Hidden Secrets Regarding Backup Storage

Introduction

A lot of time goes into picking a backup application such as Veeam, Veritas NetBackup, Commvault, HYCU, Oracle RMAN Direct, and dozens of others. However, rarely does anyone think about or truly understand the impact of the backup storage that sits behind these backup application in regards to cost, performance, scalability, and IT staff time.

Low-cost Primary Storage Disk

Many organizations use low-cost primary storage disk for their backup storage. It is simple and fast. The challenge with low-cost primary storage disk is the long-term cost depending on the retention policy. Some backup applications have no data deduplication to save disk storage and others have a low level of deduplication that will achieve 2:1, 3:1 and even a 5:1 data reduction. This definitely helps in reducing storage and saving costs. However, you cannot run aggressive 20:1 data reduced data deduplication in a backup application as it is too compute intensive. Data deduplication in the backup application slows it down so you can only afford to do a low level of deduplication in the backup application without crippling backup performance. In addition, many backup applications also do their encryption in software. The combination of data deduplication and encryption can greatly slow backups down.

Disk behind a backup application is cost effective when the retention is low, typically 2 copies to as many as 12 copies (this is a combination of weekly, monthly and yearly backups). The cost cutover point depends on the level of deduplication that the backup application achieves.

Backup application deduplication approach and cost cutover point:

- No deduplication   1:1   2 copies
- 1MB blocks   2:1   4 copies
- 128KB blocks   5:1   8-12 copies with incrementals forever

Low-cost primary storage disk is the best approach for 2, 4 and sometimes even 6 copies. Typically, at over 4 copies, dedicated appliances with data deduplication can be far less expensive than low-cost primary storage disk. This is because resources are dedicated to the deduplication process and more aggressive algorithms can be used achieving deduplication ratios of 10:1 to as high as 50:1 with an average typically around 20:1.

Often what happens is that the IT organization does not know how to calculate how much low-cost primary storage will be required, so they buy a certain amount of usable terabytes, and then 3 months later buy more, and another 3 months later buy more, etc. They start with $50,000 worth of backup storage and 3 years later end up with $500,000 worth of backup storage. You need to calculate up front how much storage will be required over the next 3-5 years so you know up front and can plan or instead chose a different approach. In hindsight, many organizations greatly overspend on storage.

Companies, such as ExaGrid, can use a straight disk calculator that shows the difference in cost between low-cost primary storage disk and appliances with deduplication, while taking into effect the retention policy (number of daily, weekly, monthly and yearly backups), the data growth rate for a specified number of years, whether replication to a DR site is required, and the cost per usable terabyte of the low-cost primary storage. This approach quickly shows if you are overspending on your backup storage.

Inline, Scale-up Deduplication Appliances

Inline, scale-up deduplication appliances such as Dell Data Domain, HPE StoreOnce and Veritas NetBackup Appliances, are all dedicated to deduplication and achieve higher deduplication ratios and therefore can greatly reduce the cost of the backup storage. However, there is a huge penalty to pay to save on storage costs.

Since the deduplication is performed inline, which means on the way to the disk between the backup application and the storage, the performance is about 15% the performance of writing directly to disk. As a result, other techniques are employed such as Data Domain Boost or HPE Catalyst in order to perform some of the deduplication work on the backup application media server. This approach improves the performance to about 1/3rd the performance of disk. So you are saving storage costs, but taking a huge hit on backup performance which increases your backup window. If you turn on software encryption for encrypted data at rest, then the backups slow down even further.

In addition, all of the data on the disk is deduplicated so that when you want to restore, the data must be rehydrated, and the restores can be up to 20x slower than restoring un-deduplicated data from disk.

Lastly, these appliances use a scale-up architecture, which means they have a fixed controller and disk shelves. As the data grows, they don't add any additional resources so the backup window grows and grows until the backup window is so long that you need to buy a larger front-end controller or additional controllers that are each a silo system increasing backup management. Scale-up storage has always forced expensive and disruptive upgrades, siloed systems and forced product obsolescence as each time new processors, disk, etc. come out, the systems need to be refreshed.

ExaGrid Tiered Backup Storage

ExaGrid looked at these problems and came to the conclusion that IT organizations want it all:

- Low-cost storage due to aggressive data deduplication
- The high performance of disk without inline deduplication
- The high performance restores or restoring from un-deduplicated data
- No impact to performance for encryption at rest
- Scale-out architecture so the backup window does not grow as data grows
- Comprehensive security and ransomware recovery
- No forklift upgrades
- No planned product obsolescence
- And knowing what their costs will be in 3, 4 and 5 years from now

Landing Zone
High Speed Backups & Restores

Repository Tier
Current Backup & Retention Data –
Deduplicated for Low Cost

Landing Zone
Repository Tier

Pros
- Fast for backups
- Fast for restores
- Adaptive deduplication – storage savings
- Ransomware recovery solution
- Scale-out
   - Fixed-length backup window
   - No forklift upgrade
   - No planned product obsolescence

ExaGrid's Tiered Backup Storage delivers on all of these goals:

- ExaGrid offers the same level of data deduplication as the Dell, HPE, Veritas appliances, and in some cases offers even more.
- ExaGrid has a unique front-end disk-cache Landing Zone where the backups are written to without any inline deduplication. This makes the backup as fast as any disk.
- ExaGrid performs encryption at rest at the disk level which occurs in nanoseconds avoiding software-based encryption that slows backups down.
- The most recent backups remain in the Landing Zone for fast restores, and as the data is not deduplicated and there is no data rehydration process or penalty.
- ExaGrid's second tier is the long-term retention repository. Data in the Landing Zone is deduplicated into the repository tier in order to keep all long-term retention data in a deduplicated form to reduce storage costs. Since the deduplication is not inline, there is no performance impact on backups or restores.
- ExaGrid uses a scale-out storage architecture. Each appliance has processor, memory, networking and disk. The combination is set for a 6-hour backup window. As data grows, full appliances with all resources are added so the backup window stays fixed in length as the data grows. If the backup window is 6 hours at 100TB it is 6 hours at 500TB, 1PB, 2PB, etc.
- Due to scalability, there are no forklift upgrades as the backup window does not grow.
- ExaGrid has eliminated planned product obsolescence. All appliance models, new and older, and models of any size can be mixed and matched in the same system. If ExaGrid stops marketing a model, it will not be end-of-life'd, and the maintenance and support stays the same.
- ExaGrid has a standard 5-year price protection program. Whatever you pay for an appliance up front is your price for additional appliances for the next 5 years. Due to the scale-out approach, you can buy what you need when you need and you know the price stays fixed. In addition, maintenance and support will not go up more than 3% per year.
- ExaGrid also includes a unique approach for ransomware recovery. Since the ExaGrid Repository Tier is not network-facing, only the ExaGrid code has access to that tier. ExaGrid delays delete requests (default is 10 days), so that if the backup data is deleted only the Landing Zone data is impacted and all long-term retention data is still intact for restore from. Low-cost primary storage or inline data deduplication appliances are network-facing and would be 100% deleted. In addition, ExaGrid also has immutable data objects to guard against encrypted data being sent to the Landing Zone or data being encrypted directly on the Landing Zone.

Summary

Low-cost primary storage disk is the best choice for shorter retention periods such as a few weeks.

Dedicated deduplication with higher deduplication ratios is required for backup storage cost savings.

Inline, scale-up deduplication appliances reduce backup performance, reduce restore performance, and don't scale. You are forced to trade off performance and scalability for storage cost savings.

ExaGrid provides the best of both worlds by providing the performance of disk for backup and restores, long-term deduplication savings for all retention, scale-out architecture for a fixed-length backup window and no forklift upgrades or planned product obsolescence. In addition, ExaGrid's architecture provides a tiered air gap, with a non-networking-facing Repository Tier with immutable data objects and delayed deletes for ransomware recovery.