Where Does the Power Go and What to do About it?

James Hamilton

2008.12.02
Architect, Data Center Futures

e: JamesRH@microsoft.com
w: research.microsoft.com/~jamesrh
w: perspectives.mvdirona.com
Agenda

• Where does the Power go & What To do about it?
  – Power Distribution Systems & Optimizations
  – Critical Load Optimizations
    • Server Design & Utilization
  – Mechanical Systems & Optimizations
• Modular Systems & Summary
PUE & DCiE

• Measure of data center infrastructure efficiency
• Power Usage Effectiveness
  – PUE = (Total Facility Power)/(IT Equipment Power)
• Data Center Infrastructure Efficiency
  – DCiE = (IT Equipment Power)/(Total Facility Power) * 100%

http://www.thegreengrid.org/gg_content/TGG_Data_Center_Power_Efficiency_Metrics_PUE_and_DCiE.pdf
Where Does the Power Go?

• Assuming a pretty good data center with PUE ~1.7
  – Each watt to server loses ~0.7W to power distribution losses & cooling

• Power losses are easier to track than cooling:
  – Power transmission & switching losses: 8%
    • Detailed power distribution losses on next slide
  – Cooling losses remainder: 100-(59+8) => 33%

• Data center power consumption:
  – IT load (servers): 1/1.7=> 59%
  – Distribution Losses: 8%
  – Mechanical load(cooling): 33%
Power Distribution

8% distribution loss
0.97^3 * 0.94 * 0.99 = 92.2%

0.3% loss
99.7% efficient

6% loss
94% efficient, >97% available

0.3% loss
99.7% efficient

0.3% loss
99.7% efficient

2.5MW Generator
~180 Gallons/hour

IT LOAD

~1% loss in switch
Gear and conductors

208V

480V

13.2kv

115kv

13.2kv

13.2kv

UPS:
Rotary or Battery
Move Power Redundancy to Geo-Level

- Over 20% of entire DC costs is in power redundancy
  - Batteries to supply up to 15 min at some facilities
  - N+2 generation (2.5MW) at over $2M each
- Instead use more, smaller, cheaper data centers
- Typical UPS in the 94% range
  - ~0.9MW wasted in 15MW facility (4,500 servers)
  - 97% available (0.45MW loss in 15MW)
Power Distribution Optimization

• Two additional conversions in server:
  – Power Supply: often <80% at typical load
  – Voltage Regulation Module: ~80% common
  – ~95% efficient available & affordable

• Rules to minimize power distribution losses:
  1. Avoid conversions (Less transformer steps & efficient or no UPS)
  2. Increase efficiency of conversions
  3. High voltage as close to load as possible
  4. Size voltage regulators (VRM/VRDs) to load & use efficient parts
  5. DC distribution potentially a small win (regulatory issues)

• Two interesting approaches:
  – 480VAC (or higher) to rack & 48VDC (or 12VDC) within
  – 480VAC to PDU and 277VAC to load
    • 1 leg of 480VAC 3-phase distribution
Cooperative Expendable Micro-Slice Servers

- CEMS: Cooperative Expendable Micro-Slice Servers
  - Correct system balance problem with less-capable CPU
    - Too many cores, running too fast, for memory, bus, disk, ...

- Joint project with Rackable Systems

<table>
<thead>
<tr>
<th>System-X</th>
<th>CEMS V3 (Athlon 4850e)</th>
<th>CEMS V2 (Athlon 3400e)</th>
<th>CEMS V1 (Athlon 2000+)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU load%</td>
<td>56%</td>
<td>57%</td>
<td>57%</td>
</tr>
<tr>
<td>RPS</td>
<td>95.92</td>
<td>75.26</td>
<td>54.27</td>
</tr>
<tr>
<td>Price</td>
<td>$2,371</td>
<td>$500</td>
<td>$685</td>
</tr>
<tr>
<td>Power</td>
<td>295</td>
<td>60</td>
<td>39</td>
</tr>
<tr>
<td>RPS/Price</td>
<td>0.04</td>
<td>0.15</td>
<td>0.08</td>
</tr>
<tr>
<td>RPS/Joule</td>
<td>0.32515254</td>
<td>1.254333333</td>
<td>1.391538462</td>
</tr>
<tr>
<td>RPS/RU</td>
<td>1918.4</td>
<td>18062.4</td>
<td>13024.8</td>
</tr>
</tbody>
</table>

- CEMS V2 Comparison:
  - Work Done/$: +372%
  - Work Done/Joule +385%
  - Work Done/RU: +941%

Update: New H/W SKU likely will improve numbers by factor of 2. CEMS still a win.
Conventional Mechanical Design

- Cooling Tower
- Heat Exchanger (Water-Side Economizer)
- A/C Condenser
- Primary Pump
- A/C Evaporator
- A/C Compressor
- Secondary Pump
- CWS Pump

 Blow down & Evaporative Loss for 15MW facility: ~360,000 gal/day

- Server fans 6 to 9W each
- Air-side Economization
- Hot/Cold Mix
- Overall Mechanical Losses ~33%
- Computer Room Air Handler
- Air Impeller

- Cold fans
- Hot
- Diluted Hot/Cold Mix
- Cold

2008/12/2  http://perspectives.mvdirona.com
Mechanical Optimization

• Simple rules to minimize cooling costs:
  1. Raise data center temperatures
  2. Tight control of airflow with short paths
  3. Cooling towers rather than A/C
  4. Air side economization (open the window)
  5. Low grade, waste heat energy reclamation

• Best current designs bring water close to load but not direct water
  – Lower heat densities could be 100% air cooled
  – density trends argue against

• Common mechanical designs: 33% lost in cooling
• PUE 1.1 to 1.2 implies cooling overhead in 5% to 15% range
• PUE under 1.0 within reach with some innovation
  – Waste heat reclamation in excess of power distribution & cooling overhead (~30% effective reclamation sufficient for sub 1.0)
Agenda

• Where does the Power go & What To do about it?
  – Power Distribution Systems & Optimizations
  – Critical Load Optimizations
    • Server Design & Utilization
  – Mechanical Systems & Optimizations

• Modular Systems & Summary
Modular Data Center

• Just add power, chilled water, & network
• Drivers of move to modular
  – Faster pace of infrastructure innovation
    • Power & mechanical innovation to 3 year cycles
  – Efficient scale-down
    • Driven by latency & jurisdictional restrictions
  – Service-free, fail-in-place model
    • 20-50% of system outages cause by admin error
    • Recycle as a unit
  – Incremental data center growth
    • Transfer fixed to variable cost
• Microsoft Chicago deployment: entire first floor with ½ MW containers
Summary

• Some inefficient facilities as low as 2.0 to 3.0 PUE
• PUE in ~1.2 attainable with care using state of the art techniques
• PUE in ~1.1 range attainable
  – aggressive air side economization
  – higher temperature
  – high voltage distribution to racks
• PUE under 1.0 within reach with some innovation
  – Waste heat reclamation in excess of power distribution & cooling overhead
    (~30% effective reclamation sufficient for sub 1.0)
• Most important gains not measured by PUE
  – Increased server efficiency with sub-component power management
  – Much higher server utilization
• Work done/$ & work done/W are what really matters (S/W issues dominate)
More Information

• These slides
  – <JRH>

• Designing & Deploying Internet-Scale Services

• Architecture for Modular Data Centers
  • http://mvdirona.com/jrh/talksAndPapers/JamesRH_CIDR.doc

• Increasing DC Efficiency by 4x
  • http://mvdirona.com/jrh/talksAndPapers/JamesRH_PowerSavings20080604.ppt

• JamesRH Blog
  – http://perspectives.mvdirona.com

• Email
  – JamesRH@microsoft.com