HP StorageWorks XP Disk Array Configuration Guide for mySAP Business Suite

First Edition

Abstract

The goal of this paper is to provide an overview of configuring HP StorageWorks XP disk arrays correctly in an SAP environment to achieve maximum I/O performance and simplified SAP management. In today's business environment, customers need the most efficient and reliable applications available. With the new HP StorageWorks Disk Array XP12000 and SAP, customers find a winning combination to ensure business operations for a competitive advantage.

First, the paper reviews how the traditional disk layout for SAP environments works well in scenarios with many smaller disks attached directly to individual database servers but does not work well in multi-terabyte SAP landscapes. The paper then presents a recommended approach for SAP database distribution on a high-end storage system for greater efficiency.

Next, general recommendations are given on implementing SAP distribution on an XP array—the first step is choosing the appropriate RAID level for striping capabilities, and the next step is choosing a suitable emulation type to define the logical devices (LDEVs) per array group.

This paper summarizes how XP array configurations in SAP environments decrease I/O bottlenecks and enable easy expansion during times of growth without any downtime. The result is an expanded storage system providing optimal load balancing, thereby increasing the efficiency of your data center.

Executive summary

Good distribution of SAP database components on the HP StorageWorks XP disk arrays ensures sufficient disk storage space and reliable data security with the highest level of I/O performance, providing the opportunity to grow the database with moderate system management efforts.

Although the XP disk array is highly adaptive and has great flexibility, a reasonable storage configuration must be chosen to satisfy the highest workloads for an SAP landscape.

This paper summarizes actual best practices for XP configurations in SAP environments.

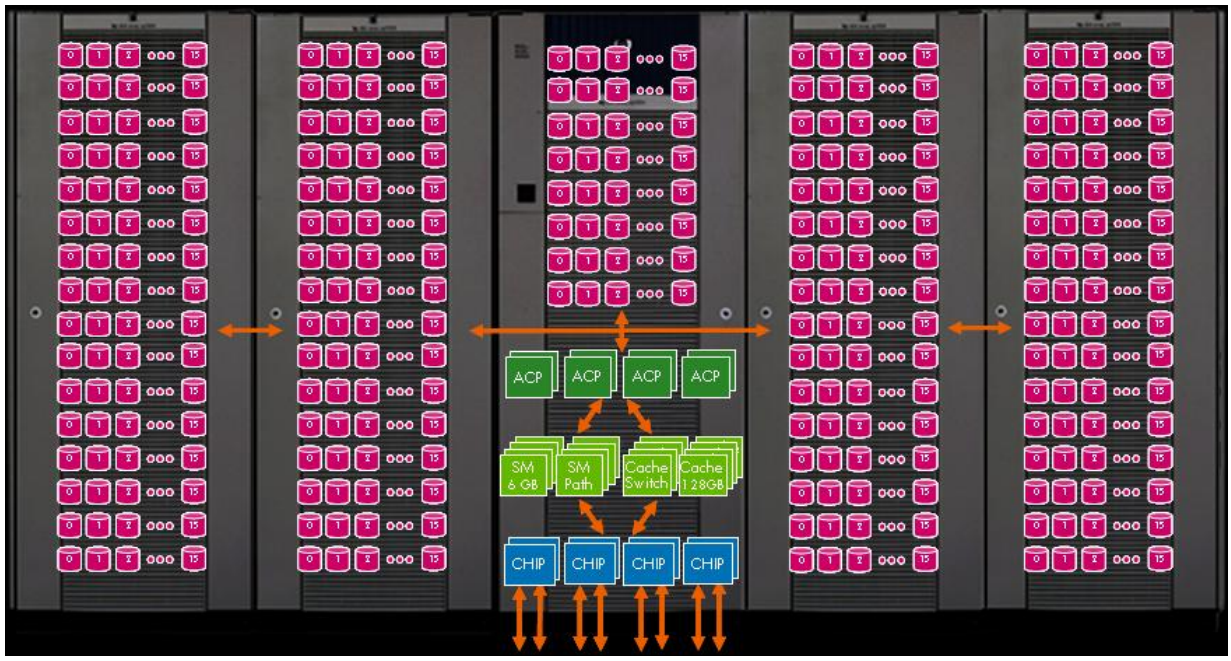
Overview

The XP disk arrays are positioned at the high end of the HP enterprise storage portfolio. In an SAP environment, this means that over 1,000 disk drives and more than 120 GB of cache can be made available within a single array to satisfy the largest SAP customer requirements.

The XP disk array family

Various factors determine why a particular disk array technology is chosen for a specific SAP solution. Performance, availability, and reliability are the main criteria, together with usable capacity and configuration flexibility, that lead to a decision to use an XP storage array to satisfy the requirements of a sophisticated SAP landscape.

Figure 1. XP12000 architecture overview and maximum configuration



XP array configuration

For information on how to configure an XP disk array, the HP Network Storage Solutions (NSS) Sizer utility is the preferred choice. To be sure that your XP configuration will reach some specific performance goals, consult the HP XP Performance Estimator Tool before working with the NSS Sizer.

The XP array in an HP SAN

The HP StorageWorks SAN Design Reference Guide defines the rules and guidelines surrounding the design of storage area network (SAN) infrastructures for XP arrays in the HP SAN with specific platform and operating system rules.

Different storage layouts for an SAP database

Traditional storage configuration for an SAP R/3 system on Oracle

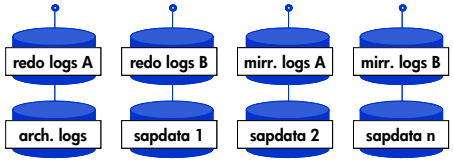
The traditional storage disk layout in SAP environments is based on the following recommendations extracted from the official SAP installation guides.

Table 1.

Placing data according to access method	The data must be placed according to its predominant access method. This means that if data is mainly to be written or read, it should be separated onto other disks.
Placing different data types	Different data objects, such as redo log files, archive redo log files, rollback segments, and temporary storage files, must use separate physical disks. These disks should preferably be mirrored for data protection.
Placing data and index files	Another rule is to separate data and indexes from each other to allow the database to access its data and index in parallel.

These recommendations have worked well for scenarios with many smaller disks directly attached to an individual database server. An example of this traditional layout is shown in Figure 2.

Figure 2. Traditional SAP storage layout



This conventional approach involves the following drawbacks when building a large-scale storage system for a multi-terabyte SAP landscape:

- The initial distribution of database objects for access load balancing requires considerable manual effort.
- The most heavily used disks cause I/O bottlenecks that impair the performance of the entire storage system. These hotspots require permanent monitoring and recurrent corrective action, such as relocating database objects to disks with lower utilization, and the administration of free disk space distributed across multiple physical devices ties up scarce IT resources.
- Database load distribution tends to be unpredictable. For instance, nightly batch runs generate heavy sequential I/O traffic, while user interaction during the day implies intense random access to disks. This results in vast unplanned load fluctuations—different disks are involved at different times. Because of the lack of predictable access patterns, it is difficult for administrators to identify the "right" storage optimization methods.

Recommended approach for SAP database distribution on high-end storage systems

The most common approach to efficiently managing rapidly growing disk capacities while keeping storage performance at maximum levels throughout the life cycle of the database storage solution is based on a set of recommendations in an SAP ATG group white paper titled "Database Layout for R/3 Installations under ORACLE—Terabyte Project."

Based on the recommendations and on benchmarking experience gained in large SAP installations of HP customers, a first version of this white paper was written to leverage the SAP approach for HP products. Subsequently, the approach presented in this white paper is supported by Oracle® in a general white paper titled "Optimal Storage Configuration Made Easy" and in an HP XP storage subsystem-related white paper titled "SAME and the HP XP512."

Maximize spindle usage through striping for optimal load balancing

All database objects, the tablespaces, log files, and temp files are distributed across as many physical disks as possible to use the performance of the whole storage system for a single application access.

Figure 3. SAME approach

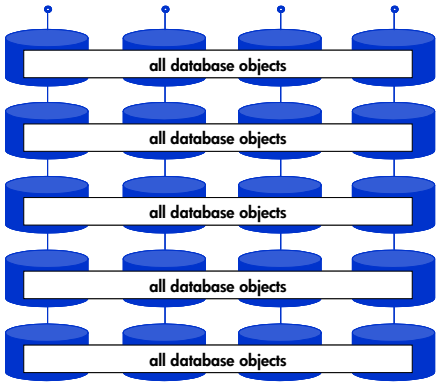


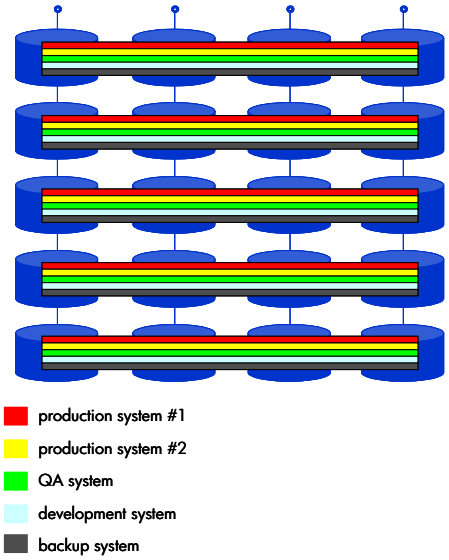
Figure 3 illustrates that all database objects of an SAP instance are evenly distributed across all available disks and controllers, with large logical volumes striped across multiple disks.

SAP landscape storage consolidation

A typical SAP landscape consists of development, quality assurance, and production systems. In a full-fledged mySAP.com environment, the number of production systems (workplace, Business Warehouse, application plus objects [APO], and so on) increases in pace with growing business requirements.

Striving for greater efficiency, large SAP customers tend to consolidate their IT infrastructures—particularly their storage facilities. The idea is to centralize data storage on a smaller number of high-performance, high-availability systems. Consequently, multiple R/3 systems can share the same storage system. This way, customers can apply common administration and high-availability processes to all their systems and achieve greater flexibility by dynamically assigning storage capacity according to business needs.

Figure 4. Storage consolidation in an SAP landscape



To prevent hotspots following hardware upgrades (after the new data is stored on newly added disks), the disk array should be equipped with the reasonable number of disks during initial installation.

If one SAP system outgrows the available disk space, one of the systems can be moved to another storage system to provide optimally distributed disk free space for data growth on the production system.

Load peaks in one system can influence the performance of other systems accessing the same storage subsystem. Therefore, it is a good idea to use the HP StorageWorks Application Policy Manager (APM). The APM enables administrators to control I/O throughput to and from each host bus adapter (HBA) directly on the XP disk array level.

General recommendations

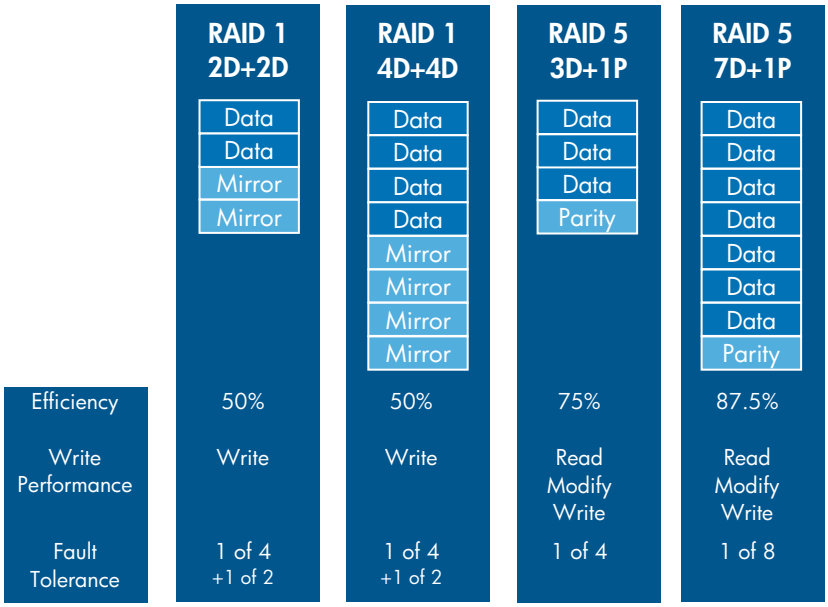
To implement the model of distributing all SAP database objects on the maximum of available spindles, the I/O distribution capabilities of both the array and the operating system level must be considered.

SAP data distribution on the XP array level

Hardware striping based on RAID levels

The first step in distributing SAP data on an XP storage array is to choose the appropriate RAID level for assigned array groups, which are the four-disk drive entities that provide striping capabilities within an XP array. Figure 5 references the possibilities of the various RAID levels on the XP architecture, outlining the pros and cons.

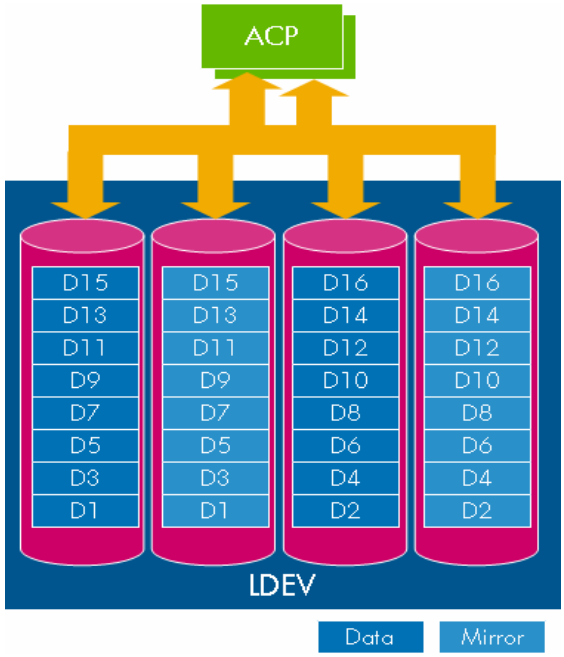
Figure 5. XP RAID levels



The striping across two disks with mirrors using XP RAID level 1 is a combination of RAID 1 (mirroring) and RAID 0 (striping). The array is first set up as a group of mirrored pairs (RAID 1) and then striped (RAID 0). A RAID level 1 array configured in this manner can sustain multiple drive failures, maintaining optimal read and write performance with the cost of additional disk drives. There is the possibility of RAID 1 4D+4D, which is a concatenation of two RAID 1 2D+2D groups (eight disks) on a single or dual Array Controller Processor (ACP) pair, forming a larger maximum LDEV, which is the host visible entity when it has been mapped to a front-end CHIP port as a logical unit (LUN).

On an XP array for which the additional I/O processor utilization associated with RAID 5 write operations does not lead to CPU bottlenecks on the ACP pairs (that is, on an XP array with enough high performance ACP pairs installed to support the random write load generated by the SAP system), RAID 5 3D+1P or RAID 5 6D+2P (as shown in Figure 7)—depending on the operating system striping capabilities—is a good choice because it offers the same I/O response times to the SAP database as RAID1, provided the same number of physical disks are being used. If one of the disk drives in a RAID 5 6D+2P group fails, it is still under protection because there is a second parity disk and the group is not degraded as under the RAID 5 7D+1P. RAID 5 is an excellent combination of performance, redundancy, and storage efficiency for SAP configurations with moderate write and update requirements.

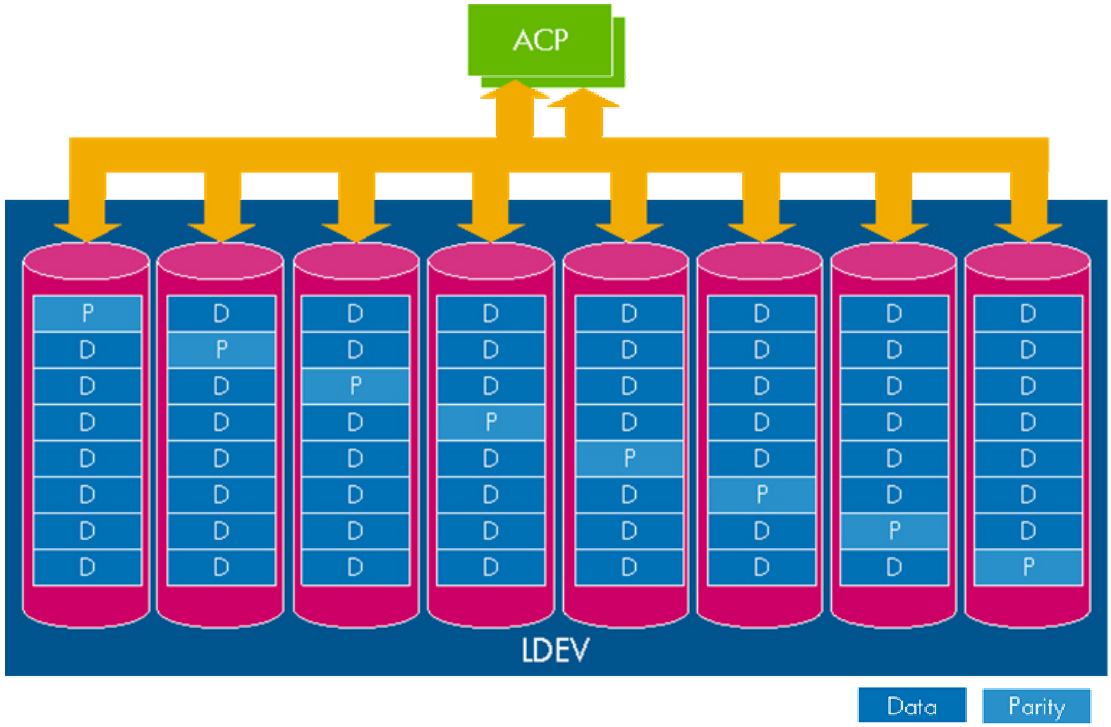
Figure 6. Distributed parity across four disks



In this context, the I/O pattern of an R/3 system is predominantly (over 85%) random read and less than15% random write. Because all writes go the XP cache, the choice of RAID level has no effect on write response times as long as the write cache utilization does not reach 100% (which should never happen under normal SAP operating conditions because of the relatively low random write load generated by an R/3 system) or unless the ACP utilization significantly exceeds 50% as a consequence of the parity calculation associated with RAID 5 writes. However, when a bottleneck on the I/O processors within the ACP pairs starts to develop, the random read response times are affected, which, in turn, impacts the overall performance of the SAP system.

The HP white paper, "Comparative Analysis of Database Storage Layouts for Performance-Critical SAP Systems," addresses the performance aspect of the different RAID levels for SAP configurations.

Figure 7. Striping with distributed parity across eight disks

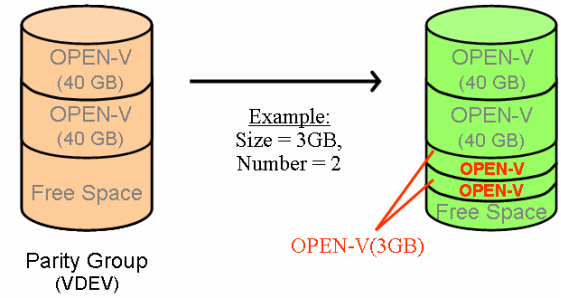


Across the disks of an array group in an XP disk array, striping functionality is implemented on the hardware level. To achieve a higher degree of data distribution on the XP disk array, individual LDEVs carved out of the four disks can be combined to form larger volumes using a capability called Logical Unit Size Expansion (LUSE). This LUSE constitutes a LUN to the host and can be distributed on the LDEV level, across multiple RAID groups. However, as the individual LDEVs are concatenated rather than striped to form a LUSE, such a dispersed LUSE does not ensure the same level of load distribution as a volume manager on the operating system level with load distribution across multiple RAID groups. Therefore, if a volume manager on the operating system level can be used for this purpose, HP discourages the use of a dispersed LUSE because it causes the configuration to become more complex without additional performance benefit.

Disk type emulation on XP for SAP

The next step in configuring XP storage for an SAP environment is to choose a suitable emulation type to define the LDEVs per array group in terms of size and number. The license-enabled OPEN-V emulation type gives the greatest flexibility in specifying volumes in mixed sizes from 48 MB to 2 TB for a single parity group, as seen in the example of Figure 8. The size and number of the LDEVs to be generated for a specific SAP configuration depends on the actual disk size and the number of available parity groups for an SAP database instance.

Figure 8. OPEN-V volume addition example



The goal is to have a reasonable number of LUNs available on the operating system level maintaining the balance with the fewest possible LUNs to minimize the administrative overhead and distribute the I/O load effectively over all components in the I/O path from the Fibre Channel Adapter (FCA) down to the disks.

Distributing SAP I/O load across all XP components

Spreading the SAP database across sufficient RAID groups over different ACPs with the correct number of LUNs ensures that the SAP application gets the available resources out of the back-end of an XP array. To avoid a bottleneck in the front-end of the XP array, sufficient CHIP ports must be mapped to the LUNs holding the data of an SAP instance. To implement multipathing functionality for availability and performance reasons, each LUN must have at least two different physical paths to an FCA in a host.

Performance considerations

The first step in having sufficient I/O capabilities at any time for an SAP environment is to understand the I/O workload for planning purposes. But there is no single I/O profile in an SAP landscape with different SAP applications, SAP system types (development, test, and production), and load types (batch and dialog) at different times. Furthermore, at rapidly growing sites, the I/O characteristics can change over time. This scenario requires a flexible SAP storage layout from the beginning, in combination with performance monitoring capabilities.

A good source in supporting the performance-related part of a sizing process is the XP Performance Estimator Tool, which is based on real measured data.

A good source in understanding XP performance measurement using HP StorageWorks Performance Advisor XP is the white paper, "XP Performance Including Performance Advisor."

Implementing an SAP storage layout on HP-UX with Oracle

The way to implement an SAP storage layout on an XP array using Oracle does not differ between HP-UX on the PA-RISC or the Intel® Itanium® processor family architecture. The HP StorageWorks Disk Array XP Operating System Configuration Guide describes the requirements and procedures for connecting the XP family of disk arrays to an HP-UX system and configuring the new disk array for operation with HP-UX.

SAP data distribution using the Logical Volume Manager

Logical volume managers (LVMs) usually provide I/O distribution on the operating system level. The LVM on HP-UX supports block level striping and extent distribution. The advantages of using large volumes created at the operating system level is the possibility of achieving optimal distribution of the I/O load across all devices, host bus controllers, and I/O channels within the storage system. For large LUSE and OPEN-V LUNs, HP recommends increasing the SCSI queue depth (from the default value of 8) using the `scsictl` command on the operating system level. This usage of LVM provides a high degree of flexibility in influencing the granularity (size of the stripes) and the degree (number of disks per stripeset) of data distribution.

LVM block level striping

Stripe sizes can range anywhere from 4 KB to 32 MB. In an SAP environment, HP recommends a stripe size of 2 to 4 MB. One drawback of this method is that it cannot be combined with LVM mirroring, so it is not possible to use the `pvmove` command for moving physical volumes from one device to another because `pvmove` is based on LVM mirroring. But this functionality is available in various flavors on the XP array level. For mirroring, HP StorageWorks Business Copy (BC) and HP StorageWorks Continuous Access (CA) can be considered. The migration of LDEVs from a source array group to target array group is delivered with HP StorageWorks Auto LUN XP functionality.

LVM extent distribution

This method uses a larger granularity of block sizes (called extent)—usually 4, 8, or 16 MB. The upper size limit of the logical volumes depends on the chosen size of extents. The advantage of extent distribution over block level striping is that the LVM can be used for mirroring. The `pvmove` command can also be used with this method. A welcome side effect of using a larger granularity is that during concurrent sequential access, it dramatically reduces disk head movement and thus helps to protect the random I/O performance—for example, during times of multiple concurrent table scans, index range scans, or both and times of online backup with multiple parallel data streams.

Database expansion in combination with enhanced XP features

Fast growing SAP configurations are faced with the challenge of data redistribution to avoid hotspots when the initial configuration runs out of space and new storage hardware should be added. Therefore, the database expansion methodologies must be analyzed in terms of I/O load of the latest data and of the disks where the latest data is stored.

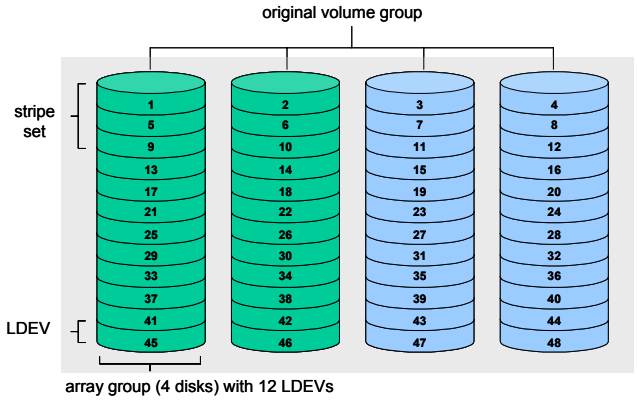
The effect of adding new disks in the course of database expansion is that new data only is stored on the new disks. Consequently, this creates hotspots on the new disks, thus impairing load balancing. To prevent or at least defer this costly process, HP recommends adding a sufficient number of disks and adequate disk space at the initial installation of the system for the foreseeable future.

In the following example, the SAP database is expanded without downtime for the application. An HP-UX server has the database server role. The XP disk array is equipped with 16 disks, organized in four array groups. There are 48 LDEVs. The logical volume is striped (block level striping with 64 KB) on 12 physical volumes (`lvcreate -i 12 -I 64`).

Initial configuration

The installation shown in Figure 9 provides optimal load balancing because the blocks are striped across all four array groups and across all disks within these array groups.

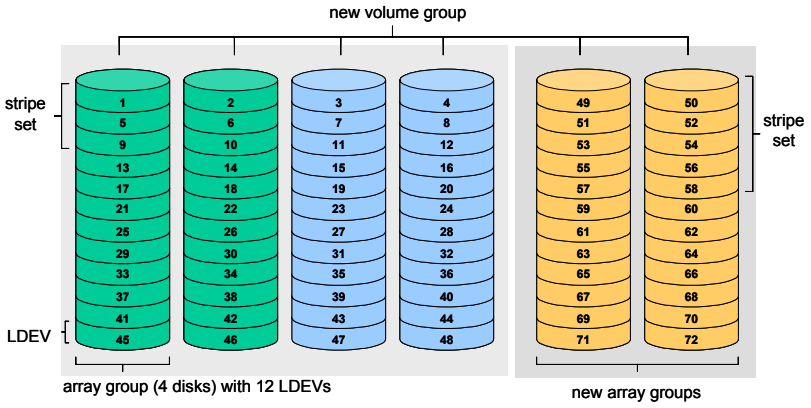
Figure 9. Initial configuration



Adding disks to keep pace with database growth

As the database grows, the storage system must be expanded by the addition of two further array groups. HP recommends adding these disks uniformly across all of the available controllers. Another critical aspect is the number of devices across which the data is striped; in the example, striping spans 12 devices. The block level striping method used implies that the total number of LDEVs must be a multiple of 12. (If extent distribution was used, the number of LDEVs does not matter.)

Figure 10. Adding physical disks

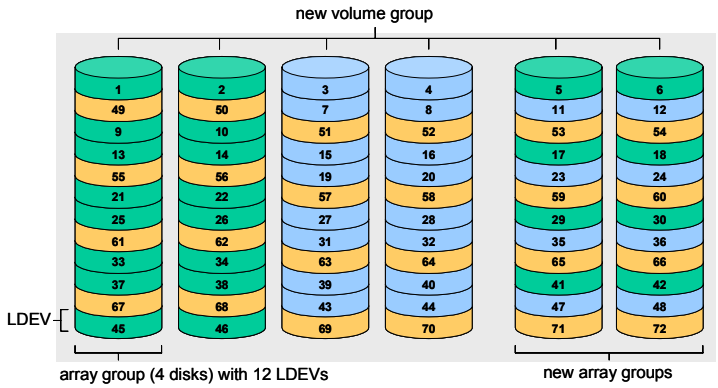


Expanding the logical volumes by adding disks can be done without downtime. However, the new data will reside on the new disks only, as shown in Figure 10. Note that the first new stripeset is located on LDEV 49 to 60, which implies that only two array groups are affected. Because the most recent data is accessed most frequently and will eventually reside on the newly added array groups only, I/O contention on these array groups is likely to occur.

Even distribution of disks for optimal load balancing

To correctly expand the XP disk array, mix the old and the new LDEVs such that an even and balanced distribution is achieved. This means that, after the new disks have been installed, some of the old LDEVs must be moved with Auto LUN XP to the new array group, as shown in Figure 11. The result is an expanded storage system, providing optimal load balancing. The drawback when using this mechanism to rebalance the database after adding array groups is that the CU:LDEV numbering is disordered, which makes the configuration more complex to understand.

Figure 11. LDEV redistribution for optimal performance



Defining the number of disks per stripe set

The reason for having one stripeset across 12 physical volumes (rather than four, which would be sufficient for a complete initial distribution in this example) is because a single stripeset spans all array groups in the initial system configuration. Because the degree of distribution of an existing striped logical volume cannot be changed, this degree should be wide enough to span the whole array in case of later storage expansions. According to this missing flexibility of block level striping, the general advice is to use extent distribution.

With four array groups, a stripeset on four physical volumes would be a sound decision. However, subsequent system expansion by two additional array groups would not allow the striped blocks to be distributed in an optimal way. Another possibility would be to stripe across six physical disks so as to have an optimal situation when the system is expanded with two additional groups. The drawback of this approach is that in the initial installation with four array groups, two stripes would be placed on two of the groups. A stripeset across 12 physical volumes is the optimal situation with four, six, and 12 array groups because in every case, the stripeset is evenly distributed across all controllers.

In the initial installation, allowance must be made for subsequent expansion of the storage system, for example, by calculating the best number of stripes within one stripeset. Whenever possible, the number of stripes per stripeset should equal the maximum number of LUNs the storage system can provide.

Using raw devices

Avoiding I/O contention on individual disks or portions of a large storage system by balancing the data across the entire storage system hardware is the most effective way of achieving good I/O performance. On HP-UX, the use of raw devices for the database is another method that can lead to performance advantages. These performance advantages can be achieved by:

- Bypassing the file system structure.
- Removing one layer of buffering inside the operating system (buffer cache). However, the very same effect can also be achieved by choosing the appropriate Journal File System (JFS) mount options ("`convosync=direct`" for JFS 3.3 and "`mincache=direct,convosync=direct`" for JFS 3.1).
- Increasing the System Global Area (SGA) by spending the RAM in the server saved on buffer cache. However, because there is no reason to increase the buffer cache of a database server in the first place, no matter whether the database files are placed on raw devices or in the file system, there is also no RAM to be gained by choosing raw devices.
- Using asynchronous I/O. The system call will return immediately when the I/O request has been passed down to the hardware or queued in the operating system. The process is not disrupted because it does not need to wait for the results of the system call. Instead, it can continue executing (usually just by issuing the next couple of I/O requests) and then receive the results of the I/O operation later, when they are available. However, besides the Oracle log writer (which can write origlogs and mirrorlogs in parallel when allowed to do asynchronous I/O), in normal operation, only the database writer can benefit from asynchronous I/O, and as has been mentioned before, the database writer is doing its work in the background. Therefore, as long as the total throughput of the database writer (or of multiple database writers or database writer I/O slaves simulating asynchronous I/O in a somewhat less efficient way) is sufficient to prevent "checkpoint not complete" waits, no performance improvement should be expected from enabling asynchronous I/O. This is especially true for applications that are so heavily read-dominated, as is the case for SAP R/3. For example, in the case of a table access by way of an index, the reading process must always wait for each I/O to complete before it can request the next index block because it is the result of each read that determines the next block to be read. Thus, asynchronous I/O can do nothing to speed such index accesses, which make up the bulk of all I/O requests in a typical SAP R/3 system.

Some issues must be kept in mind when raw devices are used. Creating physical raw disk devices, which are needed for the several database objects, can require considerable effort, particularly if storage space must be provided in defined sizes. Therefore, deployment of LVM and the use of raw logical volumes are recommended. The advantage of this method is that it achieves greater flexibility in the size of volumes. Also, the striping functionality can be used. In an HP Serviceguard environment, the use of LVM is imperative.

For situations in which log file synchronization causes performance problems, consider placing the online redo logs on raw devices. However, when there is no use of the Oracle mirrorlogs because the logs are mirrored on array level and the file system for online logs has been formatted with 1-KB block size and is avoiding the use of the file system buffer cache, there will be little benefit to using raw devices. In these situations, the usage of HP StorageWorks Cache LUN XP can be considered, but because of the price:performance relationship, the implementation of cache LUNs has not been seen often.

In summary, weighting the administrative overhead, especially for fast growing SAP configurations and the eventual performance gain, the usage of raw devices is generally not recommended.

HP-UX configuration example

Complementing the conceptual considerations from the previous sections, this section proposes a practical approach to designing a storage layout for SAP with HP-UX on an XP storage array. This sample provides a general guideline for designing a practicable layout that strikes a balance between availability, ease of administration, and performance.

The following configuration is valid for any HP 9000 server (PA-RISC) with HP-UX 11.11 as well as the HP Integrity server family with HP-UX 11.23 or higher and all supported XP configurations.

Figure 12. Sample storage layout for SAP with HP-UX on an XP disk array

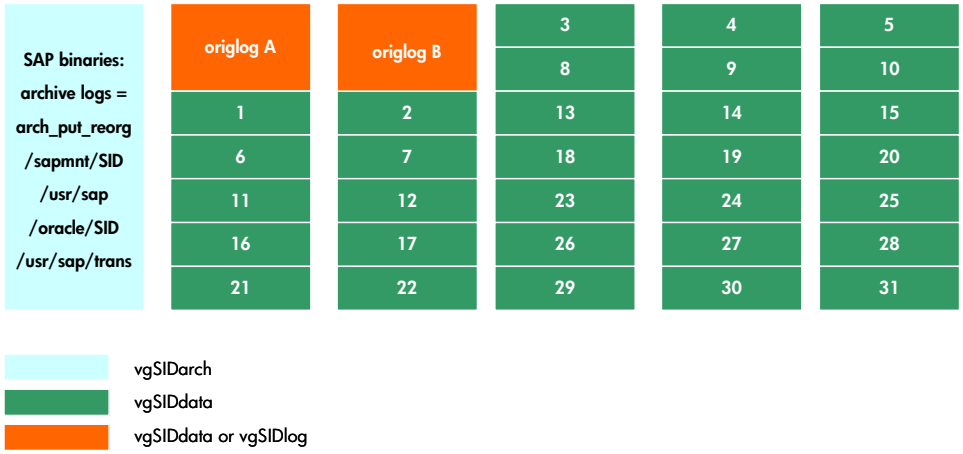


Figure 12 shows a configuration with the following characteristics:

- Each of the six columns represents a separate XP RAID group comprising four or eight disks, depending on the chosen RAID level. This gives six times the number of disks in an array group, which is normally four for RAID 3+1. For RAID 6+2, there are eight disks behind every column. The number of rows (in this example, 7, 6, and 1) is the number of generated LDEVs per array group, depending on the emulation type that has been chosen.
- Each different color represents a different LVM group.
- Oracle's origlogA and origlogB are separate logical volumes residing on dedicated CVS- or OPEN-V LUNs.
- The green area LDEVs, numbered from 1 to 31, symbolize the "sapdata" logical volume, with the numbers referring to the LVM extents of the logical volume.
- Mount options for sapdata and redo log files use JFS to bypass the buffer cache and enforcing direct I/O (`mincache=direct`, `convosync=direct`).

This configuration can be expanded as described in the section, "Database expansion in combination with enhanced XP features" on page 11.

For more information

HP links

HP Network Storage Solutions
http://www.hp.com/go/storage

HP NSS Sizer
http://h30144.www3.hp.com/

Managing MC/ServiceGuard Extension for SAP R/3
http://docs.hp.com/hpux/pdf/B7885-90013.pdf

HP StorageWorks SAN Design Reference Guide
http://h18000.www1.hp.com/products/storageworks/san/documentation.html

Comparative Analysis of Database Storage Layouts for Performance Critical SAP Systems
http://whitepapers.silicon.com/0,3800002488,39000478q-3,00.htm

XP Storage tools (HP internal)
http://storagetools.lvld.hp.com/xp

Oracle links

Optimal Storage Configuration Made Easy
http://otn.oracle.com/deploy/performance/pdf/opt_storage_conf.pdf

SAME and the HP XP512
http://otn.oracle.com/deploy/availability/pdf/SAME_HP_WP_112002.pdf

SAP links

Database Layout for R/3 Installations under ORACLE—Terabyte Project
https://websmp202.sap-ag.de/~form/sapnet?_FRAME=CONTAINER&_OBJECT=011000358700005625622000E

5982-7763EN, 08/2004

invent