TECHNICAL WHITE PAPER

Protecting Azure NetApp Files with Rubrik

Amrutha Naik, Ed Morgan, Pierre-François Guglielmi
April 2020
RWP-0502

TABLE OF CONTENTS

ABOUT THIS WHITE PAPER

The primary objective of this white paper is to study the different use cases that can be leveraged with the marriage of Rubrik & Azure NetApp Files. This white paper is intended to provide architects and data center administrators information about the implementation and benefits of Rubrik's integration with Azure NetApp Files, steps to configure Rubrik to protect Azure NetApp Files, and performance benchmarks of the solution.

EXECUTIVE SUMMARY

As the amount of unstructured data continues to grow exponentially, enterprises face the daunting task of ensuring that critical data on Network Attached Storage (NAS) systems are fully protected, and the digitization of business requires enterprises to move faster and be more agile to survive. Applying new technologies to existing business activities (e.g., leveraging AI to increase customer satisfaction) will continue to fuel the cloud paradigm. According to IDC, enterprises will spend more than $500 billion on cloud and cloud services by 2021, with 80% of application development on cloud platforms. For many enterprises, public cloud represents the ability to rapidly access resources for innovation while operating in a data-rich environment. With this exponential increase in cloud adoption, it is paramount to have a matching data protection and management solution.

For the first time, Microsoft will deliver an Azure native, first-party service for enterprise NFS/SMB file services based on NetApp ONTAP technology. This new development is driven by a strategic partnership between NetApp and Microsoft and further extends the reach of NetApp's world-class data services to Azure. This Azure cloud-native data service delivers high performance, reliability, and enterprise data management and security for customers who are moving enterprise NFS/SMB workloads to Azure.

The data service is Azure NetApp Files, the industry's first bare-metal cloud file storage and data management service. Azure NetApp Files (ANF) is a new Azure service, delivered by Microsoft and built on NetApp's industry leading ONTAP technology directly in the Azure data centers. The service allows customers to move their NFS and SMB file-share workloads into Azure – even legacy applications – without rearchitecting their applications. Users can reduce their migration from months or years, to days or weeks. Azure NetApp Files can be provisioned and managed seamlessly within the Azure portal. It has the same billing, CLI, and deployment paradigm as any other Azure service; the dedicated environment in Azure means that performance is both optimal and assured. Azure NetApp Files eliminates the need for time-consuming and expensive architectural change, delivering an easy way to seamlessly provision file-based workloads in Azure and helping organizations to meet their cloud mandate in record time. With Azure NetApp Files, any file-based workload can move to Azure, including NFS v3 and even SMB shares, with no change. To increase storage, simply choose from multiple levels of guaranteed performance seamlessly through the Azure portal – and change performance on the fly without moving data or creating any new volumes.

CUSTOMER CHALLENGES

Some very common customer challenges with NAS protection include:

- Large file servers with billions of files are often particularly hard to handle as they may take hours to days for a full backup.
- Unsatisfactory backup performance or elongated recovery times are challenges that backup administrators typically encounter when protecting file data.
- Backup administrators often begin each day by remediating failed backups from the previous night, thus increasing operational management efforts.

HOW RUBRIK SOLVES THESE CHALLENGES

Rubrik takes a modern incremental-forever approach, eliminating the need for periodic full backups. Performing incremental-forever backups reduces nightly backup windows, especially when there is a need to protect millions or billions of files. From a restore perspective, each backup looks like a full backup whether restoring a single file, a folder, or a full volume. This approach works not only for small enterprise NAS systems/file servers but even file services on cloud such as Azure NetApp Files. Operational management is minimized thus providing valuable production time for backup administrators to focus on more closely aligned business objectives.

RUBRIK OVERVIEW

Let's understand a little more about the different building blocks of Rubrik in relation to NAS protection and get familiar with the common terminologies used in Rubrik NAS protection.

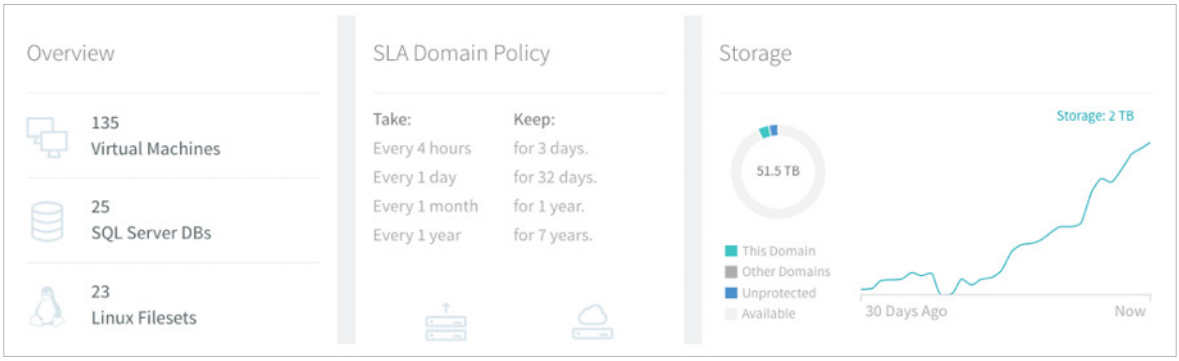
FILESET

A fileset is a rule that defines the set of folders and files that should be backed up. The benefits of protecting NAS file data with filesets include:

- **Policy-Driven Management** — filesets are linked to an SLA Domain, which provides a configurable set of policies that can be applied to groups of virtual machines, applications, and hosts to achieve specific data protection objectives.
- **Centralized Management** — providing visibility through the Rubrik UI of all file data that has been protected and ready for granular recovery.
- **Incremental-forever backups** — via block mapping and intelligent metadata algorithms to dramatically reduce local storage requirements, provide much faster backups, and reduce network usage.
- **Multiple Recovery Options** — filesets provide granular file and folder-level protection with the ability to exclude sub-paths. Users can choose to restore files or folders to the original server, export to a different server, or download to the browsing system. There are many variations and granularity available to accomplish your recovery needs.

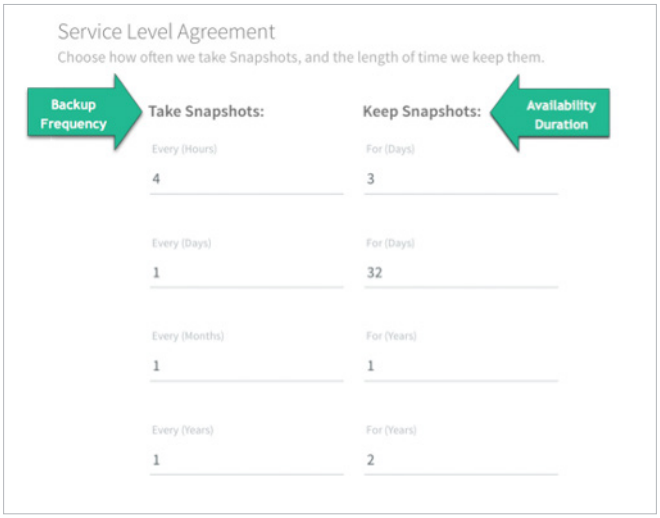
SLA DOMAIN POLICY

As with all datasets, Rubrik's data management approach begins with SLA Domain Policies. A Rubrik SLA Domain Policy is a declarative policy encompassing the core items needed for backup and recovery, replacing the need to individually configure jobs, tasks, and other items. SLA Domain Policies are a core part of Rubrik's architecture and extend across all data types, as shown below:

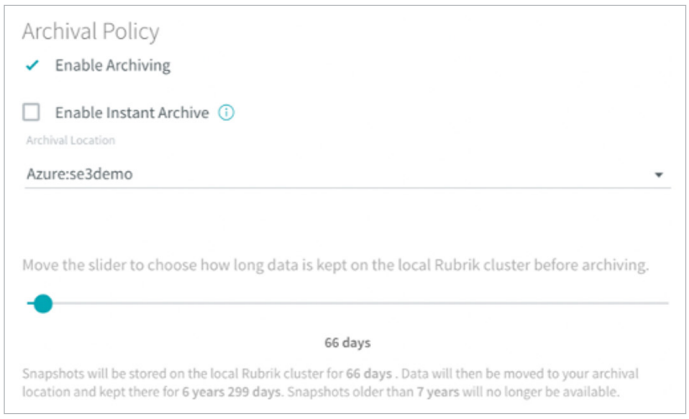


Let's walk through the pieces needed to configure an SLA Domain Policy that can apply to all data types:

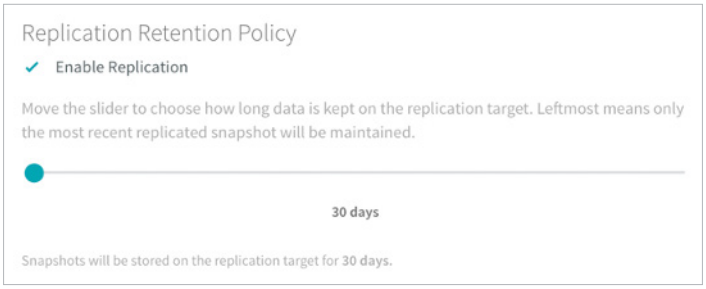
- **Backup Frequency:** Also known as the Recovery Point Objective (RPO), this describes the frequency in which backups are taken.
    - For file data, this determines how often a restore point is captured to provide recovery of a single file or multiple files. Due to the incremental-forever nature of Rubrik, some customers choose to protect more frequently than the traditional daily backup.
- **Availability Duration:** Also known as retention. Or, how long are backups retained?
    - For file data, retention can widely vary whether due to customer policy or regulatory and compliance requirements. For many customers, this often ranges from a week to 90 days but may be multiple years in order to meet regulatory compliance.



- **Archival Policy:** Archive targets can be public cloud (AWS, Azure, or Google Cloud Platform) or on-premises (S3 compatible object stores, NFS, or tape). This dictates which archive target is used and when archives are maintained solely in the cloud and not on the local Rubrik cluster. If archives are maintained solely in the cloud (past 30 days for instance), RTO is longer due to the time required to retrieve data back to the Rubrik cluster.
    - For file data, archival can dramatically reduce long-term storage costs, especially when retention is required for years due to regulatory or compliance reasons. The amount of data retained on the local Rubrik cluster is most often determined by the time frame of restore requests. For many customers, this ranges from 30 days to 1 year depending on when most recoveries tend to occur.



- **Replication Policy:** this relates to Disaster Recovery (DR). Simply put, how much replicated data should be maintained at a DR site?
    - For file data, this is often determined by how far back in time a customer would need to be able to restore files if there is a DR event. Some customers choose to keep the same amount of data on the remote Rubrik cluster as on their local Rubrik cluster, while others choose to have a shorter timeframe available on the remote Rubrik cluster to provide cost savings.

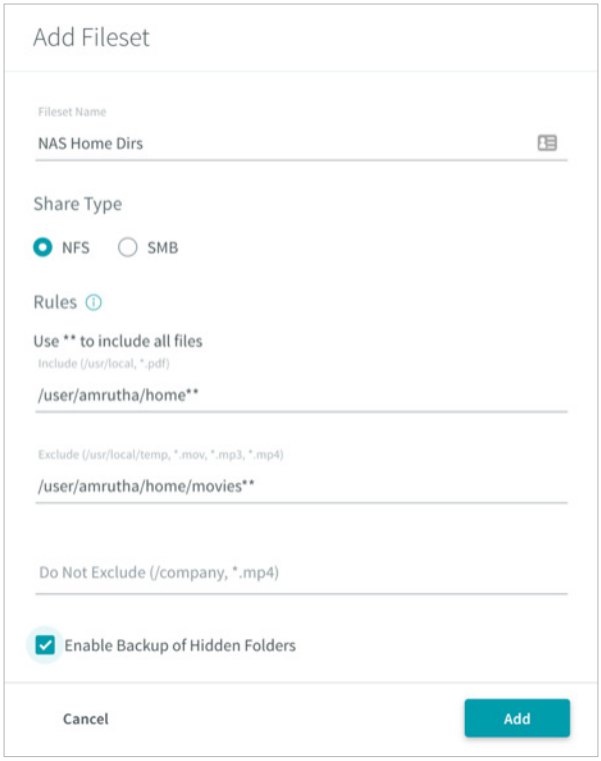


UNIVERSAL FILESET

Rubrik filesets integrate with SLA policies and are universal in that various operating systems (Linux, AIX, and Windows) and NAS devices are protected utilizing the same fileset concept. There are both similarities and differences in how Rubrik manages data from Linux, AIX, Windows and NAS hosts. The similarities exist due to the way Rubrik utilizes filesets, and the differences are managed by Rubrik without requiring customer interaction after initial configuration.

WHAT IS A RUBRIK FILESET?

As shown below, a fileset is the combination of full paths or filenames to define objects to include or exclude from backup. The Rubrik cluster interprets the syntax entered for the inclusion or exclusion fields of filesets by following a relative path rule. In the case of NAS filesets, after defining the share path, any associated filesets only need to include the desired path after the share.



WHY DID RUBRIK CHOOSE TO USE FILESETS?

Filesets were designed to give customers more granular control while still providing a construct that scales in a simple way for large environments where operational overhead is challenging.

HOW DO RUBRIK FILESETS PROTECT LARGE WORKLOADS?

If the data protected by a fileset exceeds 100 GB, Rubrik automatically subdivides incoming datasets into 100 GB partitions. Partitioning filesets provides increased ingest and restore performance as the workload can be distributed across all available nodes in the cluster. This distribution is done transparently from a customer perspective and allows Rubrik to protect many files as a single fileset. Further, filesets continue to scale horizontally as more nodes join the Rubrik CDM platform.

Filesets can be defined for Linux, AIX and Windows servers or NAS shares. For Linux, AIX and Windows (whether physical or virtual), Rubrik provides data protection for file systems through the pairing of the host with a fileset and an SLA policy. A single host can also be paired with multiple filesets, and each fileset can even be assigned to a different SLA Domain if desired.

For NAS, the Rubrik cluster pairs a fileset with NAS shares. The pairing of multiple filesets to a single share is similarly permitted and different SLA Domains for each share fileset are allowed. SLA assignment continues to match the core simplicity of the Rubrik architecture thus allowing the same SLA policy for VM's, Linux, AIX, Windows, or SQL also to be assigned to NAS datasets.

NAS FILESET PROTECTION

NAS shares are protected similarly to Linux, AIX and Windows hosts and do not rely on Network Data Management Protocol (NDMP). Not using NDMP is a conscious design choice that further solidifies Rubrik as vendor agnostic, provides true incremental-forever for NAS, and stores data in a "native" format.

Most traditional backup applications still utilize NDMP. As a legacy control protocol created to move data between a backup server and the tape drives/libraries, NDMP has significant disadvantages. The disadvantages begin with the fact that NDMP was created decades ago when datasets were much smaller and physical tape was the primary backup media. In addition, NDMP does not dictate the format of the backup stream. As a result, each vendor sends the backup stream in a proprietary format that is usually not meant to be unpacked.

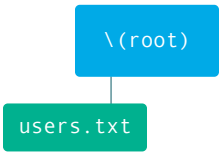
NAS file systems are exposed through either NFS or SMB (sometimes referred to as CIFS) protocols. Rubrik can backup file systems exposed through either protocol and preserve the corresponding metadata, including Access Control Lists (ACLs). Our approach is agnostic to the NFS protocol version and supports NFS 3.x and NFS 4.0. Similarly, for SMB, we support SMB 1.x, 2.x, and 3.x. This method includes storing data in a native format, backing up in an incremental-forever fashion, and instant access to all of the data management capabilities of the Rubrik platform including replication to DR, cloud archive, global search, erasure coding, and more. To further simplify datacenter operations, the Rubrik approach does not require NDMP accelerator nodes or "proxy VMs."

SAMPLE NAS FILESET

This section provides a few example use cases for defining NAS filesets:

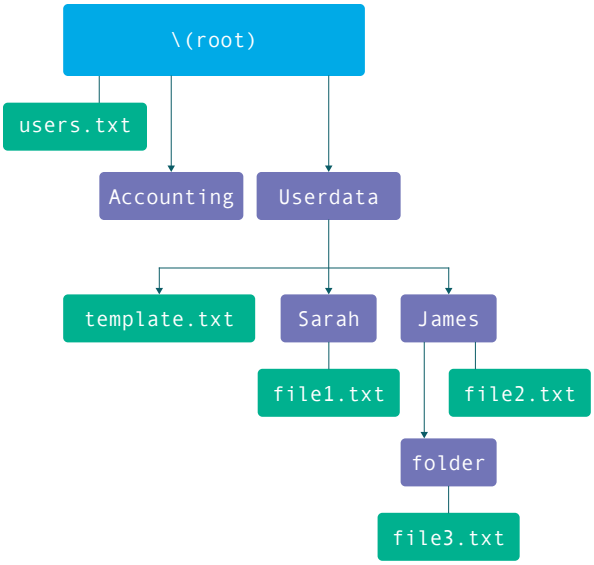
- **Perform a one-level backup of all files in the root path**

```
Include: \*
Backup data:
\users.txt
```



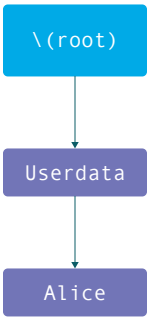
- **Perform a recursive backup of everything under the root path**

```
Include: \
Backup data:
users.txt
Accounting
Userdata
Userdata\template.txt
Userdata\Sarah
Userdata\Sarah\file1.txt
Userdata\James
Userdata\James\file2.txt
Userdata\James\folder
Userdata\James\folder\file3.txt
```



- **Perform an empty backup of directories that start with capital A (case sensitive) under \Userdata**

```
Include: \Userdata\A*
Backup data:
Userdata\Alice
```

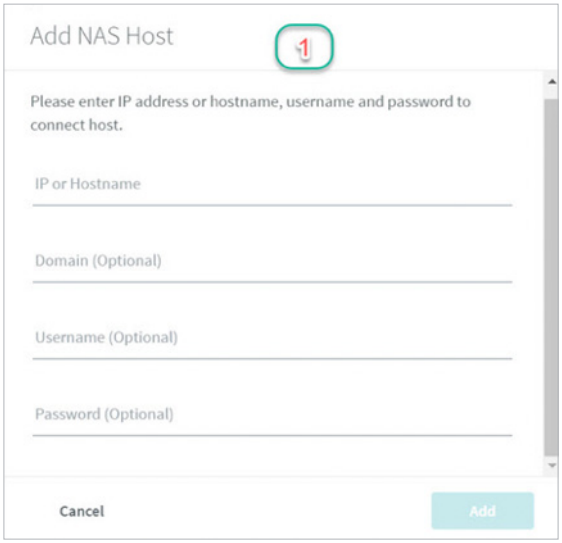


NAS SHARE BACKUP WORKFLOW

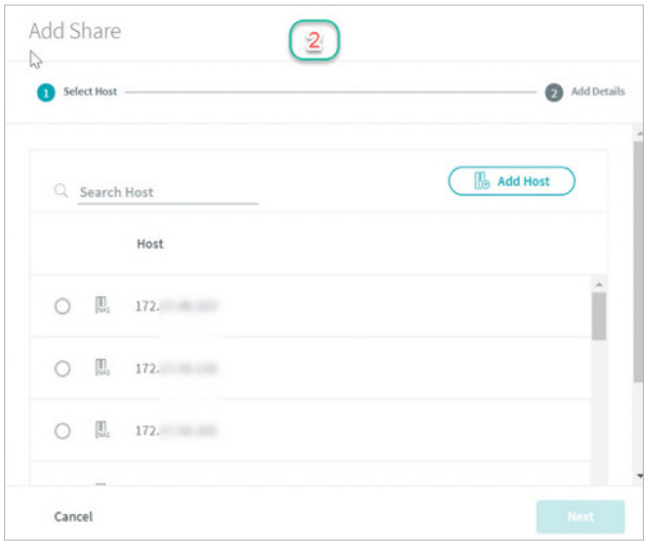
As NAS appliances do not support the installation of any third-party agents, Rubrik manages and protects data in NAS shares by connecting directly to NAS appliances. The user can begin maintaining and protecting a NAS host by adding the host to the Rubrik cluster. Read permissions across the entire NAS share are required.

Rubrik provides a streamlined process for setting up NAS backups and having the Rubrik cluster connect directly to NAS servers. Setup is simple:

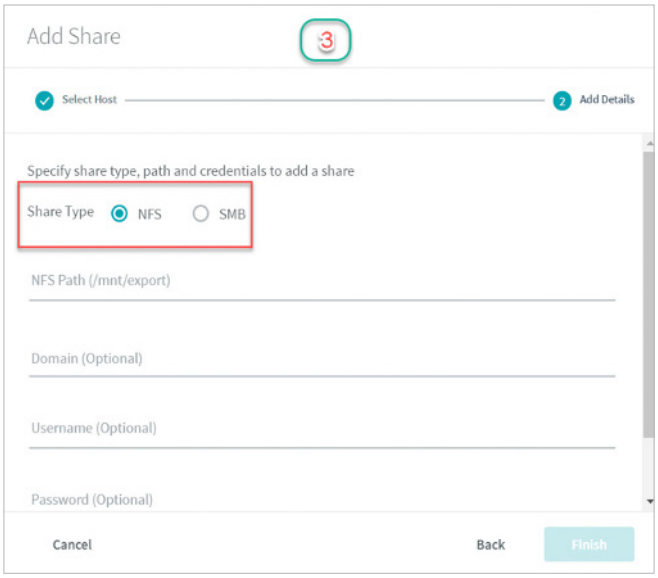
1. Add the IP address or FQDN and supply appropriate credentials:



2. Select a NAS host or add a new one:



3. Add share details such as the protocol and the path:



NAS SHARE RECOVERY FLOW

NAS Share, Windows, and Linux filesets have similar recovery methods that allow users to search or browse and select individual files or folders to restore, export, or download.

The steps below are handled automatically by Rubrik if applicable when a customer restores NAS data:

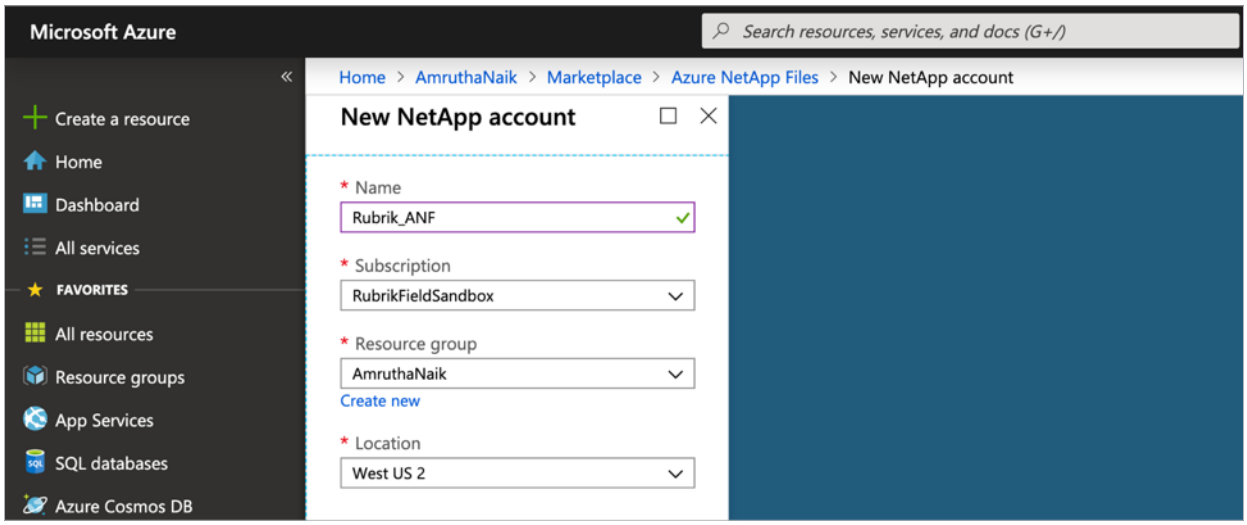
1. Based on the source path, determine which partitions need to be accessed whether a single partition or multiple partitions.
2. For each partition, assign it to a node to be executed. This allows restore scalability across many Rubrik nodes.
3. For each partition, create all directories that need to be built, including the ones in the partition boundaries.
4. Prepare writes: open all permissions and create new non-existing files.
5. Parallel writes: write all data in parallel.
6. Create all symbolic links.
7. Set all permissions and other metadata.

PROTECT AZURE NETAPP FILES USING RUBRIK

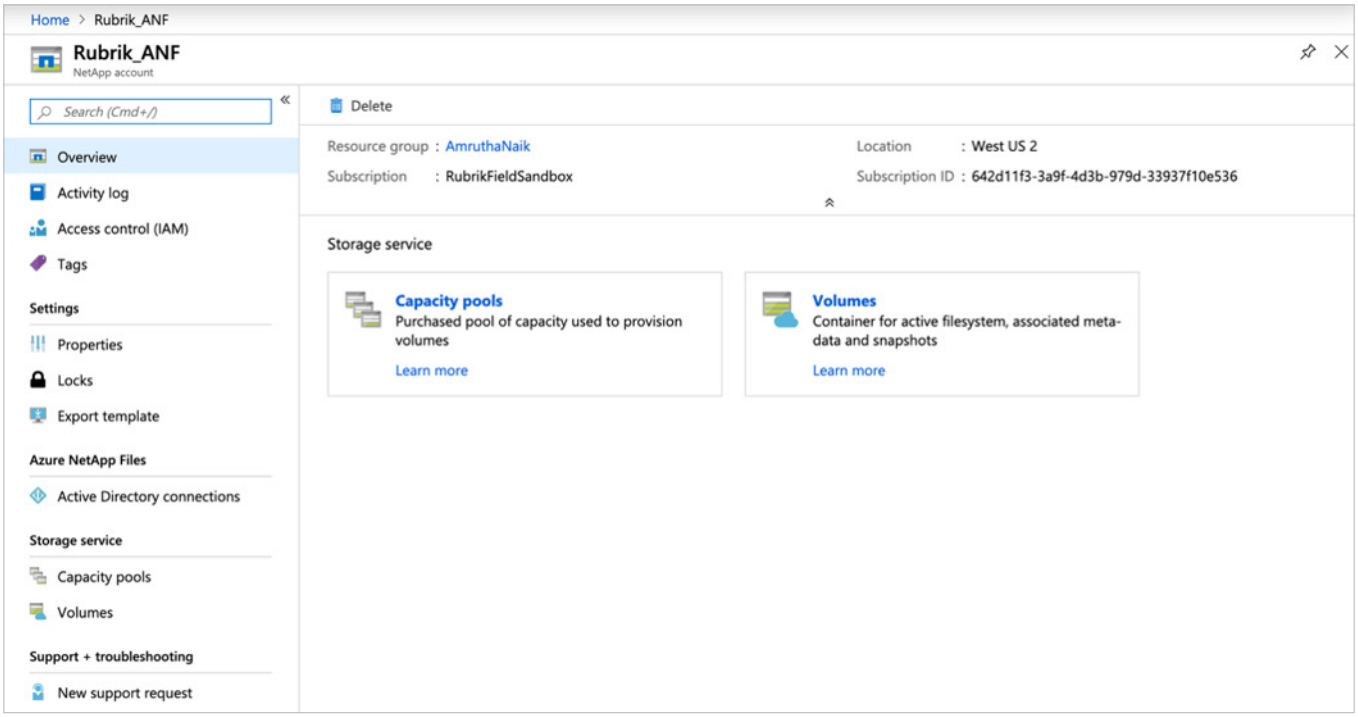
The below section will cover the following two parts:

- Creating NetApp account in Azure portal, creating capacity pool, volumes and NFS/SMB shares
- Configuring Rubrik to protect Azure NetApp Files

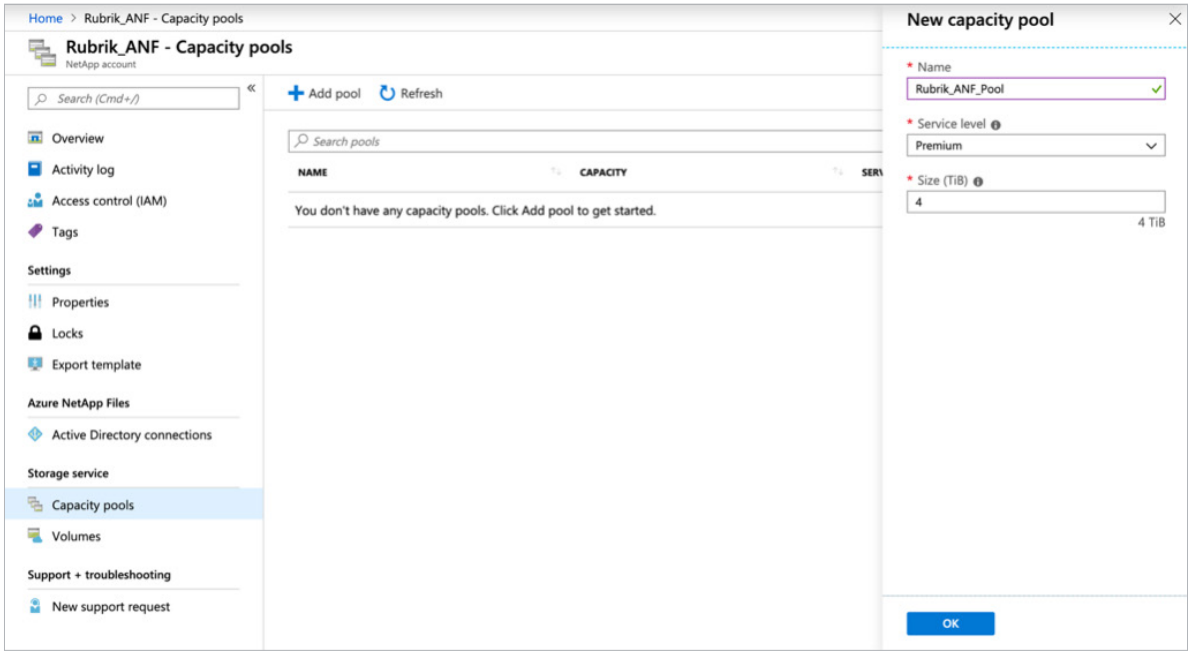
1. In the Azure portal search box, search for **Azure NetApp Files**. After launching the Azure NetApp Files wizard, create an account by pressing the **Add** button:



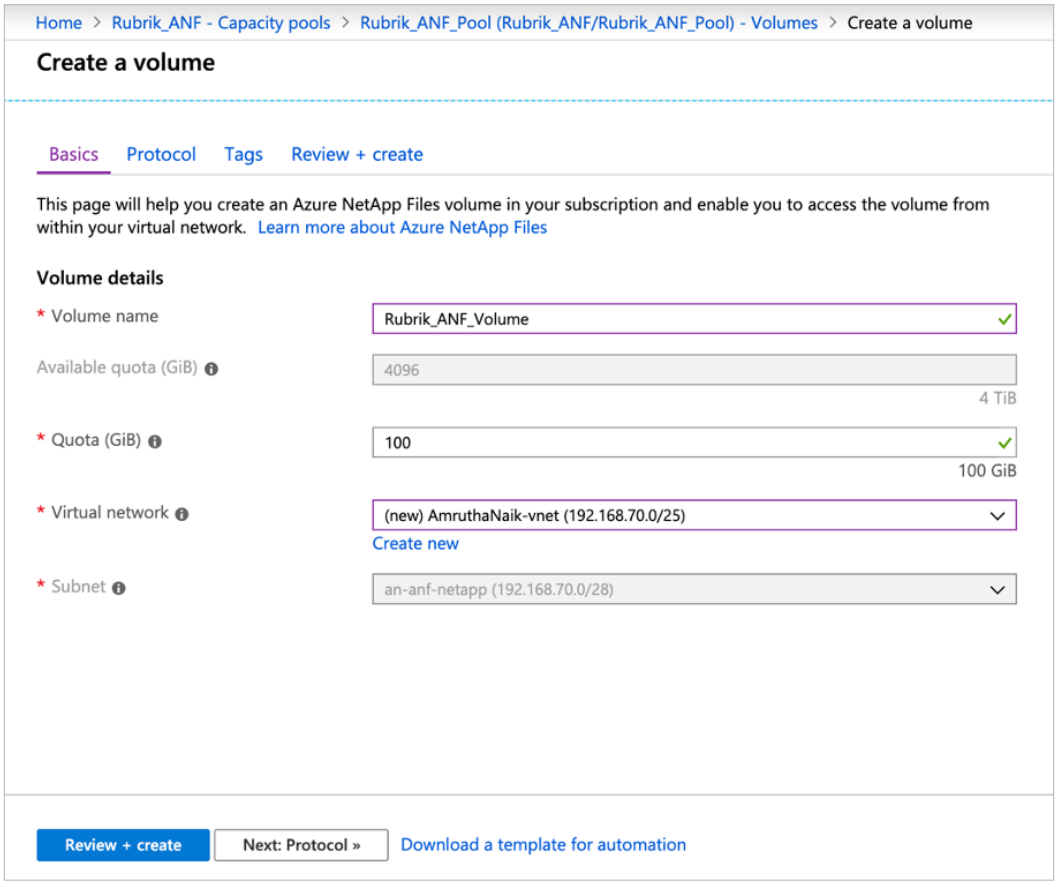
2. After the NetApp account is created, option for creating capacity pools and volumes is displayed as shown below:



3. Create a capacity pool by clicking on **Capacity pools** and click on **Add Pool**. Enter an appropriate name for the capacity pool, select the service level as per customer requirements and select the size of the pool and click on **OK**:



4. Once the capacity pool is created, the next step is to create a volume. Click on **Add Volume** and provide the Volume name and quota. Under **Virtual network**, choose your current virtual network or click **Create new** to create a new Azure virtual network (Vnet):



Then fill in the following information:

a. Enter **myvnet1** as the Vnet name.

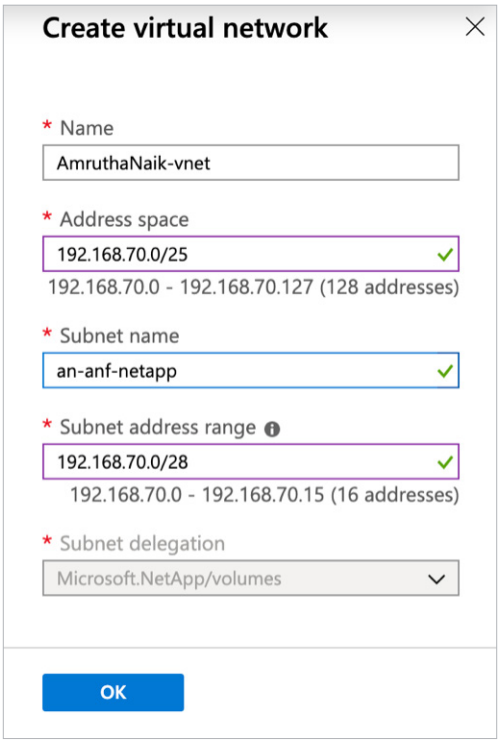
b. Specify an address space for your setting, for example, **10.7.0.0/16**.

c. Enter **myANFsubnet** as the subnet name.

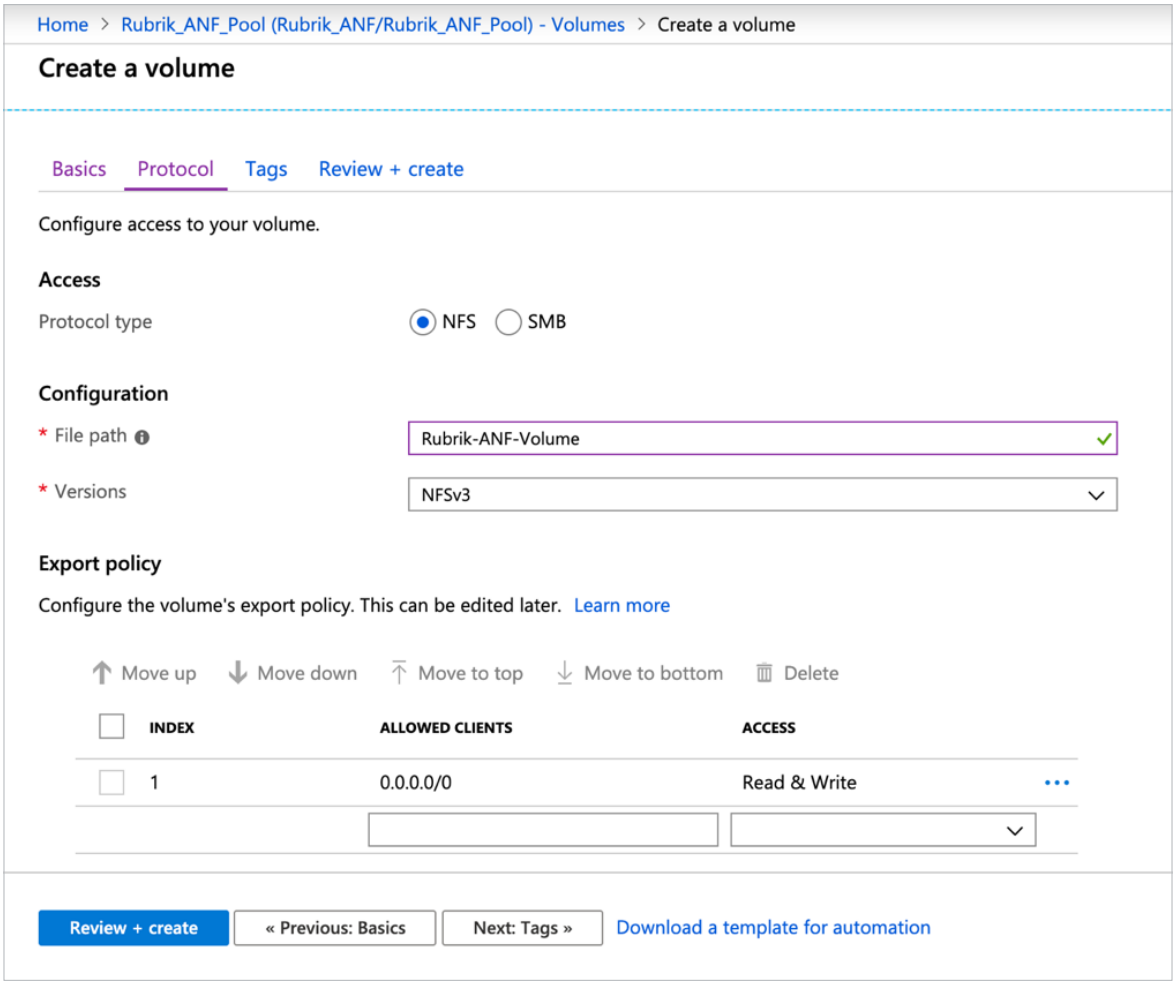
d. Specify the subnet address range, for example, **192.168.70.0/25**. Note that you cannot share the dedicated subnet with other resources.

e. Select **Microsoft.NetApp/volumes** for subnet delegation.

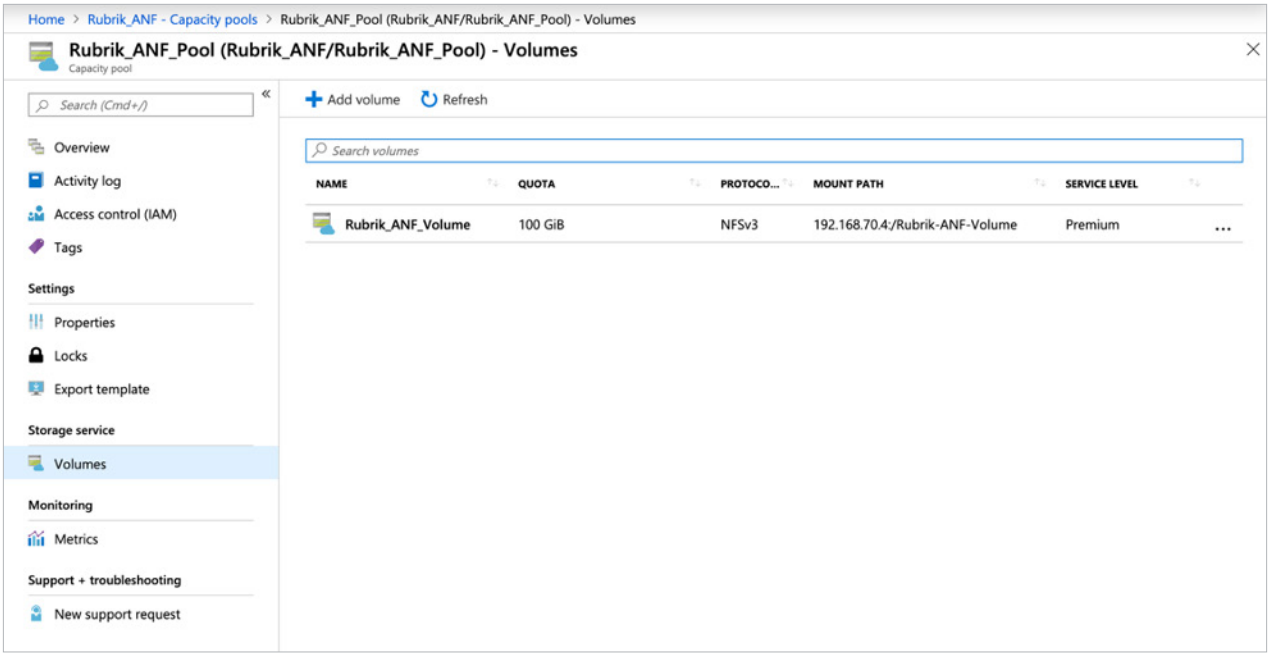
f. Click **OK** to create the Vnet.



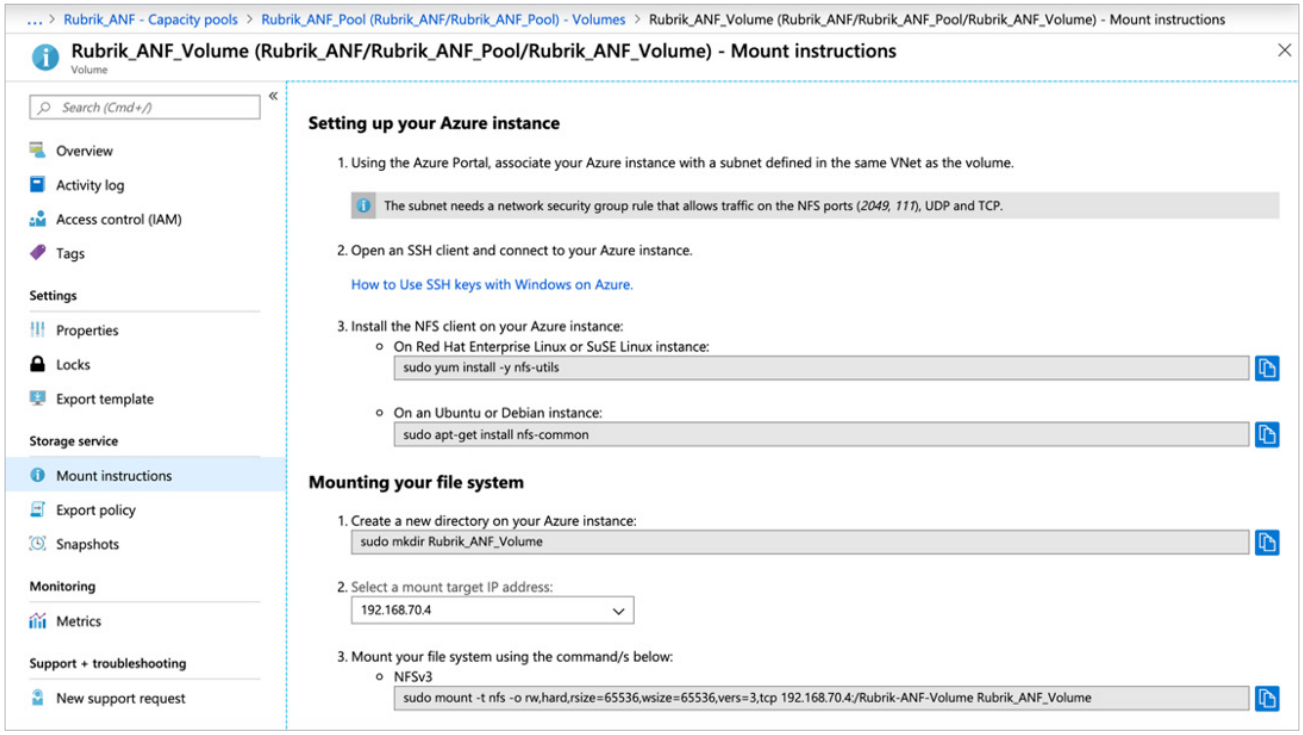
Back to the **Create a volume** wizard, select the desired file share protocol to access the share, choose the file path, and select the file share protocol version. Click **Review + create**:



Your newly created volume will now appear in the Volumes section of the **Capacity pool** blade, as shown below:



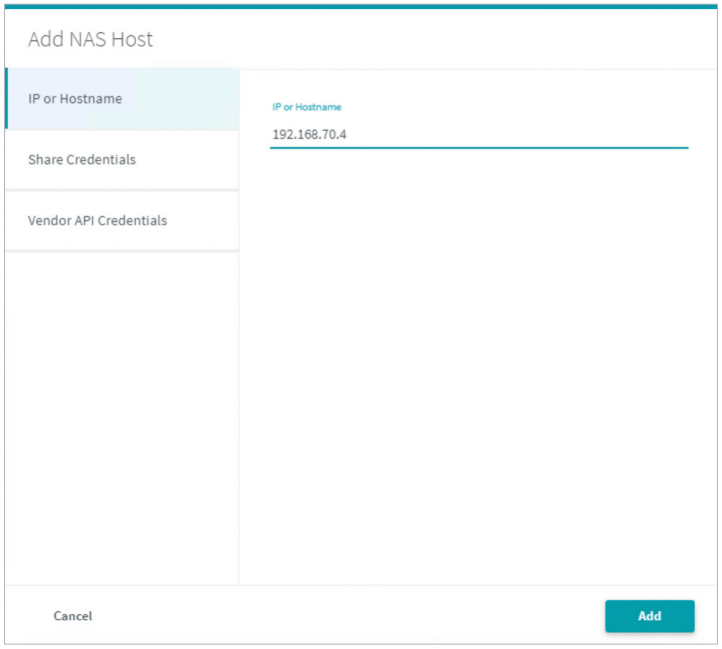
5. Click the **Volumes** blade, and then select the volume for which you want to mount. Click **Mount instructions** from the selected volume, and then follow the instructions to mount the volume:



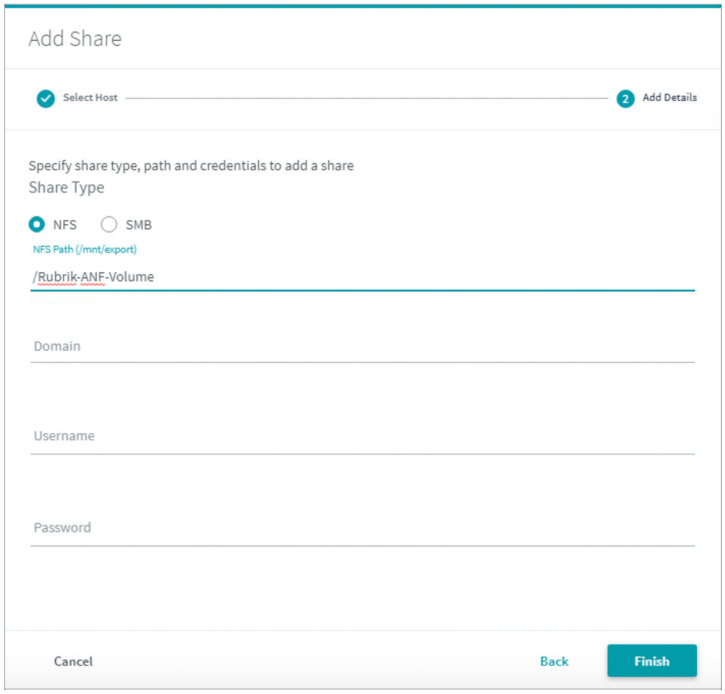
The screenshot below shows an example of mounting an Azure NetApp Files volume on a CentOS 7 client:

```
[amrutha@CentOS-7 ~]$ sudo mkdir Rubrik_ANF_Volume
[amrutha@CentOS-7 ~]$ sudo mount -t nfs -o rw,hard,rsize=65536,wsize=65536,vers=3,tcp 192.168.70.4:/Rubrik-ANF-Volume Rubrik_ANF_Volume
[amrutha@CentOS-7 ~]$ mount
sysfs on /sys type sysfs (rw,nosuid,nodev,noexec,relatime,seclabel)
proc on /proc type proc (rw,nosuid,nodev,noexec,relatime)
devtmpfs on /dev type devtmpfs (rw,nosuid,seclabel,size=4066508k,nr_inodes=1016627,mode=755)
securityfs on /sys/kernel/security type securityfs (rw,nosuid,nodev,noexec,relatime)
tmpfs on /dev/shm type tmpfs (rw,nosuid,nodev,seclabel)
devpts on /dev/pts type devpts (rw,nosuid,noexec,relatime,seclabel,gid=5,mode=620,ptmxmode=000)
tmpfs on /run type tmpfs (rw,nosuid,nodev,seclabel,mode=755)
tmpfs on /sys/fs/cgroup type tmpfs (ro,nosuid,nodev,noexec,seclabel,mode=755)
cgroup on /sys/fs/cgroup/systemd type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,xattr,release_agent=/usr/lib/systemd/systemd-cgroups-agent,name=systemd)
pstore on /sys/fs/pstore type pstore (rw,nosuid,nodev,noexec,relatime)
cgroup on /sys/fs/cgroup/perf_event type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,perf_event)
cgroup on /sys/fs/cgroup/hugetlb type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,hugetlb)
cgroup on /sys/fs/cgroup/net_cls,net_prio type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,net_prio,net_cls)
cgroup on /sys/fs/cgroup/pids type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,pids)
cgroup on /sys/fs/cgroup/cpuset type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,cpuset)
cgroup on /sys/fs/cgroup/cpu,cpuacct type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,cpuacct,cpu)
cgroup on /sys/fs/cgroup/memory type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,memory)
cgroup on /sys/fs/cgroup/devices type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,devices)
cgroup on /sys/fs/cgroup/blkio type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,blkio)
cgroup on /sys/fs/cgroup/freezer type cgroup (rw,nosuid,nodev,noexec,relatime,seclabel,freezer)
configfs on /sys/kernel/config type configfs (rw,relatime)
/dev/sda2 on / type xfs (rw,relatime,seclabel,attr2,inode64,noquota)
selinuxfs on /sys/fs/selinux type selinuxfs (rw,relatime)
systemd-1 on /proc/sys/fs/binfmt_misc type autofs (rw,relatime,fd=30,pgrp=1,timeout=0,minproto=5,maxproto=5,direct,pipe_ino=10913)
mqueue on /dev/mqueue type mqueue (rw,relatime,seclabel)
hugetlbfs on /dev/hugepages type hugetlbfs (rw,relatime,seclabel)
debugfs on /sys/kernel/debug type debugfs (rw,relatime)
/dev/sda1 on /boot type xfs (rw,relatime,seclabel,attr2,inode64,noquota)
/dev/sdb1 on /mnt/resource type ext4 (rw,relatime,seclabel,data=ordered)
tmpfs on /run/user/1000 type tmpfs (rw,nosuid,nodev,relatime,seclabel,size=815732k,mode=700,uid=1000,gid=1000)
192.168.70.4:/Rubrik-ANF-Volume on /home/amrutha/Rubrik_ANF_Volume type nfs (rw,relatime,vers=3,rsize=65536,wsize=65536,namlen=255,hard,proto=tcp,timeo=600,retrans=2,sec=sys,mountaddr=192.168.70.4,mountv
rs=3,mountport=635,mountproto=tcp,local_lock=none,addr=192.168.70.4)
[amrutha@CentOS-7 ~]$
```

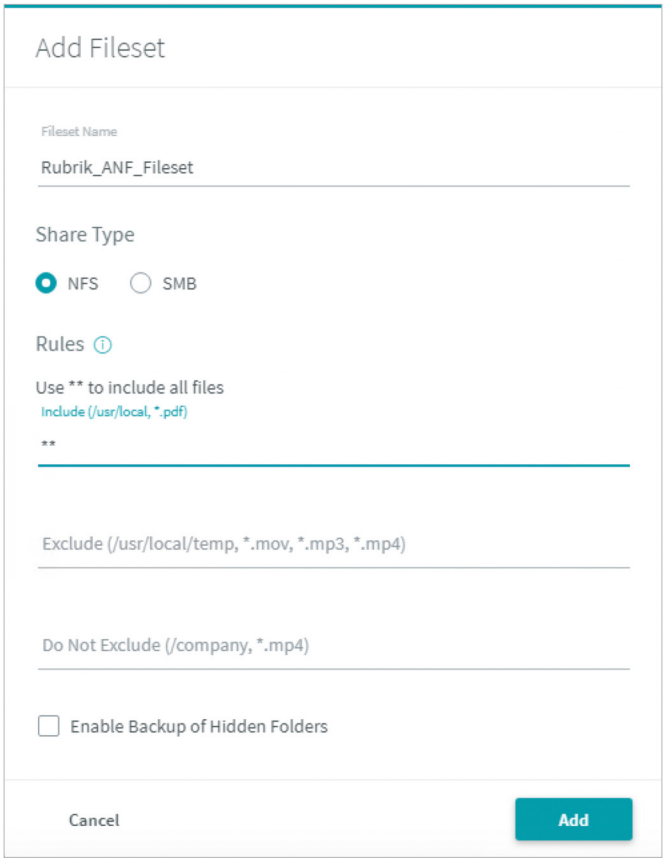
6. On Rubrik, click on **NAS Shares** under **Servers & Apps**. Click on **Add NAS Host** and provide the IP address or hostname of the NAS host retrieved from the above step (ANF mount target IP address). Click **Add**:



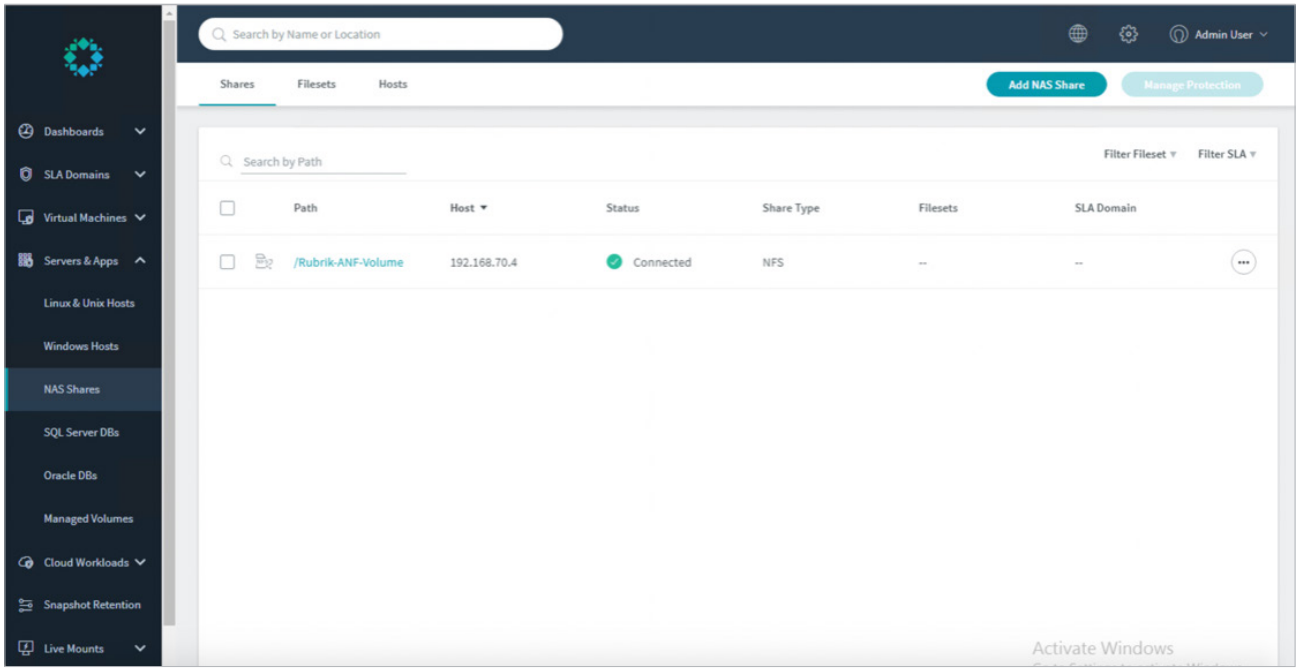
7. Add the ANF share details. Select the protocol – NFS/SMB. Add the ANF mount path and credentials (if any). Click **Finish**:



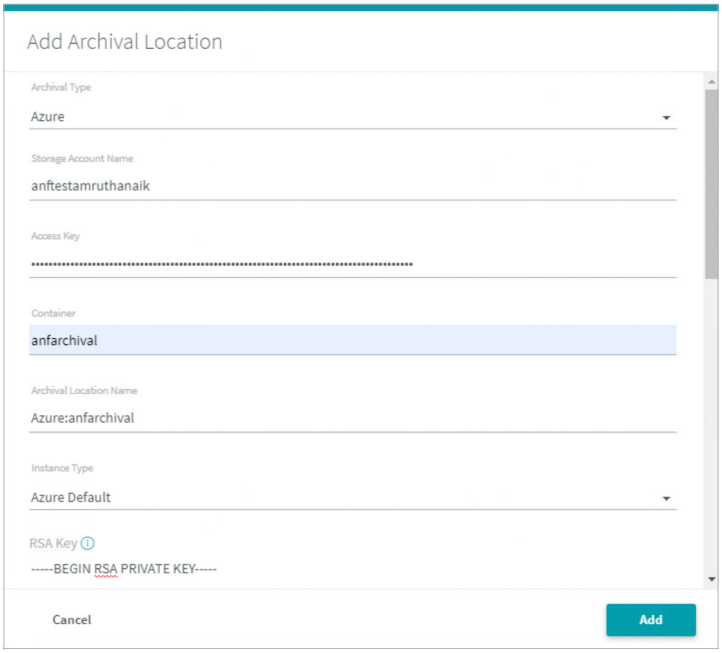
8. In the **Add Fileset** page, provide a Fileset Name, select the protocol and choose the files in the share that have to be protected. In this example, we have chosen to select all the files in the fileset. Click **Add**:



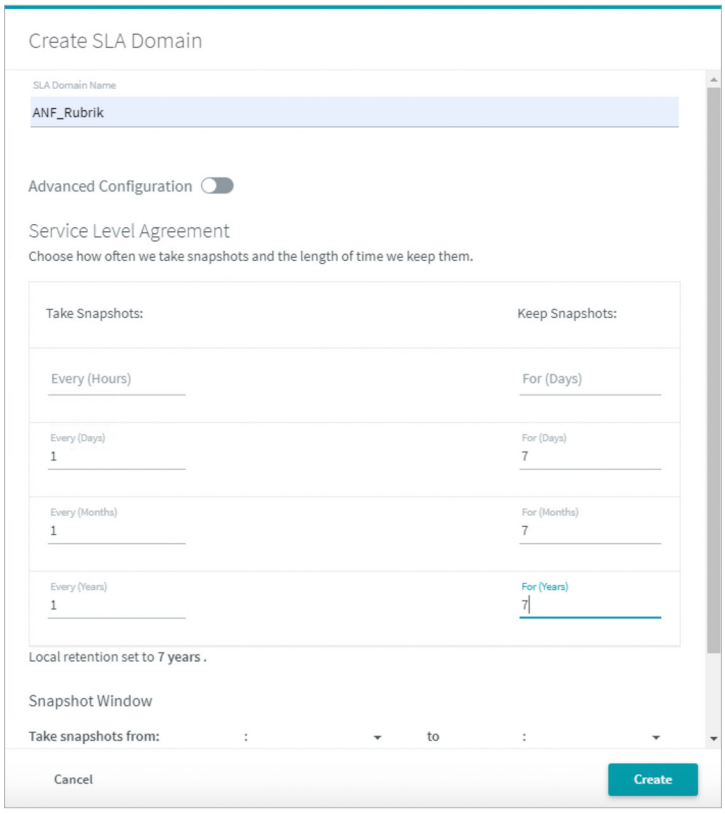
The newly added share is now listed in the Rubrik UI:



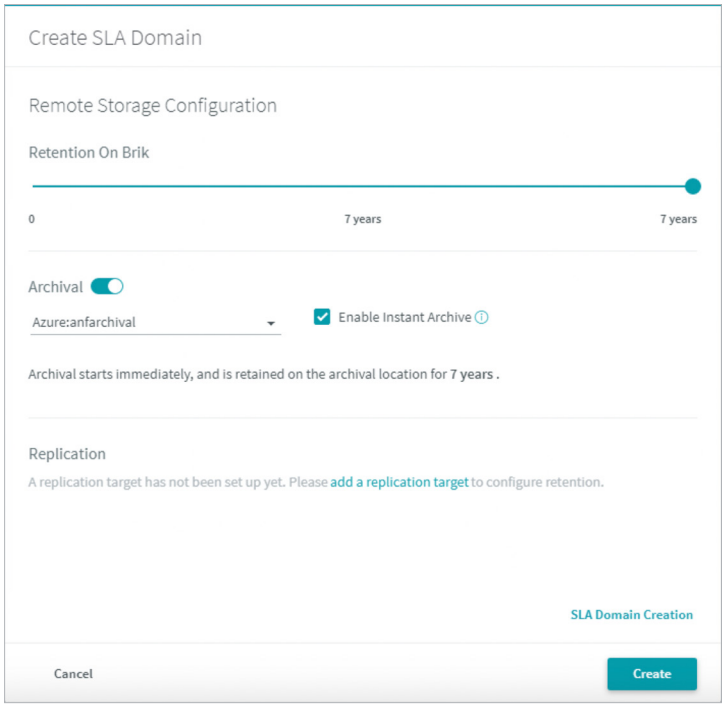
9. In the below step, we are adding an Azure container as an archival location for long-term retention. Provide Azure storage account details such as storage account name, access key, container name and click **Add**:



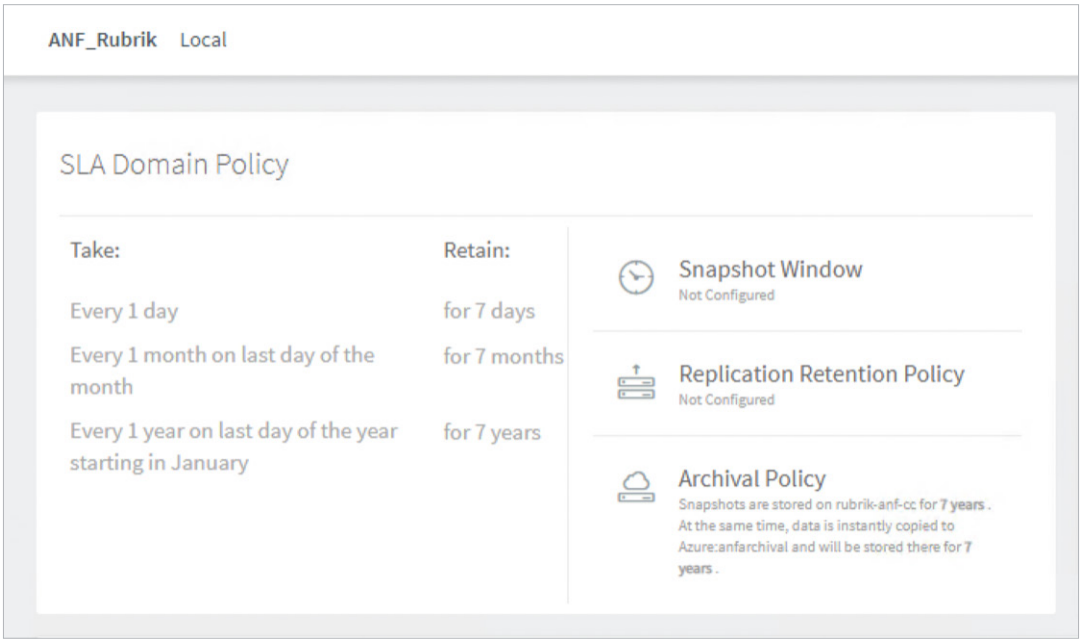
10. Create an SLA Domain policy and set the backup frequency and retention period:



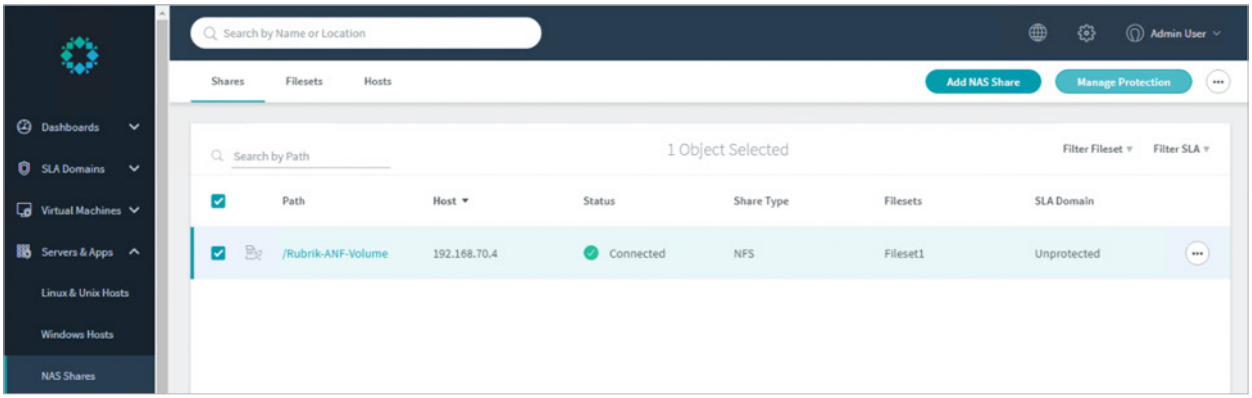
Before clicking **Create**, click the **Remote settings** button located at the bottom right-hand corner. Toggle on Archival, select the desired archival location and choose how long the backup data should be retained locally on the Rubrik cluster by using the **Retention On Brik** slider. Click **Create**:



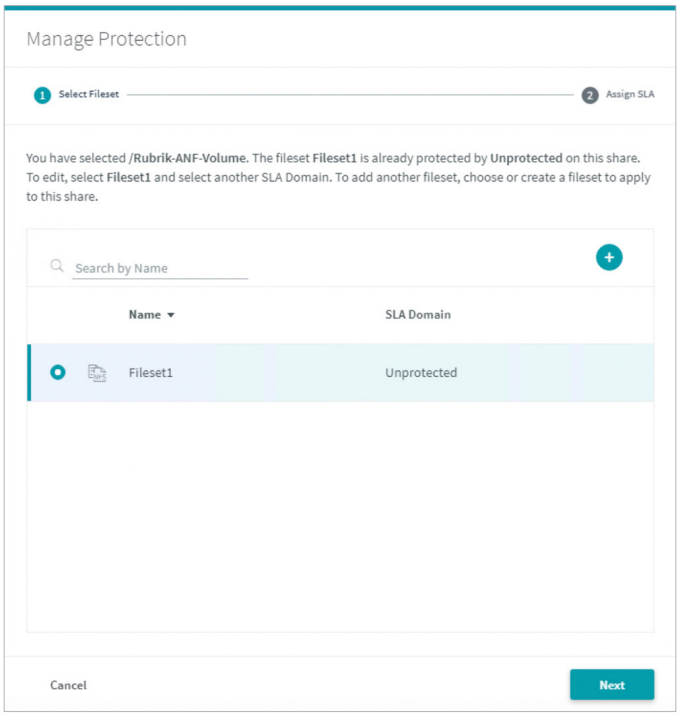
Once your new SLA Domain policy is created, you can see a summary of its settings by selecting it in the list of existing SLA Domains:



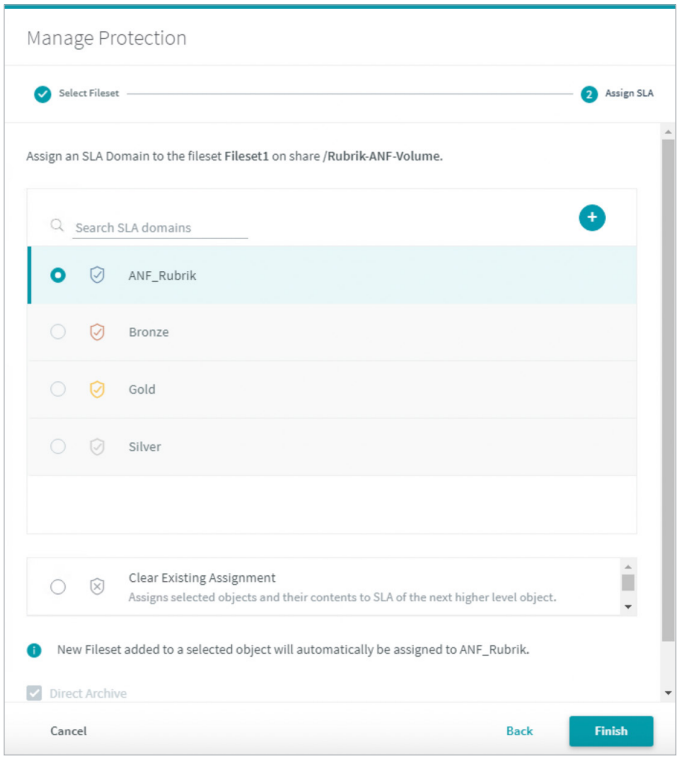
11. In the next step, assign the SLA Domain to the NAS fileset. Select the NAS fileset and click on **Manage Protection**:



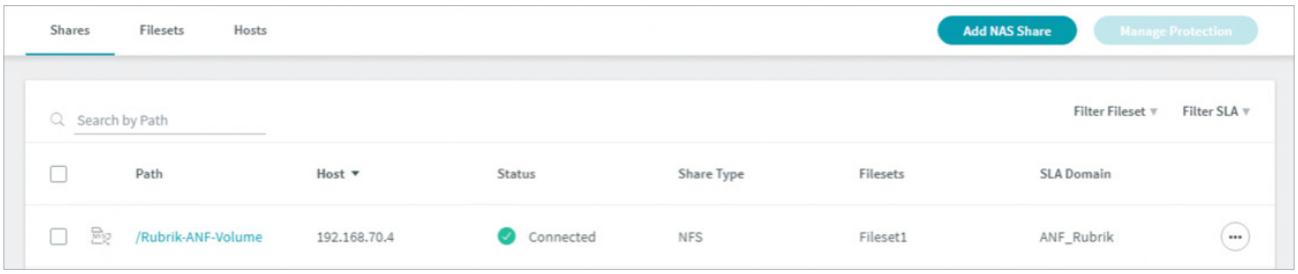
Select the Fileset that needs to be protected and click **Next**:



Assign the SLA domain policy created in the previous step and click **Finish**:



The Rubrik UI now shows which filesets are protected for this ANF volume, and by which SLA Domain policy:



BACKUP AND RESTORE PERFORMANCE

The purpose of the backup and restore performance reported in this section is to provide customers with a sense of what performance it is possible to achieve. It is by no means a guarantee that the performance one will observe in their own environment will be equal or better.

TEST ENVIRONMENT

For the purposes of the results below, the following infrastructure was used:

- 1TB of NFS data (1 million files)
- A 4 node Rubrik Cloud Edition (DS3v2 Instances)
- 40TB Capacity Pool in Azure NetApp Files Premium tier

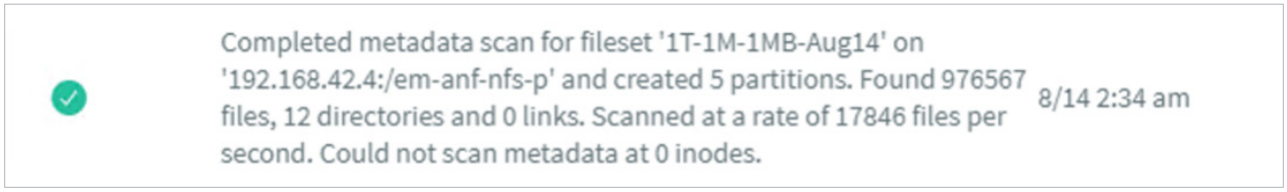
While Rubrik Cloud Edition can scale linearly to increase performance, compute and storage, these tests were performed with the default Cloud Edition installation.

BACKUP PERFORMANCE

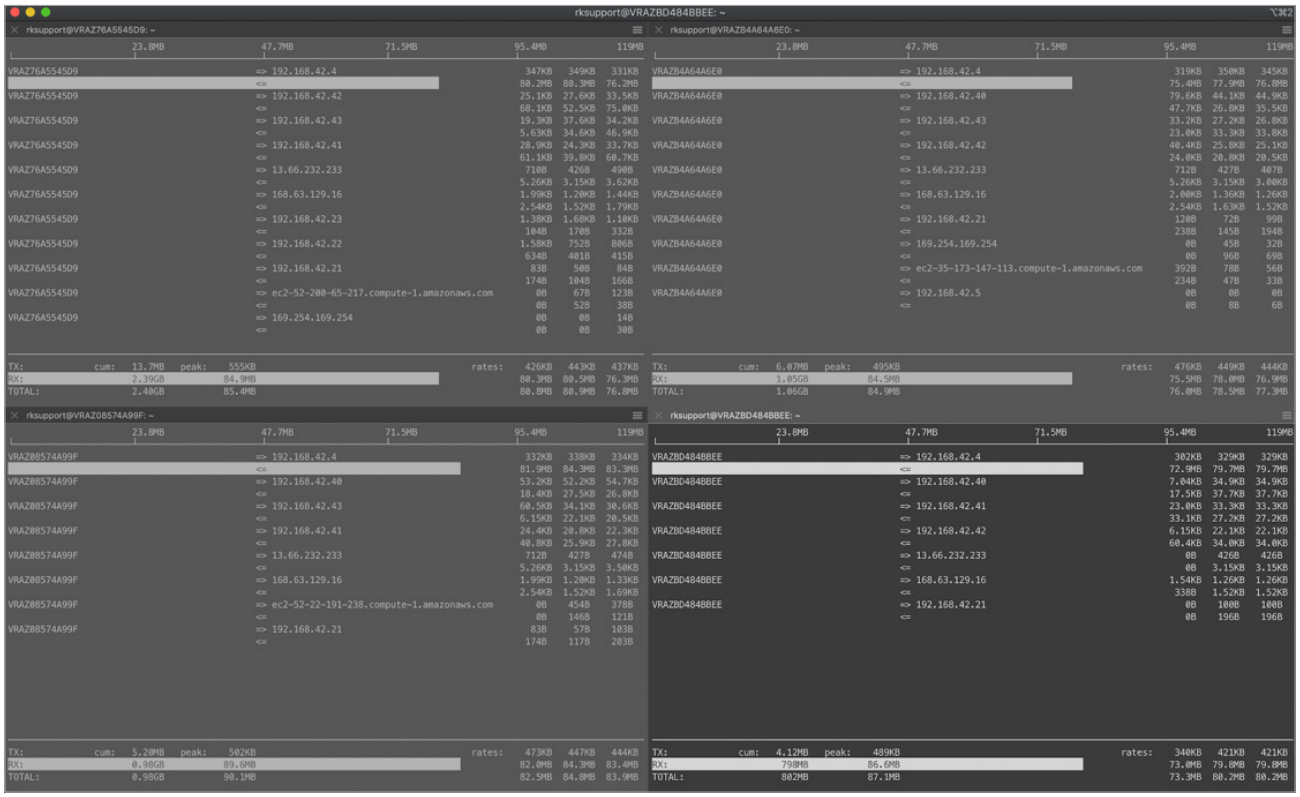
Let's take a look at the backup performance, starting with the first full backup, and followed by incremental backup.

FIRST FULL BACKUP

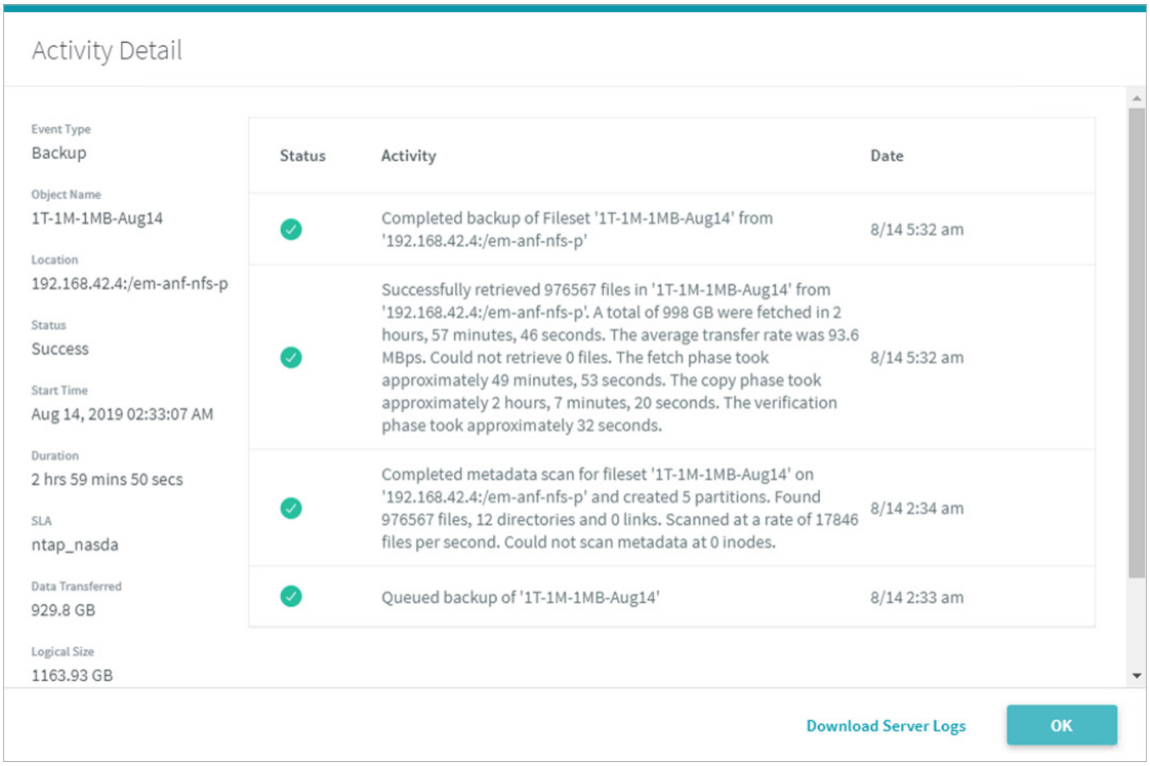
During the scan phase of the backup job, it took around 1 minute to iterate through the file system and scan the files, scanning at a rate of ~18,000 files per second. This scan performance was found to be consistent across all 3 tiers of Azure NetApp Files.



On ingest the fileset was automatically split up into partitions and ingested in parallel across the cluster at a rate of ~90MB/s per node – close to the maximum throughput achievable by a DS3v2 instance.



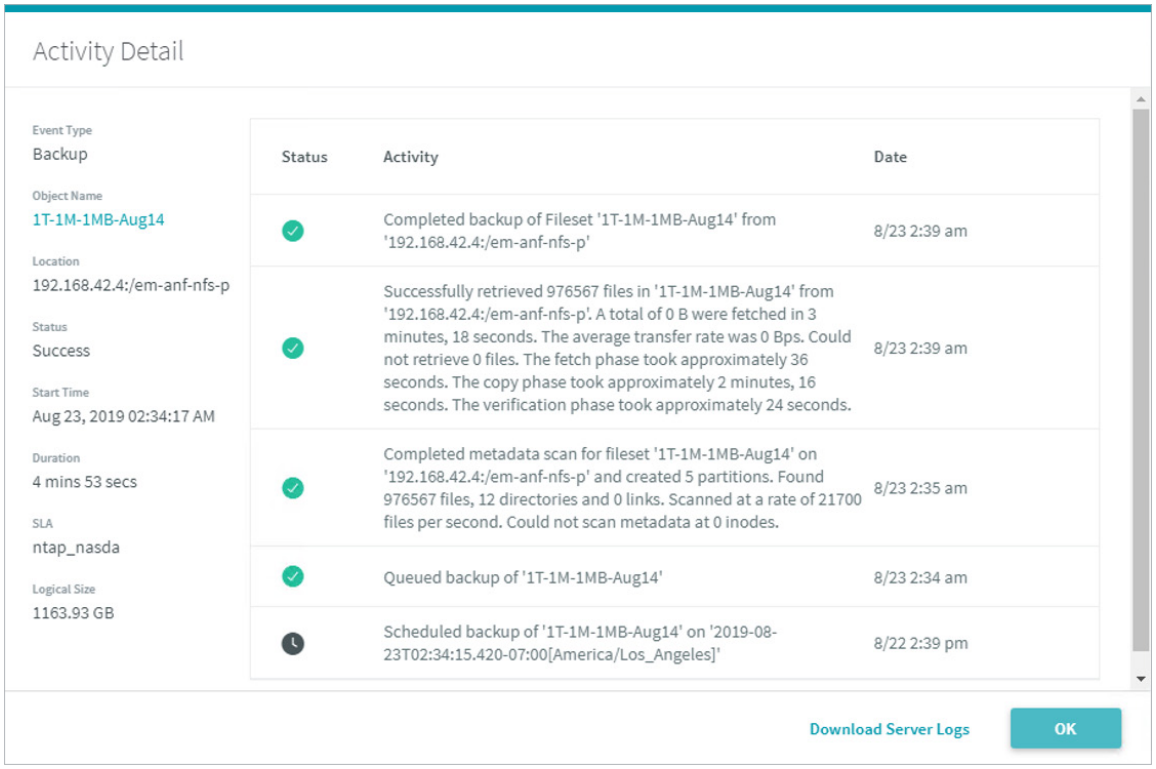
The first full backup of this dataset completed in 2hrs 59 mins, with an average end to end throughput of 93.6MB/s.



INCREMENTAL BACKUP

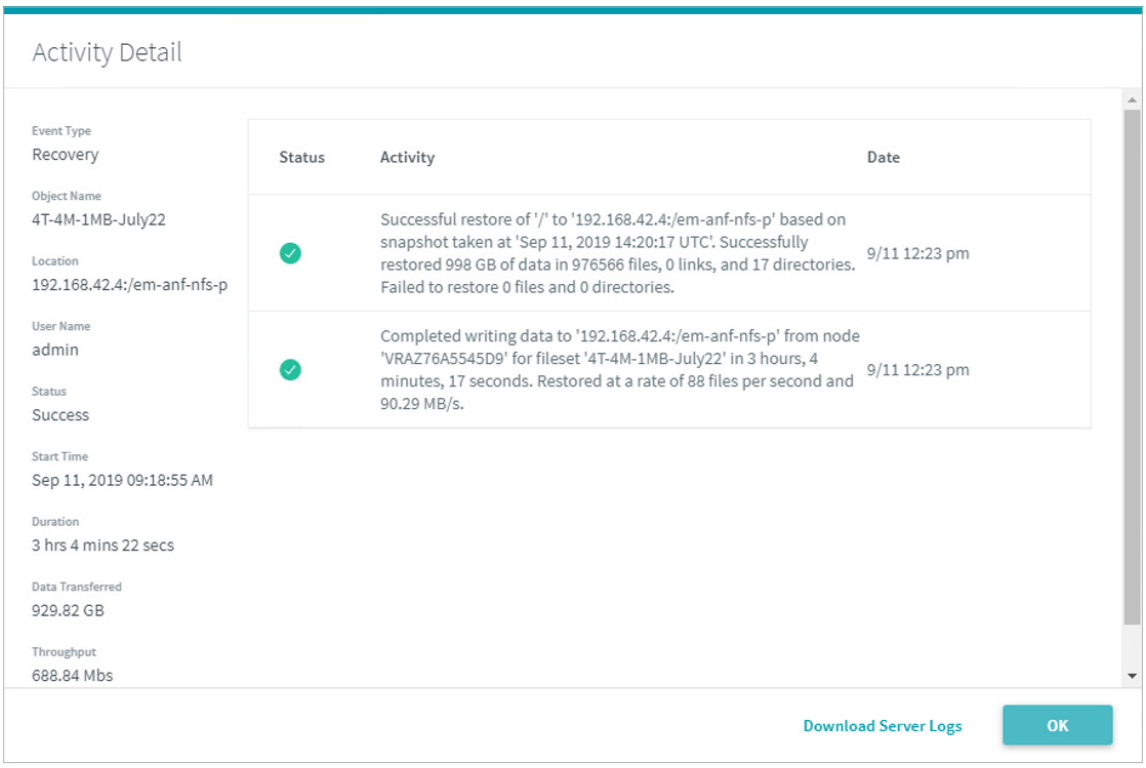
Another backup was then run on the same dataset as part of the defined SLA protecting the fileset.

Scan performance was again consistent, this time at just over ~21,000 files per second, with the backup being completed in 4 minutes 53 seconds.



RESTORE PERFORMANCE

Restoration performance of this fileset was consistent with the results from the backup. A full restore of the share was performed with data being automatically multi-streamed back to the source NFS device by the Rubrik cluster at 90.29MB/s. A full 1TB restore completed in 3 hours 5 minutes.



CONCLUSION

Azure NetApp Files was introduced jointly by Microsoft and NetApp to allow organizations to migrate existing enterprise-grade applications that use high performance file shares from on-premises to the cloud without requiring major changes, or that need high database performance such as Oracle or SAP in the cloud for instance. Running such applications in the cloud requires to have a solid data protection strategy in which backup must be fast, reliable, secure and easy to use.

Rubrik chose from day 1 to backup file shares using industry standard protocols such as NFS and SMB to overcome many of the limitations that usually come with NDMP-based solutions. Because the Azure NetApp Files service uses the same protocols to access data, it was straight forward for Rubrik to support the new data service. With the help of Rubrik CDM Cloud Edition, a virtual version of Rubrik CDM specifically built for Microsoft Azure, backups of data born in the cloud stay in the cloud.

As described throughout this document, Azure NetApp Files can be backed up by Rubrik with the same policy-based approach as usual, with extreme ease of use via the HTML5 web interface and sustainable performance for both backup and recovery.

To learn more about Rubrik's general approach of NAS backup, read our Protecting NAS at Scale with Rubrik white paper.

ABOUT THE AUTHORS

Amrutha Naik is a Senior Technical Alliances Manager at Rubrik. Prior to joining Rubrik, she worked as a Senior Product Manager at Tegile and also spent some years in Technical Marketing at EMC & NetApp gaining experience in virtualization, storage and data protection. Outside of work she enjoys painting, gardening, playing violin and using her husband as a lab rat for culinary experiments.

Ed Morgan is a Technical Product Manager at Rubrik. Before joining Rubrik he has worked in various consultancy and pre-sales roles for the past decade, with a particular focus on enterprise storage, virtualisation, and cloud, and spent 5 years serving in the British Army. Outside of work he enjoys rock music, CrossFit, endurance events, and writing about himself in the 3rd person. You can follow him on Twitter @mo6020

Pierre-François Guglielmi is the Technical Manager of Partner Solutions at Rubrik with many years of experience in virtualization and data protection. Pierre-François lives near Paris, France and enjoys CrossFit and playing music in his free time. You can follow him on Twitter @pfguglielmi.

rubrik

Global HQ
1001 Page Mill Rd., Building 2
Palo Alto, CA 94304
United States

1-844-4RUBRIK
inquiries@rubrik.com
www.rubrik.com

20200401_v2