

"Why to convince your local administrator to upgrade HTCondor."

New Features for Users in HTCondor 9.0+

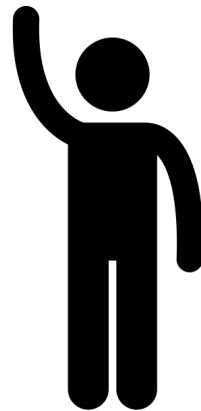
Christina Koch

Center for High Throughput Computing

I have data in
the "cloud."



amazon
S3



New File Transfer Plugins

- Improved curl, https, Box, Amazon S3, and Google Drive file transfer plugins supporting uploads and authentication.
- Enables greater flexibility for job data staging
 - Not limited by an access point with little storage space
 - Access just what you need from a full dataset without extra data movement
 - Fetch or put data into shared locations.
- Relevant Manual pages:
 - [File transfer using a URL](#)
 - [Jobs that require credentials](#)

New File Transfer Plugins

- Sample submit file syntax for an existing plugin:

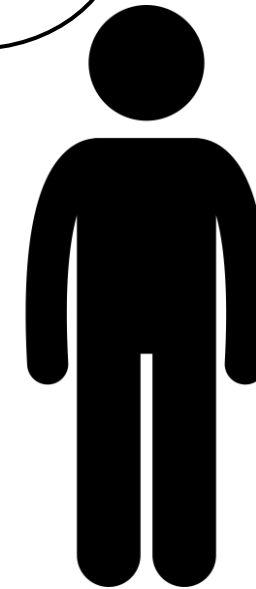
```
use_oauth_services = box
box_oauth_permissions = read:/public/
transfer_input_files = box://foo/bar.txt
```

- BYOP: Bring your own plugin!

- Users can now include their own file transfer plugin as a job specification and use it as the file transfer mechanism for other files.

```
transfer_plugins = myplugin=my_custom_plugin.py
transfer_input_files = myplugin://foo.txt
```

I want to know
more about my
job's GPU usage.



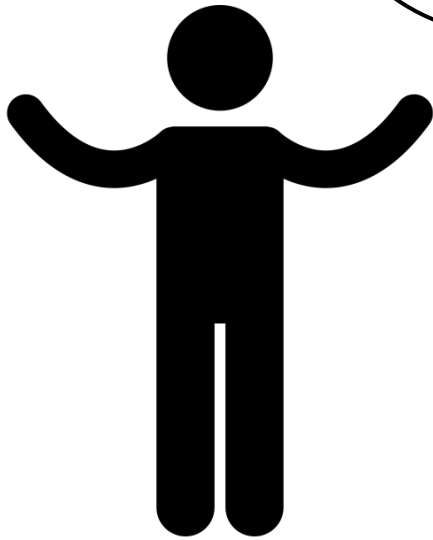
GPU Support Improvements

- Added information about GPU utilization in the job log file.

```
Partitionable Resources : Usage Request Allocated Assigned
Gpus (Average)          :      0.95          1          1 "GPU-fe797bae"
GpusMemory (MB)         :     10902
```

- Added GPU utilization attributes to the JobAd.
 - GPUsAverageUsage – GPU utilization
 - GPUsMemoryUsage – GPU memory utilization
- GPUs work with containers now (Docker + Singularity).
- More improvements coming!

I wish I could have
more real-time
updates about
my jobs.



Tools for Job Monitoring

- New `condor_watch_q` tool that efficiently provides live job status updates.

```
$ condor_watch_q
BATCH          IDLE  RUN  DONE  TOTAL  JOB_IDS
ID: 8906504    2    5    3     10   8906504.3 ... 8906504.9

[#####=====]

Total: 10 jobs; 3 completed, 2 idle, 5 running

Updated at 2021-05-24 08:57:38
```

Manual page: https://htcondor.readthedocs.io/en/latest/man-pages/condor_watch_q.html

Tools for Job Monitoring

- File transfer times are now recorded in the job log.

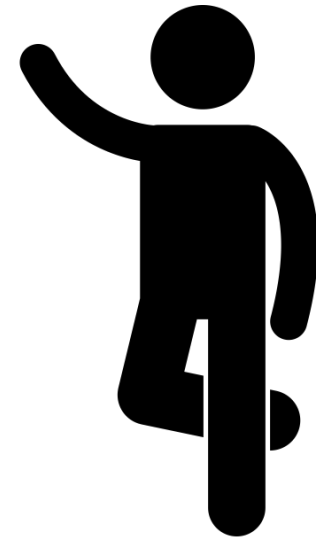
- Input transfer:

```
040 (9834570.074.000) 2021-05-20 12:03:41 Started transferring input files
      Transferring to host: <10.5.197.84:36059&alias=e0000.chtc.wisc.edu>
040 (9834570.074.000) 2021-05-20 12:03:52 Finished transferring input files
```

- Output transfer:

```
040 (9834570.074.000) 2021-05-20 12:04:02 Started transferring output files
...
040 (9834570.074.000) 2021-05-20 12:04:02 Finished transferring output files
```

I'd like to do more
with my self-
checkpointing jobs.



Checkpointing Job Support

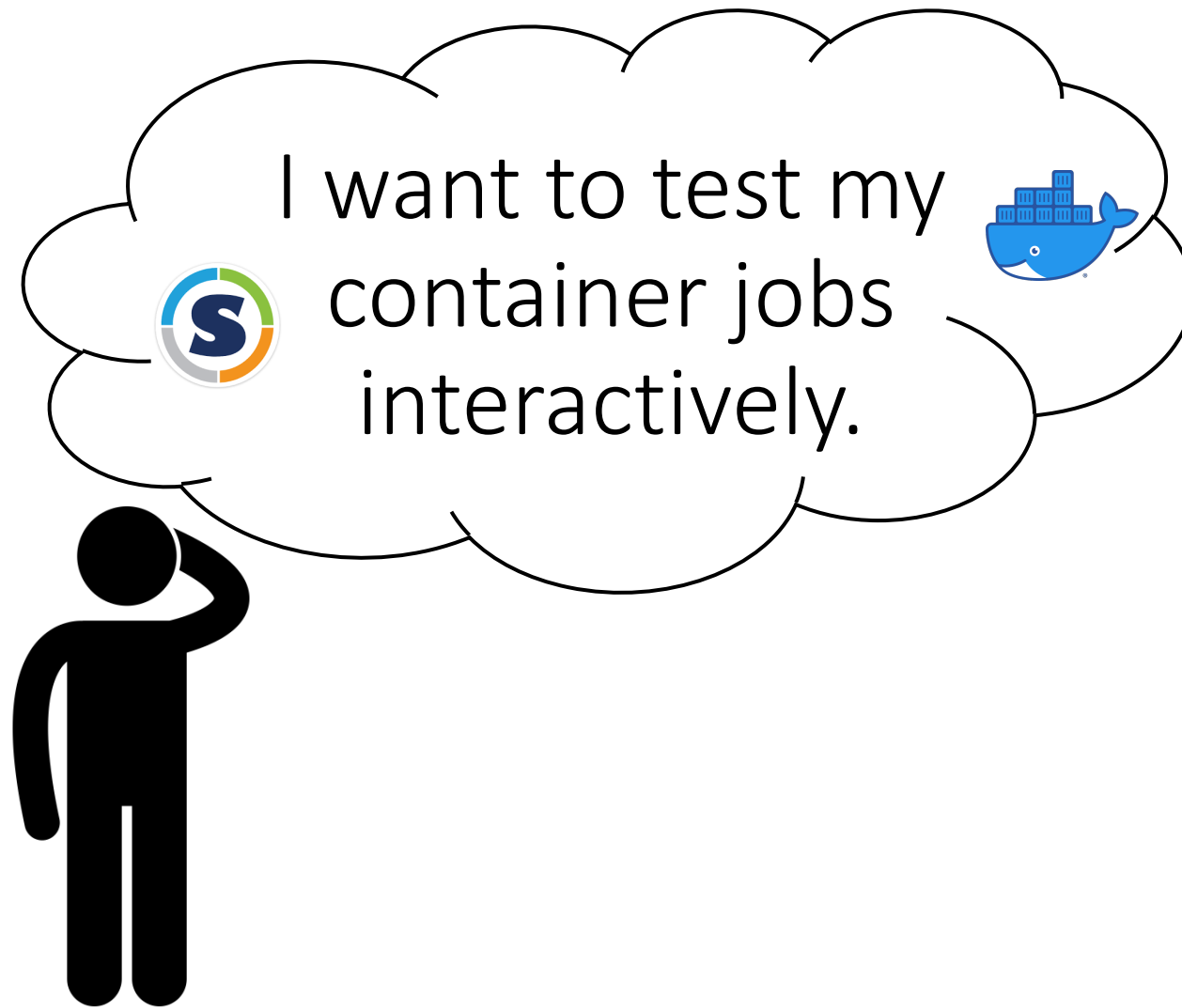
- Force jobs to transfer their checkpoint files at specific intervals by exiting with a pre-determined exit code.

```
checkpoint_exit_code = 85  
transfer_checkpoint_files = example.checkpoint,sandbox.files
```

- Jobs will automatically remain in the queue and keep running after exiting with this code, until they finish and exit normally.
 - Full details in the manual: <https://htcondor.readthedocs.io/en/latest/users-manual/self-checkpointing-applications.html>

Checkpointing Job Support

- New tool: `condor_evicted_files`
 - Use to examine the SPOOL (job sandbox files) of a previously evicted job.
 - Manual page: https://htcondor.readthedocs.io/en/latest/manual-pages/condor_evicted_files.html

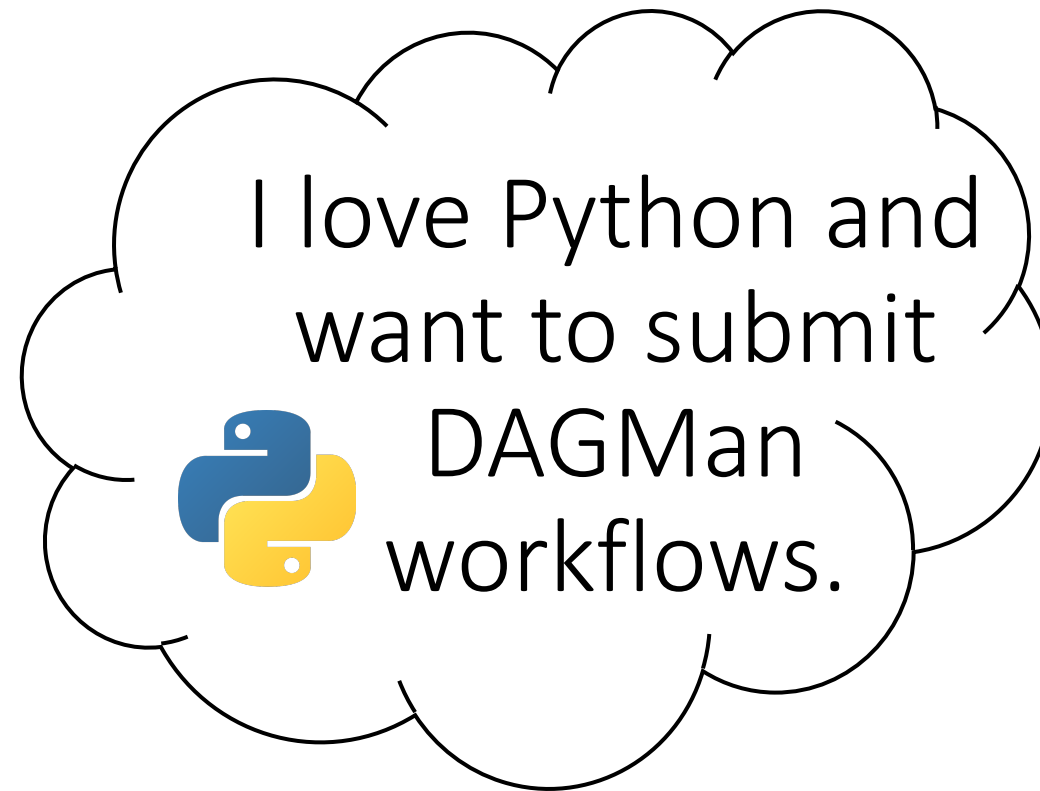
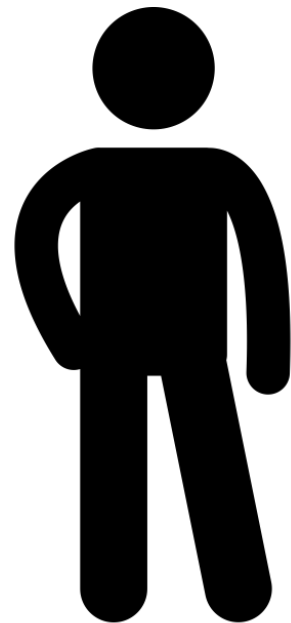


Container Jobs

- Reminder: Jobs can be containers.
 - Example submit file:

```
universe = docker
docker_image = debian:slim
```

- Can now run container jobs interactively.
 - Start a container interactively: `condor_submit -i submit.file`
 - Connect to a running container: `condor_ssh_to_job JobID`



HTCondor Python Bindings

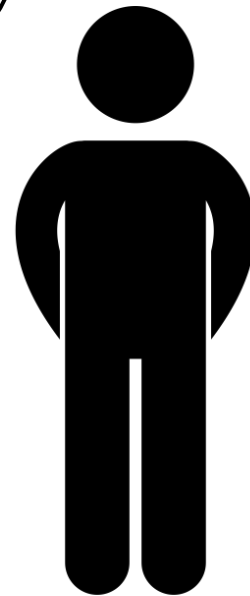
- Many improvements to the Python bindings, including new bindings for DAGMan and chirp.
- Now supports Python 3.x.
- Resources
 - [htcondor.htchirp in the manual](#)
 - [htcondor.dags in the manual](#)
 - [Python Bindings Tutorials](#)

```
dag = dags.DAG()

tile_layer = dag.layer(
    name = 'tile',
    submit_description = tile_description,
    vars = tile_vars,
)

montage_layer = tile_layer.child_layer(
    name = 'montage',
    submit_description =montage_description,
)
```


I use DAGMan and
want to use one file
to define my
workflow.



DAGMan Workflows

- Job submit descriptions can now be included inline in the DAG file, removing the need for an additional submit file.

```
JOB HelloWorld {  
    executable = /bin/echo  
    arguments = "Hello, world!"  
    output = helloworld.out  
}
```

- Coming soon: use the same inline submit description across multiple JOB nodes.

DAGMan Workflows

- DAGMan can now optionally run a script when a job goes on hold.
 - For example: email when jobs go on hold
 - Coming soon: DAGs will be able to submit jobs on hold.
- You may now change some DAGMan throttles while the DAG is running.

Questions

Acknowledgements

This work is supported by [NSF](#) under Cooperative Agreement [OAC-2030508](#) as part of the [PATH Project](#).