ORNL’s “Titan” System

- Upgrade of Jaguar from Cray XT5 to XK6
- Cray Linux Environment operating system
- Gemini interconnect
  - 3-D Torus
  - Globally addressable memory
  - Advanced synchronization features
- AMD Opteron 6274 processors (Interlagos)
- New accelerated node design using NVIDIA multi-core accelerators
  - 2011: 960 NVIDIA x2090 “Fermi” GPUs
  - 2012: 14,592 NVIDIA “Kepler” GPUs
- 20+ PFlops peak system performance
- 600 TB DDR3 mem. + 88 TB GDDR5 mem

Titan Specs

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Compute Nodes</td>
<td>18,688</td>
</tr>
<tr>
<td>Login &amp; I/O Nodes</td>
<td>512</td>
</tr>
<tr>
<td>Memory per node</td>
<td>32 GB + 6 GB</td>
</tr>
<tr>
<td># of Fermi chips (2012)</td>
<td>960</td>
</tr>
<tr>
<td># of NVIDIA “Kepler” (2013)</td>
<td>14,592</td>
</tr>
<tr>
<td>Total System Memory</td>
<td>688 TB</td>
</tr>
<tr>
<td>Total System Peak Performance</td>
<td>20+ Petaflops</td>
</tr>
<tr>
<td>Liquid cooling at the cabinet level</td>
<td>Cray EcoPHLex</td>
</tr>
</tbody>
</table>
April 15, 2012 Top 500 Submission: Jaguar

**HPL Run Statistics**

System Idle 2,935 kW

Run Start 4/15/2012: 7:17:42 AM

Run End 4/16/2012: 7:51:06 AM

Duration 24.6 hours

Sample Size 279 measurements, on 5 minute intervals, from three sources

Max kW 5,275 kW

Mean kW 5,142 kW

kW-hours 126,281
Assessment of April 15, 2012 HPL Submission

Energy Efficient HPC System Workload Power Measurement Methodology

– Aspect 1: **Level 2, Level 3 is available**
  - Eaton IQ Analyzer sampling at up to 8 times per second
  - Total energy is available directly from the unit
  - Sample is the instantaneous measurement at that time
  - The typical measurement interval for historical purposes is a 5-minute sample. Shorter measurement period of 1-minute samples are frequently used for analysis of consumption during full machine runs (HPL and others).

<table>
<thead>
<tr>
<th>Report and Analyze</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspect 1: requirements of measured values for ac measurement</td>
<td>1 instantaneous power measurement per second</td>
<td>1 instantaneous power measurement per second</td>
<td>continuously integrated total energy</td>
</tr>
<tr>
<td>Aspect 1: requirements of reported values for submission</td>
<td>one average power covering at least 20% of the run</td>
<td>time series of equally-spaced averaged power values</td>
<td>time series of equally spaced total energy values</td>
</tr>
<tr>
<td>Aspect 2: machine fraction</td>
<td>at least 1/64 of the machine or 1 kW</td>
<td>at least ? of the machine or 10 kW</td>
<td>whole machine</td>
</tr>
<tr>
<td>Aspect 3:</td>
<td>subsystems included</td>
<td>subsystems included</td>
<td>subsystems included</td>
</tr>
<tr>
<td>Aspect 3:</td>
<td>Point in power distribution where measurement is taken</td>
<td>Point in power distribution where measurement is taken</td>
<td>Point in power distribution where measurement is taken</td>
</tr>
<tr>
<td>required analyzed values for submission</td>
<td>core phase average power</td>
<td>core phase average power and whole application average power</td>
<td>core phase average power and whole application average power</td>
</tr>
</tbody>
</table>

– Aspect 2: **Level 3**
  - All 200 cabinets were measured from three main switchboards.

– Aspect 3: **Level 3**
  - Power metering is at the three switchboards, not at the individual devices.
  - More accurate assessment of total consumption, including line losses in the 200+48 480V branch circuits.
  - All 48 Liebert XDPs included in the measurement (Cray EcoPhlex closed loop cooling system)
  - Not included in the measurement:
    - Chilled water cost
    - External parallel file system
    - External login nodes
Metering Capabilities for HPC Systems at ORNL

• Every electrical service delivery system (switchboard, panel, PDU, RDU) is metered throughout the computer facility as part of the cost recovery mechanism for the facility.

• Metering at existing switchboards using Eaton IQ Analyzer (installed on main switchboards in 2009)

• IQ A Metered/Monitored Parameters
  – rms sensing.
  – Phase neutral, and ground currents.
  – Power: real, reactive, apparent (system and per phase).
  – Frequency.
  – Power factor: apparent and displacement (system and per phase).
  – Energy and demand (forward, reverse, net) real, reactive apparent at four different utility rates.
  – Individual current and voltage harmonics: magnitude, phase angle.
  – % THD: current and voltage.
  – Waveform capture.
  – ANSI C12.20 Class 0.5% revenue metering accuracy, ANSI C12.16, IEC687 Class 0.5%.

• New Capabilities (2012)
  – XFMR_S36/MSB14 (3.0MVA transformer/switchboard pair) are metered by Schneider Electric CM4000 PowerLogic Circuit Monitor
    • Highly accurate power quality monitor for critical energy systems. Substantially higher performance/capability than original equipment.
    • Provides mechanism for measuring and comparing features against original Eaton baseline, especially potentially troubling aspects including harmonics. Adds very accurate voltage transient and flicker analysis features.
  – Individual cabinet meters on two of the Cray XK6 cabinets. One meter on a non-accelerated XK6 cabinet, and a second meter on a NVIDIA Kepler-accelerated cabinet.
Recommendations

• Define the boundary of the system for measurement:
  – Disks?
  – Storage Area Network?
  – Cooling, Pumps, Chillers?
  – Transformers, UPS, AC-DC conversion?

• Remember that power measurement is a tool, not an end in itself. We use this to help inform choices, not dictate decisions.