Future of Batch Processing at CERN

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

Outline

Context

For Now: Pilot Service

Next Up: Local Jobs

Context

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

After 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

After HEPiX Spring 2014 – Annecy:

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

Pilot Service

After HEPiX Fall 2014 – Lincoln:

- Grid submissions only
- Setting up a CREAM CE
- Reviewing security

$\rightarrow$ Consolidating pilot service

For Now: Pilot Service

Setting Up an ARC CE (I)

- CREAM heavy, opaque
- Heard good things about ARC
- Simple config, single file

Setting Up an ARC CE (II)

Now we have:

- Condor setup accepting and running jobs
- User-to-VO/role mapping
- Static/dynamic information published to BDII
- HEPSPEC06 normalisation
- Puppetised configuration

Setting Up an ARC CE (III)

TODO:

- GLUE validation fails with ARC, waiting for fix
- Scale job wall time by worker node attributes
- Single queue to accept jobs
- Security review

And then evaluate HTCondor-CE?

On the Condor Front (I)

- Fairshare groups and quotas
- Accounting group injected into job submission

TODO:

- Accounting (How do *you* store it?)

Monitoring

- Still our Ganglia instance, but also...
- ...central manager, `schedds`, workers in Kibana

Next Up: Local Jobs

AFS Token Management

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

Job Submissions and Queries

Query job no matter where it's submitted from

- A `schedd` to answer all queries?
- Protection against heavy query loads

Or

- `<username>.condor.cern.ch` aliases
- Job IDs hashed to `schedds`

Group Membership Enforcement

- Submit on behalf of the group you belong to
- Post-submission checks?
- There might be plans upstream

Replacement for LSF Queues

- Interface between users and resources
- Opportunity to review what users should see
- ClassAds

Worker Node Admission

- Explicit machine list for now...
- ... with the aim of becoming more dynamic
- Adding nodes easy, removing them not so much

High Availability/Scalability

- How many schedulers?
- Multiple pools?
- Hierarchical collectors?

Conclusion

Collaboration

- European HTCondor Site Admins Meeting 2014
- Enthusiastic chats with lead developers
- HTCondor Week
- Help from RAL
- Sharing with PIC too

Outlook

- Became interested in other CE solutions
- So did experiments
- Some progress in implementing fairshare
- Grid submissions for early adopters

www.cern.ch