



Custom File Assessment

Prepared by

Jon Sims

Data Assessment for

Jerry Makamor
Acme Buildings

Date of Data Collection: 5/23/2011

Total Files Scanned: 488,304

Shares Scanned: \\streaming\streaming\$; \\streaming\streaming2\$; \\streaming\library\$; \\dl-lynx\dlat\$; \\spprodfs\users

Your potential savings \$313,854.86



Table of Contents

- Return on Investment (ROI) Analysis 3
 - Assumptions used in analyzing data calculated using industry averages..... 3
 - Direct Attached Storage..... 3
 - SAN FC Storage Disk 4
 - SAN SATA Storage Disk..... 4
 - Summary of True Business Cost (Includes Physical Capacity, Power, Cooling, etc.)..... 5
 - Summary 5
- Breakdown by Amount of Files 6
- File Classification..... 7
- File by Last Accessed 8
- File by Last Modified 9
- File by Last Accessed File Count..... 10
- File by Last Modified File Count..... 11
- File by Category..... 12
- Quantity By File Size 13
- Number of Files 14
- Final Assessment Summary 15
- Appendix..... 16



Return on Investment (ROI) Analysis

Managed or not, data will continue to grow. That is the one constant in the industry. The first step in Data Management is to first identify the "personality" of files on your network.

Today, storage growth is a concern for most companies. Companies are looking to maximize the storage they have and increase efficiency rather than just buying more physical disks. Deduplication technology is very mature today. Companies should only consider technology that provides primary storage deduplication. Technology offering primary data deduplication will also deduplicate tier 2 and backup data saving you from purchasing multiple solutions and further reduce data center cost.

Assumptions used in analyzing data calculated using industry averages

- You should have 3 copies of data to insure adequate protection against losses.
 - (Tape, Disk 2 Disk, Offsite, etc.)
- Direct Attached Storage Capacity can be over \$20 per gigabyte as a raw capacity average.
The high cost of DAS is primarily attributed to the fact you must over provision on every host which is not very efficient.
- SAN | NAS RAW capacity averages \$8.33 per gigabyte for SATA and up to \$27.49 per gigabyte for FC and SAS which reduces quickly as you scale out.

(Based on cost to store data, not only physical disk cost)

Direct Attached Storage

DAS is the most costly due to the fact it does not scale and cannot be re-allocated easily. Since you would allocate storage to a single host you have to over provision wasting storage dollars with no real flexibility. DAS also limits the ability to scale in a virtualization environment along with preventing high availability advantages found in shared storage.

TeraBytes	Cost per GB per Year
8	\$9.85
16	\$9.85
32	\$9.85
64	\$9.85
128	\$9.85

SAN FC Storage Disk

SAN storage is the most cost effective with lower cost the further you scale capacity. Fiber Channel or SAS Disk should be used anywhere you have high performance applications to insure adequate IO capability. Otherwise you will severely limit the number of users you can support and limit the ability to scale. Around 32 TB you become very close to the cost of DAS storage however you gain the ability to scale and grow storage without interruption to users. Allocating what you need to better control cost per application will dramatically increase efficiency of storage allocation and reduce over provisioning.

TeraBytes	Cost per GB per Year
8	\$27.49
16	\$18.51
32	\$11.27
64	\$6.65
128	\$3.83

SAN SATA Storage Disk

SATA is being adopted more for mass storage and suitable for around 60% of low IO applications. In a RAID set that provides at least two drives for Parity (RAID10 or RAID6) you can reduce the risk of a double disk failure and have solid reliability. This is important with high density drives where two drive failures may not as common but a second drive failure during rebuild is more common.

Note: RAID6 will use approximately ~48% fewer disk spindles than RAID10 for the same protection to further reduce storage cost, power and cooling. RAID5 does not provide adequate protection when using denser spindle sizes and will suffer performance during expansion to redistribute parity.

TeraBytes	Cost per GB per Year
8	\$8.33
16	\$6.33
32	\$4.36
64	\$3.13
128	\$2.55

Summary of True Business Cost (Includes Physical Capacity, Power, Cooling, etc..)

The cost of storage is more than the cost of a physical drive, the chart below represents the True Business cost of storage to deploy and maintain based on industry averages. Ease of management, support, performance reporting, and expansion should all be considered in any SAN environment.

	Cost per GB	Actual Business Cost
DAS	\$9.65	\$20.27
SAN FC / SAS	\$18.51	\$22.21
SAN SATA	\$6.33	\$7.60

Storage utilization rates for DAS will average 40%, in a SAN environment storage utilization rates will improve to 80% gaining a much more efficient use of your budget.

The following reports should be used as a baseline and updated after a migration or archiving project to continue to fine-tune the environment. This will allow you to maximize your storage investment and scale more efficiently for long-term growth.

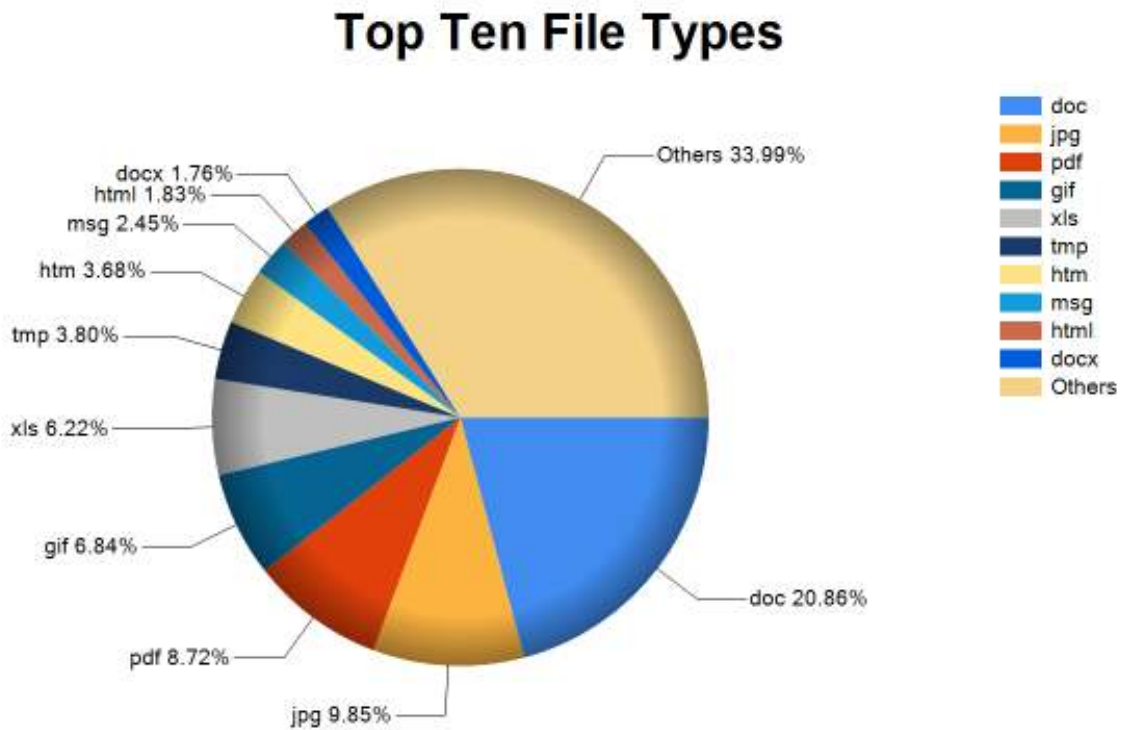
Summary

You have 50.64 terabytes of data on the share scanned at an estimated cost of \$28.00 for the storage capacity. File archiving would allow a reduction of cost by moving less used data to lower cost storage. Based on this assessment the file type "doc" was identified as a possible candidate for archiving or moving data to a lower cost tier 2 or tier 3 storage saving you an estimated \$313,854.86 in storage costs.



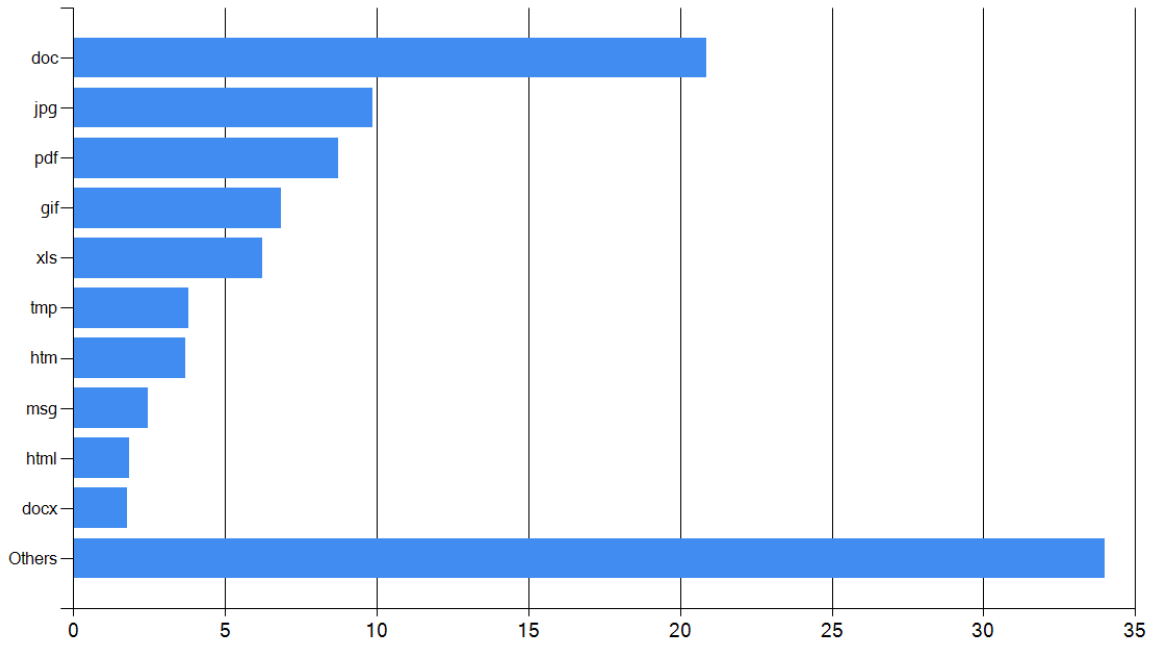
Breakdown by Amount of Files

The majority of files types consuming space are jpg with 9.85%. This is followed by pdf with 8.72%. The files outside the top 10 are labeled as Others consuming 11.09 or 33.99%.



File Classification

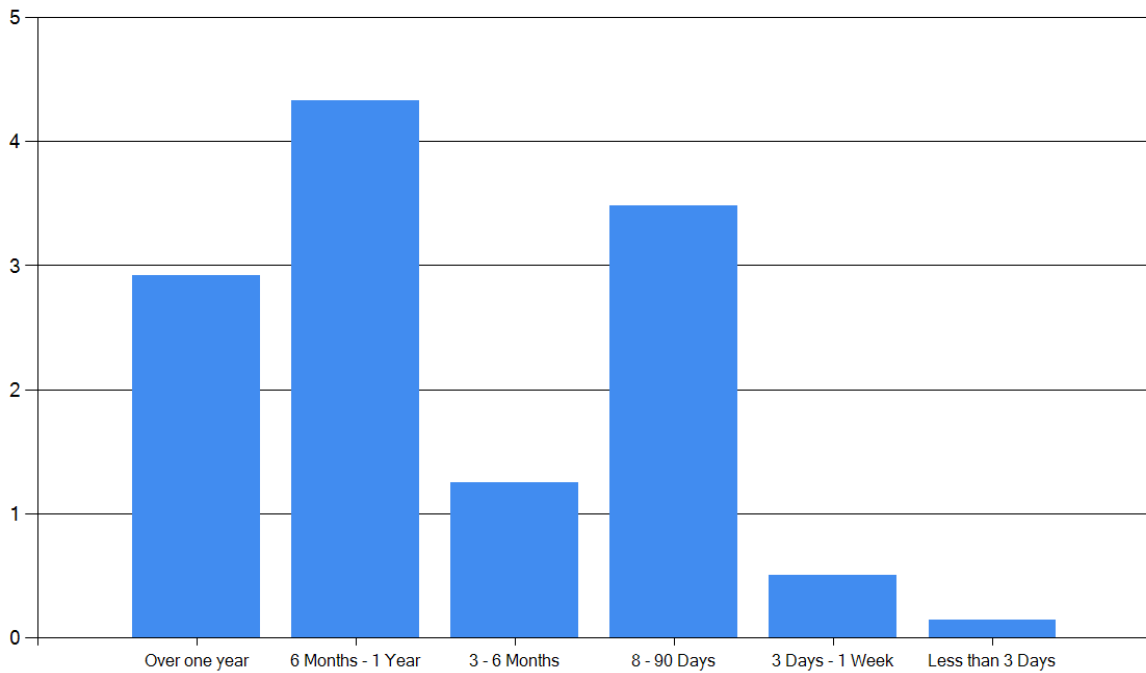
Amount of files By Percentage



File by Last Accessed

The report "File by Last Accessed" gives you an indication of the frequency data on this share is being accessed. This would help estimate the amount of capacity to plan for on tier 2 storage that would be required, and how much tier 1 storage you can save by freeing up capacity on more expensive disk.

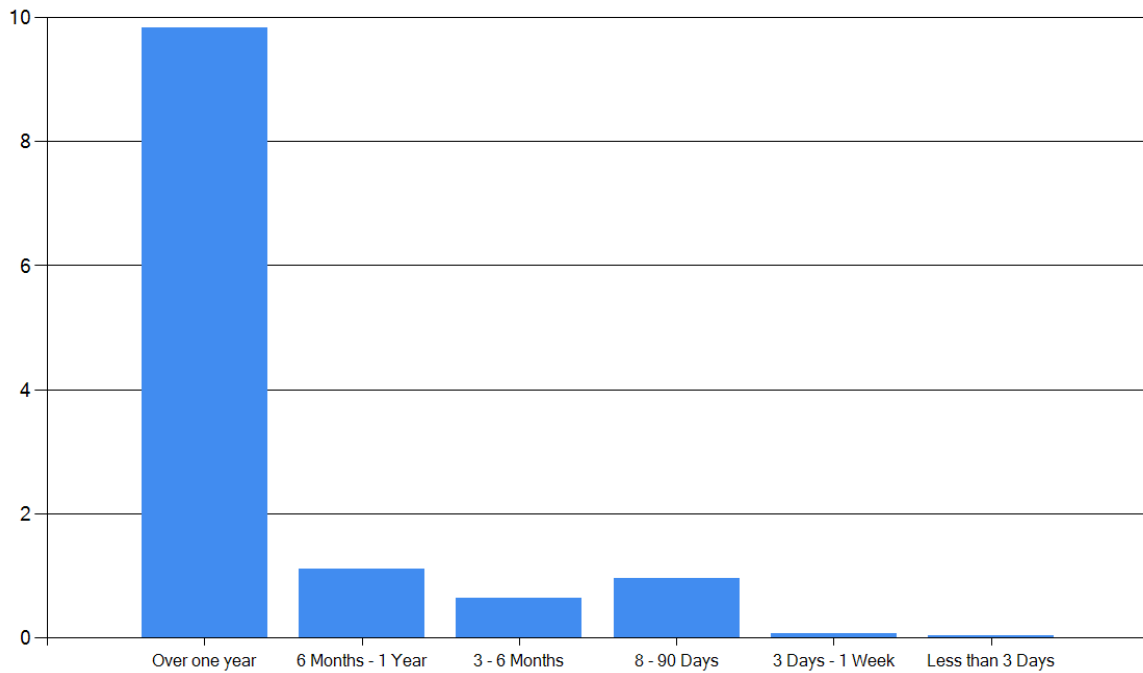
File By Last Accessed Total Size (TB)



File by Last Modified

If the value of untouched data is over 30% or greater than one year since the last modified date this would indicate the data may be important but is not changing on a frequent basis. This means write performance demands are very low and therefore this data would be ideal to put on a slower Tier 2 drive technology like SATA.

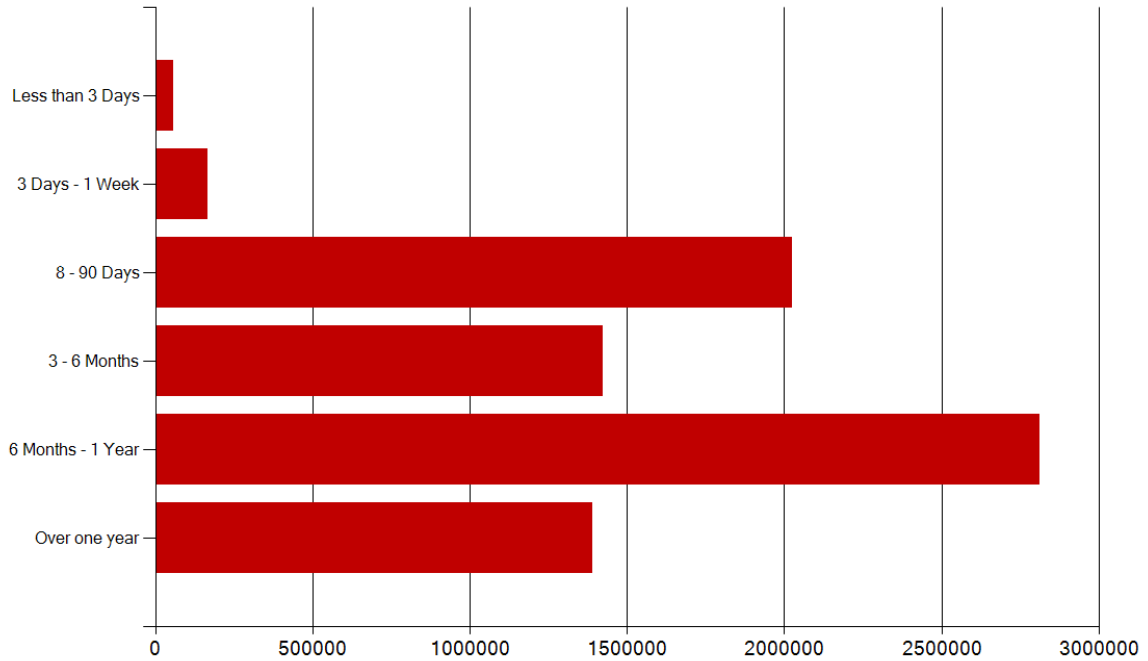
File By Last Modified Total Size (TB)



File by Last Accessed File Count

Files showing over one year since last access should be primary targets for archiving to an inexpensive off-site storage strategy, data compression or even tape. Files showing six months to one year are at a minimum great targets for a tier 2 SATA storage strategy.

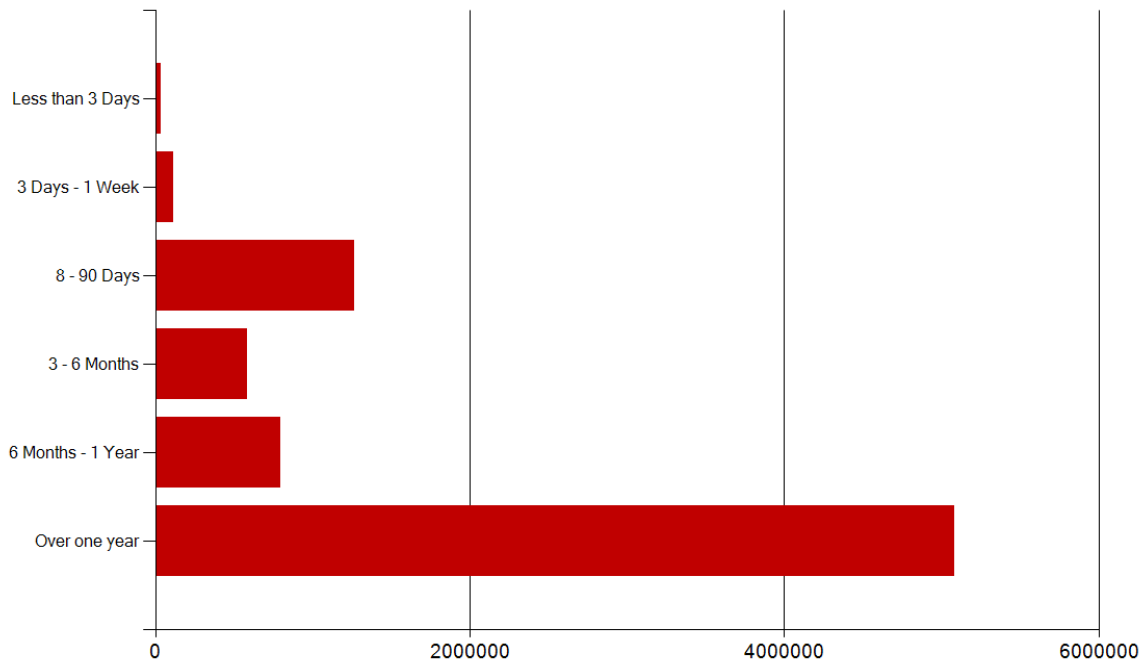
File By Last Accessed File Count



File by Last Modified File Count

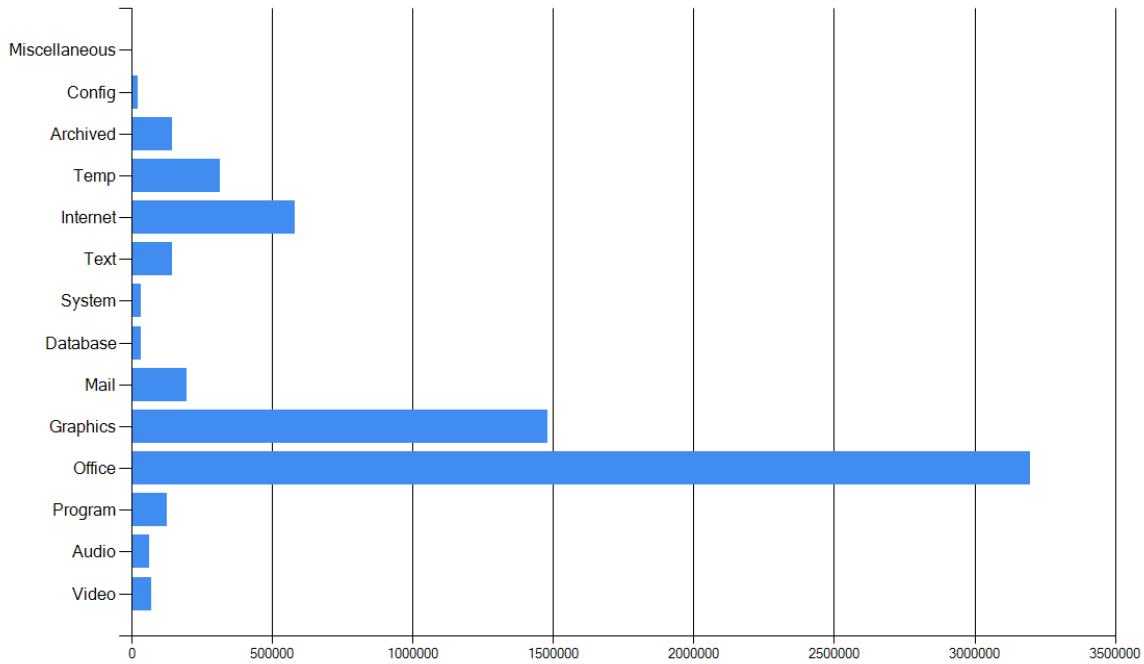
This data indicates active file usage to determine how active your users are with existing data. This can indicate an increase in write performance requirements with high results in the 3 Day to 1 Week category. Data within the one week or less category should remain in higher performance FC or SAS disk technologies.

File By Last Modified File Count



File by Category

Files by Category



Capacity growth can be hard to predict, this report indicates the historical growth over the past seven years or as far back as data is available to analyze. The largest growth occurred in 2006 with 28.41%. The lowest growth occurred in 2009 with 20.09%.

Storage Growth per Year					
Year	Total (TB)	Total Files	Files Added	File	Capacity
2005	1.37	2,415,862	2,415,862	0.00%	0.00%
2006	2.91	3,102,219	686,357	28.41%	112.36%
2007	5.20	3,935,346	833,127	26.86%	78.38%
2008	7.76	4,877,468	942,122	23.94%	49.14%
2009	10.14	5,857,112	979,644	20.09%	30.62%
2010	12.16	7,456,424	1,599,312	27.31%	19.95%
2011	12.63	7,875,523	419,099	5.62%	3.88%

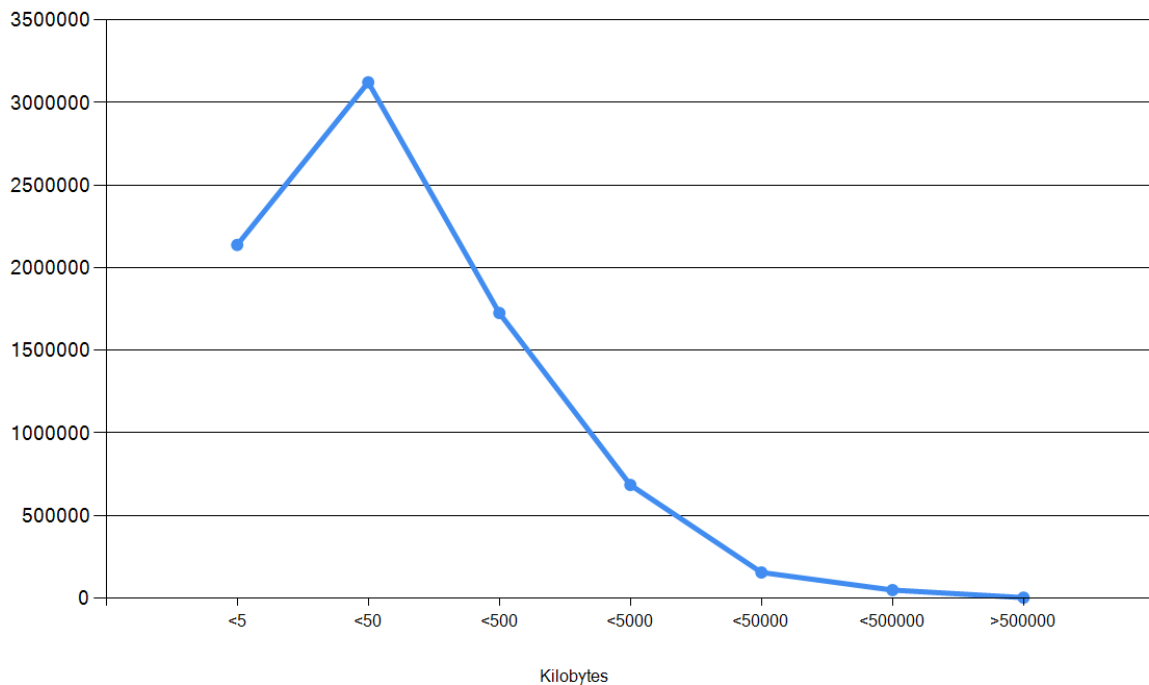
Average Yearly Capacity Growth = 49.06% Per Year

Storage Growth Consumption			
Share Scanned	GB Per Day	Days to 80%	Days to 100%
\\streaming\streaming\$	1.83	EXCEEDED	135
\\streaming\streaming2\$	0.01	215534	271199
\\streaming\library\$	0.02	131196	165079
\\dl-lynx\dlat\$	0.85	155	1883
\\spprodfs\users	5.14	7470	9573

Quantity By File Size

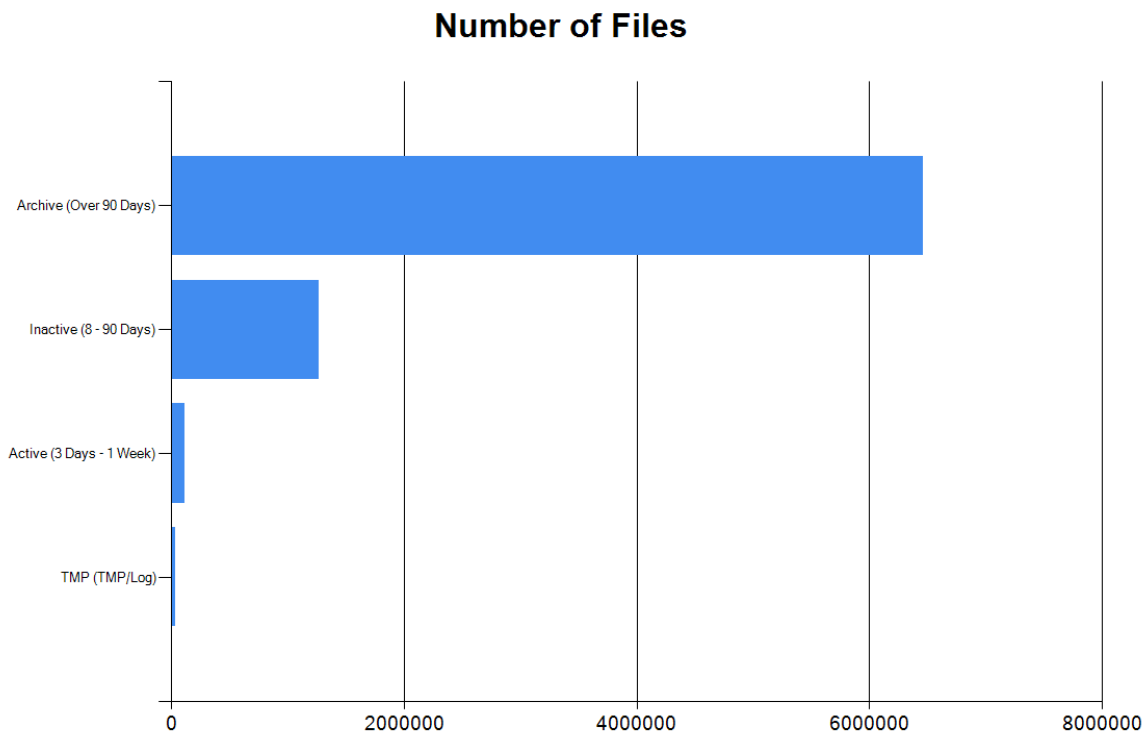
Performance of storage can be important if specific read / write profiles exist that should be compensated for. If there are primarily all small or large files this can change the recommended architecture. In the data example below we see the majority of files fall within the "<50" category which accounts for 3122437 files of the total files scanned. This may be important to the storage vendor during the pre-sales architecture or design phase.

Quantity By File Size



Number of Files

In this chart we can see the majority of data is falling into the "<50" category which is based on the number of files only. When you have a large number of files with a low change rate this could be a perfect use of Snap Shot technology. A retention schedule can be 30 - 90 days in some cases and not consume much space. Snap Shots are based on pointers so if the change rates are low you may consume little or zero additional capacity and increase file protection for improved efficiency.



Final Assessment Summary

File Classification				
File Classification	TMP TMP/Log	Active 3 Days - 1 Week	Inactive 8-90 Days	Archive Over 90 Days
Number of Files	36936	112984	1264719	6460884

Storage Growth Consumption				
Share Scanned	GB Per Day	Days to 80%	Days to 100%	
\\streaming\streaming\$	1.83	EXCEEDED	135	
\\streaming\streaming2\$	0.01	215534	271199	
\\streaming\library\$	0.02	131196	165079	
\\dl-lynx\dlat\$	0.85	155	1883	
\\spprodfs\users	5.14	7470	9573	

Combined Share Summary	
Capacity Saved By Eliminating Duplicates (TB)	2.99
Files removed by deploying single instance	3,276,092
Percent of content 1 year or older	77.86
Percent of content over 181 days suitable for archiving	8.79
Percent of duplicate data	23.70
Duplicate Reduction Savings	85,845.99
Archive Data Reduction Savings	313,854.86
Cost per GB	28.00
Shares Scanned	\\streaming\streaming\$; \\streaming\streaming2\$; \\streaming\library\$; \\dl-lynx\dlat\$; \\spprodfs\users
Total files on scanned share	488,304
Total size of scanned share (TB)	50.64

Appendix

Breakdown by Amount of Files			
File Type	File Percentage	Files Count	File Percentage
doc	20.86	1,620,800	0.22
jpg	9.85	765,205	0.60
pdf	8.72	677,933	0.40
gif	6.84	531,345	0.00
xls	6.22	483,599	0.10
tmp	3.8	295,119	0.03
htm	3.68	286,115	0.00
msg	2.45	190,342	0.03
html	1.83	142,386	0.00
docx	1.76	136,601	0.02
Others	33.99	2,641,312	11.09

Files by Category			
Category	Size (TB)	Percent	Files
Video	2.25	71.25%	70,146
Audio	2.82	89.16%	62,349
Program	0.10	3.06%	123,890
Office	1.05	33.25%	3,198,510
Graphics	0.97	30.80%	1,481,181
Mail	0.96	30.24%	194,265
Database	0.09	2.91%	31,947
System	0.01	0.29%	33,193
Text	0.04	1.12%	142,722
Internet	0.01	0.22%	578,385
Temp	0.24	7.68%	312,355
Archived	2.65	83.58%	141,289
Config	0.00	0.00%	21,952
Miscellaneous	0.24	7.58%	1,966

File Classification				
File Classification	TMP	Active	Inactive	Archive
	TMP/Log	3 Days - 1 Week	8-90 Days	Over 90 Days
Number of Files	36936	112984	1264719	6460884



File By Last Accessed		
Date Range	Files Count	Total Size (TB)
Over one year	1,390,104	2.92
6 Months - 1 Year	2,811,876	4.33
3 - 6 Months	1,423,162	1.25
8 - 90 Days	2,025,445	3.48
3 Days - 1 Week	167,033	0.50
Less than 3 Days	57,903	0.14

File By Last Modified		
Date Range	Files Count	Total Size (TB)
Over one year	5,082,130	9.84
6 Months - 1 Year	797,044	1.11
3 - 6 Months	581,710	0.63
8 - 90 Days	1,264,719	0.96
3 Days - 1 Week	112,984	0.06
Less than 3 Days	36,936	0.03

Storage Growth per Year					
Year	Total (TB)	Total Files	Files Added	File	Capacity
2005	1.37	2,415,862	2,415,862	0.00%	0.00%
2006	2.91	3,102,219	686,357	28.41%	112.36%
2007	5.20	3,935,346	833,127	26.86%	78.38%
2008	7.76	4,877,468	942,122	23.94%	49.14%
2009	10.14	5,857,112	979,644	20.09%	30.62%
2010	12.16	7,456,424	1,599,312	27.31%	19.95%
2011	12.63	7,875,523	419,099	5.62%	3.88%

Quantity By File Size							
Size of Data (KB)	<5	<50	<500	<5000	<50000	<500000	>500000
Quantity By File Size	2,137,795	3,122,437	1,725,616	684,473	154,729	48,104	2,369

Top 50 Files				
Name	Path	Size (GB)	Modified	Accessed
SaveData.MIG Backup.bkf	\spprodfs\users\ Transfer	40.21	12/22/2009	10/19/2010
	\dl-lynx\silver	13.47	6/3/2010	6/3/2010
	\spprodfs\users\COMM\Video\Media\Export video\Exports	12.83	9/26/2007	2/18/2011
2008-2009.avi	\dl-lynx\ARCHIVE\COMM 2008-2009	12.80	7/27/2009	3/4/2010
	\spprodfs\users\COMM\Video\Export video\Exports for compression	12.16	9/26/2007	2/18/2011
FinalOITPresentation.avi	\dl-lynx\ID2 Common\MISC	12.09	12/11/2007	3/2/2010
Win 7 - 64-bit.vdi	\spprodfs\users\OIT\Windows 7	10.80	10/13/2010	10/20/2010
cdd4e360-c8ab-11de-96d8-806e6f6e6963	\spprodfs\users\OIT\Smith.john\Windows 7\WindowsImageBackup\UT 137752\Backup 2010-02-03 000027	10.64	2/3/2010	10/20/2010
Ceremony.MOV	\spprodfs\users\COMM\Smith.jon\Video\Media\Export video\Exports for compression\Sombrilla Online	8.09	6/13/2008	2/19/2011
DTC10.MOV	\spprodfs\users\COMM\Smith.jon\Video\Media\Export video\Exports for compression	8.00	2/26/2008	2/19/2011
ABCD TIF 2 01.avi	\dl-lynx\dlat\$\doe.john\ernest	7.77	4/20/2009	4/23/2009
	\dl-lynx\dlat\$\doe.john\ernest	7.68	4/20/2009	4/23/2009
Higher Education 1 01.avi				
UCOM 00-V-HD	\spprodfs\users\COMM\Smith.jon\Video\Media\ABCD AdmissionsReceptions Media	7.66	9/15/2008	2/19/2011
07-V-HD	\spprodfs\users\COMM\Smith.jon\Video\Media\ABCD	7.46	10/7/2008	2/19/2011

Filters-V-HD	AdmissionsReceptions Media \spprodfs\users\COMM\Smit h.jon\Video\Media\ABCD	7.43	10/7/2008	2/19/2011
UCOM '0818-V-HD	AdmissionsReceptions Media\Rendered Media \spprodfs\users\COMM\Smit h.jon\Video\Media\ABCD	7.32	9/15/2008	2/19/2011
jyd146_node_a.log	AdmissionsReceptions Media \dl-lynx\dlat\$\WEBCT Staff\WebCT Logs\200910\COM_5623_00 1	7.29	11/12/2008	4/17/2009
ABCD Adeceptions.MOV	\spprodfs\users\COMM\Smit h.jon\Video\Media\Export video\Exports for compression	7.26	10/7/2008	2/19/2011
ABCD TIF Tape 1601 4 01.avi	\dl-lynx\dlat\$\doe.john\ernest	6.58	4/21/2009	7/14/2009
Windows XP Professional.vmdk	\spprodfs\users\CIAS\COMM ON\Training\Classes\Voice and Data Security\Version 1.23\VADS VM images\VADS XP	6.42	3/17/2010	5/22/2010
Personal Folders(1).pst	\spprodfs\users\OIT\Tom.Gr een\Exchange	6.19	3/4/2011	11/24/2010
OITpresentationV4. avi	\dl- lynx\dlat\$\Ted.Broker\avis_oi t	6.18	11/29/2007	6/9/2009
Personal Folders(1).pst	\spprodfs\users\OSP\Claudia .Garcia\Exchange	6.14	3/4/2011	12/17/2010
ABCD TIF Tape 14 01 1.avi	\dl-lynx\dlat\$\doe.john\ernest	6.13	4/20/2009	7/14/2009
ABCD TIF Romo 1 01.avi	\dl-lynx\dlat\$\doe.john\ernest	6.11	4/17/2009	4/23/2009
winxp.hdd.0.{5fbaab e3-6958-40ff-92a7- 860e329aab41}.hds	\spprodfs\users\MUSIC\Si.Mi llican\DocumentsBackup\Par allels\Microsoft Windows XP 1\winxp.hdd	6.02	2/27/2010	3/9/2010
s2 9-12.wmv	\streaming\streaming\$\Seme	5.97	10/29/2008	7/20/2010

	ster Classes\2008\Fall\MB\imbod en			
ABCD TIF 01 2.avi	\dl-lynx\dlat\$\doe.john\ernest	5.93	4/17/2009	4/23/2009
ABCD TIF 01.avi	\dl-lynx\dlat\$\doe.john\ernest	5.86	4/17/2009	4/23/2009
s2 5-8.wmv	\streaming\streaming\$\Seme ster Classes\2008\Fall\MB\imbod en	5.83	10/24/2008	7/20/2010
s2 1-4.wmv	\streaming\streaming\$\Seme ster Classes\2008\Fall\MB\imbod en	5.74	10/23/2008	7/20/2010
7-31- 09.Conversations.a vi	\dl- lynx\dlat\$\ARCHIVE\Ted.Bro ker\COM\COM\dvds	5.58	8/18/2009	11/11/2010
ABCD TIF 01 1.avi	\dl-lynx\dlat\$\doe.john\ernest	5.52	4/17/2009	4/23/2009
Sequence 01.mov	\streaming\streaming\$\Com mon\hesuer	5.23	7/14/2009	2/16/2011
Higher Education 1 01 1.avi	\dl-lynx\dlat\$\doe.john\ernest	5.19	4/20/2009	4/23/2009
Salley ValenzuelaABCD TIF 12 01.avi	\dl-lynx\dlat\$\doe.john\ernest	4.92	4/20/2009	4/23/2009
Mechanical Engineering.mov	\dl-lynx\dlat\$\College101	4.86	4/3/2009	8/27/2010
Personal Folders(1).pst	\spprodfs\users\CoSaE\ Exchange	4.86	3/4/2011	12/17/2010
personal folders(1).bak	\spprodfs\users\CoSaE\ \Exchange	4.79	2/7/2011	2/8/2011
Mechanical Engineering.mov	\dl-lynx\dlat\$\College101\FC C101\Capture Scratch\C101 & C&T	4.79	2/27/2009	4/28/2009
T.mov	\dl-lynx\dlat\$\College101\FC C101	4.78	2/27/2009	8/27/2010
ABCD TIF Tape 14 01.avi	\dl-lynx\dlat\$\doe.john\ernest	4.76	4/20/2009	4/23/2009
k.smith.pst	\spprodfs\users\PPLANT\Kat hryn\Exchange	4.57	3/5/2011	12/18/2010