

NetApp Thin Provisioning at TRP

**Synopsis:** This document provides the complete details and the standards about Thin Provisioning method used in TREx-Professional setup.

**Segment:** Infrastructure as a Service – Storage

**Authors:** Mike Arndt / Santhana Ramasamy

**Contributors:** Mike Arndt /Santhana Ramasamy/Nataraja

**Document Version:** V1.0

**Date:** 01-02-2012

Release date: 01-17-2012

**CONFIDENTIAL INFORMATION**

This document contains information proprietary to Thomson Reuters and may not be reproduced, disclosed or used in whole or part without express permission of Thomson Reuters.

© Thomson Reuters 2012

Contents

[1. Terminology and Definitions 2](#_Toc314579540)

[2. Why thin provisioning? 3](#_Toc314579541)

[3. Thin Provisioning Rules 3](#_Toc314579542)

[4. Enabling Thin Provisioning 4](#_Toc314579543)

[5. Handling Threshold Breaches 6](#_Toc314579544)

[Aggregate Nearly Full 6](#_Toc314579545)

[Aggregate Full 6](#_Toc314579546)

[Aggregate Overcommitted 6](#_Toc314579547)

[6. Threshold Breach Reaction - Examples 7](#_Toc314579548)

# Terminology and Definitions

The following terminology is used throughout this document, and must be well understood by the reader:

|  |  |
| --- | --- |
| **Terms** | **Definitions** |
| filer or controller | A 7-mode NetApp storage system is typically a HA-Pair, comprised of 2 “filers” or “controllers” that each serve data for drives assigned to them. |
| aggregate | An aggregate is a pool of physical disk drives, organized into raid groups, on which FlexVol (or volumes) are provisioned. Aggregates can only be grown in size, never reduced in size. |
| FlexVol or volume | A FlexVol or volume is a storage container that resides on a single aggregate. For NAS protocols, the volume (or a subset of it) is accessed directly by the CIFS or NFS client. |
| thick volume | A ”thick volume” has a guarantee of type “volume”, and it’s size counts against space utilization in the underlying aggregate as soon as it is provisioned, regardless of how much (if any) space is actually in use. |
| thin volume | A “thin volume” typically has a guarantee of type “none” (a guarantee of type “file” can also be used, but should not be for the TRP environment). A thin volume only consumes space in the underlying aggregate when space is actually used within the volume by an end user. |
| LUN | A LUN is a block device used by FCP or iSCSI clients, and it resides within a FlexVol or volume. A LUN can have it’s reservation enabled within the volume, but the volume guarantee will still determine whether or not storage for the LUN is pre-allocated within the aggregate (with thick volumes) or used on demand when the end user places data within the LUN (with thin volumes). |
| volume guarantee | The volume guarantee assigned to a given volume is a configuration setting that determines whether space is pre-allocated for the volume within the aggregate (with a volume guarantee of type “volume”) or used on demand when the end user places data within the volume (with a volume guarantee of type “none”). |
| DFM | DFM is a NetApp management software suite used by TRP. Each TRP datacenter has multiple DFM servers – one for CPS, and one or more for CIS (some have CIS production and CIS client on different DFM servers). |
| threshold | In the context of this document, a threshold is a configuration setting on a DFM server that determines when events are created and alarms triggered for various space related issues. |
| aggrNearlyFullThreshold | This is a DFM threshold used for thin provisioned aggregates at TRP, and is measured based on the percentage of space in used in a given aggregate (due to either thick or thin volumes). On thick provisioned aggregates, this is typically set to 100%, since space is managed at the volume level. On thin provisioned aggregates, we typically set this at 65% in order to provide an alert that a given aggregate is filling up and should be investigated in the near future. |
| aggrFullThreshold | This DFM threshold provides another level of alerting, similar to the aggrNearlyFullThreshold. The aggrFullThreshold is typically set at 75% in the TRP environment, and is indicative that immediate action should be taken on an aggregate which triggers this type of alarm. |
| aggrOvercommittedThreshold | This DFM threshold is not related to how much space is in use within a given aggregate, but rather how much space has been allocated to thick or thin provisioned FlexVols or volumes. This can be used to track when too much storage has been allocated on a given aggregate, based on the rules defined, even if the aggregate utilization is still at acceptable levels. |

# Why thin provisioning?

With thin provisioning, when you create volumes for different purposes in a given aggregate, you do not actually allocate any space for those volumes in advance. The space is allocated only when the application host needs it.

The unused aggregate space is available for the thinly provisioned volumes to expand or for creating new volumes. By allowing as-needed provisioning and space reclamation, thin provisioning can improve storage utilization and decrease storage costs.

A FlexVol volume shares its containing aggregate with other FlexVol volumes. Therefore, a single aggregate is the shared source of all the storage used by the FlexVol volumes it contains. A FlexVol volume resides on a pool of storage (the underlying aggregate), and can be sized based on how much data you want to store in it, rather than based on the size of the disks in the aggregate. This flexibility enables you to maximize the performance and capacity utilization of the storage systems. Because FlexVol volumes can access all available physical storage in the aggregates on which they reside, dramatic improvements in storage utilization are possible.

The FlexVol technology enables you to oversubscribe the free space to adapt rapidly to the changing business needs.

The benefits of using thin provisioning are as follows:

* Allows storage to be provisioned just like traditional storage, but it is not consumed until data is written.
* Storage-provisioning time is greatly reduced, because you can create the storage for an application quickly without depending on the actual physical space available.
* Through notifications and configurable threshold values, you can plan your procurement strategies well in advance and have enough storage for thin provisioned volumes to grow.

# Thin Provisioning Rules

The following rules should be followed in order to safely implement thin provisioning at TRP:

* The target ratio for over-provisioning on the aggregates is 1.2:1. This ratio may change over time. For example, using the current standard, up to 12TB of FlexVols may be allocated on an aggregate that is 10TB in size (assuming all other rules are also being followed).
* Utilization is below 65% on the aggregate, and known growth will not push utilization above 75%.
* Filers that are in the process of being tech refreshed should not have additional allocations placed on them via the thin provisioning process.
* The filer performance should be below 70% utilization based on the TRP load.
* All volumes on a given aggregate should be provisioned in the same manner – we should not mix thin provisioned and thick provisioned volumes on the same aggregate, as this makes space accounting and troubleshooting much more difficult.
* Spare disk shelves should be available in on-site at Thomson Reuters in order to grow aggregates by adding disk drives, on short notice if required.

# Enabling Thin Provisioning

The procedure for setting up thin provisioning is as follows:

1. List the volumes in the aggregate you want to thin provision using:

**> aggr status –v <aggrname>**

1. Once you have the list of volumes contained in the aggregate you want to thin provision, execute the following command against all the contained volumes in the aggregate to set the volume options as per the thin provisioning best practice.

**> vol options <volname> guarantee none**

1. Rename the aggregate name to have a “\_thin” notation:

> **aggr rename <aggrname> <aggrname>\_thin**

1. Determine the aggregate object ID to set the threshold from the respective DFM server. For example, the aggregate object ID is the first column (bolded) in the output below:

# **dfm aggr list|grep eg-naslowpe-h02**

**808** eg-naslowpe-h02:aggr1\_Prod Aggregate 32\_bit No

**4081** eg-naslowpe-h02:aggr2\_Prod Aggregate 32\_bit No

**4526** eg-naslowpe-h02:aggr3\_Prod Aggregate 32\_bit No

**4528** eg-naslowpe-h02:aggr4\_Prod Aggregate 32\_bit No

**20927** eg-naslowpe-h02:aggr5\_prod Aggregate 32\_bit No

281 eg-naslowpe-h02:vol0 Traditional 32\_bit No

1. Set the aggregate nearlyfull, full and overcommitted threshold values from the respective DFM server. Note that the aggrOvercommittedThreshold will change as the TRP Thin Provisioning Rules are modified to a more aggressive ratio than the current value of 1.2:1. For example:

# **dfm aggr set <aggr object id> aggrNearlyFullThreshold=65**

# **dfm aggr set <aggr object id> aggrFullThreshold=75**

# **dfm aggr set <aggr object id> aggrOvercommittedThreshold=120**

1. Add the aggregate (aggregate id that is obtained in the above step) as a member to the “Thin\_Aggregates” group on the respective DFM server. For example:

**# dfm group add –u netapp –p netapp Thin\_Aggregates <aggr id>**

1. After renaming the aggregate, use the “**dfm aggr list|grep <filername>**” command to see if the renames are being reflected. If this is not happening, even after waiting for 30 min or so, you will have to reboot the DFM server.
2. On the DFM server for the respective filer, execute the script show in the example below that collects all the thin aggregate utilization to see if dfm database is reflecting newly added aggregates.

**# /filers/admin/bin/aggr\_vol\_util\_thin.pl**

1. Ensure that all the volumes in the aggregates are set for “none” space guarantee using the command in the example below (as shown in **red**) on the filer in question:

**> aggr show\_space –g <aggrname>**

**eg-nascorp-h02> aggr show\_space -h aggr3\_thin\_techref**

**Aggregate 'aggr3\_thin\_techref'**

**Total space WAFL reserve Snap reserve Usable space BSR NVLOG A-SIS**

**3984GB 398GB 0KB 3585GB 0KB 0KB**

**Space allocated to volumes in the aggregate**

**Volume Allocated Used Guarantee**

**prod\_corp\_h0048\_root 6856KB 6440KB none**

**infra\_emgc1p\_n01ora1\_nosnap 73GB 72GB none**

**infra\_emgc1p\_s01ora1\_snap 674GB 669GB none**

**prod\_corp\_h0054\_root 50MB 2332KB volume**

**infra\_wi\_creativesvcsprod\_usr\_snap 2512MB 1109MB none**

**infra\_wi\_creativesvcsprod\_tsys\_nosnap 836MB 698MB none**

**infra\_wi\_creativesvcsprod\_info\_snap 755MB 616MB none**

**Aggregate Allocated Used Avail**

**Total space 751GB 744GB 2833GB**

**Snap reserve 0KB 0KB 0KB**

**WAFL reserve 398GB 40GB 357GB**

# Handling Threshold Breaches

This section will define how to react to the 3 different types of threshold breaches that may occur.

## Aggregate Nearly Full

The alarm for this threshold breach indicates that the utilization within an aggregate is higher than we would like it to be. This should be investigated with high priority, but we do not necessarily need to take an action to reduce aggregate utilization. The steps to take in this situation are:

* Open an IM for the investigation.
* Determine the volumes using DFM history to see which contained volumes are contributing more to the aggregate growth.
* Check if there is any snapshot that has caused the abnormal growth of the volumes in the aggregate, which may be responsible for the aggregate growth.
* Check the snapvault lags for the contained volumes and fix if there are any issue with snapvault update..
* Check for snapmirror lag if any for the contained volumes, update or troubleshoot if there is any issue.
* Check for recent quota growth requests that might have lead to the growth of the volume.
* Double Check for any unused/migrated volumes if they are online and accordingly create a CR to offline and destroy the same.

## Aggregate Full

The alarm for this threshold breach indicates that the utilization within an aggregate is much higher than we would like it to be. This should be investigated with high priority, and immediate action should be taken to reduce the aggregate utilization so that an out of space condition does not occur. The steps to take in this situation are:

* Take the same steps as listed in the Aggregate Nearly Full section above in an attempt to reduce the utilization of the aggregate in question. If this does not allow aggregate utilization to be reduce below 65%, continue with the next steps in this section.
* If the above steps did not resolve the issue then we need to create a CR plan for growing aggregate with the existing disks or plan for procuring disks if the aggregate is eligible for further growths, keep Storage implementation team informed.
* If there is no provision for growth of an aggregate Create a CR for migrating the volume, co-ordinate with the volume owners and plan for migrating the volumes that are having high growth rates to a less occupied thin aggregate, keep Storage implementation team informed.

## Aggregate Overcommitted

The alarm for this threshold breach indicates that the amount of space allocated on a given aggregate is higher than the TRP rules allow for. This alarm is not related to the amount of space actually in use within the aggregate, as the utilization thresholds are covered by the Aggregate Nearly Full and Aggregate Full alarms. The steps to take in this situation are:

* Rename the aggregate name from **aggrname to aggrname\_full.**
* Notify the Storage Implementation team that this aggregate was provisioned at a higher ratio than the Thin Provisioning Rules allow for.

# Threshold Breach Reaction - Examples

This section will give some examples of how to investigate a threshold breach of the type Aggregate Nearly Full or Aggregate Full.

**Step1**: You will receive three types of email alerts similar to the below one

Dfm Alert Type1: Error event on **eg-nascorp-f06:aggr1\_thin** (Aggregate Nearly Full)

=**Utilization exceeding 65%**

**Incident RTC:2 hrs**

Dfm Alert Type2: Error event on **eg-nascorp-f06:aggr1\_thin** (Aggregate Full)

**=Utilization exceeding 75%**

**Incident RTC:30 min**

Dfm Alert Type3: Error event on **eg-nascorp-f06:aggr1\_thin** (Aggregate Overcommitted)

**=Allocation exceeding 120%**

**Incident RTC:One day Minor**

**Step2**: Collect the filername/Vfilername and aggregate info from the emailed DFM alert, look for the Incident In the service manager with the collected information.

**Step3**: Check for aggregate utilization by logging into any of the DFM server using SSH with Uaccount credentials.

**DFM servers**

nidaros,nerstrand,newnan,cmp111cwq,cmp111mgc,cmp111xjq,cmp111dfm,c111ujr

Use the below two command to find the current utilization

1. DF

nidaros:rsh **eg-nascorp-f06 df –Ag aggr1\_thin**

Aggregate total used avail capacity

aggr1\_thin 13225GB 7408GB 5816GB 56%

**aggr1\_thin/.snapshot 0GB 0GB 0GB ---%**

1. Aggr

netapp@newnan:~> rsh eg-nascorp-f06 aggr show\_space -h aggr1\_thin

Aggregate 'aggr1\_thin'

Total space WAFL reserve Snap reserve Usable space BSR NVLOG A-SIS Smtape

21TB 2204GB 0KB 19TB 0KB 162MB 0KB

Space allocated to volumes in the aggregate

Volume Allocated Used Guarantee

vol0 333GB 14GB volume

eg\_nascorp\_f06\_corpvsip\_root 7220KB 6892KB none

prod\_corp\_f0161\_root 13MB 12MB none

prod\_corp\_f0158\_root 9772KB 9444KB none

infra\_c111fcu\_vms\_nosnap 19GB 144MB none

infra\_c111fcu\_vms\_snap07 44GB 20GB none

infra\_c111fcu\_vms\_snap14 24GB 1228MB none

infra\_c111fcu\_vms\_snap45 24GB 1228MB none

infra\_c111fcu\_vmw\_nosnap 27GB 8632MB none

corpshared\_snap1 4148GB 4125GB none

corpshared\_snap0 1867GB 1847GB none

corpshared\_nosnap0 1090GB 1085GB none

reserve\_techrefresh\_ct\_snap1\_10\_21\_11\_jms 46GB 1460KB none

Aggregate Allocated Used Avail

Total space 7625GB 7104GB 11TB

Snap reserve 0KB 0KB 0KB

WAFL reserve 2204GB 115GB 2088GB

**Step3a**: Check for the volumes whose space guarantee is set to volume from the above output.

**Step3b**: If any volume found with space guarantee of the volume is set to “volume” change it to none and see if the aggregate utilization came down, if yes, go to end otherwise continue with

**Step4**: Check for aggregate snapshot that is contributing to the aggregate growth using the below command

If yes, delete the snapshots using the following commands

# nidaros: rsh eg-nascorp-f06 snap list -A aggr1\_thin

Aggregate aggr1\_thin

working...

No snapshots exist.

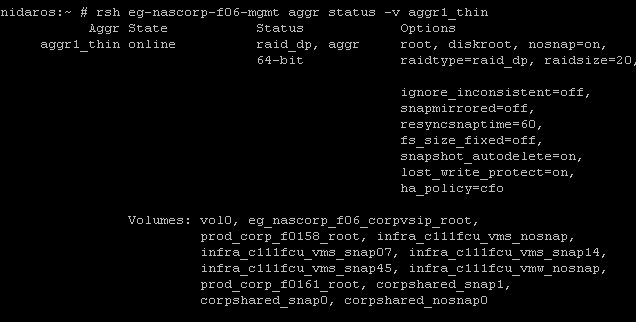
nidaros:rsh eg-nascorp-f06 “privset advanced:snap delete -A -a **aggr1\_thin:priv set”**

**Step5: Check the aggregate utilization again and wait for 5-10 min to see if the utilization gradually comedown below threshold mark, if yes go to end.**

rsh eg**-nascorp-f06 df –Ag aggr1\_thin**

**Step6**: Check for any unused/migrated volumes by logging into any of the DFM servers using U account credentials. .

With the available filername and aggregate info in the alert, identify the volumes that are already migrated , if yes then co-ordinate with storage implementation team and Schedule a CR to offline and destroy the migrated volumes.



rsh eg-nascorp-f06-mgmt vol offline migrated\_volume

rsh eg-nascorp-f06-mgmt vol offline migrated\_volume

Check the aggregate utilization, if the utilization is below threshold if yes go to end.

**Note**: Usually a migrated volume starts with a prefix “Migrated” followed by underscore and a volume name)

**Step7**: Determine the corresponding DFM server info for the filer using thin provisioned aggregate daily usage report.

|  |  |  |
| --- | --- | --- |
| **DFM Server** | **Filer** | **Aggregate Name** |
| cmp111cwq | eg-nascorp-f06 | aggr1\_thin |

**Step8**: Login to corresponding DFM server determined in step7 using your u account credentials

E building

<http://nerstrand.int.westgroup.net:8080>

<http://cmp111mgc.int.westgroup.com:8080>

H building

<http://nidaros.int.westgroup.net:8080>

<http://cmp111xjq.int.westgroup.com:8080>

F building

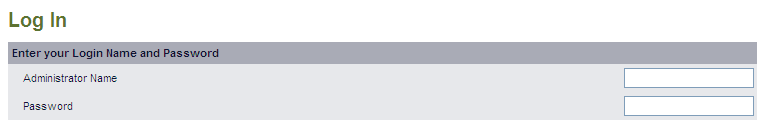
<http://newnan.int.westgroup.net:8080>

<http://cmp111cwq.int.westgroup.com:8080>

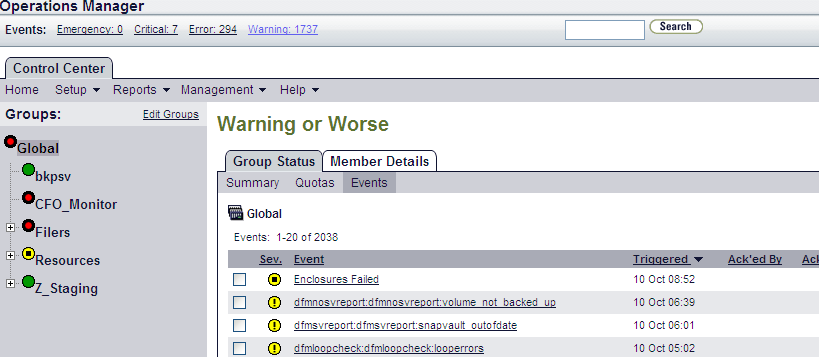
XO building

<http://cmp111dfm.int.westgroup.com:8080>

<http://c111ujr.int.westgroup.com:8080>



**Step9**: Navigate to the errors and warning sections in the top left corner to see any Volume abnormal growths, if there are no alerts for abnormal growth then go to step18.

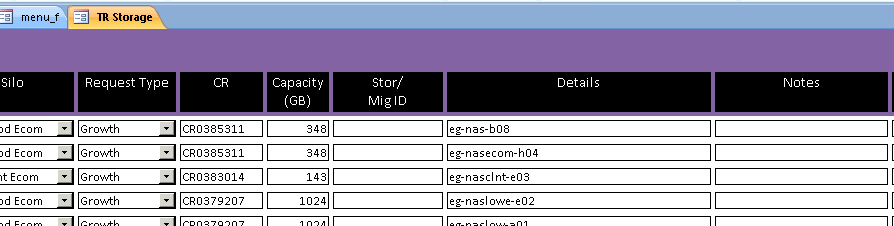


**Step10**: Determine if the volume in question belongs to the thin aggregate using the below command, if the answer is no go to step18.

rsh filername Vol container <volumename>

**Step11**: Check for recent large Growth/New allocation requests that might have lead to the growth of an aggregate using the below link. If the answer is no then go to step18.

[\\prod-corp-f0161\sg$\Ongoing Project Documentation\PipelineTracking\StorageImplementation.mdb](file:///\\prod-corp-f0161\sg$\Ongoing%20Project%20Documentation\PipelineTracking\StorageImplementation.mdb)



**Step12**: Check the snapshot utilization using below command

nidaros:~ # rsh eg-nascorp-f06 df -g corpshared\_snap1

Filesystem total used avail capacity Mounted on

/vol/corpshared\_snap1/ 4340GB 4015GB 324GB 93% /vol/corpshared\_snap1/

**/vol/corpshared\_snap1/.snapshot 1085GB 0GB 1084GB 0%**

**Step13**:If the snapshot utilization on the volume in question contained in the aggregate is exceeding 100% then check for snap autodelete if the status is set to off, set the value using the below commands.

NOTE: If the snap utilization is less than 100% go to step18.

**nidaros:~ # rsh eg-nascorp-f06 snap autodelete corpshared\_snap1**

**snapshot autodelete settings for corpshared\_snap1:**

**state : on**

**commitment : try**

**trigger : snap\_reserve**

**target\_free\_space : 20%**

**delete\_order : oldest\_first**

**defer\_delete : user\_created**

**prefix : (not specified)**

**destroy\_list : none**

**rsh eg-nascorp-f06 snap autodelete corpshared\_snap1 on**

**Step14: Update snapvault manually if the snap utilization is exceeding 100% using the below link, if the snap utilzation is less than 100% then go to step17.**

[**\\prod-corp-f0161\sg$\Procedures\NetApp\snapvault\_manualupdate.doc**](file:///\\prod-corp-f0161\sg$\Procedures\NetApp\snapvault_manualupdate.doc)

**Step15**: check for snapvault duplicate relationships using the below commands: based on the lag time ( >1000 hrs) release the duplicate relationships using the below commands.

Snapvault status

nidaros:~ # rsh eg-nascorp-f06 vfiler run prod-corp-f0158 snapvault status|grep infra\_c111fcu\_vms\_snap14

prod-corp-f0158:/vol/infra\_c111fcu\_vms\_snap14 corpe1:/vol/sv\_14\_infra\_c111fcu\_vms\_snap14/1 Source **09:22:53** Idle

On a snapvault primary:

snapvault release <primary\_path> <secondary\_filer>:<secondary\_path>

**Step16**: Check for any snapmirror relationship exists for the volume in question with a lag time > 1000 hrs, Update or break the relationship after consulting with team/members responsible for tech refresh.

Snapmirror lag

rsh eg-nascorp-f06 vfiler run prod-corp-f0158 snapmirror status|grep corpshared\_snap1

Snapmirror update

rsh eg-nascorp-f06 vfiler run prod-corp-f0158 snapmirror update corpshared\_snap1

Snapmirror break

rsh eg-nascorp-f06 vfiler run prod-corp-f0158 snapmirror quiesce corpshared\_snap1

rsh eg-nascorp-f06 vfiler run prod-corp-f0158 snapmirror break corpshared\_snap1

**Step17**: If step1 to 16 did not help in getting the aggregate utilization below FULL threshold and if the Alert is of type2,Check for available spare disks on the filer in question using the below command, if the spares are greater than 12 disks create a CR plan for growing aggregate with the existing disks execute the following commands.

nidaros:~ # rsh eg-nascorp-f06 vol status -s

Spare disks

RAID Disk **Device** HA SHELF BAY CHAN Pool Type RPM Used (MB/blks) Phys (MB/blks)

--------- ------ ------------- ---- ---- ---- ----- -------------- --------------

Spare disks for block or zoned checksum traditional volumes or aggregates

spare 5a.10.10 5a 10 10 SA:B - SAS 15000 418000/856064000 **420**156/860480768

spare 5a.10.11 5a 10 11 SA:B - SAS 15000 418000/856064000 420156/860480768

spare 5a.10.12 5a 10 12 SA:B - SAS 15000 418000/856064000 420156/860480768

spare 5a.10.13 5a 10 13 SA:B - SAS 15000 418000/856064000 420156/860480768

spare 5a.10.14 5a 10 14 SA:B - SAS 15000 418000/856064000 420156/860480768

Growing an aggregate with the size of disks available in GB

rsh eg-nascorp-f06 aggr add aggr1\_thin 420@8

Check the aggregate utilization

rsh eg-nascorp-f06 df –Ag aggr1\_thin

**Note**: Grow an aggregate in terms of Partial raid= 8 disks or full raid 16 disks depending on the spares available and the disks required for bringing down the aggregate utilization.

**Step18**: If step1 to 17 did not help in getting the aggregate utilization below FULL and if the alert is of Type2 create a CR Plan for migrating one or more large volumes to another thin aggregate,

* Identify the filer and aggregate which is less loaded and less utilized. **Refer to the thin aggregate report.**
* Create a migration plan using the template in the below link

[\\prod-corp-f0161\sg$\Procedures\Migration Templates\NAS Migration Template1.docx](file:///\\prod-corp-f0161\sg$\Procedures\Migration%20Templates\NAS%20Migration%20Template1.docx)

**\\prod-corp-f0161\sg$\Procedures\Migration Templates has other templates.**

**Step19**: If step1 to 17 did not help in getting the aggregate utilization below FULL and the Alert is of Type2, Create a CR plan for procuring disks, if the aggregate is eligible for further growths, keep Storage implementation team informed.

**Procedure need to be given by storage implementation team**

**Thin provisioning Flow chart on aggregates**

<https://theshare.thomsonreuters.com/sites/Storage/SI/Shared%20Documents/Forms/AllItems.aspx?RootFolder=%2Fsites%2FStorage%2FSI%2FShared%20Documents%2FThin%20Provisioning%2FNew%20and%20Existing%20Requests%20%28Flow%20Charts%29&InitialTabId=Ribbon%2EDocument&VisibilityContext=WSSTabPersistence>