FOR SYSTEM ADMINISTRATORS AND SAS USERS

SAS Storage Tips

A white paper series prepared by IBM ^ pSeries Solutions Enablement

Configuration Options using IBM SSA Storage for SAS

This document discusses approaches to configuring SAS storage on IBM Serial Storage Architecture (SSA) devices such as the IBM 7133 disk subsystem. It details some of the tradeoffs in performance, availability and cost for both permanent and temporary storage configurations.

SAS storage requirements

SAS applications typically require both permanent and temporary storage. SAS input and output data files can be considered permanent because these files persist for some amount of time before and after your SAS program is executed. SAS workspace, on the other hand, is temporary storage because it is used to hold files created and accessed only during program execution.

Although each application of SAS is unique, most programs input data from SAS data files, process it (i.e. edit, join, reduce, and sort), perform some kind of computationally intensive analyses, then write results to output files. The characteristics of the permanent storage used for storing input and output data and the temporary storage used for storing partially processed information can be different, exhibiting tradeoffs in performance, availability, and cost.

Performance versus availability versus cost

You must make the determination of the relative importance of SAS performance, data availability, and cost of storage. For instance, if your SAS input data can be retrieved conveniently from another computer system at any time, then high availability of input data may not be an important consideration on your SAS computer system.

Likewise, it may be easier to justify re-running your SAS analyses to regenerate lost data than to justify investing in redundant storage for output. In either case, you may wish to optimize for performance and save money on hardware you would need to assure higher availability. The RAID 0 configuration discussed below may be the right choice for you.

On the other hand, your source data may be the result of a very time-consuming monthly extract from a heavily used data warehouse; and it may be stored only on your SAS system. Or, your SAS analyses may take an extremely long time to run, and the results may be immediately critical to business operations. In these cases, you may want to sacrifice some performance to ensure that your data files are less subject to corruption or loss due to hardware failures. In this case, consider the high-availability RAID 1, RAID 5, or RAID 10 configurations.

Things to consider
Unless you have an overwhelming argument to isolate your SAS data at the expense of all else, do not dedicate a single disk drive to SAS activities. You will be assuring your users the worst possible performance.

Stripe your data (both permanent and temporary) across multiple disks. If you can't stripe, try to place SAS input and output storage on one disk, with SAS work (temporary) storage on another. If you can stripe only some of your storage, stripe your SAS work storage.

Striping is not the same as concatenation. When data is stored on concatenated disks, one disk is filled before any data is written to the next disk. In general, concatenation means only one drive is active at any time, no matter how many drives you have allocated. Striping breaks your data up into smaller pieces, writing a piece to each allocated disk, and filling each disk at the same rate. As a rule, this assures that more than one drive will be in operation at the same time. This parallelism increases the speed at which data can be read and written. It also tends to balance disk utilization and reduces the need for manual tuning.

IBM's Serial Storage Architecture and Advanced SerialRAID Plus adapter
SSA is a high performance technology used in systems such as the IBM 7133 disk subsystem to connect disk devices and host adapters in serial loops. IBM's Advanced SerialRAID Plus adapter is capable of the largest variety of SSA configuration types: Non-RAID, RAID 0, RAID 1, RAID 5 and RAID 10. Each adapter contains two initiators, each supporting its own loop of up to 48 disks. An optional 128 MByte on-board cache is also available to improve write performance to those disks.

RAID 0: maximum performance
If you chose to maximize performance above all other considerations, then you should consider a RAID 0 configuration for both permanent and temporary SAS storage. A RAID 0 configuration (otherwise known as striped disks) can be built by striping across as few as two or as many as sixteen member disks on the same SSA loop.

For most SAS applications, both permanent and temporary storage are used predominantly for transferring large amounts of data sequentially. Here, the performance benefit of RAID 0 comes from the data being striped across multiple disks. All of the disks involved can retrieve or store their data in parallel.

In some cases, temporary storage involves random accesses with short data transfers. In this situation, performance is improved by striping because spreading data across disks also spreads seek activity across all the disks.

The disadvantage of RAID 0 is its lack of any availability protection in the event of a disk failure. No redundancy exists, so when a disk fails, all the data in the entire striped array is lost, not just that of the failed disk. In addition, RAID 0 is supported in a single adapter configuration only on IBM SSA storage. The adapter bandwidth can limit the

number of disks that are kept busy. If this happens, you could achieve better performance with a different configuration that spans multiple adapters and loops.

RAID 1: high availability, but costly

RAID 1, or simple mirroring, is a great way to provide high availability if you can afford it. This configuration simply stores multiple copies of your data on two or more disk drives. If one drive fails, the data is immediately available from one of the copies, or mirrors. Data reads can be faster, since the AIX operating system can retrieve data from any mirror disk drive. Writes that are not completely cached by the disk subsystem can be slightly slower, however, because the data must be written to all target drives before the write is complete.

RAID 1 is relatively expensive. It requires at least twice the number of disk drives.

RAID 10: faster than RAID 1

RAID 10 is a configuration that stripes data across multiple drives and mirrors the striped array as well. You get the performance of striping and the high availability of mirroring. As with RAID 1, you may see a slight degradation in write performance over RAID 0 because of the additional number of drives being written to. If cost is your least worry, RAID 10 is the best solution, offering both speed and availability.

RAID 5: a good compromise

RAID 5 is a good compromise between high performance, high availability, and cost. Whereas mirroring stores multiple copies of your data in its entirety, RAID 5 stores only one complete set of data, plus a small amount of additional data (or "parity" information) to provide redundancy protection. Rather than reverting to another copy of the data when a disk fails, RAID 5 assures that lost data can be reconstructed in the event of hardware failure. By storing only the additional parity information, RAID 5 uses significantly fewer disks than RAID 1 at a significantly lower cost.

RAID 5 performance may be slower compared to other configurations if there is a high proportion of write activity. For RAID 5, each write operation normally requires two read operations (to read the old data and old parity) and two write operations (to write the new data and new parity) to member disks. On IBM SSA storage such as the 7133 disk subsystem, the performance hit isn't as severe as on other systems, because the SSA adapter – not the host computer – performs this additional write-related activity. Still, the difference may be apparent.

You can mix and match

Since SAS applications use both permanent and temporary storage, you can chose to mix and match your storage configurations. For instance, you could use RAID 1, RAID 10, or RAID 5 for permanent storage; then use RAID 0 for your temporary storage. Reads and writes to SAS input and output data would be slower than if you were to use RAID 0; but assurances of availability of that data would be higher. At the same time, you would be maximizing performance during program execution when temporary storage is used. If a disk were to fail in your temporary storage, you could simply restart your SAS

program after replacing the drive. An even safer option would be to use RAID 5 for permanent storage, and RAID 10 for temporary storage.

Fast write cache

The fast write cache on your SSA adapter allows your computer system to complete a write operation when data is written to the cache, but before it is actually written to the disk drive. You should enable this option. It will increase performance tremendously; and the probability of hardware failure is very low. Nevertheless, you should know that it is possible to disable this feature. When data integrity is your utmost concern, disabling fast write cache will ensure that no write is considered complete until the data is physically on the disk.

Other configuration options

Since the Advanced SerialRAID Plus adapter supports non-RAID disks, you can create AIX Logical Volume Manager (LVM) configurations using individual disks on an SSA loop as well. Just be aware that write operations to LVM mirrored or striped logical volumes can require more host processor resources than those required for all types of RAID implemented at the adapter level.

If your SAS applications tend to be CPU-bound, you should not consider this option. However, if your applications are I/O bound, then you could actually improve performance by sharing the I/O load between the host CPUs and the SSA adapters. For example, you may choose to implement RAID 5 configurations at the SSA adapter level. Then you could stripe – at the AIX LVM level - across hdisks originating from different SSA loops. This hybrid configuration would give you a degree of high-availability and performance not otherwise possible, especially if your SSA adapters reside on separate PCI busses.

For more information, refer to the following IBM redbooks at http://www.redbooks.ibm.com**:**

Understanding SSA Subsystems in Your Environment,
SG24-5750-00,
A Practical Guide to Serial Storage Architecture for AIX,
SG24-4599-00

Other resources:

IBM Storage web site
http://www.storage.ibm.com