Is it Time to Reboot Primary Storage?

by Curtis Preston, Senior Analyst

Primary storage has remained the same for quite some time. It comes in two primary forms: block or file, also known as SAN or NAS. Yes, each of these two forms can be filled with a mix of ingredients, including SATA drives, SAS drives, and flash drives. You have all-flash SAN/NAS, hybrid SAN/NAS, and SATA SAN/NAS devices. But in the end, it basically comes down to the same two choices we had 20 years ago, "Do you want block or file?"

In contrast, other storage markets have been heating up for a while. We have multiple new ways to store, protect, and share data that are offering customers many ways to save money while keeping their data safe at the same time. Why doesn't primary storage incorporate these features? Is it time for a reboot?

Status quo

The current primary storage products are all designed around the idea that all data resides on primary storage all the time. While companies selling primary storage may offer secondary storage products, the primary and secondary products do not typically integrate. In fact, many such companies tend to downplay their secondary storage offerings, as any acknowledgment of such products might reduce the sales of primary storage. Similarly, while there are a lot of options in the cloud, primary storage vendors tend not to acknowledge the cloud as a destination for data. Even if they offer it as a service, there is typically no direct relationship between primary storage in the cloud. The products tend to be point products designed for specific uses: primary storage, secondary storage, and the cloud.

First Get Flash Right

You would have to be hiding under a rock to not know flash storage is making significant inroads into the primary storage market. Flash storage is very fast, and it is also very reliable (although not infallible). Flash is so popular there are a number of storage products designed entirely around flash, known as all-flash arrays. But even the most staunch advocates of flash would acknowledge that not all applications need the performance flash offers. And since flash cost significantly more than magnetic disk, this is an important thing to acknowledge.

The reality is the performance needs of most data can be met with traditional disk. The challenge comes in determining which applications should go where. This is where hybrid technologies that use a combination of flash and magnetic disk come into play. They automatically place data that needs high-performance on flash, and more typical data on magnetic disk. This offers the best of both worlds, where each application gets the kind of storage that suits it best without wasting money on more expensive storage for applications that do not need it.

Storage for How Workers Work

Another interesting product that has taken off in the last decade or so is the concept of *file sync and share*. The advent of cheap computers allowed a lot of people to have their own personal laptops and PCs at home, but this created the demand to be able to synchronize your files across all of these devices, and Dropbox was born. While effective as a product, the challenge with the success of Dropbox and products like it is the creation of shadow IT. It's one thing to synchronize your personal data between two laptops that you and your spouse own. It's an entirely different product when the data being synchronized is company data. This creates compliance, intellectual property, and data protection issues. When data protection is put in the hands of those who do not fully understand it, bad things can happen. For example, consider the all too frequent example of a customer who deletes data once it has been synchronize to Dropbox, because it is now "in the cloud." (Since Dropbox is a sync client, this action deletes the files in Dropbox, too.)

That is not to say file sync and share is not a good idea for company data. What's not a good idea is file sync and share of company data with no oversight by the company's IT department and data protection professionals. What organizations need now is something we refer to as enterprise file sync and share (EFSS). The problem is most primary storage systems and NAS systems in particular ignore EFSS or the vendor tries to meet the requirement through an add-on product instead of integrating the capability directly into the primary storage system.

Of course many organizations have multiple offices and those offices need to share production data between sites. A modern primary storage system should be multi-site aware and have the ability to distribute data between sites based on policy, while at the same time enforcing version control. Multi-site distribution complements an EFSS strategy by not requiring all users have all data with them all the time. Instead they can access the data they need from a local primary storage device while in the office and only use EFSS for data they need to take with them while away from any office.

Intelligent Scaling

Another important advancement is *scale out storage*. Instead of the traditional monolithic designs of traditional storage arrays, scale out storage systems are built around many smaller systems built on industry-standard off-the-shelf hardware significantly reducing cost. This class of storage helps to reduce cost in a number of ways, the first of which is the use of off-the-shelf components instead of custom hardware. The second reason is scale out allows customers to buy what they need when they need it and add capacity when necessary. This is much better than the traditional approach of buying everything up front by using an educated guess at eventual capacity needs. To some extent though, scale out primary storage feeds the idea that all data must sit on primary storage for all time. If primary storage could be integrated with a scale out secondary offering then there may not be a need for the primary storage tier to scale out.

Object Ready Primary Storage

An example of scalable secondary storage is scale out storage is *object storage*. It's an entirely new class of storage completely different from the traditional choices of block or file. An object most closely resembles a file, in that it is discrete element that carries a collection of related data. (Most objects are storing what most would think of as files, but it is storing them as objects.) But an object storage system works very differently than a file system. Each object is given a unique identity based on its contents, not its location.

This identity is based on a cryptographic hashing algorithm that provides a value unique to that object and the metadata associated with that object. By design, it performs object level deduplication, since two objects with the same content would create the same hash and therefore have the same identity and only get stored once.

Giving each object a unique identity allows the system to automatically check the integrity of each object anytime it wants by recalculating the hash based on the current contents of the object and comparing that to the original hash. Any difference in the hash means the object is corrupt and the system needs to replace it with another copy. This makes it very safe to store objects for a very long time without fear of silent data corruption.

Unfortunately, many companies have yet to implement any kind of object storage system. This is usually because they don't yet have an application that knows how to talk to object storage. This is why we see the creation of many NFS and SMB gateways to object storage systems, allowing file-based products to continue speaking that language while the files are translated into objects stored on the object storage system and automatically protected with its features.

The final thing becoming common in newer storage systems is integrated data protection of various types. Almost all new storage systems have some type of way to recover from accidental or malicious deletion via volume level snapshots or object level versioning. A related feature is the concept of a WORM file or object that for compliance purposes even an authorized user is not allowed to delete. Also becoming more common is multisite and n-way synchronization, which allows customers to store their data in multiple geographic locations with ease.

What's lacking is a bridge between the primary storage system and its use cases and the secondary, object based storage system and its strength. IT professionals are forced to spend time identifying and moving data between the two storage types. Since time is the one thing that most IT professionals don't have, the movement of data does not occur and as a result primary storage continues to grow out of control.

Primary Storage Reboot

Imagine a storage product that provides high-performance flash storage, magnetic disk for unstructured production data, and object storage for unstructured reference data – automatically placing each file or object on the most appropriate tier, eventually automatically archiving it and reducing the consumption of physical storage. It would of course protect the data at every stage with snapshots or versioning, multi-site replication, and hash based protection for long-term storage. It would also allow customers to store data where they wanted using enterprise file sync and share services, without placing the data at risk. All those features would be worthy of a reboot.

Sponsored by [Nexsan](#)