Acronis

Deduplication in Acronis Backup Advanced

Technical Whitepaper

Table of Contents

Introduction

Powered by the Acronis AnyData Engine, Acronis Backup Advanced delivers robust, easy-to-use, unified data protection and disaster recovery for multi-system environments. Based on your business needs, you can deploy individual solutions or seamlessly blend them together into one efficient backup solution and manage them using a single unified console, the Acronis Management Server (AMS).

Once blended, Acronis Backup Advanced provides additional centralized storage options through the Acronis storage node. One of the key capabilities of the Acronis storage node is **deduplication**.

Deduplication technology helps reduce storage costs and network bandwidth utilization by eliminating duplicate data blocks when you back up and transfer data.

Acronis Backup Advanced deduplication helps you to:

1. Reduce storage space usage by storing only unique data
2. Eliminate the need to invest in data deduplication-specific hardware
3. Reduce network load because less data is transferred, leaving more bandwidth for your production tasks

This document is appropriate reading for highly technical individuals. It describes Acronis Backup Advanced's deduplication technology and how it executes in Acronis Backup 12.5 Advanced. IT professionals will find this information valuable when designing and implementing a backup infrastructure.

Storage Challenges

We are living in the era of big data.

In 1990, the hard disk of a personal computer was 10 megabytes. Now, multi-terabyte disks are the norm. Every 10 minutes, humanity creates as much data volume as was created from the dawn of civilization until the year 2000.

You must protect and back up all this data — otherwise your company can lose money, reputation, time — even your entire business can shut down.

However, 75 percent of small- to medium-sized businesses (SMBs) surveyed by Acronis and IDC admit that their data is not fully protected. The "sheer volume of data" was given as one of the primary reasons why these SMBs do not protect all data.

For example, let's look at a company with 400 employees who use desktops and laptops. An average laptop can hold from 50 to a few hundred gigabytes of data on the hard disk — so all PCs contain from 20 to 150TB of data. With a 2:1 compression ratio, the backup administrator needs to provision between 10 to 75TB for every full backup, plus have more space for incremental and differential backups. Eventually, this company may need to acquire as much as a **one petabyte of storage** for PC backups alone.

Let's assume this company invests in expensive storage for their PC backups. The next, even bigger challenge is to back up the PCs to this storage. A one-hundred-megabit network can transfer only 10 megabytes of data per second. At this rate, a full backup will take from **two to three weeks** to transfer 10 to 75 TB of data over a 100-mbit network.

Yet, every desktop has the same Windows® operating system, same applications, and often numerous copies of the same data. Storing and transferring the same data multiple times to the same storage is a waste of time and resources. If a backup solution transfers and stores only unique data, the company can decrease their storage capacity and network requirements by up to 50 times! With deduplication, your organization can realize these savings.

How Deduplication Helps

Deduplication minimizes storage space by detecting data repetition and storing the identical data only once. Deduplication reduces network load. During a backup, if data is found to be a duplicate of data previously backed up, it is not transferred over the network to storage.

When you use deduplication, Acronis Backup Advanced deduplicates the backups and saves them to a location managed by Acronis Storage Node component. A backup location where deduplication is enabled is called a **deduplicating vault**. Starting from Acronis Backup v12.5

Advanced **deduplicating vault** is presented in the product interface and documentation as "**managed location**" with enabled deduplication.

Deduplication operates at a block level and works with all operating systems supported by Acronis Backup Advanced.

Deduplication produces maximum results when you create:

- **Full backups** of similar data from different sources, such as operating systems (OSs), virtual machines, and applications deployed from a standard image
- **Full backups** of systems that you previously backed up to the same deduplicating vault
- **Incremental backups** of similar data from different sources; for example, when you deploy OS updates to multiple systems and run an incremental backup
- **Incremental backups** where the data does not change but the location of data does change; for example, when data, such as a file, circulates over the network or within one system and appears in a new place

How It Works

During deduplication, the backup data is split into blocks. Each block's uniqueness is checked through a special database, which tracks all the stored blocks' checksums. Unique blocks are sent to the storage and duplicates are skipped.

For example, if 10 virtual machines are backed up to the deduplicated vault and the same block is found in five of them, only one copy of this block is sent and stored.

This algorithm of skipping duplicate blocks saves storage space and minimizes network traffic.

The following sections describe deduplication technology and its implementation in Acronis Backup 12.5 Advanced.

Deduplication at Source

When performing a backup to a deduplicating vault, the Acronis Backup agent calculates a fingerprint or a checksum of each data block. This fingerprint or checksum is often called *a hash value*.

The data block size varies from **1 byte to 256KB for disk-level and file-level backups**. Each file that is less than 256KB is considered a complete data block. Files larger than 256KB are split into 256KB blocks.

Before sending the data block to the vault, the agent queries the storage node to determine whether the block's hash value is already stored there. If so, the agent sends only the hash value; otherwise, it sends the block itself. The storage node saves the received data blocks in a temporary file. Saving backups to a temporary file first — instead of writing those directly to the deduplication data store — increases scalability and allows parallel processing of backups from multiple agents.

Some data, such as encrypted files or disk blocks of a non-standard size, cannot be deduplicated. In these cases, the agent always transfers this data to the vault without calculating the hash values. For more information about limitations of deduplication, see the **Deduplication Limitations** section below.

Once the backup process is complete, the vault contains the resulting backup and the temporary file with the unique data blocks. The temporary file is located in a special place called the Local Data Store (LDS) and will be processed in the next stage (deduplication at target). The backup (TIB file) contains hash values and the data that cannot be deduplicated. Further processing of this backup is not needed — you can recover data from it.

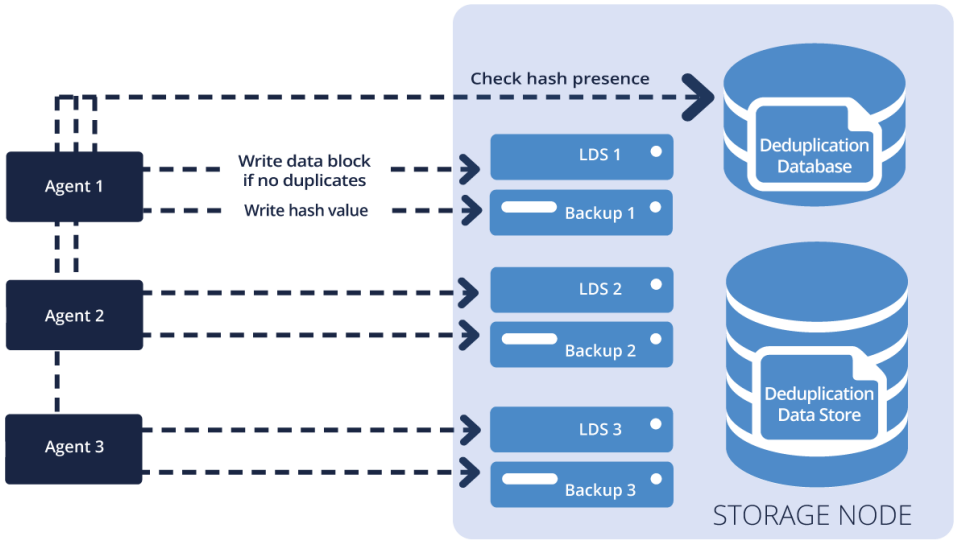

Figure 1:

Deduplication at Target

After a backup to a deduplicating vault is complete, the storage node runs the **indexing** task. This task deduplicates the data in the vault as follows:

1. Data blocks are moved from the temporary file to a special file — **the deduplication data store** — within the vault. Duplicate blocks are stored only once.
2. Hash values and links (or offsets) are saved to the **deduplication database**, so the data can be easily reassembled (rehydrated).
3. After all the data blocks are moved, temporary files (LDS1, LDS2, LDS3 on the diagrams) are deleted.

As a result, the data store contains a number of unique data blocks. Each block has one or more references from the backups that are represented by .tib files — shown as "Backup 1," "Backup 2," and "Backup 3" on the diagrams. The references are recorded in the deduplication database. The .tib files are untouched because they already contain hash values and links to the required data blocks. Before indexing is complete, these links point to the temporary LDS and when indexing finishes, they are redirected to the deduplication database. In addition, the .tib files contain the data that cannot be deduplicated.

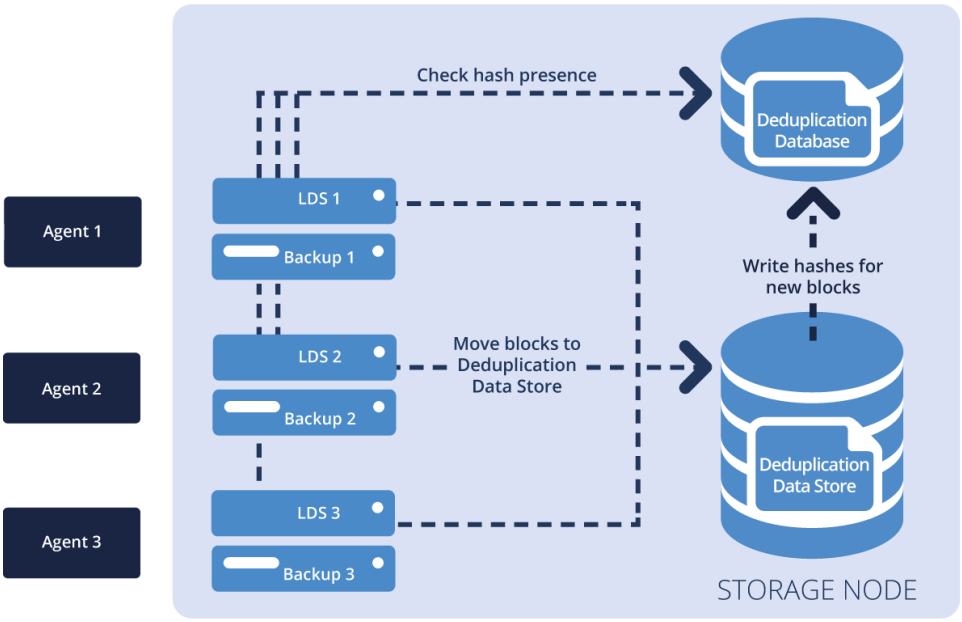

Figure 2

Figure 3 below illustrates the result of deduplication at target.

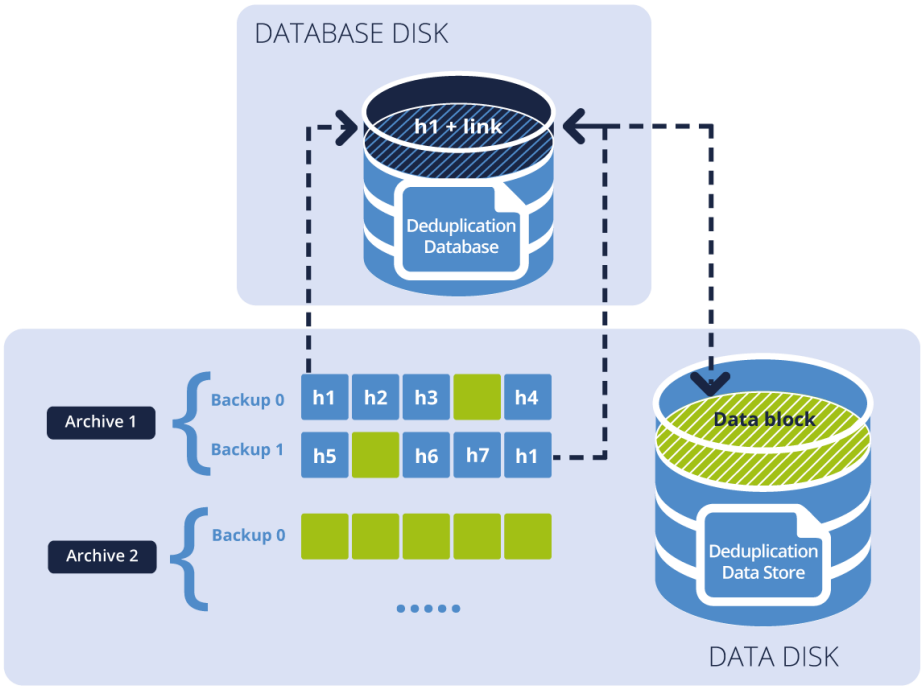

Figure 3

The diagram shows two archives. Each of them has a separate set of backups. In Archive 1, h1 through h7 — designated by blue blocks — contain hash values that are stored in the backup files. The green blocks are the data blocks that cannot be deduplicated. Archive 2 contains only data (green) blocks and is encrypted. As a result, the deduplication database contains hash values of blocks that can be deduplicated, and the deduplication data store contains data blocks from both Archive 1 and Archive 2.

According to best practices, you should store the deduplication database and deduplication data store on separate disks to achieve better performance.

The indexing task may take some time to complete and is running on background after each backup session completion.

If RAM capacity on the storage node is not sufficient to deduplicate large amounts of data, the indexing activity may fail. The backups will continue to run successfully. You can add more RAM to the storage node, or delete unnecessary backups and run the compacting task. After the next scheduled backup, the indexing will start again.

Deduplication Database

The Acronis Backup Advanced storage node with a deduplicated vault maintains the deduplication database. The deduplication database contains the hash values of all data blocks stored in the vault, except for those that cannot be deduplicated, e.g., encrypted files.

The deduplication database is stored on the storage node's local disk. You can specify the database path when you create the vault.

The size of the deduplication database is about 0.05 percent of the total size of unique data stored in the vault. In other words, each terabyte of new (unique) data adds about 500MB to the database.

If the database is corrupted or the storage node is lost, the vault itself survives. The new storage node can rescan the vault and recreate the vault database and the deduplication database.

The vault database is a tiny database, which contains the metadata of all archives stored in the vault. The metadata includes information about the backup sizes, machine names, indexing state, etc. It is stored in XML files in the **.meta** folder inside the deduplication database path.

Deduplication Data Store

The deduplication data store is a special set of files within a centralized, managed, deduplicated vault. Unique data blocks, except those that cannot be deduplicated are stored here (see the "Deduplication Restrictions" section for more details).

The Acronis storage node keeps a separate deduplication data store for each deduplicated vault. As a result, deduplication applies to the specific vaults, but not across vaults.

Recovery

During recovery, the Acronis Backup agent requests the data from the Acronis storage node through a proprietary, secure protocol. The storage node reads backup data from the vault and if a block is referenced in the deduplication data store, the storage node reads data from it. For an agent, the recovery process is transparent and independent of the deduplication.

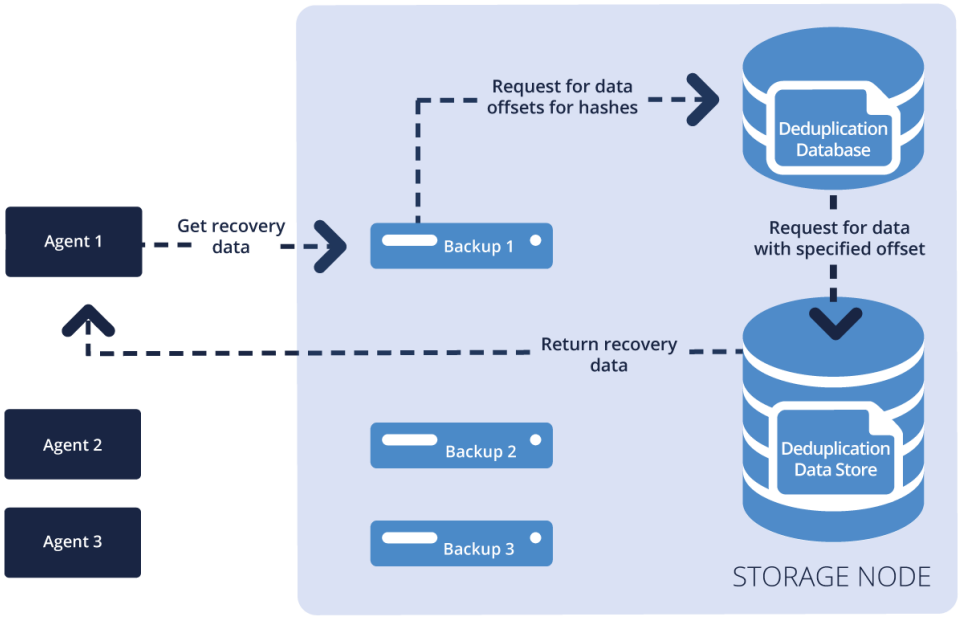

Figure 4

If the indexing task is not yet complete, then part of the data is still in the temporary file. Recovery is not affected — the only difference in this case is the storage node takes backup data from the temporary file instead of the deduplication data store.

Compacting

After one or more backups are deleted from the vault — either manually or through retention rules — the data store may contain blocks, which are no longer referenced by any backup. These *orphan* blocks are deleted by a compacting task that the storage node runs on schedule.

By default, the compacting task runs every Sunday at 03:00. You can also trigger this task manually via “acrocmd” CLI utility.

Here is how compacting works. First, the storage node scans through all backups in the vault and marks all referenced blocks as used (the appropriate hash is marked as used in the deduplication database). Second, the storage node scans through all blocks in the old deduplication data store, moves all used blocks to the newly created data store, and updates appropriate records in the deduplication database. Last, the newly created data store replaces the old one and all orphan blocks are deleted.

The compacting process may require additional system resources. That is why the compacting task runs only when a sufficient amount of data has accumulated. The threshold is determined by the **Compacting Trigger Threshold** configuration parameter.

Compression and Encryption

The Acronis Backup agent compresses the backed up data before sending it to the server. Hash values for each data block are calculated before compression. If two equal blocks are compressed with different levels of compression, they are still recognized as duplicates. Data blocks are not recompressed on the storage node, and are stored with the same compression level as specified in the backup plan.

Backups, encrypted on the client side, are not deduplicated for security reasons.

To leverage both encryption and deduplication, Acronis Backup Advanced supports encrypting the managed vault itself. Data stored in this vault is encrypted with the AES cryptographic algorithm. During recovery, the data is transparently decrypted by the storage node using a vault-specific encryption key stored on the storage node. If the storage medium is stolen or accessed by an unauthorized person, the vault cannot be decrypted without access to the storage node server.

The AES cryptographic algorithm operates in the cipher-block chaining (CBC) mode and uses a randomly generated key with a user-defined size of 128, 192 or 256 bits. The larger the key size, the longer it takes for the program to encrypt the archives that are stored in the vault and the more secure the archives are.

The key is also encrypted with AES-256 using an SHA-256 checksum that is based on a user-entered password. This checksum, used to generate the encryption key, is not stored anywhere on the disks and is only used for verification purposes. With this two-level security, the archives are protected from any unauthorized access, but it is impossible to recover a lost password.

Replicating Data

Replication is a process of copying backup data to another location. By default, the data is copied immediately after backup. The user has the option to postpone copying the backup via the replication inactivity time option.

When the data is replicated/staged to a deduplicated vault, deduplication still applies — duplicate data blocks are not resent if they are already stored in the destination vault. In this case, only the unique data is transferred between the two storage nodes. The source storage node behaves like an agent with deduplication-on-target enabled — it reads backup data blocks, generates hashes for it, and sends the request for the hashes to the target storage node. If there are some blocks with these hashes in the deduplication data store, these blocks are not sent over the network. As a result, there is the same network bandwidth economy during replication as there is when backing up data.

Deduplication Enhancements in Acronis Backup Advanced v11.7 and higher

The improvements listed below become effective when the Acronis Backup Advanced 11.7 and newer versions agents back up to deduplicating vaults created on version 11.7 and newer storage nodes. When a new agent backs up to the old vault or an old agent backs up to the new vault, the older deduplication algorithm is used. To apply the new deduplication algorithm to old backups, you need to import the backups into a newly created vault.

1. Variable block size (64-256KB) is used instead of fixed 4KB block sizes
2. 40-160MB RAM per 1TB of data is required (compared to 3,000MB RAM per 1TB in version 11.5 update 6, and 8,000MB RAM per 1TB in earlier releases)
3. Example: for 40TB of unique data, approximately 2.4GB of RAM is required (vs. 120GB in version 11.5 update 6)
4. SSD disks are not required for the deduplication database
5. Regular HDDs are sufficient to achieve optimal performance
6. Recovery from a deduplicated backup is now 40 percent faster
7. Backup speed is not impacted when the deduplication database size increases
8. The storage node startup time on large data sets is reduced to one to three minutes
9. Validating deduplicated backups is 2.5 times faster
10. Compacting the data store is 2.5 times faster

With these new enhancements, you will see a drastic improvement in the compacting duration time. See Figure 5 below.

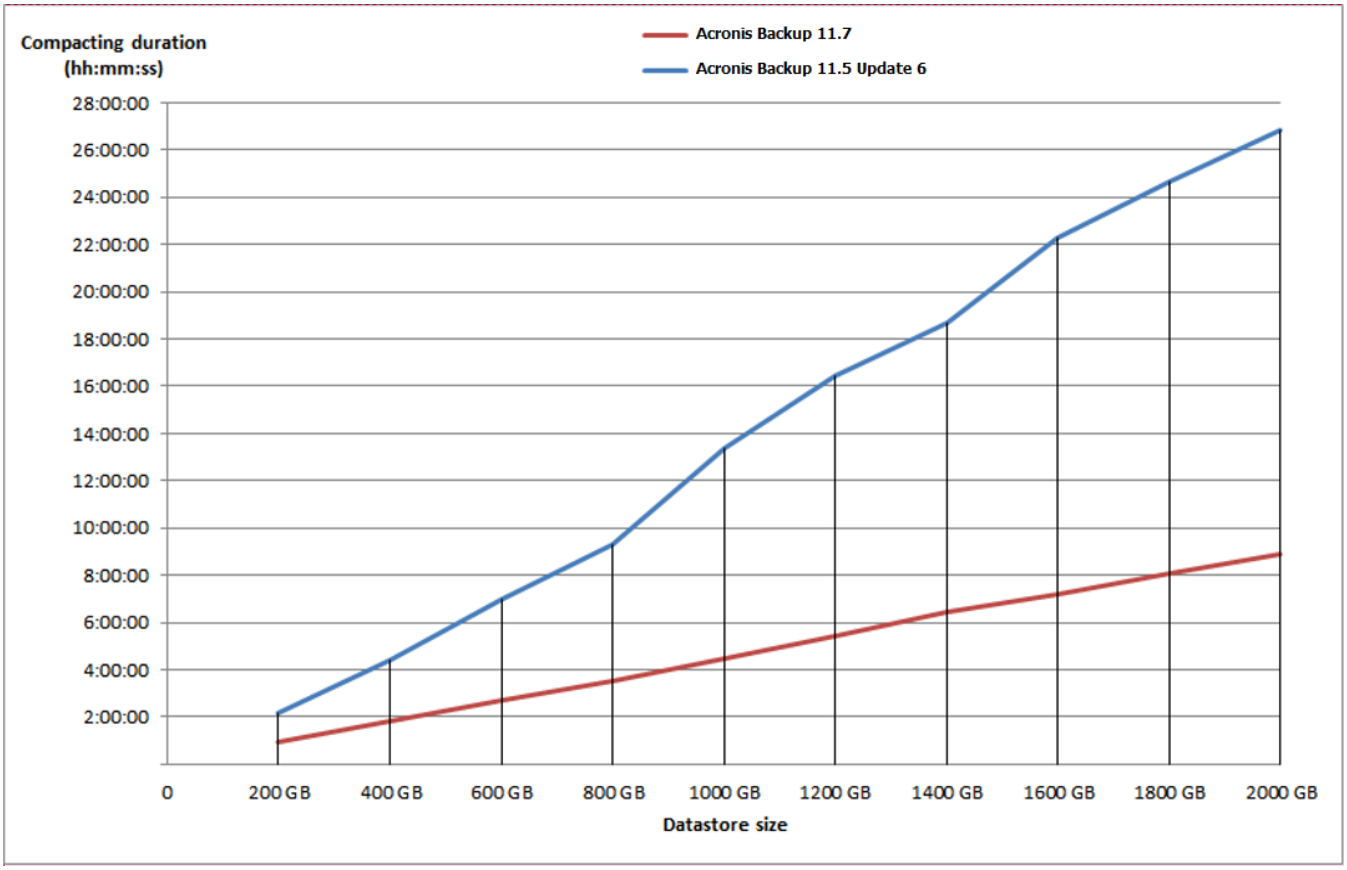

Figure 5: Compacting Improvements in Acronis Backup Advanced 11.7 and newer versions versus Acronis Backup Advanced 11.5 Update 6

When to Use Deduplication

Deduplication has the greatest impact when the deduplication ratio has the lowest value. Here is the formula for the deduplication ratio calculation (for more calculations, please refer to "*Appendix A: Estimating Required Storage Capacity*"):

Deduplication ratio = Unique data percentage + (1 – Unique data percentage) / Number of machines

This means that:

1. Deduplication is most effective in environments where there is a lot of duplicate data on each machine
2. Deduplication is most effective in environments where you need to back up a lot of similar machines/virtual machines/applications

In addition, deduplication can help in other scenarios, such as when you are trying to optimize your wide area network (WAN).

Let us look at some typical use cases.

Use Case 1: Big Environment With Similar Machines

Environment

One hundred similar workstations need to be backed up. The workstations were initially deployed using Acronis Snap Deploy.

Deduplication effect

The workstations were deployed from a single image using Acronis Snap Deploy, so the operating system and generic applications that run on all machines are identical. As a result, there are many duplicates. Deduplication is even more effective because there are a large number of workstations.

Conclusion

Deduplication is very effective in this scenario because it minimizes storage capacity and saves storage costs.

Use Case 2: WAN Optimization

Environment

Forty similar workstations in the main office need to be backed up to a remote location.

Deduplication effect

We do not know if the workstations were deployed from a single image. However, similar types of operating systems often have many similar files. Let us assume that 50 percent of the data on each PC is unique — still quite good for deduplication:

Deduplication ratio = 50% + (100% – 50%) / 40 = 51.25%

The approximate storage and network traffic savings is 48.75 percent (100% – 51.25%), which means deduplication cut these requirements almost in half. Since the systems are backed up to a remote location, the WAN connection can be relatively slow. Halving the traffic offers a big advantage.

Conclusion

Deduplication is an effective solution for this case because it optimizes the network WAN.

Use Case 3: Business-Critical Application Servers

Environment

Five application servers, all with different applications, need to be backed up. The total amount of data is 20TB.

Deduplication effectiveness

Application servers host huge amounts of data and different applications. This means there will be very few, if any duplicates at all. Moreover, the total amount of data to be backed up and processed is very high.

In this case, the storage node will index large amounts of data but little benefit will be realized because there are no duplicates. In the worst-case scenario, a single storage node may not be able to process all the backups in one day.

Conclusion

Deduplication is not effective for this case. Backing up to a simple, high-capacity, network-attached storage (NAS) is a better solution.

Deduplication Restrictions

Common Restrictions

If you protect the archive with a password and encryption, deduplication does not apply. Data blocks of password-protected, encrypted archives are stored in the backups (.tib files), just as they would be in a non-deduplicating vault.

To encrypt an archive while still deduplicating it, leave the archive unencrypted and encrypt the deduplicating vault itself with a password. You can do this when creating the vault. The data will still be safe at rest as well as in transit. The latter is enforced by the encryption of the proprietary backup protocol.

Disk-Level Backup

If the volume's allocation unit size — also known as cluster size or block size — is not divisible by 4096 (4KB), deduplication of disk blocks will not apply.

Tip: The allocation unit size on most of the New Technology File Systems (NTFSs) and ext3 volumes is 4KB, which allows for block-level deduplication. Other examples of allocation unit sizes that allow for block-level deduplication include 8KB, 16KB, and 64KB.

File-Level Backup

A file is not deduplicated if the file is encrypted by the operating system and the "**In archives, store encrypted files in decrypted state"** checkbox in the backup options is not selected. (Note: it is cleared by default).

In an NTFS file system, a file may have one or more additional sets of data associated with it — often called alternate data streams.

When these files systems are backed up, so are all its alternate data streams. However, these streams are never deduplicated — even when the file itself is.

Deduplication Best Practices[1]

Deduplication is a complex process and its speed relies on multiple factors. The most important factors that influence deduplication speed are:

- The speed of access to the deduplication database (IOPS)
- The RAM capacity of the storage node
- The number of deduplicating vaults created on the storage node

The following are recommendations on how to increase deduplication performance.

Put the Deduplication Database and Deduplicating Vault on Separate Physical Devices

To increase the speed of access to a deduplication database, the database and the vault must be located on separate physical devices.

It is best to allocate dedicated devices to the vault and the database. If this is not possible, avoid placing a vault and/or database on the same disk with the operating system. The operating system performs a large number of background hard-disk read/write operations, which significantly slows down the deduplication.

Selecting a Disk for a Deduplication Database

1. The database must reside on a fixed drive. Please do not place the deduplication database on external detachable drives or network devices, such as an over-the-LAN iSCSI.
2. The volume, where you plan to store the deduplication database, must have at least 10GB of free space. When backing up a large number of machines, you may need even more free space.
3. The disk space required for a deduplication database can be estimated by using the following formula:

$$S = U * 90 / 65536 + 10$$

where:

[1] *Note*: These best practices are applicable to Acronis Backup Advanced 11.7 and newer versions.

S – disk size, in GB

U – estimated amount of unique data in the deduplication data store, in GB.

For example, if the planned amount of unique data in the deduplication data store is U = 5TB, the deduplication database will require free disk space of:

$$S = 5 * 1024 * 90 / 65536 + 10 = 17GB$$

Selecting a Disk for a Deduplicating Vault

To avoid losing data, use a redundant array of independent disks or RAID. RAID 10 is your best option. RAID 5 and RAID 6 are not recommended because they are prone to double-rebuild failures on large, hard-disk capacities. Avoid RAID 0 because it does not have redundancy and avoid RAID 1 because it does not improve performance. You can use either local disks or SAN.

40-160MB of RAM per 1TB of Unique Source Data

When the limit is reached, deduplication will stop but backup and recovery will continue to work. In general, the more RAM you have, the more you can store larger volumes of unique data.

Only One Deduplicating Vault on Each Storage Node

Create only one deduplicating vault on a storage node. Otherwise, the RAM is divided proportionally to the number of the vaults and will impact the performance of deduplication.

Each vault consumes 1 chunk of 512MB of RAM for its deduplication database upon creation. The next chunk of 512MB will be allocated only after the first one is nearly (90% +) consumed. The size of these increment chunks to be consumed is defined in Windows registry via "*PHashIndexMemoryFullGrowthPercent*" value (Default=100%, so RAM consumption will increase by 512MB) under "HKEY_LOCAL_MACHINE\SOFTWARE\Acronis\ASN\Configuration\StorageNode" key on Acronis Storage Node machine.

64-bit Operating System

The storage node requires a 64-bit operating system. Do not share the storage node with other resource-intensive applications, such as database management systems (DBMS) or enterprise resource planning (ERP) systems.

Minimum of Quad-Core 2.5 GHz CPU

Use a minimum of 2.5 GHz CPU with at least four cores. Multi-CPU systems are also supported.

Ample Free Space in the Vault

Indexing a backup may require as much free space as the backup occupies immediately after saving it to the vault. Without compression or deduplication, it will be equal in size to the original backed up data.

Do not create a deduplicating managed vault on a FAT32 volume. The maximum file size in a FAT32 file system is 4GB. However, the vault stores the deduplicated items in two very large files that may be greater than 4GB. When this limit is reached, the storage node may stop functioning.

High-Speed LAN

A 1GB LAN is recommended to allow the storage node to receive five to six backups with deduplication in parallel.

Back Up a Typical Machine Before Backing Up Multiple Machines With Similar Content

When backing up several machines with similar content, back up one machine first and wait until indexing is complete. Once the first machine's backup has been indexed, most of the data will be in the deduplication data store and will not be transferred/stored again. This means that the other machines will back up much faster due to the efficient deduplication.

Back Up Machines at Different Times

When you back up a large number of machines, schedule the backup operations over a period of time. To do this, create several backup plans with different schedules, or use the spreading functions of the centralized backup plan in Acronis Backup Advanced 11.7 and newer versions.

(*Applicable to v11.7 and below only*) Use Fast Cataloging

Indexing starts after cataloging is complete. To reduce the time required for backup processing, switch automatic cataloging to the **fast** mode. You can start **full** cataloging manually outside of the backup window.

Useful Links

You can find more information at:

1. Acronis website: http://www.acronis.com
2. Acronis Backup 12.5 Advanced online help: http://www.acronis.com/en-us/support/documentation/AcronisBackup_12.5
3. Acronis Knowledge Base: https://kb.acronis.com

Appendix A: Estimating Required Storage Capacity

Deduplication technology saves on storage capacity and the required storage capacity depends on the following factors:

1. Backup scheme
2. Retention period
3. Compression ratio
4. Number of machines
5. Amount of data to back up per machine
6. Unique data percentage
7. Daily incremental backup size

Here is some advice on how to estimate the required storage capacity.

Backup Scheme and Retention Period

The backup scheme and retention period define how many backups will be stored after a specific period of time. The following table shows the numbers of backups (full, incremental, and differential) after specific periods, depending on the backup scheme and retention period.

Scheme	4 weeks			6 months			1 year			2 years		
	Full	Diff.	Incr	Full	Diff.	Incr	Full	Diff.	Incr	Full	Diff.	Incr
Simple	1	0	27	1	0	181	1	0	363	1	0	727
GFS[2] (keep monthly backups indefinitely)	2	3	4	7	5	4	14	5	4	27	5	4
GFS (keep monthly backups for 1 year)	2	3	4	7	5	4	14	5	4	14	5	4

[2] Grandfather – Father – Son (GFS): a monthly/weekly/daily backup scheduling and retention scheme in Acronis Backup Advanced.

The size of a full backup changes because of new data. For example, if the daily incremental change rate is 1 percent and backups run only on workdays, the full backup size after 52 weeks may be up to 3.6 times higher than the first full backup.

To estimate capacity, use the projected backup size at the end of the retention period. For more accurate results, take an average between the initial backup size and the projected, last backup size.

The incremental backup size depends on the frequency of backups. Daily incremental backup size is one of the parameters defined initially. It is possible to configure backups to run weekly, monthly, or at any other frequency so the real incremental backup size needs to be estimated.

Differential backup size depends on the amount of daily changes and the number of days from the last full backup. To estimate the size of largest differential backup, calculate the number of days between the first full and last differential backups. For example, in the case of GFS, the longest period between full and differential backups is 15 days, so the maximum differential backup size will be the size of the daily incremental data multiplied by 15. For a more accurate estimation, use the average between the first and the last differential backups of the same full backup.

Compression Ratio

The data in all types of backups is usually compressed. Default, "normal" compression levels reach from 40 to 60 percent, meaning that the compressed data is 40 to 60 percent of the original data.

Number of Machines and Amount of Data

The number of machines affects your deduplicated backup in several ways:

1. Backing up more than one machine in parallel creates additional LAN traffic
2. Parallel connections from multiple machines requires the storage node to handle multi-threaded backups
3. More machines mean more backed up data

The deduplication ratio depends on the number of machines: the more machines you back up, the more you will save on storage.

Unique Data Percentage

The amount of unique data on a machine depends on the role of the system. The "percent unique" numbers below are derived from Acronis' experience with its customers and may vary in your environment.

1. Virtual machines: 30 percent unique
2. Office workstations: 50 percent unique
3. Database servers: 65 percent unique
4. File servers: 75 percent unique

Daily Incremental Backup Size

Daily incremental backup size is the amount of data which is added or changed on the machine daily. It is equal to the size of the daily incremental backup (before compression and deduplication).

Storage Capacity Calculations

When you plan the storage capacity of your system, use the following formulas. Some of them are for initial backup planning (as it affects initial backup time) and the rest are for total capacity planning.

Initial Storage Capacity

The following formulas help you calculate the first backup size before incremental and differential backups:

*Total data size = Number of machines * Amount of data to back up per machine*

*Full non-deduplicated backup size = Total data to back up size * Compression ratio%/100%*

Deduplication ratio = Unique data percentage/100% + (1 – Unique data percentage/100%) / Number of machines

*Initial storage space = Full non-deduplicated backup size * Deduplication ratio*

Storage Space at the End of the Period

Estimating initial storage space is just part of backup storage planning. You must also plan for the storage capacity at the end of the period.

*Total backup data size = Number of machines * (Data size per machine + Daily incremental backup size * No of weeks * 5)*

*Full non-deduplicated backup size = Total backup data size * Number of full backups * Compression ratio%/100%*

*Incremental non-deduplicated backup size = Incremental backup size * Number of machines * Number of incremental backups * Compression ratio%/100%*

*Differential non-deduplicated backup size = Differential backup size * Number of machines * Number of differential backups / Compression ratio*

All non-deduplicated backup size = Full non-deduplicated backup size + Incremental non-deduplicated backup size + Differential non-deduplicated backup size

Deduplication ratio = Unique data percentage + (1 – Unique data percentage) / Number of machines

*Total backup storage space = All non-deduplicated backup size * Deduplication ratio*

These formulas assume:

1. The daily incremental backup size is a data addition rate, but in reality, some data is changed and replaces the existing data on a machine, so the actual number can be lower.
2. The period is in weeks, and five is used as the number of work days. Holidays count as workdays.

Appendix B: Upgrading to Acronis Backup Advanced 11.7 and newer versions Deduplication

To upgrade your vault created by Acronis Backup Advanced 11.5 Update 6 and lower follow these steps:

1. Upgrade your Acronis Backup Advanced management server and storage node(s) to the latest version
2. Upgrade your Acronis Backup Advanced agents to the latest version
3. Create new deduplication vault(s) on your storage node(s)
4. Edit the backup tasks to use new the deduplication vault as a target
5. (Optional) Export the old backups from the old deduplication vaults to newly created vaults to reapply the new deduplication algorithm. Note, this operation can be time-consuming and resource-intensive.

Due to a new deduplication database format, it is not possible to apply new deduplication improvements to an existing vault.

About Acronis

Acronis sets the standard for hybrid cloud data protection through its backup, disaster recovery, and secure file sync and share solutions. Powered by the Acronis AnyData Engine and set apart by its image technology, Acronis delivers easy, complete and affordable data protection of all files, applications and operating systems across any environment — virtual, physical, cloud and mobile.

Founded in 2003, Acronis protects the data of over 5 million consumers and 500,000 businesses in over 145 countries. With more than 100 patents, Acronis products have been named best product of the year, and cover a range of features, including migration, cloning and replication. Today,

Acronis solutions are available worldwide through a global network of service providers, distributors and cloud resellers. For additional information, please visit www.acronis.com

Follow Acronis on Twitter: http://twitter.com/acronis.