



White Paper

Make a Molehill out of a Mountain

Remediate Backup Tapes to Limit Liability While Reducing Storage Costs

Author – Jim McGann, Index Engines, VP of Information Discovery

Introduction

“How many backup tapes do you have?”

“I have no idea – probably thousands.”

“Do you need to keep them?”

“No.”

“Why don’t you recycle them?”

“Legal won’t let us.”

This might be a typical Storage Manager’s response when questioned about a company’s backup tape stockpile. These tapes are often created in response to a key objective of any IT organization - to protect enterprise data assets. Thus a mountain of old backup tapes has been amassed, largely of tapes that have long lived out their disaster recovery purpose. Why not recycle, or destroy all these old tapes? Federal regulations forbid it. Data on these tapes “MAY” be necessary to support current or future litigation. What data? A very, very small percent of what exists, typically less than 1%. Why then keep all these tapes? Because it has been next to impossible to separate out the useless data from what legal requires.

Sometime down the road, if not already, specific data from backup tapes will be requested by legal. Some corporate legal teams have issued a mandate to not touch tapes; others have been forced to do so. Either way stricter regulations are forcing the issue. The June 2009 California Electronic Discovery Act, for example, declares all electronically stored information is accessible, and requires it to be produced. In January 2010 Judge Scheindlin, the judge on the groundbreaking Zubulake case, issued an opinion where she denied the use of the burdensome argument and called out the plaintiffs as grossly negligent and issued sanctions for not collecting data from backup tapes to support the case. The courts are ruling more and more against firms that do not produce data, including tape data, in a timely manner. Many cases exist today where fines were imposed against botched collection of historical files and email. Will your firm be next?

Storing old tapes is not only a potential liability, but also a wasted expense. Even if it costs only a few dollars a month to store a tape - it adds up. Additionally, since these old tapes can’t be recycled, new tapes must be purchased for ongoing tape backups. This expense alongside the storage costs quickly becomes a large item in the budget. This IT expense could easily be allocated to something more useful for the organization. This paper discusses how to take a mountain of stored tapes and turn them into a molehill by extracting the relevant data and eliminating unnecessary tapes.

Consider Remediation

In the past it was far too expensive and difficult to understand the detailed content of old backup tapes. The content would need to be first restored, and then analyzed in order to determine what to keep and what can be purged. Restoring thousands or tens of thousands of tapes was out of the question. It would take too much time, money and legacy infrastructure. As a result IT departments have let the mountain of tapes grow taller every day – with no end in sight.

The problem has now been solved by eliminating the need for expensive and time consuming backup restoration and applying a more intelligent approach. New technology scans tapes and then searches and extracts specific files and email without the original backup software. This allows you to only deal with relevant files (less than 1% of the tape content) and not the bulk of useless content. Index Engines intelligent tape discovery solution has made tape remediation an achievable project. In significantly less time, an IT department can process tapes in house, find what legal needs, archive it and make it available when it is needed. This efficient, cost effective process enables tape remediation allowing IT departments to recapture tape storage budgets, while supporting legal with the data they need.

An Alternative: Automated Tape Processing

The new automated process is simple - no specialized skills or software are required. Assume a situation with 10,000 tapes in offsite storage. The first step would be to catalog the tapes to profile the content. Using a tape library, tape headers can be scanned in minutes, only requiring manpower to load the tapes. Once the scan is complete, analyze the catalog and eliminate incremental backups, as well as backups of non-user data servers and blank tapes. This typically reduces the volume by 80%, turning a 10,000 tape job into a 2,000 tape job. Stopping here eliminates 80% of the tapes and achieves significant cost savings.

Once the cataloging is done, the remaining set of tapes contains potentially responsive data that will support current and future litigation. The next step requires a full scan of the tapes. This generates a searchable index of the content and metadata without copying or modifying the existing tapes. Collaborating with legal, the search queries are defined (the management team’s email, files related to a sensitive project, intellectual property documents, etc.). Legal can then search the index, tag what they want, and request the data be extracted. IT will then run an extract job and all the tagged files and email will be ripped from tape, keeping all the content and metadata intact. When this process is complete the tapes can then be recycled. Read Index Engines “*The Anatomy of Tape*” white paper for full details on a tape remediation project.

Visual Comparison

Details of a typical Index Engines installation with 10,000 tapes are as follows:

Tape Profile:	10,000 LTO2 Tapes (a combination of NetBackup, ArcServe and BackupExec)
Hardware Environment:	Index Engines Appliance, PC with browser access, Tape Library (3 active drives and capacity for 50 cartridges in this example).
Phase 1: Cataloging:	All 10,000 tapes are scanned; one/two minutes per tape. Total cataloging time 26 days (24x7 w/ 10% downtime) or 381 tapes/day.
Phase 2: First Pass Culling:	Analyzing the catalog identifies tapes to be eliminated from further processing: blanks, backed ups from non-user data servers, and incremental backups. This is typically 80% of the total tapes; 8,000 are eliminated, leaving 2,000 tapes for full indexing.

Phase 3: Indexing:	Full scan of the 2,000 remaining tapes (approx 400TB of data). The index contains the complete text and metadata of the content. Total indexing time 17 weeks. The total index size, typically 4 to 8% of the data, is approximately 16-32 TB.
Phase 4: Discovery:	Using the browser based interface, legal can then search the index generated from the tape content and find what they need. The relevant content (typically 1% or less) will be tagged, ~4 TB
Phase 5: Extraction:	The tagged data is extracted. This is an automated process that does not require the original backup software, just the appropriate tape be loaded in the library. Total extraction time ~18hours
Phase 6: Recycle Tapes:	Once relevant content is extracted and delivered to legal, tapes can be recycled.
Total Time:	Cataloging (3.7 weeks) + Indexing (17 weeks) + Extraction (.2 weeks) = 20.9 weeks
Tape Storage Costs:	Approximately \$30/tape/year. For 10,000 tapes the storage costs equals \$300,000 annually.
Costs to Purchase New Tapes:	\$100 per tape at 25 per week would cost \$130,000 annually

If you combine the cost to store tapes offsite with the cost to acquire new tapes in support of the existing backup process it equals \$430,000 per year. As the volume of tapes is growing each week, this number will continue to increase over time. In order to compute the payback for such a project you would need to break out the cost to acquire the Index Engines product, the dedicated tape library as well as the manpower costs. The expenditure for manpower, tape libraries, hardware, and software will prove out an ROI in less than one year. This does not include any costs associated with ongoing litigation where tapes are pulled from storage for restoration. These litigation support costs could easily reach hundreds of thousands of dollars annually, which would result in a shorter period of ROI.

Conclusion

In the past it was not cost effective to remediate the mountains of tape stored offsite. New technology now makes this feasible, and is quickly becoming a best practice for any organization that is faced with constant legal events. As legal and IT work together, tape remediation is quickly becoming the preferred method to reduce corporate liability, and expand IT's ever-shrinking budget.