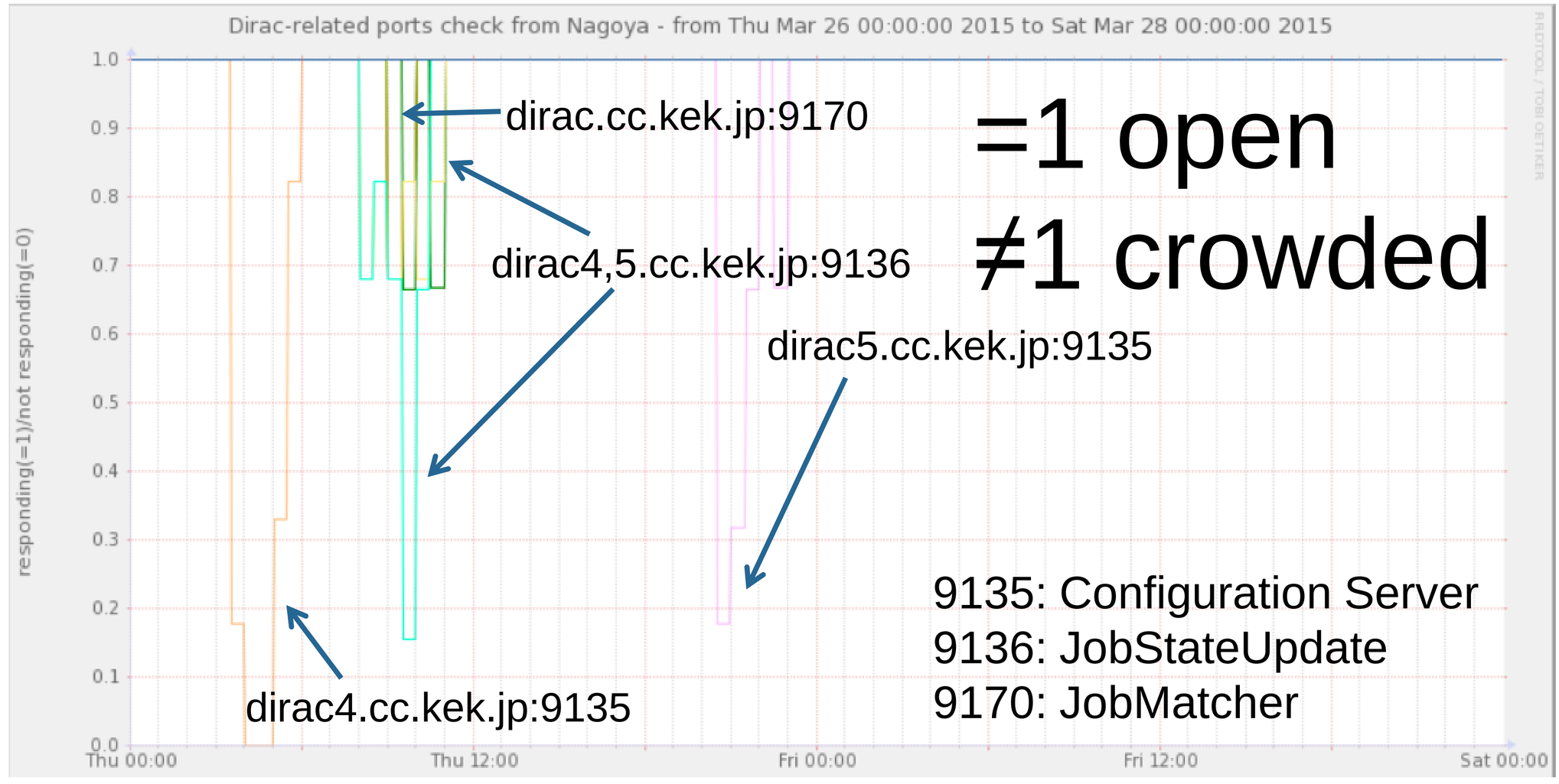
CE	queue	status	jobid	log	last updated time
gridce1.pi.infn.it	cream-lsf-belle	submission_failed	None	log	2015/04/02 02:55:45 UTC
gridce4.pi.infn.it	cream-lsf-belle	submission_failed	None	log	2015/04/02 02:55:43 UTC
gridce2.pi.infn.it	cream-lsf-belle	submission_failed	None	log	2015/04/02 02:55:42 UTC
gridce0.pi.infn.it	cream-lsf-belle	submission_failed	None	log	2015/04/02 02:55:40 UTC
gridce3.pi.infn.it	cream-lsf-belle	submission_failed	None	log	2015/04/02 02:55:39 UTC
gridce1.pi.infn.it	cream-lsf-belle	submission_failed	None	log	2015/04/02 01:55:53 UTC
gridce4.pi.infn.it	cream-lsf-belle	submission_failed	None	log	2015/04/02 01:55:50 UTC
gridce2.pi.infn.it	cream-lsf-belle	submission_failed	None	log	2015/04/02 01:55:48 UTC
gridce0.pi.infn.it	cream-lsf-belle	submission_failed	None	log	2015/04/02 01:55:46 UTC
gridce3.pi.infn.it	cream-lsf-belle	submission_failed	None	log	2015/04/02 01:55:44 UTC
gridce1.pi.infn.it	cream-lsf-belle	submission_failed	None	log	2015/04/02 00:55:48 UTC
gridce4.pi.infn.it	cream-lsf-belle	submission_failed	None	log	2015/04/02 00:55:46 UTC
gridce2.pi.infn.it	cream-lsf-belle	submission_failed	None	log	2015/04/02 00:55:43 UTC
gridce0.pi.infn.it	cream-lsf-belle	submission_failed	None	log	2015/04/02 00:55:41 UTC
gridce3.pi.infn.it	cream-lsf-belle	submission_failed	None	log	2015/04/02 00:55:39 UTC

You can get the failure reason by clicking "log".
In this case:
FATAL - Received NULL fault; the error is due to another cause: FaultString=[java.lang.NullPointerException] - FaultCode=[SOAP-ENV:Server] - FaultSubCode=[SOAP-ENV:Server]

When pilots/jobs do not run on some site, we can judge whether it is due to DIRAC or CE.

DIRAC port checker

- When DIRAC server experiences a heavy load, it sometimes gets down. But, before it, we can observe foretastes.
- One is port accessibility. DIRAC assigns ports from 9130 to 9200 function by function. When some port is crowded, it becomes difficult to open it.
- We periodically open the ports and check if the port is accessible or not. This can be job failure reason, too.



Even if the server gets down, we can reboot quickly since mail notification system to the operator is equipped.

Summary

- Belle II experiment requires the huge computing resources to process its huge data sample.
  - Belle II operates the distributed computing system using DIRAC.
  - But, having DIRAC is not an end of the story.
- Through the operation, many kinds of troubles have been experienced and they reduce the effective amount of the CPU resources.
  - To reduce the "dead time" of the CPU resources, it is important to find any problem faster and we need a good monitoring system.
  - Not only a standard monitoring system equipped in DIRAC but also our own system optimized to detect the trouble we faced are necessary.
- We have been developing the monitoring system for Belle II computing.
  - Here, we introduce the "active-way" monitorings.
  - They help to find problems on WN, CE and central DIRAC servers.

("passive-way" monitoring is discussed at Poster#337)

Future Plan

- More sophiticated/automated monitoring system
  - Combining various information collected by sub-monitoring systems and diagnose what happens automatically.
  - Provide "human-readable" operation summary:

LCG.Legnaro.it : OK
LCG.MPPMU.de : OK
LCG.McGill.ca
- from twiki: The site will be in a scheduled downtime on 2014-10-17 10:00:00 UTC. Pilot submission will be stopped by the operation team before the downtime. -- ShiftLog20141015night

LCG.Melbourne.au
- from twiki: **fixed** (ShiftLog20141014owl) -- 1012: CREAM CE looks problematic and doesn't accept new job. Issued redmine ticket (See redmine ticket 1338)

LCG.NCHC.tw
- Health checker info. : "Aborted pilot jobs" has been found since 03:20:00 UTC on 2015/04/01 (details)
- job submission check : Pilot submission failure has been found since 14:28:00 UTC on 2015/04/02. (details)

LCG.NTU.tw
- from twiki: **fixed** NTU SE seems down. (22-10-2014 10:00 UTC)

LCG.Napoli.it
- waiting pilot check : Long-waiting pilot has been found since 14:50:00 UTC on 2015/04/02. (details)

  - Automated notification system:
    - ✓To LCG site: submit GGUS tickets
    - ✓To other site: similar system (redmine?)
- Also monitoring system for SE
  - accessibility
  - read/write
  - checksum

Belle II monitoring system pursues the most efficient operation for Belle II computing by quickly finding the problems.

Kobayashi-Maskawa Institute for the Origin of Particles and the Universe

NAGOYA UNIVERSITY