Globally Distributed Enterprise File Sharing with Azure NetApp Files and NetApp Global File Cache

Evolving from StorSimple to next-gen Azure solutions.
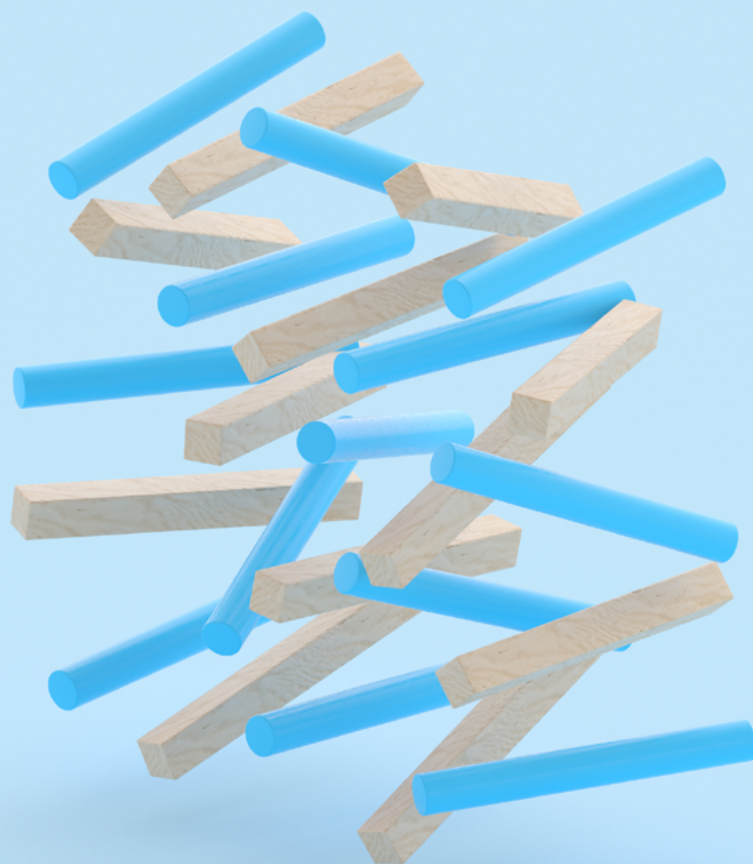
■ NetApp

Table of contents

Streamline and simplify IT storage and infrastructure by centralizing unstructured data into Microsoft Azure using Azure NetApp Files to provide fast local and geographically distributed access with NetApp Global File Cache™.

Introduction

Why did people use StorSimple...what are the primary use-cases?

- Unstructured file shares that can be more easily managed locally, backed by the flexibility/scale of the cloud

- Ability to have substantial amounts of data logically presented to a location, without the need to have a significant storage footprint at that location

As the StorSimple 8000 series appliance approaches end of support on December 31st, 2022, customers are exploring alternative solutions for managing and storing the data housed on this platform. If they do not have a plan in place to move away from StorSimple before the end of support, there could be some serious issues. If it breaks it cannot be fixed and data will be lost. There will also be no more updates for needed bug fixes. The possibility for data loss corruption, compliance breaches and more are things that should not be taken lightly for any organization. Luckily, NetApp is here to help.

Microsoft and NetApp have teamed up to present a natural evolution for StorSimple customers, that preserves their original objectives and extends incremental value.

- Global locking for 'single source of data' - everyone using the same data

- Centralized backup and compliance scan capability through Cloud Backup and Cloud Compliance

Azure NetApp Files (ANF) paired with NetApp Global File Cache (GFC) allows for customers to continue storing critical data in Azure while maintaining content at remote sites. In this document we will provide insight to the value of the NetApp solutions, cover data migration to Azure NetApp Files, provide guidance on how to deploy Global File Cache, and detail the experience to expect when using Azure NetApp Files and Global File Cache after migrating from StorSimple.

Solution overview

NetApp Global File Cache + Azure NetApp Files: a "major step" in unstructured data management for the distributed enterprise

85% of companies are in the process of adopting a cloud transformation strategy. This means combining on-premises, hybrid, and public cloud services and associated storage technologies, like file/block-based and object storage (e.g., Azure BLOB) to host both structured and unstructured data.

NetApp and Microsoft recognize the impact on the organization, end users, distributed IT strategy, datacenter, and cloud operations. The NetApp and Microsoft joint-solution approach allows for a scalable, flexible, and cost-effective solution strategy by addressing all layers of the enterprise from end users to branch offices, to the datacenter and cloud infrastructure.

With NetApp Global File Cache intelligent file caching software and Azure NetApp Files, in conjunction with Microsoft Azure, enterprises can do more than just control their data, they can revolutionize the way they manage unstructured data, both in their daily operations and how users access that data globally and in-cloud.

Drastically reduce storage footprint

Consolidation with Azure NetApp Files, enhanced with NetApp Global File Cache™, enables distributed branch offices total access to the entire directory structure, giving them streamlined access to all company data while only active data sets are cached locally. Users have immediate access to all centralized data, which could be 100s of TBs or even PBs of unstructured data, but only data that is relevant to the user population in that specific site is cached locally in the Global File Cache Edge cache. Additionally, as the active dataset ages over time, the Global File Cache intelligent file caching purging mechanism clears the least recently used (LRU) cached files from the local storage cache volume(s).

Streamline and simplify distributed IT

As organizations are aiming to centralize and consolidate their branch office IT storage assets, eliminating complexity and backups provides significant cost savings. As NetApp Global File Cache deploys transparently on a (virtual) Microsoft Windows Server instance, on traditional servers, or on virtualization platforms like Microsoft Hyper-V or VMware vSphere, enterprises can consolidate local storage and embed services like Microsoft Active Directory, DNS/DHCP, DNS, DFS Namespaces, andSoftware Distribution in their streamlined and standardized branch office IT image. Since the actual data is consolidated and stored safely and securely in Azure, a variety of data management tasks like backup and restore (BU/R), disaster recovery (D/R), and archiving can now be handled centrally as well, leveraging powerful mechanisms like time and space efficient snapshots and replication within the Azure Intelligent Cloud. This not only simplifies operations (by taking away these tasks at "the edge") but also allows for achieving more aggressive SLAs, further enhancing the business, and taking away business risks.

Ensure optimal user experience

With NetApp Global File Cache intelligent file caching software, your distributed users are guaranteed an optimal experience. By accessing and collaborating on data in real-time—transparent to all client platforms—users will feel as if they are all working in the same office, anywhere in the world, regardless of bandwidth, latency, and distance.

Unleash the power of the Azure Intelligent Cloud

While all data can be made accessible globally, the file service also allows for direct access to the data in-cloud. This means advanced (IaaS) applications, as well as a variety of platform (PaaS) services, can access the data directly in Azure NetApp Files, providing an extremely powerful, fast, and scalable capability to enhance the business once in Azure. This could entail file indexing, searchability, HPC, and analytics capabilities readily available in Azure or as 3rd-party offerings, just to name a few.

A New Data Store with Azure NetApp Files

Azure NetApp Files is a powerful, cost-effective, and easy-to-use data management solution for your enterprise cloud NAS storage workloads. With Azure NetApp Files, organizations can build a public cloud "datacenter" that is the foundation of a Data Fabric that services the needs of users worldwide. IT teams can then seamlessly manage data as it flows to wherever needed most, leveraging a centrally managed approach to all data, with the benefits and scale of public cloud.
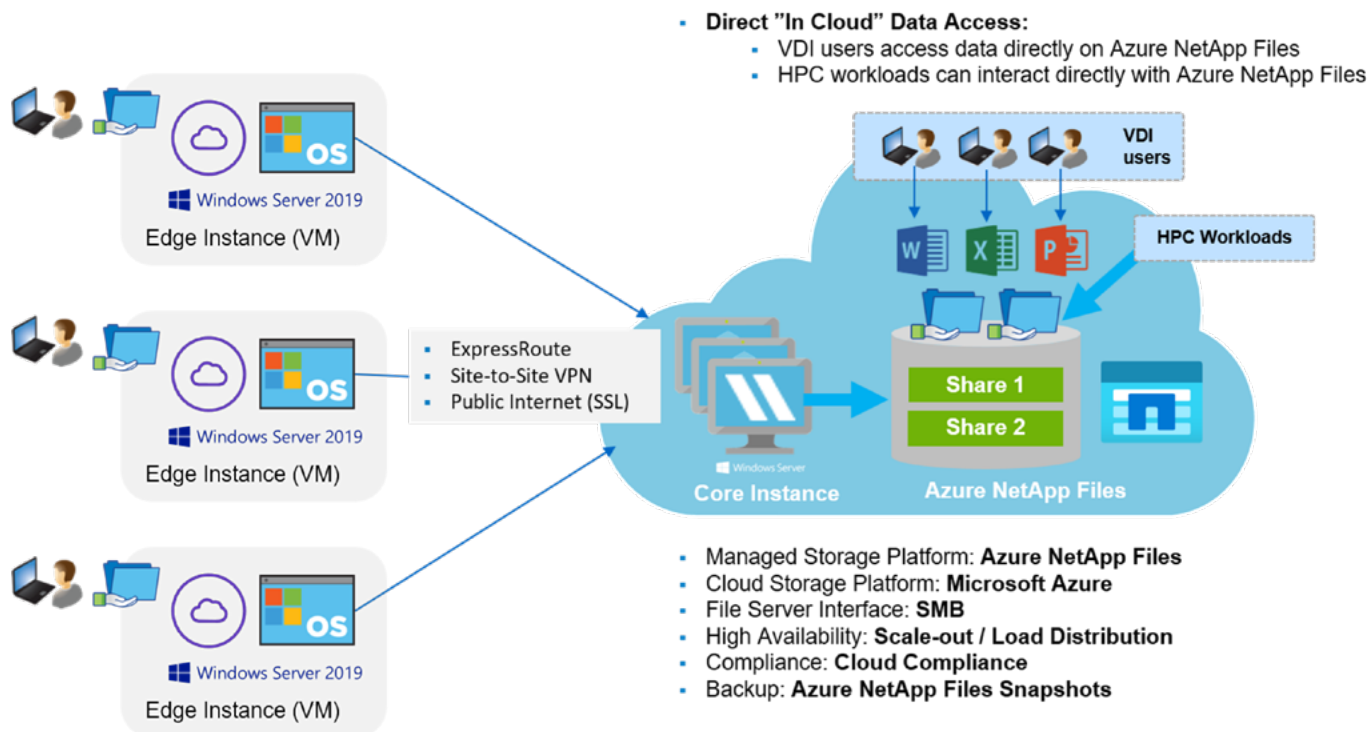
Azure NetApp Files addresses challenges the digital enterprise faces by:
- ✓ Consolidating distributed file servers into Microsoft Azure
- ✓ Managing distributed storage and IT infrastructure
- ✓ Simplifying deployment and data management
- ✓ Flexibly supporting your changing business needs
- ✓ Providing scalable, on-demand capacity and performance
- ✓ Powering your enterprise applications, both on-premises as well as in the cloud

Enterprises can simplify their storage environment with proven operational efficiency by leveraging economies of scale and the resiliency of the Microsoft Azure public cloud infrastructure and services. Data is secured with built-in encryption and a consolidated approach with local caching gives a global view of storage with a single management console and a single physical footprint.

Azure NetApp Files features
- ✓ Seamless workload migration
- ✓ Low-cost Disaster Recovery using public cloud
- ✓ Reduced storage footprint
- ✓ Grow-as-you-go file shares
- ✓ Automated DevOps environments
- ✓ Azure NetApp Files managed encryption at rest
- ✓ Cost effective data protection

- **Direct "In Cloud" Data Access:**
  - VDI users access data directly on Azure NetApp Files
  - HPC workloads can interact directly with Azure NetApp Files

- Managed Storage Platform: **Azure NetApp Files**
- Cloud Storage Platform: **Microsoft Azure**
- File Server Interface: **SMB**
- High Availability: **Scale-out / Load Distribution**
- Compliance: **Cloud Compliance**
- Backup: **Azure NetApp Files Snapshots**

Using Azure NetApp Files to Consolidate Distributed File Servers

Azure NetApp Files allows the enterprise to simplify operations by unifying data management across a consolidated data footprint. This benefit is further expanded when all enterprise data—even that used primarily by users outside of the datacenter locale—are using the centralized and consolidated infrastructure, hosted in the public cloud. This provides the ability to increase the efficiency of IT staff as they can easily move data and scale storage resources when and where they are needed most. Azure NetApp Files as a foundation for a global Data Fabric brings increased flexibility, control, and security across the public cloud storage environment. The main repository for the unstructured data is a (number of) share(s) configured on the customer's Azure NetApp Files storage account, hosted in the Microsoft Azure, providing direct SMB access. The customer's file storage solution provides volumes associated with corporate file shares hosted on Azure NetApp Files. *See figure above.*

Accessing the ANF datastore with a Global File Cache fabric

By introducing the NetApp Global File Cache software and integrating with existing Azure NetApp Files storage in the cloud, all distributed locations can use the centralized Azure NetApp Files file storage resources as if they were local. The result is a single, centralized storage footprint, versus a distributed storage architecture that requires local data management, backup, security management, storage, infrastructure footprint, etc. in each location. The NetApp Global File Cache Edge instances transparently integrate with the Global File Cache fabric in the customer's Azure subscription:

1. Distributed locations connect to the Azure instances via the NetApp Global File Cache fabric

2. Software provides a Virtual File Share and Intelligent File Cache at each location

3. Enables high performance global file sharing with real-time distributed file locking

The software overlays the Microsoft Windows File Sharing mechanism, fully integrating with Microsoft security principles like Active Directory, ACLs, and NTFS permissions, thereby allowing it to perform at a global scale, even in locations that are challenged with poor connectivity (low bandwidth or high latency).

✓ Flexible: Storage agnostic, works with any SMB infrastructure including Azure NetApp Files

✓ Intelligent: Caches only what is needed at the branch (active dataset)

✓ Zero-touch: Automatically purges "stale" cached files over time (LRU)

✓ Performant: Compresses, streams, and reduces data

✓ Consistent: Distributed file locking for enterprise applications

✓ Provides data-in-transit encryption, streaming and compression between Microsoft Azure and edge

NetApp Global File Cache Edge

Each distributed office will run an instance of the NetApp Global File Cache software, configured as an edge instance; the edge VMs provide the critical performance-enhancing functions such as file caching, file-level differencing, and local service to users. Each Edge will present the users in that location full visibility to all the cloud-resident shares to which they are authorized, allowing even smaller locations the ability to leverage all enterprise information assets without the need for that data to be physically housed on-site.

The Edge instances can run on Windows Server 2012 R2 and above, either on physical hardware or virtual infrastructure in the remote location and will employ

an intelligent file caching partition which uses an array of algorithms to retain the most active data set for that location, ensuring optimal performance for users located there.

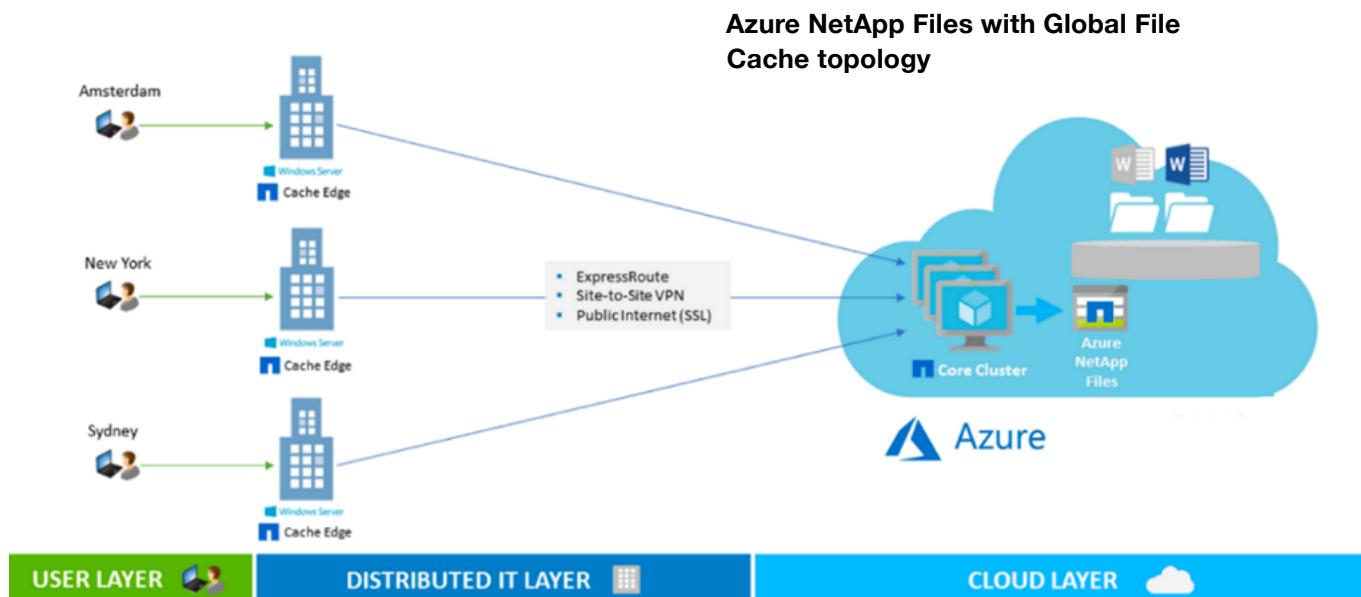### NetApp Global File Cache Edge instance

1. Software Installation Package or Virtual Appliance running on Windows Server 2012 R2 and above

2. Creates a Virtual File Share (e.g.): \\Edge\FASTData\ [datacenter]\[fileserver]\[share]\[folder]\

3. Contains the Global File Cache Intelligent File Cache

Network connectivity

Connectivity is provided by the customer's existing network infrastructure, either Microsoft Azure ExpressRoute, or a secure site-so-site or point-to-site Virtual Private Network (VPN) connection between each location needing access to the centralized data and the datacenter. With NetApp Global File Cache, there is also the choice of enabling an SSL connection between the Core and Edge instances if there is no direct connection or VPN available or possible. When using a VPN, the VPN must be capable of carrying bidirectional traffic on TCP ports 6618-6630 between the Global File Cache Edge and the corresponding Global File Cache Core instance(s).

Configuration guidelines

Please review the NetApp Global File Cache Hardware and Software requirements and Application Best practices guides at www.cloud. netapp.com to ensure optimal deployment and performance.

Azure NetApp Files with Global File Cache topology

Amsterdam — Windows Server / Cache Edge

New York — Windows Server / Cache Edge

Sydney — Windows Server / Cache Edge

- ExpressRoute
- Site-to-Site VPN
- Public Internet (SSL)

Core Cluster

Azure NetApp Files

Azure

USER LAYER | DISTRIBUTED IT LAYER | CLOUD LAYER

Deployment methodologies

The most common topology is a hub-and-spoke deployment, as it is typically used for the purpose of data centralization and consolidation of storage from distributed branch offices into a datacenter. NetApp Global File Cache with Azure NetApp Files enables consolidation of all enterprise data onto the secure and scalable platform for access by any/all users worldwide.

Operating environment summary

The topology referenced here is a hub-and-spoke model, whereby the network of distributed offices/ locations are all accessing one common set of data in the customer's public cloud platform of choice. The key points of the reference architecture are as follows:

1. Centralized data store: Azure NetApp Files service in Microsoft Azure Intelligent Cloud

2. NetApp Global File Cache fabric: Extension of the central data store to the distributed locations

    a. NetApp Global File Cache Core VM instance(s), (i.e., DS4_v3), mounting to Azure NetApp Files file shares (SMB) in Microsoft Azure

    b. NetApp Global File Cache Edge instance(s), running in each distributed location

        i. Presents a Virtual File Share that provides access to central data in real-time

        ii. Hosts the Intelligent File Cache on a custom-sized NTFS volume (D:\)

        iii. Caches "active" data on-demand or by leveraging scheduled pre-population jobs

3. Network connection

    a. Microsoft Azure ExpressRoute

    b. Virtual Private Network connectivity

    c. SSL connection

4. Integration with customer's Active Directory Domain

    a. The Active Directory (AD) implementation should be a standard AD configuration with on-premises domain controllers; Azure AD (AAD) or ADDS is not supported at this time

    b. Single domain infrastructure only is supported at this time (for multiple domain configurations via trust relationships, contact NetApp)

    c. The Service Account should be made a part of the Domain Administrators Group in the Active Directory Domain containing the Azure NetApp Files resource

5. DFS-Namespace for the use of a global namespace (recommended)

    a. Create a unified namespace for users to access data in a consistent manner, always using the local cache in the (nearest) branch office or directly in Microsoft Azure

    b. Leverage Active Directory Sites & Services to enable location awareness and failover/failback redundancy in case of local branch office outage

    c. Aggregate centralized volumes, file shares, and services into a single namespace to allow multi-PB scale

User experience

By using the NetApp Global File Cache solution to extend Azure NetApp Files centralized file shares to globally distributed users, these users have access to these shares via either a UNC path or a DFS Namespace.
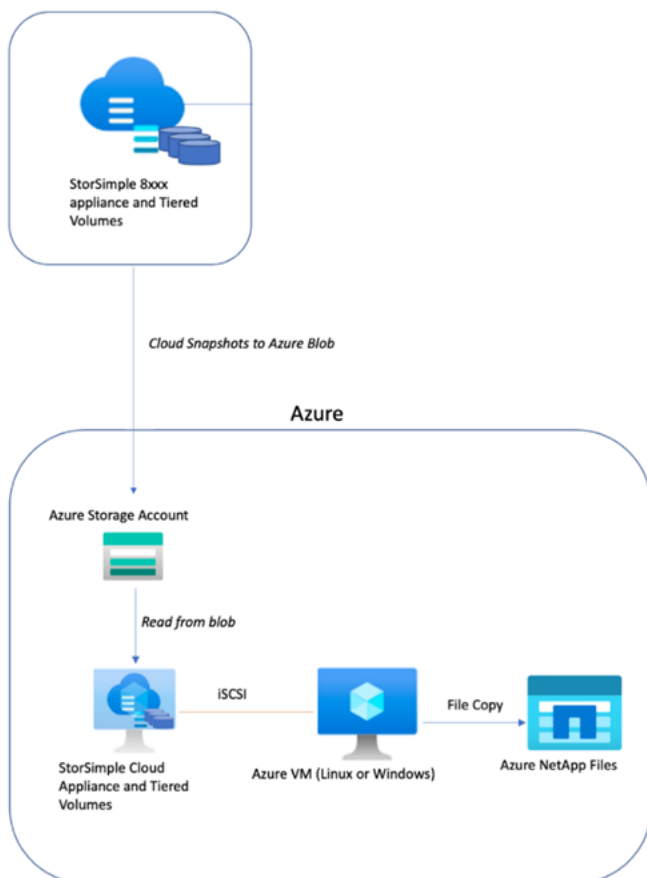
The user experience for properly configured systems is analogous to the experience of having a local file server, i.e., users or applications can navigate to a directory structure, select shares/folders, and work with files. The complete range of file operations (open, save, copy, paste, etc.) are available to the user requiring no change in workflow.

1. When a user requests use of a file in the central repository, Active Directory will authenticate that user's access rights

2. After a successful authentication, the file is opened centrally from the backend file server by the Global File Cache core, and a lock is applied (centrally on Azure NetApp Files) to that file

3. If the file has not ever been used by a user in that location, the file is served using the proprietary streaming and compression technologies inherent in the Global File Cache solution to improve performance

    a. If, however, the file has been used by any user in that location before, it is already present in the local cache; in this case the file will be served out of the local cache without incurring network transfer operations, thereby providing a high-performance experience

i. If the file version in local cache is not the most up-to-date version that is in the authoritative backend file server repository, any differences (and only the differences) will be sent to the local cache and merged with the cached version of the file upon open; this maximizes performance and minimizes network resource use

ii. *Note: the file remains locked at the central repository, and was only served after authentication and lock were performed*

4. User operations continue as normal, and any updates/changes/writes will be cached locally

5. Upon save/exit, any changes to the file will be differenced back to the authoritative central copy

6. Upon exit, after the saves are completed centrally, the file is closed, and the lock will be released and available to other users

Migrate your StorSimple data to Azure NetApp Files

Overview of Data Migration – StorSimple to Azure NetApp Files

StorSimple 8xxx appliance and Tiered Volumes

Cloud Snapshots to Azure Blob

Azure

Azure Storage Account

Read from blob

StorSimple Cloud Appliance and Tiered Volumes

iSCSI

Azure VM (Linux or Windows)

File Copy

Azure NetApp Files

Getting the data to Azure
To expedite migration, we will rely on cloud snapshots to hydrate a copy of the tiered volume in Azure via the StorSimple Cloud Appliance. This is to prevent reading data from the physical appliance which most likely has spilled over into an Azure Storage Account, thus creating a bottleneck. Deploy the 8020 cloud appliance.

Cloud snapshots are configured via backup policies. Prior to migration, ensure the appliance has completed a cloud snapshot so that the most current data set has been replicated to Azure. You can run a manual backup and monitor its progress/completion. Ensure that each tiered volume slotted for migration is protected via a cloud snapshot.

Volume Clone
Use the following guidance to clone the most current cloud snapshot to the StorSimple Cloud Appliance: https://docs.microsoft.com/en-us/azure/storsimple/storsimple-8000-clone-volume-u2

Create Azure VM
StorSimple presents a block storage device via iSCSI for host access. We will use this host to copy data into the target Azure NetApp Files volume(s) as well. Use the following guidance to create a Windows Server VM in the same region as the StorSimple Cloud Appliance: https://docs.microsoft.com/en-us/azure/virtual-machines/windows/quick-create-portal

A few things to keep in mind for the Azure VM:

• Deploy the VM in the same region as the StorSimple Cloud Appliance. This will reduce latency in accessing volumes from the host over iSCSI.

• Ensure Accelerated Networking is enabled

• Recommended minimum instance size is a DS3v2. This matches the instance size of the StorSimple Cloud Appliance and allows for consistent bandwidth between appliance and VM.

Connecting to the Cloud Appliance
The StorSimple 8020 is an iSCSI-based appliance. The Azure VM deployed in the previous step will connect via iSCSI to the appliance private IP address.

• Locate the StorSimple IP Address

• Connect Azure VM to StorSimple Cloud Appliance

Create Azure NetApp Files share

Next, create an Azure NetApp Files SMB volume that will be used as the target for the migration. You can find detailed steps on how to configure Active Directory and the SMB volume here: https://docs.microsoft.com/en-us/azure/azure-netapp-files/azure-netapp-files-create-volumes-smb

Each SMB volume represents an individual file share. A common strategy for aligning source data on StorSimple volumes

Data copy

There are multiple tools, CLI and GUI based, that can handle the task of copying data. We will list a few options here for guidance:

Robocopy: Microsoft provided CLI based utility

NetApp Cloud Sync: NetApp SaaS based copy tool (provided at no-cost for Azure NetApp Files customers)

Summary

The need to move away from StorSimple is just around the corner and NetApp is here to help. This combination of Azure NetApp Files and NetApp Global File Cache software allows enterprises to consolidate their unstructured data to a centralized "single set of data," taking advantage of the flexibility, availability, and economics of a centralized storage model in the public cloud for one of their largest use cases— unstructured data—while maintaining a reduced storage footprint at the datacenter and distributed branch locations.

The resulting effect(s) on the business are significant:

✓ A consolidated file storage environment leveraging the benefits of the public cloud and the scalability and performance from an enterprise storage service

✓ A reduced storage infrastructure footprint at distributed locations through intelligent file caching

✓ Significant cost savings in the hardware and management aspects of providing file services to distributed users

✓ The ability to leverage the enterprise-class scale, flexibility, and security of Azure NetApp Files-resident data across all users, regardless of location, both on-premises as well as in-cloud

✓ Increased flexibility and agility through enhanced global collaboration

✓ Enhanced security and compliance by removing the risk of data loss/leak at distributed locations through error, disaster, and intrusion

■ NetApp

NetApp

Copyright Information