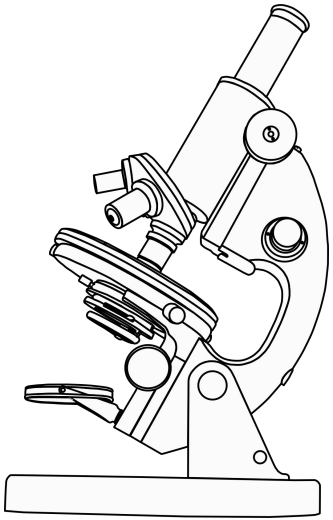
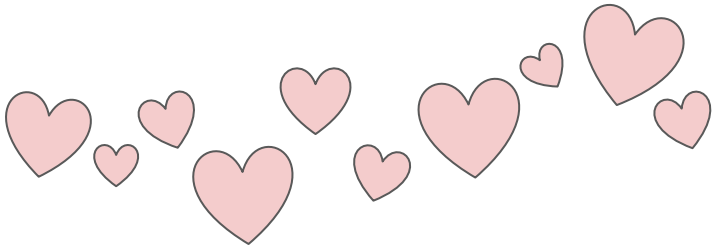
Connecting Resources with Science via HTCondor-CE

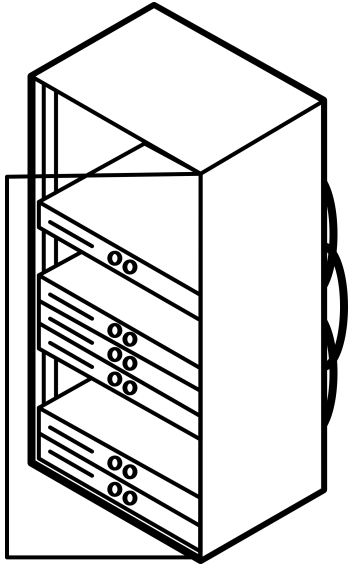
Brian Lin

OSG All Hands 2017

A fundamental problem of scientific computing at scale is matchmaking

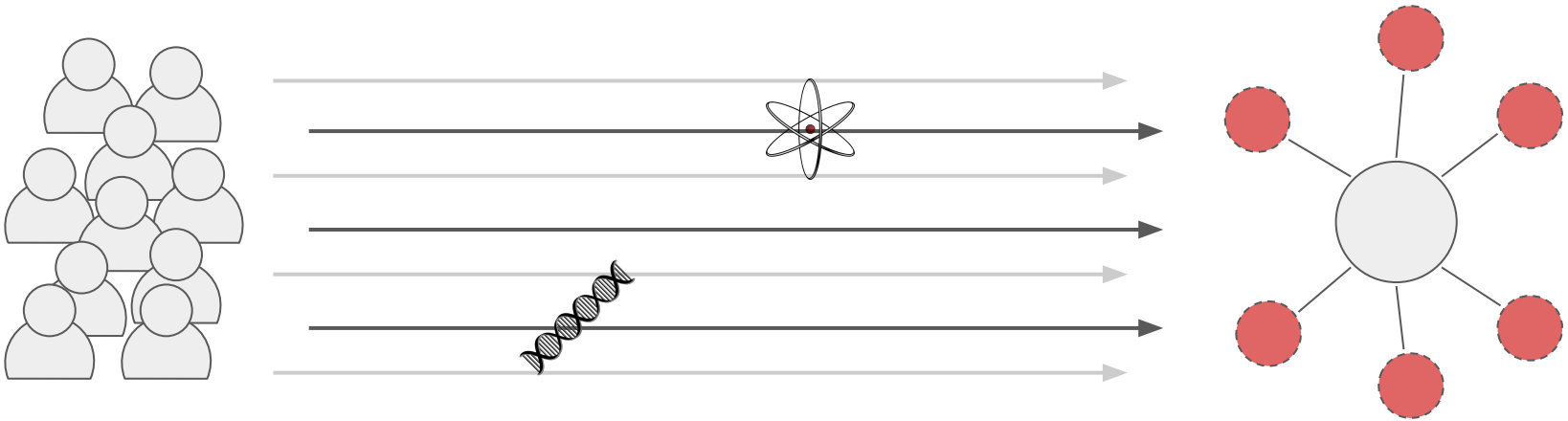




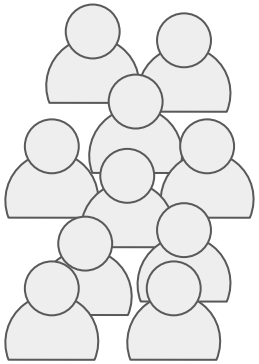


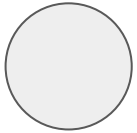
Managing Scale

Managing Scale



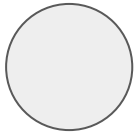
The OSG Model

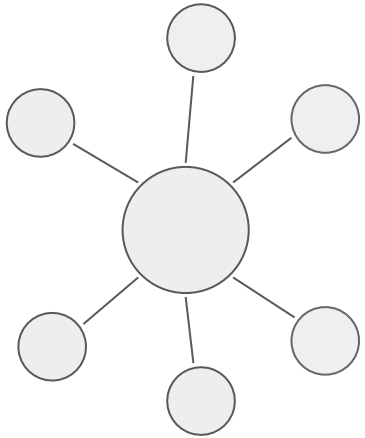




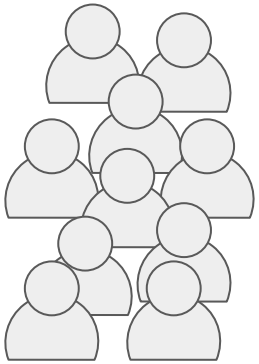
User Submit

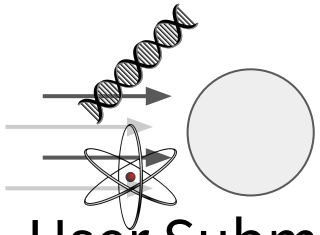
Site Gateway





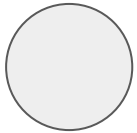
The OSG Model

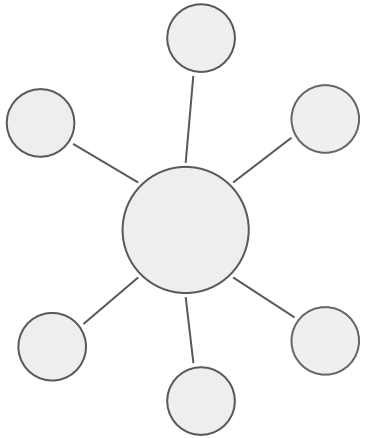




User Submit

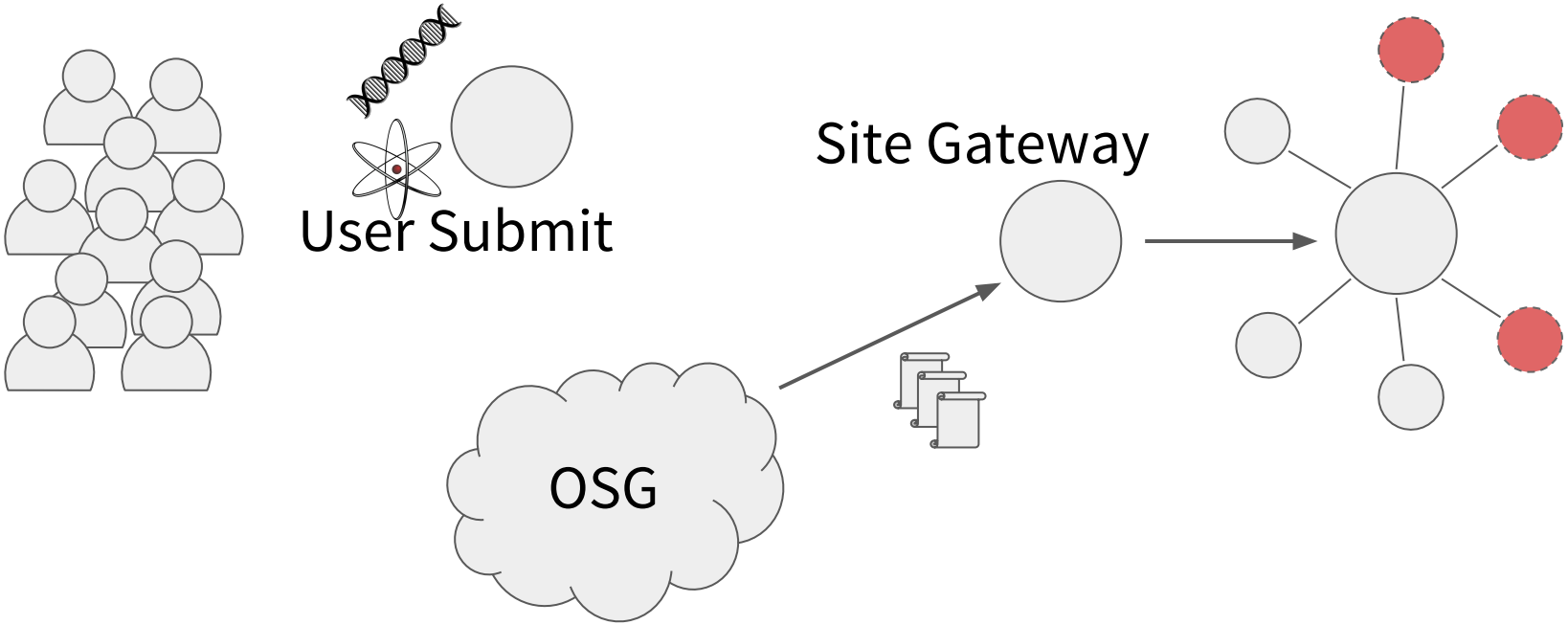
Site Gateway



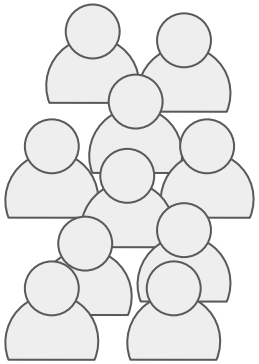


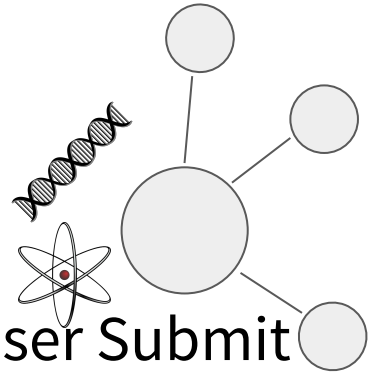
OSG

The OSG Model



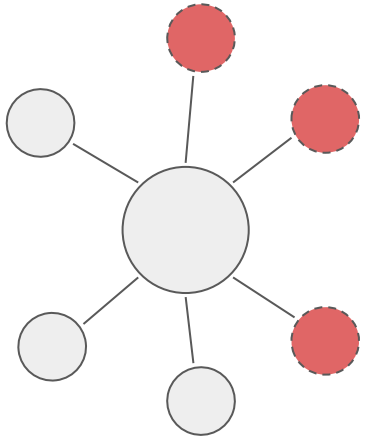
The OSG Model





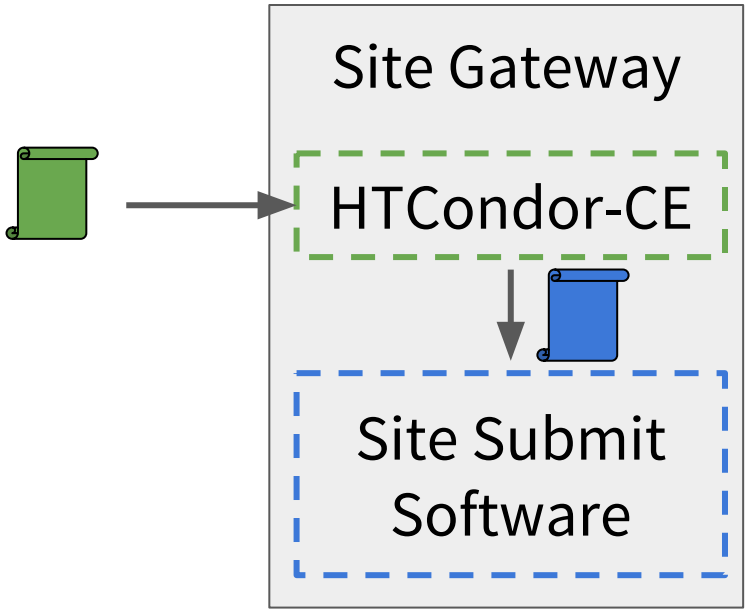
User Submit

Site Gateway

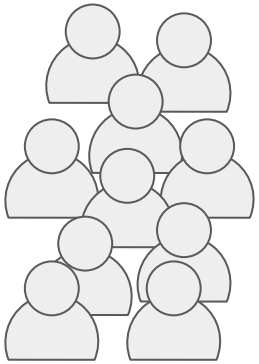


HTCondor-CE: Site Gateway

- Site gateway = HTCondor-CE on batch system submit host
- OSG entry point for pilot jobs
- Filter and transform incoming jobs for compatibility with site policy
- Based on core HTCondor features

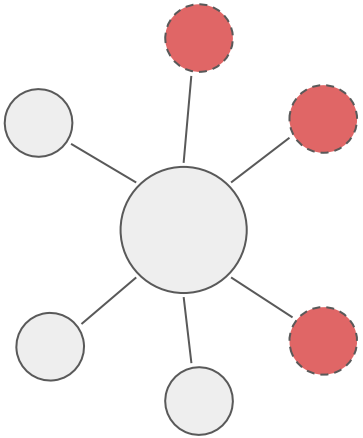


The OSG Model: HTCondor-based



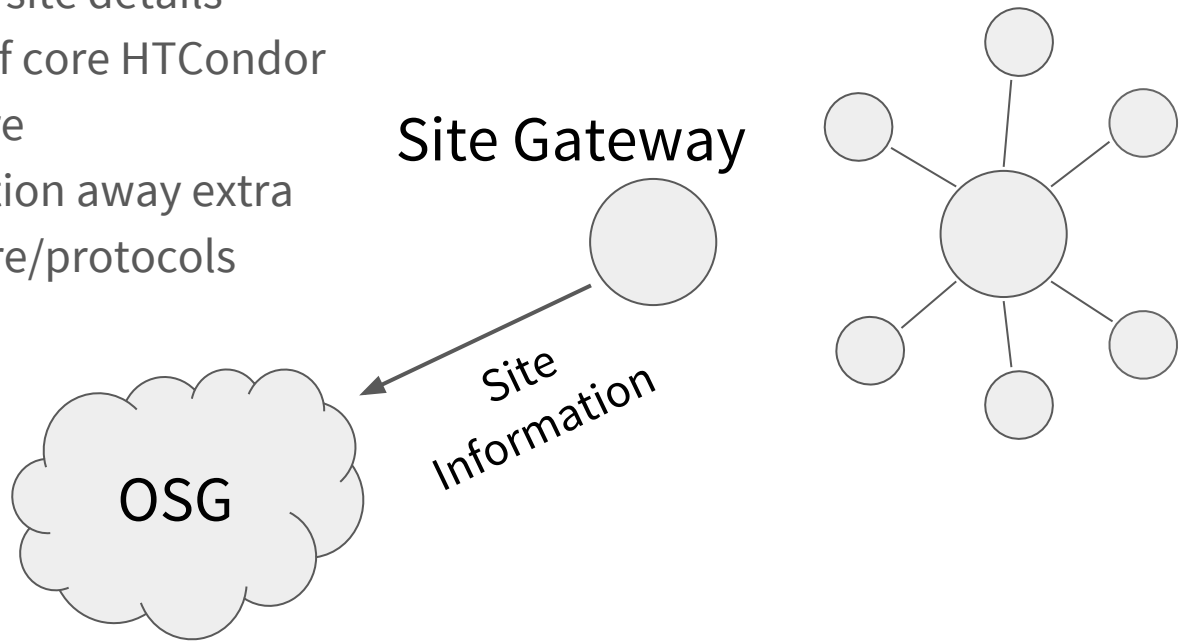
Site Gateway

HTCondor

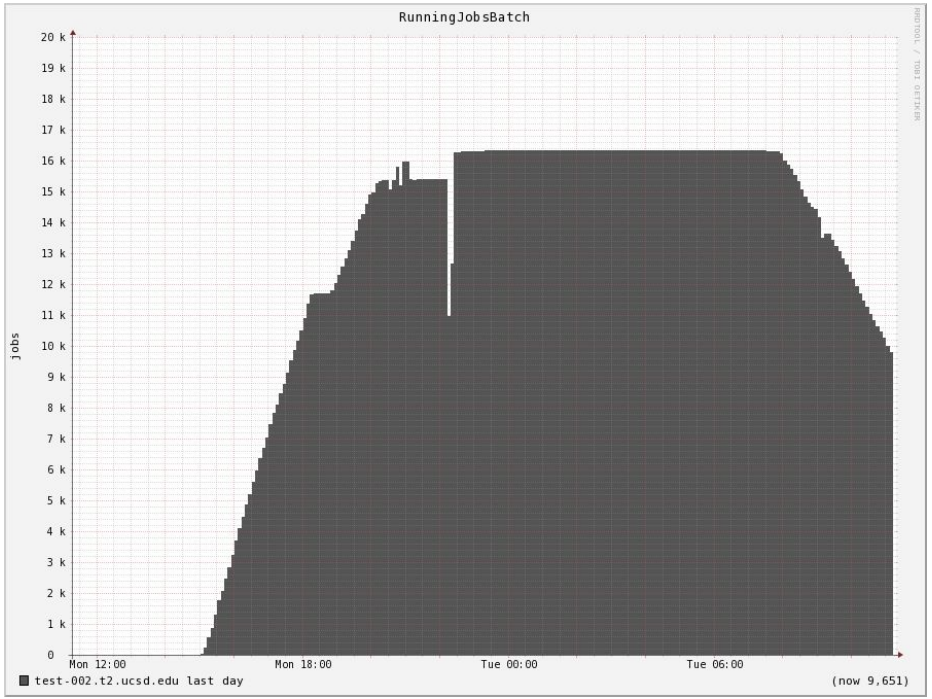


HTCondor-CE: Central Collector

- Central storage for site details
- Takes advantage of core HTCondor ‘advertising’ feature
- Allows us to transition away extra supporting software/protocols



HTCondor-CE: Scalable



- Benefit from HTCondor scale improvements
- Last round of scale tests by Edgar in 2015
- 16k* jobs, 2 ports per-job with a start-up rate of 70 jobs/min
- Scales horizontally!

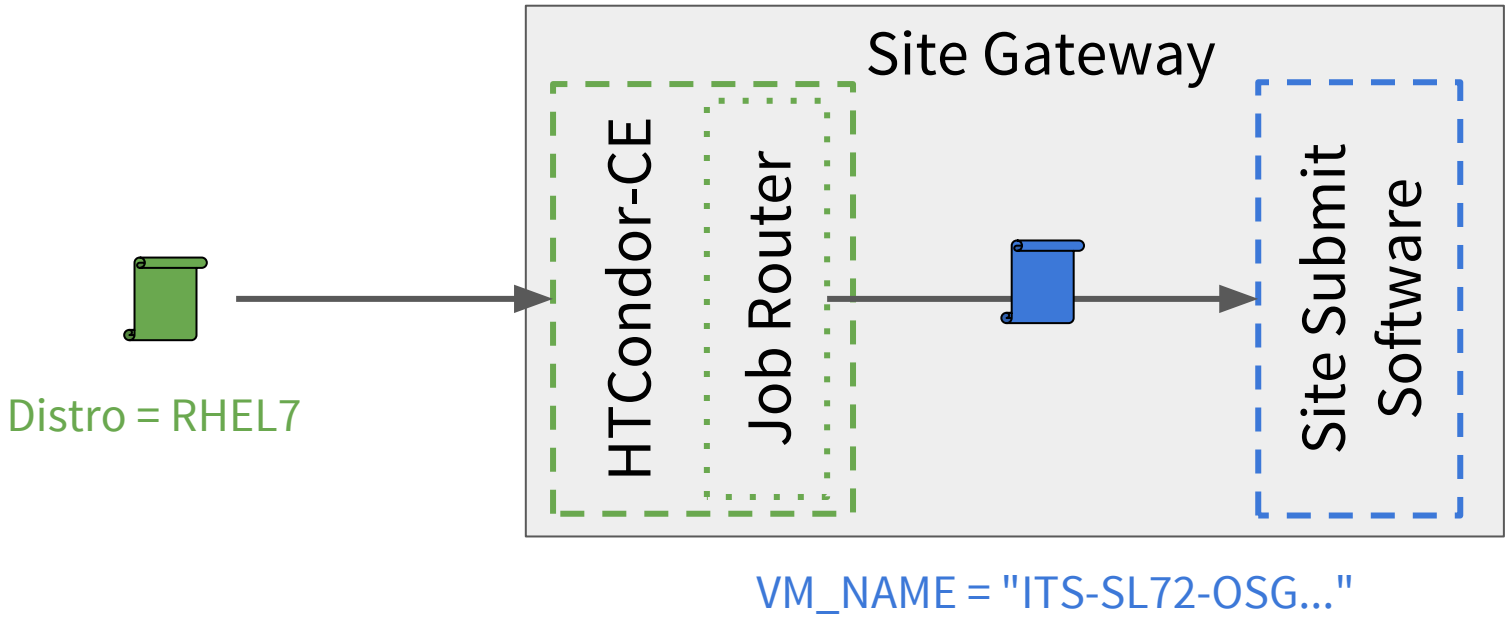
* bottlenecked by the backend cluster

HTCondor-CE: In the Wild

Site	Cluster Type	Site Policy
Vanderbilt	Slurm	Stakeholder jobs run in preferred Slurm partitions; incoming jobs modified to accommodate hyper-threading
Purdue	HTCondor	Avoid subclusters that can't run OSG jobs
	PBS	Set PBS accounting group based on job submitter
Nebraska	Slurm	GPU jobs should run under a separate Slurm partition
	HTCondor	Jobs need to run inside Docker containers
Syracuse	HTCondor	Jobs run under custom VM infrastructure
Langston University	HTCondor	Separate cluster for specific OSG jobs via chained CEs!

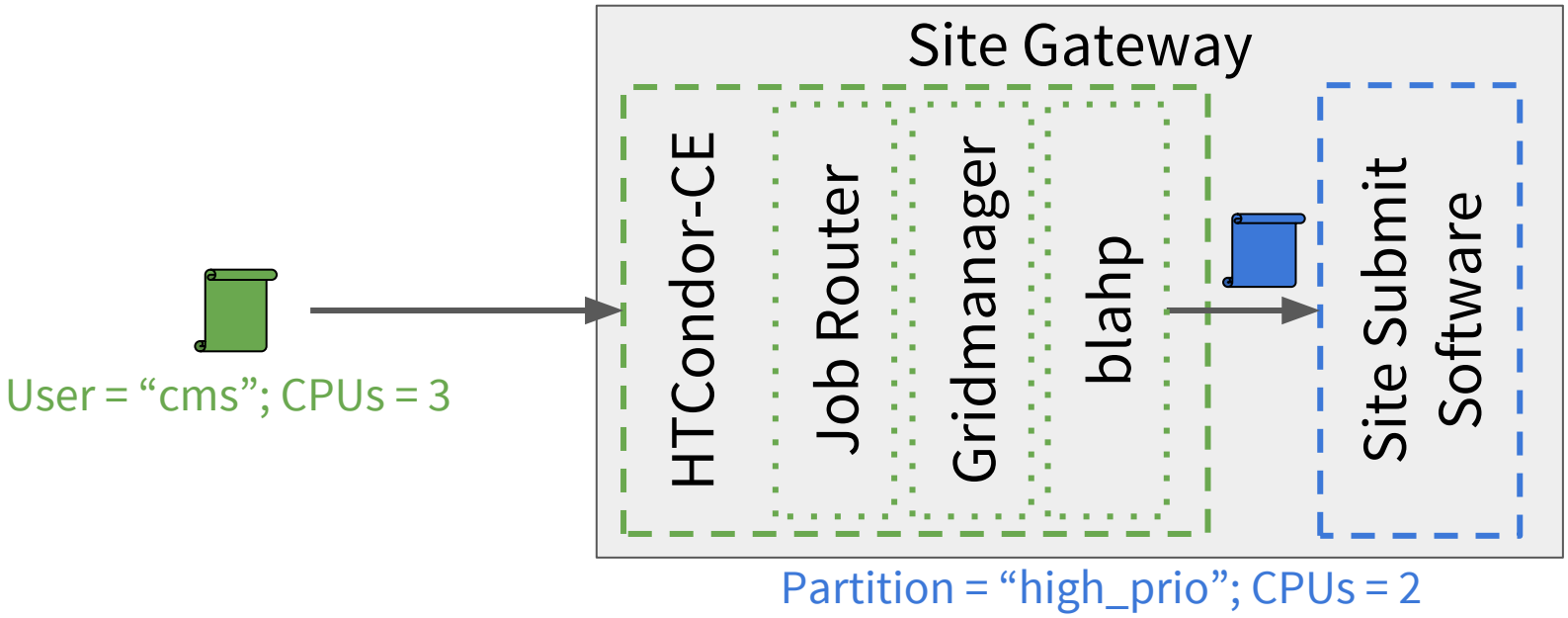
HTCondor-CE: Job Router, HTCondor backend

Syracuse	HTCondor	Jobs run under custom VM infrastructure



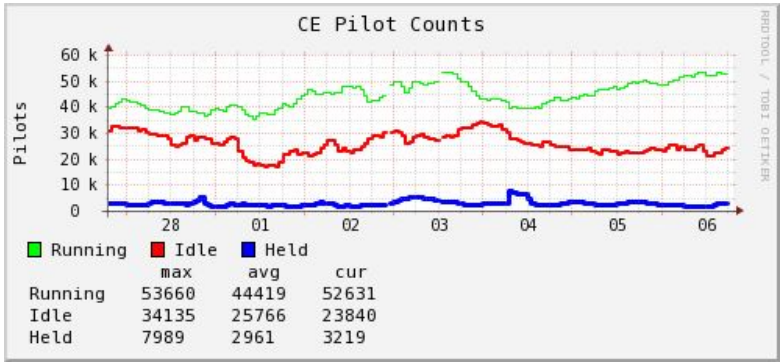
HTCondor-CE: Job Router, non-HTCondor backend

Vanderbilt	Slurm	Stakeholder jobs run in preferred Slurm partitions; incoming jobs modified to accommodate hyper-threading



HTCondor-CE: Looking Forward

- We have pilot job tracking and introspection
- Missing easy payload job introspection and history




```
blin@cs-wifiopen-92:~$ condor_status
Name                                                OpSys      Arch   State     Activity LoadAv Mem     ActvtyTime

slot1@glidein_1_174501201@cmsprod-405273.0-red-c5713.unl.edu LINUX      X86_64 Unclaimed Idle      0.120 248952  2+12:22:56
slot1_51@glidein_1_174501201@cmsprod-405273.0-red-c5713.unl.edu LINUX      X86_64 Claimed   Busy      0.000   2048  1+18:58:00
slot1@glidein_1_173601603@cmsprod-406049.0-red-c5615.unl.edu    LINUX      X86_64 Unclaimed Idle      0.230 136312  0+07:48:00
slot1_1@glidein_1_173601603@cmsprod-406049.0-red-c5615.unl.edu  LINUX      X86_64 Claimed   Busy      1.010   2048  0+07:45:27
slot1_2@glidein_1_173601603@cmsprod-406049.0-red-c5615.unl.edu  LINUX      X86_64 Claimed   Busy      1.000   2048  0+07:45:27
slot1_3@glidein_1_173601603@cmsprod-406049.0-red-c5615.unl.edu  LINUX      X86_64 Claimed   Busy      1.000   2048  0+07:45:25
slot1_4@glidein_1_173601603@cmsprod-406049.0-red-c5615.unl.edu  LINUX      X86_64 Claimed   Busy      1.000   2048  0+07:45:25
slot1_5@glidein_1_173601603@cmsprod-406049.0-red-c5615.unl.edu  LINUX      X86_64 Claimed   Busy      1.010   2048  0+07:45:24
```

HTCondor-CE: Summary

Pros

- Public, uniform job entry point
- Scalable
- Site-local, flexible configuration

Cons

- Site-local, flexible configuration
- Administrative overhead

Not ready to run your own HTCondor-CE?

See next talk on OSG-hosted CEs!

Site Admin Sessions

Office Hours - Thursday @ 9 AM

Site Installation Overview - Thursday @ 11 AM

Questions?