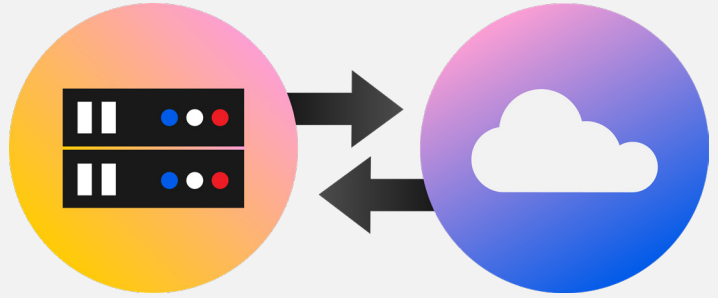


SOLUTION BRIEF

Accelerated Data Migrations

Streamlining the migration process to move data without consequences.



As unstructured data growth continues unabated, IT teams must have the ability to relocate data on demand to support the needs of the organization. From migrations driven by a need for performance improvements to those aiming for cost savings or resulting from corporate acquisitions, migrations are a frequent requirement and moving data, especially at scale, is fraught with difficulty.

Whether data is moving between on-premises storage systems, into or out of the cloud, or from one cloud provider to another the process is typically lengthy, requires in-depth understanding of a variety of storage options, demands meticulous planning, is extremely resource-intensive for IT teams, and can result in data corruption if executed incorrectly.

Assessing project requirements, selecting the migration target, and designing the migration path all add to the technical burden of those responsible for the migration.

There are so many elements that require attention within migration projects that business leaders often overlook a critical factor that can significantly impact the overall timeline and success of the operation: network speed. Data can only move at the speed of light as controlled by internet or local fiber connections, so moving large volumes of data takes a very long time, even within the same data center.

To further complicate the process, data can change in real time. This has important ramifications for data integrity during and after the migration. As a result, migration deadlines have a habit of slipping, even when the process runs uninterrupted.

Panzura Symphony is a hybrid cloud data services platform with extensive migration capabilities made to tackle the complexities of data movement, in real time. Symphony accelerates and simplifies large scale data migrations through a uniquely direct, metadata-aware approach that moves data across the shortest possible distance, with the fewest possible touchpoints, at the fastest possible speed.

This solution brief discusses key considerations for migrating data at scale with Panzura Symphony. It explores data movement to and between cloud providers, between on-premises storage systems, and also covers migration between object storage and file storage, in either direction.

When data is bound for the cloud, migrations can start immediately

Migrating to the cloud or between cloud providers is often driven by the promise of lower costs along with increased agility and flexibility. Those benefits are real and substantial, however it can take an extended period of time to realize them.

Cloud migrations are typically approached in a linear fashion. First comes the hyperscaler or cloud storage provider selection. Then follows the challenging work of architecting the perfect environment and designing the migration itself. The flexibility and scalability of the cloud makes this a complex mission as it involves multiple decisions around distinct types of storage. Some data will migrate to object storage, others to various performance levels of spinning and solid-state disk (SSD).

Rightfully, that means the design and build phase can consume months and with most migration tools, IT leaders are forced to delay the migration until these phases are completed.

However, when migrating data with Panzura Symphony, none of the decisions around the eventual data location matter. Instead, Symphony gives organizations a significant head start as it allows the longest part of the entire process — moving the data — to begin as soon as the hyperscaler has been selected and the storage bucket enabled.

The key is intelligent use of metadata

The information that informs decisions such as final data location typically sits within the metadata — it does not require content inspection to assess data relationships, permissions, and other attributes such as date last accessed or edited.

However, metadata awareness varies widely among migration tools and where it exists, it is usually limited to awareness only. This renders most tools incapable of taking action based on any intelligence they are able to surface.

Symphony is fully and comprehensively metadata-aware, enabling migration management by event, file system or object attribute, and direct API call. This means data can eventually move to its final resting place — whether file, object, or performance tier — after cloud architecture decisions are made, and without fear of losing any source file metadata, including access control lists (ACLs).

This patented approach to managing metadata doesn't rely on stubs, symlinks or any unnatural acts to migrate the data beyond simply keeping the original metadata stored within the object. It also has the potential to add queryable metadata, making searching for the files much simpler, to enable the final leg of the migration.

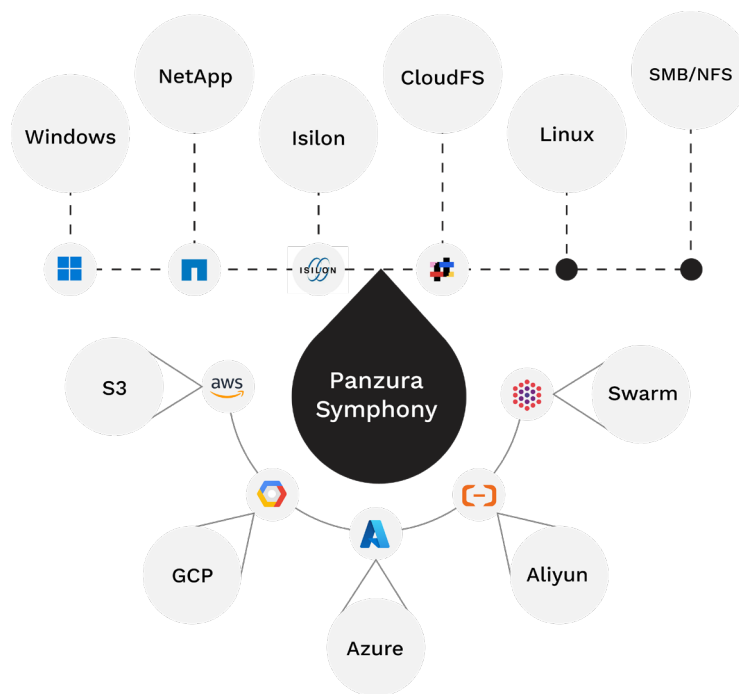
Effectively, Symphony accelerates cloud migrations by enabling two parallel streams. The first is the data or file system transfer itself, allowing the data to be moved to and stored in an inexpensive object format as soon as the hyperscaler decision is made. This means the data can already be there awaiting its final move, when your architecture team finalizes the design.

Then, once cloud architectures have been completed, Symphony is used to transfer the original dataset from object storage to the chosen cloud storage types and locations, all within the hyperscaler's network and without incurring egress fees.

By commencing the data transfer at the earliest possible moment, and running in parallel with other project tasks, Symphony reduces the cost, complexity, and duration of migration projects all while ensuring that the original dataset and metadata such as ACLs are preserved in object format.

Symphony's cloud migration features include:

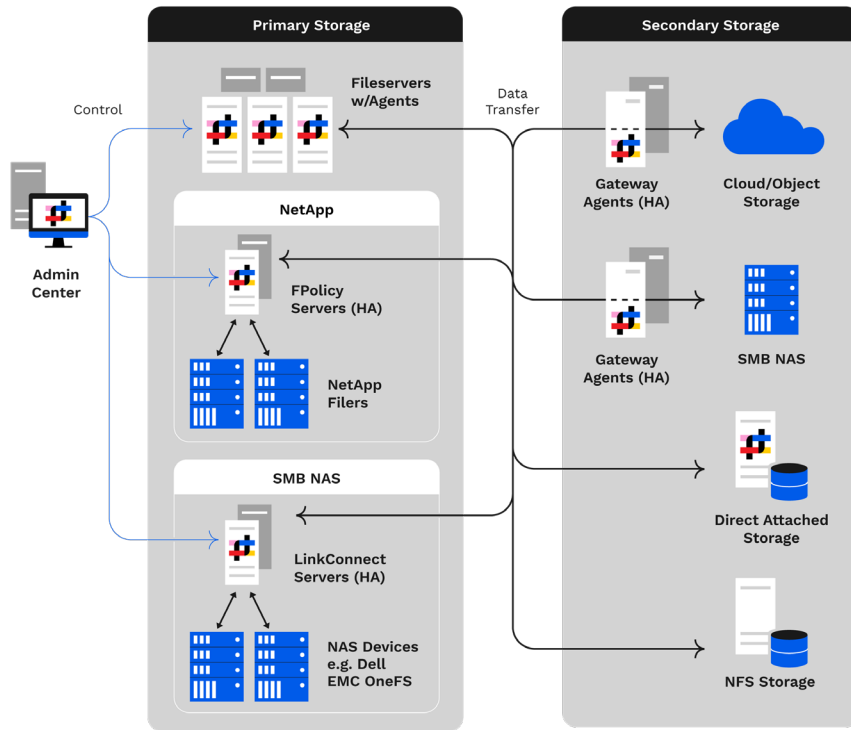
- ✓ Source support for file storage platforms and protocols including Windows Server, Dell EMC Isilon, NetApp, IBM Spectrum Scale, NFS, and SMB.
- ✓ This native support greatly improves both migration performance and reliability.
- ✓ Source support for cloud and object storage platforms including Azure, IBM Cloud Object Storage, Dell EMC ECS, NetApp StorageGRID, DataCore Swarm, Hitachi Vantara HCP, Amazon S3, Google Cloud Platform, and Alibaba OSS.
- ✓ Disintermediated, stateless, command and control architecture that places the policy server outside the data path for maximum availability and highest performance. Panzura's patents for Symphony are around the disintermediated architecture of the system. No other data management and mobilization solutions are capable of achieving Symphony's levels of performance for data migration and mobilization.
- ✓ Ongoing visibility of data demographics with detailed and exportable reports. Data migrations tend to be a black hole of information. Most tools simply tell you how many files have been transferred successfully. By contrast, Symphony provides administrators with all the relevant data during a migration: file count, capacity migrated, capacity, files waiting to be migrated, percentage of file changes, and even a time estimate to completion, based on earlier performance.
- ✓ No middleware, databases, or indexes to manage, and no single point of failure. IT leaders familiar with original data mobilization and hierarchical storage management (HSM) solutions will especially appreciate the freedom this approach affords them.
- ✓ Policy, event, and API-activated migration for files and objects.
- ✓ Detailed system dashboard, responsive UI, reporting and notifications.
- ✓ Distributed deployment scaling to exabytes of data and billions of files and objects.
- ✓ ReST Management API for custom integration with orchestration and workflow systems.



While Symphony is a complete data services platform, providing comprehensive data management tools, its migration feature can be shut down following migration to reduce IT and operational costs. Symphony migrations are self-contained and have no elements requiring features to remain “sticky” in order to decode or manage data, unlike some competitive products. This allows the tool to be removed completely after the migration has concluded.

Symphony's Architecture and Management

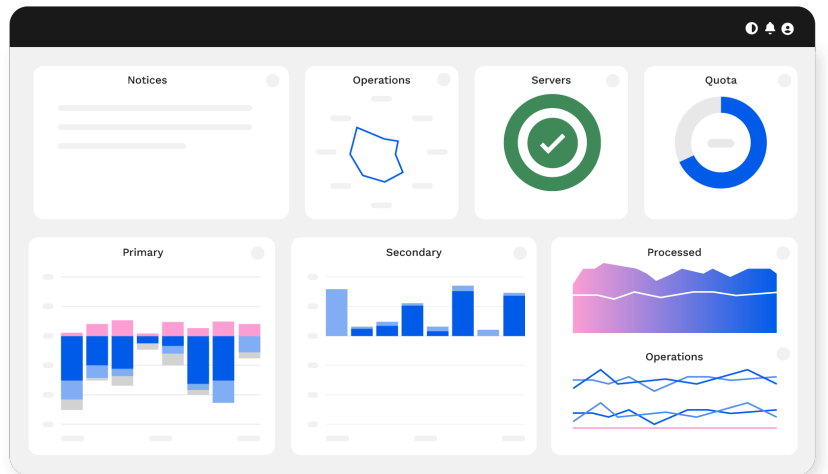
Symphony's native and protocol support for major storage systems provides seamless migration workflows for any-to-any data migrations. File and object data is migrated while retaining original metadata including ACLs.



Symphony's AdminCenter provides a single-pane-of-glass view of system operations and other important details such as migration progress and estimated completion time based on previous performance. System notices, operation types, capacity consumption, and storage server health can be easily monitored from a unified dashboard.

Cloud migration project managers can identify unexpected policy behavior, network and storage consumption and other critical data points at a glance. Intelligent widgets allow managers and system administrators to effortlessly view active and historical operations, task execution, and systemwide node health.

Using a responsive UI design, the AdminCenter dashboard can be accessed by smart phone and tablet users, as well as on desktop browsers. This makes monitoring the status of your migration and the health of your systems easier.



Streamline and boost performance through enablement, not obstruction

The performance associated with any type of migration is usually down to the speed of the wire, and the efficiency of the tool doing the migration. When moving large volumes of data, every step or interaction the migration requires creates a bottleneck that effectively throttles the maximum achievable transfer speed over the available network.

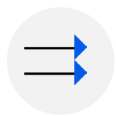
The chief culprits are latency and resource utilization. Latency comes into play for every check that must be performed before an operation takes place. For example, this includes consulting a database for information, slowing the entire process. Additionally, migrations require significant processing power, so the more complex the process, the greater the resource consumption.

Common tools used for migration, such as Robocopy, typically move anywhere from 4-5 TB per day to migrate customers from their on-premises storage systems to a cloud-native storage solution like S3, or Panzura CloudFS. **Using Symphony, Panzura customers see as much as a threefold migration performance improvement to 10-15 TB per day over the same infrastructure.**

Symphony intentionally sits outside of the data path to facilitate maximally efficient data transfer, at maximum speed. Its disintermediated architecture means it does not use indexes, databases or distributed hash tables that can cause bottlenecks and create complexity. This significantly reduces latency, optimizes resource utilization, and enables exabyte-scale performance, while also reducing the attack surface for the data itself.

Symphony does provide database integration for reporting output, logging and event chain of custody but does not require a database to operate. This 'middleware-less' architecture powers scalability and performance rates impossible to achieve with legacy platforms.

Comparing the differences between Panzura Symphony and Robocopy when used for migration:



Parallelism

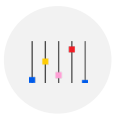
While Robocopy can be configured to support multi-threaded transfers, Symphony provides better control and visualization of workflows (serial / parallel configurations) as well as end-to-end task management. Symphony allows for configuration of the maximum number of operations that may be performed in parallel on behalf of a single policy for a given source.

Symphony allows for configuration of the total number of operations that may be performed in parallel across all policies on a gateway. This does not limit the number of policies which may be run in parallel — operations will simply be queued if necessary.



Granular control of file movement

Symphony provides highly granular rules, wildcard, and regex matching for filename and path selection. It also provides AD owner and owner group matching, as well as providing compound rules which can express complex criteria. Robocopy does not have either of these capabilities. Symphony provides sophisticated date matching options and enables include and exclude selection of nested directories, whereas Robocopy is limited in this respect. Symphony provides relational database integration for source and target dataset comparison (post transfer), while Robocopy does not have this level of integration or logging.



Task control and continuity

Symphony tasks can be restarted, paused and re-run, whereas Robocopy errors are logged but not resolved. Errors must be manually retried once error logs are reviewed. The entire Robocopy process will need to be re-run or a new process will need to be created. Symphony policies can be re-run and any previous errors will be automatically retried, but successful operations will not be repeated. This makes Symphony suitable for very large datasets even in the presence of failures such as network interruptions, due to its recoverability.



Visibility

Symphony provides a single pane of glass complete with data assessment, which is essential for planning, interactive and sharable reports, graphs, and charts, interactive cost modeling and export options including CSV, JSON and database. It also provides comprehensive, thread-aware logging with database and syslog options. Robocopy does not provide assessment, reporting, visualization, or cost modeling.



Single control pane for all data sources

Symphony provides a single pane of glass for configuration overview, source/target mapping, task control and scheduling and supports SMB, NFS, NetApp, Dell EMC Isilon, S3 and compatible sources including AWS Glacier, Azure and ADLS Blog, and Google Blob. Robocopy only supports local Windows file systems and SMB via command line. It has no dashboard or UI functionality. Additionally, Symphony provides task notification completion state and system health emails, while Robocopy does not offer these features.



Integration options

Symphony provides a number of integration options, including database, webhooks, and a comprehensive ReST API. It also enables powerful, task-level post-run actions such as task chaining, HTTP POSTs (webhooks, etc), run local programs, scripts, and run remote commands via SSH. Robocopy provides a command-line interface only.

While this solution brief has compared Robocopy in depth, there are other options for migrations, such as Komprise and Datadobi. These products are often regarded as a superior option for enterprises than Robocopy, however they suffer from similar challenges as those presented above.

Both Komprise and Datadobi utilize middleware to track changes, creating a database, distributed hash table, and index. This not only adds complexity and a single point of failure, but also make these platforms “sticky”, in that they will remain in place after the migration has completed.

By contrast, Symphony’s disintermediated architecture does not require any middleware and has no single point of failure during the migration process. Command and control is outside of the data path, and the migration servers are created as close to the data as possible. This enables the accelerated and optimized performance we have discussed here.

Optimizing on-premises migrations and storage

While time is a luxury afforded to on-premises data migrations and the process is usually both simpler and faster, there are still a number of hurdles to overcome. Additionally, there are opportunities to boost performance and achieve cost savings if the migration is done correctly.

Simply migrating datasets from one expensive storage array to another may not be the best business decision. However, lack of visibility into siloed storage environments leaves IT ops teams with little alternative and results in unnecessary capital expenditure. Typically, 80-90% of stored data is either warm or cold. This data does not need to be sitting on expensive, high-performance SSDs, although it does need to be accessible and available when needed.

Symphony's comprehensive source support for file systems, protocols and object stores provides seamless conversion and mobilization from source data type to object, allowing organizations to easily leverage cost-effective object storage for large amounts of infrequently accessed data. User-generated file datasets can be ingested into object format, preserving the structure, naming and metadata of the original dataset. Object datasets may be converted back to files, utilizing this stored information. Additionally, custom metadata may be attached during the ingest operation, making the object data more queryable.

File and object data operations may be automated by Symphony policies, triggered by ReST API integration or executed directly by migration project users. Using Symphony, file and object datasets can be managed from a single policy controller, without the need to develop custom integration or have domain knowledge of object APIs.

While this may feel like the HSM of the past, it is...and it isn't. Traditional HSM had its issues with parsing all of the metadata created by the constant motion of the data.

Symphony's real-time metadata-awareness capabilities mitigate these challenges, ushering in powerful data mobilization and management that can not only migrate data between storage tiers, but also storage types, such as object and cloud object or file.

This type of data mobilization and management can enable your company to begin the slow migration to cloud, utilizing true cost savings options such as long-term cold storage, at a fraction of the cost of on-premises storage systems.

Panzura Symphony complements and extends the data management capabilities of Panzura CloudFS. It can operate alongside or independently of CloudFS. For further information on data migrations, and to optimize your storage costs, and maximize migration speed and efficiency, visit panzura.com