Consolidating Computational Resources at IceCube

David Schultz, Gonzalo Merino, Vladimir Brik
Wisconsin IceCube Particle Astrophysics Center

European HTCondor Workshop 2017

THE ICECUBE COLLABORATION

AUSTRALIA
University of Adelaide

BELGIUM
Université libre de Bruxelles
Universiteit Gent
Vrije Universiteit Brussel

CANADA
SNOLAB
University of Alberta–Edmonton

DENMARK
University of Copenhagen

GERMANY
Deutsches Elektronen-Synchrotron
ECAP, Universität Erlangen-Nürnberg
Humboldt–Universität zu Berlin
Ruhr-Universität Bochum
RWTH Aachen University
Technische Universität Dortmund
Technische Universität München
Universität Mainz
Universität Wuppertal
Westfälische Wilhelms-Universität Münster

JAPAN
Chiba University

NEW ZEALAND
University of Canterbury

REPUBLIC OF KOREA
Sungkyunkwan University

SWEDEN
Stockholms Universitet
Uppsala Universitet

SWITZERLAND
Université de Genève

UNITED KINGDOM
University of Oxford

UNITED STATES
Clark Atlanta University
Drexel University
Georgia Institute of Technology
Lawrence Berkeley National Lab
Marquette University
Massachusetts Institute of Technology
Michigan State University
Ohio State University
Pennsylvania State University
South Dakota School of Mines and Technology
Southern University and A&M College
Stony Brook University
University of Alabama
University of Alaska Anchorage
University of California, Berkeley
University of California, Irvine
University of Delaware
University of Kansas
University of Maryland
University of Rochester
University of Texas at Arlington
University of Wisconsin–Madison
University of Wisconsin–River Falls
Yale University

FUNDING AGENCIES

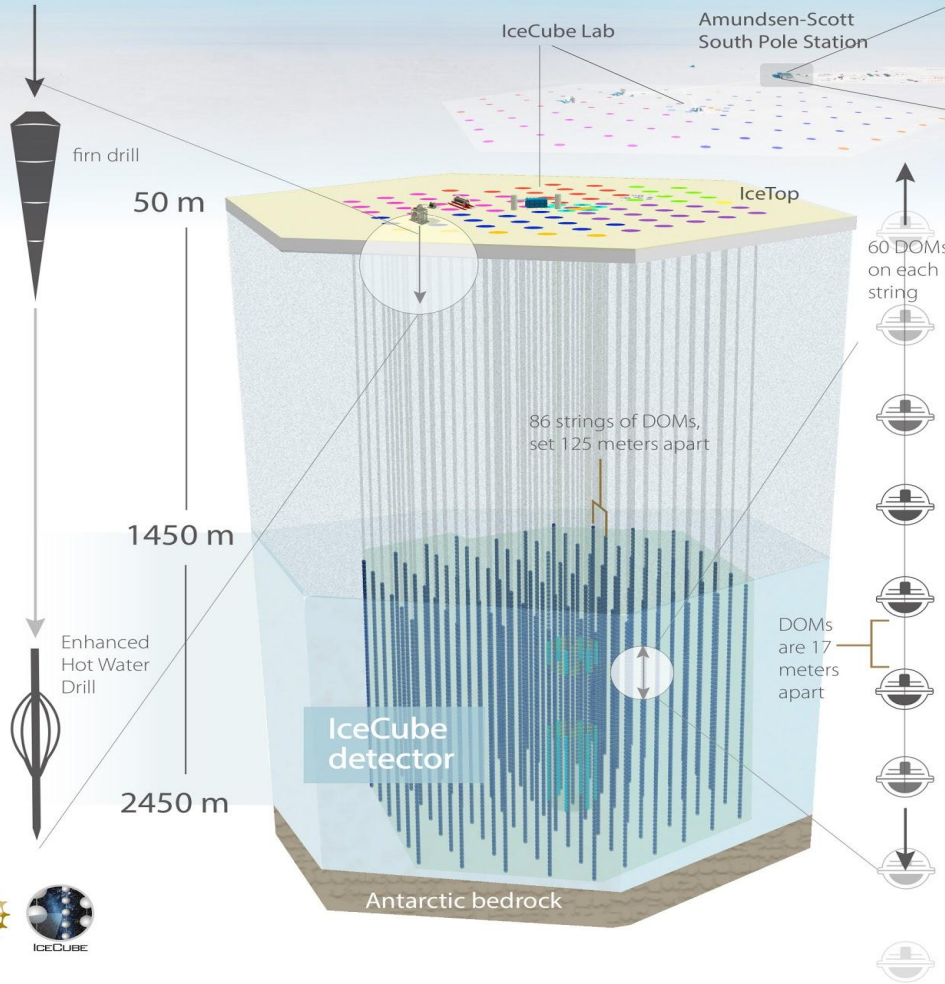
Fonds de la Recherche Scientifique (FRS-FNRS)
Fonds Wetenschappelijk Onderzoek-Vlaanderen (FWO-Vlaanderen)
Federal Ministry of Education and Research (BMBF)
German Research Foundation (DFG)
Deutsches Elektronen-Synchrotron (DESY)
Japan Society for the Promotion of Science (JSPS)
Knut and Alice Wallenberg Foundation
Swedish Polar Research Secretariat
The Swedish Research Council (VR)
University of Wisconsin Alumni Research Foundation (WARF)
US National Science Foundation (NSF)

IceCube
South Pole Neutrino Observatory

icecube.wisc.edu

The IceCube Neutrino Observatory

Design and construction

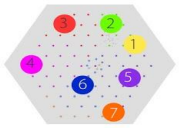


Detector Design

- 1 gigaton of instrumented ice
- 5,160 light sensors, or digital optical modules (DOMs), digitize and time-stamp signals
- 1 square kilometer surface array, IceTop, with 324 DOMs
- 2 nanosecond time resolution
- IceCube Lab (ICL) houses data processing and storage and sends 100 GB of data north by satellite daily

Detector Construction

7 seasons of construction, 2004-2011



- 28,000 person-days to complete construction, or 77 years of continuous work
- 4.7 million pounds of cargo shipped, 1.2 million of which was the drill
- 48 hours to drill and 11 hours to deploy sensors per hole
- 4.7 megawatts of drill thermal power with 200 gallons of water per minute delivered at 88 °C and1,000 psi

IceCube's Computing Ecosystem

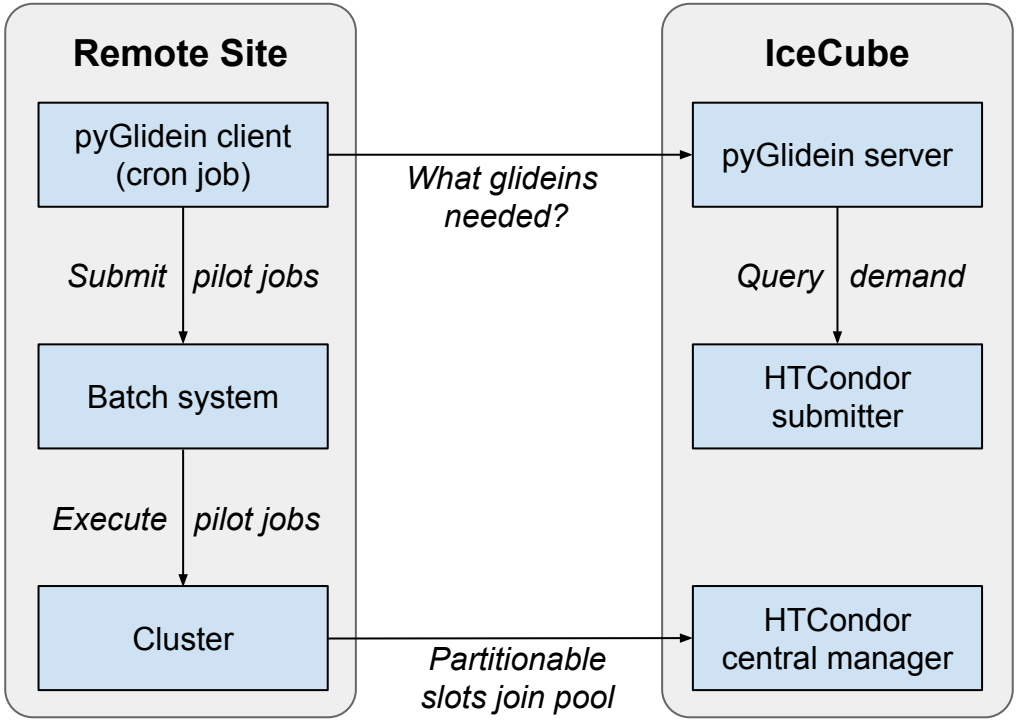
1. IceCube's HTCondor cluster at UW-Madison
   - Main analysis facility; 7300 CPU cores, 320 GPUs; *direct access to Lustre storage*
2. Flocking access to OSG, and various HTCondor pools at UW-Madison
   - Good source of CPUs, but not a lot of GPUs available
3. Clusters owned by institutions that are part of IceCube collaboration
   - Many CPUs and GPUs across about 20 heterogeneous sites
4. XSEDE allocations on Comet, Bridges, XStream supercomputers
   - Mostly interested in GPUs; can have special constraints

Goal: consolidate all available resources into a single global HTCondor pool (eventually)

IceCube's pyGlidein software

- https://github.com/WIPACrepo/pyglidein
  - Uses glideins to make many sites' resources accessible via a single HTCondor pool
  - In production since 2015, making difficult-to-access resources easy to use
  - People like it: simple, easy to set up by a non-expert, low maintenance
  - "Client" is a cron job that submits pilots to the local cluster
  - Can submit to HTCondor, PBS, SLURM, UGE, LSF
  - For technical details see David Schultz's Condor Week 2016 presentation: http://research.cs.wisc.edu/htcondor/HTCondorWeek2016/presentations/ThuSchultz_IceCubeGlideins.pdf

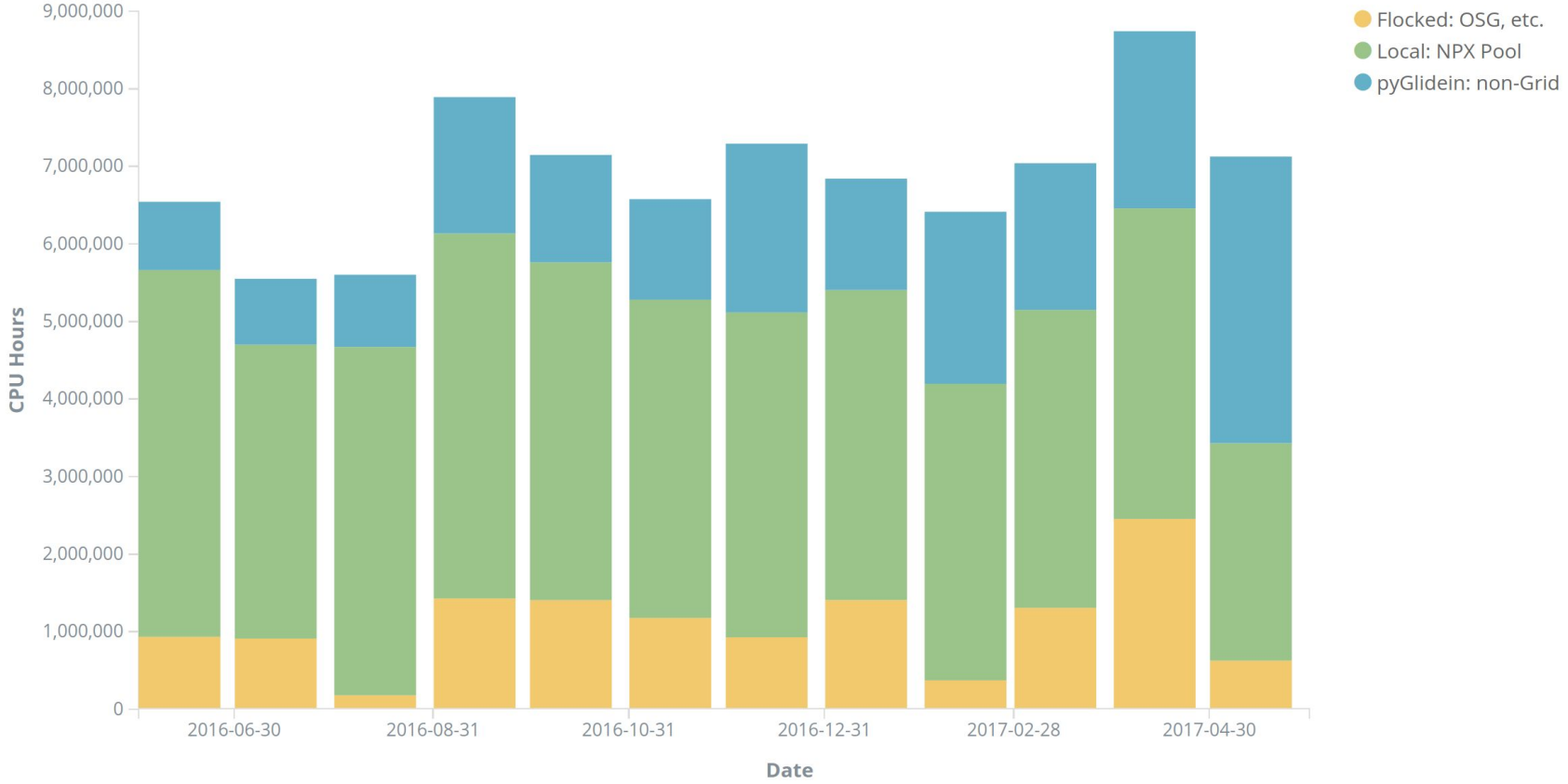
IceCube's pyGlidein software



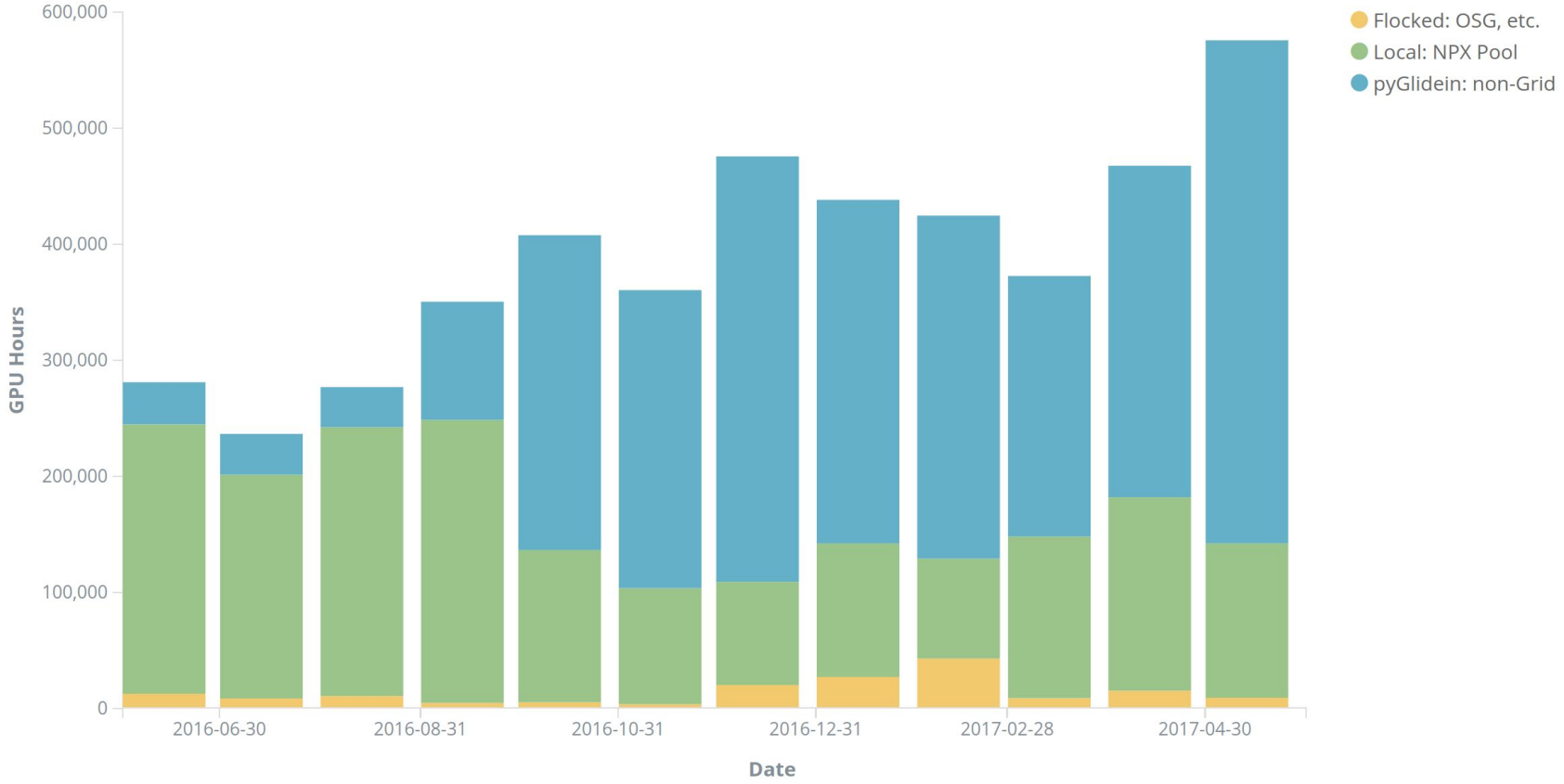
IceCube’s pyGlidein software

- Designed to make incorporation of new sites easier
  - Simplicity makes people more willing to get on-board
  - Support for site admins provided via a Slack channel
  - Local admins can specify policies at their sites
- Combines heterogeneous resources into a single HTCondor pool
  - Centralized management of priorities, accounting, monitoring
  - Up to 6000 CPUs, 600 GPUs seen in pool so far
  - Sub-collectors used to improve scalability

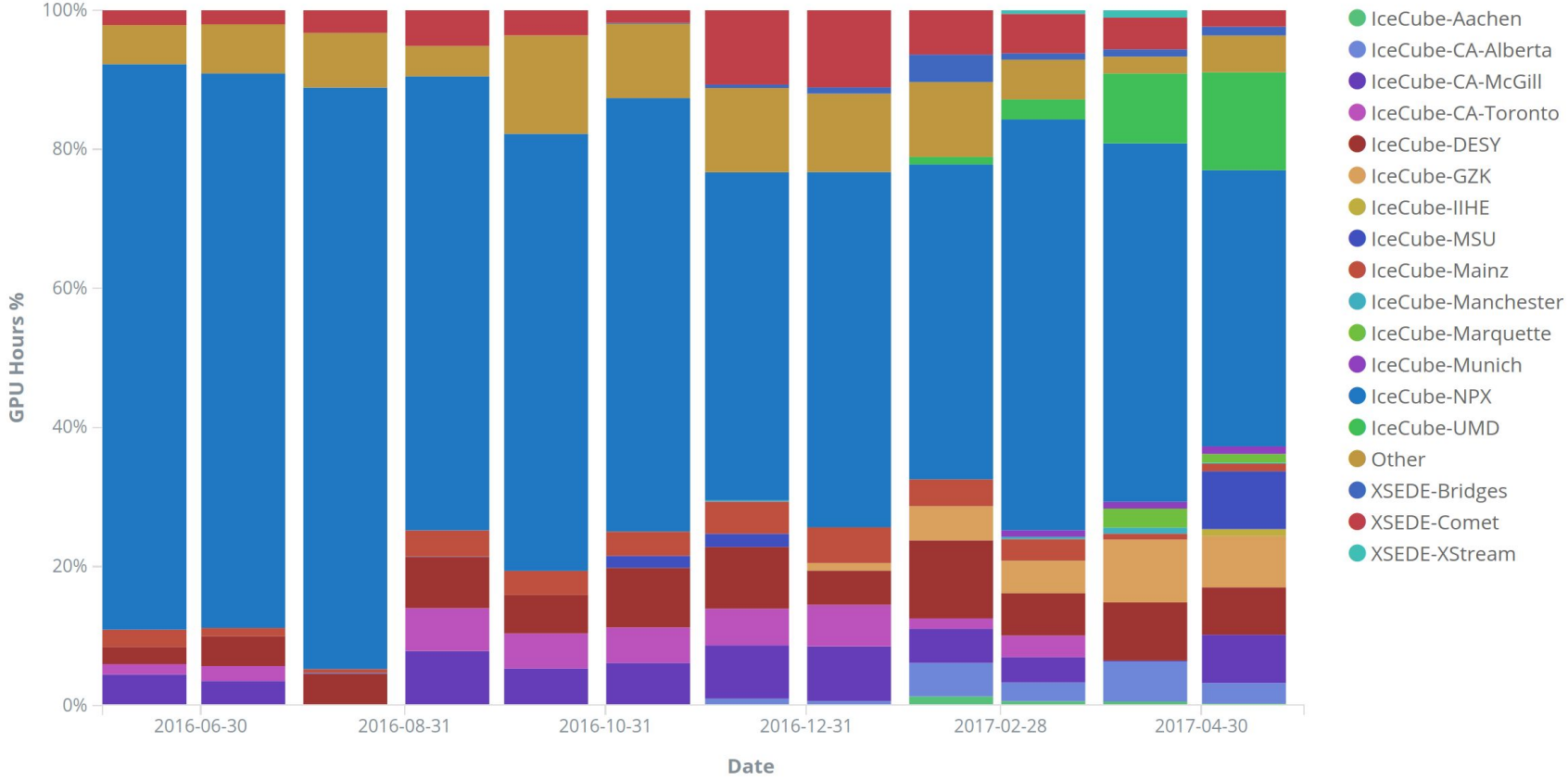
CPU hours by job type



GPU hours by job type



Relative GPU hours by site



Use case: CHTC resource allocation

- Before glideins: IceCube flocked to UW CHTC pool
  - Couldn't prioritize important workflows or UW users from our side
- Now: all CHTC resources are glideins
  - Can implement resource allocation policies on per-site basis using RANK expressions
  - IceCube submitters tag jobs based on owner's LDAP groups

```
include command : /usr/local/libexec/condor/set_affiliation.sh
SUBMIT_EXPRS = $(SUBMIT_EXPRS) Affiliation
```

  - At start-up glideins probabilistically pick a RANK to satisfy resource allocation policy

```
if [ "$rand" -lt 45 ]; then export _condor_RANK=...
elif [ "$rand" -lt 70 ]; then export _condor_RANK=...
...
```

Future work

- Auto-update pyGlidein software
  - Update process is trivial, but tracking down all admins is not
- Collect more monitoring and accounting data
  - Really want data on efficiency GPU jobs
  - Provide easy-to-use centralized monitoring for dispersed operators
- Collect glideins' HTCondor logs for debugging
  - Debugging pilot failures can be a challenge