RESEARCH COMPUTING FOR THE GEOSCIENCES
• IT Research Computing (RCAC)
• A unit of ITaP (Information Technology at Purdue) – the central IT organization at Purdue.
• Research Computing provides advanced computational resources and services to support Purdue faculty and staff researchers.

Our goal: To be the one-stop provider of choice for research computing and data services at Purdue - Delivering powerful, reliable, easy-to-use, service-oriented computing and expertise to Purdue researchers.
COMMUNITY CLUSTERS

A BUSINESS MODEL FOR HPC AT PURDUE UNIVERSITY
• You get out at least what you put in
  • Buy 1 node or 100, you get a queue that guarantees access up to that many CPUs
• But wait, there’s more!!
  • What if your neighbor isn’t using his queue?
    – You can use it, but your job is subject to preemption if he wants to run.
• You don’t have to do the work
  • Your grad student gets to do research rather than run your cluster.
    – Nor do you have to provide space in your lab for computers.
  • ITaP provides data center space, systems administration, application support.
  • Just submit jobs!
• 5 Year cycle
  • We build a cluster **every** year!
  • Vendors provide 5 year warranty
  • After 5 years, MOU with faculty says that the cluster will be retired
  • Faculty get credit for the remaining market value of their nodes, towards the next cluster.
  • Community clusters now appear to funding agencies as paying for a service – not a capital purchase.

• No more preemption
  • Replace with “standby” queue
    - You can run all the jobs you want beyond what your queue would let you, but you’re subject to a time limit of 4 hours.
COMMUNITY CLUSTERS

VITAL STATS

• 170 “owner” partners
• ~1200 active users
• 259M hours provided in 2014
• Nationally, the gold standard for condo-style computing
• HPC and HTC communities prefer different points to optimize the scheduler.
• Small but key communities (like large memory) lose benefits of standby queues when fewer nodes are spread between several clusters.
• HTC or large memory communities often have little need for HPC-specific optimizations
  • Accelerators
  • High-speed, low-latency networks

Emerging communities often don’t fit in existing model at all!

Big Data Analytics
Graphics Rendering
Nontraditional platforms (Windows, cloud)
Rice: A traditional HPC system just like Carter or Conte

20-core Xeon nodes, 64GB RAM

The same, familiar model:
• New cluster acquisition every year
• Each a distinct, non-heterogeneous system.

Nothing different for you!
Hammer – HTC
Snyder – Data-Intensive Life Science

HTC or big memory clusters expanded annually with each purchase.

Better Community Cluster Experience
• Community Clusters are all Linux-based.
  • QGIS
  • GRASS GIS
  • SAGA GIS
  • GDAL

Soil and Water models
  • SWAT
DATA STORAGE INFRASTRUCTURE FOR RESEARCH DATA
Data usage is skyrocketing
Individual researchers work on larger and larger datasets

Conte Scratch Usage, sorted by total TB

User Ranking

TB Used

TB
Researchers find all sorts of solutions
In the past, we’ve heard lots of common requests:

• I need more space than I can get in scratch
• Where can I install applications for my entire research lab?
• I’m actively working on that data/software in scratch:
  • I have to go to great lengths to keep it from being purged.
  • I shouldn’t have to pull it from Fortress over and over
• Can I get a UNIX group created for my students and I?
• Is there storage that I can get to on all the clusters I use?
• I have funding to spend on storage – what do you have to sell?
• I need storage for my instrument to write data into
• My student has the only copy of my research data in his home directory, and he graduated/went off the grid!
As a transport hub: a place where large amounts of cargo are stored, loaded, unloaded, and moved from place to place.
Research storage available for purchase!

A storage service for research to address many common requests:

- 100G available at no charge to research groups
- Mounted on all clusters and exported via CIFS to labs
- *Not scratch*: Backed up via snapshots, with DR coverage
- Data in Depot is owned by faculty member!
- Sharing ability – Globus, CIFS, and WWW
- Maintain group-wide copies of application software or shared data
Well received!

- In less than one year, over 200 research groups are participating.
  - Over 50% are not HPC users!
- Half a PB in use
- A research group purchasing space has purchased, on average, 8.6TB.
Globus:

Transfer and share large datasets....

.... With dropbox-like characteristics ....

.... *Directly from your own storage system!*
GLOBUS

WHAT IS IT?
Data moved in 2015 to date:
  211 TB transferred
  Avg 33 unique users each month

New:
  Globus interface to Fortress

https://transfer.rcac.purdue.edu
• Programming practices – Software Carpentry
• Parallel Programming – MPI, OpenMP
• Big Data
• Matlab
• Accelerators – Xeon Phi, OpenACC, CUDA
• UNIX 101
• Effective use of Purdue research clusters
Repositories for labs, managed by PI’s queue management tool.

Purdue GitHub instance in testing now!
• Need to teach students to use HPC in a course?
• Scholar cluster is available to any instructor at no cost.

Spring 2015:
- EAPS
- CS
- STAT
- AGRY
- ANSC
- CHEM
- ChemE

Just send a CRN

Purdue University
NEED HELP?

• Hard to solve problems with HPC?
• Need help building your software or optimizing your workflow?
• Need to learn what resources are available?

RESEARCH COMPUTING COFFEE BREAK CONSULTATIONS

Meet up with ITaP research computing staff and other researchers who use or are interested in High Performance Computing at Purdue. Join us for informal discussions of scientific computing along with any other topic that might spring up. We’ll be meeting at different coffee shops around campus each week.

Check the coffee web page to see this week’s location and time.
rcac.purdue.edu/coffee
Using in our expertise to help solve your hard science problems.

Our advanced user support staff can partner with your group to take advantage of the latest technology in advanced computation, more effectively use storage and compute systems, and more.

**Domain experts in:**
- Astrophysics
- Data Science
- Bioinformatics
- Chemistry
- Molecular Dynamics
- Earth and Atmospheric Sciences
- Data Visualization
In experimental system developed in 2011/2012

...theory, an HPC system for Windows users!

Found many potential users:

- Matlab
- STATA
- MPI codes
- Financial models in Excel
The big winner?

GIS

Two GIS faculty members invested in the cluster, used HPC facilities.

Others: Don’t want a batch system – use Windows HPC as a remote desktop for interactive computations!
Microsoft ended support for Windows HPC server at the beginning of 2015.

Your cluster will continue to run, but what will come after?
LOUD SERVICES

FEAR YOUR DATA

TOO COSTLY TO MOVE YOUR DATA?

firebox virtual servers

Host LAMP servers, cluster login nodes, submission portals, nontraditional HPC, or interactive desktops, all within the research infrastructure

Move your computing environment close to your research data!

ps://www.rcac.purdue.edu/services/firebox
REBOX VMS

FOR NON-UNIX USERS

We can provide research group-dedicated Windows systems for social sciences, statistical packages, and GIS. Suitable for computationally-intensive tasks!
**IT'S A GIS CLUSTER**

Even a *high-performance connection to Data Depot storage*:

- One Large, powerful system to share with others?
- Several cluster-node grade systems on which to reserve access?
  - Buy like community clusters?
- Dynamic, on-demand GIS systems via the cloud?
  - Pay for service as you go?

GIS Databases (postGIS, etc)

GPUs?

Local flash storage?
Thank you!

QUESTIONS?

We want to hear your use cases!

Questions?

Contact Us:
rcac-help@purdue.edu
@PurdueRCAC
http://www.facebook.com/PurdueRCAC