



Computation Institute

# Connecting Campus Infrastructures with HTCondor Services

Lincoln Bryant

Computation and Enrico Fermi Institutes

University of Chicago

Condor Week 2014



THE UNIVERSITY OF  
CHICAGO

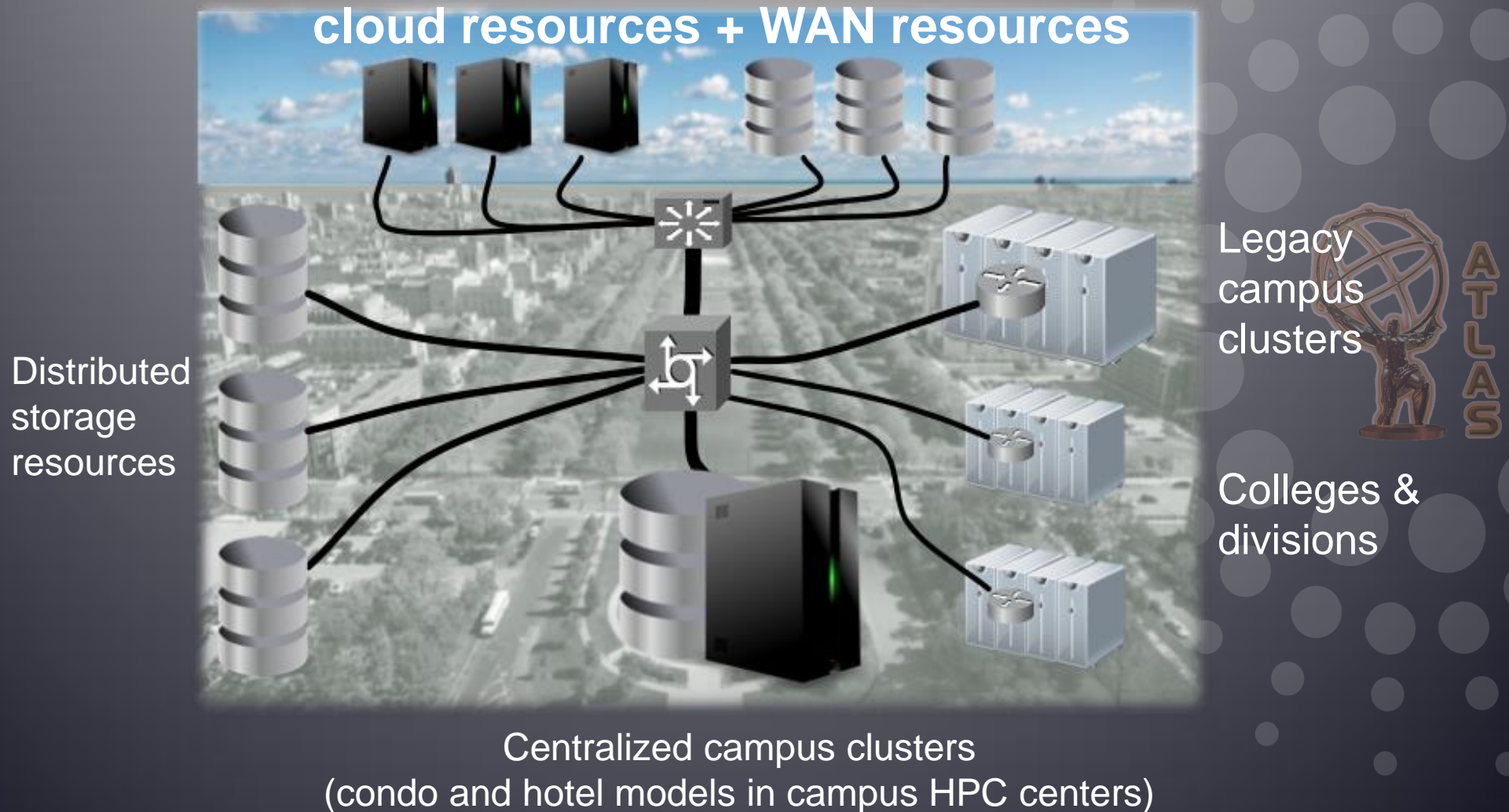
[efi.uchicago.edu](http://efi.uchicago.edu)  
[ci.uchicago.edu](http://ci.uchicago.edu)

# Outline

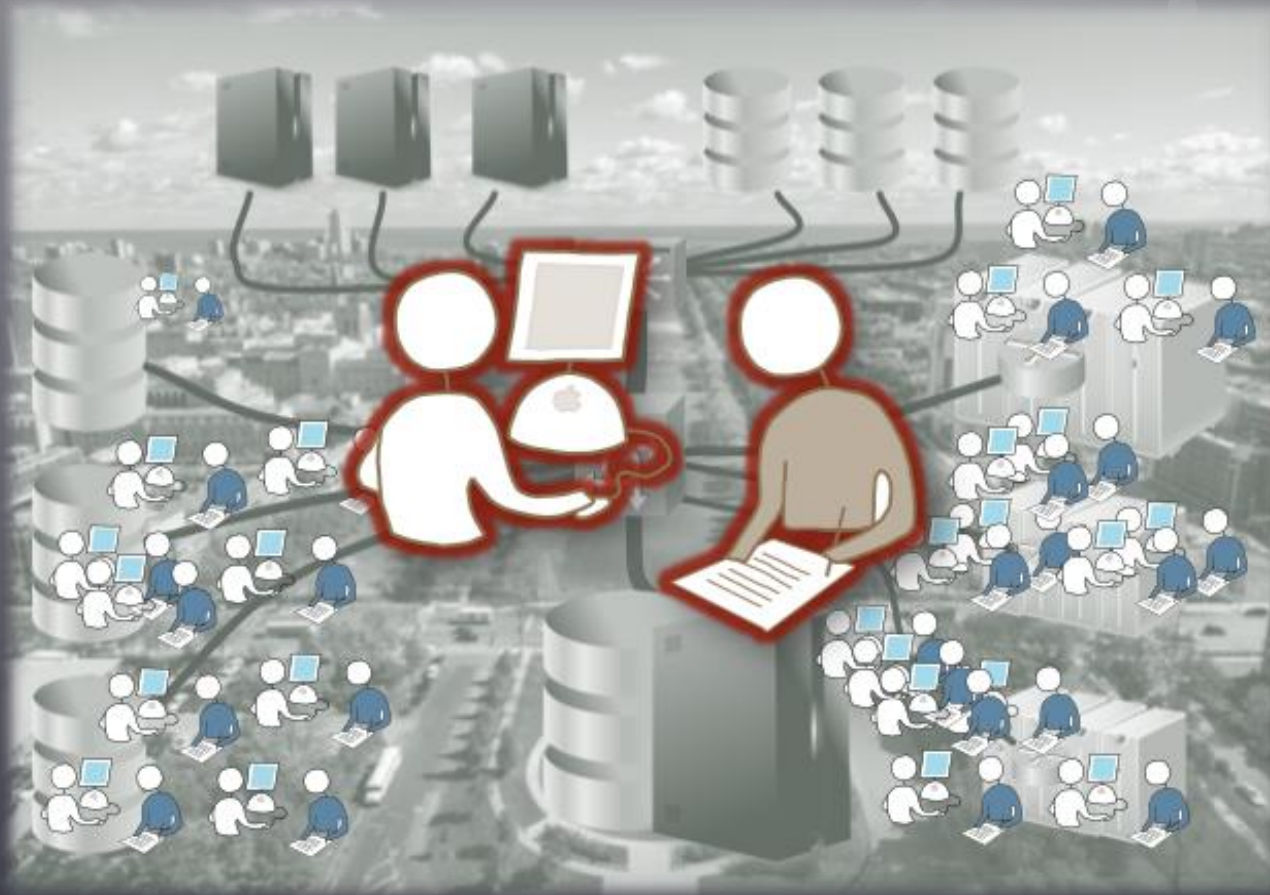
- The need for connective services for cyber infrastructure
- Technical approaches using HTCondor
- Example deployments: ATLAS Connect, Duke CI Connect
- Next steps



# Resources are distributed



# How can we transparently connect them to users?



*Approach: work from the campus perspective considering both users and resource providers*

# Providers: accountable to investors

- The BIG problem:
  - Resource providers must first meet campus investor requirements
    - Sometimes with little effort to worry about connectivity to the national ecosystem
    - They are typically slow to open up resources for sharing
- But computers depreciate in value really fast!
  - Shouldn't we obligate ourselves to make the most out of these machines?
- Can we do the heavy lifting for them?



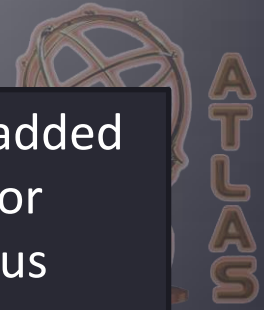
# Bring services directly to campus

**Campus  
Condo  
Cluster  
or grid**



As value added services for the campus computing center

Tightly coupled  $\leftrightarrow$  serial high throughput  
Minimize environment differences for users  
Virtually extend capacity



# Connecting Cluster Resources

- There's a large barrier to entry when we aren't even running the same scheduler software.
- The solution: Bosco
  - <http://bosco.opensciencegrid.org/>
- Bosco provides an HTCondor interface to PBS, SGE, SLURM, etc., via BLAHP<sup>[1]</sup>
- Has end-user and multi-user (service) modes
- We can provide transparent user access to multiple resources using Bosco and direct flocking



<sup>[1]</sup> [see BLAHP presentation at HTCondor Week 2013](#)

# Advantages of the Bosco approach

- From the admin perspective, we only need a user account and SSH key on the remote machine.
- Since our Bosco service appears a normal user, it's trivial to apply local policy.
- If jobs get killed, we can handle that.
  - All the better if we can use an opportunistic queue!
- Bosco also lets us have pre-job wrappers that allow us to build a comfortable nest





# Bosco as a standard tool for HTC

- In the reverse direction, we want users of remote clusters to be using our resources for spillover.
- This is an especially nice value-added service for HPC environments
  - We don't require allocations. We'll process your pleasingly parallel workloads for free.
- Everyone in HPC-land seems to be using Modules, so we have done some work to provide a Bosco module.



# Example use case:

**ATLAS**



**CONNECT**

[connect.usatlas.org](http://connect.usatlas.org)



THE UNIVERSITY OF  
**CHICAGO**

[efi.uchicago.edu](http://efi.uchicago.edu)  
[ci.uchicago.edu](http://ci.uchicago.edu)

# ATLAS Connect Service Components

- Globus Platform

- Reliable file transfer to ‘scratch’ endpoints
- Identity management
- Groups, Projects

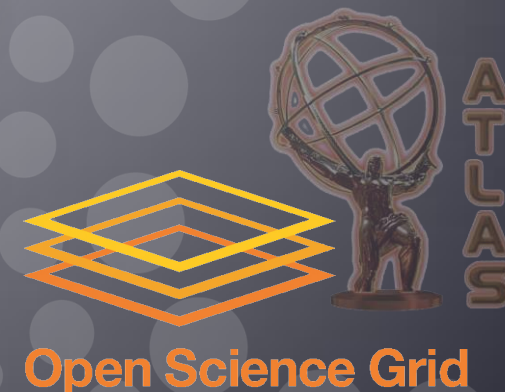


- Login host

- Auto-provisioning, quickly validating users

- Bosco-based Glidein Factories

- “Remote Cluster Connect Factories” (RCCF)
- One instance per resource target



- Gratia accounting & Cycle Server monitoring

- FAXbox storage service

- POSIX, http, XRootD, Globus access

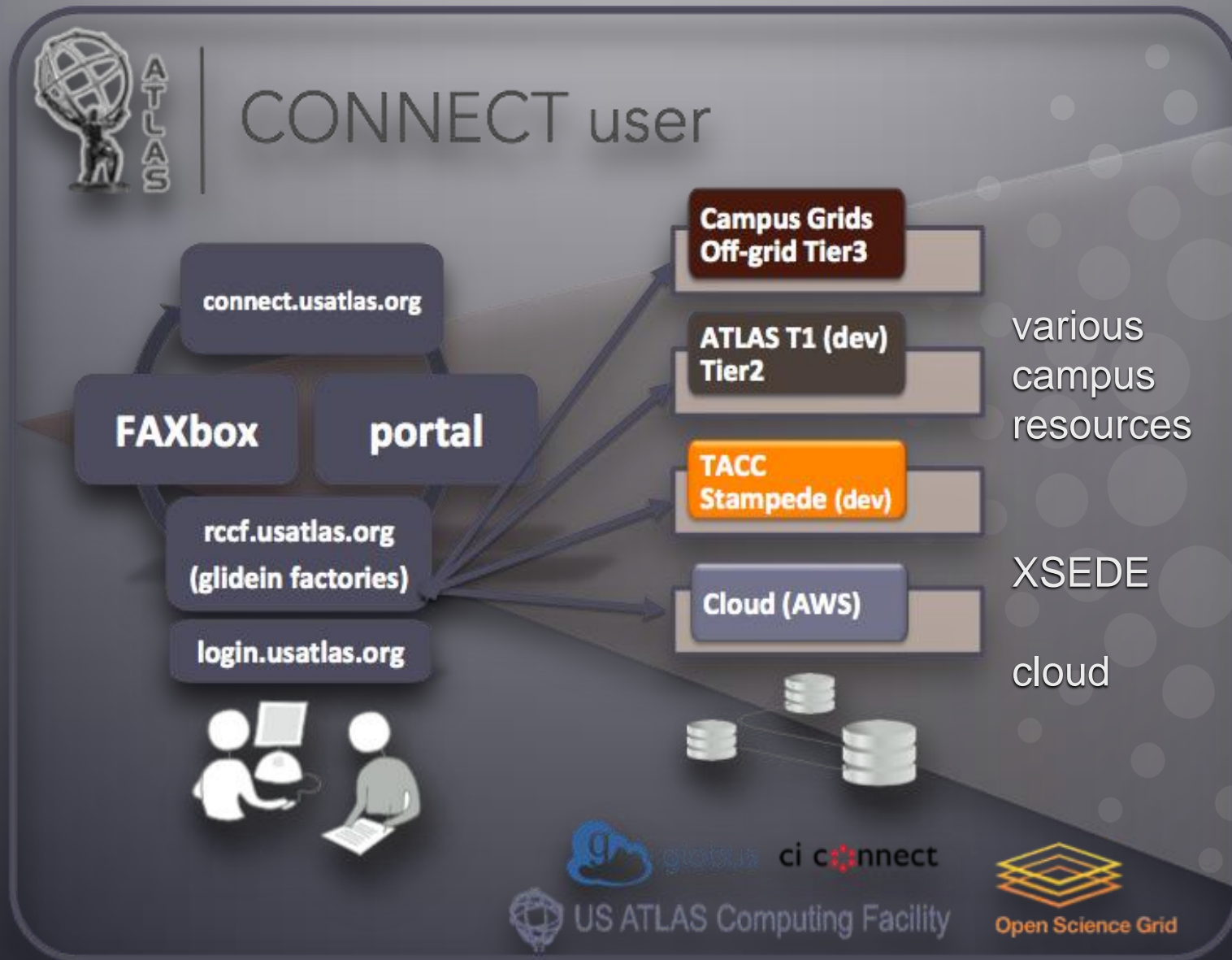


# Three Service Types for ATLAS

- **ATLAS Connect User**
  - HTCondor-based login machine for US ATLAS physicists to reach cycles at dedicated datacenters as well as departmental clusters.
- **ATLAS Connect Cluster**
  - Send jobs from local departmental cluster to ATLAS Connect infrastructure using HTCondor's flocking mechanism
- **ATLAS Connect Panda**
  - Integration with ATLAS "Panda" job workflow manager.
  - Opportunistically send simulation jobs to clouds, campus clusters, HPC resource centers



# ATLAS Connect User



users  
from 44  
institutions



THE UNIVERSITY OF  
**CHICAGO**

efi.uchicago.edu  
ci.uchicago.edu

# Looks like a very large cluster

- Users want to see quick, immediate “local” batch service
- Most Tier 3 batch use is **very** spikey
- Use opportunistic resources to elastically absorb periods of peak demand
- Easily adjust virtual pool size according to US ATLAS priorities



# Current resource targets

- Pool size varies depending on demand, matchmaking, priority at resource

## Pool Summary

Pool	Total Slots	Running	Idle	Owner	Status	Detailed View
CSU Fresno Factory	248	248	0	0		<a href="#">Usage</a> <a href="#">Jobs</a>
Great Lakes Tier 2 Factory	643	639	4	0		<a href="#">Usage</a> <a href="#">Jobs</a>
Midwest Tier 2 Factory	1992	1992	0	0		<a href="#">Usage</a> <a href="#">Jobs</a>
Southwest Tier 2 Factory	0	0	0	0		<a href="#">Usage</a> <a href="#">Jobs</a>
TACC Stampede	0	0	0	0		<a href="#">Usage</a> <a href="#">Jobs</a>
UC3	528	528	0	0		<a href="#">Usage</a> <a href="#">Jobs</a>
UChicago RCC Factory	0	0	0	0		<a href="#">Usage</a> <a href="#">Jobs</a>
<b>Total</b>	<b>3411</b>	<b>3407</b>	<b>4</b>	<b>0</b>		<a href="#">Usage</a> <a href="#">Jobs</a>



Jobs by State



Jobs by Owner



Slots by State



Slots by Owner

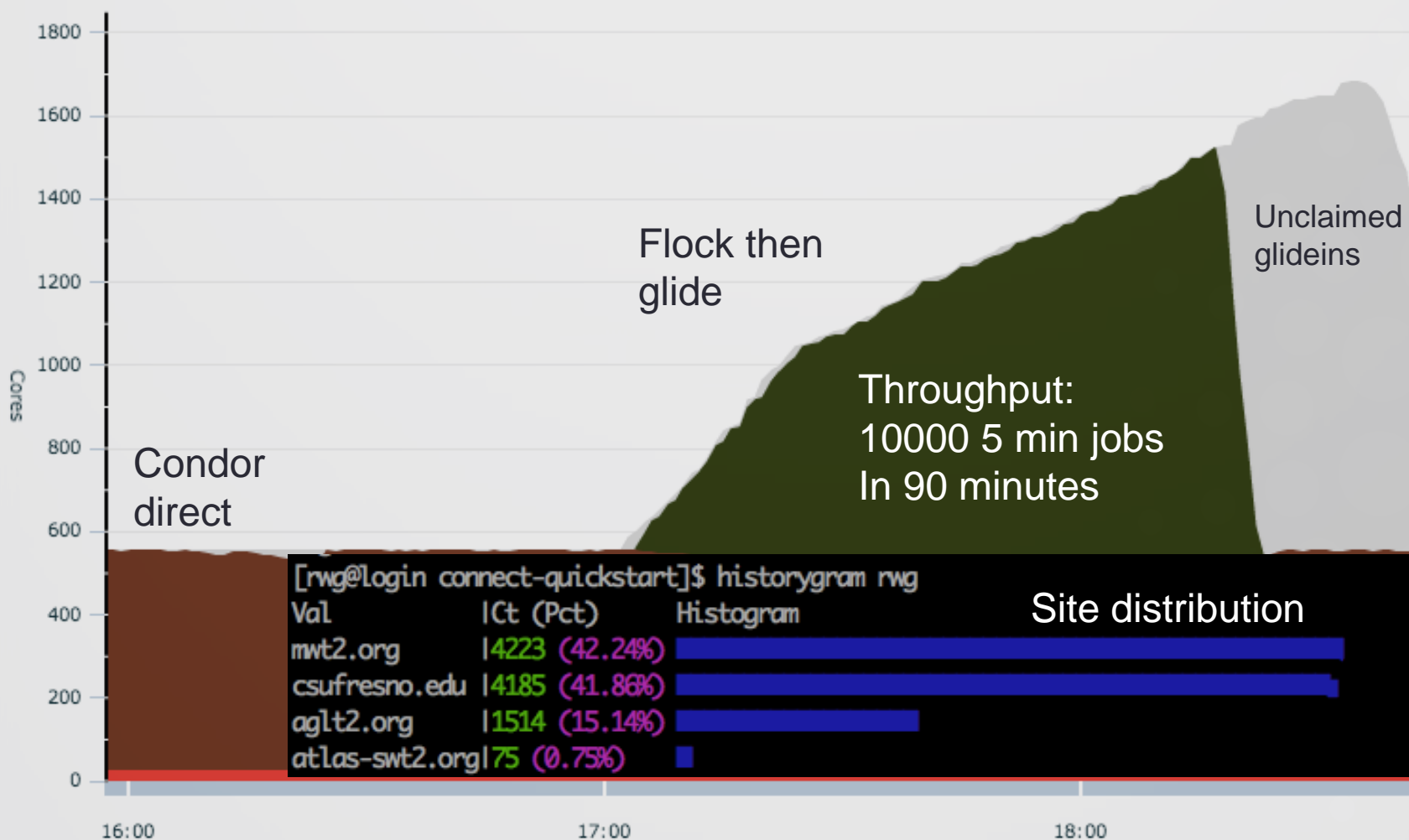


# Connect is very quick relative to grid

## Show: Historical grid usage in all pools

Time Frame: 3 Hours | Day | Week | Month

View as: Area | Line



**Legend**

- Used by owner
- vpa
- rwg
- Unclaimed

```
[rwg@login connect-quickstart]$ historygram rwg
```

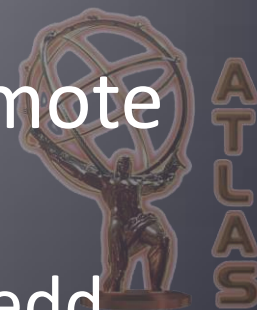
Val	Ct (Pct)	Histogram
mwt2.org	4223 (42.24%)	[Bar]
csufresno.edu	4185 (41.86%)	[Bar]
aglt2.org	1514 (15.14%)	[Bar]
atlas-swt2.org	75 (0.75%)	[Bar]

### Site distribution



# ATLAS Connect Cluster

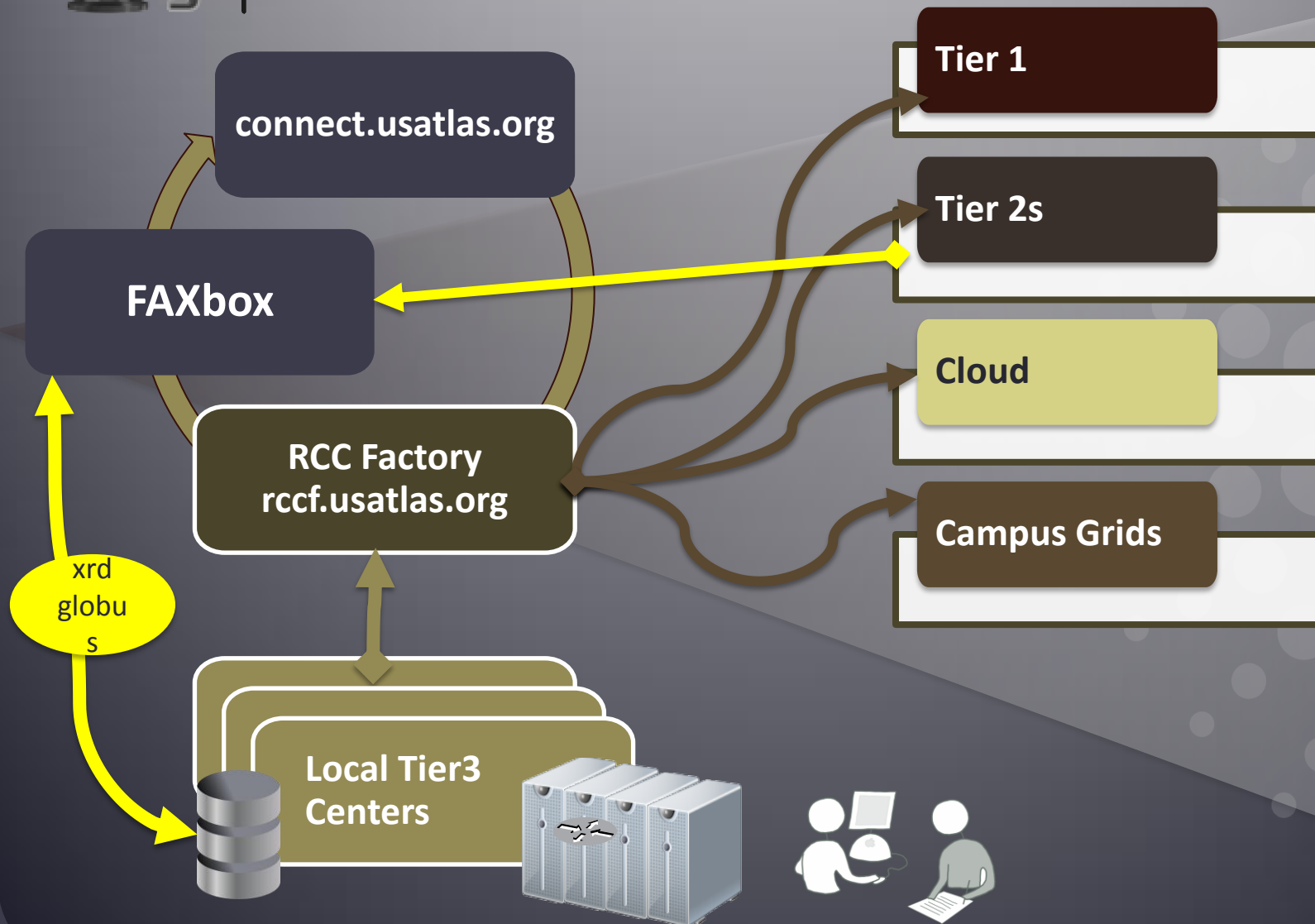
- ATLAS physicists on their institution's "Tier 3" cluster can submit into ATLAS Connect without ever leaving home.
  - Submissions can be overflow or targeted using HTCondor class ads.
- Admins configure HTCondor to flock to the Remote Cluster Connect Factories (RCCF)
  - Configuration happens on the local HTCondor schedd
  - Firewall rules etc. opened as necessary.
- The RCCF service can reach any of the targets in the ATLAS Connect system
  - Easily reconfigured for periods of high demand





ATLAS

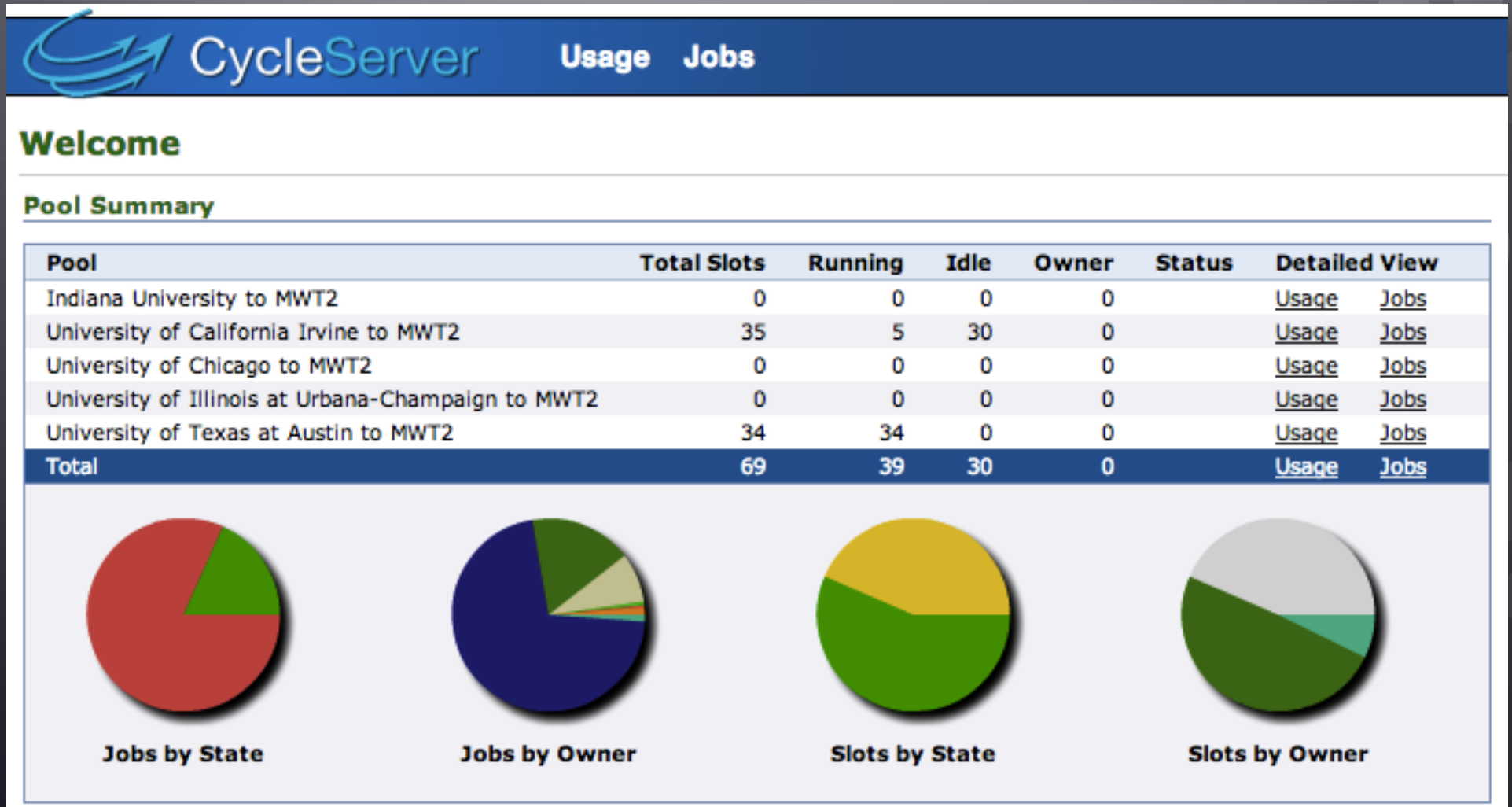
# CONNECT cluster



ATLAS

# ATLAS Connect Cluster in use

- Five clusters configured in this way so far
- Works well, very low maintenance

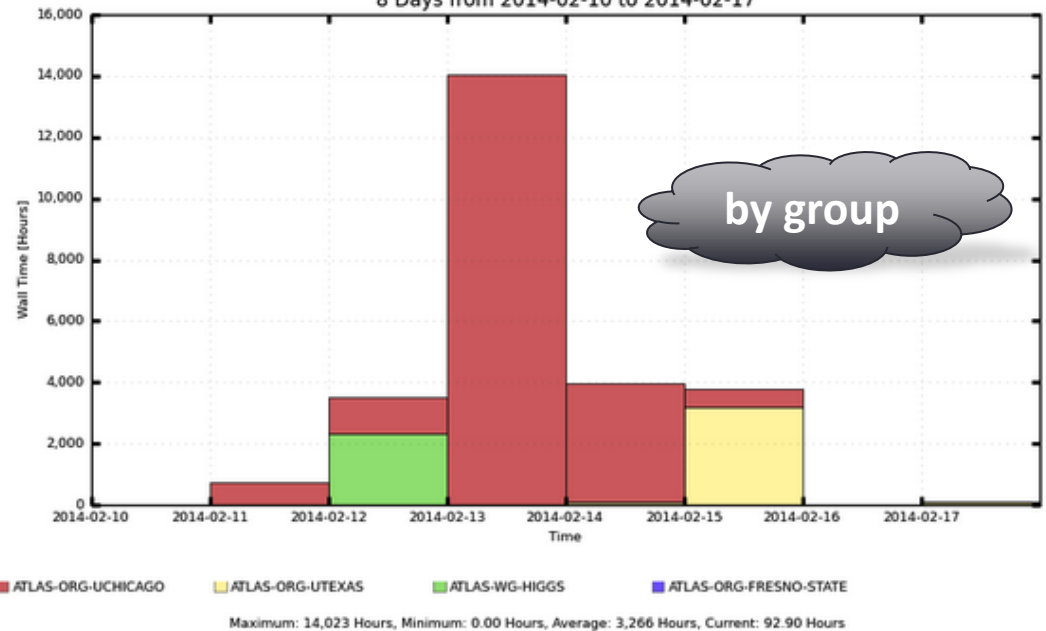


ATLAS

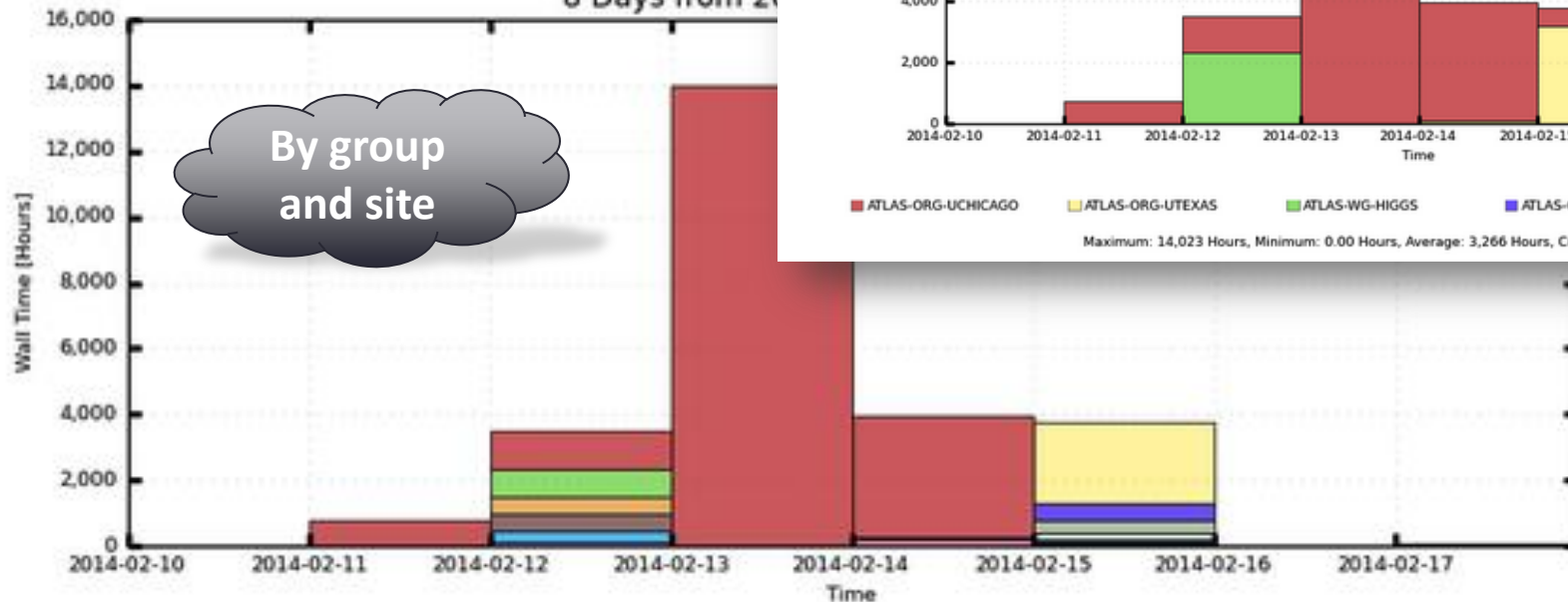
# Early adopters ramping up

Last update: Mon Feb 17 09:10:01 CST 2014

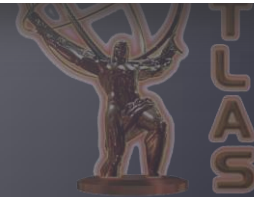
Daily Hours By Project  
8 Days from 2014-02-10 to 2014-02-17



Daily Hours  
8 Days from 2014-02-10 to 2014-02-17



- ATLAS-ORG-UCHICAGO , ruc.mwt2@uct2-gk.mwt2.org/condor
  - ATLAS-WG-HIGGS , login.atlas.ci-connect.net
  - ATLAS-WG-HIGGS , ruc.mwt2@uct2-gk.mwt2.org/condor
  - ATLAS-ORG-UTEXAS , ruc.mwt2@iut2-gk.mwt2.org/condor
  - ATLAS-ORG-UTEXAS , ruc.mwt2@uct2-gk.mwt2.org/condor
  - ATLAS-WG-HIGGS , ruc.mwt2@iut2-gk.mwt2.org/condor
  - ATLAS-ORG-UCHICAGO , fresnoatlas@t3head.atlas.csufresno.edu/condor
  - ATLAS-ORG-UTEXAS , fresnoatlas@t3head.atlas.csufresno.edu/condor
  - ATLAS-ORG-UCHICAGO , atlasconnect@gate04.aglt2.org/condor
  - ATLAS-ORG-UCHICAGO , uct2-bosco.uchicago.edu:11122?sock=collector
  - ATLAS-ORG-UTEXAS , uc3-mgt.mwt2.org
  - ATLAS-ORG-UCHICAGO , uc3-mgt.mwt2.org
  - ATLAS-WG-HIGGS , atlasconnect@gate04.aglt2.org/condor
  - ATLAS-WG-HIGGS , fresnoatlas@t3head.atlas.csufresno.edu/condor
  - ATLAS-ORG-UCHICAGO , ruc.mwt2@iut2-gk.mwt2.org/condor
  - ATLAS-ORG-UTEXAS , ruc.mwt2@mwt2-gk.campuscluster.illinois.edu/condor
  - ATLAS-ORG-FRESNO-STATE , fresnoatlas@t3head.atlas.csufresno.edu/condor
  - ATLAS-ORG-UTEXAS , atlasconnect@gate04.aglt2.org/condor
  - ATLAS-ORG-UTEXAS , login.atlas.ci-connect.net
  - ATLAS-ORG-UCHICAGO , login.atlas.ci-connect.net
- Maximum: 14,023 Hours, Minimum: 0.00 Hours, Average: 3,266 Hours, Current: 92.90 Hours



# Extending ATLAS Connect to more resources

- Some of our colleagues have access to “off-grid” resources such as supercomputers and campus clusters.
- We can’t expect any of these resources to have our software prerequisites.
- By combining HTCondor and Parrot<sup>[2]</sup>, we can run ATLAS jobs on these kinds of resources.
- Parrot allows us to:
  - Access our ATLAS software repositories
  - Play some LD\_PRELOAD tricks to access system dependencies that we need



<sup>[2]</sup> [see the Parrot homepage](#)

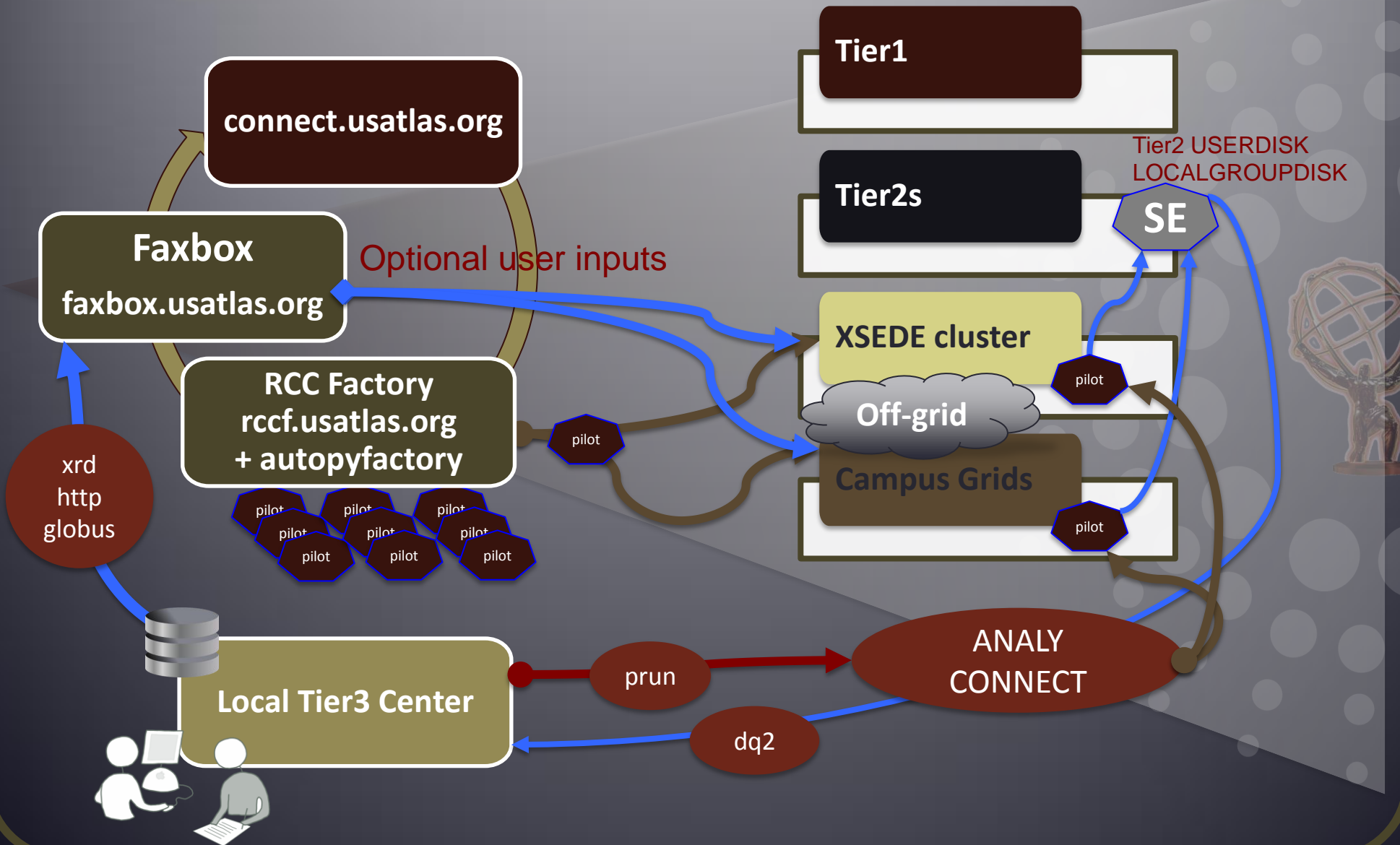
# ATLAS Connect Panda

- We've been able to integrate ATLAS Connect with Panda (ATLAS grid workflow manager)
  - ATLAS production (simulation) jobs are fairly well understood in terms of requirements.
  - A new opportunistic queue is created in the workflow system and pointed at HPC resources
  - Jobs come through AutoPyFactory<sup>[3]</sup>, and get locally submitted as HTCondor jobs
  - Pre-job wrappers use Parrot to set up an environment that looks like an ATLAS worker node for the jobs.





# CONNECT panda Analysis queue



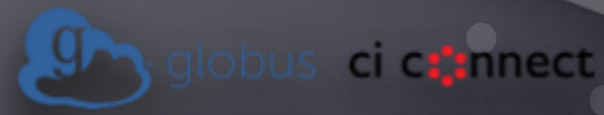
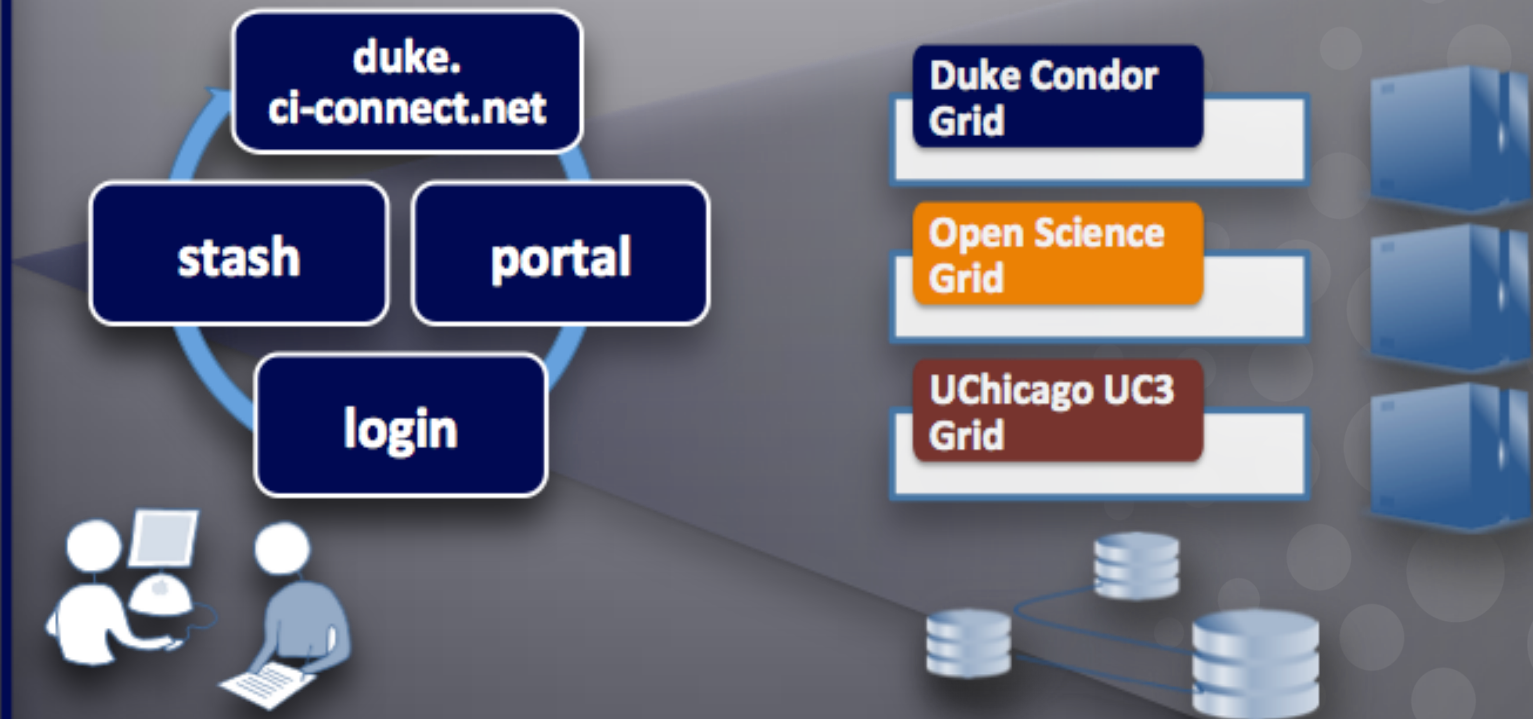
# Back to Campuses

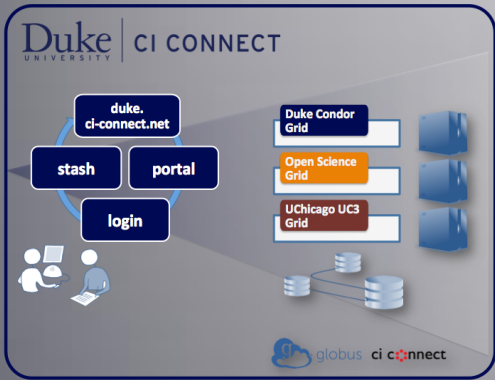
ci c:connect

- We're also providing value-added services for existing campus clusters (ci-connect.net)
- One of our early adopters: Duke University
  - Login node and scratch storage provisioned for Duke users
  - Integrated into Duke "BlueDevil Grid" campus grid.
  - Also provides a submit interface into the Open Science Grid
  - Bridged to opportunistic resources at the University of Chicago



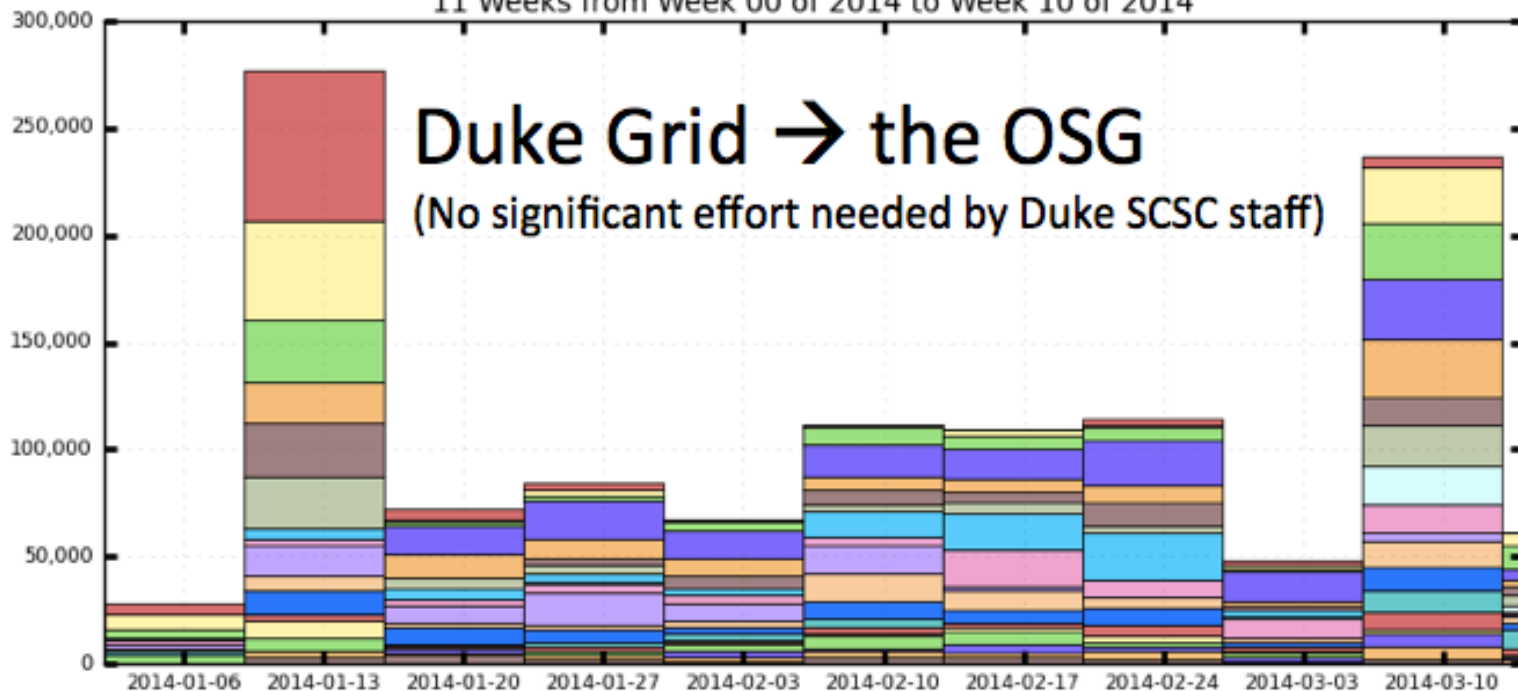






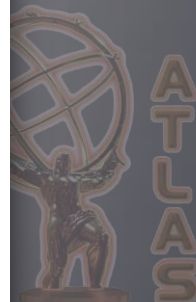
## WMS Hours Spent on Jobs By Facility (Glidein)

11 Weeks from Week 00 of 2014 to Week 10 of 2014

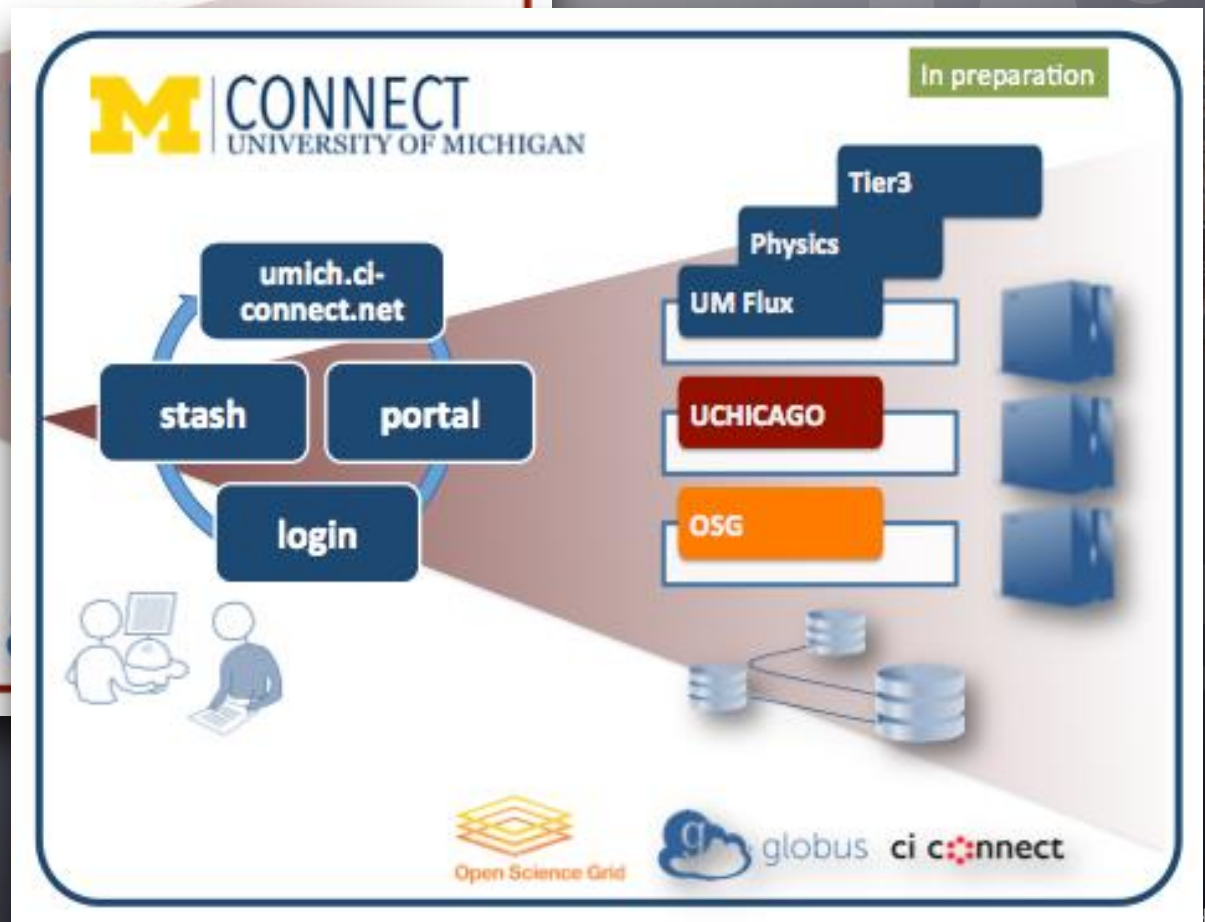
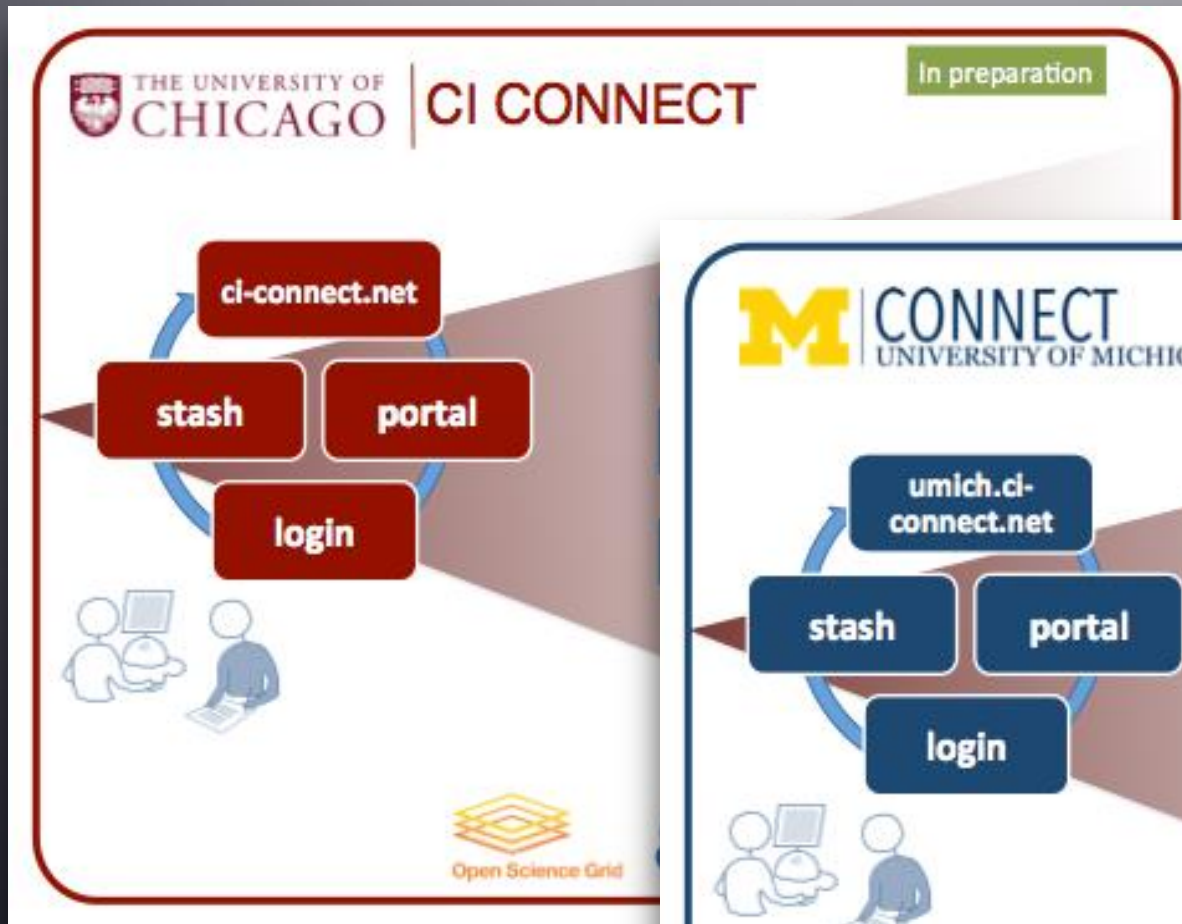


- |                   |             |        |                           |
|-------------------|-------------|--------|---------------------------|
| Crane-CE1         | Nebraska    | UCSDT2 | login.duke.ci-connect.net |
| Sandhills         | NWICG_NDCMS | Tusker | SMU_HPC                   |
| BU_ATLAS_Tier2    | GlueX_VOMRS | MWT2   | Other                     |
| GridUNESP_CENTRAL | GLOW        | AGLT2  | IU_OSG                    |
| QT_CMS_T2         | Purdue-RCAC | UCD    | cinvestav                 |

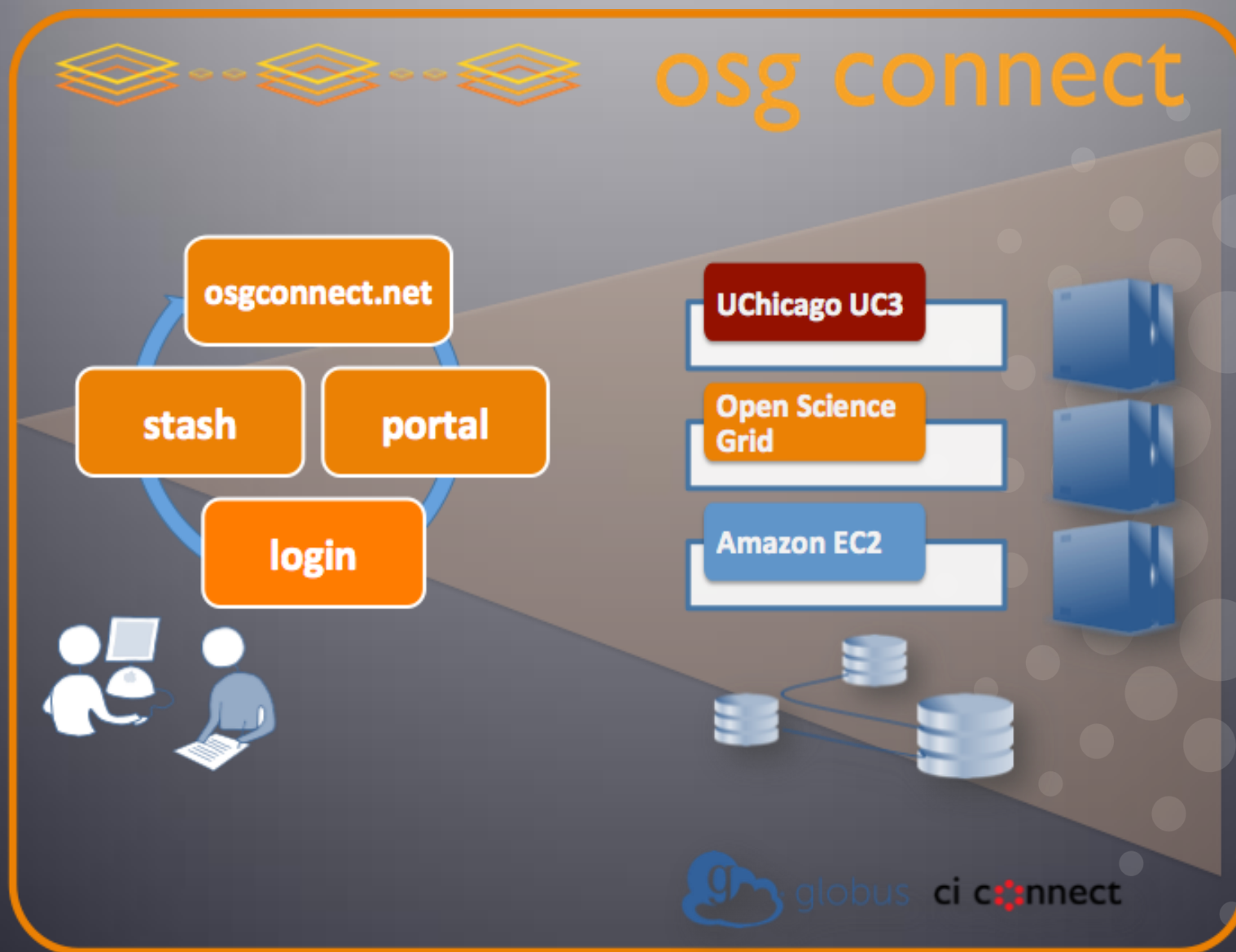
Maximum: 276,604 , Minimum: 28,237 , Average: 110,106 , Current: 61,593



# CI Connect Services in Preparation



# Where it started: a login service for OSG



# What does the next-generation look like?

- Why can't identity behave more like eduroam?
  - I want my login account to be my campus identity. Period. No new accounts.
- Can we take a hard look at solving some of the problems with software distribution?
  - Are virtual machines the way forward? What about containers?
  - We can play games with static compilation and LD\_PRELOAD, but it sure would be nice to have something that looks like your home environment!
- Data access is still not dead simple
  - Focus on data **delivery**, not data management



# Thank you!



# Acknowledgements



- Dave Lesny – UIUC (MWT2)
- David Champion – UChicago (MWT2)
- Steve Tuecke, Rachana Ananthakrishnan – (UChicago Globus)
- Ilija Vukotic (UChicago ATLAS)
- Suchandra Thapa (UChicago OSG)
- Peter Onysis – UTexas
- Jim Basney (CI-Logon) & InCommon Federation
- & of course the HTCondor and OSG teams



osg connect