



State of the Open Science Grid

OSG All Hands Meeting
March 25th, 2015

Frank Würthwein
OSG Executive Director





Many Thanks to The Organizers !!!

Especially Pamela, Kristian, and Clemmie

- Founding Member of OSG in ~2003
- Stepped down as Executive Director of OSG to be the new U.S. CMS Operations Program Manager after providing leadership to OSG from 2012 – 2015.

Many thanks Lothar !!!

- Vision
- State of the Art
- Future Progress to watch



Vision

Imagine ...

- **All clusters at Universities & National Labs are shared.**
 - Sharing policy is locally controlled.
 - All owners want to share to maximize the benefit to all.
- **Researcher use a single interface to use resources ...**
 - ... they own
 - ... others are willing to share
 - ... they have an allocation on
 - ... they buy from a commercial (cloud) provider

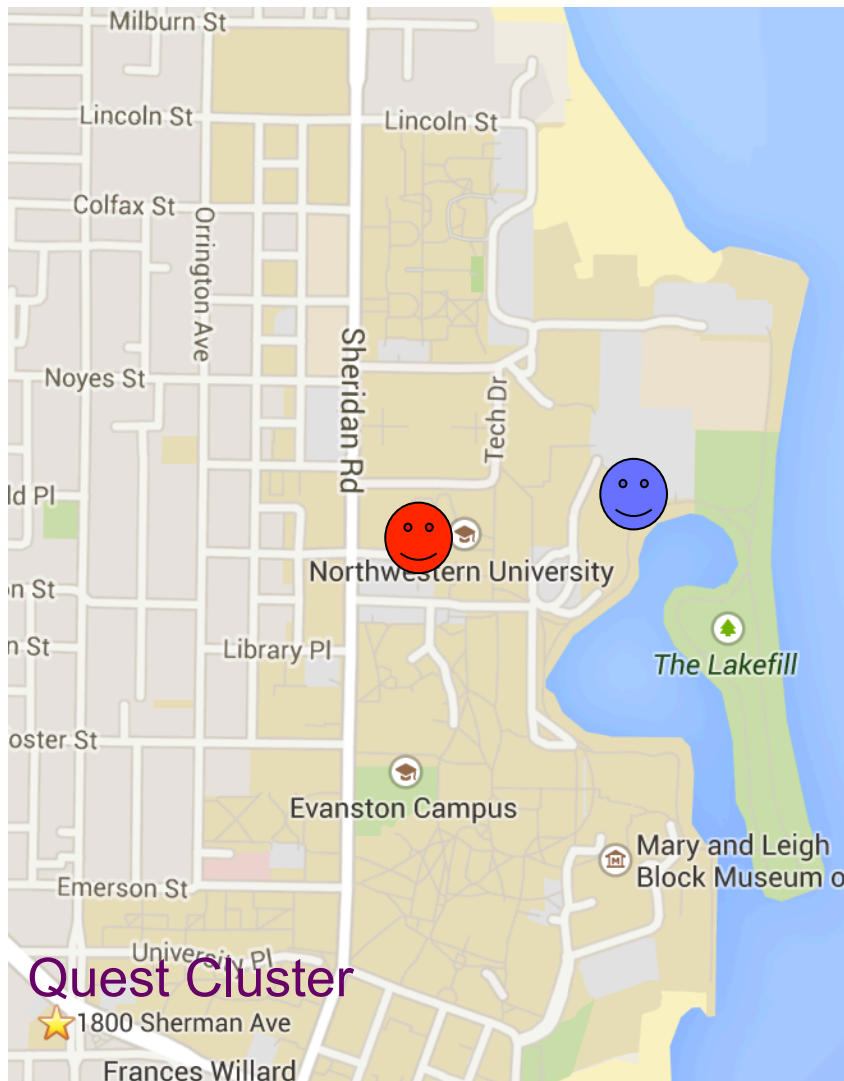
**OSG focuses on making this technically possible
for Distributed High Throughput Computing**

- Operate a shared Production Infrastructure
 - collaborate with partners that want to share their hardware => **Open Facility**
- Advance a shared Software Infrastructure
 - collaborate with partners that want to share their software => **Open Software Stack**
- Disseminate knowledge across Researchers, IT professionals & Software developers.
 - collaborate with partners that want to share their ideas => **Open Ecosystem**

3 Example Views of OSG

- **Single PI Perspective**
 - OSG-Connect (see session yesterday)
 - OSG-XD
 - OSG operates login node, disk, and provisions resources across the facility for single PIs and small groups.
 - Access either via XSEDE allocation or “word of mouth”.
- **IT Organization Perspective**
- **Large Scale Research Community Perspective**

IT Organization Perspective



Joe is an economist using DHTC to model banking networks. He develops software on a Unix server in his office and needs much larger resources to run his models.



Jane is an Astrophysicist analyzing DES data. She has a small cluster near her office, and access to Quest as well as an XSEDE allocation at SDSC.

Both Joe and Jane want to do their work transparently from the infrastructure they own, to Campus resources, to national resources, and bring home their derived data products for further analysis.

OSG provides expertise, software, and production services.

LSST will publish a 15PB dataset with primary location at NCSA.

Scientists all over the US will need different $O(1\%)$ slices of that data for their research.

Joe, a new faculty at NW works in a small team with Jane, his former advisor at Harvard.

In addition to resources at NW & Harvard, they want to use some of Joe's startup to buy additional resources at AWS.

OSG provides expertise, software, and production services.

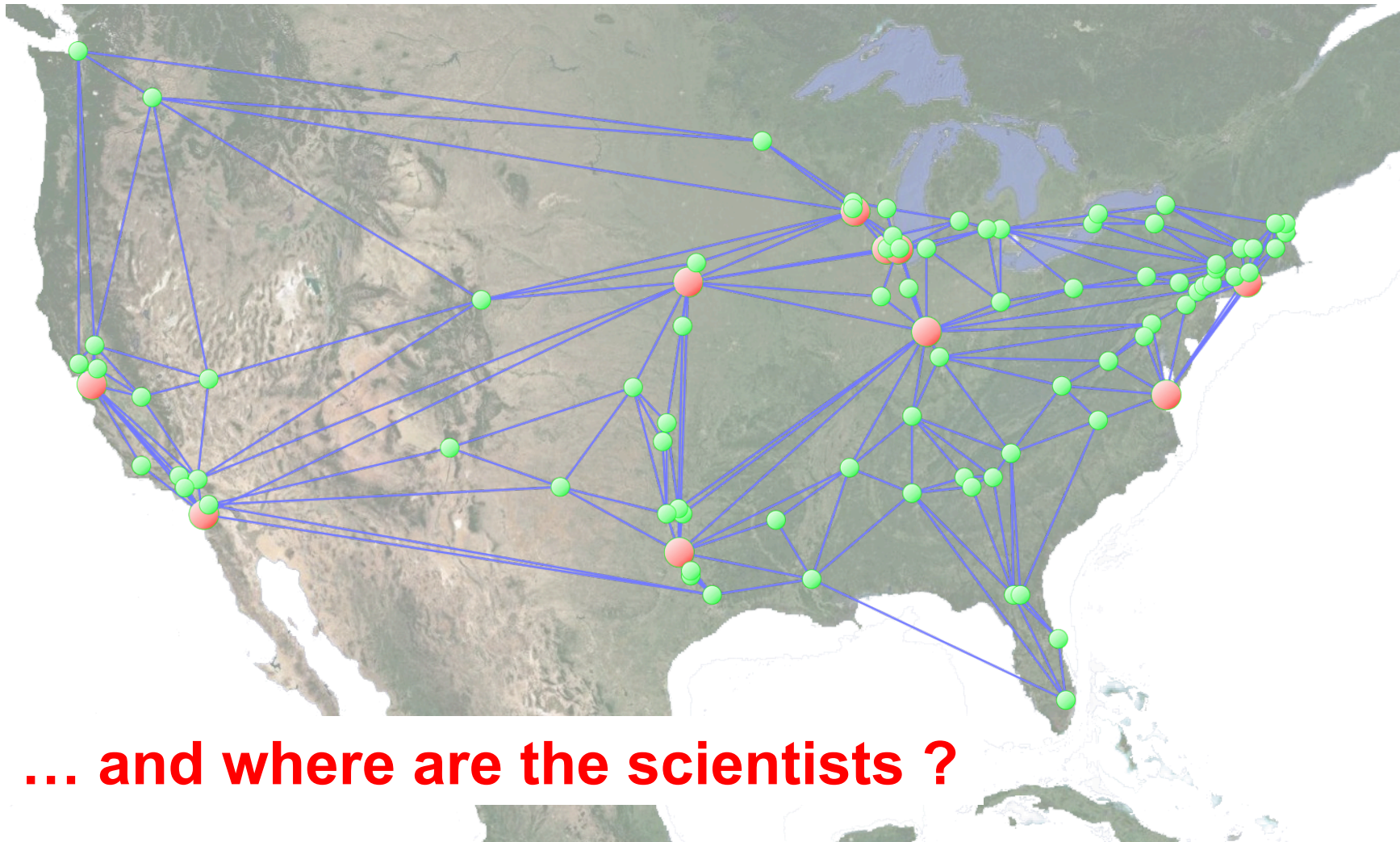


State of the Art



**Federating 140+ clusters,
most of which are located in the U.S.**

Federated Clusters in 40/52 States

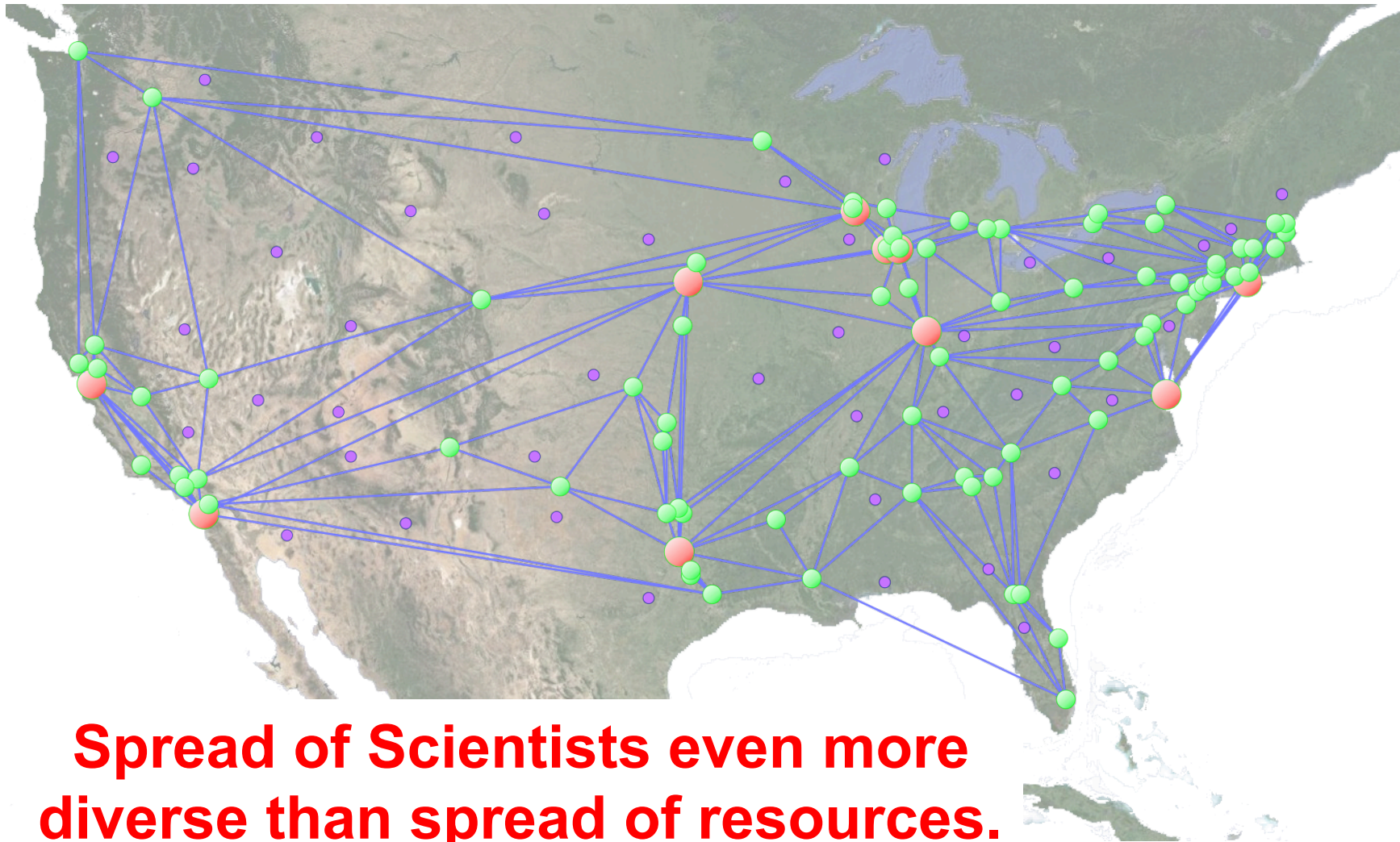


... and where are the scientists ?



Open Science Grid

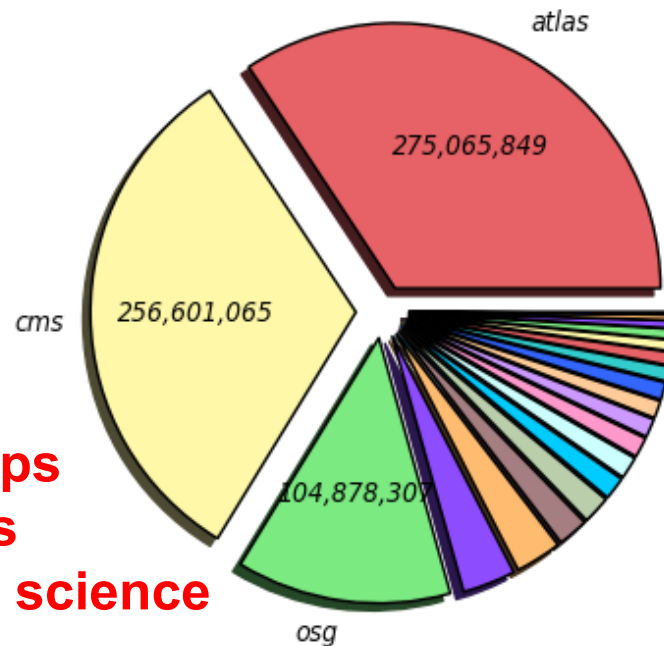
Scientists on campuses everywhere use OSG



OSG Hours in 2014

800 Million hours in 2014, or an average of ~ 100,000 cores 24x7x365

Wall Hours by VO (Sum: 802,273,116 Hours)
53 Weeks from Week 04 of 2014 to Week 04 of 2015



66% ATLAS & CMS
13% 18 other HENP groups
10% various life sciences
11% all other, incl. social science

| | | | | |
|---------------------|--------------------|-------------------|------------------------|------------------------|
| atlas (275,065,849) | cms (256,601,066) | osg (104,878,308) | glow (25,883,802) | dosar (22,625,611) |
| alice (14,895,246) | minos (13,661,032) | nova (11,104,055) | gridunesp (10,794,045) | mu2e (9,838,435) |
| minerva (8,956,031) | dzero (8,501,805) | cdf (8,443,062) | Other (7,121,334) | Other (7,121,334) |
| lbne (5,237,510) | gluex (4,935,380) | star (3,585,428) | sbgrid (2,560,975) | microboone (1,286,015) |



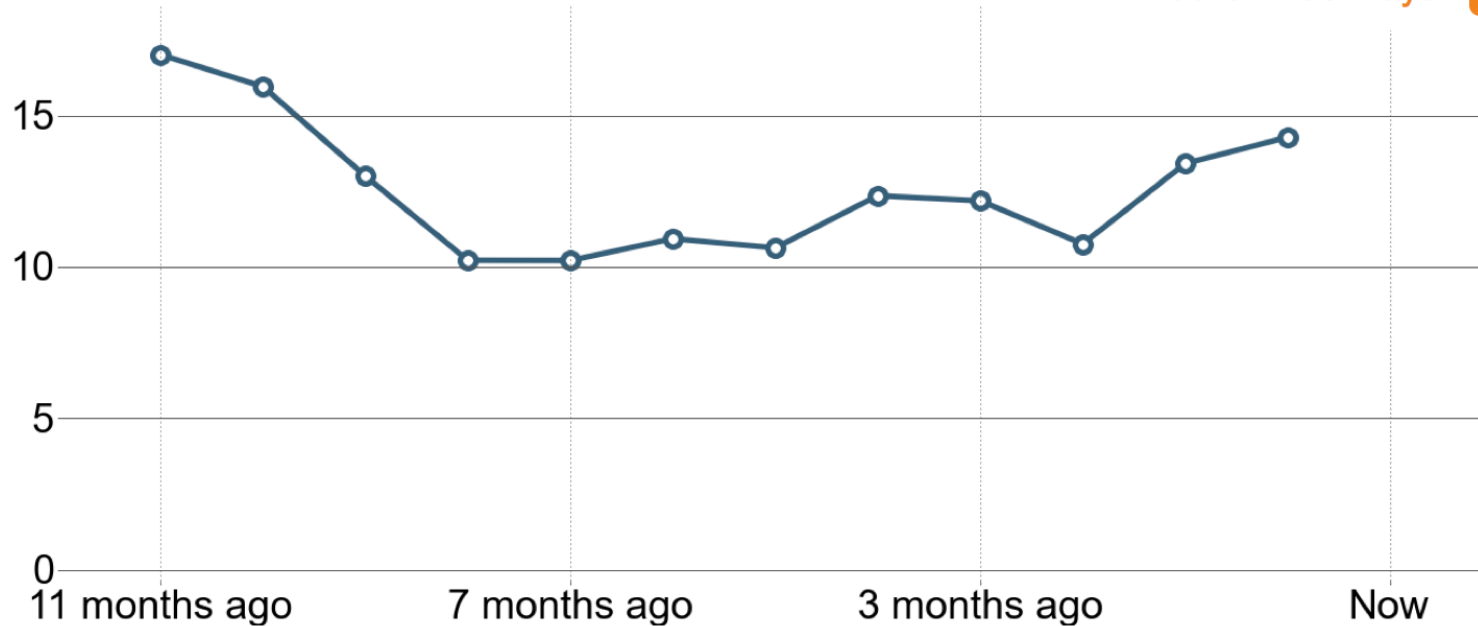
Data Transferred



Status Map Jobs CPU Hours Transfers **TB Transferred**

Petabytes of Transfers/Month

24 Hours 30 Days **12 Months**



Large Volumes are unique capability of the large LHC experiments.
.... large geek gap between LHC and the rest of Science



Future Progress to watch out for

3 Example Technology Developments

PerfSONAR Network Performance Measurements

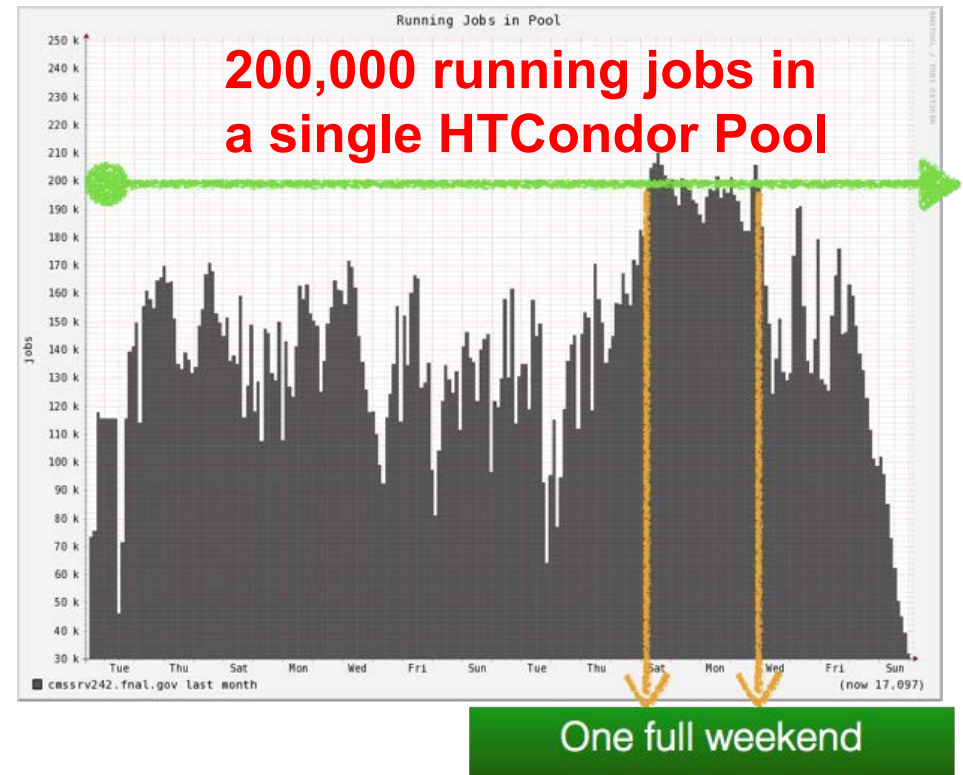


Goals:

- > Debugging Network Issues
- > Long Term performance repository
- => source of data for CS R&D

For Details, see talk by McKee later today.

- Example HTCondor:
 - until 2014, the largest pool in operations was about ~40k jobs.



OSG advancing state of the art in DHTC software practices.

Deploying new functionality

- OSG to deploy data caches to extend DHTC for all of science from ~5GB input data per workflow to ~5TB.
- Building on DOE & NSF investment in infrastructure software.
- See Talk by B.Bockelman on StashCache in Tuesday plenary for details.

OSG advancing state of the art in DHTC software practices.

- OSG continues to advance Science through DHTC & the power of sharing.
 - Ready for the start of Run 2 of the LHC
 - Ready for the Intensity Frontier as new major stakeholder
 - Ready for making a big leap forwards in shrinking the geek gap in data analysis
 - Ready to advance Bioinformatics through DHTC
 - Ready to further extend the sharing paradigm across campuses.