

データセンタのTCOを削減する DELLサーバリユース

デル株式会社
2011/04/01



Agenda

□ESSA

- 電力シミュレーションツール

□Power Capping

- 使用電力量の閾値設定

□Monitoring

- 電力量のモニタリング方法、ツール

ESSA

サーバ購入時の電力、ラックスペース問題点

必要以上に電気(アンペア)を準備してませんか？

・ データセンターの電力量

- ・ ラックあたりのアンペア数が固定されている
- ・ 最新のデータセンターでも12kVA、古いものでは数kVA
- ・ 電力量がボトルネックとなり、ラックに搭載できるサーバ数が制限

・ 定格電力による設計で無駄な設備コスト

- ・ 実運用ではサーバがフル稼働していても、定格の半分程度以下しか電力利用されない
- ・ 定格電力を基に設計された分電盤、空調、UPSが十分活用されない
- ・ 例) R610 X5570 x2 24GBでHPLを実行。(当時の最高スペック)
 - ・ R610の定格(PSU)は750W。アイドル状態では150W-200W。しかしHPL*実行時でも最大400-450Wと53-65%しか利用されない。

・ サーバの高密度化

- ・ 最近のサーバは、ブレードサーバやC6100など高密度化が進んだため、42ラックに単純にフル搭載すると、定格で20kVA-30kVA

*HPL: High Performance LInpack, スーパーコンピュータTOP500のベンチマークソフト。 CPUとメモリのアクセスが頻発し、サーバの潜在的な性能測定はもちろん、CPUとメモリに対して最も電力を消費することができるベンチマークの一つ。ただしディスクアクセスがほとんどないため、内蔵ディスクの電力を測るのには不向き。



ESSAによる購入前の電力量シミュレーション

ESSAによりサーバスペックに応じた消費電力をシミュレーション

- **Dell M1000eブレードサーバの定格電力の確認**
 - 1シャーシ / 16台 : 8100W (2700W PDU x3)
- **Dell ESSAを使った購入予定構成の最大消費電力確認**
 - ESSA (Energy Smart Solution Advisor) ツールで購入構成の100%利用率の電力量の確認
 - サーバの定格電力量では必要以上に多めの電力量が算出されることから、分電盤側に無駄な設備が発生
 - <http://www.dell.com/calc>
- **例) Dell M1000eブレードシャーシ**
 - M610 16台 (E5640 x2, 24GB mem, 146GB HDD, Mez x2)
 - 1シャーシ/16台 : 約3600W



ESSAによる購入前の電力量シミュレーション

ESSAによりサーバスペックに応じた消費電力をシミュレーション

<http://www.dell.com/calc>

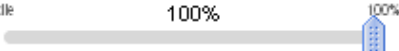
Dell Energy Smart Solution Advisor

Step 1 Add

Step 2 Configure



PowerEdge M610

CPU Loading: 

Workload: Transactional

Item (each)		
Input Power:	191.0watts	651.7btu/h
Input Current:		0.9amps
Weight:	6.4kg	14.1lbs

Limited Configuration

Full Configuration

PowerEdge M610

☒ PowerEdge M610 Blade Server; 56XX Processors

Primary Processor

- ☐ Intel Xeon X5560, 2.8Ghz, 8M Cache, 6.40 GT/s QPI, Turbo, HT
- ☐ Intel Xeon E5530, 2.4Ghz, 8M Cache, 5.86 GT/s QPI, Turbo, HT
- ☐ Intel Xeon E5506, 2.13Ghz, 4M Cache, 4.86 GT/s QPI
- ☐ Intel Xeon L5520, 2.26Ghz, 8M Cache, 5.86 GT/s QPI, Turbo, HT
- ☐ Intel® Xeon® E5503, 2.0Ghz, 4M Cache, 800MHz Max Mem
- ☐ Intel® Xeon® X5680, 3.33Ghz, 12M Cache,Turbo, HT, 1333MHz Max Mem
- ☐ Intel® Xeon® X5670, 2.93Ghz, 12M Cache,Turbo, HT, 1333MHz Max Mem
- ☐ Intel® Xeon® X5667, 3.06Ghz, 12M Cache,Turbo, HT, 1333MHz Max Mem
- ☐ Intel® Xeon® X5660, 2.8Ghz, 12M Cache,Turbo, HT, 1333MHz Max Mem
- ☐ Intel® Xeon® X5650, 2.66Ghz, 12M Cache,Turbo, HT, 1333MHz Max Mem
- ☒ Intel® Xeon® E5640 2.66Ghz, 12M Cache,Turbo, HT, 1066MHz Max Mem

Solution Configuration

Options Open Save New Print

Configuration Totals for 220 AC Input Voltage and 25°C

Total Input Power:	3586.0watts	12235.9btu/h
Total Input Current:		16.5amps
Sound Power Level:		0.0bels *
Airflow Rate:	0.0l/s	0.0CFM
Total Weight:	146.9kg	323.9lbs

Enclosures (click on model to configure) 6 error(s)

- 1 x PowerEdge M1000E Blade
- 16 x PowerEdge M610

M1000e ESSA利用によるラック高密度化

Rackあたりの電源容量 9KVAの場合の利用例

ESSAで実際の利用シーンに基づいた M1000eシャーシ単位で消費電力をシミュレーション、ラック数/UPSなどが可能！

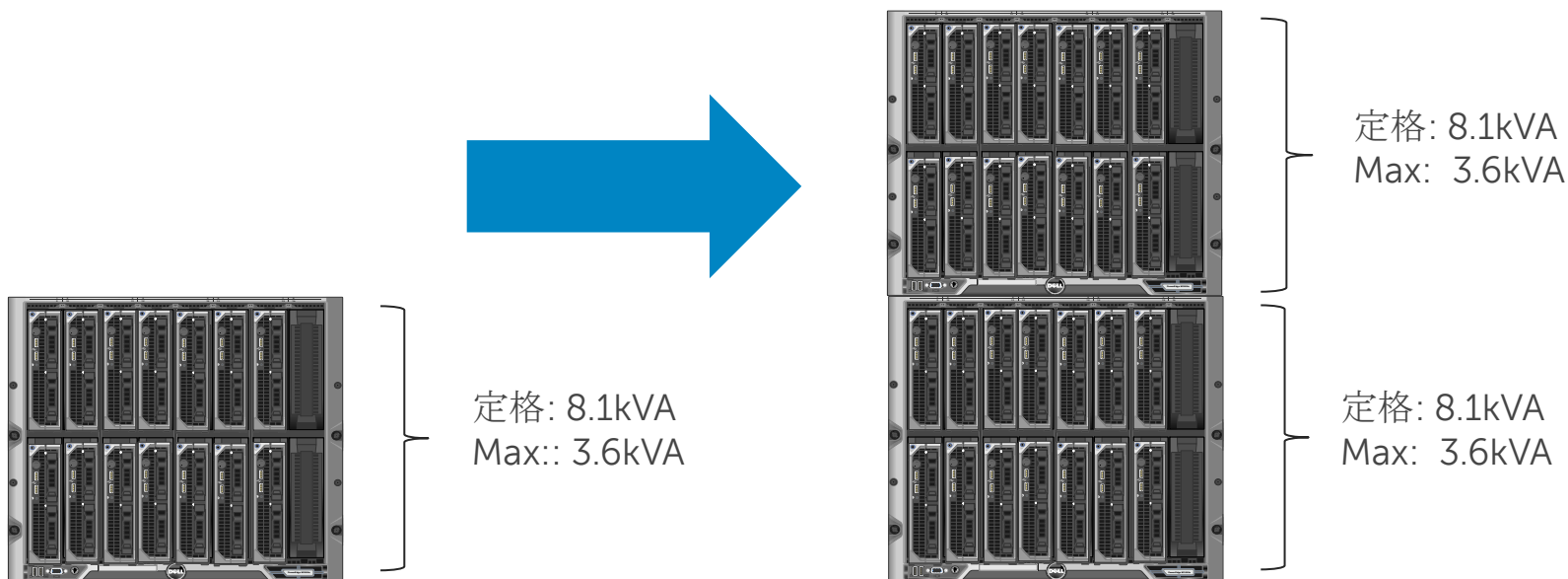
【定格電力で設計】

定格**8.1kVA** / Max: 3.6kVA
M1000e 1台/サーバ 16台

【ESSAで設計した場合】

定格 16.2kVA, **Max: 7.2kVA**,
M1000e 2台、サーバ 32台

定格で計算すればラックあたり1シャーシしか搭載できなかったが、ESSAでシミュレーションすることで2シャーシ搭載可能に。

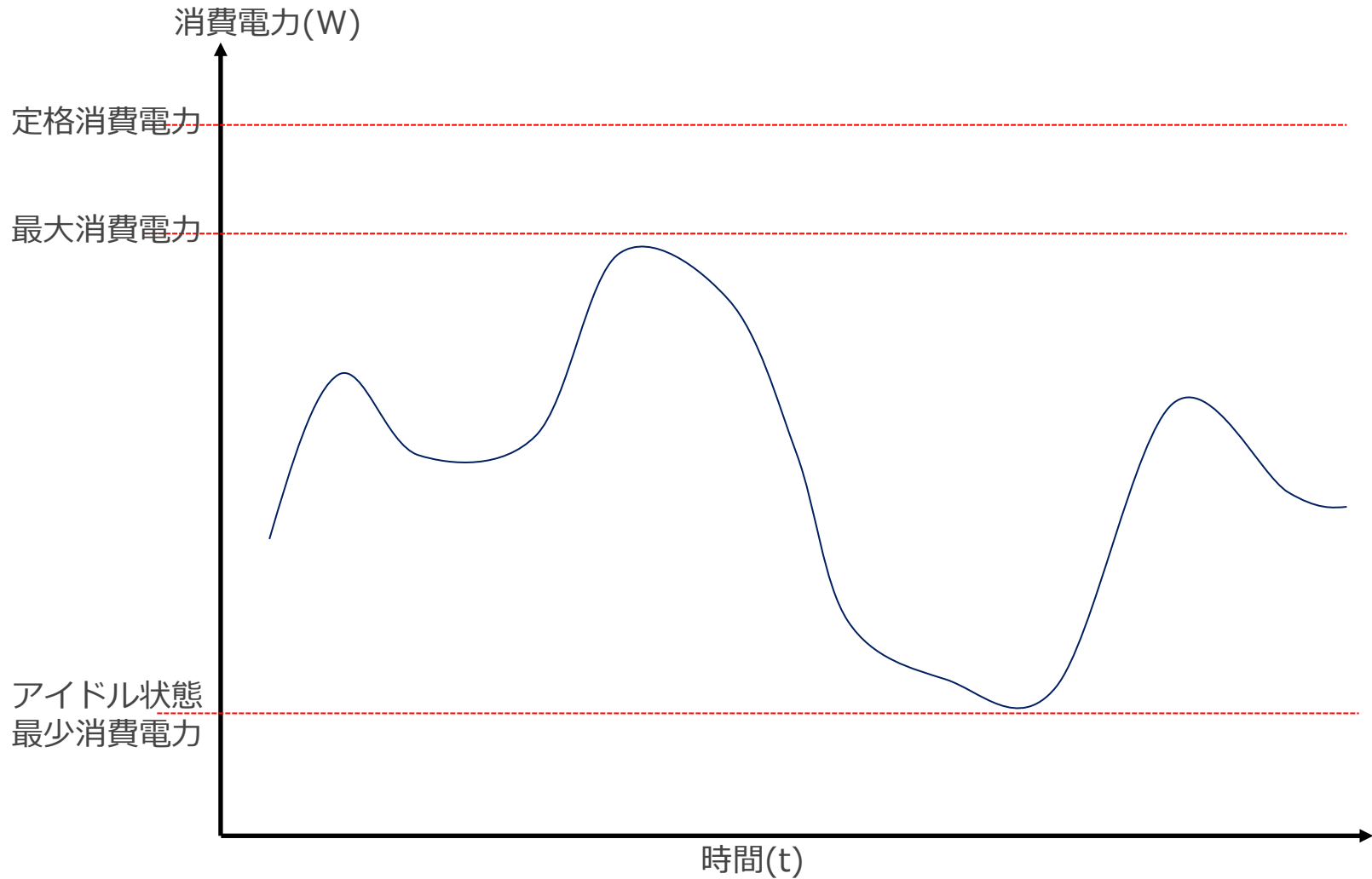


- 例) Dell M1000eブレードシャーシ、PCT M6220 x2 M6248 x4
 - M610 ブレード 16枚 E5650 2.53GHz x2、24GB MEM、146GB 15Krpm HDD x2、Doughter Card x2

Power Capping

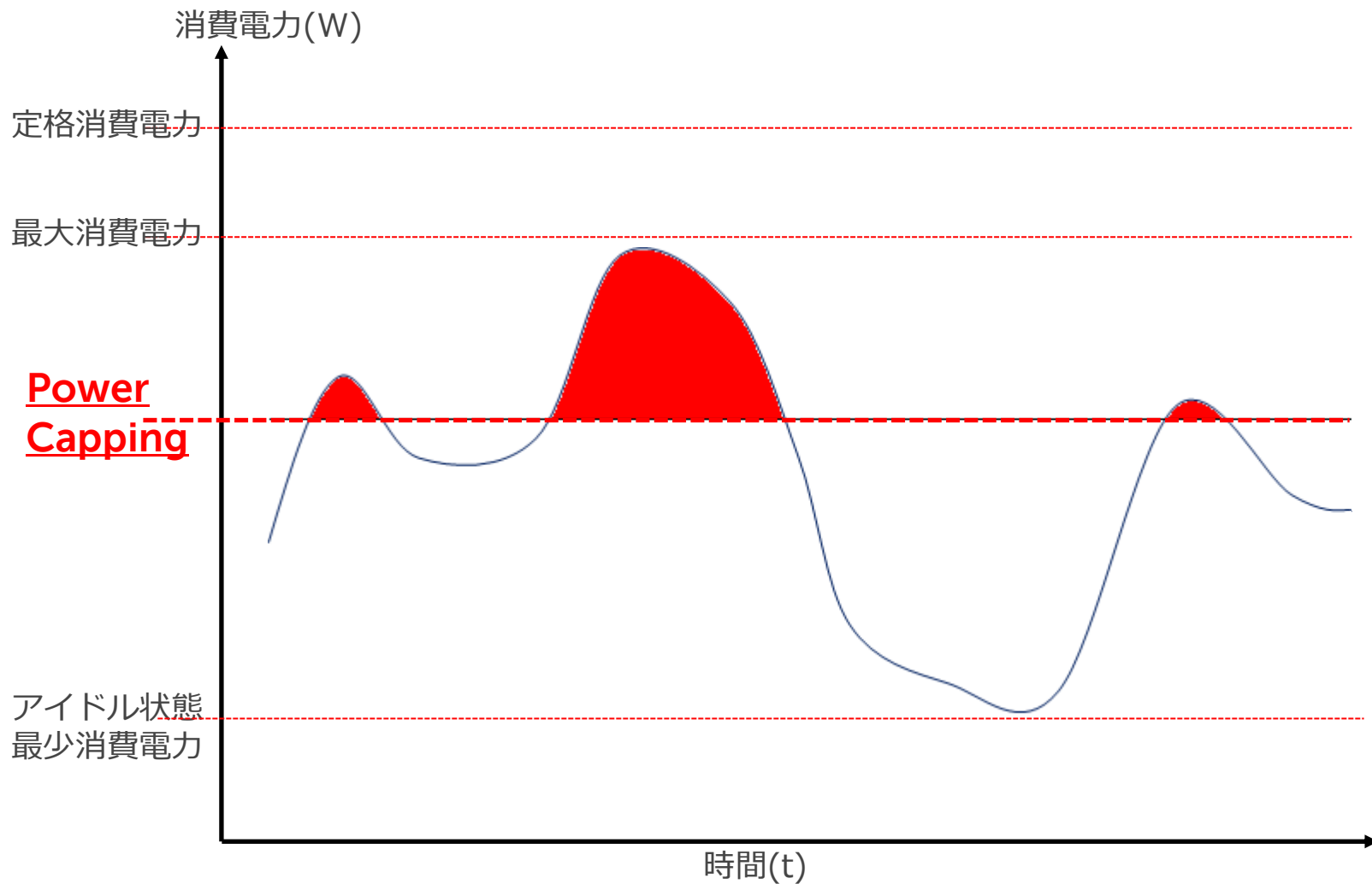
Power Cappingの効果 - 通常利用時

PowerCappingを適用しない状態の時間ごとの電力使用例



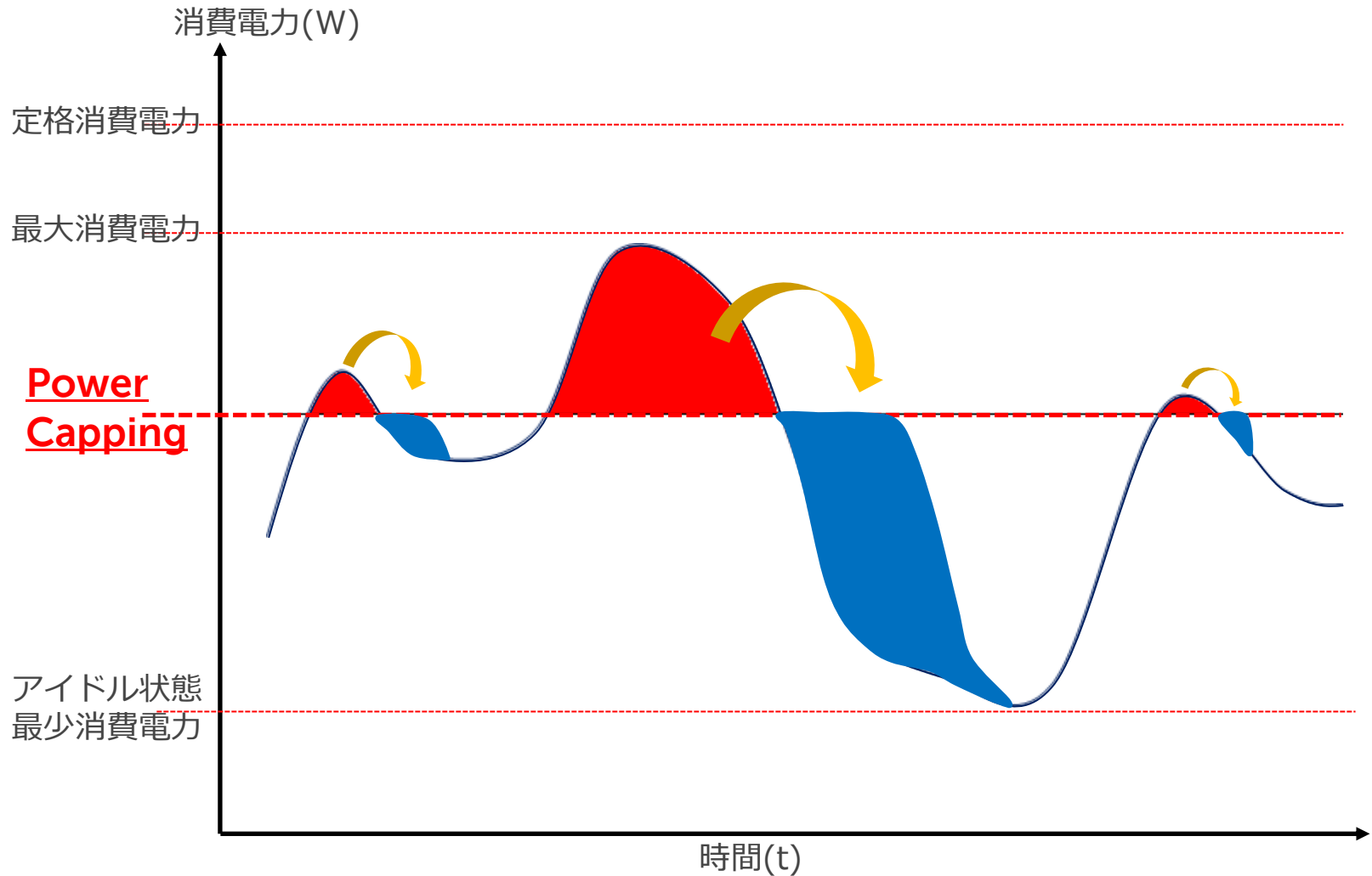
Power Cappingの効果- サーバ消費電力のピークカット

PowerCappingの閾値設定



Power Cappingの効果 - PowerCapping後の動作

ピークカットされた分の処理は負荷の低い時間に処理



- PowerCappingによって処理できなかったLoadを空いている時間に処理。設定した電力閾値を超えないようにする。

M1000e PowerCapping利用によるラック高密度化

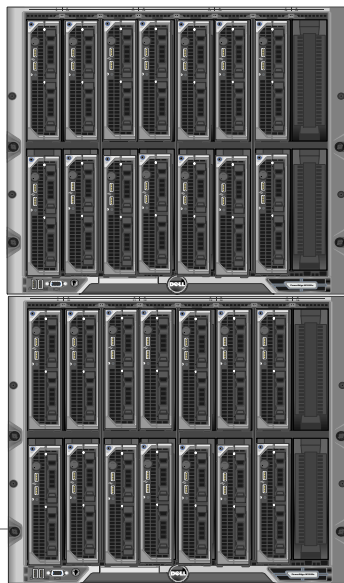
Rackあたりの電源容量 10KVAの場合の利用例

DELL Blade シャーシM1000e単位で電源上限制御をおこなうことにより
ラック数/電力削減が可能！

【通常の利用方法の場合】

Max:**7.4kVA** / Ave: 5.7kVA

M1000e 2台/サーバ 32台



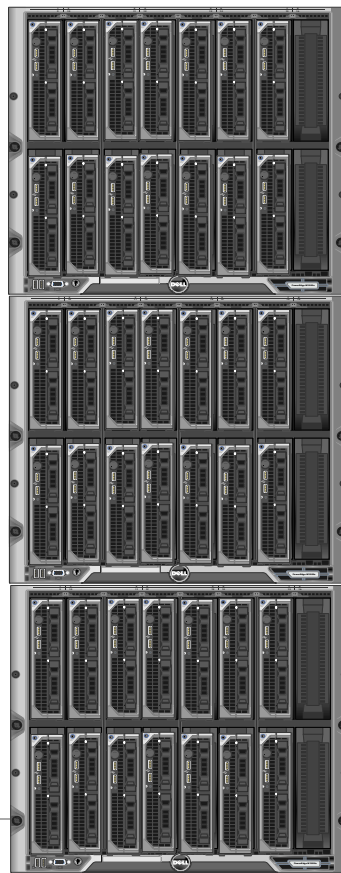
Max:3.7kVA
Ave: 2.9kVA
Min: 2.4kVA

Max:3.6kVA
Ave: 2.8kVA
Min: 2.4kVA

【PowerCap利用の場合】

Max: **9.8kVA** / Ave: 8.2kVA

M1000e **3台**/サーバ **48台**



Max:3.3kVA(Capped)
Ave: 2.9kVA
Min: 2.4kVA

Max:3.3kVA(Capped)
Ave: 2.9kVA
Min: 2.4kVA

Max:3.2kVA(Capped)
Ave: 2.8kVA
Min: 2.4kVA

Power Monitoring

ブレード管理コントローラCMCによる電力モニタリング

ブレードシャーシのリアルタイム電力モニタリング

The screenshot displays the Dell CMC-6GFRSBX Chassis Management Controller interface. The left sidebar shows a tree view of the system components, including the chassis controller, server components, and I/O modules. The main content area is titled '電源モニタ' (Power Monitor) and 'リアルタイム電力統計' (Real-time Power Statistics). A table lists various power-related attributes and their current values.

属性	値
システム入力電力	2145 W (7319 BTU/h)
ピークシステム電力	2415 W (8240 BTU/h)
ピークシステム電力開始時刻	12:56:59 04/02/2011
ピークシステム電力タイムスタンプ	13:00:53 04/02/2011
最小システム電力	2123 W (7243 BTU/h)
最小システム電力開始時刻	12:56:59 04/02/2011
最小システム電力タイムスタンプ	13:06:12 04/02/2011
システム アイドル電力	2145 W (7319 BTU/h)

- PowerCappingによって処理できなかったLoadを空いている時間に処理。設定した電力閾値を超えないようにする。

M1000e PowerCapping利用によるラック/電力削減

DELL Blade シャーシM1000e単位で電源上限値の設定が可能
サーバごとの優先順位の設定が可能

属性	値
システム入力電力の上限 (0-23% = 2715W - 100% = 11637W)	11637 W 39707 BTU/h 100 %
冗長性ポリシー	電源の冗長性 ▼
電源の冗長性を越えたサーバーパフォーマンス	☒
電源装置の動的制御を有効にする	☐
シャーシ/電源ボタンを無効にする	☐
110 VAC 操作を許可	☐

サーバーの優先順位



スロット	名前	モデル	電源状況	優先順位
1	Slot-01	PowerEdgeM600	オン	1 (高) ▼
2	Slot-02	PowerEdgeM600	オフ	2 ▼
3	Slot-03	PowerEdgeM600	オフ	1 (高) 2 3 4 5 6 7 8 9 (低)
4	Slot-04	PowerEdgeM600	オフ	4 ▼
5	Slot-05	PowerEdgeM600	オフ	
6	Slot-06	PowerEdgeM600	オフ	
7	Slot-07	PowerEdgeM600	オフ	
8	Slot-08	PowerEdgeM600	オフ	1 (高) ▼

PowerEdgeサーバのインタラクティブLDC

サーバ前面から電力モニタリングを含む迅速な情報アクセスとアクション

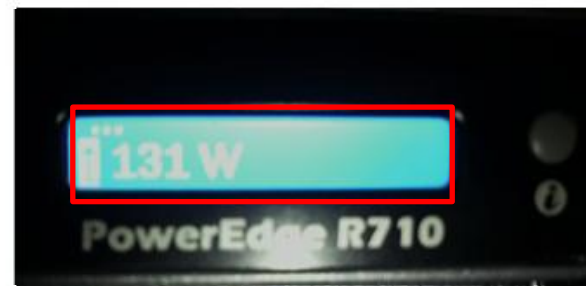


イベント発生時には、
瞬時にメッセージを表示します。

□ 操作方法

項目	ボタン	説明
1	左	カーソルが後方に1つ分移動します。
2	選択	カーソルによってハイライト表示されているメニュー項目を選択します。
3	右	カーソルが前方に1つ分移動します。 メッセージのスクロール中に次の操作ができます。 1回押すとスクロールが速くなります。 - 再び押すと停止します。 - もう1回押すとデフォルトのスクロール速度に戻ります。 - もう1回押すと同じ操作が繰り返されます。
4	システムID	システム識別モードのオン/オフを切り替えます。 すばやく押してシステム識別のオン/オフを切り替えます。POST中にシステムがハングした場合は、システム識別ボタンを5秒以上押し続けてBIOSプログレスモードに入ります。

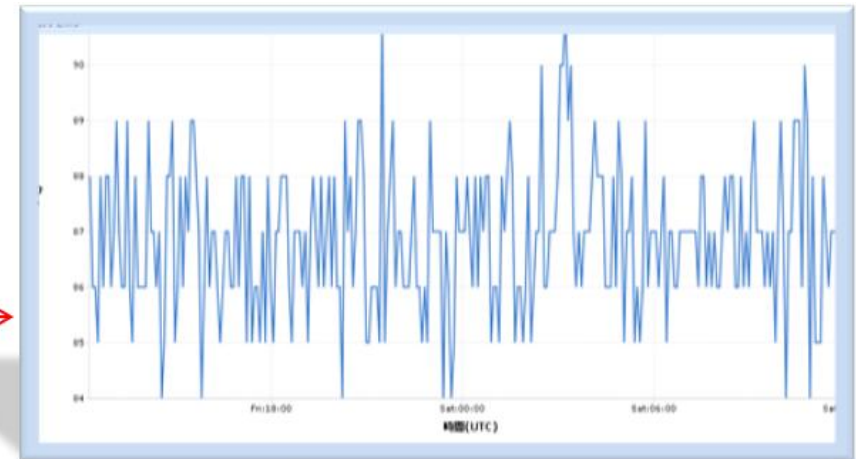
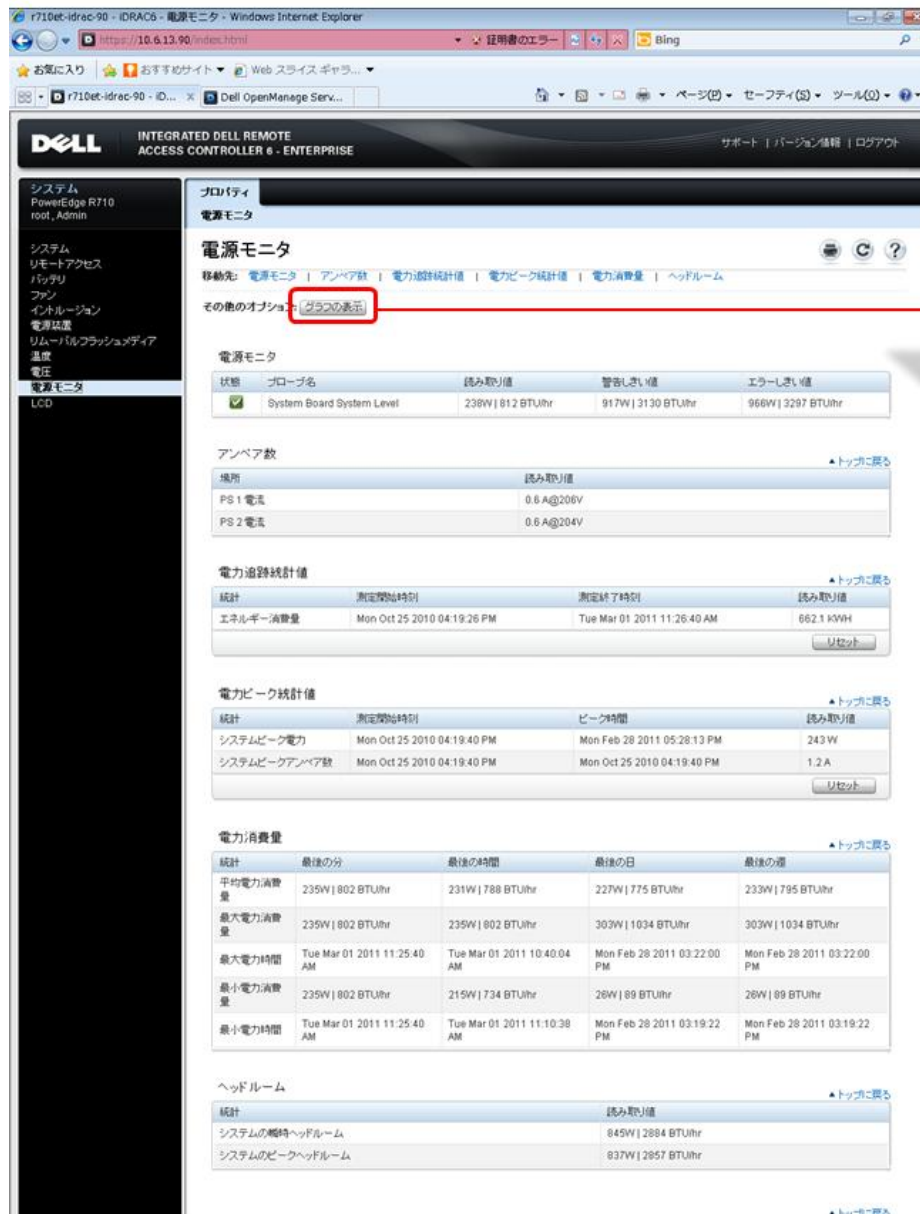
- 障害発生箇所と障害状況を知らせる
- 画面による情報閲覧と監視
 - ・ ホスト名、資産番号、サービスタグなど
 - ・ NICのMACアドレス
 - ・ リモート管理チップのIPアドレス
 - ・ 筐体内温度
 - ・ 消費電力



サーバの消費電力を簡単に確認

iDRACによる電力モニタリング

ネットワーク経由でサーバ単体のリアルタイム電力モニタリング



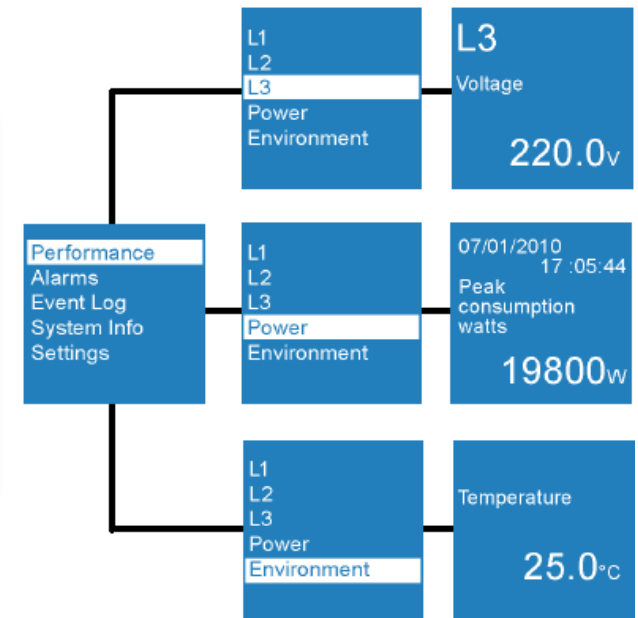
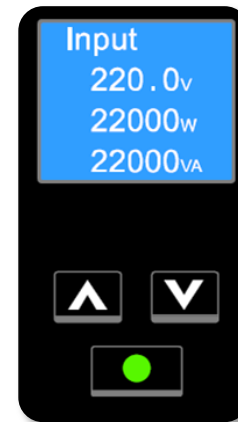
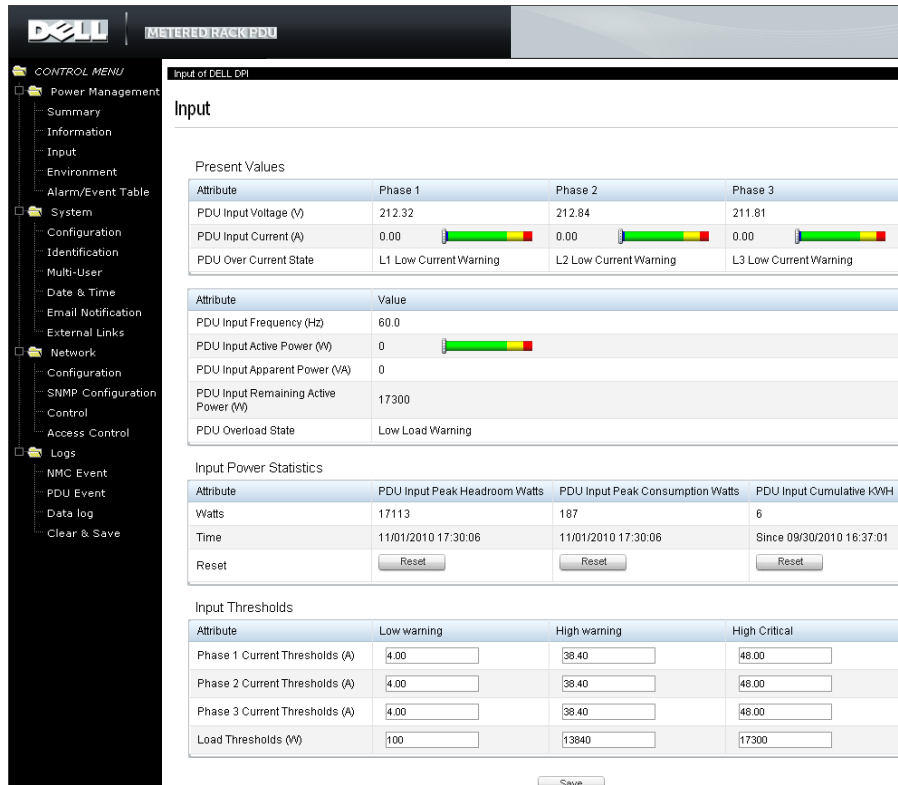
時系列で消費した電力値を表示（過去7日分）

消費電力のほか、電力の使用量（ワット×時間）やアンペア、電力のピーク値などアセスメントをする上で役立つ電力の統計情報を簡単に取得することができます。

Dell LCD Metered PDU

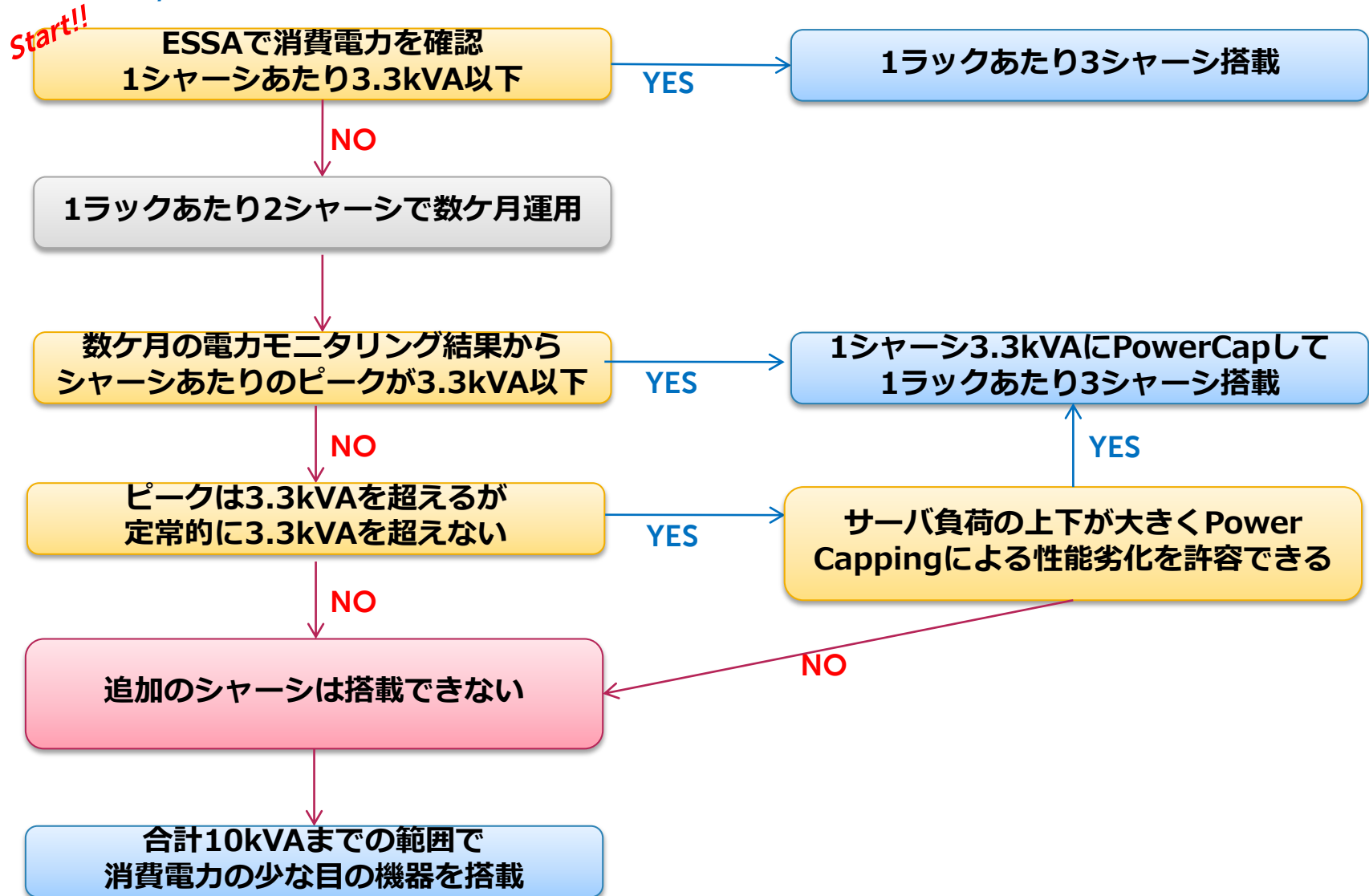
PDU側での電力モニタリング

閾値の設定/SNMP、SMTPアラートの設定が可能！



ブレードシャーシ購入計画フローチャート

10kVA/ラックに3シャーシ搭載する場合



- 力率を1.0として概算

Appendix

P-Stateによる節電効果

PowerCappingでは動的にCPUのクロックを下げることで電力を制御！

SKU		TDP	Turbo (Per # of Active Cores)					Nehalem-EP P-states Power Estimates Only (Per Processor Package)												
			1c	2c	3c	4c	P0	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13
WS 2S	W5580	130W	3.46	3.33	3.33	3.33	≥	3.20	3.06	2.93	2.80	2.66	2.53	2.40	2.26	2.13	2.00	1.86	1.73	1.60
			≤ 130W					124W	119W	113W	108W	102W	97W	92W	87W	82W	77W	72W	68W	
WS 1S	W3570	130W	3.46	3.33	3.33	3.33	≥	3.20	3.06	2.93	2.80	2.66	2.53	2.40	2.26	2.13	2.00	1.86	1.73	1.60
			≤ 130W					124W	119W	113W	108W	102W	97W	92W	87W	82W	77W	72W	68W	
	W3540		3.20	3.06	3.06	3.06	≥	2.93	2.80	2.66	2.53	2.40	2.26	2.13	2.00	1.86	1.73	1.60		
			≤ 130W					124W	118W	112W	106W	100W	95W	89W	84W	79W	73W			
	W3520		2.93	2.80	2.80	2.80	≥	2.66	2.53	2.40	2.26	2.13	2.00	1.86	1.73	1.60				
≤ 130W					123W	117W	110W	104W	98W	92W	86W	80W								
Perf	X5570	95W	3.33	3.33	3.20	3.20	≥	2.93	2.80	2.66	2.53	2.40	2.26	2.13	2.00	1.86	1.73	1.60		
			≤ 95W					90W	86W	81W	76W	72W	68W	63W	59W	55W	51W			
	X5560		3.20	3.20	3.06	3.06	≥	2.80	2.66	2.53	2.40	2.26	2.13	2.00	1.86	1.73	1.60			
			≤ 95W					90W	85W	80W	76W	71W	67W	62W	58W	54W				
	X5550		3.06	3.06	2.93	2.93	≥	2.66	2.53	2.40	2.26	2.13	2.00	1.86	1.73	1.60				
≤ 95W					90W	85W	80W	75W	70W	66W	61W	56W								
Vol	E5540	80W	2.80	2.80	2.66	2.66	≥	2.53	2.40	2.26	2.13	2.00	1.86	1.73	1.60					
			≤ 80W					76W	71W	67W	63W	59W	55W	51W						
	E5530		2.66	2.66	2.53	2.53	≥	2.40	2.26	2.13	2.00	1.86	1.73	1.60						
			≤ 80W					75W	71W	66W	62W	58W	54W							
	E5520		2.53	2.53	2.40	2.40	≥	2.26	2.13	2.00	1.86	1.73	1.60							
≤ 80W					75W	70W	66W	61W	57W											
Value	E5506	80W	2.13	2.13	2.13	2.13	≥	2.13	2.00	1.86	1.73	1.60								
			≤ 80W					75W	70W	65W	60W									
	E5504		2.00	2.00	2.00	2.00	≥	2.00	1.86	1.73	1.60									
			≤ 80W					75W	69W	64W										
	E5502		1.86	1.86	1.86	1.86	≥	1.86	1.73	1.60										
≤ 80W					74W	69W														
LV	L5520	60W	2.53	2.53	2.40	2.40	≥	2.26	2.13	2.00	1.86	1.73	1.60							
			≤ 60W					56W	52W	48W	44W	40W								
LV Value	L5506		2.13	2.13	2.13	2.13	≥	2.13	2.00	1.86	1.73	1.60								
			≤ 60W					56W	51W	47W	43W									

Notes:

- P-states are CPU performance states (power consumption & capability) within the C0 active/executing state. P0 = Turbo freq, P1 = TDP freq, P2...Pn. Power savings shown are NOT processor specifications in the EMTS.
- Only a gross approximation relative from the processor TDP. Measured power consumption in a given P-state may be higher or lower, depending on workload and variable operating voltage (mileage may vary). Other nomenclature: Demand Based Switching (DBS) w/ Enhanced Intel SpeedStep® Technology (EIST)
- Intel(R) Turbo Boost Technology, defined as P0, automatically & continuously, allows the CPU to run faster

Notes:

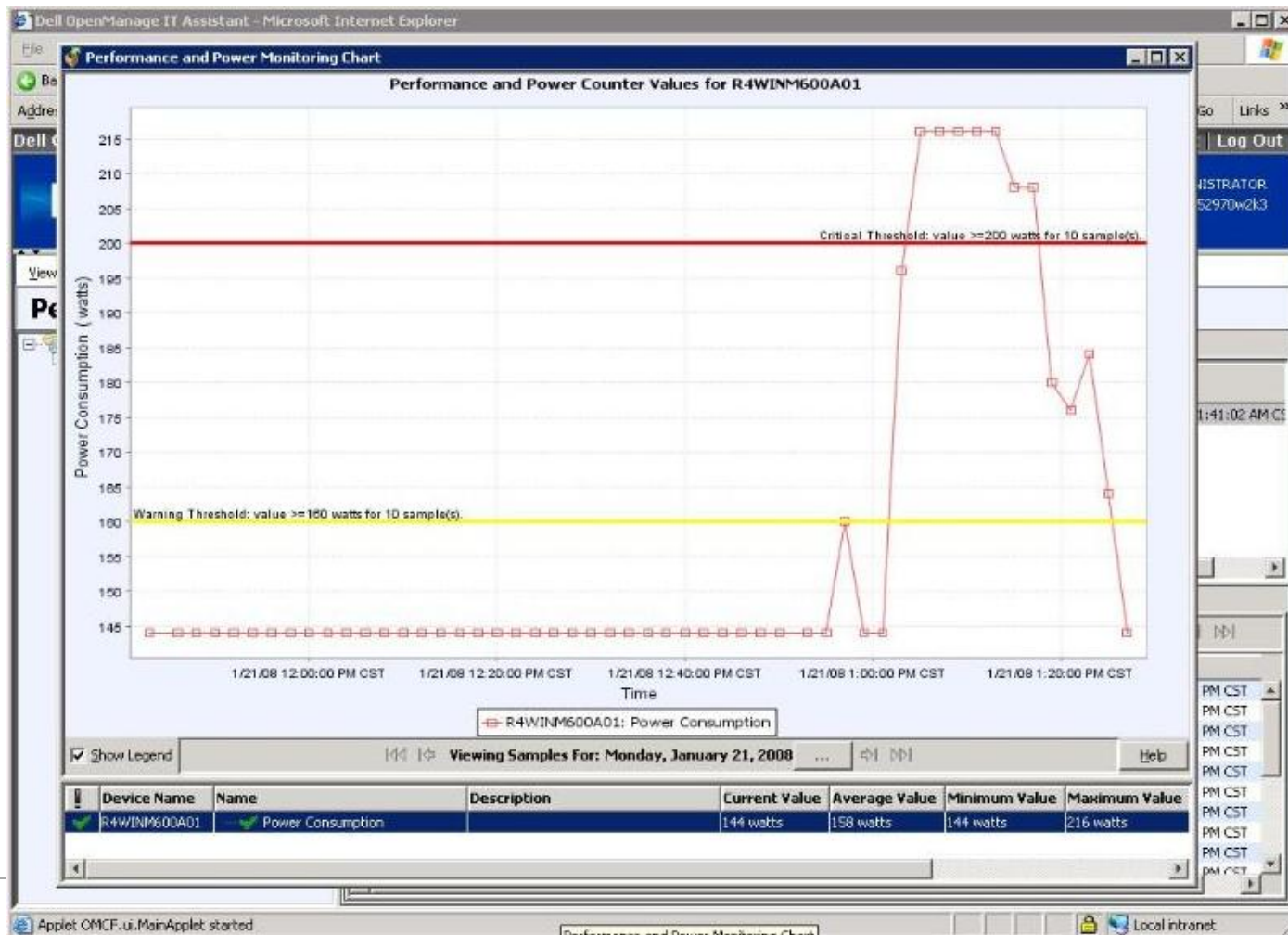
- P-states are CPU performance states (power consumption & capability) within the C0 active/executing state. P0 = Turbo freq, P1 = TDP freq, P2...Pn. Power savings shown are NOT processor specifications in the EMTs.
- Only a gross approximation relative from the processor TDP. Measured power consumption in a given P-state may be higher or lower, depending on workload and variable operating voltage (mileage may vary). Other nomenclature: Demand Based Switching (DBS) w/ Enhanced Intel SpeedStep® Technology (EIST)
- Intel(R) Turbo Boost Technology, defined as P0, opportunistically & automatically, allows the CPU to run faster



DELL ITAssistant(ITA) による電力モニタリング

ITA, DMC, による長期的な電力モニタリング

- 複数HWの障害検知、電力モニタリングが可能なサーバ添付の無償ツール



低負荷時の電源効率を最大化する M1000e DPSE

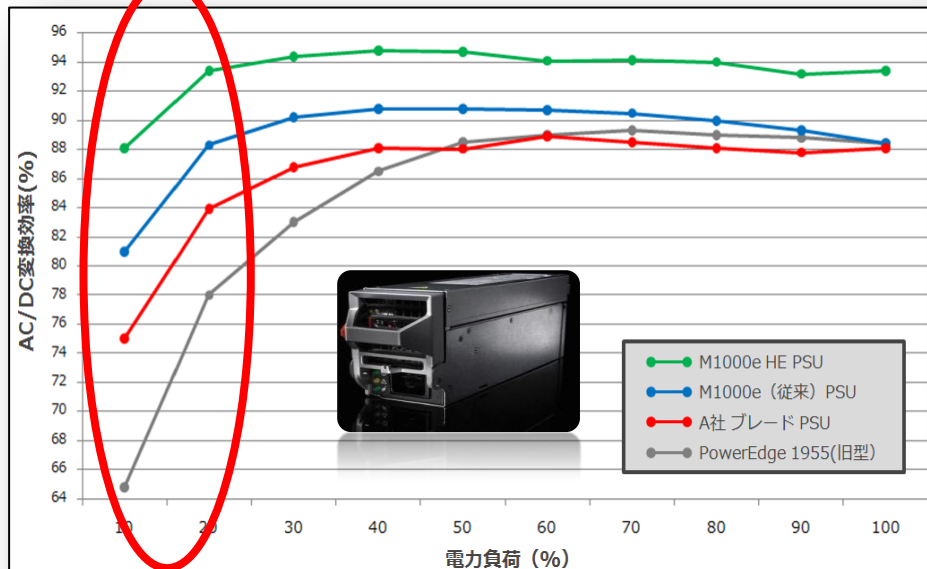
M1000eの電源動的制御機能

DPSE：ダイナミック・パワー・サプライ・エンゲージメント

CMCの標準機能！！

- 電源負荷が低場合は一般的にAC/DC変換効率は悪くなります
- DPSEは負荷が低い場合、自動で6つのPSの不要な電源をOffにする機能です
- PSあたりの負荷の向上により高いAC/DC変換効率を維持します
- DPSEをONにしてもPSの冗長化は損なわれません

【電源負荷とAC/DC変換効率】



【DPSE動作イメージ】

