IceProd Supercomputer Mode

How IceCube Production Runs on Firewalled Clusters

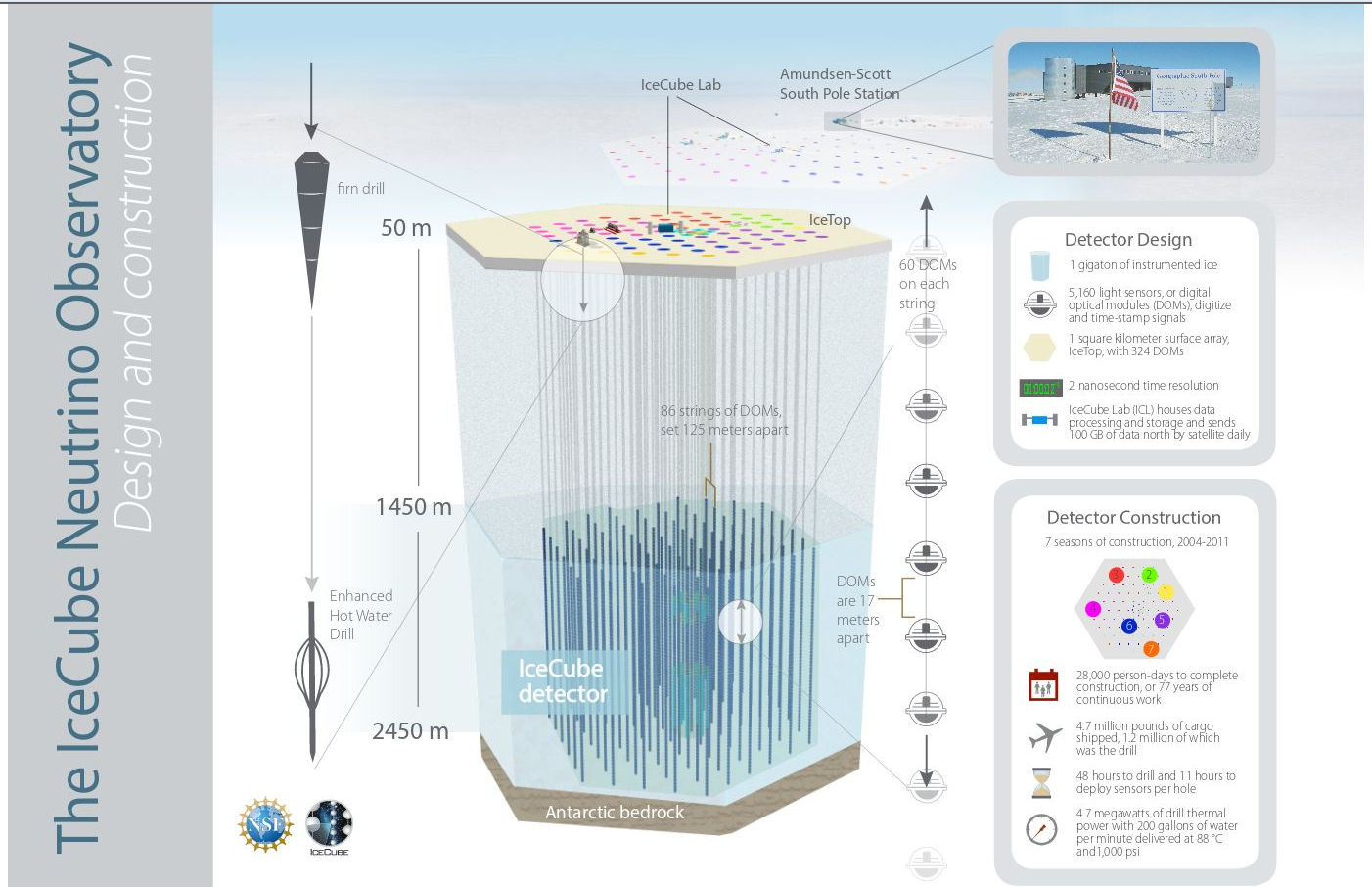
David Schultz
WIPAC, UW-Madison

Outline

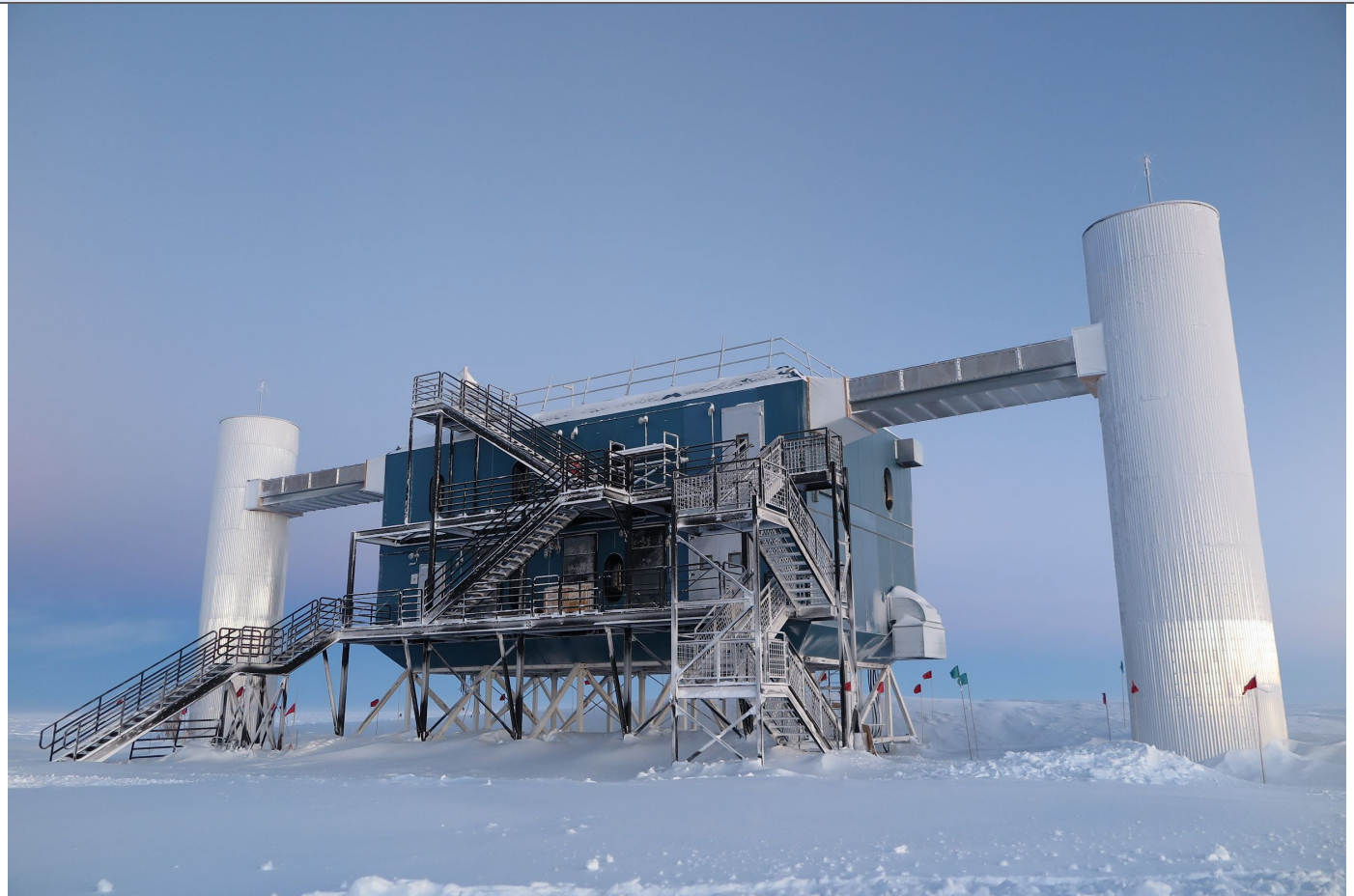
- What is IceCube / IceProd
- Supercomputer Specialties
- Current Usage
- Future Plans

IceCube Collaboration

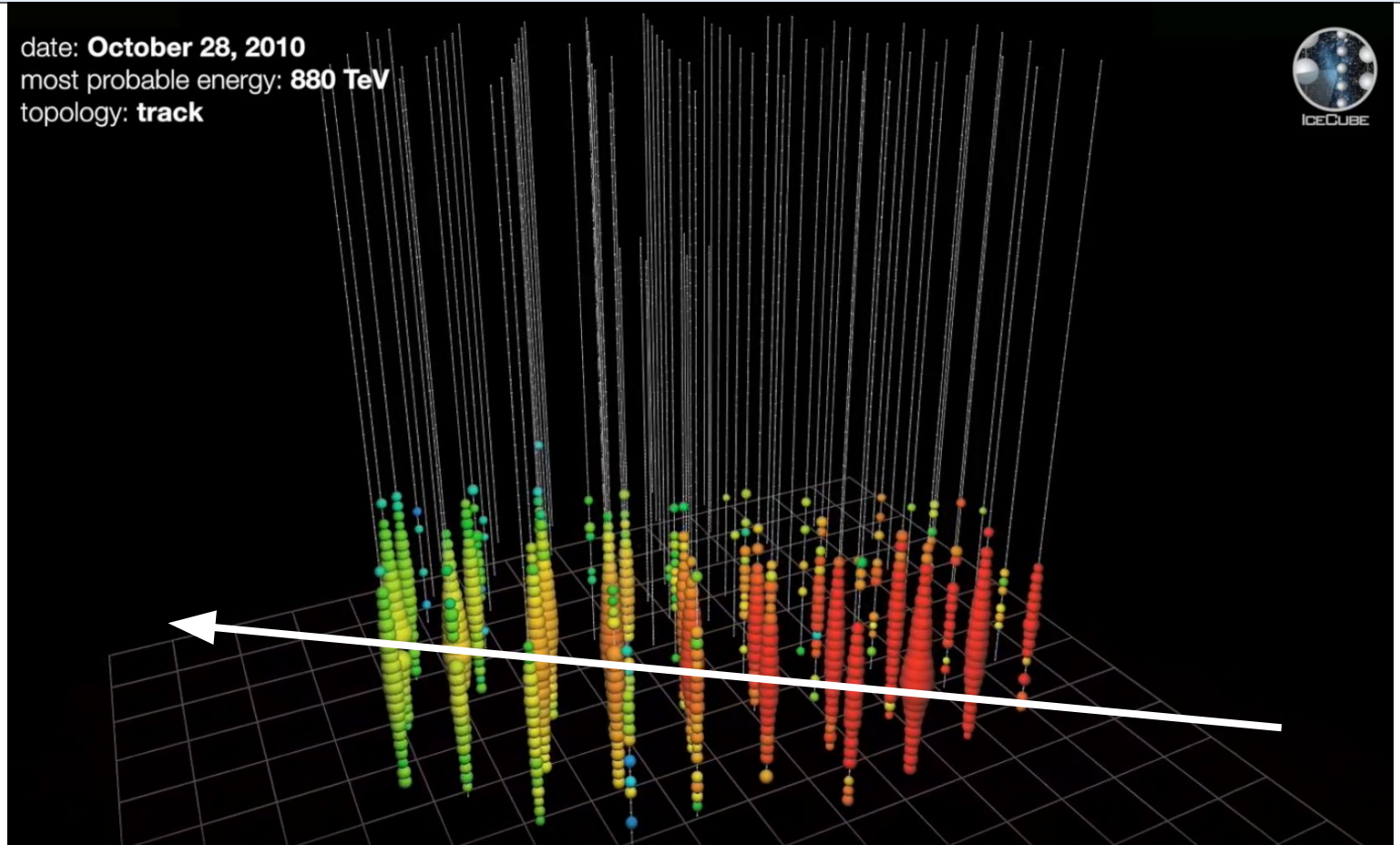
IceCube Neutrino Observatory



IceCube Neutrino Observatory



IceCube Neutrino Observatory



What is IceProd

Data provenance

- Configuration - which software, what versions, when/where it ran, …

Dataset submission

- Monitor job status, resource usage
- Retry failed jobs - resubmit with different requirements

Supercomputer Challenges

Networking

- No external access from worker nodes
- Special “data mover” nodes separate from submit nodes

Solution:

- Stage data using “data mover” nodes
  - Automated to happen before job submit, after job completion
- Edit job config to point to local paths before submission

Supercomputer Challenges

OS support

- Special flavors of operating systems not compatible with existing builds
  - Some cases where OS reports it is RHEL 7, but all libraries are different versions

Solution:

- Containers!
- Run everything in the standard OSG WN container

Supercomputer Challenges

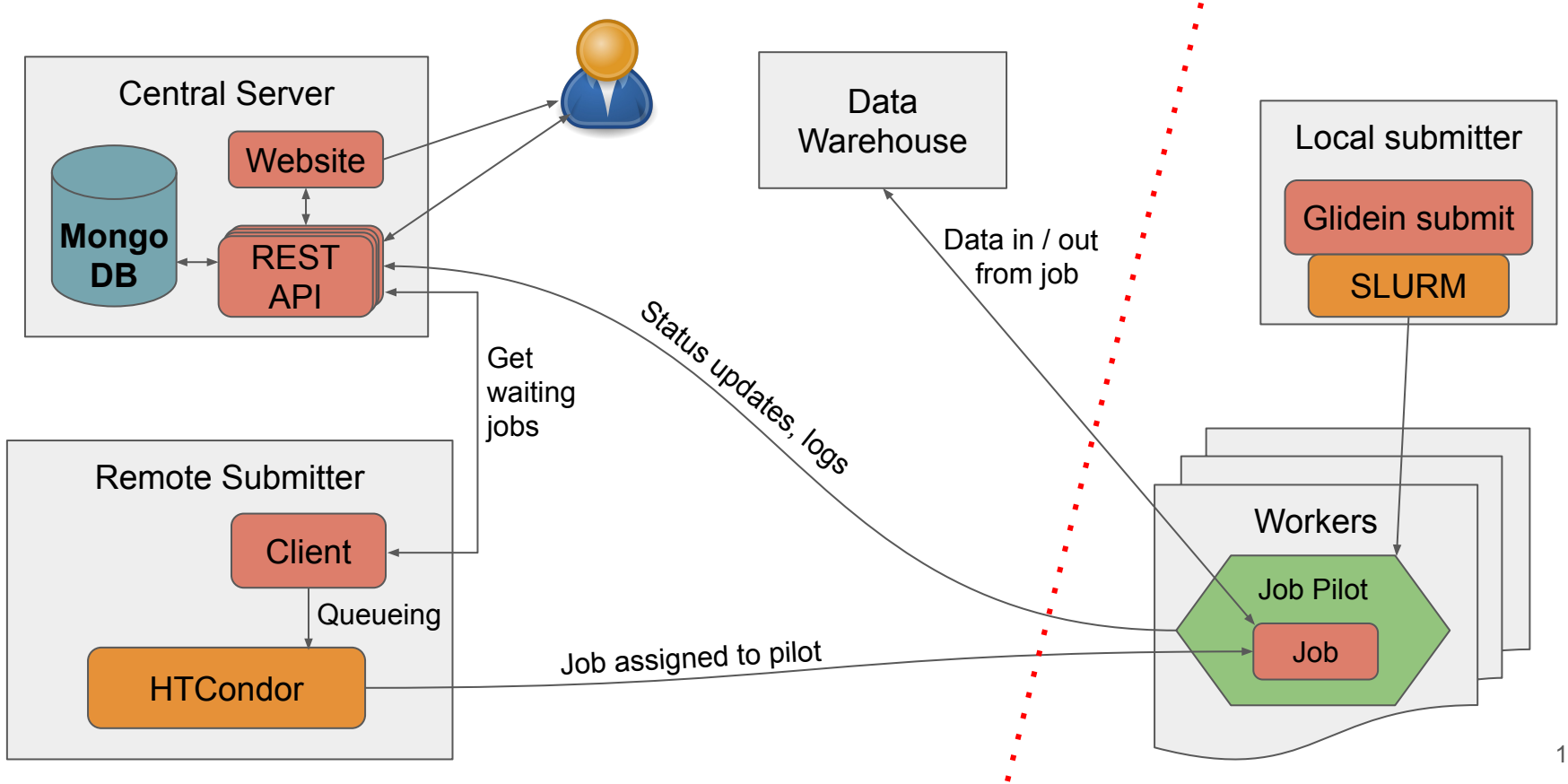
CVMFS support

- Some sites still don’t support CVMFS

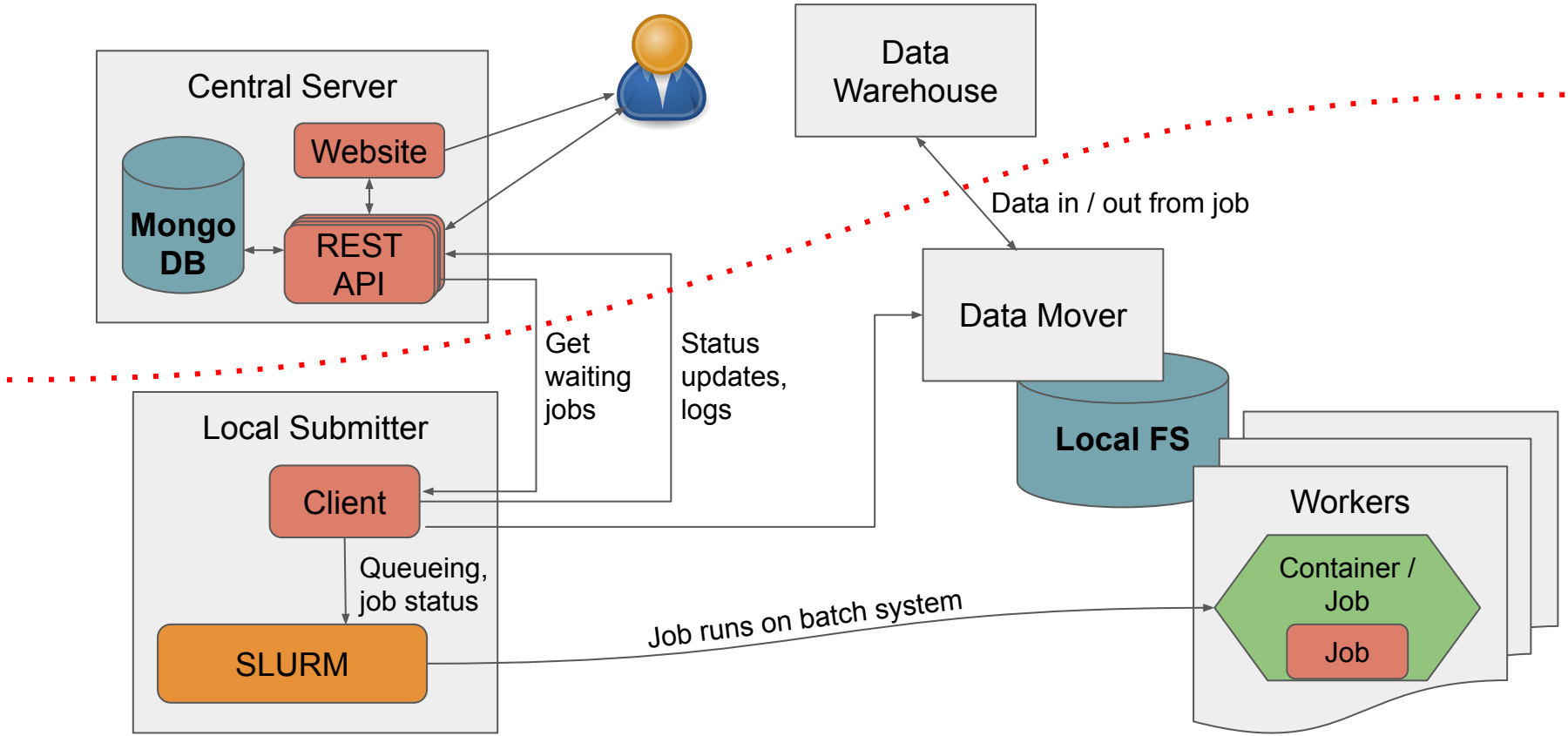
Solution:

- Containers again!
- Rsync CVMFS nightly to local FS, then bind mount into container

Typical Production Architecture



Supercomputer Architecture



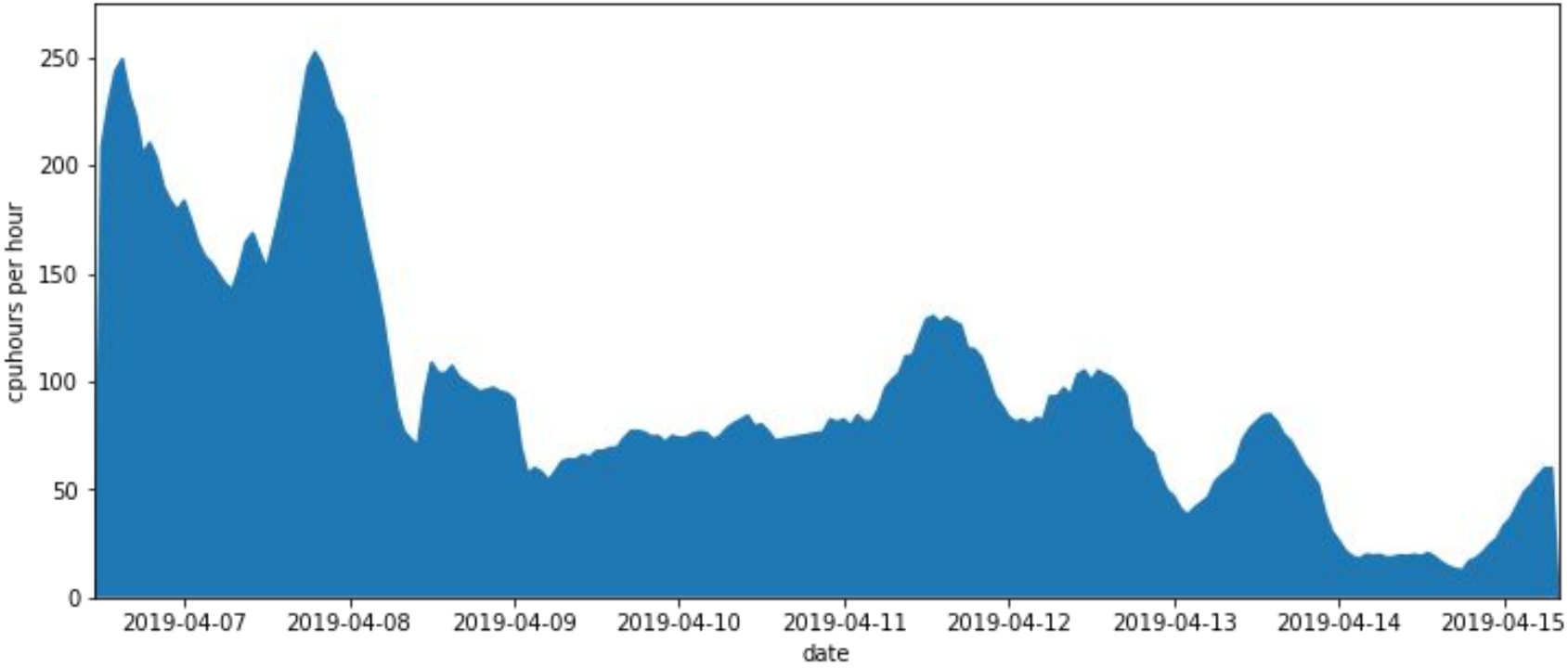
Current Sites

Graham in Compute Canada

- Part of IceCube's Compute Canada allocation
- "Small" share, about 100 CPUs continuously
- Running since April 2019

Current Sites

Graham in Compute Canada



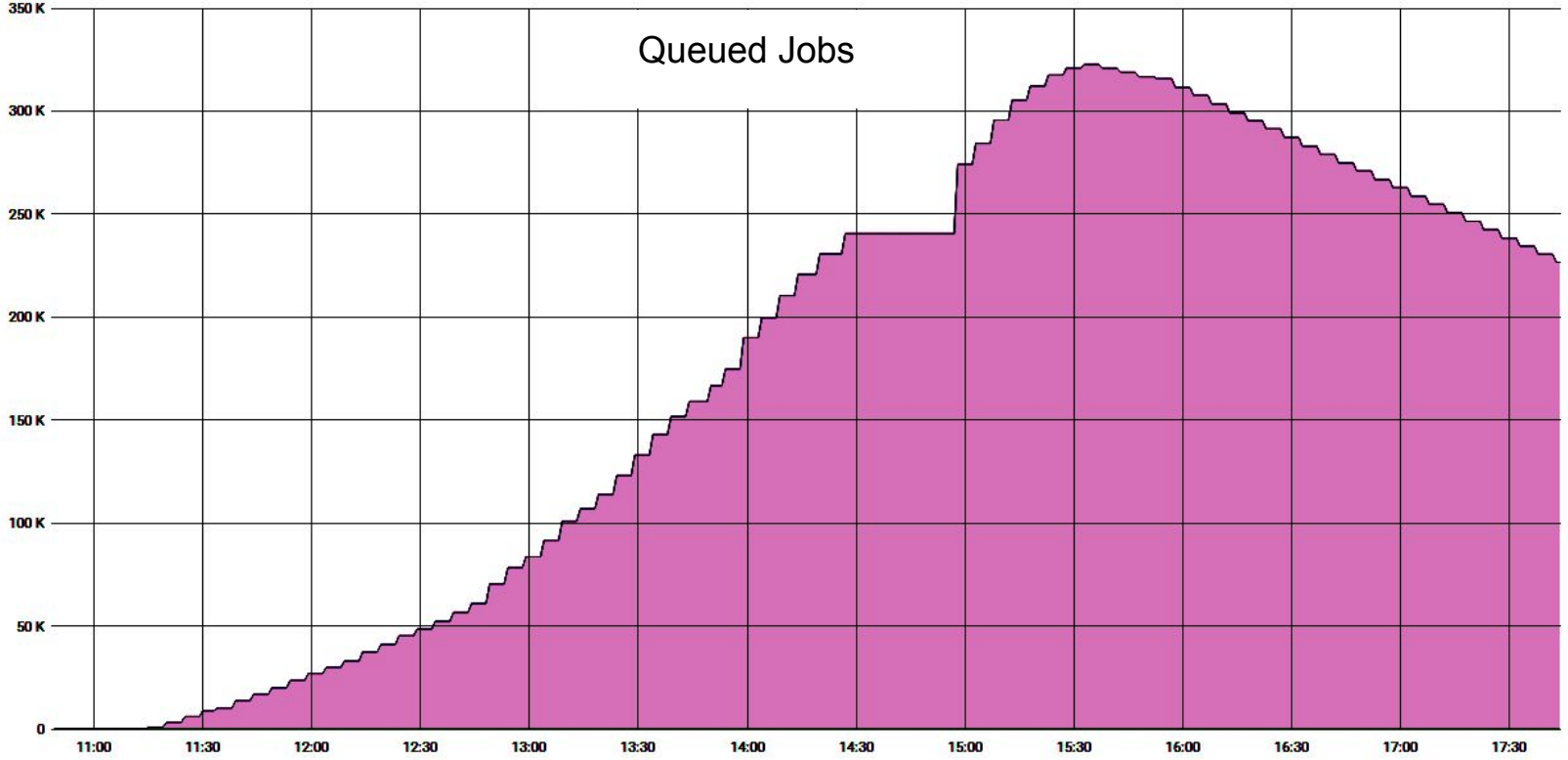
Current Sites

Exascale Cloud Demo

- NSF EAGER grant with SDSC, Internet2
- Using 3 major cloud providers: AWS, Google, Azure
- Goal: get as close to an exaflop as possible in a single condor pool, on cloud resources today
- Like a supercomputer:
  - No direct I/O – our servers can’t sustain that rate
  - No pilots – infrastructure can’t handle the load

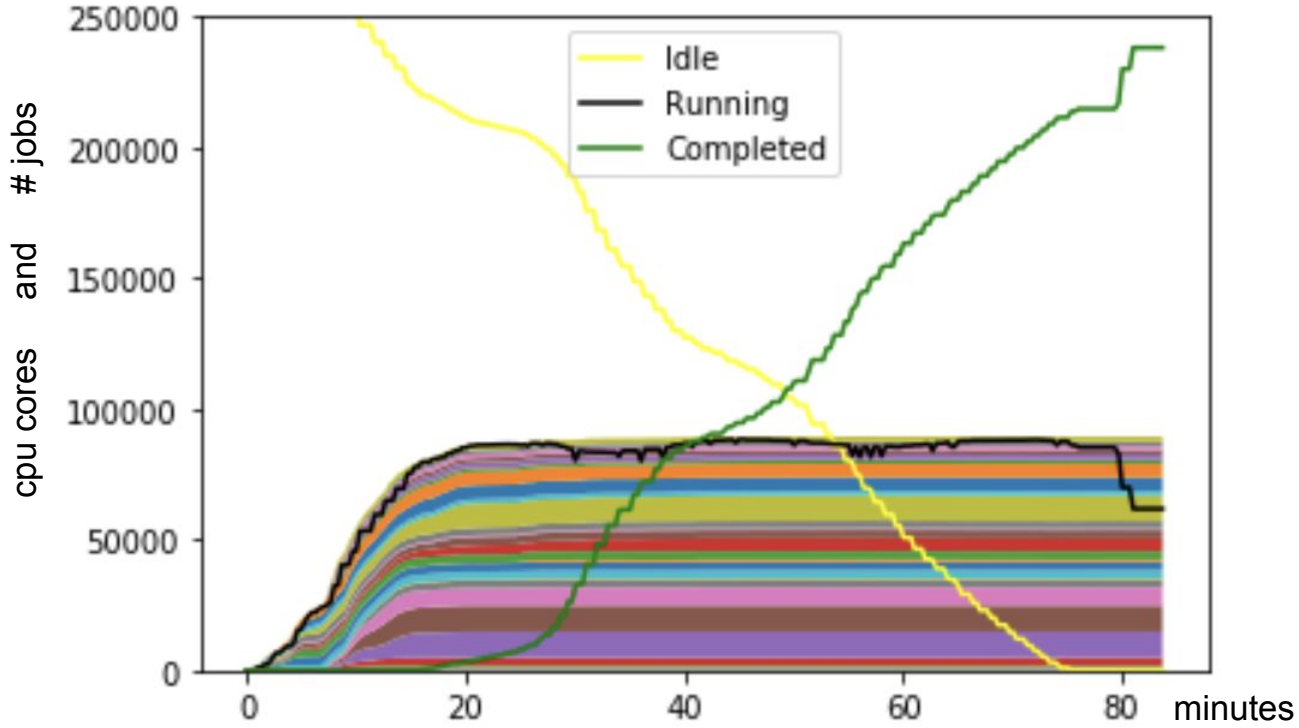
Current Sites

Exascale Cloud Demo



Current Sites

Exascale Cloud Demo



Future Supercomputer Challenges

Queueing

- Some sites have policies preferring large MPI jobs over many single-core jobs, or other special things
  - One MPI job can expand to the entire size of the supercomputer
  - Queue thus allows very few jobs per user

Still working on this

- One idea we've tested: make a local HTCondor pool out of large, long-running jobs (whole machine, 24 hours), and submit physics jobs to that pool

Conclusions

- IceCube primarily relies on a glidein model with worker node internet access
- Supercomputers disallow this, so a direct submission mode was created
- Works well so far, more sites to be added in the future