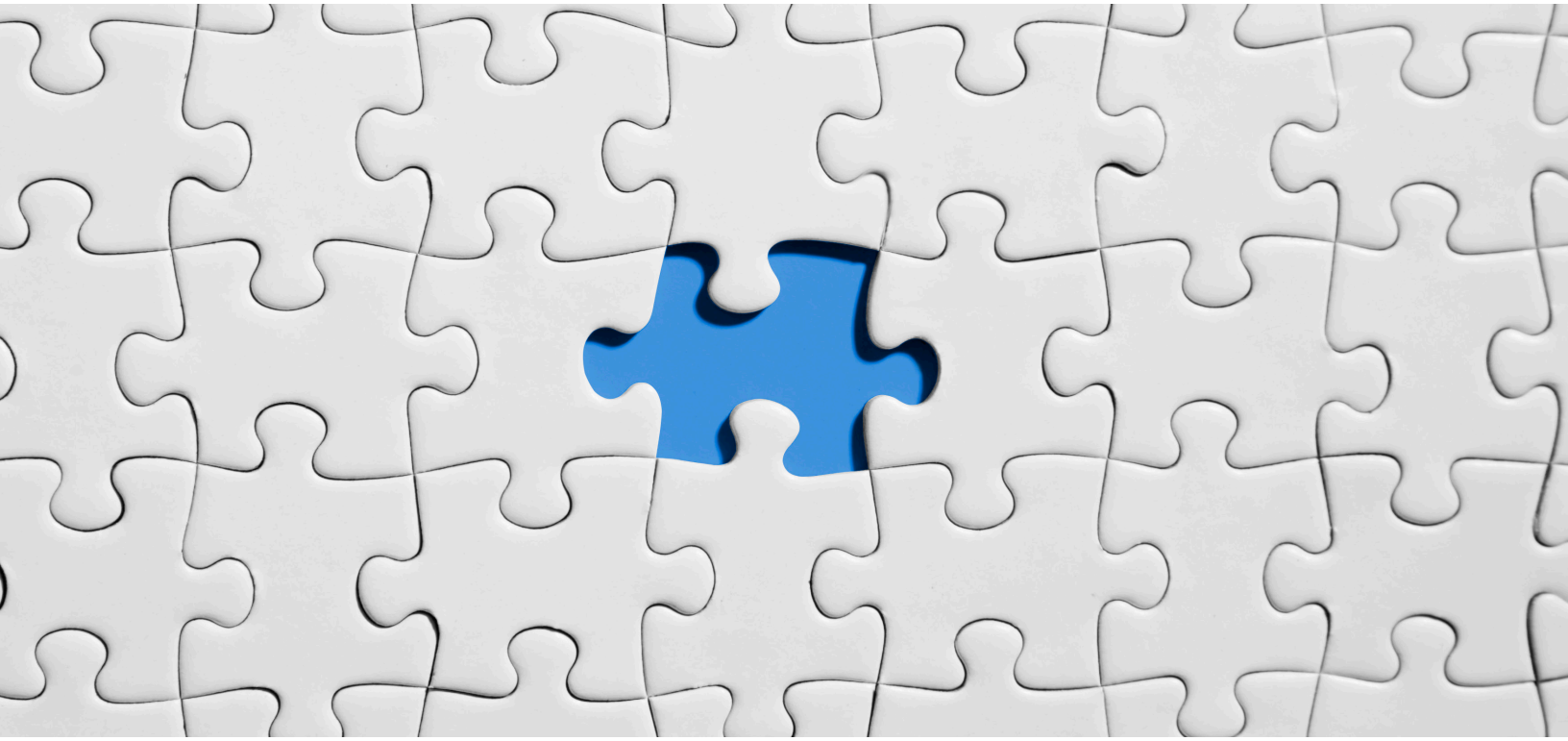
DELL Technologies

NETWORKER CLONE OPTIMIZATION



Anay Pathak

Advisory Technology Consultant

CTO Ambassador – Office of the CTO

Dell Technologies

Anay.pathak@dell.com

Deepak Verma

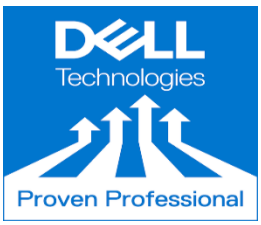
Software Senior Principal Engineer

Dell Technologies

Deepak.verma@dell.com

DELL Technologies Proven Professional

Knowledge Sharing Article



The Dell Technologies Proven Professional Certification program validates a wide range of skills and competencies across multiple technologies and products.

From Associate, entry-level courses to Expert-level, experience-based exams, all professionals in or looking to begin a career in IT benefit from industry-leading training and certification paths from one of the world's most trusted technology partners.

Proven Professional certifications include:

- Cloud
- Converged/Hyperconverged Infrastructure
- Data Protection
- Data Science
- Networking
- Security
- Servers
- Storage
- Enterprise Architect

Courses are offered to meet different learning styles and schedules, including self-paced On Demand, remote-based Virtual Instructor-Led and in-person Classrooms.

Whether you are an experienced IT professional or just getting started, Dell Technologies Proven Professional certifications are designed to clearly signal proficiency to colleagues and employers.

Learn more at www.dell.com/certification

Table of Contents

Introduction

The storage device that you use for the initial backup is often a compromise between several factors, including location, availability, capacity, speed, and cost. As a result, the backup data on initial target storage is not an ideal storage for retaining the data for a longer period of time

The need to clone data is normally driven by a requirement for additional protection, or to move data to a specific media type or location. In both cases, the priority is to secure the data as quickly as possible.

There is a higher probability that restore requests received within the first 48 hours would be linked to data corruption on primary site and in such cases, we would be able to recover data from the local backup copy. If there is a local disaster recovery or site loss, the recovery actions and objectives are likely to be very different. Selected systems and services are assigned specific priorities, recovery point objective (RPO) values, and recovery time objective (RTO) values.

Cloning and staging enables you to use storage devices more effectively by moving data between different types of devices. You can copy the data that is stored on local tape devices to other devices in remote locations without impacting the initial backup performance. You can copy backups from disk devices to tape device to facilitate offsite or long-term storage. When data is moved from disk to tape, the space reclaim is more effective

NetWorker cloning with Data Domain

The clone process allows you to:

- Create a duplicate backup and secure it offsite
- Transfer data from one location to another
- Verify backups

You can clone volumes and save sets. The clone process copies existing save sets from a volume in one device to a volume in a different device. The target volume can be the same media type or a different media type than the original.

Clone requirements & considerations

The Clone Data Domain must be running on an operating system version that is similar to or later than that of source Data Domain.

NetWorker requires two or more storage devices to perform a clone operation. One device contains the volume with the original data and the other device contains the volume to which NetWorker copies/clones the data. The clone data must reside on a volume that is different from the original volume. Each clone volume can only contain one instance of a cloned save set, even if the clone operation did not complete successfully. For example, if you want to create three clone copies of a save set, NetWorker must write each clone save set to a separate volume. As a result, you would need three separate volumes.

Note: When configuring multiple clone workflows for a scheduled clone, if a single backup pool has multiple save sets, ensure that each clone workflow is streamlined to split the list of save sets that must be cloned. If you attempt to clone a common SSID from a backup pool using multiple workflows into a single clone pool, using only backup pool as the filter, the clone action might result in a media waiting event. Ensure that the clone workflows in a single backup pool that have multiple save sets are separated

by multiple clone pools. Best practice is to configure multiple workflows using additional filters along with the backup pool.

When using a tape library with multiple devices, NetWorker Server automatically mounts the volumes required to complete the clone operation. When using standalone tape devices, you must manually mount the volumes. A message in the **Alert** tab of the **Monitoring** window indicates which volumes to mount.

Often businesses choose devices for the initial backup based on speed or cost requirements. NetWorker supports cloning or staging data to a device type that differs from the source data volume. A common cloning or staging scenario includes write to an Advanced File Type Devices [AFTD]/Dedupe storage and then clone/stage to tape or another dedupe storage. This scenario allows for an extended retention period without increasing disk space requirements. The use of deduplication can also provide efficient use of storage. Cloning to or from deduplication devices can ensure that these devices are used effectively. If the clone operation includes save sets from different devices, and you want all the save sets to be written to the same volume, include only one volume in the clone target pool.

Note: It is recommended that you do not write Network Data Management Protocol [NDMP] and non-NDMP data to the same clone volume because the number of file marks and positioning on the device differs for both data types.

Save set status

NetWorker does not clone save sets that are recyclable or eligible for recycling. If NetWorker encounters a save set that is not browsable, the save set is skipped and is not cloned. However, the clone status is successful.

Recovery scenarios

When you clone data, you provide the datazone with an alternative data recovery source, which helps protect against media loss or corruption. However, if the media is located in one of the following locations, then the second copy of the data is still vulnerable to major disasters that can affect the entire site:

- On the same tape library as the original data volume
- On a deduplication device within the same data center
- In a Data Domain environment in an onsite safe

Sometimes, you may require more copies of a save set to ensure that all the recovery scenarios are accommodated while maintaining the expected return on investment. This requirement may not apply to all clients and all data, or be practical. However, consider the reasons for cloning to ensure that the cloning strategy meets requirements and expectations.

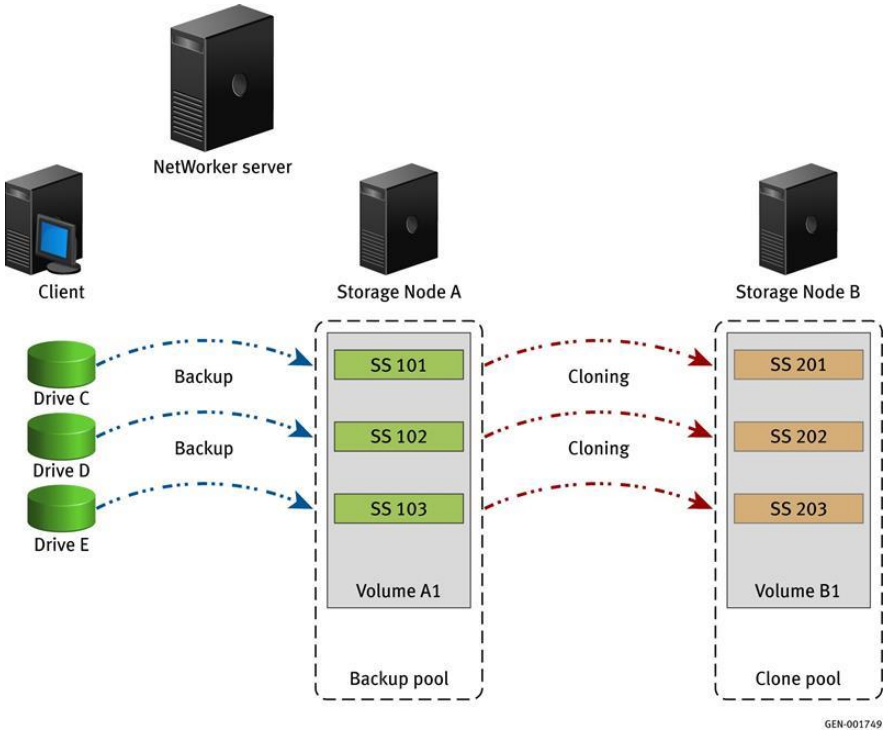
Changing the target device or moving tapes to a second location after the cloning operation completes, can provide additional protection.

NetWorker cloning example

In this example, a backup of a client with three data drives creates three save sets. These save sets are stored on a volume that is accessible through Storage Node A. Once a cloning action occurs, the copies of these save sets are sent to eligible devices in the clone pool on Storage Node B.

In this figure:

- A client performs a backup of three data drives to Storage Node A. NetWorker creates three save sets; one save set for each data drive.
- A clone operation reads the data from the volumes on Storage Node A, and then copies the save sets to Storage Node B.



Replication of backup data – Optimized approach

Today, having two backup copies is a business requirement. The objective is to have a secondary copy readily available for disaster recovery (DR) purposes. However, due to cost constraints, many organizations opt for Active-Active data center (DC) approach where half of their infrastructure runs from one DC and another half runs from another DC.

Setting up backup infrastructure and creating backup policies enables you to have backup copies of all the servers available in both DCs. This is required so that if one of DC goes down at time, backed-up data can be restored from the other DC without delay/loss.

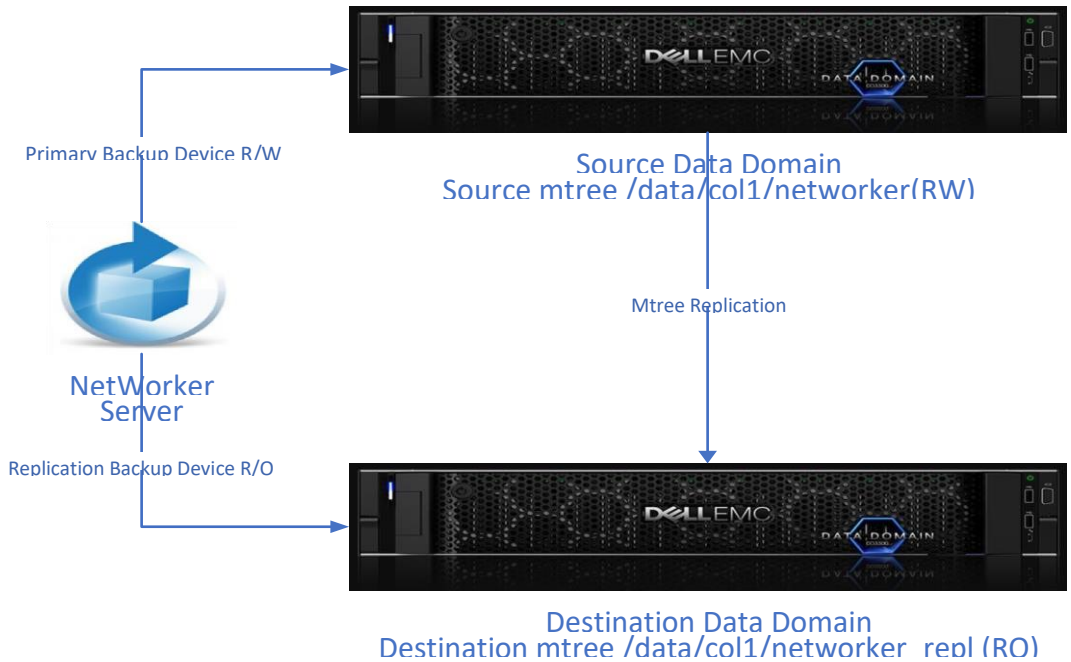
To establish the same with Dell NetWorker and Data Domain as a backup software and target storage, most of backup administrators opt for the native approach of having daily incremental and weekly full backups with clone operations.

Generally, clone action runs after completion of every backup and a policy is not marked completed until clone action finishes. There are scenarios where clone action takes more time to run as compared to backup. The backup actions are normally within the DC, i.e. the source or the client server is in the same DC where target Data Domain is placed. So, when a backup action runs, it uses intra-DC network and gets completed faster. Whereas, in the case of clone action, the job needs to write on Data Domain, which is hosted in another DC, thus, going thru inter-DC network.

Hence, to save on this extra time for clone action and still get two backup copies, the following was developed and matured to achieve two backup copies without Clone Controlled Replication (CCR).

In this concept, once the backup action is completed, the policy will be marked as 'Done' in NetWorker Management Console (NMC), whereas backed-up data will be migrated or replicated to another Data Domain using DD m-tree replication.

The devices in NetWorker are created such that NetWorker would not know about two backup copies, i.e. there will not be any information of secondary or cloned backup copy in NetWorker Catalog but still it would be able to read data from 'Read-Only' device of the second Data Domain.



Prerequisites

To establish the same, some prerequisites need to be followed:

A user should be created on both Data Domain devices and must have 'ddboost' access. ddboost account on both Data Domain devices will preferably have the same UID.

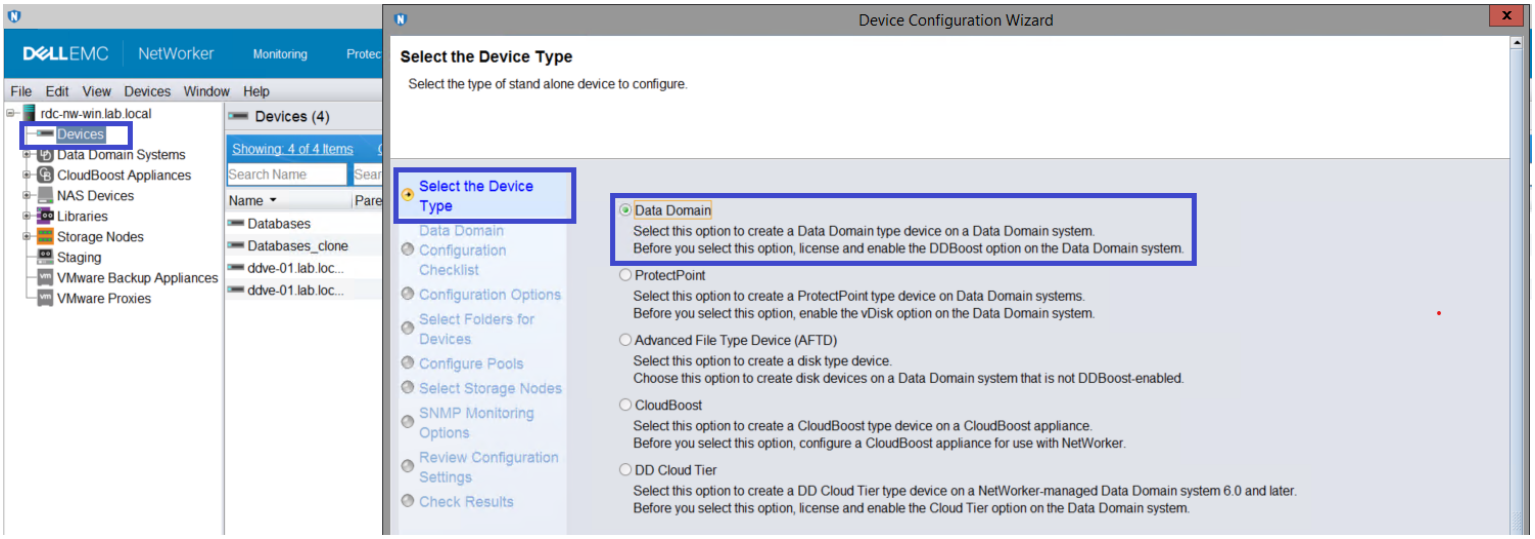
Source Data Domain

```
 sysadmin@DDVE-01# user show list
User list from node "localhost".
Name                            Uid   Role             Last Login From   Last Login Time            Status    Disable Date
------------------------------  ---   ---------------  ---------------   ------------------------   -------   ------------
sysadmin                        100   admin            10.91.137.32      Tue Jan 19 14:58:56 2021   enabled   never
prabhk1                         500   backup-operator  <unknown>         never                      locked    never
ddboost                         502   backup-operator  10.118.236.87     Thu Jan  7 20:53:39 2021   enabled   never
------------------------------  ---   ---------------  ---------------   ------------------------   -------   ------------
3 users found.
```

Procedural Steps

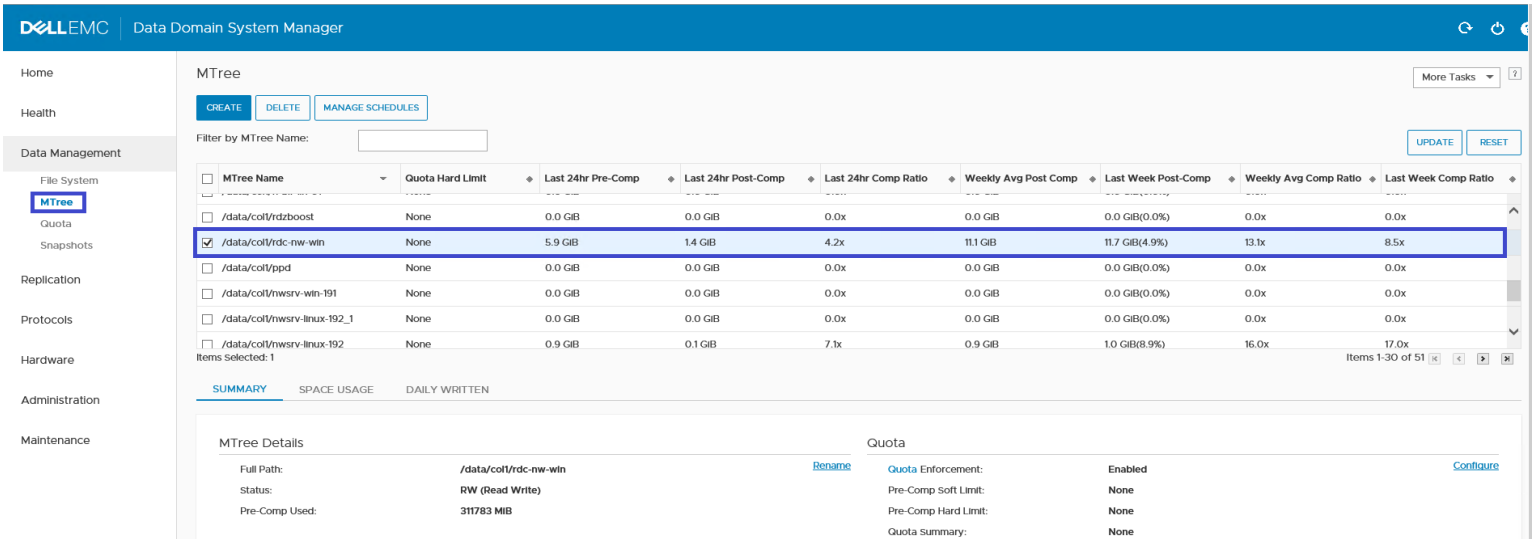
Below is the step-by-step procedure which needs to be followed to implement this design:

1. Create a backup device on NetWorker through New Device Wizard in NMC for source Data Domain along with desired Media Pool.

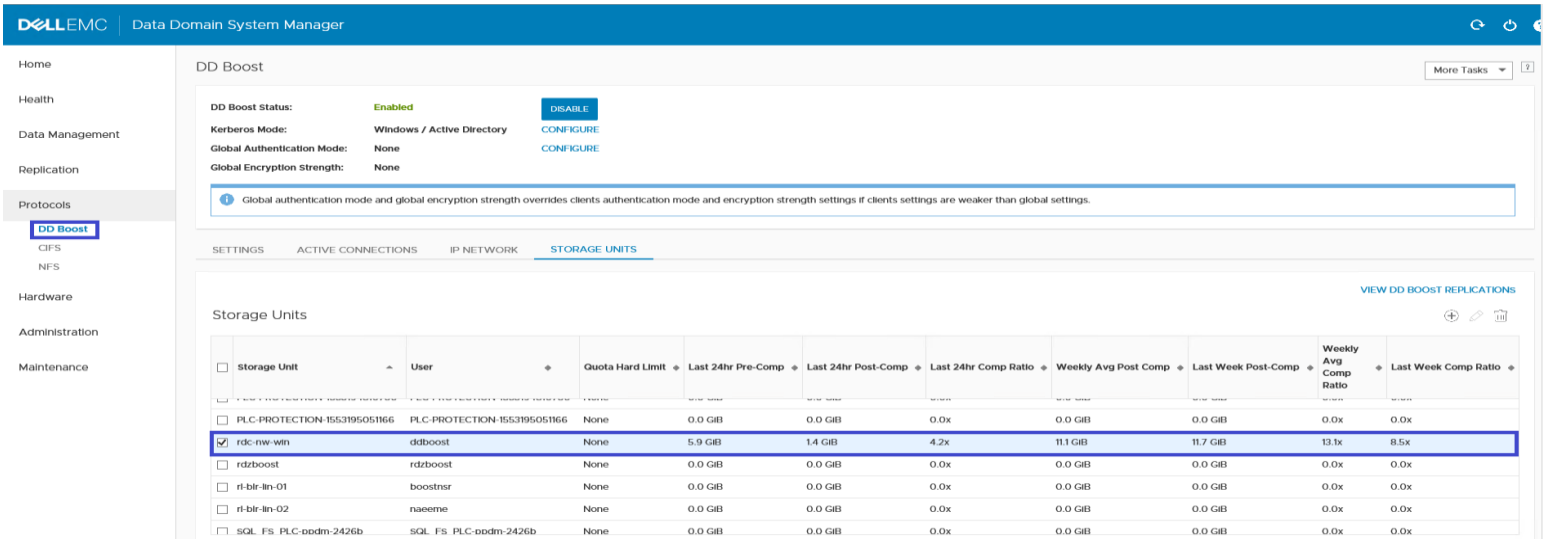


2. This will create an mtree and a corresponding DDBoost Storage Unit on source Data Domain .

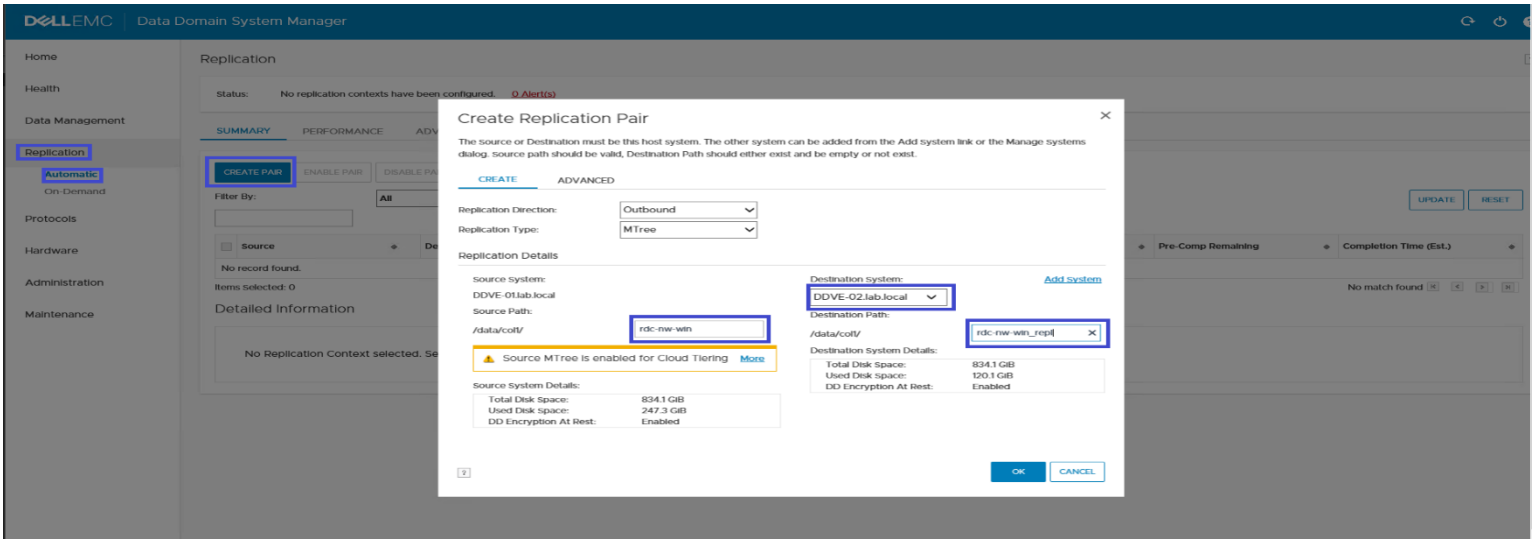
NetWorker Mtree on Data Domain



NetWorker ddboost storage unit on Data Domain



3. Configure replication of NetWorker mtree from source Data Domain to target Data Domain .



4. Once replication is setup and completed, use below command on target Data Domain Command Line Interface (CLI) to change the destination mtree to DDBoost storage unit.

   'ddboost storage-unit modify <storage unit name> user <ddboost user name>'

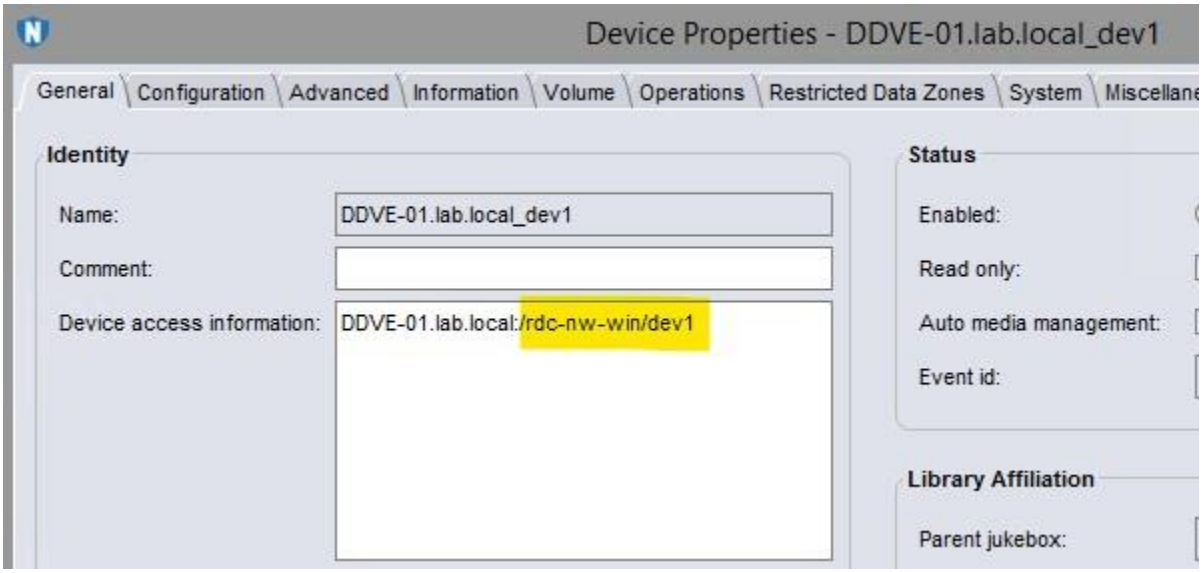
   e.g.

   sysadmin@DDVE-02# ddboost storage-unit modify rdc-nw-win_repl user ddboost

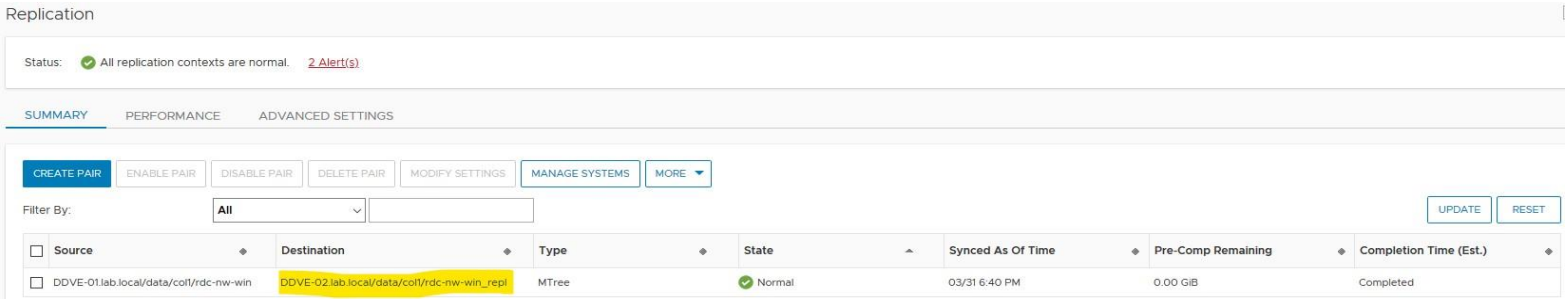
   Storage-unit " rdc-nw-win_repl " modified for user "ddboost".

5. Manually configure this target DDBoost storage unit in NetWorker as a device.

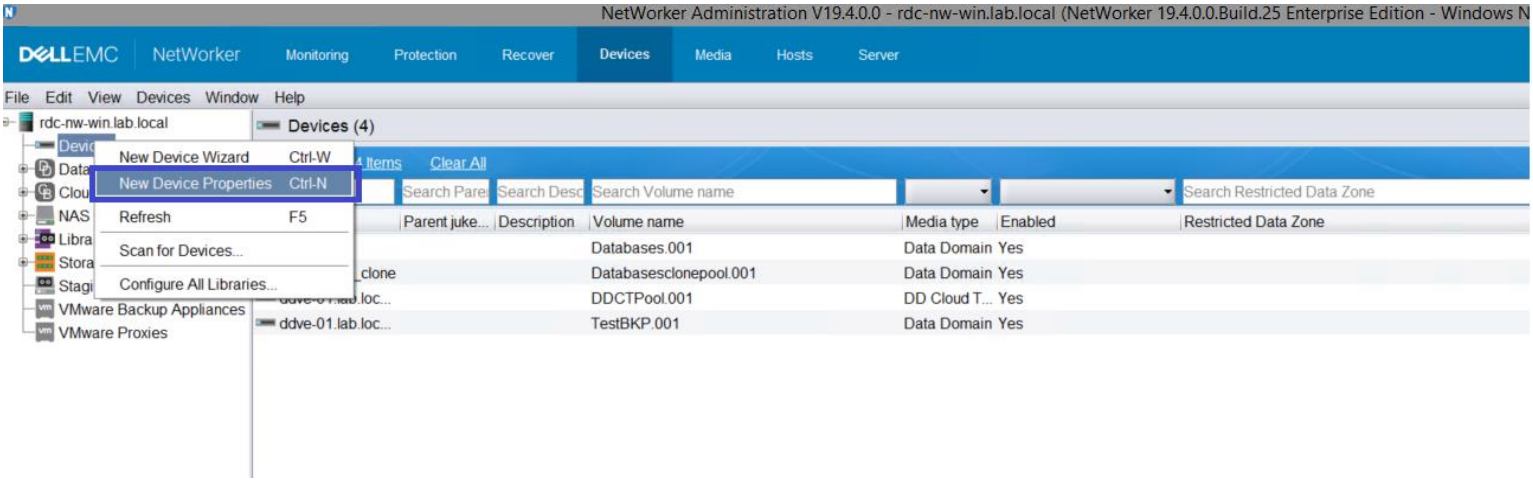
   a. Note down the folder path of the source device from the device properties in NMC.



6. Note down the replicated mtree name from Data Domain “Replication” page

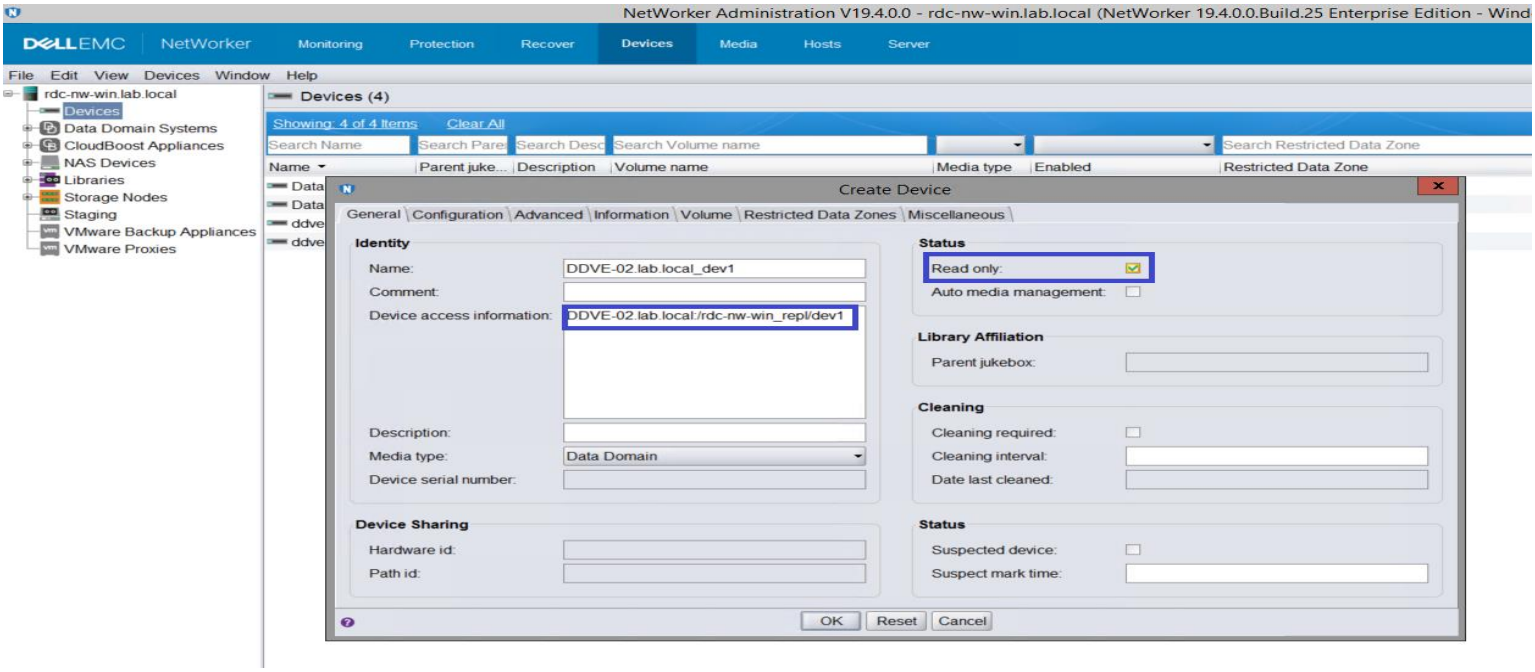


7. In NetWorker NMC on devices tab, Right click->New Device Properties.

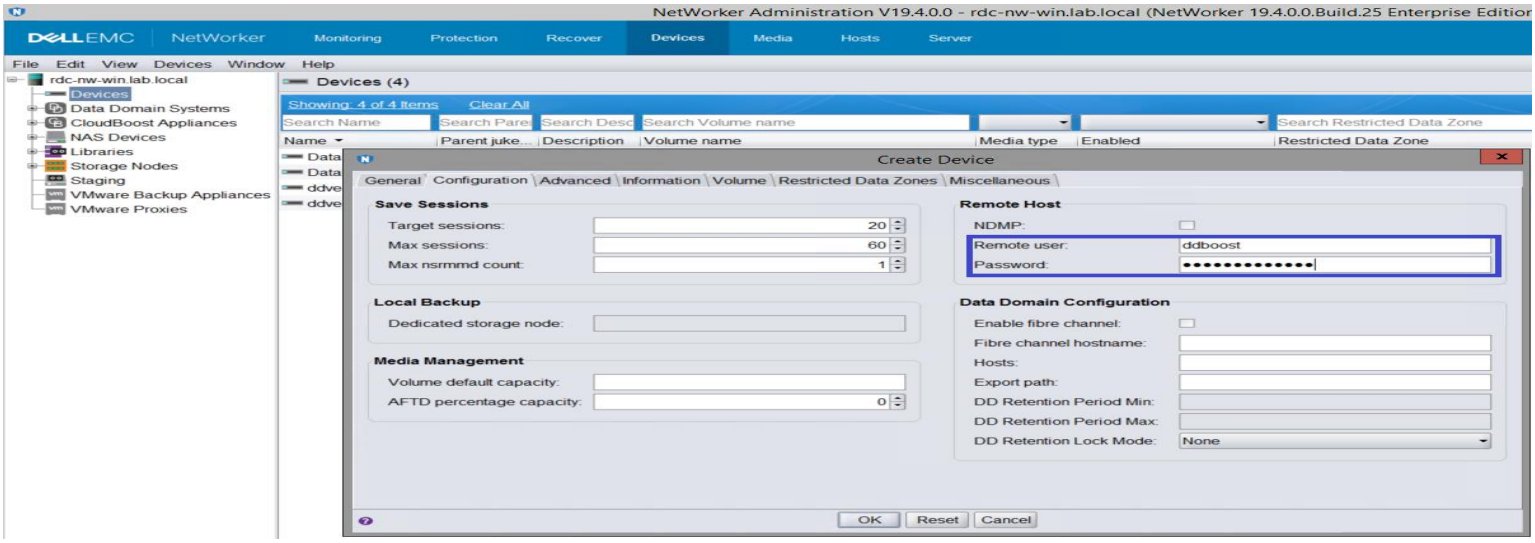


8. Under device access information, mention the path storage unit in this format DD_NAME:/replicated_mtree_name/folder_name_from_step_6.

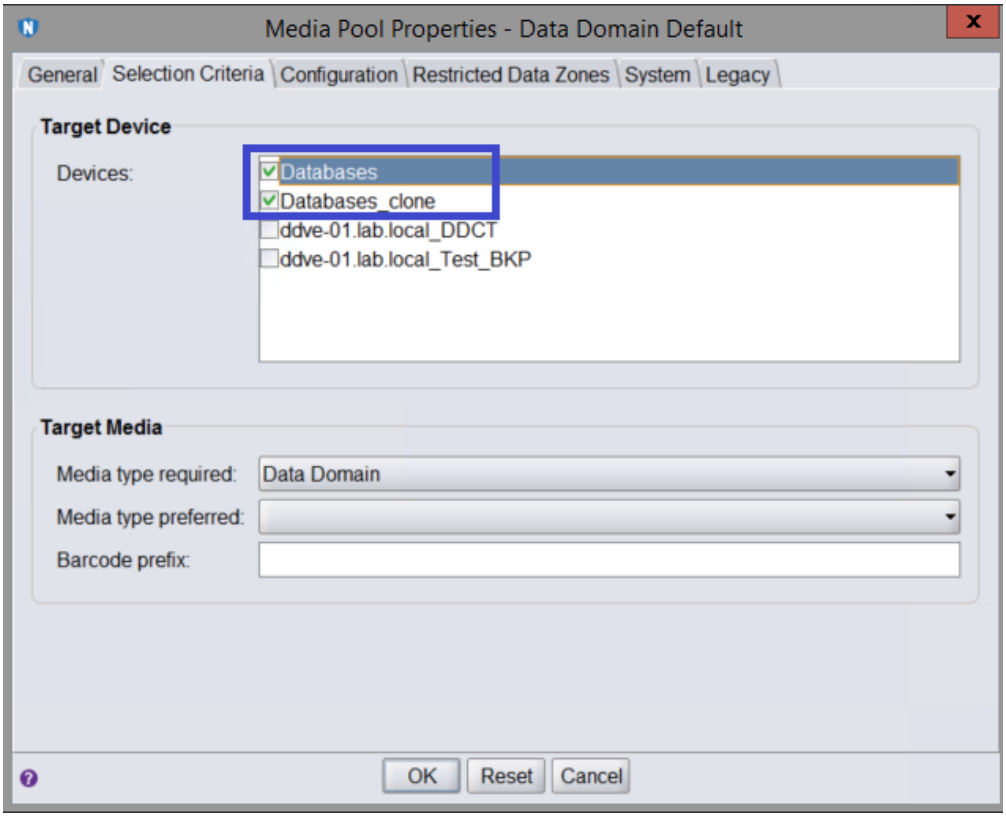
    a. Select ‘Read-only’ under status.



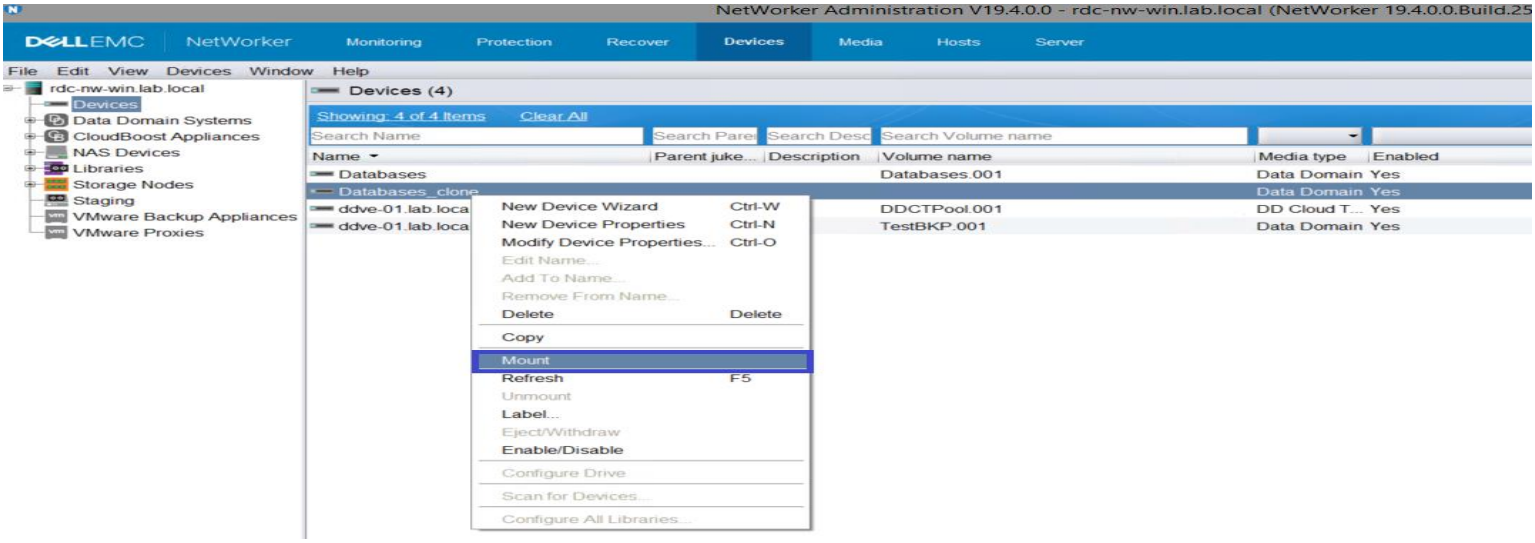
9. Go to configuration tab and key-in Remote Username and password (ddboost user as configured in pre-requisites)



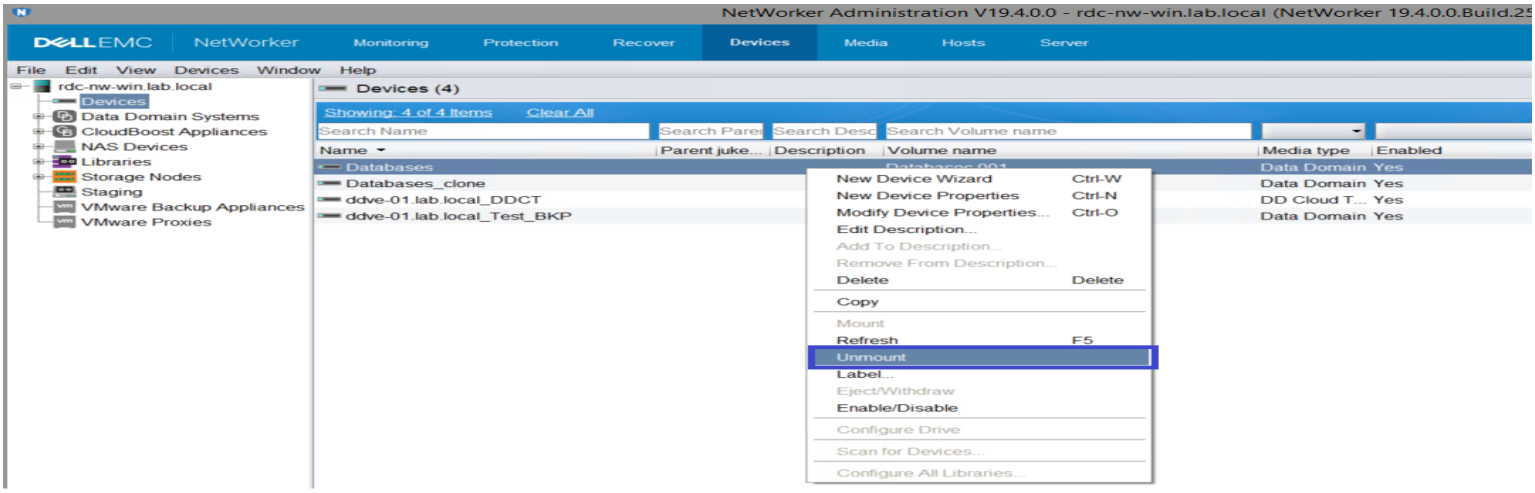
10. Click OK to create device.
11. Go to Media->Media Pools. Right click and open the Properties of Media Pool (which belongs to source device) and go to Selection Criteria.
    a. Select new device in the pool.



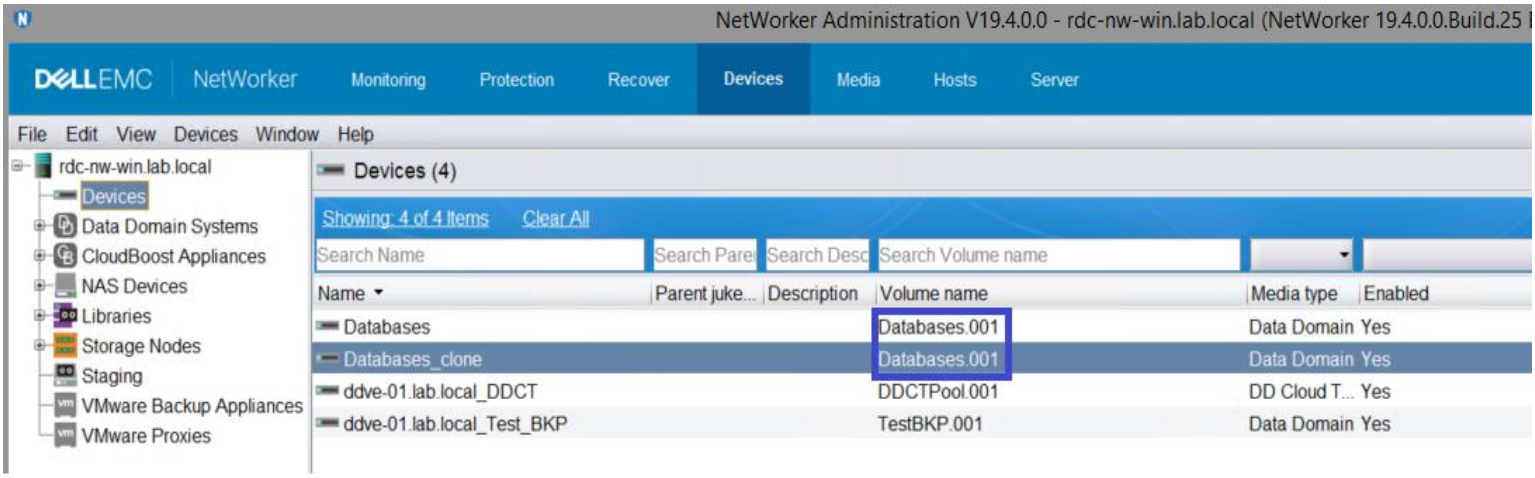
12. Mount the device.



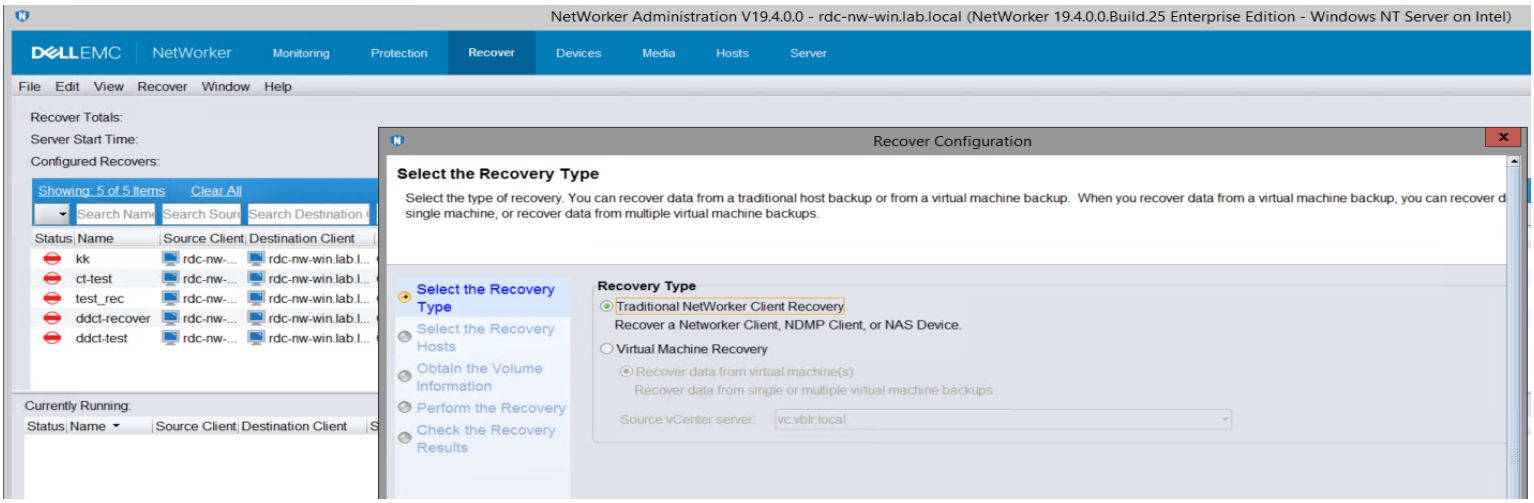
13. This device will only be used for restoring in case primary Data Domain goes down.
14. Before starting restore operation, go to Devices tab.
    a. Unmount source backup device.



15. Make sure it has the same volume name as it was on source.



16. Start restore.



Note:

Once the restore operation is completed successfully, the backup device needs to be remounted.

Advantages

- Less load on NetWorker server as clone jobs are not configured on NetWorker server
- Faster backups as there are more resources available on backup server with no load of clone.
- Light NetWorker server requirement and good competitive messaging to compete better with the likes of Commvault who will need servers at production as well as DR (Primary servers, Media servers)
- Data Domain-managed mtree level replication which should be faster than clone jobs , still giving the flexibility to end customer to maintain desired RTO/RPO and backup application has full visibility to the primary and secondary copy.
- Ability to replicate off-site via one port being opened in the firewall, and DR readiness assuming the mtree replicated system is in a different subnet.

Assumptions

- Same retention of data is needed on Primary and Second copy.
- Granular control will be lost over what gets replicated, and retention management can only be the same as what the backup are. You also can't be as selective about what gets replicated off site with mtree replication.
- If there are WAN/LAN bandwidth issues, managed file would allow for cloning select data, or specific points in time, i.e. weekly full backups. mtree is all or nothing at the mtree level. The only way to control it would be to create more mtree's on the source and splitting NetWorker backups between the mtrees as desired.

Summary

This is for customers facing issues with CCR and there is lag in clone jobs due to network issues. If retention requirements are the same at both Production and Disaster Recovery environment, this approach drastically reduces backup windows, resulting in better SLA's for backups.

References

https://dl.dell.com/content/docu101068_NetWorker_19.4_Administration_Guide.pdf?language=en_US