Yellowstone allocations and access schedule

David Hart, CISL/USS
CHAP Meeting
May 3, 2012
Anticipation and transition

- Allocation facilities
- Allocation requests thus far
  - Accelerated Scientific Discovery (ASD)
  - Climate Simulation Laboratory (CSL)
  - NCAR Strategic Capability and ASD
  - CHAP
  - WRAP
- Bluefire extension period (July-September)
- Access schedule
- Yellowstone user survey
- User environment plans
- Managing the user transition
  - System Accounting Manager (SAM)
Allocation opportunities thus far

ANTICIPATION
Yellowstone funding sources

- NSF Core Funding: 65%
- CSL: 28%
- NSF EaSM (Univ & NCAR): 7%
- University: 40% of Core
- Wyoming: 20% of Core
- NCAR: 40% of Core
Yellowstone allocation “facilities”

The segments for CSL, University and NCAR users each represent about **170 million core-hours per year**. WNA about **75 million core-hours per year**.
Allocation requests thus far

Averaging 170% over-requested
University ASD projects

- **James Kinter, COLA — 21 million core-hours**
  - Towards seamless high-resolution prediction at intraseasonal and longer timescales
- **Lance Collins, Cornell U — 19 million core-hours**
  - DNS of cumulus cloud core processes over larger volumes and for longer times
- **B. Fox-Kemper, CU-Boulder — 16 million core-hours**
  - Arrest of frontogenesis in oceanic submesoscale turbulence
- **Thomas Jordan, USC — 7.3 million core-hours**
  - Community computational platforms for developing 3-D models of earth structure
- **Michael Shay, U Delaware — 7.2 million core-hours**
  - Turbulence in the heliosphere: The role of current sheets and magnetic reconnection
Climate Simulation Lab

- 9 requests ranging from 10.5 million to 31 million core-hours
  - 6 requests were new to CSL
- Plus very large request from CESM
- Panel made six awards for 217.2 million core-hours
  - Since total recommended is less than available, plan is to shorten the award period by about two months.
**NCAR Strategic Capability & ASD**

- New approach to strategically allocate most of NCAR’s portion to merit-reviewed, project-based requests
  - Minimum size: 5 million core-hours
- **17 requests** ranging from 5.2 million to 106 million, for 281 million core-hours total
  - Encompassed both ASD and NSC projects
- **15 awards — 6 ASD**
  - 162 million core-hours awarded (60M ASD)
CHAP & WRAP

• CHAP — 30 requests (2 re-classed “small”)
  – 72 million core-hours requested
  – Ranging from 235,940 to 25 million core-hours
  – Normally ~85 million core-hours available per meeting. Less this time due to Oct 11 awards, ASD stretch, Yellowstone delay

• WRAP — 7 requests
  – 68.9 million core-hours requested
  – Ranging from 1 million to 43.2 million core-hours
  – 45 million core-hours available
CHAP request sizes
Moving from Bluefire to Yellowstone

TRANSITION
Bluefire extension (July-Sept)

- Bluefire will be around at least until September 30, 2012
- Three months unallocated due to original Yellowstone delivery schedule
- Continuing to make small University awards, emphasizing Bluefire end of life
- Making limited awards to existing large CHAP-reviewed projects to permit them to wrap-up or reach useful stopping point
- CSL, NCAR: Will be extending monthly allocations
4% of Yellowstone (~1 Bluefire-equivalent) available for other communities to begin porting and testing work during ASD.
Yellowstone user “pre-survey”

- **User survey**—Dec 2011 – Jan 2012
  - 163 respondents (88 university, 72 NCAR, 9 other)
- **Survey results in four bullet points:**
  - Scheduling—longer wall clock limit is good, but protect turnaround
  - Software—“lots” is good, but not dramatically different from what’s currently on Bluefire
  - Data movement/management—more disk for longer periods
  - Support—Better docs. You have training?
- **Compiled list of “action items” from survey questions**
  - Some items already done; to be done; should we do?
- **Discussed and refined plans with HSS**
- **Discussed plans, got “focus group” feedback from NESL, HAO**
- **Will soon be post response to survey on CISL web site, to accompany our “Transition to Yellowstone” documentation**
User environment plans

• “modules” will be used to manage software environment
  – In use on Bluefire, but critical on Yellowstone

• CSG defining software/module installation/maintenance/upgrade process
  – Using Janus as a testing ground
  – The software packages themselves aren’t the problem, but rather using them in a four-compiler environment

• tcsh will be default shell
  – User changeable option
GLADE file systems and policies

• Four file systems with different policies confusing to users (home, work, scratch, project)
  – Name of “work” space (/glade/users) adds further confusion

• Move to three file spaces only
  – /glade/home : same as now
  – /glade/scratch : same as now
  – /glade/work (combines users & project)
    • No scrubbing
    • /glade/work/[username]—500 GB automatic for all users
    • /glade/project/[project_code]—as allocated
Geyser/Caldera usage model

- All use of Geyser and Caldera managed via LSF
  - Interactive, shared or exclusive
  - Batch, shared or exclusive
- X-Forwarding will be supported but VirtualGL (i.e. vnc clients) is recommended
- Use these clusters for “share” jobs
  - At least until load dictates otherwise
- Access via LSF will be a significant change
  - Benefits include automation of workflows between HPC and DAV nodes, load balancing, and usage monitoring
### Geyser/Caldera scheduling

- **Daytime use** primarily shared (both interactive and batch), but allow some batch, exclusive use
- **Night/weekend use** less constrained, equal opportunity for batch-exclusive use
  - E.g., GPGPU code testing

<table>
<thead>
<tr>
<th>Queue</th>
<th>Wall clock</th>
<th>Job size</th>
<th>Priority</th>
<th>Q factor</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>geyser</td>
<td>24 hr</td>
<td>1-40</td>
<td>1</td>
<td>1.0</td>
<td>Interactive or batch, shared or exclusive use</td>
</tr>
<tr>
<td>bigmem</td>
<td>6 hr</td>
<td>40-640</td>
<td>1</td>
<td>1.0</td>
<td>Batch exclusive use, 4-node limit during daytime</td>
</tr>
<tr>
<td>caldera</td>
<td>24 hr</td>
<td>1-16</td>
<td>1</td>
<td>1.0</td>
<td>Interactive or batch, shared or exclusive use</td>
</tr>
<tr>
<td>gpgpu</td>
<td>6 hr</td>
<td>16-256</td>
<td>1</td>
<td>1.0</td>
<td>Batch exclusive use, 4-node limit during daytime</td>
</tr>
</tbody>
</table>
# HPC scheduling and fair share

<table>
<thead>
<tr>
<th>Queue</th>
<th>Wallclock</th>
<th>Job size</th>
<th>Priority</th>
<th>Q factor</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>capability</td>
<td>12 hr</td>
<td>16,384–65,536</td>
<td>2</td>
<td>1.0</td>
<td><em>Execution window: Friday noon – Monday 6 a.m.</em></td>
</tr>
<tr>
<td>regular</td>
<td>12 hr</td>
<td>16–16,384</td>
<td>2</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>premium</td>
<td>12 hr</td>
<td>16–16,384</td>
<td>1</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>economy</td>
<td>12 hr</td>
<td>16–16,384</td>
<td>3</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>small</td>
<td>2 hr</td>
<td>16–4,096</td>
<td>1.5</td>
<td>1.0</td>
<td>8am-5pm only</td>
</tr>
<tr>
<td>standby</td>
<td>12 hr</td>
<td>16–16,384</td>
<td>4</td>
<td>1.0 (0?)</td>
<td>30/90 hold, overrun queue</td>
</tr>
<tr>
<td>hpss</td>
<td>24 hr</td>
<td>1</td>
<td></td>
<td>n/a</td>
<td>For HPSS and data transfer</td>
</tr>
</tbody>
</table>

- Using same dynamic priority formula as Bluefire, modified accordingly for longer wall clock limits

Note: No debug or share queues.
Managing the transition

• Many (but not all) projects will be automatically migrated from Bluefire to Yellowstone
  – Will use the transition to eliminate inactive projects and user accounts, as well as unnecessary legacy system and user info

• Bluefire to Yellowstone accompanied by accounting system transition (ACC8 ➔ SAM)
  – New capabilities and interfaces
  – Considerable effort from USS/WEG/HSS to automate processes, streamline tasks, and provide a consistent user environment across resources
QUESTIONS?