ATEMPO

preserving data ecosystems

Planning the migration of your unstructured data

AN ATEMPO WHITE PAPER

Executive Summary

The volumes of data we produce are continuously increasing and this trend will accelerate exponentially with the boom of 5G, IoT and machine learning. IDC predicts that between 2018 and 2025, the volume of worldwide data will be multiplied by 5 growing from 33 ZB, to 175 ZB in 2025! 70 to 80% of this volume represents unstructured data, meaning data stored in no predefined format. It can be text data in emails or Word documents or non text data like audio files, images, etc.

To cope with this constant increase, companies are having to migrate their data from one server to another for many reasons. And the process is not without risks. This type of data (financial, marketing, sales) is a company's life blood. Any unavailability however temporary could jeopardise the activities of its teams. Hundreds of companies are arbitrarily finding themselves with several petabytes of unstructured data that they sometimes cannot control or manage properly.

This guide will help you plan your migration project to avoid the most frequent issues.

Table of Contents

Context

The question of storage migration and NAS devices appears often during the production process for new data storage, consolidation of existing storage facilities or the synchronization of several petabyte-scale NAS devices on remote sites.

The mobility of files and their independence from storage facilities is essential for companies looking to use their unstructured data over the long term, increase the capacity and performance of their storage facilities and improve their use with AI resources.

Whatever the reason behind the project, migrating very large volumes of unstructured files between two storage locations can be complex. If you must also manage petabytes of data, keep the source storage data in production or change technology, brand or storage type (from a NAS to GPFS for example), then the process can quickly become a nightmare. It is therefore essential to do the background research.

This guide is a very concise presentation of the different risks associated with migration and has a 7-step plan that will help guarantee your project's success.

Chapter 1: Identify the migration obstacles

Experience makes no difference. Migrating files between storage locations or NAS devices is still a difficult operation, even for experienced storage managers. However, there are solutions to each of the obstacles an organization may face.

OBSTACLE NO. 1: MANAGING AN UNMANAGEABLE PROJECT

Embarking on a file migration project between two storage locations often feels like a leap into the unknown. IT teams may well know their business, but the challenge is to involve the end-users. Even with expert knowledge, we often find the same challenges:

- schedule interruptions,
- replication of weak architecture in the new storage facility,
- loss of file integrity, duplicate files, or worse still, false duplicates leading to loss of data ,
- mismanaged file attributes and file rights,
- a time-consuming “trial and error” approach,
- and an end result that completely fails to meet business needs, even if there is no loss of data.

SOLUTION: SURROUND YOURSELF WITH PROFESSIONALS

A business manager is not an expert on data and even a storage manager is not an expert in leading complex projects involving multiple networks, storage/data and protocols. It is essential to surround yourself with professionals who take care of:

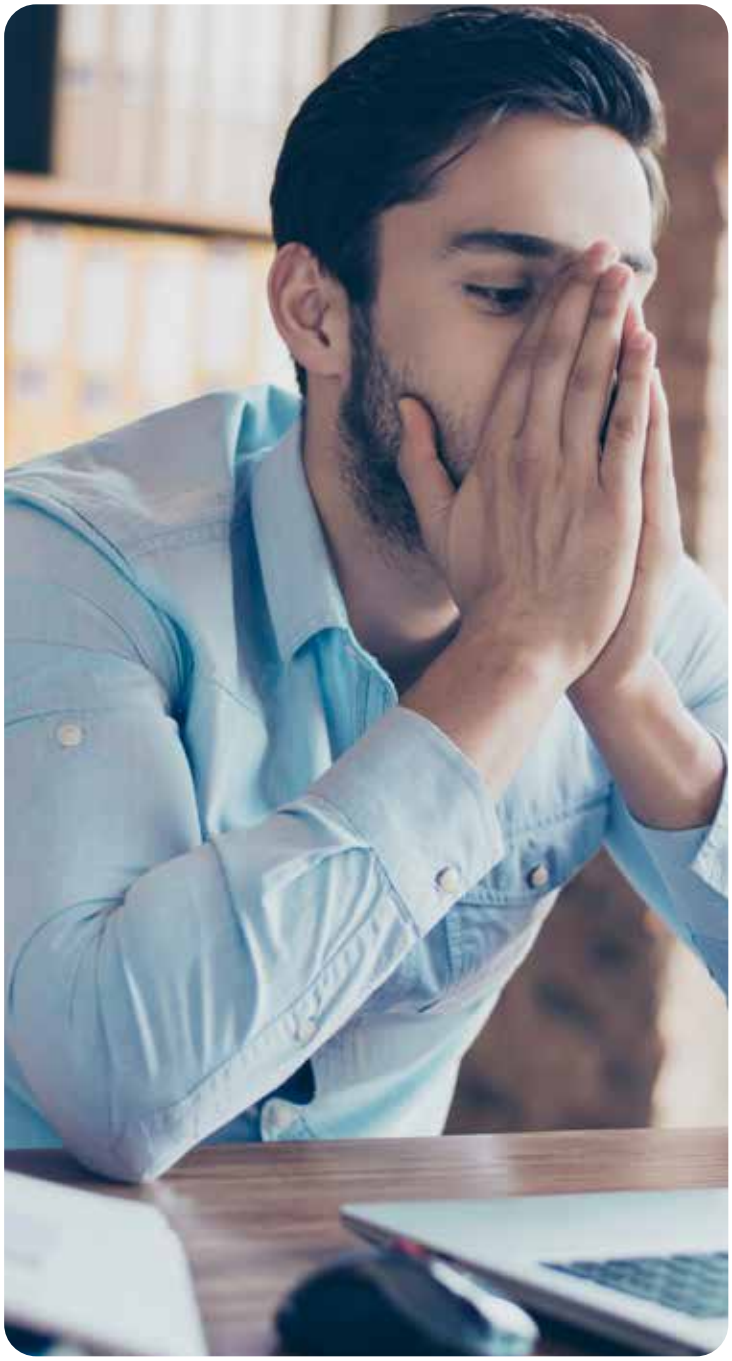
- assessing and defining the technical scope,
- preparing the migration alongside the end-users and storage administrators,
- process monitoring,
- and final acceptance of the files migrated to the new storage location.

Keep in mind that no storage migration process is valid without guaranteeing file integrity and access rights.

Involving business users will ensure the quality of the storage solution in the long term.

OBSTACLE NO. 2 - LOSS OF DATA INTEGRITY

It's clearly the worst-case scenario of file migration. The bad news is that it is not an urban myth: three-quarters of the companies surveyed by Kroll Ontrack stated that they had lost data during a migration and 23% of them never recovered it.



Even the fastest petabyte-scale migration projects often extend over several weeks. It is impossible to record everything by hand, so how can you avoid losing some data during migration and being unable to find it in the new storage device?

It is essential that the migration solution has a log listing all the transfers down to file level and that for each transfer cycle, a dashboard helps track the daily progress in the cycles of file migration to the new storage facility.

SOLUTION: TECHNOLOGY THAT'S UP TO THE CHALLENGE

When launching the process, the performance of the migration tool makes all the difference. The volume of files for migration can quickly outstrip the capacity of programs that are less sophisticated or less suitable for large volumes.

OBSTACLE NO. 3 - AN EXTENDED TRANSITION PERIOD

Because of the logistics involved and the volume of data being processed, file migrations can stretch over several weeks. But IT teams know that the longer they take, the greater the risk of disappointment and failure. A badly structured operation can take up to 180 days!

An indefinite delay can lead to a vastly inflated budget, increases the risk of compromised data access and affects the productivity of business users for longer.

SOLUTION: DEMONSTRATE SPEED AND FLEXIBILITY

Few companies can afford to stop storage production for several weeks or to endure downgraded user experience. Some migration solutions unfortunately impact user response times.

The chosen solution must have:

- the ability **to limit the impact of the file migration on the source storage** so the migration process can take place while users are working or applications are running;
- the ability to transfer data at very high speeds; for example, by adding Data Movers or filling a 10 GB network and using all the IO storage and network capacity to move as many files as possible in a limited time;
- and finally, the ability to limit the migration to certain time slots to control its impact on the network. No more migrations that take the IO and bandwidth hostage for weeks!

OBSTACLE NO. 4 - INCOMPATIBLE STORAGE PLATFORMS

One of the main difficulties when migrating between storage locations or NAS devices is foreseeable, but often still a sticking point. Moving to a new type of storage platform can cause major compatibility problems and internal file formatting issues.

Companies often find themselves stuck with a single manufacturer it is difficult to change.

They can either accept that this limits their growth capacity or embark upon an uncertain migration path **with the risk of losing their data.**

SOLUTION: OPT FOR AN INDEPENDENT MIGRATION SOLUTION

To allow data to move freely into a new NAS storage device, the migration solution converts the files into a format compatible with the system and target protocol. This conversion also preserves the ACLs and user attributes.

OBSTACLE NO. 5 - INTERRUPTIONS TO PRODUCTION

Despite the most rigorous preparation, migrating files to NAS storage devices remains an invasive process. Interrupting data generation has an impact on the whole company's performance. Data changes occuring after the initial scan may be quite small. The problem is often not migrating new data, it is scanning the filesystem to detect changes and additions. This can take days or weeks to complete.

SOLUTION: USE ITERATIVE MIGRATION CYCLES

Suspending production is not inevitable. Choose a solution that allows the source storage device to remain accessible throughout the operation, using iterative migration cycles.

- Each cycle limits the amount of bandwidth used during the storage migration, giving priority to users.
- And each cycle is automatically triggered at shorter intervals.
  At every run, the tool detects which objects have been modified and performs an incremental sync until the point of convergence between the source and target storage locations has been reached.

At this stage, it is possible to make the final cutover by briefly stopping production and changing the network. The iterative cycles ensure the files are migrated safely and minimize any impact on end-users and business processes.

Chapter 2: Identify the risks when using free tools

One of the specialities of the Atempo experts is **rescuing failed file migration.**
Our new customers come to us after the damage has been done and all tell us the same story. To keep their budget small, they chose a **free data migration tool**.

According to a Gartner survey, over 50% of data migration projects finish over budget. Some free tools, like **Robocopy** or **Rsync** satisfactorily complete simple migrations!
However, it is not uncommon that a free tool leads to more problems than solutions.

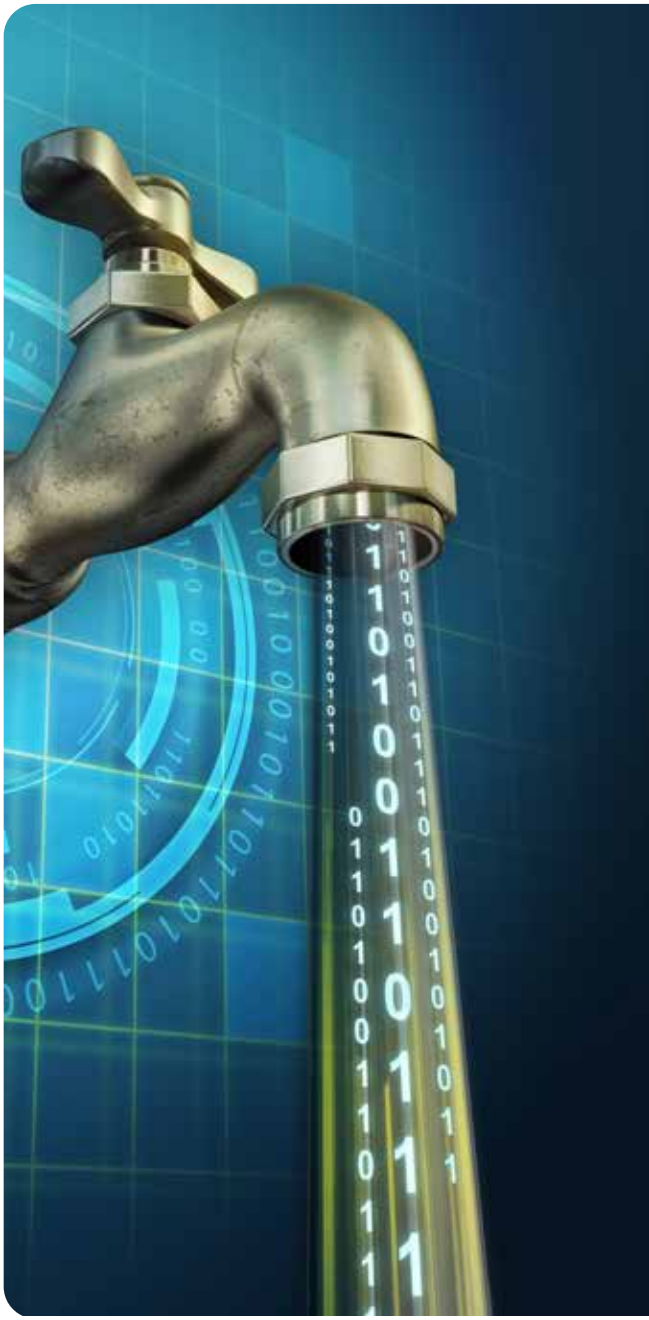
Using discussions with customers who contacted us after a failed migration project, we have produced **the 4 most common preconceptions about free tools:**

- **Preconception no. 1**: “My data is always safe with a free migration tool.”
- **Preconception no. 2**: “Any tool can manage any volume of data.”
- **Preconception no. 3:** “Free tools are for everyone!
  So logically, they can be used in all environments.”
- **Preconception no. 4:** “Free tools are the best option for a small business.”

PRECONCEPTION NO. 1: *"MY DATA IS ALWAYS SAFE WITH A FREE MIGRATION TOOL."*

FALSE - Free tools quickly become saturated during migration which can lead to occasional data losses.

The example of **MySpace** is indicative of this. Wanting to save money, the IT team migrated their data without any backups beforehand.
A high-performance tool uses **incremental migration cycles** to secure the migrated data.



The user can then:

- trace the migration history down to the file level: most of the migrations carried out by our customers are completed with the stored data in production;
- identify the stubs (the files archived in tiering) if there are any on the filesystem. It is best to ignore them whereas most free tools recall them;
- track the daily migration progress, to anticipate the cutover to the new storage location.

A storage or NAS device is of key importance to the company. Even if all the stored data is not the same size or of the same "freshness", the trace of each file must be kept, as it is or modified by a user during migration.

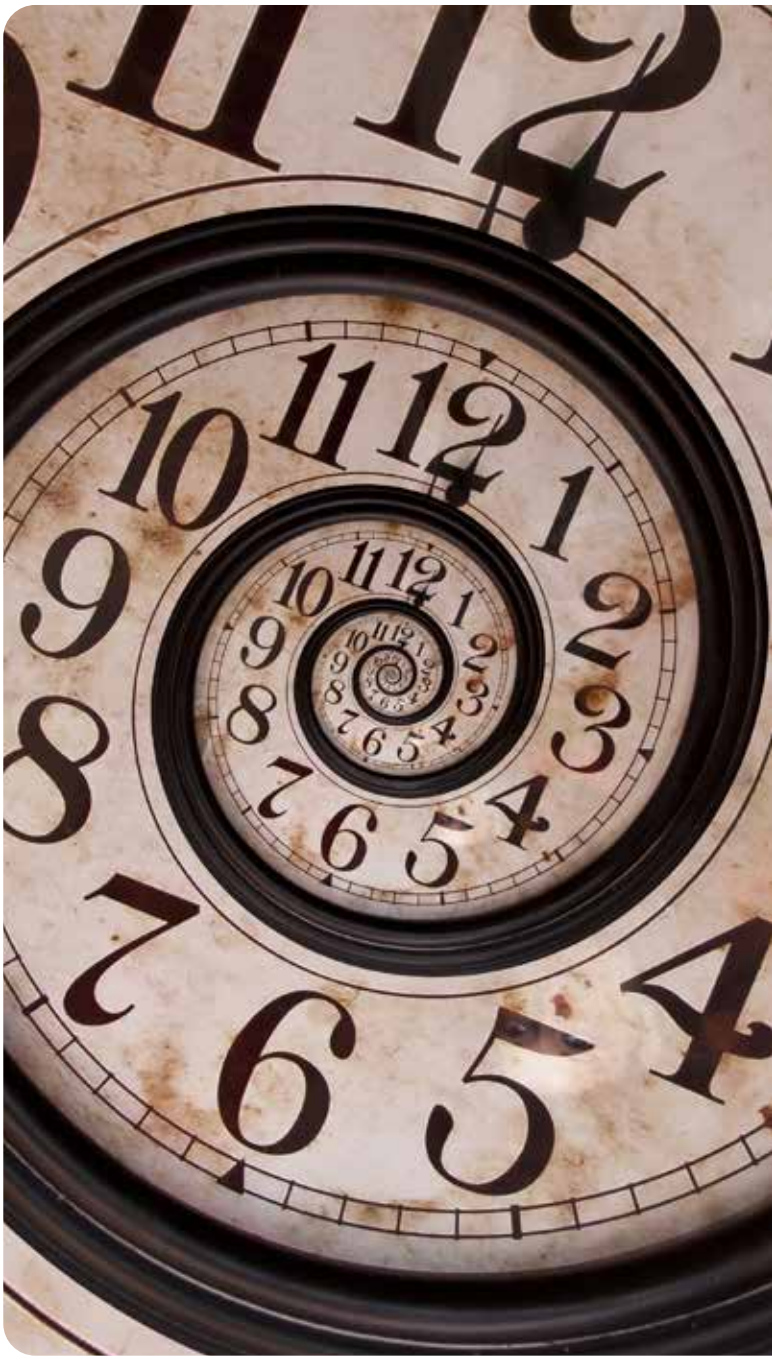
PRECONCEPTION NO. 2: *"ANY TOOL CAN MANAGE ANY VOLUME OF DATA."*

TRUE - Free tools can process large volumes of data if you accept that the data migration will have a **long and unpredictable schedule.**
One of our customers regretted the considerable amount of time Rsync required for a scan (exploration) of the NAS.

In the **absence of reliable storage synchronization** in production and replacement storage, our customer had a choice between:

- duplicate entries that represented a growing risk for the data integrity;
- or stopping production during the data migration - for an undetermined period.

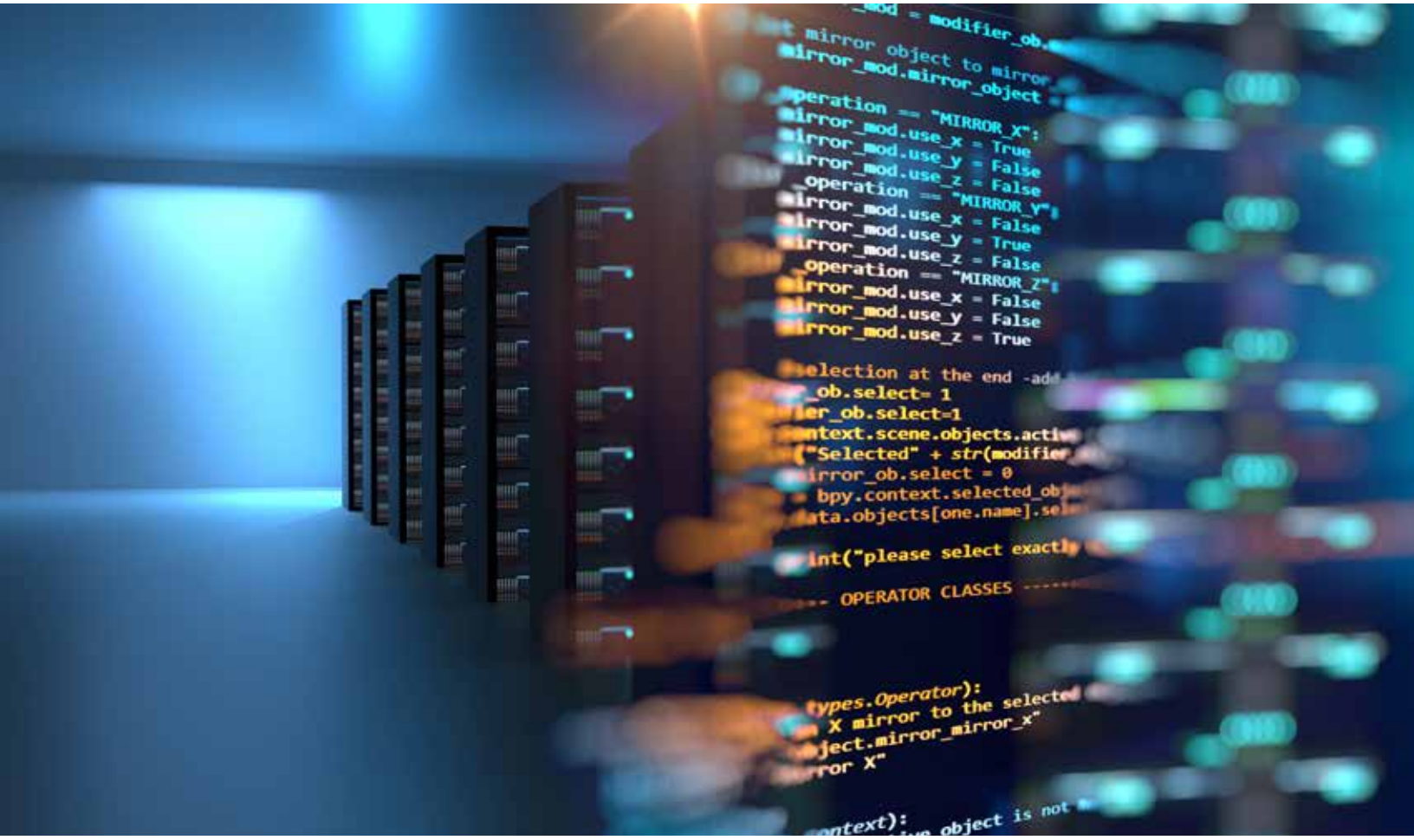
Choose a tool that guarantees **quick and secure synchronization** with the storage, even with day-to-day modifications (up to 25% for some customers). Opt for a tool that can list the changed files and create a mirror organization in the destination storage facility. Some tools can migrate **several petabytes of data** without stopping production.



PRECONCEPTION NO. 3: *"FREE TOOLS ARE FOR EVERYONE! SO LOGICALLY, THEY CAN BE USED IN ALL ENVIRONMENTS."*

FALSE - Do not confuse "unspecialized software" with "adaptable solution". To help your company control its data, choose a tool that:

- **processes unstructured data**, keeps the ACLs (Access Control Lists) and rebuilds the CIFS/NFS shares during the migration;
- adapts to the most complex IT environments. Independent supplier solutions which **facilitate migrations** between branded or different storage platforms - like Dell, Qumulo, Lustre, DDN, IBM, Huawei and Panasas.



PRECONCEPTION NO. 4: *"FREE TOOLS ARE THE BEST OPTION FOR A SMALL BUSINESS"*

FALSE - It's not just major internationals that migrate complex data! More and more start-ups, SMEs, laboratories, etc. manage **several hundred terabytes, even petabytes of data.**

It is essential you understand whether a free tool is capable of providing the right support for your migration project. This point is crucial because it will determine the costs you may have to bear if the migration fails.

The migration of unstructured data and files between storage devices is complex. Solutions such as Rsync and Robocopy are not scalable, require costly manual verification and lead to data loss. The impact on stored production data is not controlled

The pitfalls of a DIY approach

- A time-consuming "trial and error" approach,
- Files or folders that are missing or inaccessible,
- Storage service production stopped,
- Use of all the available IO and bandwidth,
- Some tools maybe free, but they come with heavy hidden costs: your storage resources will spend weeks on this ...
- Last but not least, when you need help, there is nobody there.

Chapter 3: Successful data migration in 7 stages

Every day, Atempo experts provide support with the migration of large volumes of unstructured data. Here is some advice for a successful project in 7 stages.

1- IDENTIFY THE BUSINESS REQUIREMENT

Technical challenges must not overshadow **the aim of the file migration project: meet the requirements of end-users.**

Furthermore, the involvement of the business guarantees best storage practice.

The storage manager translates business needs into technical imperatives. At this stage, he determines the type of storage migration or NAS device required:

- a massive migration **import** in one go;
- a large scale, **long-term synchronized migration**;
- an operation to **consolidate** data from several storage locations.

Deliverable:
A requirement specification document, written by the IT teams in collaboration with the end-users.
It will be a reference for the final migration delivery.

2- MAP THE MIGRATION PROJECT

The IT team continues its prerequisite inventory and focuses on:

- the nature of the data: migrations of unstructured data often involve text and multimedia files, search results, IoT...which in each case poses specific challenges depending on the size, number, organization.
- the volume of data: more and more migration projects involve several hundreds of terabytes, even petabytes of data.
- the complexity of the IT environment: silo storage, different formats and network protocols, transition to a new storage manufacturer, etc.

The precision of this preparation broadly determines the success of the migration project.

Deliverable:

An assessment of the project's risks and a list of the migration requirements that determines the target organization (governance structure, target standards, end-users and future data use, etc.).

3- PREPARE THE DATA FOR MIGRATION

These requirements must be implemented before the migration takes place! Whether you are planning a complete migration or a data consolidation, now is the time to clean up the files. This operation involves:

- **analyzing** the structure of the source data storage: number of folders and items per folder, links between files, etc.
- **correcting** the storage architecture if necessary;
- **deleting** duplicates and legacy elements - no need to transfer bad habits;
- **establishing** format standards and rules for naming files;
- **deciding** the migration rules and priorities, to make sure that the data is transferred to the right place at the right time and protected during the migration.

It is also the time for **scheduling the migration**: roll-out period, import frequency, operational times... Ideally, the business teams will be involved in this stage to reduce the impact on their activity and to **automate as many operations as possible**, to reduce the risk of error.

Deliverable:

Source data ready for migration, a list of source and target folders and shares, a prioritized migration schedule for groups of folders and the migration operation and a best practice guide for end-users for naming and storing files.

4- CHOOSE THE BEST TOOL FOR A COMPLEX DATA MIGRATION PROJECT

The market for migration support software has exploded. But most of the tools quite simply do not have the capacity to handle several petabytes of data in complex, different environments. Others which are too expensive are designed for major international groups.

Miria for Migration from Atempo makes storage migration available to all organizations and all IT environments.

- It automates the migration process and generates reports throughout the operation.
- It can handle **several petabytes of data** safely and securely.
- It rapidly detects changes or additions in your very extensive tree structure - only Miria's FastScan can do this **within the time constraints**.
- It adapts to a wide choice of file systems as it **maintains the ACLs** (Access Control Lists).
- Dell / EMC Isilon / ECS, Qumulo, Lustre, DDN GRIDScaler, IBM Spectrum Storage, Panasas, object storage, etc., independent from suppliers, Miria makes **migration between different storage platforms possible.**
- Miria **rolls out Data Movers on demand** to make data migration quick, scalable and accomodate multi-storage solutions.

Deliverable:

Complete management of your data migration project by Atempo experts.

Miria for Migration meets all the essential preparation, speed and reliability requirements for migrating large volumes of data in complex environments.

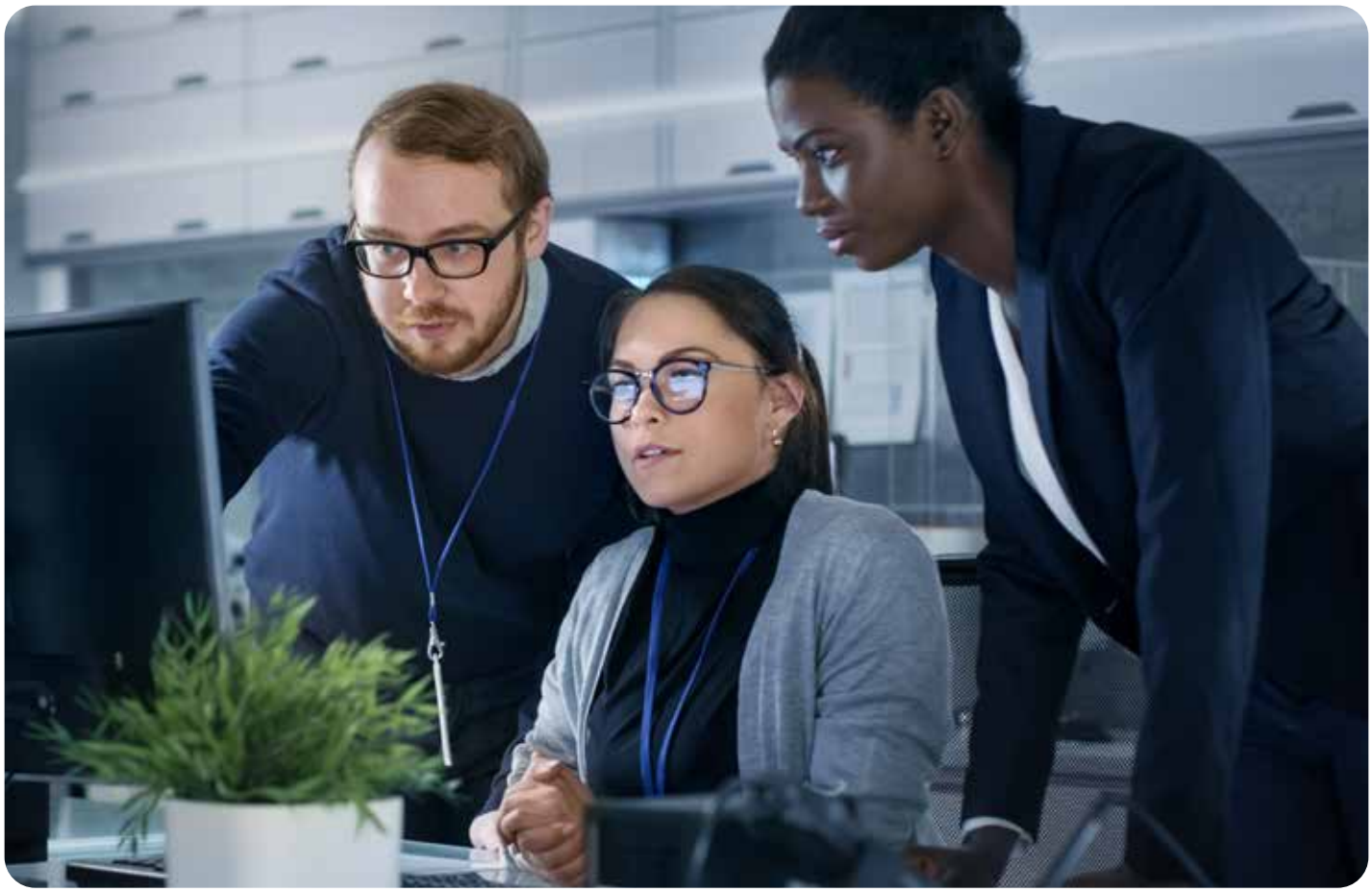
5- RUN A DATA MIGRATION PILOT TEST

A **field test** is part of the migration process recommended by Atempo. Data Management experts define the pilot test scope and launch the migration in real time.

It is also an opportunity to implement the security backups and other **risk management measures,** to eliminate any data loss.

Deliverable:

Approval of the choice of solution and migration preparation



6- IMPLEMENT THE FILE MIGRATION PROJECT

Most migration solutions cannot shorten the operational schedule. Furthermore, they monopolize storage bandwidth to the great displeasure of end-users. The tricky management of both the migration and its impact on business increases the risks of malfunction.

Miria for Migration has an innovative approach.

- Migration takes place over **selected time-frames**. Its use of bandwidth can be reduced or on the contrary pushed to saturate a 10 GB network (as an example) to accelerate file transfer.
- The files remain **accessible for analysis or processing** throughout the migration.
- In every automatic migration cycle, the tool detects the modified files in the source storage. It implements **incremental synchronization** with the target storage until both storage devices converge.

The automated processes and data flow management mean the migration is finalized in a few hours, even a few minutes, after convergence.

Deliverable:

Controlled data storage and NAS migration with no impact on production.

7- CHECK THE SUCCESS OF THE DATA MIGRATION

No file migration project is complete without a precise assessment of the target storage. This involves:

- An **inspection of the target storage** by the storage manager and the migration consultants.
- A **test of the migrated data**, using predetermined test scripts (location, access rights, etc.).
- The assessment of **end-user satisfaction**, using the business requirement specification document from the launch of the migration project.

Miria for Migration generates **automatic reports** throughout the migration project. This makes assessing the project's success even easier.

Deliverable:
An operational target storage facility, approved data migration and satisfied end-users!

Atempo solutions

- **Miria**: Backup, archiving, synchronization, migration and copy solution specific to unstructured data and very large volumes - petabyte files and storage

- **Tina (Time Navigator)**: Backup and protection of servers and applications for data centres, remote sites and distributed environments.

- **Lina (Live Navigator)**: Continuous data protection (CDP) solution for desktops, laptops and file servers.

About Atempo

Atempo is an independent software company based in Europe with an established global presence, providing solutions to protect, store, move and retrieve all critical data for thousands of businesses worldwide. With over 25 years of experience in data protection, Atempo offers a complete range of proven solutions for backing up physical and virtual servers, workstations and migrating very large volumes of data between different storage systems. Atempo's three flagship solutions, Lina, Miria and Tina, are labelled "Used by the French Armed Forces" and "France Cybersecurity".

Selected as part of the French Tech 120 government program to create 25 unicorns by 2025, the company, headquartered in Paris, has a powerful network of value-added wholesalers, resellers, manufacturers, integrators and managed service providers.

ATEMPO
preserving data ecosystems

For more information: www.atempo.com

Version February 23, 2021 - 11:50AM