

Meeting with the communities

Belle II case (extra topics)

Hideki Miyake (KEK)

May 29th , 2015
5th DIRAC user workshop@Ferrara

Outline



- ❧ **Matcher behavior under MatchingDelay**
- ❧ **OSG-CE, SSH-CE**
- ❧ **PFN definition in DIRAC-SE (dips)**
- ❧ **CS overload**

- ❧ **Reports from monitoring team**
 - ❧ SiteDirector and ComputingElement

Issues: WMS



❧ **MatchingDelay with Site=ANY**

- ❧ MatchingDelay affects both site specific job (Site=AAA) and generic job (Site=ANY)
 - ❧ Due to behavior of TQDB.__generateNotSQL()
- ❧ Since SiteDirector refers Matcher, both pilot submission and payload matching are affected by the behavior
- ❧ As a result, number of "Site=ANY" job is quite unstable (sometimes SD submits pilot once per 6 hours)
- ❧ Practically it can be avoided if site specific queue (e.g. Site="AAA") exists in TQ
 - ❧ Dirty workaround to fill TQ by DIRC Agent

❧ **SiteDirector affected by Matcher**

- ❧ Related with the issue above, SD refers Matcher for pilot submission
- ❧ It means MatchingDelay is applied for SD
- ❧ Suggest to skip MatchingDelay for pilot submission

Issues: WMS



❧ MatchingDelay with Site=ANY

❧ Site = ANY = AAA or BBB or CCC

❧ $!(AAA \text{ or } BBB \text{ or } CCC) = !AAA \text{ and } !BBB \text{ and } !CCC$

❧ Imply Site=ANY is affected by all of negative conditions

❧ Even if a pilot is submitted at Site=BBB, it is vetoed if other pilot was submitted to Site=AAA just before

❧ Same for payload matching

❧ As a result, job execution for Site=ANY is quite opportunistic

❧ Workaround: to fill TQ by Site=BBB job (at least pilot can be submitted)

❧ Payload is still affected (afak)


Matcher with Delayed Matching



Delayed Matching

- Useful to prevent SE overload
- However many pilots are wasted after the delay
- It increases both site and server loads
- Do you have any idea to reduce wasted pilots under delayed matching configuration?

PilotJobRefer...	Status	Site	ComputingEl...	Broker	CurrentJobID	OwnerGroup	LastUpdateTime [L
https://recasn...	Done	LCG.Napoli.it	recasna-ce01...	can66.cc.kek.jp	-	belle_pilot	2015-03-19 04:25
https://recasn...	Done	LCG.Napoli.it	recasna-ce01...	can66.cc.kek.jp	14892866	belle_pilot	2015-03-19 05:04
https://recasn...	Done	LCG.Napoli.it	recasna-ce01...	can66.cc.kek.jp	-	belle_pilot	2015-03-19 04:25
https://recasn...	Done	LCG.Napoli.it	recasna-ce01...	can66.cc.kek.jp	-	belle_pilot	2015-03-19 04:27
https://recasn...	Done	LCG.Napoli.it	recasna-ce01...	can66.cc.kek.jp	-	belle_pilot	2015-03-19 04:30
https://recasn...	Done	LCG.Napoli.it	recasna-ce01...	can66.cc.kek.jp	-	belle_pilot	2015-03-19 04:26
https://recasn...	Done	LCG.Napoli.it	recasna-ce01...	can66.cc.kek.jp	14900539	belle_pilot	2015-03-19 06:06
https://recasn...	Done	LCG.Napoli.it	recasna-ce01...	can66.cc.kek.jp	14896288	belle_pilot	2015-03-19 05:52
https://recasn...	Done	LCG.Napoli.it	recasna-ce01...	can66.cc.kek.jp	-	belle_pilot	2015-03-19 04:26
https://recasn...	Done	LCG.Napoli.it	recasna-ce01...	can66.cc.kek.jp	14888619	belle_pilot	2015-03-19 05:03
https://recasn...	Done	LCG.Napoli.it	recasna-ce01...	can66.cc.kek.jp	-	belle_pilot	2015-03-19 04:30
https://recasn...	Done	LCG.Napoli.it	recasna-ce01...	can66.cc.kek.jp	14891168	belle_pilot	2015-03-19 05:04
https://recasn...	Done	LCG.Napoli.it	recasna-ce01...	can66.cc.kek.jp	-	belle_pilot	2015-03-19 04:30
https://recasn...	Done	LCG.Napoli.it	recasna-ce01...	can66.cc.kek.jp	-	belle_pilot	2015-03-19 04:26
https://recasn...	Done	LCG.Napoli.it	recasna-ce01...	can66.cc.kek.jp	14895662	belle_pilot	2015-03-19 05:37
https://recasn...	Done	LCG.Napoli.it	recasna-ce01...	can66.cc.kek.jp	14897336	belle_pilot	2015-03-19 05:51
https://recasn...	Done	LCG.Napoli.it	recasna-ce01...	can66.cc.kek.jp	14891925	belle_pilot	2015-03-19 05:06
https://recasn...	Done	LCG.Napoli.it	recasna-ce01...	can66.cc.kek.jp	-	belle_pilot	2015-03-19 04:26
sshondor/b...	Done	DIRAC.UVic.ca	belles.heprc...	host-206-12-154-50.heprc.uvic.ca	-	belle_pilot	2015-03-19 04:24
sshondor/b...	Done	DIRAC.UVic.ca	belles.heprc...	host-206-12-154-50.heprc.uvic.ca	14891001	belle_pilot	2015-03-19 05:10
sshondor/b...	Done	DIRAC.UVic.ca	belles.heprc...	host-206-12-154-50.heprc.uvic.ca	-	belle_pilot	2015-03-19 04:25
sshondor/b...	Done	DIRAC.UVic.ca	belles.heprc...	host-206-12-154-50.heprc.uvic.ca	-	belle_pilot	2015-03-19 04:25
sshondor/b...	Done	DIRAC.UVic.ca	belles.heprc...	host-206-12-154-50.heprc.uvic.ca	-	belle_pilot	2015-03-19 04:25
sshondor/b...	Done	DIRAC.UVic.ca	belles.heprc...	host-206-12-154-50.heprc.uvic.ca	14888582	belle_pilot	2015-03-19 06:00
sshondor/b...	Done	DIRAC.UVic.ca	belles.heprc...	host-206-12-154-50.heprc.uvic.ca	-	belle_pilot	2015-03-19 04:25
sshondor/b...	Done	DIRAC.UVic.ca	belles.heprc...	host-206-12-154-50.heprc.uvic.ca	-	belle_pilot	2015-03-19 04:24
sshondor/b...	Done	DIRAC.UVic.ca	belles.heprc...	host-206-12-154-50.heprc.uvic.ca	14891459	belle_pilot	2015-03-19 04:29
sshondor/b...	Done	DIRAC.UVic.ca	belles.heprc...	host-206-12-154-50.heprc.uvic.ca	-	belle_pilot	2015-03-19 04:26
sshondor/b...	Done	DIRAC.UVic.ca	belles.heprc...	host-206-12-154-50.heprc.uvic.ca	-	belle_pilot	2015-03-19 04:25
sshondor/b...	Done	DIRAC.UVic.ca	belles.heprc...	host-206-12-154-50.heprc.uvic.ca	-	belle_pilot	2015-03-19 04:24
sshondor/b...	Done	DIRAC.UVic.ca	belles.heprc...	host-206-12-154-50.heprc.uvic.ca	-	belle_pilot	2015-03-19 04:26

 = Waste pilots

Issues: WMS



❧ TaskQueue and sitemask

- ❧ When a bunch of jobs (Site=ANY) is registered with TaskQueue, sites not in sitemask never matches later, even if sitemask is updated.
- ❧ Is it possible to update "Site" in TaskQueue?

❧ ARC-CE

- ❧ Currently ARCComputingElement bundled with vanilla DIRAC doesn't support EMI-3 version of ARC client (3.0.3)
 - ❧ Could you support such additional versions as well as current one?
- ❧ Also our ARC-CE site requires to append following line to xrsi:
(Runtimeenvironment="ENV/PROXY")
 - ❧ Is it possible to add such configuration through CS?
- ❧ WMSAdministrator and SiteDirector need to run on same machine due to joblist file. Can you store it on DB so that the deployment is portable?

Issues: WMS



❧ SiteDirector

- ❧ Sometimes attempt job submission even if the previous failure reason is trivial (e.g. slot is closed)
- ❧ How about to parse some obvious reasons?

❧ OSG-CE

- ❧ Previously its software stack was obsoleted and not supported
- ❧ Recently they are updated.
- ❧ Do you have a plan to support direct OSG-CE submission instead of gLite WMS?

❧ SSH-CE

- ❧ Each time scp transfers remote script to head node. It is not negligible for small sites (whose network is not good)
- ❧ Firewall could judge "attacked" → ssh multiplexing (suggested by Kiyoshi Hayasaka)
- ❧ Furthermore it is rare but several times file system was overloaded by such frequent file transfer.

Issues: DMS



File ownership

- When a job uploads a file to SE, the data transfer is performed by job owner's proxy
- In most case DMS operation utilizes DMS shifter proxy, right?
- The behavior difference sometimes bring file permission trouble unless site ACL is not carefully configured.
- What is DIRAC policy / suggestion to manage file ownership?

PFN definition of DIRAC-SE

- When we store one file on DIRAC-SE, FileCatalog gives LFN as PFN (i.e. no protocol)
- Is it collect behavior? What's best way to identify the file?

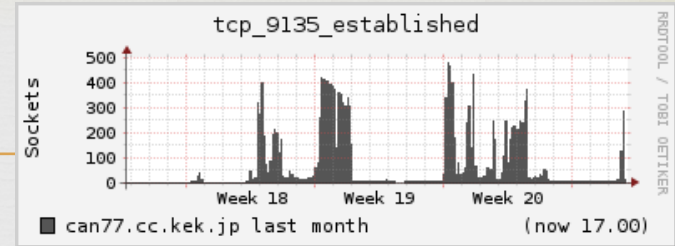
FileCatalog.getReplicas()

```
{'OK': True, 'Value': {'Successful': {'/belle/user/hideki/hogehoge/hoge1.root': {'KEK2-TMP-SE': 'srm://kek2-se01.cc.kek.jp/belle/TMP/belle/user/hideki/hogehoge/hoge1.root', 'KEK-FailoverSE': '/belle/user/hideki/hogehoge/hoge1.root'}}, 'Failed': {}}
```

```
% dirac-dms-lfn-accessURL /belle/user/hideki/hogehoge/hoge1.root KEK2-TMP-SE  
{'Failed': {}, 'Successful': {'/belle/user/hideki/hogehoge/hoge1.root': 'file:///ghi/fs01/belle/grid/storm/TMP/belle/user/hideki/hogehoge/hoge1.root'}}
```

```
% dirac-dms-lfn-accessURL /belle/user/hideki/hogehoge/hoge1.root KEK-FailoverSE  
{'Failed': {}, 'Successful': {'/belle/user/hideki/hogehoge/hoge1.root': 'dips://dirac1.cc.kek.jp:9149/DataManagement/StorageElement?=/belle/user/hideki/hogehoge/hoge1.root'}}
```


Issues: CS



☞ Recently we were suffered from heavy CS access

☞ Sometimes result in following errors

2015-05-20 05:03:55 UTC Configuration/Server ERROR: Error processing proposal Error while sending: (32, 'Broken pipe')

...

...

2015-05-20 05:04:02 UTC Configuration/Server ERROR: Error processing proposal Socket write timeout exceeded

☞ Have you seen such errors? Any suggestion to solve the issue?

☞ Tentatively increased CS instances on each server

☞ Maybe access to CS is too much. Is it possible to access CS through http (i.e. caching?)

Experience of Nagoya team from the operation of SSH sites, LCG site and SiteDirector.

SiteDirector.py

L508:

```
jobExecDir = self.queueDict[queue]
```

```
httpProxy = self.queueDict[queue]
```

should be

```
self.queueDict[queue]['ParametersDict']...
```

```
httpProxy = self.queueDict[queue]['ParametersDict']
```

Otherwise, execution directory / http proxy can not be changed.

L591:

```
'Status' should be TRANSIENT_PILOT_STATUS
```

Otherwise, huge number of jobs will be submitted to SSH site when no running jobs.

(also refer ComputingElement.py L251)

(See p7, p8)

SSHTorqueComputingElement.py

L40:

Timeout value is hardcoded. However, network is not good for some sites and timeout often happen. It should be Configured at the SiteDirector.

L102: Status 'C' should be running because in some sites, C is very long and DIRAC submit jobs.

L120,121

```
output = '%s/DIRACPilot.o%s' % ( self.batchOutput, jobStamp )  
error = '%s/DIRACPilot.e%s' % ( self.batchError, jobStamp )
```

should be

```
output = '%s/DIRACPilot.o%s' % ( self.batchOutput, jobStamp.split('.')[0] )  
error = '%s/DIRACPilot.e%s' % ( self.batchError, jobStamp.split('.')[0] )
```

Otherwise, failed to retrieve output/error files.

Resources/Computing/remote_scripts/torquece

L67:

-s R should be -s RC

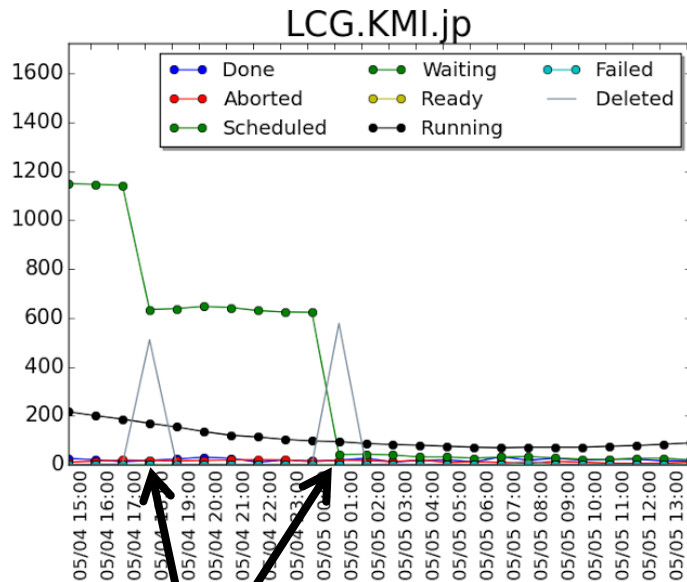
(the reason is the same as that in L102 of SSHTorqueComputingElement.py)

“Deleted” problem

When pilot becomes ‘Deleted’, pilot is removed from PilotAgentDB while the job exists in the local batch.

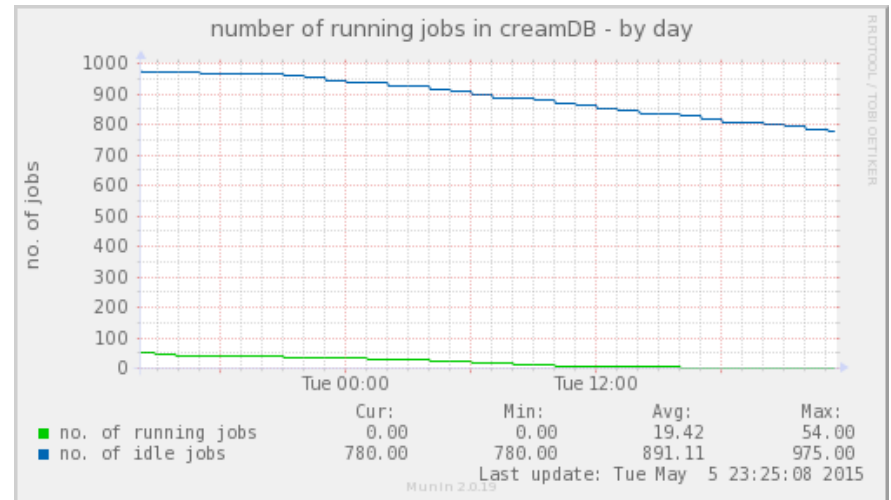
Then, DIRAC submit jobs further and local batch queue is occupied by waiting jobs. When status is changed to ‘Deleted’, job in local batch should be killed too.

of statuses taken from **PilotAgentDB**.



Jobs are deleted here and removed from PilotAgentDB.

of statuses taken from **local batch system**.



However, huge # of waiting job in local batch system.

Other things

- We want to “Delete” pilot jobs from web portal (of course, deleting from local batch, too)
- For DIRAC sites, failed to remove pilot outputs.

Supplements

SiteDirector.py ~ L591

```
def __getQueueSlots( self, queue ):  
    """ Get the number of available slots in the queue  
    """  
  
    ce = self.queueDict[queue]['CE']  
    ceName = self.queueDict[queue]['CEName']  
    queueName = self.queueDict[queue]['QueueName']  
  
    self.queueSlots.setdefault( queue, {} )  
    totalSlots = self.queueSlots[queue].get( 'AvailableSlots', 0 )  
    availableSlotsCount = self.queueSlots[queue].setdefault( 'AvailableSlotsCount', 0 )  
    if totalSlots == 0:  
        if availableSlotsCount % 10 == 0:  
  
            # Get the list of already existing pilots for this queue  
            jobIDList = None  
            result = pilotAgentsDB.selectPilots( {'DestinationSite':ceName,  
                                                'Queue':queueName,  
                                                'Status':['Running','Submitted','Scheduled']} )  
            if result['OK']:  
                jobIDList = result['Value']  
  
            result = ce.available( jobIDList )
```



Waiting is necessary here!

ComputingElement.py ~ L251

```
def available( self, jobIDList = None ):
```

```
    """This method returns the number of available slots in the target CE. The CE
       instance polls for waiting and running jobs and compares to the limits
       in the CE parameters.
```

```
       :param list jobIDList: list of already existing job IDs to be checked against
    """
```

```
# If there are no already registered jobs
```

```
if jobIDList is not None and len( jobIDList ) == 0:
```

```
    result = S_OK()
```

```
    result['RunningJobs'] = 0
```

```
    result['WaitingJobs'] = 0
```

```
    result['SubmittedJobs'] = 0
```

```
else:
```



When no running Jobs, and Waiting is ignored in SiteDirector.py L591,
DIRAC understand queue is empty and job is submitted..