

Disk-assisted backup white paper

Backup to disk and how to implement it with HP Data Protector software



Overview	2
Reduced recovery time	2
Improved backup performance	2
Increased success rate of backups completed	3
No or reduced need for multiplexing of backups to tape	3
Branch office backup consolidation	4
Object consolidation	5
Backup to disk implementation with HP Data Protector software	6
Overview	6
Synthetic full backup and virtual full backup	6
Synthetic full backup	6
Virtual full backup	6
How does virtual full position against synthetic full?	7
How HP Data Protector software synthetic backup works	7
How HP Data Protector software synthetic backup restore works	8
Distributed file media format (DFMF)	9
HP Data Protector software object consolidation prerequisites	9
Command Line details	10
Object consolidation details in HP Data Protector software GUI	11
HP Data Protector software File Library and Object Copy	14
HP Data Protector software Jukebox and Standalone file device	25
For more information	30

Overview

Advancements in technology and price reductions have recently made it possible to use disk together with tape to improve data protection. Some of the benefits that the technology promises are:

- Reduced recovery time
- Improved backup performance
- Increased success rate of backups completed
- No or reduced need for multiplexing of backups to tape
- Branch office backup consolidation

With shrinking backup windows and increasing amounts of data, users are adding more tape drives and implementing faster backup network topologies, including storage area networks (SANs). But for cost reasons, and in distributed environments, users cannot have every host on the SAN. Nor can they implement a direct-attached tape device to every host. The administration and ongoing management would be difficult and inefficient. In this type of environment, backup to disk can help. Backup to disk is essentially the process of backing up data to a disk device rather than to tape. Backups are then copied to tape for long-term storage and disaster recovery.

Combining disk and tape for backup is not necessarily a cheaper solution than using tape only, and to completely stop running backups to tape may be a risk. In any case, it is well worth the time and effort to investigate how using disk in addition to tape could potentially give you a better backup and restore solution.

Reduced recovery time

To recover a file from tape, the tape first has to be mounted and searched for the file. As tape is a sequential media, this can take several minutes depending on the technology used. To recover a file from disk is only going to take as long as it takes to copy the data over the network to the destination disk. The search time is negligible.

The tape solution will be faster when recovering larger amounts of data as tape drives can deliver more data faster than disk drives. This premise holds true as long as all data is on one tape only and not spread out. If the recovery requires the restore of a full backup and several incremental backups, it may actually be beneficial to hold the incremental backups on disk and let the backup and restore system restore the full backup from tape and the incremental data from disk.

To reduce the recovery time even further, you can hold clone or snapshot copies of data on disk. A clone is a full copy of the data you want to restore, whereas a snapshot copy only holds changed data. Restoration can be instantaneous when using snapshot copies or clones because you only have to mount the copy on the application server, rather than the primary data volume, and no data has to be copied at all during restore.

HP uses the term “Business Copies” for these data snapshots. The easiest way to manage recovery from a Business Copy is to use HP Data Protector software with the Instant Recovery option. This automatically detects and uses the Business Copy, and the administrator can use the standard HP Data Protector software interface for instant recovery of applications.

Improved backup performance

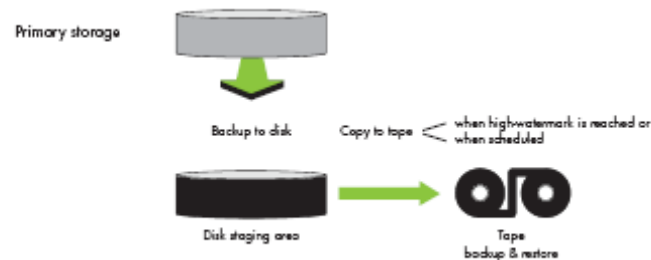
The backup window can be reduced to a minimum by taking a snapshot of the application data and moving it off-host for backup. The solution is called Zero Downtime Backup and uses either snapshots or clones of data in the disk arrays in combination with HP Data Protector software.

The backup window can also be reduced by using backup to disk, running many backups in parallel. This is especially attractive for incremental backups due to the following reason:

Tape backups are fast as long as the tape drives are streaming, that is, enough data is being sent to the drive so it does not have to stop and then re-start when it receives more data. With the ever-increasing performance of new tape drives, the requirement on the infrastructure for backups is increasing, too—faster networks, faster disks and faster backplanes. Especially when running incremental backups, many parallel streams are required to feed enough data to the drives. Even though it is copying less data than a full backup, the incremental backup can take longer to complete due to slow disks on the client side and non-streaming drives.

To resolve that problem, backup to disk can be used as a staging area. All incremental backups are stored on disk and can then be written to tape later in one pass, without delay, as all the data is in one place.

Figure 1.



Increased success rate of backups completed

Using disk as a staging area can prevent backups from failing to complete due to:

- Tape drive failure
- Backups “locking” the tape drive for longer than expected and preventing other backups from running
- Restoration jobs running in parallel

As backup to disk is not constrained to a limited set of tape drives, it is easier to accommodate changes in the backup load and the number of parallel jobs running. This lowers the risk of backup failure caused by tape drives not being available or drives being tied up with backup or restore jobs that are taking longer than expected.

No or reduced need for multiplexing of backups to tape

Multiplexing is a very useful feature that allows multiple backup streams to be written simultaneously to a tape. As mentioned previously, modern tape drives are so fast that it is almost mandatory to use multiplexing to guarantee enough data being sent to the drive to keep it running at the highest possible speed.

Multiplexing also has some drawbacks:

- Data from multiple clients and/or applications is mixed on the same tape, hence you cannot accurately identify the data that the tape holds (for example, it would be difficult to say that your Exchange data is on one tape, or that another tape is the backup of server A).
- Restoring data from a multiplexed tape can be slower because the data blocks of a particular client or application will not be aligned on the tape. Therefore, the tape drive has to search for the next block on tape.

When backing up to disk and then copying the data to tape, there is no need to multiplex the data streams as you are less dependent on a backup window and your backup design should allow fast access between backup disk and tape.

A common backup method will most likely be tape, at least for servers that have directly attached tape drives or share drives on a SAN, and incremental backups to disk. This will reduce the need for multiplexing and also optimize the number of tape drives needed.

Branch office backup consolidation

A major trend is to consolidate backups by bringing the data from remote or branch offices to a central site. This offers many benefits:

- Increased control of critical data by storing a copy centrally
- Reduced reliance on non-IT personnel in remote sites to perform backup operations
- Cost efficiency, with the recent price drop in high-speed long-distance communication links
- Improved disaster-recovery capabilities

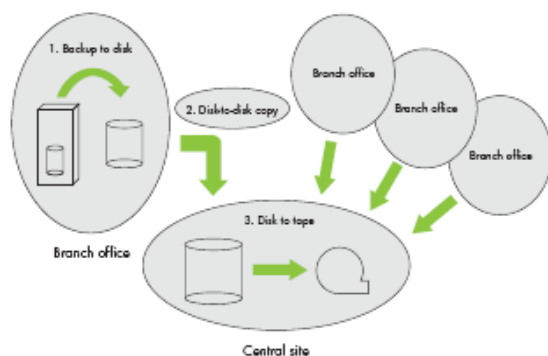
It is still common to perform local backups to direct-attached tape drives on file and print servers in local offices. With backup consolidation, these backups can be replaced either with replication of data to a central site for backup, for example, with HP OpenView Storage Mirroring software, or with local backups to disk followed by a second copy sent to a central site. The latter can be fully automated by the backup and restore software.

HP Data Protector software supports fully automated backup consolidation using backup to disk. It works in the following way:

1. In the branch or remote office, backups are performed to a local disk, which can be either internal or external. Typically, these backups are stored for one to two weeks. As the backup runs locally, it can be very efficient and quick and reduces the backup window needed.
2. After a completed backup, HP Data Protector software will copy the backup files over the network to a backup server at the central site. This operation runs in the background and only accesses backup data, not production data. At the central site, the backup data can be stored longer on disk, if required to accommodate fast file restoration.
3. The third step is to copy the backups to tape, which is done from the central backup disk. Typically, the central backup and restore environment is designed to provide high-speed backup to tape from disk. When backups have expired on disk, data can still be restored from tape directly to the client over the network.

For disaster recovery, HP Data Protector software includes a feature called Disk Delivery. This restores all server data to a disk centrally, then sends the disk to the remote office, replaces the faulty system disk, and reboots from the newly restored drive. Especially useful for remote sites, Disk Delivery can simplify restoration if the network does not have the necessary bandwidth for a full server restore.

Figure 2.



Object consolidation

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 few changed data is backed up.

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

Object consolidation removes this drawback of performing regular full backups. Object consolidation functionality enables you to merge a restore chain of a backup object, consisting of a full backup or synthetic full and at least one incremental backup, into a new, consolidated version of this object, either a synthetic full backup or a virtual full backup. During an 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.

Several advantages of object consolidation backups include:

- Space efficient—One initial full backup only; no file stored multiple times; potential to reduce required disk space and consequently tape media
- Faster backup—Only changed files need to be backed up; fast backup of file servers hosting millions of files since only incremental backup is required; allowing more frequent backup sessions
- Less time to manage—Due backup to disk no tedious tape management
- Faster restore of individual files—Single file restore from disk; tape goes off-site
- Synthetic full provides fast data restore in case of disaster recovery

Synthetic backup is an advanced backup solution that creates a synthetic full backup, which is an equivalent to a conventional full backup in terms of data, without putting stress on the production servers or the network.

In terms of restore speed, a synthetic full backup is equivalent to a conventional full backup. The restore chain consists of only one media set, thus allowing a quick and simple restore.

Backup to disk implementation with HP Data Protector software

Overview

In this second part of the document, it is detailed how HP Data Protector software can be configured and administered for backup to disk. A new device type, called File Library, is available with HP Data Protector software that provides advanced backup-to-disk functionality. This new enhancement is covered in detail, and some disk backup technologies that were already introduced in previous HP Data Protector software versions are explained.

This paper focuses on backup to disk but does not cover the data snapshot technologies that are also available with HP Data Protector software, namely Zero Downtime Backup and Instant Recovery. The three backup-to-disk technologies of HP Data Protector software that are covered include:

- HP Data Protector software synthetic backup and virtual full backup
- HP Data Protector software File Library and Object Copy
- HP Data Protector software Jukebox or Standalone file device

Synthetic full backup and virtual full backup

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

Synthetic full backup

Synthetic backup is a backup solution that eliminates the need to run regular full backups. Instead, incremental backups are run, and subsequently merged with the full backup into a new, synthetic full backup. This can be repeated indefinitely, with no need to run a full backup again. In terms of restore speed, such a backup is equivalent to a conventional full backup.

Virtual full backup

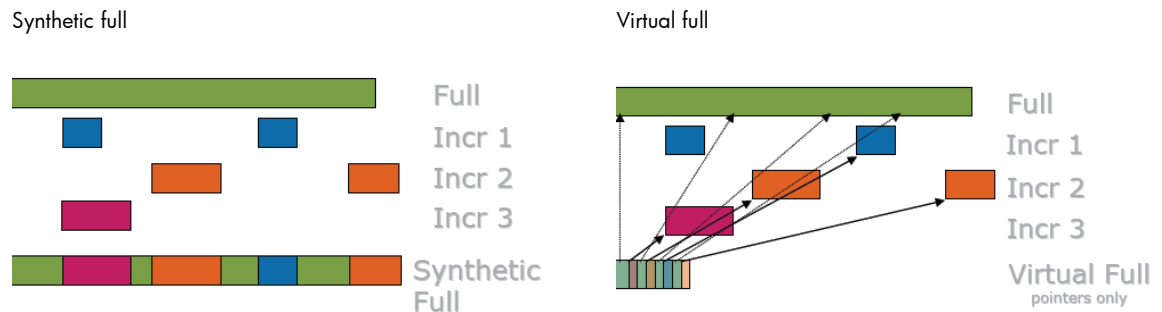
Unique in the market, virtual full backup is an even more efficient type of synthetic backup. This solution uses pointers to consolidate data rather than copy the data. As a result, the consolidation takes less time and avoids unnecessary duplication of data.

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

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

How does virtual full position against synthetic full?

Figure 3.



	Synthetic full	Virtual full
Space efficient full	No	Yes
Storage type	Disk and Tape	Disk only
Competitors	Yes	No
HP HP Data Protector software	Yes	Yes

How HP Data Protector software synthetic backup works

HP Data Protector software synthetic backup enables you to merge a full backup (or a synthetic full backup) and any subsequent incremental backup into a new, synthetic full backup. A synthetic full backup can be created from a full backup (or a synthetic full backup) that is written to a disk or tape device and incremental backups that are written to a disk-based device, called HP Data Protector software File Library. The synthetic full backup can, again, be written to a disk or tape device.

If all the backups, full and incremental, are written to the same File Library, an even more efficient type of synthetic backup is available, called virtual full 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.

The following figures explain the concept of synthetic backup and virtual full backup. They show how a synthetic full backup or a virtual full backup is created from a full backup and any number of incremental backups.

Figure 4.

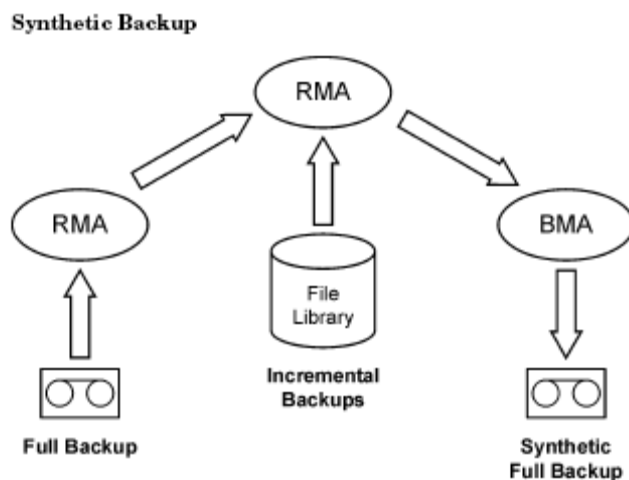


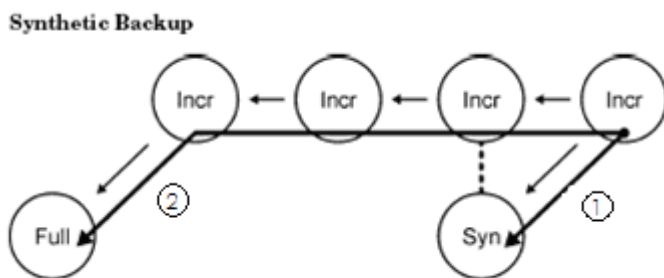
Figure 4 shows how a synthetic full backup is created. The Restore Media Agent (RMA) reads the full backup from the backup medium, which can be a tape or a disk. The data is sent to another RMA, which reads the incremental backups from the File Library and consolidates the data. The consolidated data is then sent to the Backup Media Agent (BMA), which writes the synthetic full backup to the backup medium. The backup media can be a tape or a disk.

Later on, the synthetic full backup is typically merged with subsequent incremental backups into a new synthetic backup. The procedure can be repeated indefinitely, either after each incremental backup, or at a desired interval.

How HP Data Protector software synthetic backup restore works

Restore from a synthetic full backup is equivalent to restore from a conventional full backup. The following figures present different situations, supposing you need to restore your data to the latest possible state. In all examples, a full backup and four incremental backups of the backup object exist. The difference is in the use of synthetic backup.

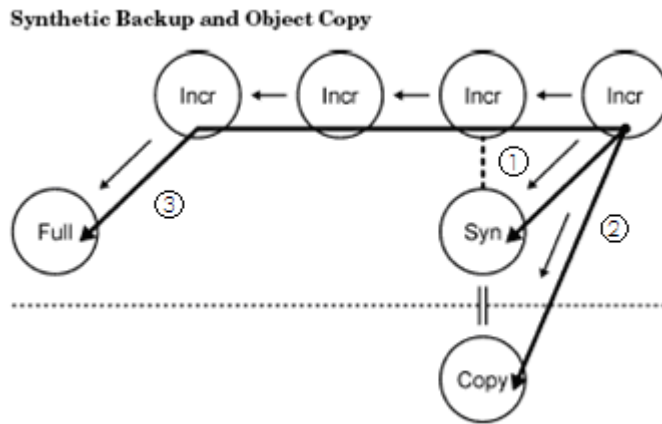
Figure 5.



In Figure 5, a synthetic full backup exists, which is used for restore by default. The restore chain consists of only two sessions, namely the synthetic full backup and the subsequent incremental

backup. The restore is significantly simpler and quicker than that without the synthetic full backup. In the figure, both possible (1+2) restore chains are shown.

Figure 6.



In Figure 6, a synthetic backup was performed and then copied. This provides additional safety. The restore to the latest possible state can use any of the three different (1, 2, or 3) restore chains shown. By default, HP Data Protector software selects the optimum restore chain, which normally includes the synthetic full backup or its copy.

Distributed file media format (DFMF)

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

Without this format HP Data Protector software writes all data and catalog segments into one file. This is done per session; hence each session creates its own file. With the new media format, pure data blocks are written into different files. This is done for each file, bigger than the used block size (default 64 KB); hence for each backed up file, a dedicated file on the File Library is created, which holds the data blocks.

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 tries to reuse those files; hence instead of copying the data blocks, they are only referred by way of pointers. Therefore consolidation sessions, creating virtual full, 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 are using pointers; normal backups, both full and incremental, are always creating new data block files.

HP Data Protector software object consolidation prerequisites

The following prerequisites must be fulfilled to perform a synthetic or virtual full backup in HP Data Protector software :

- All the backups that will be consolidated were performed with the enhanced incremental backup option enabled.
- All incremental backups that will be consolidated reside in the same File Library.

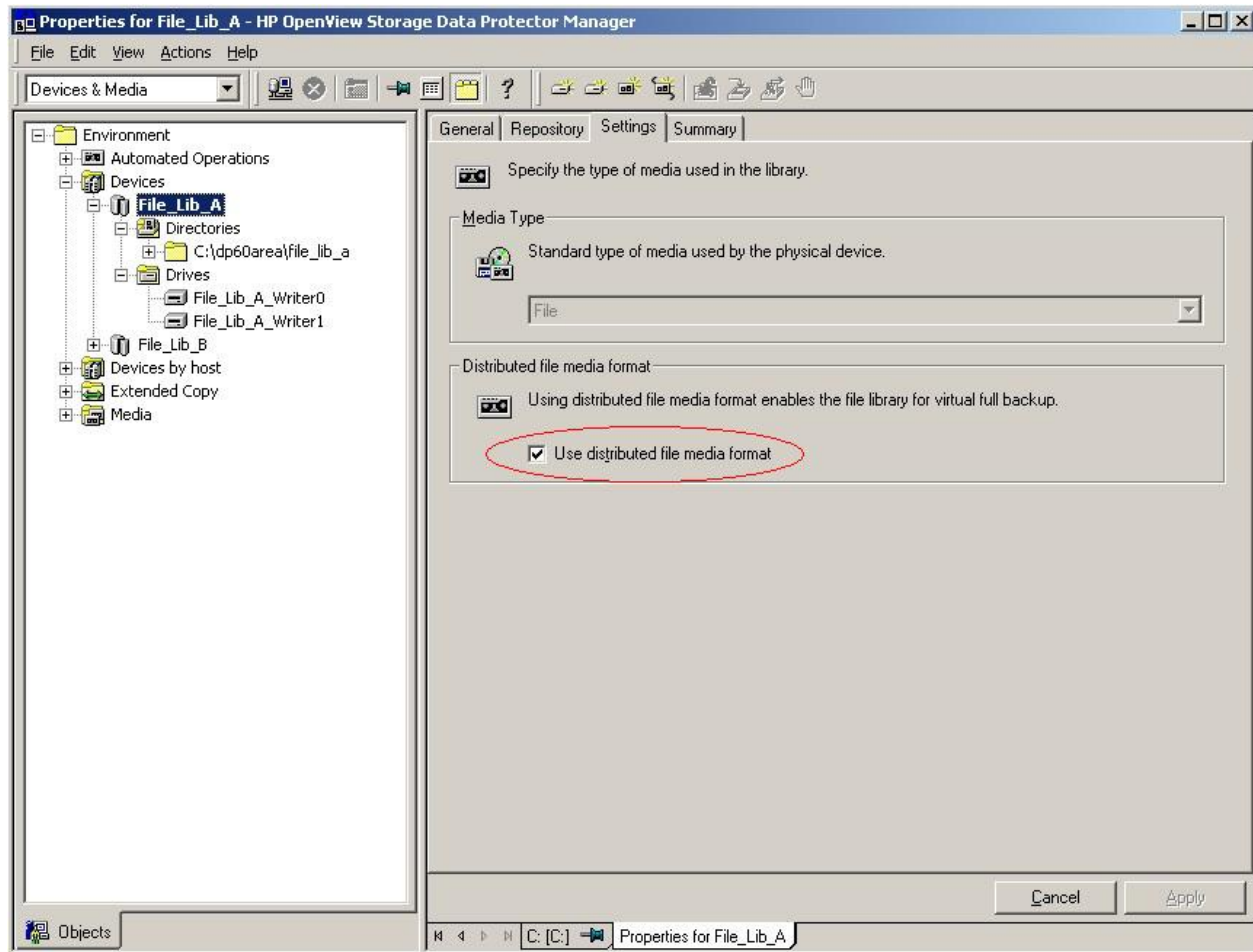
- The restore chain must be complete, meaning that all the object versions that comprise it have the status Completed or Completed/Errors and all the media holding these object versions are available.
- The necessary backup devices are configured and the media prepared.
- You need a Media Agent installed on every system that will participate in an object consolidation session.
- You need appropriate user rights for starting an object consolidation session (Start backup specification).
- To perform a virtual full backup, all the backups source (full, incremental) and target (virtual full) must reside in the same File Library that uses distributed file medium format.

Command Line details

- `omniobjconsolidate -version | -help`
- `omniobjconsolidate --consolidationlist ConsolidationSpecificationName -scheduled [GeneralOptions]`
- `omniobjconsolidate --consolidationlist ConsolidationSpecificationName -postbackup -session SessionID [GeneralOptions]`
- `omniobjconsolidate [GeneralOptions] [Device] ... Object [Object]...`

Object consolidation details in HP Data Protector software GUI

Figure 7. Defining a file device library with DFMF enabled



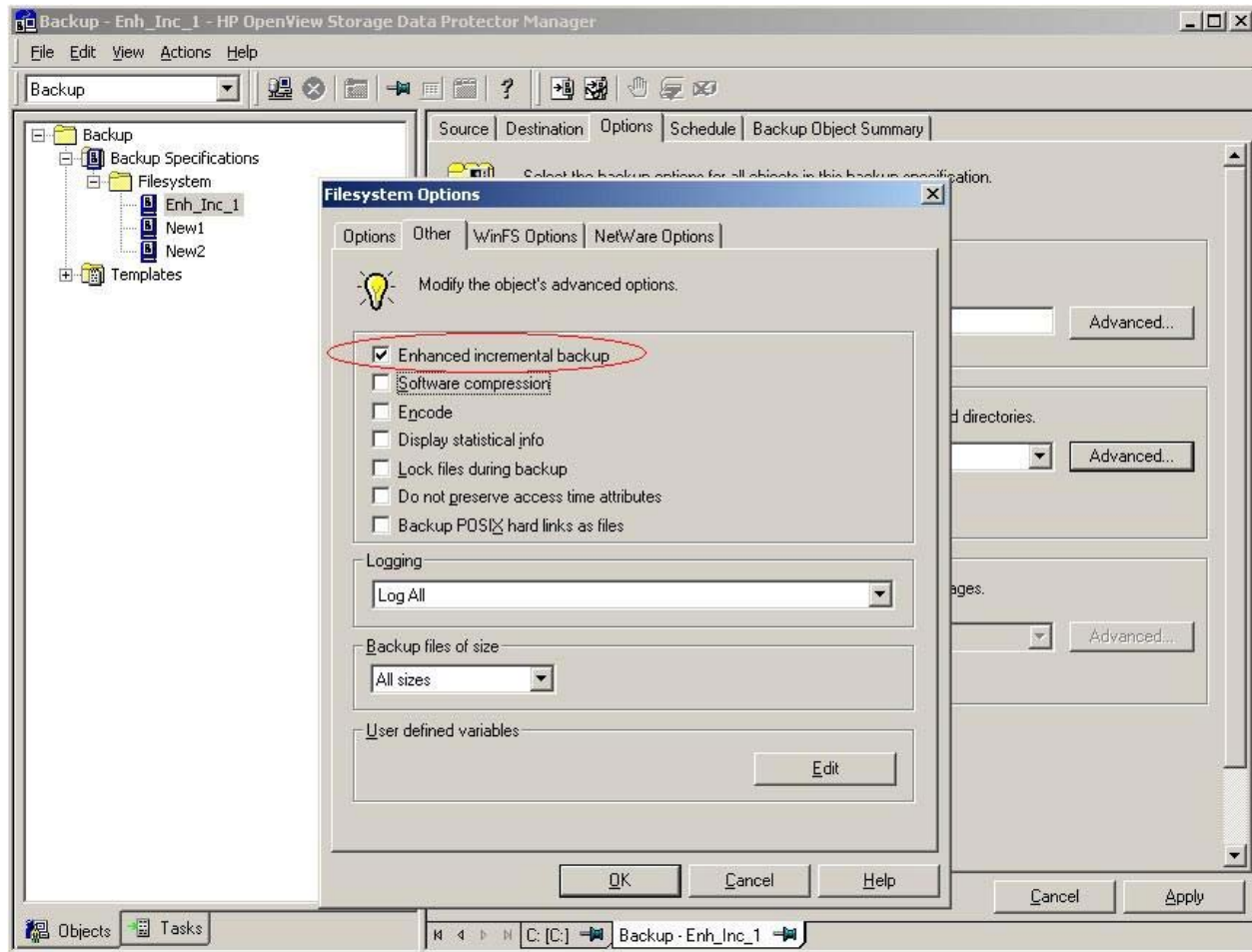
Select this option to enable the File Library for virtual full backup. Backing up to a File Library with distributed file media format is a prerequisite for virtual full backup.

If you are not planning to use virtual full backup, do not select this option.

Note

Media that use distributed file media format cannot be exported or imported.

Figure 8. Defining a backup specification with enhanced incremental backup enabled



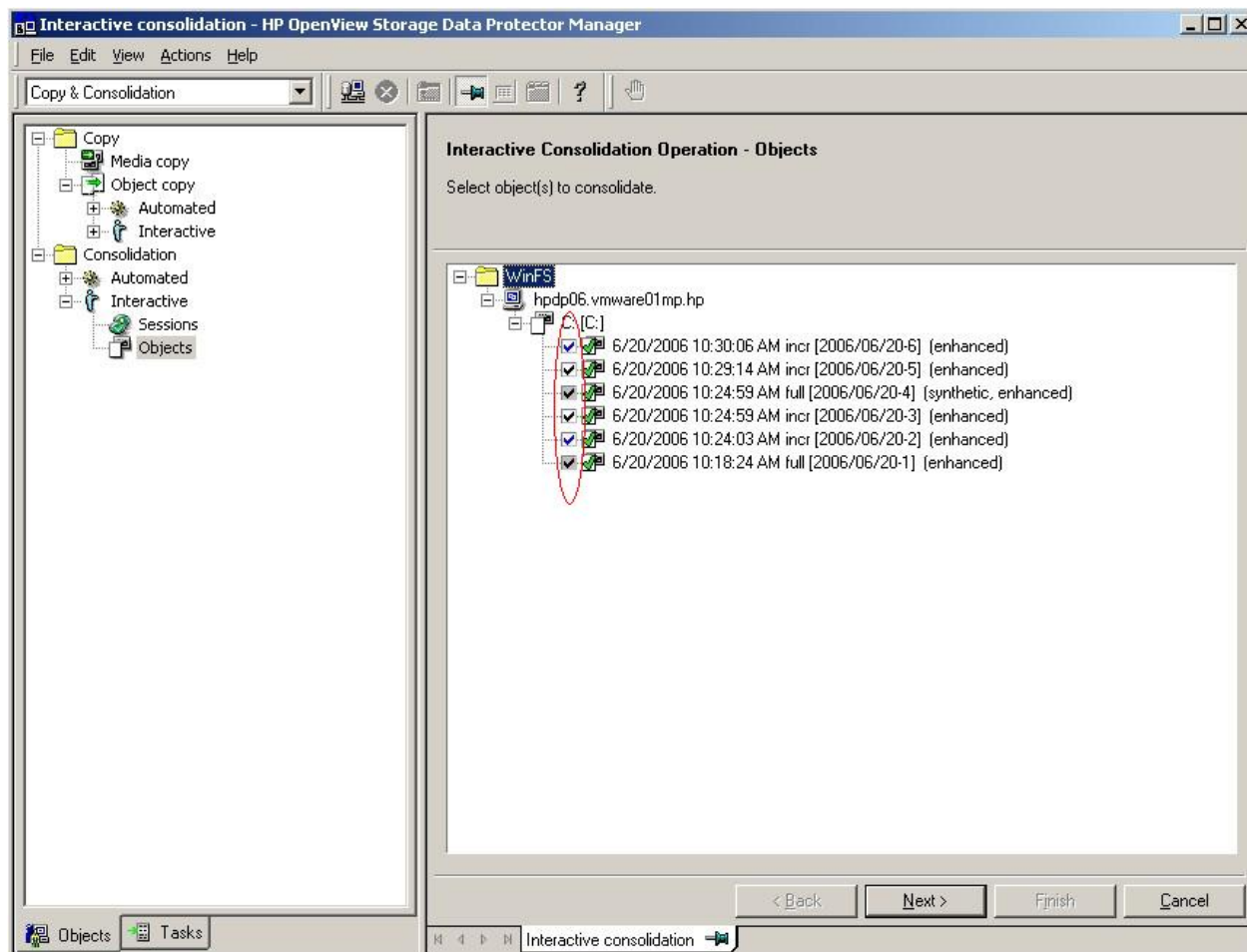
Select this option to enable enhanced incremental backup. Unlike conventional incremental backup, enhanced incremental backup reliably detects and backs up files with changes in name, location, and attributes. It is also a prerequisite for subsequent object consolidation (synthetic backup).

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, although an incremental backup is scheduled next, set the global variable `EnhIncrPromoteToFullIfNoEnhFull` to 1, which will promote the next backup to a full backup.

Figure 9. Create an interactive object consolidation after various backups of different kinds



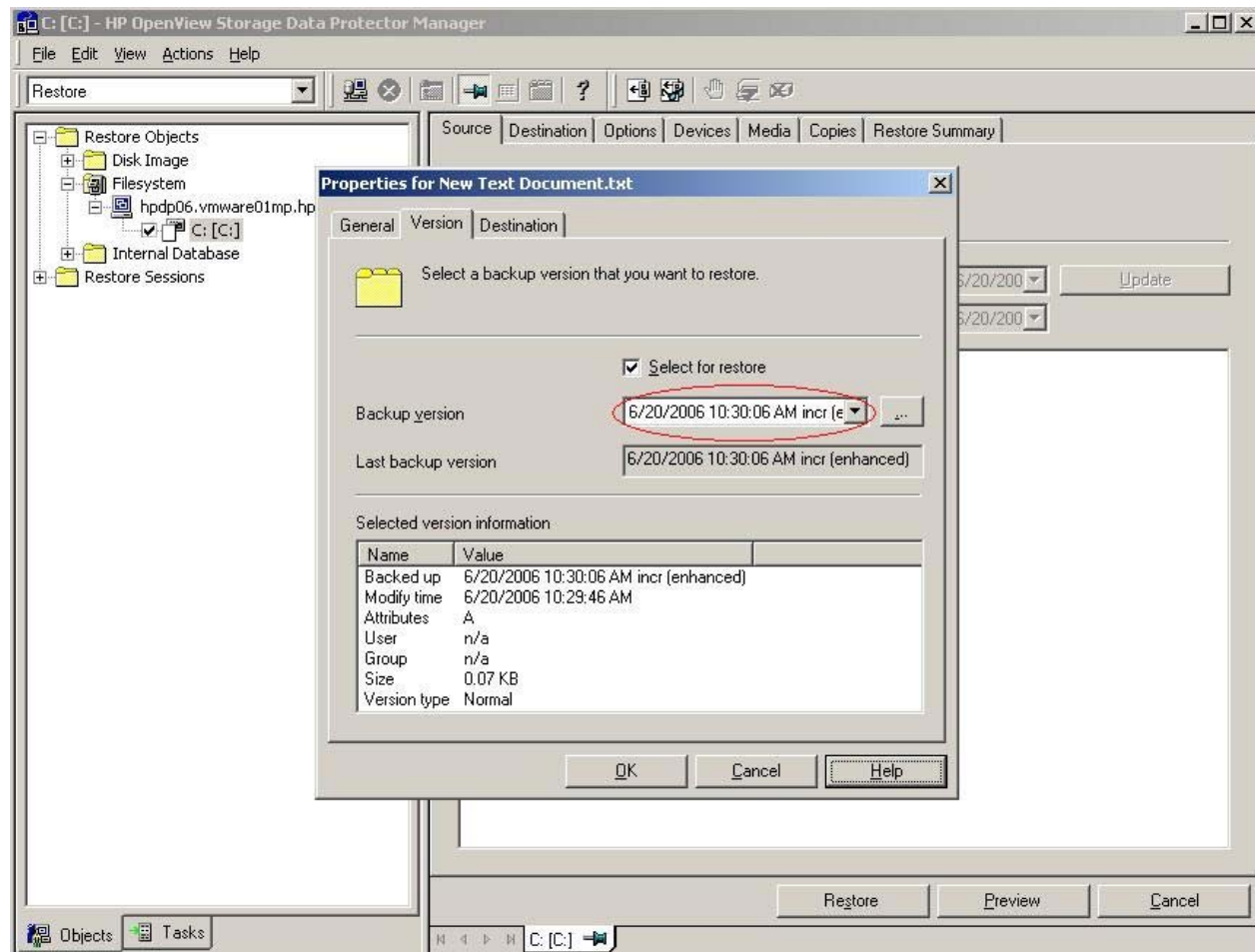
Select the points in time of the desired objects to consolidate. You cannot select full backups, because they as such cannot be consolidated.

Selecting a point in time selects the entire restore chain. If several restore chains for the same point in time exist, all of them are selected, but only one will actually be used. Your selection is marked in blue, other incrementals that comprise the restore chain are marked in black, and the corresponding full backup in gray (shaded). The blue check mark indicates the point in time that will be consolidated.

You can select several points in time for consolidation, and the restore chains may overlap. If you select a point in time that already has a black check mark, the check mark will become blue.

To clear a selected restore chain, click the blue check mark. The entire restore chain is cleared, unless some object versions are part of another restore chain, in which case they remain selected with a black check mark.

Figure 10. To restore a specific version select a backup version of a file or directory. By default, the most recent backup version is restored.



Click "... " for more information on each version.

HP Data Protector software File Library and Object Copy

HP Data Protector software Advanced Backup Disk provides a disk-staging solution, with single file restore directly from disk or tape.

Cheap disk arrays, especially the current SATA-based HP disk array, are positioned mainly as backup devices, and are the key focus for File Library devices. Should the "File Device Lib" be running out of free space, new backup capacity can be assigned automatically.

With the new HP Data Protector software File Library, you define the mount points where media is created. You can also, if required, set the number of simultaneous "writers" to be used with the library. HP Data Protector software can then use this just like any other physical backup device and will auto-create the media files on the fly, as required. This new capability is called "Advanced Backup to Disk."

An additional feature of HP Data Protector software, called "Object Copy," enables you to copy from one type of media to another. This means that you can create a backup in File Library, and then

(immediately after the backup, or at a scheduled time) automatically copy the file-based media to a physical medium such as LTO.

So this becomes a method to perform disk staging of slower backups (for example, over a slow LAN), followed by high-speed backup to a physical device. And while this is not a “virtual tape” device, it does provide file-based backup capability without the need to invest in another physical piece of hardware.

File Libraries can also be configured on a disk array in a SAN environment allowing backup across a SAN, as well as LAN.

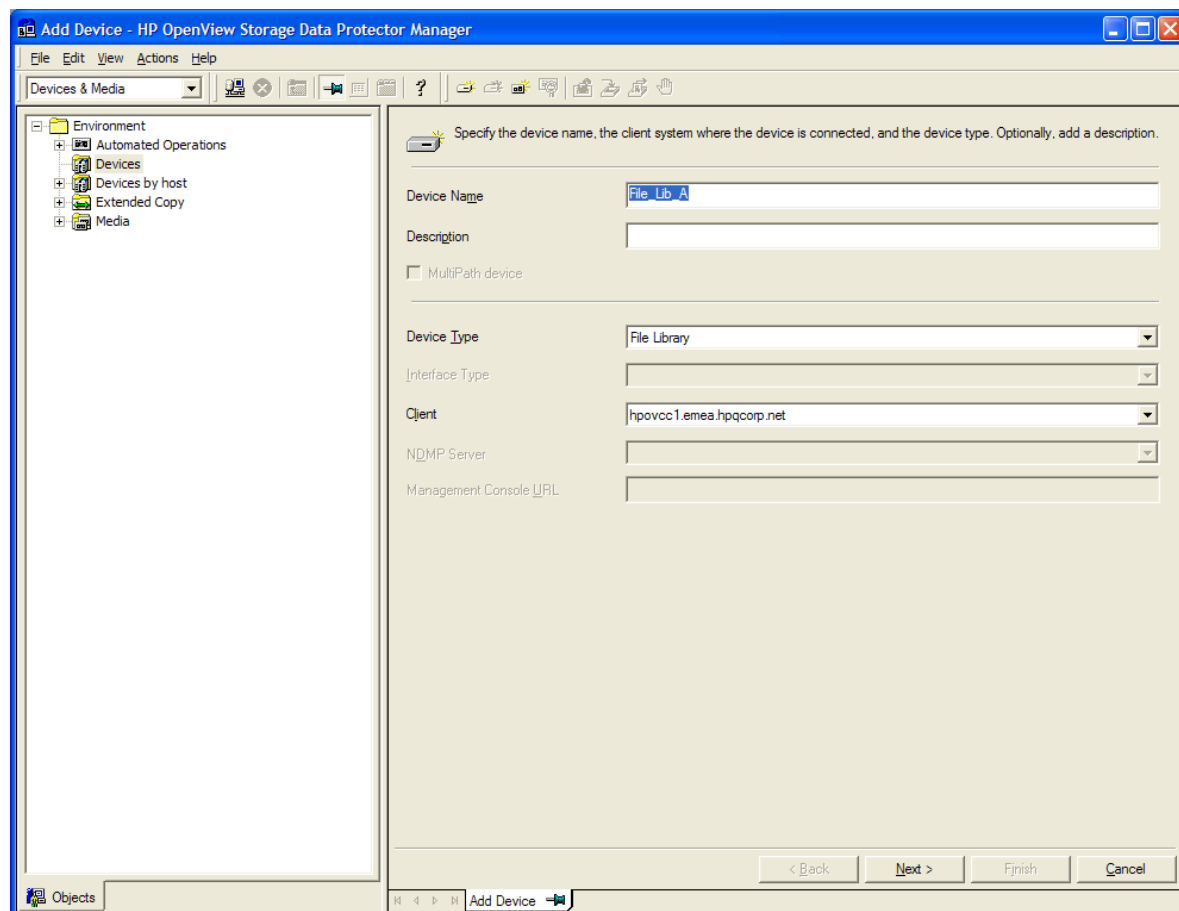
The file device writes and reads sequentially although disks are random-access media.

The file device is not limited by physical space. HP Data Protector software detects a shortage of disk space in such a way that users can still close the data segment and write the last catalog segment to disk. During the file-media formatting, there is an option to allow the pre-allocation of disk space to avoid out-of-disk-space conditions. The maximum supported file size of the File Library device is determined by the maximum size supported by the operating system. If no free slot is available, a new slot will be created.

The free disk space of the file device library is configured per directory. HP Data Protector software will check for free disk space at the creation of a new file depot and at the start of a backup. The size of the new file depot will be calculated in such a way that, even if a device becomes full, there will still be enough space to fulfil the backup requirements. To avoid problems arising when other applications are writing to the same disk, and using up the disk space so that there is not enough space available on the disk after a backup is complete, HP Data Protector software should be the exclusive user of the assigned disk space.

To configure the file device library, select “File Library” as the Device Type (Figure 11).

Figure 11.Defining a File Device Library: Device Type



Next, define the directory and the number of writers that you want to create the File Library for. If the user does not specify this number, the default value will be used. This default value is determined by the number of directories.

The directory has to be created on the disk device prior to configuring the File Library. The user is able to enter several directory names. The directories have to be on different file systems. The “Browse” option is used to select the destination directory. To configure several directories repeat the browse/select operation for each directory.

Properties can be defined to set the maximum size of the file depot, which is equivalent to the defined media size (format size). If the amount of data backed up in one session is bigger than this value, a new file depot is used to continue the backup. The default value is 50 GB.

The second option to specify is the amount of minimum free disk space to create a new file depot, with the default value being 2 MB. The third option to specify is the amount of disk space that should stay free on the disk, with the default value being 0 MB.

The last option allows the user to specify at what remaining percentage of free disk space HP Data Protector software should send a warning. During creation of the File Library, no slots/file depots will be created. They will be automatically added into the File Library during usage of the device. The user need only consider the amount of disk space available for the backup.

All “media” in the File Library are non-appendable. This is useful for efficient disk-space management. Only one session can be stored in one file depot. As soon as the protection on the session of one

depot expires, HP Data Protector software will be able to reuse this file depot. Only in the case that the customer has sessions with small amounts of data or backup of logical and archive logs should the media belong to the pool with appendable policy.

Figure 12. Defining a File Device Library: Destination Directory

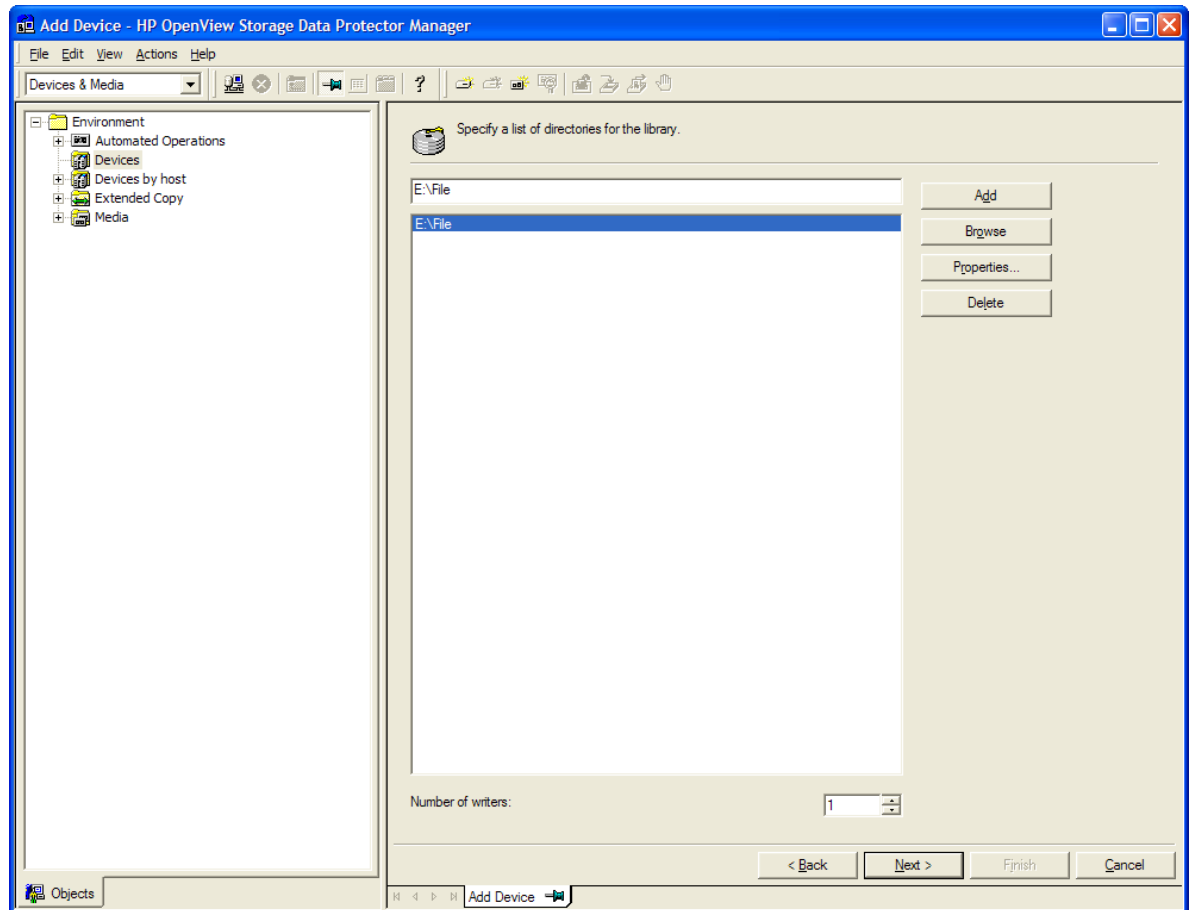


Figure 13. Defining a File Device Library: Browse Drives

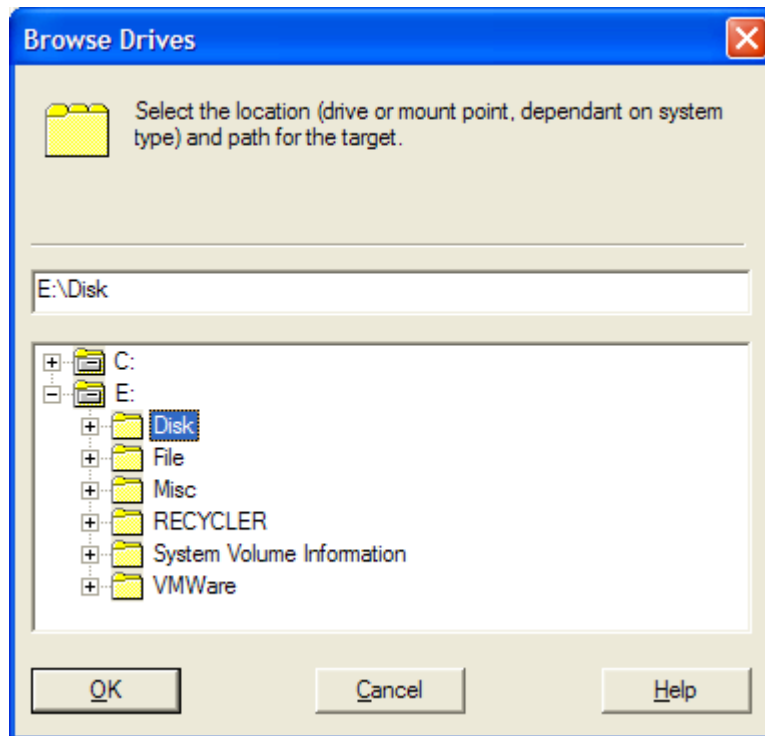
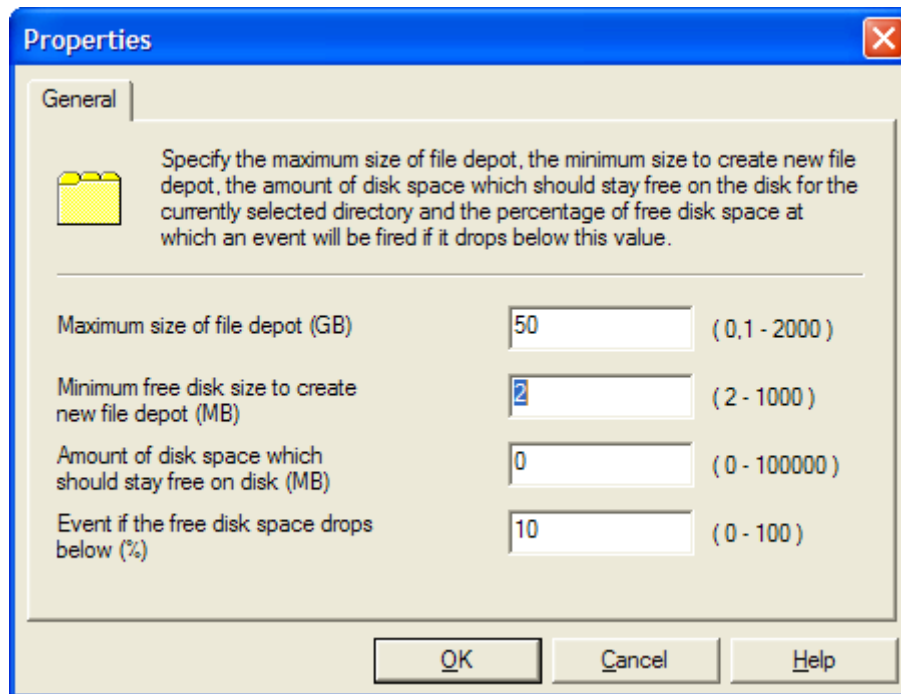


Figure 14. Defining a File Device Library: Setting Device Properties



The summary window contains a table with the following columns:

- Directory name
- Total size of the disk in which this directory is located (GB)
- Size of disk space already used by some other applications (GB)
- Size of disk space that will be used for creation of file depots (GB)
- Maximum (formatted) size of file depots (GB)

Figure 15. Defining a File Device Library: Summary

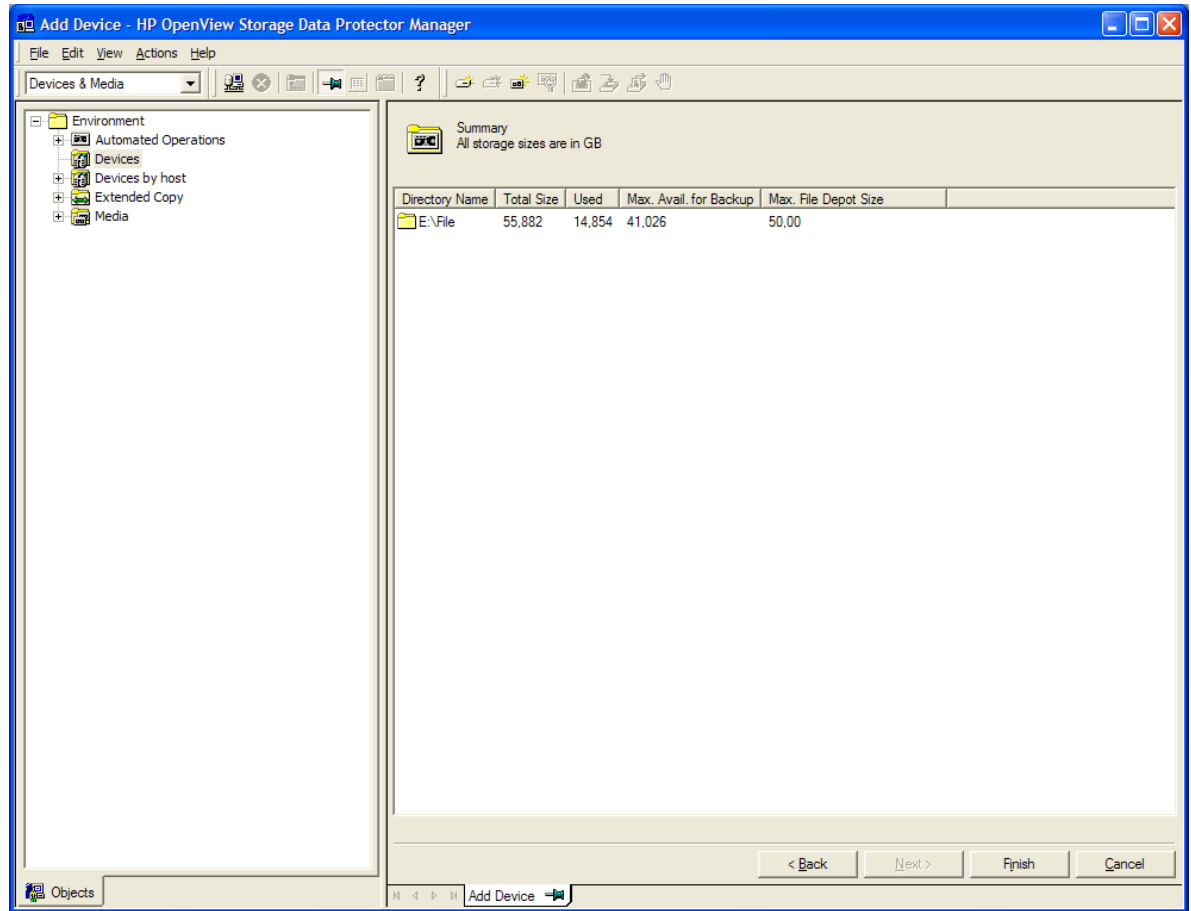
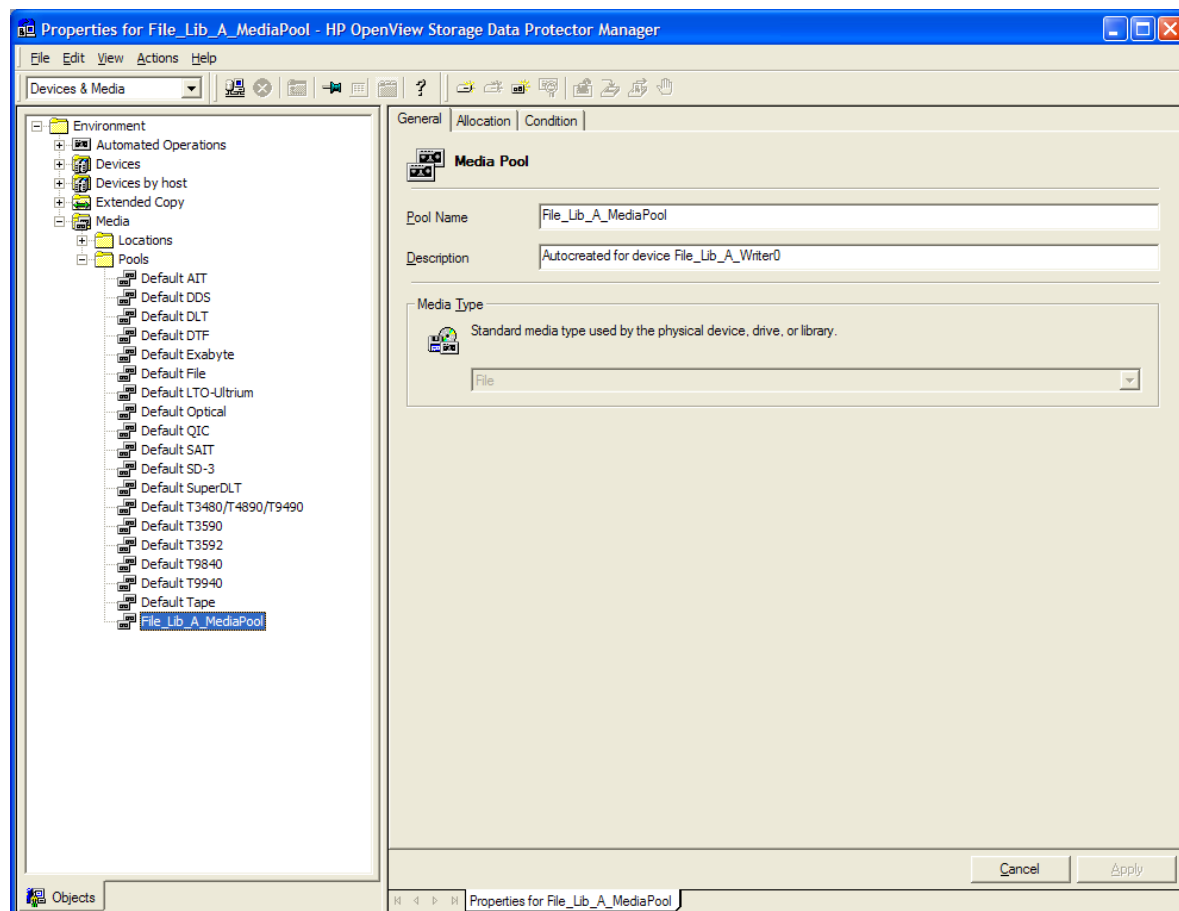


Figure 16. Defining a File Device Library: Media Pool



For each file library, a new Media Pool will be created internally.

The File Library will create new slots/file depots on demand. The user can choose when to create a new file depot/media only after all disk space is used, or only in the event that none of the already existing media is available for backup. If the first option is chosen, HP Data Protector software creates a new file depot. If this fails, HP Data Protector software looks for an existing medium with unprotected data on it. In the second option, HP Data Protector software first tries to find media that contains expired data. If no media is found, a new file depot is created.

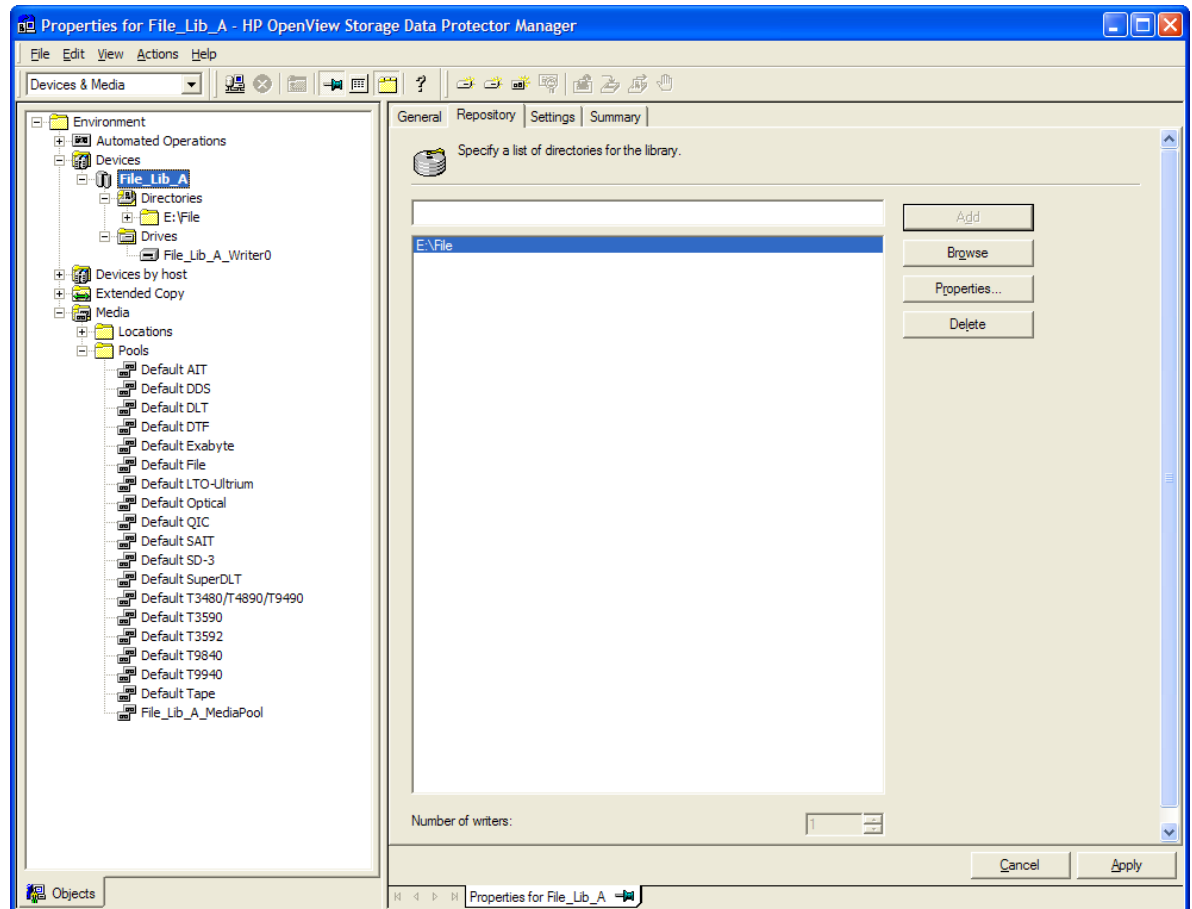
Should HP Data Protector software not be able to create a new file depot or find media with expired data on it, a mount request is issued.

The File Library has a minimum media-retention policy, which means that media with expired data on it will be deleted as soon as possible. This helps to reduce disk space.

After creating the File Library, the user can click on its properties and make changes to it within the repository window. On the Repository tab, the user can add new directories/mount points to the device. The user is also able to delete one or more directories by selecting them and pressing the

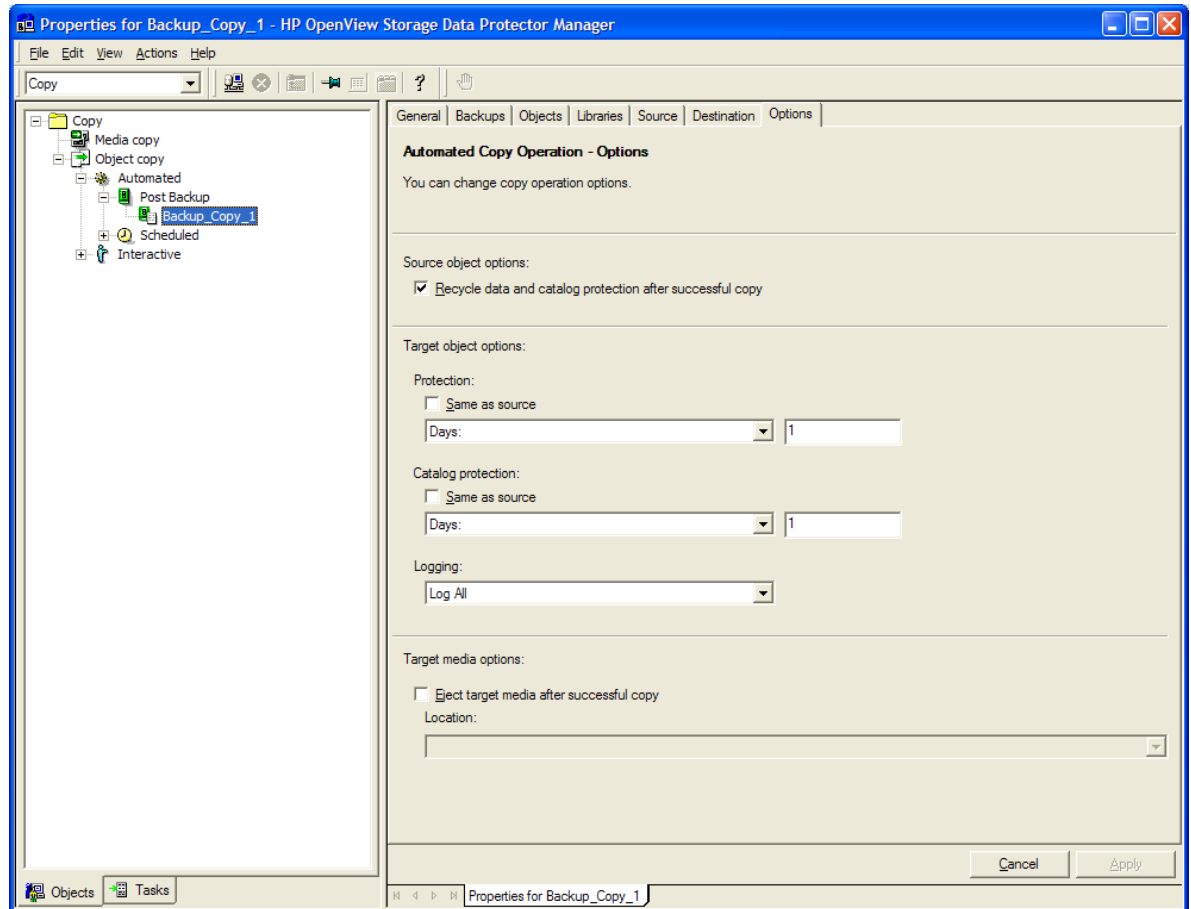
Delete button. A confirmation dialogue box will pop up so that this action can be confirmed. The user can also change the properties of the selected directory/directories by clicking the **Properties** button. A new dialogue box will pop up, in which the user can change the maximum size of the file depots that are going to be created in the selected directories, the minimum size for creation of new file depots, and the free space on the disk.

Figure 17. Defining a File Device Library: Changing File Device Properties



With “Object Copy,” a second backup from disk to tape can be configured, with options to define the source and destination backup specifications, the object filter scheduled time frame (relative or absolute), the library filter tape drives to be used, plus source and target recycling, protection, catalog, and logging options.

Figure 18. Object Copy: Disk Staging Backup to Tape



To restore a file, HP Data Protector software will, as default, choose the fastest medium to restore from. For instance, if the restoration medium is not specified, and the same file version exists on both tape and disk, HP Data Protector software automatically chooses the disk device, as this is the fastest.

Figure 19. Restore: Change Media Location Priority

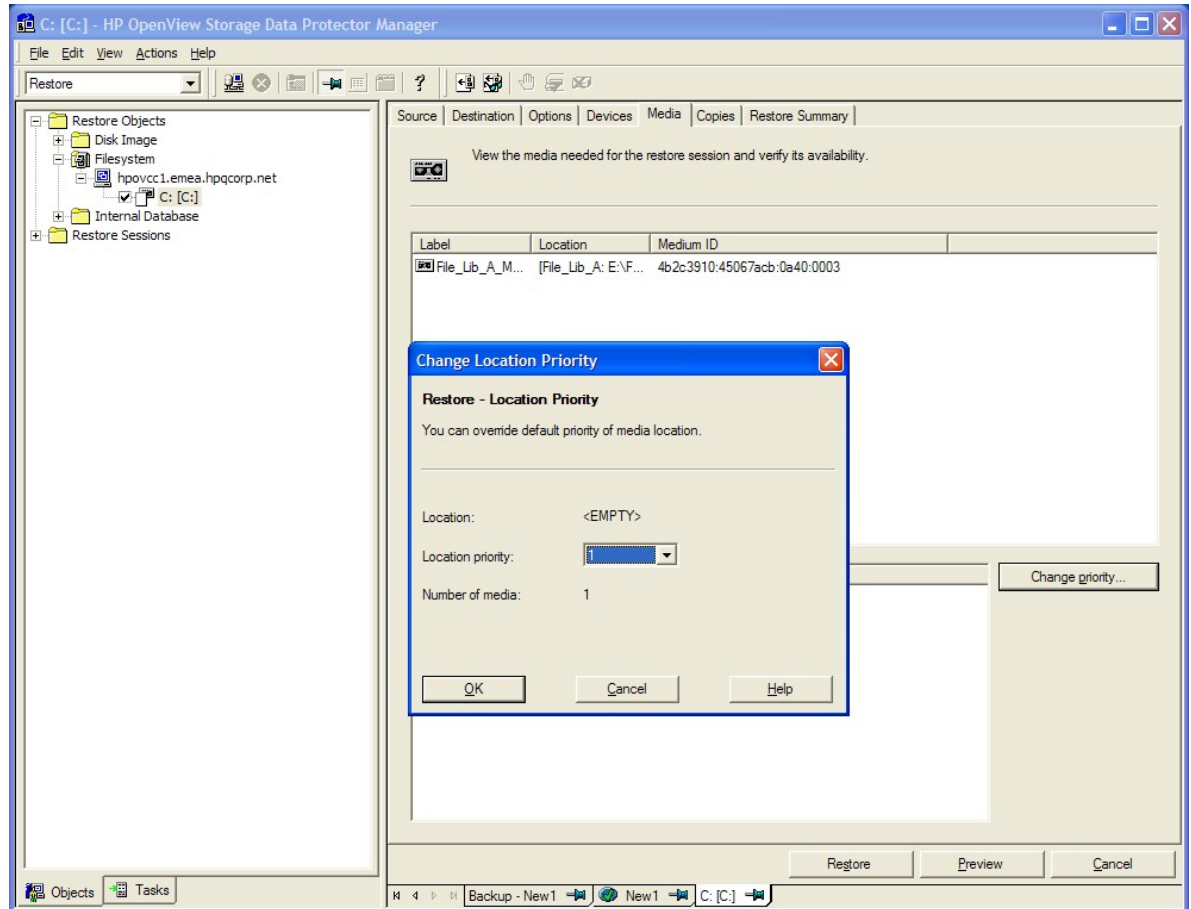
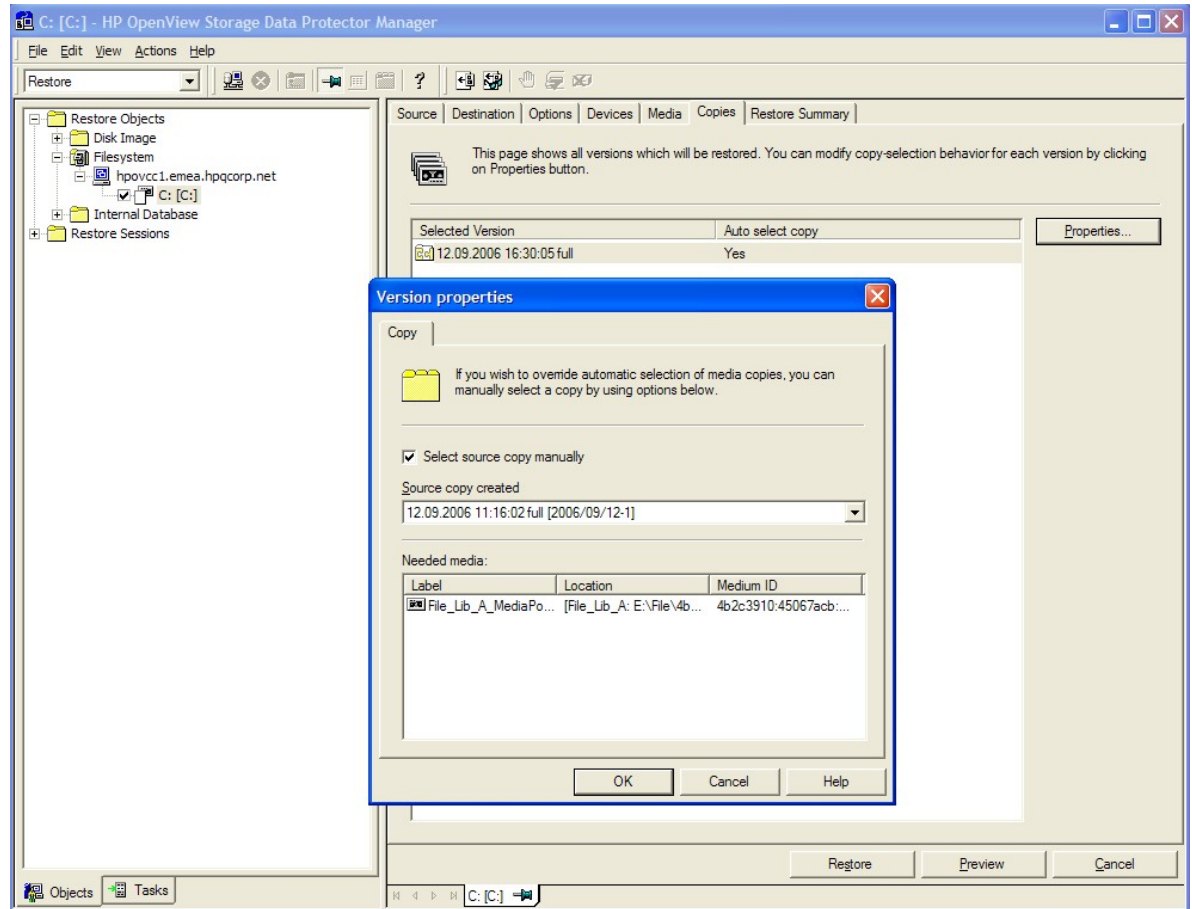


Figure 20. Restore: Auto-select or Manual Select the Source Copy



HP Data Protector software Jukebox and Standalone file device

HP Data Protector software allows you to configure a Standalone or Jukebox file device. To simulate a tape library, HP Data Protector software allows you to create a file jukebox, for example, on which you create and format file media and use as a backup device. This device type evolved from having a simple file-based test mechanism to remove hardware from a troubleshooting situation.

The jukeboxes should be optimized for maximum performance of the disk arrays, and use the same features as a tape device does, such as streaming and multiple writers. With these jukeboxes or standalone file devices, HP Data Protector software does not check for space remaining when writing to a file device. If the disk runs out of space, the backup aborts.

By backing up data to a disk device, HP Data Protector software offers the advantage of quicker file restoration. For instance, when a file has been backed up on disk, it is quicker to restore from a disk medium than from a tape medium. Jukeboxes and standalone virtual tape disk devices simulate tape devices, and would potentially show similar behavior. For instance, searching for a file would not have physical start/stop media issues, but the file slots would be searched one by one as if they were tape cartridges.

Figure 21. Defining a Jukebox: Device Type

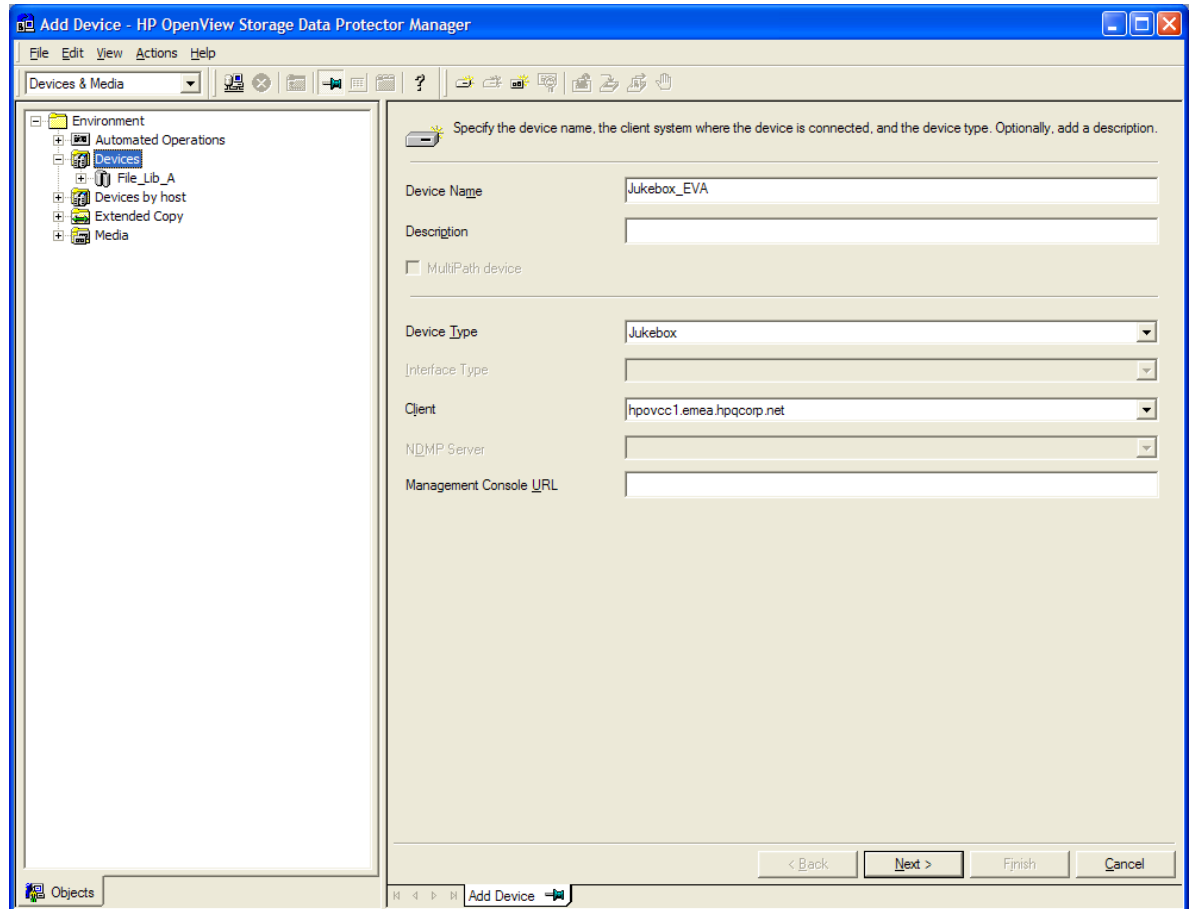


Figure 22. Defining a Jukebox: Slots

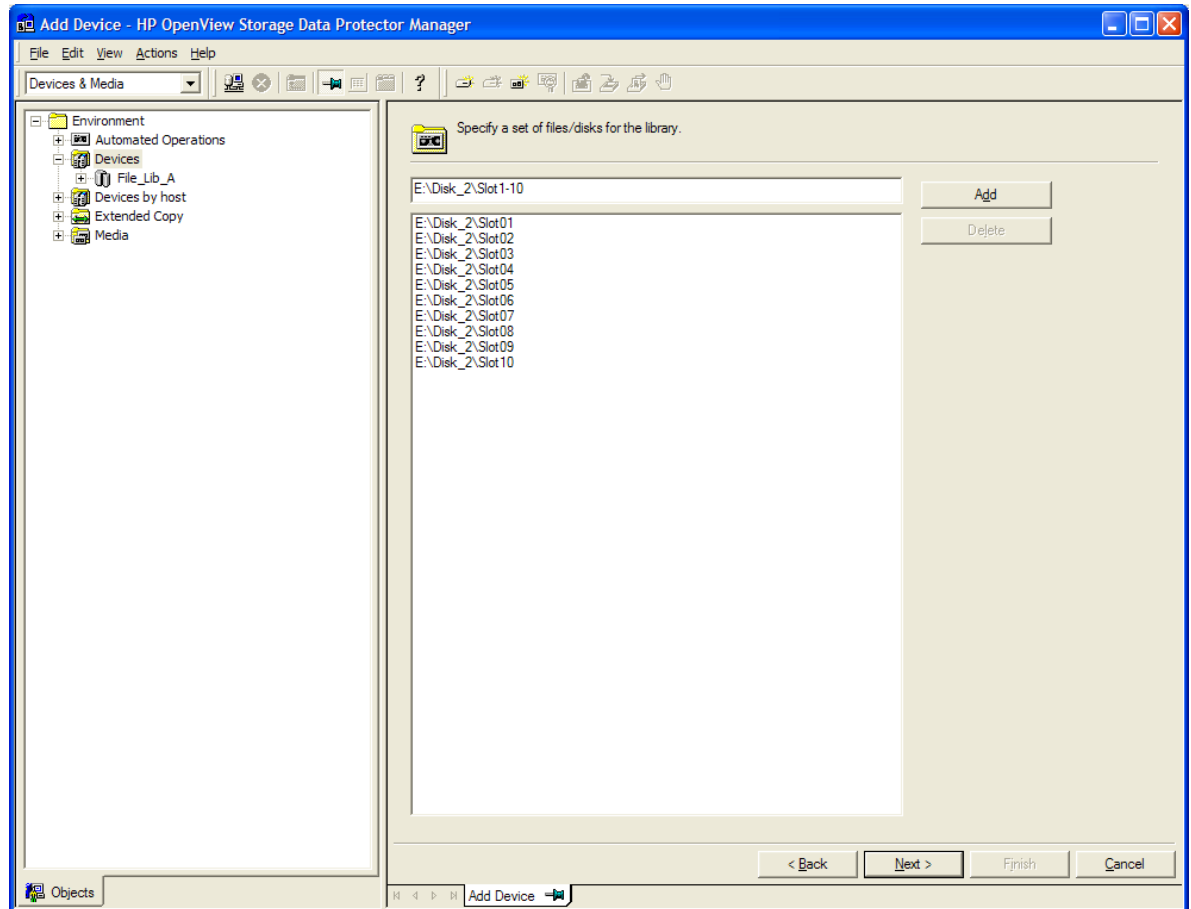


Figure 23. Defining a Jukebox: Drives

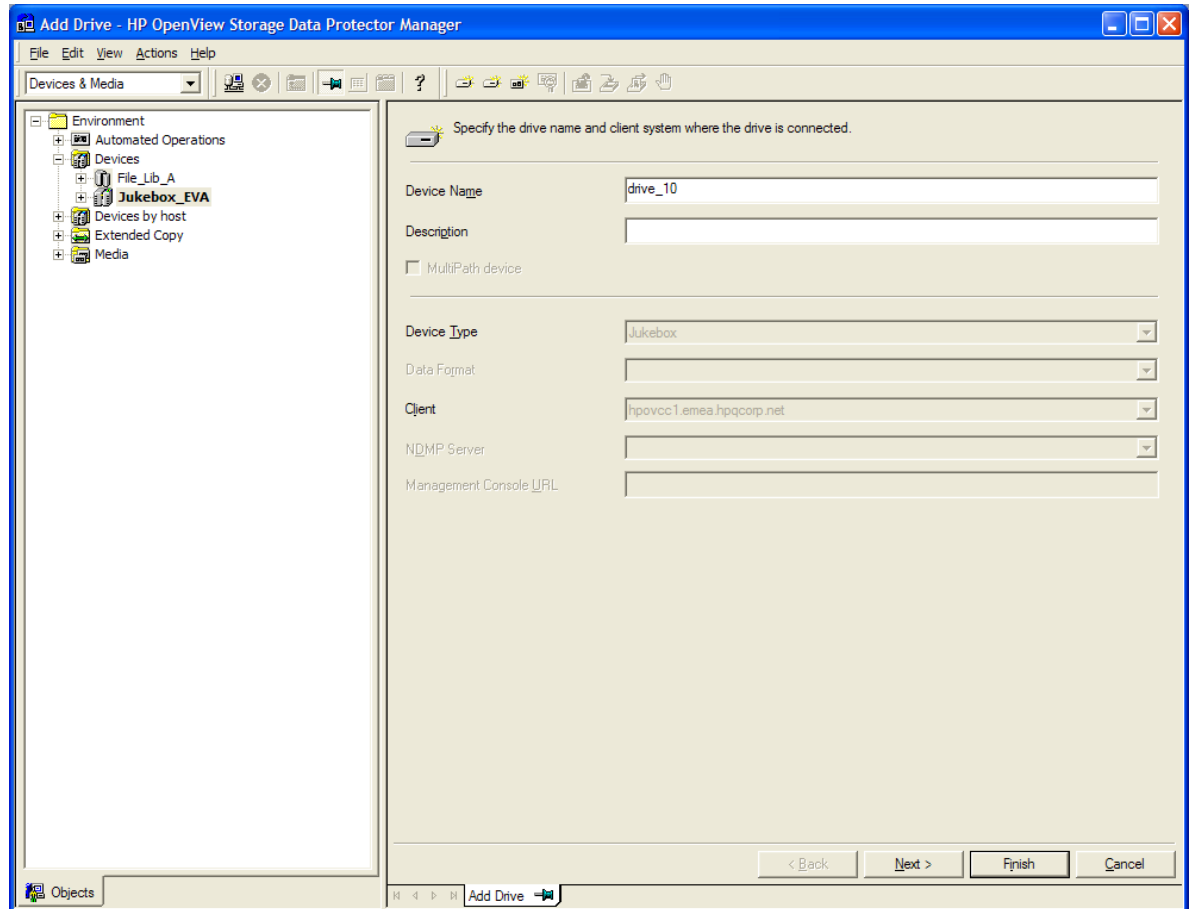
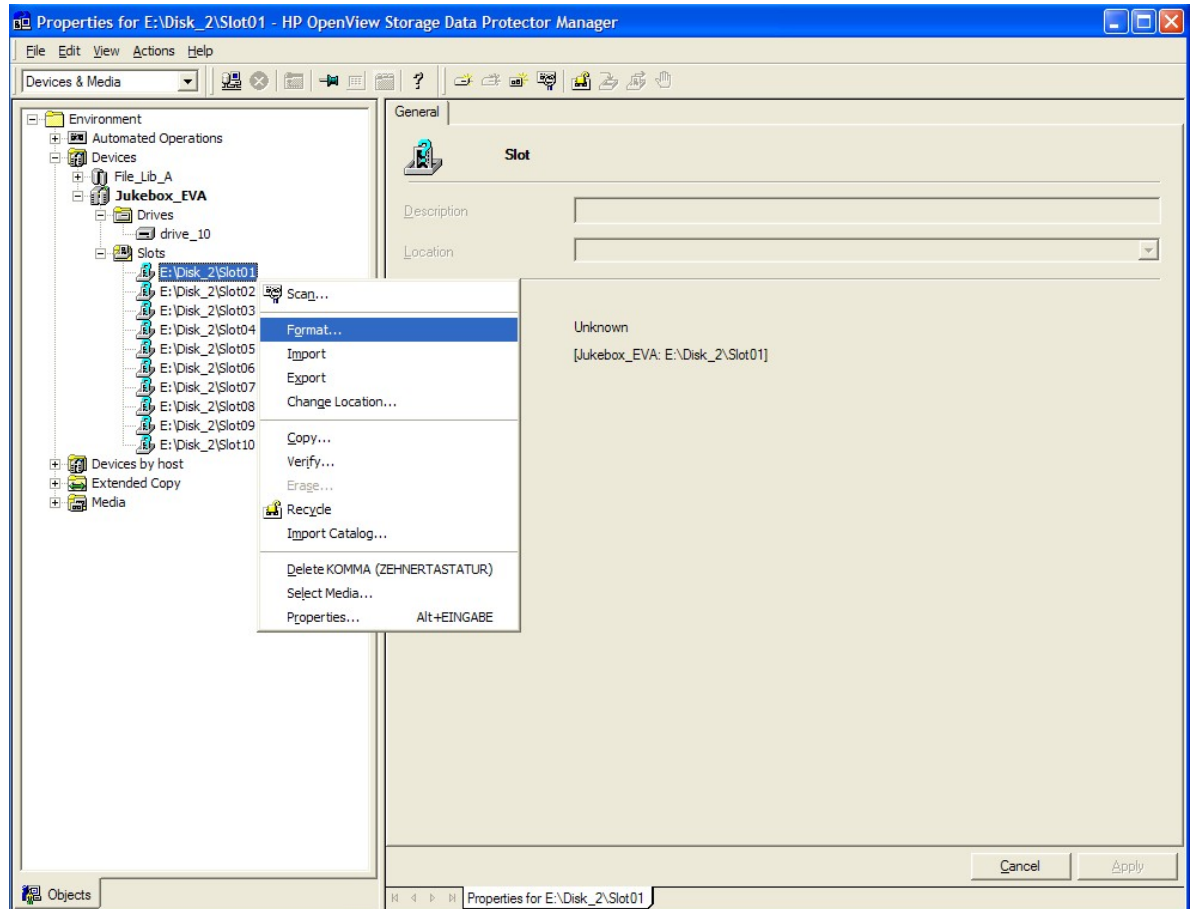


Figure 24. Defining a Jukebox: Formatting Media



For more information

For more information on HP Data Protector software, visit:

www.hp.com/go/dataprotector

© 2006 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

4AA0-7625ENW, September 2006

