



A prototype for integrating Theta supercomputer into HEPCloud

Maria Acosta Flechas – SCD/SCS/CSI

FIFE meeting

29/10/2020



Agenda

- Disclaimer
- Theta supercomputer
- Challenges
- Solution(s) around the globe
- The HTCondor split starter
- HTCondor split starter - Our three extras
- The Algorithm (roughly)
- Looking ahead

Disclaimer

The following prototype is not a final design/solution and was adapted specifically for CMS computing needs.

Theta supercomputer

- Theta is an 11.7-petaflops supercomputer based on Intel processors and interconnect technology, an advanced memory architecture, and a Lustre-based parallel file system, all integrated by Cray's HPC software stack. [1]

| Theta | Per Node | Aggregate |
|-------------------------|----------------|-------------|
| Compute Nodes | Intel KNL 7230 | 4,392 |
| Compute Cores | 64 | 281,088 |
| Compute Memory - DDR4 | 192 GiB | 843,264 GiB |
| Compute Memory - MCDRAM | 16 GiB | 70,272 GiB |
| Compute SSD | 128 GiB | 561,176 GiB |



Coming soon - Theta GPU

- ThetaGPU is an extension of Theta and is comprised of 24 NVIDIA DGX A100 nodes. Each DGX A100 node comprises eight NVIDIA A100 Tensor Core GPUs and two AMD Rome CPUs that provide 320 gigabytes (7680 GB aggregately) [1]

More resources we
can potentially use!

| COMPONENT | PER NODE | AGGREGATE |
|----------------------|----------|-----------|
| AMD Rome 64-core CPU | 2 | 48 |
| DDR4 Memory | 1 TB | 24 TB |
| NVIDIA A100 GPU | 8 | 192 |
| GPU Memory | 320 GB | 7,680 GB |

Challenges

- *COBALT: not your typical batch system.* Similar to SLURM but no official integration with HTCondor.
- Connectivity to the outside world is very limited – worker nodes don't have access to the internet, so everything needs to be done through edge login nodes. (outside of the main Cray system)
- This brings us to... Special services required by experiments:
 - CVMFS, Squid servers, voms calls, etc. are not provided or facilitated and represent an additional challenge for us.
- HPC != HTC.. We need to figure out scheduling policies compatible with the traditional HPC way of doing things: Larger nodes, less jobs

Solution(s) around the Globe

Multiple ways to solve the general problem of HPC isolation

- CINECA (Italian HPC center)[2]:
 - Open socket (tsocks) with edge/login node
 - Intercept network calls, encapsulate connections via SOCKS
- GridKA HPC (German HPC centers)[3]:
 - Proprietary, specific solutions: COBaID for scheduling, TARDIS for resource management
- BSC (Spanish supercomputer)[4]:
 - A variation of the split-starter
 - Scalability tests presented at O&C week [6], good prospect

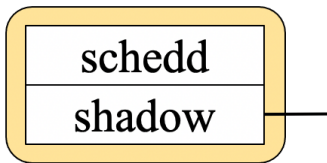


..So what about Fermilab?

The HTCondor split-starter [5]

- Proxy startd on public network
 - Accepts jobs
 - Starter writes job + data to filesystem
- On execute node
 - Job + data copied to local disk
 - Standalone starter runs job
 - Results written back to filesystem

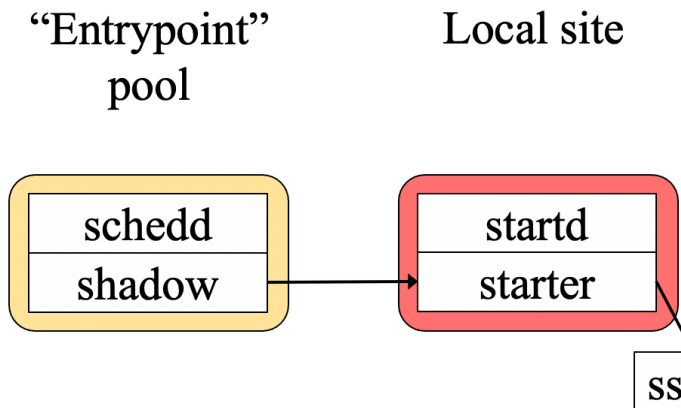
“Entrypoint”
pool



Based on diagram from: <https://indico.cern.ch/event/936993/contributions/4022104/>

The HTCondor split-starter [5]

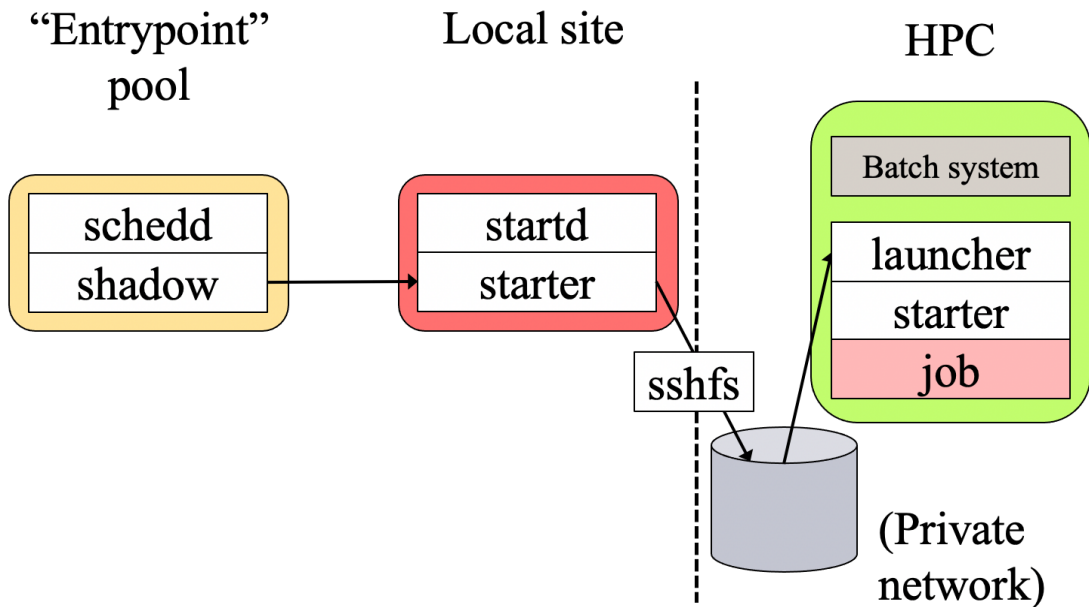
- Proxy startd on public network
 - Accepts jobs
 - Starter writes job + data to filesystem
- On execute node
 - Job + data copied to local disk
 - Standalone starter runs job
 - Results written back to filesystem



Based on diagram from: <https://indico.cern.ch/event/936993/contributions/4022104/>

The HTCondor split-starter [5]

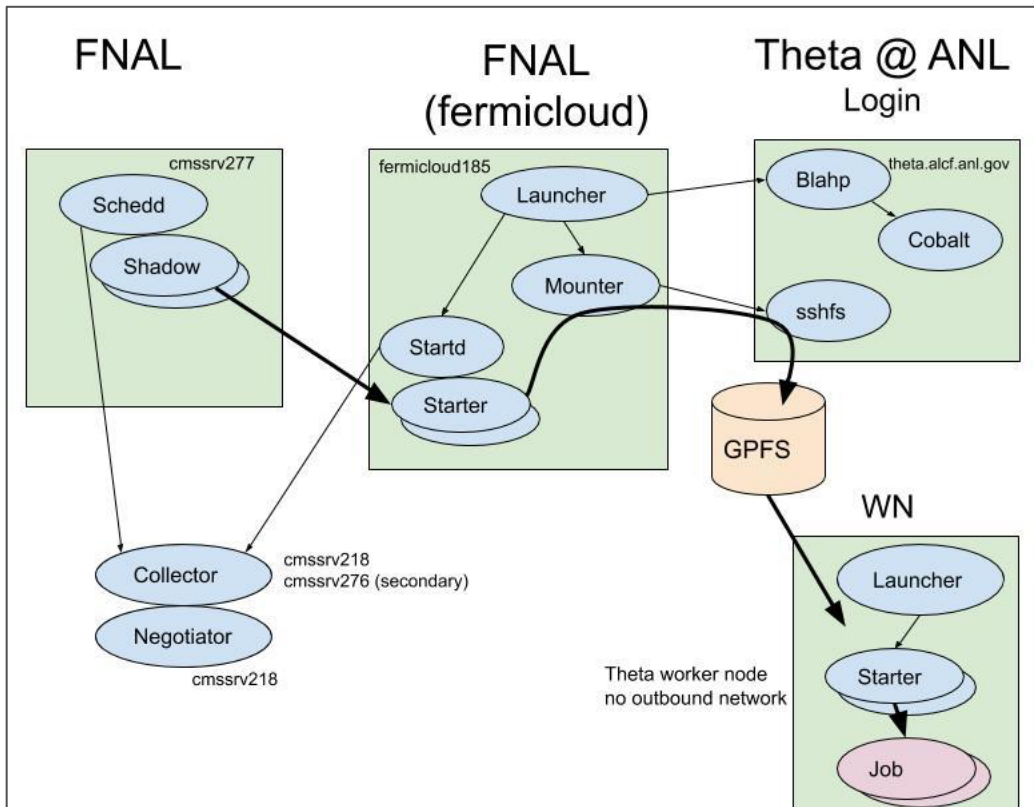
- Proxy startd on public network
 - Accepts jobs
 - Starter writes job + data to filesystem
- On execute node
 - Job + data copied to local disk
 - Standalone starter runs job
 - Results written back to filesystem



Based on diagram from: <https://indico.cern.ch/event/936993/contributions/4022104/>

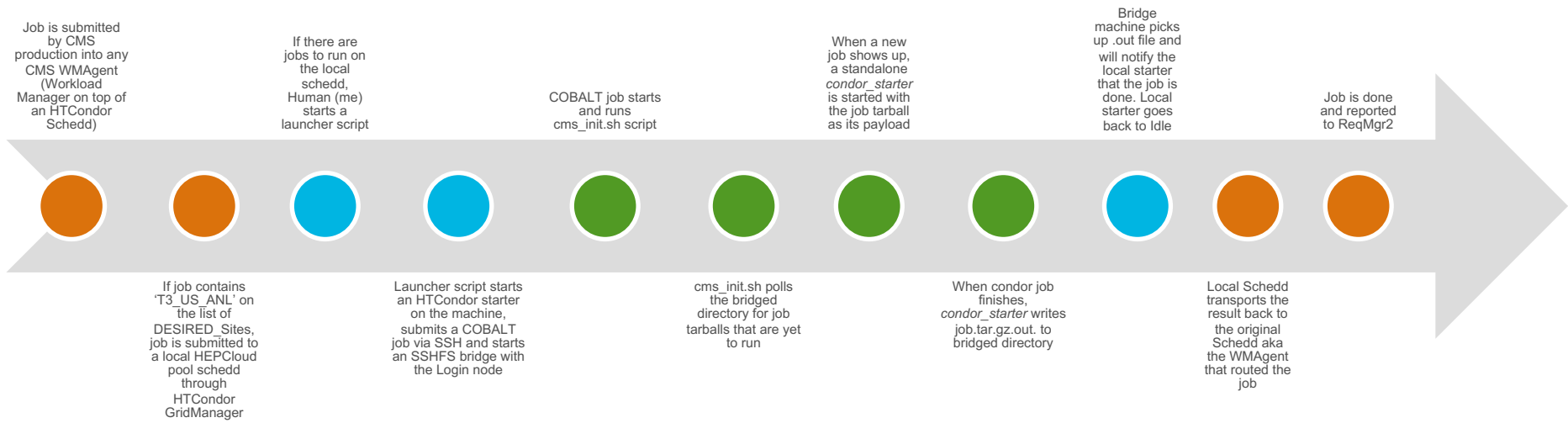
The HTCondor split-starter – Our three extras

- **Extra step: Pull jobs from CMS WMAgents**
 - Job router grabs ‘T3_US_ANL’ jobs and submits them to our local schedd
 - Job is duplicated in the local queue.
 - Original job stays Idle, copy reflects state transitions of routed job
- **Extra node “bridge node”**
 - Runs all FNAL side starters
 - Initiates interactions with Theta
 - Keeps track of jobs
 - Talks to local pool central managers
- **Extra services available to the startd**
 - Squid service for CVMFS
 - Non-FUSE, unprivileged CVMFS through cvmfsexec
 - Singularity



The Algorithm (roughly)

** Find sequence and communication diagram in backup slides



- WMAgent
- "Bridge" node
- Theta

The Algorithm (roughly) - Part 1

Job is submitted by CMS production into any CMS WMAgent (Workload Manager on top of an HTCondor Schedd)

If there are jobs to run on the local schedd, Human (me) starts a launcher script



If job contains 'T3_US_ANL' on the list of DESIRED_Sites, job is submitted to a local HEPCloud pool schedd through HTCondor GridManager

Launcher script starts an HTCondor starter on the machine, submits a COBALT job via SSH and starts an SSHFS bridge with the Login node

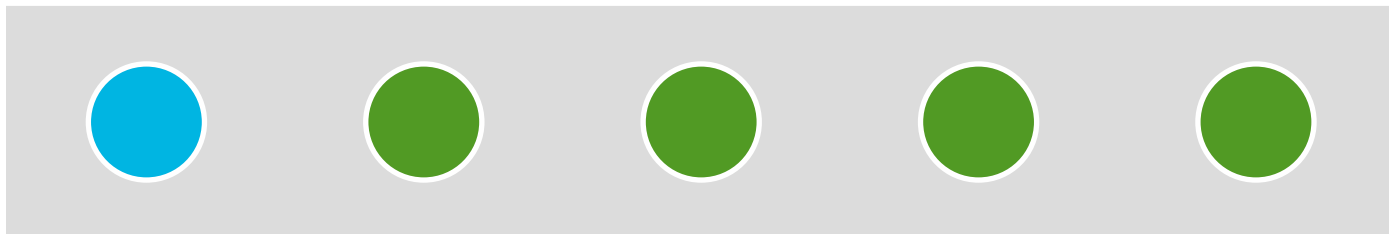
- WMAgent
- "Bridge" node
- Theta

The Algorithm (roughly) - Part 2

| | | | | | | |
|--|-------|--------|---------|------|-------|------|
| slot1_28@cobalt_15793@theta.alcf.anl.gov | LINUX | X86_64 | Claimed | Busy | 0.000 | 1280 |
| slot1_29@cobalt_15793@theta.alcf.anl.gov | LINUX | X86_64 | Claimed | Busy | 0.000 | 1280 |
| slot1_30@cobalt_15793@theta.alcf.anl.gov | LINUX | X86_64 | Claimed | Busy | 0.000 | 1280 |
| slot1_31@cobalt_15793@theta.alcf.anl.gov | LINUX | X86_64 | Claimed | Busy | 0.000 | 1280 |
| slot1_32@cobalt_15793@theta.alcf.anl.gov | LINUX | X86_64 | Claimed | Busy | 0.000 | 1280 |
| slot1_33@cobalt_15793@theta.alcf.anl.gov | LINUX | X86_64 | Claimed | Busy | 0.000 | 1280 |

COBALT job starts
and runs `cms_init.sh`
script

When a new job
shows up, a
standalone
`condor_starter`
is
started with the job
tarball as its payload



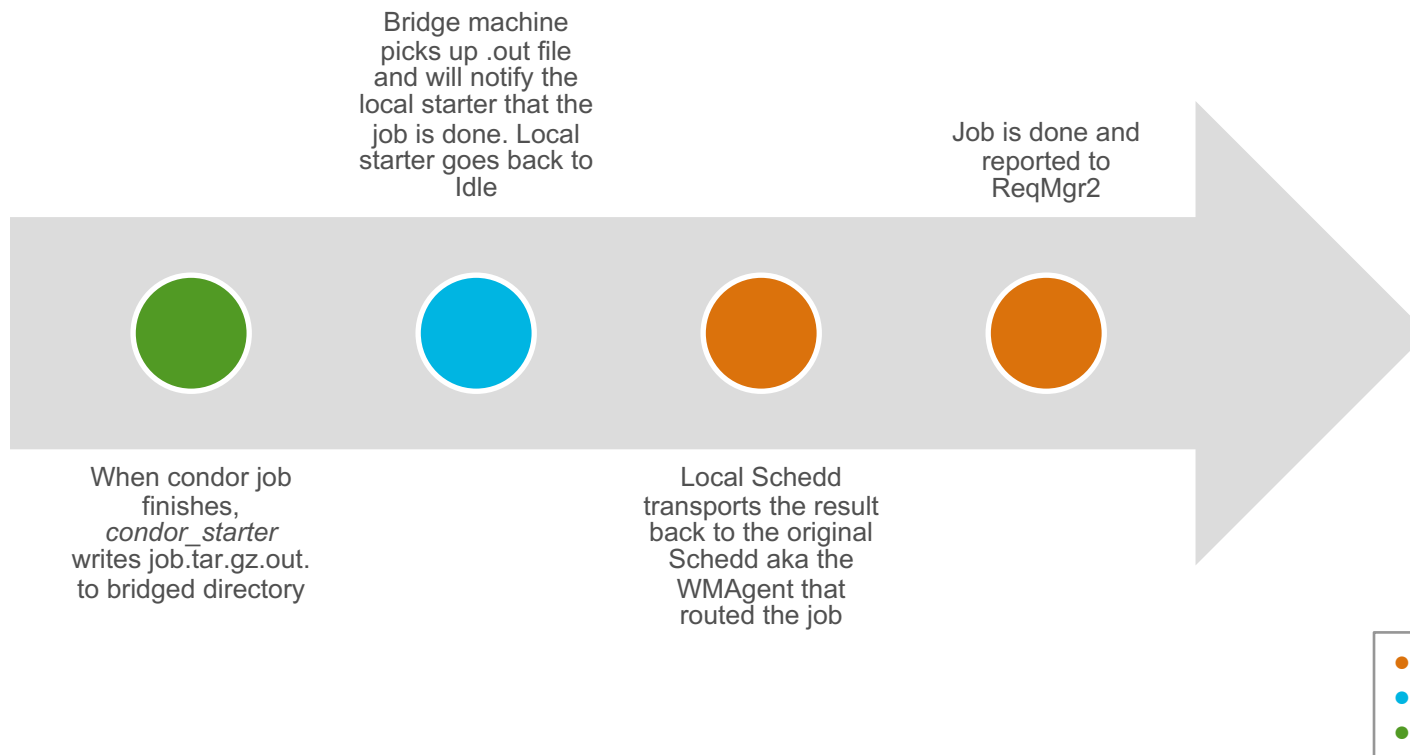
Launcher script starts
an HTCondor starter
on the machine,
submits a COBALT
job via SSH and
starts an SSHFS
bridge with the Login
node. When a job
arrives, it puts the
payload in the shared
SSHFS directory

`cms_init.sh` polls the
bridged directory for
job tarballs that are
yet to run

When condor job
finishes,
`condor_starter` writes
`job.tar.gz.out.` to
bridged directory

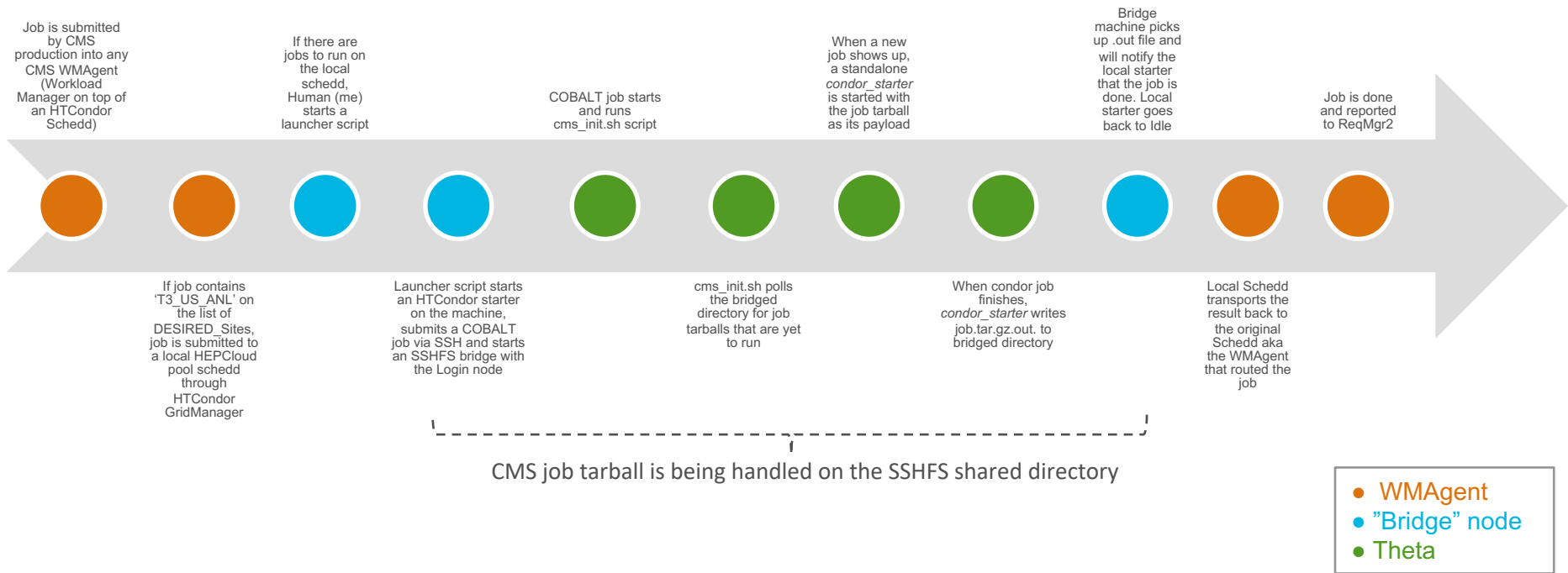
- WMAgent
- "Bridge" node
- Theta

The Algorithm (roughly) - Part 3



The Algorithm (roughly)

** Find sequence and communication diagram in backup slides



Looking ahead

- There are **as many solutions as people willing to attack the problem**.. There's no 'one fits all' so we need to invest time and technical understanding to solve our use case.
- We are **iterating** and testing different approaches along with the HTCondor team.
- Moving TBs of data through SSHFS is **not realistic nor a good practice**.
- **Data transfers problem is still a big if** as CMS is in the middle of Rucio migration. Isolated network adds complexity
- Setup is rather rudimentary, there's **room for improvement**: Exception handling, error recovery, monitoring, etc.
- **Automation!** There's still a human (me) pressing the "button" and watching jobs and startds come and go.
- **MFA** requires manual input of a random password, need to negotiate with ANL for a possible way of automating.
- There is an additional challenge to **integrate the solution we produce with the existing infrastructure**, the DE in particular



Thanks! Questions?

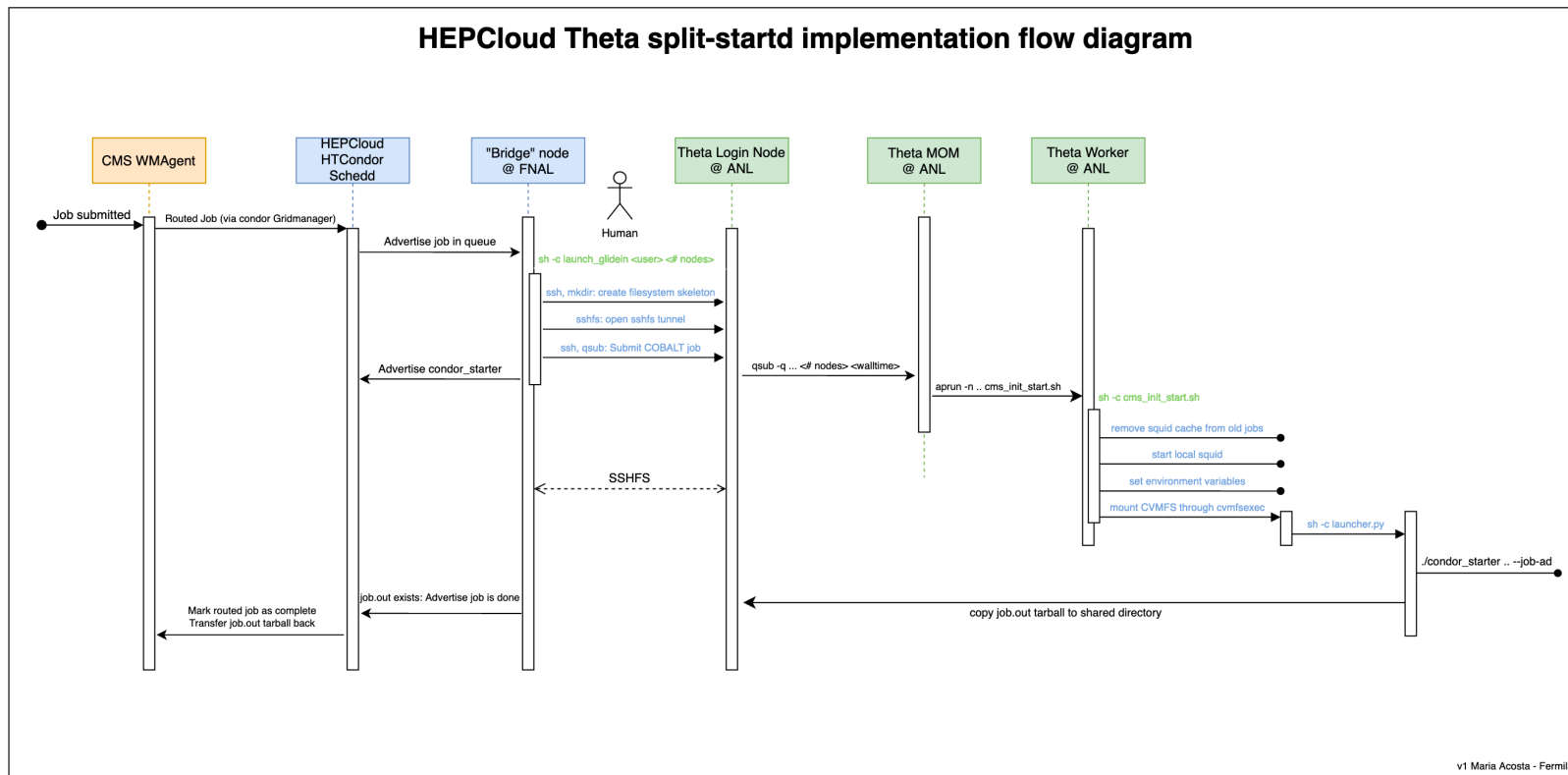
Maria Acosta - SCS/CSI
macosta@fnal.gov
@macosta on Slack

Resources

- [1] <https://www.alcf.anl.gov/support-center/theta/theta-thetagpu-overview>
- [2] https://indico.cern.ch/event/957688/contributions/4058330/attachments/2123148/3573944/RD_on_Connecting_Centers_with_limited_no_outbound_connectivity_OC_week_Oct_2020_1.pdf
- [3] https://indico.cern.ch/event/957688/contributions/4058338/attachments/2123212/3574056/2020_CMSOC_KITHPC.pdf
- [4] https://indico.cern.ch/event/957688/contributions/4058325/attachments/2123167/3573982/20201014_PIC_HTCCondor_BSC.pdf
- [5] <https://indico.cern.ch/event/936993/contributions/4022104/>
- [6] <https://indico.cern.ch/event/957688>

Backup Slides

Flowchart diagram of the setup/algorithm



Lumberjack

- “It moves the logs”
- Pause jobs in home schedd
- Transfer job queue and input files
 - Create export job queue file
- Run an HTCondor pool as HPC job
- Transfer job queue and output files
 - Merge into home schedd’s queue/spool
- Unpause jobs in home schedd

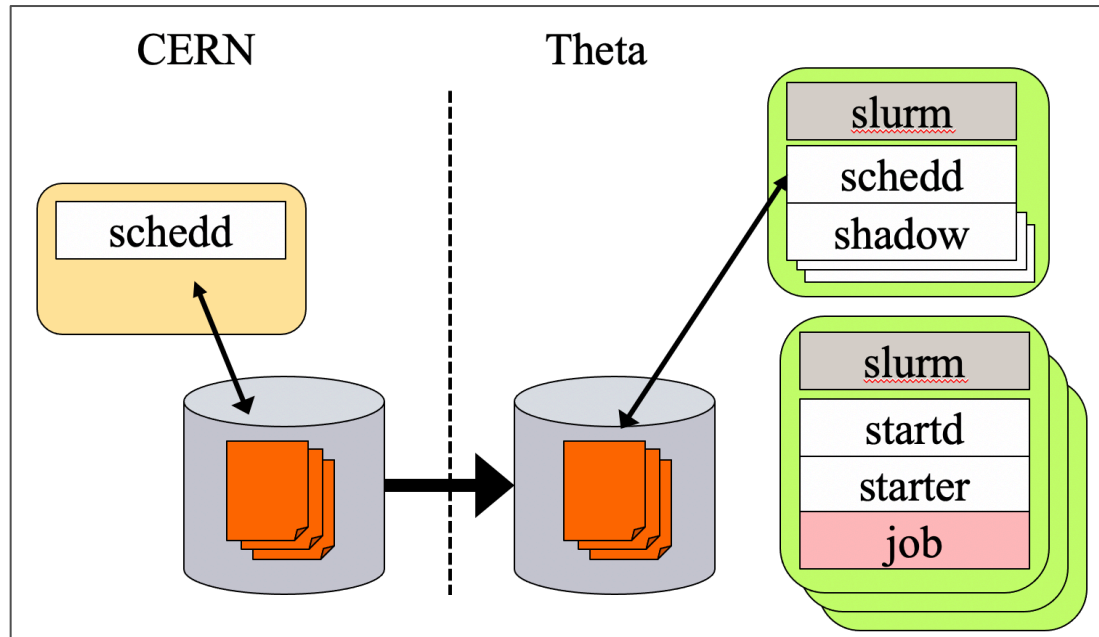


Image recovered from: <https://indico.cern.ch/event/936993/contributions/4022104/>