Overview of legacy tape based backup data migration options commonly used by organizations of any size.

ABSTRACT
This whitepaper reviews the considerations, options and best practices for migrating legacy backup data being held on tape media.

March 2015
# TABLE OF CONTENTS

**EXECUTIVE SUMMARY** .......................................................................................................................... 4  
AUDIENCE...................................................................................................................................................... 4

**LEGACY BACKUP DATA DISPOSITION & MIGRATION OPTIONS** ................................. 4
1. Do Nothing, Allow the Data to Expire in its current format and location: ............ 4
2. Shut down the legacy backup software: ................................................................. 4
3. Keep a small instance of the Legacy backup infrastructure:................................. 4
4. Deploy a Legacy Backup Data Management and Migration Tool ....................... 4
5. Backup Data Conversion and Re-ingest .............................................................. 5

**DO YOU NEED A BACKUP DATA MIGRATION PLAN?** ........................................ 5
You answered “YES” to Question 1, 2 or both: ................................................................. 5
You answered “YES” to Question 3, 4 or both: ................................................................. 5

**MORE INDEX ENGINES OPTIONS AND CONSIDERATIONS** .............................. 7

**FINAL THOUGHTS** ................................................................................................................................. 9
EXECUTIVE SUMMARY
This whitepaper provides an overview to what the typical data migration options are for dealing with the backup data being held on tape media; once a backup software migration has occurred. This whitepaper is intended to help clarify what the considerations are when choosing a data migration option and to provide enough detail so that organizations can make an informed decision to meet their needs.

AUDIENCE
This white paper is intended for guidance and understanding of the common options for dealing with legacy backup data residing on backup tapes. This document is suited for information technology workers, compliance, risk mitigation or legal personnel who are looking to understand the basics around what to do with any existing tape based backup data when you switch backup software applications or inherit tapes that are not in current production format.

LEGACY BACKUP DATA DISPOSITION & MIGRATION OPTIONS
1. Do Nothing, Allow the Data to Expire in its current format and location:
   - Pick a point in time to begin backing up the entire infrastructure with the new backup software package. Typically used if your data has a short retention of say a year or less.

2. Shut down the legacy backup software:
   - Keep a record of the tape date ranges.
   - If requested, produce the tapes to Legal or outside party. If further details are needed from these legacy tapes in small batches send tapes off site to be processed by a 3rd party service that offers tape processing.

3. Keep a small instance of the Legacy backup infrastructure: (Very Common)
   - This instance can be as small as a single Media Server attached to a single tape drive with the data center manpower to staff restores. Great for short term and light use, typically within 3 to 5 years post data creation, as after 5 years, the expertise and ability to restore old tapes even using the original infrastructure becomes untenable due to the deterioration of the tape media.

4. Deploy a Legacy Backup Data Management and Migration Tool (Best Practice for Legal Hold or Highly regulated industries that require legacy data preservation)
   - At EMC we partner with Index Engines to resell their software offering combined with either EMC or Partner Professional services for this solution offering.
   - First step is to ingest the Legacy Backup Catalog(s) and retire legacy Backup Software package including maintenance (optional). All of the backup data remains searchable, just as before. In addition, it is possible to consolidate multiple backup catalogs onto one delivering a unified catalog search.
   - Keep catalogs alive within Index Engines, participate in the EMC Data Migration work shop to find out the specifics on what data is a good target for migration or alternatively just keep the Index Engines server up and running with your legacy backup catalog ready and waiting for when someone needs a specific tape or data set restored. Index Engines does not need the legacy backup software to restore data from tape.
   - Migrate business critical data to a disk based archive for simplified retention and access in support of regulatory and compliance requirements. Typical industries where this is utilized are Banking, Oil & Gas, Government, Healthcare or Organizations with legal holds or long term backup retention requirements. Run the analysis to find out how much data would need to be moved, excluding things like differentials and incremental backups, application backups, data outside retention periods, etc. Review and decide the best course of action. In most cases 5-10% of the total data on tape media truly needs to be migrated to either legal hold repository or archive. Create a defensible data migration work flow and execute the process. Once data of value is migrated tapes can be shredded, offsite storage costs recouped, and risk of unmanaged tape data mitigated.
5. Backup Data Conversion and Re-ingest (Case by Case, not best practice for large conversions)
   - Customer personnel or EMC Professional Services would recover legacy data to a landing zone from tape or disk, then re-backup or re-archive/index it with the appropriate EMC Data Protection Suite software solution to meet the customer's future retention and recovery needs.
   - Most economical if the data sets are straightforward and the amount of data is relatively small.

DO YOU NEED A BACKUP DATA MIGRATION PLAN?

Let’s assume you have decided or are evaluating changing your existing backup software application for a new one or have a number of non-production backup environments you are maintaining. One of the most common questions that come up during this process is: what do I do with all my legacy tape data? Most industry folks agree that there are several questions that first need to be asked before the proper strategy can be selected. The following provides some guiding questions:

1. Does any of your backup data (tape, disk or other) have retention of greater than 12 months today?
2. Are you holding on to all your tapes forever because that’s what the current policy states?
3. Are you in a highly regulated industry like Healthcare, Oil and Gas, Government, Banking or Financial Services where there are strict compliance and retention requirements for certain data types?
4. Are you currently preserving large numbers of tapes for legal hold for an active or pending legal matter or continually receiving restore requests from legal and compliance to retrieve tape data onto legal hold?

If you answered no to all of these questions you are probably okay with letting the current data age off while you migrate to your new backup software application. Depending on the size and complexity of your environment and the speed of your roll out plan; will determine how you approach the matter. That said, if you said no to all of these questions you will typically turn off your existing software package or scale it down to a single server for recovery purposes only until the retention period has passed. At that point most customers decommission the old software and move on.

If you answered yes to any of the questions, then further discovery is needed to make an informed decision. As described in the brief overview of the common options above; we will now outline some more data points for the yes answers to the above questions.

You answered “YES” to Question 1, 2 or both:
Then you should consider migration options 1-3 as described above. If you have backup data that is being held longer than 12 months the most common option used widely is Migration Option #3 (Keep a scaled back single server and tape drive available Approach).

   - If you have a significant number of tapes 10K, 20K or 30K+ plus and the cost of storing them is significant then you can consider using Index Engines to migrate just the records that are business critical and eliminate all the excess as described in Migration Option #4 and below (Use Mitigation/Migration SW).
   - If the data amount being held is small like less than 50TBs in total, then you can consider / weigh the costs of doing nothing or recovering the data and re-backing it up with your new backup Software application as described in migration option #5 below. This is rarely done at scale and can be done by the end user or contracted to be done by EMC or Partner professional services.

You answered “YES” to Question 3, 4 or both:
If you have these sorts of considerations going on in your organization then yes you need a migration plan. The good news is EMC and its partner Index Engines have several solutions to consider. The objective is to find a solution that meets both your risk mitigation and time horizon needs while also reducing costs over alternative strategies. If you are in a highly regulated industry then it really comes down to the risk and liability associated with using tape as an archive and how often you are asked for legacy data retrieval and what the costs are...
associated with that data being inaccessible, slow to recover or unrecoverable. Couple those considerations with any fines or penalties that you might incur if you can’t provide the data requested in the time line required and you can produce a TCO/ROI decision matrix for the migration and disposition of data using a 3rd party software or service.

The following provides additional details to the options that can be uses to address data migration needs in legal hold or highly regulated industries.

- **Become an IT Museum:** Either have the backup software running on a server or just keep a copy of it that you can spin up. This works okay if the request are very seldom and they happen within a few years of being created. The real problem is that this type of approach means that if the data is to be maintained for a long period of time you kind of have to get into the IT museum business. By the way, this is true even if you don’t change your backup software application since the very old data being requested was backed up by systems that don’t exist anymore and tape libraries that have long since stopped functioning. Many organizations struggle to restore data from tape that is aged enough so that the servers, backup software and hardware has changed so much so that getting data from tape is a painful process.

You have to keep a server that will run the software you backed up the data with and a tape library and tape drive able to read the tapes you have. This is the unspoken truth about tapes, eventually they stop making drives in a particular format. Most tapes drives are backward competitive 2 versions. For example: if you had a LTO1 tape that could be read in a LTO1 Drive, LTO2 Drive or LTO3 Drive but not in an LTO4 Drive. So periodically you have to go through a tape migration process to migrate all the old tape format data over to a current tape format; so that if you need to read it you have a tape format that is readable in the drive format you own.

When people say tape is inexpensive they are referring to the cost of making a single copy of your data on a tape and parking it in a box. There is no RAID on tape, so best practice (not widely followed) is to have at least two copies of your data on tape as well. I think the statement should go on to say, tape is inexpensive but the process of maintaining recoverability of your data from tape makes it more expensive and time consuming as time goes on and the amount of data that needs converted periodically grows exponentially.

It’s also not talked about much, but any application data also requires you can re-create the exact environment by which the backups were created. This level of detail goes down the same SQL or Exchange server version, OS version, Active Directory version etc. So you can see after a while this will become less effective to the point of not effective at all many years from now. We have seen it take many months for clients to rebuild the appropriate backup infrastructure and restore environment to retrieve data from aged tapes.

- **Pay a 3rd Party for Tape Processing:** There are also some service providers and eDiscovery services who own software which will recover your files, email and other types of data from legacy tapes from a multitude of software vendors. They will do recovery on a per tape basis to a hard drive and mail that back to you. This works well for request that are small or happen infrequently but again can get pricy at scale or when the requests come often.

- **Leverage a Tape Data Migration Software Package:** Typically used in highly regulated industries or where customers have legal holds and lots of tapes, this software can be very effective. At EMC we have chosen to partner with Index Engines for these needs. The real advantage of using a solution such as Index Engines is it accomplishes two keys goals with a single solution. 1.) It allows you to ingest the
existing backup catalog from products like Symantec’s NetBackup, IBM’s TSM and CommVault applications so you can turn off the legacy backup software package completely while still having full access to the legacy data. Additionally for those organizations with NDMP backups their NetApp server environment can be retired as well. 2.) It allows you to scan the catalog and tapes to quickly determine what your data profile is. i.e. You have 10,000 tapes of which 4000 are incremental, 1000 contain emails, 2000 are applications like SQL and Oracle, and the rest are file data, etc. etc. You can then use this assessment to create a scope of work and TCO/ROI cost analysis to migrate just the data you care about, like just the email and file data. Most common in legal hold situations, are scenarios where the legal folks are after preservation of selective data types like email from selective individuals over a specific time period or just data from specific servers or NAS shares created by specific individuals. So the use of this software could allow you to effectively migrate just 5% or less of the overall data profile and create a single instance storage repository on disk in an ESI format and eliminate the legacy tapes along with the expense of maintaining them. The TCO/ROI vs. the alternatives is a simple process to work out so if you are in these scenarios work with your EMC representative to run the numbers and don’t forget to ask about our Index Engines starter pack and EMC Data Migration Professional Services workshop.

About Index Engines

Index Engines is a software solution that streamlines access to and management of enterprise data, including network file and email servers, backup catalogs, tape and disk images. Whether looking to assess, migrate and govern user data or for simplified access to legacy tape data and the retirement of backup catalogs, Index Engines has a solution.

Index Engines provides unprecedented file-level knowledge to manage the growing costs and risks associated with unstructured user data. Delivering the fastest indexing platform on the market today, at speeds up to 1 terabyte per hour per node, detailed metadata is captured across all data environments from primary storage to legacy backup tapes and disk.

Using this knowledge, information can be classified and dynamic reports created to analyze and identify content based on legal and data policies. Index Engines’ platform supports the disposition of data from migration, defensible deletion, tiering and archiving to policy audits and automation of governance rules. Join thousands of Index Engines users to conquer unstructured data growth, support legal and eDiscovery requests, and streamline the data center for more efficient operations.

MORE INDEX ENGINES OPTIONS AND CONSIDERATIONS

Making Sense of Legacy Backup Data

The process of determining what content has value and what is redundant, outdated and trivial is not as complex as you think. If you can develop a policy of what should be preserved, which requires input from the legal and records management team, this could result in restoring less than 1% of the legacy backup content. Even if you haven’t determined the policy of what to keep, a single instance of legacy tape data can be restored to disk and then retention policies can be defined.

The right scenario is dependent on how sound your data retention policy is. Some organizations can be very specific as to what should be preserved (based on content type, owner, and date range), others may not have a detailed policy or a “save everything” policy. Either way a solution exists to migrate and secure data of value online and eliminate the use of tape as a long-term archive.

Legacy Backup Data Migration Scenarios

Here are typical legacy backup data migration policies that allow intelligent reporting and access to backup images in order to streamline the migration of content that has value to the business:
o Single Instance of all Data: For organizations that cannot determine what should be preserved and what
no longer has value, migration of a single instance of legacy backup data from tape or disk into an
accessible and manageable online archive is the solution. Data can be classified into different categories,
each with different retention policies. This allows for legal and records to manage the data going forward
where they can determine retention periods and purge what is no longer required. For IT organizations
this represents a savings in offsite tape storage as tapes can be eliminated once the migration is
complete. This also saves ongoing tape restoration costs and provides more efficient support for
eDiscovery and compliance requirements.

o Single Instance of Email or Specific File Types: Many organizations are only concerned with legacy email
or a specific file type (i.e. PDF, Excel, etc.) as this content contains important records or sensitive
communications that must be preserved. Preserving a single instance of email or specific files from legacy
backups is a much smaller subset and simplifies the migration process, especially if you can define a date
range and not extract data that has outlived its retention requirements. This data can then be managed
according to existing legal hold and retention polices and content that no longer has value can be purged.

o Selective Culled Dataset: The most efficient method of migrating data from backup images is using a
culled dataset. A culled dataset is based on what is required for long-term preservation. If legal and
records have a defined policy as to what is required for legal hold, compliance and other regulatory
requirements, this criteria can be built into the migration strategy and only this content can be restored
and persevered. This could represent less than 1% of the tape content.

Legacy Catalog Management and Reporting

Once a strategy is defined the migration process can begin. The first phase of this process is the ingestion of the
backup catalog which significantly streamlines the process. The catalog is the key to providing knowledge and
access to the legacy tape data. Once the catalog is ingested the legacy backup software can be retired and
eliminated from the data center as it will no longer be required for data access and restoration.

Ingestion of the catalog allows for all the metadata to be indexed including the backup policies. An assessment of
the backup content can then be performed in order to further define the migration strategy.

   o Typical catalog reports that can be analyzed include:
      o Incremental vs full backups
      o Backup Data Retention Policies (vendor specific)
      o Host Reports
      o Backup Set Report
      o Backup Dates Report
      o File Accessed/Modified Dates (not available for TSM)
      o File Size/Capacity Report
      o Backup Date Report
      o File Extension Report

By utilizing these reports a detailed analysis of the tape data can be executed and a detailed disposition strategy
defined. Some of this content may be outside any retention periods, some may be a file type with no long-term
preservation value (databases, log files, etc.), some may exist on hosts or servers that have no sensitive content
that requires archiving. Disposition strategies can include migration to cloud sources, archives, network storage
and more.

Additionally, catalog reports and analysis can be used as a communication and collaboration tool with the business
user who is charged with defining data policy. With the intelligence the reports deliver the data can be easily
evaluated and policies developed based on facts versus the unknown.
Justifying Legacy Data Migration

Backup data migration from legacy tapes or backup images is not for everyone. Organizations that never restore files and email from tape or disk backups and those that have inexpensive offsite storage costs may not see value in the migration of data. However, many organizations are looking to trim ongoing data center costs and are continually spending resources and budgets restoring data from backup archives. These organizations will find value in legacy data migration.

Here are some of the key advantages associated with backup data migration:

- Retire non-production backup software: Many organizations maintain legacy backup software as a result of a migration to a new vendor, or a merger/acquisition. Once data is migrated from legacy tape there is no need to maintain this environment. This will result in cost savings through the retirement of the backup software maintenance as well as data center resources including the man power required to manage the legacy software and antiquated data restoration environments.

- Migrate to a new backup software vendor: Freeing your organization from a locked in backup provider allows for implementation of a best-in-breed platform. This new platform may provide better functionality, support or simply integrate in your environment more seamlessly. Whatever the reason a key benefit of legacy data migration is allowing your organization to make the choice that makes the most sense for it rather than remain stuck with a less than adequate solution.

- Legacy backup tape expense: As tapes pile up in offsite vaults so do the invoices from the storage vendor. Storing a single tape can be inexpensive, however, storing tens or hundreds of thousands of tapes can be significant. Additionally, the management fees associated with transporting tapes to and from offsite storage add to the overall costs. Going tapeless once data of value has been archived online will recoup an ongoing expense that is not only wasted but can be applied to other projects.

- New tape purchases: Hording tapes in storage means you can’t recycle them and must continually purchase tapes for ongoing backups. As these tapes add up, and you continue to expand tape storage capacity, the annual budget grows beyond a small line item on your budget. Recycling tapes, once they are eliminated as an archive, will add efficiency to the data center and control ongoing costs.

- Cost of restoration services to support legal requests: Many organizations outsource the restoration of tape data to their law firm or a specialized service provider. These fees may be a part of a legal or operations budget; however, the costs could be significant to the organization. Migrating data of value including what is required for legal hold and preservation would eliminate these expenses and allow internal users to access and manage this data. Additionally, making sensitive data more accessible in disk archives accelerates time-to-data in support of legal and compliance efforts.

- Long-term risk and liability: Potentially the most significant expense and pain associated with stockpiling legacy data on tape archives is the risk of the unknown. Tapes are archives of all user data, email from the CEO, and contracts from legal and manufacturing documents from R&D. Highly sensitive records that are not managed and could be requested to support litigation and regulatory compliance. These potential “smoking guns” could cost your organization millions in fines and public embarrassment. Managing this content will allow mitigation of risk and control of future potential expense.

FINAL THOUGHTS

No matter what your data migration needs are your EMC Sales and EMC Partner Sales teams have been trained on the different data migration options and can help you quickly determine what course of action or combination of options is the right fit for you and your organization.