

YES, THERE IS A MUCH BETTER WAY TO PROTECT LARGE FILE SYSTEMS!

Starfish is a unique software application for managing unstructured data at any scale, from departmental file shares to the world's largest supercomputing sites. By combining a metadata catalog with batch processors and data movers, Starfish delivers a holistic approach to unstructured data management.

Data protection is just one of many areas where Starfish has significant advantages over traditional, industry-leading point solutions especially in extreme environments. This paper focuses on the benefits of Starfish, specific to data protection.

A FRESH, SOPHISTICATED APPROACH TO DATA PROTECTION

Most data protection architects will tell you that there are two ways to protect large file systems:

- Enterprise backup with NDMP and
- NAS-to-NAS snapshot + replication

both of which are expensive and imperfect.

Starfish's holistic approach to unstructured data management enables more sophisticated and flexible data protection strategies.

The Starfish metadata catalog reveals what has changed in the file system over time, while providing insights into the relative value and data protection requirements of specific files, types of files, or curated collections of files.

Starfish data movers copy files in parallel to one or more destinations with many options for scheduling and for pre- and post-processing. As files are successfully backed up, the metadata catalog is updated, such that for every discrete file, Starfish records whether and where there are backup copies.

Recovery and retention management are similarly enabled by the insights provided by the metadata catalog and by the scalability and flexibility of the data processors and data movers.

WHAT'S WRONG WITH ENTERPRISE BACKUP WITH NDMP?

NDMP has been the best practice for backing up big enterprise NAS systems for many years, but. . .

★ **It is incredibly inefficient.**

Backup and restore operations are needlessly slow and cumbersome.

NDMP has numerous scaling limitations.

★ **Vendor lock-in**

You get locked in to both the storage vendor and the backup software vendor.

★ **Backup media encoding is proprietary.**

Files can only be accessed by first restoring from backups.

More vendor lock-in until retention periods expire.

★ **Many scale-out and HPC file systems don't support NDMP.**

WHAT'S WRONG WITH NAS TO NAS SNAPSHOT+REPLICATION?



NAS-TO-NAS SNAPSHOT and REPLICATION has been the "easy button" for shops with aggressive RTO and RPO requirements, but. . .

★ **It is expensive.**

Especially when the storage devices are only available with all-flash media.

★ **Vendor lock-in**

You must buy both devices from the same vendor.

Migrating to a new vendor is especially difficult when devices are replicated over a WAN.

★ **Retention is hard to manage in a snapshot engine.**

High-churn data gets trapped in the snapshot engine, gobbling up expensive storage space

Files that require longer-term retention need to be backed up separately.

★ **There are no tools for browsing or searching among file versions.**

★ **Many scale-out and HPC file systems do not support snapshots and/or replica on.**

STARFISH

DATA PROTECTION CLAIMS TO FAME

Starfish has many advantages over conventional data protection solutions, especially when protecting large, complex file systems. Starfish's metadata catalog delivers a superset of the functionality

commonly found in enterprise backup system catalogs, while its data mover is far more flexible and scalable than anything found in storage devices or the leading enterprise backup solutions.

Any-to-Any Backup

Backup or replicate NAS and HPC file systems to ANY storage device. Replicate between file systems from different manufacturers or to object stores that are on premises or in the cloud.

File System Protocol Support

Starfish data movers run natively on Linux and Windows enabling comprehensive support for POSIX permissions and ACLs. Starfish supports NFSv3, NFSv4, CIFS, SMB-2, SMB-3, and native clients for HPC file systems.

Object Store Protocol Support

Starfish supports S3-style, Azure, and Swift object stores as both source and target. Starfish makes it easy to specify object storage classes of service and natively supports "deep archives".

Incremental Forever Backup Strategy

Starfish tracks the data protection status of each and every file. There is no need to run full backups on files that are already protected, unless, of course, your policy calls for multiple backup copies.

Super-efficient Change Detection

Starfish tracks changes in large, complex file systems using state-of-the-art scanning and difference detection. Rapid change tracking enables much more aggressive RPOs (recovery point objectives).

Version and Point-in-time Recovery

Starfish maintains a history of the file system over time allowing for both point-in-time recovery and individual file version recovery. Starfish can leverage snapshots on both the source and target devices and take advantage of S3 versioning.

Capacity Optimization

Starfish supports compression and file-based deduplication. Starfish is sensitive to symlinks, hardlinks, and sparse files.

Unlimited Performance and Scale

Starfish moves data in parallel across any number of threads and any number of servers, making Starfish suitable for billions of small files and multiple Petabytes of mixed file sizes. Data movers can be added on the fly for bursty performance needs.

Prioritized Disaster Recovery

Starfish gives you granular control over the recovery process. Recover files in priority order or selectively promote object storage backup copies to high-performance storage for cloud-based disaster recovery. Restore to original or alternative locations.

Metadata-driven, Ultra-flexible Policies

Starfish's metadata catalog allows one to be ultra-specific about what gets backed up, where, and when. Use different rules for different types of data, projects, users, etc.

Flexible Retention Policies

Retain data forever or selectively prune backup copies based on any set of metadata in the Starfish data catalog.

No Proprietary Encoding or Obfuscated Naming

Starfish gives you control of how files are written to backup targets. Your data is never written in proprietary formats, you control how your file paths are mapped to buckets and object names.

Take Better Advantage of Secondary Storage

Conventional wisdom says that backups and archives are separate processes with separate storage. Starfish allows you to combine data protection workflows with archiving, disaster recovery, collaboration, and cloud bursting.



To learn more about how Starfish can reduce storage costs and make data more accessible and manageable, visit us at <http://www.starfishstorage.com>

Client / Industry	Source	Destination	Petabytes	File Count
Leading Medical School - Central Research Computing	EMC Isilon Scale-out NAS	EMC ECS Object Store	33	3 Billion
Leading pharmaceutical firm - Drug Discovery	Weka.IO HPC File System	Amazon S3 Object Store	6	9 Billion
Major Motion Picture Animation Studio	EMC Isilon Scale-out NAS	Qumulo Scale-out NAS	12	9.5 Billion
Biomedical Research Laboratory	EMC Isilon Scale-out NAS	Qumulo Scale-out NAS	26	6.2 Billion
Global Hedge Fund VAST All Flash Array	VAST All Flash Array	IBM COS Object Store	2	1.3 Billion

Real life examples of any-to-any backup at multi-PB scale

Below are stats from actual Starfish client sites, using a mix of enterprise NAS products, next-gen file storage solutions, on-premise object storage, and cloud object storage.