



IBM
Spectrum
Protect

File Archiving Solution with **IBM Spectrum Protect™ for Data Retention (SSAM)**

Introduction and configuration and usage guidance

White Paper

Nils Haustein, Executive IT Specialist, IBM EMEA Storage Competence Center

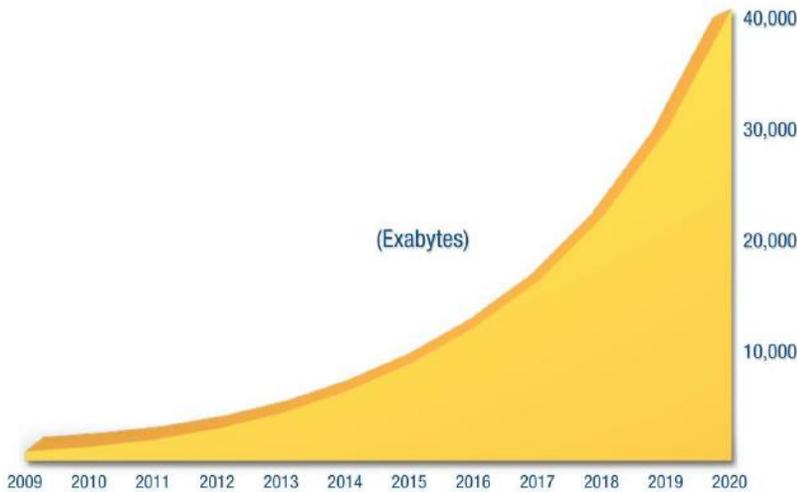
September 2017, Version 1.1

Table of Content:

Introduction	3
Overview of File Archiving	4
Requirements	4
Common Architecture	4
Archival operation	6
File Archiving Solution	7
Solution Overview	7
Spectrum Protect client (B/A client)	8
Spectrum Protect for Data Retention (SSAM)	8
SSAM Retention Policies.....	9
Solution Capabilities.....	11
Setup and Configuration.....	11
Example Configuration	12
Configuring the B/A Client	12
Configuring SSAM.....	14
Using this archive solution.....	16
Archiving.....	16
Manual Archiving	16
Using the Archive Client GUI	16
Using the Archive client command line (dsmc).....	18
Automated Archiving.....	19
Server defined schedules	19
Scheduled client commands	20
Search	21
Search via Find Menu	21
Search via Retrieve View	23
Search via Command Line Program.....	24
Retrieval	26
Managing Retention	28
Sending Retention Events via Web Client.....	28
Sending Retention Events via Command Line	29
Appendix.....	31
References	31
Disclaimer	32

Introduction

The amount of data created and copied in the digital universe is constantly growing. By 2020 IDC estimates the digital universe to be at 44 Zetabytes, this is 44×10^{21} Bytes (see figure 1).



Source: IDC's Digital Universe Study, sponsored by EMC, December 2012

Figure 1: IDC's Digital Universe Study (Source: International Data Incorporation (IDC))

The data volume is doubling in size every two years. A large portion of this data is unstructured with a majority being files stored in file systems. This presents a major challenge for file systems because of the vast amount of files. For example backing up file systems with billions of files in almost impossible because of the time it takes to crawl through the file system and identify files that require backup. Therefore it is important to archive files from a file system in an archive system that is optimized for storing and archiving huge amounts of files.

This document presents a practical solution for file archiving using the IBM Spectrum Protect Backup and Archive client (B/A client) in conjunction with IBM Spectrum Protect for Data Retention. Spectrum Protect for Data Retention was formerly named System Storage Archive Manager (SSAM) and has been rebranded to Spectrum Protect for Data Retention in 2016 [14]. In the first chapter general requirements for file archiving are discussed and a common architecture for file archiving is shown (see section [Overview of File Archiving](#)). The second chapter introduces a file archiving solution based on the B/A client and SSAM, demonstrates how it addresses the requirements for archiving and explains how to setup and configure it (see section [File Archiving Solution](#)). The third chapter provides guidance for using this solution for archiving, searches, retrieval and retention management (see section [Using this archive solution](#)).

Overview of File Archiving

This chapter describes the general requirements for file archiving and derives a common architecture for file archiving solutions. Based on this architecture the archival operations such as archive, search, retrieve and retention management are discussed.

Requirements

Requirements for file archiving come from different organizations such as users, applications, IT operations, financial officer and regulatory authorities.

Users and applications must be able to seamlessly *archive and retrieve*. Even more important is finding files years after archiving. This requires an efficient *search* function. Once file are found the user needs to be able to read the content.

IT operations guarantee the archival operations according to service level agreements and require that access to files is possible even after an *unforeseen event*.

Regulatory authorities demand that the data must be stored according to *regulatory standards* and specification. It is not subject of this document to explore details on regulations however some common regulatory requirements for data archival are:

- Data shall not be deleted or changed once archived
- Data shall be retained for a time period specified by the regulations
- Data copies shall exist in a different room or location for disaster recovery
- Data accessibility and integrity shall be demonstrable

The financial officer is concerned about the *cost* for data archival. This includes the purchase and operation of the archive system required for archiving and consisting of the components shown in Figure 2, as well as the cost of storage.

In section [Solution Capabilities](#) we explain how the solution presented in the next chapter meets these requirements.

Common Architecture

Figure 2 presents a general architecture for file archiving that is comprised of 3 layers:

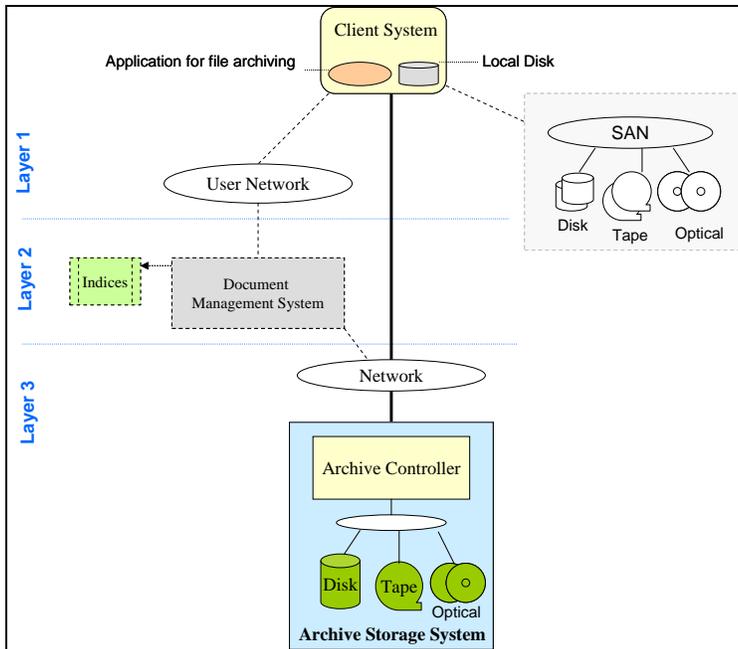


Figure 2: Common Architecture for File Archiving

On top in layer 1 in Figure 2, there is the client system used by users and applications which stores the files subject for archiving in a file system. The file system can reside on local disk drives or on disk systems attached via a Storage Area Network (SAN). A file archive application installed on the client system performs the archival operations including selection, archiving, search, retrieval and retention management (see section [Archival operation](#)). The client system can be directly connected to the archive storage system in layer 3 or it can be connected via an Enterprise Content Management system (ECM) in layer 2.

If the client system is directly connected to the archive storage the archive application archives data and metadata directly in the archive storage via the archive storage API. It may keep the metadata internally in a catalog or store the metadata with the data.

If the client system is connected via the ECM system the archive application supports the API of the ECM. The ECM system is an application comprising a metadata catalog (database) to extract and store metadata whereby the actual data is stored in the archival storage. An ECM system provides additional functions including indexing, search & discovery and support for business processes.

Metadata in this context is additional information describing the data - for example attributes of the data, owner information, access control lists and even a full text index - that can be used for subsequent searches and business processes.

The archive storage system retains the data and metadata and implements the retention policies. The data can be stored on different storage technologies attached to the archiving storage, such as flash, disk, tape and even optical storage. The archive storage may also include capabilities to migrate data from one technology to another and hence place the data on the most appropriate storage medium.

The solution presented in this document works without an ECM system. The application is the Spectrum Protect Backup and Archive client (B/A client) and the archive storage system is a solution based on Spectrum Protect for Data Retention (SSAM).

Archival operation

The archival operations within an archive system shown in figure 2 includes selection, archiving, search, retrieval and retention management.

During the *selection* files subject for archiving are being selected. The selection criteria typically depends on the business requirements. In many cases files are selected based on their name, age or type. If an ECM system is part of the architecture it typically drives the selection of files for archiving.

The *archiving* process moves selected files from the file system of the client system to the archive storage (or ECM system respectively). Files are written to the archive storage using the data interface (API) of the archive storage (or ECM system respectively). During the archive process metadata is generated that allows search and retrieval of files. The archive process is initiated and controlled by the archive application or by the ECM system. If an ECM system is part of the architecture is stores and manages the metadata, otherwise the metadata is managed by the archive application.

The *search* function is key to find files for retrieval. The search function might be as simple as searching for file and path names or as complex as searching for the content of files, leveraging a full-text index. The latter is usually possible with an ECM system. The search function is provided by the archive application in conjunction with the ECM system, if this is part of the architecture.

The *retrieval* process reads the file from the archive storage and presents it at the client system. For this purpose it might be copied from the archive system to the file system of the client.

Retention management ensures that files are retained and cannot be deleted within a specified or unspecified period of time. This is achieved by assigning a specific retention time or retention policy to each file when it is archived in the archive storage. The archive storage prevents the deletion or changes to files during the retention time. A retention policy can specify that a file is kept forever, until an event is signaled. The event causes that the file is kept for a specific retention period. Retention management is driven by the archive application (ECM respectively) and implemented by the archive storage.

File Archiving Solution

In this chapter a practical and cost-efficient file archiving solution based on the IBM Spectrum Protect product portfolio is presented. After providing an overview of the solution the implementation is discussed.

Solution Overview

Figure 3 presents an overview of the file archiving solution being discussed in this paper:

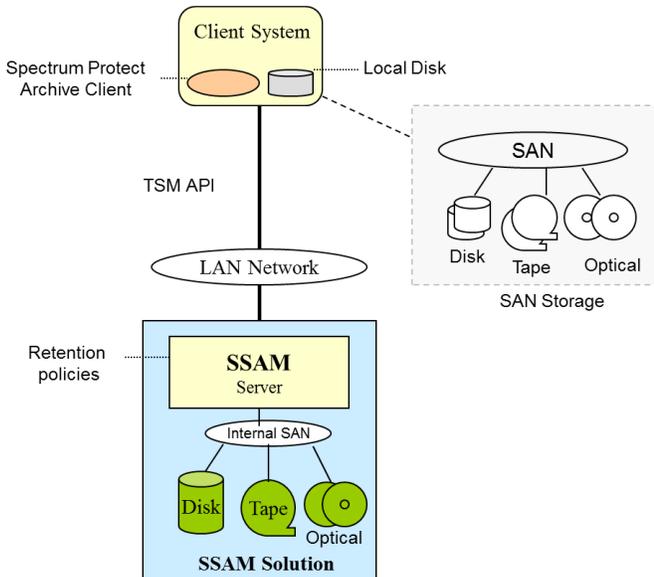


Figure 3: File archiving solution based on B/A client and SSAM solution

On the client system the Spectrum Protect Backup and Archive client (B/A client) is installed (see section [Spectrum Protect client \(B/A client\)](#)) and represents the archive application. The B/A client archives files in the SSAM solution (see section [Spectrum Protect for Data Retention \(SSAM\)](#)) that represents the archive storage. The communication between the B/A client and SSAM is based on the open TSM API that leverages TCP/IP sockets over Ethernet. Files stored in SSAM are retained according to the configured retention policies (see section [SSAM Retention Policies](#))

This solution provides basic functionality for archive, search & retrieve and retention management (see section [Using this archive solution](#)). It is cost efficient because it does not require an ECM system and the B/A client does not require a license fee when used with Spectrum Protect for Data Retention.

Archiving can be done manually using the command line program or the B/A client Graphical User Interface (GUI). In addition the archive operation can be automated, for further insight see section [Archiving](#).

The search capabilities are based on path and filenames of archive files and some additional metadata that can be given with an archive job. The search capabilities are built in to the B/A client Graphical User Interface (GUI) and command line program. For more details refer to section [Search](#).

Retrieval of files is usually done using B/A client Graphical User Interface (GUI) or the command line program. Usually a search may be done before the retrieval. Retrieval can also be automated. For more details refer to section [Retrieving](#).

The B/A client also includes functions which allow one to manage the retention period for a file or a set of files. See section [Managing Retention](#) for more details.

Spectrum Protect client (B/A client)

The Spectrum Protect backup and archive client (B/A client) is publicly available for many different platforms and operating systems and can be publically downloaded (see [1], [2] and [3]). The B/A client allows archiving of files from file system residing on local or SAN storage to the Spectrum Protect for Data Retention server (SSAM). Retrieves are done from the SSAM server to the storage of the client system.

The B/A client includes a graphical user interface (GUI) and a command line program (dsmc) which can be used to drive archival operations. Both require the B/A client to be installed and configured (see section [Configuring the B/A Client](#)).

Archiving operations can be done manually or automated by specifying file systems, directories and files. Include and Exclude lists can be used to control the selection of files to be archived based on names and extensions. In addition the archived files can be optionally deleted from the file system.

Assigning retention times and policies during file archiving is very flexible. Thereby archived files can be assigned to retention policies based on file system, directory and file name patterns.

Retrieval of files is usually a manual operation whereby multiple files can be retrieved at one time. For manual operation the GUI or the command line can be used. Retrieval can also be automated in accordance with SSAM.

The B/A client also allows searching for files using the GUI or the command line program. The GUI provides search masks for different parameters including file and directory names as well as archive dates and time. The B/A client stores the files in the SSAM server by keeping the path and file names. In addition it stores an archive description for each file being archived.

Spectrum Protect for Data Retention (SSAM)

The Spectrum Protect for Data Retention solution (SSAM solution) is built from standard hardware and software components. The high level SSAM solution architecture is shown in Figure 4:

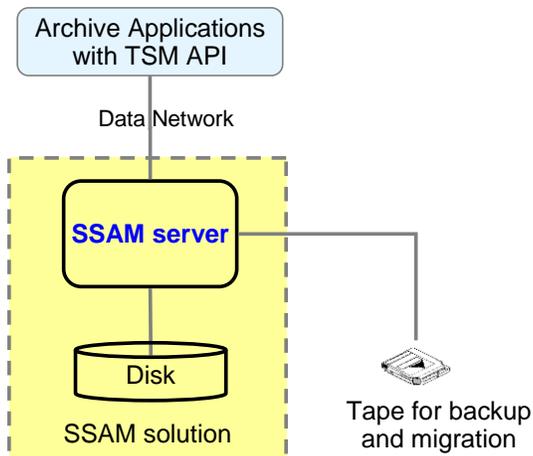


Figure 4: SSAM solution architecture

The archive application is the B/A client and manages archival operations. The B/A client uses the TSM API [5] in order to communicate with an SSAM server over an Ethernet network. The SSAM server comprises the SSAM software running on a server operating system.

The SSAM software manages the storage of the archived data objects in storage pools. Storage pools can be on disk and on tape. Typically archived data is stored on disk first and periodically backed up to tape. When the archived data is less frequently accessed but yet has to be retained it might be migrated to tape. The SSAM software can be configured to perform this automatically. Depending on the archive storage requirements the disk can be installed within the server or attached externally via a private network infrastructure (SAN).

The recommended configuration for the SSAM solution shown in Figure 4 includes IBM system p servers (POWER8) running the AIX operating system. The archive data might be stored on internal disks or external disk. External disk is realized with an IBM Storwize disk system (V7000, V5000 or V3700). The tape system might be an IBM LTO or TS1100 tape drive using read-write and WORM tape in a tape library. These components provide longevity, robustness and scalability which are key requirements for archive systems.

SSAM is designed to help companies meet regulatory requirements, manage risk, reduce cost and improve operational efficiency by providing the following functions:

- Independently assessed for compliance according to legal regulation [7]
- Business Continuity through redundancy and data protection
- Cost optimization through tiered storage function
- Build in migration functions
- Open interface made for archiving
- Flexible and scalable

For more details about the functions of an SSAM solution refer to [6].

SSAM Retention Policies

There are essentially two distinct retention policies which can be configured in SSAM as part of the archive copy group: chronological and event based. The retention policy is configured using the archive copy group parameters RETINIT, RETMIN and RETVER:

- The parameter **RETINIT** specifies the retention policy (event based or chronological). Possible values are CREATION or EVENT.

- Parameter **RETVER** specifies the number of days to retain a file. Possible values are 0 to 30,000 days or NOLIMIT.
- Parameter **RETMIN** applies only to event-based archive retention policy (RETINIT=Event) and specifies the minimum number of days to retain an object. This parameter is considered independently of the RETVER parameter. The default value is 365. Possible values are 0 to 30,000 days.

The chronological retention policy is further explained in figure 5.

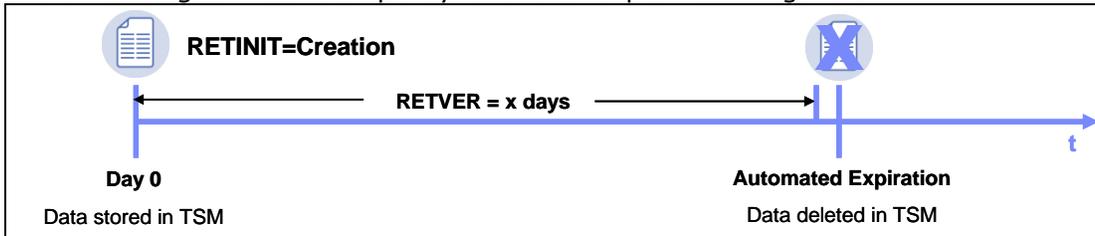


Figure 5: Chronological retention policy

When a file is stored in SSAM and assigned to a chronological retention policy (RETINIT=CREATION) on day 0 the file is retained for a fixed amount of time specified by the RETVER parameter. Once this time is expired the file is deleted automatically from SSAM.

Note, when using chronological retention policy the deletion of a file in SSAM is done automatically when the expiration time is reached. Subsequent access from the B/A client will result in failures.

The event based retention is explained in figure 6.

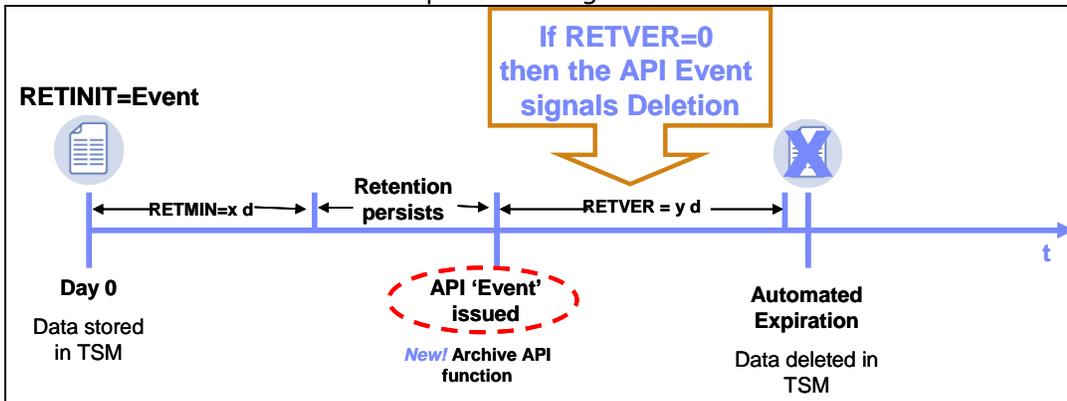


Figure 6: Event Based Retention Policy

When a file is stored in SSAM and assigned to an event based retention policy (RETINIT=EVENT) on day 0 the file is retained for an undetermined time until the event occurs. The event is a TSM API function call (dsmRetentionEvent). Upon the occurrence of the event-function SSAM calculates the retention time of the file by determining the maximum of: the time specified by RETMIN and the time of the occurrence of the event plus the time specified by RETVER. In other words, if RETMIN=5, RETVER=2 and the event occurs on day 2 after it was stored, the file is kept for 5 days, because RETMIN is greater than the time of the event plus RETVER. If for the same retention policy the event for a file is triggered after 4 days, the object is kept 6 days in total because the time of the event plus RETVER is greater than RETMIN. If the event never occurs the file is kept forever.

The meaning of an event can be configured using the parameter RETVER. For example, if RETVER=0, the file will be deleted when the event occurs, thus the event means deletion.

It is obvious that the parameter RETMIN allows enforcing retention protection. This parameter usually reflects the legal requirements in regard to the retention time for the particular kind of data.

Note, using event based retention with a RETVER=0 the deletion can be induced by the B/A Client (see [Managing Retention](#) for more details)

Retention policies are configured in the SSAM server.

Solution Capabilities

The file archiving solution presented above implements a practical approach for file archiving at low cost. It meets the basic requirements for file archiving presented in section [Requirements](#), as demonstrated below

Provides simple ways for **archiving and retrieve** using a graphical user interface, a command line or automation (see sections [Archiving](#) and [Retrieval](#)).

Provides basic **search** capabilities using different methods (see section [Search](#)). These capabilities rely on naming conventions for archived files and directories. The naming convention must be implemented by the user.

Provides **regulatory compliance** based on SSAM that has been assessed according to legal requirements of multiple countries ([7]).

Allows **automation** of file archiving, for example by using scheduled operations (see section [Automated Archiving](#)). The solution also allows different policies, one for example to delete files from the local disk after archiving. It is also possible to assign different retention policies to different files.

Provides **cost effective storage** by allowing to place the data on the most appropriate storage medium. The foundation for this is provided by SSAM and its capability to support different storage tiers (flash, disk and tape) and transparently migrate files between tiers. For more insight to cost efficiency with tiered storage solutions including a tape tier can be found in [13].

Provides **disaster protection** by an integrated backup function in SSAM allowing to create a backup copy of all data on a different storage medium in a different location. In addition the SSAM solution can be implemented as a mirrored solution across sites providing **high availability** leveraging additional cluster software such as PowerHA™ and disaster protection leveraging disk or file system replication.

Setup and Configuration

This section describes the implementation and configuration of the B/A client and SSAM solution in order to establish the file archiving solution.

Example Configuration

Figure 7 shows a simple archive solution setup including a client system running the B/A client and a SSAM solution with the SSAM software running on a server and storing data on disk.

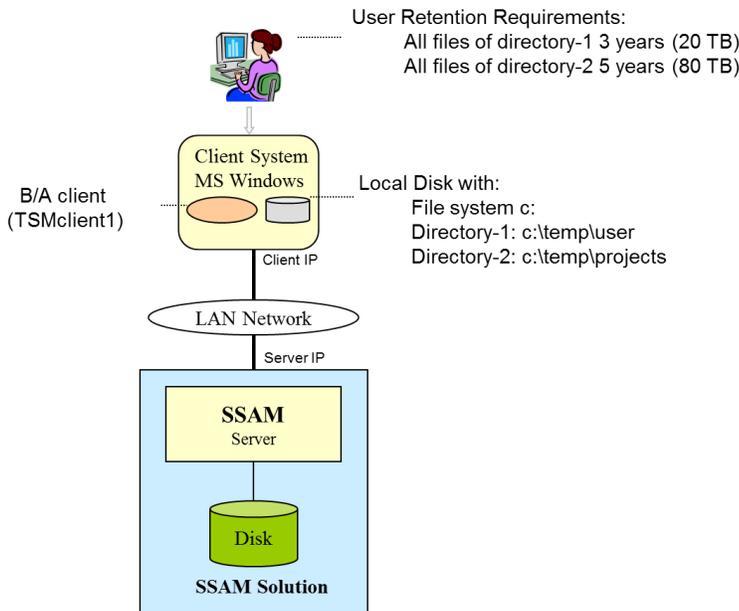


Figure 7: Example file archive solution configuration

In this example the client system is a workstation running Microsoft Windows® with the B/A client installed. The client system includes a local disk drive C: with two directories subject for archiving. The user of the client system requires three year retention for the data in directory-1 and five years retention for the files in directory-2. The amount of storage space needed is 20 TB for directory-1 and 80 TB for directory-2 – a total of 100 TB.

The client system communicates with the SSAM solution via Ethernet. Both the client and the SSAM solution have an IP address assigned in the same logical network. The client system uses the name TSMClient1 for registration and authorization. The SSAM solution stores the data on disk technology with the option to backup and / or migrate it to tape.

The implementation comprises two distinct steps: [Configuring the B/A Client](#) and [Configuring SSAM](#).

Configuring the B/A Client

The B/A client can be downloaded from IBM (see [3]). Afterwards it can be installed on the system. After installation there is a new program added to the system, as shown in figure 8.

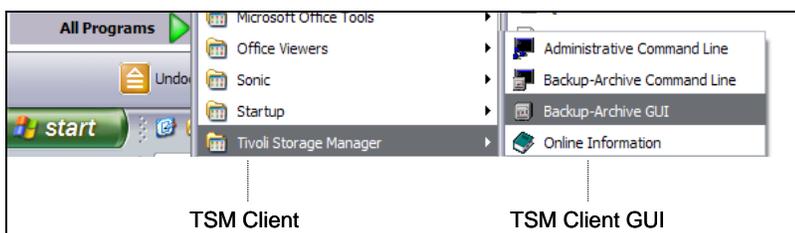


Figure 8: B/A client Program

Before the B/A client can be started via the GUI or the command line interface, the B/A client option file must be configured (see also [8]). The option file describes the connection boundaries between the B/A client and the SSAM server. The option file resides in the B/A client home directory (typically C:\Program Files\Tivoli\TSM\baclient) and is named dsm.opt. If this file does not exist, it must be created. Find below the content client option file according to the example¹:

```
Commmethod          tcpip
tcpport             1500
TCPSERVERADDRESS    server-IP
NODENAME            TSMClient1
Passwordaccess      generate
ENABLEARCHIVERETENTIONPROTECTION yes

exclude.archive c:\...\*
exclude.archive c:\temp\...\*
Include.Archive c:\temp\user\* mgmt1
Include.Archive c:\temp\Projects\ mgmt2
```

Find below a short description of the parameters in the dsm.opt:

- Commmethod: instruct the B/A client to use TCPIP as protocol
- tcpport: instruct the B/A client to use TCPIP port 1500 for communication with SSAM. Note, the SSAM service on the SSAM server must be configured on this port.
- TCPSERVERADDRESS: specifies the TCP/IP of the SSAM Server
- NODENAME: specifies the node name used by the B/A client for archiving and retrieving data
- Passwordaccess: Instructs the B/A client to automatically generate passwords.
- ENABLEARCHIVERETENTIONPROTECTION yes: is required to instruct the B/A client that it is communicating with a SSAM server
- Exclude.archive instructs the B/A client to exclude all other directories of the c: drive.
- "Include.Archive c:\<directory>\<file-pattern> mgmt1": instructs the B/A client to archive all files from <directory> satisfying the <file-pattern> and assigning the retention policy specified in management class <mgmt1>. Management class <mgmt1> is configured in SSAM.

Note: These include and exclude statements are processed from bottom to top (see [11] for more details).

The actual retention times are configured in the management classes of the SSAM server (see [Configuring SSAM](#)).

Once the B/A client is installed and configured the GUI can be launched (see figure 8) or the B/A command line interface (dsmc-program) can be invoked. The B/A client command line interface usually located in the B/A client home directory (...Tivoli\TSM\baclient). To start the dsmc-program open a command prompt on the client system, change to the directory where the dsmc-program is located and enter: dsmc. This will start the interactive mode allowing the user to enter commands and options.

The B/A client command line program (dsmc) can be run in batch- or interactive mode. Batch mode allows running client commands unattended from a batch program executed at the client system. To use the batch mode enter the command:

¹ Note, for the B/A client for MS Windows there is only a option files (dsm.opt) but no system file (dsm.sys). The system file statements are included in the option file.

```
dsmc arc -su=yes c:\temp\user\*.
```

This will archive all files from the directory c:\temp\user.

Configuring SSAM

Configuring SSAM is based on [example configuration](#) and requires the following information:

- TCPIP Address of the SSAM solution is server-IP. This address must be configured in the server running SSAM accordingly.
- The B/A client node name is TSMClient1.
- The management class mgmt1 shall be configured to retain data for 3 years.
- The management class mgmt2 shall be configured to retain data for 5 years.
- The storage requirements are 100 TB of storage space on disk.

The SSAM configuration sequence is as follows:

- Define Device Class and Storage Pool
- Define Policy Domain, Policy Set, Management Class and Archive Copy Group
- Assign a Default management class
- Validate and activate the policy set
- Register the client nodes

Here are the appropriate SSAM commands to accomplish the tasks according to this sequence:

1. Create a device class for the sequential storage pool

```
define devc FILECLASS devt=FILE mountlimit=10 maxcap=500G  
directory=/ssam/stg/fileclass
```

The device class name is FILECLASS.

The mountlimit is set to 10 because the maximum number of client sessions is below 10.

The maxcapacity is set to 500 GB and specifies the size of the volumes. A total capacity of 100 TB is required which results in a total of 200 volumes.

The location of the storage pool volumes within the SSAM server is in /ssam/stg/fileclass.

2. Create the primary storage pool to store archived objects:

```
define stg clpool1 FILECLASS pooltype=primary maxscr=200 dataformat=nonblock  
crcdata=yes reclaim=10 reclaimprocess=2 collocate=filespace
```

The name of the storage pool is clpool1.

The storage pool uses the device class FILECLASS that specifies a sequential access storage pool stored in directory /ssam/stg/fileclass.

The parameter maxscr=200 defines that a maximum of 200 volumes can be allocated. With 200 volumes of a size of 500 GB results in the total capacity of 100 TB.

The parameter dataformat=nonblock defines that small blocks are being allocated.

The parameter crcd=yes defines that a checksum is calculated for each object stored.

The parameter reclaim=10 defines that a volume is reclaimed when 10 % of the data are expired.

The parameter collocate=filespace defines that objects from the same file space are collocated.

3. Define the Policy Domain using the TSM command:

```
define domain cldom1
```

The parameter cldom1 is the name of the policy domain.

4. Define the Policy Set using the TSM command:

```
define pol cldom1 clset1
```

The parameter clpol1 is the name of the policy set.

5. Define Management Class mgmt1:

```
define mgmt cldom1 clset1 mgmt1
```

The parameter `mgmt1` is the name of the management class according to the requirements. This management class defines now the retention policy for directory-1 data.

6. Define the archive copy group for management class `mgmt1` including the retention policy and the destination storage pool:

```
define copyg cldom1 clset1 mgmt1 t=a retinit=creation retver=1095 dest=clpool1.
```

This defines the retention policy for management class `mgmt1` with a retention time of 1095 days (3 years).

7. Repeat Step 5 and 6 for management class `mgmt2`, where `retver=1825` (5 years)

Once all management classes are defined for a given policy domain, perform the following steps:

8. Assign the default management class::

```
assign defmgmt cldom1 clset1 mgmt2
```

This assigns management class `mgmt2` as the default class because it has the longest retention.

9. Validate policy set:

```
validate pol cldom1 clset1
```

There may be some warnings presented on the screen indicating that there is not backup copy group defined. This kind of warning can be ignored as long as the policy set is ready for activation.

10. Activate the policy set:

```
activate pol cldom1 clset1 mgmt1
```

There may be some warnings presented on the screen indicating that there is not backup copy group defined. At the end of the warning section the command shall indicate that the policy set has been activated.

11. In this step register the node name `TSMClient1` for the client system and assign it to policy domain `cldom1`:

```
register node tsmclient1 password domain=cldom1 maxnummp=10 passexp=0
```

The parameter `maxnummp` defines the number of mount point for this node which is equivalent to the number of sessions the application using the node-name opens in parallel.

The parameter `passexp` defines the password expiration period for this node. A value of 0 indicates that the password shall never expire and a value of 365 indicates the password expires after 365 days.

The next section elaborates some operational aspects using this archive solution.

Using this archive solution

In this section archiving, retrieval and search operations using this archiving solution are described.

Archiving

Archiving can be done using two distinct methods:

1. Manual archiving using the B/A client GUI or command line (see [Manual Archiving](#)).
2. Automated archiving is accordance to client schedules setup in the SSAM server (see [Automated Archiving](#))

The prerequisites are that the B/A client and the SSAM solution has been configured (see section [Setup and Configuration](#)).

Manual Archiving

Manual archiving can be done using the graphical user interface (GUI) of the TSM B/A client or via B/A client command line (dsmc) [10].

Using the Archive Client GUI

Start the B/A client GUI on the client system as explained in section [Configuring the B/A Client](#). The following window will be displayed (Figure 9):



Figure 9: Archive Client GUI

In order to perform an archiving operation, select **Archive**. In the archive window select **Local**, this displays the local file system of the client system (Figure 10).

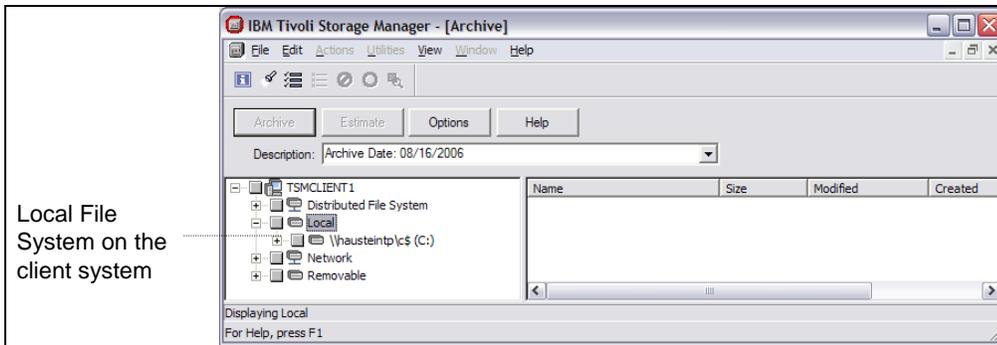


Figure 10: Local File systems for archiving

In the left panel (Figure 10) locate the directories and files subject for archiving by expanding the directory views. According to this example the directories c:\temp\user and c:\temp\projects are subject for archiving. Locate these directories as shown in Figure 11. When selecting a directory, files are presented in the right hand panel.

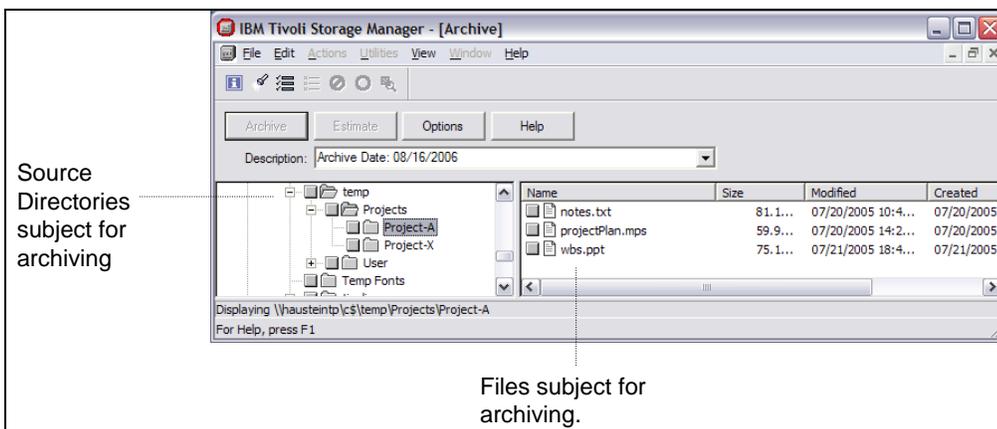


Figure 11: File view

Files or / and directories can now be chosen by selecting the gray button in front of file in the right hand panel or directories in the left hand panel as shown in Figure 12.

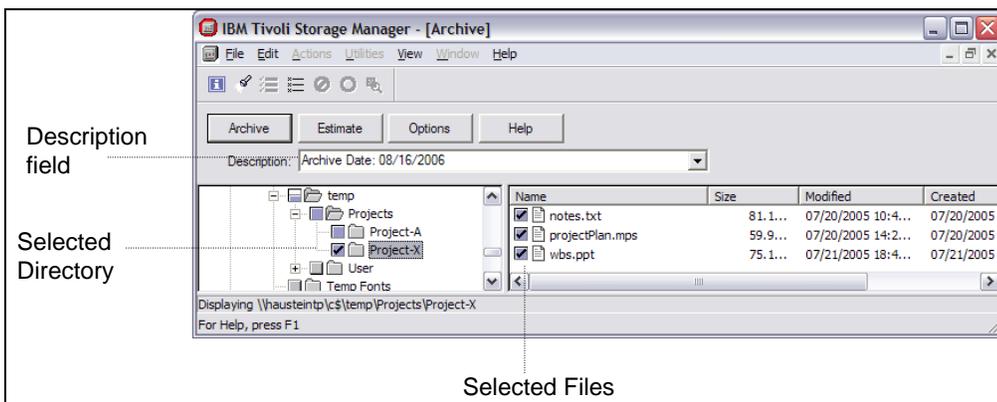


Figure 12: Selecting files and directories

Once files according to one retention policy have been selected (all files and directory under c:\temp\project should be stored 5 years) options for archiving can be selected by pressing the **Option** button. The available options are shown in Figure 13.

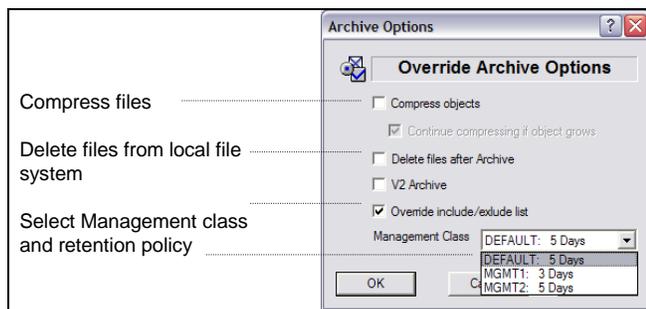


Figure 13: Available options

Select the appropriate options, in particular the option to **Delete file after archive** and the **Management Class**. Press **OK** button when done. Note, when selecting Delete file after archive the file will be physically deleted from the local file system of the client system. With the selection of the management class the associated retention policies is assigned to the selected files once archived.

The archive description field can also be changed (Figure 14). This might be important when this field is used for subsequent searches. The content of this field is also stored with the data in SSAM. Therefore it is possible to store additional metadata with the description. Please note that by default the B/A client enters a description indicating the data of archiving, this can be changed to satisfy the need of the user.

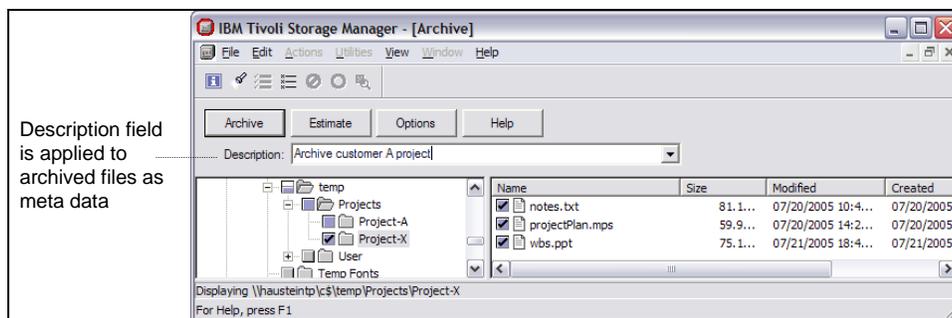


Figure 14: Description Field.

Once all this is done press the archive button and the files and directories are archived in SSAM.

Using the Archive client command line (dsmc)

The B/A client provides a command line interface which is usually installed in ...\\tivoli\tsm\baclient (/usr/tivoli/client/ba/bin in unix) and which is called *dsmc*. The command line provides all methods for archiving and retrieval. The general syntax for archiving is:

```
dsmc archive [options] file-specification
```

Important options are:

- subdir=yes: includes subdirectories
- deletfiles: deletes files after successful archiving
- archmc=(mc-name): binds all objects found to the management class *mc-name*
- compression=yes: compresses the files before transferring to SSAM

-v2archive: archives only files, no directories in the path of the file-specification
For a complete list of archiving command options see [9].

The file-specification specifies the files to be archived. Specifies path and name of the file you want to archive. You can use wildcards to specify groups of files or all the files in a directory. You can specify as many file specifications as available resources or other operating system limits permit. Separate file specifications with a space. You can also use the **filelist** option to process a list of files. The B/A client opens the file you specify with this option and processes the list of files within according to the specific command.

The example below will archive all files from directory c:\temp\user\susann and assign the retention policy of management class mgmt1. Subsequently archived files are deleted from the local file system. The archive description is "Susanns Archive".

```
dsmc archive -des='Susanns Archive' -subdir=yes -archmc=mgmt1 -del  
c:\temp\user\susann\*
```

A subsequent check shows the archived files:

```
tsm> q arch c:\*\*\susann\*  
Size  Archive Date - Time      File - Expires on - Description  
----  -  
22,158 B 08/21/2006 17:31:37  \\hausteintp\c$\temp\User\Susann\Sfile1.out  
08/24/2006 Susanns Archive  
61,348 B 08/21/2006 17:31:37  \\hausteintp\c$\temp\User\Susann\Sfile2.out  
08/24/2006 Susanns Archive
```

Automated Archiving

There are two methods to automate file archiving. With both methods all files according to include and exclude statements are archived (see section [Configuring the B/A Client](#) for more details). The two methods are:

- Server defined schedules [12]
- scheduled TSM client commands

Server defined schedules

Server defined schedules required the definition of schedules in the SSAM server and the client scheduler service to be started on the B/A client system.

To define schedules in the SSAM server the actual schedule must be defined and associated with the client node (TSMclient1). An example for creating a client schedule in the SSAM server with the name tsmclient1-archive is shown below. The schedule operation will run every day at 18:00 and archive all data from the client directory c:\temp including subdirectories.

```
define sched cldom1 tsmclient1-archive t=client action=archive options='-subdir=yes'  
-objects='c:\temp\*.*' startdate=today starttime=18:00 duration=1 priority=3  
  
define association cldom1 tsmclient1-archive tsmclient1
```

Further options which could be used for archiving are:

- subdir=yes
- deletfiles
- domain="all-local -c: -d:"
- archmc=(mc-name)

-compression=no
-v2archive

For more options refer to [9].

The client scheduler is configured via the B/A client GUI. The client scheduler can be started via the client acceptor daemon (CAD) or it can permanently be running as a service (dsmc sched). The recommended method is to start the scheduler via the client acceptor daemon. Perform the following steps to setup the client scheduler and the client acceptor daemon.

1. Start the Backup Archive Client GUI

Note this requires a minimum configuration in the option file such as:

commmethod	tcpip
tcpport	1500
TCPSERVERADDRESS	server-IP
Nodename	TSMClient1
ENABLEARCHIVERETENTIONPROTECTION	yes

2. Select **Utilities - Setup Wizard - Help me configure the TSM Client Scheduler** from the Tivoli Storage Manager main window. The TSM Scheduler Wizard window appears.
3. Select the **Install** a new or additional scheduler task and press the **Next** button.
4. Specify the **name of the Scheduler**. Select the **Local Machine** and **Use the Client Acceptor daemon (CAD)** to manage the schedule options. Click the **Next** button.
5. Specify the name of the client acceptor service that you want to manage the scheduler. If the CAD is already installed for use by the Web client, **select that CAD** from the drop down list. Otherwise, type the name you want to give the CAD, for example, TSM Client Acceptor. Click the **Next** button.
6. Follow the instructions on the screen to complete the configuration of the options file name, the HTTP port the CAD uses, authentication information, service login options, log file names, and immediate start option.

Notes:

- If the option "sessioninitiation" is set to serveronly in your client options file (dsm.opt), the client setup wizard and scheduler service might be unable to initiate authentication with the Tivoli Storage Manager server. To avoid this problem, ensure that the Contact the TSM Server to validate password checkbox on the TSM Authentication page is unchecked.
- In the Service login options window, select the option "Start automatically when Windows boots" to have the service started automatically when Windows boots so that your schedules will be run.

The client scheduler and the CAD can also be setup via the command line using the dsmutil program.

Scheduled client commands

The second method for scheduling archive operations uses the command line program dsmc. This program is usually located in the B/A client home directory (...\\Tivoli\\TSM\\baclient). This program can be used from the client system to archive files. For automation purposes the program could be scheduled within the client system for example using cron. For more information regarding the dsmc-program refer to [9].

The example below will archive all files from directory c:\\temp\\user\\susann and assign the retention policy of management class mgmt1. Subsequently archived files are deleted from the local file system. The archive description is "Susanns Archive".

```
dsmc archive -des='Susanns Archive' -subdir=yes -archmc=mgmt1 -del  
c:\temp\user\susann\*
```

Search

Searching for archived files is a vital part of an archiving solution because files are usually retained for many years. Therefore it is important to provide granular search capabilities.

Searching for archived files is based on naming conventions for path and file names in the client system. The naming conventions must be implemented by the user. The search is based on the fact that the B/A client keeps file and path names with the archived files. These names can be used as search arguments subsequently. The B/A client also allows adding an archive description (text) for each file to be archived or groups of files which can be used for searches as well.

There are three different search methods presented in here, not all have to be used. The methods are:

- Search using the **Find** menu,
- Search using the **Retrieve** view
- Search using the command line program

The table below summarizes the search arguments that can be used according to search method:

Method	File name pattern	Directory name pattern	File Size	Archive Description
Find Menu	X	X ¹	X	X ¹
Command Line	X	X	-	X ¹
Retrieve Menu ²	X	X	-	X

¹ Full qualified names without wild-cards.

² No search engine, manual search be view.

It becomes obvious from the search arguments that the archived files must include some kind of metadata in order to enable an effective search. Metadata can be based on the path and file names. It can also be based on the archive description which can be attached to archived files during [archiving](#).

The search requires a client node name with access to the archive data. Only files archived by that client node names can be searched.

Search via Find Menu

This menu allows search arguments including:

- Full qualified path or directory names,
- File names,
- File size
- Archiving dates

A search engine can be used for a fuzzy search. In order to get to the **Find** menu: start the B/A client GUI, select **Retrieve** and than select the menu item **Edit** and than **Find**. The following mask is presented (Figure 15).

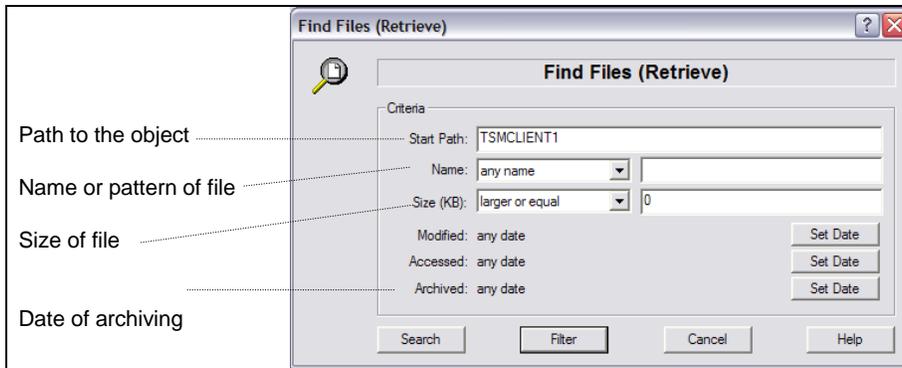


Figure 15: Search Mask

The field **Start Path** specifies the location in which to begin a file search. When opening the Find Files window, this field is pre-filled with information about the object that is selected in the tree. The **Start Path** can contain:

- Client node name
- Any fully-qualified drive or directory on your workstation

The **Name** field allows searching for file names according to different match-policies.

In the **Size** field policies for the file size to be found can be entered.

The various **Date** fields allow entering dates for archiving and access, as displayed in Figure 16. There are two dates which can be entered and a policy for the search based on dates.

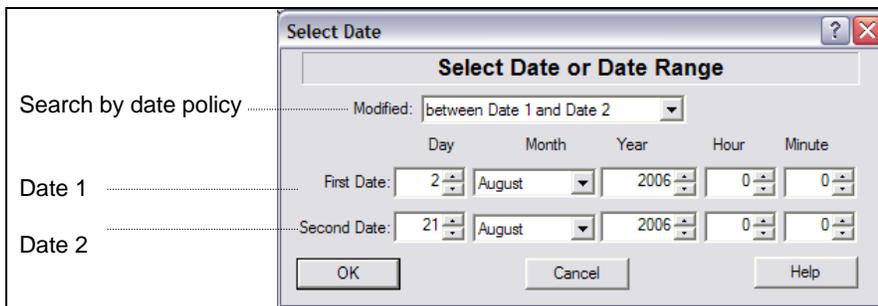


Figure 16: Search by Date

Appropriate search arguments can now be entered in the search mask. For example: when all archived files which include the pattern 'Plan' in the file name should be found, then enter the pattern "*Plan*" in the name field and select the search policy "matches mask" as shown in figure 17².

² Based on testing done by the author search strings are only applied for file names and not for directory names. Directory names have to be identified.

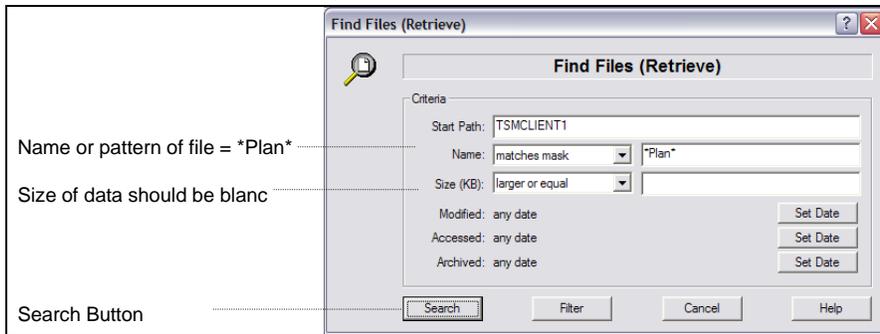


Figure 17: Search for "Plan"

By pressing the **Search** button the search is started and the results are displayed in a separate window (Figure 18). In order to retrieve one or more files, select the files by checking the gray button in front of the file name and close the search-result window as indicated in Figure 16. In the next window the file remains selected and can be retrieved by pressing the **Retrieve** button.

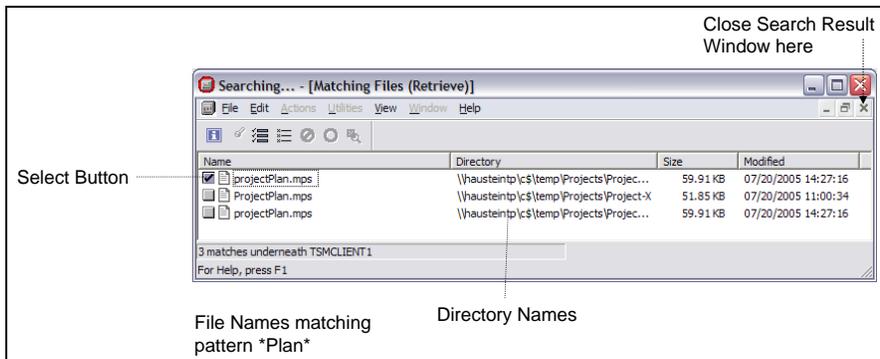


Figure 18: Search Result

Search via Retrieve View

The **Retrieve** view allows search arguments included in the path names. The path name consists of the following

- Client node name
- Archive Description
- File system
- Directory structure of the file

In order to get to the **Retrieve** view: start the B/A client GUI, select **Retrieve**. Figure 19 presents the retrieve view.

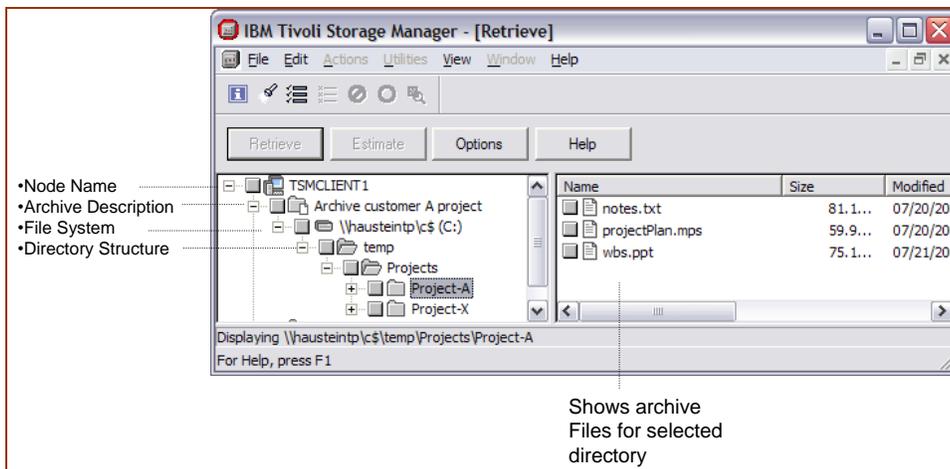


Figure 19: Search in Retrieve View

The search in the Retrieve view is not facilitated by a search engine and has to be done by identifying the appropriate items and by going through the path structure. Therefore it might be appropriate to store metadata such as the archive description with the archived files (see [archiving](#)). Metadata for example may include names of projects or other relevant names.

Search via Command Line Program

The B/A client command line interface (dsmc) can also be used for searches. When executing the dsmc-program it will use the option file which has been setup in step [Configuring the B/A Client](#). The syntax of the dsmc-program is platform-specific; this example uses the Windows Client. For searches the Windows client requires the name of the drive (such as c:\) with each file and directory name. It also accepts wildcards (*, ?). The subsequent examples uses interactive mode, whereas the syntax is the same for batch mode.

When searching with the dsmc-program the directory and file information must be given while wild-cards can be used. This however requires strict naming conventions for file- and directory names as well as the directory structure. In the example included herein all project related information has been stored in a directory structure c:\temp\projects. The names of the projects follow this path and the file names include meaningful information (metadata) such as "ProjectPlan". The dsmc-program allows the following searches:

Search for all files in a directory:

Example (1): search for all files in directory "Project-A"

```
query archive c:\temp\projects\project-A\*
```

Alternative the search can be specified as:

```
query archive c:\*\*\*project-A*
```

The result is listed below:

```
tsm> query archive c:\*\*\*project-A*

Size  Archive Date - Time      File - Expires on - Description
----  -
0 B   08/16/2006 16:47:56      \\hausteintp\c$\temp\Projects\Project-A 08/21/2006
Archive Date: 08/16/2006
0 B   08/16/2006 17:55:11      \\hausteintp\c$\temp\Projects\Project-A 08/21/2006
Archive customer A project
0 B   08/16/2006 17:55:13      \\hausteintp\c$\temp\Projects\Project-A 08/21/2006
Archive customer A project
```

```
0 B 08/16/2006 17:55:13 \\hausteintp\c$\temp\Projects\Project-A 08/21/2006
Archive customer A project
```

Example (2) Search for all files stored in directory c:\temp\user

```
query archive c:\*\user\*\*
```

The result is listed below:

```
tsm> query archive c:\*\user\*\*
Size Archive Date - Time File - Expires on - Description
-----
22,158 B 08/16/2006 16:47:57 \\hausteintp\c$\temp\User\Dr-Dan\Dfile1.out
08/21/2006 Archive Date: 08/16/2006
61,348 B 08/16/2006 16:47:57 \\hausteintp\c$\temp\User\Dr-Dan\Dfile2.out
08/21/2006 Archive Date: 08/16/2006
76,989 B 08/16/2006 16:47:57 \\hausteintp\c$\temp\User\Dr-Dan\Dfile3.out
08/21/2006 Archive Date: 08/16/2006
```

Search for specific file names:

Example (3): search for all files with the pattern *Plan* in directory c:\temp\projects

```
query archive c:\*\*\*\*plan*
```

The result is listed below:

```
tsm> query archive c:\*\*\*\*plan*
Size Archive Date - Time File - Expires on - Description
-----
61,348 B 08/16/2006 16:47:56 \\hausteintp\c$\temp\Projects\Project-
A\projectPlan.mps 08/21/2006 Archive Date: 08/16/2006
61,348 B 08/16/2006 17:55:13 \\hausteintp\c$\temp\Projects\Project-
A\projectPlan.mps 08/21/2006 Archive customer A project
53,098 B 08/16/2006 17:55:13 \\hausteintp\c$\temp\Projects\Project
-X\ProjectPlan.mps 08/21/2006 Archive customer A project
```

Search on Archive Description:

Example (4): search for archive description "Archive customer A project"

```
query archive desc="Archive customer A project"
```

The result is listed below:

```
tsm> query archive -desc="Archive customer A project" c:\*\*\*\*
Size Archive Date - Time File - Expires on - Description
-----
83,135 B 08/16/2006 17:55:11 \\hausteintp\c$\temp\Projects\Project-A\notes.txt
08/21/2006 Archive customer A project
61,348 B 08/16/2006 17:55:13 \\hausteintp\c$\temp\Projects\Project-
A\projectPlan.mps 08/21/2006 Archive customer A project
76,989 B 08/16/2006 17:55:13 \\hausteintp\c$\temp\Projects\Project-A\wbs.ppt
08/21/2006 Archive customer A project
53,098 B 08/16/2006 17:55:13 \\hausteintp\c$\temp\Projects\Project-
X\ProjectPlan.mps 08/21/2006 Archive customer A project
107,147 B 08/16/2006 17:55:13 \\hausteintp\c$\temp\Projects\Project-
X\schedule.ppt 08/21/2006 Archive customer A project
```

Search on Archival Dates

Example (4): search file archived between 18.08. and 21.08.2006

```
q ar -dateformat=4 -fromdate=18.08.2006 -todate=21.08.2006 c:\*\*\*\*
```

The result is listed below:

```
tsm> q ar -dateformat=4 -fromdate=18.08.2006 -todate=21.08.2006 c:\*\*\*\*
Size Archive Date - Time File - Expires on - Description
-----
0 B 21.08.2006 16:56:08 \\hausteintp\c$\temp\User\Nils 26.08.2006 Archive Date:
08/21/2006
61,348 B 21.08.2006 16:56:08 \\hausteintp\c$\temp\User\Nils\Nfile1.out
26.08.2006 Archive Date: 08/21/2006
```

Retrieval

File retrieval is usually done manually, after a search has finished and the appropriate file has been found. For retrieval the B/A client GUI or the command line interface (dsmc) can be used.

When using the B/A client GUI select the item **Retrieve** from the first menu. Select the file to be retrieved after the file has been found as a result of a [search](#). Figure 20 presents the **Retrieve** view with selected files and directories.

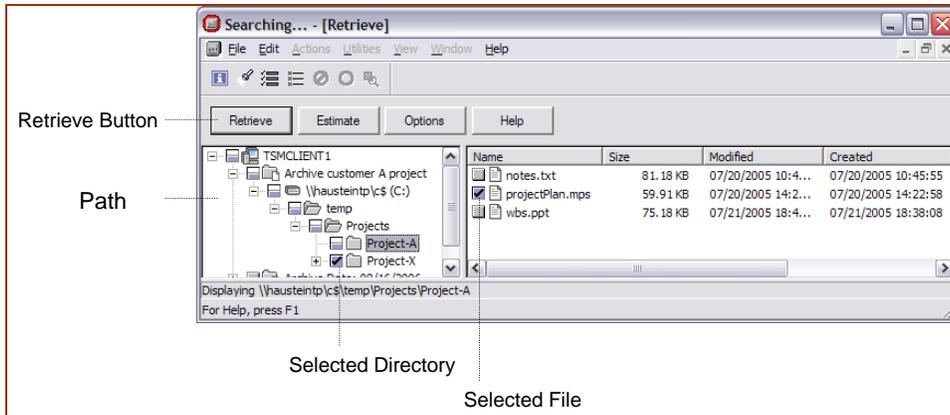


Figure 20: Retrieve view with selected files and directories.

Before retrieving data certain retrieve options can be selected by pressing the **Options** Button. The options are shown in figure 21.

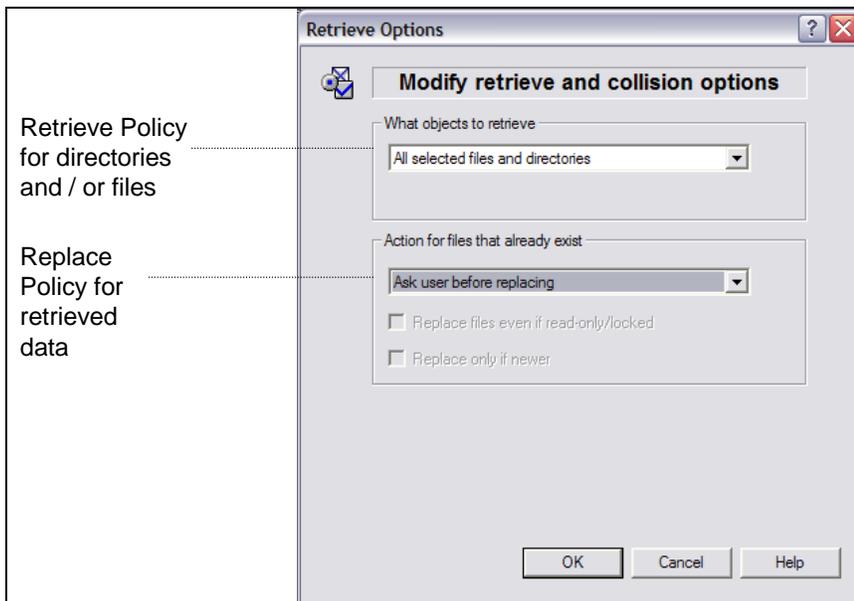


Figure 21: Retrieve Option.

It is recommended when retrieving files and directories to always ask before replacing files.

After the options have been setup accordingly press the retrieve button. Another menu may appear as shown in figure 22.

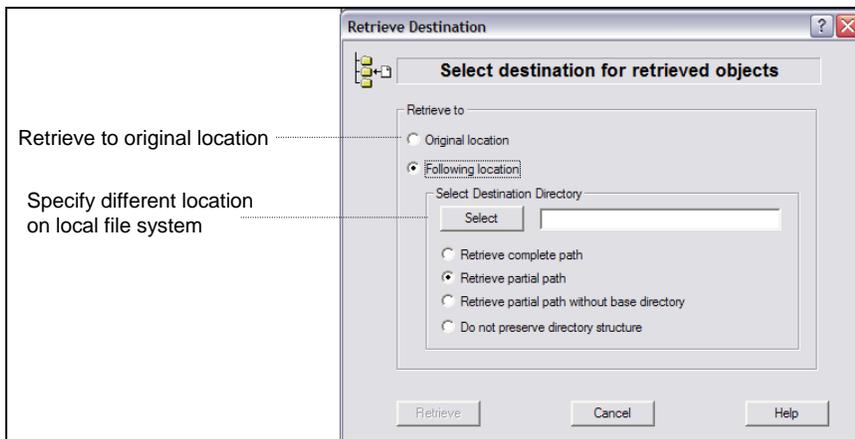


Figure 22: Retrieve Menu

In this menu it is possible to specify the location for retrieval. It is recommended selecting a location which is not identical to the original location in order to prevent confusion. It may be appropriate to setup a directory in the local file system for retrieved files.

Managing Retention

Depending on the configured retention policy a file has been assigned to during archiving the retention time can be managed differently (see section [SSAM Retention Policies](#)).

For a chronological retention policy, which has been used in the example of this document, there is nothing which can be managed: upon expiration of the retention time a file will be deleted automatically. The B/A client will subsequently not show this file anymore in the [Retrieve View](#).

For an event-based retention policy the start of the retention time can be triggered by an event (see Retention Policies) for one or more files. The event can be initiated using the B/A Web Client or the command line

Sending Retention Events via Web Client

The B/A client allows managing retention times for files assigned to event based policies through the TSM Web client. The TSM Web Client is a part of the B/A client and the service must be configured through the B/A client GUI. To accomplish this select from the main menu of the BA/client GUI (see [Archive Client GUI](#)) the menu item **Utilities – Setup Wizard** and select **Configure TSM Web Client**.

Once the TSM Web Client is configured, start the TSM Client Acceptor service. This is a service which can be found under services in administrative tools (under Windows®) The TSM Web client is launched by opening a web browser and entering the address: localhost:1581. From the main menu select **Action** and then select **Set Data Retention Events** from the main menu (see figure 23).

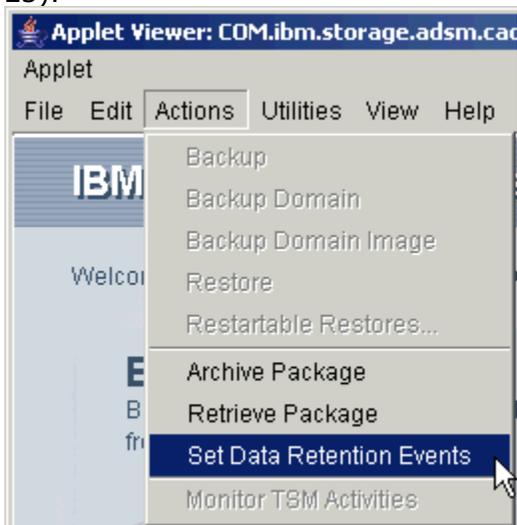


Figure 23: Managing Retention

Subsequently a view is opened showing the archives by description (see [Search via Retrieve View](#)). According to figure 24 files and / or directories and the event type can be selected. When pressing the Set Event button, the selected event will be issued to the selected files.

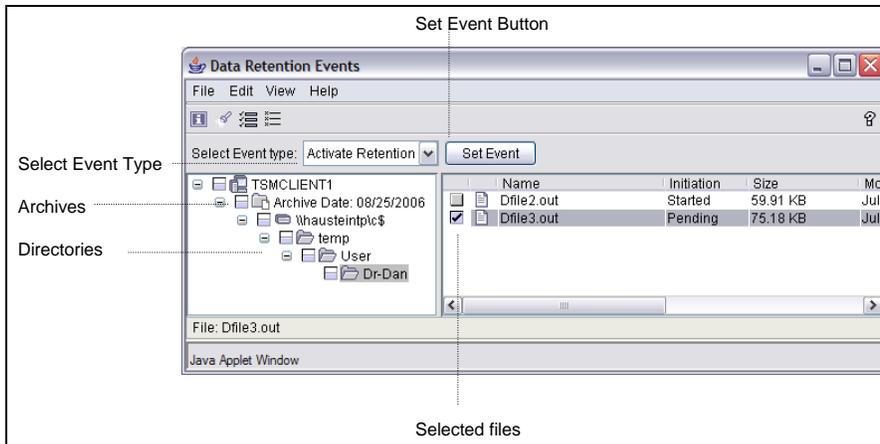


Figure 24: Setting Events via Web Client

The event type can be:

Activate Retention: activate a retention event for the selected file(s).

Hold: deletion hold causes the selected file(s) not to be deleted after expiration.

Release: deletion release causes a previously sent deletion hold to be released for the selected file(s).

Sending Retention Events via Command Line

The B/A client command line (dsmc) offers also a way to initiate an event to one or more files.

The command syntax is:

```
set event -type=a <directory and file pattern>
```

The event type can be:

A: activate a retention event for a file or set of files specified by the pattern.

H: deletion hold causing file(s) not to be deleted after expiration.

R: deletion release causing a previously sent deletion Hold to be released.

In the example below an event is sent to a file c:\User\Dr-Dan\Dfile1.out which has previously been archived and an event-based retention policy was assigned. When querying the file, the expiration time is set to "Never".

```
tsm> q ar c:\*\*\*dfile1.out
Size  Archive Date - Time      File - Expires on - Description
-----
22,158  B  08/25/2006 14:25:09  \\hausteintp\c$\temp\User\Dr-Dan\Dfile1.out  Never
Archive Date: 08/25/2006
```

Subsequently the event is sent which activates the retention time for that object:

```
tsm> set event -type=a c:\temp\user\dr-dan\dfile1.out
Rebinding-->          22,158  \\hausteintp\c$\temp\User\Dr-Dan\Dfile1.out [Sent]

Total number of objects archived:          0
Total number of objects failed:           0
Total number of objects rebound:         1
Total number of bytes transferred:        0 B
Data transfer time:                       0.00 sec
Network data transfer rate:               0.00 KB/sec
Aggregate data transfer rate:             0.00 KB/sec
Objects compressed by:                    0%
Elapsed processing time:                  00:00:03
```

When checking the object again the retention time has been adjusted reflecting an actual data. The expiration date is set to two days later than the archiving date, thus the retention time of the file is two days after sending the event.

```
tsm> q ar c:\*\*\*dfil*
Size  Archive Date - Time      File - Expires on - Description
----  -
22,158  B  08/25/2006 14:25:09    \\hausteintp\c$\temp\User\Dr-Dan\Dfil
el.out  08/27/2006 Archive Date: 08/25/2006
```

Appendix

References

- [1] Spectrum Protect archive client documentation
http://www.ibm.com/support/knowledgecenter/SSGSG7_7.1.6/client/t_arc_ret.html
- [2] Spectrum Protect client and server operating systems:
<http://www-01.ibm.com/support/docview.wss?uid=swg21243309>
- [3] Spectrum Protect Backup and Archive Client download:
<http://www-01.ibm.com/support/docview.wss?uid=swg21239415>
- [4] Spectrum Protect client and server compatibility
<http://www-01.ibm.com/support/docview.wss?uid=swg21053218>
- [5] TSM Application programming interface
http://www.ibm.com/support/knowledgecenter/SSGSG7_7.1.6/api/t_developing.html
- [6] Spectrum Protect for Data Retention solution (SSAM) whitepaper
<https://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102624>
- [7] Assessment report for SSAM Version 6.3
<http://www.kpmg.de/bescheinigungen/RequestReport.aspx?38076>
Assessment report for Spectrum Protect for Data Retention Version 8.1:
<http://www.kpmg.de/bescheinigungen/RequestReport.aspx?56B00E998BA14B31AF2E0F0FB63F0034>
- [8] Configuring the B/A client
http://www.ibm.com/support/knowledgecenter/en/SSGSG7_7.1.6/client/c_cfg.html
- [9] B/A client command line parameters:
http://www.ibm.com/support/knowledgecenter/en/SSGSG7_7.1.6/client/r_client_ref.html
- [10] Archiving files from Windows:
http://www.ibm.com/support/knowledgecenter/en/SSGSG7_7.1.6/client/c_arc_windows.html
- [11] B/A client processing option including include and exclude statements:
http://www.ibm.com/support/knowledgecenter/en/SSGSG7_7.1.6/client/c_opt_bkuparchive.html
- [12] Schedule operation with B/A client and SSAM server:
http://www.ibm.com/support/knowledgecenter/en/SSGSG7_7.1.6/client/t_sched_baclients.html
- [13] Total cost of ownership studies tiered storage solutions with tape
<http://www.clipper.com/research/TCG2013009.pdf>
<http://www.esg-global.com/blogs/active-archival-storage-a-cost-of-ownership-analysis/>
- [14] Rebranding of System Storage Archive Manager to Spectrum Protect for Data Retention:
<http://www-01.ibm.com/support/docview.wss?uid=swg21963634>

Disclaimer

This document reflects the understanding of Author on many of the questions asked about Spectrum Protect for Data Retention solutions with IBM hardware servers and software. It was reviewed by the members of the IBM Software group. This document is presented "As-Is" and IBM does not assume responsibility for the statements expressed herein. It reflects the opinions of the Author. These opinions are based on several years of joint work with the IBM Server- and Software Group. If you have questions about the contents of this document, please direct them to the Author (nils_haustein@de.ibm.com).

The Techdocs information, tools and documentation ("Materials") are being provided to IBM Business Partners to assist them with customer installations. Such Materials are provided by IBM on an "as-is" basis. IBM makes no representations or warranties regarding these Materials and does not provide any guarantee or assurance that the use of such Materials will result in a successful customer installation. These Materials may only be used by authorized IBM Business Partners for installation of IBM products and otherwise in compliance with the IBM Business Partner Agreement."

Trademarks

The following terms are registered trademarks of International Business Machines Corporation in the United States and/or other countries: AIX, IBM Spectrum Protect, DB2

A full list of U.S. trademarks owned by IBM may be found at:

<http://iplswww.nas.ibm.com/wpts/trademarks/trademar.htm>.

UNIX is a registered trademark in the United States and other countries licensed exclusively through The Open Group.

LINUX is a registered trademark of Linus Torvalds.

Windows is a trademark of Microsoft Inc. in the US and other countries.

Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. in the United States and/or other countries.

Other company, product and service names may be trademarks or service marks of others.

Acknowledgement

Special thanks to Ron Henkhaus (Spectrum Protect SWAT team, USA) for the thorough review of the document and the great additions.

I would also like to thank Klemens Poschke (Technical Sales Tivoli) for the contributions.