Future of Batch Processing at CERN

Jérôme Belleman, Ulrich Schwickerath, Iain Steers – IT-PES-PS

Outline

Context

Cluster and Setup

Running the Pilot

Local Jobs

Context

An LSF 7.0.6 Cluster

- $\approx 4\,000$ nodes
- Mostly virtual machines
- $> 65\,000$ cores
- $400\,000$ jobs/day
- $\pm 45\,000$ running jobs

An LSF ~~7.0.6~~ 9.1.3 Cluster

- $\approx 4\,000$ nodes
- Mostly virtual machines
- $> 65\,000$ cores
- $400\,000$ jobs/day
- $\pm 45\,000$ running jobs

An LSF ~~7.0.6 9.1.3~~ 7.0.6 Cluster

- $\approx 4\,000$ nodes
- Mostly virtual machines
- $> 65\,000$ cores
- $400\,000$ jobs/day
- $\pm 45\,000$ running jobs

Goals and Concerns

Goals	Concerns with LSF
30 000 to 50 000 nodes	6 500 nodes max
Cluster dynamism	Adding/Removing nodes requires reconfiguration
10 to 100 Hz dispatch rate	Transient dispatch problems
100 Hz query scaling	Slow query/submission response times

Evaluating Alternatives to LSF

HEPiX Fall 2013 – Ann Arbor:

- LSF 8/9 claims to only marginally scale higher
- SLURM showed scalability problems too
- Son of Grid Engine only briefly reviewed, as...
- ... HTCondor looked promising

Settling on Condor

HEPiX Spring 2014 – Annecy:

- Condor scaled encouragingly
- Focus on functions (grid, fairshare, auth, AFS)
- Pleasant experience

Pilot Service

HEPiX Fall 2014 – Lincoln:

- Grid submissions only
- Set up a CREAM CE
- Reviewed security

Running the Pilot Service

HEPiX Spring 2015 – Oxford:

- Set up ARC CE
- Configured fairshare groups and quotas
- Improved monitoring

$\rightarrow$ Running pilot and preparing for local submissions

Cluster and Setup

Nodes

- $\approx$ 200 worker nodes (8-slots/16-GB VMs)
- $\approx$ 1300 cores
- 4 schedulers (8-core/16-GB VMs)

Integration and Resources

- Automatic workflows (HEPiX Spring 2015)
- Node lists populated from PuppetDB queries
- Dynamic discovery and membership with Serf

HTCondor Configuration

- Currently running 8.3.8
- Single queue setup
- Fairshare: group is %-based, user is dynamic

Multicore

Multicore enabled, being tested by Experiments

Fallow, from Liverpool:

- Concerns at larger scale
- Offered patches

RAL's approach:

- Cron checking jobs
- Adjusting draining

Grid

- CEs in GOCDB
- Publication in BDII
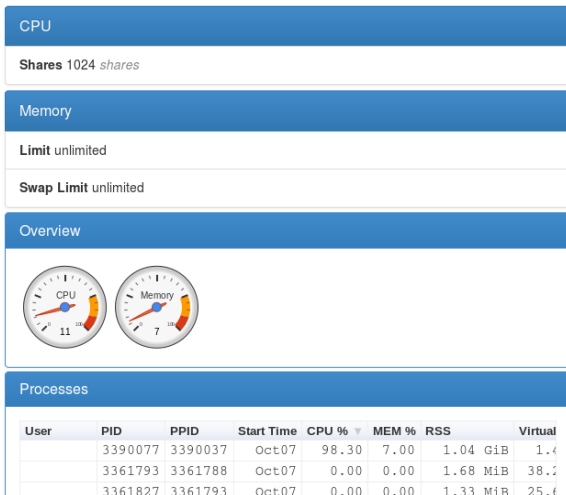- GLExec with LCMAPS plug-ins

HTCondor-CE

- Basically just a special HTCondor configuration
- HTCondor end to end
- Single area of expertise, same team for support

VO	Submission	Info
ALICE	Local `schedd`	BDII
ATLAS	Local `schedd`, PANDA	AGIS
CMS	Local `schedd`, global pool	w/o BDII
ILC	Local `schedd`	BDII
LHCb	Direct, Python API	BDII

cgroups

- Cgroups in all nodes
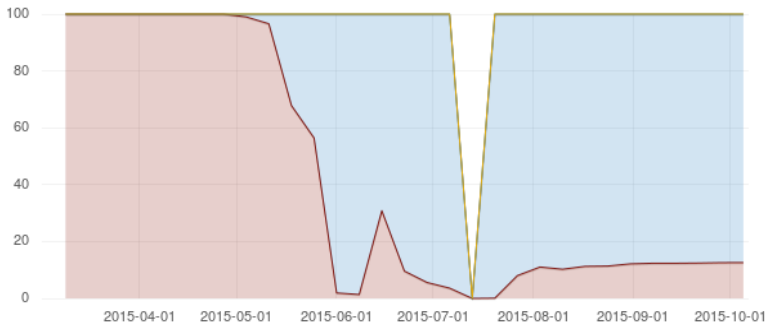- `/tmp` bind-mounts for easy pool clean-up
- Had a go at cAdvisor

CPU		
Shares 1024 *shares*		

Memory		
Limit unlimited		
Swap Limit unlimited		

Overview
CPU 11   Memory 7

Processes							
User	**PID**	**PPID**	**Start Time**	**CPU %**	**MEM %**	**RSS**	**Virtual**
	3390077	3390037	Oct07	98.30	7.00	1.04 GiB	1.(
	3361793	3361788	Oct07	0.00	0.00	1.68 MiB	38.2
	3361827	3361793	Oct07	0.00	0.00	1.33 MiB	25.(

Monitoring

Kibana/Elasticsearch

Health Checks

- Script running a `startd` cron job
- Checking various worker node exceptions
- Outputs `NODE_IS_HEALTHY` into the ClassAd
- Used by the `START` expression
- Thank you, RAL.

Accounting

- PostgreSQL, Elasticsearch, HDFS
- Publishing to APEL development server
- Python API used here too

Normalisation and Benchmarking

- HEPSPEC06 benchmarks of our worker nodes
- Looked into coordinating it with ZooKeeper
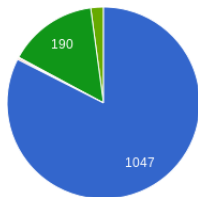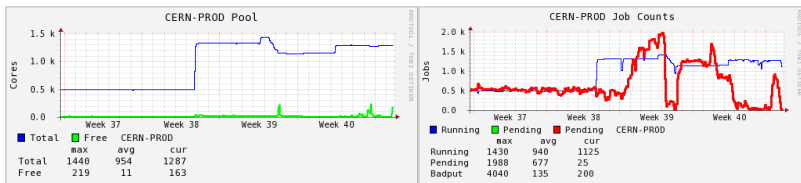- Result copied with node attributes to ClassAds

Running the Pilot

Invited Experiments to Try

- ALICE ALIEN pilots
- ATLAS HammerCloud and production jobs
- CMS tests and production
- LHCb working on DIRAC integration

Even Experiments who weren't invited, e.g. ILC

Activity

Jobs	Run	Pend	Held
482	384	95	0
453	380	73	0
369	335	34	0
219	170	30	19

Glitches

- ATLAS/ALICE jobs killed, bug in jobs? Fixed.
- Memory leak in `condor_shadow`, fixed in 8.3.6
- OpenLDAP/ARC incompatibility

Local Jobs

AFS Token Management

- There is Kerberos ticket passing
- Forging valid AFS tokens from expired ones
- Risk of credential theft
- Independence from AFS

Rewriting current system, to be based on Kerberos 5

Job Submissions and Queries

Query job no matter where it's submitted from

- `<username>.condor.cern.ch` aliases
- Consistent hashing to avoid complete remaps

Group Membership Enforcement

- Submit on behalf of the group you belong to
- Post-submission checks?
- There might be plans upstream
- Python ClassAd functions and LRU cache?

Conclusion

Collaboration

- Good partnership with Experiments
- Enthusiastic chats with lead developers
- Work on-site with Brian
- More help from RAL
- Working with PIC and DESY
- Many ideas from HTCondor Week

Outlook

- Grid service on the way to production
- HTCondor-CE a reality
- Clearer idea for local jobs

www.cern.ch