

Questions and Comments by Attendees

Monday 21 September

Before Break

Welcome, logistics (Helge Meinhard)

- (Name, affiliation): (Comment or question)

State of distributed high-throughput computing (Miron Livny)

- (Name, affiliation): (Comment or question)

A users' introduction to HTCondor and job submission (Christina Koch)

- (Name, affiliation): (Comment or question)
- (Matthew West) Question: Is there a tool to parse the job event log to get statistics about resource usage
 - Answer: Yes, there is a python log reader. And other solutions
 - Relevant links:
<https://htcondor.readthedocs.io/en/latest/apis/python-bindings/api/htcondor.html#reading-job-events> (BrianB)
 - Example application using the job event parsing:
https://github.com/bbockelm/condor_watch_q (BrianB). Note this is now packaged with HTCondor itself!
 - Another example application: https://github.com/JoshKarpel/condor_necropsy (BrianB). That one actually produces simple statistics (e.g., average / median runtime, data movement, memory usage) from a given logfile. Not in as nice a state as `condor_watch_q`, but useful nonetheless.
 - Christina: depending on what level of detail you want, something like this can give a quick summary:

```
condor_history ClusterID -limi N -af MemoryUsage DiskUsage
```
- (Jeff Templon, Nikhef) Comment: For `condor_ssh_to_job` the “if it’s available” is probably more important in Europe than in the US, maybe good to mention that. (Christina: this is a good point! `Ssh_to_job` works best in local/closely shared resources, not so much when running across the Grid.) Aha thanks, yes I was talking about the Grid stuff.
- (Chris Brew, RAL) Comment: we find interactive jobs to be useful for scheduling access to shared resources where there is not too much contention, for example our testing GPUs.

Manage workflows with HTCondor DAGMan (Lauren Michael)

- (Name, affiliation): (Comment or question)
- (Mat Selmec, UW-Madison): Question: Can you have the “SCRIPT PRE” line for a node before the “JOB” line?

- Answer: yes. You can define your node statements in any order you want, and HTCondor will figure it out once the whole file is read.
- (John Knoeller, UW Madison). DAGMan now uses a 2 pass parser so that the order of statements in the DAG file does not matter.

After Break

HTCondor deployment at CC-IN2P3 (Emmanouil Vamvakopoulos)

- (Name, affiliation): (Comment or question)
- (Chris Brew: RAL): Comment, the Job router for assigning accounting groups was a God send (or Todd send) for us allowing us replace about 30 nested IfThenElse statements that were a nightmare to maintain with about ten simple lines of config!
- (Emmanouil: CC-IN2P3) Hello Chris accountinggroup setup with usermap function is send by John Knoeller (TJ) . I test it and I slightly modified it to be coherent with htcondor-ce 4.1.0 , and condor 8.8.9. We are going to see more details on TJ talk on Tuesday afternoon.

Replacing LSF with HTCondor: The INFN-T1 experience (Stefano Dal Pra)

- (Name, affiliation): (Comment or question)

HTCondor philosophy and architecture overview (Todd Tannenbaum)

- (Name, affiliation): (Comment or question)
- Antonio Perez-Calero, CMS: on the point of having multiple negotiators in a condor pool, we indeed run 3 negotiators in our CMS Global Pool (~250k CPU cores). The original reason was to split the pool of resources into fractions of clearly separated (geographically) sets, in order to keep negotiation cycle time under control (typically below 300s). Even after the introduction of multithreaded negotiator processes, also reducing cycle time, we've kept the 3 negotiators in the central manager of the pool with high success.

Questions and comments not related directly with any presentation

- (Thain, Greg): YouTube videos of past presentations are available at <https://www.youtube.com/channel/UCd1UBXmZlqB4p85t2tu-gLw>
- (Carsten Aulbert, AEI): Regarding GPUs and multi-user systems: For the Nvidia driver, there exists the option "**NVreg_RestrictProfilingToAdminUsers**" where a non-root user can access low level counters/events for development when set to 0. However, the driver's release docs usually imply a security risk in doing so, but does not elaborate. How do HTCondor folks here handle this? Don't care or never set it?
 - (John Knoeller, UW Madison), We never set it, the GPU monitor program that would be impacted by this runs as root
 - (CA, AEI) Does this mean, on these machines, we should not run Condor on? Or at least without the GPU monitor? Setting it to 1 would not really help with non-root users (nor non CAP_SYS users) :(
 - (John Knoeller, UW Madison), I believe it means that if you wish to set this option to 1, Condor would not be affected.

- (Götz Wasch, DESY), I think users can crash the GPU if they are allowed to access the counters, similar to the performance counters of the CPU. I was asked by a user to allow this, but it was no help as that test machine had a GPU that was too old.

Tuesday 22 September

Before Break

What is new in HTCondor? What is upcoming? (Todd Tannenbaum)

- (Name, affiliation): (Comment or question)
- (Jeff Templon, Nikhef): about annexation, is it possible for an admin to add / remove nodes dynamically (already)?
- (Mihai Duta, Diamond Light Source): we have tested Singularity jobs in Condor in the vanilla universe, should we use the singularity one? What are the advantages?
 - (Götz Wasch, DESY): It knows about singularity and you can set specific options like limiting the Singularity images to use and to set specific bind mounts (although you could do that in singularity.conf as well).
 - (Matyas Selmecei, UW-Madison): Singularity is not a 'universe' like vanilla or docker -- it is something an admin sets up to force (selected) user jobs to be inside a sandbox. ['Singularity support' manual page](#)
 - (Mihai Duta, DLS): thank you. Our condor is deployed via Docker in a kubernetes cluster. Our singularity jobs do not bind mount anything, all binding is controlled via k8s manifests, and in fact, most jobs tested so far run as user nobody with "singularity --no-home" option. So, apart from "--nv", no need for much else on the singularity side. We'll give it a try, though.

Installing HTCondor (Mark Coatsworth)

- (Name, affiliation): (Comment or question)

Pslots, draining, backfill: Multicore jobs and what to do with them (Gregory Thain)

- (Name, affiliation): (Comment or question)
- Stefano, CNAF: where is a recipe to get data used to create the report on "unused slot reason"?
- Chris Brew (RAL): I never quite got:
`START = $(START) && ((RequestGPUs >= 1) || ((CPUs - RequestCPUs) > (8 * GPUs)))`
 to work. I.e start CPU jobs if there are more than 8 CPUs free for each GPU. I just have another go with it
- Chris Brew (RAL): It's also very easy to write your own draining demon using Startd Cron jobs

Comments from the chat:

Jeff: Good one Miron, scheduling for memory. It's very useful to be explicit in the beginning on how to ask for memory. In our national e-infra we have two clubs - one that says ask for memory by asking for memory, the other says ask for memory by asking for more cores. This leads to interesting phenomena.

From jose.flix.molina@cern.ch to Everyone: (3:56 PM)

yes, indeed it is also a nice topic for accounting, because those wasting cores....

From Greg Thain to Everyone: (3:56 PM)

We would encourage HTCondor jobs to request both memory and cores, so that the job focuses on describing what it wants. The site can then chose what size slot to run the job on

From Jeff to Everyone: (3:57 PM)

Sounds great to me ... I am in the "ask for memory by asking for memory" camp :)

From S Dal Pra to Everyone: (3:58 PM)

I think its also a matter of defining a "pricing policy"; currently wallclocktime * HS06 is paid, not memory, AFAIK

From Brew, Chris (STFC,RAL,PPD) to Everyone: (3:58 PM)

Jeff, I guess really you need to decouple from "Standard" Slots, so asking for more CPUs doesn't give you more memory. Just more CPUs

Chris, I agree. Lots to think about as we roll out Condor here.

From MIRON LIVNY to Everyone: (04:46 PM)

@Jeff. If we focus on Memory we need to decide how we do "power assignments" for memory and what does it mean to do "fair share" for memory ... we are still very core focused in our scheduling. In CHTC we have one server with 4 TB of memory and a fairn number of servers with more than .5 TB ... makes scheduling "interesting"

HTC at DESY (Christoph Beyer)

- (Name, affiliation): (Comment or question)

HTCondor at GRIF (Andrea Sartirana)

- (Name, affiliation): (Comment or question)
- (Brian Lin, UW Madison): (Comment) I'd love to hear more about the HTCondor-CE documentation that you'd like to see and the difficulties that you've had finding documentation that you were looking for
 - [Italiano, INFN-Bari] have you managed to use lcms ? please share configuration
 - (Brian Lin, UW Madison): Italiano, is there a reason why you want to use LCMAPS over the built-in HTCondor mapfile?
 - [Italiano, INFN-Bari] We support dynamic mapping to a pool account. We have used lcas/lcms in order to support dynamic mapping while using CREAM-CE as frontend to our computing farm. Of course I hope to switch as soon as possible to IDTOKENs. In the meanwhile We are using argus because I haven't managed to successfully configure LCMAPS
- [Emmanouil,CCIN2P3] Hello Andrea which are issues with upgrades ? ...

After Break

Archival, anonymization and presentation of HTCondor logs with GlideinMonitor (Marco Mambelli)

- (Greg Thain, UW Madison) This anonymizer is generally useful -- is it available for general use, so we can anonymize log files that are sent for debugging?

Status and plans of HTCondor usage in CMS (Marco Mascheroni)

- (Name, affiliation): (Comment or question)

Classified ads in HTCondor (Jaime Frey)

- (Name, affiliation): (Comment or question)

Job submission transformations (John Knoeller)

- (Name, affiliation): (Comment or question)

Questions and comments not related directly with any presentation

- (Name, affiliation): (Comment or question)

(Patrik Lason, Cyfronet): I have to run something instead of CREAM-CE, but with Slurm as a batch system. Should I go with ARC-CE or HTCondor? Are there any obvious pros and cons?

- (Chris Brew - RAL): Patrik, You are at a Condor workshop so the audience might be somewhat biased, for balance, we run a reasonably size and fairly complex HTCondor setup behind ArcCEs and although it's been on my roadmap to replace the ArcCEs with HTCondorCEs, I never seem to find the time to do it - Even having to rewrite the Arc config to go from Arc 5 to 6, didn't quite persuade me to find the time to deploy HTCondorCE. So I guess the upshot is that they both work, and on a technical level there is not a huge amount to choose between them. Are you embedded in a wider VO or national structure? I'm a fairly big fan of (mostly) doing what your peers are doing so you can mutually support each other.
- (Brian Lin, UW Madison): one thing to consider is that if you need it, the HTCondor-CE and APEL support is specific to HTCondor-CE with an HTCondor batch system. Though we'd certainly be interested in contributions to change that!

Wednesday 23 September

Before Break

HTCondor-CE overview (Brian Hua Lin)

- (Name, affiliation): (Comment or question)
- Jeff Templon, Nikhef. No reaction on the hand raise. I miss one function of the CE: information providing. CEs in the past published what they could do.

A followup comment/question: does an HTCondor cluster “know” how many cores it has? If it does, why is it not just a question of reporting that number (of the HTCondor batch system that the HTCondor CE is pointing to)? Why should an admin have to “give” the CE that data?

- (Matyas Selmecei, UW-Madison): an HTCondor collector knows how many cores are in the pool, but doesn't know anything about which ones are accessible -- for example, if 200 of 400 are reserved for local users via site policy, HTCondor cannot report that.
- Thanks, that is clear. I don't think we'd normally run a CE like that. Our guarantees are normally on fair-share basis which means in principle the whole cluster is available to everybody, but in practice it depends on how much you've used and how much others have used, as to how much you'll get at any given moment. If cores were really reserved, we'd put those cores in a different cluster.
- (Miron) This is my point, someone has to tell the CE what to publish. We can offer a tool that will run a “condor_status” on the local collector that selects only resources tagged as “CE enabled” and report to an external information system. In OSG we use a collector (and ClassAds) to collect information from the CEs.
- (Matyas) Policies can be arbitrarily complex -- instead of reserving for local users, for example, you could reserve for multi-core jobs, or merely *prefer* jobs from certain departments.
- Thanks and this is good. Back to the comment, I just meant that if we had cores in the batch system that were *not accessible to the HTCondor-CE at all* then we would not have them in that batch system. As I write this I realize I may have missed a hidden assumption: are your HTCondor CEs single-VO critters, or can a single HTCondor CE service handle jobs from multiple VOs?? I know the single-VO service model is much more common in the US than in the EU.
- US sites are definitely multi-VO -- which is another reason for the complication. A CMS site, for example, could run CMS jobs and opportunistic (OSG VO) jobs, but not ATLAS jobs. (I was thinking of university clusters when I mentioned the ‘local user’ restriction.)
- OK, that's an interesting case. I was thinking about our standard grid clusters in the EU. No local users at all, serving tens of different VOs using order 4 or 5 CE machines. We typically get the CE to report the total number of cores it has, regardless of whether all cores are available to each VO. Sometimes we add additional information like max running jobs per VO, so that ATLAS could find out (using a suitable query) that it has access to 1700 of the 4000 cores.
- Emmanouil, CCIN2P3, GLIDERS from a Gliding factory could use htcondor-ce collector to publish their own classadds as a local collector endpoint. If we need this functionality in the WLCG environment in order to become default , we need to be coordinated.

(Jeff): I don't understand the comment, sorry. (EV) Now is it better ? Yeah, except I haven't yet understood exactly why the gliders should do that. Or why we would need this functionality in WLCG. It's probably that I don't know enough Condor principles yet, it may well become clear later.

Replacing CREAM-CE with HTCondor-CE: The INFN-T1 experience (Stefano Dal Pra)

- (Name, affiliation): (Comment or question)

HTCondor-CE configuration (Brian Hua Lin)

- (Name, affiliation): (Comment or question)
- Emmanouil - ccin2p3 , do you have a use cases where the +maxdisk (requestdisk) is needed ?
- Stefano, CNAF, slide 28: does a file lookup happens to get the AcctGroup? (Answer: no, that's specific for slurm)

How I learned to stop worrying and love the HTCondor-CE (Max Fischer)

- (Name, affiliation): (Comment or question)
- (Jeff, Nikhef) - I could not attend the talk, I looked at the slides. Useful information, thanks!
- (Brian Lin, UW Madison): Max, if you have notes on where you were initially confused with differentiating errors between the WMS + CE + BS, I'd love to see them!
 - Primary candidates are features such as PeriodicRemove/Hold. Often VOs will report errors just with "PeriodicRemove/Hold expression ... evaluated to True" and you have to untangle the thread from there on. Tagging these with the service that killed them is quite helpful.
 - A hard-to-debug case (and probably impossible to catch by HTC) is any kind of mix-up in configuring the CE instead of the BS or vice versa.
 - Of note: This has gotten a lot better once a) we learned to keep the CE configuration simple and b) we learned to not strip any layer indicators. 9619
- (Helge Meinhard, CERN): @Andreas Haupt: Would be good if you could write down a bit more details about the confusion between UGE/Torque and HTCondor concepts; it could help improve the HTCondor documentation.
 - We started with remote submits (condor_ce_submit) to the HTCondor-CE which obviously isn't needed. You need to run a local condor schedd with local condor_submit but grid universe. Remote submits without running services locally is default in "the old grid world"
 - The concept of a schedd that manages local jobs (and handles the jobs for you) took some time to be understood. In UGE/Torque/Slurm the Scheduler is always running on the master node. So remote submits are default, one does not really expect a different concept at the very beginning
- (Brian Lin, UW Madison): Thomas, CentOS 8 (and Python 3) support for HTCondor-CE is coming soon (™). Certainly by the end of the calendar year!
 - Thanks :)
 - We have started to collect some of our experiences under <https://confluence.desy.de/display/grid/CondorCE+Setup> (but it might be still more of a logbook, so YMMV)

After Break

HTCondor-CE live installation (Brian Hua Lin)

- (Name, affiliation): (Comment or question)

HTCondor-CE troubleshooting (Brian Hua Lin)

- (Name, affiliation): (Comment or question)

What is next for the HTCondor-CE? (Brian Hua Lin)

- (Thomas, DESY): do you plan also to support other token types like Macaroons in addition to SciTokens in the long run?

Running a large multi-purpose HTCondor pool at CERN (Ben Jones)

- (Name, affiliation): (Comment or question)
- Emmanouil, ccin2p3, do you use defrag process? How many defrag daemons do you use?
 - (Ben, CERN) We have several :) In share we have 3 running. One general one, one for "cloud" and one for "preempt".
 - Things like:
 - `DEFRAG_RANK = ifThenElse(Cpus >= 8, -10, (TotalCpus - Cpus)/(8.0 - Cpus))`
 - `DEFRAG_WHOLE_MACHINE_EXPR = ((Cpus == TotalCpus) || (Cpus >= 8)) && StartJobs =?= True`
 - `DEFRAG_REQUIREMENTS = PartitionableSlot && Offline != True && StartJobs =?= True && XBatch != True`
- If yes, which is the primary key in order for the separate worker per defrag instance?
 - For the non-normal case we just have the machine export an attr, like "XBatch" or "CernPreempt"

Challenge of the migration of the RP-Coflu-Cluster @ CERN (Xavier Eric Ouvrard)

- (Name, affiliation): (Comment or question)

Questions and comments not related directly with any presentation

- (Patrik Lason, Cyfronet): I have to run something instead of CREAM-CE, but with Slurm as a batch system. Should I go with ARC-CE or HTCondor? Are there any obvious pros and cons?
- (Chris Brew - RAL): Patrik, You are at a Condor workshop so the audience might be somewhat biased, for balance, we run a reasonable size and fairly complex HTCondor setup behind ArcCEs and although it's been on my roadmap to replace the ArcCEs with HTCondorCEs, I never seem to find the time to do it - Even having to rewrite the Arc config to go from Arc 5 to 6, didn't quite persuade me to find the time to deploy HTCondorCE. So I guess the upshot is that they both work, and on a technical level there is not a huge amount to choose between them. Are you embedded in a wider VO or national structure? I'm a fairly big fan of (mostly) doing what your peers are doing so you can mutually support each other.
- (Brian Lin, UW Madison): one thing to consider is that if you need it, the HTCondor-CE BDII and APEL support is specific to HTCondor-CE with an HTCondor batch system. Though we'd certainly be interested in contributions to change that!
- (Patrik): Chris: sure, that's why I ask. It's the best audience to ask how do you compete with other solutions. I have to stay with Slurm for local, HPC users but need to run WLCG jobs too, so I try to find the best (and simple) solution. I'd like to i.e. just run a VM for Arc services, second one for HTCondor, create partitions in Slurm and forget ;-). Arc looks easier, but today I got a request to support HTCondor for Ligo/Virgo VO. I don't like the idea of having HTCondor inside Slurm, but I can't remove Slurm and I don't want to

statically split our resources.

Brian: thanks, good to know!

- (Patryk) I forgot to mention above that we also only support BDII for the HTCondor-CE + HTCondor case

Thursday 24 September

Before Break

HTCondor with containers and Kubernetes (Greg Thain)

- (Name, affiliation): (Comment or question)
- Thomas, DESY: are their plans to combine the file I/O with SciTokens as well? I.e., maybe to use Condor's SciToken management as well to request tokens for file I/O from storages and hand them over to the job? I.e., to avoid mounted namespaces altogether?
- Jeff, Nikhef : I was actually hoping the other way around - that HTCondor would help me avoid having to learn Kubernetes. It's ok to do it the other way around, that does imply that there needs to be a way to have Kubernetes talk to HTCondor, as HTCondor will know what the pressure is (scheduling) which tells Kubernetes how many of which type of containers need to run. Here I'm talking about within a single site, not about starting up Google Cloud or Amazon stuff.
 - About docker security : what I often hear is that it is easier to escape out of a container than out of a VM. I am not an expert so am not qualified to know whether you are correct or these other people.
 - Thomas: containers are in the end 'ordinary' processes within their own namespaces and live on the same kernel, i.e., you are not separating kernels from each other but isolate processes stricter including own mount namespaces
 - Matyas Selmecki, UW-Madison: it's easier, but not necessarily easy. It's important to keep your container runtime up to date, and ensure the containers do not have any more privileges than necessary. (It's harder to break out if they don't have CAP_SYS_ADMIN.)
 - I guess it depends on what one wants to do with the pods. If they are standard (EU) grid-type pods, ie no ssh access to them, no logging in, etc, then it really doesn't matter if it's a container or VM. if we let users into the pods then we really need to worry if there is sufficient isolation, given the shared nature of our clusters. It would again be different if the whole cluster was dedicated to one VO, because then a breach would only affect the breached VO, so their lack of security only affects themselves and not others. A breach on a multi-VO system where the bare OS is compromised entails shutting down the entire cluster until cleanup is done :-)
- David Schultz, UW-Madison: Is there a way to run singularity jobs inside the k8s startd to get different OS containers?
 - (Mihai Duta, Diamond Light Source): you can run Singularity images inside the other -- non-startd -- Docker container (replace sleep with singularity exec/run) just like any other job. Naturally, singularity has to be installed in the docker container.
 - David: I was hoping to avoid running the non-startd Docker container, to not give out privileges

- Christoph Beyer / DESY: You can run singularity images in a docker container - did that :)

Combining cloud-native workflows with HTCondor jobs (Clemens Lange)

- (Name, affiliation): (Comment or question)
- Chris Brew (RAL): Clemens This reminds me slightly of Analysis Description Languages. I've only seen mentions of them in passing in other talks but it looks like the HSF ran a workshop on them last year <https://indico.cern.ch/event/769263/>

HTCondor in production: seamlessly automating maintenance, OS and HTCondor updates, all integrated with HTCondor (Oliver Freyermuth)

- (Name, affiliation): (Comment or question)
- Chris Brew (RAL): Very nice, I have something very similar but will be stealing some ideas from you. One thing we have you might not is that since the nodes drain state is in their ClassAds, other nodes can use that in their decisions about whether to drain themselves. So you don't need a random smear but can put a cap on the number of draining. We then put that cap as a Collector_Attr so we could alter it and nodes could query it dynamically. So you could have a low number for your >30days but everything for NOW for the zero day remotely executable kernel flaw. Another one you may or may not have (I stole it off the RAL Tier 1) is to not take a node completely offline for a CVMFS problem but jobs stop accepting jobs from the affected VO - We've found that draining all that VOs jobs out from the node and allowing CVMFS to remount the repo is enough to clear many problems.

- (Oliver, Uni Bonn) Many thanks for these very interesting points! Indeed I immediately see the gain from capping the total number of draining nodes. Up to now, we have a very nice mix of diverse runtimes of jobs so it mostly does not harm if the draining is to some degree synchronized. For CVMFS, indeed we are still waiting for the first mount failure — we are not using it with autofs, but explicit mounts, and they have not failed us since >2 years ;-).

- (Chris Brew) We don't force periodic reboots (though I'm tempted) but do define minimal Kernel versions:

```
COLLECTOR_ATTRS = $(COLLECTOR_ATTRS), ClusterMaxDraining,
ClusterMinKernelSL6, ClusterMinKernelCentOS7
COLLECTOR.SETTABLE_ATTRS_ADMINISTRATOR = ClusterMaxDraining,
ClusterMinKernelSL6, ClusterMinKernelCentOS7
```

```
ClusterMinKernelSL6      = "2.6.32-754.30.2.el6.x86_64"
ClusterMinKernelCentOS7 = "3.10.0-1127.19.1.el7.x86_64"
ClusterMaxDraining      = 50
```

HTCondor Annex: bursting into clouds (Todd Miller)

- (Name, affiliation): (Comment or question)
- Chris Brew (RAL): Partly for the annex and partly for Miron's question at the end of the session, one things I'm seeing is the proliferation of different opportunistic resources but it does appear that many of them seem to starting to offer k8s as an option so that might be a good target for the annex to support because it could then allow us to target many different resources.

CHTC partners with Google Cloud to make HTCondor available on the Google Cloud marketplace (Cheryl Zhang)

- (Name, affiliation): (Comment or question)

After Break

HTCondor offline: running on isolated HPC systems (Jaime Frey)

- (Name, affiliation): (Comment or question)

HEPCloud use of HTCondor to access HPC centers (Anthony Tiradani)

- (Name, affiliation): (Comment or question)

HPC backfill with HTCondor at CERN (Pablo Llopis Sanmillan)

- (Name, affiliation): (Comment or question)

HTCondor monitoring at ScotGrid Glasgow (Emanuele Simili)

- (Name, affiliation): (Comment or question)

Questions and comments not related directly with any presentation

- (Matthew West): (Question) What is the distribution of workload sizes (# jobs per user, avg core per job, memory & disk usage) over the course of a year on the cluster? Is cluster usage dominated by a few large groups or is usage more flat?
- (Matthew West): (Question) What features does SLURM provide that HTCondor doesn't that makes it so often preferable for HPC systems?
 - (Götz Waschke, DESY): Slurm was built from the ground up for very large parallel jobs on many nodes. It is integrated with the job start mechanisms of the different MPI implementations.
 - (Matthew West): That makes sense. But it seems HTCondor can handle multi-node jobs in Parallel Universe. I guess its a matter of what your overall cluster workload is like.

Friday 25 September

Before Break

HTCondor at Nikhef (Jeff Templon)

- (Helge Meinhard, CERN): Rene van Dantzig also convinced me to try setting up a Condor pool on the desktops for the Chorus experiment at CERN - probably some time around 1994 or 1995.
- (Todd Tannenbaum, CHTC): If you always want "condor_q -nobatch", you can put in your condor_config "CONDOR_Q_DASH_BATCH_IS_DEFAULT=False"

- (Carsten Aulbert, AEI): @Todd - is there something like to also if I wanted `condor_q -all` as a default (other than defining a shell alias)?
- (Todd Tannenbaum, CHTC): @Carsten - I cannot think of anything off the top of my head that would give -all by default
- (Jaime Frey, CHTC): @Carsten, set CONDOR_Q_ONLY_MY_JOBS=false
- (John Knoeller, CHTC), If you want these options to apply only to yourself, you can add them to ~/.condor/user_config
- (Chris Brew, RAL) Isn't there a template file somewhere that you can use to change to the format of the output of condor_q?
- (John Knoeller, CHTC), there is a template file <https://htcondor-wiki.cs.wisc.edu/index.cgi/wiki?p=ExperimentalCustomPrintFormat> coming soon to the manual. The TL:DR is Q_DEFAULT_PRINT_FORMAT_FILE=<file>
- (Chris Brew, RAL) The thing I always thought was missing from condor_q, but someone is probably going to tell me isn't actually, is a way to force -af to fully resolve expressions, -format seems to by default but -af doesn't.
- (John Knoeller, CHTC) -format and -af are both setting up the same classad pretty printer, so they will resolve expressions in exactly the same way. -af is the equivalent of -format %v which formats string values slightly differently than -format %s does. CustomPrintFormats is another way to set up that same pretty printer.

HTCondor's Python API - the Python bindings (Jason Patton)

- (Name, affiliation): (Comment or question)

HTMap: Pythonic High Throughput Computing (Todd Tannenbaum)

- (Thomas, DESY): can I somehow(?) ensure/require a minimum runtime with htmmap? In case users map very short functions into jobs and hit the negotiator unfortunately? Todd answered it (or keep a job running for a user, that pulls all htmapped functions for some longer time?)
- (Chris Brew: RAL) I was wondering if you could supply splitting functions or hints to htmmap so the person with billions of millisecond jobs can say batch my input into chunks of about X,000 before sending the jobs out. Possibly Dask answers the question

Lightweight site-specific Dask integration for HTCondor at CHTC (Matyas Selmecci)

- (Name, affiliation): (Comment or question)

REST API to HTCondor (Matyas Selmecci)

- (Name, affiliation): (Comment or question)

After Break

HTCondor security: philosophy and administration changes (Zach Miller)

- (Name, affiliation): (Comment or question)

- (Chris Brew, RAL): (Doesn't need to be a verbal question) Is there any documentation of setting up the map file for SSL? I think all my daemons are `ssl@unmapped`
 - (Zach Miller, CHTC) replies:
 - You can find the mapfile basics here: <https://htcondor.readthedocs.io/en/latest/admin-manual/security.html#index-68>
 - Example SSL entries here ("The Classic"): <https://research.cs.wisc.edu/htcondor/CondorWeek2011/presentations/zmiller-ssl-tutorial.pdf> (Slide 12)
- (Jeff Templon, Nikhef): For the user tokens, how does the receiving machine know to what group the user belongs, if there is no mapfile? It works if your world is user-centric, not if it's more group-centric. Or I'm missing something. Disclaimer: I am a token idiot!
 - Maybe my answer is in the next talk - the Nikhef schedd looking up which group I am in sounds wrong, as I may be in multiple groups ... I want the token with which I submit to tell the system which one I want to charge my usage to and/or base my access rights on.
 - (Zach Miller, CHTC) replies:
 - The "mapfile" mentioned here was the CERTIFICATE_MAPFILE, which is separate from any group/accounting mappings.
- (Dennis van Dok, Nikhef) Can an admin trace the use of a token throughout their systems?
 - (Zach Miller, CHTC) replies:
 - There isn't an audit log specific to IDTOKENS. You can see information in various log files but there isn't a "go to" audit log. The closest thing to that are messages that the condor_schedd prints to the debug level D_AUDIT, which you can split into a separate file. But that covers only the condor_schedd and not all uses of security credentials.
- (Thomas, DESY) Could the token delegation be coupled to an AAI? Let's say, I authz myself at a AAI, that site A and side B are connected to. There I request tokens to run jobs for resource requirements X - and with the tokens can submit my jobs to the clusters at A or B?
 - (Zach Miller, CHTC) replies:
 - I'm not super-familiar with the term AAI and likely this wouldn't be possible (right now) with IDTOKENS since the condor_collector is doing the token signing and we don't have ways to call in or out of the collector to other infrastructures. But this might be possible using SciTokens, which have their own libraries and could be programmed in to something like that. A service could be written that generates a SciToken once you've authenticated/authorized yourself to the AAI, and HTCondor can use SciTokens the same as IDTOKENS (job submission, in particular). [hope that answers your question somewhat...]

From identity-based authorization to capabilities: SciTokens, JWTs, and OAuth (Jim Basney)

- (Name, affiliation): (Comment or question)
- (Jeff Templon, Nikhef): It's not clear to me (besides the x509 case) how I can have the identity and roles/memberships separated. Especially for SAML they seem to come from the same place, which isn't how it's done in EU land.
 - (Jim Basney, NCSA) It's possible (though rare) in SAML to separate the IdP and AA (Attribute Authority). We have 3 "standalone attribute authorities" in eduGAIN

(<https://technical.edugain.org/entities>). We also operate a standalone SAML AA for LIGO to add LIGO roles/memberships to assertions from campus.

- (Jim Basney, NCSA) As I mentioned to Dave, the OIDC Provider and Authorization Server have separate authorization endpoints ([OIDC Discovery](#) versus [OAuth Discovery](#)), though with CILogon we just use the same server since OIDC is based on OAuth, it's easy to support both with the same codebase.
- (Dave Dykstra, FNAL) Keep in mind that even when the same service provides both the ID Token (Open ID Provider) and the access token (Authorization Server), that service can separately defer to a site's Identity Provider to verify the user's identity. Both CILogon and IAM do this.

Allow HTCondor jobs to securely access services via OAuth token workflow (Jason Patton, Zach Miller)

- (Thomas, DESY): for the scoping, will it also be possible to limit IPs and time?
 - (Jason Patton, CHTC): For IPs, I suppose that would depend on the OAuth provider accepting that information as part of the scope or audience. For time, the token files will be automatically wiped after some (schedd-configurable) time on the submit machine, and users can clear `*all*` of their OAuth tokens on the submit machine at any time by running `condor_store_cred delete-oauth`. Revocation and expiration times, however, are up to the token providers (who hopefully provide the user with some way to revoke tokens whenever they choose). It's important to note that the act of deleting a token file does not revoke the credential!
 - (Dave Dykstra, FNAL): I think typically scopes are expected to be a little higher level of abstraction. There would probably instead be some name, and the name is translated to an ip range that the service has configured and enforces.

Workshop wrap-up (Helge Meinhard)

- (Name, affiliation): (Comment or question)

Questions and comments not related directly with any presentation

- (Name, affiliation): (Comment or question)
- (Carsten Aulbert, AEI): Brief summary of Zoom chat question. Question was about having ``condor_q -nobatch`` or ``-all`` as defaults. The configuration options for these are `CONDOR_Q_DASH_BATCH_IS_DEFAULT` and `CONDOR_Q_ONLY_MY_JOBS`, respectively.