

As methods of electronic communication continue to grow and expand, organizations have come to realize the importance of protecting saved data. To facilitate this, IT departments have implemented backup recovery solutions, usually in the form of tapes that are stored offsite for safekeeping. In the event of a disaster or system failure, IT usually can retrieve the tapes and restore the data fairly quickly and minimize loss of information. Over the years, organizations often amass large numbers of these stored tapes with all but the most current ones virtually useless for data restoration purposes.

Why doesn't the firm get rid of these tapes? Without being able to identify specific files, e-mail or documents, there really is no choice but to retain them in the event they might become discoverable in a lawsuit.

Contents of a Tape

At a basic level, backup tapes usually contain several kinds of data. This may include working documents (agreements, letters, memos, presentations, spreadsheets, etc.), e-mail, databases and system files (.exe, .dll, etc.). Communications-related data primarily consists of documents and e-mail. It is more likely to be the focus of e-discovery than transactional data or systems files.

At the typical document and e-mail file structure level, data is not very difficult to access. When backed up to tape, however, it's placed in a unique backup format, often compressed into a special container. The most common formats are LTO-2 and DLT. This special formatting makes it more challenging and complex to perform discovery as most companies do not have the appropriate hardware and software to access the data quickly.

Traditional Discovery Process for Tape Backups

Before discovery can begin, the data must be restored. This requires an inventory of tape contents as well as knowing what software (and what version) was used to back it up, what e-mail software was used, how much storage is required, etc. Tapes need to be cataloged and analyzed in order to gather this information. Adding a new dimension to the challenge, if an organization has gone through a number of mergers and acquisitions, it's possible for many flavors of backup software to be associated with the tapes that were acquired over the years.

Armed with the knowledge of how much storage is required to bring the data back online, as well as the backup software necessary, the data restoration process can begin. Restoring the data is the most time-consuming step of a traditional tape discovery process because of the sheer volume of information and the technical resources required to manage the project.

Once the data is back online, all keywords and metadata are indexed and made searchable. Speed is often an issue since the data has not been presorted and the volume of information is significant. Processing 1TB of data at 5MB/second would take about 56 hours, whereas processing at 50MB/second would get you to a much more reasonable six hours.

With the indexing complete, the actual discovery process can begin. First, system files and duplicates must be filtered out. Then, a query can be run to find responsive data. Once located, responsive data is then delivered to the legal team.

The problem with this traditional scenario is all the processing that must occur prior to running the query. It is very time-consuming and expensive. The cost to catalog a few hundred tapes, determine the contents and restore the data online could be in the millions of dollars and take many months.

At one time, these lengthy timelines and huge costs were considered an acceptable argument against the huge burden of having to produce the data. However, under the amended FRCP requirements, courts now are requiring organizations to shoulder these burdens. Legal teams must find a way to discover tape content quickly enough to please the courts, and yet remain affordable enough to not cripple their clients.

A Better Way

New technology has eliminated the need to restore tape contents completely. This can save 50 to 70 percent of the time and cost versus traditional methods. Indexing data directly from tape allows for discovery to occur quickly. In fact, the tape discovery process is completely flipped. Rather than restoring everything before locating responsive data, you can discover first then restore only what you need. If, on average, less than one percent of data on a backup tape is responsive to discovery requests, then only one percent of tape data needs to be extracted, ignoring the other 99 percent of irrelevant data.

This new approach to e-discovery of backup tape content allows archived tapes to be indexed without bulk restoration and without recreating legacy backup and application environments. It is a methodology that saves time, money and ensures compliance to electronic data handling regulations. Courts are now aware of this new, more expedient approach to tape discovery. Before you approach the bench with a burden argument in hand, be sure you're aware as well.

About our author :: :: ::

Jim McGann serves as Vice President of Marketing for Index Engines. Jim has extensive experience with the sales and marketing of enterprise software to the Fortune 2000. Prior to Index Engines, he was responsible for the sales and marketing of Scopeware at Mirror Worlds Technologies. Jim can be reached at jim.mcgann@indexengines.com.