Best Practices

Dell EMC Unity: Oracle Database Best Practices

All-flash arrays

Abstract

This document provides best practices for deploying Oracle® databases with Dell EMC™ Unity All-Flash arrays, including recommendations and considerations for performance, availability, and scalability.

June 2019

H16765

Revisions

Date	Description
November 2017	Initial release for Dell EMC Unity OE version 4.2
June 2019	Updated with new format and content for Dell EMC Unity x80F arrays

Acknowledgements

Authors: Mark Tomczik, Henry Wong

Table of contents

Executive summary

This paper delivers straightforward guidance to customers using Dell EMC™ Unity All-Flash storage systems in an Oracle® 12c database environment on Linux® operating systems. Oracle is a robust product that can be used in a variety of solutions. The relative priorities of critical design goals such as performance, manageability, and flexibility, depend on your specific environment. This paper provides considerations and recommendations to help meet your design goals.

This paper was developed using the Dell EMC Unity 880F All-Flash array, but the information is also applicable to other Dell EMC Unity All-Flash array models (x80F and x50F) unless otherwise noted. The primary Linux operating system used in this paper was Oracle Linux (OL) 7, but content is applicable to Oracle Linux (OL) 6, and Red Hat® Enterprise Linux (RHEL) 6 and 7.

These guidelines are strongly recommended by Dell EMC, but some recommendations may not apply to all environments. For questions about the applicability of these guidelines in your environment, contact your Dell EMC representative.

Dell EMC Unity x80F models provide an excellent storage solution for Oracle workloads regardless of the application characteristics and whether file or block storage is required. This paper discusses the best practices and performance of the Dell EMC Unity 880F array with block storage, but also presents best practices with native xNFS or Oracle dNFS.

In addition to providing support for file and block storage, the Dell EMC Unity x80F arrays provide a number of standard features. Some of the standard features are point-in-time snapshots, replication (local and remote), built encryption, compression, and extensive integration capabilities for an Oracle standalone or RAC environment.

Audience

This document is intended for Dell EMC Unity administrators, database administrators, architects, partners, and anyone responsible for configuring Dell EMC Unity storage systems. It is assumed readers have prior experience with or training in the following areas:

- Dell EMC Unity storage systems
- Linux operating environment
- Multipath software
- Oracle Automated Storage Management (ASM)
- Oracle standalone or RAC environment

We welcome your feedback along with any recommendations for improving this document. Send comments to StorageSolutionsFeedback@dell.com.

1 Storage configuration

Dell EMC Unity storage is a virtually provisioned, flash-optimized storage system designed for ease of use. This paper covers the All-Flash array models with emphasis on Dell EMC Unity x80F arrays. This section describes the foundational array technologies that support the application-specific sections that follow. For general Dell EMC Unity best practices, see the Dell EMC Unity: Best Practices Guide as well as the documentation listed in appendix C.

1.1 I/O module options

All Dell EMC Unity All-Flash arrays provide an embedded I/O module (optional on some models) and two optional I/O modules per storage processor (SP). This section provides an overview of the different options available in each of the I/O modules and makes a recommendation on choosing I/O modules for an Oracle environment where performance and throughput are of interest.

I/O ports used for transferring user data are available on the embedded module or additional I/O modules on each SP.

Table 1 Embedded I/O module

Array model	Converged Network Adapter (CNA)	Ethernet	SAS
300, 350F, 380, 380F 400F, 450F, 500F, 550F, 600F, 650F	2-ports: 8/16Gb Fibre Channel (FC) 4/8/16Gb FC 16Gb FC (single mode) 1/10Gb IP/iSCSI	2-ports, 10GbE Base T	2-ports, mini-HD
480, 480F, 680, 680F, 880, 880F	N/A	Optional 4-port mezzanine card: • 25GbE optical (no auto negotiation) with either 10Gb or 25Gb SFPs (mixed ok), or TwinAx (active/passive) • 10GbE BaseT RJ45 • no 4-port card (requires blank filler)	2-port 12Gb/s SAS

For Oracle environments that require FC front-end connectivity on the Dell EMC Unity 480F, 680F, or 880F models, consider the option of a filler blank in place of the embedded I/O module and use one or two optional I/O modules that support FC.

If Oracle dNFS is used, consider using the optional 4-port mezzanine card on the Dell EMC Unity 480F, 680F, or 880F models since the card can be included in both Link Aggregation Control Protocol (LACP) and fail-safe networking (FSN) configurations. LACP is discussed in the section, Configuring LACP.

Table 2 Optional I/O modules for Dell EMC Unity All-Flash arrays

Array model	Fibre Channel I/O module	Ethernet Base-T I/O module	Ethernet/iSCSI optical I/O module	SAS I/O module
300F, 350F, 400F, 450F	4-port 16Gb/s	4-port 1GbE, or 4-port 10GbE	4-port 10GbE, or 2-port 10GbE offloading	NA
500F, 550F, 600F, 650F	4-port 16Gb/s	4-port 1GbE, or 4-port 10GbE	4-port 10GbE, or 2-port 10GbE offloading	4-port mini HD (backend)
380, 380F	4-port 16Gb/s	4-port 10GbE BaseT RJ45 (auto-negotiate to 1GbE)	4-port 25GbE optical for Ethernet and iSCSI block traffic; either 10Gb or 25Gb SFPs (no auto negotiation, mixed SFPs ok), or TwinAx (active or passive)	NA
480, 480F, 680, 680F, 880, 880F	4-port 16Gb/s	4-port 10GbE BaseT RJ45 (auto-negotiate to 1GbE)	4-port 25GbE optical for Ethernet and iSCSI block traffic; either 10Gb or 25Gb SFPs (no auto negotiation, mixed SFPs ok), or TwinAx (active or passive)	4-port 12Gb SAS backend

In high-demand Oracle environments where IOPs, latency, or capacity are a concern, consider the option of using a 4-port 12Gb SAS I/O module to increase the number of configurable physical drives in the array which can help lower latency and increase IOPS and capacity.

A requirement of installing I/O modules is that they are installed in pairs (one in SPA and one in SPB) and that they are of the same type and reside in the same slots between SPA and SPB.

With Dell EMC Unity 480F, 680F, and 880F models, slot 0 I/O modules have x16 PCIe lanes while slot 1 has x8 PCIe lanes. For this reason, slot 0 should be reserved for environments needing greater bandwidths.

The Ethernet/iSCSI card can be included in both Link Aggregation Control Protocol (LACP) and fail-safe networking (FSN) configurations).

Once the Dell EMC Unity array is configured, all I/O modules are persistent and cannot change type.

1.2 Dynamic storage pools

Dell EMC Unity storage supports two types of storage pools on All-Flash storage systems: traditional pools and dynamic pools. Dynamic pools were introduced in Dell EMC Unity OE version 4.2 for all-flash storage models and became the default pool type in Dell EMC Unisphere™. While traditional pools are still supported on all-flash models, they can only be created through the Unisphere CLI or REST API. Dynamic pools offer many benefits over traditional pools. The new pool structure eliminates the need to add drives in the multiples of the RAID width. This allows for greater flexibility in managing and expanding the pool. Dedicated hot spare drives are also no longer required with dynamic pools. Data space and replacement space are spread across the drives within the pool. This allows better drive utilization, improves application I/O, and speeds up the proactive copying of failing drives and the rebuild operation of failed drives.

In general, it is recommended to create dynamic pools with large numbers of drives of the same type, and use a small number of storage pools within the Dell EMC Unity system. However, it may be appropriate to configure additional storage pools in the following instances:

- Separate workloads and resources from competing databases or applications
- Dedicate resources to meet specific performance goals
- Create smaller failure domains

Additional information can be found in the documents, *Dell EMC Unity: Dynamic Pools* and *Dell EMC Unity: Configuring Pools.*

1.2.1 Storage pool capacity

Storage pool capacity is used for multiple purposes:

- To store all data written into storage objects — LUNs, file systems, datastores, and VMware® vSphere® Virtual Volumes™ (VVols) — in that pool
- To store data that is needed for snapshots of storage objects in the pool
- To track changes to replicated storage objects in that pool

Storage pools must maintain free capacity to operate properly. By default, a Dell EMC Unity system will raise an alert if a storage pool has less than 30% free capacity, and will begin to automatically invalidate snapshots and replication sessions if the storage pool has less than 5% free capacity. Dell EMC recommends that a storage pool always have at least 10% free capacity.

Additional drives can be added to a storage pool online. However, to optimize the performance and efficiency of the storage, add drives with same specification, type, and capacity of the existing drives in the pool. Though not required, add a number of drives equal to the RAID width + 1, which allows the new capacity to be immediately available. Data is automatically rebalanced in the pool when drives are added.

Note: Once drives are added to a storage pool, they cannot be removed unless the storage pool is deleted.

1.2.2 All-flash pool

All-flash pools provide the highest level of performance in Dell EMC Unity systems. Use an all-flash pool when the application requires the highest storage performance at the lowest response time. Note the following considerations with all-flash pools:

- Consists of either all SAS flash 3 or all SAS flash 4 drives of the same capacity.
- Dell EMC FAST™ Cache and FAST VP are not applicable to all-flash pools.
- Compression is only supported on an all-flash pool.
- Snapshots and replication operate most efficiently in all-flash pools.
- Dell EMC recommends using only a single drive size and a single RAID width within an all-flash pool.

  For example: For an all-flash pool, use 800 GB SAS flash 3 drives and configure them all with RAID 5 8+1. For supported drive types in all-flash pool, see appendix B.

1.2.3 Hybrid pool

Hybrid pools (including a combination of flash drives and hard disk drives) are not supported with Dell EMC Unity All-Flash arrays.

2 Dell EMC Unity features

This section describes some of the native features available on the Dell EMC Unity platform. Not all are applicable to Dell EMC Unity All-Flash arrays and are noted in this document. Additional information on each of these features can be found in the *Dell EMC Unity: Best Practices Guide*.

2.1 FAST VP

Dell EMC FAST™ VP accelerates the performance of a specific storage pool by automatically moving data within that pool to the appropriate drive technology based on data access patterns. FAST VP is only applicable to hybrid pools within a Dell EMC Unity Hybrid flash system.

2.2 FAST cache

FAST Cache is a single global resource that can improve the performance of one or more hybrid pools within a Dell EMC Unity Hybrid flash system. FAST Cache can only be created with SAS Flash 2 drives, and is only applicable to hybrid pools. FAST Cache is not applicable to all-flash arrays.

2.3 Data reduction

Dell EMC Unity compression provides a way to reduce the amount of physical storage needed to save a dataset in an all-flash pool for block LUNs and VMFS datastores, which helps reduce the total cost of ownership of Dell EM Unity storage. This capability was added to Dell EMC Unity OE version 4.1 for thin block storage resources and was called Dell EMC Unity Compression. Thin file storage resource support was added in Dell EMC Unity OE version 4.2 for file systems and NFS datastores in an all-flash pool.

In Dell EMC Unity OE version 4.3, the Dell EMC Unity Data Reduction feature replaces compression. It provides more space savings logic to the system with the addition of zero block detection and deduplication. In Dell EMC Unity OE version 4.5, data reduction includes an optional feature called Advanced Deduplication, which expands the deduplication capabilities of the data reduction algorithm. With data reduction, the amount of space required to store a dataset for data reduction enabled storage resources is reduced when savings are achieved. Data reduction/advanced deduplication is supported on LUNs, file systems, and NFS/VMFS datastores. Starting with OE 4.5, an 8 KB Dell EMC Unity block within a resource is subject to compression and will be compressed if a 1% savings or higher can be obtained

Dell EMC Unity Data Reduction savings are not only achieved on the storage resource it is enabled on, but space savings are also realized on snapshots and thin clones of those resources as well. snapshots and thin clones inherit the data reduction setting of the source storage resource, which helps to increase the space savings that they can provide.

Dell EMC Data Reduction is easy to manage, and once enabled, is intelligently controlled by the storage system. Configuring data reduction and reporting savings is simple, and can be done through Unisphere, Unisphere CLI, or REST API.

Dell EMC Unity Data Reduction is licensed with all physical Dell EMC Unity systems at no additional cost. Data reduction is not available on the Dell EMC Unity VSA version of the Dell EMC Unity platform as data reduction requires write caching within the system. To use data reduction with block and file storage resources such as thin LUNs, thin LUNs within a consistency group, thin file systems, and thin VMware VMFS and NFS datastores, the system must be running Dell EMC Unity OE version 4.3 or later.

By offering multiple technologies of space saving, Dell EMC Unity provides flexibility for the best balance of space savings and performance.

Table 3 Dell EMC Unity all-flash arrays: data reduction/advanced deduplication

Dell EMC Unity OE	Dell EMC Unity model	Dell EMC Unity pool type	Technology
4.3/4.4	300, 400, 500, 600 300F, 400F, 500F, 600F 350F, 450F, 550F, 650F	All flash*	Data reduction
4.5	300, 400, 500, 600 300F, 400F, 500F, 600F 350F, 450F, 550F, 650F	All flash*	Data reduction
	450F, 550F, 650F	All flash*	Data reduction and advanced deduplication on dynamic pools only
5.0	300, 400, 500, 600 300F, 400F, 500F, 600F 350F, 450F, 550F, 650F 380, 480, 680, 880, 380F, 480F, 680F, 880F	All flash*	Data reduction
	450F, 550F, 650F, 380, 480, 680, 880, 380F, 480F, 680F, 880F	All flash*	Data reduction and advanced deduplication

* Resource can be created on either a traditional or a dynamic pool (for systems that support dynamic pools).

Note: Data reduction is disabled by default and needs to be enabled before advanced deduplication is an available option. After enabling data reduction, advanced deduplication is available, but is disabled by default.

While data reduction helps to optimize storage investments by maximizing drive utilization, be aware that data reduction increases the overall CPU load on the Dell EMC Unity system when storage objects service reads or writes of compressible data, and may increase latency when accessing the data.

Consider these best practices before enabling data reduction on a storage object:

- Monitor the system to ensure it has available resources to support data reduction. Refer to the “Hardware Capability Guidelines” section and Table 2 in the *Dell EMC Unity: Best Practices Guide*.
- Enable data reduction on a few storage objects at a time and then monitor the system to be sure it is still within the recommended operating ranges before enabling Data Reduction on more storage objects.
- With Dell EMC Unity x80F models, consider that data reduction will provide space savings if the data on the storage block is at least 1% compressible. Prior to the new x80F models and OE 5.0, data reduction would provide space savings if the data on the storage block was at least 25% compressible.
- Before enabling data reduction on a storage object, determine if it contains data that will compress; do not enable data reduction on a storage object if there will be no space savings.
- Contact your Dell EMC representative for tools that can analyze the data compressibility.

For more information regarding compression, see the *Dell EMC Unity: Data Reduction* document.

For additional information on Dell EMC Unity Data Reduction, see the *Dell EMC Unity: Data Reduction Overview* and *Dell EMC Unity: Data Reduction Analysis.*

2.3.1 Advanced deduplication

To increase the capacity efficiency of data reduction, advanced deduplication is an optional extension of data reduction released in OE 4.5 that can be enabled on storage and is only performed on compressed blocks. With OE 4.5, that meant that advanced deduplication would be performed on compressed blocks that sustained as little as a 1% savings or higher. In cases where a Dell EMC Unity block did not compress, advanced deduplication would not be performed. This meant that if there were blocks that contained data that was uncompressible yet multiple copies of the block existed, the duplicate copies of the block would not be reduced by a deduping process to realize further storage savings.

This restriction of not performing advanced deduplication on blocks with less than a 1% compression savings no longer exists in OE 5.0. Therefore, with OE 5.0, advanced deduplication will be able to deduplicate to an uncompressed block whenever a write or overwrite occurs on the block even if the block has 0% compression.

For more information regarding advanced deduplication, see the *Dell EMC Unity: Data Reduction* white paper and the *Dell EMC Unity: Best Practices Guide*

2.4 Data at Rest Encryption

Many Oracle database applications have data encryption requirements, specifically for data at rest. Data at Rest Encryption (D@RE) is a controller-based encryption solution that can be used for Oracle databases without requiring any database or application changes, allowing it to avoid performance impact to the database server or the applications.

Note: D@RE is a license-able feature and must be selected during the ordering processes and licensed at system initialization. D@RE can only be enabled at the time of system installation with the appropriate license and cannot be enabled later.

If encryption is enabled, Dell EMC recommends making external backups of the encryption keys after system installation, as well as immediately following any change in the system's drive configuration (such as creating or expanding a storage pool, adding new drives, or replacing a faulted drive).

For more information on D@RE, see the *Dell EMC Unity: Data at Rest Encryption* document.

2.5 Host I/O limits

Typically, a Dell EMC Unity system is used to service multiple hosts and applications. These applications can have different service levels and different storage demands. In addition, a single array can provide services to multiple environments such as development and testing as well as production. Traditionally, these scenarios have been difficult to manage to ensure that critical applications get the resources they need while managing less critical resources to make sure they do not over consume.

Host I/O limits, similar to Quality of Service (QoS), provide an excellent means to manage these types of workloads. Instead of trying to manage these workloads with multiple storage pools, host I/O limits allow LUNs to be restricted to a specified amount of IOPS or bandwidth so they do not adversely impact other applications. Also, this allows storage administrators to ensure applications and environments adhere to budgeted limits which greatly simplifies planning and management.

Host I/O limits are recommended for Oracle database environments for several reasons. First, storage administrators can ensure that demanding Oracle databases instances do not overwhelm the entire array by setting limits on database volumes. Also, if the Oracle database is the priority application, they can set limits on other LUNs on the system to ensure that the Oracle database can get the resources it requires. Another great component of host I/O limits is the ability to burst for a given limit for a specific period of time, which is user configurable. In this way, small exceptions can still be allowed while maintaining balanced performance.

In development and testing environments, it can be difficult to determine if an application meets performance requirements. Typically, test and development databases are smaller than production databases and it is not always feasible to keep a copy of production data in these environments due to storage costs or privacy concerns surrounding confidential data. The issue with smaller datasets is that the application can run much faster on a smaller database during development or testing phases, then encounter serious performance issues when deployed on a real dataset in production.

Host I/O limits can be used to restrict the I/O on smaller datasets to highlight I/O-intensive queries. Setting limits on databases in development and testing environments will help identify problem areas so they can be resolved prior to production deployment. The result is improved Oracle databases service levels and greater scalability.

For additional information, see the Dell EMC Unity: Unisphere Overview document.

3 Oracle database design considerations

The storage system is a critical component of any Oracle database environment. Sizing and configuring a storage system without understanding the I/O requirements can have adverse consequences. This section discusses the types of database workloads and some of the common tools available to measure, collect, and analyze database system performance which helps define the I/O requirements. For an Oracle environment, capacity requirements can be as important as the number of I/Os per second (IOPS) and throughput requirements.

3.1 OLTP workloads

An online transaction processing (OLTP) workload typically consists of small random reads and writes. The I/O sizes are generally equivalent to the database block size. The primary goal of designing a storage system for this type of workload is to maximize the number of IOPS while keeping the latency as low as possible. Depending on the business and application requirement, a latency of less than 1 ms is typical in a high performing environment.

Consider using 16Gb Fibre Channel (FC) or 25GbE optical I/O models in each SP. If higher drive counts are necessary to achieve higher IOPS, use 12Gb SAS IO modules for backend connectivity, such as from the storage processors (controllers) to the disk enclosures. 12Gb SAS is only available in the 480F, 680F, and 880F models and 25GbE optical is only available in Dell EMC Unity x80/x80F models.

3.1.1 OLTP performance

For best results, capture performance statistics for a period of at least 24 hours that includes the system peak workload

While every environment is unique, an OLTP workload typically consists of small, random reads and writes, and the backend storage system servicing this type of workload is primarily sized based on capacity and the number of IOPS required.

3.1.1.1 IOPS and latencies

By applying the best practices documented in this guide in an environment that has the characteristic of performing 8K random 70/30 read/write (R/W), and has a specific configuration, the Dell EMC Unity 880F array can deliver up to 29% more IOPS and up to 22% lower average latency than a Dell EMC Unity 650F array could deliver under the same configuration and load characteristic.

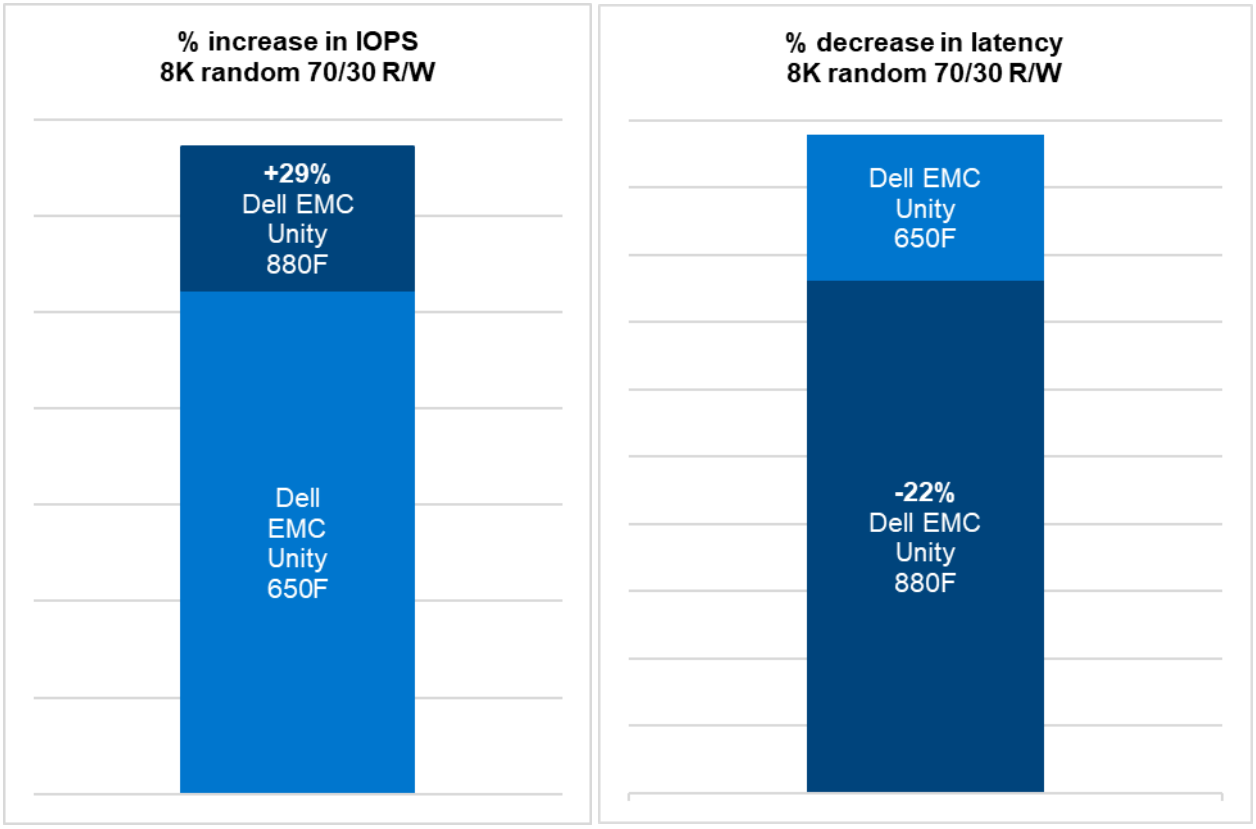

Figure 1 Percent change in average latency and IOPS when average latency is 1.5 ms: Dell EMC Unity 880F and Dell EMC Unity 650F

If a solution requires 1 ms average latencies, the Dell EMC Unity 880F array can provide up to 15% more IOPS and up to 13% lower average latencies than the Dell EMC Unity 650F array.

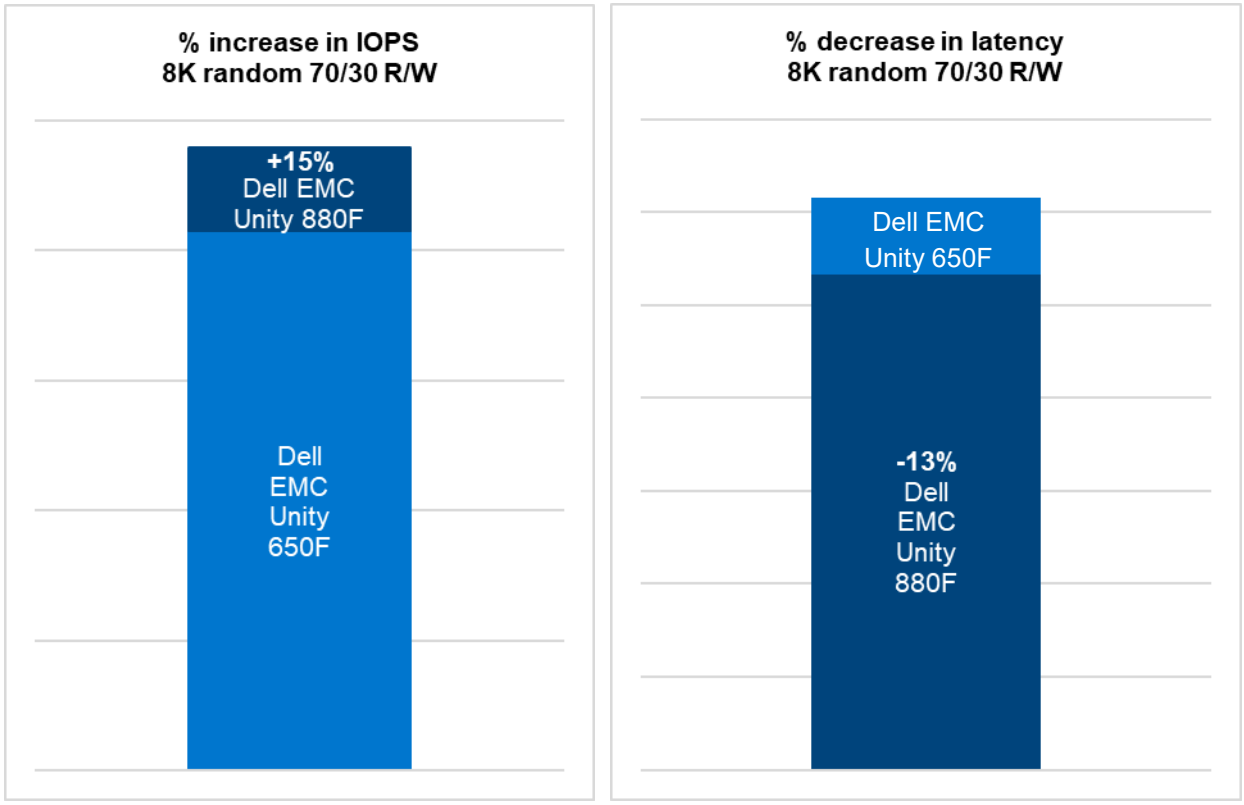

Figure 2 Percent change in average latency and IOPS when average latency is 1 ms: Dell EMC Unity 880F and Dell EMC Unity 650F

3.2 OLAP/DSS workloads

Unlike an OLTP workload, an online analytic processing (OLAP) or decision support system (DSS) workload is characterized by a relatively low volume of transactions. Most of the activities involve complex queries and aggregates a large data set. The volume of data tends to grow steadily over time and is kept available for a much longer time period. OLAP workloads generally have large sequential reads or writes.

The primary goal of designing a storage system that services this type of workload is to optimize the I/O throughput. The design needs to consider all components in the entire I/O path between the hosts and the drives in the Dell EMC Unity system. For best throughput, consider using 16Gbps FC or 25GbE optical (10GbE is also an option) iSCSI connectivity to the array, and 12Gbps SAS connectivity from the controllers to the disk enclosures. To meet high throughput requirements, multiple HBAs may be required on the server, the array, or both.

3.3 Mixed workloads

Oracle database workloads may not have I/O patterns that can be strictly categorized as OLTP or OLAP because several applications can reside within the same database. Multiple databases with different workloads can also co-exist on the same host. It is important to choose and design a storage system that can handle different types of workloads. Therefore, when testing the I/O systems, the combined workload of these databases should be accounted for and measured against the expected performance objectives.

The Dell EMC all-flash midrange storage portfolio offers storage systems that scale in both IOPS and throughput. Combined with the advanced architecture and storage-saving features like data reduction through compression, thin provisioning, thin clones, and snapshots, the Dell EMC all-flash midrange platform is ideal for any type of Oracle workload.

3.4 Storage pools

In general, it is recommended to use fewer storage pools within Dell EMC Unity systems because this reduces complexity and increases flexibility. Dell EMC recommends using a single virtual disk pool for hosting volumes for Oracle databases. This provides better performance by leveraging the aggregate I/O bandwidth of all disks to service I/O requests from Oracle databases. A single drive pool is also easier to manage, allowing an administrator to quickly and easily adapt the storage system to satisfy the ever-changing workloads that are common in Oracle databases environments. Before creating multiple storage pools to segregate workloads, understand the various Dell EMC Unity features that are available for managing and throttling specific workloads.

3.4.1 RAID configurations

By default, the Dell EMC Unity system chooses RAID 5 as the protection level when creating a new storage pool. This contradicts traditional guidance advocating RAID 1/0 for database workloads. However, this traditional guidance assumes the storage system contains spinning disks and does not consider SSDs or flash-optimized storage such as Dell EMC Unity systems. Internal testing has shown that the performance difference in most configurations between RAID 5 and RAID 1/0 in Dell EMC Unity All-Flash systems is negligible unless the workload is extremely write-intensive for an extended time period. In most cases, the small performance gain of RAID 1/0 is not worth the reduced capacity and therefore it is recommended to use the default configuration of RAID 5. For extremely heavy write workloads where maximum write performance is required, RAID 1/0 can be used.

The I/O requirements need to be clearly defined to size storage correctly. The RAID type chosen will be determined by comparing availability performance requirements. The small footprint and high IO density will typically allow a smaller drive size, reducing drive rebuild times. This means RAID 5 would be preferred over RAID 6 in most cases.

3.5 Testing and monitoring

Once the I/O requirements have been defined, I/O performance should be validated and tested before putting the environment into full production mode. After the system goes into production, it is imperative to continue to collect and analyze the performance data periodically to ensure the storage system is meeting the expected performance level. Ensure that a baseline is established and recorded so that comparison can be made.

There are many benchmarking and I/O simulation tools in the market that provide a comprehensive set of features to exercise and measure the storage system as well as other components in the I/O stack. It is up to administrators to decide the testing requirements and which tools work best for their environment. When choosing tools, consider the capabilities in the following subsections. Several performance testing utilities are shown below:

- I/O subsystem
  - dd
  - Iozone (fs benchmarking)
  - Iometer
  - ORION
  - winsat
  - FIO (fs benchmarking)
  - Vdbench
  - bonnie
- Relational Database Management System (RDBMS) level
  - SLOB
  - Oracle Database I/O calibration feature
  - DBMS_RESOURCE_MANAGER.CALIBRATE_IO
  - dbbenchmark
- Application Level testing tools (DB side)
  - Benchmark factory
  - Hammerora (hammerdb): supports the TPC-C and TPC-H workloads
  - Swingbench
  - Simora: Mines Oracle SQL Trace files and Generated SQL to be executed to reproduce the load
  - Oracle Real Application Testing: an enterprise database option from Oracle that records a database load on the source system and replays it on a destination environment
- Application Level testing tools (app side)
  - HP LoadRunner

3.5.1 Testing the I/O path

The first item to test on a new configuration is the path between the server and the array. Running a large block sequential read test using small files should saturate the path between the server and the array. This test verifies that all paths are fully functional and can be used for I/O traffic. Run this test on a dedicated server and array; using a production system could cause significant performance issues.

To validate the I/O path, run a large block sequential read test using the following guidelines as a starting point and vary as necessary:

- Create one LUN per storage processor
- Format the volumes using a 64 KB allocation unit
- Use a block size of 512 KB for the test
- Configure the test for 32 outstanding I/Os
- Use multiple threads. Eight is the recommended starting point

If the displayed throughput matches the expected throughput for the number of HBA ports in the server, the paths between the server and Dell EMC Unity array are set up correctly.

3.5.2 Testing the drives

Once the I/O path has been validated, the next step is to test the drives. For best results when testing drives on a Dell EMC Unity array, use the following guidelines when configuring the test.

- In a dual-controller system, use at least one volume per controller. This ensures that I/O will be distributed across both controllers. Using both controllers more closely simulates real world activity. For best results, use the same number of volumes on each controller. More LUNs might be better and may be required to achieve maximum performance.
- When performing I/O tests on any storage platform, it is important to use files that are larger than the Dell EMC controller cache. For more accurate results, use a file size that matches the amount of data being stored. In an environment where that is not practical due to a large data set, use a file size of at least 100 GB.
- Some I/O test tools (Oracle ORION is an example) generate files full of zeros. This behavior causes inaccurate results when testing. Avoid using test utilities that write zeros for drive validation or configure the tool to avoid writing zeros.

The purpose of this type of testing is to validate that the storage design will provide the required throughput and IOPS with acceptable latency. It is important that the test does not exceed the designed capacity of the array. For example, an array designed for a workload of 5,000 IOPS is likely to perform poorly with a workload of 10,000 IOPS. If a test is generating a workload higher than the designed capacity, adjust the workload being generated by reducing the number of threads and/or outstanding I/Os.

The results of the Live Optics analysis provide an I/O target to simulate using these tests. To get an idea of the performance capabilities of the array, run I/O tests with a range of I/O sizes commonly seen with Oracle. When testing random I/O, test with I/O sizes of 8 KB, 16 KB, and 32 KB. When testing sequential I/O, test with 8 KB, 16 KB, 32 KB and 64 KB. Since processes like read ahead scans and backups can issue much larger sequential I/O, it is a good idea to also test block sizes larger than 32 KB. To truly test the array the designed workload should be simulated at a minimum, and slightly higher if possible. To ensure the array has headroom for load spikes the throughput should be tested slightly beyond estimated production loads.

3.5.2.1 I/O simulation

The primary objective of I/O simulation is to stress the storage system. I/O simulation tools are typically easy to use and configure because they do not require a fully configured database. These tools generally allow workloads to increase or decrease during the tests by specifying different parameters:

- I/O block size
- Queue depth (outstand requests)
- Number of test files and sizes
- Number of I/O threads
- Read/write patterns (read only, write only, read/write mix)
- Ability to easily step through a series of tests with various settings (for example, run sets of 8k tests, 16k tests, 32k tests)
- Ability to write random data (non-zero) to the storage
- Ability to generate test reports or export data to other applications (such as Microsoft® Excel®)

Oracle ORION, Vdbench, and FIO are three popular I/O simulation tools available. The software is free to download and use, and the developers update the software periodically to keep up with the latest storage technologies. Oracle ORION has a unique advantage over others because it is explicitly designed to simulate Oracle database I/O workloads using the same I/O software stack as Oracle. It also provides both OLTP and OLAP simulation modes which simplify the setup and execution of the test. ORION has been bundled with the Oracle database software and can be found in the **$ORACLE_HOME/bin** directory. See appendix C for references to these tools. For more information on how to configure and run ORION, see the chapter, Calibration with the Oracle ORION Calibration Tool, in the Oracle Performance Guide.

3.5.2.2 Database transaction generation

These tools focus on generating a combination of workloads with different types of database transactions to simulate a typical OLTP, OLAP, or both. Therefore, they require a higher degree of configuration and customization in the tools and databases. SLOB, HammerDB, Swingbench, and Quest Benchmark Factory are commonly used to perform database transactional I/O benchmarks. With the exception of Quest Benchmark Factory, all of these tools are freely available for multiple operating systems.

3.5.3 Performance monitoring

Performance data can be monitored by the operating system, the Dell EMC Unity system, and in Oracle databases. Ideally, monitoring the performance continuously offers the most detail and allows in-depth analysis of the environments. At a minimum, performance statistics should be captured for at least 24 hours and during the time periods when there are heaviest activities. The following subsections describe popular software and cloud-based platforms for monitoring and analyzing performance.

3.5.3.1 Operating system monitoring

The following utilities are freely available from the operating system vendors and can perform basic system and I/O monitoring. Refer to the operating system manual and the online resources for each tool to find more information.

- sar (Linux)
- iostat (Linux)
- top (Linux)
- atop and netatop (Linux)
- collectl (Linux)
- Performance Monitor (Microsoft Windows)

3.5.3.2 Oracle database monitoring

For an Oracle database, the utilities used most are **statspack** and **AWR**. AWR is preferred to the older statspack, but either one can provide abundant performance statistics of a database. Both utilities come bundled with the database software.

Oracle Enterprise Manager (OEM) is a separate application offered by Oracle. It provides a centralized management and monitoring platform for many Oracle applications and databases. Configuration and performance data are collected through Oracle agents running on an individual host stored in a common management database. OEM provides a plethora of performance and utilization charts and many other advanced features to manage the environment.

Additional information can be found at the Oracle Enterprise Manager page.

3.5.3.3 Unisphere performance dashboard

This web-based unified management software comes with every Dell EMC Unity storage system and can manage every aspect of the storage. The performance dashboard in Dell EMC Unisphere provides both real-time and historical performance charts. Administrators can easily modify existing dashboards and charts, or add new dashboards and charts according to their needs.

Metric data ages over time and gets aggregated into longer sampling intervals. The data is kept for historical referencing for up to 90 days.

Additional information can be found in the *Dell EMC Unity: Unisphere Overview* document.

3.5.3.4 Dell EMC Live Optics

Dell's Performance Analysis Collection Kit (DPACK) has evolved into a new product called Dell EMC Live Optics and is a platform-agnostic analysis service freely available from Dell EMC. It works on Linux, Microsoft Windows, and VMware environments and collects performance data such as processor utilization, memory utilization, storage utilization, IOPS, I/O throughput, and more. Live Optics analyzes these data and provides a comprehensive in-depth report on server workloads and capacity.

Find additional information at the Dell EMC Live Optics page.

3.5.3.5 Dell EMC CloudIQ

Dell EMC CloudIQ™ is a software as a service (SaaS) application that is freely available. When it is enabled for the Dell EMC Unity storage system, it allows administrators to monitor multiple Dell EMC Unity storage systems remotely. CloudIQ provides continuous monitoring performance, capacity, configuration, and data protection, and enables administrators to manage storage proactively by receiving advanced notification for potential issues.

Find additional information in the *CloudIQ Overview* document.

4 Deploying Oracle databases on Dell EMC Unity storage

This section discusses best practices for architecture and configuration of Oracle storage for Oracle databases to realize the optimal performance and manageability of the environment.

4.1 Linux setup and configuration

Oracle databases are commonly deployed on Linux operating systems. The following subsections describe best practices when working with Dell EMC Unity storage systems on Linux operating systems.

4.1.1 Discovering and identifying Dell EMC Unity LUNs on a host

After creating and enabling host access of the LUNs in the Dell EMC Unity system, the host operating system needs to scan for these new LUNs before they can be used. On Linux, install the following rpm packages which contain useful utilities to discover and identify LUNs: **sg3_utils** and **lsscsi**.

4.1.1.1 Identifying LUN IDs on Dell EMC Unity storage

The Dell EMC Unity storage system automatically assigns LUN IDs, starting from 0 and incrementing by 1 thereafter, when enabling access to a host. Therefore, LUN 0 typically represents the very first LUN allowed access to a host.

Perform the following to view the LUN ID information:

1. In Unisphere, click Access > Hosts.
2. Select the host check box > **Host Properties** (pencil icon) > **LUNs** tab.
3. If the **Host LUN ID** column is hidden from the default view, click the columns filter (gear icon) > **Columns** > **Host LUN ID**. See Figure 3.

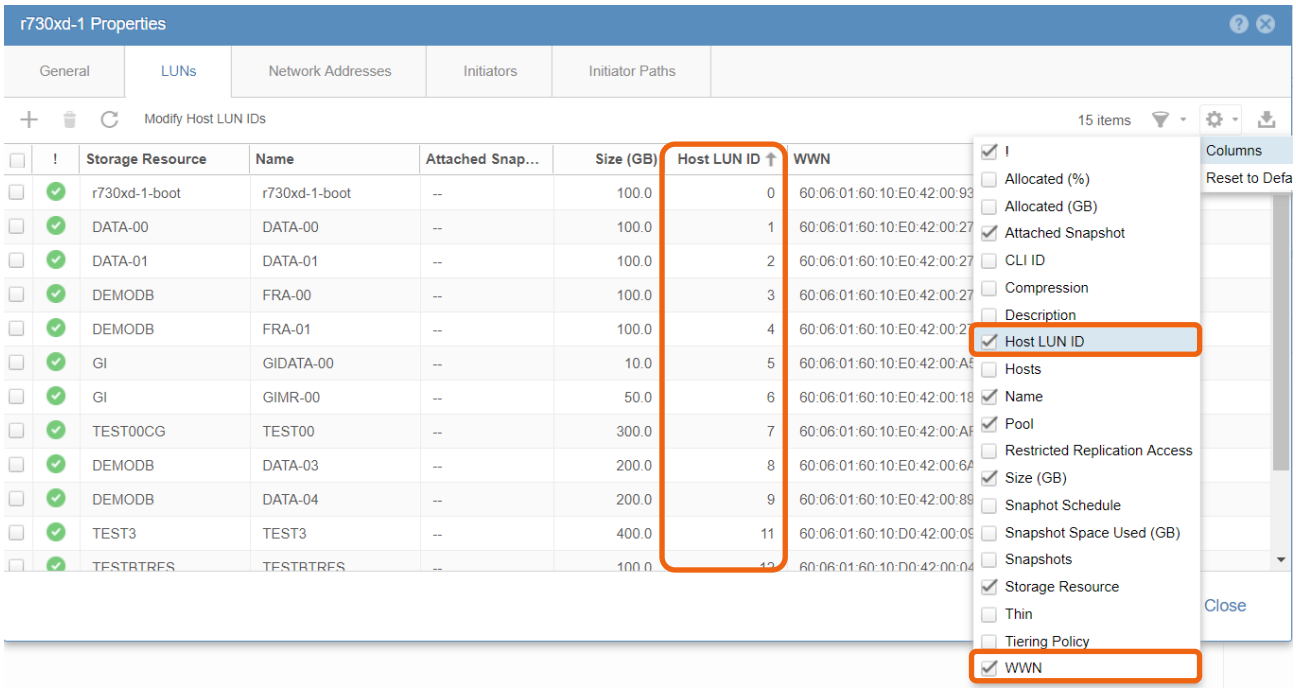

Figure 3 LUN ID information

4.1.1.2 Scanning for LUNs

Also known as a ghost LUN, LUNZ is not a real LUN but appears on the Linux operating system for the purpose of making the Dell EMC Unity system visible to the host when no LUNs are assigned to the host. LUNZ always takes up ID 0 on the host. Any I/Os sent down the LUNZ paths result in errors.

To see if there are any LUNZ on the system, run the **lsscsi** command.

```
# lsscsi|egrep DGC
[13:0:0:0]   disk    DGC      LUNZ            4201  /dev/sde
[13:0:1:0]   disk    DGC      LUNZ            4201  /dev/sdaf
[14:0:0:0]   disk    DGC      LUNZ            4201  /dev/sdd
[14:0:1:0]   disk    DGC      LUNZ            4201  /dev/sdag
[snipped]
```

When the Dell EMC Unity LUN has a LUN ID 0, use the **rescan-scsi-bus.sh** with the **--forcerescan** option which removes LUNZ and allows the real LUN 0 to show up on the host. For example:

```
# /usr/bin/rescan-scsi-bus.sh --forcerescan

# lsscsi |egrep DGC
[13:0:0:0]   disk    DGC      VRAID           4201  /dev/sde
[13:0:1:0]   disk    DGC      VRAID           4201  /dev/sdaf
[14:0:0:0]   disk    DGC      VRAID           4201  /dev/sdd
[14:0:1:0]   disk    DGC      VRAID           4201  /dev/sdag
[snipped]
```

When the Dell EMC Unity LUNs have non-zero IDs, use **-a** option instead.

```
# /usr/bin/rescan-scsi-bus.sh -a
```

Note: Omitting the **--forcerescan** option might prevent the operating system from discovering LUN 0 because of the LUNZ conflict.

4.1.1.3 Identifying LUNs by WWNs

The most accurate way to identify a LUN on the host operating system is by its WWN. The Dell EMC Unity system assigns a unique WWN for each LUN. The WWN information can be found in **Unisphere** > **Access** > **Hosts** > **Host Properties** > **LUNs**. If the **WWN** column is hidden from the default view, enable it through the columns filter. See Figure 3.

4.1.1.4 Querying WWNs using scsi_id command

To query the WWN on a Linux operating system, run the following commands against the device file.

Oracle Linux or RHEL 6.x

```
# /sbin/scsi_id --page=0x83 --whitelisted --device=<device>
```

Oracle Linux or RHEL 7.x

```
# /usr/lib/udev/scsi_id --page=0x83 --whitelisted --device=<device>
```

In these examples, **<device>** can be one of the following:

- Single path device (/dev/sde)
- Linux multipath device (/dev/mapper/mpathe)
- Dell EMC PowerPath™ device (/dev/emcpowerc)

The string returned by the **scsi_id** command indicates the WWN of the Dell EMC Unity LUN as shown in bold, appended with a 3.

36006016010e0420093a88859586140a5

4.1.1.5 Querying WWNs using multipath command

If the system has Linux device-mapper-multipath software enabled, the multipath command displays the multipath device properties including the WWN. For example:

```
# multipath -ll
mpatha (36006016010e0420093a88859586140a5) dm-0 DGC     ,VRAID
size=100G features='2 queue_if_no_path retain_attached_hw_handler' hwhandler='1
alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| |- 13:0:1:0   sdaf              65:240  active ready running
| |- 14:0:1:0   sdag              66:0    active ready running
| |- 15:0:1:0   sdbv              68:144  active ready running
| |- 16:0:1:0   sdcx              70:80   active ready running
| |- 17:0:1:0   sddz              128:16  active ready running
| `- 18:0:1:0   sdfb              129:208 active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  |- 13:0:0:0   sde               8:64    active ready running
  |- 14:0:0:0   sdd               8:48    active ready running
  |- 15:0:0:0   sdbh              67:176  active ready running
  |- 16:0:0:0   sdcj              69:112  active ready running
  |- 17:0:0:0   sddl              71:48   active ready running
  `- 18:0:0:0   sden              128:240 active ready running
```

4.1.1.6 Querying WWNs using powermt command

Similarly, when PowerPath is enabled on the system, the **powermt** command displays the multipath device properties including the WWN.

```
# powermt display dev=all
```

4.1.2 Multipathing

Multipathing is a software solution implemented at the host operating system level. While multipathing is optional, it provides path redundancy, failover, and performance-enhancing capabilities. Therefore, it is highly recommended to deploy the solution in a production environment or any environments where availability and performance are critical.

4.1.2.1 Main benefits of using an MPIO solution

- Increase database availability by providing automatic path failover and failback
- Enhance database I/O performance by providing automatic load balancing and capabilities for multiple parallel I/O paths
- Ease administration by providing persistent user-friendly names for the storage devices across cluster nodes

4.1.2.2 Multipath software solutions

There are a few multipath software choices available to choose from. It is up to the administrator's preference to decide which software solution is best for the environment. The following list provides a brief description of some of these solutions.

- Native Linux multipath (device-mapper-multipath)
- Dell EMC PowerPath™
- Symantec Veritas Dynamic Multipathing (VxDMP)

The native Linux multipath solution is supported and bundled with most popular Linux distributions in use today. Because the software is widely and readily available at no additional cost, many administrators prefer using it compared to other third-party solutions.

Unlike the native Linux multipath solution, both Dell EMC PowerPath and Symantec VxDMP provide extended capabilities for some storage platforms and software integrations. Both solutions also offer support for numerous operating systems in addition to Linux.

Only one multipath software solution should be enabled on the host and the same solution should be deployed in a cluster on all cluster hosts.

Refer to the vendor's multipath solution documentation for more information. For information on operating systems supported by Dell EMC PowerPath, see the Dell EMC Simple Support Matrix. Appendix C provides links to additional resources to these solutions.

4.1.2.3 Connectivity guidelines

The following list provides a summary of array-to-host connectivity best practices. It is important to review the documents, *Configuring Hosts to Access Fibre Channel (FC) or iSCSI Storage* and *Dell EMC Unity: High Availability*.

- Have at least two FC/iSCSI HBAs or ports to provide path redundancy.
- Connect the same port on both Dell EMC Unity storage processors (SP) to the same switch because the Dell EMC Unity system matches the physical port assignment on both SPs.
- Use multiple switches to provide switch redundancy.

4.1.2.4 Configuration file

To ease the deployment of the native Linux multipath software, it comes with a set of default settings for an extensive list of storage models from different vendors including the Dell EMC Unity system. The default settings allow the software to work with the Dell EMC Unity system right out of the box. However, these settings might not be optimal for all situations and should be reviewed and modified if necessary.

The multipath daemon configuration file needs to be created on newly installed systems. A basic template can be copied from **/usr/share/doc/device-mapper-multipath-<*version*>/multipath.conf** to **/etc/multipath.conf** as a starting point. Any settings that are not defined explicitly in /etc/multipath.conf would assume the default

values. The full list of settings (explicitly set and default values) can be obtained using the following command. Specific Dell EMC Unity settings can be found by searching for **DGC** from the output. The default settings generally work without any issues.

```
# multipathd -k"show config"
```

4.1.2.5 Creating aliases

It is generally a good idea to assign meaningful names (aliases) for the multipath devices though it is not mandatory. For example, create aliases based on the application type and environment it is in. The following snippet in the multipaths section assigns an alias of **ORA-DATA-00** to the Dell EMC Unity LUN with the WWN **36006016010e04200271a8a594a34d845**.

```
multipaths {
        multipath {
         wwid "36006016010e04200271a8a594a34d845"
         alias ORA-DATA-00
       }
}
```

4.1.2.6 Asymmetric Logic Unit Access

Dell EMC Unity systems support Asymmetric Logic Unit Access (ALUA) for host access. This allows the host operating system to recognize optimized paths from non-optimized paths. Optimized paths are the ones connected to the LUN's SP owner and they are assigned a higher priority. The default multipath settings reflect the support of **ALUA** feature on Dell EMC Unity storage. The following example shows LUN **ORA-DATA-00** has a total of 12 paths and divided into two groups. The optimized paths have a priority of 50 and non-optimized paths have a priority of 10.

```
e.g. multipath -ll shows groups of paths with different priority
# multipath -ll ORA-DATA-00
ORA-DATA-00 (36006016010e04200271a8a594a34d845) dm-18 DGC     ,VRAID
size=100G features='2 queue_if_no_path retain_attached_hw_handler' hwhandler='1
alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| |- 13:0:0:1   sdf                8:80    active ready running
| |- 14:0:0:1   sdg                8:96    active ready running
| |- 15:0:0:1   sdbi               67:192  active ready running
| |- 16:0:0:1   sdck               69:128  active ready running
| |- 17:0:0:1   sddm               71:64   active ready running
| `- 18:0:0:1   sdeo               129:0   active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  |- 13:0:1:1   sdah               66:16   active ready running
  |- 14:0:1:1   sdai               66:32   active ready running
  |- 15:0:1:1   sdbw               68:160  active ready running
  |- 16:0:1:1   sdcy               70:96   active ready running
  |- 17:0:1:1   sdea               128:32  active ready running
  `- 18:0:1:1   sdfc               129:224 active ready running
```

I/Os are sent down the optimized paths when possible. If I/Os are sent down the non-optimized paths, they are redirected by the peer SP to the primary SP through the internal bus. When the Dell EMC Unity system senses a large amount of non-optimized I/Os, it automatically trespasses the LUN from the primary SP to the peer SP to optimize the data paths.

4.1.3 LUN partition

A LUN can be used as a whole or it can be divided into multiple partitions. Certain applications, such as Oracle ASMLib, recommend partitioning over whole LUNs. Dell EMC recommends configuring whole LUNs without partitions wherever appropriate because it offers the most flexibility for configuring and managing the underlying storage.

See section 4.4.4 on choosing a strategy to grow storage space.

4.1.3.1 Partition alignment

When partitioning a LUN, it is recommended to align the partition on the 1M boundary. Either **fdisk** or **parted** can be used to create the partition. However, only **parted** can create partitions larger than 2 TB.

4.1.3.2 Creating partition using parted

Before creating the partition, label the device as GPT. Then, specify the partition offset at 2048 sector (1M). The following command creates a single partition that takes up the entire LUN. Once the partition is created, the partition file **/dev/mapper/orabin-std1** should be used for creating file system or ASMLib volume.

```
# parted /dev/mapper/orabin-std
GNU Parted 3.1
Using /dev/mapper/orabin-std
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted) mklabel gpt
(parted) quit
Information: You may need to update /etc/fstab.

# parted /dev/mapper/orabin-std mkpart primary 2048s 100%
Warning: The resulting partition is not properly aligned for best performance.
Ignore/Cancel? Ignore
Information: You may need to update /etc/fstab.
```

Note: A misalignment warning might appear. This is a known issue which can be safely ignored.

4.1.4 Partitioned devices and filesystems

When creating a filesystem, create the filesystem on properly aligned partitioned device.

```
# mkfs.ext4 /dev/mapper/orabin-std1
```

4.1.5 I/O scheduler for Oracle ASM devices

Oracle recommends using the deadline I/O scheduler for the best performance of Oracle ASM. For Oracle Linux, the deadline I/O scheduler is enabled by default in Oracle Unbreakable Enterprise Kernel. For other Linux operating systems, verify the I/O scheduler and make necessary updates if necessary.

To verify the I/O schedule use the following commands:

```
# egrep "*" /sys/block/sd*/queue/scheduler
/sys/block/sdaa/queue/scheduler:noop [deadline] cfq
/sys/block/sdab/queue/scheduler:noop [deadline] cfq
/sys/block/sdac/queue/scheduler:noop [deadline] cfq
/sys/block/sdad/queue/scheduler:noop [deadline] cfq
```

```
/sys/block/sdae/queue/scheduler:noop [deadline] cfq
/sys/block/sdaf/queue/scheduler:noop [deadline] cfq
```

To set the I/O schedule persistently, create an udev rule that updates the devices. See section 4.2.1.2 for more information about using udev to set persistent ownership and permission.

The following example shows setting the deadline I/O scheduler on all /dev/sd* devices. The rule is appended to the 99-oracle-asmdevices.rule file.

```
# cat /etc/udev/rules.d/99-oracle-asmdevices.rules
ACTION=="add|change", KERNEL=="sd*", RUN+="/bin/sh -c '/bin/echo deadline >
/sys$env{DEVPATH}/queue/scheduler'"

# udevadm control --reload-rules
# udevadm trigger
```

4.2 Oracle Automatic Storage Management

Dell EMC and Oracle recommend using Oracle Automatic Storage Management (ASM) to manage Dell EMC Unity LUNs for the database and clusterware. This section reviews the general guidelines and additional considerations for an Oracle database.

4.2.1 Preparing storage for Oracle ASM

Proper user and group ownership, and permissions must be ensured on any Dell EMC Unity LUNs that are going to be used by Oracle ASM. The LUNs should be owned by the owner of the ASM instance and have read/write privilege to them. For example, if user **grid** with primary group **oinstall** is the owner of the ASM instance, **grid:oinstall** should be assigned to the LUNs. There are different methods to set the ownership and permissions and keep these settings persistent across host reboot.

4.2.1.1 Persistent device ownership and permissions

Persistent device ownership and permission can be managed through a variety of software. The following describes some of the commonly used software on Linux host.

- Linux dynamic device management (udev)
- Oracle ASMLib
- Oracle ASMFD

4.2.1.2 Linux dynamic device management (udev)

The Linux udev facility comes with every Linux distribution and is easy to set up for persistent device ownership and permission by creating rules in the udev rule file. System rule files are located in the **/usr/lib/udev/rules.d** directory and user-defined rule files are located in **/etc/udev/rules.d**. There are many ways to define a device in the rule file. Two examples are provided as follows.

Example 1: Set device ownership and permission by WWNs

Define a rule for each Dell EMC Unity LUN using its unique WWN. With this approach, each LUN requires an udev rule. The rule file is located in **/etc/udev/rules.d/99-oracle-asmdevices.rules**. The following example shows an udev rule that sets **grid:oinstall** ownership and **660** permission on a **dm** (multipath) device that matches the WWN 36006016010d04200b584ce59557ba84a.

```
# cat /etc/udev/rules.d/99-oracle-asmdevices.rules
KERNEL=="dm-*",PROGRAM=="/lib/udev/scsi_id --whitelisted --
device=/dev/$name",RESULT=="36006016010d04200b584ce59557ba84a",ACTION=="add|chan
ge",OWNER="grid",GROUP="oinstall",MODE="0660"
…
…
…

# udevadm control --reload-rules
# udevadm trigger
```

If PowerPath is used, change **KERNEL=="dm-*"** to **KERNEL=="emcpower*".**

Example 2: Set device ownership and permission by device name pattern

The udev rule can be simplified if multipath device aliases are created with a consistent string pattern. For example, use the **ORA-** prefix in all multipath device aliases for LUNs that would be used by Oracle ASM. A single udev rule can then be used to set ownership and permission on all ORA* multipath devices. See section 4.1.2.5 on creating a multipath device alias.

```
# cat /etc/udev/rules.d/99-oracle-asmdevices.rules
KERNEL=="dm-*",ENV{DM_NAME}=="ORA*",OWNER="grid",GROUP="oinstall",MODE="0660"

# udevadm control --reload-rules
# udevadm trigger
```

The advantage of this approach is that only multipath.conf needs to be updated when new LUNs are added to the system for Oracle ASM.

4.2.1.3 Oracle ASMLib

Oracle ASMLib simplifies storage management and reduces kernel resource usage. It provides device file name, ownership, and permission persistency and reduces the number of open file handles required by the database processes. No udev is required when ASMLib is used.

When LUNs are initialized with ASMLib, special device files are created in the **/dev/oracleasm/disks** folder with proper ownership and permission automatically. When the system reboots, the ASMLib driver restarts and recreates the device files. ASMLib consists of three packages:

- oracleasm-support-*version*.*arch*.rpm
- oracleasm-*kernel*-*version*.*arch*.rpm
- oracleasmlib-*version*.*arch*.rpm

The kernel driver rpm (oracleasm-kernel-version.arch.rpm) is maintained separately by each Linux vendor. With Oracle Linux, the kernel driver is already included with Oracle Linux Unbreakable Enterprise Kernel. For more information about ASMLib and to download the software, visit http://www.oracle.com/technetwork/topics/linux/asmlib/index-101839.html.

The ownership of the ASMLib devices are defined in the **/etc/sysconfig/oracleasm** configuration file which is generated by running **/etc/init.d/oracleasm** configure initially. Update the configuration file if necessary to reflect the proper ownership and the disk scanning order.

```
# cat/etc/sysconfig/oracleasm
# ORACLEASM_ENABLED: 'true' means to load the driver on boot.
ORACLEASM_ENABLED=true

# ORACLEASM_UID: Default user owning the /dev/oracleasm mount point.
ORACLEASM_UID=grid

# ORACLEASM_GID: Default group owning the /dev/oracleasm mount point.
ORACLEASM_GID=oinstall

# ORACLEASM_SCANBOOT: 'true' means scan for ASM disks on boot.
ORACLEASM_SCANBOOT=true

# ORACLEASM_SCANORDER: Matching patterns to order disk scanning
ORACLEASM_SCANORDER="dm"

# ORACLEASM_SCANEXCLUDE: Matching patterns to exclude disks from scan
ORACLEASM_SCANEXCLUDE="sd"
```

This configuration file indicates **grid:oinstall** for the ownership and it searches for multipath devices (**dm)** and excludes any single path devices (**sd**). If PowerPath devices are used, set **ORACLEASM_SCANORDER="emcpower".**

Note: The asterisk (*) cannot be used in the value for **ORACLEASM_SCANORDER** and **ORACLEASM_SCANEXCLUDE**.

Oracle requires the LUNs to be partitioned for ASMLib use. First, create a partition with parted, then use oracleasm to label the partition. ASMLib does not provide multipath capability and relies on native or third-party multipath software to provide the function. The following example shows creating an ASMLib device on a partition of a Linux Multipath device. The **oracleasm** command writes the ASMLib header to **/dev/mapper/mpathap1** and generates the ASMLib device file in **/dev/oracleasm/disks/DATA01** with ownership as indicated in the **/etc/sysconfig/oracleasm** file.

```
# oracleasm createdisk DATA01 /dev/mapper/mpathap1
```

4.2.1.4 Oracle ASM Filter Driver

Oracle ASM Filter Driver (ASMFD) is a kernel module that sits between the operating system kernel and Oracle ASM. Oracle intends to replace Oracle ASMLib with ASMFD and recommends using ASMFD in Oracle 12c and above. ASMFD includes all the ASMLib benefits of storage device name, ownership, and permission persistency, and better kernel usage by reducing the number of open file handles. Additionally, it provides storage protection by rejecting non-oracle I/Os and hence prevents inadvertent overwrite of the ASM disks.

In a cluster environment, without ASMFD, when a cluster node is fenced, the host must be rebooted to ensure the integrity of the data. With ASMFD, the fenced node does not need to be rebooted. It is possible to restart the clusterware stack which reduces the time to recover the node.

Unlike ASMLib, ASMFD comes with the Grid Infrastructure software and there is no additional software to download. Starting with Oracle ASM 12c Release 2, the installation and configuration for Oracle ASMFD have been simplified by integrating the option into the Oracle Grid Infrastructure installation. Administrators need to select the option **Configure Oracle ASM Filter Driver** during the Grid installation.

The installation of ASMFD automatically creates an udev rule file in **/etc/udev/rules.d/53-afd.rules** that sets the afd devices with the proper ownership and permission. Do not attempt to modify or delete this file directly. Use the **asmcmd adf_configure** command to make updates instead.

```
# cat /etc/udev/rules.d/53-afd.rules
#
# AFD devices
KERNEL=="oracleafd/.*", OWNER="grid", GROUP="oinstall", MODE="0775"
KERNEL=="oracleafd/*", OWNER="grid", GROUP="oinstall", MODE="0775"
KERNEL=="oracleafd/disks/*", OWNER="grid", GROUP="oinstall", MODE="0664"
```

Either whole LUNs or LUN partitions can be used for ASMFD devices. Dell EMC recommends using whole LUNs because of certain restrictions with partitions which affect database availability during storage expansion. See section 4.2.4 for more details.

The following example shows creating an ASMFD device on a Linux multipath device. The **asmcmd afd_label** command writes the ASMFD header to **/dev/mapper/mpathb** and generates the ASMFD device **file in /dev/oracleafd/disks/DATA01**. The udev rule then ensures the afd devices are set to **grid:oinstall** and 0664 permission.

```
# asmcmd afd_label DATA01 /dev/mapper/mpatha
```

The other advantage of using ASMFD is that it supports thin-provisioned disk group starting in Oracle release 12.2.0.1.

To find out which OS platforms ASMFD is supported on, see Oracle KB Doc ID 2034681.1 at Oracle Support.

For more information on installing and configuring ASMFD, refer to the Oracle Automatic Storage Management Administrator's Guide.

4.2.2 Setting the asm_diskstring ASM instance parameter

The **asm_diskstring** ASM instance parameter tells ASM the location of the ASM devices. During the Grid Infrastructure installation, it defaults to null and it should be updated to reflect the correct location of the device files.

Table 4 Example of asm_diskstring settings

Device files	asm_diskstring setting
Linux native multipath	asm_diskstring='/dev/mapper/ORA*'
Dell EMC PowerPath	asm_diskstring='/dev/emcpower*'
Oracle ASMLib	asm_diskstring='ORCL:*'
Oracle ASMFD	asm_diskstring='AFD:*'

4.2.3 Oracle ASM guidelines

Dell EMC and Oracle recommend using Oracle ASM as the preferred storage management solution for either a single-instance database or Real Application Clusters (RAC). ASM takes place of the traditional Linux volume manager and file system. It takes over the management of the disks and creates disk groups where data files reside.

4.2.3.1 Benefits of using Oracle ASM

ASM offers many advantages over the traditional Linux storage management solution such as Logical Volume Manager (LVM). The main benefits include:

- Automatic file management
- Online data files rebalance across ASM disks
- Online addition and removal of ASM disks without downtime
- Single solution for both volume and file management integrated with Oracle software
- Improved I/O performance because ASM stripes all files across all disks in a disk group
- Seamless and transparent integration with Dell EMC Unity system features such as snapshots, thin-provisioning, thin clones, compression, and Data at Rest Encryption

4.2.3.2 ASM disk and disk group guidelines

When creating an Oracle ASM disk group, consider the following guidelines:

- For ultimate flexibility and maintaining configuration consistency, create separate disk groups for each of the following:
  - Create a disk group for the Oracle Cluster Registry (OCR) and voting files.
  - Create a disk group for Grid Infrastructure Management Repository (GIMR).
  - Use one or more disk groups for database data files for each database.
  - Use a disk group for a fast recovery area for each database.
  - Configure a database which can span across multiple disk groups but with each disk group mounted and used by one database exclusively. This provides the ability to independently optimize the storage and snapshot configuration for each individual database.
- Create LUNs with same capacity and services in the same disk group such as compression, consistency group, and snapshot schedule
- Use fewer but larger LUNs to reduce the number of objects to be managed
- Create a minimum of two LUNs for each disk group and distributing the LUNs evenly on both Dell EMC Unity storage processors to allow even I/O distribution to both processors, hence, maximizing the performance and I/O bandwidth for the environment.
- In order to take an array-based snapshot on a multi-volume Oracle database, make sure that all LUNs belonging to the same database are snapped together. This can be achieved by grouping the LUNs in a Consistency Group (see more information in section 4.2.3.3).
- While ASM can provide software-level mirroring, it is not necessary because data protection has already been provided by the built-in Dell EMC Unity RAID protection. Use External Redundancy for ASM disk groups to enable substantial storage savings, reduce overall IOPS from ASM, and results in better I/O performance.
- For best storage efficiency, create thin-provisioned LUNs in the Dell EMC Unity system for ASM use. When creating the tablespaces and data files on the ASM disk groups, administrators can set an initial size of each data file and specify the **autoextend** clause to include an extent size for growth. The Dell EMC Unity system allocates storage for the initial data file size and as the data are written to

the data files, additional space is allocated by the amount of autoextend size. An example of the **CREATE TABLESPACE** statement is shown in the following:

```
SQL> create tablespace DATATS datafile '+DATADG' size 10G autoextend on
next 1024M maxsize unlimited;
```

- By default, each LUN has unlimited I/O limits in the Dell EMC Unity system. When a certain database requires higher performance and another does not, consider creating different host I/O limit policies in Dell EMC Unisphere that limit I/O performance based on IOPS and bandwidth. Assign the policy to the LUNs corresponding to the level of performance required. The host I/O limit is applied on the LUN level.
- On Oracle 12c releases, ASMFD now supports thin provision ASM disk group. The feature allows unused space to be released back to the Dell EMC Unity system after deleting or shrinking the data files. To enable the feature, set **THIN_PROVISIONED** attribute to **'TRUE'** on the disk group. For example:

  ```
  SQL> ALTER DISKGROUP DATADG SET ATTRIBUTE 'THIN_PROVISIONED'='TRUE';
  ```

- When ASM rebalances the disk group, at the end of the rebalance is the compact phase in which the data is moved to the higher performing tracks of the spinning disks. Since the Dell EMC Unity system virtualizes the physical storage devices, and with the use of the flash devices, there is no real benefit to compacting the data. In Oracle 12c, it is now possible to disable the compact phase on individual disk group by setting the **_rebalance_compact** attribute to **'FALSE'**.

  ```
  SQL > ALTER DISKGROUP DATADG SET ATTRIBUTE '_rebalance_compact'='FALSE';
  ```

  For Oracle pre-12c releases, this can only be disabled on the ASM instance level which affects all disk groups. For database environments that have a mix of storage types, turning off the compact phase might have adverse performance implication.

  For more information on ASM compact phase rebalancing, see Oracle KB Doc ID 1902001.1 on Oracle Support.

Table 5 demonstrates an example of how ASM disk groups are organized. Figure 4 illustrates the storage layout on the database, ASM disk group, and Dell EMC Unity system levels.

Table 5 Example ASM disk group configuration

Database	ASM disk group	Number of LUNs	LUN size	Dell EMC Unity consistency group	Description
Clusterware	GIDATA	2	10 GB	N/A	Clusterware-related information such as the OCR and voting disks
Grid Infrastructure Management Repository	MGMT	2	50 GB	mgmt_cg	In 12cR2, a separate disk group created for the GI Management Repository data
Test database (testdb)	DATADG	2	200 GB	testdb_cg	Disk group that holds the database files, temporary table space, and online redo logs; contains system-related table spaces such as SYSTEM and UNDO Contains only testdb data
	FRADG	2	100 GB		Disk group that holds the database archive logs and backup data Contains only testdb logs
Development database (devdb)	DATA2DG	2	200 GB	devdb_cg	Disk group that holds the database files, temporary table space, online redo logs; contains system-related table spaces such as SYSTEM and UNDO Contains only devdb data
	FRA2DG	2	100 GB		Disk group that holds the database archive logs and backup data Contains only devdb logs

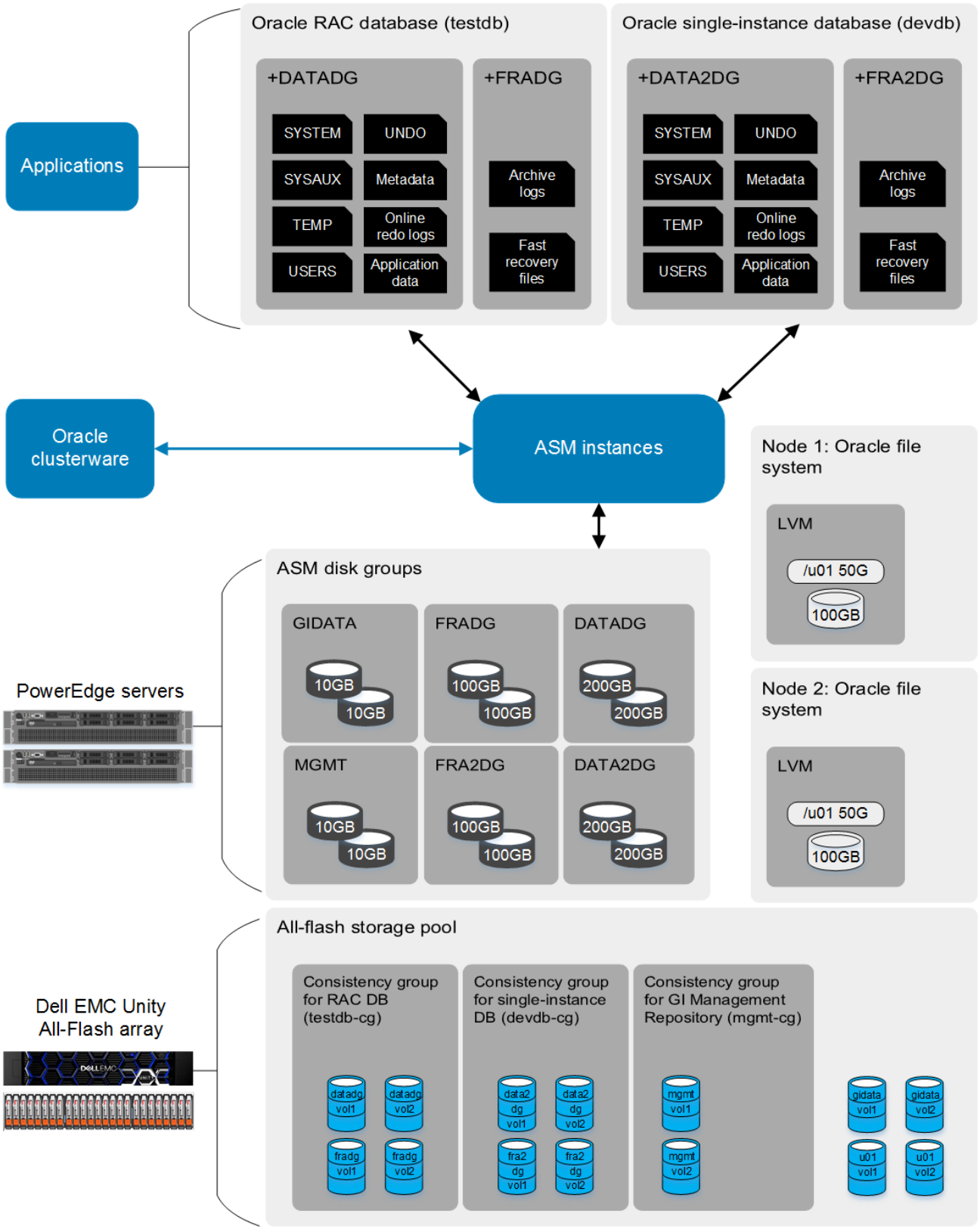

Figure 4 Oracle ASM storage layout on the Dell EMC Unity system

4.2.3.3 Consistency group

For performance reasons, it is very common for a database to span across multiple LUNs to increase I/O parallelism to the storage devices. Dell EMC recommends grouping the LUNs into a consistency group for a database to ensure data consistency when taking storage snapshots. The Dell EMC Unity system snapshot feature is a quick and space-efficient way to create a point-in-time snapshot of the entire database. Sections 8.3 and 8.4 discuss using Dell EMC Unity system snapshots and thin clones to reduce database recovery time and create space-efficient copies of the database.

In Figure 4, for example, the RAC database consists of disk group +DATADG and +FRADG. Therefore, all ASM volumes in those disk groups are configured in a single consistency group, testdb_cg. Likewise, the single instance database consists of disk groups +DATA2DG and +FRA2DG. The ASM devices of both disk groups are configured in a consistency group, devdb_cg.

The consistency group feature allows taking a database-consistent snapshot across multiple LUNs. On the database side, use the **ALTER DATABASE BEGIN BACKUP** clause before the snapshot is taken and **END BACKUP** clause after the snapshot is taken.

Note: Storage snapshots taken on a multiple-LUN database without a consistency group might be irrecoverable by Oracle during database recovery.

4.2.4 Expand Oracle ASM storage

As the storage consumption grows over time, it is necessary to increase and grow the existing storage capacity both in the Dell EMC Unity system and in the database. It is most desirable to add capacity online with minimal business interruptions. The Dell EMC Unity system has the flexibility to expand the current storage system with no interruption to the application. The following non-disruptive operations can be performed online in Unisphere:

- Adding flash devices
- Expanding the storage pool
- Increasing the size of existing LUNs
- Creating and adding new LUNs to existing hosts

The following subsections discuss the different ways to increase ASM storage capacity. Each method has its pros and cons.

4.2.4.1 Increase Oracle ASM storage by adding new LUNs

Additional storage capacity can be added to an ASM disk group by adding new LUNs to the disk group. The advantage of this method is that the process is relatively simple and safe because no changes are made to the existing LUNs.

The following outlines the general process:

1. Create new LUNs in Unisphere.
2. Ensure the size of new LUNs and other features such as compression, and that the consistency group matches the existing LUNs.
3. Allow access to new LUNs to the host systems.
4. Perform a SCSI scan on the host systems (see section 4.1.1).
5. Configure multipath for the new devices (see section 4.1.2).
6. Prepare the LUNs for ASM (see section 4.2.1).
7. Add the LUNs to the ASM disk group.

Since ASM automatically rebalances the data after new LUNs are added, it is recommended to add the LUNs in a single operation to minimize the amount of rebalancing work. The following example shows the **ALTER DISKGROUP ADD DISK** statement to add multiple devices to a disk group.

```
ALTER DISKGROUP DATADG ADD DISK 'AFD:DATADG_VOL1', 'AFD:DATADG_VOL2'
REBALANCE POWER 10 NOWAIT;
```

8. Verify the status and capacity of the disk group.

```
# asmcmd lsdsk -gk -G datadg
# asmcmd lsdg –g datadg
```

9. If the existing LUNs are in a consistency group, add the new LUNs to the same consistency group.

Note: Adding or removing LUNs in a consistency group is not allowed when there are existing snapshots of the consistency group. To add or remove LUNs in a consistency group, delete all snapshots and retry the operation.

4.2.4.2 Increase Oracle ASM storage by resizing current LUNs

The Dell EMC Unity system can extend the size of existing LUNs online. However, depending on the operating system, disk partition configuration, and Oracle software chosen, resizing ASM disks online might not be possible. Table 6 summaries the online resize capability on some configurations. It does not cover all possible configuration variations. Customers should consult with each vendor to fully understand the capability and limitation of their software.

Table 6 Resize Oracle ASM device online support matrix

Oracle version	Without ASMLib and ASMFD using non-partition LUNs	ASMFD using non-partition LUNs	ASMLib using partition LUNs
12.2.0.1	Yes	Yes	No
12.1.0.1	Yes	No	No
11.2.0.4	Yes	No	No

Note: Resizing LUNs on the OS can cause loss of data or corruption. It is recommended to back up all data before attempting to resize the LUNs.

4.2.4.3 Resize ASM devices without ASMFD and ASMLib online

Without the use of ASMFD or ASMLib, and only using whole LUNs (without partition), it is possible to resize the devices online on a wide range of OS and Oracle versions. See Table 6.

The following outlines the general steps to resize ASM devices online without ASMFD and ASMLib.

1. Take manual snapshots of LUNs that are going to be expanded. See section 8.3 for more information on taking snapshots and recovering from snapshots.
2. Expand the size of existing LUNs in Unisphere.
3. Perform a SCSI scan on the host systems and refresh partition table on each LUN path and reload multipath devices.

4. Reload multipath devices.

```
# multipathd -k"resize map DATA03"
```

For PowerPath, the new size is automatically updated.

5. Verify the new LUN size.

```
# multipath -ll ORA-TEST3 | egrep size
size=400G features='2 queue_if_no_path retain_attached_hw_handler'
hwhandler='1 alua' wp=rw

# multipath -ll ORA-TEST3 | awk '/sd/ {print $(NF-4)}' | xargs -i fdisk -l
/dev/{} | egrep "^Disk"
Disk /dev/sdfd: 429.5 GB, 429496729600 bytes, 838860800 sectors
Disk /dev/sdff: 429.5 GB, 429496729600 bytes, 838860800 sectors
Disk /dev/sdfh: 429.5 GB, 429496729600 bytes, 838860800 sectors
[snipped]
```

For PowerPath, use the following commands:

```
# fdisk –l /dev/emcpowerc

# powermt display dev=all|awk '/sd/ {print $3}'|xargs -i fdisk -l
/dev/{}|egrep "^Disk"
```

6. To determine the maximum size of the LUN, run **asmcmd lsdsk** and extract the **OS_MB** value. Use this value with the **ALTER DISKGROUP RESIZE DISK** clause.

```
# asmcmd lsdsk –k

Inst_ID  Total_MB  Free_MB   OS_MB  Name          Failgroup     Site_Name
Site_GUID                          Site_Status  Failgroup_Type  Library
Label  Failgroup_Label  Site_Label  UDID  Product  Redund   Path
      1    204800   409464  409600  TEST3DG_0000  TEST3DG_0000
00000000000000000000000000000000              REGULAR         System
UNKNOWN  /dev/mapper/ORA-TEST3
      2    204800   409464  409600  TEST3DG_0000  TEST3DG_0000
00000000000000000000000000000000              REGULAR         System
UNKNOWN  /dev/mapper/ORA-TEST3
```

Total_MB represents the current size before the resize operation.

OS_MB represents the new maximum size ASM can expand to.

7. Resize the ASM device.

```
SQL> ALTER DISKGROUP TEST3DG RESIZE DISK TEST3DG_0000 SIZE 409600M
REBALANCE POWER 10;
```

8. Verify the new ASM device size. After the resize operation completes, run **asmcmd lsdsk** to confirm the **Total_MB** value matches **OS_MB** value.

```
# asmcmd lsdsk -k
Inst_ID  Total_MB  Free_MB   OS_MB  Name          Failgroup     Site_Name
Site_GUID                          Site_Status  Failgroup_Type  Library
Label  Failgroup_Label  Site_Label  UDID  Product  Redund   Path
      1    409600   409464  409600  TEST3DG_0000  TEST3DG_0000
00000000000000000000000000000000              REGULAR         System
UNKNOWN  /dev/mapper/ORA-TEST3
      2    409600   409464  409600  TEST3DG_0000  TEST3DG_0000
00000000000000000000000000000000              REGULAR         System
UNKNOWN  /dev/mapper/ORA-TEST3
```

Run **asmcmd lsdg** to confirm the **Total_MB** value on the disk group has increased.

```
# asmcmd lsdg
Inst_ID  State    Type    Rebal  Sector  Logical_Sector  Block       AU
Total_MB  Free_MB  Req_mir_free_MB  Usable_file_MB  Offline_disks
Voting_files  Name
      1  MOUNTED  EXTERN  N         512             512   4096  4194304
409600   409464                0          409464              0
N  TEST3DG/
      2  MOUNTED  EXTERN  N         512             512   4096  4194304
409600   409464                0          409464              0
N  TEST3DG/
```

4.2.4.4 Resize ASM devices with ASMFD

When Oracle 12.2 ASMFD is used with ASM device, it is possible to resize the ASM device online without impacting the database. For Oracle version 12.1, the **afd_refresh** option that is responsible to refresh devices with new device size is not available, and there is no ability to use the feature with version 12.1 at the date when this paper is published.

Note: The **afd_refresh** option is only available in Oracle 12.2.

Resize ASM devices online:

The following outlines the general steps to resize ASM devices with ASMFD online.

1. Take manual snapshots of LUNs that are going to be expanded. See section 8.3 for more information on taking snapshots and recovering from snapshots.
2. Expand the size of existing LUNs in Unisphere.
3. Perform a SCSI scan on the host systems and refresh partition table on each LUN path and reload multipath devices.
4. Reload multipath devices.

```
# multipathd -k"resize map ORA-TEST3"
```

For PowerPath, the new size is automatically updated.

5. Verify the new LUN size.

```
# multipath -ll ORA-TEST3 | egrep size

# multipath -ll ORA-TEST3 | awk '/sd/ {print $(NF-4)}' | xargs -i fdisk -l
/dev/{} | egrep "^Disk"
```

   For PowerPath, use the following commands:

```
# fdisk –l /dev/emcpowerc

# powermt display dev=all|awk '/sd/ {print $3}'|xargs -i fdisk -l
/dev/{}|egrep "^Disk"
```

6. Refresh the ASMFD devices.

```
# asmcmd afd_refresh
```

7. To determine the maximum size of the LUN, run **asmcmd lsdsk** and extract the OS_MB value. Use this value with the **ALTER DISKGROUP RESIZE DISK** clause.

```
# asmcmd lsdsk –k
```

8. Resize the ASM device.

```
SQL> ALTER DISKGROUP DATADG RESIZE DISK DATA03 SIZE $OS_MB REBALANCE POWER
10;
```

9. Verify the new size in the ASM device and disk group. After the resize operation completes, run **asmcmd lsdsk** to confirm the Total_MB value matches OS_MB value.

```
# asmcmd lsdsk -k
```

   Run **asmcmd lsdg** to confirm the Total_MB value on the disk group has increased.

```
# asmcmd lsdg
```

Resize ASM devices offline:

As mentioned previously in this section, the online resize capability of ASMFD is available in Oracle 12.2. When running Oracle 12.1, either restart the host to refresh the LUN size, or restart the clusterware, the ASM instance, and the AFD driver on the host to minimize the outage window. In a cluster environment, this can be done in a rolling fashion to further minimize the impact of the outage.

1. Follow step 1 to step 5 in section 4.2.4.3.
2. Stop all databases on the host.
3. Stop CRS.

```
# crsctl stop crs
```

4. Reload the AFD driver.

```
# afdload stop
# afdload start
```

5. Rescan the AFD devices.

```
# asmcmd afd_scan
```

6. Restart CRS.

```
# crsctl start crs
```

7. Restart the databases.
8. Continue step 6 to step 8 in section 4.2.4.3.
9. Repeat the process on other cluster nodes.

Another alternative to restarting the node or software is to unlabel and label the AFD devices. The database associated with the devices must be stopped, and the disk groups and devices must be offlined before they can be relabeled. This approach increases the risk of data loss and corruption and requires exercise extra caution.

4.2.4.5 Resize an ASM device with ASMLib

Oracle recommends partitioning LUNs for ASMLib. To increase the size of the partition after expanding the LUN, the partition is first removed and then recreated with the new size. Therefore, the database associated with the device would be impacted.

The following outlines the general steps to resize ASM devices with ASMFD online.

1. Take manual snapshots of LUNs that are going to be expanded. See section 8.3 for more information on taking snapshots and recovering from snapshots.
2. Expand the size of the existing LUNs in Unisphere.
3. Perform a SCSI scan on the host systems, refresh the partition table on each LUN path, and reload the multipath devices.

```
# rescan-scsi-bus.sh --resize
```

4. Reload the multipath devices.

```
# multipathd -k"resize map TEST67_ASMLIB"
```

For PowerPath, the new size is automatically updated.

5. Verify the new LUN size

```
# multipath -ll TEST67_ASMLIB | egrep size

# multipath -ll TEST67_ASMLIB | awk '/sd/ {print $(NF-4)}' | xargs -i
fdisk -l /dev/{} | egrep "^Disk"
```

For PowerPath, use the following commands.

```
# fdisk –l /dev/emcpowerc

# powermt display dev=all|awk '/sd/ {print $3}'|xargs -i fdisk -l
/dev/{}|egrep "^Disk"
```

6. Stop the database across the cluster.

```
$ srvctl stop db -d demodb
```

7. Dismount the disk group on all cluster nodes.

```
SQL> ALTER DISKGROUP TEST67_ASMLIBDG DISMOUNT;
```

8. Remove and recreate the partition.

```
# parted /dev/mapper/TEST67_ASMLIB rm 1
# parted /dev/mapper/TEST67_ASMLIB mkpart primary 2048s 100%
```

9. Rescan the LUNs on all cluster nodes.

```
# rescan-scsi-bus.sh –resize
```

10. Refresh the multipath device on all cluster nodes.

```
# multipathd -k"resize map TEST67_ASMLIB"
```

For PowerPath, the new size is automatically updated.

11. Update the partition on all cluster nodes.

```
# partprobe /dev/mapper/TEST67_ASMLIB

# parted /dev/mapper/TEST67_ASMLIB u GB p
```

`parted` command should show the new size.

12. Rescan the ASMLib devices on all cluster nodes.

```
# oracleasm scandisks
# oracleasm listdisks
```

13. Mount the disk group on all cluster nodes.

```
SQL> ALTER DISKGROUP TEST67_ASMLIBDG MOUNT;
```

14. Run **asmcmd lsdsk** to extract the **OS_MB** value to determine the maximum LUN size.

```
# asmcmd lsdsk –k –g –G TEST67_ASMLIBDG
```

15. Resize the ASM device.

```
SQL> ALTER DISKGROUP TEST67_ASMLIB RESIZE DISK TEST67_ASMLIB SIZE $OS_MB
REBALANCE POWER 10;
```

16. Verify the new size in the ASM device and disk group. After the resize operation completes, run **asmcmd lsdsk** to confirm the **Total_MB** value matches the **OS_MB** value.

```
# asmcmd lsdsk -k
```

Run **asmcmd lsdg** to confirm the **Total_MB** value on the disk group has increased.

```
# asmcmd lsdg
```

17. Restart the database.

```
$ srvctl start db -d demodb
```

4.2.5 Space reclamation

The Dell EMC Unity system supports the SCSI TRIM/UNMAP feature which allows operating systems to inform which data blocks are no longer in use and can be released for other uses. For space reclamation to work, the LUNs must be thin provisioned in the Dell EMC Unity system and the Linux kernel, and Oracle ASM must also support the feature. The TRIM/UNMAP feature has been introduced in Linux kernel 2.6.28-25 and newer. With Oracle 12.2 ASMFD, thin-provisioned ASM diskgroups allow deleted space in data files to be reclaimed.

To verify the availability of the feature on the Linux operating system, query **/sys/block/$disk/queue/discard_granularity**. If the value is zero, it means the device does not support discard functionality. For example, since device sdx has a non-zero **discard_granularity** value, its free space will be reclaimed with TRIM/UNMAP.:

```
# cat /sys/block/sdx/queue/discard_granularity
8192
```

4.2.5.1 Prepare ASM disk group for space reclamation

1. Ensure LUNs are thin provisioned in the Dell EMC Unity storage system.
2. Create data files with an initial size and enable **autoextend** on the data files.
3. Set the **THIN_PROVISIONED** attribute to **'TRUE'** on the disk group.

4.2.5.2 Reclaim space in ASM disk group

The following outlines the general steps to reclaim storage space in Oracle ASM.

1. Delete rows, tables, objects, or tablespaces.
2. Enable the **ROW MOVEMENT** attribute on the tables.

   ```
   SQL> ALTER TABLE <$TABLE_NAME> ENABLE ROW MOVEMENT;
   ```

3. After deleting objects in a table, run the **ALTER TABLE SHRINK SPACE** statement to repack the rows, move down the high-water mark (HWM), and release unused extents in the data files.

   ```
   SQL> ALTER TABLE <$TABLE_NAME> SHRINK SPACE;
   ```

4. Determine the HWM of each data file and prepare the resize statements using the following script provided by Oracle. The original post can be found in the following Oracle article: https://asktom.oracle.com/pls/apex/f?p=100:11:0::::P11_QUESTION_ID:766625833673

```
# cat find_datafile_hwm.sql
set verify off line 200 pages 100
column file_name format a50 word_wrapped
column smallest format 999,990 heading "Smallest|Size|Poss."
column currsize format 999,990 heading "Current|Size"
column savings format 999,990 heading "Poss.|Savings"
break on report
compute sum of savings on report

column value new_val blksize
select value from v$parameter where name = 'db_block_size'
/

select file_name,
ceil( (nvl(hwm,1)*&&blksize)/1024/1024 ) smallest,
ceil( blocks*&&blksize/1024/1024) currsize,
ceil( blocks*&&blksize/1024/1024) -
ceil( (nvl(hwm,1)*&&blksize)/1024/1024 ) savings
from dba_data_files a,
( select file_id, max(block_id+blocks-1) hwm
from dba_extents
group by file_id ) b
where a.file_id = b.file_id(+)
/

column cmd format a85 word_wrapped

select 'alter database datafile ''' || file_name || ''' resize ' ||
ceil( (nvl(hwm,1)*&&blksize)/1024/1024 ) || 'm;' cmd
from dba_data_files a,
( select file_id, max(block_id+blocks-1) hwm
from dba_extents
group by file_id ) b
where a.file_id = b.file_id(+)
and ceil( blocks*&&blksize/1024/1024) -
ceil( (nvl(hwm,1)*&&blksize)/1024/1024 ) > 0
/
```

The prior script generates the following output.

```
VALUE
--------------------------------------------------------------------------------
8192

                                                 Smallest
                                                     Size    Curr     Poss.
FILE_NAME                                           Poss.    Size  Savings
--------------------------------------------  --------  ------  --------
+DATADG/DEMODB/DATAFILE/system.257.952265543          841     850         9
+DATADG/DEMODB/DATAFILE/demots.289.952606307       16,328  17,014       686
+DATADG/DEMODB/DATAFILE/undotbs1.259.952265593     12,932  24,708    11,776
+DATADG/DEMODB/DATAFILE/undotbs2.265.952265669         37   1,024       987
+DATADG/DEMODB/DATAFILE/demots.337.952606345       17,220  17,652       432
+DATADG/DEMODB/DATAFILE/demots.320.952606331       16,712  17,462       750
+DATADG/DEMODB/DATAFILE/sysaux.258.952265577        2,008   2,030        22
+DATADG/DEMODB/DATAFILE/users.260.952265593             1       5         4
                                                                    --------
sum                                                                   14,666

8 rows selected.

CMD
--------------------------------------------------------------------------------
alter database datafile '+DATADG/DEMODB/DATAFILE/system.257.952265543' resize 841m;
alter database datafile '+DATADG/DEMODB/DATAFILE/demots.289.952606307' resize 16328m;
alter database datafile '+DATADG/DEMODB/DATAFILE/undotbs1.259.952265593' resize
12932m;

alter database datafile '+DATADG/DEMODB/DATAFILE/undotbs2.265.952265669' resize 37m;
alter database datafile '+DATADG/DEMODB/DATAFILE/demots.337.952606345' resize 17220m;
alter database datafile '+DATADG/DEMODB/DATAFILE/demots.320.952606331' resize 16712m;
alter database datafile '+DATADG/DEMODB/DATAFILE/sysaux.258.952265577' resize 2008m;
alter database datafile '+DATADG/DEMODB/DATAFILE/users.260.952265593' resize 1m;

8 rows selected.
```

5. To resize the data files, copy and paste the **ALTER DATABASE RESIZE** statements associated with the data files. For example, the previous statements in bold shrink only the demots tablespace that resides in the DATADG disk group.
6. Manually rebalance the disk group.

   ```
   SQL> ALTER DISKGROUP DATADG REBALANCE POWER 10;
   ```

7. Confirm the release of the space in Unisphere by observing the **Capacity** and **Space Used** information on the LUN properties page. It might take several minutes to see the changes depending on the amount of data and how busy the system is at the time.

Note: Deleted space is not released until either the data files are deleted or shrunk and a rebalance operation is run against the ASM disk groups.

4.3 Linux LVM

Linux Logical Volume Manager (LVM) is a common general-purpose storage manager included in all popular Linux distributions. Since ASM does not support storing Oracle software, the software must be installed on a Linux file system that can be configured on top of LVM. LVM mirroring is not necessary because storage protection is already provided by the Dell EMC Unity system. Multiple LUNs can be grouped into a single LVM volume group. Then logical volumes must be created that span across these LUNs. When taking Dell EMC Unity system snapshots on a multi-LUN volume group, ensure the LUNs are configured in a consistency group.

A file system is created on a logical volume where the Oracle binary is installed. Additional space can be added to the volume groups, logical volumes, and file systems either by adding new LUNs or by expanding existing LUNS in the volume groups. Once volume groups and logical volumes are expanded, the file systems can be resized to the newly added space. LVM and many popular file systems, such as ext4 and xfs, allow on-demand expansion without taking down the applications.

Unlike ASM, the striping needs to be configured by administrators explicitly, and data is not rebalanced when extending the volume group.

4.3.1 LVM guidelines

- Use whole LUNs for volume groups.
- Create a dedicated volume group for storing each copy or version of Oracle software. This simplifies management and allows greater flexibility on array-based snapshots on individual Oracle software copies.
- Use two or more LUNs in a volume group when performance is of concern.
- Configure all LUNs with the same size in the same volume group and group them in the same consistency group.
- In an Oracle RAC configuration, use a dedicated local volume group for each cluster node.

4.3.2 Physical volume data alignment

When initializing LUNs in LVM, use the **--dataalignment** argument to indicate the alignment starts at 1M.

The following example shows the tasks to create an Oracle software file system on LVM:

```
# pvcreate --dataalignment 1m /dev/mapper/orabin-rac
# vgcreate vgoracle /dev/mapper/orabin-rac
# lvcreate -L 50g -n lv-oracle-bin vgoracle
# mkfs.xfs /dev/vgoracle/lv-oracle-bin (for xfs)
# mkfs.ext4 /dev/vgoracle/lv-oracle-bin (for ext4)
```

Note: If **--dataalignment** is not specified, **mkfs** might report a warning message similar to the one shown as follows. Reinitialize the LUN with **--dataalignment** to ensure proper alignment.

Misalignment warning for **mkfs.xfs**:

```
warning: device is not properly aligned /dev/vgoracle/lv-oracle-logs
Use -f to force usage of a misaligned device
```

Misalignment warning for **mkfs.ext4:**

```
mke2fs 1.42.9 (28-Dec-2013)
/dev/vgoracle/lv-oracle-bin alignment is offset by 512 bytes.
This may result in very poor performance, (re)-partitioning suggested.
```

4.4 File systems

Local file system is preferred to store Oracle software and diagnostic logs. It is also possible to store data files in a local file system but it is highly recommended to use Oracle ASM on block devices or Oracle DirectNFS in conjunction with the Dell EMC Unity NFS service instead. Sections 5, 6, and 7 discuss using the Dell EMC Unity NFS service with Oracle DirectNFS.

The Dell EMC Unity system supports a wide range of file systems on Linux. This section focuses on two popular and stable file systems: ext4 and xfs.

For additional information on supported file systems and feature limitations, see the *Dell EMC Host Connectivity Guide for Linux*.

4.4.1 File system layout

The file system can be created on top of a LUN, a LUN partition, or a logical volume in LVM. Dell EMC recommends using the whole LUN without partition or a logical volume in LVM for ease of management.

It can be beneficial to segregate the Oracle software and Oracle diagnostic logs. To do this, create a separate volume group or assign a different LUN to store Oracle diagnostic log files. The diagnostic logs can consume a large amount of space in a very short period of time. By isolating the logs in a different file system, it reduces the risk of filling up the storage space with these logs, and affects the operation of the software. Since the diagnostic logs are not mission critical to the software operation, it is not essential to enable snapshots on the LUNs used by the logs. The diagnostic logs are also good candidates to be compressed to reduce the storage consumption. Table 7 shows an example of using separate file systems for software and diagnostic logs.

Table 7 An example of file system layout for Oracle software and diagnostic logs

Volume group	Logical volume	File system mount point	Dell EMC Unity snapshot	Dell EMC Unity compression
vggrid	lv-grid-bin	/u01	Enable	Disable
vgoracle121	lv-oracle-bin	/u01/app/oracle/product/12.1.0	Enable	Disable
vgoracle122	lv-oracle-bin	/u01/app/oracle/product/12.2.0	Enable	Disable
vgoraclelog	lv-grid-log	/u01/app/grid/diag	Disable	Enable
	lv-oracle-log	/u01/app/oracle/diag	Disable	

4.4.2 File system mount options

When mounting a file system, consider the following options and guidelines.

- Identify the file system by its UUID or LVM LV device in the /etc/fstab file. Avoid using any non-persistent device paths such as /dev/sd*.
- Query UUID with the **blkid** command.

```
# blkid /dev/vgoracle/lv-oracle-rac-home
/dev/vgoracle/lv-oracle-rac-home: UUID="83cf5726-f842-448b-a143-
5f77eb0d9b37" TYPE="xfs"
```

- Include **discard** in the mount option to enable space reclamation support for the file system. More information is provided in section 4.4.4.
- Include **nofail** in the mount option if the Linux operating system experiences mount issue during system boot. This prevents interruption during the boot process which requires manual intervention.
- For the xfs file system, disable the file system check (fsck option) in /etc/fstab because it does not perform any check or repair automatically during boot time. The xfs journaling feature ensures the file system integrity and data is in a consistent state after abrupt shutdown. If a manual repair or check is necessary, use the **xfs_repair** utility to repair damaged file system.
- Set a value of **0** in the sixth field to disable fsck check. Here is an example of an xfs file system entry in /etc/fstab:

```
UUID="83cf5726-f842-448b-a143-5f77eb0d9b37"  /u01 xfs
defaults,discard,nofail 0 0
```

4.4.3 Expand storage for the file system

Certain file system types, such as ext4 and xfs, support the online resize operation. The following outlines the general steps to resize a file system online assuming non-partition LUNs are used.

1. Take manual snapshots of LUNs that are going to be expanded. See section 8.3 for more information on taking snapshots and recovering from snapshots.
2. Expand the size of existing LUNs in Unisphere.
3. Perform a SCSI scan on the host systems, refresh the partition table on each LUN path, and reload multipath devices.

```
# rescan-scsi-bus.sh –resize
```

4. Reload the multipath devices.

   ```
   # multipathd -k"resize map orabin-rac"
   ```

   For PowerPath, the new size is automatically updated.

5. Expand the logical volume if the file system is on top of LVM.

   ```
   # lvresize –L $NEW_SIZE /dev/vgoracle/lv-oracle-rac-home
   ```

6. Extend the file system size to the maximum size, automatically and online.

   ```
   # xfs_growfs –d /u01 (for xfs)
   # resize2fs /dev/mapper/orabin-rac (for ext4)
   ```

4.4.4 Space reclamation

For file system types that support the online SCSI TRIM/UNMAP command, such as ext4 and xfs, enable the **discard mount** option in **/etc/fstab** or include **–o discard** to the manual mount command. This allows space to be released back to the storage pool in the Dell EMC Unity system when deleting files in the file system. Administrators should review the file system documentation to confirm the availability of the features.

The LUNs must be thin provisioned in Dell EMC Unity storage system for space reclamation to work. As new data is written to the file system, actual space is allocated in the Dell EMC Unity system. When files are deleted from the file system, the operating system informs the Dell EMC Unity system which data blocks can be released. The release of storage is automatic and requires no additional steps. To confirm the release of space in the Dell EMC Unity system, monitor the **Total Pool Space Used** on the LUN properties page in Unisphere.

5 Dell EMC Unity file storage

Dell EMC Unity storage can serve file data through virtual file servers (NAS servers) while providing many of the advanced capabilities of Dell EMC Unity systems. Some of these capabilities are shown in the following list, while others are mentioned in the remainder of this section:

- Advanced static routing
- Packet reflect
- IP Multitenancy
- NAS server mobility
- Configurable Dell EMC Unity system parameters

Dell EMC Unity x80F storage systems support NAS connections on multiple 10GbE and 25GbE ports. In an Oracle NFS environment, 25Gb/s is recommended for the best performance. If possible, configure Jumbo frames (MTU 9000) on all ports in the end-to-end network path (NFS client interfaces, Ethernet switch interfaces, and Dell EMC Unity interfaces) to provide the best performance.

When using Oracle Direct NFS (dNFS) where high availability is needed, it is recommended to configure the Link Aggregation Control Protocol (LACP) across the same multiple Ethernet ports on each SP to provide path redundancy between clients and NAS servers. Combine LACP with redundant switches to provide the highest network availability. LACP can be configured across all available Ethernet interfaces and between the I/O modules. See Figure 34, Figure 35, and Figure 36 for examples.

For additional information pertaining to this section, see the *Dell EMC Unity: NAS Capabilities, Dell EMC Unity: Best Practices Guide,* and *Dell EMC Unity: Service Commands* documents.

5.1 Dell EMC Unity front-end Ethernet connectivity for file storage

Dell EMC Unity storage provides multiple options for 10Gb/s Ethernet front-end connectivity, through onboard ports directly on the DPE and through optional I/O modules. In general, front-end ports need to be connected and configured symmetrically across the 2 SPs to facilitate high availability and continued connectivity in case of SP failure. For best performance, it is recommended to use all front-end ports that are installed in the system so that workload is spread across as many resources as possible and use the Dell EMC Unity 10/25GbE ports for dNFS data traffic.

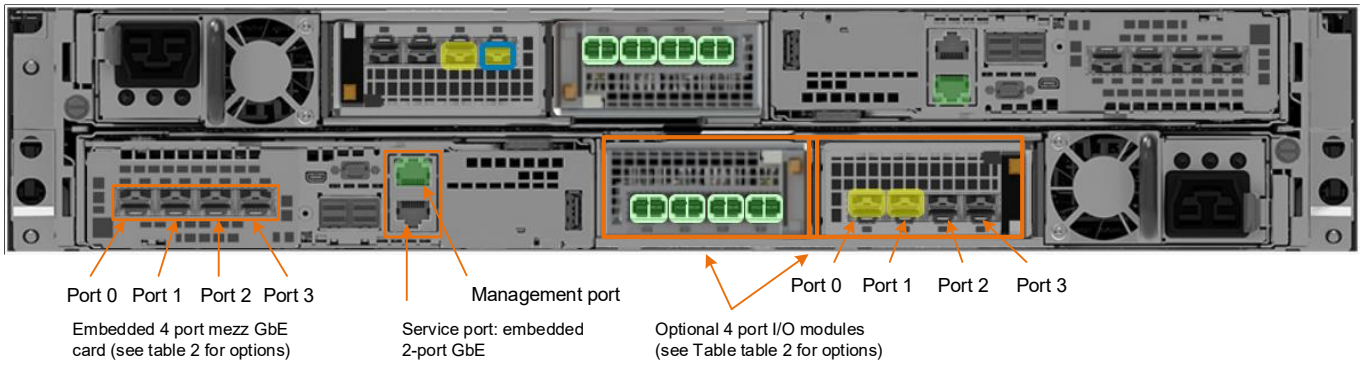

Figure 5 Dell EMC Unity 480F, 680F, and 880F front-end Ethernet ports

5.2 Dell EMC Unity NAS servers

The Dell EMC Unity virtual NAS servers are assigned to a single SP. All file systems serviced by a NAS server will have their I/O processed by the SP on which the NAS server is resident or current. If multiple NAS servers are required for multiple Oracle environments, it is recommended that NAS servers are load-balanced in a way that the front-end NFS I/O is roughly distributed evenly between the SPs. Keep in mind not to over provision either of the SPs such that in the event of failover, the peer SP does not become overloaded.

Because each NAS server is logically separate, NFS clients of one NAS server cannot access data on another NAS server. This can provide database isolation and protection across multiple NFS clients (database servers). To create a NAS server, in Dell EMC Unisphere select **File** > **NAS Servers** > **+** and supply the necessary information as shown in the following screens.

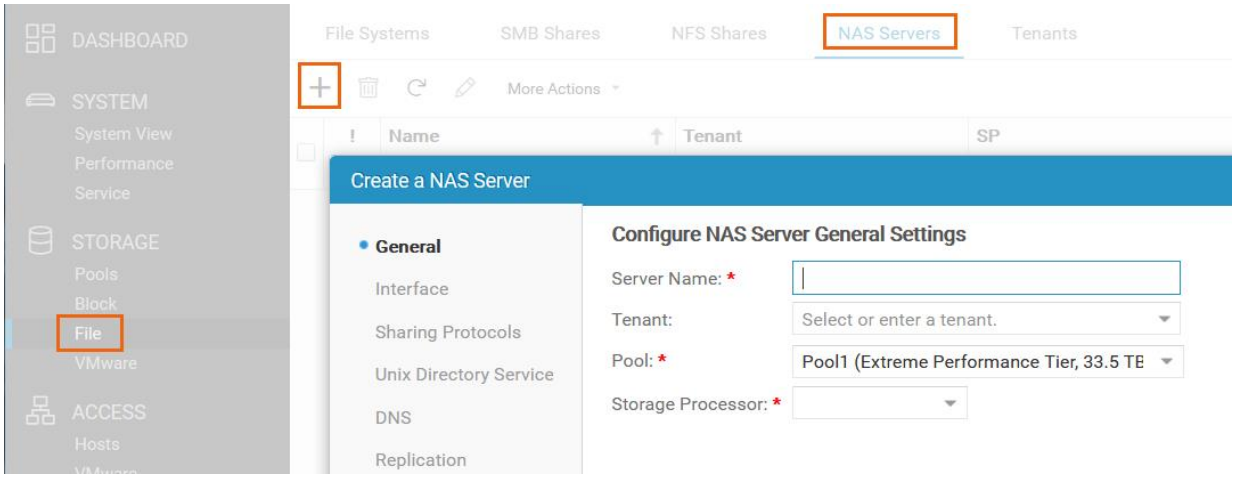

Figure 6 Starting the **Create a NAS Server** wizard

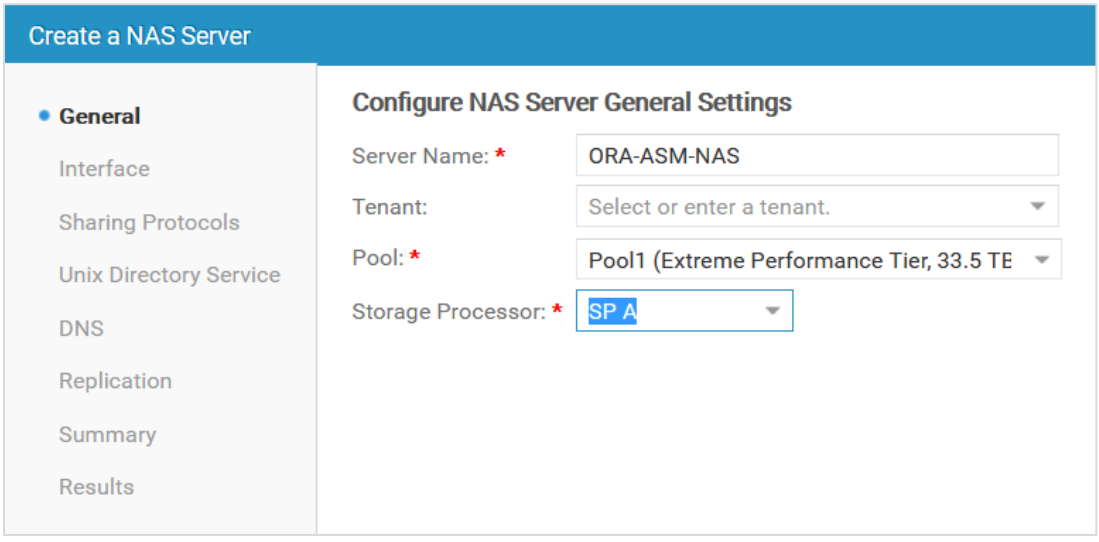

Figure 7 Creating a NAS server

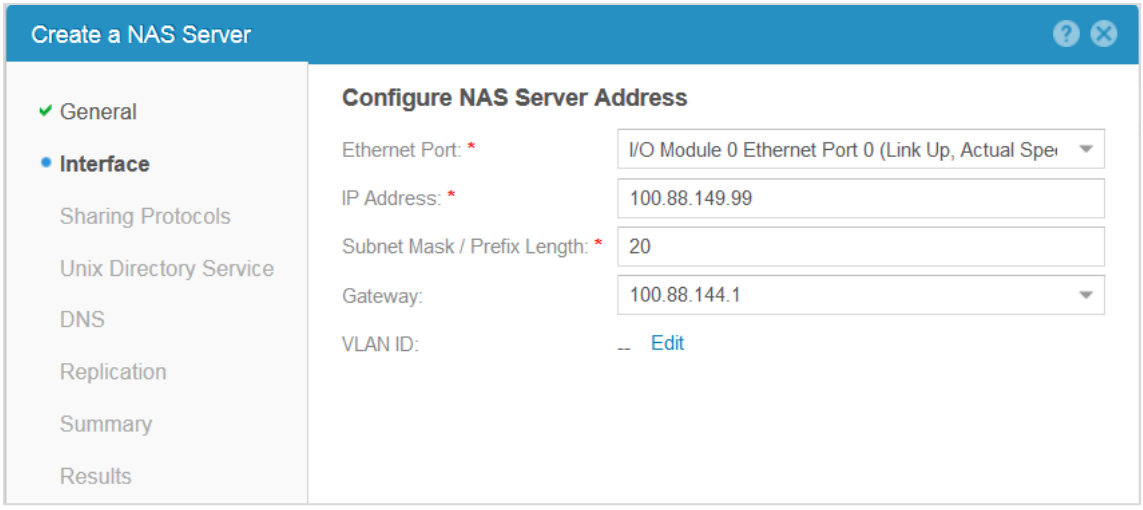

Figure 8 Specifing network information for the NAS Server interface

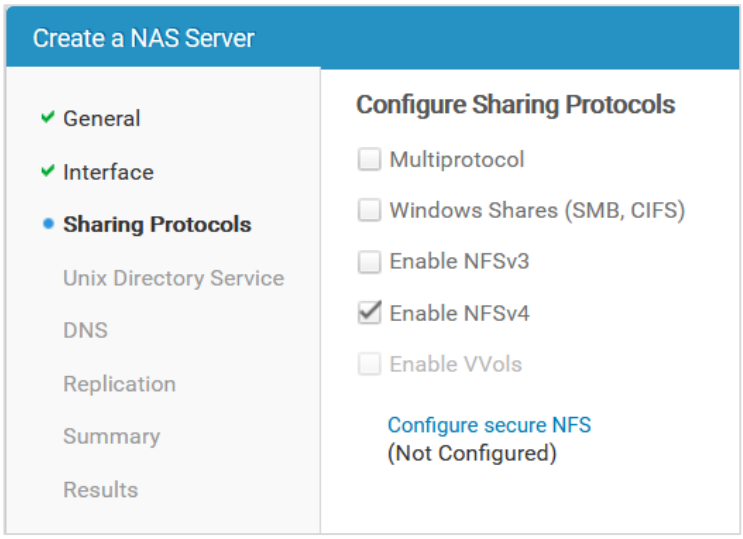

Figure 9 Defining sharing protocols for the NAS Server

When creating a NAS server for an Oracle database, enable NFSv4 if possible and then skip the steps for setting the **Unix Directory Service** and NAS server **DNS** if they are not needed. After a NAS server is created, the Dell EMC Unity NFS file systems can be created, and then Dell EMC Unity NFS shares can be created.

NAS server interfaces can either be configured as production, or backup and DR testing interfaces. The type of interface dictates the type of activity that can be performed. Table 8 displays the characteristics of the interface types.

Table 8 NAS server interface types

Interface type	Characteristics
Production	• Allows CIFS, NFS, and FTP access • Replicated during replication sessions • During replication, is active on in the source mode
Backup and DR test	• Could be used for backup and DR testing • Allows NFS access only • Not replicated during a replication session • Is active in both source and destination replication modes

If throughput will be restricted by only using one Ethernet interface, consider configuring multiple Ethernet ports for the NAS server by selecting: **File** -> **NAS Servers** -> select **checkbox for NAS server** -> **Network-> +** and adding additional Ethernet interfaces.

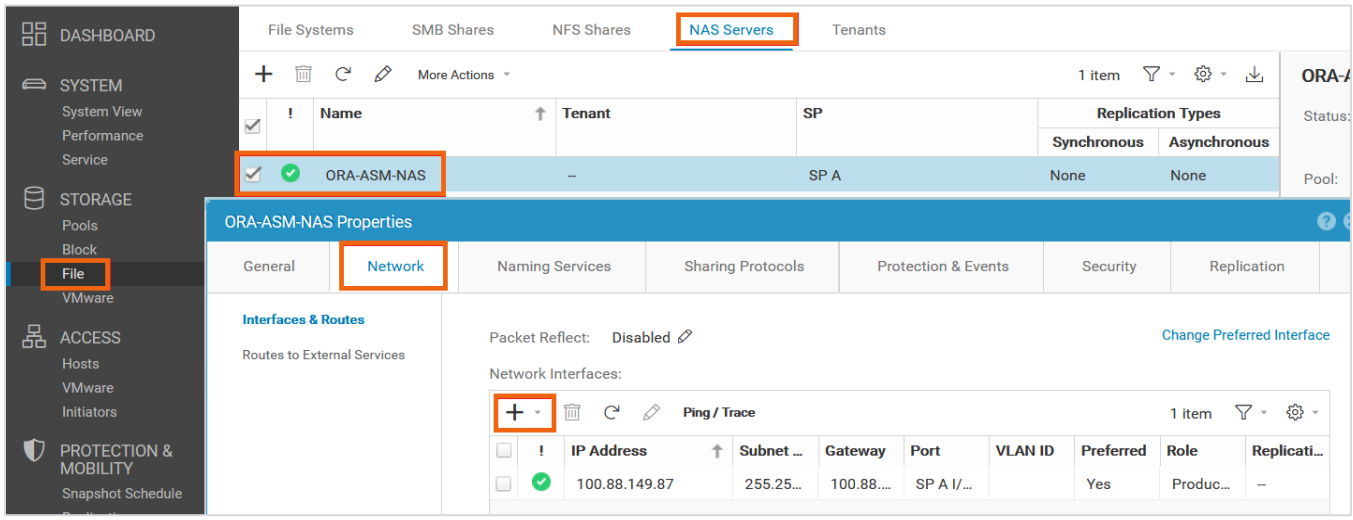

Figure 10 Defining multiple Ethernet interfaces for the NAS server

5.3 Dell EMC Unity NFS file system

The Dell EMC Unity file system contains several improvements over existing NAS file system technologies and is well suited for Oracle. The improved areas include scalability and maximum system size, flexible file system, storage efficiency, security, isolation, availability, recoverability, virtualization, and performance.

To create a file system in Unisphere, select **File** > **File Systems** > **+** and supply the desired configuration.

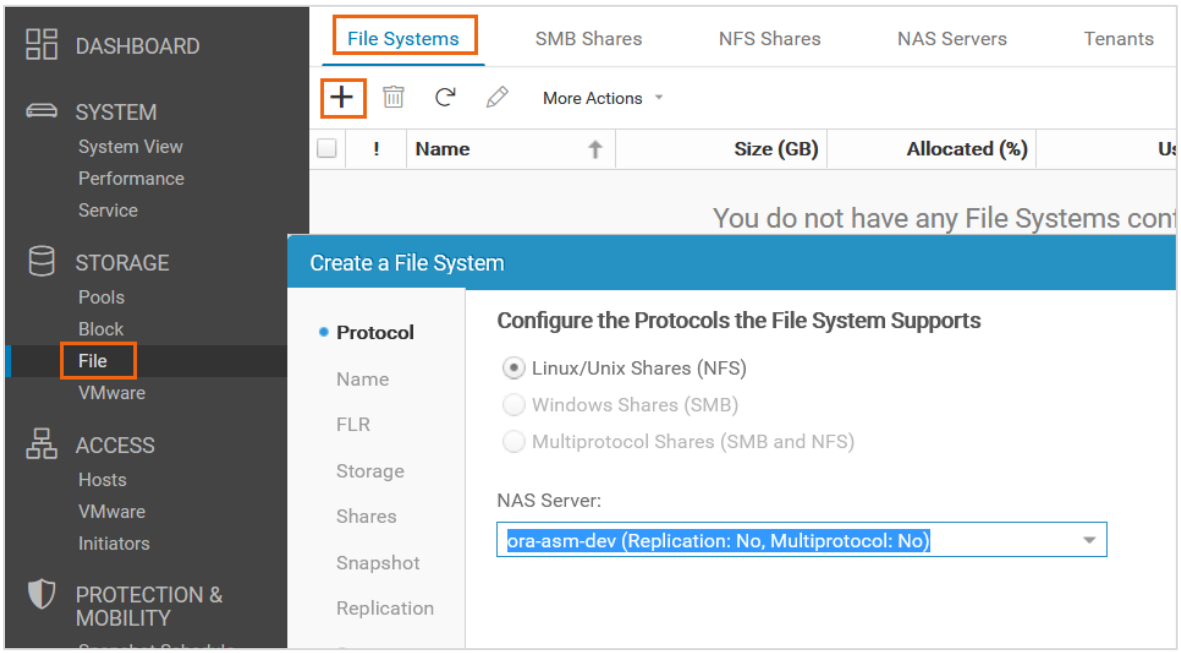

Figure 11 Creating a file system on the NAS server

With respect to the Oracle database files, the NFS file system can host Oracle datafiles that exist on ASM, file system, or both. See Figure 12 and Figure 13.

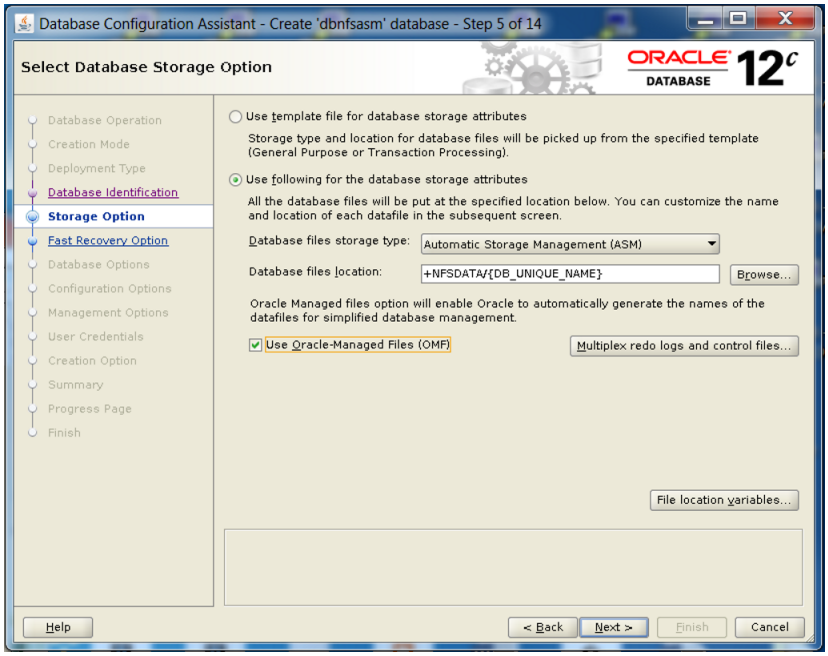

Figure 12 NFS file system hosting raw files for ASM

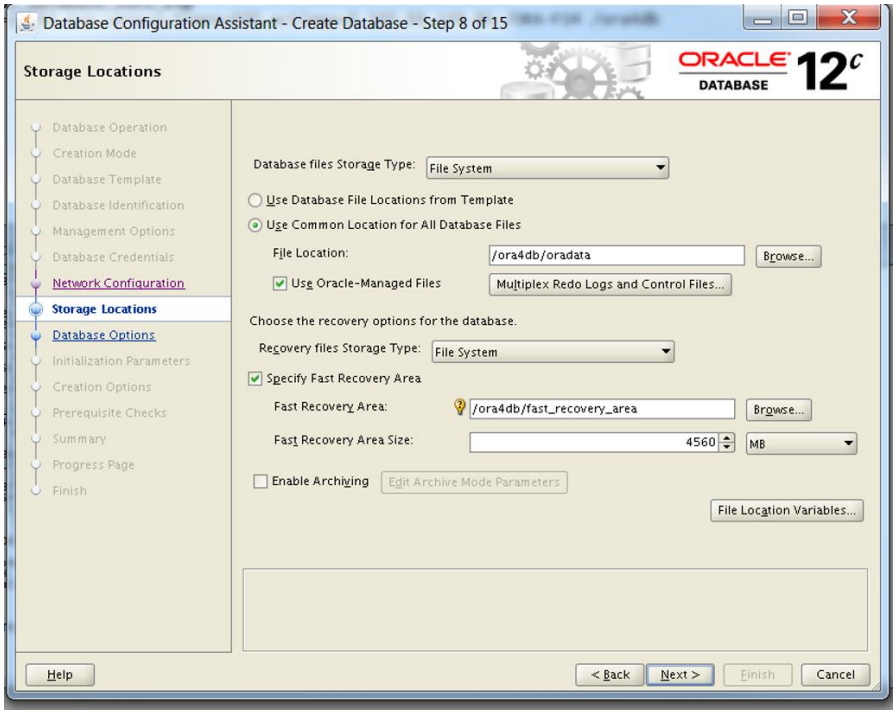

Figure 13 NFS file system hosting Oracle datafiles in a file system.

5.4 Scalability

Dell EMC Unity file systems provide scalability in a number of areas, including maximum file system size, which makes Dell EMC Unity storage ideal for Oracle environments. Dell EMC Unity OE version 4.2 increases the maximum file system size from 64 TB to 256 TB for all file systems. File systems can also be shrunk or extended to any size within the supported limits. Dell EMC recommends configuring storage objects that are 100 GB at a minimum and preferably 1 TB in size or greater.

5.5 Storage efficiency

Dell EMC Unity storage supports thin-provisioned file systems. Starting with Dell EMC Unity OE version 4.2, Unisphere also provides the ability to create thick file systems. When using Dell EMC Unity file storage with Oracle, consider using thin-provisioned file systems. Dell EMC Unity also provides increased storage flexibility by providing the ability to manually or automatically perform file system extension and shrink with reclaim.

5.6 Quotas

Dell EMC Unity storage includes full-quota support to allow administrators to place limits on the amount of space that can be consumed from a user of an NFS file system or directory, or a directory itself, in order to regulate storage consumption. When working with Oracle, quotas are not necessary in most cases. If deciding to use quotas, carefully consider their impact on managing the Oracle environment.

5.7 NFS protocol

Dell EMC Unity storage supports NFSv3 through NFSv4.1, including secure NFS.

All Dell EMC Unity OE versions support Oracle dNFS in single-node configurations. Starting with OE version 4.2, Oracle Real Application Clusters (RAC) are also supported. In order to use Oracle RAC, the nfs.transChecksum parameter must be enabled. This parameter ensures that each transaction carries a unique ID and avoids the possibility of conflicting IDs that result from the reuse of relinquished ports.

For more information about NAS server parameters and how to configure them, see the Dell EMC Unity *Service Commands* document.

NFSv4 is a version of the NFS protocol that differs considerably from previous implementations. Unlike NFSv3, this version is a stateful protocol, meaning that it maintains a session state and does not treat each request as an independent transaction without the need for additional preexisting information. With NFSv4, all network traffic is handled by underlying transport protocol as opposed to the application layer in NFSv3. This can provide savings in the overall load on the Oracle database server (NFS client). NFSv4 is preferred due to improvements over NFSv3. Some advantages of NFSv4 are:

- Ability to use TCP more thoroughly
- Ability to bundle metadata operations
- An integrated, more functional lock manager
- Conditional file delegation

While Dell EMC Unity storage fully supports the majority of the NFSv4 and v4.1 functionality described in the relevant RFCs, directory delegation and pNFS are not supported. Therefore, do not configure Oracle to use parallel dNFS (known as pNFS). For increased performance, consider using NFSv4 and Oracle Direct NFS (dNFS) with multiple network interfaces for load-balancing purposes.

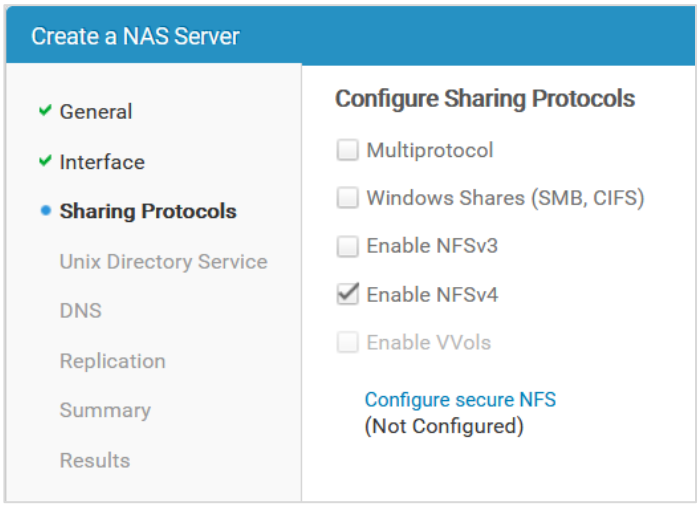

Figure 14 Sharing protocols

5.8 Dell EMC Unity NFS share

After creating the NAS server and file system, the NFS share can be created. To create the NFS share, select **File** > **NFS Shares** > **+** and supply the necessary information.

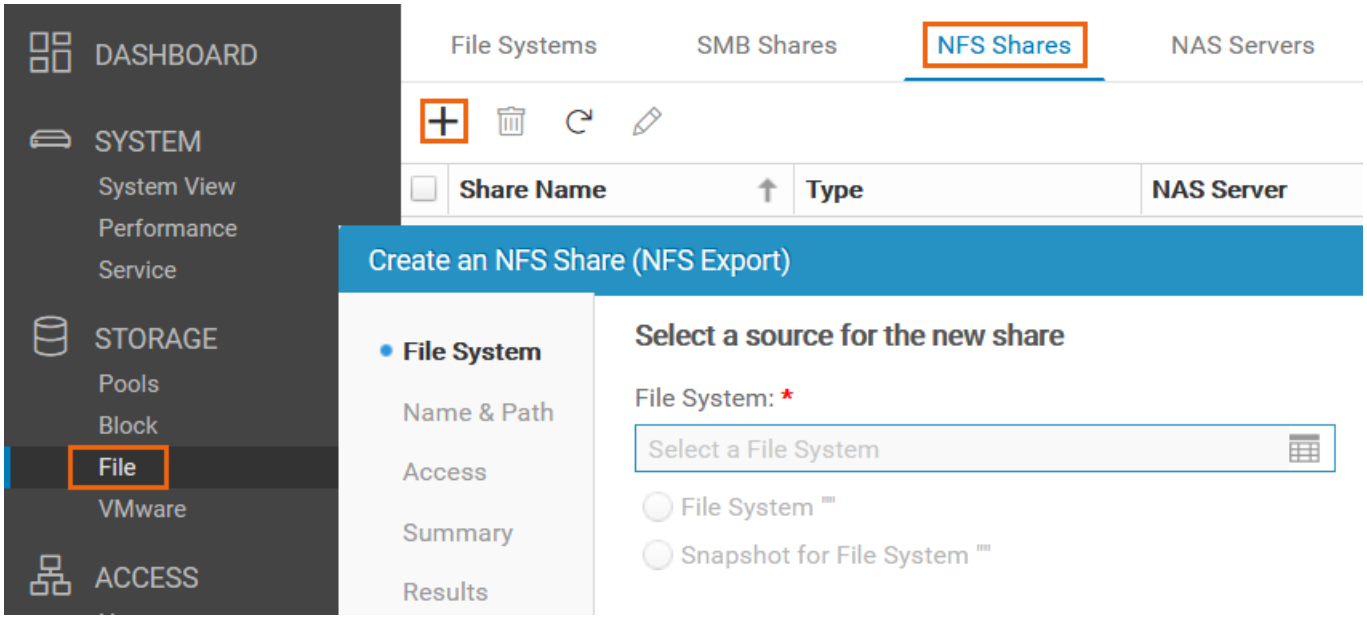

Figure 15 Creating a NFS share

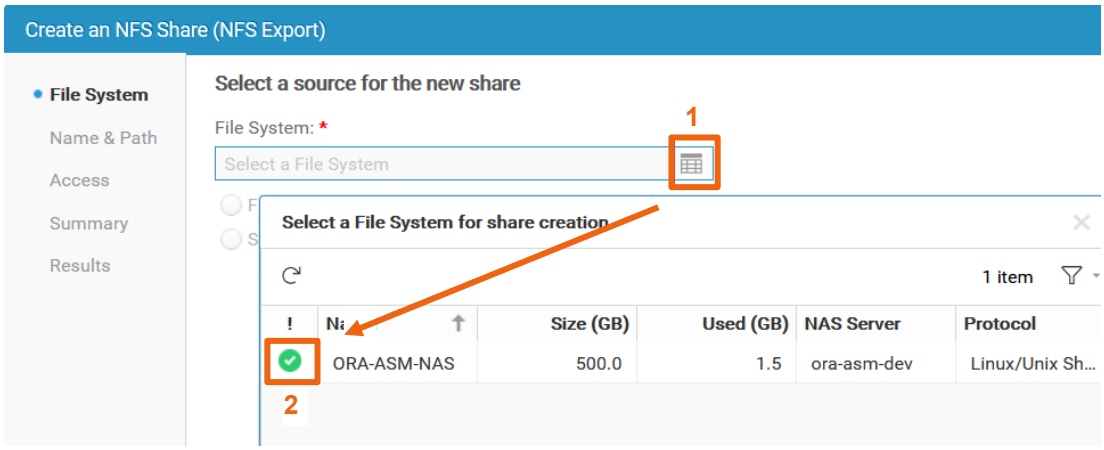

Figure 16 Assigning a file system to a NFS share

When defining the NFS share name, make sure Allow SUID is selected as this is required for Oracle software mount points.

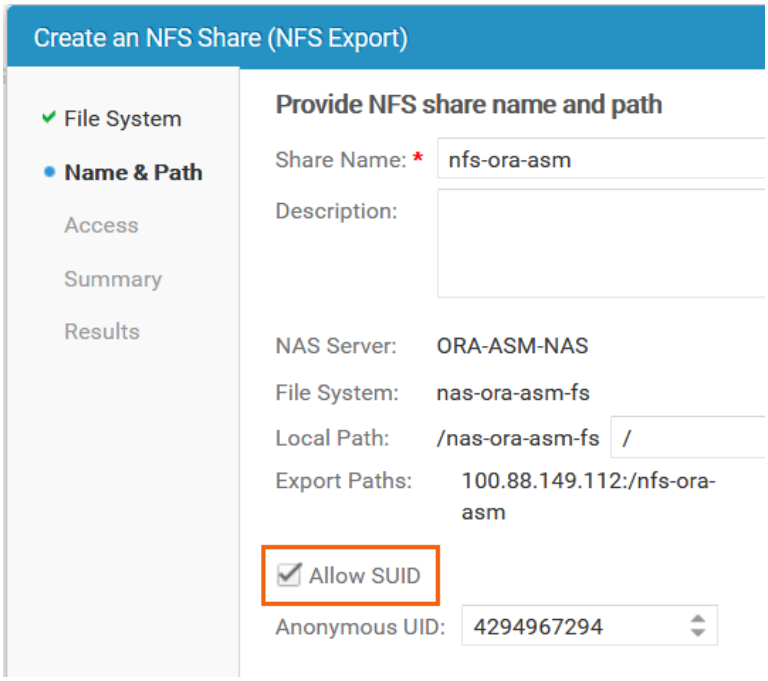

Figure 17 Allow SUID for Oracle

For NFS shares intended for Oracle, set the NFS export options for the NFS share by setting **Default Access** to **Read/Write, allow Root**.

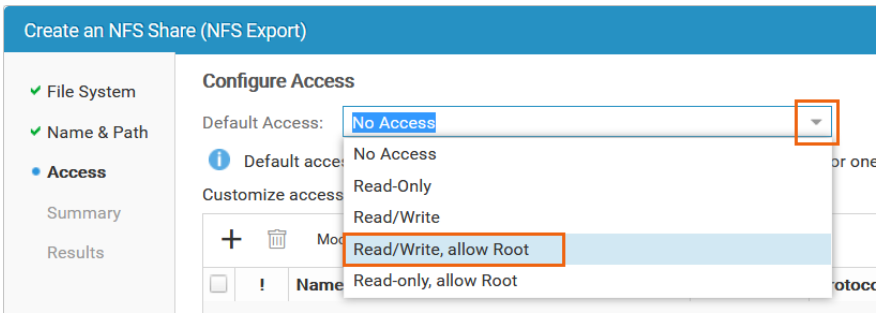

Figure 18 Speify R/W and allow root access on the NFS share

5.9 Verify access to the Dell EMC Unity NFS share

After Dell EMC Unity file storage (NAS server, NFS file system, and NFS share) has been configured for the NFS client (database server), log in to the database server and verify it has access to the NFS share through all the IPs defined for the NFS share. To verify access, use the **showmount** command in Linux on all the IPs shown in the list of **Exported Paths**. If any of the IPs do not have access to the NFS share, resolve the issue before configuring the NFS client including configuring Oracle dNFS.

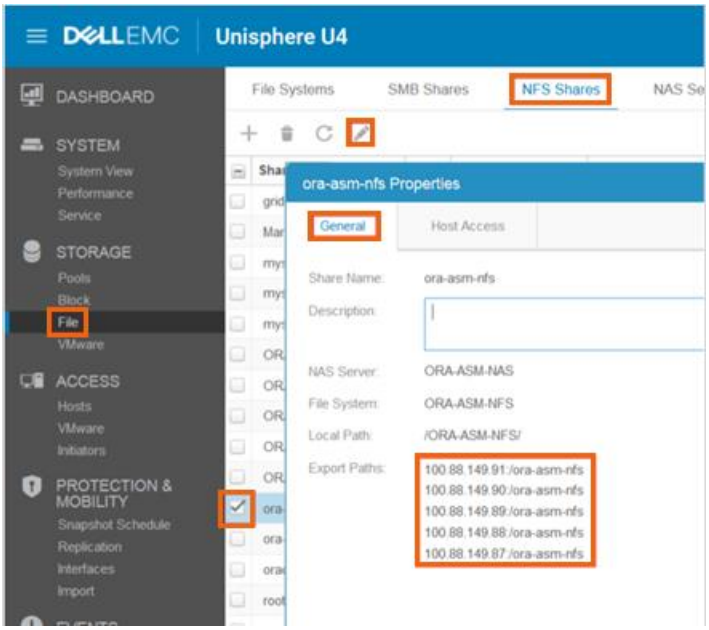

Figure 19 Configure IPs and mount names for the NFS share

The following **showmount** command only illustrates its usage on the first IP in the list of **Exported Paths**.

```
[root ~]# showmount -e 100.88.149.91
Export list for 100.88.149.91:
/ORA-ASM-NFS       (everyone)
ora-asm-nfs        (everyone)
```

5.10 Dell EMC Unity file system and Oracle ASM

To use ASM on top of the Dell EMC Unity file system, use the following process (change values where necessary):

1. Create the Dell EMC Unity NAS share.
2. Create the mount point in Linux and set the permissions and ownership on the mount point:

   ```
   mkdir /oraasmnas
   chmod 770 /oraasmnas
   chown grid:oinstall /oraasmnas
   ```

3. Mount the Dell EMC Unity NAS share.

   ```
   mount -o
   rw,bg,hard,nointr,rsize=32768,wsize=32768,tcp,vers=3,timeo=600,actimeo=0
   100.88.149.91:/ORA-ASM-NFS /oraasmnas
   ```

4. Change the permissions and ownership of the root directory on the NFS share:

```
chmod 770 /oraasmnas
chown grid:oinstall /oraasmnas
```

5. Create the raw files for the ASM disk groups and set their permissions and ownership.

```
dd if=/dev/zero of=/oraasmnas/nfsasm-ocrvote-disk01 bs=4096 count=2621440
dd if=/dev/zero of=/oraasmnas/nfsasm-data-disk01 bs=8192 count=524288
dd if=/dev/zero of=/oraasmnas/nfsasm-data-disk02 bs=8192 count=524288
dd if=/dev/zero of=/oraasmnas/nfsasm-data-disk03 bs=8192 count=524288
dd if=/dev/zero of=/oraasmnas/nfsasm-data-disk04 bs=8192 count=524288
dd if=/dev/zero of=/oraasmnas/nfsasm-data-disk05 bs=8192 count=524288
dd if=/dev/zero of=/oraasmnas/nfsasm-fra-disk01 bs=8192 count=524288
dd if=/dev/zero of=/oraasmnas/nfsasm-fra-disk02 bs=8192 count=524288
dd if=/dev/zero of=/oraasmnas/nfsasm-fra-disk03 bs=8192 count=524288
dd if=/dev/zero of=/oraasmnas/nfsasm-fra-disk04 bs=8192 count=524288
dd if=/dev/zero of=/oraasmnas/nfsasm-fra-disk05 bs=8192 count=524288

chown grid:oinstall /oraasmnas/nfsasm*
chmod 660 /oraasmnas/nfsasm*
```

6. Create the ASM disk groups.

```
SQL> create diskgroup nfsdata external redundancy disk
  2    '/oraasmnas/nfsasm-data-disk01',
  3    '/oraasmnas/nfsasm-data-disk02',
  4    '/oraasmnas/nfsasm-data-disk03',
  5    '/oraasmnas/nfsasm-data-disk04',
  6    '/oraasmnas/nfsasm-data-disk05';
```

6 Oracle Disk Manager

Oracle I/O activity and its file management infrastructure are managed by the Oracle Disk Manager (ODM) library ($ORACLE_HOME/lib/libodm12.so). ODM can also provide the ability to use NFS devices for database I/O without using the native Linux NFS kernel (kNFS), providing the ODM library containing the embedded Oracle NFS client ($ORACLE_HOME/rdbms/lib/odm/libnfsodm12.so) is enabled.

6.1 NFS traffic

Generally, NFS traffic can be either classified as control/management traffic and actual I/O traffic on application data. With respect to the OS, whether or not the Oracle ODM NFS client library is enabled, control/management of NFS devices is always managed by the native Linux NFS kernel client (kNFS) driver. When the ODM library containing the embedded Oracle NFS client is enabled, the Oracle environment is said to be using Oracle Direct NFS (dNFS) and all database I/O, NFS data traffic, flows through the dNFS driver. When the ODM library containing the embedded Oracle NFS client is disabled, all database I/O flows through the kNFS client driver.

Some examples of NFS control and management activity involve the following operations on the NFS share:

- get attribute
- set attribute
- access
- create
- mkdir
- rmdir
- mount
- umount

7 Oracle Direct NFS

Oracle Direct NFS (dNFS) is an optimized NFS client from Oracle for database I/O and resides in the ODM library as a part of the Oracle database kernel. dNFS improves the stability and reliability of NFS storage devices over TCP/IP, more so than the native Linux NFS driver (kNFS). dNFS also improves performance to NFS storage devices by bypassing the kNFS I/O stack. When mounting the database data files, Oracle will first load dNFS functionality if the Direct NFS client ODM library is enabled. If dNFS cannot access a NFS storage device, dNFS silently reverts to using the kNFS client. However, to ensure this reversion occurs, the kNFS client mount options **rsize** and **wsize** must be used.

While Dell EMC Unity 4.2, Oracle 12cR1, and 12cR2 dNFS all support NFSv3 and the stateful NFSv4 and NFSv4.1 protocols, Dell EMC Unity does not provide functionality for pNFS. Therefore, do not configure pNFS in Oracle 12cR.

It is recommended to use dNFS if NFS storage devices are used so that the performance optimizations built into Oracle can be exploited.

7.1 Benefits of dNFS

The advantage of using Oracle dNFS lies within the fact that it is part of the Oracle database kernel and all I/O to NFS storage devices are serviced by the Oracle dNFS client rather than by the kNFS client. This gives Oracle the ability to manage the best possible configuration, automatically tune itself, take advantage of the Oracle buffer cache, and appropriately use available resources for optimal multipath NFS data traffic I/O, without the overhead of the client OS kernel software.

7.2 Creating NFS client mount points

An Oracle installation requests the intended locations for storing the software and components, and is dependent on the infrastructure and application requirements. In most cases, these locations can reside on NFS shares. Some exceptions are discussed in section 7.3.

Table 9 provides examples of different Oracle directories that could reside on a NFS share. Once it is determined which NFS shares will be used by Oracle, create the necessary mount points for the NFS shares and create the NFS shares in Dell EMC Unity storage. Also, set the privileges, owner, and group of the Linux mount points and root directory on the NFS share per Oracle requirements.

Table 9 Example directories that could be serviced by NFS

Oracle directory	Environment variables and typical values	Description
Oracle base	$ORACLE_BASE=/u01/app/oracle/	The top-level directory for installations. Subsequent installations can either use the same Oracle base or a different one.
Oracle inventory	/u01/app/oraInventory/ Or $ORACLE_BASE/<srv>/oraInventory/	All Oracle installations use the same Oracle inventory directory for the installation repository metadata. If possible, Oracle recommends the inventory directory reside on a local file system: /u01/app/oraInventory If a NAS device must be used for the inventory, to prevent multiple systems from writing to the same inventory, create a unique directory for each database server: $ORACLE_BASE/<srv>/oraInventory
Oracle home	$ORACLE_HOME=$ORACLE_BASE/product/12.2.0/dbhome_1/	This directory contains the binaries, library, configuration files, and other files from a single release of one product, and cannot be shared with other releases or other Oracle products.
Database file directory	$ORACLE_BASE/oradata/	This is the location to hold the database. It is recommended to use a different NFS mount point for database files to provide the ability to mount the NFS file system with different mount options, and to distribute database I/O.
Oracle recovery directory	$ORACLE_BASE/fast_recovery_area/	Oracle recommends that recovery files and database files do not exist on the same file system.
Oracle product directory	$ORACLE_BASE/product	This mount point can be used to install software from different releases, for example: /u01/app/oracle/product/12.1.0/dbhome_1/ /u01/app/oracle/product/12.2.0/dbhome_1/
Oracle release directory	$ORACLE_BASE/product/<version>/	This mount point can be used to install different Oracle products from the same version, for example: $ORACLE_BASE/product/11gR2/dbhome_1 $ORACLE_BASE/product/11gR2/client_1 Even though this is an option, it is not recommended to install both the rdbms and client on the database server. If the client is required, it is recommended that a separate NFS be defined and a non-database server be used to host the client install.

7.3 Mount options for NFS share

Before configuring or using the dNFS driver on a NAS share, the NFS share must first be mounted using the kNFS driver. Specific mounting options are required when mounting an NFS share for dNFS usage. If the NFS volume will be used for Oracle services that need to be automatically restarted when the server restarts, the NFS volume and mount options must be specified in **/etc/fstab**; otherwise Oracle will experience issues. In an Oracle RAC cluster, ensure that all nodes in the cluster use the same mount options for each identical NFS mount point.

After the share is mounted using kNFS, dNFS mounts and unmounts the volume logically as needed. Since dNFS uses a logical mount, after it unmounts the share, the volume can still be accessed through kNFS. This guarantees that files from the share can be shared by other Oracle databases or users as necessary.

If NFS is used for database files, the NFS buffer size for reads (rsize) and writes (wsize) must be set to at least 16,384. Oracle recommends a value of 32,768. These values are set in **/etc/fstab,** or when explicitly mounting an NFS volume. Since a dNFS write size (v$dnfs_servers.wtmax) of 32,768 or larger is supported in Dell EMC Unity storage, dNFS does not fall back to the traditional kNFS kernel path. dNFS clients issue writes with v$dnfs_servers.wtmax granularity to the NFS server.

The following lists the required mount options for NFS mount points used by Oracle standalone, Oracle RAC, RMAN, and Oracle binaries running on Linux x86-64 version 2.6 and above. For additional mount options for NFS shares intended for Oracle, see the Oracle MOS note, Mount Options for Oracle files for RAC databases and Clusterware when used with NFS on NAS devices, (Doc ID 359515.1) at Oracle Support.

Linux kernel 2.6 x86-64 NFS mount options for Oracle 12c RAC and standalone:

- Mount options for binaries (ORACLE_HOME, CRS_HOME) and database files[1,2]:

  rw,bg,hard,nointr,rsize=32768,wsize=32768,tcp,vers={3|4},timeo=600,actimeo=0

- Mount options for CRS voting disk and OCR[2]:

  rw,bg,hard,nointr,rsize=32768,wsize=32768,tcp,vers={3|4},timeo=600,actimeo=0,noac

[1] The mount options are applicable only if ORACLE_HOME is shared. Oracle also recommends that the Oracle inventory directory be kept on a local file system. If it must be placed on a NAS device, create a specific directory for each system to prevent multiple systems from writing to the same inventory directory. Oracle clusterware is not certified on dNFS.
[2] Do not replace tcp with udp. Udp should never be used. dNFS cannot serve an NFS server with write size less than 32768. As desired, set option vers to either 3 or 4, and ensure the NFS sharing protocol on the Dell EMC Unity NAS server is set accordingly. In 12cR2, both OCR and voting disks must reside in ASM. See Oracle MOS note 2201844.1 for additional information. dNFS is RAC aware. Therefore, even though NFS is a shared file system, and NFS devices for Oracle have to be mounted with the noac option, dNFS automatically recognizes RAC instances and takes appropriate action for datafiles without additional user configuration. This eliminates the need to specify noac when mounting NFS file systems for Oracle datafiles or binaries. This exception does not pertain to CRS voting disks or OCR files on NFS. NFS file systems hosting CRS voting disks and OCR files, must be mounted with noac. Option noac should not be used for RMAN backup set, image copies, and data pump dump files because RMAN and data pump do not check this option and specifying it can adversely affect performance.

When configuring an Oracle RAC environment that uses NFS, ensure the entry in **/etc/fstab** is the same on each node. The following snippet from **/etc/fstab** mounts an NFS mount point for ORACLE_HOME binaries (**/u01**), and a database that will use ASM.

```
100.88.149.63:/ora-bin        /u01           nfs
rw,bg,hard,nointr,rsize=32768,wsize=32768,tcp,vers=3,timeo=600,actimeo=0,defaults 0 0
100.88.149.91:/ORA-ASM-NFS    /oraasmnas     nfs
rw,bg,hard,nointr,rsize=32768,wsize=32768,tcp,vers=3,timeo=600,actimeo=0,defaults 0 0
```

When adding multiple mount options for a specific mount point in **/etc/fstab,** do not insert spaces after options because the OS may not properly parse the options.

Mount options timeo, hard, soft, and intr control the NFS client behavior if the NFS server should become temporarily unreachable. Specifically, whenever the NFS client sends a request to the NFS server, it expects the operation to have finished after a given interval (specified in the timeout option). If no confirmation is received within this time, a minor timeout occurs and the operation is retried with the timeout interval doubled. After reaching a maximum timeout of 60 seconds, a major timeout occurs. By default, a major timeout causes the NFS client to print a message to the console and start over with an initial timeout interval twice that of the previous cascade. There is the potential for this to repeat indefinitely. Volumes that retry an operation until the server becomes available again are called hard-mounted.

See appendix A for a description of the mount options used in this paper.

7.4 Ethernet networks and dNFS

Because the performance of an Oracle database and software stored on NAS devices depend in part on the performance of the network connection between the NFS client (Oracle database server) and NFS server (Dell EMC Unity), it may be prudent to use a dedicated 1Gb/s interface between the NFS client and NAS server for NFS control traffic and a dedicated 10Gb/s interface for NFS data traffic. With Dell EMC Unity x80F systems, only 10GbE and 25GbE interfaces are available. Additional 10Gb/s Ethernet interfaces may be required between the NFS client and NFS server for increased load balancing, availability, and performance in environments with high expected NFS I/O database activity. Using 10Gb/s is the best way to utilize the full capability of Dell EMC Unity file storage. Other considerations when setting up the network are NIC speed, full duplex settings, end-to-end MTU setting, NFS data transfer buffer sizes, and using bonded NIC interface for NFS control traffic.

If NFS and network redundancy is a concern, all interfaces (database server, Ethernet switch, and Dell EMC Unity storage) used for NFS control traffic should be bonded. This bonded interface could be the bonded public network, if it exists, or even the bonded interface for the RAC interconnect in a RAC environment.

Directing NFS control and data traffic to different NICs may not always be possible because of a limited number of NICs, or infrastructure limitations. In such cases, it is possible to share an unbonded interface for both NFS control and data traffic. However, that configuration may cause network performance issues under heavy loads as the server will not perform network load balancing, and it does not provide as much database availability as multiple NICs.

When using dNFS, Oracle supports one to five network paths for NFS traffic between a NAS server and NFS client: one path for NFS control traffic and up to four paths for NFS data traffic. When using dNFS, it is always recommended to use multiple network paths and that each NFS network path belong to subnet that is not currently being used for any other NIC interface on the NFS client (database server). This includes not using

the subnet of the public network for NFS. Using unique subnets simplifies configuration of dNFS and ensures that dNFS benefits are fully exploited.

In some cases, there may be fewer available subnets than intended dNFS paths. If so, dNFS paths on the NFS client can be set up to use existing subnets already in use on the NFS client. However, this requires additional configuration in the OS network layer (relaxing ingress filtering for multihomed networks and static routing) and in Oracle (file oranfstab) if dNFS data traffic needs to be on one or more dedicated subnets. This extra configuration will disable the OS from determining the default dynamic route of the network path and will allow multiple NIC interfaces in the same server to use the same subnet. See section 7.8.1 for information on using multiple NIC interfaces in the same subnet.

If the OS chooses the dynamic route, it will invariably use the first best-matched route possible from the routing table for all paths defined. In most cases, that route will be incorrect. This results in dNFS load balancing, scalability, and failover not working as expected. Therefore, to ensure load balancing and scalability of NFS data traffic are working as expected when using multiple paths in the same subnet, configure static OS routing for each dNFS network path. See section 7.8.2 for more information.

If different subnets are used for NFS traffic, routing will be taken care of automatically by the native network driver and the default route entries in the routing table. Creating static routes are not necessary when using different subnets for dNFS traffic.

Whether or not the same or different subnets are used to define the paths between the end-point IP addresses between the Dell EMC Unity NAS server and the interfaces on the NFS client (database server), all end points must be defined in the Oracle file oranfstab when dNFS data traffic is being isolated to one or more dedicated IP addresses. For additional information, see section 7.10.

7.5 Jumbo frames

Jumbo frames, which refers to raising the maximum transfer unit (MTU) from the default of 1,500 bytes to 9,000 bytes, is advised for the entire network path: database servers (NFS client), Ethernet switch, and Dell EMC Unity storage. Utilizing Jumbo frames allows the network stack to bundle transfers into larger frames and reduce the TCP protocol overhead. The actual value used for any frame can vary depending on the immediate needs of the network session established between the NFS client and server, but raising the limit to 9,000 from end point to end point in the network path will allow the session to take advantage of a wider range of frame sizes.

7.6 Single network path for dNFS

For the simplest dNFS configuration, it is recommended to set up dNFS as a single network path between the NAS server and NFS client.

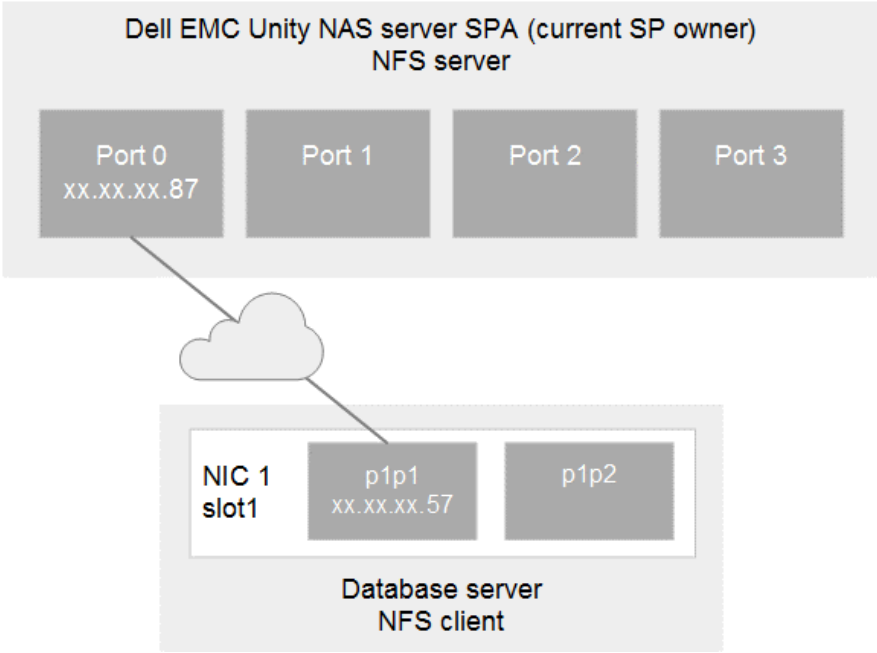

Figure 20 Single Ethernet path between the NFS client and Dell EMC Unity NAS server for NFS traffic

This path would be used by NFS control/management and data traffic and involves setting up the Dell EMC Unity NAS server and NFS share, kNFS for the NFS share, and NIC interface, and enabling dNFS.

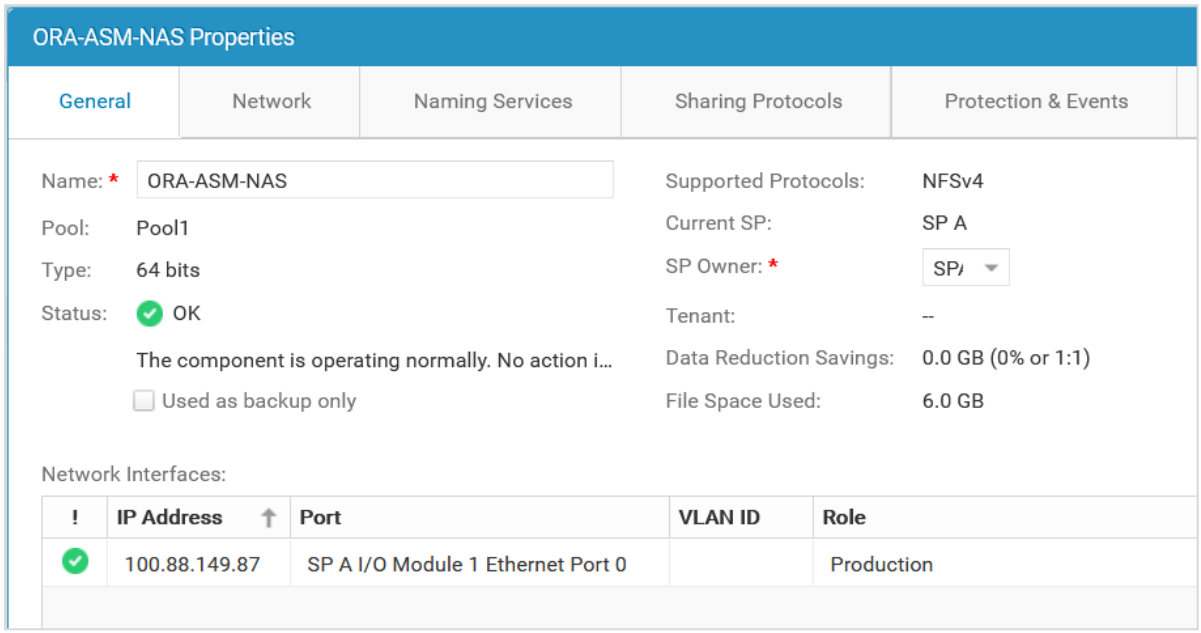

Figure 21 Dell EMC Unity NAS server with a single Ethernet path for NFS traffic

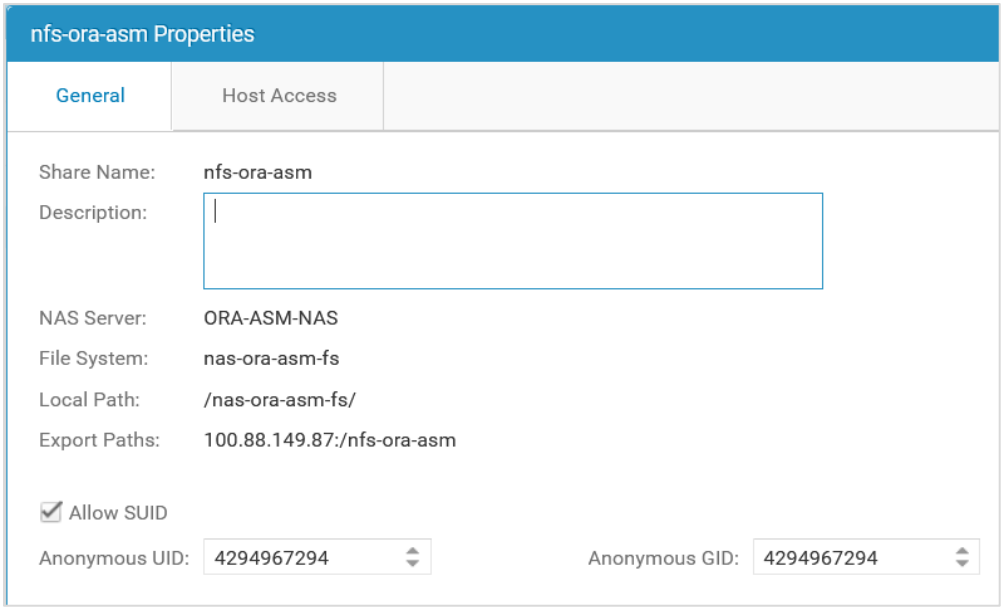

Figure 22 Dell EMC Unity NFS share

NFS client (database server) NIC interface configuration:

```
[root@r730xd-1 ~]# cat /etc/sysconfig/network-scripts/ifcfg-p1p1
<Snippet>
NAME=p1p1
DEVICE=p1p1
IPADDR=100.88.149.57
PREFIX=20
GATEWAY=100.88.144.1
```

dNFS uses two kinds of NFS mounts: the native OS mount of NFS (also referred to as kernel kNFS mount) and the Oracle database NFS mount (dNFS mount). When using a single network path for dNFS, file oranfstab is not necessary because Oracle dNFS will glean the required information for the matching mounted NFS share in file **/etc/mtab**. If dNFS is unable to find the necessary information in **/etc/mtab**, control is handed back to the database and file access is attempted through kNFS.

Information regarding Dell EMC Unity NFS share /ORA-ASM-NFS is shown in **/etc/mtab**:

```
[root ~]# grep oraasmnas /etc/mtab
100.88.149.87:/ORA-ASM-NFS /oraasmnas nfs
rw,relatime,vers=3,rsize=32768,wsize=32768,namlen=255,acregmin=0,acregmax=0,acdi
rmin=0,acdirmax=0,hard,proto=tcp,timeo=600,retrans=2,sec=sys,mountaddr=100.88.14
9.87,mountvers=3,mountport=1234,mountproto=tcp,local_lock=none,addr=100.88.149.8
7 0 0
```

If the IP used in the single network path happens to be in the same subnet used by any other NIC interface in the database server, see section 7.8.2 for additional requirements. For additional information on file oranfstab, see section 7.11.

7.7 Multiple network path for dNFS

If multiple network paths are intended for NFS traffic, consider using one path for NFS control/management traffic and the remaining NFS paths for NFS data traffic. This ensures that NFS data paths are only used for NFS data traffic.

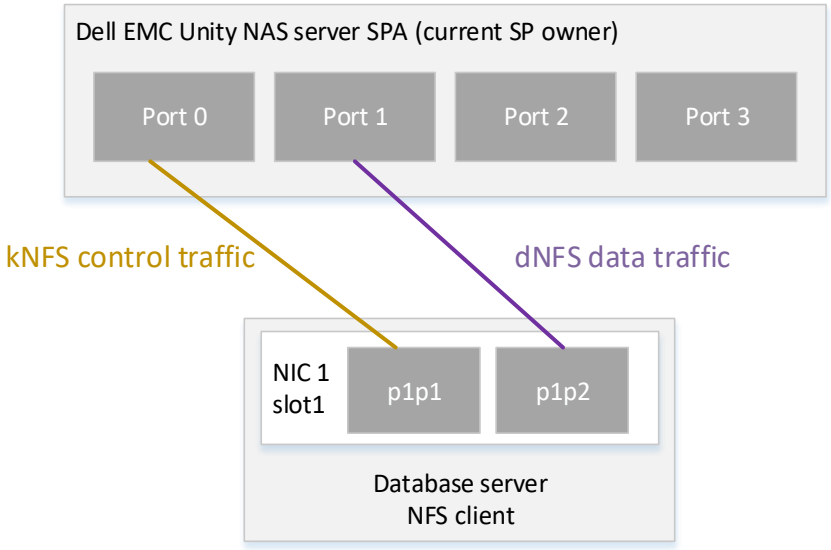

Figure 23 Dedicated interfaces for dNFS control and dNFS data traffic

If the architecture cannot support dedicated paths for all dNFS data traffic, dNFS control and data traffic can share a path.

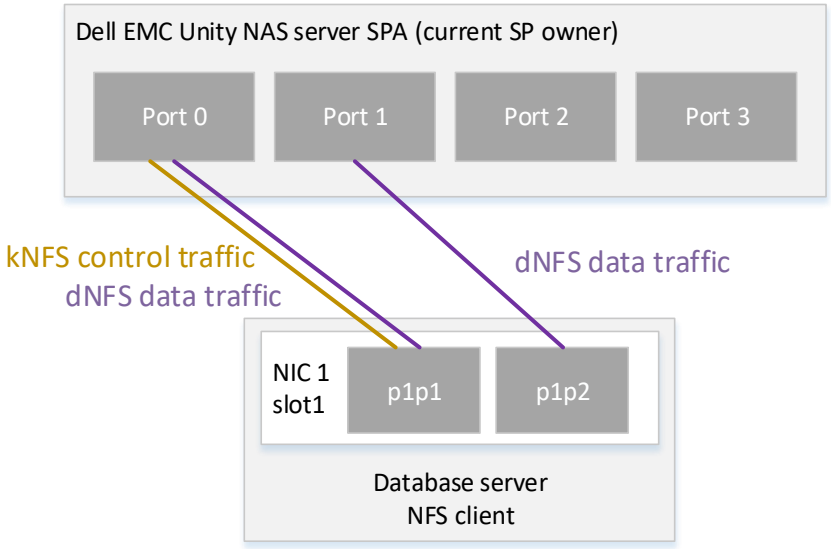

Figure 24 Shared interface for dNFS control and data traffic

If multiple dNFS paths are defined for data traffic, when a dNFS data path fails, dNFS reissues requests over any of the remaining dNFS data paths, thereby improving database availability. The multiple data paths also provide Oracle the ability to automatically tune the data paths to the NFS storage devices, thereby avoiding the need to manually tune NFS network performance at the OS level. Since dNFS implements multipath I/O internally, there is no need to configure LACP for channel-bonding interfaces for dNFS data traffic through active-backup or link aggregation. Therefore, if the LACP protocol is configured on the NIC interfaces

intended for dNFS data traffic, remove the channel-bond on those interfaces so that the interfaces operate as independent ports.

If a single interface is used for the OS kNFS mount, NFS control traffic can be blocked should the interface be down or the network cable unplugged. This blocked NFS traffic will cause the database to appear hung. To mitigate this single point of failure in the network, LACP protocol should be configured on multiple interfaces to create a channel-bonded interface for NFS control/management traffic. This is the recommended configuration and it provides increased database availability and additional network bandwidth.

For additional information on channel bonded interfaces for NFS control traffic, see section 7.9.

When configuring dNFS with multiple network paths, the recommendation is to use a unique network for each of the paths. When multiple unique networks are not available, or not desired, multiple IPs from the same subnet can be used for each of the network paths. See section 7.8 for additional requirements if a shared subnet is used for dNFS data traffic.

7.8 Shared subnets

When configuring dNFS to use a subnet that is currently being used by at least one other network interface, additional configuration is required including the Ipv4 network routing filter and static routing.

7.8.1 Ipv4 network routing filters

Linux 6 and 7 follow the recommendations of ingress filtering for multihomed networks (http://tools.ietf.org/html/rfc3704). These routing filters must be relaxed in order for multiple NIC interfaces in the same server to use the same subnet. Therefore, before configuring network interfaces to use the same subnet, make sure to relax routing filters.

If the Oracle 12c preinstall rpm is used to configure the OS before installing Oracle, the routing filters will be relaxed appropriately. Beginning with Oracle Database 12c release 2, Oracle has changed the name of this rpm so that the name corresponds to version of Oracle being installed:

- Oracle Database 12cR2: oracle-database-server-12cR2-preinstall.rpm
- Oracle Database 12cR1: oracle-rdbms-server-12cR1-preinstall.rpm

Both rpms are in the ol7_latest repository for Oracle Linux 7 on the Oracle Linux yum server as well as from ULN. Recent releases of Oracle Linux 7 by default include the proper yum configuration to install these rpms. If the rpm is missing from the OS, execute the following to install it:

Oracle 12cR1:

```
yum info oracle-rdbms-server-12cR1-preinstall
```

Oracle 12cR2:

```
yum info oracle-database-server-12cR2-preinstall
```

To verify if ipv4 routing filters have been relaxed in the current running OS, execute the following on the database server from a privileged OS user. The values of the returned parameters should be 2 if ipv4 routing filters have been relaxed.

```
[root ~]# sysctl net.ipv4.conf.all.rp_filter
net.ipv4.conf.all.rp_filter = 2
[root ~]# sysctl net.ipv4.conf.default.rp_filter
net.ipv4.conf.default.rp_filter = 2
```

If the filters are not set correctly, update **/etc/sysctl.conf** with the settings so they are persistent across reboots, and reload the system configuration:

```
[root ~]# vi /etc/sysctl.conf
net.ipv4.conf.all.rp_filter = 2
net.ipv4.conf.default.rp_filter = 2
[root ~]# sysctl -p
```

7.8.2 Static routing

When any path for dNFS traffic shares any subnet already in use on the NFS client (database server), static routing must be configured for each of the paths used by dNFS traffic. This includes defining a static route for dNFS data traffic should it use the same subnet as dNFS control traffic. If static routes are not defined, automatic load balancing and performance tuning of dNFS will not operate as expected per the dNFS path definitions in file **oranfstab,** and NFS data traffic will flow through an unexpected network path.

The remainder of this section covers two examples of static routing. The first example considers two interfaces sharing the same subnet, and the second example considers four network interfaces sharing the same subnet.

Shared subnet on two interfaces:

Figure 25 illustrates the path taken for dNFS data traffic when a subnet is shared on two interfaces with default routing. The illustration shows dNFS traffic flowing through interface em1 rather than through the intended interface p1p1.

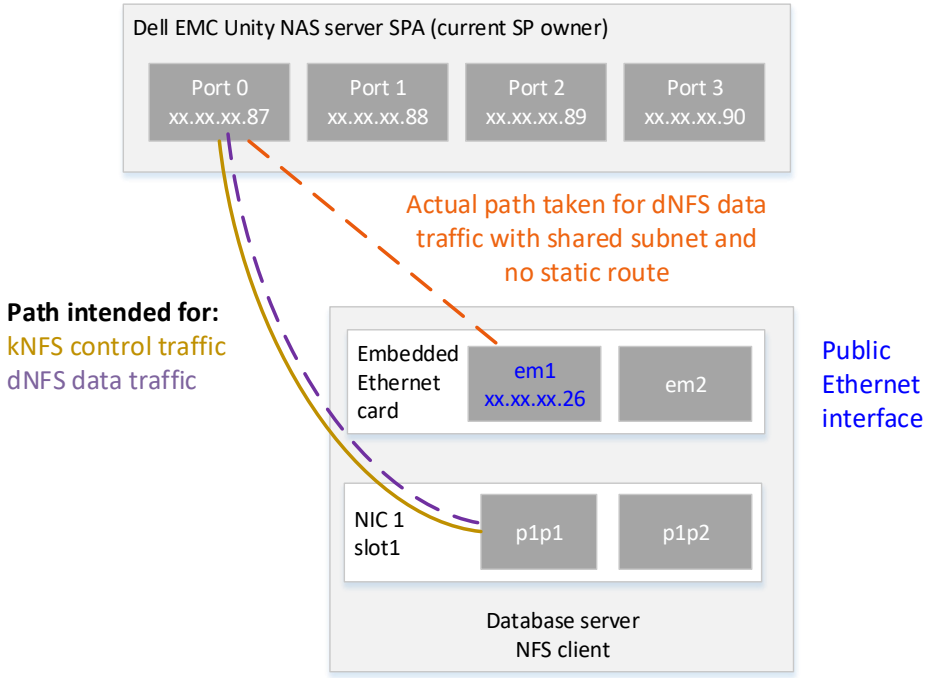

Figure 25 Incorrect path taken by dNFS data traffic on shared subnet and default routing

Default routing table – single interface for dNFS:

```
[root ~]# route -n
Kernel IP routing table
Destination     Gateway         Genmask         Flags Metric Ref    Use Iface
0.0.0.0         100.88.144.1    0.0.0.0         UG    0      0        0 em1
100.88.144.0    0.0.0.0         255.255.240.0   U     0      0        0 em1
100.88.144.0    0.0.0.0         255.255.240.0   U     0      0        0 p1p1
```

If default routing is used, the OS searches the routing table for the route that best matches the destination address and mask, and it will use that route. Since interfaces em1 and p1p1 share the same subnet value in column **Destination** and mask in column **Genmask**, the OS will consider both entries as best-matched for the target address (100.88.149.87 /20 – port 0 of Dell EMC Unity storage) and will select the first entry as the route. Since the route for em1 precedes the entry for p1p1, the OS will use the route to interface em1 for dNFS data traffic rather than the route to the intended interface p1p1.

To mitigate the issue of sending dNFS data traffic across the wrong path, a static route must be added to the route table that will force dNFS data traffic to flow between the intended p1p1 interface and target address (100.88.149.87 /20). The following command adds the necessary route to the routing table:

```
ip route add 100.88.149.87 dev p1p1
```

After adding the static route, verify the routing table is updated with the appropriate route:

```
 [root ~]# route -n
Kernel IP routing table
Destination     Gateway         Genmask         Flags Metric Ref    Use Iface
0.0.0.0         100.88.144.1    0.0.0.0         UG    0      0        0 em1
100.88.144.0    0.0.0.0         255.255.240.0   U     0      0        0 em1
100.88.144.0    0.0.0.0         255.255.240.0   U     0      0        0 p1p1
100.88.149.87   0.0.0.0         255.255.255.255 UH    0      0        0 p1p1
```

With the necessary route in place, Figure 26 shows dNFS traffic flowing through the intended interface p1p1.

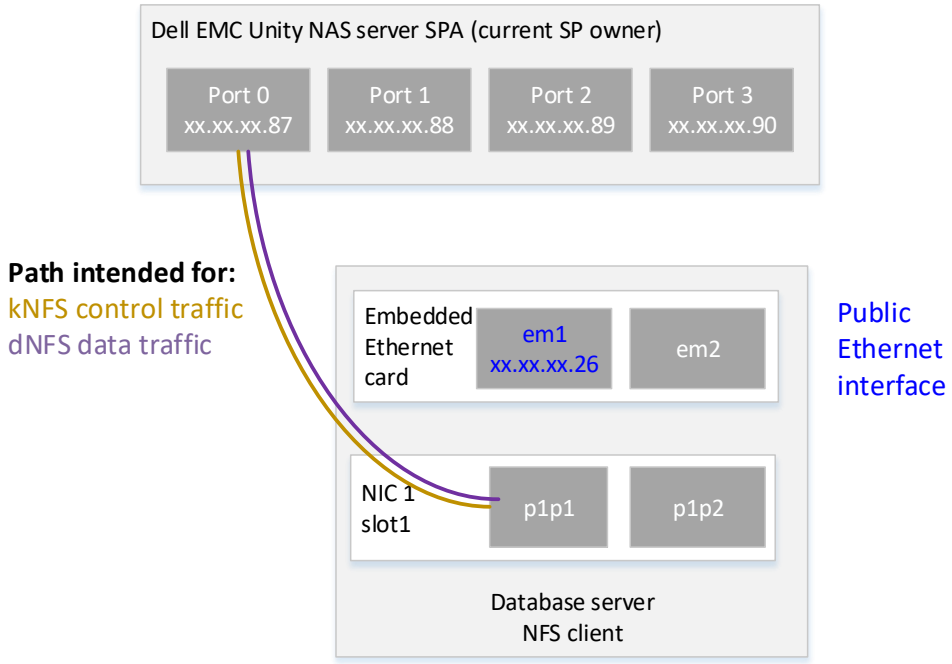

Figure 26 Correct path taken for dNFS data traffic

If modifying the route table is not desired, static routing is possible through interface routing scripts in directory **/etc/sysconfig/network-scripts**. This type of configuration will not be static across reboots.

```
echo "100.88.149.57 via 100.88.149.87" > /etc/sysconfig/network-scripts/route-p1p1/etc/sysconfig/network-scripts/ifup-routes p1p1
```

Shared subnet on four network interfaces:

This example illustrates how default and static routing change the paths taken on four interfaces (p<s>p<p>) configured with IPs from the same subnet for dNFS data traffic.

If the following default routing is used, all dNFS traffic would again flow through interface em1 because it is the first best-matched entry in the routing table.

```
[root ~]# route -n

Kernel IP routing table
Destination     Gateway         Genmask         Flags Metric Ref    Use Iface
0.0.0.0         100.88.144.1    0.0.0.0         UG    0      0        0 em1
100.88.144.0    0.0.0.0         255.255.240.0   U     0      0        0 em1
100.88.144.0    0.0.0.0         255.255.240.0   U     0      0        0 p1p1
100.88.144.0    0.0.0.0         255.255.240.0   U     0      0        0 p1p2
100.88.144.0    0.0.0.0         255.255.240.0   U     0      0        0 p2p1
100.88.144.0    0.0.0.0         255.255.240.0   U     0      0        0 p2p2
[root ~]#
```

To direct dNFS traffic to flow through all intended interfaces, the following static routes are needed:

```
ip route add 100.88.149.87 dev p1p1
ip route add 100.88.149.88 dev p1p2
ip route add 100.88.149.89 dev p2p1
ip route add 100.88.149.90 dev p2p2
```

Once the routes are defined between the interfaces and the four end-point IP addresses in the Dell EMC Unity NAS server, those traffic routes will be chosen because they best fit the destination and mask.

```
[root ~]# route -n
Kernel IP routing table
Destination     Gateway         Genmask         Flags Metric Ref    Use Iface
<snippet>
100.88.144.0    0.0.0.0         255.255.240.0   U     0      0        0 p1p1
100.88.144.0    0.0.0.0         255.255.240.0   U     0      0        0 p1p2
100.88.144.0    0.0.0.0         255.255.240.0   U     0      0        0 p2p1
100.88.144.0    0.0.0.0         255.255.240.0   U     0      0        0 p2p2
100.88.149.87   0.0.0.0         255.255.255.255 UH    0      0        0 p1p1
100.88.149.88   0.0.0.0         255.255.255.255 UH    0      0        0 p1p2
100.88.149.89   0.0.0.0         255.255.255.255 UH    0      0        0 p2p1
100.88.149.90   0.0.0.0         255.255.255.255 UH    0      0        0 p2p2
[root ~]#
```

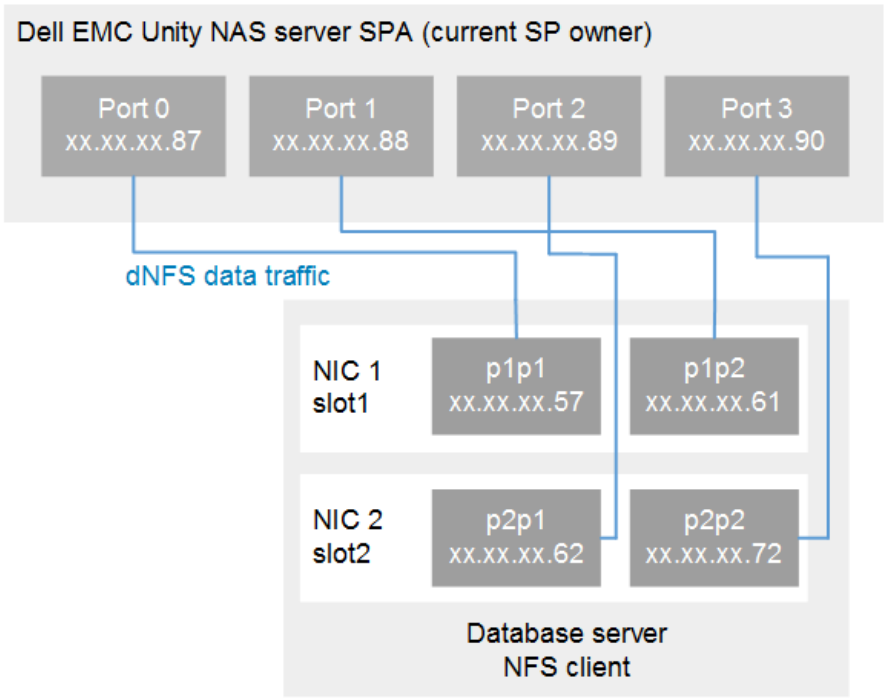

Figure 27 Correct path taken for multiple dNFS data paths

If modifying the route table is not desired, static routing is possible through interface routing scripts in the directory **/etc/sysconfig/network-scripts**. This type of configuration will not be static across reboots.

```
echo "100.88.149.57 via 100.88.149.87" > /etc/sysconfig/network-scripts/route-
p1p1
echo "100.88.149.61 via 100.88.149.88" > /etc/sysconfig/network-scripts/route-
p1p2
echo "100.88.149.62 via 100.88.149.89" > /etc/sysconfig/network-scripts/route-
p2p1
echo "100.88.149.72 via 100.88.149.90" > /etc/sysconfig/network-scripts/route-
p2p2
/etc/sysconfig/network-scripts/ifup-routes p1p1
/etc/sysconfig/network-scripts/ifup-routes p1p2
/etc/sysconfig/network-scripts/ifup-routes p2p1
/etc/sysconfig/network-scripts/ifup-routes p2p2
```

Static routing can also be defined in Dell EMC Unity storage when adding or updating the configuration of a NAS server. See section 5.2 for more information.

See Table 10 for examples of IP-address mapping from end point to end point, dNFS traffic type, and LACP.

7.9 Configuring LACP

In environments requiring high availability, a bonded NIC interface for NFS control traffic is recommended.

7.9.1 NFS client (database server) and channel-bonding configuration

If an unbonded interface is used for NFS control traffic and that interface sustains an outage, the database can appear hung under certain operations. To mitigate this single point of failure, LACP protocol should be configured on multiple interfaces to create a channel-bonded interface for NFS control/management traffic. This bonded interface could be the bonded public network or even the bonded interface for the RAC interconnect in a RAC environment. Having a dedicated bonded network for NFS control traffic should not be necessary as the NFS control or metadata traffic should be minimal.

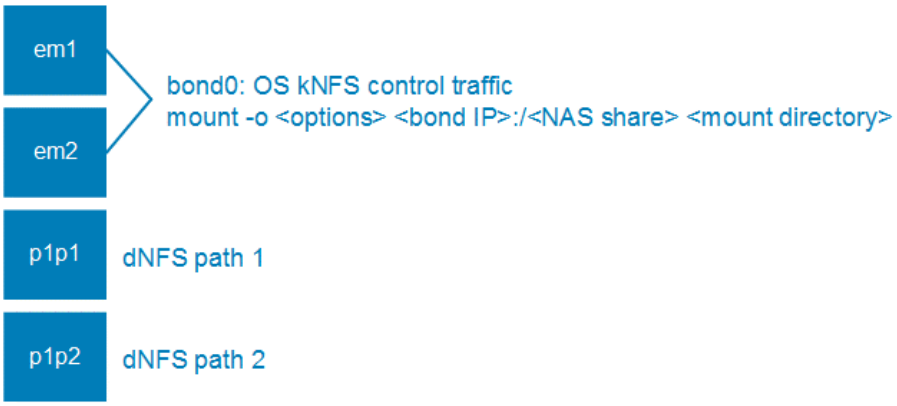

Figure 28 Bond interface for NFS control traffic

If LACP is configured on the NFS client for NFS control traffic, LACP must be configured in the Dell EMC Unity system by creating link aggregations, and by configuring port channels in the Ethernet switches connecting the Dell EMC Unity and NFS client interfaces. Link aggregations with Dell EMC Unity interfaces provide redundancy and additional bandwidth especially when multiple NFS database clients exist. In practice, link aggregations in Dell EMC Unity storage should be done only if the second link is needed for highly available configurations.

If the channel-bonded interface on the NFS client will be dedicated to NFS control traffic, it is recommended to use 1GbE network interfaces. Using 10GbE links for the dedicated channel-bonded interface for NFS control traffic may be a waste of interface resources with respect to addition bandwidth. There is benefit however from the perspective of increased availability. Should one of the interface members of the channel-bond suffer an outage, there is still another working interface in the channel-bond that traffic can flow through.

7.9.2 NAS server (Dell EMC Unity) and link aggregation configuration

If NFS traffic will flow through bonded interfaces on the NFS client (database server), front-end connectivity of the Dell EMC Unity system must also be configured appropriately to support the bonded interfaces of the NFS client. When configuring a bonded interface (link aggregate) in the Dell EMC Unity system, the candidate interfaces for the bonded interfaces from both SP A and SP B must be cabled and configured in both SP A and SP B modules before the Dell EMC Unity system will start either of the interface members of the bonded interface. If not, Dell EMC Unisphere will display a status of Link Down for the interface members of the link aggregate. The following illustration shows the Link Aggregation up in both SP A and SP B because both ports from both SPs were cabled.

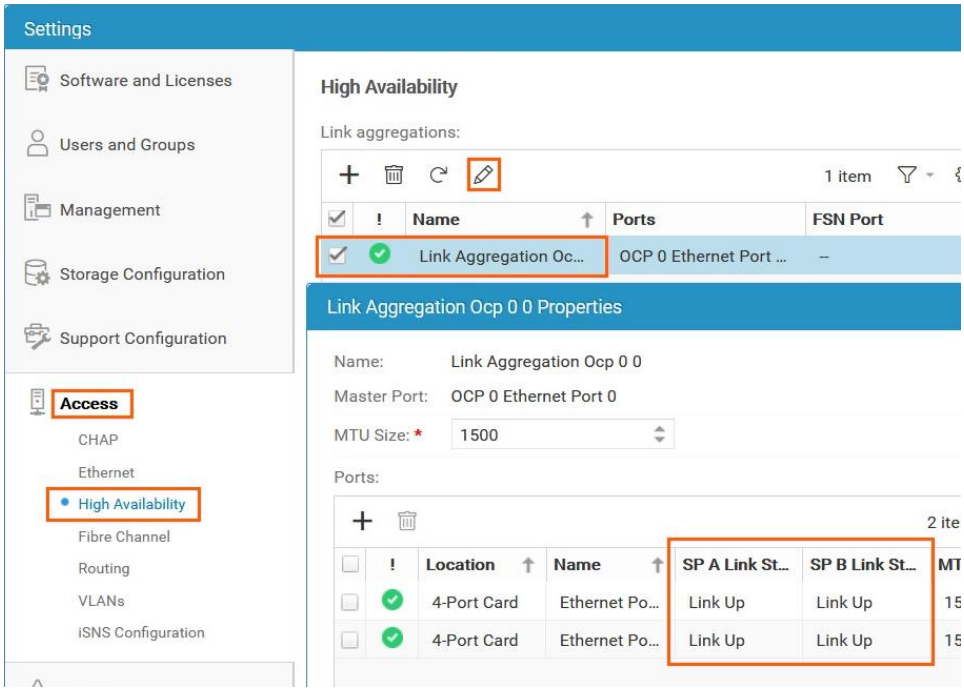

Figure 29 NAS server link aggregation

Both bonded interfaces must also use the same ports from both SPs. This is necessary because in case of failover, the peer SP uses the same ports. LACP can be configured across the ports from the same I/O module, but cannot be configured on ports that are also used for iSCSI connections. In earlier Dell EMC Unity All Flash arrays, LACP could be configured across the on-board Ethernet ports.

If a link aggregate contains two interfaces, a total of four switch interfaces will be required: two switch interfaces for the two SP A interfaces in the link aggregate, and two switch interfaces for the two SP B interfaces in the link aggregate. See Figure 34 for an illustration.

Link aggregation in Dell EMC Unity storage is configured from within the **Update system settings** wizard. To start the **Update system settings** wizard, select the gear ion in the menu bar:

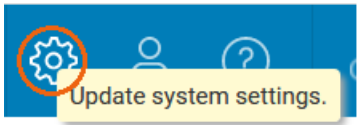

Figure 30 Update system settings wizard

In the **Settings** wizard, select **Access** > **High Availability** to manage or view link aggregations. Then, select **+** from the **Link aggregations** section to configure a bonded Dell EMC Unity interface.

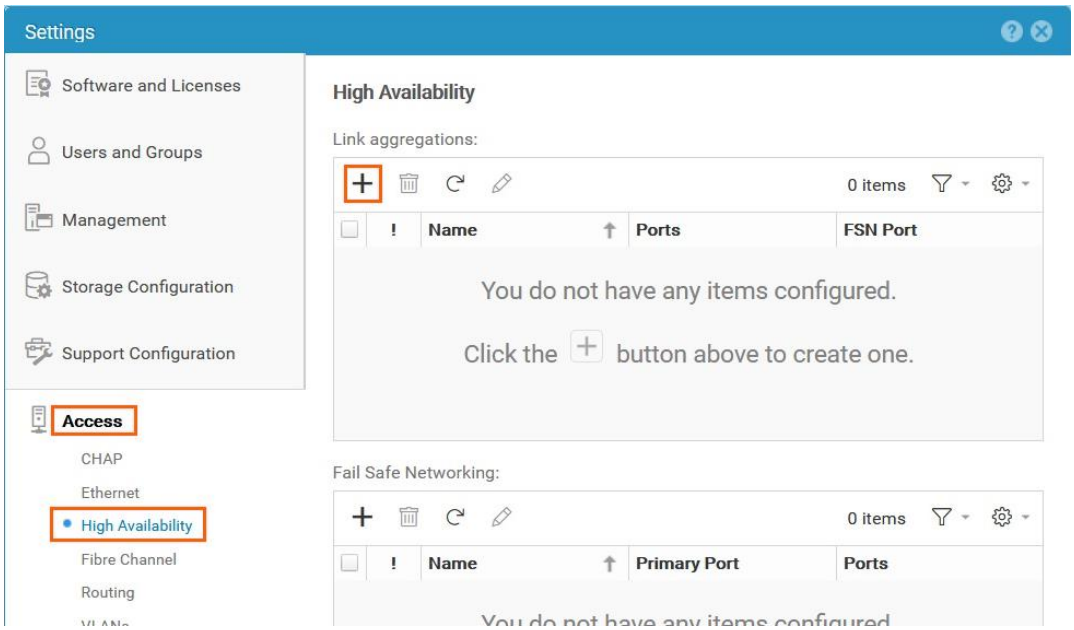

Figure 31 Creating Unity link aggregation for nfs control traffic

Setting the master and slave ports of the bonded interface will be the first steps taken.

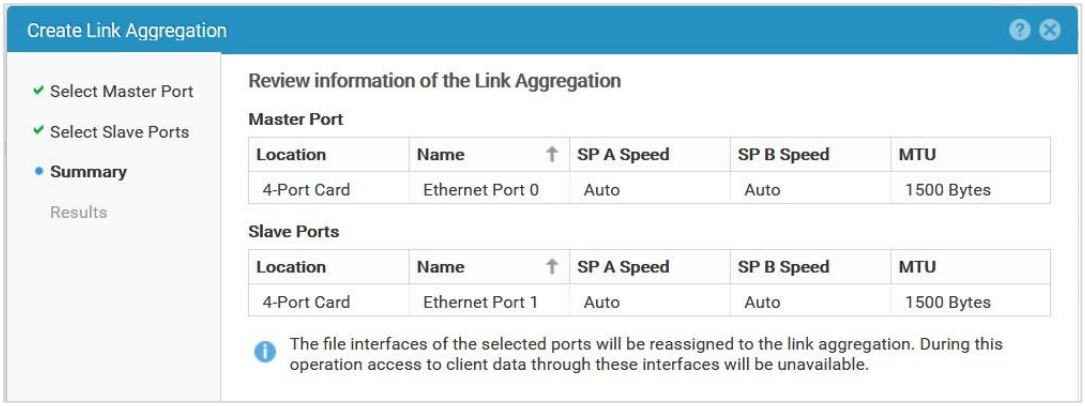

Figure 32 Unity link aggregation summary

If the bond interface is needed for dedicated NFS control traffic, MTU 1500 may be sufficient, but consider using Jumbo frames (MTU 9000). See section 7.5 for additional information.

The link aggregate can be added to the NFS server from the network properties in Unisphere: click **File** > **NAS Servers** > **edit** (pencil icon) > **Network** > **Interfaces & Routes** > **+** > **Production IP interface**. Set **Ethernet Port:** to the link aggregate created for the NFS traffic and provide the necessary networking information (IP address, subnet mask/prefix length (or CIDR), gateway) for the link aggregate.

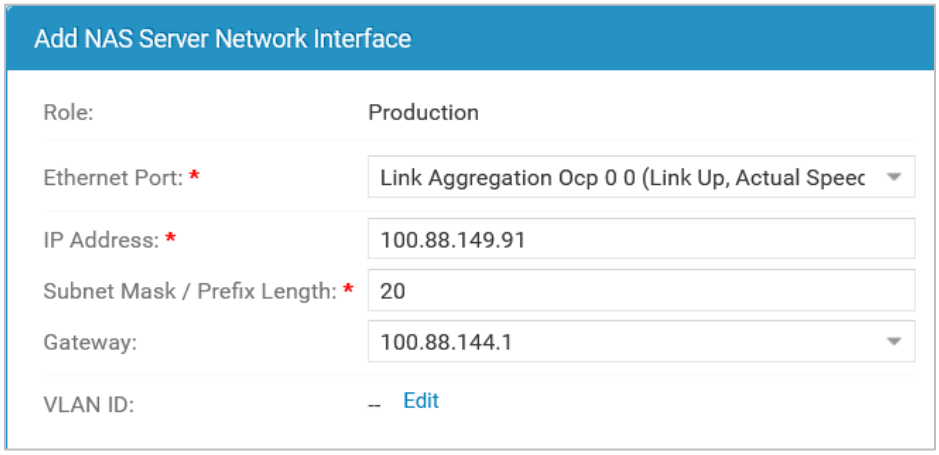

Figure 33 Defining network information for the link aggregation

Then, when mounting the NFS share on the NFS client, mount the NFS share with the IP address specified in the link aggregate interface.

```
mount –o <options> 100.88.149.91:/ora-asm-nfs-test /oraasmnas-test
```

7.9.3 Ethernet switch and port channel configuration

If NFS control traffic will flow through a bonded NFS client (database server) NIC interface and a link aggregate in Dell EMC Unity storage, Ethernet switch ports (switch interfaces) cabled to the database server NIC interfaces and Dell EMC Unity interfaces must also be configured with LACP. If the candidate switch interfaces for the bonded interfaces are in a VLAN, remove them from the VLAN before configuring the port channel.

Figure 34 illustrates how switch interfaces were configured as port channels in a Dell EMC Networking S5000 switch. The port channels will be used for NFS control traffic. Port channel 1 will be used for Dell EMC Unity SP module A and port channel 2 will be used with Dell EMC Unity SP module B.

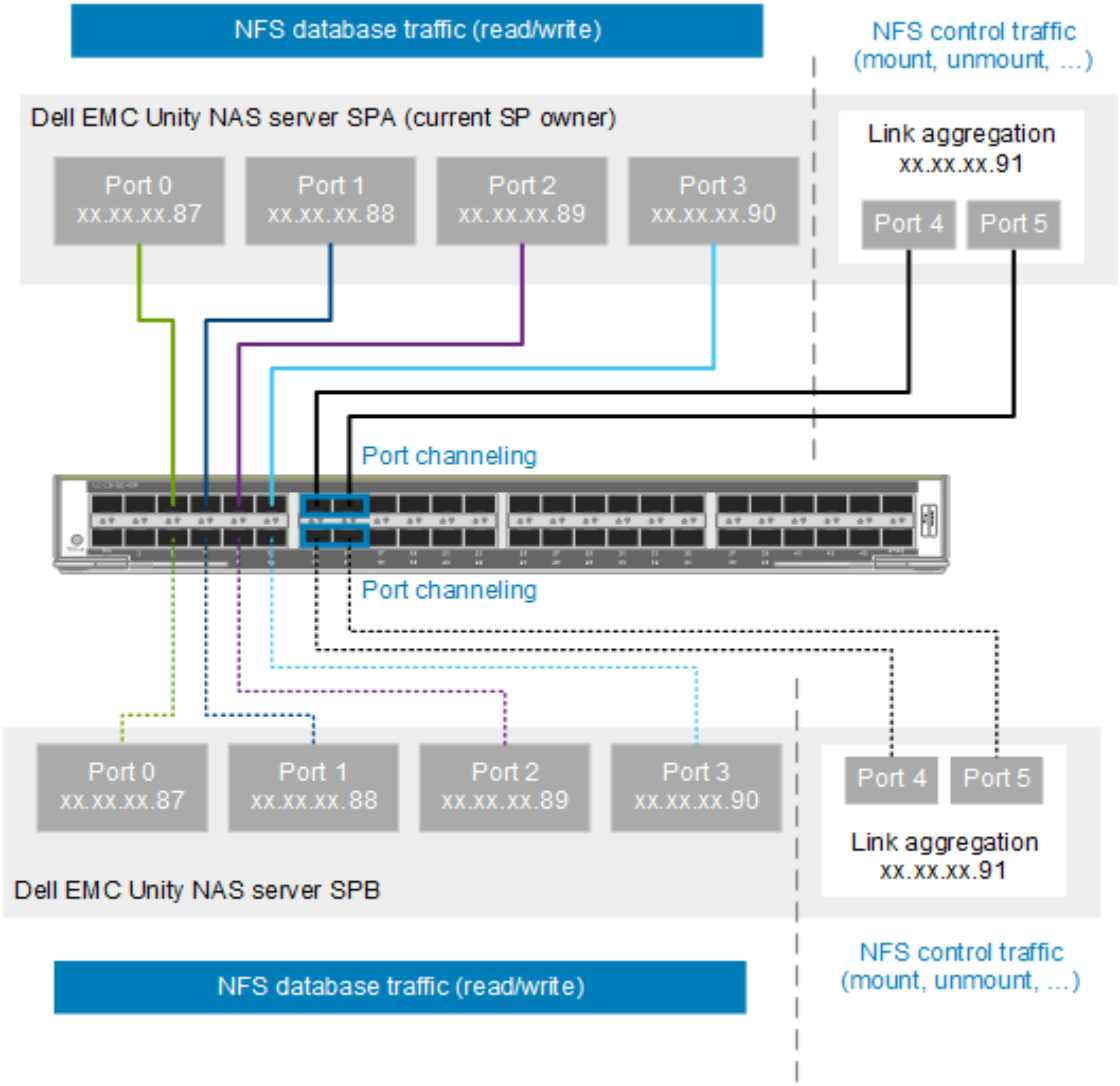

Figure 34 Cabling between Dell EMC Unity 650F storage and an Ethernet switch

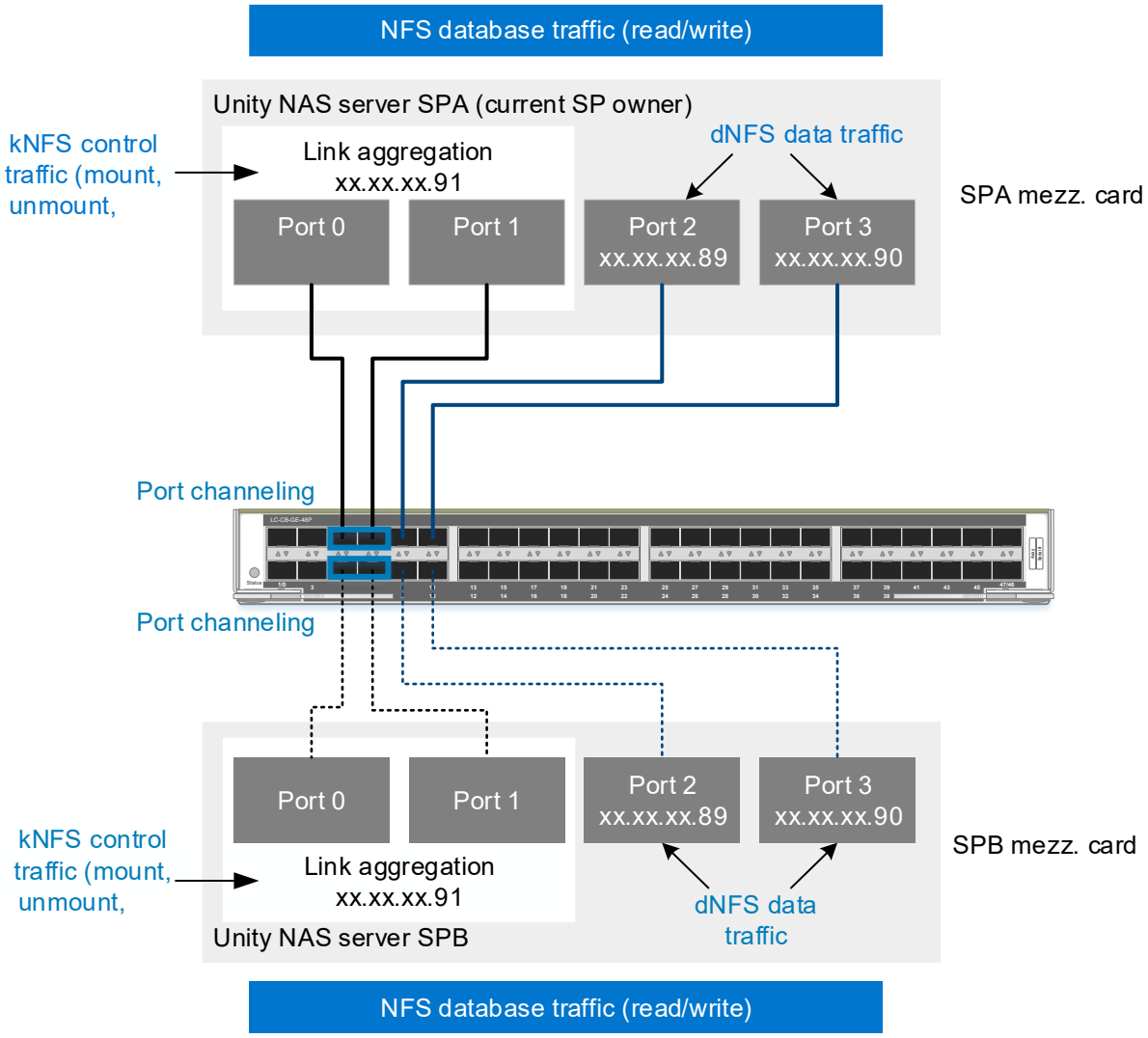

Figure 35 Cabling between Dell EMC Unity 480F, 680F, 880F storage and an Ethernet switch

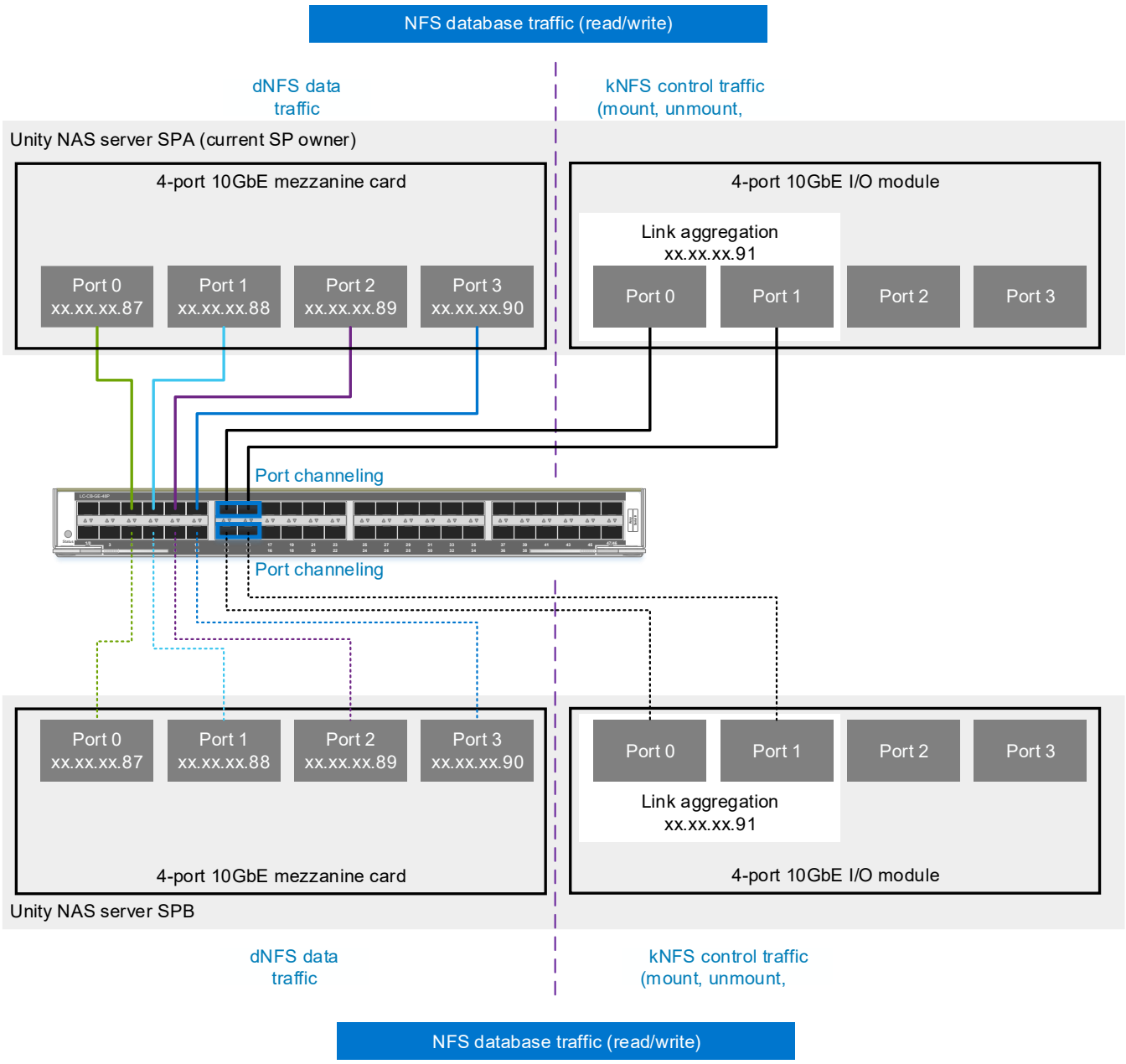

Figure 36 Cabling between Dell EMC Unity 480F, 680F, 880F storage and an Ethernet switch

Switch interfaces that will be connected to the channel-bond interfaces of the NFS client (database server) also have to be configured with LACP.

For additional network redundancy for NFS traffic, use redundant switches to provide greater network availability.

7.10 Database server: NFS client network interface configuration

For best performance with Dell EMC Unity file storage, the database server should be configured with 10Gb/s and optionally with 1Gb/s for dNFS data traffic and NFS control traffic, respectively. If possible, these ports, including all end-to-end ports servicing dNFS data traffic, should be configured for Jumbo frames (MTU 9000) to provide best performance.

For NFS control/management traffic, either 1Gb/s or 10Gb/s ports can be used. For Oracle environments that require path redundancy for NFS control traffic, it is required to use LACP across multiple interfaces from end-point to end-point.

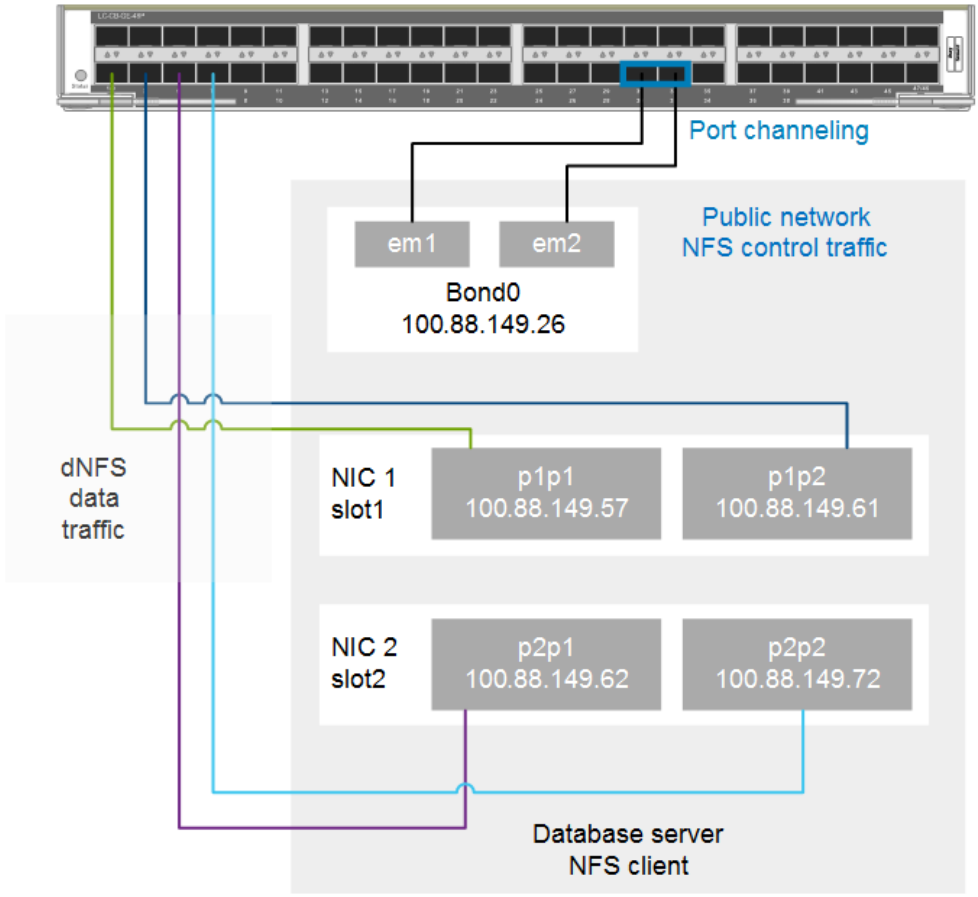

Figure 37 Ethernet connectivity between NFS client (database server) and Ethernet switch

The following snippets are for the network configuration files for the interfaces shown previously and correspond to the interface address in the OS static routes and dNFS channels defined in file oranfstab:

```
[root ~]# cd /etc/sysconfig/network-scripts
[root network-scripts]# cat ifcfg-em1
TYPE=Ethernet
DEFROUTE=yes
NAME=em1
DEVICE=em1
SLAVE=yes
MASTER=bond0
<snippet>

[root 2 network-scripts]# cat ifcfg-em2
TYPE=Ethernet
DEFROUTE=yes
NAME=em2
DEVICE=em2
SLAVE=yes
MASTER=bond0
<snippet>

[root network-scripts]# cat ifcfg-bond0
TYPE=Bond
DEFROUTE=yes
DEVICE=bond0
USERCTL=no
IPADDR=100.88.149.26
PREFIX=20
GATEWAY=100.88.144.1
BONDING_MASTER=yes
<snippet>

[root network-scripts]# cat ifcfg-p1p1
TYPE=Ethernet
DEFROUTE=no
NAME=p1p1
DEVICE=p1p1
IPADDR=100.88.149.57
PREFIX=20
GATEWAY=100.88.144.1
<snippet>
```

```
[root network-scripts]# cat ifcfg-p1p2
TYPE=Ethernet
DEFROUTE=no
NAME=p1p2
DEVICE=p1p2
IPADDR=100.88.149.61
PREFIX=20
GATEWAY=100.88.144.1
<snippet>

[root network-scripts]# cat ifcfg-p2p1
TYPE=Ethernet
DEFROUTE=no
NAME=p2p1
DEVICE=p2p1
IPADDR=100.88.149.62
PREFIX=20
GATEWAY=100.88.144.1
<snippet>

[root network-scripts]# cat ifcfg-p2p2
TYPE=Ethernet
DEFROUTE=no
NAME=p2p2
DEVICE=p2p2
IPADDR=100.88.149.72
PREFIX=20
GATEWAY=100.88.144.1
<snippet>
```

Table 10 Example of end-point mappings, dNFS traffic type, and LACP

NAS server port	NAS server IP	Host interface	Host interface IP	NFS traffic type	LACP
2	100.88.149.89	p2p1	100.88.149.75	Data	No
3	100.88.149.90	p2p2	100.88.149.76	Data	No
Link aggregation 1 (port 0)	100.88.149.91	bond0 (em1)	100.88.149.117	Control	Yes
Link aggregation 1 (port 1)	100.88.149.91	bond0 (em2)	100.88.149.117	Control	Yes

For additional information on bonded interfaces, see section 7.9.

7.11 Oracle dNFS configuration file: oranfstab

oranfstab is used by Oracle to determine which mount points are available to dNFS and how to configure dNFS network paths (referred to as channels) between the NFS servers and dNFS client.

If oranfstab does not exist and assuming the NFS file systems have been mounted, dNFS will mount and create a single dNFS channel for entries found in **/**etc/mtab that are required for the running database. The dNFS channel in Oracle will have a name equal to the IP address of the mount entry in /etc/mtab. No additional configuration is required.

The following shows the **/etc/fstab** and **/etc/mtab** entry for single NFS share:

```
[root ~]# grep ORA-ASM-NFS /etc/fstab
100.88.149.91:/ORA-ASM-NFS /oraasmnas nfs
rw,bg,hard,nointr,rsize=32768,wsize=32768,tcp,vers=3,timeo=600,actimeo=0

[root ~]# grep ORA-ASM-NFS /etc/mtab
100.88.149.87:/ORA-ASM-NFS /oraasmnas nfs
rw,relatime,vers=3,rsize=32768,wsize=32768,namlen=255,acregmin=0,acregmax=0,acdi
rmin=0,acdirmax=0,hard,proto=tcp,timeo=600,retrans=2,sec=sys,mountaddr=100.88.14
9.87,mountvers=3,mountport=1234,mountproto=tcp,local_lock=none,addr=100.88.149.8
7 0 0
SQL> select distinct svrname, path, ch_id, svr_id from v$dnfs_channels;

SVRNAME          PATH                 CH_ID     SVR_ID
---------------- ---------------- ---------- ----------
100.88.149.87    100.88.149.87             0          1
```

If multiple channels to a NAS server are needed for increased dNFS bandwidth, automatic dNFS data traffic load balancing, or automatic dNFS channel failover, the file oranfstab is required. dNFS automatically performs load balancing across all specified available channels, and if one channel fails, dNFS reissues I/O commands over any remaining available channel for that NAS server.

oranfstab can reside in either /etc or $ORACLE_HOME/dbs. If oranfstab resides in /etc, its contents will be global to all databases running on that server regardless of which ORACLE_HOME they are running from. If oranfstab resides in $ORACLE_HOME/dbs, then it will be global to any database running from that ORACLE_HOME. If ORACLE_HOME is shared between RAC nodes, all RAC databases running from the shared $ORACLE_HOME will use the same $ORACLE_HOME/dbs/oranfstab.

dNFS searches for mount entries in the following order and uses the first matching entry as the mount point:

- $ORACLE_HOME/dbs/oranfstab
- /etc/oranfstab
- /etc/mtab

If a database uses dNFS mount points configured in oranfstab, Oracle first verifies kNFS mount points by cross-checking entries in mtab and oranfstab. If a match does not exist, dNFS logs a message and fails to operate.

The following oranfstab file contains 4 dNFS data paths to two NAS server aliases, each NAS server alias is for a different database. Format of data paths can vary within oranfstab:

```
server: ORA-NAS01
local: 100.88.149.57 path: 100.88.149.63
local: 100.88.149.62 path: 100.88.149.65
local: 100.88.149.61 path: 100.88.149.64
local: 100.88.149.72 path: 100.88.149.66
mnt_timeout: 60
export: /ORA-FS1 mount: /ora1db
#
server: ORA-ASM-NFS
local: 100.88.149.57 path: 100.88.149.87
local: 100.88.149.62 path: 100.88.149.88
local: 100.88.149.61 path: 100.88.149.89
local: 100.88.149.72 path: 100.88.149.90
mnt_timeout: 60
export: /ORA-ASM-NFS mount: /oraasmnas
```

The following channels for ORA-ASM-NFS will be created. Channels for ORA-NAS01 are not shown because the current database relies only on ORA-ASM-NFS:

```
SQL> select distinct svrname
  2                , path
  3                , ch_id
  4                , svr_id
  5     from v$dnfs_channels
  6     order by ch_id;

SVRNAME         PATH                 CH_ID     SVR_ID
--------------- --------------- ---------- ----------
ORA-ASM-NFS     100.88.149.87             0          1
ORA-ASM-NFS     100.88.149.88             1          1
ORA-ASM-NFS     100.88.149.89             2          1
ORA-ASM-NFS     100.88.149.90             3          1
```

In Linux, if any NFS data path (column PATH) is defined by an IP existing in a subnet used by any other NIC interface on the database server, static routes must be defined in the OS for that NFS data path. See section 7.8 for more information.

Table 11 presents the available configuration parameters for oranfstab.

Table 11 Oranfstab configuration parameters

oranfstab directive	Description
server	This can be any name. The name uniquely identifies and is used to begin a group of directives for dNFS that controls the way in which dNFS should operate on the mounted NFS Shares indicated by the pair of **export** and **mount** values in the group. The value of **server** will also be used as an identifier in v$dnfs views and logging. For readability and supportability, it is recommended to set the value of **server** name to the name of the NAS server specified in the mount command.
local	The IP of the interface on the database server designated for NFS data traffic. The value of **local** and **path** define the end-to-end point taken for NFS data traffic. Up to four **local** and **path** pairs can be specified. If there are more than one local-path pairs, automatic load balancing and failover on dNFS data paths will be enabled.
path	The IP of the interface of the NAS server that will be used with the above **local** IP. The value of **path** and **local** define the end-to-end point taken for NFS data traffic. Up to four **local** and **path** pairs can be specified. If there are more than one local-path pairs, automatic load balancing and failover on dNFS data paths will be enabled.
export: <value> mount: <value>	This is a pair of values that cannot be broken between lines. The paired values consist of the name of the NFS Share or volume in Dell EMC Unity storage that has been exported to the NFS client (database server), and the file system mount point on the database server that will be used for the NFS share. Both values must match the appropriate corresponding paired values in **/etc/mtab** and **/etc/fstab**. The number of export-mount pairs within a server stanza is unlimited.
dontroute	**Note:** This is directive is not applicable in Linux. If specified, it will be ignored. It is intended for POSIX related OSs and instructs the OS to ignore the routes specified in the OS routing table. This guarantees that dNFS will use the routes specified by **local** and **path** in this file. To ensure proper routing occurs in Linux, use static routing. See section 7.8 for additional information.
mnt_timeout	Optional: This is the time in seconds that dNFS will wait for a successful mount before timing out. The default is 600 seconds.
nfs_version	Optional: For 12c, this specifies the version of NFS: nfsv4 or nfsv3 (default).
management	Optional. For 12c, use the management interface for SNMP.
community	Optional. For 12c, this defines the community string for SNMP.

7.12 Enabling and disabling Oracle dNFS

After installing 12c RDMBS, enabling and disabling dNFS is done by executing the following commands from the Linux user owning the ORACLE_HOME:

To enable dNFS:

```
cd $ORACLE_HOME/rdbms/lib
make -f ins_rdbms.mk dnfs_on
```

To disable dNFS:

```
cd $ORACLE_HOME/rdbms/lib
make -f ins_rdbms.mk dnfs_off
```

7.13 Verify if dNFS is being used

When there is I/O against the database, the following can be used to verify the Oracle instance is using dNFS channels and if the Ethernet network has been configured correctly.

If the alert log contains string **running with ODM**, dNFS has been enabled and the instance was started with the ODM library containing the direct NFS driver:

```
[oracle trace]$ grep 'instance running with ODM' alert_dbnfsasm.log
Oracle instance running with ODM: Oracle Direct NFS ODM Library Version 4.0
```

The local IP and path IP shown in the alert log should match all the records in oranfstab for the appropriate NAS server hosting the database. If Oracle automatically detects the local host interface because oranfstab is not defined, make sure the chosen interface is the one intended for the dNFS channel.

```
[oracle trace]$ grep 'Direct NFS: channel id' alert_dbnfsasm.log | tail -4
Direct NFS: channel id [0] path [100.88.149.87] to filer [ORA-ASM-NFS] via local
[100.88.149.57] is UP
Direct NFS: channel id [1] path [100.88.149.88] to filer [ORA-ASM-NFS] via local
[100.88.149.62] is UP
Direct NFS: channel id [2] path [100.88.149.89] to filer [ORA-ASM-NFS] via local
[100.88.149.61] is UP
Direct NFS: channel id [3] path [100.88.149.90] to filer [ORA-ASM-NFS] via local
[100.88.149.72] is UP
```

When there is database activity, there should be Ethernet activity on the interfaces corresponding to the local IPs defined in oranfstab, or on the interface used to mount the NAS share (assuming dNFS control and dNFS data traffic are routed through the same interface). The activity will be displayed as changes to (RX-OK and TX-OK) values from netstat:

The following lists send/receive (TX/RX) statistics for all interfaces:

```
[root ~]# netstat -i 5
Kernel Interface table
Iface      MTU    RX-OK RX-ERR RX-DRP RX-OVR    TX-OK TX-ERR TX-DRP TX-OVR Flg
<snippet>
p1p1      1500 138666544      0   9950 0      74959859      0      0      0 BMRU
p1p2      1500 150865538      0   9950 0      70397464      0      0      0 BMRU
p2p1      1500 151938399      0   9859 0      131360148      0      0      0
BMRU
p2p2      1500 133923499      0   9859 0      67984417      0      0      0 BMRU
<snippet>
```

The database will compare the datafile names with the NFS mount to see if the datafiles can be used by dNFS. Any datafile that dNFS can work with will reside in v$dnfs_files. Verify that all database files residing on the NFS share are seen by the database, and that there is activity on the dNFS channels.

```
select * from v$dnfs_files;
select pnum, svrname, path, local, ch_id, svr_id, sends, recvs
   from v$dNFS_channels;
```

7.14 Oracle dynamic dNFS views

Eight dNFS dynamic performance views are available in Oracle 12c to monitor ODM NFS storage devices. Four of the eight views are for standalone deployments, and four are for RAC deployments. A full description of the dynamic tables for standalone deployments can be found in the Oracle Database Reference 12cR2 at https://docs.oracle.com/database/122/REFRN/REFRN.pdf:

Table 12 Standalone deployment dNFS dynamic performance views:

dNFS dynamic performance views	Description
v$dnfs_channels	Displays open network paths/channels to servers for which dNFS is providing files
v$dnfs_files	Displays files currently open using dNFS
v$dnfs_servers	Displays servers accessed using dNFS
v$dnfs_stats	Displays performance statistics for dNFS

8 Dell EMC Unity features with Oracle databases

There are several features in the Dell EMC Unity system that provide extra enhancements and may provide additional benefits in an Oracle database environment. The following subsections provide best practices for these features and their integration with Oracle databases.

8.1 Data reduction

Data reduction is a Dell EMC Unity feature that includes both zero detection, compression, and advanced deduplication. This is the next level of space conservation. By offering multiple levels of space saving, Dell EMC provides flexibility for the best balance of space savings and performance.

Oracle provides database-level compression in its software. When database-level compression is enabled on the data, it is unlikely that the Dell EMC Unity system can further reduce consumption on these compressed data. Therefore, it is recommended that compression is applied by either the array or the database engine, but not both. Certain types of data, such as video, audio, image, and binary, usually get little benefit from compression.

Compression requires CPU resources and at high throughput levels can start to have an impact on performance. The heavy write ratio of OLAP workloads can also reduce the benefits of compression for Oracle database. File data can compress well so selective volume compression should be considered

Since both the Dell EMC Unity system and Oracle offer data compression, there are several factors to consider. There is no single recommendation since the best choice will depend on several factors such as the contents of the database, the amount of available CPU on both the storage and the database servers, and the amount of I/O resources.

The following lists the benefits of using Dell EMC Unity compression over the database-level compression:

- Dell EMC Unity compression offloads CPU resources associated with compression, allowing more CPU resources available to the OS and databases.
- Dell EMC Unity compression is completely transparent to the databases. Any versions of the database can benefit from it.
- The cost to enable compression for all applications on a Dell EMC Unity system can be lower compared to the cost to enable compression for a database.
- Dell EMC guarantees 4:1 storage efficiency for all-flash configurations. For more information, visit https://www.dellemc.com/en-us/microsites/guarantee-programs/all-flash-storage-efficiency.htm
- Oracle and Grid Linux user home directories are candidates for compression, but evaluate the benefits of compressing them.

8.2 Advanced deduplication

In addition to data reduction, advanced deduplication can be enabled if data reduction is enabled. It provides the ability to reduce the amount of storage needed for user data by keeping only a small number of copies (often just one copy) of a Dell EMC Unity data block with a given content. The deduplication scope is a single LUN, so keep this in mind when choosing the storage layout. Fewer LUNs will result in better deduplication, but more LUNs are needed for maximum performance.

This level of space saving can provide the greatest level of return in most environments, but also requires the most CPU in Dell EMC Unity storage. Because of the nature of user data, there will be duplicate data from

copies of user data. This feature should be tested with a sample of database data and workload before being enabled in production.

Advanced deduplication was first introduced in OE 4.5. It was an optional addition to the data reduction logic available with certain models and it could only be performed on Dell EMC Unity blocks that were compressed. With OE 5.0, advanced deduplication (if enabled) will deduplicate any block (compressed or uncompressed). For more information, see the *Dell EMC Unity: Data Reduction* and *Dell EMC Unity: Best Practices Guide.*

8.3 Snapshots

Snapshots provide a fast and space-efficient way to protect Oracle databases. When using snapshots with Oracle databases, there are important considerations to ensure a successful database recovery.

- All LUNs of an Oracle database must be protected as a set using the consistency group feature. The consistency group will ensure that the snapshot is taken at the exact same time on all LUNs in that group. For NFS file systems that support an Oracle database, to ensure database consistency, the entire database must exist on the file system being snapped.
- Snapshots do not replace Oracle RMAN for regular database backup. However, it offers additional protection to the database and allows offloading RMAN processing to an alternate host.
- Snapshots can be taken on demand manually or automatically based on a schedule defined on the LUN file system. It is recommended to put the database in hot backup mode before taking a snapshot and end backup mode after the snapshot is taken.

Note: Snapshots increase the overall CPU load on the system, and increase the overall drive IOPS in the storage pool. Snapshots also use pool capacity to store the older data being tracked by the snapshot, which increases the amount of capacity used in the pool until the snapshot is deleted. Consider the overhead of snapshots when planning both performance and capacity requirements for the storage pool.

- Before enabling snapshots on a storage object, it is recommended to monitor the system and ensure that existing resources can meet the additional workload requirements. (Refer to the “Hardware Capability Guidelines” section and Table 2 in the *Dell EMC Unity: Best Practices Guide*.)
- Enable snapshots on a few storage objects at a time, and then monitor the system to be sure it is still within the recommended operating ranges before enabling more snapshots. Additional information can be found in the *Dell EMC Unity: Snapshots and Thin Clones* document.

When recovering the database from a snapshot, the Dell EMC Unity system offers two methods to recover a point-in-time copy of the database: restore, and attach to host.

8.3.1 Restoring from a snapshot

With the restore method, the data of the original LUNs are replaced in place on the original server where the snapshot was taken. This is the simplest and fastest recovery method because there is no copying data, and no configuration or modification is required. The overall process to restore a snapshot is described as follows.

1. Terminate all user connections and shut down the database to be restored.
2. In Dell EMC Unisphere, identify and select the snapshot in the **LUN Snapshots** properties page or in the **Consistency Group Snapshots** properties page. See Figure 38.
3. Choose **Restore** from the **More Actions** drop-down menu.
4. After the restore operation is completed, restart the database on the host.
5. Oracle automatically performs database recovery during the startup.
6. Verify the data in the database.

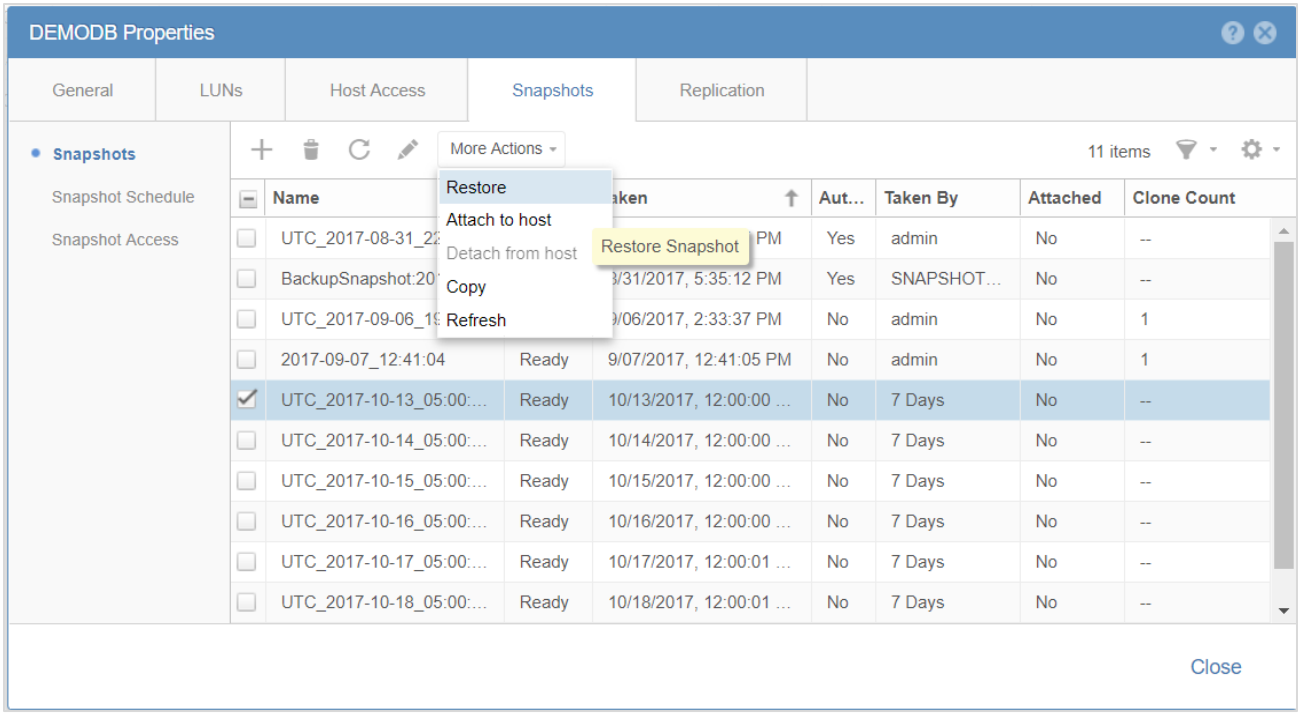

Figure 38 Snapshot properties page

8.3.2 Recovering from a snapshot

When only a subset of the data needs to be restored, a point-in-time copy of the database can be mounted on an alternate host. Administrators can then export and import the data back to the original database. The original database can remain running but access to the corrupted data should be restricted. To the operating system, the LUNs presented by the snapshot will have different WWNs than the original LUNs. However, the ASM headers are exactly the same as the original ASM devices. Therefore, it is not recommended to attach these LUNs on the same host as the original database. This will create confusion to Oracle and increase the risk of writing to the wrong LUNs accidentally. The overall recovery process involves the following:

1. Prepare the destination host to receive the LUNs.
   a. Configure the operating system similarly as the original host.
   b. Install the same version of Oracle software on the destination host.
   c. Install the same storage software (ASMLib, ASMFD) on the destination host.
2. In Unisphere, identify and select the snapshot in the **LUN Snapshots** properties page or in the **Consistency Group Snapshots** properties page. See Figure 38.
3. Choose **Attach to host** from the **More Actions** drop-down menu.
4. Select the destination host and allow Read/Write access.
5. After the snapshot is attached to the host, scan for the LUNs using **rescan-scsi-bus.sh -forcerescan** or **-a.**
6. Set proper ownership and permission on the LUNs. Since this is for temporary use, it is possible to set ownership and permission with **chown** and **chmod**. Just remember that if the host is rebooted, they will need to be set up again.

7. Scan for ASM devices.

   For ASMFD:

   ```
   # asmcmd afd_scan
   # asmcmd afd_lsdsk
   ```

   For ASMLIb;

   ```
   # oracleasm scandisks
   # oracleasm listdisks
   ```

8. Mount ASM disk groups.
9. Copy the database init parameter file to the destination host.
10. Create the database log directories on the destination host.
11. Start up the database in sqlplus.
12. Restore data using one of the following methods. Data is extracted from the destination database and imported into the original database.

    a. RMAN
    b. Datapump
    c. Copy data using the database link

    Find more information in the Oracle Backup and Recovery User's Guide.

13. Once the recovery is complete, shut down the database copy.
14. Dismount the ASM disk groups.
15. Remove the snapshot LUNs from the destination host.
16. Remove host access of the snapshot LUNs in Unisphere.

8.4 Thin clones

Thin clones are based on snapshot technology and are the preferred way to make read-write copies of databases. Similar to regular LUNs, many of the data services, such as snapshots, replications, and host I/O limit, are also available to thin clones. When thin clones are first created, they consume no storage because they share the same blocks as their parent snapshot at the beginning. As new data is written or changes are made to the existing data, new data blocks are allocated and tracked separately from the parent. The data on the thin clones are exactly the same as the parent LUNs but they have different LUN IDs and WWNs. To the operating system, they appear to be different LUNs. However, when Oracle scans for the ASM headers, they contain the same labels and disk group information as the original LUNs. Therefore, it is recommended to attach thin clones on an alternate host to avoid confusion in Oracle and risks overwriting data on the wrong LUNs.

Thin clones are suitable for various uses:

- Create full-size development and test environments from production
- Test new code, patches, or data changes in a production replica
- Offload backup and restore processing

Find additional information on thin clones can be found in the *Dell EMC Unity: Snapshots and Thin Clones* document.

8.4.1 Creating a database copy with thin clones

The following shows the overall process to create a database copy using a thin clone.

1. Prepare the destination host to receive the LUNs.
   a. Configure the operating system similarly as the original host.
   b. Install the same version of Oracle software on the destination host.
   c. Install the same storage software (ASMLib, ASMFD) on the destination host.
2. In Unisphere, identify and select the LUN or the consistency group.
3. Choose **Clone** from the **More Actions** drop-down menu.
4. Select a snapshot to clone from.

   Note: Only snapshots with no auto-delete policy and no expiration time are eligible for selection. Remove the auto-delete policy and expiration time on the snapshot before attempting the Clone action.

5. Follow the wizard to configure the thin clone name, host I/O Limit, host access, snapshot policy, and replication.
6. After the thin clone LUNs are attached to the destination host, scan for the LUNs using `rescan-scsi-bus.sh -forcerescan` or `-a`.
7. If the database clone is intended for long-term use, configure multipath and persistent ownership and permission on the thin clone LUNs.
8. Scan for ASM devices.
9. Mount ASM disk groups.
10. Copy the database init parameter file to the destination host.
11. Create the database log directories on the destination host.
12. Start up the database in sqlplus.

8.4.2 Refresh thin clones

It is possible to refresh thin clones from the same snapshot or a different snapshot. This provides an easy way to reset an environment consistently to a baseline or switch to a different point-in-time copy to examine the data. The refresh process is quick and simple as only the pointers are updated and no data is being copied. The process to refresh a thin clone is as follows:

1. Shut down the database copy using the thin clone LUNs.
2. Dismount the ASM disk groups
3. In Unisphere, select the thin clone and select the **Refresh** action in the **More Actions** menu. See Figure 39.
4. A snapshot is automatically created of the thin clone to preserve the thin clone data.
5. Select a snapshot to refresh from.

   Note: Snapshots that have auto-delete policy or expiration time set are not eligible for selection. Remove the auto-delete policy and expiration time on the snapshot first before starting the Refresh action.

6. After the refresh is completed, perform a rescan on the OS.
7. Perform a rescan on ASM devices.
8. Mount the ASM disk group.

9. Start up the database.
10. Verify the data in the database.

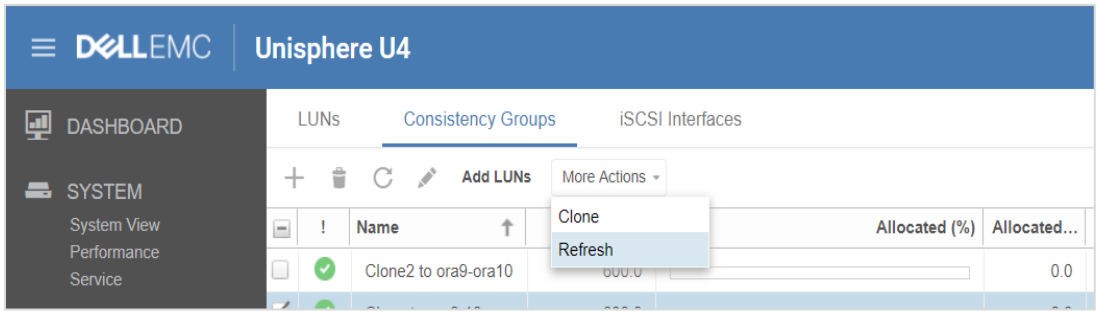

Figure 39 Refresh action in Unisphere

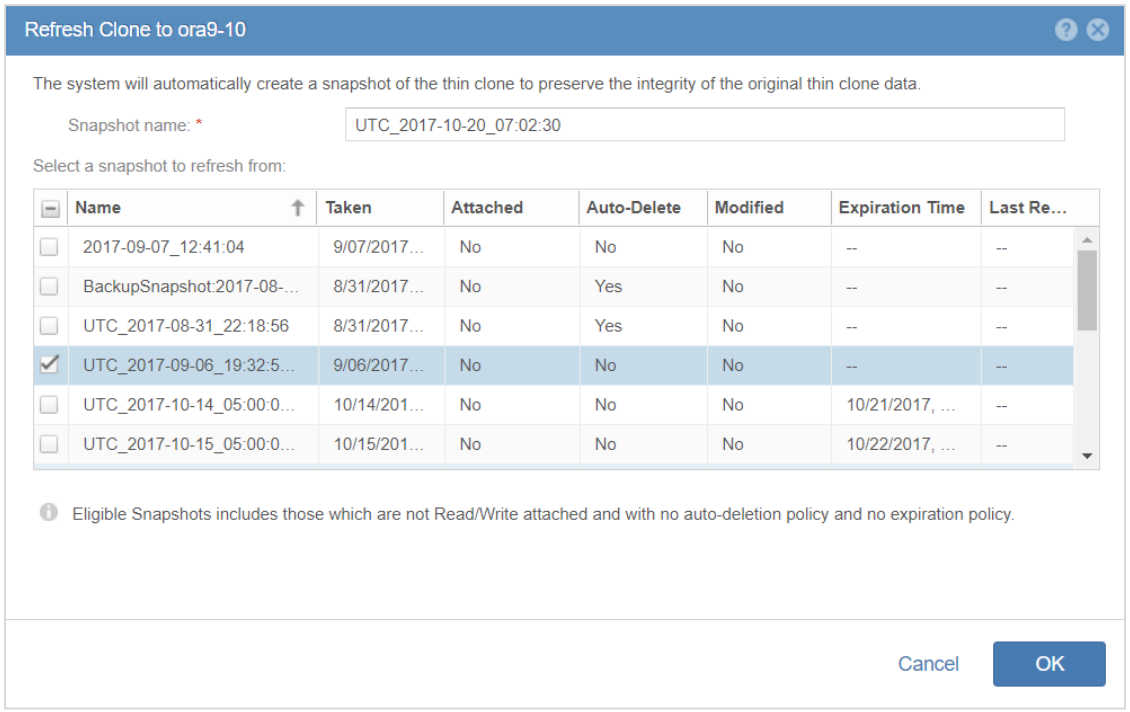

Figure 40 Refreshing a snapshot

8.5 Replication

Creating a high availability solution for Oracle databases often involves creating a copy of the data on another storage device and synchronizing that data in some manner. Dell EMC Unity replication provides data synchronization between Dell EMC Unity systems. Data is replicated at the consistency group or at the LUN and file system allowing a choice of replication settings on a per-volume basis. Using Dell EMC Unity replication can be an effective way to protect Oracle databases due to the flexibility and configuration options that it provides. The variety of options provide a robust way to develop a replication scheme that provides the proper mix of performance and bandwidth efficiency while still meeting RTO and RPO requirements.

When using Dell EMC Unity replication to protect Oracle databases that are located on multiple volumes, contain all ASM devices for a database within a consistency group. Replication is then configured on the consistency group.

Dell EMC Unity storage supports both asynchronous and synchronous replication. A flash tier is recommended (in a hybrid pool) where replication will be active. This is applicable to both the source and the destination pools.

8.5.1 Asynchronous replication

Asynchronous replication takes snapshots on the replicated storage objects in order to create the point-in-time copy, determining the changed data to transfer and maintain consistency during the transfer. Consider the overhead of snapshots when planning performance and capacity requirements for a storage pool that will have replication objects.

Setting smaller RPO values on replication sessions will not make them transfer data more quickly, but will result in more snapshot operations. Choosing larger RPOs, or manually synchronizing during non-production hours, may provide more predictable levels of performance. Additional information can be found in the *Dell EMC Unity: Replication Technologies* and *Dell EMC Unity: Configuring Replication* documents.

8.5.2 Synchronous replication

Synchronous replication transfers data to the remote system over the first Fibre Channel port on each SP. When planning to use synchronous replication, it may be appropriate to reduce the number of host connections on this port. When the CNA ports are configured as FC, CNA port 4 is defined as the synchronous replication port. If the CNA ports are configured as 10GbE, port 0 of the lowest numbered FC I/O module is the replication port. Additional information can be found in the *Dell EMC Unity: Replication Technologies* and *Dell EMC Unity: Configuring Replication* documents.

9 Data protection

In addition to the snapshots and replication provided by Dell EMC Unity systems, Dell EMC offers additional data protection software that integrates with the Dell EMC Unity data protection features. The software is optional and can be used to enhance the overall application protection.

9.1 AppSync

Dell EMC AppSync™ is a software that enables integrated Copy Data Management (iCDM) with the Dell EMC primary storage systems, including Dell EMC Unity arrays. It supports many applications, including Oracle, and storage replication technologies. For the latest support information, refer to the AppSync Support Matrix at the Dell EMC E-lab Navigator.

AppSync simplifies and automates the process of creating and using snapshots of production data. By abstracting the underlying storage and replication technologies, and through application integration, AppSync empowers application owners to manage data copy needs themselves. The storage administrator, in turn, need only be concerned with initial setup and policy management, resulting in a more agile environment.

Additional information on AppSync can be found in the *AppSync User and Administration Guide* and the *AppSync Performance and Scalability Guidelines*.

9.2 RecoverPoint virtual edition

Dell EMC RecoverPoint™ virtual edition provides continuous data protection with multiple recovery points to restore applications instantly to a specific point in time. RecoverPoint virtual edition consists of RecoverPoint Appliance (RPA) software deployed as a virtual appliance in an existing VMware® ESXi® VM environment. RecoverPoint virtual edition is a flexible deployment option which offers maximum simplicity with no dependency on a physical appliance, able to lower TCO.

A File system mount options

The following table describes the file system mount options used in this paper.

Mount option	Description
rw	Mounts the file system for both reading and writing operations
bg	Defines a background mount to occur if a timeout or failure occurs. Bg causes the mount command to fork a child which continues to attempt to mount the export and the parent process immediately returns with a zero status
hard	Explicitly marks the volume as hard-mounted and determines the recovery behavior of the NFS client after an NFS request times out. This is enabled by default and prevents NFS from returning short write errors by retrying the request indefinitely. Short writes cause the database to crash; otherwise they will continue retrying at timeo=<nn> intervals. The server will report a message to the console when a major timeout occurs and will continue to attempt the operation indefinitely.
nointr	Without this option, signals like kill -9 which can be used to interrupt an NFS call will cause data corruption in datafiles because the in-flight writes will be abruptly terminated.
rsize	Specifies the maximum size (bytes) used by NFS clients on read requests, that the NFS client can receive when reading data from a file on an NFS server. The default depends on the version of kernel, but is generally 1,024 bytes. The actual data payload size of each NFS read request is equal to or smaller than the rsize setting, with a maximum payload size of 1,048,576. Values lower than 1,024 are replaced with 4,096, and values larger than 1,048,576 are replaced with 1,048,576. If the specified value is within the supported range but not a multiple of 1,024, it is rounded down to the nearest multiple of 1,024. If a value is not specified, or if the value is larger than the supported maximum on either the client or server, the server and client negotiate the largest rsize they can both support. The rsize specified on the mount appears in /etc/mtab. However, the effective rsize negotiated by the server and client appears in /proc/mounts. With respect to Oracle, the value must be set to equal to or a larger multiple of the Oracle block size (init: db_block_size, default 8k) to prevent fractured blocks in Oracle. rsize must be set to at least 16,348. However, Oracle recommends setting the value to 32,768.
wsize	Identical to rsize, but for write requests sent from the NFS client. wsize must be set to at least 16,348. However, Oracle recommends setting the value to 32,768. Oracle dNFS clients issue writes at wtmax granularity to the NFS filer. If the dNFS client is used and the NFS server does not support a write size (wtmax) of 32,768 or larger, NFS will revert back to the native kernel NFS path.
tcp	Defines the transport protocol name and family the NFS client uses to transmit requests to the NFS server and also controls how the mount command communicates with the server's rpcbind and mountd services. If an NFS server has both and IPv4 and an IPv6 address, using a specific netid will force the user of IPv4 or IPv6 networking to communicate with the server. Specifying tcp forces all traffic from the mount command and the NFS client to use TCP. The tcp option is an alternative to specifying proto=tcp. DO NOT use UDP NFS for ANY REASON
vers	Specifies the NFS protocol version number used to contact the server's NFS service. Use either a value of 3 or 4. Option vers is an alternative to option nfsvers and is provided for compatibility with other OSs.

Mount option	Description
timeo	Defines the time (in tenths of a second) that an NFS client will wait for a request to complete before it retires the request. With NFS over TCP, the default value is 60 seconds; otherwise the default value is 0.7 seconds. If a timeout occurs, the behavior will depend on whether hard or soft was used to mount the file system.
actimeo	This option is required whenever the possibility exists to AUTOEXTEND. It ensures the behavior of AUTOEXTEND is propagated to all nodes in a cluster by disabling all NFS attribute caching (actimeo sets the values of acregmin, acregmax, acdirmin, and acdirmax to the same value). Without this option, NFS will cache the old filesize, causing inappropriate behavior. Currently, Oracle is dependent on file system messaging to advertise a change in size of a datafile; therefore this setting is necessary.
noac	Prevents NFS clients from caching file attributes so that applications can more quickly detect file changes on the NFS server.

B Dell EMC Unity x80F specifications

The following table lists specifications of Dell EMC Unity x80F All-Flash arrays.

Table 13 Dell EMC Unity x80F All-Flash array specifications

	Dell EMC Unity 380F	**Dell EMC Unity 480F**	**Dell EMC Unity 680F**	**Dell EMC Unity 880F**
CPU per SP	1x 6C @ 1.7G Broadwell	2x 8C @ 1.8G (4108) Skylake	2x 12C @ 2.1G (4116) Skylake	2x 16C @ 2.1G (6130) Skylake
Memory per SP	64GB (4x 16GB)	96GB (12x 8GB)	192GB (12x 16GB)	384GB (12x 32GB)
Min/Max drives	5 / 500	5 / 750**	5 / 1000**	5 / 1500*
Embedded SAS ports per SP	2x 4 lane 12Gb/s SAS	2x 4 lane 12Gb/s SAS	2x 4 lane 12Gb/s SAS	2x 4 lane 12Gb/s SAS
Optional SAS ports per SP	N/A	4x 4 lane or 2x 8 lane 12Gb/s SAS I/O Module	4x 4 lane or 2x 8 lane 12Gb/s SAS I/O Module	4x 4 lane or 2x 8 lane 12Gb/s SAS I/O Module
Embedded 10GbE Base T ports per SP	2 Ports	N/A	N/A	N/A
Embedded CNA ports per SP	2 Ports: 8/16Gb FC, 10GbE Opt, 1GbE BaseT, or Empty	N/A	N/A	N/A
Support 4-port mezz cards per SP	N/A	4x 10/25GbE Opt, 4x 10GbE BaseT, or Empty	4x 10/25GbE Opt, 4x 10GbE BaseT, or Empty	4x 10/25GbE Opt, 4x 10GbE BaseT, or Empty
Support I/O modules (2 slots per SP)	4x 10GbE BaseT 4x 16Gb FC 4x 10/25GbE Opt	4x 10GbE BaseT 4x 16Gb FC 4x 10/25GbE Opt 4x 12Gb/s SAS		
Supported DAEs	2.5” 25-Drive, 3.5” 15-Drive, 2.5” 80-Drive			

*Requires 4-port 12Gb SAS backend I/O module to reach max drive count.

C Technical support and resources

Dell.com/support is focused on meeting customer needs with proven services and support.

Storage technical documents and videos provide expertise that helps to ensure customer success on Dell EMC storage platforms.

The Dell EMC Unity Info Hub provides helpful links to documents and tools.

C.1 Related resources

The following referenced or recommended Dell EMC publications and resources are located at DellEMC.com.

- Dell EMC Unity: Introduction to the Dell EMC Unity Platform - A Detailed Review
- Dell EMC Unity: Best Practices Guide
- Dell EMC Unity: Dynamic Pools - Overview
- Dell EMC Unity Family Configuring Pools
- Dell EMC Unity: Compression - Overview
- Dell EMC Unity: Data Reduction – Overview
- Dell EMC Unity: Data Reduction Analysis
- Dell EMC Unity Family Guide to Configuring LUNs
- Dell EMC Unity: Performance Metrics - A Detailed Review
- Dell EMC Unity Family Monitoring System Performance
- Dell EMC Unity: Replication Technologies - A Detailed Review
- Dell EMC Unity Family: Configuring Replication
- Dell EMC Unity: Snapshots and Thin Clones - A Detailed Review
- Dell EMC Unity: Unisphere Overview - Simplified Storage Management
- Dell EMC Unity: Data at Rest Encryption - A Detailed Review
- Dell EMC Unity Drive Support Matrix
- Dell EMC Host Connectivity Guide for Linux
- Dell EMC Unity Family Configuring High Availability Guide
- Dell EMC Unity: NAS Capabilities - A Detailed Review
- Dell EMC Unity Family Guide to Configuring NFS File Sharing
- Unity Family Configuring Hosts to Access NFS
- Dell EMC Unity Family: Configuring Multiprotocol File Sharing
- Dell EMC Unity Series Configuring Hosts to Access Fibre Channel (FC) or iSCSI Storage
- Dell EMC Unity Family Unisphere Management REST API Programmer's Guide
- Dell EMC Unity Unisphere CLI User Guide
- Dell EMC Unity Family Service Commands
- Dell EMC Unity: High Availability - A Detailed Review
- AppSync User and Administration Guide
- AppSync Performance and Scalability Guidelines
- PowerPath Installation and Administration Guide
- PowerPath Load Balancing and Failover
- CloudIQ Overview

The following referenced or recommended Veritas resources are located at Veritas Online Support:

- Storage Foundation Administrator's Guide
- Storage Foundation Tuning Guide

The following referenced or recommended Oracle resources are located at the Oracle Online Documentation Portal:

- Oracle Automatic Storage Management Administrator's Guide
- Oracle Database Administration Documentation Library
- Oracle Performance Guide

The following referenced or recommended Oracle notes are located at My Oracle Support (Oracle support license required):

- Mount Options for Oracle files for RAC databases and Clusterware when used with NFS on NAS devices (Doc ID 359515.1)
- Creating File Devices On NAS/NFS FileSystems For ASM Diskgroups. (Doc ID 1620238.1)
- Direct NFS: FAQ (Doc ID 954425.1)
- How to configure DNFS to use multiple IPs (Doc ID 1552831.1)
- How to Setup Direct NFS Client Multipaths in Same Subnet (Doc ID 822481.1)
- How to configure DNFS to use multiple IPs using different subnets (Doc ID 1528148.1)
- Best Practices: How to configure DNFS client using Single and Multiple Subnets (Doc ID 2246252.1)
- Step by Step - Configure Direct NFS Client (dNFS) on Linux (Doc ID 762374.1)
- How To Setup dNFS (Direct NFS) On Oracle Release 11.2 (Doc ID 1452614.1)

Referenced or recommended I/O benchmark resources:

- vdbench download and documentation
- FIO download
- FIO output explained
- atop and netatop homepage
- collectl sourceforge homepage
- SLOB Blog
- HammerDB homepage
- Swingbench homepage