A Primer: Managing Power Consumption and Capping in Dell PowerEdge C6320 Servers

Version 1.0
February 2016

Balaji Rao, Principle Engineer, Platform Software
Pushkala Iyer, Principle Engineer, Platform Software

Dell | Enterprise Solutions Group

Table of contents

Executive summary 3
Introduction 4
Chassis View and Sled View of Power Consumption 4
Sled Power Consumption 4
Chassis View of Sled Power Consumption 4
Chassis and Sled Level Power capping 5
Chassis Level Power Capping 5
Sled Level Power Capping 9
Dell OEM IPMI Command Reference 12
Get Chassis Enclosure Power Capping – NetFn 0x30, Cmd 0xC5 12
Set Chassis Enclosure Power Capping – NetFn 0x30, Cmd 0x17 13
Get Sensor Info – NetFn 0x30, Cmd 0x16 14
Get Chassis Power Readings – NetFn 0x30, Cmd 0x2E 16
Additional Resources 17

February 2016 | Version 1.0

Executive summary

As deployments scale, the demand for power and cooling increase, driving the need for proactive management of power consumption.

The Dell PowerEdge C6320 is designed to handle the most demanding high-performance and scale-out workloads. By utilizing the PowerEdge server's integrated Dell Remote Access Controller (iDRAC), IT administrators can monitor power consumption and manage that consumption with power capping.

This document describes PowerEdge C6320 power monitoring and capping utilizing the iDRAC GUI and the open source ipmitool program.

Introduction

This Dell Technical White paper provides information on monitoring and managing the power consumption of the Dell PowerEdge C6320 server. The examples used throughout this white paper are based upon the use of iDRAC firmware Version 2.23.23.21 or greater. The paper also assumes that all processor sleds of the C6320 are provisioned with the iDRAC Enterprise License which is required for the power capping feature.

Chassis View and Sled View of Power Consumption

The PowerEdge C6320 is composed of a C6320 chassis and up to four C6320 server sleds along with storage devices, peripherals, fans, and power supplies. Chassis power consumption can be classified into two parts: sled power consumption and shared infrastructure power consumption. Sled power consumption plus the normalized shared infrastructure power consumption are thus depicted as a sled view of power consumption and the Chassis View of Power Consumption.

Sled Power Consumption

Sled Power Consumption is the actual power consumed by a given sled independent of the shared infrastructure. This view gives the actual power consumption of the sled as measured by the Intel NodeManager. Sled Power Consumption can be seen in the Power Monitoring page of the iDRAC graphical user interface (GUI), shown below in Figure 1 as the "Present Reading" value.

Chassis View of Sled Power Consumption

Chassis View of Sled Power Consumption is the Sled Power Consumption along with the *normalized* shared infrastructure power consumption. Normalization means the total shared infrastructure power divided by the number of sleds in the chassis. The "Chassis Controller View of Individual Sled Power" can be seen in the Power Monitoring page of the iDRAC GUI as shown below in Figure 1.

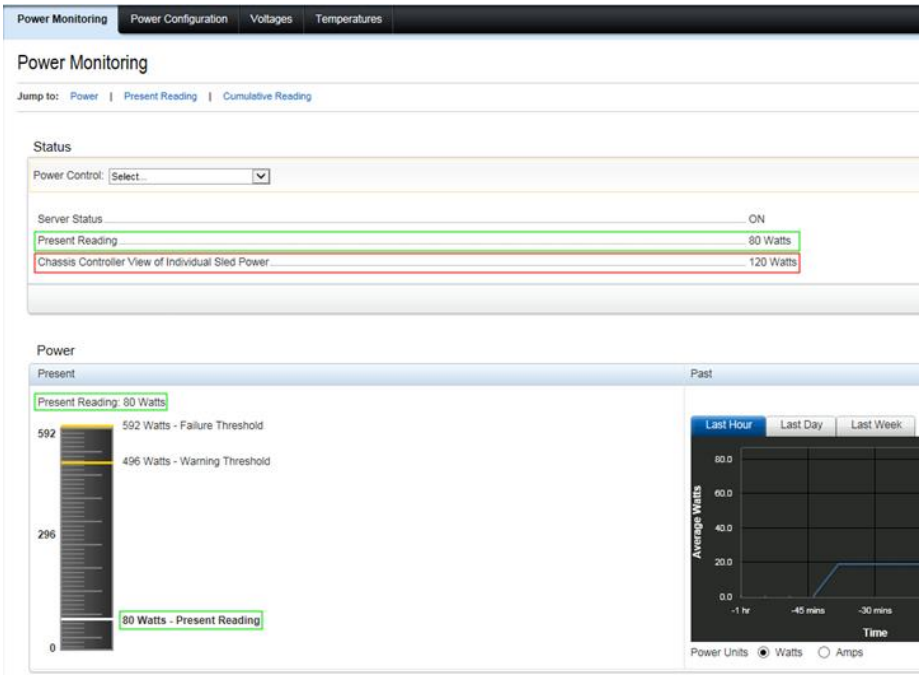


Figure 1 iDRAC GUI Power Monitoring page

Chassis and Sled Level Power capping

As a data center scales in size, power consumption and the related heat generation can become a concern. In extreme cases, over consumption of power or over heating can lead to unplanned interruption of workloads or equipment failure. To help prevent such issues, servers such as the PowerEdge C6320 provide power capping to limit power consumption.

The PowerEdge C6320 provides power capping through the iDRAC, utilizing the industry-standard IPMI protocol. The C6320 iDRAC supports two power capping methods: Chassis Level Power capping and Sled Level Power capping.

NOTE: Enabling Chassis Level and Sled Level power capping is strongly discouraged. Enabling both methods at the same time <u>may</u> result in inconsistency in power capping.

Chassis Level Power Capping

Chassis Level Power Capping is used to set a power consumption limit for the entire chassis. It is implemented through the "Chassis Power cap" Satellite Controller-Baseboard Management Controller (SC-BMC) command. The SC-BMC specification defines the interaction between the C6320 Chassis and the sleds housed within the chassis. Internally, the chassis translates this limit to individual power limits taking into account the number of powered-on sleds as well as the overall power loading.

The following examples use the open source "ipmitool" to illustrate setting Chassis Level Power Capping to 800 watts and to 500 watts, respectively. For additional details on ipmitool usage, refer to the "Dell OEM IPMI Command Reference" section.

Setting Chassis Power Cap to 800 Watts

1. Clear the previous Chassis Power Capping. Byte 1 of the *command data* indicates cap clearing

   ipmitool -I lanplus -H <iDRAC ip> -U <iDRAC admin username> -P <iDRAC admin user password> raw 0x30 0x17 *0x00 0x00 0x00 0x02 0x02 0x00 0x00 0x02 0x02*

2. Ensure Chassis Power Capping is cleared. Byte 1 of *response data* indicates that Chassis Power Capping is disabled

   ipmitool -I lanplus -H <idrac ip> -U <idrac admin username> -P <idrac admin user password> raw 0x30 0xC5
   0x00 0x00 0x00 0x02 0x00 0x00 0x00 0x02 0x00

3. Set Chassis Level Power Capping (here to 800W). Byte 4 of the *command data* indicates Chassis Power Capping over Sled Power capping

   ipmitool -I lanplus -H <iDRAC ip> -U <iDRAC admin username> -P <iDRAC admin user password> raw 0x30 0x17 *0x01 0x20 0x03 0x02 0x02 0x00 0x00 0x02 0x02*

4. Check Power Consumption using Get Sensor Info: Get Sensor Info shows the power supply units (PSU) power outputs (POUTs) not exceeding the capped value (in this case 256W and 219W)

   ipmitool -I lanplus -H <iDRAC ip> -U <iDRAC admin username> -P <iDRAC admin user password> raw 0x30 0x16

   01 0c 26 00 17 00 79 00 09 be 03 00 00 01 db 00

   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00

   0c c0 c1 c1 c1 db c8 90 56

5. Get Chassis Power Reading to read the Chassis Power Reading: this example shows that Chassis Total Power Consumption is 475 W (which approximately adds up to PSU POUTs)

   ipmitool -I lanplus -H <iDRAC ip> -U <iDRAC admin username> -P <iDRAC admin user password> raw 0x30 0x2e

   dc c8 90 56 03 db 01 9e 00 ff ff ff ff

   Figure 2 illustrates viewing the Power Configuration page within the iDRAC GUI. This shows that the Sled Power Limit is being managed by the Chassis (SC) and is verified by the Active Power Cap Policy showing DCMI.

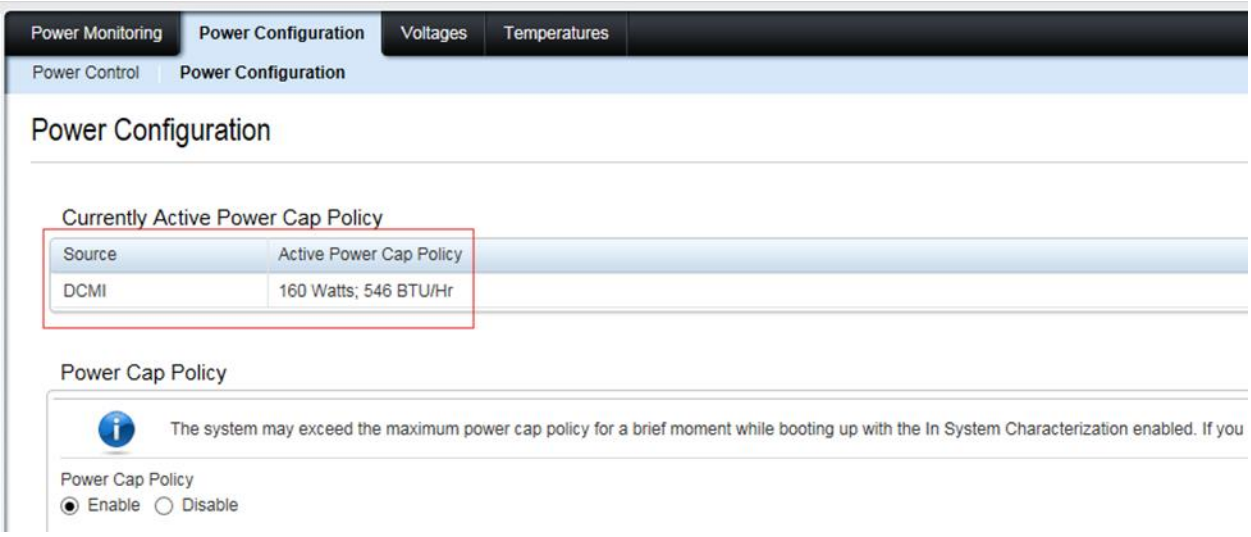


Figure 2 Viewing active Power Capping Policy via iDRAC GUI

Set Chassis Level Power Capping to 500 W

1. Clear the previous Chassis Power Capping. Byte 1 of the *command data* indicates cap clearing

   ipmitool -I lanplus -H <iDRAC ip> -U <iDRAC admin username> -P <iDRAC admin user password> raw 0x30 0x17 *0x00 0x00 0x00 0x02 0x02 0x00 0x00 0x02 0x02*

2. Ensure Chassis Power Capping is cleared. Byte 1 of *response data* indicates that Chassis Power Capping is disabled

   ipmitool -I lanplus -H <idrac ip> -U <idrac admin username> -P <idrac admin user password> raw 0x30 0xC5
   0x00 0x00 0x00 0x02 0x00 0x00 0x00 0x02 0x00

3. Set Chassis Level Power Capping (here to 500 W). Byte 4 of *command data* indicates Chassis Power Capping over Sled Power capping.

   ipmitool -I lanplus -H <idrac ip> -U <idrac admin username> -P <idrac admin user password> raw 0x30 0x17 *0x01 0xf4 0x01 0x02 0x02 0x00 0x00 0x02 0x02*

4. Get Sensor Info shows the PSU POUTs not exceeding the capped value (in this case 257W and 218W)

   ipmitool -I lanplus -H <iDRAC ip> -U <iDRAC admin username> -P <iDRAC admin user password> raw 0x30 0x16

   01 21 26 00 17 00 7b 00 09 be 03 00 01 01 da 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 0c c0 c1 c3 c1 be cc 90 56

5. Get Chassis Power Reading3 to read the Chassis Power Reading: Here it shows that Chassis Total Power Consumption is 471 W (which adds up to PSU POUTs approximately)

   ipmitool -I lanplus -H <idrac ip> -U <idrac admin username> -P <idrac admin user password> raw 0x30 0x2e

   bf cc 90 56 03 d7 01 9a 00 ff ff ff ff

Viewing the Power Configuration page within the iDRAC GUI in Figure 3 shows that the Sled Power Limit is being managed by the Chassis. This can be seen as the Active Power Cap Policy shows DCMI. From the previous example of 800 W, it can be seen that the sled power cap has reduced.

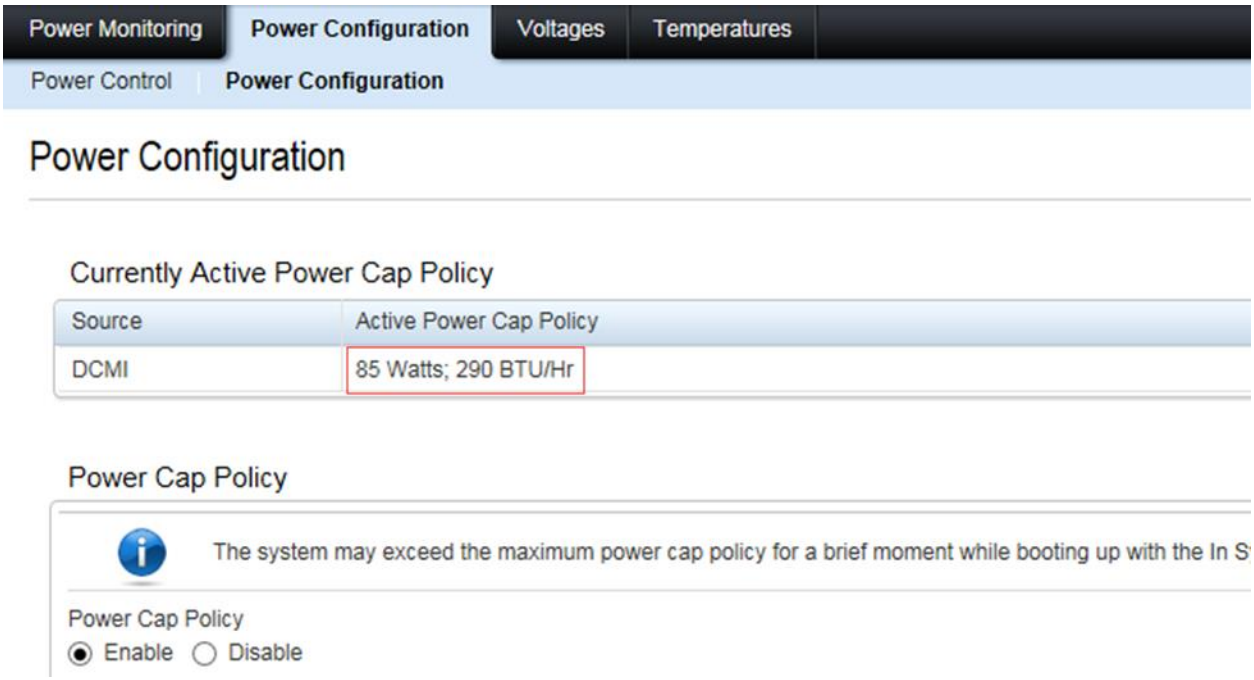


Figure 3 iDRAC GUI Power Capping Policy after cap reduction to 500W

Sled Level Power Capping

Using the Sled Level Power capping method, the administrator can set the power cap for the individual sleds. Since iDRAC firmware version 2.14.14.12, the power consumption shown in the iDRAC GUI and the RACADM command line interface provide the Chassis view of the sled power consumption which includes a portion of the shared infrastructure power consumption. In the following example ipmitool and the iDRAC GUI are used to illustrate setting Sled Level Power Capping. For additional details on ipmitool usage, refer to the "Dell OEM IPMI Command Reference" section.

Set Sled Level Power Capping to 280 W using iDRAC GUI

1. Clear the previous Chassis Power Capping; Byte 1 of the *command data* indicates cap clearing

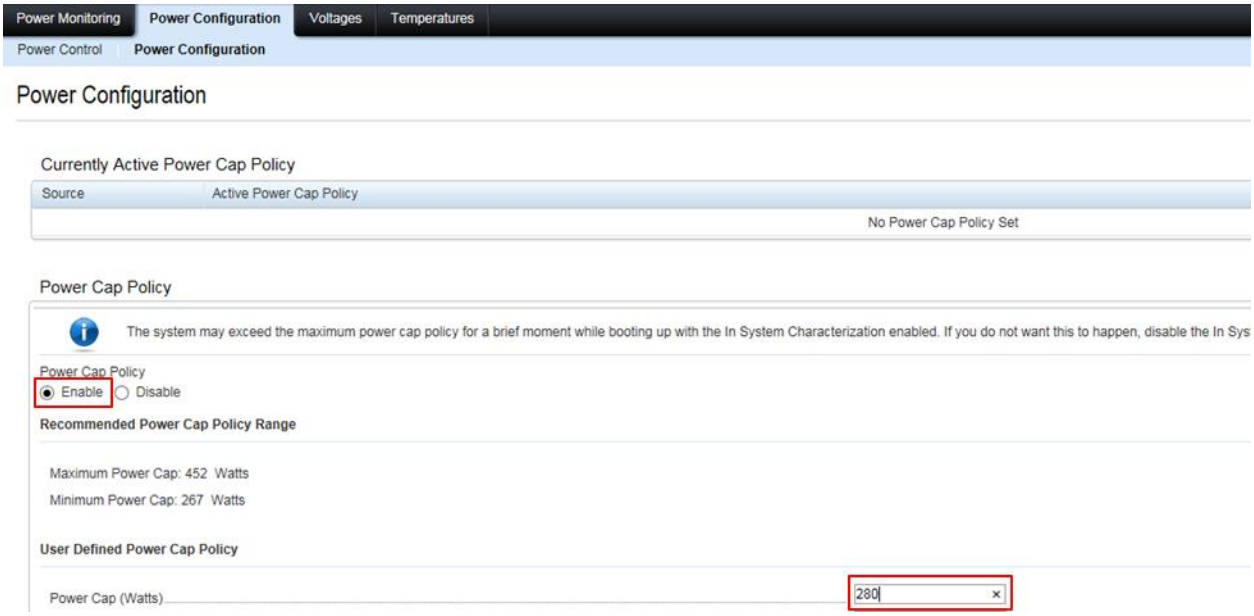
   ipmitool -I lanplus -H <iDRAC ip> -U <iDRAC admin username> -P <iDRAC admin user password> raw 0x30 0x17 *0x00 0x00 0x00 0x02 0x02 0x00 0x00 0x02 0x02*

2. Ensure that no Chassis Level Power Capping is set

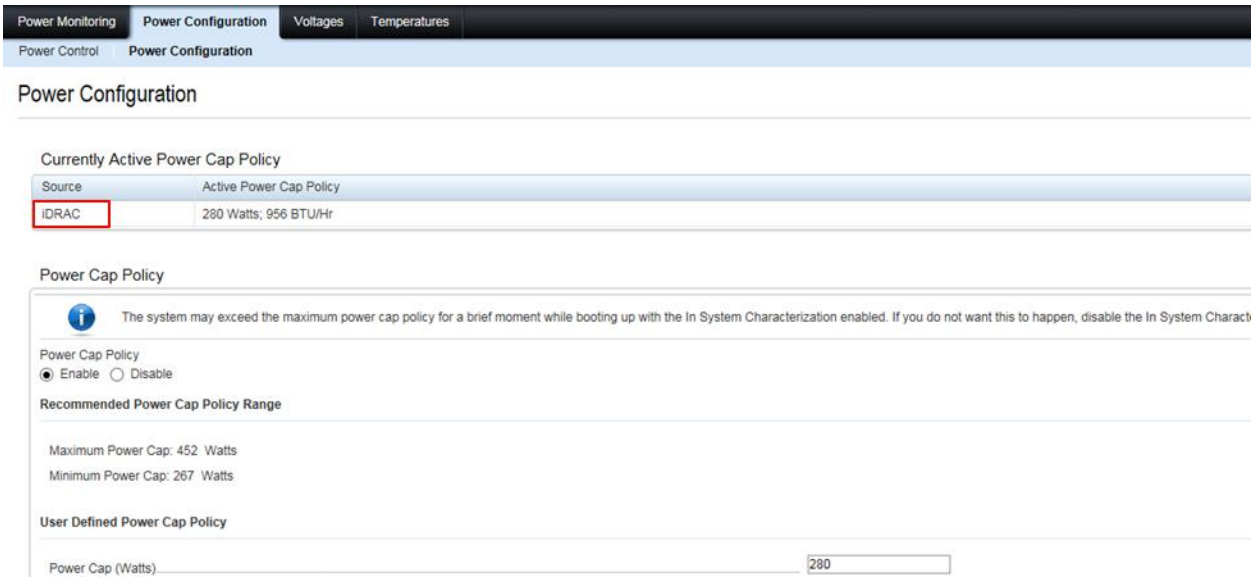
   ipmitool -I lanplus -H <iDRAC ip> -U <iDRAC admin username> -P <iDRAC admin user password> raw 0x30 0xC5. (Byte 1 of *Response Data* shows no Chassis Level Power Capping Set)

   00 00 00 02 02 00 00 02 02

3. Set the Sled Level Power Capping on individual sleds using the iDRAC GUI



4. View the Power Configuration page to verify that the Sled Power Limit is being managed by iDRAC. This is shown from the value of the Active Power Cap Policy set to "iDRAC".



Set Sled Level Power Capping to 300 W using DCMI

1. Clear the previous Chassis Power Capping; Byte 1 of the *command data* indicates cap clearing

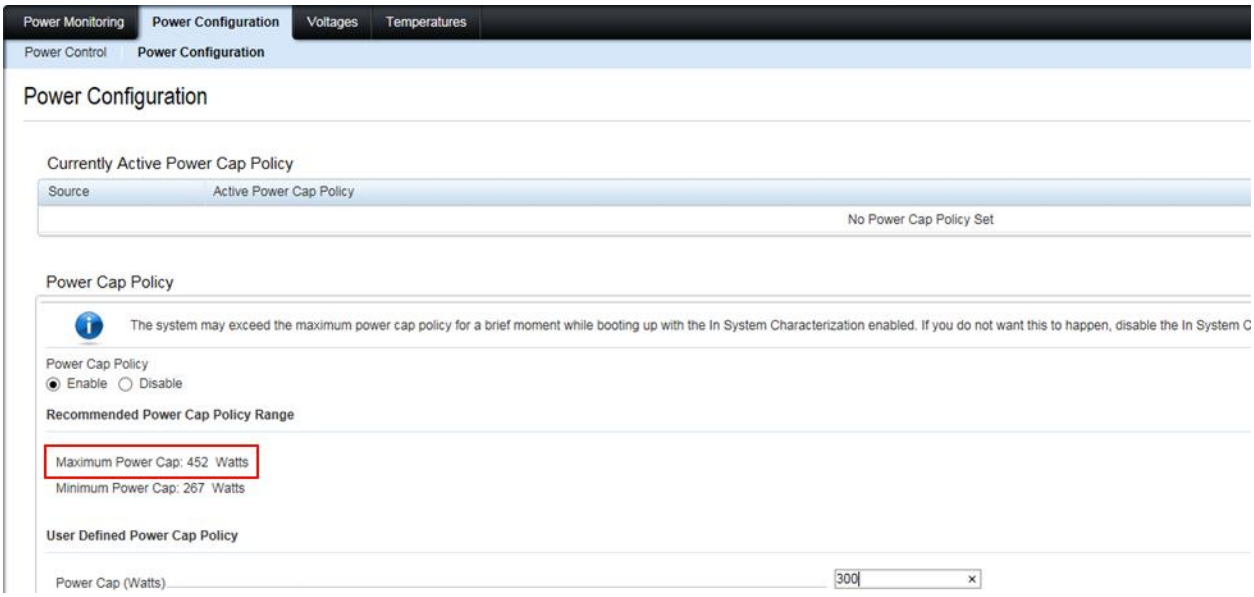
   ipmitool -I lanplus -H <iDRAC ip> -U <iDRAC admin username> -P <iDRAC admin user password> raw 0x30 0x17 *0x00 0x00 0x00 0x02 0x02 0x00 0x00 0x02 0x02*

2. Ensure that no Chassis Level Power Capping is set

   ipmitool -I lanplus -H <iDRAC ip> -U <iDRAC admin username> -P <iDRAC admin user password> raw 0x30 0xC5. (Byte 1 of *Response Data* shows no Chassis Level Power Capping Set)

   00 00 00 02 02 00 00 02 02

3. Find the Max Power Cap value by viewing the iDRAC GUI Power Configuration page. The sled power cap must be set less than this value.



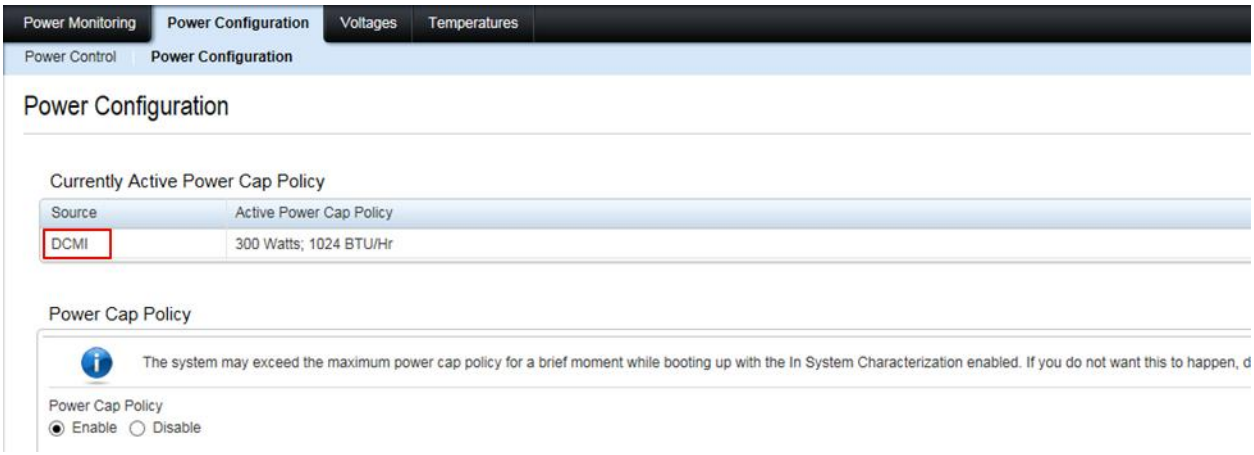
4. Set the Sled Level Power Capping on the individual sleds using DCMI Set Power Limit to 300W.

   ipmitool -I lanplus -H <iDRAC ip> -U <iDRAC admin username> -P <iDRAC admin user password> raw 0x2c 0x04 *0xdc 0x00 0x00 0x00 0x00 0x2c 0x01 0x00 0x00 0x00 0x00 0x00 0x00 0x00 0x00*
   dc

5. Activate Power Cap using DCMI Activate Power Limit; for details on DCMI, refer to the "Additional Resources" section

   ipmitool -I lanplus -H <iDRAC ip> -U <iDRAC admin username> -P <iDRAC admin user password> raw 0x2c 0x05 *0xdc 0x01 0x00 0x00*
   dc

6. View the iDRAC GUI Power Configuration page to verify that the Sled Power Limit is set using DCMI, as can be seen from the value of the Active Power Cap Policy.



Dell OEM IPMI Command Reference

This section highlights some of the Dell OEM IPMI commands that are relevant to Power Capping and Consumption on the C6320. For standard IPMI commands / DCMI commands, please refer to the respective command reference.

Get Chassis Enclosure Power Capping – NetFn 0x30, Cmd 0xC5

ipmitool -I lanplus -H <iDRAC ip> -U <iDRAC admin username> -P <iDRAC admin user password> raw 0x30 0xC5

Request Data: None

Response Data[1]:

Byte	Data Field
0	**Completion Code** 00h = Success 80h = BMC has not yet received this information from SC
1	**Chassis Power Capping Status** 00h – disabled (default) 01h – enabled
2:3	**Chassis Enclosure Power Limiting Value** In Watts, LSB First.
4	**Chassis Power Capping Action** 00h – Reserved 01h – Throttle via PROCHOT 02h – Throttle via software mechanism, such as NM or APML
5	**Sled level Power Capping exception** 00h – Reserved 01h – follow its own Sled level Power Capping 02h – follow Chassis level Power Capping
6:7	**Reserved**
8	**Emergency Capping Action (Chassis Policy)** 00h – do nothing 01h – throttle via PROCHOT 02h – throttle via software mechanism, such as NM 03h – Power Off
9	**Emergency Capping Action (Sled Policy)** 00h – follow Chassis Policy byte above 01h – throttle via PROCHOT 02h – throttle via software mechanism, such as NM 03h – Power Off Others, reserved.

[1] Note: ipmitool response does not show a success response code of 0x00 in the above examples

Set Chassis Enclosure Power Capping – NetFn 0x30, Cmd 0x17

ipmitool -I lanplus -H <iDRAC ip> -U <iDRAC admin username> -P <iDRAC admin user password> raw 0x30 0x17 0x01 0x90 0x01 0x02 0x02 0x00 0x00 0x01 0x01

Request Data:

Byte	Data Field
1	**Chassis Power Capping Status** 00h – disabled (default) 01h – enabled
2:3	**Chassis Enclosure Power Limiting Value** In Watts, LSB First.
4	**Chassis Power Capping Action** 00h – Reserved 01h – Throttle via PROCHOT 02h – Throttle via software mechanism, such as NM or APML
5	**Sled level Power Capping exception** 00h – Reserved 01h – follow its own Sled level Power Capping 02h – follow Chassis level Power Capping
6:7	**Reserved**
8	**Emergency Capping Action (Chassis Policy)** 00h – do nothing 01h – throttle via PROCHOT 02h – throttle via software mechanism, such as NM 03h – Power Off
9	**Emergency Capping Action (Sled Policy)** 00h – follow Chassis Policy byte above 01h – throttle via PROCHOT 02h – throttle via software mechanism, such as NM 03h – Power Off Others, reserved.

Response Data[2]:

Byte	**Data Field**
0	Completion Code 00h = Success 80h = BMC has not yet received configuration information from SC

[2] Note: ipmitool response does not show a success response code of 0x00 in the above examples

Get Sensor Info – NetFn 0x30, Cmd 0x16

ipmitool -I lanplus -H <iDRAC ip> -U <iDRAC admin username> -P <iDRAC admin user password> raw 0x30 0x16

Request Data: None

Response Data[3]:

Byte	Data Field
0	**Completion Code** 00h - Success 80h - BMC has not yet any information from SC 81h - SC Not Reporting for more than 10 consecutive seconds, but SC was alive before.
1	**SC BMC Communication Protocol Version** 01h for this specification.
2	**Checksum Byte** A zero checksum of the data start from the next byte to the end of data, with length specified in the Request Length Byte.
3	**Request Length Byte** The number of bytes following, excluding this Length Byte.
4	**SC Firmware Update Status Byte**. 00h – None. Firmware is working properly. 01h – Firmware image corrupted. 02h – Fan Table corrupted. 03h – Firmware update failed. 04h – Fan Table update failed.
5	**Chassis Inlet Temperature** In signed, 2's complement to represent between -127 to +127 degree Celsius. 00h - Inlet Temperature Sensor is not supported
6	**Chassis Exhaust Temperature** In signed, 2's complement to represent between -127 to +127 degree Celsius. 00h - Exhaust Temperature Sensor is not supported
7	**Sled Power Reading LSB** 00h - Sled Power Reading is not supported
8	**Sled Power Reading MSB** 00h - Sled Power Reading is not supported
9	**Sled IIN Current Reading** 00h - Input Current IIN is not supported
10	**Sled 12V VIN Voltage Reading** 00h - if 12V VIN Voltage is not supported
11	**PSU Presence Byte** 0b = Missing, 1b = Present [7] – PSU 8 [6] – PSU 7

[3] Note: ipmitool response does not show a success response code of 0x00 in the above examples

	[5] – PSU 6 [4] – PSU 5 [3] – PSU 4 [2] – PSU 3 [1] – PSU 2 [0] – PSU 1
12	**PSU Fault Byte** 0b = normal, 1b = error [7] – PSU 8 [6] – PSU 7 [5] – PSU 6 [4] – PSU 5 [3] – PSU 4 [2] – PSU 3 [1] – PSU 2 [0] – PSU 1
13:14	**PSU1 POUT, LSB first**. 00h - if this PSU is not supported.
15:16	**PSU2 POUT, LSB first.** 00h - if this PSU is not supported.
17:18	**PSU3 POUT, LSB first.** 00h - if this PSU is not supported.
19:20	**PSU4 POUT, LSB first.** 00h - if this PSU is not supported.
21:22	**PSU5 POUT, LSB first.** 00h - if this PSU is not supported.
23:24	**PSU6 POUT, LSB first.** 00h - if this PSU is not supported.
25:26	**PSU7 POUT, LSB first.** 00h - if this PSU is not supported.
27:28	**PSU8 POUT, LSB first.** 00h - if this PSU is not supported.
29	**Chassis Identification LED Status** 02h - Chassis Identification LED is Blinking 01h - Chassis Identification LED is Solid ON 00h - Chassis Identification LED is OFF / Chassis Identification LED is not supported or controlled by SC.
30	**Chassis Fault LED Status** 02h - Chassis Fault LED is Blinking 01h - Chassis Fault LED is Solid ON 00h - Chassis Fault LED is OFF / Chassis Fault LED is not supported or controlled by SC.
31	**PSU AC Loss Byte** 0b = normal, 1b = error [7] – PSU 8 [6] – PSU 7 [5] – PSU 6 [4] – PSU 5 [3] – PSU 4 [2] – PSU 3 [1] – PSU 2 [0] – PSU 1
32	**Reserved**

	00h
33	**Fan Control scheme** (1b = current algorithm in effect, 0b = not in effect.) [7:4] – Reserved. [3] – Closed loop [2] – Open loop [1] – Emergency mode (Fan Full Speed due to one or many Thermal Failure conditions) [0] – Fixed Fan Speed by user request. Note: This byte is a bit field to support the possible scenario where SC performs both closed loop and open loop fan controls simultaneously and takes the higher FAN PWM OUT as the actual output value.
34-N-1	**Fan m Reading** Where the value of m (1-base) depends on the number of fans specified in Set Chassis Configuration command.
N:N+4	**BMC Local Timestamp at the time when Set Sensor Info command was Received**. Please refer to IPMI 2.0, section 37, Timestamp Format.

Get Chassis Power Readings – NetFn 0x30, Cmd 0x2E

ipmitool -I lanplus -H <idrac ip> -U <idrac admin username> -P <idrac admin user password> raw 0x30 0x2E

Request Data: None

Response Data[4]:

Byte	Data Field
0	**Completion Code** 00h - Success 80h - BMC has not yet any information from SC 81h - SC Not Reporting for more 10 consecutive seconds, but SC was alive before.
1:4	**IPMI Timestamp of when the statistics are collected**
5	**Support bitmask** [7:4] – Reserved [1] – Chassis Cooling Power Consumption is available. [0] – Chassis Total Power Consumption is available.
6:7	**Chassis Total Power Consumption in watts. (LSB first)**
8:9	**Chassis Cooling Power Consumption in watts. (LSB first)**
10:13	**Reserved. Set as FFh.**

[4] Note: ipmitool response does not show a success response code of 0x00 in the above examples

Additional Resources

The iDRAC 8 User Guide has additional commands and usage scenarios for other iDRAC administration tasks:

http://en.community.dell.com/techcenter/systems-management/w/wiki/3204.dell-remote-access-controller-drac-iDRAC#manuals

IPMI Specification 2.0 provides the details of the IPMI standard:

http://www.intel.com/content/www/us/en/servers/ipmi/ipmi-intelligent-platform-mgt-interface-spec-2nd-gen-v2-0-spec-update.html

DCMI Specification Version 1.5 details the Data Center Manageability Interface including power management:

http://www.intel.com/content/www/us/en/data-center/dcmi/dcmi-v1-5-rev-spec.html