



Intelligent Document Segregation & Migration

The Problem

A mid-sized investment bank had contracted for the sale of one of its internally developed, but peripheral, businesses. Ninety days before closing firm leadership realized that the business they were divesting held more than 5 terabytes of digital assets and a serious percentage of those assets were proprietary to their firm and could not be legally transferred to the acquirer. Discussions with the firm's IT outsourcer and multi-national consulting firms resulted in proposals requiring 6 months to more than a year of time to segregate and transfer the digital assets -- and with costs in the millions of dollars. As the acquirer could not continue to service the clients of the divested business without access to employee emails, financial models, legal documents, et al, the deal was at considerable risk.

The bank's outside counsel introduced the bank to CBIZ Technology (then Marcum Technology) a long-term OpStack partner who brought us in to assess the situation and propose a solution.

The challenges that we needed to overcome were:

- The entire project was schedule driven as we had fewer than 90 days until closing. Delaying the closing would have been at an unacceptable cost.
- Millions of documents needed to be inspected for content in order to determine if they stayed with the bank or were to be transferred with the business being acquired. We needed to inspect then segregate or transfer over 35 million email messages with attachments totaling petabytes in size as well as millions of files in both SharePoint and OneDrive. Quick checks with the two leading eDiscovery platforms found that setup and (most critically) data transfer of the files and messages to be indexed, searched, and segregated would require more than the 90-days available before the deal's closing date.
- There were only two senior officers of the client who had authorization to review the email and files for transfer, as of the start of the project they did not have either criteria or process for review defined. Optimizing the process so that manual reviews would not be a bottleneck was a critical success factor.
- The processing and review elements needed to segregate and transfer data would require multiple parallel workflows in order to complete the work within the project's tight timeline.

Key Success Factors

The OpStack team provided requirements for a full inventory of all the digital assets in scope and while that was collected by client's MSP, began a discussion with the client on the criteria that would need to be applied

to determine the future ownership of any given document or message. It was determined that a combination of external email addresses and a specific and limited set of keywords would be used to determine if any specific message or document needed to be retained and not transferred to the acquirer. In parallel, OpStack experimented with mail and file movement at volume to determine achievable rates given Microsoft's provision of bandwidth for data exfiltration.

With that information in hand the OpStack team determined that the solution required:

- Data transfer speeds were going to be the bottleneck on meeting deadlines. Movement of files from the client's Azure subscription be done only once, Microsoft throttling of data movement precluded making multiple move or copies of files.
- Search speed was going to be critical for determining the status of each document.
- The messages and documents were going to need to be indexed in place, there wasn't time to move them before indexing.
- No single data transfer tool had the flexibility needed to handle each of the content types that needed to be moved, multiple tools were going to need to be part of the workflow.
- With uncertainty as to the appropriate search criteria for separation of messages and documents, we would need an agile approach to creating and evolving lengthy and complex search routines.

The Solution

The OpStack team selected X1 Enterprise as the search and discovery solution. X1's combination of in-place search, centrally maintained indices, user-friendly review interface, and enterprise management console made it the heart of the information segregation process. This was complimented by best-in-class tools from both Microsoft and third party vendors

The OpStack team scripted an orchestrated workflow that allowed for the parallel execution of:

- indexing, done in tranches by time period, asset-type, and owner to feed the pipelines
- complex search using a constantly evolving scoping query, with software building queries optimized for the search engine
- Transfer of logical "search hits" for senior management and counsel review
- Post-review processing of hits to create tool-appropriate manifest files for segregation and deletion
- Execution of transfer jobs with monitoring -- the monitoring was critical as Microsoft rate restricted network throughput and would kill jobs randomly throughout the entire project.

Tuning of the solution continued throughout the project, optimizing the efficiency of indexing, of search, of deletion, and segregation. By the last month of the project, the performance choke point was Microsoft's undocumented algorithm-triggered actions to throttle I/O throughput on the client's subscription. Breaking operations into smaller than otherwise optimal batches and putting wait timers in some activities lessened the impact, particularly minimizing the notice-free killing of jobs.

The Result

- The project allowed the last two years of documents and email messages to be cleaned and transferred to the acquiring organization so that the business could restart in its new home on the first business day after closing.
- Additional assets that substantially extended the scope of the data to be cleaned and transferred were identified late in the project and were transferred in the month after closing.
- The entire project was completed in one quarter the time and one eighth of the cost that has been estimated by the one big-four firm that had bid the project.