**Title: Storage Area Network, an Introduction of basic concepts**  
  
  
**1**  
Storage Area Network, an Introduction of basic  
concepts

* Antonella Corno CCIE Storage CM

**2**  
Summary

* What is a SAN
* Basic Building Blocks of a SAN
* A zoom into the Storage Architectures
* SAN elements and architecture
* Basic Protocols and Mechanisms
* Who is who Standardization Bodies and Industry  
  Organizations

**3**  
Summary

* What is a SAN
* Basic Building Blocks of a SAN
* A zoom into the Storage Architectures
* SAN elements and architecture
* Basic Protocols and Mechanisms
* Who is who Standardization Bodies and Industry  
  Organizations

**4**  
Definition

* A SAN (Storage Area Network) is a network  
  designed to transfer data from servers to  
  targets, and it is alternative to a directly  
  attached target architecture, or to a DAS  
  architecture, where the storage is connected to  
  the servers on general purpose networks
* Additional definitions of a SAN imply that the  
  SAN should also be highly performing, and should  
  be such to enable storage devices to communicate  
  with one another and with computer systems

**5**  
Different technologies

* Multiple technology can be used when building a  
  SAN traditionally the dominant technology is  
  Fiber Channel, but IP based solutions are also  
  quite popular for specific applications
* The concept of SAN is also independent from the  
  devices that are attached to it. Can be disks,  
  tapes, RAIDs, file servers, or other

**6**  
SAN and NAS network and node

* SAN vs NAS while a SAN is a network connecting  
  storage subsystems, the NAS is a storage  
  subsystem, making use of a general purpose  
  network.
* The SAN is an extension of the disk channel  
  outside the server, while the NAS is a disk  
  subsystem connected to the servers, in most cases  
  via an IP network.

**7**  
Summary

* What is a SAN
* Basic Building Blocks of SAN
* A zoom into the Storage Architectures
* SAN elements and architecture
* Basic Protocols and Mechanisms
* Who is who Standardization Bodies and Industry  
  Organizations

**8**  
SAN interconnections

* As said, different technologies can be used to  
  interconnect the network nodes, extending the  
  Disk interface outside the server
* Fiber Channel is a dedicated channel based high  
  performance and highly available network based on  
  Fiber Channel Protocols
* iSCSI is SCSI protocol carried over an IP  
  network. In this case the network infrastructure  
  can be shared with other applications
* SCSI network is an extension of the internal SCSI  
  bus, used for short distances due to its parallel  
  architecture

**9**  
Initiator and Target in FC SANs

* Fiber Channel Node can be the source or the  
  destination of information
* If the node is an Initiator (source), it is  
  usually connected to the network via an HBA (Host  
  Bus Adaptor), which is the physical connection  
  interface, and can be based either on electrical  
  or (more often) optical technology
* If the node is a target (destination), it can be  
  a JBOD (Just a Bunch of Disks), a RAID (Redundant  
  Array of Independent Disks), or a Storage array

**10**  
Hard Drives

* The basic storage element is an Hard Drive. They  
  are made into complex devices composed of  
  platters, heads, cylinders and tracks
* The Logical Block Addressing (LBA) addresses the  
  sector within the disk. Modern drives have 512  
  byte sectors
* File systems arrange files into sectors so that  
  they can be stored and retrieved
* The File system usually deals with clusters of  
  blocks and uses a FAT (File Allocation Table) to  
  map a file to the sectors

**11**  
FAT  
**12**  
JBODs and RAIDs

* While a Jbod is a group of disks packaged in an  
  enclosure and connected via a FC loop, a RAID is  
  a more sophisticated device, that may improve  
  performance and/or reliability of the storage  
  device
* RAID is improving performances reading/writing  
  information from a set of disks at the same time,  
  and reliability adding parity and/or mirroring  
  information on multiple disks of the array
* RAID can be performed in HW via a controller  
  embedded in the enclosure or software on the host

**13**  
RAID 0 or striping

* Data are split onto different disks for  
  performance increase performances depend on  
  information unit size vs stripe size
* No redundancy added
* Cost is limited (no additional hardware)

**14**  
RAID 1 - mirroring

* Data are replicated on multiple disks for  
  redundancy
* Performance may be impacted if copy is done  
  serially
* Increase of cost proportional to the amount of  
  redundancy
* More complex algorithm to manage multiple copies

**15**  
RAID 3

* Data protection via ECC (Error Correction Control  
  code)
* Good redundancy
* Performance not changed for reads but lower for  
  writes since the ECC need to be calculated
* Cost is only 1 extra disk for the entire logical  
  array

**16**  
RAID 5

* Data protection with ECC, but parity is spread on  
  the array
* Good redundancy
* Same speed reads, slower writes
* One disk per array of added cost

**17**  
Summary

* What is a SAN
* Basic Building Blocks of a SAN
* A zoom into the Storage Architectures
* SAN elements and architecture
* Basic Protocols and Mechanisms
* Who is who Standardization Bodies and Industry  
  Organizations

**18**  
RAID array vs DAS vs NAS vs SAN (I)

* A RAID array is an enclosure containing a set of  
  disks and a RAID controller providing in hardware  
  the features of a RAID 0 5 and usually some  
  caching engine
* A DAS (Direct Attached Storage) is an  
  architecture for which the storage is privately  
  attached to the servers cannot be shared, it is  
  hard to scale, expensive and complex to manage.  
  80 of the market it is still DAS

**19**  
RAID array vs DAS vs NAS vs SAN (II)

* NAS (Netwrok Attached Storage) is an architecture  
  for which the storage is attached to the servers  
  via a multi-purpose network, and it is accessed  
  at a file level via protocols like CIFS or NFS
* The network is usually an IP network
* TCP can be tuned to optimize storage transport

**20**  
DAS vs NAS architecture  
**21**  
SAN architecture

* Storage is accessed at block level not at file  
  level
* Very high performances
* Storage is shared
* Good management tools
* Interoperability issues

**22**  
Summary

* What is a SAN
* Basic Building Blocks of a SAN
* A zoom into the Storage Architectures
* SAN elements and architecture
* Basic Protocols and Mechanisms
* Who is who Standardization Bodies and Industry  
  Organizations

**23**  
SAN

* SAN is the convergence of an almost-error-free,  
  very reliable interface and the advantage of a  
  networking infrastructure
* It takes advantage of a controlled and reliable  
  data transfer interface and adds the capability  
  of switching and sharing proper of the networks
* Relies on at least 10 to -12 error probability  
  infrastructure

**24**  
SAN basic requirements

* The concept of SAN is to bring a block level  
  interface on a very reliable support, and add the  
  advantages of a networking infrastructure
* Need to be scalable in terms of speed, fast and  
  efficient getting rid of all the low level error  
  recovery mechanisms will treat the errors as  
  exceptions at application or firmward level
* Allow extensions like long distance, multi  
  protocol support (SCSI, ESCON, TCP/IP..), rely on  
  great simple connectivity

**25**  
SAN supported topologies point to point

* Point to point is the simplest topology for very  
  limited connectivity needs
* It guarantees in order delivery and full  
  bandwidth access
* The application can handle any multipath  
  connectivity to a set of disks in case this is  
  provided, since no other elements are present in  
  this topology

TX  
TX  
RX  
RX  
**26**  
SAN supported topologies arbitrated loop

* Designed to scale to a limited number of nodes  
  (up to 127)
* Low cost (no interconnecting devices needed)
* Arbitration protocol is designed to manage media  
  sharing across nodes may be disruptive when a  
  node gets added/removed from loop and loop  
  initialization protocol kicks in
* A arbitrating hub can be used instead of a  
  distributed protocol

Each node can be a server or a storage device  
**27**  
SAN supported topologies switched fabric

* In a switched fabric topology, switching element  
  get added to the nodes to allow interconnections  
  via point-to-point links
* Extended number of devices (potentially  
  thousands) and greater distances can be achieved
* Scalable, robust and reliable architecture, but  
  the cost of the interconnection devices adds on

**28**  
Summary

* What is a SAN
* Basic Building Blocks of a SAN
* A zoom into the Storage Architectures
* SAN elements and architecture
* Basic Protocols and Mechanisms
* Who is who Standardization Bodies and Industry  
  Organizations

**29**  
Fiber Channel and the others

* Fiber Channel and SCSI they are not  
  alternatives FC will act as a transport layer  
  for SCSI, amongst other protocols
* Fiber Channel and IP sometimes that are  
  alternatives, sometimes not. While FC is used  
  instead of a TCP/IP network to connect reliable  
  the servers to the storage, IP can still be used  
  to extend FC networks over long geographical  
  distance in turn in some applications, FC can be  
  used to carry IP traffic (mostly for in-band  
  management)

**30**  
FC protocols and services

* FC offers a layer 2 transport service, and uses  
  layer 3 services to provide it, and to map upper  
  layer protocols like SCSI, HIPPI, IP or others
* FC0 defines the physical interface (data rates,  
  connectors, media, distances, power and so on
* FC1 deals with encoding/decoding of bits on the  
  physical interface and low level signalling
* FC2 defines framing and classes of services
* FC3 provides common services like multicasting or  
  possibly encryption
* FC4 maps upper layer protocols onto Fiber Channel
* FC-0 to 2 are the core of the FC link protocol,  
  and will be examined in more detail in the next  
  set of slides

**31**  
FC0 physical interface

* FC supports many different variations of physical  
  interfaces Electrical and Optical both Multi  
  Mode and Single Mode. They vary in terms of cost  
  and performances (distance and speed)
* The basic physical access to a FC port is so  
  called GBIC (Gigabit Interface Connector) which  
  is a hot swappable connector supporting all types  
  of media, and comes in different form factors.  
  The MIA (Media Interface Adaptor) is an  
  electrical DB-9 to optical interface converter,  
  that can be applied externally to the enclosure  
  (disk or host)
* Fiber channel connections are always  
  bi-directional, point to point, serial. No drops  
  or taps are considered in FC
* All links should be capable of BER (Bit Error  
  Rate) of 10 to -12 or less

**32**  
FC-1 Line encoding

* FC-1 is in charge of bit transmission on the  
  line. Since FC uses a serial transmission, an  
  encoding in needed in order to transmit the clock  
  signal along with the data. This is achieved  
  mapping any pattern of 8 bits into a 10 bit  
  pattern that will guarantee enough signal  
  transitions, and so enough clock signal power on  
  the line. The clock signal is filtered and  
  recovered at the receiving point, and the bits  
  are decoded
* Not all the possible 10 bits sequences are used  
  only the ones that guarantee enough transitions  
  are allowed. The others are invalid chars. Each 8  
  bits char has 2 coding on 2 10 bit patterns with  
  opposite disparity (numbers of 0s and 1s), or 1  
  coding if the 10 bit pattern has no disparity  
  (Same number of 0s and 1s)
* To avoid accumulation of DC at the receiver, the  
  encoder remembers every time it has transmitted a  
  char with disparity (different numbers of ones  
  and zeros) and will send the next char coded with  
  opposite disparity. This mechanism is called  
  running disparity

**33**  
FC-1 characters

* FC is a word oriented architecture 4 characters  
  form a word
* All possible 256 chars are mapped into 10 bit  
  strings, and further 10 bits strings are  
  available outside the coding space. Those are  
  called special characters.
* The encoder can encode data char as well as data  
  char, and an input signal to the encoder will  
  signal the encoder itself if the coming pattern  
  is data or special char (K/D)

**34**  
FC-1 Transmission words and Ordered sets

* The transmitted words are composed by a sequence  
  of 4 characters.
* The first character of the sequence will tell if  
  the remaining three need to be interpreted as  
  control signals or data. If it is a K28.5 then  
  the next 3 char are to be considered signals, and  
  the word is called ordered set
* The ordered set are divided into Primitive  
  Signals (fill words and non-fill words), Frame  
  delimiters (SOF and EOF) and Primitive Seqiences  
  (if they are repeated multiple time on the line  
  like NOS, OLS, LR, LRR, LIP, LPB, LPE)

**35**  
FC-2 the frames

* Before any data is echanged, a session need to be  
  established between the end ports, FC-2 is in  
  charge of login session
* Once the link is established, FC-2 will handle  
  frames by using sequences and exchanges
* FC-2 is also defining the frame format for FC
* It also handles flow control and class of services

**36**  
FC-2 Session

* Before any data can be exchanged between end  
  points, a session need to be established. During  
  login the end nodes enchange information like  
  identification, credits (number of outstanding  
  frames supported)
* When the session is not needed, it gets cleared  
  by a logout operation
* Within the session each operation happens within  
  an exchange, that is a set of frames logically  
  related to each other (example all the frames  
  implementing a SCSI operation)
* A set of frames that constitutes a high-level  
  protocol information unit, is grouped into a  
  sequence. An exchange can host multiple sequences
* An FC frame is 2112 bytes long, and contains  
  addressing fields as well as control fields like  
  exchange ID, sequence ID

**37**  
FC-2 flow control

* Flow control is performed in FC via use of  
  Credits.
* At the beginning of a transmission session, the  
  two end point will negotiate how many frames the  
  sender can send before they get acknowledged by  
  the receiver (end to end credits end to end  
  flow control)
* If there are connection elements in between the  
  two end point the credits will also be negotiated  
  at each link (buffer to buffer credits)
* Each end point of a link will acknowledge receipt  
  of the frames sending back credits (link by link  
  or end to end) to the sender, which is now  
  allowed to send more frames

**38**  
FC switching fabric architecture services

* FC switching fabric relies upon a number of  
  software applications to deliver the transport  
  services to the mapped upper level protocols
* Amongst the many the most relevant ones are  
  Zoning, Alias services, Domain Controller, Name  
  Server, FSPF
* Those services allow the fabric to function with  
  no disruption, and maintain a consistent set of  
  distributed information across the nodes of the  
  fabric, and the end nodes (hosts and storage)

**39**  
Summary

* What is a SAN
* Basic Building Blocks of a SAN
* A zoom into the Storage Architectures
* SAN elements and architecture
* Basic Protocols and Mechanisms
* Who is who Standardization Bodies and Industry  
  Organizations

**40**  
Standardization Bodies and Industry Organizations

* FC is standardized by ANSI, in T11 committee  
  (www.T11.org)
* SCSI is standardized by ANSI, in T10 committee  
  (www.T10.org)
* IETF is dealing with iSCSI, FCIP, IPFC (IP/FC  
  protocols) requests for comments (RFC)  
  (www.ietf.org)
* SNIA is the Storage Networking Industry  
  Association and it is not in charge of  
  standardization (www.snia.org)
* FCIA is the Fiber Channel Industry Association  
  and it is not in charge of standardization  
  (www.fibrechannel.org)
* UNH is University of New Hampshire, and it is  
  traditionally involved with interoperability  
  third party testing. It is not a standard body  
  (http//www.unh.edu)