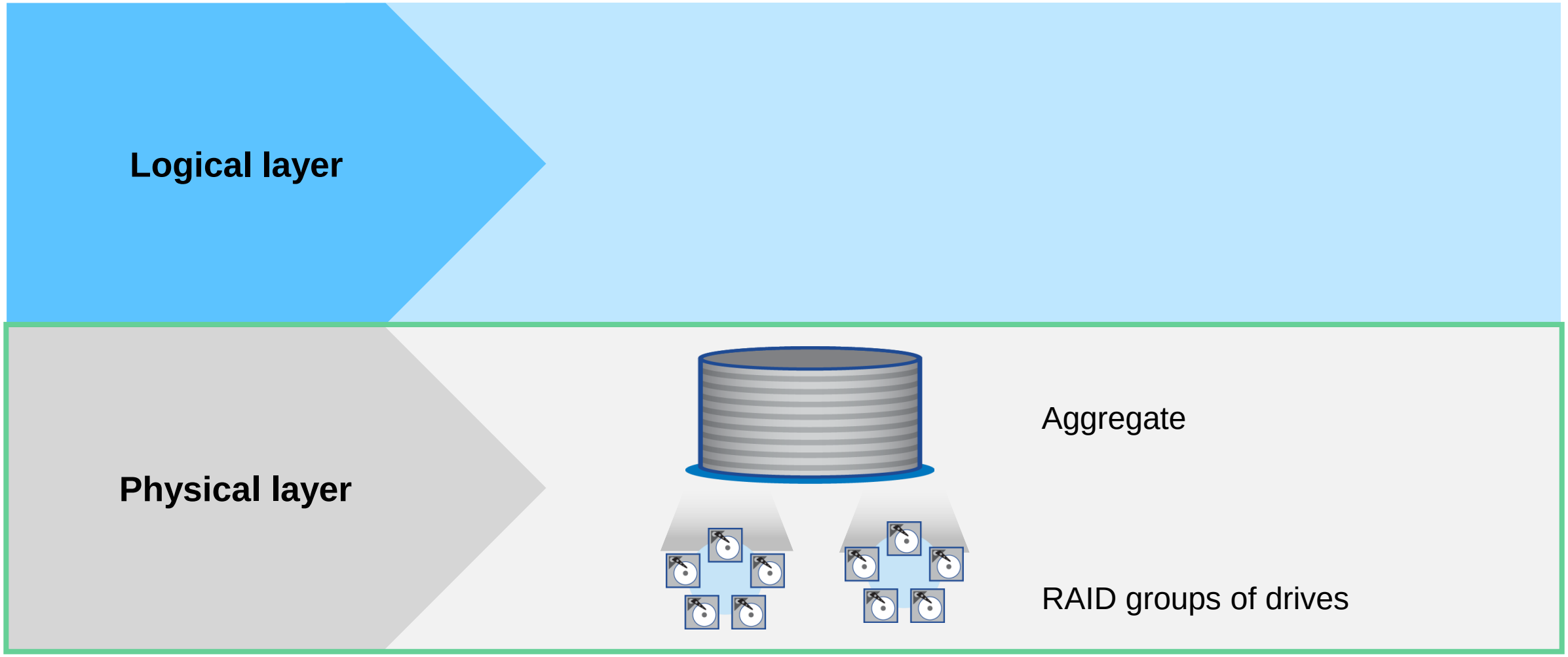
Module 5
Physical storage management

About this module

This module focuses on enabling you to do the following:

- Recognize NetApp ONTAP storage architecture concepts
- Manage physical storage resources, including drives, RAID groups, and aggregates
- Create data aggregates
- Create Flash Pool aggregates
- Set up FabricPool aggregates

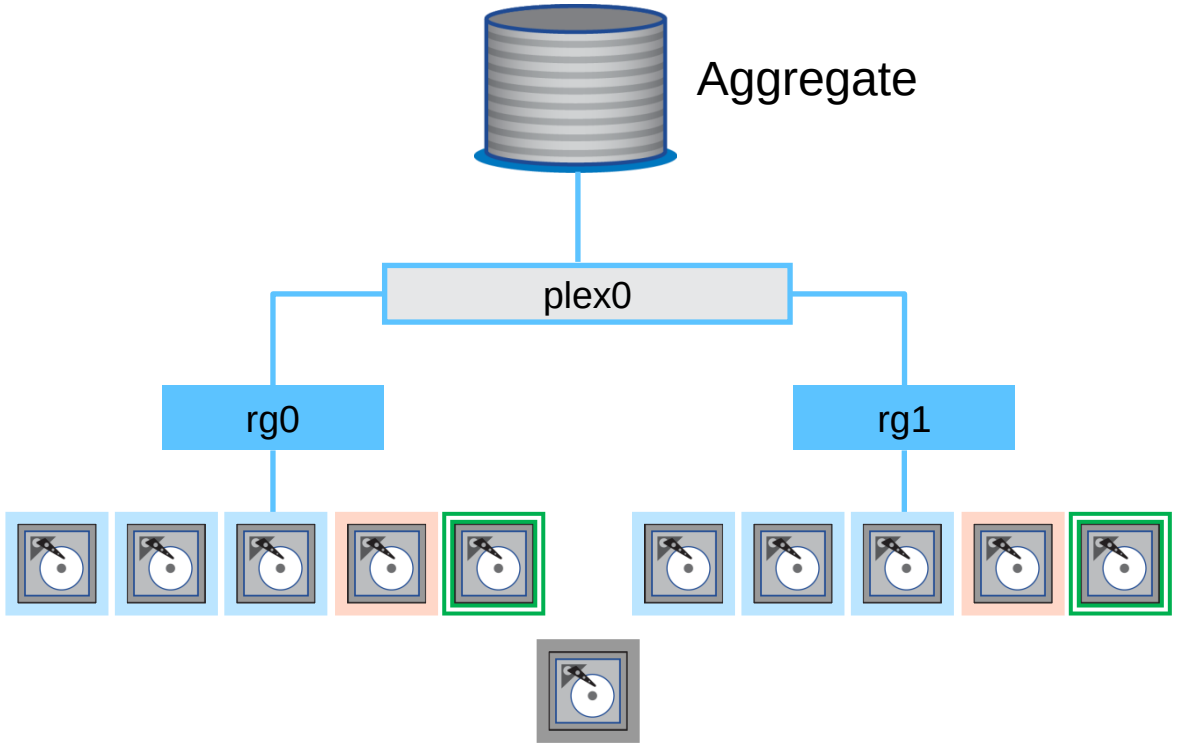
ONTAP storage architecture



Lesson 1
Drives, RAID, and aggregates

Physical storage hierarchy

- **Drive:**
  HDD or SSD
- **RAID group:**
  Drive-level protection
- **Plex:**
  Container for RAID groups
  Used by mirrored aggregates
- **Aggregate:**
  Pool of storage space



Drive types

NL-SAS HDD

Near-Line Serial Attached SCSI

- Same technology that is used in consumer disk drives
- Dual I/O path
- High capacity but moderate IOPS

SAS HDD

Serial Attached SCSI

- Point-to-point serial protocol
- Multipath I/O
- Moderate capacity but high IOPS

SAS SSD

Solid-State Drive

- No spinning platter
- Based on flash memory chip technology that is like USB flash drives
- High IOPs with low latency

Can also be used as an aggregate-specific cache

NVMe SSD

Nonvolatile Memory Express

- Solid state flash memory drives that are accessed by using the NVMe protocol
- Extreme IOPS for demanding workloads

Drive ownership

- A drive is unusable until the drive is assigned ownership to a storage controller.
  - By default, ownership is automatically assigned.
  - Ownership can be manually assigned or changed.
  - Software disk ownership is made persistent by writing the ownership information onto the drive.

- Spare drives can be reassigned or unassigned.

```
::> storage disk show -container-type unassigned
                 Usable             Container
Disk             Size   Shelf Bay   Type       Position  Aggregate Owner
---------------- ------ ----- ----- ---------- --------- --------- -----
9.11.18               -    11    18 unassigned present   -         -
```

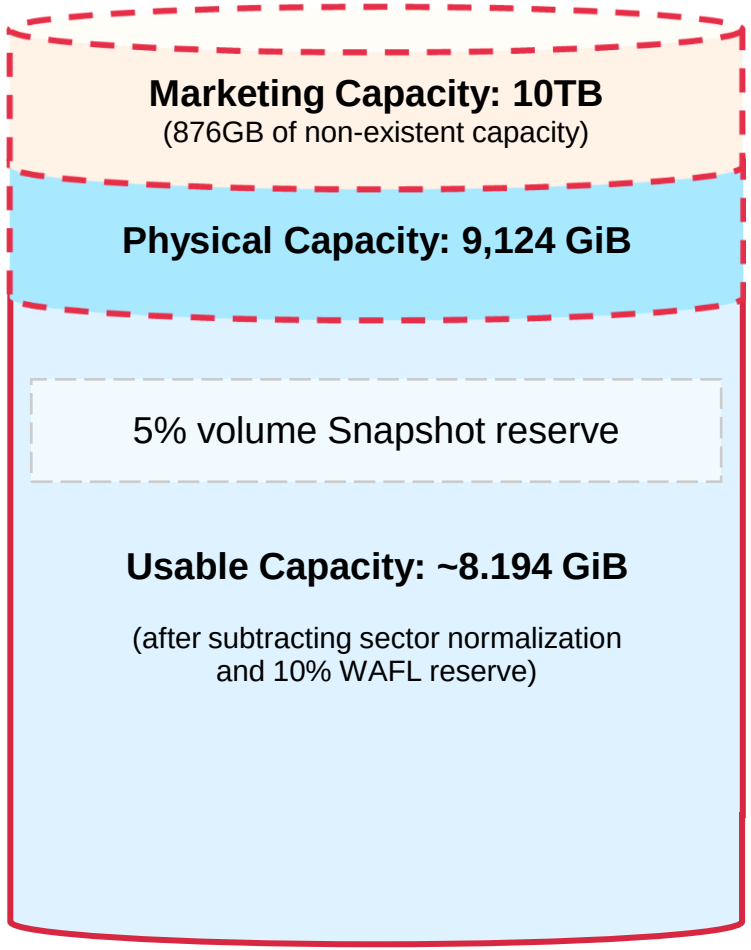

Drive capacity

Marketing and physical capacity

Drive capacity is a confusing and contentious subject for many reasons. Consider a 10TB disk drive:

- Marketing capacity: Base-10 number rounded up to an even number
- Physical or raw capacity: Available space after sector formatting in base-2 numbering
  After formatting, a 10TB drive is really a 9124GiB drive.
- Usable capacity: Space available to NetApp WAFL in base-2 numbering
  - Sector normalization: ONTAP software rightsizes all “10TB” drives equally, so they have the same number of sectors. This arrangement might result in 9105GiB.
  - NetApp WAFL reserve:
    This reserve is 10% of capacity that is set aside for WAFL metadata.

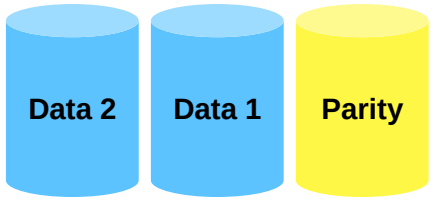
  The space available for data is now ~ 8,194GiB,
  which appears to an end-user as if 2TB has vanished.
- Snapshot reserve: FlexVol volumes reserve 5% for storing Snapshot copies, which the customer might perceive as more lost space.

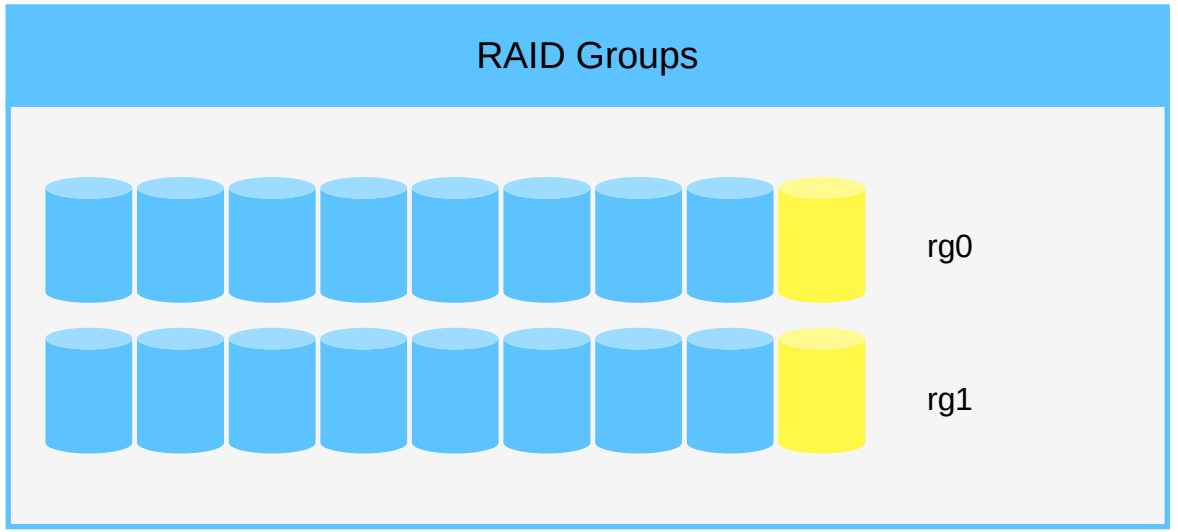


Drive roles

Parity drive

Parity drive: Stores row parity information that is used for data reconstruction when a single drive fails within the RAID group

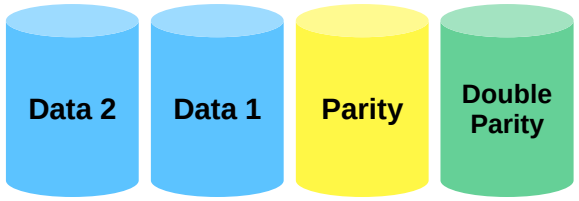


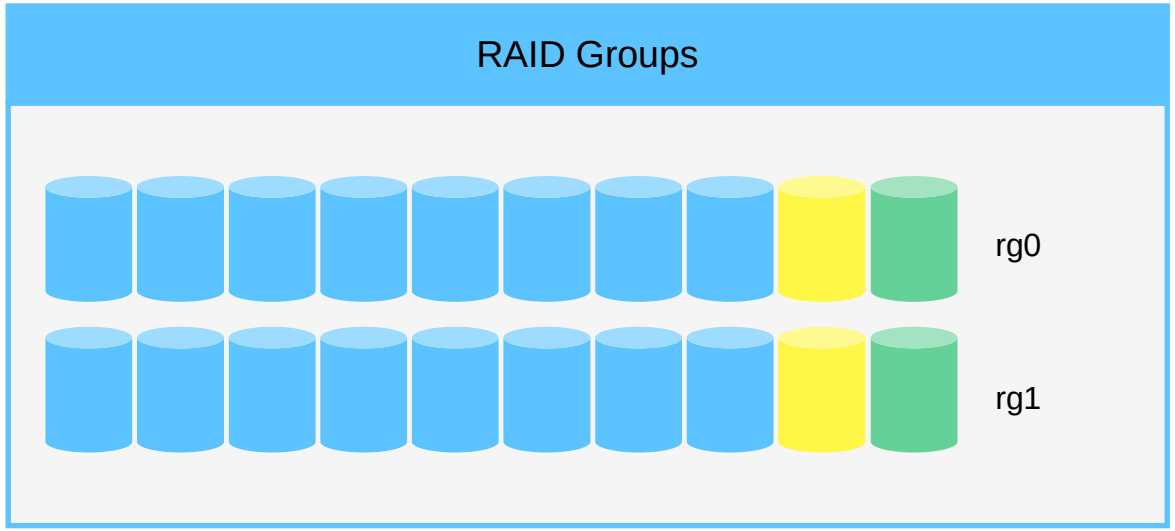


Drive roles

Double parity drive

dParity drive: Stores diagonal parity information that is used for data reconstruction when two drives fail within the RAID group

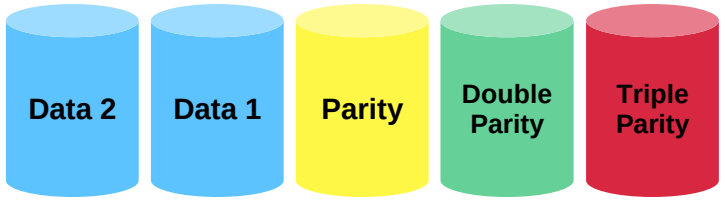


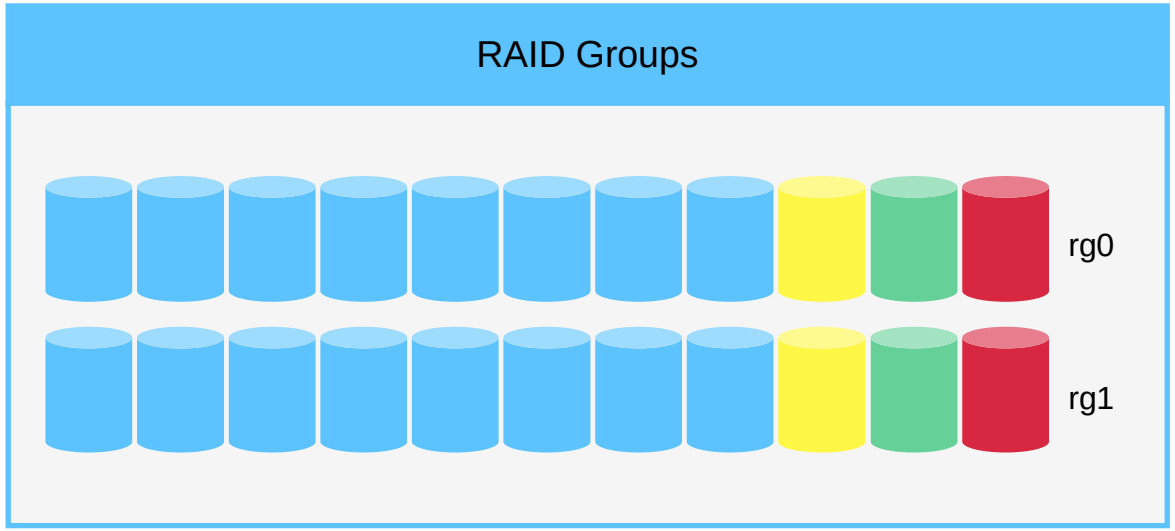


Drive roles

Triple parity drive

tParity drive: Stores anti-diagonal parity information that is used for data reconstruction when three drives fail within the RAID group





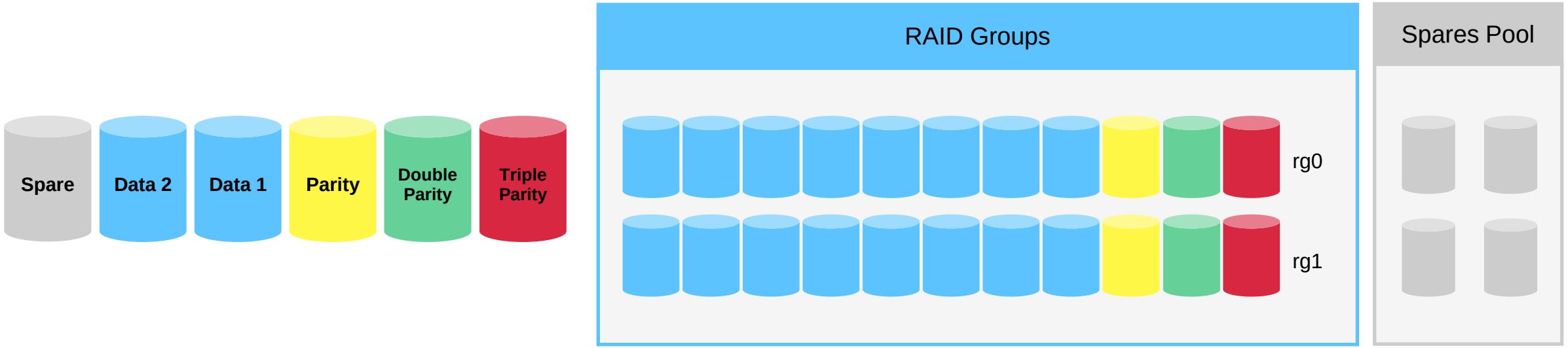
Drive roles

Spare drive

Spare drive:

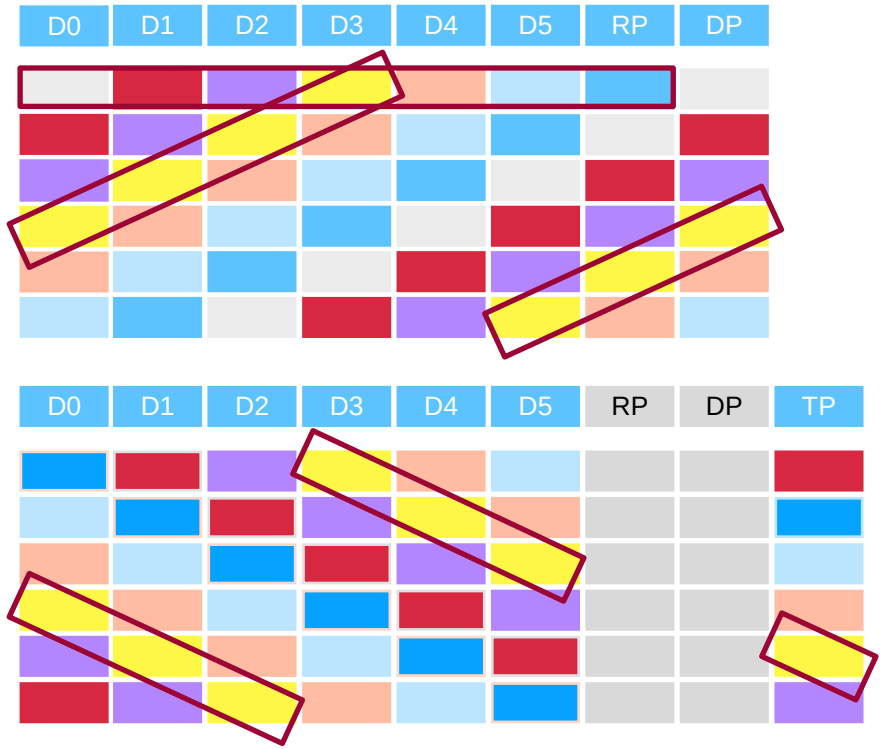
- Assigned to a storage system but not in use by a RAID group
- Used to create aggregates, add capacity to aggregates, and to replace failing drives

  Spare drives must be “zeroed” before use.



ONTAP RAID technologies

Description



- RAID 4 (row parity)
  - Adds a *row parity* drive
  - Protects against single-disk failure or media error
- RAID DP (double parity) technology
  - Adds a *diagonal parity* disk to a RAID 4 group
  - Protects against two concurrent drive failures within a RAID group
- RAID-TEC (triple erasure coding) technology
  - Adds a *triple-parity* disk to a RAID DP group
  - Protects against three concurrent drive failures

RAID group sizes

Default RAID group sizes:

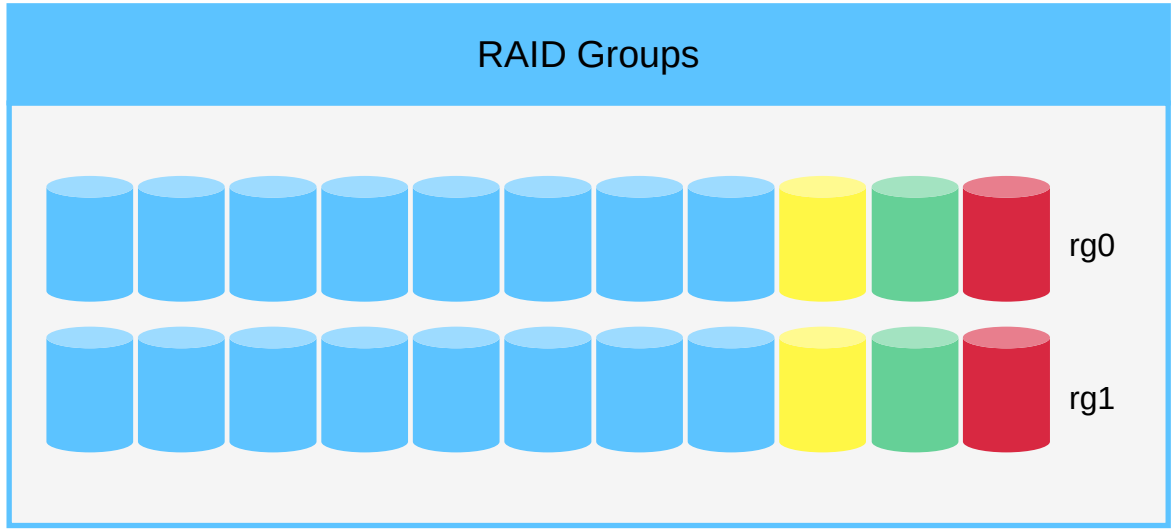
- 21 drives for SATA or NL-SAS drives
- 24 drives for SAS, SSD, or NVMe drives

When you expand an aggregate, always add the equivalent of half the RAID group size (7 to 14 drives) to avoid a degradation in performance.

Disk Type	Group Type	Default	Maximum
NL-SAS	RAID4	7	7
	RAID DP	14	20
	RAID-TEC	***21***	***29***
SAS	RAID4	8	14
	RAID DP	16	28
	RAID-TEC	***24***	***29***
SSD	RAID4	8	14
	RAID DP	23	28
	RAID-TEC	***24***	***29***

RAID group recommendations

- Drives must be the same type:
  - SAS, NL-SAS, or SSD
- Drives should be the same size.
- HDD should be the same rotational speed:
  - SAS 10K RPM
  - NL-SAS 7.2K RPM
- You should provide sufficient hot spares.



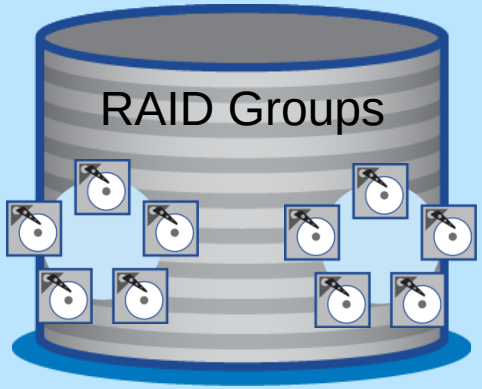
Topic for discussion

How many spare drives are needed?

Aggregates

- Aggregates are logical containers for the drives that are managed by a node.
- Aggregates consist of one or more RAID groups.
- You can use aggregates to do the following:
  - Isolate workloads with different performance demands
  - Tier data with different access patterns
  - Segregate data for regulatory purposes
- A single node owns an aggregate, but ownership can be transferred to the partner in a high-availability (HA) pair.

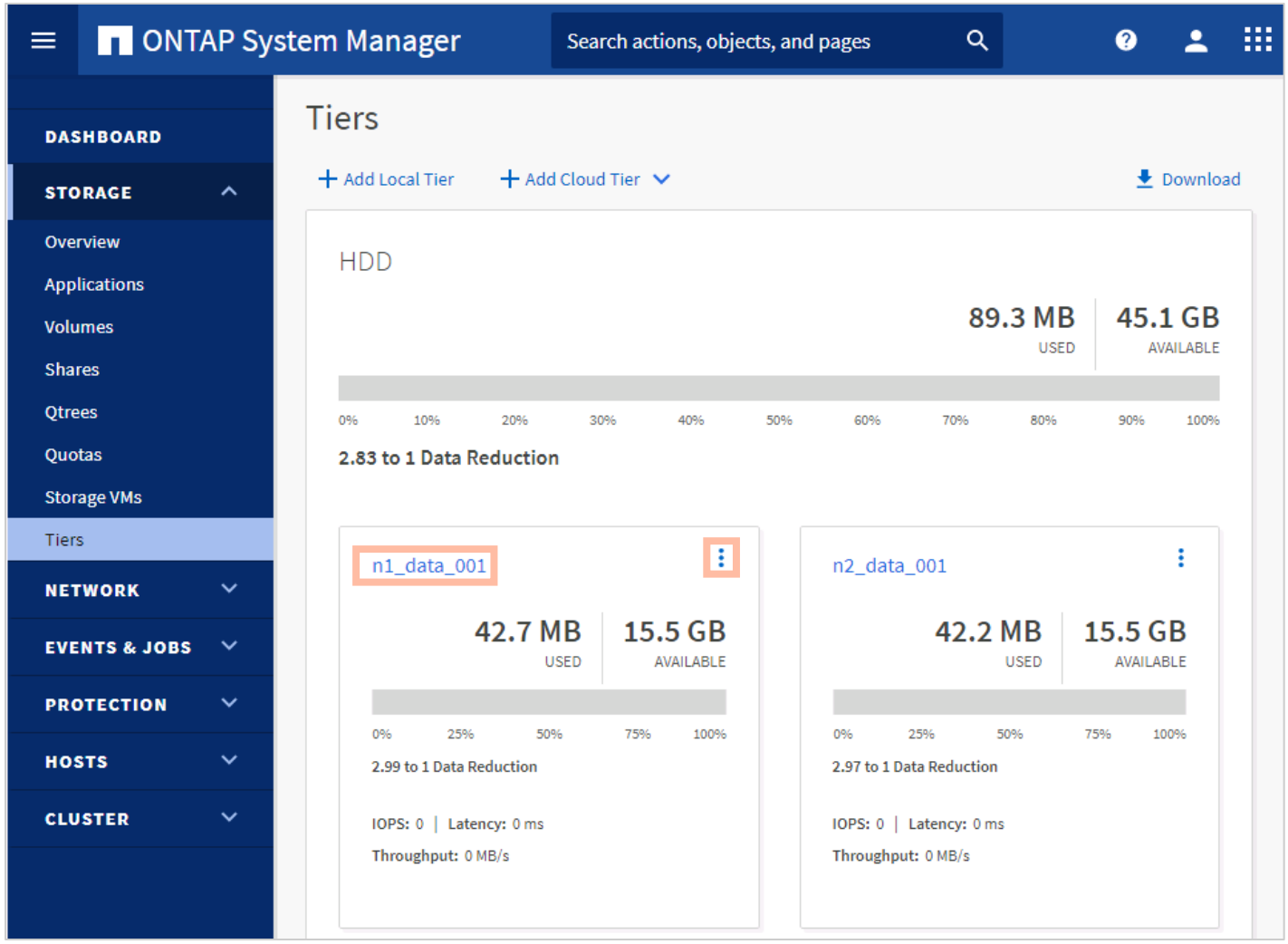
  During an HA failover, aggregate ownership is temporarily transferred to the surviving partner.



Aggregate

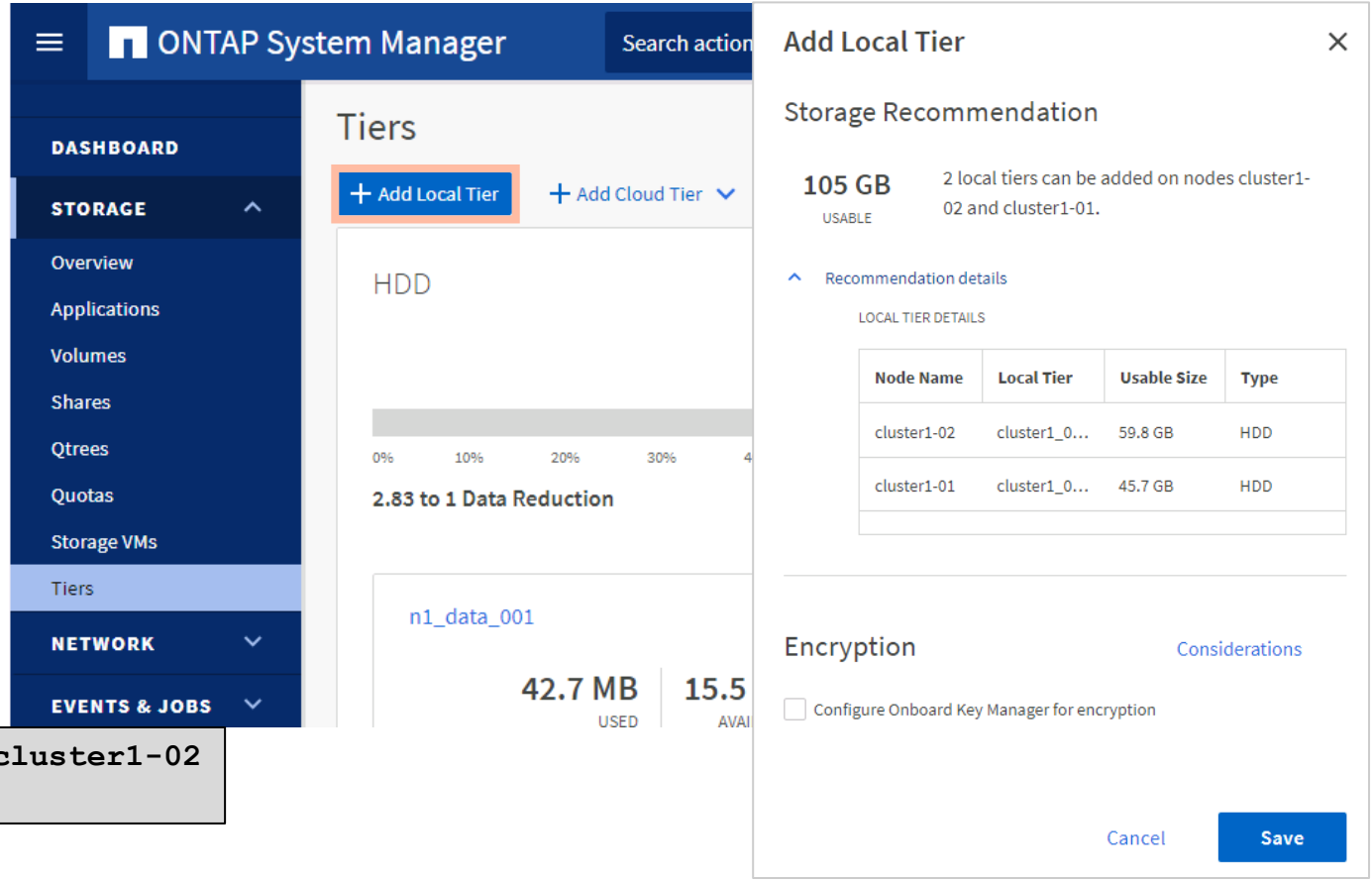
Viewing aggregates in System Manager

- Navigate to the NetApp ONTAP System Manager Tiers page: **Storage > Tiers**
- ONTAP System Manager refers to aggregates as ”local tiers.”
- Local tiers are grouped by drive type (HDD, SSD, and Flash Pool).



Create an aggregate

- System Manager creates only best practice conformant aggregates.
  - Generally, one aggregate per node per drive type is created.
  - Click **Recommendation details** to view the planned aggregates.
  - Click **Save** to create the aggregates.
- Use the CLI or API to create non-conformant aggregates.




```
cluster2::> aggr create -aggregate n2_data_02 -node cluster1-02
            -disktype ssd -diskcount 8
```

Try this task

Use cluster1 in your exercise kit to try the following tasks:

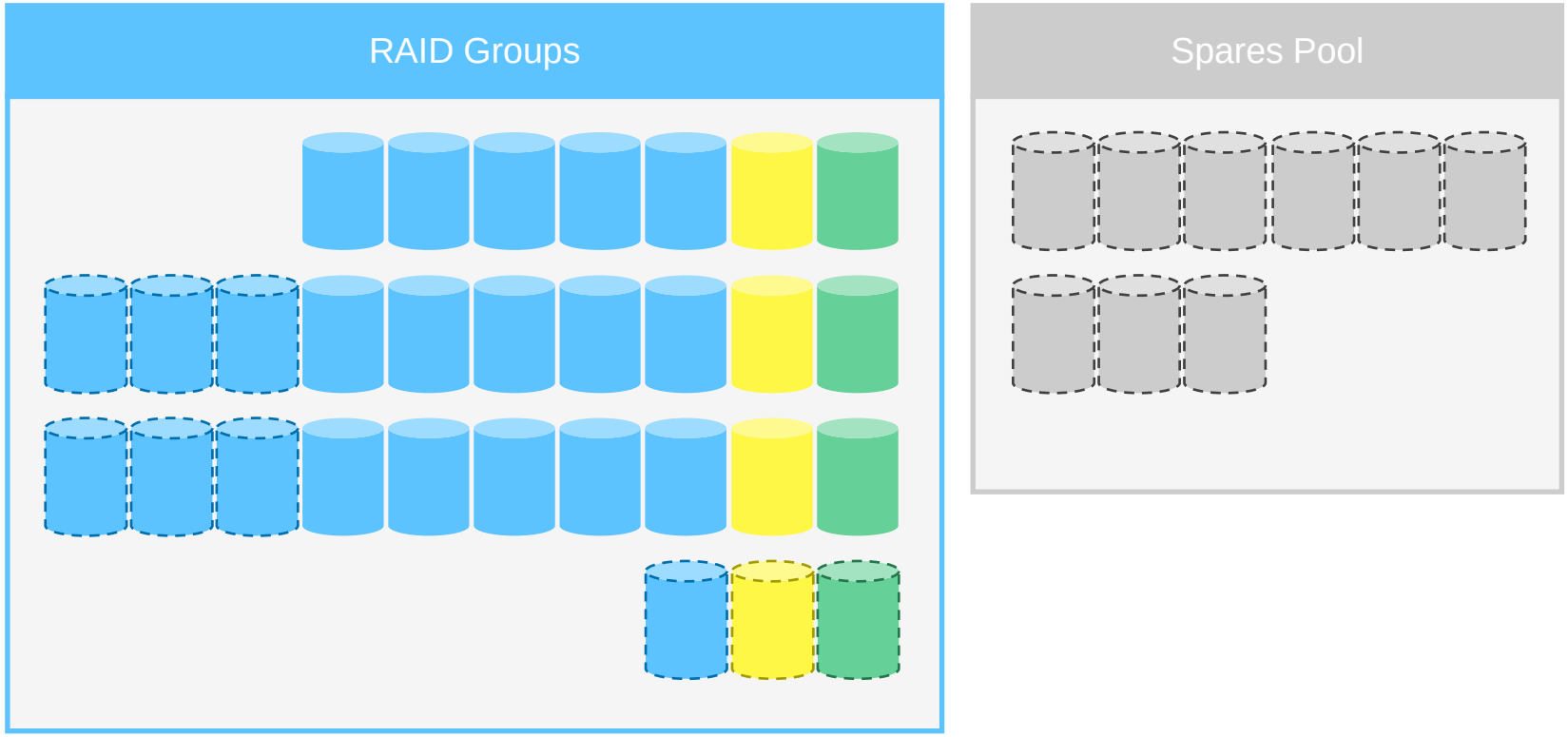
- Open a PuTTY session and use the `aggr show` command.
  - Can you tell which node owns the aggregate?
  - What is the RAID status?
  - How can you determine how many disks are in each aggregate?
- Different commands show similar things in different ways:
  - Enter `aggr show –aggregate aggr0_n1`.
  - Enter `storage disk show –aggr aggr0_n1`.
  - How do the outputs of the commands differ?

Adding drives to an aggregate

To add capacity to an aggregate, you add more drives. Careful planning ensures that you use the fewest drives to add the maximum amount of capacity.

Example: Aggregate composed of 4TB drives

- Three drives add 12TB of capacity and fill out the RAID group.
- To add 16TB, you need six drives.
  - You have no more spares.
  - The “runt” RAID group decreases performance because IOPS is serviced by a single drive.

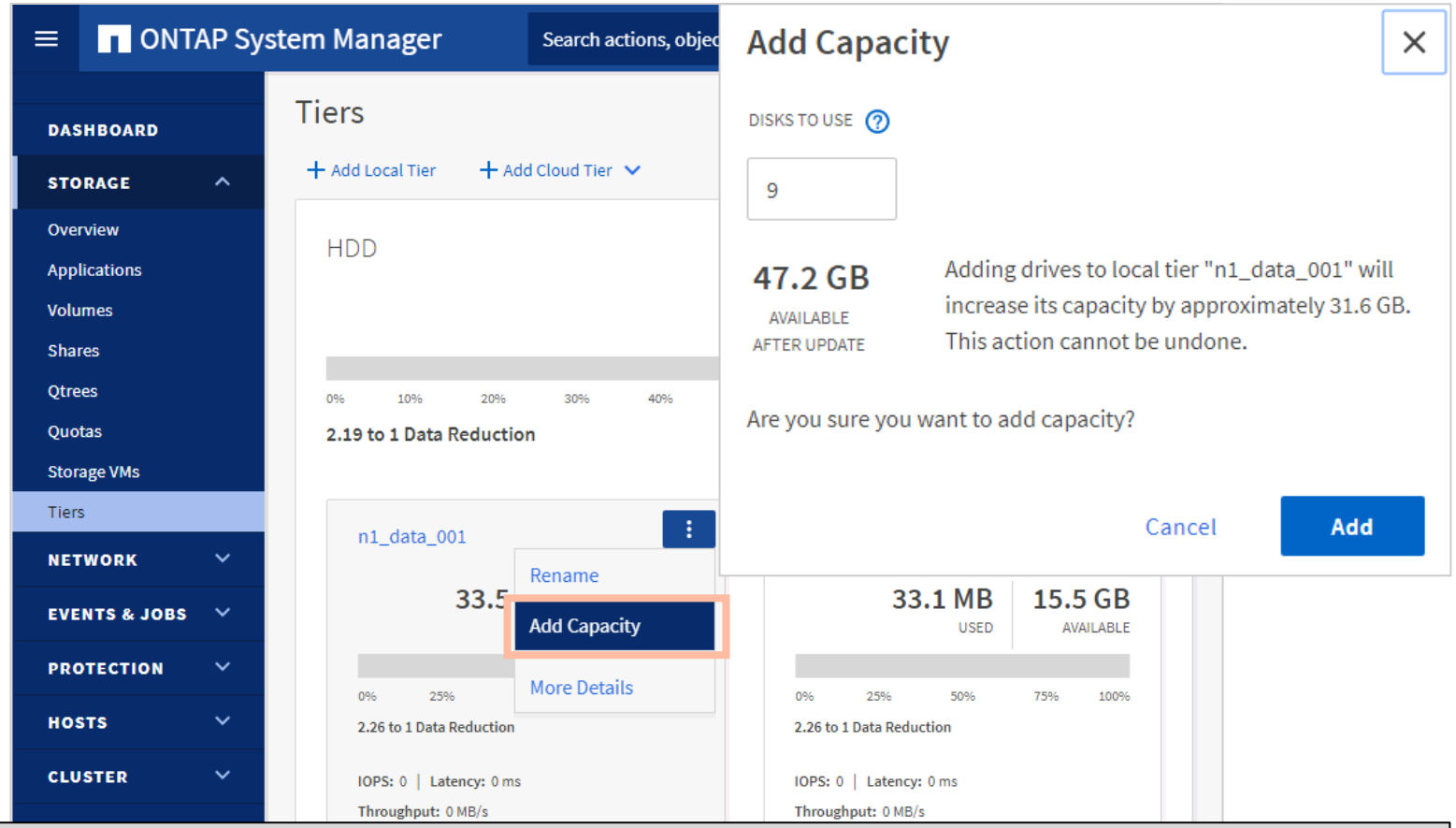


Adding capacity to aggregates

Provide the following information:

- Aggregate name
- Disks to add:
  - Disk count
  - Disk type
  - Disk class
  - Disk size
  - Disk list

You cannot shrink aggregates.

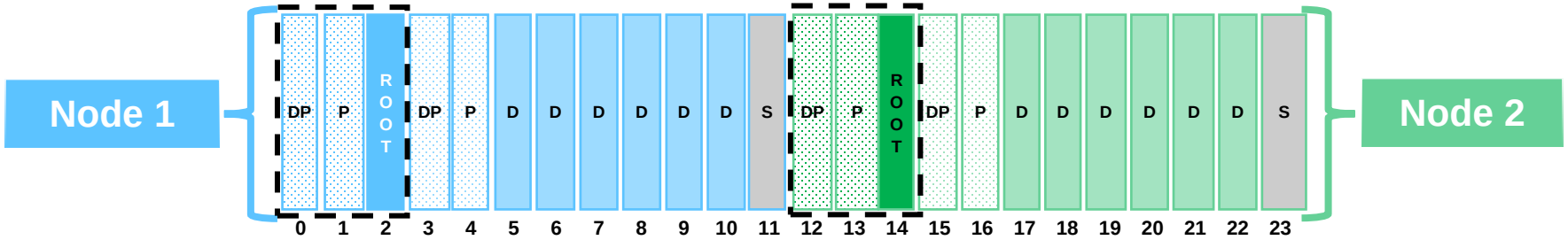



```
::> storage disk show -spare -owner cluster1-01
::> storage aggregate add-disks -aggr n1_data_001 -diskclass capacity -diskcount 3
```


Lesson 2
Advanced Disk Partitioning

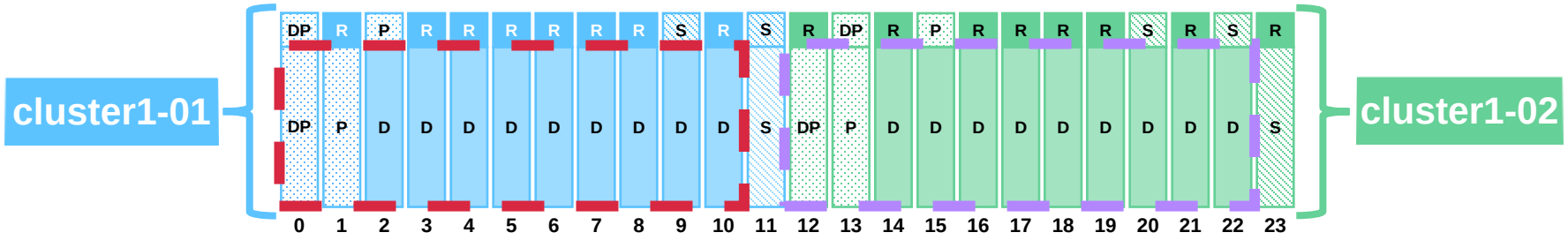
Why slice drives into partitions?

Before ONTAP 8.3 software, the following is how entry-level HA pairs used their drives.



- Of the 24 drives in the chassis, each node can use only 6 drives to store data:
  - 4 drives are used for parity.
  - 1 drive is reserved as a spare.
  - 6 drives are used to store the node root aggregates (one RAID4 aggregate per node).
  - 6 drives are available to store data.
- Efficiency was limited to about 40%.

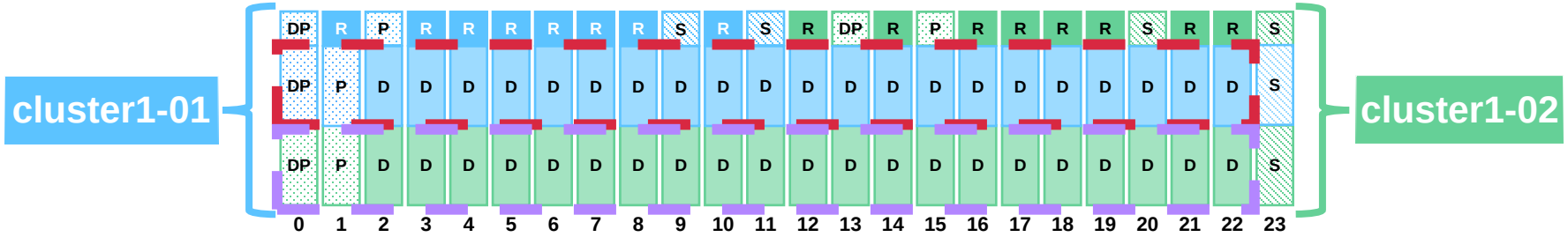
Root-data Advanced Disk Partitioning



- Drives are partitioned into one small root partition and one large data partition.
- Standard aggregate configuration per node is as follows:
  - A root aggregate RAID group of 8 data partitions + 2 parity partitions and 2 spare root partitions
  - A data aggregate RAID group of 9 data partitions + 2 parity partitions and 1 spare data partition
- Total usable capacity is 18 data partitions out of a total of 24, which achieves 75% efficiency.

Root-data-data Advanced Disk Partitioning

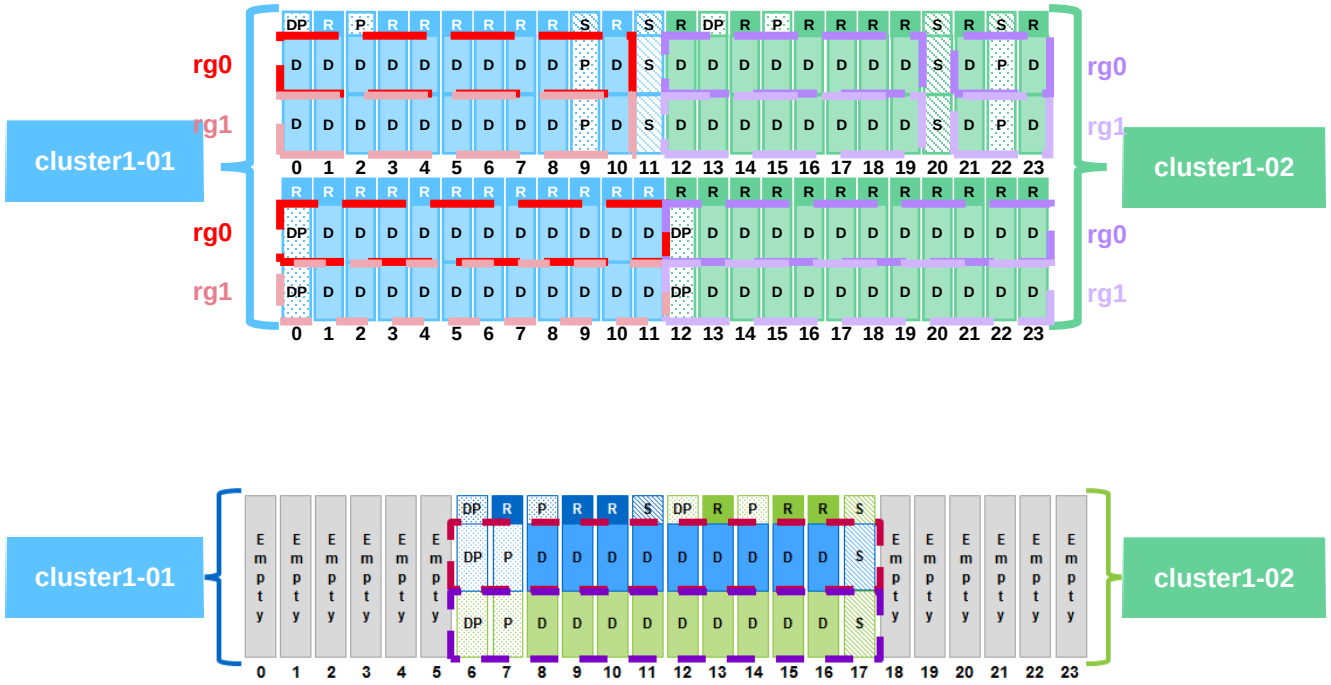
ONTAP 9 and later software



- SSDs are partitioned into one small root and two half-sized data partitions.
- The standard aggregate configuration per node is as follows:
  - A root aggregate RAID group of 8 data partitions + 2 parity partitions and 2 spare root partitions (no change from root-data partition)
  - A data aggregate RAID group of 21 data partitions + 2 parity partitions and 1 spare data partition
- The total usable capacity is 42 data partitions out of a total of 48: 87.5% efficiency, or 16.7% more usable capacity (0.875 / 0.75).

Root-data-data Advanced Disk Partitioning

Additional root-data-data partitioning information

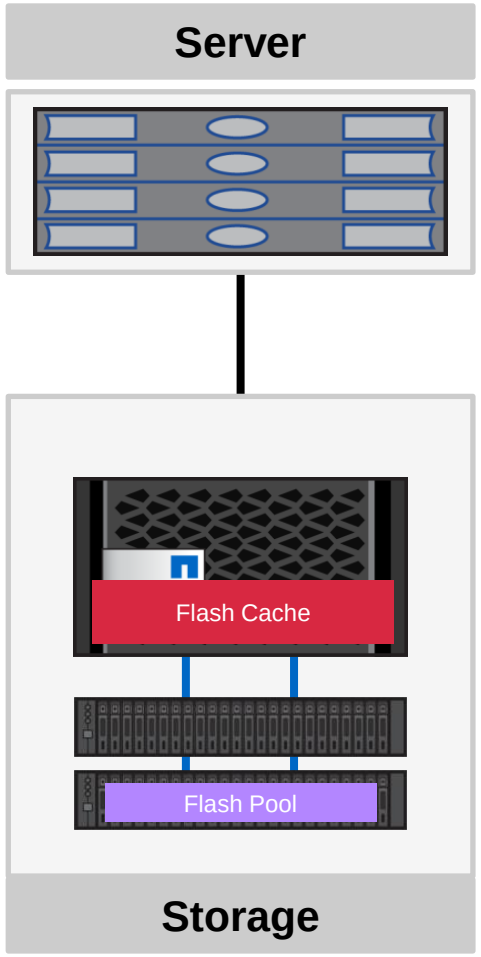


- Root-data-data partitioning is supported on only SSD systems:
  - Default root aggregate provisioning method for AFF systems and SSD-only FAS systems
  - Unsupported on entry-level FAS or AFF MetroCluster FC software
- Data partition assignments with two shelves are like root-data partitioning:
  - Data partitions on an SSD are assigned to the same node.
  - Twice as many RAID groups are used.
- Half-shelf AFF systems have 50% more usable capacity than with root-data partitioning.

Lesson 3
Flash Cache and Flash Pool features

Accelerate I/O performance

- NetApp ONTAP data caching:
  - Is intelligent policy-based caching of data and metadata
  - Delivers high-speed data access for all protocols
  - Maintains deduplicated blocks in the cache
- Flash Cache feature
  - Is controller-based cache that is shared by all volumes on a node
  - Is the best choice for multiple heterogeneous workloads
  - Is simple to deploy and use
- Flash Pool feature
  - Is aggregate-level cache
  - Delivers high-speed data access for specific workloads
  - Provides cached data persistence through failovers
  - Is optimal for database and transactional applications



Create a Flash Pool aggregate

Use the CLI to convert a traditional HDD aggregate into a Flash Pool aggregate.

Provide the following information:

- Existing aggregate name
- Cache source or drive type
- Number of drives
- RAID type (RAID_4 by default)

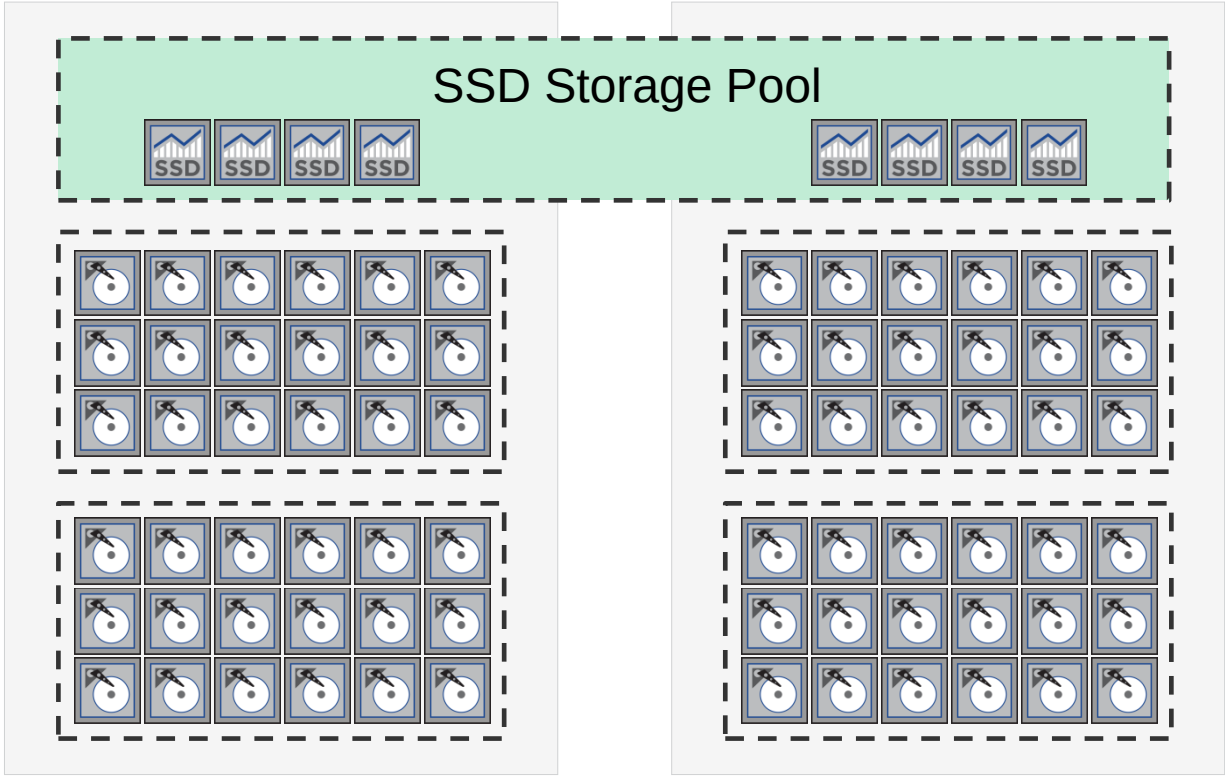
```
::> aggr modify -aggregate cluster2_01_FC_1 -hybrid-enabled true
::> aggr add-disks -aggr cluster2_01_FC_1 -disktype SSD -diskcount 8
```

? More info in Addendum

SSD storage pool for Flash Pool caching

SSDs can be dedicated to a single Flash Pool aggregate or shared through an SSD storage pool.

- Increased storage use for SSDs in Flash Pool aggregates
- Ability to share spares between HA partners
- Better use of SSD performance

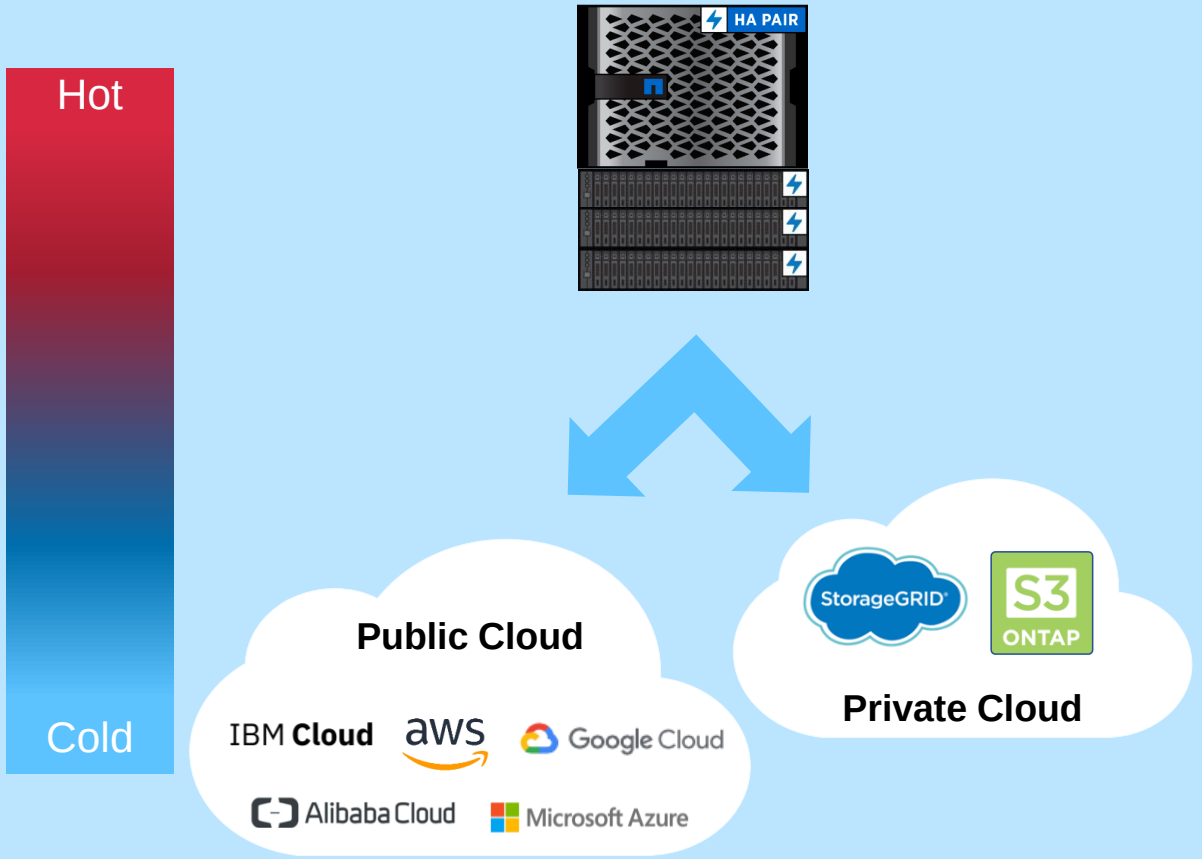


Lesson 4
FabricPool aggregates

FabricPool aggregates

Overview

- What FabricPool aggregates contain:
  - A **performance tier** for frequently accessed (“hot”) data, which is on an all-SSD aggregate
  - A **cloud tier** for infrequently accessed (“cold”) data, which is on an object store
- How FabricPool technology can enhance the efficiency of your storage system:
  - Automatically tier data based on frequency of use
  - Move inactive data to lower-cost cloud storage
  - Make more space available on primary storage for active workloads
  - View how much data in a volume is inactive by using inactive data reporting



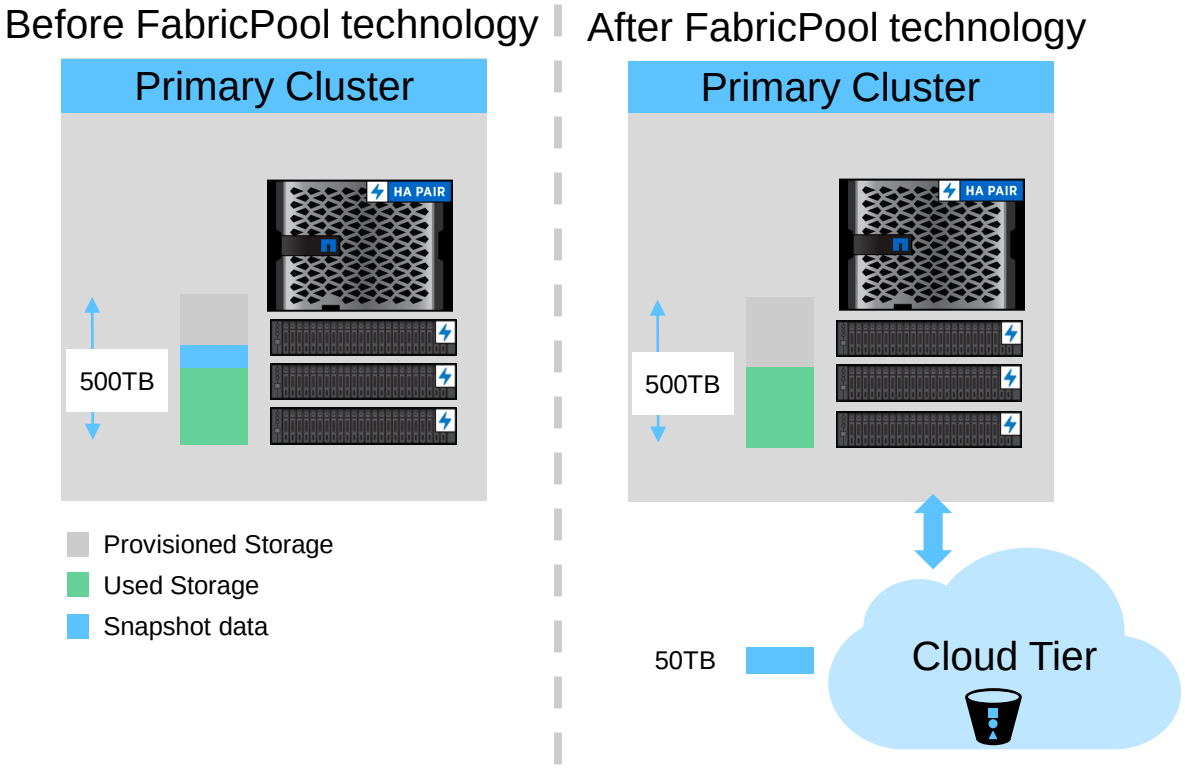
Tiering policies

Define what data is tiered and applied to individual volumes

None	Snapshot-Only	Auto	All
Data always remains in the performance tier.	This policy is the default policy.	This policy moves “cold” data blocks that are held in both Snapshot copies and the active file system.	All active and Snapshot data is written directly to the cloud tier.
There is no cooling period.	“Cold” Snapshot copy blocks that are not shared with the active file system are tiered.	There is a 31-day minimum cooling period.	There is no cooling period.
	There is a 2-day minimum cooling period.		This policy is designed for SnapMirror or SnapVault target volumes.

Make room for active workloads on primary storage

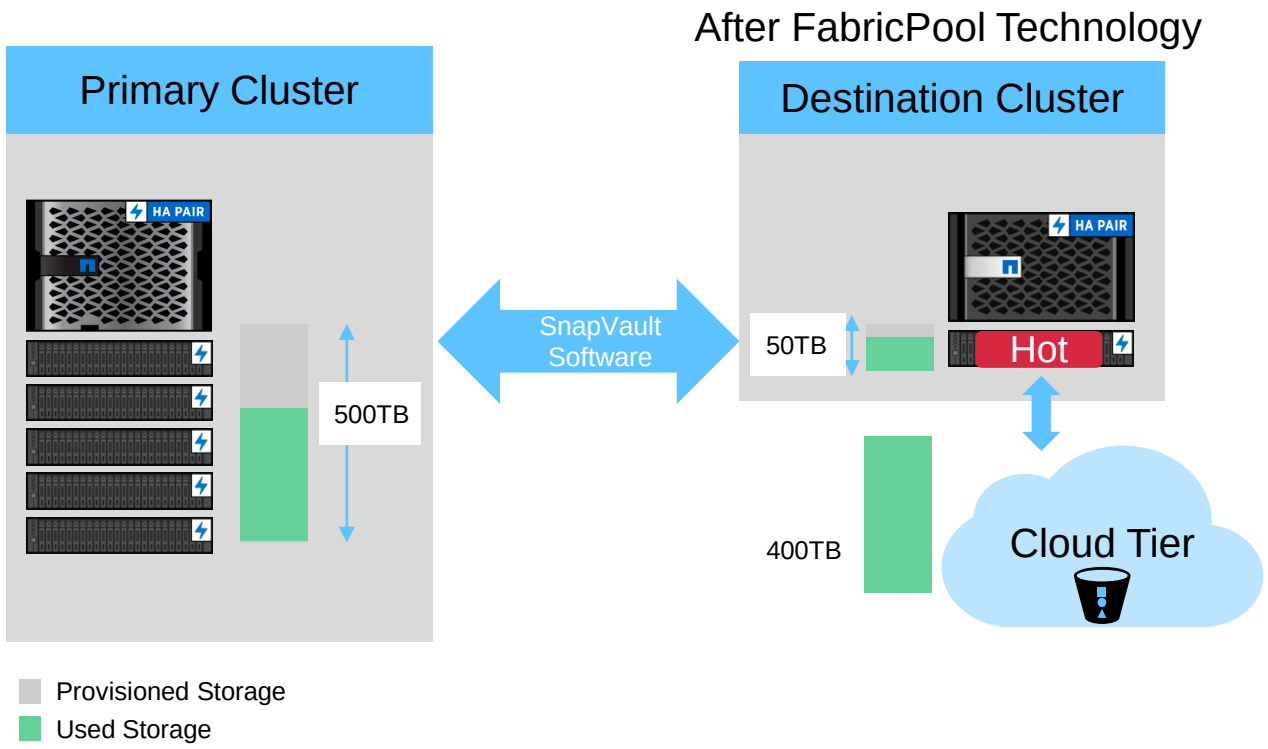
Snapshot-only tiering to the cloud



- In this example, Snapshot copies consume ~10% of used capacity.
- Moving Snapshot data enables active workloads to use the performance drives (SSDs) more effectively.

Shrink your secondary storage footprint

Tier backup data to the cloud

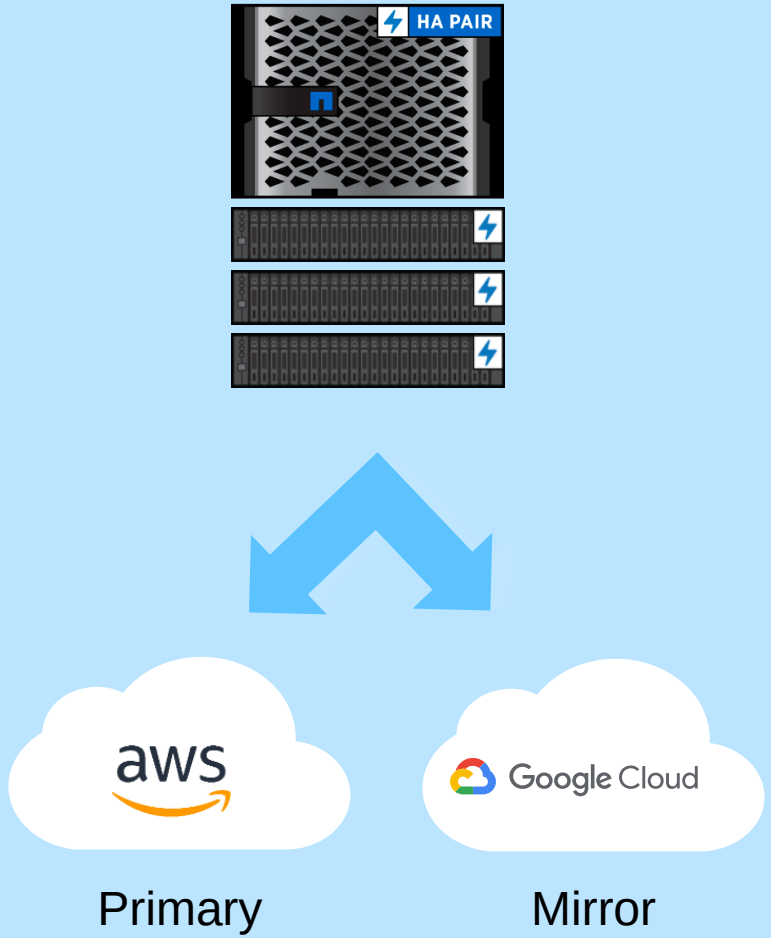


- Expand the capacity of a backup destination cluster by automatically tiering data to the cloud.
- The secondary data center footprint is reduced by up to 90%.
  Hot data (~10-20%) stays on premises, and the remaining 80-90% goes to the cloud object store.
- This method requires no changes to existing data protection policies.
  It works seamlessly.

? More info in Addendum

FabricPool mirrors

- Protect against disaster by synchronously replicating data to two object stores.
- Recover from an outage by promoting the mirror object store to primary.
- Enable easy migration of data between an on-site object store and the cloud and between public cloud providers.



Knowledge check

Module 5: Physical storage management

Knowledge check

Which statement is true of Advanced Disk Partitioning?

a. Both nodes must have a root partition or a data partition assigned.

b. Both nodes must have a root partition assigned.

c. Data partitions can be assigned to any node in a cluster.

d. Root partitions can be assigned to any node in a cluster.

Knowledge check

Which statement is true of Advanced Disk Partitioning?

a. Both nodes must have a root partition or a data partition assigned.

b. Both nodes must have a root partition assigned.

c. Data partitions can be assigned to any node in a cluster.

d. Root partitions can be assigned to any node in a cluster.

Knowledge check

What does a Flash Pool aggregate contain?

a. HDDs only

b. SSDs only

c. HDDs for data storage and SSDs for caching

d. HDDs and SSDs that are all used for data caching

Knowledge check

What does a Flash Pool aggregate contain?

a. HDDs only

b. SSDs only

c. HDDs for data storage and SSDs for caching

d. HDDs and SSDs that are all used for data caching

References

- NetApp Hardware Universe
  http://hwu.netapp.com

- ONTAP 9 Documentation Center
  http://docs.netapp.com/ontap-9/index.jsp
  - ONTAP 9 Disks and Aggregates Power Guide
  - ONTAP 9 System Administration Reference
  - ONTAP 9 Concepts

- TR-4070: NetApp Flash Pool Design and Implementation Guide
  https://www.netapp.com/us/media/tr-4070.pdf

- TR-4598: FabricPool Best Practices
  https://www.netapp.com/us/media/tr-4598.pdf

References
Videos

- ONTAP 9 Feature Overview: FabricPool
https://www.youtube.com/watch?v=5WRe4wkku10
- FabricPool Using OnCommand System Manager 9.5
https://www.youtube.com/watch?v=Fy30d36HxBU
- Cloud Tiering with FabricPool in ONTAP 9.4
https://www.youtube.com/watch?v=RF_qh9LEjzo
- Archiving Volumes with FabricPool
https://www.youtube.com/watch?v=5tDJAkqN2nA

Module summary

This module focused on enabling you to do the following:

- Recognize ONTAP storage architecture concepts
- Manage physical storage resources, including disks, RAID groups, and aggregates
- Create data aggregates
- Create Flash Pool aggregates
- Set up FabricPool aggregates

Complete an exercise

Module 5
Physical storage management

Managing physical storage

Exploring RAID-TEC and creating a flash pool

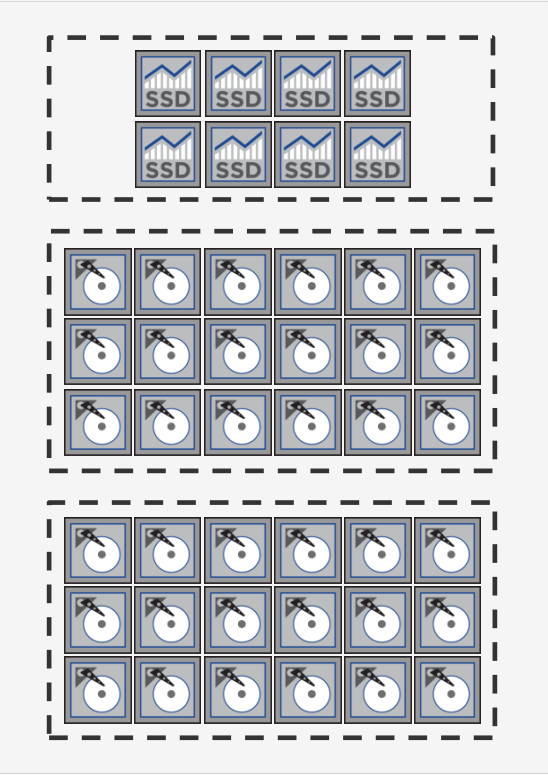
- Access your lab equipment.
- Open your Exercise Guide, Module 5.
- Complete Exercises 1 and 2.
- Share your results.

This exercise requires approximately **40 minutes**.

Addendum
Caching policies

Setting caching policies

- Caching policies determine how data and metadata are cached in Flash Cache modules and Flash Pool aggregates.
- Caching policies can be applied to storage VMs (storage virtual machines, also known as SVMs), volumes, LUNs, or files.
- The caching policy should be changed only if a different policy provides better performance for your workload.
- If the wrong caching policy is configured, volume performance might degrade severely, and performance degradation could increase gradually over time.




```
cluster1::> volume modify -vserver svm1 –volume vol1 –caching-policy random_read
```

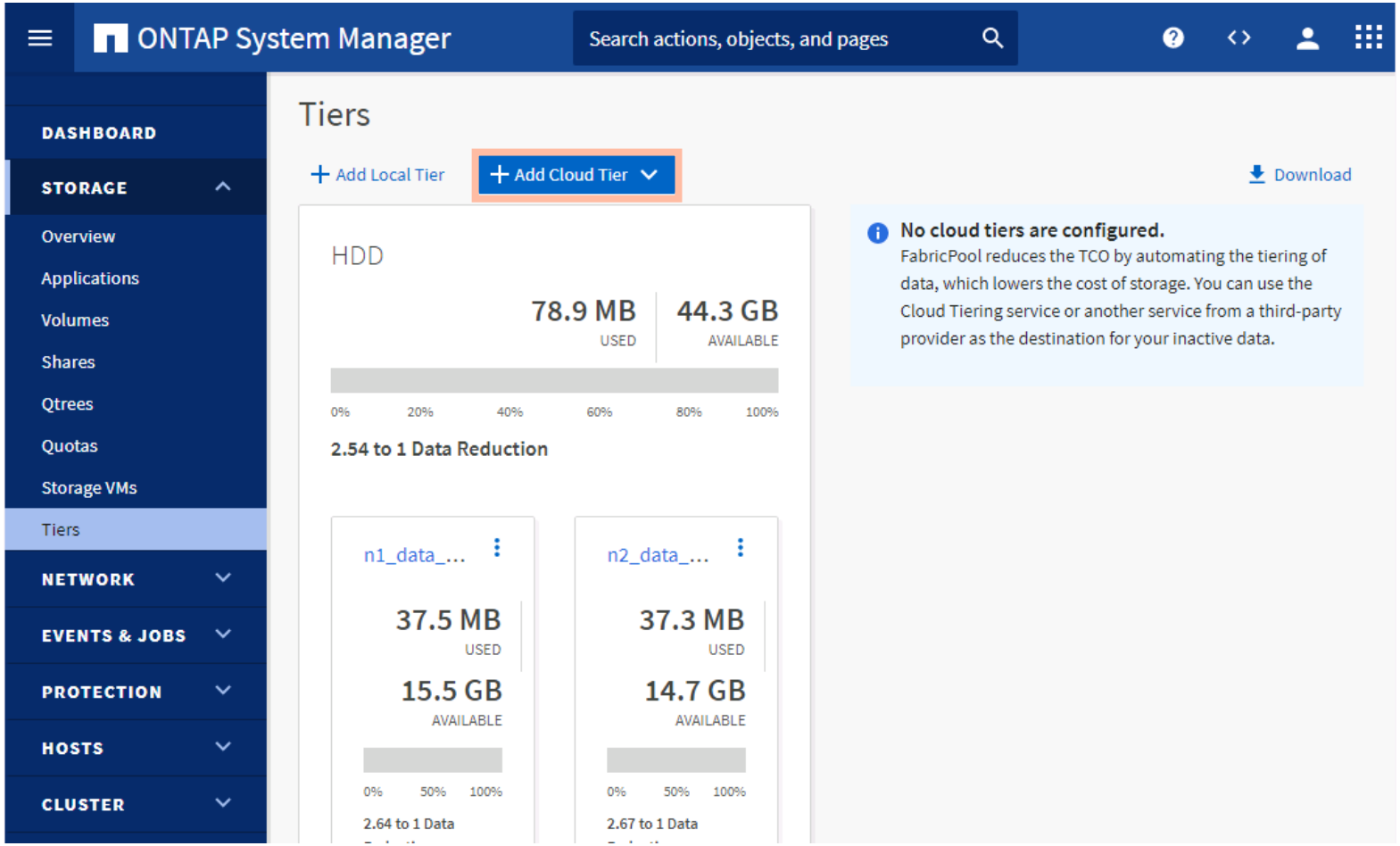

Caching policies

Policy Name	Insertions Using Read Caching Policy				Write Insertions
	Random Reads	Sequential Reads	Random Writes	Sequential Writes	
auto	Yes	No	No	No	Yes
none	No	No	No	No	No
random_read	Yes	No	Yes	No	No
noread-random_write	No	No	No	No	Yes
meta	Metadata only	No	No	No	No
meta-random_write	Metadata only	No	No	No	Yes
random_read_write	Yes	No	Yes	No	No
random_read_write- random_write	Yes	No	Yes	No	Yes
all_read	Yes	Yes	No	No	No
all_read-random_write	Yes	Yes	Yes	No	Yes
all	Yes	Yes	Yes	Yes	No
all-random_write	Yes	Yes	Yes	Yes	Yes

Addendum
FabricPool in ONTAP System Manager

FabricPool technology in ONTAP System Manager

Adding cloud tiers

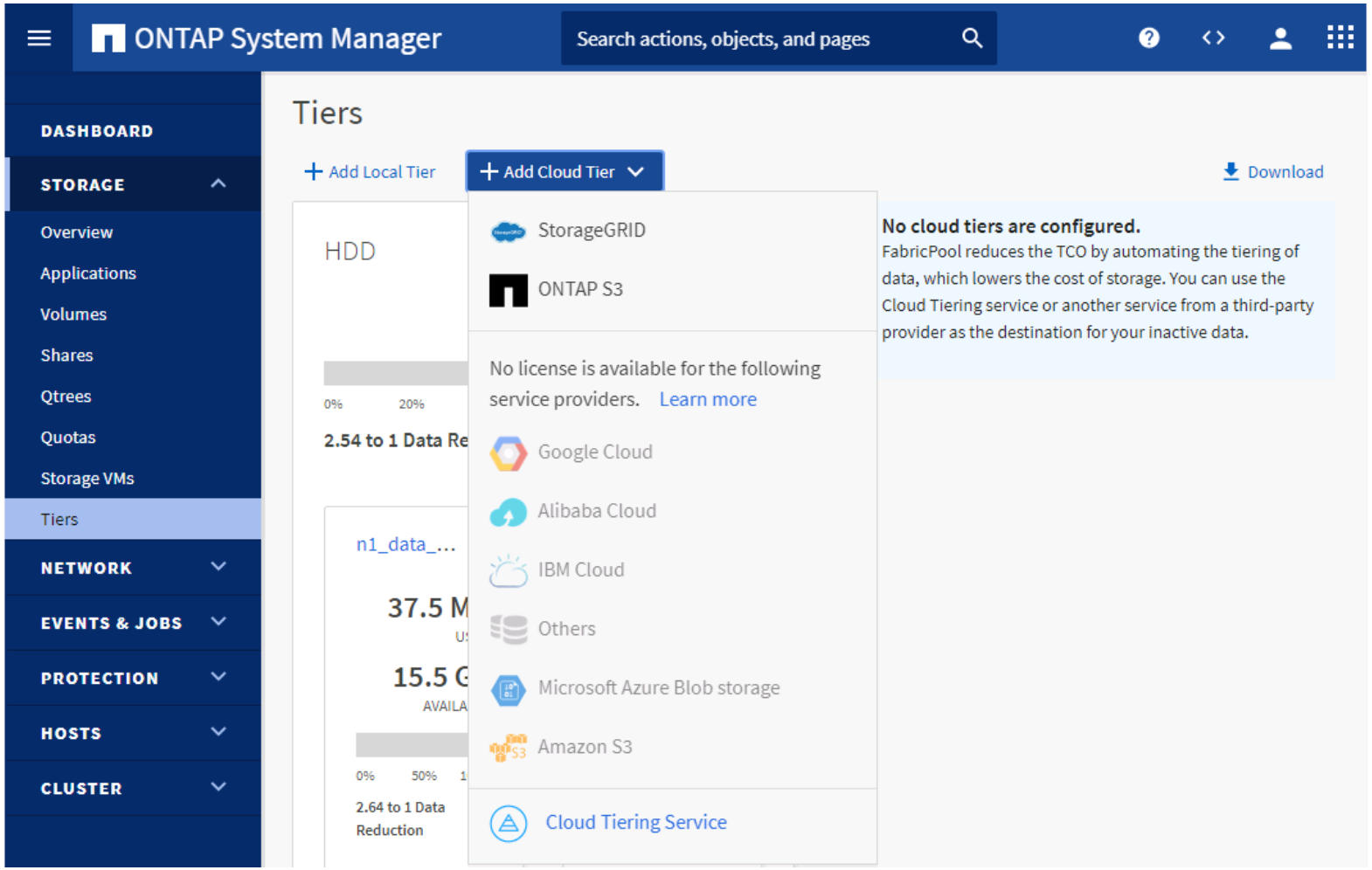


FabricPool technology in ONTAP System Manager

Add an external capacity tier

Select the object store provider for the cloud tier.

Tiering ONTAP volumes to another ONTAP system, to a StorageGRID solution, or to NetApp Cloud Tiering service does not require an ONTAP FabricPool license.

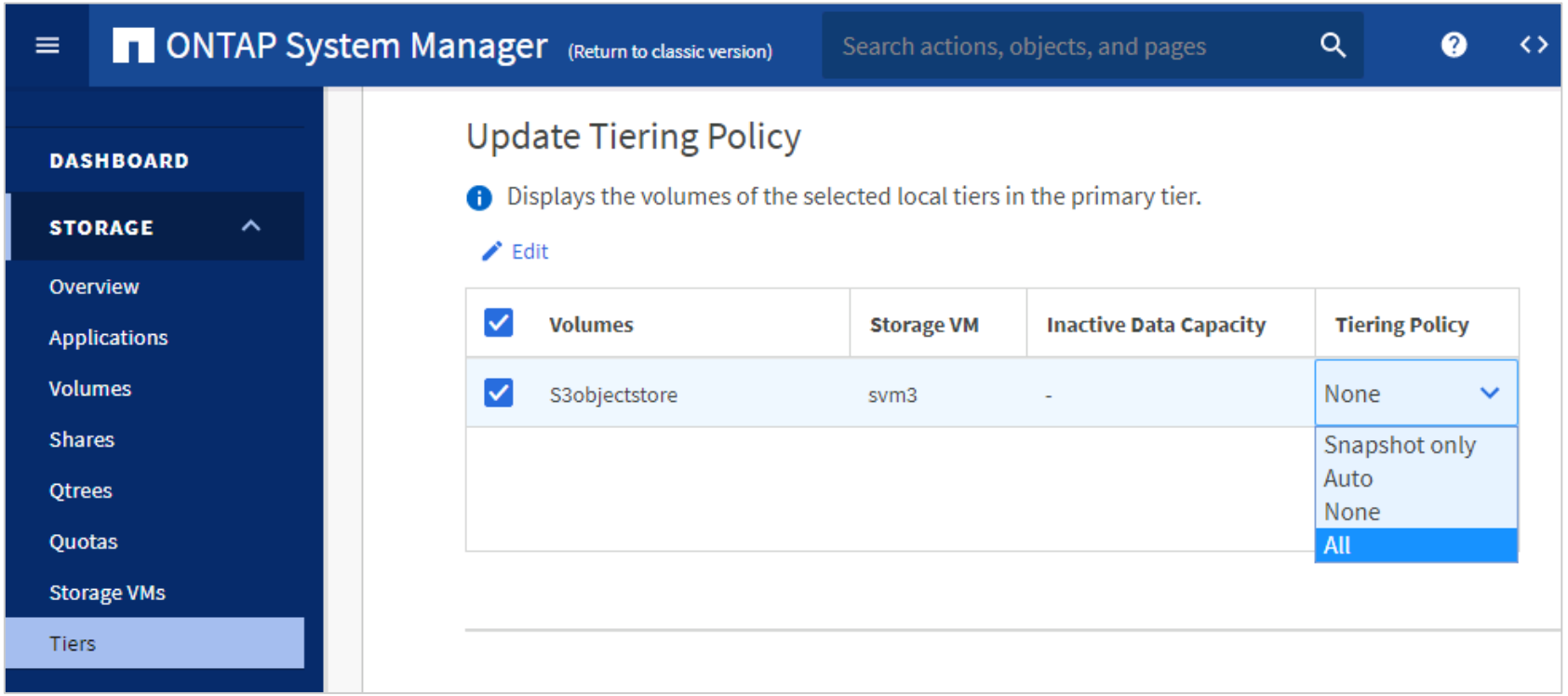


FabricPool technology in ONTAP System Manager

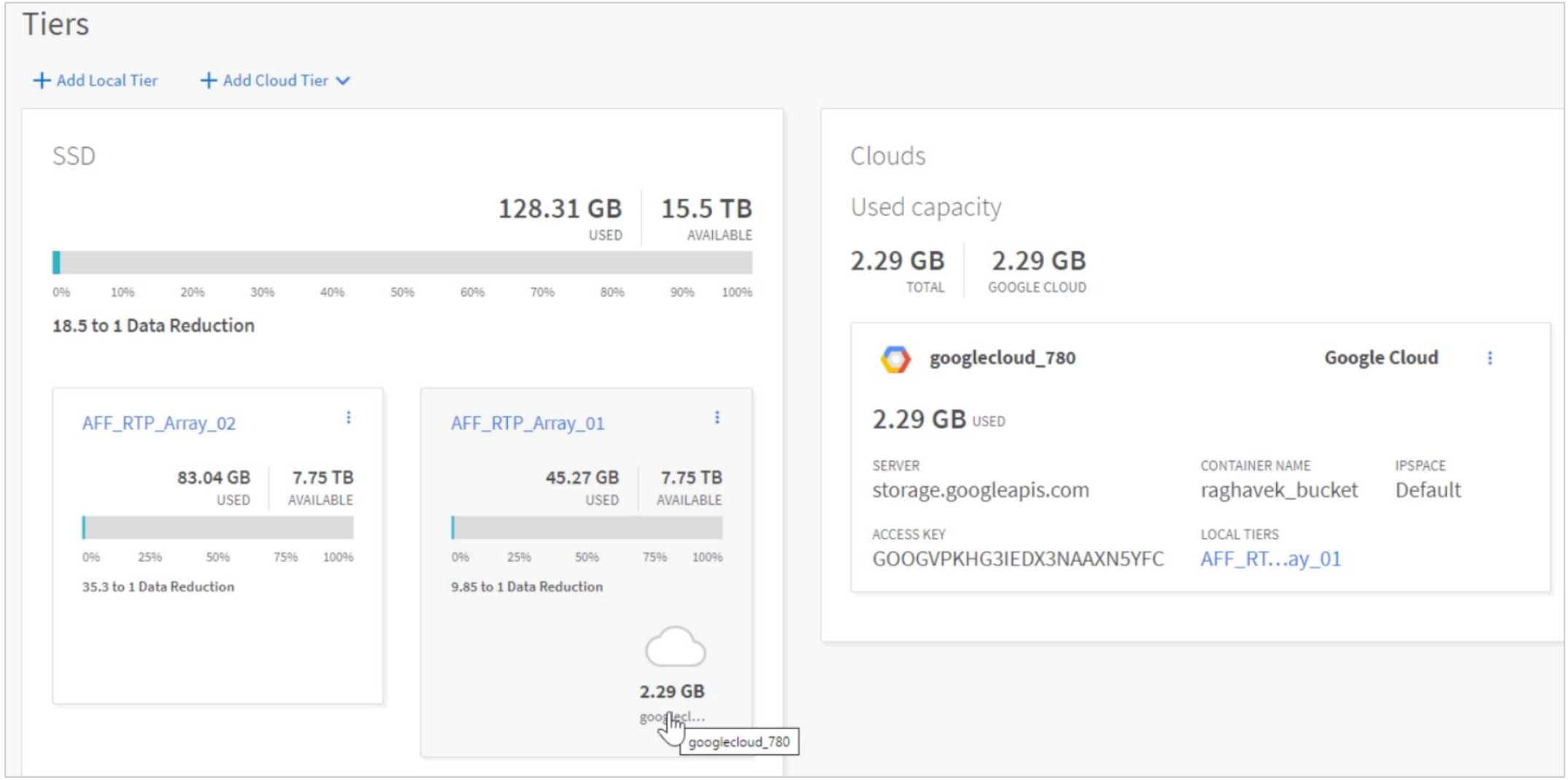
Volume tiering policy

When you add a cloud tier, be aware of the following:

- You should update the volume tiering policy for any volumes that exist in the aggregate.
- The default policy for existing volumes is None.
- Changing the tiering policy of a volume causes data to be migrated to the cloud tier.



FabricPool technology in ONTAP System Manager



Volume tiering policy

When volumes are created on a FabricPool-enabled aggregate, be aware of the following:

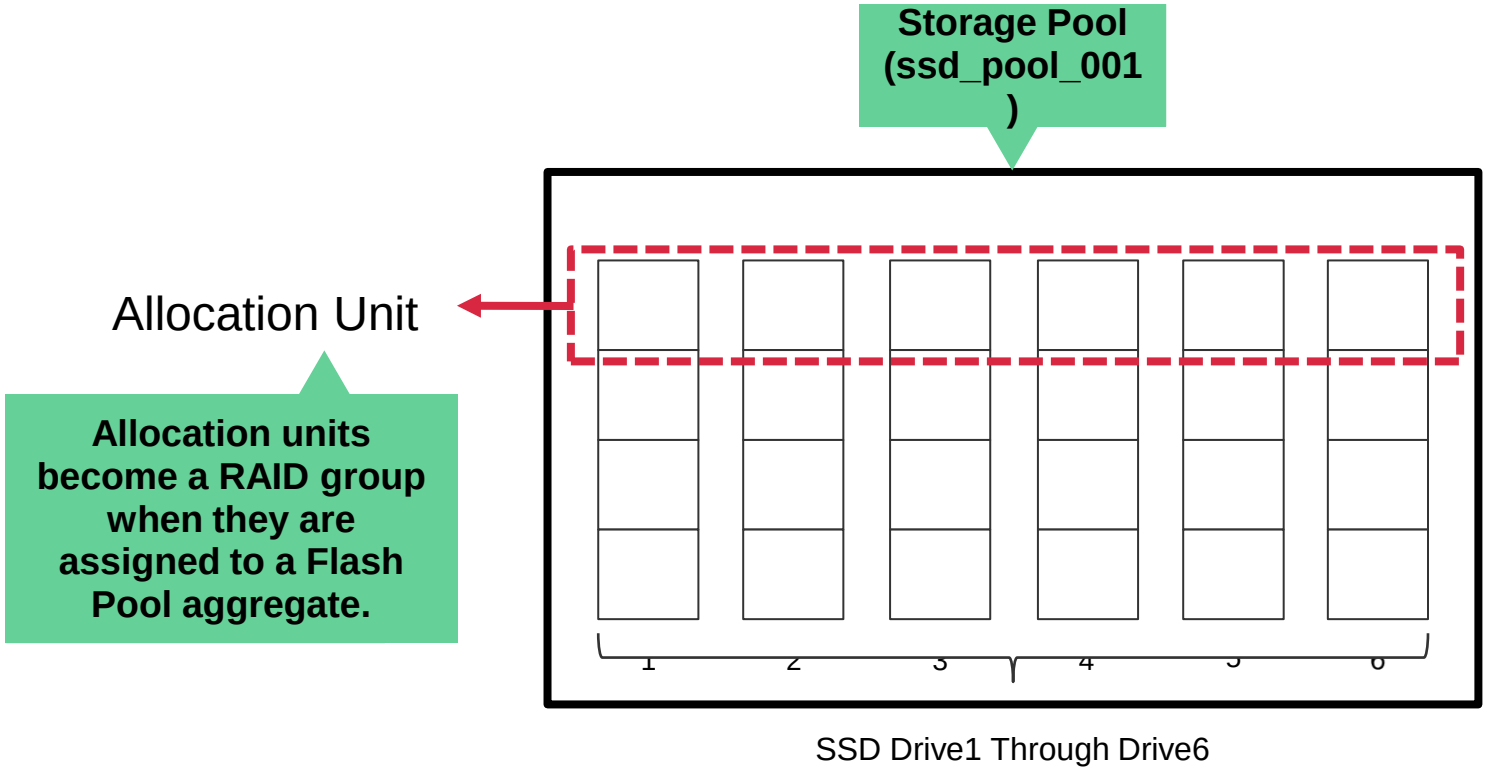
- You should select a tiering policy. The default policy is Snapshot-only.
- Changing the tiering policy of a volume after creation might cause data to be migrated to the cloud tier.
- You can change the cooling off period for only volumes with the Snapshot-only or auto tiering policy.

```
cluster1::> volume modify -vserver svm1 –volume vol1 –tiering-policy auto
–tiering-minimum-cooling-days 15
```


Addendum
Flash Pool SSD storage pools

SSD partitioning for Flash Pool cache

Creation



Create an SSD storage pool

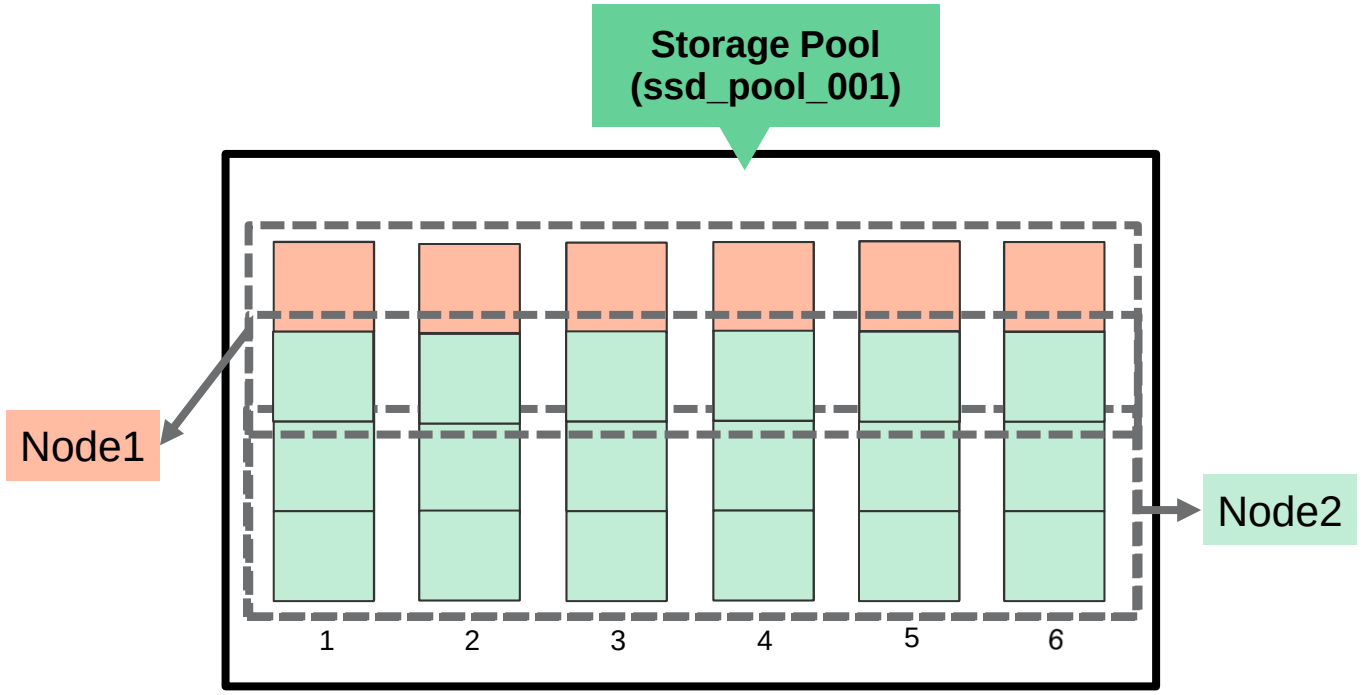
Provide the following information:

- Storage pool name
- Number of drives
- Size of SSDs from the HA pair
  (if multiple sizes are available)

```
::> storage pool create -storage-pool ssd_pool_001 -disk-count 3
```


SSD partitioning for Flash Pool cache

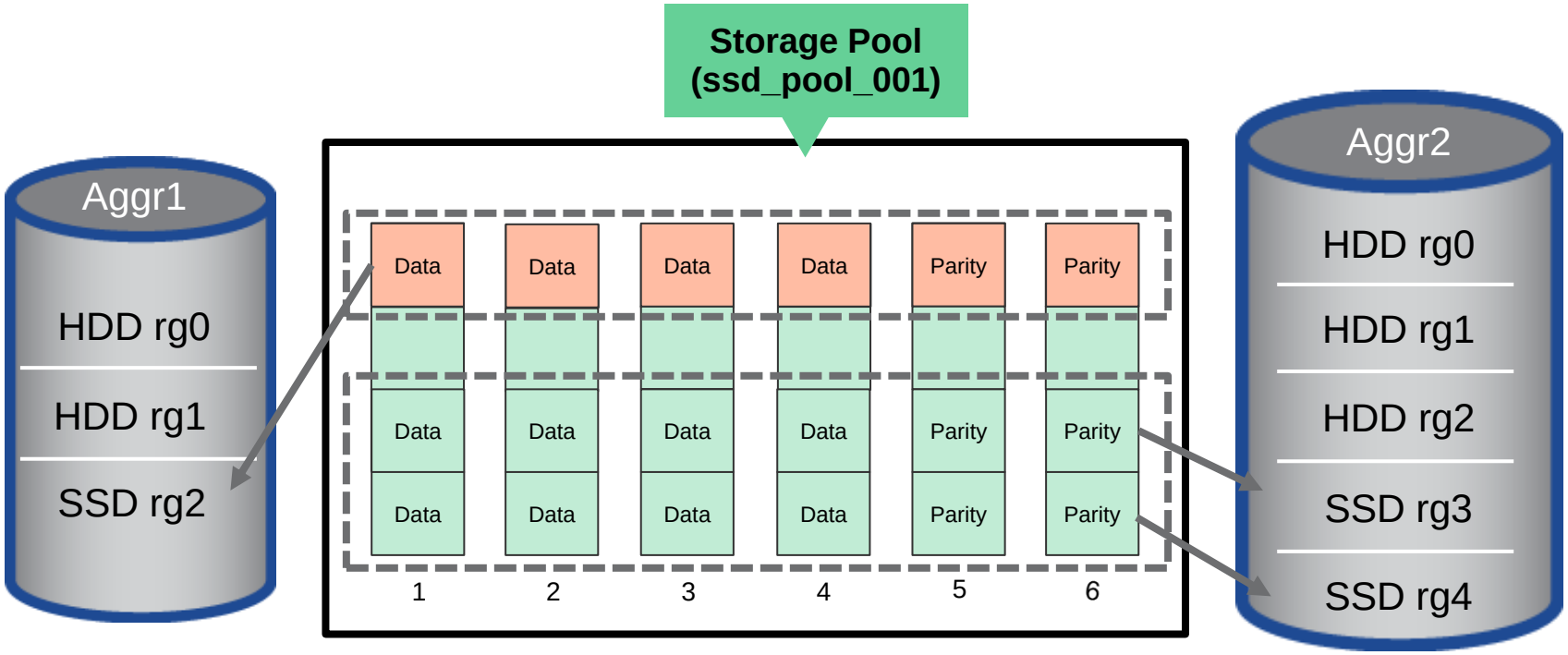
Ownership




```
::> storage pool reassign -storage-pool ssd_pool_001
    -from-node cluster1-01 -to-node cluster1-02 -allocation-units 1
```

SSD partitioning for Flash Pool cache

Reassignment



Create a Flash Pool aggregate that uses an SSD storage pool

Use the `storage aggregate add-disk` command to create a Flash Pool aggregate that uses an SSD storage pool.

Provide the following information:

- An existing aggregate name
- A storage pool name
- The number of allocation units to add

```
::> storage aggregate add-disks -aggregate rtp01_fcal_002
    -allocation-units 1 -storage-pool ssd_pool_001
```

