HP StorageWorks 4000/6000/8000 Enterprise Virtual Array configuration best practices white paper

Abstract

A business value of the HP StorageWorks Enterprise Virtual Array (EVA) is simplified management. A storage system that is simple to administer saves management time and money, and reduces configuration errors. You can further reduce errors and unnecessary expense by implementing a few best practices and optimizing your EVAs for their intended applications. This paper highlights common configuration rules and tradeoffs for optimizing HP StorageWorks EVAs for cost, availability, and performance. Getting the most from your enterprise class storage has never been easier.

Background

Two design objectives for the HP StorageWorks EVA were to provide maximum real-world performance and to reduce storage management costs. These objectives resulted in the design of an intelligent controller that minimized the number of user controllable tuning parameters. In contrast, traditional disk arrays typically have many tunable settings for both individual logical unit numbers (LUNs) and the controller. Although tunable settings might appear to be desirable, they pose potential issues:

- It is difficult and time consuming for administrators to set the parameters appropriately. Many settings require in-depth knowledge of controller internal algorithms and specific knowledge of the workload presented to the array.
- Storage administrators often do not have the time and resources to attain the expertise necessary to maintain a traditional array in optimum configuration.
- As the I/O workload changes, many parameters that were previously set might no longer be appropriate. Overcoming the effects of change requires continual monitoring, which is impractical and costly in most situations.

Because of such concerns, HP StorageWorks EVA algorithms minimize the parameters that users can set, opting instead to embed intelligence within the controller. The controller has a better view of the workload than most administrators and can be far more dynamic in responding to workload changes. The result is an array that is both easy to configure and high performing.

Overview

Although the HP StorageWorks EVA is designed to work in a wide range of configurations, configuration options can influence performance, usable capacity, or availability. With the necessary information about the configuration options for the HP StorageWorks EVA, a storage administrator can optimize its configuration for a specific application.

These configuration choices include:

- Number of disks
- Number of disk groups
- Type and number of disks in a disk group
- Vraid levels (0, 1, and 5)
- Disk failure protection level (none, single, and double)
- DR groups, snapshot, and clones
- Application configuration and third-party disaster recovery solutions
- Cache settings
- Capacity management

Each of these topics is detailed in the following sections. Note that it may **not** be possible to simultaneously optimize a configuration for cost, performance, and availability. Conflicting recommendations require one objective to yield to the demands made by another. For example, Vraid0 is clearly the best solution from a strict cost standpoint because nearly all storage is available for user data. However, Vraid0 offers no protection from a disk failure; thus, either Vraid1 or Vraid5 is a better choice for availability. Other tradeoffs can be more complex by comparison but are worth understanding because there is no best choice in all situations. "Best" depends on the priorities of a particular environment.

The best practices in this paper are based on controller software version 5.020.

Best practices summary

The following table summarizes typical best practices for optimum availability, performance, and cost. As with most generalizations, they are not the best choice for all applications. For detailed information, see the associated sections of this paper. As best practices, they are recommendations, not requirements. The HP StorageWorks EVA and HP support a wide variety of EVA configurations, and all supported configurations provide availability and performance features. These best practices have been developed to help you make good configuration choices when alternatives exist. For configuration requirements, see the EVA user manual or installation guide.

Table1.

Best practice	Source	Discussion
Disks in disk groups in multiples of eight	Availability	Multiples of eight disks per disk group allow the EVA to optimize the distribution of disks in the Redundancy Storage Set (RSS).
Use disks of same size and speed in a disk group	Performance	Ease of management and cost utilization. This configuration avoids any issues with access-density of the disk group.
As few disk groups as possible	Performance, Cost	Performance is optimized when the array is allowed to stripe data to as many disks as possible.
Protection level of one	Cost	In most installations, a protection level of one provides adequate availability. See detailed discussion of mitigating conditions.
Separate disk group for database logs	Availability	Provides consistent and current database restore from external media if data/table space disk group is corrupted.
Use 15K-rpm disks	Performance	15K-rpm disks have equal or higher performance than 10K-rpm disks. However, they are the most expensive. See details for discussion on price-performance optimization.
Load balance demand to controllers	Performance	Balancing the workload as evenly as possible to both controllers provides the highest performance utilization.
Vraid1	Availability, Performance	Vraid1 provides the highest level of data protection and availability. For most workloads, Vraid1 provides the best performance.
Vraid5	Cost	Vraid5 provides the lowest cost of protected storage.
Capacity management	Availability	Proper settings for the protection level, occupancy alarm, and available free space will provide the resources for the array to respond to capacity-related faults.
HP StorageWorks Continuous Access EVA or host-based mirroring	Availability	Real-time mirroring to an independent array provides the highest levels of data protection and availability. Geographically disperse arrays provide disaster protection.
External media backup	Availability	All data center best practices include processes to regularly copy the current data set to external media or near-line devices.
ISEE	Availability	Provides automated messaging of abnormal EVA status to HP Support or your internal IT personnel.

First best practices

The first best practices fall into the common sense category, but are often overlooked.

- Read the EVA user manual. Always operate the array in accordance with the user manual. In particular, never exceed the environmental operation requirements.
- HP continually improves the performance, reliability, and functionality of the EVA. Your EVA can benefit from these investments only if it is using the latest controller software and disk firmware.
- Deploy the array only in supported configurations. In many cases, HP does not support a configuration because it failed our testing. Do not risk the availability of your critical applications to unsupported configurations.

✓ **First best practice: Read and adhere to the user manual.**

✓ **First best practice: Stay as current as possible with the latest XCS controller software and disk firmware.**

✓ **First best practice: Deploy the array in supported configurations only.**

Best practices to optimize cost

The term "cost" in this paper refers to the cost per unit of storage. The cost is obtained by dividing the total cost of the storage by the usable data space as seen by the operating system. Cost is typically expressed in $/MB or $/GB. This section discusses options for improving the total usable capacity for a given configuration, thus lowering the cost per unit of storage.

Mixed disk capacities influence the cost of storage

The HP StorageWorks EVA can simultaneously support disks of several different capacities. Larger disks are usually more expensive but offer a lower price per unit of storage. In addition, disk technology typically doubles the capacity points every 18 months. The result of these market and technical factors is that storage arrays tend to be configured, or at least there is interest to configure arrays with disks of different capacities. The EVA greatly simplifies the management of these configurations. However, understanding EVA configuration rules can help you optimize a solution using disks with different capacities.

The disk failure protection level is a selectable parameter that defines the number of disk failure and autoreconstruction cycles a disk group can tolerate before failed disks are replaced. The protection level can be dynamically assigned to each disk group as a value of none, single, or double. Conceptually, it reserves space to handle 0 (none), 1 (single), or 2 (double) disk failure-reconstruct cycles. The space reserved is specific to a particular disk group and cannot span disk group boundaries.

The software algorithm for reserving reconstruction space finds the largest disk in the disk group; doubles its capacity; multiplies the result by 0, 1, or 2 (the selected protection level); and then removes that capacity from free space. Unlike traditional arrays, the EVA does not reserve physical disks. The reconstruction space is distributed across all disks in the disk group so that all disks remain available for application use. This is called distributed sparing. The largest disk is used even though there might only be a few large disks in a disk group. This enables the protection of the largest disks and the smaller disks. The reason the algorithm doubles the disk count is that Vraid1 recovery algorithm requires Vraid1 spare space to be in predetermined disk pairs. When a member of a pair fails, the remaining contents are moved to a new pair; thus, twice the capacity is reserved. The advantages of distributed sparing are two-fold. First, the performance value of all disks is used, and second, unlike a traditional spare, there is no risk that a reserved disk is unusable (failed) when needed.

Because reconstruction space is not shared across disk groups, it is more efficient to have the least number of disk groups possible, thus minimizing the reconstruction overhead. When the first few larger disks are introduced into a disk group, the resulting usable capacity is the similar as if they were smaller disks until the protection level requirements are met. Then the additional usable capacity is in line with the physical capacity of the new disks. Even though the resulting capacity for the first few disks is not very efficient, the alternative of creating two disk groups provides less usable capacity.

✓ **Best practice to minimize the cost of storage: Mix disk sizes within a single disk group.**

Number of disk groups influences the cost of storage

Disk groups are independent protection domains, As such, all data redundancy information and reconstruction space must be contained within the disk group. Unlike traditional RAID 5 arrays, where the stripe depth and redundancy set are the same, the EVA supports many disks in the stripe set (a disk group), but the redundancy set (RSS) is always 6 to 11 disks. (Redundancy sets are discussed in the section on optimizing availability.) As a result, multiple disk groups are not needed for Vraid5 availability and would provide excessive reconstruction space for the small number of disks that would be in each disk group.

As with traditional arrays, multiple disk groups can result in stranded capacity. Stranded capacity occurs when the capacity that could be used to create a virtual disk is distributed to many disk groups and, therefore, cannot be used to create a single LUN. (A virtual disk must be created within a single disk group.) Unlike traditional disk arrays, the EVA can easily support very large disk groups, thus eliminating stranded capacity issues.

✓ **Best practice to minimize the cost of storage: Create as few disks groups as possible.**

Note
To understand the tradeoffs associated with the number of disk groups, see the discussion of disk groups in the sections on optimizing performance and availability. Before acting on any best practice, you should understand the effect on price, performance, and availability.

Number of disks influences the cost of storage

The lowest cost per storage unit is achieved by amortizing the cost of the controllers over as many disks as possible. The lowest cost per storage unit for disks is typically on the largest disks. Thus, the lowest cost solution is to fill the array with as many of the largest disks as are supported. The HP StorageWorks 8000 Enterprise Virtual Array (EVA8000) can support up to 240 disks. This configuration results in the lowest cost of storage per MB.

✓ **Best practice to minimize the cost of storage: Fill the EVA with as many disks as possible, using the largest-capacity disks.**

Note
Before increasing the disk count, see the sections within this document on optimizing performance and optimizing availability.

Disk performance influences the cost of storage

Larger disks usually offer better price per capacity than smaller disks. Although prices continuously change, more capacity can be purchased for the same price by purchasing larger drives. Conversely, higher performance drives, such as 15K-rpm drives, are generally more expensive than their lower performance 10K-rpm counterparts.

✓ **Best practice to minimize the cost of storage: Use lower performance, larger capacity disks.**

Cross-RAID snapshot influence the cost of storage (and the perceived availability of storage)

The EVA allows the target LUN of a space-saving or fully allocated snapshot to be a different Vraid type than the source. For example, a Vraid1 LUN can have an associated Vraid0, Vraid1, or Vraid5 snapshot LUN. With space-saving snapshots, only the data that has changed is copied. This means that the unchanged data remains in the source LUN (and RAID level), and the changed data resides in the target LUN (and RAID level).

The availability characteristics of a cross-RAID snapshot LUN are those of the lower (in an availability sense) RAID level of either the source or the target LUN. Therefore, it does not make economic sense to assign the target LUN a RAID level with greater availability than the source. The snapshot LUN would consume more capacity without providing greater availability.

The hierarchy of RAID levels for the EVA is:

1. Vraid1—highest availability, uses the most raw capacity
2. Vraid5
3. Vraid0—lowest availability, uses the least raw capacity

✓ **Best practice to minimize the cost of storage: Use equal or lower RAID level for the target LUN of a Vraid snapshot.**

FATA disks influence the cost of storage

Fibre Attached Technology Adapted (FATA) disks are low-cost, low-performance disks for use in the EVA. Because of the design of these disks, HP recommends a reduced duty cycle to meet business-critical availability requirements. Use FATA disks where random access performance and continuous operation are not required.

The EVA requires that FATA disks be organized in separate disk groups.

The best application for FATA disks is for the online part of your backup and recovery solution. Snapclones assigned to FATA disk groups provide the lowest cost solution for zero-downtime backup and fast recovery storage. HP software solutions, such as HP StorageWorks Fast Recovery Solutions (FRS) for Exchange and HP OpenView Storage Data Protector, and other industry backup software, include the management of FATA-based snapclones for backup and recovery.

✓ **Best practice for FATA disk and the cost of storage: Use FATA disks to augment near-line storage usage.**

Note
FATA disks and snap copies are not a replacement for offline backup. Best practice is to retain data on an external device or media for disaster recovery.

Best practices to optimize availability

The HP StorageWorks EVA is designed for business-critical applications. Redundancy features enable the array to continue operation after a wide variety of failures. The EVA is also designed to be flexible. Some configurations allow higher availability by limiting exposure to failures that exceed the fault-tolerant design of the array. The goal is to create configurations that have the most independent protection domains so that a failure in one domain does not reduce the resiliency of another domain.

All supported configurations have some tolerance to failures but not necessarily all failures. As an example, Vraid0 offers no protection to disk failures, but it is still resilient (as an example) to most back-end Fibre Channel loop failures.

The following guidelines (1) address configurations to improve availability in sequential and multiple simultaneous failure scenarios and (2) discuss system and application configurations to improve availability.

RAID level influences availability

While Vraid5 provides availability and data protection features sufficient for most high-availability applications, some applications may require the additional availability and data-protection features of Vraid1. Vraid1 configurations can continue operation in failure scenarios where Vraid5 cannot. A statistical model of the EVA shows that, for an equivalent usable capacity, Vraid1 provides over four times the data protection of Vraid5[1].

This additional redundancy comes with additional cost caused by additional storage overhead. Nevertheless, some applications or file sets within an application warrant this additional protection.

If cost constraints do not allow a total Vraid1 configuration, consider using Vraid1 for critical files or data sets. For example:

- In database applications, select Vraid1 for log files.
- In snapclone/snapshot applications, select Vraid1 for active data sets and Vraid5 for snapclones and snapshots.

✓ **Best practice for highest availability: Vraid1 provides the highest levels of availability and data protection.**

Note
The higher availability and data protection capabilities of Vraid1 or Vraid5 should not be considered a replacement for good backup and disaster recovery processes. The best practices for business-critical applications always include frequent data backup to other near-line or offline media or devices.

Shelf and disk organization influences availability

Within a disk group, the EVA creates multiple subgroups of disks called the RSS. Each RSS contains sufficient redundancy information to continue operation in the event of a disk failure within that RSS. The EVA can thus sustain multiple simultaneous disk failures while not losing user data, as long as no more than one disk per RSS fails. RSSs are created when a disk group is created, and additional sets are created as necessary when disks are added to the disk group. RSSs are created and managed by the EVA controllers, with no user intervention required.

[1] A statistical analysis is a prediction of behavior for a large population; it is not a guarantee of operation for a single unit. Any single unit may experience significantly different results.

The target size of each redundancy set is eight disks, with a minimum of six and a maximum of 11. As disks are added to a disk group, the RSS automatically expands until it reaches 12 disks. At that point, it splits into two sets of six disks each. As more disks are added, one set increases from six to eight (the target size); then the remaining set increases. After all disks have been added to a disk group, each RSS contains eight disks, with the possible exception of the last set, which contains between six and 11 disks. This is why it is a best practice to add disks in groups of eight.

To maximize array availability, disks should be arranged vertically across shelves and distributed as evenly as possible within shelves. The number of disks in the shelves should not differ by more than one. The specific best practices vary depending on whether virtual disks are configured as Vraid1 or Vraid5. If you mix Vraid1 and Vraid5 in a single disk group, follow the best practices for Vraid5. If you use only Vraid1, follow the best practices for Vraid1.

✓ **Vraid5 best practices to optimize availability:**

- **Arrange all disks in a vertical fashion and distribute them among the shelves as evenly as possible.**
- **Keep the total number of disks in the array and the total number of disks in a disk group to multiples of eight.**
- **When creating a disk group, let the EVA choose which disks to place in the group.**

With Vraid1, the EVA controller attempts to place the members of a mirror pair on separate shelves.

✓ **Vraid1-only best practices to optimize availability:**

- **Arrange all disks in a vertical fashion and distribute them among the shelves as evenly as possible.**
- **Keep the total number of disks in the disk group to a multiple of two.**
- **When creating a disk group, let the EVA choose which disks to place in the group.**

Vraid0 influences availability

Unlike Vraid1 or Vraid5, Vraid0 has no data redundancy. Vraid0 is optimized for applications where data protection is not a requirement. Because Vraid0 has no redundancy, data in Vraid0 requires less physical capacity, and performance is not affected by additional operations required to write redundancy information. Thus, Vraid0 provides the best performance for write-intensive workloads and the lowest cost of storage but the least availability.

✓ **Vraid0 best practice to optimize availability: Vraid0 is not advised for availability. Vraid0 provides no disk failure protection.**

Note
For Vraid0, increasing the protection level does not increase the availability of the virtual disk. A single disk failure renders it inoperable even when single or double protection is enabled.

Replacing a failed disk influences availability

Following the rules for shelf and disk organization is the best protection against potential data loss and loss of availability due to disk failure. However, when a disk fails, additional steps should be followed to minimize the risk of data loss or unavailability.

HP service engineers are trained on the proper EVA repair procedures and are alert to abnormal conditions that warrant additional steps to ensure continued operation. The best practice to maximize availability is to call for HP service. If HP service is not an option or is unavailable, use the following rules.

When a disk fails, the EVA rebuilds the failed disk data through a process known as reconstruction. Reconstruction restores the disk group resiliency to another disk failure. After reconstruction or after a new disk is added to a disk group, the EVA redistributes the data proportionately and reorganizes redundancy sets to the active disks.

✓ **Best practice to optimize availability: Use the following procedure for disk replacement:**

✓ **Wait for the reconstruction to complete before removing the failed disks. This is signaled by an entry in the event log.**

1. **Use HP StorageWorks Command View EVA to ungroup the disk. This assures the disk is not a member of a disk group.**
2. **Replace the failed disk. The new disk must be inserted into the same slot as the failed disk.**
3. **Manually add the new disk into the original disk group. Ensure the disk addition policy is set to manual mode.**

Protection level influences availability

The protection level defines the number of disk failure–autoreconstruction cycles that the array can accomplish without replacement of a failed disk. Following a disk failure, the controller re-creates the missing data from the parity information. The data is still available after the disk failure, but it is not protected from another disk failure until the reconstruction operation completes.

For example, “single” protection level provides continued operation in the event of two disk failures, assuming the reconstruction of the first failed disk completes before the second disk fails.

For Vraid1 and Vraid5, protection level “none” provides resilience to a single disk failure; however this is not a best practice configuration. Vraid0 provides no protection to any disk failure.

Conversely, the statistical availability of disks and the typical service time to replace a failed disk (MTTR[2]) indicate that “double” protection level is unnecessary in disk groups of fewer than 168 disks in all but the most conservative installations. A mitigating condition would be a service time (MTTR) that exceeds seven days. Then a protection level of ”double” might be considered for groups of less than 168 disks.

✓ **Best practice to optimize availability: Use ”single” protection level.**

Note
Protection level reserved capacity is not associated with the occupancy alarm setting. These are independent controls.

Number of disk groups influences availability

Although the EVA offers numerous levels of data protection and redundancy, a catastrophic failure[3] can result in loss of a disk group. An example would be the failure of a second disk in an RSS before the reconstruction operation is complete. The probability of these events is low; however, installations requiring the highest levels of data availability may require creating multiple disk groups for independent failure domains[4]. Multiple groups result in a slightly higher cost of ownership and potentially lower performance, but may be justified by the increased availability.

[2] MTTR—Mean Time to Repair

[3] Defined as multiple, simultaneous failures that exceed the architectural redundancy

[4] Independent failure domain. A failure in one domain does not affect the availability characteristics of the other domain.

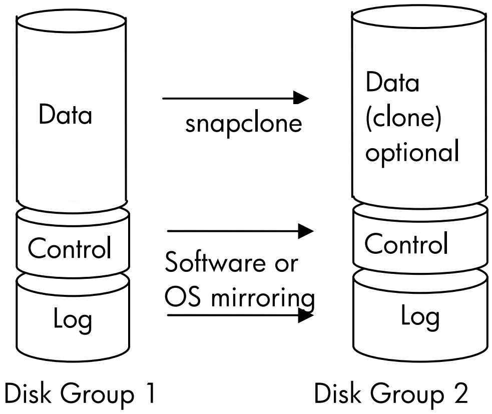
The strategy for multiple disk groups is to keep recovery data in a separate disk group from the source data. The typical use is to keep either a snapclone of the database or the database log files in a separate disk group from the application. If the primary disk group fails, the recovery disk group may remain available. Additional work is required to restore the application, but the recovery data is online, thus reducing the recovery time.

For two disk groups to prevent data loss, each disk group must contain sufficient independent information to reconstruct the entire application data set from the last backup. A practical example of this is a database that contains data files, configuration files, and log files. In this instance, placing the data files in one group and duplexing the log files and control files (duplexing or mirroring is a feature of some databases) to both the data file disk group and another group ensure that loss of a single disk group does not prevent recovering the data.

For example:

- Disk Group 1 contains data files, a copy of online redo logs, a copy of the control file, and an optional copy of archived logs (if supported by either the database or OS).
- Disk Group 2 contains a copy of online redo logs, a copy of the control file, the primary archive log directory, and an optional snapclone of the data files for Disk Group 1.

Figure 1.

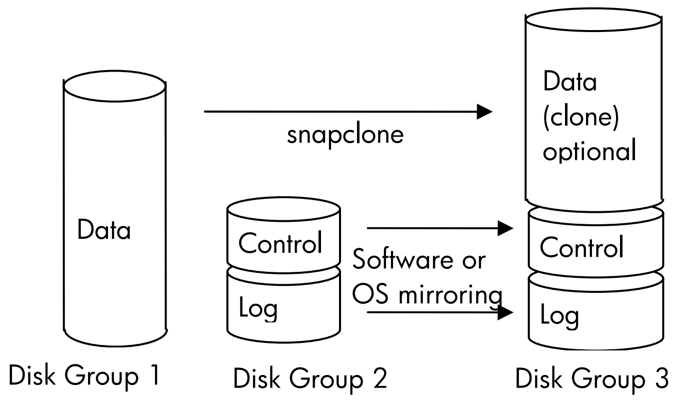


If Disk Group 2 does not contain the snapclone of the data files, the number of disks in Disk Group 2 should be determined by the sequential performance demand of the log workload. Typically, this results in more usable capacity than is required for the logs. In this case, choose Vraid1 for the log disks. Vraid1 offers the highest availability and, in this case, does not affect the cost of the solution.

A variation on this configuration is two separate disk groups for the log and control files, and a third for the data files. This configuration has a slightly higher cost but appeals to those looking for symmetry in the configuration. In this configuration:

- Disk Group 1 contains database data files.
- Disk Group 2 contains the database log, control file, and archived log files.
- Disk Group 3 contains a database log copy, control file copy, and archived log files copy (if supported).

Figure 2.



Disk groups can be shared with multiple databases. It is not a best practice to create a separate disk group for each database.

- ✓ **Best practice to optimize availability: For critical database applications, consider placing data files and recovery files in separate disk groups.**
- ✓ **Best practice to optimize availability: Assign snapclones to a separate disk group.**

Note
Creating multiple disk groups for redundancy and then using a volume manager to stripe data from a single application across both disk groups defeats the availability value of multiple disk groups.

If, for cost or capacity reasons, multiple disk groups are not implemented, the next best practice is to store the database log, control file, and log archives in Vraid1 LUNs. Vraid1 provides greater protection to disk failures than Vraid5 or Vraid0.

Number of disks in disk groups influences availability

Disks assigned to disk groups in multiples of eight provide optimum placement of RSS disk sets. In this case, the controller can place the optimum number of disks in each RSS.

- ✓ **Best practice to optimize availability: Size disk groups in multiples of eight disks.**

Capacity management and optimized availability

Free space, the capacity that is not allocated to a virtual disk, is used by the EVA controller for multiple purposes. Although the array is designed to operate fully allocated, functions like snapshot, reconstruction, leveling, remote replication, and disk management either require or work more efficiently with additional free space.

Three controls manage free space in the EVA: the protection level, the capacity occupancy alarm, and the capacity reported as available for virtual disk creation. Successful capacity planning requires understanding specific requirements for availability and cost, and setting the appropriate protection level, capacity occupancy alarm, and total virtual disk capacity.

Set the protection level for the disk group. See the previous discussion of protection level and availability.

Additional reserved free space—as managed by the occupancy alarm and the total virtual disk capacity—affect leveling, remote replication, local replication, and proactive disk management. The following best practice addresses the occupancy alarm setting and the total virtual disk capacity.

The occupancy alarm is set for each disk group as a percentage of the raw capacity. Base the occupancy alarm setting on the unique installation requirements for proactive disk management, remote replication, and leveling.

Proactive disk management. Proactive disk management (PDM) is a request by a customer or HP Services to ungroup a disk, or it is a predictive disk failure request by the EVA to ungroup a disk. In either case, the array migrates the contents of the disk to free space before removing the disk from use. PDM can occur only if sufficient free space is available. PDM operation capacity is independent of protection level capacity. Customers who desire the highest levels of availability elect to reserve additional free space for PDM.

The capacity used for PDM is twice the largest disk in the disk group for each PDM event anticipated. Typical choices are none, one, or two events. The greater the disk count in a disk group, the greater the opportunity for a PDM event.

✓ **Best practice to optimize availability. Set the occupancy alarm to allow space for one or two PDM events per disk group.**

Remote replication. HP StorageWorks Continuous Access EVA uses free space for the DR group log (also known as the write history log). The DR group log is activated when the remote link fails or is suspended. Until the link is reestablished, the EVA controller records changes locally in the DR group log. For free space management, allow for the maximum size of the write history logs in each disk group. The size of the write history log is specified when the DR group is created. The default size is combined size of the DR group's virtual disks.

Leveling and reconstruction. Leveling and reconstruction performance can be optimized with a minimum of 5 GB of free space per disk group.

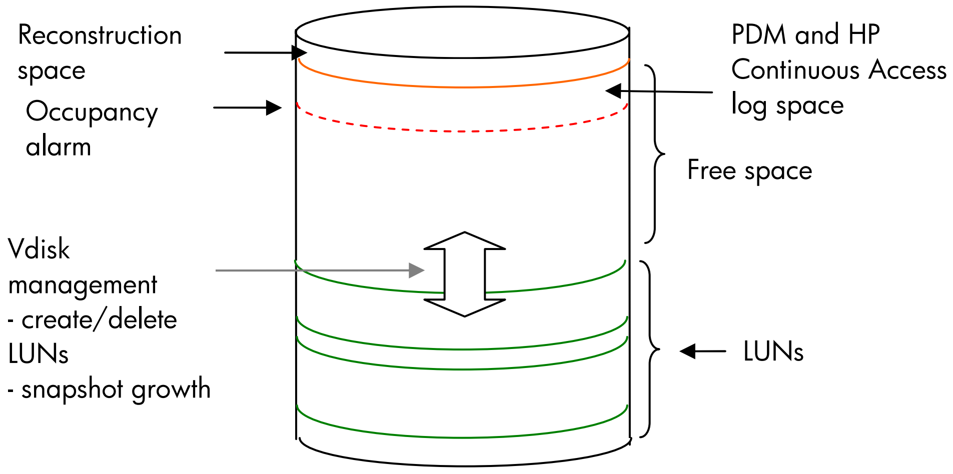
✓ **Best practice to optimize availability. Set the occupancy alarm to the larger of the capacity required for PDM or the total HP Continuous Access EVA write history log capacity, plus 5 GB. This capacity is converted into a percentage of the raw capacity and then rounded to the next largest whole number. The pseudo-Excel formula would be (see footnotes for description of functions):**

Occupancy_Alarm = 100 – ceiling[5] ((max[6] (PDM_capacity, HP Continuous Access_log_sum) + 5GB) / total_disk-group_raw_capacity))

[5] Ceiling function – compute the next largest whole number

[6] Choose the maximum of the following

Figure 3.



A logical view of the capacity of a disk group (not to scale)

Remaining free space is managed by the creation of virtual disks, and is measured by the capacity available to create additional virtual disks. This free-space capacity is used by space-efficient snapshots. It is critical that sufficient free-space is available for the space-efficient snapshot copies; else all snapshot copies in the disk group become inoperative[7]. (Fully allocated snapshot vdisks and snapclone vdisks will continue to be available.)

Snapshots use copy-on-write technology. Copy-on-write occurs only when either the original virtual disk or the snapshot virtual disk is modified (a write); then an image of the associated blocks is duplicated into free space. For any given block, the data is copied only once. As snapshot copies diverge, the capacity available to create virtual disks decreases.

The actual capacity required for a space-efficient snapshot depends on the divergence of the original virtual disk and the snapshot. This value is unique for each application, but can range from 0 percent to 100 percent. The suggestion for the initial usage of snapshot (that is, when you do not know the actual physical capacity required by the snapshot) is to reserve (do not allocate to virtual disks) 10 percent of the capacity of the parent virtual disks times the number of snapshots per parent virtual disk. For example, if you need to create two space-efficient snapshot vdisks of a 500-GB virtual disk, you need to ensure that 100 GB (500 GB*10 percent*2) of usable capacity is available. Compute the usable capacity using the RAID level selected for the snapshot vdisk.

[7] Space-efficient snapshot vdisks become inoperative individually as each attempts to allocate additional free-space. In practice the effect is that all become inoperative together.

- ✓ **Best practice to optimize availability: Leave unallocated virtual disk capacity, the capacity available to create virtual disks, equal to the sum of the capacities required for all space-efficient snapshot copies within a disk group.**

- ✓ **Best practice to optimize availability: Respond to an occupancy alarm; evaluate what changed, replace failed disks, add disks, or reevaluate space-efficient snapshot vdisk usage (delete snapshot vdisks). Extended operation of the array in an occupancy alarm condition is not a best practice.**

See examples for free space management in Appendix A.

Fault monitoring to increase availability

The best protection from downtime is to avoid the failure. Good planning and early warning of problems can minimize or eliminate many issues. For the EVA, Instant Support Enterprise Edition (ISEE) is a free service that forwards EVA faults and warnings directly to HP Services through a secure virtual private network. HP can evaluate and diagnose problems remotely, possibly avoiding more serious issues. If an array requires service, ISEE greatly increases the probability that the HP Service engineer arrives on-site with the correct replacement parts to minimize the time to repair.

If site security policies exclude direct communication to HP Services, ISEE can be set up to report warnings to a customer contact. If ISEE is not used, a custom automated alert process based on HP Web-based Enterprise Services (WEBES) or similar tool can be developed. It is critical that alarms from the EVA are not ignored.

- ✓ **Best practice to optimize availability: Install and use ISEE or equivalent tools to monitor and alert administrators to changes in EVA status.**

Integrity checking to increase availability

Exchange, Oracle®, and other databases include tools to verify the integrity of the database. These tools check the database for consistency between records and data structures. Use these tools as part of the ongoing data center processes, as you would data backup and recovery testing. Proper use of these tools can detect database corruption early, when recovery options are still available.

- ✓ **Best practice to optimize availability: Run database integrity checks as an ongoing date center process.**

Backup processes to optimize recovery

Independent of the storage device, include a proven backup and recovery process in array and data center management procedures. The EVA is supported by numerous backup applications. HP OpenView Storage Data Protector and similar third-party backup software are supported on a variety of popular operating systems. They support the EVA directly or are integrated through Oracle, Microsoft® SQL, or other databases to provide zero-downtime backup.

Along with the backup data, save a copy of the EVA configuration files. An exact description of the array configuration greatly reduces recovery time. These should be stored on media not associated with the array.

Do not consider array-based copies the only form of backup. Snapshot and snapclones copies complement a backup strategy that includes full copies to offline or near-line storage. In this application, snapclones can provide alternatives for reducing recovery time by providing the first option for recovery.

Perform regular backups and be sure to test the restore process twice a year. The EVA greatly simplifies the testing process by providing a simple process to create and delete disk groups or virtual disks. Capacity used for testing can be easily reused for other purposes.

✓ **Best practice to maintain data recovery: Perform regular data backup and biannual recovery tests.**

✓ **Best practice to maintain data recovery: Include a copy of the EVA configuration with the backup data. This can be accomplished with the SSSU Capture Configuration utility.**

Remote mirroring and availability

HP StorageWorks Continuous Access EVA is an optional feature of the array that enables real-time remote mirroring. HP Continuous Access EVA protects against catastrophic EVA or site failures by keeping simultaneous copies of selected LUNs at local and remote EVA sites. This feature can work standalone or in combination with system clustering software.

✓ **Best practice to optimize for the highest levels of data availability: Consider array-based HP Continuous Access EVA or operating system-based replication software to provide real-time mirroring to a second EVA.**

HP Continuous Access EVA and Vraid0 influence availability

When HP Continuous Access EVA is used, LUNs can be logically associated into DR groups. DR groups allow a database to recover a remote copy with transaction integrity.

Two characteristics of remote mirroring are consistency and currency. Currency refers to the time difference between the remote and local copies. Consistency refers to the content difference between the mirrored LUNs and the local LUNs. A consistent remote LUN is either equal to the content of the local LUN or equal to the content of the local LUNs at a past point in time. Synchronous HP Continuous Access EVA provides mirror copies that are both current and consistent. Asynchronous operation provides mirror copies that are consistent, but may not be current (they may be slightly delayed). To accomplish asynchronous consistency, HP Continuous Access EVA maintains the remote write ordering within each DR group. The order of the remote writes is identical to the order of the local writes. The writes may be delayed by the transmission distance, but the content of the LUNs in a DR group matches the current or a previous state of the associated local LUNs.

If a remote LUN becomes unavailable, HP Continuous Access EVA cannot continue to write to the remaining LUNs in the DR group without losing DR group consistency. If the EVA continued to write to the remaining LUNs and the unavailable LUN became available, its contents would not be consistent with the other LUNs of the DR group and a database recovery at this point would lose transaction integrity. Thus, the availability of a remote DR group is tied to the availability of the individual LUNs. If the remote DR group contains a Vraid0 LUN, mirroring to the entire DR group is tied to the availability of that Vraid0 LUN. While it is desirable to use Vraid0 as the target of a remote mirror to save costs, the solution must tolerate the resulting loss of availability.

✓ **Best practice to optimize availability: Do not use Vraid0 as a target for HP Continuous Access EVA mirror.**

System fault-tolerant clusters and availability

Many operating systems support optional fault-tolerant configurations. In these configurations, multiple servers and arrays are grouped together with appropriate software to enable continued application operation in the event of system component failure. These solutions span the replication continuum from simple local mirroring to complex disaster recovery solutions.

HP Continuous Access EVA can be used for simple mirroring to increase data availability or as a component of a complete disaster recovery solution. A complete disaster recovery solution automatically (or manually) transfers the applications to the remaining functional systems in the DR group in the event of a site or component failure.

The HP StorageWorks EVA is supported in many disaster recovery solutions. HP provides two disaster recovery products: Cluster Extensions for Windows and Linux, and Metrocluster for HP-UX. For more details, contact HP or your operating system vendor. For applications requiring the highest availability, these system-level solutions should be considered.

- ✓ **Best practice to optimize application availability: Consider disaster recovery solutions to provide continuous application operation.**

Best practices to optimize performance

Unless otherwise noted, the performance best practices in this document refer to configurations without HP Continuous Access EVA. Additional HP Continuous Access EVA performance and configuration information can be found in the HP StorageWorks Continuous Access EVA administrator guide, and the HP StorageWorks Continuous Access EVA planning guide.

Experience shows that high performance and low cost generally have an inverse relationship. While this continues to be true, the HP StorageWorks EVA virtualization technology can significantly reduce the cost of high performance. This section outlines configuration options for optimizing performance and price-for-performance, although sometimes at the expense of cost and availability objectives.

Array performance management typically follows one of two strategies: contention management or workload distribution. Contention management is the act (or art) of isolating different performance demands to independent array resources (that is, disks and controllers). The classic example of this is assigning database table space and log files to separate disks. The logic is that removing the contention between the two workloads improves the overall system performance.

The other strategy is workload distribution. In this strategy, the performance demand is evenly distributed across the widest set of array resources. The logic for this strategy is to reduce possible queuing delays by using the most resources within the array.

Before storage devices with write caches and large stripe sets, like the EVA, contention management was a good strategy. However, with the EVA and its virtualization technology, workload distribution is the simplest technique to maximizing real-world performance. While you can still manage EVA performance using contention management, the potential for error in matching the demand to the resources decreases the likelihood of achieving the best performance.

This is a key concept and the basis for many of the best practices for EVA performance optimization. This section explores ways to obtain the best performance through workload distribution.

Optimizing performance raises the issue of demand versus capability. Additional array performance improvements have very little effect on application performance when the performance capabilities of the array exceed the demand from applications. An analogy would be tuning a car engine. There are numerous techniques to increase the horsepower output of a car engine. However, the increased power has little effect if the driver continues to request that the car travel at only 30 MPH. If the driver does not demand the additional power, the capabilities go unused.

The best results from performance tuning are achieved when the analysis considers the whole system, not just the array. However, short of a complete analysis, the easiest way to determine if an application could take advantage of array performance tuning is to study the queues in the I/O subsystem on the server and in the array. If there is little or no I/O queuing, additional array performance tuning is unlikely to improve application performance. The suggestions are still valuable, but if the array performance capabilities exceed the demand (as indicated by little or no queuing), the suggestions may yield only a modest gain.

Number of disks influences performance

In general, adding more disks provides better performance under typical workloads. However, there is diminishing return from additional disks as the combined performance capability of the disks approaches one or more of the performance limits of the controller. The performance limits of the controller depend on the characteristics of the application workload. For example, sequential workloads require fewer disks than random workloads to reach the point of diminishing return.

Where additional disks are needed only for capacity and the application does not require additional performance, increasing the number of disks in the array to the maximum supported is a sensible solution. An example of this is when the demand is below the capability of the current EVA configuration. This can be determined by examining the disk queue depths. If there is little or no disk queuing delay, additional disks will not improve the performance of the array or application. For these situations, adding disks to the existing EVA provides the lowest cost solutions.

However, when the performance demand of an application exceeds the current performance capability of the array and the goal for the additional disks is to increase the application performance, consider a performance evaluation before changing the configuration. As a general rule of thumb, when the array contains 168 or more disks, performance should be reviewed to determine the preferred strategy for increasing array capacity. The performance review studies the characteristics of the workload and the array configuration and then predicts expected array performance with the additional disks. This prediction can then be compared with application requirements and alternative solutions for performance improvements.

✓ **Best practice to maximize single array performance: Fill the EVA with as many disk drives as possible.**

✓ **Best practice to maximize application performance: Consider a performance analysis before increasing the total disk count above 168.**

Number of disk groups influences performance

The number of disk groups has no effect on the performance capabilities of the EVA. An EVA can achieve full performance with a single disk group.

For typical workloads, an increased number of disk drives in a virtual disk improves the performance potential of the virtual disk. Because a virtual disk can exist only within a single disk group, it follows that having a single disk group maximizes the performance capability.

Large disk groups allow the workload to be distributed across many disks. This distribution improves total disk utilization and results in the most work being completed by the array. However, some application performance is measured by low response times for small, random I/Os, not the total work completed. In this case, sequential workloads can interfere with the response time of the random component. For these environments, a separate disk group for the sequential workload can reduce the impact on the random I/O response time.

An alternative approach to maximize performance with multiple disk groups is operating system software (that is, a volume manager) that can stripe data across multiple LUNs on multiple disk groups. This provides a similar distribution of the workload to the disks as multiple LUNs on a single disk group. However, this solution provides lower capacity utilization than a single disk group.

✓ **Best practice to optimize performance: Configure as few disk groups as possible.**

Note
Before considering a single disk group, see the discussion in this paper on the number of disk groups and availability.

Disk performance influences array performance

For applications that perform large-block sequential I/O, such as data warehousing and decision support, disk speed has little or no effect on the net performance of the EVA. Disk data sheets confirm that the average sustained large-block transfer rates are similar for both 10K- and 15K-rpm disks of the same generation. Accordingly, large capacity 10K-rpm disks make the most sense for large-block sequential workloads.

For applications that issue small-block random I/O, such as interactive databases, file and print servers, and mail servers, higher speed disk drives offer a substantial performance advantage. Workloads such as these can see gains of 30 percent to 40 percent in the request rate when changing from 10K-rpm to the equivalent number of15K-rpm disks.

Although it seems contradictory to use 10K-rpm disks for better performance for small-block random I/O workloads, there are instances in which 10K-rpm disks provide either better performance or better price-for-performance. The key is the relationship of the cost of the disk, the performance of the disk, and the quantity of disks. Because the performance gain from a 15K-rpm disk ranges from 30 percent to 40 percent, if the 15K-rpm disks are more than 30 to 40 percent more expensive than the 10K-rpm disk, then it makes sense to purchase a larger number of 10K-rpm disks.

The performance optimization with 10K-rpm disks can be achieved only when the workload is striped across the disks. Unlike traditional arrays, the EVA automatically stripes the data across all the disks in the disk group, making this optimization easy to achieve.

- ✓ **Best practice to optimize performance: 15K-rpm disks provide the highest performance.**

- ✓ **Best practice to optimize price-performance: For the equivalent cost of using 15K-rpm disks, consider using more 10K-rpm disks.**

Vraid level influences performance

Performance optimization is a complex activity; many workload characteristics and array configuration options can influence array performance. Without a complete workload analysis, array performance is difficult to predict. However, given this caveat, in general, Vraid1 provides better performance characteristics over a wider range of workloads then Vraid5. However, Vraid5 can provide superior performance for some sequential write workloads. The workloads that are candidates for Vraid5 contain a high percentage of sequential write operations, and the write record size must be in multiples of 8K bytes. The larger the record size, the greater the Vraid5 advantage.

- ✓ **Best practice for Vraid level and performance: Vraid1 provides the best performance over the widest range of workloads; however, Vraid5 is better for some sequential write workloads.**

Vraid0 influences performance

Vraid0 provides the best random write workload performance; however, Vraid0 provides no protection from disk failure. When the risk is well understood, Vraid0 can improve overall application performance. As an example of the risk associated with Vraid0, a Vraid0 virtual disk using 16 physical disks can expect two or three data loss events during the 5-year life of the array

Some applications can rely on other means for data protection. Replicated or test databases or temporary storage used by a database are examples. However, avoiding data loss by external replication only provides protection for the data. Loss of a virtual disk can interrupt service and require manual steps to recover the virtual disk and resume application operation.

✓ Best practice for Vraid0 and performance: Accepting the possible data and availability loss (Vraid0 provides no protection from disk failure), consider Vraid0 for noncritical storage needs only.

Mixing disk performance influences array performance

The EVA is designed to support heterogeneous disk types in capacity or performance. Mixing drives of different speeds in the same disk group does not slow down access for all the disks to the speed of the slowest drive in the group. Each disk is independent.

Traditional array management suggests or requires the separation of heterogeneous disks into separate groups. Following this suggestion on the EVA would negate one of the most powerful performance features of the EVA: the ability to easily stripe across many disks to improve the performance of the array.

Although grouping drives by type and speed may seem easier to manage, it is actually difficult to balance the demand to individual disk groups. Errors in this balance can result in a disk group being under- or over-utilized. Although the individual disks in a disk group can be slower, the ability to realize the aggregate performance of a single disk group is easier than when using two disk groups.

✓ Best practice to optimize the performance of an array containing disks of different performance characteristics: Combine disks with different performance characteristics in the same disk group. Do not create separate disk groups to optimize performance.

Mixing disk capacities influences performance

The EVA supports disk groups consisting of disks with different capacities, with the exception that FATA disks must be in their own disk group. Mixed-capacity disk groups are the recommended best practice for the optimization of the cost of storage. However, for performance optimization, it is recommended that disk groups contain only disks of the same capacity.

The EVA stripes LUNs across all the disks in a disk group. The amount of LUN capacity allocated to each disk is a function of the disk capacity. When disks of different capacities exist in the same disk group, the larger disks have more LUN data. Since more data is allocated to the larger disks, they are more likely to be accessed. In a disk group with a few, larger-capacity disks, the larger disks become fully utilized (in a performance sense) before the smaller disks. Although the EVA does not limit the concurrency of any disk in the disk group, the workload itself may require an I/O to a larger disk to complete before issuing another command. Thus, the perceived performance of the whole disk group can be limited by the larger disks.

When the larger disks are included in a disk group with smaller disks, there is no control over the demand to the larger disks. However, if the disks are placed in a separate disk group, the storage/system administrator can **attempt** to control the demand to the disk groups to match the available performance of the disk group.

Separate disk groups require additional management to balance the workload demand so that it scales with the performance capability of each disk group. For a performance optimized environment, however, this is the best alternative.

✓ Best practice for performance optimization: Use disks with equal capacity in a disk group.

✓ Best practice to optimize performance with disks of unequal capacity: Create separate disk groups for each disk capacity and manage the demand to each disk group.

Read cache management influences performance

Read caching can be set either at virtual disk creation or dynamically. This parameter affects both random-access read caching and sequential (prefetch) caching, although the algorithms and cache usage are different for these workloads.

Both algorithms come into play only when they will have a positive effect on performance. Random-access caching is enabled and disabled automatically as the I/O workload changes, while prefetch caching comes into play only when a sequential read stream is detected. Because of this dynamic response to changing I/O, cache efficiency is maintained at a high level, and there is no negative impact on either cache usage or performance when an I/O workload is "cache unfriendly."

Because there is no negative impact of leaving cache enabled and there is always a chance of a performance gain through caching, read cache should always be left enabled.

✓ **Best practice to optimize performance: Always enable read cache.**

Controller balancing influences array performance

With XCS versions 5.020 and greater, the EVA can present LUNs simultaneously through both controllers. This allows the EVA to be compatible with popular dynamic load balancing and path management software, such as Windows MPIO, HP-UX pvlinks, and VERITAS DMP, as well as HP StorageWorks Secure Path (contact HP for the latest compatibility matrix). Like previous versions of the EVA, either controller can own a LUN, but unlike previous versions, a LUN can now be simultaneously accessed through either controller.

Even though LUNs are presented through both controllers, it is still important to balance the workload between controllers. This is accomplished by balancing the LUN ownership between the controllers. Although ownership should be distributed by workload demand, you can initially assign LUNs to controllers by capacity. Performance data from EVAPerf can help you optimize the controller load balance.

The path through the controller that owns a LUN typically provides the best performance. If dynamic path management software is used; select the shortest-service-time option. If static path management is the only option, configure the path to use a port on the owning controller for each LUN.

When a controller fails, the LUNs owned by the failing controller are automatically switched to the remaining controller. The failover is automatic; however, the failback is not. After a failed controller is replaced, LUN ownership should be manually reassigned.

✓ **Best practice to optimize performance: Manually load balance LUN ownership to both controllers. If dynamic path management software is used, select the shortest-service-time option. Use EVAPerf to measure the load on each controller, and then redistribute LUN ownership as required.**

✓ **Best practice to optimize performance: Reassign LUN ownership after a failed controller has been repaired.**

Note
Striping LUNs within a disk group on the same controller provides no additional performance value. The EVA automatically stripes each LUN across all disks in a disk group.

Note
HP Continuous Access EVA requires that all LUNs within a DR group be owned by the same controller. Load balancing is performed at the DR group level and not the individual LUN. Additional configuration information can be found in the HP StorageWorks Continuous Access EVA administrator guide and the HP StorageWorks Continuous Access EVA planning guide.

LUN count influences performance

The EVA can achieve full performance with only a few LUNs per controller. However, the default queuing settings for some operating systems and/or host bus adapters (HBA) can constrain the performance of an individual LUN. In this case, additional LUNs or an increased OS/HBA default queue depth per LUN eliminates this constraint.

Pay particular attention to queue depth management for HP-UX. For HP-UX, the default queue depth is eight I/Os per LUN. This is insufficient for typical configurations that use only a few LUNs. You can find additional information on this subject in publications within the HP-UX community.

To optimize overall performance of arrays with snapclones, minimize the number of virtual disks that are issued a snapclone request. In this case HP recommends creating the minimum number of virtual disks and modifying the host operating system and/or HBA queue depth setting to provide sufficient queue with the small number of LUNs.

In Microsoft Windows® environments, attention to the Windows version and the HBA is required to understand the default I/O queue management operation. Recent releases of Windows do not have this issue.

Do not overlook the performance value of queue settings and LUN count. This is a common configuration error that can dramatically reduce performance.

- ✓ **Best practice to optimize performance: Follow operating system and application requirements for LUN count. If snapclones are used, then create as few LUNs as possible and manage the operating system and Fibre Channel adapter queuing appropriately.**

Transfer size influence on sequential performance

Applications, such as data warehouse and business intelligence, that have a high percentage of sequential write application can improve array performance by ensuring that the write transfer size is greater or equal to 32K and is a multiple of 8K (for example, 32K, 40K, 64K, 72K, 80K, … 128K). These transfer sizes simplify the cache management algorithms (which reduce the controller overhead to process a command) and reduce the total number of I/Os required to transfer a given amount of data. Storage systems typically choose these write transfer sizes, but operating systems, file systems, and databases also provide settings to manage the default transfer size.

- ✓ **Best practice to improve sequential write performance: Tune the write transfer size to be a multiple of 8K.**

Snapshots and snapclones influence performance

The EVA can make local copies of selected virtual disks. There are two types of copies: a full duplication (snapclones) and copy-on-write snapshots (either space efficient or fully allocated). Both types are easy to use and integrate well into typical data center processes.

The simplicity of use of the snap operations masks the internal complexity and data movement required to execute the copy commands. There are three phases to the execution of internal virtual disk copies: the metadata management, the write cache flush, and the data movement. Metadata management is the work the controller needs to perform to create the internal data structures to manage the new virtual disks. Data movement is the operation of copying the data. Metadata management is similar for all local copy operations; however, data movement differs.

A snapclone, as the name implies, makes a complete copy of an existing virtual disk. Snapclone creation places an additional workload on the disks of the target disk groups—the actual data copy. This workload interferes with the external workload during the creation of the snapclone. The observed impact is an increased command response time and a decrease in the maximum IOPS that the disk group can maintain. This performance impact continues until the virtual disk is completely copied. When the cloning operation completes, the performance impact ceases.

Snapshots take advantage of the HP StorageWorks EVA virtualization technology and copy only changes between the two virtual disks. This typically reduces the total data movement and associated performance impact relative to snapclones. However, there remains a performance impact. Snapshot uses copy-on-write technology, meaning that a block is copied during the host write operation. A block is copied only once. After it diverges from the original virtual disk, it is not copied again. Like snapclones, each block copied interferes with the host workload. However, for many applications, less than 10 percent of the data in a virtual disk typically changes over the life of the snapshot, so when this 10 percent has been copied, the performance impact of the copy-on-write ceases. Another characteristic of typical workloads is that the performance impact exponentially decays over time as the virtual disks diverge. In other words, the performance impact is greater on a new snapshot than on an aged snapshot.

With XCS 5.xx and greater, the metadata management, write cache flush, and data movement phases can be managed with separate commands—the three-phase snapshot and snapclone. This allows the overhead for the metadata management (the creation of a LUN) to be incurred once (and reused) and during low-demand periods. Using three-phase snapshot and snapclone commands can greatly improve response times during snap creation.

Remember, the snapclone operation is independent of the workload; the copy operation is initiated by the snapclone request, whereas the snapshot is driven by the workload and, by its design, must compete with the workload resources (that is, the disks).

To make a consistent copy of a LUN, the data on the associated disks must be current before the internal commitment for snapshot or snapclone. This requires that the write cache for a snap LUN be flushed to the disks before the snap commitment. This operation is automatic with the snap command. However, the performance impact of a snap operation can be minimized by transitioning the LUN to the write-through mode (no write caching) before issuing the snap copy command, and then reenabling write caching after the snap is initiated. The performance benefit is greatest when there are multiple snaps for a single LUN.

✓ Best practice for snapshot and snapclone performance: Use three-phase snapshots and snapclones.

✓ Best practice for snapshot and snapclone performance: Create and delete snapshots and snapclones during low-demand periods, or size the array to meet performance demands during snap activities.

✓ Best practice for snapshot and snapclone performance: Transition source LUN to write-through cache mode before snap initiation, and reenable write caching after the snap is initiated.

✓ Best practice for snapclone performance:

- Keep virtual disks as small as possible.
- Minimize the concurrent snapclone operations (use fewer virtual disks). Organize clone operations into consistency groups of virtual disks, and then clone consistency groups sequentially.

✓ Best practice for snapshot performance:

- Minimize the number of virtual disks with active snapshot copies. Use fewer virtual disks (it is better to have a few large virtual disks than many small virtual disks).
- Minimize the number of snapshot copies for a virtual disk. Do not keep extra snapshot copies without reason or plan for their use.
- Minimize the life of a snapshot copy. If snapshot copies are used for backup, consider deleting the snapshot virtual disk at the completion of the copy to tape.
- Delete snapshot virtual disks in order of age, oldest first.

Space-efficient snapshots use free space (capacity not reserved for normal or snapclone virtual disks) to store changed data. All space-efficient snapshot virtual disks in a disk group become inoperative when any space-efficient snapshot virtual disk in that disk group is denied a request to use additional free space. Always monitor free space. If the availability of space-efficient snapshot virtual disks is critical for the whole application availability, then overestimating the requirements for free space may be warranted.

In addition to the normal LUN divergence consumption of free space, a disk failure and the subsequent reconstruction can also compete for free space. After a reconstruction, the reserved space requirements for the protection level can cause the existing snapshot virtual disks to exceed the available free space and thus cause the snapshot virtual disks to become inoperative. See the best practice for capacity management and optimized availability to avoid this condition.

HP Continuous Access EVA and snapshots influence performance

Improvements in overall system performance can be realized when making snapshot copies of the remote target of a CA pair by temporarily suspending its DR group at the initiation of the remote snapshot copy.

The performance impact of an improperly sized (in a performance sense) snap copy LUN can cause HP Continuous Access EVA to suspend replication. System and application performance and availability can be improved by making this possible condition a planned event rather than an unplanned disruption.

✓ Best practice to optimize performance when making snapshots of remote virtual disks in DR groups: Suspend the DR Group, create the snapshots, and then resume the DR group.

Miscellaneous management best practices

Increasing capacity of the array

To minimize false indications of excessive errors, insert multiple disks carefully and slowly, pausing between disks. This carefulness allows the initial bus interruption from the insertion and the disk power-on communication with the controller to occur without the potential interruption from other disks. In addition, this process sequences leveling to not start until all the new disks are ready.

Although the array supports replacing existing smaller disks with larger disks, this process is time consuming and disruptive and can result in a nonoptimum configuration. Do this only if the option to build new disk groups and move existing data to the new disks is unavailable.

✓ **Best practice to optimize availability when adding disks to an array:**

- **Set the add disk option to manual.**
- **Add disks one at a time, waiting 60 seconds between disks.**
- **Distribute disks vertically and as evenly as possible to all the shelves.**
- **Unless otherwise indicated, add new disks to existing disk groups using the HP StorageWorks SSSU add multiple disks command.**
- **Add disks in groups of eight.**
- **For growing existing applications, if the operating system supports virtual disk growth, increase virtual disk size. Otherwise, use a software volume manager to add new virtual disks to applications.**

Disks groups and data security

Disk groups are self-contained components of the array; that is, the storage resources required for a disk group are contained completely within each disk group. A given disk group can be accessed only through LUNs created from that disk group. Lastly, a LUN can contain capacity from only one disk group.

Given these characteristics, applications or data centers that require data isolation for security objectives can accomplish these objectives by assigning unique security domains to separate disk groups.

These characteristics also make disk groups useful for tracking and allocating assets to specific groups within an organization.

✓ **Best practices for data security: Assign application and/or servers to separate disks groups. Use selective LUN presentation to limit access to approved servers.**

Note
Multiple disk groups increase the cost of the storage as well as a possible impact to the performance capability of the array. See the sections on disk group usage.

Best practice folklore, urban legends, myths, and old best practices

Urban legend: Prefilling new LUNs improves performance

There are two phases to the creation of a nonsnapshot LUN on an HP StorageWorks EVA. The first phase creates the metadata data structures. The second phase writes zeros to the associated sectors on the disks. Access to the new LUN is restricted during the first phase, but access is allowed during the second phase. During this second phase, host I/Os compete with the zeroing operation for the disks and controller, and performance is impacted.

When the zeroing completes, the array is capable of normal performance. The duration of the zeroing operation depends on the size and Vraid level of the LUNs, the number of disks in the disk group, and the host demand. In a test case using 168 36-GB disks to create 64 46-GB LUNs, zeroing required 30 minutes to complete. In this test case, there was no other load on the EVA.

Zeroing is a background operation, and the time required to complete zeroing increases when other workloads are present on the array. The virtual disk is fully accessible during this time; however, the performance of the virtual disk during zeroing does not represent future performance.

For nonsnapshot virtual disks, the EVA always maps data on the disks in its logical order. Unlike other virtual arrays for which the layout is dynamic and based on the write order, the EVA data structure is predictable and repeatable.

Given the new-LUN zeroing operation and the predictable data layout, there is no reason (with the exception of benchmarking) to preallocate data by writing zeros to the virtual disk on the EVA.

Note
Prefilling a new LUN before a performance benchmark allows the internal zeroing operation to complete before the benchmark begins.

Urban legend: Use a 12-shelf system to maximize availability

The logic for this legend is that the largest RSS can be 11 disks, and thus a 12-shelf array somehow assures optimum RSS distribution. That is not how the EVA functions. RSS distribution works with any number of supported shelves. There is no advantage to any specific number of shelves. **A better best practice is to build disk groups in multiples of eight disks.**

Urban legend: Disk groups in multiples of eight disks if only eight shelves

This myth attempts to optimize availability by matching disk groups and shelves. This is an incorrect assumption about the RSS distribution algorithm. **The best practice is always to group eight disks, regardless of the number of shelves.**

Urban legend: Disk groups of 12 or fewer disks if only six shelves

This legend attempts to increase availability by optimizing RSS to disk shelves. However, many small disk groups have poor capacity utilization and possibly poor performance. If the objective is higher availability than the default RSS management, a better practice would be to **use as few disk groups as possible and Vraid 1.**

Urban legend: Number of disks in a disk group is based on shelf count

There is no availability advantage created by this rule, and the many disk groups that can result can decrease the capacity utilization of the array and hamper performance management. **Disk groups should be in groups of eight disks, as large as possible and as few as possible.**

Most misunderstood best practice: 5 GB of free space is sufficient for array optimization in all configurations, or 90 percnet/95 percent LUN allocation is a good rule of thumb for free space.

These rules are over simplifications and are correct for only some configurations. Five GB is sufficient only if no snapshots or DR groups exist and some availability tradeoffs are accepted (proactive disk management events). Ninety percent is frequently more than is really required for optimum operation, thus unnecessarily increasing the cost of storage. **Read and follow the free space management best practices.**

Old best practice: Microsoft DiskPar utility is required for optimum Windows performance with the EVA.

Previous versions of EVA firmware poorly handled the case when transfers were not 2K-aligned. This alignment requirement was eliminated with the 5.xxx versions of XCS. **The use of DiskPar neither enhances nor detracts from EVA sequential performance.**

Appendix A

Example of capacity planning process

An array is configured with two disk groups. Disk Group 1 has 30 300-GB drives. Disk Group 2 has 100 146-GB drives.

The capacity management process is as follows:

Disk Group 1 does not use HP Business Copy EVA or HP Continuous Access EVA. The requirement for this disk group is to tolerate one drive failure, and one proactive disk management event.

First, set the protection level to single; the remaining raw capacity is now about 8.4 TB - (30-2)*300 GB.

Next, set the occupancy alarm. One PDM event will require 2*300 GB or 600 GB. Leveling optimization requires an additional 5 GB. The total 605 GB is 7.2 percent of the raw capacity. Rounding up, the occupancy alarm should be set to 92 percent.

No snapshot or remote replication capacity is required, so all remaining capacity can be allocated to LUNs.

Disk Group 2 uses HP Business Copy EVA but not HP Continuous Access EVA. Ten Vraid1 parent LUNs of 500 GB each are required with an associated Vraid1 snapshot of each LUN. The requirement for this disk group is to tolerate two disk failures and two proactive disk events.

The total raw capacity of the disk group is 14,600 GB.

First, set the protection level to double. The remaining raw capacity is 14,016 GB.

Next, set the occupancy alarm. Two PDM events require 584 GB of raw capacity. Leveling optimization requires another 5 GB. The total 589 GB represents 4.2 percent of the raw capacity. Rounding up, set the occupancy alarm to 95 percent. The occupancy alarm is now at 13,300 GB raw.

The 10 500-GB Vraid1 LUNs require approximate 10,000 GB of raw capacity. The snapshot capacity requires 20 percent of 10 TB, or 2,000 GB raw. Total required is 12,000 GB.

The required capacity of 12 TB is less than the occupancy alarm at 13.3 TB, so all is good.

If, or when, the occupancy alarm is activated, the cause of the alarm must be determined and resolved. If a disk failure or a proactive disk event is in progress, the failing disk must be replaced after the reconstruction or ungroup is completed. If neither of these events is in progress, the snapshots are consuming more capacity than anticipated. In this case, additional disks need to be added to the disk group, or LUNs must be deleted. Failure to address the occupancy alarm may result in the deactivation of all snapshot LUNs or other availability issues.

Summary

All of the preceding recommendations can be summarized in a table. This not only makes it relatively easy to choose between the various possibilities, but it also highlights the fact that many best practice recommendations contradict each other. **In many cases, there is no single correct choice** because the best choice depends on whether the goal is cost, availability, or performance. In some cases, a choice has no impact.

Table 2. Summary

	Cost	Availability	Performance
Mixed disk capacities in a disk group	Yes	-	No
Number of disk groups8	1	>1	As few as possible
Number of disks in a group	Maximum	Multiple of 8	Maximum
Total number of disks	Maximum	Multiple of 8	Maximum
Higher performance disks	No	-	Probability
Mixed disk speeds in a disk group	Yes	-	Acceptable
Redundancy level	0	1 or 2	1
LUN count9	-	-	-
Read cache	-	-	Enabled
LUN balancing	-	-	Yes

[8] Consult application or operating system best practices for minimum number of disk groups.

[9] Check operating system requirements for any special queue depth management requirements.

Glossary

access density	A unit of performance measurement, expressed in I/Os per second per unit of storage. Example, I/Os per second per GB.
data availability	The ability to have access to data.
data protection	The ability to protect the data from loss or corruption.
disk group	A collection of disks within the array. Virtual disks (LUNs) are created from a single disk group. Data from a single virtual disk is striped across all disks in the disk group.
free space	Capacity within a disk group not allocated to a LUN.
leveling	The process of redistributing data to existing disks. Adding, removing, or reconstruction initiates leveling.
LUN	Logical unit number. An addressable storage collection. Also known as a virtual disk (Vdisk).
occupancy	The ratio of the used physical capacity to the total available physical capacity of a disk group.
physical space	The total raw capacity of the number of disks installed in the EVA. This capacity includes protected space and spare capacity (usable capacity).
protection level	The protection level defines the reserved space used to rebuild the data after a disk failure. A protection level of none, single, or double is assigned for each disk group at the time the disk group is created.
protection space	The capacity that is reserved based on the protection level.
reconstruction, rebuild, sparing	Terms used to describe the process of recreating the data on a failed disk. The data is recreated on spare disk space.
reserved space	Same as protected space.
RSS	Redundancy Storage Set. A group of disks within a disk group that contain a complete set of parity information.
usable capacity	The capacity that is usable for customer data under normal operation.
virtual disk	A LUN. The logical entity created from a disk group and made available to the server and application.
workload	The characteristics of the host I/Os presented to the array. Described by transfer size, read/write ratios, randomness, arrival rate, and other metrics.

For more information

The intent of this paper is to provide technical details to optimize the HP StorageWorks Enterprise Virtual Array (EVA) for specific applications. It is not intended to be a general-purpose tutorial on EVA operation. HP and its partners provide additional information and services to help you optimize your EVA. For more information on the HP StorageWorks EVA, go to http://www.hp.com or contact your local HP sales representative.

© Copyright 2005 Hewlett-Packard Company, L.P. The information contained herein is subject to change without notice and is provided "as is" without warranty of any kind. The warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

Microsoft and Windows are U.S. registered trademarks of Microsoft Corporation. Oracle is a registered U.S. trademark of Oracle Corporation, Redwood City, California. Linux is a U.S. registered trademark of Linus Torvalds.

4AA0-2787ENW, October 2005