Straight Talk

ABOUT DISK BACKUP WITH DEDUPLICATION

Avoid the 10 mistakes you could be stuck with for years.

Bill Andrews

Straight Talk

ABOUT DISK BACKUP WITH DEDUPLICATION

Avoid the 10 mistakes you could be stuck with for years.

Bill Andrews

Contents

Introduction

The movement from using tape for backup to using disk for backup is well underway, with about 70% of IT organizations now using disk staging in front of a tape library onsite and approximately 15% of IT organizations having replaced their onsite tape backups with disk-based backup with data deduplication.[1] Industry sales of tape libraries have been on the decline for the past four years, while sales of disk-based backup with deduplication are seeing double-digit growth year over year.

Many factors should be considered when making the move to disk with deduplication, as it is not as straightforward as simply buying some disk. Unlike primary storage, from a technology perspective, disk backup solutions are based on hundreds of thousands of lines of software code to as many as a million lines of software code. This much software is required to properly work with a backup application, deduplicate the data, store the deduplicated data and replicate the data. Disk-based backup is not merely a commodity storage solution or a NAS share with storage plus deduplication. Disk backup requires a purpose-built approach that positively impacts backup performance, the backup window, the time to restore traditional or image backups, make offsite tape copies, perform instant recoveries and recover from a site disaster. The purpose-built approach also impacts the long-term costs of expansion over time. Some IT organizations have taken the simple route of purchasing a disk backup system with deduplication from their primary storage vendor, backup application vendor, tape library vendor, or server vendor only to find that they did not ask a fraction of the important questions before buying.

1 2012 CIO Pulse survey of 1,200 IT professionals (IDG Research)

In this guide, we first explain the differences between the various backup approaches. We then provide the list of questions you should ask each vendor about every solution to avoid experiencing the following ten consequences:

1. Inability to operate with or support all of your current backup applications and utilities or those you might use in the future
2. Missing support for all of the pertinent features of your backup application
3. Slow backups, or backups that start fast when you deploy the solution but then degrade over time as data grows, resulting in an expanding backup window
4. Expensive forklift upgrades to replace the controller in order to maintain performance and keep the backup window short as data grows
5. Slow restores of traditional full backups or images
6. Slow auxiliary copy or offsite tape copy
7. Lack of support for Instant Recovery solutions for files, objects, and VMs could mean recoveries take hours versus minutes
8. High cost due to increased equipment needs over time
9. Constant additional purchases of disk capacity due to a poor deduplication rate
10. Use of additional expensive bandwidth due to poor deduplication rates

This guide explains the various backup complexities, enabling you to ask the right questions and make the right decision for your specific environment and requirements.

Chapter 1

Backup to Tape

For as long as most IT professionals can remember, backup has been accomplished by having a backup application copy data every night from all primary storage on the network to tape. The organizations still backing up primarily to tape keep tapes onsite for local restores and store tapes offsite in case of a site disaster.

Most organizations keep multiple weeks to years of backups due to the need to go back to certain points in time and do any of the following:

- Restore from a failure or corruption
- Restore deleted or overwritten files
- Recover data for legal discovery
- Recover data for external audits or recover historical files to meet other user and business requirements

As a result, the amount of backup data can be 30 to as much as 100 times or more the size of the primary storage.

Example:

If the primary storage is 20TB and you keep 10 weeks of retention onsite and 3 years of monthlies offsite, then you are keeping 46 copies of the primary storage in backup storage. This would require almost 1PB of backup storage (1PB of backup storage to keep retention for 20TB of primary storage).

The massive amount of data in backup storage driven by the number of copies kept meant that tape was the only cost-effective solution. Since tape per gigabyte is far less expensive than disk, backup was always written to tape for cost reasons.

However, although tape media is inexpensive, it has many deficiencies:

- Backups fail for a host of reasons.
- Restores fail because of failed backups, corrupted tapes, missing files, blank tapes, and other problems.
- IT staff spends excessive time managing manual backup processes.
- Security is poor as data is leaving the building when tapes are transported offsite. Even onsite, tapes may not be in a secure area.
- Tapes wear and tear. As tapes are reused, they wear or stretch, increasing the chance of failed backups and subsequent restores.
- Tape libraries are more prone to breakage compared to other hardware in the data center because they have many moving parts.

To address the problems of tape, the solution is clearly backup to disk. If the cost of straight disk per gigabyte was the same as tape, everyone would use disk for backup for the following reasons:

- Backups complete successfully since writing to disk is reliable.
- Restores work since the data is reliably backed up to disk, and therefore the restore is also reliable.
- Manual movement and handling of tapes is avoided, as disk drives remain in data center racks.
- Security issues are negated since disk sits in the data center rack behind physical lock and key and within network security.
- Disk is not damaged by heat or humidity since disk is in a hermetically sealed case in the temperature and humidity-controlled data center.
- Uptime is greater because spinning disks fail far less often than tape drives and robotics.

The benefits of disk are clear. The goal is to use disk for backup—and the way to do it is to have the required disk approximate the cost of tape, including libraries, drives, media and offsite storage.

Chapter 2

Disk Staging

Since disk has historically been too expensive compared to tape, most IT organizations found a middle ground by using a small amount of disk between the backup/media server and the tape library to hold the most recent one to two weeks of backup data.

This allowed for faster and more reliable backups as well as faster and more reliable restores. However, keeping more than two weeks of onsite backups on disk would be cost-prohibitive.

Example:

Assume a Monday through Thursday rotation with:

- Full backups of email and databases each night
- Incremental backups on files, where only the files that have changed since the last backup are backed up each night

If the nightly backup is 5TB (about 25% of the full backup) and the weekend full backup (equal to the primary storage) is 20TB, then to keep one week of backups would require:

5TB x 4 nightlies (20TB) + the weekend full (20TB) = 40TB

To keep two weeks of backups for the same 20TB of primary storage, you would need 80TB of disk. For each additional week, you would need to add another 40TB. This is the reason why disk staging only augments tape but does not replace it, and why most organizations keep just one or two weeks of backups using disk staging and then put all the rest of the onsite retention and offsite disaster recovery on tape.

Disk staging is a pragmatic interim step and has eliminated some of the challenges of tape, but it is nonetheless a band-aid and has not eliminated all of the problems. Past the second week of retention, tapes are still used onsite and transported offsite for disaster recovery. Since tape is still used both onsite and offsite, the IT organization has to deal with the same failure rate issues, security challenges, IT management burden and downtime concerns.

The ideal goal is to use disk exclusively, but to do so at the cost of tape.

Chapter 3

Enter Data Deduplication

To comprehensively address the problems of tape, the optimal solution is to back up all data to disk onsite, replicate the backup data to disk offsite, and entirely eliminate tape along with its associated drawbacks—provided, as discussed previously, that the cost is equivalent to that of tape.

Since disk costs more per gigabyte than tape, the only answer is to find a way to use far less disk to store the same amount of data. If straight disk costs 20 times the price of an equivalent-sized tape library plus tapes, then if you can store the total amount of data on 1/20th of the disk, the costs are now equivalent.

Example:

Assume you have 20TB of primary data onsite and you keep 2 weeks of nightly backups at 5TB/night (25% of the total backup) with 10 weekly full backups.

Total disk required onsite to back up 20TB of primary data is:

Two weeks of nightly backups (8 x 5TB)	40TB
Ten weekly full backups (10 x 20TB)	200TB
Total:	240TB

It is clearly not reasonable to place 240TB of disk behind the backup server to back up 20TB of primary storage. The disk cost alone is prohibitive, not to mention rack space, power and cooling.

This is where data deduplication comes into play.

In the preceding example, each week you are backing up a 20TB full and after ten weeks, you have 240TB of data just for the full backups. However, as you back up from week to week, industry statistics show that only about 2% of the data actually changes. So, with straight disk, in effect you are storing the entire 20TB each week when only 400GB each week has changed.

Disk-based backup with deduplication compares incoming backup data to backup data previously stored, and *only* stores the changes or unique data from backup to backup. *This dramatically changes the economics of disk backup*. This is explained in the following example.

Example:

To keep it simple, we will only illustrate the weekly full backup as an example. If in week one, you send in a 20TB full backup, and then each subsequent week you send in the same 20TB with only 2% of the data (400GB) having changed, then you can just store the 400GB that is changing each week.

Total disk required with deduplication onsite to back up 20TB of primary data with 10 weekly fulls is:

First full backup deduplicated by 2:1 (10TB)	10TB
9 subsequent weekly fulls deduplicated by 50:1 (400GB each)	3.6TB
Total:	13.6TB

In this simple example, the 13.6TB of storage with deduplication requires 7% of the 200TB with straight disk if all 10 weeks were stored at 20TB each.

Of course, there is much more to this, which is explained in the sections that follow. In fact, the deduplication algorithms on the market in the leading products offer a 10:1 data reduction ratio to as much as a 50:1 reduction, with an average reduction of 20:1.

Deduplication Approaches and Deduplication Ratios

The range of deduplication from 10:1 to as much as 50:1 is due to two factors:

1. The deduplication product itself: some products offer a more effective and refined approach to data deduplication than others.
2. The mix of data types, which changes the overall deduplication ratio:
    - Compressed data and encrypted data does not deduplicate at all, so the deduplication ratio is 1:1.
    - Database data has a very high deduplication ratio and can get ratios of 100s to 1 to even 1,000s to 1.
    - Unstructured data may get 7:1 to 10:1.
    - Data from virtualized servers often has a significant amount of redundancy and can get very high deduplication ratios.

When you have a normal mix of databases, unstructured data, VMs, compressed or encrypted data, for the leading products you will see an average deduplication ratio of 20:1.

Deduplication is accomplished in three major ways in the industry, as described in the table below for an average mix of data types:

Approach	Block Size	Description	Average Data Reduction
Zone-level deduplication with byte change	~8MB block with variable length content splitting	Identifies large zones using a signature pattern; only saves bytes that change from previous backups	20:1
Block-level deduplication w/ variable length content splitting	~8KB block with variable length content splitting	Identifies block using a signature pattern; only saves blocks that change from previous backups	20:1
Block-level deduplication with fixed block length	64KB or 128KB fixed block sizes	Used on a shared server that is serving multiple purposes (e.g. a media server)	7:1–large block size and fixed block length
	4KB block size	Instead of using 8KB and variable length, uses a much smaller block size to improve deduplication rates for fixed block	13:1 to 15:1–no variable length content splitting

The reason why the average data reduction matters is that a greater average data reduction uses less disk for longer term retention as data grows. It also impacts how much bandwidth is required to replicate data offsite for disaster recovery. Some may look at the up-front cost of the system and make their buying decision based on initial price. However, if the lowest priced system has a poor deduplication ratio, then it will prove to be far more expensive over time due to the cost of disk (since the amount of data and retention will continue to grow) as well as the additional WAN bandwidth that will be required.

In evaluating disk backup with deduplication solutions, you need to first understand the approach the product is using. In order to save both disk capacity and WAN bandwidth over time, choose a product that gets you the greatest deduplication ratio. Ask about the specific deduplication approach, as some marketing literature will report high deduplication ratios using the best data type and not the mix of data that would be encountered in a real production environment. For example, industry marketing literature can say deduplication ratios of “up to” 50:1. The key phrase is “up to,” as these ratios are using a mix of data (e.g. databases) that achieve a high deduplication ratio. Your organization needs to know the deduplication ratio for a standard mix of product data, including databases and VMs as well as unstructured, compressed, and encrypted data.

To summarize, not all products achieve the same deduplication ratios since they use different algorithms for performing deduplication. The *lower* the deduplication ratio, the *more* disk capacity and WAN bandwidth is required over time resulting in *higher* overall costs. *The true cost is the price up front and the cost of disk and bandwidth over time.*

Disk Backup Architectures and Scalability

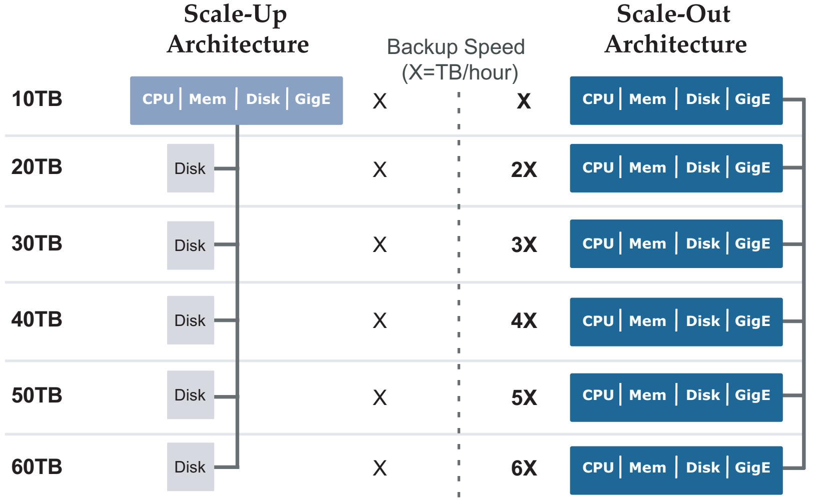
Once you understand the various deduplication approaches and their impact on disk and bandwidth required, then you need to understand the different architectures for implementing disk-based backup with deduplication. The choice of overall architecture is important because it makes a significant difference in overall backup performance, keeping a fixed-length backup window that doesn’t expand as data grows, and the ability of the system to scale without any expensive system forklift upgrades along the way.

In the average IT environment, primary storage data is growing about 30% per year and in some cases up to 50% per year. At 30% per year, the data will double about every 2.5 years. So, if you have 30TB of data today, in 2.5 years you will have 60TB. The data growth needs to be taken into account to ensure the system can handle your data now as well as your data growth in the future. Avoid choosing a system that hits the wall and requires you to replace it before it's been amortized.

Before discussing the pros and cons of the two different architectural approaches—"scale-up" and "scale-out"—let's first define the terms:

- "Scale-up" typically refers to architectures that use a single, fixed resource controller for all processing. All data flows through a single front-end controller. To add capacity, you attach disk shelves behind the controller up to the maximum for which the controller is rated. As data grows, only disk capacity is added; however, no additional compute resources (processor, memory, bandwidth) are added to deduplicate the increasing data load.
- "Scale-out" typically refers to architectures that scale compute performance and capacity in lockstep by providing full servers with disk capacity, processor, memory and network bandwidth. As data grows, all four resources are added: disk capacity, processor, memory and network bandwidth.

See the following diagram for depictions of scale-up versus scale-out approaches.



Unlike primary storage—where data is simply stored, and as your data grows you just add more storage—in the case of disk backup with deduplication, all of the data must be compared before deduplication. This is the first time in storage where processing of the data takes place *before* storing. As a result, this requires compute resources (processor, memory and bandwidth) in addition to storage. Furthermore, as the data grows, the amount of existing data to be compared against grows; therefore, an increasing amount of compute resources are required (processor, memory and bandwidth).

Because the deduplication process of comparing incoming backup data to previously stored data is computationally intensive, it is logical that the more data that comes into the disk-based backup system, the more compute power is required to compare and deduplicate the incoming backup data against the increasing amount of data already stored. To ingest, compare, and deduplicate twice the data with a fixed amount of compute resources would take twice the time as the original data amount. Therefore, systems that do not scale with additional processor, memory, and bandwidth will cause backup window expansion as data grows. The only way to avoid backup window expansion with data growth is to add compute, memory and bandwidth resources along with disk capacity.

In the **scale-up model** (front-end controller with disk shelves), as data grows, the backup window gets increasingly long until the backup window is so long, it runs into end-user production time. To bring the backup window back into the allotted backup timeframe, the front-end controller needs to be upgraded to a much more powerful and faster controller. This is called a "forklift upgrade." The cost to do a forklift upgrade can be as much as 70% of the original cost of the system, since the controller is the most expensive component, and the new more powerful controller is more expensive than the original less powerful controller. The total cost of this approach is not just the initial up-front cost, but also the cost of the subsequent forklift upgrade.

Scale-up systems usually quote raw storage. As a general rule, for systems that quote raw storage, you can actually only store about 60% of the raw storage as a full backup. For example, if the raw storage quoted is 32TB, then you can back up about a 20TB full backup.

Example:

Let's say that you purchase a scale-up system that can expand to 76TB of raw disk storage capacity. The system will be able to take in a full backup that is about 60% of the raw storage or about 47TB. If you have a 28TB full backup and your data grows at 30%, then you will be facing a forklift upgrade in about 23 months.

The net is that in 23 months you will need to replace the front-end controller with a faster, more powerful (and far more expensive) new front-end controller.

In the **scale-out model**, since full servers are added as data grows, if the data doubles, then the disk, processor, memory and bandwidth all double. If data triples, then all four resources triple and this continues as you keep growing data and adding scale-out appliances. Servers are added into a grid architecture with a single management interface and automatic load balancing to distribute the data and processing across all the servers. This approach has two major advantages. The first is that the backup window stays fixed in length—as the workload increases, so do all the resources including disk and compute capacity. The second advantage is that there are no forklift upgrades. The system comes with disk and associated compute capacity, thereby scaling easily with data growth.

The architectural approach impacts planning up front and over time. With a scale-up system, you risk over-buying or under-buying. You need to forecast data growth to ensure that you are not buying a costly system with excessive capacity relative to your needs today, or buying an under-sized system that will need a forklift upgrade in just three to six months as your data grows. With a scale-out approach, you don't need to plan, because as your data grows, the appropriate compute resources are supplied along with disk capacity. This approach, therefore, naturally scales with data growth.

The net is that the architectural approach is critical to the backup window and whether it stays fixed in length or keeps expanding. The architectural approach also affects long-term costs, since hidden down-the-road controller forklift upgrades are extremely expensive.

Disk Backup Architectures and Restore Performance

The key reason for performing backup is so that critical data can be recovered quickly in the event of human error or a system failure. Full system restores are the most critical restores, as hundreds to thousands of employees at a time can be down when a full system is down. The longer it takes to recover, the more lost hours of productivity. Two disk backup architectures are commonly used, with different resulting implications for restore performance. *Inline deduplication* is performed between the backup/media server and the disk, and data is stored in a deduplicated form. *Post-process deduplication* is performed after data lands to disk, and a full copy of the most recent backup is stored in a landing zone.

Inline deduplication is used by nearly all primary storage vendors and backup software-based deduplication implementations. With this approach, incoming backup data is deduplicated on the fly, before it lands on disk. This approach is faster than deduplication on the media server because it offloads processing from the media server and lets it stay dedicated to its task. However, inline deduplication can potentially cause a bottleneck at the point where data is streaming into the deduplication appliance, resulting in slower performance and a longer backup window. The greater impact is seen by users in recovery because this deduplication method stores all data in a deduplicated form and no full copy of backup data is kept. This slows down restores because it requires that a given backup be "rehydrated" or put back together from its deduplicated state before it can be restored. This approach also does not support instant recovery techniques available from an increasing number of backup applications today. Instant recovery allows for a granular level restore of a single file, single VM or object down to an individual email. To rehydrate the data, it can take an hour just to restore a single file.

Post-process deduplication with a landing zone allows backups to land on disk prior to deduplication. Deduplication is performed after the backup is complete, so backups are faster. This approach keeps the most recent backup on disk in a landing zone, in its complete and un-deduplicated full form. The approach of keeping a full copy of the most recent backup in a landing zone in its full form–versus only deduplicated data that always needs to be rehydrated–produces significantly faster traditional full backup and image restores, much faster tape copies, and instant recoveries of files, objects, and VMs that take minutes versus hours.

Where and How Deduplication is Implemented: Key Implications

Deduplication solutions fall into four categories based on where deduplication is performed in the backup process as well as the underlying system architecture. Each of those four categories is detailed below, with the associated pros and cons.

Backup client or agent: only moves changed data over network

Pros	Cons
• Reduces traffic over the network and reduces backup window	• Very compute-intensive • IT has to manage running the process on production servers, with associated higher operational risk

Media server: software running on media server or pre-loaded on a storage server; 64KB or 128KB fixed block sizes

Pros	Cons
• One console and one vendor to work with • Support their own data types well	• Slows down backup process due to shared resources • Uses more disk and bandwidth over time; deduplication ratio and data reduction are lower since deduplication methods must use media server CPU and memory very efficiently • Cannot be used with multiple backup applications, utilities, and database dumps

Target-side appliance that deduplicates between backup/media server and disk: block-level deduplication with inline processing and scale-up architecture

Pros	Cons
• Faster than media server approach because it keeps media server dedicated to its task • Higher deduplication ratio via dedicated target-side appliance	• Slower traditional restores, slower offsite tape copy, and recoveries in hours versus minutes because deduplicated data must be rehydrated when needed • Backup window re-expands as data grows because scale-up architecture only adds disk, not processor and memory to keep pace with workload • Costly forklift upgrades

Target-side appliance that deduplicates after data lands to disk: zone-level deduplication with byte-level deduplication and scale-out architecture

Pros	Cons
• Faster than media server approach because keeps media server dedicated to its task • Higher deduplication ratio via dedicated target-side appliance • Backup window stays fixed in length due to scale-out approach of adding compute with disk • No forklift upgrades • Fastest backups; backups write directly to disk with no deduplication in the way • Fastest restores and offsite tape copy, instant recoveries in minutes versus hours from full copy of backup kept on disk	• Uses additional disk for landing zone; however, no disadvantage to user since cost is equal or less than other solutions and rack space and power/cooling are equal • Potential in certain scenarios that recovery point objective (RPO) at DR site could be slightly less up-to-date

Chapter 4

Sizing a Disk Backup System

Just as many factors must be considered in evaluating the architectural implications of different disk backup with deduplication products, many aspects of your environment are a part of the equation to ensure that you are sizing the system correctly.

In primary storage you can simply say, “I have 8TB to store and so I will buy 10TB.” In disk-based backup with deduplication, a sizing exercise must be conducted based on a number of factors so that you avoid the risk of buying an undersized system which quickly exceeds capacity.

Data Types

As discussed in Chapter 3, the data types you have directly impact the deduplication ratio and therefore the system size you need. If your mix of data types is conducive to deduplication and has high deduplication ratios (e.g. 50:1), then the deduplicated data will occupy less storage space and you need a smaller system. If you have a mix of data that does not deduplicate well (i.e. 10:1 or less data reduction), then you will need a much larger system.

What matters is what deduplication ratio is achieved in a real-world environment with a real mix of data types.

Deduplication Method

As discussed in Chapter 3, the deduplication method has a significant impact on deduplication ratio. All deduplication approaches are not created equal.

- **Zone-level** with byte comparison or alternatively 8KB block-level with variable length content splitting will get the best deduplication ratios. The average is a 20:1 deduplication ratio with a general mix of data types.
- **64KB and 128KB fixed block** will produce the lowest deduplication ratio, as the blocks are too big to find many repetitive matches. The average is a 7:1 deduplication ratio.
- **4KB fixed block** will get close to the above but often suffers a performance hit. A 13:1 deduplication ratio is the average with a general mix of data types.

Retention

The number of weeks of retention you keep impacts deduplication ratio as well. The reason is that the longer the retention, the more the deduplication system is seeing repetitive data. Therefore, the deduplication ratio increases as the retention increases. Most vendors will say that they get a deduplication ratio of 20:1, but when you do the math, that is typically if the retention period is about 16 weeks. If you keep only two weeks of retention, you may only get about a 4:1 reduction.

Example:

If you have 10TB of data and you keep four weeks of retention, then without deduplication you would store about 40TB of data. With deduplication, assuming a 2% weekly change rate, you would store about 5.6TB of data, so the deduplication ratio is about 7.1:1 (40TB ÷ 5.6TB = 7.1:1).

However, if you have 10TB of data, and you keep 16 weeks of retention, then without deduplication you would store about 160TB of data (10TB x 16 weeks). With deduplication, assuming a 2% weekly change rate, you would store about 8TB of data, which is a deduplication ratio of 20:1 (160TB ÷ 8TB = 20:1).

Rotation

Your backup rotation will also impact the size of the disk-based backup with deduplication system you need. If you are doing rolling full backups each night, then you need a larger system than if you are doing incremental backups on files during the week and then a weekend full backup.

Rotation schemes are usually:

Database and Email

- Full backup on Monday, Tuesday, Wednesday, Thursday, weekend

File Data

- Incrementals forever or optimized synthetics
  - Copies only changed files each night, no weekend full
- Incrementals
  - Copies changed files each night, full backup of all files on the weekend
- Differentials
  - Copies files each night that have changed since the last full backup, full backup of all files on the weekend
- Rolling Fulls
  - Breaks total full backup into a subset and backs up a portion of the full backup each night (e.g. if the full backup is 30TB, then back up 10TB each night and keep rotating on a three-day schedule)

Because the backup rotation scheme you use changes how much data is being sent to the disk-based backup with deduplication system, this also impacts the system size you require.

Cross Protection

Sizing Scenario 1: You are backing up data at site A and replicating to site B for disaster recovery. For example, if site A is 10TB and site B is just for DR, then a system that can handle 10TB at site A and 10TB at site B is required.

Sizing Scenario 2: However, if backup data is kept at both site A (e.g. 10TB) and at site B (e.g. 6TB) and the data from site A is being replicated to site B while the data from site B is being cross-replicated to site A, then a larger system on both sides is required.

Bottom Line for Sizing a System

In summary, dozens of possible scenarios impact the sizing of a system, including:

- How much data is in your full backup?
  - What percentage of the data is compressed (including media files), encrypted, database, unstructured?
- What is the required retention period in weeks/months onsite?
- What is the required retention period in weeks/months offsite?
- What is the nightly backup rotation?
- Is data being replicated one way only or backed up from multiple sites and cross-replicated?
- Other considerations unique to your environment

When working with a vendor, ensure they have a sizing calculator and that they calculate the exact size of the system you need based on all of the above.

The mistake often made is that the system is acquired and in a few short months, it is full because the system was undersized, retention was longer, the rotation scheme put more data into the system, the deduplication method had a low deduplication ratio, or the data types were such that they could not deduplicate well.

The truly knowledgeable vendors understand that disk-based backup with deduplication is not simply primary storage; therefore, they have the proper tools to help you size the system correctly.

Chapter 5

Considerations and Questions for Vendors

A number of questions should be asked before you acquire a disk-based backup system to avoid the following possible undesired outcomes:

- Certain features in your backup application may not be supported.
- You may not have disk backup support for other utilities in your environment.
- Disk usage may rise quickly with each backup due to a poor deduplication ratio, and within 6 to 18 months you could need additional capacity. The poor deduplication ratio can be a result of the actual algorithm, certain data types that do not deduplicate, or both.
- Purchase of additional WAN bandwidth may be required between the primary site and DR site because of a poor deduplication ratio. The poor deduplication may be due to the actual algorithm, certain data types that do not deduplicate, or both.
- You may be faced with a forklift upgrade where the front-end controller needs to be upgraded if the system does not add processing resources with disk as data grows.
- Your restores could be slow.
- Your tape copies could be slow and could take several days to a week to complete.
- Time to recover from a disaster could be excessive.

- Vendor technical support delays in troubleshooting or resolving problems may result in downtime.

Here are some of the questions you need to ask vendors about their disk backup solution and support:

Backup Application Support

1. Does the solution support your backup application?
2. If you are thinking of changing backup applications in the future, are those other backup applications also supported?
3. For databases, does the solution support direct to disk for Oracle® RMAN dumps or SQL dumps?
4. For environments with some UNIX, does the solution support direct UNIX TAR files?
5. If it is a Symantec™ Backup Exec or NetBackup environment, which levels of OST are supported?
6. If it is a Symantec NetBackup environment, are advanced features, including AIR and optimized synthetics, supported?
7. If it is an IBM® Tivoli Storage Manager environment, are progressive incrementals supported? How is reclamation performed?
8. If the environment includes a backup application that performs deduplication (e.g. Veeam® Backup and Replication), does the disk-based backup product work with that deduplication and further deduplicate the data?
9. Can you bring in data from multiple backup applications and utilities? If so, which ones?
10. If the vendor also sells their own backup application, does their disk backup solution offer robust, complete capabilities to support other company's backup applications?

Deduplication Approach/Architecture, Impact on Backup and Recovery Operations

1. What method of deduplication does the solution use (as disk usage and bandwidth needs will vary greatly)?
    - Zone with byte-level
    - 8KB block with variable length splitting

- 4KB fixed block
- 64KB or 128KB fixed block

2. How long does it take for a traditional full backup or image backup to restore?
3. Does the solution support instant recovery techniques, such as Instant VM Recovery?
4. How long does it take to make a full backup offsite tape copy?
5. As data grows, does the system scale by adding processor, memory and bandwidth resources along with more disk capacity? If not, then as data grows, how quickly will the backup window re-expand?
6. Is it possible to speak with several customer references to verify the restore, recovery, tape copy, and scalability capabilities above?
7. How is RPO (recovery point objective) achieved at the disaster recovery site?
8. How is RTO (recovery time objective) achieved at the disaster recovery site?

System Sizing and Pricing

1. When sizing, did the vendor ask you the following for sizing the system?
   - Weekend full backup amount
   - Weekly backup rotation (incrementals, rolling fulls, differentials, etc.)
   - Annual data growth rate
2. How much of your data is comprised of each of the following?
   - Compressed
   - Encrypted
   - Unstructured
   - Database
3. What retention in weeks, months and years is required onsite and offsite?
4. Are any large data amounts expected to be added in the short term due to factors such as bringing on new customers, merging businesses, or other changes?

5. If more than one location is to be backed up, how much data is to be backed up at each location so additional system sizing can be taken into consideration for multi-site cross protection for disaster recovery?
6. Are you turning on deduplication with the backup application such that the overall deduplication rate will be impacted? If so, what does that mean to disk and bandwidth usage?
7. At your current data growth rate, when will you have to buy additional capacity?
8. If the solution is a controller/disk-shelf architecture (scale-up), at your current data growth rate, when will you need to upgrade the front-end controller via forklift upgrade? What will that upgrade cost?
9. Is the vendor giving you a low price up front to gain your business, knowing they will keep selling you more and raise the price later? Or is the vendor willing to put in writing that the price you pay up front is the same that you will pay moving forward?

Management/Reporting and Technical Support

1. What level of user- and management-level reporting is provided?
2. Can the system be easily installed with phone support or does it require an installation team and associated costs?
3. Is support handled by senior-level second-line staff to quickly resolve problems, or is support handled by junior-level first-line staff unable to quickly resolve problems?
4. Is technical support staff skilled in specific backup applications and assigned based on expertise with your backup application and environment?
5. Is technical support based in the country where the system is deployed and managed to avoid poor quality connections and language barriers?
6. Does the technical support and disk backup system include automated health-check emails from the system to the vendor to proactively alert about capacity issues, hardware diagnostics, or other problems?

The above list is not meant to be all inclusive, but rather to demonstrate the complexity of disk-based backup and why it is important to not view it as merely a NAS target, disk, and deduplication. A disk backup system is a “system”—with software required to make the solution work properly in all existing backup environments, allow for fast backups and restores, allow for cost-effective scalability, operate reliably in a real-world production environment, and be properly supported.

The following is highly recommended when acquiring a disk-based backup product:

- Ask all of the important questions.
- Receive an in-depth product presentation.
- Ask for actual customer user stories with the companies and IT staff named.
- Ask to speak with three to five references.
- Do a side-by-side test if you have the staff, lab and time.
    - Use a large amount of data to stress the systems as all products look good with a small backup job.
- Ask about the support model and ask the provided customer references how satisfied they are with the vendor's support.
- Get a future price guarantee in writing to ensure you are not getting a low-ball price up front only to have the price increased later.

Conclusion

In summary, disk backup with deduplication is not simply primary storage or a media server with a deduplication feature. Based on the choices of deduplication methods and the scaling architecture, a great deal in your backup environment can be impacted, with consequences such as:

- Slow backups
- Expanding backup windows
- Undersized systems that hit capacity limits within months
- Costly forklift upgrades for systems that don't scale as data grows
- Slow restores of traditional full or image backups
- Slow offsite tape copy
- Slow instant recoveries of files, objects, and VMs—in hours versus minutes

The choice of architecture for disk backup with deduplication can help you improve your backup environment, or it can lead you to simply replace the old tape-based challenges with new, more expensive disk-based challenges.

With each vendor, take the time to ask many questions and do a thorough comparison, as all disk backup systems are truly not created equal.

About the Author

Bill Andrews is President and CEO of ExaGrid Systems, and has over 30 years of experience in IT data center technology.

ExaGrid offers disk-based backup with deduplication appliances.

For more information, contact ExaGrid Systems at:

800.868.6985
www.exagrid.com
info@exagrid.com

www.exagrid.com