

# STORAGE SWITZERLAND REPORT

## FILE VIRTUALIZATION A BRIDGE TO THE CLOUD



### Eric Slack, Senior Analyst

Cloud storage is an attractive option for a growing number of organizations. The potential benefits from outsourcing storage infrastructure are pretty clear, ranging from incremental storage expenditures to near-limitless capacity and reduced overhead. The issue for many, however, is implementation. How to get the appropriate data identified and physically moved to the cloud is not as clear. File virtualization offers a solution with cross-platform storage tiering. By consolidating files onto a single tier based on access frequency, this technology can prepare an organization's data infrastructure internally and provide a bridge to the cloud.

There are a number of definitions for the term "cloud storage", referring to everything from the service itself to the storage infrastructure that supports that service or even the components that make up that infrastructure. The benefits of the cloud concept are fairly consistent, however. Cloud storage promises almost limitless scalability and a 'pay as consume' format, particularly appealing for lower tier and archive storage. Clouds are physically remote, enabling organizations to get their data off-site for DR protection. Being an out-sourced service, they can also reduce storage management and the logistics of housing storage infrastructure.

While the cloud concept makes sense to many organizations, implementation can present some

challenges. WAN latency due to available bandwidth will be unacceptable for many data sets. Concerns about data integrity and data security (even if they're not totally valid) will make the cloud unsuitable for some customers and some data sets. Releasing control of one's data to an out-sourced entity could present issues for some organizations. The logistics of transporting data to the cloud while still maintaining user access will need to be worked out. There will also need to be a consolidated access point to the cloud, as separate connections from each storage platform aren't practical. Most importantly, there needs to be a way to automatically identify data that is appropriate to move to the cloud.

Based on the restrictions of current cloud technologies (including bandwidth), only data with a higher tolerance for latency is suitable for the cloud. This could include data that are the least frequently accessed, examples of which are retained copies of backups, reference archives, supporting data from projects that need to be kept, etc. A storage tiering mechanism which identifies the data for each tier and migrates it accordingly is essential for cloud implementation. For most organizations an incremental implementation is most appropriate, meaning the tiering system must incorporate existing storage - not 'rip and replace' with a new infrastructure.

## Storage Tiering

Storage tiering can be implemented in a number of ways. It could be something as simple as an archive package which moves files to an archive tier (or library) leaving a stub file behind. Many storage arrays include optional software that can identify data appropriate for each tier and move it based on predetermined usage policies. This can be to shelves of disk drives (SATA and SAS/FC, for example) or SSDs within the array architecture itself or tiers on other platforms by the same manufacturers. While there are many platform-based storage tiering options available, these don't address the needs of many organizations. All the data that's under consideration for migration to the cloud probably won't be on a single storage platform, for example.

Most of the current migration solutions depend on an understanding of the storage hardware or file system that the data will reside on. When it comes to the cloud storage, that understanding will likely not be available and the infrastructure cloud providers use at any given time may evolve as well. And, consistent with the out-sourced philosophy, users may want the freedom to switch cloud providers and not be 'locked in' to one provider. These factors point to the need for a cross-platform, file-based system for managing data across different tiers of storage from different manufacturers.

### File Virtualization

File virtualization technologies, like those from [F5's ARX](#), can provide a solution. File virtualization abstracts the physical location of a file from the end user requesting it. Similar to DNS, it removes their requirement to know the physical location of each file. The file virtualization appliance receives all file requests from users and routes them to the storage devices that currently hold each file. The application to storage tiering is obvious, as each back-end storage device can house a different storage tier, regardless of manufacturer. And the virtualization appliance can apply file moves between these storage devices as

needed according to tiering policies - independent of and unknown to the end user.

Implementation can be as simple as setting up the existing storage as tier 1 and creating a second tier from an older array. With the implementation of a file virtualization appliance, the tiers could be established, a threshold set at a certain age for file access and the system be allowed to move files to the second tier as they 'age out'. Compared with a separate data migration project, this method is much more accurate, as it only moves files based upon actual history, not IT's 'best guess'. In addition, this method will continue to automatically migrate files that meet the criteria over time.

The important thing is to get started. Like saving for retirement, taking the first step can be the hardest. With a simple access-date policy on data movement, starting the tiering process can be very simple. There's no need to organize data, no need to bother data owners or set up a comprehensive 'cradle to grave' scheme to manage the data lifecycle. File virtualization automatically moves data down in tier when its use drops and back up when it's accessed again. Most importantly, data owners are unaware that anything has changed - they can continue to access their data in the same manner regardless of where it's stored.

Organizations can continue to add more tiers as they become more familiar with file virtualization. For example, a 'tier 0' SSD array can be utilized to add a performance dimension. Like the process to set up the original archive tier, the hottest data can simply be pulled off and put on solid state drives. Similarly, additional lower tiers can be added as needed to fine tune the access times for various file types and maximize savings. The point to make is that once the file virtualization system is in place, adding tiers means little extra cost to the overall system.

When the time comes to implement the cloud, it too amounts to just adding another tier of storage. The internal infrastructure is already in place. All files in the environment that meet the requirements for the new 'cloud tier' will be moved, same as if it was another internal tier. Most cloud archive providers will supply the customer with an appliance that acts as a gateway to the internet attached storage. These appliances typically translate from the internally understood NFS and CIFS protocols to a more internet friendly protocol like WEBDAV or REST. In the future the file virtualization appliances themselves may be able to do this translation and eventually provide robust support for the cloud vendor's API set. Doing so would eliminate the need for additional appliances, enable cloud mirroring and allow the insertion of custom metadata to improve retrieval from an aging archive.

## **Conclusion**

Cloud storage is an ideal extension and possibly eventual replacement to internally archived data. The number of storage options will continue to grow at both the high performance, premium cost end of the storage spectrum and the high capacity, low cost end of the spectrum. The key to supporting these options, whether a cloud storage initiative is planned or not, is to lay the foundation to be able to support multiple tiers of storage. File virtualization is a simple, automated way to lay that foundation.

### **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>.