Reliable Power and Thermal Management in The Data Center

Dева Bodas
Intel Corporation
April 19, 2004
Deva.Bodas@Intel.com
Agenda

- Data center manageability challenges & trends
- Current state of power & thermal management
- New approaches
- Integrated management architecture
- Summary
Data Center Overview

- Exponential data growth
- Server Growth
- Higher performance driving increased power
- Shrinking form factors driving increased power density
- Increasing complexity
- Increasing TCO
- Flat IT Spending
Complexity & Mission Criticality Drive TCO

Growth, complexity, focus to reduce TCO driving new architectures

Source: Intel, Gartner studies (www.intel.com/tco)
Data Center Trends

Virtualization
- Dynamic partitioning
- Logical compute, I/O & storage
- Dynamic resource allocation

Virtualization
- Dynamic partitioning
- Logical compute, I/O & storage
- Dynamic resource allocation

Automation
- Self-configuring
- Self-optimizing
- Self-healing
- Auto Power & Thermal mgt
- Performance optimization
- Auto recovery

Modularity
- Open standards & protocols
- Independent scaling

Focus of this presentation
HDDC 4-19-2004

SOFTWARE
Manages Complexity

HARDWARE
Delivers Flexibility And Value

Independent scaling

App logic, server, & network

Focus of this presentation
HDDC 4-19-2004
Agenda

- Data center manageability challenges & trends
- **Current state of power & thermal management**
- New approaches
- Integrated management architecture
- Summary
Why do you want new room, you are not using all the space you already have?

Power & cooling limits number of systems in a rack
Power & Cooling Infrastructure

2KW 42 U Rack

Empty Space

300 Watts/Sq Ft? You are not fully using the 100 Watt/Sq Ft we already have.

Using nameplate, provisioning for the worst case
Rate of Power Consumption

Rack is consuming same power even if utilization varies between 0% & 100%!

0 KW 1KW 2KW

Power Meter

2KW 42 U Rack

Empty Space

Intel Labs

HDDC 4-19-2004
Fractured Power & Thermal Management

- Fractured monitoring
  - Separate monitoring systems for various elements
  - Near zero integration
- Lack of knowledge on power consumption
  - Who is consuming power, when & how much?
  - No system power monitoring
- Limited controls for management
  - Systems shut down at time of crisis

Lack of integrated power & thermal manageability
How Do We Address These Problems?

- Improve energy efficiency in existing racks
  - DBS & EPTM
- Introduce smarter scheme for allocating power & cooling
  - ACPC & EPTM
- Establish integrated architecture
  - EPTM
Today’s Approach?

- Method to allocate power & cooling
  - Nameplate
  - De-rating
  - Sample measurement

- Near future: heat load guidance
  - ASHRAE TC9.9

- Components at the same power/performance level irrespective of workload

No scientific approach for power budget & usage
Agenda

- Data center manageability challenges & trends
- Current state of power & thermal management
- New approaches
- Integrated management architecture
- Summary
Wasted Energy

Power consumption is high while application needs are low!

Wasted Energy
Demand Based Switching

Principle
- Optimize processor speed based upon utilization

Ingredients
- Support for P states
- Power-performance policy
- OS directed changes

Example with hypothetical numbers

Change performance to match utilization, Power will follow
Benefit of Demand Based Switching

Estimated power profile for the “hypothetical example”

System energy consumption drops up to 30%

Assumptions:
All components except processor consume fixed amount of power
Wasted Space & Infrastructure

Problem statements:
- How to put more systems in a rack, reliably?
- How to take advantage of reduced power with DBS?

Power Meter

2KW

42 U Rack

Empty Space

350W

350W

350W

550W

550W
Automatic Control of Power Consumption

**Principle**
- Set power threshold
- System level power manager to ensure power consumption will not exceed a set threshold

**Ingredients**
- DBS
- Power supply monitoring
- ACPC policy

Adjustable threshold provides multiple operating power states to systems
ACPC: Benefits

- Better space & infrastructure efficiency
  - Removes nameplate dependency
  - Takes advantage of reduced power consumption with DBS
- Don’t have to shut down system while in crisis
- Opportunity for smaller power supply
  - Lower system cost
  - Better power efficiency
    - Without ACPC: 500 watts system running at 200 watts @ 50-55% efficiency
    - With ACPC: (a) use 250 watts power supply (save ~$55), (b) 200 watts @ 75% efficiency

ACPC increases rack space utilization 2 to 3x
Automation & Integration

- Automate setting of threshold
- Dynamic setting of threshold
  - Power & thermal capacity
- Policy based setting upon integrated power & thermal monitored events
Enterprise Power & Thermal Manager

- Rack power & cooling capability
- (Hotspots) Thermocouples
- Status of Cooling systems
- Status of Utility Power
- Status of Backup Power
- Administrative Policy

Enterprise level integrated power and thermal management

EPTM
Enterprise Power and Thermal Manager

Systems with ACPC

Intel Labs
EPTM: Principle of Operation

Available Power
Cooling capacity
Hot spots
System utilization

Policy based dynamic allocation of power & cooling

Set thresholds for systems at time t
Controls for power & cooling systems

Workload Predictor
Cluster Manager
Load Balancer

Policy based management maximizes use of space, power & cooling
Agenda

- Data center manageability challenges & trends
- Current state of power & thermal management
- New approaches
- Integrated management architecture
- Summary
Integrated Management Architecture

Architecture that provide integration and promotes automation
Agenda

- Data center manageability challenges & trends
- Current state of power & thermal management
- New approaches
- Integrated management architecture
- Summary
What Can You Do Now?

- Look for future announcement on server support for DBS & ACPC
- Don’t use nameplate to allocate power & cooling
  - Use ASHRAE TC9.9 guidelines for defining heat/power loads ([http://www.ashrae.org](http://www.ashrae.org))
  - Use dynamic, policy based allocation in future
- Join industry efforts
  - Common industry interfaces
  - Integration of manageability solutions
Summary

- Power & thermal manageability essential to reduce TCO & improve reliability
- DBS, ACPC & EPTM provide automated & scientific approaches for allocating of power & cooling
- Coordinated industry efforts needed to realize integrated management architecture
Backup
ACPC Operation

1. User selects a power level

2. ACPC Programs Thresholds
   - \( P_{\text{HIGH}} = 400 \)
   - \( P_{\text{LOW}} = 350 \)
   - \( P_{\text{LIMIT}} = 500 \)

3. Power Monitor generates interrupt when power consumption exceeds a threshold

4. ACPC manager forces change of P state

5. Lowering power states reduces system power consumption

Power Supply With Power Monitoring

PTARGET

watts

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>450</td>
<td>400</td>
<td>350</td>
</tr>
<tr>
<td>300</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ACPC Manager

Intel Labs

HDDC 4-19-2004
ACPC: Components

- User Interface
- OS
- ACPC Policy

ACPC (Software) Interface

- SM Bus
- Power Monitoring Controller
- System Power Supply

- +12V
- +5V
- +3.3V
- -12V

Intel Labs

HDDC 4-19-2004
## Top IT Problems

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Asset Management</strong></td>
<td>Asset ID, active location info, sys profile info persistent across installation of new OS images</td>
</tr>
<tr>
<td><strong>Interoperable management model &amp; interfaces</strong></td>
<td>Consistent management info model, Common API, &amp; command line interface across all server states: Pre-boot, OS functional, OS not-functional, etc.</td>
</tr>
<tr>
<td><strong>Remote in-band &amp; Out of Band (OOB) management</strong></td>
<td>Power on/off, BIOS install, OS boot/reboot, Set diagnostic levels, Run diagnostics – CPU, Chipset, memory, IO, storage</td>
</tr>
<tr>
<td><strong>Security</strong></td>
<td>Access control, error/virus detection, isolation, &amp; inoculation</td>
</tr>
<tr>
<td><strong>Power &amp; thermal management</strong></td>
<td>Policy-based power allocation/de-allocation, usage monitoring, integrated power &amp; thermal management standards and tools,</td>
</tr>
<tr>
<td><strong>Simplify data center management</strong></td>
<td>More intelligent and integrated tool, scripts that scale, &amp; work across multiple platforms</td>
</tr>
<tr>
<td><strong>Solution scalability</strong></td>
<td>Increased automated management (server provisioning, auto fault detection, prediction, &amp; correction)</td>
</tr>
</tbody>
</table>