Enterprise Backup and Restore technology and solutions

LESSON VII

HP Data Protector
Overview and Concepts
Part IV

Veselin Petrunov
Backup and Restore team / Deep Technical Support

HP Bulgaria Global Delivery Hub
Global Operations Center
November, 2013

HP Data Protector

Overview and Concepts

Part IV

Veselin Petrunov / November 2013

Contents

Part IV

HP Data Protector

How Data Protector operates

HP Data Protector

Data Protector processes or services

Service name	Description
Inet	The Data Protector Inet service runs on each Windows system in the Data Protector cell. Inet is responsible for communication between systems in the cell and starts other processes needed for backups and restores. The Data Protector Inet service is started when Data Protector is installed on a system. On UNIX systems, the system inet daemon (INETD) starts the Data Protector Inet process.
CRS	The CRS (Cell Request Server) process (service) runs on the Data Protector Cell Manager. It starts and controls backup and restore sessions. The service is started when Data Protector is installed on the Cell Manager system and is restarted each time the system is restarted.
KMS	The KMS (Key Management Server) process (service) runs on the Cell Manager and provides key management for the Data Protector encryption functionality. The process is started when Data Protector is installed on the Cell Manager.
MMD	The MMD (Media Management Daemon) process (service) runs on the Data Protector Cell Manager and controls media management and device operations. The process is started by the Cell Request Server process (service).
RDS	The RDS (Raima Database Server) process (service) runs on the Data Protector Cell Manager and manages the IDB. The process is started when Data Protector is installed on the Cell Manager.
UIProxy	The Java GUI Server (UIProxy service) runs on the Data Protector Cell Manager. It is responsible for communication between the Java GUI Client and the Cell Manager, moreover, it performs business logic operations and sends only important information to the client. The service is started as soon as Data Protector is installed on the Cell Manager.

HP Data Protector

Backup sessions

What is a backup session?

When a backup specification is started it is called a backup session. The backup session copies data from a source, typically a hard disk, to a destination, typically tape media. The result of a backup session is a copy of data on the backup media, the media set.

Scheduled backup session

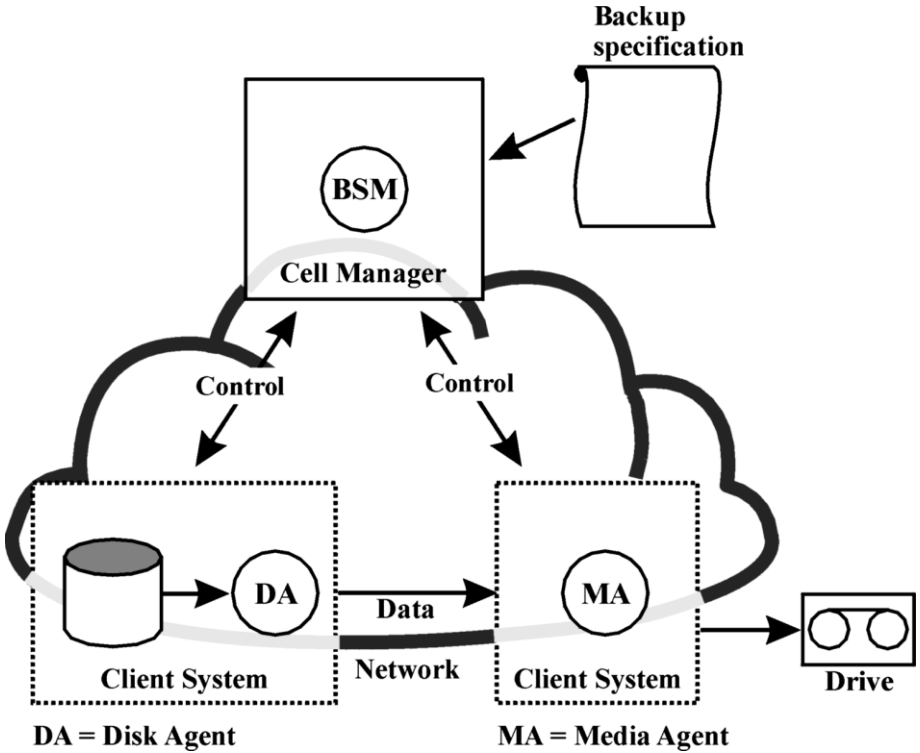
A scheduled backup session is started by the Data Protector Scheduler at the time you have specified. You can view the progress of the scheduled backup session in the Data Protector monitor.

Interactive backup session

An interactive backup session is started from the Data Protector user interface directly. The Data Protector monitor starts immediately and you can view the progress of the backup session. Note that multiple users can monitor the same backup session. You may want to stop monitoring by disconnecting the user interface from the session. The session will then continue in the background.

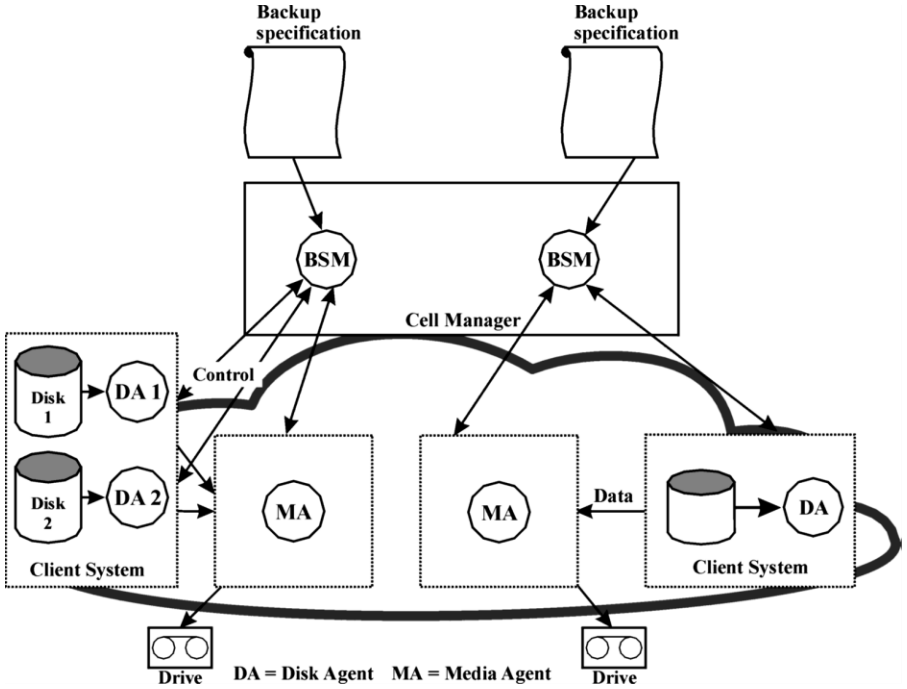
HP Data Protector

Backup session data flow and processes



HP Data Protector

How many sessions can run concurrently?



HP Data Protector

Pre-exec and post-exec commands

Data Protector pre-exec commands enable you to execute some actions before a backup or a restore session. Data Protector post-exec commands enable you to execute some actions after a backup or a restore session. A typical pre-exec action would be to shut down a database to put data in a consistent state.

The pre-exec and post-exec commands can be set for a backup specification and, as such, executed on the Cell Manager system, or they can be specified as a backup object option and thus executed on the client system where the respective Disk Agent is running.

Pre-exec and post-exec script commands can be written as executables or shell scripts. These are not supplied by Data Protector and must be written separately by, for example, the backup operator.

HP Data Protector

Queuing of backup sessions

Timeout

When a backup session is started, Data Protector tries to allocate all needed resources, such as devices. The session is queued until the required minimum resources are available. If the resources are still unavailable after the timeout, the session is aborted. The timeout period can be set using the *SmWaitForDevice* global option.

Optimizing the load

To optimize the load on the Cell Manager, Data Protector can, by default, start up to five backup sessions at the same time. The default value can be modified in the global options file. If more are scheduled at the same time, the extra sessions are queued and started subsequently as others are finished.

HP Data Protector

Mount requests in backup sessions

What is a mount request?

A mount request in a backup session appears when Data Protector needs a new medium for backup and the medium is not available.

Data Protector issues a mount request for one of the following reasons:

- There is not enough space on the backup media and there are no new media available.
- Data Protector media allocation policy for backup requires a medium that is not available in the device.
- The order of media used for backup is defined in the pre allocation list and media are not available in this order.

Responding to a mount request

Responding to a mount request includes providing the required media and telling Data Protector to proceed with the backup.

Data Protector allows you to configure what happens when a mount request is issued:

- Sending notification to an operator
- Automating a mount request

HP Data Protector

Backing up with disk discovery

What is disk discovery?

In backing up with disk discovery, Data Protector creates a detailed list of disks on the target system when the backup session is started, and backs up all disks. Therefore, all local disks on the system are backed up even though they were not present on the system when the backup was configured. Backup with disk discovery is particularly useful in dynamic environments, where configurations change rapidly. It enables you to select or exclude specific directories in the backup.

How does it compare to a standard backup?

In a standard backup, you explicitly configure specific disks, directories or other objects for backup by configuring them in the backup specification. Therefore, only these objects are backed up. If you add new disks to the system or want to back up some other objects, you must manually edit the backup specification and these new objects. You can select, as you configure the backup, the method you want to use - disk discovery or standard backup.

HP Data Protector

Restore sessions

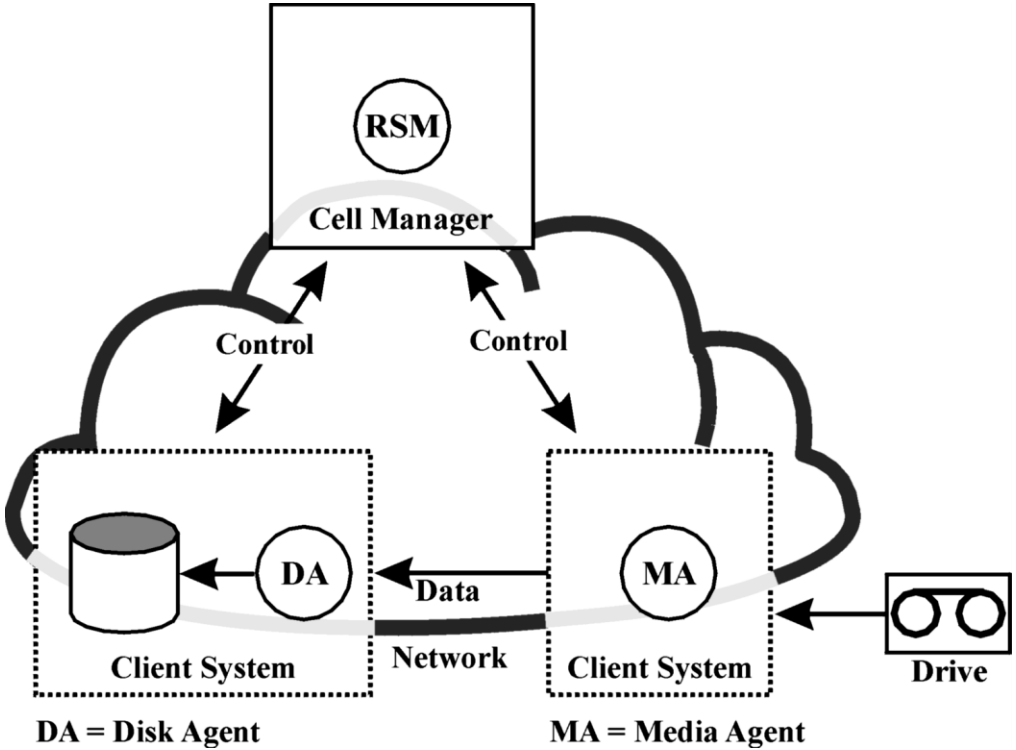
What is a restore session?

In a restore session, data is copied from a backup copy, typically on a tape medium, back to a disk.

A restore session is started interactively. You tell Data Protector what to restore, let Data Protector determine the needed media, select some options and start the restore. You and other users can monitor the progress of the session.

HP Data Protector

Restore session data flow and processes



HP Data Protector

Restore sessions

How many restore sessions can run concurrently?

A number of restore sessions can run in the cell at the same time. This number is limited by resources in the cell, such as the Cell Manager and systems with connected devices.

Queuing of restore sessions / Timeout

When a restore session is started, Data Protector tries to allocate all needed resources, such as backup devices. The session is queued for as long as the required minimum resources are not yet available. Data Protector tries to allocate the resources for a specific period of time, the timeout. Timeout is user configurable. If the resources are still unavailable after the timeout, the session is aborted.

Mount requests in a restore session

What is a mount request?

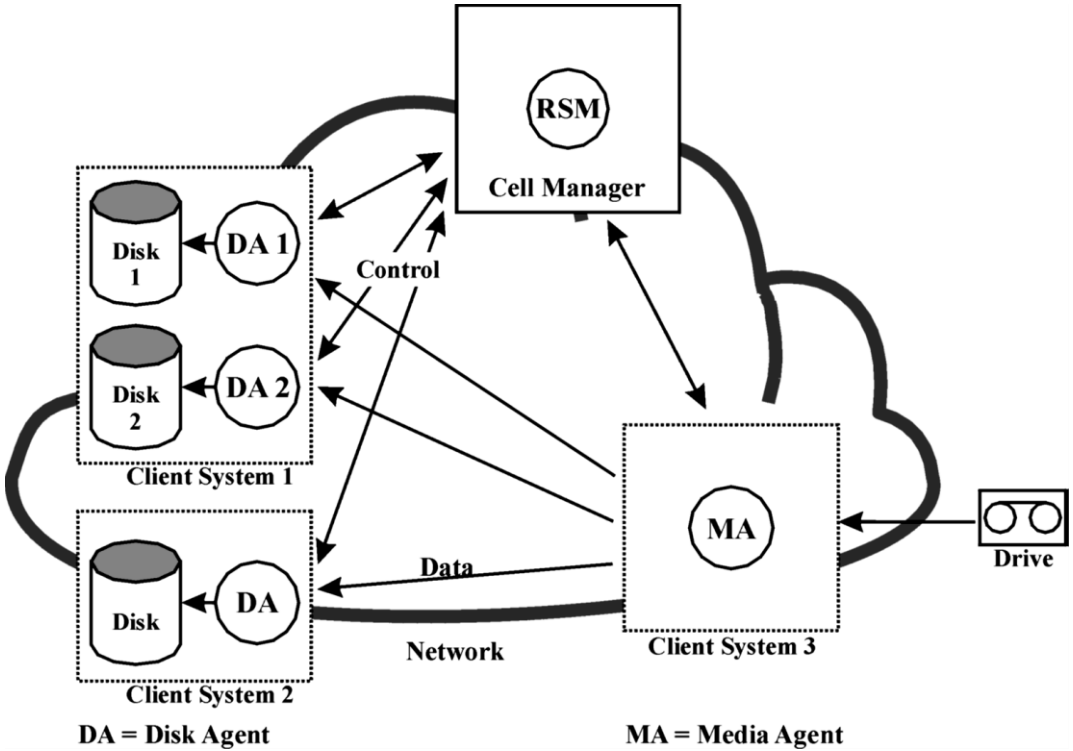
A mount request appears in a restore session when the media needed for restore are not available in the device. Data Protector allows you to configure a desired action that should happen when a mount request appears.

Responding to a mount request

Responding to a mount request includes providing the required media or any copy of media and telling Data Protector to proceed with the restore.

HP Data Protector

Parallel restores



HP Data Protector

Restore sessions

Fast multiple single file restore

Data Protector uses discontinuous object restore to improve restore performance. After restoring a specific file or tree, Data Protector repositions itself directly on the next file or tree on the medium, if there's at least a single segment between the files or trees, and continues the restore.

Within an individual restore object you can start multiple Disk Agents. This way the restoring of multiple single files that are located all over the medium is much faster than if Data Protector were to traverse the medium.

Resuming restore sessions

Restore sessions that did not complete successfully (for example, due to some network problems) can be resumed using the Data Protector resume session functionality. When you resume a failed session, Data Protector continues with the restore in a new session, starting right from where the failed session left off.

HP Data Protector

Object copy sessions

What is an object copy session?

An object copy session is a process that creates an additional copy of the backed up, copied, or consolidated data on a different media set. During an object copy session, the selected backed up, copied, or consolidated objects are copied from the source to the target media.

Automated and interactive object copy sessions

Automated object copy session

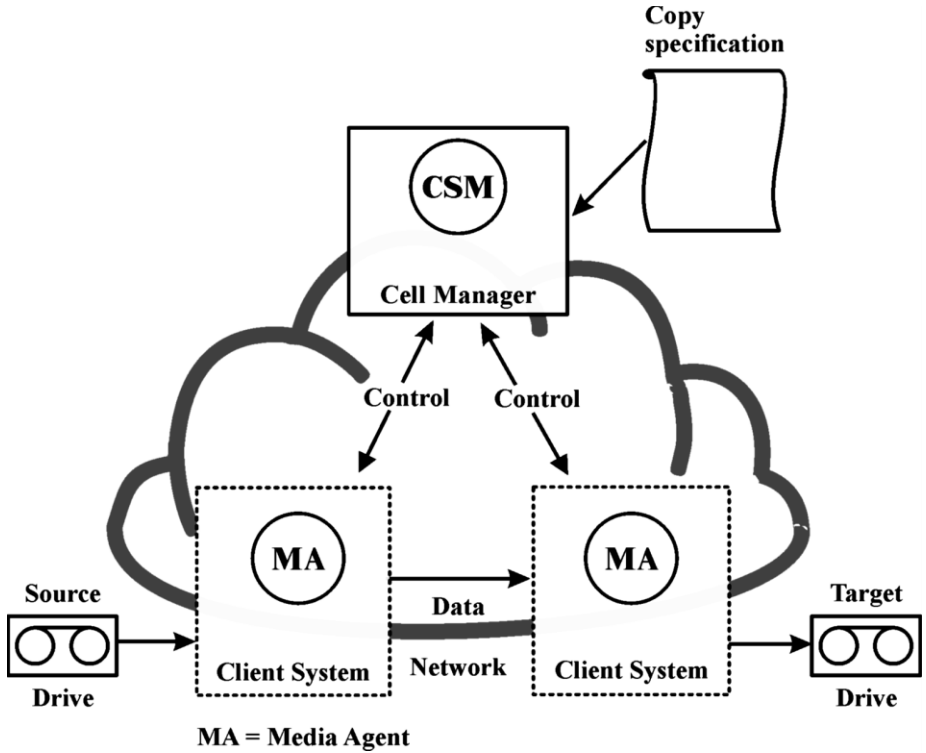
An automated object copy session can either be scheduled or started immediately after a backup, object copy, or object consolidation. A scheduled object copy session is started at the time you have specified using the Data Protector Scheduler. A post-backup or a post-copy or a post-consolidation object copy session is started after the specified session finishes. You can view the progress of the automated object copy session in the Data Protector monitor.

Interactive object copy session

An interactive object copy session is started from the Data Protector user interface directly. The Data Protector monitor starts immediately and you can view the progress of the session. Multiple users can monitor the same object copy session. You may want to stop monitoring by disconnecting the user interface from the session. The session will then continue in the background.

HP Data Protector

Object copy session data flow and processes



HP Data Protector

Object copy sessions

How many sessions can run concurrently?

A number of object copy sessions can run in the cell at the same time. This number is limited by the resources in the cell, such as the Cell Manager and the systems with connected devices. However, it is not possible to run two or more object copy sessions from the same object copy specification in parallel.

Queuing of object copy sessions / Timeout

When an object copy session is started, Data Protector tries to allocate all needed resources. The session is queued until the required minimum resources are available. If the resources are still unavailable after the timeout, the session is aborted. The timeout period can be set using the *SmWaitForDevice* global option.

Mount requests in an object copy session

What is a mount request?

A mount request in an object copy session is issued when a source or a target medium needed for the object copy operation is not available.

Responding to a mount request

Responding to a mount request includes providing the required medium and confirming the mount request. If the required source medium has media copies, you can provide a copy instead of the original medium.

HP Data Protector

Object consolidation sessions

What is an object consolidation session?

An object consolidation session is a process that merges a restore chain of a backup object, consisting of a full backup and at least one incremental backup, into a new, consolidated version of this object. During an object consolidation session, Data Protector reads the backed up data from the source media, merges the data, and writes the consolidated version to the target media.

Automated and interactive object consolidation sessions

Automated object consolidation session

An automated object consolidation session can either be scheduled or started immediately after a backup. A scheduled object consolidation session is started at the time you have specified using the Data Protector Scheduler. A post-backup object consolidation session is started after the specified backup session finishes. You can view the progress of an automated object consolidation session in the Data Protector monitor.

Interactive object consolidation session

An interactive object consolidation session is started from the Data Protector user interface directly. The Data Protector monitor starts immediately and you can view the progress of the session. Multiple users can monitor the same object consolidation session. You may want to stop monitoring by disconnecting the user interface from the session. The session will then continue in the background.

HP Data Protector

Object consolidation session data flow and processes

When an object consolidation session is started, the following happens:

1. The Copy and Consolidation Session Manager (CSM) process is started on the Cell Manager system. This process reads the object consolidation specification for information on what to consolidate and which options, media, and devices to use. It controls the object consolidation session.
2. The CSM opens the IDB, reads the information about the needed media, and writes the information about the object consolidation session, such as generated messages, to the IDB.
3. The CSM locks the devices. The session is queued until all read Media Agents and the minimum required write Media Agents are locked, with the same timeout as for backup. If the resources are still unavailable after the timeout, the session is aborted.
4. The CSM starts the Media Agents on the systems with devices that will be used in the session. The Media Agents load the source and target media allocated according to the backup policies. If destination devices are not specified per object, Data Protector selects them automatically from those you selected in the object consolidation specification according to the following criteria in the order of priority:
   - destination devices with the same block size as source devices are selected before those with a different one
   - locally attached devices are selected before network attached devices
5. One Media Agent reads the full object version. It sends the data to another Media Agent that reads incremental object versions. The latter Media Agent does the actual consolidation and sends the data to the Media Agent that writes the data to the target media. If the full backup and the incremental backups reside in the same file library, the same Media Agent reads all the backups and consolidates them. If the block size of the source device is smaller than that of the destination device, blocks are repackaged.
6. When the object consolidation session is completed, the CSM closes the session.

HP Data Protector

Object consolidation sessions

How many sessions can run concurrently?

A number of object consolidation sessions can run in the cell at the same time. Object consolidations sessions are treated like backup sessions and their number is limited by the same factors.

Queuing of object consolidation sessions / Timeout

When an object consolidation session is started, Data Protector tries to allocate all needed resources. The session is queued until the required minimum resources are available. If the resources are still unavailable after the timeout, the session is aborted. The timeout period can be set using the *SmWaitForDevice* global option.

Mount requests in an object consolidation session

What is a mount request?

A mount request in an object consolidation session is issued when a source or a target medium needed for the object consolidation operation is not available.

Responding to a mount request

Responding to a mount request includes providing the required medium and confirming the mount request. If the required source medium has media copies, you can provide a copy instead of the original medium.

HP Data Protector

Object verification sessions

What is an object verification session?

An object verification session is a process that verifies the media segments allocated to a specified object or specified objects, checking the information in the header segments and reading the data blocks in the data segments to verify their format. If a cyclic redundancy check (CRC) was performed during the original backup, it also recalculates the CRC and compares it with the original.

Data Protector can perform the verification on the host that was the source of the backup, effectively verifying the Data Protector components in the restore path, on another host, verifying restore capability to a different location, or directly on the host with the media agent involved, verifying the data only.

Automated and interactive object verification sessions

Automated object consolidation session

You can specify an automatic object verification session to run at a specified time, using the Data Protector Scheduler, or to run as a post-backup object verification session immediately after completion of a specified backup, object copy, or object consolidation session. You can view the progress of such sessions in the Data Protector monitor.

Interactive object consolidation session

You can start an interactive object verification session directly from the Data Protector user interface. The Data Protector monitor starts immediately and you can view the progress of the session. Multiple users can monitor the same object verification session. You can perform other operations with the user interface and let the session continue in the background, if required.

HP Data Protector

Object verification session data flow and processes

When an object verification session is started, the basic process flow is as follows:

1. The Restore Session Manager (RSM) process is started on the Cell Manager system, triggered either by:
   - the Data Protector Scheduler, for a scheduled session
   - the End of Session event, for post-backup sessions
   - the user from the GUI or the CLI, for interactive sessions

   This process controls the verification session.
2. The RSM opens the IDB, reads the information about the objects to be verified, and writes information about the verification session, such as generated messages, to the IDB.
3. The RSM starts the Media Agents (MA) on the source systems involved in the verification. For each drive used in parallel, a new Media Agent is started.
4. Verification of the data is performed by the Disk Agents (DA) on the destination hosts, so the RSM starts a Disk Agent for each destination disk in parallel. The actual number of Disk Agents started depends on the objects you selected for verification. The process is similar to that for restore.
5. The Media Agents read the object data from the media and send it to the Disk Agents that perform the verification. The RSM monitors the progress of the session and starts new Disk Agents and new Media Agents as necessary.
6. When the object verification session is completed, the RSM closes the session.

HP Data Protector

Media management sessions

What is a media management session?

A media management session is used to perform a certain action on the media, such as initializing media, scanning the content, verifying data on the media, and copying media.

Logging to the IDB

Information about a media management session, such as generated messages, is stored in the IDB.

Data Protector monitor and media management session

A media management session can be viewed in the monitor window. If you close the Data Protector GUI, the session will continue in the background.

HP Data Protector

Media management session data flow

When a media management session is started, the following happens:

1. The Media Session Manager (MSM) process is started on the Cell Manager system. This process controls the media session.
2. The MSM starts the Media Agents (MAs) on the system that has devices used for the media management session.
3. Media Agents perform the requested operation and send generated messages to the Data Protector user interface, where you can track the progress. The session is also stored in the IDB.
4. When the session is complete, the MSM closes the session.

How many sessions can run?

A number of media management sessions can run in the cell at the same time if they do not use the same resources, such as devices or media.

HP Data Protector

How Data Protector operates

HP Data Protector

Integration with applications

HP Data Protector

Integration with database applications

From the user's perspective, a **database** is a set of data. Data in a database is stored in **tables**. Relational tables are defined by their columns and are given a name. Data is stored in rows in the table. Tables can be related to each other, and the database can be used to enforce these relationships. Data can thus be stored in **relational format** or as **object-oriented** structures such as abstract data types and methods. Objects can be related to other objects, and objects can contain other objects. A database is usually managed by the server (manager) process that maintains data integrity and consistency.

Whether you use relational structures or object-oriented structures, databases store data in **files**. Internally, these are database structures that provide a logical mapping of data to files, allowing different types of data to be stored separately. These logical divisions are called **tablespaces** in Oracle, **dbspaces** in Informix Server, and **segments** in Sybase.

HP Data Protector

Relational database

Data files are physical files that contain all of a database's data. They change randomly and can be very large. They are internally divided into pages.

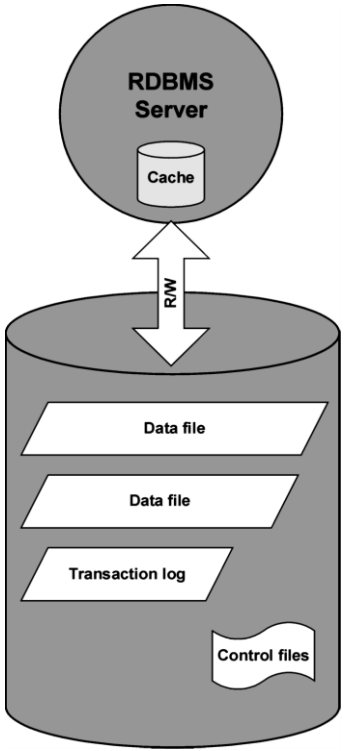
Transaction logs record all database transactions before they are further processed. Should a failure prevent modified data from being permanently written to data files, the changes can be obtained from log files. Any kind of recovery is done in two parts: **roll forward**, which applies transaction changes into the main database and **roll back**, which removes uncommitted transactions.

Control files hold information about the physical structure of the database, such as, database names, names and locations of a database's data files and log files, and the time stamp of the database's creation. This control data is kept in control files. These files are critical for the operation of the database.

The **cache** of the database server process contains the most-often used pages of the data files.

The following is the standard flow of transaction processing:

1. A transaction is first recorded into the transaction log.
2. Changes required in the transaction are then applied to cached pages.
3. From time to time sets of modified pages are flushed to data files on disk.



HP Data Protector

Filesystem backup of databases and applications

Databases are constantly changing while they are online. Database servers consist of multiple components that minimize response time for connected users and increase performance. Some data is kept in the internal cache memory and some in temporary log files, which are flushed at **checkpoints**.

Because data in a database can change during a backup, a filesystem backup of database files makes no sense without putting the database server into a special mode or even offline. Saved database files have to be in a consistent state, otherwise the data is of no use.

The following steps are required to configure a filesystem backup of the database or application:

- identify all data files
- prepare two programs that are able to shut down and start up the database, respectively
- configure the filesystem **backup specification** with all the data files included and specify the shut-down program as a **pre-exec command** and the start-up program as a **post-exec command**

This method is relatively simple to understand and configure but has one key disadvantage: *the database is not accessible during the backup, which is unacceptable for most business environments*.

HP Data Protector

Online backup of databases and applications

To overcome the necessity to shut down the database during a backup, database vendors have prepared interfaces that can be used to put databases temporarily into special modes to save the data to tapes. Server applications are thus online and available to users during the backup or restore process. These application-specific interfaces allow backup products, like Data Protector,

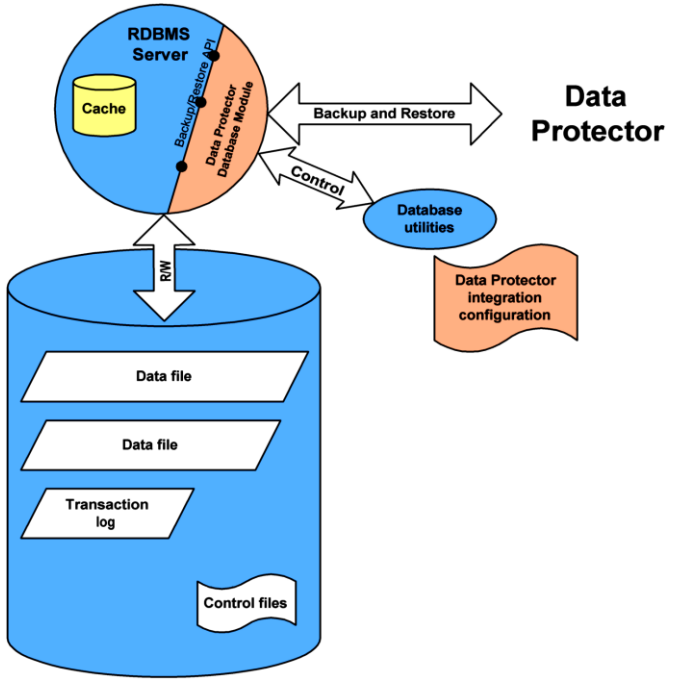
to back up or restore logical units of the database application. The functionality of the backup APIs varies depending on the database vendor. Data Protector integrations are available for major databases and applications. For a detailed list of supported integrations, see the HP Data Protector Product Announcements, Software Notes, and References.

The essence of the backup interface is that it provides the backup application with consistent data (even if it may not be consistent on the disk) while at the same time keeping the database operational.

HP Data Protector

Data Protector integration with databases

The picture shows how a relational database is integrated with Data Protector. Data Protector provides a **Database Library** that is linked in to the database server. The database server sends data to Data Protector and requests data from it. Database utilities are used to trigger backup and restore operations.



HP Data Protector

Data Protector integration with databases

A typical procedure to configure the backup of a database through the Data Protector integration is as follows:

1. A database/application-specific agent is installed on the database system

2. The Data Protector integration is configured for each database. Data needed for Data Protector to work with this database are stored on the database system (into configuration files or registry entries). Typically, this includes pathnames and user names/passwords.

3. The backup specification is prepared using the Data Protector user interface.

Besides the key advantage of the database being **online** all the time there are also other benefits of using the Data Protector integrations with the databases:

- There is no need to specify the location of data files. These can be located on different disks.
- The logical structure of the database can be browsed. It is possible to select only a subset of the database.
- Applications are aware of backup operation and keep track of which parts are backed up.
- Several modes of backup are possible. Besides **full** backups, users can select (block level) **incremental** backups or only the backup of transaction logs.
- Several modes of restore are possible and after the restore of data files, the database can automatically restore transaction logs and apply them as configured.

HP Data Protector

Integration with virtualization environments

Offline filesystem backup of virtual machines

As virtual machines are constantly changing while they are online, you must put the virtual machines in a special mode or even shut them down them before you start a filesystem backup.

Files on the disk that belong to a virtual machine have to be in a consistent state, otherwise the backup image created for that virtual machine is of no use.

To configure a filesystem backup of a virtual machine, you must identify all virtual machine files, create two programs that are able to shut down and start up the virtual machine, and create a filesystem backup specification with all the virtual machine files included and specify the shut-down program as a pre-exec command and the start-up program as a post-exec command.

This method is relatively straightforward, but has one key disadvantage: the virtual machine cannotbe used actively during the backup.

HP Data Protector

Integration with virtualization environments

Online backup of virtual machines

Data Protector can use specific interfaces provided by the virtualization environments to perform a backup of virtual machines while they are running (online backup). Depending on the virtualization environment, applications inside the virtual machines can also be put into a consistent state before the backup starts.

Besides the key advantage of the virtual machine being online all the time there are also other benefits of using the Data Protector integrations with the databases:

- There is no need to specify the location of data files.
- Virtualization environments are aware of the backup operation and keep track of which parts have been backed up.
- Several modes of backup are possible.
- Several modes of restore are possible.

Q&A

Thank you