

HP StorageWorks Enterprise Virtual Array configuration best practices — white paper



| | |
|--|----|
| Executive summary..... | 3 |
| Introduction..... | 3 |
| Overview..... | 3 |
| Best practices to get started..... | 3 |
| Detailed best practices | 5 |
| The first best practice..... | 5 |
| Best practices to optimize the cost of storage..... | 5 |
| Mixed disk capacities influence the cost of storage..... | 5 |
| Number of Disk Groups influences the cost of storage..... | 5 |
| Quantity of disks influences cost of storage..... | 6 |
| Disk performance influences the cost of storage | 6 |
| FATA disks influence the cost of storage | 6 |
| Best practices to optimize availability | 7 |
| RAID level influences availability | 7 |
| Shelf and disk organization influences availability | 7 |
| Replacing a failed disk influences availability | 8 |
| Protection Level influence on availability..... | 9 |
| Number of Disk Groups influences availability..... | 9 |
| Number of disks in Disk Groups influences availability..... | 10 |
| Fault monitoring to increase availability | 10 |
| Backup processes to optimize recovery | 10 |
| Remote mirroring and availability..... | 10 |
| System fault-tolerant clusters and availability | 11 |
| Best practices to optimize performance..... | 11 |
| Number of disks influences performance | 11 |
| Number of Disk Groups influences performance..... | 12 |
| Disk performance selection influences array performance..... | 12 |
| Write cache management influences performance..... | 13 |
| Vraid 0 influence on performance | 14 |
| Mixing disk performance influences array performance | 14 |
| Mixing disk capacities influences performance..... | 14 |
| Read cache management influence on performance | 15 |
| Controller balancing influence on array performance | 15 |
| LUN count influences performance..... | 16 |
| Snapshot and Snapclone performance management..... | 16 |
| Windows data alignment and performance | 17 |

| | |
|---|----|
| Best practices to manage capacity | 18 |
| Increasing capacity of the array | 18 |
| Capacity management and optimized operation..... | 18 |
| Disks groups and data security..... | 19 |
| Best practice folklore, urban legends, and myths..... | 19 |
| Pre-allocation | 19 |
| Use a 12-shelf system to maximize availability..... | 19 |
| Disk groups in multiple of eight disks if only eight shelves..... | 19 |
| Disk groups of 12 or less disks if only six shelves | 19 |
| Number of disks in a Disk Group based on the shelf count..... | 20 |
| Summary | 20 |
| Glossary..... | 21 |
| For more information..... | 22 |

Executive summary

A business value of storage virtualization is simplified management. Storage systems that are simple to administer reduce configuration errors and save management time and money. Even with simplified storage management, knowledge of the best practices for configuration and management tasks can further optimize the storage for a specific application. Getting the most from your enterprise class storage has never been easier. This paper highlights the most common configuration rules and trade-offs. Following these guidelines will help ensure each HP StorageWorks Enterprise Virtual Array (EVA) is optimized for its intended application.

Introduction

One of the many design objectives for the HP StorageWorks EVA program was to provide maximum real-world performance and a reduction of the storage management costs. This objective resulted in the design of an intelligent controller that minimized the number of user controllable tuning parameters. This is in sharp contrast to traditional disk arrays, which typically have many tunable settings for both individual logical unit numbers (LUNs) and the controller itself. Although tunable settings might appear to be desirable, there are problems associated with them:

- It is difficult and time consuming for administrators to set the parameters appropriately. Many of the 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 level of expertise necessary to maintain a traditional array in optimum configuration.
- As the I/O workload changes, many of the parameters that were previously set might no longer be appropriate. Overcoming this problem requires continual monitoring, which is impractical and costly in most situations.

Because of such concerns, the HP StorageWorks EVA algorithms reduce the number of parameters that can be set by the user, opting instead to embed intelligence within the controller itself. The controller has a better view of the workload than most administrators and can be far more dynamic in responding to changes in the workload. 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, some of the 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 manage the configuration to optimize its attributes for a specific application.

These configuration choices include:

- Number of Disk Groups
- Type of Disk Group
- Type and number of disks in a Disk Group
- Vraid levels (0, 1, and 5)
- Disk failure protection level (none, single, and double)
- Cache settings

Each of these topics is covered in more detail in the following sections. Note that it might **not** be possible to simultaneously optimize a configuration for cost, performance, and availability. **There might be conflicting recommendations** because one of the three choices must yield to demands made by another. For example, Vraid0 is clearly the best solution from a strictly cost standpoint because nearly all the storage is available for user data. But because Vraid0 offers no protection from a disk failure, either Vraid1 or Vraid5 would be a better choice for availability. Other trade-offs are minor by comparison but must sometimes still be made. There is no “best” choice in these situations because “best” depends on the priorities of a particular environment.

Best practices to get started

The best practices itemized in this document can be daunting for the uninitiated user. The following table attempts to summarize a subset of best practices that are typically selected as an optimization of availability, performance,

and cost. The objective of this list is to provide a simple starting point. However, as with any generalization, they might not be the best choice for all applications. Detailed information can be obtained by the associated sections within this document.

The following table, and the remainder of this document, are best practice recommendations; they are not configuration or usage requirements. The HP StorageWorks EVA and HP support a wide variety of EVA configurations. All supported configurations prove availability and performance features.

Table 1.

| Best practice | Source | Discussion |
|---|---|--|
| Shelves of eight or more | Availability Optimization | Eight or more shelves allow the EVA to optimize the distribution of disks in the Redundancy Storage Set (RSS). |
| Disks in multiples of eight | Availability Optimization | Allows the assignment of disks in multiples of eight to a Disk Group. Disk should be distributed as evenly as possible to all shelves. |
| Disks in Disk Groups in multiples of eight | Availability Optimization | Eight disks per Disk Group allow the EVA to optimize the distribution of disks in the RSS. |
| Disks of same size and speed in a Disk Group | Performance Optimization | Ease of management and performance utilization. This configuration avoids any issues with access-density of the Disk Group. |
| As few Disk Groups as possible | Performance Optimization | Performance is optimized when the array is allowed to stripe data to as many disks as possible. |
| Protection Level of One | Availability Optimization | In most installations, Protection Level of one provides adequate availability. See detailed discussion of mitigating conditions. |
| Separate Disk Group for database logs | Availability Optimization for Database applications | Provides consistent and current database restore from external media if data/table space Disk Group is corrupted. |
| Separate Disk Groups with FATA disks | Cost Optimization | FATA disk provide the lowest cost of storage per megabyte. Separate Disk Groups provide increased data protection capability. |
| Use 15K-rpm disks | Performance Optimization | 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 Optimization | Load balancing the workload as evenly as possible to both controller provides the highest performance utilization. |
| Use Vraid1 | Availability Optimization | Vraid1 provides the highest level of data protection and availability. |
| Use Vraid5 | Cost Optimization | Vraid5 provides the lowest cost of protected storage. |
| HP StorageWorks Continuous Access EVA or host-based mirroring | Availability Optimization | 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 Optimization | All data center best practices include processes to regularly copy the current data set to external media or near-line devices. |
| ISEE | Availability Optimization | Provides automated messaging to HP Support or your internal IT personnel of abnormal EVA status. |

Detailed best practices

The best practices itemized here are just that, best practices. They are not configuration requirements; HP supports the EVA in many configurations not stated here. The user manual or installation guide will define the configuration requirements. The best practices have been developed to help make good configuration choices when alternatives exist.

The first best practice

The first best practice is to stay as current as possible with the latest Version Controller Software (VCS) release. HP continually improves the performance, reliability, and functionality of the EVA. Your EVA can only benefit from these investments if it is current with the latest VCS release.

The best practices in this paper are based on VCS 3.020.

✓ **First best practice: Stay as current as possible with the latest VCS release.**

Best practices to optimize the cost of storage

The term “cost,” as used 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 useable data as seen by the operating system and 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 Enterprise Virtual Array (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, the array with disks of different capacities. The EVA greatly simplifies the management of these configurations. However, understanding of the configuration rules of the EVA can help optimize a solution using disks with different capacities.

The disk failure protection level is a storage administrator-selectable parameter for each Disk Group that defines the number of disk failure then auto-reconstruction cycles the Disk Group can tolerate. The protection level can be dynamically assigned to each Disk Group to a value of none, single, or double. Conceptually, it reserves space to handle 0 (none), 1 (single), or 2 (double) disk failure-restore cycles. The space reserved is specific to a particular Disk Group and cannot span Disk Group boundaries.

The software algorithm for reserving rebuild space finds the largest disk in the Disk Group; doubles its capacity; multiplies the result by 0, 1, or 2; and then removes that capacity from free space. The net effect is the rebuild space is distributed across all disks in the Disk Group. Unlike traditional arrays, the EVA does not reserve physical disks; 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, as well as the smaller disks. The Vraid1 recovery algorithm requires Vraid1 spare space be in predetermined disk pairs—when one of the pairs fail, the contents is moved to a new pair, thus twice the capacity is reserved. The advantages of the distributed sparing are two-fold. First, the performance value of all the disks is used, and second, unlike a traditional spare, there is no risk that a reserved disk is unusable (failed) when needed.

Because rebuild space is not shared across Disk Groups, it is more efficient to have the least number of Disk Groups possible, thus minimizing the rebuild space. However, when the first few larger disks are introduced into a Disk Group, their resulting usable capacity is the same as if they were smaller disks. This continues 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 usage 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 fundamental building blocks of the EVA. Disk groups are self-contained protection domains, meaning the data redundancy information for any disk must be contained within the associated Disk Group. In

addition, the EVA implements distributed spare space, meaning the spare space for reconstruction is not reserved by whole disks, but distributed to many disks. Because of disk protection domains and the distributed sparing; spare space cannot span Disk Groups. 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 six to 11 disks. Redundancy sets are discussed in more detail in the Availability section. The net effect is multiple Disk Groups are not needed for Vraid5 availability. The result of these architectural design features is multiple Disk Groups consume additional usable capacity for spare space but deliver disk sparing capabilities that exceed the statistical requirements given the number of disks in a Disk Group.

Available capacity is the space in a Disk Group that remains after space has been reserved for the protection level. It is used for creation of virtual disks, snapshots, and Snapclones, as well as temporary space used by the EVA for certain internal operations. Available capacity decreases when virtual disks, snapshots, or Snapclones are created or when physical disks are removed. Available capacity increases when virtual disks or snapshots are deleted or when additional physical disks are added to the Disk Group.

As with traditional arrays, multiple Disk Groups can lead to what is known as “stranded capacity.” Vdisks must be created from a single Disk Group, thus the array might contain the desired usable capacity, but the capacity is distributed to many Disk Groups. Unlike traditional disk arrays, the EVA can easily support very large Disk Groups, thus eliminating the stranded capacity issues.

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

Note: This would be a good place to point out to the reader that the discussion on Disk Groups within the performance and availability sections should be studied to understand the trade-offs associated with the Disk Group count. As an example, a single Disk Group is not a best practice for many data bases applications. Before acting on any best practice suggestion, the price, performance, and availability of the suggestion should be understood.

Quantity of disks influences cost of storage

The lowest cost per storage unit solution 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 the largest disks. Thus, the lowest cost of storage solution is to fill the array with as many of the largest disks that are supported. The HP StorageWorks Enterprise Virtual Array 5000 (EVA5000) can support up to 240 disks and up to 35 TB useable capacity. 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 disk drives as possible, using the largest capacity disks.**

Note: The section within this document on disk count and performance optimization and the section on disk count and availability should be studied before increasing the disk count.

Disk performance influences the cost of storage

Larger disks usually offer better price per capacity. Although prices continuously change, it is reasonable to assume that you can purchase more capacity for the same price when using larger drives than with smaller drives. Similarly, higher performance drives, such as 15,000-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.**

FATA disks influence the cost of storage

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

The EVA requires that FATA disks be organized in a separate Disk Group.

The best application match for FATA disks is to assist online backup and recovery processes. 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 available 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 should not be considered a replacement for off-line backup. All best practices include the retention of data on an external device or media for disaster recovery.

Best practices to optimize availability

The HP StorageWorks Enterprise Virtual Array (EVA) is designed for business critical applications. It has designed-in redundancy features that enable the array to continue operation in the event of a wide variety of failures. The EVA is also designed to be flexible. In a statistical and design sense, some configurations allow a higher level of availability by limiting the probability of exposures to failure scenarios that exceeds the fault-tolerant design characteristic of the array. The goal is to create configurations that have the greatest number of orthogonal protection domains so that a failure in one domain does not reduce the resiliency of another domain.

All supported configurations have some degree of tolerance to failures but not necessarily all failures. As an example, Vraid 0 is supported, it 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 address configuration options to improve the availability characteristics of sequential and multiple simultaneous failure scenarios and discuss system and application configurations to improve availability.

RAID level influences availability

While Vraid 5 provides availability and data protection features sufficient for most high-availability applications, some applications should consider the additional availability and data-protection features of Vraid1. Vraid1 configurations can continue operation in failure scenarios where a Vraid5 configuration would be incapable of providing protection. A statistical model of the EVA, for an equivalent usable capacity, shows that Vraid 1 provides over four times the data protection of Vraid 5¹.

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 usage for Vraid1 for critical files or data sets. For example:

- For database applications, select Vraid 1 for log files.
- For Snapclone/snapshot applications, select Vraid 1 for active data sets and Vraid 5 for Snapclone/snapshot usage.

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

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

Shelf and disk organization influences availability

Within a Disk Group, the EVA creates a subgroup of disks for redundancy purposes—the Redundancy Storage Set (RSS). Each of these individual redundancy sets contains sufficient redundancy information to enable continued operation in the event of a disk failure within that redundancy set. In this manner, the EVA can sustain multiple, simultaneous disk failures while not losing user data—as long as no more than one disk per redundancy set fails. These redundancy sets are built when a Disk Group is created, and additional sets are created as necessary by the controller firmware when disks are added to the Disk Group.

The target size of each redundancy set is eight disks, with a minimum size of six, and a maximum size of 11. As disks are added to a Disk Group, the redundancy set automatically increases in size until it reaches 12 disks. At that point, it is split 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

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.

redundancy set contains eight disks, with the possible exception of the last set, which contains between six and 11 disks. All of this is managed by the EVA controllers, with no user intervention required.

To maximize the array availability, disks should be arranged vertically within the shelves, distributed as evenly as possible in the shelves. The number of disks in the shelves should not differ by more than one. The highest level of redundancy is achieved when eight or more disk shelves are used. In this way, the array minimizes the conditions in which the RSS has two disks in the same disk shelf.

The steps taken to accomplish this vary, depending on whether the array contains Vdisks configured only as Vraid1 or Vraid5. If a mixture of Vraid1 and Vraid5 are contemplated for a single Disk Group, then the guidelines for Vraid5 should be followed. If only Vraid1 will be used, then the Vraid1 guidelines are appropriate.

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

- **Use a minimum of eight shelves in a configuration for Vraid5. If the array contains fewer than eight shelves, consider Vraid 1 for highest availability.**
- **All disks should be arranged in a vertical fashion, distributed among the shelves as evenly as possible.**
- **The total number of disks and the total number of disks in a Disk Group should be a multiple of eight.**
- **When creating a Disk Group, let the EVA choose which disks to place in the group.**

With Vraid1, the EVA firmware attempts to place the individual members of a mirror pair on different shelves. Because of this, the guidelines are much simpler, and the suggested number of shelves can be less than eight.

✓ **Vraid1 best practice to optimize availability:**

- **All disks should be arranged in a vertical fashion, distributed among the shelves as evenly as possible.**
- **The total number of disks and Disk Groups should be a multiple of two.**
- **When creating a Disk Group, let the EVA choose which disks to place in the group.**

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 the additional operations required to write the redundancy information. Thus, Vraid 0 provides the best performance for write-intensive workloads and provides the lowest cost of storage.

✓ **Vraid0 best practice to optimize availability: Use of Vraid0 is not advised as a best practice for availability.**

Replacing a failed disk influences availability

Following the rules on shelf and disk organization is the best protection against potential data loss and loss of availability in the event of a disk failure. However, when a disk fails, there are additional steps to follow to minimize the risk of data loss or unavailability.

HP service engineers are trained on the proper EVA repair procedures and are on the alert for abnormal conditions that might 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, these basic rules should be followed.

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 the RSSs to the active disks.

✓ **Disk replacement availability best practice: Use the following procedure:**

1. **Wait for the reconstruction to complete. This is signaled by an entry in the event log.**
2. **Use Command View to Ungroup the disk—this assures the disk is not a member of a Disk Group.**
3. **Replace the failed disk. The new disk must be inserted into the same slot as the failed disk.**
4. **Add the new disk into the original Disk Group; ensure the Disk Addition Policy is set to Manual mode.**

- ✓ **Best practice to optimize availability: The failed disk should remain in the array until reconstruction has completed. The replacement disk should be inserted into the same slot as the failed disk.**
- ✓ **Best practice to optimize availability: The Disk Addition Policy should be set to manual; the new disks should be added to the Disk Group of the failed disk.**

Protection Level influence on availability

The Protection Level defines the number of disk failure auto reconstruction cycles the array can accomplish without physically replacing any failed disks. An important concept to understand is the cycle of a disk failure, followed by the controller re-creating (or rebuilding) 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.

As an example, a Protection Level of 1 provides continued operation in the event of two disk failures, assuming the restore from the first failure completes before the second disk fails.

Although the HP StorageWorks Enterprise Virtual Array (EVA) is resilient to a single disk failure with a Protection Level of 0, it is not a best practice configuration.

Conversely, the statistical availability of disks and the typical service time to replace a failed disk (MTTR²) would indicate that a Protection Level of two would be unnecessary in Disk Groups of fewer than 168 disks in all but the most conservative installations. A mitigating condition would be if the service time (MTTR) might exceed seven days, then a Protection Level of 2 might be considered.

- ✓ **Best practice for Protection Level and availability: Use a Protection Level of one.**

Number of Disk Groups influences availability

Although the EVA offers numerous levels of data protection and redundancy, a catastrophic failure³ 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 can complete. The probability of these events is very low; however, installations requiring the highest levels of data availability should consider creating separate Disk Groups for independent failure domains. Multiple groups result in a slightly higher cost of ownership and potentially lower performance utilization, but the increase in availability might be the correct decision for a very high availability application.

The strategy is to keep application recovery data on a separate Disk Group from the source data. The typical use is either Snapclones or database log files created in a separate Disk Group. If the primary Disk Group fails, the recovery Disk Group is still 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 level data set from the last backup. A practical example of this is a database that contains both table space and log files. In this instance, placing the table space in one group and duplexing the log files (a typical feature of some databases) to both the table space Disk Group and another group ensures that loss of an entire Disk Group does not prevent recovering the data.

As an example:

- Disk Group 1: Database table space, database log, and index files (if duplexed by database)
- Disk Group 2: Database log and index files

The number of disks in Disk Group 2 should be sized by the sequential performance demand of the log workload. Typically, this will result 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, will not affect the cost of the solution.

A variation on this configuration is to create two separate Disk Groups for the log and index, and a third for the table space. This configuration has a slightly higher cost but appeals to those looking for symmetry in the configuration.

- Disk Group 1: Database table space
- Disk Group 2: Database log and index files
- Disk Group 3: Database log and index files (if duplexed by database)

² MTTR—Mean Time to Repair

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

This logic also applies to Snapclones used for recovery. Assigning a Snapclones to a different Disk Group helps ensure a recoverable image if the original Disk Group becomes unavailable.

- ✓ **Best practice to optimize availability: For critical database applications, consider placing data and log files in separate Disk Groups.**
- ✓ **Best practice to optimize availability: Assign Snapclones to a different 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 would defeat the availability value of multiple Disk Groups.

Number of disks in Disk Groups influences availability

Disks assigned to Disk Groups in multiples of eight provide the controller the opportunity to optimize the placement of the RSS disk sets. In this case, the controller will have the optimum number of disks in each RSS of a Disk Group. See the discussion on [shelf and disk optimization](#).

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

Fault monitoring to increase availability

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

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

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

Backup processes to optimize recovery

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

Vsnap should not be considered the only form of backup. Vsnap and Snapclone are a complement to any backup strategy that includes full copies to off-line or near-line storage. In this application, Snapclones can provide alternatives for reducing recovery time by providing the first option for recovery.

It is important to not only perform regular backups, but to also test the restore process twice a year. The EVA greatly simplifies the testing process by providing a simple process to create Disk Groups or Vdisks. This capacity can be used for the testing, then easily deleted and the storage used for other purposes.

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

Remote mirroring and availability

HP StorageWorks Continuous Access EVA is an optional feature of the array that enables real-time remote mirroring. Continuous Access EVA can keep multiple, simultaneous copies of selected LUNs at local or remote EVA sites. Catastrophic EVA or site failures can be protected against with the second copy of the data. Continuous Access EVA can work as a stand-alone product or in combination with system software.

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

System fault-tolerant clusters and availability

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

Continuous Access EVA can be used for simple mirroring to increase data availability or as a component of a complete DR solution. Unlike Continuous Access EVA only, a DR 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 Enterprise Virtual Array (EVA) is supported in many of these solutions. Contact HP or your operating system vendor for more details. For those applications requiring the highest level of availability, these system level solutions should be considered.

- ✓ **Best practice to optimize for the highest levels of application availability: Consider disaster recovery solutions to provide continuous application operation.**

Best practices to optimize performance

Experience shows that high performance and low cost are generally an inverse relationship. While this continues to be true, the HP StorageWorks EVA virtualization technology can significantly reduce the cost of high performance. The following section outlines configuration options for optimizing performance and price: performance, although sometimes at the expense of either, or both, cost and availability.

Performance management typically follows into 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, using the EVA and taking advantage of its virtualization technology, workload distribution is the simplest technique to maximizing real-world performance. While it is still possible to manage performance using contention management with the EVA, the opportunity to incorrectly match the demand to the resources is large, thus not achieving the best performance or utilization.

This is a key concept and the basis for many of the best practices for EVA performance optimization. The following sections explore how to obtain the best performance using the workload distribution technique.

One additional point about performance optimization—demand versus capability. The suggestions in this paper discuss improving the performance capabilities of the array. Additional array performance improvements will have very little effect on application performance, when the performance capabilities of the array exceed the demand from the 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 will have little effect if the driver continues to request the car travel at only 30 MPH. If the driver does not demand the usage of the additional power, the capabilities will go unused.

The easiest way to determine if an application could take advantage of array performance tuning is to study the queues in the I/O system and the array. If there is little or no queuing, then additional performance tuning is unlikely to yield better application performance. This is not to say that these suggestions are not valuable. The point is if the array performance capabilities exceed the demand, then these suggestions might yield only a modest gain.

Number of disks influences performance

In general, adding more disks provides better performance under typical workloads. However, there is a diminishing return in performance of an additional disk 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 are

dependent on the characteristics of the application workload. As an example, sequential workloads require fewer disks than random workloads to reach the point of diminishing return.

For some applications, where the additional disks are needed only for capacity and the application does not require additional performance, increasing the number of disks in the array to its 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 delays, the additional disks will not increase 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, a performance evaluation should be considered. As a general rule of thumb, when the array is configured with 168 or more disks, it is advisable that a performance analysis review be completed to determine the preferred strategy for increasing the capacity of the array. The performance review will study the characteristics of the workload and array configuration and then provide a prediction of expected array performance with the additional disks. This prediction can then be compared with the application requirements and with other 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 Vdisk will improve the performance potential of a Vdisk. Because a Vdisk can only exist 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 the total disk utilization, and thus, 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 does 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: The discussions within this document on the number of Disk Groups and availability should be studied before considering a single Disk Group.

Disk performance selection 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. An examination of the disk data sheets would confirm that the average, sustained, large-block transfer rates are similar for both 10K- and 15K-rpm disks of the same generation. As such, 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% to 40% in the request rate when changing from 10K-rpm to the equivalent number of 15K-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: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 is in the range of 30% to 40%, if the 15K-rpm disks are more than 30 to 40% more expensive than the 10K-rpm disk, then it makes sense to purchase a larger number of 10K-rpm disks. As an example (and remember, this is only an example, contact HP for the current pricing), consider a system with a total budget of \$100,000. The goal is to get the best performance possible for this amount of money. The results are shown in Table 2, with an assumed cost for disk storage, and where the “IOPS” field represents the request rate at a response time of 30 ms for a typical I/O workload.

Table 2. Disk price:performance analysis example (not current prices)

| | IOPS | Cost | \$/IOP | Disk count at \$100,000 budget | Total IOPS at \$100,000 budget |
|-----------------------|-------------|----------------|---------------|---------------------------------------|---------------------------------------|
| 72 GB, 10K rpm | 110 | \$2,300 | \$20.91 | 43 | 4,730 |
| 72 GB, 15K rpm | 150 | \$3,400 | \$22.67 | 29 | 4,350 |
| Best choice | 15K rpm | 10K rpm | 10K rpm | | 10K rpm |
| Best choice advantage | 35% | 48% | 9% | | 9% |

As can be seen, although a 15K-rpm disk offers higher performance than the equivalent capacity 10K-rpm disk, the lower cost of the 10K-rpm disk means that more 10K-rpm disks can be purchased for the same total cost. The increased number of disks translate directly into higher overall performance. The preceding table illustrates a price:performance analysis; however, disk pricing can be very volatile, so you should recompute the price:performance with current price and performance data.

- ✓ **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.**

Write cache management influences performance

The purpose of cache mirroring is to prevent data loss in the unlikely event of a controller failure. In operation, writes from the host are sent to the controller that manages the target LUN. That controller stores the write data in its non-volatile cache and sends the data across the mirror port to the other controller. The second controller stores a copy of the data in its cache and then signals completion to the original controller. The original controller then signals completion of the request to the host. The actual write to the disks will occur at a later time based on the write cache flush algorithms. Because there are two independent copies of the data maintained in separate caches, failure of a cache memory board does not result in loss of the data.

The data movement operation to mirror (copy) the write data requires controller resources, thus there is an impact on the total performance capability of the controller during write operations. The level of impact depends on many factors, including the percentage of write operations, the size of the write requests, and the overall write request rate as presented by the various hosts.

From a performance perspective, it is possible to obtain gains in performance when a LUN is configured with cache mirroring disabled. It is important to understand that disabling cache mirroring can result in data loss if the controller that contains the data that has yet to be written to disk should fail.

However, there are applications in which the data is replicated through other means, and the performance advantages gained through disabling write caching outweigh the additional management time to recover the lost data. The following are examples of applications that can warrant disabling write cache mirroring:

- Initial load, or restore from tape operation
- Replicated databases
- Synchronous Continuous Access replication applications
- Temporary storage used by databases

Write cache mirroring is a dynamic setting and can be enabled or disabled after the Vdisk is created.

- ✓ **Best practice to optimize performance: Accepting the possible availability loss of such a configuration and disabling write cache mirroring can result in increased write performance.**

Vraid 0 influence on performance

Vraid 0 provides the best random write workload performance; however, Vraid 0 has no protection from a disk failure. In some cases, when the risk is well understood, appropriate use of Vraid 0 might improve overall application performance. However, for planning purposes, you should plan on a Vraid 0 Vdisk with 16 disks to lose data two or three times in the life of the array.

Like disabling write cache, but more likely to lose data than a controller failure, 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 the use of external replication only ensures that no data is lost. Loss of a Vdisk will cause an interruption in service and require manual steps to recover the Vdisk and application operation.

- ✓ **Best practice for Vraid 0 and performance: Accepting the possible data and availability loss, Vraid0 should be considered for non-critical storage needs only.**

Mixing disk performance influences array performance

Although grouping drives by type and speed might seem easier to manage, there are situations in which adding higher performing disk drives into an existing group can be beneficial. From a performance standpoint, 15K-rpm disks will perform 30% to 40% faster than 10K-rpm disks with a small-block, random access workload. Mixing drive speeds within a single group would then result in host-level read latencies that could vary depending on which drive processes the I/O request. (Remember, writes are acknowledged as complete after they are accepted into the write cache.) Although variations in performance are generally not desirable, it is easier to assure higher total performance utilization if the disks are all in a single Disk Group. It is also important to note that mixing drives of different speeds in the same Disk Group does not slow down the accesses to the speed of the slowest drive in the group.

The basic rule of thumb is that the overall performance of a LUN is dependent on the number of physical disks underneath that LUN. If drives of differing speeds are separated into different Disk Groups, then it follows that LUNs in each Disk Group will have fewer drives than a single Disk Group. Thus, the performance of a LUN will be lower than it would have been with all drives in the same group. Because it is difficult to manually balance the I/O load between different LUNs (striping would be the same as adding the disks in a single LUN), the realistic total performance of two LUNs in separate Disk Groups will be less than the total performance of those same two LUNs in a single large Disk Group.

- ✓ **Best practice to optimize performance: It is acceptable to create a single Disk Group with mixed performance disks.**

Mixing disk capacities influences performance

This is probably the most complex option to create a best practice recommendation. Many variables can influence the "right" choice.

For the initial configuration or when a choice is available, it is better to create or expand Disk Groups using disks of equal capacity. This eliminates any chance of inconsistent performance caused by potential hot-spots or the access density of a separate Disk Group.

Unfortunately, if there is a need to grow the array after its installation, the optimum choice for the new disk purchase will often be newer generation, larger disk. In this case, there is a choice—create a separate Disk Group or add the new disks to an existing Disk Group. If the growth plans require a separate Disk Group for availability requirements, then this would take priority over a decision made to optimize performance. If there is no requirement for a separate Disk Group, then the decision will be based on the storage administrator's knowledge of the workload.

Creating new Disk Group will most likely result in an imbalance in the access density with the existing Disk Groups. If there is a specific plan to match the demand with the capability of each Disk Group, then an additional Disk Group can be a good choice. But, if the storage administrator is uncertain of the workload demand, then adding the disks to an existing Disk Group will be the best choice.

In a manner similar to disk speeds, disks with different capacities would seem to be better placed in separate Disk Groups. As with the preceding advice, more disks usually produce better performance, so placing disks of differing capacities in a single Disk Group results in overall higher performance for most applications.

For example, an existing HP StorageWorks Enterprise Virtual Array (EVA) is configured with a single Disk Group of 48 36-GB, 15K-rpm disks. The goal is to increase the capacity and performance of the application by adding

16 72-GB, 15K-rpm disks. For this example, both the 36-GB and 72-GB 15K-rpm disks are capable of about 125 I/Os per second.

- The existing Disk Group is 1.7 TB and rated at 6,000 IOPS, which equals 3.5 IOPS per GB.
- Option 1—Add the new disks to the same group. The result is one Disk Group: 2.8 TB, 8,000 IOPS, or 2.8 IOPS per GB.
- Option 2—Create an additional new Disk Group with 16 72-GB, 15K-rpm disks. The result is 1.1 TB, 2,000 IOPS, and 1.7 IOPS per GB.

Table 3.

| | Disk | | | Disk group | | |
|---------------------|------------|------|-----------------|----------------|-------------|---------|
| | # of disks | IOPS | Capacity | Total capacity | Total IOP/s | IOPS/GB |
| Existing Disk Group | 48 | 125 | 36 GB | 1.7 TB | 6,000 IOPS | 3.5 |
| Option 1 | 64 | 125 | 36 GB and 72 GB | 2.8 TB | 8,000 | 2.8 |
| Option 2 | 16 | 125 | 72 GB | 1.1 TB | 2,000 | 1.7 |

For option 2, the new Disk Group has half the capacity and a quarter of the performance of the original Disk Group. Unless there is a specific plan for data that will be accessed at half the rate of the original Disk Group, applications assigned to the new Disk Group run the risk of poor latency (response time) caused by queuing delays on the disks.

For option 1, the data is leveled to all the disks in the Disk Group based on the capacity of each disk. The result is that the bigger disks contain more data. But because the workload is distributed to all the disks, it is far less likely that any single Vdisk will suffer poor performance, and it is far more likely that the disks will have a higher total utilization than option 2—thus, better array performance and better application performance.

This is a key example of how the EVA provides superior real-world performance. Traditional disk arrays and volume management software do not redistribute the data to newly inserted disks. Thus, as traditional arrays grow capacity, they create “hot disks” that require additional management to eliminate.

- ✓ **Best practice for performance optimization: Purchase and create Disk Groups using equal capacity disks.**
- ✓ **Best practice to optimize performance when disks of unequal capacity are present: Unless there is a specific plan to manage the differences in access density of the Disk Groups, it is preferable to combine disks with different capacities in the same Disk Group.**

Read cache management influence on performance

One of the parameters that can be set, at either Vdisk creation or dynamically, is read caching. 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 are designed to 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 only comes into play if a sequential read stream is detected. Because of this dynamic response to changing I/O workload conditions, cache efficiency is maintained at a high level, and there is no negative impact on either cache usage or performance in the presence of an I/O workload that 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 leave read caching enabled.**

Controller balancing influence on array performance

Although both controllers in an EVA can access all physical disks, a Vdisk is assigned to be managed by only one controller at a time. Consequently, a single Vdisk can only use the resources of a single controller, such as cache, data path bandwidth, Fibre Channel ports, or processor cycles. Under heavy workload demand, it is possible that only one of the controllers in an EVA pair will reach one of its resource limits if the demand is not equally distributed across both controllers. Distributing the workload evenly across both controllers ensures the maximum total performance of the array.

There are two methods to distribute the access; assignment and striping. In assignment, the application or system administrator will assign different files or file sets to different Vdisks through different controllers. For striping, the

administrator will use operating system- or application-based software to stripe the data evenly to the controllers. Striping provides the best performance distribution of the workload.

Here is an example of striping across controllers. Two Vdisks are created within a single Disk Group. Each Vdisk is assigned access through a different controller. Host-based software (like a volume manager) stripes the two Vdisks to form a single logical disk presented to the file system or application. The host-based striping ensures both controllers are accessed equally.

The configuration of load balancing across controllers is unique for each operating system and possibly unique to the version of volume management or path management software within a single operating system. Although it is beyond the scope of this paper to describe the unique details to accomplish controller level balancing, it is recommended that controller load balancing be addressed to optimize for performance.

- ✓ **Best practice to optimize performance: Attempt to balance the workload demand evenly across both controllers on an EVA.**

Note: Striping LUNs within a Disk Group on the same controller provides no additional performance value. The EVA will automatically stripe each LUN across all the disks in a Disk Group.

LUN count influences performance

LUN count has no effect on native EVA performance—the maximum EVA performance can be demonstrated with a single LUN per controller. However, because of the default operating system or host bus adapter (HBA) queue management in some systems, an application (such as a database) can achieve better performance with multiple LUNs, or they might require special queue depth tuning to achieve maximum performance with a small number of LUNs.

Overall array performance optimization with Snapclones can be achieved by minimizing the number of Vdisks that are issued a Snapclone request. Thus, in this case, the recommendation would be to create the minimum number of Vdisks and modify the host operating system and/or HBA queue depth setting to provide sufficient queue with the small number of LUNs.

Particular attention should be made for HP-UX and queue depth management. The default queue depth is eight I/Os per LUN. This is insufficient for typical configurations that use only a few LUNs. There are numerous other publications within the HP-UX community that address this issue.

In Microsoft Windows® environments, attention to the Windows version and the HBA is required to understand the default I/O queue management operation.

The performance value of this best practice should not be overlooked. This is a common configuration error that can result in dramatic reduction in performance.

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

Snapshot and Snapclone performance management

The EVA has the capability of making local copies of selected Vdisks. There are two types of copies: a full duplication and a space efficient snapshot copy. Both types of copies 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 two components to the execution of internal Vdisk copies: the metadata management and the data movement. Metadata management is the work the controller needs to perform to create the internal data structures to manage the new Vdisks. Data movement is the operation of copying the data. Metadata management is similar for both Vsnap and snapshot and Snapclone operations; however, data movement differs.

A Snapclone, as the name implies, makes a complete copy of an existing Vdisk. Snapclone creation places an additional workload on the disks of the target Disk Groups. 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 the Disk Group can maintain. This performance impact continues until the Vdisk has been completely copied. When the cloning operation completes, the performance impact ceases.

Vsnaps and snapshots take advantage of the HP StorageWorks EVA virtualization technology and only copy changes between the two Vdisks. This typically results in a reduction in 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 only copied during the write operation. A block is copied only once. When it diverges from the original Vdisk, it is not copied again. For typical applications, less than 10% of the data in a Vdisk typically changes, so when this 10% diverges, the performance impact ceases.

Another characteristic of typical workloads is that the performance impact has an exponential decay as the Vdisks diverge. In other words, the performance impact is greater on a newly created snapshot than on an aged set of snapshot Vdisks.

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, disks).

- ✓ **Best practice for snapshot and Snapclone performance: Create and delete snapshots and Snapclones during low demand periods, or scale performance sizing for operation during snap activities.**

- ✓ **Best practice for Snapclone performance:**
 - **Keep Vdisks as small as possible.**

 - **Minimize the concurrent Snapclone operations (use fewer Vdisks). Organize clone operations into consistency groups of Vdisks, and clone groups sequentially.**

- ✓ **Best practice for Vsnapshots and snapshot performance:**
 - **Minimize the number of Vdisks with active snapshot copies. Use fewer Vdisks (it is better to have few large Vdisks than more smaller Vdisks).**

 - **Minimize the number of snapshot copies for a Vdisk. Do not keep extra snapshot copies without reason or plan for their usage.**

 - **Minimize the life of a snapshot copy. If snapshot copies are used for backup, consider deleting the snapshot Vdisk at the completion of the copy to tape.**

 - **Delete snap Vdisks in order of age, oldest first.**

Capacity-free snapshots use free space (capacity not reserved for normal or Snapclone Vdisks) to store the changed data. All capacity-free snapshot Vdisks will become inoperative when any capacity-free snapshot Vdisk is denied a request to use additional free space, thus it is wise to always monitor free space usage. If capacity-free snapshot Vdisk availability is critical for the whole application availability, then over-estimating the requirements for free space might 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. It is possible that after a reconstruction, the reserved space requirements for the protection level will cause the existing snapshot Vdisks to exceed the available free space, thus causing the snapshot Vdisks to go inoperative.

- ✓ **Best practice to assure snapshot availability: Use the following sequence to compute free space:**
 1. **At initial configuration, create primary Vdisks but no snap Vdisks.**
 2. **Set the Protection Level to two.**
 3. **Use the remaining available capacity, as computed by the snapshot target Vraid level, for remaining free space for snapshot value.**
 4. **Return the Protection Level to one.**

Windows data alignment and performance

It is possible, in a Windows system, that the array transfer requests are not aligned to 2K boundaries. This is caused by a characteristic of the Master Boot Record (MBR). As a result, when programs transfer data to or from disks that have more than 63 sectors per track, misalignment can occur at the track level, with allocations beginning at a sector other than the starting sector. This misalignment can defeat system optimizations of I/O operations designed to avoid crossing track boundaries.

- ✓ **Best practice to optimize performance in Windows environments: Use the Microsoft DiskPar utility to ensure data block alignment.**

Best practices to manage capacity

Increasing capacity of the array

To minimize false indications of excessive errors, insertion of multiple disks should be done carefully and slowly, with a pause between inserting disks. This process 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 ensures Leveling does not start until all the new disks are ready.

Although the array will support the process of replacing existing smaller disks with new larger disks, this process is both time consuming and performance disruptive and can result in a non-optimum configuration. This process should be used only if the option to build new Disk Groups and move the existing data to the new disks is unavailable.

- ✓ **Best practice to optimize availability when adding disks to an array:**
 - **Set add disk setting to Manual.**
 - **Add disks one at a time, following the “wait 60 seconds” rule between disks.**
 - **Distribute disks vertically and evenly as possible to all the shelves.**
 - **Unless otherwise indicated, add new disks to existing Disk Groups. Use the HP StorageWorks Command View add multiple disks command (see command line interface [CLI] commands).**
 - **Add disks in groups of eight.**
 - **For growing existing applications, if the operating system supports Vdisk growth, increase Vdisk size. Otherwise, use a software volume manager to add new Vdisks to applications.**
- ✓ **Best practice to add capacity: Avoid replacing existing disks with larger capacity disks. Either retain the old disks in the system or create a new Disk Group and migrate the existing data.**

Capacity management and optimized operation

Free space, available capacity that is not allocated for use by a Vdisk, is used by the EVA controller for multiple purposes. Although the array is designed to operate at full capacity allocation, functions like Vsnap, reconstruction, leveling, and disk management work more efficiently with additional free capacity.

This is particularly true when Vsnap Business Copies are in use. Improper estimation of free capacity can result in application errors that may result in a loss of service—either the application or the backup function that is using the Vsnap. For this reason, this best practice is unique for Vsnap usage.

Vsnap space efficient local copies use a “copy-on-write” technology. Copy-on-write occurs only when either the original virtual disk or the Vsnap virtual disk is modified (a write), and an image copy of the associated blocks is duplicated. For any given block, the data is only copied once.

The actual physical capacity required for a Vsnap is dependent on the divergence of the original virtual disk and the Vsnap, which is unique for each application. The suggestion for the initial usage of Vsnap (initial meaning you are unfamiliar with the actual physical capacity required by the Vsnap) is to reserve (do not allocate to Vdisks) a capacity equal to 10% of the capacity of the parent Vdisks times the number of Vsnap per parent Vdisk. For example, if you need to create two Vsnap of a 500-GB Vdisk, you need to ensure 100 GB (500 GB*10%*2) of usable capacity is available. The usable capacity should be computed using the RAID level selected for the Vsnap.

- ✓ **Best practice for initial capacity management when Vsnap are present: Keep an unallocated usable capacity equal to 10% to 15% of the total usable capacity of the Vdisks that will use Vsnap.**
- ✓ **Best practice for capacity management: Always keep a minimum of 5 GB of Vraid 0 capacity unallocated per Disk Group, and always use a Protection Level greater than zero.**

For existing arrays that do not meet this best practice, the addition of disks into each Disk Group without the associated LUN creation automatically increases the free space or allows increased Protection Levels.

Previous EVA best practices have indicated that the array should be kept below 90% to 95% capacity utilization. Current studies have concluded that the 5 GB and a Protection Level of 1 is sufficient to optimize leveling.

Disks groups and data security

Disk groups are wholly 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. And, lastly, a LUN can contain capacity from only one Disk Group.

Given these characteristics of Disk Groups, applications or data centers that require data isolation for security objectives can accomplish these objectives through the usage of multiple Disk Groups. In this case, each unique security domain would be assigned to a different Disk Group.

These characteristics of a Disk Group can also be used for asset tracking and allocation of assets to a specific group 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.**

Best practice folklore, urban legends, and myths

Pre-allocation

There are two phases to LUN creation on an HP StorageWorks Enterprise Virtual Array (EVA): the first phase creates the metadata data structures; the second phase writes zeros to the associated sectors on the disks. Access to the newly created LUN is restricted during the first phase by Command View, but access is allowed during the second phase. During the second phase, host I/Os compete with the zeroing operation for the disks and controller, thus performance is impacted during this period.

During a test case on a 168, 36-GB disk EVA, creating 64 46-GB LUNs, a sequential write performance benchmark was started at the completion of the LUN creations. It was observed that 30 minutes after the last LUN was created, the array was capable of maximum performance.

In this test case, there was no other load on the EVA. Zeroing is a background operation, the time required to complete zeroing would be extended if other workloads were present on the array. The Vdisk is fully accessible during this time; however, the performance of the Vdisk during zeroing would not be representative of future performance.

For non-snapshot Vdisks, the EVA always maps the 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 both these characteristics—the new LUN zeroing operation and the predictable data layout—there is no reason to pre-allocate data by writing zeros to the Vdisk on the EVA.

- ✓ **Urban legend for pre-allocation: There is no benefit to pre-fill a Vdisk on the EVA.**

Note: a performance benchmark would show inconsistent results during the zeroing period. A pre-fill of the Vdisk will ensure performance benchmarks are representative of long term performance.

Use a 12-shelf system to maximize availability

There is a good basis for this “best practice.” The largest an RSS can be is 11 disks. This ensures that for any number of disks in the array, the controller can attempt to optimize the placement of the RSS to shelves. A better best practice is to follow the eight or more shelves and disks and disks in Disk Groups in multiple of eight.

Disk groups in multiple of eight disks if only eight shelves

This “best practice” attempts to optimize the availability by matching Disk Groups and shelves. This is a correct assumption, but the “best” best practice is always to have disks in groups of eight, independent of the number of shelves.

Disk groups of 12 or less disks if only six shelves

This “best practice” attempts to increase availability by the optimization of RSS to disk shelves. However, many, small Disk Groups have poor capacity utilization and possibly the possibility poor performance utilization. If the objective is higher availability than the default RSS management, then a better option would be to use as few Disk

Groups as possible and Vraid 1. Vraid 1 will provide higher availability with only a modest decrease in capacity utilization for arrays with only six shelves.

Number of disks in a Disk Group based on the shelf count

This rule does not ensure an optimum layout, plus the multiple Disk Groups will decrease the capacity utilization of the array and possibly make performance management more difficult. Disk groups should be in groups of eight disks, as large as possible and as few as possible.

Summary

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

Table 4. Summary

| | Cost | Availability | Performance |
|---------------------------------------|---------|---------------|--------------------|
| Mixed disk capacities in a Disk Group | No | No | Acceptable |
| Number of Disk Groups ⁴ | 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 |
| Disable write cache mirroring | - | No | Maybe |
| Mixed disk speeds in a Disk Group | - | - | Acceptable |
| Redundancy level | 0 | 1 or 2 | 1 |
| LUN count ⁵ | - | - | - |
| Read cache | - | - | Enabled |
| LUN balancing | - | - | Yes |

⁴ Consult application or operating system best practices for minimum number of Disk Groups.

⁵ Check operating system requirements for any special queue depth management requirements.

Glossary

| | |
|----------------------------------|---|
| access density | A unit of performance measurement. I/O 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 in a Disk Group 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 out of 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 was to provide technical details to optimize the HP StorageWorks Enterprise Virtual Array (EVA) for specific applications. It was not intended to be a general purpose tutorial on EVA operation. HP and its partners provide additional information and services to help train and optimize your EVA. For more information on the HP StorageWorks Enterprise Virtual Array, go to <http://www.hp.com/go/eva> or contact your local HP sales representative.

© Copyright 2004 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.

5982-9140EN 10/2004

