CASAHL Crawler and Analyzer for Enterprise Content

Companies that take the time to understand their legacy content always find opportunities to organize, optimize, and streamline their systems. It's the first step in a migration, or in deciding if migration is right for your enterprise. CASAHL is here to help – we've perfected a set of products that let enterprises identify and understand the content in their legacy systems quickly and painlessly.

These tools are based on the DART lifecycle, CASAHL's guiding principle. Time and time again, we saw that enterprise migrations were only successful when the enterprises understood their content and made data-based, realistic migration plans, so we built a product suite to make that possible. The CASAHL Crawler is the Discovery component of our product suite. It triages content across a variety of legacy systems (shown below with Office 365), including SharePoint, Google Drives and Sites, Confluence, Exchange public folders, Box, and Dropbox, and collects information on many different metrics. The CASAHL Analyzer, the Analysis component of DART, takes the information collected by Crawler and turns it into detailed reports. Together, Crawler and Analyzer enable enterprises to conduct a thorough assessment of all eligible legacy systems and make factual, data-based decisions to determine the best path forward.

Streamline Office 365 Migration
CASAHL Complements FastTrack for Full Office 365 Deployments

Site and App Platforms
- SharePoint Apps
- SharePoint Online
- Notes/Domino/Quickr
- Google Sites
- Confluence
- Exchange Public Folders

File-Centric Systems
- Google Drive
- Box
- Dropbox
- File Shares

Enterprise DBMSs
- Access, SQL Server...

Other Sources

Discovery → Analysis → Rationalization → Transition

Office 365
- SharePoint Online
- OneDrive · Power BI
- PowerApps · Delve
- Custom Apps · Groups

Microsoft Azure
- SharePoint Server
- Azure SQL Database
- Other Targets

On-Premises
- SharePoint Server
- SQL Server
- Other Targets

Proven with Over 650 Enterprises

How Does It Work?

Crawler examines the client's content systems to discover sources of information and collect key usage statistics from those sources, then compiles this information and passes the data it discovered to Analyzer. The Analyzer examines the data and provides detailed summary reports used to help enterprises make informed, effective decisions about their legacy deployments.

The concrete set of benefits provided by Crawler and Analyzer offer enterprises an understanding of legacy content that helps shine light on the best paths forward. Enterprises that decide not to migrate still find opportunities to reduce deployment footprints and save on licensing; enterprises that decide to migrate to the cloud learn how best to pursue a smooth, painless, cost-effective migration. By assessing content before migrating, enterprises become able to flag unused content for archival or retirement, accurately estimate the cost and duration of a migration, identify potential migration blockers before migrating, build better migration plans, and decide what type of migration, if any, is the best solution for their needs.

CASAHL Crawler for Enterprise Content
What is Crawler? How does it play into the overarching DART product suite?

The CASAHL Crawler's function is to discover all the sources of data in an enterprise's legacy systems, including collaborative applications, and then to collect key information such as usage history, metadata, custom parts, users, and content statistics from the data sources it identifies. Candidate sources for Crawler include legacy versions of SharePoint, file shares including Box and Dropbox, Confluence, Exchange public folders, Google Drives and Sites, and CMIS-compliant systems, among others. There is also a specialized version of Crawler for IBM / Lotus Notes products as part of the DART suite for Lotus Notes Applications.

There are four unique aspects of Crawler that make it thorough, efficient, and unobtrusive:

- Its *small footprint* – Crawler is a simple Windows-based piece of software. It can be installed on any PC that has the client software for the underlying system that needs analysis. After a simple 30-minute training session, it can even be run by the enterprise's own system administrators.
- Its *non-intrusive* nature - Crawler only collects design-level information and does not examine the data itself. It only requires reader-level access permissions and does not need to be installed on the servers. As a result, it is both non-intrusive and secure, making it easy to continue high-performance work while Crawler is running.
- Its *multiple content system capability* – Crawler works with a wide variety of content systems, including legacy SharePoint, Google Drives, Exchange public folders, file shares, CMIS-compliant systems, Dropbox and Box, and Confluence.
- Its *parallel data collection* feature – multiple instances of Crawler can be installed to simultaneously access and collect design and usage data from any number of servers and locations. It can also be scheduled to run in hands-free mode, making it possible to

inventory an entire collection of content sources without tying up your enterprise's IT staff and resources.

This screenshot shows the Crawler in action for SharePoint. The same drag-and-drop UI can be used for any content system Crawler runs in, including file shares like Box and Dropbox, Exchange public folders, Google Drives, and ECM systems, which are often created as "shadow IT" and managed separately from more traditional enterprise systems.

Once each instance of Crawler has finished its run and collected a batch of data, that data is uploaded to CASAHL's secure cloud. From there, the batches are merged and prepared for analysis by the next component of CASAHL's DART suite, Analyzer.

CASAHL Analyzer for Enterprise Content

What is Analyzer? How does it play into the overarching DART product suite?

The CASAHL Analyzer's ultimate goal is to help the client figure out which content to migrate to the target deployment, and which sets of data should take priority during the migration. To that end it takes the batches of data collected by the CASAHL Crawler and generates highly-detailed reports on the complexity, volume, and usage statistics of the data sources it identified.

These reports make it easy for transition planners to identify candidate sites that would be straightforward and/or high-priority targets for migration. They identify low-value resources, such as applications and sites that have low (or no) collaboration and high complexity, allowing transition planners to focus transition efforts on the most popular collaborative applications, identify resources that can be archived or deleted, and form preliminary transition plans and effort estimates. Using the data in these reports also lets enterprises determine migration priorities or set up simple content to migrate first.

Analyzer compiles its reports from data collected by Crawler, meaning it often has to handle data from multiple legacy sources simultaneously. In those cases the reports are broken down system by system in order to better report the relevant details for each legacy deployment. Some of the details recorded for each system type are listed below.

SharePoint

For SharePoint, Analyzer delivers a PowerPoint report that provides a high-level executive summary of the selected environment, highlighting key takeaways on sizes, distributions, usage, and properties of sites, files, users, and more. A detailed Excel report contains site-level scale/complexity scores, users, usage, and listings of all SharePoint objects (such as InfoPath forms, workflows, and more). Lastly, an Overall Summary Report (Excel) provides views of the environment organized by usage, file types, and author. Pivot charts in the summary report allow you to customize the layout of the information for easier interpretation or presentation.

SharePoint Analysis: Reported Data (by site)

Site Size	Analyzer compiles data on the total size of the entire site, the total size of the files in that site, and the sizes of each individual file larger than 250MB.
Site Users	Crawler reports on the author and editor activity for each site.
Complexity Rating	Analyzer rates SharePoint sites on their size, their complexity, and the amount of work required to redesign them for modernized systems.
Scale Rating	Each SharePoint site that has been assigned a Complexity Rating is given a Scale Rating that indicates the estimated time needed to redesign that site for a modernized system.
Site Collection, Sites and Sub-Sites, and ACL	Analyzer lists the settings and properties for each site and sub-site, as well as the permission specifications in the form of access control lists.
Lists	Analyzer collects data on the settings, properties, templates, and object counts for all lists and libraries.
Folder Path Depths	Analyzer records the amounts of nesting in, and the layouts of, file trees.
Solutions and Sandbox Solutions	Analyzer records data for both types of custom SharePoint apps, including how they are customized, what their permissions are, and how they are deployed.
Workflow Templates and Instances	Analyzer examines and records the types of templates used and the frequency of their use to determine which are most important to migrate.
Content Types	Crawler provides a compiled list of the lists, alarms, custom SharePoint sites, workflow templates, web apps, and other content it discovers.
Alerts	Analyzer compiles data on the alerts found in the system, including their event types, who and what they are used by, and the location of items used during the alerts.
Farms	Analyzer maps out the layout and configuration of on-site servers in the system.
Documents, List, Items, and Attachments	Analyzer provides a compiled list of documents, list items and attachments to those items, the properties of the list items, any files with blocked file extensions, and any duplicate documents.
InfoPath Templates and Instances	Crawler reports on the use of InfoPath in conjunction with the enterprise's SharePoint systems.
Event Receivers	Crawler reports on the use of Event Receivers in conjunction with the enterprise's SharePoint systems.

SharePoint Analysis: Reported Data (by site)

Web Apps	Analyzer reports on the location of the app in the SharePoint server farm, any blocked file extensions the app may specify, the time zone the app is set to, and any specific web configuration mods the app uses.
Permission Inheritance	Analyzer provides a summary of the ownership of and access to account permissions as spread across the site.
Usage and Age	Crawler lists the date created, versions created, and modifications by location for documents and list items, with the data organized by month.
Pages and Web Parts	Crawler collects data on the web-centric uses of SharePoint for both Pages and Web Parts.
Features	Crawler collects data on all SharePoint Features in the system, including their properties and the SharePoint items associated with each feature.

Google Drive

Analyzer delivers a high-level executive PowerPoint summary of the selected environment highlighting key takeaways on sizes, distributions, usage, and properties of sites, files, users, and more. An Overall Summary Report in Excel details views of the environment organized by usage, file types, and author/editor. Pivot charts allow you to customize the information and view information based on documents, items, users, and sites/containers.

Google Drive Analysis: Reported Data

Drive Folders	Analyzer reports on the settings, properties, and total files found in each Google Drive for which an instance of Crawler was run.
Folder Path Depths	Analyzer examines and records the amounts of nesting and layouts of file trees in the systems.
Drive Users	Analyzer compiles list of author and editor activity for each Google Drive.
Usage / Age	Analyzer compiles a report on the date files were created and the modifications made to them over time.
File Properties, File Types, and Total Files	Analyzer compiles a list of files, the properties of those files, the types of those files, and any instances of duplicate files.
Total File Size	Analyzer compiles data on the total size of the files in each Drive, and the sizes of each individual file larger than 250MB.
Most Used Office File Types by Count	Analyzer compiles counts of file instances as categorized by file type as well as data on how often those file types are used.
Files by File Family	Analyzer compiles a list of all Office documents in named groupings of file types, e.g., .docx, .doc, .xls, and .xlsx for Office file types, etc.

File Shares, Exchange Public Folders, and ECM Systems

CASAHL delivers a high-level PowerPoint executive summary of the selected environment highlighting key takeaways on sizes, distributions, usage, and properties of sites, files, users, and more. An Excel Overall Summary Report details views of the environment organized by usage, file types, and

author/editor. Pivot charts allow you to customize the information. View information based on documents, items, users, and sites/containers.

File Shares, Exchange Public Folders, and ECM Systems Analysis: Reported Data

Source (Drive/Folders/etc.)	Analyzer reports on the settings, properties, and total files found in each system for which an instance of Crawler was run, as well as the permission specifications for the source in the form of access control lists.
Folder Path Depths	Analyzer examines and records the amounts of nesting and layouts of file trees in the systems.
Users	Analyzer compiles a list of author and editor activity for each location in the system.
Usage / Age	Analyzer provides a report detailing the date the files were created and the modifications made to them over time.
File Properties, File Types, and Total Files	Analyzer compiles a list of files, the properties of those files, the types of those files, and any instances of duplicate files.
Total File Size	Analyzer compiles data on the total size of the files in each Drive, and the sizes of each individual file larger than 250MB.
Most Used Office File Types by Count	Analyzer compiles counts of file instances as categorized by file type as well as data on how often those file types are used.
Files by File Family	Analyzer compiles a list of all Office documents in named groupings of file types, e.g., .docx, .doc, .xls, and .xlsx for Office file types, etc.

Enterprises that want to assess legacy Notes/Domino deployments should see our specialized versions of Crawler and Analyzer for Lotus Notes & Domino for more information on working with those specific systems.

Next Steps and Other Information

Once the CASAHL Analyzer has compiled its reports and decision-making among users, managers, decision-makers, and developers is under way, it's time for enterprises that want to migrate to start the rationalization process. Rationalization helps enterprises collaboratively develop, refine, and manage transition plans. The next product in the DART suite, the DART Lifecycle Manager (DLM), takes the reports compiled by Analyzer and puts the information into a custom workspace where rationalization takes place, which business users and IT personnel across the enterprise can use to construct a migration plan.

Click here to read more about the DART Lifecycle Manager, or learn more about the DART suite on our site. For more information on the assessment process, the benefits it can offer your enterprise, or the next steps if your enterprise doesn't necessarily want to migrate, feel free to contact us. Enterprises that want the benefits of assessment without running the entire assessment themselves are encouraged to explore CASAHL's Pre-Migration Assessment service. Enterprises that have already completed an assessment should see our rationalization product or rationalization service for their next steps.

6

About CASAHL

CASAHL helps enterprises optimize their content/collaboration deployments and move to the cloud, starting with legacy system assessment and continuing all the way through a complete migration. We work with source enterprise systems including Lotus Notes, SharePoint, Office 365, Confluence, Google Sites, file sharing systems (file servers, Box, Dropbox, Exchange public folders, and Google Drive), enterprise content management systems, and database systems. CASAHL's product focus includes modernizing and migrating critical content and collaborative applications from legacy systems to the cloud, keeping related resources optimized post-migration, and helping enterprises attain peak value from their new investments.

CASAHL is a proud repeat winner of Microsoft Partner of the Year for migration solutions and has been exclusively focused on content/collaboration optimization products and solutions since 1993. Together with systems integrator partners including Microsoft, Sogeti, Perficient, Avanade, and others, we have migrated the content and apps of over 650 enterprise accounts into SharePoint and Office 365 over the past eight years. To learn more about engaging CASAHL's migration and ongoing optimization solutions, please contact us, and see our blog for perspectives on up-and-coming enterprise collaboration/content topics.

CASAHL TECHNOLOGY, INC.
2400 Camino Ramon
Building K, Suite 355
San Ramon, CA 94583

Phone: 925-328-2828
Fax: 925-328-1188
Email: info@casahl.com