

# STORAGE SWITZERLAND REPORT

## INTEGRATING SSD AND ARCHIVE



George Crump, Senior Analyst

Storage is experiencing a fundamental technology shift, at the heart of which is the mechanism that's used to physically record data. For the better part of a decade the storage mechanism of choice has been fibre channel (FC) drives, essentially a derivative of SCSI based drives, which came to market a decade prior to that. While this lineage will continue with SAS based technology the shift will come as Solid State Disk (SSD) and SATA enabled Disk Archive systems continue to mature. These technologies offer significant performance and cost advantages, over a purely FC environment, but integrating multiple tiers of storage into existing infrastructures will require some changes.

### **New Storage Technologies**

SSD, essentially banks of DRAM or FLASH memory designed to look like a hard disk, can replace the inefficient practice of striping data across a large number of disk drives to generate performance. The result is actually better performance and a much lower cost per I/O. Disk based archive couples cost-effective SATA drive technology with deduplication, compression and potentially, power-managed drives, to deliver the lowest cost per GB at the lowest power consumption. Trays of fibre channel storage can now be replaced with a single, optimized, disk based archive.

The application landscape for storage has also changed. At the beginning of the decade the storage challenge was databases; how to reliably store their information, provide sufficient capacity and deliver adequate performance. While these issues are still not totally resolved, they are manageable. Unstructured data, or data outside of a database, now presents a greater challenge for IT. The growth in size and number of files has led to file server or NAS 'sprawl', excessive storage purchases and significant time investment managing and protecting the environment.

### **New Requirements, Data Placement**

Most unstructured environments could benefit from multiple tiers of storage, such as putting currently highly active data on SSD, modestly active data on FC and dormant or reference data on disk archive. The challenge however, is no single storage manufacturer has wrapped these three technologies together, into a single, self-managing package, focused on unstructured data. Certainly, some vendors offer SSD, FC and SATA drives, but the integration has historically been lacking. What users are looking for is seamless integration between these three storage types and potentially, other future storage technologies as they mature.

What this integration means is essentially movement of data between storage tiers. In order to be functional and cost-effective, multi-tiered storage systems must have the right data on the right storage at the right time. This 'data placement' issue is central to the infrastructure changes that were mentioned earlier.

Historically, there's also been resistance to multi-tiered storage in many environments due largely to the management required. Getting data onto the right tier often requires manually identifying and then copying the relevant data between the classes of storage involved. In addition to increasing IT workload, moving data to multiple locations can create problems with end users as it complicates their efforts to navigate the appropriate file systems. While there are multiple methods to deal with SSD and disk archive within the context of structured databases, technology for efficient data placement in the unstructured realm has been lacking.

#### New Solution, File Virtualization

File virtualization solutions like those from [F5](#) may be the ideal way to solve the problem of efficient data placement and integrate these different technologies from different vendors in to a smoother, more automated work flow. These solutions allow multiple storage tiers on one, or multiple, NAS platforms to be viewed by the user as a single storage pool, similar to a DNS server. For example, most people don't know the IP addresses of their favorite web sites, they just type the name and it appears. File virtualization performs a similar function but does so on files instead of IP addresses. The user or application requests a file, then the file virtualization system routes the request to where the data is physically stored.

File virtualization systems enable the transparent placement of data to both applications and users between different storage locations. The data placement capabilities are also nearly 'autonomic', essentially

driven by a 'set and forget' policy engine that just executes as needed, initiating at the right time and running in the background. It allows for data placement decisions to be made for the storage manager at the speed changes happen, but still provide the manager with some level of control. As an example, the ability to exclude certain data, based on type or user allows the manager to control data movement in this automated, policy-based system.

Data placement for SSD technology in particular has a real-time component to it. This may require an inline decision making capability that can run at network speed, like a file virtualization system. While the cost per GB of SSD has dropped significantly over the last few years, it's still more expensive than fibre technology. Maximization of the investment is critical. Wasted or even unused space, at SSD prices, is unacceptable. This demands that the right data, the most active data, be on SSD and that the tier be near capacity.

File virtualization also provides a new level of granularity, with decisions enabled at the file level. Older technologies, like some Global File Systems or NAS appliances that make decisions at a directory or file system layer, are too coarse for these tiered storage infrastructures. SSDs are too expensive to hold any extra files that are in a directory with the essential ones. Similarly, placing an entire directory that includes some active files, on a disk archive tier may not only affect the user experience, but may also impact the storage itself. Having to constantly read and write through a deduplication engine can degrade the performance of that system. If the disk archive is also using power-managed drives (MAID), re-waking drives will negate expected power savings.

Another benefit of file virtualization is flexibility. Performance is not solely dictated by the speed of the storage but also the processing power and bandwidth of the NAS. The SSD technology can be on a high speed, high I/O NAS, while the disk archive appliance is typically focused on cost per GB, instead of performance. File virtualization solutions like those from F5 can support a variety of NAS vendors integrating them into a single storage entity. That flexibility then enables the storage manager to select the best of breed solution for each technology.

File virtualization is the solution that can finally make storage tiering a reality. With the addition of SSD and archive tiers into a traditional fibre channel environment, the problem of data placement, or getting the right data on the right tier at the right time, is critical. Unstructured data is overtaking databases as the primary growth area for storage, and new technologies are needed to provide transparent, file-level movement of these data, at network speed. File virtualization solutions, like those from F5, can bring flexibility and performance to multi-tiered infrastructures and support the new requirements of these very different storage technologies.

### **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>.**