

Job monitoring on DIRAC for Belle II distributed computing

Yuji Kato, Kiyoshi Hayasaka, Takanori Hara, Hideki Miyake, Ikuo Ueda
for Belle II computing group



Monitoring?

Huge amount of ...

- Resources

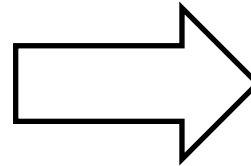
~100k cores
several 100PB

- Interfaces

authentication
network
batch..

- Users

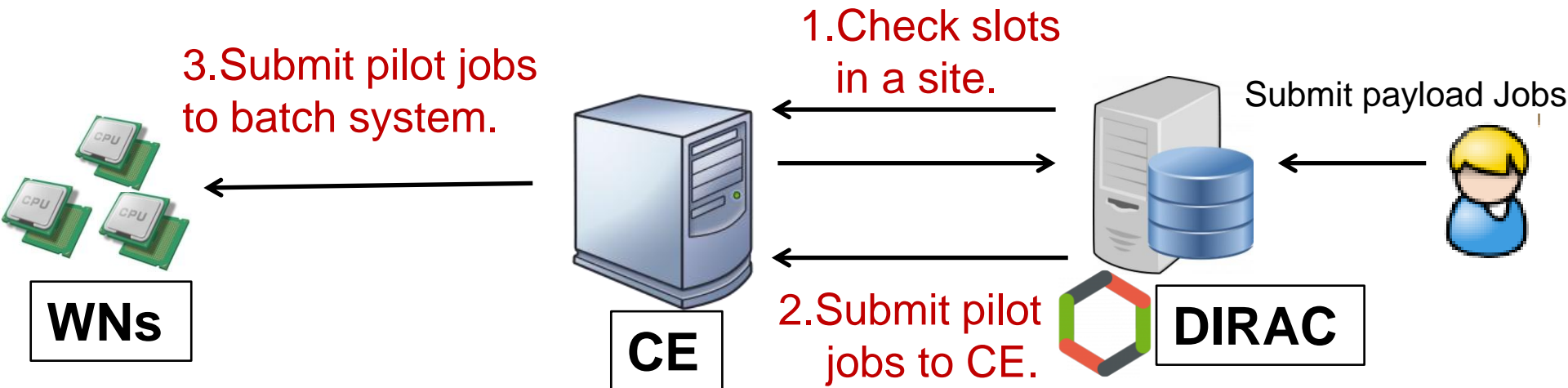
~600 members



Huge amount of
troubles!

**Developed a system which detect the trouble quickly.
Monitor team is composed of site maintainers.**

Workload management flow in the DIRAC



Pilot Job

4. Perform sanity checks
5. Execute payload Job

·
·

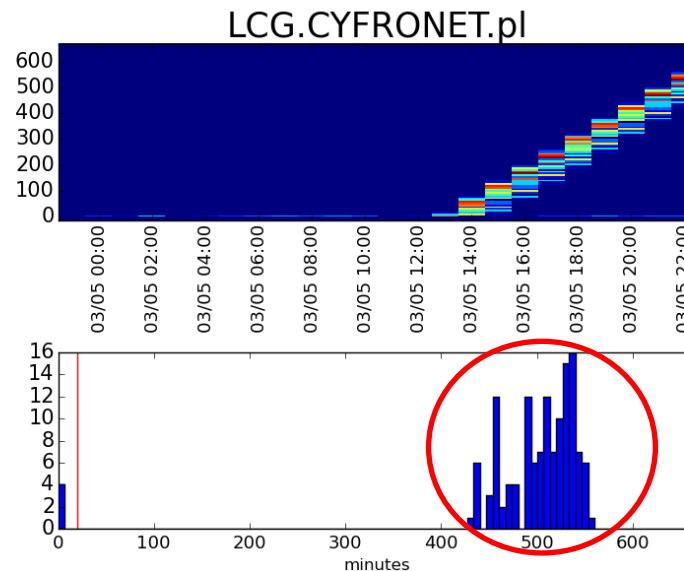
Many steps → Need to detect problems in each step

1. Check slots in a site

4

Sometimes, CE reports incorrect # of running pilot jobs due to the problem of CREAM etc.

→ **Characterized by “long silent pilots.”**



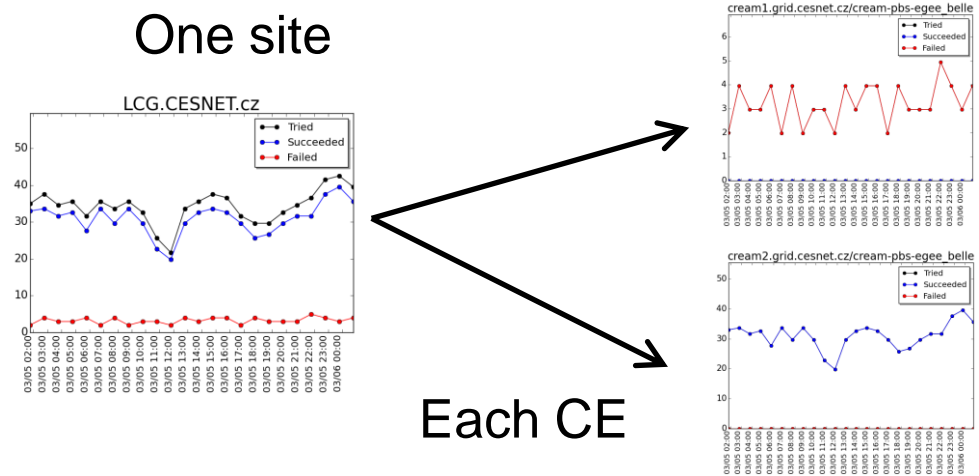
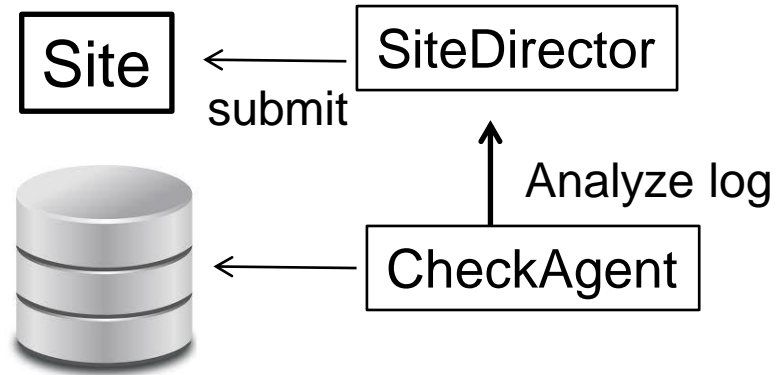
Pilot silent time distribution (in minutes).

Red line shows the possible maximum silent time for normal pilot jobs.

In this case, CREAM-CE recognizes finished job as running.

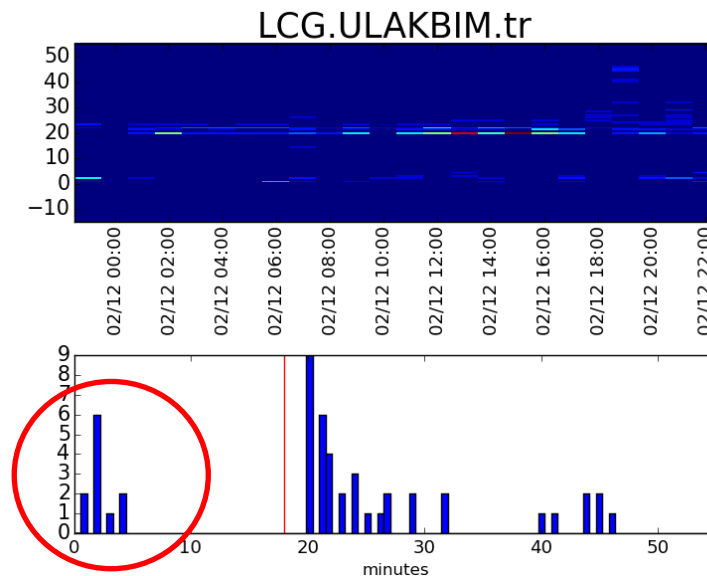
2. Submit pilot jobs to CE

Submission of pilot jobs to CE often fails because of CE down or problem proxy expire etc. Pilot jobs are sent by “SiteDirector”. DIRAC agent to monitor the activity of SiteDirector is developed and visualized.



4. Perform sanity checks

At the beginning of the pilot job, sanity check of WN is performed. If a problem is found, the pilot job stops immediately. Ex. CVMFS not properly mounted, disk full, failed to download DIRAC client etc..
→ problem on WN is characterized by **short pilots**.



Pilot life time distribution (in minutes).

Redline is possible minimum life time for normal pilot jobs 6

Acknowledgement

Just after the banquet...

I run in the convenience store to buy..

Ukon no tikara

Decompose alcohol quickly!



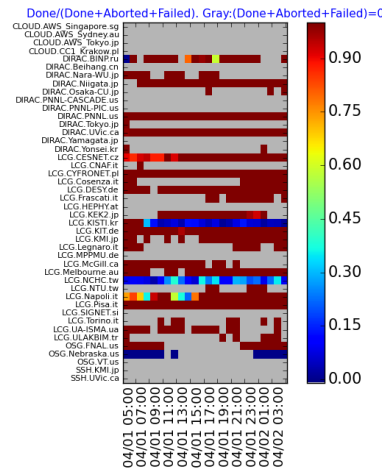
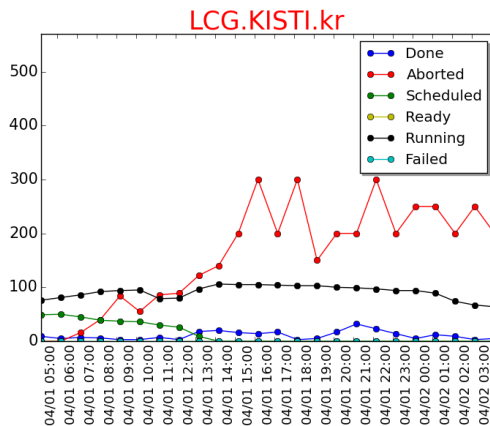
Yunker

Give energy!



3. Submit pilot jobs to batch system

Submission to batch server often fails because of **problem on the batch system**. If it is failed, status of pilot job becomes “Aborted”.



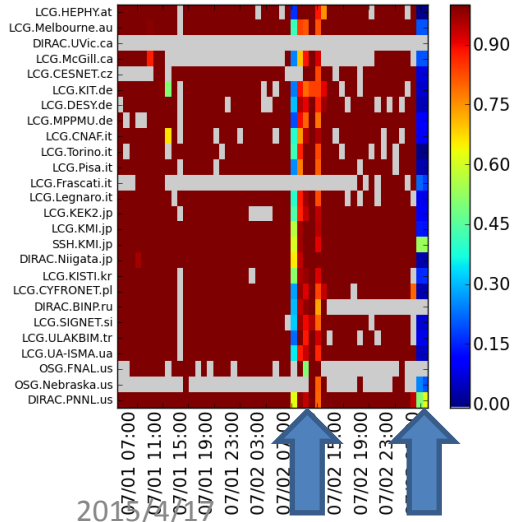
Example of error message:

[BLAH error: submission command failed (exit code = 1) (stdout: (stderr:qsub: Queue is not enabled MSG=queue is disabled.)

5. Execute Belle II Jobs

Payload jobs may fail with many reasons. **For example, failed to contact meta data server (AMGA), failed to handle input/output files, and problem on program itself.**

$(\text{Done})/(\text{Done}+\text{Failed})$ for each site. Gray means $(\text{Done}+\text{Failed})=0$



“Job efficiency” for each site.
Simultaneous failure for all the site means problem on central server.
In this case, AMGA was down.