HP Data Protector software object consolidation

"Best Practices"

Table of contents

Executive summary

This white paper provides information on how to use and configure the object consolidation feature introduced in HP Data Protector software version 6.0 effectively.

Solution description

HP Data Protector software A.06.00 introduces an advanced backup solution called synthetic full backup. This solution enables you to create synthetic full backups and virtual full backups with an operation called object consolidation allowing implementing an incremental forever backup strategy.

The incremental forever paradigm means that except for the first backup where a full backup is performed, only incremental backups are executed. This concept presents the most efficient way of backups; only the changed data is backed up.

However, without object consolidation, the restore process would last far too long since all backup sessions would have to be restored. Due to this behavior, regular full backups are required and all backups would need to be protected permanently. Object consolidation removes this drawback.

During the object consolidation session, HP Data Protector software reads the backed up data from the source media, merges the data, and writes the consolidated version to the target media.

An object consolidation can result either in a synthetic full or a virtual full backup.

A synthetic full backup is the result of an object consolidation operation where a restore chain of backup objects is merged into a new, synthetic full version of this object. A synthetic full backup is equivalent to a conventional full backup in terms of restore speed.

A virtual full backup is an efficient type of synthetic backup where data is consolidated using pointers instead of being copied. It is performed if all the backups (the full backup, incremental backups, and the resulting virtual full backup) are written to a single HP Data Protector software file library using distributed file media format.

There are a number of advantages performing object consolidation:

- **Space efficient if virtual full**: Only one initial full backup; no file is stored multiple times, and therefore, has the potential to reduce required backup space/media.
- **Faster backup**: Only changed files need to be backed up enables fast backup of file servers hosting millions of files allowing more frequent backup.
- **Less time to manage**: No tedious tape management with backup to disk.
- **Faster restore of individual file**s: Single file restore from disk means that the tapes can be stored off-site.

Most importantly, Synthetic Full also provides fast disaster recovery.

Object consolidation prerequisites

The following prerequisites are required for successful object consolidation:

- All the backups that will be consolidated have to be performed with the enhanced incremental backup option enabled.
- All incremental backups that will be consolidated must reside in one file library.
- The restore chain should be complete, meaning that all the object versions that comprise it have the `Completed` or `Completed/Errors` status and all the media holding these object versions are available.
- The necessary backup devices have to be configured and the media prepared.
- Media Agents have to be installed on every system that will participate in an object consolidation session.
- Appropriate user rights for starting an object consolidation session need to be configured. The same user rights apply as for backup.
- All the backups—full, incremental, and virtual full—must reside in one file library that uses distributed file media format to perform a virtual full backup.

Object consolidation limitations

The object consolidation has certain limitations:

- The destination devices must have the same or a larger block size than the source devices.
- The same medium cannot be used as a source medium and as a target medium in the same object consolidation session.
- While the source media are being read, they are unavailable for restore.
- The Reconnect functionality between agents is not available in an object consolidation session.
- Object consolidation is not supported on Novell NetWare and MPE/iX systems
- Object consolidation is supported only for file system backup
- Media that use distributed file media format cannot be exported or imported
- File libraries configured with `Appendable` allocation policy are not supported
- ZDB is not supported
- Cluster environments requires manual configuration changes
- Virtual full backups are tracked as synthetic full backups inside the GUI
- Object consolidation does not support HP Data Protector software Disk Agent (DA) encrypted data

Note:
Whenever you change the setting of the software compression or encode option in the backup specification, a full backup must be performed as a base for subsequent object consolidation. Setting the encode option can also turn AES encryption on if AES encryption is enabled. Object consolidation with Disk Agent software AES encryption is not supported.

Also the same compression/encode plug-in should be used for full backups and related incremental backups.

Backup retention management

Backup retention management is an important aspect to consider when the incremental forever backup strategy is being followed.

A successful consolidation process relies on the availability of the object versions to be consolidated.

If backup media containing a specific version is not available, or an object version which belongs to the consolidation chain has expired, the consolidation for that object will fail.

While scheduling backups, you may have set the same protection period for full and incremental backups. This means that incremental backups are protected for the same duration as the relevant full backup. The consequence of this is that your data will actually be protected only until the full backup expires.

You cannot restore or consolidate incremental backups that have been based on expired full backups. Configure the protection for your full backups so that they are protected for longer than your incremental backups.

The time difference between the protection for the full backup and the incremental backup should be the amount of time between the full backup and the last incremental backup before the next full backup. For example, if you run incremental backups Monday through Friday and full backups on Saturday, you should set the protection of the full backup to at least 6 days more than for the incremental backups. This will keep your full backup protected and available until your last incremental backup expires.

Data Protection: Specifying how long data is kept on the media

Configuring protection policies is extremely important for the safety of your data and for successful management of your environment. See the HP Data Protector Concepts Guide for more detailed information on how to define these policies.

Based on your company data protection policies, you have to specify how long your backed up data is kept on the medium. For example, you may decide that data is out of date after three weeks and can be overwritten during a subsequent backup.

Types of object consolidation

You can start an object consolidation session interactively or specify an automated start of the session.

Automated object consolidation

HP Data Protector software offers two types of automated object consolidation: post-backup object consolidation and scheduled object consolidation.

Post-backup object consolidation

Post-backup object consolidation takes place after the completion of a backup session. The finished backup session that triggers the object consolidation is specified by the relevant name of the backup specification in the automated post-backup object consolidation specification. It consolidates selected objects according the automated object consolidation specification.

Scheduled object consolidation

Scheduled object consolidation takes place at a user-defined time. Objects backed up during different backup sessions can be consolidated in a single scheduled object consolidation session.

Alternatively you can start an interactive object consolidation.

Interactive object consolidation

You can select objects for interactive consolidation from the Objects or Sessions starting point, depending on your needs. You cannot save an interactive object consolidation specification; you can only start an object consolidation session.

How to Consolidate Objects

To begin with, create an object consolidation specification. In the specification, select the object versions you want to consolidate, the media and devices you want to use, and the session options.

Selection of devices

You may need separate devices for reading full backups, incremental backups, and writing synthetic full backups. Optionally, the same device can be used for reading full backups and incremental backups. The destination devices can have a larger block size than the source devices. However, to avoid impact on performance, it is recommended that the devices have the same block size and are connected to the same system.

Devices that are not available at the beginning of a session cannot be used in that session. If a media error occurs, the device with errors will be avoided within that session.

Devices need to be carefully selected. See device mapping details in the Device Mapping section further down in this document.

Object consolidation options

You can specify data protection, catalog protection, and logging level in the object consolidation specification. Equivalent options are used for backup as well.

Selection of the media set

If an object version that will participate in consolidation has copies residing on different media sets, any of those media sets can be used as a source. By default, HP Data Protector software automatically selects the most appropriate media set. You can influence the media set selection by specifying the media location priority only for object copy versions but not for different media sets within a restore chain. Preferably, you could specify a media for consolidation that is not vaulted offsite.

HP Data Protector software will use the media set with the highest priority (priority 1 is the highest, priority None is the lowest) if more than one media set equally matches the conditions of the media set selection algorithm.

The overall process of media selection is the same as for restore. While consolidating objects interactively, you can manually select the media set to be used. You cannot select media when configuring automated object consolidation as the backup of the objects is often performed at a later time.

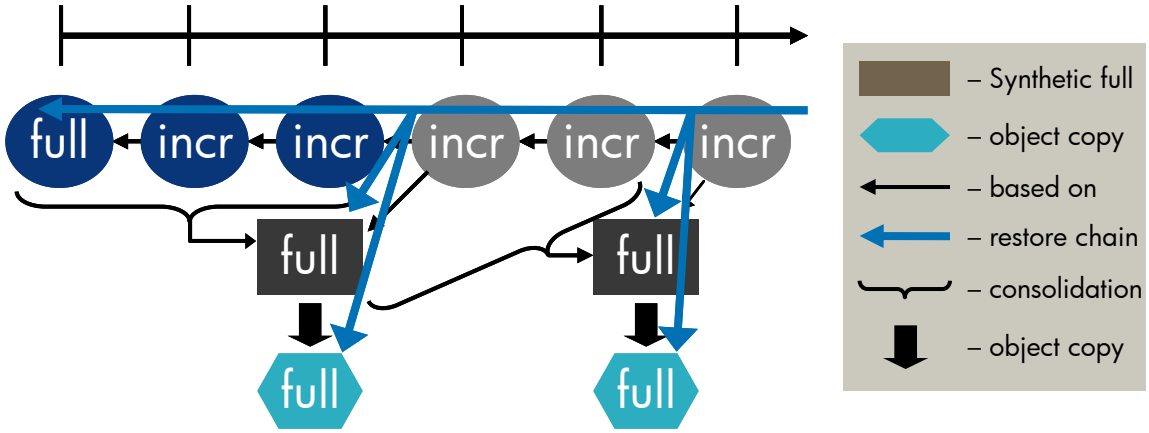
In HP Data Protector software 6.0, automatic media-set-selection (AMSS) is enhanced to consider besides object version copy also different restore chains. This result was the following algorithm:

1. Start with m media-sets
2. Filter out all unusable media-sets, for example, media which are poor or have expired
3. Select the shortest restore chain
4. If the restore-chain has alternatives (object copy), pick according to location priority
5. End up with one media-set

Point 3 can be swapped with point 4 by setting the (undocumented) `UserSpecifiedMediaPriorityHasHigherImportance` global parameter. The usage of this parameter is not recommended since it gets very difficult to predict which restore chain is used for the restore. It should be used only for troubleshooting purposes.

Automatic media-set-selection (AMSS) example

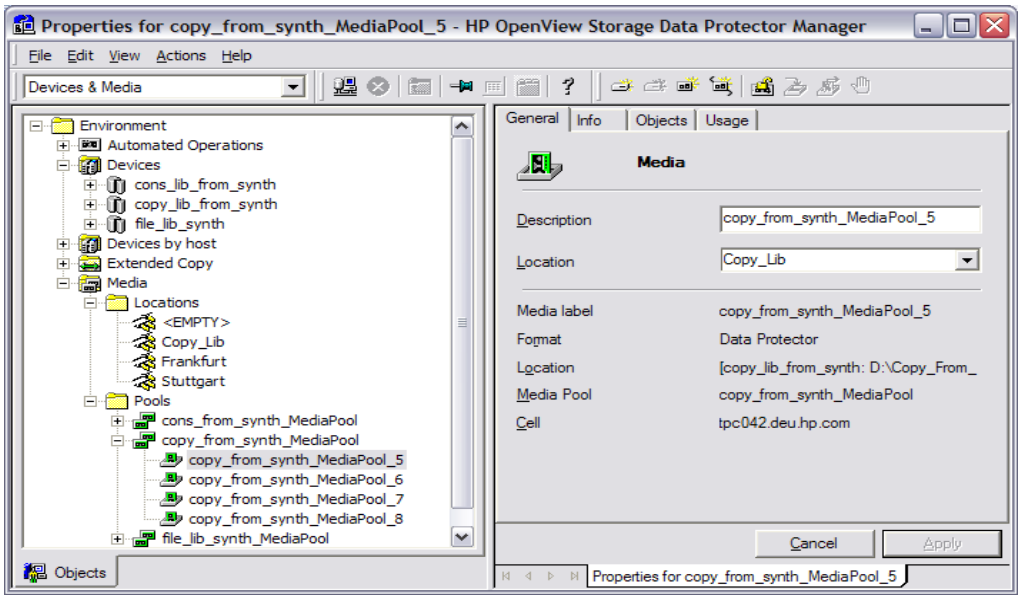
Figure 1. Restore chains



This example shows a backup sequence of one full backup with subsequent incremental backups. After each second incremental backup, a synthetic full is created which is copied to tape. There exist now five restore chains. In case of a restore, HP Data Protector software can now choose between five different restore possibilities. After filtering out all unusable media, HP Data Protector software checks for the shortest restore chain. Since some of the items within the chain (synthetic full) have a copy, it must be determined which one to use. This is done based on a manual selection or on the location priority.

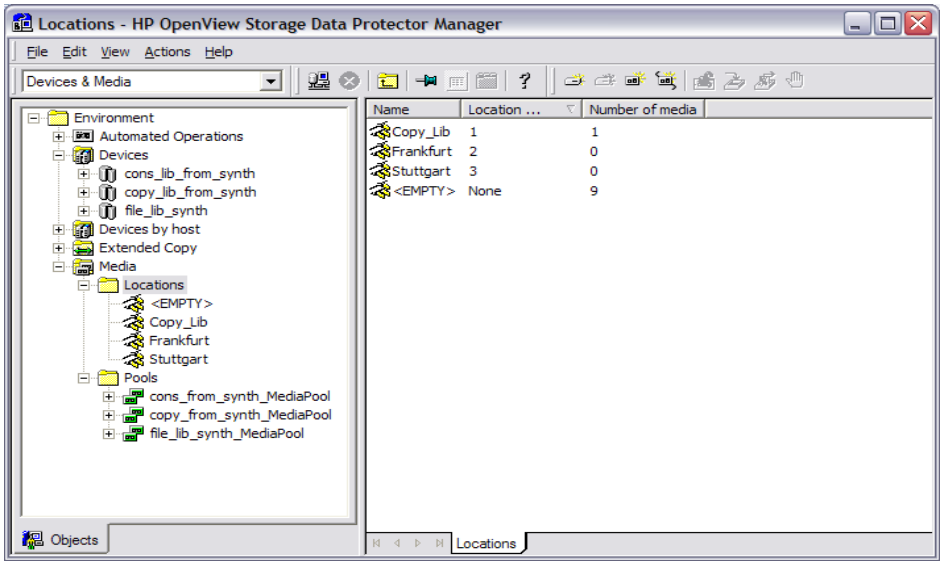
Figure 2 and Figure 3 shows the location priority relevant configuration details in the HP Data Protector software GUI.

Figure 2. File Library Copy Media at `Copy_lib` offsite location



The location detail cannot be assigned to a file library media pool but only to individual media within a file library pool. If a media location priority change is a requirement, a post-backup script can help to change the priority automatically upon finishing a backup run.

Figure 3. Media Location priority properties



Device Mapping

For the backup consolidation process, HP Data Protector software will use the devices to read from full backups and incremental backups that have been used to write during the backup process. In addition, a new device (writer) will be assigned to write the new consolidated backup to media.

If full backups and incremental backups reside in separate file libraries, a device mapping needs to be in place prior to the consolidation process. This is also true if consecutive interactive object consolidations are being performed.

While configuring an object consolidation specification, the HP Data Protector software creation wizard asks to define the source devices used to read the full backups and all required incremental backups.

All devices used for the incremental backups and for all full backups, which are part of the complete restore chains are, by default, reserved as source devices and hence can not be used as target device. This means that each further backup consolidation requires an additional target device. The consequence of this fact is that there might be no more target devices to select from.

In order to solve this issue, the source devices reserved to read the full can be mapped to another device, and therefore, can be freed up to be used as a target device.

Figure 4. Device selection without device mapping

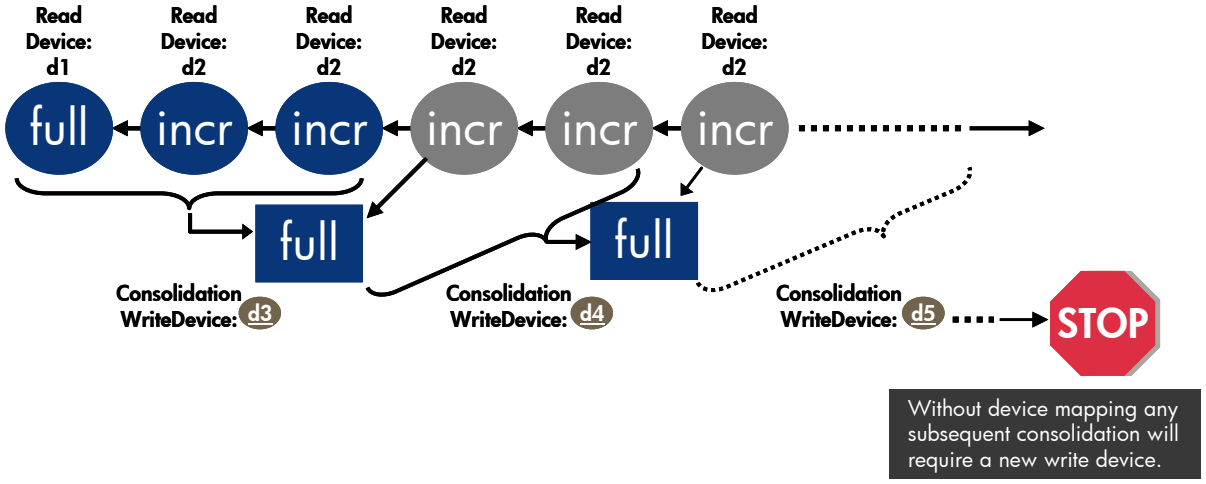
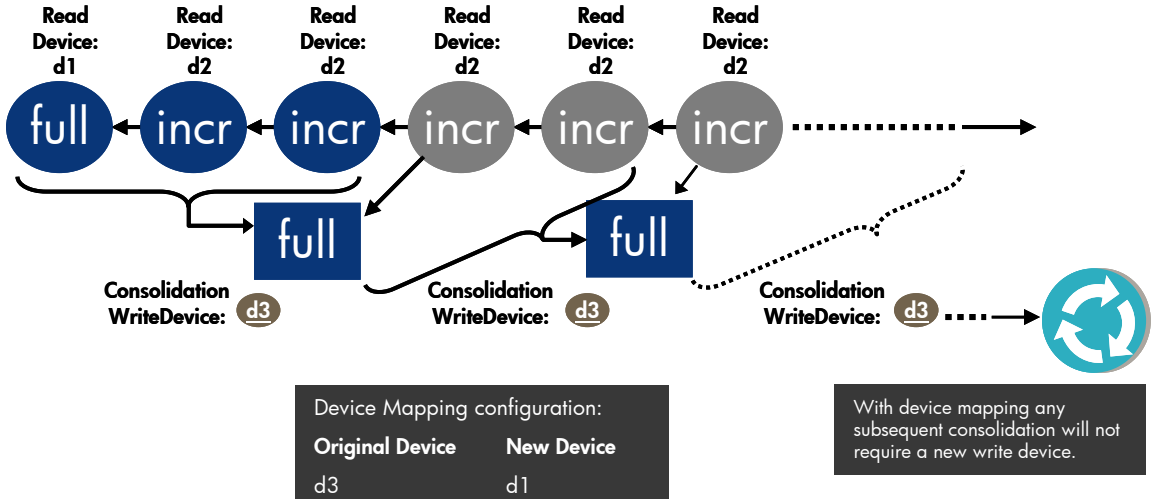


Figure 5. Device selection with device mapping



Device mapping example: Interactive consolidation

The following example of an interactive object consolidation with and without device mapping is based on a sequence (Figure 6) of full, incremental, incremental, synthetic, incremental, synthetic, incremental backups into a single file library with four writers.

Figure 6. Backup, consolidation sequence

Note: HP Data Protector software will automatically select all sessions/media that can potentially be consolidated into a new synthetic full backup. The choice of session/media finally being used for consolidation is defined by the selection algorithm of the Automatic Media Set selection described earlier in this document. Typically the shortest restore chain (session 2007/07/26–22 and session 2007/07/26–21 in this case) would be selected for consolidation.

Each synthetic full backup had used up one additional writer (Figure 7). The next consolidation routine will use the remaining writer3 to write the synthetic full backup (Figure 8).

A consequence of this is that no more writers will be available for any subsequent consolidation routines after the next one.

Configuration without device mapping

Figure 7. Read devices without mapping

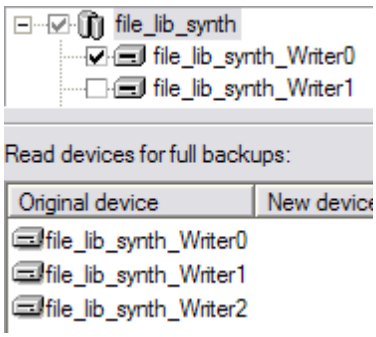
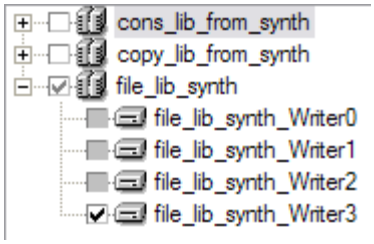


Figure 8. Writer device without mapping



With device mapping this issue can be solved easily.

Configuration with device mapping

Figure 9. Read devices with mapping

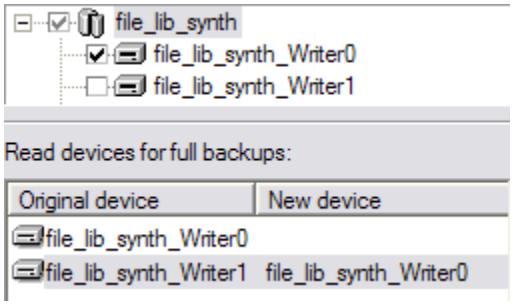
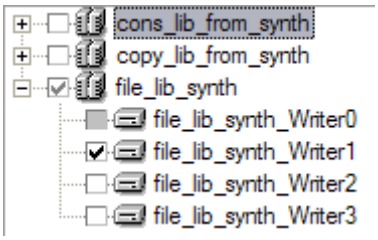


Figure 10. Writer device with mapping



The very first full backup was done to the `file_lib_synth_Writer0` device and the first synthetic full backup was done to the `file_lib_synth_Writer1` device. Without mapping, both devices are reserved and therefore can not be used for this consolidation. Hence, an additional device is required. The third consolidation requires also an additional device and so on.

The `file_lib_synth_Writer1` source device is mapped to `file_lib_synth_Writer0` as shown in Figure 9. This frees up this device, and therefore, can be used for the target device as shown in Figure 10.

The device mapping task can be very complex when many Read and Write devices are listed in the configuration menu. This circumstance needs to be considered and device selection needs to be done carefully.

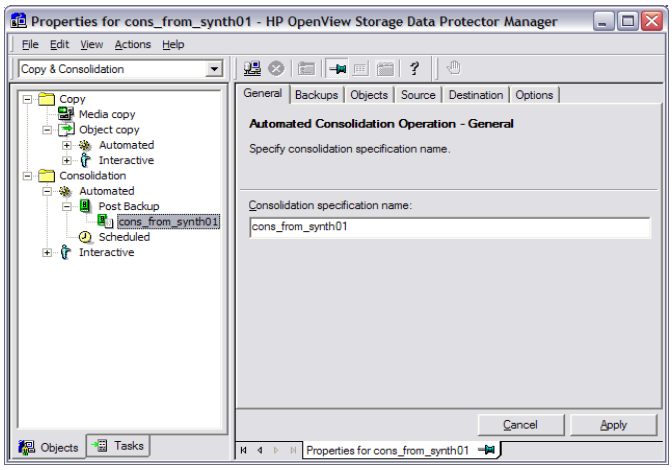
Note: The mapping within an interactive consolidation can only be done after the first interactive object consolidation. The reason for this is that the writer being used to write the consolidation is assigned after the first consolidation run.

Device mapping example: Automated post backup consolidation

This example demonstrates the full configuration sequence of an automated post-backup consolidation definition.

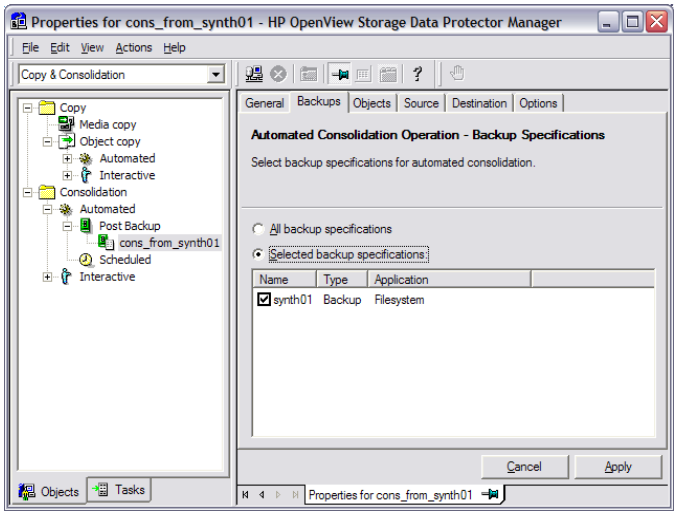
Figure 11 shows the General tab of the automated consolidation configuration wizard in which the name of the consolidation task is specified.

Figure 11. Consolidation specification name



Specify the backup datalist from which you would like to consolidate data from.

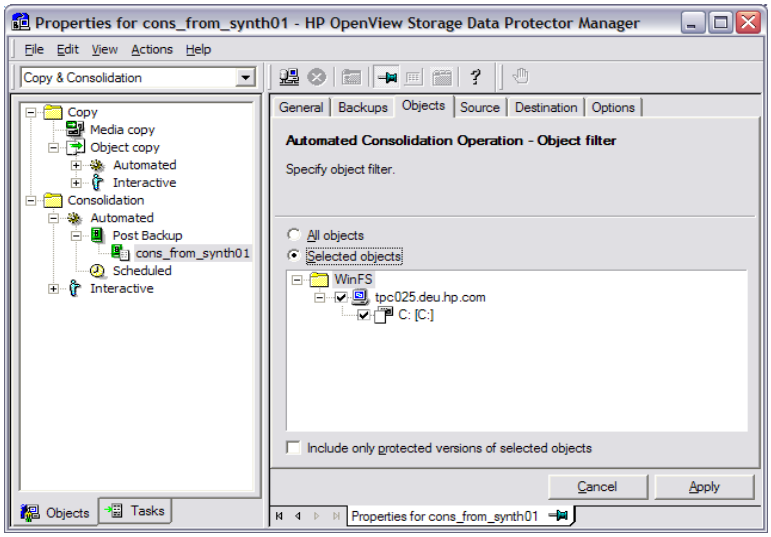
Figure 12. Backup datalist selection



Select objects from which you would like to consolidate data from.

Note: Initially only “All Objects” can be selected.

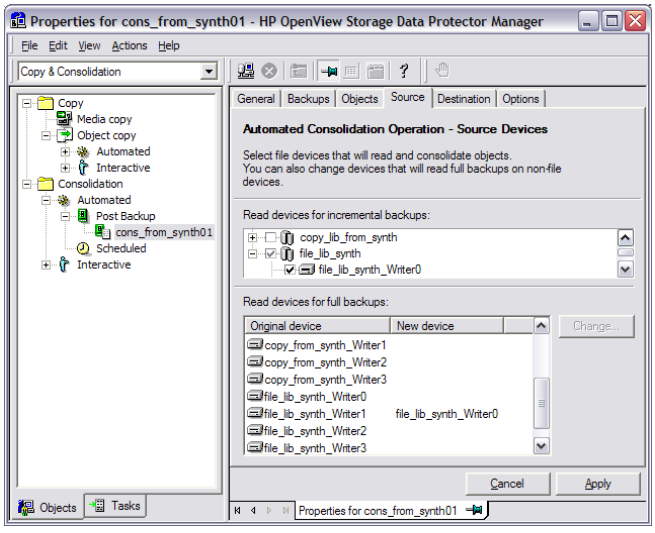
Figure 13. Backup object selection



In the Consolidation Source Devices dialog, you can specify the devices to read from during the object consolidation. For the configuration of the automated consolidation operation, the list of devices for full backups is populated with all file library readers configured in HP Data Protector software. A mapping can be done even without having done an initial consolidation before.

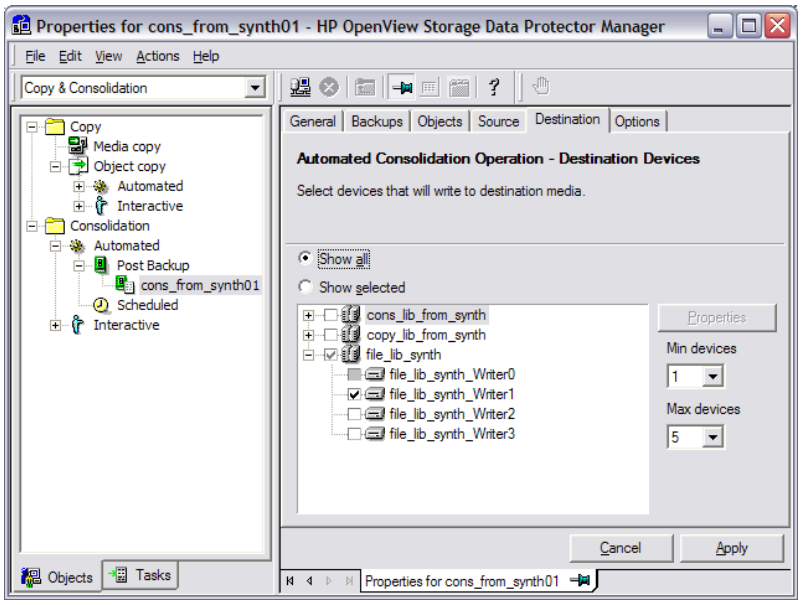
Note: A device mapping is required if source and target are in different file libraries or for interactive consolidations.

Figure 14. Consolidation source devices



Due to proper device mapping as shown in the previous screen, there are always enough writer devices to write the consolidated object.

Figure 15. Object consolidation writer

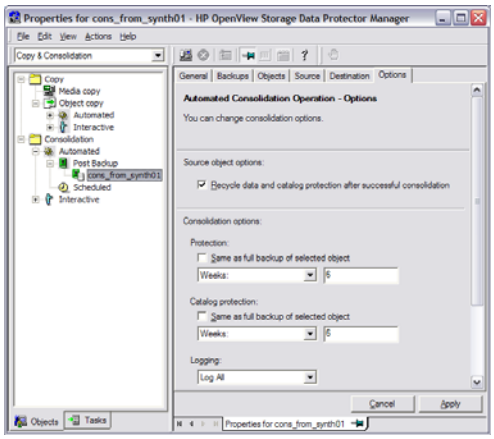


In the consolidation options dialog, specify how backup data and catalog data is being processed after a successful consolidation.

Option: Recycle data and catalog protection after successful consolidation

Select this option to remove data and catalog protection of the objects on the source media. When there are no more protected objects on the media, the media can be overwritten.

Figure 16. Consolidation options



Important
If you recycle data protection of source objects, the recycled points in time will no longer be available. Unless copies of these points in time exist, you will be able to restore only to the latest (consolidated) point in time.

In addition a different protection period than what the regular full can be specified for the consolidated object.

Specifying a relatively short protection period for full and incremental backups frees up disk space as used media for those backup types will be released by HP Data Protector software.

Enhanced Incremental Backup

With conventional incremental backup, the main criterion for determining whether a file has changed or not since a previous backup is the file's modification time. However, there are cases where this criterion is not effective. For example, if a file has been renamed or moved to a new location, or if some of its attributes have changed, its modification time does not change. Consequently, the file is not always backed up in an incremental backup. Such files are backed up in the next full backup.

Enhanced incremental backup reliably detects and also backs up renamed and moved files, as well as files with changes to their attributes.

Detection of certain changes (such as changes of permissions or ACLs) also depends on the setting of the following options in the backup specification. The preferred setting enables maximal detection of changes with enhanced incremental backup.

- **Windows:** Do not use `archive` attribute.
  The default is `OFF` (`archive` attribute is used). This is the preferred setting.
- **UNIX:** Do not preserve `access time` attributes.
  The default is `OFF` (`access time` attributes are preserved). The preferred setting is `ON`.

Enhanced incremental backup also eliminates unnecessary full backups of entire backup objects while adding new trees to a backup object. For example, if an additional directory is selected for backup since the last backup, a full backup of this directory (tree) is performed, whereas the backup of the rest is incremental.

The first full backup with an enabled enhanced incremental backup creates a small database on each client to be backed up. For each directory to be backed up, two files are created in that database. Both contain a hash-key, one for the directory and the other for all the files of this directory. Additionally a timestamp is stored.

The hash-key contains the properties of the file. This information allows HP Data Protector software to detect nearly all changes to files and thus to be able to add all modified files into an incremental backup. This is done by comparing the stored hash-key with the current hash-key that is generated whenever an incremental backup is done.

Note that the hash-key does not contain the ACL information. Therefore, the enhanced incremental backup detects changes on file permissions not through the hash-key and must still rely on OS flags, like attribute flag on Windows (see also limitation of enhanced incremental backup).

The database contains binary files that are stored within a directory structure. Under <`data protector home`>`/enhincrdb` directory, a new directory is created (see table below) for each mountpoint. On UNIX each directory gets a "_" (underscore) as first letter. This means that the directory representing the root directory is named "_". See the table below with some more examples.

Under each of those directories, 256 subdirectories are created named 00 through FF (hex). Those directories store the hash-keys files that are distributed according to an algorithm, which ensures an even distribution.

Table 1. Enhanced incremental db naming example

Windows	Type	Name of the directory
	Drive letter example: "F:\"	As the drive letter example: "F"
	Mounted directory "C:\mnt\disk1"	Each directory path separated with underline "C_mnt_disk1"
Unix	root directory	"_"
	Example: /home /home/mntpoint	 "_home" "_home_mntpoint"

Detection

Depending on the platform, enhance incremental backup mode is able to detect the following potential changes on a file system since the last related backup:

- **All supported platforms:**
  - Changes to file contents
  - Renamed files/directories
  - Moved files/directories
- **Posix-only platforms (different UNIX flavors that support "stat" system function):**
  - Any change to file/directory permissions, owner id, or owner group id if "Do not preserve access time attributes" option is turned off (default)
  - Any change to file attributes if "Do not preserve access time attributes" option is turned on.
- **Windows:**
  - Any change to permission list
  - Any change to ownership
  - Any change to attributes hidden, read only, system, encryption, and compression
- **OpenVMS:**
  - ACL list
  - access level
  - backlink ID
  - blocks allocated
  - caching options
  - directory file ID
  - PATHWORKS MS-DOS flags
  - file ID
  - file characteristic (stream file, record file, text file, and so on)
  - file protection
  - journal control flag
  - attributes modified date
  - creation date
  - expiration date
  - revision date

- file owner
- record attribute
- extended record attributes
- record protection
- file version limit

Why use enhanced incremental backup mode
Use enhanced incremental backup mode to:

- Ensure incremental backup of files with changes in name, location, or attributes.
- Eliminate unnecessary full backups if some of the selected trees change.
- Enable subsequent object consolidation (synthetic full backup).

Impact on disk space consumption
Enhanced incremental backup uses a small database on each client that is backed up. The impact on disk space on the client is typically less than 1% of the size of the files selected for backup.

Impact to backup duration
In most of the cases, the impact by setting the enhanced incremental mode is very small and therefore, negligible.

As the HP Data Protector software cell manager has to maintain a hash database on each client to be backed up, the performance impact to full/incremental backups can result in a 1%–5% performance decrease based on a 10% change to the data backed up.

The automatic or scheduled object consolidation is a separate process and does not influence backup runs as the consolidation does not run necessarily on the backup client.

Impact to backup client main memory consumption
Enhanced incremental backup produces higher memory requirements on a client system. Typically, all filenames of all directories in a path being processed are kept in memory. This means that if the Disk Agent (DA) is backing up the directory `tree/` and works currently at `/home/user_a/testdir`, all filenames of `home`, `user_a`, and `testdir` are kept in memory. This is due to the fact that the Disk Agent (DA) first performs an in-depth treewalk and must maintain the next item available after returning from lower level.

CPU and I/O requirements are not affected significantly.

Limitations
Enhanced incremental backup is not supported on Novell NetWare and MPE/iX systems.

With HP Data Protector software 6.0, incremental ZDB is now supported. However, it is not possible to perform enhanced incremental ZDB.

Enhanced incremental backup of clustered systems is supported, but requires manual configuration steps.

Detection method relies on system time off the client.

With the `Enhanced incremental backup` option enabled, hardlinks are not supported anymore. This means that all hardlinks are backed up as normal files.

How HP Data Protector software handles hardlinks in general is defined through two options:

- Backup POSIX hard links as files (default not selected)
- Detect NTFS hardlinks (default not selected)

This means that, by default, HP Data Protector software backs up hardlinks on NTFS as normal files and POSIX hard links as links.

Using enhanced incremental backup can now cause a conflict in case hardlinks should be backed up as links. In such a case the following warning is listed inside the session output:

```
"[Warning] From: BSM@system1 Time: 8.5.2006 15:01:12
[61:20005] Hardlink detection is not compatible with enhanced backups.
Hardlinks for object 'bbn.deu.hp:/' will be backed up as files."
```

To suppress this warning in the future, modify the backup specification so that it will not contain hardlink detection options.

- For Windows Filesystem objects, uncheck the "Detect NTFS hardlinks" option.
- For other Filesystem objects, check the "Backup POSIX hard links as files" option.

Please note that a backup of the enhanced incremental database is not recommended since, in case the database is restored, it is probably out of sync and can not be used anymore. If HP Data Protector software detects entries inside the enhanced incremental database that are older than the last incremental backup, an enhanced incremental backup cannot use the local database anymore, and therefore, HP Data Protector software switches for all out of sync directories to backup type full. This backup type recreates the hash-keys.

The session contains the following warnings:

```
[Warning] From: VBDA@tpc003.dp3.com "C:" Time: 5/5/2006 10:50:08 AM
[81:1421]   Enhanced incremental database information is out of date for
directory C:\hr\fs_test.
[Warning] From: VBDA@tpc003.dp3.com "C:" Time: 5/5/2006 10:50:08 AM
[81:1406]   One or more directories were backed up in full mode to
prevent potential data loss.
```

Note the same behavior happens in the following cases:

1. While performing a full and several incremental backups, all with enhanced incremental enabled.
2. While deleting a directory from disk and perform an incremental backup again.
3. While adding the same directory again and performing an incremental. The added directory is detected by HP Data Protector software as out of date and a full backup of this directory tree is performed.

Note that a restore of the enhanced incremental database is only useful if all of the corresponding data (all mount points) are being restored, for example in case of a disaster recovery.

Enabling Enhanced Incremental Backup

This procedure describes how to enable enhanced incremental backup in an existing backup specification, but you can also enable it while creating a new backup specification.

To enable enhanced incremental backup of shared mount points in a cluster environment, additional steps are required.

Steps

1. In the Context List, click **Backup**.
2. In the Scoping Pane, expand **Backup Specifications** and then expand **Filesystem**. All saved backup specifications are displayed.
3. Click on the backup specification that you want to modify.
4. Click the **Options** tab. Under **Filesystem** Options, click **Advanced**.
5. In the Filesystem Options dialog box, click the **Other** tab. Select Enhanced incremental backup and click **OK**.
6. Click **Apply**.

Tip

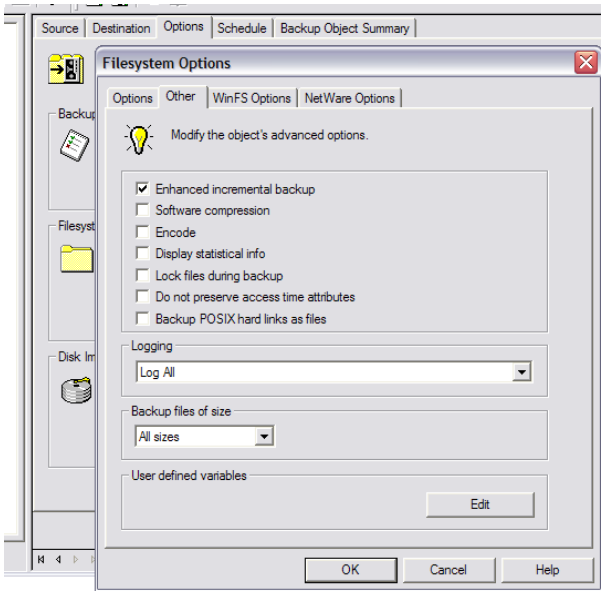
You can set this option for individual objects. In the Backup Object Summary tab, select the desired object, click Properties, click the Other tab, and select Enhanced incremental backup.

Note

Incremental backup will run in the enhanced mode only after a full backup is performed.

To start running enhanced backups with the next backup, although an incremental backup is scheduled next, set the `EnhIncrPromoteToFullIfNoEnhFull` global variable to `1`, which will promote the next backup to a full backup.

Figure 17. Enhanced incremental backup mode



Enabling Enhanced Incremental Backup in a Cluster Environment

For backing up shared volumes inside a cluster environment, a volume can be backed up from any (physical) cluster system. This requires that the incremental local database for each shared volume is always available on that physical system where it is currently mounted. This can not be guaranteed in case the enhanced incremental database is stored inside the HP Data Protector software home directory.

In order to solve this issue, the enhanced incremental database is moved to the shared volume and appropriate links pointing to this shared volume are created on all cluster nodes.

The following steps have to be performed before the first enhanced incremental backup is performed:

Create a repository directory and links as described below:

Microsoft Cluster Server

1. On the shared mount point that will be backed up, create the following directory:

```
mkdir <drive letter of the shared volume>:\enhincrdb

mkdir <drive letter of the shared volume >:\enhincrdb\<drive letter
of the shared volume >

Example: drive letter = F

F:\enhincrdb

F:\enhincrdb\F
```

2. On each node, create a link to the enhanced incremental database by running:

```
linkd

 “C:\Program Files\OmniBack\enhincrdb\< drive letter of the shared
volume>”

 “<drive letter of the shared volume >:\enhincrrep\< drive letter
of the shared volume >

Example: drive letter = F

linkd “C:\Program Files\OmniBack\enhincrdb\F” “F:\enhincrdb\F”
```

The `linkd.exe` utility is available in the Windows resource kit.

Now you can perform backups of data on this mount point in the enhanced mode.

UNIX

1. On the shared mount point that will be backed up, create the following directory:

```
mkdir /<name of the shared mount point>/enhincrdb

mkdir /<name of the shared mount point>/enhincrdb/_<name of the
shared mount point>

Example: name of the shared mount point = shared_disk

mkdir /shared_disk /enhincrdb

mkdir /shared_disk /enhincrdb/_shared_disk
```

2. On each node, create a link to the enhanced incremental database by running:

```
ln –s /<name of the shared mount point>/enhincrdb/_<name of the
shared mount point> /var/opt/omni/enhincrdb/_<name of the shared
mount point>

Example: name of the shared mount point = shared_disk

ln -s /shared_disk/enhincrdb/_shared_disk
/var/opt/omni/enhincrdb/_shared_disk
```

Now you can perform backups of data on this mount point in the enhanced mode.

Distributed File Media Format (DFMF)

Unique in the market, virtual full backup is an even more efficient type of synthetic backup. This solution uses pointers to consolidate data rather than copy the data. As a result, the consolidation takes less time and avoids unnecessary duplication of data.

Virtual full backup is performed in the same way as normal synthetic backup with the following three requirements. In an object consolidation session, HP Data Protector software automatically performs a virtual full backup if:

- All backups, the full backup, incremental backups, and the resulting virtual full backup are written to one File Library
- The File Library uses distributed file media format (DFMF)
- All backups (full + incremental) are configured to use the "enhanced incremental backup" option

With HP Data Protector software 6.0, a new media format named distributed file media format (DFMF) was introduced. This format can only be used with the HP Data Protector software File Library and is, by default, not enabled.

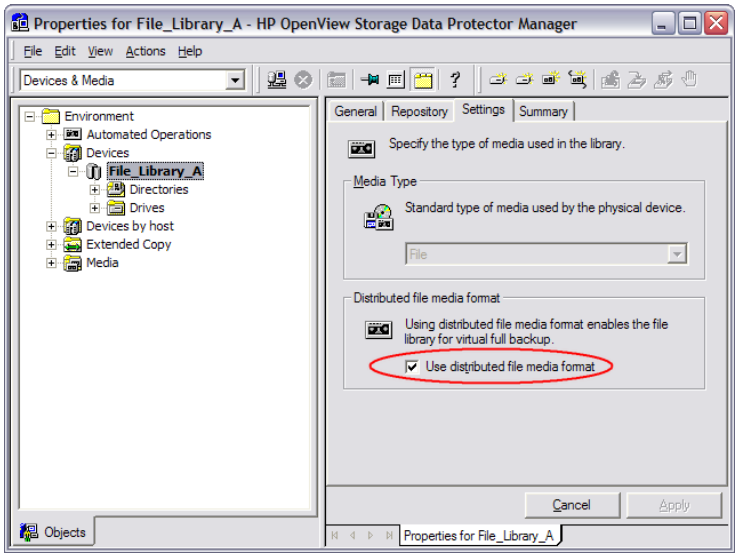
With the new media format, pure data blocks are written into different files. This is done for each file, larger than the used block size (default 64 KB). Hence, for each backed up file, a dedicated file on the File Library that holds the data blocks is created.

Without this format, HP Data Protector software media is represented by one file in a file library. This is done per session; hence, each session creates its own file(s).

If a consolidation session is performed on backups that are all located in the same file library, the data that will be consolidated is already stored in one or more media files. The new DFMF concept will reuse those files. Hence, instead of copying the data blocks, they are only referred by way of pointers. Therefore, consolidation sessions creating virtual full backups do not copy the files hosting the data blocks. Instead, the new session only refers to them by way of pointers. Note that only consolidation sessions use pointers; normal backups, both full and incremental, always create new data block files.

To enable the DFMF format for a HP Data Protector software file library, the relevant file library configuration setting needs to be changed. Figure 18 depicts the device property settings.

Figure 18. Defining a file device library with DFMF enabled



Synthetic Backup and Media Space Consumption

Frequent synthetic backups typically mean significant space consumption on the backup media.

However, if virtual full backup is performed, the backup media space consumption is reduced.

With virtual full backup, the space consumption largely depends on the size of the backed up files. If the files are significantly larger than the block size used, virtual full backup achieves maximum savings of the space compared to normal synthetic backup. On the other hand, if the files are smaller than the block size, the savings are rather small. As a rule of thumb, one can calculate the additional space needed for a virtual full like:

(Total size of all files that are smaller than DP block size) + (DP block size * number of files larger than DP block size)

OMNIRC variables

The omnirc variables are useful for troubleshooting or overwriting other settings affecting the behavior of the Data Protector client. However, use them only if your operating environment demands it. The Disk Agents and Media Agents use the values of these options.

The omnirc variables can be also be set on each client only in the file:

Windows: <Data_Protector_home>\omnirc

UNIX: /opt/omni/.omnirc

Novell NetWare: sys:\usr\omni\omnirc

Object Consolidation Cache parameters

When an object consolidation session starts, the Restore Media Agent (RMA) reads all source sessions required for the consolidation. This always includes one full backup (can also be a synthetic or virtual full) and several (at least one) incremental backup session. During this step, HP Data Protector software tracks all segments (or records in the segments) that are required for the object consolidation. This information is stored in memory, and if required in temporary files.

These files are stored in the HP Data Protector software temp directory (UNIX: /tmp; Windows <DP home>\tmp).

Note that for each directory level, one temporary file is created. The location of temporary files can be redirected with the OB2CONSOLIDATIONCACHEPATH omnirc variable.

The size of those files is rather small; a consolidation test with 5.5 million files has shown a size of 500 MB.

Each file takes: 52 bytes + [filename length + 1] * 2 bytes.

The temporary files are created when cache page size on directory level is exceeded (4K by default). The maximum cache page size can be changed with the OB2CONSOLIDATIONCACHEPAGESIZE omnirc variable.

The filename starts with fts_dirlevel and ends with some numbers.

Object Consolidation file exception handling

There can be number of reasons why a file was not backed up while it was present in the object directory list. From backup session point of view, the file content can be backed up in whole, partly (file ends with error record), or not backed up at all. From the consolidation point of view, the following cases need to be handled:

- The file content was not backed up in any session.
- The file content was partly backed up in some sessions and not backed up in other sessions.
- The file content was partly backed up in some (one or many) sessions and backed up in whole in some sessions.
- The file was backed up in whole in some sessions.

The last case is the expected behavior when backup sessions succeed, while the others have to be handled as exceptions.

The consolidation process can run in two different modes to handle the above mentioned exceptions.

If the OB2CONSOLIDATIONPRESERVECOMPATIBILITY omnirc variable is set to 1, the compatibility with a normal restore operation is preserved. That means that the object restored from restore chain is exactly the same as if it would be restored from a synthetic object build from the same restore chain. If the compatibility mode is disabled, the synthetic object holds more useful information as it can be retrieved from restore chain as whole. This is also default behavior.

If restore compatibility mode is disabled, the RMA tries to use the latest complete copy of a file. This means that any non-complete versions of a file are overwritten by older full versions of a file. If the compatibility mode is enabled, the latest version of a file is used no matter if the version is complete or not.

The default behavior is = compatibility mode disabled (0)

Distributed File Media Format (DFMF) Parameters

DFMF uses the <medium id>.fd HP Data Protector software medium file that is used to host all the catalog data and the first data block. All further data blocks are written into a separate file named Fxxxx, where xxxx is a sequential number in hex starting from 0000 to FFFF. Those files are grouped inside directory named Dxxxx where xxxx is used in the same manner. Note that each Fxxxx file hosts data only from a certain file.

The Dxxxx directories are grouped together in a directory called "repository" and is named rp_<medium id>. This directory is created for the first backup (medium id refers to that medium which is used by that backup) and is used by all further sessions. Another repository directory is created only in case the currently-used directory is full (the max. number of Dxxxx and Fxxxx is reached). If the repository is full during a backup, the medium is reported as full and a new medium is created.

The maximum number of Dxxxx directories per repository is 16384 with 16384 Fxxxx files per directory. This ends up with the max number of 4,294,967,296 Fxxxx files per repository. This is the maximum permitted configuration. The repository can be configured to hold less directories/files with OB2DFMFDIRECTORIES and OB2DFMFFILES omnirc variables. Note that omnirc variables are used only during creation of the repository. These values are than written into repository map file (.rm) and used afterwards.

Distributed File Media Format data verification

The analysis phase of a consolidation session does not need the information stored in offloaded files within the DFMF repository. Offloaded file are files holding pure data blocks only. By setting the OB2DFMFREADOFFLOAD variable to 1, the reading of the whole DFMF repository (including offloaded files) will be enforced. This way the operation will take more time, but will verify the medium during the session. By default (set to 0), the offloaded files will not be read.

Object Consolidation Performance Aspects

Object consolidation and Object copy

During an object consolidation or object copy session, many of the same processes run as during a backup or a restore session. Basically, data is read from source media as if it was restored and written to target media as if it was backed up. An object consolidation or object copy session has the same effect on the IDB operation and hardware utilization as backup and restore.

The performance implications of a backup and restore process are described below.

Backup

When a backup session is started, a session record is created in the IDB. Also, for each object in the session, an object version record is created. Both records are stored in the CDB (catalog db) part and have several attributes. The Backup Session Manager updates media during a backup. All media records are stored in the MMDB (media management db) part and are allocated for a backup depending on policies.

When a data segment (and a catalog segment after it) is written on the tape, a media position record is stored in the CDB for each object version that was part of this data segment. In addition, the catalog is stored in the DC (Detail Catalog) binary file. One DC binary file is maintained per Data Protector medium. The DC binary file is named `<MediumID>_<TimeStamp>.dat`. The name is not changed when backups append to the same medium. If a medium is overwritten during a backup, its old DC binary file is removed and a new DC binary file is created.

All session messages generated during backups are stored in session messages binary files (the SMBF part).

Restore

When configuring restore, Data Protector performs a set of queries in the CDB and DCBF parts to enable users to browse virtual file systems of backed up data. These browse queries are done in two steps. The first step is to select a specific object (filesystem or logical drive). If this object has many backup versions stored, this can take some time because Data Protector scans the DCBF to build a lookup cache for later browsing. The second step is browsing directories.

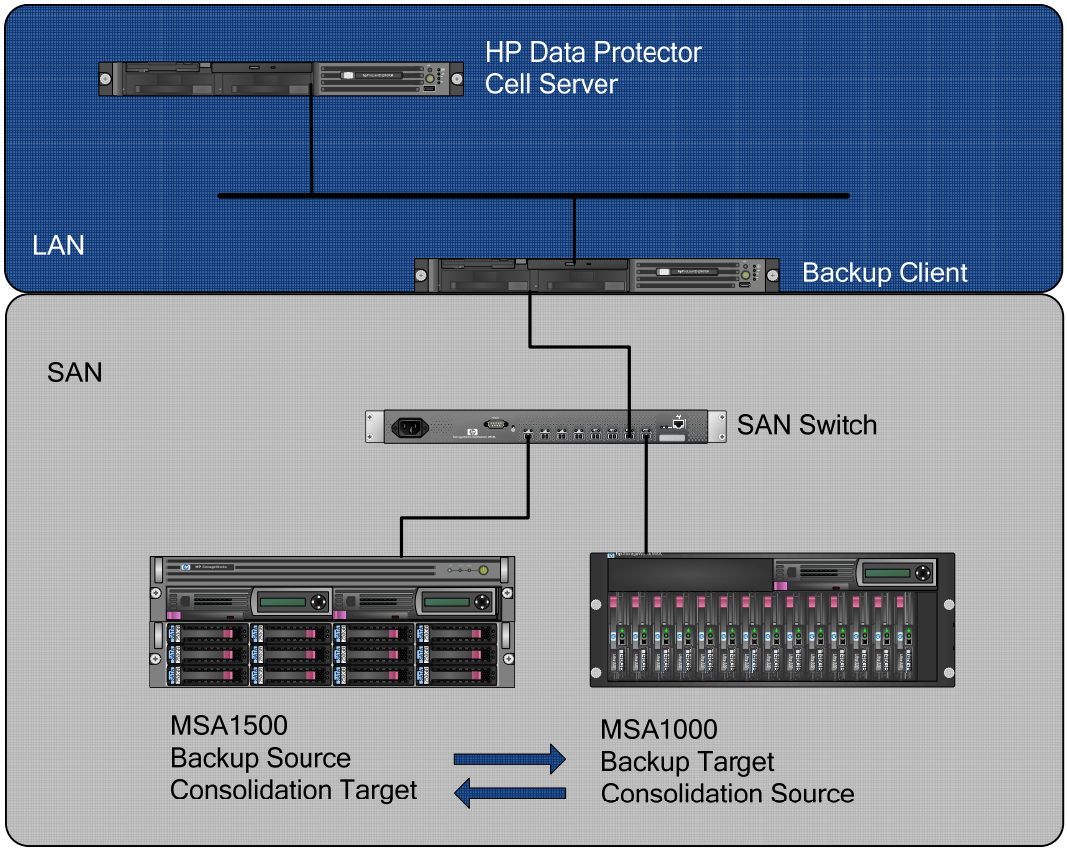
After specific versions of files are selected, Data Protector determines the required media and locates media position records used by the selected files. These media are read by the Media Agents and data is sent to the Disk Agents that restore the selected files.

Performance Test Case

In order to demonstrate the performance of object consolidation, the environment depicted in Figure 19 was setup.

The HP Data Protector software cell server and the backup client are two HP ProLiant DL380 hosts running with one Intel Xeon 3.2GH CPU each; both clients are connected through Ethernet LAN. Each client is equipped with 1GB main memory. The backup client on the other side is connected to a SAN switch to which a HP MSA1500 and a HP MSA1000 are connected to.

Figure 19. Consolidation Test setup

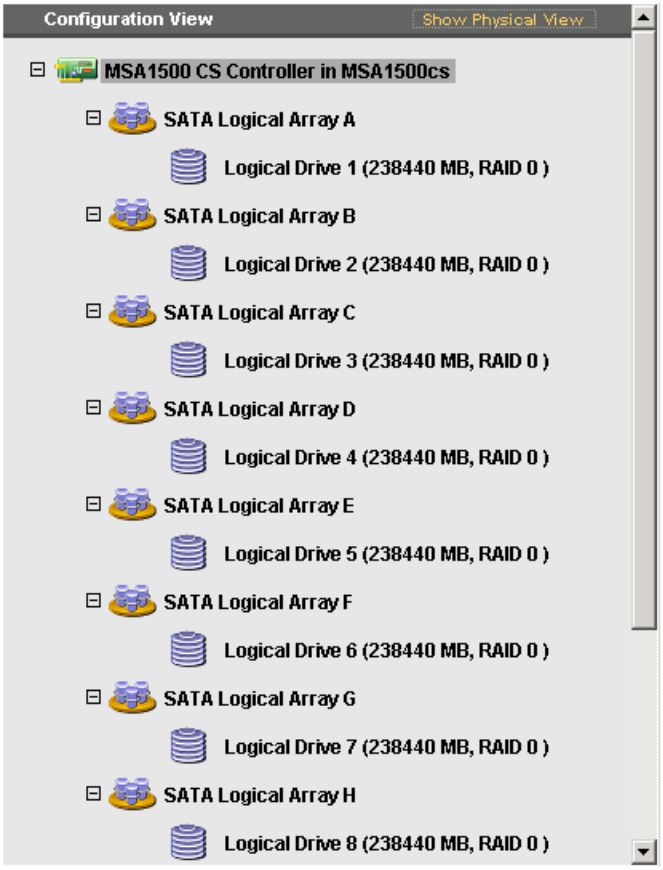


The MSA1000 will also be consolidation target for Virtual Full Backups.

The performance test setup should analyze the performance impact of various Data Protector software configuration settings rather absolute performance numbers.

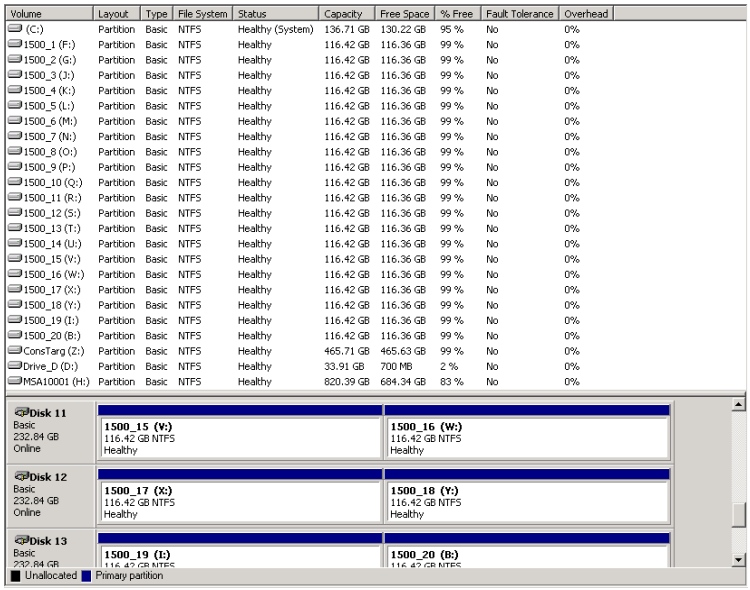
The MSA1500 has 10 LUN configured each with a total capacity of approximately 250 GB SATA disk drives.

Figure 20. MSA1500 capacity



The MSA1500 disks are presented to the backup client on which two primary partitions per LUN have been created. Each partition can hold up to approximately 116 GB of data. The MSA1500 acts as a backup data source in the first phase of the test and as a destination for consolidated backups in the second phase of the test.

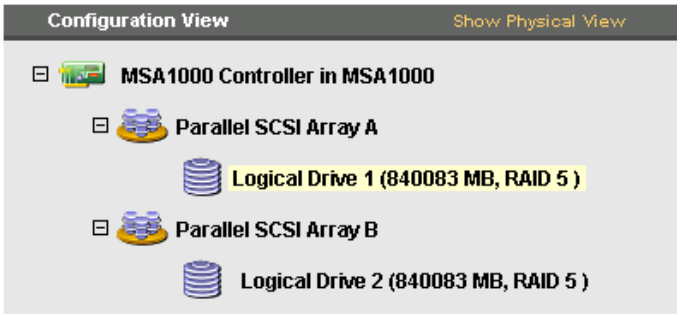
Figure 21. Backup client disk configuration



A total of twenty objects represented by disk drives can now be processed by HP Data Protector software. In addition, the MSA1500 acts as a destination target during the second phase—object consolidation—of the test.

The MSA1000 has two LUN configured each having a total capacity of approximately 840 GB SCSI disk drives. One LUN represents the backup destination device.

Figure 22. MSA1000 capacity



The MSA1000 works in RAID5 mode.

The MSA1000 is also presented to the backup client and acts as a backup destination device in the first phase of the test and as a source for backup consolidation in the second phase of the test. The MSA1000 will also act as a DFMF formatted library for Virtual Full Backups.

This setup was chosen to not have concurrent read/write operations on one modular smart array during the test. This setup also ensures that DA, MA, RMA, and BMA will operate on one single system during the test.

This test case evaluates the following cases:

- Backups were created using one writer with concurrency set to twenty (20)
- Backups were created using twenty writers with concurrency set to one (1)
- Consolidation was done using one writer with concurrency set to twenty (20)
- Consolidation was done using twenty writers with concurrency set to one (1)

The number of read devices for the consolidation process was in all cases set to twenty (20) to allow parallel processing of the data.

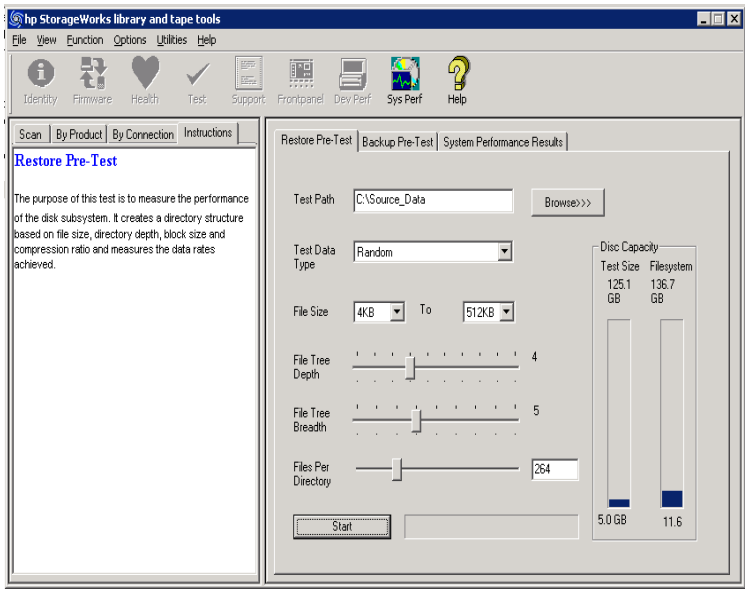
Performance-tuning recommendations given for Modular Smart Arrays in the "HP Data Protector software Advanced Backup to Disk performance white paper" found at http://www.hp.com/go/dataprotector were followed in the discussed setup in the sense that the maximum size of file depots was set to 10 GB.

Data generation

The test data to be backed up was created using HP StorageWorks Library and Tape Tools (L & TT).

HP StorageWorks Library and Tape Tools is a free download from the web and can be found at: http://h18006.www1.hp.com/products/storageworks/ltt/index.html?jumpid=reg_R1002_USEN

Figure 23. Library and Tape Tool User Interface



Each of the twenty directories (objects) was filled with approximately 1 GB of data. The file size was between 4 KB and 512 KB. The total number of files per object was 8432 and spread over 30 folders.

Figure 24. Object Capacity

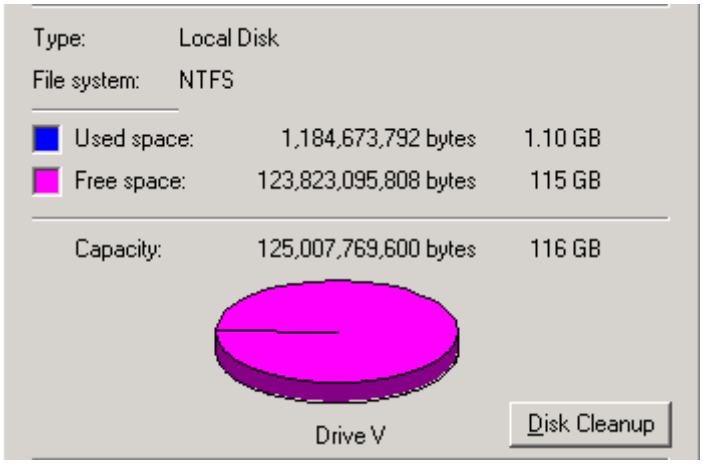
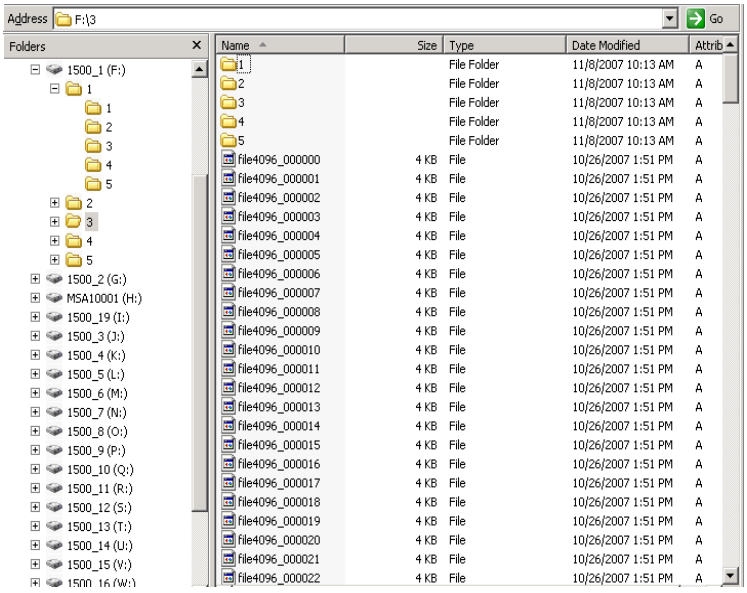
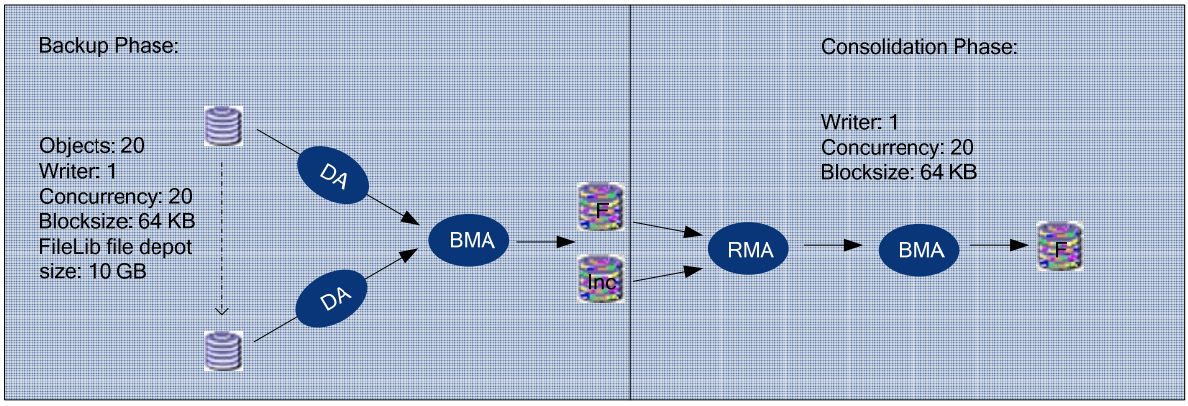


Figure 25. Object Layout



Performance Test Results

Case 1:



In this test case, one initial full backup and one incremental backup was performed. The data change rate from full backup to incremental backup was 3.5%.

During the backup phase, only one writer was used to write into the HP Data Protector file library. The concurrency for that writer was set to twenty (20) meaning twenty Disk Agents (DA) writing simultaneously to that file library creating a highly multiplexed data stream into the file library. In total, four file depots were written. During the consolidation phase, one writer with concurrency set to twenty was used to write the synthetic full backup. In total, three file depots were written. For the restore, an object which was close to the end of the consolidation process was chosen.

Synthetic full backup mode

	20 objects Full Backup	20 objects Incremental Backup	Consolidation	1object Restore	Read Data Kbytes/ Restore	Used Media to restore
Elapsed time	0:15:30	0:03:01	2:08:02	0:12:31	21745600	3
CPU utilization	~60%	~50%	~6%	~10%		
I/O utilization	↑	↑	→	↓		
Memory utilization	→	→	→	↓		

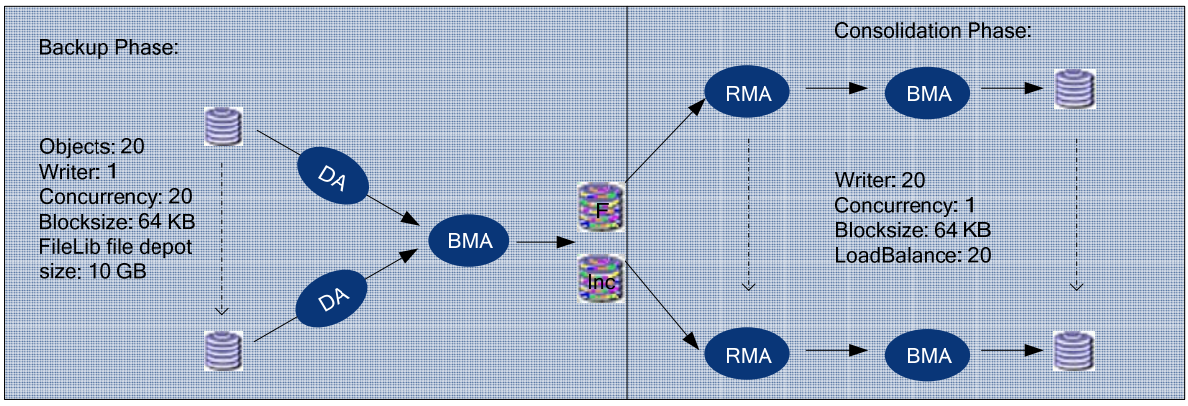
Virtual full backup mode

	20 objects Full Backup	20 objects Incremental Backup	Consolidation	1object Restore	Read Data Kbytes/ Restore	Used Media to restore
Elapsed time	0:36:08	0:03:17	0:17:34	0:05:52	7307584	1
CPU utilization	~60%	~50%	~10%	~10%		
I/O utilization	→	↓	↓	↓		
Memory utilization	→	↓	→	↓		

Legend of test equipment resource utilization icon during respective operation.

↓ Low → Medium ↑ High

Case 2:



In this test case, one initial full backup and one incremental backup was performed. The data change rate from full backup to incremental backup was 3.5%.

During the backup phase, only one writer was used to write into the HP Data Protector file library. The concurrency for that writer was set to twenty (20) meaning twenty Disk Agents (DA) writing simultaneously to that file library creating a highly multiplexed data stream into the file library. In total, four file depots were written. During the consolidation phase, twenty writers with concurrency set to one were used to write the synthetic full backup. One file depot per object was created. For the restore, an object which was close to the end of the consolidation process was chosen.

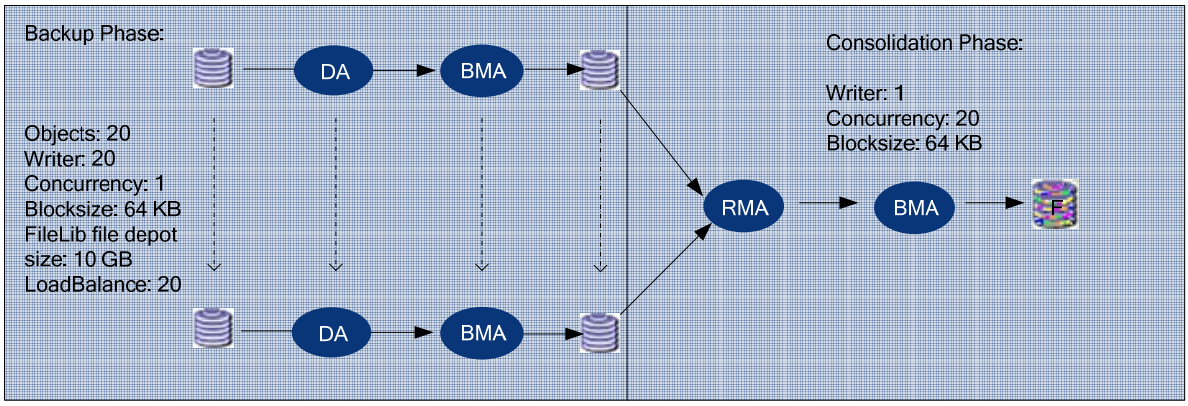
Synthetic full backup mode

	20 objects Full Backup	20 objects Incremental Backup	Consolidation	1object Restore	Read Data Kbytes/ Restore	Used Media to restore
Elapsed time	0:15:22	0:03:05	1:37:01	0:03:57	1089664	1
CPU utilization	~60%	~40%	~15%	~13%		
I/O utilization	↑	↑	→	↓		
Memory utilization	→	→	→	↓		

Virtual full backup mode

	20 objects Full Backup	20 objects Incremental Backup	Consolidation	1object Restore	Read Data Kbytes/ Restore	Used Media to restore
Elapsed time	0:36:08	0:03:11	0:24:25	0:04:00	1089664	1
CPU utilization	~60%	~55%	~15%	~10%		
I/O utilization	↓	↓	↓	↓		
Memory utilization	→	↓	→	↓		

Case 3:



In this test case, one initial full backup and one incremental backup was performed. The data change rate from full backup to incremental backup was 3.5%.

During the backup phase, twenty writers were used to write into the HP Data Protector file library. The concurrency for the writers was set to one (1) per writer meaning twenty Disk Agents (DA) writing simultaneously through twenty Media Agents(MA) to that file library creating one file depot per object. In total, twenty file depots were written. During the consolidation phase, one writer with concurrency set to twenty was used to write the synthetic full backup. Three file depots were created. For the restore, an object which was close to the end of the consolidation process was chosen.

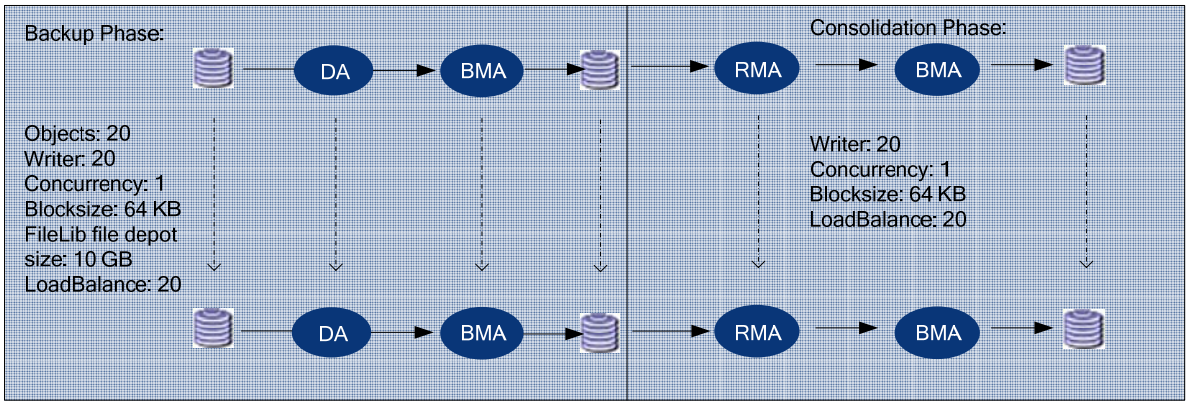
Synthetic full backup mode

	20 objects Full Backup	20 objects Incremental Backup	Consolidation	1object Restore	Read Data Kbytes/ Restore	Used Media to restore
Elapsed time	0:14:06	0:04:46	0:18:48	0:12:13	21745600	3
CPU utilization	~70%	~20%	~35%	~10%		
I/O utilization	↑	→	↑	↓		
Memory utilization	→	→	→	↓		

Virtual full backup mode

	20 objects Full Backup	20 objects Incremental Backup	Consolidation	1object Restore	Read Data Kbytes/ Restore	Used Media to restore
Elapsed time	0:37:52	0:04:48	0:13:49	0:05:01	7484288	1
CPU utilization	~65%	~40%	~25%	~10%		
I/O utilization	↓	↓	→	↓		
Memory utilization	→	→	→	↓		

Case 4:



In this test case one initial full backup and one incremental backup was performed. The data change rate from full backup to incremental backup was 3.5%.

During the backup phase, twenty writers were used to write into the HP Data Protector file library. The concurrency for the writers was set to one (1) per writer meaning twenty Disk Agents (DA) writing simultaneously through twenty Media Agents(MA) to that file library creating one file depot per object. In total, twenty file depots were written. During the consolidation phase, twenty writers with concurrency set to one were used to write the synthetic full backup. One file depot per object was created. For the restore, an object which was close to the end of the consolidation process was chosen.

Synthetic full backup mode

	20 objects Full Backup	20 objects Incremental Backup	Consolidation	1object Restore	Read Data Kbytes/ Restore	Used Media to restore
Elapsed time	0:14:19	0:04:47	0:20:57	0:04:03	1089664	1
CPU utilization	~70%	~35%	~25%	~10%		
I/O utilization	↑	→	↑	↓		
Memory utilization	→	→	↑	↓		

Virtual full backup mode

	20 objects Full Backup	20 objects Incremental Backup	Consolidation	1object Restore	Read Data Kbytes/ Restore	Used Media to restore
Elapsed time	0:37:24	0:04:51	0:17:29	0:03:58	1089664	1
CPU utilization	~40%	~30%	~15%	~10%		
I/O utilization	↓	↓	→	↓		
Memory utilization	→	→	↑	↓		

Performance Test Observations

The backup/restore timing and system resource utilization behavior observed in this test scenario comes close to what is documented in the "HP Data Protector software performance white paper" found at http://www.hp.com/go/dataprotector.

As already mentioned in this document, test cases as described in the previous section have been evaluated. The HP Data Protector software configuration used during the tests might not be optimal for the most customer environments. However, it highlights where to make modifications in the configuration of HP Data Protector software to get the best performance where Synthetic Full or Virtual Full backup is a topic.

In all four test cases, the elapsed time for the synthetic full backup mode scenario was quite similar. It was between 14 minutes and 16 minutes. The two cases where the elapsed time was closer to 16 minutes were caused when only one backup writer was used. The writer concurrency was twenty meaning twenty Disk Agents writing simultaneously to the file depot. This setup leads to a highly multiplexed file depot on the disk subsystem. I/O utilization was very similar in all four cases whereas CPU utilization showed to be slightly higher when the writer configuration was set to twenty writers with concurrency one.

Full backups of this type are only relevant initially. They are then no longer needed as Synthetic Full or Virtual Full will be the preferred backup method.

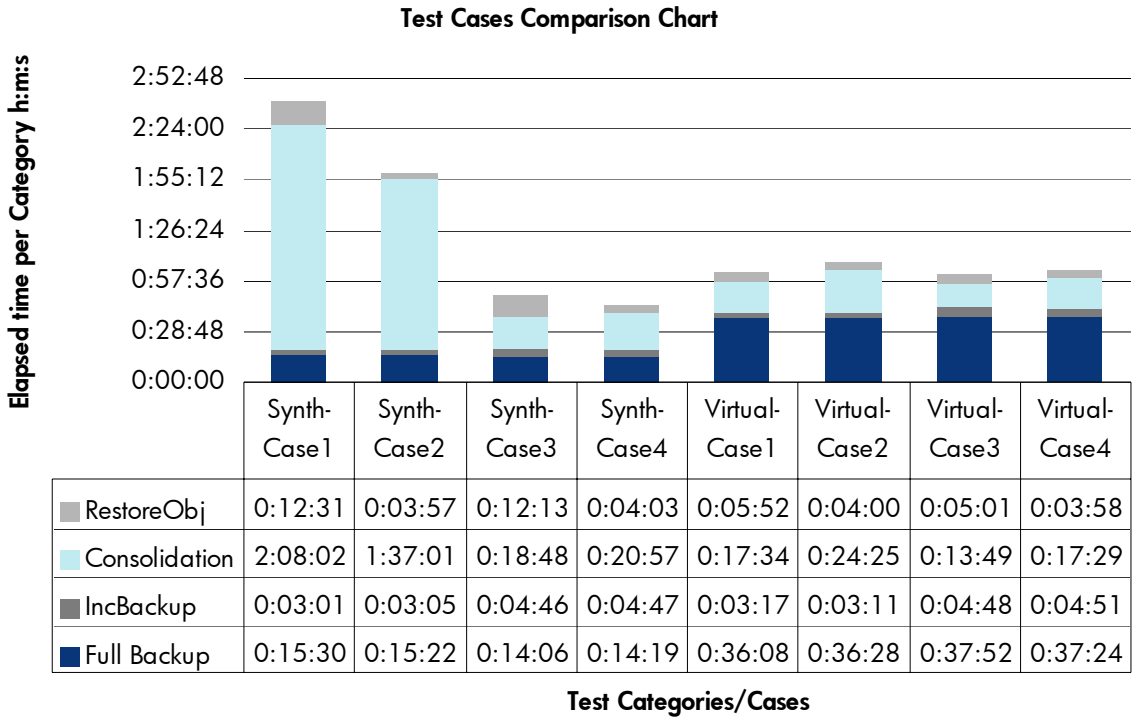
The elapsed time for incremental backups showed to be faster where a high concurrency per writer was used. The slower backup performance in the low concurrency setting can be related to a higher number of writer processes running on the backup client as well as simultaneous write attempts in the disk array. Memory utilization and disk array performance have had an impact to the incremental backup performance. For the consolidation process it was advantageous to use backup media with non multiplexed objects on it. Significant time savings up to 80% were observed for the consolidation process. The restore process of an object showed to be clearly faster when the consolidation process had used twenty writers with concurrency set to one per consolidated object. The restore of one object required only one media whereas up to three media had to be read where one writer with a concurrency setting of twenty has created the multiplexed synthetic full backup.

All four test cases were performed a second time in which a DFMF formatted disk library was used.

The elapsed time for the initial full backups on the DFMF library was more or less the same in all four DFMF test cases. However, it ran for almost double the time as it ran in the synthetic full backup mode scenario in all four scenarios. The performance of the disk sub system mainly influenced the elapsed time because of concurrent read/write operations within the disk sub system. Incremental backups were faster were one writer with concurrency set to twenty was used but there was no significant difference compared to the synthetic full backup mode scenario.

Consolidation was definitely faster on the DFMF-formatted library as the advantage of the Distributed File Media Format kicked in here. Since the selected standard block size in the DFMF library is 64 KB, 5100 files out of 8432 per object having file sizes between 4 KB and 64 KB were copied during the consolidation process. The space saving was not significant. However, there was no big advantage visible compared to case three and four of the synthetic full backup mode scenario. Demultiplexed media helped to keep the elapsed time for the consolidation process low.

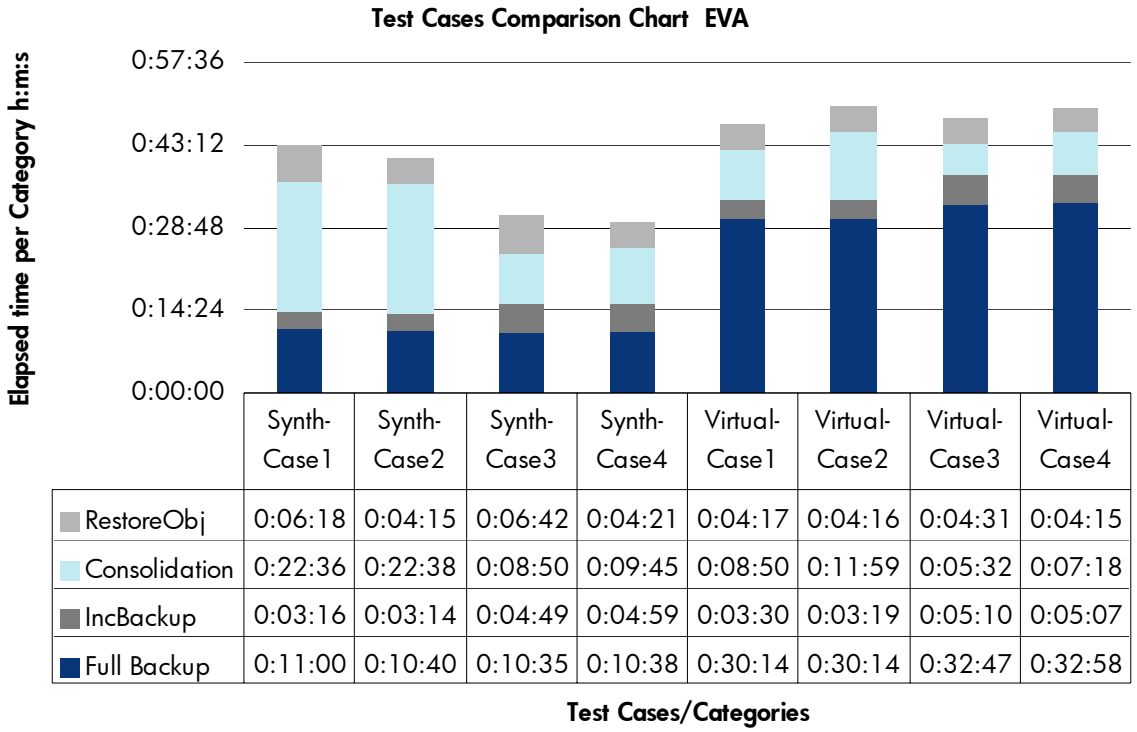
Figure 26. Performance chart measured on MSA hardware



	Synth-Case1	Synth-Case2	Synth-Case3	Synth-Case4	Virtual-Case1	Virtual-Case2	Virtual-Case3	Virtual-Case4
RestoreObj	0:12:31	0:03:57	0:12:13	0:04:03	0:05:52	0:04:00	0:05:01	0:03:58
Consolidation	2:08:02	1:37:01	0:18:48	0:20:57	0:17:34	0:24:25	0:13:49	0:17:29
IncBackup	0:03:01	0:03:05	0:04:46	0:04:47	0:03:17	0:03:11	0:04:48	0:04:51
Full Backup	0:15:30	0:15:22	0:14:06	0:14:19	0:36:08	0:36:28	0:37:52	0:37:24

Additional tests conducted on EVA (enterprise virtual array) hardware using the previously mentioned scenarios showed similar performance characteristics.

Figure 27. Performance chart measured on EVA hardware



	Synth-Case1	Synth-Case2	Synth-Case3	Synth-Case4	Virtual-Case1	Virtual-Case2	Virtual-Case3	Virtual-Case4
RestoreObj	0:06:18	0:04:15	0:06:42	0:04:21	0:04:17	0:04:16	0:04:31	0:04:15
Consolidation	0:22:36	0:22:38	0:08:50	0:09:45	0:08:50	0:11:59	0:05:32	0:07:18
IncBackup	0:03:16	0:03:14	0:04:49	0:04:59	0:03:30	0:03:19	0:05:10	0:05:07
Full Backup	0:11:00	0:10:40	0:10:35	0:10:38	0:30:14	0:30:14	0:32:47	0:32:58

Conclusions

Backup and restore performance heavily depends on the configured backup infrastructure. As described in the HP Data Protector software performance white paper in the Performance Bottlenecks section, several components influence the performance in such a solution. These are:

- Disk array hardware and/or disk array configuration
- Raid Level
- Backup server hardware (sufficient main memory, CPU power, I/O performance)
- Network infrastructure
- File System architecture
- Tape drive hardware (LTO, DDS, DLT)
- Compression (software compression)
- Encryption (software encryption)

Some customers might have a focus on fast performing backups and some might focus on fast restore processes. Depending on those requirements or existing SLAs (Service Level Agreements), one or the other backup/restore method might become the preferred method. However a backup/restore process can only be as fast as the slowest component in such a solution. This fact also applies to object consolidation. As an object consolidation also requires the underlying backup infrastructure, it becomes obvious that, from a component point of view, a slow performing device will affect an object consolidation process as well. In addition to the hardware aspect, the way a backup and an object consolidation have been configured also affects consolidation processing time.

A consolidation process in general requires a full backup containing x objects and in addition y incremental backups to be read. A large number of streamed, highly multiplexed objects to a file library or tape device will most likely result in a long-lasting object consolidation process. In an optimal setup, from a consolidation point of view, a non/low multiplexed backup will have a positive influence to the consolidation processing time. See elapsed time for object consolidation in Synth-Case3, Synth-Case4, Virtual-Case3 and Virtual-Case4 in the previous comparison charts.

In general, the following aspects need to be considered prior to the implementation of Synthetic Full and Virtual Full as a backup strategy:

- **Quantities of objects in one consolidation run**: Limit the maximum number of objects to 32 for best writer (Media Agent) utilization. Set concurrency per writer (Media Agent) to one (1) to get non-multiplexed backup media. Consider to split up object consolidation runs into smaller consolidation jobs for backup media where the number of objects is beyond 32 objects. Run consolidation sessions in sets of 32 objects. Avoid running those sessions in parallel as they might access the same disk subsystem.

- **Size of objects in one consolidation run**: Check the size of objects in one consolidation since the size of objects matters for the backup duration as well as for the consolidation process. A consolidation will take about the same time to a restore an object(s) plus processing overhead.

- **Quantity of object versions (Incremental backups)**: Perform frequent consolidation sessions to keep the total number of object versions in a consolidation session low.

- **Concurrency with which the backup media (full and incremental) was written**: Modify the number of Disk Agents started for each Media Agent (called Disk Agent (backup) concurrency) using the Advanced options for the device or when configuring a backup if required. The concurrency set in the backup specification takes precedence over the concurrency set in the device definition. Fast backup device technology requires a constant data stream to stay in streaming mode. This can be achieved through a large number of Disk Agents started for that media. A side effect of such a configuration is a backup media with a high degree of interleaved data. This circumstance affects data restore performance as theoretically the entire tape or file depot need to be read for a single file.

- **Quantity of devices (Media Agent)**: A Media Agent is a process that controls reading from and writing to a device, which reads from or writes to a medium (typically a tape). During a backup session, a Media Agent receives data from the Disk Agent and sends it to the device for writing it to the medium. During a restore session, a Media Agent locates data on the backup medium and sends it to the Disk Agent. The Disk Agent then writes the data to the disk. A Media Agent also manages the robotics control of a library. Typically, devices are licensed in HP Data Protector software, meaning each tape drive in use requires a valid license. With the introduction of the HP Data Protector software file library, a new licensing schema for such devices has been introduced. The Advanced Backup to Disk LTU is a capacity based license for such devices. These licenses are available in 1xTB/10xTB/100xTB versions. This circumstance allows configuring a large number of devices (writers) within a HP Data Protector software file library to utilize their capabilities mentioned earlier in conjunction with a low concurrency.

- **Load balancing of devices (Media Agent)**: Balance the load (usage) of devices automatically by default so that they are used evenly. Load balancing optimizes the device usage by balancing the number of objects backed up to each device. Since load balancing is done automatically during backup time, you do not need to manage the assignment of objects to devices, so that all assigned devices stay busy during the backup session. Just specify the devices to be used for this session and Data Protector will dynamically assign them.

  Min
  The minimum number of available devices (devices that are not being used by another Data Protector session and have the license to be started) required for starting the session. If fewer devices are available than specified here, the session will queue. The default is 1.

  Max
  The maximum number of available devices that Data Protector will use in the session at the same time. The highest number you can specify here is 32. The default is 5.

  For object consolidation purposes, it is beneficial to increase the Max number to the number of configured writers.

- **User rights**: Set appropriate user rights for starting an object consolidation session. The same user rights apply as for backup.
  User rights or access rights are the permissions needed to perform specific Data Protector tasks. Configuring a backup, starting a backup session, or starting a restore session are typical user tasks Users have the access rights of the user group to which they belong.

- **Availability of media and/or devices**: For a successful object consolidation, all objects and objects versions along with their associated devices must be present and available. An object consolidation for an object will not be successful and abort if any of the required items are not available.

- **Source of data to be consolidated**: For an object consolidation, the full backup can reside on any storage media such as a local attached device (DDS, DLT, LTO) or Library or VTL device or file library on disk. The incremental backup media however must reside within a HP Data Protector software file library. It is beneficial to use libraries with robotics or VTL devices or disk file libraries as potential media mount requests of local attached devices might interrupt the consolidation process. Matching block sizes need to be considered while devices are being configured. Disk fragmentation should be verified as disk head repositioning influences consolidation processing time.

- **Target of data to be consolidated**: For a synthetic full, attempt to use physically-separate source and target devices as the consolidation session process might invoke a high number of read/write cycles on the disk subsystem. Block size should be equal or larger than source block size.

- **Avoid block size repackaging**: A medium is divided into data segments, catalog segments, and a header segment. Header information is stored in the header segment, which is the same size as the **block size**. Data is stored in data blocks of data segments.
  Segments are not written as a whole unit, but rather in smaller subunits called blocks. The device hardware processes data it receives using a **block size** specific to the device type.
  Data Protector uses a default device block size regarding different device types. The block size applies to all devices created by Data Protector and to Media Agent running on the different platforms.
  Increasing the block size can improve performance. You can adjust the blocks sent to the device while configuring a new device or when changing the device properties using the Advanced options for the device.
  You should change the block size before formatting tapes. The device block size is written on a medium header so that Data Protector knows the size to be used. If the device block size differs from the medium block size, an error occurs.

However, before changing the block size for the device, you need to check the supported block size of the used host adapter. The minimum block size for old SCSI cards, such as Adaptec 2940, used to be 56 KB. Currently, the minimum block size that is mainly used with newer SCSI cards is 64 KB.
You can increase the maximum block size on a Windows Media Agent client by modifying its Registry. Before changing the block size for a particular SCSI card, refer to the SCSI vendor documentation or contact the vendor support.
For object consolidation purposes avoid data block repackaging by making sure target device block size is equal or larger that source device block size.

- **HP Data Protector software file library size**: A file library is a device which resides on an internal or external hard disk drive defined by the user. The file library device can include one or several directories. Only one directory can be located on a file system. There is no maximum capacity for the file library device that is set by Data Protector. The only limit on the size of the device is determined by the maximum size of the file system where the directory is located. For example, the maximum size of the file library device running on LINUX would be the maximum size you can save on the file system.

- **File Depot Size**: File depots are the files containing the data from a backup to a file library device. The size of file depots is defined when you initially create the file library device. During this process you specify all sizing properties for the device, including the maximum size of the file depots. The sizing properties of the file depots, although only entered once, are globally applied to each file depot. If the size of data to be backed up within one session is larger than the originally specified file depot size, Data Protector automatically creates more file depots until the allocated disk space for the file library device has been consumed. On Windows, the default file depot/slot size is 50 GB, although the file library device has been tested on Windows with file depots of up to 600 GB. On UNIX, the maximum file depot size is 2 TB. A good choice is a size of 10 GB on any platform. Data Protector automatically creates file depots until there is no more disk space available for the device. The amount of space which must stay free for the file library device is defined in the device properties when the device is initially being set up.
  If the total disk space available to the file library device goes below a user specified level, a notification is issued.

  In situations where the file depots are located on a variety of disks, it is not recommended to put file depots from two different file library devices on a single disk. This is owing to the fact that if the properties are different, it can cause a conflict in Data Protector (for example, if on one file library device the remaining disk space for the file depot is specified as 20 MB and on the other file library device 10 MB).

- **Scheduled or Post-Backup consolidation device mapping**: The device mapping for a successful scheduled or post-backup object consolidation should to be done before the first consolidation run. The mapping can be complex as all writers from all configured devices will be displayed in the configuration screen.

Command Line Interface

omniobjconsolidate

`omniobjconsolidate` is a command intended to interface consolidation and is similar to `omniobjcopy`. It can be used to start interactive as well as preconfigured consolidation specifications.

Syntax

```
Syntax 1: omniobjconsolidate -consolidationlist
ConsolidationSpecificationName -scheduled [General Options]

Syntax 2: omniobjconsolidate -consolidationlist
ConsolidationSpecificationName -postbackup -session sessionId [General
Options]

Syntax 3: omniobjconsolidate -version | -help

Syntax 4: omniobjconsolidate -no_monitor Object [Object .... ] [General
Options]

Object Syntax:

{-filesystem | -winfs | -vbfs | -netware | -omnidb } Client:MountPoint Label -session
SessionID [–copyid copyId] [-sourcedevice LogicalDevice] -consolidatedevice
LogicalDevice -targetdevice LogicalDevice [-protect {none | weeks n | days n
| until Date | permanent}] [-keepcatalog {weeks n | days n | until Date |
permanent}] [-[no_]log | -log_dirs | -log_file]] [-recycle]

Global Options:

-report {warning | minor | major | critical}
```

Options explained:

- **-version**: Displays the version of the `omniobjconsolidate` command.
- **-help**: Displays the usage synopsis of the `omniobjconsolidate` command.
- **-no_monitor**: Monitors the session and displays all messages by default. If this option is used, the command displays only the sessionID.
- **-consolidationlist**: Selects certain consolidation specification.
- **-scheduled**: Starts immediately automated (scheduled) consolidation specification.
- **-postbackup**: Starts immediately an automated (post-backup) consolidation specification related to the referenced sessionID specified by the `-session` option.
- **-session**: Provides sessionID for post-backup consolidation in Syntax 2 and provides version information of an object in Object Syntax.
- **-recycle**: Recycles source objects data and catalog protection after successful compaction.
- **-filesystem**: Selects restore chain of a filesystem identified with `Client:MountPoint` Label for consolidation.
- **-winfs**: Selects restore chain of a Windows filesystem identified with `Client:MountPoint` Label for consolidation.
- **-copyid**: Identifies certain copy if present. If not present, Data Protector performs automatic media-set selection to select best copy as source-object-version.
- **-sourcedevice**: Specifies logical device other than default that will read full backup object from medium. By default, the device that is used to read object is the same as has been used to write object.

• **-consolidatedevice**: Specifies logical file device that will read incremental object-versions and perform consolidation.

• **-targetdevice**: Specifies logical device that will write synthetic object-version to medium.

• **-protect**: Specifies consolidated object data protection if present. If not present, same as full backup of restore-chain.

• **-keepcatalog**: Specifies consolidated object-version catalog protection if present. If not present, same as full backup of restore-chain.

• **-[no_]log, -log_dirs, -log_file, -log_all**: Specifies consolidated object log level if present. If not present, `-log_all` is used.

• **-report**: Sets the level of error notification for the session.

• **Runtime**: `omniobjconsolidate` behaves exactly as `omniobjcopy` during runtime. This means:

  • console is exactly the same (abort, mount request, ...)

  • session messages are displayed the same

Examples:

• To start a consolidation session that consolidates the `WinFS` object versions for `OBJECT1` on the `system1.company.com` host to the point in time defined with the `2006/09/06-1` session ID, using the device `LTO3`" as the source device and the `FILEDEV1` file library as the consolidation device, and writes the consolidated objects to the `LTO4` device, use:
```
omniobjconsolidate -winfs system1.company.com:/C 'OBJECT1' –session
2006/09/06-1 -sourcedevice 'LTO3' -consolidationdevice 'FILEDEV1' -
targetdevice 'LTO4'
```

• To start an interactive consolidation session for the `system1.company.com:/ 'Label42'` filesystem object from the `2006/07/01-2` session, using the `DEV1` device to read the source object and the `DEV2` device to consolidate the object, and write the consolidated object to the `DEV3` device, use:
```
omniobjconsolidate -filesystem system1.company.com:/ 'Label42' -session
2006/07/01-2 -sourcedevice 'DEV1' –consolidationdevice 'DEV2' -
targetdevice 'DEV3'
```

• To start a post-backup consolidation specification named `post_BU1` immediately for the `2006/04/03-1` session, use:
```
omniobjconsolidate -consolidationlist post_BU1 -postbackup –session
2006/04/03-1
```

• To start a scheduled consolidation specification named `ConsolidationSpec` immediately, use:
```
omniobjconsolidate -consolidationlist ConsolidationSpec -scheduled
```

Configuration files and options

consolidation specification files

Locations:

Automated consolidation specifications are stored in two locations:

1. automated post-backup consolidation specifications:
   `<DP_HOME>\Config\Server\consolidationlists\afterbackup`
2. automated scheduled consolidation specifications:
   `<DP_HOME>\Config\Server\consolidationlists\scheduled`
3. Automated scheduled copy operations schedules are stored in:
   `<DP_HOME>\Config\Server\consolidationlists\scheduled\schedules`

The name of the schedule is the same as the name of the automated scheduled consolidation specification.

consolidation specification options

```
OPTIONS
{
[-protect {none | weeks n | days n | until date | permanent | source}]
[-keepcatalog {weeks n | days n | until date | permanent | source }]
[-[no_]log | -log_dirs | -log_file | -log_all]]
[-dynamic _min_ _max_]
[-locationpriority _location]
...
[-locationpriority _location]
[-recycle]
}
READ_DEVICE_MAPPING
{
_device1_ _device2_
...
_deviceN-1_ _deviceN_
}
CONSOLIDATION DEVICES
{
_device1_
...
_deviceN_
}
FILTER (datalist | protection | timeframe | objects)
```

```
{
(_datalist_ _datalist_ ... _datalist_) |
(-reltime _startTimeT_ _durationTimeT_ | -abstime _startTimeT_
_endTimeT_) |
(_objectname_ _objectname_ ... _objectname_)
}
OBJECT _objectName_ _sessionId_
{
[-sourcedevice _deviceName_]
[-consolidatedevice _deviceName_]
[-copyid N]
[-protect {none | weeks n | days n | until date | permanent | source}]
[-keepcatalog {weeks n | days n | until date | permanent | source }]
[-[no_]log | -log_dirs | -log_file | -log_all]]
[-recycle]
}
DEVICE _deviceName_
{
/* options supported in backup specifications go here */
}
```

Use Cases

Use case #1: Daily file system incremental backup with object consolidation afterwards.

This scenario describes a change from an existing weekly full backup with daily incremental backups in between to an incremental forever file system backup using one initial full backup with subsequent incremental daily backups. An object consolidation is performed automatically after each daily incremental backup.

Usage: File Servers with high availability and fast response time requirements. High data change rate in file system. Short restore chain, short restore time requirement.

Note: Capacity of backup data can increase quite a bit assuming the retention policy is the same as without object consolidation.

Virtual Full might rectify capacity demands as long as data stays on File Library.

Figure 28 depicts the current backup strategy which is weekly full plus daily incremental backups, whereas Figure 29 shows the new backup strategy of "Incremental Forever".

Figure 28. Weekly full plus daily incremental backups

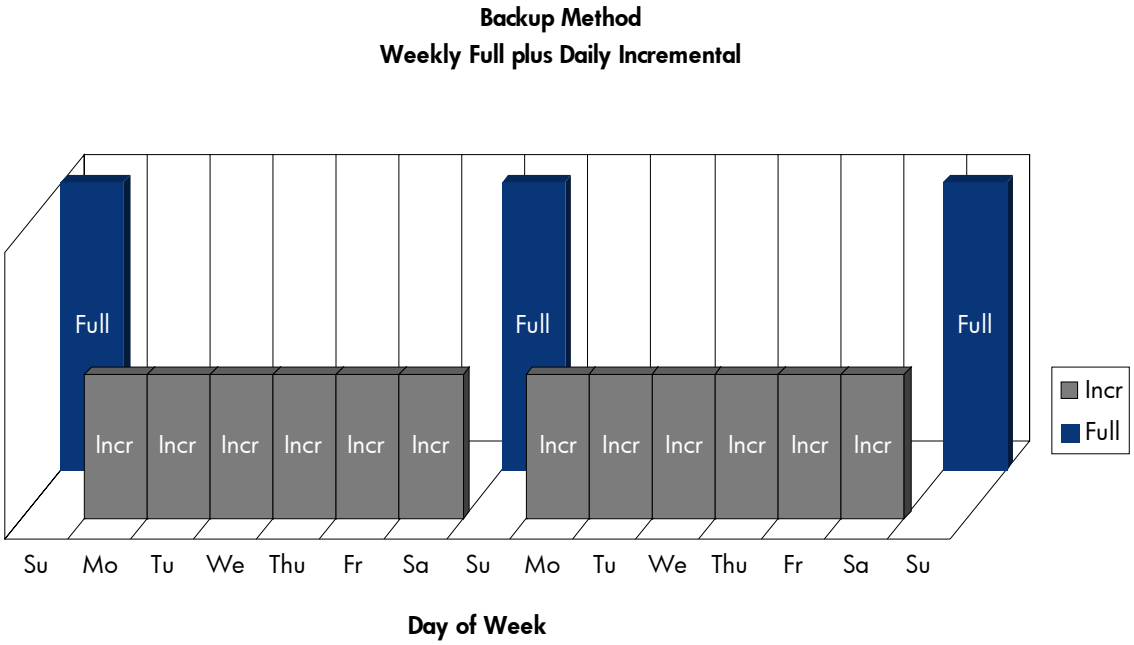
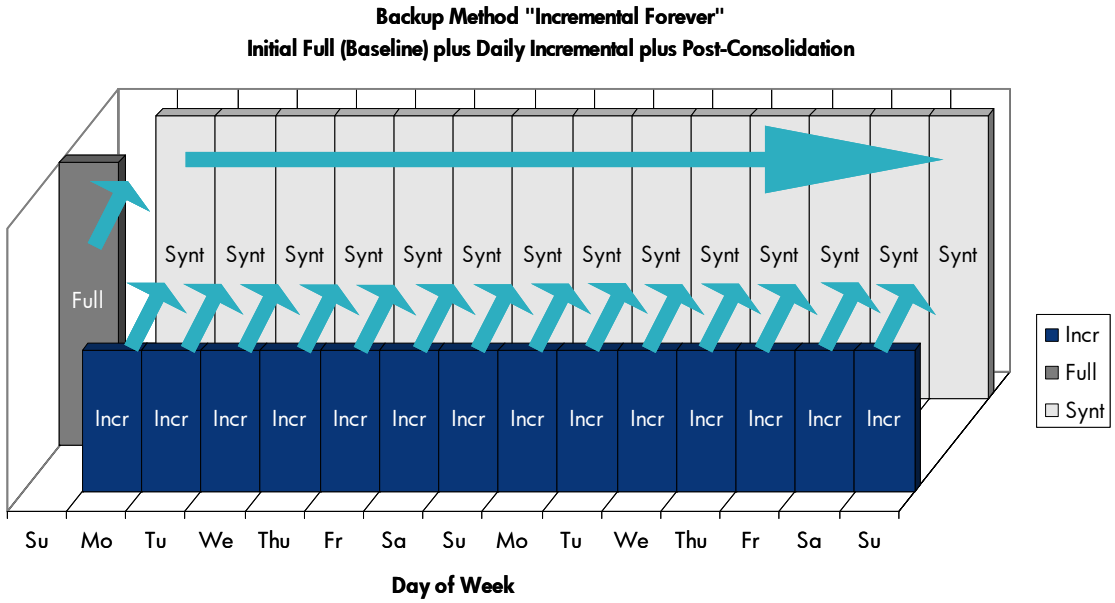


Figure 29. New backup strategy "Incremental Forever"



For the new backup strategy, a new backup retention policy should be considered as the data volume grows significantly as the consolidation always includes the predecessor backup which includes the full backup. The full backup and the incremental backup media might get a relatively short protection period. Monthly or quarterly full backups provide increased data security.

The advantage of this backup strategy is definitely the relatively short incremental backup run which introduces minimal additional processor load on the server that the HP Data Protector software Disk Agent runs on.

The data consolidation itself takes place on the backup host and can be performed as either a post-backup job or a scheduled job in periods where the consolidation task might not affect important running or scheduled backup jobs on the HP Data Protector software cell server.

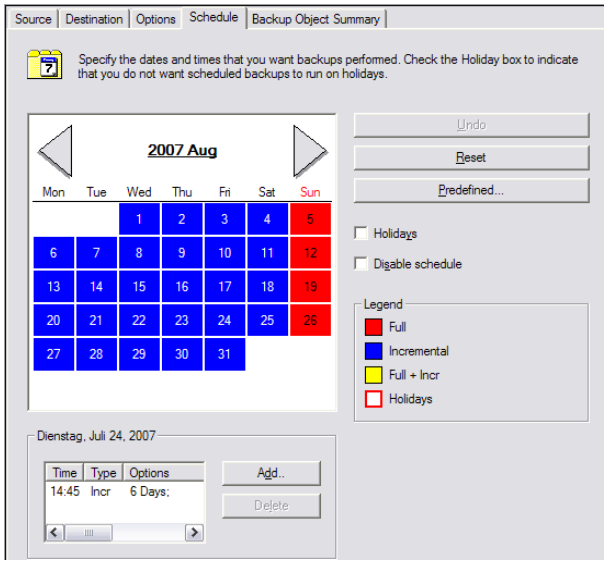
Configuration

A change of the existing backup schedule is required. Figure 30 shows the backup scheduler configuration screen before the change to the new incremental forever backup strategy.

Incremental backups with a protection period of six days were scheduled for regular working days. Full backups were scheduled for Sundays.

Alternatively a new backup specification can be configured and deployed.

Figure 30. Backup scheduler



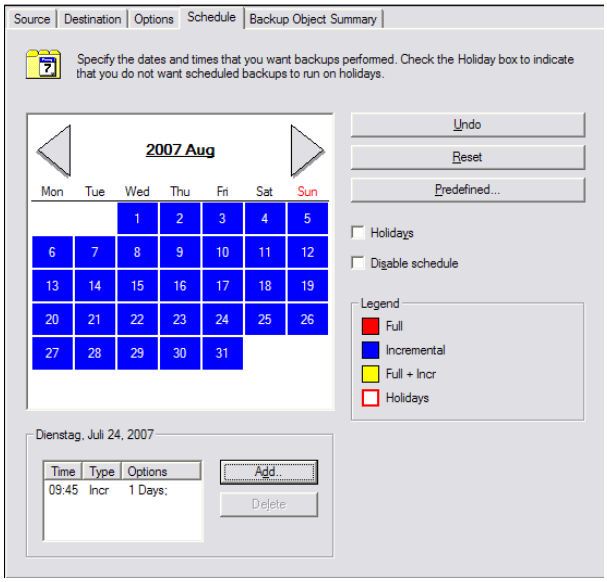
Use the Reset button to delete any existing backup schedule.

The Add button will open a configuration dialog in which the new backup schedule can be defined.

The image in Figure 31 shows incremental backups configured for every day of the week.
The protection is set to one day.

Note: The weekly full backup on Sundays has been replaced by an incremental backup.

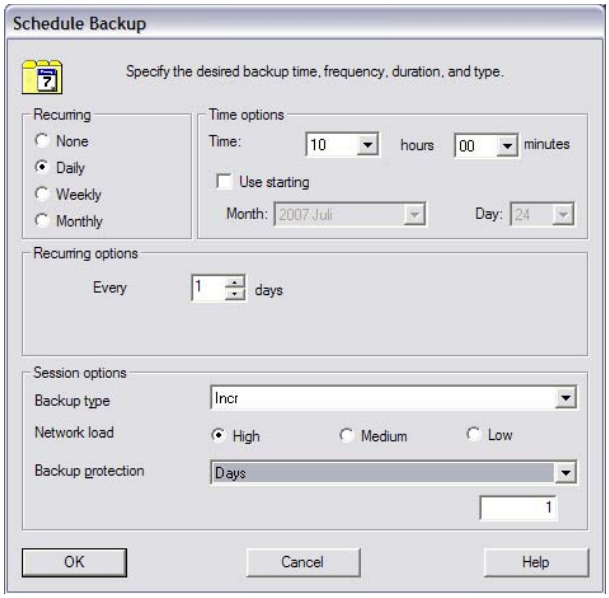
Figure 31. Incremental forever schedule



In the Schedule Backup configuration dialog, select the appropriate schedule, backup type and backup protection.

The protection period was set to one day as there will be an object consolidation automatically triggered after each successfully run incremental backup.

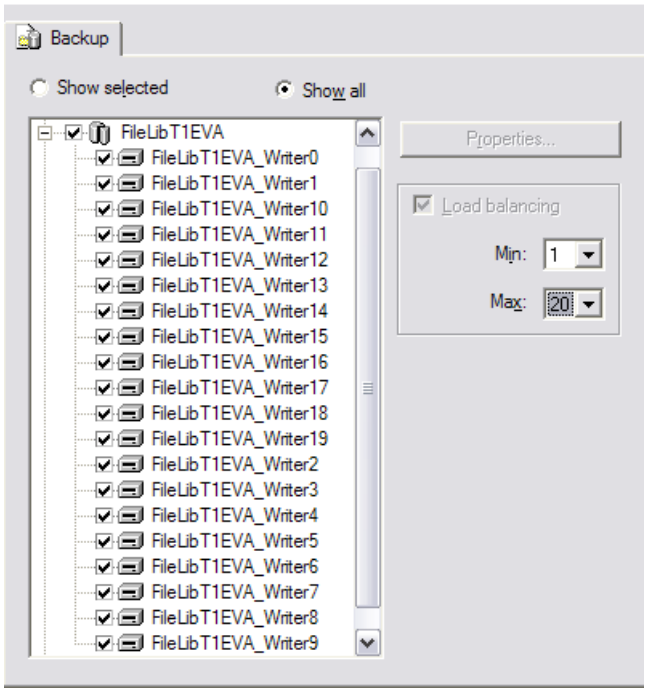
Figure 32. Schedule backup configuration dialog



The backup specification is changed to utilize a higher number of writers. In this case, the number of writers matches the number of objects. The writer configuration matches the configuration of the consolidation writer (block size: 64 KB, concurrency: 1).

The Max: value is set to 20 to allow twenty concurrent Media Agent processes writing the incremental backup to the HP Data Protector software file library.

Figure 33. Backup writer setting

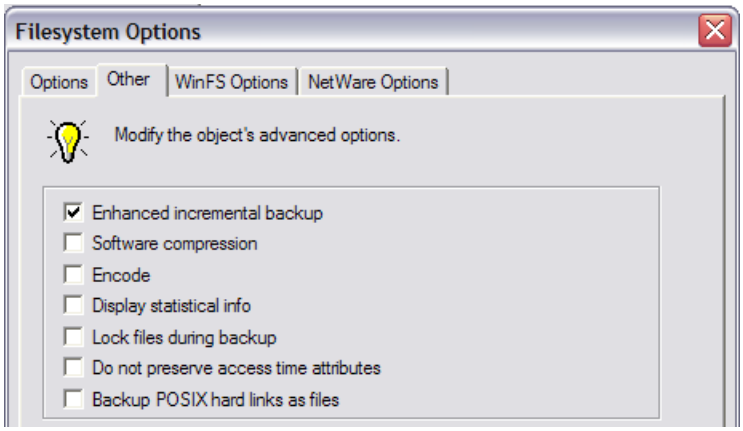


Note

By default there is only one writer per directory.

In addition the backup mode needs to be set to "enhanced incremental" if it was not set to that mode.

Figure 34. Enhanced incremental backup



Note

After you select this option, incremental backup will run in the enhanced mode only after a full backup is performed.

To start running enhanced backups with the next backup when an incremental backup is scheduled next, set the `EnhIncrPromoteToFullIfNoEnhFull` global variable to `1`, which will promote the next backup to a full backup.

HP Data Protector software will recognize the backup mode change and will issue the following warning upon running the next incremental backup:

```
[Warning] From: BSM@tpc042.deu.hp.com "mptest1" Time: 12.12.2007 13:39:29
 Missing Enhanced incremental full backup for
 WinFS tpc025.deu.hp.com:/N "N:".
 Incremental backup will not run in the enhanced mode until
 a full backup is performed.
```

As a consequence, a full backup in enhanced incremental mode need to be performed.

A consolidation is not possible until the first enhanced incremental full backup becomes available.

For the post-consolidation task, in this scenario, a consolidation writer setting was chosen to achieve maximum consolidation and restore per object performance. Test case 4 reflects the optimal setting for this requirement.

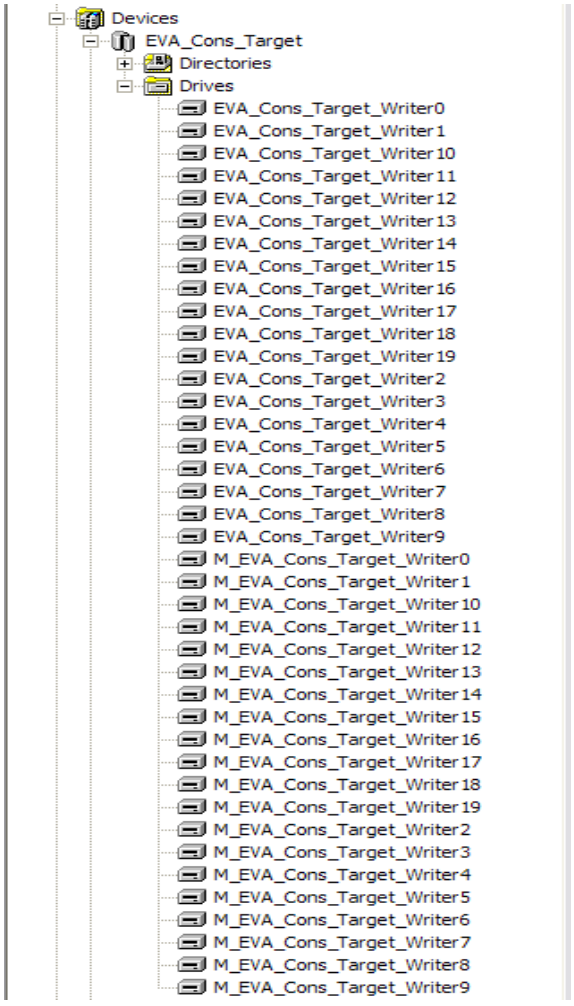
For every writer for the consolidation file library a mapping device represented by `M_EVA_Cons_Target_Writer` was setup using the same configuration parameters such as block size and concurrency.

The consolidation file library has now 40 writers in total. The backup specification defines 20 objects to be backed up.

Figure 35 depicts the consolidation device writer configuration. Each writer has the standard file library block size of 64kB and a reconfigured concurrency of 1.

The remaining settings have been left untouched.

Figure 35. Consolidation File library writer



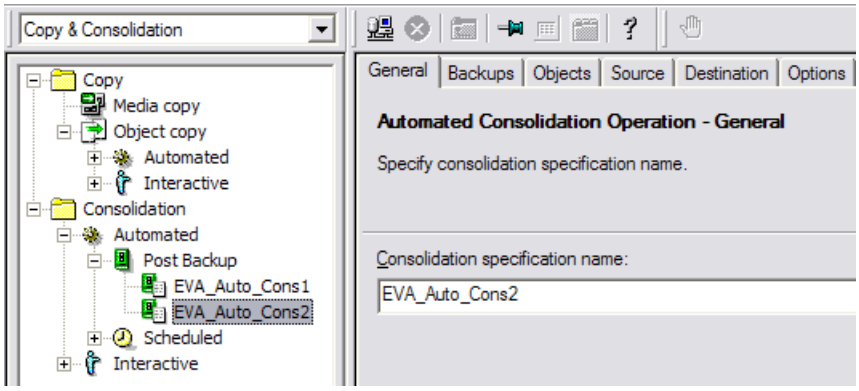
A new post-backup consolidation configuration will consolidate objects automatically upon finishing the backup job.

Post-backup object consolidation takes place after the completion of a backup session that is specified by the name of the backup specification in the automated object consolidation specification.
It consolidates objects backed up in that particular backup session that match the specified criteria.

Steps

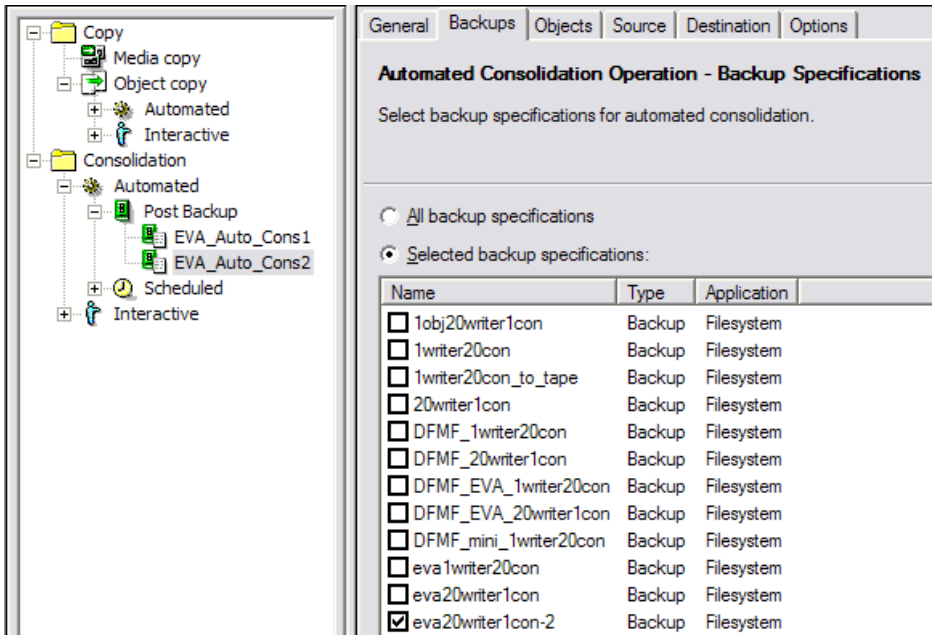
1. In the Context List, click **Copy & Consolidation**.

Figure 36. Post Backup Consolidation Context



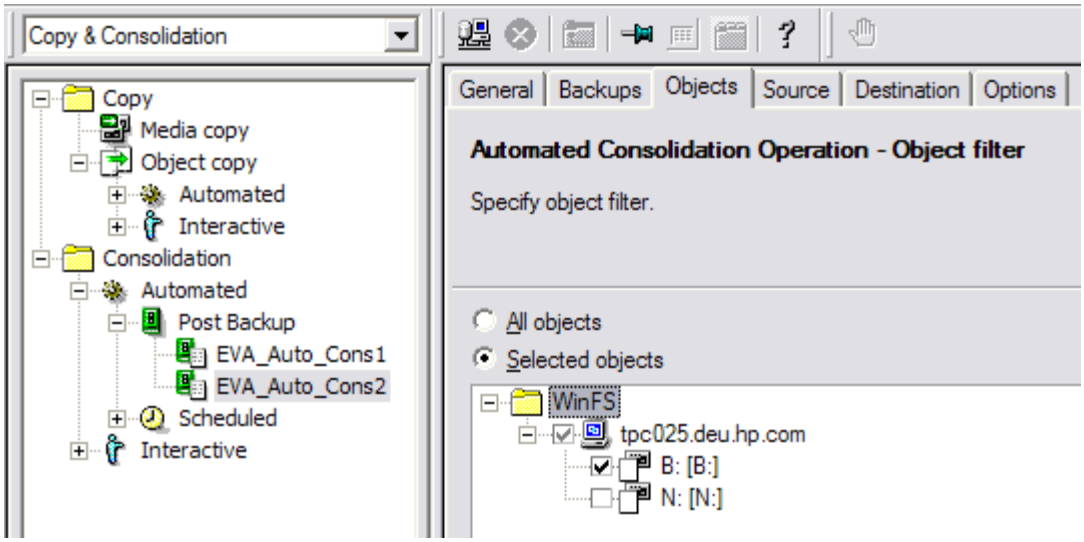
2. In the Scoping Pane, expand **Consolidation**, and then expand **Automated.**
3. Right-click **Post Backup** and click **Add** to open the wizard.
4. In the Consolidation specification name text box, type a name for the object consolidation specification. Click **Next**.
5. Select the backup specifications that contain the objects you want to consolidate. Click **Next**.

Figure 37. Select Backup Specification



6. Specify the object filter for the object consolidation operation. Click Next.

Figure 38. Select Objects to be consolidated

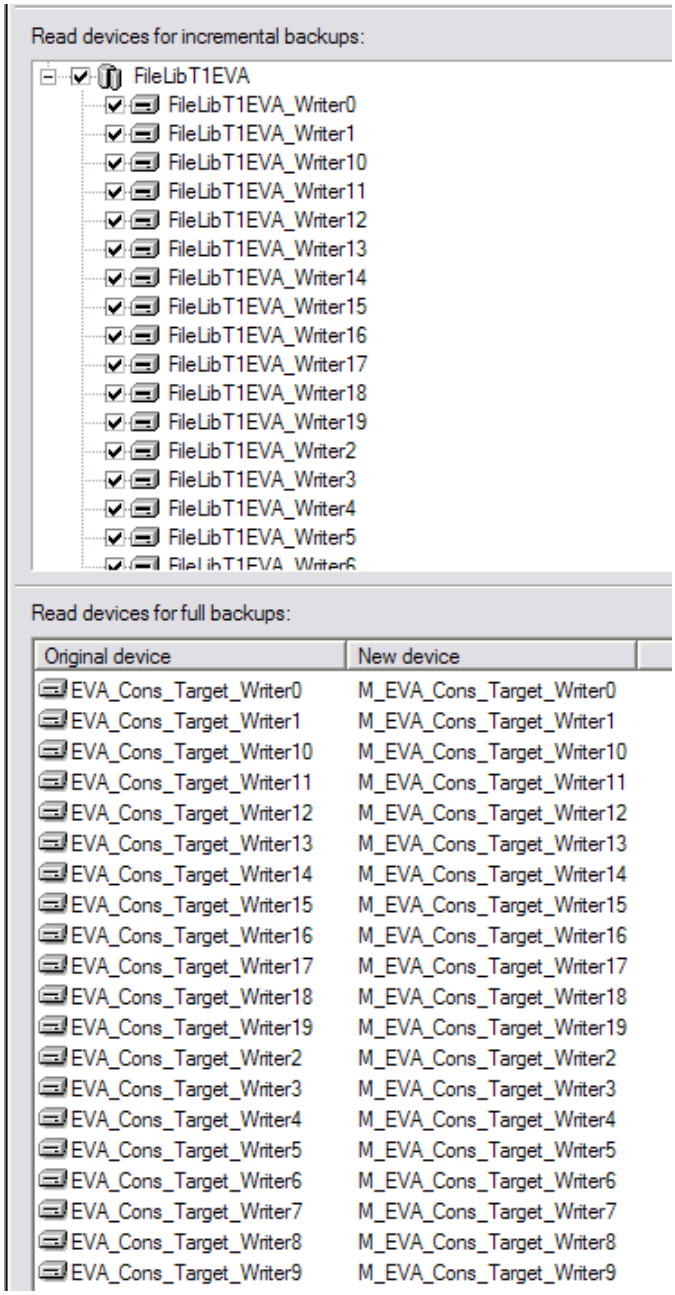


7. Specify the devices that will read the incremental backups and the full backups.

Limit the object consolidation to specific file libraries by selecting these libraries as read devices for incremental backups. Only objects residing in the specified file libraries will be consolidated.

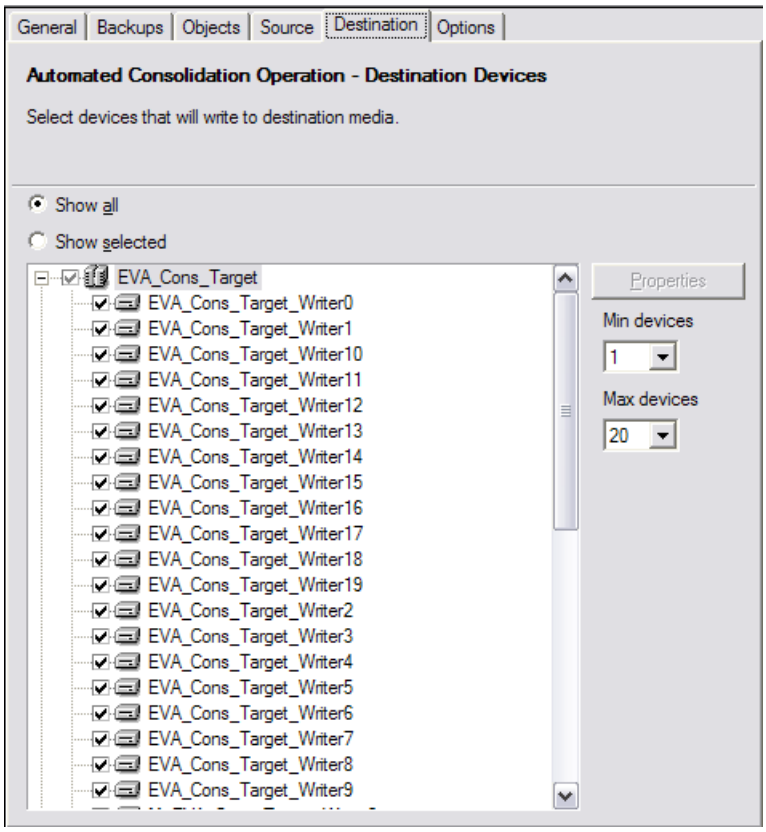
By default, the read devices for full backups are those used for backup in the selected backup specifications. You can change them here if desired. In the section for read devices for full backups, a device mapping was done to always have enough writers for the consolidation process. The original writer is mapped to the "mapping device".
Click **Next.**

Figure 39. Read devices mapping



8. Select the destination devices for the object consolidation operation. Data Protector software will select the most suitable devices from those you specify here. Make sure to match the Max devices setting to the number of selected writers.
   Please note: Max. devices can not exceed 32.
   Click **Next**.

Figure 40. Destination Devices



9. Specify options as desired. To save disk space, the check box “Recycle data and catalog protection after successful consolidation” can be selected.

   Select this option to remove data- and catalog-protection of the objects on the source media. When there are no more protected objects on the media, the media can be overwritten.

   Important
   If you recycle data protection of source objects, the recycled points in time will no longer be available. Unless copies of these points in time exist, you will be able to restore only to the latest (consolidated) point in time.

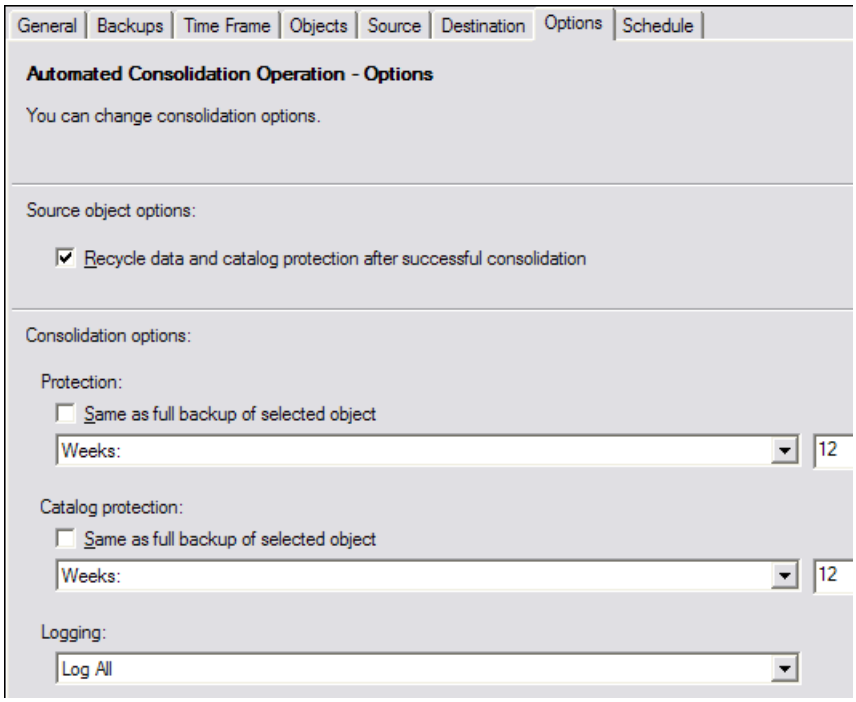
   On checking the check box, HP Data Protector software will respond with a warning message similar to the one shown in Figure 42.

Option: Catalog protection
This option defines how long information about the consolidated objects (such as file names and file versions) is kept in the IDB.

By default, the catalog protection of the consolidated objects is the same as the catalog protection of the full backup of the objects. If a relative period of catalog protection was set for the full backup, such as "n" days or weeks, the same protection period is counted from the creation time of the synthetic full backup.

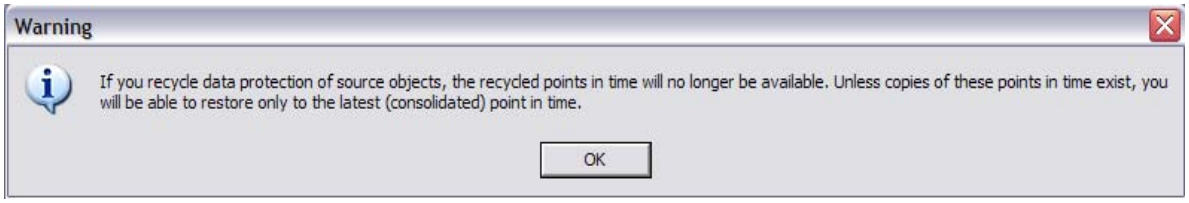
To specify a different protection period, deselect the **Same as full backup of selected object** option and select one of the following:
None: The catalog will not be protected.
Until: The catalog protection will last until the specified date.
Days: The catalog protection will last for the specified number of days.
Weeks: The catalog protection will last for the specified number of weeks.
Same as data: The catalog protection will be the same as data protection.
Click **Next.**

Figure 41. Consolidation Options



NOTE: You may specify a new backup/catalog protection time as the source of the object consolidation might have a too short protection period.

Figure 42. Warning message



10. Click Finish to exit the wizard.

Backup Results of Use Case #1

Figure 43. Restore Session information

Name	Status	Backup Specification	Backup Type
2007/12/13-58	Completed	eva20writer1con-2	incr
2007/12/13-59	Failed	EVA_Auto_Cons2	full
2007/12/13-60	Completed	eva20writer1con-2	incr
2007/12/13-61	Completed	EVA_Auto_Cons2	full
2007/12/13-62	Completed	eva20writer1con-2	incr
2007/12/13-63	Completed	EVA_Auto_Cons2	full
2007/12/13-64	Completed	eva20writer1con-2	incr
2007/12/13-65	Completed	EVA_Auto_Cons2	full
2007/12/13-66	Completed	eva20writer1con-2	incr
2007/12/13-67	Completed	EVA_Auto_Cons2	full
2007/12/13-68	Completed	eva20writer1con-2	incr
2007/12/13-69	Completed	EVA_Auto_Cons2	full
2007/12/13-70	Completed	eva20writer1con-2	incr
2007/12/13-71	Completed	EVA_Auto_Cons2	full

The `2007/12/13-58` session was the initial session for this backup specification. The backup type is `incr` (incremental) as it was configured incremental in the scheduler configuration.

The session message of the `2007/12/13-58` session shows the following warning:

```
[Warning] From: BSM@tpc042.deu.hp.com "eva20writer1con-2" Time:
13.12.2007 11:45:05
[61:11018]  Could not find related full backup of the object
  tpc025.deu.hp.com:/B named "B:".
  Backup mode is changed to full.
```

Reason: As this backup specification was newly deployed, there was no initial full backup for this object to be found. HP Data Protector software, therefore, switched the backup mode to “full” automatically.

The `2007/12/13-59` session shows “Failed” session status.

The session message of the 2007/12/13-59 session shows the following critical message:

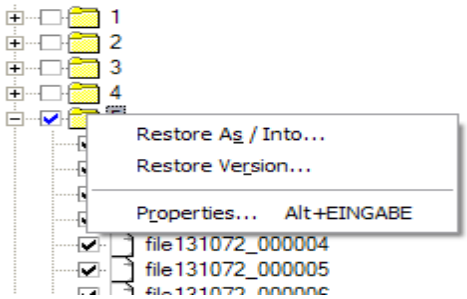
```
[Critical] From: CSM@tpc042.deu.hp.com "EVA_Auto_Cons2" Time: 13.12.2007
11:46:11

      No objects to consolidate. The session will abort.
```

Reason: The automated post-backup consolidation got started right after the initial full backup. There was no incremental backup for that object available which could have been consolidated into a new synthetic full backup.

The restore object pane allows restoring a specific version of an object. Right click on the relevant item to restore and select “Restore Version…”.

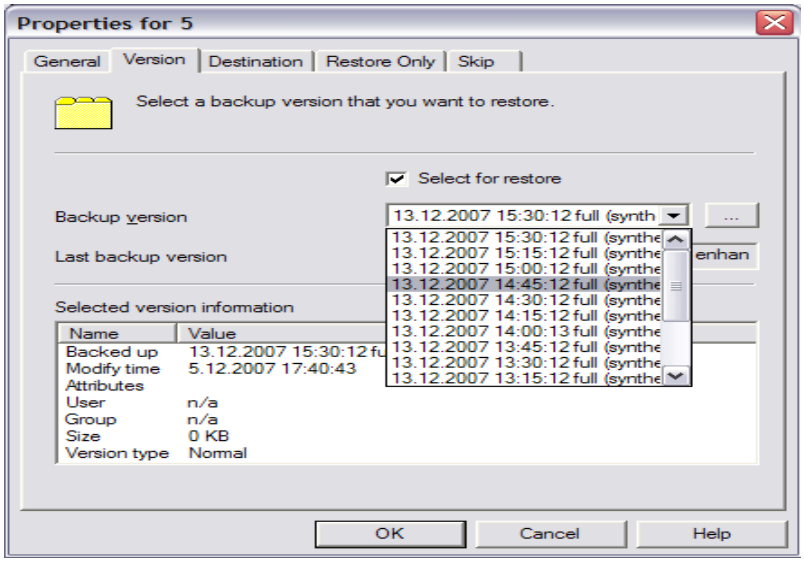
Figure 44. Restore Version dialog



A Property dialog box will open from which the desired object version can be selected.

Figure 45 depicts the Property dialog box.

Figure 45. Property dialog box



The drop-down box shown in Figure 45 allows you to select from a point-in-time backup version to restore represented by a synthetic full backup.

Note: There is only one incremental backup performed between each synthetic full. HP Data Protector software automatically hides the single instance of the incremental backup for restore. HP Data Protector software shows one version for each point-in-time.

Disk Space Consumption Traditional Incremental vs. Synthetic Full Backup vs. Virtual Full Backup

Figure 46, Figure 47, and Figure 48 represent the actual disk space consumption after a series of incremental backup and object consolidation runs. As already mentioned earlier, a synthetic full back up always includes the initial full backup and the consolidated version of n object versions. Space requirements can grow enormous.

Figure 46. Incremental Backup Disk Space Consumption

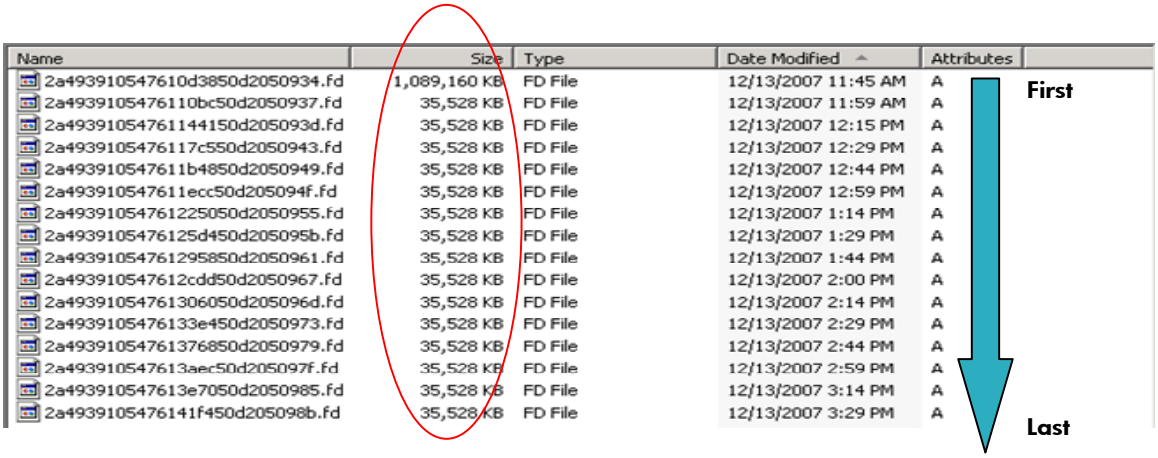
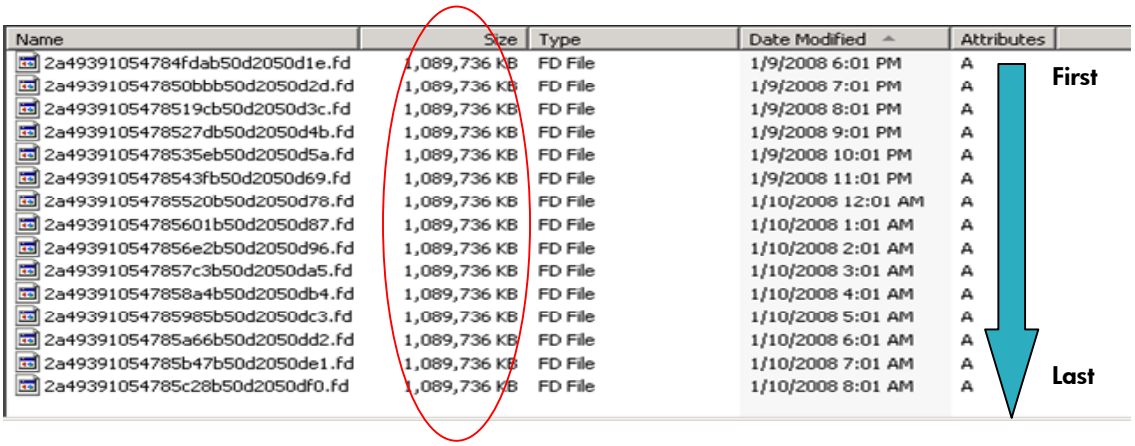


Figure 47. Synthetic Full Disk Space Consumption



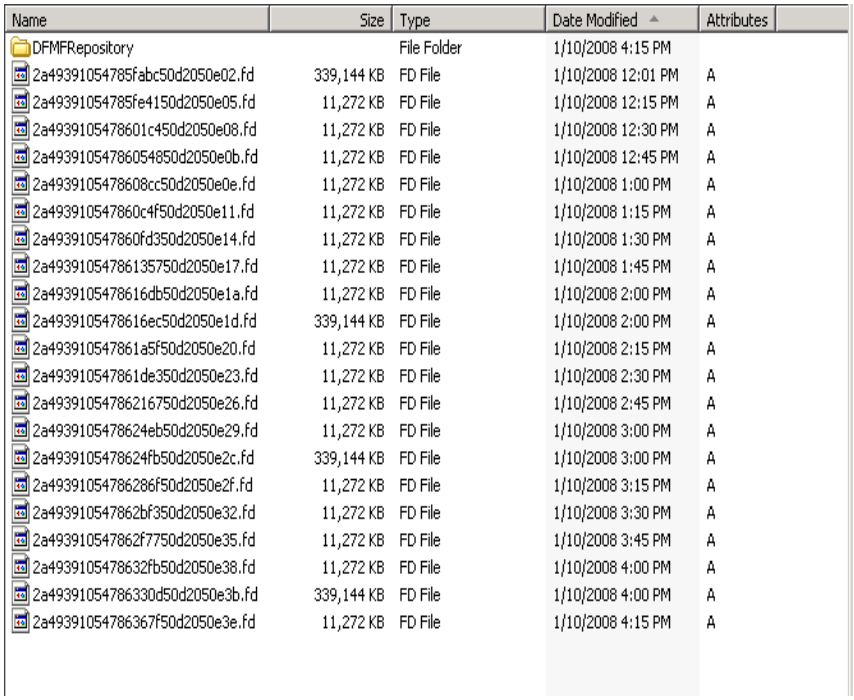
Proper planning of the retention period of the synthetic full and incremental backup becomes vital to control the disk space usage of the HP Data Protector software backup solution.

An alternative would be using a DFMF formatted HP Data Protector software file library allowing Virtual Full backups or an automated object copy run which is discussed in use case #3.

With virtual full backup, the space consumption largely depends on the size of the backed up files. If the files are significantly larger than the block size used, virtual full backup achieves maximum savings of the space compared to normal synthetic backup. On the other hand, if the files are smaller than the block size, the savings are rather small. As a rule of thumb one can calculate the additional space needed for a virtual full like:

(Total size of all files that are smaller than DP block size) + (DP block size * number of files larger than DP block size)

Figure 48. Virtual Full Disk Space Consumption without recycling of source data

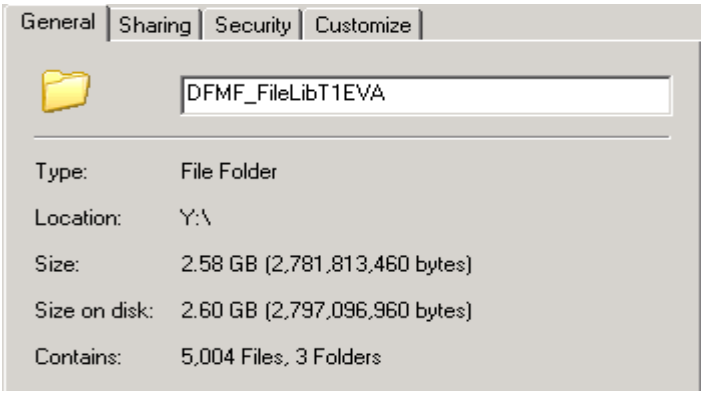


Name	Size	Type	Date Modified	Attributes
DFMFRepository		File Folder	1/10/2008 4:15 PM	
2a493910547 85fabc50d2050e02.fd	339,144 KB	FD File	1/10/2008 12:01 PM	A
2a493910547 85fe4150d2050e05.fd	11,272 KB	FD File	1/10/2008 12:15 PM	A
2a4939105478601c450d2050e08.fd	11,272 KB	FD File	1/10/2008 12:30 PM	A
2a4939105478605 4850d2050e0b.fd	11,272 KB	FD File	1/10/2008 12:45 PM	A
2a4939105478608cc50d2050e0e.fd	11,272 KB	FD File	1/10/2008 1:00 PM	A
2a493910547860c4f50d2050e11.fd	11,272 KB	FD File	1/10/2008 1:15 PM	A
2a493910547860fd350d2050e14.fd	11,272 KB	FD File	1/10/2008 1:30 PM	A
2a4939105478613 5750d2050e17.fd	11,272 KB	FD File	1/10/2008 1:45 PM	A
2a4939105478616db50d2050e1a.fd	11,272 KB	FD File	1/10/2008 2:00 PM	A
2a4939105478616ec50d2050e1d.fd	339,144 KB	FD File	1/10/2008 2:00 PM	A
2a493910547861a5f50d2050e20.fd	11,272 KB	FD File	1/10/2008 2:15 PM	A
2a493910547861de350d2050e23.fd	11,272 KB	FD File	1/10/2008 2:30 PM	A
2a4939105478621 6750d2050e26.fd	11,272 KB	FD File	1/10/2008 2:45 PM	A
2a4939105478624eb50d2050e29.fd	11,272 KB	FD File	1/10/2008 3:00 PM	A
2a4939105478624fb50d2050e2c.fd	339,144 KB	FD File	1/10/2008 3:00 PM	A
2a4939105478628 6f50d2050e2f.fd	11,272 KB	FD File	1/10/2008 3:15 PM	A
2a493910547862bf350d2050e32.fd	11,272 KB	FD File	1/10/2008 3:30 PM	A
2a493910547862f7750d2050e35.fd	11,272 KB	FD File	1/10/2008 3:45 PM	A
2a4939105478632fb50d2050e38.fd	11,272 KB	FD File	1/10/2008 4:00 PM	A
2a4939105478633 0d50d2050e3b.fd	339,144 KB	FD File	1/10/2008 4:00 PM	A
2a4939105478636 7f50d2050e3e.fd	11,272 KB	FD File	1/10/2008 4:15 PM	A

Without proper data retention period handling, the disk space consumption shown in figure 48 will steadily grow over time.

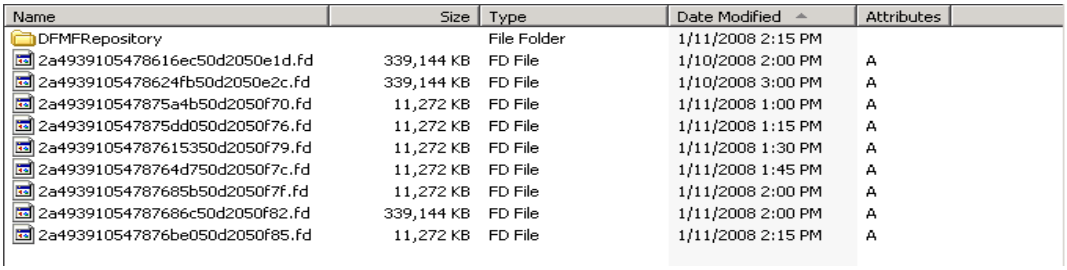
The space consumption shown in Figure 48 is depicted in Figure 49.

Figure 49. Disk space usage



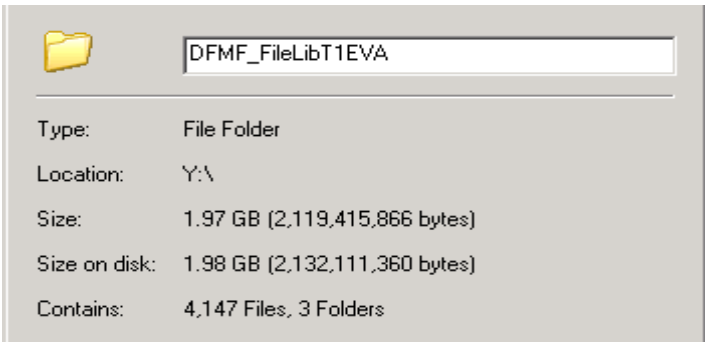
After considering a retention policy, the backup data can now be recycled automatically in predefined intervals.

Figure 50. Virtual Full Disk Space Consumption with recycling of source data



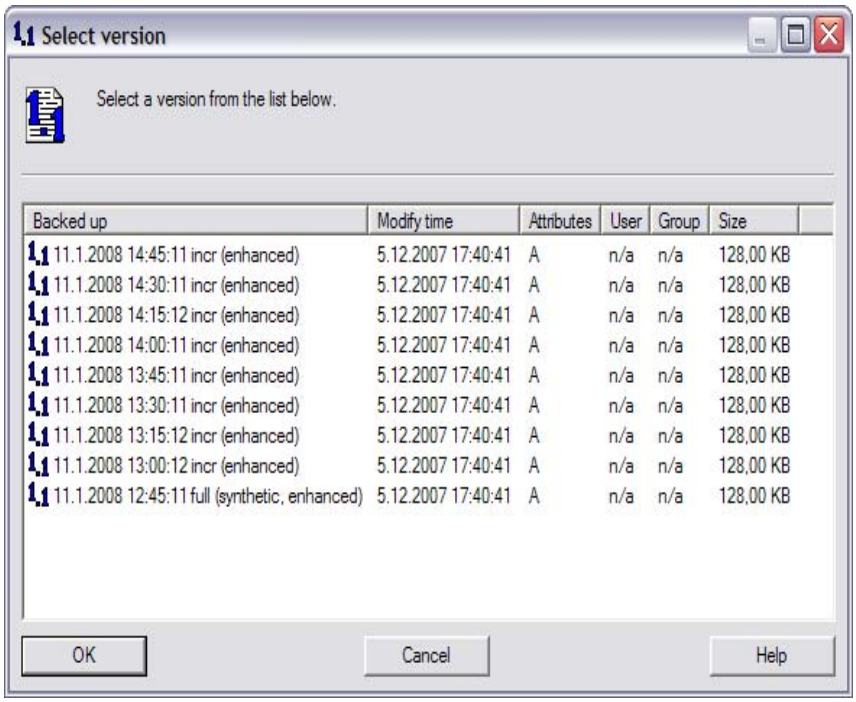
Name	Size	Type	Date Modified	Attributes
DFMFRepository		File Folder	1/11/2008 2:15 PM	
2a4939105478616ec50d2050e1d.fd	339,144 KB	FD File	1/10/2008 2:00 PM	A
2a4939105478624fb50d2050e2c.fd	339,144 KB	FD File	1/10/2008 3:00 PM	A
2a493910547875a4b50d2050f70.fd	11,272 KB	FD File	1/11/2008 1:00 PM	A
2a493910547875dd050d2050f76.fd	11,272 KB	FD File	1/11/2008 1:15 PM	A
2a4939105478761535 0d2050f79.fd	11,272 KB	FD File	1/11/2008 1:30 PM	A
2a4939105478764d750d2050f7c.fd	11,272 KB	FD File	1/11/2008 1:45 PM	A
2a49391054787685b50d2050f7f.fd	11,272 KB	FD File	1/11/2008 2:00 PM	A
2a49391054787686c50d2050f82.fd	339,144 KB	FD File	1/11/2008 2:00 PM	A
2a493910547876be050d2050f85.fd	11,272 KB	FD File	1/11/2008 2:15 PM	A

Figure 51. Disk space usage



In the restore object selection dialog, one can still choose from the last consolidated version or point-in-time versions of data.

Figure 52. Restore object version dialog for virtual full backups



Note: The GUI does not distinguish between virtual full and synthetic full backup. The virtual full backup is listed as “full (synthetic, enhanced)”.

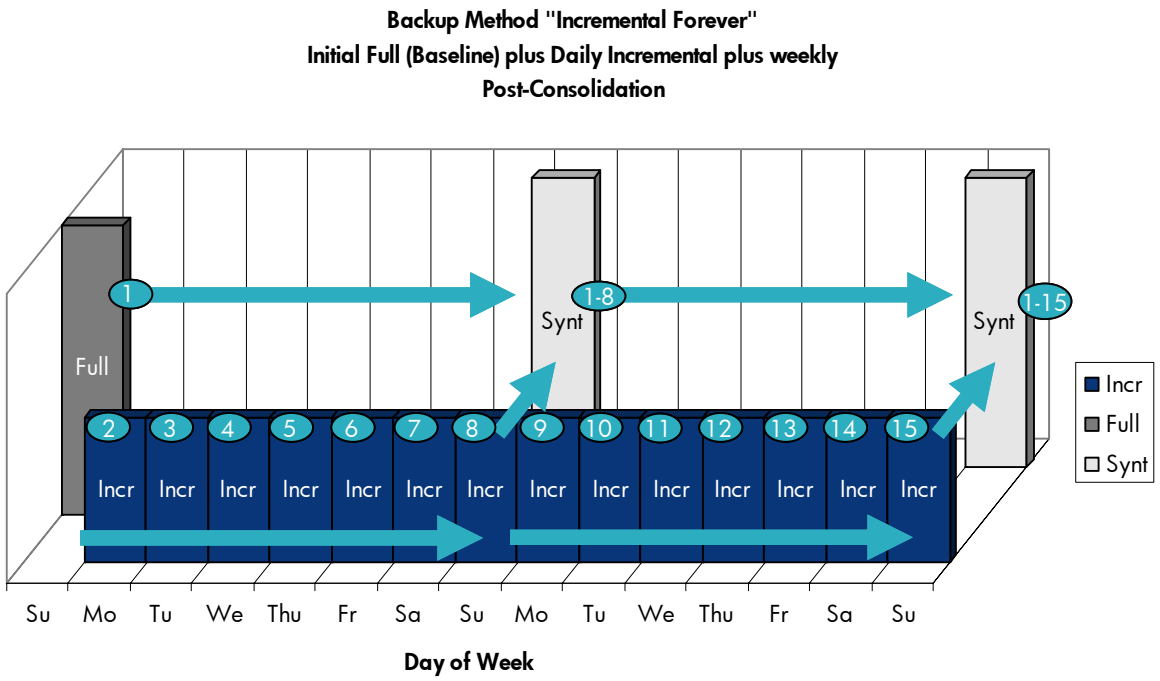
Use case #2: Daily file system incremental backup with weekly object consolidation.

This scenario describes a change from an existing weekly full backup with daily incremental backups in between to an incremental forever file system backup using one initial full backup with subsequent incremental daily backups. An object consolidation is performed automatically every week.

Usage: File Servers with high availability and fast response time requirements. High data change rate in file system. Short restore chain, short restore time requirement. Point-in-time restore possibility using incremental backups.

Figure 53 depicts the new backup strategy which is daily incremental backups plus weekly object consolidation.

Figure 53. Weekly Post-Consolidation



Note: The weekly full backup on Sundays was replaced by an incremental backup and a synthetic full backup.

The backup schedule for this scenario can be copied from use case #1. The protection period for the incremental backups should be increased to match desired protection to be able to consolidate objects every week.

A new scheduled backup consolidation setup will consolidate objects automatically on scheduling the object consolidation.

Scheduled object consolidation takes place at a user-defined time. It consolidates objects that match the specified criteria. Objects backed up during different backup sessions can be consolidated in a single scheduled object consolidation session.

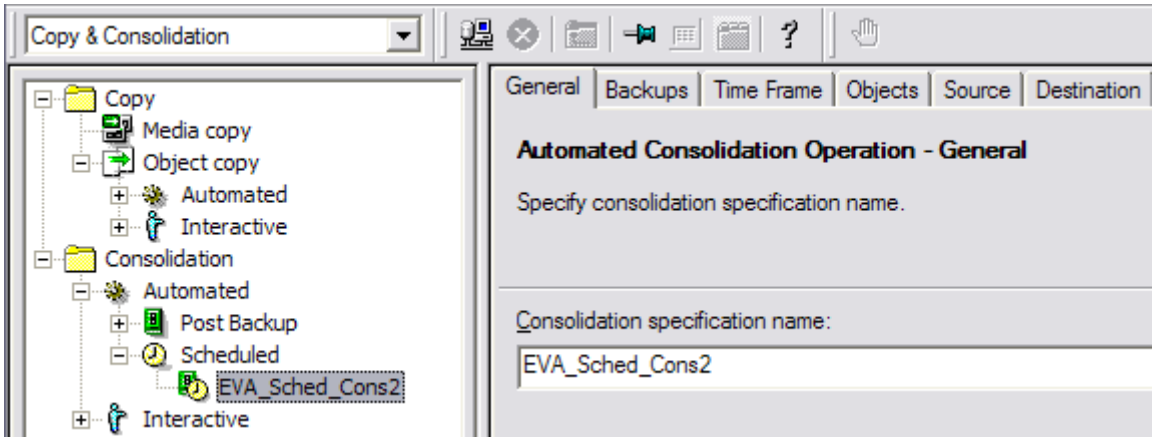
When there are many possible restore chains to select from, HP Data Protector software consolidates the one containing the object version with the latest point in time. For example, backup sessions: Full, Incr1, Incr2, Incr2, Incr2 result in three restore chains but HP Data Protector software consolidates only the one consisting of Full, Incr1, and the latest Incr2.

Configuration

Steps

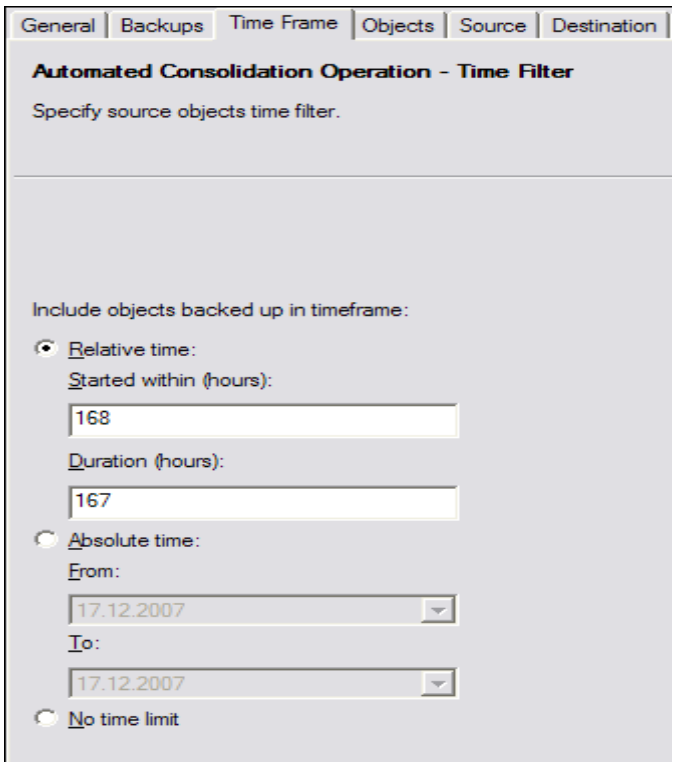
In the Context List, click **Copy & Consolidation**.

Figure 54. Scheduled Consolidation Context



1. In the Scoping Pane, expand **Consolidation**, and then expand **Automated**.
2. Right-click **Scheduled** and click **Add** to open the wizard.
3. In the Consolidation specification name text box, type a name for the object consolidation specification. Click **Next**.
4. Select the backup specifications that contain the objects you want to consolidate. Click **Next**.
5. Specify the time filter for the object consolidation operation. Only objects that were backed up in the specified time frame will be consolidated. Click **Next**.

Figure 55. Time Filter



Relative time
Select this option to set a relative period of time and then specify the time frame. The first number specifies the beginning of the time frame, and the second number the duration of the time frame. For example, if you specify 168 in the first field and 167 in the second field, and the operation is scheduled Sundays at 9 P.M., HP Data Protector software will consolidate objects from the sessions that took place between 9 P.M. a week ago and 8 P.M. the Sunday where the consolidation takes place. (An additional "Relative time" period example is further down in this document.)

Absolute time
Select this option to set an absolute period of time. Specify the starting and the end date of the period. Click the drop-down arrows to display the calendar.

No time limit
Select this option to include all sessions, regardless of when they were performed.

6. Specify the object filter for the object consolidation operation. Click **Next**.
7. Specify the devices that will read the incremental backups and the full backups.

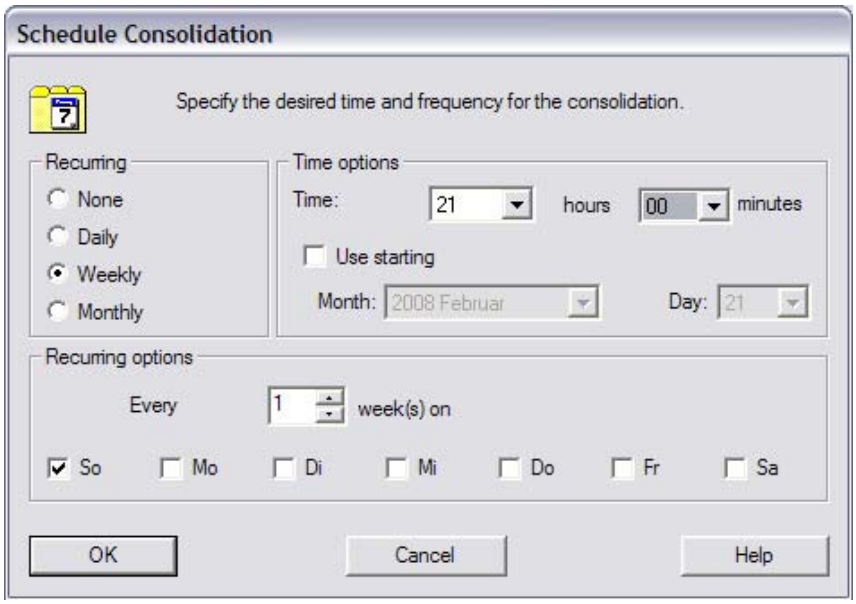
   Limit the object consolidation to specific file libraries by selecting these libraries as read devices for incremental backups. Only objects residing in the specified file libraries will be consolidated.

   By default, the read devices for full backups are those used for backup in the selected backup specifications. You can change them here if desired. Click **Next**.

8. Select the destination devices for the object consolidation operation. Data Protector will select the most suitable devices from those you specify here. Click **Next**.
9. Specify options such as data recycling and backup data protections desired. Click **Next**.

Right-click a date and click **Schedule** to display the Schedule Consolidation dialog box. Specify the options as desired and click **OK**.

Figure 56. Schedule Consolidation



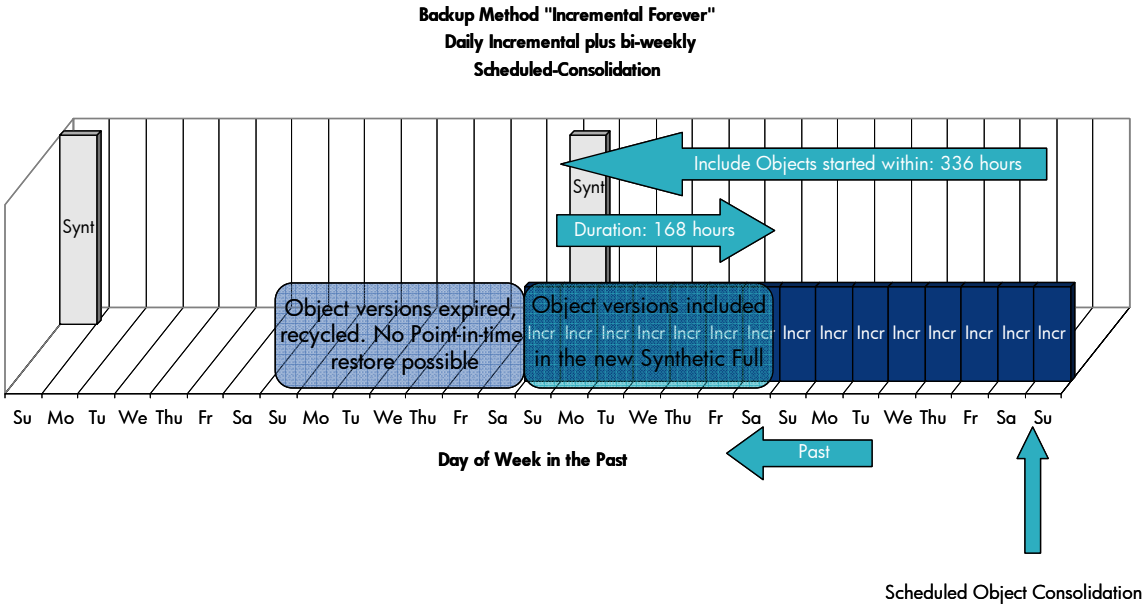
10. Click **Finish** to exit the wizard.

Additional "Relative time" period example:

In this, scheduled object consolidation objects from the week before the last week are consolidated into a synthetic full back up. Point-in-time restores from the past week remain available. Sources of object versions included into the synthetic full backup have been recycled and are no longer available for point-in-time restore operations. This operation frees up disk space and provides point-in-time restore possibilities as desired.

This method of object consolidation might be used where disk space might become a bottleneck on one side and point-in-time requirements are a necessity on the other side.

Figure 57. Relative time period example



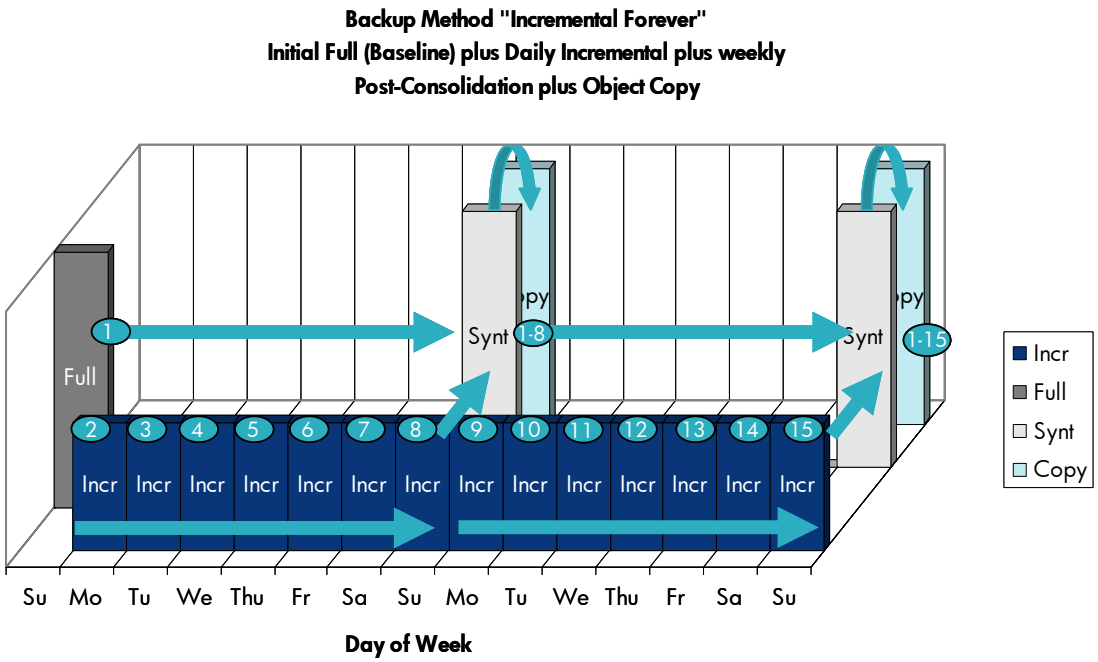
Use case #3: Daily incremental file system backup plus weekly object consolidation with object copy afterwards.

This scenario describes a change from an existing weekly full backup with daily incremental backups in between to an incremental forever file system backup using one initial full backup with subsequent incremental daily backups. An object consolidation is scheduled every week. For disk space clean up purposes an automated object copy operation is performed in which in addition incremental backup source file depots are recycled.

Usage: File Servers with high availability and fast response time requirements. High data change rate in file system. Short restore chain, short restore time requirement. Point-in-time restore possibility using incremental backups. Disk space constrains.

Figure 58 depicts the new backup strategy which is daily incremental backups plus weekly object consolidation plus a weekly object copy.

Figure 58. Weekly Post-Consolidation plus Post-Copy



The backup schedule for this scenario can be copied from use case #1. The protection period for the incremental backups should be increased to match the desired protection to be able to consolidate objects every week.

The object consolidation configuration of use case #2 can be copied.

A new automated post backup object copy configuration need to be generated.

The HP Data Protector software object copy functionality enables you to copy selected object versions to a specific media set. You can select object versions from one or several backup sessions or object consolidation sessions. During the object copy session, HP Data Protector software reads the backed up data from the source media, transfers the data, and writes it to the target media.

The result of an object copy session is a media set that contains copies of the object versions you specified.

The following characterizes the object copy functionality:

- **Start of session**: You can start an object copy session interactively or automatically.
- **Selection of media**: You can use original media sets containing backups, media sets containing object copies, or media sets that are media copies as source media.
  However, the selection of the media set is not possible after the start of the object copy session.
  In case of a mount request, you need to provide the specific medium that is requested by HP Data Protector software, or its identical copy (created using the media copy functionality).

- **Media type**: You can copy objects to media of a different type. Furthermore, the block size of the destination device can be the same or larger than the block size of the source device.
- **Media policy**: You can append data to media already containing backups or object copies.

You can start an object copy session interactively or specify an automated start of the session.

Automated object copying

In an automated object copy specification, you can specify one or more criteria for the selection of object versions that will be copied:

- **Backup specifications**: To copy only object versions backed up using specific backup specifications.
- **Data protection**: To copy only protected object versions.
- **Number of existing copies**: To copy only object versions that do not have more than the specified number of successful copies.
- **Libraries**: To copy only object versions located on the media in the specified libraries.
- **Time frame (only in a scheduled object copy specification)**: To copy only object versions backed up in the specified period of time.

HP Data Protector software offers two types of automated object copying: post-backup object copying and scheduled object copying.

Post-backup object copying

Post-backup object copying takes place after the completion of a session that is specified in the automated object copy specification. It copies objects selected according to the automated object copy specification that were written in that particular session.

Scheduled object copying

Scheduled object copying takes place at a user-defined time. Objects from different sessions can be copied in a single scheduled object copy session.

Configuration

Post-backup object copying takes place after the completion of a backup session or object consolidation session that is specified by the name of the backup or object consolidation specification in the automated object copy specification. It copies objects from that particular session that match the specified criteria.

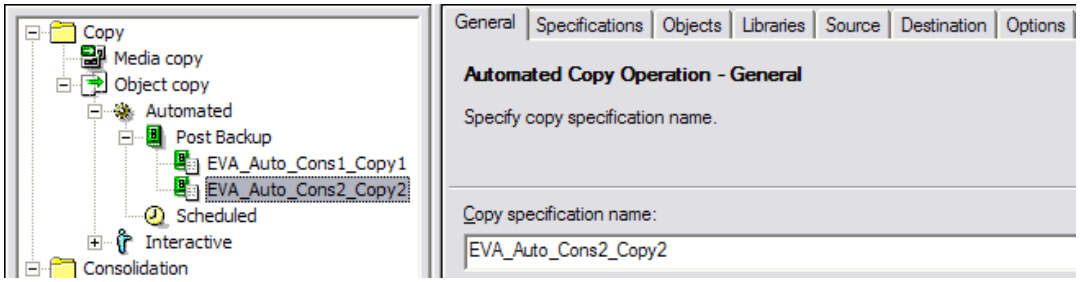
A post-backup object copy session does not start if the backup session failed. If the backup session has been aborted but contains completed objects, a post-backup object copy session copies the completed objects by default. To disable the copying of aborted sessions, set the `CopyStartPostBackupOnAbortedSession` global option to `0`.

Steps

1. In the Context List, click **Copy & Consolidation**.
2. In the Scoping Pane, expand **Copy**, then **Object copy**, and then **Automated**.
3. Right-click **Post Backup** and click **Add** to open the wizard.

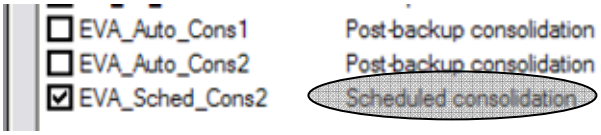
In the Copy specification name text box, type a name for the object copy specification. Click **Next**.

Figure 59. Automated Copy Operation Dialog



Select the specifications that contain the objects you want to copy. Click **Next.**

Figure 60. Select specification



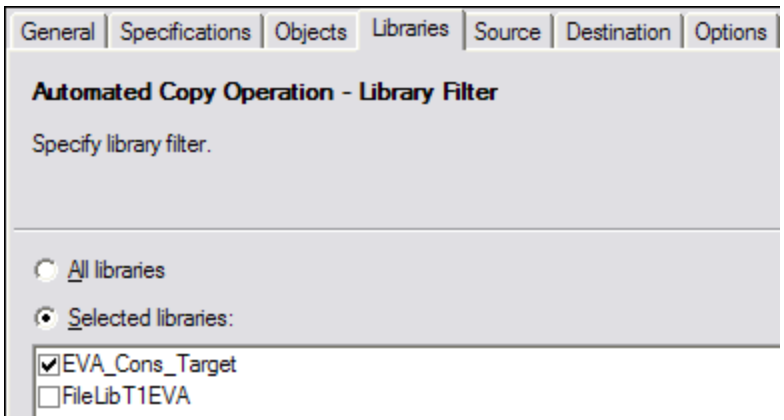
4. Specify the object filter for the object copy operation. Only objects that match the specified criteria will be copied. Click **Next**.

   Option: Include only protected objects
   Select this option to copy only objects with data protection.

   Option: Include only objects with number of copies less than
   Select this option to copy only object that do not have more than a certain number of copies. Specify the number. The maximum is 10.

5. Specify the library filter for the object copy operation. Only objects residing on media in the specified libraries will be copied. Click **Next**.

Figure 61. Library Filter

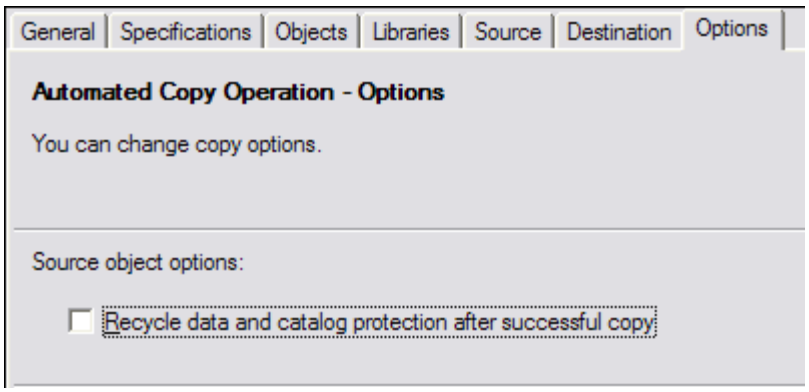


6. The devices used for backup in the selected backup specifications are used as source devices in the object copy operation by default. You can change the source devices here if desired. Click **Next**.
7. Select the destination devices for the object copy operation. HP Data Protector software will select the most suitable devices from those you specify here. Click **Next**.

Figure 62. Copy Operation destination device

8. Specify the source object options, target object options, and target media options as desired.

Figure 63. Copy Operation Options



For further disk space savings the "Recycle data.." check box might be selected here.

Note: In this example, consolidated backups are transferred to tape and are no longer available for fast restore as the tape might be vaulted offsite.

If a tape gets a new "location" identifier the priority of media selection might get changed as well. This needs to be considered for eventual restores, copy, or consolidation tasks.

Virtual full backups, when copied to tape, will cause data blocks to be transferred to tape rather than links. This needs to be considered while planning tape capacities.

Click **Finish** to exit the wizard.

Backup Results of Use Case #3

The restore context will provide the following restore session information for the scenario discussed previously.

Figure 64. Restore session information

Name	Status	Backup Specification	Backup Type	Start Time
2007/12/20-...	Completed	EVA_Auto_Cons2_Copy2	full (copy)	20.12.2007 14:01:39
2007/12/20-...	Completed	EVA_Sched_Cons2	full	20.12.2007 14:00:06
2007/12/20-...	Completed	eva20writer1con-2	incr	20.12.2007 14:00:05
2007/12/20-...	Completed	eva20writer1con-2	incr	20.12.2007 13:45:05
2007/12/20-...	Completed	eva20writer1con-2	incr	20.12.2007 13:30:05
2007/12/20-...	Completed	eva20writer1con-2	incr	20.12.2007 13:15:05
2007/12/20-...	Completed	eva20writer1con-2	incr	20.12.2007 13:00:05
2007/12/20-...	Completed	eva20writer1con-2	incr	20.12.2007 12:45:05
2007/12/20-...	Completed	eva20writer1con-2	incr	20.12.2007 12:30:06
2007/12/20-...	Completed	eva20writer1con-2	incr	20.12.2007 12:15:07
2007/12/20-...	Failed	EVA_Sched_Cons2	full	20.12.2007 12:00:06
2007/12/20-...	Completed	eva20writer1con-2	incr	20.12.2007 12:00:06

The session status for each session is as described in use case #1.

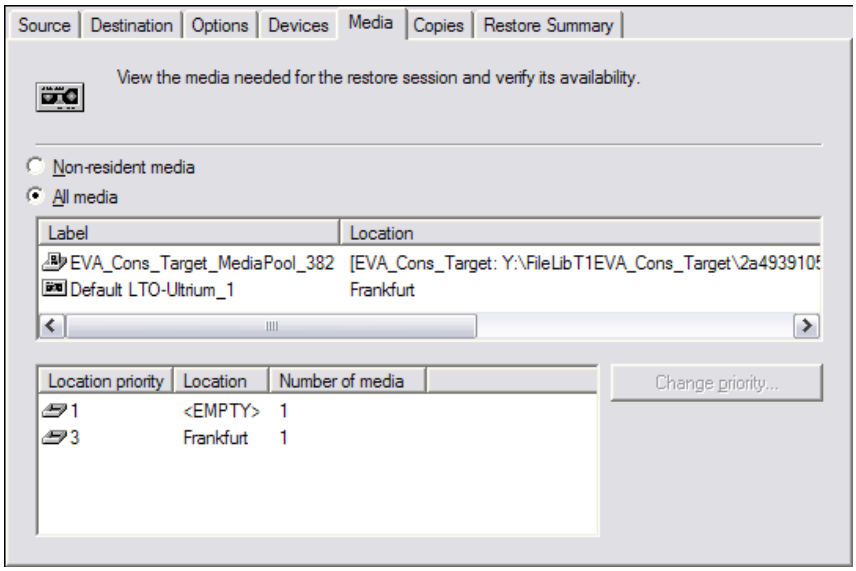
The top entry in Figure 64 shows the object copy session which is also marked as “full (copy)” backup type.

A copy process means that the source object(s) need to be read (similar to a restore operation) and written (similar to a backup operation) to a new media. This need to be considered as the associated devices for that copy process will be locked and unavailable to other planned tasks.

In the media tab of the restore objects context, consolidated objects reside on multiple media now.

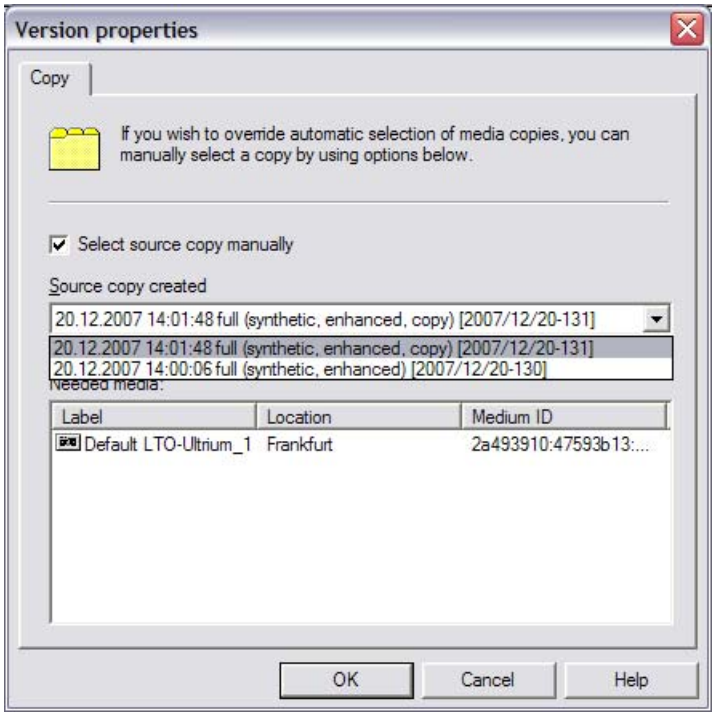
Note: The <EMPTY> location was changed to location priority 1. Priority 1 is the highest priority. This setting ensures that a restore operation for an object existing on either media will be restored from the media with the higher priority.

Figure 65. Media location and priority



If circumstances require a restore from the copy, one can select the copy in the copy version properties dialog.

Figure 66. Copy Versions Property Dialog



The copy media needs to be available (mounted) to avoid potential mount requests.

Glossary

- Cell Manager—The main system in the cell where the essential software for HP Data Protector software is installed and from which all backup and restore activities are managed.
- Concurrency—The number of Disk Agents started for each Media Agent is called Disk Agent (backup) concurrency
- Disk Agent (DA)—Disk Agent is a HP Data Protector software component needed on a client to back it up and restore it. The Disk Agent controls reading from and writing to a disk. During a backup session, the Disk Agent reads data from a disk and sends it to the Media Agent, which then moves it to the device. During a restore session, the Disk Agent receives data from the Media Agent and writes it to the disk.
- EVA—The HP StorageWorks Enterprise Virtual Arrays are the mid-range to enterprise leading high performance, high capacity and high availability "virtual" array storage solutions.
- GUI (graphical user interface)—A HP Data Protector software–provided GUI is a cross-platform (HP-UX, Solaris, Windows) graphical user interface, for easy access to all configuration, administration, and operation tasks.
- IDB—The HP Data Protector software IDB is an internal database, located on the Cell Manager, that keeps information regarding what data is backed up, on which media it resides, the result of backup, restore, copy, object consolidation, and media management sessions, and which devices and libraries are configured.
- LAN (local area network)—A computer network covering a small geographic area, like a home, office, or group of buildings.
- Load Balancing—HP Data Protector software automatically balances the usage of devices selected for backup, so that they are used evenly. Load balancing optimizes the device usage by balancing the number of objects written to each device.
- Media Agent (MA, BMA, RMA)—A HP Data Protector software process that controls reading from and writing to a device, which reads from or writes to a medium (typically a tape). During a backup session, a Media Agent receives data from the Disk Agent and sends it to the device for writing it to the medium. During a restore session, a Media Agent locates data on the backup medium and sends it to the Disk Agent. The Disk Agent then writes the data to the disk. A Media Agent also manages the robotics control of a library.
- SAN (storage area network)—A SAN is an architecture to attach remote computer storage devices such as disk array controllers and tape libraries to servers in such a way that to the operating system the devices appear as locally attached devices.

For more information

- Disk-Assisted Backup Whitepaper
- HP Data Protector software Performance White Paper
- HP Data Protector software Advanced Backup to Disk Performance White Paper

Visit:
www.hp.com/go/dataprotector

4AA1-8956ENW, March 2008

invent