*E-Guide*

# Cloud Storage and Backup Market Rundown

ComputerWeekly.com

# In this E-Guide:

Your organization's information is arguably its most valuable asset, which means that the amount you rely on your data should be reflected in the care you take protecting it.

Whether you are an expert or just starting out, there are elements and trends of cloud storage, replication, and snapshots that are always changing and improving, as well as vendors and solutions that can improve  the quality and efficiency of your storage and data protection strategy.

Check out this e-guide for a comprehensive overview of some of the storage industry's most fundamental topics to make sure you are taking advantage of everything they have to offer your organization.

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous

# Cloud storage 101: NAS file storage on AWS, Azure and GCP

*Antony Adshead, Storage Editor*

Despite the many and rapid changes in storage in recent years, such as the advent of flash and the move to cloud storage, there are still some fundamentals in place. Among these are the basics of how data is accessed, whether by file, block or object.

Object storage has been a rising star among these. It forms the basis of much of the basic storage provision offered by public cloud services. In the case of Amazon Web Services S3, it has even become something of a de facto standard that is in use more widely than just the AWS cloud.

But file and block access storage are still needed for particular use cases and make up the vast bulk of stored data, in the datacentre at least.

Yet organisations also want to use cloud compute and storage capacity and to burst workloads to the cloud when necessary. In many cases, that will involve applications that haven't been developed as cloud-native, and so file and block storage will be needed.

In this first in a series of articles, we will look at file access storage provided in the big three public clouds: AWS, Microsoft Azure and Google Cloud Platform (GCP).

Other articles will look at virtual storage appliances and cloud instances from storage players in the cloud, as well as NAS gateways and distributed file systems that offer file access cloud storage by other methods.

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous

# Overview: Similarities and differences

All of the big three public cloud providers – AWS, Azure and GCP – offer native network-attached storage (NAS) storage services.

All three also offer higher-performing file storage based on NetApp storage.

Where Azure is different is that it provides file storage caching, aimed at providing low-latency access to a set of files in a single namespace, and it provides these in a number of service levels.

## AWS

Amazon's two main file storage offers – EFS (Elastic File Storage) and FSx (for Windows and Lustre) – are both Posix-compliant, which means they work with applications that demand, for example, file permissions, file locking capabilities, and a hierarchical directory structure via NFSv4.

Use cases targeted include big data analytics, web serving and content management, application development and testing, media workflows, database backups, and container storage.

EFS is NFS access file storage for Linux applications that can run on AWS compute instances or on-premises servers. It can scale to petabytes and comes in two service levels – standard and infrequent access (IA), with automated tiering between the two to place files in the most appropriate for their usage profile.

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous

AWS says access to files is parallelised to achieve “high levels” of throughput (10GBps quoted) and input/output (I/O) performance (500,000 IOPS). It says the cost can be 8c per GB per month, assuming an 80/20 split between IA and Standard storage.

Amazon FSx for Windows File Server provides file storage accessible via the Windows-native SMB protocol and delivers features such as Access Control Lists (ACLs), user quotas, user file restore and Active Directory (AD) integration. Flash and spinning disk hard disk drive (HDD) media options are possible, and FSx storage is accessible from Windows, Linux, and MacOS compute instances and on-premise hardware.

Claimed performance comprises sub-millisecond latency, tens of GB per second throughput and millions of input/output operations per second (IOPS).

Data is encrypted at rest and the service claims compliance with ISO, PCI-DSS, SOC and HIPAA.

Amazon FSx for Lustre is targeted at file-based use cases such as machine learning and high-performance computing (HPC). It integrates with AWS S3 as a bulk data store at more cost-effective rates, with data presented in file format in FSx for Lustre.

Data is accessible from EC2 instances and from on-premise locations.

## Microsoft Azure

Azure’s cloud file storage options include native and NetApp-based performance options as well as varying levels of caching services.

Azure File provides fully managed file shares in the cloud accessible via Server Message Block (SMB) or REpresentational State Transfer (REST) that can support cloud or on-premise deployments of Windows, macOS and Linux.

Two service levels are offered in Azure File – standard and premium.

Being a Microsoft service you get the integrations you'd expect, such as Active Directory, and Azure positively encourages "lift and shift" of applications and data that can use Azure Files.

Meanwhile, Azure NetApp Files is billed as "enterprise grade" and provides file storage for Linux and Windows compute based on NetApp storage in the Azure cloud. It is aimed at performance-intensive applications such as SAP HANA, databases, HPC apps and enterprise web applications.

Access is via SMB and NFS and there are three performance/cost tiers available – standard, premium and ultra.

Microsoft Azure also offers some file storage caching services that are intended to provide speedier access to data for high performance workflows.

Azure HPC Cache is an NFS-connected service that provides single namespace storage for on-premise NAS or Azure-located application data, which can be file or Blob (object).

Meanwhile, as a result of Microsoft's acquisition of Avere in 2018, Azure offers a couple of file-based caching type services based on its technology.

Avere vFXT for Azure is billed as "a high-performance caching service" and is a software-based service iteration of the FXT Edge Filer. The idea is that vFXT is used as a cloud-

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous

based file access cache that can allow HPC applications to run without being re-factored for the cloud. It is optimised for read-heavy workloads and presents a single namespace to applications.

Azure FXT Edge Filer is a hardware product and so falls slightly out of this survey. It is something like co-located hardware, offered as a service and is presumably the underpinning for the vFXT.

FXT Edge Filer works with customer NAS and Azure Blob and Amazon S3 storage to act as a high performance cache for HPC workloads. It will scale up to 24 nodes to provide claimed millions of IOPS and hundreds of GBps throughput. FXT comes in two models that differ chiefly in the amount of RAM and storage capacity.

## Google Cloud Platform

GCP's Cloud Filestore offers two performance tiers of NFS-connected file storage with up to 64TB of capacity per share. Premium offers much higher throughput and IOPS than standard, with 1.2GBps vs 100MBps read for the former and 60,000 vs 5,000 IOPS for the latter. Stated availability is 99.9% for both tiers.

Google is a bit more modest in its proposed use cases than some of the AWS and Azure cloud file storage offers. GCP targets video rendering, application workloads, web content management and home directories.

If you want more than the basic file storage offered by GCP, NetApp Cloud Volumes are also available. This is NFS and SMB-connected for Linux and Windows application workloads.

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous

NetApp Cloud Volumes on GCP comes in three performance/cost tiers – standard, premium and extreme at $0.10, $0.20 and $0.30 per GB per month and range from 4,000 to 32,000 IOPS and throughput of 16MB to 128MB per TB.

**Read more about cloud storage**

- Cloud storage 101: File, block and object storage from the big three public cloud providers: AWS, Azure and GCP. We look at what's on offer and the use cases they are aimed at.
- Avoid the cloud compliance trap. When you hand over data to a cloud provider, you don't hand over responsibility for legal and regulatory compliance. Beware of falling into a cloud compliance trap.

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous

# Snapshots 101: Array vs backup software

*Antony Adshead, Storage Editor*

Nearly all enterprise and mid-range storage arrays come with snapshots. But that's not the only place you can manage them from.

Backup software also comes with snapshot functionality, and in fact some of the more recent entrants to the market have built their approach to backup around snapshots.

Snapshots capability in backup products starts with the ability to manage and copy storage hardware makers' snapshots. This is the case with Veritas, IBM's Spectrum Protect and EMC's Networker. Things get more sophisticated with, for example, Commvault, which can manage a range of array makers' snapshots from a single console.

Then there is the more recent wave of data protection products that come as appliances and are somewhat akin to hyper-converged nodes. These products – such as those from Cohesity and Rubrik – could be said to be snapshot-centric and base their backups around them. Veeam also sits somewhere near these suppliers and their approaches, but without the hardware form factor.

Before looking at backup suppliers' snapshot capabilities, let's run over some snapshots fundamentals.

## Snapshots basics

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous

A snapshot is a copy of the state of a unit of storage (drive, volume, etc) at a specific point in time.

To be more accurate, snapshots are often comprised of pointers to an existing copy or copies – an original and/or subsequent snapshots – as well as actual storage blocks, such as those that have been deleted and need to be kept to rebuild the full picture at a specific point.

Storage arrays usually keep a set amount of snapshots (256 or 1,024, for example) and customers can roll back to a previous version if required. Depending on how often they have been made, that could give several days' worth of coverage to restore to.

**Where is it best to manage snapshots from – the array or your backup product?**

Snapshots taken at the array are likely to be quick to take, with a lower input/output (I/O) tail as they are written locally to the same storage.

They will also be quick to access for the same reasons.

But snapshots on the array could become inaccessible if the array is effectively a single point of failure. Array-native snapshots are also likely to be crash-consistent, with a possibility that they lack essential components to allow for a full rehydration.

Snapshots made from backup software are likely to have a higher overhead on resources as they are managed from and written to other locations on the network.

They are also likely to take slightly longer and have more impact on production because they will be application-consistent. In other words, the snapshot will ensure the application is quiesced and everything needed to successfully rebuild data to that point in time is taken.

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous

# Array makers' snapshot functionality

All arrays come with snapshots and, in general, they offer a similar set of choices.

Customers can set how many shots they want to retain (up to certain limits), whether they want them to be copy-on-write or redirect-on-write and in some cases (such as NetApp) whether they want regular snapshots, archive copies or clones that can be used for disaster recovery (DR) failover.

# Snapshots in backup products

### Cohesity

Cohesity's approach is based on its DataPlatform and SpanFS file system, with snapshots and its SnapTree functionality operating as a metadata-based functionality within it.

In SnapTree, where normally recovery from snapshots would require the rebuilding of a large chain, Cohesity shifts some of that onto a chain of metadata pointers to gain a claimed boost in recovery times.

It calls SnapTree "distributed redirect-on-write". Changes are written to new blocks every time and the "distributed" part refers to the fact changes are written across the file system.

Cohesity has native snapshot management integration with Pure Storage arrays, as well as HPE Nimble flash storage and Cisco HyperFlex NAS.

### Commvault

In addition to traditional backup, Commvault allows customers to take backups of array snapshots. These can be stored elsewhere in case of problems on the storage array that cut access to snapshots held there.

Also, Commvault has IntelliSnap, which allows customers to manage multiple supplier array- and cloud-based snapshots through a single console and claims it can be used to base data protection on snapshots rather than traditional backup.

IntelliSnap can discover snapshot capabilities across the storage estate and cloud, and create long-term snapshot backups and archives from snapshots using commodity disk, cloud or tape.

### EMC

EMC offers Avamar and NetWorker, both of which will work with VMware snapshots – as will most mainstream backup products – but of the two, the enterprise-level NetWorker has the more developed snapshot management capability. Its Snapshot Management automates the generation of point-in-time data snapshots and cloning on supported storage arrays such as EMC VNX, XtremIO, and Symmetrix.

NetWorker's snapshot capability includes its module for Microsoft applications that uses Volume Shadow Copy Service (VSS) to protect Exchange, SQL Server, Active Directory, SharePoint and Hyper-V. Then there are its PowerSnap Modules for SAP, Oracle, SQL Server and IBM DB2 applications and databases.

NetWorker also has what it calls "snapshot-assisted backups" although these appear to be snapshots that are stored elsewhere for added protection or, for example, disaster recovery purposes.

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous

### IBM

IBM's Spectrum Protect (formerly Tivoli Storage Manager, or TSM) features Spectrum Protect Snapshot, which allows backups to be created from a range of storage array makers' snapshots, as well as those from database and applications such as enterprise resource planning (ERP) software. But it appears to need IBM storage hardware or a General Parallel File System (GPFS) as a target.

### Rubrik

Snapshots are at the core of Rubrik's approach to backup and data protection, using them to create a stable system image to make backups from.

In VMware environments it uses VMware's vSphere APIs (application programming interfaces) for Data Protection (VADP) to create a snapshot of virtual machines (VMs) from which it makes backups.

For Windows environments, Rubrik wrote its own VSS agent for application-consistent snapshots. Writing its own VSS agent helps gives a claimed boost in immunisation against failed snapshots.

Also, Rubrik leverages snapshots in NAS backup where it works with NetApp OnTap snapshot and differential information to know what new blocks/files to backup since the previous job.

### Veeam

Veeam has a range of functionality that aims to make storage array makers' snapshots easier to work with. This includes the ability to copy storage snapshots to other locations and to restore items from it at fine levels of granularity.

You can configure a backup job to maintain a snapshot chain on the storage system in addition to regular backup files.

It is also possible to create snapshot-only jobs to build a chain of snapshots on the primary storage array and on secondary storage as an option.

Additionally, you can set preferences for numbers of snapshots retained and automatic deletion beyond set limits.

Meanwhile, Veeam's Explorer for Storage Snapshots has integrations with Dell EMC, NetApp, HPE and IBM products.

Veeam also has Snapshot Hunter, which can detect and remove orphaned snapshots that remain after backup or replication job sessions to save space.

### Veritas

In NetBackup, Replication Director uses the Veritas OpenStorage API to manage snapshots from hardware arrays and use them as roll-back points. Snapshots can also be replicated to other storage locations.

Meanwhile, Veritas's multicloud CloudPoint product integrates with NetBackup and includes snapshot management and orchestration for storage array-based protection and in the cloud.

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous

CloudPoint can connect to array-based snapshots in Hitachi Data Systems G-Series, HP 3PAR, Pure Storage, Dell EMC Unity and NetApp FAS arrays.

Incidentally, NetBackup uses its Snapshot Client to create an image of a client volume, then backs up data from the snapshot. According to Veritas, that means users can access primary data without interruption while data on the snapshot volume is backed up.

**Read more about snapshots**

- Cloud snapshots and backups: You often need more data protection than native cloud services will give. But should you choose backup or snapshots? And what about third-party backup in the cloud?
- Storage 101: Snapshots vs backup: We go over the basics of snapshots – they're a quick and accessible way of protecting data, but they're not a substitute for backup. So how do you combine the two?

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous

# Storage 101: Replication vs backup, and synchronous vs asynchronous

*Antony Adshead, Storage Editor*

Backups, snapshots, cloning and replication are all valuable ways of protecting the organisation's data.

In this article, we'll look at replication, in particular between storage arrays. Key to this will be to define it and present the pros and cons of replication with reference to other methods of data protection.

All too often in IT there's a lack of clarity over what exactly a technology is, or does. The latter is the important bit, because it's how different technologies function that can determine how they fit together.

## Replication versus snapshots

Replication is fundamentally a method of producing a clone of a unit of storage. In other words, it is a replica of a drive, volume or logical unit number (LUN), for example. In most cases, what is being strived for is an exact copy – maybe almost immediately, maybe just eventually.

That makes a clone or replica different from a snapshot, because snapshots in most cases can only become a usable replica following some sort of rebuilding process. That's because snapshots comprise an original copy of the drive or volume plus updates to it, as well as

perhaps deleted blocks that have to be reincorporated to create an accurate copy from a previous point in time.

The idea is that snapshots can be re-built and rolled back pretty quickly, but they're not there as an alternative, usable copy of the source media. Meanwhile, clones and replicas often are.

The simplest clone/replica of all is when, for example, a developer needs a database to run some test queries on. They can clone an exact copy of an existing production database and do what they want with it in the test environment. That clone will be an exact replica of the database at the point in time it was created, but it will not likely ever reflect any further changes to the source copy.

But at the other end of the scale in terms of creating an available, working clone is synchronous replication. This sees data written to two or more units of storage as near to simultaneously as possible to provide a working copy that can be failed over to on-the-fly.

Obviously this comes at a price in terms of cost and technical complexity and there are limitations, as we shall see. But this is often what we mean when we talk about replication.

## Replication versus backup

Can replication replace backups? The simple answer is no. Backups and replication (and maybe snapshots too) have to complement each other.

Because replication can be almost continuous and creates a near real-time copy, it can also make a replica of corrupted or infected files. In that case, you need a version to roll back to.

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous

That could be derived from a snapshot, but then they also need to be underpinned by backups – and replication is often costly, so it may be that only certain datasets are replicated while everything is backed up.

## Synchronous versus asynchronous array replication

In synchronous replication, data can be written to the second site as soon as it hits cache in the primary site. On receipt, the second site sends an acknowledgement to the primary site storage and the host where the change originated. It's the method of replication that comes as close to writing multiple copies of data as near to simultaneously as possible.

Synchronous replication is often the preserve of the most high-end block storage arrays.

Asynchronous replication adds a stage to the process, by acknowledging the host at the primary site when the data is written. Then the write is sent to the second site, which acknowledges that write back to the primary site array. Asynchronous replication is found in a wider range of storage products, such as iSCSI storage, network-attached storage (NAS), and so on.

Replication over great distances starts to suffer from about 1 millisecond of latency per 100 miles, and suppliers often recommend no more than a few hundred miles round trip.

For that reason, synchronous replication can have more of an impact on application performance. It demands acknowledgement before the next input/output (I/O) can take place, whereas asynchronous replication acknowledges locally so the next change can take

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous

place, with movement of data delayed. Of course, that also means the two data sets will differ for a longer time.

A real-world replication strategy might use a combination of synchronous replication – for the most critical elements of an application such as redo logs – while less critical data that could be restored goes via asynchronous. Snapshots could form part of the mix too, but it would all need to be underpinned with regular backups.

## Host, hypervisor and cloud replication

Here we have dealt primarily with synchronous and asynchronous replication in storage arrays.

Other forms of replication exist, such as:

- Host replication – between servers, perhaps of individual applications, databases or the entire server.
- Hypervisor replication – Replication managed at hypervisor level and comprised of its elements, such as individual virtual machines (VMs), and virtual storage, for example.
- Cloud replication – This could be replication to the cloud or multiple clouds as a target, or between clouds.
- Geo-replication – This is where data is stored in multiple remote locations, potentially very distant from each other. This can be for reasons of disaster recovery or to enhance availability. Replication over such long distances is not likely to be synchronous.

**Read more about replication**

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous

- Replication won't protect VMs against ransomware. Seamless replication is among the benefits of virtualisation, but many organisations fail to back up virtual machines properly.
- Cloud snapshots and backups: How to protect data in the cloud. You often need more data protection than native cloud services will give. But should you choose backup or snapshots? And what about third-party backup in the cloud?

Cloud storage 101: NAS file storage on AWS, Azure and GCP

Snapshots 101: Array vs backup software

Storage 101: Replication vs backup, and synchronous vs asynchronous