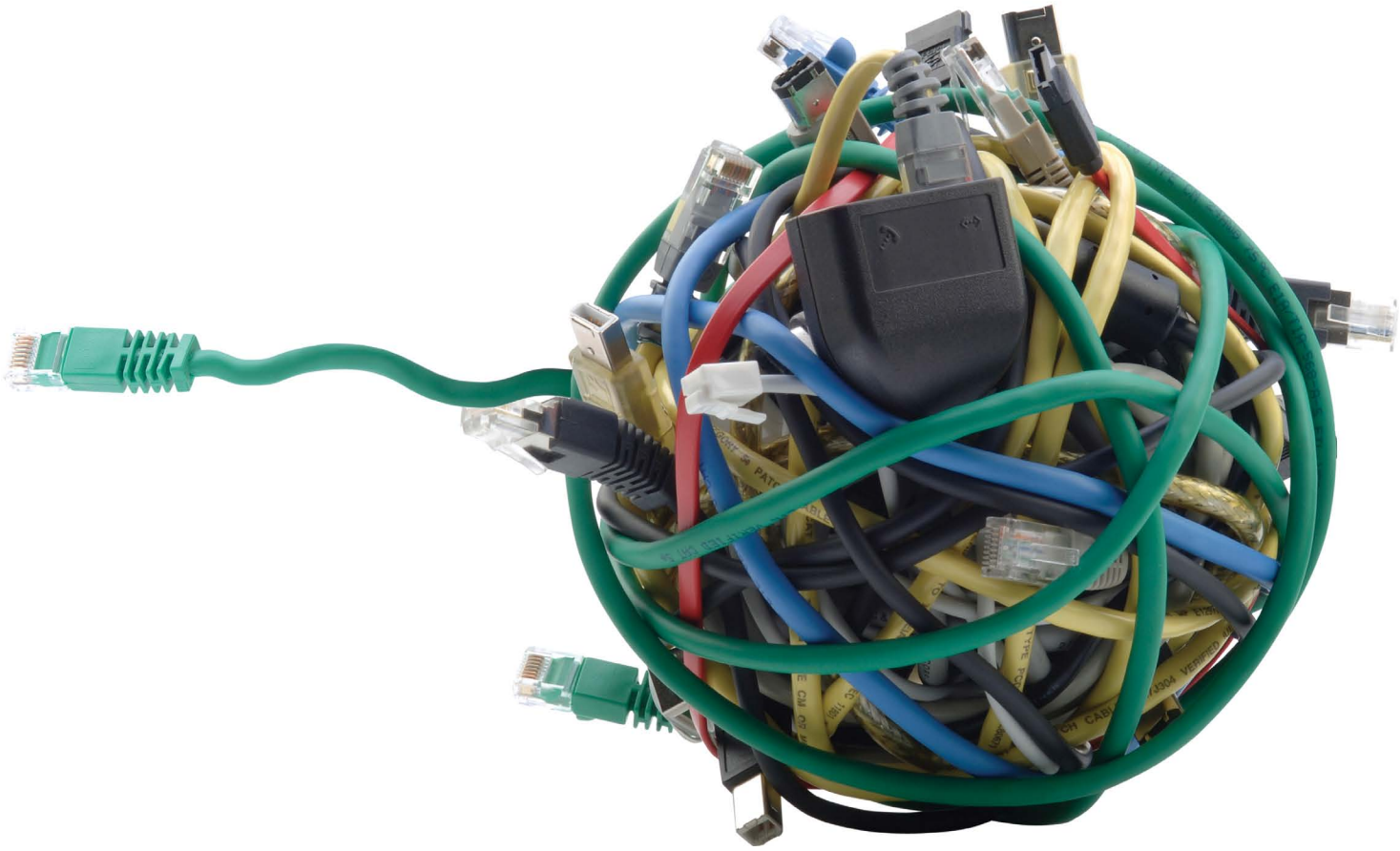
NextIO

White Paper:

Best Practices That Address High Profile Pain Points and Challenges in Today's Dynamic Data Center



Background

Data center server administrators are facing increased challenges around their server I/O requirements. Virtual server sprawl, mobility of applications, and new high performance technologies are combining together to demand a new way of thinking about deploying servers inside the rack. Several key challenges for the data center server admin are described below, along with best practices for implementation.

Challenge #1

Virtual server hosts consume large numbers of Ethernet ports.

Considering the increasing rates of data center virtualization deployments and the fact that VMware has 84% of that market, the implications are huge and the pain is considerable.

Typical server deployments using 1GbE use 10-12 Ethernet ports; larger servers can take up to 20-24 1GbE ports. This not only impacts the servers from an adapter standpoint, it also increases the number of cables in the rack as well as the number of Top of Rack (TOR) leaf switches required. 10GbE improves the situation somewhat, but VMware deployments using 10GbE still require two to three adapters. In both cases, the number of adapters is driven by:

- The number of virtual adapters required by each Virtual Machine (VM)
- The bandwidth required for performing virtual server migration in a timely fashion
- The predominant use of NAS storage in VMware deployments (largely due to the superiority of VMware's NAS provisioning resources vs. those for block storage)
- The number of physical network connections to each server required by each VM

Painful Impacts:

- The high implementation and ongoing management costs incurred by deploying vMotion using multiple links limits the total number of applications that can economically deploy VMware. This means that many servers in the data center that aren't using VMware are significantly underutilized.
- Data center managers that do choose to deploy VMware have to deal with managing huge numbers of TOR leaf switches, significant cable management issues (think of 240 Ethernet cables in the back of a single rack), and high acquisition costs.

Best Practice Solutions:
One way to reduce TCO (both CapEx and OpEx), as well as the increased complexity associated with logistical issues around a VMware deployment, would be to consolidate the I/O at the top of rack in a single appliance. This approach would create a pool of I/O resources that could be shared and reallocated among servers as workloads dictate.

However, for the solution to be most effective, one must enable multiple ports for different networks to be available to all VM hosts. It should also allow for additional virtual adapters and/ or more bandwidth to be dynamically allocated. It must also address system level needs for increased performance and flexibility by enabling the addition of more physical adapters as more virtual adapters are deployed in the rack. As a best practice, the solution needs to allow data center managers to add other technologies besides 1Gb or 10Gb Ethernet, such as Fibre Channel, InfiniBband or SAS. And of course, the solution must be able to handle redundancy requirements.

Challenge #2

Deploying and maintaining multi-fabric data centers is proving to be a very expensive approach from both an acquisition (CapEx) and operations (OpEx) perspective.

While server hardware costs are decreasing rapidly and lower end CPU's are drastically cheaper, acquisition costs for 10GbE and Fibre Channel adapters remain high, on average, $1000 for 10GbE, and $1300 for Fibre Channel. This brings CapEx costs for 10GbE and Fibre Channel inside the rack to $3k or more per server when adding in Ethernet NICs, FC HBAs, and corresponding switches for a 32-server configuration. Doing the math for example, a 32-server rack configuration would cost $96k. For enterprise data centers with10 or more racks of 32 servers, the costs add up very quickly. For managed service providers whose revenue is dependent upon standing up server racks frequently and quickly, this becomes an expensive proposition. Multiple fabrics also require specific management skill sets. Most organizations today staff at least one group per technology. They have LAN administrators, SAN administrators and server administrators. This makes operating costs more expensive because of the duplication in functions associated with each fabric and often in the delays associated with transitioning a task between silo organizations and different skill sets.

Painful Impacts:

- High acquisition and operating costs
- Potentially lost revenue due to delays in standing up new server racks

Best Practice Solutions:
A best practice solution would be one that reduces both CapEx and OpEx costs while minimizing the operational complexities associated with a multi-fabric data center. A way to achieve this would be to create shared pools of I/O resources at the top of rack. By consolidating and virtualizing the I/O devices, the need for costly multiple physical adapters, leaf switches and cables is often eliminated or significantly reduced. Independent Ethernet and Fibre Channel connections per server could be consolidated over a single PCI Express (PCIe) connection going directly into the top of rack appliance. A standard rack configuration would involve two PCIe cables from each server, set up in redundant connections to two top of rack appliances.

With this approach, managing multiple fabrics becomes much easier from a staffing perspective, enabling organizations to stand up new servers much more quickly and accelerate time to revenue.

Challenge #3

Brand name servers with capacity for multiple GPU's are expensive

GPU hybrid computing has moved from a niche market to a major force in High Performance Computing (HPC). Most of the world's top supercomputers are being deployed with clusters of GPUs for hybrid computing. Customers realize tremendous computing value with GPUs because the floating-point performance of a GPU can be 10X or more the performance of a standard X86 processor. As customers migrate from the development/research phase on GPU computing to full deployments, they often require the clustering of multiple GPUs on a single X86 compute node for maximum performance. Unfortunately, reliable servers from brand name vendors are more expensive when purchased with the slot capacity for multiple GPUs. This drives up the CapEx cost of the cluster dramatically and limits the options on cluster deployment. It may also force a qualification of new servers, driving up the OpEx costs of deploying the cluster.

Painful Impacts:

- High acquisition costs
- Reduced selection of server products
- Potential OpEx increases through qualification of new severs

Best Practice Solutions:

A best practice solution for deploying GPU clusters is one that enables a wide variety of server choices from reliable, brand name vendors. The simplest way to achieve this is to separate the GPUs from the x86 server. The GPUs can be contained in a chassis separate from the x86 server and connected via industry standard PCIe Express cables. This allows a customer to determine how many GPUs per server are optimal in that environment. Customers can deploy 1 GPU per server up to 16 GPUs per server without having to change the preferred server product. Additionally, tier one servers have high performance PCI Express slots. The external GPU model enables a single GPU cluster box to interoperate with many servers from many top vendors.

Challenge #4

Dense server form factors cannot host more than one or two GPUs, preventing them from optimal use in hybrid compute clusters

Enterprise data centers are under constant pressure to add more computing power in the same power, cooling, and real estate. It is especially important in High Performance Computing environments where maximum performance within those constraints is a requirement. This pressure has dramatically increased the adoption of dense blade, 1U, and ½ U servers as the form factor of choice for High Performance Computing. However, as GPU hybrid computing increases in popularity, the dense server form factors become problematic. Optimal GPU hybrid computing requires multiple GPUs per server. The dense server form factors generally cannot hold more than 1 one or 2 two GPUs per server. This creates a dilemma for the data center manager. A different, sub-optimal server must be utilized for hybrid computing, or a sub-optimal GPU to server ratio must be used. In either case, the computing power for a rack is reduced and the total space required for a cluster increases. Additionally, the data center manager must now deal with management and qualification of different servers for the standard enterprise functions and the HPC hybrid computing functions.

Painful Impacts:

- Increased power, cooling, and real estate per server
- Increased OpEx costs due to management and qualification of different servers

Best Practice Solutions:
The optimal way to deploy GPUs for hybrid computing with dense servers is with external GPU appliances. These appliances separate the server from the cluster of GPUs, allowing a large number of GPUs to be connected to a single, dense server. By deploying external GPUs with a PCIe attached appliance, a single PCIe slot in a dense server can connect to as many as 16 GPUs. Now, the data center manager is free to choose an optimal GPU to CPU ratio in any form factor server. Maximum rack density is possible in this configuration, yielding the best performance for a given power, cooling, and real estate budget.

As data center server administrators seek to deploy more virtual servers and HPC GPU systems, they will encounter new challenges in the I/O infrastructure of those servers. Administrators who deploy rack level I/O architecture will find these solutions to be simultaneously low cost, flexible, and offer high performance. A technology solution that allows independent choices for the server and I/O platforms will yield optimum results in cost, management, and upgradeability.

For more information about NextIO products and solutions, contact info@nextio.com