WHITE PAPER

Complete File Solution for NetApp NAS

Snapshot management, file catalog and search, long term archive and file reporting

Summary

NetApp has long been an industry leader in providing Network Attached Storage (NAS) file services. With its multi-protocol ONTAP data management software, NetApp provides high-performance, scalable file services (NFS, SMB/CIFS).

Catalogic has been delivering value-adding storage solutions built on NetApp technology for over twenty years. By combining two of our solutions, Catalogic provides a complete NAS offering that includes snapshot management, a searchable file index for operational file recovery, export of data to disk, tape or cloud for long-term archiving, and file analytics.

The Problem: Endless File Growth

The major challenge with NAS file data is simple: there's just too much of it. File data is growing exponentially while regulations and compliance rules are making it necessary to keep files for long periods of time. This combination creates a host of problems.

- There's so much file backup data that you can't easily find files for short-term, operational recovery
- Backing up files takes more and more time, making a mess of your backup windows
- Long-term retention for compliance is too expensive to keep on disk, so tape is still a necessity. And is cloud really an option?
- The contents of the file environment are unknown. Do you know how many files you have? What kinds of files you're storing? How old files are? Most organizations have no good idea of what they are storing.

This is a complex problem with many moving parts, but Catalogic has cost-effective tools that can have you resting easy.

Solution Overview

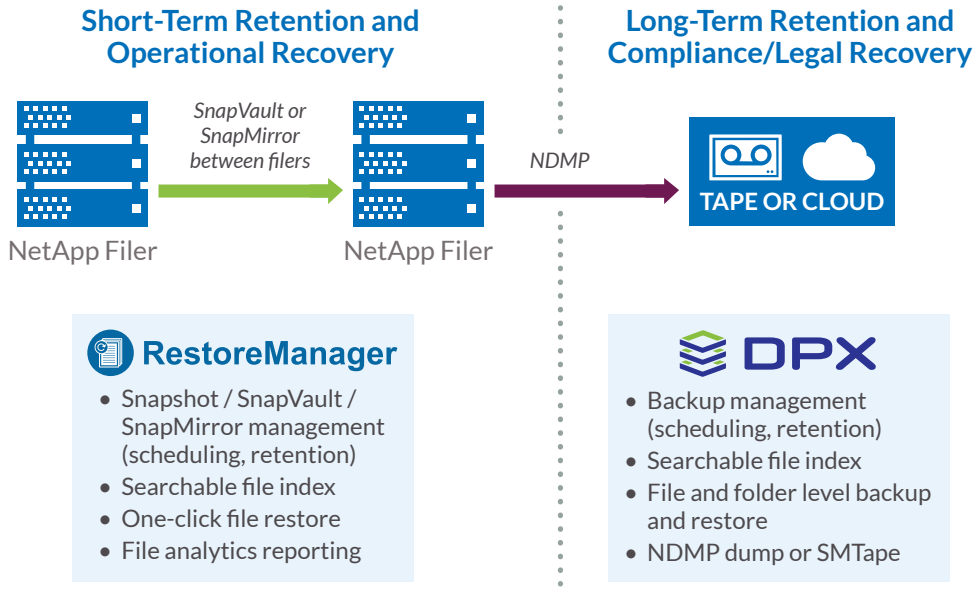
This section will provide an overview of the Catalogic solution for NetApp file environments. The following sections will provide details.

The combination of NetApp ONTAP copy technology and Catalogic software easily handles even the largest file challenge. Let's break the challenge into four parts and see how Catalogic solutions and NetApp handle the challenge. Catalogic combines RestoreManager by ProLion with Catalogic DPX for an end-to-end solution.

1. **Eliminating the File Backup Window with Snapshots**
   If file data has outgrown the ability to easily back it up every day, consider using NetApp snapshot technology. Snapshots capture file data in seconds, letting you protect files multiple times per day. RestoreManager includes NetApp Snapshot management, letting you set copy schedules and data retention.
2. **Finding and Restoring Files Easily**
   RestoreManager provides a scalable file catalog for NetApp NAS data. It indexes all copies of your files including Snapshot, SnapVault and SnapMirror copies. The index is searchable using multiple filters and wildcard options. Files are restored with a single click.
3. **Archiving File Data for Long-Term Retention**
   Moving data off disk is essential for long-term data retention. Catalogic DPX can move data from NetApp storage to tape and cloud while retaining a file-level index.

4. Understanding Your File Environment

Effective file management requires data awareness. RestoreManager provides a customizable data analytics dashboard that lets you understand what's happening in your NetApp file infrastructure.



This diagram provides a high-level look at the Catalogic solution. RestoreManager software provides file search and restore for daily operational recovery, while DPX provides long term retention of data on tape or cloud. SnapVault and SnapMirror are optional and NDMP backups can be taken directly from the primary NetApp filer.

RestoreManager: Snap Management, File Search and Operational Recovery

NetApp Snapshots, along with SnapVault and SnapMirror, are the best way to protect large file environments that are outgrowing traditional backup. But the missing piece is a file index that lets you find files from among the millions of files and versions of files spread across multiple copies on both primary and secondary storage. You may also have files stored on a NetApp AltaVault device.

RestoreManager creates a central, online file index of every NetApp snapshot, giving you a single catalog-based view into your files. You can search snapshots using multiple criteria and restore files and folders right from within RestoreManager with a single click.

What's more, RestoreManager indexes both primary and secondary storage. That's important because primary snapshots are usually only maintained for a few days. By also indexing SnapMirror and SnapVault destination volumes, you can find older versions of files that have been moved off your primary storage. RestoreManager can also index files stored on AltaVault.

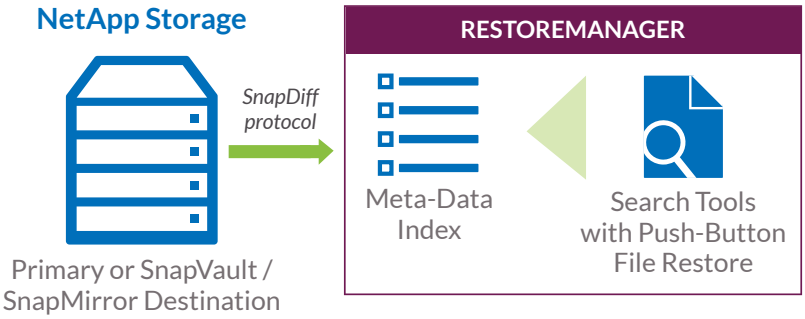
Immediately after a new Snapshot of a volume has been generated, RestoreManager uses the SnapDiff API to gather the relevant metadata from the files and folders and loads this data to its central database. Searching is now easy with this central index in place. A single click restores files to a specific folder or to their original location.

Multiple filters enable targeted searching:

- By file name, parts of the name or file path, with wildcards being permitted
- By data type or file ending: jpg, xls, doc, ppt, etc.
- By deletion period
- By creation period
- By file size
- Etc.

With RestoreManager, you can truly use NetApp Snapshot, SnapVault and SnapMirror technologies for filer backup and get away from slower, streaming backup methods that can't keep up with your file growth. RestoreManager fills the gap of not having a searchable file index.

RestoreManager Indexing



RestoreManager uses the SnapDiff protocol to create a meta-data index of all copies on NetApp primary and secondary storage. The index can be easily searched, and one or more files recovered with a single click.

DPX: Long Term File Data Archiving

While RestoreManager solves the challenges of short-term file recovery, long-term file retention is not well served by storage on spinning disk. There are many downsides to consider about disk for long term storage.

- Not cost effective for off-line or long-term storage
- Relatively short lifespans and high failure rates
- Requires special housing facilities which add to overall costs (power, cooling)
- High use of floorspace relative to storage capacity
- Frequent technology and interface updates which make older generation hardware unusable

You may keep as much as a year's worth of data on disk, but beyond that it starts to get very expensive. Even so, we often see organizations with five- and ten-year-old data sitting on disk untouched for years.

Clearly, tape and cloud are more cost-effective options. Which to choose? There are pros and cons for each, and it depends on your specific requirements. See the comparison table for details.

Tape	Cloud
Pros	
• Cost-effective in terms of storage capacity (lowest price per TB) • Low external costs (power, cooling, floor space) • Large capacity media, e.g. 12 TB per tape with LTO-8 (30 TB compressed) • Easy transport and off-site storage • 30-year shelf life (with appropriate handling) • Built-in encryption and compression • Standardized long-lifespan connection interface (SCSI, FCL) • Established technology with well-defined growth path for several future generations (new generation approx. every 2-3 years) • Backward compatible for two generations	• Pay-as-you-go model requires no initial large investment • Extreme capacity, effectively unlimited • Easily to grow or reduce capacity • No IT maintenance required • High reliability • Choice of data redundancy options • Provides long-term storage options • Technology changes are transparent to the customer • Data is easily shared

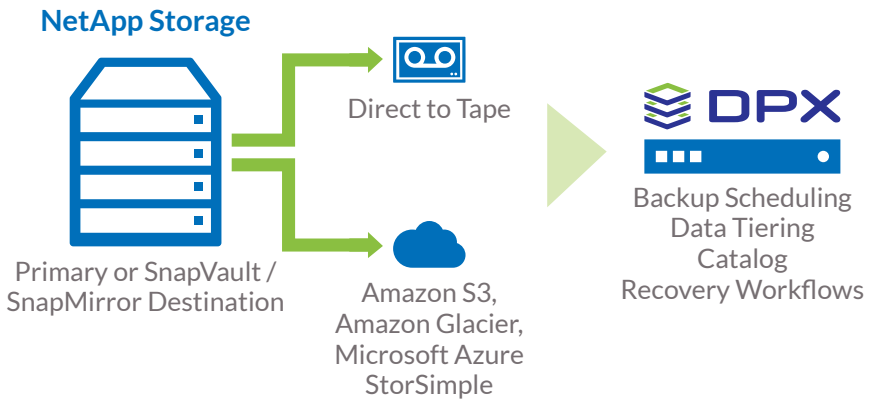
Tape	Cloud
Cons	
• Requires special tape drive or tape library • High initial investment, but quick breakeven point • Proper storage conditions required to achieve maximum shelf life • Drives and libraries require periodic maintenance and cleaning • Not as easily plugged in and brought online as disk • If re-used often (approx. 100 times), tapes are prone to wear and tear. Therefore, better suited for long-term archive. • Linear, sequential data storage plus physical loading into drive – may take several minutes to load tape, locate data and start recovery.	• On-going expenses • Some cloud-based storage may charge for retrieving data (egress charge) • Slow data access due to data being across WAN (some direct access options may mitigate this) • Faster access may require increase in WAN capacity and cost • Not suitable for quick access of large data sets, which may also add a cost premium • Data is held outside of organizational control – complete dependence on the provider • Any loss of WAN connectivity means no access to data

While there are many aspects to consider between tape and cloud, one of the key issues is how often you expect to recover data. Retrieving data from the cloud can be costly. But if you find that the majority of restores are from data center storage, cloud may be a good choice for long-term storage that only needs to be accessed occasionally. You may even consider using both, splitting your file archiving across both tape for the very long term, and cloud for medium-term.

Fortunately, Catalogic DPX can help with both.

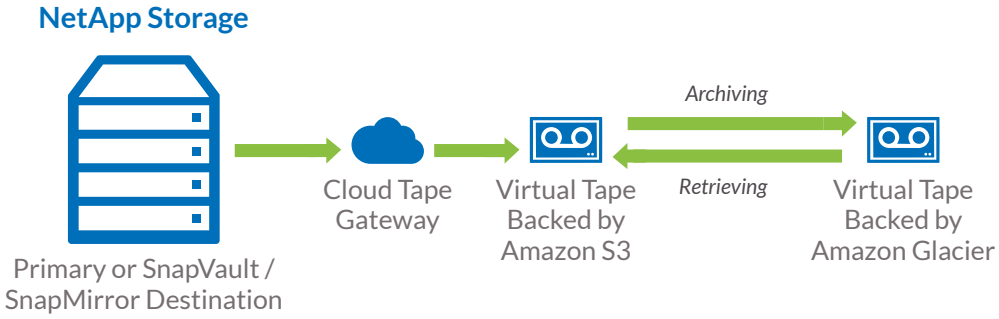
DPX provides archiving of NetApp file data to tape or cloud, including full file indexing and the ability to restore individual files. DPX handles the backup scheduling to offload data and maintains a file catalog for individual file restore.

When sending data to tape, DPX supports high-speed data transfer using a tape drive or library connected directly to the NetApp filer via iSCSI or fibre channel. Other solutions such as Veeam require a Windows proxy server to transfer data, adding significant latency which slows data transfer rates.



DPX provides easy backup of NetApp file data to either tape or cloud. The DPX server manages the backup scheduling and data tiering and maintains a catalog of files. Recovery workflows are also managed by DPX, whether restoring from tape or from cloud.

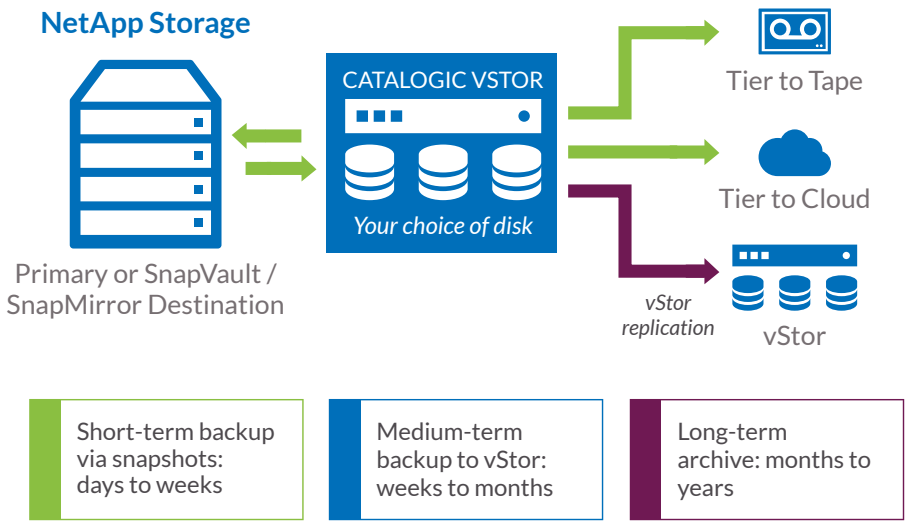
DPX takes advantage of the VTL capabilities of Amazon and Azure to provide tape-like data management in the cloud. For Amazon, there is a disk-based VTL tier in S3 and a tape-based tier in Glacier which is lower cost. DPX manages the tape lifecycle and can move backups from S3 to Glacier and also retrieve tapes as needed.



If you wish to use Amazon as an archive destination, you have the option to tier backup data from S3 to Glacier. Glacier is suited for very long-term storage. DPX will control the process and lets you recover seamlessly from either tier.

If file level recovery from tape/cloud isn't required, DPX can use a volume-based backup (SMTape) which provides faster data transfer. However, this method requires recovering the entire data volume in order to access files, so it should be considered carefully before implementing.

As an alternative to direct backup to tape or cloud, file data can be staged on disk using the Catalogic vStor software-defined backup appliance. This allows you to move data from NetApp FAS storage onto lower cost, high capacity disk, such as the NetApp E-Series. This near-line backup data can be quickly accessed for file restores. As data ages, it can be tiered off to tape or cloud. You can even use vStor to replicate data from one site to another.



With Catalogic DPX, you have full archive flexibility to make use of whatever backup media best serves your organization's needs.

RestoreManager: Understanding Your Files with File Analytics Reports

You can't fix what you don't understand. Too often, file storage is a big unknown. Most organizations can give a good guess as to how much disk capacity is taken up by files, but they have little to no idea about the files themselves.

If you don't know what you have, you don't know if you have a problem. For example, what if over half your file data hasn't been accessed in years? What if a large percentage consists of

file types that you shouldn't be storing? Can you easily tell which departments or users are consuming the most capacity?

RestoreManager can solve these mysteries and many more. As it gathers data for file search and restore, RestoreManager makes use of the information to generate a host of valuable reports. Multiple pre-built reports are included. But even better, the Kibana reporting tool allows you to customize reports to pinpoint the precise information that you need. Report templates can be easily transferred and Catalogic will assist with creating whatever special reports you need to optimize your NetApp file environment.

Conclusion

File environments are a challenge to properly protect and manage in the face of relentless file growth. With Catalogic solutions designed specifically for NetApp file environments, you can find cost-effective solutions that will protect your critical file data now and into the future.

50 Tice Boulevard, Suite 110
Woodcliff Lake, NJ 07677 U.S.A.
201.249.8980
catalogicsoftware.com

United Kingdom: +44 (0) 207 712 1667
Germany: +49 (0) 2154 880829 0