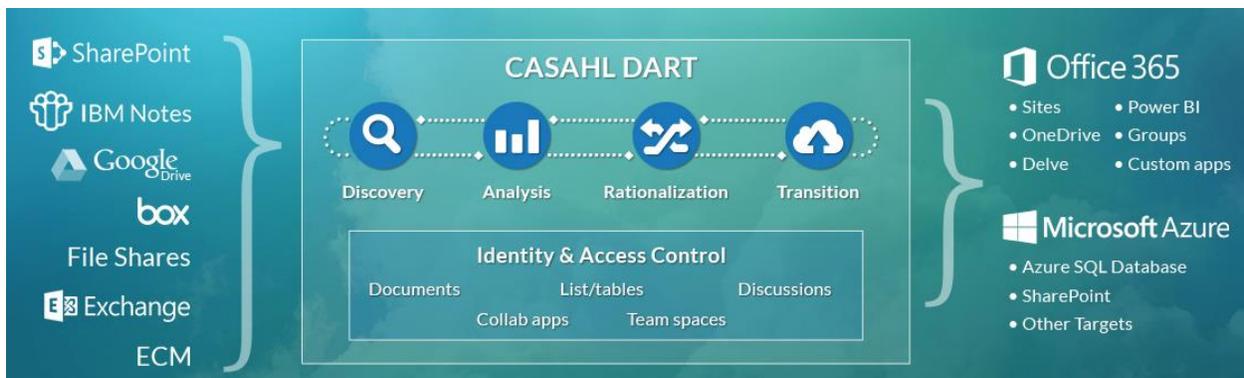


CASAHL Crawler and Analyzer for Enterprise Content

Enterprises need to know what essential data to modernize for a systems migration to be successful, but many businesses have so much data spread across multiple content systems that they aren't even sure what data they have. CASAHL's DART suite makes assessment of that data easier than ever so that you can get the most out of your migration.

The CASAHL Crawler for Enterprise Content, DART's Discovery component, is the first step toward making the best use of your SharePoint and Office 365 deployment. Crawler examines the client's content systems to discover sources of information, then collects key usage statistics from those sources. Once Crawler has finished compiling this information, it passes the data it discovered to the CASAHL Analyzer, the Analysis component of DART, which examines the data and provides detailed summary reports used to help make your migration as efficient as possible.



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 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, Exchange public folders, Google Drives, and CMIS-compliant systems, among others. There is also a specialized version of Crawler for IBM 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, among others.
- And 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 all of your enterprise’s IT staff.

The screenshot shows the Crawler application interface. At the top, there are buttons for 'Batch' and 'Log'. Below that are 'Refresh', 'Remove', and 'Clear All' buttons. The main part of the interface is a table with the following data:

Name	Target	Start Time	End Time	Folders	Files
Consulting	http://panama/Consulting	1/4/2012 7:45 PM	1/4/2012 7:48 PM	16	197
Development	http://panama/Development	1/4/2012 7:45 PM	1/4/2012 7:48 PM	63	356
Documentation	http://panama/Documentation	1/4/2012 7:48 PM	1/4/2012 7:48 PM	10	33
Human Resources	http://panama/HR	1/4/2012 7:48 PM	1/4/2012 7:48 PM	3	212
Marketing	http://panama/Marketing	1/4/2012 7:48 PM	1/4/2012 7:49 PM	25	204
Sales	http://panama/sales	1/4/2012 7:48 PM	1/4/2012 7:49 PM	0	3
Strategic Tasks	http://panama/StrategicTasks	1/4/2012 7:49 PM	1/4/2012 7:49 PM	0	7
Technical Services	http://panama/Training	1/4/2012 7:49 PM	1/4/2012 7:49 PM	0	5

Below the table is a log window showing the following text:

```

1/4/2012 7:45:14 PM =====
1/4/2012 7:45:14 PM Started Crawling...
1/4/2012 7:48:02 PM Total Folders Crawled: 63
1/4/2012 7:48:02 PM Total Files Crawled: 356
1/4/2012 7:48:02 PM Completed
1/4/2012 7:48:02 PM =====

```

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, Exchange public folders, Google Drives, and ECM 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 modernized SharePoint or Office 365 system, 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 with CASAHL’s [Automated Fixed-Fee Migration service](#). The same reports identify low-value resources, such as applications and sites that have low (or no) collaboration and high complexity. This allows transition planners to focus transition efforts on the most popular collaborative applications, to identify resources that can be archived or deleted, and to form preliminary transition plans and effort estimates using the data in the CASAHL Analyzer’s reports.

SharePoint

For SharePoint, Analyzer delivers a high-level executive summary in PowerPoint format 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 allow you to customize the information.

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	SharePoint sites are rated 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	The settings and properties for each site and sub-site are listed, as well as the permission specifications in the form of access control lists.
Lists	Data is collected on the settings, properties, templates, and object counts for all lists and libraries.
Folder Path Depths	A record of the amounts of nesting in and the layouts of file trees.
Solutions and Sandbox Solutions	Data is recorded on 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.

SharePoint Analysis: Reported Data (by site)

Content Types	A compiled list of the lists, alarms, custom SharePoint sites, workflow templates, web apps, and other content discovered by Crawler.
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	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.
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	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	Data on the web-centric uses of SharePoint are collected for both Pages and Web Parts.
Features	Data are collected on all SharePoint Features in the system, including their properties and the SharePoint items associated with each feature.

Google Drive

Analyzer delivers a PowerPoint high-level 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.

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	A list of author and editor activity is compiled for each Google Drive.
Usage / Age	A report on the date files were created and the modifications made to them over time.
File Properties, File Types, and Total Files	A compiled list of files, the properties of those files, the types of those files, and any instances of duplicate files.

Google Drive Analysis: Reported Data

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	All Office documents are listed 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, 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	A list of author and editor activity is compiled for each location in the system.
Usage / Age	A report on the date files were created and the modifications made to them over time.
File Properties, File Types, and Total Files	A compiled 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	All Office documents are listed in named groupings of file types, e.g., .docx, .doc, .xls, and .xlsx for Office file types, etc.

Once the CASAHL Analyzer has compiled its reports and decision-making among users, managers, decision-makers, and developers is under way, it's time to start the rationalization process. The next product in the DART suite, the DART Lifecycle Manager (DLM), takes the reports compiled by Analyzer and puts the information in a DART site, where business users and IT personnel across the enterprise use it to construct a migration plan. [Click here](#) to read more about the DART Lifecycle Manager.

About CASAHL

Since its inception in 1993, CASAHL has maintained an exclusive focus on providing enterprise-class solutions that optimize and modernize collaboration and content environments such as SharePoint, IBM Notes/Domino, Google Docs, Office 365, Enterprise Content Management, DBMS, and ERP systems. CASAHL has worked with more than 2,000 enterprise customers worldwide.

CASAHL won Microsoft Partner of the Year in 2009 and 2010 for Migration solutions. Platform vendors including IBM, Microsoft, and Sybase have OEMed products from CASAHL to optimize and modernize customer deployments for their respective platforms.

CASAHL's unique solution provides automatic discovery and analysis of existing systems, along with automated generation of transition, coexistence, and integration tasks among previously deployed environments and modern alternatives such as SharePoint and Office 365.



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