Best Practices

# Dell EMC SC Series: Oracle Best Practices

## Abstract

Best practices, configuration options, and sizing guidelines for Dell EMC™ SC Series storage in Fibre Channel environments when deploying Oracle®.

May 2020

# Revisions

| Date | Description |
|---|---|
| January 2012 | Initial release |
| April 2012 | Content and format change |
| July 2013 | Added ORION information provided by Copilot |
| April 2015 | Content and format change |
| July 2017 | Major rewrite. Made document agnostic with respect to Dell EMC SC series arrays. Format changes. |
| August 2018 | Content changes for preallocation |
| May 2020 | Changed setting of port down |

# Acknowledgments

Author: Mark Tomczik

The information in this publication is provided “as is.” Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

# Table of contents

# Executive summary

Managing and monitoring database storage, capacity planning, and data classification for storage are some of the daily activities and challenges of database administrators (DBAs). These activities also have impacts on the database environment, performance, data tiering, and data archiving. With traditional storage systems, DBAs have limited ability to accomplish these activities in an effective and efficient manner, especially from a storage perspective. Storage and system administrators typically perform all necessary management and monitoring activities of physical storage in the storage system and operating system (OS), while DBAs are typically challenged with provisioning and configuring the storage within the Oracle® environment. This includes determining how and where to use the storage within the database.

Dell EMC™ SC Series arrays address these challenges and provide DBAs more visibility into and management of the storage environment, such that it frees their time to address other pressing production and database infrastructure needs. Along with storage management comes determining best practices, which are not always obvious. The intent of this document is to help readers understand best practices, configuration options and sizing guidelines of SC Series storage in Fibre Channel environments when deploying Oracle to deliver a cost-effective alternative for demanding performance and fault-tolerant storage requirements of Oracle deployments.

# How to use this document

The ideas and best practices in this document are applicable to other software versions, but the commands may have to be modified to fit other OS and software versions. Also, the best practices and recommendations presented in this document should be evaluated and adjusted accordingly for each environment.

This document is not intended to provide a step-by-step configuration or be an exact sizing guide. It also should not be considered an exhaustive or authoritative source on any component discussed. Performance tuning of any component within the document is not in scope of this document. Actual configuration and sizing will vary based on individual business, application, infrastructure and compatibility requirements.

For information on SC Series arrays, see the *Dell Storage Manager Administrator's Guide* available on the Dell Support website. For detailed information on Oracle, see the My Oracle Support or Oracle Help Center sites. Additional resources are listed in appendix B.1.

# Audience

This document is intended for information technology professionals seeking to deploy and configure a cost-effective Oracle database environment using SC Series arrays.

# Prerequisites

Readers should have formal training or advanced working knowledge of the following:

- RAID, Fibre Channel, multipathing, serial-attached SCSI (SAS), and IP networking administration
- Operating and configuring SC Series storage arrays
- Installing, configuring, and administrating Linux®
- Dell EMC server architecture and administration
- General understanding of SAN technologies
- Operating Dell Storage Manager (DSM)
- Oracle architecture
- Installing, configuring, and administering Oracle 11g and 12c; Automated Storage Management (ASM), Real Application Clusters (RAC), or single instance database

# 1 Introduction

When designing the physical layer of a database, DBAs must consider many storage configuration options. In most Oracle deployments, storage configuration should provide redundancies to avoid downtime for such events as component failures, maintenance, and upgrades. It should also provide ease of management, performance, and capacity to meet or exceed business requirements.

SC Series arrays provide a powerful and complete set of features and wizards to create such a configuration. Determining best practices for the configuration will often be subject to subtle nuances based on business and infrastructure requirements and standards.

## 1.1 Manageability

SC Series arrays can be configured and easily managed to adapt to ever-changing business and database requirements. Storage administrators can easily manage volumes, move data between storage tiers, and add disks to the array to increase performance and capacity without sustaining an application outage. By automating storage-related tasks, SC Series storage reduces the risk of human error and allows storage administrators and DBAs to concentrate on business-critical administration functions and services.

## 1.2 High availability

Storage configurations must protect against failure of storage hardware such as disks, power, controller, host bus adapters (HBAs), and fabric switches. SC Series arrays were designed with high availability and fault tolerance to protect against unplanned outages due to component failure. Controller firmware upgrades, hardware expansions, or hardware replacements can occur without forcing an application outage.

## 1.3 Performance

When talking about storage media, high-performance and higher-cost storage devices like flash-based solid-state drives (SSD) are deployed in many storage solutions available today. With the release of Storage Center OS (SCOS) 6.4 and the SC Series All-Flash array, an entire Oracle deployment can efficiently exist in an all-flash array. Should business requirements require less-expensive high-capacity storage, SC Series arrays can be configured with only mechanical hard disk drives (HDD), or as a hybrid array containing both SSDs and HDDs. Whether SSDs or HDDs are configured in an SC array, additional drives can be added dynamically. Also, a variety of connectivity options exist, including 16Gbps Fibre Channel and 10Gbps iSCSI, to ensure that data move quickly and efficiently between the database server and the SC series array.

SC Series arrays can be configured and sized to have the power and scalability to host an Oracle deployment and deliver the necessary IOPS, TPS, latencies, and throughput. The SC Series all-flash array can be deployed with one or more SSD media types organized in different storage tiers. An array of this type is known as an all-flash optimized array. For example, an all-flash optimized array could have a mixture of single-level cell (SLC) SSDs and multi-level cell (MLC) SSDs:

- Tier 1 (T1) containing 6 x 400GB write-intensive SLC drives; these drives are optimized for write operations and are higher performing as compared to read-intensive MLC drives
- Tier 2 (T2) containing 6 x 1.6TB read-intensive MLCs drives; these drives are best fit for read operations

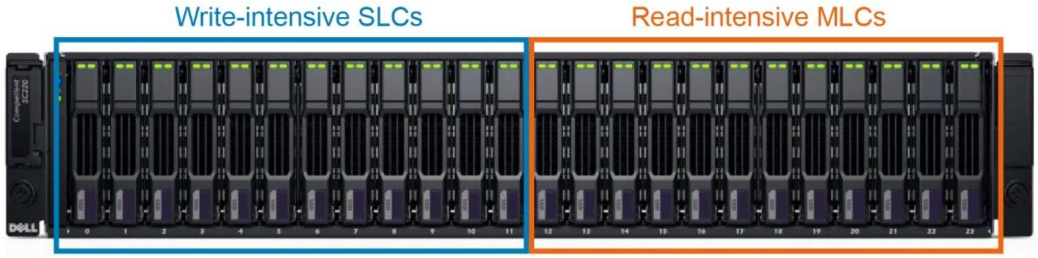


Figure 1 Dell SC220 enclosure with MLC and SLC drives

Read-intensive drives (MLCs) provide greater capacity and lower costs than write-intensive drives (SLCs), but SLCs have greater endurance rates than MLCs. This makes SLCs optimal for workloads with heavy write characteristics, and MLCs optimal for workloads with heavy reads characteristics.

An SC Series hybrid array can be deployed with a combination of SSDs and HDDs, with each media type in its own storage tier.

Table 1 SC Series hybrid array tier configuration examples

| Configuration example | Storage tier | Media type |
|---|---|---|
| 1 | 1 | Triple level cell (TLC) SSD |
|  | 3 | 15K HDD |
| 2 | 1 | Write-intensive SLC |
|  | 2 | Read-intensive MLC |
|  | 3 | 15K HDD |
| 3 | 1 | TLC |
|  | 2 | 15K HDD |
|  | 3 | 7K HDD |

SC Series arrays can also be deployed with only HDDs, with each media type in its own storage tier.

Table 2 SC Series spinning media array tier configuration examples

| Configuration example | Storage tier | Media type |
|---|---|---|
| 1 | 1 | 15K HDD |
|  | 2 | 10K HDD |
|  | 3 | 7K HDD |
| 2 | 1 | 15K HDD |
|  | 2 | 10K HDD |
| 3 | 1 | 10K HDD |
|  | 2 | 7K HDD |

When deploying Oracle on an SC Series array, database, system, storage and network administrators have to configure other components of the infrastructure stack. Some of the components are discussed in the remainder of this document.

# 2 Fibre Channel connectivity

SC Series arrays have been tested to work with Dell, Emulex®, and QLogic® HBA cards. They also support simultaneous transport protocols including Fibre Channel (FC), iSCSI, and FCoE. Although this document was created for FC environments with QLogic HBAs, much of it should also apply to iSCSI and FCoE. For information on other transport protocols supported by SC Series arrays, see the *Dell Storage Manager Administrator's Guide* available on the Dell Support website. The *Dell EMC Storage Compatibility Matrix* provides a list of compatible arrays and switches.

To maximize throughput and HA protection, Dell EMC recommends using at least two 8 Gbps quad-port HBAs, or at least two 16 Gbps dual- or quad-port HBAs in database servers, and two 16 Gbps quad-port FC HBAs in the array. Additional HBAs may be required in the server for greater throughput requirements.

## 2.1 Zoning

Zoning that delivers port-level redundancy and high availability in the event of a SAN component outage is recommended. One way to provide this level of service is to implement the following:

- Multiple fabrics
- Multiple Fibre Channel switches (one or more switches per fabric VSAN and SC fault domain)
- SC Series dual controllers
- Multiple Fibre Channel HBAs per SC Series controller
- SC Series virtual port mode
- Multiple dual- or quad-port 16Gb Fibre Channel server HBAs per server
- Server multipathing
- Soft zones

### 2.1.1 Soft (WWN) zoning

Dell EMC recommends using soft zoning rather than hard zoning. A soft zone is based on the WWN of a single initiator (server HBA port) and multiple front-end virtual ports (targets) from a HBA in the SC Series array controller as illustrated in Table 3.

Table 3 Soft zone example

| FC zone | Fabric | WWN | Description |
|---|---|---|---|
| MT_r730xd_1_s4 | 1 | 2001000e1ed020c6 | Server HBA 1, port 1 |
| | | 5000d3100495464e | Controller 1 front-end virtual port 1 |
| | | 5000d3100495464f | Controller 1 front-end virtual port 2 |
| | | 5000d31004954650 | Controller 2 front-end virtual port 1 |
| | | 5000d31004954651 | Controller 2 front-end virtual port 2 |

Soft zones provide the ability to move the server or the SC Series array to another switch without updating any zone. SC Series virtual port mode requires N_Port ID Virtualization (NPIV) be enabled on the attached FC switch. A soft zone should exist for each server HBA port. Soft zones in each fabric share the same targets, but have a different initiator.

An example of soft zoning with a SC8000 is illustrated in Figure 2.

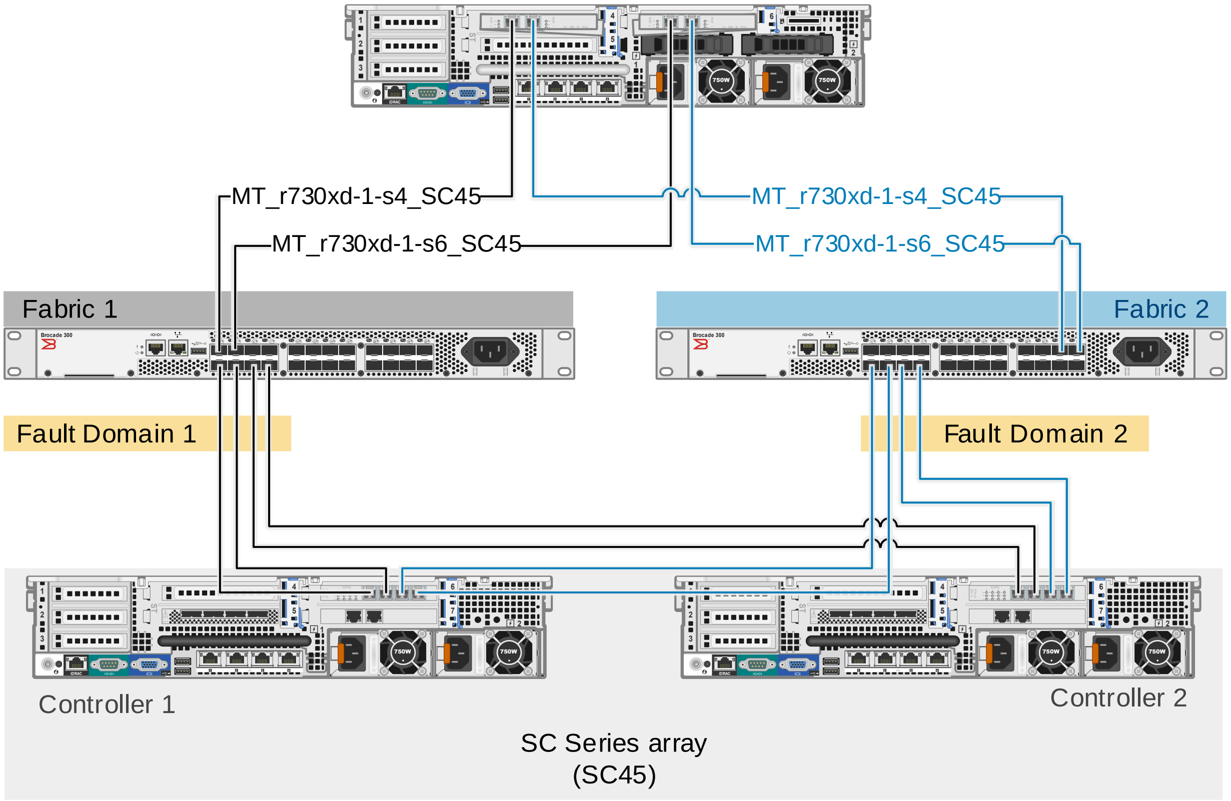
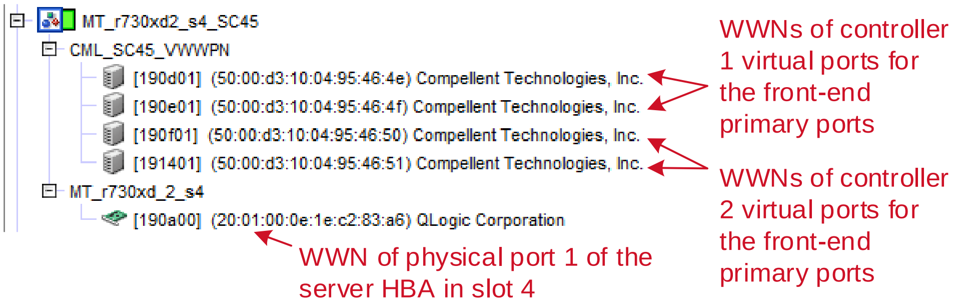


Figure 2 Two FC fabrics and four zones

An example of a soft zone in Brocade® switch Explorer is shown in Figure 3:



```
MT_r730xd2_s4_SC45
  CML_SC45_VWWPN
    [190d01] (50:00:d3:10:04:95:46:4e) Compellent Technologies, Inc.
    [190e01] (50:00:d3:10:04:95:46:4f) Compellent Technologies, Inc.
    [190f01] (50:00:d3:10:04:95:46:50) Compellent Technologies, Inc.
    [191401] (50:00:d3:10:04:95:46:51) Compellent Technologies, Inc.
  MT_r730xd_2_s4
    [190a00] (20:01:00:0e:1e:c2:83:a6) QLogic Corporation
```

WWNs of controller 1 virtual ports for the front-end primary ports

WWNs of controller 2 virtual ports for the front-end primary ports

WWN of physical port 1 of the server HBA in slot 4

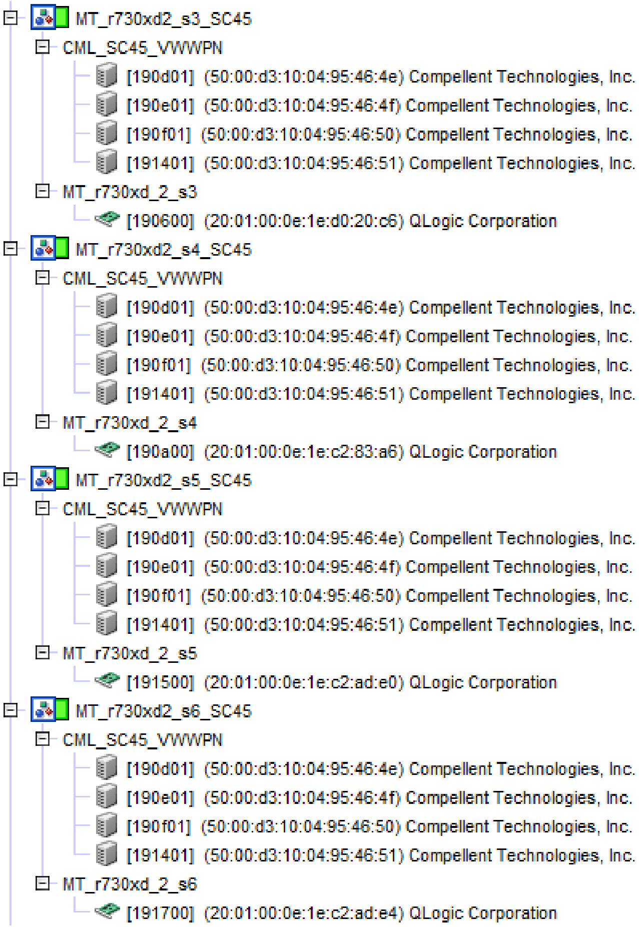
Figure 3 Soft zone

When soft zoning with multiple controllers in virtual port mode, the ports from each controller must be equally distributed between the two fabrics. It does not matter which controller ports are chosen for a fabric as long as the ports are chosen in the same manner for all zones in the same fabric (Table 4). Adhering to a cabling and zoning standard will create an easily maintainable fabric.

When soft zoning with multiple controllers in virtual port mode, create the following zones in both fabrics:

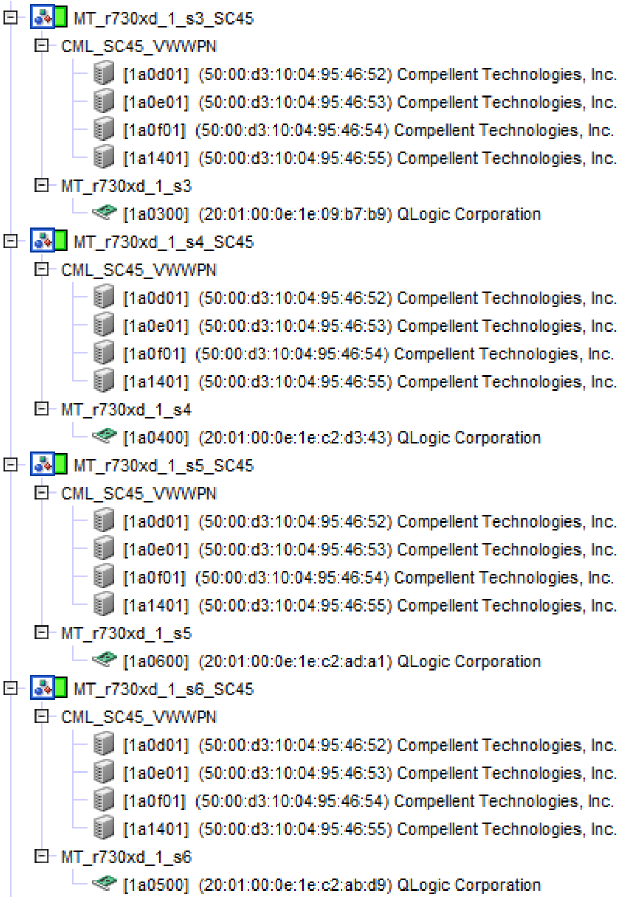
- A zone that includes half the physical ports from both controllers in one fabric, and a zone that includes the remaining ports from both controllers in the other fabric. For example: one zone could have ports 1 and 2 from both controllers, the other zone could have ports 3 and 4 from both controllers.
- A zone that includes the virtual ports from half the physical ports from both controllers in one fabric, and a zone that includes the remaining virtual ports from the other half of the physical ports from both controllers in the other fabric. Virtual ports need to be equally divided between the zones based on their physical ports,
- A zone for each server HBA port that will connect to the SC Series array. The zone must include the WWNs of the server HBA port and the multiple SC Series virtual ports that are on the same FC switch.

Figure 4 and Figure 5 illustrate dual fabrics with four dual-port server HBAs and dual quad-port controllers.



```
MT_r730xd2_s3_SC45
  CML_SC45_VWWPN
    [190d01]  (50:00:d3:10:04:95:46:4e) Compellent Technologies, Inc.
    [190e01]  (50:00:d3:10:04:95:46:4f) Compellent Technologies, Inc.
    [190f01]  (50:00:d3:10:04:95:46:50) Compellent Technologies, Inc.
    [191401]  (50:00:d3:10:04:95:46:51) Compellent Technologies, Inc.
  MT_r730xd_2_s3
    [190600]  (20:01:00:0e:1e:d0:20:c6) QLogic Corporation
MT_r730xd2_s4_SC45
  CML_SC45_VWWPN
    [190d01]  (50:00:d3:10:04:95:46:4e) Compellent Technologies, Inc.
    [190e01]  (50:00:d3:10:04:95:46:4f) Compellent Technologies, Inc.
    [190f01]  (50:00:d3:10:04:95:46:50) Compellent Technologies, Inc.
    [191401]  (50:00:d3:10:04:95:46:51) Compellent Technologies, Inc.
  MT_r730xd_2_s4
    [190a00]  (20:01:00:0e:1e:c2:83:a6) QLogic Corporation
MT_r730xd2_s5_SC45
  CML_SC45_VWWPN
    [190d01]  (50:00:d3:10:04:95:46:4e) Compellent Technologies, Inc.
    [190e01]  (50:00:d3:10:04:95:46:4f) Compellent Technologies, Inc.
    [190f01]  (50:00:d3:10:04:95:46:50) Compellent Technologies, Inc.
    [191401]  (50:00:d3:10:04:95:46:51) Compellent Technologies, Inc.
  MT_r730xd_2_s5
    [191500]  (20:01:00:0e:1e:c2:ad:e0) QLogic Corporation
MT_r730xd2_s6_SC45
  CML_SC45_VWWPN
    [190d01]  (50:00:d3:10:04:95:46:4e) Compellent Technologies, Inc.
    [190e01]  (50:00:d3:10:04:95:46:4f) Compellent Technologies, Inc.
    [190f01]  (50:00:d3:10:04:95:46:50) Compellent Technologies, Inc.
    [191401]  (50:00:d3:10:04:95:46:51) Compellent Technologies, Inc.
  MT_r730xd_2_s6
    [191700]  (20:01:00:0e:1e:c2:ad:e4) QLogic Corporation
```

Figure 4 Four zones (one for port 1 from each HBA) in fabric 1



```
MT_r730xd_1_s3_SC45
  CML_SC45_VWWPN
    [1a0d01]  (50:00:d3:10:04:95:46:52) Compellent Technologies, Inc.
    [1a0e01]  (50:00:d3:10:04:95:46:53) Compellent Technologies, Inc.
    [1a0f01]  (50:00:d3:10:04:95:46:54) Compellent Technologies, Inc.
    [1a1401]  (50:00:d3:10:04:95:46:55) Compellent Technologies, Inc.
  MT_r730xd_1_s3
    [1a0300]  (20:01:00:0e:1e:09:b7:b9) QLogic Corporation
MT_r730xd_1_s4_SC45
  CML_SC45_VWWPN
    [1a0d01]  (50:00:d3:10:04:95:46:52) Compellent Technologies, Inc.
    [1a0e01]  (50:00:d3:10:04:95:46:53) Compellent Technologies, Inc.
    [1a0f01]  (50:00:d3:10:04:95:46:54) Compellent Technologies, Inc.
    [1a1401]  (50:00:d3:10:04:95:46:55) Compellent Technologies, Inc.
  MT_r730xd_1_s4
    [1a0400]  (20:01:00:0e:1e:c2:d3:43) QLogic Corporation
MT_r730xd_1_s5_SC45
  CML_SC45_VWWPN
    [1a0d01]  (50:00:d3:10:04:95:46:52) Compellent Technologies, Inc.
    [1a0e01]  (50:00:d3:10:04:95:46:53) Compellent Technologies, Inc.
    [1a0f01]  (50:00:d3:10:04:95:46:54) Compellent Technologies, Inc.
    [1a1401]  (50:00:d3:10:04:95:46:55) Compellent Technologies, Inc.
  MT_r730xd_1_s5
    [1a0600]  (20:01:00:0e:1e:c2:ad:a1) QLogic Corporation
MT_r730xd_1_s6_SC45
  CML_SC45_VWWPN
    [1a0d01]  (50:00:d3:10:04:95:46:52) Compellent Technologies, Inc.
    [1a0e01]  (50:00:d3:10:04:95:46:53) Compellent Technologies, Inc.
    [1a0f01]  (50:00:d3:10:04:95:46:54) Compellent Technologies, Inc.
    [1a1401]  (50:00:d3:10:04:95:46:55) Compellent Technologies, Inc.
  MT_r730xd_1_s6
    [1a0500]  (20:01:00:0e:1e:c2:ab:d9) QLogic Corporation
```

Figure 5 Four zones (one for port 2 from each HBA) in fabric 2

Table 4 Zones: Two dual-port server HBAs and two SC Series controllers with quad front-end ports

| FC Zone | Fabric | Server HBA | Server HBA port | SC Series controller | SC Series controller ports |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1, 2 or 1, 3 |
| 1 | 1 | 1 | 1 | 2 | 1, 2 or 1, 3 |
| 2 | 1 | 2 | 1 | 1 | 1, 2 or 1, 3 |
| 2 | 1 | 2 | 1 | 2 | 1, 2 or 1, 3 |
| 3 | 2 | 1 | 2 | 1 | 3, 4 or 2, 4 |
| 3 | 2 | 1 | 2 | 2 | 3, 4 or 2, 4 |
| 4 | 2 | 2 | 2 | 1 | 3, 4 or 2, 4 |
| 4 | 2 | 2 | 2 | 2 | 3, 4 or 2, 4 |

### 2.1.2 Hard (port) zoning

Hard zoning is based on defining specific ports in the zone. Because the zone is based on ports, if the server or SC Series array is moved to a different port or switch, the fabric will require an update. This can cause issues with the manageability or supportability of the fabric. Therefore, Dell EMC does not recommend hard zoning with SC Series arrays.

## 2.2 QLogic settings

Configure the server HBAs according to the recommendations in the *Dell Storage Manager Administrator's Guide* available on the Dell Support website to improve connection speeds between the database server and SC Series arrays.

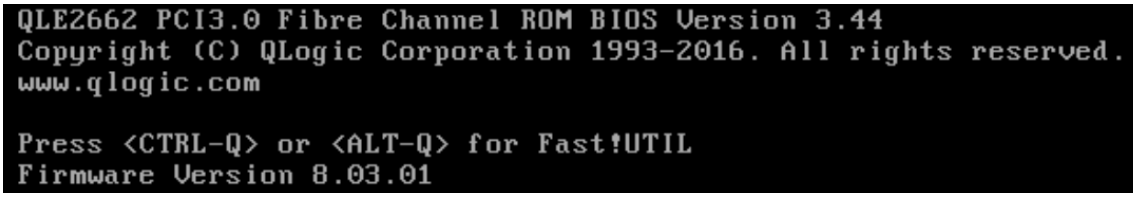
**Note:** Both the BIOS settings and OS driver module for QLogic control the performance of an HBA. Settings specified in the driver module take precedence over the BIOS settings.

For new systems, be sure to review the Dell EMC Storage Compatibility Matrix for a supported QLogic adapter model, driver, boot code version, and firmware version. Then, update the HBA firmware, boot code, and driver to the applicable versions before doing any configuration. For existing systems, Dell EMC recommends verifying the existing configuration against the compatibility matrix for supportability. Then, verify the functionality of the latest firmware, driver, and QLogic settings in a test environment before promoting the changes to a production environment.

### 2.2.1 Server FC HBA BIOS settings

Using QLogic Fast!UTIL, configure the HBA BIOS during the power on system test (POST) of the server.

1. During the POST, press **[Ctrl**] + **[Q]** process to start Fast!UTIL.



```
QLE2662 PCI3.0 Fibre Channel ROM BIOS Version 3.44
Copyright (C) QLogic Corporation 1993-2016. All rights reserved.
www.qlogic.com

Press <CTRL-Q> or <ALT-Q> for Fast!UTIL
Firmware Version 8.03.01
```

2. Select an HBA in Fast!UTIL. Repeat steps 2–5 for all HBAs (or initiators) that are zoned to an SC Series array and are displayed by Fast!UTIL.
3. Reset the adapter to factory defaults. Select **Configuration Settings** > **Restore Default Settings**. Press **[ESC]** multiple times to display the **Fast!UTIL Options** menu.
4. Select **Configuration Settings** > **Adapter Settings** and make the changes shown in Table 5. Press **[ESC]** multiple times to display the main Fast!UTIL menu.
5. Select **Scan Fibre Devices**. Press **[ESC]** to display the main Fast!UTIL menu.
6. Select **Configuration Settings** > **Advanced Adapter Settings** and make the changes shown in Table 5.

Table 5 QLogic HBA BIOS Settings

| QLogic BIOS menu | QLogic BIOS attribute | Value |
|---|---|---|
| Adapter Settings | Host Adapter BIOS | Enable |
| | Connection Options | QLe25xx and earlier: 1 (for point-to-point only)<br>QLe26xx and later: Default |
| Advanced Adapter Settings | Login retry count | 60 |
| | Port down retry count | 60 |
| | Link down timeout | 30 |
| | Execution Throttle | 256 |
| | LUNs per Target | 128 |
| | Enable LIP Reset | Yes |
| Selectable Boot Settings (Each HBA port has two paths to the boot volume. The WWN for each path should be selected, except when installing and configuring Dell EMC PowerPath™. Then, only server initiator port from one HBA should be enabled for boot.)[2] | Selectable Boot [1] | Enable |
| | Boot Port Name, Lun (nnnnn,0) (This was configured after the FC fabric zones were created) [2] | WWN for the 1st boot volume path, and Lun should correspond to the Lun chosen during the mapping operation in DSM of the volume to the server. For boot from SAN, Dell EMC Linux best practices say Lun/LUN needs to be 0. |
| | Boot Port Name, Lun (nnnnn,0) (This was configured after the FC fabric zones were created) [2] | WWN for the 2nd boot volume path, and Lun should correspond to the Lun chosen during the mapping operation in DSM of the volume to the server. For boot from SAN, Dell EMC Linux best practices say Lun needs to be 0. |

[1] If business requirements allow, it is recommended to configure boot from SAN. In such cases, apply the recommended configurations as allowed. If using EMC PowerPath for multipathing, see the *EMC Host Connectivity Guide for Linux* and *PowerPath for Linux Installation and Administration Guide* available on Dell Support for specific instructions for configuring storage connectivity and QLogic selectable boot settings for boot from SAN.

After configuring the QLogic HBA cards, configure the QLogic driver settings.

**Note:** Record the WWNs of all HBAs identified and enabled in QLogic and zoned to the SC array. WWNs are needed when creating a logical server object in Dell Storage Manager (DSM). See section 3.7.

### 2.2.2 Server FC HBA driver settings: timeouts and queue depth

Configure the link down timeout, and if necessary, the queue depth in Linux after backing up the original QLogic adapter configuration file.

The timeout value determines the time a server waits before the server destroys a connection after losing connectivity. The timeout should be set to 60 seconds to provide enough time for the WWN of the failed port to transfer to a port on the other controller (assuming a dual controller SC Series configuration is deployed). Either Dell EMC PowerPath or native Linux multipathing (Device-Mapper Multipathing) is recommended.

The default queue depth value of 32 may be adequate, but a value of 64 or 128 may work well too. The optimal queue depth value is dependent on a number of parameters, including creating snapshots (Replays) of the Oracle database. Determining the optimal value is out of scope of this document.

```
options qla2xxx qlport_down_retry=60
options qla2xxx ql2xmaxqdepth=<value>
```

For additional information on how to configure the QLogic adapter configuration files, see the references in appendix B.

# 3 SC Series array

Storage administrators have to make complex decisions daily on storage configuration, usage, and planning. For example, when creating a volume, the question may be asked: Will it be sized appropriately? If all volumes are oversized in a traditional array, there is the added issue of over provisioning the array. SC Series storage provides solutions to these complex decisions with a robust set of features that provide an easy-to-manage storage solution. The features can be used in an Oracle database environment to assist the DBA with data tiering, database performance, efficient use of storage, and database activities like database refreshes and offloading database backups to a dedicated RMAN server. The remainder of this section discusses an overview of SC Series features, benefits and SC configuration best practices for an Oracle deployment.

## 3.1 SC Series features

**Dynamic block architecture** records and tracks metadata for every block of data and provides system intelligence on how those blocks are being used. The metadata enables SC Series storage to take a more sophisticated and intelligent approach to storing, recovering, and managing data.

**Storage virtualization** virtualizes enterprise storage at the disk level, creating a dynamic pool of storage resources shared by all servers. Because read/write operations are spread across all available drives within the same tier, multiple requests are processed in parallel, boosting system performance

**Dynamic capacity (thin provisioning)** delivers high storage utilization by eliminating allocated but unused capacity. It completely separates storage allocation from utilization, enabling users to create any size virtual volume upfront, and only consume actual physical capacity when data is written.

**Data instant replay (DIR)** is a snapshot technology that provides continuous, space-efficient data protection. A snapshot taken of a volume creates a point-in-time copy (PITC) of the volume by making all written pages read-only. Any further changes to the volume get written to new pages (active data). When the volume is read, SC Series storage seamlessly presents the read-only pages from the snapshot and any active data. Consistent snapshots in an Oracle environment can be effective for database backups, recoveries, and cloning. For more information on consistent snapshots, see sections 3.20, 3.21, and 3.22.

**Data Progression (DP)** is a Dell Fluid Data™ storage or automated tiered storage feature that automatically migrates data to the optimal storage tier based on a set of predefined or custom policies called storage profiles. Data Progression eliminates the need to manually classify and migrate data to different storage tiers while reducing the number and cost of drives and reducing the cooling and power costs.

**Fast Track** technology enhances automated tiered storage by dynamically placing the most frequently accessed data on the fastest, or outer, tracks of each hard disk drive. Fast Track does not require any manual configuration and it is licensed separately. To see how RAID groups of a storage tier are allocated between standard and fast tracks, see the **Track** column in Figure 18. A value of **Fast** indicates that Fast Track is being used within the RAID group. If Fast Track is licensed, it is enabled by default and will be utilized behind the scenes. No manual configuration is required.

For additional documentation on these features, see the references in appendix B.1.

## 3.2 Benefits of deploying Oracle on SC Series storage

Some of the benefits of deploying Oracle databases on SC Series storage are listed in Table 6.

Table 6 Benefits of Oracle database on SC Series storage

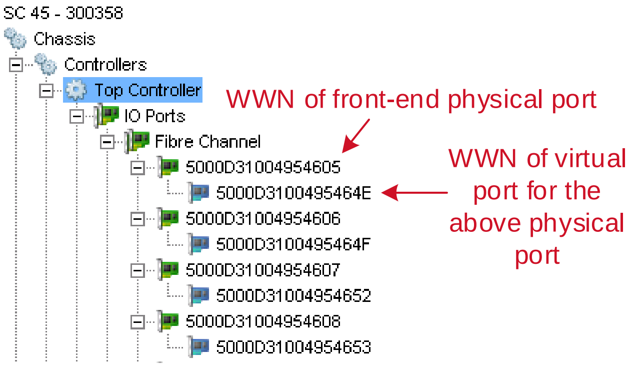
| Benefit | Details |
|---|---|
| Lower total cost of ownership (TCO) | Reduces acquisition, administration, and maintenance costs |
| Greater manageability | Ease of use, implementation, provisioning, and management |
| Simplified RAC implementation | Provides shared storage (raw or file systems) |
| High availability and scalability | Clustering provides higher levels of data availability and combined processing power of multiple server for greater throughput and scalability |
| Dell Information Life Cycle (ILM) benefits | Provides tiered storage, dynamic capacity, Data Progression, thin provisioning, snapshots, and more |
| Expand database buffer cache beyond the SGA in main memory | Provides the ability to extend Database Smart Flash Cache to tier 0 (T0) or T1 storage |

## 3.3 SC Series back-end connections

Should external disk enclosures be used, consider using either SC400 or SC420 as they support 12 Gbps SAS connectivity from the controllers to the enclosures. SC220 only supports 6Gbs SAS,

## 3.4 Virtual port mode

Dell EMC strongly recommends using SC Series virtual port mode over legacy port mode to provide redundancy at the front-end port and storage-controller level. If virtual port mode is used, the WWNs of the front-end virtual ports must be used in the zone rather than the front-end physical ports.



```
SC 45 - 300358
Chassis
Controllers
  Top Controller
    IO Ports
      Fibre Channel
        5000D31004954605
          5000D3100495464E
        5000D31004954606
          5000D3100495464F
        5000D31004954607
          5000D31004954652
        5000D31004954608
          5000D31004954653
```

WWN of front-end physical port
WWN of virtual port for the above physical port

Figure 6 WWN of physical and virtual ports of a SC Series controller

In the event of a controller or controller-port failure, virtual ports will fail over to any physical port within the same fault domain. Once the failure is resolved and ports are rebalanced in the SC Series array, the virtual ports return to their preferred physical port. Virtual port mode with FC networks provide benefits over legacy port mode including increased connectivity and bandwidth, and port failover for improved redundancy.

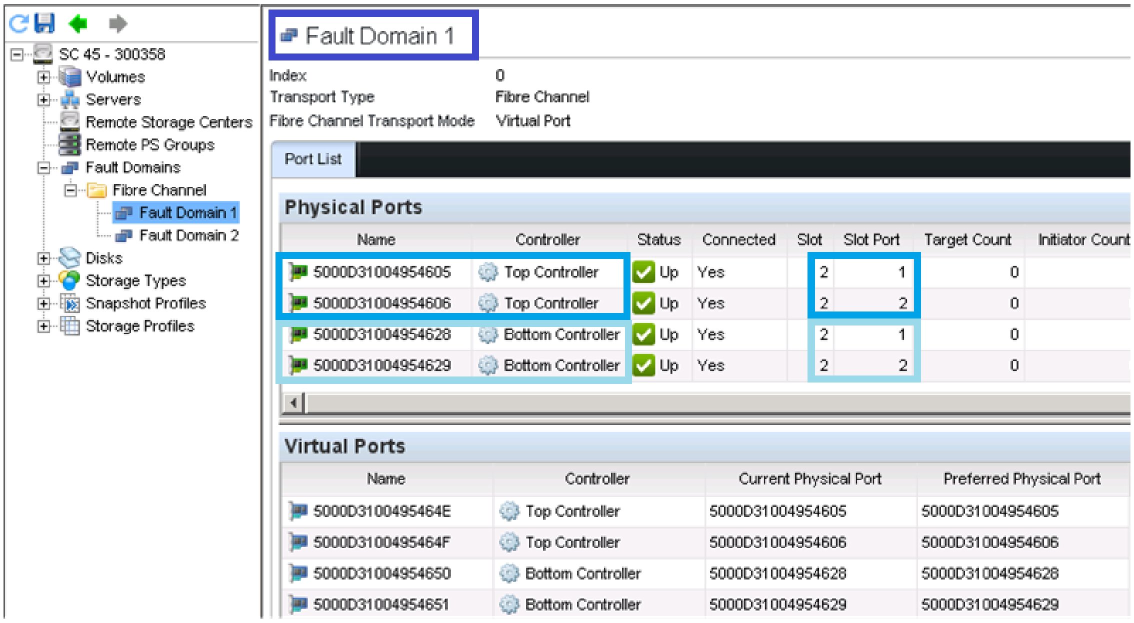
The actions taken in virtual port mode when a controller or controller port failure occurs are shown in Table 7.

Table 7 Failover mechanics with virtual ports

| Scenario | Action taken |
|---|---|
| Normal operation | All ports pass I/O |
| A controller fails in a dual-controller array | The virtual ports on the failed controller move to the physical ports on the remaining controller |
| Port failure | An individual port fails over to another port in the same fault domain |

## 3.5 Fault domains in virtual port mode

Fault domains are generally created when SC Series controllers are first configured using the **Configure Local Ports** wizard. If a fault domain needs to be changed after the initial controller configuration, refer to *Dell Storage Manager Administrator's Guide* or contact Dell support for guidance. Fault domains group SC Series controller front-end ports that belong to the same network and transport media (FC or iSCSI, but not both). Each front-end port must belong to a fault domain (see Figure 7 and Figure 8). Ports that belong to the same fault domain can fail over to each other because they have the same connectivity and transport protocol. Dell EMC recommends evenly distributing connections from each controller to each fault domain.



Fault Domain 1

| | |
|---|---|
| Index | 0 |
| Transport Type | Fibre Channel |
| Fibre Channel Transport Mode | Virtual Port |

Port List

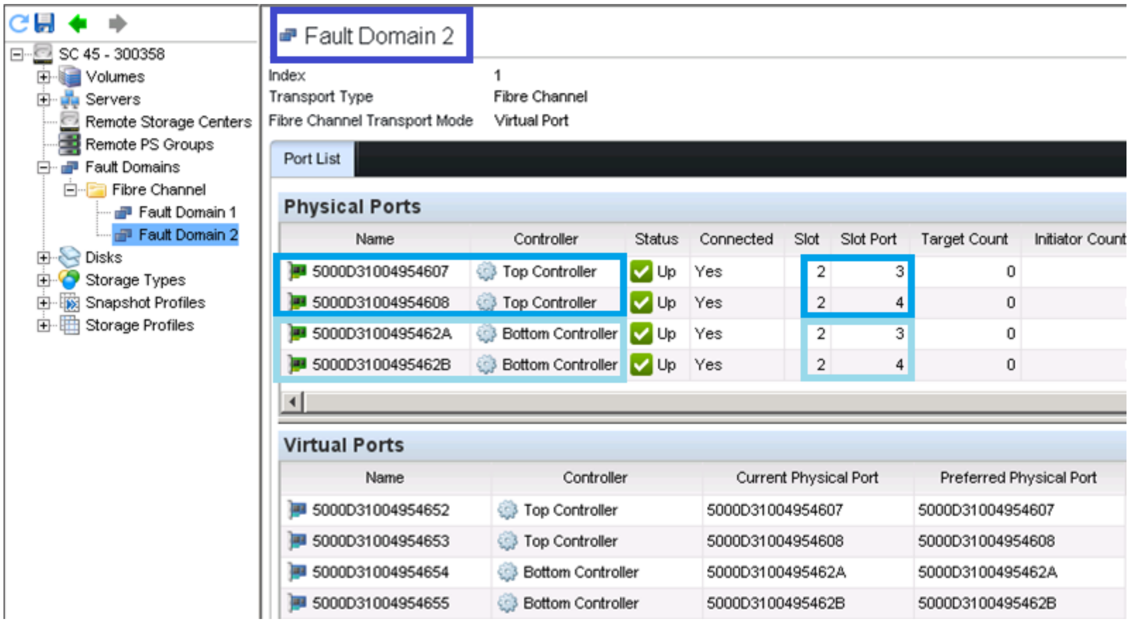
Physical Ports

| Name | Controller | Status | Connected | Slot | Slot Port | Target Count | Initiator Count |
|---|---|---|---|---|---|---|---|
| 5000D31004954605 | Top Controller | ✓ Up | Yes | 2 | 1 | 0 | |
| 5000D31004954606 | Top Controller | ✓ Up | Yes | 2 | 2 | 0 | |
| 5000D31004954628 | Bottom Controller | ✓ Up | Yes | 2 | 1 | 0 | |
| 5000D31004954629 | Bottom Controller | ✓ Up | Yes | 2 | 2 | 0 | |

Virtual Ports

| Name | Controller | Current Physical Port | Preferred Physical Port |
|---|---|---|---|
| 5000D3100495464E | Top Controller | 5000D31004954605 | 5000D31004954605 |
| 5000D3100495464F | Top Controller | 5000D31004954606 | 5000D31004954606 |
| 5000D31004954650 | Bottom Controller | 5000D31004954628 | 5000D31004954628 |
| 5000D31004954651 | Bottom Controller | 5000D31004954629 | 5000D31004954629 |

Figure 7 SC Series fault domain 1 and physical/virtual port assignments



Fault Domain 2

|  |  |
|---|---|
| Index | 1 |
| Transport Type | Fibre Channel |
| Fibre Channel Transport Mode | Virtual Port |

Physical Ports

| Name | Controller | Status | Connected | Slot | Slot Port | Target Count | Initiator Count |
|---|---|---|---|---|---|---|---|
| 5000D31004954607 | Top Controller | ✓ Up | Yes | 2 | 3 | 0 |  |
| 5000D31004954608 | Top Controller | ✓ Up | Yes | 2 | 4 | 0 |  |
| 5000D3100495462A | Bottom Controller | ✓ Up | Yes | 2 | 3 | 0 |  |
| 5000D3100495462B | Bottom Controller | ✓ Up | Yes | 2 | 4 | 0 |  |

Virtual Ports

| Name | Controller | Current Physical Port | Preferred Physical Port |
|---|---|---|---|
| 5000D31004954652 | Top Controller | 5000D31004954607 | 5000D31004954607 |
| 5000D31004954653 | Top Controller | 5000D31004954608 | 5000D31004954608 |
| 5000D31004954654 | Bottom Controller | 5000D3100495462A | 5000D3100495462A |
| 5000D31004954655 | Bottom Controller | 5000D3100495462B | 5000D3100495462B |

Figure 8 SC Series fault domain 2 and physical/virtual port assignments

## 3.6 Redundancy for SC Series front-end connections

The following types of redundancy are available:

**Storage controller redundancy**: The ports on an offline storage controller move to the remaining available storage controller.

**Storage controller port redundancy**: I/O activity on a failed port moves to another available port in the same fault domain (providing virtual port mode is enabled).

**Server path redundancy**: Multipathing is configured on the server and multiple paths exist between the SC Series array and server, allowing the server to use multiple paths for I/O. If an I/O path becomes unavailable, the server continues to use the remaining active paths.

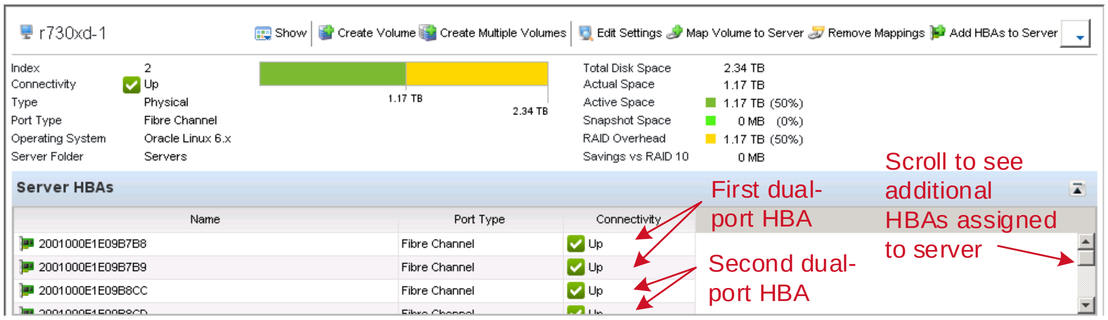
Dell EMC strongly recommends that all types of redundancy be provided in Oracle environments.

## 3.7 Creating server objects in SC Series arrays

Each physical server that uses SC Series storage must have a corresponding logical server object defined in DSM. To simplify defining the logical server objects in DSM, create them after the physical server has been configured with the appropriate hardware, racked and cabled, has been zoned into the fabric, powered on, and HBAs have been enabled in QLogic BIOS. Powering on the server will deliver power to the HBAs and allow DSM to see the enabled HBAs in the server as DSM interrogates the fabric.

Dell EMC recommends that at least two 8 Gb or 16 Gb dual-port HBAs be installed in the physical server. Two dual-port HBAs provide redundancy of initiator ports and HBAs. Additional HBAs or HBA ports may be necessary to provide the bandwidth necessary to support the expected database I/O performance requirements. To see how the number of initiators can benefit performance, see *Dell EMC SC Series Storage and Oracle OLTP* and *Optimizing Dell EMC SC Series Storage for Oracle OLAP Processing*.

Figure 9 shows a server object in DSM with multiple initiators.



| | |
|---|---|
| Index | 2 |
| Connectivity | ✓ Up |
| Type | Physical |
| Port Type | Fibre Channel |
| Operating System | Oracle Linux 6.x |
| Server Folder | Servers |

| | | |
|---|---|---|
| Total Disk Space | | 2.34 TB |
| Actual Space | | 1.17 TB |
| Active Space | 🟩 | 1.17 TB (50%) |
| Snapshot Space | 🟩 | 0 MB (0%) |
| RAID Overhead | 🟨 | 1.17 TB (50%) |
| Savings vs RAID 10 | | 0 MB |

| Name | Port Type | Connectivity |
|---|---|---|
| 2001000E1E09B7B8 | Fibre Channel | ✓ Up |
| 2001000E1E09B7B9 | Fibre Channel | ✓ Up |
| 2001000E1E09B8CC | Fibre Channel | ✓ Up |
| 2001000E1E09B8CD | Fibre Channel | ✓ Up |

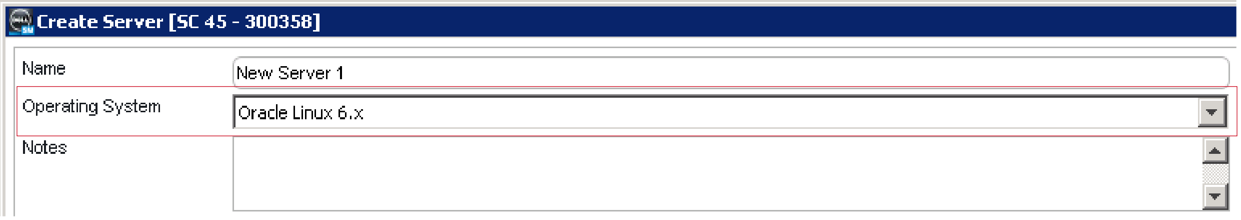
First dual-port HBA

Second dual-port HBA

Scroll to see additional HBAs assigned to server

Figure 9 HBAs assigned to DSM server object r730xd-1

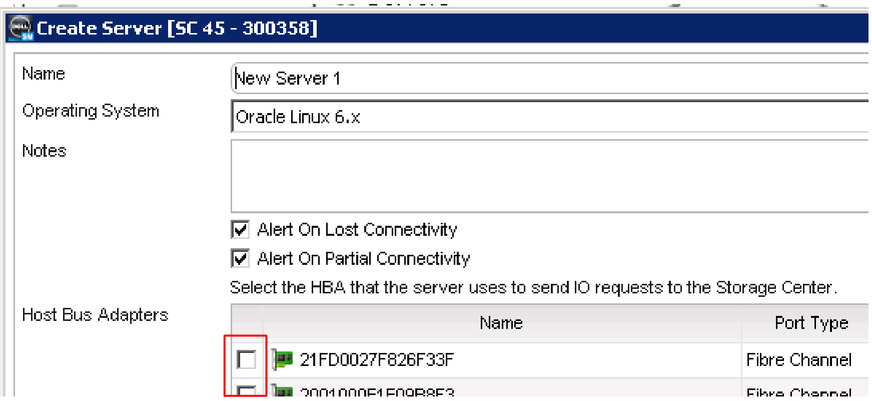
An important DSM server object attribute is **Operating System** (Figure 10). By setting **Operating System** to the type of OS intended to be installed on the physical server, SC Series storage implements a set of specific OS rules to govern the automated process of mapping volumes to the server object. For more information on mapping, see section 3.25.



| | |
|---|---|
| Name | New Server 1 |
| Operating System | Oracle Linux 6.x |
| Notes | |

Figure 10 Selecting the desired OS for a server object in DSM

All server HBAs transporting data to and from the SC Series array need to be defined in the corresponding DSM server object. To add a server HBA to the server object, select the check boxes associated with the server HBAs (Figure 11) that have been zoned to the SC Series array (see section 2.1), and enabled and scanned in QLogic BIOS (see section 2.2.1).



| | |
|---|---|
| Name | New Server 1 |
| Operating System | Oracle Linux 6.x |
| Notes | |

☑ Alert On Lost Connectivity

☑ Alert On Partial Connectivity

Select the HBA that the server uses to send IO requests to the Storage Center.

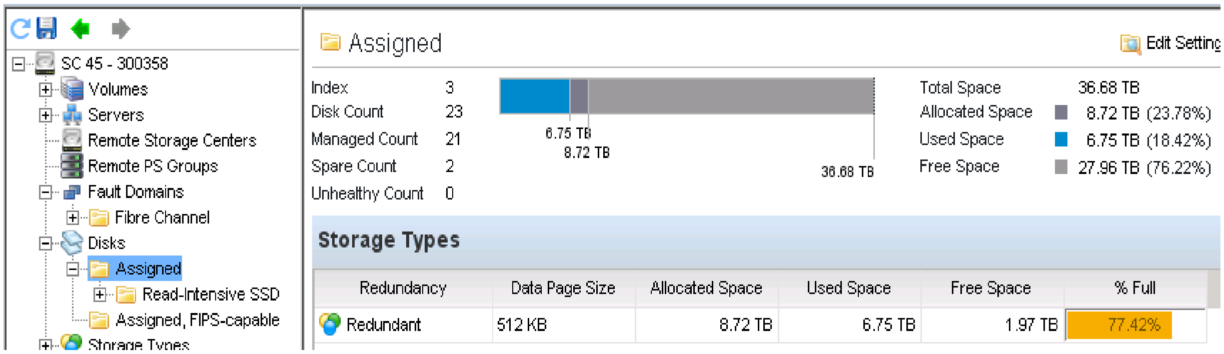
Host Bus Adapters

| | Name | Port Type |
|---|---|---|
| □ | 21FD0027F826F33F | Fibre Channel |
| □ | 2001000E1E09B8F3 | Fibre Channel |

Figure 11 Adding server HBAs to a server object

## 3.8 Disk pools and disk folders

In most configurations, regardless of the disk's performance characteristics, all disks should be assigned to a single disk folder to create one virtual pool of storage. The virtual pool of storage is referred to as the pagepool and it corresponds to a disk folder in the SC Series system. The default disk folder is called **Assigned.**



SC 45 - 300358
Volumes
Servers
Remote Storage Centers
Remote PS Groups
Fault Domains
Fibre Channel
Disks
Assigned
Read-Intensive SSD
Assigned, FIPS-capable
Storage Types

Assigned

Edit Setting

| | |
|---|---|
| Index | 3 |
| Disk Count | 23 |
| Managed Count | 21 |
| Spare Count | 2 |
| Unhealthy Count | 0 |

| | | |
|---|---|---|
| Total Space | | 36.68 TB |
| Allocated Space | ■ | 8.72 TB (23.78%) |
| Used Space | ■ | 6.75 TB (18.42%) |
| Free Space | ■ | 27.96 TB (76.22%) |

**Storage Types**

| Redundancy | Data Page Size | Allocated Space | Used Space | Free Space | % Full |
|---|---|---|---|---|---|
| Redundant | 512 KB | 8.72 TB | 6.75 TB | 1.97 TB | 77.42% |

Figure 12 Assigned disk folder

There are some cases where multiple disk folders (multiple pagepools) may be needed. In such cases, caution should be used as there are tradeoffs between single and multiple pagepools, the most notable one being performance. If the number of spindles is split between multiple pagepools, there will be a reduction in the maximum performance potential a pagepool can deliver.

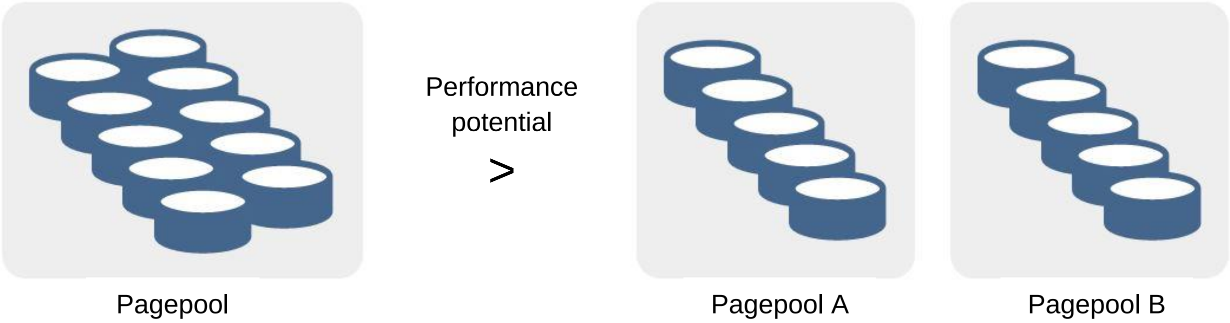
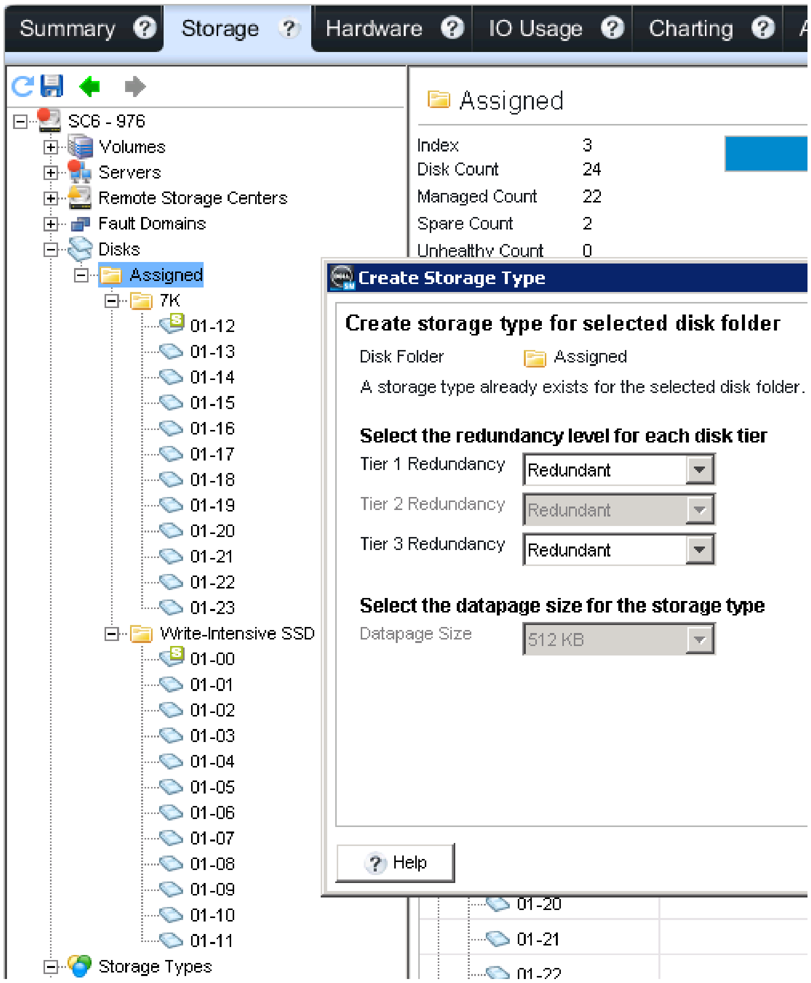


Figure 13 Pagepool performance potential

When dealing with high-performance database systems in which performance is affected by number of available spindles, Dell EMC strongly recommends only one pagepool (the **Assigned** disk folder) be defined. Only in rare circumstances do the benefits of multiple disk folders outweigh the disadvantages. Also, single pagepools are easier to manage and can accommodate an ever-changing workload. Another reason against using multiple pagepools is that Data Progression does not migrate storage across disk folders.

## 3.9 Storage types

When creating the **Assigned** disk folder, the SC Series system requires a storage type be defined and set for the folder (Figure 14). A storage type defines the type of **Tier Redundancy** applied to the available storage tiers, and the **Datapage Size**. Default values for **Tier Redundancy** and **Datapage Size** are heuristically generated and are appropriate for most deployments. Contact Dell Support for advice should a change from the default value be desired.



| | |
|---|---|
| Index | 3 |
| Disk Count | 24 |
| Managed Count | 22 |
| Spare Count | 2 |
| Unhealthy Count | 0 |

Create storage type for selected disk folder

Disk Folder: Assigned

A storage type already exists for the selected disk folder.

Select the redundancy level for each disk tier

Tier 1 Redundancy: Redundant

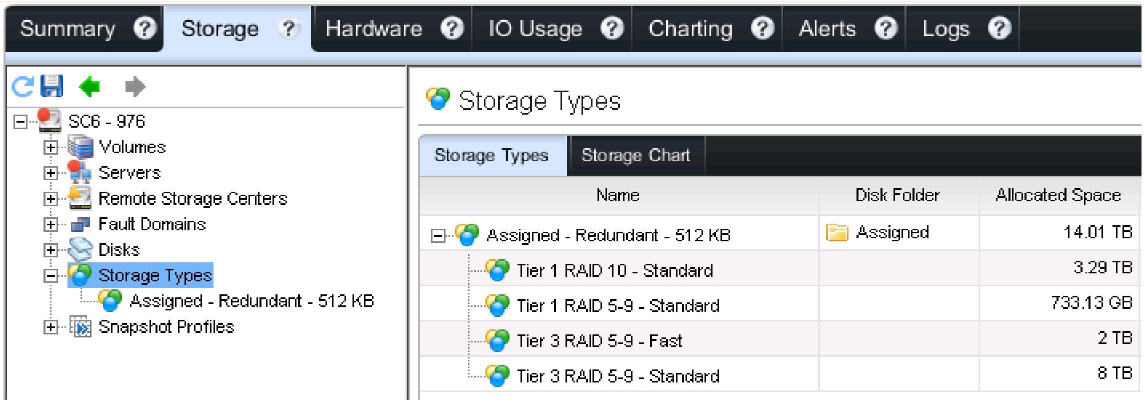
Tier 2 Redundancy: Redundant

Tier 3 Redundancy: Redundant

Select the datapage size for the storage type

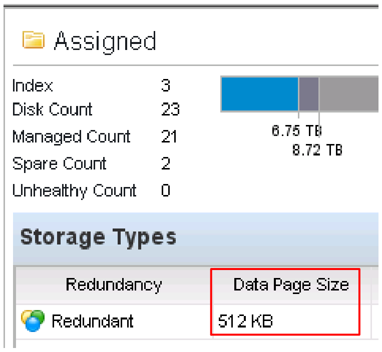
Datapage Size: 512 KB

Figure 14 Creating a storage type



| Name | Disk Folder | Allocated Space |
|---|---|---|
| Assigned - Redundant - 512 KB | Assigned | 14.01 TB |
| Tier 1 RAID 10 - Standard | | 3.29 TB |
| Tier 1 RAID 5-9 - Standard | | 733.13 GB |
| Tier 3 RAID 5-9 - Fast | | 2 TB |
| Tier 3 RAID 5-9 - Standard | | 8 TB |

Figure 15 Displaying defined storage types



| | |
|---|---|
| Index | 3 |
| Disk Count | 23 |
| Managed Count | 21 |
| Spare Count | 2 |
| Unhealthy Count | 0 |

Storage Types

| Redundancy | Data Page Size |
|---|---|
| Redundant | 512 KB |

Figure 16 **Assigned** disk folder configured with redundant 512 KB pages

## 3.10 Data page

The size of a data page is defined by the storage type and is the space taken from a disk folder and allocated to a volume when space is requested. In a default SC configuration, space is allocated from the **Assigned** disk folder. In a non-standard SC configuration where multiple pagepools are defined, space is allocated from the pagepool specified by the Storage Type selected for a volume. Possible data page sizes are:

**512 KB**: Appropriate for applications with high performance needs, or environments in which snapshots are taken frequently under heavy I/O. Selecting this size reduces the amount of space DSM can present to server objects. All flash, flash-optimized and hybrid storage types use 512 KB by default.

**2 MB**: Default value for non-flash arrays. Appropriate for most application program needs.

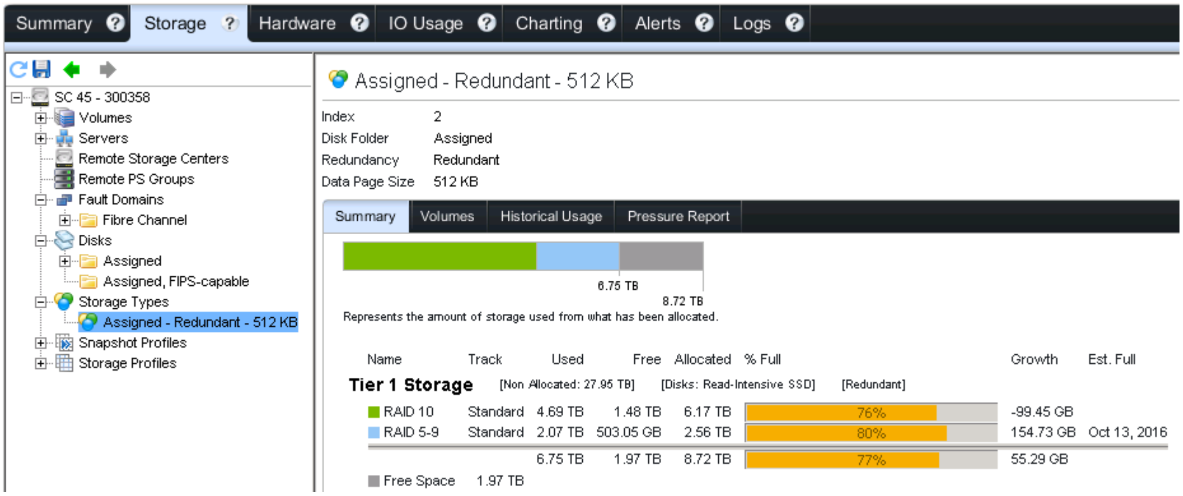
**4 MB**: Appropriate for configurations that use a large amount of disk space with infrequent snapshots.

**Note:** Dell EMC strongly recommends using the default data page size. If there is a need to change the data page size, consult with Dell Support to assess the performance impact and to ensure system resources remain balanced after changing the data page size.

When creating a volume, SC Series arrays assign the volume a maximum size as requested. Once the volume is presented to a server object, space from the pagepool will either be assigned to the volume all at once (preallocation) or when data is written to the volume. See section 3.26 for information on preallocation.

## 3.11 Tiered storage

All disk space within the pagepool is allocated into at least one, and up to three storage tiers. A tier defines the type of storage media used to save data. When only two types of disks are used in an SC Series array, the array automatically creates two tiers of storage. The fastest disks are placed in tier 1 (T1), and higher capacity, cost-efficient disks with lower performance metrics are assigned to tier 3 (T3). When three types of disks are used, the array automatically creates three tiers of storage. When using SC Series all-flash arrays with one type of media, one storage tier is created (T1) and it is protected by RAID 10 and RAID 5.



Assigned - Redundant - 512 KB

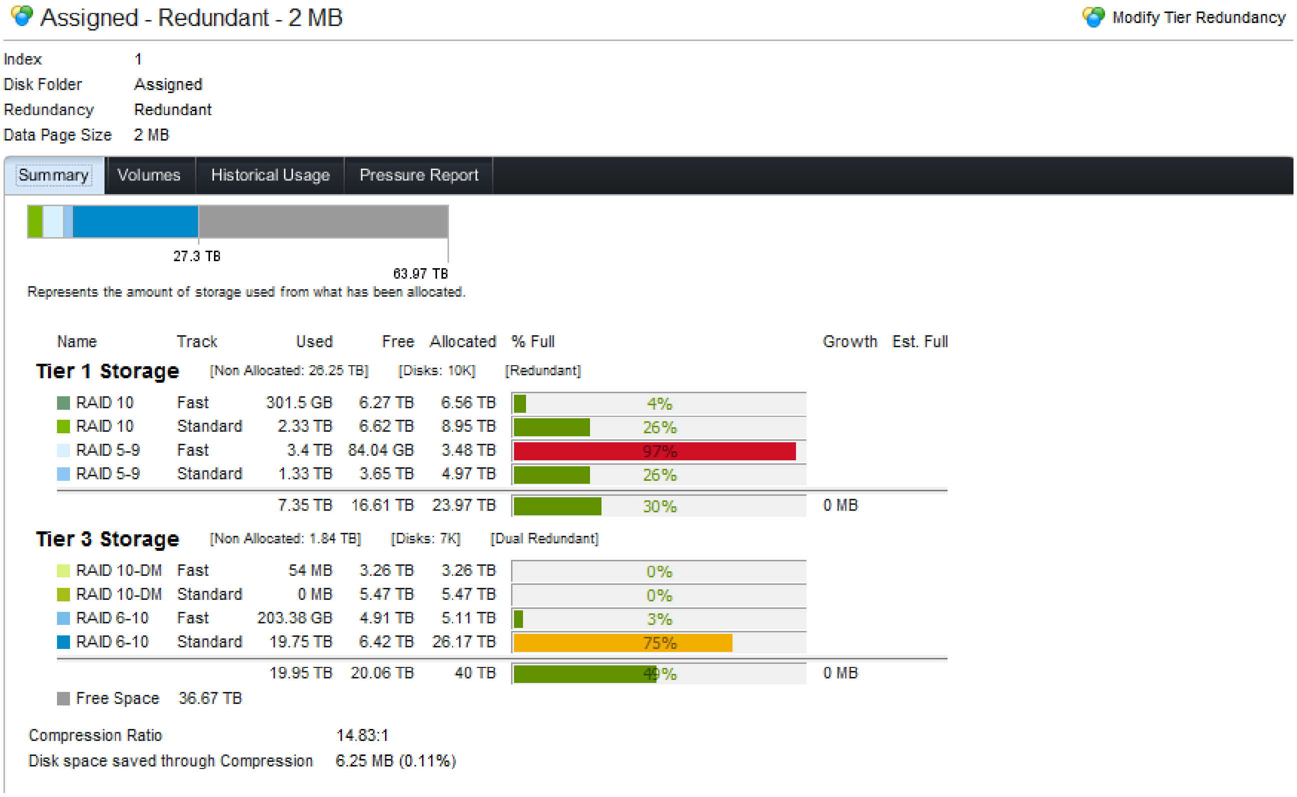
| | |
|---|---|
| Index | 2 |
| Disk Folder | Assigned |
| Redundancy | Redundant |
| Data Page Size | 512 KB |

Represents the amount of storage used from what has been allocated.

| Name | Track | Used | Free | Allocated | % Full | Growth | Est. Full |
|---|---|---|---|---|---|---|---|
| **Tier 1 Storage** [Non Allocated: 27.95 TB] | | | | [Disks: Read-Intensive SSD] | [Redundant] | | |
| 🟩 RAID 10 | Standard | 4.69 TB | 1.48 TB | 6.17 TB | 76% | -99.45 GB | |
| 🟦 RAID 5-9 | Standard | 2.07 TB | 503.05 GB | 2.56 TB | 80% | 154.73 GB | Oct 13, 2016 |
| | | 6.75 TB | 1.97 TB | 8.72 TB | 77% | 55.29 GB | |
| ▨ Free Space | 1.97 TB | | | | | | |

Figure 17 Storage types in SC Series all-flash arrays

When using multiple storage tiers, frequently accessed data remains on tier 1, and data that has not be accessed for the last 12 Data Progression cycles is gradually migrated to tier 2 (should it exist) and then to tier 3 (should it exist), providing Data Progression has been licensed.



Assigned - Redundant - 2 MB

| | |
|---|---|
| Index | 1 |
| Disk Folder | Assigned |
| Redundancy | Redundant |
| Data Page Size | 2 MB |

Represents the amount of storage used from what has been allocated.

| Name | Track | Used | Free | Allocated | % Full | Growth | Est. Full |
|---|---|---|---|---|---|---|---|
| **Tier 1 Storage** | [Non Allocated: 26.25 TB] | [Disks: 10K] | [Redundant] | | | | |
| RAID 10 | Fast | 301.5 GB | 6.27 TB | 6.56 TB | 4% | | |
| RAID 10 | Standard | 2.33 TB | 6.62 TB | 8.95 TB | 26% | | |
| RAID 5-9 | Fast | 3.4 TB | 84.04 GB | 3.48 TB | 97% | | |
| RAID 5-9 | Standard | 1.33 TB | 3.65 TB | 4.97 TB | 26% | | |
| | | 7.35 TB | 16.61 TB | 23.97 TB | 30% | 0 MB | |
| **Tier 3 Storage** | [Non Allocated: 1.84 TB] | [Disks: 7K] | [Dual Redundant] | | | | |
| RAID 10-DM | Fast | 54 MB | 3.26 TB | 3.26 TB | 0% | | |
| RAID 10-DM | Standard | 0 MB | 5.47 TB | 5.47 TB | 0% | | |
| RAID 6-10 | Fast | 203.38 GB | 4.91 TB | 5.11 TB | 3% | | |
| RAID 6-10 | Standard | 19.75 TB | 6.42 TB | 26.17 TB | 75% | | |
| | | 19.95 TB | 20.06 TB | 40 TB | 49% | 0 MB | |
| Free Space | 36.67 TB | | | | | | |

| | |
|---|---|
| Compression Ratio | 14.83:1 |
| Disk space saved through Compression | 6.25 MB (0.11%) |

Figure 18 Multiple storage tiers

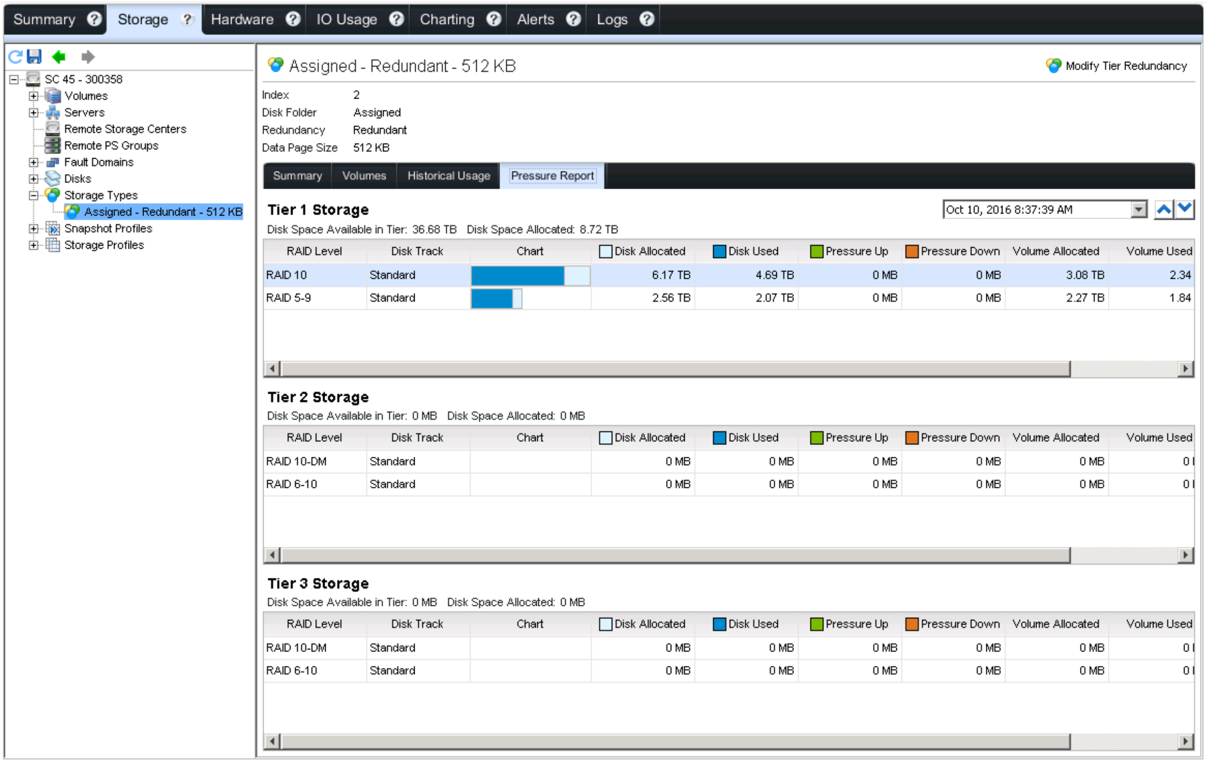
## 3.12 Tier redundancy and RAID

Data within tiers is protected by redundancy through the implementation of RAID technology. RAID requirements for each disk tier are based on the type, size, and number of disks in the tier, and will result in either single or dual redundancy of the data on a volume. In rare cases, redundancy for a tier can be disabled by using RAID 0, but caution is advised. For RAID 0 usage, contact Dell Support for advice.

When a page is requested from the page pool, SC Series applies the necessary dynamic RAID level, sometimes referred to as the redundancy level, to the page and assigns it to the tier of storage that requested it. During the life of the page, SC Series dynamically changes its RAID level as necessary. Each storage tier has two redundancy levels: single and dual. Within each redundancy level, multiple RAID levels can be defined. One RAID level is used when writing data to the page, and another for Data Progression and snapshot data. It should be noted that by default, any tier with a drive size of 900GB or larger is configured for dual redundancy, but single redundancy is also available.

Table 8 Tier redundancy and RAID types

| Tier redundancy | Description |
|---|---|
| Non-redundant | SC Series arrays will use RAID 0 in all classes, in all tiers. Data is striped but provides no redundancy. If one disk fails, all data is lost. Dell EMC does not recommend using non-redundant (RAID 0) storage unless data has been backed up elsewhere, and then only in specific cases after a thorough evaluation of business requirements. Contact Dell Support for advice. |
| Single-redundant | Protects against the loss of any one drive. Single-redundant tiers can contain any of the following types of RAID storage:<br>• RAID 10<br>• RAID 5-5 (requires 6-disk minimum)<br>• RAID 5-9 (requires 10-disk minimum) |
| Dual-redundant | Protects against the loss of any two drives. Disks larger than 900 GB should use dual redundancy and in some cases it is mandated. Dual-redundant tiers can contain any of the following types of RAID storage. Should the tier have fewer than the required 7 disk minimum for RAID 6, the default tier redundancy is set to single redundancy RAID 5-5 for any disk drive size.<br>• RAID 10 dual mirror (DM)<br>• RAID 6-6 (requires 7-disk minimum)<br>• RAID 6-10 (requires 11-disk minimum) |



Assigned - Redundant - 512 KB

Index 2
Disk Folder Assigned
Redundancy Redundant
Data Page Size 512 KB

Oct 10, 2016 8:37:39 AM

**Tier 1 Storage**

Disk Space Available in Tier: 36.68 TB Disk Space Allocated: 8.72 TB

| RAID Level | Disk Track | Chart | □ Disk Allocated | ■ Disk Used | 🟩 Pressure Up | 🟧 Pressure Down | Volume Allocated | Volume Used |
|---|---|---|---|---|---|---|---|---|
| RAID 10 | Standard | | 6.17 TB | 4.69 TB | 0 MB | 0 MB | 3.08 TB | 2.34 |
| RAID 5-9 | Standard | | 2.56 TB | 2.07 TB | 0 MB | 0 MB | 2.27 TB | 1.84 |

**Tier 2 Storage**

Disk Space Available in Tier: 0 MB Disk Space Allocated: 0 MB

| RAID Level | Disk Track | Chart | □ Disk Allocated | ■ Disk Used | 🟩 Pressure Up | 🟧 Pressure Down | Volume Allocated | Volume Used |
|---|---|---|---|---|---|---|---|---|
| RAID 10-DM | Standard | | 0 MB | 0 MB | 0 MB | 0 MB | 0 MB | 0 |
| RAID 6-10 | Standard | | 0 MB | 0 MB | 0 MB | 0 MB | 0 MB | 0 |

**Tier 3 Storage**

Disk Space Available in Tier: 0 MB Disk Space Allocated: 0 MB

| RAID Level | Disk Track | Chart | □ Disk Allocated | ■ Disk Used | 🟩 Pressure Up | 🟧 Pressure Down | Volume Allocated | Volume Used |
|---|---|---|---|---|---|---|---|---|
| RAID 10-DM | Standard | | 0 MB | 0 MB | 0 MB | 0 MB | 0 MB | 0 |
| RAID 6-10 | Standard | | 0 MB | 0 MB | 0 MB | 0 MB | 0 MB | 0 |

Figure 19 Tier redundancy and RAID types

Storage profiles define tier redundancy, or the RAID level used to protect data on a volume. In most Oracle environments, using the default tier redundancy provides appropriate levels of data protection, good I/O performance, and storage conservation for all types of database applications. Therefore, Dell EMC recommends using the default tier redundancy and evaluating its suitability before attempting to change it. If changing the default tier redundancy, do so only after evaluating application requirements during a RAID rebalance, should one be required. Consult Dell Support in assessing a change.

With high-performance Oracle applications, it is also recommended to evaluate gains in storage efficiencies with dual redundancy against increased latencies caused by the extra RAID penalty of dual redundancy. If the increased latencies are a concern, evaluate the suitability of using single redundancy. Keep in mind that single redundancy will provide greater storage efficiencies, less latency, but will provide less data protection in the event of multiple drive failures.

For information on redundancy requirements for each disk tier, see the *Dell Storage Manager Administrator's Guide* available on the Dell Support website.

## 3.13 Tier redundancy and media type

Media type used within a tier can influence the type of redundancy enforced on the tier.

Table 9 Media type and redundancy requirements

| Media type | Redundancy requirements |
|---|---|
| HDD | By default, disks under 966 GB are set to single redundancy.<br>Dual redundancy is recommended for 966 GB to 1.93 TB disks.<br>Dual redundancy is required when adding 1.93 TB or larger disks to a new disk pool. |
| SSD | By default, disks under 1.7 TB are set to single redundancy.<br>Dual redundancy is recommended for 1.7 TB to 2.78 TB disks.<br>In an SC Series system with new or existing disk pools, adding 2.79 TB or larger disks requires dual redundancy (single redundancy option is disabled). |

Dell EMC recommends using SSDs whenever possible for Oracle environments. In the event SSDs are not available, 15K HDDs are recommended. For additional information, see section 5.8.

## 3.14 RAID stripe width

The stripe width for RAID 5 is either 5 or 9, and for RAID 6 it is either 6 or 10. Modifying stripe width updates the corresponding RAID 5 or 6 selections for all predefined storage profiles. User-created storage profiles are not affected. Table 10 shows the stripe widths for all RAID levels in SC Series arrays.

In most Oracle implementations, the default stripe width should deliver the expected I/O performance and storage conservation. Therefore, Dell EMC recommends using the default stripe width within a tier and evaluating its suitability before attempting to change it. If changing the RAID stripe width, do so only after evaluating application requirements. Should a change be desired, consult Dell Support in assessing the change and the performance impact and to ensure system resources remain balanced during the RAID rebalance. A RAID rebalance will be required after it has been modified if the tier contains any disks

**Note:** A RAID rebalance should not be performed unless sufficient free disk space is available within the assigned disk folder, and only should be done when most appropriate for application requirements.

Table 10 RAID stripe width in SC Series arrays

| RAID level | Stripe width description |
|---|---|
| RAID 0 | Stripes across all drives in the tier with no redundancy |
| RAID 10 | Stripes data along with one copy across all drives in the tier |
| RAID 10-DM | Stripes data along with two copies across all drives in the tier |
| RAID 5-5 | Distributes parity across five drives (4 data segments, 1 parity segment for each stripe); tier requires at least 6 drives (5 for RAID, and one for spare) |
| RAID 5-9 | Distributes parity across nine drives (8 data segments, 1 parity segment for each stripe); tier requires at least 10 drives (9 for RAID, and one for spare) |
| RAID 6-6 | Distributes parity across six drives (4 data segments, 2 parity segments for each stripe); tier requires at least 7 drives (6 for RAID, and one for spare) |
| RAID 6-10 | Distributes parity across ten drives (8 data segments, 2 parity segments for each stripe); tier requires at least 11 drives (10 for RAID, and one for spare) |

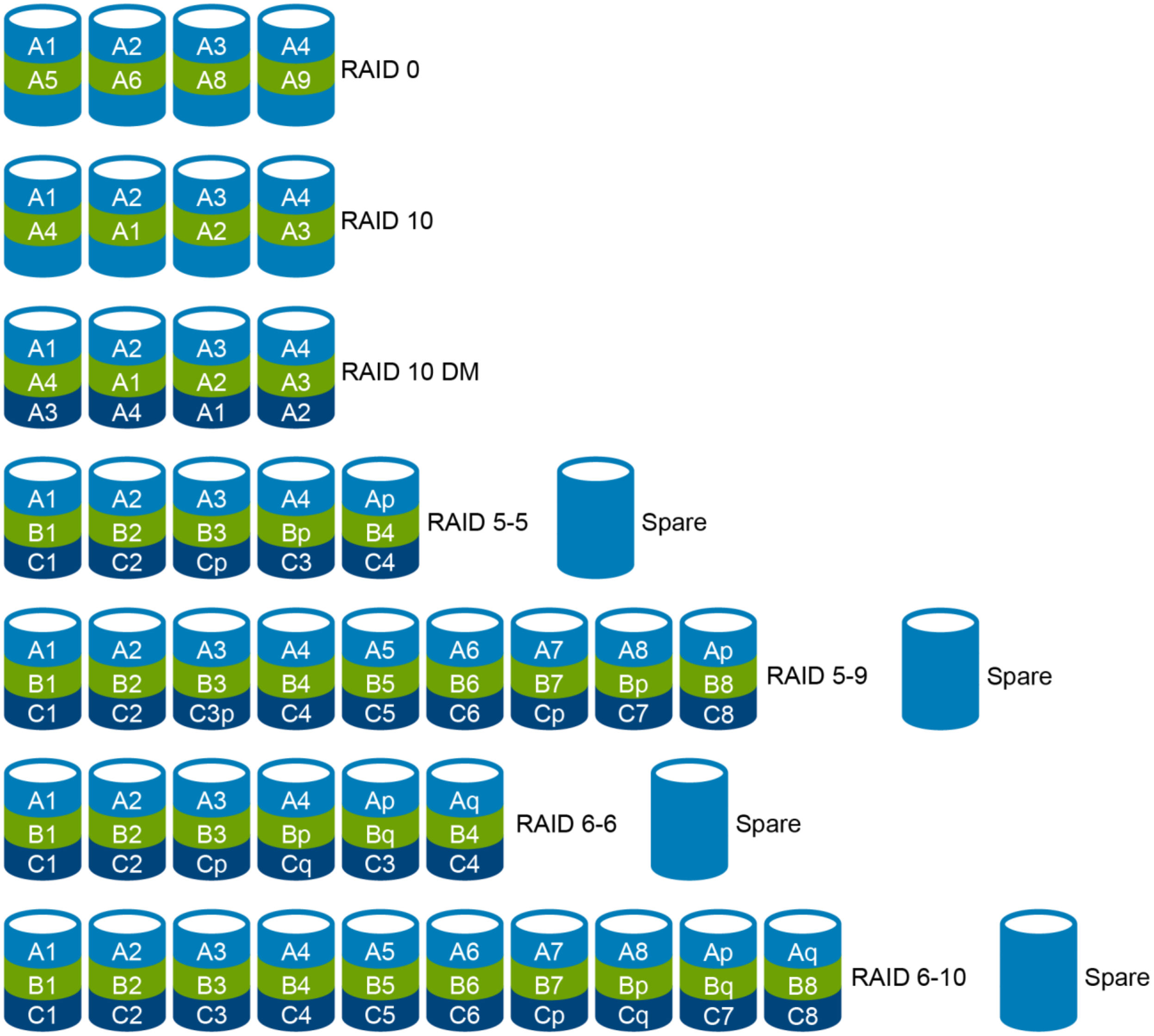
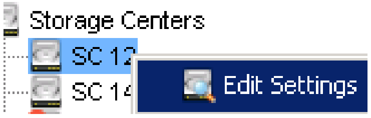
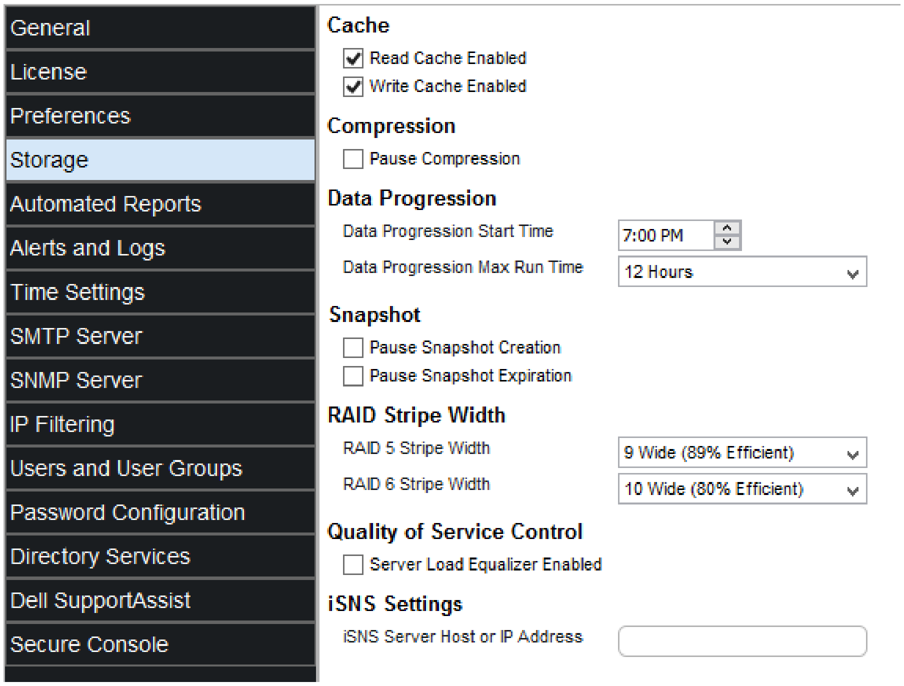


Figure 20 RAID stripes in SC Series arrays

SC Series arrays store most active data on RAID 10, and least active data on RAID 5 or RAID 6. Distributing data across more drives is marginally less efficient, but decreases vulnerability. Conversely, distributing data across fewer drives is more efficient, but marginally increases vulnerability.

To view RAID stripe widths and efficiencies, in DSM right-click the array, select **Edit Settings**, and click **Storage**.





Cache

☑ Read Cache Enabled

☑ Write Cache Enabled

Compression

□ Pause Compression

Data Progression

Data Progression Start Time: 7:00 PM

Data Progression Max Run Time: 12 Hours

Snapshot

□ Pause Snapshot Creation

□ Pause Snapshot Expiration

RAID Stripe Width

RAID 5 Stripe Width: 9 Wide (89% Efficient)

RAID 6 Stripe Width: 10 Wide (80% Efficient)

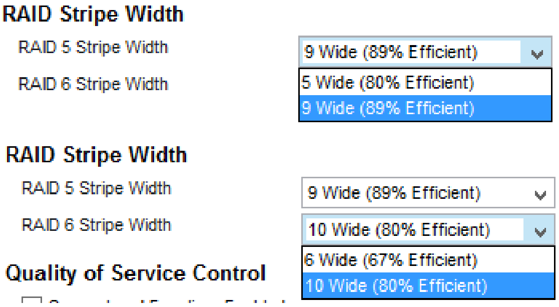
Quality of Service Control

□ Server Load Equalizer Enabled

iSNS Settings

iSNS Server Host or IP Address:

The **RAID Stripe Width** drop-down fields show the available stripe widths for RAID 5 and RAID 6.



RAID Stripe Width

RAID 5 Stripe Width: 9 Wide (89% Efficient)

RAID 6 Stripe Width

RAID Stripe Width

RAID 5 Stripe Width: 9 Wide (89% Efficient)

RAID 6 Stripe Width: 10 Wide (80% Efficient)

Quality of Service Control

Figure 21 RAID efficiencies of stripe widths

## 3.15 RAID penalty

Depending on the RAID level implemented, data and parity information may be striped between multiple disks. Before any write operation is considered complete, the parity must be calculated for the data and written to disk. The time waiting for the parity information to be written to disk, or the number of extra I/Os required for parity, is referred to the RAID (or write) penalty. The penalty only comes into play when a write is required. With RAID 0, there is no penalty because there is no parity. When there is no penalty, the write or RAID penalty is expressed as a 1 (the one write required to write the original data to disk).

Table 11 lists the RAID penalty and description for each SC Series RAID level.

Table 11 RAID penalty

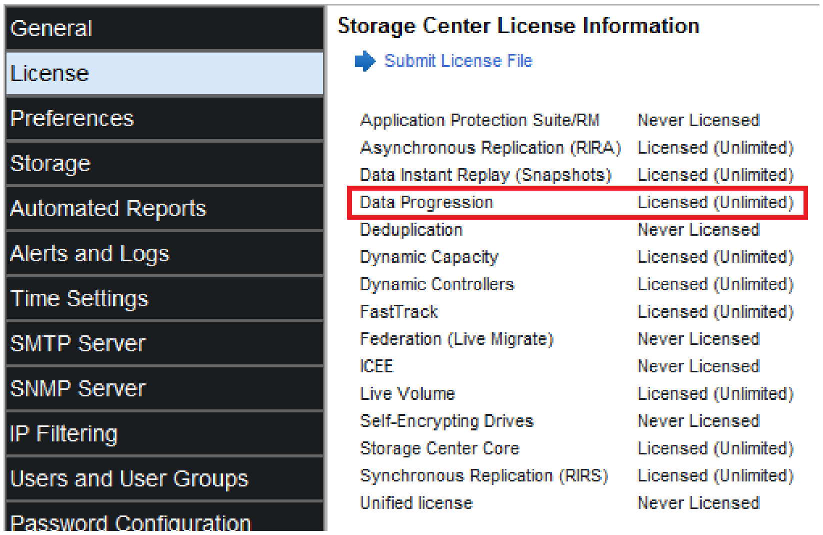
| RAID level | RAID penalty | I/O description |
|---|---|---|
| RAID 0 | 1 | 1 write |
| RAID 10 | 2 | 2 writes (one for data and one for copy of data) |
| RAID 10-DM | 3 | 3 writes (one for data and two for copy of data) |
| RAID 5-5<br>RAID 5-9 | 4 | 2 reads (one for data and one for parity),<br>2 writes (one for data and one for parity) |
| RAID 6-6<br>RAID 10-10 | 6 | 3 reads (one for data and two for parity),<br>3 writes (one for data and two for parity) |

Although RAID 0 has the lowest RAID penalty, Dell EMC does not recommend using it because there is no redundancy. SC Series storage uses RAID 10 by default for every write in all storage profiles.

## 3.16 Data Progression

Data Progression is a separately licensed SC Series feature that provides the maximum use of lower-cost drives for stored data, while maintaining high-performance drives for frequently-accessed data. The behavior of Data Progression is determined by the storage profile applied to each volume. If storage profile **Recommended (All Tiers)** is used in an all-flash system, snapshot data is immediately moved to T1 RAID 5-9 or T1 RAID 6-10. Writes to the volume still occur to T1 RAID 10 or T1 RAID 10 DM. This allows T1 space to be used more efficiently. Dell EMC strongly recommends the use of Data Progression for Oracle deployments.

To see if **Data Progression** is licensed, right-click a SC Series array in DSM, select **Edit Settings**, and select **License**.

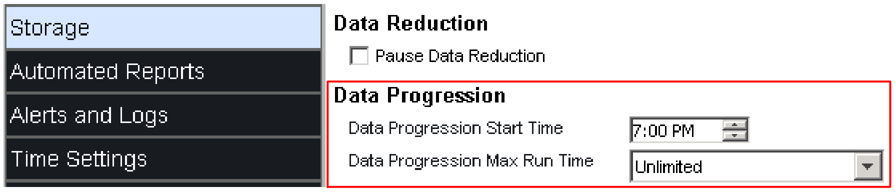


Storage Center License Information

| | |
|---|---|
| Application Protection Suite/RM | Never Licensed |
| Asynchronous Replication (RIRA) | Licensed (Unlimited) |
| Data Instant Replay (Snapshots) | Licensed (Unlimited) |
| Data Progression | Licensed (Unlimited) |
| Deduplication | Never Licensed |
| Dynamic Capacity | Licensed (Unlimited) |
| Dynamic Controllers | Licensed (Unlimited) |
| FastTrack | Licensed (Unlimited) |
| Federation (Live Migrate) | Never Licensed |
| ICEE | Never Licensed |
| Live Volume | Licensed (Unlimited) |
| Self-Encrypting Drives | Never Licensed |
| Storage Center Core | Licensed (Unlimited) |
| Synchronous Replication (RIRS) | Licensed (Unlimited) |
| Unified license | Never Licensed |

Figure 22 Verifying Data Progression license

By default, Data Progression runs every 24 hours at 7 PM system time. This schedule start time can be changed to avoid any resource contention between heavy I/O activity produced by databases and the activity generated by Data Progression cycles. A maximum elapsed time of Data Progression cycles can also be set. It is recommended to consult with Dell Support to assess the appropriateness of a change. To update the scheduled time of Data Progression, right-click a SC Series array, select **Edit Settings**, select **Storage**, and set the start and maximum elapsed time for Data Progression.



□ Pause Data Reduction

**Data Progression**

Data Progression Start Time: 7:00 PM

Data Progression Max Run Time: Unlimited

Figure 23 Setting the start time and maximum elapsed run time for Data Progression cycles

SC Series arrays also use Data Progression to move snapshots. When a snapshot is created, the data is frozen and moved immediately to the tier specified by the storage profile to hold snapshots.

Table 12 Snapshot storage classes used by Data Progression

| Storage profile | Tier 1 | Tier 2 | Tier 3 |
|---|---|---|---|
| Recommended (All Tiers) | RAID 5-9, RAID 6-10 | RAID 5-9, RAID 6-10 | RAID 5-9, RAID 6-10 |
| High Priority (Tier 1) | RAID 5-9, RAID 6-10 | | |
| Flash Optimized with Progression (Tier 1 to All Tiers) | | RAID 5-9, RAID 6-10 | RAID 5-9, RAID 6-10 |
| Write Intensive (Tier 1) | RAID 10, RAID 10-DM | | |
| Flash Only with Progression (Tier 1 to Tier 2) | | RAID 5-9, RAID 6-10 | |
| Low Priority with Progression (Tier 3 to Tier 2) | | RAID 5-9, RAID 6-10 | RAID 5-9, RAID 6-10 |
| Medium Priority (Tier 2) | | RAID 5-9, RAID 6-10 | |
| Low Priority (Tier 3) | | | RAID 5-9, RAID 6-10 |

Snapshots can occur in multiple ways:

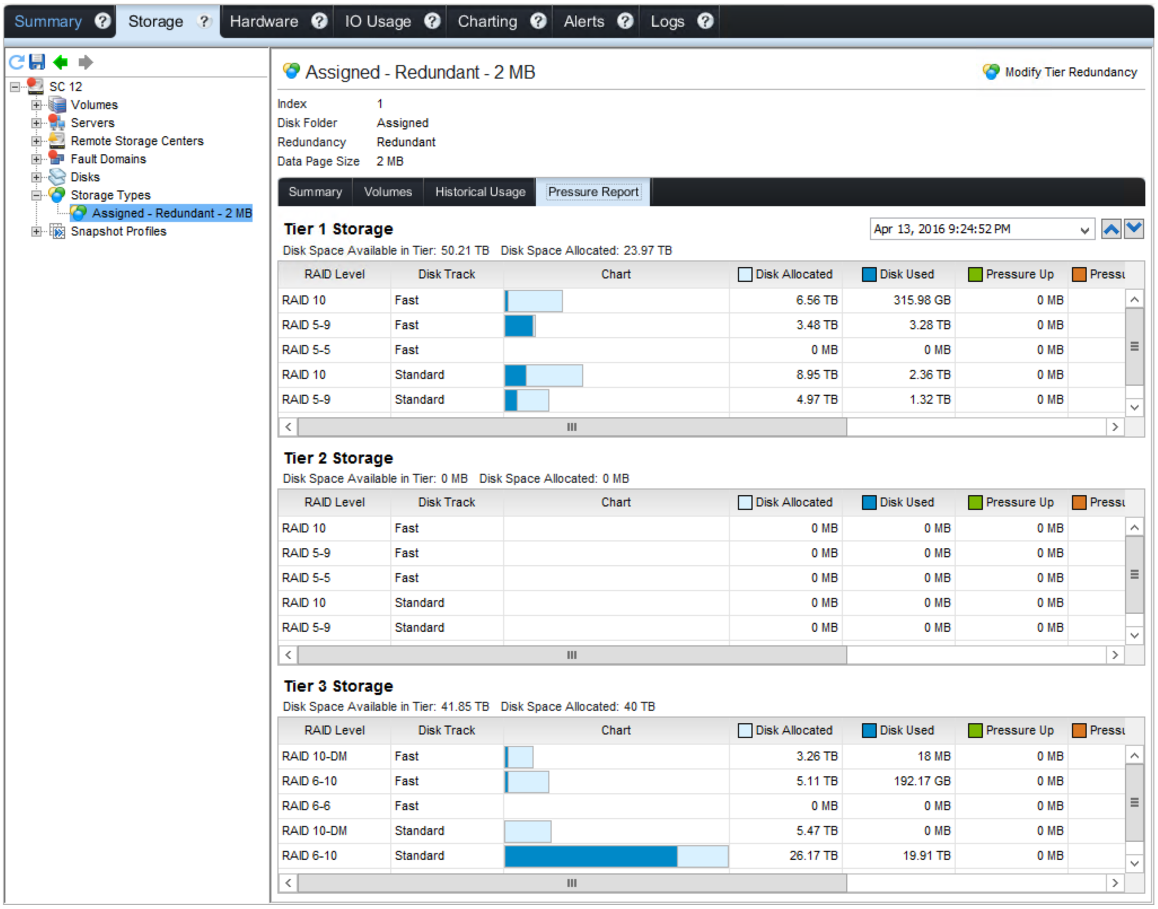
- As a scheduled event according to the snapshot profile
- By manual selection through DSM
- On demand by the SC Series array to move data off T1

## 3.17 Data Progression pressure reports

A tier can become full through normal usage, by data movement from Data Progression cycles, or from frequent database snapshots with long retention periods. When a tier becomes full, the SC Series array writes data to the next lower tier which can cause performance issues because of the RAID penalty. Therefore, Dell EMC recommends using Data Progression pressure reports to monitor disk usage. Data Progression pressure reports display how space is allocated and consumed across different RAID types and storage tiers for the previous 30 days. This trending information can assist in fine-tuning snapshot profiles and retention periods for databases and cloning operations.

To view the Data Progression pressure report, perform the following:

1. In DSM, select the **Storage** view.
2. In the **Storage** pane, select a SC Series array.
3. Click the **Storage** tab.
4. In the navigation pane, expand **Storage Type.**
5. Select the individual storage type to examine.
6. Select the tab, **Pressure Report**.



Assigned - Redundant - 2 MB

| | |
|---|---|
| Index | 1 |
| Disk Folder | Assigned |
| Redundancy | Redundant |
| Data Page Size | 2 MB |

**Tier 1 Storage**

Disk Space Available in Tier: 50.21 TB  Disk Space Allocated: 23.97 TB

| RAID Level | Disk Track | Chart | □ Disk Allocated | ■ Disk Used | 🟩 Pressure Up | 🟧 Pressu |
|---|---|---|---|---|---|---|
| RAID 10 | Fast | | 6.56 TB | 315.98 GB | 0 MB | |
| RAID 5-9 | Fast | | 3.48 TB | 3.28 TB | 0 MB | |
| RAID 5-5 | Fast | | 0 MB | 0 MB | 0 MB | |
| RAID 10 | Standard | | 8.95 TB | 2.36 TB | 0 MB | |
| RAID 5-9 | Standard | | 4.97 TB | 1.32 TB | 0 MB | |

**Tier 2 Storage**

Disk Space Available in Tier: 0 MB  Disk Space Allocated: 0 MB

| RAID Level | Disk Track | Chart | □ Disk Allocated | ■ Disk Used | 🟩 Pressure Up | 🟧 Pressu |
|---|---|---|---|---|---|---|
| RAID 10 | Fast | | 0 MB | 0 MB | 0 MB | |
| RAID 5-9 | Fast | | 0 MB | 0 MB | 0 MB | |
| RAID 5-5 | Fast | | 0 MB | 0 MB | 0 MB | |
| RAID 10 | Standard | | 0 MB | 0 MB | 0 MB | |
| RAID 5-9 | Standard | | 0 MB | 0 MB | 0 MB | |

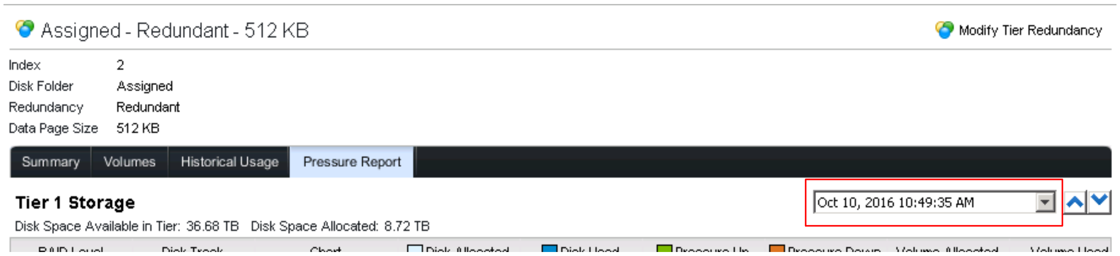
**Tier 3 Storage**

Disk Space Available in Tier: 41.85 TB  Disk Space Allocated: 40 TB

| RAID Level | Disk Track | Chart | □ Disk Allocated | ■ Disk Used | 🟩 Pressure Up | 🟧 Pressu |
|---|---|---|---|---|---|---|
| RAID 10-DM | Fast | | 3.26 TB | 18 MB | 0 MB | |
| RAID 6-10 | Fast | | 5.11 TB | 192.17 GB | 0 MB | |
| RAID 6-6 | Fast | | 0 MB | 0 MB | 0 MB | |
| RAID 10-DM | Standard | | 5.47 TB | 0 MB | 0 MB | |
| RAID 6-10 | Standard | | 26.17 TB | 19.91 TB | 0 MB | |

Figure 24 Data Progression Pressure Report

A drop-down field can be used to report on a specific date.



Assigned - Redundant - 512 KB

| | |
|---|---|
| Index | 2 |
| Disk Folder | Assigned |
| Redundancy | Redundant |
| Data Page Size | 512 KB |

**Tier 1 Storage**

Disk Space Available in Tier: 36.68 TB Disk Space Allocated: 8.72 TB

Oct 10, 2016 10:49:35 AM

Figure 25 Specific date period for pressure report

For more information on Data Progression pressure reports, see the *Dell Storage Manager Administrator's Guide* available on the Dell Support website.

## 3.18 Volume distribution reports

Dell EMC recommends reviewing volume metrics using subtabs **Volumes**, **Volume Growth**, and **Storage Chart** after selecting a volume in DSM. (See Figure 26).

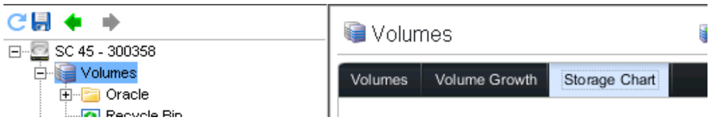
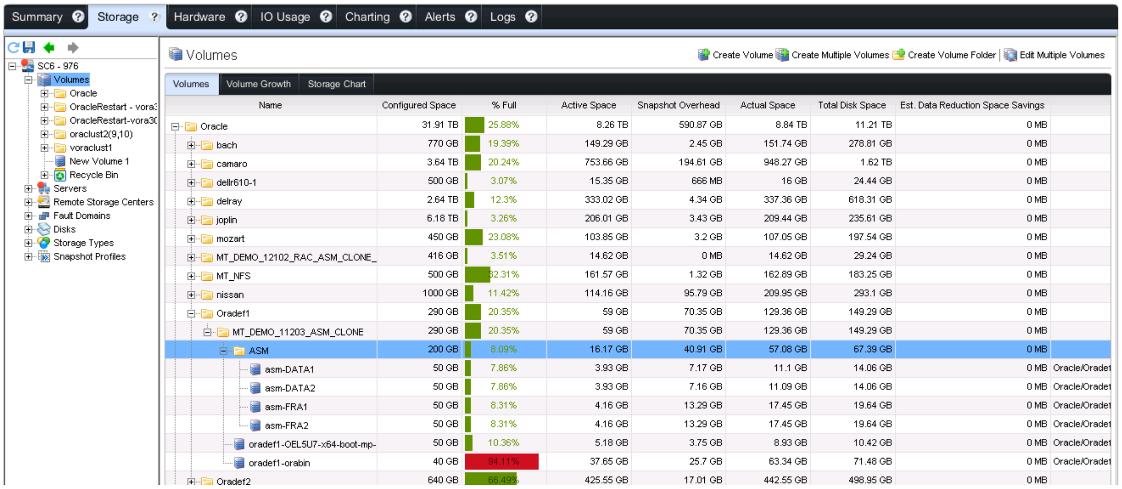


Figure 26 Volume distribution reports

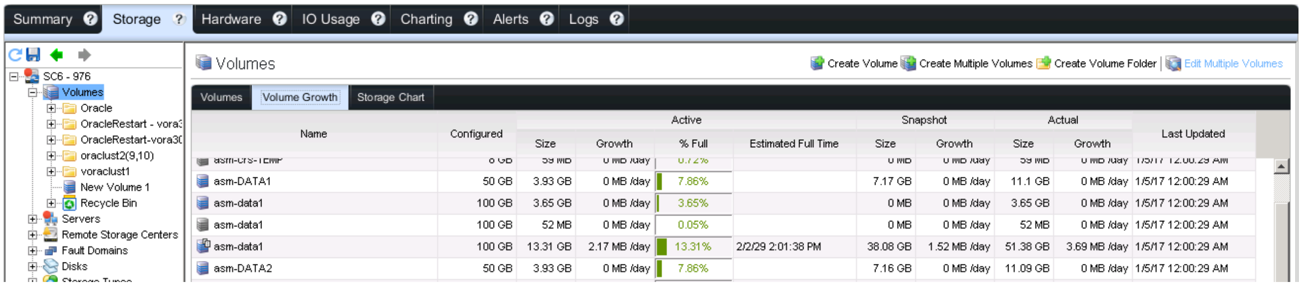
The metrics help to determine:

1. Growth trends (logical and physical space consumed and allocated space for a volume)
2. Volumes which have the largest overhead of snapshot data
3. Any necessary snapshot cleanup or snapshot profile adjustments justified
4. The amount of logical and physical space consumed and allocated space for a volume



| Name | Configured Space | % Full | Active Space | Snapshot Overhead | Actual Space | Total Disk Space | Est. Data Reduction Space Savings | |
|---|---|---|---|---|---|---|---|---|
| Oracle | 31.91 TB | 25.88% | 8.26 TB | 590.87 GB | 8.84 TB | 11.21 TB | 0 MB | |
| bach | 770 GB | 19.39% | 149.29 GB | 2.45 GB | 151.74 GB | 278.81 GB | 0 MB | |
| camaro | 3.64 TB | 20.24% | 753.66 GB | 194.61 GB | 948.27 GB | 1.62 TB | 0 MB | |
| dellr610-1 | 500 GB | 3.07% | 15.35 GB | 666 MB | 16 GB | 24.44 GB | 0 MB | |
| delray | 2.64 TB | 12.3% | 333.02 GB | 4.34 GB | 337.36 GB | 618.31 GB | 0 MB | |
| joplin | 6.18 TB | 3.26% | 206.01 GB | 3.43 GB | 209.44 GB | 235.61 GB | 0 MB | |
| mozart | 450 GB | 23.08% | 103.85 GB | 3.2 GB | 107.05 GB | 197.54 GB | 0 MB | |
| MT_DEMO_12102_RAC_ASM_CLONE_ | 416 GB | 3.51% | 14.62 GB | 0 MB | 14.62 GB | 29.24 GB | 0 MB | |
| MT_NFS | 500 GB | 32.31% | 161.57 GB | 1.32 GB | 162.89 GB | 183.25 GB | 0 MB | |
| nissan | 1000 GB | 11.42% | 114.16 GB | 95.79 GB | 209.95 GB | 293.1 GB | 0 MB | |
| Oradef1 | 290 GB | 20.35% | 59 GB | 70.35 GB | 129.36 GB | 149.29 GB | 0 MB | |
| MT_DEMO_11203_ASM_CLONE | 290 GB | 20.35% | 59 GB | 70.35 GB | 129.36 GB | 149.29 GB | 0 MB | |
| ASM | 200 GB | 8.09% | 16.17 GB | 40.91 GB | 57.08 GB | 67.39 GB | 0 MB | |
| asm-DATA1 | 50 GB | 7.86% | 3.93 GB | 7.17 GB | 11.1 GB | 14.06 GB | 0 MB | Oracle/Oradef |
| asm-DATA2 | 50 GB | 7.86% | 3.93 GB | 7.16 GB | 11.09 GB | 14.06 GB | 0 MB | Oracle/Oradef |
| asm-FRA1 | 50 GB | 8.31% | 4.16 GB | 13.29 GB | 17.45 GB | 19.64 GB | 0 MB | Oracle/Oradef |
| asm-FRA2 | 50 GB | 8.31% | 4.16 GB | 13.29 GB | 17.45 GB | 19.64 GB | 0 MB | Oracle/Oradef |
| oradef1-OEL5U7-x64-boot-mp- | 50 GB | 10.36% | 5.18 GB | 3.75 GB | 8.93 GB | 10.42 GB | 0 MB | Oracle/Oradef |
| oradef1-orabin | 40 GB | 94.11% | 37.65 GB | 25.7 GB | 63.34 GB | 71.48 GB | 0 MB | Oracle/Oradef |
| Oradef2 | 640 GB | 66.49% | 425.55 GB | 17.01 GB | 442.55 GB | 498.95 GB | 0 MB | |

Figure 27 Volume report



| Name | Configured | Active |  |  |  | Snapshot |  | Actual |  | Last Updated |
|---|---|---|---|---|---|---|---|---|---|---|
|  |  | Size | Growth | % Full | Estimated Full Time | Size | Growth | Size | Growth |  |
| asm-crs-TEMP | 8 GB | 59 MB | 0 MB /day | 0.72% |  | 0 MB | 0 MB /day | 59 MB | 0 MB /day | 1/5/17 12:00:29 AM |
| asm-DATA1 | 50 GB | 3.93 GB | 0 MB /day | 7.86% |  | 7.17 GB | 0 MB /day | 11.1 GB | 0 MB /day | 1/5/17 12:00:29 AM |
| asm-data1 | 100 GB | 3.65 GB | 0 MB /day | 3.65% |  | 0 MB | 0 MB /day | 3.65 GB | 0 MB /day | 1/5/17 12:00:29 AM |
| asm-data1 | 100 GB | 52 MB | 0 MB /day | 0.05% |  | 0 MB | 0 MB /day | 52 MB | 0 MB /day | 1/5/17 12:00:29 AM |
| asm-data1 | 100 GB | 13.31 GB | 2.17 MB /day | 13.31% | 2/2/29 2:01:38 PM | 38.08 GB | 1.52 MB /day | 51.38 GB | 3.69 MB /day | 1/5/17 12:00:29 AM |
| asm-DATA2 | 50 GB | 3.93 GB | 0 MB /day | 7.86% |  | 7.16 GB | 0 MB /day | 11.09 GB | 0 MB /day | 1/5/17 12:00:29 AM |

Figure 28 Volume growth report

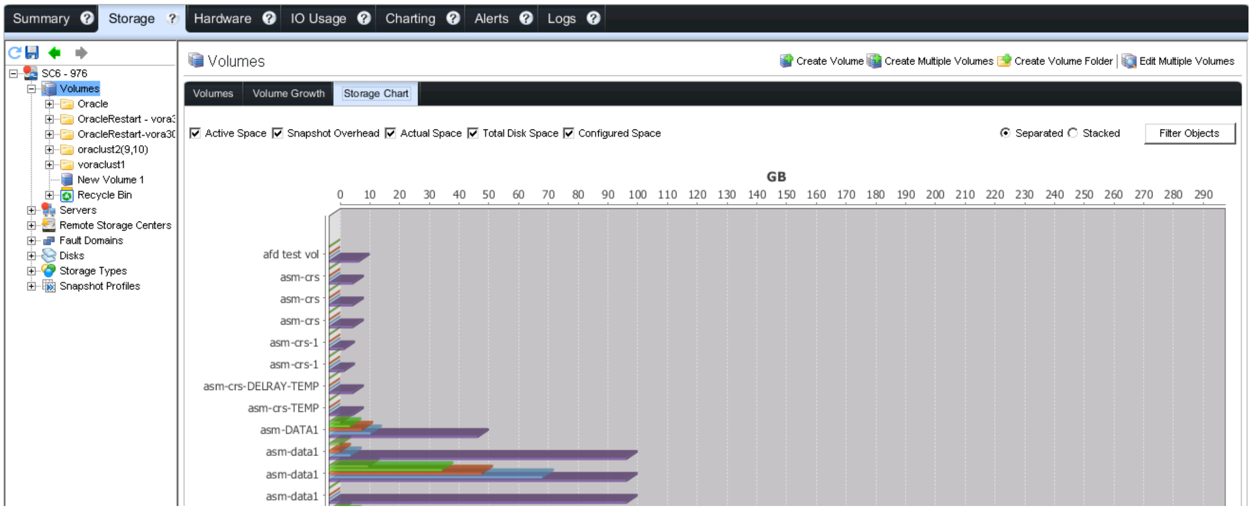


Figure 29 Volume growth chart

## 3.19 Storage profiles

Dell Fluid Data™ storage automatically migrates data (Data Progression) to the optimal storage tier based on a set of predefined or custom policies called storage profiles. Storage profiles:

- Define which tier of disk is used to accept initial writes, or new pages, to a volume
- Determine how to move data between and across tiers
- Set the RAID level used to protect data on a volume
- Apply profiles to each volume, by default

In most cases, using the recommended default storage profile provides appropriate levels of data protection, good performance metrics for all types of database volumes by ensuring all writes occur to T1 RAID 10, and movement of snapshots and less-active data to other RAID levels, lower storage tiers, or both. Therefore, Dell EMC recommends using the default storage profile and evaluating its appropriateness before attempting to use a different storage profile. If there is a need to pin the entire database, or part of it (active and less-active data) to T1, consider the impacts on T1 capacity and Data Progression caused by complete data rebuilds from nightly extract, transform, load (ETL) operations, index rebuilds, and associated snapshots.

Depending on the type of storage media used in SC series arrays, storage profiles will vary. The default standard storage profiles are described in Table 13.

Table 13 Default standard profiles in SC Series arrays (spinning media or hybrid arrays)

| Name | Initial write tier | Tier and RAID levels | Progression |
|---|---|---|---|
| Recommended (All Tiers) | 1 | Writeable:<br>• T1 RAID 10, RAID 10-DM<br>• T2 RAID 10, RAID 10-DM<br>Snapshots:<br>• T1 RAID 5-9, RAID 6-10<br>• T2 RAID 5-9, RAID 6-10<br>• T3 RAID 5-9, RAID 6-10 | To all tiers* |
| High Priority (Tier 1) | 1 | Writeable:<br>• T1 RAID 10, RAID 10-DM<br>Snapshots:<br>• T1 RAID 5-9, RAID 6-10 | No |
| Medium Priority (Tier 2) | 2 | Writeable:<br>• T2 RAID 10, RAID 10-DM<br>Snapshots:<br>• T2 RAID 5-9, RAID 6-10 | No |
| Low Priority (Tier 3) | 3 | Writeable:<br>• T3 RAID 10, RAID 10-DM<br>Snapshots:<br>• T3 RAID 5-9, RAID 6-10 | No |

* Available only when Data Progression is licensed.

**Recommended (All Tiers):** This is the default profile in which all new data is written to T1 RAID 10 or T1 RAID 10-DM. Data Progression moves less-active data to T1 RAID 5-9/RAID 6-10 or a slower tier based on how frequently the data is accessed. In this way, the most active blocks of data remain on high-performance SSDs or Fibre Channel drives, while less active blocks automatically move to lower-cost, high-capacity SAS drives. This requires the Data Progression license. If SSDs are used in a hybrid array, they are assigned to this storage profile.

**High Priority (Tier 1):** This storage profile provides the highest performance by storing written data in RAID 10 or RAID 10 DM on T1, and snapshot data in RAID 5-9/RAID 6-10 on T1. SC Series arrays do not migrate data to lower storage tiers unless T1 becomes full. If Data Progression is not licensed, this is the default storage profile. Without Data Progression, volumes must be configured to use a specific tier of storage, because data will not migrate between tiers. If SSDs are used in a hybrid array without Data Progression, they are assigned to this storage profile. Spinning media is assigned to **Medium Priority (Tier 2)** or **Low Priority (Tier 3)**. SC Series does not write data to a lower storage tier unless the intended tier becomes full.

**Medium Priority (Tier 2):** This storage profile provides a balance between performance and cost efficiency by storing written data in RAID 10 or RAID 10 DM on T2, and snapshot data in RAID 5-9/RAID 6-10 on T2. The SC Series array does not migrate data to other storage tiers unless T2 becomes full. SC Series storage does not write data to a lower storage tier unless the intended tier becomes full.

**Low Priority (Tier 3):** This storage profile provides the most cost-efficient storage by storing written data in RAID 10 or RAID 10 DM on T3, and snapshot data in RAID 5-9/RAID 6-10 on T3. The SC Series array does not migrate data to higher tiers of storage unless T3 becomes full. SC Series does not write data to a higher storage tier unless the intended tier becomes full.

Table 14 Default standard profiles in SC Series all-flash arrays

| Name | Initial write tier | Tier and RAID levels | Progression |
|---|---|---|---|
| Flash Optimized with Progression (Tier 1 to All Tiers) | 1 | Writeable:<br>• T1 RAID 10, RAID 10-DM<br>Snapshots:<br>• T2 RAID 5-9, RAID 6-10<br>• T3 RAID 5-9, RAID 6-10 | To all tiers |
| Write Intensive (Tier 1) | 1 | Writeable:<br>• T1 RAID 10, RAID 10-DM<br>Snapshots:<br>• T1 RAID 10, RAID 10-DM | No |
| Flash Only with Progression (Tier 1 to Tier 2) | 1 | Writeable:<br>• T1 RAID 10, RAID 10-DM<br>Snapshots:<br>• T2 RAID 5-9, RAID 6-10 | To tier 2 only |
| Low Priority with Progression (Tier 3 to Tier 2) | 3 | Writeable:<br>• T3 RAID 10, RAID 10-DM<br>Snapshots:<br>• T2 RAID 5-9, RAID 6-10<br>• T3 RAID 5-9, RAID 6-10 | To tier 2 only |
| Low Priority (Tier 3) | 3 | Writeable:<br>• T3 RAID 10, RAID 10-DM<br>Snapshots:<br>• T3 RAID 5-9, RAID 6-10 | No |

**Flash Optimized with Progression (Tier 1 to All Tiers):** This is the default profile which provides the most efficient storage for both read-intensive and write-intensive SSDs. All new data is written to write-intensive T1 drives, snapshot data is moved to T2 RAID 5-9/RAID 6-10, and less-active data progresses to T3 RAID 5-9/RAID 6-10. When T1 reaches 95 percent capacity, the SC Series array creates a space management snapshot and moves it immediately to T2 to free up space on T1. The space management snapshot is moved immediately and does not wait for a scheduled Data Progression. Space management snapshots are marked as created on demand and cannot be modified manually or used to create View Volumes. Space-management snapshots coalesce into the next scheduled or manual snapshot. The SC Series array creates only one on-demand snapshot per volume at a time. This requires the Data Progression license.

**Write Intensive (Tier 1):** This storage profile directs all initial writes to write-intensive SSDs on T1 RAID 10/RAID 10-DM. This data does not progress to any other tier. This profile is useful for storing Oracle redo logs and temporary database files and tablespaces.

**Flash Only with Progression (Tier 1 to Tier 2):** This storage profile performs initial writes to T1 RAID 10/RAID 10 DM on high-performance drives (write-intensive SLC SSDs). On-demand Data Progression then moves snapshot and less-active read-intensive data to T2 RAID 5-9/RAID 6-10, but remains on SSDs where it benefits from read-intensive MLC SSDs read performance. This profile is useful for storing volumes with data that require optimal read performance, such as Oracle database golden images or clones for rapid database deployments. Requires Data Progression license.

**Low Priority with Progression (Tier 3 to Tier 2):** This storage profile directs initial writes to T3 RAID 10/RAID 10 DM on less-expensive drives, and then allows frequently accessed data to progress from T3 RAID 5-9/RAID 6-10 to T2 RAID 5-9/RAID 6-10. This profile is useful for migrating large amounts of data to

Oracle, or infrequent large Oracle data loads without overloading T1 SSDs. This is also referred to the cost-optimized profile. Requires Data Progression license.

Because SSDs are automatically assigned to T1, profiles that include T1 allow volumes to use SSD storage. If there are volumes that contain data that is not accessed frequently, and do not require the performance of T1 SSDs, use profiles Medium Priority (Tier 2) or Low Priority (Tier 3), or create and apply a new profile that does not include high-performance disks.

In hybrid SC Series arrays (SSDs and HDDs), if there are not enough SSDs to retain the entire database, use custom storage profiles to separate the data between the different media, then monitor the performance to determine a best fit for each media type. When using SLCs with automated tiering, the tier needs to be sized for 100 percent IOPS and 100 percent capacity of the volumes using the SSD tier. If it is not and the tier becomes full, performance can degrade considerably. In cases like this, redistribute the data to more appropriate storage tiers.

Flash-optimized SC Series arrays (SLCs and MLCs) provide the benefit of delivering high capacity and performance. SLCs in this configuration should be sized to support the expected IOPS, and MLCs should be sized to support the expected capacity.

**Note:** For best performance with Data Progression, Dell EMC recommends using the storage profiles provided by the SC Series array. If a custom profile is needed, engage Dell Support to discuss the performance impact and to ensure that system resources remain balanced with a custom profile. Custom storage profiles are beyond the scope of this document.

Storage profiles provided by the SC Series array cannot be modified and by default cannot be seen in the DSM Storage navigation tree:

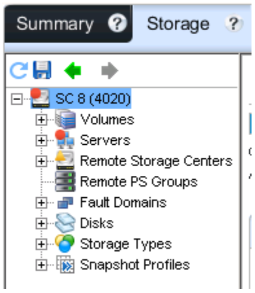
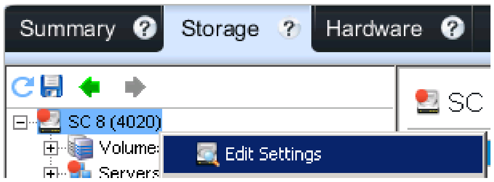


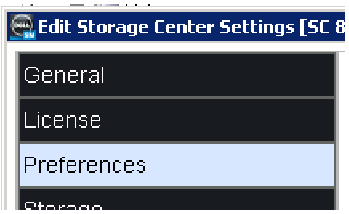
Figure 30 Storage Profiles not shown in the navigation tree

To display Storage Profiles in the navigation tree, or to be able to create custom storage profiles, perform the following:

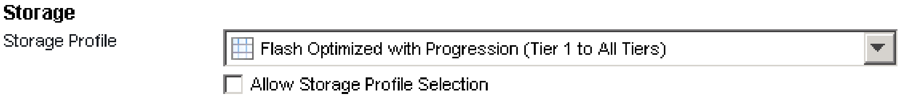
1. Right-click the SC Series array and select **Edit Settings.**



2. Select **Preferences**.



3. Select the check box, **Allow Storage Profile Selection**.



**Storage**

Storage Profile: Flash Optimized with Progression (Tier 1 to All Tiers)

□ Allow Storage Profile Selection

---

**Note:** Once the ability to view and create Storage Profiles has been enabled, it cannot be disabled.

---

## 3.20 Snapshot profiles

A snapshot profile is a policy-based collection of rules describing the type of schedule (once, daily, weekly, or monthly), a date/time, the interval to create the snapshot, the volumes to snapshot, an expiration time for the snapshot, if the snapshot should be write-consistent across all volumes in the snapshot profile, and the snapshot method. Regardless of the policy-based rules, each snapshot profile is limited to 40 volumes in version 7.2.1 or earlier of SCOS. Starting in version 7.2.10, the limit has been increased to 100 for all SC series except for SC5020, where it is 50 volumes.

Any snapshot method can be used for creating a snapshot of a database that has been shutdown cleanly. The result of the snapshot would be similar to a cold backup in Oracle terminology. However, if the database is open and spans multiple volumes, a snapshot must be created using a consistent snapshot profile while the database is in BEGIN BACKUP mode.

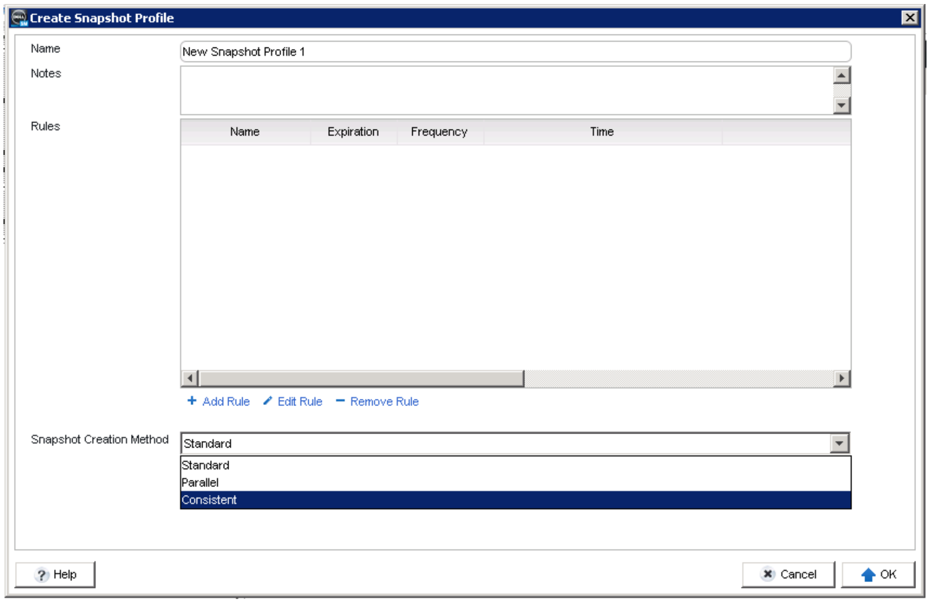
A consistent snapshot profile (consistency group) in part guarantees the atomicity of the database if all the volumes (LUNS) that make up the database are members of the same consistent snapshot profile and the open database is placed in BEGIN BACKUP mode before the snapshot is made. Oracle redo logs should be switched before BEGIN BACKUP and after END BACKUP. Without the combined use of a consistency group and placing the database in BEGIN BACKUP mode, any snapshot created of the opened database cannot be used later for database cloning or database restores because:

- Consistency groups guarantee SC Series volume consistency, but do not guarantee application consistency
- Oracle BEGIN BACKUP guarantees database consistency

Once the consistent snapshot is made, the database can be taken out of BEGIN BACKUP mode.

If the entire database (datafiles, system, sysaux, control files, redo log files, etc.) resides on a single volume and is opened, a standard snapshot of that volume can be made, but the database must still be placed in BEGIN BACKUP mode prior to the creation of the snapshot. The use of a single LUN for the entire database is strongly discouraged, but is mentioned here only to illustrate the significance of using BEGIN BACKUP.

Write-consistency and data integrity across all volumes in a snapshot profile is governed by defining the snapshot profile as a consistent snapshot profile (see Figure 31).



Create Snapshot Profile

Name: New Snapshot Profile 1

Notes:

Rules:

| Name | Expiration | Frequency | Time |
|---|---|---|---|

Snapshot Creation Method: Standard

Figure 31 Consistent snapshot profile

Since a snapshot is taken on a set of volumes, use care when defining the snapshot profile. For example, if there are multiple databases that reside in a set of volumes, and a consistent snapshot is taken of that volume set, it will contain an image of all the databases. If a snapshot of only one of the databases was needed, disk space will be wasted by the other database snapshots. Also, with respect to flexibility in recovering a database without affecting other databases, if the same snapshot was used to recover one of the databases, all the other databases in the snapshot would be recovered to the same point in time (PIT). This could leave the other databases in an unusable state. Therefore, Dell EMC recommends that each Oracle database be placed on its own set of volumes, and in its own consistent snapshot profile. A volume can exist in multiple snapshot profiles, so if there's a need to create a snapshot of the oracle binaries along with the database, the volume containing the Oracle binaries could be placed in each consistent snapshot profile.

When taking snapshots of an Oracle database, if the entire database resides on only one volume, a standard or parallel snapshot profile could be used in lieu of the consistent snapshot profile. However, if an Oracle database spans multiple volumes, a consistent snapshot profile must be used to guarantee write-consistent PITC of the SC volumes used by the database. To simplify the management of snapshot profiles, Dell EMC recommends using the consistent snapshot method for all snapshot profiles that will be used with Oracle databases, regardless of single or multi-volume databases. That way, should a single-volume Oracle database that uses a standard snapshot profile expand to multiple volumes, a consistent snapshot profile does not need to be created. Simply adding the new volume to the existing consistent snapshot profile will suffice. Dell EMC also recommends that each Oracle database be placed on its own set of volumes and in its own consistent snapshot profile.

**Note:** Consistent snapshot profiles provide write-consistency and data integrity across all SC Series volumes in the snapshot. They do not provide write-consistency and data integrity at the application level. For that, Oracle BEGIN/END BACKUP must be used with consistent snapshot profiles. For information on the process of creating a consistent snapshot and using BEGIN/END BACKUP, see 3.22

Expiration times can be set on snapshots. When snapshots expire, the SC Series array automatically releases the space used by the snapshot back to the pagepool.

To create a snapshot profile:

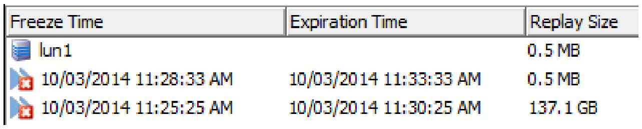
1. In the DSM system tree, right-click **Snapshot Profiles** and select **Create Snapshot Profile**.
2. In drop-down field **Snapshot Creation Method**, select value **Consistent**.
3. Select and set **Timeout Snapshot creation after** seconds and **Expire incomplete Snapshot sets if Snapshot creation timeout is reached** as appropriate.
4. Provide a **Name** and optional **Notes** for the volume.
5. Select an on-demand or scheduled snapshot. On-demand snapshot profiles are useful when creating a backup or database clone from within a shell script that executes from a job schedule. If creating a scheduled snapshot, select **Add Rule**.
6. Specify the type of schedule: **Once**, **Daily**, **Weekly**, or **Monthly**.
7. Specify the date and time, the frequency, and an expiration time for the snapshot. Click **Continue**.

If adding multiple time policies to the snapshot profile, repeat steps 5-7. After all time policies have been added to the schedule, click **OK** and click **Close**.

## 3.21 Snapshots (Replays)

The snapshot (Replay) feature is a licensed technology available on SC Series arrays that provides continuous space-efficient data protection. Snapshots create space-efficient, write-consistent, point-in-time copies (PITC) of one or more volumes based on a collection of rules that exist in a snapshot profile. Snapshots can be used for immediate recovery from data loss caused by hardware failures or logical errors. SC Series snapshot technology is different from other traditional PITCs because when a snapshot is taken data is not copied, it is only frozen. Should a write request occur to a frozen block, a new RAID 10 or RAID 10 DM page is allocated to the volume and the write will occur in that new page.

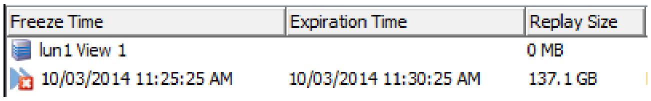
Initial space consumed by a snapshot and corresponding View Volume is much less than the size of the source volume. In Figure 32, the source volume (represented by the snapshot frozen at 11:25:25) is 137.1 GB, while a later snapshot frozen at 11:28:33 is only 0.5 MB in size, a 99 percent reduction in space consumption. Actual space reduction is dependent on each environment.



| Freeze Time | Expiration Time | Replay Size |
|---|---|---|
| lun1 | | 0.5 MB |
| 10/03/2014 11:28:33 AM | 10/03/2014 11:33:33 AM | 0.5 MB |
| 10/03/2014 11:25:25 AM | 10/03/2014 11:30:25 AM | 137.1 GB |

Figure 32 Size of snapshots

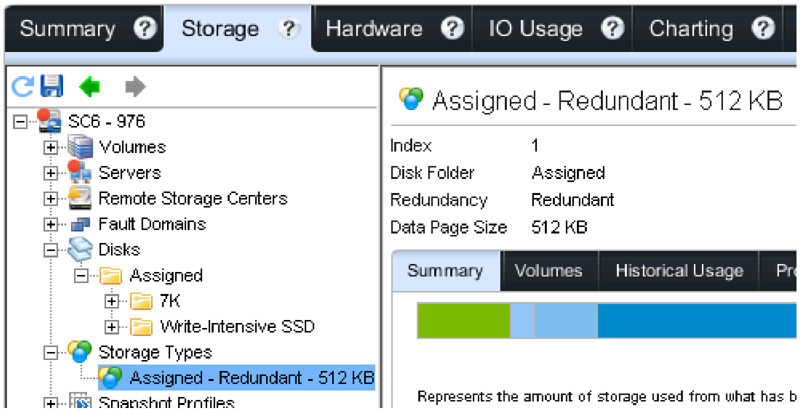
Once a snapshot is created, view volumes can be created from it and presented to a server DSM object. The server can then use the view volume as it would with any other volume. Once the view volume is created, no additional space is allocated to it as long as there are no writes being written to the view volume. See **lun 1 View 1** in Figure 33.



| Freeze Time | Expiration Time | Replay Size |
|---|---|---|
| lun1 View 1 | | 0 MB |
| 10/03/2014 11:25:25 AM | 10/03/2014 11:30:25 AM | 137.1 GB |

Figure 33 Size of View Volume: lun 1 View 1 — zero bytes

If writes are issued against a view volume, SC Series allocates new pages (**Data Page**) from the corresponding storage type and pagepool to the view volume. The size of the new page will be the size defined for the corresponding Storage Type.



Assigned - Redundant - 512 KB

| | |
|---|---|
| Index | 1 |
| Disk Folder | Assigned |
| Redundancy | Redundant |
| Data Page Size | 512 KB |

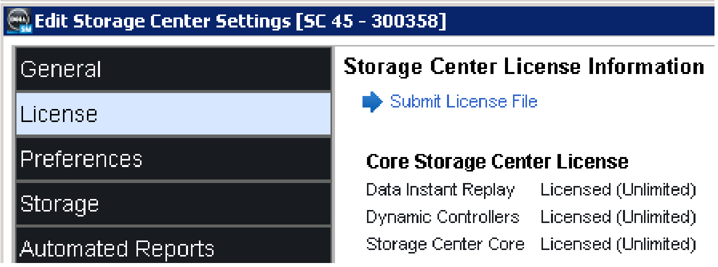
Represents the amount of storage used from what has b

If every Oracle block in a View Volume is written, then the total space required for that View Volume will be the size of the source volume. Therefore, when sizing the array, the number and intended use of snapshots needs to be considered.

Dell EMC strongly recommends Snapshots be licensed in an Oracle environment. It dramatically simplifies both the effort and required resources to implement backup and recovery strategies. It also provides an easy methodology for cloning Oracle databases.

To see if the **Data Instant Replay** (snapshots) feature is licensed, perform the following:

1. Select an SC Series array from the main DSM window.
2. Right-click the array in the navigation tree, and select **Edit Settings**.
3. Select **License.**
4. If **Data Instant Replay** is licensed, it will appear under section **Core Storage Center License.**



Edit Storage Center Settings [SC 45 - 300358]

Storage Center License Information

Core Storage Center License

| | |
|---|---|
| Data Instant Replay | Licensed (Unlimited) |
| Dynamic Controllers | Licensed (Unlimited) |
| Storage Center Core | Licensed (Unlimited) |

Figure 34 Verifying Data Instant Replay license

Dell EMC recommends creating snapshots under these conditions:

- Immediately before a database goes live, or before and after an upgrade, major change, repair, or any major maintenance operation
- Once per day to allow Data Progression to move age-appropriate data more efficiently and effectively to other storage types and tiers
- On a schedule that satisfies appropriate business requirements for recovery point objective (RPO) and recovery time objective (RTO). The longer the time between snapshots, the greater the RPO becomes. For high-performance, business-critical systems, this may be a concern. To mitigate that concern, simply take more frequent snapshots. If the RPO is unknown, a snapshot taken once per day of each Oracle database might be sufficient until the RPO is defined
- Dramatically simplify both the effort and required resources to implement or augment conventional backup and recovery strategies and provide immediate recovery from data loss caused by hardware failures or logical errors to the last known state of any unexpired snapshot
- Provide ability to immediately create a copy of a database
- In an all-flash array, move data from T1 RAID 10 to T1 RAID 5 or RAID 6 using on-demand snapshots, should Oracle require more space on T1 RAID 10
- Provide lower RTO and RPO than conventional backup/recovery strategies
- Provide the ability to replicate a database to a DR location

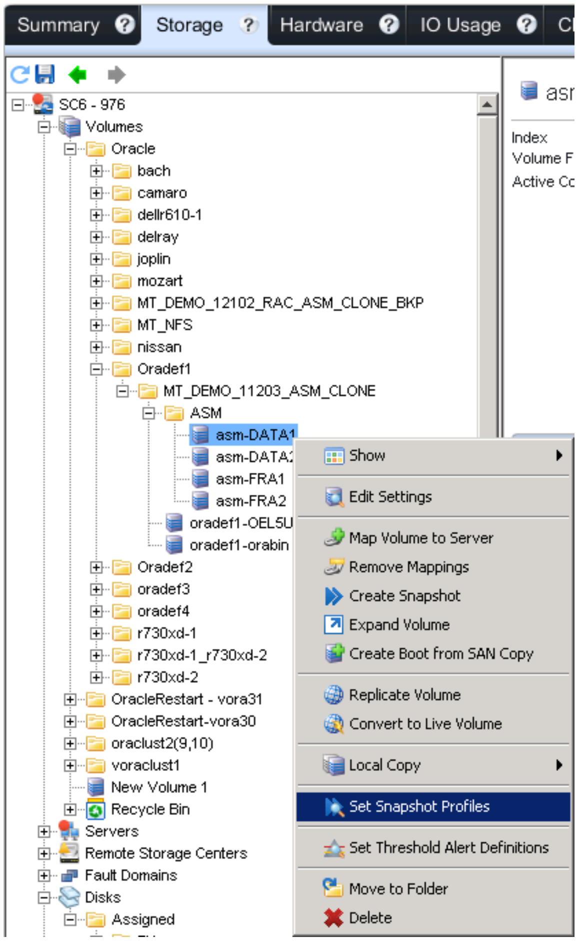
Snapshots can be performed using different methods:

**Standard** instructs SC Series to take snapshots in series for all volumes associated with the snapshot.

**Parallel** instructs SC Series to take snapshots simultaneously for all volumes associated with the snapshot profile.

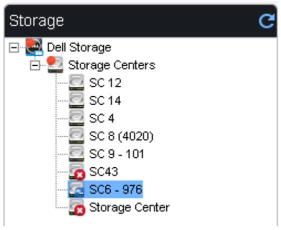
**Consistent** instructs SC Series to halt I/O and take snapshots for all volumes associated with the snapshot. Once the snapshots are created, I/O resumes on all volumes associated with the snapshot. This is the recommended snapshot method for Oracle databases.

To assign a volume to a snapshot profile, see section 3.25. To assign snapshot profiles to a volume after the volume has been created, right-click the volume from the navigation tree and select **Set Snapshot Profiles.**

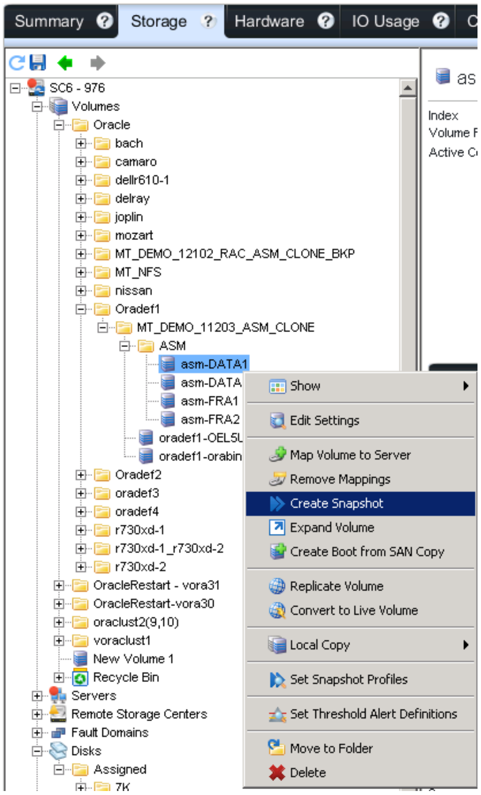


To create a snapshot of an Oracle database, see section 3.22. To create a snapshot of a non-Oracle database volume, follow these steps:

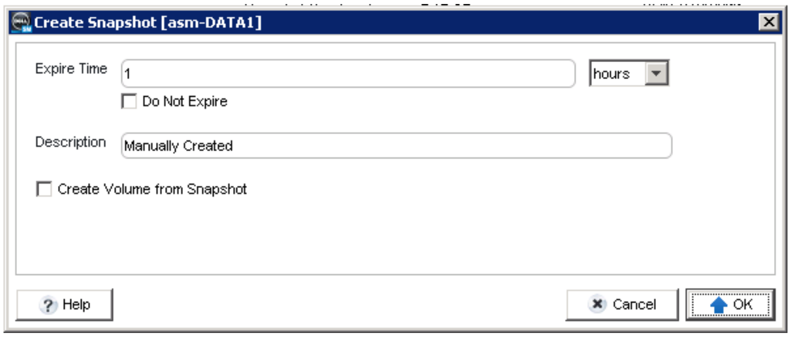
1. Make sure the volume belongs to a snapshot profile.
2. In DSM, select the **Storage** view and select the SC Series array containing the desired volume.



3. Select the **Storage** tab, right-click the desired volume, and select **Create Snapshot**.



4. Follow the remaining instructions in the Create Snapshot wizard.


Create Snapshot [asm-DATA1]

Expire Time: 1 hours

□ Do Not Expire

Description: Manually Created

□ Create Volume from Snapshot

## 3.22 Consistent snapshot profiles and Oracle

Consistent snapshots are recommended in Oracle environments and should be considered under the following conditions:

- Immediately before the database goes live, or before and after an Oracle upgrades, major change, repair, or any major maintenance operations.
- Once per day to allow Data Progression to move age-appropriate data more efficiently and effectively to other storage types and tiers.
- On a schedule that satisfies appropriate business requirements for recovery point objective (RPO) and recovery time objective (RTO); the longer the time between snapshots, the greater the RPO becomes; for high-performance, business-critical systems, this may be a concern; to mitigate that concern, take more frequent snapshots. If the RPO is unknown, a snapshot taken once per day of each Oracle database might be sufficient until the RPO is defined.
- Dramatically simplify both the effort and required resources to implement or augment conventional backup and recovery strategies and provide immediate recovery from data loss caused by hardware failures or logical errors to the last known state of any unexpired snapshot.
- Provide ability to immediately create a database clone.
- Provide the ability to replicate a database to a DR location.

To create a snapshot of an open database, perform the following steps:

1. In Oracle, make sure the database is running in archive log mode.
2. In Oracle, execute:

```
ALTER SYSTEM ARCHIVE LOG CURRENT;
ALTER DATABASE BEGIN BACKUP;
```

3. In SC Series, create a consistent snapshot using the consistent snapshot profile for the database.
4. In Oracle, execute:

```
ALTER DATABASE END BACKUP;
ALTER SYSTEM ARCHIVE LOG CURRENT;
```

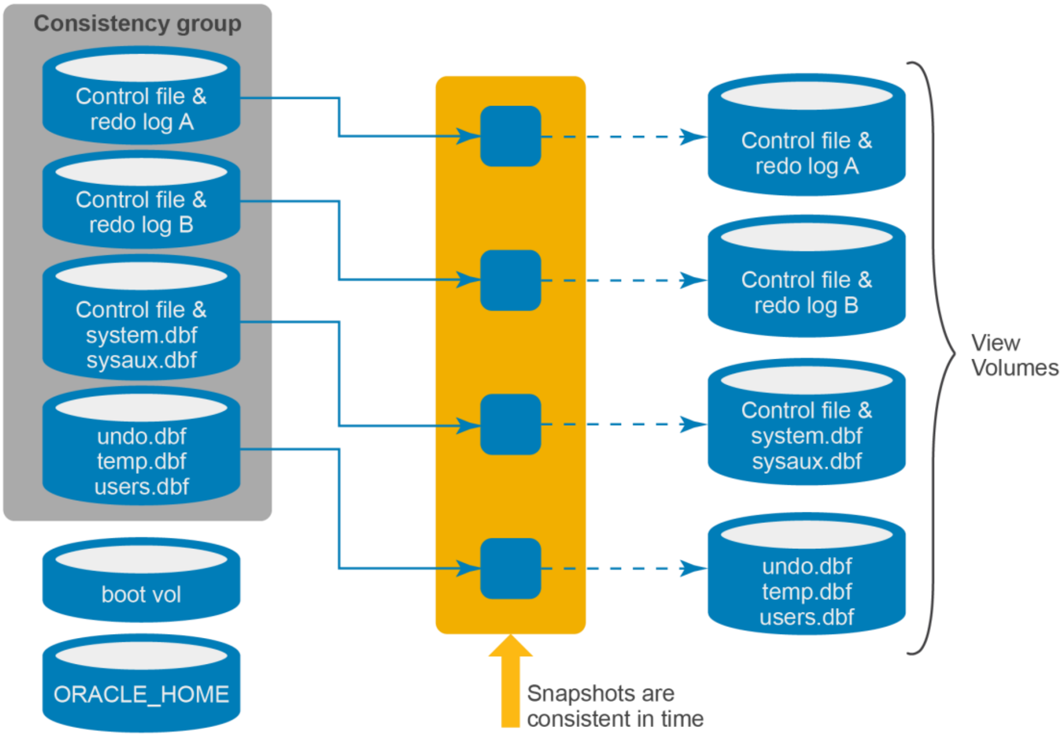
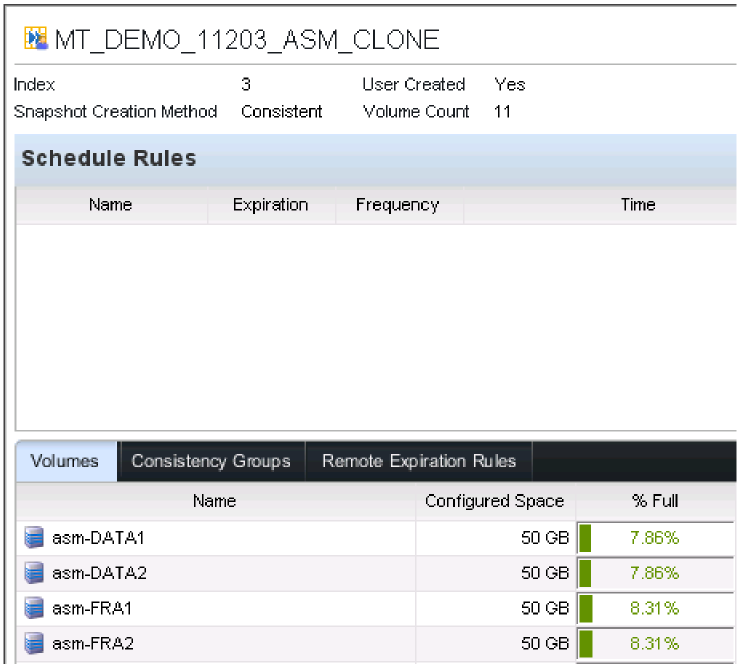

Consistency group
Control file & redo log A
Control file & redo log B
Control file & system.dbf sysaux.dbf
undo.dbf temp.dbf users.dbf
boot vol
ORACLE_HOME
Control file & redo log A
Control file & redo log B
Control file & system.dbf sysaux.dbf
undo.dbf temp.dbf users.dbf
View Volumes
Snapshots are consistent in time

Figure 35 Consistent snapshot


MT_DEMO_11203_ASM_CLONE

| | | | |
|---|---|---|---|
| Index | 3 | User Created | Yes |
| Snapshot Creation Method | Consistent | Volume Count | 11 |

Schedule Rules

| Name | Expiration | Frequency | Time |
|---|---|---|---|

| Name | Configured Space | % Full |
|---|---|---|
| asm-DATA1 | 50 GB | 7.86% |
| asm-DATA2 | 50 GB | 7.86% |
| asm-FRA1 | 50 GB | 8.31% |
| asm-FRA2 | 50 GB | 8.31% |

Figure 36 Consistent snapshot profile containing ASM volumes for one database

## 3.23 Using Data Progression and snapshots

In order to utilize Data Progression effectively, snapshots must be used. The frequency of when the snapshots are created will be dependent on each environment, but a good place to start would be to create them at least once a week on all volumes. With over 2,000 snapshots available in the SC4000, SC5000, SC7000, SC8000, and SC9000 Series arrays, and 2,000 snapshots available in the SCv2000 Series arrays, this allows for more than one snapshot per week on each volume.

Data Progression moves snapshot data according the storage profile used by a volume. If the storage profile, Recommended (All Tiers), is used in an all-flash system, snapshot data is immediately moved to T1 RAID 5-9/RAID 6-10. Writes to the volume still occur to T1 with RAID 10/RAID 10DM. This allows T1 space to be used more efficiently. See sections 3.16 and 3.19 for more information.

Dell EMC also recommends reviewing Data Progression pressure reports periodically to see space consumption in a tier for the previous 30 days. The report will be useful in fine tuning snapshot profiles and retention periods for Oracle databases and capacity planning of the array. See section 3.17 for more information.

## 3.24 SC Series cache settings

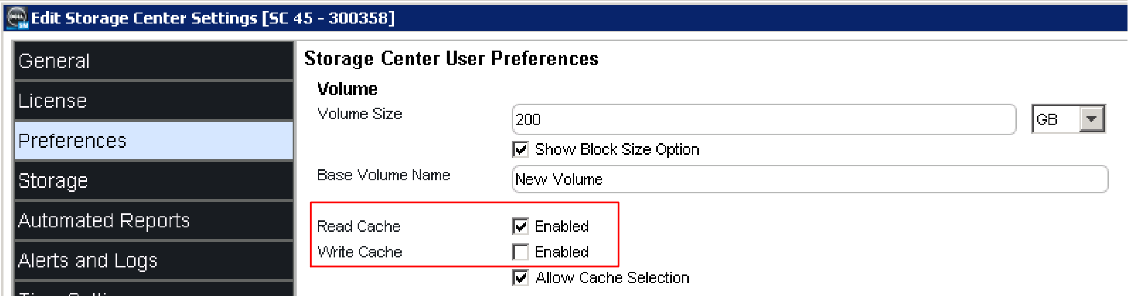
SC Series storage provides both a read-ahead and write cache to improve performance of I/O. Read-ahead cache anticipates the next read and holds it in quick volatile memory, thus improving read performance. Write cache holds written data in volatile memory until it can be safely stored on disk, thereby increasing write performance. By default, both read and write cache are enabled globally.

Because SSDs are memory devices, Dell EMC recommends disabling write cache in an SC Series all-flash or all-flash optimized array to maximize performance. Disabling read cache in an SC Series all-flash or all-flash optimized array may or may not improve performance. In an SC Series hybrid array, or an array of all spinning media, Dell EMC recommends enabling read and write caches.

Prior to changing the state of a cache, Dell EMC recommends testing applications to establish a baseline of performance metrics that can be compared against the same application tests with a different cache state. Because a change in either cache setting can have impact on performance, consult Dell Support in assessing the appropriateness of a change.

To view or change the global cache state, perform the following in DSM:

1. Select the **Storage** view.
2. In the **Storage** navigation tree, right-click a SC Series array and select **Edit Settings**.
3. Select **Preferences**. DSM displays the global cache properties.
4. Select or clear the **Read Cache** or **Write Cache** check boxes to enable or disable read or write cache.
5. Select **OK.**



Storage Center User Preferences

Volume

Volume Size: 200 GB

☑ Show Block Size Option

Base Volume Name: New Volume

Read Cache: ☑ Enabled

Write Cache: □ Enabled

☑ Allow Cache Selection

Figure 37 Read-ahead and write cache global settings

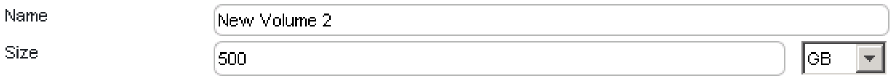
For information on setting cache for a volume, see section 3.27.

## 3.25 Creating volumes (LUNs) in DSM

When creating LUNs for an Oracle database, it is suggested to create an even number of volumes that are distributed evenly between SC Series controllers in a dual controller array. The even LUN distribution across controllers distributes I/O across both controllers.

To create a volume, perform the following in DSM:

1. Select the **Storage** view.
2. From the **Storage** navigation tree, select an SC Series array.
3. Expand **Volumes** from the SC Series navigation tree.
4. Right-click the location within the **Volumes** navigation tree where the volume should reside and select **Create Volume**.
5. Specify the name, size, and units for the volume.



Name: New Volume 2

Size: 500 GB

Oracle recommends using ASM with fewer large volumes rather than many small volumes. Use the Oracle recommendation for sizing as a starting point; the number and sizes of the volumes depend on business requirements and the version of Oracle. When creating volumes for a database, it is recommended to initially create the volumes (datafiles, archived redo logs, and flash recovery area) larger than needed to allow for database growth. Since SC Series has the ability to thinly and dynamically provision storage, disk space is not taken up until the actual data has been written to it. Oracle tablespaces can be created with the AUTOEXTEND parameter so space is only used when needed. Oracle ASM 12c Administrator’s guide recommends the number of LUNs for each ASM disk group be at least equal to four times the number of active I/O paths. For example, if a disk group has two active I/O paths, then a minimum of eight LUNs be used. The LUNs should be of equal size and have the same performance characteristics.

The maximum size of a volume and total storage is dependent on the version of Oracle:

- If the COMPATIBLE.ASM and COMPATIBLE.RDBMS ASM disk group attributes are set to 12.1 or greater:
  - PB maximum storage for each Oracle ASM disk with the allocation unit (AU) size equal to 1 MB
  - 8 PB maximum storage for each Oracle ASM disk with the AU size equal to 2 MB
  - 16 PB maximum storage for each Oracle ASM disk with the AU size equal to 4 MB
  - 32 PB maximum storage for each Oracle ASM disk with the AU size equal to 8 MB
  - 320 exabytes (EB) maximum for the storage system
- If the COMPATIBLE.ASM or COMPATIBLE.RDBMS ASM disk group attribute is set to 11.1 or 11.2:
  - 2 terabytes (TB) maximum storage for each Oracle ASM disk
  - 20 petabytes (PB) maximum for the storage system
- If the COMPATIBLE.ASM or COMPATIBLE.RDBMS ASM disk group attribute is set to less than 11.1, see Oracle documentation for maximum disk size.

As every environment is different, testing needs to be done to determine the optimal number and size of LUNs and the optimal SC Series configuration to ensure performance meets expected business requirements. A number of different performance analysis tools are available, some of which require the candidate database be fully created and populated. It is recommended to work with a Dell EMC storage architect when interpreting test results so the SC Series array can be configured appropriately for the expected system loads.

For a discussion on Oracle ASM and SC series storage, see section 6.

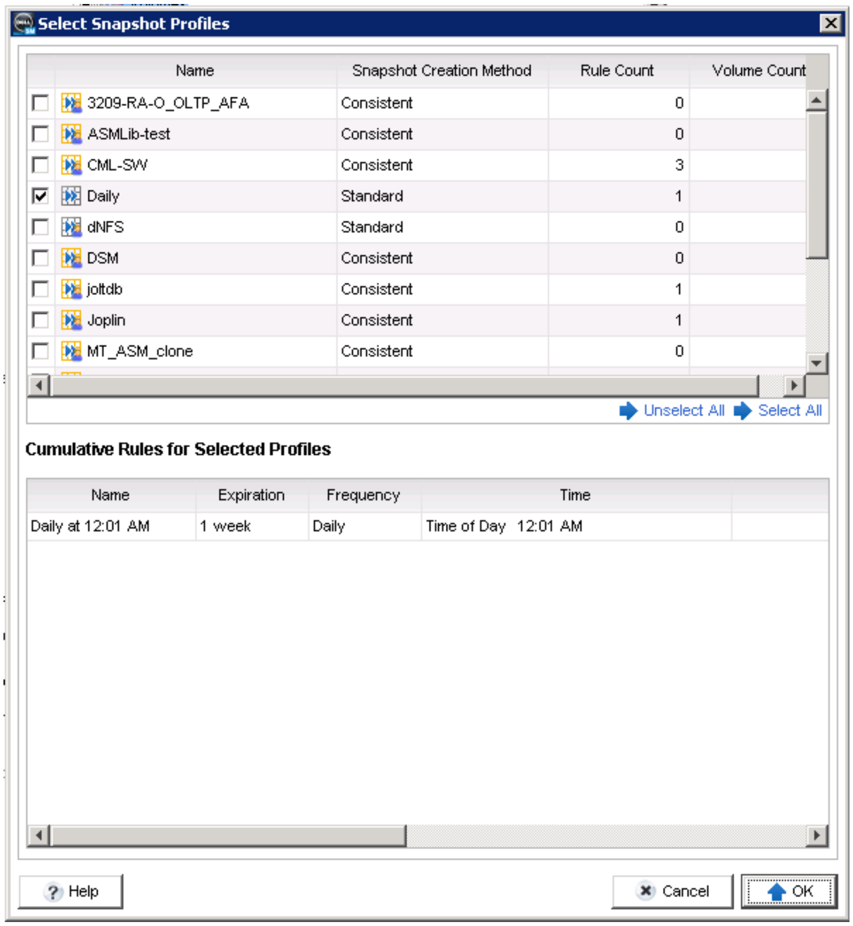
6. Add optional notes tagged to the volume.

Notes

7. Select the **Change** option that corresponds to **Snapshot Profiles** to assign the volume to a snapshot profile.

Snapshot Profiles Daily Change

8. In most Oracle deployments, the database volumes should be removed from the **Daily** profile schedule. However, before removing them, assess the change.



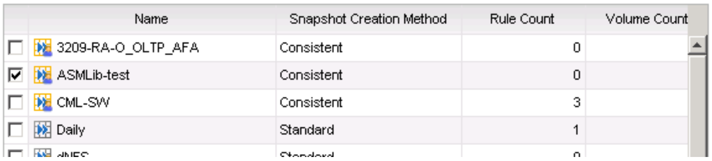
Select Snapshot Profiles

| | Name | Snapshot Creation Method | Rule Count | Volume Count |
|---|---|---|---|---|
| □ | 3209-RA-O_OLTP_AFA | Consistent | 0 | |
| □ | ASMLib-test | Consistent | 0 | |
| □ | CML-SW | Consistent | 3 | |
| ☑ | Daily | Standard | 1 | |
| □ | dNFS | Standard | 0 | |
| □ | DSM | Consistent | 0 | |
| □ | joltdb | Consistent | 1 | |
| □ | Joplin | Consistent | 1 | |
| □ | MT_ASM_clone | Consistent | 0 | |

Cumulative Rules for Selected Profiles

| Name | Expiration | Frequency | Time |
|---|---|---|---|
| Daily at 12:01 AM | 1 week | Daily | Time of Day 12:01 AM |

9. Uncheck **Daily and** select the consistent snapshot profile for the database. Select **OK**.

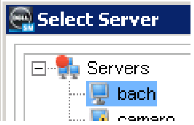


| | Name | Snapshot Creation Method | Rule Count | Volume Count |
|---|---|---|---|---|
| □ | 3209-RA-O_OLTP_AFA | Consistent | 0 | |
| ☑ | ASMLib-test | Consistent | 0 | |
| □ | CML-SW | Consistent | 3 | |
| □ | Daily | Standard | 1 | |
| □ | dNFS | Standard | 0 | |

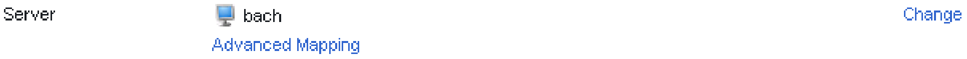
10. To present the volume to a server, select the **Change** option that corresponds to **Server**.

Server
Change

11. Select the server and click **OK**.

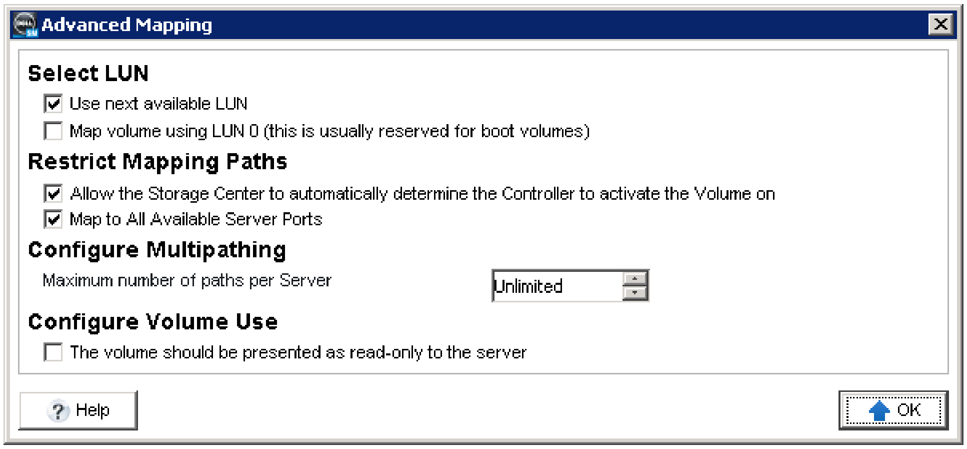


12. Select **Advanced Mapping**.



Server bach Advanced Mapping Change

13. In Oracle deployments, most SC Series volumes should be presented to database servers using multiple paths to avoid single point of failure should a device path fail. The default settings in section **Restrict Mapping Paths** and **Configure Multipathing** of the **Advanced Mapping** wizard were left at default settings to ensure all appropriate mappings would be used and to balance the volume mappings between both controllers.



Advanced Mapping

Select LUN

☑ Use next available LUN

□ Map volume using LUN 0 (this is usually reserved for boot volumes)

Restrict Mapping Paths

☑ Allow the Storage Center to automatically determine the Controller to activate the Volume on

☑ Map to All Available Server Ports

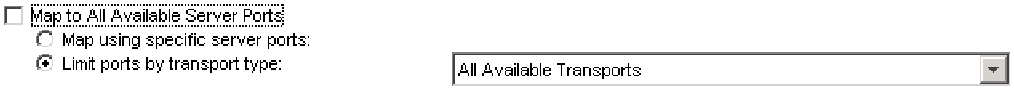
Configure Multipathing

Maximum number of paths per Server: Unlimited

Configure Volume Use

□ The volume should be presented as read-only to the server

14. If there is a need to restrict the mapping to specific server HBAs or transport type, uncheck **Map to All Available Server Ports**.

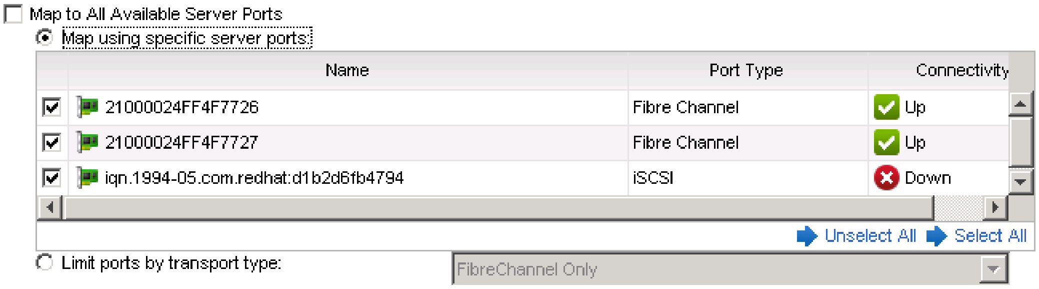


□ Map to All Available Server Ports

○ Map using specific server ports:

● Limit ports by transport type: All Available Transports

15. Select **Map using specific server ports** to restrict server HBA mappings.



□ Map to All Available Server Ports

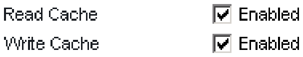
● Map using specific server ports:

| | Name | Port Type | Connectivity |
|---|---|---|---|
| ☑ | 21000024FF4F7726 | Fibre Channel | ✓ Up |
| ☑ | 21000024FF4F7727 | Fibre Channel | ✓ Up |
| ☑ | iqn.1994-05.com.redhat:d1b2d6fb4794 | iSCSI | ✗ Down |

○ Limit ports by transport type: FibreChannel Only

When finished, click **OK**.

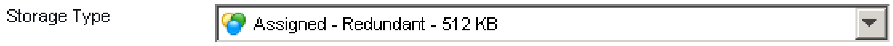
16. Set the appropriate read-ahead and write cache options for the volume.



Read Cache ☑ Enabled

Write Cache ☑ Enabled

17. If multiple pagepools are defined, select the storage type associated with the pagepool from which space should be allocated for the volume.



Storage Type: Assigned - Redundant - 512 KB

18. As appropriate, choose the options to create the volume as a replication, Live Volume, or to preallocate storage, and click **OK**.

After a volume has been created and presented to the SC server object, a device scan must be performed on the physical server before the server can use the volume. Scanning for devices can be accomplished several different ways. See sections 3.25.1 and 3.25.2 for information on scanning devices.

Starting with SCOS 6.2.1, SC presents volumes as having 4 KB physical bocks and 512 byte logical blocks.

```
[root]# multipath -ll
<Snippet>
oraasm-crs1 (36000d3100003d0000000000000000241) dm-4 COMPELNT,Compellent Vol
size=8.0G features='1 queue_if_no_path' hwhandler='0' wp=rw
`-+- policy='round-robin 0' prio=1 status=active
  |- 7:0:2:10 sdj  8:144  active ready running
  |- 7:0:3:10 sdn  8:208  active ready running
  |- 8:0:2:10 sdx  65:112 active ready running
  `- 8:0:3:10 sdab 65:176 active ready running
[root]# blockdev --getss /dev/sdj
512
[root]# blockdev --getpbsz /dev/sdj
4096
```

Figure 38 4KB sector disk in emulation mode

Prior to SCOS 6.2.1, volumes were presented as having 512 byte physical and logical blocks.

See sections 6.1 and 6.2 for more information regarding 4 KB sectors.

### 3.25.1 Scanning devices in PowerPath environments

To scan devices in environments using PowerPath, use command `emcplun_linux`. See section 4.1 for more information.

### 3.25.2 Scanning devices in non-PowerPath environments

The following script provided by Linux package sg3_utils, can be used to scan for devices:

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

`rescan-scsi-bus.sh` requires LUN0 be the first mapped logical unit. If LUNs are mapped for the first time, `rescan-scsi-bus.sh` must be run twice. The first scan adds LUN0. Other LUNS are added during a second scan. Refer to [Red Hat Enterprise Linux 6 Storage Administration Guide]() for additional information on `rescan-scsi-bus.sh` usage. If package sg3_utils is not available, the following can be used:

```
for a in $(ls -1 /sys/class/scsi_host)
do
echo "- - -" > /sys/class/scsi_host/$a/scan
done
```

## 3.26 Preallocating storage

The SC Series array can preallocate storage to a volume when the volume is initially presented to an SC Series server object. This preallocation can be performed if the **Preallocate Storage** option is enabled system-wide in the SC Series array and selected during SC Series volume creation. Preallocating storage causes the SC Series array to take from the pagepool the maximum size specified for the volume, assign it to the volume, and write zero-filled data to the pages. After initial preallocation, should the volume size be increased in the SC Series array, any write that requires a new page be allocated to the volume will cause the array to thinly provision a new non-zero page. Also, should a snapshot be created on the volume, any write to a frozen page will cause the array to thinly provision a new non-zero page to the volume.

Therefore, expanding the size of a preallocated volume, or creating snapshots on the preallocated volume, negates the benefit of preallocating storage. With respect to snapshots, the reason for this is that when a snapshot of a volume is created, all pages (with zero-filled or non-zero data) within the volume are frozen. Should there be an attempt to write to a frozen page, a new page we be thinly provisioned to the volume and the modification will occur on the new page and not on the preallocated zero-filled frozen page. This negates any benefits from preallocation.

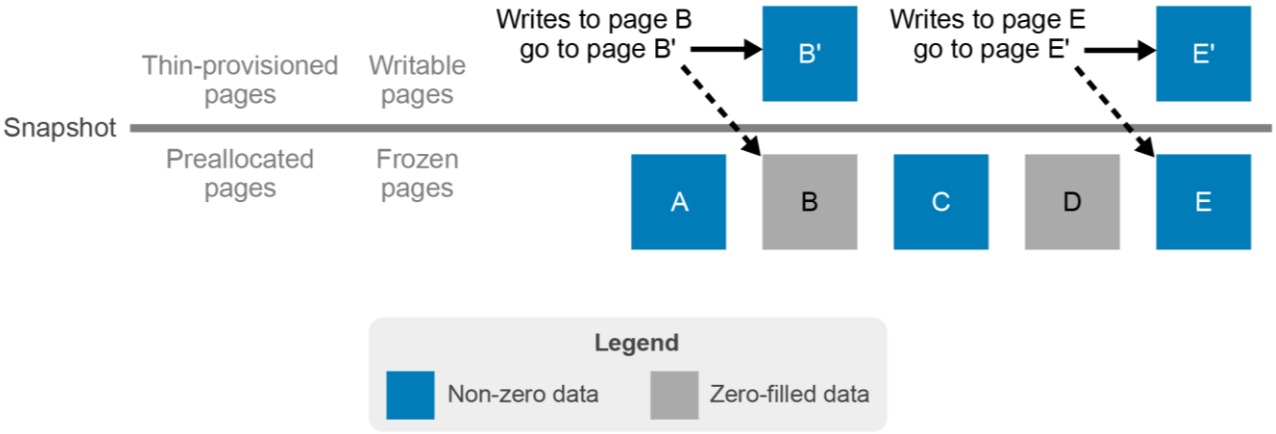


Figure 39 Frozen preallocated pages

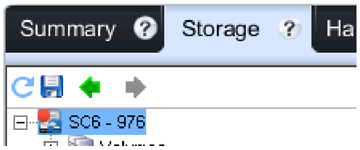
Therefore, for high-performance database systems, where performance can be affected by dynamic block allocation, consider the tradeoff between preallocating and not preallocating space, especially if snapshots are a requirement. Also, consider the tradeoff between expanding or not expanding a preallocated volume. Choosing to preallocate upon volume creation means that applying a snapshot profile to the volume or expanding the volume size at any point in the future will undo the benefit of preallocation. For this reason, Dell EMC recommends that snapshots should never be used on preallocated volumes, and that preallocated volumes should never be expanded.

**Note**: The option to **Preallocate Storage** is only available after the volume is presented to a server, and only if the **Allow Preallocate Storage Selection** option has been enabled in DSM and no data has been written to the volume.

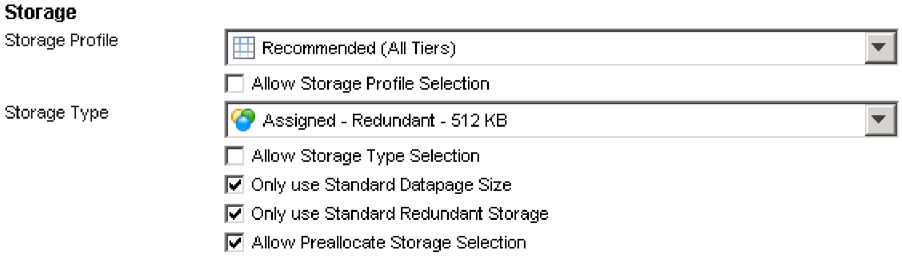
To enable the feature to preallocate space to a volume, perform the following in DSM:

1. Select the **Storage** view.
2. From the **Storage** navigation tree, select an SC Series array.

3. Select the **Storage** tab and in the main navigation tree, right-click the SC Series array.



4. Select **Edit Settings**.
5. Select **Preferences**.
6. Under **Storage > Storage Type**, select the check box, **Allow Preallocate Storage Selection**, and click **OK**.



**Storage**

Storage Profile: Recommended (All Tiers)

☐ Allow Storage Profile Selection

Storage Type: Assigned - Redundant - 512 KB

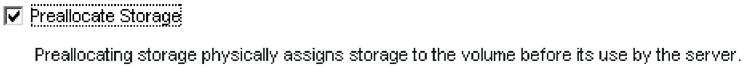
☐ Allow Storage Type Selection

☑ Only use Standard Datapage Size

☑ Only use Standard Redundant Storage

☑ Allow Preallocate Storage Selection

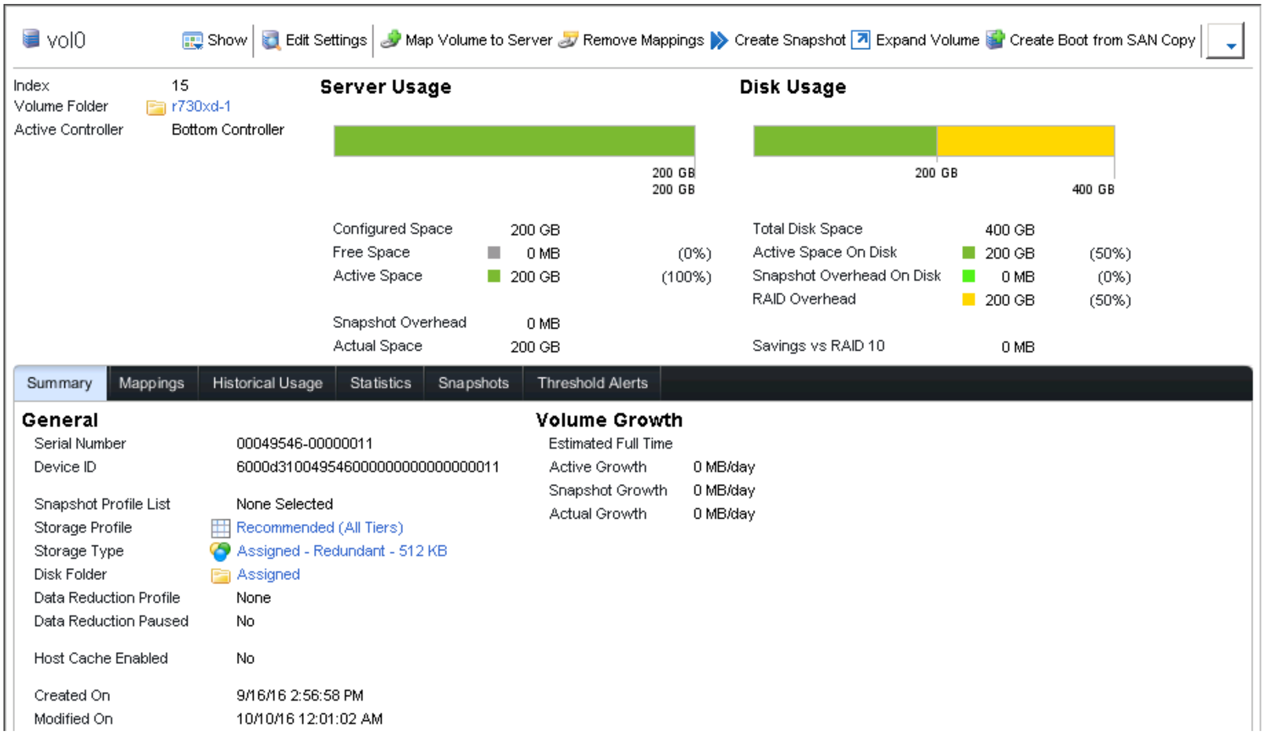
After the feature has been enabled, a volume can be preallocated by selecting check box **Preallocate Storage** after the volume has been presented to a server object and selecting the remaining steps from the **Create Volume** wizard.



☑ Preallocate Storage

Preallocating storage physically assigns storage to the volume before its use by the server.

Figure 40 Selecting preallocation on a volume



| | |
|---|---|
| Index | 15 |
| Volume Folder | r730xd-1 |
| Active Controller | Bottom Controller |

**Server Usage**

| | | | |
|---|---|---|---|
| Configured Space | | 200 GB | |
| Free Space | ■ | 0 MB | (0%) |
| Active Space | 🟩 | 200 GB | (100%) |
| Snapshot Overhead | | 0 MB | |
| Actual Space | | 200 GB | |

**Disk Usage**

| | | | |
|---|---|---|---|
| Total Disk Space | | 400 GB | |
| Active Space On Disk | 🟩 | 200 GB | (50%) |
| Snapshot Overhead On Disk | 🟩 | 0 MB | (0%) |
| RAID Overhead | 🟨 | 200 GB | (50%) |
| Savings vs RAID 10 | | 0 MB | |

**General**

| | |
|---|---|
| Serial Number | 00049546-00000011 |
| Device ID | 6000d310049546000000000000000011 |
| Snapshot Profile List | None Selected |
| Storage Profile | Recommended (All Tiers) |
| Storage Type | Assigned - Redundant - 512 KB |
| Disk Folder | Assigned |
| Data Reduction Profile | None |
| Data Reduction Paused | No |
| Host Cache Enabled | No |
| Created On | 9/16/16 2:56:58 PM |
| Modified On | 10/10/16 12:01:02 AM |

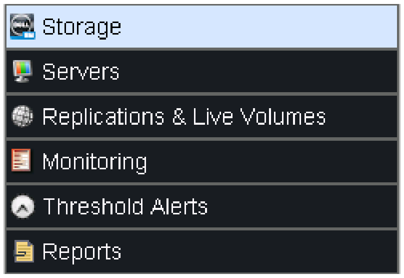
**Volume Growth**

| | |
|---|---|
| Estimated Full Time | |
| Active Growth | 0 MB/day |
| Snapshot Growth | 0 MB/day |
| Actual Growth | 0 MB/day |

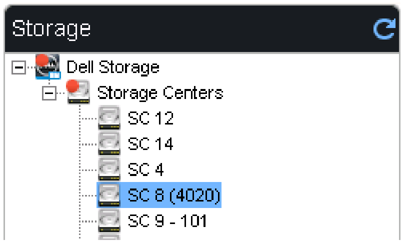
Figure 41 Volume metrics after preallocation

To see the progress of preallocation, perform the following steps to view the SC Series background processes:

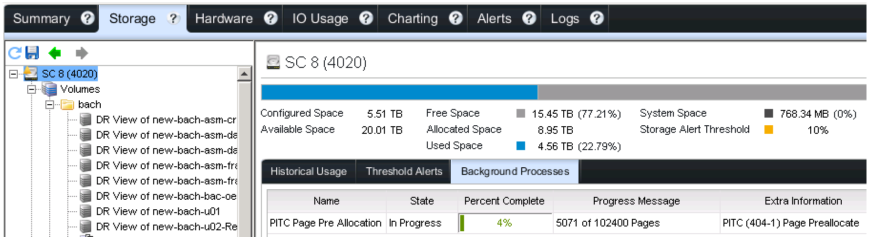
1. Select the **Storage** view.



2. From the **Storage** navigation tree, select an SC Series array.



3. Select tab **Storage** > the array object folder > **Background Processes**.


Summary
Storage
Hardware
IO Usage
Charting
Alerts
Logs
SC 8 (4020)
Volumes
bach
DR View of new-bach-asm-cr
DR View of new-bach-asm-da
DR View of new-bach-asm-da
DR View of new-bach-asm-fra
DR View of new-bach-asm-fra
DR View of new-bach-bac-oe
DR View of new-bach-u01
DR View of new-bach-u02-Re
Historical Usage
Threshold Alerts
Background Processes

SC 8 (4020)

| | | | | | | |
|---|---|---|---|---|---|---|
| Configured Space | 5.51 TB | Free Space | 15.45 TB (77.21%) | System Space | 768.34 MB (0%) |
| Available Space | 20.01 TB | Allocated Space | 8.95 TB | Storage Alert Threshold | 10% |
| | | Used Space | 4.56 TB (22.79%) | | |

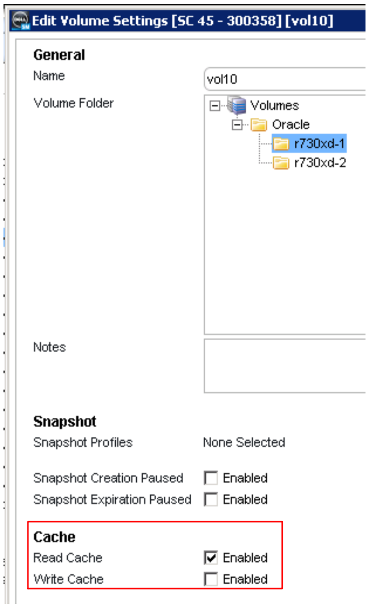
| Name | State | Percent Complete | Progress Message | Extra Information |
|---|---|---|---|---|
| PITC Page Pre Allocation | In Progress | 4% | 5071 of 102400 Pages | PITC (404-1) Page Preallocate |

## 3.27 Volume cache settings

Dell EMC recommends disabling write cache for arrays containing flash. Disabling read-ahead cache in all-flash arrays will be application-specific and may or may not improve performance. In any event, Dell recommends testing applications prior to disabling write or read-ahead cache to establish a baseline of performance metrics which can be compared to the same applications tests after a cache is disabled.

When creating a volume, the volume inherits the global cache setting. Should the volume require a cache setting different than the global cache setting, perform the following:

1. Select the **Storage** view.
2. In the navigation tree, right-click a volume, and select **Edit Settings**.
3. Select or clear the **Read Cache** or **Write Cache** check boxes to enable or disable the read-ahead or write caches and click **OK**.



Edit Volume Settings [SC 45 - 300358] [vol10]

**General**

| | |
|---|---|
| Name | vol10 |
| Volume Folder | |
| Notes | |

**Snapshot**

| | |
|---|---|
| Snapshot Profiles | None Selected |
| Snapshot Creation Paused | ☐ Enabled |
| Snapshot Expiration Paused | ☐ Enabled |

**Cache**

| | |
|---|---|
| Read Cache | ☑ Enabled |
| Write Cache | ☐ Enabled |

Figure 42 Read-ahead and write cache volume settings

## 3.28 Data reduction with deduplication and compression

As the demand for storage capacity increases, so does the pursuit of having the ability to perform data reduction and compression. This can be shown by considering database data that are candidate for archival or perhaps retained in the database for long periods of time and undergoing infrequent changes. When considering storage as a whole instead of at the database level, in most cases, about 80 percent of SAN data will be candidate for archival.

Until the release of SCOS 6.5.1, DBAs and storage administrators had to rely on Oracle and OS means to reduce data on the arrays. Starting with the release of SCOS 6.5.1 in 2014, data compression was included as a native feature that worked in concert with Data Progression on multi-tier systems to compress frozen inaccessible data at the block level on each volume. With the release of SCOS 6.5.10, data compression was extended to include support for SC Series arrays with only one tier of storage. With the release of SCOS 6.7, data compression was further enhanced to extend compression support to include both frozen accessible and frozen inaccessible data, and to support data compression on flash drives in all-flash arrays. The release of SCOS 7.0 introduces data deduplication that works in tandem with compression to further reduce the amount of data stored on an array. Data that remains after deduplication is compressed to reduce the storage footprint even further.

Data reduction can be enabled and disabled on multiple volumes simultaneously, or on a volume-by-volume basis. It can also be paused system-wide, on multiple volumes simultaneously, and volume-by-volume.

While reading reduced data requires fewer resources than actually reducing it, there is some extra load on the controller in real-time from rehydrating the reduced data requested by hosts. When approaching the full I/O potential of the controllers, do not allow the hottest volumes to use data reduction on active pages with flash.

When data reduction runs on an SC Series LUN that has had its data reduced by Oracle or the OS, additional data reduction by SC series array may not be realized on the volume. Dell EMC recommends evaluating if enabling data reduction in both the SC array and Oracle provides additional benefit. Enabling data reduction at the array level provides several benefits:

- The ability to pause all data reduction on the array.
- The ability to offload the server CPU overhead required for data reduction to dedicated processor cores on SC Series arrays.

The amount of data reduction realized is largely dependent of the type of data being reduced. Volumes with files that are text-based typically achieve a higher reduction ratio. Volumes with data consisting of files that are already compressed to some degree (such as audio or video files), typically achieve a lower reduction ratio.

For additional information on data reduction with deduplication and compression feature and supported SC environments, see the *Dell Storage Center OS 7.0 Data Reduction with Deduplication and Compression*.

# 4 Server multipathing

An I/O path generally consists of an initiator port, fabric port, target port, and LUN. Each permutation of this I/O path is considered an independent path. Dynamic multipathing/failover tools aggregate these independent paths into a single logical path. This abstraction provides I/O load balancing across the HBAs, as well as non-disruptive failovers on I/O path failures. The use of dynamic multipath I/O combined with SC Series storage technologies yields the best high availability and performance storage and is recommended for any Oracle environment. To provide dynamic multipath I/O redundancy at the path level in Linux, the server must have at least two active and configured FC initiator ports. See Figure 2 for a server with 4 active FC initiator ports.

Two dynamic multipathing tools that provide this functionality are Dell EMC PowerPath and native Linux Device-Mapper Multipath (DM-Multipath). Either can be used with Oracle. Of the two, PowerPath is more robust. Using both multipathing tools concurrently is not a supported configuration. The remainder of this section will discuss both multipath tools. The single logical path created by the multipathing tool is assigned a pseudo-device name that must be persistent across server reboots. This will be discussed in the remainder of this section and in section 6.

In multipathing environments, the multipath logical device (pseudo-device) must be used in Oracle rather than the Linux device names of each single path. The reason for this are:

- Linux device names, `/dev/sd<a>,` are not consistent across server reboots nor across nodes in a cluster. If they were used in Oracle, they could cause issues,
- DM-Multipath devices `/dev/dm-<n>` are not persistent or consistent across server reboots, or across nodes of an Oracle RAC cluster, and
- Linux sees multiple paths to the same disk, or LUN, and will create an entry in the SCSI device table for each path. This is problematic for ASM, since ASM cannot handle two or more devices mapping to the same LUN. As a result, ASM produces an error. Therefore, in multipath environments, ASM's instance parameter ASM_DISKSTRING must be configured to discover only pseudo-devices. ASM_DISKSTRING and additional ASM configuration will be discussed in sections 6.2, 6.3, and 6.4.

The installation of PowerPath and DM-Multipath are beyond the scope of this document. Refer to Oracle documentation for additional information and recommended best practices when using multipath devices. Single-path devices can also be used by Oracle, but since they are not recommended, they are not covered in this document.

## 4.1 Dell EMC PowerPath

Dell EMC PowerPath pseudo-device names are dynamically assigned during the loading of the HBA driver. Therefore, any changes to the configuration may result in changes in the pre-existing device naming association and render some existing mount table inaccurate if you do not update the mount points to correspond to the new device configuration and its device naming association. Pseudo-devices are created with naming convention `/dev/emcpower<x>[<partition>]`, where `<x>` and `<partition>` are the device and partition indicator (should the device be partitioned):

```
# ls -ltr /dev | grep emc
crw-r--r--   1 root root     10,  58 Mar  9 10:54 emcpower
brw-rw----   1 root disk    120,   0 Mar  9 10:54 emcpowera
<snippet>
brw-rw----   1 root disk    120,  96 Mar  9 10:54 emcpowerg
```

Figure 43 Unpartitioned PowerPath pseudo-device names

PowerPath maintains persistent mappings between pseudo-devices and their corresponding back-end LUNs and records the mapping information in configuration files residing in directory `/etc`. With PowerPath 4.x and higher, mapping configuration exists in several files:

```
/etc/emcp_devicesDB.dat
/etc/emcp_devicesDB.idx
/etc/powermt_custom.xml
```

Figure 44 PowerPath configuration files

Dell EMC recommends saving and exporting pseudo-device and LUN mappings prior to any PowerPath or kernel upgrade or PowerPath reconfiguration:

- To save the PowerPath configuration to the default file `/etc/powermt_custom.xml`, execute:

  ```
  powermt save [-f <filename>.xml]
  ```

- To export the PowerPath configuration to a named file in xml format, execute:

  ```
  emcpadm export_mappings -x -f <filename>.xml
  ```

- To export the PowerPath configuration to a named file in flat text format, execute:

  ```
  emcpadm export_mappings -f <filename>.txt
  ```

After creating volumes in DSM and mapping them to a SC server object, a device scan on the physical server needs to be performed followed by PowerPath configuration. There are a number of ways to accomplish the scan and configuration. The following command will scan all HBAs for new devices and provides the administrator the option to configure and save the new PowerPath configuration:

```
/etc/opt/emcpower/emcplun_linux scan hba
```

To see how PowerPath pseudo-devices are mapped to SCSI devices, execute:

```
/sbin/powermt display dev=all

# powermt display dev=all
<snippet>
Pseudo name=emcpowerg
SC ID=5000D3100003D000 [SC6 - 976]
Logical device ID=6000D3100003D00000000000000001FB [ora-data1]
state=alive; policy=ADaptive; queued-IOs=0
==============================================================================
--------------- Host ---------------   - Stor -  -- I/O Path --   -- Stats ---
###  HW Path               I/O Paths    Interf.  Mode     State   Q-IOs Errors
==============================================================================
   8 qla2xxx                sdaf       976-33    active   alive      0      0
   7 qla2xxx                sdac       976-31    active   alive      0      0
   8 qla2xxx                sdi        976-34    active   alive      0      0
   7 qla2xxx                sdc        976-32    active   alive      0      0
```

Figure 45 Display all Dell EMC PowerPath pseudo-devices

In RAC environments, there may be occasions where PowerPath pseudo devices are not consistent across nodes. If consistent device names are desired, Dell EMC's utility emcpadm can be used to rename the devices. Some emcpadm command options that are helpful in renaming devices are:

- emcpadm getusedpseudos: identifies pseudo devices in use
- emcpadm getfreepseudos: displays available pseudo-device names
- emcpadm renamepseudo: renames a pseudo-device
- emcpadm export_mappings: Saves the current PowerPath configuration
- emcpadm import_mappings: Loads PowerPath configuration from a file
- emcpadm check_mappings: Compares the current mappings with the mappings contained in a file

DELL EMC recommends using the same PowerPath pseudo-device for a shared disks on all nodes of the cluster. For information on consistent pseudo-device names across nodes of a RAC cluster, see section 6.

## 4.2 Linux native DM-Multipath

When DM-Multipath /etc/multipath.conf is configured with:

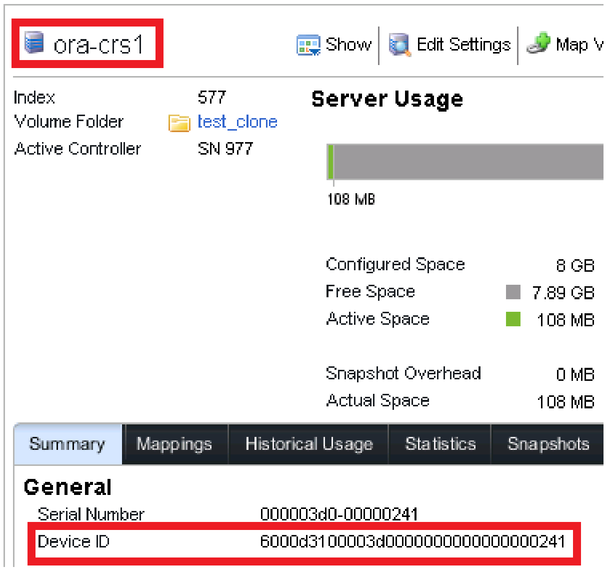
```
defaults {
   user_friendly_names yes
   <Snippit>
}
```

pseudo-devices are created with naming convention /dev/mapper/mpath<device>[<partition>], where <x> and <partition> are the device letter and partition indicator (should the device be partitioned).

```
# ls -ltr /dev/mapper
<snippet>
lrwxrwxrwx 1 root root       7 Apr 14 07:06 mpathe -> ../dm-3
lrwxrwxrwx 1 root root       7 Apr 14 07:06 mpathep1 -> ../dm-8
```

When using DM-Multipath devices for ASM, pseudo-devices names need to be used in Oracle rather than Linux device names /dev/sd<a> or device-mapper devices /dev/dm-<n>. The reason for this is Linux device names and device-mapper devices are not persistent or consistent across server reboots, or across nodes (see Figure 48) of an Oracle RAC cluster. For additional information on consistent pseudo-device names across nodes of a RAC cluster, see section 6.

Pseudo-device names can be changed into something more readable using directive ALIAS in /etc/multipath.conf. DM-Multipath maintains persistent mappings between pseudo-devices and back-end LUNS in configuration file /etc/multipath.conf. Dell EMC recommends backing up /etc/multipath.conf prior to any DM-Multipath reconfiguration. The persistent alias name is associated with a SC volumes Device ID (wwid) (Figure 46).



ora-crs1

Index 577
Volume Folder test_clone
Active Controller SN 977

Server Usage

108 MB

| | | |
|---|---|---|
| Configured Space | | 8 GB |
| Free Space | ▨ | 7.89 GB |
| Active Space | 🟩 | 108 MB |
| Snapshot Overhead | | 0 MB |
| Actual Space | | 108 MB |

General

Serial Number 000003d0-00000241
Device ID 6000d3100003d0000000000000000241

```
defaults {
        user_friendly_names yes
        <Snippet>
}
<Snippet>
multipaths {
        <Snippet>
        multipath {
                wwid 36000d3100003d0000000000000000241
                alias oraasm-crs1
        }
        multipath {
                wwid 36000d3100003d0000000000000000242
                alias oraasm-data1
        }
        multipath {
                wwid 36000d3100003d0000000000000000243
                alias oraasm-data2
        }
        multipath {
                wwid 36000d3100003d0000000000000000244
                alias oraasm-fra1
        }
        multipath {
                wwid 36000d3100003d0000000000000000245
                alias oraasm-fra2
        }
}
```

Figure 46 DM-Multipath devices for ASM

Once `/etc/multipath.conf` has been updated, reload DM-multipath:

```
/etc/init.d/multipathd reload
```

Dell EMC recommends using a naming convention for all DM-Multipath pseudo-device aliases that provides easy device identification and management. The naming convention used in this document includes:

- a prefix (oraasm and oasm) to uniquely identify all pseudo-device intended for ASM,
- ASM disk group name, and
- a suffix to designate an index number for LUNs indented for the same ASM disk group

```
# ls -ltr /dev/mapper/oraasm*[^p][12]
<Snippet>
lrwxrwxrwx 1 root root 7 May 16 12:49 /dev/mapper/oraasm-fra1 -> ../dm-1
lrwxrwxrwx 1 root root 7 May 16 12:49 /dev/mapper/oraasm-data1 -> ../dm-0
lrwxrwxrwx 1 root root 7 May 16 12:49 /dev/mapper/oraasm-data2 -> ../dm-3
lrwxrwxrwx 1 root root 7 May 16 12:49 /dev/mapper/oraasm-crs1 -> ../dm-4
lrwxrwxrwx 1 root root 7 May 16 12:49 /dev/mapper/oraasm-fra2 -> ../dm-6
```

Figure 47 Aliased DM-Multipath pseudo-devices (DM-Names)

To see mappings between DM-Multipath pseudo-devices and SCSI devices, execute:

```
multipath -ll
```

Although it is not a requirement in RAC environments, DELL EMC recommends using the same DM-Multipath alias for a shared disk on all nodes of the cluster. For additional information on consistent pseudo-device names across nodes of a RAC cluster, see section 6.

For additional information on DM-Multipath with ASM, see section 6, information provided at My Oracle Support, Red Hat Enterprise Linux 6 DM multipath documentation, Red Hat Enterprise Linux 7 DM multipath documentation and Dell SC series best practices document for RHEL 6x or 7x.

# 5 Sizing an SC Series array for Oracle

Many of the recommended settings for configuring SC Series arrays for Oracle are mentioned in section 3. This section covers additional information on sizing a SC Series array for Oracle deployments.

In a balanced system, all components from processors to disks are orchestrated to work together to guarantee the maximum possible I/O performance metrics. These are described as follows:

**IOPS** describe the number of reads and writes occurring each second. This metric is used for designing OLTP databases and is key for determining the number of required disks in an array while maintaining accepted response times. If the array uses SSDs, the array will typically provide enough IOPS once capacity and throughput are met.

**Throughput** describes the amount of data in bytes per second transferred between the server and storage array. It is primarily used to define the path between the server and array as well as the number of required drives. A small number of SSDs can often meet IOPS requirements but may not meet throughput requirements. It can be calculated using IOPS and average I/O size:

$$Throughput\ MBs = IOPS * IO\ size$$

**Latency** describes the amount of time an I/O operation takes to complete. High latencies typically indicate an I/O bottleneck.

## 5.1 Configure I/O for bandwidth and not capacity

Storage configurations for a database should be chosen based on I/O bandwidth or throughput, and not necessarily storage capacity. The capacity of individual disk drives is growing faster than the I/O throughput rates provided by those disks, leading to a situation in which a small number of disks can store a large volume of data, but cannot provide the same I/O throughput as a larger number of smaller capacity disks.

## 5.2 Stripe far and wide

The guiding principle in configuring an I/O system for a database is to maximize I/O bandwidth by having multiple disks and channels accessing the data. This can be done by striping the database files. The goal is to ensure each Oracle tablespace is striped across a large number of disks so data can be accessed with the highest possible I/O bandwidth. When using SC Series arrays, striping is accomplished automatically at the storage level. Oracle ASM also provides stripping at the application level.

## 5.3 Random I/O compared to sequential I/O

Because sequential I/O requires less head movement on disk, spinning media can perform more sequential I/O than random I/O. SC Series arrays may cause sequential I/O requests from the server to become large block random I/O on the disks. Since large block random I/O provides comparable performance to sequential I/O, performance will still be good on SC Series arrays. When sizing an SC Series array, assume all I/O is random. This will ensure both random and sequential I/O meets expectations.

## 5.4 OLTP and OLAP/DSS workloads

OLTP systems usually support predefined operations on very specific data, and their workloads generally have small, random I/Os for rapidly changing data. As such, SC Series arrays should be primarily sized on the number of IOPS. For mechanical drives, smaller-capacity, faster spinning drives are able to provide more

IOPS than larger-capacity, slower spinning drives. Since SSDs have no mechanical parts, and are best suited for random I/O, consider using SSDs for best performance in OLTP workloads.

Data warehouses are designed to accommodate ad-hoc queries, OLAP, DSS, and ETL processing. Their workloads generally have large sequential reads. Storage solutions servicing workloads of this type are predominantly sized based on I/O bandwidth or throughput and not capacity or IOPS. When sizing for throughput, the expected throughput of the entire system and each component in the I/O path (CPU cores, HBAs, FC connections, FC switches and ports, disk controllers, and disks) must be known. Then, the entire I/O path between the server and physical disks needs to be sized appropriately to guarantee a balanced use of system resources needed to maximize I/O throughput and provide ability to grow the I/O system without compromising the I/O bandwidth. In some cases, throughput can easily be exceeded when using SSDs, so it is important to understand the characteristics of SSDs and the expected I/O pattern of Oracle.

## 5.5 General sizing recommendation

There are several recommendations and points to consider when sizing:

- Dell EMC recommends evaluating tier sizing (including the RAID write penalty and snapshots) before actually implementing the tier in any environment. When sizing an array, assume all I/O will be random. This will yield best results.
- Before releasing any storage system to production, Dell recommends using the Dell Performance Analysis Collection Kit (DPACK) on a simulated production system during at least a 24-hour period that includes the peak workload. The simulation will help define I/O requirements. It might also be possible to use IOMeter to simulate the production system. After production begins, repeat the analysis on the production system.
- Understand what level of ASM disk group redundancy (external, normal, high) is being considered and how much of the database will reside in each redundancy type. Understand if extended distance clusters are being considered.
- Have a good understanding of the application workloads (OLTP, OLAP, or hybrid).
- For most I/O patterns, SSDs provide considerably better performance than spinning media.
- Understand the tradeoffs between multi- and single-level cell SSDs (MLC, SLC). SLCs are better suited for write-intensive workloads, and MLCs are better suited for read-intensive workloads.
- Understand the required performance metrics of the servers connected to the SC Series array. The IOPS and throughput will help determine the number of disks required in the array, and throughput will help define the paths between the SC Series array and server.
- Accessible data pages on a volume are not eligible to move to a lower tier until they are at least 12 progression cycles old. In databases where tables undergo frequent mass updates or reloads, or frequent index rebuilds, data pages may never reach an age of 12 progression cycles. Therefore, these pages will always reside in T1. In cases like this, T1 needs to be sized for 100 percent of the capacity and support the expected performance metrics of the servers and applications.

Dell recommends factoring in the RAID penalty when determining the number of disks in an SC Series array. If this is not considered, the SC Series array will be undersized. To include IOPS for write penalty, use the following formula:

$$\boldsymbol{Total\ IOPS} = Number\ of\ IOPS + (Number\ of\ IOPS * Write\ Percentage * (RAID\ Penalty - 1))$$

Where:

$$Number\ of\ IOPS = Total\ number\ of\ IOPS\ generated\ by\ the\ server$$

$$Write\ Percentage = Percent\ of\ the\ I/O\ attributed\ to\ writes$$

$$RAID\ Penalty = Number\ of\ I/Os\ required\ to\ complete\ a\ write$$

For example, consider a server performing 20,000 IOPS with a mix of 80 percent reads and 20 percent writes to RAID 10. A total of 24,000 IOPS, (20,000 + (20,000 * 0.20 * (2 -1))), would be required for T1 and the array should be sized with enough disks to support the IOPS.

### 5.5.1 Sizing for multiple tiers

When sizing an SC Series array with multiple tiers, use the following suggestions as a starting point:

**Tier 1** can be sized for 100 percent of total IOPS (including RAID penalty) with at least 20 to 30 percent of the total capacity of all the Oracle databases the array will service. Sizing needs to accommodate storage required by snapshots retained in T1, should the storage profile require it.

**Tier 3** can be sized for 20 percent of total IOPS (including RAID penalty) with at least 70 to 80 percent of the total capacity of all the Oracle databases and snapshots the array will service. Sizing needs to accommodate storage required by snapshots retained in T3, should the storage profile require it.

### 5.5.2 Sizing single tier for an SC Series all-flash array

When sizing an SC Series all-flash array for an Oracle environment, use the following as a starting point:

**Tier 1** should be sized for 100 percent of all the Oracle databases and total expected throughput (and RAID penalty). T1 sizing also needs to include the amount of storage required by snapshots retained in T1 when using default storage profiles: Recommended (All Tiers), High Priority (Tier 1), Write Intensive (Tier 1), or custom storage profiles defined to use T1 for snapshots.

## 5.6 Test the I/O system before implementation

Before creating a database, I/O bandwidth and IOPS should be tested on dedicated components of the I/O path to ensure expected performance is achieved. On most operating systems, this can be done using one of the following:

- Testing acceptable response times with large amounts of data being loaded within a window of time
- Testing acceptable response times with large amounts of data being accessed by queries during peak production times
- Using throughput numbers and experience from an existing identical configured environment

Using the first option, testing could be performed with simple scripts to measure the performance of reading and writing large test files that perform large block sequential reads and writes with large test files using Linux command `dd` or Oracle's ORION. If a dual controller SC Series array is being tested, two large test files should be used with each volume owned by a different controller. The test will verify if all I/O paths are fully functional. If the resulting throughput matches the expected throughput for the components in the I/O path, the paths are set up correctly. Caution should be exercised should the test be run on a live system as the test could cause significant performance issues.

To help define I/O requirements, Dell EMC recommends using DPACK on a simulated production system during at least a 24-hour period that includes the peak workload. In the event it is not possible to simulate the production system, it might be possible to use Iometer to simulate the production system. For other available testing tools, see Table 15.

Table 15 Performance analysis tools:

| Category | Tool | Vendor |
|---|---|---|
| I/O subsystem | DPACK | Dell EMC |
| | ORION | Oracle |
| | iometer | Iometer Org |
| | fio | Freecode and Sourceforge |
| | ioping | Free Software Foundation |
| | dd | Linux OS native |
| RDBMS | SLOB | Kevin Closson |
| | Oracle PL/SQL package<br>DBMS_RESOURCE_MANAGER.CALIBRATE_IO | Oracle |
| Transactional | Benchmark Factory | Quest™ |
| | HammerDB | Open Source |
| | SwingBench | Dominicgiles |
| | Oracle Real Application testing | Oracle |

The performance analysis tools should have the ability to:

- Configure block size
- Specify number of outstanding requests
- Configure test file size
- Configure number of threads
- Support multiple test files
- Not write blocks of zeros, or have an option to override writing zeros

If it is not possible to run DPACK or another testing tool, test the paths between the server and a dedicated SC Series array by running a large block sequential read test using small files (one file per volume, per controller). Tests should use multiple threads and use 512KB blocks and queue depth of 32. This should saturate all paths between the array and server, and verify that all paths are functioning and will yield the I/O potential of the array. If the throughput matches the expected throughput for the number of server HBA ports, the I/O paths between the SC Series array and the server are configured correctly. If the test is run on a SC Series array not dedicated to this test, it could cause significant performance issues. If smaller block sizes are used for the test, as might be required or used by an application that will use the array, IO saturation rates of all paths may not be achievable and therefore the test may not verify that all paths are functioning and yield the IO potential of the array.

Dell EMC recommends repeating this test and validating the process on the production server after go-live to validate and establish a benchmark of initial performance metrics.

Once a design can deliver the expected throughput requirement, additional disks can be added to the storage solution to meet capacity requirements. But the converse is not necessarily true. If a design meets the expected capacity requirements, adding additional disks to the storage solution may not make the design meet the required throughput requirements. This can be illustrated by considering the following. Since the capacity of individual disk drives is growing faster than the I/O throughput rates provided by the disks, a situation can occur where a small number of disks can store a large volume of data, but cannot provide the same I/O throughput as a larger number of smaller disks.

After validating throughput of I/O paths between the SC Series array and the server, and meeting capacity requirements, test the disk I/O capabilities for the designed workload of the SC Series array. This will validate that the storage design provides the required IOPS and throughput with acceptable latencies. The test must not exceed the designed capacity of the array, otherwise the test results will be misleading. If the test does exceed the designed capacity, reduce the number of test threads, outstanding I/Os, or both. Testing random I/O should be done with I/O sizes of 8KB and 16KB as a starting point and adjust from there. When testing sequential I/O, I/O sizes should be 8KB, 16KB, 64KB, or larger.

Dell EMC recommends repeating this test and validating the process on the production server after go-live to validate and establish a benchmark of initial performance metrics.

This testing methodology assumes the guidelines mentioned in previous sections have been followed and modified according for business requirements.

The principle of stripe-far-and-wide needs to be used in tandem with data warehouses to increase the throughput potential of the I/O paths. For information on stripe-far-and-wide and ASM Stripe and Mirror Everything (S.A.M.E) methodology, see sections 5.2 and 6.1 respectfully.

## 5.7 Plan for growth

A plan should exist for future growth of a database. There are many ways to handle growth, and the key consideration is to be able to grow the I/O system without compromising on the I/O bandwidth.

## 5.8 Disk drives and RAID recommendations for Oracle

Table 16 shows some permutations of disk drive types and RAID mixes for Oracle implementations. The information is provided only for illustration and not as a set of rules that govern the actual selection of disk drives and RAID for Oracle file types. Business needs will drive the selection and placement process.

Table 16 Disk drive types and RAID recommendations

| Description | RAID 10 SSD[1] | RAID 10 FC/SAS 15K[1] | RAID 10 FC/SAS 10K | RAID 5 FC/SAS 15K | RAID 5 FC/SAS 10K | RAID 5 FC/SAS 7.2K | All RAID[2] FC/SAS/SATA |
|---|---|---|---|---|---|---|---|
| Data files[3] | Recommended | Recommended | OK | DP[4] | DP[4] | DP[4] | DP[4] |
| Control files | Recommended | Recommended | OK | Avoid | Avoid | Avoid | Avoid |
| Redo logs | Recommended | Recommended | OK | Avoid | Avoid | Avoid | Avoid |
| Archived redo logs | Recommended | Recommended | OK | Not required | Not required | Not required | DP[4] |
| Flashback recovery[5] | Recommended | Recommended | OK | Not required | Not required | Not required | DP[4] |
| OCR files / voting disks | Recommended | Recommended | OK | Avoid | Avoid | Avoid | Avoid |

[1] Dell EMC recommends using SSDs whenever possible for Oracle environments. In the event SSDs are not available, 15K HDDs are recommended.

[2] RAID 6 is available for 900+ GB drives.

[3] Data files include: system, sysaux, data, indexes, temp, undo.

[4] Use single redundancy (RAID 5) for Data Progression (DP) unless business needs justify using dual redundancy (RAID 6, or RAID 10 DM). A Dell EMC storage architect can assist in configuring the correct redundancy for the environment.

[5] Flashback recovery assumes that ASM is used. If not, flashback recovery becomes the location of database disk image backups, archived redo logs, and other infrequently accessed Oracle database files. In a hybrid array (SSDs and HDD), place the flashback recovery area on RAID 10 FC/SAS 15K media after performing an analysis of the array and business requirements.

# 6 SC Series array and Oracle storage management

Two Oracle features to consider for managing disks are Oracle Managed Files and ASM, a feature introduced with Oracle 10g. Without these features, a database administrator must manage database files in native file systems or as raw devices. This could lead to managing hundreds or even thousands of files. Oracle Managed Files simplifies the administration of a database by providing functionality to automatically create and manage files. ASM provides additional functionality for managing not only files but also the disks. ASM handles the tasks of striping and providing disk redundancy at the operating system level, including rebalancing the database files when new disks are added or removed from the system. The use of dynamic multipath I/O combined with ASM and SC Series arrays yields high availability and performance for a database server. ASM is the recommended and the preferred storage solution of Oracle.

When preparing disks for ASM, several different storage resource can be used:

- Logical unit numbers (LUNs)

  LUNs are presented to the Linux server by SC series storage. Oracle recommends that hardware RAID functionality is used within a LUN.

- Entire raw disk or partition of the raw disk

  A disk partition can be the entire disk or part of the disk. However, the partition used by ASM cannot be the partition that contains the partition table because Oracle would overwrite the partition table. With SC series storage, raw disk or partitions of raw disk are created from LUNs.

- Logical volumes (LVM)

  LVMs are typically used in less complicated environments and are mapped to a LUN or used raw disk or partitions of a raw disk. As stated in Oracle's ASM administrator's guide, Oracle does not recommend LVMs because they create duplication of functionality, and adds a layer of complexity that is unnecessary with Oracle ASM. Also, Oracle does not recommend using LVMs for mirroring as ASM provides mirroring. Since LVMs are not recommended by Oracle, they are not covered in this document.

- NFS files

  ASM disk groups for database files can be created from NFS and Oracle Direct NFS (dNFS). NFS or dNFS files are not supported for Oracle Clusterware files. NFS files are available with the Dell EMC FS8600 and FS7600 NAS appliances and are not covered in this document. See the document *Running Oracle over NFS with FS8600 scale-out file system* on Dell TechCenter for additional information.

## 6.1 ASM

After SC volumes are presented to a SC server object, the corresponding physical server can discover the volumes as devices and condition them if necessary. ASM can then be used to define pools of storage from these devices. These storage pools are called ASM disk groups and are comparable to LVM volume groups. Once a disk group is created, the Oracle kernel manages naming and placement of the database files in the disk group. ASM administrators can change the storage allocation (adding or removing disks) in disk groups with SQL commands `create diskgroup`, `alter diskgroup`, and `drop diskgroup`. ASM administrators can also manage disk groups with Oracle Enterprise Manager (EM), ASM command-line utility (asmcmd), ASM Configuration Assistant (asmca), and Database Configuration Assistant (dbca).

For improved I/O bandwidth, create disk groups with an even number of LUNs having the same performance characteristics and capacity, with LUNs evenly distributed between the dual SC controllers. Under normal conditions, the LUNs should also belong to the same SC storage type, SC storage profile, and have the same SC volume characteristics. An even number of LUNS per disk group also allows the database server to take advantage of ASM stripe and Mirror Everything (S.A.M.E) methodology, and allows throughput and IOPS be distributed between dual SC controllers. If two LUNS per disk group do not satisfy database capacity requirements, consider initially creating larger LUNs. Dell EMC recommends creating disk groups with fewer larger volumes rather than many small volumes. For information on maximum LUN sizes for ASM see section 3.25. In Oracle's ASM Administrator's guides, Oracle recommends the number of LUNs per ASM disk group be at least equal to four times the number of active I/O paths. This may not be necessary for all environments and should be evaluated before implementing.

For LUNs intended for ASM, there are several tools, at the OS level, that can be used to manage them and provide consistent device naming and permission persistency: ASMLib, ASM Filter Driver (ASMFD), and UDEV. Each will be discussed in sections 6.2, 6.3, and 6.4 respectfully.

Starting with Oracle 11g, ASM disks are owned by the Grid Infrastructure (GI) owner which belongs to system group OSASM. Any user belonging to group OSASM, can read and write ASM disks. In some cases, this can be a concern and can lead to a data security breach and possible accidental data corruption of the ASM disks. In situations like this, ASMFD can be used to mitigate these concerns from occurring. In the 12c ASM Administrator's guide, Oracle recommends using ASMFD.

Table 17 Benefits of using ASM

| Benefit | Description |
|---|---|
| Simplifies storage administration | ASM is essentially a volume manager and integrated cluster files system, so it eliminates the need for third-party volume managers and file systems. Additionally, there is no need to specify and manage filenames. Wherever a file is traditionally specified in Oracle, an ASM disk group can be specified instead. Every new file automatically gets a new unique name. This prevents using the same filename in two different databases. Disk group naming avoids using two different names for the same file. |
| Provides storage reliability features and database availability | A reliability policy is applied on a file basis, rather than on a volume basis. Hence, the same disk group can contain a combination of files protected by mirroring, parity, or not protected at all. |
| Improved and predictable performance | ASM maximizes performance by automatically distributing database files across all disks in a disk group while removing the need for manual I/O tuning (spreading out the database files to avoid hotspots). It has the performance of raw disk I/O without the inconvenience of managing raw disks. It also eliminates the overhead at the file system layer. Unlike logical volume managers, ASM maintenance operations do not require that the database be shut down. This allows adding or dropping disks while the disks are in use. |
| Improved storage flexibility and scalability | ASM helps DBAs manage a dynamic database environment by allowing them to increase or decrease storage allocation without a database outage. |

ASM requires its own dedicated instance for the purpose of managing disk groups. When Oracle RAC is deployed, an ASM instance must exist on each node in the RAC cluster.

### 6.1.1 ASM instance initialization parameter ASM_DISKSTRING

An important ASM instance initialization parameter is ASM_DISKSTRING. It specifies an operating system-dependent pathname used by ASM to restrict the set of disks considered for discovery by ASM, providing ASM has permission to open the pathname and read the disks.

If ASM disks are solely managed by either ASMFD or ASMLib, let ASM_DISKSTRING assume the default value (a NULL string). However, if UDEV is used to manage ASM disks, set ASM_DISKSTRING to the location of the candidate ASM disks as determined by UDEV rules. The following are possible values if UDEV is solely used with either DM-Multipath or PowerPath pseudo-devices:

| | |
|---|---|
| UDEV and Powerpath: | `ASM_DISKSTRING='/dev/emcpower*'` |
| UDEV and DM-Multipath: | `ASM_DISKSTRING='/dev/mapper/*'` |

In a RAC environment, each ASM instance can have a different value of ASM_DISKSTRING, providing the value allows ASM to discover the same shared ASM disks. Figure 48 presents a shared DM-Multipath partitioned pseudo-device having a different name on two nodes of the same cluster. Both pseudo-devices have the same World Wide Identifier name (WWID 36000d3100003d0000000000000000241). So even though the pseudo-device names (`oraasm-crs1p1` and `oasm-crsp1`) are different, they are the same disk.

```
[node1]# ls -ltr /dev/mapper/ora*crs*p1
lrwxrwxrwx 1 root root 8 Apr 14 15:17 /dev/mapper/oraasm-crs1p1 -> ../dm-14
[node2]# multipath -ll | grep crs
oraasm-crs1 (36000d3100003d0000000000000000241) dm-4 COMPELNT,Compellent Vol


[node2]# ls -ltr /dev/mapper/oa*crs*p1
lrwxrwxrwx 1 root root 7 Apr 14 15:16 /dev/mapper/oasm-crsp1 -> ../dm-1
[node2]# multipath -ll | grep crs
oasm-crs1 (36000d3100003d0000000000000000241) dm-1 COMPELNT,Compellent Vol
```

Figure 48 Different DM-Multipath pseudo-devices referencing the same partitioned disk device:

Figure 49 illustrates this shared device being used in ASM disk group CRS:

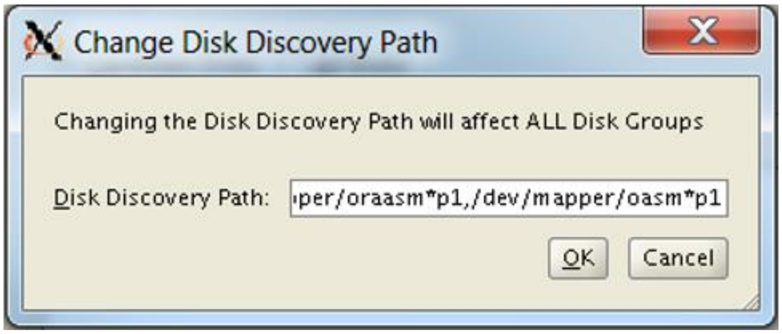
```
INSTANCE_N HOST_NAME  DGNAME           DSKNAME      PATH
---------- ---------- ---------------- ------------ -------------------------
+ASM1      node1      CRS              CRS_0000     /dev/mapper/oraasm-crs1p1

INSTANCE_N HOST_NAME  DGNAME           DSKNAME      PATH
---------- ---------- ---------------- ------------ -------------------------
+ASM2      node2      CRS              CRS_0000     /dev/mapper/oasm-crs1p1
```

Figure 49 ASM using a shared disk having different pseudo-device names in a cluster

When using UDEV in a RAC environment, it is recommended that ASM_DISKSTRING be set to the location and names of the shared device on all nodes. Here is an example of a possible setting when using different pseudo-devices on a two node cluster:

```
ASM_DISKSTRING='/dev/mapper/oraasm*p1','/dev/mapper/oasm*p1'
```

From an Oracle GUI:



Change Disk Discovery Path

Changing the Disk Discovery Path will affect ALL Disk Groups

Disk Discovery Path: per/oraasm*p1,/dev/mapper/oasm*p1

A simpler Disk Discovery Path in this example is:

```
ASM_DISKSTRING='/dev/mapper/o*p1'
```

If PowerPath were used with partitioned devices, the Disk Discovery Path could be set to:

```
ASM_DISKSTRING='/dev/emcpower*1'
```

If ASM_DISKSTRING is not set to identify the shared disks on all nodes of the cluster, issues will arise during at least the installation of GI on the non-primary cluster nodes:

```
Disk Group CRS creation failed with the following message:
ORA-15018: diskgroup cannot be created
ORA-15031: disk specification '/dev/mapper/oraasm-crs1p1' matches no disks


Configuration of ASM ... failed
```

### 6.1.2 Shared device names – RAC environment

Although ASM has no difficulty identifying shared disks having different names on nodes of the same cluster, not all RAC components share that ability.

If a device does not share the same name across all nodes of a cluster, issues could arise. One such condition occurs when installing or cloning 11g Grid Infrastructure (GI). Oracle records the device location and name used for the CRS disk group from the node performing the install in file `${GI_HOME}/crs/install/crsconfig_params` on all nodes (Figure 50):

```
[node2]$ grep ASM_DIS /u01/app/11.2.0/grid/crs/install/crsconfig_params
ASM_DISK_GROUP=CRS
ASM_DISCOVERY_STRING=/dev/mapper/oraasm*p1,/dev/mapper/oasm*p1
ASM_DISKS=/dev/mapper/oraasm-crs1p1
```

Figure 50 Node 2’s crsconfig_params references the shared LUN (ASM_DISKS) from node 1

Therefore, although it is not a requirement to use the same device name for shared block devices across all nodes of a cluster, it may be beneficial to use them. Some reasons to consider using consistent names are:

- Ensure the voting disk can be accessed on all nodes. It is possible to use different asm_disksting per nodes, but it is not a commonly used practice.
- Easier device management and administration when having a large number of disks from block devices, or
- Ensures all components of Oracle can identify, record and reference the same block device on all nodes.

### 6.1.3 Using unpartitioned or partitioned devices

The choice to use either the entire disk device or a single partition that spans the entire disk device in ASM is dependent upon the environment and Oracle requirements. This also applies to the pseudo-device in a multipath environment. For example, if ASMLib is used in a multipath environment, only partitioned pseudo-devices can be used. If ASMFD or UDEV is used to manage device persistence, either the entire pseudo-device or a partitioned pseudo-device can be used. Some reasons to use unpartitioned or partitioned devices or pseudo-devices are:

- Partitions provide a way to identify a device or pseudo-device with fdisk. If the disk can be identified, it could be assumed that the disk is in use, and therefore prevent accidental reuse of the disk. Having a partition provides a method to track a device using fdisk, and prevents accidental use of a disk.
- Not using a partition allows for easier administration and management of the volume if there is a need to change the geometry of the existing ASM disk.
- Using partitioned devices or pseudo-devices will introduce misalignment issues. With unpartitioned devices or pseudo-devices, everything will be aligned.
- Both 11g and 12c Grid Infrastructure install and database installation guides say to use a single partition that spans the entire device or pseudo-device.

**Note:** Dell EMC recommends using partitioned devices. However, since partitioned devices may not be suitable in all environments, evaluate their need and use them accordingly.

If using partitions, Oracle recommends not specify multiple partitions on a single physical disk as a disk group device as Oracle ASM expects each disk group device be on a separate physical disk. The installation guides also say to use either fdisk or parted to create a single whole-disk partition on each disk device for ASM.

Partitioning the device will not impact performance as Oracle uses I/O_direct in Linux, which bypasses the kernel FS buffers to access the database. See section 6.2.1 and 6.2.2 for examples of partitioning a PowerPath or DM-Multipath pseudo-device.

### 6.1.4 Creating and aligning disk partitions on ASM disks

If using partitioned devices for ASM, create the partition so it starts on the boundary of a 4KB physical block. In general, offsets should start on multiples of 2, equal to or greater than 64KB. Oracle recommends using an offset of 1MB (first sector = 2048 when sectors are 512 bytes) as stated in their best practices and starter kits. The following example uses fdisk to create a partition with an offset of 1MB. Larger offsets can be used too, especially for larger LUNs. For example, if partitions are aligned at 16MB (first sector = 32768, when sectors are 512 bytes), I/O will be aligned for stripe widths up to 16MB. For volumes with a size of 2TB or larger, `parted` must be used because `fdisk` cannot handle volumes larger than 2TB.

```
# fdisk /dev/mapper/oraasm-data1

WARNING: DOS-compatible mode is deprecated. It's strongly recommended to
         switch off the mode (command 'c') and change display units to
         sectors (command 'u').

Command (m for help): u
Changing display/entry units to sectors
Command (m for help): n
Command action
   e   extended
   p   primary partition (1-4)
p
Partition number (1-4): 1
First sector (61-1048575, default 61): 2048
Last sector or +size or +sizeM or +sizeK (2048-1048575, default 1048575):
Using default value 1048575
Command (m for help): w
The partition table has been altered!

Calling ioctl() to re-read partition table.
Syncing disks.
# partprobe /dev/mapper/oraasm-data1
# ls -ltr /dev/mapper/oraasm-data1p1
lrwxrwxrwx 1 root root 7 Apr 14 11:56 /dev/mapper/oraasm-data1p1 -> ../dm-13
```

For additional information on aligning disk partitions, see references in appendix B.1

### 6.1.5 ASM disks and Linux I/O scheduler

For optimal Oracle ASM performance, Oracle recommends the Linux I/O scheduler be set to `deadline` on the devices targeted for ASM. This section assumes DM-Multipath devices, but can be applied to PowerPath devices with little change.

To see the current I/O schedule of a DM-Multipath device, display the schedule of the corresponding device-mapper (dm-) device:

```
[root]# ls -ltr /dev/mapper |grep dm-[01346]$
lrwxrwxrwx 1 root root       7 May 16 12:49 oraasm-fra1 -> ../dm-1
lrwxrwxrwx 1 root root       7 May 16 12:49 oraasm-data1 -> ../dm-0
lrwxrwxrwx 1 root root       7 May 16 12:49 oraasm-data2 -> ../dm-3
lrwxrwxrwx 1 root root       7 May 16 12:49 oraasm-crs1 -> ../dm-4
lrwxrwxrwx 1 root root       7 May 16 12:49 oraasm-fra2 -> ../dm-6
[root]# cat /sys/block/dm-[01346]/queue/scheduler
noop [deadline] cfq
noop [deadline] cfq
noop [deadline] cfq
noop [deadline] cfq
noop [deadline] cfq
```

If `[deadline]` appears in the output, it signifies the schedule default value is deadline. If the IO scheduler is not set to deadline, use UDEV to set the schedule.

```
cat /etc/udev/rules.d/60-oracle-schedulers.rules
ACTION=="add|change", KERNEL=="dm-
*",ENV{DM_NAME}=="oraasm*",ATTR{queue/rotational}=="0",ATTR{queue/scheduler}="de
adline"
```

**Note:** The rule begins with ACTION and occupies one line in the rule file. Because the margins of this document do not support the length of the rule, the above rule wraps to a second and third line.

Load the rule in the kernel and activate it.

```
/sbin/udevadm control --reload-rules
/sbin/udevadm trigger --type=devices --action=change
```

On clustered environments, copy this rule file to all nodes of the cluster and activate the rule on each node.

### 6.1.6 512 byte or 4KB sector disks with ASM

Historically, disks offered lower capacities and a physical and logical sector size of 512 bytes. During that same time, ASM and ASMLib were introduced with Oracle 10g R1.

Over the last decade, the industry has been moving toward disks with greater storage capacity and 4KB sector sizes. In anticipation of the move to 4KB drives, Oracle 11g and ASM were designed to detect disk sector size. Prior to this, ASM was designed to work with 512 byte sector disks.

Starting with SCOS 6.2.1, when Dell EMC SC Series presents a volume to a physical Linux server, the volume is presented with 4KB physical bocks and 512 byte logical blocks (4KB emulation mode).

```
[root]# multipath -ll
<Snippet>
oraasm-crs1 (36000d3100003d000000000000000241) dm-4 COMPELNT,Compellent Vol
size=8.0G features='1 queue_if_no_path' hwhandler='0' wp=rw
`-+- policy='round-robin 0' prio=1 status=active
  |- 7:0:2:10 sdj  8:144  active ready running
  |- 7:0:3:10 sdn  8:208  active ready running
  |- 8:0:2:10 sdx  65:112 active ready running
  `- 8:0:3:10 sdab 65:176 active ready running
[root]# blockdev --getss /dev/sdj
512
[root]# blockdev --getpbsz /dev/sdj
4096
```

Figure 51 4KB sector disk in emulation mode

When running 4KB emulation mode with Oracle 11g or 12c, Oracle recommends 512 byte or 4KB block size for redo logs, and 4KB or larger for db_block_size for datafiles.

Also, when using ASMLib to manage 11g or 12g ASM disks, Oracle recommends configuring ASMLib to identify and use the correct sector size of the disk. By default, Oracle will create diskgroups with a sector size of 4KB. Unless the correct set of patches are applied to Oracle, this will cause issues when using Oracle DBCA and OUI for building ASM and databases.

```
ORA-01078: failure in processing system parameters
ORA-15081: failed to submit an I/O operation to a disk
ORA-27091: unable to queue I/O
ORA-17507: I/O request size 512 is not a multiple of logical block size
```

These issues can be mitigated by applying all necessary Oracle patches and configuring ASMLib as necessary, using advanced cookbooks to manually build the Oracle environment, or avoiding using ASMLib.

See section 6.2 for more information.

## 6.2 ASMLib

To configure ASMLib, execute the following from a super user account:

```
/etc/init.d/oracleasm configure
```

ASMLib configuration settings will be recorded in `/etc/sysconfig/oracleasm`. After ASMLib configuration, three additional ASMLib directives need to be added to `/etc/sysconfig/oracleasm`.

- **ORACLEASM_SCANORDER** specifies disks that ASMLib should scan first. The directive takes a whitespace-delineated list of simple prefix strings to match (for example, string1 string2 … stringn). Wildcards are not allowed. The value of ORACLEASM_SCANORDER will be discussed in sections 6.2.1 and 6.2.2.

- **ORACLEASM_SCANEXCLUDE** specifies the disks that ASMLib will ignore. The directives takes a whitespace-delineated list of simple prefix strings to match (for example, string1 string2 … stringn). Wildcards are not allowed. For most environments and environments that use dynamic multipath devices, use the following to instruct ASMLib to not scan individual block devices, including each path of a multipath pseudo device

  ```
  ORACLEASM_SCANEXCLUDE="sd"
  ```

- **ORACLE_USE_LOGICAL_BLOCK_SIZE** was added to ASMLib for 4KB sector disks. This directive is available in rpms oracleasm-support-2.1.8-1 and kernel-uek 2.6.39-400.4.0 or later. The directive takes a boolean value of true or false and specifies whether or not ASMLib should use the logical or physical block size of the device. To configure this directive, execute `oracleasm configure [-b|-p]`, or manually edit `/etc/sysconfig/oracleasm`.

  To use the logical block size of the device, set the value to TRUE:

  ```
  oracleasm configure -b
  ```

  To use the physical block size of the device, set the value to FALSE:

  ```
  oracleasm configure -p
  ```

  For additional information on ORACLE_USE_LOGICAL_BLOCK_SIZE, see documents 1530578.1 and 1500460.1 in My Oracle Support.

If manually editing `/etc/sysconfig/oracleasm`, make sure the link to `/etc/sysconfig/oracleasm-_dev_oracleasm` is not broken.

```
# ls -ltr /etc/sysconfig/oracleasm*
lrwxrwxrwx. 1 root root  24 Jan 19 13:34 /etc/sysconfig/oracleasm -> oracleasm-
_dev_oracleasm
-rw-r--r--. 1 root root 978 Mar  3 14:13 /etc/sysconfig/oracleasm-_dev_oracleasm
```

After ASMLib is configured, ASMLib can be enabled and started:

```
# /etc/init.d/oracleasm enable
Writing Oracle ASM library driver configuration: done
Initializing the Oracle ASMLib driver:                        [  OK  ]
Scanning the system for Oracle ASMLib disks:                  [  OK  ]
```

After starting ASMLib, ASM administrators need to instruct ASMLib to mark the candidate devices as ASM disks by stamping them with a label, assigning an alias to the devices and placing the devices in directory `/dev/oracleasm/disks/`.

In most Oracle deployments when only using ASMLib to manage ASM disks, ASM instance initialization parameter ASM_DISKSTRING should not be changed from the default value. The default value is equivalent to setting `ASM_DISKSTRING='ORCL:*'`. Disks managed by ASMLib are automatically discovered and identified by alias `ORCL:<ASM diskname>`.

```
[root]# /usr/sbin/oracleasm-discover
Using ASMLib from /opt/oracle/extapi/64/asm/orcl/1/libasm.so
[ASM Library - Generic Linux, version 2.0.12 (KABI_V2)]
Discovered disk: ORCL:CRS1 [16771797 blocks (8587160064 bytes), maxio 1024,
integrity none]
Discovered disk: ORCL:DATA1 [104856192 blocks (53686370304 bytes), maxio 1024,
integrity none]
Discovered disk: ORCL:DATA2 [104856192 blocks (53686370304 bytes), maxio 1024,
integrity none]
Discovered disk: ORCL:FRA1 [104856192 blocks (53686370304 bytes), maxio 1024,
integrity none]
Discovered disk: ORCL:FRA2 [104856192 blocks (53686370304 bytes), maxio 1024,
integrity none]
```

When using ASMLib, setting ASM_DISKSTRING to any value other than the default or 'ORCL:*' (for example, `'/dev/oracleasm/disks/*')` will disable or bypass ASMLib. Bypassing ASMLib may be useful if using Oracle 10g or 11g and using 4K sector disks. See the My Oracle Support document, 1500460.1, for more information.

For information on installing ASMLib and additional configuration, see information in My Oracle Support and references in appendix B.1

### 6.2.1 ASMLib and PowerPath

After a partition is created on a PowerPath pseudo-device and the partition table updated, the partitioned pseudo-device is visible:

```
# fdisk /dev/emcpowerb
<Snippet>
# partprobe /dev/emcpowerb
# ls -ltr /dev | grep emcpowerb1
brw-rw----   1 root disk    120,  97 Mar  9 10:58 emcpowerb1
```

To use the partitioned pseudo-device with ASMLib, use the following process:

1. Configure ASMLib to scan PowerPath pseudo-devices (`/dev/emcpower*1`) before scanning other SCSI devices by adding the following directive to /`etc/sysconfig/oracleasm` and restarting ASMLib. If the environment is a RAC environment, add the directive to file /`etc/sysconfig/oracleasm` on all nodes of the cluster. Then, restart ASMLib on all nodes of the cluster:

   ```
   ORACLEASM_SCANORDER="emcpower"

   # /etc/init.d/oracleasm restart
   Dropping Oracle ASMLib disks:                              [  OK  ]
   Shutting down the Oracle ASMLib driver:                    [  OK  ]
   Initializing the Oracle ASMLib driver:                     [  OK  ]
   Scanning the system for Oracle ASMLib disks:               [  OK  ]
   ```

2. Now use ASMLib to stamp the partitioned pseudo-device as an ASM disk. If the environment is a RAC environment, perform this step on only one of the nodes.

   ```
   # /etc/init.d/oracleasm createdisk data1 /dev/emcpowerb1
   Marking disk "data1" as an ASM disk:                       [  OK  ]
   ```

   Once the partitioned pseudo-device is stamped as an ASM disk, the new ASM disk appears:

   ```
   # ls -ltr /dev/oracleasm/disks/
   total 0
   brw-rw---- 1 grid oinstall 120, 97 Mar  9 12:25 DATA1
   ```

3. To verify the ASM disk is using the partitioned pseudo-device, execute either `ls -l`, or `ASMLib querydisk`, with the `–d` switch, against the ASM disk:

   ```
   # ls -ltr /dev/oracleasm/disks
   total 0
   brw-rw---- 1 grid oinstall 120, 97 Mar  9 14:42 DATA1

   # /etc/init.d/oracleasm querydisk -d DATA1
   Disk "DATA1" is a valid ASM disk on device [120,97]
   ```

   The major and minor device number, 120 and 97 respectfully, of the ASM disk needs to match the major and minor device number of the partitioned pseudo-device:

   ```
   # ls -ltr /dev | grep emcpowerb1
   brw-rw----   1 root disk    120,  97 Mar  9 14:42 emcpowerb1
   ```

```
# grep emcpowerg1 /proc/partitions
 120       97   52428784 emcpowerb1
```

ASM should also have access to all SCSI paths of the PowerPath partitioned pseudo-device:

```
# /etc/init.d/oracleasm querydisk -p DATA1
Disk "DATA1" is a valid ASM disk
/dev/sdc1: LABEL="DATA1" TYPE="oracleasm"
/dev/sdi1: LABEL="DATA1" TYPE="oracleasm"
/dev/sdac1: LABEL="DATA1" TYPE="oracleasm"
/dev/sdaf1: LABEL="DATA1" TYPE="oracleasm"
/dev/emcpowerb1: LABEL="DATA1" TYPE="oracleasm"
```

4. In a RAC environment, after all the ASM disks have been created, execute the following command on the other nodes of the cluster:

```
/etc/init.d/oracleasm scandisks
```

For more information on using Dell EMC PowerPath with Oracle ASM see *Using Oracle Database 10g's Automatic Storage Management with EMC Storage Technology*, a joint engineering white paper authored by Dell EMC and Oracle.

## 6.2.2 ASMLib and Linux DM-Multipath

After a partition is created on a DM-Multipath pseudo-device and the partition table updated, the DM-Multipath partitioned pseudo-device is visible

```
# fdisk /dev/mapper/oraasm-data1
<Snippet>
# partprobe /dev/mapper/oraasm-data1
# ls -ltr /dev/mapper/oraasm-data1p1
lrwxrwxrwx 1 root root 7 Apr 14 11:56 /dev/mapper/oraasm-data1p1 -> ../dm-13
```

To use the partitioned pseudo-device with ASMLib, use the following process:

1. For consistent and persistent device mappings of DM-Multipath devices, ASMLib must be configured to only scan device-mapper devices (`/dev/dm-<n>`) before scanning other devices (e.g. the SCSI devices.). Devices in `/dev/mapper/<name>` are created by UDEV for improved human readability and are not known by the kernel. To only scan device-mapper devices, make sure /`etc/sysconfig/oracleasm` contains the following ASMLib directive and value, then restart ASMLib. If the environment is a RAC environment, add the directive to file oracleasm on all nodes of the cluster. Then, restart ASMLib on all nodes of the cluster:

```
ORACLEASM_SCANORDER="dm-"

# /etc/init.d/oracleasm restart
Dropping Oracle ASMLib disks:                              [  OK  ]
Shutting down the Oracle ASMLib driver:                    [  OK  ]
Initializing the Oracle ASMLib driver:                     [  OK  ]
Scanning the system for Oracle ASMLib disks:               [  OK  ]#
```

2. Now use ASMLib to stamp the partitioned pseudo-device as an ASM disk: If the environment is a RAC environment, perform this step on only one of the nodes.

```
# /etc/init.d/oracleasm createdisk data1 /dev/mapper/oraasm-data1p1
Marking disk "data1" as an ASM disk:                       [  OK  ]
```

   Figure 52 Create ASM disks from partitioned DM-Multipath devices.

   Once the partitioned DM-Multipath pseudo-device is stamped, the new ASM disk appears:

```
# ls -ltr /dev/oracleasm/disks/
total 0
brw-rw---- 1 grid oinstall 252, 13 Apr 14 12:29 DATA1
```

3. To verify that the ASM disk is using the partitioned DM-Multipath device, execute a series of `ls -l` and `ASMLib querydisk -p` commands against the ASM disk and compare the major and minor device numbers. In the following example, the major and minor device numbers are 252 and 13 respectfully:

```
# ls -ltr /dev/oracleasm/disks/DATA1
brw-rw---- 1 grid oinstall 252, 13 Apr 14 12:44 /dev/oracleasm/disks/DATA1

# /etc/init.d/oracleasm querydisk -p data1 | grep mapper
/dev/mapper/oraasm-data1p1: LABEL="DATA1" TYPE="oracleasm"

# ls -ltr /dev/mapper/oraasm-data1p1
lrwxrwxrwx 1 root root 7 Apr 14 12:43 /dev/mapper/oraasm-data1p1 -> ../dm-
13

# ls -ltr /dev/dm-13
brw-rw---- 1 grid oinstall 252, 13 Apr 14 12:43 /dev/dm-13
```

   The major and minor device numbers of the ASM disk needs to match the major and minor device number of the partitioned DM-Multipath device.

   ASM should have access to all SCSI paths of the DM-Multipath partitioned pseudo-device:

```
# /etc/init.d/oracleasm querydisk -p data1
fDisk "DATA1" is a valid ASM disk
/dev/sdc1: LABEL="DATA1" TYPE="oracleasm"
/dev/sdf1: LABEL="DATA1" TYPE="oracleasm"
/dev/sdy1: LABEL="DATA1" TYPE="oracleasm"
/dev/sdab1: LABEL="DATA1" TYPE="oracleasm"
/dev/mapper/oraasm-data1p1: LABEL="DATA1" TYPE="oracleasm"
```

4. In a RAC environment, after all the ASM disks have been created, execute the following command on the other nodes of the cluster:

```
/etc/init.d/oracleasm scandisks
```

## 6.3 ASMFD configuration

ASMFD is a feature available on Linux starting with Oracle 12c Release 1 (12.1.0.2.0) and is installed by default with Oracle Grid Infrastructure. ASMFD resides in the I/O path of the Oracle ASM disks. For ASMFD usage with Oracle 11g Release 2 and Oracle 12c Release 2, see information on My Oracle Support.

ASMFD driver module information and configuration resides in directory:

```
/sys/module/oracleafd
```

By default, the ASMFD driver module is configured for advanced format drives (4k byte sectors) with a boolean value of 0:

```
# cat /sys/module/oracleafd/parameters/oracleafd_use_logical_block_size
# 0
```

If SC Series storage is configured with 512e drives, execute the following command to configure the ASMFD driver so that ASM can correctly define the sector size of the disks as 512 bytes:

```
# echo 1 > /sys/module/oracleafd/parameters/oracleafd_use_logical_block_size
```

After the driver is configured for the type of drives, configure ASMFD by executing the following command from a super user account. The command will install AFD software, load AFD drivers, create UDEV rules for AFD, and additional configuration:

```
asmcmd afd_configure
```

ASMFD marks the device as an ASM disk by stamping the device with a label, assigning an alias to the device and placing the device in directory `/dev/oracleafd/disks`.

When using only ASMFD to manage ASM disks, ASM instance initialization parameter ASM_DISKSTRING should not be changed from the default value, a NULL string. The default value is equivalent to setting ASM_DISKSTRING='AFD:*'. Disks managed by ASMFD are automatically discovered and identified by alias `AFD:<ASM diskname>`.

```
[grid]$ sqlplus / as sysasm
<Snippet>
SQL> show parameter diskstring
NAME                                 TYPE        VALUE
------------------------------------ ----------- ------------------------------
asm_diskstring                       string
```

Figure 53    ASM_DISKSTRING set to default value of NULL

```
asmcmd -p
ASMCMD [+] > dsget
parameter:
profile:
ASMCMD [+] > dsset 'AFD:*'
ASMCMD [+] > dsget
parameter:AFD:*
profile:AFD:*
ASMCMD [+] > quit
```

Figure 54    Using asmcmd to set ASM_DISKSTRING for ASMFD

ASMFD records ASM_DISKSTRING and its value in `/etc/afd.conf`. ASMFD uses the value similar to the ORACLEASM_SCANORDER directive of the ASMLib driver module as it instructs ASMFD to scan only the devices identified by the value.

```
# cat /etc/afd.conf
afd_diskstring='AFD:*'
```

When using multipath devices with ASMFD, ASMFD must be configured to use only pseudo-devices. For additional information on ASMFD, see information in My Oracle Support and in appendix B.1

### 6.3.1 ASMFD and PowerPath

Either partitioned or unpartitioned devices can be used with ASMFD. The content of this section discusses the process of using partitioned devices, but it can be easily modified for unpartitioned devices.

1. Create a partition that spans the entire disk and update the partition table, after which, the partitioned pseudo-device will be visible:

   ```
   # fdisk /dev/emcpowerg
   <snippet>
   # partprobe /dev/emcpowerg
   # ls -ltr /dev | grep emcpowerg
   brw-rw----   1 root disk    120,  96 Mar  9 10:54 emcpowerg
   brw-rw----   1 root disk    120,  97 Mar  9 10:58 emcpowerg1
   ```

2. Configure ASMFD so that it scans only PowerPath pseudo-devices (`/dev/emcpower*`) and not other SCSI devices. To achieve this, make sure `/etc/afd.conf` contains the line below. If the environment is a RAC environment, add the directive to file `/etc/afd.conf` on all nodes of the cluster:

   ```
   afd_diskstring='/dev/emcpower*'
   ```

3. Stop and start the ASMFD driver:

   ```
   /etc/init.d/afd stop
   /etc/init.d/afd start
   ```

4. Now use ASMFD to stamp the pseudo-device as an ASM disk: If the environment is a RAC environment, perform this step on only one of the nodes.

   ```
   asmcmd afd_label DATA1 /dev/emcpowerg1
   ```

   Once the pseudo-device is stamped, the new ASM disk appears:

   ```
   # ls -ltr /dev/oracleafd/disks/
   total 0
   brw-rw---- 1 grid oinstall 120, 97 Mar  9 12:25 DATA1
   ```

5. To verify that the ASM disk is using the pseudo-device, compare the major and minor device number of the ASM disk to the major and minor device number of the pseudo-device. If the major and minor numbers match, the correct device is used.

```
# ls -ltr /dev | grep emcpowerg1
brw-rw----   1 root disk    120,  97 Mar  9 14:42 emcpowerg1
# grep emcpowerg1 /proc/partitions
 120       97   52428784 emcpowerg1
```

6. In a RAC environment, after all the ASM disks have been created, on the other nodes of the cluster, execute:

```
asmcmd afd_scan '/dev/emcpower*'
```

For more information on using Dell EMC PowerPath with Oracle ASM see *Using Oracle Database 10g's Automatic Storage Management with EMC Storage Technology*, a joint engineering white paper authored by Dell EMC and Oracle.

### 6.3.2 ASMFD and Linux DM-Multipath

Either partitioned or unpartitioned devices can be used with ASMFD. The content of this section discusses the process of using partitioned devices, but can be easily modified for unpartitioned devices.

1. Create a partition that spans the entire disk and update the partition table, after which, the partitioned pseudo-device will be visible:

```
# fdisk /dev/mapper/oraasm-data1
<snippet>
# partprobe /dev/mapper/oraasm-data1
# ls -ltr /dev/mapper/oraasm-data1p1
lrwxrwxrwx 1 root root 7 Apr 14 11:56 /dev/mapper/oraasm-data1p1 -> ../dm-7
```

2. Configure ASMFD so that it scans only DM-Multipath pseudo-devices (`/dev/mapper/*`) targeted for ASM and no other SCSI devices. To achieve this, make sure `/etc/afd.conf` contains the line below. If the environment is a RAC environment, add the following directive to file `/etc/afd.conf` on all nodes of the cluster. (See section 4.2 for a discussion of prefix.)

```
afd_diskstring='/dev/mapper/<prefix>*'
```

3. Stop and start the ASMFD driver:

```
/etc/init.d/afd stop
/etc/init.d/afd start
```

4. Use ASMFD to stamp the pseudo-device as an ASM disk. If the environment is a RAC environment, perform this step on only one of the nodes.

```
asmcmd afd_label DATA1 /dev/mapper/oraasm-data1
```

Once the DM-Multipath device is stamped, the new ASM disk appears:

```
# ls -ltr /dev/oracleafd/disks/
total 0
brw-rw---- 1 grid oinstall 252,  7 Apr 14 12:29 DATA1
```

5. To verify that the ASM disk is using the pseudo-device, compare the major and minor device number of the ASM disk to the major and minor device number of the pseudo-device. If the major and minor numbers match, the correct device is used.

   ```
   # ls -ltr /dev/oracleafd/disks/DATA1
   brw-rw---- 1 grid oinstall 252, 9 Apr 14 12:44 /dev/oracleasm/disks/DATA1
   # ls -ltr /dev/mapper/oraasm-data1p1
   lrwxrwxrwx 1 root root 7 Apr 14 12:43 /dev/mapper/oraasm-data1p1 -> ../dm-9
   # ls -ltr /dev/dm-9
   brw-rw---- 1 grid oinstall 252, 9 Apr 14 12:43 /dev/dm-9
   ```

6. In a RAC environment, after all the ASM disks have been created, on the other nodes of the cluster, execute:

   ```
   asmcmd afd_scan '/dev/mapper/<prefix>*'
   ```

For more information on using Dell EMC PowerPath with Oracle ASM see *Using Oracle Database 10g's Automatic Storage Management with EMC Storage Technology*, a joint engineering white paper authored by Dell EMC and Oracle.

## 6.4 UDEV configuration

Through a set of user-defined rules, UDEV can also be used to manage device persistence, ownership, and privileges of ASM disks. Providing an exact series of steps and UDEV rules are unrealistic because the steps and rules are dependent on the customer's storage configuration and requirements. However, several examples of UDEV rules, corresponding dynamic multipath, and ASM_DISKSTRING configurations are provided in this section for reference only. For additional information, see the references in appendix B.1.

When using UDEV, Dell EMC recommends:

- Using the DM-Multipath or PowerPath partitioned or unpartitioned pseudo-device
- Set ASM instance initialization parameter ASM_DISKSTRING to point to the location of the ASM disks referenced or named by UDEV

If partitioned pseudo-devices are used, do not use `$parent` in the device name in UDEV rules because it instructs UDEV to apply the rule to the parent device of a partitioned device and not the partition.

To use UDEV, a custom rules file must be created. The file must contain UDEV configuration directives and it must reside in directory `/etc/udev/rules.d`. Since rule files are read in lexical order, create a rule file with a higher number like `99-oracle-asm-devices.rules`. When specifying configuration directives for a device in a rule, make sure directives OWENR, GROUP and MODE are specified in this order and before any other characteristic. They should also be set to the owner and group of the Linux user that performed the GI install, and privileges must be set to 0660. If multipath pseudo devices are used, reference the pseudo device name in the rule.

## 6.4.1 UDEV and PowerPath

This section provides UDEV examples for PowerPath pseudo devices intended for ASM.

For more information on using Dell EMC PowerPath with Oracle ASM see *Using Oracle Database 10g's Automatic Storage Management with EMC Storage Technology*, a joint engineering white paper authored by Dell EMC and Oracle.

### 6.4.1.1 UDEV with un-partitioned PowerPath pseudo devices

Unpartitioned PowerPath pseudo-devices can be identified in UDEV with a single rule by setting the KERNEL directive to the candidate pseudo-device names targeted for ASM

```
# cat /etc/udev/rules.d/99-oracle-asm-devices.rules
ACTION=="add|change", KERNEL=="emcpower[kgfbc]", OWNER:="grid",
GROUP:="oinstall", MODE="0660"
```

If a single UDEV rule is not granular enough, a UDEV rule can be constructed for each candidate pseudo-device. There are a number of ways to accomplish this, but the example in Figure 55 uses the UUID of the pseudo-devices.

```
# cat /etc/udev/rules.d/99-oracle-asm-devices.rules
# CRS
ACTION=="add|change", KERNEL=="emcpowerf", PROGRAM=="/lib/udev/scsi_id --
page=0x83 --whitelisted --device=/dev/%k",
RESULT=="36000d3100003d0000000000000000241", OWNER:="grid", GROUP:="oinstall",
MODE="0660"
# DATA1
ACTION=="add|change", KERNEL=="emcpowerb", PROGRAM=="/lib/udev/scsi_id --
page=0x83 --whitelisted --device=/dev/%k",
RESULT=="36000d3100003d0000000000000000242", OWNER:="grid", GROUP:="oinstall",
MODE="0660"
# DATA2
ACTION=="add|change", KERNEL=="emcpowerc", PROGRAM=="/lib/udev/scsi_id --
page=0x83 --whitelisted --device=/dev/%k",
RESULT=="36000d3100003d0000000000000000243", OWNER:="grid", GROUP:="oinstall",
MODE="0660"
# FRA1
ACTION=="add|change", KERNEL=="emcpowerk", PROGRAM=="/lib/udev/scsi_id --
page=0x83 --whitelisted --device=/dev/%k",
RESULT=="36000d3100003d0000000000000000244", OWNER:="grid", GROUP:="oinstall",
MODE="0660"
# FRA2
ACTION=="add|change", KERNEL=="emcpowerg", PROGRAM=="/lib/udev/scsi_id --
page=0x83 --whitelisted --device=/dev/%k",
RESULT=="36000d3100003d0000000000000000245", OWNER:="grid", GROUP:="oinstall",
MODE="0660"
```

Figure 55 Unique rule for each candidate PowerPath pseudo device

**Note:** Each rule begins with ACTION and occupies one line in the rule file. Because the margins of this document do not support the length of the rule, the previous rules wrap to a second and third line.

UDEV rules need to be added to the kernel and activated:

```
/sbin/udevadm control --reload-rules
/sbin/udevadm trigger --type=devices --action=change
```

After the rule set becomes active, device-persistence is set on the un-partitioned pseudo-devices:

```
# ls -ltr /dev | grep emcpower[bcfgk]
brw-rw----   1 grid oinstall 120, 160 Apr 13 09:11 emcpowerk
brw-rw----   1 grid oinstall 120,  16 Apr 13 09:11 emcpowerb
brw-rw----   1 grid oinstall 120,  96 Apr 13 09:11 emcpowerg
brw-rw----   1 grid oinstall 120,  32 Apr 13 09:11 emcpowerc
brw-rw----   1 grid oinstall 120,  80 Apr 13 09:11 emcpowerf
```

Figure 56 UDEV set ownership, group and privileges correctly on kernel PowerPath devices

```
# /sbin/powermt display dev=all | egrep 'Pseudo|Logical'
<snippet>
Pseudo name=emcpowerb
Logical device ID=6000D3100003D0000000000000000242 [ora-data1]
Pseudo name=emcpowerc
Logical device ID=6000D3100003D0000000000000000243 [ora-data2]
Pseudo name=emcpowerf
Logical device ID=6000D3100003D0000000000000000244 [ora-fra1]
Pseudo name=emcpowerg
Logical device ID=6000D3100003D0000000000000000245 [ora-fra2]
Pseudo name=emcpowerk
Logical device ID=6000D3100003D0000000000000000241 [ora-crs]
```

When using UDEV, set ASM instance initialization parameter ASM_DISKSTRING to the location and name of the pseudo-devices.

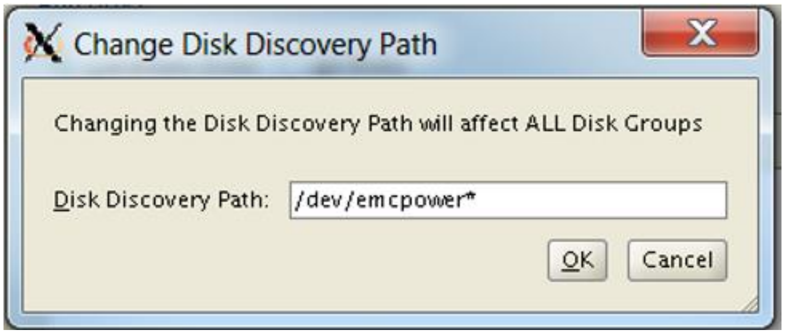
```
asm_diskstring='/dev/emcpower*'
```

```
[grid]$ sqlplus / as sysasm
<snippet>
SQL> alter system set asm_diskstring='/dev/emcpower*';
SQL> show parameter asm_diskstring
NAME                                 TYPE        VALUE
------------------------------------ ----------- ------------------------------
asm_diskstring                       string      /dev/emcpower*
```

```
asmcmd -p
ASMCMD [+] > dsset '/dev/emcpower*'
ASMCMD [+] > dsget
parameter:/dev/emcpower*
profile:/dev/emcpower*
```

To change the disk discovery path in any of the appropriate Oracle GUI tools (such as runInstaller, config.sh, and asmca), select **Change Discovery Path**.

Then, change the default value to the appropriate value for the environment.



Changing the Disk Discovery Path will affect ALL Disk Groups

Disk Discovery Path: /dev/emcpower*

### 6.4.1.2 UDEV with partitioned PowerPath pseudo-devices

The procedure for using partitioned pseudo-devices with UDEV is very similar to using unpartitioned pseudo-devices.

Create a single primary partition on the PowerPath pseudo-device and update the Linux partition table:

```
# fdisk /dev/emcpowerd
<snippet>
# partprobe /dev/emcpowerd
# ls -ltr /dev/emcpowerd*/
brw-rw---- 1 root disk 120, 48 Mar 13 10:30 /dev/emcpowerd
brw-rw---- 1 root disk 120, 49 Mar 16 14:28 /dev/emcpowerd1
```

Figure 57 PowerPath pseudo-device with its primary partition

Partitioned PowerPath pseudo-devices can be identified in UDEV with a single rule by setting the UDEV KERNEL directive to all partitioned PowerPath pseudo-devices targeted for ASM. Since only one partition exists on the pseudo-devices, 1 is used as the suffix in the KERNEL pseudo-device name.

```
# cat /etc/udev/rules.d/99-oracle-asm-devices.rules
ACTION=="add|change", KERNEL=="emcpower[kgebc]1", OWNER:="grid",
GROUP:="oinstall", MODE="0660"
```

**Note**: Each rule begins with ACTION and occupies one line in the rule file. Because the margins of this document do not support the length of the rule, the previous rule wraps to a second line.

If a single UDEV rule is not granular enough for the environment, a set of UDEV rules can be constructed to identify each candidate pseudo-device targeted for ASM. There are a number of ways to accomplish this. The example shown in Figure 58 uses the UUID of the pseudo devices.

```
# cat /etc/udev/rules.d/99-oracle-asm-devices.rules
# CRS
ACTION=="add|change", KERNEL=="emcpowere1", PROGRAM=="/lib/udev/scsi_id --
page=0x83 --whitelisted --device=/dev/%k",
RESULT=="36000d3100003d00000000000000001fa", OWNER:="grid", GROUP:="oinstall",
MODE="0660"
# DATA1
ACTION=="add|change", KERNEL=="emcpowerb1", PROGRAM=="/lib/udev/scsi_id --
page=0x83 --whitelisted --device=/dev/%k",
RESULT=="36000d3100003d00000000000000001fb", OWNER:="grid", GROUP:="oinstall",
MODE="0660"
# DATA2
ACTION=="add|change", KERNEL=="emcpowerc1", PROGRAM=="/lib/udev/scsi_id --
page=0x83 --whitelisted --device=/dev/%k",
RESULT=="36000d3100003d0000000000000000206", OWNER:="grid", GROUP:="oinstall",
MODE="0660"
# FRA1
ACTION=="add|change", KERNEL=="emcpowerk1", PROGRAM=="/lib/udev/scsi_id --
page=0x83 --whitelisted --device=/dev/%k",
RESULT=="36000d3100003d00000000000000001fc", OWNER:="grid", GROUP:="oinstall",
MODE="0660"
# FRA2
ACTION=="add|change", KERNEL=="emcpowerg1", PROGRAM=="/lib/udev/scsi_id --
page=0x83 --whitelisted --device=/dev/%k",
RESULT=="36000d3100003d0000000000000000207", OWNER:="grid", GROUP:="oinstall",
MODE="0660"
```

Figure 58 Unique rule for each candidate PowerPath pseudo device

**Note**: Each rule begins with ACTION and resides on a single line in the rule file. Because the margins of this document do not support the length of the rule, the rule wraps to a second and third line.

Once the rule set is defined, add the rule set to the kernel and activate it:

```
# /sbin/udevadm control --reload-rules
# /sbin/udevadm trigger --type=devices --action=change
```

After the rule set becomes active, device-persistence is set on the desired partitioned pseudo-devices:

```
# ls -ltr /dev/ | grep emcpower
<snippet>
brw-rw----   1 root disk     120, 160 Apr  7 12:03 emcpowerk
brw-rw----   1 root disk     120,  96 Apr  7 12:03 emcpowerg
brw-rw----   1 root disk     120,  64 Apr  7 12:03 emcpowere
brw-rw----   1 root disk     120,  32 Apr  7 12:03 emcpowerc
brw-rw----   1 root disk     120,  16 Apr  7 12:03 emcpowerb
brw-rw----   1 grid oinstall 120,  65 Apr  7 12:03 emcpowere1
brw-rw----   1 grid oinstall 120,  97 Apr  7 12:03 emcpowerg1
brw-rw----   1 grid oinstall 120, 161 Apr  7 12:03 emcpowerk1
brw-rw----   1 grid oinstall 120,  33 Apr  7 12:03 emcpowerc1
brw-rw----   1 grid oinstall 120,  17 Apr  7 12:03 emcpowerb1
```

```
# /sbin/powermt display dev=all | egrep 'Pseudo|Logical'
<snippet>
Pseudo name=emcpowere
Logical device ID=6000D3100003D0000000000000000241 [ora-crs]
Pseudo name=emcpowerg
Logical device ID=6000D3100003D0000000000000000242 [ora-data1]
Pseudo name=emcpowerb
Logical device ID=6000D3100003D0000000000000000243 [ora-data2]
Pseudo name=emcpowerc
Logical device ID=6000D3100003D0000000000000000244 [ora-fra1]
Pseudo name=emcpowerk
Logical device ID=6000D3100003D0000000000000000245 [ora-fra2]
```

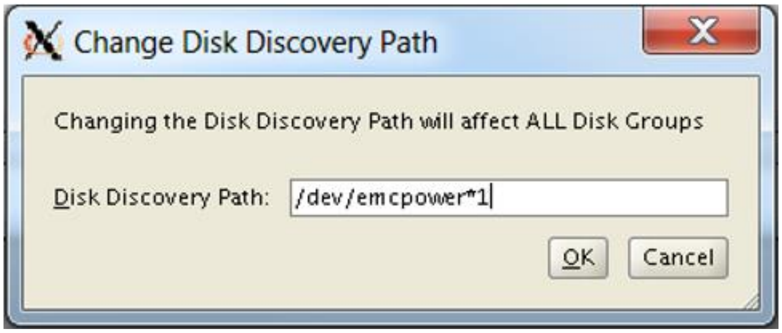
When using UDEV, ASM instance initialization parameter ASM_DISKSTRING should be set to the location and name of the partitioned pseudo-devices rather than to the name of the parent pseudo-device.

```
[grid]$ sqlplus / as sysasm
<snippet>
SQL> alter system set asm_diskstring='/dev/emcpower*1';
SQL> show parameter asm_diskstring
NAME                                 TYPE        VALUE
------------------------------------ ----------- ------------------------------
asm_diskstring                       string      /dev/emcpower*1
```

```
[grid]$ asmcmd -p
ASMCMD [+] > dsset '/dev/emcpower*1'
ASMCMD [+] > dsget
parameter:/dev/emcpower*1
profile:/dev/emcpower*1
```

To change the disk discovery path in any of the Oracle GUI tools (such as runInstaller, config.sh, and asmca), select **Change Discovery Path**.

Then change the default value to the appropriate values for the environment.



Changing the Disk Discovery Path will affect ALL Disk Groups

Disk Discovery Path: /dev/emcpower*1

## 6.4.2 UDEV and DM-Multipath

This section provides UDEV examples for DM-Multipath pseudo devices intended for ASM. The snippet from /etc/multipath.conf in Figure 46 is used for the examples in this section.

For additional usage and configuration information on DM-Multipath I/O with ASM and UDEV, see information provided at My Oracle Support, as well as Red Hat Enterprise Linux 6 DM multipath documentation and Red Hat Enterprise Linux 7 DM multipath documentation.

### 6.4.2.1 UDEV with unpartitioned DM-Multipath pseudo devices

Assuming a prefix and suffix were used when naming the DM-Multipath device alias (Figure 46), a single UDEV rule can be used to manage device persistence of these devices. All candidate devices can be identified in the UDEV rule by testing the equivalency of:

- Device environment key DM_NAME against the unique prefix of the DM-Multipath alias names
- KERNEL against all kernel device-mapper devices (dm-*)

```
# cat /etc/udev/rules.d/99-oracle-asm-devices.rules
ACTION=="add|change", KERNEL=="dm-
*",ENV{DM_NAME}=="oraasm*",OWNER="grid",GROUP="oinstall",MODE="0660"
```

**Note:** Each rule begins with ACTION and occupies one line in the rule file. Because the margins of this document do not support the length of the rule, the previous rule wraps to a second line.

If a single UDEV rule is too generic for the environment, a UDEV rule can be constructed for each candidate pseudo-device targeted for ASM. The example shown in Figure 60 uses the UUID of the pseudo devices shown in Figure 59.

```
# cd /dev/mapper
# for i in oraasm*[^p][12]
>   do
>   echo $i
>   udevadm info --query=all --name=/dev/mapper/$i | grep -i DM_UUID
> done
oraasm-crs1
E: DM_UUID=mpath-36000d3100003d0000000000000000241
oraasm-data1
E: DM_UUID=mpath-36000d3100003d0000000000000000242
oraasm-data2
E: DM_UUID=mpath-36000d3100003d0000000000000000243
oraasm-fra1
E: DM_UUID=mpath-36000d3100003d0000000000000000244
oraasm-fra2
E: DM_UUID=mpath-36000d3100003d0000000000000000245
```

Figure 59 Display UUIDs of DM-Multipath pseudo devices

```
# cat /etc/udev/rules.d/99-oracle-asm-devices.rules
ACTION=="add|change", ENV{DM_UUID}=="mpath-36000d3100003d0000000000000000241",
GROUP="oinstall", OWNER="grid", MODE="0660"
ACTION=="add|change", ENV{DM_UUID}=="mpath-36000d3100003d0000000000000000242",
GROUP="oinstall", OWNER="grid", MODE="0660"
ACTION=="add|change", ENV{DM_UUID}=="mpath-36000d3100003d0000000000000000243",
GROUP="oinstall", OWNER="grid", MODE="0660"
ACTION=="add|change", ENV{DM_UUID}=="mpath-36000d3100003d0000000000000000244",
GROUP="oinstall", OWNER="grid", MODE="0660"
ACTION=="add|change", ENV{DM_UUID}=="mpath-36000d3100003d0000000000000000245",
GROUP="oinstall", OWNER="grid", MODE="0660"
```

Figure 60 Unique rule for each candidate DM-Multipath pseudo device

**Note:** Each rule begins with ACTION and occupies one line in the rule file. Because the margins of this document do not support the length of the rule, the previous rules wrap to a second line.

After the rules are created, stop DM-Multipath or remove any device specified in the UDEV rule from DM-Multipath. Then, reload UDEV and either reload or start DM-Multipath.

```
multipath -f oraasm-fra1
multipath -f oraasm-fra2
multipath -f oraasm-data2
multipath -f oraasm-data1
multipath -f oraasm-crs1

/sbin/udevadm control --reload-rules
/sbin/udevadm trigger --type=devices --action=change
/etc/init.d/multipathd reload
```

After the rules are active, verify the kernel device-mapper devices (dm-) for the DM-Multipath pseudo device aliases have the owner, group and privilege defined appropriately:

```
[root]# ls -ltr /dev | grep 'dm-[01346]$'
brw-rw----   1 grid oinstall 252,   1 May 16 12:49 dm-1
brw-rw----   1 grid oinstall 252,   0 May 16 12:49 dm-0
brw-rw----   1 grid oinstall 252,   3 May 16 12:49 dm-3
brw-rw----   1 grid oinstall 252,   4 May 16 12:49 dm-4
brw-rw----   1 grid oinstall 252,   6 May 16 12:49 dm-6
```

Figure 61 UDEV set ownership, group and privileges correctly on kernel dm devices

When using UDEV, set ASM instance initialization parameter ASM_DISKSTRING to the location and prefix of the alias pseudo device:

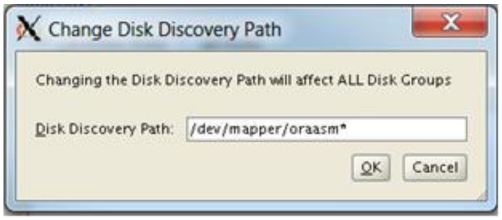
```
asm_diskstring='/dev/mapper/oraasm*'
```

```
[grid]$ sqlplus / as sysasm
<snippet>
SQL> alter system set asm_diskstring='/dev/mapper/oraasm*';
SQL> show parameter asm_diskstring
NAME                                 TYPE        VALUE
------------------------------------ ----------- ------------------------------
asm_diskstring                       string      /dev/mapper/oraasm*
```

```
[grid]$ asmcmd -p
ASMCMD [+] > dsset '/dev/mapper/oraasm*'
ASMCMD [+] > dsget
parameter: /dev/mapper/oraasm*
profile: /dev/mapper/oraasm*
```

To change the disk discovery path in any of the appropriate Oracle GUI tools (such as runInstaller, config.sh, and asmca), select **Change Discovery Path**.

Then change the default value to the appropriate values for the environment.



Changing the Disk Discovery Path will affect ALL Disk Groups

Disk Discovery Path: /dev/mapper/oraasm*

When using UDEV in a RAC environment, set ASM_DISKSTRING to the locations and names of the pseudo-devices that represent the shared disks on all nodes.

### 6.4.2.2 UDEV with partitioned DM-Multipath pseudo devices

Assuming a prefix and suffix were used when naming the DM-Multipath device alias (see Figure 46), create a single primary partition on the DM-Multipath device and update the Linux partition table with the new partition:

```
# fdisk /dev/mapper/oraasm-data1
<snippet>
# partprobe /dev/mapper/oraasm-data1
# /etc/init.d/multipathd reload
Reloading multipathd:                                     [  OK  ]
# ls -ltr /dev/mapper/ | grep data
lrwxrwxrwx 1 root root       7 Mar  7 11:10 oraasm-data1 -> ../dm-8
lrwxrwxrwx 1 root root       8 Mar  7 11:10 oraasm-data1p1 -> ../dm-11
```

Figure 62 DM-Multipath pseudo device with its primary partition

DM-Multipath partitioned pseudo-devices can be identified in UDEV by a single rule that tests:

- The device environment key DM_NAME against the unique prefix and suffix of the DM-Multipath partitioned device. Since only one partition exists on the DM-Multipath device, p1 is used as the suffix.
- KERNEL against all kernel device-mapper devices (dm-*)

```
# cat /etc/udev/rules.d/99-oracle-asm-devices.rules
ACTION=="add|change", KERNEL=="dm-
*",ENV{DM_NAME}=="oraasm*p1",OWNER:="grid",GROUP:="oinstall",MODE="0660"
```

**Note:** Each rule begins with ACTION and occupies one line in the rule file. Because the margins of this document do not support the length of the rule, the previous rule wraps to a second line.

If a single UDEV rule is too generic for the environment, a UDEV rule can be constructed for each candidate pseudo-device targeted for ASM. The example shown in Figure 64 uses the UUIDs of the DM-Multipath partitioned pseudo devices in Figure 63. Notice the prefix of the DM_UUID value, part1-, which stands for partition 1.

```
# cd /dev/mapper
# for i in oraasm*p1
>   do
>   echo $i
>   udevadm info --query=all --name=/dev/mapper/$i | grep -i DM_UUID
> done
oraasm-crs1p1
E: DM_UUID=part1-mpath-36000d3100003d0000000000000000241
oraasm-data1p1
E: DM_UUID=part1-mpath-36000d3100003d0000000000000000242
oraasm-data2p1
E: DM_UUID=part1-mpath-36000d3100003d0000000000000000243
oraasm-fra1p1
E: DM_UUID=part1-mpath-36000d3100003d0000000000000000244
oraasm-fra2p1
E: DM_UUID=part1-mpath-36000d3100003d0000000000000000245
```

Figure 63 Display UUIDs of DM-Multipath partitioned pseudo devices

```
# cat /etc/udev/rules.d/99-oracle-asm-devices.rules
#oraasm-crs1p1
ACTION=="add|change", ENV{DM_UUID}=="part1-mpath-
36000d3100003d0000000000000000241", GROUP="oinstall", OWNER="grid", MODE="0660"

#oraasm-data1p1
ACTION=="add|change", ENV{DM_UUID}=="part1-mpath-
36000d3100003d0000000000000000242", GROUP="oinstall", OWNER="grid", MODE="0660"

#oraasm-data2p1
ACTION=="add|change", ENV{DM_UUID}=="part1-mpath-
36000d3100003d0000000000000000243", GROUP="oinstall", OWNER="grid", MODE="0660"

#oraasm-fra1p1
ACTION=="add|change", ENV{DM_UUID}=="part1-mpath-
36000d3100003d0000000000000000244", GROUP="oinstall", OWNER="grid", MODE="0660"

#oraasm-fra2p1
ACTION=="add|change", ENV{DM_UUID}=="part1-mpath-
36000d3100003d0000000000000000245", GROUP="oinstall", OWNER="grid", MODE="0660"
```

Figure 64 Unique rule for each candidate partitioned pseudo device

**Note:** Each rule begins with ACTION and resides on a single line in the rule file. Because the margins of this document do not support the length of the rule, the rule wraps to a second line.

After the rules are created, stop DM-Multipath or remove any device specified in the UDEV rule from DM-Multipath. Then, reload UDEV and either reload or start DM-Multipath.

```
multipath -f oraasm-fra1
multipath -f oraasm-fra2
multipath -f oraasm-data2
multipath -f oraasm-data1
multipath -f oraasm-crs1

/sbin/udevadm control --reload-rules
/sbin/udevadm trigger --type=devices --action=change
/etc/init.d/multipathd reload
```

After the rules are active, verify the kernel device-mapper devices (dm-) for the DM-Multipath pseudo device aliases have the owner, group and privilege defined appropriately:

```
# ls -ltr /dev/mapper/oraasm*p1
lrwxrwxrwx 1 root root 8 Apr 14 15:17 /dev/mapper/oraasm-fra1p1 -> ../dm-12
lrwxrwxrwx 1 root root 8 Apr 14 15:17 /dev/mapper/oraasm-fra2p1 -> ../dm-16
lrwxrwxrwx 1 root root 8 Apr 14 15:17 /dev/mapper/oraasm-crs1p1 -> ../dm-15
lrwxrwxrwx 1 root root 7 Apr 14 15:17 /dev/mapper/oraasm-data2p1 -> ../dm-14
lrwxrwxrwx 1 root root 7 Apr 14 15:17 /dev/mapper/oraasm-data1p1 -> ../dm-13

# ls -ltr /dev | egrep "dm-[89]|dm-1[034]"
brw-rw----   1 grid oinstall 252,  15 Apr 14 15:17 dm-15
brw-rw----   1 grid oinstall 252,  16 Apr 14 15:17 dm-16
brw-rw----   1 grid oinstall 252,  12 Apr 14 15:17 dm-12
brw-rw----   1 grid oinstall 252,  14 Apr 14 15:17 dm-14
brw-rw----   1 grid oinstall 252,  13 Apr 14 15:17 dm-13
```

When using partitioned DM-Multipath pseudo-devices and UDEV, set Oracle ASM instance parameter ASM_DISKSTRING to the location and aliased pseudo device, rather than to the name of the parent DM-Multipath logical device or kernel device-mapper (dm-) device.

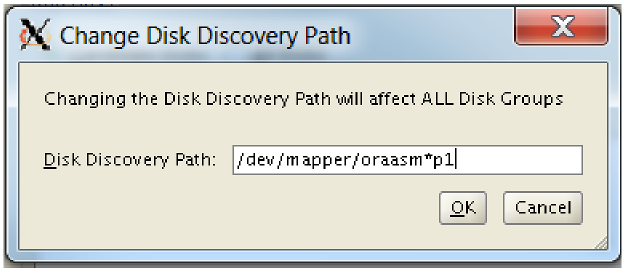
```
asm_diskstring=’/dev/mapper/oraasm*p1’
```

```
[grid]$ sqlplus / as sysasm
<snippet>
SQL> alter system set asm_diskstring='/dev/mapper/oraasm*p1';
SQL> show parameter asm_diskstring
NAME                                 TYPE        VALUE
------------------------------------ ----------- ------------------------------
asm_diskstring                       string      /dev/mapper/oraasm*p1
```

```
[grid]$ asmcmd -p
ASMCMD [+] > dsset '/dev/mapper/oraasm*p1'
ASMCMD [+] > dsget
parameter: /dev/mapper/oraasm*p1
profile: /dev/mapper/oraasm*p1
```

To change the disk discovery path in any of the Oracle GUI tools (such as runInstaller, config.sh, and asmca), select **Change Discovery Path**.

Then change the default value to the appropriate value for the environment.



Changing the Disk Discovery Path will affect ALL Disk Groups

Disk Discovery Path: /dev/mapper/oraasm*p1

## 6.5 ASM disk groups

Oracle generally recommends using no more than two disk groups per database:

**Database area:** Some of the data typically saved in this disk group are: database files containing application data and indexes, database metadata, control files, online redo logs.

**Flash Recovery Area (FRA):** This disk group contains recovery-related files such as multiplexed copies of control files, online redo logs, backup set, and flashback logs.

However, this may not be adequate. Dell EMC recommends additional disk groups be considered for each type of SC Series storage profile used by the database, different disk groups for the different I/O patterns in the database, and one disk group for Grid Infrastructure CRS should CRS be required. Care must be used when choosing the number of disk groups, because as the number of disk groups increases so does the complexity of the system.

Figure 65 shows how ASM disk groups could reside in a storage enclosure of a flash-optimized array.

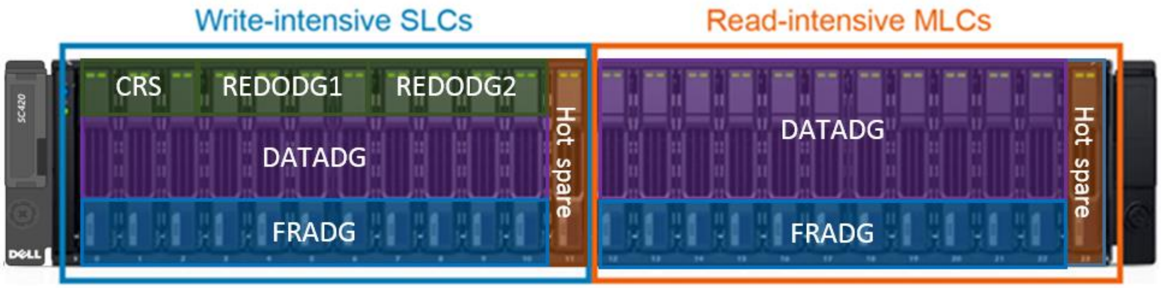


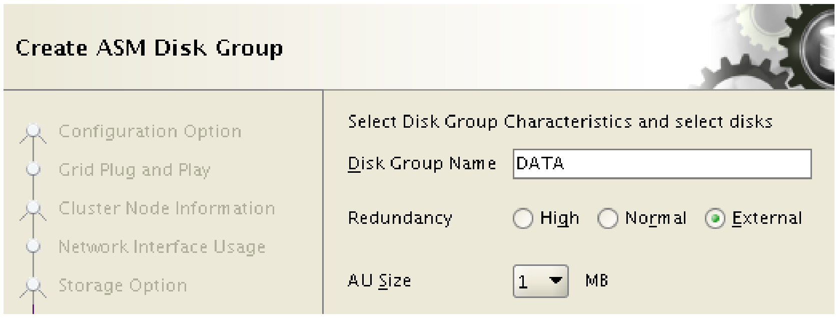
Figure 65 ASM disk groups

ASM provides three different types of redundancy that can be used within each ASM disk group:

**Normal Redundancy**: Provides two-way mirroring in ASM. This requires two ASM failure groups and is the default redundancy. This is the default.

**High Redundancy**: Provides three-way mirroring in ASM. This requires three ASM failure groups.

**External Redundancy**: Provides no mirroring in ASM. This level of redundancy requires some form of RAID protection on the storage arrays. Since external redundancy leverages RAID protection provided by the storage array, the database server processor and other resources will be less consumed, therefore increasing processor cycles for database operations.



Create ASM Disk Group

Select Disk Group Characteristics and select disks

Disk Group Name DATA

Redundancy ○ High ○ Normal ● External

AU Size 1 MB

Figure 66 ASM disk group redundancy

The following use cases will help determine the type of ASM redundancy that should be used:

- Normal or high redundancy must be selected for Oracle stretch clusters.
- If a server cannot access an ASM LUN at the operating system level due to operating system errors, the ASM instance will fail. If this is a primary concern, then consider configuring ASM disk groups with normal (two ASM failure groups) or high redundancy (three ASM failure groups).
- Mission-critical applications may warrant high redundancy disk groups rather than normal or external redundancy.
- External redundancy should be used if the RAID protection provided by the SC Series array is sufficient in meeting business requirements. In most Oracle environments, this will be the preferred and recommended redundancy when using SC series storage.

When using normal or high ASM redundancy for a disk group, ensure there is enough free space in the disk groups. Oracle needs this free space to automatically rebuild the contents of a failed drive, to other drives in the failure group belonging to the disk group. The amount of space required in a disk group for this redundancy can be found in column `V$ASM_DISKGROUP.REQUIRED_MIRROR_FREE_MB`.

When sizing the SC Series array, take in to account the type of ASM disk group redundancy (external, normal, or high) required and the volume of data expected in each of the different redundant disk groups:

- **External redundancy:** 1x size of expected data retained in disk groups of this redundancy.
- **Normal redundancy:** 2x size of expected data retained in disk groups of this redundancy.
- **High redundancy:** 3x size of expected data retained in disk groups of this redundancy.

If Oracle extended distance clusters are required, a second SC Series array will be required and needs to be sized and configured appropriately. Ideally, both SC series arrays and I/O paths should be sized and configured identically.

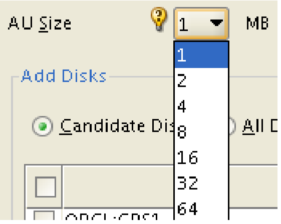
If dynamic multipath devices are used, Dell EMC recommends using pseudo-device names and not Linux device names when creating ASM disk groups. Dell EMC also recommends each ASM disk group that contains database data or indexes be created with an even number of ASM disks with the same capacity and performance characteristic, and where each ASM disk within the same disk group is active on a different controller. This allows both controllers to participate in servicing the IO requests from the originating ASM disk group, and it allows ASM to stripe. If ASM disk groups containing other types of Oracle data exist, evaluate whether or not they should contain an even number of disks, for example a disk group for CRS.

When a disk is added to a disk group, an ASM disk name is assigned and written to the whole-drive device or the partition. This name is not the same as the operating system device or pseudo-device name. The ASM disk name allows the device or pseudo-device name be different on nodes of the same cluster, provided that ASM_DISKSTRING is set to the location of the device.

### 6.5.1 ASM allocation units (AU) and OS device tuning considerations

Every Oracle ASM disk is divided into allocation units (AU) and is the fundamental unit of allocation within a disk group. A file extent consists of one or more allocation units. An Oracle ASM file consists of one or more file extents.

When creating a disk group, set the Oracle ASM allocation unit size to 1, 2, 4, 8, 16, 32, or 64 MB, depending on the specific disk group compatibility level. Larger AU sizes typically provide performance advantages for data warehouse applications that use large sequential reads and very large databases (VLDBs). See section 3.25 for information on how AU size impacts the size of a database.


AU Size 1 MB

Since Oracle 11g, Oracle recommends 4 MB for the AU size for a disk group. 8 MB AUs might also be a good choice. Some benefits of using 4 MB AU and 8 MB AU are:

- Increased I/O throughput of the I/O subsystem
- Reduced SGA size to manage the extent maps in the database instance.
- Faster datafile initialization
- Increased file size limits
- Reduced database open time
- 8 MB AU may deliver improved performance if the database is large as there are fewer AUs to manage

The objective in setting AU size is to define it small enough that the AUs do not become too hot and large enough for optimal sequential reads, while reducing the number of AUs to manage.

### 6.5.2 ASM diskgroups with ASMLib and pseudo devices

When using ASMLib with pseudo devices from either EMC PowerPath or native DM-Multipath disk groups should reference ASMLib managed disks (Figure 67) rather than pseudo devices:

```
$ /etc/init.d/oracleasm listdisks
CRS1
DATA1
DATA2
FRA1
FRA2
```

Figure 67 Display ASMLib managed disks

When identifying ASM disks during the creation of an ASM disk group, specify a search string that points to a subset of the disks returned by the disk discovery (ASM_DISKSTRING). Any discovered disk that matches the specified search string will be added to the disk group. For example, with ASM_DISKSTRING value (ORCL:*), the ASM instance will discover all ASM disks in /dev/oracleasm/disks.

Then, when creating the disk group, if the disk string (asmcmd: dsk string. SQL*Plus: DISK clause) identifies a subset of the discovered disks, the subset of disks will be added to the disk group. The following will add all disks having a name starting with DATA from the discovery path to the disk group.

```
$ cat dg_data_config-ASMLib-disks.xml
<dg name="data" redundancy="external">
<dsk string="ORCL:DATA*"/>
<a name="compatible.asm" value="11.2"/>
<a name="compatible.rdbms" value="11.2"/>
<a name="au_size" value="4M"/>
</dg>


$ asmcmd lsdg
State     Type    Rebal  Sector  Block       AU  Total_MB  …   Name
MOUNTED   EXTERN  N         512   4096  1048576      5115  …   CRS/
$ asmcmd mkdg dg_data_config-ASMLib-disks.xml
$ asmcmd lsdg
State     Type    Rebal  Sector  Block       AU  Total_MB  …   Name
MOUNTED   EXTERN  N         512   4096  1048576      5115  …   CRS/
MOUNTED   EXTERN  N         512   4096  4194304    102398  …   DATA/
```

Figure 68 Create disk group with a subset of discovered disks using asmcmd xml file

```
SQL> CREATE DISKGROUP FRA
  2    EXTERNAL REDUNDANCY
  3    DISK 'ORCL:FRA*'
  4    ATTRIBUTE 'compatible.asm'='11.2'
  5            , 'compatible.rdbms'='11.2'
  6  /
```

Figure 69 Create disk group with a subset of discovered disks using SQL*Plus

```
SQL> select GROUP_NUMBER, DISK_NUMBER, … from v$asm_disk;

Grp Dsk Mount   Mode
Num Num Status  Status  Name            FailGrp         Label           Path
--- --- ------- ------- --------------- --------------- --------------- --------
---------
  0   2 CLOSED  ONLINE                                  FRA1
ORCL:FRA1
  0   3 CLOSED  ONLINE                                  FRA2
ORCL:FRA2
  1   0 CACHED  ONLINE  CRS1            CRS1            CRS1
ORCL:CRS1
  2   0 CACHED  ONLINE  DATA1           DATA1           DATA1
ORCL:DATA1
  2   1 CACHED  ONLINE  DATA2           DATA2           DATA2
ORCL:DATA2
```

Figure 70 Disk group 2 and 0 were created with a subset of discovered disks

If it is desired to uniquely reference discovered disks while creating the disk group, name each ASM disk:

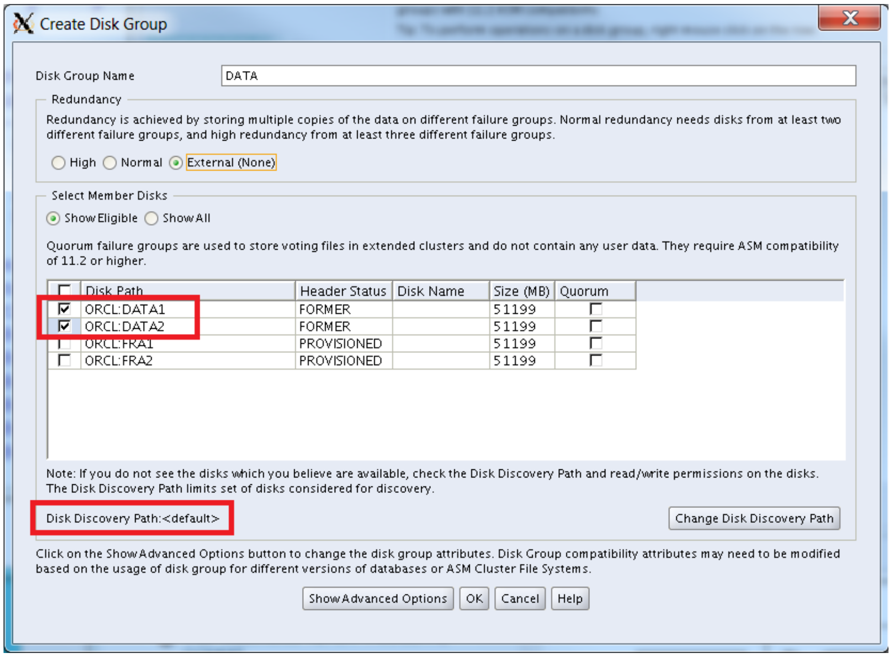
```
$ cat dg_data_config-ASMLib-multiple_disks.xml
<dg name="data" redundancy="external">
<dsk name="DATA1" string="ORCL:DATA1"/>
<dsk name="DATA2" string="ORCL:DATA2"/>
<a name="compatible.asm" value="11.2"/>
<a name="compatible.rdbms" value="11.2"/>
</dg>
$ asmcmd mkdg dg_data_config-ASMLib-multiple_disks.xml
```

Figure 71 Create disk group with specific set of discovered disks using asmcmd xml file

```
SQL> CREATE DISKGROUP FRA
  2    EXTERNAL REDUNDANCY
  3    DISK 'ORCL:FRA1'
  4       , 'ORCL:FRA2'
  5    ATTRIBUTE 'compatible.asm'='11.2'
  6             , 'compatible.rdbms'='11.2'
  7  /
```

Figure 72 Create disk group with specific set of discovered disks using SQL*Plus

If using asmca to create ASM disk groups, select the check boxes associated with the appropriate disks for the disk group:



Disk Group Name: DATA

Redundancy

Redundancy is achieved by storing multiple copies of the data on different failure groups. Normal redundancy needs disks from at least two different failure groups, and high redundancy from at least three different failure groups.

○ High ○ Normal ● External (None)

Select Member Disks

● Show Eligible ○ Show All

Quorum failure groups are used to store voting files in extended clusters and do not contain any user data. They require ASM compatibility of 11.2 or higher.

| □ | Disk Path | Header Status | Disk Name | Size (MB) | Quorum |
|---|---|---|---|---|---|
| ☑ | ORCL:DATA1 | FORMER | | 51199 | □ |
| ☑ | ORCL:DATA2 | FORMER | | 51199 | □ |
| □ | ORCL:FRA1 | PROVISIONED | | 51199 | □ |
| □ | ORCL:FRA2 | PROVISIONED | | 51199 | □ |

Note: If you do not see the disks which you believe are available, check the Disk Discovery Path and read/write permissions on the disks. The Disk Discovery Path limits set of disks considered for discovery.

Disk Discovery Path:<default>

Click on the Show Advanced Options button to change the disk group attributes. Disk Group compatibility attributes may need to be modified based on the usage of disk group for different versions of databases or ASM Cluster File Systems.

## 6.5.3 ASM diskgroups with ASMFD and pseudo devices

When using ASMFD with pseudo devices from either PowerPath or native DM-Multipath, disk groups should reference ASMFD managed disks:

```
/dev/oracleafd/disks/<name>
```

When identifying ASM disks during the creation of an ASM disk group, specify a search string that points to a subset of the disks returned by the disk discovery (ASM_DISKSTRING). Any discovered disk that matches the specified search string will be added to the disk group. For example, with ASM_DISKSTRING value (`AFD:*`), the ASM instance will discover all ASM disks in `/dev/oracleafd/disks`.

Then, when creating the disk group, if the disk string (asmcmd: dsk string. SQL*Plus: DISK clause) identifies a subset of the discovered disks, the subset of disks will be added to the disk group. The following will add all disks having a name starting with DATA from the discovery path to the disk group.

```
$ cat dg_data_config-ASMLib-disks.xml
<dg name="data" redundancy="external">
<dsk string="AFD:DATA*"/>
<a name="compatible.asm" value="11.2"/>
<a name="compatible.rdbms" value="11.2"/>
<a name="au_size" value="4M"/>
</dg>


$ asmcmd lsdg
State    Type    Rebal  Sector  Block       AU  Total_MB  …   Name
MOUNTED  EXTERN  N         512   4096  1048576      5115  …   CRS/
$ asmcmd mkdg dg_data_config-ASMLib-disks.xml
$ asmcmd lsdg
State    Type    Rebal  Sector  Block       AU  Total_MB  …   Name
MOUNTED  EXTERN  N         512   4096  1048576      5115  …   CRS/
MOUNTED  EXTERN  N         512   4096  4194304    102398  …   DATA/
```

Figure 73 Create disk group with a subset of discovered disks using asmcmd xml file

```
SQL> CREATE DISKGROUP FRA
  2    EXTERNAL REDUNDANCY
  3    DISK 'AFD:FRA*'
  4    ATTRIBUTE 'compatible.asm'='11.2'
  5            , 'compatible.rdbms'='11.2'
  6  /
```

Figure 74 Create disk group with a subset of discovered disks using SQL*Plus

```
SQL> select GROUP_NUMBER, DISK_NUMBER, … from v$asm_disk;

Grp Dsk Mount   Mode
Num Num Status  Status  Name            FailGrp         Label           Path
--- --- ------- ------- --------------- --------------- --------------- --------
---------
  0   2 CLOSED  ONLINE                                  FRA1            AFD:FRA1
  0   3 CLOSED  ONLINE                                  FRA2            AFD:FRA2
  1   0 CACHED  ONLINE  CRS1            CRS1            CRS1            AFD:CRS1
  2   0 CACHED  ONLINE  DATA1           DATA1           DATA1
AFD:DATA1
  2   1 CACHED  ONLINE  DATA2           DATA2           DATA2
AFD:DATA2
```

Figure 75 Disk group 2 and 0 were created with a subset of discovered disks

If it is desired to uniquely reference discovered disks while creating the disk group, name each ASMFD disk:

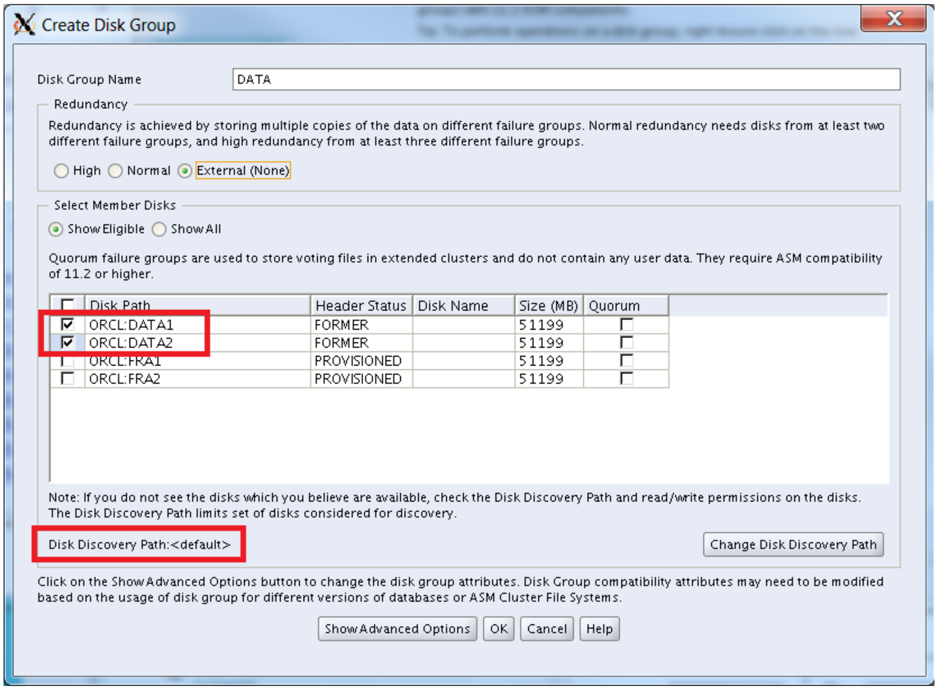
```
$ cat dg_data_config-ASMFD-multiple_disks.xml
<dg name="data" redundancy="external">
<dsk name="DATA1" string="AFD:DATA1"/>
<dsk name="DATA2" string="AFD:DATA2"/>
<a name="compatible.asm" value="11.2"/>
<a name="compatible.rdbms" value="11.2"/>
</dg>
$ asmcmd mkdg dg_data_config-ASMFD-multiple_disks.xml
```

Figure 76 Create disk group with specific set of discovered disks using asmcmd xml file

```
SQL> CREATE DISKGROUP FRA
  2    EXTERNAL REDUNDANCY
  3    DISK 'AFD:FRA1'
  4       , 'AFD:FRA2'
  5    ATTRIBUTE 'compatible.asm'='11.2'
  6             , 'compatible.rdbms'='11.2'
  7  /
```

Figure 77 Create disk group with specific set of discovered disks using SQL*Plus

If using asmca to create ASM disk groups, select the check boxes associated with the appropriate disks for the disk group:



Create Disk Group

Disk Group Name: DATA

Redundancy

Redundancy is achieved by storing multiple copies of the data on different failure groups. Normal redundancy needs disks from at least two different failure groups, and high redundancy from at least three different failure groups.

○ High ○ Normal ● External (None)

Select Member Disks

● Show Eligible ○ Show All

Quorum failure groups are used to store voting files in extended clusters and do not contain any user data. They require ASM compatibility of 11.2 or higher.

| ☐ | Disk Path | Header Status | Disk Name | Size (MB) | Quorum |
|---|---|---|---|---|---|
| ☑ | ORCL:DATA1 | FORMER | | 51199 | ☐ |
| ☑ | ORCL:DATA2 | FORMER | | 51199 | ☐ |
| ☐ | ORCL:FRA1 | PROVISIONED | | 51199 | ☐ |
| ☐ | ORCL:FRA2 | PROVISIONED | | 51199 | ☐ |

Note: If you do not see the disks which you believe are available, check the Disk Discovery Path and read/write permissions on the disks. The Disk Discovery Path limits set of disks considered for discovery.

Disk Discovery Path:<default>

Click on the Show Advanced Options button to change the disk group attributes. Disk Group compatibility attributes may need to be modified based on the usage of disk group for different versions of databases or ASM Cluster File Systems.

## 6.5.4 ASM diskgroups with PowerPath and UDEV

This section refers to partitioned pseudo-devices. If unpartitioned devices are used, simply remove any reference to the partition and <partition-indicator> value from the remainder of this section.

ASM instance initialization parameter ASM_DISKSTRING is used by ASM to discover all candidate ASM disks that could be added to a disk group. With `ASM_DISKSTRING='/dev/emcpower*1'`, ASM will discover the first partition of all EMC partitioned devices in /dev that are owned by the GI owner, belong to the GI install group, and have read/write privileges for owner and group:

```
# ls -ltr /dev/emcpower*1 | grep grid
brw-rw---- 1 grid oinstall 120,  81 Apr 13 06:20 /dev/emcpowerf1
brw-rw---- 1 grid oinstall 120,  33 Apr 13 06:20 /dev/emcpowerc1
brw-rw---- 1 grid oinstall 120,  97 Apr 13 06:20 /dev/emcpowerg1
brw-rw---- 1 grid oinstall 120,  17 Apr 13 06:20 /dev/emcpowerb1
brw-rw---- 1 grid oinstall 120, 161 Apr 13 06:40 /dev/emcpowerk1
```

```
SQL> select group_number, disk_number, … from v$asm_disk;

Grp Dsk Mount   Mode
Num Num Status  Status  Name            Path
--- --- ------- ------- --------------- --------------------
  0   1 CLOSED  ONLINE                  /dev/emcpowerg1
  0   2 CLOSED  ONLINE                  /dev/emcpowerf1
  0   3 CLOSED  ONLINE                  /dev/emcpowerc1
  0   4 CLOSED  ONLINE                  /dev/emcpowerb1
  1   0 CACHED  ONLINE  CRS_0000        /dev/emcpowerk1
```

When identifying ASM disks during the creation of an ASM disk group, specify a disk string that identifies a subset of the disks returned by the ASM disk discovery process. Any disk that matches the specified search string will be added to the disk group. This can lead to ASM disks not having the same order in the ASM disk group as indicated by the SC volume name. For example, Figure 78 uses a disk string that identifies multiple pseudo-devices and it creates ASM disk misnomers: SC Series volume ora-data1 (pseudo-device emcpowerb) was intended to be stamped as the first ASM disk DATA_0000, but it was stamped as DATA_0001.

```
$ cat dg_data_config-UDEV-pp-multiple_disks.xml
<dg name="data" redundancy="external">
<dsk string="/dev/emcpower[bc]1"/>
<a name="compatible.asm" value="11.2"/>
<a name="compatible.rdbms" value="11.2"/>
</dg>
$ asmcmd lsdg
State    Type    Rebal  Sector  Block       AU  Total_MB  …   Name
MOUNTED  EXTERN  N         512   4096  1048576      5115  …   CRS/
$ asmcmd mkdg dg_data_config-UDEV-pp-multiple_disks.xml
$ asmcmd lsdg
State    Type    Rebal  Sector  Block       AU  Total_MB  …   Name
MOUNTED  EXTERN  N         512   4096  1048576      5115  …   CRS/
MOUNTED  EXTERN  N         512   4096  1048576    102398  …   DATA/


SQL> CREATE DISKGROUP DATA
  2   EXTERNAL REDUNDANCY
  3   DISK '/dev/emcpower[bc]1'
  4   ATTRIBUTE 'compatible.asm'='11.2'
  5           , 'compatible.rdbms'='11.2';

Diskgroup created.

SQL> select GROUP_NUMBER, DISK_NUMBER, … from v$asm_disk;

Grp Dsk Mount   Mode
Num Num Status  Status  Name            Path
--- --- ------- ------- --------------- --------------------
  0   1 CLOSED  ONLINE                  /dev/emcpowerg1
  0   2 CLOSED  ONLINE                  /dev/emcpowerf1
  1   0 CACHED  ONLINE  CRS_0000        /dev/emcpowerk1
  2   0 CACHED  ONLINE  DATA_0000       /dev/emcpowerc1
  2   1 CACHED  ONLINE  DATA_0001       /dev/emcpowerb1
```

Figure 78    Disk string identifies multiple PowerPath pseudo-devices with ASM disk misnomers

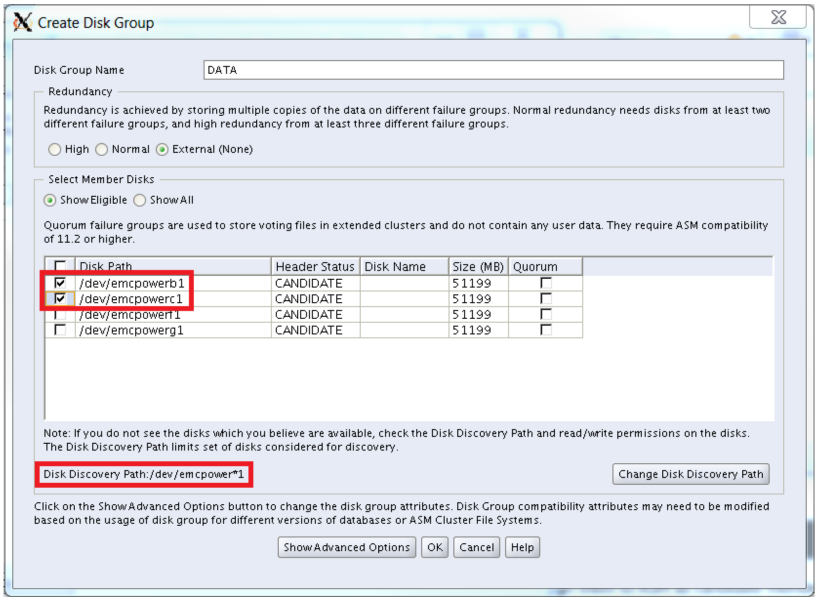
If SC Series volume names have an ordering that needs to be maintained within in the disk group, Dell EMC recommends the disk string uniquely identify each pseudo-device rather than a subset of pseudo-devices and then name each disk separately to ensure the correct disks are used. Figure 79 shows how to resolve the misnomer:

```
CREATE DISKGROUP DATA
  EXTERNAL REDUNDANCY
  DISK '/dev/emcpowerb1' NAME DATA_0000
     , '/dev/emcpowerc1' NAME DATA_0001
  ATTRIBUTE 'compatible.asm'='11.2'
          , 'compatible.rdbms'='11.2';
```

```
<dg name="fra" redundancy="external">
<dsk name="DATA_0000" string="/dev/emcpowerb1"/>
<dsk name="DATA_0001" string="/dev/emcpowerc1"/>
<a name="compatible.asm" value="11.2"/>
<a name="compatible.rdbms" value="11.2"/>
</dg>
```

Figure 79 PowerPath pseudo-devices, unique disk strings and disk names

If using asmca to create ASM disk groups, use the Disk Discovery Path that references the partitioned or unpartitioned pseudo-devices devices, select the check boxes associated with the appropriate pseudo-devices for the disk group, then name the disks appropriately:



Disk Group Name: DATA

Redundancy

Redundancy is achieved by storing multiple copies of the data on different failure groups. Normal redundancy needs disks from at least two different failure groups, and high redundancy from at least three different failure groups.

○ High ○ Normal ● External (None)

Select Member Disks

● Show Eligible ○ Show All

Quorum failure groups are used to store voting files in extended clusters and do not contain any user data. They require ASM compatibility of 11.2 or higher.

| □ | Disk Path | Header Status | Disk Name | Size (MB) | Quorum |
|---|---|---|---|---|---|
| ☑ | /dev/emcpowerb1 | CANDIDATE | | 51199 | □ |
| ☑ | /dev/emcpowerc1 | CANDIDATE | | 51199 | □ |
| □ | /dev/emcpowerf1 | CANDIDATE | | 51199 | □ |
| □ | /dev/emcpowerg1 | CANDIDATE | | 51199 | □ |

Note: If you do not see the disks which you believe are available, check the Disk Discovery Path and read/write permissions on the disks. The Disk Discovery Path limits set of disks considered for discovery.

Disk Discovery Path:/dev/emcpower*1

Click on the Show Advanced Options button to change the disk group attributes. Disk Group compatibility attributes may need to be modified based on the usage of disk group for different versions of databases or ASM Cluster File Systems.

When identifying ASM disks during the creation of an ASM disk group, use the absolute path of and the partitioned pseudo-device names, `/dev/emcpower<device><partition-indicator>`, and not the partitioned Linux device names (`/dev/sd<x><p>`).

### 6.5.5 ASM diskgroups with DM-Multipath and UDEV

This section refers to partitioned logical DM-Multipath aliased devices. If unpartitioned devices are used, simply remove any reference to the partition-indicator value from the remainder of this section.

ASM instance initialization parameter ASM_DISKSTRING is used by ASM to discover all candidate ASM disks that could be added to a disk group. With ASM_DISKSTRING='/dev/mapper/oraasm*p1', ASM will discover all DM-Multipath partitioned devices that are owned by the GI owner, belong to the GI install group, and have read/write privileges for owner and group:

```
[node1]# ls -ltr /dev/mapper/oraasm*p1
lrwxrwxrwx 1 root root 8 Apr 14 15:17 /dev/mapper/oraasm-fra1p1 -> ../dm-10
lrwxrwxrwx 1 root root 8 Apr 14 15:17 /dev/mapper/oraasm-fra2p1 -> ../dm-14
lrwxrwxrwx 1 root root 8 Apr 14 15:17 /dev/mapper/oraasm-crs1p1 -> ../dm-13
lrwxrwxrwx 1 root root 7 Apr 14 15:17 /dev/mapper/oraasm-data2p1 -> ../dm-8
lrwxrwxrwx 1 root root 7 Apr 14 15:17 /dev/mapper/oraasm-data1p1 -> ../dm-9
[node1]# ls -ltr /dev | egrep "dm-[89]|dm-1[034]"
brw-rw----   1 grid oinstall 252,   9 Apr 17 12:33 dm-9
brw-rw----   1 grid oinstall 252,   8 Apr 17 12:33 dm-8
brw-rw----   1 grid oinstall 252,  10 Apr 17 12:33 dm-10
brw-rw----   1 grid oinstall 252,  14 Apr 17 12:33 dm-14
brw-rw----   1 grid oinstall 252,  13 Apr 17 14:39 dm-13
[root@oradef3 ~]#



SQL> select group_number, disk_number, … from v$asm_disk;

Grp Dsk Mount   Mode
Num Num Status  Status  Name            Path
--- --- ------- ------- --------------- ----------------------------
  0   0 CLOSED  ONLINE                  /dev/mapper/oraasm-fra2p1
  0   2 CLOSED  ONLINE                  /dev/mapper/oraasm-data1p1
  0   3 CLOSED  ONLINE                  /dev/mapper/oraasm-data2p1
  0   4 CLOSED  ONLINE                  /dev/mapper/oraasm-fra1p1
  1   0 CACHED  ONLINE  CRS_0000        /dev/mapper/oraasm-crs1p1

SQL>
```

When identifying ASM disks during the creation of an ASM disk group, specify a disk string that identifies a subset of the disks returned by the ASM disk discovery process. Any disk that matches the specified search string will be added to the disk group. This can lead to ASM disks not having the same order in the ASM disk group as indicated by the SC volume name. For example, Figure 80 uses a disk string that identifies multiple pseudo-devices and it creates ASM disk misnomers: SC Series volume ora-data1 (pseudo-device emcpowerb) was intended to be stamped as the first ASM disk DATA_0000 and not DATA_0001.

```
$ cat dg_data_config-UDEV-DM-Multipath_disks.xml
<dg name="data" redundancy="external">
<dsk string="/dev/mapper/oraasm-data*p1"/>
<a name="compatible.asm" value="11.2"/>
<a name="compatible.rdbms" value="11.2"/>
</dg>
$ asmcmd lsdg
State    Type    Rebal  Sector  Block       AU  Total_MB  …   Name
MOUNTED  EXTERN  N         512   4096  1048576      5115  …   CRS/
$ asmcmd mkdg dg_data_config-UDEV-DM-Multipath_disks.xml
$ asmcmd lsdg
State    Type    Rebal  Sector  Block       AU  Total_MB  …   Name
MOUNTED  EXTERN  N         512   4096  1048576      5115  …   CRS/
MOUNTED  EXTERN  N         512   4096  1048576    102398  …   DATA/
$
```

```
SQL> CREATE DISKGROUP DATA
  2   EXTERNAL REDUNDANCY
  3   DISK '/dev/mapper/oraasm-data*p1'
  4   ATTRIBUTE 'compatible.asm'='11.2'
  5           , 'compatible.rdbms'='11.2'
  6  /

Diskgroup created.

SQL> select GROUP_NUMBER, DISK_NUMBER, … from v$asm_disk;

Grp Dsk Mount   Mode
Num Num Status  Status  Name            Path
--- --- ------- ------- --------------- ----------------------------
  0   0 CLOSED  ONLINE                  /dev/mapper/oraasm-fra2p1
  0   4 CLOSED  ONLINE                  /dev/mapper/oraasm-fra1p1
  1   0 CACHED  ONLINE  CRS_0000        /dev/mapper/oraasm-crs1p1
  2   0 CACHED  ONLINE  DATA_0000       /dev/mapper/oraasm-data2p1
  2   1 CACHED  ONLINE  DATA_0001       /dev/mapper/oraasm-data1p1
```

Figure 80 Disk string identifies multiple DM-Multiapth pseudo-devices with ASM disk misnomers

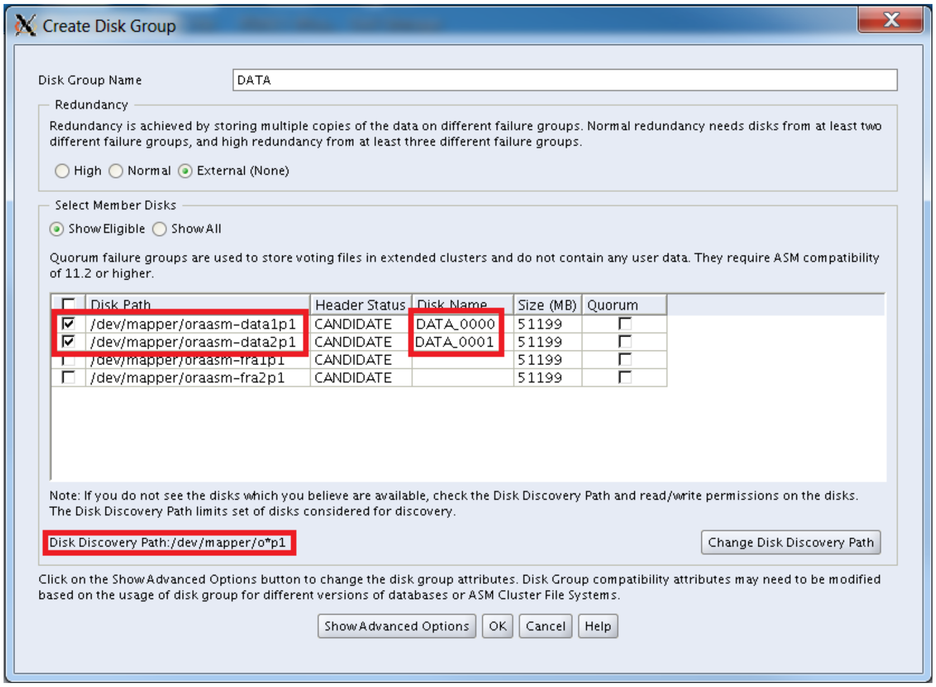
If SC Series volume names have an ordering that needs to be maintained within in the disk group, Dell EMC recommends the disk string uniquely identify each pseudo-device rather than a subset of pseudo-devices and then name each disk separately to ensure the correct disks are used. Figure 81 shows how to resolve the misnomer:

```
CREATE DISKGROUP DATA
  EXTERNAL REDUNDANCY
  DISK '/dev/mapper/oraasm-data1p1' NAME DATA_0000
     , '/dev/mapper/oraasm-data2p1' NAME DATA_0001
  ATTRIBUTE 'compatible.asm'='11.2'
          , 'compatible.rdbms'='11.2'
/
```

```
<dg name="fra" redundancy="external">
<dsk name="DATA_0000" string="/dev/mapper/oraasm-data1p1"/>
<dsk name="DATA_0001" string="/dev/mapper/oraasm-data2p1"/>
<a name="compatible.asm" value="11.2"/>
<a name="compatible.rdbms" value="11.2"/>
</dg>
```

Figure 81 DM-Multipath pseudo-devices, unique disk strings and disk names

If using asmca to create ASM disk groups, use the Disk Discovery Path that references the partitioned logical DM-Multipath aliased devices, select the check boxes associated with the appropriate partitioned pseudo-devices for the disk group, then name the disks appropriately:



Disk Group Name: DATA

Redundancy

Redundancy is achieved by storing multiple copies of the data on different failure groups. Normal redundancy needs disks from at least two different failure groups, and high redundancy from at least three different failure groups.

○ High ○ Normal ● External (None)

Select Member Disks

● Show Eligible ○ Show All

Quorum failure groups are used to store voting files in extended clusters and do not contain any user data. They require ASM compatibility of 11.2 or higher.

| □ | Disk Path | Header Status | Disk Name | Size (MB) | Quorum |
|---|---|---|---|---|---|
| ☑ | /dev/mapper/oraasm-data1p1 | CANDIDATE | DATA_0000 | 51199 | □ |
| ☑ | /dev/mapper/oraasm-data2p1 | CANDIDATE | DATA_0001 | 51199 | □ |
| □ | /dev/mapper/oraasm-fra1p1 | CANDIDATE | | 51199 | □ |
| □ | /dev/mapper/oraasm-fra2p1 | CANDIDATE | | 51199 | □ |

Note: If you do not see the disks which you believe are available, check the Disk Discovery Path and read/write permissions on the disks. The Disk Discovery Path limits set of disks considered for discovery.

Disk Discovery Path:/dev/mapper/o*p1

Click on the Show Advanced Options button to change the disk group attributes. Disk Group compatibility attributes may need to be modified based on the usage of disk group for different versions of databases or ASM Cluster File Systems.

When identifying the ASM disks during the creation of a disk group in ASM, use the absolute path of and the partitioned pseudo-device names that represents the single logical path of the single ASM partitioned device:

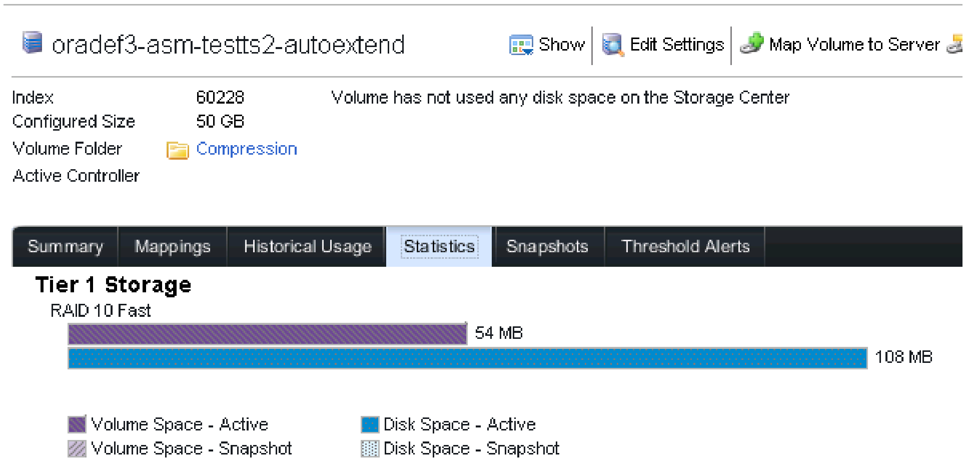
```
/dev/mapper/<alias><partition-indicator>
```

Do not use the `/dev/sd<x>` or `/dev/dm-<n>` name that make up the logical device. This is because dm-<n> and sd<x> device names are not consistent across reboots nor across servers in a RAC cluster.

## 6.6 ASM and SC Series thin provisioning

Just as SC Series storage provides thin provisioning on storage within the array, Oracle provides thin provisioning within a database when using an Oracle feature called autoextend. Autoextend provides the ability for Oracle to extend a database file should the need arise. To take advantage of database thin provisioning with SC Series storage, Dell EMC recommends using the Oracle autoextend feature.

The following example illustrates SC Series storage and autoextend working in concert when using ASM. The example uses a 50GB SC Series volume (oradef3-asm-testts2-autoextend) in tier 1 with RAID 10. The volume is added to Oracle diskgroup +TESTTS which was defined with external redundancy.



oradef3-asm-testts2-autoextend

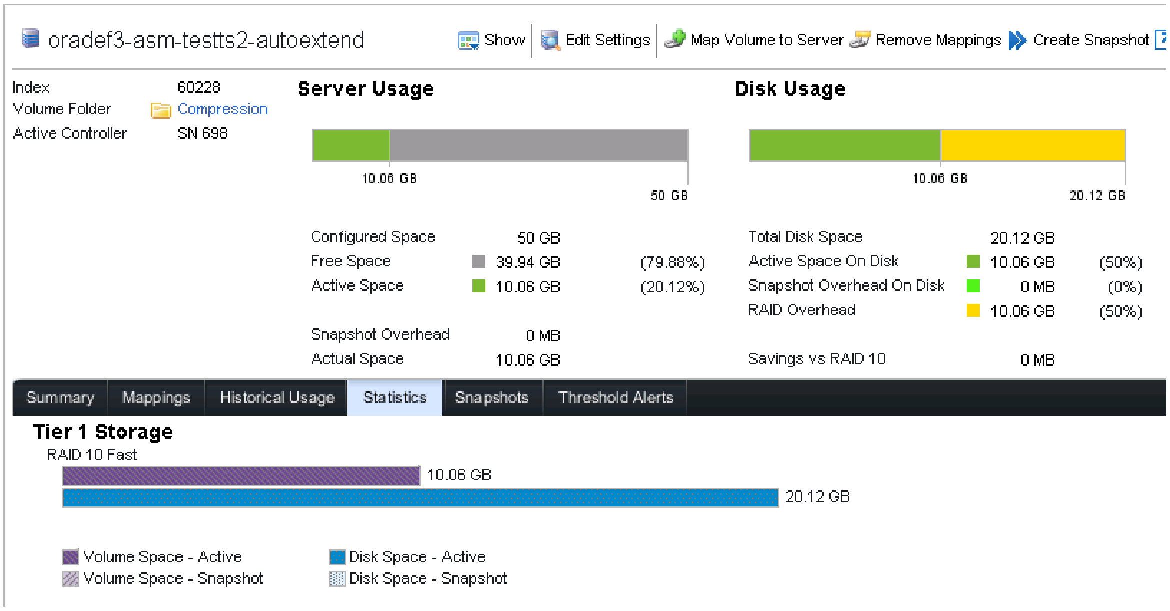
Index 60228 Volume has not used any disk space on the Storage Center
Configured Size 50 GB
Volume Folder Compression
Active Controller

Figure 82 50 GB SC Series volume after adding it to an ASM disk group +TESTTS.

After the LUN is added to the ASM disk group, tablespace TESTTS is created in the diskgroup:

```
CREATE BIGFILE TABLESPACE "TESTTS"
  DATAFILE '+TESTTS' SIZE 10G
  AUTOEXTEND ON
  NEXT 128K MAXSIZE UNLIMITED
  EXTENT MANAGEMENT LOCAL SEGMENT SPACE MANAGEMENT AUTO DEFAULT NOCOMPRESS
/
```

At this point, DSM displays the amount of space allocated to the SC Series volume equal to the sum of the size specified in the CREATE TABELSPACE command and any additional space needed for the RAID level of the storage tier. Since the tablespace is defined to be 10GB and RAID 10 redundancy is used, a total of 20GB physical disk space is allocated from the SC Series pagepool to this volume (see **Disk Usage** in Figure 83). The usable allocated space for the volume is shown under **Server Usage** in Figure 83:



oradef3-asm-testts2-autoextend

| | |
|---|---|
| Index | 60228 |
| Volume Folder | Compression |
| Active Controller | SN 698 |

**Server Usage**

| | | | |
|---|---|---|---|
| Configured Space | | 50 GB | |
| Free Space | ■ | 39.94 GB | (79.88%) |
| Active Space | ■ | 10.06 GB | (20.12%) |
| Snapshot Overhead | | 0 MB | |
| Actual Space | | 10.06 GB | |

**Disk Usage**

| | | | |
|---|---|---|---|
| Total Disk Space | | 20.12 GB | |
| Active Space On Disk | ■ | 10.06 GB | (50%) |
| Snapshot Overhead On Disk | ■ | 0 MB | (0%) |
| RAID Overhead | ■ | 10.06 GB | (50%) |
| Savings vs RAID 10 | | 0 MB | |

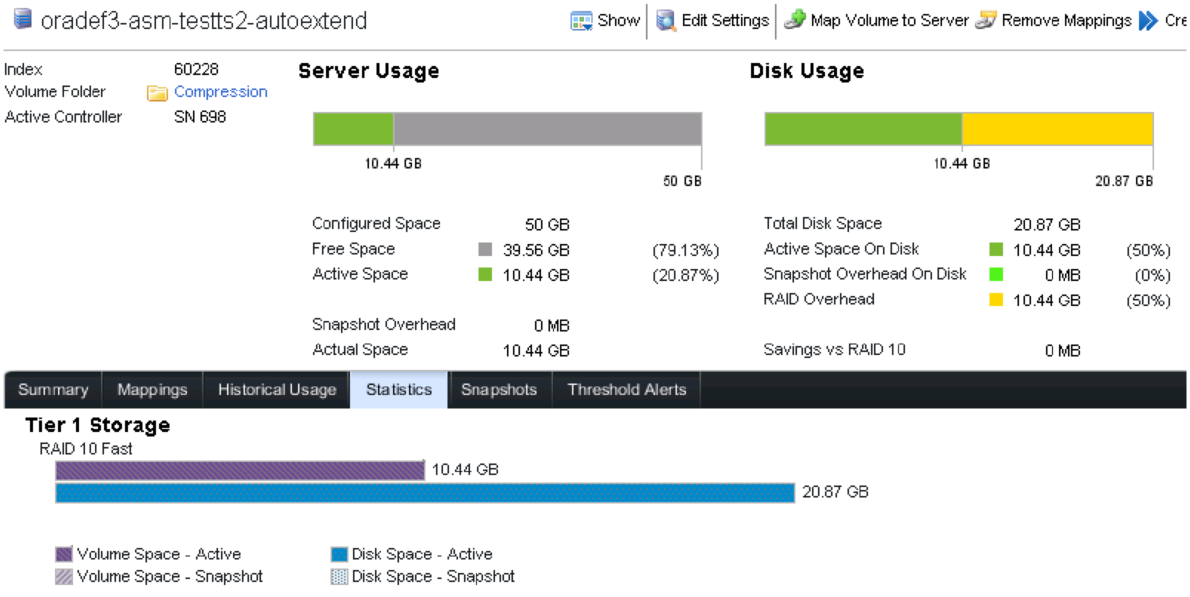
**Tier 1 Storage**

RAID 10 Fast

▨ Volume Space - Active
▨ Volume Space - Snapshot
■ Disk Space - Active
▨ Disk Space - Snapshot

Figure 83 SC Series volume statistics: Server Usage (logical disk usage) and Disk Usage (physical disk usage)

After completely filling up the 10GB of allocated space in the LUN with user data, Figure 84 shows an additional 384MB allocated to the volume after more data was added to the tablespace.



oradef3-asm-testts2-autoextend

| | |
|---|---|
| Index | 60228 |
| Volume Folder | Compression |
| Active Controller | SN 698 |

**Server Usage**

| | | | |
|---|---|---|---|
| Configured Space | | 50 GB | |
| Free Space | ■ | 39.56 GB | (79.13%) |
| Active Space | ■ | 10.44 GB | (20.87%) |
| Snapshot Overhead | | 0 MB | |
| Actual Space | | 10.44 GB | |

**Disk Usage**

| | | | |
|---|---|---|---|
| Total Disk Space | | 20.87 GB | |
| Active Space On Disk | ■ | 10.44 GB | (50%) |
| Snapshot Overhead On Disk | ■ | 0 MB | (0%) |
| RAID Overhead | ■ | 10.44 GB | (50%) |
| Savings vs RAID 10 | | 0 MB | |

**Tier 1 Storage**

RAID 10 Fast

▨ Volume Space - Active
▨ Volume Space - Snapshot
■ Disk Space - Active
▨ Disk Space - Snapshot

Figure 84 Thin provisioning an additional 348MB

## 6.7 ASM and SC Series thin reclamation

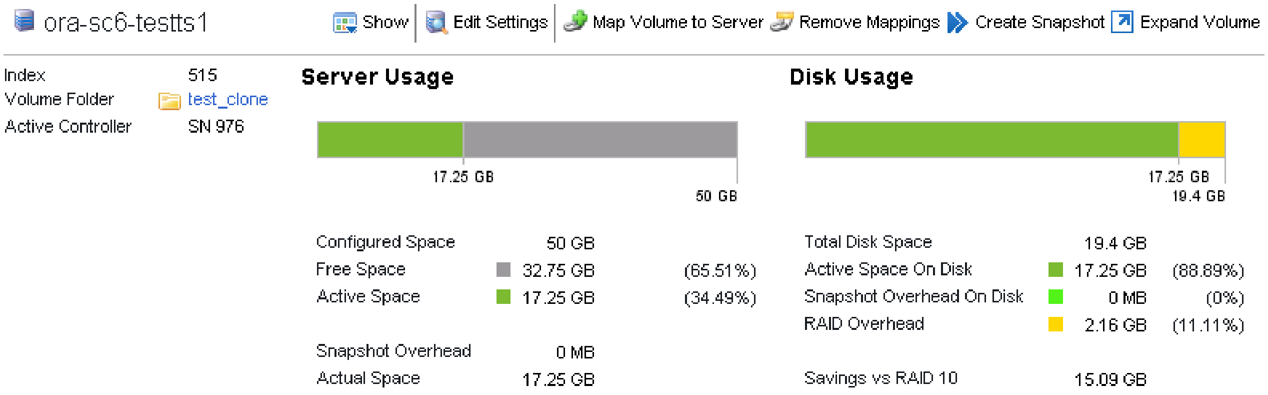
Over time, utilization of allocated space within a thinly provisioned LUN can decrease as changes are made to the database through operations like:

- Dropping a tablespace
- Resizing an oracle datafile after shrinking a tablespace
- Dropping a database

In operations like these, when space is released from an ASM disk group, space that had previously been allocated inside the array is also deallocated, provided that SCOS is at version 6.5.1 or later and one of the following conditions are met:

- If Oracle ASM Space Reclamation Utility (ASRU) is used
- If Oracle 12c and ASMFD are used with ASM diskgroups created with attribute THIN_PROVISIONED set to true

The following example illustrates how Oracle ASRU in conjunction with SC Series storage can be used to deallocated thinly provisioned storage in SC Series storage when it is no longer needed in Oracle ASM.



ora-sc6-testts1

| | |
|---|---|
| Index | 515 |
| Volume Folder | test_clone |
| Active Controller | SN 976 |

**Server Usage**

| | | | |
|---|---|---|---|
| Configured Space | | 50 GB | |
| Free Space | ■ | 32.75 GB | (65.51%) |
| Active Space | 🟩 | 17.25 GB | (34.49%) |
| Snapshot Overhead | | 0 MB | |
| Actual Space | | 17.25 GB | |

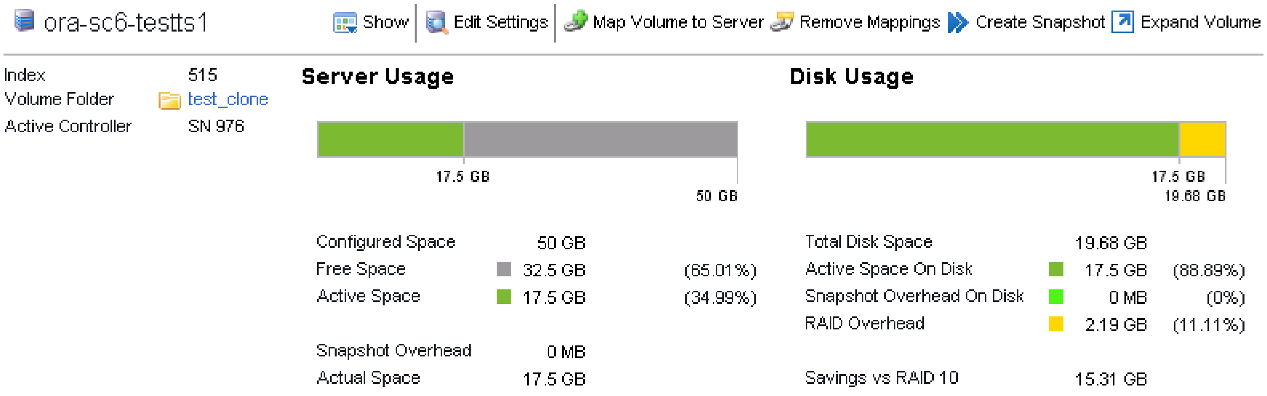
**Disk Usage**

| | | | |
|---|---|---|---|
| Total Disk Space | | 19.4 GB | |
| Active Space On Disk | 🟩 | 17.25 GB | (88.89%) |
| Snapshot Overhead On Disk | 🟩 | 0 MB | (0%) |
| RAID Overhead | 🟨 | 2.16 GB | (11.11%) |
| Savings vs RAID 10 | | 15.09 GB | |

Figure 85 SC Series volume (ora-sc6-testts1) containing 17.25 GB of data in Oracle tablespace TESTTS

After dropping tablespace TESTTS, which is consuming 17.25 GB of data, SC Series storage still shows the 17.25 GB of space still being consumed by the volume. See Figure 86.

```
SQL> drop tablespace testts including contents and datafiles;

Tablespace dropped.

SQL>
```



| | |
|---|---|
| Index | 515 |
| Volume Folder | test_clone |
| Active Controller | SN 976 |

**Server Usage**

| | | | |
|---|---|---|---|
| Configured Space | | 50 GB | |
| Free Space | ▨ | 32.5 GB | (65.01%) |
| Active Space | 🟩 | 17.5 GB | (34.99%) |
| Snapshot Overhead | | 0 MB | |
| Actual Space | | 17.5 GB | |

**Disk Usage**

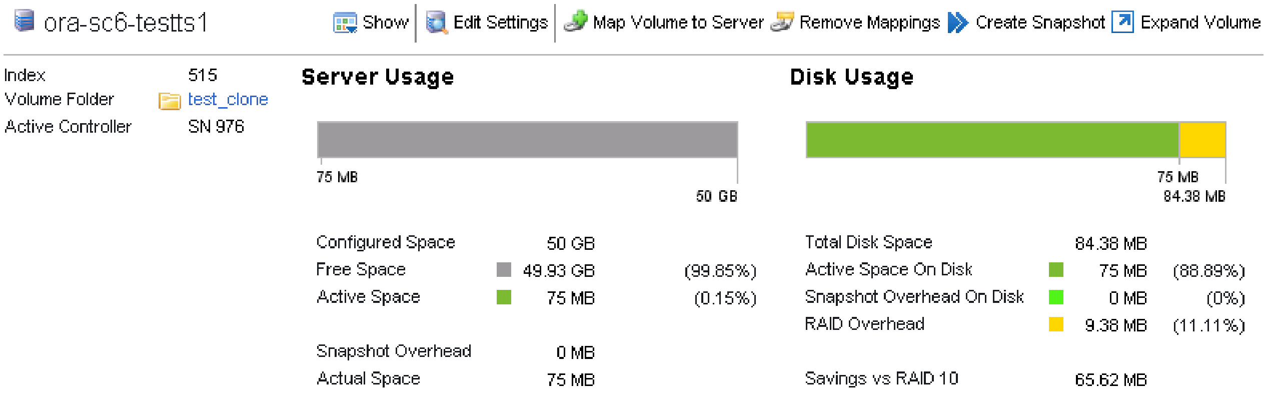
| | | | |
|---|---|---|---|
| Total Disk Space | | 19.68 GB | |
| Active Space On Disk | 🟩 | 17.5 GB | (88.89%) |
| Snapshot Overhead On Disk | 🟩 | 0 MB | (0%) |
| RAID Overhead | 🟨 | 2.19 GB | (11.11%) |
| Savings vs RAID 10 | | 15.31 GB | |

Figure 86 Released ASM space is not deallocated without calling ASRU

To instruct SC Series arrays to release the space, ASRU must be called to write zeros to the space once occupied by tablespace TESTTS.

```
$ /u01/app/11.2.0/grid/perl/bin/perl ./ASRU.pl TESTTS
Checking the system ...done
Calculating the sizes of the disks ...done
Writing the data to a file ...done
Resizing the disks...done
Calculating the sizes of the disks ...done

/u01/app/11.2.0/grid/perl/bin/perl -I /u01/app/11.2.0/grid/perl/lib/5.10.0
/u01/sw_grid/ASRU/zerofill 1 /dev/mapper/oraasm-testts1 75 51200
51125+0 records in
51125+0 records out
53608448000 bytes (54 GB) copied, 65.1528 s, 823 MB/s

Calculating the sizes of the disks ...done
Resizing the disks...done
Calculating the sizes of the disks ...done
Dropping the file ...done
$
```

After executing ASRU, thin reclamation will occur in the SC Series array (see Figure 87)



| | |
|---|---|
| Index | 515 |
| Volume Folder | test_clone |
| Active Controller | SN 976 |

**Server Usage**

| | | | |
|---|---|---|---|
| Configured Space | | 50 GB | |
| Free Space | ▨ | 49.93 GB | (99.85%) |
| Active Space | 🟩 | 75 MB | (0.15%) |
| Snapshot Overhead | | 0 MB | |
| Actual Space | | 75 MB | |

**Disk Usage**

| | | | |
|---|---|---|---|
| Total Disk Space | | 84.38 MB | |
| Active Space On Disk | 🟩 | 75 MB | (88.89%) |
| Snapshot Overhead On Disk | 🟩 | 0 MB | (0%) |
| RAID Overhead | 🟨 | 9.38 MB | (11.11%) |
| Savings vs RAID 10 | | 65.62 MB | |

Figure 87 Thinly provisioned space used by ASM has been deallocated in SC Series storage

If it is necessary to deallocate thinly provisioned storage, Oracle recommends using ASMFD or ASRU, with a preference towards ASMFD, providing their usage has been evaluated and any side effects from using them are understood. For information on installing ASRU, its configuration, issues or usage exceptions, see information on the My Oracle Support site. If ASMFD or ASRU are not used, space can be reclaimed with data pump export, creating new SC Series volumes and assigning them to the ASM diskgroup, dropping the old LUNs from the disk groups, and then importing the data pump export.

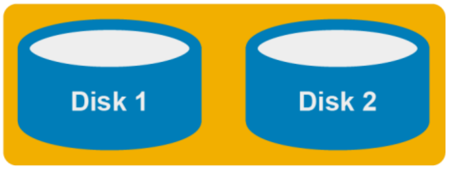
## 6.8 Oracle ASM on a SC Series array

The following diagram illustrates an example configuration of SC Series storage profiles with Oracle ASM on an all-flash array. It is recommended to alter this configuration (storage profile and disk group placement) to fit the business requirements. Refer to sections 5.8 and 3.19 and Table 16 for placement of ASM disks in other SC series array configurations.

Storage profile: **Flash Optimized with Progression (Tier 1 to all Tiers)**

| Tier | Drive type | Writeable data | Snapshot data | Progression |
|---|---|---|---|---|
| 1 | SSD write-intensive | RAID 10 | RAID 10 | N/A |

**ASM DATA disk group**

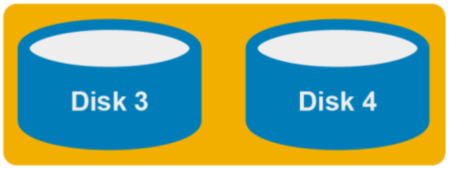


Frequently accessed files
Control files
Online redo logs

Storage profile: **Flash Only with Progression (Tier 1 to Tier 2)**

| Tier | Drive type | Writeable data | Snapshot data | Progression |
|---|---|---|---|---|
| 1 | SSD write-intensive | RAID 10 | RAID 10 | N/A |
| 2 | SSD read-intensive | N/A | RAID 5 | RAID 5 |

**ASM FRA disk group**



Active redo logs
Flashback logs
Multiplexed control files
Multiplexed online redo logs
RMANJ backup sets

## 6.9 Oracle with cooked file systems

A cooked filesystem is a non-raw file system that contains data that describes the storage of data. Several cooked filesystems supported by Oracle are listed below:

- btrfs
- ext2, ext3, and ext4
- ocfs and ocfs2
- xFS and VxFS
- gfs
- nfs and dnfs
- ReiserFS
- vfat

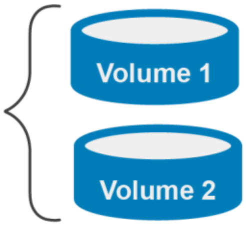
For a list of supported file systems supported by Oracle, see Doc ID 236826.1 on My Oracle Support and *Oracle Linux Administrator's Guide*.

## 6.10 Cooked filesystems with spinning media

When using cooked files systems, consider the following file placements as a starting point.

Storage profile: **High Priority (Tier 1)**

| Tier | Drive type | Writeable data | Snapshot data | Progression |
|---|---|---|---|---|
| 1 | FC/SAS 15K | RAID 10 | RAID 5, 6 | N/A |

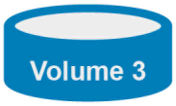


**Tablespaces:**
system, sysaux,
temp, undo

**Other tablespaces:**
X, Y, Z, control files,
online redo logs

Storage profile: **Recommended (all tiers)**

| Tier | Drive type | Writeable data | Snapshot data | Progression |
|---|---|---|---|---|
| 1 | FC/SAS 15K | RAID 10 | RAID 5, 6 | N/A |
| 2 | FC/SAS 15K | RAID 10 | RAID 5, 6 | N/A |
| 3 | SATA | RAID 10 | RAID 5, 6 | N/A |



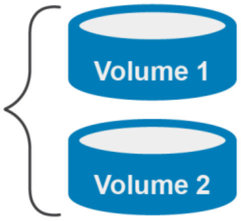
Archive redo logs,
flashback logs,
RMAN backup sets

## 6.11 Cooked filesystems with multiple SSD media types

In a hybrid SC Series array with multiple types of SSD media, consider the following file placement as a starting place.

Storage profile: **Flash Optimized with Progression (Tier 1 to all Tiers)**

| Tier | Drive type | Writeable data | Snapshot data | Progression |
|---|---|---|---|---|
| 1 | SSD write-intensive | RAID 10 | N/A | N/A |
| 2 | SSD read-intensive | N/A | RAID 5, 6 | N/A |
| 3 | FC/SAS 15K | N/A | RAID 5, 6 | RAID 5, 6 |

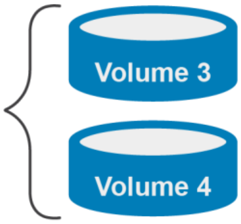


Online redo logs,
control files

Multiplexed control files,
multiplexed online redo files

Storage profile: **Flash Optimized with Progression (Tier 1 to all Tiers)**

| Tier | Drive type | Writeable data | Snapshot data | Progression |
|---|---|---|---|---|
| 1 | SSD write-intensive | RAID 10 | N/A | N/A |
| 2 | SSD read-intensive | N/A | RAID 5, 6 | N/A |
| 3 | FC/SAS 15K | N/A | RAID 5, 6 | RAID 5, 6 |

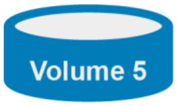


**Tablespaces:**
system, sysaux,
temp, undo

**Other tablespaces:**
X, Y, Z

Storage profile: **Flash Optimized with Progression (Tier 1 to all Tiers)**

| Tier | Drive type | Writeable data | Snapshot data | Progression |
|---|---|---|---|---|



Archive redo logs
flashback logs
RMAN backup sets

| | | | | |
|---|---|---|---|---|
| 1 | SSD write-intensive | RAID 10 | N/A | N/A |
| 2 | SSD read-intensive | N/A | RAID 5, 6 | N/A |
| 3 | FC/SAS 15K | N/A | RAID 5, 6 | RAID 5, 6 |

## 6.12 Direct I/O and async I/O

Oracle recommends using both direct I/O and async I/O. For information on enabling direct and async I/O in the Oracle environment, refer to Oracle documentation.

## 6.13 Raw devices and Oracle

Starting with Oracle 11g, Oracle began the process of desupporting raw storage devices. With Oracle 11gR2, raw storage could no longer be used by Oracle Universal Installer (OUI) for new installations of Oracle clusterware, and DBCA could no longer be used to store database files on raw devices. But raw devices were supported only when upgrading an existing installation using the partitions already configured.

For additional information on using raw devices, see information published on the My Oracle Support.

# 7 Conclusion

SC Series arrays provide a cost-effective storage solution regardless of the type of media used within the array. This becomes more apparent with SC Series arrays using flash that can outperform high-performance arrays of 15K spinning disks for the same cost. When adding Oracle ASM to the configuration, implementing and maintaining an Oracle deployment is even easier. The flexibility and agility of SC Series arrays allows an administrator to simply add or remove disks based on the required workload, and Oracle ASM eliminates the need for third-party volume management for database files. This allows DBAs and the system administrators more time to concentrate on other important tasks and reduce overlapping responsibilities.

Data in SC Series volumes are automatically striped depending on which RAID level has been selected for the volume. If ASM is not used, use an OS striping mechanism (for example, LVM, or VxVM) to get better performance because of multiple disk queues at the OS level.

In summary, use the recommendations outlined in this document for improving manageability and performance of the SC Series array and Oracle environment:

- Configure I/O for bandwidth and not capacity.
- Use direct I/O and async I/O options; set filesystemio_options to setall or DirectIO.
- Use an even number of LUNs per database per logical file grouping.
- SC Series volumes for ASM should follow the Oracle recommendation on LUN sizing.
- When using only SSDs, disable SC Series write cache globally or on SC Series database volumes.
- Use the default storage profiles, unless detailed analysis dictates otherwise.
- Use Dell EMC PowerPath or native Linux device mapper.
- For device persistence on ASM disks, use either ASMLib, ASMFD, or UDEV.
- Use Oracle ASM and Oracle managed files.
- Use ASM external redundancy, unless business needs require normal or high redundancy.
- Configure ASM disk groups to sustain striping far and wide and to reduce Linux kernel contention accessing and queuing for the same disk.
- Perform due diligence on I/O system tests before database implementation.
- Assign all volumes used by a specific database to the same consistent snapshot profile dedicated to the database.
- For SC Series array sizing, consider:
  - The number of snapshots retained on the array and how the snapshots will be used
  - Type of ASM redundancy that will be used: 1x size of disk for external, 3x size of disk for normal ASM redundancy, and 5x size of disk for high ASM redundancy
  - RAID penalty and storage profiles used by the ASM disk groups

# A Oracle performance testing with ORION

The ORION (ORacle I/O Numbers) calibration tool is similar to the Iometer tool developed by the Intel® Corporation and dd for Linux. Like Iometer and dd, ORION runs I/O performance tests, however it is tailored to measure Oracle RDBMS workloads using the same software stack that Oracle uses. It is not required that Oracle or even a database be installed in order to use ORION.

ORION is available in Windows, Linux, and Solaris platforms for use in predicting the performance of an Oracle database, typically prior to installing Oracle. ORION is primarily made available for use in a pre-production or test environment because the use of ORION in a live environment is inherently dangerous. This is due to the fact that the testing of writes is potentially destructive to existing data.

ORION was made available as a separate download with Oracle 10g, but has now been included in Oracle 11g. ORION can be used to test Storage Area Networks, Network-Attached Storage or Direct-Attached Storage. Configure the test to reflect the specific hardware and software for more accurate predictive results than the preconfigured test profiles. This will also help determine the type of database needed. Typically, databases fall into two main categories (OLTP and OLAP, described as follows), depending upon how they are used. It is very important to determine which type is needed so that the ORION test can be tailored accordingly and produce meaningful results.

**OLTP:** The Online Transaction Processing (OLTP) database, characterized by many small transactions (often 8k), reads and writes in rapid succession and allows multi-user access. An OLTP database generally consists of highly normalized data ordered with a non-clustered index. This can be thought of as a bank ATM database. The databases are optimized to maximize the speed of a myriad of small transactions; the throughput is generally scrutinized according to the number of I/Os per Second (IOPS).

**OLAP:** The Online Analytical Processing (OLAP) database typically processes fewer and much larger (1MB) transactions, primarily reads. With an OLAP database, data is generally de-normalized, often using a star schema and organized with a clustered index. An example would be a data warehouse, used by business intelligence for data mining or reporting. These databases are designed to run complex queries on large amounts of data as efficiently as possible. Throughput considerations here focus on Megabytes per Second (MBPS). Often, a single database provides both functionalities. The ORION workload options can be configured to test the performance expected from a database that has to bear such a mixed workload.

Getting started with ORION is relatively simple. With 11g and beyond, ORION is part of the Oracle RDBMS install directory, so there is no need to download the software.

To begin, log in to the Oracle Linux account and change the working directory for the source of the test runs and results. Pick a unique name for the test run (iotest in this example) and create a file named **iotest.lun** in the working directory. Determine which LUNs or volumes will be used for the test and add them to file iotest.lun. Make sure each volume is on a separate line. Each line must not contain a trailing terminator and must be terminated with a carriage return. The resulting contents of the file iotest.lun may be similar to this:

```
/dev/raw/test1
/dev/raw/test2
/dev/raw/test3
/dev/raw/test4
```

**Note:** Tests which include a write component can destroy data. Never include a volume that contains important data to the source LUN file (iotest.lun). Rather, create new volumes for testing purposes. Check to be sure Oracle software has access to all the test volumes listed in iotest.lun. If proper access is not granted, ORION will fail.

Since the ORION test is dependent on async I/O, ensure the platform is capable of async I/O. With Windows, this capability is built in, but in Linux or Solaris, library libaio needs to be in one of the standard lib directories on the library path within the environment. From the command line, navigate to the ORION application directory and run the test. There are three mandatory parameters: run, testname, and num_disks (see Table 18).

Table 18 Mandatory parameters

| Parameter | Description |
|---|---|
| -run | Type of workload to run:<br>• simple: random small (8k) I/Os, then tests random large (1MB) I/Os<br>• normal: random small (8k) and random large (1MB) I/Os simultaneously<br>• oltp: random small (8k) only<br>• dss: (decision support system) tests random large (1MB) only<br>• advanced: run workload with parameters specified by user |
| -testname | The unique name of the test run; change each run or output files get overwritten |
| -num_disks | The number of actual disks (spindles); if not specified, it will default to number of volumes in .lun input file |

A minimalistic test run on volume(s) comprising of five physical disks would look like this:

```
orion -run simple -testname iotest -num_disks 5
```

The output would be similar to this:

```
ORION: ORacle IO Numbers -- Version x.x.x.x.x
Test will take approximately xx minutes
Larger caches may take longer
```

When the test is complete, the results will exist in the test output files listed in Table 19.

Table 19 ORION output files

| File | Description |
|---|---|
| iotest_summary.txt | Recap input parameters + latency, MBPS and IOPS results. |
| iotest_mbps.csv | Large I/Os in MBPS , performance results |
| iotest_iops.csv | Small I/Os in IOPS, performance results |
| iotest_lat.csv | Latency of small I/Os |
| iotest_tradeoff.csv | Large MBPS / small IOPS combinations possible with minimal latency |
| iotest_trace.txt | Extended, unprocessed output |

The .txt files are text files with summary data and unprocessed data, respectively. The summary text file is useful, providing the maximum value for each I/O type.

The .csv files are a list of comma-separated values that can be loaded into Microsoft Excel, or another database application, and then converted into a graph for easier evaluation of results.

An ORION test can be tailored more closely to the environmental variables by using the advanced designator with the `–run` option. Available options are listed in Table 20.

Table 20 ORION environment testing variables

| File | Description |
|---|---|
| -size_small | Small I/O size in KB (default = 8) |
| -size_large | Large I/O size in KB (default = 1024) |
| -type | Type of large I/Os |
| rand: random (default) | Seq: sequential streams |
| -num_streamIO | Concurrent I/Os per stream; for example, Threads (default = 1)<br>Number of CPUs x number of parallel threads per CPU<br>**Note:** This only works if type is specified as seq |
| -simulate | How ORION forms the virtual test volumes:<br>• concat: serial concatenation of volumes (default) if devices are already striped<br>• raid0: raid 0 mapping across all volumes if devices are not already striped. raid0 stripe size in KB (default = 1024) |
| -stripe | |

| File | Description |
|---|---|
| -write | Percentage of writes (default = 0); a typical percentage would be 'write 20'<br>**Note**: Write tests destroy data on volumes. |
| -cache_size | Size in MB of an array cache.<br>cache_size 0 disables cache warming (recommended) |
| -duration | Duration in seconds for data point (default = 60)<br>**Note:** If –num_disks > 30, set –duration 120 |
| -matrix | Defines the data points to test:<br>• basic: test first row and first column<br>• detailed: entire matrix tested<br>• col: defined small I/O load with varying large I/O load (-num_small)<br>• row: defined large I/O load with varying small I/O load (-num_large)<br>• max: tests varying small and large I/O loads up to set limits (-num_small, -num_large) |
| -num_small | Number of outstanding small I/Os ( matrix must be point, col, or max) |
| -num_large | Number of outstanding large I/Os (matrix must be point, col, or max) |
| -verbose | Prints tracing information to standard output (default = not set) |
| -datainput | By default, ORION writes zeros. This can lead to inaccurate performance metric. To force ORION to write non-zeros, set `–datainput` to a file containing data that should be used for writes by ORION. This is only available with ORION that comes with the 12c distribution. |
| -storax | Specifies the API to use for I/O testing. Values are:<br>• skgfr: Use operating system I/O layer<br>• asmlib: Use ASMLib disk devices based storage API for I/O |
| -datainput | Available starting in 12c. Specified a file containing the data that should be written by ORION during write requests. Without this option, ORION writes zero which can skew test results. See Oracle documentation for additional information. |

An elaborate custom workload test setting in ORION is shown in this example:

```
orion -run advanced -testname iotest -num_disks 5 \
-simulate raid0 -stripe 1024 -write 20 -type seq \
-matrix row -num_large 0
```

To force ORION (12c) to write non-zeros, use the following steps:

```
base64 /dev/urandom | head -c 100000000 > file.txt
nohup /u01/app/oracle/product/11.2.0/dbhome_1/bin/orion -run advanced -matrix
col -num_small 128 -size_large 1024 -write 90 -type rand -duration 60 -testname
run54a -datainput file.txt
```

**Note:** To get useful test results, it is often advisable to disable the array cache and set `-cache_size 0`, or at least set the cache size variable to a fraction of the actual cache size. This is because in a production environment, Oracle rarely monopolizes the entire cache, since there are other production demands. Testing with access the array cache fully enabled and dedicated will tend to yield results that are too good to be meaningful outside a protected test environment.

The ORION tool is meant to be used to test a pre-production environment, prior to installing Oracle and does not require that Oracle or a database be in place at the time of testing. ORION also tests writes as well as reads; use caution as the data residing on the volumes will be destroyed.

A more detailed description of the use of ORION can be accessed from within ORION, as follows:

```
orion -help
```

Also, an excellent introduction to ORION can be found on the Oracle website at:
http://docs.oracle.com/cd/E11882_01/server.112/e16638/iodesign.htm#BABFCFBC

The information in this appendix was compiled primarily with information from these two sources.

# B 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.

## B.1 Additional resources

Referenced or recommended Dell EMC publications:

- Dell EMC SC Series: Red Hat Enterprise Linux Best Practices
- Flash-optimized Data Progression
- Optimizing Dell EMC SC Series Storage for Oracle OLAP Processing
- Dell Storage Center OS 7.0 Data Reduction with Deduplication and Compression

Referenced or recommended Red Hat publications:

- Red Hat Enterprise Linux 6 Storage Administration Guide
- Red Hat Enterprise Linux 6 DM Multipath, DM Multipath Configuration and Administration
- Deploying Oracle Database 12c on RHEL 7 - Recommended Practices
- Deploying Oracle Database 12c on RHEL6 - Recommended Practices
- Deploying Oracle Database 11g on RHEL 6 - Recommended Practices
- Deploying Oracle RAC 11g R2 Database on RHEL 6 - Recommended Practices
- Red Hat customer portal support note 222473

Other recommended publications:

- Standard RAID levels

Other recommended publications: (Oracle manuals):

- Installing Oracle ASMLib
- Configuring Oracle ASMLib on Multipath DisksConfiguring Oracle ASMLib on Multipath Disks
- Using Oracle Database 10g's Automatic Storage Management with EMC Storage Technology.
- Oracle Database Storage Administrator's Guide 11g Release 1 (11.1)
- Oracle Automatic Storage Management, Administrator's Guide, 11gR2
- Oracle Automatic Storage Management, Administrator's Guide, 12cR1
- Oracle Database Concepts, 12c Release 1
- Oracle Database, Installation Guide, 11g Release 2 (11.2) for Linux, E47689-10
- Oracle Database Administrator's Reference 12c Release 1 (12.1) for Linux and UNIX-Based Operating Systems E10638-14
- Oracle Database Administrator's Guide 12c Release 1
- Oracle® Grid Infrastructure Installation Guide 11g Release 2 for Linux
- Oracle® Grid Infrastructure Installation Guide 12c Release 1 for Linux
- Oracle Database Data Warehousing Guide 12c Release 1 (12.1)
- Oracle Database Data Warehousing Guide 11g Release 1 (11.1)
- Oracle Linux Administrator's Guide for Release 6
- Oracle Linux Administrator's Guide for Release 7

Other recommended publications (with My Oracle Support Doc IDs):

- Oracle Support Doc 1077784.1: Can I create an 11.2 disk over the 2 TB limit
- Lun Size And Performance Impact With Asm (Doc ID 373242.1)
- Oracle Support Doc 1601759.1: Oracle Linux 5 — Filesystem & I/O Type Supportability
- Oracle Database — Filesystem & I/O Type Supportability on Oracle Linux 6
- Oracle Support Doc 1487957.1: ORA-1578 ORA-353 ORA-19599 Corrupt blocks with zeros when filesystemio_options=SETALL on ext4 file system using Linux ()
- ORA-15040, ORA-15066, ORA-15042 when ASM disk is not present in all nodes of a Rac Cluster. Adding a disk to the Diskgroup fails. (Doc ID 399500.1)
- How To Setup Partitioned Linux Block Devices Using UDEV (Non-ASMLIB) And Assign Them To ASM? (Doc ID 1528148.1)
- ASMFD (ASM Filter Driver) Support on OS Platforms (Certification Matrix). (Doc ID 2034681.1)
- How to Install ASM Filter Driver in a Linux Environment Without Having Previously Installed ASMLIB (Doc ID 2060259.1)
- How To Setup Partitioned Linux Block Devices Using UDEV (Non-ASMLIB) And Assign Them To ASM? (Doc ID 1528148.1)
- How To Setup ASM on Linux Using ASMLIB Disks, Raw Devices, Block Devices or UDEV Devices? (Doc ID 580153.1)
- Oracle Linux 6 - ASM Instances Fail with 4K Sector Size LUN (Doc ID 2211975.1)
- Supporting 4K Sector Disks [Video] (Doc ID 1133713.1)
- Supporting ASM on 4K/4096 Sector Size (SECTOR_SIZE) Disks(Doc ID 1630790.1)
- New Block Size Feature for Oracle ASM and oracleasm-support (Doc ID 1530578.1)
- Alert: After SAN Firmware Upgrade, ASM Diskgroups ( Using ASMLIB) Cannot Be Mounted Due To ORA-15085: ASM disk "" has inconsistent sector size. (Doc ID 1500460.1)
- Supported and Recommended File Systems on Linux (Doc ID 236826.1)
- RAC and Oracle Clusterware Best Practices and Starter Kit (Linux) (Doc ID 811306.1)