SCALITY
NAS ARCHIVER

SCALITY

Archiving NAS Data to Object Storage

Archiving NAS Data
to Object Storage

I. Highlights

Network-attached storage (NAS) has become dominant in the enterprise for a wide variety of file-based collaboration use cases. Customers depend on NAS "filer" systems for their reliability in protecting data and the fast interactive performance they deliver for the right use cases. However, ownership of "tier-1" NAS systems can become costly, and those systems must, therefore, be deployed as efficiently as possible for the use cases and data for which they were designed. A well-known fact in the enterprise is that only a small percentage of data is actually active (accessed frequently) on most storage systems, with some studies showing that only somewhere between 20-40% of data is typically active.

This means that up to 80% of data on NAS systems may be inactive, dormant (never accessed), and could be archived. Archiving that data to a highly reliable, always-on and scalable object storage system can bring the best of both worlds: retaining ready availability while offloading the expensive NAS. This paper describes a data protection model using Scality RING as the archive target for inactive NAS data to achieve 50% and higher TCO savings versus traditional NAS data protection, along with freeing up a commensurate amount (50%+) of NAS tier-1 storage capacity.

II. NAS Use Cases in the Enterprise

Over the last 25 years, NAS has become the standard and incumbent storage technology for a wide variety of enterprise file use cases. The classical NAS use cases are those where file data needs to be stored and shared by groups of office users, such as home directories ("shares") and content repositories. NAS has also become standard for teams of knowledge workers collaborating on file data for design automation applications (CAD/CAM for electronic or mechanical designs), media design, source code repositories and also for image, video and static file content for websites and other media delivery.

As content has grown in capacity and variety, NAS has extended into scale-out architectures to support growth from initial 10s to 100s of terabytes of capacity into petabyte-scale. Because file systems do incur high overhead for maintaining all of the requisite folder, file inode and associated metadata, scalability beyond a few petabytes of capacity or tens of thousands of files per file system on a single scale-out NAS has remained challenging. These limitations have driven the growth of object storage as a more scalable solution for high-capacity (multi-petabyte) applications and use cases.

NAS filers perform admirably for the use cases for which they were designed, and enterprises depend on them for these types of deployments. While storage area network (SAN) technology is often deemed as "tier-1" high-performance storage, it is completely appropriate to classify NAS as tier-1 storage for the use cases where it fits. Lower cost and lower performing tier-2 and 3 storage solutions include nearline systems, backup appliances, cold storage and tape libraries — none of which are appropriate for the NAS offload scenario described above.

III. Data Protection for NAS

Most modern NAS systems provide built-in snapshot technology to protect data against inadvertent loss from accidental deletion, overwrites due to user or application error, or even from viruses. Snapshots were popularized by NetApp® who also provided low-overhead (copy-on-write) technology for space savings, plus fast snapshot restores to provide an efficient and fast way for end-users to retrieve lost files. A typical data protection strategy for NAS starts with a snapshot policy, such as hourly or daily snapshots, based on the criticality of the data and the resulting recovery point objective (RPO). It is important to note that snapshots do indeed consume space on the NAS filer, and space is reserved from the usable space on the NAS file system volumes. Despite that, the TCO savings model below does not consider the overhead of snapshots since the majority of the TCO savings will result from the next part — traditional backups.

To backup NAS file data, enterprise IT administrators will use a backup software solution from any of the popular vendors, such as Veritas, Commvault, IBM, Rubrik and others. Commonly, the backup software is licensed based on the volume (capacity) of data to be backed up, so as more NAS data needs to be managed by the backup software, the more expensive the solution.

A typical enterprise backup strategy might consist of:

- **Weekly backups:** Maintaining one month of weeklies, so this implies storing 4 full backups
- **Monthly backups:** Maintaining on year of monthlies, hence 12 full backups
- **One yearly backup:** One additional full backup

Deduplication and compression can be used to reduce the space requirements of all of these backups, as can incremental backups. A reasonable assumption is that a business can achieve 50% reduction in space required to store all of these backups.

For many businesses, these backups are important enough that an additional copy must be stored offsite for disaster protection and compliance purposes. Many backup solutions, whether software- or appliance-based, provide data replication capabilities that enable them to store this "D/R" copy of backups offsite in a second data center data center or a D/R copy in the cloud.

Superior Data Protection Approach:
NAS Offload Archiving to Scality RING

Instead of the traditional backup approach described above, we propose a solution to offload inactive data from the NAS and maintain it on Scality RING distributed object storage solution as an extended online archive.

Scality RING provides extensive data durability and availability capabilities to ensure that archives are protected for long-term storage and at high scale. A combination of capabilities for data protection—including erasure coding, local replication of objects (files), self-healing after disk or server failures, a redundant design with no single points of failure—ensures that data is maintained reliably and is always available. Most industry experts agree that object storage solutions, such as Scality RING, are so durable that traditional backups are not required to further protect the data stored on the system. For more than a decade, Scality RING has provided mission-critical applications with proven nonstop operations. Scality also provides a 100% uptime guarantee when combined with the Dedicated Care Service (DCS) support agreement.

To transparently detect and migrate inactive data to the RING, Scality provides the Scality NAS Archiver as an optimal solution for NAS offload. The Scality NAS Archiver provides a wide range of migration policies based on file size, file types, file age and more, and can provide either stub-based or stubless archiving from the NAS filer to the RING. Stub-based archiving ensures that the NAS filer continues to operate as normal— and that applications can retain seamless access to any data that has been moved from the NAS to Scality RING — simply by referencing the file pointer (stub) maintained in the file system. For use cases where files on the NAS are very small, stubless archiving can help further reduce space overhead on the NAS by not imposing the small overhead of the stubs themselves in the NAS file system.
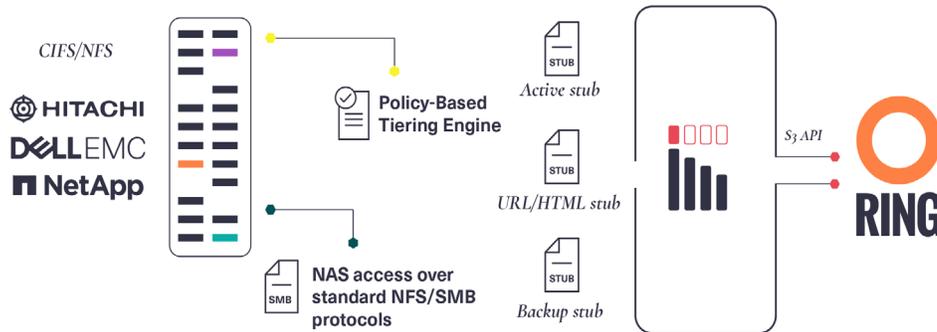
This solution fully preserves normal application and end-user access to files stored on the NAS for both active and inactive data. Therefore, we now have a fully transparent tiered solution consisting of a fast tier-1 NAS file system, coupled with a scalable, online and cost-effective tier-2 archive.

To be clear, the remaining active data on the NAS filer should still be protected using normal techniques, including snapshots, backup and offsite replication (if needed). However, since inactive files are now offloaded from the NAS to the RING, we now have several immediate benefits from the offload:
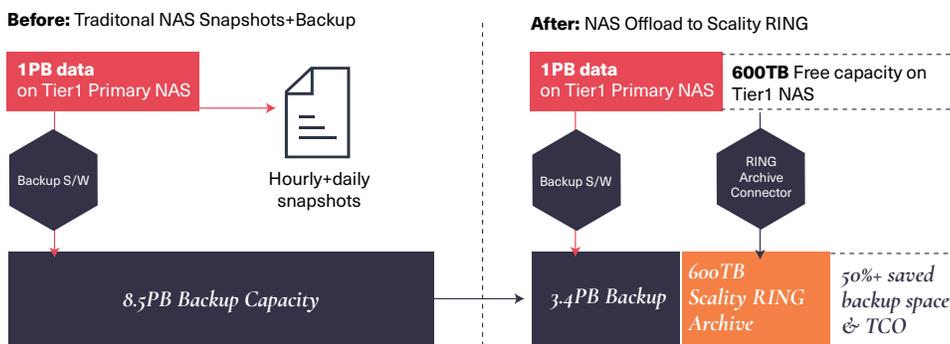
- The NAS file system now has 50-80% capacity free. This costly tier-1 free space can now be used for new file data that truly requires NAS access and performance characteristics.

- The backup software employed can be right-sized for 50-80% lower capacity licensing, thereby reducing CapEx.

- The archive data on the RING is always online and always up-to-date since the Scality NAS Archiver will transparently move files that hit the migration policy criteria.

- This implies we have a super-optimized data protection strategy with a zero RPO and zero RTO (recovery time objective) for all data moved from the NAS to the RING.

- If the NAS filer fails or becomes inaccessible, all of the data moved to RING can be accessed directly and natively using standard file protocols over RING scale-out file system connectors (NFS or SMB), thereby providing an ideal business continuance solution in the event the NAS fails or is out of service.

IV. TCO Savings Example

Here, we model a simple scenario where 1PB of file data is stored on a NAS filer. Assuming 50% space reduction from compression and deduplication, a standard data protection policy consisting of hourly and daily snapshots, with the backup policy described earlier, would incur the following space overhead for protecting the data:

I.	4 x weekly backups: 4 x 1PB = 4PB	» 2PB (50% reduction)
II.	12 x monthly backups: 12 x 1PB = 12PB	» 6PB (50% reduction)
III.	1 x yearly backup: 1 x 1PB = 1PB	» 0.5PB (50% reduction)
Total capacity for data protection =		8.5PB

Such an approach requires 8.5PB of capacity to protect the 1PB of file data on the NAS. Note that we choose to ignore the offsite data replication aspect of data protection, which would obviously further increase the storage requirements.

Before: Traditonal NAS Snapshots+Backup   **After:** NAS Offload to Scality RING

1PB data on Tier1 Primary NAS → Backup S/W → Hourly+daily snapshots

8.5PB Backup Capacity

1PB data on Tier1 Primary NAS   **600TB** Free capacity on Tier1 NAS

Backup S/W   RING Archive Connector

3.4PB Backup   *600TB Scality RING Archive*   *50%+ saved backup space & TCO*

To look at the TCO savings using the offload of inactive files to the Scality RING archive, we can make the following assumptions:

- 60% of the data is inactive (a conservative estimate, as this may be 80% or more in many environments)

- Hence, 600TB of data is moved to the RING using the Scality NAS Archiver (with stubs maintained on the NAS to ensure transparent access is maintained to these files)

- The remaining 40% of the data is, therefore, active (400TB) and maintained on tier-1 NAS. It is still protected using the same snapshot+backup strategy as above (with the same 50% space reduction approach)

I.	4 x weekly backups: 4 x 0.4PB = 1.6PB	» 0.8PB (50% reduction)
II.	12 x monthly backups: 12 x 0.4PB = 4.8PB	» 2.4PB (50% reduction)
III.	1 x yearly backup: 1 x 0.4PB = 0.4PB	» 0.2PB (50% reduction)
Total capacity for data protection =		3.4PB

If we now add the 3.4PB for traditional backups to the 600TB of archives stored on the Scality RING, we have a total data protection footprint of 4.0PB of capacity. This is over a 50% reduction in space (and cost) as compared to the 8.5PB of data protection capacity required in the traditional scenario.

V. Summary

This Scality RING NAS archiving solution provides such compelling TCO savings for users of enterprise NAS systems that it pays for itself and more. It provides not only a transparent two-tier storage solution, preserving the normal performance and access characteristics of the NAS filers for users and applications, but it also introduces a scalable, long-term, durable and cost-effective solution for archiving on the RING object storage solution.

Perhaps more important than the direct TCO savings, the NAS filer now has 60-80% free capacity, which can be used for additional tier-1 file data. This opens up new applications and use cases that the enterprise can use for its key business goals. Moreover, while all data archived to Scality RING is fully and seamlessly accessible through the NAS without any changes, this also provides a super optimal zero RPO/RTO solution in the event the NAS filer fails or becomes inaccessible.

For additional information on the Scality RING and NAS Archiver, please visit: https://www.scality.com/products/ring.