What's new in the HTCondor Software Suite (HTCSS) ? What's coming up?

HTCondor Week 2022

Todd Tannenbaum
Center for High Throughput Computing
University of Wisconsin-Madison

Some New Terminology

- Releases
  - Long Term Support (LTS) Releases Channel
  - Feature Releases Channel
- HTCondor Software Suite (**HTCSS**)
  - Access Point (AP)
  - HTCondor Pool =

    Central Manager +

    Execution Point(s) (EP)
  - HTCondor Compute Entrypoint (HTCondor-CE)
- OSG

Release Channels

- Long-Term Support (LTS) Releases (*formerly 'stable series'*)
  - Only bug fixes
  - vMajor.0.Update (e.g. 9.0.1, 9.0.2, … 9.0.15, 10.0.0, 10.0.1, 10.0.2, …)
  - Last year: v9.0 just released with 71 documented enhancements
  - Today: v9.0.12
- Feature Releases *(formerly 'developer series')*
  - Bug fixes plus new features
  - vMajor.Minor.Update (e.g. 9.1.0, 9.2.0, 9.3.0, 9.3.1, 9.4.0, …)
  - Today: v9.8.1

Update since last year

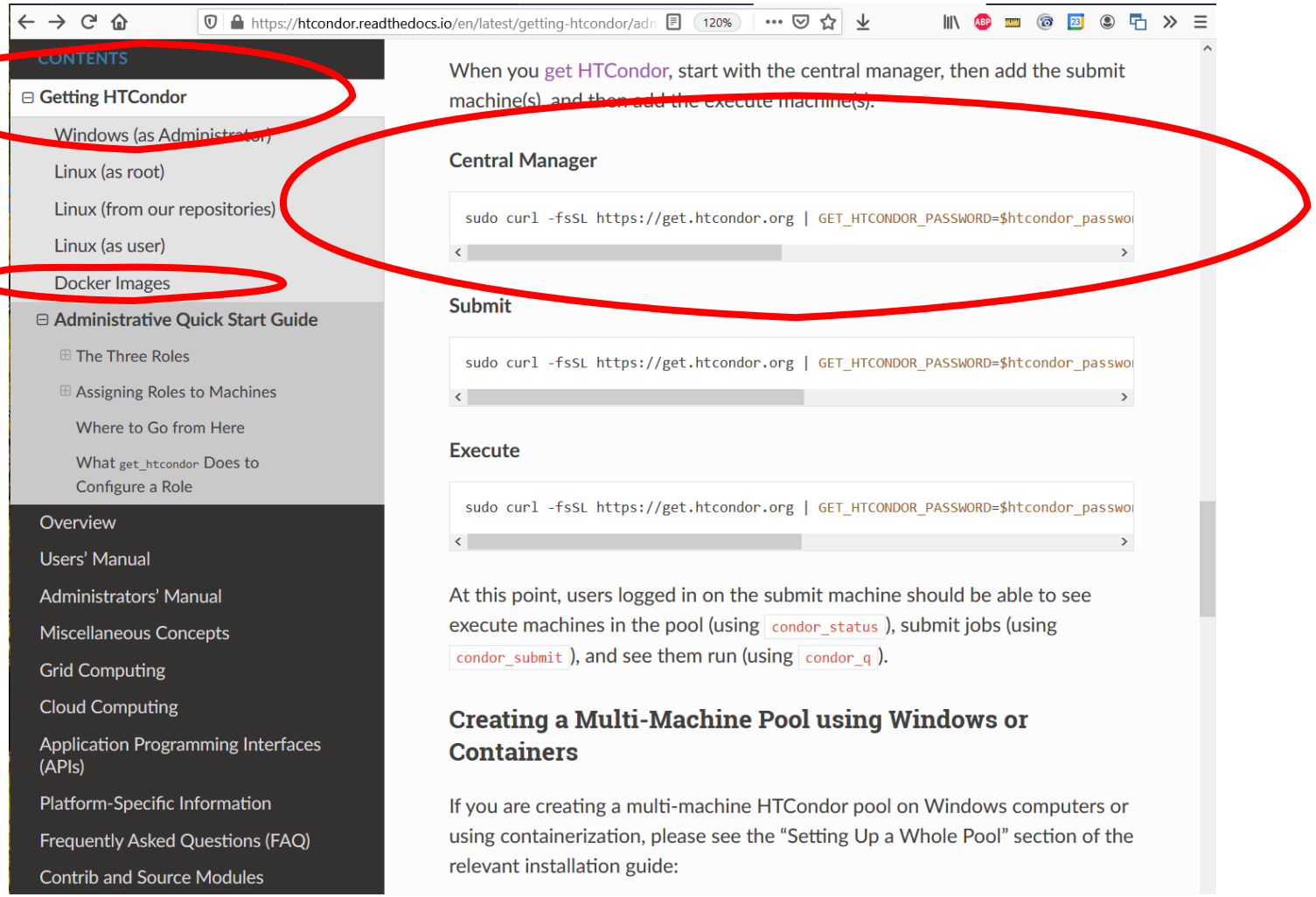
- Just going to focus on some "big" items keeping us "off the streets"
- For a complete list of "what has changed"
  - HTCondor
    - Highlights at: https://htcondor.org/htcondor/release-highlights/
    - Details at: https://htcondor.readthedocs.io/en/latest/version-history/development-release-series-91.html
  - HTCondor CE
    - Highlights at: https://htcondor.org/htcondor-ce/release-highlights/
    - Details at: https://htcondor.com/htcondor-ce/v5/releases/#htcondor-ce-5-version-history

What was introduced in HTCSS LTS v9.0.x?

Fast "one-line" install & configure

- Secure by default using tokens
- Watch demo of install and setup in a three minutes (demo starts at minute 4:00):

https://youtu.be/BsE0xvts-iA



New in v9.0 for end-users

- Transfer job data to/from web servers, Box.com, Amazon S3, Google Drive, MS OneDrive via file transfer plugins now supporting uploads, authentication, and improved error handling
- Improved support for GPUs, including GPU utilization info and support for using GPUs in containers
- New condor_watch_q tool that efficiently provides live job status updates
- New tools and mechanisms to support jobs that checkpoint
- Fixed interactive submission of Docker jobs
- Many improvements to the Python API, including Python 3 support and new bindings for DAGMan and chirp
- DAGMan now provides a method for inline jobs to share submit descriptions
- You may now change some DAGMan throttles while the DAG is running
- You can assign priorities to DAGMan (scheduler universe) jobs
- And more…

New in v9.0 for end-users

- Transfer job data to/from web servers, Box.com, Ama… rive, MS OneDrive via file transfer plugins now supportin… and improved error handling
- Improved support for GPUs, inclu… using GPUs in containers
- New condor_watch… updates
- New tools a… t
- Fixed …
- Many … ing Python 3 support and new bindings
- DAGMan… od for inline jobs to share submit descriptions
- You may … ome DAGMan throttles while the DAG is running
- You can as… priorities to DAGMan (scheduler universe) jobs
- And more…

Topic of Christina's talk on YouTube at https://youtu.be/tQR2KovkSA8

New in v9.0 for sys admins

- Fundamental security model changes:
  - Secure by default
  - Authorization via Tokens: IDTOKENS, SciTokens
  - Oauth2 workflow to interoperate with other services
- New packaging
  - Native packaging has releases, release candidates, and daily builds
  - Official containers on Docker Hub designed to work well with k8s
  - Conda
- Push job information into Elastic Search
- More powerful classad transform language for HTCondor-CE, schedd
- Dataflow mode for jobs
- Container support improvements
- Cgroup limit improvements
- Cap the number of cores allocated to individual users
- New platforms (*e.g.* EL8, Ubuntu 20)
- And more…

New in v9.0 for sys admin

- Fundamental security model changes:
  - Secure by default
  - Authorization via Tokens: IDTOKENS, SciT
  - Oauth2 workflow to interoperate with
- New packaging
  - Native packaging has release
  - Official containers on D
  - Conda
- Push jo
- More po
- Dataflow
- Container
- Cgroup lim
- Cap the num allocated to individual users
- New platforms (*e.g.* EL8, Ubuntu 20)
- And more…

Topic of Greg's talk on YouTube at https://youtu.be/djh_URo-q40

What have we been up to this past year?

What have we been up to this past year?

HTWORDLE

- Same Rules as Wordle but
  - Winning word something commonly used w/ HTCondor
  - Team offers suggestions to the team captain, captain enters the word
  - 20 second timeout

After each guess, the color of the tiles will change to show how close your guess was to the word.

WEARY

The letter **W** is in the word and in the correct spot.

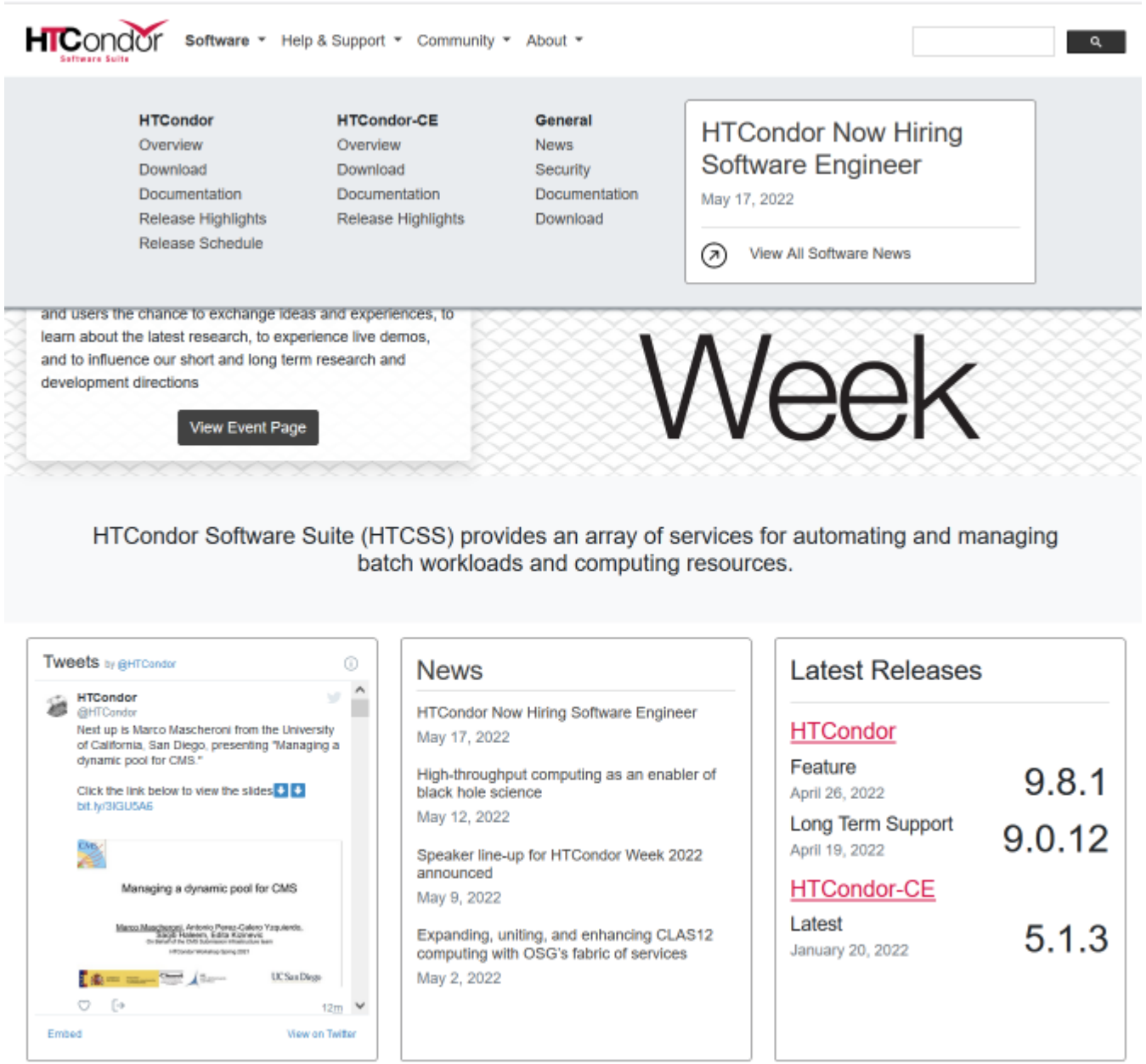
PILOT

The letter **L** is in the word but in the wrong spot.

VAGUE

The letter **U** is not in the word in any spot.

Updates to Web and Manual to reflect new terminology and the Suite idea



Upcoming OS and Architecture Support

Enterprise Linux 8 / Stream 8

- ETA: Available Now
- Currently based on Rocky8, planning to switch to Alma

ARM and PowerPC architectures

- ETA: Summer

Enterprise Linux 9 / Stream 9

- Challenges due to shifts in several libraries (e.g. OpenSSL, cgroups)
- Interoperability with HTCSS v8.x may be limited (upgrade to v9.x!)
- ETA: End of the year

Stop Runaway ~~Trains~~ Jobs!

- Stop runaway jobs!
- Following can go in your submit file:
  - allowed_execute_duration = <seconds>
    (from exec to exit)
  - allowed_job_duration = <seconds>
    (include job transfers)

Job Sets

- Users want to think about a set of jobs as it relates to their mental model (and NOT based upon when they submit). For instance:
  - Set of jobs analyzing genome 52
  - Set of jobs analyzing images captured on date xxx
- Manage as a set
- Aggregated statistics across all jobs in the set

Job Sets, cont

- See "Job Sets" section in the HTCondor Manual… it appeared in HTCondor 9.4.0!
- Job Set description file contains a name, an iterator, and at least one job.

Example:

```
name = MyJobSet

iterator = table inputfile,foo,bar {
    input_A.txt,0,0
    input_B.txt,0,1
    input_C.txt,1,0
    input_D.txt,1,1
}

job x=foo,y=bar my-job.sub

job {
    executable = a.out
    arguments = $(inputfile) $(foo) $(bar)
    transfer_input_files = $(inputfile)
}
```

Based on this job set description, with two job descriptions (which become two job clusters), you would expect the following output when submitting this job set:

```
$ htcondor jobset submit my-jobs.set
Submitted job set MyJobSet containing 2 job clusters.
```

Experimenting with new command line user interface

- *htcondor <noun> <verb>*
  - *"htcondor job submit", "htcondor job status", …*
  - *"htcondor dag submit", "htcondor dag status", …*
  - *"htcondor jobset submit", "htcondor jobset status", …*
  - Legacy tools (condor_q, condor_submit, …) not going anywhere…

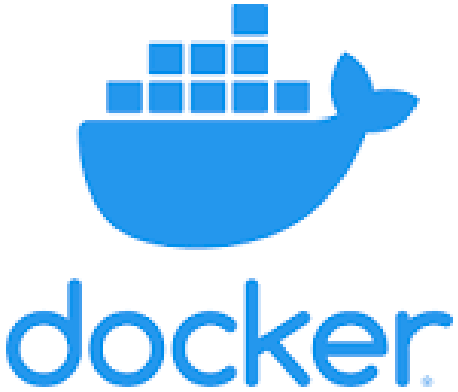
New Command Line User Interface, cont

$ htcondor job status 123.45
Job 123.45 is currently running.
It started running 2.1 hours ago.
It was submitted 3.6 hours ago.
Its current memory usage is 2.5 GB out of 4.0 GB requested.
Its current disk usage is 3.8 GB out of 5.5 GB requested.
It has restarted 2 times.

$ htcondor jobset status MyJobSet
MyJobSet currently has 6 jobs running, 9 jobs idle, 4 jobs complete, and 1 job held.
MyJobSet has so far used 210.4 hours of wall clock time and 190 hours of CPU time.
MyJobSet was created 4.8 hours ago.
MyJobSet contains the following job clusters with 10 processes each: 135, 136

Containers

- Docker and Singularity (Apptainer) support in HTCondor is very popular
- *BUT… does the user really care about the CR?*



Current world order:

```
universe = docker
docker_image = Debian
```

Or

```
universe = vanilla
requirements = HAS_SINGULARITY == True
+SingularityImage = "/cvmfs/…"
```

Containers

- Docker and Singularity (Apptainer) support in HTCondor is very popular
- *BUT… does the user really care about the CR?*

Current world order:

```
universe = docker
docker_image = Debian
```

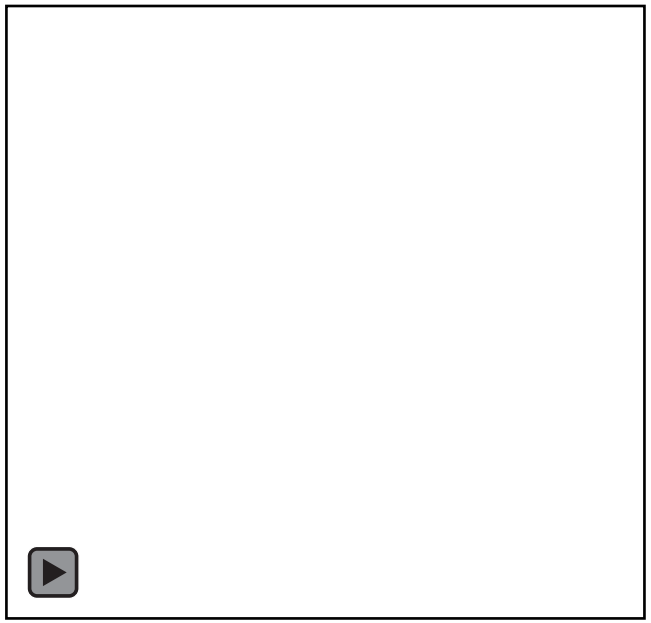
Or

extended_submit_commands

```
universe = vanilla
requirements = HAS_SINGULARITY == True
+SingularityImage = "/cvmfs/…"
```

New "Container Universe"

- EP advertises container runtimes available, and uses whichever one can get the job done

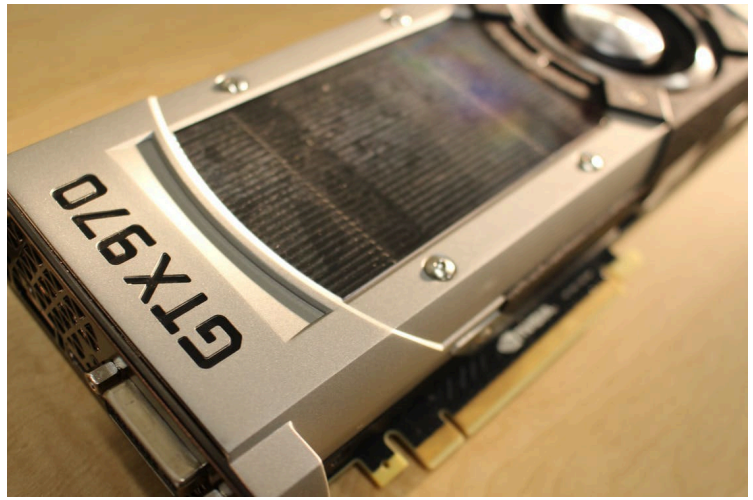


- New world order:

```
universe = container
container_image = /cvmfs/my/image/dir/
# Or container_image = docker://Debian
# Or container_image = myImage.sif
# Or container_image = http://xxx/image.sif
```

GPU Scheduling Improvements

- HTCondor has long been able to detect GPU devices and schedule GPU jobs
- More recently also:
  - Monitor/report job GPU processor utilization
  - Monitor/report job GPU memory utilization

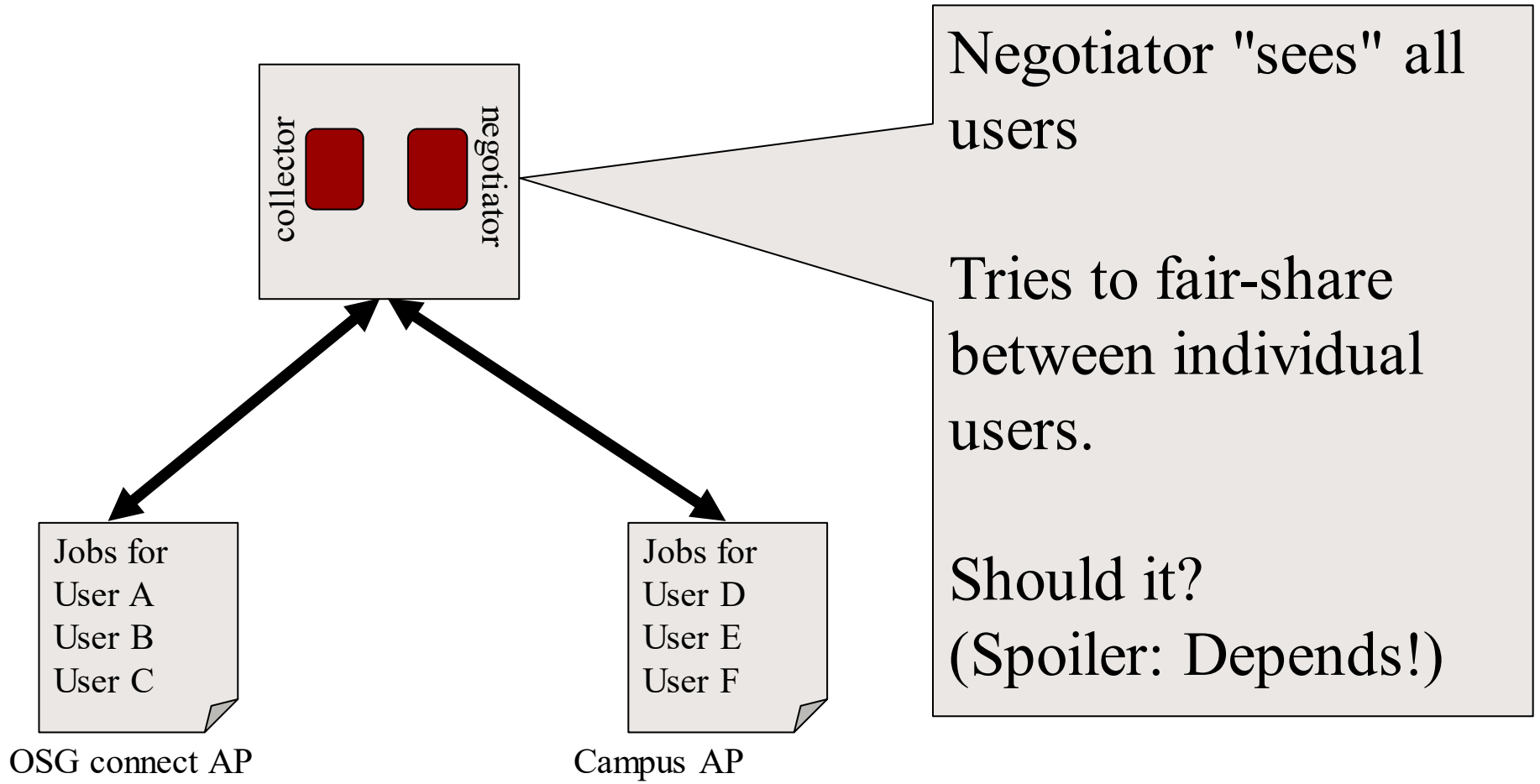


- ***New in v9.8.0: Support for heterogenous GPUs in one server***
  - E.g. a server with two different models of GPU cards
  - NVIDIA Multi-Instance GPU (MIG) partitioning
- Currently working on concurrent jobs on one device
  - User will specify a concurrency level per job set

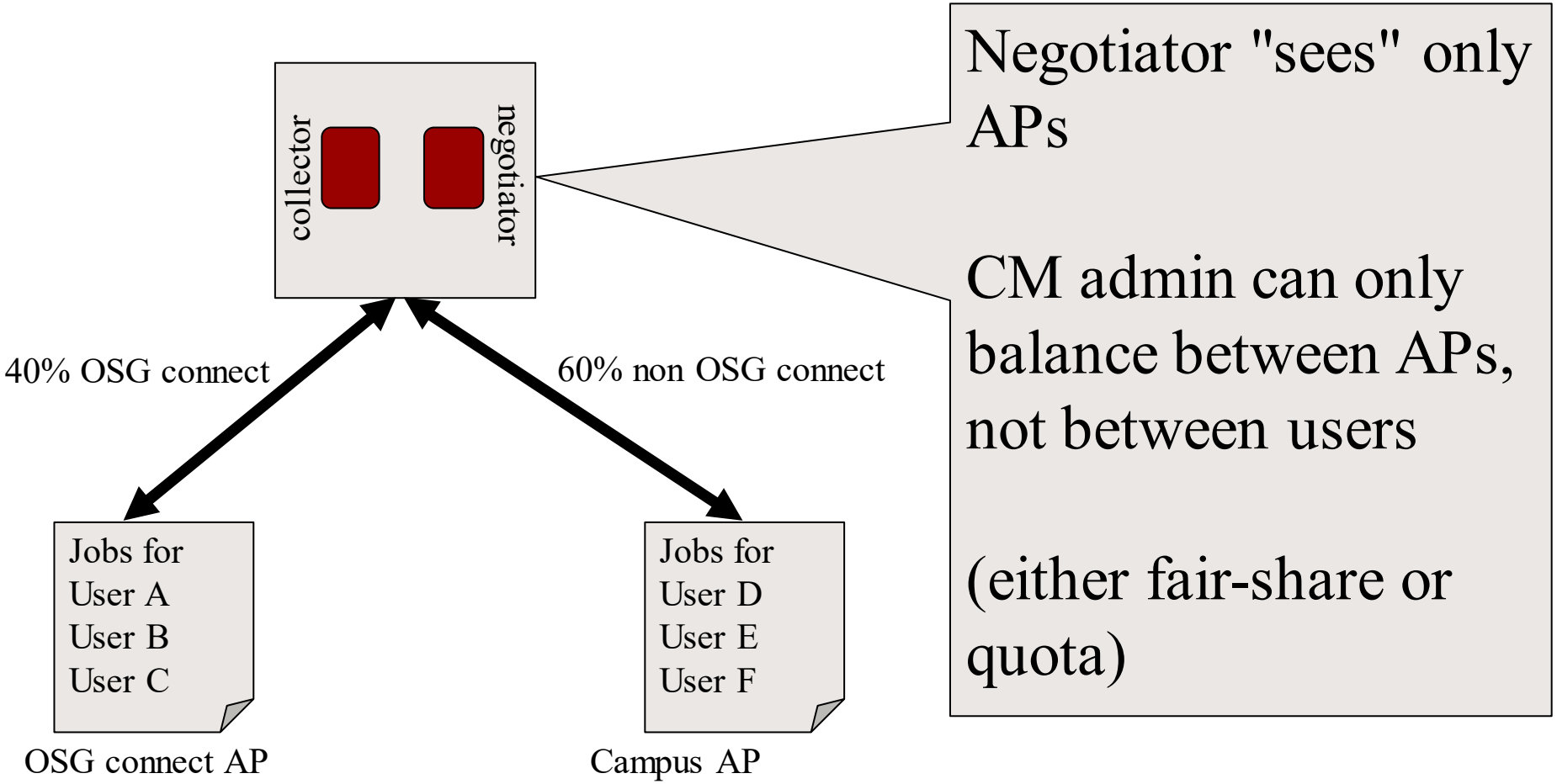
Submit File Example:

```
Executable = foo.exe
RequestGPUs = 1
RequireGPUs =
    Capability > 7.0
Queue
```

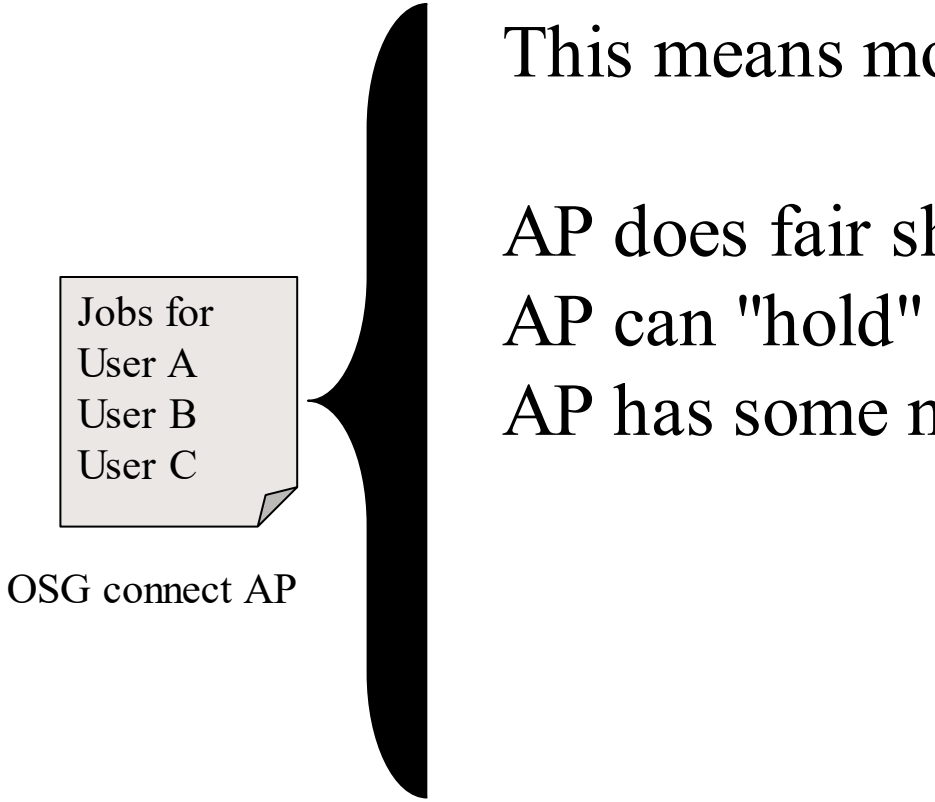
AP/CM interactions: As it is Today



AP/CM interactions: where are going…



AP/CM interactions: where are going…



This means more power (and work) for AP:

AP does fair share (controlled by AP admin)
AP can "hold" slots, do draining and merging
AP has some notion of accounting

Bring Your Own Capacity

- Add additional Execution Points that your Access Point can leverage to run your jobs!
- Continue work on mechanisms to enable end-users to provision computing capacity from
  - Local Clusters (e.g. campus Slurm HPC clusters)
  - Clouds
  - HPC Centers / Supercomputers
- Run a job set as a campaign at an HPC site
  - Now available for users in OSG OSPool
  - More presentations and demos today and tomorrow

Security Enhancements

- Host-based security is no longer the default. ☺
- Turn-key installation with "get_htcondor" tool sets up all the security knobs for you
- For IDTOKENS authentication, no longer need a signing key on every EP host in your pool
  - EP belongs to a pool and trusts the Central Manager
  - Central Manager can give capabilities for admin tools
- TOFU ! "Trust on First Use" ability for SSL
- Improvements against replay attacks

GSI to Tokens Transition

- Grid Community Toolkit support is going away, so HTCSS use of GSI authentication phased out
  - Use tokens instead (SCITOKENS, IDTOKENS)
  - Full details here:
    https://htcondor-wiki.cs.wisc.edu/index.cgi/wiki?p=PlanToReplaceGridCommunityToolkit
- Mostly done, a few improvements remain todo
- NOTE: HTCSS will still delegate and manage an X.509 proxy with the job, but the GSI authentication method is gone from v9.4 +

HTCondor-CE Work

- New syntax for writing job routes (classad transform syntax) has nice benefits including:
  - Statements evaluated in the order the are written
  - Use of variables that are not included in resultant job ad
  - Use of simple case statements
  - See https://htcondor.com/htcondor-ce/v5/configuration/writing-job-routes/
- Mechanism for a (hosted) CE to access capacity needing 2fa
- Ability for a job route to insert an IDTOKEN

Did we miss something you need?

TOWN HALL on upcoming Needs/Trends

Follow us on Twitter!
https://twitter.com/HTCondor

This work is supported by NSF under Cooperative Agreement OAC-2030508 as part of the PATh Project. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING