

Experiences with HTCondor-CE

How did it get here, where is it going and what's it all about?

Ste Jones, Liverpool
GridPP42
April 2019

Topics

- I'll talk about what I've learned and try to answer some of these questions.
 - Why is it interesting?
 - What does it do?
 - Where does it come from?
 - How does it work?
 - What batch systems does it use?
 - What parts were missing?
 - Was it easy to install?
 - Is it any good?
 - Does anyone else use it?
 - Should I use it?
 - What else am I doing?
 - Where is it all written up?

Why the interest?

- Because it's there.
- An alternative to ARC CE.
- A replacement for CREAM CE.
- The track record of HT-Condor.
- Supported long-term by a committed team of experts.
- Used at CERN, I'm told.
- Major experiments (inc. ATLAS, LHCb) submit to it.

What does it do? Where does it come from?

- It's open source. It's well regarded. And it's part of OSG. There is some documentation here:

<https://bbockelm.github.io/docs/compute-element/htcondor-ce-overview/>

- It comes from University of Wisconsin, Madison ... the same place as HTCondor. There's this website about it all:

<https://research.cs.wisc.edu/htcondor/>

- And here's a recent picture of the team... they're a good bunch, very helpful and friendly.

What does it do? Where does it come from?



How does it work? What Batch Systems does it use?

- It's a special configuration of HTCondor (batch system.)
- All the usual condor_* commands have matching condor_ce_* commands. E.g. submission is via condor_ce_submit (or via condor_schedd on vbox.)
- It has some of the usual CE features like job reception, authn/authz, delegation to the batch system.
- It can talk to several batch systems; HTCondor, PBS, LSF, SGE, SLURM)

What was missing?

- APEL. But new work on the accounting is now tested/functional/accurate.
- We're still doing some development with APEL (Adrian Coveney) and HTCondor-CE (Brian Lin, Brian Bockelmann) to “smooth it out”.
- For APEL, we're removing a pointless BDII query.
- And for HTCondor-CE, we're going to RPMise the interface code, possibly puppetise the install and basically make it slick(er) to use.

What else is missing?

- No GLUE1 bdii, but the VOs don't seem to notice. GLUE2 only.
- Apart from that, it's all there. You can use it like it is ... it's fine. Plenty of jobs come.
- I made this PDF about it.

hep.ph.liv.ac.uk/~sjones/Gridpp42/HTCONDOR_CE_APEL.pdf

Was it easy to install?

- To keep things simple, we used HTCondor as the batch system. We were already using that with our ARC system, so installation of the batch system was a no-brainer.
- We used the same setup (e.g. accounting groups, fair shares, making users, scaling, WN setup, etc. as described in the ARC/Condor example build) but stripped away all the ARC bits.

https://www.gridpp.ac.uk/wiki/Example_Build_of_an_ARC/Condor_Cluster

Was it easy to install?

- So then we could focus only on the CE installation. Two questions:
- Was the CE really easy to install?
 - Nope. I had to “try things out” a lot to find a way to make it install cleanly.
- But is it easier to install now?
 - Yes. I’ve written up all the steps. You’re welcome to try it, or find your own way if you want (but please write it up if you do.)

Steps

- My steps are here (please edit it if you find errors or better ways to do things.)

https://www.gridpp.ac.uk/wiki/Example_Build_of_an_HTCondor-CE_Cluster

- I give two approaches; manual and with Puppet.
- The manual install was obtained by doing a Puppet install, and identifying all the packages and files that we added or changed.
- I extracted the list of packages (and archives of files).

More steps

- Then I applied those to a Plain Vanilla Head Node.
- And the approach worked – it functioned as a HTCondor-CE.
- All the material for the manual install was extracted and put in a git repo, e.g.
`git clone https://github.com/gridpp/htcce.git`
- Follow the README.md files that lie within.

More steps

- The Puppet section of the example build gives some detail on:
 - Obtaining and installing the htcondor_ce puppet module to install htcondor_ce.
 - Some minor edits that were needed to adapt the module.
 - The Yum Repositories required.
 - The Hiera parameters needed by the module.
 - A list of things to do once the module has finished: a bit of extra config, improvements to the figures in the BDII, the installation of host certificates.
 - The workings of the ARGUS Integration (external callout.)
 - The APEL integration. The BDII Installation (only GLUE2 at Liv.)
 - A small set of tests for the new system.
 - GOCDB Registration.

What's it like now?

- So, I think it might be better now.
- We need some foolhardy chump/eager volunteer to verify the work and see what comes out.
- If you try it out and find new things, please include them in the wiki document.

Is it any good?

- Apart from a “fat fingers” moment last week, when I added the new worker nodes to both our pools at the same time, we’ve had no issues at all in the ~ three months we’ve been using it.
- Not many VOs have used it; just LHCb and ATLAS (mcore, score).
- Slot utilisation is excellent. It doesn’t seem to need any explicit defragging to run MCORE. No Drainboss/Fallow etc. That’s a big mystery.

Is it efficient?

- It runs in a 3 cpu VM, with 3 GB RAM.
- We've got about 1,800 slots on it.
- It's presently running ~ 1000 tasks.
- Load average is low, 0.05, 0.11, 0.13
- Only half the RAM is used.
- It's been up for 72 days.
- ~ 170,000 jobs; 1500 failures that I know about (see above.)
- I think it's OK.

Does anyone else use it?

- I know that OSG sites use it (don't know which).
- CERN reputedly run much of their work via HTCondor_CE (see talks by Iain Steers.)
- The Tier 1 at PIC uses it.
- I've had HTCondor-CE APEL questions from IN2P3, so I know they use it.
-

Does anyone else use it?

- And I trawled the BDII to find this list of end points:
 - `tweetybird04.cern.ch`
 - `ce13.pic.es`, `ce14.pic.es`, `ifaece04.pic.es`
 - `cms-t2-ce01.sdfarm.kr`
 - `condorce1.ciemat.es`, `condorce2.ciemat.es`
 - `osgce1.farm.particle.cz`, `osgce2.farm.particle.cz`
- And that's all I could find.

Should I use it?

- Your choice but CREAM is going sour.
- ARC would be one choice. Works OK.
- But I like a bit of competition.
- So there's a bit of a contest here. We run both. It would not be right to have only one CE in the world.
- So let's see what happens.
- My money is on ... not saying!

What else are you doing?

- Since I started this, it's never ending...
- Fix a small BDII slot count issue (patch given in Example Build.)
- For APEL, fix a problem to circumvent a pointless BDII query.
- RPMs for the interface scripts.
- Get another site set up so we can test it all (esp. APEL) independently.
- Create Puppet module to install APEL client for HTCondor-CE.

What else are you doing?

- Figure out a split head-node version, with batch on one system, CE on another.
- Try out another batch system ... SLURM?
- Figure out how to route jobs to (e.g.) “smaller sites” that don’t have a head-node.
- Build a lightweight APEL interface with pure Python that dispenses with the need for an SQL client database (it’s a heavy hit just to send some accounting down the wire, and it’s a simple problem; see VAC for similar solution.)

Where is it written up?

Brian Bockelmann's documents:

- <https://bbockelm.github.io/docs/compute-element/htcondor-ce-overview/>

GridPP “steps”:

- https://www.gridpp.ac.uk/wiki/Example_Build_of_an_HTCondor-CE_Cluster

HTCondor-CE APEL Accounting:

- <https://twiki.cern.ch/twiki/bin/view/LCG/HtCondorCeAccounting>
- <https://twiki.cern.ch/twiki/bin/view/LCG/HtCondorCeAccountingDesign>

CERN talks, Iain Steers, HTCondor workshop:

- <https://indico.cern.ch/event/467075/contributions/1143797/>
- <https://indico.cern.ch/event/467075/contributions/1143807/>

Summary

- I spoke about the HTCondor-CE product.
- Where it comes from. What it does. How it works. Who created it.
- How we installed it. Bits we had to do ourselves.
- Whether it's any good. Who else uses it.
- Where the documents are.
- I hope that's enough to get you started if you want to adopt it. So good luck.

Questions? Ideas? Suggestions?

Criticisms? Insults?

....