HTCondor ~~Architecture~~ Vocabulary

Christina Koch

HTCondor Week @ Throughput Computing 2023

July 12, 2023

Did you know the manual has a glossary?
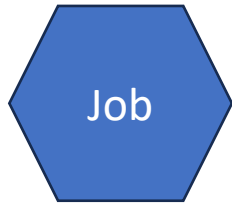
https://htcondor.readthedocs.io/en/latest/codes-other-values/glossary.html
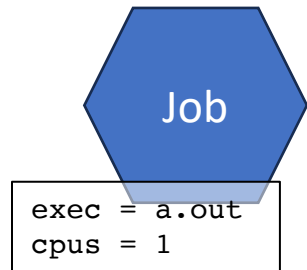
Job

What is a job?

Job

"Atomic unit of work"
One executable (script, binary)
Like y = f(x), transforms inputs into outputs

How does HTCondor describe jobs?

ClassAd

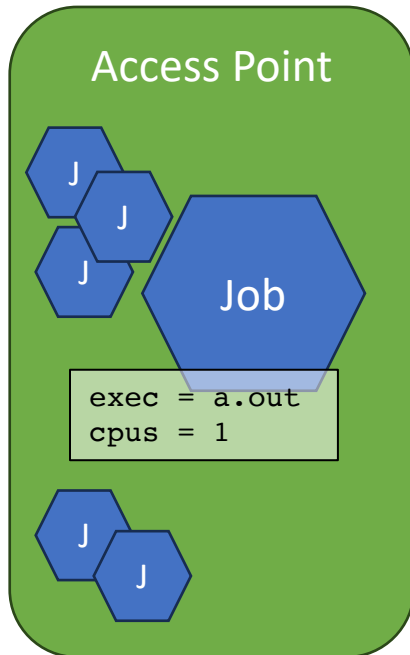**What is a ClassAd?**

Job

```
exec = a.out
cpus = 1
```

"Key-value pairs"

Metadata for a job (or other HTCondor object)

Values can be Booleans, numbers, strings, or expressions

Can add custom keys/values
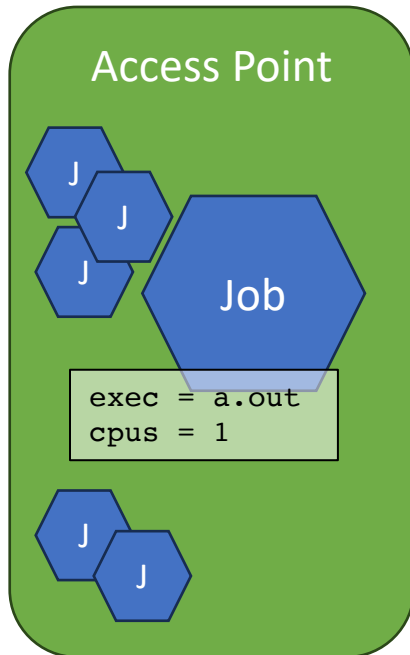
Access Point (AP)

What is an Access Point (AP)?

"machine where users place jobs to be run"
The home for jobs
Often where users log in
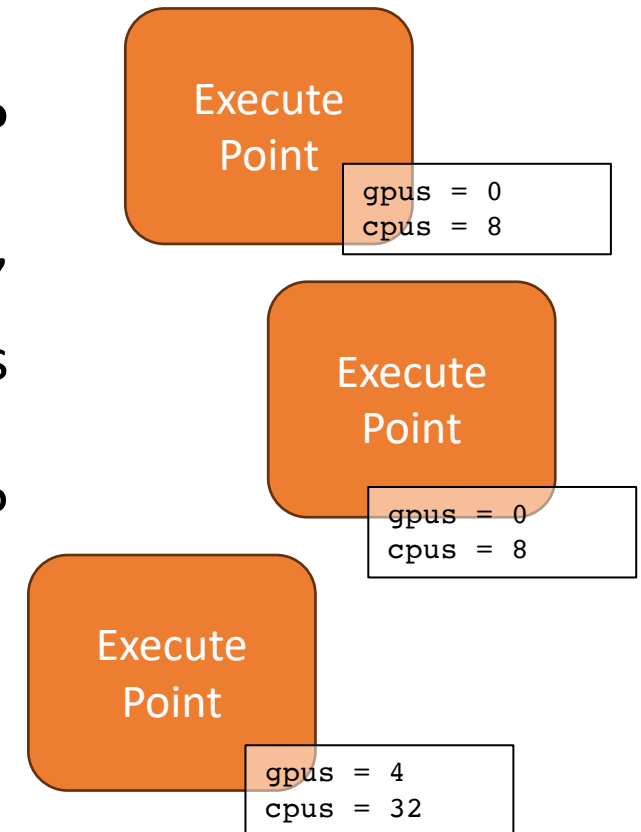
Execution Point (EP)

Access Point

J
J
J

Job

```
exec = a.out
cpus = 1
```

J
J

What is an Execution Point (EP)?

"where jobs run"
ALSO described using ClassAds

…how do the APs and EPs talk to each other?

Execute Point

```
gpus = 0
cpus = 8
```

Execute Point

```
gpus = 0
cpus = 8
```

Execute Point

```
gpus = 4
cpus = 32
```

Daemons

condor_schedd

Access Point
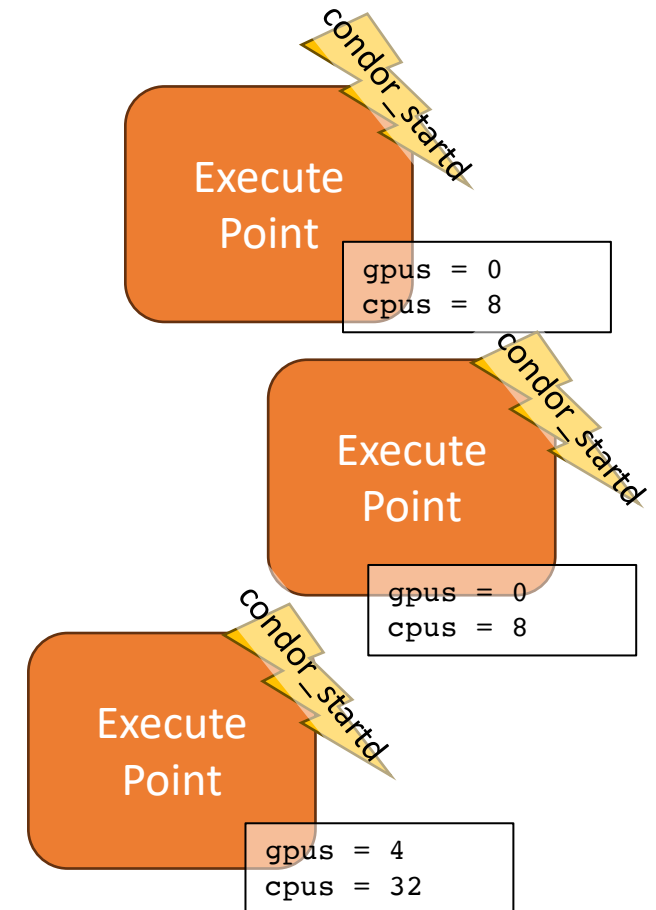
J
J
J

Job

```
exec = a.out
cpus = 1
```
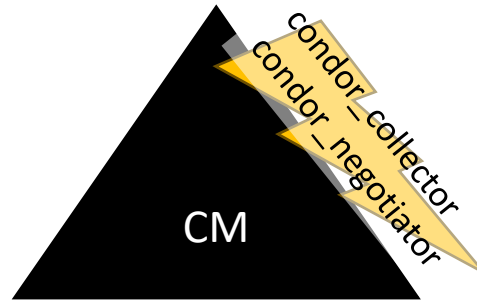
J
J
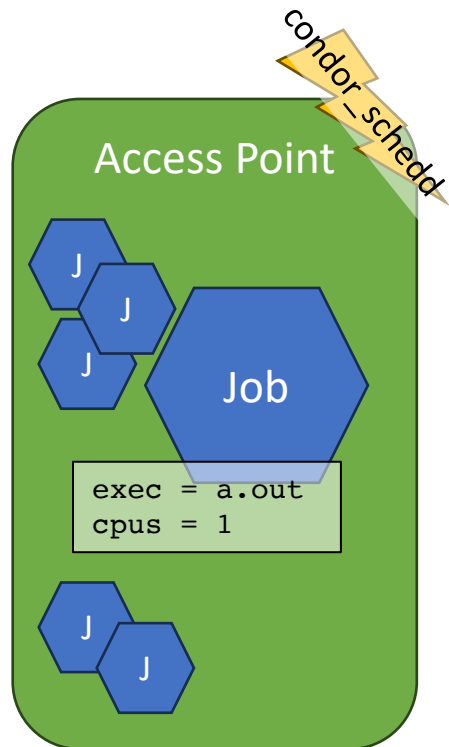
First…What are Daemons?

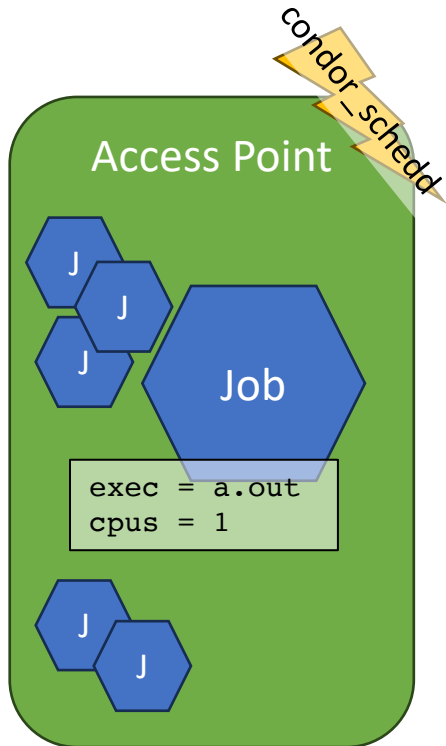"A long-running process; an older term for 'service.'"

HTCondor entities (AP, EP, etc.) are associated with a particular daemon

Now, "how do APs and EPs find each other?"
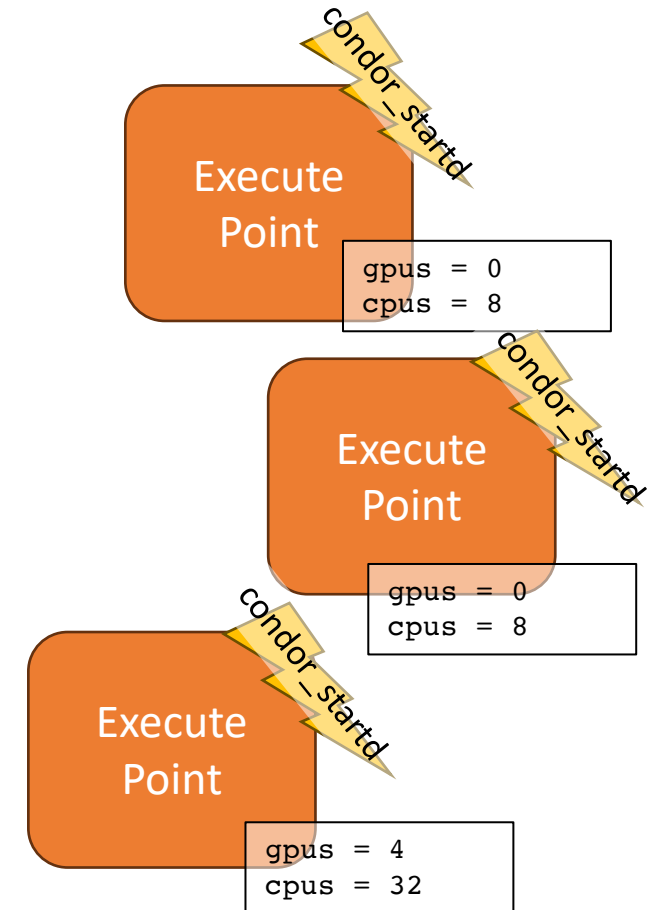
condor_startd

Execute Point

```
gpus = 0
cpus = 8
```

condor_startd

Execute Point

```
gpus = 0
cpus = 8
```

condor_startd

Execute Point

```
gpus = 4
cpus = 32
```

CHTC   HTCondor Software Suite   PATh

Central Manager (CM)

Access Point

J
J
J

Job

```
exec = a.out
cpus = 1
```

J
J

condor_schedd

condor_collector
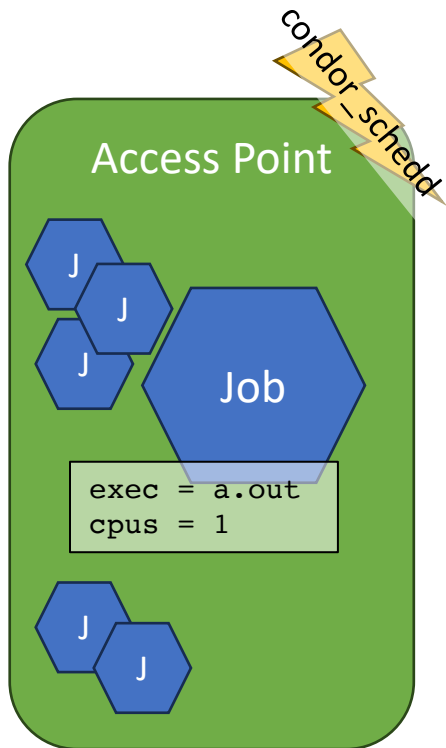condor_negotiator

CM

What is the Central Manager?
"machine with central database of all the services, an accountant and negotiator"

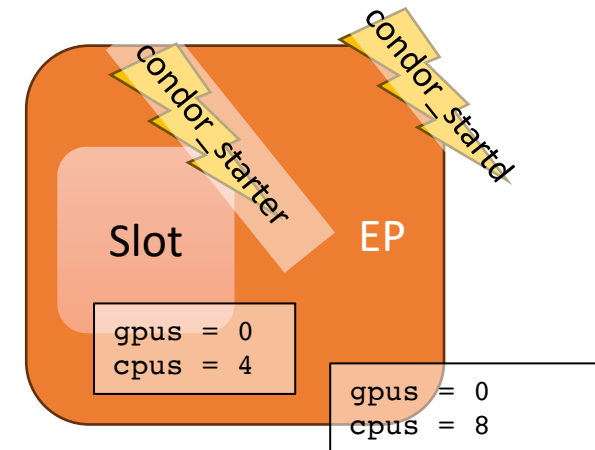It is responsible for "matching" jobs on the AP with execute points

Execute Point

condor_startd

```
gpus = 0
cpus = 8
```

Execute Point

condor_startd

```
gpus = 0
cpus = 8
```

Execute Point

condor_startd

```
gpus = 4
cpus = 32
```

Slot

condor_schedd

Access Point
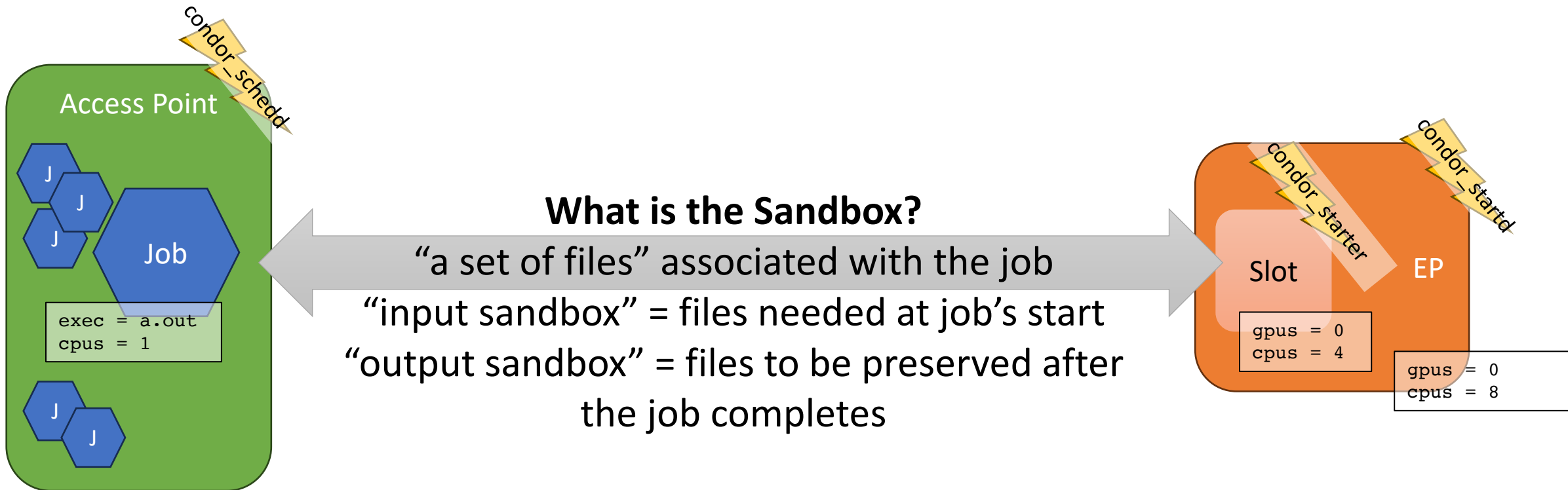
J
J
J

Job

```
exec = a.out
cpus = 1
```

J
J

What is a Slot?
"the location on the EP where the job runs"
Think of it as a "slice" (or sometimes the entirety!) of a server/machine.

Created by the *condor_starter* daemon

condor_starter

condor_startd

Slot

EP

```
gpus = 0
cpus = 4
```

```
gpus = 0
cpus = 8
```

CHTC

HTCondor
Software Suite

PATh

Sandbox

Access Point

condor_schedd

J
J
J
Job

```
exec = a.out
cpus = 1
```

J
J

What is the Sandbox?
"a set of files" associated with the job
"input sandbox" = files needed at job's start
"output sandbox" = files to be preserved after the job completes

condor_startd

condor_starter

Slot

EP

```
gpus = 0
cpus = 4
```

```
gpus = 0
cpus = 8
```

Learn More

- Architecture
  - https://indico.cern.ch/event/936993/contributions/4022092/
- ClassAds
  - https://indico.cern.ch/event/936993/contributions/4022129/
- User Tutorial
  - https://agenda.hep.wisc.edu/event/1733/contributions/25486/
- HTCondor Philosophy
  - https://agenda.hep.wisc.edu/event/1579/contributions/23050/

What's Missing?

Acknowledgements