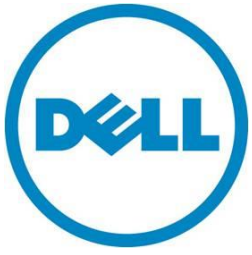
Simplifying networking on the journey to data center convergence

A Dell White Paper



Author: Manash Kirtania
July 2015

Contents

Abstract

In today's data-driven and demanding data centers, customers are actively looking to enhance their offerings using cost-effective infrastructure and solutions. Data centers are transforming in response to a data explosion which is driving newer deployment models, better speed and feed, virtualization, multi-tenancy, I/O convergence, workload optimization and simpler management solutions. At the center of this transformation is one key concept: simplicity. There is a clear trend towards organization convergence (IT silos to generalists) driven by shrinking or flat IT budgets, by a shift of workload onto the cloud and by strong cloud economics. As this shift occurs, more server and application administrators are becoming more familiar with networking concepts. Previously, the network access layer was controlled by network administrators. As access shifts to a different group of staff, operational simplicity is the key.

The next set of requirements is related to space/cooling, serviceability, uptime and IT productivity. Modularity can solve these requirements, but networking intricacies and the associated learning curve for data center operators are not easy challenges to overcome. Cable consolidation and I/O aggregation and switching are the fundamental networking pillars to maximize benefits from modularity and overall IT savings in general. The deployment of small or medium modular systems, rather than one large one, minimizes the fault domain. Again, this approach calls for networking simplicity. Convergence of traffic types provides the next bump to IT savings, but in the face of customer skepticism and lack of understanding, convergence simplicity is a must.

Dell offers a broad spectrum of enterprise server, storage, networking, and software solutions to make sure customers' data centers are future ready. Fundamental to each of these solutions is simplicity. In order to offer customers more efficient data center building blocks for their ever-demanding and changing workloads, Dell recently launched a 2RU modular PowerEdge FX2/FX2s platform with an array of blade servers, blade I/O modules and a direct-attached storage block for those platforms. In this white paper, we discuss how the FX2/FX2s products help customers to address the evolving needs of their data centers, with a particular focus on the networking aspects. We discuss the value proposition of blade I/O modules, which makes networking "easy as 1-2-3" for server administrators, application administrators and IT generalists by hiding all the networking intricacies and offering a super-charged networking for networking admins and specialists with comprehensive L2/L3 software features.

The networking modules primarily address simplicity, versatility, and performance, delivering the benefits of cable simplification while lowering the total cost of ownership compared to competitive offerings. The device is simple: install, deploy and forget. No training is required. The device is user-friendly for all users, including networking administrators pressed for time and server/IT generalists tip-toeing into networking. The device is versatile, offering a wide range of speeds and feeds, various levels of I/O convergence and a range of working modes supporting complete automation to complete customization. The device is powerful, packing advanced software features that provide high performance and scalability.

Dell PowerEdge FX2/FX2s platforms

The Dell PowerEdge FX converged architecture provides a simple, flexible, modular platform that can be easily customized to match the various workloads that data centers deploy. This architecture allows customers to take advantage of both the rack servers (which offer simplicity, smaller footprint, smaller fault domain and lower cost) and blade systems (which offer shared infrastructure, higher density and higher efficiency).

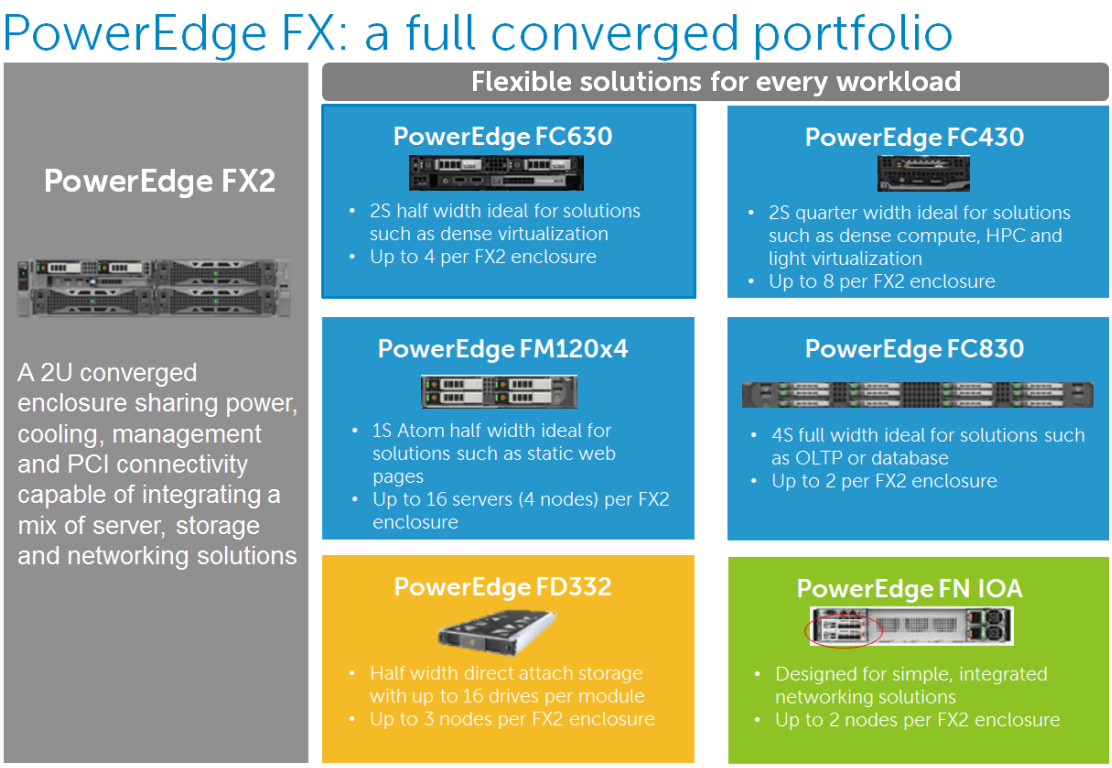


Figure 1: PowerEdge FX2/FX2s system and components

FX2/FX2s is the first incarnation of FX architecture, offering a 2RU modular chassis with building blocks of quarter-slot, half-slot and full-slot servers, I/O aggregation networking devices and a direct-attached storage (DAS) module (Figure 1). The FX2s chassis includes 8 x PCIe slots, connected to the servers, in the back of the FX2s chassis. The FX2 chassis does not include those PCIe slots. For a detailed description of these platforms, refer to the white paper "The PowerEdge FX Architecture Portfolio Overview [1]."

Dell PowerEdge FN I/O Modules

In the back of FX2/FX2s chassis, there are two networking I/O slots that offer redundant networking connections to the servers within the chassis. Customers can choose between several networking I/O modules for these slots (Figure 2):

- FN 410S I/O Module: offers 8 x 1Gb/10GbE internal connections to the servers and 4 x 10GbE SFP+ ports for external connections.
- FN 410T I/O Module: offers 8 x 1Gb/10GbE internal connections to the servers and 4 x 10G Base-T ports for external connections.

- FN 2210S I/O Module: offers 8 x 1Gb/10GbE internal connections to the servers and 2 x 10GbE SFP+ ports and 2 x 2/4/8G FC ports for external connections. Note that these FC ports can be converted to 10GbE SFP+ ports using software reboot of the I/O module.

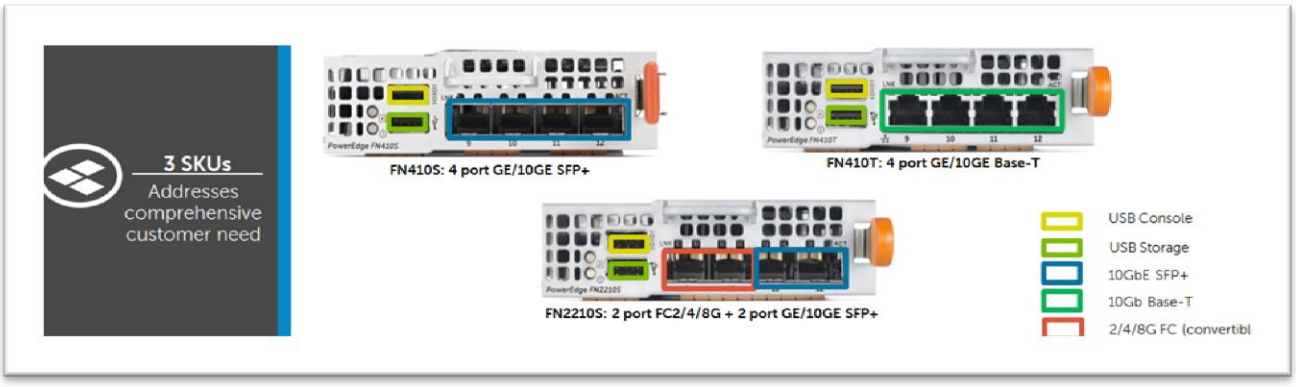


Figure 2: PowerEdge FN IOM flavors

We recommend same kind of FN I/O module in redundant I/O fabrics (2 slots) in the chassis. All the IOMs are powered by the Dell Networking Operating System 9.x, which has been battle-hardened over the last decade to address data center features and requirements. These IOMs support I/O aggregation capabilities, which today are primarily L2 features. The I/O modules come with an automated default setting to enable plug-n-play management. These modules are capable of supporting full L2/L3 switching features (L2 today and L3 in upcoming release in Q3 2015).

Why Dell PowerEdge FN I/O Module?

FN I/O module is a powerful device full of data-center class features which ultimately allows customers to lower TCO and protect investments for years to come. Figure 3 shows the high-level value propositions of FN IOM.

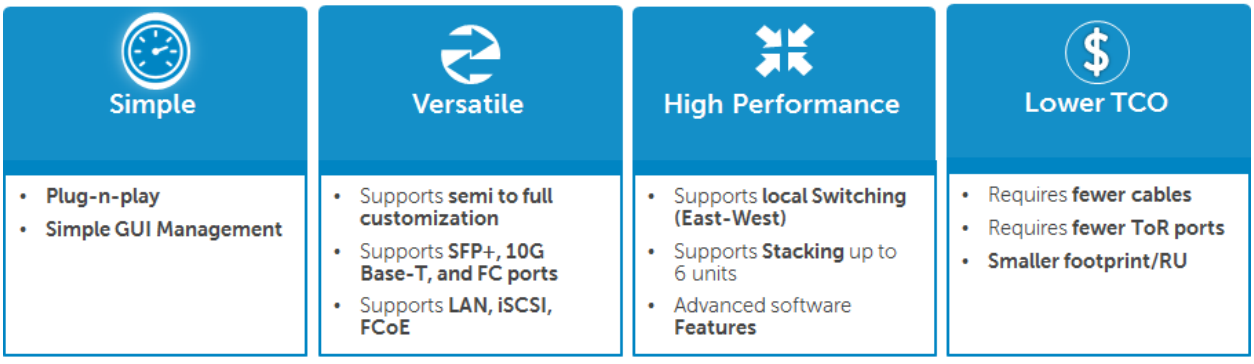


Figure 3: Value proposition summary of FN IOM

Simple

Simple plug-n-play mode of operation:

Given the combination of the data explosion and the shift toward virtualized infrastructures, data center operators today face the challenge of how to seamlessly manage a wide array of physical and virtualized resources, as well as a multitude of workloads that demand connectivity to the data. With simplicity in mind, Dell introduced FN IOM. The default plug-n-play mode means that customers do not need to configure any networking settings on the device. This mode ("Standalone Mode") was specifically built with the server or application administrator (or any administrator who wants to save time upfront, deploy and get benefits of cable consolidation and I/O aggregation) in mind. These users do not want to configure complex networking settings on the device; rather, they simply want server connectivity within and outside the chassis while doing I/O aggregation. This mode is perfect for the use case in which the customer simply turns on the device and the device works. The FN device automatically configures all VLANs on all ports and creates a LACP LAG using the uplinks which automatically carry the superset of all server-facing VLANs. No customization is required and the default settings meet the needs of most users.

Simple GUI management:
Customers have various deployments in their data centers, including typical LAN, converged LAN and iSCSI, and converged LAN and FC. Using CMC, today's user can change the VLAN configurations, upgrade software, and monitor status of such basic capabilities as ports status, speed and type. Using CMC and IOM GUI[1] interface, the future user can support all the above as well as various other features such as TACACS+/RADIUS security settings, port up/down, etc. Using OME[2], the user can clone FN networking configurations to a number of other FN devices, enabling a large scale deployment of FN devices.

Versatile

FN devices are shipped in plug-n-play mode. However, the FN device offers capabilities to support semi-customization to full-customization. Essentially, the FN device offers two levels of capabilities: I/O aggregation, which is offered by L2 switching functionality and full-switch[3], which provides full L2/L3 switching functionality. I/O aggregation can be done in several modes: standalone, VLT, stack and PMUX.

[1] IOM GUI is expected to be available in Q3CY2015.
[2] OME cloning capability for FN devices is expected in a future release.
[3] Full-switch mode is expected to be available in Q3CY2015.

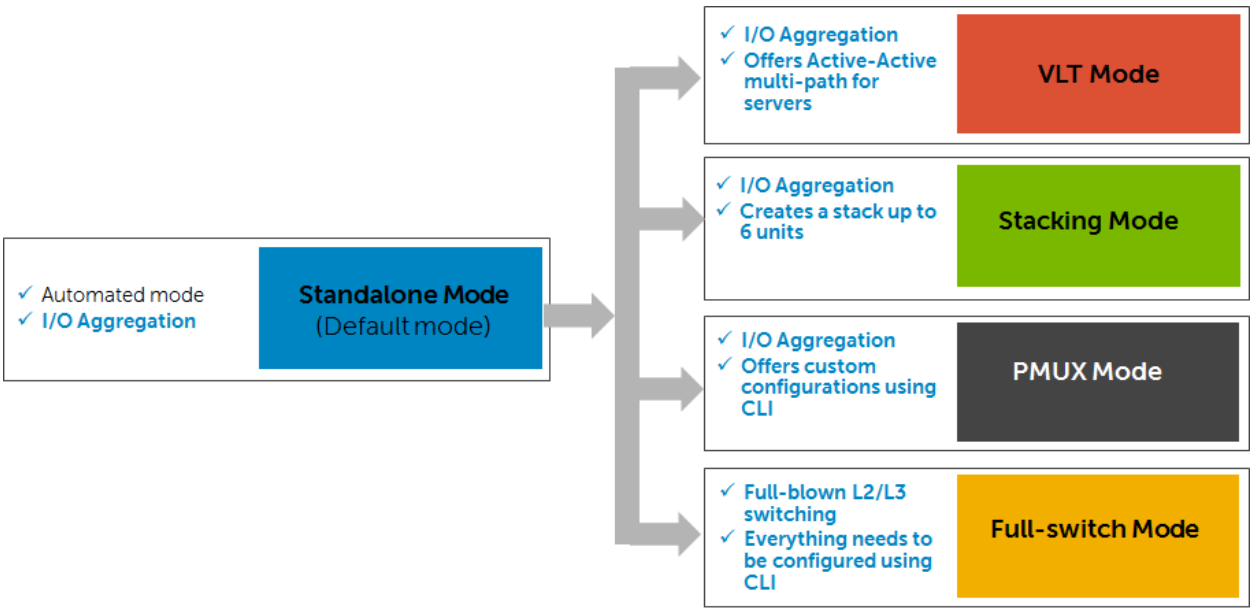


Figure 4: Various modes available on FN I/O Modules

In true standalone mode, the user uses the FN device without enabling high resiliency. When high resiliency is a concern, user can put two FN devices in the chassis in VLT (Virtual Link Trunking) mode. In this use case, the customer wants all the automated settings with both IOMs connected through an external port to form a VLT offering an active-active multi-path for the servers in the chassis. In this case, failover is protected and customer can individually upgrade IOM software.

To further simplify cabling among multiple chassis, reduce traffic to ToR switches, and thereby improve East-West I/O performance, customer can use FN devices in "stack mode". This allows customer to connect up to six IOMs in a ring or daisy-chain configuration. In this use case, the goal is to keep traffic within the stacked chassis instead of sending them to upper layer ToR and thereby handle east-west traffic within the stacked chassis. We recommend all the IOMs in the same slot across multiple chassis (not to exceed 6) to form a stack. The customer would then get two stacks across the chassis (discussed later in this paper). The two stacks ensure service continuity against any software upgrade disruption on a stack.

The "PMUX Mode" (programmable/CLI mode) allows the customer to access CLI on the I/O Aggregator. In this use case, the customer wants I/O aggregation capabilities but needs more customization.

Finally, we offer a full L2/L3 capable "Full-Switch Mode". This is targeted for the network administrators and the use case primarily includes provisioning L3 and advanced software capabilities such as QoS/ACL/OpenFlow on the IOM that are not available in I/O aggregation mode.

Using CMC and IOM GUI[4] interface, the user can provision various modes of the FN device, VLAN management and TACACS+/RADIUS security settings and monitor various states of the IOM device including the port status, port type and port speed. Note that the GUI does not support configuring various features in "full-switch" mode, as it is intended for the CLI users.

[4] IOM GUI is expected to be available in Q3CY2015.

These modes including a plug-n-play mode and simple GUIs to support these modes, offer a wide spectrum from complete automation to complete customization of FN IOM device to simplify deployment.

Supports a wide range of deployment scenarios with Converged LAN and SAN:
FN devices run the same software as another successful Dell blade switching product line, the MXL/IOA for M1000e chassis. This software supports all the converged software building blocks: iSCSI TLVs, Data Center Bridging (DCB), Fibre Channel over Ethernet (FCoE) Snooping Bridge (FSB), FCoE NPIV Proxy Gateway (NPG) and the F_port[5] feature which allows customer to connect F2210S directly to FC storage, eliminating the need for a SAN switch (Figure 5).

These fundamental software capabilities allow users to keep pace with their convergence needs:

- Small (entry-level) deployment with few servers in FX2 chassis directly connected to storage array
- Medium-size installation with dozens of servers where FN devices can simply carry FCoE traffic (in and out) to ToR FC switch
- Larger deployment which encompasses multiple racks of servers. FN devices (in NPG mode) can directly be connected to ToR FC switch.

Customers can re-purpose the FN devices to suit their growing needs:

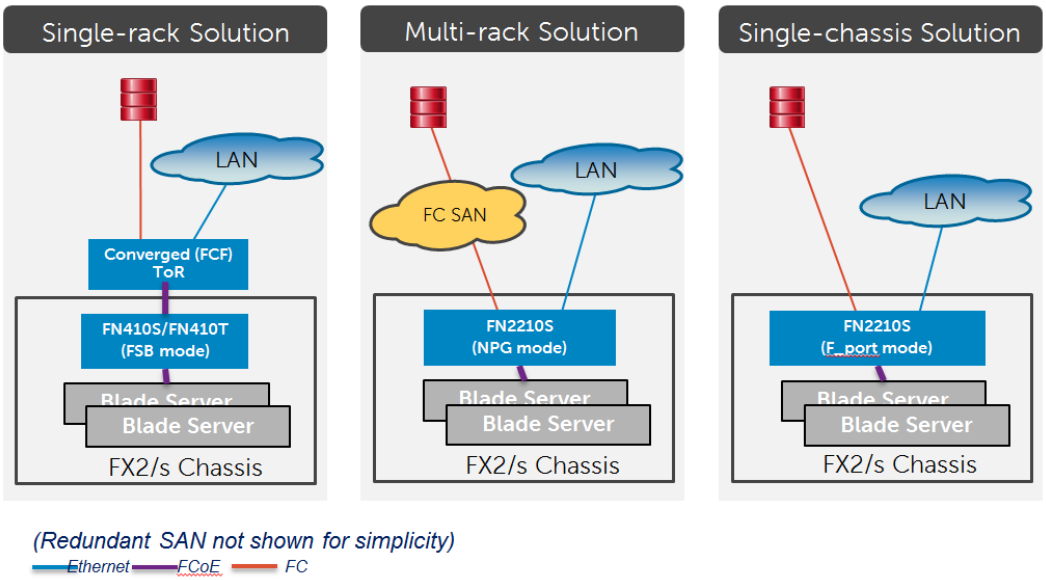


Figure 5: FCoE/FC deployments with FN IOM

The F_port feature enables direct connectivity between FN2210S FC ports and FC arrays. This enables a phased roll-out strategy. For small deployment, it makes sense to connect FC arrays directly to the FN2210S. This eliminates the need for an external SAN switch, ensuring simplicity and lower cost. When customer needs (hosts, target) grow, the same FN2210S device can be used in NPG mode to connect to a FC switch. Thus, the customer can achieve a wide range of converged options for future growth while protecting their investment.

Customers are looking at ways to minimize their CapEx and OpEx. Convergence addresses these goals by reducing the number of adapters, number of switches and number of cables. There are three versions of FN IOMs which address data center deployments with FX2/FX2s systems. The most common version is SFP+, which allows customer to address typical LAN, converged LAN + iSCSI, and

[5] F_port feature is expected to be available in Q3CY2015

converged LAN + FCoE (FSB) deployments. Next, the 10G Base-T version allows customer to address the above deployments in a copper environment. Finally, the 2210S version addresses converged LAN + FC (NPG) scenarios.

Customers have full flexibility of deploying FN I/O modules with Dell ToRs or 3rd party ToRs to address various deployment scenarios. Figure 6 highlights several deployment scenarios with an all-Dell solution.

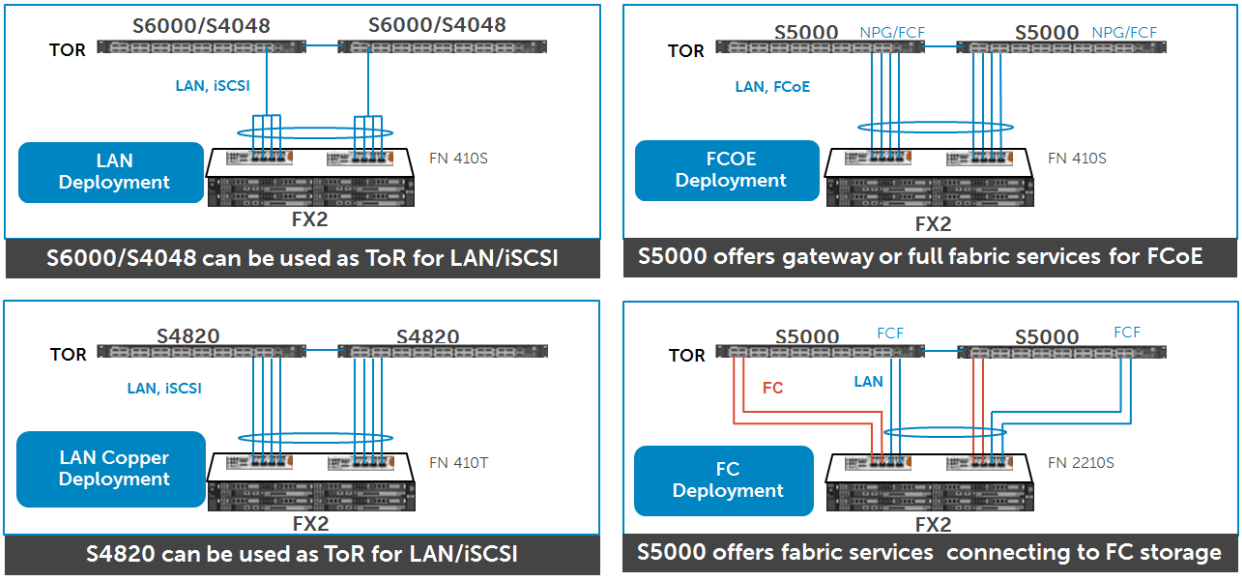


Figure 6: Few data center topologies using various flavors of FN IOMs

Use case 1: A customer wants to deploy a typical LAN scenario with iSCSI storage arrays

In this case, FN410S is connected to Dell S6000/S4048/S4800 ToRs which are connected to iSCSI arrays. FN IOM ports are configured in a LAG with the ToRs configured with VLT. Note that when a Dell EqualLogic PS array is connected to the ToRs, ToRs must be configured with VLT.

Use case 2: A customer wants to deploy a typical LAN scenario with iSCSI storage arrays but wants to leverage copper infrastructure

In this case, FN410T is connected to Dell 10G Base-T ToR S4820T. The iSCSI arrays are connected to S4820Ts and they can be put in VLT (or not) depending on the iSCSI array requirements.

Use case 3: A customer wants to deploy a converged LAN and FC scenario and use a converged ToR

In this case, FN410S is used and set in FCoE FSB (FIP Snooping Bridge) mode, transparently passing FCoE traffic between the server and the ToR. In this example, we use Dell converged ToR S5000. S5000 operates in one of two modes: NPG (NPIV Proxy Gateway) or Full Fabric Switch mode. Note that this situation can also arise in order to leverage 3rd party converged ToR which is capable of splitting FC/Ethernet.

Use case 4: A customer wants to deploy a converged LAN and FC scenario and wants to split FC and Ethernet at the chassis level

In this case, FN2210S is used, offering 2 FC ports and 2 Ethernet ports. FN2210S operates in NPG mode, splitting FC and Ethernet at the IOM. FC ports are connected to a Full Fabric switch rather than

a converged ToR. Here we use Dell S5000 as a full fabric switch that is connected to FC arrays. Note that this situation can also arise in order to leverage connectivity between FN2210S and 3rd party FC switches.

Supports deployment scenarios with non-converged LAN and SAN:
In many cases, customers may proceed with a non-converged deployment of the LAN and FC SAN scenario. FN IOM can be used for LAN and PCIe slots can be used for SAN. In this case, the customer can choose adapters to support a specific SAN that goes in the PCIe slot in the back of the FX2s chassis. Note that depending on the server type (e.g. FC430 or FC630), PCIe connectivity to the server varies. Figure 7 shows a scenario in which a user wants to use separate LAN and FC using HBA. With the FC630, two PCIe slots are available per server and the HBAs on those are connected to SAN A and SAN B FC switches.

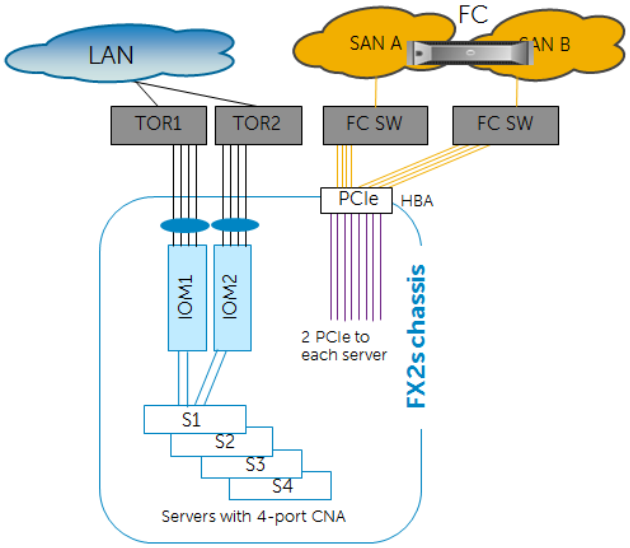


Figure 7: Non-converged LAN and SAN with FX2s chassis and FN IOM

High Performance

Advanced Software features and full-blown L2/L3 switch mode:
FN IOMs are fully loaded with data center-class software features that have been battle hardened over the last decade in the Dell Networking Operating System. One advanced feature to call out is the Virtual Link Trunking (VLT) feature which allows user to connect two FNIOMs in a FX2/s chassis to offer an active-active multipath from servers to the FNIOMs. This feature improves deployment resiliency by having dual brains, syncing all the ARPs, MACs, etc., between the FNIOMs, offering a faster failover should a link fail and allowing individual software upgrades on the FNIOMs.

The "full-switch" mode allows a customer to have full L2/L3 capabilities on these devices. If the customer is looking to deploy routing protocols on FN devices or on OpenFlow or to install various ACLs or QoS, this mode is needed. In this mode, the customer can enable L3 routing (e.g., OSPF v4/v6), multicasting features (e.g., PIM-SM/SSM), OpenFlow offering software defined networking based control, ACLs and QoS. The powerful full-switch mode allows FN devices to be transformed into a data center class L2/L3 switch just by changing the software mode from the default mode (standalone automated mode) to full-switch mode. Typically, full-switch mode is targeted at the network administrators who are comfortable with ground-up configuration using CLI. The automated

mode is targeted at server administrators who do not want to deal with networking intricacies and prefer the automatic settings.

Addresses growing East-West traffic using Local Switching:

A dominant trend in today's data centers is that east-west bound traffic is growing consistently. Server virtualization, multi-tier applications, distributed applications and cloud based applications are all driving this growth in east-west traffic. In the context of the FX2/FX2s chassis and servers, customers can keep server-to-server traffic local to the chassis using the L2 switching capabilities of FN IOMs (Figure 8). With standalone typical rack servers using NICs/CNAs and FX2 systems with pass-through modules, the customer would need to send all traffic to the ToR to get switched. This requires more uplink bandwidth and imposes higher latency compared to what customers can achieve with a FX2 system with FN IOMs. Note that Dell offers two Pass-Through modules for FX2/FX2s chassis: the 1GbE Pass-Through module (8x 1GbE server facing ports + 8x 1GbE external facing ports) and the 10GbE Pass-Through module (8x 10GbE server facing ports + 8x 10GbE external facing ports). These Pass-Through modules simply provide extensions of connection between server ports and ToR ports. They do not offer any switching capabilities.

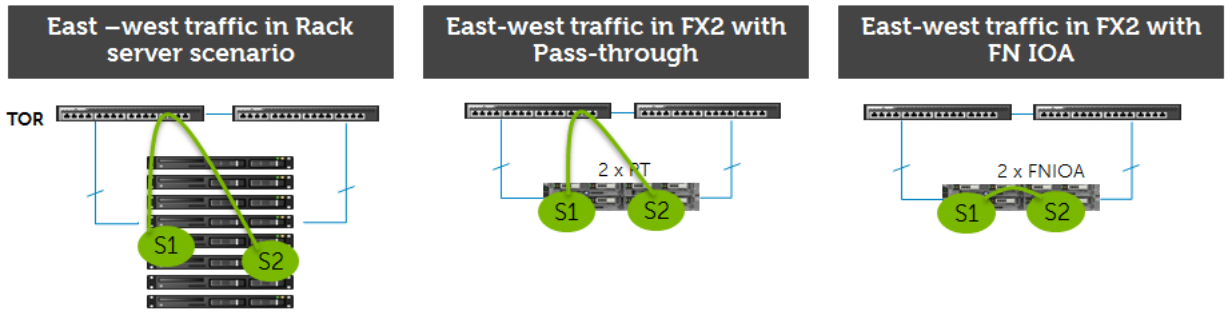


Figure 8: East-west traffic is better architected with FN IOMs in FX2

Due to the local L2 switching capability of the FN IOM, servers within the chassis can communicate via FN IOM. This server-to-server traffic does not leave the chassis, which helps in terms of lower latency and reduced uplink bandwidth which eventually translates into fewer ToR ports and fewer cables.

Addresses growing East-West traffic using Stacking:

FN IOM not only does the local switching of east-west traffic but also offers stacking of up to 6 units of FX2 chassis in a ring or daisy-chain configuration. These stacked units keep traffic local to the stack and do not send traffic destined for the servers within the stack to ToR to get switched. This reduces the number of ToR ports required. At least two stacks across the chassis are recommended when using the IOMs in the same slots. This enables higher resiliency and ensures service continuity in case of a software upgrade on a stack. Figure 9 shows an example in which two ports on each IOM are used to configure a ring stack. The other two ports on each IOM are used for uplinks.

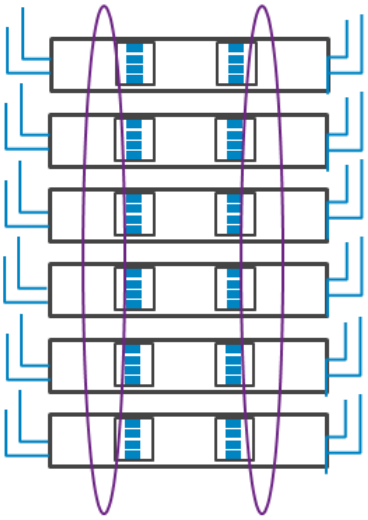
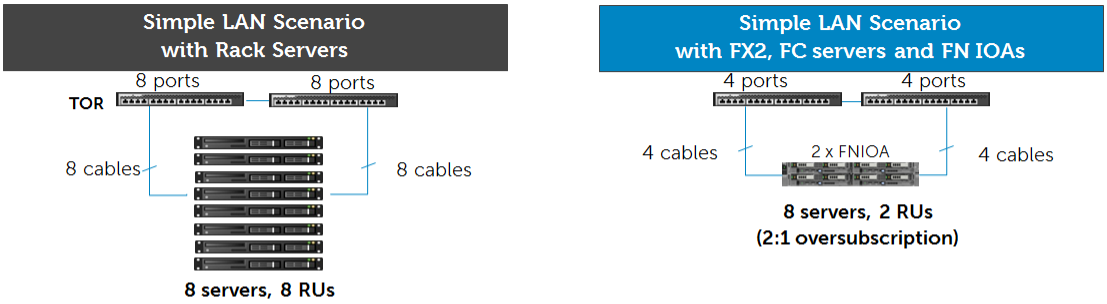


Figure 9: Stacking with FN IOMs

Lower TCO

Cable reduction, adapter reduction, ToR ports reduction and space/cooling cost savings:
A PowerEdge FX2/FX2s chassis can have up to 8 X FC430 quarter-slot servers, 4 x FC630 half-slot servers or 2 x FC830 full-slot servers. The FC430 servers are equipped with 2 x 10GbE CNA with one port on the CNA connected to each FNIOM. The FC630 servers can take either a 2 x 10GbE CNA with one port on the CNA connected to each FN IOM or a 4 x 10GbE CNA with two ports connected to each FN IOM. The FC830 servers can take up to two 4 x 10GbE CNA with up to four ports connected to each FN IOM. When customers have these servers with various workloads driving traffic on all eight internal ports on the I/O module, they have an opportunity to aggregate them before sending them to the Top of Rack (ToR) or next level networking device. Pass-through I/O modules do not aggregate traffic and simply provide link extensions between server CNA ports and ToR ports.

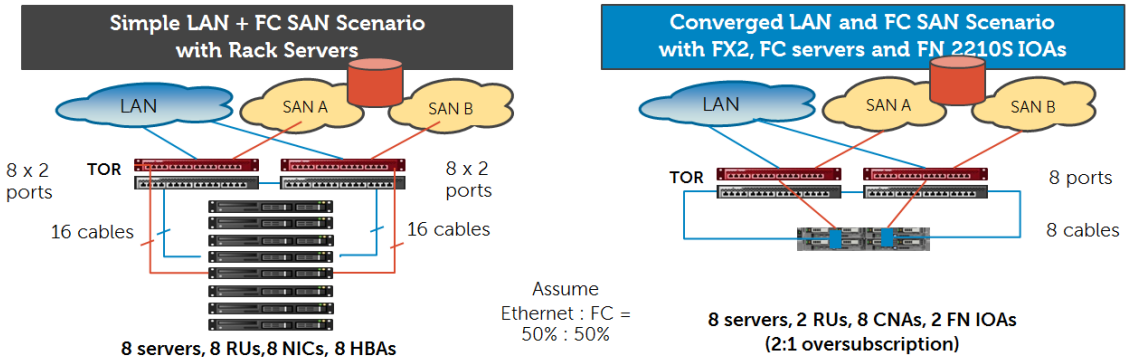
With I/O aggregation, customers can achieve an up to 8-to-1 cable reduction. To explore cable reduction further, consider a typical rack server scenario which has 8 x 1RU rack servers with 2 x 10GbE CNA vs. 1 x PowerEdge FX2 chassis fully populated with 8 x FC430 server blades. Assuming that 2:1 oversubscription satisfies bandwidth requirements, customers can easily cut down the cables by 50% (Figure 10). Note that this reduction is not limited to cables. Much bigger savings come with reducing rack units (and associated cooling costs) by 75%. Additionally, in most deployments, the sides where these cables are connected are ToR ports which are typically more expensive than the access ports. When customers can reduce the number of cables, they directly reduce the ToR ports costs. In the above case, customers can reduce the expense by 50%. If customers aggregate traffic using 8:1 over-subscription, the cost savings will be much higher.



	With Rack Servers	With FX servers and FN IOAs	FX solution
RU	8	2	75% less footprint which also implies less power, less cooling
Number of Cables	16	8	50% reduction
Number of ToR ports	16	8	50% reduction

Figure 10: FN IOM cuts down cables, rack spaces and ToR footprints

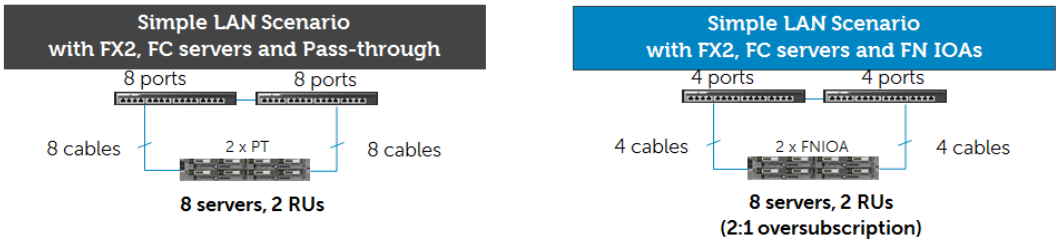
In case of FC deployments, these benefits become even more pronounced. A typical rack server in traditional LAN/FC deployment has a NIC and a HBA connected to LAN and FC SAN networks, respectively. Because FN IOM offers both Ethernet and FC ports, this dramatically reduces cables, adapters and ToR ports (both LAN and SAN). Figure 11 shows these two scenarios.



	With Rack Servers	With FX servers and FN IOAs	FX solution
RU	8	2	75% less footprint which also implies less power, less cooling
Adapters	8 NICs + 8 HBAs	8 CNAs	50% reduction
Number of Cables	32	8	75% reduction
Number of ToR ports	32	8	75% reduction

Figure 11: FN IOMs in converged LAN/FC scenario offer even more benefits

Comparing an FX with pass-through modules vs. the same with FN IOMs, many of these same benefits apply. While there are no savings with rack unit/space both taking 2RU, there are savings due to the reduction of cables and associated ToR ports using FNIOMs (Figure 12).



	With FX Servers and PT	With FX servers and FN IOAs	FX/FNIOA solution
Number of Cables	16	8	50% reduction
Number of ToR ports	16	8	50% reduction

Figure 12: Cost savings with FN IOM over pass-through modules in a FX2 chassis

Cost-savings comparison with competitors:
While FNIOMs are loaded with software and many other features, the purpose of this white paper is to prove that these features eventually help customers from an economic standpoint, compared to other competitive offerings.

The first scenario (Figure 13) compares Cisco UCS deployment and FX2 deployment using FNIOM for 128 servers with 2x10GbE CNA. Support of 128 servers using Cisco UCS architecture requires the following hardware:

- 16 x UCS 5108 Chassis (6RU each, 96RU total)
- 32 x UCS 2204XP FEX (4x10G uplinks)
- 2 x UCS 6296UP FI
- B200 M3 servers

Support of 128 servers using Dell hardware requires the following:

- 32 x FX2 Chassis (2RU each, 64 RU total)
- 64 x FN410S SFP+ IOM
- 2 x S6000
- FC630 servers

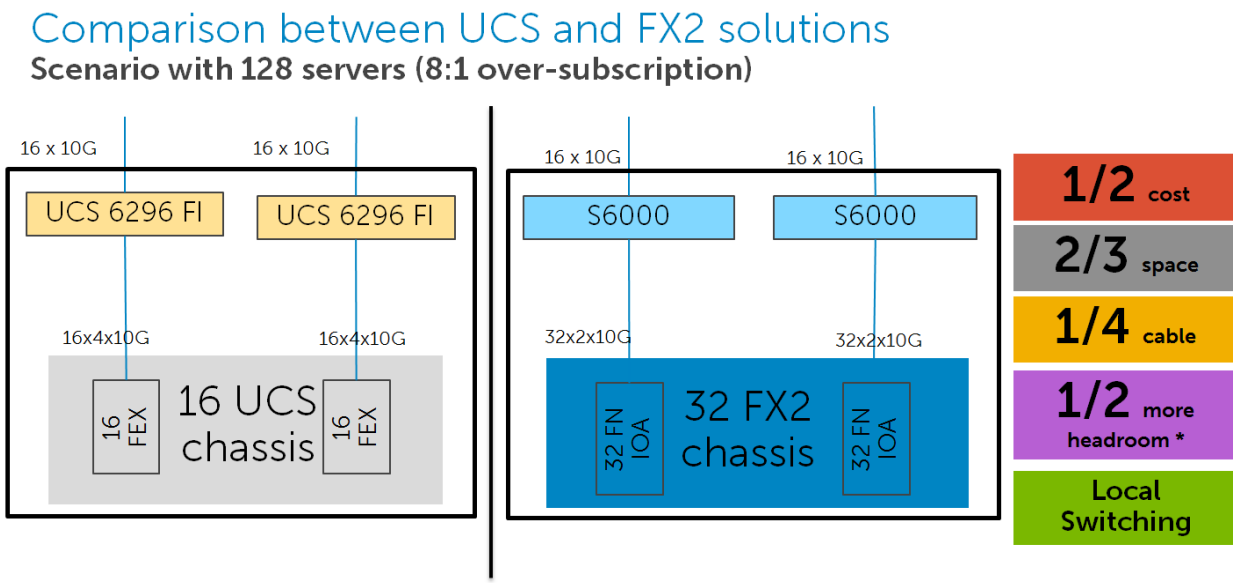


Figure 13: Comparing Cisco UCS vs Dell FX2 solutions for 128 servers

Comparing the list prices shows that the Dell FX2 solution in this example is 50% less expensive than the Cisco solution, requires 33% less RU and 75% less cables, and still has 50% headroom (unused ports on the ToR). Moreover, FNIOM offers local switching. FEX cannot locally switch traffic, rather passing them to FI to get switched.

The scenario in Figure 14 shows two alternatives that customers might face:

- FX2 chassis with Pass-Through modules connected to Nexus 2248PQ FEX connected to Nexus 5672UP ToR
- FX2 chassis with FN IOMs connected to Nexus 5672UP ToR

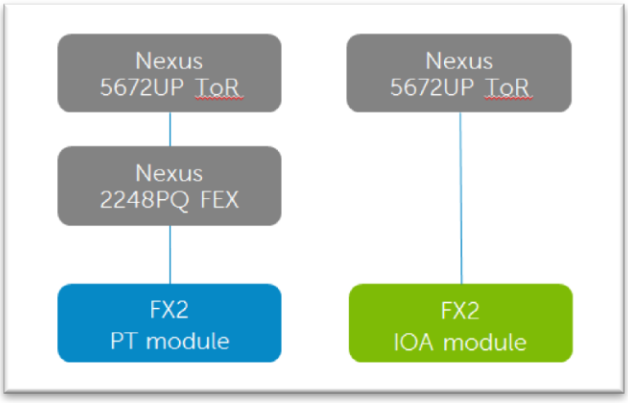


Figure 14: Pass-Through vs IOM with Cisco ToR

The above scenario is expanded in Figure 15 which shows the two scenarios in which customers have FX2 chassis fully populated with 4 x FC630 servers.

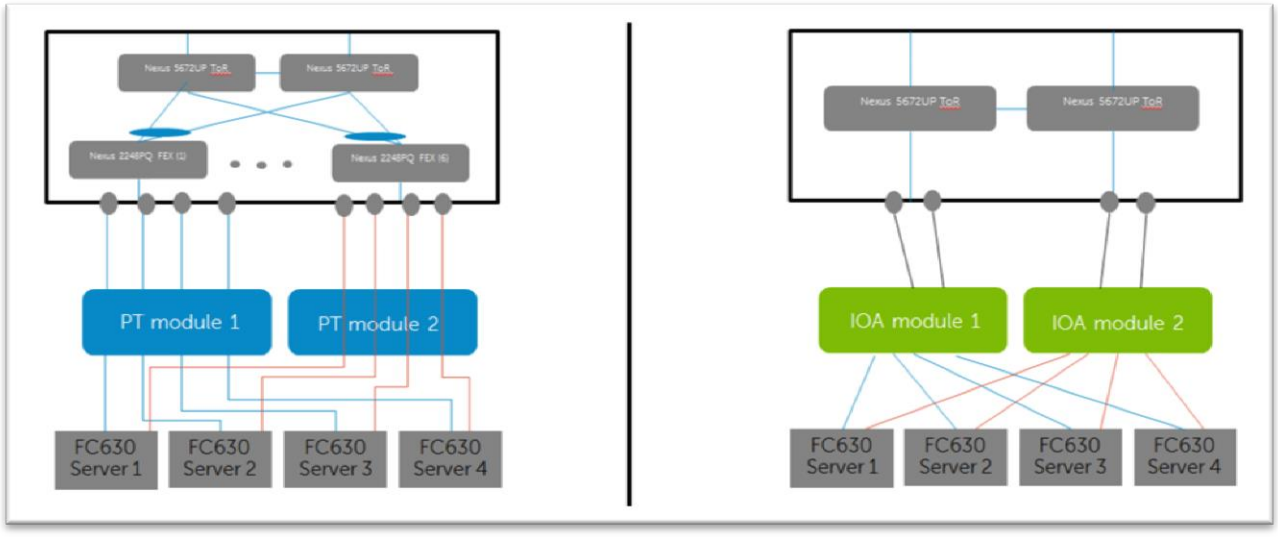


Figure 15: Pass-Through and FNIOM scenarios with Cisco ToR

For the purposes of comparison, this scenario considers the Cisco devices (as mentioned above) as black boxes and determines how much a server facing 10GbE port for each of these scenarios would

cost. Assuming similar oversubscription from the black box perspective, the costs are $502[6] and $560 per port respectively using current list prices. Populating these numbers with list prices for FX2 and components list, we find that the FN IOM scenario can save 19% of CapEx. This can quickly translate into a significant cost savings as customers deploy a large number of FX2/s chassis. The more servers customers have in the chassis, the better value they can extract from the FNIOMs. Figure 16 shows the savings a customer can achieve using FNIOM over Pass-through in a FX2 chassis. In this example, if customers have three or more servers in the FX2/s chassis, they can save ~30% CapEx. If they plan to deploy 1-2 servers in FX2/s chassis, they should consider future growth. FNIOMs can certainly protect their investment for many years to come.

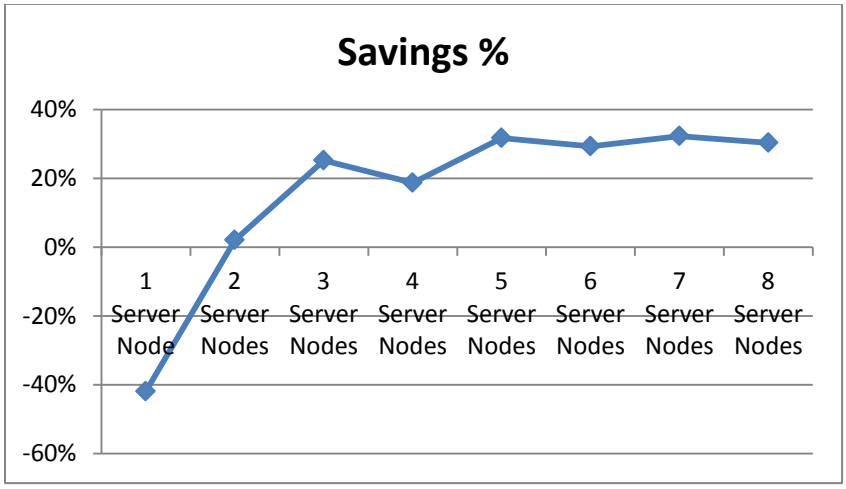


Figure 16: FNIOM savings over Pass-Through for the FX2/s chassis with servers

Summary

This white paper looked at how FN IOM can be a valuable addition to FX2 investment to address data center needs. FN IOM reduces TCO through reduction of cables and footprint in terms of rack unit as well as number of ToR ports. It also offers optimized east-west architecture, advanced software features and automation that seamlessly integrate with customer data center deployments whether used with other Dell switches or with 3rd party switches. Dell PowerEdge FX2/s with FN IOM is fully equipped to meet the needs of data center customers both today and for many years to come, addressing evolving workloads through a high performance, configurable and modular solution.

[6] List prices may change over time.

References

[1] The PowerEdge FX Architecture Portfolio Overview
http://www.dell.com/learn/us/en/05/shared-content~data-sheets~en/documents~fx-architecture-portfolio-overview.pdf

[2] Dell PowerEdge FX2 Solution Saved Time and Labor on Deployments
http://www.principledtechnologies.com/Dell/PowerEdge_FX2_deployment_0115.pdf

[3] VDI Performance Comparison: Dell PowerEdge FX2 and FC430 with VMWare Virtual SAN
http://www.principledtechnologies.com/Dell/PowerEdge_FX2_FC430_VMware_VSAN_VDI_0215_v3.pdf

[4] Oracle RAC Performance: Dell PowerEdge FX2S with Fluid Cache for SAN vs Competing Cache Solutions
http://en.community.dell.com/techcenter/extras/m/white_papers/20441035