



CMS Diverse use of clouds

David Colling



List of Authors

Anastasios Andronidis, Daniela Bauer,
Olivier Chaze, David Colling, Marc
Dobson, Maria Girone, Claudio Grandi,
Adam Huffman, Dirk Hufnagel, Farrukh
Aftab Khan, Andrew Lahiff, Alison
McCrae, Massimo Sgaravatto, Duncan
Rand, Anthony Tiradani, Xiaomei Zhang
+ support from many other people

Clouds playing an increasing role in
CMS activities



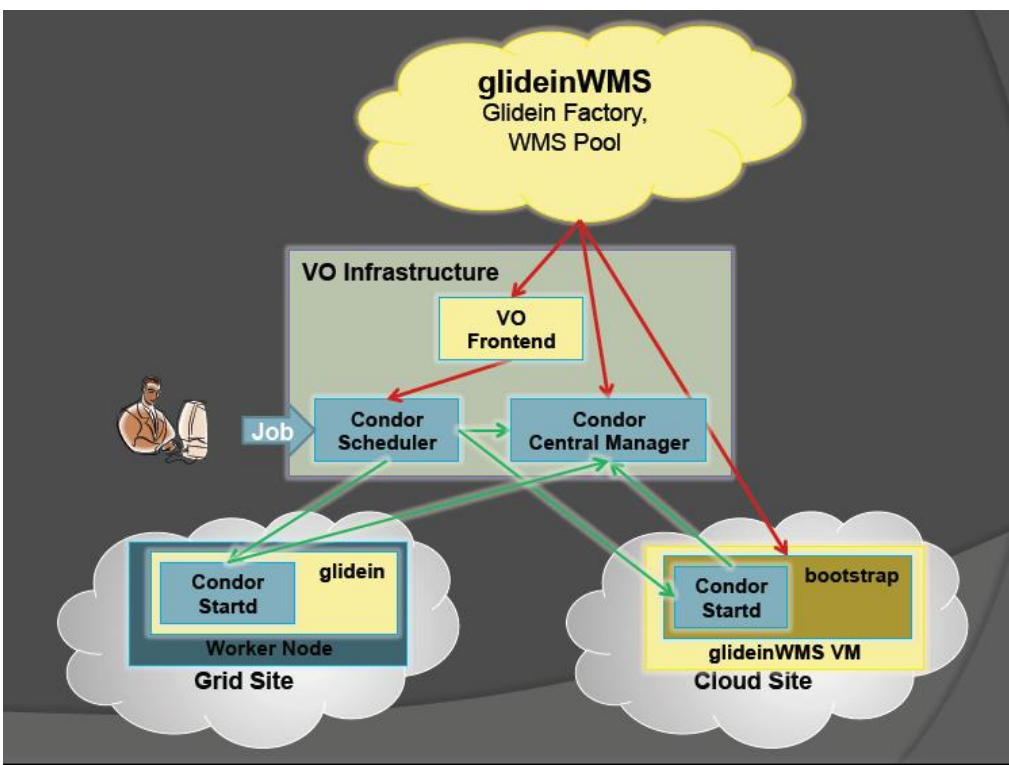
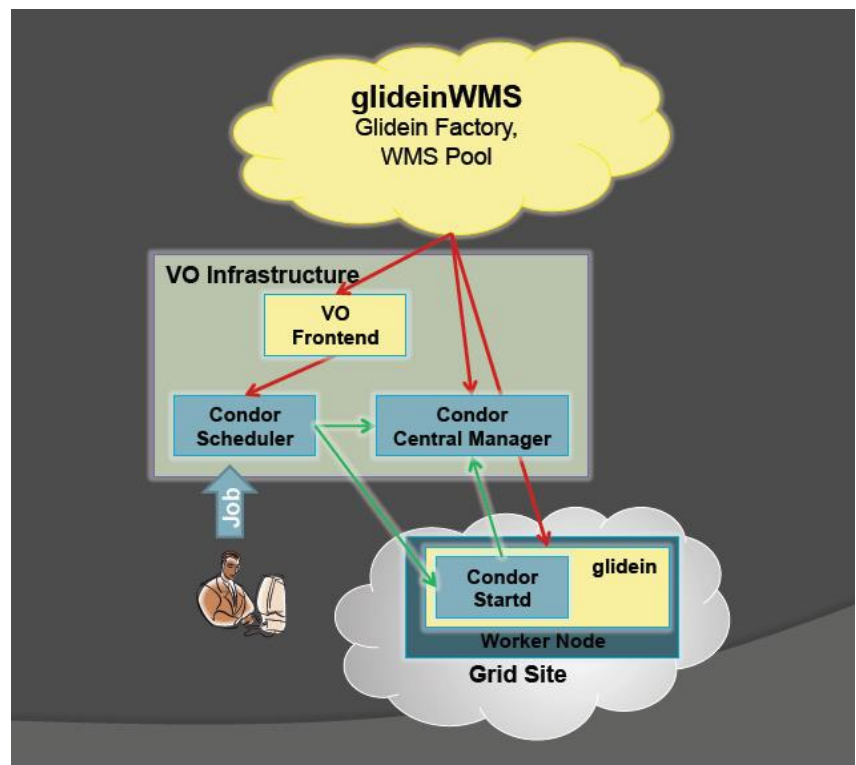
Content

- Default CMS interaction with Clouds
- T0 on the CERN Agile Infrastructure (AI)
- Using the CMS High Level Trigger (HLT) as a Cloud
- Using regional resources as clouds



How CMS uses Clouds

GlideinWMS used by CMS for job submission to Grids has been modified to also operate with EC2 Clouds. This has been in place for a number of years and works well. However, required modifications to allow for some of the “features” of early OpenStack releases.





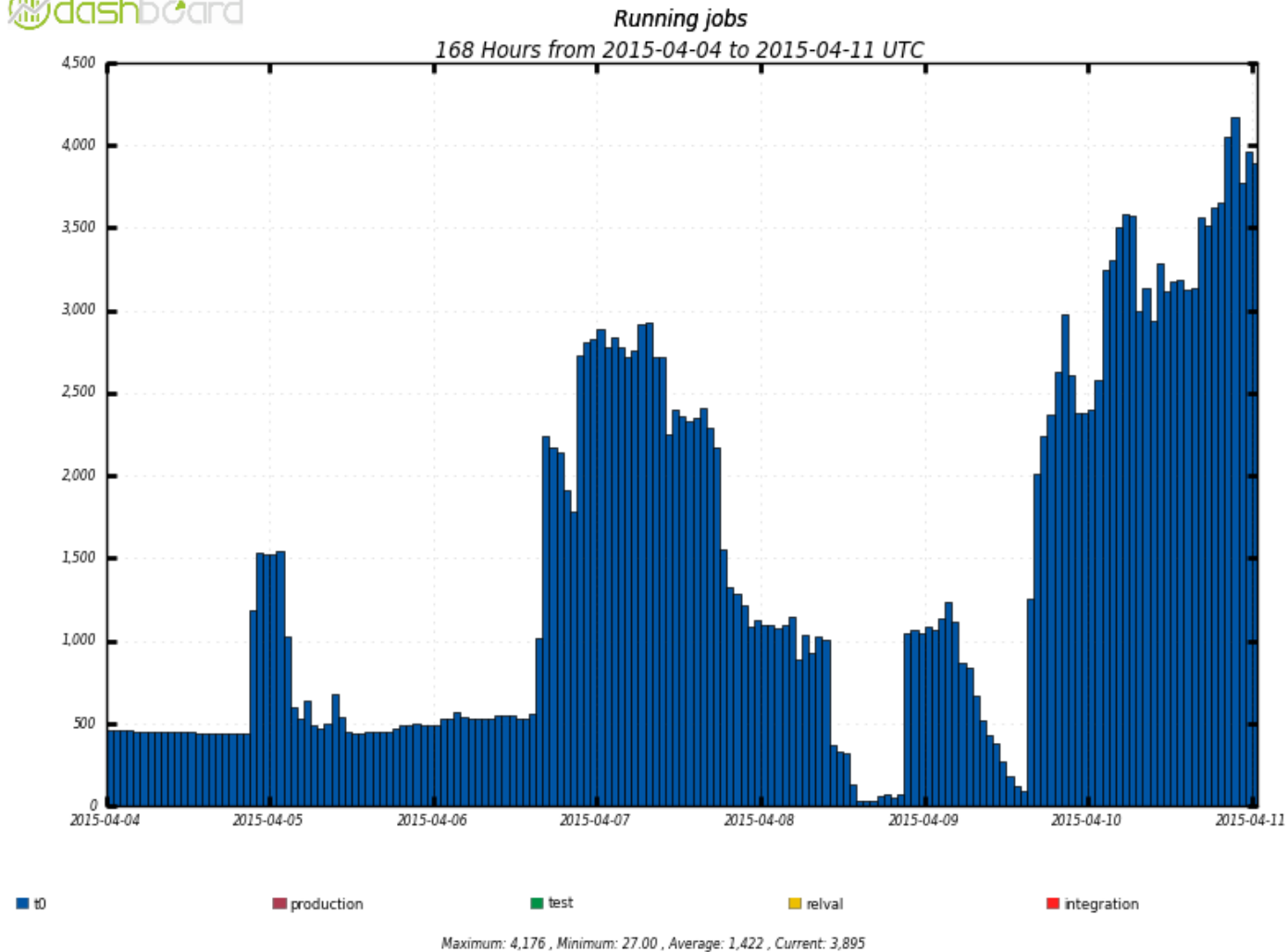
CMS T0 on AI¹

- T0 completely moved from LSF to AI (OpenStack)
- Approximately 9000 cores now (-> 15000= \sim 220 kHS06)
- 8 core/ 16GB VMs
- Images built by CMS, using CERN-IT automated build system.
- VM are not very dynamic but almost static (1 month duration) – not very agile
- Resources are part of the T0 Pool and shared with other uses. However, T0 runs with very high priority and so will push out other users quickly if needed.
- Currently scale testing with multicore jobs.

¹ See Dirk Hufnagel's talk for details



CMS T0 on AI





Running the HLT as Cloud

- The HLT is a considerable CPU resource. At startup:
 - Dell C6100 : 288 nodes, type = 12 cores 24GB ram (will be decommissioned end of 2015 and replaced)
 - Dell C6220 : 256 nodes, type = 16 cores 32GB ram (will be decommissioned end of 2016 and replaced)
 - Megware : 360 nodes, type = 24 cores 64GB ram (are the brand new machines)
- **Total 489kHS06** (cf. 300kHS06 for all of T1 request in 2015)
- No usable disk mass storage (only local disk to each machine)
- **When needed to be the HLT it must be the HLT and alternative use must not interfere** – However when not the HLT this is an extremely valuable resource. Cloud solution chosen (based on OpenStack) with VMs that can be killed completely if needed.

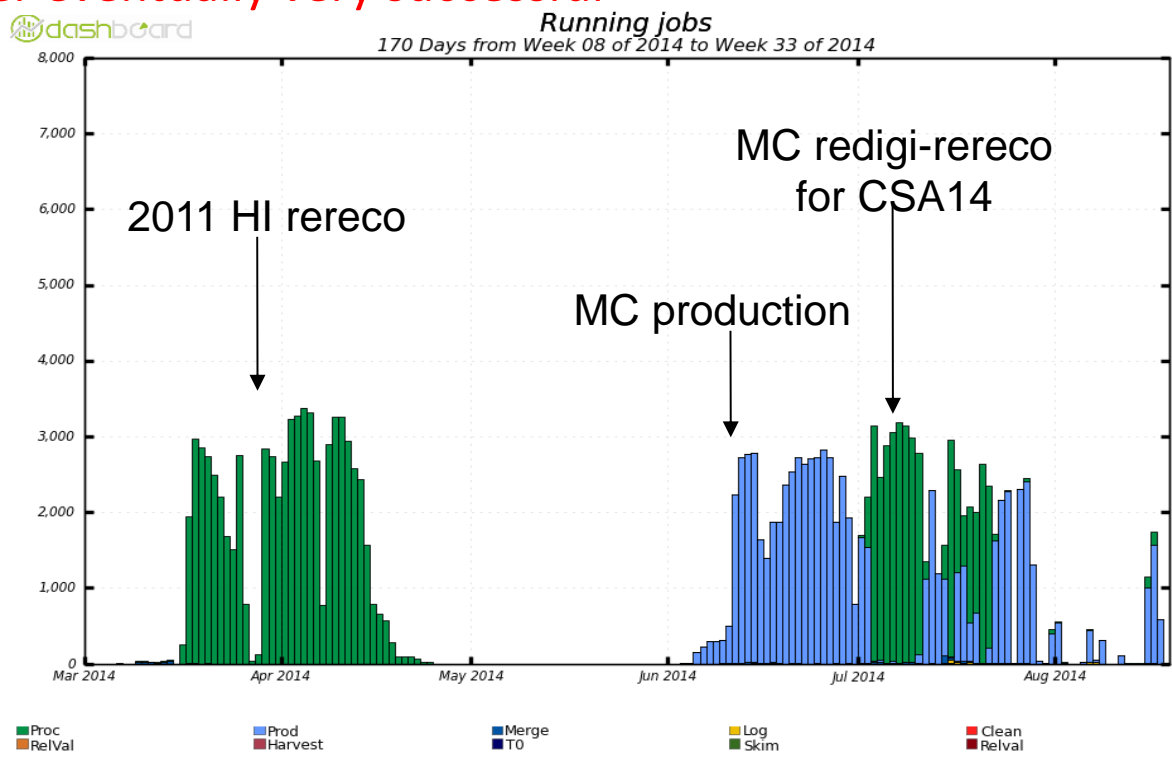


Running the HLT as Cloud

During LS1 operated HLT as a Cloud. Doing this we learnt a lot and overcame many minor but temporarily show stopping issues, including:

- Squid configuration
- “features” of early OpenStack
- ...

However eventually very successful





HLT as a Cloud in Run 2

We will clearly run as in the Technical Shutdowns. 1 week in ~10 in 2012.

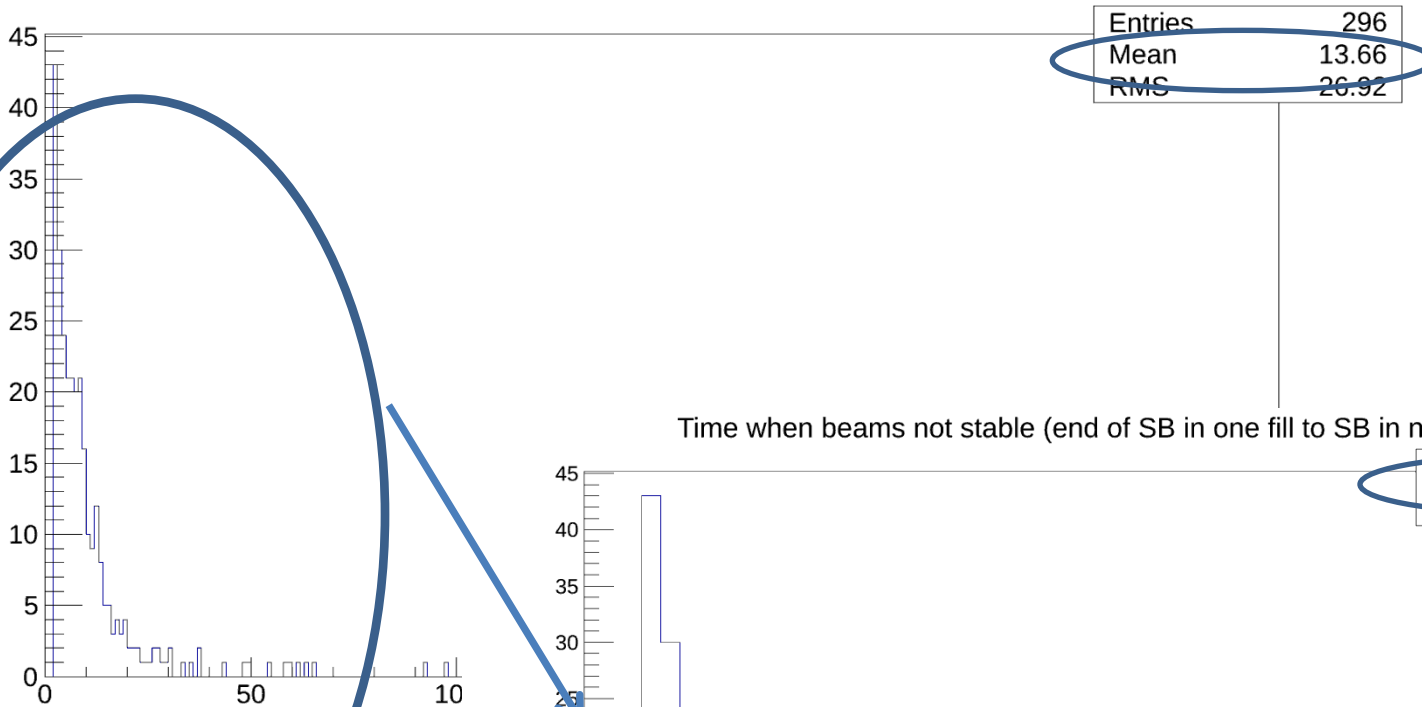
But when else?



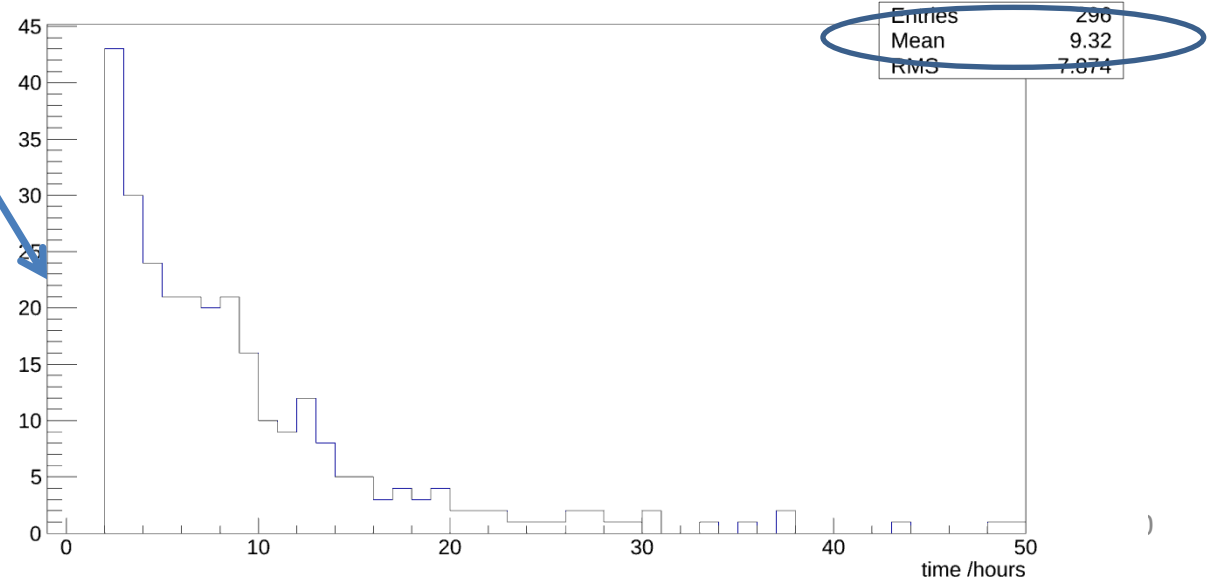
HLT as a Cloud in Run 2

- Between stable beams in 2012

Time when beams not stable (end of SB in one fill to SB in next)



Time when beams not stable (end of SB in one fill to SB in next)



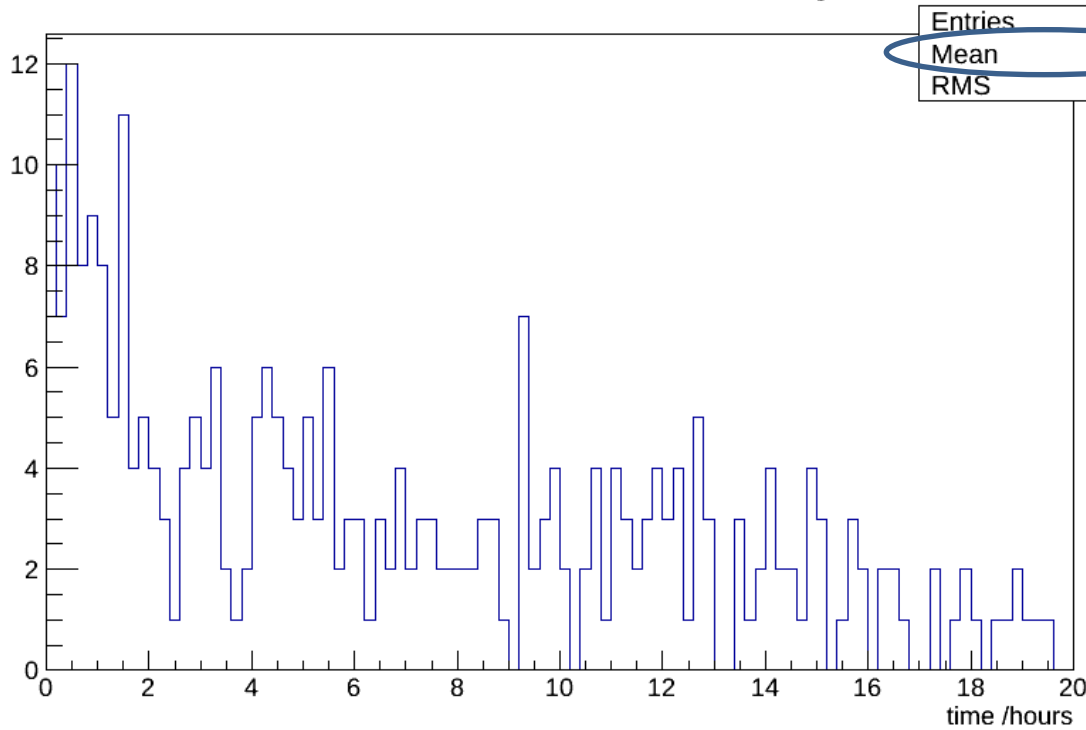
14/04/2015

)



HLT as a Cloud in Run 2

Duration of stable beam running

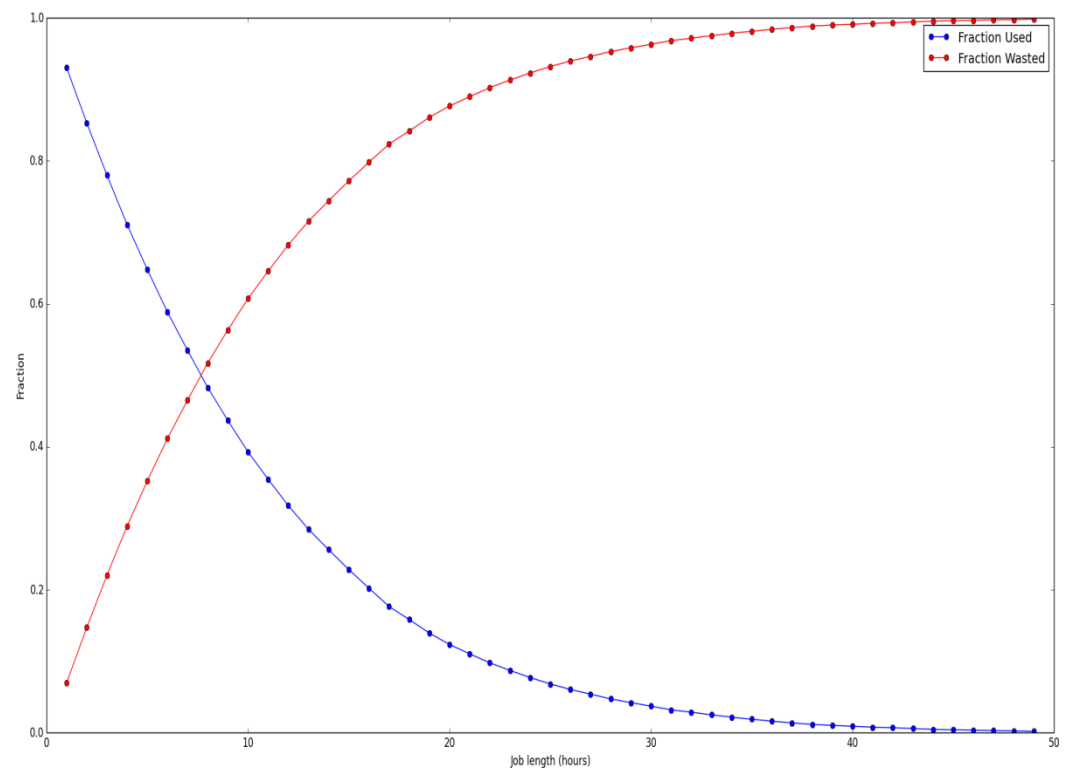


So even during stable running we have no stable more of the time that we have stable beam.

Even with expected overheads we can expect typically ~6 hours of usable time between runs. But this time is randomly distributed.



HLT as a Cloud in Run 2

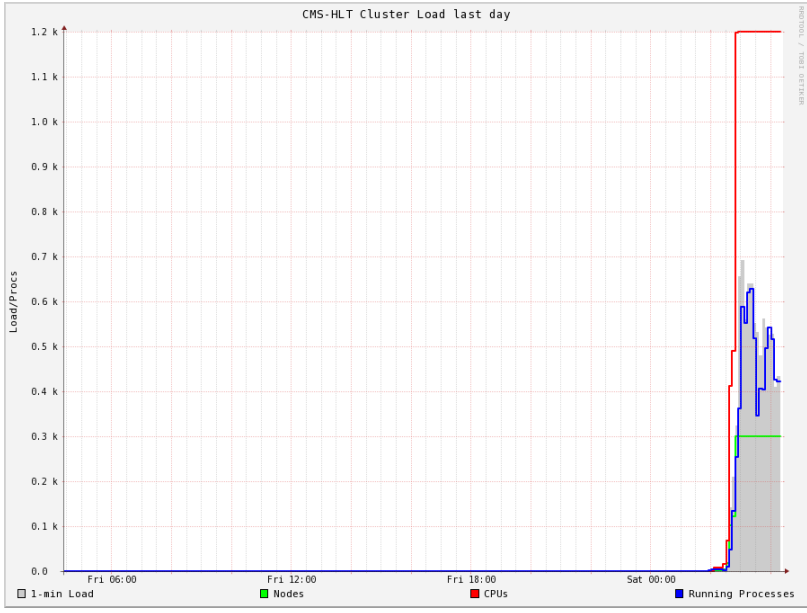


Assuming an exponentially decaying time between fills (but with nothing shorter than 1 hour) and with mean of 6 hours – i.e. roughly 2012 running. Assuming that all compute are jobs are killed when need to return to HLT. Clear and rapid trade off between complexity of many short jobs and inefficiencies caused by killing longer jobs.

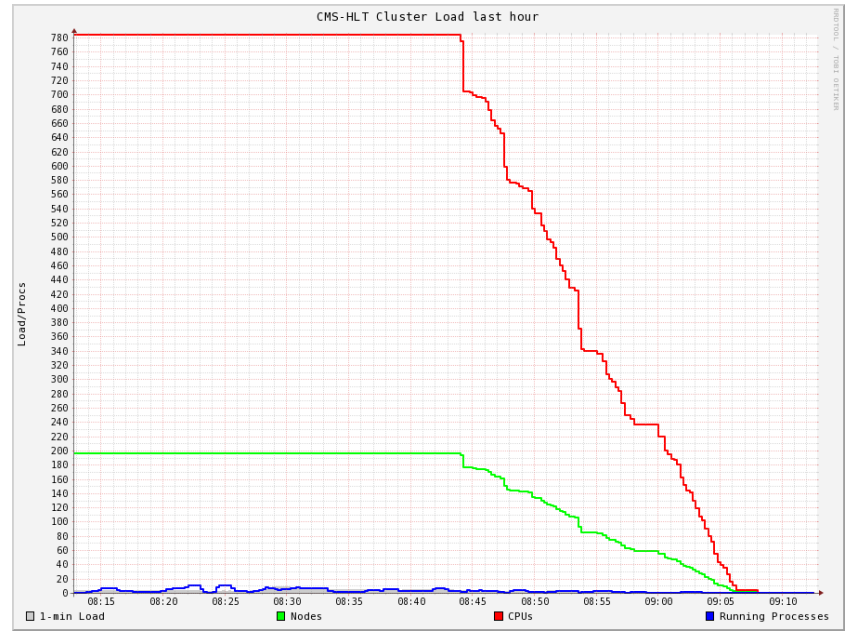
So anticipate running ~2hour jobs, but this will be adjusted as we gain experience.

HLT as a Cloud in Run 2

Also need to be able to start and stop quickly (two example tests)



Can start 1200 cores in
~10 minutes



Can shutdown 700 cores in ~20
minutes

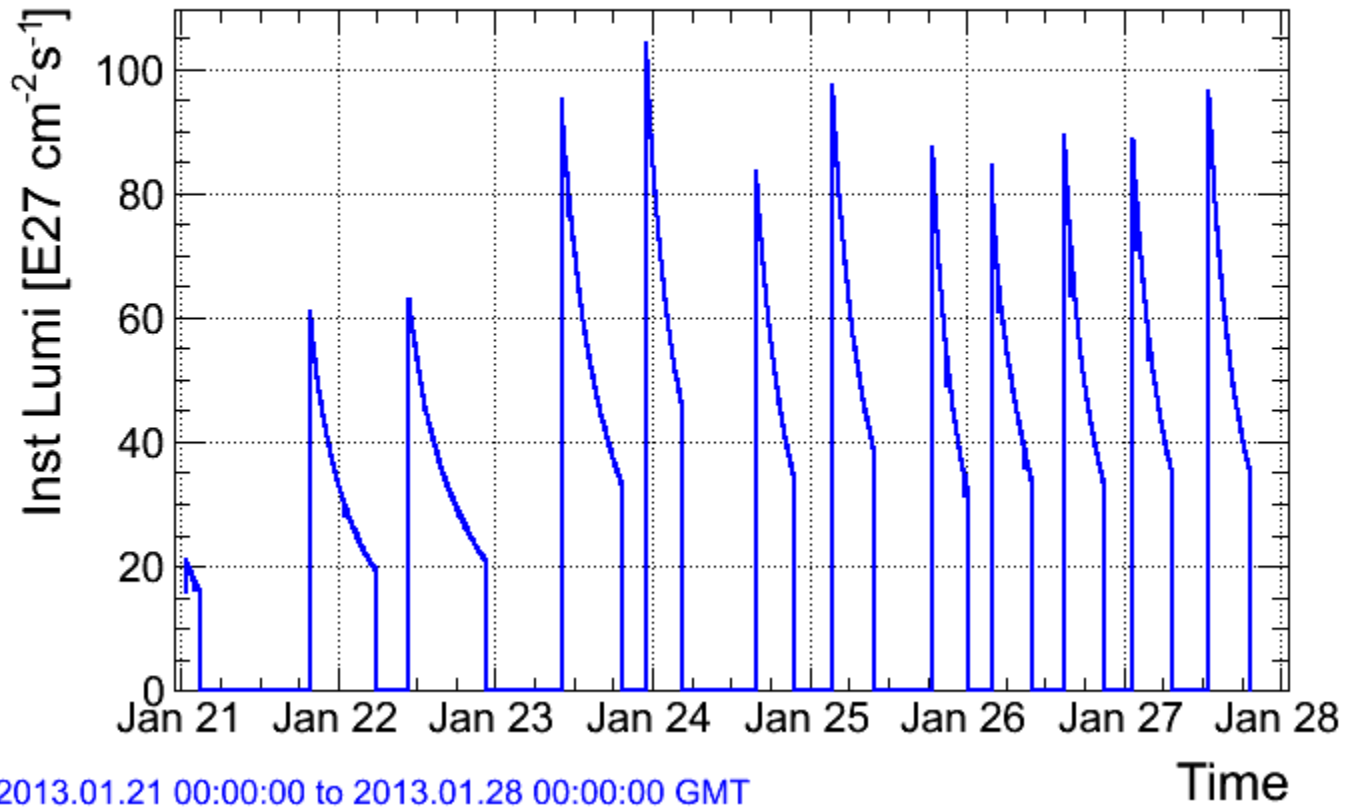


HLT as a Cloud in Run 2

Even during fills use of the HLT is not constant so should be possible to sometimes use part of the HLT even during data taking.

Then have VMs starting on HLT nodes.

Week03 : CMS HF Inst. Lumi. (online) [p-Pb] $\sqrt{s_{NN}} = 5.02$ TeV



2013.01.21 00:00:00 to 2013.01.28 00:00:00 GMT



HLT as a Cloud in Run 2

So we have a 3 level plan:

- Start by using HLT as a Cloud during Technical Stops. If that works ...
- Start to use the HLT as a Cloud between fills. If that works ...
- Start to use parts of the HLT as a Cloud during fills at low luminosity.

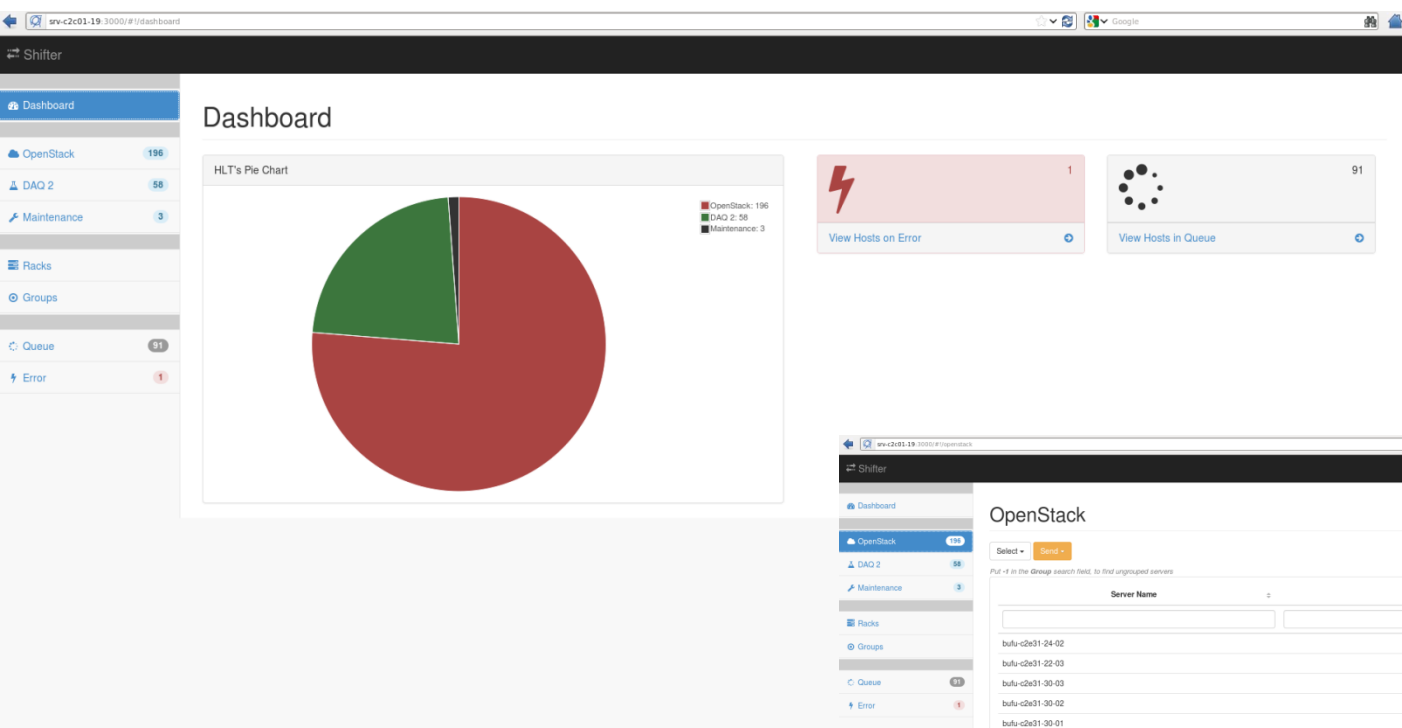
However, in this model the HLT team must be in control.

So a new tool, Cloud-igniter, used by the online team to start and kill VMs rather than requests from coming from the factory. New VMs connect to the Global pool.



HLT as a Cloud in Run 2

Current DAQ operator view in Cloud-ignitor



The OpenStack interface displays a table of servers with the following columns: Server Name, Rack, and Group.

Server Name	Rack	Group
bfu-c2b31-24-02		Openstack-reserved
bfu-c2b31-22-03		Openstack-reserved
bfu-c2b31-30-03		Openstack-reserved
bfu-c2b31-30-02		Openstack-reserved
bfu-c2b31-30-01		Openstack-reserved
bfu-c2b31-30-04		Openstack-reserved
bfu-c2b31-24-04		Openstack-reserved
bfu-c2b31-20-03		Openstack-reserved
bfu-c2b31-20-01		Openstack-reserved
bfu-c2b31-22-01		Openstack-reserved
bfu-c2b31-26-02		Openstack-reserved
bfu-c2b31-24-01		Openstack-reserved
bfu-c2b31-26-01		Openstack-reserved
bfu-c2b31-26-03		Openstack-reserved
bfu-c2b31-26-04		Openstack-reserved
bfu-c2b31-20-04		Openstack-reserved
bfu-c2b31-22-02		Openstack-reserved



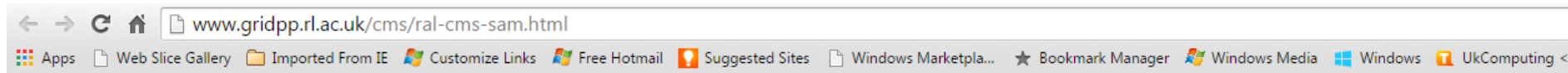
Institution Clouds

- While there has been activity in this area over the last two years it has been less of a focus than the CERN based activity. Have performed significant testing but never joined production infrastructure. However this is changing...
- Have been working with clouds in China, Italy and the UK.
- Have run user analysis using GlideinWMS installed in Italy and the UK.
- Have mainly tested Institutional Clouds using the “traditional” CMS cloud approach. However, have carried tests using VAC and Vcycle and using cernvm connecting directly to HTCondor pools.
- Started to look at “Cloud bursting”



Institution Clouds

- Developing procedure to move resources into production through testing.



RAL CMS SAM Tests

Site	Last job run
CLOUD_GridPP	-
CLOUD_RAL	-
HTCONDOR_RAL	2015-04-11 05:08:12
VAC_Manchester	2015-04-11 05:48:20
VAC_Oxford	2015-04-11 05:43:14
VCYCLE_GridPP	2015-04-11 05:07:23
VCYCLE_CERN	-
VCYCLE_EGIFedCloud	-

Colour coding (completed jobs over past 24 hours): <30% <60% <90% >=90%

Test jobs are the CE-cms-analysis SUM test, and 2 are submitted every 2 hours for each site. CLOUD_* rely on VMs being instantiated by glideinWMS; all others rely on self-generated VMs.

GlideinWMS monitoring: [factory frontend](#)



Conclusions

- CMS is using clouds resources in different ways at different in different situations
- So far the primary focus has been on Clouds at CERN with test being performed at institutional sites outside CERN.
- Developing procedure to bring institutional clouds into full production.