Active File Archiving on Windows 2012 R2

Isaac Barton - Assistant Director of IT for Enterprise Infrastructure Services
& **Cliff Giddens** - Systems Support Associate - Enterprise Infrastructure Services

Business Need – *What is the problem we are solving?*

-constantly having to buy more storage & that's expensive

-the need to have archived data accessible-

From Wikipedia, the free encyclopedia

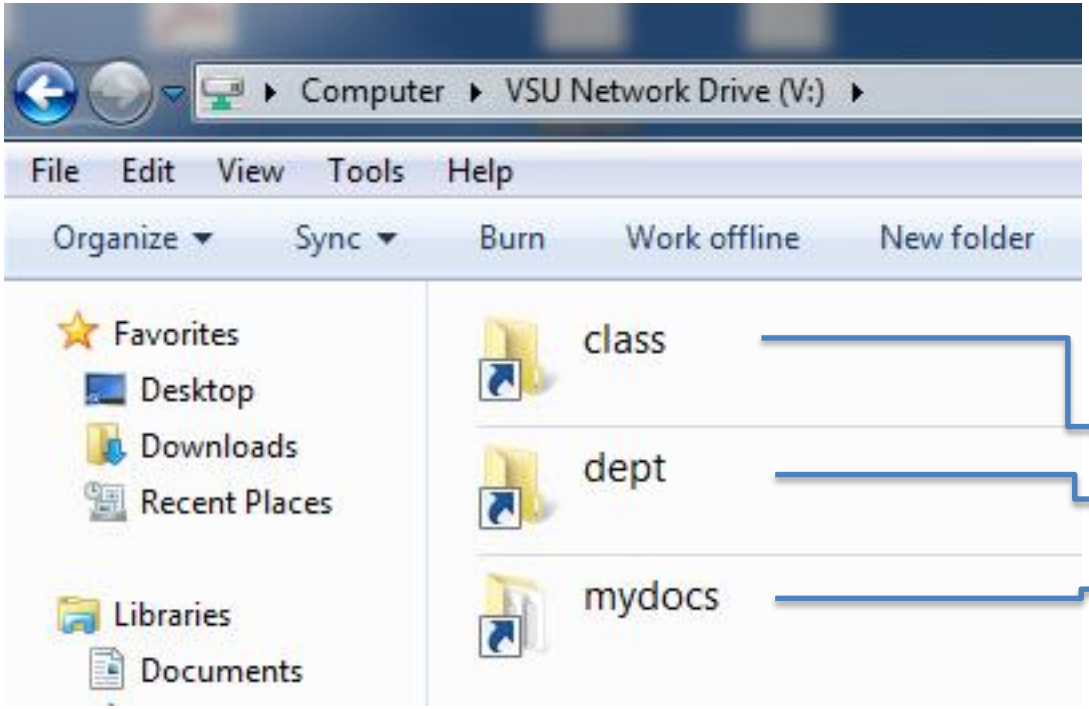
Active Archive

Active Archive is a method of tiered storage which gives the user access to data across a virtualized file system that migrates data between multiple storage systems and media types including solid-state drive/flash, hard disk drives, magnetic tape, optical disk, and cloud. The result of an active archive implementation is that data can be stored on the most appropriate media type for the given retention and restoration requirements of that data.[1] This allows less time sensitive or infrequently accessed data to be stored on less expensive media, and eliminates the need for an administrator to manually migrate data between storage systems. Additionally since storage systems such as tape libraries have very low power consumption, the operational expense of storing data in an active archive is greatly reduced.

[2] Active archives provide organizations with a persistent view of the data in their archives and make it easy to access files whenever needed. Active archives take advantage of metadata in order to keep track of where primary, secondary, and sometimes tertiary copies of data reside within the system, in order to maintain online to near-online accessibility to any given file in a file system, regardless of the storage medium being utilized.

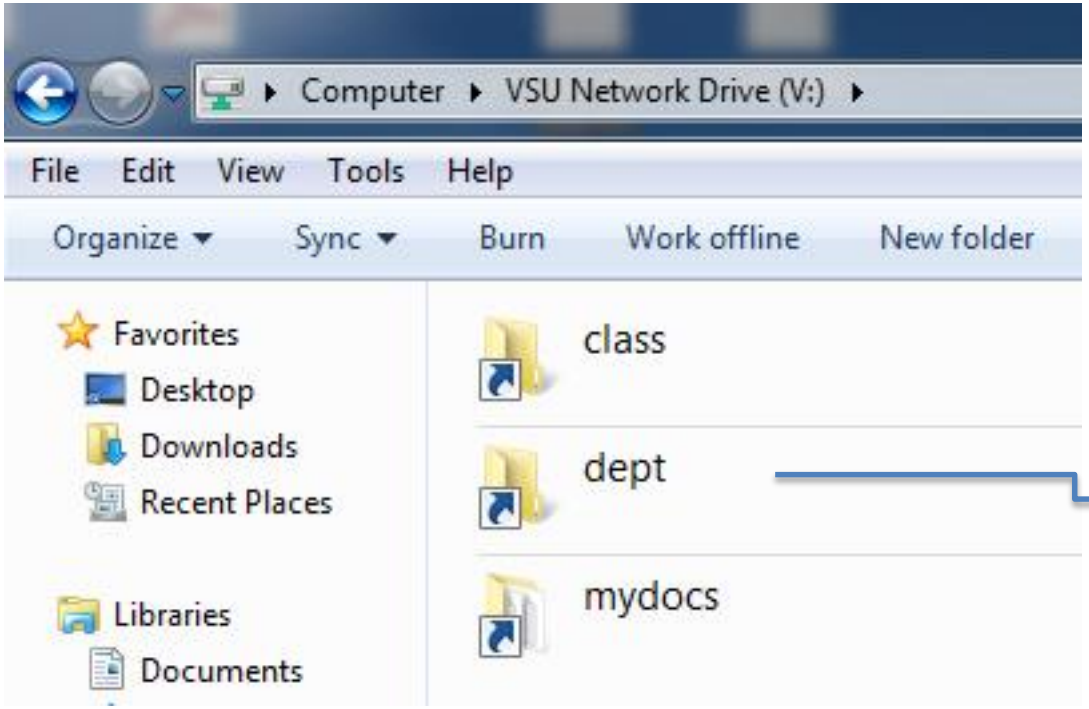
Active Archiving at VSU

Our “V: Drive”



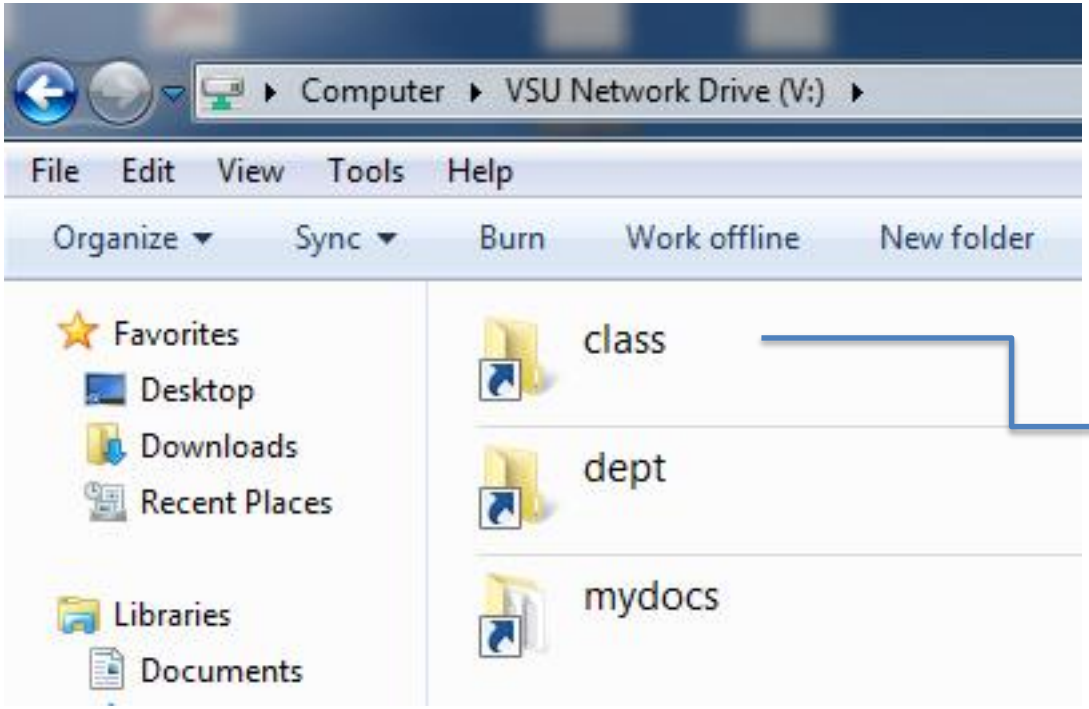
Look like folders, but are DFS links to virtual servers that provide various terabytes of storage

Active Archiving at VSU



FOUND:
10+ year old files – likely never to be touched again.

Active Archiving at VSU

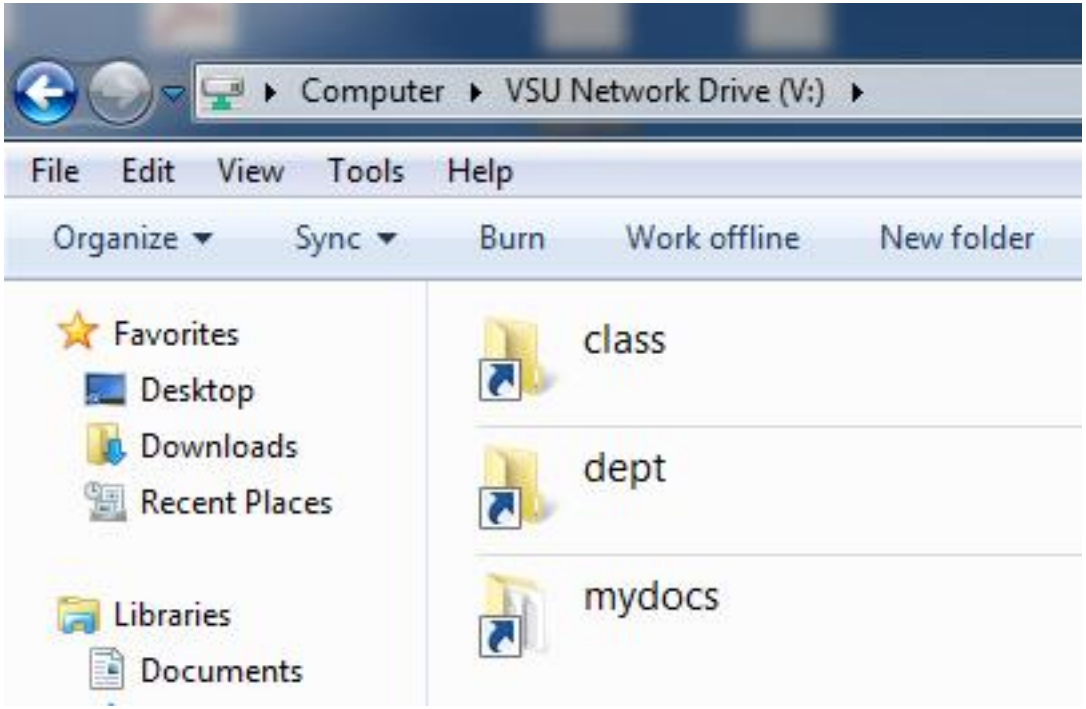


100's of new class folders **each** semester!

Current Class Example:
ACCT3000-FALL2015

Old Class Example:
ACCT3000-FALL2010

Active Archiving at VSU



This folder will be reduced to less than 1MB (on disk) from 100's of megabytes!

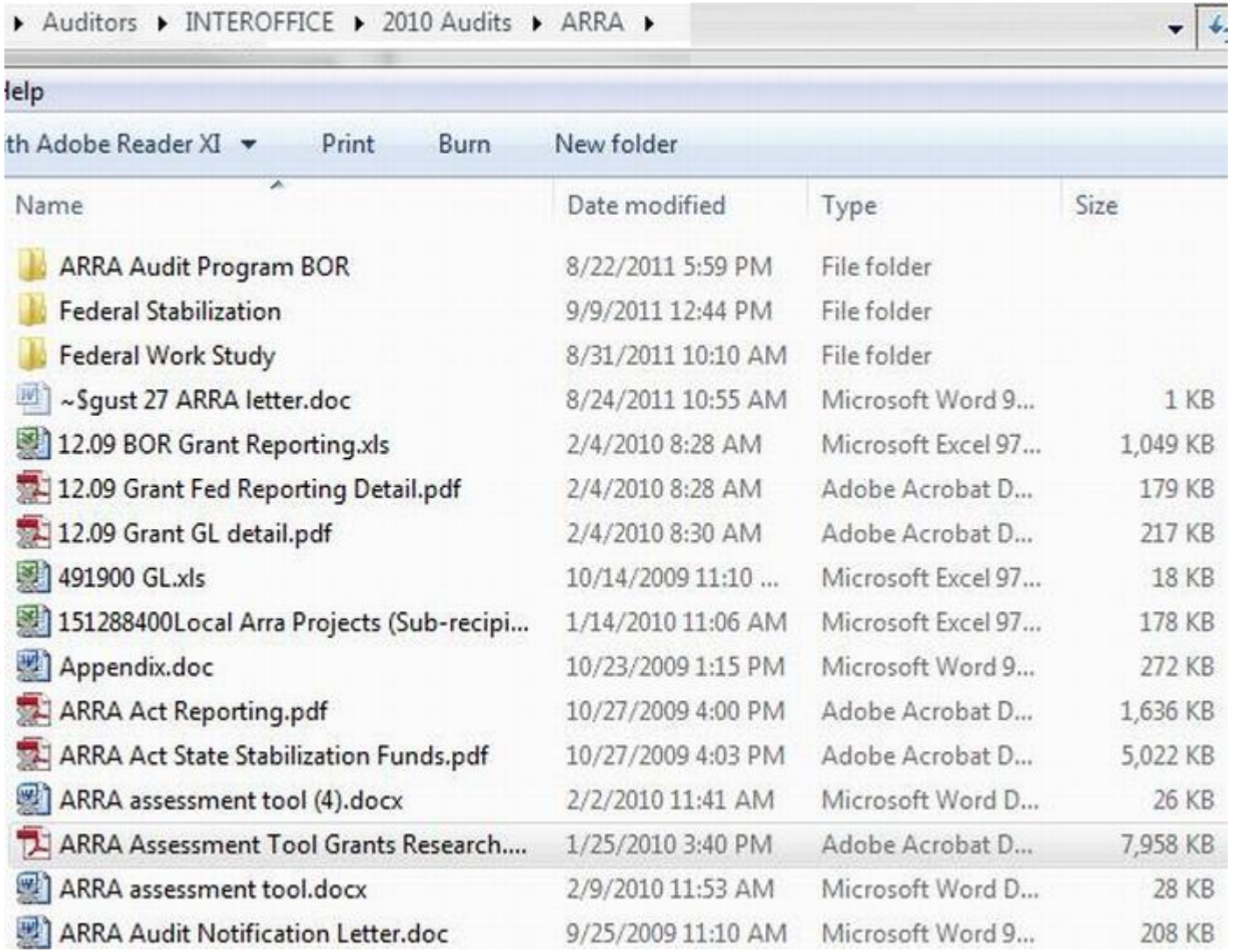
ACCT3000-FALL2010

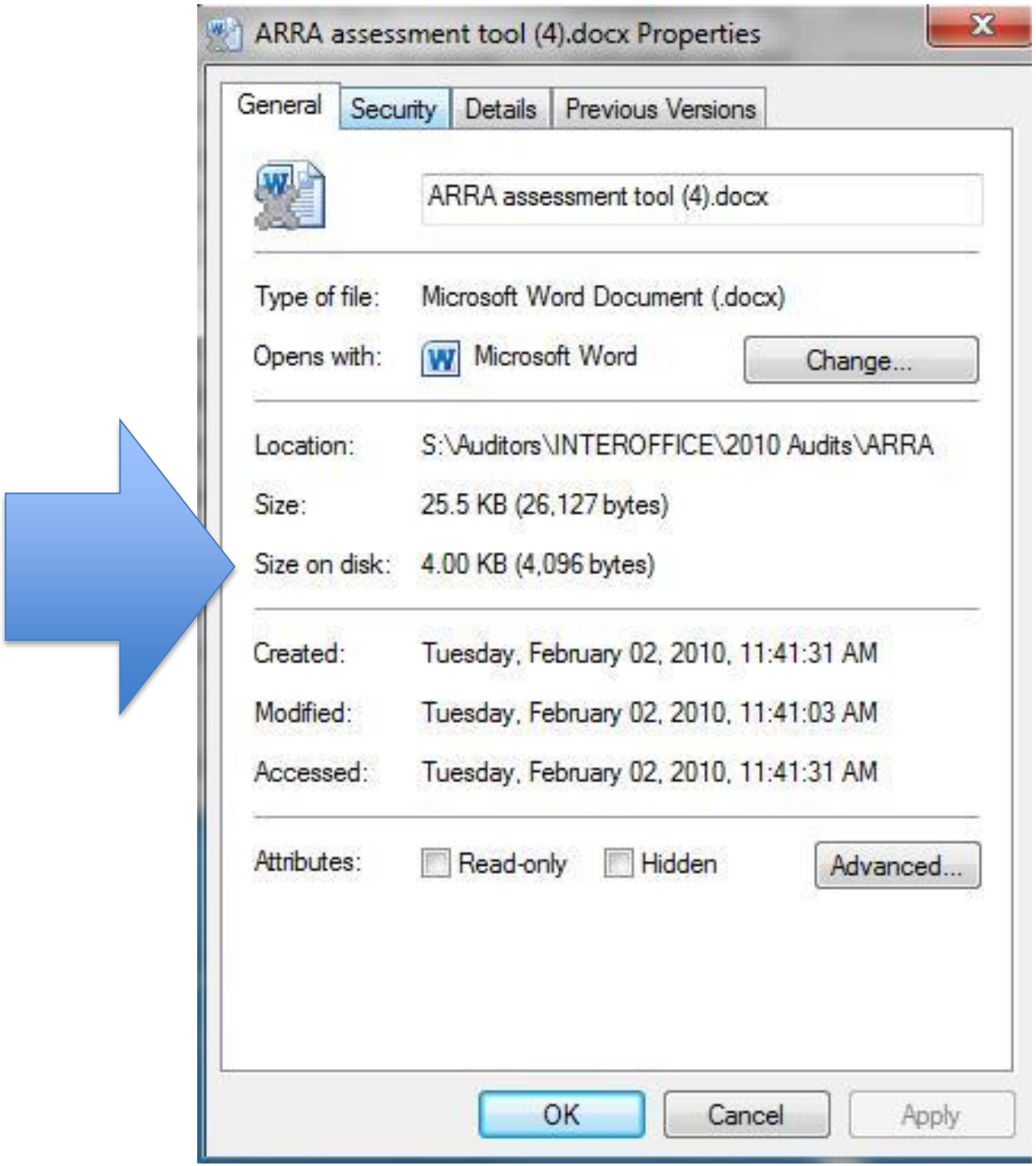
Active Archiving at VSU

To the end users all of the files still appear to be in place, while older files are likely **only** on a disk cache or in a tape library!

*archived indicator (grey X)

*Recalled file



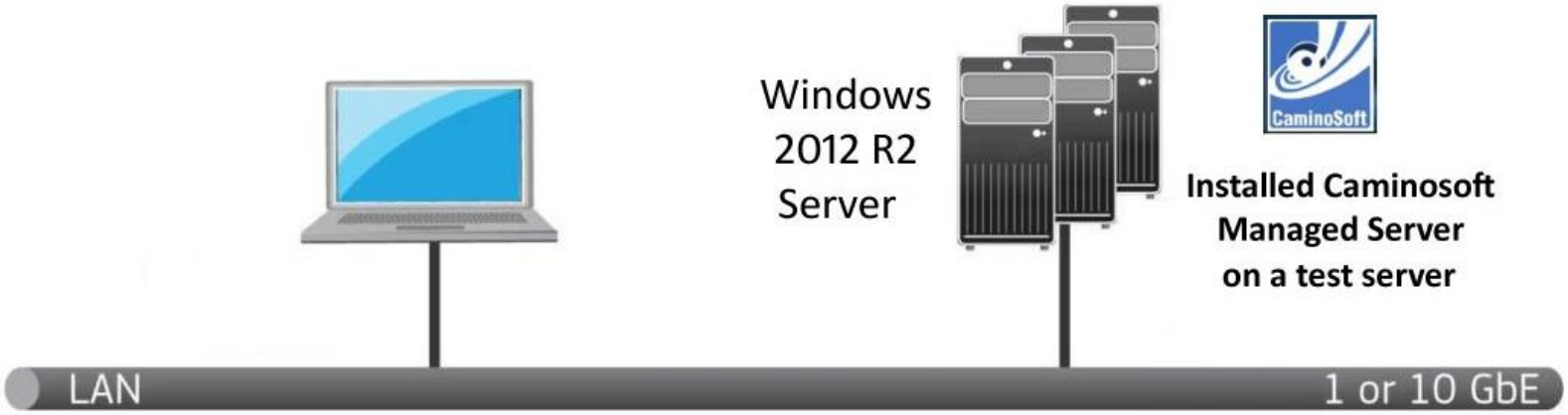

ARRA assessment tool (4).docx Properties
General
Security
Details
Previous Versions
ARRA assessment tool (4).docx
Type of file: Microsoft Word Document (.docx)
Opens with: Microsoft Word
Change...
Location: S:\Auditors\INTEROFFICE\2010 Audits\ARRA
Size: 25.5 KB (26,127 bytes)
Size on disk: 4.00 KB (4,096 bytes)
Created: Tuesday, February 02, 2010, 11:41:31 AM
Modified: Tuesday, February 02, 2010, 11:41:03 AM
Accessed: Tuesday, February 02, 2010, 11:41:31 AM
Attributes: Read-only Hidden
Advanced...
OK
Cancel
Apply

How did VSU implement this?

Installed on each server

Connected to servers via CIFS shares

LTO6 tape library connected to Strongbox



We installed Caminosoft Managed Server on a test server (using a mirror of our real data.)
We then went through an exhaustive testing process in a Proof of Concept (POC.)
Made sure of the following:

- that it was actually saving disk space
- that files would be recalled automatically upon a double click of the stubs
- that the performance was acceptable
- that backing up stubs will not recall the actual files
- that restored stubs still worked correctly

STRONGBOX

LTFS NAS Archive

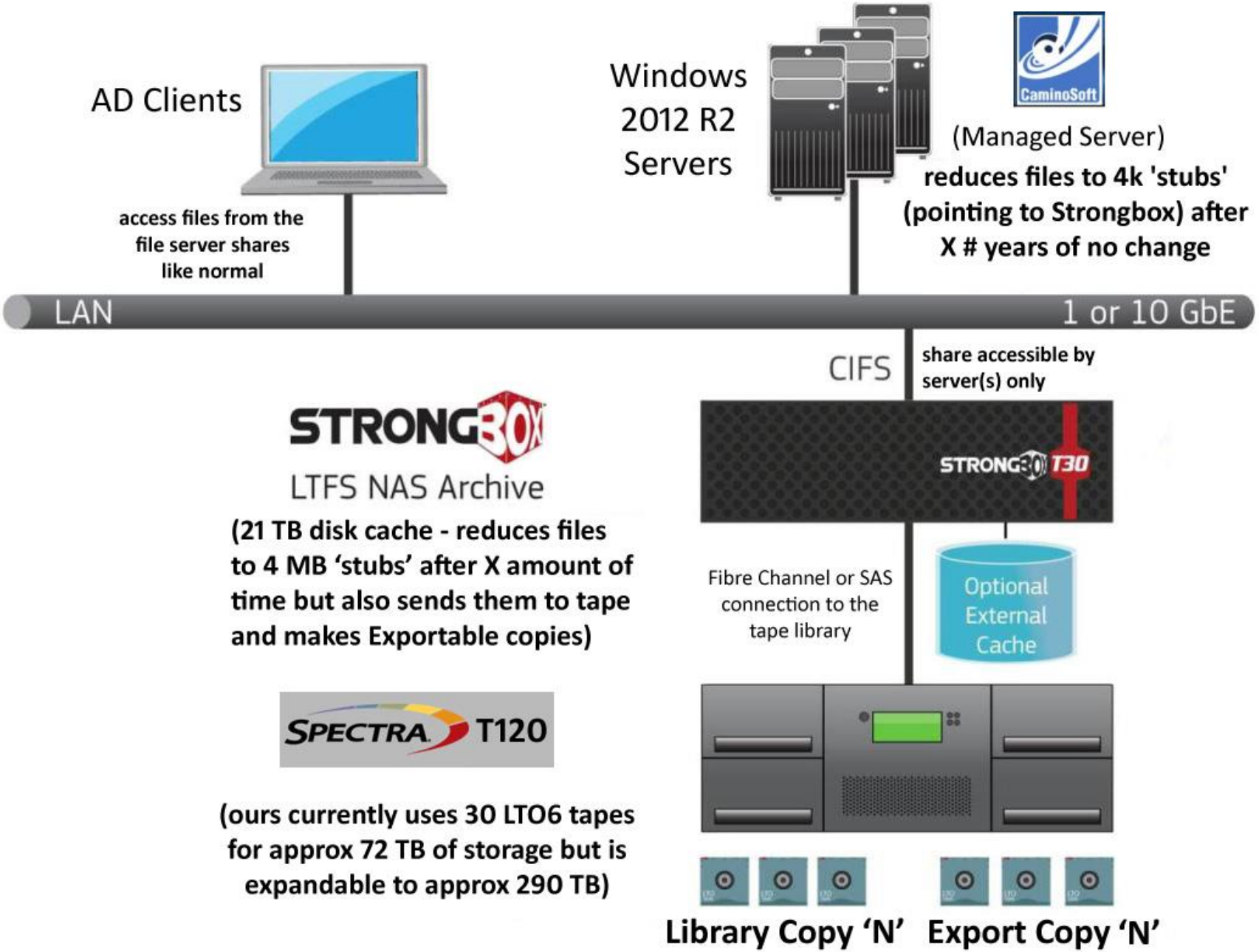
- LTFS network attached storage
- 21 TB disk cache
- 3 rack units
- Linux backend
- Web GUI for configuration
- Added to Active Directory as a server
- File transfers enabled through CIFS (shares)
- Archived files are instantly written to tape

STRONGBOX
LTFS NAS Archive

Eventually the files on disk cache get truncated - kind of like Caminosoft does (to save space) EXCEPT the truncated file size is 4MB by default.

The reason they keep larger “stubs” is so that when a client recalls a file - Strongbox sends a 64K “chunk” every 14 seconds (by default) to keep the client “happy” so it won’t timeout. This gives Strongbox time to retrieve the rest of the file from the tape library - put it back on the cache and send the remainder.

*Note: Files smaller than 4MB remain “un-truncated” - so if they are recalled – they are instantly sent back to the server from disk.


AD Clients
access files from the file server shares like normal
Windows 2012 R2 Servers
CaminoSoft
(Managed Server)
reduces files to 4k 'stubs' (pointing to Strongbox) after X # years of no change
LAN
1 or 10 GbE
CIFS
share accessible by server(s) only
STRONGBOX
LTFS NAS Archive
(21 TB disk cache - reduces files to 4 MB 'stubs' after X amount of time but also sends them to tape and makes Exportable copies)
STRONGBOX T30
Fibre Channel or SAS connection to the tape library
Optional External Cache
SPECTRA T120
(ours currently uses 30 LTO6 tapes for approx 72 TB of storage but is expandable to approx 290 TB)
Library Copy 'N'
Export Copy 'N'

Summary of Savings

(archiving 2+ year old files)

Class	6 TB (initially 54% used – now 50%)
Dept	13 TB (initially 89% used – now 57%)
Libdata	7 TB (initially 73% used – now 41%)
Mydocs	16 TB (initially 95% used – now 76%)

Total: 42 TB

Archived: 19 TB (since February 2015)