



# Campus Update: Holland Computing Center

Dr. David R. Swanson  
Director, Holland Computing Center



# HCC @ NU

- Holland Computing Center has a University-wide mission (UNL, UNO, UNMC) to
  - Facilitate computational science and engineering research and other scholarly activity that requires advanced CyberInfrastructure
  - Provide expertise in and maintain local resources for HPC, HTC, large data sets and advanced networking
  - Engage and train NU researchers, students, and other state communities

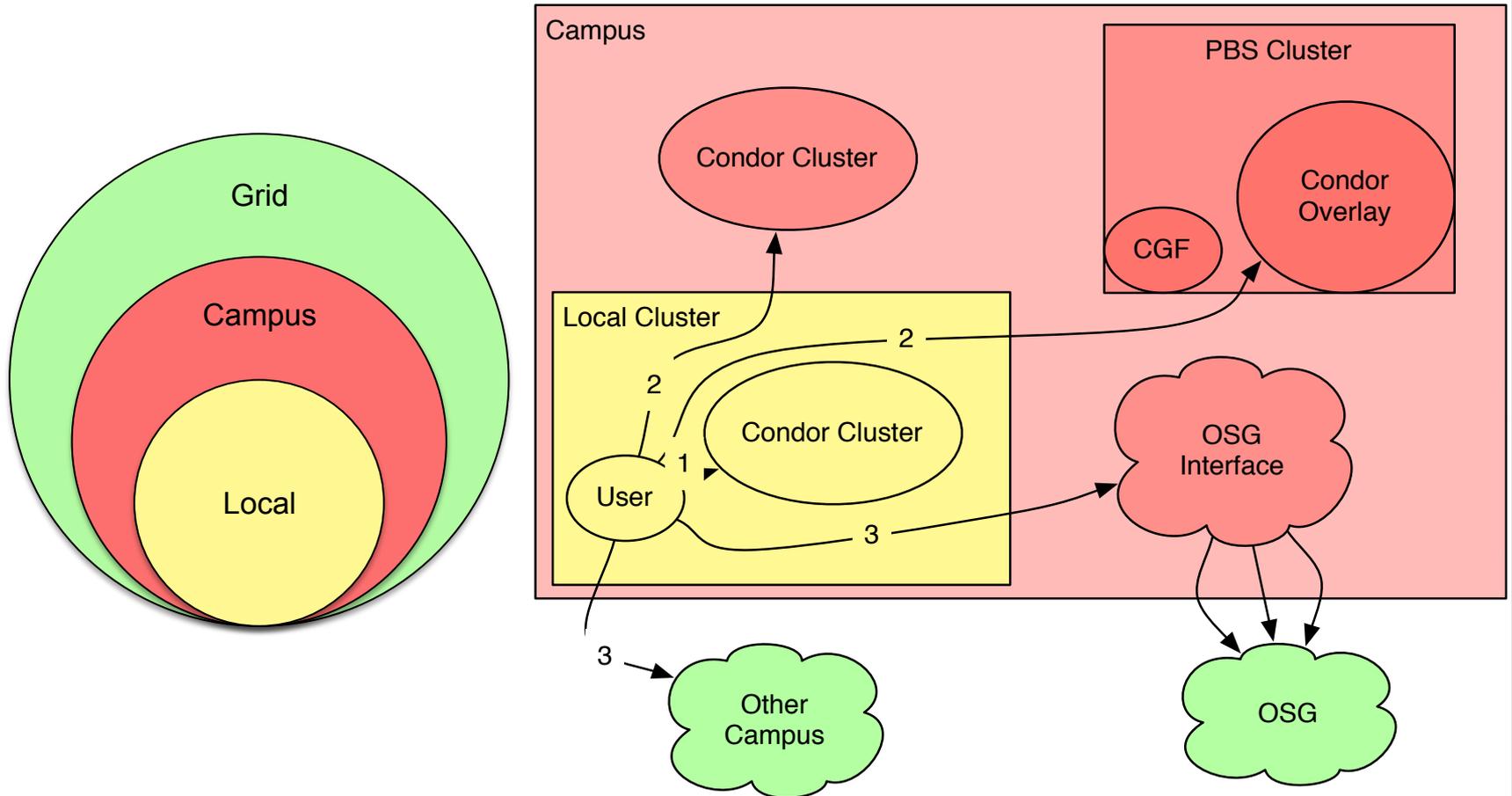


# HCC @ NU

- Holland Computing Center resources are combined into an HTC campus grid
  - 10,000 cores, 500 TB in Omaha
  - 5,000 cores, 3 PB in Lincoln
  - All tied together via a single submission protocol using OSG software stack
  - Straightforward to expand to OSG sites across the country, as well as to EC2 (cloud)
  - HPC jobs get priority; HTC ensures high utilization

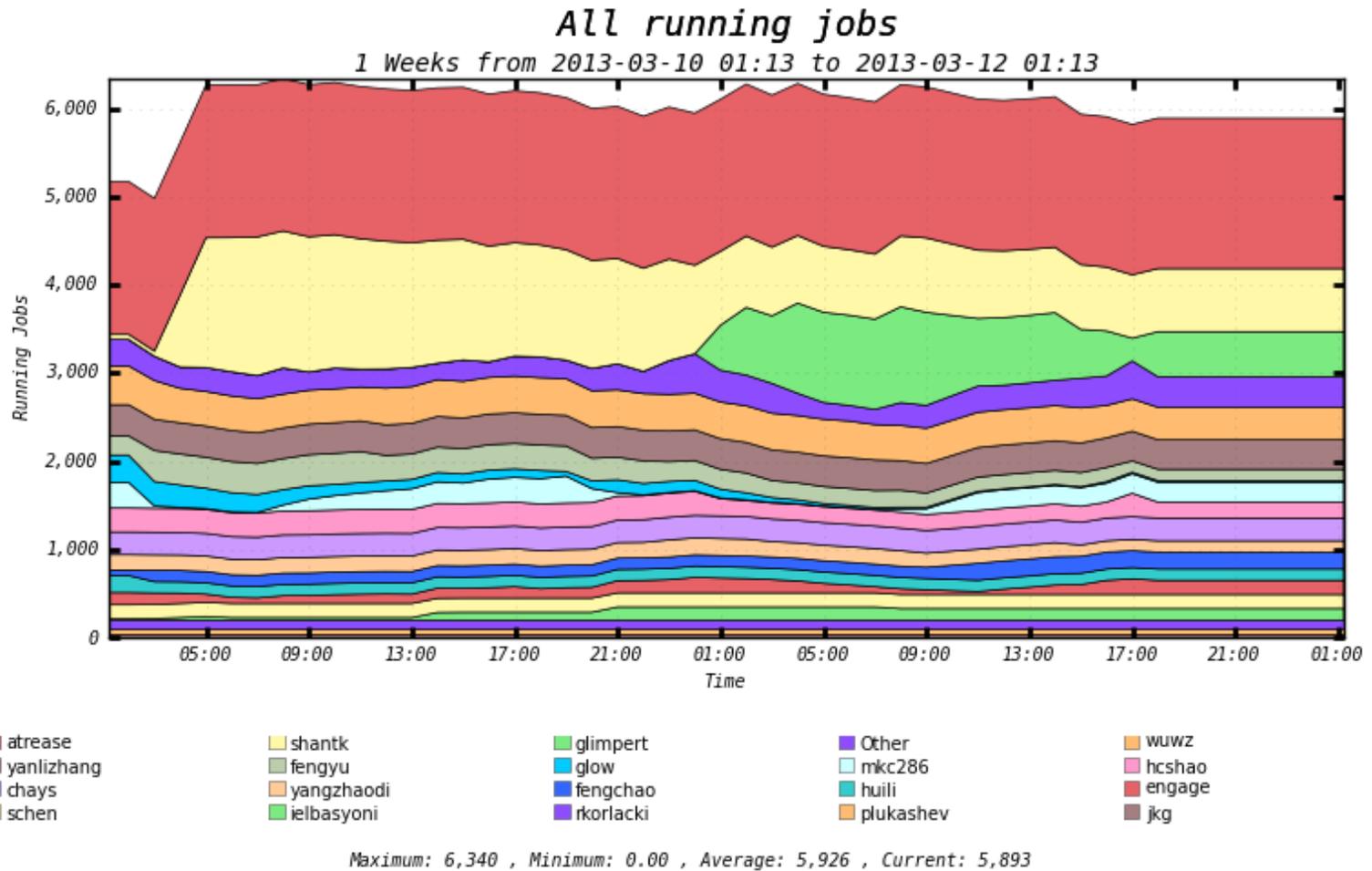


# HCC Model for a Campus Grid



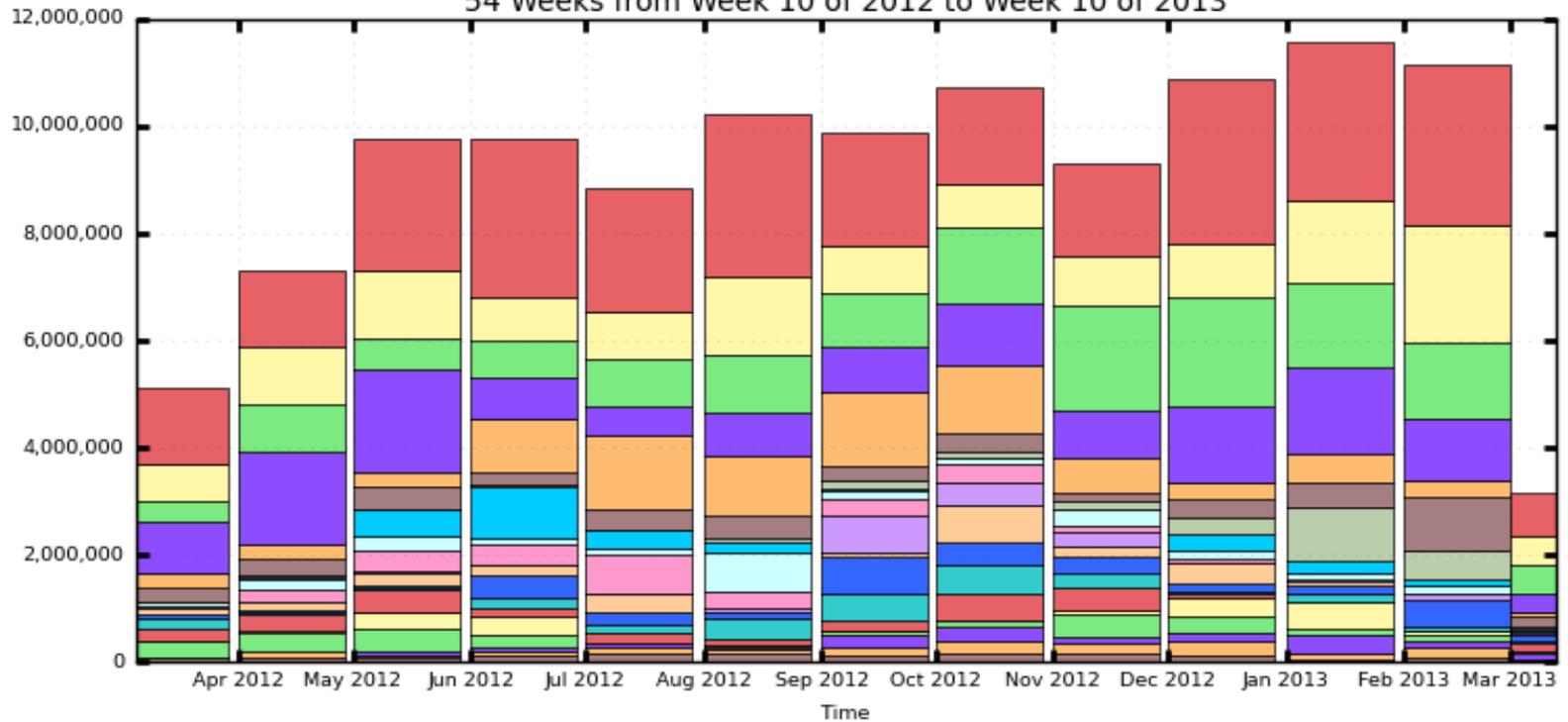
Me, my friends and everyone else

# It Works!



# Monthly Wall Hours per VO

54 Weeks from Week 10 of 2012 to Week 10 of 2013



Maximum: 11,592,467 , Minimum: 3,159,406 , Average: 9,060,807 , Current: 3,159,406



# High Performance Computing

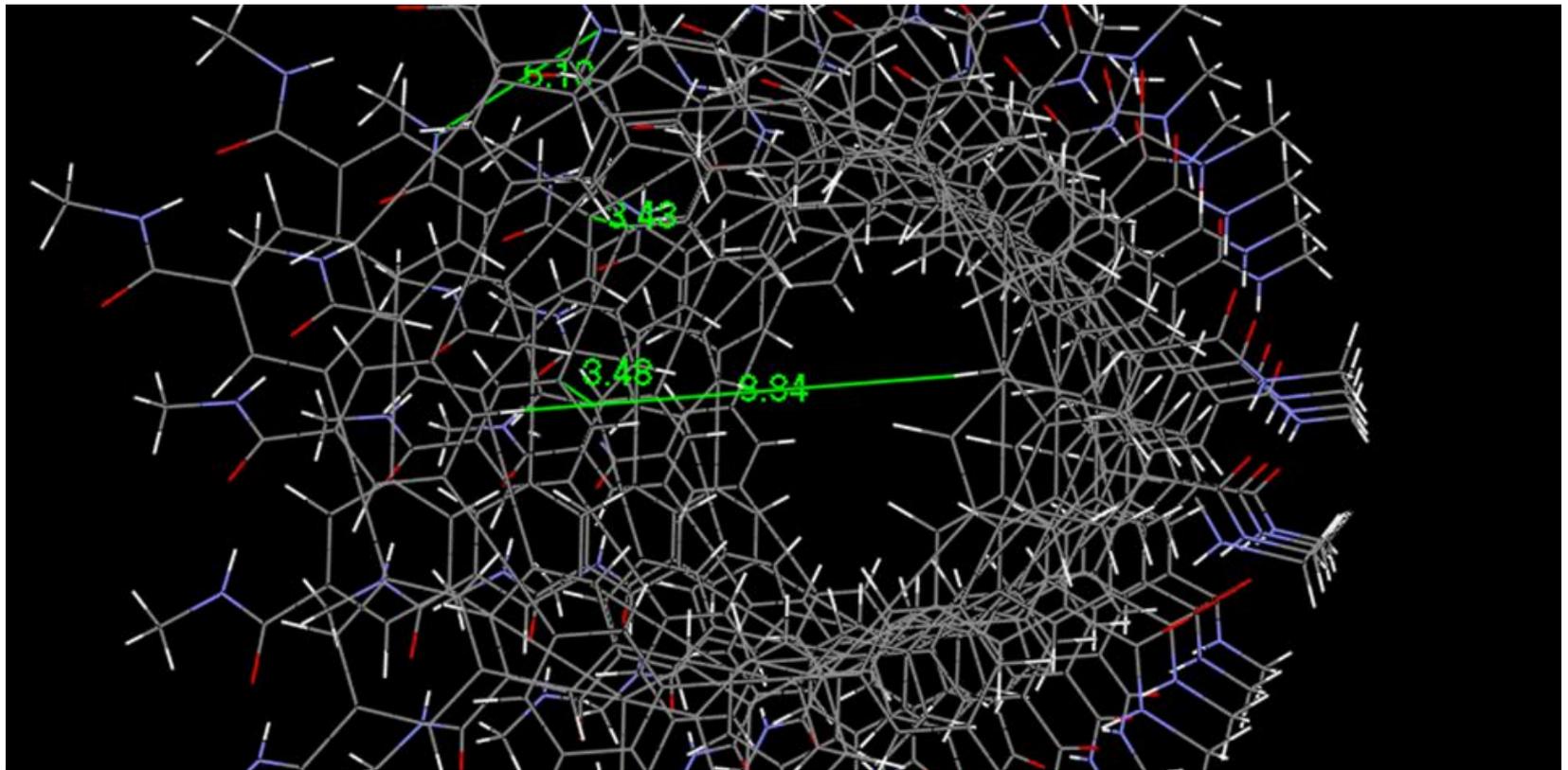
- Xiao Zeng, Chemistry, UNL
- DFT and Car Parrinello MD
- HPC – tightly coupled codes
- Requires expensive low-latency local network (infiniband)
- Requires high-performance storage (Panasas, Lustre)
- Requires highly reliable hardware



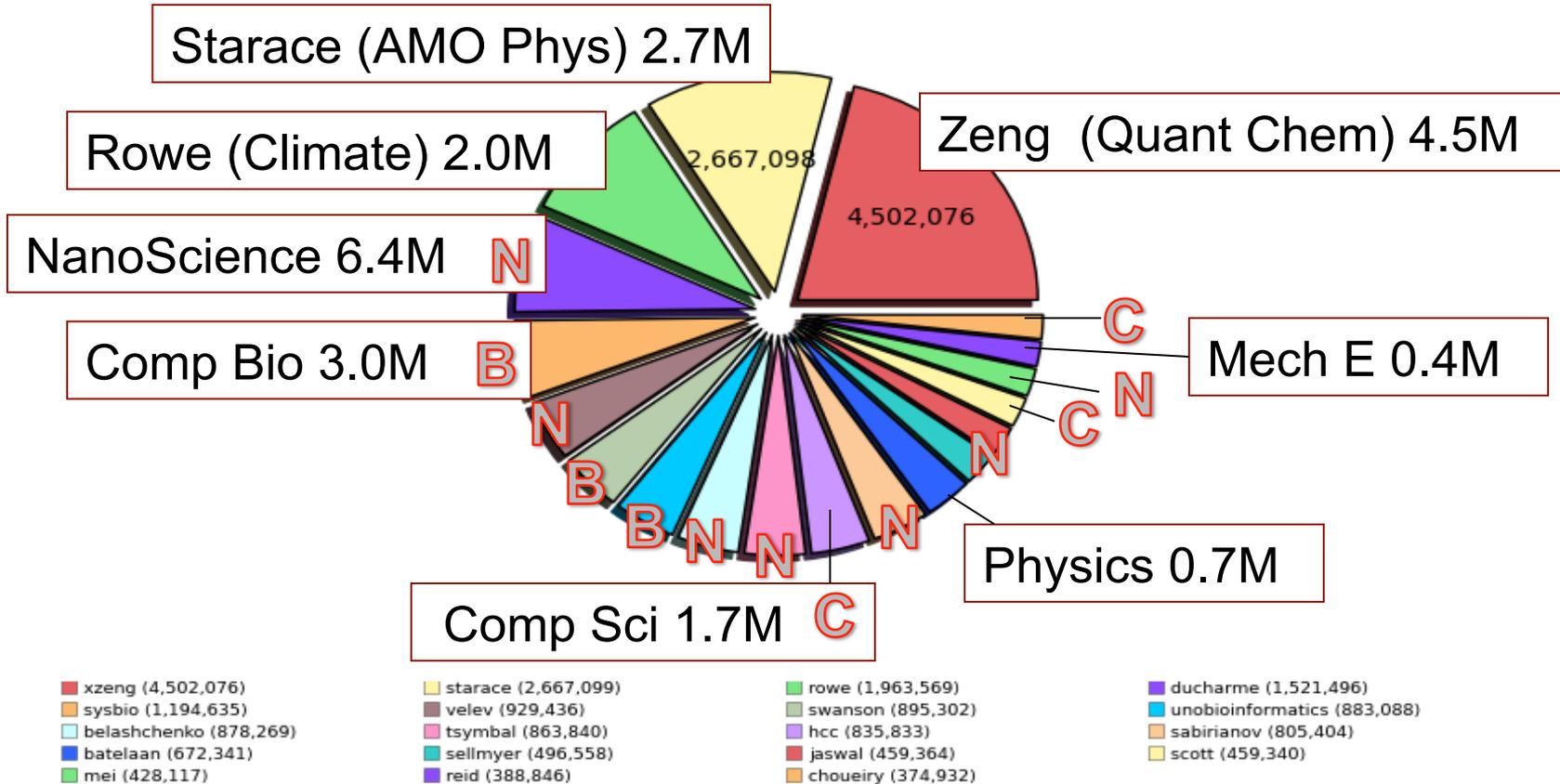
# Gold! PNAS Cover, May 30, 2006



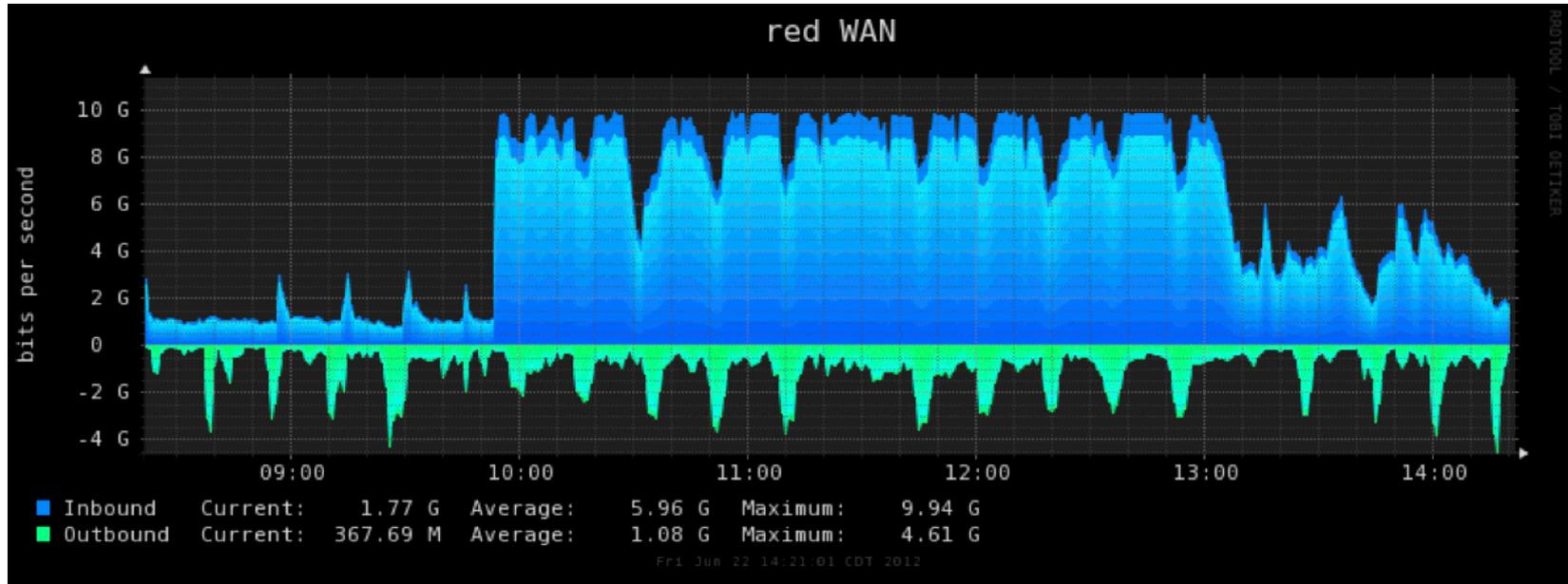
# Nature Communications, July 17, 2012



# HCC CPU Hour Usage 2012



# HCC Networking



- 2 locations connected at 30 gbps, 10 gbps to I2
- DYNES
- LARK project

# HCC Team

- CMS: David Swanson, Brian Bockelman, Ken Bloom, Carl Lundstedt, Aaron Dominguez, Garhan Attebury
- OSG/HCC: Ashu Guru, Adam Caprez, Jun Wang, Derek Weitzel
- Not pictured: *John Thiltges*, Josh Samuelson, Tom Harvill, Brad Hurst, Nick Palermo, Jenny Shao, *Wayne Huang*



# Eureka! A Higgs! (or at least something mostly indistinguishable)

- Tier2 brought HCC into OSG
- Over 1 PB currently at HCC's Tier2, 4400 cores
- UNL closely linked to KU, KSU physicists via a jointly hosted Tier3
- AAA with UCSD, Wisconsin, others
- HDFS



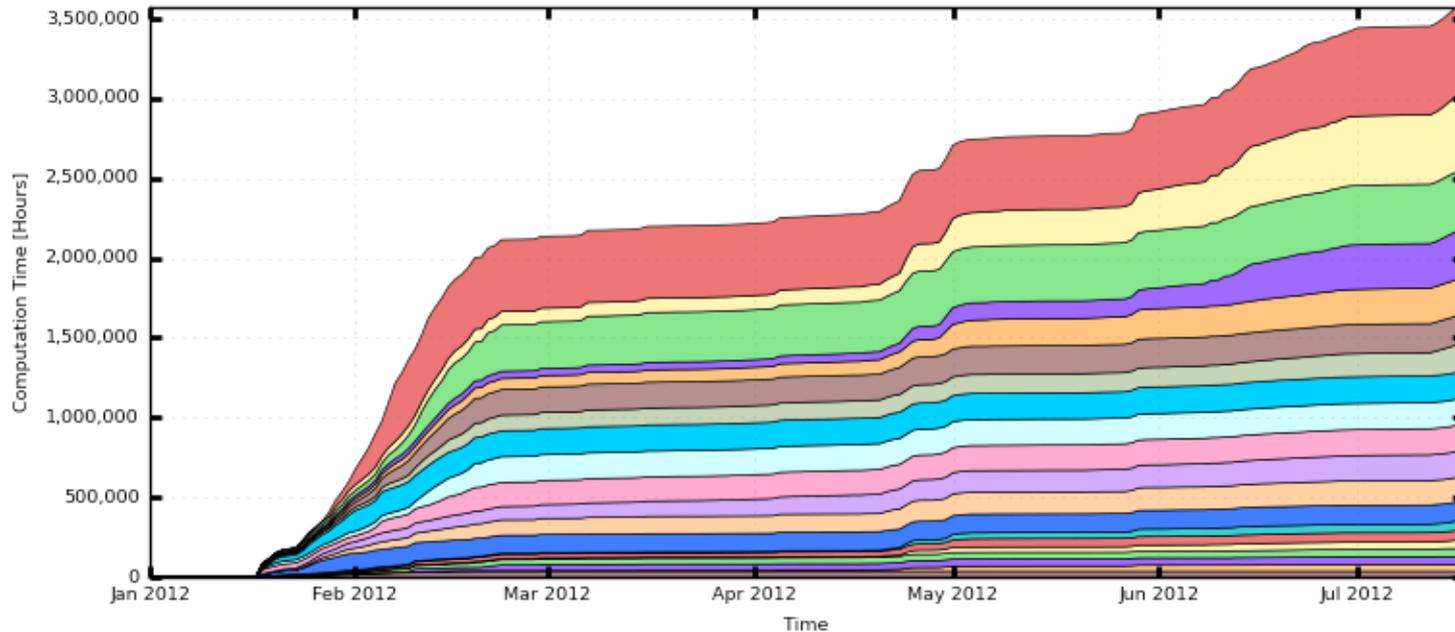
# HCC VO

- Campus VO
- Ca. 3M hours use of OSG last year by HCC
  - Very bursty
  - Local resources handle most of HT needs
- Resource provider



# OSG Resources

Cumulative Hours Spent on Jobs By Facility  
198 Days from Week 01 of 2012 to Week 29 of 2012



USCMS-FNAL-WC1-CE (554,781)	FNAL_FERMIGRID (476,557)	MWT2 (373,559)	FNAL_GPGRID_3 (301,869)
Other (227,328)	GridUNESP_CENTRAL (182,292)	AGLT2_CE_2 (171,933)	Purdue-RCAC (167,532)
CIT_CMS_T2 (164,039)	GLOW (163,948)	UCSDT2 (163,394)	NERSC-Carver (152,819)
Purdue-Steele (120,297)	FNAL_CDFOSG_1 (61,676)	UConn-OSG_CE (59,405)	FNAL_CDFOSG_2 (53,641)
SPRACE (49,962)	MIT_CMS (48,325)	BNL_ATLAS_2 (43,752)	OU_OSCER_ATLAS (35,943)

Total: 3,573,063 Hours, Average Rate: 0.21 Hours/s

# Working philosophy

- Use what we buy
  - Depreciation is immediate
  - Leasing is still more expensive (for now)
- Share what we aren't using
  - Share opportunistically – retain local ownership
  - Consume opportunistically – there is more to gain!
  - Collaborators, not just consumers
  - Greater good vs. squandered opportunity



# Engaging Users

- Meeting with HCC staff near a whiteboard
- Decision:
  - (1) Condor DAG
  - (2) other -- MPI, OpenMP, huge shared filesystem, Windows code, ...
- if (1), deploy on campus grid (!)
  - Scale out by flocking
  - Scale out to OSG
- if (2), Maui/Torque or SLURM

# Recent Examples

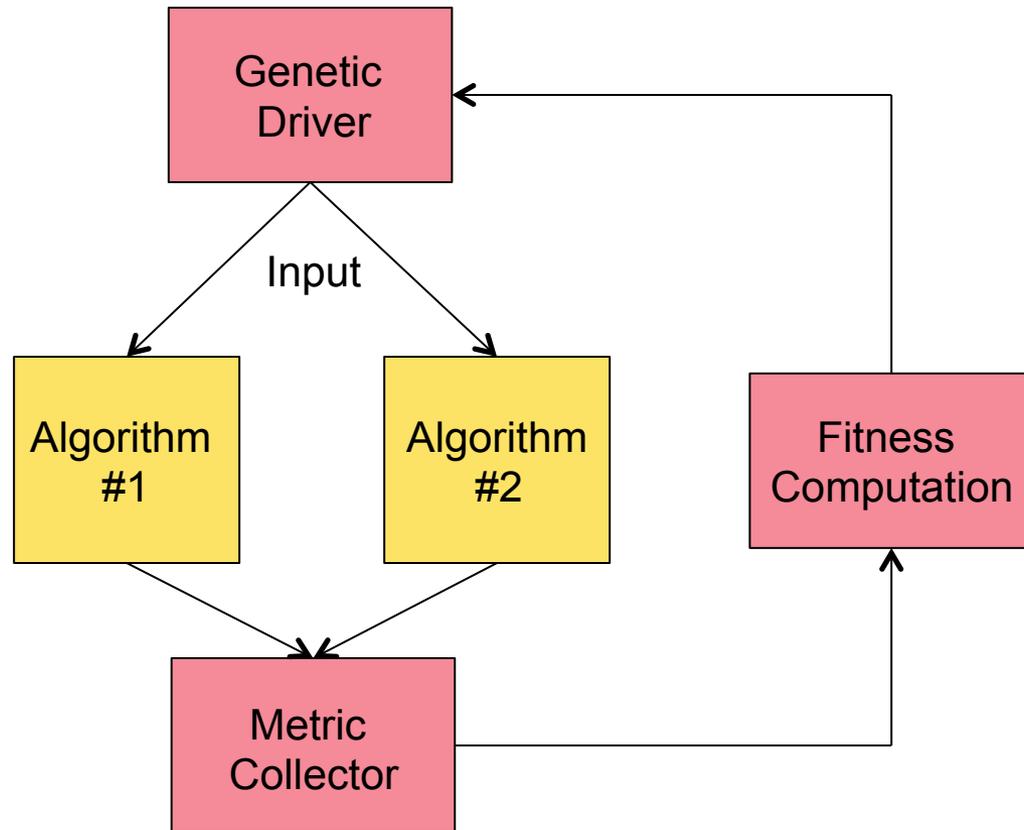
- Genetic algorithm (Parrot/Chirp)
- Render farm (Portal + Bosco)
- ISPTM (Portal + Pegasus)
- Quantum Espresso (Partitionable slots)
  
- Derek Weitzel, Adam Caprez, Ashu Guru and Jun Wang



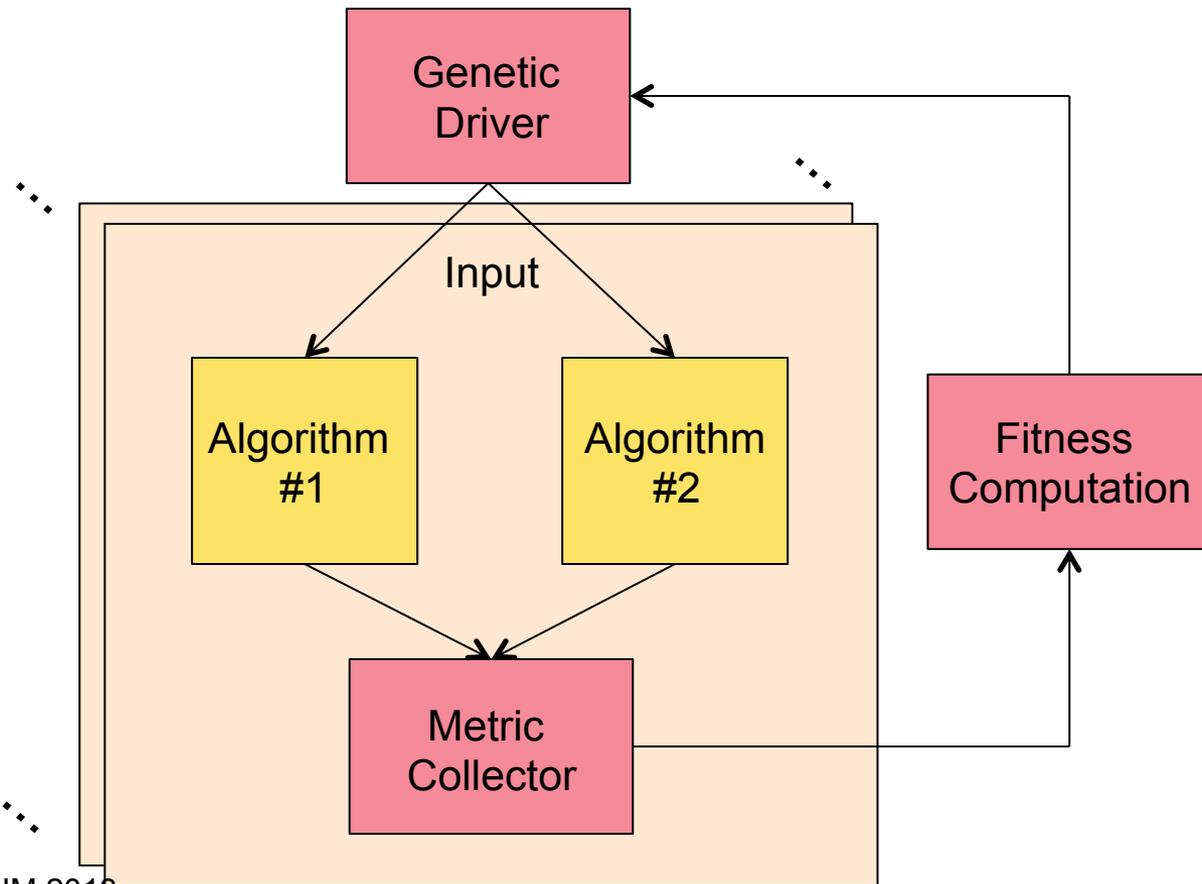
# Differential Genetic Algorithm-Based Software Testing

- Genetic Algorithm
  - Problem is encoded as a population of chromosomes, each chromosome representing a candidate solution
  - Compute the fitness of an entire generation at once, as opposed to single chromosomes in a generation

# Differential Genetic Algorithm-Based Software Testing



# Differential Genetic Algorithm-Based Software Testing



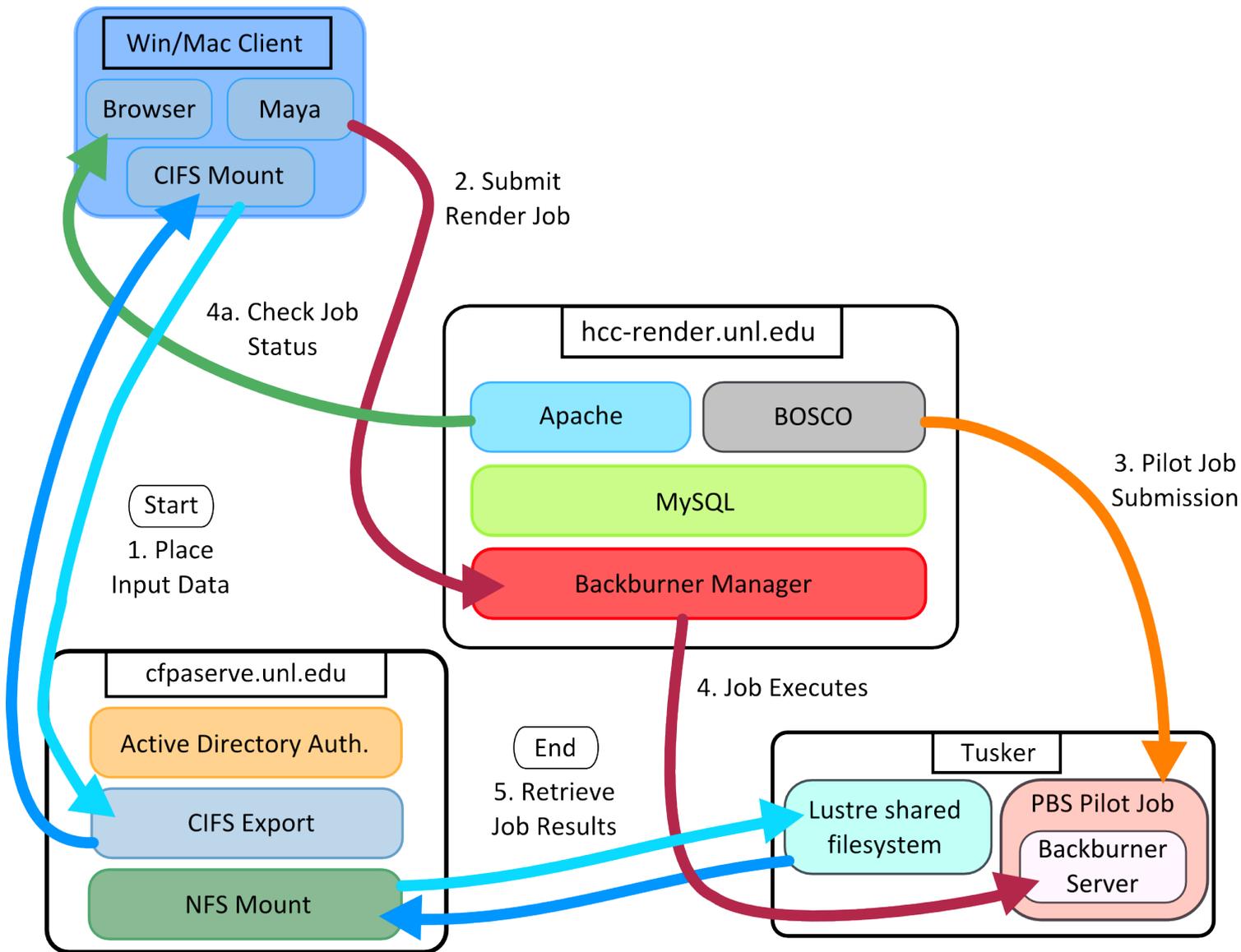
# Differential Genetic Algorithm-Based Software Testing

- Require dynamic and constant communication with Sandhills
- Parrot/Chirp
  - Nothing for user to install, no special permissions required
- Run a “chirp server” on Sandhills
- Interface with Sandhills by writing and reading files to and from a virtual chirp directory

# Render Farm

- Faculty, students use Maya for Windows or Mac
- Comes with workflow manager known as Backburner
- The above lends itself to running only as one user “Maya” on cluster
- Use NFS/CIFS to map local file system to cluster
- Finally, use BOSCO to submit glideins for backburner jobs

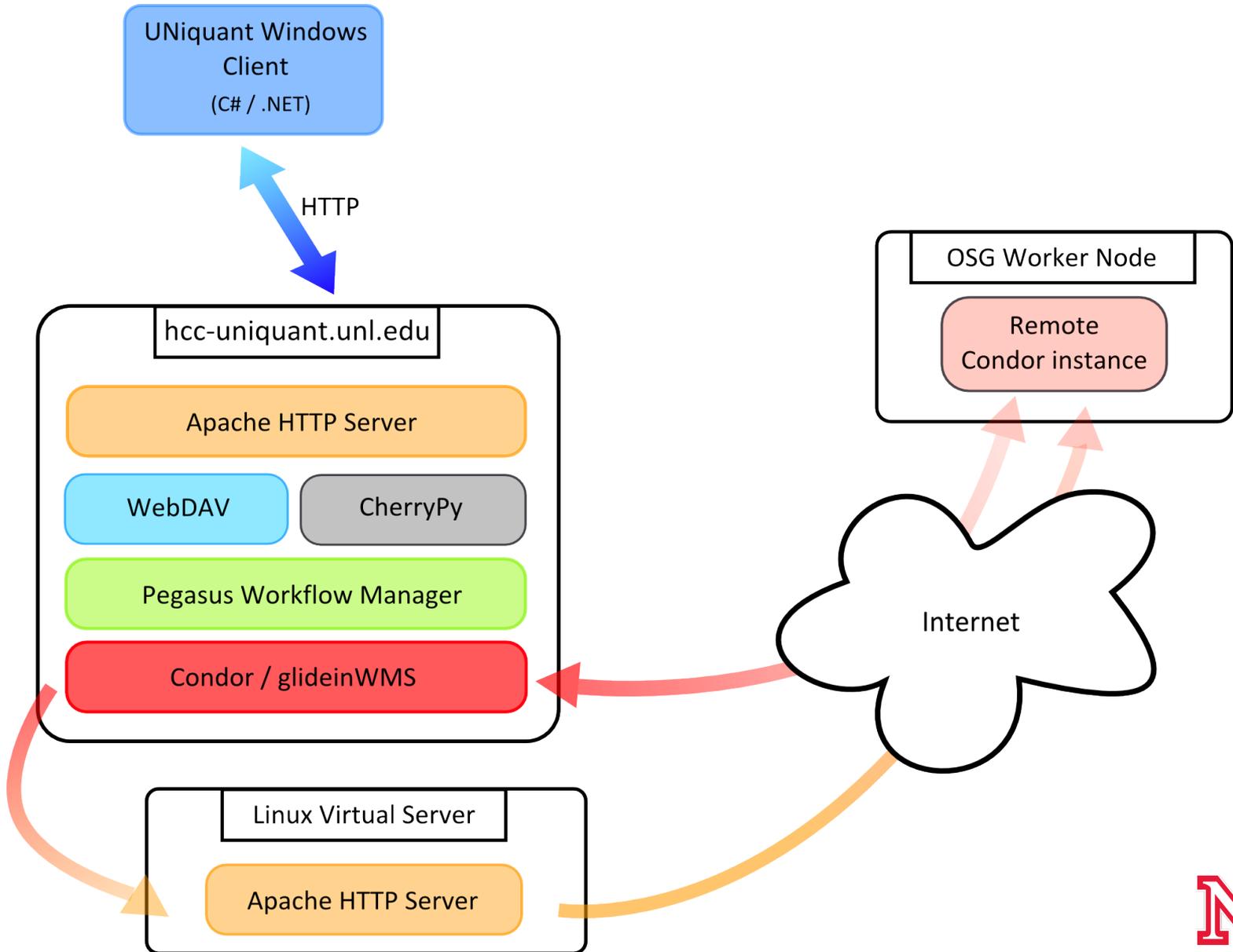




# ISPTM – Iterative Search for Post Translational Modifications

- Analyzes Mass Spectra data to find peptides that are altered after translation
- Wanted app for job creation/submission
- Needed to scale approach – iterative, time-consuming





# Quantum Espresso

- Quantum Chemistry package
- Too big of memory footprint, too long run time to run in serial
- Derek Stewart (Cornell) wanted to run via OSG over multiple clusters
- Configured Partitionable slots on 8-cores
- <http://derekweitzel.blogspot.com>



# Summary

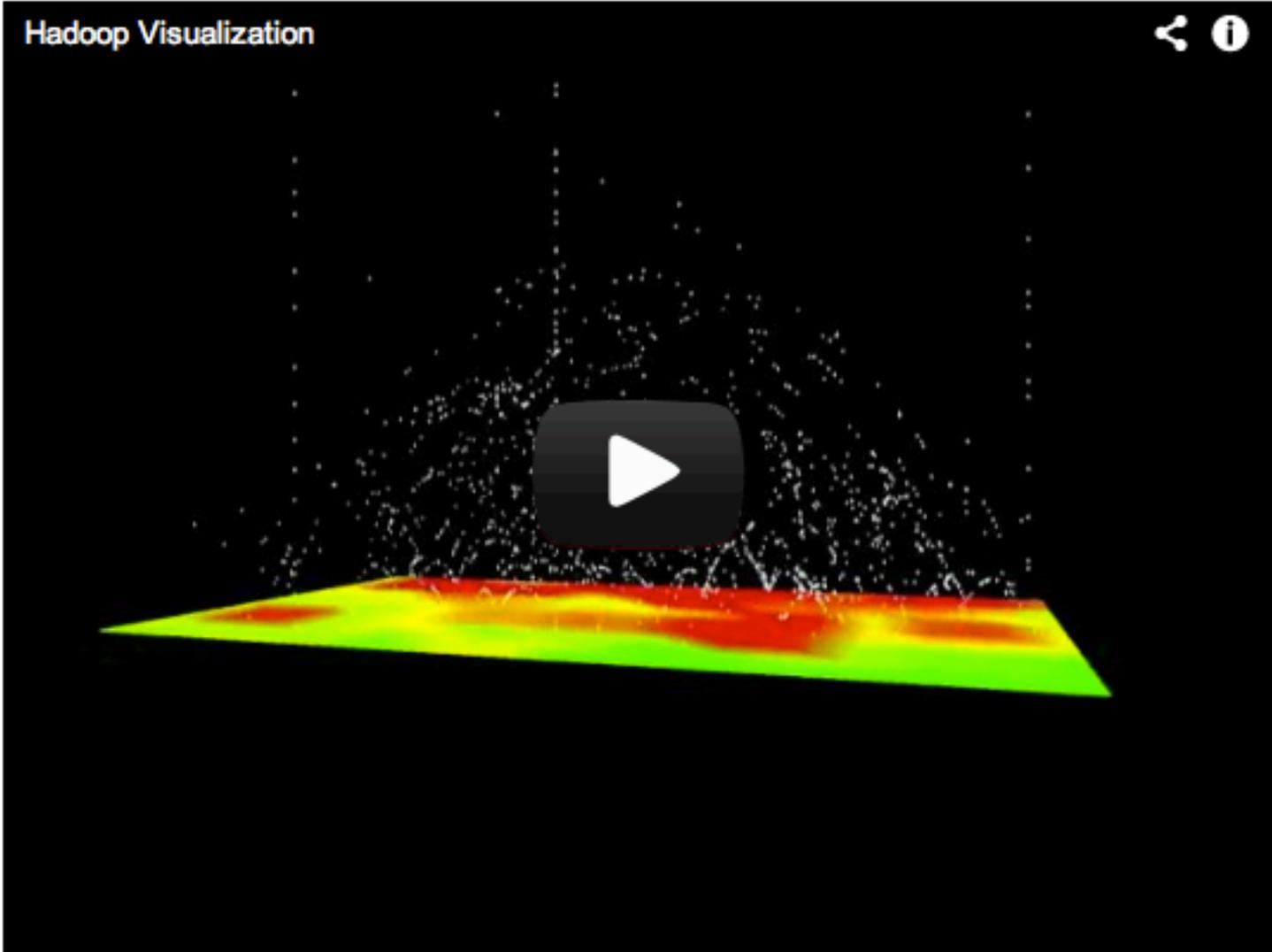
- HPC and HTC coexist peacefully at HCC
  - HPC dominates local submissions
  - HTC/OSG provides burst capability
- Several examples of recent usage of Campus Grid
  - Effort required
  - Portal vs. direct submit
  - Linear combination of familiar approaches
  - Whiteboard not optional



# **A FEW THINGS YOU MAY HAVE SEEN...**

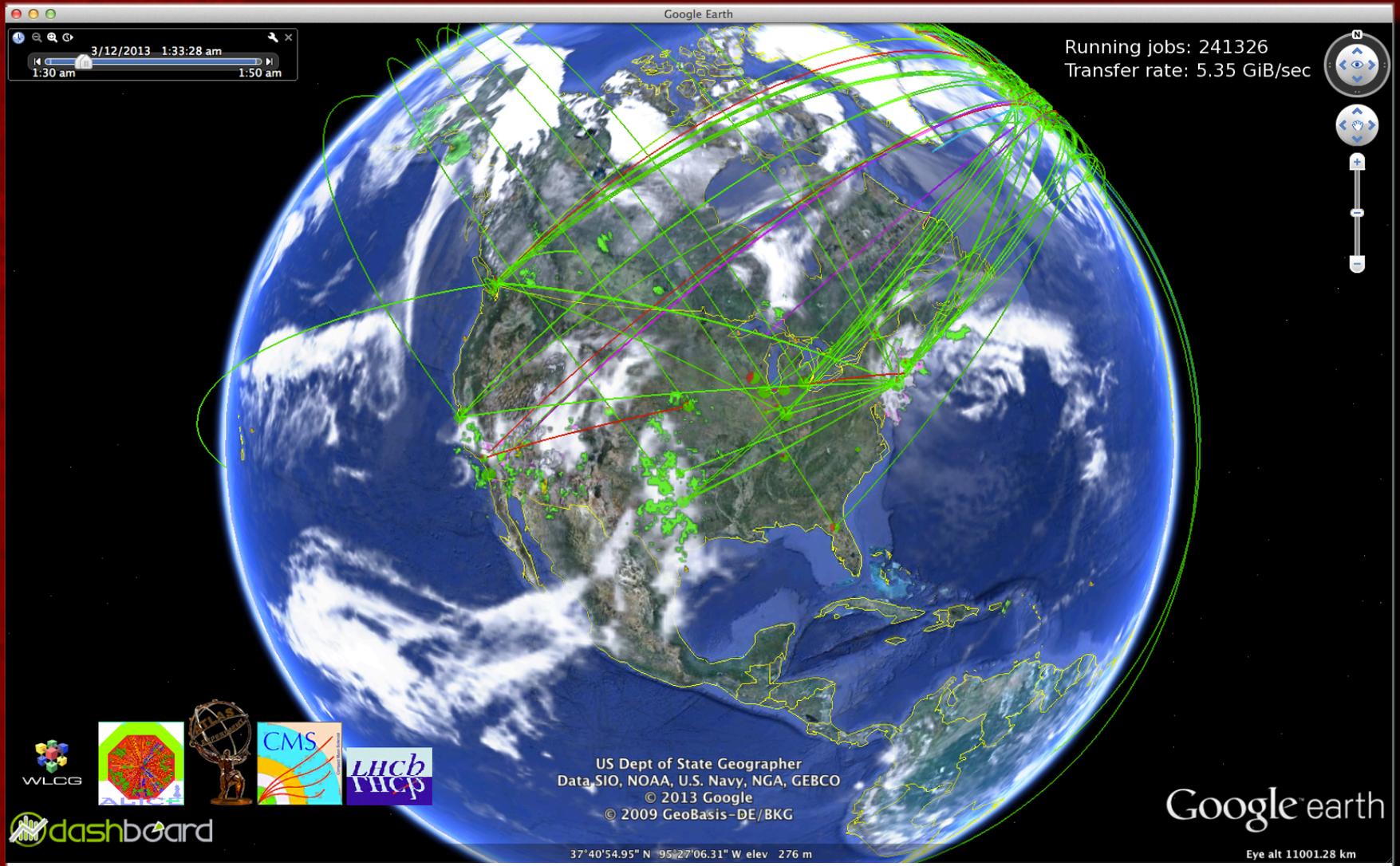


Hadoop Visualization

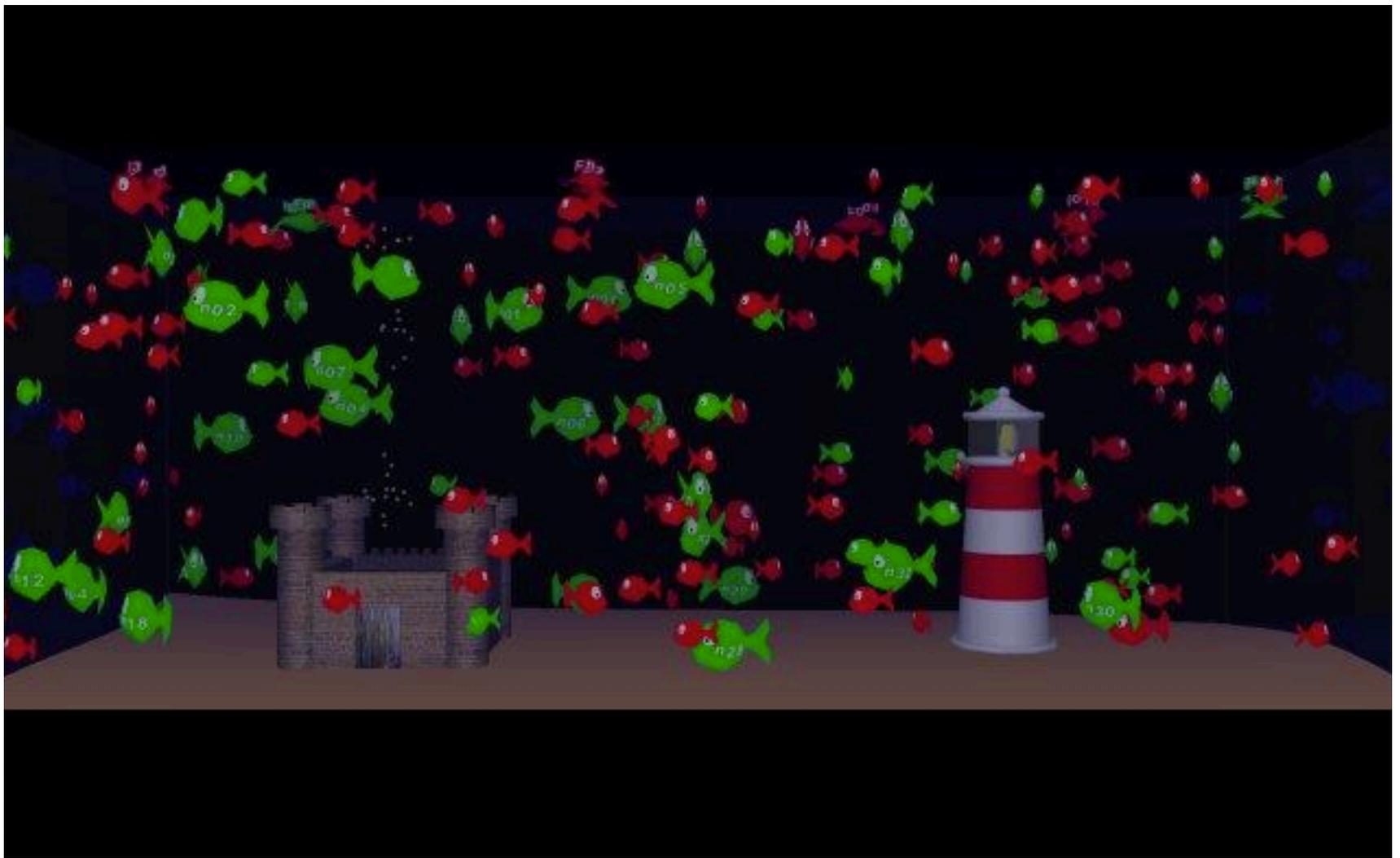


<https://www.infoworld.com/d/big-data/facebook-pushes-the-limits-of-hadoop-206108>





<https://www.opensciencegrid.org/bin/view/Main/InstallingOSGGoogleEarth> 



<http://sourceforge.net/projects/fish-monitoring/>





NU Administration (UNL, NRI)  
NSF, DOE, EPSCoR, OSG  
Holland Foundation

CMS: Ken Bloom, Aaron Dominguez

HCC: Drs. Brian Bockelman, Adam Caprez, Ashu Guru,  
Carl Lundstedt, Nick Palmero, Jun Wang.

Garhan Attebury, Tom Harville, William Hurst,  
Josh Samuelson, John Thiltges.

Chen He, Derek Weitzel, Zhe Zhang  
Xiao Zeng

©2007 The Board of Regents of the University of Nebraska

UNIVERSITY OF  
**Nebraska**  
Lincoln