UNIT III

STORAGE NETWORKING TECHNOLOGIES AND VIRTUALIZATION

Block-Based Storage System, File-Based Storage System, Object-Based and Unified Storage. Fibre Channel SAN: Software-defined networking, FC SAN components and architecture, FC SAN topologies, link aggregation, and zoning, Virtualization in FC SAN environment. Internet Protocol SAN: iSCSI protocol, network components, and connectivity, Link aggregation, switch aggregation, and VLAN, FCIP protocol, 140 connectivity, andconfiguration. Fibre Channel over Ethernet SAN: Components of FCoE SAN, FCoE SAN connectivity, Converged Enhanced Ethernet, FCoE architecture.

BLOCK-BASED STORAGE SYSTEM

- ***Block storage is for flexible, fast access***
- ***Block storage*** is a form of ***cloud storage*** that is used to store data, often on storage area networks (SANs).
- Data is stored in blocks, with each block stored separately based on the efficiency needs of the SAN.
- Each block is assigned a unique address, which is then used by a management application controlled by the server's operating system to retrieve and compile data into files upon request.
- Block storage offers efficiency due to the way blocks can be distributed across multiple systems and even configured to work with different operating systems.
- Block storage also offers an impressive level of flexibility because it can be accessed by different operating systems as mounted drive volumes and has the ability to use operating system–specific file systems (such as the New Technology File System (NTFS) for Windows and Virtual Machine File System (VMFS) for VMware).
- This makes using block storage quite similar to storing data on a hard drive within a server, except the data is stored in a remote location rather than on local hardware.

How block storage works?

- A block is a fixed-size amount of memory within storage media that’s capable of storing a piece of data. The size of each block is determined by the management system.
- The block size is generally too small to fit an entire piece of data, and so the data for any particular file is broken up into numerous blocks for storage.
- Each block is given a unique identifier without any higher-level metadata; details such as data format, type, and ownership are not noted.
- The operating system allocates and distributes blocks across the storage network to balance efficiency and functionality.
- When a file is requested, the management application uses addresses to identify the necessary blocks and then compiles them into the complete file for use.

By enabling storage across multiple environments, block storage separates data from the limitations of individual user environments. As a result, data can be retrieved through any number of paths to maximize efficiency, with high input/output operations per second (IOPS). The result is an approach that offers a higher level of efficiency than other cloud storage methods, making it ideal for high-performance applications or applications that require constant writing and retrieval.

Benefits of block storage

Block storage is a common and popular cloud storage choice because of its numerous benefits.
High efficiency: Block storage's high IOPS and low latency make it ideal for applications that demand high performance.
Compatibility: Block storage works across different operating systems and file systems, making it compatible for enterprises whatever their configuration and environment.
Flexibility: With block storage, horizontal scaling is extremely flexible. Cluster nodes can be added as needed, allowing for greater overall storage capability.
Large file efficiency: For large files, such as archives and video files, data must be completely overwritten when using file or object storage. With block storage, the management application identifies only the block targeted for change within the large file, increasing the efficiency of data updates.

Limitations of block storage

Like any technology platform, block storage comes with limitations despite its numerous benefits.
Greater cost: While block storage is easily scalable, it can also be expensive due to the cost of SANs. In addition, managing block storage requires more-specialized training for management and maintenance, increasing the overall expense.
Performance limitations: With block storage, metadata is built in and hierarchical, and it is defined by the file system. Because data is broken up into blocks, searching for a complete file requires the proper identification of all its pieces. This can create performance issues for operations accessing the metadata, particularly with folders featuring a large number of files. While the tipping point is usually about 10,000 files, some issues are seen with directories containing only 1,000 files.

Block storage use cases

As with object storage and other types of cloud storage, block storage works best inspecific circumstances based on user needs and given parameters.
The following are just several of many effective block storage use cases: **Containers:** Block storage supports the use of container platforms such as Kubernetes,creating a block volume that enables persistent storage for the entire container. Thisallows for the clean management and migration of containers as needed.
Email servers: Email servers can take advantage of block storage's flexibility and scalability. In fact, in the case of Microsoft Exchange, block storage is required due to the lack of support for network-attachedstorage.

Databases: Block storage is fast, efficient, flexible, and scalable, with support for redundant volumes. This allows it to support databases, particularly those that handle a heavy volume of queries and where latency must be minimized.
Disaster recovery: Block storage can be a redundant backup solution for nearline storage and quick restoration, with data swiftly moved from backup to production through easy access.

Need for *block storage :*

· Block storage continues to be an efficient and flexible cloud storage option forenterprises require high-performance workloads or need to manage large files. Learn more about how Oracle delivers block storage solutions with Oracle Cloud Infrastructure.

FILE-BASED STORAGE SYSTEM

- File storage—also called file-level or file-based storage—is a hierarchical storage methodology used to organize and store data on a computer hard drive or on network-attached storage (NAS) device.
- In file storage, data is stored in files, the files are organized in folders, and the folders are organized under a hierarchy of directories and subdirectories.
- To locate a file, all you or your computer system need is the path—from directory to subdirectory to folder to file.
- Hierarchical file storage works well with easily organized amounts of structured data. But, as the number of files grows, the file retrieval process can become cumbersome and time-consuming. Scaling requires adding more hardware devices or continually replacing these with higher-capacity devices, both of which can get expensive.
- To some extent, you can mitigate these scaling and performance issues with cloud-based file storage services. These services allow multiple users to access and share thesame file data located in off-site data centers (the cloud).
- You simply pay a monthly subscription fee to store your file data in the cloud, and you can easily scale-up capacity and specify your data performance and protection criteria. Moreover, you eliminate the expense of maintaining your own on-site hardware since this infrastructure is managed and maintained by the cloud service provider (CSP) in its data center.
- This is also known as Infrastructure-as-a-Service (IaaS).

File storage benefits

If your organization requires a centralized, easily accessible, and affordable way to store files and folders, file-level storage is a good approach. The benefits of file storage include the following:
Simplicity: File storage is the simplest, most familiar, and most straightforward approach to organizing files and folder on a computer’s hard drive or NAS device. You simply name files, tag them with metadata, and store them in folders under a hierarchy of directories and subdirectories. It is not necessary to write applications or code to access your data.

File sharing: File storage is ideal for centralizing and sharing files on a Local Area Network (LAN). Files stored on a NAS device are easily accessible by any computer on the network that has the appropriate permission rights.

Common protocols: File storage uses common file-level protocols such as Server Message Block (SMB), Common Internet File System (CIFS), or Network File System (NFS). If you utilize a Windows or Linux operating system (or both), standard protocols like SMB/CIFS and NFS will allow you to read and write files to a Windows-based or Linux-based server over your Local Area Network (LAN).

Data protection: Storing files on a separate, LAN-connected storage device offers you a level of data protection should your network computer experience a failure. Cloud-based file storage services provide additional data protection and disaster recovery by replicating data files across multiple, geographically-dispersed data centers.

Affordability: File storage using a NAS device allows you to move files off of expensive computing hardware and onto a more affordable LAN-connected storage device. Moreover, if you choose to subscribe to a cloud file-storage service, you eliminate the expense of on-site hardware upgrades and the associated ongoing maintenance and operation costs.

File storage use cases

File storage is a good solution for a wide variety of data needs, including the following: **Local file sharing:** If your data storage needs are generally consistent and straightforward,such as storing and sharing files with team members in the office, consider the simplicity offile-level storage.

Centralized file collaboration: If you upload, store, and share files in a centralized library—located on-site, off-site, or in the cloud—you can easily collaborate on files with internal and external users or with invited guests outside of your network. **Archiving/storage:** You can cost-effectively archive files on NAS devices in a small data center environment or subscribe to a cloud-based file storage service to store and archive yourdata.

Backup/disaster recovery: You can store backups securely on separate, LAN-connected storage devices. Or you can subscribe to a cloud-based file storage service to replicate your data files across multiple, geographically-dispersed data centers and gain the additional data protection of distance and redundancy.

OBJECT-BASED STORAGE SYSTEM

- Object storage, also known as object-based storage, is a computer data storage architecture designed to handle large amounts of unstructured data.
- Unlike other architectures, it designates data as distinct units, bundled with metadata and a unique identifier that can be used to locate and access each data unit.
- These units—or objects—can be stored on-premises, but are typically stored in the cloud, making them easily accessible from anywhere.
- Due to object storage's scale-out capabilities, there are few limits to its scalability, and it's less costly to store large data volumes than other options, such as block storage.

- Much of today’s data is unstructured: email, media and audio files, web pages, sensor data, and other types of digital content that do not fit easily into traditional databases. As a result, finding efficient and affordable ways to store and manage it has become problematic.
- Increasingly, object storage has become the preferred method for storing static content, data arches, and backups.

- Object storage is a data storage architecture for storing unstructured data, which sections data into units—objects—and stores them in a structurally flat data environment.
- Each object includes the data, metadata, and a unique identifier that applications can use for easy access and retrieval.

Definition of Object storage

How does object storage work?

- With object storage, the data blocks of a file are kept together as an object, together with its relevant metadata and a custom identifier, and placed in a flat data environment known as a storage pool.
- When you want to access data, object storage systems will use the unique identifier and the metadata to find the object you need, such as an image or audio file.
- You can also customize metadata, allowing you to add more context that is useful for other purposes, such as retrieval for data analytics.
- You can locate and access objects using RESTful APIs, HTTP, and HTTPS to query object metadata. Since objects are stored in a global storage pool, it’s fast and easy to locate the exact data you need. Plus, the flat environment enables you to scale quickly, even for petabyte or exabyte loads.
- Storage pools can be spread across multiple object storage devices and geographical locations, allowing for unlimited scale. You simply add more storage devices to the pool as your data grows.
- The benefits of object storage, like its elasticity and scalability, have made it an ideal fit for managing unstructured data in cloud infrastructure. So, what is object storage inthe cloud? It’s exactly what it sounds like—object-based storage as an on-demand cloud service.
- In fact, cloud object storage is the primary storage format for most major cloud service providers.

Benefits of object storage:

Massive scalability

You can easily scale out the flat architecture of object storage without suffering from the same limitations as file or block storage. Object storage size is essentially limitless, so data can scale to exabytes by simply adding new devices.

Reduced complexity

Object storage has no folders or directories, removing much of the complexity that comes with hierarchical systems. The lack of complex trees or partitions makes retrieving files easier as you don’t need to know the exact location.

Searchability

Metadata is part of objects, making it easy to search through and navigate withoutthe need of a separate application. It’s also far more flexible and customizable. You can tag objects with attributes and information, such as consumption, cost, and policies for automated deletion, retention, and tiering.

Resiliency

Object storage can automatically replicate data and store it across multiple devices and geographical locations. This can help protect against outages, safeguard against data loss, and help support disaster recovery strategies.

Cost efficiency

Object storage was created with cost in mind, providing storage for large amounts of data at a lower price than file- and block-based systems. With object storage, you only pay for the capacity you need, allowing you to control costs even for large amounts of data.

UNIFIED STORAGE :

- Also known as multiprotocol storage, unified storage allows multiple types of data to be stored in the same device.
- It combines block and file storage protocols, such as iSCSI, NFS, and SMB, into a single platform, making it easier for IT administrators to manage and maintain their storage infrastructure because they have it all in one place.
- With unified storage, users can access their data from different applications andplatforms via a single interface, which helps streamline workflows and reduce storagecomplexity.

Unified Storage Architecture

- A unified storage architecture is the design and framework that underpins thefunctionality of unified storage systems.
- It uses a single storage pool to store data, which can be allocated dynamically to different storage tiers based on the data's usage.
- The architecture also includes features like data deduplication, compression, and encryption, which further enhance the efficiency and security of the storage system.

How a Unified Storage Architecture Works

- A unified storage architecture works by consolidating different types of storage into a single system. It provides a single point of management for all storage-related tasks, including provisioning, allocation, backup, and recovery.
- It also uses advanced features such as thin provisioning, snapshots, and cloning to reduce storage waste and improve efficiency.
- Furthermore, a unified storage architecture can scale horizontally and vertically to meet the growing demands of data-intensive applications and workloads.

Components of Unified Data Storage

- A unified data storage system contains several components, including storagecontrollers, storage arrays, network interfaces, and management software.

- The storage controllers manage the storage access protocols and data services, while the storage arrays contain the physical storage devices such as hard drives and solid- state drives.
- Network interfaces connect the storage system to the network, and management software provides a GUI or command-line interface for administrators to manage and monitor the storage environment.

Support for Multiple Storage Protocols

- One of the key benefits of unified storage is its support for multiple storage protocols. This allows users to access their data from a variety of platforms and applications, regardless of the underlying storage technology.
- For example, a user can access their files from their desktop PC, laptop, or mobile device, without worrying about the file format or location. Unified storage also supports multiple operating systems, including Windows, Linux, and Mac OS.

Benefits and Advantages of Unified Storage

- Unified storage offers several benefits and advantages to businesses and networks. It's simplified and cost-effective, for one.
- It also offers scalable and flexible architecture and improved data protection and security. Let's take a closer look at each of these benefits.

Simplified and Cost-effective

By consolidating different types of storage into a single platform, unified storage simplifies storage management and reduces operational costs. It eliminates the need for separate storage systems for different types of data, reducing the time and effort required to manage and maintain them.

Additionally, unified storage can help businesses save money on storage hardware and software licenses, as they no longer need to purchase separate systems for block and file storage.

Scalable and Flexible

A unified storage architecture is designed to scale both horizontally and vertically tomeet the growing demands of data-intensive applications and workloads.

It uses tiered storage, thin provisioning, and other techniques to optimize the use ofavailable storage resources, increasing the system's capacity and performance.

A unified storage architecture is also flexible enough to support different storage technologies, including hard disk drives, solid-state drives, and cloud storage.

Potential Downsides of Unified Storage

While unified storage offers many benefits, it also has potential downsides that businesses should consider. Two of the most significant downsides are performance and complexity issues and vendor lock-in.

Performance and Complexity Issues

Unified storage can be more challenging to manage and maintain than separate systems for block and file storage.

Furthermore, unified storage may require additional resources, like network bandwidth and processing power, to handle the diverse storage workloads effectively.

Vendor Lock-in

Another potential downside of unified storage is that it can be challenging to switch vendors or migrate to different storage technologies once you have committed to a particular system.

This can result in higher costs and being locked in to specific hardware and software.

Unified Storage Vendors and Providers

There are several vendors and providers that claim to provide unified storage solutions, including NetApp, Dell, and Hewlett Packard. Each vendor offers a different set of features, performance, and pricing options.

However, historically these legacy storage systems have only offered unified storage via compromises and retrofits leveraging architectures built for the era of spinning disk. They lack native multi-protocol support for block and file on all-flash storage.

FlashArray is the first truly unified block and file platform of its kind. Not only doyou get a global storage pool that can be used across block and file, but you also can expand it non-disruptively on the fly with unlimited file system sizes. FlashArray also offers high performance, scalability, and flexibility, making it an excellent choice for businesses and networks with demanding storage needs.

Block Storage vs. File Storage vs. Unified Storage

Block storage is a type of storage that manages data in large, fixed-sized blocks, typically used for databases, virtual machines, and backup data. File storage, on the other hand, organizes data into files and folders, typically used for documents, images, and video files. Unified storage combines both block and file storage protocols into a single platform, providing a more versatile and flexible storage solution.

Object storage vs. file storage vs. block storage

- Over time, the world's data storage needs have evolved with the introduction of the internet and an expanding list of data sources and types. Traditional file storage and block storage aren't well-suited to handle the enormous amount of data being generated, especially unstructured data that is not made to fit into structured data storage methods.

So, how does object storage compare to file storage and block storage?

File storage

File storage stores and organizes data into folders, similar to the physical files you might store in a paper filing system in an office. If you need information from a file, you'll need to know what room, cabinet, drawer, and folder contains that specific document. This same hierarchical storage structure is used for file storage, where files are named, tagged withmetadata, and then placed in folders.

To locate a piece of data, you'll need to know the correct path to find it. Over time, searching and retrieving data files can become time-consuming as the number of files grows. While scalability is more limited, it is a simple way to store small amounts of just about any type of data and make it accessible to multiple users at once.

Block storage

Block storage improves on the performance of file storage, breaking files into separate blocks and storing them separately. A block-storage system will assign a unique identifier to each chunk of raw data, which can then be used to reassemble them into the complete file when you need to access it. Block storage doesn't require a single path to data, so you can store it wherever is most convenient and still retrieve it quickly when needed.

Block storage works well for organizations that work with large amounts of transactional data or mission-critical applications that need minimal delay and consistent performance. However, it can be expensive, offers no metadata capabilities, and requires an operating system to access blocks.

Object storage

Object storage, as discussed earlier, saves files in a flat data environment, or storage pool, as a self-contained object that contains all the data, a unique identifier, and detailed metadata that contains information about the data, permissions, policies, and other contingencies. Object storage works best for static storage, especially for unstructured data, where you write data once but may need to read it many times.

While object storage eliminates the need for directories, folders, and other complex hierarchical organization, it's not a good solution for dynamic data that is changing constantly as you'll need to rewrite the entire object to modify it. In some cases, file storage and block storage may still suit your needs depending on your speed and performancerequirements.

FIBRE CHANNEL SAN

- Fibre Channel is a high-speed network technology that runs on high-speed optical fiber cables (preferred for front-end SAN connectivity) and serial copper cables (preferred for back-end disk connectivity).
- The FC technology was created to meet the demand for increased speeds of data transfer among computers, servers, and mass storage subsystems.

SOFTWARE DEFINED NETWORKING (SDN)

- SDN is an approach to networking that uses software controllers that can be driven by application programming interfaces (APIs) to communicate with hardware infrastructure to direct network traffic. Using software, it creates and operates a series of virtual overlay networks that work in conjunction with a physical underlay network.
- SDNs offer the potential to deliver application environments as code and minimize the hands-on time needed for managing the network.

Types of SDN

There are four primary types of software-defined networking (SDN):

- **Open SDN** – Open protocols are used to control the virtual and physical devices responsible for routing the data packets.
- **API SDN** – Through programming interfaces, often called southbound APIs,organizations control the flow of data to and from each device.

·

- **Overlay Model SDN** – It creates a virtual network above existing hardware, providing tunnels containing channels to data centers. This model then allocates bandwidth in each channel and assigns devices to each channel.
- **Hybrid Model SDN** – By combining SDN and traditional networking, the hybrid model assigns the optimal protocol for each type of traffic. Hybrid SDN is often used as an incremental approach to SDN.

SDN Architecture

The architecture of software-defined networking (SDN) consists of three main layers: the application layer, the control layer, and the infrastructure layer. Each layer has a specific role and interacts with the other layers to manage and control the network.

Infrastructure Layer: The infrastructure layer is the bottom layer of the SDN architecture, also known as the data plane. It consists of physical and virtual network devices such as switches, routers, and firewalls that are responsible for forwarding network traffic based on the instructions received from the control plane.

Control Layer: The control layer is the middle layer of the SDN architecture, also known as the control plane. It consists of a centralized controller that communicates with the infrastructure layer devices and is responsible for managing and configuring the network. The controller interacts with the devices in the infrastructure layer using protocols such as OpenFlow to program the forwarding behaviour of the switches and routers. The controller uses network policies and rules to make decisions about how traffic should be forwarded based on factors such as network topology, traffic patterns, and quality of service requirements.

Application Layer: The application layer is the top layer of the SDN architecture and is responsible for providing network services and applications to end-users. This layer consists of various network applications that interact with the control layer to manage the network.

Examples of applications that can be deployed in an SDN environment include network virtualization, traffic engineering, security, and monitoring. The application layer can be used to create customized network services that meet specific business needs.

The main benefit of the SDN architecture is its flexibility and ability to centralize control of the network. The separation of the control plane from the data plane enables network administrators to configure and manage the network more easily and in a more granular way, allowing for greater network agility and faster response times to changes in network traffic.

Advantages of SDN:

Software-defined networking (SDN) offers several advantages over traditional networking architectures, including:

o ***Centralized Network Control:*** One of the key benefits of SDN is that it centralizes the control of the network in a single controller, making it easier to manage and configure the network. This allows network administrators to define and enforce network policies in a more granular way, resulting in better network security, performance, and reliability.

o ***Programmable Network:*** In an SDN environment, network devices are programmable and can be reconfigured on the fly to meet changing network requirements. This allows network administrators to quickly adapt the network to changing traffic patterns and demands, resulting in better network performance and efficiency.

o ***Cost Savings:*** With SDN, network administrators can use commodity hardware to build a network, reducing the cost of proprietary network hardware. Additionally, the centralization of network control can reduce the need for manual networkmanagement, leading to cost savings in labor and maintenance.

o ***Enhanced Network Security:*** The centralized control of the network in SDN makes it easier to detect and respond to security threats. The use of network policies and rules allows administrators to implement fine-grained security controls that can mitigate security risks.

o ***Scalability:*** SDN makes it easier to scale the network to meet changing traffic demands. With the ability to programmatically control the network, administrators can quickly adjust the network to handle more traffic without the need for manual intervention.

o ***Simplified Network Management:*** SDN can simplify network management by abstracting the underlying network hardware and presenting a logical view of the network to administrators. This makes it easier to manage and troubleshoot the network, resulting in better network uptime and reliability.

Disadvantages of SDN

While software-defined networking (SDN) has several advantages over traditional networking, there are also some potential disadvantages that organizations should be aware of. Here are some of the main disadvantages of SDN:

o ***Complexity:*** SDN can be more complex than traditional networking because it involves a more sophisticated set of technologies and requires specialized skills to manage. For example, the use of a centralized controller to manage the network requires a deep understanding of the SDN architecture and protocols.

o ***Dependency on the Controller:*** The centralized controller is a critical component of SDN, and if it fails, the entire network could go down. This means that organizations need to ensure that the controller is highly available and that they have a robust backup and disaster recovery plan in place.

o ***Compatibility:*** Some legacy network devices may not be compatible with SDN, which means that organizations may need to replace or upgrade these devices to take full advantage of the benefits of SDN.

o ***Security:*** While SDN can enhance network security, it can also introduce new security risks. For example, a single point of control could be an attractive target for attackers, and the programmability of the network could make it easier for attackers tomanipulate traffic.

Vendor Lock-In: SDN solutions from different vendors may not be interoperable, which could lead to vendor lock-in. This means that organizations may be limited in their ability to switch to another vendor or integrate new solutions into their existing network.

o ***Performance:*** The centralized control of the network in SDN can introduce latency, which could impact network performance in certain situations. Additionally, the overhead of the SDN controller could impact the performance of the network as the network scales.

COMPONENTS OF FC SAN

- A SAN consists of **three basic components: servers, network infrastructure, and storage.**
- These components can be further broken down into the following key elements: ***node ports, cabling, interconnecting devices (such as FC switches or hubs), storage arrays, and SAN management software.***

Node Ports

- In fibre channel, devices such as hosts, storage and tape libraries are all referred to as nodes. Each node is a source or destination of information for one or more nodes.
- Each node requires one or more ports to provide a physical interface forcommunicating with other nodes. These ports are integral components of an HBA and the storage front-end adapters.
- A port operates in full-duplex data transmission mode with a transmit (Tx) link and a receive (Rx) link (see Figure 6-3)

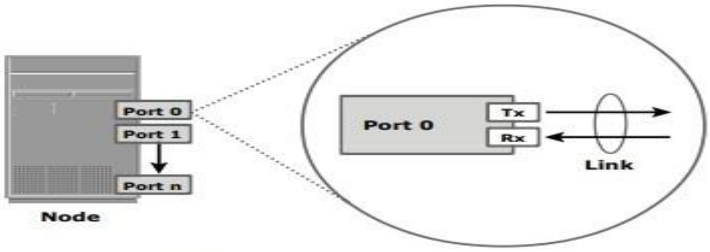


Figure 6-3: Nodes, ports, and links

Cabling

- SAN implementations use optical fiber cabling. Copper can be used for shorter distances for back-end connectivity, as it provides a better signal-to-noise ratio for distances up to 30 meters.
- Optical fiber cables carry data in the form of light.
- There are two types of optical cables, multi-mode and single-mode. Multi-mode fiber (MMF) cable carries multiple beams of light projected at different angles simultaneously onto the core of the cable (see Figure 6-4 (a)).
- Based on the bandwidth, multi-mode fibers are classified as OM1 (62.5µm), OM2 (50µm) and laser optimized OM3 (50µm). In an MMF transmission, multiple light beams traveling inside the cable tend to disperse and collide.
- This collision weakens the signal strength after it travels a certain distance — a process known as modal dispersion. An MMF cable is usually used for distances of upto 500 meters because of signal degradation (attenuation) due to modal dispersion.

Single-mode fiber (SMF) carries a single ray of light projected at the center of the core(see Figure 6-4 (b)).

- These cables are available in diameters of 7–11 microns; the most common size is 9 microns. In an SMF transmission, a single light beam travels in a straight line throughthe core of the fiber.

- The small core and the single light wave limits modal dispersion. Among all types of fibre cables, single-mode provides minimum signal attenuation over maximumdistance (up to 10 km).
- A single-mode cable is used for long-distance cable runs, limited only by the power of the laser at the transmitter and sensitivity of the receiver

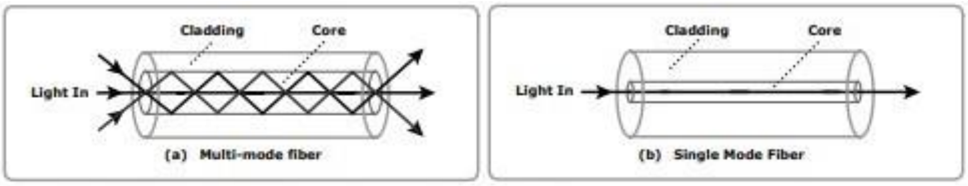


Figure 6-4: Multi-mode fiber and single-mode fiber

- MMFs are generally used within data centers for shorter distance runs, while SMFsare used for longer distances. MMF transceivers are less expensive as compared to SMF transceivers.
- A Standard connector (SC) (see Figure 6-5 (a)) and a Lucent connector (LC) (see Figure 6-5 (b)) are two commonly used connectors for fiber optic cables.
- An SC is used for data transmission speeds up to 1 Gb/s, whereas an LC is used for speeds up to 4 Gb/s. Figure 6-6 depicts a Lucent connector and a Standard connector.
- A Straight Tip (ST) is a fiber optic connector with a plug and a socket that is locked with a half-twisted bayonet lock (see Figure 6-5 (c)).
- In the early days of FC deployment, fiber optic cabling predominantly used STconnectors. This connector is often used with Fibre Channel patch panels

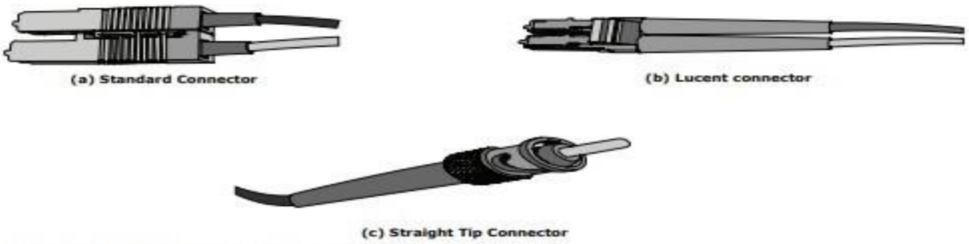


Figure 6-5: SC, LC, and ST connectors

- The Small Form-factor Pluggable (SFP) is an optical transceiver used in optical communication. The standard SFP+ transceivers support data rates up to 10 Gb/s.

Interconnect Devices

- Hubs, switches, and directors are the interconnect devices commonly used in SAN. Hubs are used as communication devices in FC-AL implementations. Hubs physically connect nodes in a logical loop or a physical star topology.
- All the nodes must share the bandwidth because data travels through all the connection points. Because of availability of low cost and high performance switches, hubs are no longer used in SANs.
- Switches are more intelligent than hubs and directly route data from one physical port to another.

- Therefore, nodes do not share the bandwidth. Instead, each node has a dedicated communication path, resulting in bandwidth aggregation
- Directors are larger than switches and are deployed for data center implementations. The function of directors is similar to that of FC switches, but directors have higher port count and fault tolerance capabilities.

Storage Arrays

- The fundamental purpose of a SAN is to provide host access to storage resources.
- The large storage capacities offered by modern storage arrays have been exploited in SAN environments for storage consolidation and centralization.
- SAN implementations complement the standard features of storage arrays by providing high availability and redundancy, improved performance, business continuity, and multiple host connectivity.

SAN Management Software

- SAN management software manages the interfaces between hosts, interconnect devices, and storage arrays.
- The software provides a view of the SAN environment and enables management of various resources from one central console. It provides key management functions, including mapping of storage devices, switches, and servers, monitoring and generating alerts for discovered devices, and logical partitioning of the SAN, called zoning.
- In addition, the software provides management of typical SAN components such as HBAs, storage components, and interconnecting devices.

➔ FC ARCHITECTURE

- The FC architecture represents true channel/network integration with standard interconnecting devices. Connections in a SAN are accomplished using FC.
- Traditionally, transmissions from host to storage devices are carried out over channel connections such as a parallel bus. Channel technologies provide high levels of performance with low protocol overheads.
- Such performance is due to the static nature of channels and the high level of hardware and software integration provided by the channel technologies.
- However, these technologies suffer from inherent limitations in terms of the number of devices that can be connected and the distance between these devices. Fibre Channel Protocol (FCP) is the implementation of serial SCSI-3 over an FC network. In the FCP architecture, all external and remote storage devices attached to the SAN appear as local devices to the host operating system.

The key advantages of FCP are as follows:

- ■ Sustained transmission bandwidth over long distances.

- Support for a larger number of addressable devices over a network.Theoretically, FC can support over 15 million device addresses on a network.
- Exhibits the characteristics of channel transport and provides speeds up to 8.5 Gb/s (8 GFC).

- The FC standard enables mapping several existing Upper Layer Protocols (ULPs) to FC frames for transmission, including SCSI, IP, High Performance Parallel Interface (HIPPI), Enterprise System Connection (ESCON), and Asynchronous Transfer Mode (ATM).

Fibre Channel Protocol Stack

- It is easier to understand a communication protocol by viewing it as a structure of independent layers.
- FCP defines the communication protocol in five layers: FC-0 through FC-4 (except FC-3 layer, which is not implemented). In a layered communication model, the peer layers on each node talk to each other through defined protocols.

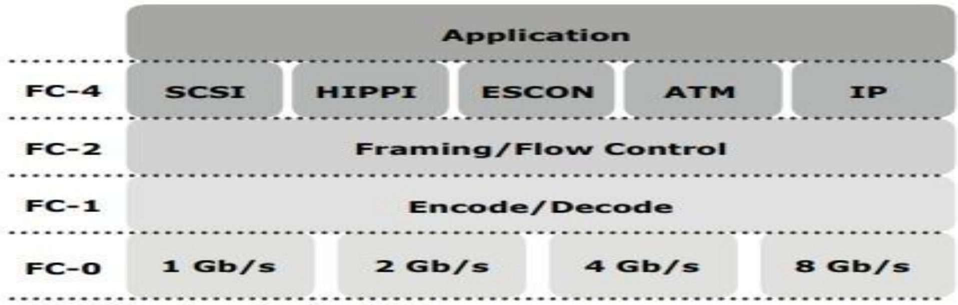


Figure 6-13: Fibre channel protocol stack

- Figure 6-13 illustrates the fibre channel protocol stack

FC-4 Upper Layer Protocol

- FC-4 is the uppermost layer in the FCP stack. This layer defines the application interfaces and the way Upper Layer Protocols (ULPs) are mapped to the lower FC layers. The FC standard defines several protocols that can operate on the FC-4 layer (see Figure 6-7).
- Some of the protocols include SCSI, HIPPI Framing Protocol, Enterprise Storage Connectivity (ESCON), ATM, and IP

FC-2 Transport Layer

- The FC-2 is the transport layer that contains the payload, addresses of the source and destination ports, and link control information.
- The FC-2 layer provides Fibre Channel addressing, structure, and organization of data (frames, sequences, and exchanges).
- It also defines fabric services, classes of service, flow control, and routing.

FC-1 Transmission Protocol

- This layer defines the transmission protocol that includes serial encoding and decoding rules, special characters used, and error control.

- At the transmitter node, an 8-bit character is encoded into a 10-bit transmissions character. This character is then transmitted to the receiver node.
- At the receiver node, the 10-bit character is passed to the FC-1 layer, which decodes the 10-bit character into the original 8-bit character.

FC-0 Physical Interface

- FC-0 is the lowest layer in the FCP stack. This layer defines the physical interface, media, and transmission of raw bits.
- The FC-0 specification includes cables, connectors, and optical and electrical parameters for a variety of data rates. The FC transmission can use both electrical and optical media.

Fibre Channel Addressing

- An FC address is dynamically assigned when a port logs on to the fabric. The FC address has a distinct format that varies according to the type of node port in the fabric. These ports can be an N_port and an NL_port in a public loop, or an NL_portin a private loop.
- The first field of the FC address of an N_port contains the domain ID of the switch (see Figure 6-14).
- This is an 8-bit field. Out of the possible 256 domain IDs, 239 are available for use;the remaining 17 addresses are reserved for specific services. For example, FFFFFC is reserved for the name server, and FFFFFE is reserved for the fabric login service.
- The maximum possible number of N_ports in a switched fabric is calculated as 239 domains × 256 areas × 256 ports = 15,663,104 Fibre Channel addresses

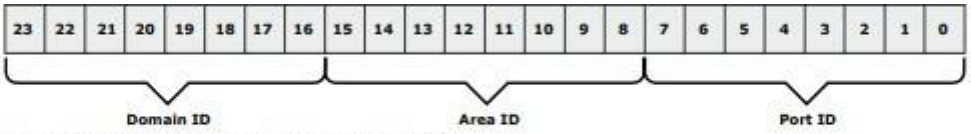


Figure 6-14: 24-bit FC address of N_port

- The area ID is used to identify a group of F_ports. An example of a group of F_ports would be a card on the switch with more than one port on it. The last field in the FC address identifies the F_port within the group.

FC Address of an NL_port

- The FC addressing scheme for an NL_port differs from other ports.
- The two upper bytes in the FC addresses of the NL_ports in a private loop are assigned zero values. However, when an arbitrated loop is connected to a fabric through an FL_port, it becomes a public loop.
- In this case, an NL_port supports a fabric login. The two upper bytes of this NL_port are then assigned a positive value, called a loop identifier, by the switch. The loop identifier is the same for all NL_ports on a given loop.
- Figure 6-15 illustrates the FC address of an NL_port in both a public loop and a private loop. The last field in the FC addresses of the NL_ports, in both public and

private loops, identifies the AL-PA. There are 127 allowable AL-PA addresses; oneaddress is reserved for the FL_port on the switch

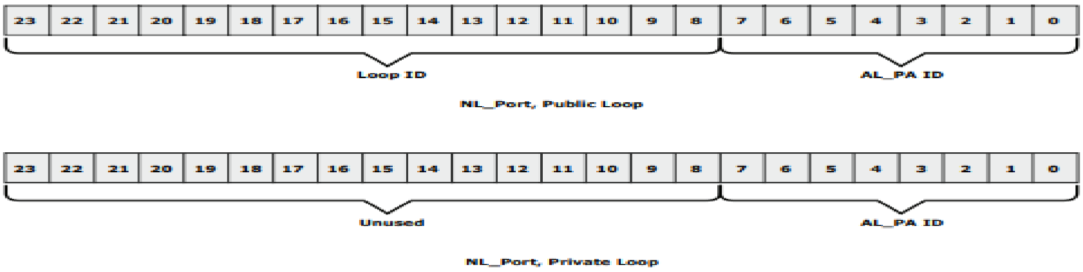
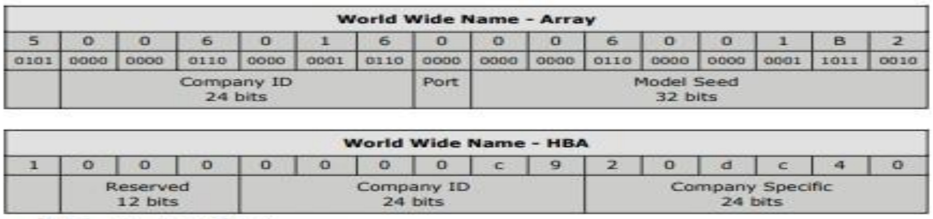


Figure 6-15: 24-bit FC address of NL_port

World Wide Names

- Each device in the FC environment is assigned a 64-bit unique identifier called the World Wide Name (WWN). The Fibre Channel environment uses two types of WWNs: World Wide Node Name (WWNN) and World Wide Port Name (WWPN). Unlike an FC address, which is assigned dynamically, a WWN is a static name for each device on an FC network.
- WWNs are similar to the Media Access Control (MAC) addresses used in IP networking. WWNs are burned into the hardware or assigned through software. Several configuration definitions in a SAN use WWN for identifying storage devices and HBAs.
- The name server in an FC environment keeps the association of WWNs to the dynamically created FC addresses for nodes. Figure 6-16 illustrates the WWNstructure for an array and the HBA.



World Wide Name - Array															
5	0	0	6	0	1	6	0	0	0	6	0	0	1	B	2
0101	0000	0000	0110	0000	0001	0110	0000	0000	0000	0110	0000	0000	0001	1011	0010
	Company ID 24 bits						Port	Model Seed 32 bits							

World Wide Name - HBA															
1	0	0	0	0	0	0	0	c	9	2	0	d	c	4	0
	Reserved 12 bits			Company ID 24 bits						Company Specific 24 bits					

Figure 6-16: World Wide Names

FC Frame

- An FC frame (Figure 6-17) consists of five parts: start of frame (SOF), frame header, data field, cyclic redundancy check (CRC), and end of frame (EOF).
- The SOF and EOF act as delimiters.
- In addition to this role, the SOF is a flag that indicates whether the frame is the first frame in a sequence of frames. The frame header is 24 bytes long and contains addressing information for the frame.
- It includes the following information: Source ID (S_ID), Destination ID (D_ID),Sequence ID (SEQ_ID), Sequence Count (SEQ_CNT), Originating Exchange ID (OX_ID), and Responder Exchange ID (RX_ID), in addition to some control fields

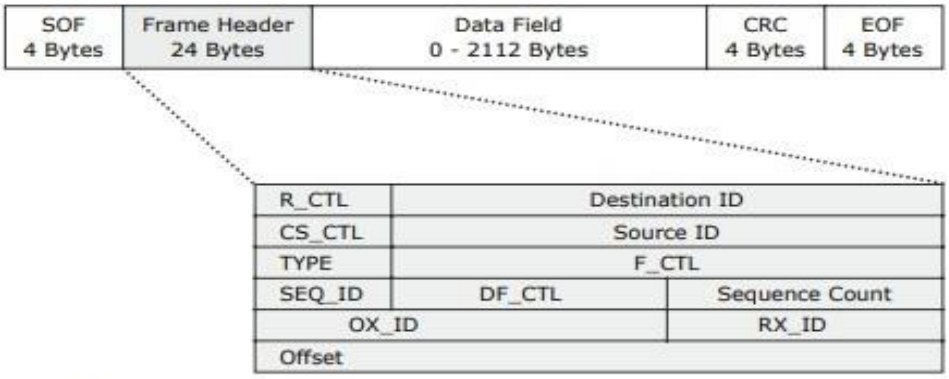


Figure 6-17: FC frame

- The S_ID and D_ID are standard FC addresses for the source port and the destination port, respectively. The SEQ_ID and OX_ID identify the frame as a component of a specific sequence and exchange, respectively.

- ***The frame header also defines the following fields:***

■ Routing Control (R_CTL): This field denotes whether the frame is a link control frame or a data frame. Link control frames are nondata frames that do not carry any payload. These frames are used for setup and messaging. In contrast, data frames carrythe payload and are used for data transmission.

■ Class Specific Control (CS_CTL): This field specifies link speeds for class 1 and class 4 data transmission.

■ TYPE: This field describes the upper layer protocol (ULP) to be carried on the frame if it is a data frame. However, if it is a link control frame, this field is used to signal an event such as “fabric busy.” For example, if the TYPE is 08, and the frame isa data frame, it means that the SCSI will be carried on an FC.

■ Data Field Control (DF_CTL): A 1-byte field that indicates the existence of any optional headers at the beginning of the data payload. It is a mechanism to extend header information into the payload.

■ Frame Control (F_CTL): A 3-byte field that contains control information related to frame content. For example, one of the bits in this field indicates whether this is the first sequence of the exchange

Structure and Organization of FC Data

- In an FC network, data transport is analogous to a conversation between two people, whereby a frame represents a word, a sequence represents a sentence, and an exchangerepresents a conversation.

■ **Exchange operation:**

An exchange operation enables two N_ports to identify and manage a set of information units. This unit maps to a sequence. Sequences can be both unidirectional and bidirectional depending upon the type of data sequence exchanged between the initiator and the target.

■ **Sequence**:

A sequence refers to a contiguous set of frames that are sent from one port toanother. A sequence corresponds to an information unit, as defined by the ULP.

- **Frame:**

A frame is the fundamental unit of data transfer at Layer 2. Each frame cancontain up to 2,112 bytes of payload.

Flow Control

- Flow control defines the pace of the flow of data frames during data transmission. FC technology uses two flow-control mechanisms: buffer-to-buffer credit (BB_Credit) and end-to-end credit (EE_Credit).

BB_Credit

- FC uses the BB_Credit mechanism for hardware-based flow control.
- BB_Credit controls the maximum number of frames that can be present over the link at any given point in time. In a switched fabric, BB_Credit management may take place between any two FC ports.
- The transmitting port maintains a count of free receiver buffers and continues to send frames if the count is greater than 0. The BB_Credit mechanism provides frame acknowledgment through the Receiver Ready (R_RDY) primitive.

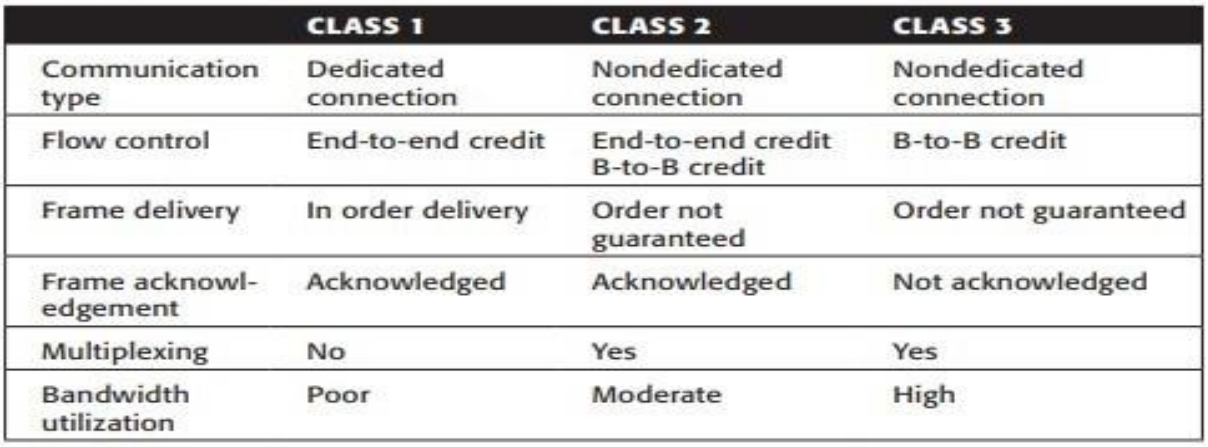
EE_Credit

- The function of end-to-end credit, known as EE_Credit, is similar to that of BB_ Credit.
- When an initiator and a target establish themselves as nodes communicating with each other, they exchange the EE_Credit parameters (part of Port Login).
- The EE_Credit mechanism affects the flow control for class 1 and class 2 traffic only

Classes of Service

- The FC standards define different classes of service to meet the requirements of a wide range of applications.
- The table below shows three classes of services and their features (Table 6-1)

Table 6-1: FC Class of Services



	CLASS 1	CLASS 2	CLASS 3
Communication type	Dedicated connection	Nondedicated connection	Nondedicated connection
Flow control	End-to-end credit	End-to-end credit B-to-B credit	B-to-B credit
Frame delivery	In order delivery	Order not guaranteed	Order not guaranteed
Frame acknowledgement	Acknowledged	Acknowledged	Not acknowledged
Multiplexing	No	Yes	Yes
Bandwidth utilization	Poor	Moderate	High

- Another class of services is class F, which is intended for use by the switches communicating through ISLs. Class F is similar to Class 2, and it provides notification of non-delivery of frames.

➔ FC TOPOLOGIES :

- Fabric design follows standard topologies to connect devices.
- Core-edge fabric is one of the popular topology designs.
- ***Variations of core-edge fabric and mesh topologies*** are most commonly deployed in SAN implementations.

Core-Edge Fabric

- In the core-edge fabric topology, ***there are two types of switch tiers in this fabric.*** Theedge tier usually comprises switches and offers an inexpensive approach to adding more hosts in a fabric.
- The tier at the edge fans out from the tier at the core. The nodes on the edge can communicate with each other.
- The core tier usually comprises enterprise directors that ensure high fabric availability. Additionally all traffic has to either traverse through or terminate at this tier. In a two- tier configuration, all storage devices are connected to the core tier, facilitating fan- out.
- The host-to-storage traffic has to traverse one and two ISLs in a two-tier and three-tier configuration, respectively.
- Hosts used for mission-critical applications can be connected directly to the core tier and consequently avoid traveling through the ISLs to process I/O requests from these hosts.

- The core-edge fabric topology increases connectivity within the SAN while conserving overall port utilization. If expansion is required, an additional edge switch can be connected to the core.
- This topology can have different variations. In a single-core topology, all hosts are connected to the edge tier and all storage is connected to the core tier.
- Figure 6-21 depicts the core and edge switches in a single-core topology.

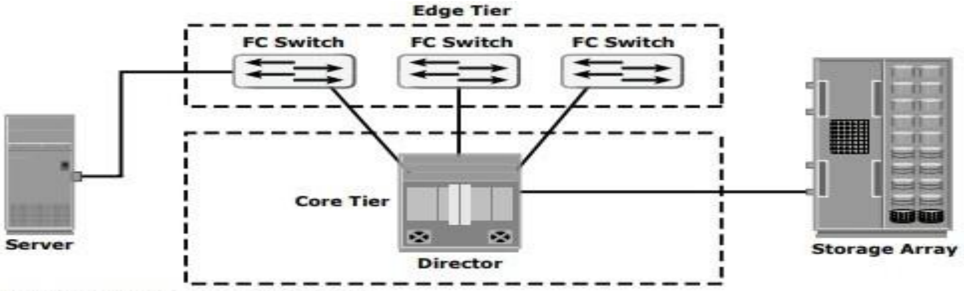


Figure 6-21: Single core topology

- A dual-core topology can be expanded to include more core switches.
- However, to maintain the topology, it is essential that new ISLs are created to connecteach edge switch to the new core switch that is added.
- Figure 6-22 illustrates the core and edge switches in a dual-core topology.

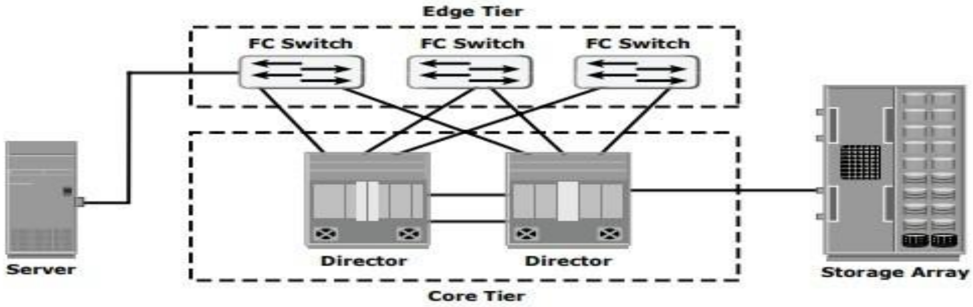


Figure 6-22: Dual-core topology

Benefits and Limitations of Core-Edge Fabric

- The core-edge fabric provides one-hop storage access to all storage in the system. Because traffic travels in a deterministic pattern (from the edge to the core), a core- edge provides easier calculation of ISL loading and traffic patterns.
- Because each tier's switch is used for either storage or hosts, one can easily identify which resources are approaching their capacity, making it easier to develop a set of rules for scaling and apportioning.
- A well-defined, easily reproducible building-block approach makes rolling out new fabrics easier. Core-edge fabrics can be scaled to larger environments by linking core switches, adding more core switches, or adding more edge switches.
- This method can be used to extend the existing simple core-edge model or to expand the fabric into a compound or complex core-edge model.
- However, the core-edge fabric may lead to some performance-related problems because scaling a core-edge topology involves increasing the number of ISLs in the fabric.

- As more edge switches are added, the domain count in the fabric increases. A common best practice is to keep the number of host-to-storage hops unchanged, at one hop, in acore-edge.
- Hop count represents the total number of devices a given piece of data (packet) passes through. Generally a large hop count means greater the transmission delay between data traverse from its source to destination.
- As the number of cores increases, it may be prohibitive to continue to maintain ISLs from each core to each edge switch. When this happens, the fabric design can be changed to a compound or complex core-edge design.

Mesh Topology

- In a mesh topology, each switch is directly connected to other switches by using ISLs. This topology promotes enhanced connectivity within the SAN.
- When the number of ports on a network increases, the number of nodes that can participate and communicate also increases.
- A mesh topology may be one of the two types: full mesh or partial mesh. In a full mesh, every switch is connected to every other switch in the topology. Full mesh topology may be appropriate when the number of switches involved is small.
- A typical deployment would involve up to four switches or directors, with each of them servicing highly localized host-to-storage traffic.

- In a full mesh topology, a maximum of one ISL or hop is required for host-to-storage traffic. In a partial mesh topology, several hops or ISLs may be required for the traffic to reach its destination.
- Hosts and storage can be located anywhere in the fabric, and storage can be localized to a director or a switch in both mesh topologies.
- A full mesh topology with a symmetric design results in an even number of switches, whereas a partial mesh has an asymmetric design and may result in an odd number of switches.
- Figure 6-23 depicts both a full mesh and a partial mesh topology.

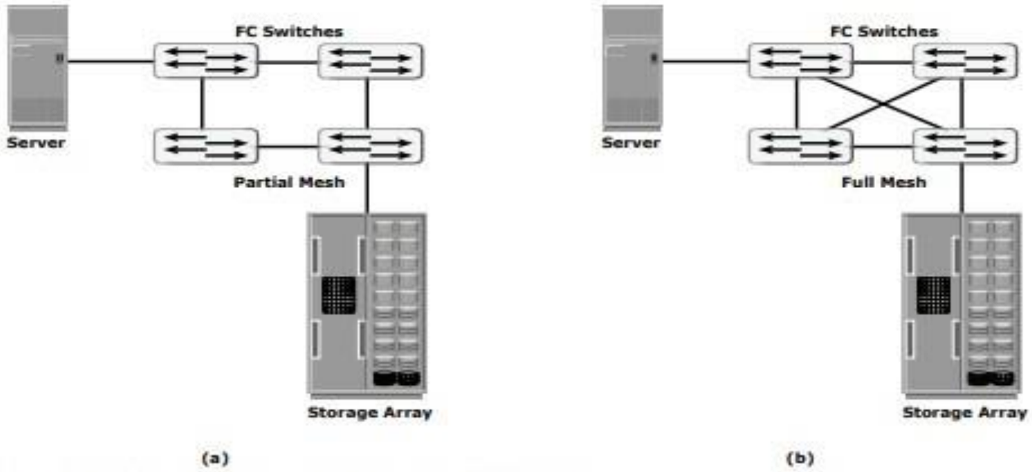


Figure 6-23: Partial mesh and full mesh topologies

ZONING

- Zoning is an FC switch function that enables nodes within the fabric to be logically segmented into groups that can communicate with each other (see Figure 6-18).
- When a device (host or storage array) logs onto a fabric, it is registered with the name server.When a port logs onto the fabric, it goes through a device discovery process with other devices registered in the name server.
- The zoning function controls this process by allowing only the members in the same zone to establish these link-level services.

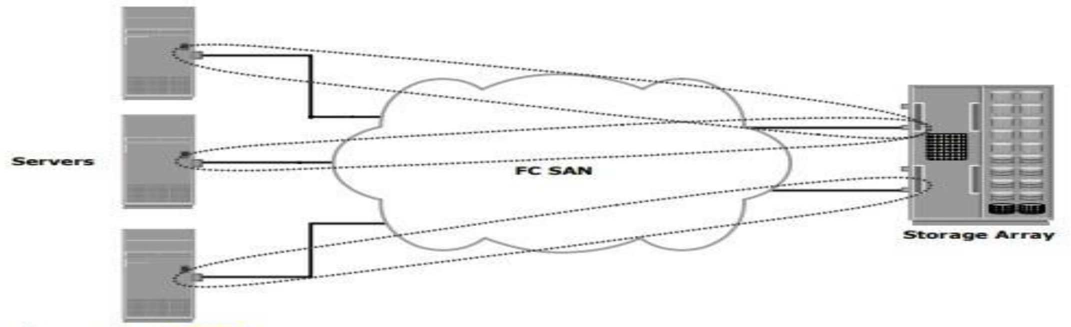


Figure 6-18: Zoning

- Multiple zone sets may be defined in a fabric, but only one zone set can be active at a time. A zone set is a set of zones and a zone is a set of members. A member may be in multiple zones. Members, zones, and zone sets form the hierarchy defined in the zoning process (see Figure 6-19).
- Members are nodes within the SAN that can be included in a zone. Zones comprise a set of members that have access to one another. A port or a node can be a member of multiple zones.
- Zone sets comprise a group of zones that can be activated or deactivated as a single entity in a fabric. Only one zone set per fabric can be active at a time.
- Zone sets are also referred to as zone configurations

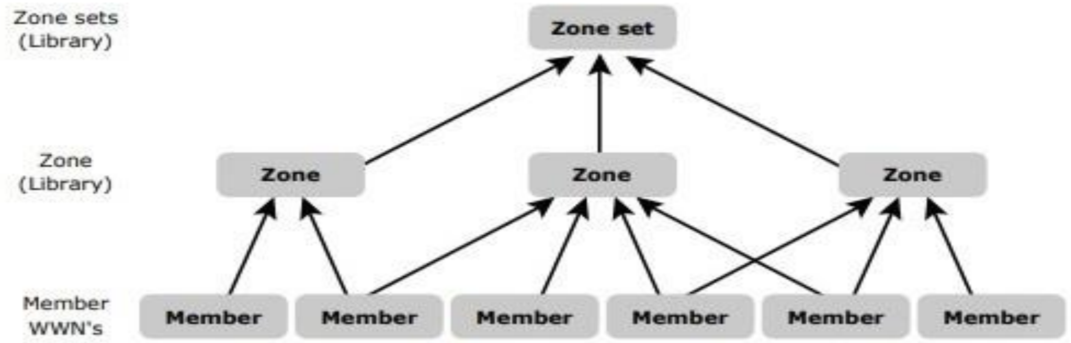


Figure 6-19: Members, zones, and zone sets

Types of Zoning

- ***Zoning can be categorized into three types:***

■ **Port zoning:** It uses the FC addresses of the physical ports to define zones.

In port zoning, access to data is determined by the physical switch port to which anode is connected.

The FC address is dynamically assigned when the port logs on to the fabric.

Therefore, any change in the fabric configuration affects zoning. Port zoning is alsocalled *hard zoning*.

Although this method is secure, it requires updating of zoning configurationinformation in the event of fabric reconfiguration.

■ **WWN zoning:** It uses World Wide Names to define zones. WWN zoning is also referred to as soft zoning. A major advantage of WWN zoning is its flexibility. It allows the SAN to be recabled without reconfiguring the zone information. This is possible because the WWN is static to the node port.

■ **Mixed zoning**: It combines the qualities of both WWN zoning and port zoning. Using mixed zoning enables a specific port to be tied to the WWN of a node.

Figure 6-20 shows the three types of zoning on an FC network.

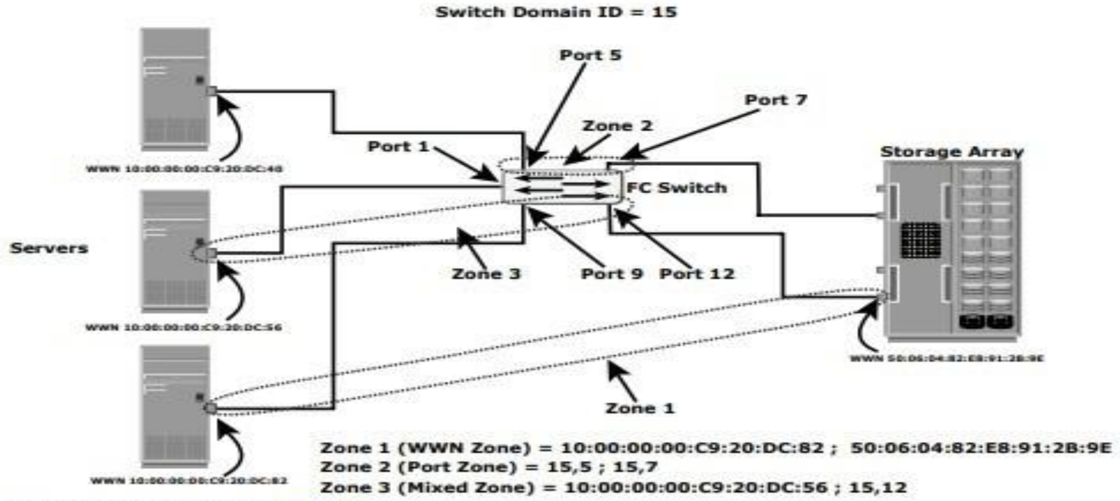

Switch Domain ID = 15
Port 5
Port 7
Zone 2
Port 1
Storage Array
FC Switch
WWN 10:00:00:00:C9:20:DC:40
Servers
Zone 3
Port 9
Port 12
WWN 10:00:00:00:C9:20:DC:56
WWN 50:06:04:82:E8:91:2B:9E
Zone 1
WWN 10:00:00:00:C9:20:DC:82
Zone 1 (WWN Zone) = 10:00:00:00:C9:20:DC:82 ; 50:06:04:82:E8:91:2B:9E
Zone 2 (Port Zone) = 15,5 ; 15,7
Zone 3 (Mixed Zone) = 10:00:00:00:C9:20:DC:56 ; 15,12

Figure 6-20: Types of zoning

- Zoning is used in conjunction with LUN masking for controlling server access to storage. However, these are two different activities.
- Zoning takes place at the fabric level and LUN masking is done at the array level

FC SAN VIRTUALIZATION

- For SAN virtualization, the available virtualization features in the IBM Storage portfolio is described.
- These features enable the SAN infrastructure to support the requirements of scalability and consolidation, combining them with a lower TCO and a higher return on investment (ROI):

IBM b-type Virtua✚Fabrics CISCO Virtual SAN (VSAN)
✚N_Port ID Virtualization (NPIV) support for virtual nodes

IBM b-type Virtual Fabrics

- The Virtual Fabric of the IBM b-type switches is a licensed feature that enables the logical partitioning of SAN switches. When Virtual Fabric is enabled, a default logical switch that uses all of the ports is formed.
- This default logical switch can be then divided into multiple logical switches by grouping them together at a port level. Figure 6-6 shows the flow of Virtual Fabric creation

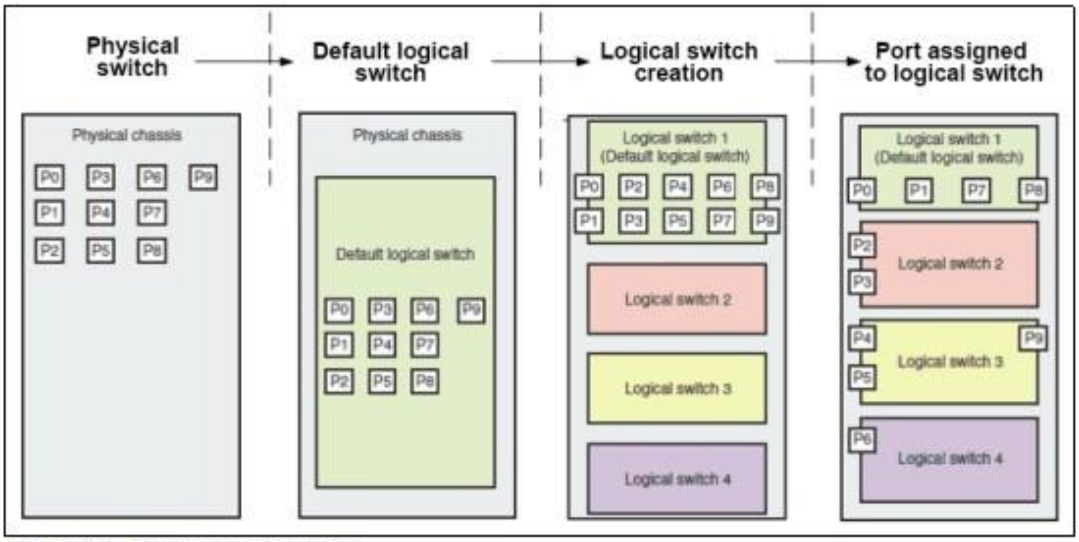


Figure 6-6 Virtual Fabric creation

- Logical fabric When the fabric is formed with at least one logical switch, the fabric iscalled a logical fabric.
- Two methods of fabric connectivity are available for logical fabrics:

❖ A logical fabric is connected with a dedicated inter-switch link (ISL) toanother switch or a logical switch.

Figure 6-7 shows a logical fabric that is formed between logical switchesthrough a dedicated ISL for logical switches.

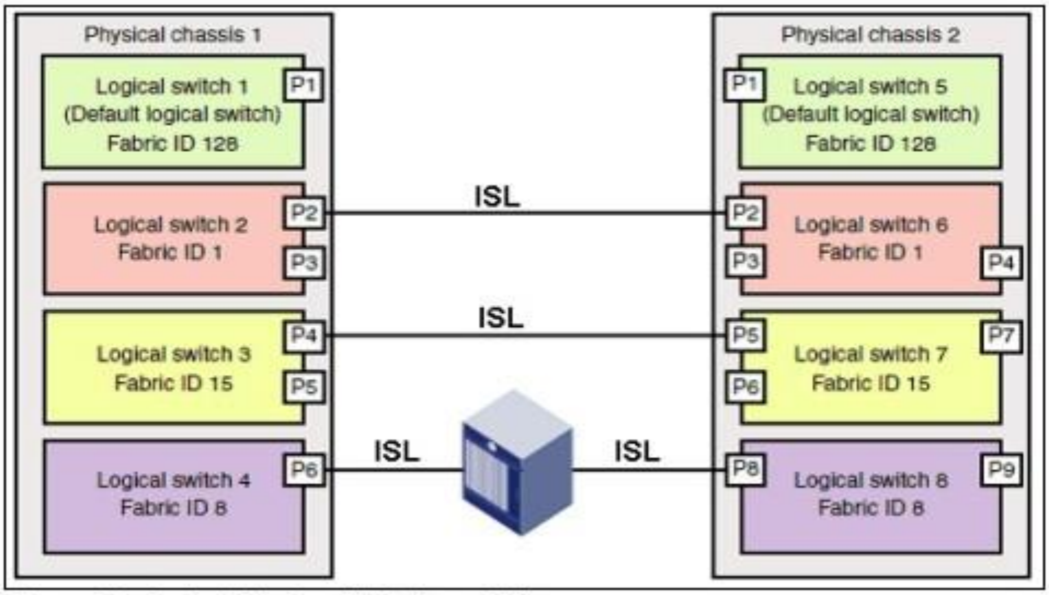

Physical chassis 1
Logical switch 1 (Default logical switch) Fabric ID 128
P1
Logical switch 2 Fabric ID 1
P2
P3
Logical switch 3 Fabric ID 15
P4
P5
Logical switch 4 Fabric ID 8
P6
ISL
ISL
ISL
ISL
Physical chassis 2
Logical switch 5 (Default logical switch) Fabric ID 128
P1
Logical switch 6 Fabric ID 1
P2
P3
P4
Logical switch 7 Fabric ID 15
P5
P6
P7
Logical switch 8 Fabric ID 8
P8
P9

Figure 6-7 Logical fabrics with dedicated ISL

❖ Logical fabrics are connected by using a shared ISL, which is called an extended ISL (XISL), from a base logical switch. In this case, the separate logical switch is configured as a base switch.

This separate logical switch is used only for XISL connectivity and not for device connectivity.

Figure 6-8 shows a logical fabric that is formed through the XISL in the baseswitch

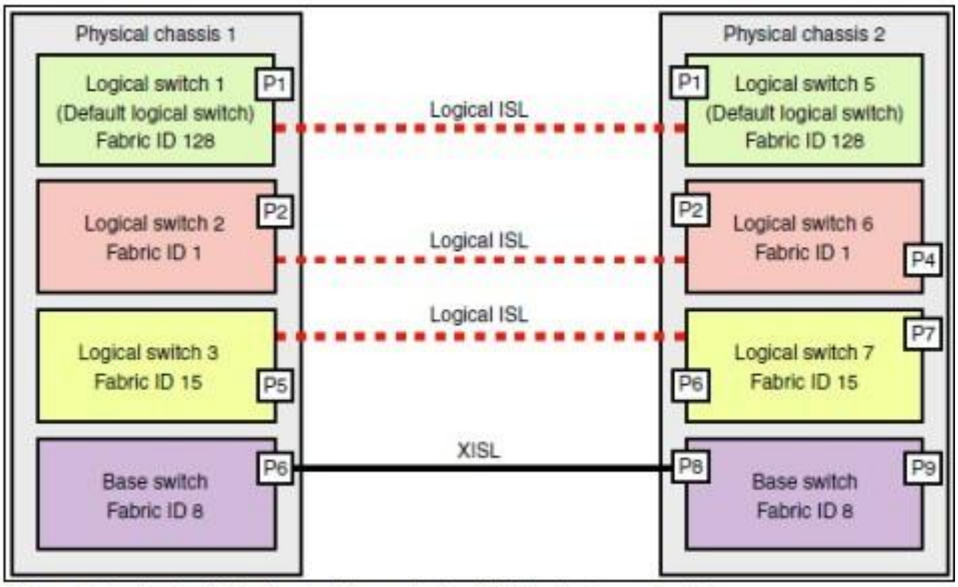


Figure 6-8 Logical ISL formed through the XISL in the base switch

- Cisco virtual storage area network Cisco virtual storage area network (VSAN) is a feature that enables the logical partition of SAN switches. A VSAN provides the flexibility to partition, for example, a dedicated VSAN for disk and tape.
- Or, a VSAN can provide the flexibility to maintain production and test devices in separate VSANs on the same chassis.
- Also, the VSAN can scale across the chassis, which allows it to overcome the fixed port numbers on the chassis.
- Virtual storage area network in a single storage area network switch With VSAN, you can consolidate small fabrics into the same chassis.
- This consolidation can also enable more security by the logical separation of the chassis into two individual VSANs.
- Figure 6-9 shows a single chassis that is divided into two logical VSANs.

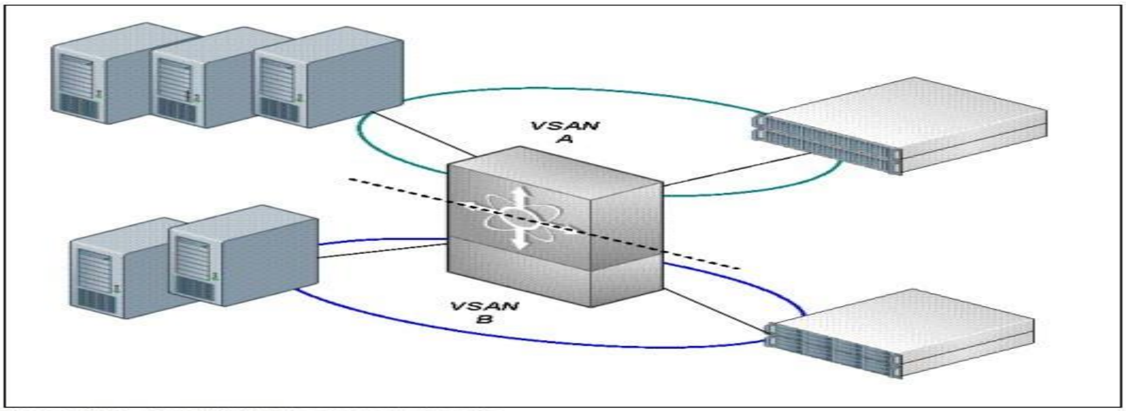


Figure 6-9 Two VSANs in a single chassis

- Virtual storage area network across multiple chassis In multiple chassis, the virtual storage area network (VSAN) can be formed with devices in one chassis to devices in another switch chassis through the extended inter-switch link (XISL).

- *Figure 6-10 shows the VSAN across chassis with an enhanced inter-switch link (EISL) for VSAN communication.*

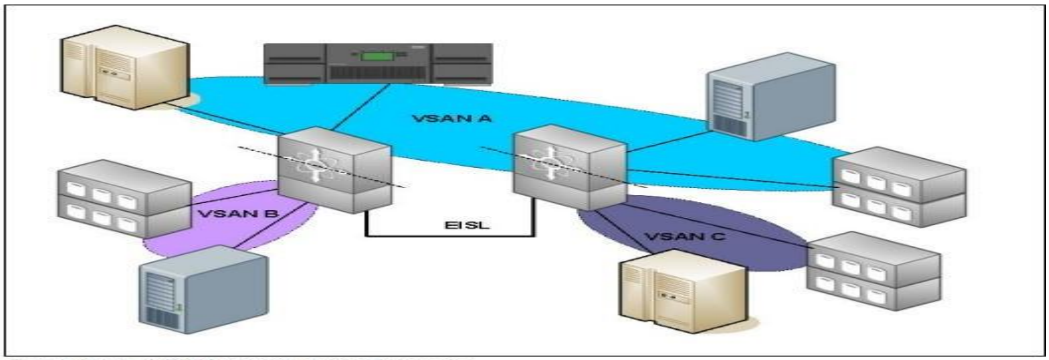


Figure 6-10 VSAN across multiple chassis

- N_Port ID Virtualization Server virtualization with blade servers provides enhanced scalability of servers. This scalability is supported equally in the SAN with N_Port ID Virtualization (NPIV).
- NPIV allows SAN switches to have one port that is shared by many virtual nodes, therefore, supporting a single HBA with many virtual nodes.
- *Figure 6-11 shows sharing a single HBA by multiple virtual nodes. In this case, the same HBA is defined with multiple virtual worldwide node names (WWNNs) and worldwide port names (WWPNs).*

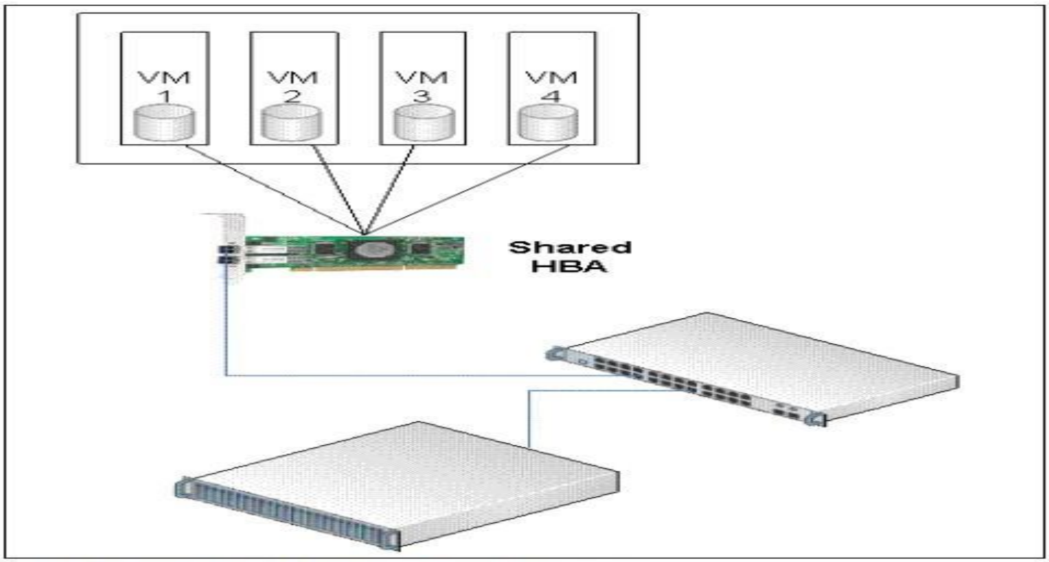


Figure 6-11 Single HBA with multiple virtual nodes

- NPIV mode of blade server switch modules On blade servers, when they are enabled with the NPIV mode, the FC switch modules that connect to an external SAN switch for access to storage act as an HBA N_port (instead of a switch E_port).
- The back-end ports are F_ports that connect to server blade modules.

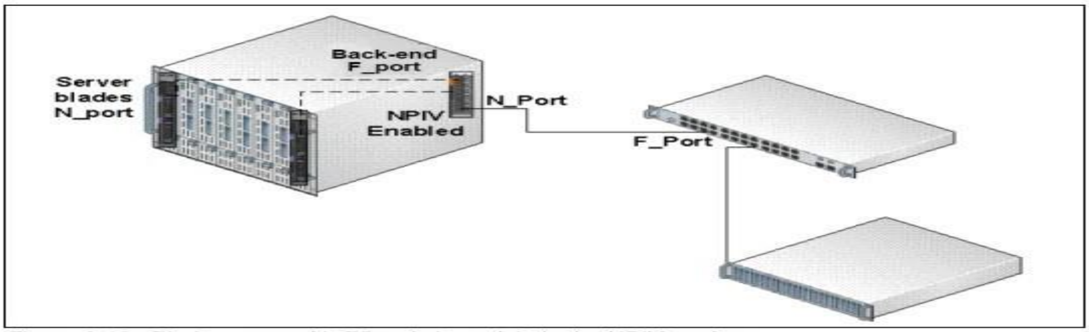


Figure 6-12 Blade server with FC switch module in the NPIV mode

- With the NPIV mode, we can overcome the interoperability issues of merging external switches that might come from separate vendors to the blade server switch module. Also, management is easier because the blade switch module becomes a node in the fabric.
- And, we can overcome the scalability limitations of many switch domains for a switch module in blade servers.

➔ INTERNET PROTOCOL SAN (IP SAN)

- 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 and software.
- 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.
- Figure 8-1 illustrates the co-existence of FC and IP storage technologies in an organization where mission-critical applications are serviced through FC, and business critical applications and remote office applications make use of IP SAN.
- Disaster recovery solutions can also be implemented using both of these technologies.
- Two primary protocols that leverage IP as the transport mechanism are iSCSI and Fibre Channel over IP (FCIP).

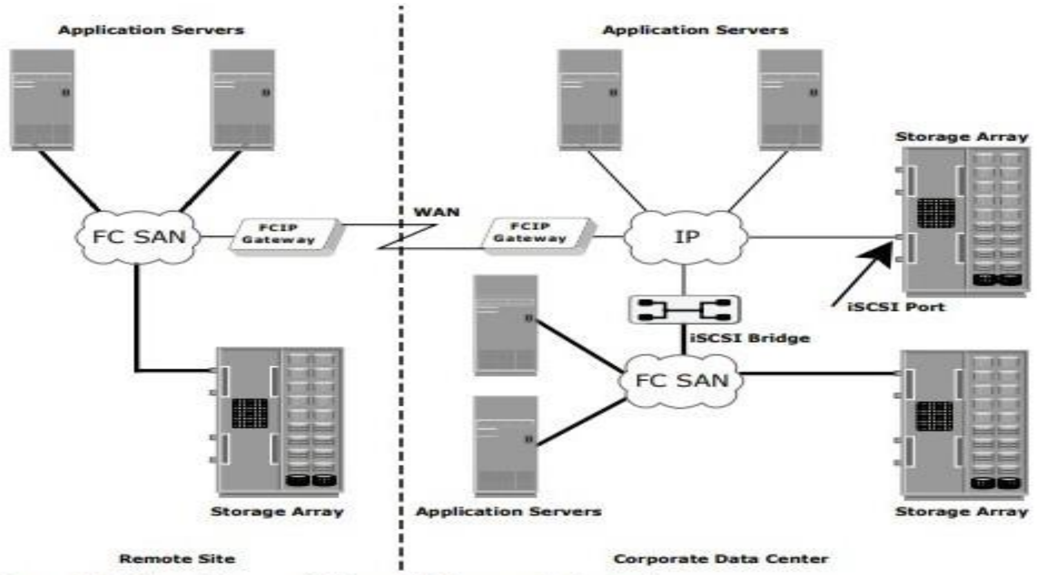
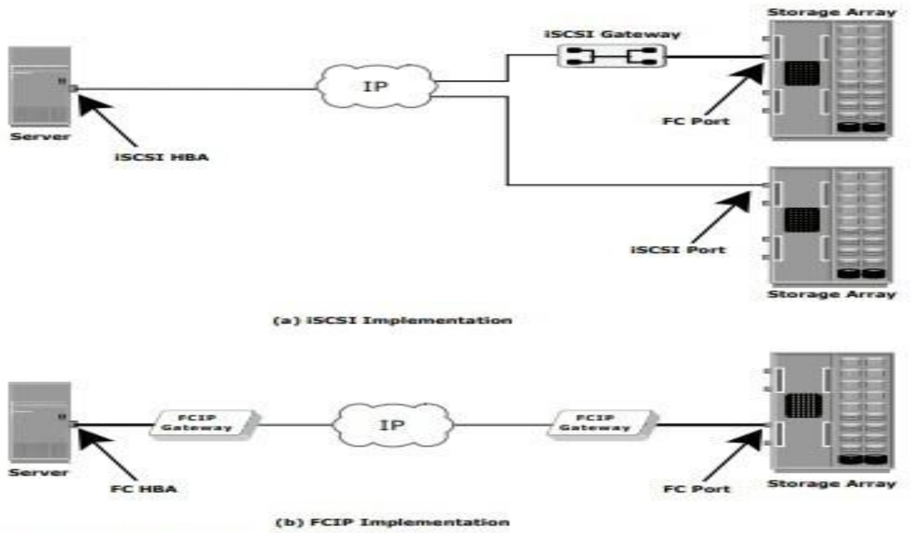


Figure 8-1: Co-existance of FC and IP storage technologies

- iSCSI is the host-based encapsulation of SCSI I/O over IP using an Ethernet NIC card or an iSCSI HBA in the host.
- As illustrated in Figure 8-2 (a), IP traffic is routed over a network either to a gateway device that extracts the SCSI I/O from the IP packets or to an iSCSI storage array.
- The gateway can then send the SCSI I/O to an FC-based external storage array, whereas an iSCSI storage array can handle the extraction and I/O natively. FCIP uses a pair of bridges (FCIP gateways) communicating over TCP/IP as the transport protocol.
- FCIP is used to extend FC networks over distances and/or an existing IP-based infrastructure, as illustrated in Figure 8-2 (b).
- Today, iSCSI is widely adopted for connecting servers to storage because it is relatively inexpensive and easy to implement, especially in environments where an FCSAN does not exist.
- FCIP is extensively used in disaster-recovery implementations, where data is duplicated on disk or tape to an alternate site.



(a) iSCSI Implementation

(b) FCIP Implementation

Figure 8-2: iSCSI and FCIP implementation

➔ iSCSI

- iSCSI is an IP-based protocol that establishes and manages connections between storage, hosts, and bridging devices over IP. iSCSI carries block-level data over IP- based networks, including Ethernet networks and the Internet.
- iSCSI is built on the SCSI protocol by encapsulating SCSI commands and data in order to allow these encapsulated commands and data blocks to be transported using TCP/IP packets.

➔ Components of iSCSI

- Host (initiators), targets, and an IP-based network are the principal iSCSI components. The simplest iSCSI implementation does not require any FC components. If an iSCSI-capable storage array is deployed, a host itself can act as an iSCSI initiator, and directly communicate with the storage over an IP network.
- However, in complex implementations that use an existing FC array for iSCSI connectivity, iSCSI gateways or routers are used to connect the existing FC SAN. These devices perform protocol translation from IP packets to FC packets and vice- versa, thereby bridging connectivity between the IP and FC environments.

➔ iSCSI Host Connectivity

- iSCSI host connectivity requires a hardware component, such as a NIC with a software component (iSCSI initiator) or an iSCSI HBA. In order to use the iSCSI protocol, a software initiator or a translator must be installed to route the SCSIcommands to the TCP/IP stack.
- A standard NIC, a TCP/IP offload engine (TOE) NIC card, and an iSCSI HBA are the three physical iSCSI connectivity options. A standard NIC 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. However, the NIC provides no external processing power, which places additional overhead on the host CPU because it is required to perform all the TCP/IP and iSCSI processing.

- If a standard NIC is used in heavy I/O load situations, the host CPU may become a bottleneck. TOE NIC help alleviate this burden. A TOE NIC offloads the TCP management functions from the host and leaves 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, requiring host CPU cycles.
- An iSCSI HBA is capable of providing performance benefits, as it offloads the entire iSCSI and TCP/IP protocol stack from the host processor.

- Use of an iSCSI HBA is also the simplest way for implementing a boot from SAN environment via iSCSI. If there is no iSCSI HBA, modifications have to be made tothe 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 very similar to the functionality of an FC HBA, but it is the most expensive option.
- A fault-tolerant host connectivity solution can be implemented using host based multipathing software (e.g., EMC Power Path) regardless of the type of physical connectivity. Multiple NICs can also be combined via link aggregation technologies toprovide failover or load balancing.
- Complex solutions may also include the use of vendor-specific storage-array software that enables the iSCSI host to connect to multiple ports on the array with multiple NICs or HBAs.

Topologies for iSCSI Connectivity

- The topologies used to implement iSCSI can be categorized into two classes: native and bridged. Native topologies do not have any FC components; they perform all communication over IP.
- The initiators may be either directly attached to targets or connected using standard IP routers and switches.
- Bridged topologies enable the co-existence 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 inan FC SAN.

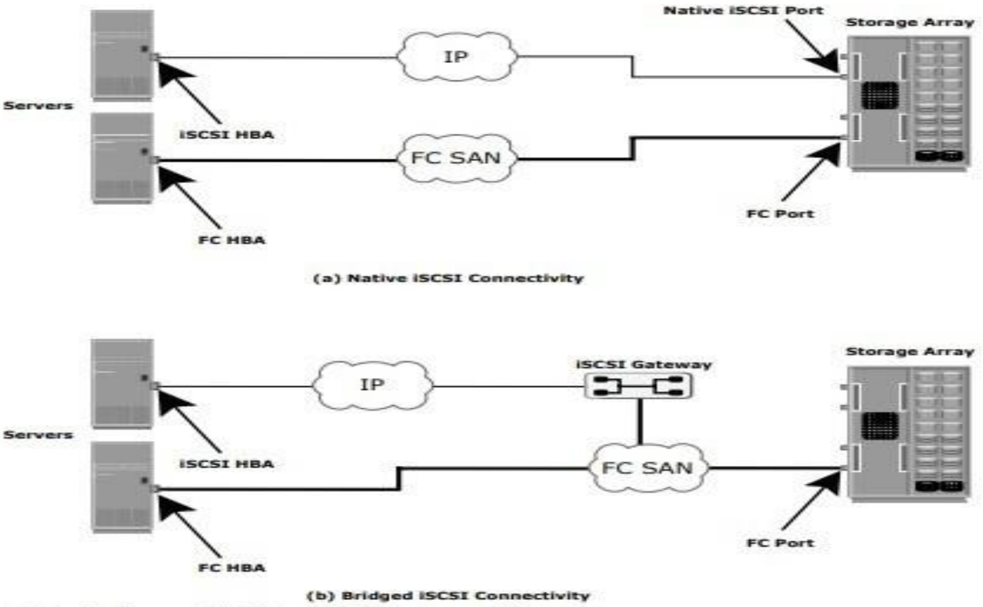
Native iSCSI Connectivity

- If an iSCSI-enabled array is deployed, FC components are not needed for iSCSI connectivity in the native topology.
- In the example shown in Figure 8-3 (a), the array has one or more Ethernet NICs that are connected to a standard Ethernet switch and configured with an IP address and listening port.
- Once a client/ initiator is configured with the appropriate target information, it connects to the array and requests a list of available LUNs.
- A single array port can service multiple hosts or initiators as long as the array can handle the amount of storage traffic that the hosts generate.
- Many arrays provide more than one interface so that they can be configured in a highly available design or have multiple targets configured on the initiator. Some NAS devices are also capable of functioning as iSCSI targets, enabling file-level and block- level access to centralized storage.
- This offers additional storage options for environments with integrated NAS devices or environments that don’t have an iSCSI/FC bridge.

Bridged iSCSI Connectivity

- A bridged iSCSI implementation includes FC components in its configuration.

- Figure 8-3 (b) illustrates an existing FC storage array used to service hosts connected through iSCSI.
- The array does not have any native iSCSI capabilities—that is, it does not have any Ethernet ports.
- Therefore, an external device, called a bridge, router, gateway, or a multi-protocol router, must be used to bridge the communication from the IP network to the FC SAN.
- These devices can be a stand-alone unit, or in many cases are integrated with an existing FC switch. In this configuration, the bridge device has Ethernet ports connected to the IP network, and FC ports connected to the storage.
- These ports are assigned IP addresses, similar to the ports on an iSCSI-enabled array. The iSCSI initiator/host is configured with the bridge's IP address as its target destination.
- The bridge is also configured with an FC initiator or multiple initiators. These are called virtual initiators because there is no physical device, such as an HBA, to generate the initiator record



(a) Native iSCSI Connectivity

(b) Bridged iSCSI Connectivity

Figure 8-3: Native and bridged iSCSI connectivity

Combining FCP and Native iSCSI Connectivity

- A combination topology can also be implemented.
- In this case, a storage array capable of connecting the FC and iSCSI hosts without theneed for external bridging devices is needed (see Figure 8-3 [a]).
- These solutions reduce complexity, as they remove the need for configuring bridges.
- However, additional processing requirements are placed on the storage array becauseit has to accommodate the iSCSI traffic along with the standard FC traffic

➔ iSCSI Protocol Stack

- The architecture of iSCSI is based on the client/server model.
- Figure 8-4 displays a model of the iSCSI protocol layers and depicts the encapsulation order of SCSI commands for their delivery through a physical carrier.

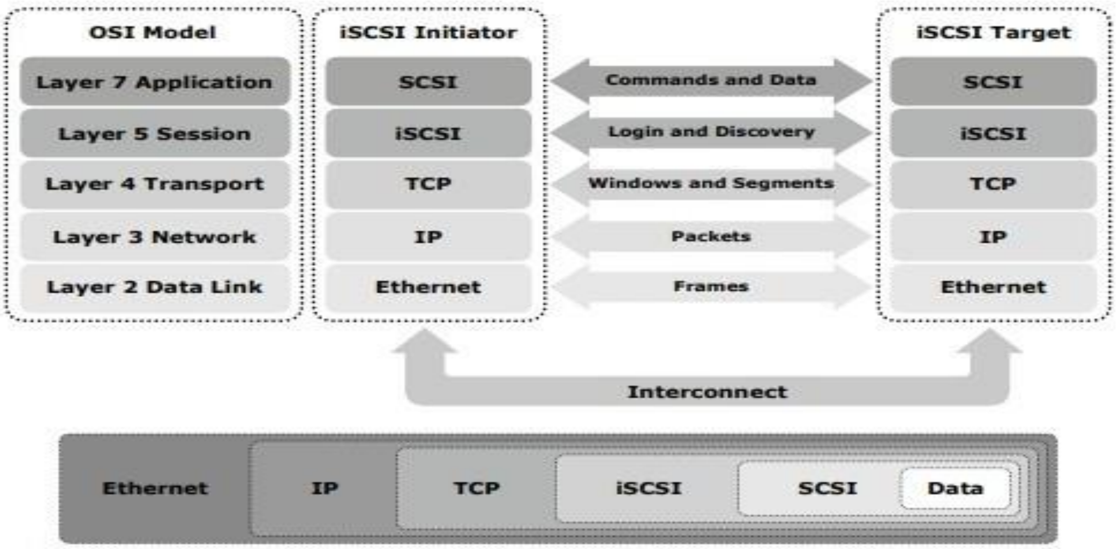
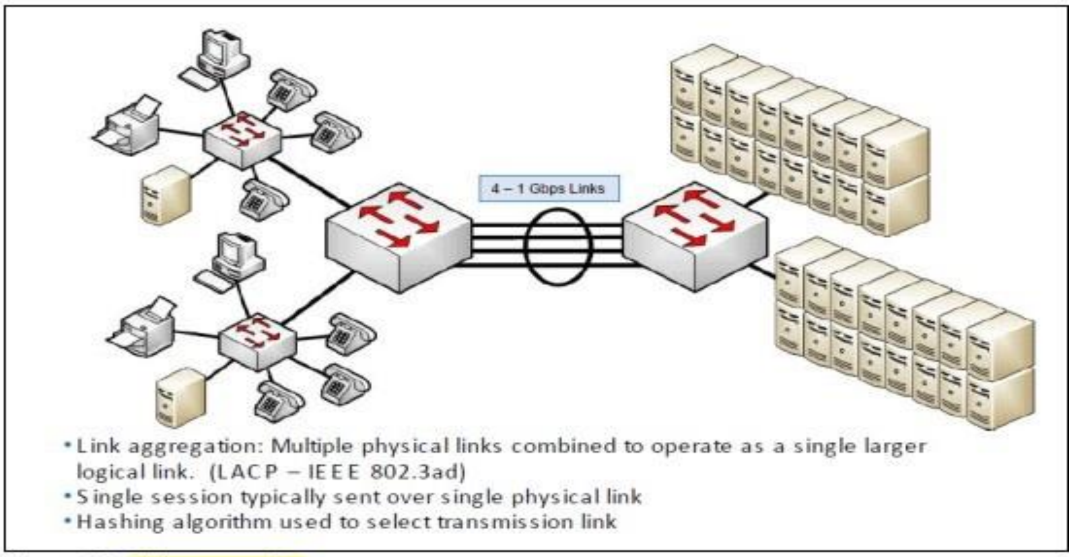


Figure 8-4: iSCSI protocol stack

- SCSI is the command protocol that works at the application layer of the OSI model. The initiators and targets use SCSI commands and responses to talk to each other. The SCSI command descriptor blocks, data, and status messages are encapsulated into TCP/IP and transmitted across the network between initiators and targets.
- iSCSI is the session-layer protocol that initiates a reliable session between a device that recognizes SCSI commands and TCP/IP.
- The iSCSI session-layer interface is responsible for handling login, authentication, target discovery, and session management. TCP is used with iSCSI at the transport layer to provide reliable service.
- TCP is used to control message flow, windowing, error recovery, and retransmission.
- 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 for each hop through a separate physical network.

➔ LINK AGGREGATION

- Link aggregation combines multiple physical links to operate as a single larger logical link.
- The member links no longer function as independent physical connections, but as members of the larger logical link (Figure 4-9).



- Link aggregation: Multiple physical links combined to operate as a single larger logical link. (LACP – IEEE 802.3ad)
- Single session typically sent over single physical link
- Hashing algorithm used to select transmission link

Figure 4-9 Link aggregation

- Link aggregation provides greater bandwidth between the devices at each end of the aggregated link.
- Another advantage of link aggregation is increased availability because the aggregated link is composed of multiple member links.
- If one member link fails, the aggregated link continues to carry traffic over the remaining member links. Each of the devices that is interconnected by the aggregated link uses a hashing algorithm to determine on which of the member links the frames will be transmitted.
- The hashing algorithm might use various information in the frame to make the decision.
- This algorithm might include a source MAC, destination MAC, source IP, destinationIP, and more. It might also include a combination of these values.

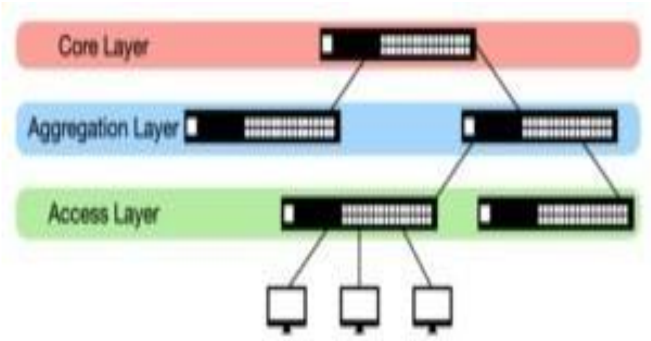
➔ SWITCH AGGREGATION:

- An aggregation switch is a networking device that allows multiple network connections to be bundled together into a single link. This enables increased bandwidth and better network performance.
- Typically, aggregation switches use link aggregation protocols, such as Link Aggregation Control Protocol (LACP) and Ethernet Aggregation to combine multiple links into a single, logical connection.
- Therefore, they can offer great flexibility and scalability, allowing for quick and easy network expansion or reconfiguration.
- In most cases, aggregation switches are used in networks with high-traffic levels or large numbers of users, as they can efficiently distribute data across multiple links.

Role of the Aggregation Switch in the Network

- The aggregation switch is located in the middle of the network architecture, which is equivalent to a middle-level manager of a company.

- It needs to be responsible for managing the data from the lower layer (the access layer switch), and at the same time, it also reports data to the upper layer (the core layer switch).
- Usually, when the aggregation switch receives data from the access switch, it will perform local routing, filtering, traffic balancing, and QoS priority management. Then it will process the security mechanism, IP address translation, and multicast management of the data.
- Finally, it will forward the data to the core layer switch or perform local routing processing according to the processing result to ensure the normal operation of the core layer.
- It can be seen from the above that the aggregation switch has functions such as source address, destination address filtering, real-time policy, security, network isolation, and segmentation.
- Compared with access switches, aggregation switches have better performance and higher switching speeds.



The aggregation switch is located in the middle of the network architecture

- However, in practical applications, some network architectures only have access switches and core switches without aggregation switches.
- The reason is that the network is small, simple, and has a short transmission distance. Users do not deploy aggregation switches to reduce network costs and maintenance burden.
- However, if the number of network users exceeds 200, and the number of users will continue to grow in the future, it is recommended to deploy aggregation switches.

Aggregation Switches vs Access Switches: What's the Difference?

The main distinction between access switches and aggregation switches is their level of operation and performance.

Operational Level. The access switch is located in the network where users can directly connect to or access the network. As for the aggregation switch, it is used to reduce the load on the core layer equipment, and it performs uploading and distributing as well as other functions such as policy implementation, security, and working group access.

Features. Aggregation switches require more performance than access layer switches to handle all traffic from access layer devices, fewer interfaces, and faster switching rates. And the access switch primarily provides adequate bandwidth for access layer access and includes user management functions such as address authentication, user authentication, and user information collection.

Things to Consider When Selecting Aggregation Switches

Backplane Bandwidth and Packet Forwarding Rate

If the backplane bandwidth and packet forwarding rate are limited, the switch's data processing capability will be compromised, resulting in congestion.

As a result, when selecting an aggregation switch, it is sufficient to select the appropriate one based on the actual needs in order to avoid resource waste.

Port Type and Port Number

Data from several access switches must be combined by the aggregation switch before being forwarded to the core switch. The kind and quantity of uplink ports on the network's access switches must be taken into account while choosing the aggregation switch.

When choosing an aggregation switch, it is best to take the network's scale into account. As the network upgrades and expands, the company should select an enterprise-level switch with scalable ports as the aggregation switch.

Port Speed

The uplink and downlink should be considered when determining the aggregation switch port rate. The speeds of the ports can be the same or different.

For instance, when a 10G aggregation switch needs to be interconnected with a 10G access switch, a 10G downlink port must be selected.

Functional Management

Link Aggregation. High-bandwidth aggregation links connecting to core switches are required for aggregation switches. Therefore, link aggregation must be supported by aggregation switches in order to ensure that the access layer has enough bandwidth and that it can continue to function even if one of the links is cut.

Quality of Service (QoS). To assure the quality of service for a certain type of traffic, the QoS priority policy might give it a priority during transmission. The performance and quality of audio and video communications cannot be guaranteed during networktransmission unless you choose an enterprise-level switch that supports QoS as an aggregation switch when you make your purchase.

***Security Strategy*.** The aggregation switch can select an enterprise-level switch that supports security controls to stop unauthorized access and malicious information from entering the network. The ACL (Access Control List) can specify the kinds of traffic communication that are permitted, effectively stop some forms of traffic from being transmitted, and maintain network security.

Redundancy

The network's safety is guaranteed by redundancy capability. Power redundancy is crucial for aggregation switches because it allows them to continue to function even if one of their power supplies fails or needs to be replaced, preventing any disruption to the network's regular operations.

VIRTUAL LOCAL AREA NETWORK (VLAN)

- A virtual local area network (VLAN) is a networking concept in which a network is logically divided into smaller virtual LANs.
- The Layer 2 traffic in one VLAN is logically isolated from other VLANs (Figure 4-5)

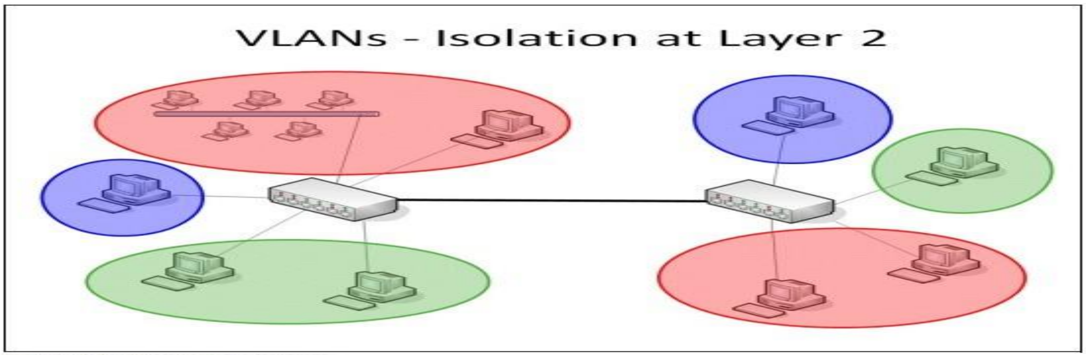
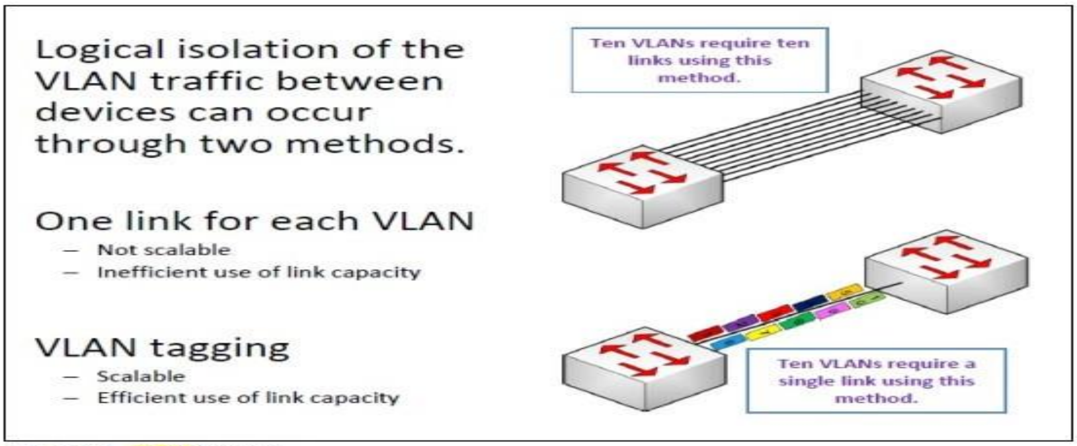


Figure 4-5 Isolation at Layer 2


Logical isolation of the VLAN traffic between devices can occur through two methods.

One link for each VLAN
- Not scalable
- Inefficient use of link capacity

VLAN tagging
- Scalable
- Efficient use of link capacity

Ten VLANs require ten links using this method.

Ten VLANs require a single link using this method.

Figure 4-6 VLAN tagging

- The first method uses a single link for each VLAN. This method does not scale well because it uses many ports in networks that have multiple VLANs and multipleswitches.
- Also, this method does not use link capacity efficiently when traffic in the VLANs is not uniform.
- The second method is VLAN tagging over a single link in which each frame in tagged with its VLAN ID. This method is highly scalable because only a single link is required to provide connectivity to many VLANs.
- This configuration provides for better utilization of the link capacity when VLAN traffic is not uniform.
- The protocol for VLAN tagging of frames in a LAN environment is defined by the IEEE 802.1p/q standard (priority tagging and VLAN identifier tagging).
- Inter-switch link (ISL): ISL is another protocol for providing the VLAN tagging function in a network. This protocol is not compatible with the IEEE 802.1p/q standard.

Tagged frames

- The IEEE 802.1p/q standard provides a methodology for information, such as VLAN membership and priority, that is added to the frame (Figure 4-7).

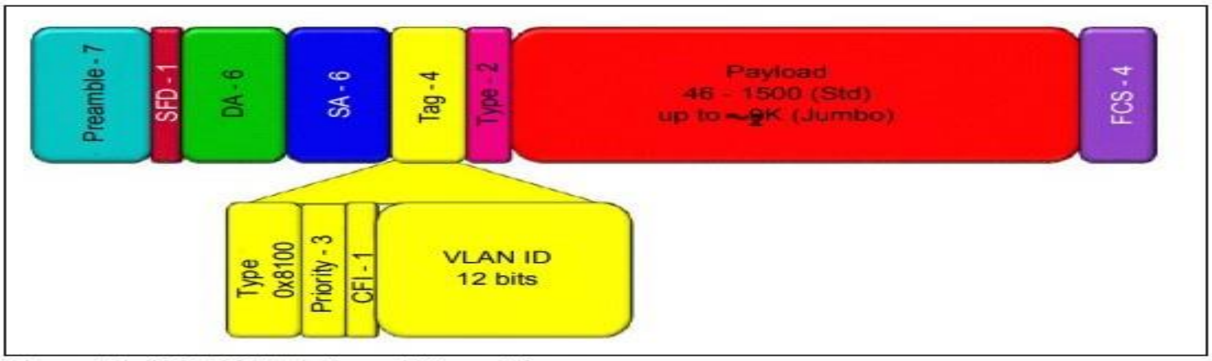


Figure 4-7 IEEE 802.1p/q tagged Ethernet frame

- The standard provides an additional 4 bytes of information to be added to each Ethernet frame. A frame that includes this extra information is known as a tagged frame.
- The 4-byte tag has four component fields:

✓ The type field is 2 bytes and has the hexadecimal value of x8100 to identify the frame as an 802.1p/q tagged frame.

✓ The priority field is 3 bits and allows a priority value of eight different values to be included in the tag. This field has the “P” portion of the 802.1p/qstandard.

✓ The Canonical Format Indicator field is 1 bit and identifies when the contents of the payload field are in canonical format.

✓ The VLAN ID field is 12 bits and identifies the VLAN that the frame is a member of, with 4,096 different VLANs that are possible.

➔ <u>FCIP PROTOCOL</u>

- Organizations are now looking for new ways to transport data throughout the enterprise, locally over the SAN as well as over longer distances, to ensure that data reaches all the users who need it.
- One of the best ways to achieve this goal is to interconnect geographically dispersed SANs through reliable, high-speed links.
- This approach involves transporting FC block data over the existing IP infrastructure used throughout the enterprise.
- The FCIP standard has rapidly gained acceptance as a manageable, cost effective way to blend the best of two worlds: FC block-data storage and the proven, widely deployed IP infrastructure.
- FCIP is a tunneling protocol that enables distributed FC SAN islands to be transparently interconnected over existing IP-based local, metropolitan, and wide-area networks.
- As a result, organizations now have a better way to protect, store, and move their data while leveraging investments in existing technology. FCIP uses TCP/IP as its underlying protocol.
- In FCIP, the FC frames are encapsulated onto the IP payload, as shown in Figure 8-9. FCIP does not manipulate FC frames (translating FC IDs for transmission).

- When SAN islands are connected using FCIP, each interconnection is called an FCIP link.
- A successful FCIP link between two SAN islands results in a fully merged FC fabric.

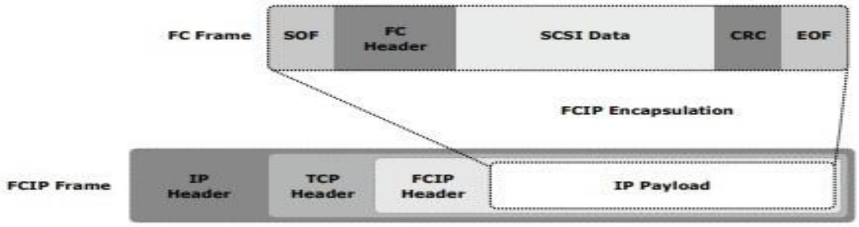


Figure 8-9: FCIP encapsulation

➔ FCIP Topology (FCIP CONNECTIVITY AND CONFIGURATION)

- An FCIP environment functions as if it is a single cohesive SAN environment. Before geographically dispersed SANs are merged, a fully functional layer 2 network exists on the SANs.
- This layer 2 network is a standard SAN fabric.
- These physically independent fabrics are merged into a single fabric with an IP link between them. An FCIP gateway router is connected to each fabric via a standard FC connection (see Figure 8-10).
- The fabric treats these routers like layer 2 fabric switches. The other port on the router is connected to an IP network and an IP address is assigned to that port. This is similar to the method of assigning an IP address to an iSCSI port on a gateway.

- Once IP connectivity is established, the two independent fabrics are merged into a single fabric.
- When merging the two fabrics, all the switches and routers must have unique domain IDs, and the fabrics must contain unique zone set names. Failure to ensure these requirements will result in a segmented fabric.
- The FC addresses on each side of the link are exposed to the other side, and zoning or masking can be done to any entity in the new environment.

-

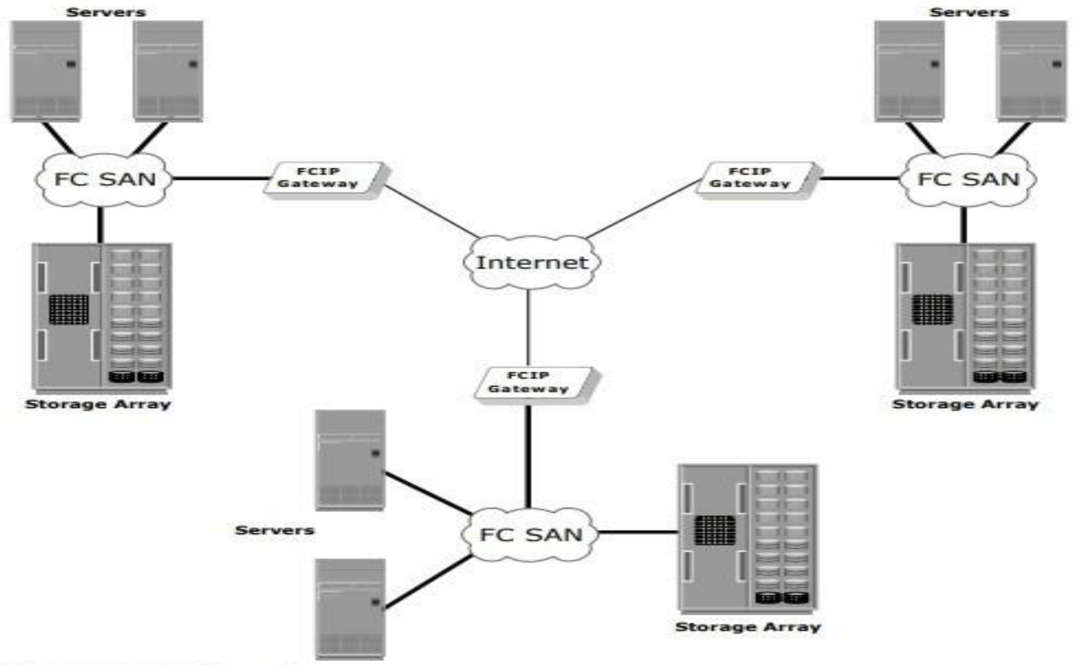


Figure 8-10: FCIP topology

FCIP Performance and Security

- 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 increasedbandwidth.
- In a scenario of extended distance, the IP network may be a bottleneck if sufficient bandwidth is not available.
- 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 host timeouts.
- 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 databased on the router’s support.
- IPSec is one such security measure that can be implemented in the FCIP environment.

➔ FCOE (FIBRE CHANNEL OVER ETHERNET)

- FCoE (Fibre Channel over Ethernet) is a storage protocol that enables Fibre Channel (FC) communications to run directly over Ethernet.

- FCoE makes it possible to move Fibre Channel traffic across existing high-speed Ethernet infrastructure and converges storage and IP protocols onto a single cable transport and interface.
- The goal of FCoE is to consolidate I/O (input/output) and reduce switch complexity, as well as to cut back on cable and interface card counts.
- Adoption of FCoE has been slow, however, due to a scarcity of end-to-end FCoE devices and a reluctance on the part of many organizations to change the way they implement and manage their networks.
- Traditionally, organizations have used Ethernet for Transmission Control Protocol/Internet Protocol (TCP/IP) networks and FC for storage networks.
- Fibre Channel supports high-speed data connections between computing devices that interconnect servers with shared storage devices and between storage controllers and drives.
- FCoE shares Fibre Channel and Ethernet traffic on the same physical cable or lets organizations separate Fibre Channel and Ethernet traffic on the same hardware.
- FCoE uses a lossless Ethernet fabric and its own frame format.
- It retains Fibre Channel's device communications but substitutes high-speed Ethernet links for Fibre Channel links between devices.

➔ FCOE SAN COMPONENTS

The key FCoE SAN components are:

✓ Network adapters such as Converged Network Adapter (CNA) and softwareFCoE adapter

✓ Cables such as copper cables and fiber optical cables

✓ FCoE switch

Converged Network Adapter (CNA)

- The CNA is a physical adapter that provides the functionality of both a standard NIC and an FC HBA in a single device.
- It consolidates both FC traffic and regular Ethernet traffic on a common Ethernet infrastructure.
- FC traffic onto Ethernet frames and forwarding them to FCoE switches over CEE links.
- They eliminate the need to deploy separate adapters and cables for FC and Ethernet communications, thereby reducing the required number of network adapters and switch ports.
- A CNA offloads the FCoE protocol processing task from the compute system, thereby freeing the CPU resources of the compute system for application processing.
- It contains separate modules for 10 Gigabit Ethernet (GE), FC, and FCoE Application Specific Integrated Circuits (ASICs).

Software FCoE Adapter

- Instead of a CNA, a software FCoE adapter may also be used. A software FCoE adapter is OS or hypervisor kernel-resident software that performs FCoE processing.

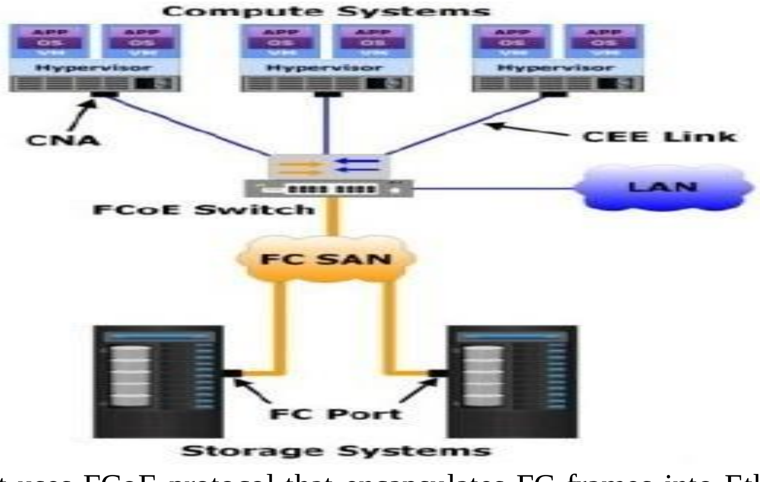
- The FCoE processing consumes hosts CPU cycles.
- With software FCoE adapters, the OS or hypervisor implements FC protocol insoftware that handles SCSI to FC processing.
- The software FCoE adapter performs FC to Ethernet encapsulation. Both FCoE traffic (Ethernet traffic that carries FC data) and regular Ethernet traffic are transferred through supported NICs on the hosts.

FCOE Switch

- An FCoE switch has both Ethernet switch and FC switch functionalities. It has a Fibre Channel Forwarder (FCF), an Ethernet Bridge, and a set of ports that can be used for FC and Ethernet connectivity.
- FCF handles FCoE login requests, applies zoning, and provides the fabric services typically associated with an FC switch.
- It also encapsulates the FC frames received from the FC port into the Ethernet frames and decapsulates the Ethernet frames received from the Ethernet Bridge to the FC frames.
- Upon receiving the incoming Ethernet traffic, the FCoE switch inspects the Ethertype of the incoming frames and uses that to determine their destination.
- If the Ethertype of the frame is FCoE, the switch recognizes that the frame contains anFC payload and then forwards it to the FCF.
- From there, the FC frame is extracted from the Ethernet frame and transmitted to the FC SAN over the FC ports.
- If the Ethertype is not FCoE, the switch handles the traffic as usual Ethernet trafficand forwards it over the Ethernet ports.

➔ CONVERGED ENHANCED ETHERNET (CEE) (or)(FCoE SAN CONNECTIVITY)

- FCoE SAN is a Converged Enhanced Ethernet (CEE) network that is capable of transporting FC data along with regular Ethernet traffic over high speed (such as 10 Gbps or higher) Ethernet links.



- It uses FCoE protocol that encapsulates FC frames into Ethernet frames. FCoE protocol is defined by the T11 standards committee.

- FCoE is based on an enhanced Ethernet standard that supports Data Center Bridging (DCB) functionalities (also called CEE functionalities). DCB ensures lossless transmission of FC traffic over Ethernet.
- FCoE SAN provides the flexibility to deploy the same network components for transferring both server-to-server traffic and FC storage traffic. This helps to mitigate the complexity of managing multiple discrete network infrastructures. FCoE SAN uses multi-functional network adapters and switches.
- Therefore, FCoE reduces the number of network adapters, cables, and switches, along with power and space consumption required in a data center.

Some of the CEE enhancements to Ethernet include:

Priority Flow Control: Focused on developing a standard mechanism that can control the flow for each traffic class of service independently. The idea is to ensure zero loss when a traffic class gets congested in data center bridging networks.

Data Center Bridging exchange: Focused on developing a standard mechanism that can ensure interoperability.

Priority-Based Packet Scheduling: Used to develop a standard mechanism to set scheduling priorities for a set of traffic classes.

Some of the advantages of CEE convergence include:

- CEE enhances the network-attached storage and Internet small computer interface by offering traffic differentiation at the link layer.
- Fiber channel over CEE enables new servers in the data center to use a single link for both Ethernet and fiber channel communications, thereby reducing cable costs.
- CEE technology can be used to converge a variety of applications such as local area networks, storage area networks and high-performance computing.

➔ FCoE ARCHITECTURE

- Fibre Channel over Ethernet (FCoE) is a method of supporting converged Fibre Channel (FC) and Ethernet traffic on a data center bridging (DCB) network.
- FCoE encapsulates unmodified FC frames in Ethernet to transport the FC frames over a physical Ethernet network.
- An FCoE frame is the same as any other Ethernet frame because the Ethernet encapsulation provides the header information needed to forward the frames. However, to achieve the lossless behavior that FC transport requires, the Ethernet network must conform to DCB standards.
- DCB standards create an environment over which FCoE can transport native FC traffic encapsulated in Ethernet while preserving the mandatory *class of service* (CoS) and other characteristics that FC traffic requires.
- Supporting FCoE in a DCB network requires that the FCoE devices in the Ethernet network and the FC switches at the edge of the SAN network handle both Ethernet and native FC traffic. To handle Ethernet traffic, an FC switch does one of two things:

- ✓ Incorporates FCoE interfaces.

✓ Uses an FCoE-FC gateway such as a QFX3500 switch to de-encapsulate FCoE traffic from FCoE devices into native FC and to encapsulate native FC traffic from the FC switch into FCoE and forward it to FCoE devices through the Ethernet network.

FCoE Devices

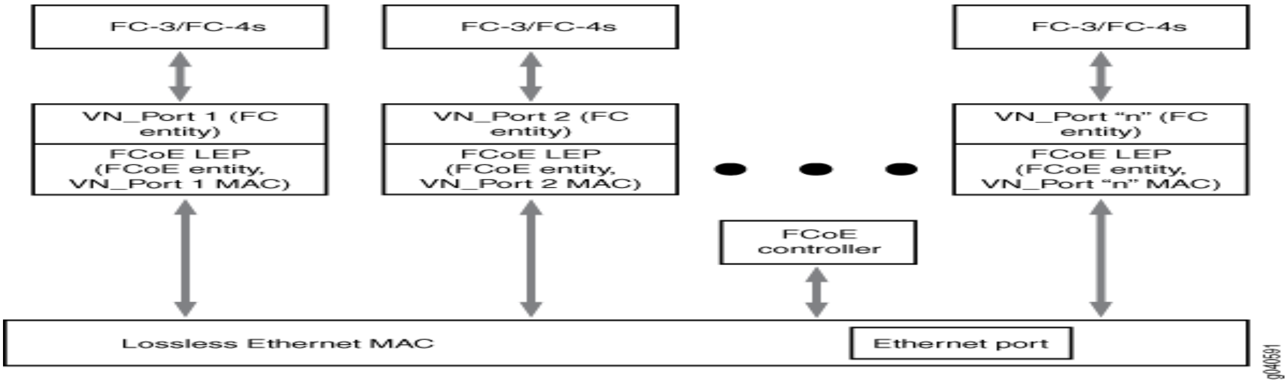
- Each FCoE device has a converged network adapter (CNA) that combines the functions of an FC host bus adapter (HBA) and a lossless Ethernet network interface card (NIC) with 10-Gbps Ethernet ports.
- The portion of the CNA that handles FCoE traffic is called an FCoE Node (ENode). An ENode combines FCoE termination functions and the client part of the FC stackon the CNA.
- ENodes present virtual FC interfaces to FC switches in the form of virtual N_Ports (VN_Ports). A VN_Port is an endpoint in a virtual point-to-point connection called a virtual link.
- The other endpoint of the virtual link is an FC switch (or FCF) port. A VN_Port emulates a native FC N_Port and performs similar functions: handling the creation, detection, and flow of messages to and from the FC switch.
- A single ENode can host multiple VN_Ports. Each VN_Port has a separate, unique virtual link with a FC switch.
- ENodes contain at least one lossless Ethernet media access controller (MAC). Each Ethernet MAC is paired with an FCoE controller. The lossless Ethernet MAC is a full-duplex Ethernet MAC that implements Ethernet extensions to avoid frame loss due to congestion and supports frames of at least 2500 bytes.
- The FCoE controller instantiates and terminates VN_Port instances dynamically as they are needed for FCoE sessions. Each VN_Port instance has a unique virtual linkto an FC switch.
- Nodes also contain one FCoE link end point (LEP) for each VN_Port connection. An FCoE LEP is a virtual FC interface mapped onto the physical Ethernet interface.
- An FCoE LEP:

✓ Transmits and receives FCoE frames on the virtual link.

✓ Handles FC frame encapsulation for traffic going from the server to the FCswitch.

✓ Performs frame de-encapsulation of traffic received from the FC switch.

Figure 1 shows a block diagram of the major ENode components.



FCoE Frames

- The FCoE protocol specification replaces the FC0 and FC1 layers of the FC stack with Ethernet, but retains the FC frame header. Retaining the FC frame headerenables the FC frame to pass directly to a native FC SAN after de-encapsulation.
- The FCoE header carries the FC start of file (SOF) bits and end of file (EOF) bits inan encoded format. FCoE supports two frame types, control frames and data frames. FCoE Initialization Protocol (FIP) carries all of the discovery and fabric login frames.
- FIP control frames handle FCoE device discovery, initializing communication, and maintaining communication.
- They do not carry a data payload. FIP has its own EtherType (0x8914) to distinguish FIP traffic from FCoE traffic and other Ethernet traffic.
- To establish communication, the ENode uses the globally unique MAC address assigned to it by the CNA manufacturer.
- After FIP establishes a connection between FCoE devices, the FCoE data frames handle the transport of the FC frames encapsulated in Ethernet.
- FCoE also has its own EtherType (0x8906) to distinguish FCoE frames from other Ethernet traffic and ensure the in-order frame handling that FC requires. FCoE framesinclude:

✓ 2112 bytes FC payload

✓ 24 bytes FC header

✓ 14 bytes standard Ethernet header

✓ 14 bytes FCoE header

✓ 8 bytes cyclic redundancy check (CRC) plus EOF

✓ 4 bytes VLAN header

✓ 4 bytes frame check sequence (FCS)

- The payload, headers, and checks add up to 2180 bytes. Therefore, interfaces that carry FCoE traffic should have a configured maximum transmission unit (MTU) of 2180 or larger. An MTU size of 2180 bytes is the minimum size; some network administrators prefer an MTU of 2240 or 2500 bytes.

Virtual Links

- Native FC uses point-to-point physical links between FC devices. In FCoE, virtual links replace the physical links.

- A virtual link emulates a point-to-point link between two FCoE device endpoints,such as a server VN_Port and an FC switch (or FCF) VF_Port.
- Each FCoE interface can support multiple virtual links.
- The MAC addresses of the FCoE endpoints (the VN_Port and the VF_Port) uniquely identify each virtual link and allow traffic for multiple virtual links to share the same physical link while maintaining data separation and security.
- A virtual link exists in one FCoE VLAN and cannot belong to more than one VLAN. Although the FC switch and the FCoE device detect a virtual link as a point-to-point connection, virtual links do not need to be direct connections between a VF_Port anda VN_Port.
- A virtual link can traverse one or more transit switches, also known as passthrough switches.
- A transit switch can transparently aggregate virtual links while still appearing and functioning as a point-to-point connection to the FCoE devices. However, a virtual link must remain within a single Layer 2 domain.

FCoE VLANs

- All FCoE traffic must travel in a VLAN dedicated to transporting only FCoE traffic. Only FCoE interfaces should be members of an FCoE VLAN. Ethernet traffic that is not FCoE or FIP traffic must travel in a different VLAN.
- FCoE traffic cannot use a standard LAG because traffic might be hashed to different physical LAG links on different transmissions. This breaks the (virtual) point-to-point link that Fibre Channel traffic requires.
- If you configure a standard LAG interface for FCoE traffic, FCoE traffic might be rejected by the FC SAN.