MODULE III

IP SAN and FCoE

Traditional SAN environments ***allow block I/O over Fibre Channel***. IP offers easier management and better interoperability. When block I/O is run over IP, the existing network infrastructure can be leveraged, which is more economical than investing in new SAN hardware andsoftware.

Many long-distance, disaster recovery (DR) solutions are already leveraging IP-based networks. In addition, many robust and mature security options are now available for IP networks. With the advent of block storage technology that leverages IP networks (the result is often referred to as IP SAN), organizations can extend the geographical reach of their storage infrastructure. IP SAN technologies can be used in a variety of situations.

Primary protocols that leverage IP as the ***transport mechanism are iSCSI and Fibre Channel over IP (FCIP). iSCSI is the host-based encapsulation of SCSI I/O over IP using an Ethernet NIC card or an iSCSI HBA in the host.***

2.1 iSCSI

- Internet Small Computer Systems Interface (iSCSI) is an IP based protocol that establishes and manages connections between host and storage over IP, as shown in Fig2.21.
- iSCSI encapsulates SCSI commands and data into an IP packet and transports them using TCP/IP.
- iSCSI is widely adopted for connecting servers to storage because it is relatively inexpensive and easy to implement, especially in environments in which an FC SAN does not exist.

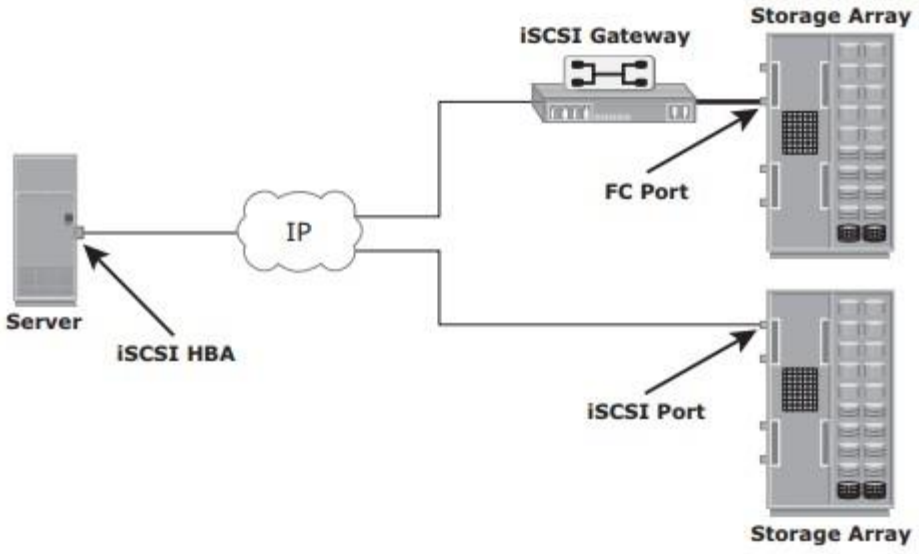


Fig 2.21: iSCSI implementation

2.9.1 Components ofiSCSI 7) About ISCI and components of ISCI

- An initiator (host), target (storage or iSCSI gateway), and an IP-based network are the key iSCSIcomponents.
- If an iSCSI-capable storage array is deployed, then a host with the iSCSI initiator can directly communicate with the storage array over an IPnetwork.
- However, in an implementation that uses an existing FC array for iSCSI communication, an iSCSI gateway isused.
- These devices perform the translation of IP packets to FC frames and vice versa, thereby bridging the connectivity between the IP and FCenvironments.

2.9.2 iSCSI Host Connectivity

The three iSCSI host connectivity options are:

- A standard NIC with software iSCSIinitiator,
- a TCP offload engine (TOE) NIC with software iSCSIinitiator,
- an iSCSIHBA

- The function of the iSCSI initiator is to route the SCSI commands over an IPnetwork.
- A **standard NIC with a software iSCSI** initiator is the simplest and least expensive connectivity option. It is easy to implement because most servers come with at least one, and in many cases two, embedded NICs. It requires only a software initiator for iSCSI functionality. Because NICs provide standard IP function, encapsulation of SCSI into IP packets and decapsulation are carried out by the host CPU. This places additional overhead on the host CPU. If a standard NIC is used in heavy I/O load situations, the host CPU might become a bottleneck. TOE NIC helps reduce thisburden.
- A **TOE NIC** offloads TCP management functions from the host and leaves only the iSCSI functionality to the host processor. The host passes the iSCSI information to the TOE card, and the TOE card sends the information to the destination using TCP/IP. Although this solution improves performance, the iSCSI functionality is still handled by a software initiator that requires host CPUcycles.
- An **iSCSI HBA** is capable of providing performance benefits because it offloads the entire iSCSI and TCP/IP processing from the host processor. The use of an iSCSI HBA is also the simplest way to boot hosts from a SAN environment via iSCSI. If there is no iSCSI HBA, modifications must be made to the basic operating system to boot a host from the storage devices because the NIC needs to obtain an IP address before the operating system loads. The functionality of an iSCSI HBA is similar to the functionality of an FCHBA.

8) About different ISCSI topologies

2.9.3 iSCSI Topologies

- Two topologies of iSCSI implementations are **native** and**bridged**.
- Native topology does not have FCcomponents.
- The initiators may be either directly attached to targets or connected through the IPnetwork.
- Bridged topology enables the coexistence of FC with IP by providing iSCSI-to-FC bridging functionality.
- For example, the initiators can exist in an IP environment while the storage remains in anFC environment.

Native iSCSI Connectivity

- FC components are not required for iSCSI connectivity if an iSCSI-enabled array is deployed.
- In Fig 2.22(a), the array has one or more iSCSI ports configured with an IP address and is connected to a standard Ethernetswitch.
- After an initiator is logged on to the network, it can access the available LUNs on the storage array.
- A single array port can service multiple hosts or initiators as long as the array port can handle the amount of storage traffic that the hostsgenerate.

Bridged iSCSI Connectivity

- A bridged iSCSI implementation includes FC components in itsconfiguration.
- Fig 2.22(b), illustrates iSCSI host connectivity to an FC storage array. In this case, the array does not have any iSCSI ports. Therefore, an external device, called a gateway or a multiprotocol router, must be used to facilitate the communication between the iSCSI host and FCstorage.
- The gateway converts IP packets to FC frames and viceversa.
- The bridge devices contain both FC and Ethernet ports to facilitate the communication between the FC and IPenvironments.
- In a bridged iSCSI implementation, the iSCSI initiator is configured with the gateway's IP address as its targetdestination.
- On the other side, the gateway is configured as an FC initiator to the storagearray.
- **Combining FC and Native iSCSI Connectivity:** The most common topologyis a

combination of FC and native iSCSI. Typically, a storage array comes with both FC and iSCSI ports that enable iSCSI and FC connectivity in the same environment, as shown in Fig 2.22(c).

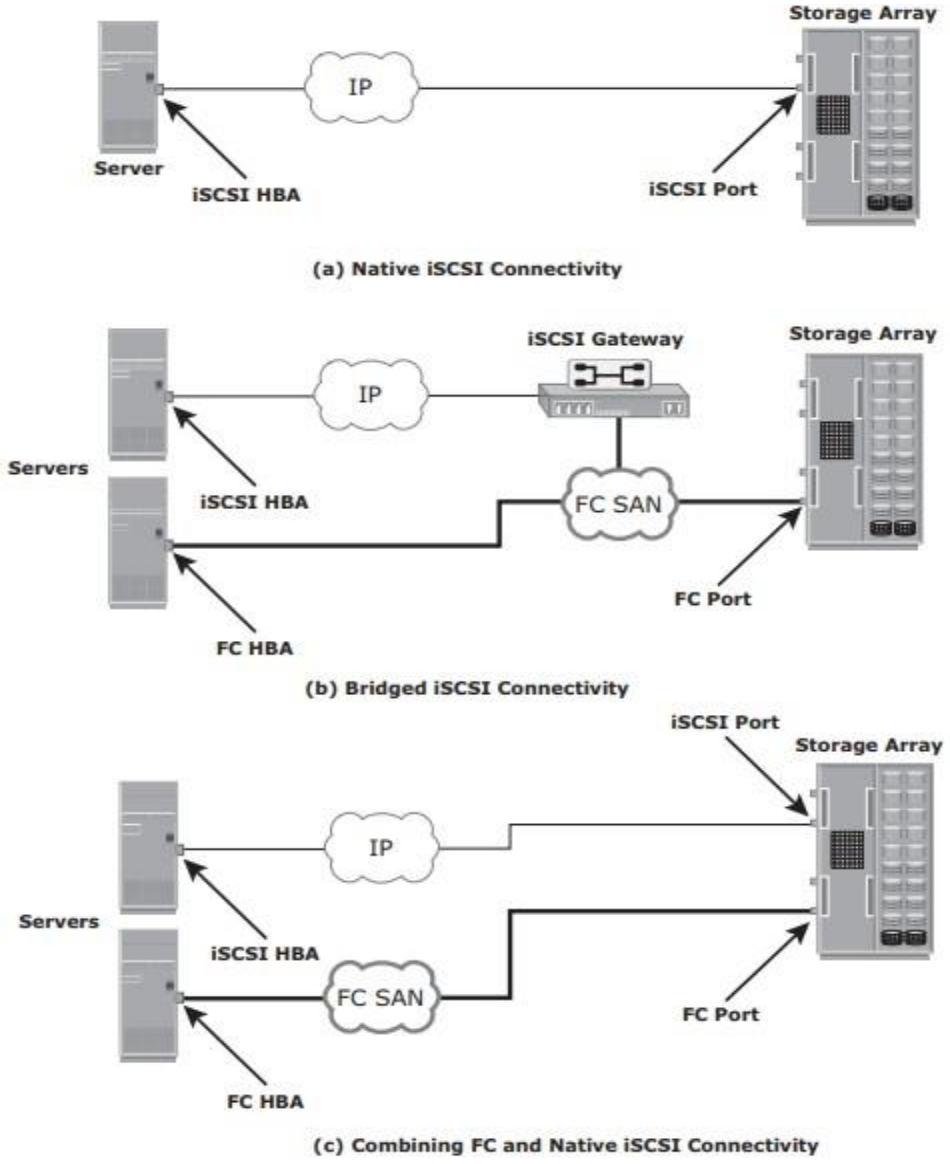


Fig 2.22 : iSCSI Topologies

9) About ISCSI and ISCSI protocol stack with a neat diagram

2.9.4 iSCSI ProtocolStack

- Fig 2.23 displays a model of the iSCSI protocol layers and depicts the encapsulation order of the SCSI commands for their delivery through a physicalcarrier.

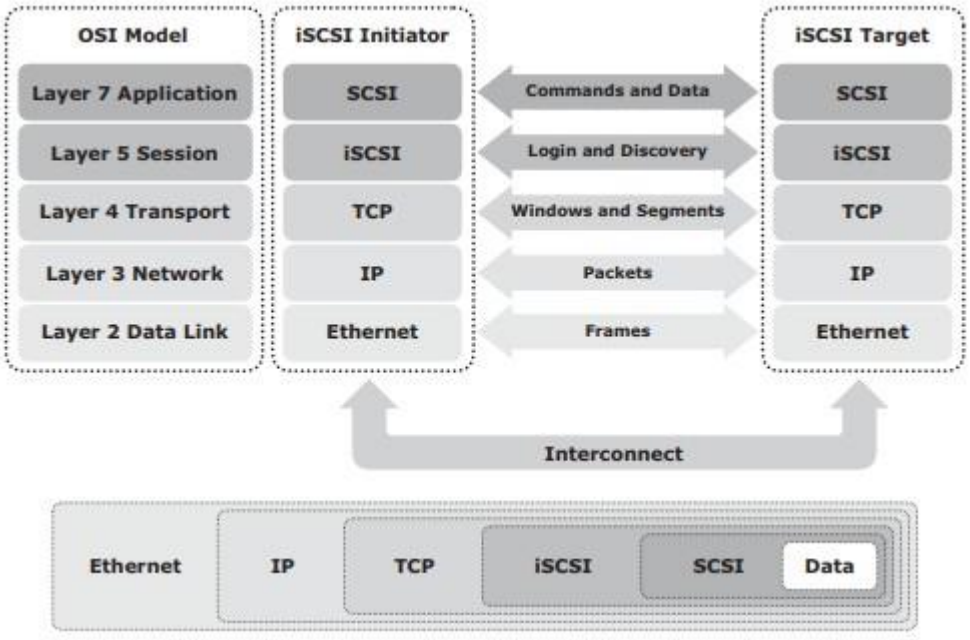


Fig 2.23: iSCSI protocol stack

- SCSI is the command protocol that works at the application layer of the Open System Interconnection (OSI)model.
- The initiators and targets use SCSI commands and responses to talk to eachother.
- The SCSI command descriptor blocks, data, and status messages are encapsulated into TCP/IP and transmitted across the network between the initiators andtargets.
- iSCSI is the session-layer protocol that initiates a reliable session between devices that recognize SCSI commands andTCP/IP.
- The iSCSI session-layer interface is responsible for handling login, authentication, target discovery, and sessionmanagement.
- TCP is used with iSCSI at the transport layer to provide reliabletransmission.

- TCP controls message flow, windowing, error recovery, andretransmission.
- It relies upon the network layer of the OSI model to provide global addressing and connectivity.
- The Layer 2 protocols at the data link layer of this model enable node-to-node communication through a physical network.

2.9.5 iSCSI PDU

- A *protocol data unit* (PDU) is the basic "information unit" in the iSCSIenvironment.
- The iSCSI initiators and targets communicate with each other using iSCSI PDUs. This communication includes establishing iSCSI connections and iSCSI sessions, performing iSCSI discovery, sending SCSI commands and data, and receiving SCSIstatus.
- All iSCSI PDUs contain one or more header segments followed by zero or more data segments.
- The PDU is then encapsulated into an IP packet to facilitate thetransport.
- A PDU includes the components shown in Fig2.23.
- The IP header provides packet-routing information to move the packet across anetwork.
- The TCP header contains the information required to guarantee the packet delivery to the target.
- The iSCSI header (basic header segment) describes how to extract SCSI commands and data for the target. iSCSI adds an optional CRC, known as the *digest*, to ensure datagram integrity. This is in addition to TCP checksum and EthernetCRC.
- The header and the data digests are optionally used in the PDU to validate integrity and data placement.

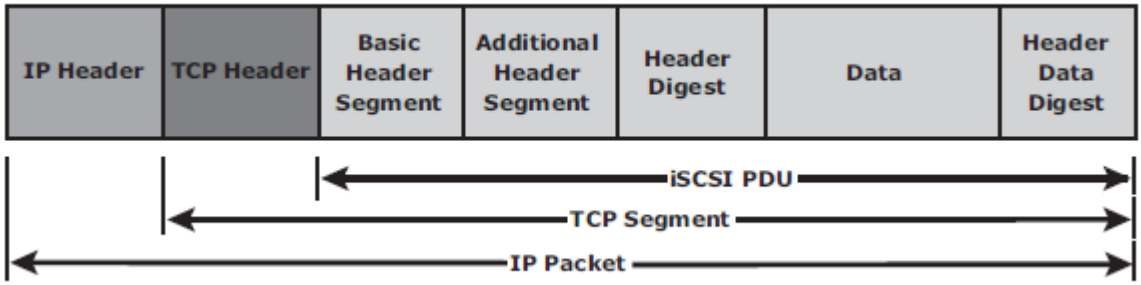


Fig 2.23 iSCSI PDU encapsulated in an IP packet

- The iSCSI header describes how to extract SCSI commands and data for the target. iSCSI adds an optional CRC, known as the *digest*, beyond the TCP checksum and Ethernet CRC to ensure datagram integrity.
- The header and the data digests are optionally used in the PDU to validate integrity, data placement, and correct operation As shown in Figure 8-7.

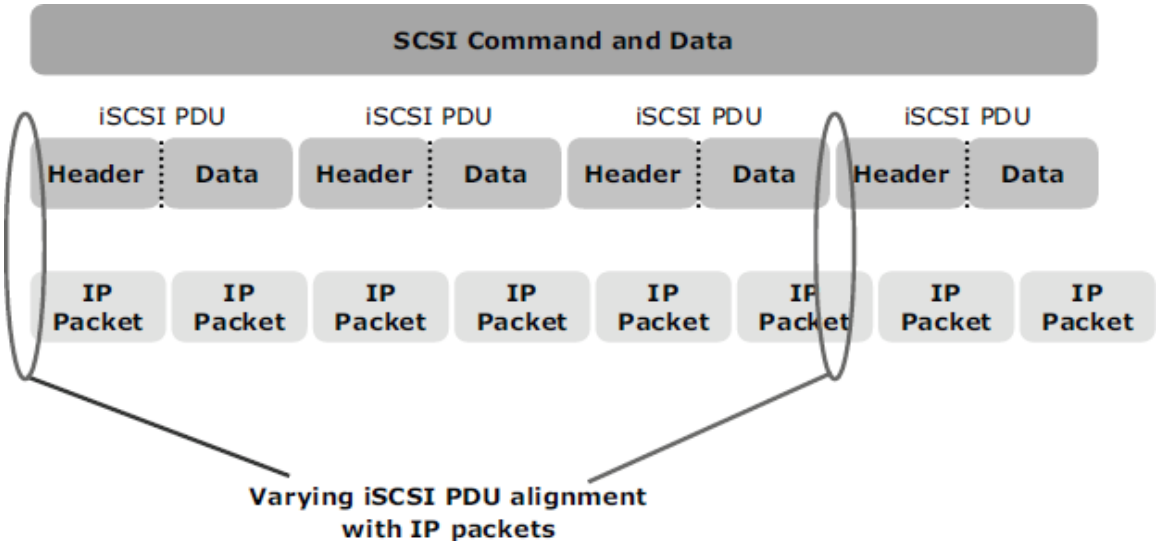


Figure 8-7: Alignment of iSCSI PDUs with IP packets

Each iSCSI PDU does not correspond in a 1:1 relationship with an IP packest.

Depending on its size, an iSCSI PDU can span an IP packet or even coexist with another PDU in the same packet.

Therefore, each IP packet and Ethernet frame can be used more efficiently because fewer packets and frames are required to transmit the SCSI information.

2.9.6 iSCSI Discovery

- An initiator must discover the location of its targets on the network and the names of the targets available to it before it can establish asession.
- This discovery can take place in twoways:
  - **SendTargetsdiscovery**
  - **internet Storage Name Service(iSNS).**
- In *SendTargets discovery*, the initiator is manually configured with the target's network portal to establish a discovery session. The initiator issues the SendTargets command, and the target network portal responds with the names and addresses of the targets available to thehost.
- iSNS (Fig 2.24) enables automatic discovery of iSCSI devices on an IP network. The initiators and targets can be configured to automatically register themselves with the iSNS server. Whenever an initiator wants to now the targets that it can access, it can query the iSNS server for a list of availabletargets.
- The discovery can also take place by using service location protocol (SLP). However, this is less commonly used than SendTargets discovery andiSNS.

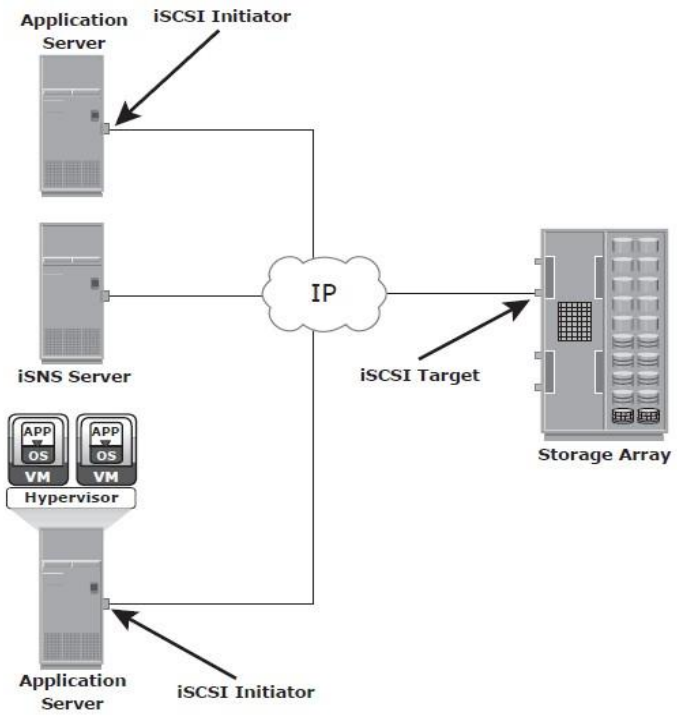


Fig 2.24 : Discovery using iSNS

2.9.7 iSCSI Names

- A unique worldwide iSCSI identifier, known as an *iSCSI name*, is used to identify the initiators and targets within an iSCSI network to facilitatecommunication.
- The unique identifier can be a combination of the names of the department, application, or manufacturer, serial number, asset number, or any tag that can be used to recognize and manage thedevices.
- Following are two types of iSCSI names commonlyused:
    - **iSCSI Qualified Name(IQN):**
    - **Extended Unique Identifier(EUI)**
- **iSCSI Qualified Name (IQN):** An organization must own a registered domain name to generate iSCSI Qualified Names. This domain name does not need to be active or resolve to an address. It just needs to be reserved to prevent other organizations from using the same domain name to generate iSCSI names. A date is included in the name to avoid potential conflicts caused by the transfer of domainnames.

  An example of an IQN is iqn.2008-02.com.example:*optional_string.* The *optional_string* provides a serial number, an asset number, or any otherdevice identifiers.
- **Extended Unique Identifier (EUI):** An EUI is a globally unique identifier based on the IEEE EUI-64 naming standard. An EUI is composed of the eui prefix followed by a 16-character hexadecimalname, suchaseui.0300732A32598D26.
- In either format, the allowed special characters are dots, dashes, and blankspaces.

2.9.8 iSCSI Session

- An iSCSI session is established between an initiator and a target, as shown in Fig2.25.
- A session is identified by a session ID (SSID), which includes part of an initiator ID and a target ID.
- The session can be intended for one of thefollowing:
    - The discovery of the available targets by the initiators and the location of a specific target on anetwork
    - The normal operation of iSCSI (transferring data between initiators andtargets)

- There might be one or more TCP connections within each session. Each TCP connection within the session has a unique connection ID(CID).
- An iSCSI session is established via the iSCSI login process. The login process is started when the initiator establishes a TCP connection with the required target either via the well-known port 3260 or a specified targetport.
- During the login phase, the initiator and the target authenticate each other and negotiate on variousparameters.
- After the login phase is successfully completed, the iSCSI session enters the full-feature phase for normal SCSI transactions. In this phase, the initiator may send SCSI commands and data to the various LUNs on thetarget.
- The final phase of the iSCSI session is the connection termination phase, which is referred to as the logoutprocedure.
- The initiator is responsible for commencing the logout procedure; however, the target may also prompt termination by sending an iSCSI message, indicating the occurrence of an internal error condition.
- After the logout request is sent from the initiator and accepted by the target, no further request and response can be sent on that connection.

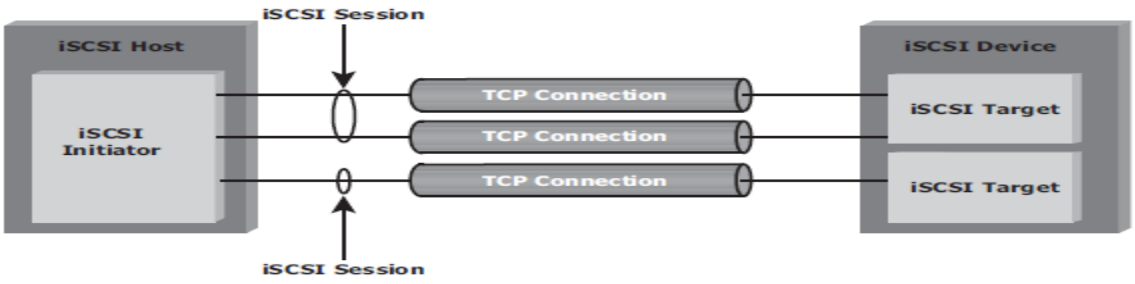


Fig 2.25 iSCSI session

iSCSICommandSequencing

- The iSCSI communication between the initiators and targets is based on the request-response commandsequences.A command sequence may generate multiplePDUs.
- A ***command sequence number*** **(CmdSN**) within an iSCSI session is used for numbering all initiator-to-target command PDUs belonging to thesession.
- This number ensures that every command is delivered in the same order in which it is transmitted, regardless of the TCP connection that carries the command in thesession.
- Command sequencing begins with the first login command, and the CmdSN is incremented by one for each subsequentcommand.
- The iSCSI target layer is responsible for delivering the commands to the SCSI layer in the order of theirCmdSN.
- Similar to command numbering, a ***status sequence number*** **(StatSN)** is used to sequentially number status responses, as shown in Fig2.26.
- These unique numbers are established at the level of the TCPconnection.
- A target sends ***request-to-transfer*** **(R2T)** PDUs to the initiator when it is ready to acceptdata.
- A *data sequence number* (DataSN) is used to ensure in-order delivery of data within the same command.
- The DataSN and R2TSN are used to sequence data PDUs and R2Ts,respectively.

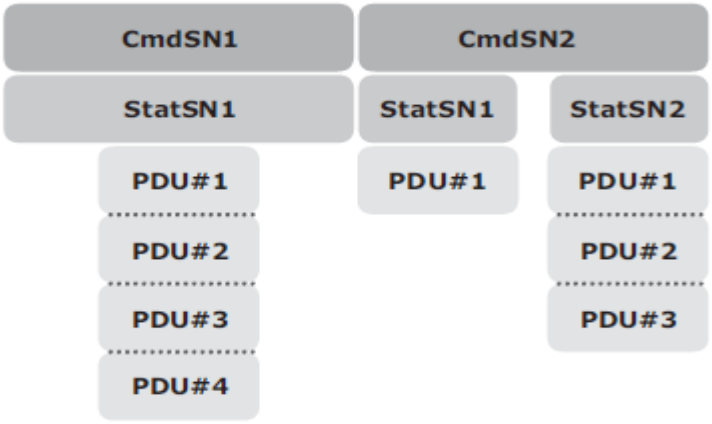


Fig 2.26 Command and status sequence number

10) About FCIP and fiber channel protocol stack with a neat diagram

2.2 FCIP (Fibre channel overIP)

- FCIP is a IP-based protocol that is used to connect distributed FC-SANislands.
- Creates virtual FC links over existing IP network that is used to transport FC data between different FCSANS.
- It encapsulates FC frames into IPpacket.
- It provides disaster recoverysolution.

2.10.1 FCIP ProtocolStack

- The FCIP protocol stack is shown in Fig 2.27. Applications generate SCSI commands and data, which are processed by various layers of the protocolstack.
- The upper layer protocol SCSI includes the SCSI driver program that executes the read-and-writecommands.
- Below the SCSI layer is the Fibre Channel Protocol (FCP) layer, which is simply a Fibre Channel frame whose payload isSCSI.
- The FCP layer rides on top of the Fibre Channel transport layer. This enables the FC frames to run natively within a SAN fabric environment. In addition, the FC frames can be encapsulated into the IP packet and sent to a remote SAN over theIP.

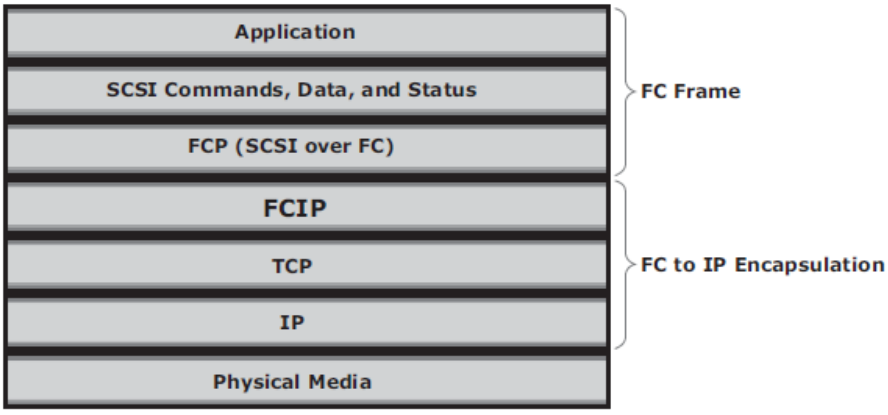


Fig 2.27 FCIP protocol stack

- The FCIP layer encapsulates the Fibre Channel frames onto the IP payload and passes them to the TCP layer (see Fig 2.28). TCP and IP are used for transporting the encapsulated information across Ethernet, wireless, or other media that support the TCP/IPtraffic.
- Encapsulation of FC frame into an IP packet could cause the IP packet to be fragmented when the data link cannot support the maximum transmission unit (MTU) size of an IPpacket.
- When an IP packet is fragmented, the required parts of the header must be copied by all fragments.
- When a TCP packet is segmented, normal TCP operations are responsible for receiving and re-sequencing the data prior to passing it on to the FC processing portion of thedevice.

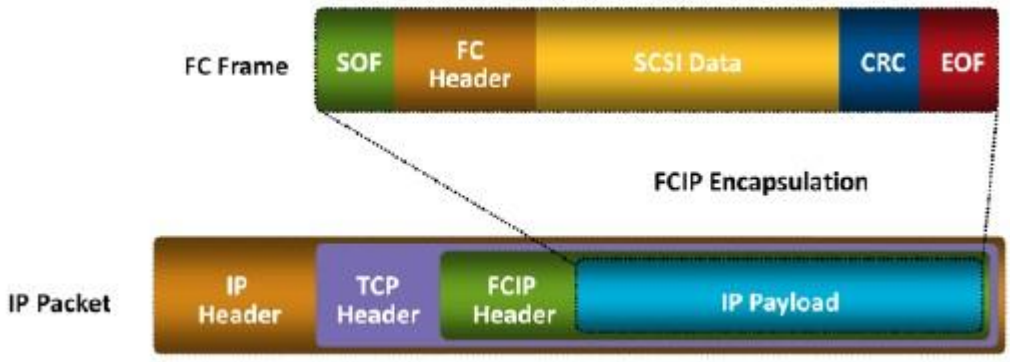


Fig 2.28 FCIP encapsulation

2.10.2 FCIPTopology

- In an FCIP environment, an FCIP gateway is connected to each fabric via a standard FC connection (Fig2.29).
- The FCIP gateway at one end of the IP network encapsulates the FC frames into IPpackets.
- The gateway at the other end removes the IP wrapper and sends the FC data to the layer 2fabric.
- The fabric treats these gateways as layer 2 fabricswitches.
- An IP address is assigned to the port on the gateway, which is connected to an IP network. After theIPconnectivityisestablished,thenodesinthetwoindependentfabricscancommunicate

with each other.

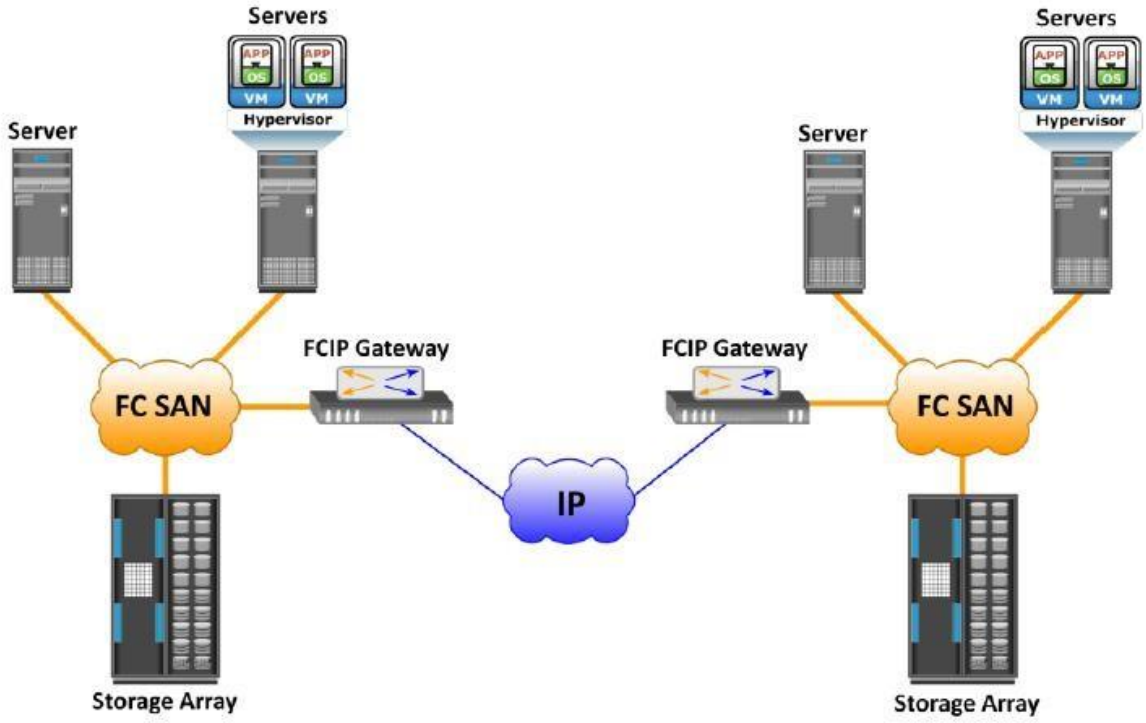


Fig 2.29 FCIP topology

FCIP Performance andSecurity

- Performance, reliability, and security should always be taken into consideration when implementing storage solutions.
- The implementation of FCIP is also subject to the same consideration. From the perspective of performance, multiple paths to multiple FCIP gateways from different switches in the layer 2 fabric eliminates single points of failure and provides increased bandwidth. In a scenario of extended distance, the IP network may be a bottleneck if sufficient bandwidth is notavailable.
- In addition, because FCIP creates a unified fabric, disruption in the underlying IP network can cause instabilities in the SAN environment. These include a segmented fabric, excessive RSCNs, and hosttimeouts.
- The vendors of FC switches have recognized some of the drawbacks related to FCIP and have implemented features to provide additional stability, such as the capability to segregate FCIP traffic into a separate virtual fabric.
- Security is also a consideration in an FCIP solution because the data is transmitted over public IP channels.
- Various security options are available to protect the data based on the router's support. IPSec is one such security measure that can be implemented in the FCIP environment

11) About NAS and the benefits of NAS

.

NETWORK ATTACHED STORAGE(NAS)

- File sharing, as the nameimplies,enables users to share files withotherusers.
- Traditionalmethodsoffilesharinginvolvecopyingfilestoportablemediasuchasfloppy diskette, CD, DVD, or USB drivesanddeliveringthemtootherusers withwhomitisbeing shared.
- However, this approach is not suitable in an enterprise environment in which a large number of users at different locations need access to common files.
- Network-based file sharing provides the flexibility to share files over long distances among a large number of users. File servers use client- server technology to enable file sharing over a network.
- To address the tremendous growth of file data in enterprise environments, organizations have been deploying large numbers of fileservers.
- These servers are either connected to direct-attached storage (DAS) or storage area network (SAN)-attached storage.
- This has resulted in the proliferation of islands of over-utilized and under-utilized file servers and storage. Inaddition, such environments have poor scalability, higher management cost, and greater complexity.
- *Network-attached storage* (NAS) emerged as a solution to these challenges.

Benefits of NAS

NAS offers the following benefits:

- Comprehensive access to information: Enables efficient file sharing and supports many-to-one and one-to-many configurations. The many-to-one configuration enables a NAS device to serve many clients simultaneously. The one-to-many configuration enables one client to connect with many NAS devices simultaneously.
- Improved efficiency: NAS delivers better performance compared to a general-purpose file server because NAS uses an operating system specialized for file serving.
- Improved flexibility: Compatible with clients on both UNIX and Windows platforms using industry-standard protocols. NAS is flexible and can serve requests from different types of clients from the same source.
- Centralized storage: Centralizes data storage to minimize data duplication on client workstations, and ensure greater data protection

- Simplified management: Provides a centralized console that makes it possible to manage file systems efficiently
- Scalability: Scales well with different utilization profiles and types of business applications because of the high-performance and low-latency design
- High availability: Offers efficient replication and recovery options, enabling high data availability. NAS uses redundant components that provide maximum connectivity options. A NAS device supports clustering technology for failover.
- Security: Ensures security, user authentication, and file locking with industry-standard security schemas
- Low cost: NAS uses commonly available and inexpensive Ethernet components.
- Ease of deployment: Configuration at the client is minimal, because the clients have required NAS connection software built in.

File Sharing Environment

- File sharing enables users to share files with otherusers
- In file-sharing environment, the creator or owner of a file determines the type of access to be given to other users and controls changes to the file.
- When multiple access a shared file at the same time, a locking scheme is required to maintain data integrity and also make this sharing possible. This is taken care by file-sharingenvironment.
- Examples of file sharingmethods:
    - File Transfer Protocol(FTP)
    - Distributed File System(DFS)
    - Network File System (NFS) and Common Internet File System(CIFS)
    - Peer-to-Peer(P2P)

What is NAS?

- NAS is an IP based dedicated, high-performance file sharing and storagedevice.
- Enables NAS clients to share files over an IPnetwork.

- Uses network and file-sharing protocols to provide access to the filedata.
- Ex: Common Internet File System (CIFS) and Network File System(NFS).
- Enables both UNIX and Microsoft Windows users to share the same dataseamlessly.
- NAS device uses its own operating system and integrated hardware and software components to meet specific file-serviceneeds.
- Its operating system is optimized for file I/O which performs better than a general-purpose server.
- A NAS device can serve more clients than general-purpose servers and provide the benefit of server consolidation.

2.12.1 Components ofNAS

- NAS device has *two* key components (as shown in Fig 2.33): **NAS head** and**storage**.
- In some NAS implementations, the storage could be external to the NAS device and shared with otherhosts.
- NAS head includes the followingcomponents:
    - CPU andmemory
    - One or more network interface cards (NICs), which provide connectivity to the client network.
    - An optimized operating system for managing the NAS functionality. It translates file-level requests into block-storage requests and further converts the data supplied at the block level to filedata
    - NFS, CIFS, and other protocols for filesharing
    - Industry-standard storage protocols and ports to connect and manage physical disk resources

- The NAS environment includes clients accessing a NAS device over an IP network using file-sharingprotocols.

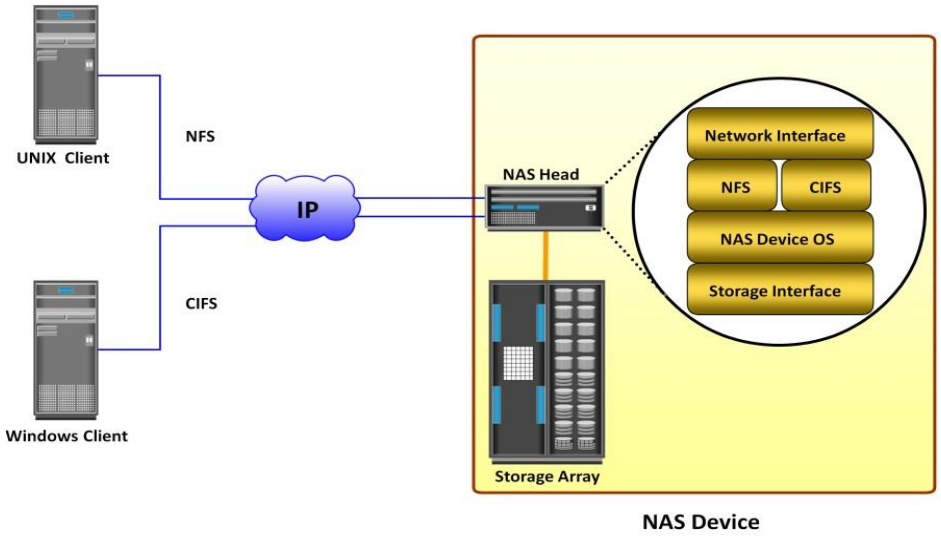


Fig 2.33 Components of NAS

2.12.2 NAS I/OOperation

- NAS provides *file-level data access* to its clients. File I/O is a high-level request that specifies the file to beaccessed.
- Eg: a client may request a file by specifying its name, location, or other attributes. The NAS operating system keeps track of the location of files on the disk volume and converts client file I/O into block-level I/O to retrievedata.
- The process of handling I/Os in a NAS environment is asfollows:

1. The requestor (client) packages an I/O request into TCP/IP and forwards it through the network stack. The NAS device receives this request from thenetwork.
2. TheNASdeviceconvertstheI/Orequestintoanappropriatephysicalstoragerequest, which is a block-level I/O, and then performs the operation on the physical storage.
3. When the NAS device receives data from the storage, it processes and repackages the data into an appropriate file protocolresponse.
4. The NAS device packages this response into TCP/IP again and forwards it to the client through thenetwork.

➢ Fig 2.34 illustrates the NAS I/O operation

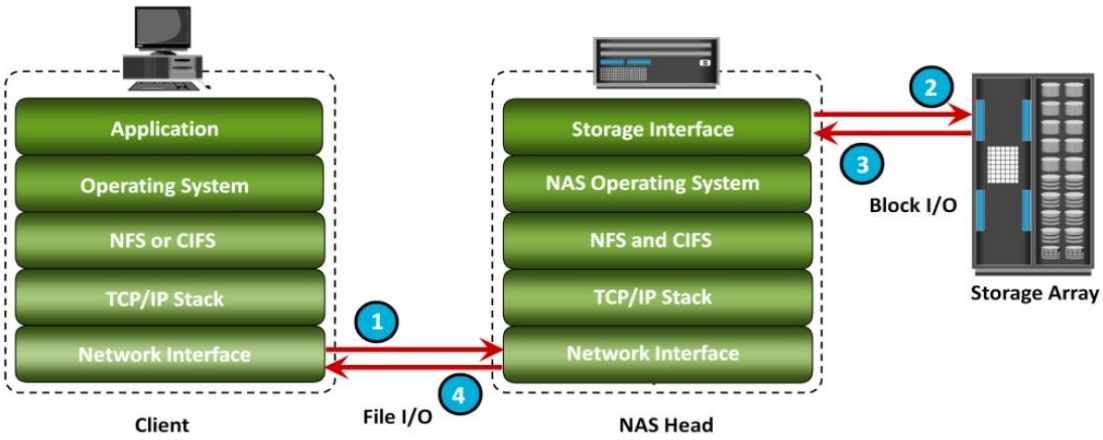


Fig 2.34 NAS I/O Operation

13) About the NAS implementations in detail(Unified NAS connectivity and NAS Gate Way Connectivity)

NAS Implementations

- Three common NAS implementations are unified, gateway, and scale-out. The unified NAS consolidates NAS-based and SAN-based data access within a unified storage platform and provides a unified management interface for managing both the environments.
- In a gateway implementation, the NAS device uses external storage to store and retrieve data, and unlike unified storage, there are separate administrative tasks for the NAS device and storage.
- The scale-out NAS implementation pools multiple nodes together in a cluster. A node may consist of either the NAS head or storage or both. The cluster performs the NAS operation as a single entity.

Unified NAS

- Unified NAS performs file serving and storing of file data, along with providing access to block-level data.
- It supports both CIFS and NFS protocols for file access and iSCSI and FC protocols for block level access.
- Due to consolidation of NAS-based and SAN-based access on a single storage platform, unified NAS reduces an organization's infrastructure and management costs.
- A unified NAS contains one or more NAS heads and storage in a single system. NAS heads are connected to the storage controllers (SCs), which provide access to the storage.
- These storage controllers also provide connectivity to iSCSI and FC hosts. The storage may consist of different drive types, such as SAS, ATA, FC, and flash drives, to meet different workload requirements.

Unified NAS Connectivity

- Each NAS head in a unified NAS has front-end Ethernet ports, which connect to the IP network.
- The front-end ports provide connectivity to the clients and service the file I/O requests. Each NAS head

has back-end ports, to provide connectivity to the storage controllers.

- iSCSI and FC ports on a storage controller enable hosts to access the storage directly or through a storage network at the block level. Figure 7-5 illustrates an example of unified NAS connectivity.

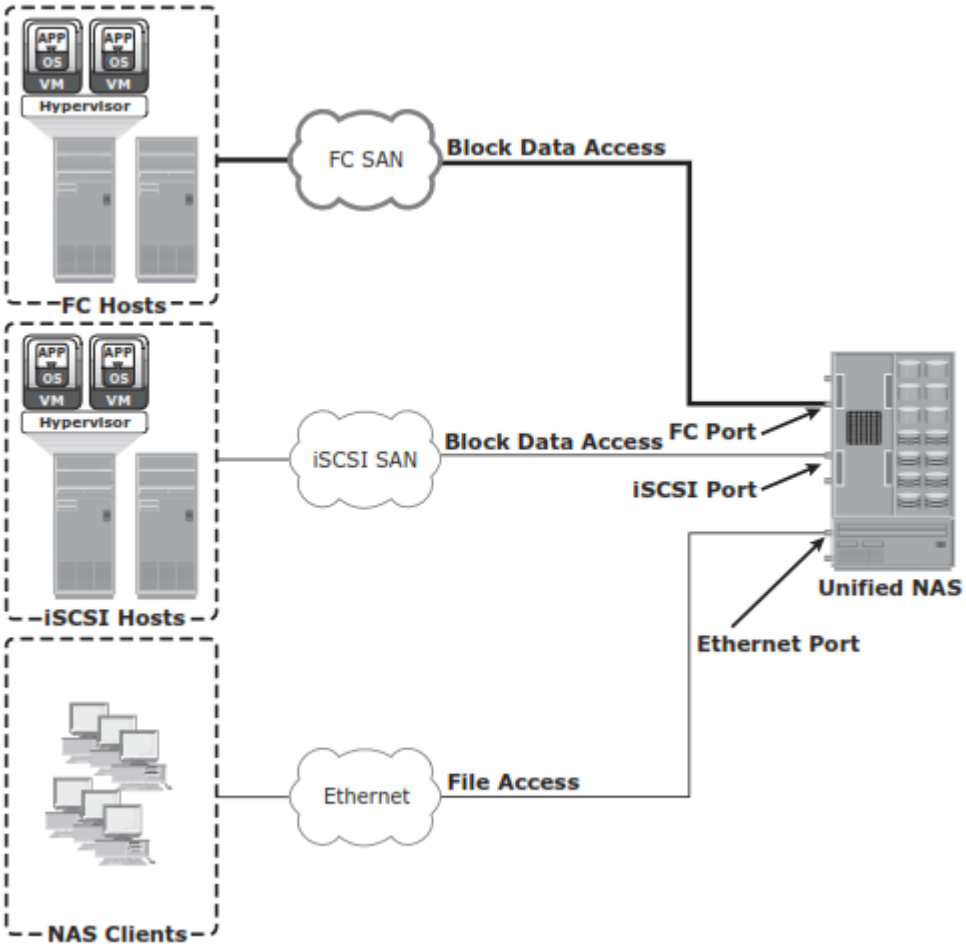


Figure 7-5: Unified NAS connectivity

Gateway NAS

- A gateway NAS device consists of one or more NAS heads and uses external and independently managed storage.
- Similar to unified NAS, the storage is shared with other applications that use block-level I/O. Management func- tions in this type of solution are more complex than those in a unified NAS environment because there are separate administrative tasks for the NAS head and the storage.
- A gateway solution can use the FC infrastructure, such as switches and directors for accessing SAN-attached storage arrays or direct- attached storage arrays.
- The gateway NAS is more scalable compared to unified NAS because NAS heads and storage arrays can be independently scaled up when required.
- For example, NAS heads can be added to scale up the NAS device performance. When the storage limit is reached, it can scale up, adding capacity on the SAN, independent of NAS heads.
- Similar to a unified NAS, a gateway NAS also enables high utilization of storage capacity by sharing it with the SAN environment.

Gateway NAS Connectivity

- In a gateway solution, the front-end connectivity is similar to that in a unified storage solution.
- Communication between the NAS gateway and the storage system in a gateway solution is achieved through a traditional FC SAN.
- To deploy a gateway NAS solution, factors, such as multiple paths for data, redundantfabrics, and load distribution, must be considered. Figure 7-6 illustrates an example of gateway NAS connectivity.

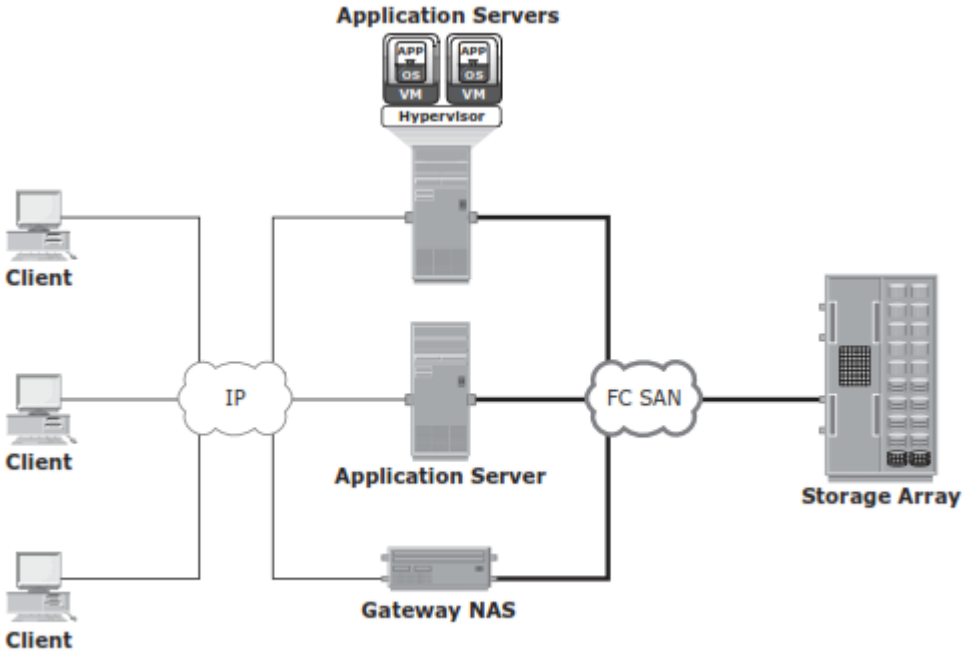


Figure 7-6: Gateway NAS connectivity

- Implementation of both unified and gateway solutions requires analysis of the SAN environment.
- This analysis is required to determine the feasibility of combining the NAS workload with the SAN workload.
- Analyze the SAN to determine whether the workload is primarily read or write, and if it is random or sequential. Also determine the predominant I/O size in use. Typically, NAS workloads are random with small I/O sizes.
- Introducing sequential workload with random workloads can be disruptive to the sequential workload.
- Therefore, it is recommended to separate the NAS and SAN disks. Also, determine whether the NAS workload performs adequately with the configured cache in the storage system.

Scale-Out NAS

- Both unified and gateway NAS implementations provide the capability to scale- up their resources based on data growth and rise in performance requirements.
- Scaling up these NAS devices involves adding CPUs, memory, and storage to the NAS device. Scalability is limited by the capacity of the NAS device to house and use additional NAS heads and storage.
- Scale-out NAS enables grouping multiple nodes together to construct a clustered NAS system.
- A scale-out NAS provides the capability to scale its resources by simply adding nodes to a

clustered NAS architecture. The cluster works as a single NAS device and is managed centrally.

- Nodes can be added to the cluster, when more performance or more capacity is needed, without causing any downtime.
- Scale-out NAS provides the flexibility to use many nodes of moderate performance and availability characteristics to produce a total system that has better aggregate performance and availability.
- It also provides ease of use, low cost, and theoretically unlimited scalability.
- Scale-out NAS creates a single file system that runs on all nodes in the cluster.
- All information is shared among nodes, so the entire file system is accessible by clients connecting to any node in the cluster.
- Scale-out NAS stripes data across all nodes in a cluster along with mirror or parity protection. As data is sent from clients to the cluster, the data is divided and allocated to different nodes in parallel.
- When a client sends a request to read a file, the scale-out NAS retrieves the appropriate blocks from multiple nodes, recombines the blocks into a file, and presents the file to the client.
- As nodes are added, the file system grows dynamically and data is evenly distributed to every node. Each node added to the cluster increases the aggregate storage, memory, CPU, and network capacity. Hence, cluster performance also increases.
- Scale-out NAS is suitable to solve the "Big Data" challenges that enterprises and customers face today.
- It provides the capability to manage and store large, high-growth data in a single place with the flexibility to meet a broad range of performance requirements.

Scale-Out NAS Connectivity

- Scale-out NAS clusters use separate internal and external networks for back-end and front-end connectivity, respectively.
- An internal network provides con- nections for intracluster communication, and an external network connection enables clients to access and share file data. Each node in the cluster connects to the internal network.
- The internal network offers high throughput and low latency and uses high-speed networking technology, such as InfiniBand or Gigabit Ethernet.
- To enable clients to access a node, the node must be connected to the external Ethernet network. Redundant internal or external networks may be used for high availability.
- Figure 7-7 illustrates an example of scale-out NAS connectivity.

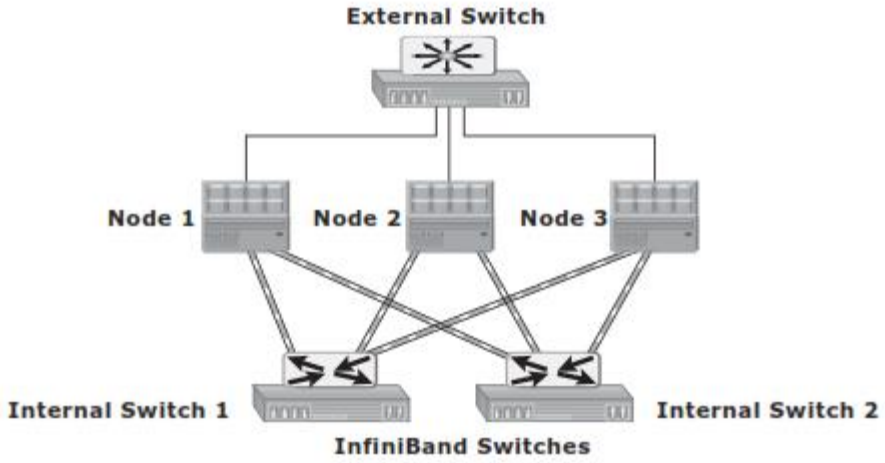


Figure 7-7: Scale-out NAS with dual internal and single external networks

2.12.3 NAS File Sharing Protocols

- NAS devices support multiple file-service protocols to handle file I/Orequests
- Two common NAS file sharing protocolsare:
  - Common Internet File System(CIFS)
  - Network File System(NFS)
- NAS devices enable users to share file data across different operatingenvironments
- It provides a means for users to migrate transparently from one operating system toanother

Network File System (NFS)

- NFS is a **client-server protocol** for file sharing that is commonly used on **UNIXsystems**.
- NFS was originally based on the connectionless *User Datagram Protocol(UDP).*
- It uses *Remote Procedure Call (RPC)* as a method of inter-process communication between two computers.
- The NFS protocol provides a set of RPCs to access a remote file system for the following operations:
    - Searching files anddirectories
    - Opening, reading, writing to, and closing afile
    - Changing fileattributes
    - Modifying file links anddirectories
- NFS creates a connection between the client and the remote system to transferdata.
- NFSv3 and earlier is a statelessprotocol
- It does not maintain any kind of table to store information about open files and associated pointers. Each call provides a full set of arguments - a file handle, a particular position to read or write, and the versions of NFS - to access files on the server.
- Currently, three versions of NFS are inuse:
    1. **NFS version 2 (NFSv2):** Uses *UDP* to provide a *stateless* network connection between a client and a server. Features, such as locking, are handled outside theprotocol.
    2. **NFS version 3 (NFSv3):** Uses *UDP or TCP*, and is based on the *stateless protocol* design. It includes some new features, such as a 64-bit file size, asynchronous writes, and additional file attributes to reducerefetching.
    3. **NFS version 4 (NFSv4):** Uses TCP and is based on a *stateful protocol* design. It offers enhanced security. The latest NFS version 4.1 is the enhancement of NFSv4 and includes some new features, such as session model, parallel NFS (pNFS), and dataretention.

Common Internet File System (CIFS)

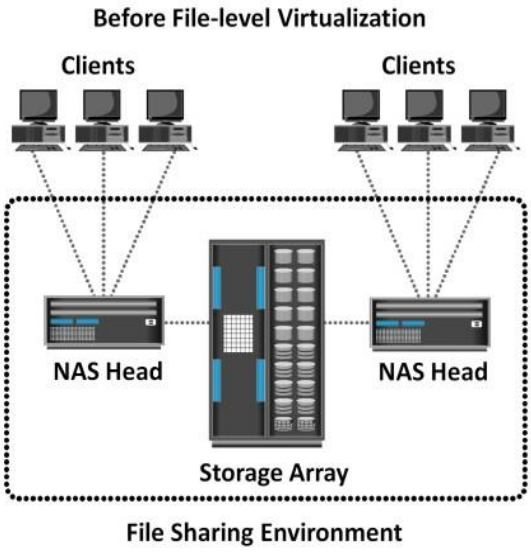
- CIFS is a *client-server application* protocol
- It enables clients to access files and services on remote computers over**TCP/IP**.
- It is a public, or open, variation of **Server Message Block (SMB)**protocol.
- It provides following features to ensure dataintegrity:
    - It uses file and record locking to prevent users from overwriting the work of another user on a file or arecord.
    - It supports fault tolerance and can automatically restore connections and reopen files that were open prior to an interruption. This feature depends on whether an application is written to take advantage ofthis.
    - CIFS is a stateful protocol because the CIFS server maintains connection information regarding every connected client. If a network failure or CIFS server failure occurs, the client receives a disconnection notification. User disruption is minimized if the application has the embedded intelligence to restore the connection. However, if the embedded intelligence is missing, the user must take steps to reestablish the CIFS connection.
- Users refer to remote file systems with an easy-to-use file-namingscheme:
- Eg: \\server\share or\\servername.domain.suffix\share

2.3 File-level Virtualization

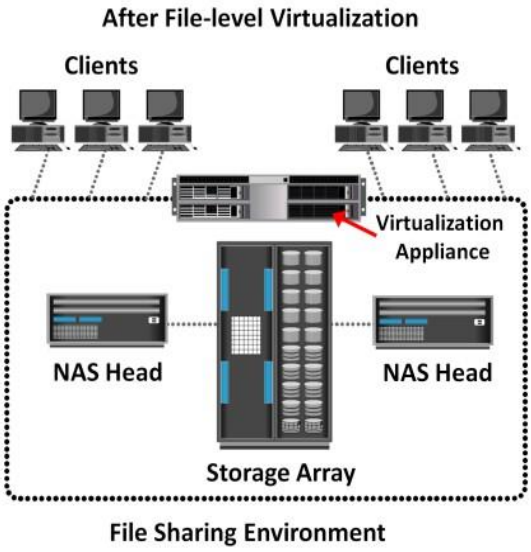
- File-level virtualization, implemented in NAS or the file server environment, provides a simple, non disruptive file-mobilitysolution.
- It eliminates the dependencies between data accessed at the file level and the location where the files are physicallystored.
- It creates a logical pool of storage, enabling users to use a logical path, rather than a physical path, to accessfiles.
- Aglobalnamespaceisusedtomapthelogicalpathofafiletothephysicalpathnames.File-

level virtualization enables the movement of files across NAS devices, even if the files are being accessed.

Before and After File-level Virtualization



- Dependency between client access and file location
- Underutilized storage resources
- Downtime is caused by data migrations



- Break dependencies between client access and file location
- Storage utilization is optimized
- Non-disruptive migrations