



Why is Virtualization creating Storage Sprawl?



by George Crump, Lead Analyst

Desktop and server virtualization have brought many benefits to the data center. These two initiatives have allowed IT to respond quickly to the needs of the organization while driving down IT costs, physical footprint requirements and energy demands. But there is one area of the data center that has actually increased in cost since virtualization started to make its way into production... storage. Because of virtualization, more data centers need flash to meet the random I/O nature of the virtualized environment, which of course is more expensive, on a dollar per GB basis, than hard disk drives. The single biggest problem however is the significant increase in the number of discrete storage systems that service the environment. This “storage sprawl” threatens the return on investment (ROI) of virtualization projects and makes storage more complex to manage.

Why is Storage Sprawl Worse in Virtual Environments?

Storage sprawl has become worse in organizations that have made a significant investment in virtualization technologies. For example most virtual desktop infrastructures (VDI) have an allflash array to handle desktop images, preventing boot storms, logout storms and maintaining acceptable performance throughout the day. But most VDI environments also need a file store for user home directories. There is little to be gained if this data is placed on the all-flash array, but certainly data centers need to provide storage to their users to support user created data. As a result most organizations end up buying a separate Network Attached Storage (NAS) device to support user home directories and other types of unstructured data.

The virtual server environment will also see multiple storage systems implemented to support its operation. This includes a high performance system, typically an all-flash array, to make sure that critical applications experience a performance profile similar to their bare metal days. There is also a workhorse type of system that today is often a mix of flash and hard disk drives, a hybrid system. The role of this system is to support the mid-tier applications and services.

Finally there is a third type of storage often used for old VMs and sometimes old user data, as well as data being collected from big data sources like sensors. This system is often capacity focused using high capacity hard drives to create an affordable storage area for this “at rest” data.

Does Storage Sprawl Matter?

Some vendors will claim that storage sprawl no longer matters in virtualized environments, that the hypervisor can manage the movement of virtual machine data between types of storage, just like it moves virtual machines between different types of compute servers. This line of thinking encourages the purchase of the above purpose specific storage systems; an all-flash array for performance sensitive VMs, a hybrid array for more modest VMs



and a hard disk based system for VMs that have no use for the performance that a flash based system could provide.

What is not explained is that these separate silos of storage still need to be managed. They often use their own unique snapshot, replication and other data services per system. That means that an IT administrator needs to learn each storage system interface and scripting language. It also means that when data is transferred between these systems, that the application may see a performance slow down, and may cause an outage, while this migration occurs. The transfer needs to copy or move data from the primary storage system to the alternative storage system and that means, at a minimum, transferring data across the storage network and potentially the general purpose network. Finally, there is an optimization issue. Similar to direct attached storage, silos of shared storage can't be used optimally because there is rarely a balance between performance focused applications and capacity based ones.

Storage sprawl results in three issues for IT managers. First is wasted IT administrator time even though something like an all-flash or even a hybrid array was purchased to reduce time spent tuning storage performance. The second factor is unpredictable performance deliverables, even though these systems were implemented to resolve that issue. The need to transfer data across a network will alter performance characteristics. And third the imbalance of performance and capacity resources among the storage silos.

Can Sprawl Be Stopped?

The solution, a high-performance mixed workload storage system that supports a wide variety of workloads and a wide variety of storage protocols while delivering the performance and storage economics that the various workloads need. The key to full return on storage consolidation investment is a single system that can be fully utilized while still delivering on the specific demands required by each workload it supports.

Why Stop Virtualization's Storage Sprawl

For a storage system to stop storage sprawl there are several key requirements which it must deliver, and if it can successfully deliver these requirements, then the benefits of a single system are numerous. First a single unit is simply easier to manage. No matter how much software is layered over the management of multiple storage systems, it can't hide the fact that there are indeed multiple storage systems there that require more time to manage.

The second benefit is that a single storage system can be used more efficiently. A data center with a single storage system never has to worry about storage system B running out of storage capacity while storage system A is running out of storage performance.

The third benefit is economics. Even if the independent storage systems are priced aggressively, the organization is still paying for multiple storage controllers, storage capacity and storage software features. And of course, each will have its own service agreement that has to be maintained.

The Requirements for Consolidated Storage

The first requirement is for mixed media. While flash has become less expensive, it is still not as competitive as capacity disk pricing. And a storage system that is trying to consolidate ALL storage will be required to support data types that simply don't belong on flash. At the same time, performance demands need to be addressed



by the consolidated storage system. These systems will also have to leverage flash storage to meet those demands. Further they should than what flash can provide it is able to deliver that too.

The second requirement is for a high performance HDD tier. Too many hybrid storage systems leverage flash and then only high capacity HDDs as their second tier. The problem is that high capacity hard drives are slower, and because fewer of them are required to meet the capacity demands of the environment there are less of the drives to dedicate to performance.

The consolidated system needs to leverage faster performing drives of more moderate capacity. This also means that more drives will be used. This combination will deliver respectable HDD performance without increasing cost. More importantly, there will be less of a drop off between flash and HDD performance in the case of a cache miss.

The third requirement is that the consolidated storage system also support multiple protocols. This means going beyond the classic SAN/NAS support and providing both Object and even Mainframe access. For storage consolidation to make sense it needs to move beyond the virtualized use case and be able to provide capacity for analytics and other workload types.

The fourth requirement is that the consolidated system be able to keep up with performance and capacity demands as they continue to increase over time. This means that the system should have the ability to scale-out instead of scale up. A scale-out system allows the organization to avoid costly fork-lift upgrades to their storage systems. This is even more critical as workloads are consolidated since the used capacity of the system will be so much higher. High capacity means longer migrations; longer migrations mean more downtime. Scale-out eliminates the need for migration.

Finally, and most important, is the need for a highly reliable system with multiple points of redundancy. If the storage system is truly going to be the only storage system in the environment, then it can't fail or experience any outages. This reliability should be adjustable by application or workload so that mission critical workloads could survive multiple outages.

Conclusion

A single consolidated storage system should bring many benefits to the organization, but those benefits come at a risk of variable application performance and also a greater risk of failure. To mitigate these risks the consolidating storage system needs to leverage multiple types of storage media and have multiple points of redundancy built into the system. If this combination can be delivered, the virtual environment should become simpler to administrate while also being less expensive to run.

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.

Storage Switzerland is the leading storage analyst firm focused on the emerging storage categories of memory based storage (Flash), big data, virtualization, cloud computing and data protection. The firm is widely recognized for its blogs, white papers, and videos on such current technologies like all-flash arrays, deduplication, software-defined storage, backup appliances, and storage networking. The "Switzerland" in the firm's name indicates our pledge to provide neutral analysis of the storage marketplace, rather than focusing on a single vendor or approach.