Managing aggregates

ONTAP 9

NetApp
March 09, 2022

Table of Contents

Managing aggregates

RAID protection levels for disks

ONTAP supports three levels of RAID protection for aggregates. Your level of RAID protection determines the number of parity disks available for data recovery in the event of disk failures.

With RAID protection, if there is a data disk failure in a RAID group, ONTAP can replace the failed disk with a spare disk and use parity data to reconstruct the data of the failed disk.

- RAID4

  With RAID4 protection, ONTAP can use one spare disk to replace and reconstruct the data from one failed disk within the RAID group.

- RAID-DP

  With RAID-DP protection, ONTAP can use up to two spare disks to replace and reconstruct the data from up to two simultaneously failed disks within the RAID group.

- RAID-TEC

  With RAID-TEC protection, ONTAP can use up to three spare disks to replace and reconstruct the data from up to three simultaneously failed disks within the RAID group.

Related information

NetApp Technical Report 3437: Storage Subsystem Resiliency Guide

Determine the number of disks or disk partitions required for an aggregate

You must have enough disks or disk partitions in your aggregate to meet system and business requirements. You should also have the recommended number of hot spare disks or hot spare disk partitions to minimize the potential of data loss.

Root-data partitioning is enabled by default on certain configurations. Systems with root-data partitioning enabled use disk partitions to create aggregates. Systems that do not have root-data partitioning enabled use unpartitioned disks.

You must have enough disks or disk partitions to meet the minimum number required for your RAID policy and enough to meet your minimum capacity requirements.

> In ONTAP, the usable space of the drive is less than the physical capacity of the drive. You can find the usable space of a specific drive and the minimum number of disks or disk partitions required for each RAID policy in *Hardware Universe*. You can also use the `storage aggregate show-spare-disks` command to find the usable space of a specific disk.

In addition to the number of disks or disk partitions necessary to create your RAID group and meet your capacity requirements, you should also have the minimum number of hot spare disks or hot spare disk

partitions recommended for your aggregate:

- For all flash aggregates, you should have a minimum of one hot spare disk or disk partition.

  The AFF C190 defaults to no spare drive. This exception is fully supported.

- For non-flash homogenous aggregates, you should have a minimum of two hot spare disks or disk partitions.
- For SSD storage pools, you should have a minimum of one hot spare disk for each HA pair.
- For Flash Pool aggregates, you should have a minimum of two spare disks for each HA pair. You can find more information on the supported RAID policies for Flash Pool aggregates in the Hardware Universe.
- To support the use of the Maintenance Center and to avoid issues caused by multiple concurrent disk failures, you should have a minimum of four hot spares in multi-disk carriers.

Related information

NetApp Hardware Universe

NetApp Technical Report 3838: Storage Subsystem Configuration Guide

Correct misaligned spare partitions

When you add partitioned disks to an aggregate, you must leave a disk with both the root and data partition available as spare for every node. If you do not and your node experiences a disruption, ONTAP cannot dump the core to the spare data partition.

What you'll need

You must have both a spare data partition and a spare root partition on the same type of disk owned by the same node.

Steps

1. Display the spare partitions for the node:

   `storage aggregate show-spare-disks -original-owner` *`node_name`*

   Note which disk has a spare data partition (spare_data) and which disk has a spare root partition (spare_root). The spare partition will show a non-zero value under the `Local Data Usable` or `Local Root Usable` column.

2. Replace the disk with a spare data partition with the disk with the spare root partition:

   `storage disk replace -disk` *`spare_data`* `-replacement` *`spare_root`* `-action start`

   You can copy the data in either direction; however, copying the root partition takes less time to complete.

3. Monitor the progress of the disk replacement:

   `storage aggregate show-status -aggregate` *`aggr_name`*

4. After the replacement operation is complete, display the spares again to confirm that you have a full spare disk:

`storage aggregate show-spare-disks -original-owner node_name`

You should see a spare disk with usable space under both `Local Data Usable` and `Local Root Usable`.

Example

You display your spare partitions for node c1-01 and see that your spare partitions are not aligned:

```
c1::> storage aggregate show-spare-disks -original-owner c1-01

Original Owner: c1-01
 Pool0
  Shared HDD Spares
                                                                 Local    Local
                                                                  Data     Root
Physical
 Disk                       Type     RPM Checksum         Usable   Usable
Size
 -------------------------- ----- ------ -------------- -------- --------
--------
 1.0.1                      BSAS    7200 block           753.8GB       0B
828.0GB
 1.0.10                     BSAS    7200 block                0B  73.89GB
828.0GB
```

You start the disk replacement job:

```
c1::> storage disk replace -disk 1.0.1 -replacement 1.0.10 -action start
```

While you are waiting for the replacement operation to finish, you display the progress of the operation:

```
c1::> storage aggregate show-status -aggregate aggr0_1

Owner Node: c1-01
 Aggregate: aggr0_1 (online, raid_dp) (block checksums)
  Plex: /aggr0_1/plex0 (online, normal, active, pool0)
   RAID Group /aggr0_1/plex0/rg0 (normal, block checksums)
                                                                    Usable
Physical
     Position Disk                      Pool Type     RPM     Size
Size Status
     -------- ------------------------- ---- ----- ------ --------
-------- ----------
     shared   1.0.1                      0   BSAS    7200  73.89GB
828.0GB (replacing, copy in progress)
     shared   1.0.10                     0   BSAS    7200  73.89GB
828.0GB (copy 63% completed)
     shared   1.0.0                      0   BSAS    7200  73.89GB
828.0GB (normal)
     shared   1.0.11                     0   BSAS    7200  73.89GB
828.0GB (normal)
     shared   1.0.6                      0   BSAS    7200  73.89GB
828.0GB (normal)
     shared   1.0.5                      0   BSAS    7200  73.89GB
828.0GB (normal)
```

After the replacement operation is complete, you confirm that you have a full spare disk:

```
ie2220::> storage aggregate show-spare-disks -original-owner c1-01

Original Owner: c1-01
 Pool0
  Shared HDD Spares
                                                            Local    Local
                                                             Data     Root
Physical
 Disk                      Type     RPM Checksum         Usable   Usable
Size
 ------------------------- ----- ------ -------------- -------- --------
--------
 1.0.1                     BSAS    7200 block           753.8GB  73.89GB
828.0GB
```

Determine drive and RAID group information for an aggregate

Some aggregate administration tasks require that you know what types of drives compose the aggregate, their size, checksum, and status, whether they are shared with other aggregates, and the size and composition of the RAID groups.

Step

1. Show the drives for the aggregate, by RAID group:

   `storage aggregate show-status` *`aggr_name`*

   The drives are displayed for each RAID group in the aggregate.

   You can see the RAID type of the drive (data, parity, dparity) in the `Position` column. If the `Position` column displays `shared`, then the drive is shared: if it is an HDD, it is a partitioned disk; if it is an SSD, it is part of a storage pool.

Example: A Flash Pool aggregate using an SSD storage pool and data partitions

```
cluster1::> storage aggregate show-status nodeA_fp_1

Owner Node: cluster1-a
 Aggregate: nodeA_fp_1 (online, mixed_raid_type, hybrid) (block checksums)
  Plex: /nodeA_fp_1/plex0 (online, normal, active, pool0)
   RAID Group /nodeA_fp_1/plex0/rg0 (normal, block checksums, raid_dp)
                                                                   Usable
Physical
     Position Disk                       Pool Type     RPM     Size
Size Status
     -------- -------------------------- ---- ----- ------ --------
-------- -------
     shared   2.0.1                       0   SAS    10000  472.9GB
547.1GB (normal)
     shared   2.0.3                       0   SAS    10000  472.9GB
547.1GB (normal)
     shared   2.0.5                       0   SAS    10000  472.9GB
547.1GB (normal)
     shared   2.0.7                       0   SAS    10000  472.9GB
547.1GB (normal)
     shared   2.0.9                       0   SAS    10000  472.9GB
547.1GB (normal)
     shared   2.0.11                      0   SAS    10000  472.9GB
547.1GB (normal)

   RAID Group /nodeA_flashpool_1/plex0/rg1 (normal, block checksums,
raid4) (Storage Pool: SmallSP)
                                                                   Usable
Physical
     Position Disk                       Pool Type     RPM     Size
Size Status
     -------- -------------------------- ---- ----- ------ --------
-------- -------
     shared   2.0.13                      0   SSD        -  186.2GB
745.2GB (normal)
     shared   2.0.12                      0   SSD        -  186.2GB
745.2GB (normal)
8 entries were displayed.
```

Relocate aggregate ownership within an HA pair

Relocate aggregate ownership

You can change the ownership of aggregates among the nodes in an HA pair without interrupting service from the aggregates.

Both nodes in an HA pair are physically connected to each other's disks or array LUNs. Each disk or array LUN is owned by one of the nodes. Although ownership of disks temporarily changes when a takeover occurs, the aggregate relocation operations either permanently (for example, if done for load balancing) or temporarily (for example, if done as part of takeover) change the ownership of all disks or array LUNs within an aggregate from one node to the other. The ownership changes without any data-copy processes or physical movement of the disks or array LUNs.

About this task

- Because volume count limits are validated programmatically during aggregate relocation operations, it is not necessary to check for this manually.

  If the volume count exceeds the supported limit, the aggregate relocation operation fails with a relevant error message.

- You should not initiate aggregate relocation when system-level operations are in progress on either the source or the destination node; likewise, you should not start these operations during the aggregate relocation.

  These operations can include the following:

  - Takeover
  - Giveback
  - Shutdown
  - Another aggregate relocation operation
  - Disk ownership changes
  - Aggregate or volume configuration operations
  - Storage controller replacement
  - ONTAP upgrade
  - ONTAP revert
- If you have a MetroCluster configuration, you should not initiate aggregate relocation while disaster recovery operations (*switchover*, *healing*, or *switchback*) are in progress.
- If you have a MetroCluster configuration and initiate aggregate relocation on a switched-over aggregate, the operation might fail because it exceeds the DR partner's volume limit count.
- You should not initiate aggregate relocation on aggregates that are corrupt or undergoing maintenance.
- Before initiating the aggregate relocation, you should save any core dumps on the source and destination nodes.

Steps

1. View the aggregates on the node to confirm which aggregates to move and ensure they are online and in good condition:

   `storage aggregate show -node` *`source-node`*

The following command shows six aggregates on the four nodes in the cluster. All aggregates are online. Node1 and Node3 form an HA pair and Node2 and Node4 form an HA pair.

```
cluster::> storage aggregate show
Aggregate     Size Available Used% State   #Vols  Nodes  RAID Status
--------- -------- --------- ----- ------- ------ ------ -----------
aggr_0     239.0GB   11.13GB   95% online       1 node1  raid_dp,
                                                        normal
aggr_1     239.0GB   11.13GB   95% online       1 node1  raid_dp,
                                                        normal
aggr_2     239.0GB   11.13GB   95% online       1 node2  raid_dp,
                                                        normal
aggr_3     239.0GB   11.13GB   95% online       1 node2  raid_dp,
                                                        normal
aggr_4     239.0GB   238.9GB    0% online       5 node3  raid_dp,
                                                        normal
aggr_5     239.0GB   239.0GB    0% online       4 node4  raid_dp,
                                                        normal
6 entries were displayed.
```

2. Issue the command to start the aggregate relocation:

   `storage aggregate relocation start -aggregate-list` *`aggregate-1`*`,` *`aggregate-2`*`…`
   `-node source-node -destination` *`destination-node`*

   The following command moves the aggregates aggr_1 and aggr_2 from Node1 to Node3. Node3 is Node1's HA partner. The aggregates can be moved only within the HA pair.

```
cluster::> storage aggregate relocation start -aggregate-list aggr_1,
aggr_2 -node node1 -destination node3
Run the storage aggregate relocation show command to check relocation
status.
node1::storage aggregate>
```

3. Monitor the progress of the aggregate relocation with the `storage aggregate relocation show` command:

   `storage aggregate relocation show -node` *`source-node`*

   The following command shows the progress of the aggregates that are being moved to Node3:

```
cluster::> storage aggregate relocation show -node node1
Source Aggregate   Destination    Relocation Status
------ ----------- -------------  ------------------------
node1
       aggr_1      node3          In progress, module: wafl
       aggr_2      node3          Not attempted yet
2 entries were displayed.
node1::storage aggregate>
```

When the relocation is complete, the output of this command shows each aggregate with a relocation status of `Done`.

Commands for aggregate relocation

There are specific ONTAP commands for relocating aggregate ownership within an HA pair.

If you want to…	Use this command…
Start the aggregate relocation process	`storage aggregate relocation start`
Monitor the aggregate relocation process	`storage aggregate relocation show`

Related information

ONTAP 9 commands

Assigning aggregates to SVMs

If you assign one or more aggregates to a storage virtual machine (SVM, formerly known as Vserver), then you can use only those aggregates to contain volumes for that SVM. Assigning aggregates to your SVMs is particularly important in a multi-tenancy environment.

What you'll need

The SVM and the aggregates you want to assign to that SVM must already exist.

About this task

Assigning aggregates to your SVMs helps you keep your SVMs isolated from each other; this is especially important in a multi-tenancy environment..

Steps

1. Check the list of aggregates already assigned to the SVM:

   ```
   vserver show -fields aggr-list
   ```

   The aggregates currently assigned to the SVM are displayed. If there are no aggregates assigned, "-" is

displayed.

2. Add or remove assigned aggregates, depending on your requirements:

If you want to…	Use this command…
Assign additional aggregates	`vserver add-aggregates`
Unassign aggregates	`vserver remove-aggregates`

The listed aggregates are assigned to or removed from the SVM. If the SVM already has volumes that use an aggregate that is not assigned to the SVM, a warning message is displayed, but the command is completed successfully. Any aggregates that were already assigned to the SVM and that were not named in the command are unaffected.

Example

In the following example, the aggregates aggr1 and aggr2 are assigned to SVM svm1:

```
vserver add-aggregates -vserver svm1 -aggregates aggr1,aggr2
```

Determine space usage in an aggregate

You can view space usage by all volumes in one or more aggregates with the `aggregate show-space` command. This helps you see which volumes are consuming the most space in their containing aggregates so that you can take actions to free more space.

The used space in an aggregate is directly affected by the space used in the FlexVol volumes it contains. Measures that you take to increase space in a volume also affect space in the aggregate.

The following rows are included in the `aggregate show-space` command output:

- `Volume Footprints`

  The total of all volume footprints within the aggregate. It includes all of the space that is used or reserved by all data and metadata of all volumes in the containing aggregate.

- `Aggregate Metadata`

  The total file system metadata required by the aggregate, such as allocation bitmaps and inode files.

- `Snapshot Reserve`

  The amount of space reserved for aggregate Snapshot copies, based on volume size. It is considered used space and is not available to volume or aggregate data or metadata.

- `Snapshot Reserve Unusable`

  The amount of space originally allocated for aggregate Snapshot reserve that is unavailable for aggregate Snapshot copies because it is being used by volumes associated with the aggregate. Can occur only for aggregates with a non-zero aggregate Snapshot reserve.

- `Total Used`

  The sum of all space used or reserved in the aggregate by volumes, metadata, or Snapshot copies.

- `Total Physical Used`

  The amount of space being used for data now (rather than being reserved for future use). Includes space used by aggregate Snapshot copies.

The following example shows the `aggregate show-space` command output for an aggregate whose Snapshot reserve is 5%. If the Snapshot reserve was 0, the row would not be displayed.

```
cluster1::> storage aggregate show-space

                      Aggregate : wqa_gx106_aggr1

      Feature                                 Used      Used%
      --------------------------------    ----------     ------
      Volume Footprints                      101.0MB         0%
      Aggregate Metadata                       300KB         0%
      Snapshot Reserve                        5.98GB         5%

      Total Used                              6.07GB         5%
      Total Physical Used                    34.82KB         0%
```

Determine which volumes reside on an aggregate

You might need to determine which volumes reside on an aggregate before performing operations on the aggregate, such as relocating it or taking it offline.

Steps

1. To display the volumes that reside on an aggregate, enter

   `volume show -aggregate aggregate_name`

   All volumes that reside on the specified aggregate are displayed.

How you can determine and control a volume's space usage in the aggregate

You can determine which FlexVol volumes are using the most space in the aggregate and specifically which features within the volume. The `volume show-footprint` command provides information about a volume's footprint, or its space usage within the containing aggregate.

The `volume show-footprint` command shows details about the space usage of each volume in an

aggregate, including offline volumes. This command bridges the gap between the output of the `volume show-space` and `aggregate show-space` commands. All percentages are calculated as a percent of aggregate size.

The following example shows the `volume show-footprint` command output for a volume called testvol:

```
cluster1::> volume show-footprint testvol

      Vserver : thevs
      Volume  : testvol

      Feature                                     Used    Used%
      --------------------------------    ----------    -----
      Volume Data Footprint                    120.6MB       4%
      Volume Guarantee                          1.88GB      71%
      Flexible Volume Metadata                 11.38MB       0%
      Delayed Frees                             1.36MB       0%
      Total Footprint                           2.01GB      76%
```

The following table explains some of the key rows of the output of the `volume show-footprint` command and what you can do to try to decrease space usage by that feature:

Row/feature name	Description/contents of row	Some ways to decrease
`Volume Data Footprint`	The total amount of space used in the containing aggregate by a volume's data in the active file system and the space used by the volume's Snapshot copies. This row does not include reserved space.	• Deleting data from the volume. • Deleting Snapshot copies from the volume.
`Volume Guarantee`	The amount of space reserved by the volume in the aggregate for future writes. The amount of space reserved depends on the guarantee type of the volume.	Changing the type of guarantee for the volume to `none`.
`Flexible Volume Metadata`	The total amount of space used in the aggregate by the volume's metadata files.	No direct method to control.
`Delayed Frees`	Blocks that ONTAP used for performance and cannot be immediately freed. For SnapMirror destinations, this row has a value of `0` and is not displayed.	No direct method to control.

`File Operation Metadata`	The total amount of space reserved for file operation metadata.	No direct method to control.
`Total Footprint`	The total amount of space that the volume uses in the aggregate. It is the sum of all of the rows.	Any of the methods used to decrease space used by a volume.

Related information

NetApp Technical Report 3483: Thin Provisioning in a NetApp SAN or IP SAN Enterprise Environment

Methods to create space in an aggregate

If an aggregate runs out of free space, various problems can result that range from loss of data to disabling a volume's guarantee. There are multiple ways to make more space in an aggregate.

All of the methods have various consequences. Prior to taking any action, you should read the relevant section in the documentation.

The following are some common ways to make space in an aggregate, in order of least to most consequences:

- Add disks to the aggregate.
- Move some volumes to another aggregate with available space.
- Shrink the size of volume-guaranteed volumes in the aggregate.

  You can do this manually or with the `autoshrink` option of the autosize capability.

- Change volume guarantee types to `none` on volumes that are using large amounts of space (large volume-guaranteed volumes with large reserved files) so that the volumes take up less space in the aggregate.

  A volume with a guarantee type of `none` has a smaller footprint in the aggregate than a volume with a guarantee type of `volume`.

- Delete unneeded volume Snapshot copies if the volume's guarantee type is `none`.
- Delete unneeded volumes.
- Enable space-saving features, such as deduplication or compression.
- (Temporarily) disable features that are using a large amount of metadata .

Commands for managing aggregates

You use the `storage aggregate` command to manage your aggregates.

If you want to…	Use this command…
Display the size of the cache for all Flash Pool aggregates	`storage aggregate show -fields hybrid-cache-size-total -hybrid-cache-size -total >0`
Display disk information and status for an aggregate	`storage aggregate show-status`
Display spare disks by node	`storage aggregate show-spare-disks`
Display the root aggregates in the cluster	`storage aggregate show -has-mroot true`
Display basic information and status for aggregates	`storage aggregate show`
Display the type of storage used in an aggregate	`storage aggregate show -fields storage-type`
Bring an aggregate online	`storage aggregate online`
Delete an aggregate	`storage aggregate delete`
Put an aggregate into the restricted state	`storage aggregate restrict`
Rename an aggregate	`storage aggregate rename`
Take an aggregate offline	`storage aggregate offline`
Change the RAID type for an aggregate	`storage aggregate modify -raidtype`

Related information

ONTAP 9 commands