

Report Provided by:



ICT Analytics

File Analysis Services Example Report

Commercial in Confidence

Contents Copyright ©

All Rights Reserved



Confidentiality, conditions and copyright issues

The information contained in this document is proprietary to ICT Analytics and is supplied on the understanding that it shall not be copied, stored in a retrieval system, or transmitted in any form, by any means whether electronic, mechanical, photocopying, recording or otherwise supplied to a third party without the prior written consent of ICT Analytics Ltd.

Copyright ICT Analytics



1 EXECUTIVE SUMMARY

The table below summarises the data collection from all of the file systems:

Total Used Storage - TB	25.45
Average File Size - KB	684
Max File Size - GB	55
Total Number of Files	39,038,102

The table below provides a summary of estimated savings.

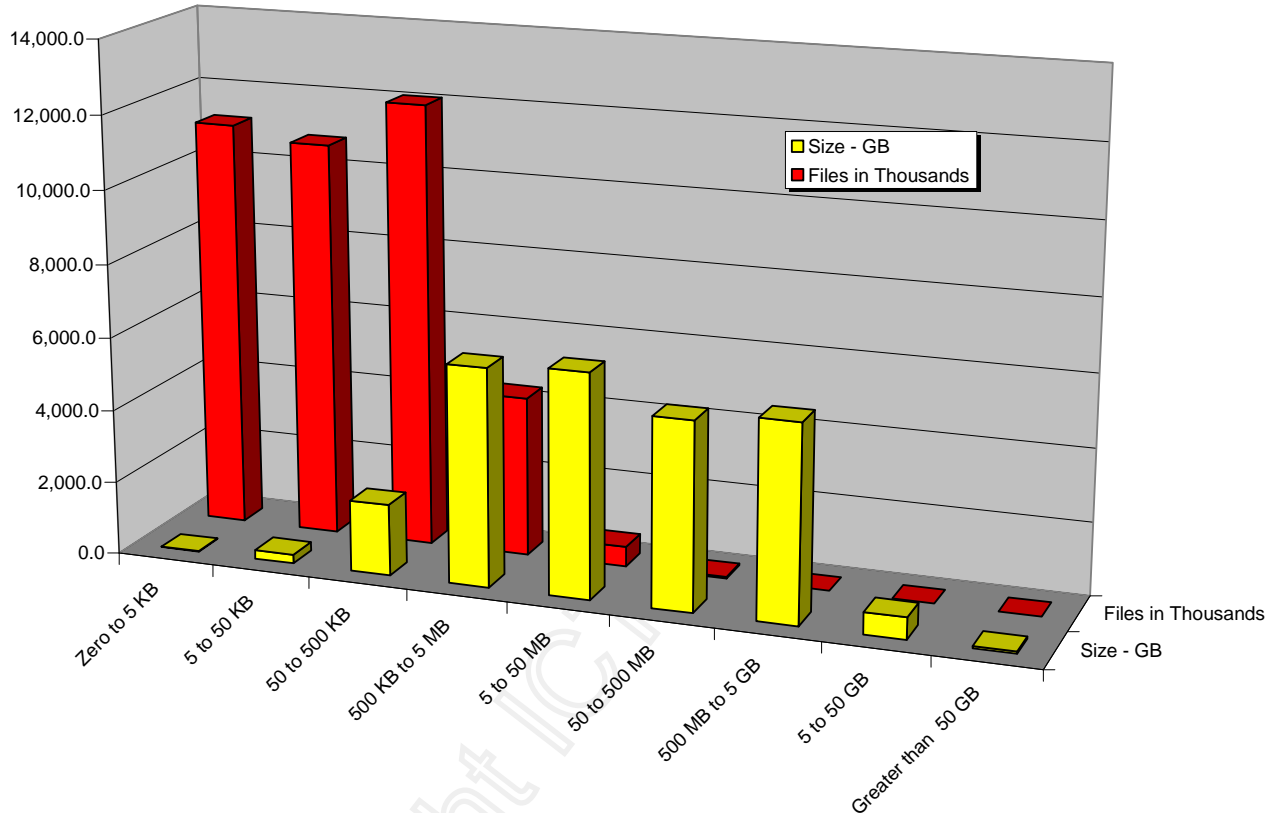
Description	Estimated Saving
PST & MSG - 50% reduction in space used	2,844
ZIP Archives - 50% reduction in space used	872
Archive of data not accessed in the last year - 50% reduction in space used	10,867
Compression of remaining data	3,346
De-duplication of remaining data	2,507
Total Saving	20,436

This represents a potential saving of around 80% in online storage; there would also be a significant impact and cost reduction in the backup and recovery infrastructure. Even, with more modest savings, the reduction in space and performance will be significant. Additionally, it is estimated that over 7 millions files can be removed from the file servers.



2 FILE DISTRIBUTION

The chart below shows the file distribution in terms of number and size.



2.1 KEY OBSERVATIONS

- The key observation is that the file system is made up of many small files; these are in the range zero to 500 kb in size and account for a total of over 32 million files. Although they do not account for a large amount of space, they will have a big impact on file system performance – particularly for backup.
- Similarly, a relatively small number of files account for the majority of the physical space. Detailed analysis to identify the types of files and opportunity for archive will save significant amounts of storage space.

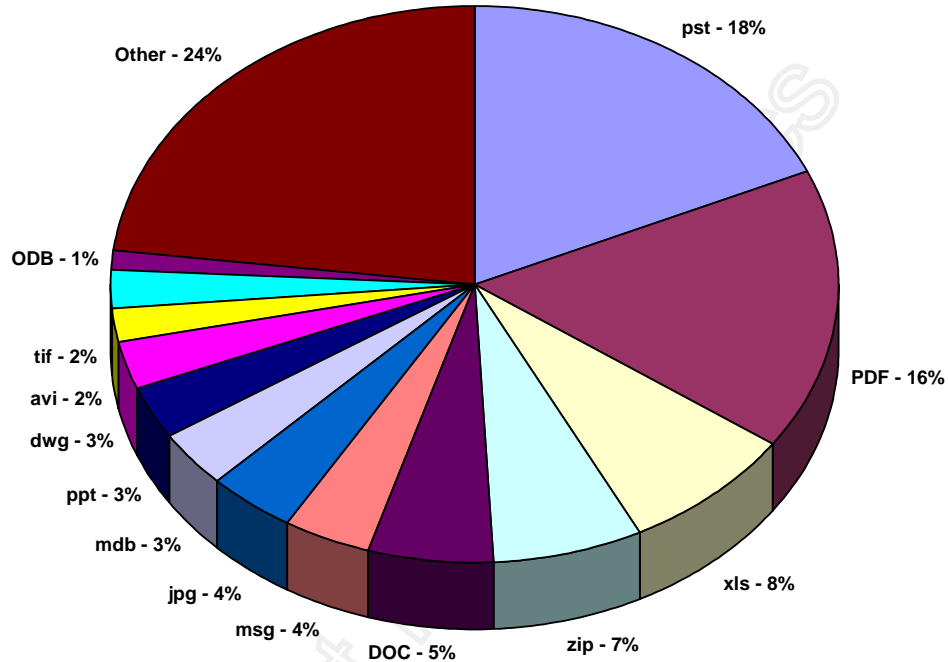
Therefore, this high level profile and associated recommendations have been used to focus detailed analysis on all of the data collected.



3 FILE TYPE PROFILE

3.1 OVERVIEW OF FILE PROFILE

The chart below summarises file type distribution:



Extension	Description	Size - GB	Cumulative %	Number of Files
pst	Personal Folder File (Microsoft Outlook)	4,684	18%	104,988
pdf	Portable Document file (Adobe Acrobat)	4,137	35%	5,613,578
xls	Worksheet file (Microsoft Excel)	1,973	42%	3,761,628
zip	Zip file Compressed archive	1,744	49%	485,694
doc	Microsoft Word document	1,353	55%	4,541,807
msg	Mail message (Microsoft)	1,003	59%	2,955,113
jpg	JPEG bitmap	994	62%	1,604,957
mdb	Database (Microsoft Access)	850	66%	48,572
ppt	PowerPoint presentation (Microsoft)	774	69%	301,779
dwg	AutoCAD drawing	691	72%	1,282,802
avi	Microsoft Audio Video Interleaved file for Windows movie	556	74%	6,665
tif	Tag image bitmap file	549	76%	450,468
odb	OpenDocument Database	281	77%	2,858



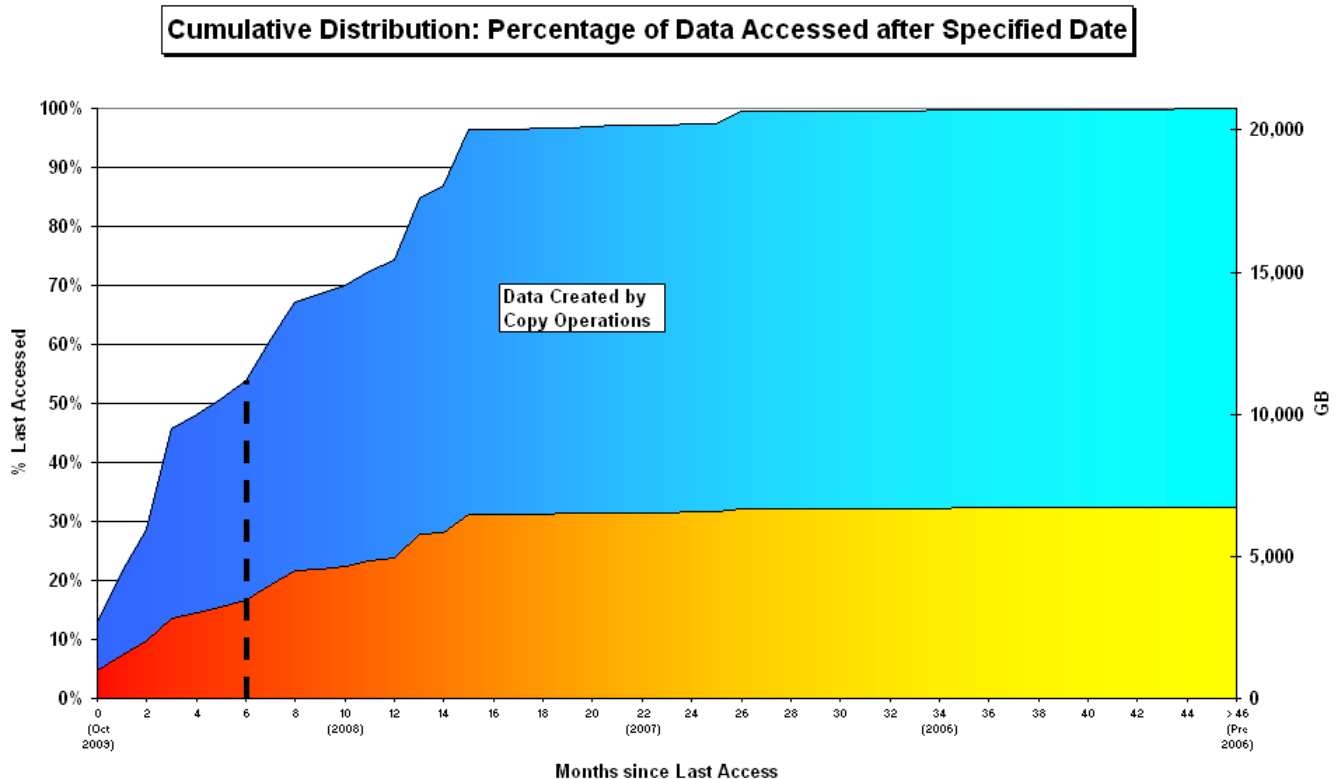
3.2 OVERVIEW OF FILE ACCESS PROFILE

The graph below looks at the profile of files based on an estimation of the last access date; this is displayed as cumulative distribution going back in time. Analysis of the "last access date" is not a straight forward task as there are many user actions - such as just looking at the properties of a file, or even just selecting a file in Explorer - that will update the last access date even if the file was not opened. Analysis of the data shows "group updates" to last access date occurs where files in the same directory structures are accessed simultaneously which clearly is not as a result of real access. These factors need to be taken into account for any strategic decisions for automated policy based archiving.

The graph below also excludes pst files which need to be addressed separately as already stated. The File Access analysis takes into account all creation, modification and change dates to derive a more accurate estimation of last access date of data grouped into a monthly distribution. The plot in blue indicates where files have been created as part of copy operations (not move operations). This is clearly a frequent activity involving much of the data content.

Viewing the graph, picking any date shows the amount of data that has been accessed since that date by its intersection with the area graph.

A dotted line on the graph indicates data that has not been accessed in over 6 months. The total is a little over 50%, but a large majority is created by copy and other operations that do not in fact open the file for "real" access. The more realistic access percentage is likely to be closer to 20%.



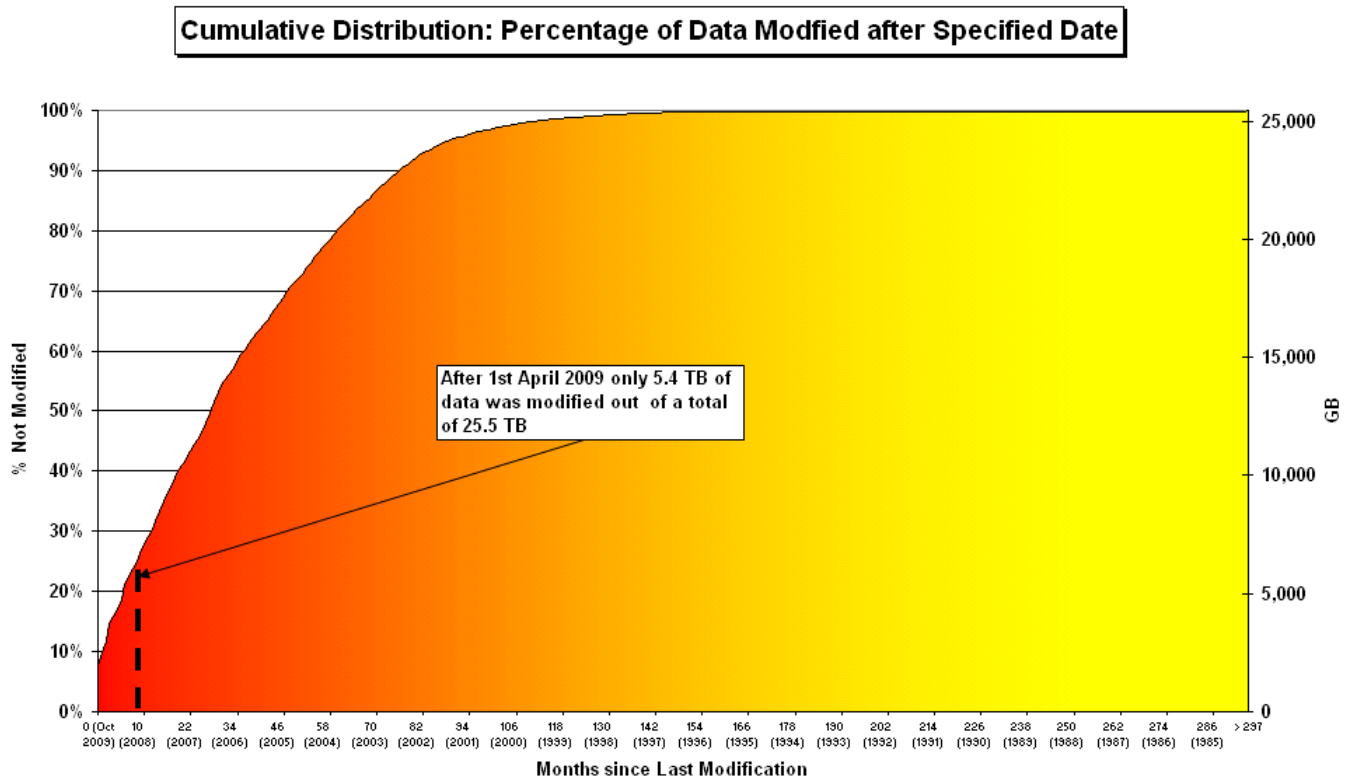


4 FILE MODIFICATION PROFILE

4.1 OVERVIEW OF FILE MODIFICATION PROFILE

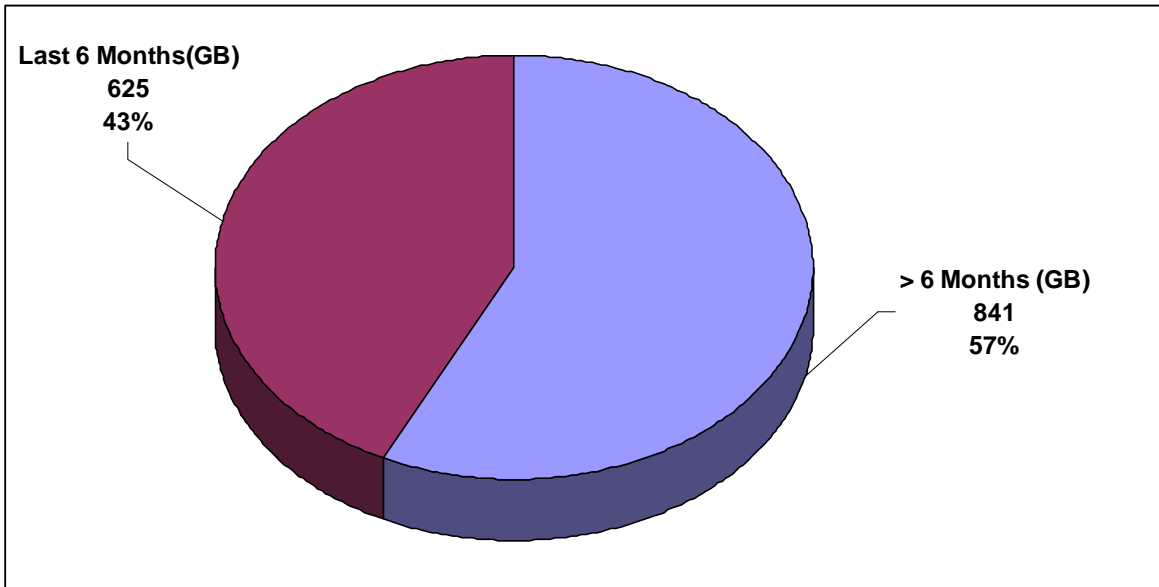
The graph below is a cumulative distribution based on the last modification or change date of a file. As in the previous section, viewing the graph and picking any date shows the amount of data that has been modified since that date by its intersection with the area graph.

A dotted line on the graph indicates data that has not been modified in the last 6 months. The total is a little over 20% indicating the majority of data is relatively static. As has already been observed, over 6 TB of data of a static type anyway –e.g. pdf; jpg etc. This has significant implications on backup which is discussed in more detail in the report.





Based on detailed analysis, and automated changes to directory structures, a 57% reduction in data backed up can be achieved. This is illustrated in the pie chart below:



5 FILE ATTRIBUTES PROFILE

5.1 OVERVIEW OF FILE ATTRIBUTES PROFILE

The tables below summarises file attributes broken down into each category:

Attribute	GB	Number of Files
Archive	14,421	18,159,068
Compressed	86	73,502
Directory	0	0
Hidden	52	352,829
Normal	10,499	19,578,673
Offline	0	1
Read only	682	1,973,273
System	10	209,889
Temporary	2	6,248



5.2 OVERVIEW COMPRESSION PROFILE

Minimal data is currently compressed on disk. Of the data that is compressed, an average compression ratio of 1.83 is achieved. The table below uses these measured compression ratios to predict the storage saving from compression. No data was available for odb so a ratio of 2 was used as this is typically compressible data.

Type	Size - Disk	Ratio	Saving
pst	4,685	1.20	781
pdf	4,139	1.15	529
xls	1,973	2.51	1,187
zip	1,745	1.00	0
doc	1,355	1.61	515
msg	1,003	1.32	246
jpg	994	1.01	6
mdb	850	3.84	629
ppt	774	1.08	60
dwg	691	1.66	275
avi	556	1.00	0
tif	549	1.12	60
odb	281	2.00	141
Other	5,854	1.83	2,655
			7,083



6 FILE DUPLICATION PROFILE

6.1 OVERVIEW OF FILE DUPLICATION PROFILE

In terms of exact file duplication, the following table summarises the key statistics:

Total baseline files with one or more with duplicates:	2,192,168
Total file duplicate count (including baseline):	17,531,970
Total baseline space used – GB	1,246
Total duplicated space used (including baseline) - GB	6,005

TO RECEIVE YOUR FREE FILE SYSTEM ANALYSIS OVERVIEW REPORT

VISIT WWW.ICTANALYTICS.COM

DOWNLOAD THE FAS UTILITY

RUN THE UTILITY ON YOUR FILE SERVERS

SEND THE COLLECTED DATA BACK FOR ANALYSIS

BY IMPLEMENTING THE RECOMMENDATIONS IN A FULL FAS REPORT
YOU WILL TYPICALLY REDUCE YOUR UNSTRUCTURED ON-LINE DATA
STORAGE FOOTPRINT BY 60%

REDUCING YOUR STORAGE FOOTPRINT WILL SAVE YOU MONEY BY:

- REDUCING YOUR NEED FOR ADDITIONAL STORAGE PURCHASES
 - REDUCING THE TIME TAKEN TO BACKUP YOUR SYSTEMS
 - REDUCING THE AMOUNT OF REPETITIVE OFF-LINE STORAGE
 - REDUCING THE TIME TO RECOVER INDIVIDUAL FILES
- REDUCING YOUR RECOVERY TIME IN THE EVENT OF SYSTEM FAILURE