

Share-to-Share Backup

Background

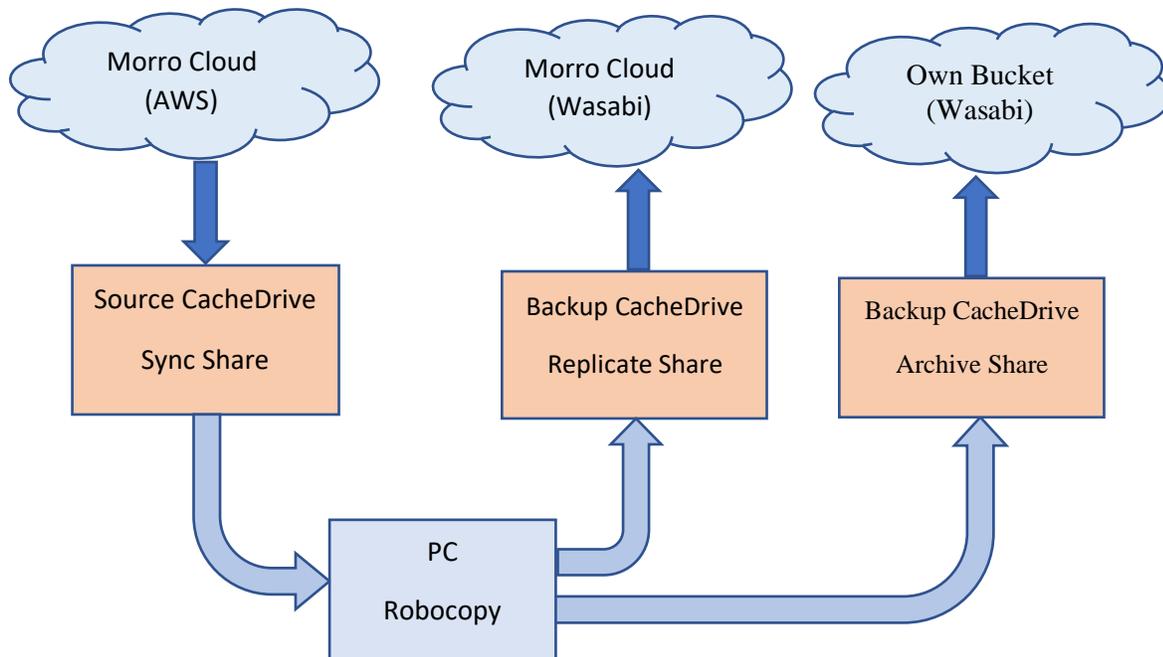
Morro Global File System stores data in the cloud storage provider's object storage such as AWS S3. AWS S3 provides high data endurance by maintaining redundant data copies in multiple Availability Zones. However, some IT organizations may have the policy to backup data from one cloud storage provider to a second provider. CloudNAS provides such cloud redundancy by using backup between different share types.

Types of Backup Target Shares

Sync shares are used by most of the CloudNAS users to share files among one or multiple sites. There are two ways to backup a Sync share from one cloud storage provider to another:

Backup to Replicate Share: The first approach is to create a second redundant Morro file system for the share. Replicate share is used for this purpose and the user runs backup software periodically to incrementally copy from Sync Share to Replicate Share.

Backup to Archive Share: The second approach is to backup to an object storage bucket which the customer owns. In this approach, a second copy of the file system is created. Archive Share is used for this purpose and the user runs backup software periodically to incrementally backup Sync Share to Archive Share.



Comparison of the two methods:

	<i>Backup to Replicate Share</i>	<i>Backup to Archive Share</i>
File Attributes	YES Attributes such as permissions and timestamps are stored in the Morro Global File System.	YES Attributes are stored in the object's meta data.
File Size	Unlimited Morro Global File System has no file size limit.	Limited up to the storage provider's object size, usually in the TB range.
File Name	No Restriction Same as source Sync Share	Restricted by object storage provider's naming rule, original name in kept in object's metadata.
Access of backup data	Multi-site Replicate share can be converted to Sync share for full multi-site sync.	Single-site Only from the target CacheDrive
DR without Morro Data service	YES Can access the full backup dataset without Morro if the backup target CacheDrive is larger than the backup dataset.	YES Data (without metadata) can be accessed from the storage provider directly but need CacheDrive access to retrieve both data and metadata.

Share-to-Share Backup with Robocopy

In the following example, we show the use of Robocopy to perform the backup between the source and target shares.

Setup

- Have a dedicated CacheDrive as the backup target. This can avoid cache thrashing of using the same CacheDrive as source and target.
- Make sure the admin or user has enough privilege to perform the copy operations.
- Create the Replicate or Archive share as backup target. You can set the backup schedule to either "Continuous" or "Scheduled". In a tight space scenario, Continuous mode will free up disk space as soon as files are uploaded.
- For the following example, map the source Sync share as S: drive and the target share as T: drive.

Run Robocopy

Run Robocopy with the following options to copy data from S: to T:

```
robocopy S:\ T:\ /COPY:DATSO /PURGE /XO /E /FP /NP /V /LOG:robocopy.log
```

Several essential options are used here:

Option	Explanation
/COPY:DATSO	Copy D=Data, A=Attributes, T=Timestamps, S=NT ACL, O=Owner Info
/PUGE	Delete destination files/dirs. Which no longer exist in source.
/XO	Exclude older (or the same) files
/E	Copy subdirectories including empty ones.
Do not use /R:n	Number of retries. Recommend to use the default 1 million retries. If you ever use /REG option to save in registry, set n to large number.
Do not use /W:n	Wait time between retries. Recommend to use default 30 seconds.

With these options, it copies only the new or modified files. Also, it will delete files and directories in the target share if they are deleted in the source share. Since the source is a live share, the file tree may be changed when data copy is going on. But the subsequent run will copy the incremental changes.

More Technical Details

Robocopy with the above command line option will do incremental copy. The first run may take long depending on the total data size. However, the successive incremental runs should be faster depending on the size of data changes in between Robocopy runs.

Source CacheDrive may incur cache miss and trigger file download. If the file is large, it may cause timeout error on the backup software. Robocopy waits until the file content is downloaded. For the second issue, Robocopy will retry at 30-second intervals.

During the Robocopy copy operation, the target CacheDrive may encounter temporary disk full. (You will see ERROR 112 (0x00000070) until disk space is available.) However, since we use the “Continuous Mode”, uploaded files are swapped out to increase the disk available space. Therefore, CacheDrive with cache size smaller than the backup dataset can be used, although the Robocopy operation may take longer.

We recommend cache size (usable space) to be at least 3 times larger than largest file size. If the largest file size is 500GB, then you will need 500GB x 3 + 200GB (OS, workspace, etc.), so 2TB is good size. But we prefer minimum size of 1TB in CacheDrive.

Conclusion

CloudNAS provides multiple ways to enhance data endurance as well as data availability by replicating data between cloud storage providers as well as between Morro Data and customer’s own cloud storage bucket.