Introducing Gardner
Center for Research Informatics

- Established in 2011 to support BSD research

- Mission:
  - To provide informatics resources and service to the BSD, to participate in clinical and biomedical research of the highest scientific merit, and to support and promote research and education in the field of informatics
Resources and Services

• Clinical data for research
• Bioinformatics data analysis
• Computing infrastructure
  – Storage
  – HPC
  – Virtual Servers
• Research data management tools
• Custom-built applications
• Educational opportunities

http://cri.uchicago.edu
CRI Infrastructure Team

- **Director**
  - Thorbjorn Axelsson

- **High Performance Computing**
  - Mike Jarsulic
  - Tony Aburaad

- **Virtual Servers**
  - Andy Brook
  - Sneha Jha

- **Storage**
  - Olumide Kehinde

- **Utility Infielder**
  - Dan Sullivan
About Me

- Lived in Pittsburgh for about 32 years
- Attended the University of Pittsburgh (at Johnstown)
- Bettis Atomic Power Laboratory (2004-2012)
  - Scientific Programmer (Thermal/Hydraulic Design)
  - Analyst - USS Gerald R. Ford
  - High Performance Computing
- University of Chicago (2012 – present)
USS Gerald R. Ford
A Few Weeks Ago!!!

“Gerald R. Ford USS, what a place...it really feels like a place.”
About Tony

• Masters student in computer science at UChicago
  – Completing coursework in machine learning, distributed computing, and iOS
• Spent last summer at the Computation Institute working on a caching tool for the Open Science Grid
• Has been helping with Gardner at the CRI since November
• Dislikes mimes
CRI HPC Clusters
September 2012

• Prudential Data Center
  – BRDFCLUSTER
  – IBICLUSTER
  – IBIBMEM

• Kenwood Data Center
  – BIOCluster
Tarbell

- Purchased by the CRI in 2012 by the previous staff
- Dell cluster utilizing AMD Bulldozer processors
- Infiniband QDR
- 110 TB Scratch Space
- Why named Tarbell?
Who was Harlan Tarbell?

• Born in Delavan, IL
• Grew up in Groveland, IL
• Magician
• Doctor of Naprapathy
• Futurist

Themes:
• Beginner mistakes
• Predicting the future
• Quackery
Beginner Mistakes

• Scratch space
  – Set up poorly where the system would become unstable
  – Utilized only 60 TB of space initially
  – Hardware had low RAM (24 GB per node)

• Login node
  – Only one (fixed)
Predicting the Future

• Compute Nodes
  – Only one tier of memory (fixed)

• Infiniband
  – Expecting QDR to stick around forever
  – Poor strategy for future clusters
AMD Bulldozer

- Did not live up to expectations
- Shared Floating Point Unit
- Lawsuit
Quackery
Tarbell Metrics

• Since December 2013
  – 234 Users
  – Total User Jobs: 4.6 Million
  – Total CPU Hours: 18.29 Million
  – Average Queue Hours: 2.94 Hours
  – Average Job Efficiency: 65%
  – Average Wall Clock Accuracy: 11%
Who was Martin Gardner?

- Graduate of the University of Chicago
- Yeoman on the USS Pope during WWII
- Amateur Magician
- Mathematical Games
- Skepticism
- Literature
- Art
Mathematical Games

- Flexagons
- Polynominoes
- Game of Life
- Newcomb’s Paradox
- Mandelbrot's Fractals
- Penrose Tiling
- Public Key Cryptography
- Best bet for simpletons paradox
Mathematical Games
Skepticism

• Original founders of CSICOP

• Critic of:
  – Lysenkoism
  – Homeopathy
  – Chiropractic
  – Naturopathy
  – Orgone Chambers
  – Dianetics
Literature and Art
Also of Interest...

![The Flight of Peter Fromm by Martin Gardner](image1)

![The Whys of a Philosophical Scrivener by Martin Gardner](image2)
What is HPC?
## Node Count Comparison

<table>
<thead>
<tr>
<th>Node Type</th>
<th>Tarbell</th>
<th>Gardner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Compute Nodes</td>
<td>34</td>
<td>88</td>
</tr>
<tr>
<td>Mid-Tier Compute Nodes</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>Large Memory Nodes</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>GPU Nodes</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Xeon Phi Nodes</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Interactive Nodes</td>
<td>2</td>
<td>2 (eventually 4)</td>
</tr>
<tr>
<td>Remote Viz Nodes</td>
<td>0</td>
<td>Possibly 2</td>
</tr>
</tbody>
</table>
# Core Count Comparison

<table>
<thead>
<tr>
<th>Node Type</th>
<th>Tarbell</th>
<th>Gardner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Compute Nodes</td>
<td>2176</td>
<td>2464</td>
</tr>
<tr>
<td>Mid-Tier Compute Nodes</td>
<td>0</td>
<td>784</td>
</tr>
<tr>
<td>Large Memory Nodes</td>
<td>80</td>
<td>112</td>
</tr>
<tr>
<td>GPU Nodes</td>
<td>0</td>
<td>140</td>
</tr>
<tr>
<td>Xeon Phi Nodes</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>2256</strong></td>
<td><strong>3528</strong></td>
</tr>
</tbody>
</table>
# Standard Node Comparison

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Tarbell</th>
<th>Gardner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>AMD Opteron 6274</td>
<td>Intel Haswell E5-2683 v3</td>
</tr>
<tr>
<td>Clock Speed</td>
<td>2.2 GHz</td>
<td>2.0 GHz</td>
</tr>
<tr>
<td>Processors per Node</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Cores per Processor</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>Instructions per Cycle</td>
<td>8 (or 4)</td>
<td>16</td>
</tr>
<tr>
<td>RAM per Core</td>
<td>4 GB</td>
<td>4.5 GB</td>
</tr>
</tbody>
</table>
## Mid-Tier Compute Nodes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Gardner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>Intel Haswell E5-2683 v3</td>
</tr>
<tr>
<td>Clock Speed</td>
<td>2.0 GHz</td>
</tr>
<tr>
<td>Processors per Node</td>
<td>2</td>
</tr>
<tr>
<td>Cores per Processor</td>
<td>14</td>
</tr>
<tr>
<td>Instructions per Cycle</td>
<td>16</td>
</tr>
<tr>
<td>RAM per Core</td>
<td>16 GB</td>
</tr>
</tbody>
</table>
# Large Memory Node Comparison

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Tarbell</th>
<th>Gardner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>Intel Westmere E7-4860</td>
<td>Intel Haswell E5-2683 v3</td>
</tr>
<tr>
<td>Clock Speed</td>
<td>2.27 GHz</td>
<td>2.0 GHz</td>
</tr>
<tr>
<td>Processors per Node</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Cores per Processor</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>Instructions per Cycle</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>RAM per Core</td>
<td>25.6 GB</td>
<td>45.7 GB</td>
</tr>
</tbody>
</table>
## GPGPU Nodes

<table>
<thead>
<tr>
<th>CPU Attribute</th>
<th>Gardner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>Intel Haswell E5-2683 v3</td>
</tr>
<tr>
<td>Clock Speed</td>
<td>2.0 GHz</td>
</tr>
<tr>
<td>Processors per Node</td>
<td>2</td>
</tr>
<tr>
<td>Cores per Processor</td>
<td>14</td>
</tr>
<tr>
<td>Instructions per Cycle</td>
<td>16</td>
</tr>
<tr>
<td>RAM per Core</td>
<td>8 GB</td>
</tr>
<tr>
<td>Accelerator</td>
<td>Nvidia Tesla K80</td>
</tr>
<tr>
<td>GPU</td>
<td>Tesla GK210 (x2)</td>
</tr>
<tr>
<td>Cores per GPU</td>
<td>2496</td>
</tr>
<tr>
<td>RAM per Accelerator</td>
<td>24 GB</td>
</tr>
</tbody>
</table>
## Xeon Phi Nodes

<table>
<thead>
<tr>
<th>CPU Attribute</th>
<th>Gardner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>Intel Haswell E5-2683 v3</td>
</tr>
<tr>
<td>Clock Speed</td>
<td>2.0 GHz</td>
</tr>
<tr>
<td>Processors per Node</td>
<td>2</td>
</tr>
<tr>
<td>Cores per Processor</td>
<td>14</td>
</tr>
<tr>
<td>Instructions per Cycle</td>
<td>16</td>
</tr>
<tr>
<td>RAM per Core</td>
<td>8 GB</td>
</tr>
<tr>
<td>Accelerator</td>
<td>Intel Xeon Phi 5110P (x2)</td>
</tr>
<tr>
<td>Cores per Accelerator</td>
<td>60</td>
</tr>
<tr>
<td>RAM per Accelerator</td>
<td>8 GB</td>
</tr>
</tbody>
</table>
# Scratch Space Comparison

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Tarbell</th>
<th>Gardner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>Intel Westmere E5620</td>
<td>Intel Haswell E5-2623 v3</td>
</tr>
<tr>
<td>Clock Speed</td>
<td>2.4 GHz</td>
<td>3.0 GHz</td>
</tr>
<tr>
<td>Processors per Node</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Cores per Processor</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Instructions per Cycle</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>RAM per Node</td>
<td>24 GB</td>
<td>64 GB</td>
</tr>
<tr>
<td>Cache Pool</td>
<td>N/A</td>
<td>200 GB</td>
</tr>
<tr>
<td>Usable Space</td>
<td>110 TB</td>
<td>350 TB</td>
</tr>
<tr>
<td>Interconnect Bandwidth</td>
<td>40 Gb/s</td>
<td>56 Gb/s</td>
</tr>
</tbody>
</table>
Benchmarking

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Tarbell</th>
<th>Gardner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical Performance</td>
<td>44.2 TFLOPs</td>
<td>112.8 TFLOPs</td>
</tr>
<tr>
<td>Actual Performance</td>
<td>21.2 TFLOPs</td>
<td>97 TFLOPs</td>
</tr>
<tr>
<td>GPU Theoretical Performance</td>
<td>N/A</td>
<td>14.5 TFLOPs</td>
</tr>
<tr>
<td>GPU Actual Performance</td>
<td>N/A</td>
<td>11.4 TFLOPs</td>
</tr>
<tr>
<td>Xeon Phi Theoretical Performance</td>
<td>N/A</td>
<td>2 TFLOPs</td>
</tr>
<tr>
<td>Xeon Phi Actual Performance</td>
<td>N/A</td>
<td>1.7 TFLOPs</td>
</tr>
</tbody>
</table>

FLOPs = Nodes * Number of Cores/Node * Frequency * Operations per Cycle
Software

• Compilers
  – Intel
  – PGI
  – GNU
  – Java 7 and 8
  – DLang

• MPI
  – OpenMPI
  – MPICH
  – Intel MPI

• Software Environment
  – Lmod

• Scheduler
  – Moab 9.1

• Resource Manager
  – Torque 6.1
What is Going to Happen To?

• Tarbell
  – Decommissioned: 3/31/17

• LMEM-CRI
  – Decommissioned

• Stats
  – Repurposed
  – X enabled login nodes for the cluster
  – Commercial software: SAS, Stata, MATLAB, etc.

• Galaxy
  – Decommissioned with Tarbell
Obtaining an Account

• Prerequisites: BSD Account
• Sign up for and account
  – [http://cri.uchicago.edu](http://cri.uchicago.edu)
  – Early Access
    • Email Address for Job Output
    • Emergency Phone Number
    • Software Requests
    • Level of Experience
  – Collaborator Accounts
Being a Good HPC Citizen

1. Do not run analysis on the Login Nodes!
2. Cite the cluster and the software used in your publications.
3. Try to be accurate with your resource requests.
4. Allow the CRI to install open source software for you.
5. If you are going to run an analysis that is much larger than normal, let us know in advance.
Being a Good HPC Citizen

6. Provide feedback.
7. Clean up your Scratch Storage.
8. If using a script to submit, sleep for a few seconds in between each submission.
9. Be sure to release memory in your scripts.
10. If you have a question, don’t hesitate to ask us.
11. If you notice a problem, report it.
Citations

• The continued group and support of the CRI’s HPC program is dependent on demonstrable value.

• Citing the cluster allows us to justify purchasing faster clusters with more capacity in the future.

• Sample Citation:
  – This work utilized the computational resources of the Center for Research Informatics’ Gardner HPC cluster at the University of Chicago (http://cri.uchicago.edu).

• Make sure you site the software used as well!
Software Installation

• Software request can be submitted via the Resource Request forms at http://cri.uchicago.edu

• Advantages to allowing the CRI to install open source software:
  – Other users can utilize it
  – Support nightmare
  – Portability

• Disadvantages
  – It may take a few days (let us know the priority)
How to Get Support

• Call the CRI Help Desk
  – 773-834-8475
• Email [hpc@rt.cri.uchicago.edu](mailto:hpc@rt.cri.uchicago.edu) to submit a ticket or use the Request Forms on the CRI Website
• Meet with Mike at our Peck Office (N161)
  – Tuesday and Thursday Afternoons
  – Schedule an appointment
• User Group Meetings
  – Once a month at Peck
Examples

- Get an account
  - Resource Request Form
- Have software installed
  - Resource Request Form
- Job extension
  - Email hpc@rt.cri.uchicago.edu
  - CC: Mike (mjarsulic@bsd.uchicago.edu)
- Major problem on the cluster
  - Call Help Desk
  - Email hpc@rt.cri.uchicago.edu
Logging In

• On Campus
  – ssh to gardner.cri.uchicago.edu

• Off Campus
  – VPN
    • CVPN (CNET Account Required)
    • BSDVPN
  – ssh to gardner.cri.uchicago.edu
Storage

- Home Directories (/home/<userid>)
  - Permanent, Private, Quota’d, Not Backed Up
  - 1 Gb/s

- Lab Shares (/group/<lab_name>)
  - Permanent, Shared, Quota’d, Backed Up
  - 1 Gb/s

- Scratch Space (/scratch/<userid>)
  - Purged, Private, Not Quota’d, Not Backed Up
  - 56 Gb/s
  - Purged every 6 months (to start)
Software Environment

• Tarbell -> Environment Modules
  – Flat module system
  – Modules written in TCL
  – Last Update: December 2012

• Gardner -> Lmod
  – Hierarchical module system
  – Modules written in Lua
  – Last Update: August 2016
Lmod Basics

• See which modules are available to be loaded
  ─ module avail

• Load packages
  ─ module load <package1> <package2>

• See which packages are loaded
  ─ module list

• Unload a package
  ─ module unload <package>
Scheduling Jobs (Defaults)

• Maximum Amount of Walltime
  – 14 Days

• Maximum Amount of Processors
  – 500 concurrent

• Maximum amount of jobs
  – 500 concurrent

• Maximum amount of memory
  – 2 TB
Job Scheduling (Queues)

• Route
  – Default Queue (non-executable)

• Express
  – 1 node; 1 proc; <= 4 GB RAM; <= 6 hours

• Standard
  – Multi-node; Multi-proc; <= 8 GB RAM
Job Scheduling (Queues)

• Mid
  – Multi-node; Multi-proc; > 8 GB RAM; <=24 GB RAM

• High
  – Multi-node; Multi-proc; > 24 GB RAM
Torque Client Commands

• Submit a job
  – qsub <scriptname>

• Delete a job
  – qdel <jobid>

• Job status
  – qstat

• Extended Job Status
  – qstat -f
Torque Directives

• Specify a Job Name
  – #PBS -N <JobName>

• Specify nodes and cores
  – #PBS -l nodes=x:ppn=y

• Specify wall clock time limit
  – #PBS -l walltime=[dd:[hh:[mm:]]]ss

• Specify the memory limit
  – #PBS -l mem=<x>gb
Torque Directives

• Specify the shell to execute the script
  – #PBS -S <path_to_shell>

• Specify the STDOUT location
  – #PBS -o <path>

• Specify the STDERR location
  – #PBS -e <path>
qsub Arguments

• Run and interactive job
  - qsub -I

• Submit a job and immediately hold it
  - qsub -h <jobscript>
Volume of a Molecule
Other Possible New Features

- Web Portal (w/ Templates)
- Remote Visualization
- Data Staging
- NUMA controlled jobs
- Improved checkpointing