

# Meeting the HTC Demand with Diagrid and XSEDE

Kimberley Dillman

Research Programmer / Purdue XSEDE  
Campus Champion

Purdue University

email: [kadillma@purdue.edu](mailto:kadillma@purdue.edu)



**PURDUE**  
UNIVERSITY

What “HTC” resources do we have at  
Purdue?

# HTC “Traditional” Methods

- Community Cluster “standby” queues
- Diagrid
- Open Science Grid (OSG)

# Community Cluster “standby” Queues

- Must “purchase” at least one node to gain access to any “community cluster” with the exception of Radon (recycled cluster)
- Allows access to any/all the CPUs in the cluster for up to 4 hours when idle (not in use by “owner”)
- Jobs in this queue have lower priority than jobs in “owner” queues
- Can run serial jobs and “share” nodes with other jobs or “claim” the entire node

# DiaGrid

- A large, high-throughput, distributed computing system
- Operated by Rosen Center (RCAC)
- Uses Condor to manage jobs and resources
- Good for running serial computations on a large number of processors
- Utilizing idle cycles
- Including all Purdue clusters, lab computers, department computers, desktop, totaling 50,000+ cores
- Purdue leading a partnership of 10 campuses and institutions

# DiaGrid Partners



Instant access to high-throughput computing

[Resources](#)

[Members](#)

[Explore](#)

[About](#)

[Support](#)

## Partners



**PURDUE**  
UNIVERSITY



UNIVERSITY OF  
**Nebraska**  
Lincoln  
*Pioneering new frontiers.*

University of  
Louisville



University of  
Notre Dame

Indiana State  
University

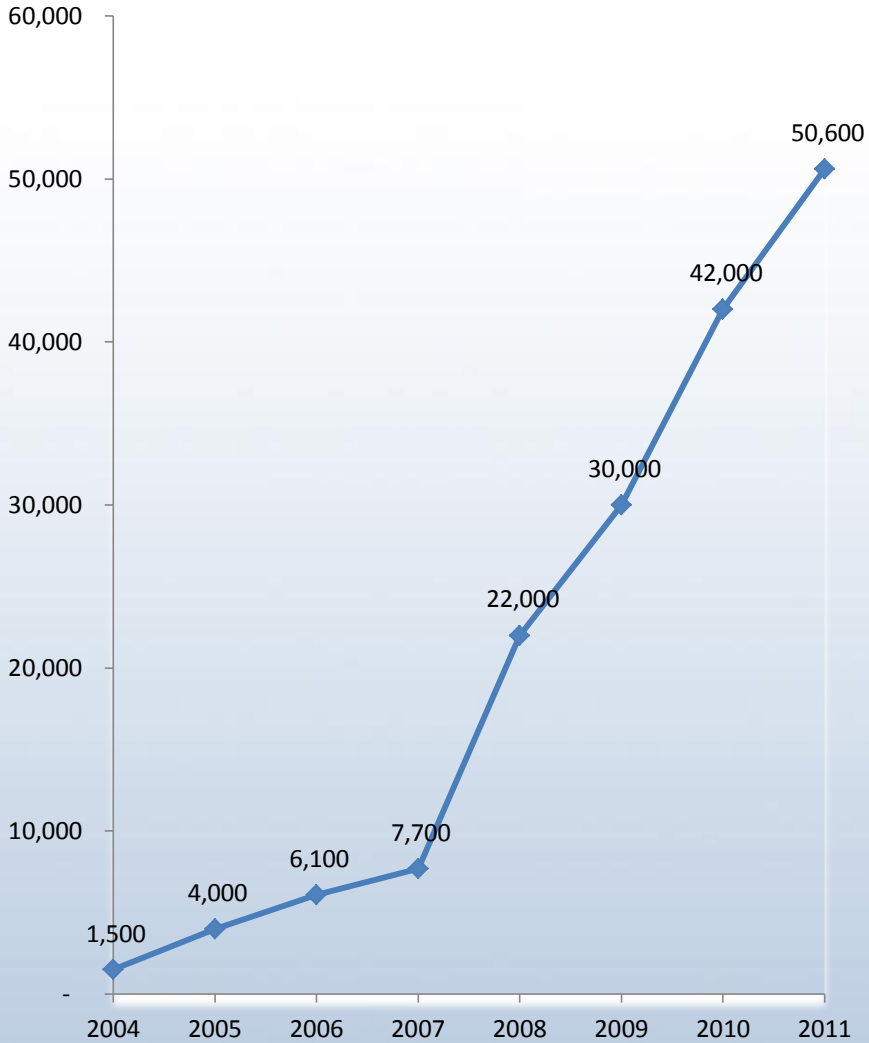


Purdue University  
North Central

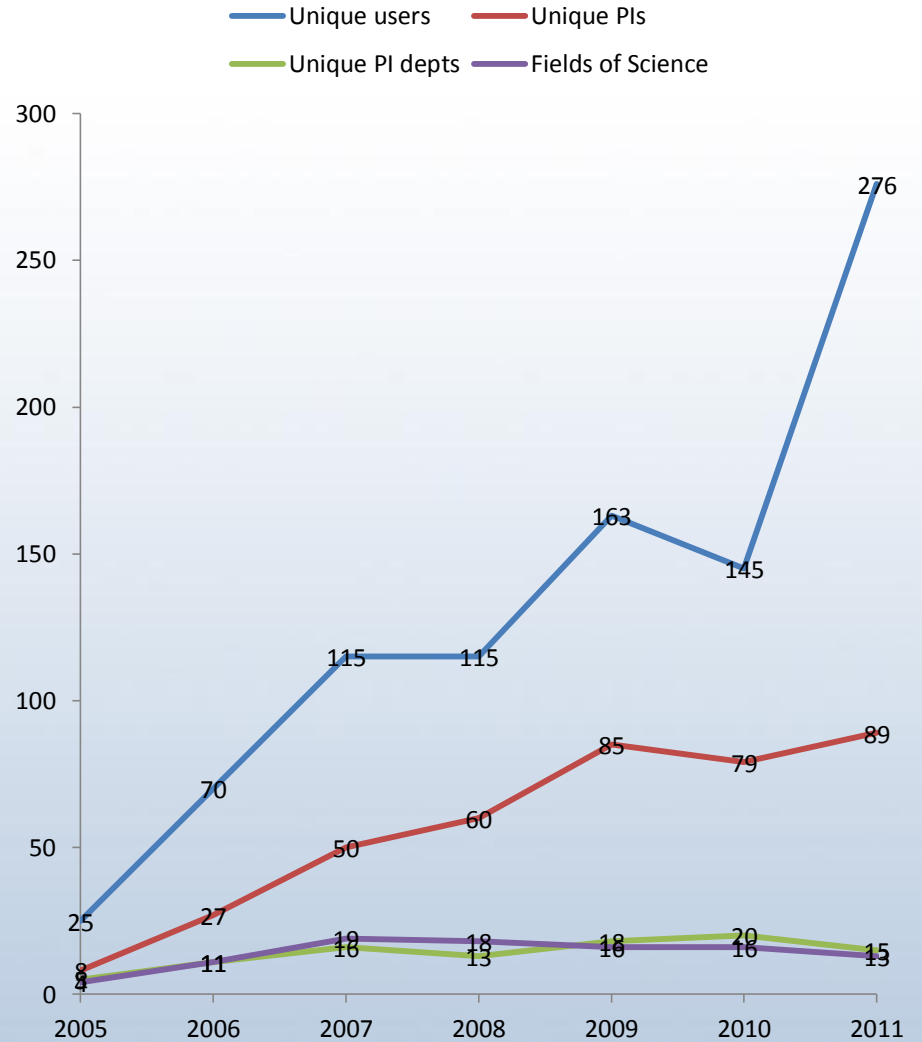
**PURDUE**  
CALUMET

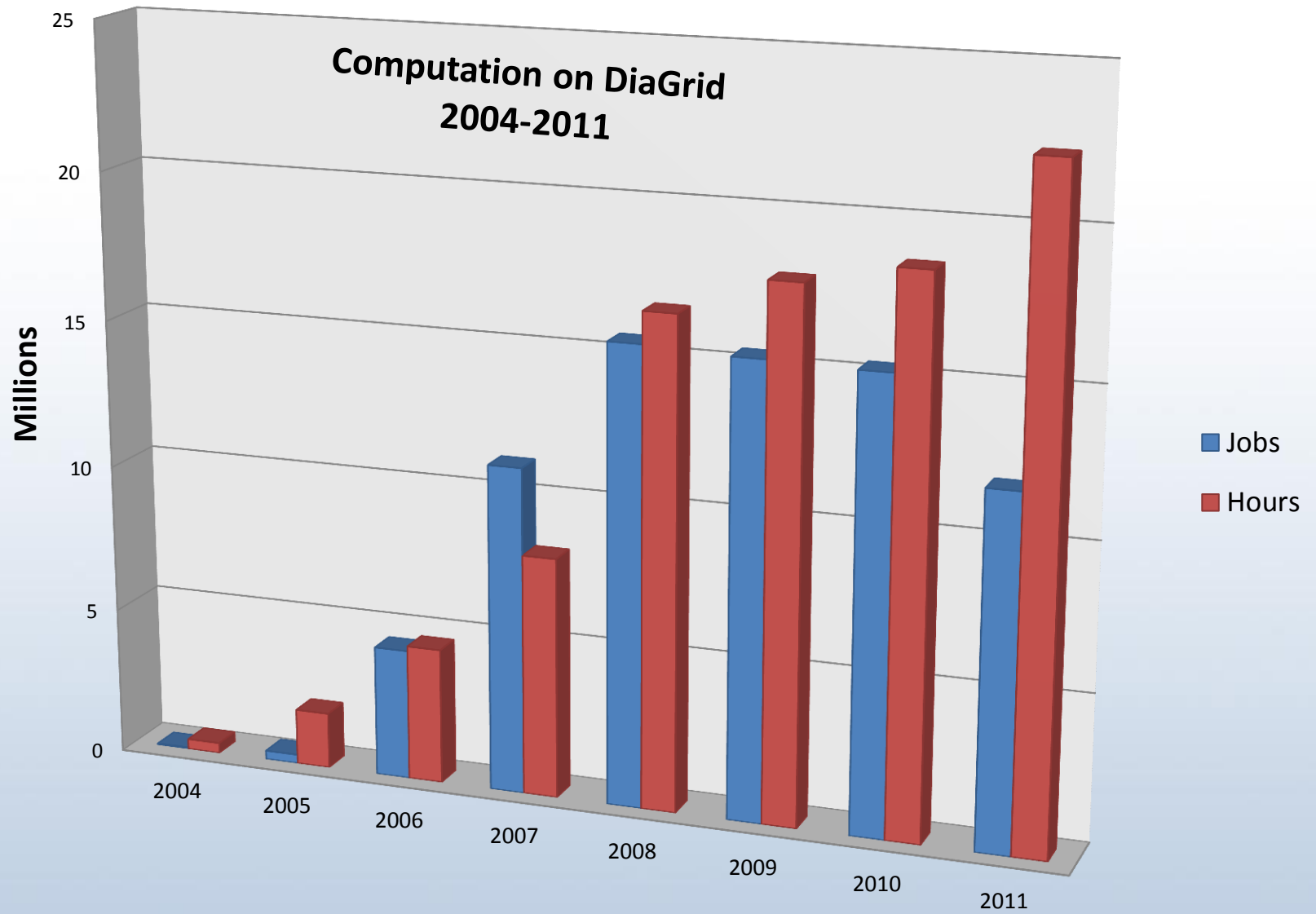
Become a  
Partner Today  
»

## Growth of DiaGrid 2004-2011 (number of cores)



## DiaGrid User Count







# Open Science Grid (OSG)

- Some Current Users at Purdue
  - CMS Tier 2
  - NEES (via NEESHub)
- Methods of access
  - “traditional” VO submit host
  - “submit” capability enabled within a HUBZero hub (i.e. NEES)

# Job Priority Order

- Node “owner” (highest priority)
- Cluster “standby” queue (second priority)
- Diagrid/OSG (lowest priority)

# Why isn't the "traditional" method good enough?

- Not all jobs are "serial" or are "short enough" in duration to fit well into a "cycle scavenging" mode of operation
- Not all scientists are or want to be "computer experts"
  - They don't want to have to know or understand the different methods of job submission and syntax (i.e. no command line)
  - They just want to get their science done!
- Users usually don't care where or how the job runs...just that it does so successfully...

# Enter....Diagrid Hub

- What is it?
  - Access to the Diagrid Pool of resources plus a dedicated queue on the Hansen cluster
  - GUI “front end” that “hides” the details of job submission from the user
- What applications are available?
  - Blast
  - R
  - CryoEM image processing
- Who uses it?

# Science-as-a-Service



No Forms. No waiting. Just instant access to high-throughput computing

89 million Condor jobs run to date  
492 hub users in past 12 months  
117 simulation users

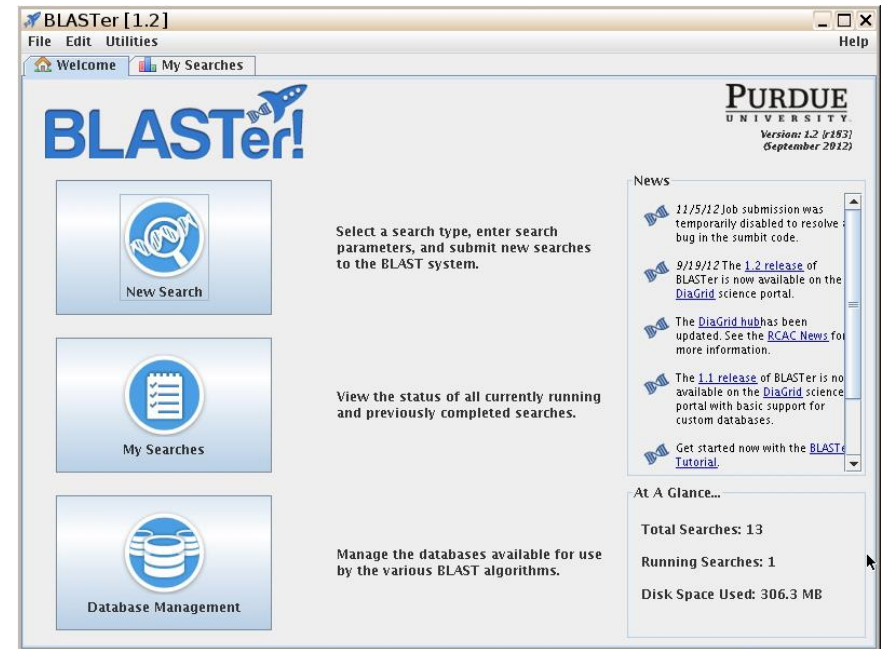


# Bioinformatics – BLASTer for large scale sequence alignment



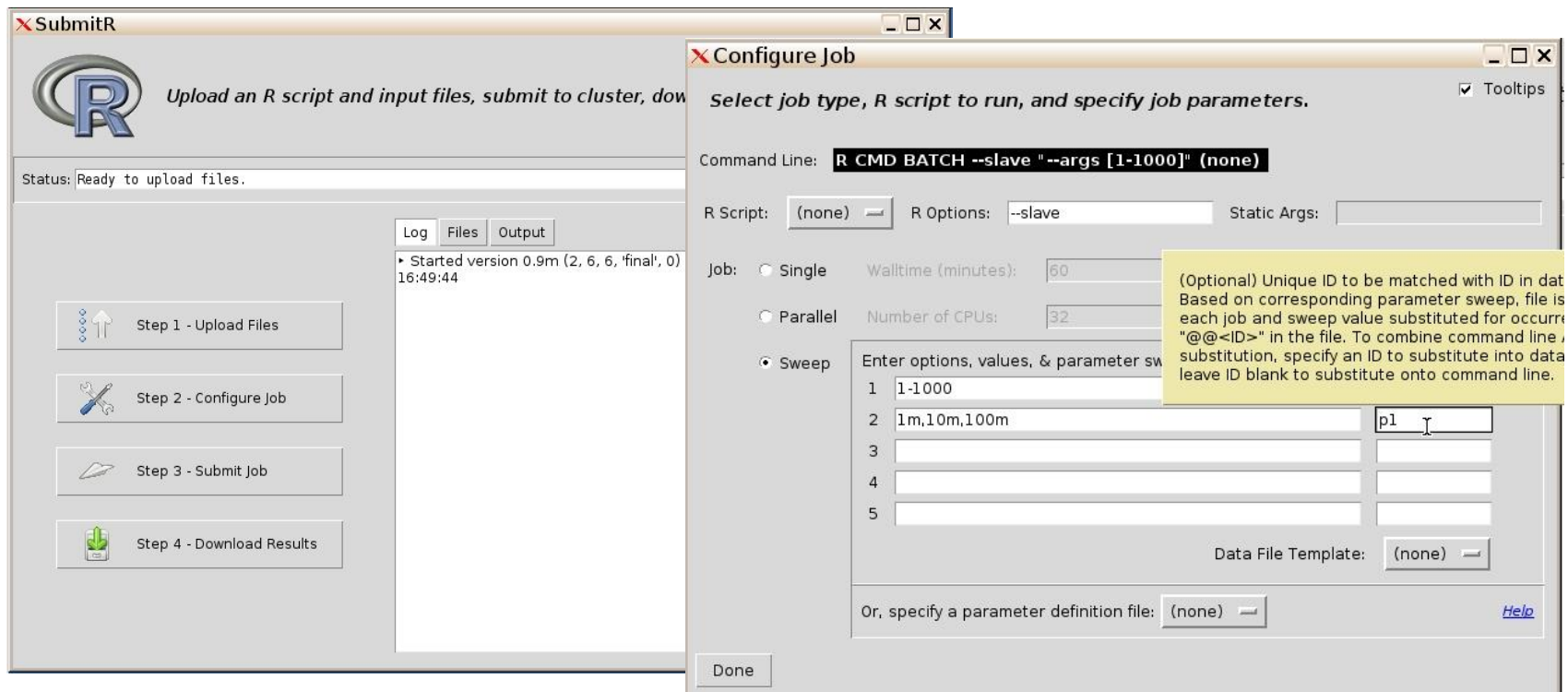
J. Andrew DeWoody, Nick Marra,  
Forestry & Natural Resources

- Using Blaster to annotate assembly of gene sequences (50,534 contigs) from E51K Illumina in study of gene evolution
- 8 days in the lab → less than 3 hours on DiaGrid



- Blaster has completed 1 million search jobs (equivalent to searches of tens of millions of sequences against public and custom databases)
- Currently 44 research users of Blaster
- April, Sept campus wide presentations
- Nov. presentation to Coll. Pharmacy

# R for everyone – Data analysis, parameter sweeps, parallel applications



- Beta release of SubmitR in October 2012
- A single interface allowing users to access Condor and cluster resources
- 1.1 M processor hours has been used by R applications from DiaGrid
- Community forum on November 28 (21 researchers attended)



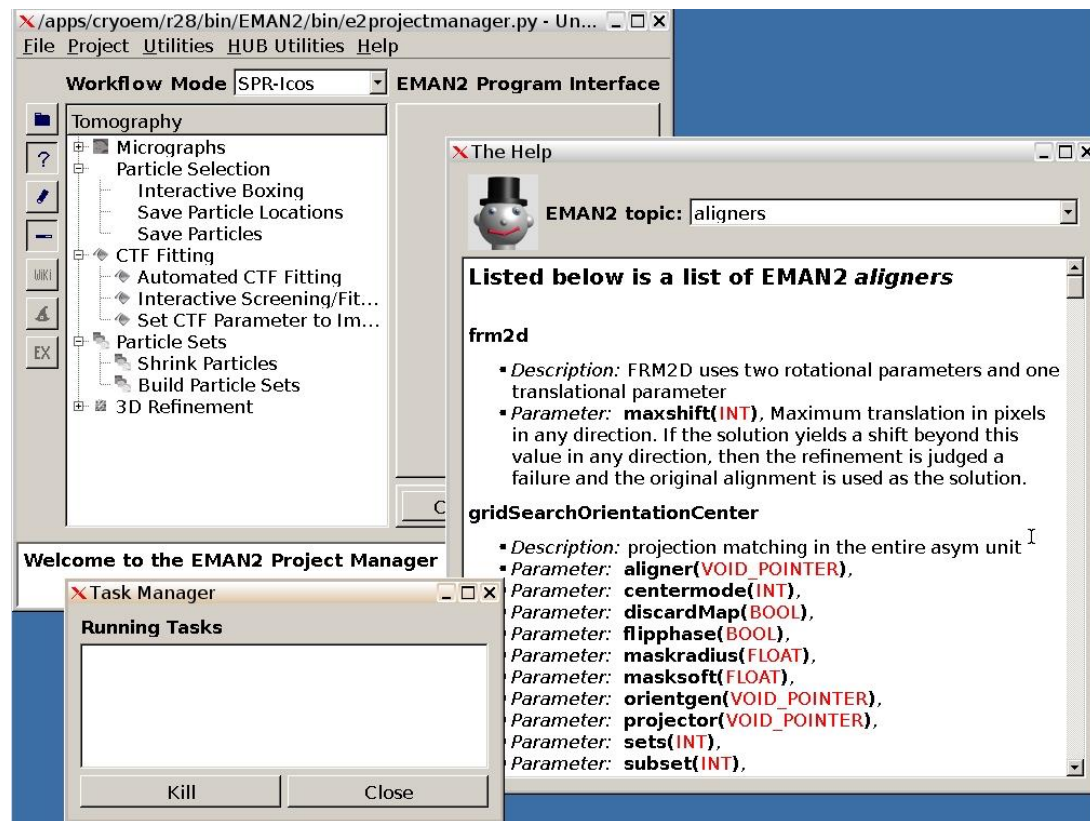
# DiaGrid is also for sharing scientific tools – CryoEM 3D Reconstruction

- Powerful tool created by research group in Biology
- Adapted to DiaGrid hub to share with the larger research community
- Still in development, prototype tested with small class in 2012



Wen Jiang, Biology

Slide Courtesy of Carol Song - Purdue





# Diagrid Hub – Latest Stats..

- SubmitR (via diagrid-a queue on Hansen cluster)
  - Total Jobs: 1013
  - Total Wall Processor Hours: 2,574,001
- Blaster
  - Total Jobs: 258,961
  - Total Wall Processor Hours: 1,092,839
- CryoEM
  - Used in “beta mode” in small class settings
  - Being integrated with Pegasus (workflow)

# Diagrid Hub – What's next..

- Enhancements to existing tools
- Add new tools
- Pegasus is now available in HUBZero...use it to manage large job workflows...

Where do you go when you need more?

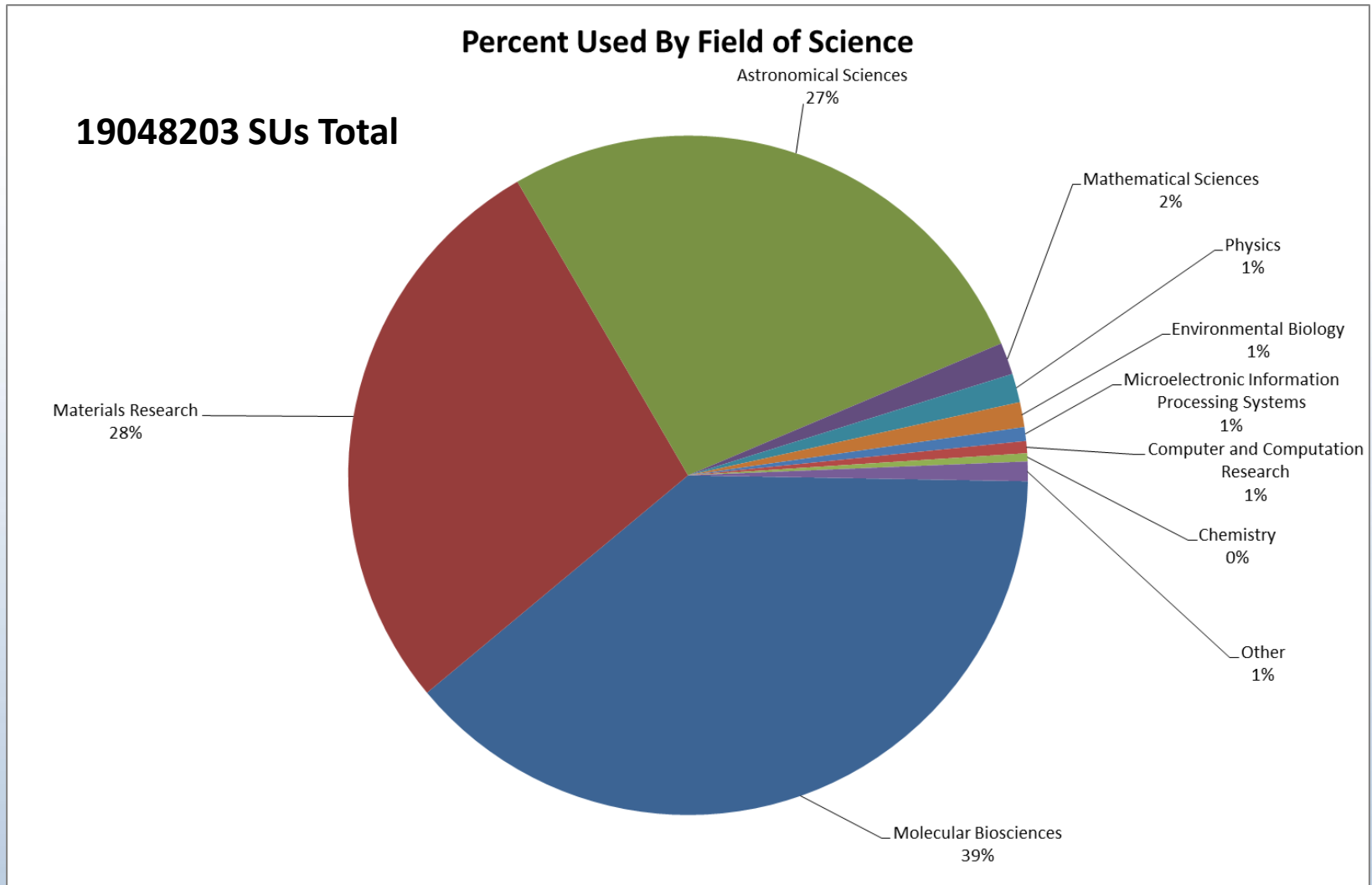
# XSEDE (and what is available)

- Large supercomputers
- Medium clusters
- Viz resources
- Diagrid
- OSG

# XSEDE Resources for HTC

- Diagrid (aka Purdue Condor Pool)
  - No longer available via XSEDE allocations after July 2013
- Open Science Grid
  - XSEDE has a submit host for login and submit access to OSG
- Some TACC resources
  - Stampede and Lonestar have special “submit scripts” that can aid users with “bundling” and submitting serial jobs so that they “look like” large parallel jobs to the system

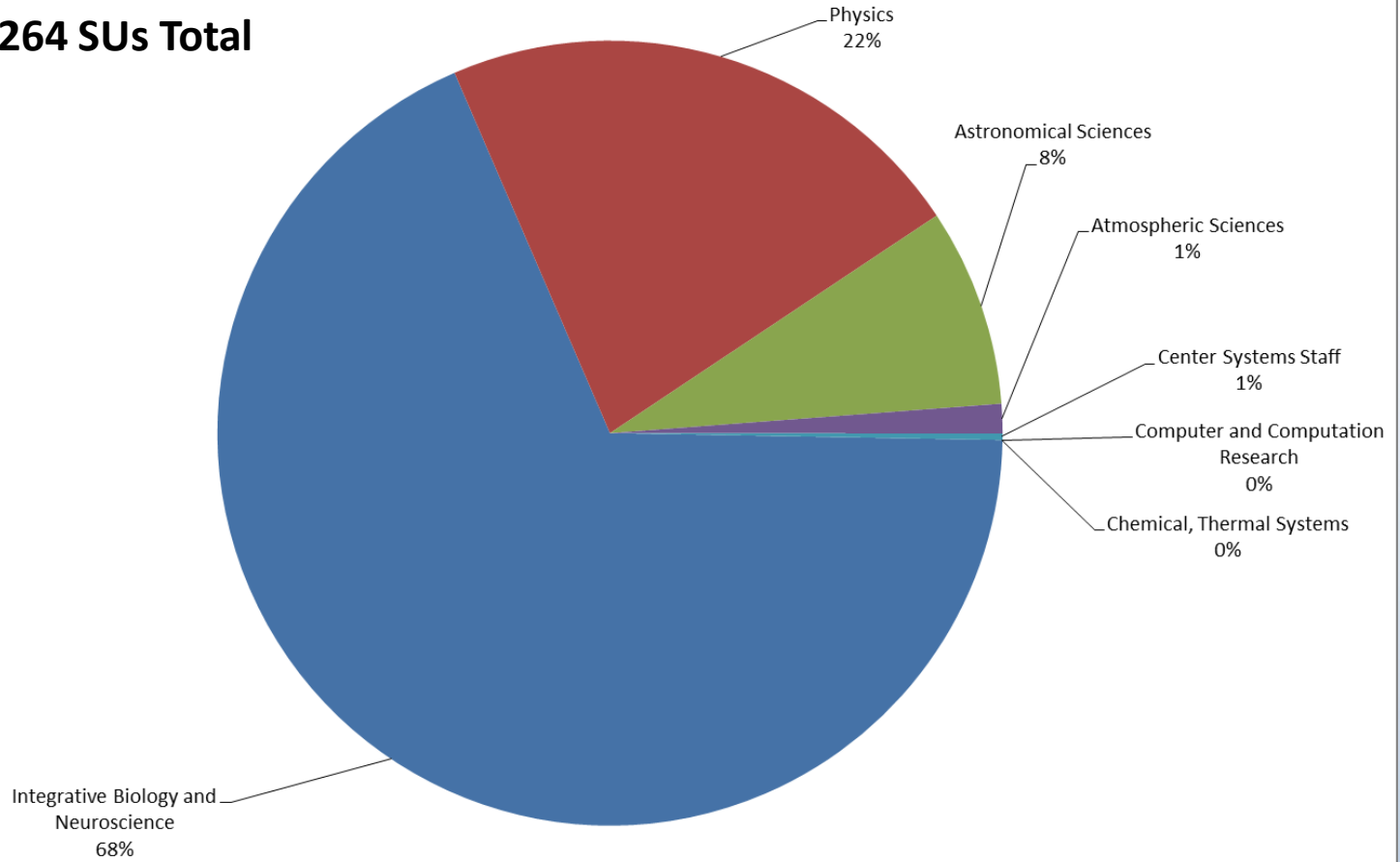
# Purdue Condor Pool Usage



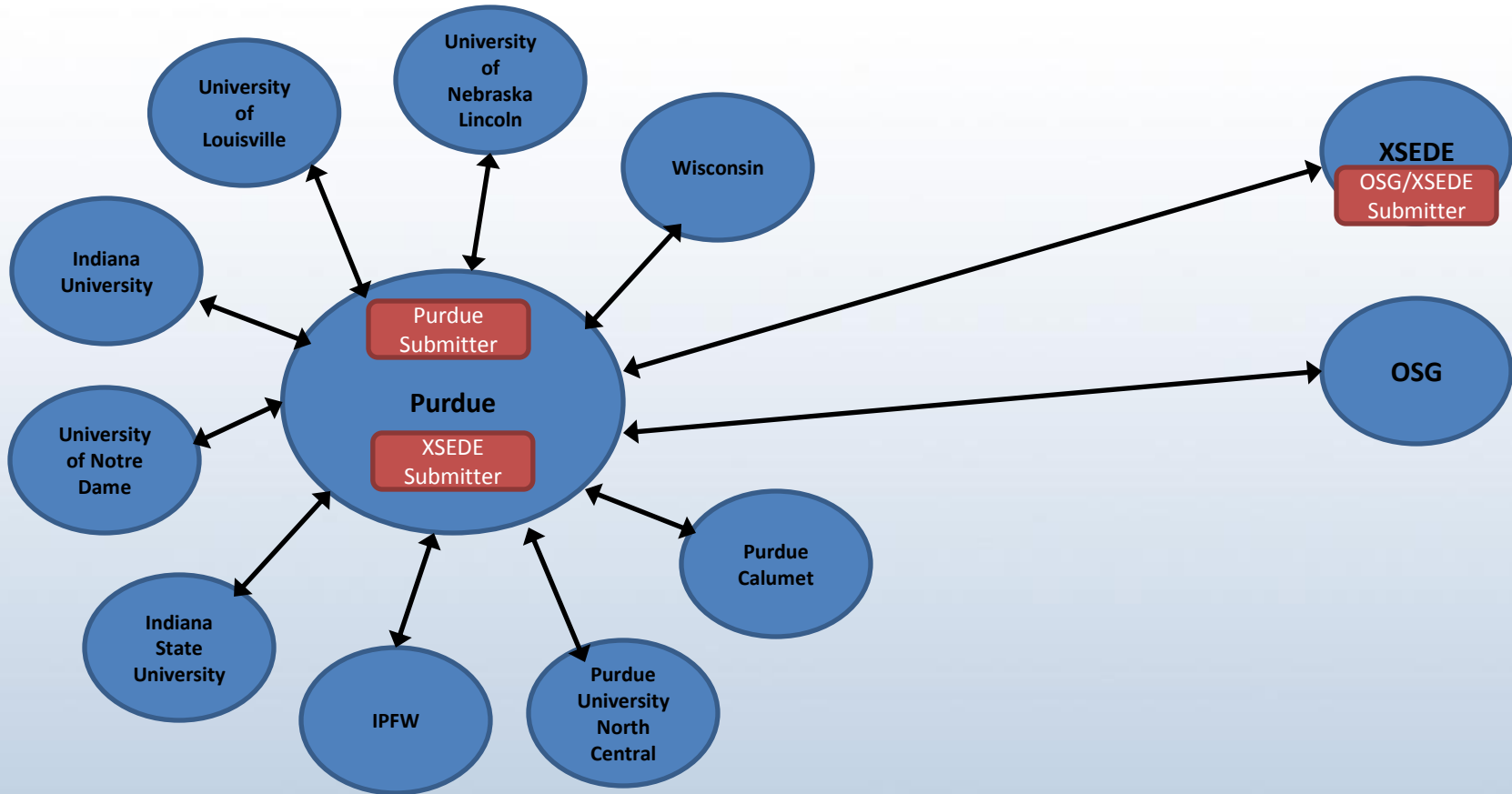
# OSG Usage

Percent Used By Field of Science

4800264 SUs Total



# Diagrid/XSEDE/OSG Connectivity Diagram





XSEDE has another “resource” to help  
Campuses help their users....

## The Campus Champions Program!

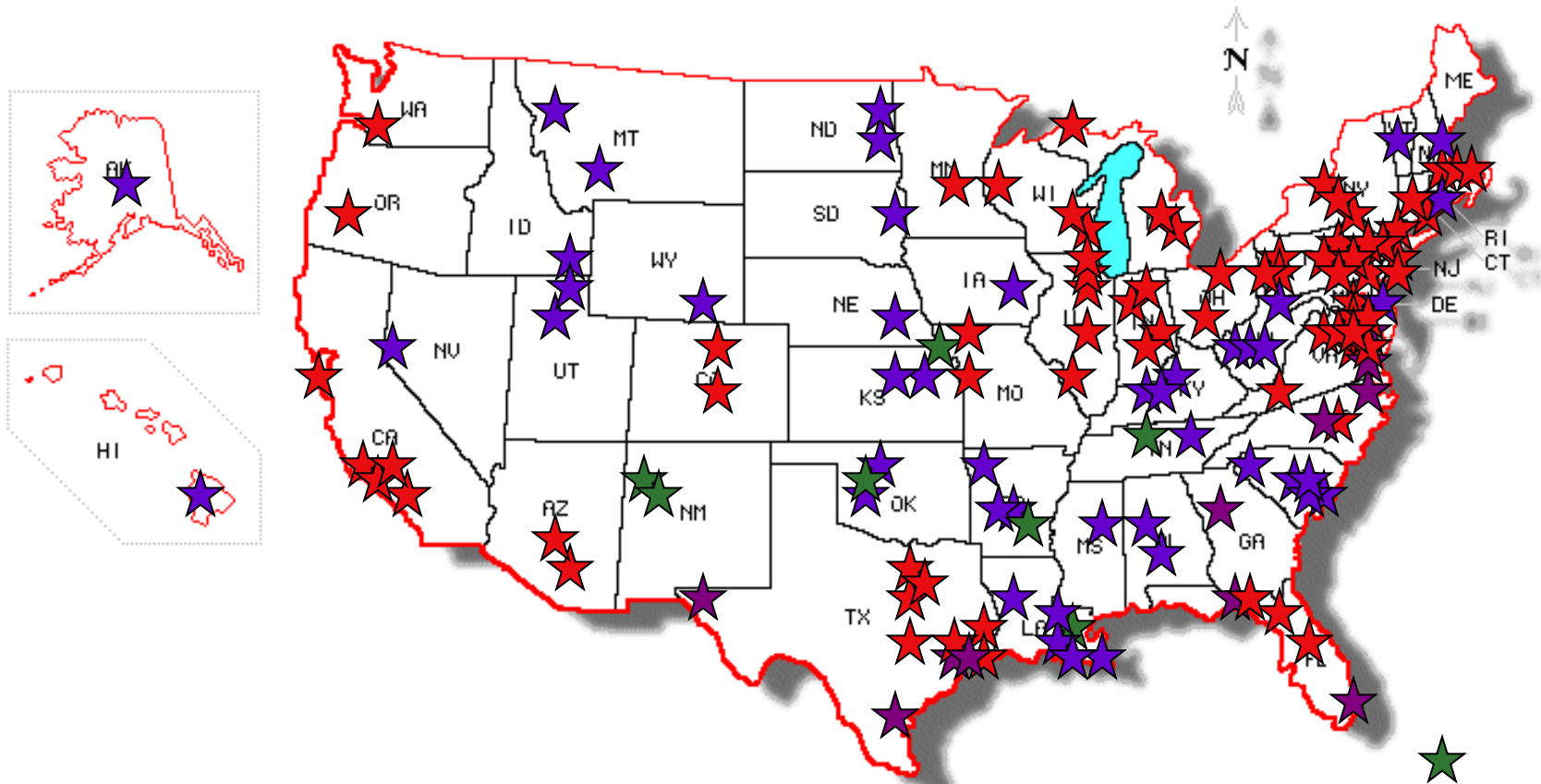
# Campus Champion Program





- Kay Hunt, Purdue, coordinates the program
- Launched in 2007
- More than 100 member campuses today



# Campus Champion Institutions

March 8, 2013



-  Current Campus Champion Institutions (unclassified) – 73
  -  Current Campus Champion Institutions (EPSCoR states) 45
  -  Current Campus Champion Institutions (Minority Serving Institutions)--11
  -  Current Campus Champion Institutions (both EPSCoR and MSI) – 8
- Total Number of Campus Champion Institutions Overall -- 137**

# XSEDE Support of Champions

- Support provided from across XSEDE
- Champions provided monthly training and updates
- XSEDE staff as liaisons for Champions
- Champions provided with start-up account
- Champions are members of User Services team
- Forum for sharing and interactions
- Access to information on usage by their users
- Waive registrations for annual XSEDE Conference

# Campus Champion Role

- Raise awareness locally
- Provide training
- Get users started with access quickly
- Represent needs of local community
- Provide feedback to improve services
- Attend annual XSEDE conference
- Share campus training and education materials
- Build community among Champions



# Strategies of Success

- CI Days events on campus
- Act as regional source of information
- Champions co-present at conferences to promote and grow the program
- Champion focused sessions at SC'xx and XSEDE'xx
- Significant increase in number of new XSEDE users
- Large number of under-represented institutions have joined



# XSEDE Campus Bridging Vision

- Provide software and training tools for interoperation with XSEDE infrastructure
- Make better use of the nation's aggregate CI resources
- Campus Bridging is a set of tools, techniques, and consulting
- Tools for doing this:
  - Installers
  - Documentation & training
  - Ability to contribute community resources for greater good



# The “program” from a Champion’s perspective...

- A Champion’s role is to assist a user in determining the “best” resources to further their “science” including resources other than XSEDE (including local resources if available)
- Champions “network” with each other to ask and answer questions about topics that benefit their local institution and user base even if it is not directly related to XSEDE resources and services
- Champions serve as an important “resource” to XSEDE by providing “insight” into the needs and problems encountered by themselves and their users



Questions?