Technical Report

SANtricity OS Dynamic Disk Pools

Feature description and best practices

Charles Binford and Mitch Blackburn, NetApp

November 2024 | TR-4652

Abstract

With NetApp® Dynamic Disk Pools (DDP), storage administrators can group sets of like disks into a pool topology in which all the drives in the pool participate in the I/O workflow. This technology provides faster drive rebuilds than with RAID 5 or RAID 6 and removes the complications of RAID group configurations, so that storage administrators can focus on capacity allocation. This technical report provides a detailed description of the DDP feature.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Overview

NetApp Dynamic Disk Pools (DDP) technology represents a significant advancement in storage system data protection and management. As disk capacities continue to grow without corresponding increases in data transfer rates, traditional RAID rebuild times are getting longer, even up to several days. Slow rebuilds result in much more time with degraded performance and exposure to additional disk failures.

With five issued patents, DDP technology is designed to deliver worry-free storage through effortless management and self-optimization while maintaining predictable performance under any conditions, including recovery from drive failures. With rebuild times that are up to four-times faster than previous methods, DDP technology significantly reduces exposure to multiple cascading disk failures, providing excellent data protection.

The following list identifies the key DDP attributes that enable these benefits:

- Simplified management:
  - Distributed hot spare capacity (known as preservation capacity) eliminates the need for dedicated idle hot spare drives.
  - You can add drives to a pool without reconfiguring the RAID level.
  - The protection scheme and stripe size are automatic; you do not need to configure them.
- Predictable performance:
  - A deterministic algorithm dynamically distributes data, spare capacity, and protection information across a pool of drives.
  - If a drive fails, segments are recreated elsewhere, which reduces the magnitude and duration of the performance disruption.
  - The large pool of drives reduces hot spots.
- Reduced exposure to multiple disk failures:
  - Through segment relocation, the system returns to an optimal state faster.
  - Through prioritized reconstruction, any stripes that experience multiple drive failures are given the highest priority.

This technical report provides a high-level overview of DDP technology and best-practice guidelines for using pools.

Intended use

This information is for NetApp customers, partners, and OEMs.

Technical overview

With DDP technology, NetApp SANtricity OS and management software allows you to create pools in addition to traditional volume groups (generally referred to as RAID groups). A pool can range in size from a minimum of 8 drives to as large as all the drives in a storage system, which is up to 480 drives in the NetApp E5700 system. Pools can consist of either hard disk drives (HDDs) or solid-state drives (SSDs). In addition, pools and volume groups can coexist in the same system. Pools consisting of HDDs must have at least 11 drives and only support RAID 6 volumes.

As with volume groups, pools and pool volumes can also be configured through the CLI or through the REST API. Two configurations are only available with SSDs and if configured through the REST API:

- RAID 1 volumes in pools using less than 11 drives

- RAID 1 volumes in the same pool as RAID 6 volumes

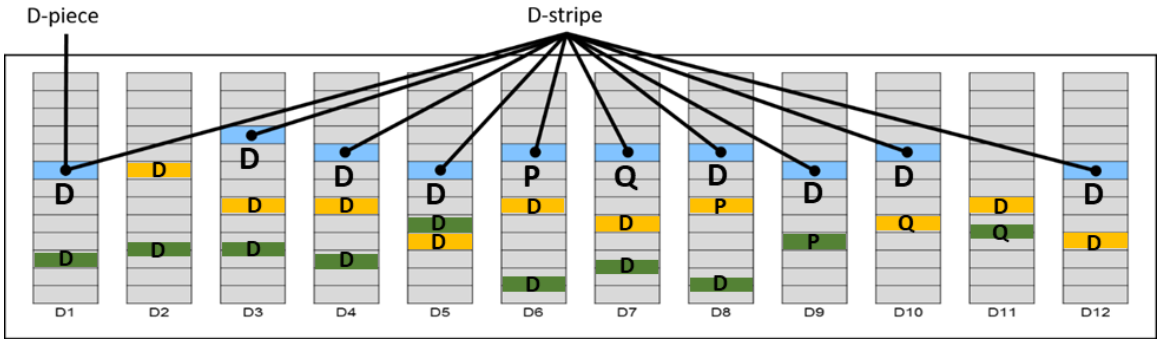
An SSD based pool can be created with as few as 8 SSD drives. If a pool contains 8-10 drives, then the volumes created on that pool can only be RAID 1 in a 3+3 configuration. If a pool of less than 11 drives is expanded to 11 or more drives, the existing RAID 1 3+3 volumes keep their 3+3 configuration, but new volumes created on the larger pool will be either RAID 6 8+2 or RAID 1 5+5.

RAID 1 and RAID 6 volumes can be mixed and matched as needed with SSD pools. RAID 1 has more parity overhead, less protection against data loss, and better performance for writes when compared to RAID 6.

Data layout

Within a pool, volume data is distributed across all drives, regardless of how many drives are assigned to the pool. A volume comprises many virtual stripes, known as D-stripes. Each D-stripe resides on either 6 or 10 drives, depending on the RAID level, that are distributed throughout the pool by an intelligent optimization algorithm. The portion of each D-stripe that resides on a single drive is called a D-piece. Each D-piece is a contiguous section of the physical drive. Figure 1 shows an example of how a D-stripe might be laid out for a RAID 6 volume in a pool. In this case, the pool consists of 12 drives, but even if it had more, the D-stripe would still only be divided into 10 D-pieces. Note that the D-pieces do not necessarily reside in the same portion of each drive. More information on D-stripes and D-pieces and their capacities can be found in the glossary of terms.

Figure 1) D-piece and D-stripe example for RAID 6 volume.



Each D-stripe consists of 8,192 RAID volume stripes on NVMe drives and 4,096 RAID volume stripes on SAS drives. Each volume stripe is composed of eight 128KiB data segments (RAID 6), for a total of 1MiB. As shown, eight of the segments are data (D), one is parity (P), and one is the RAID 6 Q value.

Note: As of SANtricity OS 11.90 each D-stripe consists of 8,192 RAID volume stripes on SAS drives. This is to accommodate large capacity NL-SAS drives such that the sum of capacities of all pools in the system (maximum pool capacity) is increased to 12PiB. There will be no mixing of the D-stripe sizes. A system that starts out with D-stripes of 4,096 RAID volume stripes will continue to create new DDP with that same size despite upgrading the SANtricity OS to 11.90. To reuse old drives and get 8GB D-stripes, all existing DDP using 4GB D-stripes would need to be deleted and recreated.

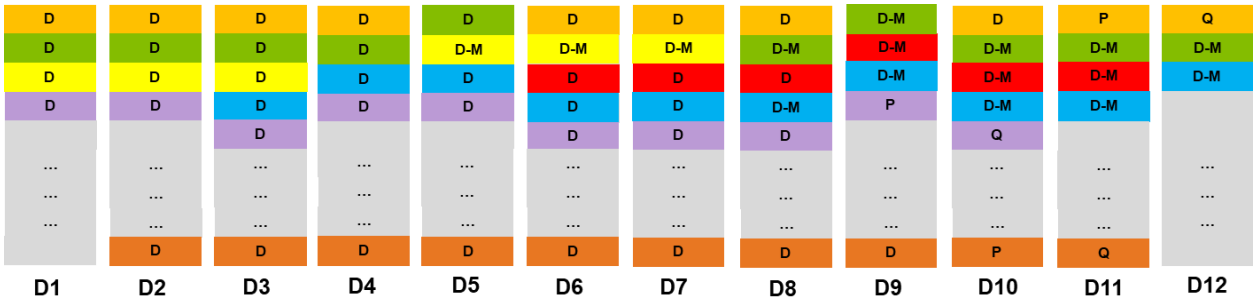
After a pool has been created, a volume can be created within the pool. This volume consists of some number of D-stripes spread across all the drives within the pool, with the number of data D-stripes equaling the defined volume capacity divided by the D-stripe size. For example, a 2TiB RAID 6 volume consists of 256 data D-stripes on NVMe drives and 512 data D-stripes on SAS drives. Allocation of D-stripes for a given volume is performed starting at the lowest available range of logical block addresses (LBAs) for a given D-piece on a given drive.

Unlike a volume group configuration with a fixed number of drives, the number of drives in a Dynamic Disk Pool can change without “breaking” the pool, it’s just smaller or larger. When adding a drive to say a

100-drive pool, 1% of the data from the existing drives will be dynamically rebalanced on to the new drive. When a drive is lost, those effected segments are regenerated by all the drives and again dynamically rebalanced across the remaining drives. This all happens automatically, and because it is a function of all the drives in the pool reading and writing, the recovery time is significantly decreased (hours vs. days) as is the performance impact compared to a volume group configuration where one previously idle spare is being exposed to the rest of the volume group writing to that single drive. For a RAID 1 volume in a pool, there is a mirror set of drives. So, for a pool of 8 to 10 drives the RAID 1 volume in the pool is laid out in a 3+3 pattern of D-pieces and D-stripes requiring three data drives and three parity drives. For eleven drives or more a 5+5 pattern of D-pieces and D-stripes requiring five data drives and five parity drives is used.

RAID 1 3+3 volumes, RAID 1 5+5 volumes, and RAID 6 volumes can all coexist in a pool on the same set of drives. An example is shown in Figure 2.

Figure 2) Multiple pool configurations existing on a single set of drives.

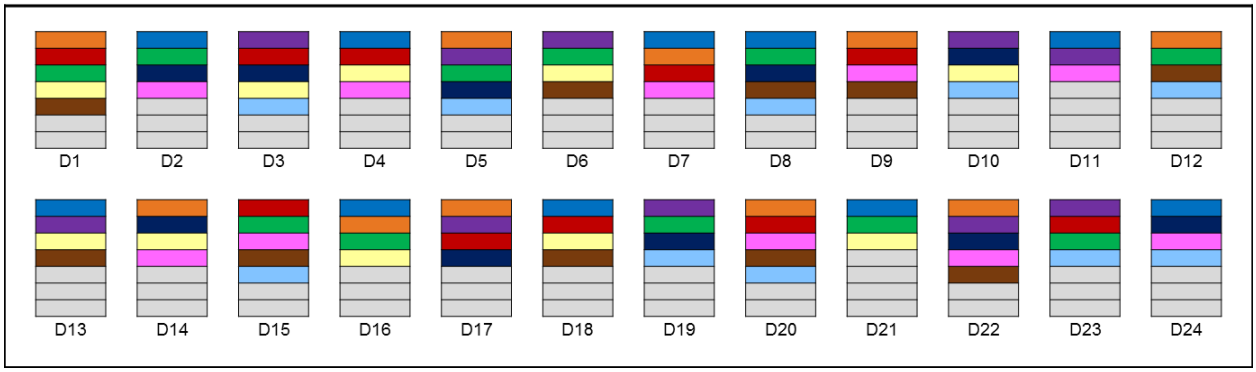


Operation when a drive fails

A major benefit of DDP technology is that, rather than using dedicated stranded hot spares, the pool itself contains integrated preservation capacity to provide rebuild locations for potential drive failures. This feature simplifies management because you no longer need to plan for or manage individual hot spares. It also greatly improves the time of rebuilds and enhances the performance of the volumes themselves during a rebuild.

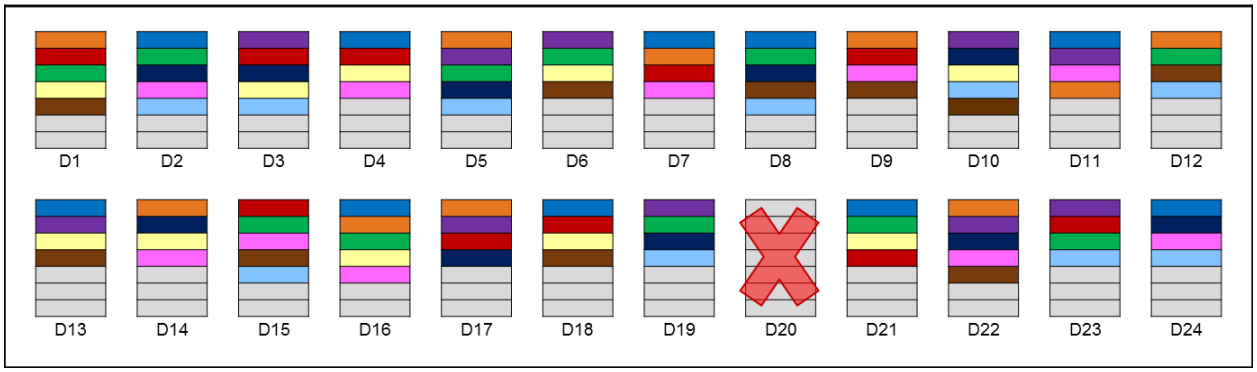
To begin discussion of DDP operation when a drive fails, consider the 24-drive pool that is depicted in Figure 3. Each different color in the diagram represents a D-stripe, each of which contains 10 D-pieces of the same color. The D-pieces are distributed over the pool by the DDP intelligent algorithm, as previously noted.

Figure 3) 24-drive pool.



Now suppose that one of the drives in the pool fails, as in Figure 4.

Figure 4) 24-drive pool with one drive that has failed.



When a drive in a pool fails, the D-pieces from the failed drive are reconstructed segment by segment, using the same mechanism that is normally used by RAID. The intelligent algorithm chooses other drives in the pool on which to write the rebuilt D-pieces, confirming that no single drive contains two D-pieces from the same D-stripe. The individual D-pieces are reconstructed at the lowest available LBA range on the selected drive.

In Figure 4, drive D20 has failed, and its D-pieces have been rebuilt and written to other drives. The rebuild operation runs in parallel across all drives. Because multiple drives participate in the effort, the overall performance effect of this situation is reduced, and the length of time that is needed to complete the operation is also dramatically reduced.

Multiple drive failures

Based on the striping previously described, there exist critical D-pieces when more than one drive fails. For example, drives D8 and D20 in Figure 3 have two D-pieces in common, the brown and the light blue. If these two drives fail, then the brown and the light blue D-pieces are critical to be restored before another drive fails. Since only these failed D-Pieces in these two D-stripes must be restored before another drive failure occurs, the time of exposure to data loss is minimized. If D11 fails, there are D-pieces that are on D8 and D20 that are also on D11. If these D-pieces have not been moved yet, then these are critical D-pieces and must be moved immediately, before another drive fails.

Note: If no D-pieces are shared between the three failed drives, then there are no critical D-pieces.

Note: For a RAID 1 DDP any D-Stripe with a failed drive is considered critical. You can have only one drive fail at a time. Once two drives fail then volumes will fail.

To minimize data availability risk, if multiple drives fail within a pool, any D-stripes that are considered critical are given priority for reconstruction. This approach is called critical reconstruction. After critically affected D-stripes are reconstructed, the rest of the necessary data is then reconstructed.

From a controller resource allocation perspective, there are two user-modifiable reconstruction priorities within the pool:

- Degraded reconstruction priority is assigned for instances in which only a single D-piece must be rebuilt for the affected D-stripes. The default priority for this instance is high.
- Critical reconstruction priority is assigned for instances in which a D-stripe has two missing D-pieces (one missing D-piece for RAID 1 DDP) that must be rebuilt. The default priority for this instance is highest.

For very large pools with two simultaneous disk failures, only a relatively small number of D-stripes are likely to encounter the critical situation in which two D-pieces must be reconstructed. As discussed previously, these critical D-pieces are identified and reconstructed initially at the highest priority. This approach returns the pool to a degraded state very quickly so that further drive failures can be tolerated.

As an example, assume that a pool of 192 drives has been created and has two drive failures. In this case, it is likely that the critical D-pieces would be reconstructed in less than one minute and, after that minute, an additional drive failure could be tolerated. From a mathematical perspective, with the same 192-drive pool, only 5.2% of D-stripes would have a D-piece on one drive in the pool and only 0.25% of the D-stripes would have two D-pieces on those same drives. Therefore, only 48GiB of data must be reconstructed to exit the critical stage. A very large pool can continue to maintain multiple sequential failures without data loss until there is no additional preservation capacity to continue the rebuilds.

After the reconstruction, the failed drive or drives can be subsequently replaced, although replacement is not specifically required. Fundamentally, this replacement of failed drives is treated in much the same way as an online capacity expansion of the pool. Failed drives can also be replaced before the pool exits from a critical or degraded state.

Drive rebuilds

Figure 5 and Figure 6 illustrate the difference in rebuild times between RAID 6 and DDP technology. Figure 5 shows data from an E2800 system with various numbers of HDDs, while Figure 6 presents similar data for SSDs. Both were run with the rest of the system in an idle I/O state. These charts show that a pool rebuilds faster than a RAID 6 volume group, and, as the pool spindle count increases, DDP rebuild times go down compared to RAID 5 and RAID 6 rebuild times.

Figure 5) Example of time to rebuild an HDD.

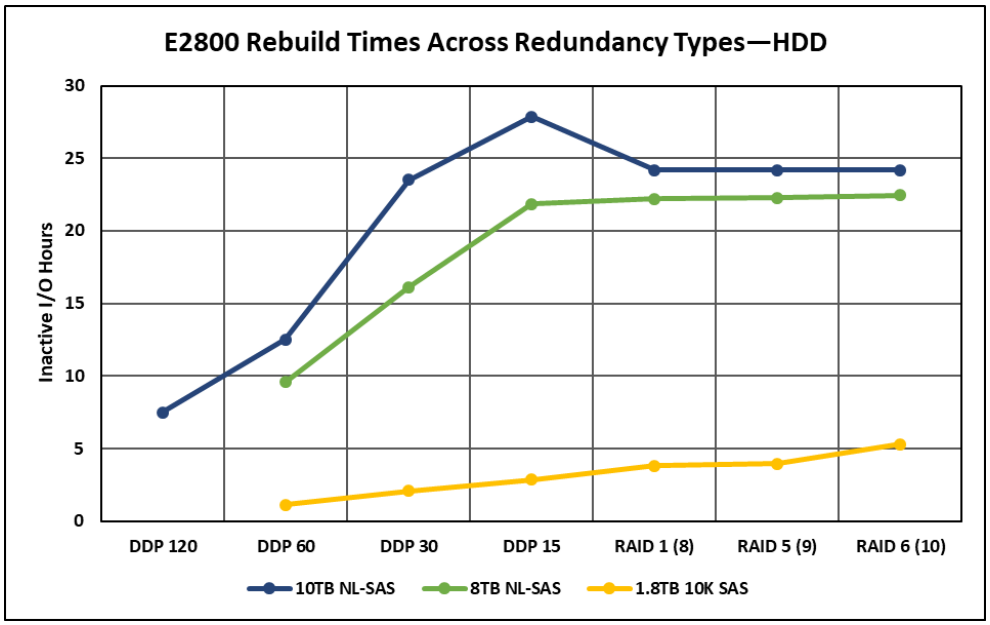
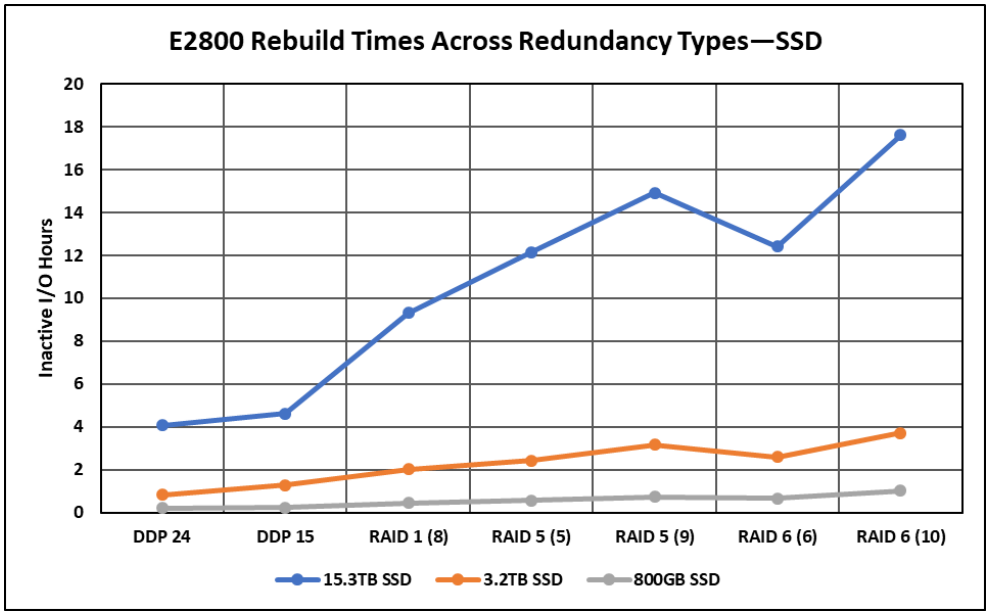


Figure 6) Example of time to rebuild an SSD.



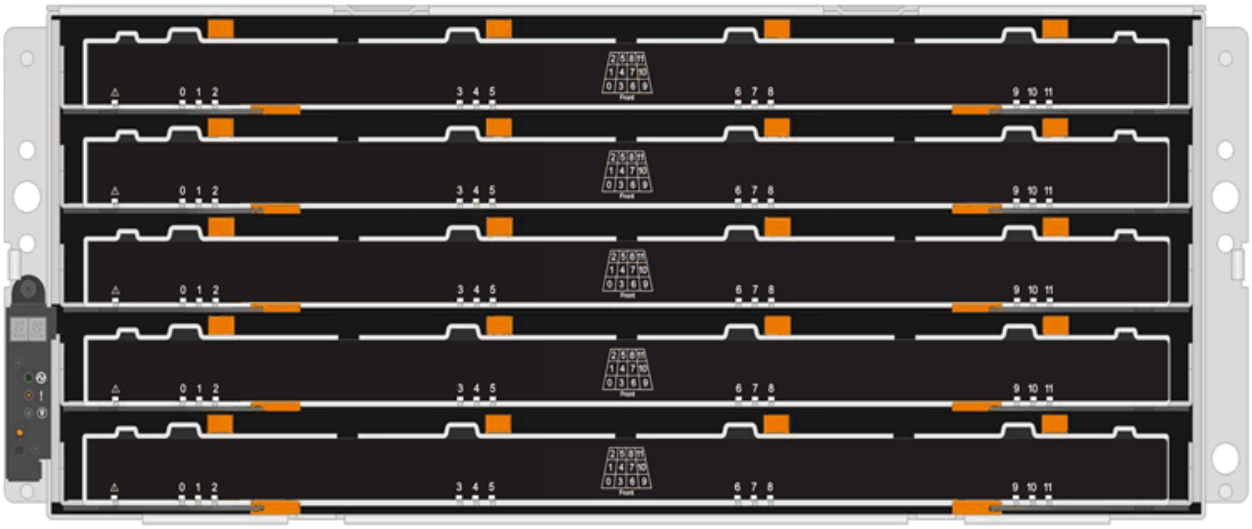
Shelf loss protection

The pool must include drives from at least five shelves and there must be an equal number of drives in each shelf. Shelf loss protection is not applicable to high-capacity shelves (4U60); if your system contains high-capacity shelves, refer to Drawer Loss Protection.

Drawer loss protection

The NetApp 4U60 drive shelf holds up to 60 drives in 4U of rack space. The drives are organized into five drawers, with each drawer containing up to 12 drives. See Figure 7.

Figure 7) NetApp 4U60 drive shelf.



As of SANtricity OS 11.25, it is possible to achieve drawer loss protection (DLP) within a single 4U60 shelf. Drawer loss protection refers to the ability of a pool to withstand the loss of an entire drawer and still maintain I/O operations with no loss of availability.

To enable DLP with a single shelf, the configuration must have at least 15 drives, equally distributed among the drawers. SANtricity System Manager presents DLP candidates during pool creation. The number of drives in these candidates is always a multiple of five. When you add drives to increase pool capacity, you should add them in groups of five, one per drawer to maintain DLP with the new capacity.

DLP can also be enabled for systems with multiple 4U60 shelves. In this case, drives with similar characteristics (drive type, capacity, data assurance, and security) must be distributed equally across all drawers so that proper candidates are presented. All DLP candidates that are presented through the management interfaces consist of an equal number of drives per drawer for all drawers in the pool.

In some complex configurations, it might be necessary to create a smaller DLP pool initially. This situation can occur if similar drives are not equally distributed among the drawers because some of the drawers include other drive types. In that case, you can expand the pool by adding five drives at a time to maintain DLP.

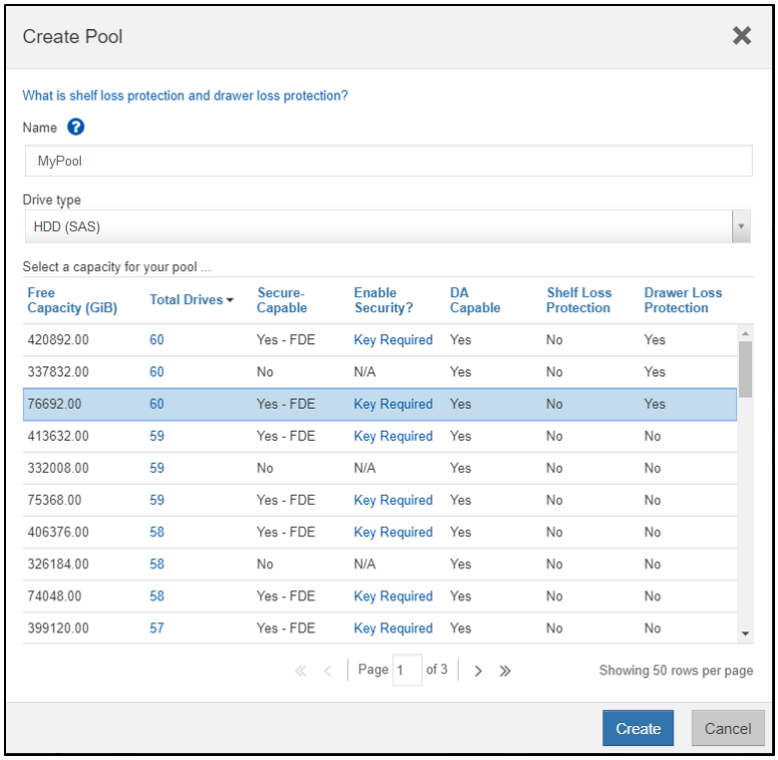
Management

Configuring pools is simpler than configuring traditional volume groups. You do not have to choose RAID levels, segment sizes, or global hot spares because they are determined by system defaults. The primary decisions for the administrator are what type of drives, how many drives, and whether the pool will have security or DLP.

Note: The administrator can also choose specific drives for the DDP but only through the REST API or SMcli.

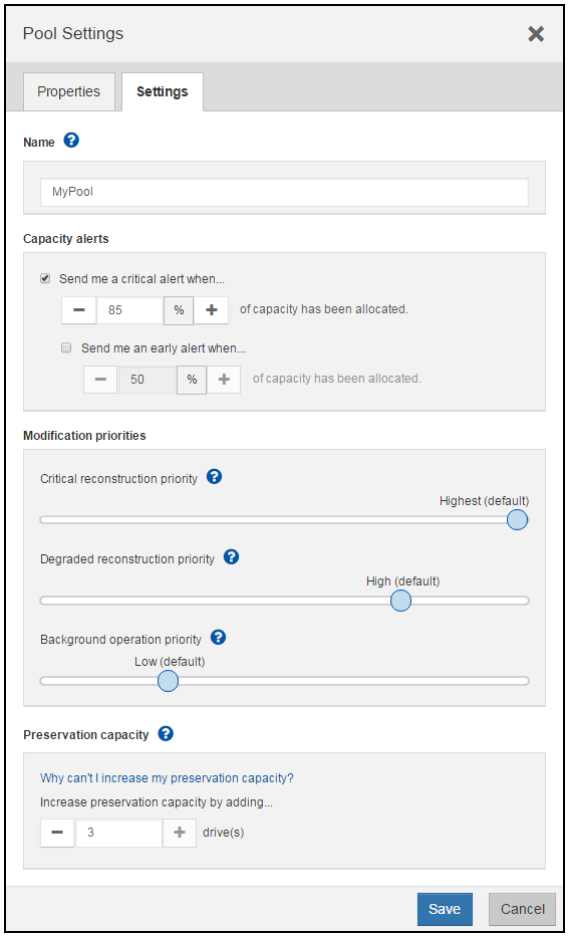
Figure 8 shows pool creation in SANtricity System Manager. In this case, the administrator has selected a candidate that has 60 HDDs that offer both DLP and full disk encryption (FDE) security.

Figure 8) Create Pool in SANtricity System Manager.



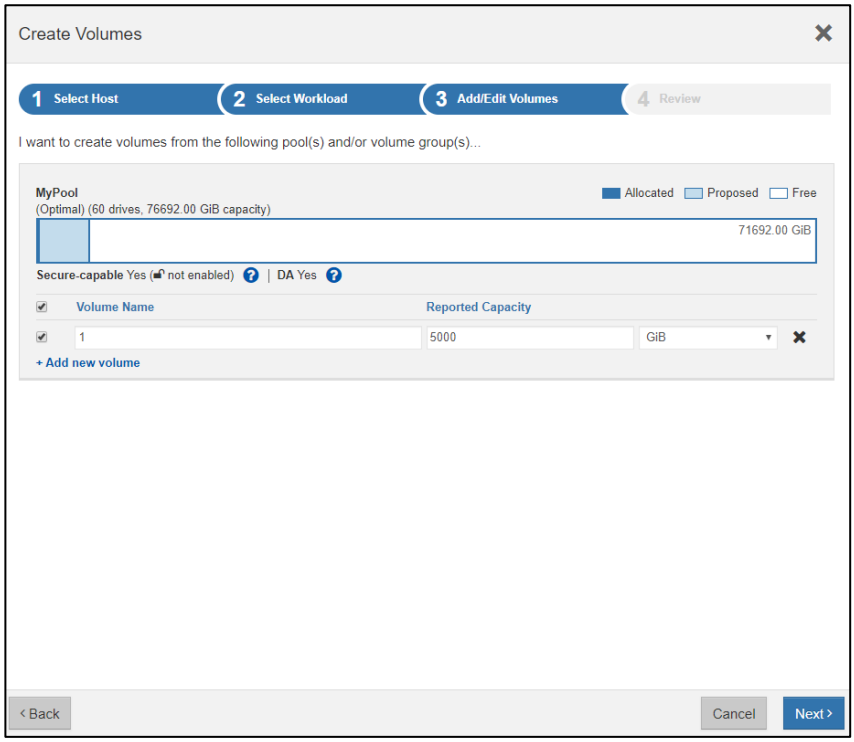
You can change pool settings, but changes are generally not required. Figure 9 shows various pool settings in SANtricity System Manager. The default settings for reconstruction and other background operations are set here. The administrator can also change the preservation capacity equivalent number of drives. In this case, the default is 3, but the user can select any number from 0 up to a maximum of 10 drives or 20% of the drives, whichever is less.

Figure 9) Pool Settings in SANtricity System Manager.



After you create a pool, you can create volumes within the pool in much the same way as with traditional volume groups, as seen in Figure 10.

Figure 10) Create Volumes in SANtricity System Manager.



As with volume groups, pools and pool volumes can also be configured through the CLI or through the REST API.

Addition of larger capacity drives

It can become necessary to use larger capacity drives in your DDP than what was originally installed. Assume the pool had 11 x 8TB drives when created. When new drives are required, if 8TB drives are no longer available you must install larger capacity drives such as 12TB drives. The 12TB drives will now be downsized to 8TB to work with the drives that already make up the pool.

Now assume you want to recapture the capacity that is unused in the 12TB drives. The only way to do that is to replace all the 8TB drives with 12TB drives. If even one 8TB drive exists as part of the pool the additional capacity of the 12TB drives will be unavailable. When the last 8TB drive is replaced with a 12TB drive, then all the capacity of the 12TB drives will become available for use.

At this point, existing volumes can be dynamically expanded into the additional capacity, or new volumes can be created.

REST API features

The following additional configurations are only available if configured through REST API (not available on the UI):

- Create RAID 1 volumes in the same pool as RAID 6 volumes (11 drive minimum).
  - SSD only (not available for HDD pools).
  - Mix and match RAID 6 / RAID 1 volumes as needed.
  - RAID 1 volumes are 5+5 (five data, five parity per stripe, compared to eight data, two parity for RAID 6).
  - RAID 1 has more PARITY overhead than RAID 6 (50% versus 20% for RAID 6 8+2).

    - RAID 1 has less protection against data loss than RAID 6.
    - A single drive failure creates critical stripes.
    - RAID 1 has lower latency (better performance) for writes.
- An eight-drive pool minimum.
    - SSD only (not available for HDD pools).
    - Pools with drive counts between eight and ten (that is, less than the historical DDP minimum of eleven) can only be RAID 1 in a 3+3 configuration. This allows automatic rebuild to spare capacity in the event of a drive failure.
    - If a small pool is expanded to 11 or more drives, the existing 3+3 volumes stay at their 3+3 configuration. New volumes created on the larger pool are the usual RAID 6 8+2 or RAID 1 5+5.

Comparison of DDP and volume groups

Pools and the volumes within them allow several operations that are like operations in traditional volume groups and offer some features that are unique to DDP technology.

Table 1 shows a brief feature comparison between DDP and volume groups.

Table 1) Comparison between DDP and volume groups.

Feature	DDP and pool volumes	Volume groups and volume group volumes
NetApp Snapshot™ technology	Yes	Yes
Volume copy	Yes	Yes
Synchronous mirroring	Yes, except for EF600 and EF300	Yes, except for EF600 and EF300
Asynchronous mirroring	Yes	Yes
Dynamic volume expansion	Yes	Yes
Online capacity expansion and reduction	Yes, add or remove up to 60 drives at a time (11.41)	Partial, add a maximum of 2 drives, no reduction
Dynamic RAID migration	No	Yes
Dynamic segment sizing	No, segments are always 128KiB	Yes
Hot spare	No, uses distributed preservation capacity	Yes, dedicated global hot spare or spares
Dynamic redistribution	Yes, nondisruptive background operation	No, fragmentation and stranded capacity due to deletions
Drive evacuator	Yes	Yes
Shelf and drawer loss protection	Yes	Yes
SSD support	Yes	Yes

Configuration Guidelines

Table 2 lists several important considerations for when you configure pools.

Table 2) DDP configuration guidelines.

Description	Configuration
Maximum number of pools per system	20

Description	Configuration
Minimum number of drives per pool	8 RAID 1 3+3, 11 RAID6 or RAID 1 5+5
Minimum number of drives per pool for drawer loss protection (DLP) with a single 4U60 shelf*	15 **Note:** Available in SANtricity OS 11.25 or later
Maximum pool capacity** (sum of capacities of all pools in the system)	Prior to SANtricity OS 11.40.1: • E2800—2PiB • E5700—2PiB As of SANtricity OS 11.40.1: • E2800—6PiB • E5700—6PiB As of SANtricity OS 11.60.1: • EF300—12PiB • EF600—12PiB As of SANtricity OS 11.90: • All systems—12PiB
Maximum volume size	Prior to SANtricity OS 11.30: • All systems—64TiB As of SANtricity OS 11.30: • All systems—2PiB As of SANtricity OS 11.50: • All systems—4PiB
Default preservation capacity by pool size (number of equivalent drives)	• 11 drives: 1 • 12–31 drives: 2 • 32–63 drives: 3 • 64–127 drives: 4 • 128–191 drives: 6 • 192–255 drives: 7 • 256–384 drives: 8 • 385–480 drives: 10
RAID 1 3+3 default preservation capacity	The RAID 1 3+3 configuration requires a minimum of 8 drives and a maximum of 10. • 8 – 10 drives: 2
Drive types supported **Note:** All drives in a pool must be of the same type and have the same characteristics (data assurance, security). To avoid losing the capacity of larger drives, all drives should have the same capacity.	SAS, NL-SAS, SSD, NVMe
Online addition to or removal from a pool	Up to 60 drives at a time

* To maintain DLP as capacity is added to a disk pool, drives must be added in groups of five, one drive added per drawer in a 4U60 shelf.

** Maximum pool capacity includes RAID protection overhead, pool preservation capacity, usable capacity, and a small DDP-specific reserve based on the size of the pool.

Performance

DDP configurations are generally not the highest-performing configuration. However, with flash-based pools the delta when compared with RAID 5 and RAID 6 varies from insignificant with 100% read workloads to slightly less performance when running DDP with 100% write workloads. Figure 11 provides a general performance comparison using an EF570 array running 16KB and 64KB I/O as the workloads change from 100% write to 100% read.

Figure 11) Performance with 16KB and 64KB I/O as the workload varies from 100% write to 100% read.

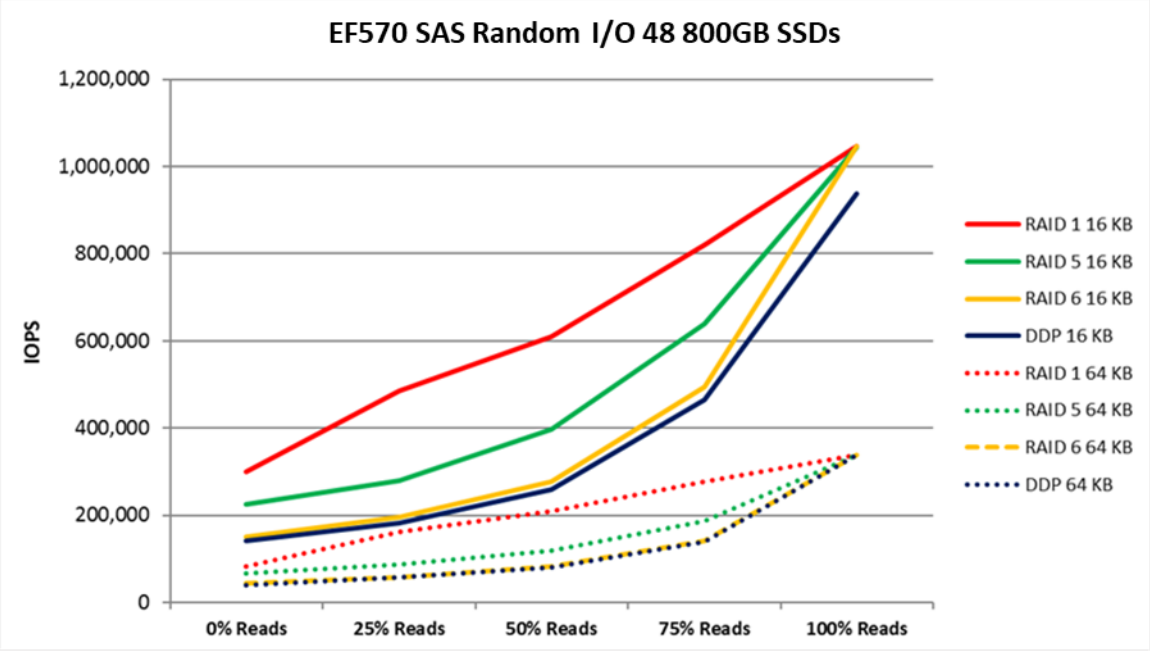
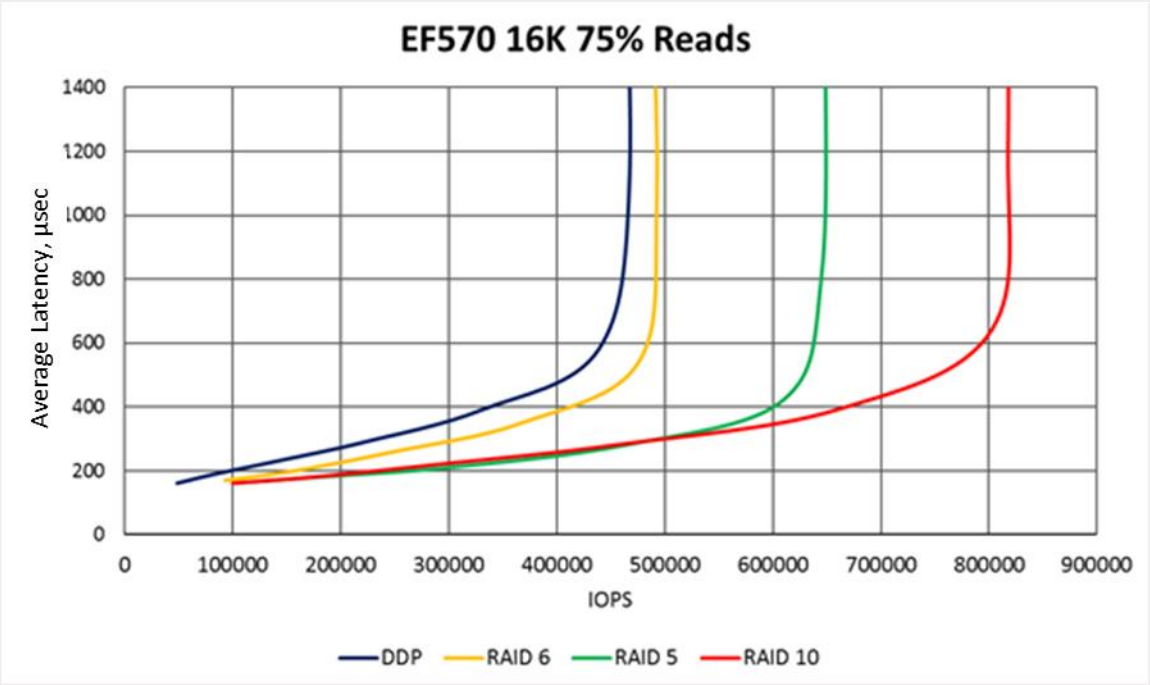


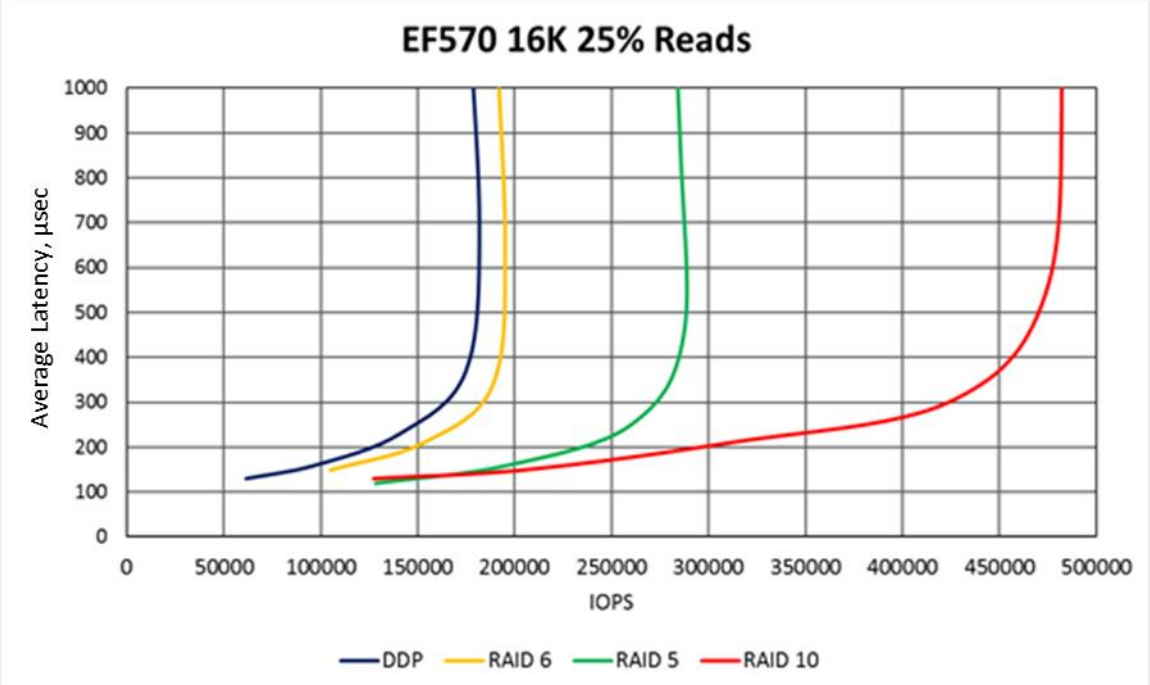
Figure 12 shows a comparison of latency between RAID 6 DDP, RAID 5, RAID 6, and RAID 10 by using a 16KB I/O size and a 75% read workload.

Figure 12) DDP to RAID 5 and RAID 10 latency comparison with a 75% read, 16KB I/O workload.



As the write component increases in the workload, the IOPS drop as expected, but the latency remains low, as shown in Figure 13.

Figure 13) DDP latency compared to RAID 5, RAID 6, and RAID 10 using a 25% read, 16KB I/O workload.



These results suggest that, when using DDP technology, IOPS and latency performance, especially at the lower end of the performance range, is very close to other standard RAID choices. At the high end of the performance range, standard RAID offers some performance advantage.

In most cases, performance comes down to a small trade-off of top-end performance for a significant improvement in drive rebuild time that is offered by pools of 30 drives or more.

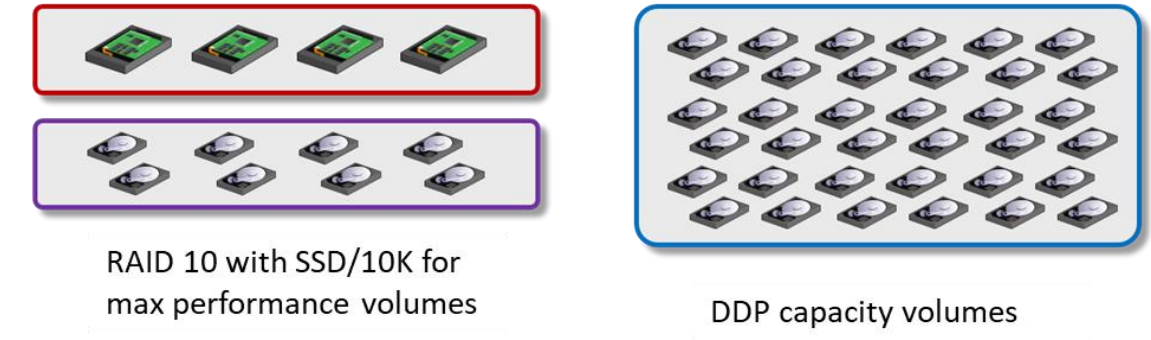
Best practices for configuring pools

In general, it is a best practice to use DDP for homogeneous environments or for hosting more than one application on a single storage system. DDP technology is designed to perform best in a random workload environment, and several applications on a single storage system produce a random profile to the storage array.

For environments with one or more bandwidth-hungry applications that use the storage system, such as streaming media with HDDs, configuring one traditional volume group per application is the best practice.

For storage systems with a mix of SSDs, high-performance HDDs, and NL-SAS drives that are used for both high-performance and high-capacity workloads, you might want to configure a mix of volume groups and pools. See Figure 14 for an example.

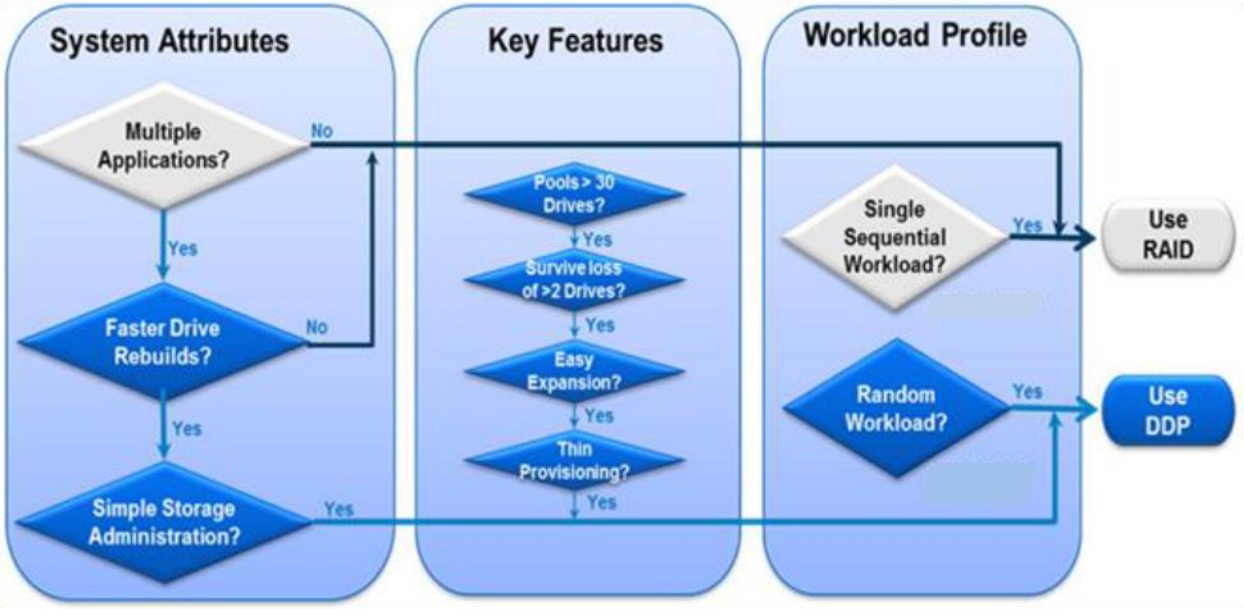
Figure 14) Mixing volume groups and pools.



Choosing between DDP technology and traditional RAID

DDP technology offers several key features that you should consider when you select a storage virtualization technology. Figure 15 compares DDP features with traditional RAID features.

Figure 15) Selecting DDP technology or RAID.



Note: If a double drive failure occurs, DDP technology rebuilds critical segments first, allowing the system to survive another drive failure within minutes of the initial double drive failure.

How large should the pool be?

From an architectural standpoint, a pool can be expanded up to the total number of disks in the system, assuming that all drives are of the same type. Expansion is limited only by the maximum pool capacity (shown in Table 2) for the specific array model and software version. However, the best practice in multiworkload environments is to create multiple pools of up to 90 drives for midrange system performance requirements and up to 120 drives for low-range system performance requirements.

With all NetApp E-Series RAID choices, performance varies based on the drive type that is used; for example, SSDs are much faster than NL-SAS drives. This section provides best practice guidelines to help administrators, or anyone who is planning a proof of concept, to provision E-Series arrays to achieve performance goals for IOPS, throughput, and drive rebuilds. Although it is possible to create even larger pools, for NL-SAS HDDs, 60 to 120 drives per pool maintains an optimal balance between performance and reliability.

A single large pool might be a better fit for long-term archive use cases. This approach keeps administrative overhead low while maintaining the data on a very reliable storage platform.

Analytics best practices: Small-block random workloads

Both traditional relational databases and NoSQL databases generate a largely random, small-block read/write workload. This is also the case for OLTP and hypervisor workloads. DDP technology is tuned to perform very well under these types of random workloads. In this type of environment, it is best practice to put all volumes in a single pool. This approach simplifies system administration and database administration and meets the performance needs of both database log and data storage. Given the higher capacities of disks and the premium on floor space, it's no longer efficient to segment workloads and to isolate database log files to single RAID 1 volumes. Such segmentation creates islands of stranded storage and increases administrative overhead.

Another best practice is to ensure balance between the two RAID controllers, not just the number of volumes but also the function of the volumes. When you use NetApp SANtricity management software to

create volumes, the system automatically balances the volumes between the controllers. The Automatic Load Balancing feature can also dynamically adjust volume ownership based on controller workload. The automatic load-balancing feature requires SANtricity OS 11.30 or later and appropriate host type selection on Windows, VMware, or Linux (kernel 3.10 or higher).

Backup and video surveillance best practices: Sequential workloads

Consider a video surveillance workload that involves multiple cameras recording at once, each in a sequential manner. As multiple streams are written to different parts of the volume, the workload appears random to the array. For this scenario, DDP performs well, as discussed in the section “Analytics best practices: Small-block random workloads.”

For performance-sensitive workloads that are sequential and that typically have larger block transfer sizes, the best practice is to keep these workloads isolated and on their own traditional RAID volume group, rather than configuring pools. This type of workload describes many backup and surveillance applications. However, you should use performance sizing tools to determine whether volume groups are a requirement, given the rebuild benefits of DDP technology. Some backup implementations have shown that a DDP configuration meets the performance requirements while also delivering the rebuild benefits of DDP technology.

If performance is your main goal, the use of a traditional volume group with one volume is the best practice. If shorter rebuild times, better degraded-mode performance, and ease of administration are your goals, the use of pools of 12 to 30 drives can better meet your requirements.

Technical computing best practices: Large-block sequential workloads

Technical computing workloads such as Lustre or IBM Spectrum Scale generate large-block sequential reads and writes, demanding the most from the underlying media. The best practice for this type of environment is to use a traditional RAID volume group with one volume per volume group.

Configuring pools with equal capacity volumes

In some environments dividing a pool into equal volumes is required. Because SANtricity System Manager and the CLI require the administrator to enter the desired capacity for each volume to be created, the user must calculate these capacities before configuring the system.

To begin this calculation, it is useful to know that the usable capacity of a pool is always a multiple of the D-stripe size. The calculation requires you to determine the number of D-stripes per volume and then multiply by the D-stripe size to obtain the number of gibibytes per volume. As an example, suppose that the pool has 161,456GiB of usable capacity and the user wants to carve it into five equal volumes. The calculation is as follows:

1. Find the number of D-stripes in the pool: Total capacity in GiB / D-stripe size = 161,456/4 = 40,364.
2. Determine number of D-stripes per volume: 40,364/5 = 8,072.8.
3. Round down to a whole number of D-stripes per volume = 8,072.
4. Multiply the number of D-stripes by 4 to obtain the gibibytes per volume: 8,072*4 = 32288GiB.

   Note: In this example, four D-stripes are left over because of the remainder in step 2. If the total number of D-stripes is not divisible by the number of volumes that you want, you can’t create equal volumes that use all the pool capacity.

Figure 16 demonstrates this example in SANtricity System Manager, showing the five equal volumes in the pool. Figure 17 shows the pool capacity and the allocated capacity. Note the free capacity of 16.00GiB as described in this example. It is also important to note that for this simple calculation to work properly, preferences must be set to display capacity values in GiB.

Figure 16) Equal volumes in SANtricity System Manager.

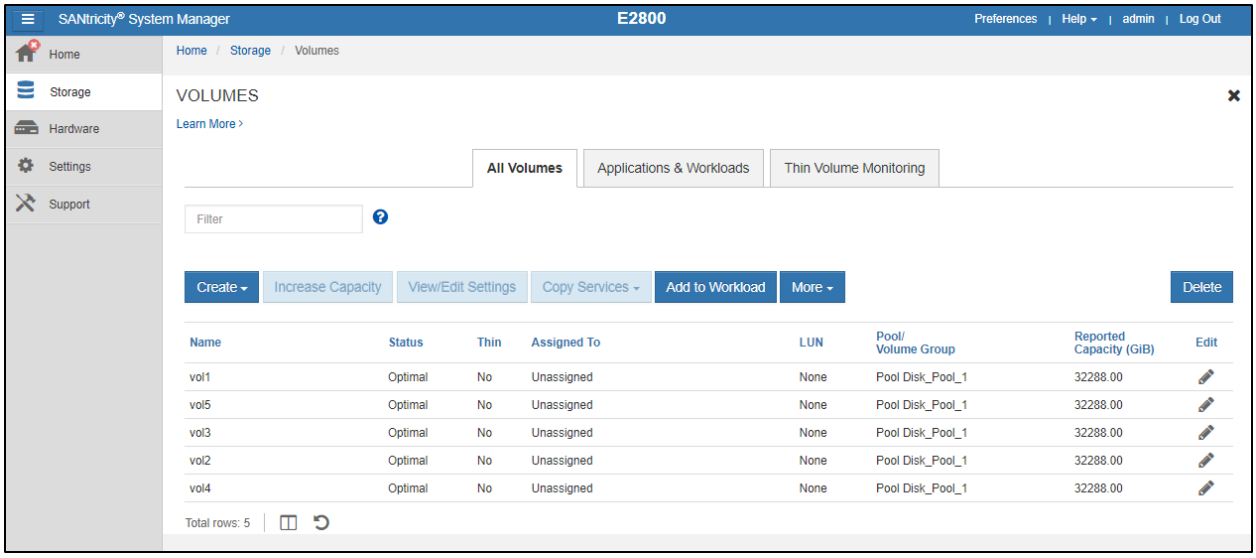
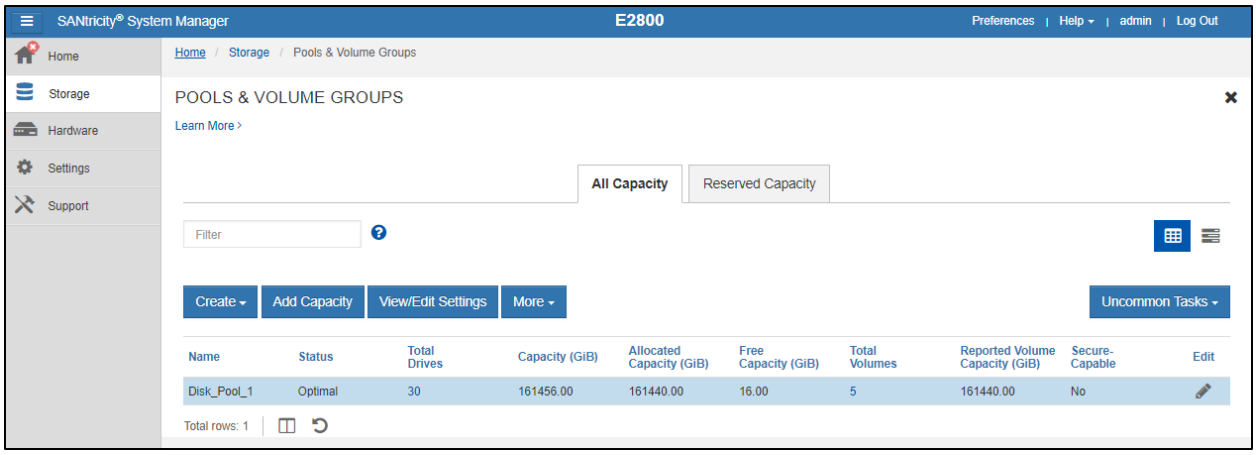


Figure 17) Pool capacity with equal volumes in SANtricity System Manager.



Using the CLI to create equal volumes is like using SANtricity System Manager. To illustrate, consider an example of a disk pool with 229.996TB of usable capacity and nine equal volumes. The following command displays the pool attributes:

```
show diskPool ["Disk_Pool_1"];
```

The portion of the result that shows capacity is as follows:

```
Total capacity:                              234.346 TB
        Preservation capacity:                     4,455.000 GB(3 Drives)
        Usable capacity:                           229.996 TB
```

To create nine equal volumes, enter the following command nine times with different volume names:

```
create volume diskPool="Disk_Pool_1"
userLabel="vol1"
capacity=26168GB;
```

To see the result, display the pool attributes again:

```
show diskPool ["Disk_Pool_1"];
```

The values that were returned for capacity are as follows:

```
Total capacity:                              234.346 TB
         Preservation capacity:                    4,455.000 GB(3 Drives)
         Usable capacity:                          229.996 TB
         Unusable capacity:                        0.000 MB

      Used capacity:                               229.992 TB
         Volumes:                                  (9), 229.992 TB
         Repositories:                             (0), 0.000 MB
      Free Capacity:                               (1), 4.000 GB
      Percent full:                                99%
```

Pools with equal capacity volumes can also be configured with the API documentation. To illustrate this, consider an example of a disk pool with 8672 GiB of usable capacity and eight equal volumes.

In the Volumes section, execute GET /storage-systems/{system-id}/storage-pools on the desired pool. If successful, then look for "freeSpace": in response and that value is the number of bytes in the pool:

```
"largestFreeExtentSize": "9311489097728",
    "raidStatus": "optimal",
    "freeSpace": "9311489097728",
    "drivePhysicalType": "sas",
    "driveMediaType": "ssd",
    "normalizedSpindleSpeed": "spindleSpeedSSD",
    "diskPool": true,
    "id": "04000000600A098000A09BE60000D65C64125BE9",
    "name": "TestPool"
```

Using an online converter, convert the number of bytes into GiB and then use the process above to find the number of Gibibytes per volume. Execute POST /storage-systems/{system-id}/volumes to create the desired number of volumes:

```
{
  "poolId": "04000000600A098000A09BE60000D65C64125BE9",
  "name": "EqualVolume1",
  "sizeUnit": "gb",
  "size": "1084",
  "raidLevel": "raid6"
}
```

Once again execute GET /storage-systems/{system-id}/storage-pools to check that all capacity has been used equally:

```
    "blkSizeRecommended": 512,
    "usedSpace": "9311489097728",
    "totalRaidedSpace": "9311489097728",
    "extents": [],
    "largestFreeExtentSize": "0",
    "raidStatus": "optimal",
    "freeSpace": "0",
    "drivePhysicalType": "sas",
    "driveMediaType": "ssd",
    "normalizedSpindleSpeed": "spindleSpeedSSD",
    "diskPool": true,
    "id": "04000000600A098000A09BE60000D65C64125BE9",
    "name": "TestPool"
```

Reconstruction priority setting

The reconstruction priority setting can be changed to optimize for the fastest rebuild times (highest-priority setting) or to minimize how storage system performance is affected during a drive rebuild (lowest-priority

setting). The default setting and best practice are high priority, balancing a fast rebuild time with maintaining acceptable system performance.

The rebuild of a degraded pool is faster than the rebuild of a traditional volume group. Additionally, during this degraded state, the I/O performance to the host server or servers is affected very little compared with a traditional volume group. With larger disk pool sizes, there is less effect on the performance with DDP technology. Rebuilds and the transition from a degraded or critical state happen faster than those of a traditional volume group.

This performance delta occurs because DDP technology decouples drive loss protection from a total drive rebuild. The DDP feature rebalances the data across the remaining drives, affecting only a portion of the drives in the array that are required to calculate parity. Traditional RAID affects all the drives in the volume, reducing overall performance. The expected performance degradation and length of time to complete the required rebuild operations can vary based on many factors, including workload pattern and rebuild priority settings.

Conclusion

Drive capacities continue to increase and add to rebuild times for disk failures, leaving storage systems at the risk of data loss. Prolonged drive rebuilds negatively affect the performance of the system for extended periods of time, making it difficult for administrators to meet SLAs and affecting business application response needs. With the introduction of SANtricity Dynamic Disk Pools technology, E-Series administrators now have a choice that provides many advantages over traditional volume groups. This technology:

- Substantially improves reconstruction times (by up to 8 times) and limits exposure to additional drive failures by prioritizing critical segment reconstruction
- Reduces the performance impact during rebuild
- Enables online addition or removal of drives, up to 60 at a time
- Eliminates dedicated idle hot spares

E-Series storage systems support both DDP and traditional volume groups. They can be mixed and matched in a single system to provide superior configuration flexibility to meet any workload requirement.

Appendix A: Glossary of terms

The following table defines terms that are used in this document that might be unfamiliar to the reader.

Term	Description
Critical segment reconstruction	When a D-stripe has two D-pieces that are affected by multiple drive failures, the system must re-create them on other drives in the pool. They are re-created by RAID 6 reconstruction of each of the segments within the affected D-pieces. These segments are called critical segments.
D-piece	The portion of a D-stripe that is contained on one drive.
D-piece size	The D-piece size depends on the drive. For NVMe drives, a D-piece is 1 GiB in total capacity. For all SAS drives, a D-piece is 512 MiB in total capacity.
Drawer loss protection (DLP)	When a pool has DLP, it can withstand failure or removal of one entire drawer of a drive shelf without losing availability.
D-stripe	The DDP element that comprises volumes. Each D-stripe resides on 10 drives for a RAID 6 volume or a RAID 1 5+5 volume in the pool, regardless of pool size. For a RAID 1 3+3 volume each D-stripe resides on 6 drives. D-stripes are distributed throughout a pool by an intelligent algorithm to optimize performance and failure tolerance.

Term	Description
D-stripe size – NVMe drives	The D-Stripe size depends on the RAID level and the D-piece size: • RAID 6: 1 GiB D-piece x 8 data drives = 8GiB • RIAD 1 5+5: 1 GiB D-piece x 5 data drives = 5GiB • RIAD 1 3+3: 1 GiB D-piece x 3 data drives = 3GiB
D-stripe size – SAS drives	The D-Stripe size depends on the RAID level and the D-piece size: • RAID 6: 512 MiB D-piece x 8 data drives = 4GiB • RIAD 1 5+5: 512 MiB D-piece x 5 data drives = 2.5GiB • RIAD 1 3+3: 512 MiB D-piece x 3 data drives = 1.5GiB
GB, MB, TB, PB	These units are used when referring to capacity by base 10 values. I/O throughput values and raw drive capacities are expressed in base 10 values.
GiB, MiB, TiB, PiB	These units are used when referring to capacity with base 2 values. Pool, volume group, and volume capacities are all expressed in base 2 values.
Logical unit number (LUN)	When a volume is assigned to a host, it acquires a reference number by which the host knows the volume.
Pool	The container in DDP technology that houses volumes.
RAID stripe size	The RAID stripe size is a function of the segment size for DDP (128KiB) and the number of data drives: • RAID 6: 128KiB x 8 data drives = 1MiB • RAID 1 5+5: 128KiB x 5 data drives = 640KiB • RAID 1 3+3: 128KiB x 3 data drives = 384KiB There are 8,192 RAID stripes in a D-stripe on NVMe drives and 4,096 RAID stripes in a D-stripe on SAS drives.
RAID group	An industry term that refers to containers that house volumes and that use traditional RAID. For the E-Series, NetApp refers to them as volume groups.
Segment	The portion of a RAID stripe that resides on a single drive. A segment is 128KiB in size, and there are 8,192 segments in a D-piece on NVMe drives and 4,096 segments in a D-piece on SAS drives.
Volume	The unit of capacity that is used to store application data. A volume resides in a volume group or in a pool and is assigned a LUN when it is associated with a host.
Volume group	The NetApp E-Series name for a RAID group.

Units' convention

In this document, IEC binary units are used when referring to base-2 values, and decimal units are used for base-10 values. Following are examples of binary units:

- **KiB.** Kibibyte, or 1,024 bytes
- **MiB.** Mebibyte, or $1,024^2$ bytes
- **GiB.** Gibibyte, or $1,024^3$ bytes
- **TiB.** Tebibyte, or $1,024^4$ bytes
- **PiB.** Pebibyte, or $1,024^5$ bytes

Following are examples of decimal units:

- **KB.** Kilobyte, or 1,000 bytes
- **MB.** Megabyte, or $1,000^2$ bytes
- **GB.** Gigabyte, or $1,000^3$ bytes
- **TB.** Terabyte, or $1,000^4$ bytes

- **PB.** Petabyte, or $1{,}000^5$ bytes

Note: NetApp SANtricity® System Manager uses binary labels for binary values.

Appendix B: Thin provisioning

E-Series and EF-Series arrays support thin provisioning when they are used with the DDP feature, but for most workloads, a thick volume is a better choice.

Note: Thin provisioning is not supported with RAID 10, RAID 5, RAID 6, nor on the EF300/EF600.

The thin provisioning feature provides overcommit capacity, but it does not deliver the same level of IOPS or throughput performance that thick volumes do. As a result, thin volumes are generally not recommended for EF-Series arrays or for any transactional workload. The maximum thin-provisioned volume size is 256 TiB.

If you originally configured thin volumes and want to change to thick volumes, you must copy existing data to a new thick volume. Therefore, you should carefully plan the time to copy data, the associated cost of conversion capacity, and application cutover logistics.

Thin provisioning is only available through the REST API. Here an example of creating a thin volume through the API Documentation:

```
{
  "poolId": "04000000600A098000A09BE60000D64D63FFF589",
  "name": "thinVolumeTEST",
  "sizeUnit": "gb",
  "virtualSize": "50",
  "repositorySize": "100",
  "maximumRepositorySize": "200",
  "dataAssuranceEnabled": false
}
```

Appendix C: Implementation of DDP Features only available through REST API

Creating a RAID 1 pool

As mentioned earlier in this document there are two features available through REST API that are not available in the UI. These features are an eight-drive pool minimum and creating RAID 1 volumes in the same pool as RAID 6 volumes.

As seen in Figure 18, a user can create an 8-10 Drive DDP which can have RAID 1 (3+3) volumes on it. If the pool is expanded to 11 or more drives, then RAID 6 (8+2) or RAID 1 (5+5) volumes can be created while the first RAID 1 volume will stay 3+3, as seen in Figure 19.

Figure 18) An 8-drive pool created through REST API.

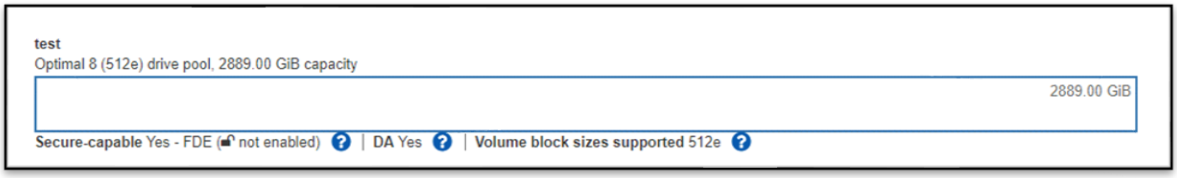
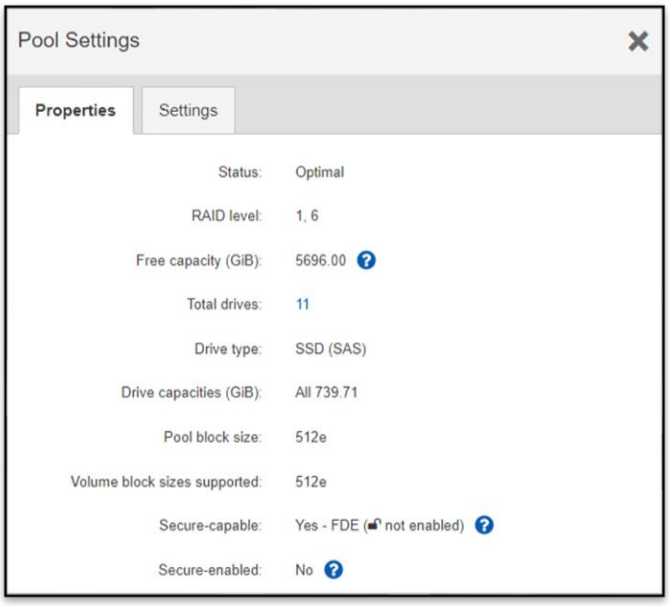


Figure 19) A pool with 11 drives that contains RAID 1 and RAID 6 volumes.



This appendix will show the options for using these features through Swagger Docs, Python Scripting, and SMcli.

Swagger Documentation

This section will contain the process of creating an 8-drive DDP using the REST API swagger documentation in SANtricity System Manager. This works only with SSDs. We will also cover how RAID 1 and RAID 6 volumes can be created on the same pool.

Gathering Information

This first part is to acquire the necessary information to create and edit storage pools and volumes.

Storage Array World-Wide Identifier

- Go into SANtricity system manager.
- Click on Support on the left side of the page.
- Click on Support Center.
- The storage array world-wide identifier will be found towards the middle of the page.
- Make sure to copy this and paste it into a separate file to use later.

Disk Drive IDs

- Go into SANtricity system manager.
- Press “help” dropdown button in the top right corner and then select “API Documentation”.
- Click on the Hardware section then on the GET button titled “/storage-systems/{system-id}/drives”.
- Press on the “Try it out” button on the right side then paste your storage array world-wide identifier into the system-id box.
- After pressing the blue Execute button you will have the response body which contains the drive IDs.
- For each drive copy the “driveRef” value into a file to use later.

8-Drive DDP and RAID 1 Volume

Now we will cover how to create a DDP with only 8 drives and create a RAID 1 3+3 volume from that pool.

Creating a Storage Pool

- Go to the Volume Section and Click on the first POST where you can create a storage pool.
- Click on the “Try it out” on the right side of the screen.
- Then you will need to input the WWN of the system in the first text box.
- In the second box you will need to input raidLevel as raidDiskPool to create a DDP, the driveIds, whether you want to eraseSecuredDrives and the name of the pool.
- Then press Execute.
- If you are successful, you will get a server response with code 200, and now you will want to go and copy the “volumeGroupRef” somewhere to use later to create volumes from this pool.

Creating a Volume

- In the Volume Section, click on the POST where you can create a volume.
- Click on the “Try it out” on the right side of the screen.
- Then you will need to input the WWN of the system in the first text box.
- In the main box you will place the “volumeGroupRef” we got earlier into “poolId”.
- You will also need to input a name for the volume, the desired size, if you want data assurance enabled and the raid level. (For an 8-drive pool only RAID 1 volumes can be created)
- Then press Execute.

Expanding to 11-Drive or more DDP and Creating a RAID 1 and 6 Volumes

Now we will expand the capacity of the pool to 11 or more drives. Now any new volumes will be RAID 6 8+2 or RAID 1 5+5 while the existing volumes will stay at RAID 1 3+3.

Expanding the Capacity of a Storage Pool

- In the Volume Section, click on the POST where you can expand the capacity of a storage pool.
- Click on the “Try it out” on the right side of the screen.
- Then you will need to input the WWN of the system in the first text box.
- You will also need to input the storage pool id which is also the same as the volumeGroupRef.
- In the body you will need to add the drive ids. (I put 3 drives to expand the pool to an 11-drive DDP)
- Then press Execute.

Creating a Volume from the expanded pool

I then created a RAID 6 and RAID 1 volume on this 11-drive DDP verifying that RAID 1 and RAID 6 volumes can be created on the same pool.

System Manager Command Line Interface (SMcli)

Now we will use the SMcli to create a storage pool and create volumes on that pool. This method has a simple setup and lots of documentation that can help the user.

- Go into the SANtricity system manager GUI, and on the left side click on setting, then click on system.
- Scroll down until you see Command Line Interface which is in the Add-ons section.
- Download the SANtricity command line interface and make sure to extract the contents in the desired location.

- Run the command prompt as administrator and then cd into the location of the SANtricity CLI's bin folder. (For me it was: C:\Users\marcosv\Downloads\SMcli\SMcli-01.60.00.9002\bin)
- To create an 8-drive pool, a command like this example can be executed:

```
SMcli 10.115.24.25 10.115.24.26 -u XXX -p XXX -k -c "create diskPool
driveType=SAS userLabel=\"poolTest\" driveCount=8;"
```

- To create a volume on that pool, a command like this example can be executed:

```
SMcli 10.115.24.25 10.115.24.26 -u XXX -p XXX -k -c "create volume
diskPool=\"poolTest\" userLabel =\"volume1\" capacity=10GB
thinProvisioned=FALSE cacheReadPrefetch=FALSE raidLevel=1;
```

- I then expanded the 8 Drive DDP made above into an 11 Drive DDP in the GUI and then added a RAID 6 volume and a RAID 1 volume to that pool using the SMcli.

Scripting through Web Services Proxy

There is also the option of using python scripts with the Web Services Proxy. Sample scripts are provided which can be edited and used. This is one of the most effective methods once code is understood.

- Download E-Series SANtricity Unified Manager and Web Services Proxy.
- Download and install Unified Manager (Version 6 and above) and add an array.
- In browser go to https://127.0.0.1/ to access NetApp SANtricity Web Services Proxy.
- From Here you can access Unified Manager or the interactive API swagger which is just like the SANtricity System Manager API documents.
- Make sure you have python installed with the modules: requests, urllib3, charset-normalizer, certifi, and idna.
- There are also sample python scripts which can be ran from the command line. (make sure sample files "restlibs.py" and "configuration.py" are in the same location as your scripts, you also want "idna.py" in your script location from C:\Program Files\Python38\Lib\encodings)
- Here is an example where a storage pool is created and then a volume pool is created to it:

```
from restlibs import generic_delete, generic_post, generic_get, array_controller
from configuration import defaultAddresses
from pprint import pprint
with array_controller(defaultAddresses) as array:

    #Use the drive selection endpoint to request # of drives and type
    drives_req = {"driveCount": 8, "interfaceType": "sas"}
    drives = generic_post('drives', drives_req, array_id=array['id'])

    if(drives is not None):
        driveSet = list(map(lambda d: d['driveRef'], drives))
        pool = generic_post('pools', {'diskDriveIds' : driveSet,
                                      'name' : 'testPool',  #make sure to input desired name and
raid level
                                      'raidLevel' : "raidDiskPool"}, array_id=array['id'])

    volume_data = {'name' : 'volume1',                  #make sure to input the details for
volume
                   'poolId' : pool['volumeGroupRef'],
                   'segSize' : 512,
                   'size' : 100,
                   'sizeUnit' : 'gb',
              'raidLevel': 'raid1'
                   }

    #create the volume
    volume = generic_post('volumes', data=volume_data, array_id=array['id'])
```

- I verified our story that if this 8-drive pool has 3 drives added then RAID 1 and RAID 6 volumes can be created on this now 11 drive pool while the original volume will stay RAID 1.

Where to find additional information

To learn more about the information that is described in this report, see the following documents:

- E-Series engineering documents (not available to the public): statement of work, feature description, Engineering Requirements (ER) document
- NetApp University online training courses
- E-Series and EF-Series systems engineer technical training deck that is published on the NetApp Field Portal

Version history

Version	Date	Document version history
Version 1.0	December 2017	Initial release.
Version 2.0	September 2019	Updated for 11.60.1.
Version 3.0	January 2020	Updated for video surveillance.
Version 4.0	June 2022	Added REST API only features.
Version 5.0	July 2023	Updated for 11.80 and added appendix on implementation of REST API only features.
Version 6.0	November 2024	Updated for 11.90 to modify DDP limits.

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

TR-4652-1124

NetApp