

HTCondor-CE

Brian Bockelman, EGI Workshop, May 2019



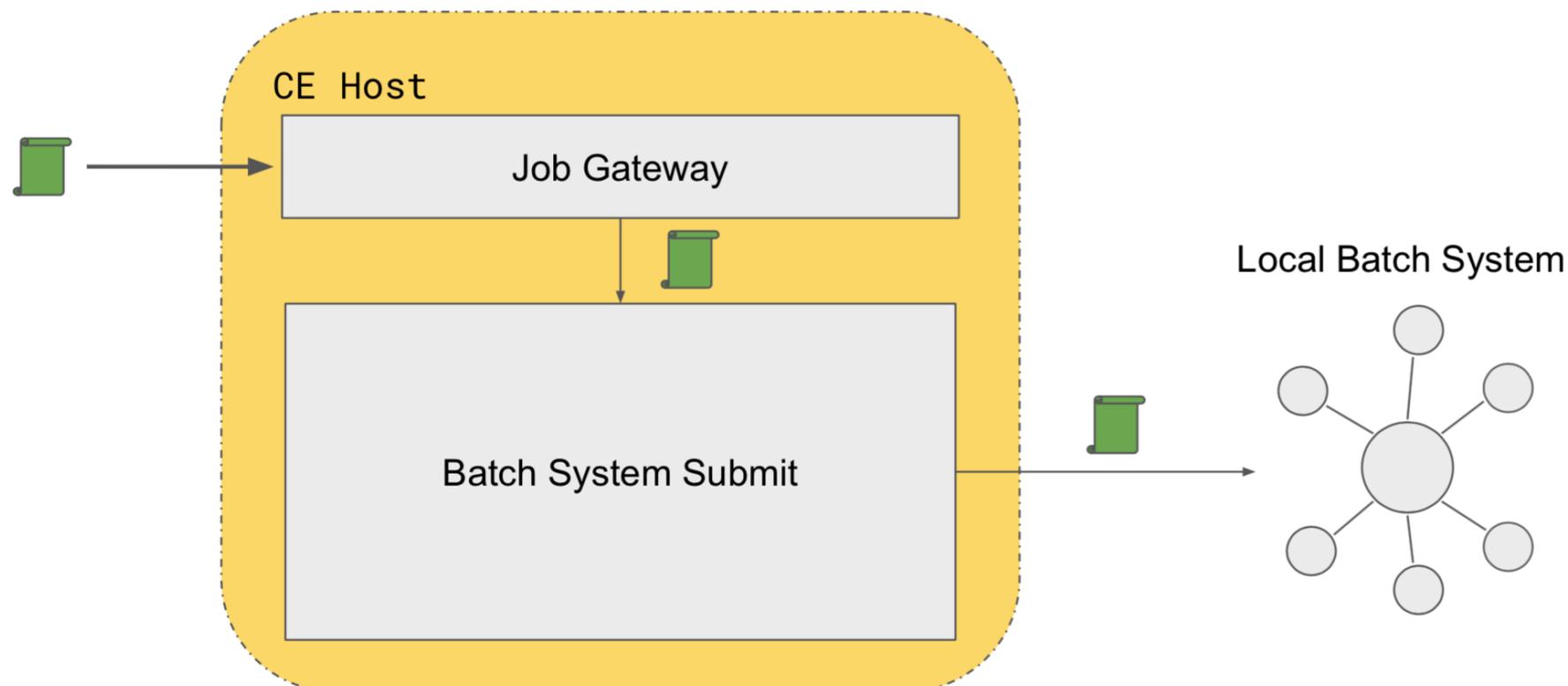
This project is supported by National Science Foundation under Cooperative Agreement OAC-1836650. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

What's in a CE?



MORGRIDGE
INSTITUTE FOR RESEARCH

- A compute element (CE) serves as the entry point to your local compute resources. It:
 - Exposes a **remote API** for resource acquisition
 - Provides **authentication and authorization**
 - Maps the resource acquisition requests 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

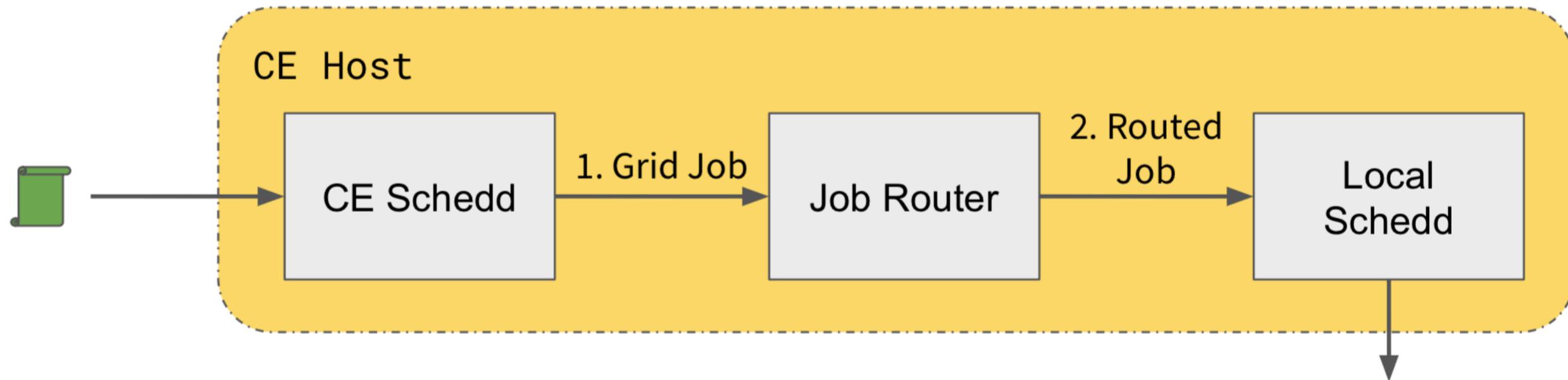
HTCondor-CE



MORGRIDGE
INSTITUTE FOR RESEARCH

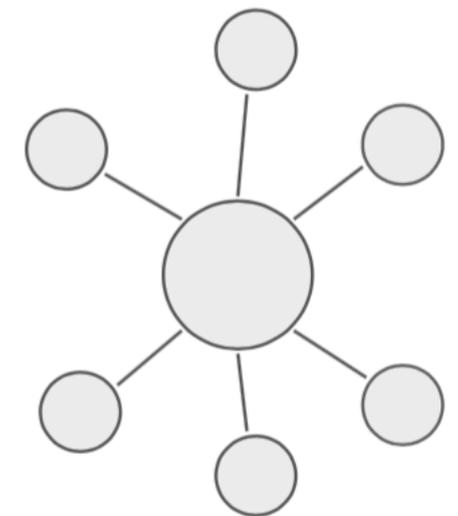
- If you look hard enough at the previous slide, you might realize that HTCondor itself fulfills all these needs!
 - HTCondor offers a remote API, has extensive auth{z,n} features, and can transform / submit jobs to an underlying batch system (another HTCondor, SLURM, PBS, etc).
- Effectively, we took a normal HTCondor submit host, enabled GSI configuration, enabled the built-in “job router” for transformations, and used the blahp to integrate with site batch systems.
 - Yes, that’s the same blahp used in CREAM!

Internals

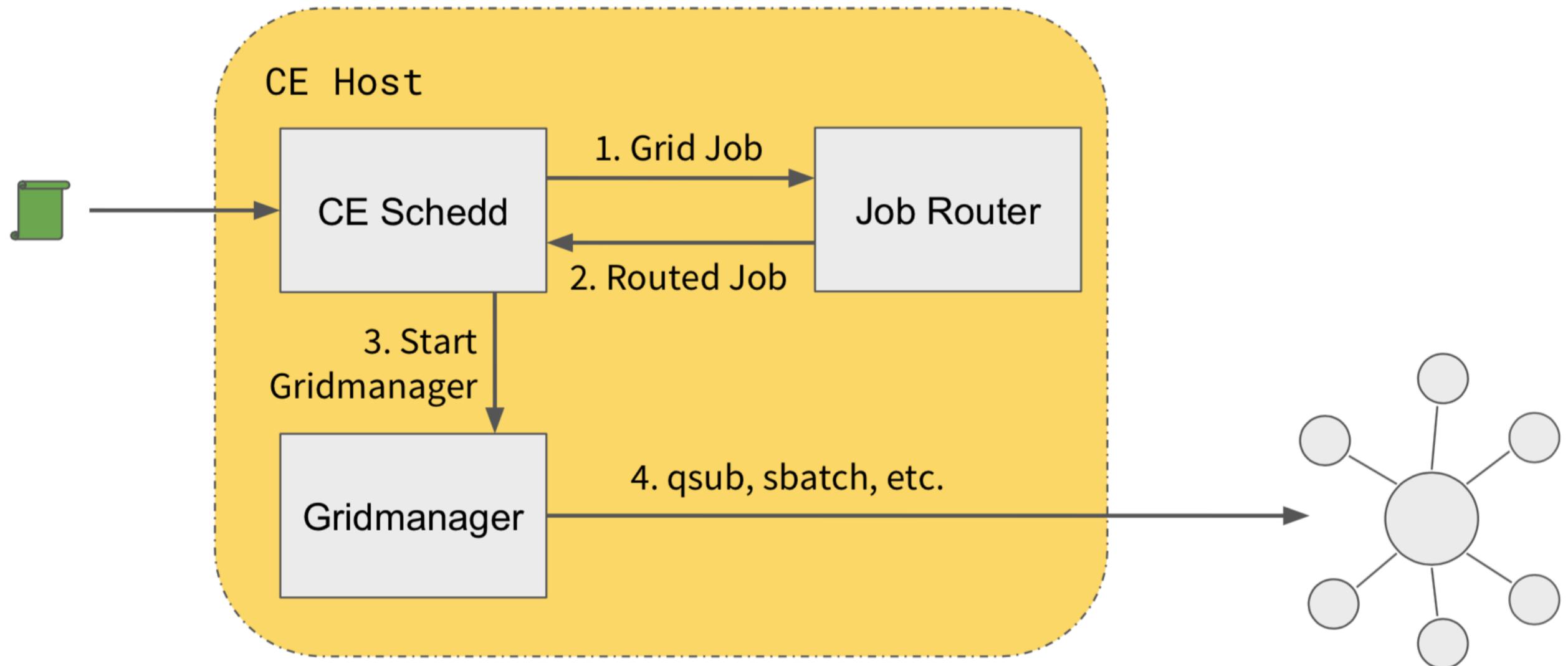


pstree output:

```
├─condor_master├─┬─condor_collector│   └─condor_schedd│   └─condor_shared_port│   └─condor_job_router
```



Internals



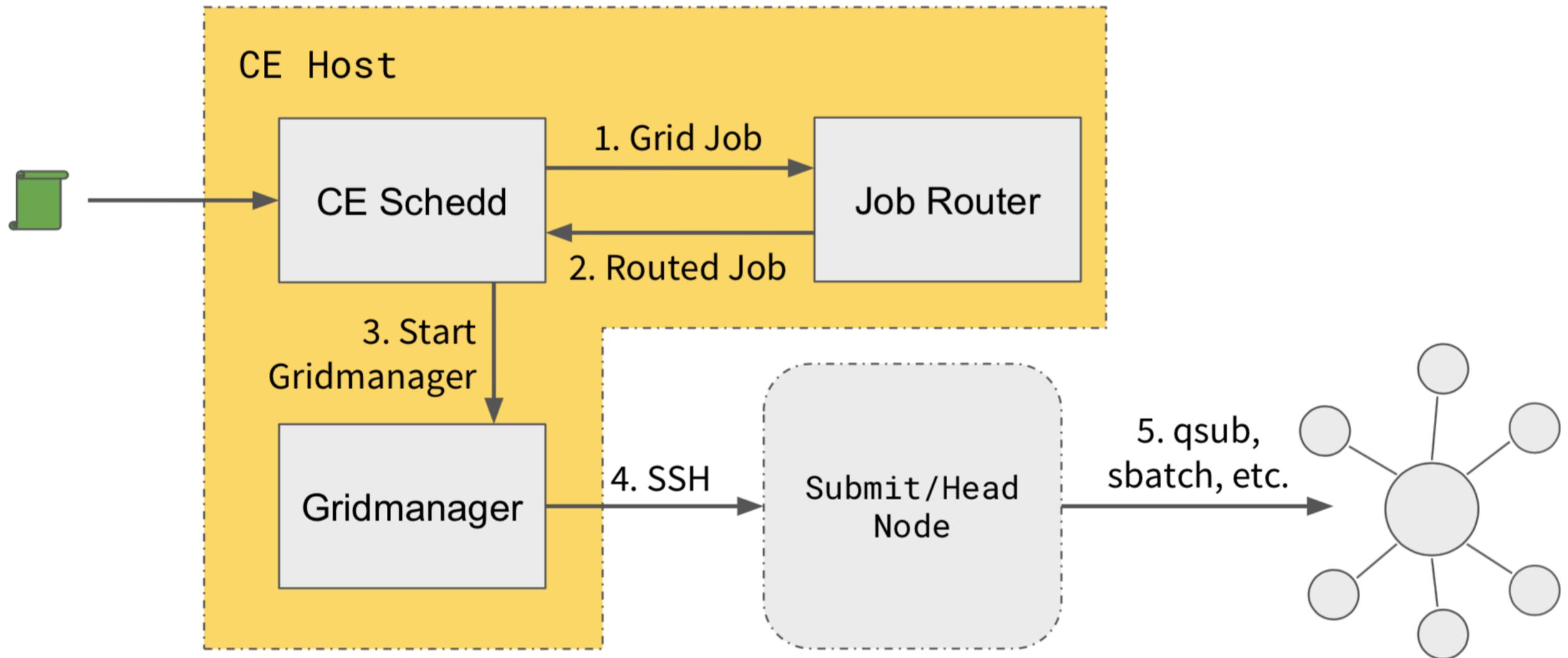
ps tree output:

```
├─condor_master──┬─condor_collector  
│               │   └─condor_schedd──┬─condor_gridmanager───blahp  
│               │                   │   └─condor_shared_port  
│               │                   └─condor_job_router
```

Internals



MORGRIDGE
INSTITUTE FOR RESEARCH



Why?



MORGRIDGE
INSTITUTE FOR RESEARCH

- The HTCondor-CE was born from an OSG investigation into a replacement for Globus GRAM.
 - In that instance, Globus got a renewal but the idea stuck...
- HTCondor-CE is, in the end, a special configuration of HTCondor.
 - In addition to configuration files, there is some packaging and helper scripts (help debug, invoke CLI, etc).
- This means the product can benefit from the expertise, support, and community around HTCondor. No dependency on another external team winning (or not winning...) a grant.
 - HTCondor *is* grant funded, but has a proven track record: it is over 30 years old.
- If you know HTCondor, then you know the CE!

Here's an example "binary" from the CE, `condor_ce_q`:

```
Executable File | 7 lines (5 sloc) | 163 Bytes  
1  #!/bin/sh  
2  
3  . /usr/share/condor-ce/condor_ce_env_bootstrap  
4  . /usr/share/condor-ce/condor_ce_client_env  
5  export GSI_AUTHZ_CONF=/dev/null  
6  exec condor_q "$@"
```

Integrating the HTCondor-CE



MORGRIDGE
INSTITUTE FOR RESEARCH

- The HTCondor-CE started within the OSG but almost immediately became broader than that, especially with OSG's collaboration with CERN.
 - **Additional backends:** SLURM support was a noticeable gap in the original version in 2013!
 - **BDII integration:** HTCondor-CE came about around the time OSG retired the BDII; original implementation.
 - **Authorization plugin:** OSG uses LCMAPS exclusively; CERN used Argus. Both use the same plugin interface, but there's a lot of configuration details to figure out.
 - **Accounting:** Integrates cleanly with the OSG accounting system but the interfaces for APEL are very different.
- Work still needed on the last item - contributions welcome!

Continued Work



MORGRIDGE
INSTITUTE FOR RESEARCH

- Over the last year, we began separating out the OSG-specific configuration from the base CE RPMs.
 - HTCondor-CE is available from the HTCondor yum repositories.
- In the next year or two, we want to move the files and release process into HTCondor team itself.
 - Not just config files, but also documentation and support!
 - There's a huge overhaul needed in the internal security configuration as well — configuration specialized for a “2013” version of HTCondor, can be greatly simplified.

<https://htcondor-wiki.cs.wisc.edu/index.cgi/wiki?p=InstallHtcondorCe>

Future Work - Authorization



MORGRIDGE
INSTITUTE FOR RESEARCH

- I have been working on some exciting new authentication mechanisms in HTCondor:
 - *Token Auth*: Allows the CE to generate a JWT that can be used for remote authentication. To appear in 8.9.2.
 - *SciTokens Auth*: Allows a VO to generate a signed JWT and have it authenticated at the CE. To appear in 8.9.2 or 8.9.3.
- A VO could generate a SciToken and use it to acquire a CE token.
 - The CE token has lower privilege: no matter what, it is only useful to talk to the destination CE.
 - This work will also be the basis of any future WLCG token auth.

PAYLOAD: DATA

```
{  
  "exp": 1557223795,  
  "iat": 1557223195,  
  "iss": "hcc-briantest7.unl.edu",  
  "scope": "condor:/READ",  
  "sub": "bbockelm@hcc-briantest7.unl.edu"  
}
```

```
$ condor_token_fetch -lifetime 600 -authz READ  
eyJhbGciOiJIUzI1NiIsImtpZCI6IiBPT0wifQ.eyJleHAiOjE1NTcyMjM0TUUsImVudCI6IjE5NTYyMzE5NSwiLCJpdiI6IjE5NTYyMzE5NSwiLCJhdWQiOiJhcnRpdDdyVCKyMWvHYpgi4Zc  
Njb3B1IjoiaY29uZG9yO1wvUkVBRCIsInN1YiI6ImJib2NrZWxtQGhjYy1icmlhbnRlc3Q3LnVubC5lZHUifQ.xFNdJOGS6XpHngwTw2hG0sRQPdDyVCKyMWvHYpgi4Zc
```

Future Work - Integration



MORGRIDGE
INSTITUTE FOR RESEARCH

- As the breadth of resources used by HEP widens, there's increased interest in using HPC resources and non-traditional resources:
 - There are a few places (e.g., Argonne National Lab) that have done a HTCondor-CE in front of their HPC center. Not clear how many will go this route.
 - At PIC, work is ongoing to provide more direct integration of payload jobs for Barcelona Supercomputing Center — would allow for pure HTCondor infrastructures to integrate directly without requiring a network service.
 - DODAS has been working on CE-less pilots - presentation later this session. I believe this work could be extended and widely adopted!

Future Work - Integration



MORGRIDGE
INSTITUTE FOR RESEARCH

- One newer HTCondor feature is “late materialization” — can instruct HTCondor to generate as many similar instances of a jobs as required.
- Would allow a sysadmin to setup a simple “NoCE” instance on a normal HTCondor-CE (if they have access to a security token such as a GSI proxy).
- An admin would need to submit a pilot job once to a HTCondor-CE, then only adjust the number of desired running jobs.

Building a Community



MORGRIDGE
INSTITUTE FOR RESEARCH

- What's important going forward is starting to build a larger community around the HTCondor-CE.
 - OSG is happy to help all of our partners, but doesn't have a clear mandate for some tasks (e.g., APEL integration is tough!).
 - Similarly, the HTCondor team doesn't have the effort to do the integration with every infrastructure out there.
- But we can help serve as a common watering hole where everyone can gather.
- As a community, we've done a poor job of coalescing around accounting, for example. I'm aware of about 4-5 sites that each did their own APEL accounting implementation.
 - Let's get this built-in. Thanks to Stephen who is starting to hammer this home!
- What other things can our team do to make the community feel welcome?

Want More?

**This has been a short overview -
We are planning a more in-depth session at the 2019
European HTCondor Week in Ispra, Italy.**

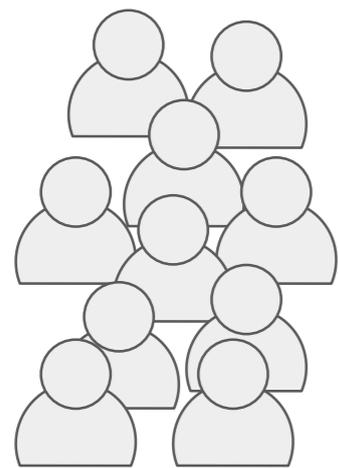


<https://indico.cern.ch/e/htcondor2019>

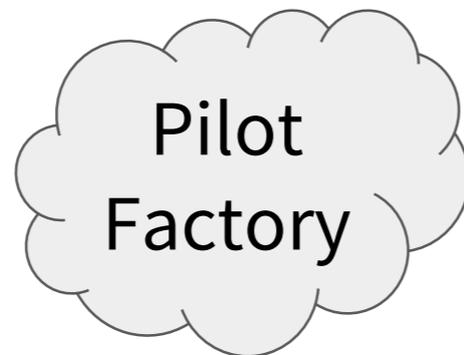
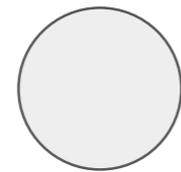
Save the Date! 24-27 September

BACKUP

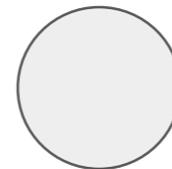
The Pilot Overlay Model



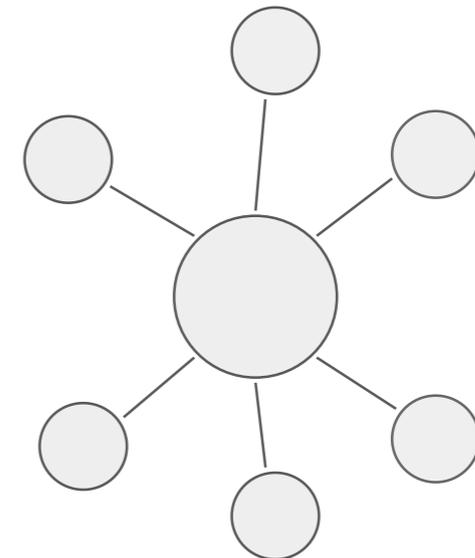
User Submit



Compute Element

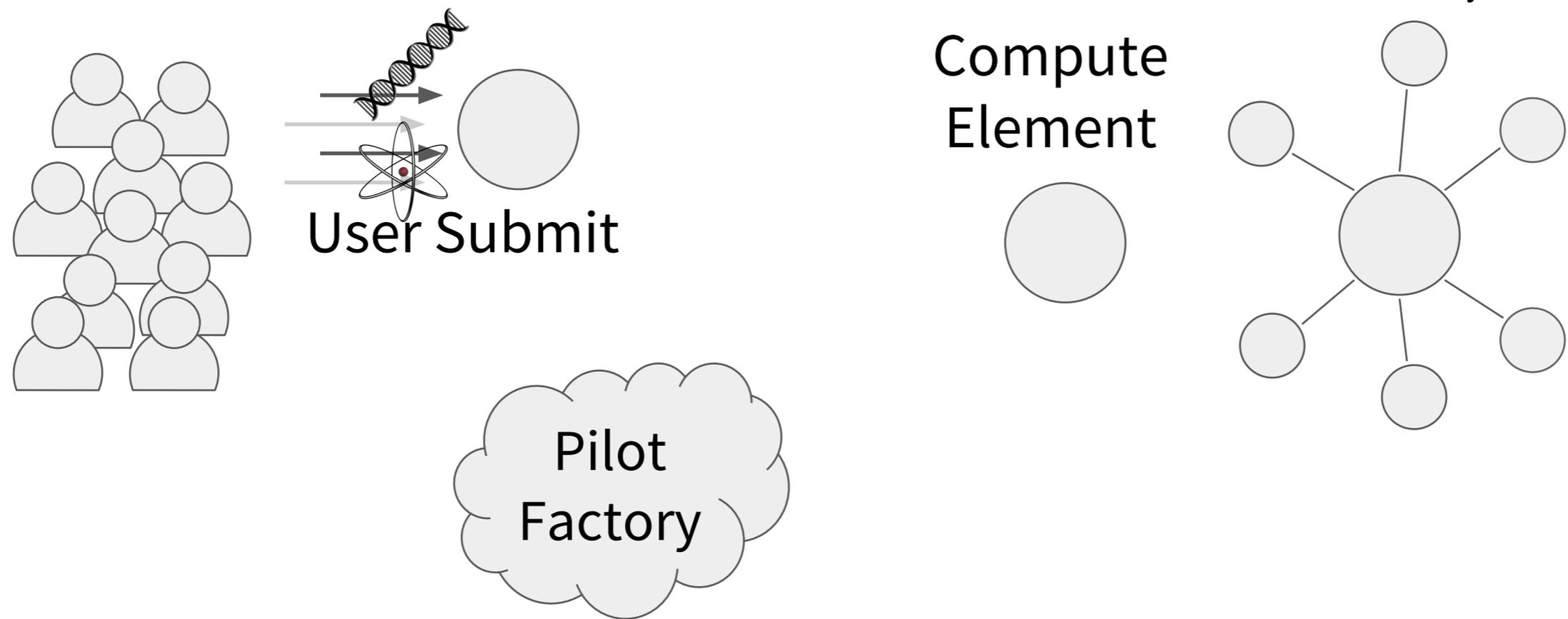


Local Batch System

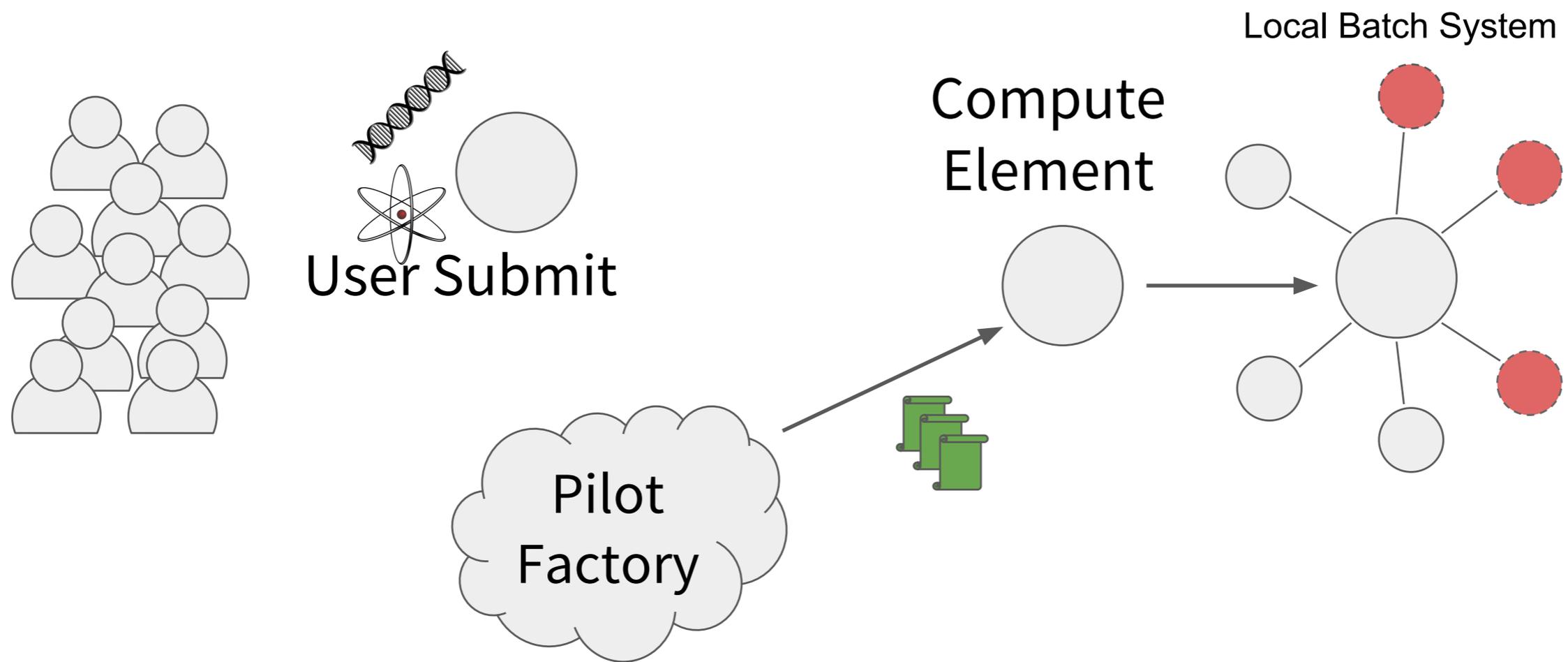


Slide (and prior diagrams) from Brian Lin

The Pilot Overlay Model



The Pilot Overlay Model



The Pilot Overlay Model

