

CAS Data Light for Information Discovery

Content Analysis at *human speed*



PRODUCT HIGHLIGHTS

- An integrated platform for collection, identification, preservation, management, and eventual purge of your ESI.
- Included in the 2016 Gartner Magic Quadrant for Enterprise Information Archiving, and Market Guides for eDiscovery Solutions and File Analysis
- Use optional GPU hardware to speed analysis faster than 1,000x, enabling culldown in minutes, not days (pg. 4).
- Make fully interactive decisions faster than before with breakthrough performance on largest data stores.
- Make better decisions with higher confidence in correctness via an intuitive graphical interface powered by our proprietary machine-learning algorithms.
- Automatically OCR image files, including signature block detection to identify contracts and other priority documents.
- Data Light is optimized for various workflow skins including Information Governance, eDiscovery, Dark Data Cleanup and Compliance.
- Accesses data in hundreds of formats; Capax can also create custom connectors if needed.

CAS Data Light for Information Governance enables organizations to move and analyze data *from* anywhere *to* anywhere, with Data Light acting as the data police in between. Going to the cloud? Use Data Light to prevent confidential information from leaving your organization. Merging with another company? Use Data Light to analyze the available data before taking on someone else's risk. Client data strewn about? Compare to your client list and pull back anything sensitive to a known controlled location. Executing an official Information Governance initiative? Lack of comprehensive Information Governance controls leaves organizations struggling to be certain they can tell the courts that they've found "everything."

Data Light uses groundbreaking machine-learning algorithms that run on massively-parallel GPU hardware to enable organizations to locate, analyze and most importantly, take prompt action on their oceans of unstructured data.

Data Light powers an effective information governance strategy with a customizable dashboard that displays rule sets graphically so you see what data is affected by the rules you're building, either enterprise-wide or repository-by-repository. Trainable clustering technology speeds the identification of similar documents for defensibly applying retention and deletion policies *en masse*. Policies are actionable, following flexible workflows for proper data management, moving information to your location of record while applying appropriate retention rules. Consistent policy application with detailed reporting provides additional defensibility and confidence in your deletion and purging activities. Once a policy is applied, data can be purged or held in place, or be moved/copied to the cloud, to a more secure location like our CAS archive with its fine-tuned retention controls, or to any other location you need. You have complete control over data location and retention.

Corporations can't rely on IT to provide and maintain a patchwork constellation of different tools to compile responsive information. Outside service providers using proprietary tools may take too long and may require steep learning curves.

While IT and outside services can help, organizations must increasingly take ownership of ensuring that the tools and processes for eDiscovery will deliver success.

CAS Data Light for Information Governance makes you successful with features for:

- *Fast and complete analysis:* Scan metadata only for fast preliminary results. Apply deep indexing for more complete analysis, searching all content across hundreds of file types with minimal performance impact. Data Light even accesses Active Directory user information to identify original owners/creators.
- *De-duplication* of data, corporate-wide or within a single custodian.
- *“Find more like this:”* train Data Light using existing documents to find those that are conceptually similar (e.g. contracts).
- *Discover and cluster* similar documents together across the entire corpus without an example of a responsive document. Investigate individual clusters, and create sub-clusters for more granular analysis.
- *Classify* individual documents and clusters into relevant/not relevant buckets or other categories quickly via an intuitive interface.

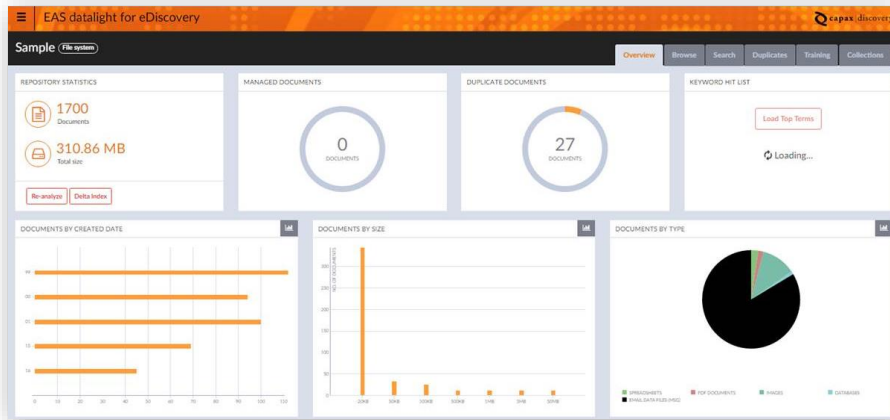


Figure 1: CAS Data Light Main Dashboard

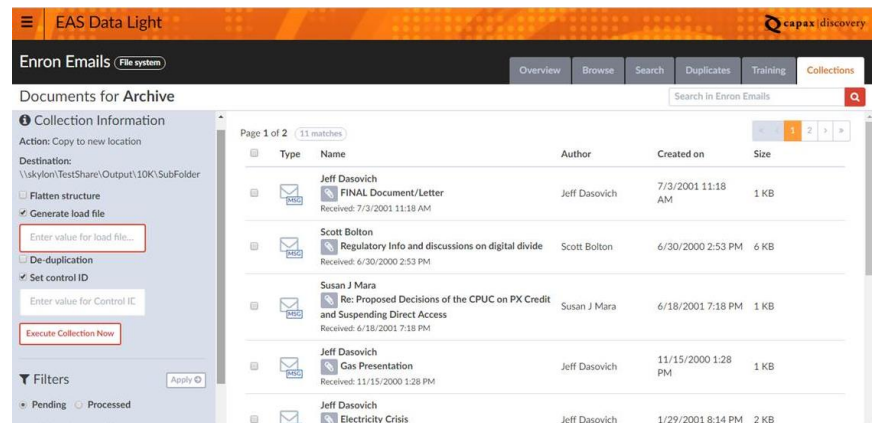


Figure 2: eDiscovery Collections & Export Screen



Figure 3: Cluster Map of Entire Collection

We've made strong performance claims for CAS Data Light, with tests showing faster than 1,000x performance improvements over software-only products. In this day and age, that's an unusually strong claim. And this speed is core to our ability to help customers take control of their data, by letting them try out proposed policy changes, seeing the impact of a change in seconds rather than hours. That's critical to both productivity and correctness of eDiscovery projects.

How do we deliver such a significant performance breakthrough?

GPU Hardware: Breakthrough Boost in Performance and Quality

CAS Data Light is the first in the industry to support GPU (Graphics Processing Unit) hardware, using it to speed index generation and applying the machine learning algorithms that cluster ingested documents. Recently, Capax Discovery's engineers saw the potential for emerging GPU hardware to perform dramatically, and began redesigning our products to support GPUs. The inexpensive GPU boards we support slide into industry standard servers, and can be easily installed in your own machines, or we can provide them as part of our cloud offerings.

With such an immense performance boost, Data Light increases the scope and accuracy of its operations, indexing 100% of documents and 100% of each document's content instead of relying on sampling of a small percentage of words within a subset of the document corpus. We thus deliver the highest-quality searches and help you make optimal retain/discard decisions while still running fast enough for the most demanding customers who need to handle multiple petabytes of data.

Data Light's unique ability to deliver higher quality results at breakthrough speeds means users can confidently take action on their data in real time. You save storage costs while being certain that you're making defensible retain/discard decisions.

What is GPU Technology?

Graphics Processing Units (GPUs) are specialized chips originally built to offload processing of graphics from the Central Processing Unit (CPU) of a computer system. GPUs were originally developed for high-end graphics workstations. Over the last 20 years, inexpensive GPU chips have become the heart of mass-market gaming consoles, enabling the creation of rich, immersive virtual environments.

Because graphical operations involve performing the same relatively simple calculation on millions of graphical elements, GPUs have thousands of processor cores whereas CPU chips have only handfuls. This means that, for the right application, inexpensive GPUs can realistically speed up application throughput by hundreds or thousands of times.

In the last few years, cutting-edge physics labs and other scientific users have adapted commodity GPUs to make the world's fastest supercomputers. Recently, general-purpose programming toolkits have emerged to simplify software development on GPU-based servers, enabling them to be used for general-purpose computing applications.

Because indexing and search involves performing the same relatively simple operations on potentially billions of documents, eDiscovery is a natural use of GPU technology.