Open Systems SAN Sales Checklist

The Solution Assurance information, tools and documentation ("Materials") are being provided to IBM Business Partners to assist them with customer installations. Such Materials are provided by IBM on an "as-is" basis. IBM makes no representations or warranties regarding these Materials and does not provide any guarantee or assurance that the use of such Materials will result in a successful customer installation. These Materials may only be used by authorized IBM Business Partners for installation of IBM products and otherwise in compliance with the IBM Business Partner Agreement.

Getting Started

Initially you will want to learn something about SANs, how they can be used, and so on. You will also want to understand the various reasons a customer my decide to implement a SAN. Then you can look at actually putting a solution together.

IBM Global Services has many SAN-related services that can help you along the way, and these should be reviewed to see which are appropriate. In addition to normal consulting, design and implementation services, for instance, IGS has a Fiber Transport Service to handle the cabling needs for a SAN, and a Storage Assessment Service that can take an inventory of used (and unused) storage and make recommendations based on application needs and potential savings and benefits.

Documentation

The following documents will give you a lot of information about SANs in general, and IBM's Strategy for SAN:

- Start at http://www.storage.ibm.com/ibmsan/
  - ❖Links to different product categories at left
  - ❖Links to documentation at right
- Sample SAN Configurations paper found at www.ibm.com/suppor/techdocs document # WP100197
- Redbooks: (at http://www.redbooks.ibm.com search on "Storage" and "Area")
  - ❖Introduction to Storage Area Network (SAN) SG24-5470
  - ❖Designing an IBM Storage Area Network, SG24-5758
  - ❖Storage Area Networks: Tape Future in Fabrics , SG24-5474
  - ❖Planning and Implementing an IBM SAN , SG24-6116-00
  - ❖Many more available
- SAN Data Gateway Cookbook on this Techdocs website (w3.ibm.com/support/techdocs)
- "Designing Storage Area Networks" by Thomas Clark SR23-9397

Education

Course descriptions can be found at http://w3.education.ibm.com . Both the customer and local field personnel need to be trained.

- **SS700** - a 1 1/2 day intro to SAN and IBM's SAN Strategy is the current "starting point" for SAN education.
- Computer-based training
  - ❖**SE62000W** - Fibre Channel and Storage Area Networks Fundamentals (approx 5 hours)
  - ❖**SE61000W**- IBM 2109 Fibre Channel Switch (approx 7 hours)
  - ❖**SE11155P** - 2108-G07 Storage Area Network Data Gateway (approx 1 hour)
- **S9E27** - Fibre Channel and SAN Fundamentals with Hands-on Lab (5 days) This course goes into Fibre Channel protocols in depth, re-enforced with lab exercises. Negligible IBM content. Use of FC Analyzer in labs.
- **IHC0464C** - SAN Technology and Troubleshooting - 3 day PD/PSI course intended for Help Desk personnel

Customer Requirements/Goals and IBM products

Sometimes customers have a specific problem they are trying to solve, and sometimes they just want to start implementing Storage Area Networks as a longer term strategy. In either case, the question "What do you want to do with a SAN?" becomes important. Following is a list of SAN uses and benefits and the IBM products to consider to solve these problems.

- **Distance and Connectivity** - This is the main function of Fibre Channel networks. This connectivity is then taken advantage of by various pieces of software to provide numerous other benefits (such as those below.) IBM provides Fibre Channel switches (3534-F08, many 2109 models, 2031s, 2032s and 2042s) that can allow hundreds of servers and storage devices to be interconnected. Maximum distances supported by IBM between a given server and a storage device is 70km, but in some cases this is plenty enough to make a difference. A single hop can be 10km over single-mode fiber or up to 500 meters over multi-mode fiber. This is much more flexible than SCSI's limit of 25 meters. Servers can access storage without having to be moved, off-site tapes can be written to directly, and remote mirroring is also possible at these distances. This easy connectivity also allows...
- **Storage Consolidation** - probably the biggest driver toward SAN solutions today. Customers want to manage storage separately from servers. Consider a large storage subsystem such as the Enterprise Storage Server (ESS) or one of the FAStT products. These products can attach many hosts, and storage can be parceled out to the servers as needed. Some storage can be left unassigned, and then given to servers as their requirements increase. Storage can also be added to these subsystems without interrupting current processing. In essence, these subsystems provide the benefits of storage consolidation within a single box.

  By providing connectivity for servers to multiple disk subsystems, a SAN can provide storage consolidation benefits across multiple subsystems. By providing connectivity from multiple servers to tape drives, a SAN can allow tape pooling or tape sharing between applications such as a backup application. (Remember, the SAN only provides the connectivity, it is up to the software in the servers to handle sharing requirements.
- **Disk Sharing** - implementing clusters that share volumes is easier due to SAN connectivity, but care must be taken to ensure that software in the servers are available to handle this. For instance, you do not want different operating systems to see the same drive. Even when using the same

operating system, cluster configuration work is necessary to keep things straight. For instance, for NT, additional software (Microsoft Cluster Services) is required.

- **Tape Sharing** - currently Tivoli Storage Manager is the IBM software required to share tapes between multiple servers. Other backup products also allow tape sharing. The SAN infrastructure mentioned in the Distance and Connectivity bullet above will provide the connectivity to allow Tape Sharing.
- **File Sharing** - although once thought to be one of the "later" benefits of SAN, IBM can do this today with the Tivoli SANergy product. Standard locking mechanisms are used between heterogeneous servers to allow concurrent access to files over a SAN. Once again the infrastructure provides the connectivity, and Tivoli SANergy provides the software to take advantage of it.
- **LAN-Free Backup** - Tivoli Storage Manager *Clients* - which are typically customers' file and application server - can do LAN-free backup today.
- **Server-less Backup** - Tivoli Storage Manager has announced use of Outboard Copy commands (in limited environments today) to allow backups to flow across a SAN using some other data mover not in any Tivoli Client, Agent, or Server. The Tivoli software will make the request of a third party in the SAN (today, a DataMover component of a SAN Data Gateway which is withdrawn from marketing and cannot be ordered after 12/13/2002), and that third party will handle the data movement across the SAN and then acknowledge the completion of it.
- **Centralized Management** - Tivoli will provide software (Tivoli SAN Resource Manager) to perform many management functions. These include besides Tivoli Storage Manager function mentioned above:
  ❖SAN Discovery including SAN connectivity
  ❖Centralized control of which server can get to which disk drives. The customer will be able to view, allocate, remove, and authorize access to storage either via GUI, or automatically via policy.
  ❖File System monitoring and extension
- **Virtualization** - IBM can do some SAN-based virtualization using DataCore's SAN Symphony software today. Future products from IBM will provide block-level virtualization through SAN connectivity, and a SAN File System to facilitate heterogeneous sharing of data.

Design Issues

The first and foremost questions to answer when designing SAN solutions is to answer the following questions:

- What servers need to get to what storage?
- How much bandwidth is required by the server applications, and how much is available from the storage devices?
- What kind of high availability requirements are there?

While these questions may seem obvious, they are often forgotten when the Siren Song of SAN clouds the minds of otherwise clear-thinking people. It is worth going into them in a little detail...

1. **What servers need to get to what storage?** - This is a matter of application requirements, and sharing requirements. In most cases, actual storage volumes or drives (be they physical drives or

logical ones such as those defined in a disk subsystem) need not be shared between servers. In some cases they will, but for any given volume there are usually more servers to be *prevented* from accessing the storage than need to be allowed. This means that you don't have to jump through hoops figuring out how to achieve any-to-any connectivity. It is all right to start out with more than one physical fabric - set of interconnected switches. (Most customers have two separate fabrics in a high availability configuration. Devices and servers have connections to both fabrics and thus various failures will not stop processing. Any server or device can be attached to more than one fabric by using more than one Fibre Channel adapter/port. In any case, answering this question is part of what allows you to translate business need to an actual configuration.

2. **How much bandwidth is required by the server applications, and how much is available from the storage devices?** - If you only look at storage needs, you can end up attaching far too many servers to a single storage subsystem. Subsystems such as the ESS and the FAStT subsystems have well know maximum aggregate bandwidth capability, and you need to configure enough subsystems to satisfy the aggregate bandwidth needs of the servers. It is possible in some cases for the aggregate bandwidth needs of the servers to be so low that you can reasonably attach many many servers to one subsystem, but in most instances server bandwidth requirements will limit how many (and which) servers can connect to a single ESS or FAStT box.
3. **What kind of high availability requirements are there?** - Where is the customer willing to tolerate a single point of failure? Most high availability configurations have multiple adapters in servers connected to separate SANs connected to multiple adapters in the storage devices. Typically this means that there are now multiple paths from a server to a drive. *There must be software in the servers to handle this.* With the FAStT devices that means the appropriate RDAC software and with the ESS that means the IBM Software Device Driver (SDD).

NOTE: There is no support for use of SDD through a SAN Data Gateway.

Server Requirements

The table shown on the next page is a convenient way to collect information required to answer the previous questions. Make as many copies of this table as needed to include all servers participating in the SAN. The columns are to be filled out with the following information:

- Server - Identify the server with any unique name that is meaningful to you
- Operating System and Level - What level of what operating system is running
- Current adapters - number and type of adapters today (e.g. SCSI, SSA, Fibre Channel)
- Proposed adapters - number and type of adapters after SAN migration (this may change as more detailed design is done
- Storage required - How many Gigabyte or Terabytes of disk storage do you initially want to assign to this server?
- Bandwidth required - At peak load, what bandwidth in Megabytes/sec is needed in and out of this server.
- Device Sharing - Will this server share devices with any other server? Which ones? (This does not imply that it will use the devices all of the time, what matters is whether it can "see" devices that

another server “sees.” However, work must be done to ensure that all servers “seeing” the same device have been configured correctly including any necessary extra software. For instance, two NT servers need Cluster Services software correctly configured.)

- Multiple Path - Will the server have multiple paths to the same device (LUN)? Any unique pair of server-adapters and storage-port that can access the same LUN represents a different path to that LUN.)Only one of the two endpoints need be different for it to be a different path.
- Other devices - will this server have access to other devices such as tape or already existing storage subsystems? List them.

Server	Operating System and Level	Current Adapters	Proposed Adapters	Storage Required	Bandwidth Required	Device Sharing?	Multiple paths?	Other Devices?

Which switch philosophy is appropriate?

Often, when choosing a solution, one consideration is whether to go with switches or directors. Switches typically have little or no redundant hardware, except perhaps an additional power supply. The term "director" refers to a fully-redundant switch with no single point of failure except for the ports themselves. This is really like having 2 switches in one box, one of them in hot standby to take over non-disruptively should the active one fail. This non-disruptive failover can be deliberate, as when upgrading to a new level of code that has been downloaded to the backup hardware. Typically, all control logic, power, fans, shared memory and the switching mechanism(s) themselves are duplicated, so that only a port or port-card failure requires manual intervention to recover from. As long as there is a spare port card in the director then only cable(s) need be moved from failing port(s).

Directors are more expensive than switches, and typically have more ports as well. Fabrics built with just switches can achieve redundancy through using multiple switches with multiple paths, such that a switch outage does not remove all paths between any two devices. Before discussing how one would decide which to use, let's look at the switches and directors IBM can provide.

IBM provides switches and directors from three different vendors. From Brocade, IBM OEMs the 2109 switch in several models. The 2109-S08 and S16 are 8 and 16-port switches with 1Gb ports and have been withdrawn from marketing. The 3534-1RU was a specialized switch for private-loop servers and is also withdrawn from marketing. These switches can attach to most any kind of device and are the most widely deployed today. IBM currently sells the 2109-F16 which is a 16-port switch with 2Gb ports, and the 3534-F08 which is an 8-port 2Gb switch. (There are more differences between Brocade's 1Gb and 2Gb switches then just speed, but this is not discussed in this paper.) The 2109-M12 is a large chassis based device - it is actually two 64-port switches - that has redundant hardware like a director, but as of 10/28/02, the failover requires a 20-30 second outage - causing path changes etc. - and so for now the 2109-M12 is still not quite in the director class. Brocade is working on a microcode upgrade to reduce this outage to zero time (as with directors from McDATA and INRANGE).

IBM resells the McDATA product line which includes a 64 port director (2032-064), a 140-port director (2032-140), and a 32 and 16 port switch (2031-232 and 231-216) none of which support FC_AL. IBM also sells the 2031-242, a 24-port switch that can support FC_AL. All of these switches/directors also offer 2Gb ports. A previous 1Gb FC_AL "loop switch" (2031-L00) is also available. (Most McDATA switches and directors do not support loop connection as do the 2109s, so loop-only devices such as 3590 tape drives must either connect to a 2031-224, or to a 2031-L00 which then has an "uplink" to a switch or director.) There are also 1Gb products that have been withdrawn from marketing. These are the 2032-001 director, and the 2031-032 and 2031-016 switches. McDATA directors (2032s) can be configured to run FICON and/or FCP protocols. The switches (2031s) are FCP only.

IBM also resells a 64-port, a 128-port, and a 256-port INRANGE director, that can also run FICON and/or FCP. The INRANGE directors can support FCP loop devices as well.

.

Some customers will be willing to pay for the “industrial strength” reliability of the directors (indeed some demand it.). Others will opt for a switch fabric with sufficient extra switches to provide high availability. In essence, directors take the “ESCON approach” to high availability, and switches take a “switched LAN approach” to high availability.

The main issues in choosing between switches and directors are the size and complexity of the fabric, and the “pain” associated with downtime or outages. For example, can you tolerate the outages associated with microcode upgrades? Can you live with fewer paths until a new switch is acquired and cabled up?

Since directors have more ports than switches, large fabrics are easier to build with directors. Large complex fabrics using switches will need at least twice as many switches as the same fabric built with directors and depending on how many ports must be dedicated to inter-switch links, and how big the directors are, the difference can be a factor of 3 to 1, 4 to 1, or even higher. Usually, the cost of a fabric will be lower when using switches (you should always run different configs and then compare the price - prices change rapidly!), but it does dramatically increase the complexity involved. Every time you add new switches to a fabric you have to figure out how many inter-switch links (ISLs) are needed to handle the traffic, this design effort is not trivial, and is greatly simplified when using fewer, higher-port directors

Restoring an installation back to a full configuration after an outage will take longer with switches. Even if extra switches are “on the shelf” to replace a failing switch, much physical re-cabling is required to substitute a switch. With directors, the redundant switch in the chassis takes over immediately (with no loss of data), and the failing switch can then be replaced non-disruptively, basically to “replace the spare tire.”

In a port failure, there will often be some loss of data, and there is also the need for manual intervention here to move a cable. For an entire SAN to be fault-tolerant, you need extra HBAs in servers and extra ports on devices. You connect these either to different switches or to different port cards on a director (or to different directors of course.)

Microcode upgrades to 2109s cause short outages, and may need to be scheduled in maintenance windows, whereas with INRANGE and McDATA products, after the new code is downloaded, a non-disruptive upgrade to the new level of code can be done. Fallback to the previous level of code is just as easy. (With directors, the old code may still be in the backup control processor, allowing a fallback to the old code without having to do another download.)

There is no one correct answer regarding switches and directors. It depends on the costs and pain to a customer for various kinds of outages. Some customers can tolerate short outages quite easily and do not really need non-disruptive code upgrades and immediate non-disruptive failover. Others experience so much pain from any outage (or from the size and complexity of their SAN) that the additional cost of the directors is easily justified.

Today, with the McDATA product line IBM can provide a mix of directors and switches, in which the industrial strength directors are in the core of a SAN (where an outage affects many many servers and devices), and the less expensive "edge switches" will be on the boundary. This kind of core/edge topology is also available with 2109s using the M12 in the core. In the future, when switch-to-switch standards are implemented, multi-vendor fabrics will be possible, yielding more options for mixed director and switch networks.

NOTE: Very large configurations are possible with all vendors because it is not required that all switches be interconnected. By putting multiple adapters in storage devices and servers you can have them connected to multiple different fabrics (separate, not-connected islands of interconnected switches) and still get all the connectivity you want without exceeding the maximum fabric size for a given switch. Fibre Channel architecture allows for up to 239 switches in a single fabric. Vendors typically have smaller maximums that they have tested and certified, although they will often agree to support larger fabrics when asked. Care should be taken not to exceed whatever maximum IBM is currently supporting.

Staying out of trouble

Items to consider/potential pitfalls

- **Fibre Channel hosts need to access non-Fibre Channel devices?** - The customer may want to use non-FC storage devices with their FC SAN. Typically these are either SCSI or SSA devices. There are products available to provide this, care must be taken to ensure that enough are ordered.
  - **SCSI devices** may not me upgrade-able to FC. In some cases with newer devices, the native Fibre Channel adapters are not available yet. The IBM SAN Data Gateway allows SCSI devices (Magstar tapes, VSS, and ESS) to connect to Fibre Channel servers. The IBM SAN Data Gateway Router is a smaller less expensive gateway for tape only. (These are being withdrawn from marketing and cannot be ordered after 12/31/2002.) Enough gateways/routers/bridges need to be configured to handle this. Support will be an issue as these devices reach end of life.
  - **SSA disk** can be used by Fibre Channel servers through 7140 SLIC Routers.
- **Any HP or iSeries servers?** - Older HP adapters, and iSeries prior to V5R2, were private loop only and could only access devices attached to its loop or to devices attached to Quickloop ports on 2109 switches, or TL-ports on INRANGE directors. The need for private-loop server support is going away, but may still be out there. In many cases, support for the older adapters requiring this support has been dropped so check carefully when dealing with these.
- **Any devices or servers loop only?** - Older server adapters, older devices, and most current tape drives only support loop attachment. Older versions of the SAN Data Gateway only support loop attachment. These cannot attach to many McData devices (only 2031-L00 and 2031-224).
- **Management needs** - Are Storwatch Specialists enough? Experts needed? Tivoli Framework?
- **Sharing** - if multiple servers can see the same volumes, the appropriate software must be included and configured - (e.g. MS Cluster Services - HACMP)

- **"Un-sharing"** - by default SANs are any-to-any. Provisions must be made with zoning and/or LUN masking and/or physical path separation to ensure servers only see the LUNs they are supposed to see.
- **Multi-pathing** - multiple paths to the same LUN from the same server require appropriate software in that server (e.g. IBM Subsystem Device Driver)
- **High Availability Requirement** - if HA is required ensure no single points of failure. This usually means extra adapters in hosts and devices, requiring more FC ports in the SAN *and* appropriate software for multi-pathing support. Use of two separate SANs for HA is appropriate.
- **Potential "function interference"** - it is quite likely that a single server may need many or all of the functions specified earlier in this paper. Some combinations of software may not be supported (e.g. Multi-pathing and clustering software often interfere, and you typically cannot use two different multi-pathing add-ons in the same server, certainly not on the same HBA
- **Verify support for the configuration, and identify all software and microcode requirements. Many pieces may need to be upgraded for your configuration.** - Check the supported server matrices. Typically these are found by going to the external product site, and looking for the link to "Interoperability" or "Interoperability Matrix" or sometimes "Connectivity" or "Supported Servers." The first place to look is the site for the specific storage devices you are interested in connecting. Then look switch sites and/or HBA sites pointed to, and also the SDG site if you are using that. Additional information can be found in the IBM Salesmanual pages available at http://www.ibmlink.ibm.com/ussman.
- **Is there SSA storage that the customer wishes to re-use?** Either use with an ESS (if the SSA storage is the appropriate model) or use the 7139 SLIC router to go from Fibre Channel hosts to SSA disk. If re-using SSA disk through an ESS data needs to be migrated.
- **Are there distances requiring hops greater than 500 meters?** These require longwave connections using single-mode fiber.
- **Are there distances requiring hops greater than 10km?** In some cases there are special optics available (transceivers) to plug into a switch point instead of the normal transceiver that will allow you to go longer distances. *Typically, these are just for inter-switch links as you need these on both sides of the link.* There is also limited support for DWDM approaches. In some cases, a simple RPQ will allow longer distances. In **all** cases, you will want to ensure sufficient buffers available on each side of the link to drive it to full utilization. At 1Gb you need 6 buffers for each 10km of distance in the link. At 2Gb, you need 12 buffers for each 10km of distance.
- **Software requirements** - each device has its own Specialist for configuration. 2032s and 2042s use a client/server approach using a dedicated, supplied PC for the server machine (much like the ESS Specialist). The The SAN Data Gateway also uses client/server approach but the server piece can run on any machine. 2031s can be managed via a browser, as can the 2109s and 3534.(The "Specialist" is a webserver built into the switch microcode.) 2109s, 2108s, and the 3534 have Telnet capability for a command-line interface, while 2031s, 2032s, and 2042s only have limited command-line capability through their serial ports.

Switch to switch connections

Deciding on switch to switch connections normally requires some understanding of bandwidth needs or capabilities of the servers and devices. Different vendors have different restrictions on the maximum size

of a fabric, and who many hops are available. These maximums are constantly being upgraded, and thus should be checked each time you are putting together a solution.

Parallel links and alternate paths can be used to add bandwidth with the following considerations:

- Switches will only load balance between paths of equal hops. If there is only one "shortest path" between a server and storage device then that is the only path that is used.
- Thought must be given to how much data might actually travel across these links. For instance, if a single ESS is the only subsystem on the SAN, with F models you are going to have something less than 400 Megabytes/per second total bandwidth coming out of that Shark and you would only need to configure to handle that much bandwidth

Getting more help

SANs represent a rapidly changing environment, and it is not uncommon in a given situation, to have a question that is not answered by the on-line documentation. Here is a partial list of places to go for more help:

- **Partner-Line -** IBM Business Partners have access to these people for questions not answerable via on-line documentation
- **Techline -** IBM field employees, and PartnerLine can go to Techline for assistance
- **QASearch and QAAuthor -** Business Parnters and IBMers can research and create questions about IBM products
- **TechExpress -** For complex situations, a TechExpress request can be built, and appropriate experts will be gathered for a conference call, design review, or whatever may be needed