

Storage Virtualization for Rock Stars

STORAGE VIRTUALIZATION FOR ROCK STARS

Part 3: Storage in the Fast Lane – Achieving “Off-the-Charts” Performance Management

When most business or IT managers are queried for their definition of storage efficiency, the answer quickly turns to performance. Storage needs to be nimble, delivering not only capacity sufficient to meet the growing volume of data but also the speed to write data as rapidly as it is presented and to retrieve it immediately upon requests from applications and decision-makers. Hardware vendors have long marketed speed enhancements as a key differentiator and main selling point of their latest storage array, switch or interconnect, charging a premium for the latest, fastest technology.

Storage virtualization has the potential to break speed barriers without breaking the storage budget.



Jon Toigo
Chairman, The Data Management Institute LLC

STORAGE VIRTUALIZATION FOR ROCK STARS

Storage in the Fast Lane: Achieving “Off-the-Charts” Performance Management

STORAGE PERFORMANCE: THE MEASURE OF IT EFFICIENCY?

Rightly or wrongly, the yardstick most commonly applied to evaluating how well IT departments are delivering services to their users is speed. The speed at which databases post transactions (“write operations”) and the speed with which requested information is painted onto the screens of user client devices (“read operations”) shape the perceptions of users regarding the efficiency of IT operations overall.

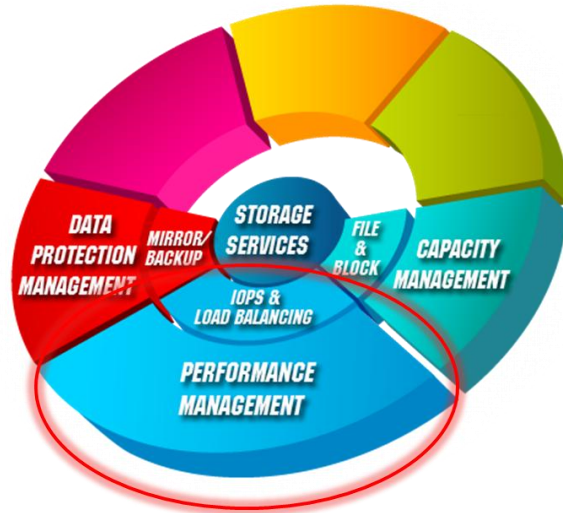
Input/output speed is, of course, a relative thing and hardly a precise or accurate measure of overall IT performance. I/O performance is influenced by many variables. At root, it reflects the balance that has been struck between processors, networking and storage infrastructures – three technologies that are constantly changing in terms of their performance characteristics. But speed is also influenced by a broad set of factors linked to workload, traffic patterns, and operating system and application software stack design.

Truth be told, when one considers the successful completion of the trillions of transactions that traverse IT infrastructure each working day, the volume of digital data that is processed and handled by information technology, and the myriad transformations to which the data is subjected as it moves between physical hardware and logical abstractions, one could readily conclude that IT administrators are the modern equivalent of miracle workers.

Still, I/O speed is a “touch point” between IT and the business it serves. As such, it creates and reinforces perceptions of overall IT performance with practical ramifications for strategies, budgets, and careers. Speed is also leveraged by vendors to promote their wares, with boasts about “the fastest performance of any server/switch/interconnect/array” frequently cited in marketing literature and sales campaigns to differentiate competing products in the market.

Discussions of speed tend to narrow quickly to a focus on storage infrastructure. These days, mechanical storage media manifests much slower I/O rate than networks and backplanes or processors and memories. Yet, to borrow the line of a contemporary storage vendor, storage is where information lives.

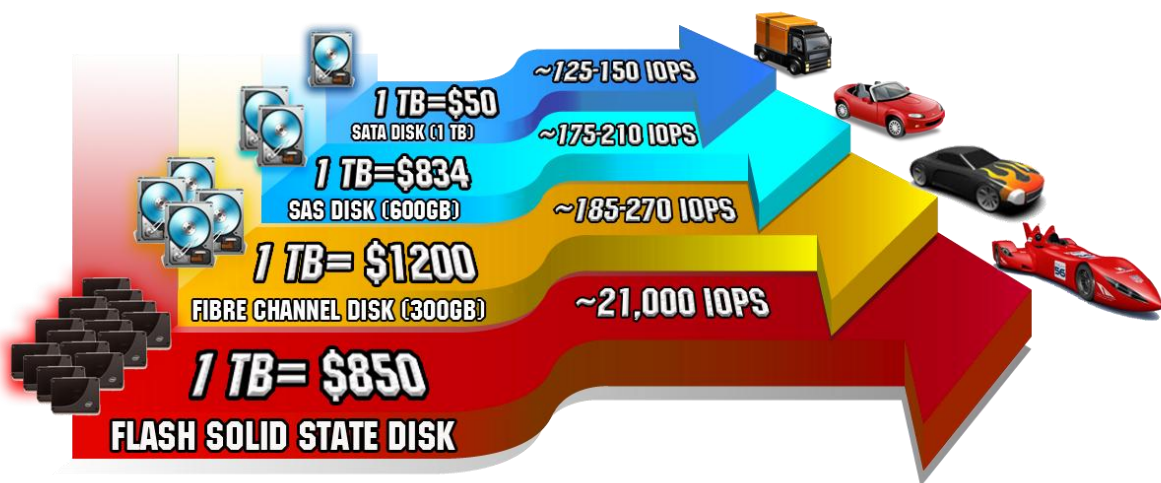
Thus, considerable effort has been made to optimize storage performance at the device level and to architect storage infrastructure to get the most speed for the money. Managing storage performance over time is therefore a key management service expected of IT administrators.



Storage performance management is a multifaceted task. It actually begins with the assessment of application

requirements and the selection of appropriate media to host application data. Key criteria in media selection can include data output characteristics -- whether the data being generated is transactional in nature, consisting of short block writes, or if it comprises of large files that may extend beyond the size of a specific disk drive. Access frequency and numbers of concurrent accesses expected of data, and the duration of high frequency and high concurrency accesses (in seconds, minutes, days, or weeks) are also important criteria.

Cost is also important in media selection. Generally speaking, the faster the storage, the more expensive it is, so striking a balance between speed and expense usually sets pragmatic parameters on media selection. Differences between media performance and cost are illustrated below.



For over 30 years, one approach used to accelerate the speed of storage infrastructure has been to house many disks in an array, then to create logical drive comprising many individual physical drives so they are presented as a single unit. In its simplest form, the intention is to apply the greatest number of read/write heads to the task of capturing data to the disk. Spreading out the workload in a parallel fashion provides an overall performance capability that exceeds the write speeds of any one disk drive.

Additionally, data may be written only to a part of each drive – an area requiring the smallest amount of read/write head movement – so that latency (the amount of time needed to orient the head over the writing platter surface – can be minimized. This is called “short stroking” and, while effective at delivering speed improvement, is very wasteful of disk media, expensive in terms of the numbers of disk that need to be deployed, and heavy on power consumption.

In addition to simple disk arrays and short stroking, caching has also been used to optimize disk performance. A cache controller, consisting of memory and logic, provides a means to capture and acknowledge writes received at a disk array and to queue actual data placement operations so that the performance of the application is not hampered by the slower speed of disk operations themselves.

Read caching is slightly different. Data that is accessed frequently from disk may be read into memory on the cache controller and served to applications from that location, where performance is several orders of magnitude faster than the speed at which data is located and read from physical disk. Considerable work has been done on read/write caching over the years to help optimize disk performance, contributing to the costliness of arrays as much as to their performance.

The most recent innovation in the area of disk performance acceleration is the industry’s current initiatives around flash random access memory-based solid state disk (FLASH SSD). Some vendors have chosen to create FLASH SSD in the same form factor as 3.5 or 2.5 inch disk drives and use these devices as targets for hosting data directly. Writes and reads are significantly faster using solid state memory than using rotational magnetic media, but there are caveats. For one, data written to the cells of a FLASH SSD must be erased before the cells can be written again. (With a hard disk, new data simply overwrites old data.) That additional erase operation causes FLASH SSD to present non-linear performance.

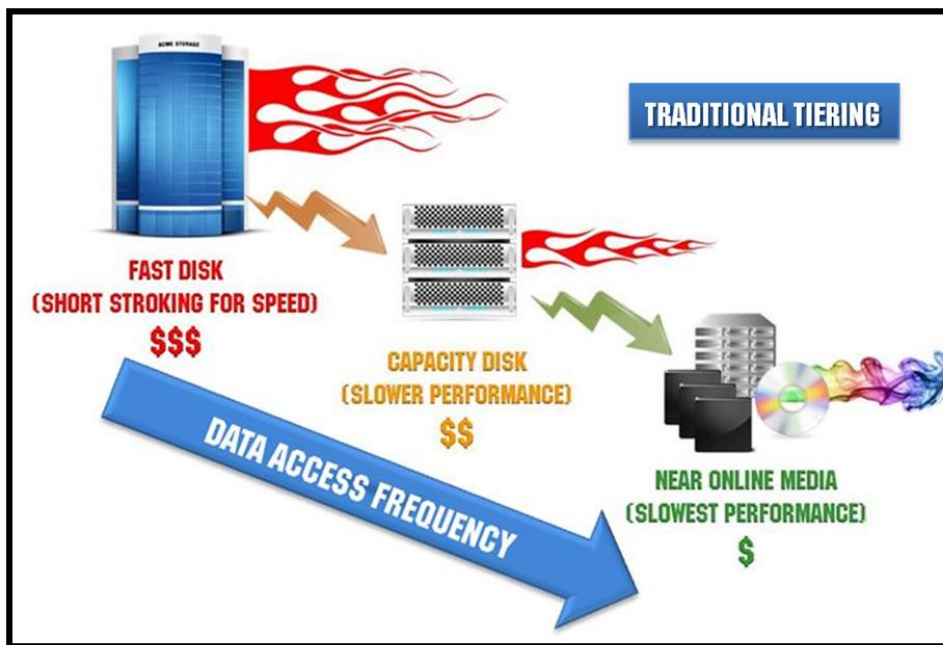
Also, FLASH SSD currently manifests a memory wear problem: a maximum of 250,000 writes can be made to a single cell location before it wears out. Vendors typically work around these and other issues by delivering far more capacity on FLASH SSD than is

advertised, then swapping out written cell banks and worn out cells with capacity hidden on the drive.

Bottom line: while the advertised drive life of a FLASH SSD drive is said to be on par with magnetic disk (about five years), they may burn out or operate at much reduced performance much more quickly than a disk drive depending on the workload. In a laptop computer operating personal productivity applications, a FLASH SSD might be a comfortable, if pricey, replacement for a conventional mechanical disk drive. However, as storage for a high performance transaction processing application generating millions of write operations per second, FLASH SSD units may need to be replaced weekly.

An alternative use of FLASH SSD to accelerate performance is to use it in a disk optimization role. Some vendors are using confusing terminology to describe this architecture, calling it *auto-tiering*, so a bit of explanation may be required to understand it.

The confusion is around the word tiering. Traditional storage tiering is a protocol, a way of doing things with data over time to optimize the available capacity of disk resources. Traditional tiering conceives of different classes of storage ranging from fast (high speed, low capacity, short stroking) disk arrays intended to capture data from demanding transactional applications such as databases as quickly as possible to slower and slower arrays comprising higher and higher capacity media with slower and slower performance characteristics.



The traditional tiering protocol sees data moving between these storage classes over time as the frequency with which data is being accessed declines. The protocol addresses the problem of capacity, performance and cost elegantly by placing data on storage that is most appropriate to its use.

Some vendors of storage equipment have sought to automate this traditional tiering process and call the resulting capability of their array controller “auto-tiering.” This is a bit of an oxymoron, since the vendor typically sells an array with both fast and slow media in the same cabinet, but obviates the cost-savings that accrue by charging extra for the auto-tiering software on the array controller. In a few cases, vendors have begun adding a stand of FLASH SSD drives as the fastest tier of storage, “Tier 0,” thereby moving each class of mechanical disk down a notch in the tiering protocol. In general, the effect is to dramatically increase the cost of the overall storage platform.

To confuse matters more, some vendors use the term “auto-tiering” in an altogether different way. What they are describing is a method for using a memory device, typically a FLASH SSD drive, as a disk optimizer.

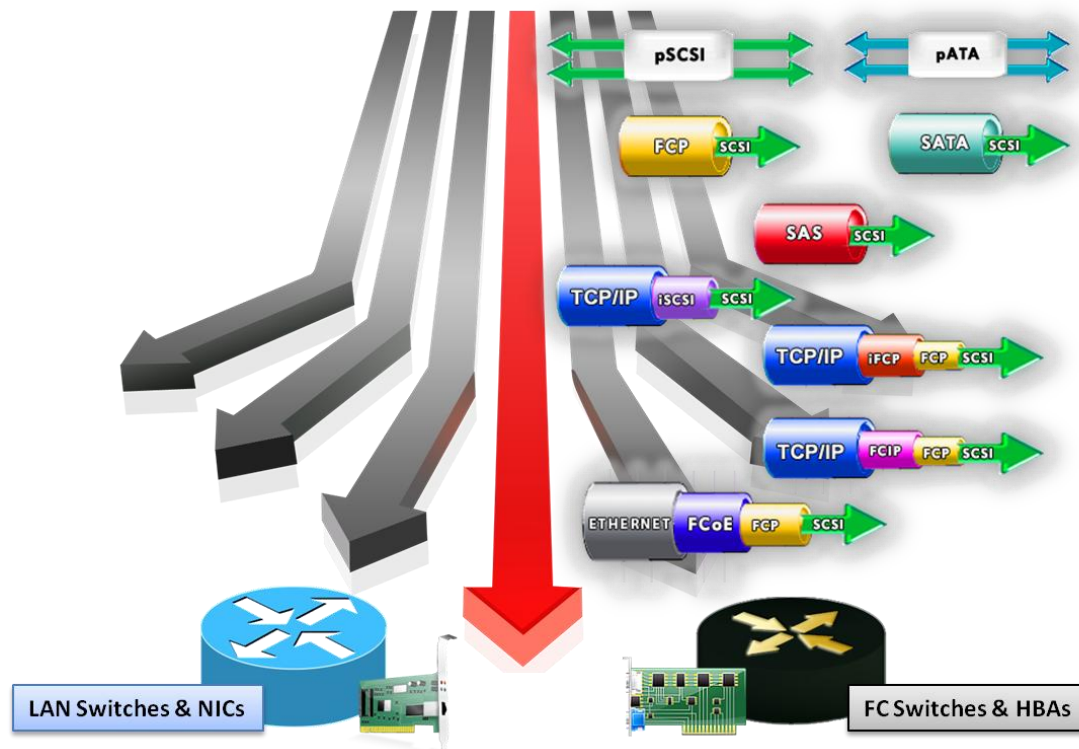
In this approach, data is written by applications to fast disk. However, if the data is determined to be “hot” – accessed frequently or by many concurrent read requests – it is temporarily copied to the FLASH SSD device so it can be serviced at much higher speed. Once the data “cools” and access frequency or concurrency is reduced, read requests are re-pointed back to the hard disk, where normal I/O requests are handled.



These disk optimizations, depending on the data being serviced, can be a cost-effective use of FLASH SSD technology. It remains to be seen whether the technique is best applied at the physical array level, or better exposed as a sharable service across all arrays in the infrastructure. More discussion on this point is provided below.

Just as media selection and protocols such as caching and tiering can influence storage infrastructure performance, so can the choice of the interconnect used to attach storage to servers. Contemporary storage has no shortage of interconnect standards. Most

derive from the Small Computer Systems Interface (SCSI) standard developed over thirty years ago.



In the late 1990s, a proliferation of interconnect technologies began to occur based on the serialization of parallel SCSI to enable more devices to be connected via a single interconnecting bus. The most popular interconnects today – Fibre Channel Protocol, Serial Attached SCSI, and even iSCSI – are all manifestations of the same thing: SCSI serialized. The key differences have been the transport mechanism employed for sending SCSI commands and data serially across a wire. Fibre Channel uses copper or fiber optic cable, specialized switches and host bus adapters installed on servers to extend the backplane of the server over distance – an architecture known as a fabric, but commonly misrepresented as a “network” (as in Storage Area Network or SAN). Serial Attached SCSI is pursuing the same general approach today.

Some vendors preferred to use already ubiquitous Ethernet networks and the Internet Engineering Task Force protocol suite, TCP/IP, as the transport for serialized SCSI, producing a standard called iSCSI (SCSI over IP). A network interface card and LAN switching equipment are used to create this kind of interconnect between storage and servers, with iSCSI operating as an application across a specialized and usually

segregated LAN. Again, this is not a real SAN; it simply uses a LAN as a transport for extending the SCSI bus backplane to numerous target devices.

The key takeaway from the above is that all storage is direct-attached and all storage interconnects are essentially extensions of the bus backplane of a server, rather than “networked storage.” This is important to understand only because of its impact on issues such as storage management. A true network protocol has a management layer: a FC SAN does not.

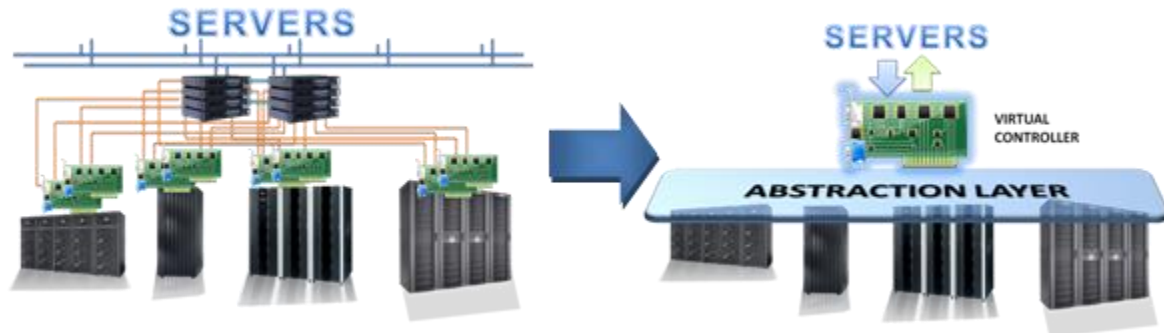
Bottom line: there may be many paths through the storage interconnect fabric but there is nothing in the underlying interconnecting protocols to help manage traffic across paths or to balance the load. Evidence of this deficit can be found in just about every SAN deployed today, where the typical utilization efficiency of a given port on a FC switch hovers at about 15% of optimal. Even iSCSI Ethernet switch ports are underutilized. Performance management requires that paths between servers and storage be optimized so that some ports and links do not become congested, slowing the delivery of commands and data to and from servers and storage, while other paths go largely unused.

In a nutshell, performance management entails the management of media, protocols and paths to achieve the greatest possible speed from the storage infrastructure in response to data from applications and end users. Achieving managed performance in contemporary storage infrastructure confronts several challenges including infrastructure complexity, array isolation, lack of classification of target storage arrays according to performance characteristics, and the absence of quality of service (QoS) and load balancing functionality.



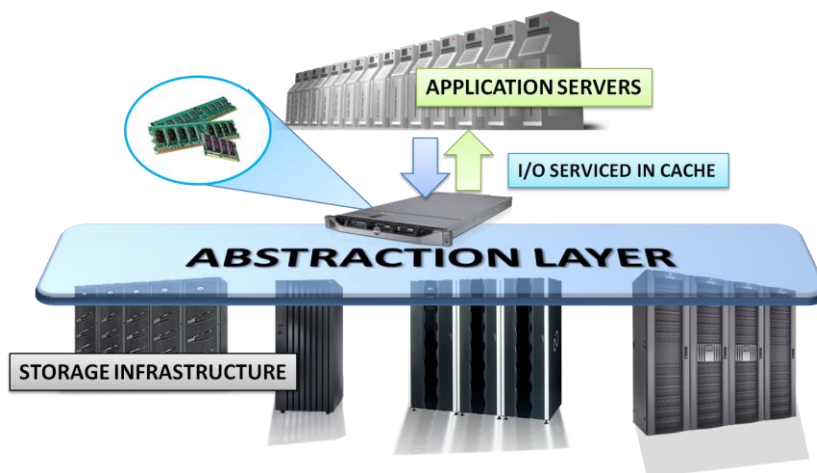
PERFORMANCE MANAGEMENT WITH STORAGE VIRTUALIZATION

The good news is that storage virtualization can help with many of these issues. As noted in previous papers in this series, storage virtualization is the establishment of a logical or virtual controller above the physical storage infrastructure layer – including interconnecting switches, cabling and arrays. Storage virtualization is a technology for simplifying the presentation of storage resources so that those resources can be managed efficiently in terms of capacity, performance and availability.



Speed through Caching

The first thing many administrators notice after virtualizing their storage is a significant performance bump. All I/O is serviced faster than it was when storage was presented directly from physical layer devices. The speed improvement differs from shop to shop, but a 200% improvement is pretty common, and some companies using DataCore Software's SANsymphony-V have reported increases of up to 300% from the storage infrastructure below the virtual controller.



The truth is that this performance bump relates to caching. In the DataCore approach to storage virtualization: the virtualization engine takes the form of software loaded on a commodity server running the Microsoft Windows™ Server 2008

server operating system. Essentially, SANsymphony-V aggregates all of the storage

mounted to the server, enabling the creation of one or more storage pools, then handles all writes to and reads from these logical storage pools using server memory as a cache. As with caching in a physical array, reads and writes to/from the SANsymphony-V memory cache are always faster than reads and writes to/from a physical disk drive: hence, the performance improvement.

Of course, given the diversity of workload directed to the virtual controller and its increasingly arbitrary characteristics – especially as companies embrace server and desktop virtualization – a “static” cache (one that simply queues all write requests) is no guarantor of improved performance over time. In a virtual server environment, for example, many applications may share the same I/O paths and create different traffic levels and unpredictable peak loads that make intelligent provisioning of storage challenging.

DataCore Software’s innovation, called “adaptive caching,” addresses this issue by intelligently balancing workloads based on performance characteristics and other factors. Metaphorically, adaptive caching can be viewed as shock absorbers that normalize differences in terrain so that horsepower and torque are translated effectively into maximum wheel speed.

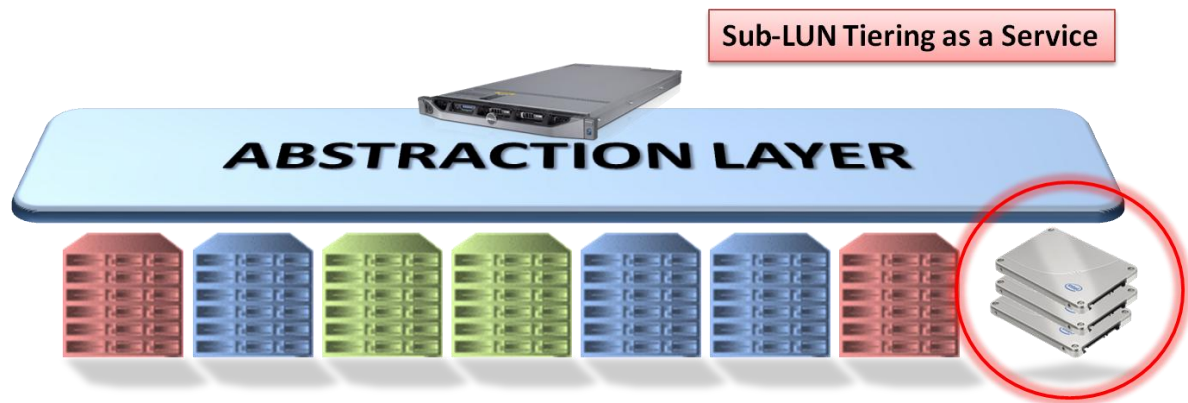
A potential side benefit of this raw speed improvement is the possibility it enables for eliminating cost-ineffective short stroking arrays. With a speed improvement of up to 3X, short-stroke arrays may become a footnote in storage history.

Speed through Tiering

In addition to improving the overall speed of storage infrastructure using adaptive caching, storage virtualization also enables the tiering of storage without requiring the re-cabling or re-addressing of arrays in the infrastructure. Virtualization enables the administrator to set up multiple pools of storage based on criteria such as the performance characteristics of member arrays. Fast storage may be pooled together to provide capture storage for transactional workload, while slower storage arrays may be pooled into a resource that is more appropriate for files. Arrays with high capacity but slow spindles may be pooled together and targeted as archival repositories.

With pools of storage thus defined, migrating data between them in accordance with some sort of hierarchical storage management or archive application is greatly simplified in a storage virtualization setting. No storage is isolated and grouping together arrays with common speeds and feeds characteristics, or arrays interconnected by different interconnecting technologies (Fibre Channel, iSCSI, SAS, etc.) is easily

accomplished. That's good news for administrators who want to deploy traditional storage tiering.



As mentioned above, current advances in tiering, so-called sub-LUN tiering, are gaining mindshare as a mechanism for optimizing the performance of disk drives in an array. This is in addition to caching; it is a purposeful movement of selected data to faster media based on the amount of access being made to that data. When the data cools off, they are demoted to more cost effective disks.

With storage virtualization, sub-LUN tiering can be delivered on a much grander scale. For example, with its burgeoning auto-tiering technology, DataCore Software enables administrators to leverage FLASH SSD drives in its virtualization engine, on the SANsymphony-V server itself, to provide the equivalent of sub-LUN tiering as a service for all of the storage arrays that the administrator has deployed to date. This can save companies the cost of refreshing existing hardware with new SSD-ready gear, while delivering the performance optimization promised from sub-LUN tiering techniques.

Speed through Path Management

The third way that storage virtualization can help improve performance is by helping to resolve the performance of the “plumbing” of the storage infrastructure – the wires, cables, switches and protocols that are used to connect the storage to the servers. At this infrastructure layer, it is typical to find terrible inefficiencies, with some data paths and switch ports hardly used at all, while others are overloaded and congested.

Very few companies have deployed “taps” and meters, or other performance monitoring technologies to understand the I/O traffic traversing their Ethernet and Fibre Channel infrastructure. Often, this reflects the fact that storage infrastructure has evolved over time, using different interconnects and models for connecting equipment together. Today, especially with the onset of server virtualization, the old plumbing models are creating major slowdowns in storage I/O.

How can storage virtualization help to right the situation? In the case of DataCore SANsymphony-V, all connections between storage arrays and the storage virtualization nodes are inventoried when the storage is added to a virtual pool. So the administrator knows how equipment is connected together – what kinds of interconnects and protocols are being used and what switch ports are handling traffic.

Next, DataCore uses a multipathing I/O driver in order to leverage all interconnecting paths between the SANsymphony-V server and physical storage in an intelligent way. SANsymphony-V also examines queue depths to determine which paths to the physical storage are congested, and routes I/O in a balanced way across all available paths to the storage in a target pool. This load balancing is performed behind the scenes, though reporting and monitoring is extensive. So, administration of traffic across links does not require a deep knowledge of link speeds and feeds or fabric design.

FINAL THOUGHTS

The above is not intended to suggest that performance management is a simple matter, even in a virtualized storage environment. Many issues that might cause performance problems have nothing whatsoever to do with storage infrastructure or interconnects. Applications, server hypervisors and even operating systems can impact I/O performance.

The good news is that with the right storage virtualization solution, performance issues that have to do with storage infrastructure can be resolved readily. At a minimum, the technology can eliminate most explanations for I/O slowdown so that troubleshooting can move upstream to other potential explanations. If the impediments to IT performance are linked to storage, virtualizing storage with products such as DataCore Software's SANsymphony-V can help administrators deliver off-the-charts storage performance.

