The Guide to **Choosing Backup Storage**

Considerations, key requirements, and
options for modern day IT shops

EXAGRID®

Tiered Backup Storage

Contents

Introduction

Most IT organizations put significant thought and research into choosing a backup application. However, most also assume you can simply put any kind of storage behind the backup application, and it will just work.

Choosing the wrong backup storage can impact:

- Backup performance and resulting backup window time
- Restore performance and the impact on users
- Whether the backup window stays fixed length as the data grows
- How much time is spent on managing the backups
- How open the storage is to a security or ransomware attack
- The overall cost over time.

In primary storage, you simply buy the usable terabytes or petabytes that you need, including some growth, however unlike primary storage, backup storage is unique with:

- Large backup jobs
- Backup rotation: incrementals, differentials, fulls, synthetic fulls, reconstituted fulls, etc.
- Data deduplication which impacts performance, storage, and system sizing, etc.

Backup storage is much more complex. While backup applications are usually top of mind, backup storage can often be overlooked and leads to poor performance, and overpaying for storage in the backup environment. In addition, the importance of utilizing backup storage as a key part of the recovery process especially after a ransomware attack. Being prepared for a recovery from a site disaster is also a big part of the equation.

- Is the storage file system maximized for:
    - standard files?
    - database transactions?
    - OR truly optimized for large backup jobs?
- Is the storage integrated with advanced protocols specifically designed for large backup jobs such as Veritas NetBackup OST, Veeam Data Mover, etc.?
- Can your backup storage perform job concurrency (jobs running in parallel)?
- Is the backup storage "scale-up", so the backup window grows as data grows or is it "scale-out", which keeps a fixed length backup window as data grows?
- How can you ensure the storage is sized correctly for:
    - backup rotation
    - backup retention: dailies, weeklies, monthlies, yearlies
    - cross replication for multiple sites
    - second site disaster recovery
    - data growth for the next 3 to 5 years
    - whether the backup application or backup storage has data deduplication
        - What is the impact on performance (as deduplication is compute intensive)?
        - What is the deduplication ratio and resulting storage savings as each vendor is different?
        - How much of the data is encrypted or compressed as target side deduplication will not bring any further deduplication?

Core Requirements

What are the core requirements for backup storage?	
✔	Fast for backups for the shortest backup window
✔	Fast for restores to keep users productive
✔	Fixed-length backup window as the data grows
✔	Secure against attacks and ready to recover from an attack such as ransomware
✔	Sized correctly so that if you plan on 3 years of growth, you are not re-buying in 9 months
✔	Low cost up front
✔	Low cost over time – no forklift upgrades, no planned product obsolescence
✔	Easy to use & manage, with low touch time
✔	Exceptional support that is knowledgeable, available and knows both backup applications and backup storage

Primary Storage - HDD

Primary Storage is designed for user files and not backup. It is often slow for large backup jobs, as it is storage only and not an appliance/server. Primary storage HDD cannot run backup application advanced transport protocols and becomes very expensive very quickly because it does not have any additional data deduplication above and beyond the backup application. Backup application deduplication is typically 1:1, 2:1, 3:1 up to 5:1. That will save some storage, but will not save the maximum amount of storage, resulting in high storage costs for longer term retention.

Considerations:

- Not maximized for backup performance (files system, advanced backup protocols, job concurrency, etc.). This results in longer backup windows, that grow over time.
- Restores will be as fast as disk as long as the data is not deduplicated by the backup application. If deduplicated by the backup application, then the deduplicated blocks must be put back together or rehydrated which slows restores down.
- The storage is typically not scale-out, so the backup window grows as data grows, so you are chasing an ever-expanding backup window.
- Becomes very expensive with longer term retention of over 6 copies. The storage is often under-sized as the vendors don't know how to size for backup including backup rotation, data growth, retention policy, cross replication, etc. Always compare apples to apples for sizing as storage is often under-sized to look less expensive. (You could be buying more storage in 6 to 9 months.)
- The storage is network facing and highly vulnerable to security attacks.
- The storage is often discontinued, forcing you to buy something else. (Product obsolescence)
- Cost over time is much larger than anticipated, as upfront sizing is typically inadequate.
- Managing silos of primary storage takes IT staff time.
- The support organizations do not understand backup storage, resulting in finger pointing between the backup vendor and the storage provider.

Low Cost Disk Servers - HDD

Low-cost storage servers are designed for user files and not backup. It is often slow for large backup jobs, as it is storage only and not an appliance/server. Low-cost disk servers cannot run backup application advanced transport protocols and becomes very expensive very quickly because it does not have any additional data deduplication above and beyond the backup application. Backup application deduplication is typically 1:1, 2:1, 3:1 up to 5:1. That will save some storage but will not save the maximum amount of storage. In addition, low-cost storage will have very large HDD drives with a single front-end controller in order to drive down the cost and resulting price. Because the drives are larger, there is less spindle count which results in a large performance impact. If a server had 10 x 4TB drives, there are 10 spindles to simultaneously write to. If the server has 4 x 10TB drives, there are only 4 spindles to write to. In addition, the front end-controller is a bottleneck. If you look at the rated ingest rate and divide by the number of TBs to derive the performance per TB, it is slow for backup.

Considerations:

- Not maximized for backup performance (files system, advanced backup protocols, job concurrency, etc.). This results in longer backup windows.
- Low spindle count with bottleneck single controllers.
- Restores will be slower than primary storage disk if the data is not deduplicated by the backup application due to spindle and controller bottlenecks. If deduplicated by the backup application, then the deduplicated blocks must be put back together or rehydrated which slows restores down.
- The storage is not scale-out, so backup window grows as data grows so you are chasing an ever-expanding backup window.
- Because it is low cost, it does help the economics but still becomes more expensive at a certain retention point. Storage is often under-sized as the vendors don't know how to size for backup including backup rotation, data growth, retention policy, cross replication, etc. Always compare apples to apples for sizing as often under-sized systems look less expensive.(You could be buying more storage in 6 to 9 months.)
- The storage is network-facing and highly vulnerable to security attacks.
- The storage is often discontinued, forcing you to buy something else. (Product obsolescence)
- Cost over time is much larger than anticipated as upfront sizing is typically inadequate.
- Managing silos of primary storage takes IT staff time.
- The less you pay, the lower the level of support that is provided. The support organizations do not understand backup storage, resulting in finger pointing between the backup vendor and the storage provider.

Solid State Storage - SSD

Solid State Storage – SSD is designed for database transactions. It is often slower for large backup jobs, as it is storage only and not an appliance/server. Solid state storage cannot run backup application advanced transport protocols and becomes very expensive very quickly because it does not have any additional data deduplication above and beyond the backup application. Backup application deduplication is typically 1:1, 2:1, 3:1 up to 5:1. That will save some storage but will not save the maximum amount of storage. Lastly, SSD is typically 5X the cost per TB versus standard HDD disk.

Considerations:

- Not maximized for backup performance (files system, advanced backup protocols, job concurrency, etc.) This results in longer backup window. This sounds counter intuitive for SSD so test before you buy, and you will see that file systems need to be optimized for large backup files.

- Restores will be faster than HDD disk as long as the data is not deduplicated by the backup application. If deduplicated by the backup application, then the deduplicated blocks must be put back together or rehydrated which slows restores down but will still be faster than HDD disk.

- These systems are typically scale-out, however compute resources must grow at the same rate the data grows, or the backup window grows as data grows so you are chasing an ever-expanding backup window.

- Extremely expensive with longer term retention as SSD is typically 5X the price of HDD. SSD is extremely expensive for backup storage. Often under-sized as the vendors don't know how to size for backup including backup rotation, data growth, retention policy, cross replication, etc. Always compare apples to apples for sizing as often under-sized systems look less expensive. (You could be buying more storage in 6 to 9 months.) SSD may look closer to the price of HDD but you might not be buying the required capacity.

- The storage is network facing and highly vulnerable to security attacks. Need to buy additional storage and tools which further drives up cost.

- Cost over time is much larger than anticipated as upfront sizing is typically inadequate.

- The support organizations do not understand backup resulting in finger pointing between the backup vendor and the storage provider.

Microsoft ReFS and Linux XFS

For an organization utilizing ReFS and XFS, they use block cloning which can double the deduplication ratio from 2:1 to 4:1. As a result, it will cut the storage required by 50%. To lower the cost of backup storage, IT departments will couple low-cost disk servers or SSD storage with either ReFS or Linux XFS. This works reasonably well for small businesses that do not have a lot of data, a lot of data growth, and do not have a need for performance. However, for larger organizations, 50TB to petabytes, the impact on backup is staggering.

Considerations:

- Has all the challenges of low-cost disk servers or SSD described early in this document.
- In addition, ReFS and XFS have a further performance impact, so the performance is less than the native disk file system.
- Does not scale well as ReFS was designed for smaller IT shops.
- Take a lot of management time, especially for Linux XFS.
- Good for small IT shops with 30TB or less, may not be good for IT shops with 30TB to 50TB, and is not a viable solution for the upper mid-market, small enterprise, enterprise, and large enterprise.

Inline Deduplication Appliances

Inline deduplication appliances have advanced data deduplication and are typically 4 to 10 times more efficient than deduplication in the backup applications, which greatly reduces the backup storage and resulting storage costs for longer term retention. However, these appliances do the data deduplication on the way to the disk, which significantly slows down backups to about 1/6th the speed of disk. They deploy software that runs on the media server/agent to do some of the work on the backup servers but that still only achieves performance about 1/3rd the speed of disk and now their code on the media server/agents takes a performance hit because compute is being used for pre-deduplication work. Since the deduplication is performed inline, all of the data is stored as deduplicated blocks and they have to be put back together (rehydrated) for every request. Restores are about 1/20th the performance of restores for a non-deduplicated file on standard disk. In addition, these solutions are scale-up, so as data grows there are no additional memory, compute, and networking resources added resulting in an ever-growing backup window. The net is that this first-generation approach saves storage but at the cost of backup performance, restore performance, and scalability. You have to lose to win.

Considerations:

- Slow performance as inline deduplication is compute intensive. Even with deduplication software that runs on the media server the performance is still inadequate. In addition, they don't all do job concurrency for parallel backup jobs. Inline deduplication appliances save storage costs but are the slowest approach for ingest performance.

- Restores are painfully slow, as the data has to be put back together or rehydrated for each request. VM boots can take an hour versus a few minutes.

- These solutions are not scale-out so backup window grows as data grows so you are chasing an ever-expanding backup window. This results in expensive and disruptive forklift upgrades.

- The backup storage is network facing and highly vulnerable to security attacks. You have to add additional storage and software to increase security which virtually doubles the cost.

- The storage is often discontinued, forcing you to buy something else. (product obsolescence)

- Cost over time is much larger than anticipated as upfront sizing is typically inadequate. Often vendors undersize to lower their entry price.

- The support organizations do not understand backup storage, resulting in finger pointing between the backup vendor and the storage provider.

End-to-end Backup Application to Backup Storage Solutions

Over the decades, IT technology swings from distributed systems to all-in-one solutions and then back to distributed systems. With end-to-end backup solutions, you are betting that the vendor has "best of breed" for the backup application and for the backup storage. This is never the case in backup. If you decide to change the backup storage you can't and now you are locked into an inadequate and very expensive backup solution. Many end-to-end solutions hide their true costs by recommending that you keep 2 to 4 weeks on site and then keep all longer-term retention in the cloud. IT organizations see the low cost for the onsite system but have no idea of the cloud costs for the next 36 to 60 months. Then, when they get the cloud bill in about 9 months, they see the true costs of the storage and now they are LOCKED IN.

Considerations:

- Backup performance is slow because they do inline data deduplication, compression, and encryption all in the software which slows backups down.

- Restores are slow because they have to rehydrate the deduplicated data blocks for each request.

- They are scale-out, however; it is very expensive because they use erasure coding to achieve redundancy which adds a lot more cost.

- Becomes very expensive with longer term retention because the deduplication ratios are between 3:1 to 5:1 so they use a lot of disk storage. They are often under-sized as the vendors don't know how to size for backup including backup rotation, data growth, retention policy, cross replication, etc. In addition, there is no easy way to determine the monthly cloud costs for retention storage. Always compare apples to apples for sizing as often under-sized systems look less expensive. However, you will most likely be buying more storage in 6 to 9 months.

- They tout immutable data objects for ransomware recovery however that assumes there is no way to get to the immutable data objects. They do not have a non-network-facing tier (tiered air gap).

- Often, they have you buy the hardware from a separate vendor.

- The support organizations do not understand backup storage, resulting in finger pointing between the backup vendor and the storage provider.

ExaGrid Tiered Backup Storage

ExaGrid Tiered Backup Storage is designed for backup. It takes into consideration backup performance, restore performance, fixed-length backup window as data grows, non-network-facing tier for security, upfront cost, cost over time, and support that understands both the backup application and the backup storage.

Considerations:

- Unique front-end Landing Zone. Backups write direct to disk with no inline deduplication to slow backups down. Performance is fast due to:
  - No inline deduplication
  - A file system optimized for large backup jobs
  - Use of advanced protocols for performance such as the Veeam Data Mover or Veritas NetBackup OST, etc.
  - Use of job concurrency for parallel backup jobs
  - Use of integrated backup application functionality for front-end job load balancing (Veeam SOBR, NetBackup Single Storage Disk Pool, Commvault Spill & Fill, Oracle RMAN Channels, HYCU Scale-out, etc.)
  - Data at rest encryption at the drive level (takes nanoseconds)
  - Typically, 2 times faster than SSD and 3 to 5 times faster than inline deduplication appliances.

- Restores will be as fast as disk, as the data is in the Landing Zone in the native backup application format ready to restore.

- Scale-out where compute is added, with capacity, the system scales linearly keeping the backup window fixed-length as data grows.

- Data is deduplicated from the Landing Zone into a non-network-facing repository tier where a copy of the most recent data in the Landing Zone and all retention data is stored in a highly deduplicated format to greatly lower the storage requirement and resulting cost. ExaGrid allows Veeam and Commvault to keep their native data deduplication turned "on" and ExaGrid will further deduplicate that data.

- All the data is stored in a deduplicated format in the non-network-facing repository tier that threat actors cannot see or access. The system employs a delayed delete policy to ensure data deleted in the Landing Zone is not immediately deleted in the repository. In addition, the system can survive encrypted data ingest because all the previous repository data objects are immutable, which means they are not changed, deleted, or modified, protecting the most recent backup and all retention points.

- No planned product obsolescence as ExaGrid allows any age or size appliance to be mixed or matched in the same system. So even if a model is no longer available it can still be used with current models.

- Sizing for backup is extremely complicated because unlike primary storage there are a dozen impact items such as rotation, retention, data types, etc. ExaGrid has extensive sizing tools and prides itself on and is committed to not undersizing.

- ExaGrid's system is designed for ease of use. Customers constantly say they spend the least time on ExaGrid.

- ExaGrid assigns a level 2 support engineer to each customer account so that the IT person works with the same senior support engineer all the time. You don't get the run around and you don't have to repeat yourself over and over. ExaGrid's Level 2 support engineers are experts in backup applications, networking, and ExaGrid Tiered Backup Storage.

Summary

Backup storage is very complicated, and it is easy to make a backup buying decision mistake.

ExaGrid highly recommends that you:

1. Take the time to truly listen to all the ins and outs of all the solutions.

2. That you talk to at least 3 to 5 reference accounts who are using the solution, ensure that it is just you and the customer, and the vendor is not on the call. Remember, everyone has customer success stories, talk live to customers.

3. Test, test, test. There is a lot of marketing hype out there and a lot of salespeople telling interesting stories. Nothing beats testing side by side and seeing ingest performance, restore performance, and scalability for yourself. Also, do a mock security attack and see which system is still standing.

4. Always, double, and triple check the sizing. There are thousands of stories where customers bought 3 years of data growth up front and in 9 months, they hit the wall and have to buy additional capacity and now the long-term cost was much higher than expected or imagined.

	Primary Storage HDD Disk	Low Cost Disk with ReFS or XFS	Primary Storage SSD Disk	Deduplication Appliances	End-to-End Backup	EXAGRID Tiered Backup Storage
Fastest performance for shortest backup window	●	●	●	✕	●	✔
Fastest restores for user productivity uptime	✔	●	✔	✕	●	✔
Fixed length backup window as data grows – scale-out	✕	✕	✔	✕	✔	✔
Low cost up front for longer term retention	✕	✔	✕	✕	✕	✔
Low cost over time – no forklift upgrade or obsolescence	✕	●	●	✕	✕	✔
Comprehensive security	✕	✕	●	✕	✔	✔
Strong recovery from ransomware Non-network-facing tier/ tiered air gap/delayed deletes	✕	✕	✕	✕	●	✔

● = adequate

EXAGRID®