

Terminology

This section defines the terms discussed in the positioning section and provides references to additional information

1. **What is Storage Networking?**

The practice of creating, installing, administering, or using networks whose primary purpose is the transfer of data between computer systems and storage elements and among storage elements.

2. **What is a Storage Area Network (SAN)?**

A network whose primary purpose is the transfer of data between computer systems and storage elements and among storage elements. A SAN consists of a communication infrastructure, which provides physical connections, and a management layer, which organizes the connections, storage elements, and computer systems so that data transfer is secure and robust. SANs are typically used to connect block storage devices but can also be used for system interconnection in clusters and to connect NAS. Block and object-based storage devices use SCSI-3 command sets to access data.

3. **What is Storage Virtualization?**

Storage virtualization, sometimes called SAN virtualization, is the abstraction of storage, decoupling application data from its storage. In-band, or symmetric, implementations are host-independent but typically suffer from performance limitations regardless of whether they include switching hardware ("SAN Controller") or not ("SAN Server"). Out-of-band, or asymmetric, implementations may separate data from control traffic to improve performance and scalability ("Meta Server") or use clustered file systems for good price/performance but limited scalability ("SAN Cluster"); however, both require host software and server-to-server communication.

4. **What are the benefits of storage virtualization?**

The key benefits are:

- better storage utilization and ability to satisfy peak demands by pooling,
- increased staff productivity since it is easier to manage a unified view of storage than separate products,
- improved scalability and availability by decoupling applications from disruptive storage events,
- enhanced security by not exposing access to physical storage systems,
- greater flexibility and reduced risk of obsolescence by supporting products from multiple vendors using multiple technologies, and
- reduced risk of vendor lock-in with the concomitant costs of ownership.

5. **What is Network Attached Storage (NAS)?**

Storage elements that connect to a network and provide file access services to computer systems. A NAS storage element consists of an engine which implements the file services, and one or more devices, on which data is stored. The file engine "head" and the backend storage may be combined in a single "filer" or decoupled. Clients use file protocols like DAFS, NFS, CIFS, SAMBA, HTTP, and FTP to access data in filers.

6. **What is a Cluster?**

A collection of computers that are interconnected for the purpose of improving reliability, availability, serviceability and/or performance. Often, clustered computers have access to a common pool of storage, and run special software to coordinate the component computers' activities. *"In Search of Clusters, Second Edition"* by Gregory F. Pfister is a great introduction to clustering – its cover art inspired the name of Microsoft "Wolfpack" cluster server.

7. **Which networking technologies can be used for storage networks and clusters?**

Fibre Channel, Gigabit Ethernet, and InfiniBand networks are the primary standard technologies being used or considered for storage networking and clustering. Proprietary technologies played a significant role in niche markets but the cost and interoperability advantages of standards-based networking has reduced their acceptance.

Storage Area Networks originated in the S/390 mainframe environment with the introduction of ESCON serial channels and directors in 1990. However, Fibre Channel is the dominant technology used to implement SANs in the much larger "open systems" market. Ethernet is the dominant technology to connect NAS (to LANs) and is being promoted as a SAN for IP Storage. Clustering is implemented primarily using high-performance proprietary connections or using Ethernet but Fibre Channel is also well-suited for interprocessor communication. InfiniBand is a new contender for intra- and inter-processor communication as well as for storage networking. Although designed to coexist with Ethernet and Fibre Channel networks initially, its long-term deployment strategy is to displace both within the electronic business data center.

Asynchronous Transfer Mode (ATM) is a link layer telecommunications protocol that uses T1 services or SONET as a physical layer. It may be used to connect SANs over MAN/WAN distance.

8. **What is SCSI?**

The Small Computer Systems Interface is both a command set and a bus hardware specification. It originated from the IBM S/360 selector channel, was scaled down into the Shugart Associates Systems Interface, and became an ANSI standard in

1986. SCSI-2, released in 1994, had higher performance, lower overhead, improved functionality, broader device support, and greatly improved compatibility. SCSI-3 is the latest set of standards designed to support both parallel bus and serial network interfaces. SCSI standards are developed by the NCITS T10 Technical Committee (<http://www.t10.org/>).

The SCSI Trade Association (www.scsita.org) was formed in 1995 to promote parallel SCSI.

Parallel SCSI started using a narrow bus (50 pin connector) and grew into a wide bus (68 pin or 80 pin SCA-2 backplane connector). The maximum bus length depends on the number of devices attached on the bus; table shows length for 8 (narrow) and 16 (wide) attached devices.

Parallel SCSI	Max MB / sec	Bus width (bits)	Bus length (meters)		
			SE	HVD	LVD
SCSI-1	5	8	6	25	na
SCSI-2 Fast/10	10 20	8 16	3 3	25 25	na
SPI Fast/20 Ultra SCSI	20 40	8 16	1.5 na	25 25	na
SPI-2 Fast/40 Ultra2 SCSI	40 80	8 16	na	12	12
SPI-3 Fast/80 Ultra3 SCSI	160	16	na	na	bp
SPI-4 Fast/160 Ultra320	320	16	na	na	bp

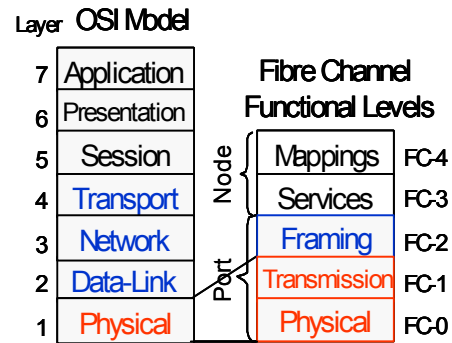
The SCSI-3 Architecture specifies a variety of command sets, transport protocols, and physical interconnects. Command sets are available for block, stream, jukebox, CD-ROM, RAID

controllers, enclosures, object-based storage devices and primary (in common) commands. Transport protocols include SCSI Parallel Interface (SPI), Serial Bus Protocol (SBP over IEEE 1394), Serial Storage Architecture SCSI-3 Protocol (SSA-S3P), SCSI over Scheduled Transfer (SST), Fibre Channel Protocol (FCP), and the SCSI RDMA Protocol (SRP).

9. What is Fibre Channel (FC)?

The Fibre Channel standard specifies how to serially transport multiple upper level protocols using a common physical medium over campus-wide distances. The spelling of "Fibre" was chosen to indicate support for copper media as well as glass fiber. The standard is specified using five functional levels, FC-0 through FC-4. Most specifications are developed by the NCITS T11 Technical Committee (<http://www.t11.org/>).

These levels do not correspond cleanly to the 7 layer OSI model but some similarities can be noted. FC-0 handles the physical interface and media and FC-1 defines encoding/decoding and serial transmission protocol. These physical and signalling properties are specified in OSI layer 1. FC-2 packages data into sequences and exchanges as well as in frames, like OSI layer 2, provides routing information and segmentation/reassembly processes like OSI layer 3, and reliability features (such as a 32-bit cyclic redundancy check, buffer-to-buffer and/or node-to-node flow control, classes of service, and a recovery protocol) like OSI layer 4. Devices that support the Fibre Channel standard are called nodes. Nodes contain one or more ports which connect to ports in other nodes. Port information, corresponding to levels FC-0, FC-1, and FC-2, is specified in the FC-PH standard first approved in 1994 and later separated into FC-PI (FC-0) and FC-FS (FC-1 and FC-2).



Node information corresponds to level FC-3 and level FC-4. FC-4 standards include mappings of channel, network, and other protocols onto Fibre Channel. Many FC-4 upper layer protocols (ULPs) are developed by T11 (e.g., FC-VI) but some are specified by T10 (e.g., FCP, SRP) or the IETF (<http://www.ietf.org>) (e.g., IPFC).

There are three basic types of ports, the Node (N), Fabric (F), and Expansion (E) port. N ports reside on servers or storage systems. F and E ports reside on switches. The term L port means that a port can participate in an arbitrated loop. When arbitrated loop capabilities are added to the basic port types they are called NL, FL, and G ports. A Generic (G) port can act as an E port, an F port, or an FL port, depending on what connects to it.

A brief history of organizations promoting the benefits of Fibre Channel may also be useful.

The Fibre Channel Systems Initiative (FCSI) was a temporary organization formed in 1991 by IBM, HP, and Sun to kickstart the market. Its duties were handed over to the Fibre Channel Association (FCA) in 1995. The U.S. FCA later merged with the Fibre Channel Loop Community (FCLC) to become what is known as the Fibre Channel Industry Association (FCIA). See <http://www.fibrechannel.org/>.

The Open Standards Fabric Initiative (OSFI) was a consortium of vendors (Brocade, McDATA, Ancor, Vixel, Gadzoox) promoting interoperability between Fibre Channel switches. The efforts of OSFI concluded in May 2000 with a demo at N+I and the submission of OSRP as a routing protocol to FC-SW-2. OSRP has been replaced with FSPF (Fabric Shortest Path First) from which OSRP had been derived. FSPF specifies a common method for routing and moving data among Fibre Channel switches. The 6 elements of interoperability are Link Initialization; Principal switch selection and Domain ID assignment; Distributed Name Server; Distributed State Change Notification; Zoning; and Routing.

Channels typically operate in a closed, predictable environment where error-free delivery is paramount. Networks often operate in an open, unstructured environment where on-time delivery (e.g. video) takes precedence. Fibre Channel was designed to efficiently support both environments. Fibre Channel also supports both optical (glass) and electrical (copper) media.

Physical Media core/cladding (2.5mm jacket)	Transmitter (LW laser can use SM or MM)	Distance (meters feet)
9/125 micron single-mode	Long wave laser (1300 -1550 nm)	10,000 32,808
50/125 micron multi-mode	Short wave laser (780 -850 nm)	500 1,640
62.5/125 micron multi-mode	Short wave laser (780 -850 nm)	175 574
Twinax (STP)	Electrical (ECL) 150 ohm	30 98.4
Video coax	Electrical 75 ohm	25 82
Mini coax	Electrical 75 ohm	10 32.8

The most common media types are 50 μ optical multi-mode short wave laser and twinax copper. Connector types are dual SC, LC, CG, or MT-RJ for optical and DB-9 or HSSDC for copper.

Fibre Channel products that support net data rates of 100 and 200 MBps in each direction (using 1.0625 and 2.125 Gbps signaling rates) are currently available. Designs for higher rates (4.25 Gbps and 10 Gbps) are in development.

The Interoperability Laboratory at the University of New Hampshire (<http://www.iol.unh.edu>) develops test suites for Fibre Channel standards compliance. The Laboratory for Computational Science and Engineering at the University of Minnesota (<http://www.lcse.umn.edu>) provides a facility to test and apply Fibre Channel solutions.

Unlike parallel SCSI which requires direct bus attachment, Fibre Channel supports device connection in point-to-point (dedicated bandwidth), loop/hub (shared bandwidth), and switched fabric (scaled bandwidth) topologies.

10. How is a FC arbitrated loop initialized?

Initialization is how each port is dynamically assigned one of 127 Arbitrated Loop Physical Addresses (AL_PA). A loop initialization primitive (LIP) is propagated around a loop after a loop device (L_Port) is powered on or detects a failure. The device with lowest port name (or the fabric if present) becomes the loop master (by sending a LISM frame). A bitmap is sent around the loop using four frames (LIFA, LIPA, LIHA, LISA) which allow devices to assign their AL_PA by priority. If all devices support them two additional frames are sent (LIRP and LILP). Initialization is complete once the loop master sends a close primitive signal. Once a device gains control of the loop (by sending an

arbitrate primitive signal; lowest AL_PA wins) it establishes a point to point link to a target device (by sending an open primitive signal); other devices in the loop simply repeat the data.

11. What are ESCON and FICON?

Enterprise System CONNECTION (ESCON) and Fiber CONNECTION (FICON) are the IBM names for the SBCON and FC-SB-2 single-byte command code set standards used for mainframe I/O. The ESCON serial architecture superseded the OEMI parallel channel. It uses half duplex 200 Mbps optical cables and MT-RJ connectors. FICON is an FC-4 ULP that retains the strengths of ESCON while benefiting from the faster, full-duplex, and multiple concurrent I/O operations capabilities of Fibre Channel.

12. What is Ethernet?

The Ethernet cabling and signaling scheme was co-invented by Robert Metcalfe and David Boggs at Xerox's Palo Alto Research Center in the early 1970s. Ethernet is often used to refer to any local area network using a carrier sense multiple access with collision detection (CSMA/CD) MAC protocol to share the media. Ethernet originally supported 10 Mbps over shielded coaxial cable with a maximum segment length of 500 m. The Institute of Electrical and Electronics Engineers 802.3 specification was developed in 1980 based on Ethernet. However, whereas Ethernet defined a single physical layer protocol, 802.3 specified several including 10Base5 (similar to Ethernet) and 10Base2 (thin coax). In 1990 the IEEE finalized 10BaseT supporting twisted pair, star topology and 100 m segment length. In 1995, the IEEE standardized 100BaseT (Fast Ethernet) increasing bandwidth to 100Mbps while retaining the 802.3 CSMA/CD protocol. As a result, 100BaseT supports the

applications and networking software running on 802.3 networks.

13. What is Gigabit Ethernet (GbE)?

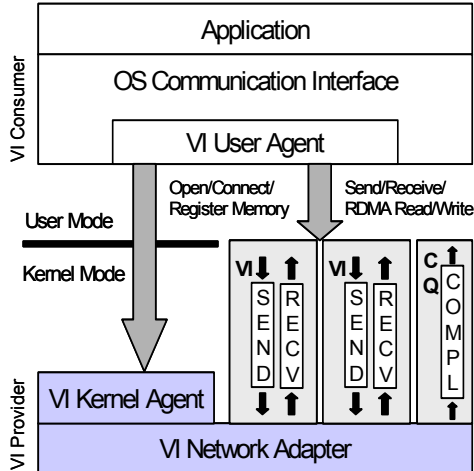
Gigabit Ethernet refers to extensions to the IEEE 802.3 standard that increase LAN speed to 1000 Mbps. There are two different standards but both look identical to Ethernet from the data link layer upward. The 1998 IEEE 802.3z standard grafted the IEEE 802.2 LLC and IEEE 802.3 full or half duplex MAC sublayers on top of the FC-1 and FC-0 Fibre Channel levels. 802.3z supports single-mode fiber (1000BaseLX, 5 km) multi-mode fiber (1000BaseSX, 550 m for 50 μ , 275 m for 62.5 μ), and copper STP (1000BaseCX, 25 m). The IEEE 802.3ab standard was added in 1999 to support the most common cabling, category 5 copper UTP (1000BaseT, 100m). GbE also added a Gigabit Media Independent Interface (GMII) between the MAC and physical layers to support the different encode/decode schemes used by 802.3z and 1000BaseT.

The Gigabit Ethernet Alliance has further details (<http://www.gigabit-ethernet.org/>). The Gigabit Ethernet Consortium (<http://www.iol.unh.edu>) was formed in 1997 to conduct conformance and interoperability testing for both GbE standards.

14. What is Virtual Interface (VI) Architecture?

The Virtual Interface (VI) Architecture specifies a high performance network interface consisting of Virtual Interfaces, Completion Queues, VI Providers, and VI Consumers. A VI provides a protected, directly accessible interface to a VI Network Interface Card without the overhead of traditional network architectures. A VI consists of a Send Queue and a Receive Queue which hold Descriptors of data movement requests. VI pairs are connected to support bi-directional, memory to memory

transfer. A "doorbell", typically implemented as a memory-mapped register on a VI NIC, is used to notify a VI NIC that work has been placed on one of the Work Queues.



A Completion Queue holds information about completed Descriptors from multiple Work Queues. A VI Provider instantiates a Virtual Interface using a VI NIC and a VI Kernel Agent. The VI Kernel Agent is a device driver for the VI NIC that registers application memory and manages VIs. The VI Consumer communicates using a VI. It consists of an application program, a communications facility, and a VI User Agent. The VI User Agent is the software that abstracts the underlying VI NIC in accordance with an interface defined by the Operating System communication facility. An implementation of the VI User Agent is called a Virtual Interface Provider Library (VIPL).

VI/TCP (Internet VI) is an IETF Internet-Draft document that defines enhancements to the VIPL API which support VI functionality during operation over TCP/IP. These enhancements allow the VI Provider to supply Descriptor Flow Control and negotiate the Maximum Transfer Unit Size downward. Though designed to be compliant with the VI Architecture and the VIPL API it has not gained

broad support so the WARP protocol has been proposed as a replacement.

WARP is a data transfer mechanism that uses self-describing packets to reduce memory requirements. WARP separates upper layer protocol headers and data by using separate messages. WARP may be layered on top of TCP or used as an extension to SCTP. In both cases each stream allows mixing both RDMA and Send semantics. iSCSI can also be accelerated using WARP using either Send or RDMA transfer mode for each iSCSI PDU, as specified by WARP.

The VI Architecture specification is available at (<http://viarch.org>). The VIPL specification is standardized by the VI Developers Forum (VIDF).

15. What is RDMA?

Remote Direct Memory Access (RDMA) is a high performance mechanism to read and write data from or to remote memory without intervention by the remote processor. Protection tags are used to ensure that processes can only access memory to which they have been explicitly authorized. In contrast, in the traditional Send/Receive model, the sender specifies where to find a message in its local memory and the receiver determines where to place the message in its remote memory.

RDMA Read copies data from a virtually continuous remote memory region into local memory according to a scatter list. When used with reliable delivery, an RDMA read response indicates data sent previously has been received on the remote side. Periodic zero length RDMA reads can be used as a cluster heartbeat since the remote side is alive if it completes.

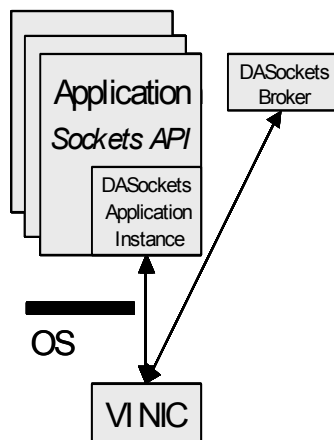
RDMA Write copies data from local memory regions according to a gather list into a continuous remote memory region. RDMA operations usually do not require pre-posting a receive descriptor on the remote processor.

The exception is when immediate data is included in an RDMA write operation in which case a receive descriptor is required to hold the 4 bytes of data.

RDMA operations are supported by VI and InfiniBand. RDMA Read is only supported for reliable service types. RDMA Write is also supported for unreliable connections.

16. What is (Fast) Direct Access Sockets?

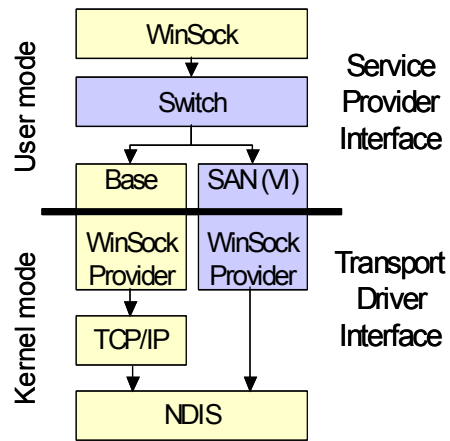
Direct Access Sockets (previously called Fast Sockets, or FSock) is an emulation of the BSD sockets interface for networking applications. Whereas traditional sockets implementations assume packetization, lossy media, and operation in a protected mode, DASockets builds atop VI to provide low-latency, high-throughput, communication. DASockets Application Instances intercept an application's use of sockets using a Broker to discover remote DASockets nodes. A set of protocols collectively called the DASockets Session Protocol controls the operation of Application Instances and Brokers. DASockets also allows administrators to specify which sockets are eligible for acceleration and to capture and analyze DASockets statistics.



17. What is Microsoft WinSock Direct?

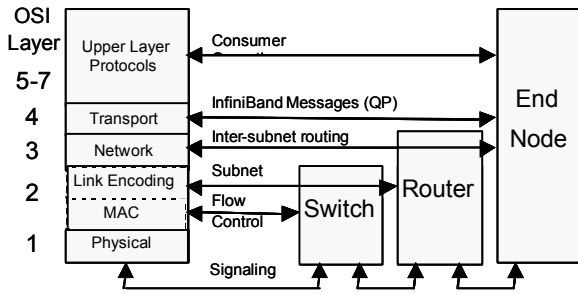
Windows Socket Direct allows applications using the WinSock API to

transparently benefit from a high performance System Area Network (SAN) interconnection. A software Switch is used to choose, on a per-connection basis, whether to use the standard TCP/IP provider or a "direct path" SAN (VI) Provider. Microsoft suggests that ERP solutions, e-commerce applications, database vendors, and technical and scientific developers will benefit most. WinSock Direct is the basis for Sockets over InfiniBand (SoIB).



18. What is InfiniBand?

InfiniBand defines a switched communications fabric for connecting multiple end nodes with high bandwidth and low latency in a protected, remotely managed environment. It defines the host behavior (verbs) and memory operation such that the channel adapter can be located as close to the memory complex as possible. The InfiniBand point-to-point switched fabric supports reliable messaging (send/receive) and memory manipulation (remote DMA) without software in the data path. It provides enhancements to VI. Operation of the InfiniBand Architecture can be described as a series of layers corresponding to layers 1-4 of the OSI Model, as depicted in the next figure.



An InfiniBand subnet is composed of end nodes, switches, routers, and subnet managers interconnected by links. Each device may attach to one or more switches and/or directly with each other. Multiple links can exist between any two devices. An end node can communicate using multiple ports and multiple paths to provide fault tolerance and increased bandwidth.

management agents referred to as General Service Agents that can be accessed through a well known interface called the General Service Interface.

Work is underway to layer traditional network and storage protocols over InfiniBand. IP over InfiniBand (IPoIB) activities are underway within the IETF to allow InfiniBand to act as the link layer for IP and ARP. The InfiniBand Trade Association Software Working Group is designing a Sockets Direct Protocol (SDP) to support the general requirement for stream Sockets over InfiniBand (SoIB). The NCITS T10 technical committee is also specifying the SCSI RDMA Protocol (SRP) to standardize SCSI over InfiniBand.

IB Link Width	Laser (nm), Connector	Optic Fiber (microns)	Distance (meters, feet)
1X-SX	850 nm, dual LC	62.5/125 μ m multimode	2 - 125, 6.56 - 410.1
		50/125 μ m multimode (beige)	2 - 250, 6.56 - 820.2
1X-LX	1300 nm, dual LC	9/125 μ m singlemode (blue)	2 - 10,000, 6.56 - 32,808
4X-SX	850 nm, MPO (MTP)	62.5/125 μ m multimode	2 - 75, 6.56 - 246
		50/125 μ m multimode	2 - 125, 6.56 - 410.1
12X-SX	850 nm, dual MPO	62.5/125 μ m multimode	2 - 75, 6.56 - 246
		50/125 μ m multimode	2 - 125, 6.56 - 410.1

The signaling rate for 8b/10b encoded data is 2.5 Gbps, resulting in a 2.0 Gbps raw data rate. In addition to providing support for both copper and optical media like Fibre Channel and Gigabit Ethernet, InfiniBand also supports three link widths. The 1X, 4X, 12X links provide for simultaneous transmit and receive of one, four, or twelve bits of encoded data respectively on copper differential pairs or on fiber optic cables.

19. What is a distributed file system?

An InfiniBand Subnet Manager is responsible for configuring and managing switches, routers, and channel adapters. Each node provides a Subnet Manager Agent that the Subnet Manager accesses through a well known interface called the Subnet Management Interface. Each node may also contain additional

Distributed file systems allow one computer on a network to use the files and peripherals of another networked computer as if they were local. Distributed file system protocols are typically developed by one vendor and licensed to other vendors to provide compatibility. Example protocols include Sun Microsystem's Network File System (NFS was standardized as RFC 1094 in March 1989; NFS V3 was standardized as RFC 1813 in June 1995; NFS V4 was submitted as an Internet Draft in June 2000) and IBM's Server Message Block subsequently developed further by Microsoft and others. The Common Internet File

System (CIFS) is an enhanced version of the SMB protocol used as the native Windows file sharing protocol. Samba is a popular open source software suite that provides file and print services to SMB/CIFS clients. Each component of the Samba suite is described in a separate manual page (see <http://www.samba.org>). More advanced distributed file systems include AFS, ARLA, DCE DFS, Coda, Ficus, InterMezzo, Sprite/Zebra, xFS, GFS, and Lustre.

20. What is the Direct Access File System?

The Direct Access File System is a shared file access protocol designed to take advantage of high performance RDMA mechanisms. DAFS is designed to allow high-speed, fault-tolerant, consistent views of files to servers that may be running different operating systems. DAFS protocol and API specifications are available at <http://www.dafscollaborative.org>.

DAFS clients may be implemented in user space or protected (kernel) mode so applications may use DAFS in an optimized fashion or transparently. Some implementations export a device interface as well as a file interface so applications expecting block storage can also use DAFS filers, further blurring the distinction between SAN and NAS.

21. What is IP Storage?

The use of IP-based networks to transport block storage traffic. Encapsulations of SCSI (iSCSI, EtherStorage), FCP (SoIP), and Fibre Channel (FCIP), are being considered by the Internet Engineering Task Force (IETF) but are not yet standards. Adaptec EtherStorage, 3ware NSU, IBM TotalStorage, FalconStor IPStor, and Cisco SN5420 are example IP storage solutions. Other vendors are using IP encapsulation to create wide area storage networks. Examples include Nishan Systems "Storage over

IP", Entrada Networks "SANs over IP and light", SANcastle InfiniLink, CNT UltraNet, Pirus, and partnerships between Brocade/Cisco Catalyst 6000 and Vixel/Lucent OptiStar. Proponents of IP storage argue that using Fibre Channel to build a SAN entails substantial retraining, additional investments, and a second network to provide out-of-band management. A more pragmatic approach would apply the mix of proven technologies best suited to address specific business challenges (such as availability and performance) rather than relying on prejudiced assumptions or proprietary implementations.

22. What is iSCSI?

Internet SCSI describes a transport protocol that maps the SCSI remote procedure invocation model on top of the TCP protocol. The SCSI layer builds/receives SCSI Command Data Blocks and relays/receives them with the remaining parameters to/from the iSCSI layer that builds/receives iSCSI Protocol Data Units and relays/receives them to/from one or more TCP connections that form an initiator-target "session". Use of VI/TCP (RDMA over TCP) has been proposed as a solution to concerns about reducing iSCSI host overhead, especially for data phase transfers. iSCSI is supported by Adaptec, Agilent, Alacritech, Cisco, HP, IBM, Intel, Quantum, SANGate, SANRAD, and SAN Valley.

23. What protocols does EtherStorage use?

The SCSI Encapsulation Protocol is an IETF Internet-Draft offered by Adaptec to demonstrate that Gigabit Ethernet can be used as a storage interconnect. SEP is a session-layer protocol that assumes a reliable transport protocol such as STP (SAN Transport Protocol) or TCP. The specification allows internet-wide access to storage using the SCSI command set but is intended to be used in the data center using

Gigabit Ethernet as the link layer. The network stack from SEP on down is intended to be processed in a host adapter.

24. What is Storage over IP (SoIP)?

Storage over IP is a trademark of Nishan Systems referring to a set of protocols designed to implement a fibre channel fabric on an IP network. Two variants of a storage transport protocol (OSI layer 4 encapsulations of FCP) are defined: **mFCP** for metro-area and local-area SANs based on UDP (informational document) and **iFCP**, a gateway-to-gateway protocol for interconnecting Fibre Channel and mFCP SANs via the TCP/IP public network. The Internet Storage Name Service (**iSNS**), companion protocol is designed to meet the requirements of iFCP and mFCP and has been specially tailored to support iSCSI. iSNS can be used with either UDP or TCP transport protocols. SoIP is supported by ATL, ADIC, BakBone Software, BMC, Chaparral, DataCore, Emulex, Eurologic, IBM, JNI, Legato, Nortel, QLogic, Quantum, Seek Systems, Spectra Logic, StorageNetworks, Sun Microsystems, Veritas, and XIOtech.

25. What is Fibre Channel over IP (FCIP)?

Fibre Channel over TCP/IP (**FCIP**) specifies a way to encapsulate Fibre Channel frames over TCP/IP and describes mechanisms that allow islands of FC SANs to be interconnected over IP-based networks. FCIP relies upon TCP for congestion control and management and upon both TCP and FC for data error and data loss recovery. FCIP is designed to transport all classes of FC frames as datagrams using an IP network with as good or better bit-error-rates and line speeds as the Fibre Channel environments being bridged. Note that FCIP cannot carry Class 1 FC traffic or primitives.

FCIP assumes a high bandwidth, high reliability, low loss link level transport such as Gigabit Ethernet, SONET, ATM, or DWDM. FCIP is an IETF Internet-Draft sponsored by the IPFC Working Group including Brocade, Gadzoox, Lucent, LightSand, McDATA, QLogic, and Vixel.

Note: FCIP should not be confused with IPFC. IPFC is an FC-4 upper layer protocol that allows IP traffic to run over Fibre Channel networks.

26. What is flow control and how is it implemented by different technologies?

Flow control is used to prevent buffer overruns. Parallel SCSI uses REQ/ACK and disconnect for this purpose. Fibre Channel uses end-to-end and buffer-to-buffer credit-based mechanisms. A credit represents the ability of a device to accept a frame. EE credits are established during node login and are replenished by acknowledgements. BB credits are established during fabric login or set to zero in a loop topology. BB credits are replenished by receiver-ready ordered sets. Gigabit Ethernet uses IEEE 802.3x flow control. If a receiver becomes congested it can send a pause frame to the other end of the point to point link, instructing the sender to stop sending packets for a specific time period. InfiniBand provides features in both the link and transport layers to manage network congestion. The link layer offers virtual lanes, credit based flow control, and static rate control. Virtual lanes create independent connections sharing the same physical link. Credits indicate the number of 64 byte units that a receiver periodically indicates it is prepared to accept. Injection rate control constrains a channel adapter to transmit at or below a maximum bandwidth. Fractional injection rate controls can also be provided by limiting the rate at which credits are issued. The InfiniBand end-to-end flow control mechanism is optional. The sender keeps track of the number

of work queue entries is has consumed and may not send more data than allowed by the credits granted by the receiving channel adapter. Some implementations can also limit the number of unacknowledged packets in reliable connections.

27. What is class of service (CoS) and how is it implemented by different technologies?

Class of service refers to different levels of data delivery guarantees, connectivity, and bandwidth used to satisfy different applications.

Fibre Channel defines five classes of service though only classes 2 and 3 are widely implemented. Class 1 guarantees in-order delivery of all frames using a *dedicated connection*. Class 2 does not require a dedicated connection but does *acknowledge frame delivery*. If multiple routes are available traffic congestion could result in out-of-order delivery unless in-order delivery is requested. Class 2 is suitable for mission-critical workloads and bursty transactions like database updates. Class 3 provides connectionless *datagram* services that do not acknowledge frame delivery. Class 3 is required and is used by SCSI-FCP. It is widely used in arbitrated loop configurations where in-order delivery results from the topology. However, Class 3 frames may be discarded in congested switched fabrics, triggering FC-4 level recovery. Class 4, defined for fabric topology only, defines *virtual circuits* to guarantee access among multiple destinations without allocating the full bandwidth of a Class 1 connection. It is suitable for time-sensitive applications but is not widely implemented due to complexity and cost. The idea for Class 5 involved an isochronous, just in time service but it is not specified in the Fibre Channel standard. Class 6 is similar to Class 1 but it uses a fabric multicast service. A multicast server replicates and

forwards frames to all N_Ports registered in a multicast group. Applications requiring *multicast* capabilities can use either Class 3 or, if acknowledged delivery is required, Class 6.

The TCP/IP protocol suite provides a reliable connection-oriented service (TCP) and datagram service (UDP), similar to FC Classes 2 and 3. The InfiniBand transport layer provides reliable and unreliable services for both connections and datagrams. Reliable service guarantees that messages are delivered at most once, in-order, and without corruption.

InfiniBand allows other protocols to be carried by a subnet. Datagrams that encapsulate such traffic are referred to as raw datagrams. Multicast is also supported.

28. What is quality of service (QoS) and how is it implemented by different technologies?

Quality of Service means providing consistent, predictable data delivery service. The focus of QoS is providing a predictable share of available bandwidth during periods of congestion and being able to measure and report on this service. The QoS Forum (<http://www.qosforum.com/>) is an international, industry forum accelerating the adoption of standards-based QoS technologies including DIFFSERV, RSVP, MPLS, and 802.1q.

The DMTF defines QoS as "A collection of technologies that allows applications and/or users to request and receive predictable network service levels in terms of data throughput capacity (bandwidth), latency variations (jitter) and/or propagation latency (delay)."

There are three basic levels of QoS: Best Effort Service (*lack of QoS*) - basic connectivity, no guarantees.

Differentiated Service (*Soft QoS*) - Some traffic is treated better than the rest. This is a statistical preference, not a hard and fast guarantee.
Guaranteed Service (*Hard QoS*) - An absolute reservation of network resources for specific traffic.

There are two different models of QoS: Provisioned QoS (Differentiated Services) detects priority and manages packet processing by it.
Signaled QoS (Integrated Services) reserves resources per-flow.

IP provides a "best effort" service which means that it can make no guarantees about when data will arrive, or how much it can deliver. In fact, it requires higher level protocols like TCP to simply ensure that data is not lost. This is acceptable for web browsing and e-mail but is not adequate for new applications like audio/video streaming, conferencing, storage networking, or clustering.

ATM uses the term QoS to refer to a set of performance characteristics: Peak-to-peak Cell Delay Variation (CDV), Maximum Cell Transfer Delay (maxCTD), Cell Loss Ratio (CLR), Cell Error Ratio (CER), Severely Errored Cell Block Ratio (SECBR), Cell Misinsertion Rate (CMR).

29. What is a firewall?

Computer hardware or software that prevents unauthorized access to private data. Firewalls can be used in several places. A firewall between the Internet and a site's web servers can limit the ports and protocols open for use by the public. Another firewall can be used to allow access to the database servers only from the application servers. Firewalls can also be used to allow mission-critical corporate data centers to safely connect to and update their web sites.

30. What is web caching?

Storing web content closer to the consumer so browsers have fast access to data even when the Internet is congested or a web server is unavailable. *Software-only* caching products are inexpensive but don't scale well. *Caching servers*, either prepackaged servers or specialty appliances, are better suited for larger networks and are offered with a wide variety of disk and network connections. *Caching service providers* replicate web content on servers located around the world and provide tools to balance load and manage distribution.

31. What is SNMP and RMON?

Simple Network Management Protocol is a widely used method of monitoring, configuring, and controlling networked devices. Many IETF RFC documents specify the network management framework consisting of a management protocol (e.g. SNMPv1), a definition of management Information and events referred to as Structure of Management Information (e.g., SMIV1), sets of common management information (e.g., MIB-II), and an administrative framework based on communities (one agent and one or more managers). SNMP operations are performed on variable identity/value pairs. Monitoring is provided by retrieving pairs based on exact (GET) or approximate (GETNEXT and GETBULK) match of a variable's identity. The TRAP and INFORM operations report events on a managed system to a list of managers. The SET operation is provided to configure and control a managed system. Request/response messages are used to transport the SET/GET operations between managers and agents over a stateless datagram service like Internet's UDP, Novell's PEP, AppleTalk's DDP, or OSI's CLTS. Agents may use a trap message to report events to a manager or

inform/response messages may be used to report events with a manager.

There are many RFC documents which define common information about devices, interfaces, and protocols. These RFCs specify thousands of objects in Management Information Base (MIB) modules. The objects and events for Remote Network Monitoring (RMON) are contained in five MIB modules specified in RFCs 1757, 1513, and 2021 and Internet-drafts HC-RMON and SWITCH-RMON. RFC 2074 is also relevant. The (nearly 900) objects defined in these MIBs are intended as an interface between an RMON agent and management application and are not intended for human manipulation. Organizations may use remote network monitoring instruments, often called monitors or probes, one per network segment, to manage its networks. RMON objects are arranged into several groups. If a probe implements a group, it must implement all objects in that group. RMON enhances remote network management by:

Allowing a probe to perform diagnostics and collect statistics continuously
 Allowing probes to add significant value to the data it collects, and
 Attempting notification of a management station when exceptional conditions occur
 Supporting multiple management stations

32. What is CIM?

The Common Information Model (CIM) is an object-oriented information model designed to unify, rather than replace, existing management protocols that have been successful in limited management domains. For example, although SNMP is the dominant technology used for network management, DMI, CMIP, and RMI are more widely used for desktop, telecom, and Java application management. CIM enables the tight integration of products that use

standard and/or proprietary methods of managing resources. Popular implementations of CIM include Windows Management Instrumentation (WMI) and Solaris WBEM Services. The SNIA also provides an open source Java implementation. CIM excels in its ability to describe relationships between managed objects as well as describing properties of the object itself. The Distributed Management Task Force (<http://www.dmtf.org>) standardizes CIM with input from industry groups seeking to enhance its use in large data center environments.

33. Where can I find additional terminology information?

The following technical glossaries may be of general interest:

- A Dictionary of Storage Networking Terminology (SNIA):

<http://www.snia.org/English/Resources/Dictionary.html>
- A Glossary of Networking Terms (RFC 1208):

<http://andrew2.andrew.cmu.edu/rfc/rfc1208.html>
- Internet Users' Glossary (RFC 1392):

<http://stiwww.epfl.ch/utile/rfc1392.html>
- American National Standard Dictionary of Information Technology:

http://www.ncits.org/tc_home/k5htm/WD.htm
- IBM Terminology:

<http://www-3.ibm.com/ibm/terminology/>

Positioning Technologies

This section describes technical differences between storage networking technologies and suggests where each is best applied. The general approach is to select the best tool for each job. It is possible to use a hammer to drill a hole but it is neither cost-effective nor likely to produce the desired results.

34. Which networking technology is better?

Organizational issues often take precedence over technical issues so some level-setting may add perspective. Network technicians suggest that coalescing all networks on a familiar TCP/IP technology is best. Storage technicians suggest that optimizing networks for data transmission requires different technology than that for client access. Wise business decisions should select whatever technology best meets their challenges most affordably at a given point in time. TCP/IP is ubiquitous for client/server connectivity; Fibre Channel is dominant for storage networking. InfiniBand is positioned to both complement and challenge both technologies in data center environments. Fibre Channel and Gigabit Ethernet use similar physical components so arguments about inventory savings may not be compelling. InfiniBand uses different physical components but focuses on processor/memory connections. Industrial-strength, affordable, and available solutions are the "best". Today, TCP/IP is best for client access, Fibre Channel is best for storage networking and VI is best for cluster interconnect. FC-VI implements VI over the standard FC interconnect. VI implementations over proprietary networks are likely to be tactical solutions. In the future, InfiniBand promises a standard connect for data center environments but is not yet available. InfiniBand will initially

connect to existing client access and storage networks but seeks to eventually replace both within the data center. Risk-averse data center managers tend to change gradually so as to ensure service levels and protect investments. This suggests that all three technologies will coexist for many years.

35. How manageable is each technology?

SAN solutions may require skill to configure and tune but offer comprehensive, flexible, and affordable functionality. NAS solutions tradeoff ease of installation with a decentralized and less scalable approach. As each storage architecture matures, the operational distinctions between them will likely blur. In either case, the important management standards are SNMP and CIM.

36. Compare WinSock Direct, DASockets, and DAFS.

All three technologies can use Virtual Interface to accelerate network communication. WinSock Direct is integrated in Windows 2000 Datacenter Server to allowing users of the WinSock API to transparently use multiple high performance interconnects including VI. DASockets and WinSock Direct emulate TCP/IP so any application that already uses a sockets interface can automatically benefit from VI. DASockets is designed to use VI on Linux, Solaris, and Windows NT. DAFS on the other hand is a distributed file access method optimized for a local file sharing environment. DAFS client code is used to access the file services provided by a DAFS server. TROIKA participates in the evolution and support of all of these interfaces using a unified, standards-based infrastructure in order to simplify storage networking and clustering throughout the enterprise. When available, WinSock Direct, DASockets,

and DAFS should all be promoted, particularly over Fibre Channel and when available, InfiniBand.

37. Is block I/O or file I/O better?

Applications access objects, records, or byte streams saved in databases or files which reside on devices which store data in blocks. Since devices are block-oriented, block I/O is always performed. However, most applications perform their functions best using higher level data services provided by other software layers. An application uses the services of a database or file system to perform its I/O but eventually it uses block I/O to access the storage device. The file system may reside on the same host as the application or be moved elsewhere. NAS filers package the file system with the storage and hide the block I/O. The advantage of this approach is that multiple hosts may share files over a standard network. However, since networks were designed to rely on software to provide reliable transmission, data access over a network is often slower than using channels designed for data transmission. Applications that access file systems on different servers may benefit from faster network speeds. Gigabit Ethernet benefits applications using NAS filers but is unsuitable for access to traditional block-oriented storage devices. Fibre Channel networks were designed to support both types of access and therefore are suitable for accessing data stored in databases, file systems, and filers.

38. Where is Fibre Channel best applied?

Fibre Channel is optimized for reliable and high performance data delivery. For the foreseeable future, it will remain the best technology for storage access, particularly for block storage devices and very large volumes, e.g., imaging. This should not be surprising since Fibre Channel was designed to support more devices, higher

performance and availability, greater distance, simpler cabling, improved scalability, and a broader set of applications than SCSI.

Fibre Channel should be promoted for data center storage networking and clustering today. TROIKA has a significant window of opportunity providing the benefits of a standards-based converged network today, while the similar promises of InfiniBand technology gradually emerge and mature over time. The TROIKA Zentai Controller supports block level access (using FCP) and file level access (using IPFC and FC-VI) operating simultaneously over a single Fibre Channel network.

39. Where is ATM best applied?

Asynchronous Transfer Mode (ATM) is a link layer telecommunications protocol that uses T1 services or SONET as a physical layer. ATM is a "lossy" protocol designed to provide on-time, in-order delivery of isochronous data (voice and video) over long distance. However, it can also be used to carry FC traffic over WAN distances.

40. Where is Gigabit Ethernet best applied?

Gigabit Ethernet is expected to be initially deployed for switch connections to speed backbone and server network connections. Specialized workgroups may also be upgraded but most desktops will remain unchanged. IDC notes that backbones have just been upgraded to support 100 Mbps desktops and it will take a few PC upgrade cycles before GbE is deployed.

Although ATM can also be used to build campus-wide backbones at a moderate price, the key benefit of ATM has been seen in metropolitan and wide area networks.

Many companies are currently exploring the use of Gigabit Ethernet to access block storage. However, the benefits of IP Storage, namely

leveraging current networking skills and lower cost components, remain to be proven once standards are developed and interoperability is demonstrated. Organizational challenges may limit the networking staff from assuming data storage responsibilities and the fact that fiber optic components for Fibre Channel and Gigabit Ethernet are similar may mitigate the touted cost advantages. If the technical challenges (bandwidth, processing overhead, and data integrity) of using Gigabit Ethernet for block storage access don't limit its useful application within the data center, it is possible that InfiniBand will. It is likely that IP Storage will find the best fit with applications that require remote block storage access.

41. What is IP Storage used for?

IP Storage is intended to allow block-oriented storage devices to be accessed using protocols that can run on Gigabit Ethernet. Several proposed standards have been proposed to the IETF but it is likely that product offerings will remain proprietary for a few years before any of the proposals are standardized to enable multi-vendor interoperability.

Fibre Channel over Internet Protocol is intended to connect islands of FC SANs by encapsulating FC frames over IPv4-based networks that may span LAN, MAN, and WAN distances. The motivation behind connecting remote sites includes backup, mirroring, replication, and distance extension.

42. Where is InfiniBand best applied?

InfiniBand will likely appear first within servers as a PCI or PCI-X replacement. It was designed to remove limitations on processor performance imposed by traditional I/O architectures. When the technology matures, it will likely be a good fit for clustered applications, primarily displacing proprietary and Ethernet connections used for interprocessor communication. If and

when native InfiniBand storage subsystems appear, the tradeoffs compared to using native Fibre Channel devices should be evaluated. It is likely that servers with InfiniBand capabilities, starting with backend servers, will predominantly access storage over switched fabrics that route access through existing Fibre Channel storage networks. Customers accessing data from NAS devices are the most likely candidates to use InfiniBand networks for storage access, primarily by displacing Gigabit Ethernet.

43. How are web sites designed for availability?

The first aspect is to eliminate single points of failure. Redundant network access means using multiple paths between each web site and ISP(s) and ensuring the use of multiple network access points from each ISP and the Internet backbone. Protection from site-wide outages also requires multiple geographically-dispersed web sites updated over multiple paths to the data center. Web sites can also be made more available by using fault-tolerant servers or by using clusters of servers that serve as backups for each other. Balancing traffic between web sites and to the server farms within a web site also provides availability and scalability. Server farms are well-suited to handle stateless transactions like http requests. Other aspects of availability include protecting data from user or application errors and from unauthorized access or deletion. Backups, firewalls, user authentication, resource authorization, encryption, logging, monitoring, and periodic reviews by outside auditors can help provide this protection. A high level of availability also requires competent operational staff following proper site management practices including change and problem management.

44. How do data centers provide availability?

Continuous availability seeks to provide both high availability (no unplanned outages) and continuous operations (no planned downtime). Technology can provide protection from component failures through redundancy (e.g. RAID, n+1 power supplies, clustering, IP load balancing) and protection from application errors through versioning (e.g. backup/restore). However, continuous availability requires more than technology; it requires that applications be designed for availability and that people follow processes designed to reduce human error. Availability service levels may also specify performance metrics because degraded applications may be considered unavailable.

Disaster-tolerant architectures are based on redundancy of components, data (mirroring), and servers (clustering) and often assumes some level of geographically-dispersed hardware. Campus architectures use Fibre Channel, Ethernet, or FDDI networks.

45. How can these technologies be used to build a better data center?

Picture an Internet Data Center with racks of high-density InfiniBand servers in the middle, racks of GbE network devices for client connections on the left, and racks of Fibre Channel storage on the right. Client transactions (web pages, email, orders, streaming media requests) are routed through a gateway, firewall, and potentially cached and over a VPN to an Ethernet backbone to web or application servers. The servers, using PCI (or PCI-X or InfiniBand in the future) for intraprocessor connections, may be clustered for availability and performance using TCP/IP or VI over any wire transport. Application servers may access database servers, perhaps over another firewall, and that server accesses block storage or filers. Block data flows over Fibre Channel to the servers (perhaps between the data center and an SSP). File transfers use NFS/SMB (typically over Ethernet) or

DAFS (typically over FC) and may flow over an InfiniBand network in the future. Remote mirroring, replication, bulk data transfer, and geographically-dispersed clustering may also use the MAN/WAN gateway to access other data centers.

Adding network, processor, or storage capacity to today's data center may require that new connections be established between each component that is to have access to a shared resource pool. As the number of elements to be connected increases, the cost and complexity of proper configuration and management rises quickly. Even with switched fabric connections, scaling may be limited by server slot limitations, software restrictions, or available skills.

For example, adding new storage capacity may require connecting the storage subsystems to one or more Fibre Channel switches, perhaps through a SCSI to Fibre Channel bridge, and connecting each host system to these switches by adding one or more Host Bus Adapters to a peripheral bus on each server. Coordinated configuration tasks are also necessary on each server, switch, and storage subsystem along each path to ensure secure access with data integrity.

Building a better data center should allow linear scaling, so only one (or two for redundancy) connection between the new component and the current switched fabric would be required, thus reducing the cost and complexity of adding new processor or I/O capacity. Of course, new technology must coexist within the existing IT infrastructure to protect current investments and reduce deployment risks. This is an ongoing activity as new products emerge, mature, and eventually displace older technology that has served its useful life. This chart illustrates how InfiniBand and faster versions of Ethernet and Fibre Channel can be deployed to upgrade the capabilities of a more affordable, manageable, available, and secure data center.

Market Opportunities

This section quantifies the volumes or revenues associated with various market segments to help prioritize networking technology opportunities.

46. How large is the SAN opportunity?

According to IDC, the worldwide SAN market will grow 58% annually from \$3.4 billion in 2000 to \$13.4 billion in 2004. A June 2001 IDC report projected SAN-attached non-mainframe disk market revenues to reach \$18B by 2004. Chase H&Q reports "The primary drivers of SAN deployment include reduced management costs, improved LAN performance, and increased reliability and availability".

47. How large is the SAN Virtualization opportunity?

A Morgan Keegan report published in July 2000 estimated \$1.2 billion software revenue in 2003 based on port count forecasts and assumptions about ports per SAN, software revenue per SAN, and SAN virtualization software adoption rates.

48. How large is the NAS opportunity?

Dataquest estimates the market for NAS will grow 56% annually from \$1.4 billion in 2000 to \$8.3 billion in 2004 (of which, 70% are high end, 17% midrange, and 13% entry level solutions). Chase H&Q reports "End users install NAS appliances to offload traffic from application servers, for heterogeneous file sharing, and for management simplicity."

49. How large is the clustering opportunity?

IDC estimates the server clustering market will grow to \$1.1 billion in 2004, a CAGR of nearly 25% from \$368 million in 1999. Market leaders in 1999 were Microsoft (20%), EMC (14%), Veritas (13%), Compaq (10%), Legato (7%), Qwest (6%),

Novell and Unisys (each 5%). IDC classifies clustering products within Serverware, a broader segment reaching \$2.7 billion in 1999.

50. How large is the SCSI disk opportunity?

The IDC 1999 *Worldwide Disk Systems Market Forecast and Review* reported 1999 worldwide disk system supplier revenue of over \$28 billion. Storage interconnects were 84.5% SCSI, 7.8% Fibre Channel, 7.1% SSA, and 0.6% other. IDC forecasts a smooth transition from SCSI to Fibre Channel external storage systems through 2003. The worldwide market, driven almost entirely by Windows 2000 and Unix systems, will increase at a 1999–2003 CAGR of 12% to \$46.4 billion.

51. How large is the Fibre Channel opportunity?

Components that implement Fibre Channel standards may reside in servers, storage networks, storage systems, and test equipment.

The Host Bus Adapter (HBA) market in 2000 was \$536M in factory revenue with 627K units according to IDC's *2001 WW FC HBA Forecast and Analysis, 1998-2004*. IDC expects market revenues to grow 42% annually from \$697M in 2000 to \$2.8B in 2004 with port shipments growing 70% annually to 6M.

HBA factory revenue market share leaders in 2000 were Emulex (34.8%), JNI (18.4%), QLogic (17.0%), Compaq (11.8%), Agilent (7.3%), Sun (6.4%), Interphase (2.3%), and others (1.9%). HBA unit shipment leaders in 2000 were Emulex (32.1%), Qlogic (29.1%), Agilent (12.8%), JNI (11.8%), Compaq (7.2%), Interphase (2.7%), Sun (2.6%), and others (1.9%). 97% of HBAs shipped were single port.

IDC forecasts 2003 worldwide hub/switch ports to grow to 5,297,351 representing \$2,788 million in factory

revenues or \$4,159 million in market revenues.

A Chase H&Q report published in June 2000 estimated revenues in 2003 of the host adapter, router, hub, switch, and director segments as \$1.9 billion, \$375 million, \$140 million, \$1.65 billion, and \$1.45 billion respectively.

IDC reported 1999 worldwide factory revenues of \$15.6 million for 184,336 entry hub ports, \$53.8 million for 207,129 managed hub ports, \$14.1 million for 23,003 loop switch ports, and \$100.7 million for 131,755 fabric switch ports, and \$52.4 million for 20,960 director switch ports, and \$269.5 million for host bus adapters. Hub/switch revenue market share leaders were Brocade (38.8%), McData (22.1%), Gadzoox (16.9%), and Vixel (9.8%).

52. How large is the Ethernet opportunity?

IDC reported that more than 85% of all installed network connections were Ethernet by the end of 1997. IDC projected 48 million NICs and 48 million hub ports would be shipped in 1998 compared to combined shipments of about 5 million NICs for Token Ring, FDDI, and ATM. In 1998, the average price per Fast Ethernet port was \$432 for switches and \$73 for hubs/NICs. The Dell'Oro Group estimated that the 1999 price per multimode fiber GbE port will be about \$1610 for switches and \$700 for hubs, compared with \$4800 for 622 Mbps ATM, \$1860 for switched FDDI, \$390 for switched Fast Ethernet, and \$85 for Fast Ethernet hubs.

A June 2001 IDC report stated "there will be very few shared environments where Ethernet is a shared wire/fibre for storage and message traffic. IDC does not expect to see Ethernet SANs using 10Mbs. Or 100Mbs.

53. How large is the VI opportunity?

Virtual Interface is an enabling technology which has many potential applications such as cluster interprocessor communication, TCP application acceleration, and high-performance file serving.

54. How large is the InfiniBand opportunity?

IDC has forecast worldwide InfiniBand revenues to increase from \$236 million in 2002 to \$2 billion in 2004. The 2002 forecast consists of \$129 million for switch ports, \$99 million for servers < \$100,000, and \$7.8 million for target ports. The 2004 forecast consists of \$1 billion for switch ports, \$483 million for target ports, and \$468 million for servers < \$100,000.

55. How large is the DAFS opportunity?

DAFS is positioned to capture a large part of the high-performance segment of the NAS market. IDC reported \$21.4 billion of worldwide supplier revenue for open systems disk storage systems in 1999, \$822 million (3.8%) of which was NAS, \$1.363 billion (6.4%) was SAN, and \$19.2 billion (89.9%) was directly attached. With device interfaces to DAFS filers the opportunity increases to include database and email storage traditionally handled with SAN storage.

56. How large is the software market?

The IDC *Worldwide Software Market Forecast Summary: 2000-2004*, reported the worldwide packaged software revenues were \$156.8 billion in 1999, 78.4% from the U.S., categorized into three primary markets: applications software (\$70.8 billion), system infrastructure software (\$49.5 billion), and applications development and deployment software (\$36.5 billion).

57. How large is the storage software market opportunity?

IDC reported the worldwide market for storage software was \$5.0 billion in 1999. Revenues were classified as \$2.5 billion for the backup, restore, archive, and HSM segment, \$933.5 million for storage utilities, \$819 million for Storage Resource Management, and \$703.2 million for data replication / availability software.

Market definitions, sizings, and forecasts differ and it may be useful to compare IDC figures with those of other leading market analysts.

A 2001 DataQuest report forecast a 26% annual growth in the worldwide storage management market from \$5.3 billion in 2000 to \$16.7 billion in 2005. New license revenue leaders as EMC (25.5%), Veritas (16.3%), IBM / Tivoli (16.1%), Computer Associates (11.7%), BMC (4.3%), Network Appliance (3%), StorageTek (2.7%), Compaq (2.7%), Legato (2.7%), HP (1.7%), and others (13.4%).

The Data Management (backup, archive, and HSM) segment had 44% share with 19% CAGR, shrinking segment share to 34% in 2005.

The Storage Infrastructure (file systems, volume managers and replication) segment ranked next with 37% share with a 30% annual growth rate, growing to a leading segment share of 43% in 2005.

The Enterprise Storage Resource Management (device administration, media and library management, and SRM) was the smallest segment (19% share) and fastest growing (32%), but still smallest segment share of the total 2005 storage software market.

58. How large is the xSP opportunity?

IDC projects spending on managed storage services to grow from \$140 million in 2000 to \$4.8 billion in 2003. Gartner/DataQuest estimates the SSP

opportunity to grow from \$10 million in 1999 to \$476 million in 2001 to \$6.1 billion in 2004. (Enterprise Storage Group forecasts SSP spending to reach \$11.5 billion in 2003.) Gartner forecasts ASP revenues to grow from \$900 million in 1998 to \$23 billion in 2003.

Competitors

This section lists vendors involved in the storage networking and related industries. Familiarity with the vendors participating in each segment can help assess the likely impact of different technologies and understand market dynamics.

59. Who are Fibre Channel SAN vendors?

Host adapter vendors include Adaptec, Agilent, ATTO, Cambex, Compaq, ConnectCom, Delphi Engineering, Genroco, Great River, Emulex, Interphase, JNI, LSI Logic, Qlogic, Sun, SBS, TROIKA, and VMIC. Switch/director vendors include Brocade, McData, Inrange, QLogic(Ancor), Gadzoox, and Vixel. Hub vendors include Vixel, Gadzoox, Emulex, and STK(NSG). Router vendors include Crossroads, Chaparral, Pathlight, ATTO, and StorageTek. Test equipment vendors include Finisar, Ancot, I-TECH, PTI, and Xyratex. Storage vendors include Compaq, EMC, IBM, Hitachi, MTI, Sun (Cobalt), Dot Hill, Eurologic, Quantum (DSS), StorageTek, Advanced Digital Information (ADIC), and Exabyte. SAN training is provided by Infinity I/O, Solution Technology, FSI Consulting, Connectivity Solutions, and product vendors including McDATA, EMC, Exabyte, Legato, and Veritas.

60. Who are Storage Virtualization vendors?

Morgan Keegan categorizes SAN (storage) virtualization products as SAN server, SAN controller (adding switching capabilities), SAN metaserver, or SAN cluster architectures. The first two approaches centralize virtualization intelligence (symmetric) while the last two remove it from the data path (asymmetric). SAN server vendors include DataCore, StorageApps, TrueSAN, and Veritas. SAN controller vendors include StorageTek,

DataDirect, and Dell. SAN metaserver vendors include Compaq, EMC (CrosStor), and StoreAge. SAN cluster vendors include HP (Transoft), Veritas, and StoreAge. Vendors that implement storage virtualization in SAN appliances include Vicom and Crossroads.

61. Who are Gigabit Ethernet vendors?

Network Interface Card (NIC) vendors include 3Com, Alteon, Cabletron, Essential, HP, Intel, Jato, Packet Engines, Silicon Graphics, Solidum, Sun, XaQti, and ZNYX. Switch vendors include 3Com, Acclaim, Alteon, Bay Networks, Cabletron, Compaq, Extreme, Foundry, GigaLabs, HP, Hitachi Cable, IBM, Intel, LANNET, Lucent, NetVantage, ODS, Performance Technologies, Plaintree, Samsung, Siemens, and XLNT. Test equipment vendors include HP, Netcom, Network Associates, XaQti, and Xyratex. Vendors of attached servers include Network Appliance and Silicon Graphics.

62. Who are the leading NAS vendors?

Network Appliance, EMC, Auspex, Compaq, IBM, Procom, Dell, Artecon, ECCS, LSI Logic, Unisys, LAND-5, Raidtec, and ANDATAACO had the largest revenue share in 1999. Snap Appliance (Quantum), Maxtor, Sun (Cobalt), Connex, and StorLogic also offer NAS systems.

63. Who are the leading software vendors in the systems infrastructure market?

IDC ranked the top five vendors, based on 1999 revenues in this segment, as Microsoft (\$8.1 billion), IBM (\$7.7 billion), Computer Associates (\$4.4 billion), HP (\$2.1 billion), and Novell (\$980 million). Other popular vendors participating in this segment include Sun (\$920 million), BMC (\$766 million), Veritas (\$674million), EMC (\$612 million), Compaq (\$498 million), Candle (\$360

million), SCO (\$185 million), Legato (\$183 million), and StorageTek (\$173 million).

64. Who are the leading backup vendors?

IDC ranked vendors according to worldwide revenues for backup software (\$2.546 billion in 1999) as follows: Computer Associates/Sterling (\$763/\$145 million), Veritas (\$389 million), IBM (\$336 million), Legato (\$185 million), HP (\$78 million), Sun (\$39.5 million), CommVault (\$33.5 million), and Compaq (\$25.6 million). All other vendors served less than 1% of the market.

The top three backup platforms in terms of revenues were Unix (\$816 million), Windows (\$737 million) and mainframe (\$670 million). By 2004, backup revenues on Windows (\$2.6 billion) is expected to surpass revenues on Unix (\$1.76 billion) and mainframe (\$782 million). Backup revenues on Linux will grow rapidly (36% CAGR) but remain relatively small, reaching \$137 million in 2004 from \$29 million in 1999.

65. Who are the leading SRM vendors?

IDC ranked vendors according to worldwide 1999 revenues for the \$703 million Storage Resource Management market (in \$ millions): EMC (\$557.5), STK (\$41), Fujitsu (\$26.2), IBM (\$14.3), Ontrack (\$6.2), Sun (\$5.8), Legato (\$5.3), NSI (\$5), HP (\$2.5), Compaq (\$2.4), Veritas (\$2.3), Proginet (\$1.8), NCR (\$1.6), QSTAR Technologies (\$1), and all other (\$30).

The top three SRM platforms in terms of revenues were mainframe (\$232 million), Unix (\$232 million) and Win32 (\$151 million). By 2004, SRM revenues for both Unix (\$516 million) and Win32 (\$430 million) are projected to surpass revenues for mainframe products (\$272 million). SRM revenues on Linux will remain

relatively small, growing to \$43 million by 2004.

66. Who are the leading database vendors?

A May 2000 Dataquest estimated the worldwide database industry will grow from \$8 billion in 1999 to \$12.7 billion by 2004 with Windows and Unix platform revenues nearly equal by then. The market will be driven by Internet-related applications, e-commerce, content management, integrated business intelligence, and new mobile applications. 1999 market share leaders were Oracle (31%), IBM (30%), Microsoft (13%), Informix (4%), Sybase (3%), and others (18%).

67. Who are the leading service providers?

There are many types of service providers impacting the storage networking industry. Storage, Internet, Network, and Application Service Providers are collectively referred to as xSPs. Aggregators partner with other service providers to offer sophisticated solutions with the convenience of one-stop shopping. Storage Service Providers (SSPs) manage data storage infrastructures on a pay-as-you-grow basis. Examples include Storage Networks, ManagedStorage, Storability, WorldStor, StorageWay, Arsenal Digital, Scale Eight, CreekPath, StorageProvider, Centripetal, Comdisco, SunGard, Iron Mountain, Loudcloud, and MangoSoft. SSPs that provide services to ASPs are called ASP Infrastructure Providers (AIPs). Examples include Allegrix, Chapter 2, Esoftglobal, Exenet, Mimecom, and New Moon. Internet Service Providers (ISP) outsource Internet operations and host server co-location. Example ISPs include AT&T Global Network, Concentric Network, DIGEX, EarthLink, Enron, Exodus, Frontier,

GTE Internetworking, Level3, MCI Worldcom UUNet, PSINet, and Verio. Application Service Providers (ASPs) deliver and manage applications and computer services from remote data centers to multiple users via the Internet or a private network. Example ASPs include Applicast, AristaSoft, Breakaway Solutions, Corio, Interliant, NaviSite, Qwest Cyber.Solutions, ReSourcePhoenix.com, USinternetworking, and ViStorm. The ASP Industry Consortium (<http://www.aspindustry.org>) is an international ASP advocacy group. Management Service Providers (MSPs) deliver system management services, rather than business applications, on a recurring fee basis. Examples include 2ndWave, InteQ, and Nuclio. Refer to <http://www.mspassociation.org/>. Services are also provided by system integrators such as Accentre (Anderson Consulting), EDS, and PricewaterhouseCoopers, by software vendors including J.D. Edwards, Microsoft, Oracle, PeopleSoft, SAP, and Siebel Systems, by major technology providers like Cisco, IBM, Intel, and Nortel, and by portals, such as AOL and Yahoo.

68. Who are cluster vendors?

Digital VAX Clusters appeared in 1983 using proprietary interprocessor communication (IPC). In the late 1980s, Data General, HP, IBM, NCR, Pyramid, Sequent, Sun, and Tandem offered single vendor clusters for high availability but were not very scalable. IBM Parallel Sysplex was introduced in 1994 and has gained widespread acceptance but is restricted to IBM mainframe environments. Oracle and Informix introduced database clustering using standard hardware in the mid 1990s. Microsoft, Novell, Compaq, SCO, CLAM (now Availant), Veritas, and Legato also provide lower cost clustering products.

Internet technology has allowed server farms to be configured for availability by balancing traffic rather than by

clustering servers but may not meet performance expectations or be suitable for transactions with state, e.g., shopping carts. IP load balancing software vendors include Cisco, Microsoft, PolyServe, Resonate, Platform Computing, and Bright Tiger.

A March 2000 high availability study from D.H. Brown ranked UNIX cluster solutions as follows: Compaq TruCluster Server for Tru64 UNIX, IBM HACMP for AIX, HP MC/ServiceGuard for HP-UX, EMC DG/UX Clusters, Sun Clusters for Solaris, and Sequent ptx/Clusters for Dynix/ptx. Other products include Compaq OpenVMS Cluster and AT&T/NCR Lifekeeper (Unix/NT).

69. Which vendors are members of the InfiniBand Trade Association?

Steering Committee member companies are Compaq, Dell, Hewlett-Packard, IBM, Intel, Microsoft and Sun Microsystems. Sponsoring member companies are 3Com, Adaptec, Agilent, Brocade, Cisco, EMC, Fujitsu-Siemens, Hitachi, Lucent, NEC and Nortel Networks. Refer to <http://www.infinibandta.org> for a complete list of members, currently listed at over 200 vendors.

70. Which vendors are involved in DAFS?

Emulex, Intel, Network Appliance, Troika, and Veritas hold leadership positions in the DAFScollaborative. Startups like Broadband Storage and several universities are also active. Over 77 software and hardware vendors have signed the contribution agreement and many others have been actively monitoring the development of the specification. Refer to <http://www.dafscollaborative.org>.

71. What is optical networking and which vendors compete in this market?

ONI Systems categorizes optical networking products into four broad categories:

Long-haul transmission focuses on moving data and voice traffic across several hundred kilometers. Vendors of products in this category include Ciena, Corvis, Lucent, and Nortel.

Wide-area optical cross-connects (also called bandwidth managers) focus on switching signals between the long-haul segments. Vendors of products in this category include Lightera (Ciena), Lucent, Monterey Networks (Cisco), Sycamore Networks, and Tellium.

Metropolitan Area Networks (MAN) transmit signals between long-haul segments and enterprises with a focus on adding value in areas such as flexibility of connections and speeds, services, protection, and manageability. Vendors of products in this category include Ciena, Lucent, Nortel, ONI, and Tellabs.

Access networks are concerned with access to data and voice signals by end users such as homes, offices, and factories. Vendors of products in this category include Amber, Appian, Atmosphere, Chromatis (Lucent), Cyras, Geysler, Kestrel, LuxN, MAYAN, Nortel, ONI, Quantum Bridge, and Terawave.

72. Which vendors provide web caching?

The most widely used *software-only caching* product is the open source Squid project. Commercial products are provided by Inktomi (Traffic Server, based on Squid), Microsoft (Proxy Server), AOL/Netscape (Proxy Server 3.5), Novell (Internet Caching System), and WebSpective (WebSpective). *Prepackaged servers* usually bundle Squid on a UNIX variant. These products are available from Sun/Cobalt (CacheQube), Eolian

(InfoStorm), and PacketStorm (WebSpeed). Another group of prepackaged servers bundle Inktomi's Traffic Server with a server plus network hardware. These products are available from Nortel/Alteon, Compaq, and Foundry. *Specialty appliances*, typically run operating systems, file systems, and applications specifically designed for caching, and provide additional functions. Example vendors include CacheFlow (CacheFlow), Cisco (Cache Engine), Infolibria (DynaCache), Network Appliance (NetCache), and Lucent (WebCache), and Entera (TeraNode). Vendors in the *caching service provider* space include Akamai (FreeFlow), Digital Island (GlobalCache, which uses software from Inktomi and WebSpective), Mirror Image (Central Cache Connection), Sandpiper Networks (Footprint), and Skycache (uses satellites).



Troika Networks, Incorporated
2829 Townsgate Road
Westlake Village, CA 91361-3017
<http://www.TroikaNetworks.com>
Tel: +1 805.371.1377