Storage Switzerland, LLC

New Requirements for Enterprise Backup

Prepared by: George Crump, Lead Analyst
Prepared on: September 2014

For a disk backup appliance to be deemed "enterprise" it had to be able to perform well and provide more capacity expansion than non-enterprise class appliances, while still being almost as cost effective as its mid-range brethren. But with the growth of virtualization, the explosion of unstructured data and the proliferation of big database applications, the game is changing. The future of disk backup appliances will require that they provide more than just these basic features, instead adding five new capabilities to keep pace with enterprise demand.

#1 - Throughput

The ability to move data into and, more importantly, out of the disk backup appliance will become increasingly critical. While software solutions have provided many features, like deduplication and change block tracking, to reduce the network bandwidth and ingest performance requirements, none of these features are effective during a recovery. In most all cases the data needs to be transferred back to the primary host. A key requirement for future enterprise backup appliances will be to backup and recover data at extremely high rates of speed, in the range of hundreds of TBs per hour, instead of dozens.

TBs per hour is the universal way to express the performance of a disk appliance. It is somewhat misleading, however, since most organizations don't need to transfer hundreds of TBs each hour. But what this means is that a 500GB virtual machine can be restored in a few minutes.

#2 - Mixed Use Case, More than just Backup

Enterprise disk backup appliances must also be able to store more than just backup, with the obvious use case the merging of archive and backup into a single storage system. This can be driven by separate archive and backup applications storing data on the same device, or by allowing the backup application to act as the archive application as well.

In the past the 'best practice' was to keep the backup and archive applications (both software and hardware) as separate entities. Archive software required a more sophisticated means of tracking historical data and archive hardware needed to manage millions, if not trillions, of small files. On the other hand, backup software historically needed a simple means of tracking data and backup hardware needed to deal with large sequential I/O.

Now, modern backup software applications have very robust database architectures that can track years', or even decades', worth of information. This means that the next generation enterprise backup appliance needs to be able to support these two different types of I/O.

Another function that's becoming increasingly common is the ability to support the recovery-in-place capabilities of some VMware-specific backup applications. This means, again, that the backup appliance must have the raw horsepower to provide respectable performance for any virtual machine that the system is supporting, plus have high availability features so that the VM being recovered has the same safety net of protection that it has when running on primary storage.

#3 - High Availability

The future enterprise disk backup appliance will be more than a destination for a second copy of data. If it is also going to be used for data archiving and for emergency VM hosting then its ability to continue to keep data available, even during a failed state, is critical. This means more than eliminating single points of failure. For the enterprise it means multiple points of redundancy which allows for multiple components to fail without the loss of data or even access

to data. It also needs to have automated healing to restore high availability with remaining resources, even prior to replacement of failing components.

#4 - Incremental Capacity Scaling

The data center-wide growth in storage capacity requirements is driving the demand for petabyte-scale storage systems. But the growth in unstructured data, data often generated by machines instead of humans, is of greater concern. The problem with storing this data is often compounded by the fact that much of it needs to be stored indefinitely, so it can be mined in the future.

At the same time, IT budgets remain flat and data centers can't afford to purchase hundreds of TBs of capacity upfront. They need to incrementally add more storage as the demand warrants to support growth as well as changes in business conditions such as mergers or new application requirements. The ability to start small and grow is critical, as is the ability to add that capacity without having to redirect backup jobs.

#5 - No More RAID

One of the single biggest challenges of providing true high availability is the ability to recover from a hard disk drive failures. The problem for most disk backup appliances, even ones that currently claim to be "enterprise", is that they use traditional RAID 5 or 6 as their protection from drive failures. These same systems also use high capacity disk drives to keep costs down; and to meet the capacity demands described above they use a high number of those drives. The net effect is that drive failures can be a frequent occurrence and the time it takes to rebuild the data that was on one of these large drives can take days or even weeks.

The future enterprise-class disk backup appliance will almost certainly need to have some form of erasure coding in order to meet the demand for high availability. RAID's limitation is that its lowest level of granularity is the drive itself. If the drive fails the entire drive has to be read from end to end, meaning an empty drive takes as long to recover as one that's full. With erasure coding just the data that was on that drive needs to be rebuilt. This differential can be significant when dealing with multi-TB sized drives.

Summary

Ironically, the future of enterprise class disk backup appliances has less to do with their ability to store backups and more with their ability to respond to a changing environment. This an environment that may call on that appliance to occasionally host a virtual machine, store petabytes of data or to provide new levels of backup **and** recovery performance. Companies like NEC with their HYDRAstor HS8 line of enterprise backup appliances are well equipped to meet the new enterprise demands and deserve consideration as IT planners look for their next generation disk backup appliance.