Advanced Concepts and Terminology

Module 9

Objectives

This module describes advanced concepts and terminology essential to understanding how the Enterprise Virtual Array virtualization software operates. It defines additional terms of the EVA and then explores more deeply the concepts behind the terms.

After completing this module, you should be able to:

- Describe distributed sparing.
- Describe disk failure protection levels.
- Describe redundant storage sets and their benefits.

Distributed sparing

Distributed sparing is the process of using available space in the disk group for several sparing functions. Because the EVA distributes capacity across all members of a disk group, no spare physical disks exist as with conventional arrays. The EVA uses the spare or unassigned capacity on all the disks within the disk group.

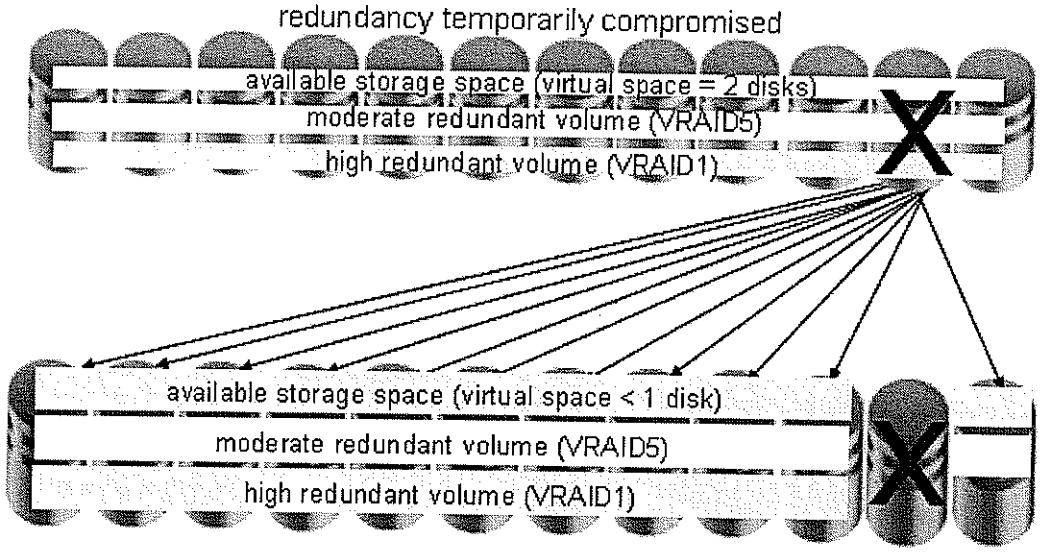
The EVA uses unassigned capacity as needed for items such as:

- Virtually Instantaneous Snapclones (Vsnaps) writes
- New virtual disk creation
- Freeing a physical disk for removal or reassignment
- Data reconstruction after disk failure or disks failures
  - Applies to VRAID1 and VRAID5
  - VRAID0 data is lost

Unassigned capacity increases when virtual disks are deleted or new physical disks are added to the disk group.

Data reconstruction

For VRAID5, XCS rebuilds failed (or removed) data on unassigned space using the parity information.



For VRAID1, the data is first copied to unassigned space on different disks from the surviving disk of the mirrored pair. The data is then moved from the surviving disk to the mirrored pairs of the recipient disks.

If one disk of a mirrored pair fails, the other can no longer be used for VRAID1. If a new disk is added, the surviving disk can be paired with it.

For VRAID1, two specific disks are always paired to decrease the likelihood of loosing the VRAID1. Both the paired disks must fail simultaneously. If copies of the data were randomly striped across any disk, eventually the mirrored disk would be linked to every disk in the disk group. Then, a simultaneous failure of any two disks would fail the VRAID1

Disk failure protection levels

Disk failure protection reserves unallocated disk space in a disk group to recover from physical disk failures. Distributed sparing uses free space first; the failure protection level merely reserves free space to ensure it is available for only disk recovery and not for virtual disk creation or Vsnap writes.

You reserve spare capacity by selecting a disk protection level when creating a disk group. The protection level can also be changed after the disk group has been established.

Three disk failure protection levels:

- None — Reserves no physical disk capacity. Instead, unassigned capacity, if available, is used for sparing.
- Single — Reserves enough unassigned spare capacity to recover from a **single** disk failure. The capacity of two disks must be reserved to recover a single VRAID1 disk if the disk group is full with only VRAID1 Vdisks.
- Double — Reserves enough unassigned spare capacity to recover from a **double** disk failure. The capacity of four disks must be reserved to recover from a double VRAID1 disk failure if the disk group is full with only VRAID1 Vdisks.

Protection levels do not assign specific capacity and do not change the amount of unassigned capacity, but ensure that new virtual disk creations and Vsnap writes do not use any of the free space needed for reconstruction. The capacity available for new Vdisks decreases if a higher protection level is selected for the disk group.

Protection level examples

Protection level with equal disk capacities

Computing maximum virtual disk sizes with same-sized disk capacities is straightforward; multiply the number of disks by the capacity, then subtract the number of disks required for the protection level.

Note

The capacities given in the examples are close to real life; they account for disk formatting, disk group metadata overhead, and virtual disk overhead. For example, a 36GB disk, when formatted, yields approximately 33.91GB of usable capacity.

Example 1

A disk group of 10 physical disks at 36GB each, with a protection level of **None**, has a maximum virtual disk size of 338GB.

Example 2

A disk group of 10 physical disks at 36GB each, with a protection level of **Single**, has a maximum virtual disk size of 270GB (338 – 68).

Example 3

A disk group of 10 physical disks at 36GB each, with a protection level of **Double**, has a maximum virtual disk size of 202GB (338 – 136).

Protection level with mixed disk capacities

When computing maximum virtual disk sizes for disk drives of mixed sizes, you must use the following rules:

- The Single protection level for a disk group requires two times the **largest** drive capacity of any drive in the group.
- The Double protection level for a disk group requires four times the **largest** drive capacity of any drive in the group.

Example

If a disk group has ten 36GB disks and one 72GB disk, and you need double protection, the space of four 72GB drives is set aside for spare capacity. The maximum virtual disk size would be 134GB (406GB – 272GB).

Redundant storage sets

A redundant storage set (RSS) is a subgrouping of drives within a disk group that reduce the risk of a multidisk failure resulting in the loss of user data.

Data is segmented into 8MB redundant stores (RStores). Physical data is allocated for an RStore based on the VRAID type—that is, 8MB for VRAID0, 10MB for VRAID5, and 16MB for VRAID1.

The physical storage for a given RStore is fully contained within a single RSS. Failed disk drive recovery is restricted to the affected RSS only which makes recovery more efficient by reducing the access to the number of disks.

Because all the data from an RStore is contained within an RSS, a VRAID5 could withstand multiple disk failures if those failed disks reside in different RSSs. The simultaneous failure of two disks within an RSS would fail a VRAID5.

Note

A VRAID1 can withstand multiple disk failures as long as two of its paired disks do not fail simultaneously.

Because RSSs are contained within a disk group, the more disks in the disk group, the more RSSs will be created. The more RSSs created, the more likely that two simultaneous disk failures will be in different RSSs. Therefore, in a VRAID5 environment, you should make your disk group as large as possible.

RSS calculations

The HSV controllers create RSSs automatically with no input from the user. The rules for RSS creation are as follows:

- Create 8-member RSSs until there are fewer than 16 drives.
- If more than 11 drives remain, create two evenly split RSSs, if possible. An RSS always has an even number of disks except for the remaining disks in the disk group. This supports VRAID1 pairing.
- If the number of remaining drives is 11 or fewer, create one RSS.
- If an RSS falls to fewer than six members, the system will try to merge it with another RSS.

Note

Disks within each RSS are identified by a zero-based number called an RSS index. The RSS and RSS index are displayed by Command View EVA 3.0 and higher.

The following are some examples of RSS calculations.

Example 1

A 38-disk disk group would consist of four 8-member RSSs and a 6-member RSS (8 + 8 + 8 + 8 + 6 = 38).

Example 2

A 50-member disk group would contain five 8-member RSSs and one 10-member RSS (8 + 8 + 8 + 8 + 8 + 10).

- This configuration could allow up to six drive failures before failing a VRAID5—if every failure occurred in a different RSS.
- This configuration could allow up to 25 drive failures before failing a VRAID1—if there were no simultaneous failures of paired disks and there was enough free space to reconstruct the failed drives.

Note

If a disk group member is removed (or fails), the RSSs are adjusted to ensure that each RSS has at least six members. This adjustment results in a re-leveling of the virtual disk's data.

Virtual disk behavior review

For all VRAID levels, space is allocated across all members of the disk group.

Consider the following behaviors for each VRAID type:

- VRAID0
  - Failure of any disk group member results in a virtual disk failure.
  - Unassigned capacity in the disk group does not help you.
- VRAID1
  - Each disk is paired (or married) to another specific disk.
  - A disk group with an odd number of disks (for example, a nine-member disk group) would have one unpaired disk. This means one disk would not have any data for the VRAID1 written to it.
  - HP recommends that disk groups have an even number of disks.
- VRAID5
  - Data is written in a 4+1 parity scheme.
  - Odd or even numbered RSS memberships do not affect the writing scheme.

Disk failures with VRAID1

If a disk group member failure occurs with a VRAID1 virtual disk, and assuming the disk group has spare space allocated or the disk group has enough available free space:

- Two copies of the failed disk are reconstructed into the free space of the disk group.
  - The EVA uses the spare space only if there is insufficient free space.
- If the unpaired disk were to fail during the reconstruction, the VRAID1 virtual disk would be lost.
- After the data has been reconstructed, the data is releveled across the remaining members of the disk group.
  - The releveling may take hours to complete.
  - For odd-numbered disk groups, one disk is not used.
- If another disk fails (not the unpaired disk) during the releveling:
  - With adequate disk group free space or allocated spare space, the reconstruction process would repeat itself.
  - Without adequate space or allocated spare space, the failed disk cannot be reconstructed; there is no place to put the reconstructed data.
  - Failure of another disk group member only destroys the VRAID1 if it happens to be the unpaired disk of the original lost disk.

When one or more disk group members are added back into the disk group:

- New disk pairings occur.
- RSSs are recalculated.
- Data is releveled across the disk group.

Disk failures with VRAID5

If a disk group member failure occurs with a VRAID5 virtual disk, and the disk group has allocated spare space or enough available free space:

- Data from the failed disk is reconstructed into the free space.
  - The EVA uses the spare space if there is insufficient free space.
  - The EVA chooses where to put the reconstructed data based on what will maintain the highest level of redundancy.
- If another member of the same RSS were to fail during the reconstruction, in most circumstances, the VRAID5 virtual disk would be lost.
- After the data has been reconstructed,
  - RSSs for the disk group are recalculated as necessary.
  - The data is releveled across all remaining members of the disk group.
  - The leveling may take hours to complete.
- If during the releveling, another member of the same RSS were to fail:
  - With adequate disk group free space or allocated spare space, the reconstruction process would repeat itself.
  - Without adequate disk group free space or allocated spare space, a failed disk group member cannot be reconstructed; there is no place to put the reconstructed data.
    - RSSs are not recalculated.
- Failure of another disk group member in another RSS does not disable the VRAID5.

When one or more disk group members are added back into the disk group:

- RSSs are recalculated as necessary.
- Data is releveled across the disk group.