Big Red II Overview

Robert Henschel – henschel@iu.edu
Manager, Scientific Applications and Performance Tuning

Research Technologies, UITS
Indiana University

October 2nd, 2013
Contents

• Agenda and logistics
• Organizational overview
• Big Red II overview
• Big Red II details
• Application performance
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00 - 8:15</td>
<td>Coffee, Q&amp;A</td>
<td></td>
</tr>
<tr>
<td>8:15 - 9:00</td>
<td>Big Red II Overview</td>
<td>Robert Henschel (IU)</td>
</tr>
<tr>
<td>9:00 - 10:00</td>
<td>Compiling and Running Applications</td>
<td>Dave Strenski (Cray)</td>
</tr>
<tr>
<td>10:00 - 10:15</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>10:15 - 11:15</td>
<td>Data Capacitor 2 and RT Storage</td>
<td>Stephen Simms (IU)</td>
</tr>
<tr>
<td>11:15 - 12:15</td>
<td>Maintaining Applications on GPU Platforms</td>
<td>John Stone (UIUC)</td>
</tr>
<tr>
<td>12:15 - 13:15</td>
<td>Lunch</td>
<td></td>
</tr>
<tr>
<td>13:15 - 14:15</td>
<td>Introduction to CUDA</td>
<td>Jonathan Bentz (NVIDIA)</td>
</tr>
<tr>
<td>14:15 - 14:30</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>14:30 - 15:30</td>
<td>Introduction to OpenACC</td>
<td>Jonathan Bentz (NVIDIA)</td>
</tr>
<tr>
<td>15:30 - 16:30</td>
<td>Introducing ParaView and VTK</td>
<td>William Sherman (IU)</td>
</tr>
<tr>
<td>16:30 - 17:00</td>
<td>Q&amp;A</td>
<td></td>
</tr>
</tbody>
</table>
Thank you!

• Thank you for joining us today, let’s not waste your time!
  – 128 participants, 31 departments

• Thanks to NVIDIA and Cray for sponsoring the event.

• Thanks to everyone who helped with the workshop.

• There will be hands-on sessions, and people to help.
First Hands-On Session ;-)  

• Who has attended the Pre-Launch workshop?
• Who has used Big Red, Quarry or Mason?
• Who knows what the Data Capacitor is?
• Who has used the SDA?
• Who has run jobs on Big Red II?
• Who has a GPU code ready to run?
• Who would like to develop code for Big Red II?
OVPIT/ UITS / RT

• Office of the Vice President of Information Technology
  http://ovpit.iu.edu/

• University Information Technology Services
  http://uits.iu.edu/

• Research Technologies
  http://rt.uits.iu.edu/
Contents

• Agenda and logistics
• Organizational overview
• Big Red II overview
• Big Red II details
• Application performance
Big Red II Overview

- **Cray XE6 / XK7 hybrid supercomputer**

<table>
<thead>
<tr>
<th>System Size – 1020 compute nodes</th>
<th>Interconnect: Gemini</th>
</tr>
</thead>
<tbody>
<tr>
<td>• 344 XE6 CPU compute nodes</td>
<td>Topology: 11 x 6 x 8 3D Torus</td>
</tr>
<tr>
<td>2 x 2.5 GHz 16-core AMD</td>
<td>Peak Performance: 1.0003 PFLOPS</td>
</tr>
<tr>
<td>32 cores per node</td>
<td>Total XE6 Cores: 11,008</td>
</tr>
<tr>
<td>64 GB system memory</td>
<td>Total XK7 Cores: 10,816</td>
</tr>
<tr>
<td>• 676 XK7 GPU compute nodes</td>
<td>Total x86-64 Cores: 21,824</td>
</tr>
<tr>
<td>1 x 2.3 GHz 16-core AMD</td>
<td></td>
</tr>
<tr>
<td>16 cores per node</td>
<td></td>
</tr>
<tr>
<td>32 GB system memory</td>
<td></td>
</tr>
<tr>
<td>One NVIDIA K20 GPU (Kepler)</td>
<td></td>
</tr>
<tr>
<td>5 GB video memory</td>
<td></td>
</tr>
</tbody>
</table>
Big Red II Details

- 344 CPU nodes (11,008 cores)
  - 8 BR nodes are now one node!
  - BRII CPU nodes are equivalent to more than 2,700 BR nodes.

- 676 GPU nodes (10,816 cores, 676 NVIDIA GPUs)
  - CUDA, OpenACC and OpenCL support

- Cray/GNU/PGI/Intel compilers
- High Throughput Computing (HTC) tools to pack nodes
**ESM – Extreme Scalability Mode**
- No compromise *scalability*
- Low-Noise Kernel for scalability
- Native Comm. & Optimized MPI
- Application-specific performance tuning and scaling

**CCM – Cluster Compatibility Mode**
- No compromise *compatibility*
- Fully standard x86/Linux
- Standardized Communication Layer
- Out-of-the-box ISV Installation
- ISV applications simply install and run

CLE run mode is set by the user on a job-by-job basis to provide full flexibility
Similar Systems

- 13 PFLOPS Blue Waters at the National Center for Supercomputing Applications
- 20 PFLOPS Titan at Oak Ridge National Laboratory
Data Capacitor II Overview

- 2 SFA12K40 with 10 84-slot chassis each
- 1680 total 3 TB SATA drives
  - 3.5 PB usable capacity

- Lustre file system

- Bandwidth
  - Up to 5 GB/s via Ethernet (Mason, Quarry)
  - >40 GB/s via InfiniBand (BRII)
Why should you care?

• Big Red II has replaced Big Red
  – Quarry and Mason will continue

• Big Red II is standard x86 architecture
  – Will run standard Linux applications (Matlab)

• Big Red II has GPU acceleration
  – GPU-enabled libraries and applications
Why should you care? (cont’d)

• 32 cores per CPU node with 64 GB of RAM

• 16 cores per GPU node and one NVIDIA K20 GPU

• Data Capacitor II has 3.5 PB of storage capacity, with more than 40 GB/s of bandwidth to BRII
Contents

• Agenda and logistics
• Organizational overview
• Big Red II overview
• Big Red II details
• Application performance
Access

• Create an account:
  – [https://itaccounts.iu.edu/](https://itaccounts.iu.edu/)
  – Graduate students with hourly positions may not see BRII listed, please contact us for an account!

• ssh to [bigred2.uits.iu.edu](bigred2.uits.iu.edu)
  – IU network ID and pass phrase
  – Read the message of the day

• [https://cybergateway.uits.iu.edu](https://cybergateway.uits.iu.edu)
Environment

• Home directory:
  – /N/u/<username>/BigRed2

• Scratch space for temporary files:
  – /N/dc2/scratch/<username>

• Bash shell
  – Changes via hps-admin@iu.edu
Environment (cont’d)

• module command for adding software to your environment
  – module avail
  – module list
  – module load <key>
  – module unload <key>

• Permanently configure your environment via the .modules file
Environment (cont’d)

• Try to be explicit when loading modules, to make sure you get what you think you are getting!
  – module load namd
  – module load namd/gnu/gpu/2.9
  – module load namd/gnu/mpi/2.9
Batch System

• Shared access via a batch system
  – PBS, like Quarry and Mason
• Show all available queues, and their definitions:
  – `qstat -Q`
  – `qstat -Qf`
• List jobs: (man qstat / man showq)
  – `qstat -a -u <userID>`
  – `showq -i`
  – `checkjob -v <jobID>`
Interactive Jobs

• Interactive jobs:
  – `qstat -I`
  – Followed by `aprun` or `ccmrun/ccmlogin`
Applications on BRII

• On the UITS cyber gateway

• module avail
  – And then “manually” parsing the output.
Single Job Mode

- Only one job can run on a compute node at a time.
- If you have a serial application, you need to put more than one on a node, to make efficient use of the system.
- [http://kb.iu.edu/data/bdka.html](http://kb.iu.edu/data/bdka.html) – On Big Red II at IU, how do I use PCP to bundle multiple serial jobs to run them in parallel?
- BigJob is also available, via python/2.7.5
Contents

• Agenda and logistics
• Organizational overview
• Big Red II overview
• Big Red II details
• Application performance
Schrödinger / Jaguar

Single Node Performance

Prof. M. Baik – IUB Chemistry
BLASTn

Single Node Performance

Runtime in Minutes

Cores

Prof. M Radovich – IUPUI School of Medicine
NAMD

Time to Simulate one Nanosecond

Runtime in Days

Nodes (Cores)

1 (32)

2 (64)

4 (128)

Quarry
BR II (CPU)
BR II (GPU)

NAMD APOA1 Benchmark – University of Illinois
GROMACS

Gromacs using 1 Mio. Atoms

Days per Nanosecond

Nodes (Cores)

Prof. P. Ortoleva – IUB Chemistry
High Performance Linpack (Top500)

Linpack Performance (CPU)

- % Peak
- TFLOPS
- Nodes (Cores)

- 1 (32)
- 2 (64)
- 4 (128)
- 8 (256)
- 16 (512)
- 32 (1024)
- 64 (2048)
- 128 (4096)

- Efficiency
- Performance

HPL Benchmark – University of Tennessee
Prof. C. Horowitz – IUB Physics

Runtime per Time Step CPU

Runtime per Time Step GPU

IUMD
IUMD (cont’d)

- Sample experiment: 1,000,000 time steps
- Runtime
  - GPU
  - CPU
  - 1 node
  - 40 vs. 580 h
- Decrease runtime to less than 8 hours
  - 8 vs. 128 nodes
- Overall usage
  - 43 vs. 606 node hours
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Nodes</td>
<td>proc/</td>
<td>thrd/</td>
<td>Total</td>
<td>GPUs</td>
<td>2k</td>
<td>4k</td>
<td>8k</td>
<td>10k</td>
<td>18k</td>
<td>27k</td>
<td>54k</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>0.045896</td>
<td>0.182727</td>
<td>0.729334</td>
<td>1.137855</td>
<td>3.679151</td>
<td>8.773932</td>
<td>33.113196</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>16</td>
<td>16</td>
<td>0</td>
<td>0.023123</td>
<td>0.092122</td>
<td>0.365574</td>
<td>0.568937</td>
<td>1.842539</td>
<td>4.396315</td>
<td>16.581987</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>16</td>
<td>0</td>
<td>0.023186</td>
<td>0.091823</td>
<td>0.365442</td>
<td>0.569187</td>
<td>1.964987</td>
<td>4.139379</td>
<td>16.597283</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>16</td>
<td>0</td>
<td>0.023160</td>
<td>0.091625</td>
<td>0.365315</td>
<td>0.569853</td>
<td>1.842069</td>
<td>4.184544</td>
<td>16.574141</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>32</td>
<td>32</td>
<td>0</td>
<td>0.011822</td>
<td>0.046241</td>
<td>0.183337</td>
<td>0.286768</td>
<td>0.924731</td>
<td>2.207135</td>
<td>8.295711</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>2</td>
<td>16</td>
<td>32</td>
<td>0</td>
<td>0.011830</td>
<td>0.046270</td>
<td>0.183541</td>
<td>0.287081</td>
<td>0.982747</td>
<td>2.075265</td>
<td>8.297569</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>4</td>
<td>8</td>
<td>32</td>
<td>0</td>
<td>0.011709</td>
<td>0.046098</td>
<td>0.183198</td>
<td>0.285327</td>
<td>0.983472</td>
<td>2.215595</td>
<td>8.289325</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>8</td>
<td>4</td>
<td>32</td>
<td>0</td>
<td>0.011733</td>
<td>0.046111</td>
<td>0.183382</td>
<td>0.285431</td>
<td>0.923707</td>
<td>2.074159</td>
<td>8.295398</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.001537</td>
<td>0.004111</td>
<td>0.013736</td>
<td>0.020748</td>
<td>0.064023</td>
<td>0.141498</td>
<td>0.557932</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>1</td>
<td>32</td>
<td>64</td>
<td>0</td>
<td>0.006128</td>
<td>0.023609</td>
<td>0.092408</td>
<td>0.144740</td>
<td>0.466411</td>
<td>1.046428</td>
<td>4.151496</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>2</td>
<td>16</td>
<td>64</td>
<td>0</td>
<td>0.006031</td>
<td>0.023278</td>
<td>0.092026</td>
<td>0.143824</td>
<td>0.462768</td>
<td>1.038550</td>
<td>4.147990</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>64</td>
<td>0</td>
<td>0.006036</td>
<td>0.023286</td>
<td>0.092111</td>
<td>0.144023</td>
<td>0.462772</td>
<td>1.038941</td>
<td>4.147361</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>8</td>
<td>4</td>
<td>64</td>
<td>0</td>
<td>0.005949</td>
<td>0.023354</td>
<td>0.091967</td>
<td>0.144039</td>
<td>0.462065</td>
<td>1.038406</td>
<td>4.148921</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0.001063</td>
<td>0.002470</td>
<td>0.007368</td>
<td>0.010933</td>
<td>0.032796</td>
<td>0.071724</td>
<td>0.280199</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td>1</td>
<td>32</td>
<td>128</td>
<td>0</td>
<td>0.003255</td>
<td>0.011923</td>
<td>0.046414</td>
<td>0.072679</td>
<td>0.233137</td>
<td>0.523592</td>
<td>2.077741</td>
</tr>
<tr>
<td>19</td>
<td>4</td>
<td>2</td>
<td>16</td>
<td>128</td>
<td>0</td>
<td>0.003184</td>
<td>0.011934</td>
<td>0.046509</td>
<td>0.072695</td>
<td>0.233136</td>
<td>0.522597</td>
<td>2.078094</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>128</td>
<td>0</td>
<td>0.003144</td>
<td>0.011781</td>
<td>0.046215</td>
<td>0.072241</td>
<td>0.232211</td>
<td>0.520004</td>
<td>2.075744</td>
</tr>
<tr>
<td>21</td>
<td>4</td>
<td>8</td>
<td>4</td>
<td>128</td>
<td>0</td>
<td>0.003134</td>
<td>0.011797</td>
<td>0.046256</td>
<td>0.072319</td>
<td>0.232157</td>
<td>0.520388</td>
<td>2.076702</td>
</tr>
<tr>
<td>22</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0.000917</td>
<td>0.001632</td>
<td>0.004268</td>
<td>0.006118</td>
<td>0.017334</td>
<td>0.037145</td>
<td>0.142272</td>
</tr>
<tr>
<td>23</td>
<td>8</td>
<td>1</td>
<td>32</td>
<td>256</td>
<td>0</td>
<td>0.001753</td>
<td>0.006185</td>
<td>0.023723</td>
<td>0.036952</td>
<td>0.117868</td>
<td>0.263093</td>
<td>1.044285</td>
</tr>
<tr>
<td>24</td>
<td>8</td>
<td>2</td>
<td>16</td>
<td>256</td>
<td>0</td>
<td>0.001675</td>
<td>0.006087</td>
<td>0.023390</td>
<td>0.036717</td>
<td>0.117014</td>
<td>0.261629</td>
<td>1.040367</td>
</tr>
<tr>
<td>25</td>
<td>8</td>
<td>4</td>
<td>8</td>
<td>256</td>
<td>0</td>
<td>0.001655</td>
<td>0.006109</td>
<td>0.023516</td>
<td>0.036674</td>
<td>0.116959</td>
<td>0.261822</td>
<td>1.043256</td>
</tr>
<tr>
<td>26</td>
<td>8</td>
<td>8</td>
<td>4</td>
<td>256</td>
<td>0</td>
<td>0.001647</td>
<td>0.006035</td>
<td>0.023373</td>
<td>0.036363</td>
<td>0.117569</td>
<td>0.260787</td>
<td>1.041389</td>
</tr>
<tr>
<td>27</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>0.000875</td>
<td>0.001117</td>
<td>0.002559</td>
<td>0.003516</td>
<td>0.009211</td>
<td>0.019288</td>
<td>0.072216</td>
</tr>
<tr>
<td>28</td>
<td>16</td>
<td>1</td>
<td>32</td>
<td>512</td>
<td>0</td>
<td>0.000961</td>
<td>0.003330</td>
<td>0.012028</td>
<td>0.018520</td>
<td>0.059188</td>
<td>0.131969</td>
<td>0.522726</td>
</tr>
</tbody>
</table>
More Information


- Research Tech Expo, next Tuesday and Thursday.

- We are happy to present to your research group, or department!
- Telephone and in-person meetings are no problem!
Even More Information

- System support: hps-admin@iu.edu
- Application support: sciapt@iu.edu
- Data Capacitor: hpfs-admin@iu.edu
- SDA: store-admin@iu.edu
- Visualizations: vishelp@iu.edu
- General help: researchtech@iu.edu
Thank You!

Questions?