Tivoli® software

IBM

IBM Tivoli Storage Manager Suite for Unified Recovery – Archive Option

Usage and Licensing Best Practices Guide

June 2012

Table of Contents

Disclaimer:

Introduction

The IBM Tivoli Storage Manager family of products is a leader in unified recovery management, providing the ability to protect a wide range of systems including virtual machines, file servers, email servers, databases, mainframes, and even desktops from a single administration interface. Tivoli Storage Manager has also been an innovation leader in data lifecycle management, with a robust range of automated data placement, migration, retention and expiration capabilities.

The IBM Tivoli Storage Manager Suite for Unified Recovery (TSM SUR) provides storage and backup administrators with a simplified, alternative method for procuring and deploying the data protection and retention solutions, throughout the infrastructure, that meet the needs of their particular environment. With this bundled offering, you can deploy any of ten different solution components, in any location and quantity, with a simplified license that measures only the amount of data being managed. License costs for backup data can be reduced through the use of built-in source and target data deduplication, and there is no charge for duplicate copies of the data.

This guide describes the usage of the Archive Option licensing tier available with TSM SUR, and how users can take best advantage of the potential cost savings that it offers.

To benefit from this new pricing option, existing TSM SUR customers might need to change the way they manage their long-term data retention processes, to utilize the archiving feature within TSM rather than just storing backup data sets forever. Without making these changes, customers using this archiving option are likely to end up paying more in software license fees, because the data that they put into archive will also remain in backup pools.

Note: IBM is recognized by many industry analysts as a leader in archive software, and Tivoli Storage Manager's archiving capabilities contribute to this recognition.

To learn more, please visit: http://www.ibm.com/software/data/smart-archive/

IBM Tivoli Storage Manager Suite for Unified Recovery – Archive Option

The licensing for TSM SUR is based on the amount of data (number of terabytes) stored in TSM Primary Storage Pools and any TSM FastBack repositories. Data in Copy Pools or otherwise replicated is not counted for license purposes. As the data in the licensed pools grows, the price of additional TB licenses is reduced according to a 7-tier structure, as shown here:

Tier (TB)	Price Reduction
1-100	
101-250	10%
251-500	20%
501-750	40%
751-1250	50%
1251-2000	60%
2001+	70%
Archive Option	80%

The Archive Option adds an eighth tier, priced at an 80% reduction from the first tier, for data that is stored by TSM as archive data (from an archive operation by the TSM backup-archive client or from an archive transfer via the TSM API) and stored on tape or virtual tape.

The use of the archive tier can potentially save customers a significant amount of licensing costs, but care must be taken in the way that data is managed in order to achieve these savings and to avoid the potential for unintended cost increases.

Backup vs. Archive – the basics

Backup = operational recovery of current data
Archive = long-term retention and retrieval of non-operational data

Many organizations – analysts estimate more than 50% - use their data backup systems for long-term retention as well as short-term operation recovery capability. This is appropriate if the data in the backup system can be properly categorized and have appropriate retention/expiration policies applied. Tivoli Storage Manager has supported this concept since its inception almost 20 years ago and it is an important part of its Unified Recovery Management capability.

But retaining standard backup data sets for long periods to meet 'archive' requirements is not an ideal practice. It can create an ever-increasing amount of duplicate data, raising overall costs (TCO) and complicates the administrator's ability to retrieve the right files, or the right version of files, when requested by the business. And within the capacity-based licensing model of Tivoli Storage Manager Suite for Unified Recovery, keeping backups long term is not cost efficient.

Much of the data that is archivable is not duplicate, but simply old. The benefit of archiving vs. long-term retention in the production environment for this type of data it is not intuitively obvious.   Big data can have a significant, crippling effect on every "downstream" process that touches this data, including backup operations. Aged data that remains on the host server or application without proper archiving can cause congestion in the entire data protection stream, impacting networks, primary storage, secondary storage, replication, and disaster recovery. This is true of both unstructured data (which TSM archiving can address) and structured data (which IBM and other vendor's Enterprise Content Management products can address).

A proper archive strategy, which moves older, non-operational data from production systems to long-term storage media such as tape or virtual tape, will free up (more expensive) production storage capacity for new data, reduce the amount of data in the backup stream, and simplify long-term retention and expiration processes.

Removing archivable data from production systems is the key to maximizing cost savings, as shown in this simple scenario:

1. A new file, MYWORK_V1.DOC, is created and backed up during the next backup cycle

2. You edit MYWORK_V1.DOC and save it as MYWORK_V2.DOC. MYWORK_V1.DOC is now eligible for archiving, since the edited file is now the working version.

3. If you do nothing with the production data, you will retain both copies of the file in your backup system forever. But if you archive MYWORK_V1.DOC and delete it from the

production system, it will eventually be deleted from the backup repository according to your backup expiration policy.

4. If, however, you archive MYWORK_V1.DOC but do not delete it from the production system, you will have a copy of the same file in both the backup and archive repositories, resulting in increased costs and complicated data lifecycle management.

A common approach to archiving is to create a full copy of current data on a monthly, quarterly or yearly basis, depending on business requirements. If you do this using the TSM archiving functionality, you can set your backup expiration policies on the same schedule, to automatically eliminate data that has been archived from the backup repository after it has been deleted from the production system.

Following a proper archiving strategy, where data that is no longer needed operationally is archived and deleted from the production system, allows you to take full advantage of the preferential pricing of the Tivoli Storage Manager Suite for Unified Recovery – Archive Option.

TSM SUR Archive Option – use cases

Customers can benefit from the purchase of TSM SUR – Archive Option licenses in the following scenarios:

1. Archive, and then delete the original data
   - The greatest benefit from implementing TSM archive (from a cost perspective) is realized when customers choose to archive data to TSM and ultimately delete this data from the source (either at the time of archive or shortly thereafter). This is true for both new and existing TSM customers.

   - This may be done by using the TSM backup-archive client (for unstructured data) or by using a specific content management or archive application (for structured data) which then ultimately sends its data to the TSM storage hierarchy for long-term retention. This practice has many advantages. It removes data from the existing backup stream, reduces the amount of future TSM SUR backup license required, and frees up capacity on the source server or workstation.

   - Customers would then purchase an appropriate TSM SUR - Archive Option license at a lower cost, thus realizing significant savings vs. using the higher-cost TSM SUR backup licenses for this archived data. Any backup licenses which are freed up during this process may be applied to future data growth realized on the source servers or workstations.

2. For new TSM SUR customers
   - New customers moving to TSM SUR have the option of sizing their required capacity (and software licenses) based upon some combination of backup and archive data. These customers would purchase a portion of their licenses as backup, and a portion as archive.

- The exact mix can be determined by calculating the amount of data stored for short-term retention (typically at least 30 days to as long as 180 days) and the amount of data stored for long-term retention (everything beyond the short-term retention period).

- Much, if not all of the short-term data would be stored on TSM disk, and the appropriate deduplication rate would be applied as required. Much, if not all of the long-term data would be stored on media other than TSM disk, and no deduplication would be considered in the license calculation.

- New customers can schedule periodic archive operations (weekly, monthly, quarterly, or as needed) to meet existing business requirements. These archives are essentially "full" copies of data. As a best practice, the data that is archived and no longer needed on the server or workstation should be deleted from the source system.

3. For existing TSM SUR customers

- Existing customers currently running TSM archive operations in addition to their TSM backup operations can purchase and immediately apply TSM SUR Archive Option licenses against the total amount of archive data in the tape or virtual tape (VTL) primary pools as reported by TSM. These customers would then "free up" any licenses previously purchased for TSM SUR backup, and reuse these for future growth of their backup workloads. This would potentially defer the need to purchase new TSM SUR licenses for backup in the near term.

- Existing customers might want to begin using TSM archive for the first time in order to meet compliance or retention requirements of the business or some outside regulatory agency. These customers would create and schedule archive operations in accordance with their business needs (weekly, monthly, quarterly, or as needed).

- After the archive operations have been running for a period of time (for example, 90 days), the customer can choose to discontinue retention of extra versions of their corresponding backup data. This partially offsets the additional capacity that is consumed by the new archive workload.

  - **Attention:** When reducing backup retention periods, older versions of files that are no longer in the production systems will be deleted. Ensure that these files are not needed for potential regulatory or legal discovery in the future before shortening the retention settings.

  - TSM does not include a function to automatically move data from a backup status to an archive status. Existing customers with large amounts of old backup data can consider restoring that data to a temporary repository, and then archiving it, to take advantage of the reduced pricing of the TSM SUR Archive Option.

  - It is possible that a customer would initially see a net increase in overall TSM capacity driven mostly by new archive data, and a small decrease in TSM backup capacity resulting from the change to their retention policies. Over time, as older data continues to be placed in archive rather than backup, significant cost savings may be achieved.

When to backup and when to archive

When the Tivoli Storage Manager backup-archive client backs up or archives a file, it sends a copy of the file and its associated attributes to the TSM server; however, backups and archives have different goals.
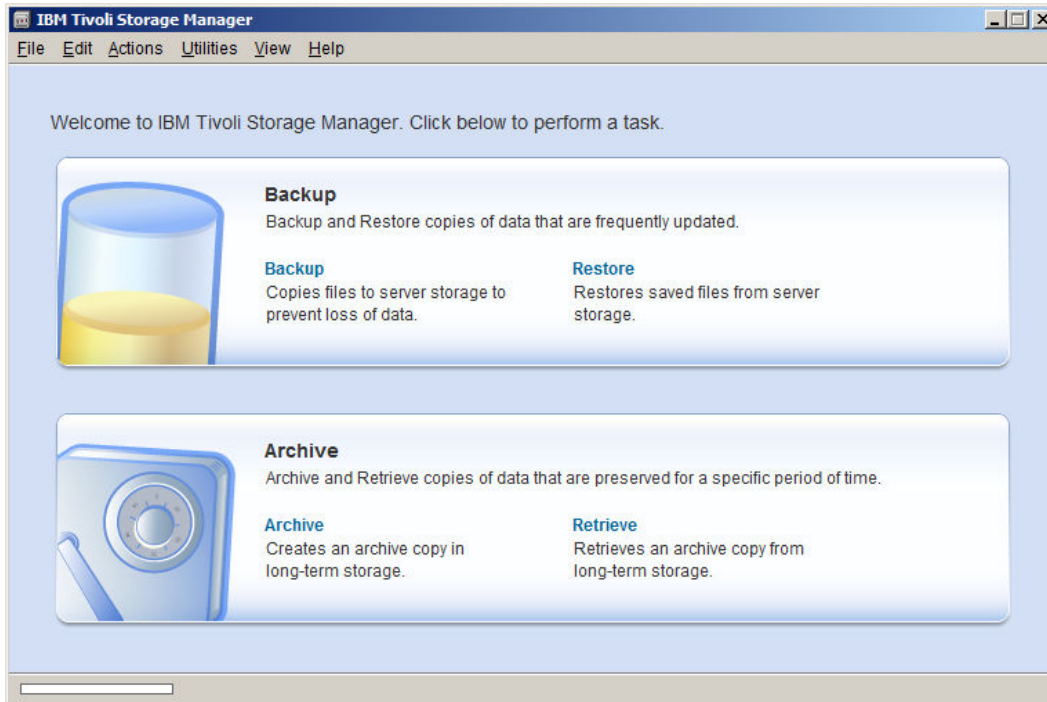
Use backups to protect against unforeseen damage to your files, and use archives for maintaining more permanent versions of your files. Backup data is managed by version, using predetermined policy based rules. Using these rules, the Tivoli Storage Manager administrator can control the following:

- The number of versions
- The number of days each additional backup copy is kept
- What happens to backup data versions when the file is deleted on the client system

Each copy of the file stored on the server is considered to be a separate and unique version of the file.

Archive is a powerful and extremely flexible mechanism for storing long term data. Archive data is kept for a specified number of days or until an event occurs. Archive has no concept or support for versions. The user or Tivoli Storage Manager administrator is responsible for determining what files get added to an archive.

Tip: If archive is run multiple times against a file using the same archive description, a new copy of the file is added to the archive each time that archive is run. To simplify retrieve, store only one copy of a file in each archive.

Backups protect against file damage or loss that can occur through accidental deletion, corruption, disk crashes, and so forth. The server maintains one or more backup versions for each file that you back up. Older versions are deleted as newer versions are made. The number of backup versions that the server maintains is set by your administrator.

Archive copies are saved for long-term storage. The server can store an unlimited number of archive versions of a file. The archive retention period can be set by length of time from when the file was created, by an event, or by a combination of both.
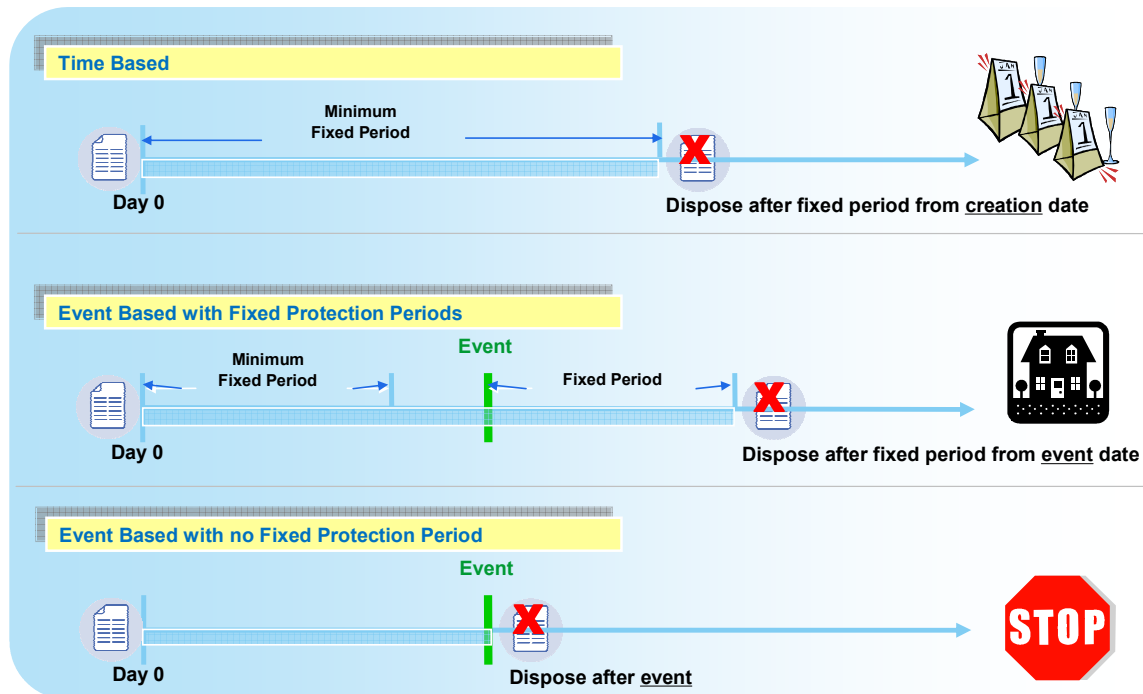
Archives are useful if you need to go back to a particular version of your files, or you want to delete a file from your server or workstation and retrieve it at a later time, if necessary. For example, you might need to save spreadsheets for seven years for tax purposes, but because you are not using them, you do not want to leave them on your workstation.

An advanced Storage Resource Management tool, such as IBM Tivoli Storage Productivity Center, can help scan and identify data files that are appropriate for archiving or deletion, based on criteria that you set, such as length of time since the file was created, modified or last accessed.

Not only does Tivoli Storage Productivity Center guide you toward the right data to archive, it also allows you to begin automating this process so that the administrators do not have to repeatedly deal with this step in the information lifecycle management (ILM) process.

Potential Savings with the TSM SUR Archive Option

How much money can you save with the Archive Option? The answer to this question depends on a wide range of variables, including how much data you archive and delete, your backup and retention policies, the rate of data growth and change, and so on.

In the best case as you are starting out with TSM SUR and its capacity based licensing, every terabyte of data that goes into archive instead of backup could save you up to 80% on your license and S&S costs because the license cost for the Archive Option licensing tier is 80% less than the first backup tier. That percentage will decrease as you grow into the higher backup licensing tiers because those tiers are less expensive than the first tier.

And the use of data deduplication on your backup data, which can reduce your backup licenses by 40% or more depending on the nature of the data being protected, will also narrow the savings derived from archiving.

Note: Tivoli Storage Manager does not support data deduplication on tape. We view deduplication on physical tape as a potentially dangerous practice, especially when your archive data spans multiple tape cartridges, as it makes it difficult (if not impossible) to guarantee the ability to keep consistent data sets over time. It also can significantly impact retrieval performance, even on a single tape, as the deduplicated chunks will need to be read from different places on the tape(s) in order to re-hydrate the file. If data deduplication is desired in the tape tier, it is best performed by a virtual tape library such as IBM ProtecTIER.

Tivoli Storage Manager Suite for Unified Recovery – Archive Option also supports archiving data to Virtual Tape Libraries (VTL), such as IBM ProtecTIER. In these cases, the data may be deduplicated by the VTL, though TSM will have no visibility into that process. The amount of archive data that would be licensed with the Archive Option is for the fully hydrated (non deduplicated) files whether the data is sent to tape or virtual tape.

Additional Resources:

For estimating data capacity for the Tivoli Storage Manager Suite for Unified Recovery – Archive Option, a modified version of the capacity script is available. This script and instructions for running it are available here:
https://www-304.ibm.com/support/docview.wss?uid=swg21500482&wv=1

Backup and archive data applicable to TSM SUR and TSM SUR – Archive Option pricing will be measured and reported as shown in the example report below.

```
************************************************************
**** Tivoli Storage Manager Suite for Unified Recovery ****
****          Terabyte (TB) Capacity Report          ****
****        For Use with V6.1 or V6.2 Servers        ****
****************** (macro version: 1.4) ******************

Capacity Report for server:  SS2_ADSM_GROUP_SERVER

Report generation date:  2012-02-17

Data Managed by TSM Suite for Unified Recovery
Total TB within PRIMARY storage pools: 17.96

*---------------- TB by Backup or Archive ----------------*
Total BACKUP TB within PRIMARY storage pools: 14.69

Total ARCHIVE TB within PRIMARY storage pools: 3.26

*---------------- Deduplication Benefits ----------------*
TSM Data deduplication resulted in
  TB being excluded from measurement: 11.49

*---------------- Notes and Explanations ----------------*
* Copy Pool data represents DR copies of data and
   is excluded from measurement
* Active Data Pools are a special type of Copy pool
   and Active Data Pools data is excluded from measurement
* Virtual Volume data is data from a source TSM Server stored
   as an archive on a target TSM server. For reporting
   purposes Virtual volume data is only counted on the source server
* Data from TSM for SharePoint and HSM for Windows are
   excluded from the capacity measurement. These are
   OEM products that are purchased separately
* Data Replicated from FastBack to a TSM Primary Storage
   pools is also excluded from measurement, this data
   represent DR copies of data and is excluded
   from TSM Suite for Unified Recovery capacity
* Other Duplicated Data represents data reported in
   PRIMARY storage that is a duplication
   of data already stored through other means.

*---------------- Details on Excluded Data ---------------*

      Storage pools of type COPY or ACTIVE: 29.87 TB
```

For more information, please visit the IBM Tivoli Storage Manager Information Center:
http://publib.boulder.ibm.com/infocenter/tsminfo/v6r3/index.jsp

TSM Backup/Archive Client Installation and User's Manuals:
        UNIX-based platforms   Windows platforms

Whitepaper: Anatomy of an Archiving Project

Appendix A: Implementing the TSM Server for the TSM SUR-Archive Option

To take advantage of the TSM Archive Option, the TSM server must be set up correctly. The archive data must reside on VTLs, or tape, or both. Therefore you will need to define the necessary VTL and tape device classes, and then the primary storage pools that use those device classes. You will then need management classes with archive copy groups that have those storage pools as destinations. If necessary create one or more management classes for archiving data, and specify the VTL/tape primary storage pools as destinations.

The example below illustrates how to set up a TSM server using the TSM Administration Center GUI. You can of course also perform all of the set up steps using the TSM administration command line interface. The steps shown assume you have already logged into the Admin. Center and have navigated to Server Policy Domains.

Start by choosing to modify the necessary Server Policy Domain...

Next create a new or modify an existing management class for archiving to VTL or tape by selecting the appropriate action from the Action pull-down menu...
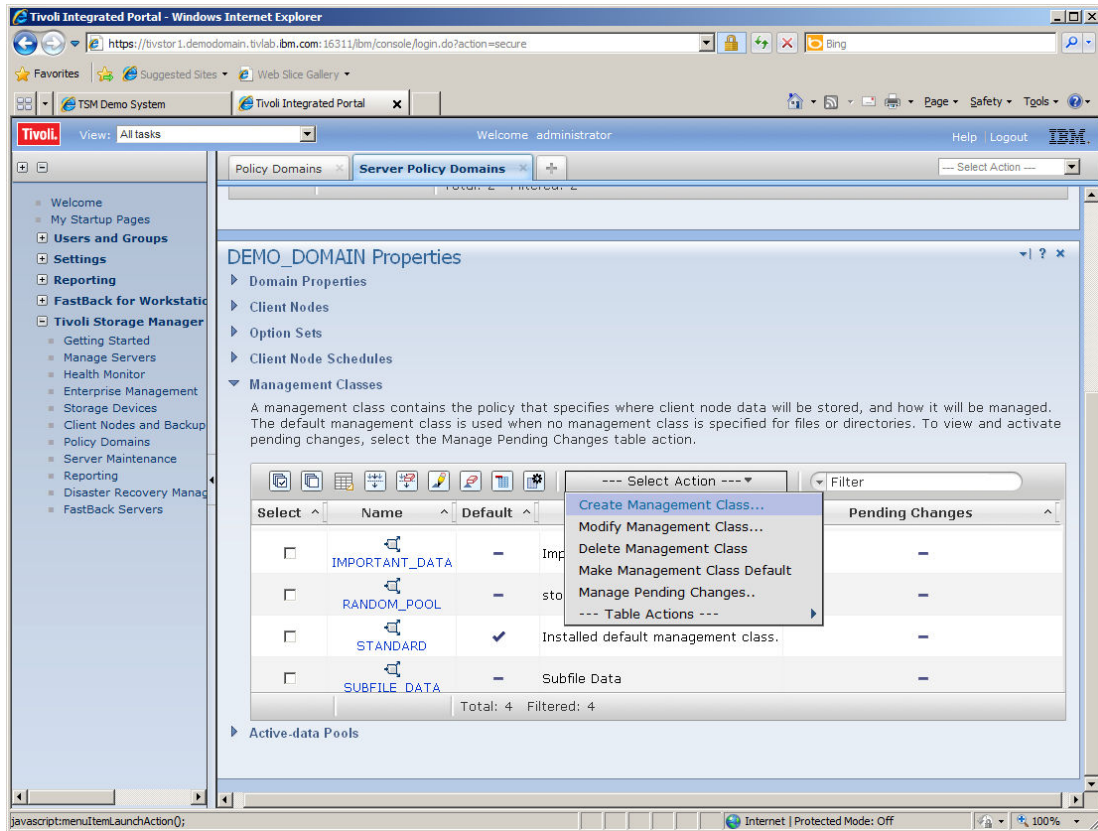
You can also choose either event-based retention or creation-based retention. Creation-based retention, also called chronological or time-based retention, is the default. In this example we're setting it up for archive only and using creation-based retention…
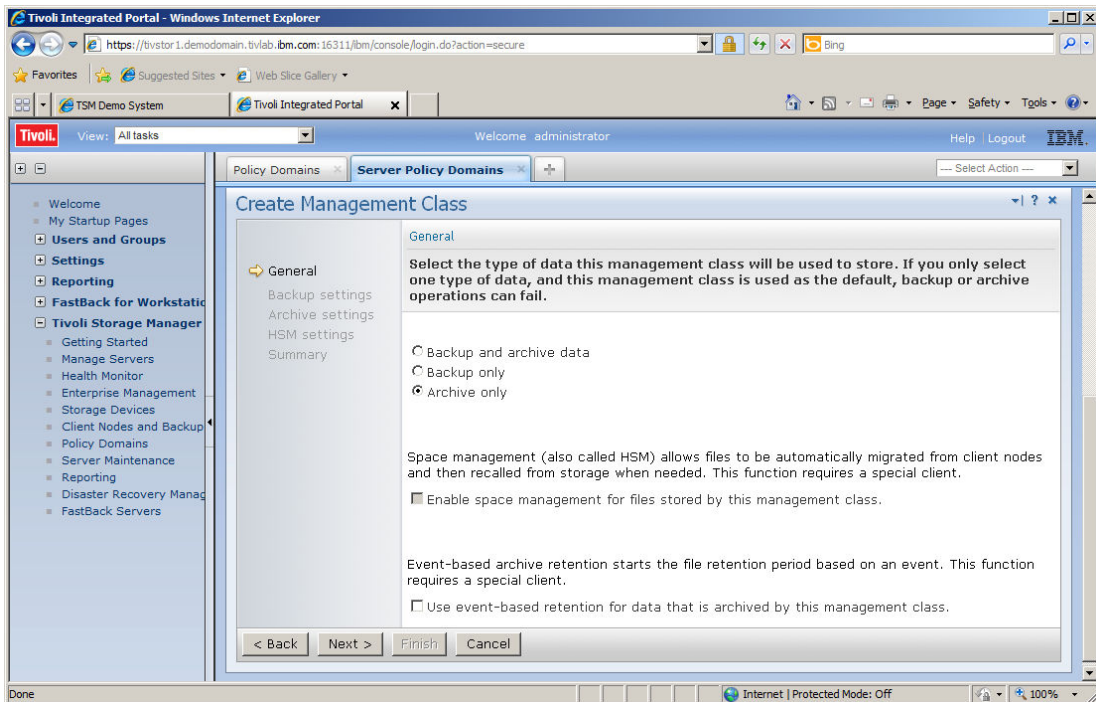

Give the management class a name and description. This management class will send the archives directly to the VTL.
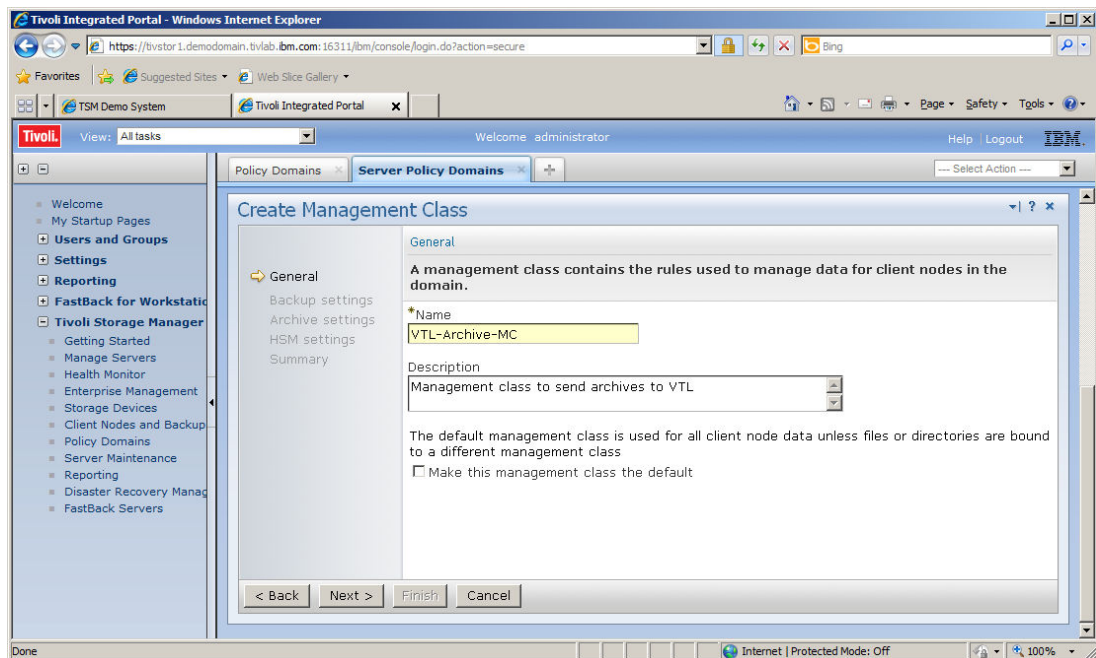
Next you choose the destination storage pool and retention period, as shown below. The destination storage pool in this example is VTL_POOL, and the retention period is 5 years (1826 days).
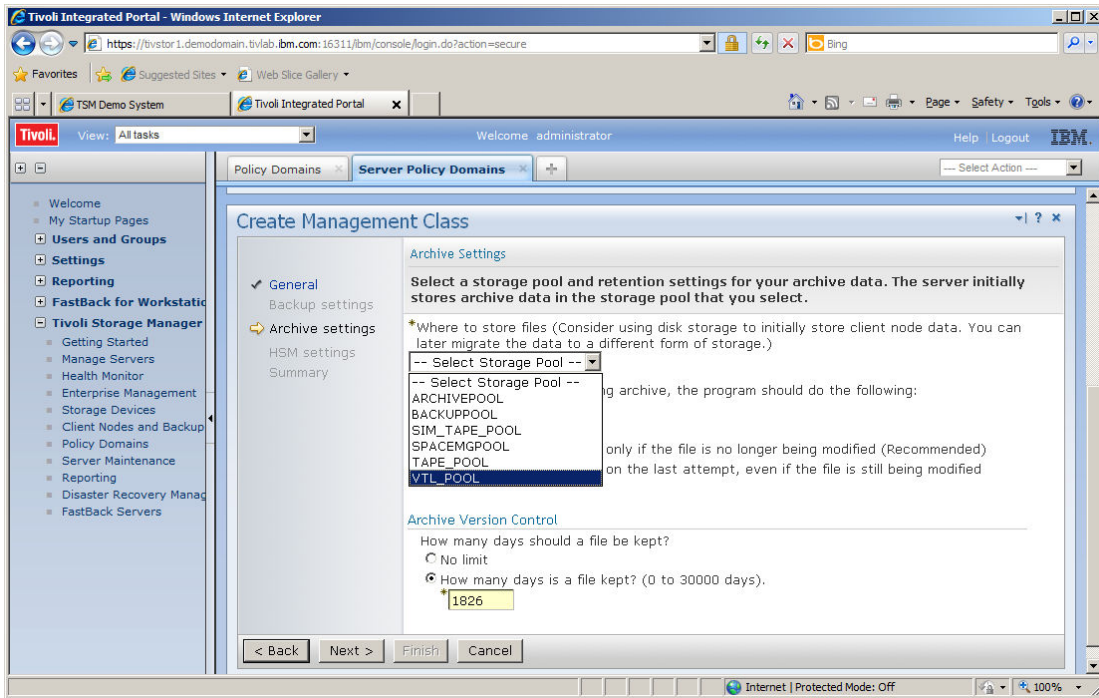

After you click Next, you should get a screen like the following confirming the creation of the management class.
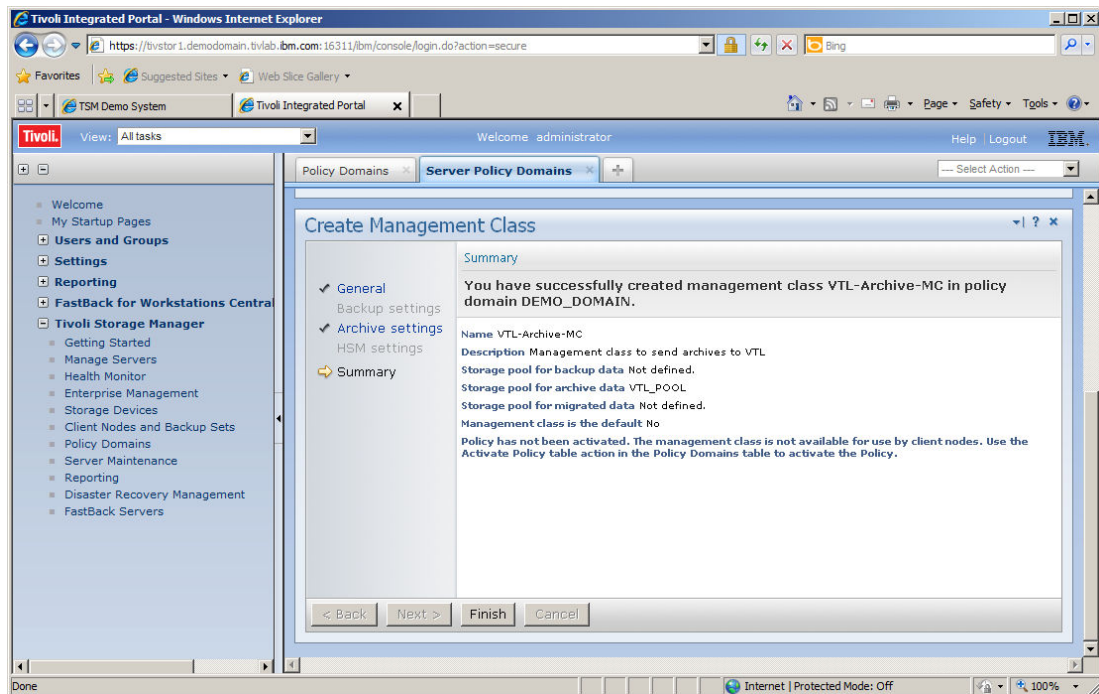
Simply click Finish, and this will take you back to the Management Classes screen. Although the new Management Class has been created, it has not been activated. You can create more management classes before performing the activation. For example, create one for sending archives to tape. At this point, though, we will do the activation by clicking Manage Pending Changes.
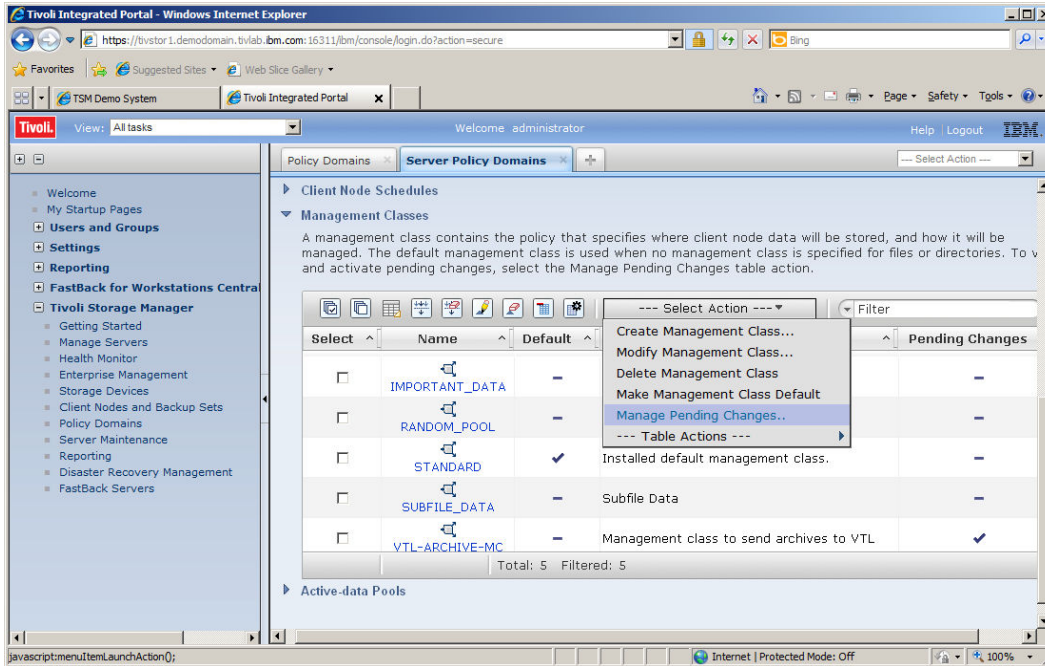
Then click the Activate Changes button, and your server will be ready for archiving...
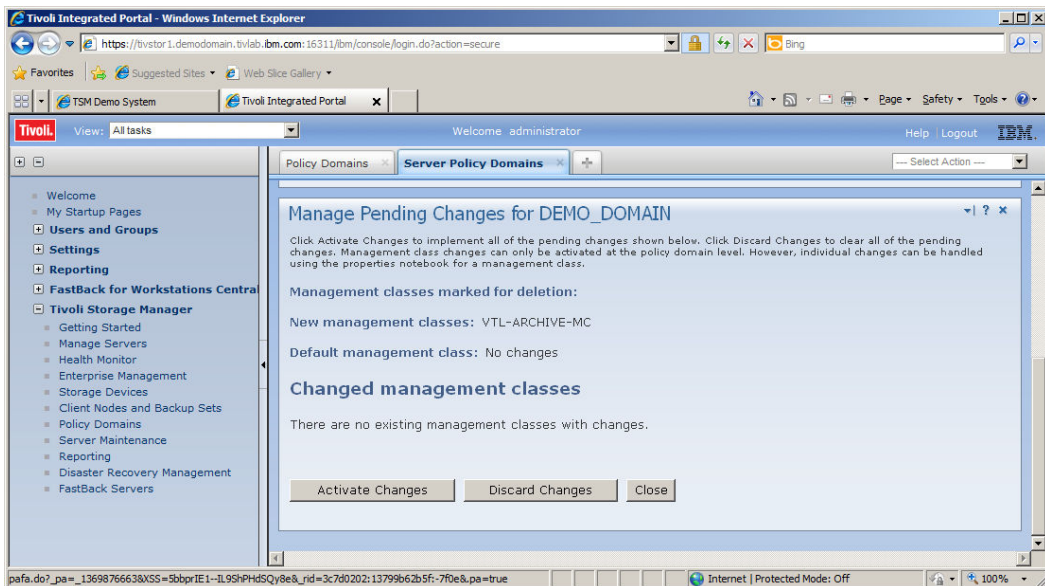
Additional information:

Management classes and copy groups:
http://publib.boulder.ibm.com/infocenter/tsminfo/v6r3/topic/com.ibm.itsm.client.doc/c_mgtc_using.html

## Appendix B: Exploiting the TSM SUR-Archive Option

How do you exploit the TSM Archive Option, to save on capacity licensing costs? You must archive files but do it in a way that stores those files on VTL or tape, and if necessary, you must reduce or prevent creating unnecessary copies of backups of those files.

When should you archive files? Over time, a server or workstation can accumulate many files, and possibly many versions of the same file, which may be accessed infrequently, if ever again. These files are good candidates for archival. You can also archive files proactively, as files change, and not accumulate an excess of backup copies. The example given assumes the first type of situation.

In the example below (shown by the Restore screen of the backup-archive client GUI), we have 8 versions of the same file, ALargeChangingDoc. One is active and the other 7 are inactive. The number of inactive versions retained over time is controlled by the VEREXISTS and RETEXTRA parameters of the management class backup copy group. In this example, the pertinent management class keeps 10 versions of an existing file.
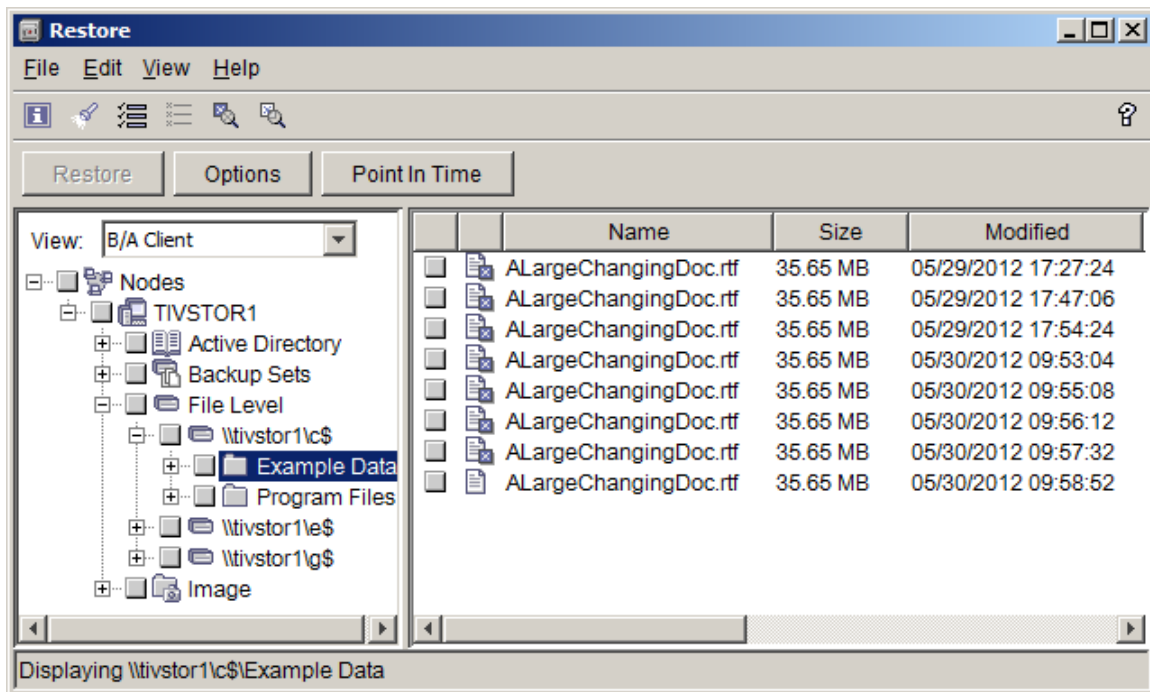
To archive this file, to start the process of exploiting Archive Option licensing, click on the Archive option of the backup-archive client GUI …
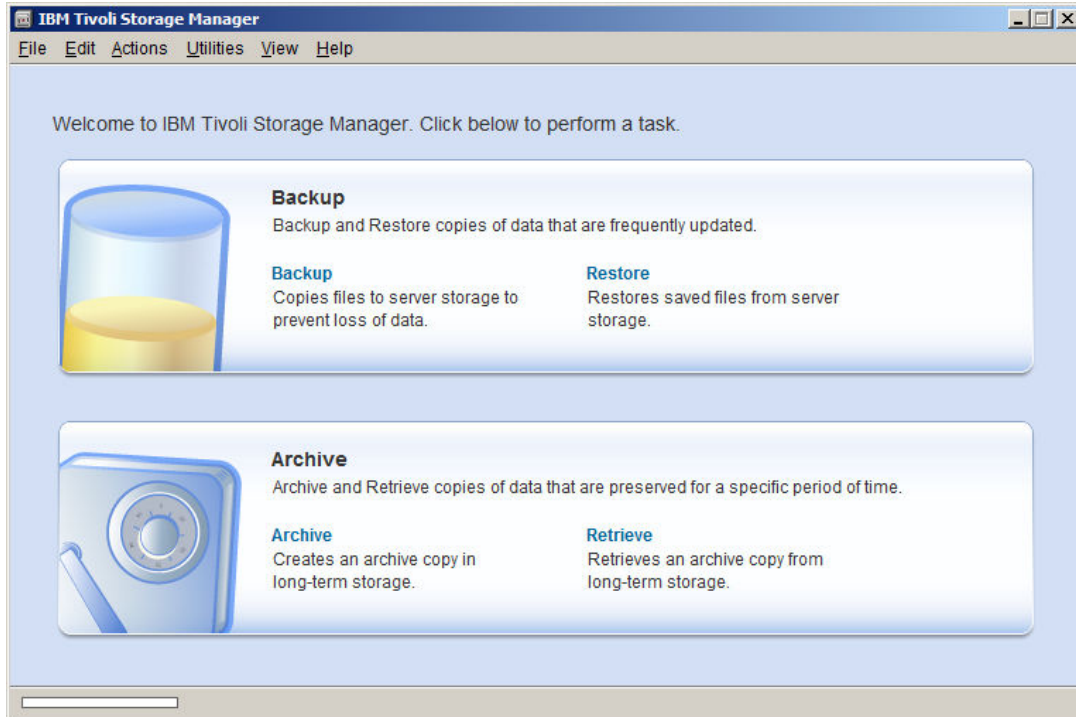

Then navigate to the files to be archived as shown below. The example also shows a customized description. The description can be used later to search for a particular archive.
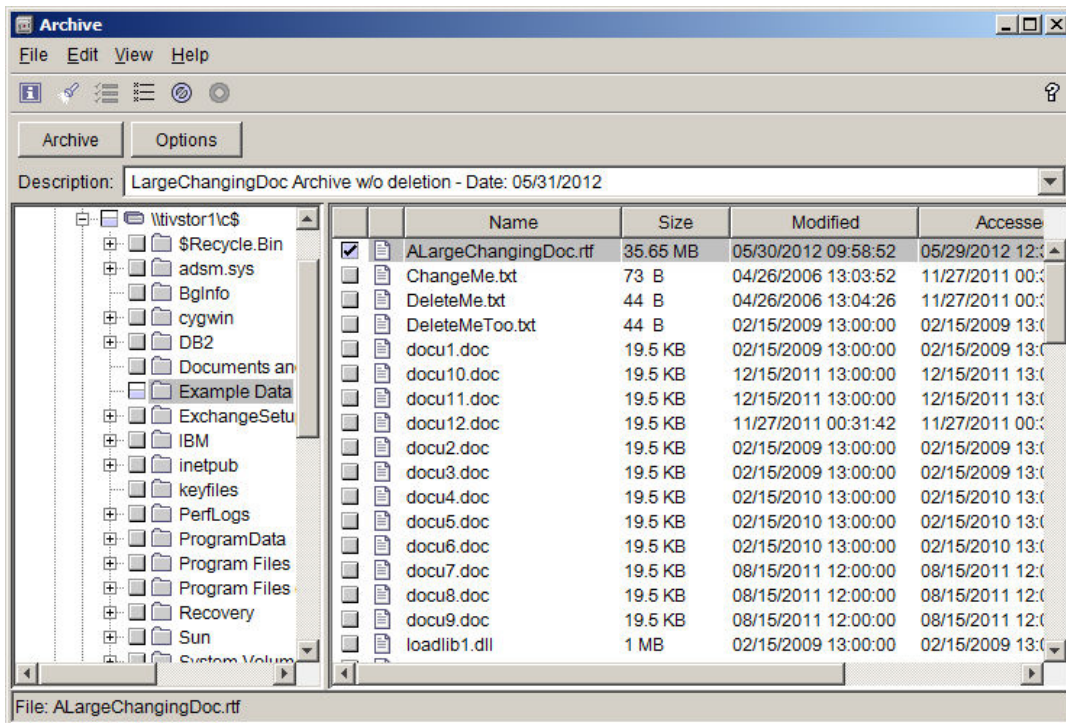
In order for the archive to be eligible for Archive Option licensing, the archive must reside on a VTL or tape storage pool. The archive can go directly to one of those storage pools, or it can go first to disk and then migrate to one of them. In this example, we send the archive directly to the VTL storage pool that was set up in the Server Implementation example (Appendix A).
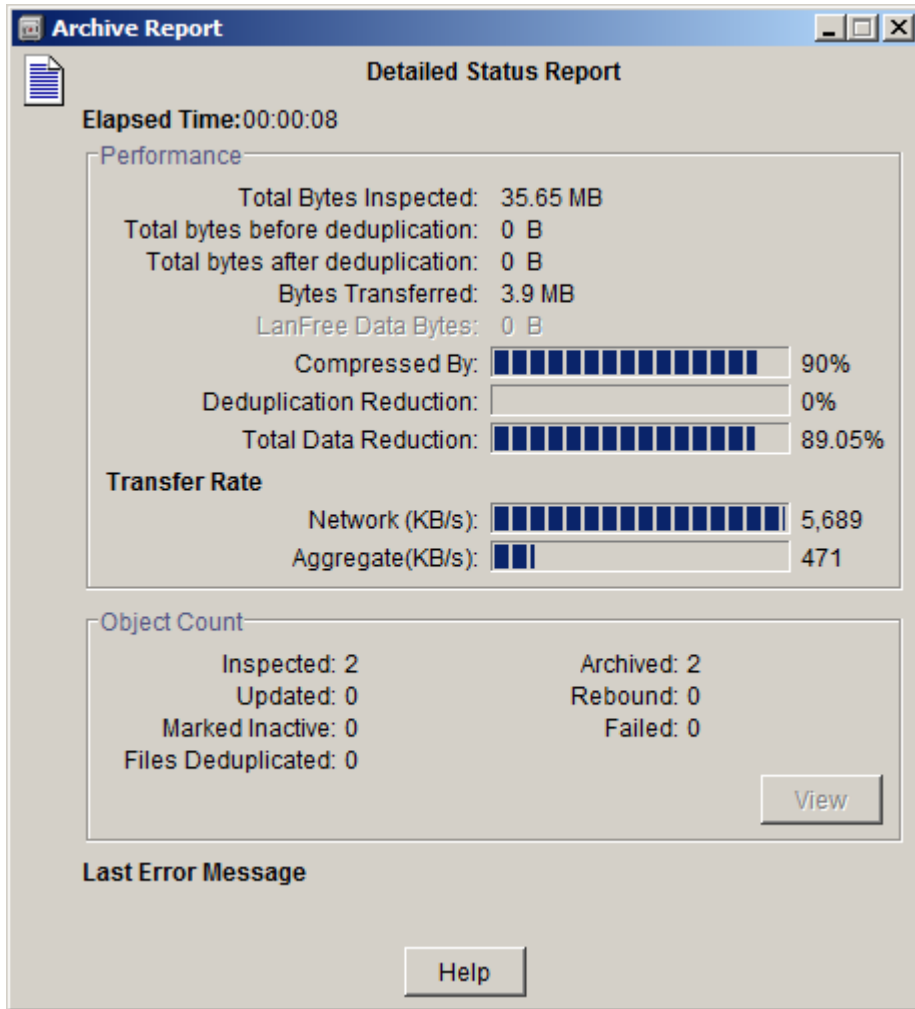
Getting the archive to go directly to the correct storage pool requires associating (also called binding) the archive to a management class defined with the archive destination set to one of those pools. From our Server Implementation example, that management class is VTL-ARCHIVE-MC. If the appropriate management class is not the default and not automatically associated with the files being archived (through the Include/Exclude list), then you will need to select the management class. Use the Options button, which takes you to "Override Archive Options".

In the screenshot below you can see we checked "Override include/exclude list"; then we selected the desired management class from the pull-down menu. The other options are the defaults.
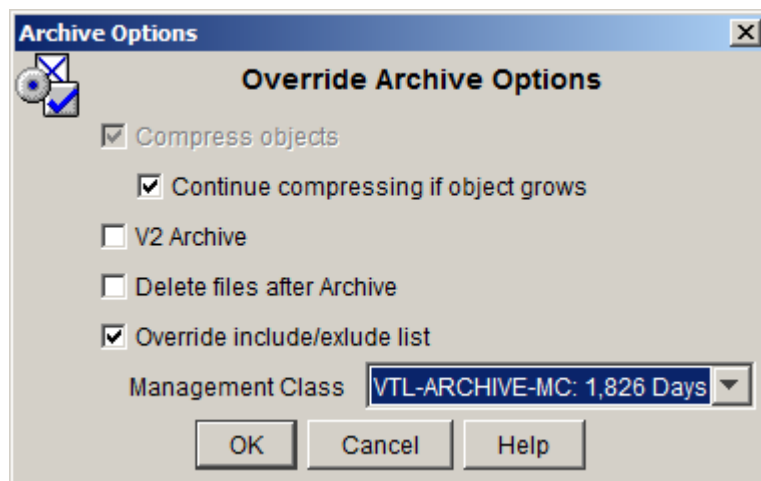

We click OK and then back on the Archive screen, click the Archive button.

The screenshot below shows a successful archive.

Archive Report

Detailed Status Report

Elapsed Time: 00:00:08

Performance

Total Bytes Inspected:	35.65 MB
Total bytes before deduplication:	0 B
Total bytes after deduplication:	0 B
Bytes Transferred:	3.9 MB
LanFree Data Bytes:	0 B
Compressed By:	90%
Deduplication Reduction:	0%
Total Data Reduction:	89.05%

Transfer Rate

Network (KB/s):	5,689
Aggregate(KB/s):	471

Object Count

Inspected: 2	Archived: 2
Updated: 0	Rebound: 0
Marked Inactive: 0	Failed: 0
Files Deduplicated: 0	

View

Last Error Message

Help

If you want to retrieve a file from this archive, you choose Retrieve from the backup-archive client GUI and then navigate to the file by drilling down into the archive, as shown below.
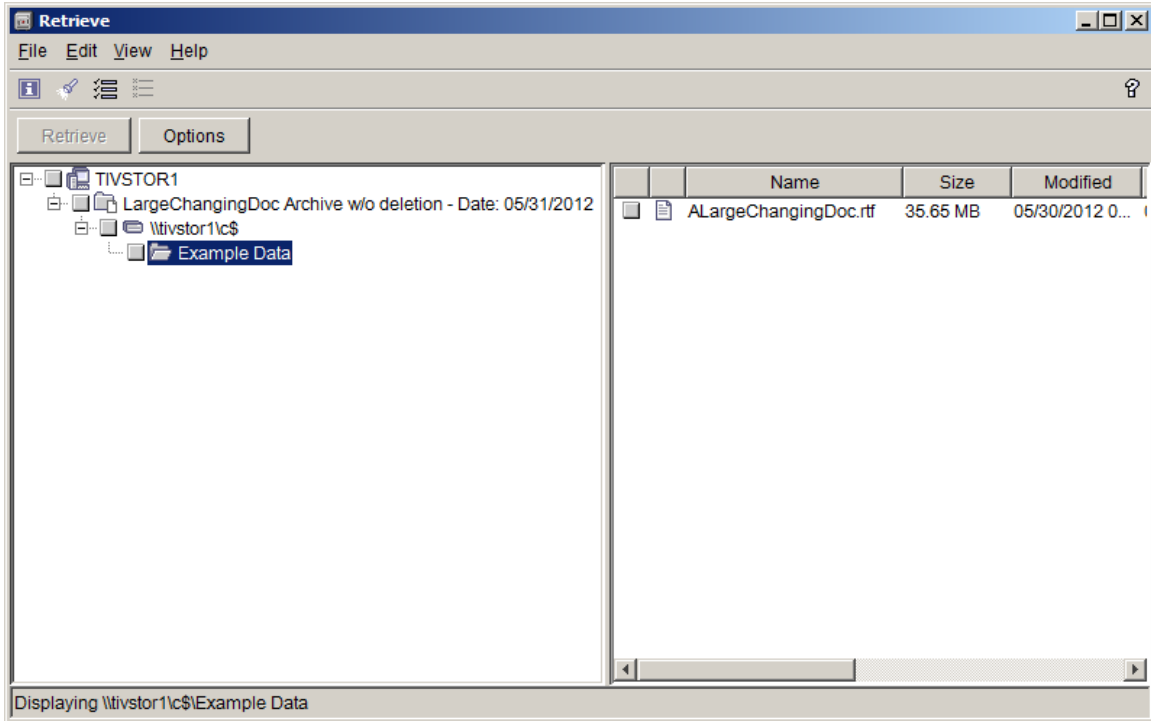

We now have an archive of the file, and it resides on a VTL; so the licensing cost for its capacity consumption is at the Archive Option rate. However, have we reduced our overall license cost? The answer is no. The reason is that we still have all the backup copies of the file, as shown below. So our license cost for backup (that is, regular SUR) capacity consumed by this file has not decreased, and we now also have Archive Option capacity used by the archive of the file. How do we reduce the overall capacity license cost? Answer begins on the next page.
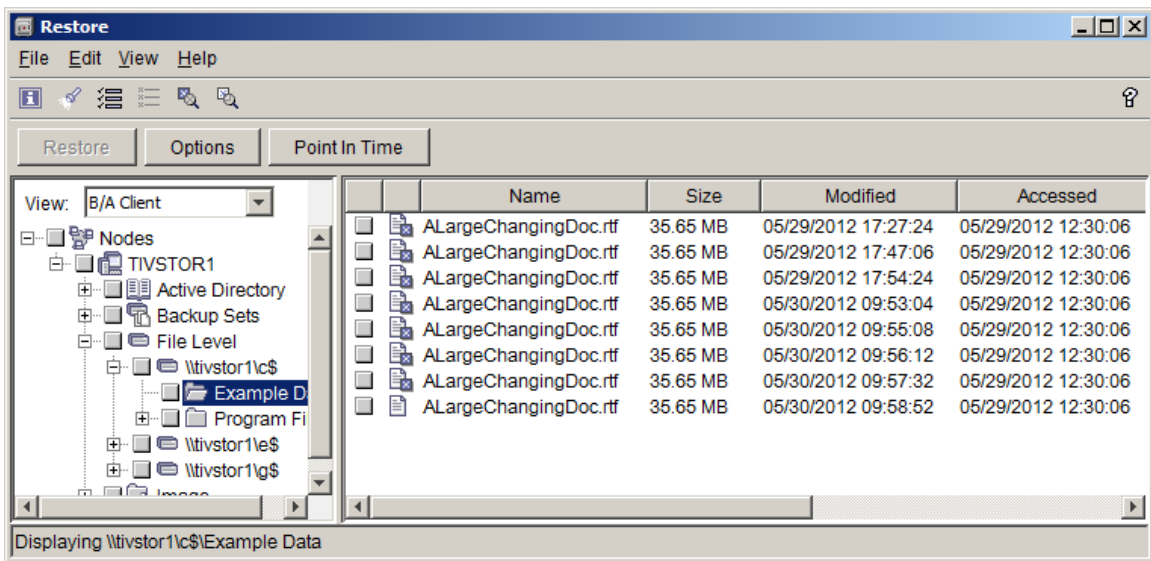
The way to reduce the overall capacity license cost is to eliminate the unnecessary backup copies. This can be accomplished by deleting the file from the file server or workstation. Of course, this is consistent with the earlier decision to archive the file (because it is no longer actively used). After the file is deleted, the TSM policy controls for managing retention of backups of deleted files takes over. In the case of our example file, the policy says to keep one copy of the last revision of the deleted file instead of 10 copies for an existing file (VERDELETED vs VEREXISTS).

Attention: Before deleting a file, and thereby losing some older backup versions, consider the possibility of whether there are bits of information in earlier versions of the file that are no longer in the most recent version. This could be important if the file has some legal or corporate governance value, for example, the information in the file might be needed for regulatory compliance. If this is the case, the earlier version or versions of the file should be restored (possibly to a temporary area) and archived along with the most recent version, all before the file is deleted.

The file can be deleted manually after a successful archive, or we can have the TSM backup-archive client automatically delete the file as part of the archive operation, after the archive completes successfully. We will use the automatic method. First, here is the screen for archiving, with a slightly different Description.
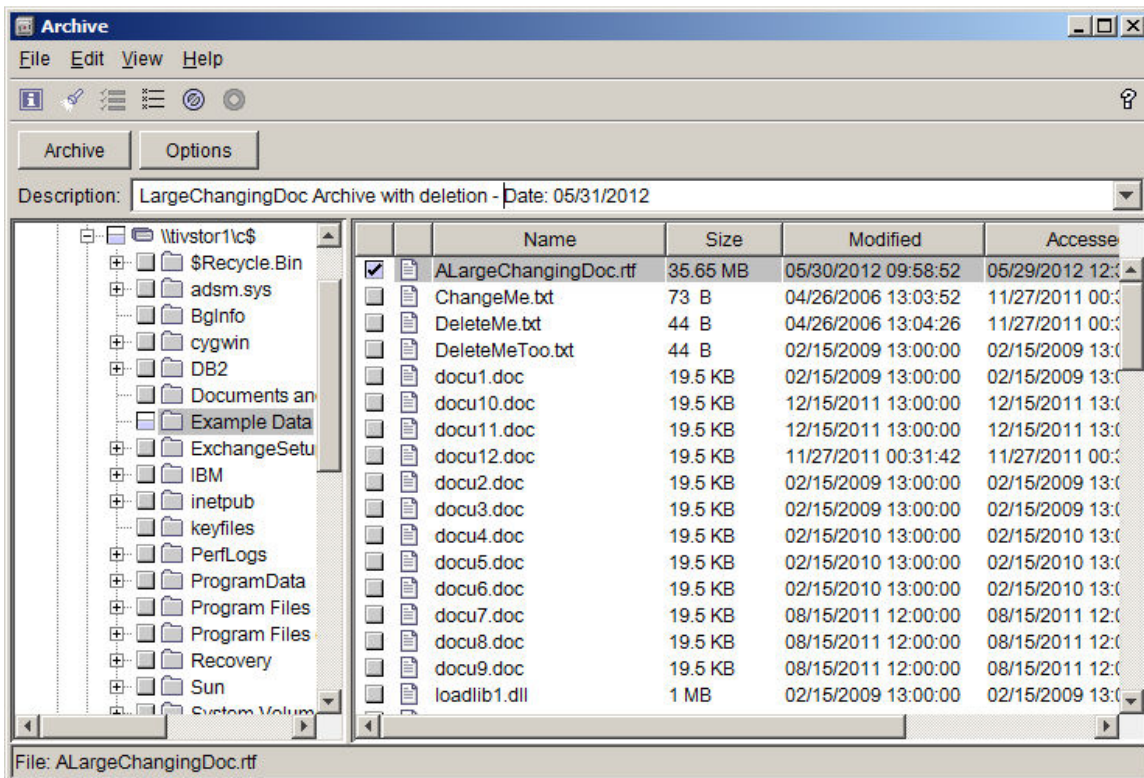
Next we click on the Options button again. This time we put a check mark in the "Delete files after Archive" box, along with overriding the Management Class, again selecting the VTL Management Class.
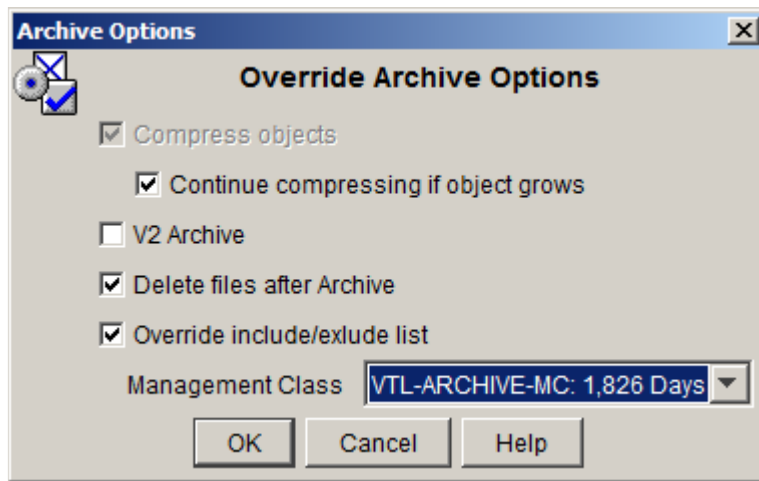
We click OK and then back on the Archive screen, click the Archive button.

After the archive operation successfully completes, we check to see if the file has been deleted. As shown in the Windows Explorer screenshot below, it has. But what has happened to all of the backup copies we had before? If you check after the archive, you will see they still exist. How do we get rid of them? Answer below...
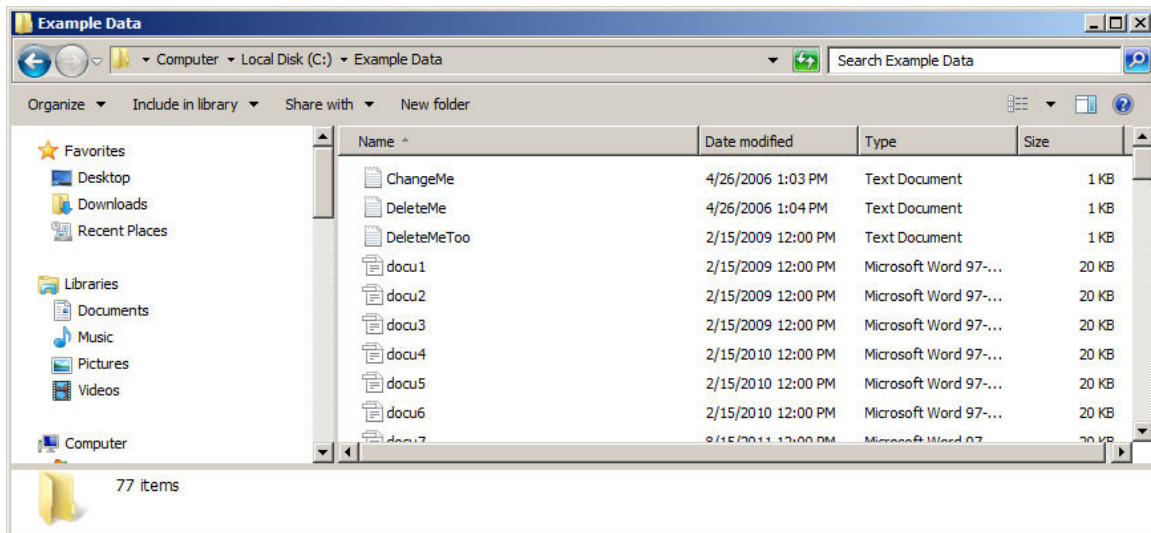
Two more steps have to occur before the extra backup copies go away completely and are no longer consuming regular licensing capacity:

1) We must run an incremental backup of at least the directory or folder in which the file existed. Usually, the regular incremental backup of the client will take care of this, eliminating the need to run some manual or extra incremental backup. The incremental backup is required so that TSM recognizes that the file no longer exists. After this, the extra versions of the backup are no longer available, as shown in the Restore window screenshot below. We just have the one version kept because of the VERDELETED setting.
   (The remaining backup copy of the deleted file will remain for the number of days specified by the RETONLY parameter of the Management Class's Backup copy group.)
2) The TSM server database Expire Inventory process must run to eliminate references to those backup copies. Normally, Expire Inventory runs as part of daily TSM server housekeeping (for example, one of the administrative schedules), or it can be run manually.

At this point, we are successfully exploiting the Archive Option. The file is archived on VTL, consuming capacity at the Archive Option rate, and instead of eight copies we now have one backup copy (and some number of days later, zero copies) consuming capacity at the regular SUR capacity rate.
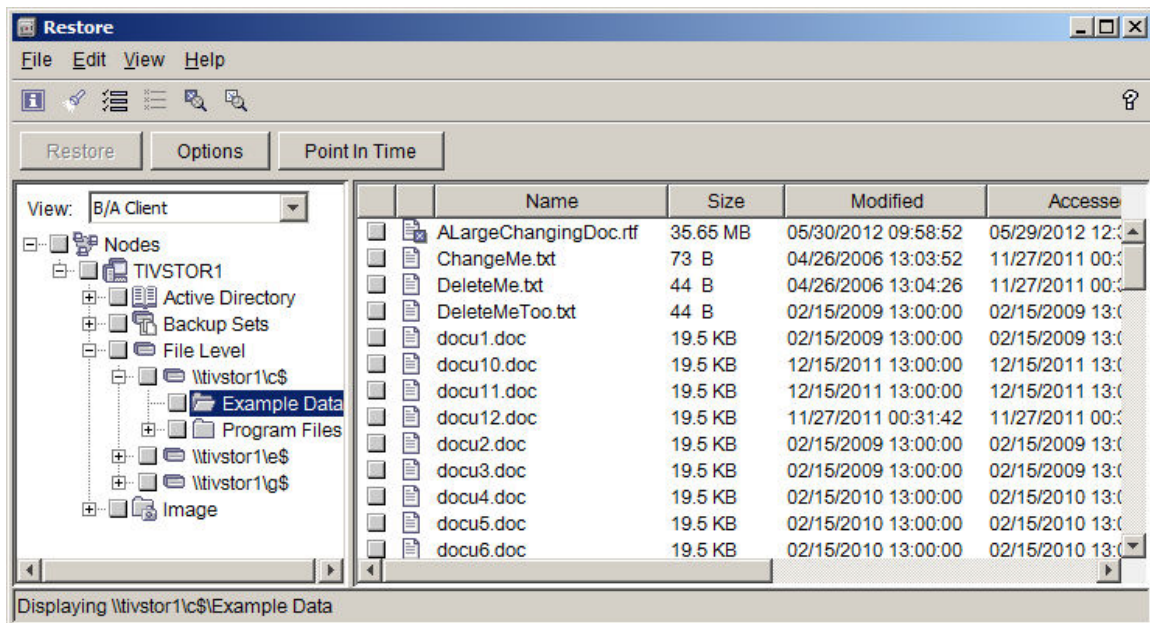
Additional Information:

Archive and retrieve data with backup-archive clients:
http://publib.boulder.ibm.com/infocenter/tsminfo/v6r3/topic/com.ibm.itsm.client.doc/t_arc_ret.html