Intel® Intelligent Power Node Manager

A dynamic approach to managing power in the data center.

...the need for more computing resources and data handling relentlessly marches on— at an astonishing rate.

David Jenkins
Server Technology
Marketing Manager,
Server Platforms Group

Introduction

One of the greatest challenges for data center operators today is the increasing cost of power and cooling as a portion of the total cost of operation. Power and cooling costs in data centers have risen sharply over the past several years—as much as 400 percent in the span of a decade.[1] In many data centers, electricity now represents as much as half of total operating expenses.[2] Of equal concern is capacity. In 2007, IDC found that 40 percent of data center owners said they would exceed power capacity within 12 to 24 months without expansion.[3] Three years later, IDC stated that data center power, cooling, and space problems will continue and even accelerate beyond those earlier projections.[4] Another 39 percent of data center owners say they will exceed cooling capacity in 12 to 24 months.[5] Paradoxically, many data centers lack precise data on their power and cooling usage and the finer controls to make the most of power and cooling resources.

Meanwhile, the need for more computing resources and data handling relentlessly marches on—at an astonishing rate. IDC expects the amount of digital information produced by the world will increase fivefold from 2009 to 2012.[6]

Many data center owners have a very practical reason for fearing they won't be able to keep up with this data overload: cost. Data center expansion is expensive. A 2007 IDC Datacenter Trends Survey found that data center construction costs are projected to exceed $1,000 per square foot or $40,000 per rack. Faced with these issues, organizations have three basic choices:
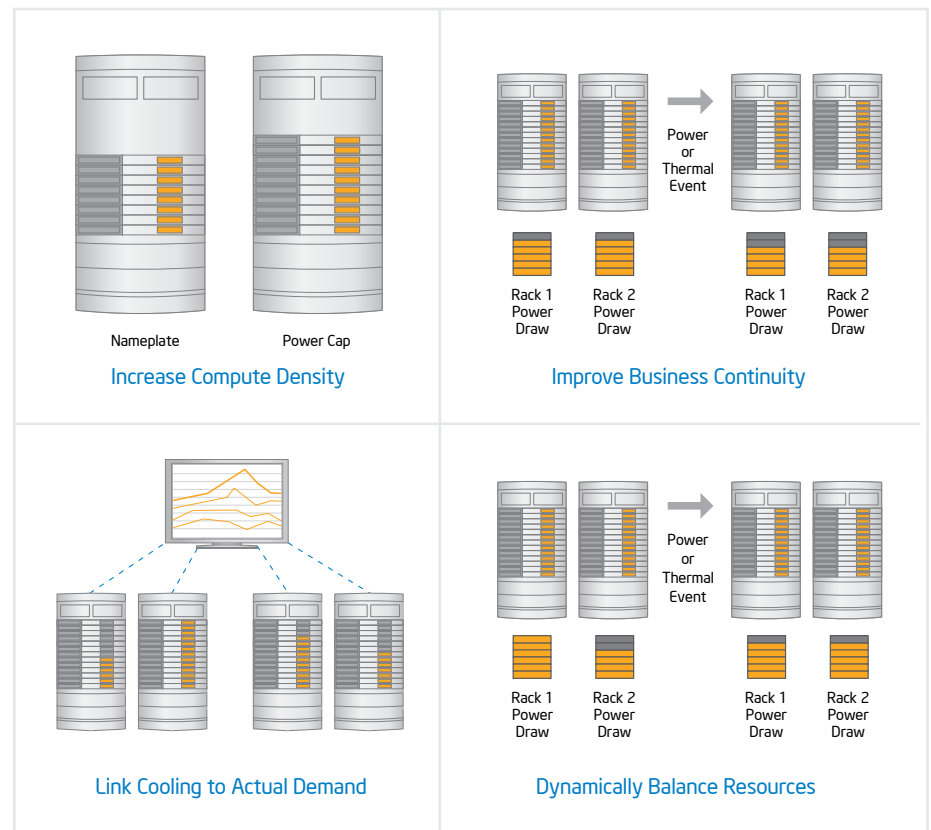
- Build new data centers despite the expense,
- Expand power and cooling capacity of existing data centers, or
- Improve usage of existing capacity through better power management.

Table of Contents

The first and second choices require major capital expenditure. The third offers a way to squeeze more performance out of existing resources. This merits a closer look. In fact, recent developments in server power capping solutions make this choice particularly compelling.

Server power capping solutions enable IT managers to set a power limit (cap) in watts for each server. Server power capping is rapidly gaining momentum in the marketplace for its usefulness and effectiveness in dealing with a number of common data center challenges. Coupled with better power monitoring and reporting of power consumption under load, server power capping enables IT managers to more precisely utilize existing power and cooling resources.

Increase Compute Density

Improve Business Continuity

Link Cooling to Actual Demand

Dynamically Balance Resources

Server power capping can be used to:

- **Increase rack density.** By maintaining server budget policies set according to real workload by a data center management system, power capping enables stranded (unused) power to be rescued, allowing more servers to be added to the rack.

- **Maximize work during power and thermal spikes.** By dynamically capping power to shed load, data center managers can allow critical jobs to continue running during power and thermal events.

- **Reduce power consumption.** Having actual server power reported on instrumented systems enables cooling capacity to be more closely linked to actual rack temperatures.

- **Facilitate power-based load balancing.** Power consumption is becoming part of the load balancing equation in virtual environments. Power capping enables data centers to dynamically manage power- and thermal-constrained systems and racks.
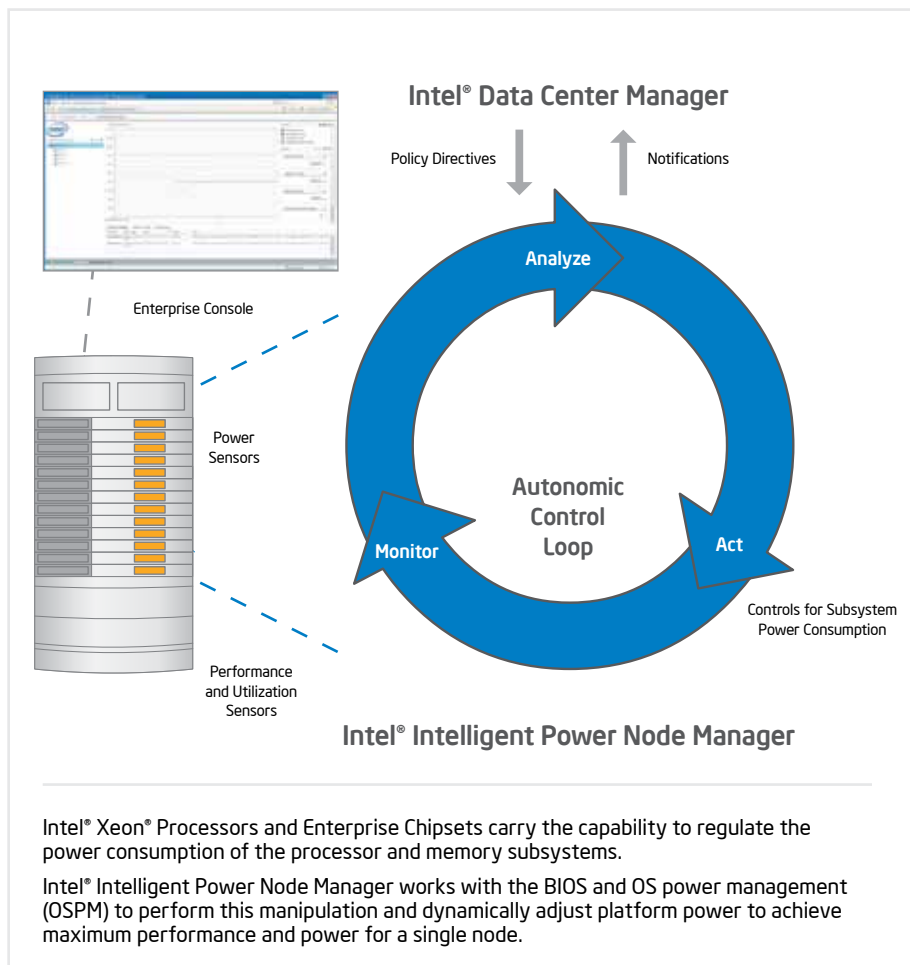
Introducing Intel® Intelligent Power Node Manager

Data center designers have often had to rely on inadequate power usage data in deploying servers and racks. Lacking actual power usage information for each server based on its intended load, they stay on the conservative side and populate server racks based on available server power information, using the declared power rating from the server nameplate. This nearly always leads to an overstatement of a server's actual power need, resulting in underutilization of the budgeted rack power. Likewise, relying on overstated power demands often leads to providing more cooling than is necessary for the actual loads. In fact, data center designers often use overstated power demand and heat generation figures instead of actual temperatures. The price they pay for overestimation is additional costs for unnecessary cooling.

To enable data center owners to better utilize rack power and optimize server density by better controlling server performance and power consumption, Intel has developed an innovative solution known as Intel® Intelligent Power Node Manager. Used with data center management tools, Intel Intelligent Power Node Manager allows data center managers to set a power budget for a server, enabling up to 40 percent denser deployments.[7] It's a key ingredient to dynamic power management. In fact, when paired with Intel® Data Center Manager (Intel® DCM)—a software technology that provides power and thermal monitoring and management for servers, racks, aisles, and other groups of servers—the combined solution can help optimize entire data centers.

Intel Intelligent Power Node Manager is an out-of-band (OOB) power management policy engine available with the Intel® Xeon® processor E3, 5600, and E7 series families. It enables regulation of power consumption (power capping) through modulation of the power of the processor and memory subsystems. Intel Intelligent Power Node Manager works with the BIOS and operating system power management (OSPM) to regulate and dynamically adjust platform power to achieve maximum performance within the defined power envelope for a single node.

Intel® Data Center Manager

Policy Directives ↓ ↑ Notifications

Analyze

Autonomic Control Loop

Monitor **Act**

Enterprise Console

Power Sensors

Performance and Utilization Sensors

Controls for Subsystem Power Consumption

Intel® Intelligent Power Node Manager

Intel® Xeon® Processors and Enterprise Chipsets carry the capability to regulate the power consumption of the processor and memory subsystems.

Intel® Intelligent Power Node Manager works with the BIOS and OS power management (OSPM) to perform this manipulation and dynamically adjust platform power to achieve maximum performance and power for a single node.

Intel Intelligent Power Node Manager features include:

- **Dynamic power monitoring.** Intel Intelligent Power Node Manager measures actual power consumption of a server platform within acceptable error margin of +/-10 percent.[8] Gathering information from the Power Management Bus (PMBus) Interface, Intel Intelligent Power Node Manager provides real-time power consumption data singly or as a time series, and reports through Intelligent Platform Management Interface (IPMI).

- **Platform power capping.** Intel Intelligent Power Node Manager makes it possible to set platform power to a targeted power budget while maintaining maximum performance for the given power level. Intel Intelligent Power Node Manager receives power policy from an external management console through IPMI and maintains power at the targeted level by dynamically adjusting processor and memory power.

- **Power threshold alerting.** Intel Intelligent Power Node Manager enables you to monitor platform power against the targeted power budget. When the target power budget cannot be maintained, Intel Intelligent Power Node Manager sends out alerts to the management console.

Taking Full Advantage of Intel's Latest Microarchitecture

It's especially significant that Intel Intelligent Power Node Manager is now available in all Intel® Xeon® processor families. Based on Intel's latest 32nm microarchitecture, this processor series helps lower energy costs with automated energy efficiency features that deliver significant improvements over the first Intel® quad-core server processors—30% less power than previous generation servers.[9]

Intel Intelligent Power Node Manager enables data center managers to set a power cap for the energy-efficiency features of Intel® Xeon® processor 5600 series-based platforms to work within. These features include:

- **Intel® Intelligent Power Technology.** This technology minimizes power consumption when server components are not fully utilized. To do this, it employs two Intel® microarchitecture (Nehalem) features:

  – Automated low-power states that automatically put processor and memory into the lowest available power states meeting the requirements of the current workload. The result is that system power consumption is based on real-time load.

  – Integrated power gates allow individual idle cores to be reduced to near-zero power independent of other operating cores. This cuts idle power consumption to 10 watts, versus 16 or 50 watts in earlier generations of Intel quad-core processors.[10] It also reduces server idle power consumption by up to 50 percent versus the previous generation of two-socket server processors.[11]

- **Intel® Turbo Boost Technology.** This technology (in combination with Intel Intelligent Power Technology), delivers performance on demand by allowing processors to operate above the rated frequency to speed specific workloads and drop back down to reduce power consumption during low utilization periods.

Beyond the processor, Intel Intelligent Power Node Manager provides dynamic power management capabilities for the chipset and other subsystems, such as the Intel® QuickPath Interconnect (Intel® QPI), PCI Express* (PCIe*) system, memory, fan, and instrumented power supplies. For instance, Intel Intelligent Power Node Manager helps tap specific Intel® 5600 Chipset and memory controller features. These include:

- **Memory power management.** This feature automatically places DIMMs into a lower power state when not being used.[12] DIMMs are also automatically idled when all processor cores in the system are idle.[13]

- **Chipset power management.** This feature puts Intel QuickPath Interconnect links and PCIe lanes in power reduction states when not active.[14] The chipset is also capable of placing PCIe cards in the lowest power state possible.[15]

Intel Intelligent Power Node Manager Requirements

Intel Intelligent Power Node Manager is a vendor-agnostic solution that exposes its capabilities through standards-based IPMI interface from supported Baseboard Management Controllers (BMC).

The requirements to deploy Intel Intelligent Power Node Manager include:

- Intel Xeon processor E3, 5600, and E7 series-based servers supporting Intel Intelligent Power Node Manager. These servers deliver the enhanced capabilities in the processor, I/O, memory, and fan subsystems that serve as the basis for understanding what is going on within the system and provide the required "control knobs."

- Operating system with power management capability.

- Management console or scripting enabled with Intel Intelligent Power Node Manager IPMI commands.

To enable companies to build management consoles that can use Intel Intelligent Power Node Manager to cap power consumption by individual servers or a group of servers, Intel offers a software development toolkit named Intel DCM. Intel DCM can be used to aggregate Intel Intelligent Power Node Manager-enabled systems (servers, racks, and groups of servers). This allows IT to monitor, cap, and manage each system's power to work inside of rack- or data center-level power and thermal constraints. Intel DCM can be attached to existing system management software as a console or as a Web service. Intel DCM provisions power consumption through communication with Intel Intelligent Power Node Manager. The middleware instructs Intel Intelligent Power Node Manager to set power limits for servers based on the level of activity. For example, using Intel DCM, an IT manager could set up a rack to dynamically cap power consumption on inactive servers while raising the power bar on active servers.

To learn more about Intel Intelligent Power Node Manager requirements, see: http://software.intel.com/sites/datacentermanager/requirements.php
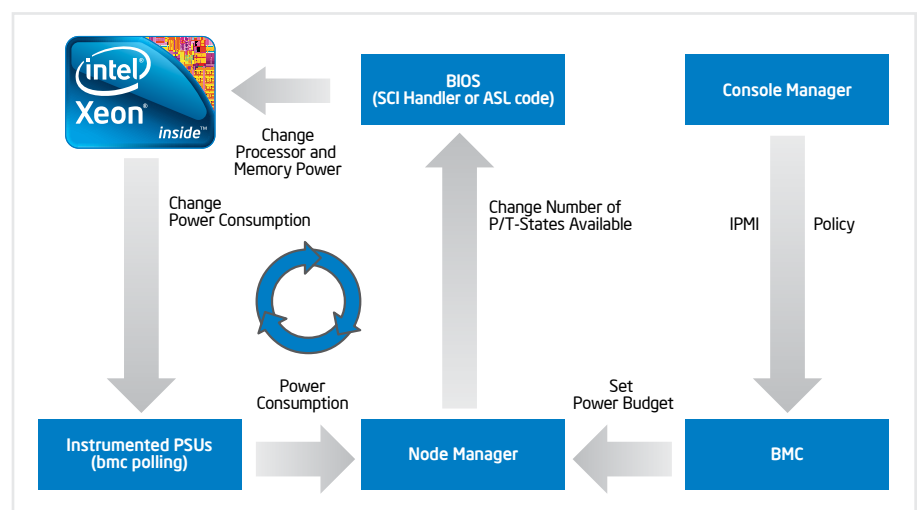
...Intel Intelligent Power Node Manager utilizes a closed-loop algorithm to determine the optimal processor performance and corresponding power consumption to meet the target power level.

How Intel Intelligent Power Node Manager Works

Once Intel Intelligent Power Node Manager is integrated with a management console, an IT manager can use the console to: 1) monitor the real workload requirements of an Intel Intelligent Power Node Manager-enabled server, and 2) set policy. For instance, an IT manager might evaluate a server for a few weeks and notice it has never gone above 200 watts. In such a circumstance, the IT manager could safely lower the budget from 500 watts to 300 watts without restricting the performance capability of the workload. By doing so, the IT manager will have freed up 200 watts. By freeing up 200 watts each in several servers in a rack, the data center would be able to deploy more servers into that rack to add more compute perfor-mance—all while remaining within the rack's power budget.

In a similar way, Intel Intelligent Power Node Manager could be used to strategi-cally throttle workloads to reduce cooling needs in a thermal event or to keep a server within its power limits.

For example, to protect against a workload spike, a data center manager could send commands via the management console to set a power cap. The console communi-cates the new power budget policies via the baseboard management controller, or BMC, to the Intel Intelligent Power Node Manager, which is continuously monitoring real-time server power consumption. The current power level is then compared to policy. When the current workload requires more power than the cap, Intel Intelligent Power Node Manager utilizes a closed loop algorithm to determine the optimal processor performance and corresponding power consumption to meet the target power level. This feedback loop ensures that server power consumption remains continuously controlled and predictable, and allows the server to perform as much work as possible while the policy is in place. If the policy can't ultimately be met through adjustments, Intel Intelligent Power Node Manager reports an exception. Once actual power falls below the policy budget, the server returns to its normal state.

Demonstrated Savings

In 2008 Intel's Digital Enterprise Group partnered with Baidu.com (the largest search company in China) on a proof-of-concept (POC) project using Intel Intelligent Power Node Manager to dynamically optimize server performance and power consumption to maximize the server density of a rack under a simulated Baidu production environment.[16] The results demonstrated savings of up to 40 watts per system without performance impact when an optimal power management policy is applied. At the rack level, an increase of up to 20 percent additional capacity could be achieved within the same rack-level power envelope when an aggregated optimal power management policy is applied. Compared with today's data center operation at Baidu, Intel Intelligent Power Node Manager could provide an improvement of approximately 20 percent to 40 percent in rack density.

The POC showed that Intel Intelligent Power Node Manager reduced power consumption at node level by approximately 40 to 50 watts with negligible impact to Baidu workload performance. When the same concept is applied at rack level with a combination of Intel Intelligent Power Node Manager and Intel Data Center Manager working together, the rack-level capacity could increase approximately 20 percent within the same power envelope without performance impacts.

How to Put Intel Intelligent Power Node Manager to Work for You

The ever-increasing amounts of digital data inundating virtually every organization portend one thing: Data centers will continue to need to increase compute performance. Intel Intelligent Power Node Manager-enabled Intel Xeon processor 5600 series-based platforms provide an excellent way to squeeze additional compute performance out of existing power, cooling, and floor space resources. The instrumentation delivered by these platforms enables industry-leading power capping solutions that help deliver the most out of every dollar spent and kilowatt consumed in the data center, and extend the life of existing data centers. What's more, through power capping, these platforms help data center managers achieve a lower total cost of ownership for each server by allowing them to limit each server's power consumption over its service life—so each server consumes just the power it needs to do its share of the work. Multiplied by hundreds of servers, this can save significant amounts of money.

To start making better use of current power budgets, IT managers should learn more about Intel Intelligent Power Node Manager-enabled Intel Xeon processor 5600 series-based platforms through their server vendor or the additional resources listed here.

Compared with today's data center operation at Baidu, Intel Intelligent Power Node Manager could provide an improvement of approximately 20 percent to 40 percent in rack density.

Intel® Intelligent Power Node Manager

Additional Resources

For more information on the Intel Intelligent Power Node Manager and other power-saving technologies, see:

Power-capping animation
communities.intel.com/community/openportit/server/blog/2009/05/11/
intel-xeon-processor-5500-instrumentation-enables-power-capping

"A Dynamic Approach to Power Budgeting." *Intel® Software Insight Magazine,* August, 2008.
softwarecommunity.intel.com/articles/eng/3802.htm

"Baidu Proof of Concept White Paper"
communities.intel.com/docs/DOC-1492

Intel® Xeon® processor E3 family product brief
www.intel.com/Assets/PDF/prodbrief/325197.pdf

Intel® Xeon® processor 5600 series product brief
www.intel.com/Assets/en_US/PDF/prodbrief/323501.pdf

Intel® Xeon® processor E7 family product brief
www.intel.com/Assets/PDF/prodbrief/325213.pdf

For additional information go to:
www.intel.com/cloudcomputing

[1] "Dynamic Data Center Power Management: Trends, Issues, and Solutions." Filani, David; Jackson He, Sam Gao, et al. Intel® Technology Journal, Feb. 21, 2008, pg. 59.

[2] "Power Surge. The heat is rising—and costs, too—as tightly packed servers consume gobs of electricity," Darrell Dunn, InformationWeek, Feb. 27, 2006.

[3] "Datacenters heading for cash crunch," InfoWorld, March 26, 2008.

[4] http://www.gartner.com/it/page.jsp?id=1368614.

[5] "Datacenters heading for cash crunch," InfoWorld, March 26, 2008.

[6] "As the economy contracts, the digital universe expands," a multimedia white paper, IDC 2009. See: www.emc.com/collateral/demos/microsites/idc-digital-universe/iview.htm.

[7] Source: Dynamic Power Optimization for Higher Server Density Racks – A Baidu Case Study with Intel® Dynamic Power Technology (Intel, 2008).
See http://software.intel.com/sites/datacentermanager/intel_node_manager_v2e.pdf for more details

[8] Intel does not guarantee the power accuracy delivered by any server. Power accuracy will vary based on server designer and manufacturer design, implementation and component selection. Contact the server vendor of choice for details regarding power reading accuracy associated with a specific server SKU.

[9] Source: Fujitsu Performance measurements comparing Xeon L5650 vs X5570 SKUs using SPECint_rate_base2006.
See http://docs.ts.fujitsu.com/dl.aspx?id=0140b19d-56e3-4b24-a01e-26b8a80cfe53 and http://docs.ts.fujitsu.com/dl.aspx?id=4af74e10-24b1-4cf8-bb3b-9c4f5f177389.

[10] Depending on processor SKU.

[11] Intel internal measurements of 221 W at idle with Supermicro 2xE5450 (3.0 GHz 80 W) processors, 8x2 GB 667 MHz FBDIMMs, 1x700 W PSU, 1x320 GB SATA hard drive vs. 111 W at idle with Supermicro software development platform with 2xE5540 (2.53 GHz Intel microarchitecture (Nehalem) 80 W) processors, 6x2 GB DDR3-1066 RDIMMs, 1x800 W PSU, 1x150 GB 10 k SATA hard drive. Both systems were running Windows* 2008 with USB suspend select enabled and maximum power savings mode for PCIe link state power management. Measurements as of Feb. 2009.

[12] Using DIMM CKE (Clock Enable).

[13] Using DIMM self refresh.

[14] Using L0s and L1 states.

[15] Using cards enabled with ASPM (Active State Power Management).

[16] For more details, see the white paper "Dynamic Power Optimization for Higher Server Density Racks – A Baidu Case Study with Intel® Dynamic Power Technology" at: http://communities.intel.com/docs/DOC-1492.