



Archive Migration Best Practices: Planning for Success

A Kanatek-Globanet Whitepaper

August, 2014



Table of Contents

| | |
|--|----|
| Introduction and Summary | 3 |
| Planning: The Key to a Successful Migration | 3 |
| Migration Sizing | 5 |
| Data Volumes | 5 |
| Item Counts | 6 |
| Effective Performance | 7 |
| Calculating Migration Times | 7 |
| Migration Batching | 8 |
| Effective Error Handling | 10 |
| Processing Errors: The Decision Process | 11 |
| Disaster Recover Planning | 12 |
| Chain of Custody Management and Verification | 13 |
| Chain of Custody | 14 |
| Verification | 14 |
| Winding Down Migrations | 14 |
| Kanatek Archive Data Migration Service | 16 |
| About Kanatek | 16 |
| About Globanet | 16 |



Introduction

Data migrations can be long complex projects and once were not for the faint of heart. Today, however, next generation migration tools and services for email archive migration have come on the market and significantly reduced the difficulty, the time and the cost. The result is that data migration is now being seen as a worthy investment, an opportunity to streamline archives, reduce storage costs and improve efficiency. Where organizations once avoided data migration projects, they are now looking at the best email archive platform for their needs, how a migration project can leave them in a better position and the right partners to help them get there.

This whitepaper will help you understand today's best practices for migrating email archives, providing guidance from two of the industry's leading migration experts – Kanatek and Globanet – on the key considerations, including:

- **Planning:** setting expectations around the migration process and helping to size the migration.
- **Performance:** setting performance expectations, looking at metrics and how to calculate migration durations.
- **Error Handling:** making fast decisions about how to handle errors, knowing how many times to retry errors, and putting a clear error handling process in place.
- **Chain of Custody and Verification:** reporting and making sure that the items have been cleanly migrated from source to target and without any changes or corruption.
- **Winding Down:** cleanly finishing the migration and removing the migration products from the environment.



Planning: The Key to a Successful Migration

Before starting a migration, it is important first to plan clearly how the migration will proceed from start to finish. Well thought out plans are major contributors to successful migrations and answer the following questions:

PHASES OF MIGRATION

Have you considered all the phases that go into a migration? Migrations typically have several phases:

- **Planning:** writing the migration plan and establishing how your migration will run.
- **Pilot:** running a set of archives (usually 10-15) to get a speed estimation for the migration and to validate the migration runs correctly and without error, as well as the effect, if any, on the production environment.
- **Full Migration:** running a first pass of all archives for migration, attempting every item in scope (which includes processes like Shortcut Conversion).
- **Error Resolution:** resolving errors, a phase which usually runs almost in parallel to the Full Migration phase.
- **Reporting:** running all final reports for the migration and finalizing the migration process.
- **Winding-up:** Removing the migration tool, backing up the database and decommissioning the source system.

SIZING INFORMATION

How large is your migration? How much data actually needs to be migrated and what size is estimated to end up on disk in the target at the end of the process? See the “Migration Sizing” section below for more information about single and multiple-instance storage techniques.



PERFORMANCE

What are the best, worst and average migrations speeds? Start by finding out how many items exist in total (see “Migration Sizing”), and then run a small pilot migration of 10-15 archives to the target. This will give you your expected “average” rate and will also take into account your worst-case and best-case speeds. Migrations vary significantly over the course of time, so calculating the maximum time your migration can take – the worst-case speed – is important.

PILOT GROUPS

How many pilot groups will I need? Pilots are always a good idea, even more so if this is your first migration. Pilots will help you identify potential issues and allow you plan in advance. As a pilot, migrating shared archives or legacy users who have left the organization or project repositories that are no longer used can provide a good sample set of data to test how the migration process works, including validating verifications, generating reports, establishing average speeds and understanding performance and error rates. This is a best practice rule – without running a pilot, never assume how fast or how well the migration will go!

ERROR MANAGEMENT

How will you handle errors so that they don’t bring everything to a halt each time? Effectively handling errors and making fast decisions about errors keeps migration performance and efficiency high. Poorly performing migration projects are usually stopped very frequently to resolve or investigate errors, harming the entire concept of continual progress. Ensure that you assess probable error handling time requirements and establish an error decision process before starting.

DISASTER RECOVERY TECHNIQUES

How will you be ready in case disaster strikes? Making sure your migration is accessible and continues in the event of hardware or serious environmental failure is critical. Don’t overlook this important step in your planning. See the “Disaster Recovery” section below more information and for best practices.



REPORTING FREQUENCY

How often are you going to report to your stakeholders? Which stakeholders get which report? How are those reports delivered to those stakeholders? A best practice plan includes a matrix of who gets which migration reports – and how often.

USER INVOLVEMENT

How will you communicate with users and how could they be affected by the migration? Any migration affects end users in some form, whether from components like email shortcuts or links in SharePoint, or from the source system impacting the user experience because it is being placed under load by the migration process. Best practice is to have a plan to communicate with users frequently, and especially an email or other notification to let them know when the migration of their archive starts and finishes!

MIGRATION PARTNER AND VENDOR COMMUNICATIONS

How will you communicate with and manage the partners and vendors helping you with the migration? Establish a good support channel and partnership with your migration vendor or partner, and understand how to contact them for support, when they are available (including 24/7 optionally) and what the Service Level Agreements are in place for support turnaround and diagnostics.

Migration Sizing

Sizing your migration appropriately during the planning stage is essential to calculating an estimate of how long the migration will take.

DATA VOLUMES

Understanding and planning data volumes in a migration might seem trivial at first, but can become a complex task when one starts assessing the storage and archiving requirements of the target servers, service or archiving system. When an organization is asked “how much data do you



have in your archiving system?” that organization will normally produce a number that can be based on a number of different things, specifically:

- the compressed, single-instance size of the data stored by the system on disk (SID – Single Instance on Disk).
- the uncompressed, multi-instance data that was archived by the system (MIS – Multi-Instance at Source).
- the compressed, multi-instance data stored in each archive (MIA – Multi-Instance in Archives).

Which figure is used during the planning stage is critical, particularly when migrating between Multi-Instance systems (such as Microsoft Office 365) and Single-Instance systems such as Symantec Enterprise Vault. The more efficient Symantec Enterprise Vault will store the data in Single Instance Format, and therefore an archive on Enterprise Vault that is only 1 TB will become, taking into account decompression and fan-out factors, may be 3 TB (MIA) of data in Microsoft Office 365.

Clearly some systems – such as PST file format, EML and MSG file formats – have no concept of SID storage since they don’t apply to single instance. When planning a migration from or to these systems, remember that only MIS and MIA figures are available (and in the case of the formats specified above, are usually the same subject to PST compression for PST files).

Given the huge differences between systems, understanding the differences in volume figures is essential to making sure that your plan ensures an effective and efficient migration.

ITEM COUNTS

Item counts are the best practice and most effective way to estimate the speed and length of a migration. That is why it is critical to get the item count for a migration as quickly as possible before starting the project because it is the most effective way to assess progress and speed.

Using data volumes (rather than item count) to assess migration sizing introduces some significant variances caused by the storage method of



source or target systems. For example, the size of the item in the source may or may not include special representations, indices and formatting that the target system uses. Moreover, the form and structure of the metadata will be different.

The bad news is that getting the item count from the source archiving or hosted system can often be non-trivial, similar to establishing data volumes. However, there are two options to achieve a good count:

- **Multi-instance count (MIC)** – the count of items in each archive, summed to a total count of all items.
- **Singe-instance count (SIC)** – the count of items actually stored on disk by the source archiving system or service.

The relationship between MIC and SIC is usually the approximate fan-out factor, that is, the number of average recipients receiving an item. This is not a strict rule however, as other factors can and will affect how a system stores and single instances items. Establishing SIC and MIC counts is a perfect example of where a migration service vendor can be of real assistance.

The MIC is the necessary figure to plan your migration successfully. Many common and popular archiving systems out on the market (Office 365, Symantec Enterprise Vault, Microsoft SharePoint, PST files, EML or MSG files, IBM Lotus Notes, and others) can only take one item at a time through there API. This is one item per user, repository, archive or similar, and most system APIs do not allow the item to be assigned to more than one entity (i.e. a user archive) per time. In most cases, therefore, most migration tools end up doing a multi-instance migration, and hence the MIC number is hugely important to assessing how long the migration will take.

Once a pilot migration has been run and an average obtained, you can begin to work out the length of time take for the migration as shown in Table 1.



| | | |
|----------------------------|------------|---|
| Average Msg Size (KB) | 50 | Your vendor can assist you finding this figure |
| Total Volume (GB) | 2500 | The total MIA figure in your source system(s) |
| Total Items (MIC) | 50,000,000 | $= (MIA \times 1024 \times 1024) / \text{Avg Msg Size}$ |
| Average Speed (msgs/sec) | 15 | Run a pilot migration run to find this figure |
| Hours/Day Migration Window | 8 | The maximum amount of time per day you can run migration |
| Total Number of Days | 116 | $= MIC / (\text{Average Speed} \times 3600 \times \text{Window})$ |

Table 1: Duration Estimation

Running a pilot migration is necessary to find out exactly how fast your migration will go and always consider that this migration speed will change as the migration progresses.

Effective Performance

When managing a migration, performance optimization is at the heart of the process. With millions (or even billions) of items to move, ensuring that you mitigate any expected environmental or configuration bottlenecks is essential.

CALCULATING MIGRATION TIMES

When migrating data, it's critical to obtain the highest item count per second successfully migrated. The difference between a migration running at an average of 3 msgs/sec and 4 msgs/sec may sound unimportant, but putting this in the context of a migration running at 8 hours per day for 50 Million items – that single sustained difference in the average item rate will account for an additional 145 days (almost 5 months).



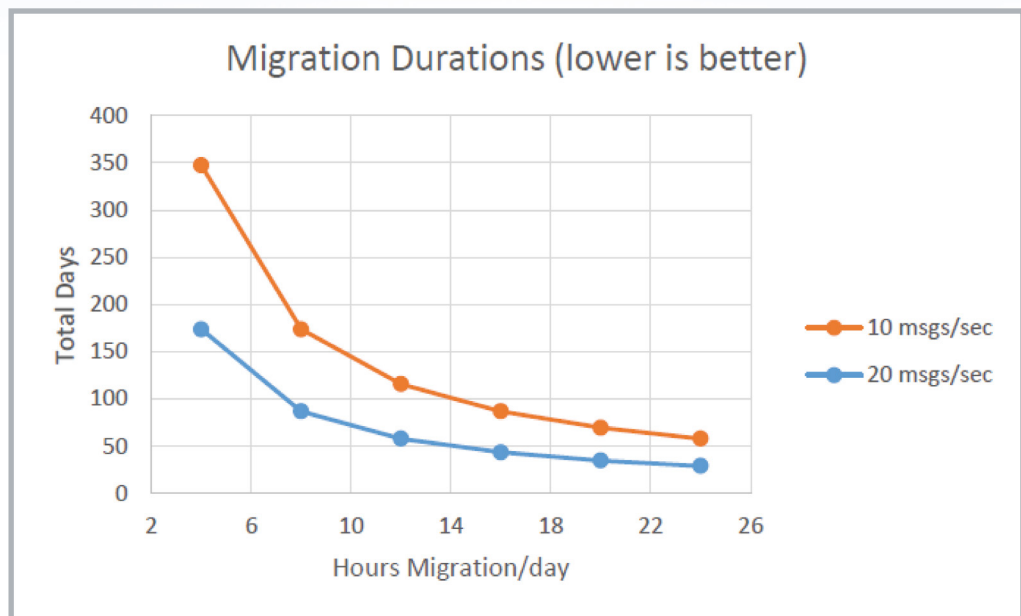


Figure 1: The difference between 10 msgs/sec and 20 msgs/sec, with different migration windows

It's important to ensure that the migration vendor you use not only clearly shows you the average and historic migration rates clearly, but also allows you to finely tune migration performance by scaling out nodes and adjusting number of archives migrated at one time. They should also use a tool that performs single instancing for migrations. Enhanced migration tools on the market will also perform automatic migration tuning – monitoring the source and target – to help ensure the smoothest and best possible migration. The migration product used should automatically adjust the migration rates, threads and migration pressure depending on the resource usage and contention of the source and targets involved – without any intervention. Make sure that you data migration service partner is using the right migration tools.

MIGRATION BATCHING

The order in which archives are migrated is actually very critical to a migration, not because it makes any difference to the data being migrated, but because it will affect overall performance. The goal for a migration is to ensure that all threads, nodes and servers are 100% engaged for the entire migration window – whether that is 8 hours/day, or 24 hours/day.



Anything less than 100% utilization of migration resources results in lost time – and makes the difference between migration rates as we discussed earlier. Let's take an example with 7 archives – 6 users, and 1 Journal.

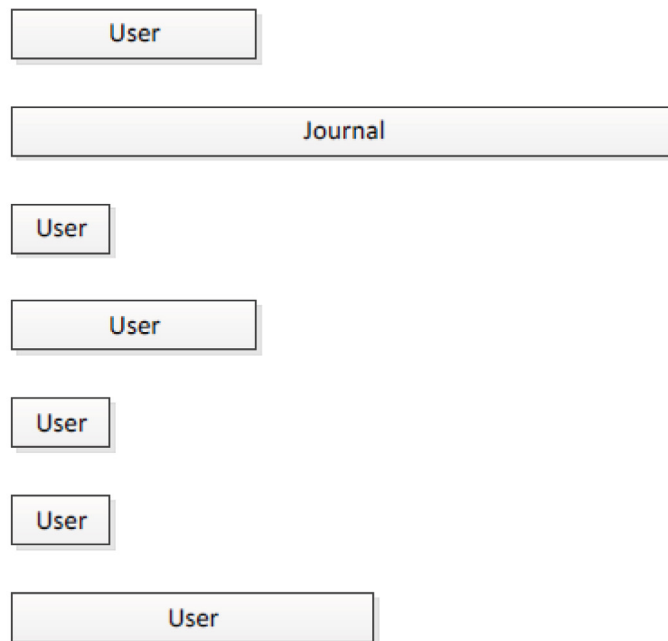


Figure 2: Sample archive listing, indicating archive size

Let's assume for a second that a migration tool migrates these archives in the order they are in the chart, and that each archive (for the purposes of this explanation) migrates at the same rate. This is what will happen in terms of thread activity:

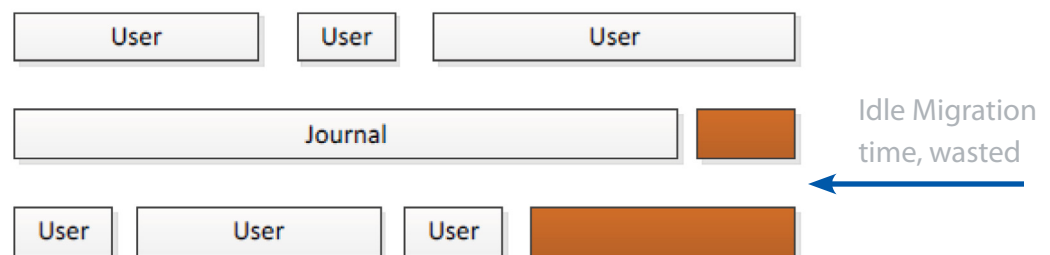


Figure 3: Idle time wasted by non-optimal migration



Some migration tools attempt to deal with this idea with “mailbox splitting.” However, splitting an archive doesn’t solve this issue – either the archives become too large in number to manage for reporting and audit purposes, or the split archives remain too large and suffer from the same problem.

Now let’s look at what happens when intelligent ordering is applied:

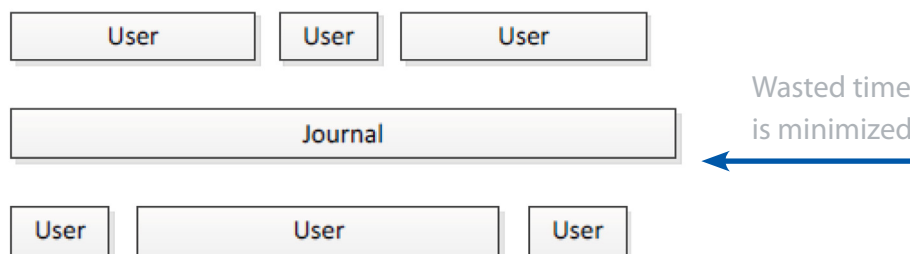


Figure 4: Optimized migration

However, some products take this even further and remove a significant limitation of current migrations – removing the limitation of one thread migrating one archive. Advanced migration tools are able to migrate a single archive without splitting and with multiple threads at a time, dynamically moving each thread onto the next archive in the set. This removes idle or wasted time in a migration process altogether and achieving the 100% migration utilization target.

Effective Error Handling

Handling errors in the migration process is critical to ensuring that performance is maximized and audit and manual intervention needs are significantly reduced as far as possible. In any migration, it is important to make sure that, as discussed earlier, you have an effective and rapid decision process in place for errors. When errors occur, however, it is important to manage those errors during the migration process quickly and efficiently post decision.



Aside from ignoring the errors altogether, there are typically three ways to handle archives with errors once they have occurred:

- **Re-run the task(s) after primary pass:** the task(s) that contain the archives with one or more errors, just re-running the task-as is.
- **Move the archive(s) to another group or task:** isolating those archives with errors that you wish to re-run and re-run them in parallel to the next batch of archives.
- **Re-run all archives at the end of the migration process:** waiting until a primary pass of all archives is completed before attempting to address errors.

Best practice dictates that the way to handle errors is to deal with them and make a decision on those errors as soon as they occur, without slowing down progress, as explained earlier. Using Error Grouping reports from migration products makes it easily possible to see the errors as they occur and group them.

Moving the archives to another group or task and then immediately re-running them in parallel to other archives as part of the mainstream migration means that you are addressing both errors and migrating new archives at the same time – achieving the best of both situations where possible.

Leaving all errors after a primary run of all archives introduces a significant risk to the project – since it is unknown how long those errors will take to resolve depending on the type and state of the error. This effectively introduces a task at the end of the migration project undefined in duration and resource, and so it is important to make sure risk is mitigated by tackling errors at the first opportunity in your migration project.

PROCESSING ERRORS: THE DECISION PROCESS

Making fast and efficient decisions about any items that end up in a state of error in the migration process is important to keep making efficient progress, minimizing idle time. For every error that occurs, an item has



already been extracted from the source, decompressed, and possibly even converted – and yet will not make it into the target archive and complete the full migration cycle. To put errors in perspective, if a significant error rate of 10% of all items failed migration, an effective reduction of 10% of the migration rate applies. For a 50-million item migration, with an 8-hour per day migration window, running at a very slow average speed of 4 items/second, that increases the migration by an incredible 48 days. Resolving errors and making decisions is therefore critical.

Best practice for error handling is to decide – before starting the migration – on an error handling process all stakeholders in the organization are signed up to and agree upon – allowing quick and efficient decisions to be made. If an organization has to stop migrating every time an error or group of errors occur, migration rates decrease significantly. Below is an example of a best-practice flowchart:

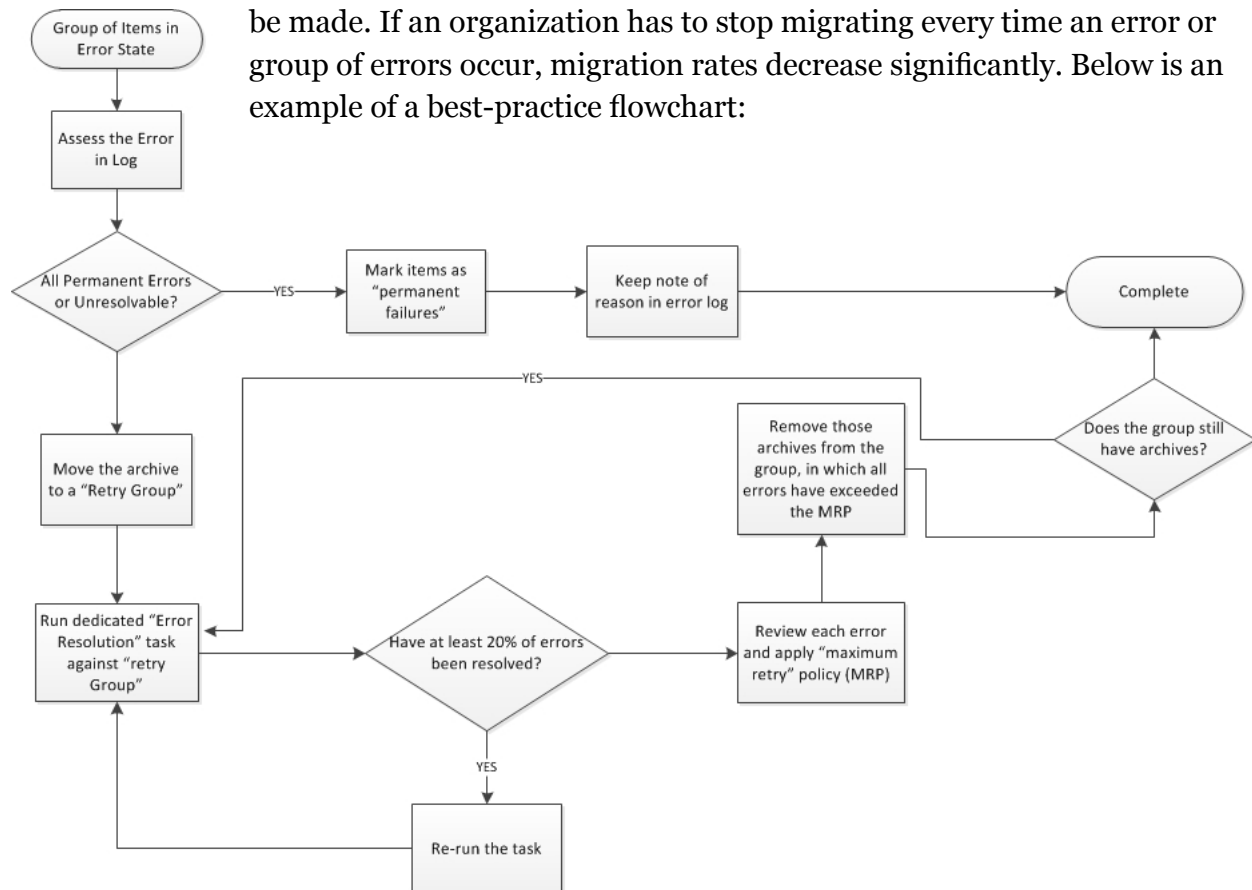


Figure 5: Best practice error handling



Make sure that before you start your project, you create your own error handling process – and enable fast error handling decisions. Make sure to get input from the migration vendor, too.

DISASTER RECOVERY PLANNING

Disaster recovery might seem far, far away from being needed as part of a migration because it is not a real-time or “critical business system,” but nothing could be further from the truth. Given that a migration is usually copying or modifying core data (typically email, SharePoint or files), it is absolutely essential that an organization knows where data is and what state it is in at any given time.

Imagine that a huge outage occurred right in the middle of a migration. Shortcuts for 20 archives were being converted, and now those 20 mailboxes are in a hybrid state. That has implications for your project, and now a decision needs to be made about those mailboxes. But what that decision should be is unclear because the state of the mailboxes is also unclear since the migration server, database and reports are gone.

Making sure you have some level of a disaster recovery plan around your migration is sensible. Some migration tools have the option of purchasing a hosted migration – where your migration partner can host databases and all management components, taking over full responsibility for Disaster Recovery. However, some organizations have a need to continue to host migrations themselves.

Figure 6: Disaster Recovery – example architecture

In Figure 6, you’ll notice two levels of disaster recovery have been configured – one at the migration database level, and another at the migration management component level. While it is critical to ensure the metadata (including chain of custody, audit and item state records) are replicated and available, it is also essential to make sure the management components are available for reporting purposes and to have a defined plan to start-up or switch the management components quickly and efficiently to the backup system or database so migration or reporting can continue.



While disaster recovery is critical, instant failover is not needed, and can often be a hindrance to the migration process.

Chain of Custody Management and Verification

When migrating items from a source to a target, the whole project's success rests on the assumption that the items will successfully and completely get to the target unmodified and without any issues or omissions. For user migrations, where a user's own archive or data is being migrated, legal compliance and Chain of Custody may not seem so important, while for journal or compliant data, they are. Verification, then, is at the heart of every migration, regardless of type, making sure that the items are indeed fully migrated to the target archive.

CHAIN OF CUSTODY

Ensuring total Chain of Custody during the migration means the every item migrated has been:

- Identified in the source system as an independent item ready for migration.
- Tracked throughout the entire migration process through each component from source to target.
- Hashed at source and hashed at target to uniquely identify each item being migrated.
- Subject to hash comparison based on both the source and target hashes being compared to verify the item has not changed in any form.
- Audited and can be reported on in a Chain of Custody report.

All migration products must offer at least basic chain of custody, including reporting subject and date and time as well as verifying the item was migrated. Your partner and/or migration tool should validate the item's hashes on both sides to verify the item did not change during data migration.



VERIFICATION

Verification is a distinctly different concept from Chain of Custody. While the two are related, verification is not about keeping a record proving the item's path through the migration or providing a source and target hash, but rather verifies that the item was not changed by the system after it was ingested by its API.

In many migration products, the item is handed over to the Target API after the target hash is taken – and unless the API reports an error, the item is assumed to be correctly ingested. Verification goes further, retrieving the item again after ingestion, and provided a post-ingestion comparison between the item's hash before ingestion into the target, and the retrieved item's hash. If they do not match, verification has failed – because it is clear the item has either not been ingested correctly by the target system, or because the target system has changed the item during indexing and storage.

Verification is an advanced concept for most migration products today and is only offered by a few providers with significant experience of migration. Advanced migration products provide two options for verification – basic and advanced, with advanced verification looking at every property on the item, and basic looking at the critical properties which “make up” the item. Combining Chain of Custody capabilities with Verification provides a bullet-proof migration process. Anything less and migrated items are at risk of being incomplete.

Winding Down Migrations

Winding down the migration begins immediately after completing the primary pass of the final set of archives. After that point, all that remains to migrate are either items that are in a state of error or have been skipped or excluded during the migration process.

A best practice wind-down process ensures that:

- Decisions have been made on all error types at the start of the wind-down process.



- Tasks have been re-run containing incomplete Archives (archives which have items in a state of error or skip) Items
- All final errors in the migration are understood and explained clearly, and final error report summaries have been run.
- The source system has been appropriately decommissioned and is no longer online (if appropriate) .
- The reports have been run, and all exported for compliance purposes, including archive summary reports, Chain of Custody reports and,
- The migration tool has been removed or uninstalled – if it is no longer needed for dynamic reporting, and migration database(s) fully backed up in case of a future need.

During the wind-down process, it is also essential to make sure that:

- **Everything has actually been migrated:** that archives or groups have not been accidentally ignored. Use a simple summary report to verify the items have actually been migrated.
- **The source system is no longer in use:** sometimes end-users of the source system might have unexpected links or still have non-standard routes to search and extract data from the system. Check logs (for example Internet Information Services or Firewall logs) to see if requests or activity is still occurring, and run a temporary system shutdown for 1 week to allow quick stand up if users report problems.
- **All users can retrieve and search correctly:** this verifies shortcut retrieval software (if relevant) is installed, permissions are set correctly, and archives have been fully enabled.
- **Full Reports for compliance have been run:** all reports necessary for the migration have been exported and saved, and they cover all items. This is critical to meet the compliance needs of your project – ensuring that you have all the compliance and chain of



custody information you need to prove items have been migrated (or were in a permanent state of error).

- **Database Backup:** the migration database has been backed up in case it is needed in the future for audit purposes.

Formally finishing a migration project can be a daunting task, but the involvement of a professional data migration services partners will help ensure that the entire migration has been completed and no details missed or left out.

ABOUT KANATEK'S DATA ARCHIVE MIGRATION SERVICE

Developed in partnership with Globanet, a leading provider of migration tools, Kanatek's Data Archive Migration Service is specifically designed for migrating large data archives to or within Symantec Enterprise Vault™ environments quickly, accurately and securely. Starting with a thorough review to understand the requirements and using Globanet's world-class Migrate product, Kanatek:

- Reduces the time needed to migrate.
- Reduces the risk of lost or unavailable data.
- Provides a seamless user experience.
- Reduces both the cost of migration and data storage.
- Eliminates redundant and corrupt data.
- Optimizes your Enterprise Vault experience.
- Maintains regulatory compliance.

While Kanatek's new Data Archive Migration Service makes it easier for Symantec users to migrate their Enterprise Vault installations, it is also invaluable to organizations looking to move to Symantec Enterprise Vault from other archiving products, whether as a strategic decision or to support merger and acquisition activities. Enterprise Vault is the industry leader in integrated content archiving and enables customers to store, manage and discover unstructured information while deduplicating it at the source to reduce on-going storage costs.





ABOUT KANATEK

Kanatek has been helping clients get the most from their investment in Symantec Information Management products since 1982 by delivering, supporting and managing complex enterprise storage, data management and service solutions. Kanatek solutions answer the most demanding, mission-critical big data challenges facing global enterprises today, helping clients understand and integrate new technologies by leveraging their existing investments and infrastructure. Kanatek's approach is based on its proprietary AIMS methodology, which provides a cohesive set of services in four key areas: Assessment, Implementation, Management and Support. With headquarters in Ottawa and offices in Toronto, Kanatek is a Symantec Platinum Partner and the only partner in Canada authorized by Symantec to deliver support. www.kanatek.com



ABOUT GLOBANET

Globanet is a leading developer and reseller of archiving, data migration, compliance and eDiscovery solutions worldwide. Founded in 1996, the company has been a pioneer in intelligent information governance and has developed a portfolio of software and services to help organizations manage data from creation to expiry. Globanet's proprietary solutions include the Globanet Merge1™ message capture platform and Globanet Migrate™ data migration software. A Symantec Platinum Partner with Master Specialization in Archiving and eDiscovery, its professional services team has extensive experience with industry-leading Enterprise Vault and Clearwell. Globanet's broad range of services includes policy and solution design, installation and configuration, data migration, custom add-ons and project-based eDiscovery consulting.

