

Design criteria for optimized high availability of IBM® RS/6000® SP® systems with SSA-Disks

Summary:	This article points out the possible Single points of Failure of an IBM® RS/6000® SP® and recommends concepts and means in order to ensure and optimize high availability of SP® clusters and the SSA disks
Author:	Richard Handforth, UK richard@dedgug.freemove.co.uk
Edition:	1.0
Date:	1999/09/30
Reference:	GE9002 (www.mannherz.de)

The contents of these pages must not be altered, disseminated in public or duplicated for commercial purposes without the prior written consent of the publisher. This also applies to the use of the contents of these pages on other Web sites or networked computers.

These pages and the information they contain have been compiled to the best of our ability and using the best knowledge available. However, no legal obligation can be implied therefrom. Liability for the correctness, completeness and reliability of information, data and documents shall be restricted to cases of gross negligence or malicious conduct for all losses or damage (including consequential losses) incurred through a user's trust in the information obtained during the use of the service. Liability shall also be restricted to cases of gross negligence or malicious conduct for all losses or damage (including consequential losses) incurred through use of these pages in any other way (for example, through downloading Web pages, or similar).

Insofar as we are under contractual obligation to certain customers to provide certain contents on our Web pages, liability shall be restricted to cases of gross negligence or malicious conduct for all losses or damage (including consequential losses), whereby this does not apply to negligent violation of duties essential to performance of the relevant contracts.

The author is responsible for the contents of the document.

Contents

- 1. Design Considerations for Disk and Data Availability.....3
 - 1.1 Availability of SSA Disks.....3
 - 1.2 Quorum and Mirroring.....3
 - 1.2.1 2-disk Volume Group.....4
 - 1.2.2 4-disk Volume Group.....5
 - 1.2.3 6-disk Volume Group.....5
 - 1.2.4 8-disk Volume Group.....5
 - 1.2.5 Quorum and Mirroring Disk Combinations.....6
 - 1.2.6 Mirroring with Odd Numbers of Disks.....6
 - 1.2.7 Location of Mirrored Copies.....6
 - 1.2.8 Extending Logical Volumes Across Disks.....7
 - 1.3 Disk Reintegration.....7
 - 1.3.1 Replacing Failed Disks.....7
 - 1.3.2 Resynchronising Stale Disk Partitions.....7
- 2. Single Points of Failure in SP® Hardware.....8
 - 2.1 Control Work Station.....8
 - 2.2 SP® Frame.....8
 - 2.3 SSA Disk Racks.....9
 - 2.4 SSA Disk Subsystems.....9
 - 2.4.1 Power Units.....9
 - 2.4.2 Subsystem Signal/Bypass Cards.....10
 - 2.5 SSA Disk Loops.....10
- 3. Single Points of Failure in SP® Networks.....11
 - 3.1 Internal Network.....11
 - 3.2 Client Networks.....11
 - 3.3 Switch network.....12
- 4. Trademarks.....13

1. Design Considerations for Disk and Data Availability

1.1 Availability of SSA Disks

The location of mirrored copies of logical volumes on disks, and whether or not the quorum should be turned on or off, are important design considerations in any HACMP environment. Poor planning and bad design can mean that disk failures become a Single Point of Failure (SPoF), resulting in application failures, data corruption, and in some cases system crashes. When a system crashes then normally a node takeover will occur, but if data has already been corrupted then the application may still be unusable even after a takeover. Also, if an application fails, then this may not necessarily be detected by HACMP and there will be no takeover.

All SSA disk configurations must have mirrored logical volume copies to ensure the highest level of availability, although this may not be possible if there are cost constraints. If mirroring is not used, then should disks fail, or disk subsystems become inaccessible for whatever reason, it will not be possible to replace failed disks or reintegrate stale disks into their respective volume groups without considerable downtime. This may be acceptable in some environments, but generally speaking this is not desirable since it defeats the objective of trying to attain high availability for an application.

The location of disks and mirrored copies in and across subsystems, and their connections within SSA loops, must be carefully considered in order to achieve the highest possible availability of data within a cluster, and reduce downtime to a minimum when the reintegration of failed components is required.

SSA disk configurations must be such that no matter which node has control of cluster resources, high availability of disks and data can only be achieved if one of the following criteria is valid when a loop is broken:

- A cluster node can access all disks in the resource group it currently controls. This occurs when a loop has been broken, but all disks are still accessible by using alternative routes around the loop; or
- A cluster node can access a complete set of mirrored disks within the resource group. In this case either single disks may have failed, or multiple disks have become inaccessible because of adapter, subsystem, or node failure.

In clusters containing nodes with single SSA adapters a node takeover may be required before these design criteria can be achieved. SSA loop failures will occur in an SSA disk configuration if:

- Any node fails.
- Any SSA adapter fails.
- Any SSA cable fails.
- Any disk subsystem fails.
- Any subsystem disk fails.
- Any signal/bypass card fails.

The resulting disk access (or lack of it) will be determined by what has failed, and where the failure occurs; the configuration must be designed to minimise the effects of any of these failures.

1.2 Quorum and Mirroring

When a volume group is varied on and quorum is turned on, for the volume group to remain accessible, AIX

requires that more than 50% of the Volume Group Descriptor Areas (VGDA) contained on the disks are available; the VGDA contains information about the volume group, and also time stamps which are compared between disks to determine whether logical volumes have become stale.

A single disk volume group contains 2 VGDA on the disk. In a 2 disk volume group, one disk contains 2 VGDA and the second disk contains a single VGDA. Any volume group containing three or more disks has single VGDA on each disk.

In a 2 disk volume group, if you lose the disk containing 2 VGDA, then you will lose quorum and the volume group will vary off and data will no longer be accessible, even though mirroring may be in place. In both three and four disk volume groups, you can only lose a single disk for the volume group to remain active since you must at all times have more than 50% of the VGDA available.

It is possible to turn quorum off for a volume group, and this should only be done when mirroring is in place. This effectively means that you could lose all disks except one and the volume group would still remain active, although the data may not be usable depending on the particular application, and under these circumstances there is also the possibility of data corruption.

In AIX®, mirroring is at the logical volume level with either one or two mirrored copies possible. All copies should be located on separate disks, although this is not essential. It is possible to place all the copies on a single disk but this defeats the object of mirroring. Having three copies obviously provides greater protection, but offset against this is the increase in cost.

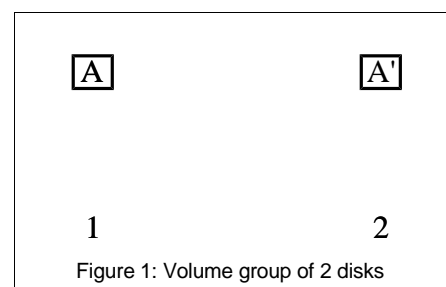
Ideally in an HACMP environment, mirroring should always be present and quorum should **always** be on in order to protect data integrity. Unfortunately, in volume groups which contain just two disks (one disk mirrored) then this could lead to a potential reduction in availability if quorum remained on and the disk containing the 2 VGDA were lost.

Although the two disk volume group is a special case, in all other cases the balance has to be drawn between the need for availability and the potential for data corruption should quorum be switched off. This is especially true when an even number of disks are spread evenly across only 2 disk subsystems where the loss of one subsystem will automatically mean the loss of quorum.

If quorum is on, disks can be arranged across SSA subsystems in the following manner to ensure that a volume group remains active in the event of a subsystem failure. It should be noted that the mirroring will be at logical volume level and not at disk level, although in the following examples it is assumed that all logical volumes on each disk are mirrored to the same location on the corresponding mirrored disk.

1.2.1 2-disk Volume Group

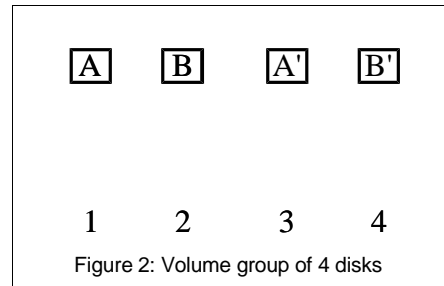
It is not possible to arrange two disks across two subsystems to ensure that the volume group will always remain active if a subsystem fails while quorum is turned on. It is recommended that quorum be turned off since this will allow the volume group to continue being accessed no matter which disk fails. Data corruption is unlikely to occur since if a further disk is lost then all disks will be missing and no data can be written or read. In the above situation it is essential that the system clocks between all cluster nodes are accurate so that should there have been an HACMP takeover prior to the subsystem rejoining, and the missing subsystem again becomes available either during or after the takeover, then re-synchronisation of the logical volumes will be performed from the accurate up-to-date copy to the stale copy, and not the other way around.



The alternative to turning quorum off is to add a third „quorum-buster“ disk to the volume group and locate it in a third disk subsystem. Although this will involve added expense, it would then be possible to spread the logical volumes across the three disks as shown in the example below using an odd number of disks.

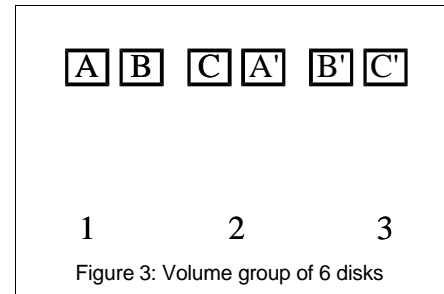
1.2.2 4-disk Volume Group

A four disk volume group must be spread across four subsystems to ensure that the volume group remains active no matter which subsystem fails. If 4 subsystems are not available, then a quorum-buster disk should be added and the disks could then be spread across the subsystems in a 2-2-1 configuration.



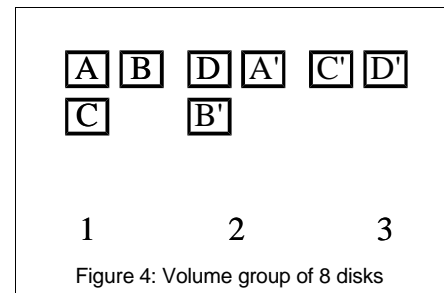
1.2.3 6-disk Volume Group

A six disk volume group must be spread across three subsystems to maintain quorum in the event of a subsystem failure.



1.2.4 8-disk Volume Group

An eight disk volume group must be spread across three subsystems to maintain quorum in the event of a subsystem failure.



1.2.5 Quorum and Mirroring Disk Combinations

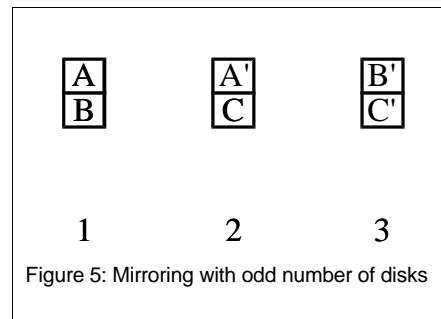
The following table shows the ideal locations of disks and their corresponding mirrors, so that, apart from the 2 disk situation, quorum can remain on:

Disks total	Disks in each Subsystem				Quorum
	1	2	3	4	
2	A	A'			off
4	A	B	A'	B'	on
6	A, B	C, A'	B', C'		on
8	A, B, C	D, A', B'	C', D'		on
10	A, B, C, D	E, A', B'	C', D', E'		on
12	A, B, C, D	E, F, A', B'	C', D', E', F'		on
14	A, B, C, D, E	F, G, A', B', C'	D', E', F', G'		on
16	A, B, C, D, E, F	G, H, A', B', C'	D', E', F', G', H'		on
18	A, B, C, D, E, F	G, H, I, A', B', C'	D', E', F', G', H', I'		on
20	A, B, C, D, E, F, G	H, I, J, A', B', C', D'	D', E', F', G', H', I', J'		on

Table 1: Ideal disk locations

1.2.6 Mirroring with Odd Numbers of Disks

The above examples have all assumed an even number of disks where it is possible to mirror all the logical volumes on a disk to the exact same locations on the corresponding mirrored disk. It is also possible to provide mirroring, with quorum turned on, across an odd number of disks, but in such a situation careful thought must be given to the placement of the logical volumes on each disk. Figure 5 represents three logical volumes mirrored across disks in separate subsystems in such a way as to maintain quorum in the event of a subsystem loss, and also to allow access to any of the logical volumes no matter which subsystem fails. This design principle can be extended to a volume group containing any odd number of disks.



Having an odd number of disks (apart from 1) has the advantage that only three subsystems need ever be used, but the added disadvantage is that much greater thought, planning and administrative overhead is required in the placement of the logical volumes.

1.2.7 Location of Mirrored Copies

To ensure highest possible availability of data, mirrored copies must be located in separate subsystems, and each subsystem containing mirrored copies of logical volumes must be located in different disk racks. Ideally three subsystems should be used in the manner described above, so that in the event of failure of the power supply to a disk rack then quorum will still be maintained and data will still be accessible through the mirrored copies.

1.2.8 Extending Logical Volumes Across Disks

In a volume group containing an even number of mirrored disks, logical volumes can extend across multiple disks since if a single disk containing part of the logical volume is lost, or a complete disk subsystem fails, then a mirrored copy will always be accessible.

In volume groups containing odd numbers of disks, extending logical volumes across disks can become much more complex, particularly when you want to use more than half the disks in the volume group since the mirroring must now be considered at the physical partition level and the use of partition map files is required. In such cases a carefully planned distribution of partitions will be required to ensure that the logical volume is totally accessible if a disk or subsystem fails. This should only be considered as a last resort since there could be performance implications in addition to the administrative complexity.

With five or more disks you can easily extend across two (or more) disks without resorting to map files provided that they are within the same subsystem.

1.3 Disk Reintegration

Mirroring will help to preserve the integrity and high availability of data, but should disks fail, or logical volumes become stale, a considerable amount of effort may be required to ensure that the system is restored to its original condition.

1.3.1 Replacing Failed Disks

SSA disks are hot-pluggable which means that they can be replaced while a system continues to operate without having to shut down the system. Mirrored copies of logical volumes will still have to be rebuilt on the new disk followed by synchronisation of the data.

Although it is extremely unlikely that multiple disks will fail at the same time, should this occur then their replacement could be time consuming and resynchronisation of the data onto the new disks may affect existing users since this very I/O intensive operation is likely to increase user response times.

1.3.2 Resynchronising Stale Disk Partitions

Whenever access to a disk is lost without the disk actually failing, for example when the power to a disk subsystem fails, then it is most likely that logical volumes will become stale; the number of stale partitions will depend on the amount of I/O write activity that continues while the disk is inaccessible. When access to disks again becomes available, then these stale logical volumes must be resynchronised.

There are two ways to resynchronise the data on disks, and the method to be used is dependent on how important it is to keep the system up and running.

The first method, which is usually the quickest, but means that users will temporarily lose access to their application, is to shut down the application, vary off its volume group, and then vary it back on again. During vary on the data will automatically be resynchronised, and the length of time to resynchronise the data will depend on the number of disks involved, and how stale the logical volumes are. If power to a whole frame has been lost, then the resynchronisation could take some considerable time. After resynchronisation the application can be restarted.

The second method can be considerably time consuming and involves removing all affected disks from the volume group, adding them in again, and rebuilding the logical volume structure and data on each disk. If map files are kept regarding the location of each logical volume on each disk then the recreation of the logical volumes can be speeded up. The rsynchronisation of data can again be time consuming, but the only effect on users is likely to be increased response times, although rebuilding of the data could be done when user access is at a minimum.

2. Single Points of Failure in SP® Hardware

2.1 Control Work Station

The control workstation itself is a Single point of Failure, but is not a critical component as far as HACMP clusters are concerned. Should it fail, the operation of individual nodes will be unaffected, but the following system management restrictions will be noticed:

- It will not be possible to control the SP® hardware.
- The System Data Repository will be unavailable.
- No configuration changes will be possible.
- Software installations will not be possible from the control workstation.
- Should a switch fault occur, reset processing will not be completed.
- Error logging of alerts raised by nodes will be lost, although the information will still be logged on individual nodes.
- Administrative tasks using PSSP will not be possible.

HACWS can be used to circumvent these potential problems automatically without user intervention.

As an alternative, a standby system could be used, which could be installed with a `mksysb` image of the failed control workstation and then cabled into the SP® to allow administration functions and SP® monitoring tasks to be performed. In such a configuration, this data should be stored on an external volume group:

- SP® management data
- AIX® system images
- PSSP and related software install filesets (`/spdata` filesystem)
- NIM configuration files, and any other software install filesets

This would ensure up to date configuration information would be accessible by the standby machine, which a `mksysb` image from the previous day may not provide.

2.2 SP® Frame

Power is supplied to a frame's three power units (SEPBU) via a single three-phase power cable. This is a Single Point of Failure and if the cable fails or is disconnected, nodes will lose their power supply and will be inaccessible. To ensure high availability, HACMP clusters should be configured across frames, so that if the power supply cable to a single frame fails, takeovers will be initiated for all nodes in clusters contained in the frame.

Although the frame supervisor and the RS232 cables from the frame supervisor to the nodes and the control workstation are Single Points of Failure, a failure of any of these components will mean system management restrictions only, either for a single node or multiple nodes depending on the failure. Client access and the normal operation of cluster nodes will be unaffected.

Since each frame is shipped with redundant power supply units, these are not Single Points of Failure.

2.3 SSA Disk Racks

Each external disk rack has a single power cable and a single internal power distribution unit. Both of these components for the rack itself are Single Points of Failure. If the logical volumes on the disks in each cluster were mirrored across racks, which have their own independent power supplies, then for the cluster as a whole they would not be Single Points of Failure.

Should a power distribution unit fail, or the rack power cable fail or be disconnected, then power will be lost to all the subsystems in the rack and all the logical volumes on every disk would become stale. Reintegration of all these disks into their respective clusters and the synchronisation of data on all the logical volumes would affect user access.

It is recommended that a second power distribution unit be installed, cabled from a separate power supply, so that the failure of either a single power cable or internal power distribution unit will have no immediate effect on cluster operations.

Should one of the dual power distribution units fail, then messages would be placed in the error logs of all nodes connected to SSA disks in the rack, provided that the SSA subsystem power units themselves were split between the dual power distribution units as discussed below.

2.4 SSA Disk Subsystems

2.4.1 Power Units

Each SSA disk subsystem can be supplied with three power units. One provides power for the front 8 disks, a second provides power for the rear 8 disks and the optional third is a standby in the event that one of the other two fails. The individual power units themselves will not be a Single Point of Failure provided that all three units are installed.

The power cable supplying power to these units, however, is a Single Point of Failure for the SSA subsystem since it is connected to all three power units. If we consider a cluster as a whole, then the cable may not be a Single Point of Failure provided that mirrored disks in another subsystem in a separate disk rack are still accessible.

If the power cable fails, or the rack power distribution unit fails, then all disks in the subsystem will become stale and the subsequent reintegration into the cluster and synchronisation of the data when the power again becomes available may affect user access.

Since it has been recommended that each disk rack should have a second power distribution unit, then it is further recommended that the current single cable to the three subsystem power units be replaced with two cables, each coming from a separate power distribution unit, with one cable connected to two subsystem power units, and the second to a single power unit. Figure 6 suggests a possible power cabling solution. A and B are rack power distribution units. F is the subsystem power unit for the front 8 disks, R is the subsystem power unit for the rear 8 disks, and S is the standby power unit. If A fails then only the rear 8 disks in each subsystem will still be accessible. If B fails, then the standby will take over the power supply to the rear 8 disks and all disks in the subsystem will be accessible. The above diagram shows just two subsystems in a rack, but the principle can be extended to all subsystems in the rack.

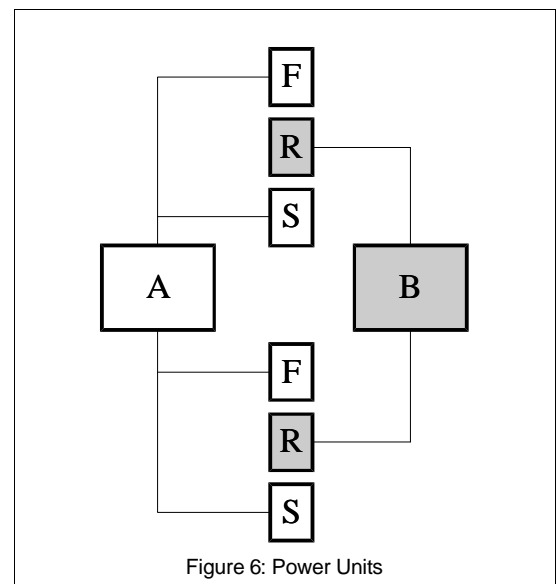


Figure 6: Power Units

2.4.2 Subsystem Signal/Bypass Cards

Cabling from an adapter is attached to „J“ connectors (ports) on the subsystem, which in turn connect to the internal SSA disks. The „J“ connectors are attached in pairs to signal or bypass cards, which are not themselves Single Points of Failure provided that disk access can still be achieved through an alternative loop path using another bypass card, or the mirrored copies are accessible through a second loop.

The bypass cards in 7133-600 SSA subsystems contain „J“ connectors which are numbered to indicate the first disk in the bank of four to which they are connected. The pairing combinations are 1 and 16, 4 and 5, 8 and 9, 12 and 13. The bypass cards can also operate in two modes.

In bypass mode, if a node which is connected to both J connectors on the bypass card fails, then the connectors on the card close the loop and effectively join together the two banks of 4 disks for which they provide access. These two banks may contain dummy disks to fill out the disk drive slots and allow the loop to remain unbroken, but should more than three dummy disks be connected consecutively, then the loop will break and disks will become inaccessible; this is a distinct possibility with bypass mode.

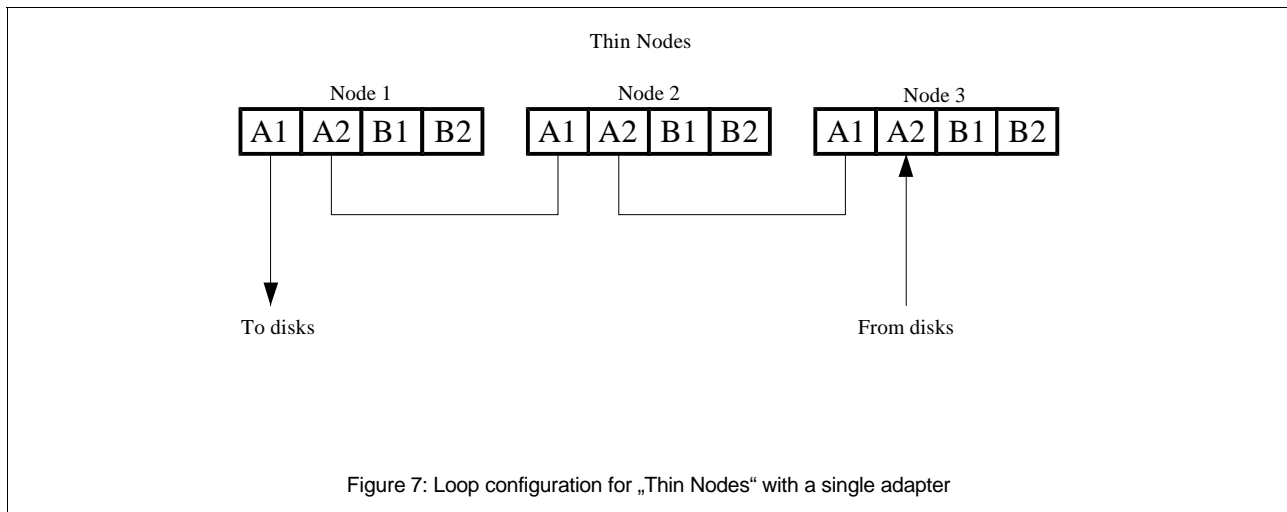
In forced inline mode, if a node connected to the „J“ connectors on a single bypass card fails, then the loop will be broken and remain broken, but the disks will still be accessible along the alternative loop path. Since multiple nodes are connected to subsystems in each cluster, then all the bypass cards on each subsystem should be set to forced inline mode.

2.5 SSA Disk Loops

All external disks must be connected to all nodes which could possibly require access in the event of a takeover and must be cabled such that when a disk fails, or an SSA cable fails and the loop is broken, then the loop does not become a Single Point of Failure and any active disks in the loop can still be accessed via an alternative loop path, or the mirrored copies are accessible through a second loop.

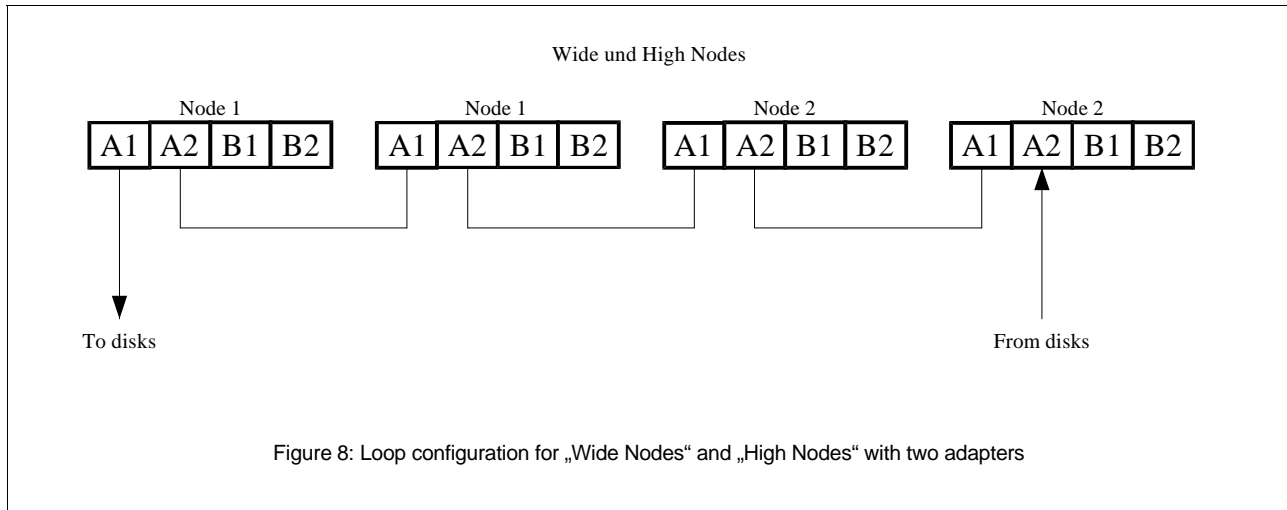
In thin nodes it may not be possible to have more than one SSA adapter so that if the adapter fails then no data will be accessible from the node and a takeover will be initiated. Only two SSA loops are possible connecting all nodes to three subsystems, and both these must be configured to provide the best availability and also to take advantage of the full band width of the SSA adapters.

A loop configuration similar to the following should be used for thin nodes with single adapters. Both the A and B loops will be similar, each loop connected to different mirrored disks. Both 2 and 3 node clusters will be similarly cabled.



Wide and high nodes will most likely have two SSA adapters and the loops must be cabled such that when an SSA adapter fails, or when a disk fails, then all the disks in the loop (apart from a failed disk) will still be accessible via the alternative loop path.

A loop configuration similar to the following should be used for wide and high nodes with 2 SSA adapters. Both the A and B loops will be similar, although connected to different disks. Both 2 and 3 node clusters will be similarly cabled:



3. Single Points of Failure in SP® Networks

3.1 Internal Network

The internal network for the SP® is a Single Point of Failure, and the failure of a single segment of the network will become a global network failure. If the network fails then applications will not be affected but system administration tasks will not be possible and the heartbeat demons, which run on all nodes and monitor their health, will not be able to communicate and determine the current status of the nodes. HACMP will register the internal network failure but do nothing further.

The SP® internal network cannot be made redundant, but the effect of network failure can be minimised by restructuring using routers (or switches) so that each node has its internal network adapter connected directly to the router, and not to a LAN with just a single connection to a router. If a network cable then failed for a single node connection, then only this node would be affected and the SP® would not suffer a global network failure. If a router/switch is used, then there should also be a backup router/switch facility so that the router/switch itself does not become a Single Point of Failure in the system.

3.2 Client Networks

A client network allows user access via workstations to applications residing on server nodes, and in most cases is likely to be a Single Point of Failure since backup networks for client access are very rare due to the expense involved. To minimise the effects of network failure, each client network should be designed so that it can be broken down into segments, either by connections using routers, intelligent switches, hubs, or some other means, so that multiple routes should be available through the network to cluster nodes.

If the portion of a network directly connected to a node fails, then there can be no access from any client workstations to any applications currently running on the node and so a takeover to another cluster node will be

initiated. If the takeover node is on the same segment, then HACMP will register a global network failure for the cluster and no takeover will occur. To prevent this happening, it is recommended that all nodes in a cluster be on different segments of the network.

If a network segment other than one directly attached to a node fails, then provided that there is an alternative route around the failed segment, clients on other segments will still be able to access the node. It is recommended that client workstations be spread across as many segments as possible to lessen the impact of the failure of a single segment.

The design of client networks where DNS or NIS is configured must also take into account what would happen if a primary/master server is longer accessible due to network failure. DNS and NIS primaries and secondaries must be located on separate segments of the network to ensure that at least one is available when a network segment fails.

3.3 Switch network

Since there is only one switch network in an SP®, this is a Single Point of Failure. The latest SP® switch technology has built-in redundancy and recoverability which means that switch faults are local rather than global faults, and a total switch failure is now a rare event. If data access over the switch is critical, then consideration should be given to the installation of a backup high speed network, such as FDDI, to be used in the event of a global switch failure.

To determine how the switch could fail, and to minimise the effects of failures, it is necessary to look at individual components of the network.

The SP® switch network consists of the switch board in each frame, which contains various chips, switch adapters in the nodes, and internal and external cabling connecting these components. The switch board receives power from the frame power units, which have built-in redundancy. Since switch faults tend to be local, in the event of switch cable failure the only affect is on the link where the fault occurred. Switch cable failure will thus mean a node takeover.


Each SP® has a primary and backup node for the E-primary node, which is the node that initialises the entire switch network and recovers when switch faults are detected. The E-primary is thus eliminated as a Single Point of Failure, but it is recommended that the primary node and the backup primary be located in separate frames since if this were not the case and the frame containing both nodes lost its power supply, then a global switch network failure would occur.

In each frame there are four switch chips handling communications to nodes; each of these chips is responsible for four separate nodes. If one of the chips fails, then communication through the switch to all four nodes serviced by the chip will be lost. For disk subsystems to be highly available, they need to have at least two paths for communication to protect against node and adapter failure. If the nodes that the disk subsystem are connected to are on the same switch chip, then there is the possibility that switch communication of data on the disk subsystem could be lost. Disk subsystems are protected against single chip failures by having the disks connected to nodes in separate frames, which contain different switch chips.

A switch clock is used to ensure proper data reception and also that the current time of day across the switch system is the same. There is one master switch board in an SP® and any additional switch slave boards will obtain clock information from the master; the master switch board is usually in the first frame. On each master switch board, there is a master chip, which provides the clock function to the other chips, which in turn provide connections to the nodes and other switch boards.

The SP® switch is designed with a backup master chip to ensure optimum availability. If the master chip failed without a backup, the SP® would lose its switch time synchronisation, and thus globally fail. The recovery process to the backup is manual and it is recommended that the backup be configured on a slave board in a separate frame to eliminate this potential Single Point of Failure.

4. Trademarks

- The  logo is a registered trademark in Germany of Mannherz EDV-Dienstleistungen
- AIX is a registered trademark of IBM Corp. in the United States and/or other countries
- IBM is a registered trademark of IBM Corp. in the United States and/or other countries
- MVS is a trademark of IBM Corp.
- OS/2 is a registered trademark of IBM Corp. in the United States and/or other countries

All other names of products or companies mentioned on these pages are trademarks or registered trademarks of their respective owner.