NetApp

Work with Flash Pool aggregates

ONTAP 9

NetApp
March 21, 2022

Table of Contents

Work with Flash Pool aggregates

Flash Pool caching policies and SSD partitioning

How Flash Pool aggregate caching policies work

Caching policies for the volumes in a Flash Pool aggregate let you deploy flash as high performance cache for your working data set while using lower-cost HDDs for less frequently accessed data. If you are providing cache to two or more Flash Pool aggregates, you should use Flash Pool SSD partitioning to share SSDs across the aggregates in the Flash Pool.

Caching policies are applied to volumes that reside in Flash Pool aggregates. You should understand how caching policies work before changing them.

In most cases, the default caching policy of `auto` is the best caching policy to use. The caching policy should be changed only if a different policy provides better performance for your workload. Configuring the wrong caching policy can severely degrade volume performance; the performance degradation could increase gradually over time.

Caching policies combine a read caching policy and a write caching policy. The policy name concatenates the names of the read caching policy and the write caching policy, separated by a hyphen. If there is no hyphen in the policy name, the write caching policy is "none", except for the `auto` policy.

Read caching policies optimize for future read performance by placing a copy of the data in the cache in addition to the stored data on HDDs. For read caching policies that insert data into the cache for write operations, the cache operates as a *write-through* cache.

Data inserted into the cache by using the write caching policy exists only in cache; there is no copy in HDDs. Flash Pool cache is RAID protected. Enabling write caching makes data from write operations available for reads from cache immediately, while deferring writing the data to HDDs until it ages out of the cache.

You can change the caching policy for a volume that resides on a Flash Pool aggregate by using the `-caching-policy` parameter with the `volume create` command. When you create a volume on a Flash Pool aggregate, by default, the `auto` caching policy is assigned to the volume.

If you move a volume from a Flash Pool aggregate to a single-tier aggregate, it loses its caching policy; if you later move it back to a Flash Pool aggregate, it is assigned the default caching policy of `auto`. If you move a volume between two Flash Pool aggregates, the caching policy is preserved.

How Flash Pool SSD partitioning works for Flash Pool aggregates using storage pools

If you are providing cache to two or more Flash Pool aggregates, you should use Flash Pool Solid-State Drive (SSD) partitioning. Flash Pool SSD partitioning allows SSDs to be shared by all the aggregates using the Flash Pool. This spreads the cost of parity over multiple aggregates, increases SSD cache allocation flexibility, and maximizes SSD performance.

For an SSD to be used in a Flash Pool aggregate, the SSD must be placed in a storage pool. You cannot use

SSDs that have been partitioned for root-data partitioning in a storage pool. After the SSD is placed in the storage pool, the SSD can no longer be managed as a stand-alone disk and cannot be removed from the storage pool unless you destroy the aggregates associated with the Flash Pool and you destroy the storage pool.

SSD storage pools are divided into four equal allocation units. SSDs added to the storage pool are divided into four partitions and one partition is assigned to each of the four allocation units. The SSDs in the storage pool must be owned by the same HA pair. By default, two allocation units are assigned to each node in the HA pair. Allocation units must be owned by the node that owns the aggregate it is serving. If more Flash cache is required for aggregates on one of the nodes, the default number of allocation units can be shifted to decrease the number on one node and increase the number on the partner node.

You can use only one spare SSD for a storage pool. If the storage pool provides allocation units to Flash Pool aggregates owned by both nodes in the HA pair, then the spare SSD can be owned by either node. However, if the storage pool provides allocation units only to Flash Pool aggregates owned by one of the nodes in the HA pair, then the SSD spare must be owned by that same node.

The following illustration is an example of Flash Pool SSD partitioning. The SSD storage pool provides cache to two Flash Pool aggregates:

Storage pool SP1 is composed of five SSDs and a hot spare SSD. Two of the storage pool's allocation units are allocated to Flash Pool FP1, and two are allocated to Flash Pool FP2. FP1 has a cache RAID type of RAID4. Therefore, the allocation units provided to FP1 contain only one partition designated for parity. FP2 has a cache RAID type of RAID-DP. Therefore, the allocation units provided to FP2 include a parity partition and a double-parity partition.

In this example, two allocation units are allocated to each Flash Pool aggregate. However, if one Flash Pool aggregate required a larger cache, you could allocate three of the allocation units to that Flash Pool aggregate, and only one to the other.

Determine Flash Pool candidacy and optimal cache size

Before converting an existing aggregate to a Flash Pool aggregate, you can determine whether the aggregate is I/O bound, and what would be the best Flash Pool cache size for your workload and budget. You can also check whether the cache of an existing Flash Pool aggregate is sized correctly.

What you'll need

You should know approximately when the aggregate you are analyzing experiences its peak load.

Steps

1. Enter advanced mode:

   ```
   set advanced
   ```

2. If you need to determine whether an existing aggregate would be a good candidate for conversion to a Flash Pool aggregate, determine how busy the disks in the aggregate are during a period of peak load, and how that is affecting latency:

   ```
   statistics show-periodic -object disk:raid_group -instance raid_group_name
   -counter disk_busy|user_read_latency -interval 1 -iterations 60
   ```

   You can decide whether reducing latency by adding Flash Pool cache makes sense for this aggregate.

   The following command shows the statistics for the first RAID group of the aggregate "aggr1":

   ```
   statistics show-periodic -object disk:raid_group -instance /aggr1/plex0/rg0
   -counter disk_busy|user_read_latency -interval 1 -iterations 60
   ```

3. Start Automated Workload Analyzer (AWA):

   ```
   storage automated-working-set-analyzer start -node node_name -aggregate
   aggr_name
   ```

   AWA begins collecting workload data for the volumes associated with the specified aggregate.

4. Exit advanced mode:

   ```
   set admin
   ```

   Allow AWA to run until one or more intervals of peak load have occurred. AWA collects workload statistics for the volumes associated with the specified aggregate, and analyzes data for up to one rolling week in duration. Running AWA for more than one week will report only on data collected from the most recent week. Cache size estimates are based on the highest loads seen during the data collection period; the load does not need to be high for the entire data collection period.

5. Enter advanced mode:

   ```
   set advanced
   ```

6. Display the workload analysis:

   ```
   storage automated-working-set-analyzer show -node node_name
   ```

7. Stop AWA:

```
storage automated-working-set-analyzer stop node_name
```

All workload data is flushed and is no longer available for analysis.

8. Exit advanced mode:

```
set admin
```

Create a Flash Pool aggregate using physical SSDs

You create a Flash Pool aggregate by enabling the feature on an existing aggregate composed of HDD RAID groups, and then adding one or more SSD RAID groups to that aggregate. This results in two sets of RAID groups for that aggregate: SSD RAID groups (the SSD cache) and HDD RAID groups.

What you'll need

- You must have identified a valid aggregate composed of HDDs to convert to a Flash Pool aggregate.
- You must have determined write-caching eligibility of the volumes associated with the aggregate, and completed any required steps to resolve eligibility issues.
- You must have determined the SSDs you will be adding, and these SSDs must be owned by the node on which you are creating the Flash Pool aggregate.
- You must have determined the checksum types of both the SSDs you are adding and the HDDs already in the aggregate.
- You must have determined the number of SSDs you are adding and the optimal RAID group size for the SSD RAID groups.

  Using fewer RAID groups in the SSD cache reduces the number of parity disks required, but larger RAID groups require RAID-DP.

- You must have determined the RAID level you want to use for the SSD cache.
- You must have determined the maximum cache size for your system and determined that adding SSD cache to your aggregate will not cause you to exceed it.
- You must have familiarized yourself with the configuration requirements for Flash Pool aggregates.

About this task

After you add an SSD cache to an aggregate to create a Flash Pool aggregate, you cannot remove the SSD cache to convert the aggregate back to its original configuration.

By default, the RAID level of the SSD cache is the same as the RAID level of the HDD RAID groups. You can override this default selection by specifying the `raidtype` option when you add the first SSD RAID groups.

Steps

1. Mark the aggregate as eligible to become a Flash Pool aggregate:

```
storage aggregate modify -aggregate aggr_name -hybrid-enabled true
```

If this step does not succeed, determine write-caching eligibility for the target aggregate.

2. Add the SSDs to the aggregate by using the `storage aggregate add` command.

   You can specify the SSDs by ID or by using the `diskcount` and `disktype` parameters.

   If the HDDs and the SSDs do not have the same checksum type, or if the aggregate is a mixed-checksum aggregate, then you must use the `checksumstyle` parameter to specify the checksum type of the disks you are adding to the aggregate.

   You can specify a different RAID type for the SSD cache by using the `raidtype` parameter.

   If you want the cache RAID group size to be different from the default for the RAID type you are using, you should change it now, by using the `-cache-raid-group-size` parameter.

Create a Flash Pool aggregate using SSD storage pools

Determine whether a Flash Pool aggregate is using an SSD storage pool

You can configure a Flash Pool aggregate by adding one or more allocation units from an SSD storage pool to an existing HDD aggregate.

You manage Flash Pool aggregates differently when they use SSD storage pools to provide their cache than when they use discrete SSDs.

Step

1. Display the aggregate's drives by RAID group:

   `storage aggregate show-status aggr_name`

   If the aggregate is using one or more SSD storage pools, the value for the `Position` column for the SSD RAID groups is displayed as `Shared`, and the name of the storage pool is displayed next to the RAID group name.

Create an SSD storage pool

You can create solid state drive (SSD) storage pools to provide SSD cache for two to four Flash Pool aggregates.

About this task

- You must supply a disk list when creating or adding disks to a storage pool.

  Storage pools do not support a `diskcount` parameter.

- The SSDs used in the storage pool should be the same size.

Steps

1. Determine the names of the available spare SSDs:

   `storage aggregate show-spare-disks -disk-type SSD`

   The SSDs used in a storage pool can be owned by either node of an HA pair.

2. Create the storage pool:

```
storage pool create -storage-pool sp_name -disk-list disk1,disk2,…
```

3. **Optional:** Verify the newly created storage pool:

```
storage pool show -storage-pool sp_name
```

Results

After the SSDs are placed into the storage pool, they no longer appear as spares on the cluster, even though the storage provided by the storage pool has not yet been allocated to any Flash Pool caches. You cannot add SSDs to a RAID group as discrete drives; their storage can be provisioned only by using the allocation units of the storage pool to which they belong.

Create a Flash Pool aggregate using SSD storage pool allocation units

You can configure a Flash Pool aggregate by adding one or more allocation units from an SSD storage pool to an existing HDD aggregate.

What you'll need

- You must have identified a valid aggregate composed of HDDs to convert to a Flash Pool aggregate.
- You must have determined write-caching eligibility of the volumes associated with the aggregate, and completed any required steps to resolve eligibility issues.
- You must have created an SSD storage pool to provide the SSD cache to this Flash Pool aggregate.

  Any allocation unit from the storage pool that you want to use must be owned by the same node that owns the Flash Pool aggregate.

- You must have determined how much cache you want to add to the aggregate.

  You add cache to the aggregate by allocation units. You can increase the size of the allocation units later by adding SSDs to the storage pool if there is room.

- You must have determined the RAID type you want to use for the SSD cache.

  After you add a cache to the aggregate from SSD storage pools, you cannot change the RAID type of the cache RAID groups.

- You must have determined the maximum cache size for your system and determined that adding SSD cache to your aggregate will not cause you to exceed it.

  You can see the amount of cache that will be added to the total cache size by using the `storage pool show` command.

- You must have familiarized yourself with the configuration requirements for Flash Pool aggregates.

About this task

If you want the RAID type of the cache to different from that of the HDD RAID groups, you must specify the cache RAID type when you add the SSD capacity. After you add the SSD capacity to the aggregate, you can no longer change the RAID type of the cache.

After you add an SSD cache to an aggregate to create a Flash Pool aggregate, you cannot remove the SSD

cache to convert the aggregate back to its original configuration.

Steps

1. Mark the aggregate as eligible to become a Flash Pool aggregate:

   ```
   storage aggregate modify -aggregate aggr_name -hybrid-enabled true
   ```

   If this step does not succeed, determine write-caching eligibility for the target aggregate.

2. Show the available SSD storage pool allocation units:

   ```
   storage pool show-available-capacity
   ```

3. Add the SSD capacity to the aggregate:

   ```
   storage aggregate add aggr_name -storage-pool sp_name -allocation-units
   number_of_units
   ```

   If you want the RAID type of the cache to be different from that of the HDD RAID groups, you must change it when you enter this command by using the `raidtype` parameter.

   You do not need to specify a new RAID group; ONTAP automatically puts the SSD cache into separate RAID groups from the HDD RAID groups.

   You cannot set the RAID group size of the cache; it is determined by the number of SSDs in the storage pool.

   The cache is added to the aggregate and the aggregate is now a Flash Pool aggregate. Each allocation unit added to the aggregate becomes its own RAID group.

4. Confirm the presence and size of the SSD cache:

   ```
   storage aggregate show aggr_name
   ```

   The size of the cache is listed under `Total Hybrid Cache Size`.

Related information

[NetApp Technical Report 4070: Flash Pool Design and Implementation Guide](#)

Determine the impact to cache size of adding SSDs to an SSD storage pool

If adding SSDs to a storage pool causes your platform model's cache limit to be exceeded, ONTAP does not allocate the newly added capacity to any Flash Pool aggregates. This can result in some or all of the newly added capacity being unavailable for use.

About this task

When you add SSDs to an SSD storage pool that has allocation units already allocated to Flash Pool aggregates, you increase the cache size of each of those aggregates and the total cache on the system. If none of the storage pool's allocation units have been allocated, adding SSDs to that storage pool does not affect the SSD cache size until one or more allocation units are allocated to a cache.

Steps

1. Determine the usable size of the SSDs you are adding to the storage pool:

   ```
   storage disk show disk_name -fields usable-size
   ```

2. Determine how many allocation units remain unallocated for the storage pool:

   ```
   storage pool show-available-capacity sp_name
   ```

   All unallocated allocation units in the storage pool are displayed.

3. Calculate the amount of cache that will be added by applying the following formula:

   (4 — number of unallocated allocation units) × 25% × usable size × number of SSDs

Add SSDs to an SSD storage pool

When you add solid state drives (SSDs) to an SSD storage pool, you increase the storage pool's physical and usable sizes and allocation unit size. The larger allocation unit size also affects allocation units that have already been allocated to Flash Pool aggregates.

What you'll need

You must have determined that this operation will not cause you to exceed the cache limit for your HA pair. ONTAP does not prevent you from exceeding the cache limit when you add SSDs to an SSD storage pool, and doing so can render the newly added storage capacity unavailable for use.

About this task

When you add SSDs to an existing SSD storage pool, the SSDs must be owned by one node or the other of the same HA pair that already owned the existing SSDs in the storage pool. You can add SSDs that are owned by either node of the HA pair.

The SSD you add to the storage pool must be the same size as disk currently used in the storage pool.

Steps

1. **Optional:** View the current allocation unit size and available storage for the storage pool:

   ```
   storage pool show -instance sp_name
   ```

2. Find available SSDs:

   ```
   storage disk show -container-type spare -type SSD
   ```

3. Add the SSDs to the storage pool:

   ```
   storage pool add -storage-pool sp_name -disk-list disk1,disk2…
   ```

   The system displays which Flash Pool aggregates will have their size increased by this operation and by how much, and prompts you to confirm the operation.

Commands for managing SSD storage pools

ONTAP provides the `storage pool` command for managing SSD storage pools.

If you want to…	Use this command…
Display how much storage a storage pool is providing to which aggregates	`storage pool show-aggregate`
Display how much cache would be added to the overall cache capacity for both RAID types (allocation unit data size)	`storage pool show -instance`
Display the disks in a storage pool	`storage pool show-disks`
Display the unallocated allocation units for a storage pool	`storage pool show-available-capacity`
Change the ownership of one or more allocation units of a storage pool from one HA partner to the other	`storage pool reassign`

Related information

ONTAP 9 commands

Determine whether to modify the caching policy of Flash Pool aggregates

Determine whether to modify the caching policy of Flash Pool aggregates overview

You can assign cache-retention policies to volumes in Flash Pool aggregates to determine how long the volume data remains in the Flash Pool cache. However, in some cases changing the cache-retention policy might not impact the amount of time the volume's data remains in the cache.

About this task

If your data meets any of the following conditions, changing your cache-retention policy might not have an impact:

- Your workload is sequential.
- Your workload does not reread the random blocks cached in the solid state drives (SSDs).
- The cache size of the volume is too small.

The following steps check for these conditions. The task must be done in advanced privilege mode.

Steps

1. View the workload volume:

```
statistics start -object workload_volume
```

2. Determine the workload pattern of the volume:

```
statistics show -object workload_volume -instance volume-workload -counter
sequential_reads
```

3. Determine the hit rate of the volume:

```
statistics show -object wafl_hya_vvol -instance volume -counter
read_ops_replaced_pwercent|wc_write_blks_overwritten_percent
```

4. Determine the `Cacheable Read` and `Project Cache Alloc` of the volume:

```
system node run -node node_name wafl awa start aggr_name
```

5. Display the AWA summary:

```
system node run -node node_name wafl awa print aggr_name
```

6. Compare the volume's hit rate to the `Cacheable Read`.

   If the hit rate of the volume is greater than the `Cacheable Read`, then your workload does not reread random blocks cached in the SSDs.

7. Compare the volume's current cache size to the `Project Cache Alloc`.

   If the current cache size of the volume is greater than the `Project Cache Alloc`, then the size of your volume cache is too small.

Modify caching policies of Flash Pool aggregates

You should modify the caching policy of a volume only if a different caching policy is expected to provide better performance. You can modify the caching policy of a volume on a Flash Pool aggregate.

What you'll need

You must determine whether you want to modify your caching policy.

About this task

In most cases, the default caching policy of `auto` is the best caching policy that you can use. The caching policy should be changed only if a different policy provides better performance for your workload. Configuring the wrong caching policy can severely degrade volume performance; the performance degradation could increase gradually over time. You should use caution when modifying caching policies. If you experience performance issues with a volume for which the caching policy has been changed, you should return the caching policy to `auto`.

Step

1. Modify the volume's caching policy:

```
volume modify -volume volume_name -caching-policy policy_name
```

Example

The following example modifies the caching policy of a volume named "vol2" to the policy `none`:

```
volume modify -volume vol2 -caching-policy none
```

Set the cache-retention policy for Flash Pool aggregates

You can assign cache-retention policies to volumes in Flash Pool aggregates. Data in volumes with a high cache-retention policy remains in cache longer and data in volumes with a low cache-retention policy is removed sooner. This increases performance of your critical workloads by making high priority information accessible at a faster rate for a longer period of time.

What you'll need

You should know whether your system has any conditions that might prevent the cache-retention policy from having an impact on how long your data remains in cache.

About this task

The task must be done in advanced privilege mode.

Steps

1. Change the privilege setting to advanced:

   ```
   set -privilege advanced
   ```

2. Verify the volume's cache-retention policy:

   By default the cache retention policy is `normal`.

3. Set the cache-retention policy:

ONTAP Version	Command
ONTAP 9.0, 9.1	`priority hybrid-cache set volume_name read-cache=read_cache_value write-cache=write_cache_value cache-retention-priority=cache_retention_policy` Set `cache_retention_policy` to `high` for data that you want to remain in cache longer. Set `cache_retention_policy` to `low` for data that you want to remove from cache sooner.
ONTAP 9.2 or later	`volume modify -volume volume_name -vserver vserver_name -caching-policy policy_name.`

4. Verify that the volume's cache-retention policy is changed to the option you selected.

5. Return the privilege setting to admin:

```
set -privilege admin
```