Fast computers, big/fast storage, fast networks

Marla Meehl
Manager, Network Engineering and Telecommunications, NCAR/UCAR, Manager of the Front Range GigaPoP
Computational & Information Systems Laboratory
National Center for Atmospheric Research
## Current HPC Systems at NCAR

<table>
<thead>
<tr>
<th>System Name</th>
<th>Vendor System</th>
<th># Frames</th>
<th># Processors</th>
<th>Processor Type</th>
<th>GHz</th>
<th>Peak TFLOPs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Production Systems</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bluefire (IB)</td>
<td>IBM p6 Power 575</td>
<td>11</td>
<td>4096</td>
<td>POWER6</td>
<td>4.7</td>
<td>77.00</td>
</tr>
<tr>
<td><strong>Research Systems, Divisional Systems &amp; Test Systems</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firefly (IB)</td>
<td>IBM p6 Power 575</td>
<td>1</td>
<td>192</td>
<td>POWER6</td>
<td>4.7</td>
<td>3.610</td>
</tr>
<tr>
<td>frost (BG/L)</td>
<td>IBM BlueGene/L</td>
<td>4</td>
<td>8192</td>
<td>PowerPC-440</td>
<td>0.7</td>
<td>22.935</td>
</tr>
</tbody>
</table>
High Utilization, Low Queue Wait Times

- Bluefire system utilization is routinely >90%
- Average queue-wait time < 1hr

Average Queue Wait Times for User Jobs at the NCAR/CISL Computing Facility

<table>
<thead>
<tr>
<th>Peak TFLOPs</th>
<th>Bluefire since Jun’08</th>
<th>Lightning since Dec’04</th>
<th>Blueice Jan’07-Jun’08</th>
<th>Bluevista Jan’06-Sep’08</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>77.0 TFLOPs all jobs</td>
<td>1.1 TFLOPs all jobs</td>
<td>12.2 TFLOPs all jobs</td>
<td>4.4 TFLOPs all jobs</td>
</tr>
<tr>
<td>Premium</td>
<td>00:05</td>
<td>00:11</td>
<td>00:11</td>
<td>00:12</td>
</tr>
<tr>
<td>Regular</td>
<td>00:36</td>
<td>00:16</td>
<td>00:37</td>
<td>01:36</td>
</tr>
<tr>
<td>Economy</td>
<td>01:52</td>
<td>01:10</td>
<td>03:37</td>
<td>03:51</td>
</tr>
<tr>
<td>Stand-by</td>
<td>01:18</td>
<td>00:55</td>
<td>02:41</td>
<td>02:14</td>
</tr>
</tbody>
</table>

Bluefire System Utilization (daily average)

- Spring ML Powerdown
- May 3 Firmware & Software Upgrade
Bluefire Usage

- **Climate**: 49.3%
- **Accelerated Scientific Discovery**: 18.0%
- **Oceanography**: 4.0%
- **Atmospheric Chemistry**: 4.0%
- **Basic Fluid Dynamics**: 2.6%
- **Astrophysics**: 4.4%
- **Cloud Physics**: 1.3%
- **Upper Atmosphere**: 0.3%
- **Miscellaneous**: 0.3%
- **Weather**: 10.0%

**Total Usage**: 100.0%
CURRENT NCAR COMPUTING >100TFLOPS

Peak TFLOPs at NCAR (All Systems)
CISL Computing Resource Users
1347 Users in FY09

- Univ Faculty, 191, 14%
- Graduate and Undergrad Students, 343, 25%
- NCAR Scientists & Proj Scientists, 160, 12%
- NCAR Assoc Sci & Support Staff, 252, 19%
- University Research Associates, 280, 21%
- Government, 108, 8%
- Other, 13, 1%

CISL Computing Resource Users in FY09
1347 Users
Universities with largest number of users in FY09

Users are from 114 U.S. Universities
Disk Storage Systems

- Current Disk Storage for HPC and Data Collections:
  - 215 TB for Bluefire (GPFS)
  - 335 TB for Data Analysis and Vis (DAV) (GPFS)
  - 110 TB for Frost (Local)
  - 152 TB for Data Collections (RDA, ESG) (SAN)

- Undergoing extensive redesign and expansion.
  - Add ~2PB to (existing GPFS system) to create large central (x-mounted) filesystem
Data Services Redesign

Near-term Goals:
- Creation of unified and consistent data environment for NCAR HPC
- High-performance availability of central filesystem from many projects/systems (RDA, ESG, CAVS, Bluefire, TeraGrid)

Longer-term Goals:
- Filesystem cross-mounting
- Global WAN filesystems

Schedule:
- Phase 1: Add 300TB (now)
- Phase 2: Add 1 PB (March 2010)
- Phase 3: Add .75 PB (TBD)
Mass Storage Systems

- NCAR MSS (Mass Storage System)
  - *In Production since 70s*

- Transitioning from the NCAR Mass Storage System to HPSS
  - *Prior to NCAR-Wyoming Supercomputer Center (NWSC) commissioning in 2012*
  - *Cutover Jan 2011*
  - *Transition without data migration Silos/Tape Drives*

- Tape Drives/Silos
  - *Sun/StorageTek T10000B (1 TB media) with SL8500 libraries*
  - *Phasing out tape silos Q2 2010*
  - *Oozing 4 PBs (unique), 6 PBs with duplicates*
    - “Online” process
    - Optimized data transfer
    - Average 20 TB/day with 10 streams
Estimated Growth

Total PBs Stored (w/ DSS Stewardship & IPCC AR5)
Facility Challenge Beyond 2011

- **NCAR Data Center**
  - NCAR data center limits of power/cooling/space have been reached
  - Any future computing augmentation will exceed existing center capacity
  - Planning in progress

- **Partnership**

- **Focus on Sustainability and Efficiency**

- **Project Timeline**
NCAR-Wyoming Supercomputer Center (NWSC) Partners

- NCAR & UCAR
- University of Wyoming
- State of Wyoming
- Cheyenne LEADS
- Wyoming Business Council
- Cheyenne Light, Fuel & Power Company
- National Science Foundation (NSF)

http://cisl.ucar.edu/nwsc/
Focus on Sustainability

- Maximum energy efficiency
- LEED certification
- Achievement of the smallest possible carbon footprint
- Adaptable to the ever-changing landscape of High-Performance Computing
- Modular and expandable space that can be adapted as program needs or technology demands dictate
Focus On Efficiency

- Don’t fight mother nature
- The cleanest electron is an electron never used
- The computer systems drive all of the overhead loads
  - *NCAR will evaluate in procurement*
  - *Sustained performance / kW*
- Focus on the biggest losses
  - *Compressor based cooling*
  - *UPS losses*
  - *Transformer losses*
Design Progress

- Will be one of the most efficient data centers
  - *Cheyenne Climate*
  - *Eliminate refrigerant based cooling for ~98% of the year*
  - *Efficient water use*
  - *Minimal transformations steps electrically*

- Guaranteed 10% Renewable

- Option for 100% (Wind Power)

- Power Usage Effectiveness (PUE)
  - *PUE = Total Load / Computer Load = 1.08 ~ 1.10 for WY*
  - *Good data centers 1.5*
  - *Typical Commercial 2.0*
Expandable and Modular

- **Long term**
  - 24 - acre site
  - Dual substation electrical feeds
  - Utility commitment up to 36MW
  - Substantial capability fiber optics

- **Medium term – Phase I / Phase II**
  - Can be doubled
  - Enough structure for three or
  - four generations
  - Future container solutions or other advances

- **Short term – Module A – Module B**
  - Each Module 12,000 sq.ft.
  - Install 4.5MW initially can be doubled in Module A
  - Upgrade Cooling and Electrical systems in 4.5MW increments
Project Timeline

- 65% Design Review Complete (October 2009)
- 95% Design Complete
- 100% Design Complete (Feb 2010)
- Anticipated Groundbreaking (June 2010)
- Certification of Occupancy (October 2011)
- Petascale System in Production (January 2012)
HPC services at NWSC

- PetaFlop Computing – 1 PFLOP peak minimum
- 100’s PetaByte Archive (+20 PB yearly growth)
  - HPSS
  - Tape archive hardware acquisition TBD
- PetaByte Shared File System(s) (5-15PB)
  - GPFS/HPSS HSM integration (or Lustre)
  - Data Analysis and Vis
  - Cross Mounting to NCAR Research Labs, others
- Research Data Archive data servers and Data portals
- High-Speed Networking and some enterprise services
NWSC HPC Projection

Peak PFLOPs at NCAR

- ARCS Phase 4
- ICESS Phase 1
- ICESS Phase 2
- bluesky
- bluefire
- frost

- IBM POWER6/Power5 (bluefire)
- IBM POWER5+/p575/H (blueice)
- IBM POWER5/p57 (bluevista)
- IBM BlueGene/L
- IBM Opteron/Linux (pegasus)
- IBM Opteron/Linux (lightning)
- IBM POWER4/Co (bluesky)
NCAR Data Archive Projection
Total Data in the NCAR Archive (Actual and Projected)

Petabytes

Jan-99
Jan-00
Jan-01
Jan-02
Jan-03
Jan-04
Jan-05
Jan-06
Jan-07
Jan-08
Jan-09
Jan-10
Jan-11
Jan-12
Jan-13
Jan-14
Jan-15

Total
Unique
Tentative Procurement Timeline

- NDA's
- RFI
- RFI Evaluation
- RFP Draft
- RFP Refinement
- Benchmark Refinement
- RFP Release
- Proposal Evaluation
- Negotiations
- NSF Approval
- Installation & ATP
- Award
- NWSC Production
- Installation & ATP
- Award
- NSF Approval
- Negotiations
- Proposal Evaluation
- RFP Release
- Benchmark Refinement
- RFP Draft
- RFI Evaluation
- RFI
- NDA's

July 2009 (Jul-09)
January 2010 (Jan-10)
July 2010 (Jul-10)
December 2010 (Dec-10)
July 2011 (Jul-11)
January 2012 (Jan-12)
## NWSC Information Crib sheet

### Structural

<table>
<thead>
<tr>
<th>Description</th>
<th>Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size of structure</td>
<td>153,302 square feet</td>
</tr>
<tr>
<td></td>
<td>17,680 penthouse</td>
</tr>
<tr>
<td></td>
<td>170,982 square feet</td>
</tr>
<tr>
<td>Height of structure</td>
<td>62 feet to top of penthouse</td>
</tr>
<tr>
<td>Square footage of raised floor</td>
<td>2 raised floor modules, 12,000 square feet each</td>
</tr>
<tr>
<td>Depth of raised floor</td>
<td>10 feet</td>
</tr>
<tr>
<td>Height from the raised floor surface to the raised floor area ceiling</td>
<td>12 feet</td>
</tr>
<tr>
<td>Height of interstitial space above raised floor ceiling space to hard roof</td>
<td>Average 9 feet</td>
</tr>
<tr>
<td>Size of Administration / Visitor Center space</td>
<td>9,550 first floor</td>
</tr>
<tr>
<td></td>
<td>6,445 second floor</td>
</tr>
<tr>
<td></td>
<td>15,995 net square feet</td>
</tr>
<tr>
<td>Size of the Network Operations Center (NOC)</td>
<td>3,560 net square feet</td>
</tr>
<tr>
<td>Size of the HPSS tape archival (Mass Store) area</td>
<td>2,380 net square feet</td>
</tr>
<tr>
<td>Size of the HPSS expansion area</td>
<td>4,026 nsf²</td>
</tr>
<tr>
<td></td>
<td>includes exercise room</td>
</tr>
<tr>
<td>Size of the Mechanical / Electrical Utility Plant (CUP) space</td>
<td>39,846 - Chiller floor</td>
</tr>
<tr>
<td></td>
<td>11,394 - Main floor</td>
</tr>
<tr>
<td></td>
<td>31,142 - Electrical floor</td>
</tr>
<tr>
<td></td>
<td>17,654 - Penthouse</td>
</tr>
<tr>
<td></td>
<td>100,036 - Total CUP</td>
</tr>
<tr>
<td>No elevation changes (ramps) from the loading dock to the equipment raised floor space</td>
<td>0</td>
</tr>
</tbody>
</table>
Front Range GigaPoP (FRGP)

- 10 years of operation by UCAR
- 15 members
- NLR, I2, Esnet peering @ 10Gbps
- Commodity – Qwest and Level3
- CPS and TransitRail Peering
- Intra-FRGP peering
- Akamai
- 10Gbps ESnet peering
- www.frgrp.net
Fiber, fiber, fiber

- NCAR/UCAR intra-campus fiber – 10Gbps backbone
- Boulder Research and Administration Network (BRAN)
- Bi-State Optical Network (BiSON) – upgrading to 10/40/100Gbps capable
  - Enables 10Gbps NCAR/UCAR Teragrid connection
  - Enable 10Gbps NOAA/DC link via NLR lifetime lambda
- DREAM – 10Gbps
- SCONET – 1Gbps
Bi-State Optical Network (BiSON)

Map by Bryan Anderson
February 9, 2010

GCS_WGS_1984
Datum:D_WGS_1984

LEGEND

- Level 3 Fiber Co. - 156
- Level3 ICG - 40
- PRPA - 30
- Mcleod Fiber - 64
- Broadwing Fiber - 64
- Build Route - 8
Western Regional Network (WRN)

- A multi-state partnership to ensure robust, advanced, high-speed networking available for research, education, and related uses
- Increased aggregate bandwidth
- Decreased costs
Questions