TRIM Integration with Data Protector

Table of Contents

Introduction

This paper explains how HP Data Protector can be used to back up an HP TRIM system.

It is intended for TRIM system administrators who have no knowledge of Data Protector but have a reasonable knowledge of TRIM's architecture and data organization.

Prerequisites

For the practical parts in the second half of this paper, you need access to a running TRIM system or the knowledge of how to set one up.

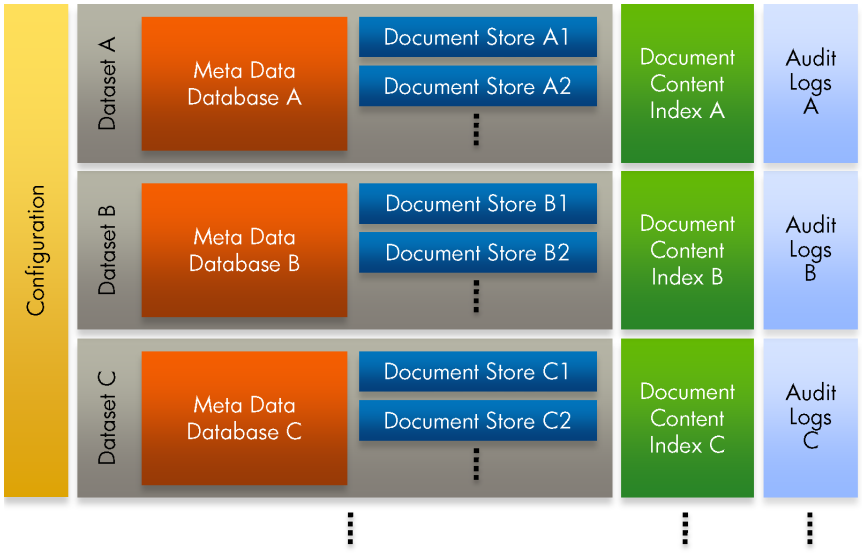
TRIM Internals

You can derive your backup strategy directly from TRIM's architecture and data organization. Since TRIM's architecture is quite flexible, this section gives an example configuration of a TRIM system and a corresponding sample backup.

TRIM's Data Organization

Figure 1 illustrates TRIM's data organization.

Figure 1: TRIM's Data Organization



The core components of TRIM are its *Datasets*. All Datasets are configured in one global configuration file.

Each Dataset is identified by a Dataset Identifier and Dataset Name, and defines the following:

- The type and location of a Metadata Database

  Metadata Databases store all the records, retention rules, workflows, locations and so on needed for the TRIM system. These databases are the most important data. Without them, the Document Stores are meaningless to the system.

- One or more Document Stores

  Document Stores contain the actual files linked to individual records. The physical locations of the stores are defined in the corresponding Metadata Database.

For each Dataset, you can also instruct TRIM to maintain the following:

- A Document Content Index.

  The index facilitates fast indexed searches. It can always be rebuilt from the Document Stores, but this can take a considerable amount of time. The location of the index is held in the global configuration file.

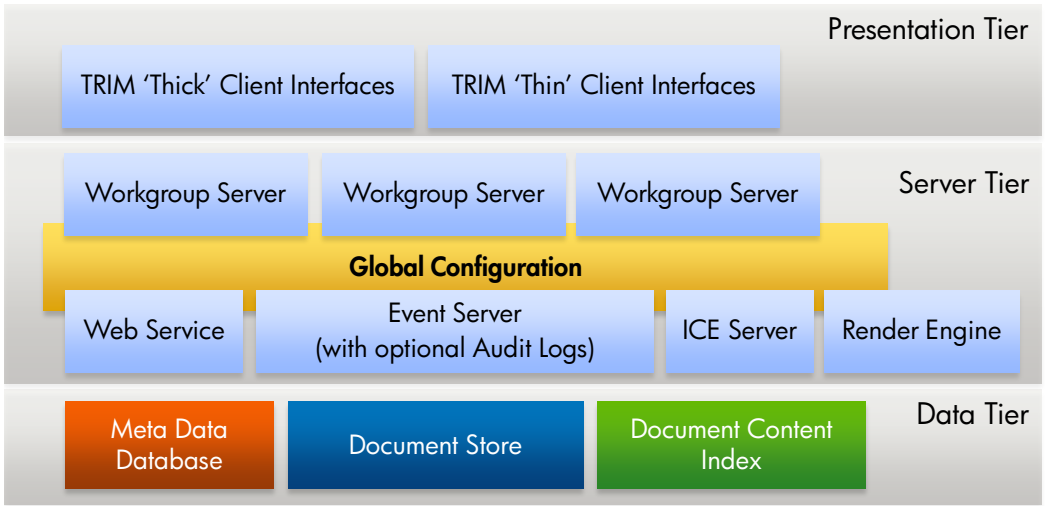
- Audit Logs.

The logs store information about every transaction made in TRIM. By default they are created on an Event Server. All this information is specified in one global configuration file.

TRIM's Architecture

Figure 2 illustrates how these data parts fit into TRIM's architecture. A TRIM system is divided into three *Tiers*: the *Data Tier*, the *Server Tier* and the *Presentation Tier*. The Datasets and associated elements mainly belong to the Data Tier of TRIM's architecture, while the configuration and Audit Logs are stored within the Server Tier.

Figure 2: TRIMS's Architecture



Every block in this figure represents one or more servers (some of which are optional).

The global configuration file resides on all the servers in the server tier, so it is sufficient to back it up only from one of the servers. While not mandatory, maintaining a backup of the configuration file significantly eases the restoration procedure.

Implications for Backup

The Data Tier is the most important part from a backup perspective, and all its components should be backed up regularly. In addition, it is highly recommended to back up the global configuration file and the Audit Logs.

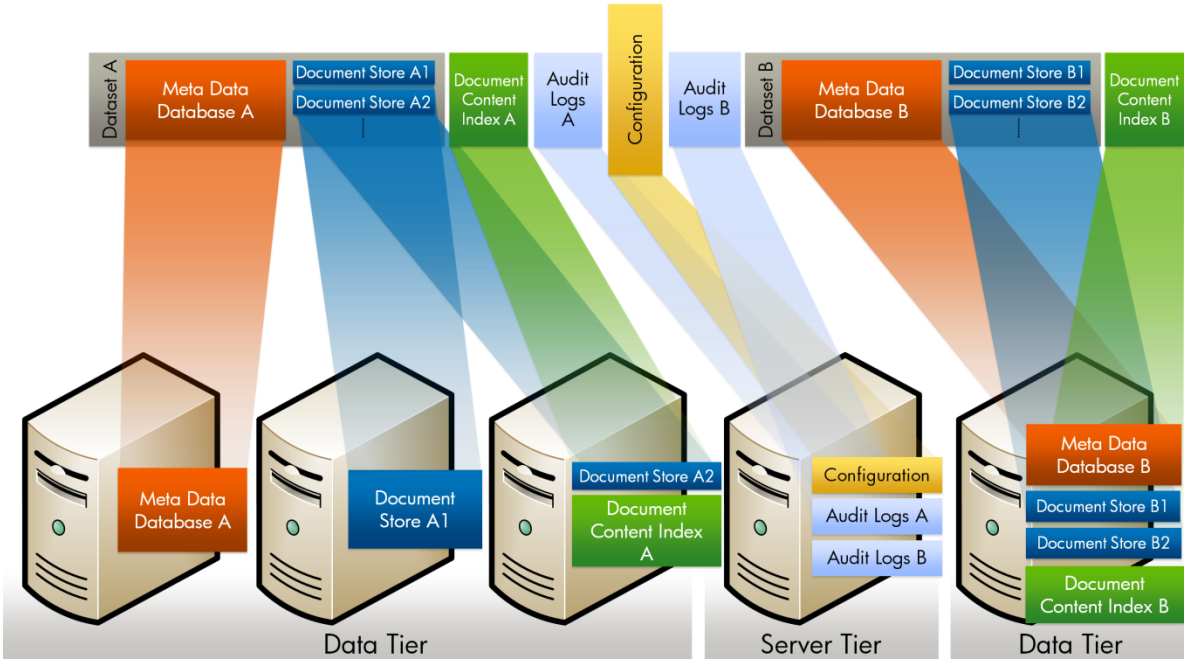
Sample Configurations

Sample Configuration of a TRIM System

As an example to explain how to design the backup strategy, Figure 3 shows the part of a sample TRIM system that is relevant for backups.

Dataset A is distributed over several hosts. Dataset B is completely set up on one host. Different Datasets could also share hosts.

Figure 3: Example Configuration of a TRIM System

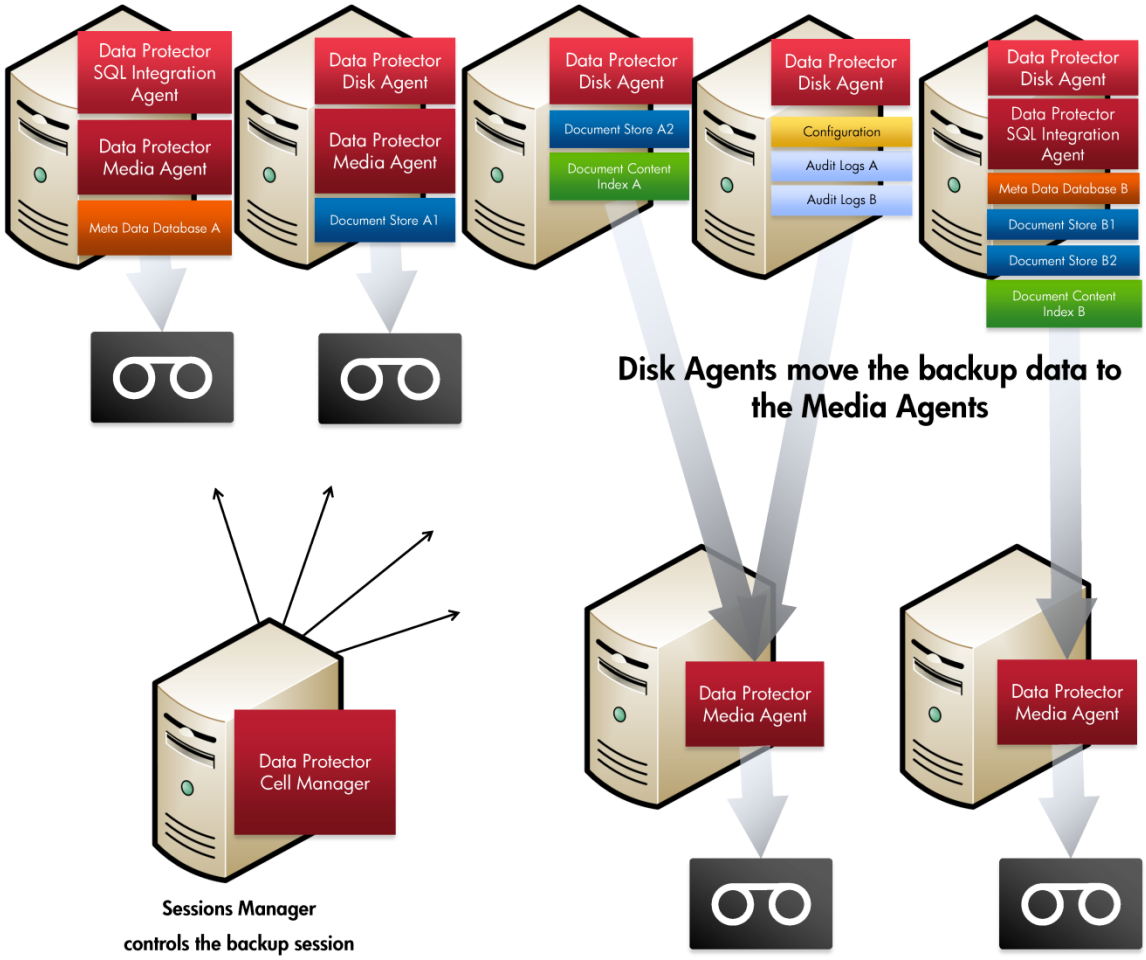


Sample Backup Configuration

A possible backup approach is illustrated in Figure 4. A setup of Data Protector requires a Disk Agent on every machine that contains data as files and an SQL Integration Agent on every machine that contains databases. This Disk Agent/SQL Integration Agent transfers all the data to a Media Agent which then writes it to the backup medium (such as tape). All actions are managed by the Cell Manager, which in this case is installed on a separate machine. It could also be installed on one of the other machines, alongside the Disk Agents, Integration Agents or Media Agents.

When the schedule in the Cell Manager requires a backup to be triggered, the Session Manager is started, which in turn communicates with the Disk Agents, SQL Integration Agents and Media Agents and instructs them to connect to each other.

Figure 4: Example Backup Configuration



Step-by-Step Guide for a TRIM Backup

The following sections use the concepts explained in the previous sections to explain how to accomplish a backup using Data Protector step-by-step, including installation of Data Protector itself. Make sure that you have familiarized yourself with these concepts, otherwise the steps in the following sections will not be clear.

Installing Data Protector

Installing Data Protector is explained in detail in the *Data Protector Installation and Licensing Guide* and will not be covered here. However, you can use Figure 4 to determine the components that need to be installed on the individual hosts. For example:

- A Cell Manager is required in the network to control all operations.
- A Disk Agent is required on every host containing data to be backed up.
- A Media Agent is needed on every host serving as a backup target (and which might be connected to a tape library or similar).

Depending on the database used as a Metadata Database, you need to install additional components on the relevant Disk Agents to ensure proper database backups. Data Protector has off-the-shelf integrations for Oracle, MS SQL Server 2005, DB2 and Sybase.

Figure 5 and Figure 6 show an example of how to install these integrations. In the example, we assume that the TRIM installation uses a MS SQL server as Metadata database. To install these components using *push installation,* make sure that you have installed the DP Installation Server.

Figure 5: Installing additional components on a client

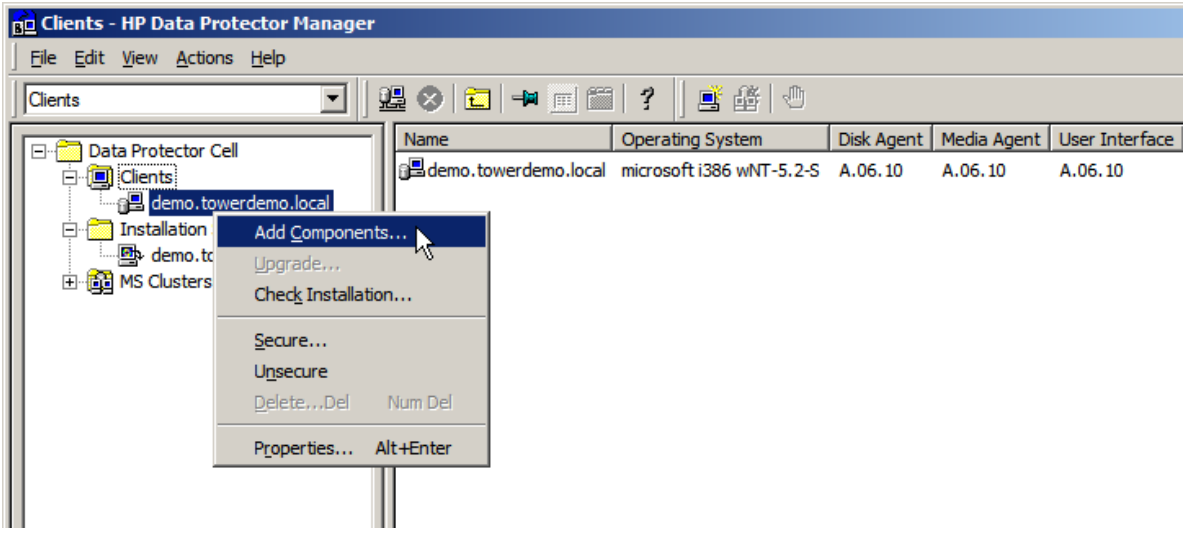
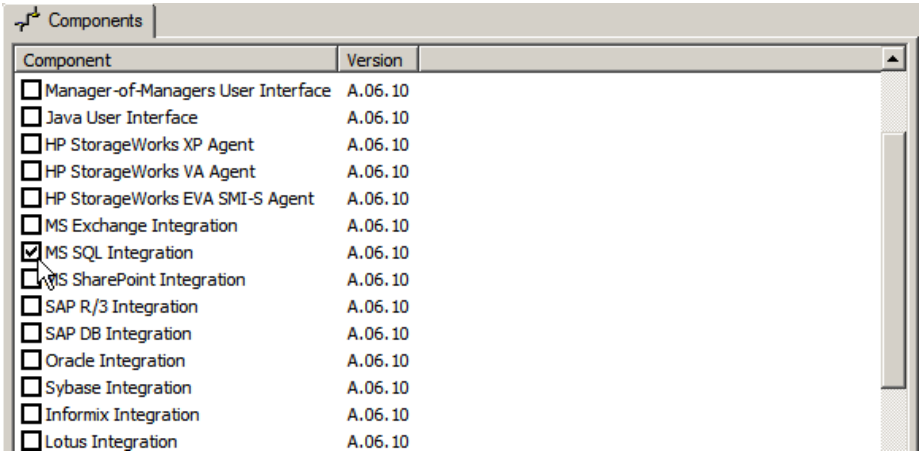


Figure 6: installing MS SQL Server Integration



Preparing Data Protector for Backup

The backup preparation process is divided into three tasks:

- Setting up the target devices
- Finding the locations of TRIM data components
- Creating backup specifications for the different data components

The following is a detailed description of each of the tasks.

Setting up Devices

To prepare Data Protector for backup, first set up logical devices on the Media Agents to serve as targets for backups. A logical device is one abstraction layer above the real backup device (such as a tape library) and helps to define different ways to access one individual device.

To create devices, switch to the Devices & Media context, right-click on Devices and then click on **Add Device**. The rest of the procedure largely depends on the backup hardware capabilities available in the environment. Refer to the *Data Protector Concept Guide*, chapter 3 "Media management and devices" for more information on logical devices.

Finding the Locations of Data Components

Before you can set up a backup specification that contains data source locations and target devices you need to obtain the data source location from TRIM.

1. To retrieve **types and locations of Metadata Databases** using TRIM Enterprise Studio, select the properties of the relevant Datasets as shown in Figure 7. In the Connection tab, the type and location of the database are defined, as shown in
2. Figure 8. In this case it is an MS SQL Server located on the host "DEMO" and the database's name is "Corp".

Figure 7: TRIM Enterprise Studio

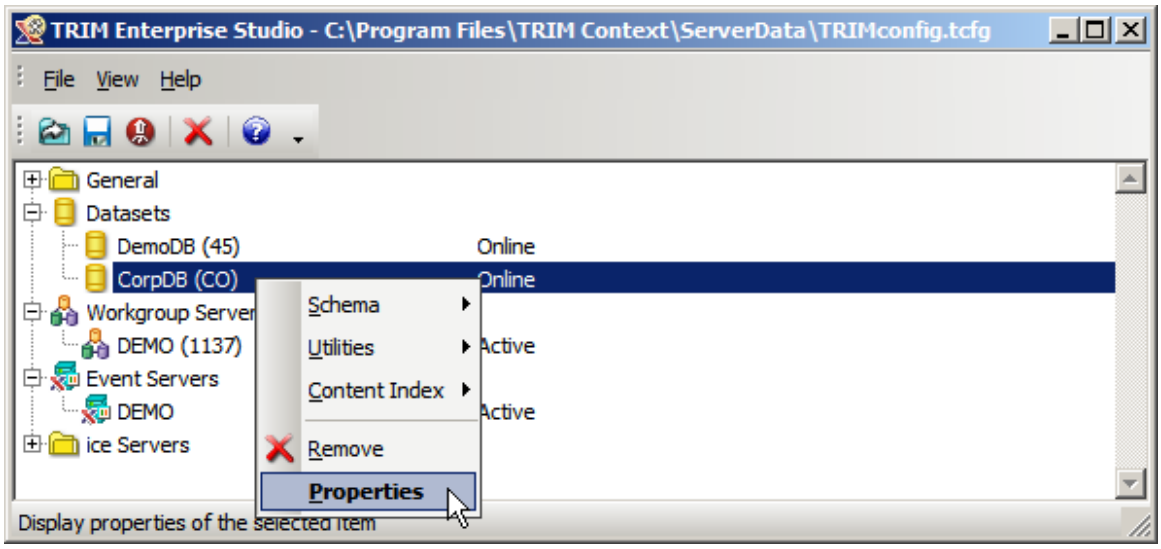
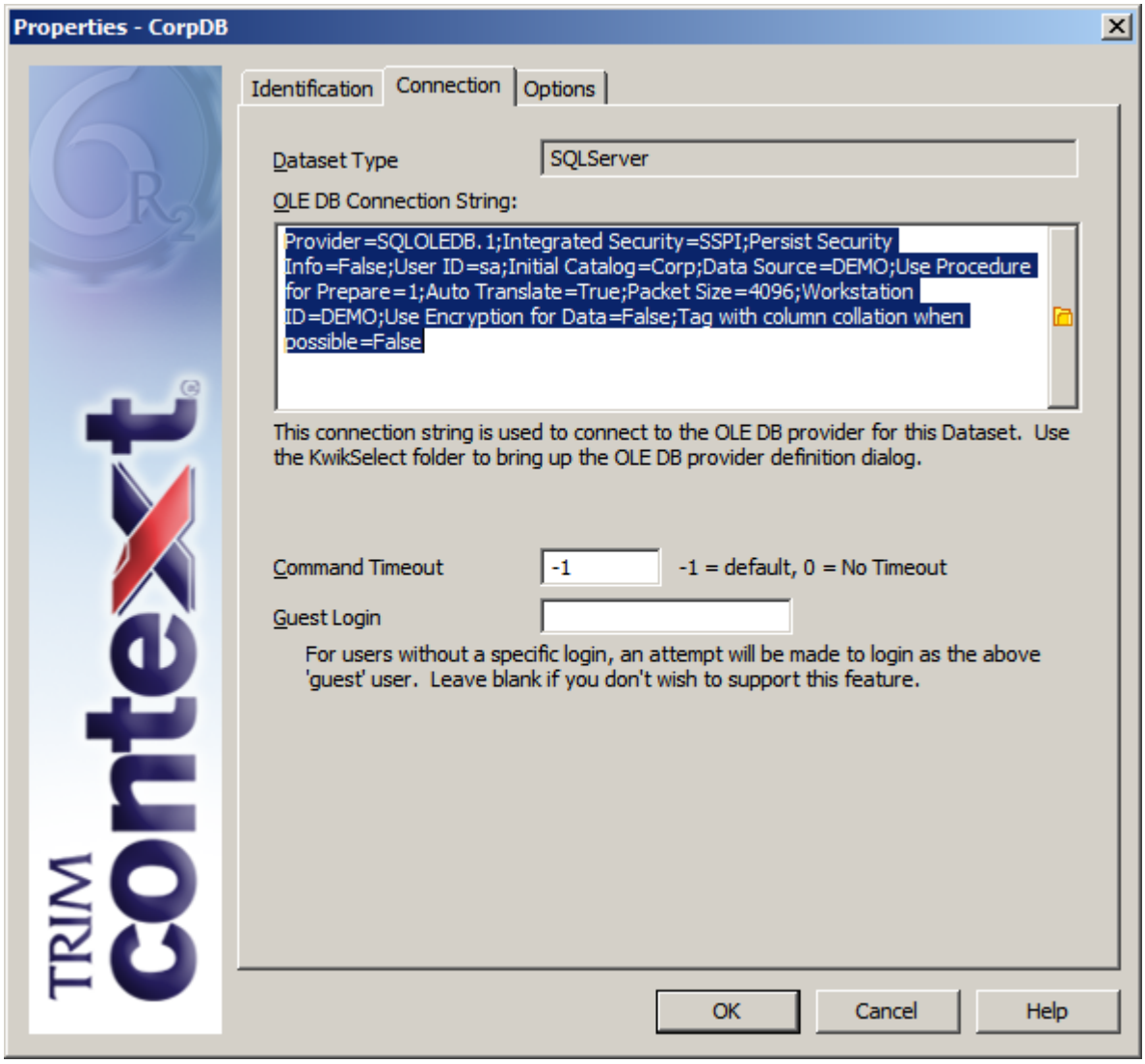
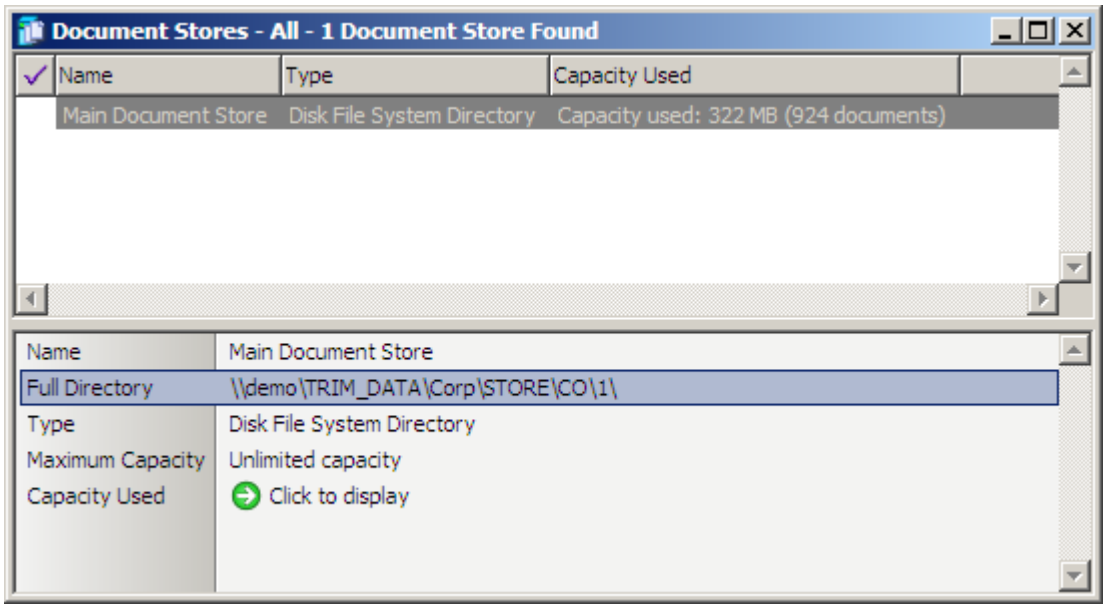


Figure 8: Database type and location



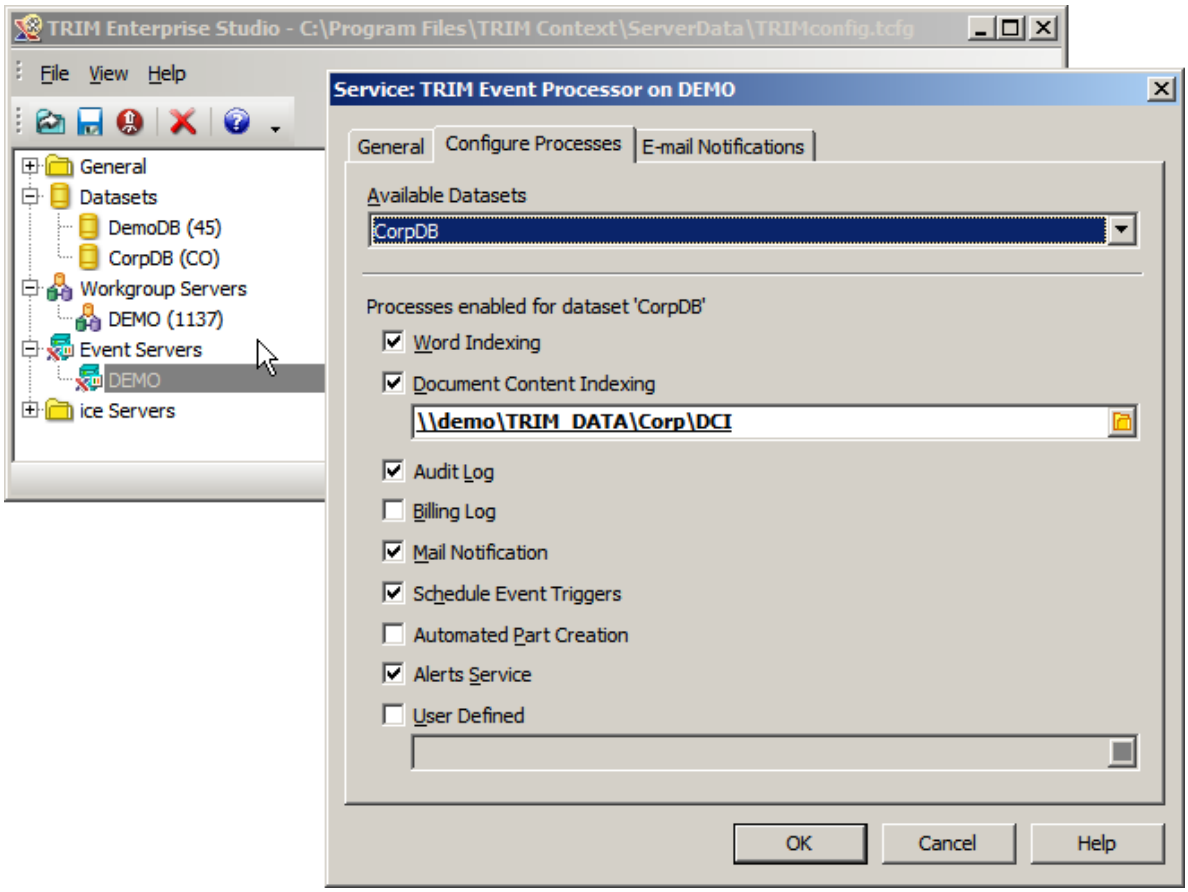
3. To find the **locations of the Document Stores**, open the TRIM client with an administrator login and access the store management through **Tools → Context Administration → Document Store…**. The Full Directory field will tell you the exact location of each store. Here it is \\demo\TRIM_DATA\Corp\STORE\CO\1 (see Figure 9).

Figure 9: Location of Document Stores



4. To find the **location of the Document Content Index for each Dataset,** open properties of your event server as shown in Figure 10, and select the Configure Processes tab. If an index is configured, the complete path is displayed in the Document Content Indexing field. In this case it is \\demo\TRIM_DATA\Corp\DCI\CO, which is the path we will use in our example.

Figure 10: Properties of Event Server



5. If Audit Log is enabled, as shown in Figure 10, by default the **location of the Audit Logs** is on the Event Server in \Program Files\TRIM Context\ServerData\<Dataset Identifier>\AuditLogs\. In this case it is \\demo\Program Files\TRIM Context\ServerData\CO\AuditLogs\.
6. The **global configuration file** can be found on every server in the Server Tier in \Program Files\TRIM Context\ServerData\TRIMconfig.tcfg.

Creating a Backup Specification for the Database

You can now create a backup specification for the Metadata Database using the location obtained in the previous section. In our example, an MS SQL Server 2005 is used as the Metadata Database. For this, as well as for other databases (Oracle, DB2, Sybase, and so on), Data Protector provides special integrations to handle database backups properly. Within these integrations there are multiple methods for backing up the database. Deciding which method to use depends on the particular circumstances. Consult the appropriate Data Protector guide for information:

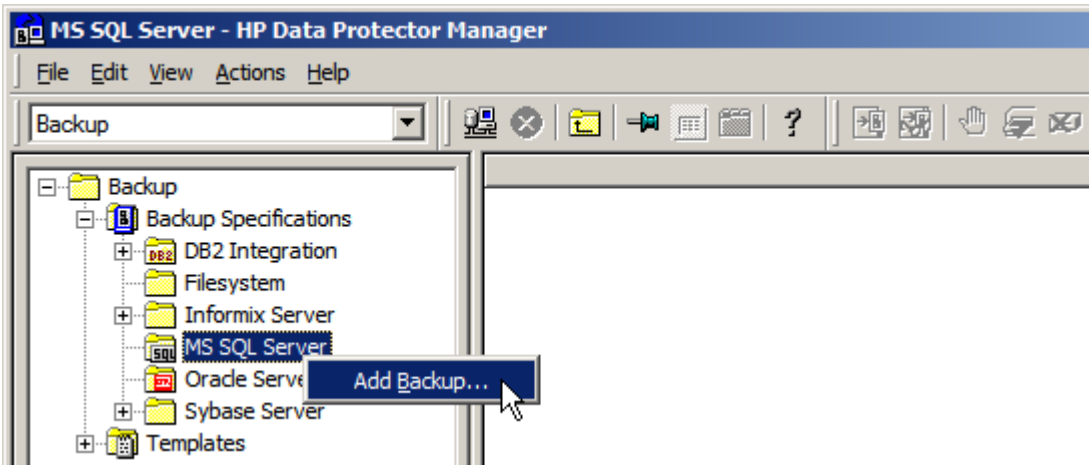
- *Integration Guide for Microsoft Applications*
- *Integration Guide for Oracle and SAP*
- *Integration Guide for IBM Applications*

- *Integration Guide for VMware, Sybase, NNM and NDMP*

To create a backup specification:

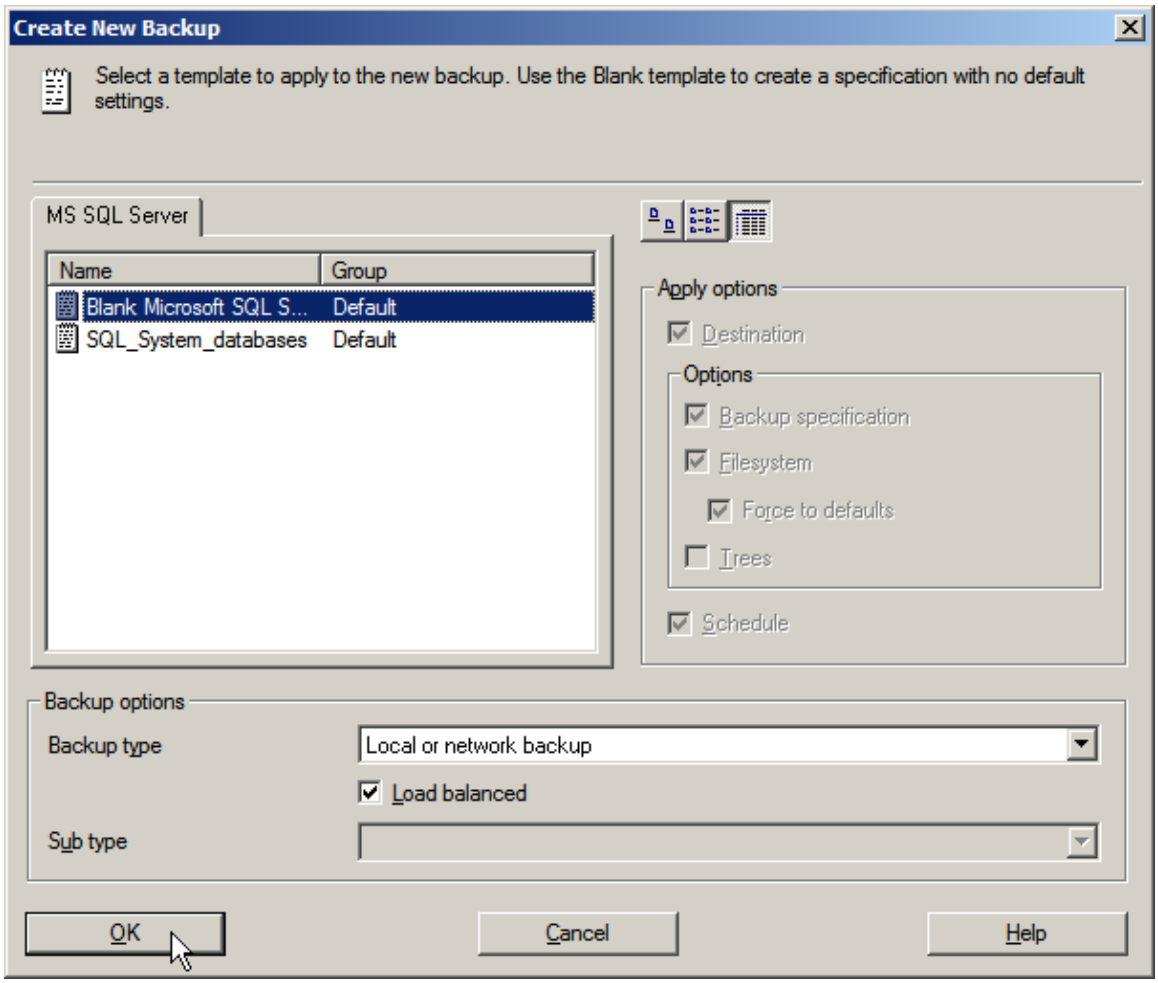
1. In Data Protector, in the Context List, click **Backup**.
2. In the Scoping Pane, expand **Backup Specifications**.
3. Right-click the type of item you want to back up (MS SQL Server in our example), and click **Add Backup**. See Figure 11.

Figure 11: MS SQL Server Add Backup



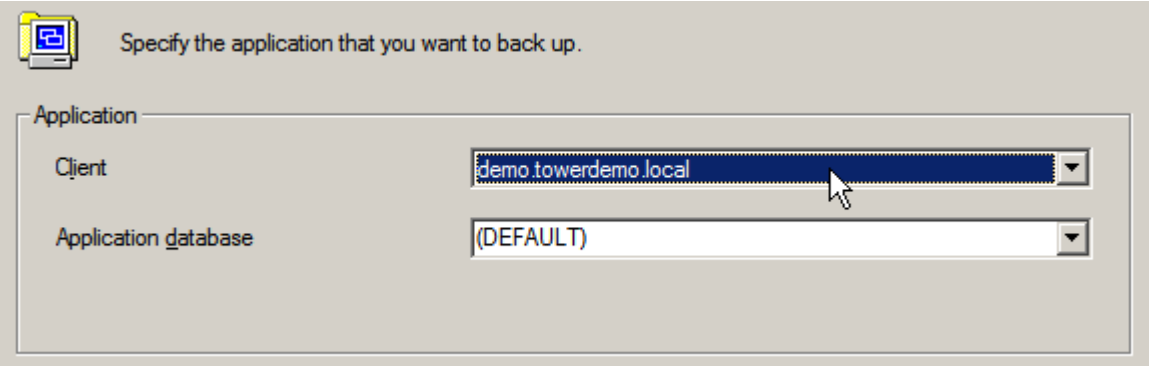
4. In the Create New Backup dialog box, leave the defaults and click **OK**. See Figure 12.

Figure 12: Create New Backup Dialog



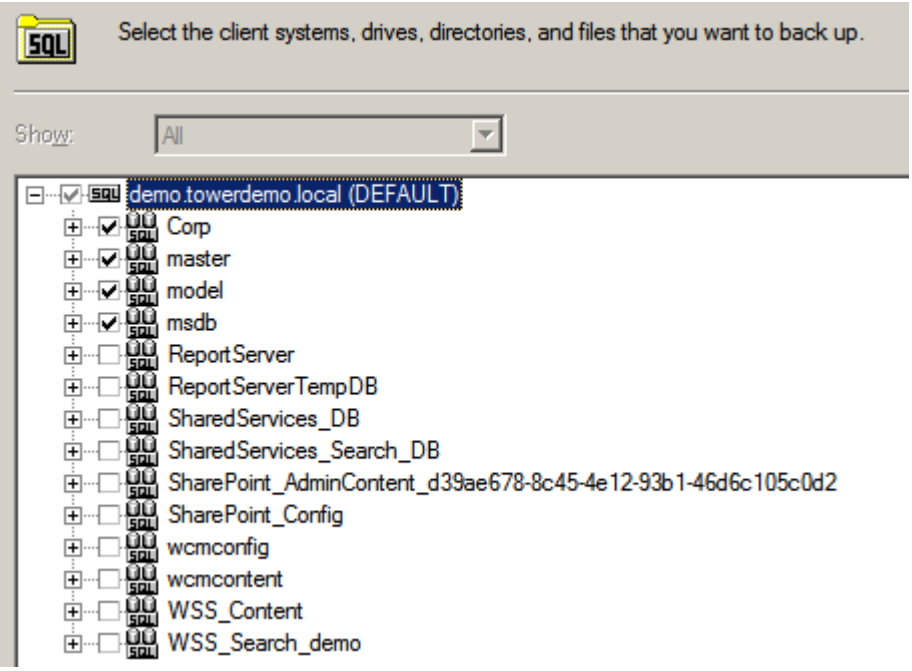
5. Select the appropriate client and the application database and click **Next**. See Figure 13.

Figure 13: Choosing the database



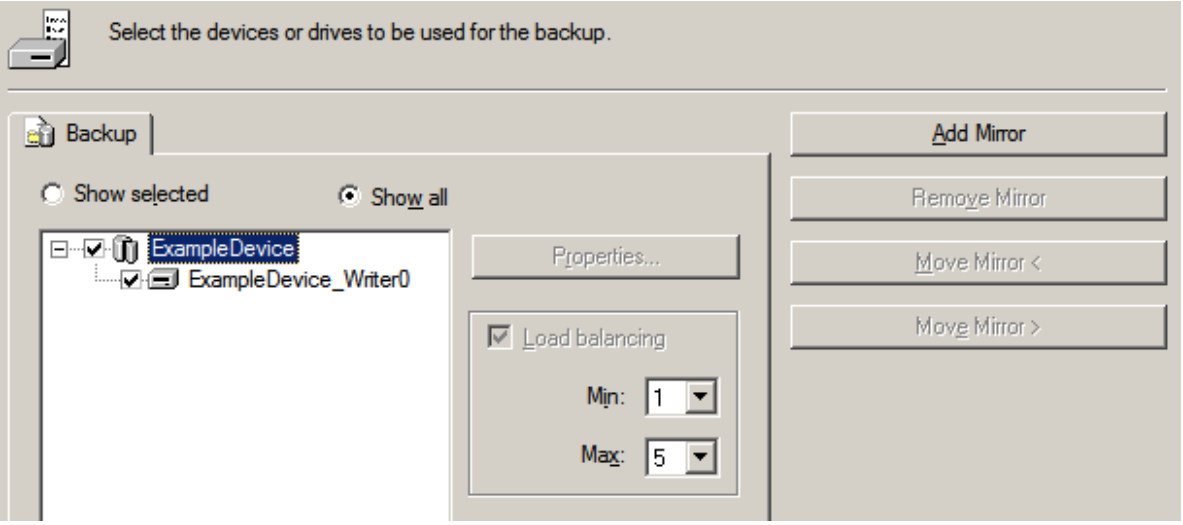
6. In the Source property page, expand the system that contains the objects that you want to back up and the select the sources you want backed up. See Figure 14.

Figure 14: Adding backup specification: select sources



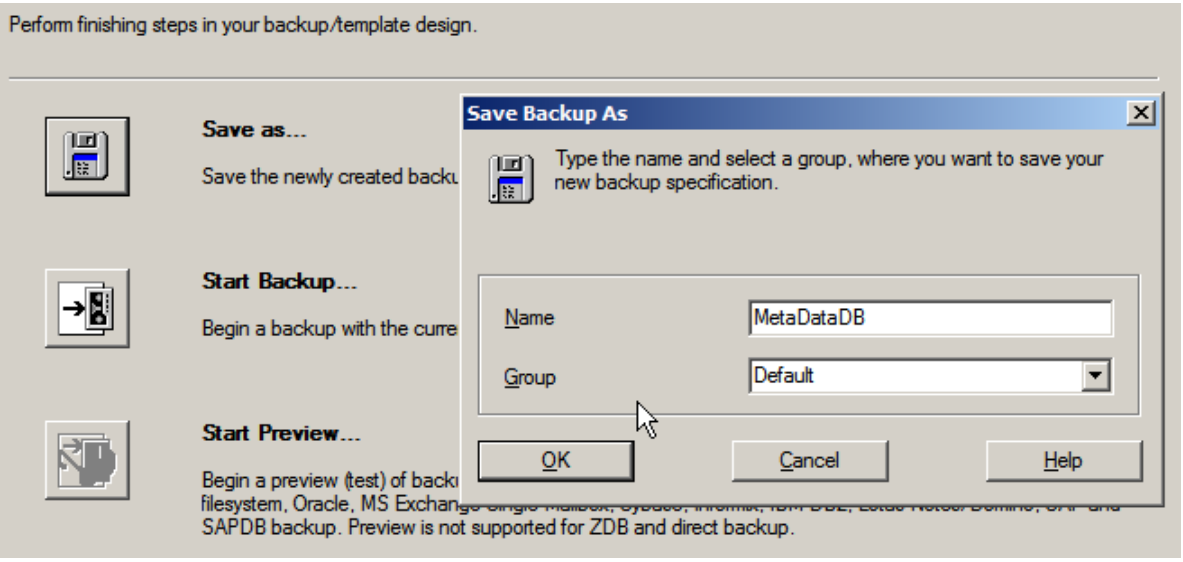
7. In addition to the Metadata Database itself (here "Corp") it is highly advisable to back up the system databases up as well. For further considerations concerning the system databases see the *Data Protector Integration Guide for Microsoft Applications*.
8. In the Destination property page, select the device you will use for your backup. See
9. Figure 15. Devices should have already been set up as described above.

Figure 15: Adding backup specification: choose device



10. The remaining steps are the same as usual. Refer to the *Data Protector Concept Guide*, Chapter 2, "Planning your backup strategy" for more information.
11. Finally save the backup specification, specifying an appropriate name. See Figure 16.

Figure 16: Saving the backup specification



Creating a Backup Specification for the Document Store, Document Content Index, Audit Logs and Global Configuration File

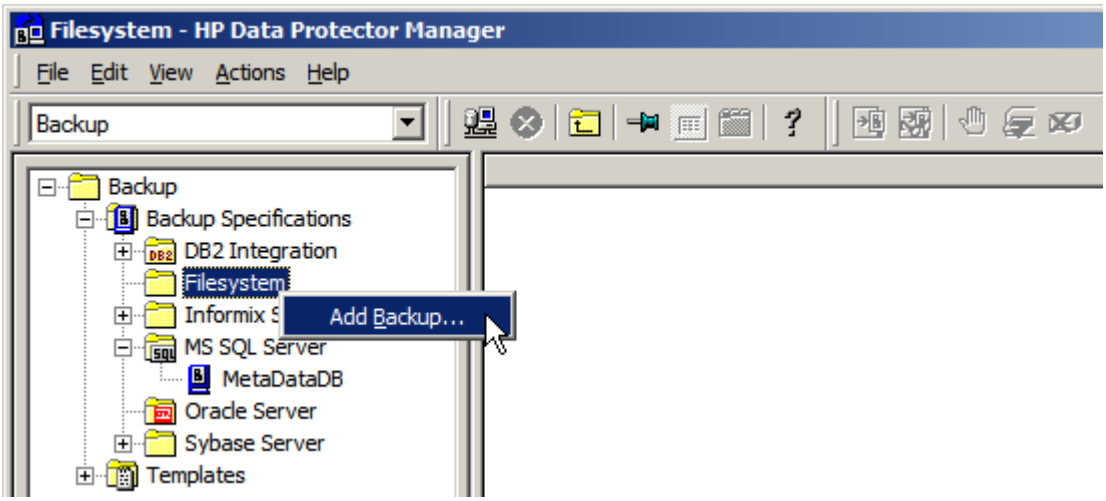
Creating a backup specification for the Document Store, Document Content Index, Audit Logs and global configuration file is similar to creating a database backup specification, except that this time the sources are flat file systems instead of a database.

In this paper, for simplicity, the Document Store, Document Content Index, Audit Logs and global configuration file will be backed up using one backup specification. In most cases it is better to put these into separate specifications, for example, if the different components have different schedules. See also "Further Considerations" later in this paper on this subject.

To create a backup specification for file systems:

1. Right-click on Filesystem in the Backup Context, and click **Add Backup**. See Figure 17.

Figure 17: Adding a backup specification



2. In the Create New Backup dialog box, leave the defaults and click **OK**.
3. Figure 18 shows the file tree structures. Select which directories and files Data Protector should back up, as shown in Figure 19.

Figure 18: Selecting sources for the backup

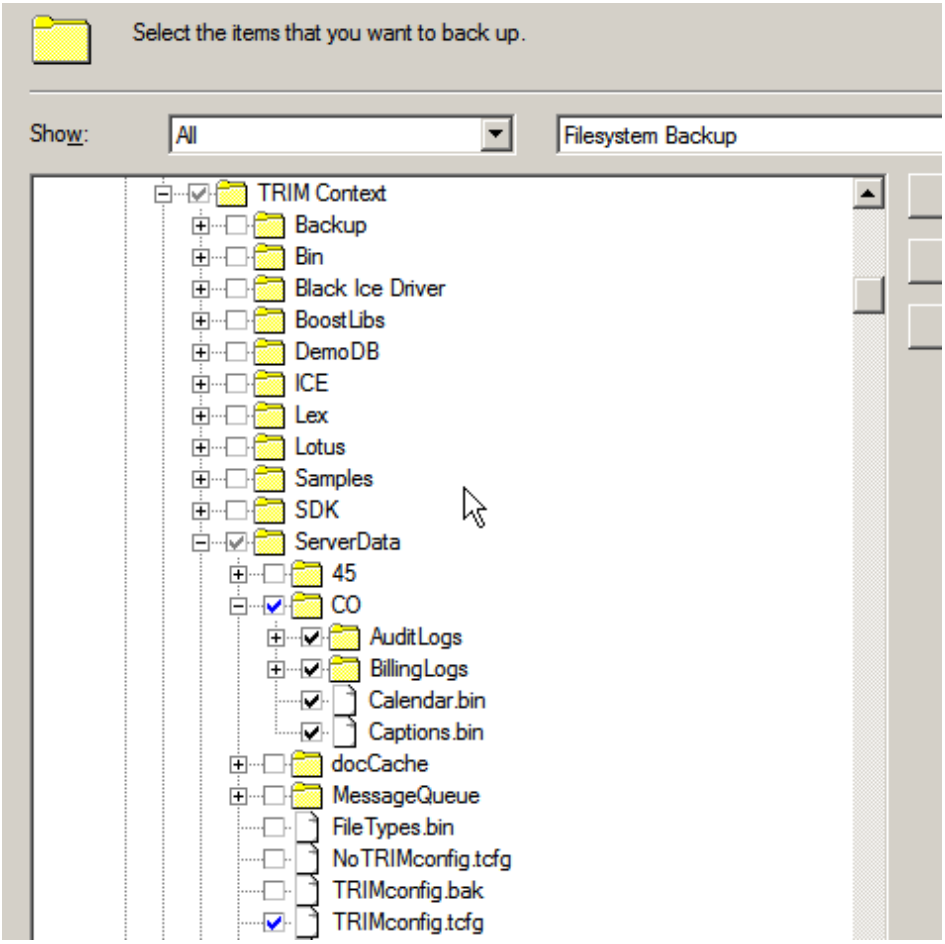
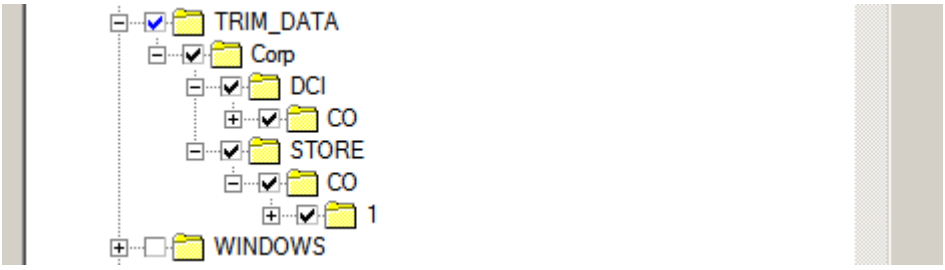


Figure 19: Selecting sources for the backup

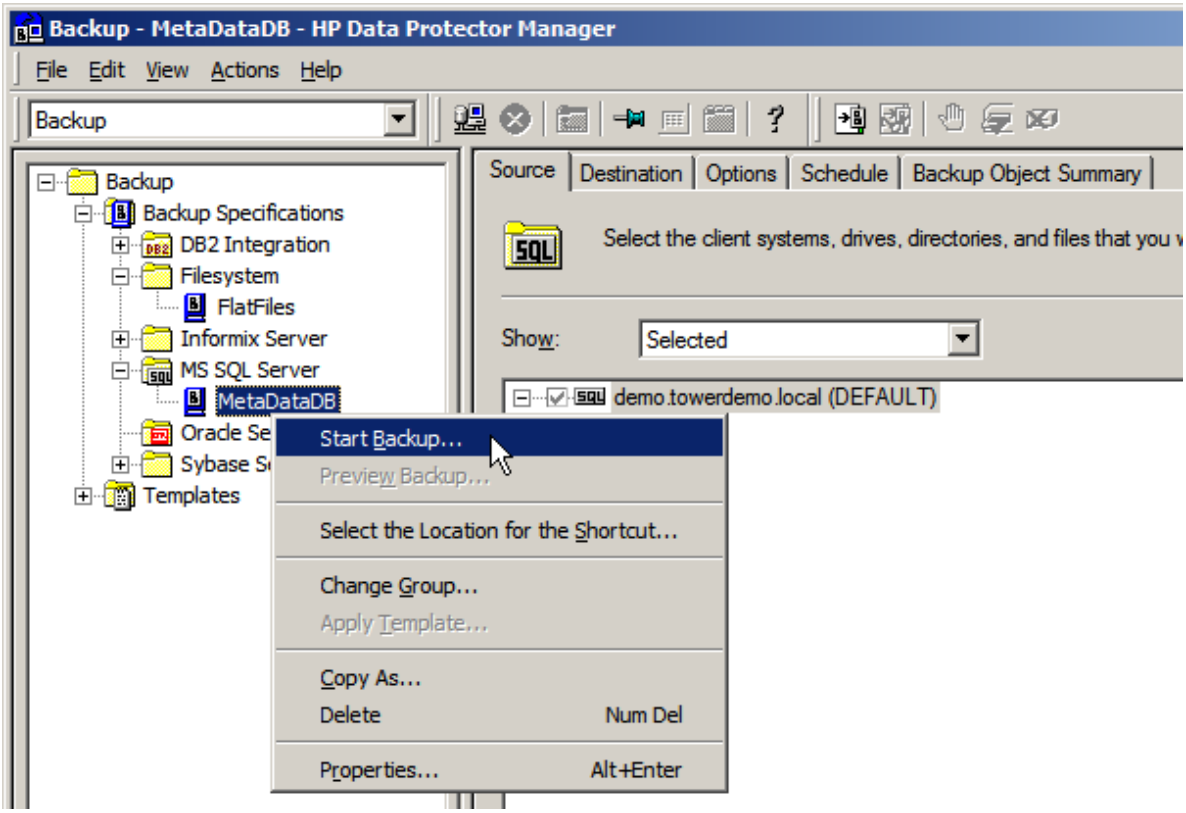


4. The remaining steps are as usual. Again the final step is to save the backup specification with an appropriate name.

Performing a Backup

Backups are either triggered automatically if specified in the backup specification or you can start them manually by right-clicking each backup specification and selecting **Start Backup** as shown in Figure 20.

Figure 20: Starting a backup



For more information on the options in the Start Backup dialog (Figure 21), see the Data Protector online Help index "starting backup sessions". In this paper a full backup will be performed.

For each backup that is performed, Data Protector should show a backup session output similar to the one in Figure 22. **When backups have been successfully performed for both the database and the file systems the backup is complete.**

Figure 21: Starting a Backup dialog

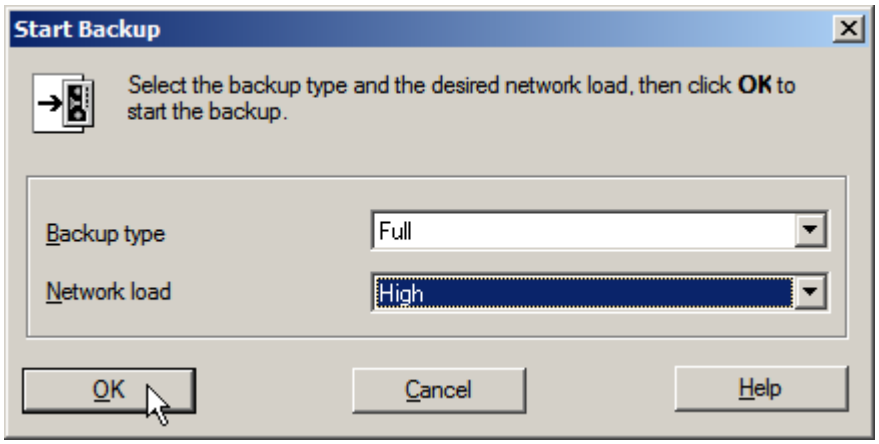


Figure 22: Backup session output

```
[Normal] From: BSM@demo.towerdemo.local "FlatFiles"  Time: 3/17/2009 5:53:57 AM
      Backup session 2009/03/17-1 started.

[Normal] From: BMA@demo.towerdemo.local "ExampleDevice_Writer0"  Time: 3/17/2009 5:54:04 AM
      STARTING Media Agent "ExampleDevice_Writer0"

[Normal] From: BMA@demo.towerdemo.local "ExampleDevice_Writer0"  Time: 3/17/2009 5:54:05 AM
      Loading medium from slot C:\FileLibrary\8185a8c0549bf733d50a8850001.fd to device ExampleDevice_Writer0

[Normal] From: BMA@demo.towerdemo.local "ExampleDevice_Writer0"  Time: 3/17/2009 5:54:05 AM
      C:\FileLibrary\8185a8c0549bf733d50a8850001.fd
      Initializing new medium: "ExampleDevice_MediaPool_3"

[Normal] From: VBDA@demo.towerdemo.local "C:"  Time: 3/17/2009 5:54:11 AM
      STARTING Disk Agent for demo.towerdemo.local:/C "C:".

[Normal] From: VBDA@demo.towerdemo.local "C:"  Time: 3/17/2009 5:54:15 AM
      COMPLETED Disk Agent for demo.towerdemo.local:/C "C:".

[Normal] From: BMA@demo.towerdemo.local "ExampleDevice_Writer0"  Time: 3/17/2009 5:54:15 AM
      Unloading medium to slot C:\FileLibrary\8185a8c0549bf733d50a8850001.fd from device ExampleDevice_Writer0

[Normal] From: BMA@demo.towerdemo.local "ExampleDevice_Writer0"  Time: 3/17/2009 5:54:15 AM
      COMPLETED Media Agent "ExampleDevice_Writer0"

[Normal] From: BSM@demo.towerdemo.local "FlatFiles"  Time: 3/17/2009 5:54:15 AM

      Backup Statistics:

            Session Queuing Time (hours)          0.00
            -------------------------------------------
            Completed Disk Agents ........           1
            Failed Disk Agents ...........           0
            Aborted Disk Agents ..........           0
            -------------------------------------------
            Disk Agents Total  ...........           1
            ===========================================
            Completed Media Agents .......           1
            Failed Media Agents ..........           0
            Aborted Media Agents .........           0
            -------------------------------------------
            Media Agents Total  ..........           1
            ===========================================
            Mbytes Total .................       11 MB
            Used Media Total .............           1
            Disk Agent Errors Total ......           0


================================================================================
                     Session completed successfully!
================================================================================
```

Performing a Restore

Restoring from data loss is done in four steps:

1. Reinstall the necessary applications (TRIM Context and/or the database application, such as MS SQL Server) on the affected hosts.
2. Restore the lost data.
3. If necessary, perform actions to get consistent Datasets.
4. Restart the TRIM services.

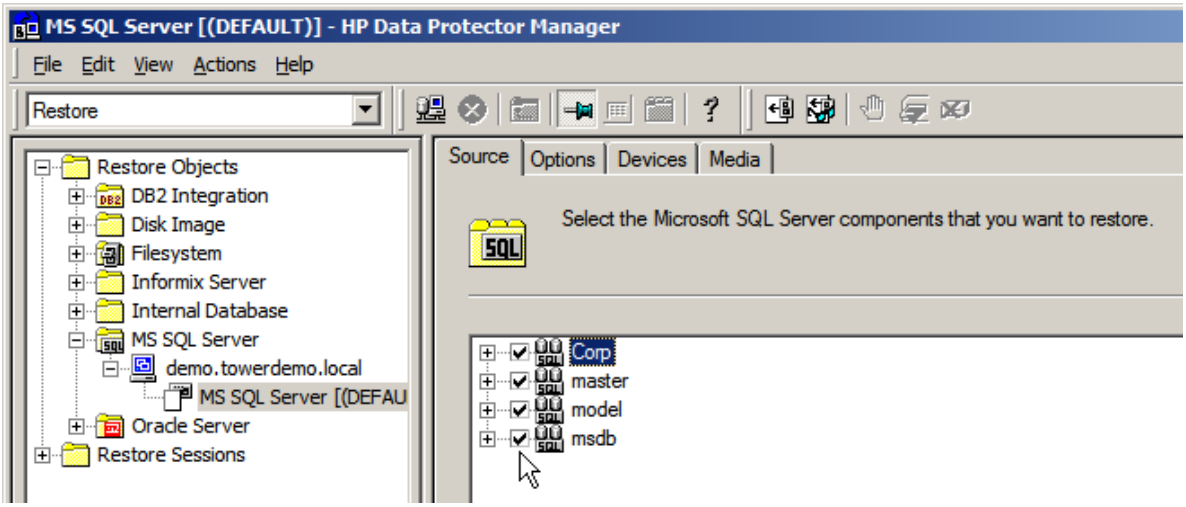
It is assumed that you know how to reinstall the necessary applications, so the following assumes that pristine installations of the applications exist.

Since Step 3 is only necessary in certain cases it is discussed in detail in “Further Considerations” later in this paper.

Restoring the Metadata Database

1. In the Context List, click **Restore**.
2. In the Scoping Pane, under Restore Objects, expand the appropriate data type (for example, **MS SQL Server**).
3. Expand the client system with the data you want to restore and then click the object that has the data.
4. In the Source property page, expand the object and then select all the databases to be restored, as shown in Figure 23.
5. Click **Restore**.

Figure 23: Restoring the database



This will only restore the Metadata Database. However, TRIM will not know about the database until the global configuration file is restored as well.

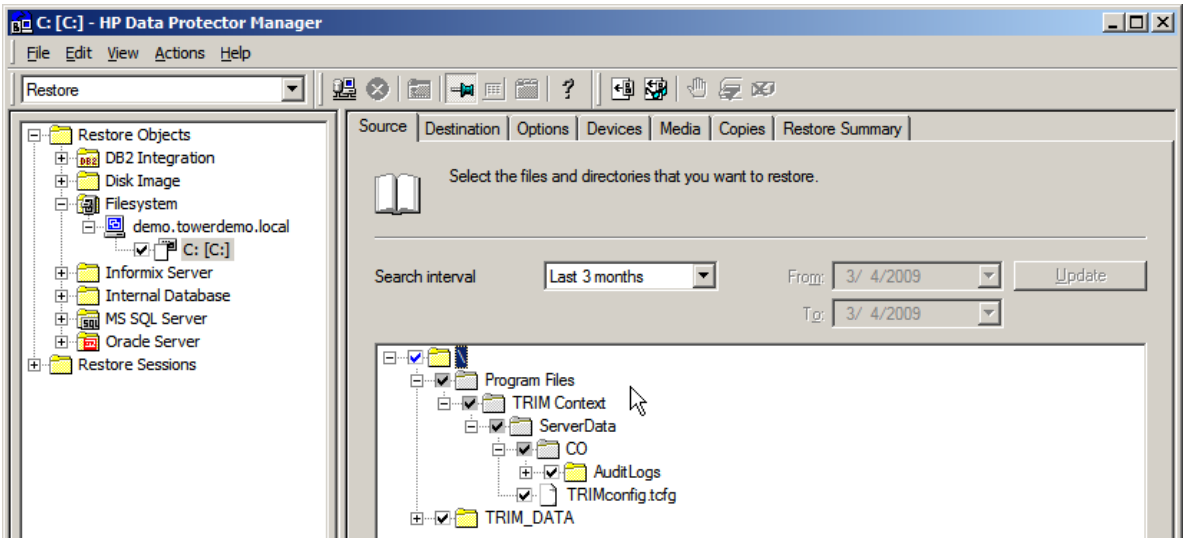
Consult the appropriate Data Protector guide for more information:

- *Integration Guide for Microsoft Applications*
- *Integration Guide for Oracle and SAP*
- *Integration Guide for IBM Applications*
- *Integration Guide for VMware, Sybase, NNM and NDMP*

Restoring the Document Store, Document Content Index, Audit Logs and Global Configuration File

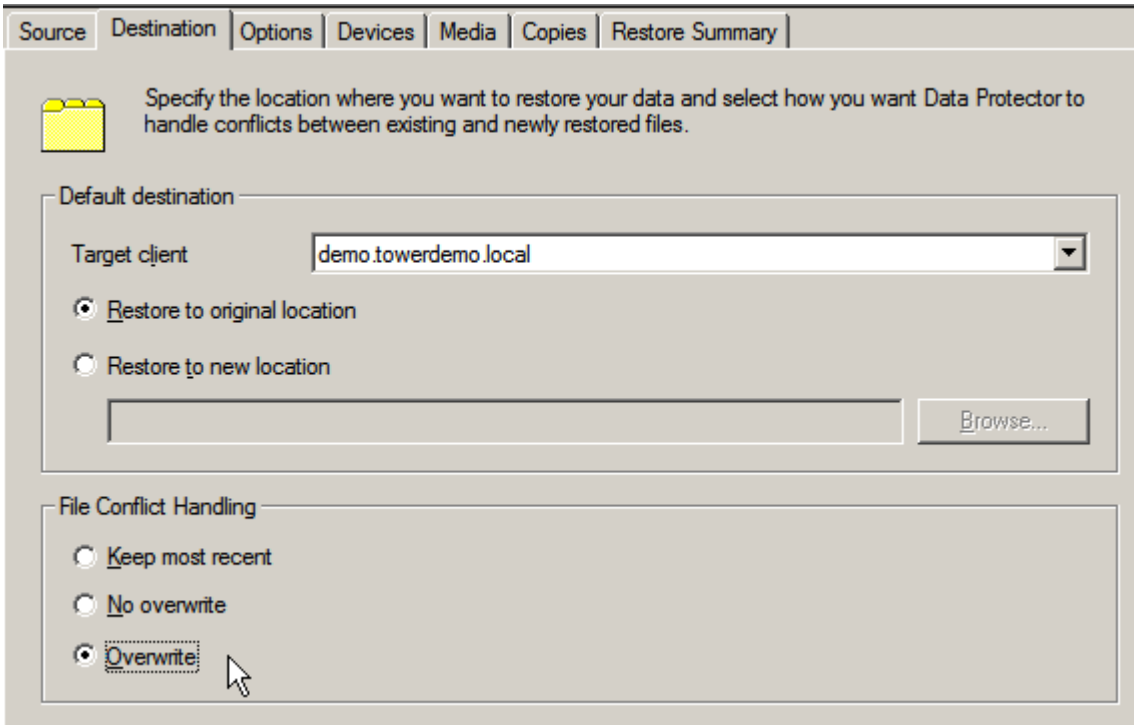
1. In the Scoping Pane, expand the Filesystem branch until the desired backup object appears.
2. In the Source property page, expand the object and select directories and files that you want to restore. In this example, everything will be restored. See Figure 24.

Figure 24: Selecting restore sources



3. Open the Destination tab, and select **Overwrite** in the File Conflict Handling options. This will ensure that the global configuration file, automatically installed while reinstalling TRIM, will be overwritten by the backed-up version. Otherwise Data Protector will keep the new one.

Figure 25: Conflict handling



If everything performs correctly, Data Protector should show a restore session output similar to that in Figure 26.

Figure 26: Restore Session Output

```
[Normal] From: RSM@demo.towerdemo.local ""  Time: 3/4/2009 11:24:14 AM
      Restore session 2009/03/04-4 started.

[Normal] From: RMA@demo.towerdemo.local "ExampleDevice_Writer0"  Time: 3/4/2009 11:24:14 AM
      STARTING Media Agent "ExampleDevice_Writer0"

[Normal] From: RMA@demo.towerdemo.local "ExampleDevice_Writer0"  Time: 3/4/2009 11:24:15 AM
      Loading medium from slot C:\FileLibrary\010aa8c0549ae854b514a850004.fd to device ExampleDevice_Writer0

[Normal] From: VRDA@demo.towerdemo.local "C:"  Time: 3/4/2009 11:24:21 AM
      STARTING Disk Agent for demo.towerdemo.local:C:

[Warning] From: VRDA@demo.towerdemo.local "C:"  Time: 3/4/2009 11:24:23 AM
      C:\Program Files\TRIM Context\ServerData\TRIMconfig.tcfg
      Option OVERWRITE => Old file will be replaced !

[Normal] From: VRDA@demo.towerdemo.local "C:"  Time: 3/4/2009 11:24:23 AM
      Share information successfully restored for directory.
      C:\TRIM_DATA


[Normal] From: VRDA@demo.towerdemo.local "C:"  Time: 3/4/2009 11:26:17 AM
      COMPLETED Disk Agent for demo.towerdemo.local:C:

[Normal] From: RMA@demo.towerdemo.local "ExampleDevice_Writer0"  Time: 3/4/2009 11:26:24 AM
      Unloading medium to slot C:\FileLibrary\010aa8c0549ae854b514a850004.fd from device ExampleDevice_Writer0

[Normal] From: RMA@demo.towerdemo.local "ExampleDevice_Writer0"  Time: 3/4/2009 11:26:24 AM
      COMPLETED Media Agent "ExampleDevice_Writer0"


================================================================================
                      Session completed successfully!
================================================================================
```

Restarting Services

Now that all data and the global configuration file are restored, you must tell TRIM to use the restored data and configuration by restarting TRIM's services. To do this, select **Start → Administrative Tools → Services** and then restart all TRIM services.

TRIM should now be fully recovered.

Further Considerations

Backup Frequencies

In planning a backup environment, you need to balance the protection provided by frequent backups against the impact on performance. The following two scenarios illustrate this.

High Protection Using High Frequency Backups

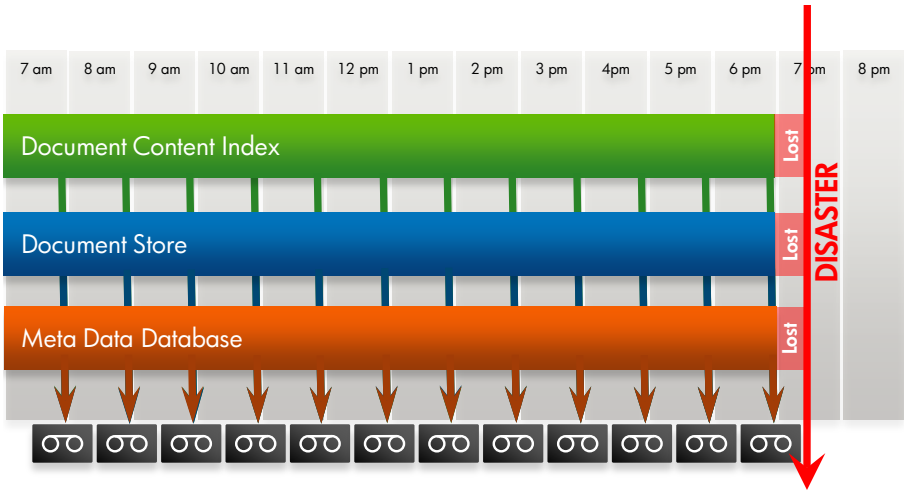
This scenario is recommended for systems that need to reduce the risk of data loss to a minimum. The scheduler is configured to do an *hourly* backup of *all* TRIM components.

High frequency backups have a high impact on the system that is to be backed up. To ensure that normal TRIM services are not affected, the environment running TRIM must be set up accordingly. Taking into account the size of the Database and Document Store, you need to consider:

- backing up to disk instead of tape
- how to provide sufficient network bandwidth
- servers that can handle large volumes of traffic

Figure 27 illustrates such a strategy and what happens when a disaster occurs. When the disaster occurs, half-an-hour's data is lost. To reduce the impact on the system, you could configure the backup to do a full backup nightly and an incremental backup every hour.

Figure 27: High frequency backup strategy



Benefits of this strategy:

- The worst case is the loss of data of one hour.
- There is an extremely low risk of data inconsistency (see "Procedures Following an Inconsistent Data Recovery" below for more information on recovering consistency).

High Performance/Low Impact Setup Using Low Frequency Backups

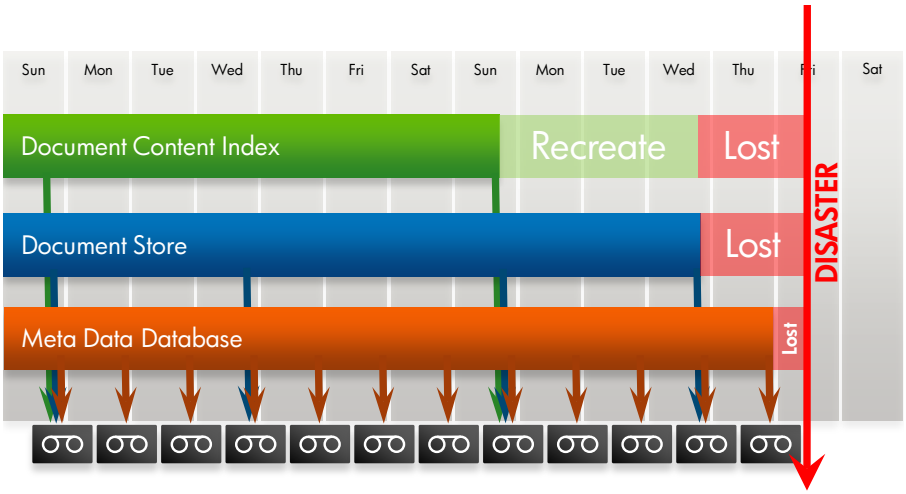
In this scenario we assume that backup to disk, large network bandwidth or servers capable of large traffic volumes are not available.

To keep the impact on the system as low as possible, consider creating separate backup specifications and schedules for the individual components of TRIM. Depending on the TRIM setup, it may be advantageous to back up the Metadata Database more often than the Document Store. For example, the Document Store could be too large to back up daily, while the Metadata Database is too important *not* to back it up daily. This can result in inconsistency, with a Metadata Database more recent than the Document Store after restoration. However, TRIM can easily handle such inconsistency as explained in section *"Procedures Following an Inconsistent Data Recovery"*.

It may also sometimes be sensible to back up the Document Content Index even less frequently than the Document Store. This can be handled by TRIM very well, since it is possible to recreate the Document Content Index from the Document Store for a certain time span only. This approach combines the advantages of not having to back up the Document Content Index too often, and keeping the recreation time to a minimum.

Figure 28 shows an example how such a backup strategy could be set up. In this example, the very important Metadata Database is backed up daily, the Document Store is backed up twice a week, and the Document Content Index only once a week. If a disaster happens as illustrated, the database loses only half a day, and the Document Store loses one and a half days. Although the Document Content Index loses four and a half days, it can quickly be recreated from the Document Store so effectively it also only misses one and a half days of data. The recreation of the Document Content Index is quick because it only needs to be done for three days.

Figure 28: Low frequency backup strategy



Procedures Following an Inconsistent Data Recovery

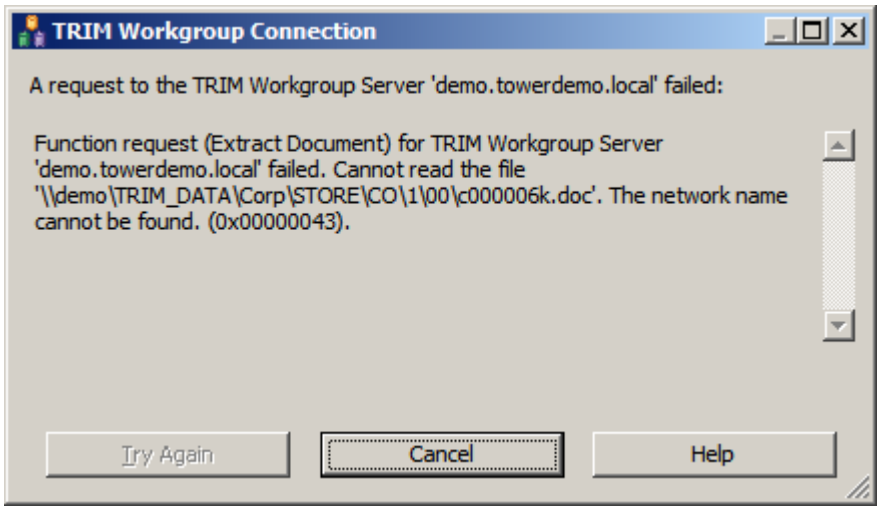
In the event of a recovery, you may need to decide whether a more complete but inconsistent recovery is more advantageous than an older but consistent recovery. If you choose an inconsistent recovery, there are a number of procedures within TRIM to bring the system back to a consistent state.

Scenario 1: Database is recovered but documents missing in the Document Store

In this case the system is operational but some documents are not available.

If a file referenced by the Metadata Database does not exist in the Document Store due to data loss, TRIM gives the error message shown in Figure 29, but the rest of the system will still work.

Figure 29: Error handling non-existent referenced file

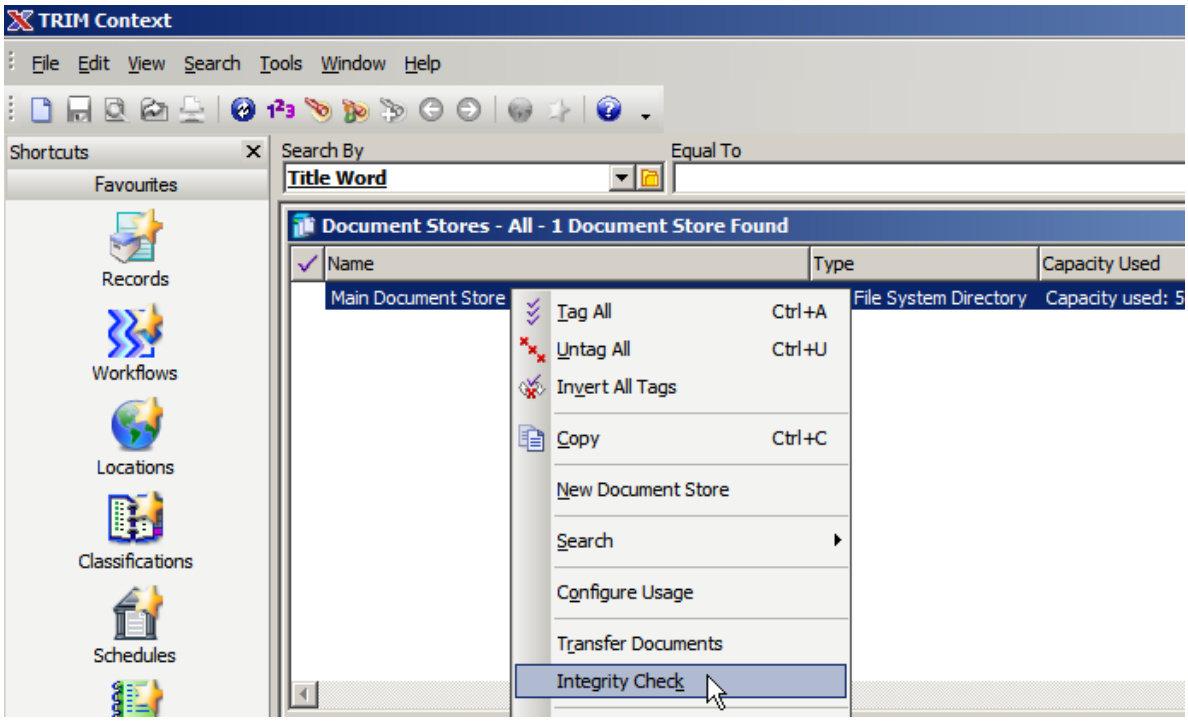


TRIM provides a Document Store Integrity Check which checks for any missing documents, although it takes time to execute in a large system.

To invoke the check:

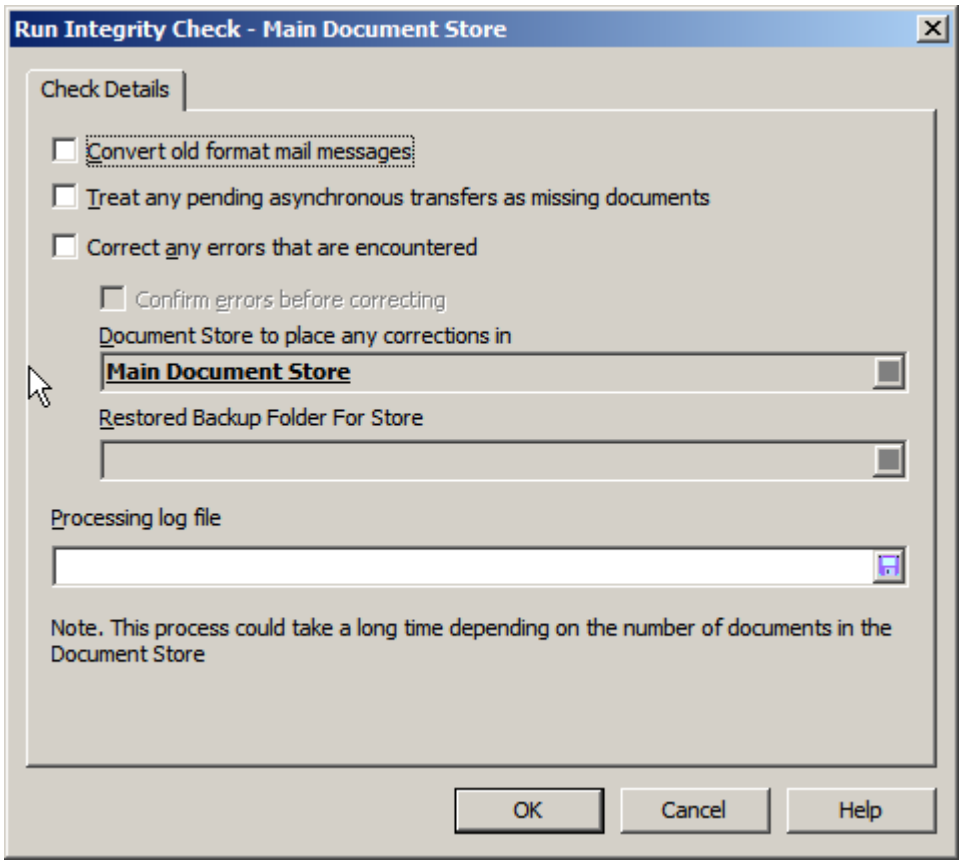
1. In the TRIM Context's main menu, open the Document Stores window by selecting **Tools → Context Administration → Document Stores…** See Figure 30.

Figure 30: Opening Integrity Check dialog



2. In the Run Integrity Check dialog (see Figure 31) click **OK** to start the check.

Figure 31: Run Integrity Check dialog



Note that you can source documents from any document caches in the system (if available) and simply copy them to the document store.

Scenario 2: All documents in the store but database recovered to a previous point in time

There is no reverse document store integrity check in TRIM but the documents with no metadata need to be removed before normal operation can be resumed. The file names are generated by TRIM in sequential order. To find out the last filename used by the database there are two different ways, depending on the Dataset schema version of TRIM used (to get your version open TRIM Context and select **Help → Setup Information...** and click on **Network Connections**):

- ***Before schema 118***: The table TSNumbers was the key to generating estoreIDs, using the SQL statement

```
SELECT nbrLastNumber  FROM TSNUMBERS WHERE nbrType = ")
        + (long)TRIM::nt_elecStoreURIs
        + " AND nbrRef = " + storeUri;
```

- ***After schema 118:*** It is purely a query of the TSELECSTOR.esitems column.

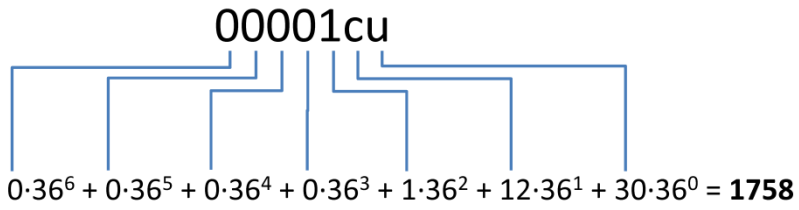
This will provide you with a number, such as 1758. This means the last filename c00001cu.doc stored would be worked out as follows:

- the initial c which indicates this is a TRIM document

- the file extension (here, ".doc"), which indicates the type of document and is the same as the original document that is registered.
- the part of the file name between the leading c and the file extension is a unique base 36 number, used by the database in the reSID column of the TSRECELEC table to associate the document with its record.

Figure 32 shows how the base 36 number and the decimal number relate to each other.

Figure 32: Conversion between base 36 number and decimal number



There are a few tools on the web to do decimal to any radix base number conversions, for example, at the time of writing, at http://www.easysurf.cc/cnver17.htm#b10toba.

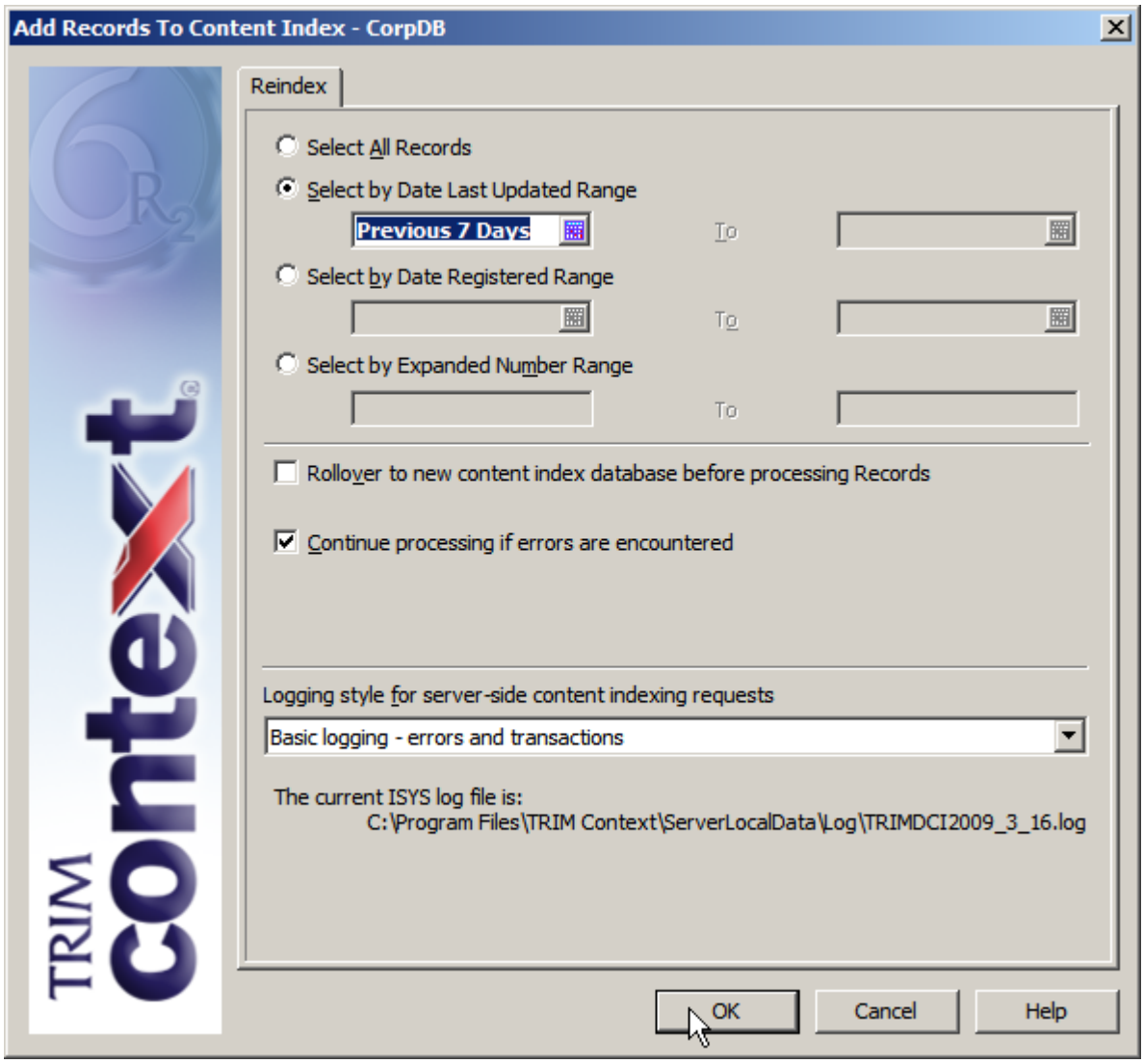
You need to remove all documents after this in the sequence. However if there are audit logs available, you may be able to re-catalogue them back into TRIM using the standard interfaces.

Scenario 3: All documents and metadata recovered but content index out of date.

Document Content Indexes within TRIM can be separated physically into chains which logically operate as one index. In the event of a recovery only the updatable chains need to be reindexed. Read-only chains will be able to be recovered as they are.

The fastest way to reindex the updatable chains is as a batch process using the TRIM Enterprise Studio tool as shown in Figure 33. To open this dialog, right-click on the Dataset in TRIM Enterprise Studio and select **Content Index → Reindex**. This tool allows you to reindex a selection of documents, such as those updated in the last 7 days. The Event Processor must not be running against these indexes while the manual reindex is in progress.

Figure 33: Reindex dialog



Legal Considerations

When backing up a TRIM system, be aware that legal regulations can state the records must be deleted after a certain time. For example, the law might insist that you delete backups older than 2 years. How you do this depends heavily on the particular circumstances. One solution would be to set the backup frequency and backup retention time at a very small fraction of the legally required retention time. It is your responsibility to conform to legal regulations.

For more information

www.hp.com/go/dataprotector

4AA2-4885ENA, May 2009