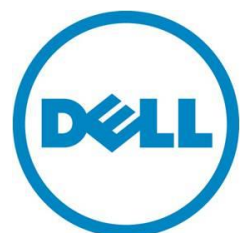


Power and Thermal Considerations for FX Platform Configurations

This white paper describes and illustrates factors to take into account when configuring FX infrastructures.

FLEXIBLE.PERFORMANCE.INNOVATIVE.POSSIBILITY
CONVERGED.**POWER**.DYNAMIC.STORAGE.SLEEK
AND
OD
LO
PH
SCALABLE.NETWORKING.**EDGE**.REVOLUTIONARY.
WORKLOADS.VIRTUALIZATION.DENSE.MANAGEABLE.



Power and Thermal Considerations for FX Platform Configurations

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.

©2015 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. Microsoft, Windows, and Windows Server are either trademarks or registered trademarks of Microsoft Corporation in the United States and/or other countries. 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.

November 2015 | Version 1.22

Introduction

The tremendous flexibility and modularity of the FX platform allows you to achieve new levels of density as you address many different workloads in your data center. This density is one of the major factors in what makes the FX platform so cost efficient. But, along with this incredible density comes the need to pay closer attention to power and thermal considerations as you configure FX solutions.

This document is intended to familiarize you with the power and thermal related factors that can influence the configuration of an FX infrastructure. With a better understanding of the power/thermal environment you can achieve better results as you work with your Dell team to determine the final configuration details of FX-based infrastructures.

Power Considerations

When configuring an FX system, two of the primary considerations are how much power will be required to run the chosen servers and storage at the desired level of availability, and what configuration you need to ensure enough headroom for spikes in workload demand. To get the right answer for your specific datacenter, there are some things you need to know:

- FX systems support dual power supplies of either 2000W or 1600W (the 1100W PSU is available with FM120 micro-server configurations only). The FX2 enclosure consumes AC power and distributes the load across the active power supply unit.
- To provide power redundancy the 1600 & 2000W PSUs for FX were designed to support running at 150% of a single PSU's rated capacity (e.g. 2400 & 3000W respectively) **for short periods of time**, so that if one PSU fails, the system has time to throttle back CPU/Memory power to enable operations to continue with just one PSU.
- An FX2 chassis with two 2000W PSUs can deliver a maximum of up to 3000W of continuous AC power (connection-imposed chassis limitations prevent the full 4000W). That power is allocated to server modules and the associated enclosure infrastructure. However, this capacity varies based on the power redundancy policy selected. Power can be configured to one of three levels of redundancy:
 1. 1+1 Grid Redundant – power draw never to exceed 1 PSU's capacity
 2. Redundancy Alerting Only (RAO) – allowed to exceed 1 PSU's capacity, but alerts on overage occurrence
 3. 2+0 Non-redundant – power draw allowed to exceed 1 PSU's capacity, but does NOT trigger an alert.

These will be explained more fully below.

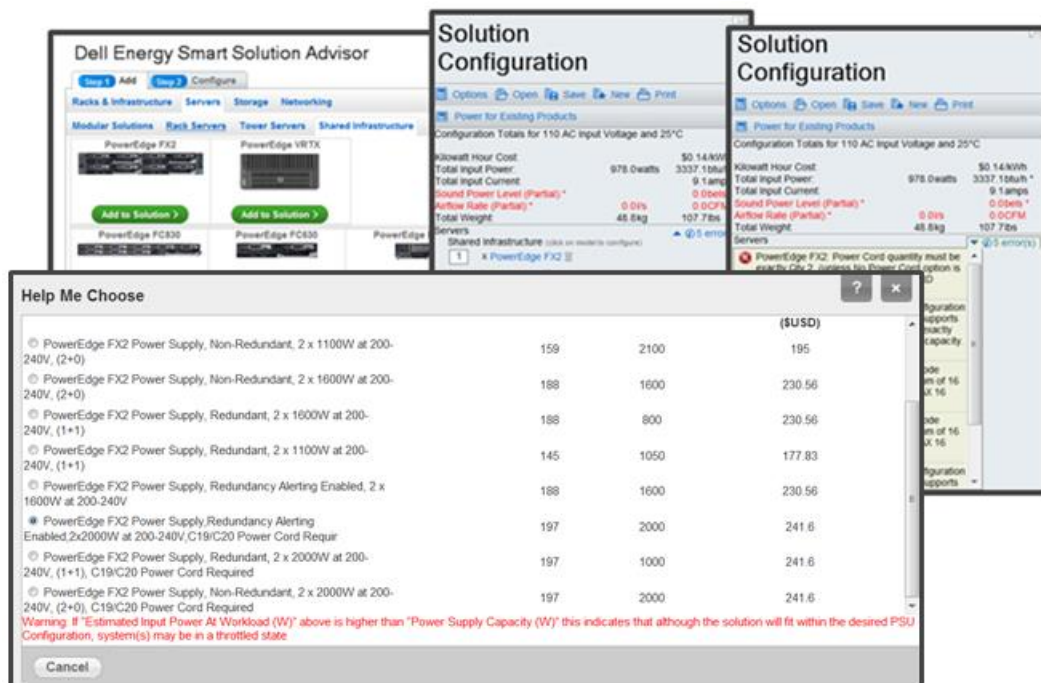
- The major draws on power in an FX platform are (in rough order of magnitude): processors, amount of memory and number and type of storage devices. Cooling fans, networking cards, and other components in the chassis impact power draw to a much lesser degree.
- CPU utilization is a major factor in power draw, and may be heavily influenced by workload. For example, a file server may create very little power draw, while an HPC application may push the CPU to 90% of its capacity.

- While the FX platform is designed to be a data center system, and most implementations will use high-line 208-240v input power sources, there are cases where a low-line 100-120V input is used, and in those cases the effective output power from the PSUs is cut in half (2000W -> 1000W / 1600W -> 800W).

The ESSA Tool: Pre-configure your FX system online first

Dell provides a tool - the [Energy Smart Solution Advisor](#) (ESSA) - that can help you plan your FX platform (or any Dell system) for maximum efficiency by letting you [to build out \(online\) potential configurations of systems](#). Along with other configuration guidance, it determines what the constructed system's power draw will be.

In a graphical interface, this tool enables you to select an FX2 chassis and load it with FX servers and storage blocks.



Simple Screen Shots of the ESSA Tool

It gives detailed messages that help you build the correct configuration as you add components. It uses the same configuration validation rules used by your sales person and can be an invaluable assistance while constructing your infrastructure to fit your workload and OPEX requirements. Note that power transitions in the processor (i.e., turbo-mode) can have an important, yet hard to predict, impact on power draw – all the more reason to use the ESSA tool.

We encourage you to use ESSA and work with your Dell sales reps to finalize the details of more complex configurations, once you get to that point. To get started with ESSA, you can refer to the [ESSA FAQ and User Guide](#).

Redundancy Policy: Availability level and dual PSU operating modes

The Chassis Management Controller (CMC) in the FX platform lets you set one of three levels of PSU operating modes that determine the level of system redundancy and what action is taken if there is a PSU failure.

The three levels are as follows:

1+1 Grid Redundant Power Supplies mode

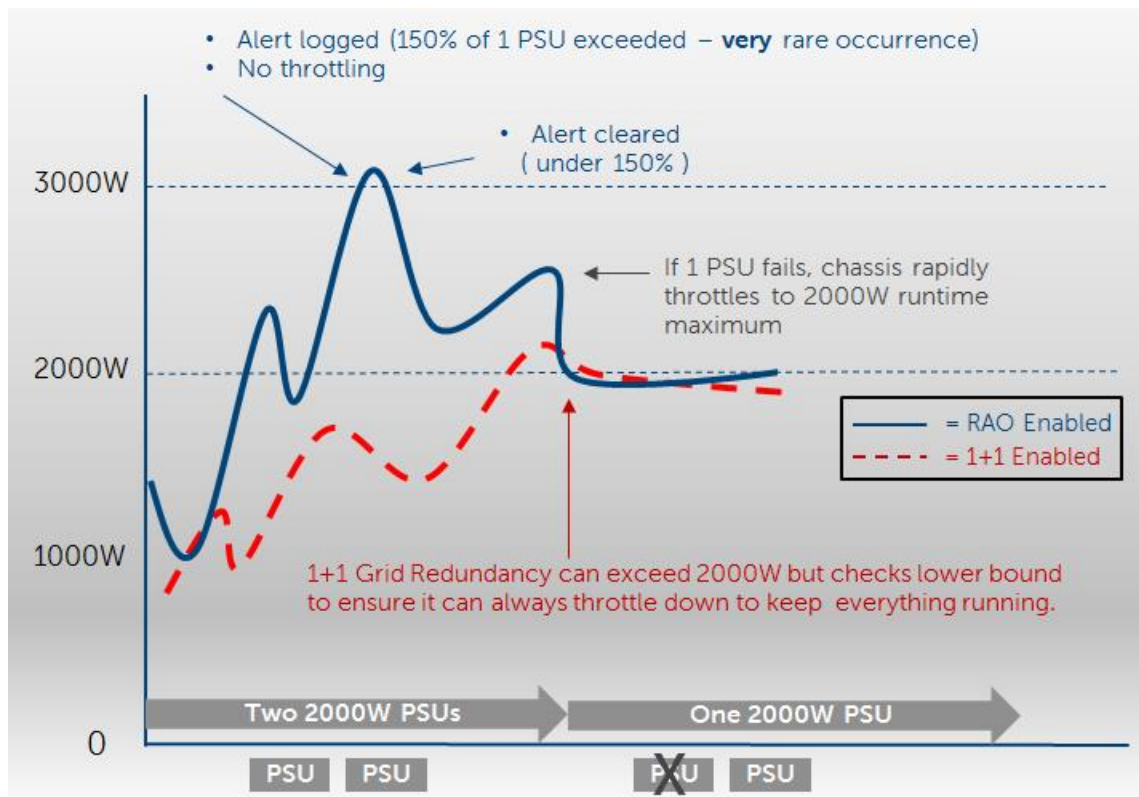
When set to this mode, system power (draw) is limited to the “name-plate” capacity supportable by a single PSU (**not** the 150% short-term capacity). This guarantees continued application availability in the case of PSU/grid failures. For example, an FX platform with two 2000W PSUs can only be configured to draw a peak power of 3000W. During system boot a “power budget check” is run to determine (and possibly limit) the number of nodes allowed to operate in the chassis. While 1+1 mode provides complete power redundancy, you must accept the limitation of the maximum power capacity.

Redundancy Alerting Enabled (RAO) mode

When set to this mode, system power draw can exceed the maximum capacity of a single PSU, and use the capability of the second PSU. Note that when a PSU fails with this setting, up to 3000W power draw can be maintained for a few milliseconds (using the extra 150% short-term capacity of the surviving PSU) while the system workload is throttled back to the rated 2000W level of the single remaining PSU. In this mode, an alert is sent when single PSU capacity is exceeded - so that the administrator is made aware of the (relatively small) potential risk to availability. The alert is cleared when the power draw drops back down to the redundant level.

RAO is the recommended (and default) setting for FX platforms. It poses very low risk potential to FX infrastructures (failure rates of today’s PSUs is very low), while offering the extra power capability of the second PSU for occasional spikes in demand.

Dell recommends that you monitor RAO enabled systems to ensure that any power “excursions” are very short-lived and infrequent enough to be considered an acceptable risk to the business. If a system is exceeding redundant capacity regularly and/or for long periods of time, you should take action to migrate workloads to other systems. In RAO mode you may also want to take advantage of OpenManage Power Center (OMPC) capabilities to put in place policy/controls that throttle power draw when extended excursions occur - to further minimize your exposure to potential downtime.



An illustration of how a dual 2000W PSU system operates- before and after possible PSU failure.

2+0 Non-Grid Redundant mode

This mode is identical to RAO mode - but without alert logging. This mode would typically be used for extremely high performance applications, like HPC applications. Users of this mode have decided that the need for more power exceeds the potential risk of downtime caused by PSU/grid failure.

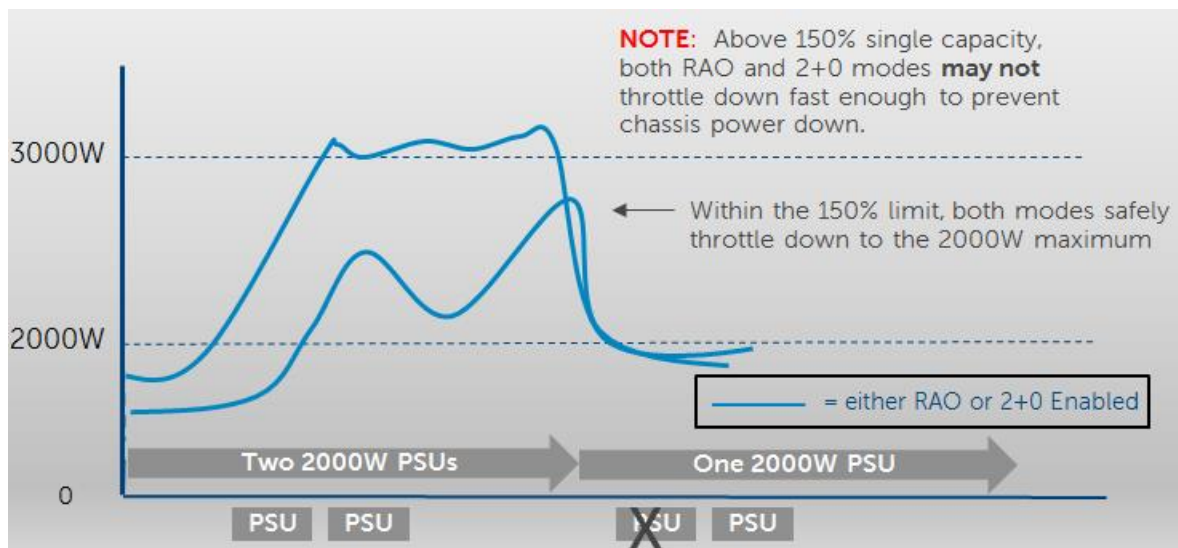


Illustration of throttling behavior for RAO and 2+0 modes on a dual 2000W PSU system

Examples of Configurations/Power modes

Many FX configurations will run in fully redundant (1+1) mode. The table below shows **some** of the possible configurations that do:

	FC430	FC630	FC630	FC830	FC630 & FD332
Node Qty	8	4	4	2	2x FC630
Chassis	FX2S	FX2S	FX2S	FX2S	FX2S
IOM	2x 1GbE PT	2x FN410S	2x 1GbE PT	2x 1GbE PT	2x FN410S
NDC	1GbE DP	10GbE DP	1GbE QP	2x 1GbE QP	10GbE DP
CPU	2x E5-2630L	2x E5-2630	2x E5-2670	4x E5-4620	2x E5-2670
DIMMs	8x 8GB	16x 32GB	16x 32GB	48x 32GB	24x 32GB
HD	2x SSD	2x 15k SAS	2x 15k SAS	2x 15k SAS	2x SSD
PCIe	1GbE DP	10GbE DP	1GbE QP	2x FC16 DP	10GbE DP
RAID		H730	H730	H730	H730
PSU	2000W 1+1	2000W 1+1	2000W 1+1	2000W 1+1	2000W 1+1

However, as illustrated in the next table, in some FX configurations the number of processors and the processor speed can have a definite impact power-wise and will require that you utilize Redundant Alert Only as your PSU redundancy mode.

	FC430	FC430	FC430	FC430
Node Qty	8	8	8	8
Chassis	FX2S	FX2S	FX2S	FX2S
IOM	2x 1GbE PT	2x 1GbE PT	2x FN410S	2x FN410S
LOM	1GbE DP	1GbE DP	10GbE DP	10GbE DP
CPU	2x E5-2630L	2x E5-2650	2x E5-2650	2x E5-2670
DIMMs	8x 8GB	8x 16GB	8x 32GB	8x 32GB
HD	2x SSD	2x SSD	2x SSD	2x SSD
PCIe	1GbE DP			10GbE DP
RAID				
PSU	2000W 1+1	2000W RAO	2000W RAO	2000W RAO

Power management on FX systems

Once an FX system is running in your data center, you can use Dell system management tools to monitor and control power.

You can use the CMC web interface or RACADM command line interface to manage and configure power controls, for example, you can:

- View the status and power use for the chassis, servers and PSUs.
- Configure power budget and redundancy policy for the chassis.
- Execute power control operations (turn on, turn off, system reset, power-cycle) for the chassis

The iDRAC8 Enterprise provides server-level management that monitors, reports and controls power consumption at the processor, memory and system level.

OpenManage Power Center (OMPC) is another useful management console that communicates with iDRAC and the Intel Node Manager to dynamically manage server power, and system input power capacity and to deliver group power management at the rack, row and data center level. It can also report the power consumption over time for servers, power distribution units and uninterruptible power supplies.

You can use OMPC (with an iDRAC Enterprise license) to set the power cap limit for your systems to limit the system power consumption. With its millisecond-fast power capping OMPC can also effectively increase the number of chassis in a rack by preventing tripping a circuit breaker. It can also be used to keep the total chassis consumption under 3000W, where servers can throttle quickly enough to avoid failure due to a loss of one of the 2000W PSUs

Thermal Considerations

The PowerEdge FX2 chassis includes highly efficient power supplies and fans, and a layout optimized for balanced cooling of the compute, PCIe, I/O, and power subsystems. The optimized hardware design coupled with sophisticated power-management capabilities built into the Chassis Management Controller (CMC), power supplies and iDRAC allow you to further enhance a power-efficient server environment. Designed with full closed-loop control, it modulates the fan speed to deliver adequate airflow to all of the components and ensure that they remain in their designed operating temperature windows.

Fresh Air options

It is possible to configure Fresh Air compliant FX platforms that are able to operate up to 40°C (104°F) for 99% of the year plus additional excursion operation up to 45°C (113°F) with humidity up to 90% (29°C maximum dew point). With Fresh Air systems customers can save on OpEx by running data center chillers fewer hours annually,

Each of the following FX servers can run in compliance with Fresh Air 2.0, using a subset of the available processors and with a number of limitations (i.e., PCIe SSDs not supported):

FC630 2cpu 1.8" drive configurations	FC630 2cpu 2.5" drive configurations
FC830 2cpu 1.8" drive configurations	FC830 2cpu 2.5" drive configurations
FC830 4cpu 1.8" drive configurations	

The FC 430 server and the FC830 4cpu 2.5"drive configuration are NOT supported

The online ordering pages have the full listings of specific compliant processors and limitations.

Conclusion

While the FX architecture allows you to consolidate more (and more varied) workloads into a smaller data center footprint, it has been designed in a way that takes the complexity and risk out of constructing and maintaining future ready infrastructures that can more efficiently handle increasingly demanding IT requirements. It does this by implementing innovative technologies and cutting edge automation that allows you to smoothly transition your IT operations into the emerging model of software defined data centers and converged infrastructure. Learn how Dell can help you do more with converged infrastructure at:

<http://www.dell.com/en-us/work/learn/converged-infrastructure>