Blue Waters System Overview
Greg Bauer
The Blue Waters EcoSystem

Petascale Education, Industry and Outreach

Petascale Applications (Computing Resource Allocations)

Petascale Application Collaboration Team Support

Outstanding User and Production Support
WAN Connections, Consulting, System Management, Security, Operations, ...

Value added hardware (external networking, IDS, nearline storage, import/export, etc) and software (Eclipse, HPSS, Visualization, computational libraries, etc)

Blue Waters Base System – Processors, Memory, Interconnect, Online Storage, System Software, Programming Environment

National Petascale Computing Facility

Great Lakes Consortium for Petascale Computing
Blue Waters Computing System

- **Aggregate Memory** – 1.5 PB
  - Scuba Subsystem - Storage Configuration for User Best Access
  - 120+ Gb/sec
  - 100-300 Gbps WAN

- **10/40/100 Gb Ethernet Switch**
- **External Servers**
- **IB Switch**
- >1 TB/sec

- **Spectra Logic**: 300 usable PB
- **Sonexion**: 26 usable PB
National Petascale Computing Facility

- Modern Data Center
  - 90,000+ ft² total
  - 30,000 ft² 6 foot raised floor
  - 20,000 ft² machine room gallery with no obstructions or structural support elements

- Energy Efficiency
  - LEED certified Gold
  - Power Utilization Efficiency, PUE = 1.1–1.2
  - 24 MW current capacity – expandable
  - Highly instrumented
Note – UIUC is only institution leading more than one PRAC (4)

UIUC has Co-PIs on several others
The Movement of Data
Cray “Gemini” High Speed Network

- 40 GbE Switch
- 440 Gb/s Ethernet from site network
- 55 GB/s
- 40GbE Switch
- 10 GbE LAN/WAN

Protocols
- LNET
- TCP/IP (10 GbE)
- GridFTP (TCP/IP)
- SCSI (FCP)

Near Line
- 1.2PB Disk
- 300+PB Tape
- 100 GB/s
- 100 GB/s
- 100 GB/s

All storage sizes given as the amount usable

- Online disk >25PB
- /home, /project, /scratch

- Core IB Switches
- 1200 GB/s
- 300 GB/s
- 100 GB/s
- 100 GB/s
- 100 GB/s
Blue Waters Nearline/Archive System

- Spectra Logic T-Finity
  - Dual-arm robotic tape libraries
  - High availability and reliability, with built-in redundancy
- Blue Waters Archive
  - Capacity: 380 PBs (raw), 300 PBs (usable)
  - Bandwidth: 100 GB/sec (sustained)
  - Redundant arrays of independent tapes RAIT for increased reliability
Largest Nearline environment

- HPSS disk cache: 1.2 PB
- HPSS (S)
- 100 GB/s
- 100 GB
- 1.2PB Disk
- 300+PB Tape
- IB
- FC8

HPSS Data mover

LAN

10GigE

40GigE

QDR IB switch

HPSS core (DB2)

HPSS disk cache

1.2 PB

366 IBM TS1140 Jag tape drives
Six 15,930-slot dual-arm libraries
300+PB data storage
10GigE

LAN
Online Storage

- Cray Sonexion with Lustre for all filesystems.
- All visible from compute nodes.

- **home**: 144 OSTs: 2.2 PB usable: 1 TB quota
- **projects**: 144 OSTs: 2.2 PB usable: 3 TB group quota
- **scratch**: 1440 OSTs: 22 PB usable: 500 TB group quota
GO with Globus Online

- GridFTP client development for IE and HPSS nodes.
- Enabled data striping with GridFTP.
- Managed file transfers.
- Command line interface.
- Globus connect for sites without GridFTP endpoints.
Blue Waters High Speed Network

Blue Waters 3D Torus
Size
23 x 24 x 24

Service Nodes
- Operating System
  - Boot
  - System Database
- Login/Network
  - Login Gateways
  - Network
- Lustre File System
  - LNET Routers

Compute Nodes
- Cray XE6 Compute
- Cray XK7 Accelerator

Interconnect Network

InfiniBand
Login
Server(s)
Network(s)
GigE
Fibre Channel

Lustre

Boot Raid
Blue Waters XE6 Node

<table>
<thead>
<tr>
<th>Node Characteristics</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Core Modules*</td>
<td>16</td>
</tr>
<tr>
<td>Peak Performance</td>
<td>313 Gflops/sec</td>
</tr>
<tr>
<td>Memory Size</td>
<td>64 GB per node</td>
</tr>
<tr>
<td>Memory Bandwidth (Peak)</td>
<td>102 GB/sec</td>
</tr>
<tr>
<td>Interconnect Injection Bandwidth (Peak)</td>
<td>9.6 GB/sec per direction</td>
</tr>
</tbody>
</table>

*Each core module includes 1 256-bit wide FP unit and 2 integer units. This is often advertised as 2 cores, leading to a 32 core node.

Blue Waters contains 22,640 XE6 compute nodes
Cray XK7 and a Path to the Future

Blue Waters contains 3,072 NVIDIA Kepler (GK110) GPUs

<table>
<thead>
<tr>
<th>XK7 Compute Node Characteristics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Host Processor</td>
<td>AMD Series 6200 (Interlagos)</td>
</tr>
<tr>
<td>Host Processor Performance</td>
<td>156.8 Gflops</td>
</tr>
<tr>
<td>Kepler Peak (DP floating point)</td>
<td>&gt; 1.3 Tflops</td>
</tr>
<tr>
<td>Host Memory</td>
<td>32GB 51 GB/sec</td>
</tr>
<tr>
<td>Kepler Memory</td>
<td>6GB GDDR5 capacity &gt; 180 GB/sec</td>
</tr>
</tbody>
</table>
Summary

• Outstanding Computing System
  • The largest installation of Cray’s most advanced technology
  • Extreme-scale Lustre file system with advances in reliability/maintainability
  • Extreme-scale archive with advanced RAIT capability
• Most balanced system in the open community
  • Blue Waters is capable of addressing science problems that are memory, storage, compute, or network intensive or any combination.
  • Use of innovative technologies provides a path to future systems
• Illinois/NCSA is a leader in developing and deploying these technologies as well as contributing to community efforts.