STORAGE SWITZERLAND REPORT

STORAGE PERFORMANCE SPRAWL VS. CONSOLIDATION

George Crump, Senior Analyst

Consolidation is a strategic project underway in many data centers today. There is consolidation of servers with virtualization, consolidation of infrastructure built on 10GbE technology and consolidation of storage around larger, highly scalable storage systems. What there is not, however, is storage performance consolidation. In fact, storage performance seems to be going in the opposite direction. There's a 'storage performance sprawl' occurring in today's data centers.

Storage performance sprawl refers to the addition of multiple distributed components in the storage infrastructure designed to improve specific performance pain points. This is being caused because the mechanical hard drive has been stuck at 15K RPMs since 1999. To get around the performance limitation of the individual drive, users and storage manufacturers have been forced to get creative in order to wring additional performance out of this now decade-old limitation.

Work-arounds, like high drive-count RAID groups that stripe data across all the drives in the array and then subsequently short stroke those drives, have lead to poor utilization, high cost and power inefficiencies. These band-aides as well as decreasing costs have brought about the adoption of solid state disk (SSD) to decrease mechanical drive count and further increase performance. However, the SSD adoption leads to potentially further complications as a new tier is implemented and the decisions are made about what data should be moved to that tier, how it should be moved and when.

These limitations have lead to a 'throw more hardware at it' approach to improving storage performance. Other than adding the ability to manage these large array groups, storage systems suppliers have been primarily focused on providing additional data services like snapshots, space optimization and thin provisioning. Increasing storage I/O performance has centered mostly on increasing drive count, increasing I/O bandwidth and adding the aforementioned SSD technology. Quiet apart from storage hardware, there has been little work on enhancing the performance attributes and intelligence of the storage software stack.

The challenge with using more hardware to fix a performance problem is that it must be applied in a near-universal fashion and that it's expensive. For example if there is a NAS I/O performance issue the typical approach is to buy a newer NAS unit with faster storage processors, higher bandwidth network adapters and potentially, some form of solid state disk.

Beyond the obvious cost issues, the problem is that this only increases the performance potential of the applications and data that are hosted on that particular NAS, not the throughput of the environment as a whole. Also each NAS device, each server and each user accessing that data likely needs to have individual bandwidth increases to take advantage of that added performance. Especially in the case of SSD technology, the data services (snapshots, clones, optimization) may actually become the bottleneck. With SSD there is no longer the 'luxury' of a highly latent storage technology giving the storage systems the ability to provide other data services during 'spare' cycles. Finally, all these different components of performance must be managed independently and manually, placing additional burden on IT staffs.

Traditionally, when designing performance bottlenecks out of an environment the architecture has to be optimized for when the I/O demands are at their highest. This peak-load design means most of the components in the environment must be upgraded for that performance improvement to be realized. In reality high performance demands are seldom constant. They come from a variety of sources at different times and there is almost always 'quiet time', or periods when demand is low. The result is much of the additional performance architecture is underutilized for most of the day.

A potential solution may be to develop a performance services appliance like those offered by Avere Systems. These solutions allow for the consolidation of NAS performance behind a single system without replacing the existing NAS devices nor upgrading them with faster bandwidth and more expensive drive configurations or technologies.

These performance consolidators like Avere's are often stand-alone appliances that can be clustered in some form for availability. They will often have a combination of DRAM, SSD and SAS based drive technology to distribute among the active data sets, dependent on the access patterns of that data. The storage manager however has to deal with just one I/O target. Users and applications are redirected to the appliance, the appliance communicates with the NAS device (s) that now become back-end data service devices for storing and protecting the near-active and dormant datasets.

The result is that performance is consolidated on the appliance and no changes are made to the NAS hardware or O/S. Performance is increased without the need for upgrades to hardware I/O, drive count or drive technology. Additionally, all the bandwidth capabilities within those NAS units remain the same. Instead, performance of active data is centralized on the performance consolidation system when load is high and then distributed back to the legacy NAS layer when the load is low.

With this approach the front-end performance consolidation appliance serves as a staging area for high speed access to the most active data. It automatically promotes and demotes data to the staging area from the existing NAS systems in the environment. As data settles and applications go quiet, the active data is copied back to the NAS systems. The legacy NAS systems then serve as the storage area for all of the inactive data as well as an additional copy of the active data. The storage manager can leverage the NAS systems' rich data services to make snapshots, perform optimization and protect the data. Most importantly though, these data services can all be performed without affecting the performance layer.

The effect of performance consolidation is that it can eliminate performance sprawl. With it, back-end NAS systems can remain in more standard configurations, without having to upgrade drives or bandwidth. By consolidating performance of the entire NAS infrastructure, future purchases can become more capacity focused instead of performance focused. Basic, more economical NAS systems, that normally would not perform adequately for the environment, can be implemented without sacrificing performance.

Performance consolidation reduces the cost of ongoing NAS investments. It also simplifies the management of the environment by centralizing performance technologies like SSD, high speed SAS and 10GBE bandwidth into a single system.

About Storage Switzerland

Storage Switzerland is an analyst firm focused on the virtualization and storage marketplaces. For more information please visit our web site: http://www.storage-switzerland.com.