CASAHL Case Study: File Share Assessment

Introduction: Enterprise Concerns and Assessment Goals

Successful migrations start with migration plans, and effective migration plans start with assessment. CASAHL is often approached by companies that need someone to help them make sense of their expansive legacy deployments before they can plan for and carry out a migration. The enterprise that provided the data for this case study had two interrelated concerns: they wanted to migrate to a new system, but they were worried that migrating their legacy file share content without the legacy clutter would be difficult, and that migrating their content would just recreate existing problems in their next deployment.

By engaging with CASAHL for an assessment, the enterprise gained full visibility into their existing file shares and the insights needed to build a migration plan. The enterprise used the data-driven reports provided by the assessment to inform their decision-making process, enabling them to quickly understand their legacy content and plan cost-effective system optimizations that would keep their content organized and accessible in a new deployment.

The Problems: Understanding Content, Reducing Migration Scope, and Planning a Migration

The enterprise wanted to migrate to Office 365, but was concerned about keeping all the details and metadata of the content in their file shares intact. Part of their apprehension also came from an insufficient understanding of the content in their file shares, and how much of that content was actively used. Enterprises with legacy file share deployments often feel that they don't understand the content in their deployments, and worry that any new deployments they migrate to will quickly become as cluttered and disorganized as their legacy file shares.

To avoid carrying unused and disorganized content into their new deployment, this enterprise had to assess and understand the content in their existing file shares. Once active content was identified and unused or duplicate content was filtered out, the enterprise was able to identify potential problems that could impact their target deployment, and decide what content to migrate in order to build and carry out a migration plan.

The Solution: Assessment and Informed Decision-Making

CASAHL did an assessment to determine how much content the enterprise had across their legacy file shares, then provided them with detailed reports and recommendations to help them set criteria for a successful migration.

1

The first step in addressing the enterprise's migration challenge was to provide full visibility into the resources stored in the enterprise's file shares. CASAHL's assessment tools were used to inventory the content in their file share deployment and identify the resources that were still active and valuable. The initial assessment results found nearly 6 million files spread across more than 16,500 root folders. It also turned up several obstacles to migration that the enterprise had suspected, but not known, that they needed to address, including large amounts of duplicate content and structural issues caused by folder organization.

## Active Content Identification: Focusing on Key Content

A key part of any assessment is discovering which legacy content is still active or has become inactive. In the case of this enterprise, examining the levels of activity in root folders over time revealed a clear trend:
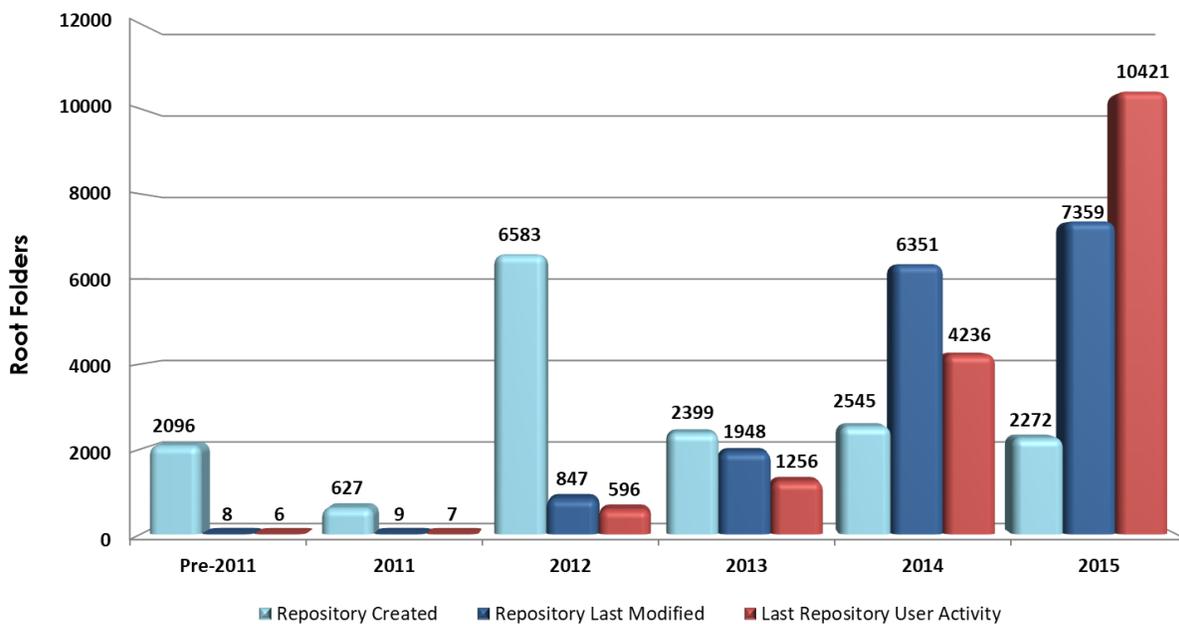
More and more folders were becoming inactive every year, even as the number of new root folders created stayed consistently low after 2012. Root folders that were abandoned between 2012 and 2015 added up to mean that a significant portion of total folders were going unused. Identifying active folders, or folders that had been modified within specific timeframes, gave the enterprise insight into content that should be considered for migration.

It also gave the enterprise insight into content that could safely be retired or archived:

| Doc Range	Root Folders
0	6,902
1-20	5,055
21-50	1,255
51-100	715
101-200	514
201-500	699
501+	1,382
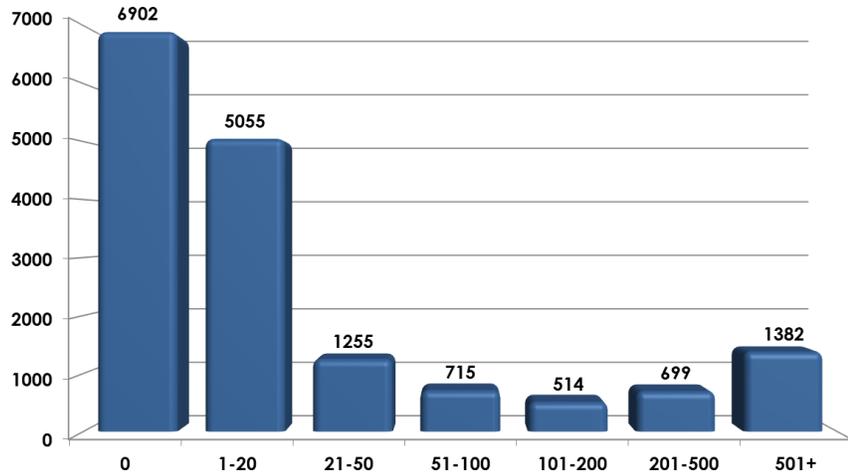
The number of root folders that did not contain files, or contained a maximum of twenty files, suggested that a combined seventy percent of root folders either held content that was infrequently used and possibly abandoned, or held no content at all. Once the enterprise became aware of this, they were able to determine whether any content and folders among that 70% were worth migrating, and prepare the rest for archival.

Archiving unused content on this scale allowed the enterprise to sharply reduce the scope, and therefore the cost, of their migration. Identifying active folders gave the enterprise the insights needed to identify potential structural issues and trends that could be used to avoid recreating those structural issues in their new deployment. Identifying inactive content also gave the enterprise the opportunity to greatly reduce the scope and cost of the migration by excluding those sites from the migration.

Duplicates and File Types: Reducing Migration Scope and Preventing Future Clutter

CASAHL's assessment reports always include a breakdown of file types found in the assessed deployment. This helps enterprises to better map the content they have against the needs of any future deployments by getting a clearer sense of what they have to work with, and makes it easier to identify possible sources of clutter in future deployments, such as duplicate files.

| Family Type	Files
Adobe	575,542
Archives	18,707
Audio and Video files	3,722
Blocked	2,945
Data	861,172
Google documents	0
Image files	879,940
Office documents	1,609,785
Other documents	144,685
Scripts	69,460
SharePoint Design files	201
Source Code	66,888
System	145,212
Unknown	1,275,631
Web Documents	412,344
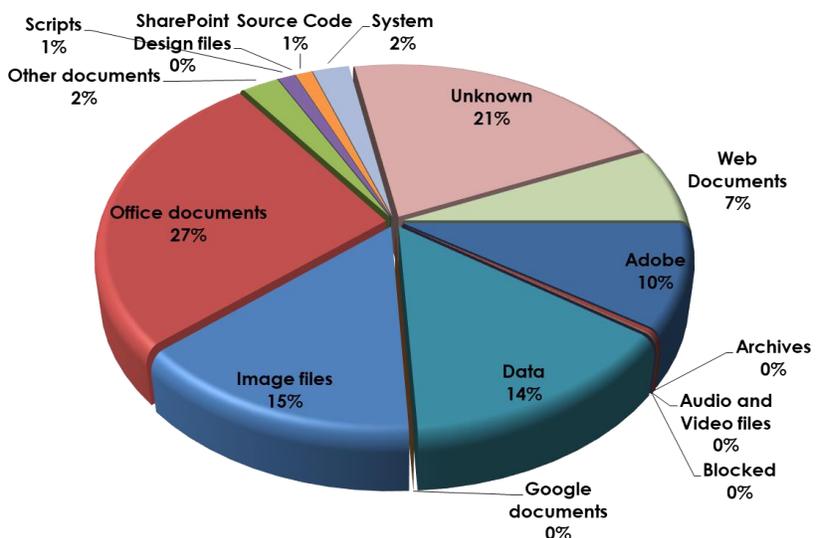
Some types of content, such as Office documents, are more likely to produce duplicate content than others. Duplicate files are a common source of clutter, especially in legacy on-premises deployments and file shares. The assessment results revealed 2,829,388 potentially duplicate files in this enterprise's file share deployment. Identifying this duplicate content, along with other unused content, allowed the enterprise to sift out extraneous files and reduce the scope of their migration, which in turn reduced the cost and effort required to migrate.

Identifying duplicate content also helped the enterprise preemptively clean up their new deployment. In CASAHL's experience, users often save personal copies of files they received by email, but rarely refer to them again; this is one of the most common causes of file duplication.

Identifying duplicate content is a great way for enterprises to reclaim storage space whether or not they plan to migrate, but it's especially helpful for enterprises that want to migrate to modern cloud systems like Office 365. Since the structure of Office 365's cloud-based document libraries makes this type of file duplication largely obsolete, identifying and deleting duplicate files helps keep their next deployment that much more efficient by removing the need to create duplicate files in the first place. By identifying and cutting down on duplicate content now, the enterprise was able to both reduce their eventual migration cost & effort, and prevent the same problem from recurring in the future.

Active Users: Gathering Input

Once an enterprise has identified its active content and applications, it's time to turn to the report section on active users. Engaging an enterprise's most active users provides valuable insights when planning for a migration and, after the migration, promotes better adoption of the target deployment. Moving power users and their content to new deployments first also gives them a chance to adjust to the new system, thus building momentum for the new system from the inside.
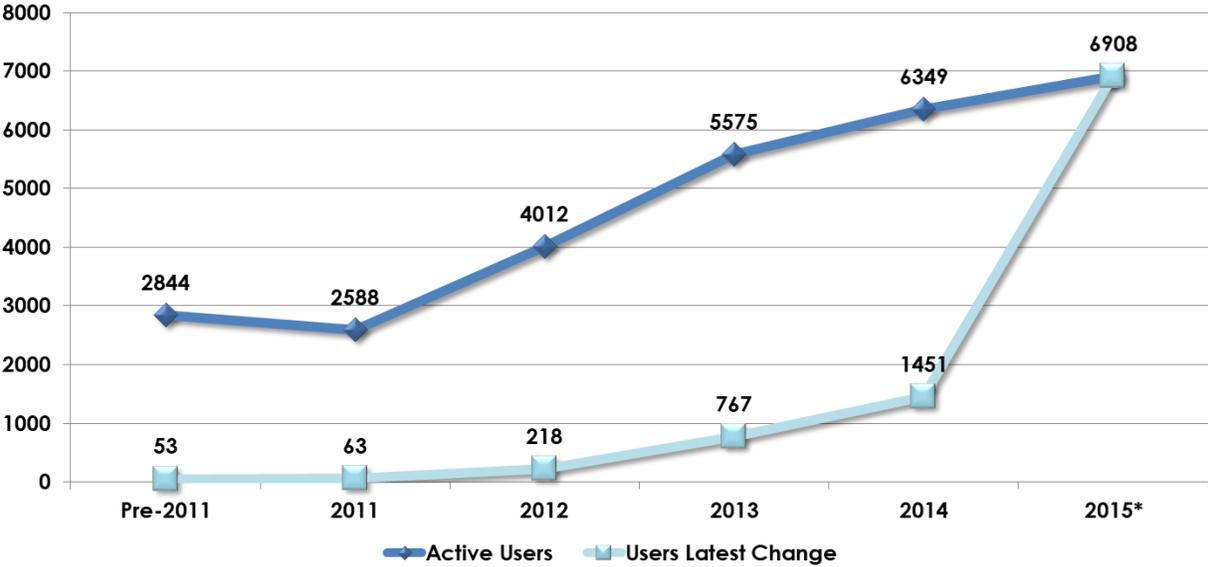
By identifying the most prolific users and inviting them to sit in on the migration discussion, the enterprise ensured that no important content was overlooked. Key users whose concerns are addressed during migration planning often become champions of the new system among their peers, as was the

case with this enterprise, whose active users helped ease the transition for other users and increased adoption of the new deployment after the migration.

Conversely, the enterprise was able to save tremendously on software licensing by identifying inactive user accounts, whether from previous employees, accidental redundancy, or users who simply don't use legacy systems. The gap between active users and total user licenses illustrates the gap between total users and active users, up until the enterprise removed the unused user licenses in 2015.

By soliciting the input of these key users during planning, the enterprise was able to make their migration more thorough and therefore more effective. Experience shows that enterprises willing to collect active user input during planning *and* migrate their simple, highly-used content first find that those users are among the first to be active in the new platform and quickly become productive in their new home. User data and input often helps enterprises remediate complex migration problems or identify areas for improvement as well, as was the case with this enterprise; the next section describes some of the problems identified in this way.

Illegal Files, Folders, and Pathways: Preemptively Resolving Migration Blockers

Duplicate content wasn't the only obstacle between this enterprise and the efficient, orderly cloud deployment they wanted to migrate to. By using the assessment report's information on potential migration blockers, the enterprise was able to identify illegal filenames, illegal folder names, and other structural issues, and resolve those before they caused problems during their migration.

Problems with illegal folder and file names crop up most frequently when a file has an 'illegal' name – a name that has characters the target deployment won't support, is too long, or has certain character patterns. To avoid a bunch of unpleasant surprises during a migration, enterprises need to be aware of any restriction differences between their source and target systems, and to have a way to identify content that doesn't meet those restrictions. CASAHL's assessment report gives enterprises the information they need to handle both.

Category	Count	Significance
Files with an illegal name	220,469	File and folder names can be illegal in SharePoint due to unsupported characters, names longer than 128 characters, or certain character patterns.
Folders with an illegal name	94,444	
Files with a long path	25,947	SharePoint supports URLs up to 260 characters. If folder + file paths exceed 200 characters, then it runs the risk of exceeding this limit after factoring in SharePoint server, site, and library names.

Files and folders with long paths can also cause problems. While they don't generally run afoul of the same system requirements as illegal file and folder names, overly lengthy file and folder names can make the paths for the files and folders too long, and therefore accidentally exclude folders & the content in them from migration. CASAHL's assessment maps the extent to which folders are nested in a given deployment to help enterprises deal with this potential pitfall:
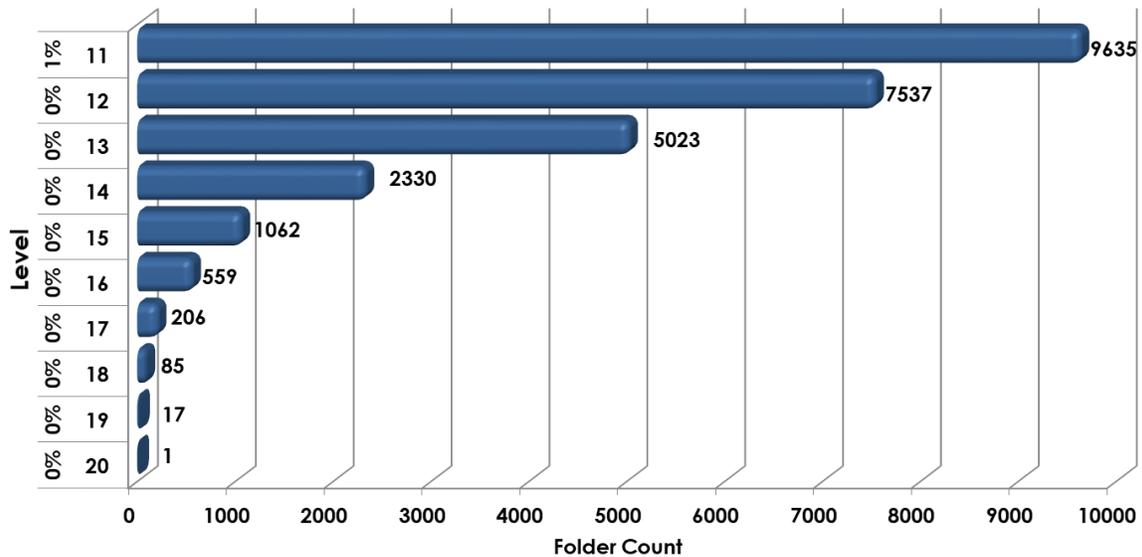
In this case, assessment revealed 220,469 files with an illegal name, 94,444 folders with an illegal name, and 25,947 files with a too-long path. By running an assessment before migrating, the enterprise caught these migration issues so they could be resolved before the actual migration, and gained insights into better ways to organize the next deployment.

Results and Next Steps

In the end, the enterprise reviewed the information in CASAHL's assessment report and decided to go through with the migration. The detailed reports helped the enterprise's migration team decide which content needed to be migrated, what could be archived or deleted, and what should be optimized for their new deployment Office 365. Because the enterprise assessed their content before migrating, they were able to profile their environment, use that profile to identify and remove migration blockers, and generate estimates of the effort and time investment the migration required. These details and the data included in the reports offered ways for the enterprise to formulate a migration plan while immediately reducing the scope, and therefore time and expense, of their eventual migration.

Overall, this enterprise engaged CASAHL to assess roughly 3.3 TB of data spanning over 16,522 root folders. The results of CASAHL's assessment allowed the enterprise to begin planning a more effective and efficient migration that addressed concerns about organization and efficiency in their next deployment. The next steps for the enterprise were to consult their most active users to ensure that user input was taken into account, and then to build a migration plan based on the data provided by the assessment.

If you are interested in learning more about this enterprise's migration experience or other details of file share assessment, please contact us for more information. If you are interested in learning more about the next stage of the DART migration process, please ask your CASAHL representative for other case studies. If your enterprise is interested in learning more about CASAHL's automated migration techniques or the assessment of multiple file shares with one solution, read more by visiting our site. If you have questions about CASAHL's assessment service and would like to learn more about engaging CASAHL to do the same for your enterprise, please contact us at info@casahl.com.