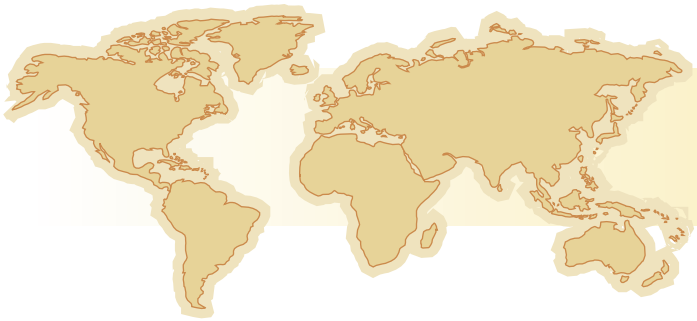


DCIG INC

SOLUTION BRIEF

Disk-Based Data Protection: Five Essential Enterprise Considerations

December 2006

Datacenter Infrastructure Group Inc.
Jerome M. Wendt
President and Lead Analyst

As a former storage administrator and engineer, I have lived through my share of backup nightmares. Now in my role as an industry analyst, I continue to hear stories from users about their ongoing struggles with backup. If you are one of the lucky ones, you have never experienced a glitch but more than likely, your level of frustration with managing backups equals or exceeds mine.

My frustrations with backup go back to my earliest days as an administrator. For me, it was a mind numbing exercise driven by an underlying sense of fear. Daily tasks involved verifying backups completed, labeling tapes, sending them offsite, rotating out old ones and then repeating the whole process again the next day. However, since I rarely had the time, budget or equipment to verify a backup tape worked, there was always this nagging fear in the back of my mind about how I was going to recover the data or if I even could recover the data. So whenever I did need to recover data, I always hoped against hope that I could locate the correct tape and recover the necessary data from it.

Of course then along came Sarbanes-Oxley, SEC 17a-4 and a host of other regulations. Now suddenly every lost tape became a matter of national security since misplacing or losing any tape with sensitive data had the potential of making national headlines. In one instance, a company I worked for shipped tapes containing sensitive data from one site to another that somehow got lost. The worst of it was no one knew exactly where in the shipping process the tapes were lost or why it happened. My company knew only that somewhere between point A and point B, the tapes turned up missing, and it was our responsibility to find them. This loss started a chain of events that involved individuals flying around the country checking warehouses, postal facilities, airport hangars — even snow banks — trying to find the missing tapes.

While I wish I could say incidents like this are a rarity, now that I work as an industry analyst, I know that is not the case. Consider what has occurred only recently: in September 2006, Chase Card Services revealed that in July 2006, they had mistakenly discarded five computer data tapes containing the personal information of 2.6 million Circuit City Cardholders, and these tapes were later found in a landfill. Also in that same time frame, Nelnet Inc, a student loan administrator based in Lincoln, Nebraska, lost a tape containing the personal information of 188,000 student loan customers. And it goes without saying, there are a lot more of stories like this.

So after living, working and breathing these experiences for many years, I am now using my hands-on exposure to these technologies in my new role as an industry analyst, dedicating myself to helping users find better data protection options. In respect to improving data protection, most organizations have already figured out that disk backup is the way they want to do their backups short- and long-term. But two attitudes prevail in corporations.

We love disk backup!

Disk-based backups and restores provide corporations with numerous advantages over tape. The speed and reliability of disk backups shorten backup windows, allow backups to complete successfully and greatly reduce the amount of time and stress that administrators experience troubleshooting failed backups. But once disk-based data protection technologies are deployed and administrators start working with them, the downsides of disk-based data protection emerge and with that emerges the other prevailing attitude.

We need a better way!

Current disk-based technologies such as Virtual Tape Libraries (disk-as-tape) and disk-as-disk appliances cost money – a lot of money. With users keeping anywhere from five to 20 or more backup copies of their data and the cost ranging anywhere from $5-20GB for disk storage, the cost for extra disk capacity can rapidly mount up. In fact, a recent IDC Study cited cost as the top reason why disk-based backup and restore is not used or used only in a limited fashion in data centers.[1]

Why not just buy more tape?

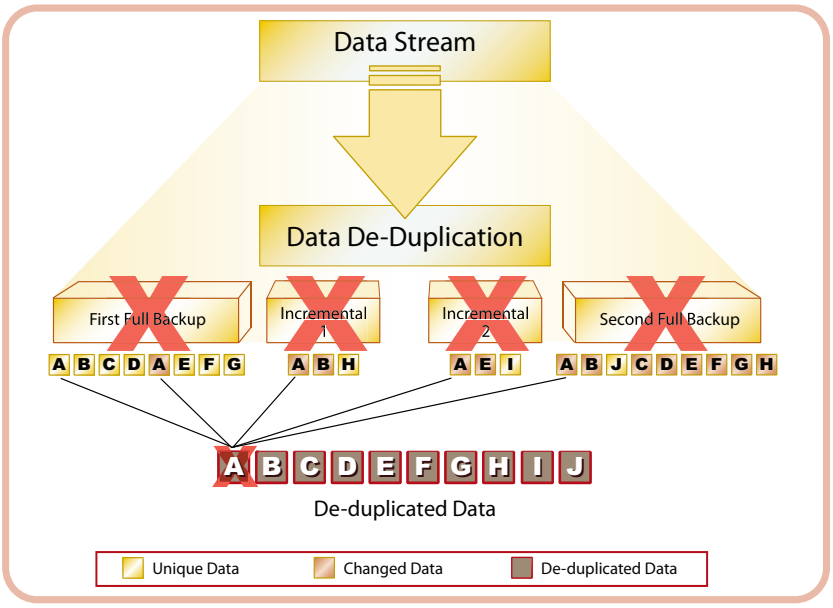
Tape drives are faster and new tapes hold more capacity than ever. The problem is, servers can't feed backup data streams (typically 30-50MB/sec and <10MB/sec in environments with lots of files) fast enough to take advantage of these increases in speed and capacity.

Upgrading to new tape technology may actually result in slower backups than before due to shoe shining and back hitch problems.

Once the data is backed up to disk, users have no easy way to move the data offsite. Replicating the data to another disk array only amplifies the cost issue and introduces network bandwidth considerations. Users can also opt to copy the data off to tape and move it offsite that way but that re-introduces the tape management problem which was part of the company's motivation for eliminating tape in the first place.

New de-duplication technologies partially address both of these problems, but they have their own set of challenges. Current vendor implementations are usually unacceptably slow as they can't provide sufficient processing power to de-duplicate incoming data, and they stretch the reliability limits of RAID. In the case of RAID, the failure of multiple SATA disks in a RAID group could cause the loss of a chunk of backup data that would in turn compromise the integrity of hundreds of backups. Figure 1 illustrates the inherent risks with relying upon RAID technology for data resiliency.

[1] IDC Special Study, Disk-Based Data Protection, 2006: Multiclient Study, Final Report, Doc #202680, August 2006



Data de-duplication streamlines disk utilization and thus increases cost effectiveness, but can create unacceptable risk for data loss.

The failure of multiple SATA disks in a RAID group could cause the loss of a chunk of backup data that would in turn compromise the integrity of hundreds of backups.

In this figure, if de-duplicated data chunk "A" is compromised, you would lose the key to your first full backup, your first incremental backup, your second incremental backup, AND your second full backup as well.

Figure 1. Data De-duplication Risk

So with none of today's technologies truly satisfying all enterprise data protection requirements, where can we go from here to resolve the backup and restore challenge? To make that decision, I recommend stepping back and evaluating disk-based data protection solutions based on five essential considerations critical to making a long-term, strategic investment. The following five considerations allow today's IT departments to balance their current responsibilities of reducing cost and risk while assuming new roles of enabling corporate growth:

- Manageability
- Scalability
- Availability
- Future Proofing
- Affordability

Manageability

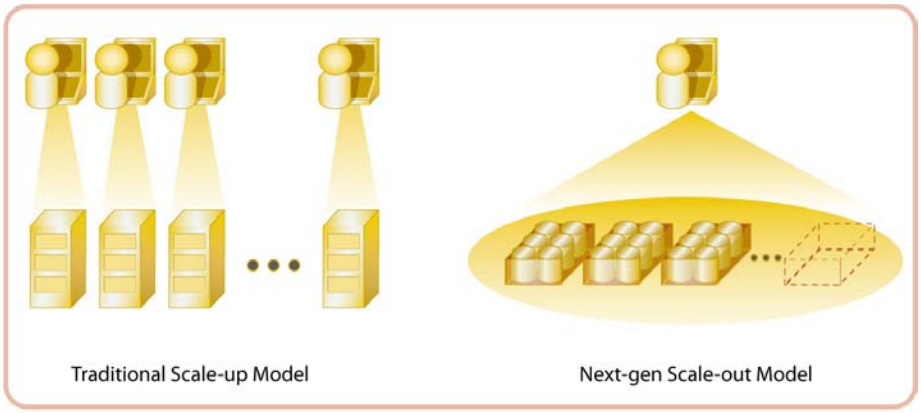
Recommendation: *Organizations should look for products that are self-managing, self-tuning and provide single-instance management.*

Manageability is a key consideration because companies often resist deploying new technologies precisely because they know they require a lot of staff time to manage, configure and monitor. With staff already stretched thin and new staff difficult to justify for functions that are often viewed as non-strategic, you can't afford to deploy a sub-optimal product. New products need to minimize routine and time-consuming tasks such as capacity planning, LUN creation and allocation, and performance optimization to free them to devote more time to more strategic tasks such as optimizing and improving their current environment.

Management complexity is also a critical consideration, and one that organizations typically overlook. If "loved the first, hated the third" sounds familiar, you are not alone. Even the simplest-to-manage device becomes a burden to manage when there are multiple physical instances to manage. Rather than just completing a task and moving on to other tasks that add business value, your job somehow devolves into device management. New disk-based data protection platforms should support the ideal of "One Instance" management. In this way as growth occurs, management complexity and staffing overhead do not, thereby avoiding the management nightmares most current approaches create.

Key questions to ask when evaluating products for manageability are:

- What administrative tasks are required for daily system management?
- How difficult is it to add more storage capacity?
- As it grows, does it retain its "One Instance" management feature?
- Can it keep current as technologies evolve without requiring forklift upgrades?



Traditional Scale-up Model

- Islands of space & performance capacity
- Labor intensive
- High cost management

Next-gen Scale-out Model

- Single capacity pool
- Self- managing automation
- "One-Instance" management

Figure 2. Ideal Disk-based Data Protection Architecture

Scalability

Recommendation: *Start small and grow based on business needs but verify that you can scale capacity and performance components independently.*

When evaluating a disk-based data protection technology, scalability has two dimensions: capacity and performance. Scaling capacity is required to support data growth while scaling performance is needed to support shrinking backup and restore windows in the face of continual data growth. With most organizations experiencing year-over-year data growth rates in the 20% or more range, and every TB of online data equating to 10 – 25 TBs of backup data, you need to verify that your data protection technology can keep up with your corporation's growth rate.

The ability to start small is another critical feature. You'll want to start with a small "taste" of disk-base backup/restore, and then add capacity as your use expands. Since it's difficult to know whether or not a particular product will work in your environment until you install it, the risk of failure could overwhelm its potential benefits.

Using an architecture that allows you to start small and grow large allows you to understand the idiosyncrasies of the system, experience some initial successes and not put your entire enterprise (or pocketbook) at risk. Then as performance or capacity demands for backup or restore functions grow, or as other systems come up for technology refreshes, you could grow the system at that time.

Also critical to scalability is a system's ability to easily and non-disruptively grow. This feature is especially important for a data protection platform as you have backups running throughout the night with restores and replication tasks occurring throughout the day. This leaves no time for downtime to add new capacity to a system.

Key questions to ask when evaluating products for scalability are:

- What are the product's practical capacity and performance growth limits?
- Can you scale capacity and performance independently of one another?
- What is the performance hit if data de-duplication or data compression is introduced?
- Can you non-disruptively add capacity or performance?

Availability

Recommendation: *Choose a system that satisfies your organization's hardware availability and data availability requirements.*

Backups are like insurance policies — they provide data protection in the event something goes wrong. However, would you buy an insurance policy for your house or car knowing that it lapses for certain periods of time? Of course not, yet that's how most organizations operate today — not really knowing for sure if their backups are available or how to fix the situation if they are not.

One aspect of availability is the system itself – the system must remain up and running so it can support backup and restore operations. In order to meet these needs, the disk-based data protection platform must not have any single point of failure; must support online software upgrades; and must provide diagnostic capabilities that allow it to pro-actively identify failing or faulty components and either shut them down or heal them. The system also needs to send out alerts to initiate the repair process of the defective part in order to prevent future system impact and ensure continued 24x7x365 uptime.

Data availability is the second element of availability. In this case, it means keeping backup data readily available if a restore is needed. This feature is especially pertinent with disk-based data protection platforms since de-duping introduces new levels of risk that most users are not aware of or, if they are aware of them, fail to understand the true potential impact. With de-duped data, the loss of a single block of backup data can impact dozens upon dozens of backup images and prevent the restoration of any of them. The loss of an entire disk magnifies this problem as it can prevent the restore of hundreds of backup images and, as disks grow in size, so too will the magnitude of the potential impact.

Many users believe that the hardware RAID built into many subsystems today provides sufficient protection for their data. But in today's de-duplication world, traditional RAID configurations fail to provide adequate levels of protection. Disk-based data protection architectures need to account for these new risks by providing higher levels of data resiliency without increasing storage overhead to the point that it negates the cost advantages of using de-duping technology.

Key questions to ask when evaluating products for availability are:

- How many hard drive failures can the system sustain before data is lost?
- What is the storage overhead for the level of protection provided?
- What is the performance degradation during a disk rebuild?
- Can the system be used for backups or restores during a disk rebuild?
- What health checks and monitoring standards are supported by the system?

Future Proofing

Recommendation: *Choose a system that provides long-term technical and financial investment protection.*

One angle that organizations often overlook when selecting any given technology is: how often do you want to revisit a specific technology choice, especially one as critical as disk-based data protection? Once any technology is selected, this event starts a chain of other events in the background that become difficult and costly to undue. For instance, users are trained on the device's interface. Interoperability with other applications is tested and verified. Support personnel become familiar with the corporation's daily operations and support procedures. Scripts are written that take advantage of advanced features of the device.

Once these processes are in place, the last thing anyone wants to do is derail them. But unless the disk-based data protection platform offers non-disruptive hardware upgrades to allow a data center to take advantage of new technologies such as larger disks and faster processors at better prices, replacements down the road become inevitable. So organizations need to think strategically about whatever disk-based protection product they purchase because, ideally, it will be with them a long time.

The ideal product should account for these ongoing and inevitable evolutions in technology. Doing so provides an organization with needed levels of future technical and financial protection and requires a large vendor who offers the resources to support both the product and your company short- and long-term.

Key questions to ask when evaluating products for future proofing are:

- How large and financially stable is the company offering the product?
- Can the company provide adequate 24X7x365 support?
- What is the supported upgrade path between system releases?
- Does the system support heterogeneous disk sizes and speeds?

Affordability

Recommendation: *Choose a disk-based data protection whose value proposition is as viable in its fifth year as it was in its first*.

Affordability underpins all of the previous points and almost goes without saying when referencing them. For instance:

- ***Manageability*** means less staff time spent monitoring system utilization and optimizing system resources. That frees individuals to spend more time pursuing more strategic business objectives such as lowering other operational costs or opening up doors for new revenue-producing opportunities.
- ***Scalability*** translates into less time spent planning technology refreshes and allows users to add capacity in a "pay as you grow" fashion. It also allows organizations to offer disk-based data protection at a price level that almost any department in the organization can afford and easily utilize.
- ***Availability*** provides higher levels of uptime during periods of backups and restores, thus reducing the time required to troubleshoot failed backups and restores.
- ***Future proofing*** minimizes the time companies need to spend evaluating competing technologies. It also allows the organization to focus on taking advantage of the product's full capabilities without worrying that the time they invest in developing them will be negated by a sudden shift in technology.

Key questions to ask when evaluating products for affordability are:

- ✔ What amount of disk overhead is required to ensure data resiliency?
- ✔ Are hardware upgrades as economical in year five as there were in year one?
- ✔ As it grows, will you manage the product or will it manage you?

Introducing HYDRAstor from NEC

Being well acquainted with most disk-based data protection products on the market and having written about them on multiple occasions, I was of the opinion that no single product addresses every enterprise concern; however, HYDRAstor from NEC has changed my opinion. A grid-based storage architecture, HYDRAstor is built from the ground up to address the manageability, scalability, availability and affordability needs of data centers while positioning the IT department as an enabler of future growth within their corporation.

HYDRAstor satisfies the five essential enterprise considerations for disk-based data protection in the following ways:

- ***Manageability.*** HYDRAstor is self-managing, self-tuning and provides the "One Instance" management needed as the product grows.
- ***Scalability.*** HYDRAstor scales out as a single instance from 200 MB/sec to 70 TB/hour in performance throughput and from TBs to PBs in total storage capacity.

- ***Availability.*** HYDRAstor ensures data availability and integrity by allowing the failure of up to three disk failures (or even more, if you set its "resiliency dial" higher) without data corruption or loss, and without paying a huge overhead penalty, while also delivering hardware availability by enabling organizations to perform non-disruptive hardware and software upgrades.
- ***Future Proofing.*** HYDRAstor is able to ride the continuing hardware and software technology curve through non-disruptive and ongoing upgrades of hardware and software.
- ***Affordability.*** At under a dollar a GB, HYDRAstor's TCO is at or below that of tape while eliminating much of the management overhead and hassle associated with tape.

With HYDRAstor's low cost per GB for disk, it opens the door for companies to completely eliminate tape from their environment in all areas, not just primary backup and recovery. By incorporating patent-pending de-duplication technology with an evolutionary grid-based architecture, HYDRAstor allows corporations to nearly infinitely scale out capacity and performance at the primary site while economically configuring a secondary site that can serve as a remote replication target.

Furthermore, through inline de-duplication at the sub-file file, HYDRAstor solves two data protection issues – efficient storage utilization and efficient WAN replication – in one fell swoop. HYDRAstor thus effectively nets your organization "two for one" cost savings by eliminating the need to implement a WAN compression solution on top of your data protection solution – you get both in one package.

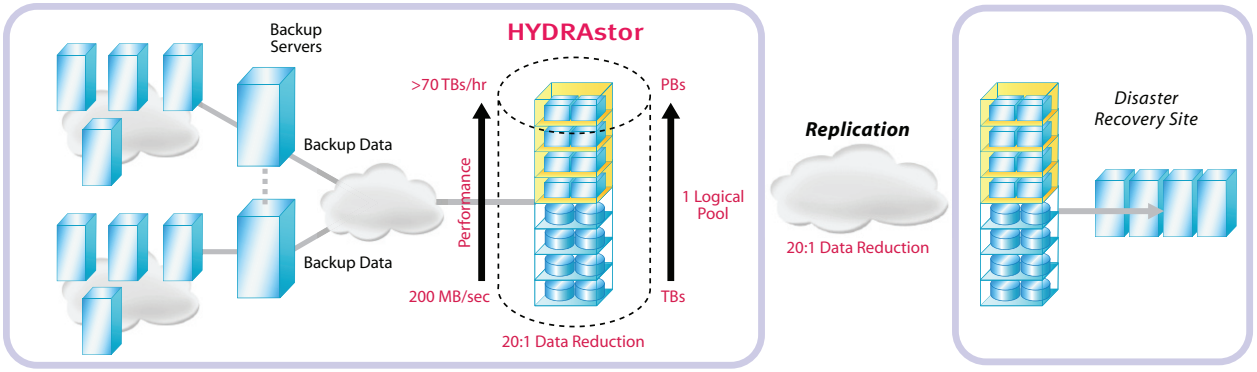


Figure 3. HYDRAstor platform, a scaleable, manageable, affordable backup solution.

But just as important as the technology is the company that stands behind it. NEC is not a fly-by-night outfit operating on a shoestring budget that hopes to use your data center as a test bed for HYDRAstor. NEC operates on a global basis and understands the nuts and bolts of data centers and the unique requirements that a disk-based data protection product must satisfy in those environments. Having invested a significant amount of research and engineering in HYDRAstor, NEC is presenting a truly innovative offering from a Global 200 company that combines the best of cutting-edge technology and old fashioned support. And in a market filled with a lot of noise from products that create as many problems as they solve, HYDRAstor from NEC is one product that may finally signal the beginning of the end for user frustrations with backup and restore.

DCIG Inc.
Jerome M. Wendt
7511 Madison Street
Omaha NE 68127
Office 402.884.9594
Cell 402.203.1181
Email jerome.wendt@att.net