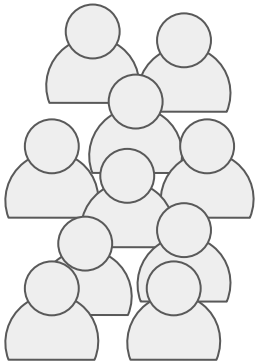
# HTCondor-CE: Basics and Architecture

ISGC 2019 - Taipei, Taiwan

Brian Lin
University of Wisconsin — Madison

# The Pilot Overlay Model

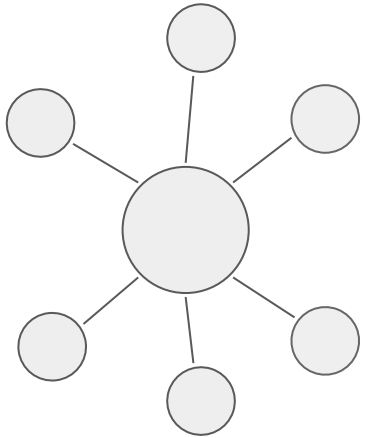


User Submit

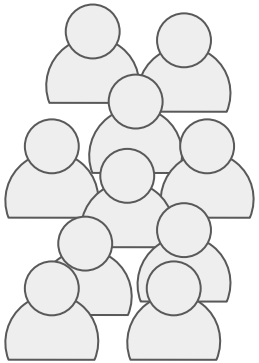
Pilot Factory

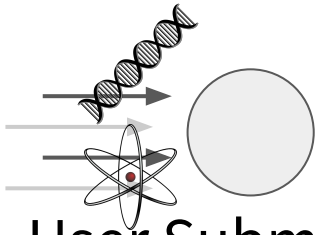
Compute Element

Local Batch System



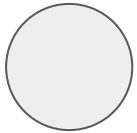
# The Pilot Overlay Model



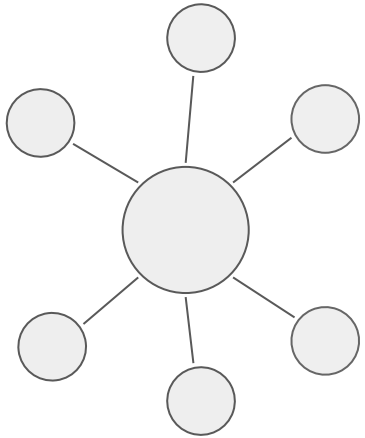


User Submit

Compute Element

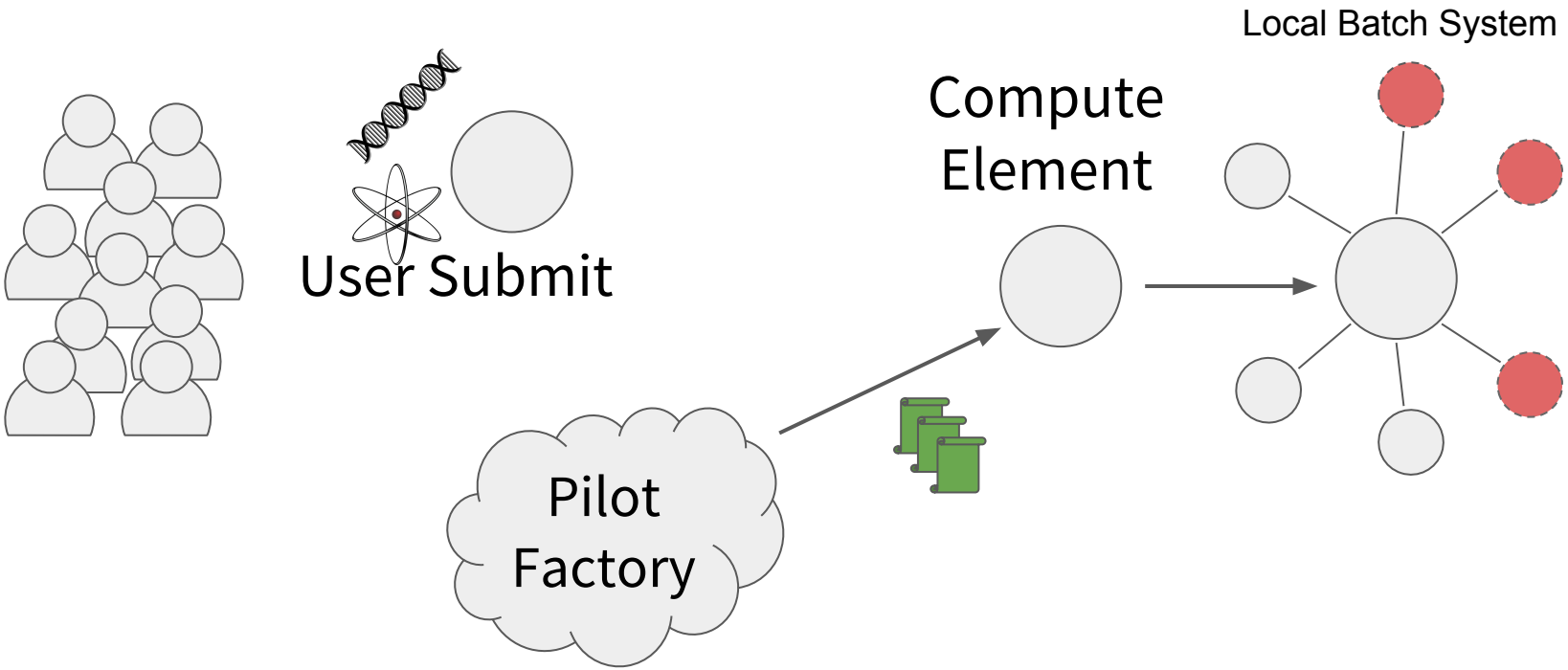


Local Batch System

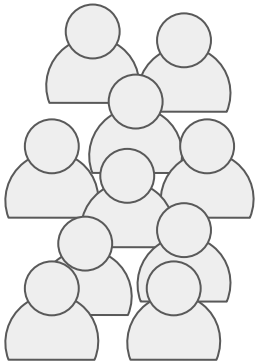


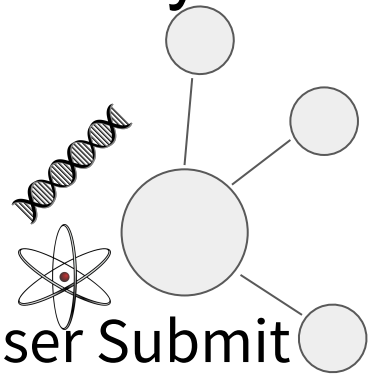
Pilot Factory

# The Pilot Overlay Model



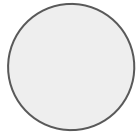
# The Pilot Overlay Model



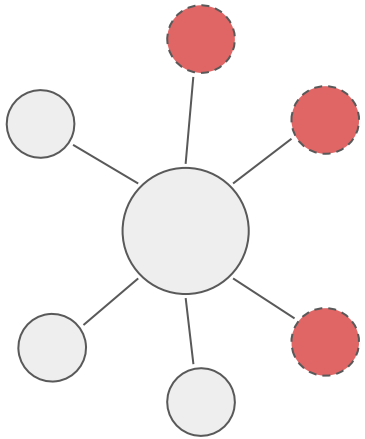


User Submit

Compute Element



Local Batch System

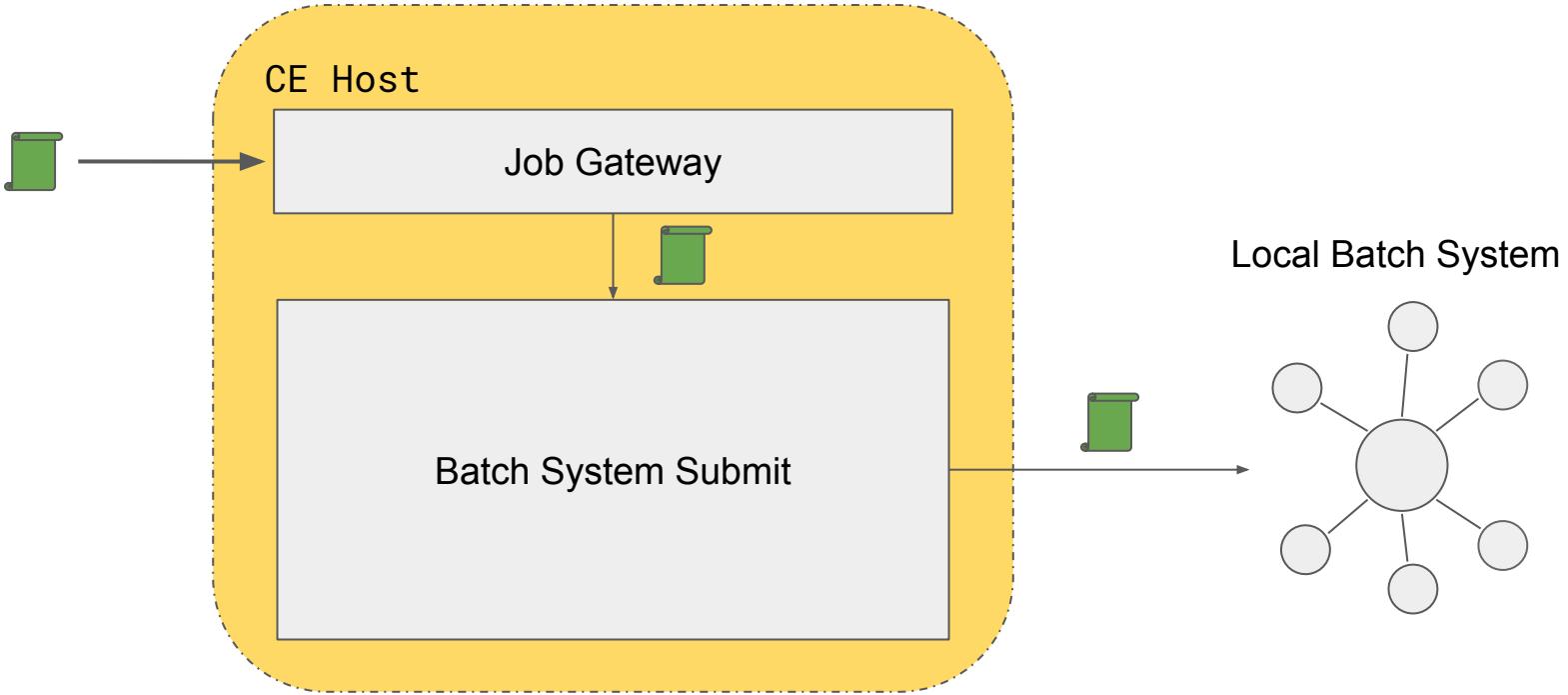


Pilot Factory

# What's a CE?

- A compute element (CE) serves as the entry point to your local compute resources
  - Exposes a **remote API** for resource acquisition
  - Provides authentication and **authorization**
  - Interacts with the **resource layer** (i.e. batch system)
- A CE is made up of a thin layer of **job gateway** software installed on a host that can submit and manage jobs in your local batch system
- Designed to support the pilot job overlay model (i.e. resource provisioning requests) and is generally not intended for direct user submission

# Compute Element Architecture



# HTCondor as a Job Gateway

HTCondor-CE is HTCondor configured as a job gateway

- Same HTCondor binaries, ClassAds, and configuration language to provide the **remote API**
- Relevant tools wrapped to use the HTCondor-CE configuration (e.g., `condor_ce_q`, `condor_ce_status`, etc.)
- Separate `condor_master` and service (`condor-ce`)
- Provides GSI authentication and uses HTCondor security for **authorization**
- Supports interaction with the following **resource layers**...
  - HTCondor batch systems directly
  - Slurm, PBS, SGE, and LSF batch systems via the Gridmanager daemon
  - Also with the above batch systems over SSH
- Available via HTCondor or OSG Yum repositories

# Job Router

- The Job Router is responsible for taking a job, creating a copy, and changing the copy according to a set of rules
  - When running an HTCondor batch system, the copy is inserted directly into the site batch schedd
  - Each chain of rules is called a “route” and is defined by a ClassAd
  - This set of rules reflects a site’s policy
- Once the copy has been created, attribute changes and state changes are propagated between the source and destination jobs

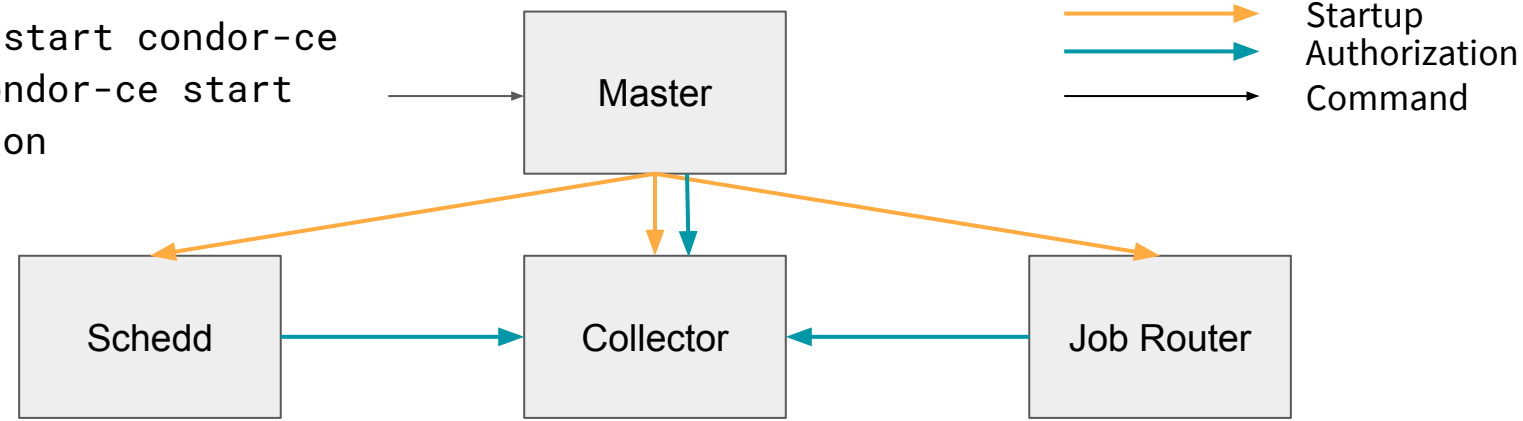
# HTCondor-CE + HTCondor Batch System

- Two sets of HTCondor daemons
  - Two sets of configuration: `/etc/condor-ce/config.d/` and `/etc/condor/config.d/`
  - Two sets of logs: `/var/log/condor-ce/` and `/var/log/condor/`
- Note the lack of the `condor_negotiator` for the CE set of daemons. HTCondor-CE doesn't manage any worker nodes so it doesn't need to do matchmaking!

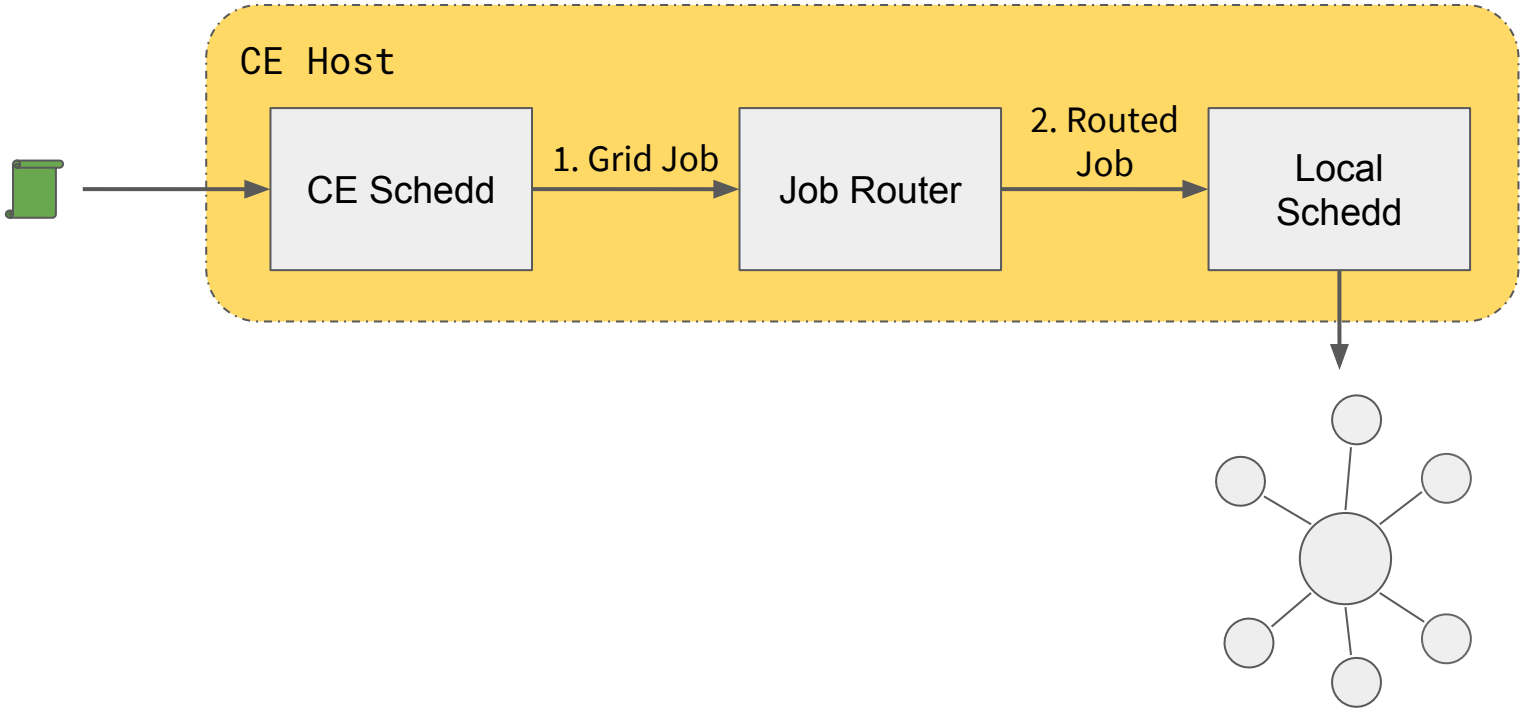
```
# pstree
[...]
  ├─condor_master─┬─condor_collector
  |               ├─condor_negotiator
  |               ├─condor_procd
  |               ├─condor_schedd
  |               ├─condor_shared_port
  |               └─condor_startd
  ├─condor_master─┬─condor_collector
  |               ├─condor_job_router
  |               ├─condor_procd
  |               ├─condor_schedd
  |               └─condor_shared_port
  |
[...]
```

# HTCondor-CE Daemons

```
systemctl start condor-ce
service condor-ce start
condor_ce_on
```



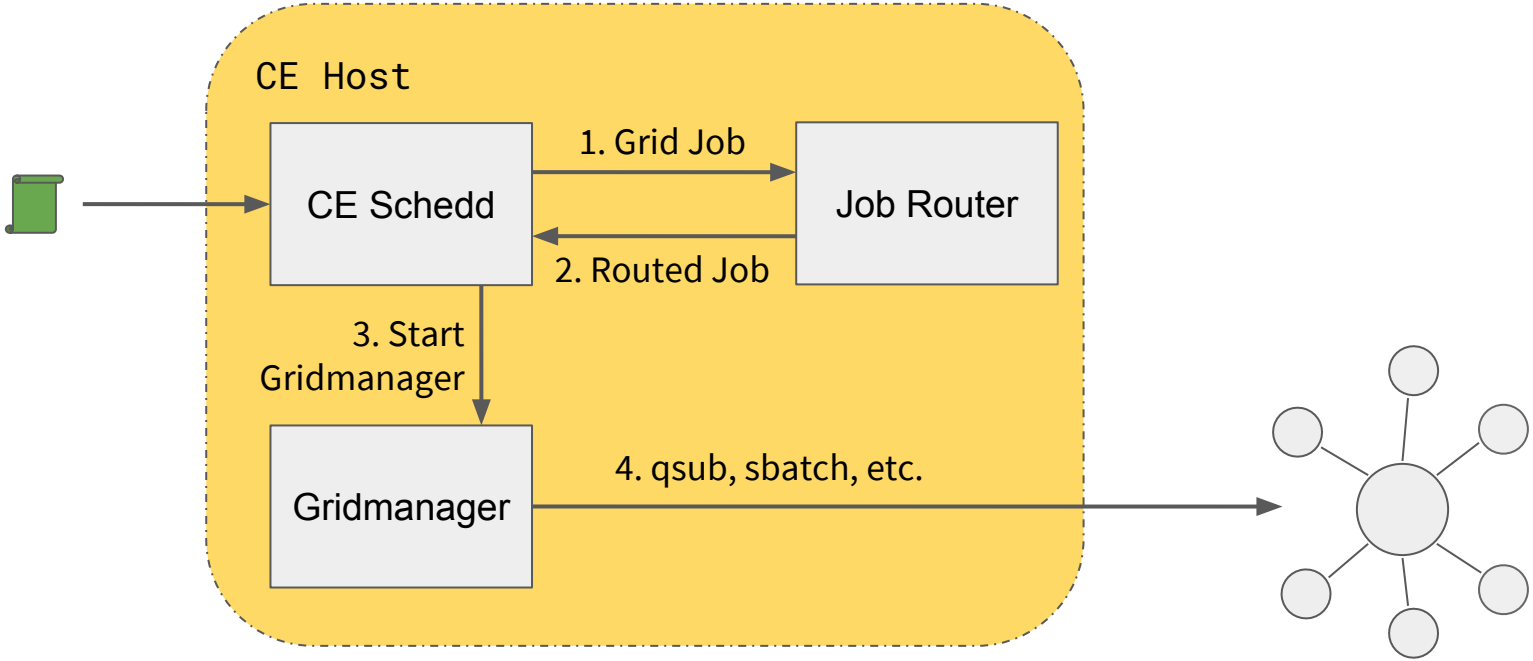
# HTCondor-CE + HTCondor Batch System



# HTCondor-CE + Non-HTCondor Batch System

- Since there is no local batch system schedd, jobs are routed back into the CE schedd as “Grid Universe” jobs
- Grid universe jobs spawn a Gridmanager daemon per user with log files: `/var/log/condor-ce/GridmanagerLog.<user>`
- Requires a shared filesystem across the cluster for pilot job file transfers

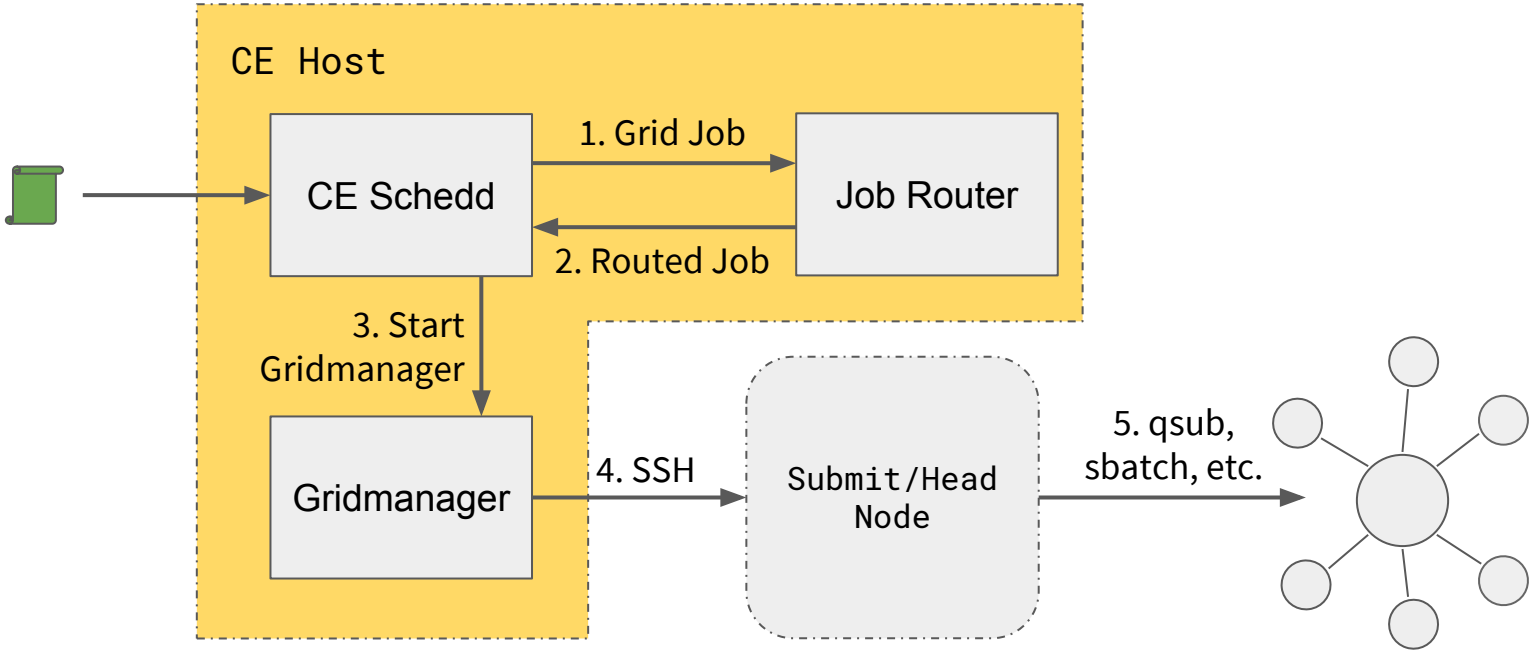
# HTCondor-CE + Non-HTCondor Batch System



# HTCondor-CE + SSH

- Using BOSCO (https://osg-bosco.github.io/docs/), HTCondor-CE can be configured to submit jobs over SSH
  - Requires SSH key-based access to an account on a node that can submit and manage jobs on the local batch system
  - Requires shared home directories across the cluster for pilot job file transfer
- The Open Science Grid (OSG) uses HTCondor-CE over SSH to offer the OSG Hosted CE as a service
- Can support up to ~10k jobs concurrently

# HTCondor-CE + SSH



# HTCondor-CE Requirements

- Open port (TCP) 9619
- Shared filesystem for non-HTCondor batch systems for pilot job file transfer
- CA certificates and CRLs installed in `/etc/grid-security/certificates/`
  VO information installed in `/etc/grid-security/vomsdir/`
- Ensure mapped users exist on the CE (and across the cluster)
- Minimal hardware requirements
  - Handful of cores
  - HTCondor backends should plan on ~½ MB RAM per job
  - Expecting high rates of jobs? HTCondor-CE `SPOOL` dir should live on an SSD
    Default: `/var/lib/condor-ce/spool` (`condor_ce_config_val -v SPOOL`)
- For example, our Hosted CEs run on 2 vCPUs/2GB RAM

# HTCondor-CE Information Services

- Reporting the information required by pilot job factories
  - Contact information (hostname/port)
  - Access policy (authorized virtual organizations)
  - What resources can be accessed?
  - Debugging info (site batch system, site name, versions) for humans
- Each HTCondor-CE in a grid can be configured to report information to a central HTCondor-CE collector
  - Install a central collector via the `htcondor-ce-collector` package
  - Run via the `condor-ce-collector` service
- BDII support for HTCondor batch systems via the `htcondor-ce-bdii` package (supports multiple HTCondor-CEs): https://github.com/opensciencegrid/htcondor-ce/tree/master/contrib/bdii

# Why Consider this CE?

- If you are using HTCondor for batch:
  - One less software provider - same thing all the way down the stack.
  - HTCondor has an extensive feature set - easy to take advantage of it (i.e., Docker universe).
- Regardless, a few advantages:
  - Can scale well (up to at least 16k jobs; maybe higher).
  - Declarative ClassAd-based language.
- But disadvantages exist:
  - Non-HTCondor backends are finicky outside PBS and SLURM.
  - Declarative ClassAd-based language.

# Additional Resources

- Installation Guide
  https://htcondor-wiki.cs.wisc.edu/index.cgi/wiki?p=InstallHtcondorCe
- Overview
  https://opensciencegrid.org/docs/compute-element/htcondor-ce-overview/
- Job Router Configuration Guide
  https://opensciencegrid.org/docs/compute-element/job-router-recipes/
- Troubleshooting Guide
  https://opensciencegrid.org/docs/compute-element/troubleshoot-htcondor-ce/
- Additional Help
  htcondor-users@htcondor.org