What are the Requirements to Consolidate Storage?

by George Crump, Lead Analyst

In the perfect world, a data center would have a single storage system. One that would provide high performance and high capacity in a single platform that doesn't take up much data center floor space. Ideally, this system would be reliable enough that it could back itself up to an even more cost-effective storage tier but one that is again part of the same system.

Unfortunately, the data center is not a perfect world. It has to respond constantly to new and often unexpected requests, and "throwing hardware" at the problem seems like the fastest way to appease the demand, especially in the new "agile" data center. As a result, there is more storage sprawl in the modern data center than ever. But the data center should not give up on the consolidation dream. In fact, the need to respond quickly is all the more reason to invest in a consolidated architecture, but that architecture has to meet certain requirements so that the data center does not have to compromise its agility.

What Causes Storage Sprawl?

Most data centers end up with multiple storage systems so that they can address the "need of the day." Typically, the original storage system is inflexible and can't adapt to these new demands on performance or capacity. At the same time, the IT staff is stretched too thin and does not have the time to focus on finding a particular storage system to meet the needs of the entire data center. As a result, they are susceptible to believing the claims of the last vendor in the door.

A request for new storage, or complaints about existing storage, combined with an overworked IT staff ends up leading to the purchase of a new storage system to address that particular problem. But this quick answer leads to management chaos and cost overruns as well as exposed data protection policies.

Making Storage Consolidation a Reality

Despite the current state of affairs in the data center, storage consolidation can again be a reality. The advent of flash and high-capacity hard drives allows a storage system to meet the extremity of mixed workloads. But these vastly different technologies have to be used intelligently so that predictable performance for mission critical workloads can be maintained. A storage system that positions itself as a consolidated storage solution though, has a higher burden of responsibility than the various point solutions that permeate the market today. It has to provide high performance and high reliability in as small a data center footprint as possible.

The Requirements for Storage Consolidation

Performance

The agile data center can't compromise on performance, and a move to a consolidated architecture should not force them to make a choice between performance or consolidation. But at the same time it is not cost efficient nor practical that the consolidated storage system be all flash as some vendors propose. There is simply too much data that does not require the performance, nor the expense, of all flash storage.

The problem is that most hybrid storage systems have garnered a bad reputation for inconsistent performance because they often skimp on the use of cache, and they use a small number of high capacity drives. An inadequate cache means more frequent misses, and a small number of high capacity hard disks means that there are less available to respond to the cache miss.

Instead, the consolidated storage system should leverage a very large cache area that is built from both DRAM and Flash. Increasing the size significantly reduces the chances of a cache miss and leveraging DRAM further boosts storage performance. Also the consolidated system benefits by being able to use a high number of high capacity drives appropriately since it is now storing all the data for the data center. This means that there are enough hard disk spindles to still sustain acceptable performance when there is, a now rarer, cache miss.

Finally, the storage system should provide the ability to deliver prioritization for specific workloads so that mission critical workloads can be guaranteed the performance they need. Performance assurance can be achieved by leveraging software built into the storage system or leveraging the quality of service capabilities within the operating environment.

Multi-Petabyte Scale in Less Space

Since the consolidated storage system handles all production workloads as well as potentially all backup and archive workloads, it needs to be able to provide multi-petabyte scale. But this scale needs to be done more space efficiently than legacy storage systems. Using hard disk drives that are 6TB or 8TB (most point storage systems use 2TB or 4TB) in capacity delivers a more space efficient scale without having to count on technologies like deduplication or compression.

While these high capacity drives do reduce the storage system's footprint, they also raise concerns about performance. As mentioned above, the performance can be offset by leveraging a large cache front-end and by a high number of drives. Using a large memory storage area and efficiently using high capacity disk storage is, again, now practical if one storage system is going to meet an entire data center's needs.

Unprecedented Reliability

The high capacity drives used to reduce footprint also create concerns with recoverability. The data protection scheme deployed in the consolidated storage system has to be intelligent, so when a drive fails only the actual data on that drive has to be recovered. In other words, the system needs to avoid the beginning-to-end read of the drive, standard in RAID5/6 designs, to return to a protected state. This same technology also allows for multiple drives in the system to assist in the rebuild process. The result should be a return to a fully protected state in a matter of minutes instead of hours or days.

Reliability goes far beyond drive protection though. Most storage systems today are as much software as they are hardware. A software focused storage system means more frequent updates so that the system can meet the ever changing needs of the data center. As a result, these systems need to be able to be non-disruptively upgraded.

Non-disruptive upgrades need to be thought about differently. The typical non-disruptive upgrade is done one storage node at a time in a sequential manner. Each time a node is upgraded the hosts, connected to that node must undergo a path failure recovery process. The path failure recovery process is handled by a multi-pathing drive on the host and typically causes application latency that may require an application restart or at least user concern.

The next generation of consolidated storage systems need to provide a method to upgrade the entire system at once. It has to be done very quickly, below the path time out, so that hosts don't need to go through a path recovery process. The result is the entire storage system upgrade process is less intrusive to the overall environment. While this approach is more technologically challenge for the storage system maker, once complete, it provides the customers with a safer and less disruptive upgrade process.

This type of non-disruptive upgrade is often best deployed in a scale-out architecture built from a cluster of storage servers or nodes. The scale-out architecture also allows for additional performance and capacity to be added in the same non-disruptive fashion. Finally, a scale-out architecture more easily enables a triple or better level of redundancy across components, with the final result being a six or even seven 9's of uptime, instead of the more typical five.

Totally Unified Storage

Lastly, to consolidate all the workloads in the environment also means supporting the various protocols that the data center requires. This means both "old" and "new" protocols need to be supported beyond the standard File (CIFS/NFS) and block support. A consolidated system will need to eventually support legacy mainframe connectivity as well as provide object storage support. Finally, interfacing with operating environments like vCenter and OpenStack will be important.

The important aspect in connectivity is that the consolidated storage systems have the ability to evolve over time. New protocols are being introduced all the time. The storage software needs to be able to be updated, again non-disruptively, in order to add that support.

Conclusion

The goal of consolidating storage seems like a distant memory thanks to the data center's need to respond to a rapidly changing business environment. However, the reality is that the very technology that seems to justify storage sprawl, flash and high capacity hard drives, is the very technology that justifies consolidation. These techniques need to be brought together in a storage system that has the intelligence and the scale to leverage correctly each of their attributes.

Sponsored by INFINIDAT

INFINIDAT has brought to market InfiniBox, a new generation of highly reliable, scalable and efficient storage systems designed specifically to support a wide variety of workload types, including those in virtual architectures. InfiniBox allows an organization to consolidate down to a single solution that has the performance, capacity and affordability to support most of an organization's workloads.