Power Budgeting Protection
Mechanisms in the PowerEdge Server Portfolio

A Dell technical white paper detailing power budgeting protection mechanisms in Dell’s 12th generation servers, which ensure reliability and protect customer data
This document is for informational purposes only and may contain typographical errors and technical inaccuracies. The content is provided as is, without express or implied warranties of any kind.

© 2013 Dell Inc. All rights reserved. Dell and its affiliates cannot be responsible for errors or omissions in typography or photography. Dell, the Dell logo, and PowerEdge are trademarks of Dell Inc. Intel and Xeon are registered trademarks of Intel Corporation in the U.S. and other countries. IBM is a registered trademark of International Business Machines Corporation. Other trademarks and trade names may be used in this document to refer to either the entities claiming the marks and names or their products. Dell disclaims proprietary interest in the marks and names of others.

May 2013 | Version 1.0
Contents

Executive summary .................................................................................................................. 4
Introduction ............................................................................................................................ 4
Why is power budgeting necessary? ....................................................................................... 4
Configuration dynamic power range .................................................................................... 4
Power capacity configuration “right-sizing” ......................................................................... 6
  Rack and tower server capacity options ....................................................................... 6
  Blade enclosure capacity options ................................................................................ 6
  Capacity expanding features ....................................................................................... 7
Summary ................................................................................................................................. 7
What is power budgeting? ..................................................................................................... 7
Order advisement .................................................................................................................. 7
Proactive vs. reactive ............................................................................................................. 8
  Proactive controls ........................................................................................................ 8
  Reactive controls .......................................................................................................... 9
Power consumption limits .................................................................................................... 10
  Rack and tower server limits ...................................................................................... 10
  Blade server limits ....................................................................................................... 11
Conclusion ............................................................................................................................ 11
Additional information ......................................................................................................... 11
About the authors .................................................................................................................. 11

Figures

Figure 1. Example of server model dynamic power range ......................................................... 5
Figure 2. Dell’s latest rack-mount and tower server PSU options .............................................. 6
Figure 3. M1000e blade enclosure power capacity options .................................................... 7
Figure 4. Proactive PSU capacity check ................................................................................ 8
Figure 5. PowerEdge M1000e shared power infrastructure ................................................... 9
Figure 6. Extended Power Range ....................................................................................... 10
Power Budgeting Protection Mechanisms in the PowerEdge Server Portfolio

Executive summary

Dell’s latest generation of servers provide a broad portfolio of options that enable a vast array of customer solutions. Each customer application is unique, but minimizing power consumption remains a key focus for many customers. Dell’s “right-sized” power delivery options enable customers the flexibility to choose a configuration that meets their needs efficiently. Going beyond a hardware only offering, Dell’s order advisement tools enable customers to choose the right power delivery option for their needs. Additionally, Dell’s in-system power budgeting mechanisms ensure that customers can have confidence in the reliability of their systems and the safety of their data in any configuration.

Introduction

Power capacity sizing is an intrinsic aspect of designing a server solution to meet customer needs. Historically, server hardware was offered with only one power supply unit (PSU) configuration option, which would support the maximum potential consumption of all subsystems. While effective, this approach was very inefficient for the majority of workloads, translating to a greater operating cost. In more recent product generations, power capacity configuration offerings have expanded significantly to better serve customers’ specific needs efficiently, such as Dell’s High Efficiency Power Supply options. There are also an increasing number of power management features that have an impact on power capacity sizing, such as Dell’s Hot Spare PSU efficiency feature or various features of the M1000e blade server infrastructure. While these features offer customers excellent options to tailor solutions to their needs, they also make it complex to answer the basic question “Do I have enough capacity?” Over generations, Dell has developed a series of power budgeting mechanisms to answer this basic capacity question both at the point-of-sale and during operation. The latest Dell™ PowerEdge™ server portfolio extends power budgeting leadership with enhanced accuracy through comprehensive inventories, as well as advanced reactive power mechanisms to control excursions and protect data.

Why is power budgeting necessary?

There are two main factors that contribute to the need for power budgeting. First, for each server model, there is a widening gap between the configuration that consumes maximum power and the configuration that consumes minimum power. Second, as more power supply capacity and redundancy options are offered, there are an increased number of configurations in which server power consumption can exceed available PSU capacity.

Configuration dynamic power range

For any given server configuration, power consumption can vary dramatically based on the amount of “stress” or “utilization” applied to each subsystem (such as processor, memory, and storage), which are collectively referred to as a “workload.” The same hardware under a light workload can potentially consume hundreds of watts less than when it is under a maximum workload. The difference between a configuration’s maximum and minimum power consumption across all potential workloads is that configuration’s “dynamic power range.” Generally, the dynamic power range increases each generation as subsystems consume less under light loads and consume more when at their maximum workload.
This is particularly true in Dell PowerEdge 12\textsuperscript{th} generation servers that support the Intel\textsuperscript{®} Xeon\textsuperscript{®} processor E5 product family with Intel Turbo Boost Technology 2.0, which provides an \textit{increased capability} to use power headroom even in excess of the processor-rated upper limit for short durations.

As the configuration dynamic power range expands, it becomes increasingly complex for customers to predict worst-case consumption. An undersized power capacity configuration may initially seem stable; but later, perhaps years in the future, improved higher performance workloads can push the system harder, possibly even to the point of shutdown if initial power capacity budgeting was inadequate.

When considering power consumption, it is also important to keep in mind that each server model is highly configurable, and some of those options can have an impact on power consumption. Dell’s latest PowerEdge servers offer greater system component configurability options than ever before. This allows our customers an excellent opportunity to customize a solution for their needs, but it also means that a given server model has an \textit{even larger} dynamic power range to consider. This increased range can be particularly important to customers who deploy non-homogenous configurations, and even more so when components are added after the point-of-sale. We recommend that customers perform capacity-sizing analysis for each unique configuration each time changes are made to that configuration considering the factors mentioned in this paper.

\textbf{Figure 1. Example of server model dynamic power range}
Power capacity configuration “right-sizing”

In order to serve a more diverse spectrum of customer solutions, Dell continues to expand its portfolio of “right-sized” power capacity configuration options. “Right-sized” capacity configurations for both rack-mount and blade servers offer the maximum capacity and optional redundancy that customers need for their individual solutions, eliminating the power efficiency penalty of a “one size fits all” solution.

Rack and tower server capacity options

There are currently up to eight hot-pluggable common form factor PSU options for Dell’s latest rack-mount and tower servers, and many more when redundant and non-redundant options are considered.

Blade enclosure capacity options

Dell’s M1000e blade enclosures also offer a rich selection of power capacity options, including five redundancy configurations of the 2700W PSUs, and many more when considering legacy support for previously offered PSUs.
Power Budgeting Protection Mechanisms in the PowerEdge Server Portfolio

Figure 3. M1000e blade enclosure power capacity options

<table>
<thead>
<tr>
<th>Enclosure</th>
<th>2700W Capacity Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1000e</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Grid Redundant</td>
</tr>
<tr>
<td></td>
<td>PSU Redundant</td>
</tr>
<tr>
<td></td>
<td>Non-Redundant</td>
</tr>
<tr>
<td>3+3</td>
<td>3+1</td>
</tr>
<tr>
<td>3+0</td>
<td></td>
</tr>
<tr>
<td>4+1</td>
<td></td>
</tr>
<tr>
<td>5+1</td>
<td></td>
</tr>
</tbody>
</table>

Capacity expanding features

Dell also offers innovative capacity expanding features such as Extended Power Range in the latest generation PSUs, and Dynamic Power Supply Engagement (DPSE) in the M1000e blade enclosure. These features and options make Dell’s server portfolio increasingly versatile and flexible, but they can also make it more challenging for customers to determine which power capacity option is right for their needs.

Summary

The widening gap in the configuration dynamic power range, and the broadening selection of Dell configuration options and capacity expanding features provide excellent benefits, when power budgeting is planned appropriately. Dell supports power budget planning by offering purchasing tools to provide guidance on which option is best for each customer’s unique needs. Customers also benefit from protection mechanisms designed into the servers themselves to ensure power consumption remains within the available power capacity during the entire service life of the server.

What is power budgeting?

In order to account for the wide variety of possible configurations in a server model dynamic power range, and due to the need to ensure that power consumption is always less than power capacity, it is necessary to create a “power budget” for each server configuration. A “power budget” is an amount of power that a server configuration can be expected to consume. Dell’s latest servers implement a mixture of proactive and reactive algorithms to maintain power budgets for hardware protection, customer applications, and to prevent shutdowns. In all cases, power consumption limits are established to determine necessary actions.

Order advisement

Dell provides customers with several tools to understand the power capacity requirements of PowerEdge servers. Using these tools, a customer can model typical and maximum power consumption, and they can also receive component selection recommendations to optimize for parameters such as efficiency or headroom. For more details on these tools, check out Power Efficiency “How To” for the Dell PowerEdge Server Portfolio.
Proactive vs. reactive

Sources of power consumption in the server such as the CPU, memory, and GPGPUs (general-purpose graphics processing units) are capable of generating a significant increase in power within a very small time window. Some of these sources can be reactively throttled when a workload exceeds the limits of the PSU or shared power infrastructure, but current technology is limited in how quickly it can detect and respond to fast, transient events. As a result, Dell provides a mixture of proactive and reactive controls to provide holistic coverage and high reliability that configurations will not shut down under unanticipated high utilization workloads.

Proactive controls

Some power transients occur so quickly and with such magnitude that they cannot be controlled reactively with existing technologies. To address this, Dell’s latest servers perform a proactive mechanism where an inventory and budget calculation is performed during the system power-on process. In the case of rack-mount and tower servers, a check is performed during system boot based on the capacity and redundancy configuration of installed PSUs. If this check fails, the system will halt and log error messages to warn the user that the configuration could exceed the maximum PSU capacity, which could cause a server shutdown and potential data loss if a worst-case workload is applied.

Similarly for blade servers, in order to ensure the chassis has enough power available to support the active modules, the chassis may temporarily prevent power-on for new modules introduced to the chassis environment. Server modules have an automatic power-on retry mechanism that allows them to retry powering on, and the chassis will grant power if it is available at retry. This proactive chassis level power check is in place to ensure that PSUs continue to operate at their rated capacity in order to avoid an unwanted shutdown. Additionally, in order to share infrastructure power more effectively, the chassis management controller (CMC) allocates power based on blade server usage. This CMC allocation process is done in steps, with the blade server requesting additional power in predetermined intervals based on workload requirements. This algorithm allows chassis power to be available for other blade servers to power on or to increase their workloads.
PowerEdge servers use available system inventory information as well as lab characterized power information to calculate server power budgets. The Integrated Dell Remote Access Controller version 7 (iDRAC7) embedded systems management agent performs a system inventory to determine information about the installed components, such as CPU, memory, PCIe adapters, storage, and fans. The CPU maximum power consumption values take into account the Turbo Boost capabilities that are offered by Intel (see Configuration dynamic power range). The power budget, based on the system inventory, is then tuned based on extensive measurements captured and analyzed by Dell engineering development teams to generate a characterized power budget that is comprehensive for all workloads.

Based on this characterized power budget, Dell.com has ordering tools to help validate server configurations at the time of purchase. Additionally, the server power characterization process also gathers typical workload power measurements. These measurements form a basis of the guidance provided by the Energy Smart Solution Advisor (ESSA) tool, which calculates power consumption based on typical customer workloads (see Order advisement).

Reactive controls

The latest Dell PowerEdge servers use the Intel Node Manager technology as well as the capabilities of the Intel Xeon processor E5 product family in order to reactively limit power. Power consumption limits are established based on the output power capacity of the rack/tower PSU or blade shared power infrastructure (see Power consumption limits). The latest Dell PowerEdge servers also implement optimized power monitoring functionality so that real-time power consumption can be continuously compared to the established PSU output or CMC allocation limits. Based on these limits, the Intel Node Manager will reactively manage server power consumption using graceful software controls. These controls are efficient and use CPU and memory Running Average Power Limiting (RAPL) to allow for very granular control of power to maximize performance under a power limit. However, RAPL controls have limitations in their response times and to accommodate for faster possible transients, the Dell design implements fast-acting hardware mechanisms that make use of processor throttling capabilities to quickly control power. Even with these industry leading technologies, there are some devices that cannot be throttled and some power transients that are too fast to be addressed reactively. In these cases, Dell’s implementation of proactive controls fills the gap, ensuring a holistic, reliable solution.
Power consumption limits

Dell defines power consumption limits dynamically for a given configuration based on the output power capacity of the rack server PSU configuration or the available power from the blade shared-power infrastructure. The aforementioned proactive and reactive mechanisms use these limits to determine when actions such as warning messages or throttling are necessary to inform and protect the end user from unexpected shutdowns and potential data loss.

Rack and tower server limits

The latest Dell PowerEdge rack-mount and tower servers establish power budget limits based on the rating of the installed PSU configuration, and whether or not the previously mentioned Extended Power Range (EPR) feature is supported (see Capacity expanding features). EPR temporarily extends the output capability of the power supply using reactive controls, which allows a higher limit to be used for power budgeting purposes (see Figure 6).

![Figure 6. Extended Power Range](image)

The limits used for rack-mount and tower servers also take into account the customer’s selected redundancy setting for the PSU configuration. In non-redundant configurations, all available capacity is used for power budgeting limits, whereas in redundant configurations, the capacities of supplies that are reserved for redundant failover are not included. By taking into account the various capacity options and the Extended Power Range feature, Dell’s power budget limits for rack-mount servers afford customers flexibility, while maintaining protection.
Blade server limits

Blade server power budgeting limits are determined by the CMC power allocation mechanism. In the shared power infrastructure, the CMC is responsible for designating power for each module in the chassis. For servers, the embedded management agent (iDRAC7) communicates the server’s Configuration Dynamic Power Range to the CMC, and the CMC determines allocation. It is the responsibility of each blade server to ensure that appropriate power budgeting limits are set in order to ensure that the allocation granted by the CMC is maintained. These individual server power consumption limits are in place to ensure that shared power infrastructure PSUs continue to operate at their rated capacity.

Conclusion

As the number of server options and configurations continue to expand, it becomes increasingly important for the power budget of each customer’s unique configuration to be considered to ensure that customer workloads stay within safe bounds. Through the combination of order advisement with proactive and reactive controls, Dell’s comprehensive power budgeting solution provides guidance to customers at the point-of-sale, and continues to protect the customer once the server hardware is onsite and deployed. Using Dell’s power budget solution, customers can deploy servers with confidence, and the simple piece of mind that their hardware and data are safe, both now and in the future.

Additional information

For additional details and information, see the following documents and resources:

- Power Consumption Reduction: High Efficiency Power Supplies
- Power Consumption Reduction: Hot Spare
- Increasing Energy Efficiency through Modular Infrastructure
- Intel® Turbo Boost Technology
- Power and Cooling Innovations in Dell PowerEdge Servers
- Dell PowerEdge Overview on Power: Extended Power Range and Hot Spare Technologies
- Power to Dell PowerEdge M1000e Blade Server Enclosures
- Power Efficiency “How To” for the Dell PowerEdge Server Portfolio
- Data Center Power Management with Intel® Node Manager
- Dell.com/PowerAndCooling

About the authors

Kyle Cross is a Senior Engineer in the Dell Enterprise Solutions Group specializing in the design of Node Power features for use across the PowerEdge server portfolio. Kyle has worked at Dell for ten years, including previous positions in Advanced Engineering and Concept Development. He has a B.S. in Electrical Engineering from The University of Texas at Austin, and currently holds two patents.

Reema Aron is a system engineer in the Dell Enterprise Solutions Group specializing in Node Power features for the blade server portfolio. Reema has worked at Dell for three years, including a previous position in Modular Continued Server Development. Reema has a B.S. in Electrical Engineering from The University of Texas at Austin.