MOOR INSIGHTS & STRATEGY

Software Defined Networking and Emerging Server Form Factors

Does the answer to network complexity live in network virtualization or rethinking the server chassis?

> *Virtualization on standardized hardware is a key IT trend that began in the late 1990's with the consolidation and virtualization of storage using SAN and NAS technology. Costs plummeted, customers had greater control and deploying/ reprovisioning became a more seamless and agile process. Then in the early 2000's, compute became virtualized on x86 platforms, bringing those same benefits to the processing front. Today, the final step of virtualization – network virtualization – is in vogue but as this technology comes into prime time, there may be differences in how it is deployed and how quickly customers move to it. The intersection of network virtualization with changing server form factors, most notably in the largest cloud customers, may present some interesting challenges.*

Networking Is at the Breaking Point

With the rise of cloud computing, virtualization and big data, businesses are looking for more flexibility and agility. The markets are moving faster and it is more difficult to keep pace if your systems are holding you back. Now that systems and storage are virtualized, networking needs to change in order to keep pace with the market.

In the world of networking, we're seeing the rise of the Software Defined Networking (SDN) concept as a way to tackle the problems of today's networks. Replace the congestion, cable sprawl and sluggish reprovisioning processes with a virtual overlay and automation. It sounds great, and people are rallying around the concepts. Some larger cloud data centers are already deploying it. But is that the only way? What if the solution (or the mitigation) is in the server chassis and not the network?

In this paper we look at some ways that companies are trying to tackle this problem that don't necessarily create a new virtual network on top of their existing physical networks. Some of these challenges can easily be solved with SDN, but there are other ways that customers also need to consider as they look at reducing their networking burden.

As customers investigate SDN they need to understand how chassis form factors can alter the potential ROI of such a solution. The chassis/form factor choices here do not eliminate the need for SDN, but they can have an impact on the benefits.

Server Sprawl is Causing Many of the Problems

As we look at the cause of today's networking problems, they don't always tie back to networking equipment. Actually, server hardware and workloads have been driving the complexity of networking. As the server cost came down, component sizes decreased and virtualization took over, we saw three big dynamics collide: more servers were being deployed, those servers were denser and each of those servers was running a larger number of virtual machines instead of a single workload, this necessitated more physical network connections. With virtualization, customers liked to have separate communications channels for data networking, storage networking and even live migration. Suddenly the typical rack went from 10-20 servers with a pair of network connections each to 30-40 servers with up to 8 or 10 physical connections each. What were once 20-40 physical network cables per rack has exploded to well north of 200-300 cables; the networking was having a hard time keeping up. Throwing more network equipment at the problem both helped and hurt the situation. The primary problems were in traffic management (bottlenecks), load balancing, physical provisioning and change management.

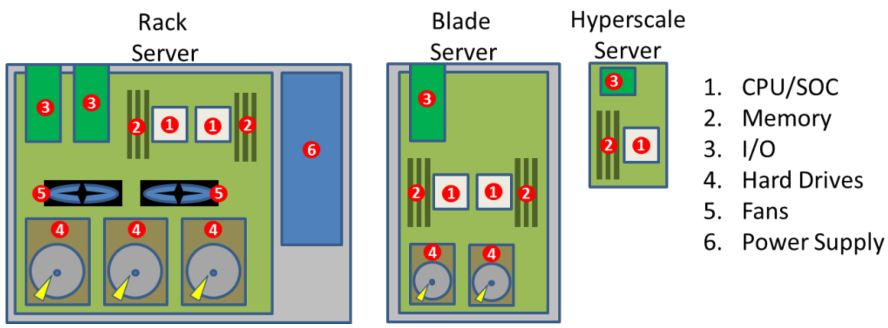
SDN promises to remove these pain points by making it easier for customers to manage their networks, adding a virtual layer on top of the physical network. However, that is not the only solution to this problem, and one of the biggest challenges for SDN – having to either completely deploy at scale or manage multiple network methodologies – means that there may be other ways for customers to try to minimize the issues that they face.

At the server/physical layer there are network consolidation/fabric solutions that, while they do not replace SDN altogether, may actually bring some benefit for the near term. Those can take some of the wind out of the SDN sails by reducing the magnitude of some of the SDN benefits (while not completely eliminating the need.) Anything that is done at the rack level today to consolidate or create a fabric will, in essence, stretch out the payback of an SDN solution by reducing the benefits and increasing the amount of time required to recoup the investment. The more mainstream these technologies become the greater the probability that they will impact the adoption rate of SDN.

Defining the Server

Before jumping into a discussion of how the server form factors are changing, it is first important to define what a server is, as that interpretation may also be in flux according to some. For purposes of this paper we will use the following definitions.

Server – A collection of processor/SOC sockets that share a common main memory pool with full memory coherence. Large cloud data centers are predominantly deploying 1P (single socket) servers, while traditional IT loads are being run mainly on 2P (dual socket) servers. The primary pieces that help define a server are the CPU, memory and I/O subsystem; outside of these three components, all other components can be pooled or shared at the chassis level depending on the form factor.



Chassis – The chassis is a physical enclosure that will hold one or more servers. Standard rack servers are 19" wide and generally 2 rack units high (2U) with 1U and 1/2U (twins) being the next most popular chassis form factors. (A rack unit is an EIA standard, 1.75" high.) Blade chassis will hold several servers like books on a bookshelf; the servers slide into the chassis and share some common components like fans and power supplies. Hyperscale servers are card-based and typically installed into a highly proprietary dense chassis. These servers share a much higher level of integration than blades. Tower servers, popular in emerging economies and remote/branch office locations have been on a continual slide down in share. Because they are often deployed as standalone products, there is little need to discuss them in the greater context of network or server consolidation.

(L to R) Rack mount server (Dell PowerEdge R720), Blade Chassis (IBM BladeCenter), Hyperscale server (HP Moonshot 1500)

Overall the future trend for servers focuses on greater density and more modularity as the data center continues to drive for greater agility and faster deployment.

Server-based Solutions

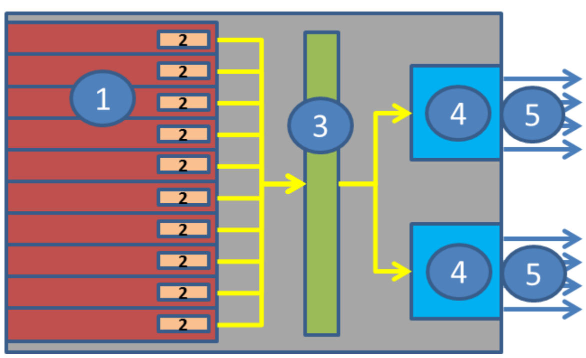
If servers are the problem, then perhaps servers can also be the solution. Here are a few ways that customers have been tackling deployment and consolidation issues that do not require changing the network overlay:

Blade Servers

Blades have been deployed in some of the largest data centers and have taken approximately 17% of the market. While blade growth had flattened out dramatically and is now actually shrinking, blade-based servers are already solving some of the networking problems for customers who have chosen to deploy.

Cisco, Dell, HP and IBM all have blade solutions but each is proprietary to its own company. These systems typically have a midplane (mezzanine) that connects all of the blade servers in a chassis to the networking on the back. This gives an instant aggregation point for the servers and can reduce the dozens of cables needed down to only a handful. If one was to put 3-4 blade chassis in a rack they may have 30 or fewer 10GbE (or soon, 40GbE) connections coming out of the rack. With such a consolidation, many of the traditional challenges are mitigated. Unfortunately, at the server level each blade includes its own network controller and typically an I/O slot for a second controller, increasing the size of the blade and adding extra cost.

Complicating the situation is the fact that every blade chassis comes with its own management software to handle the distribution of I/O to the servers in the chassis. This adds an extra level of complexity for management, especially if the blade networking modules are not OpenFlow compatible.



Blades (1) each have their own network adapter (2) but aggregate network traffic through a common backplane (3) and then utilize shared network switching (4) for uplink (5)

The consolidation factor for blades has already reduced some of the pressure on data center networking. Those customers that have standardized on blades are generally using blades for the majority of their infrastructure now. But in looking through the current internal blade networking modules from the leading vendors, while there are promises of support for OpenFlow, internal OpenFlow compliant switch modules are not mainstream. Even when they are available, there is an existing install base to consider. This means customers who have deployed blades will not get the full SDN benefits without significant switching costs and will create extra levels of manageability that they must address. Deploying SDN may still make sense for these customers but there is a tradeoff on the amount of abstraction the customer is willing to accept in order to get the SDN advantages.

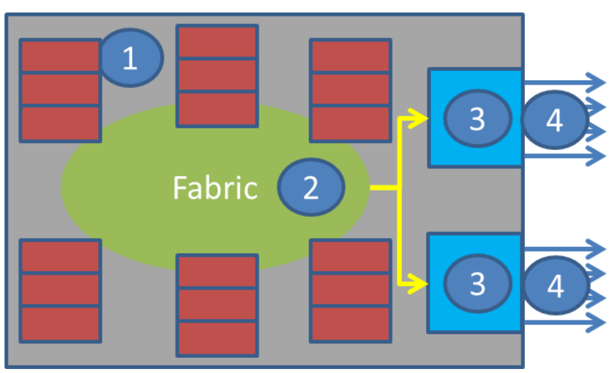
In looking at blades, the natural question might be "what about 'twins'?" Twins are smaller servers, typically 2 servers in a 1U rack height chassis. However these typically have all of their I/O ports duplicated on the back of each server, so there is no aggregation with these products. Twins do not solve any of the networking problems; they only address server density issues.

Hyperscale Servers

Hyperscale servers, highly dense “fabric in a chassis” products, solve many of the networking issues by consolidating/aggregating/managing network services for all of the miniaturized servers that are contained in a single chassis. With hyperscale servers dozens to hundreds of servers are consolidated down to a single chassis, allowing for extreme density. These products, from companies like AMD, Calxeda, Dell HP and others are designed mainly for large cloud data centers, typically handling web or big data workloads.

Hyperscale servers tend to be chosen for more homogeneous environments where blade servers are more configurable than hyperscale servers and tend to be used with heterogeneous environments.

Hyperscale servers are different from blades in that they don’t rely on a midplane technology and instead create a 3D fabric inside of the chassis. This fabric connects smaller server cards that contain a subset of the server components (typically memory, I/O and compute) while the storage and network services are typically handled through the fabric (that is shared across all of the servers.) In contrast, blades include their own I/O subsystems and have discrete networking components and drives. By reducing the components and sharing at the chassis level, hyperscale servers can greatly reduce cost, power and footprint.



Microservers use server cards (1) tie into a fabric (2) and utilize virtualized network connections; an integrated shared switch (3) aggregates traffic and provides uplink (4)

Hyperscale servers usually feature integrated L2 switching through their fabric. This allows all of the servers to handle their east-west traffic within the chassis through the fabric, minimizing the amount of traffic outside of the chassis.

Interestingly, these hyperscale servers will generally be deployed in large cloud data centers, the same places that are using SDN today; so to some degree hyperscale customers may have to choose between SDN or hyperscale servers, they probably won’t use both unless the Hyperscale supports OpenFlow.

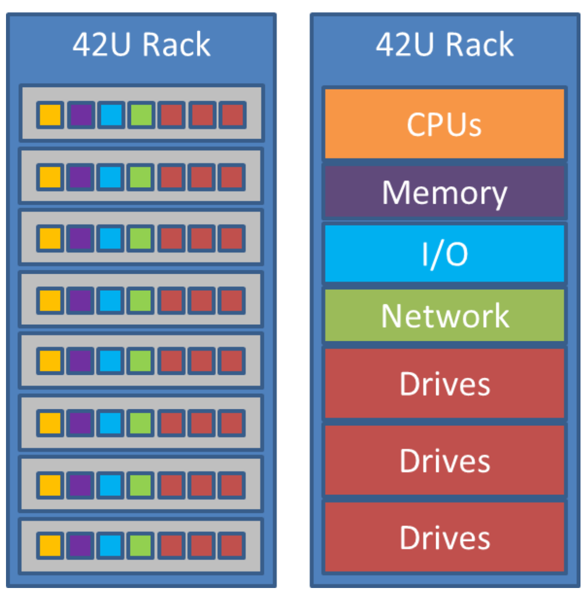
Many of the hyperscale server vendors claim that utilizing their products allows one to remove the top of rack networking and save capital deployment costs in that manner, but as we see later in this paper, that simplistic view ignores some of the challenges networking faces that brought the industry to the point that it is at today.

Disaggregated Racks

Recently Intel announced a disaggregated rack strategy with future products built around silicon photonics technology. This "rack-based disaggregation" will allow a single rack to act more like a large-scale system, where CPU, memory and I/O can be separated and added in increments that make sense for the applications running in the rack.

As the servers in a rack are all being fed by a 100Gb/s silicon photonics pipeline for the network, there is little need for more than a handful of cables coming out of the back of the rack to provide the networking capabilities for the full rack based on today's workloads and those of the near future. Assuming that all of the actual networking will have to be based on a virtual model (a pair of redundant high-bandwidth physical optical cables will provide the network connectivity), then there will need to be some management built down into the rack to provide the "disaggregated" network services. This solution would allow the east-west traffic to remain inside the rack, or potentially even distributed in a rack to rack basis depending on how the technology is implemented (there is ~300M limitation on silicon photonics and ~100M limitation on Ethernet CAT-6 already today.)

Because this technology will need to have some chassis/system level management tools, there will be yet one more management console, with no "single pane of glass" to manage the network.



Networking-based Solutions

When looking at a hardware-based solution, while consolidating servers can definitely help minimize the networking problems through aggregation or architecture, the other way to address networking pain points with hardware is through the networking components themselves. Clearly SDN tackles the provisioning/agility issues for networking, but some of the other large issues that customers face (that are also SDN pain points) can be addressed through other means today that do not require re-architecting the network. These are more short-term tactical answers that address one or two points and do not represent the type of change that SDN can bring in agility.

Cost

Tackling the cost factor has been a key driver for many SDN investigations. The thought of using software to manage your network instead of relying on more expensive hardware from the likes of Cisco or Juniper means that one can reduce capital expenditures. For those customers looking to merely reduce the cost of their networking, a host of white box companies like Accton can provide 10GbE network switches at a fraction of the cost traditional vendor solutions. Customers might not have the same manageability and feature set, but there will be a tremendous cost savings. However, this does not address any of the other issues. But, if budgets are the only concern, the trend towards white box networking can alleviate that pain.

Bandwidth/Cabling

If limited bandwidth or cable sprawl is a key pain point, the affordability threshold for 10GbE will be crossed this year as more customers start deploying 10GbE over its lower bandwidth cousin 1GbE. This means the typical server today, with 8-10 1GbE cables can be serviced by a pair of 10GbE cables (for redundancy), bringing greater bandwidth and a significant reduction in cabling. The ability to deploy this on a rack-by-rack basis means that customers can either deal with the newer technology when bandwidth/cabling matter or as new racks are being deployed. Best of all, 10GbE can live together along with the 1GbE infrastructure under the same management framework. Reducing the cabling and increasing the bandwidth. However, cabling alone will not address the other issues that SDN can resolve. If customers require different physical networks for storage, communications and VM migrations, then moving to 10GbE will not necessarily eliminate as many cables as hoped.

Removing Tiers

Typical data center architectures have 3 tiers, with a set of core switches, end of row switches for each row of racks and top of rack switches to service the devices in each rack. One way to reduce cost and complexity is to remove the top of rack switching and route the devices in the rack directly to the end of row. Typically each rack has a pair of TOR switches for redundancy. Each TOR switch typically is running 10GbE today, and typical uplink is 80-120Gb, meaning

8-12 10GbE cables per switch. This is 16-24 total cables per rack today and as 40GbE becomes more mainstream this number could potentially come down.

With standard rack servers, the ability to remove top of rack switching is not an option. However, with blade and hyperscale servers already consolidating within the rack, the cabling to the EOR switches becomes more reasonable. With 100 meter distances for Ethernet, top of rack switching could be eliminated, sending the chassis directly to the EOR switches. This would reduce capital expenditures by $50-80K per rack, but would increase cabling management and cost in the process. From an operational standpoint there would be fewer physical devices to manage, which would reduce operational costs. However, with fewer tiers, there would be less of an ability to fine-tune some of the lower level devices (which may or may not be something customers are doing today.) Additionally, failures can become more widespread as the physical domain spread becomes larger with the removal of TOR switching. Troubleshooting is both simplified and complicated as there are fewer layers to check, but each layer has a larger reach when something goes wrong.

As networking bandwidth becomes more robust and 40Gb Ethernet becomes more mainstream, the ability to reduce tiers becomes more workable (though still not necessarily more affordable.)

Management

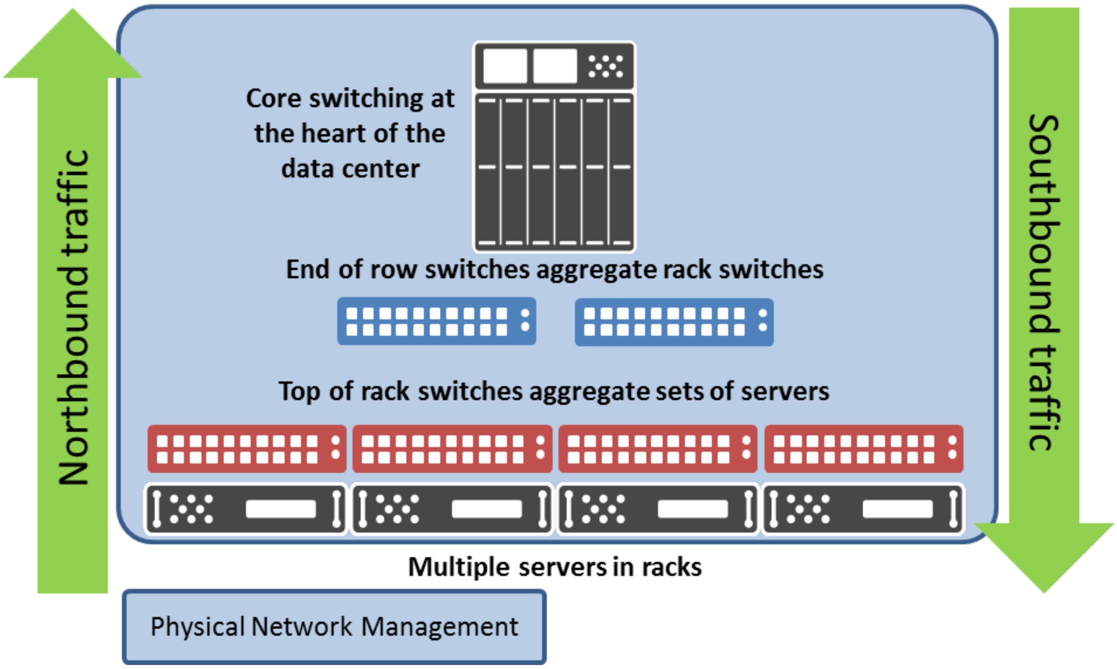
We continue to stress management as a key consideration when looking at SDN versus traditional networking or some type of server aggregation strategy. While the proponents of SDN describe the benefits of having a virtual network, they often gloss over the fact that there is still a physical network underneath that also needs to be managed simultaneously.

One of the challenges of SDN today is that there are two layers that must be managed: a physical layer and a virtual layer. This is similar to an airline that has to deal with two separate logistics systems which are joined together, yet independent (one to handle the passengers/baggage and one to handle the planes themselves.) While one system is concerned with safety inspections, fueling and maintenance, there is a completely different system that handles seating, meals and customer status. Some functions are tightly inter-related (like fuel and passenger load) while others (safety inspection and entertainment choices) are completely independent. Networks are no different; there are some aspects of the physical and virtual network that are inter-related and others that are completely independent.

Now add into the management mix some kind of aggregated system at the other end. Be it blades, hyperscale servers or some other type of shared hardware solution, if OpenFlow (the SDN protocol) can’t reach all the way down to those integrated switches or that chassis fabric, there is an island on the other end that needs to be dealt with. And typically it is dealt with through a proprietary management tool from the vendor.

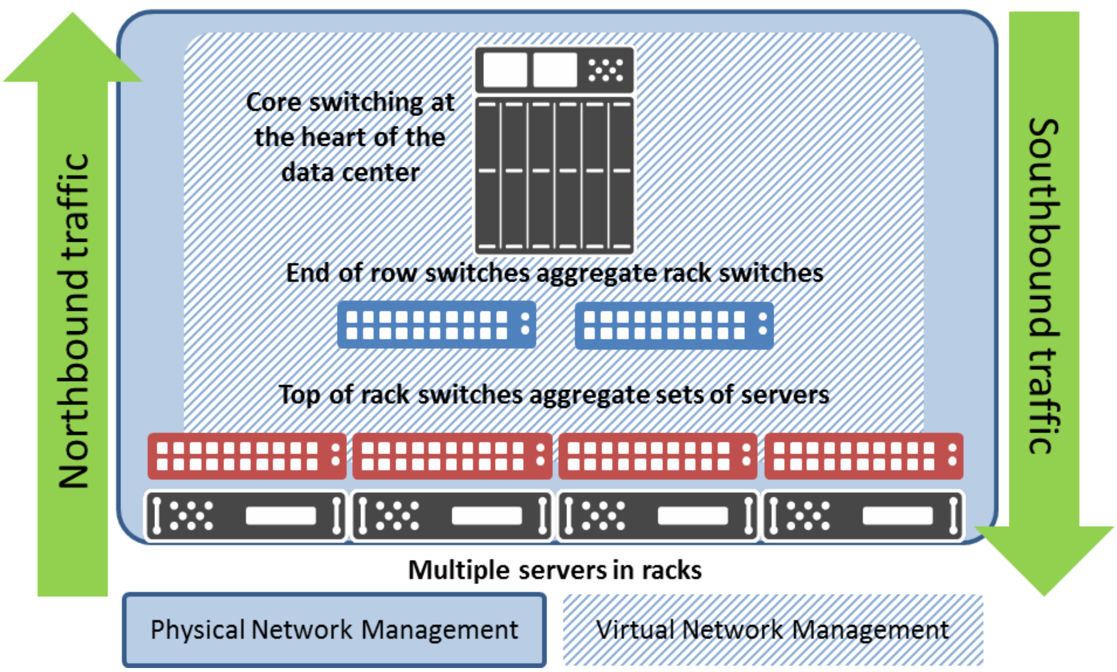
In a typical environment, the management console covers all of the layers:

Network Management

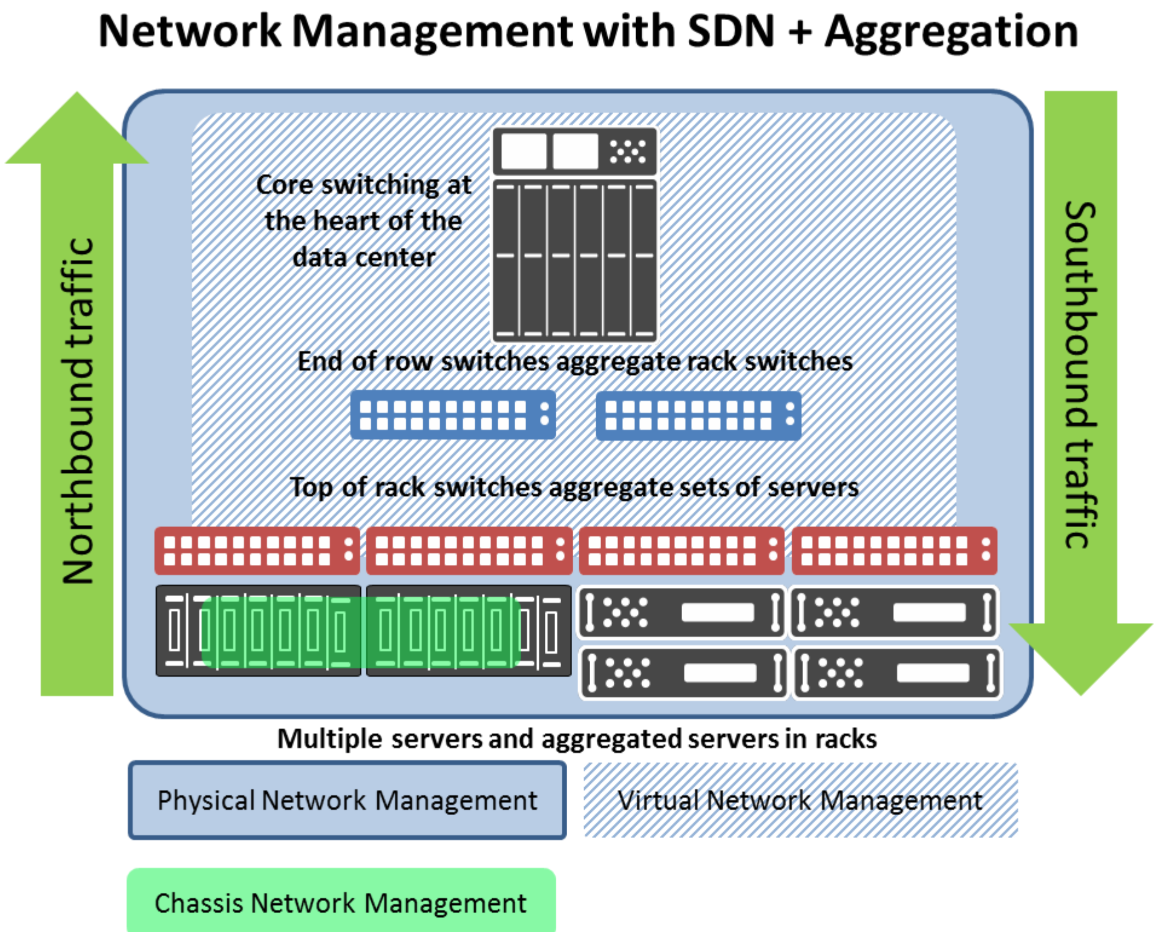


In an SDN environment, the SDN management tool manages the virtual layer and the standard management tool manages the underlying physical layer:

Network Management with SDN



Adding another fabric at the server level that is not SDN/OpenFlow compliant means an even more complicated management scenario with three different tools involved:



Bandwidth at the Server and Chassis Level

In looking at the different scenarios, we see that there are several considerations around bandwidth, cabling and density that need to be considered in determining the form factors for deployment.

	Chassis Per Rack	Servers Per chassis	Servers Per Rack	Cables per Rack	Cables Per Server	Chassis Internal Bandwidth	Chassis External Bandwidth	Bandwidth per Server (external)
Standard rack servers – GbE	42	1	42	336-420	8-10	N/A	N/A	8-10Gb
Standard Rack Servers – 10GbE	42	1	42	84-168	2-4	N/A	N/A	20-40Gb
HP Blade Chassis	4	16	64	32-48	.5-1.3	320Gb	480Gb	5-7.5Gb
HP Moonshot	9	45	360	108	0.3	90Gb	120Gb	2.66Gb
AMD SeaMicro	4	64 or 256	256 or 1024	48	0.047-0.188	2048Gb	160Gb	0.47-1.875Gb

Clearly, as densities increase, bandwidth per server decreases.

While this is an important measure for servers, keep in mind that over the years workloads have begun to change for servers. With the rise of more connected applications, more enterprise application integration and more services-based computing resources, servers are generating significantly more east-west traffic rather than mainly north-south. For traditional workloads that run on Xeon or Opteron processors, both N-S and E-W traffic are required so bandwidth needs to be higher. With large numbers of VMs being hosted, significant network traffic is being generated. For hyperscale workloads there is less dense compute happening (both in terms of the type of work and the amount of CPU throughput); much of the east-west traffic is handled within the chassis and storage is localized, so there is less need to step outside of the fabric.

For hyperscale servers, the cost of infrastructure is lower, along with the power, so while bandwidth is reduced, it is generally in proportion to the other factors. Traditional servers, which are more often running a large number of VMs, will demand the higher bandwidth, hyperscale servers, when running on smaller core CPUs will stand up fewer VMs per core, so again, the proportional math works.

In relation to SDN, managing bandwidth (which is one of the many reasons for deploying SDN) is minimized when dealing with a hyperscale server infrastructure because the workloads are less bandwidth-dependent. For servers running a larger number of complex VMs, having more bandwidth (and managing the bandwidth and access) is far more critical than in the hyperscale server environment where the chassis is somewhat bandwidth challenged (externally) and the workloads are not as bandwidth driven. Those servers are much more in need of SDN because of the complexity and throughput requirements.

We believe that hyperscale servers (in these dense chassis) are not a great target for SDN solutions today. The manufacturers' claims of being able to remove layers of networking at the top of the rack probably don't hold as much weight because of factors mentioned above. So while these servers may not need SDN, they still need something at the top of the rack.

The Bottom Line

As the market looks to SDN as a solution to the networks' inability to keep pace with today's business needs, there are other forces at play within the server space that will have an impact on the potential viability for SDN.

Customers with standard data centers (rackmount servers) are the best equipped to take on SDN with the least amount of management complexity (beyond the fact that you are now managing two separate levels of networking, the physical and virtual.)

Customers looking to deploy blades, hyperscale servers or some other aggregated/disaggregated form factors should consider the management complexity before tackling SDN. The greater the percentage of these products in the overall data center mix, the more customers need to take them into consideration when making SDN decisions.

But for those looking to fundamentally re-invent their IT infrastructure with alternative strategies like hyperscale compute platforms, investigating the OpenFlow support of these platforms will be critical before making a move to SDN.

We cannot stress enough that management is a key component of the entire solution and something that should not be quickly dismissed. With each new layer that is added to the data center, the ability to manage across layers becomes more critical.

Again, as we have said before, SDN is a promising technology, but there are still many pieces that are in flux at this moment. Until more of the market is sorted out, investigation of SDN should continue, but deployment may not be the best strategy at this moment unless all of the pieces that are required for your data center are available.

In the mean time we recommend that those investigating SDN factor in any chassis changes and dense form factors to ensure that they have the proper financial metrics and the proper systems management visibility during the project scoping.

Important Information About This Paper

Author
John Fruehe, Guest Analyst at Moor Insights & Strategy.

Reviewer
Paul Teich, CTO and Senior Analyst at Moor Insights & Strategy

Editor
Patrick Moorhead, President & Principal Analyst at Moor Insights & Strategy.

Inquiries
Please contact us here if you would like to discuss this report and Moor Insights & Strategy will promptly respond.

Citations
This note or paper can be cited by accredited press and analysts, but must be cited in-context, displaying author's name, author's title and "Moor Insights & Strategy". Non-press and non-analysts must receive prior written permission by Moor Insights & Strategy for any citations.

Licensing
This document, including any supporting materials, is owned by Moor Insights & Strategy. This publication may not be reproduced, distributed, or shared in any form without Moor Insights & Strategy's prior written permission.

Disclosures
Moor Insights & Strategy provides research, analysis, advising, and consulting to many high-tech companies mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this documented.

DISCLAIMER
The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions and typographical errors. Moor Insights & Strategy disclaims all warranties as to the accuracy, completeness or adequacy of such information and shall have no liability for errors, omissions or inadequacies in such information. This document consists of the opinions of Moor Insights & Strategy and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice.

Moor Insights & Strategy provides forecasts and forward-looking statements as directional indicators and not as precise predictions of future events. While our forecasts and forward-looking statements represent our current judgment on what the future holds, they are subject to risks and uncertainties that could cause actual results to differ materially. You are cautioned not to place undue reliance on these forecasts and forward-looking statements, which reflect our opinions only as of the date of publication for this document. Please keep in mind that we are not obligating ourselves to revise or publicly release the results of any revision to these forecasts and forward-looking statements in light of new information or future events.