ECE590-03
Enterprise Storage Architecture

Fall 2018

Storage Area Network (SAN)

Tyler Bletsch

Duke University

Adapted from the course "Information Storage and Management v2"
(module 5-6), published by EMC corporation.

Includes additional content cited inline.

Fibre Channel SAN

Module 5: Fibre Channel Storage Area Network (FC SAN)

Lesson 1: Overview of FC SAN

During this lesson the following topics are covered:

- Evolution of FC SAN
- Components of FC SAN
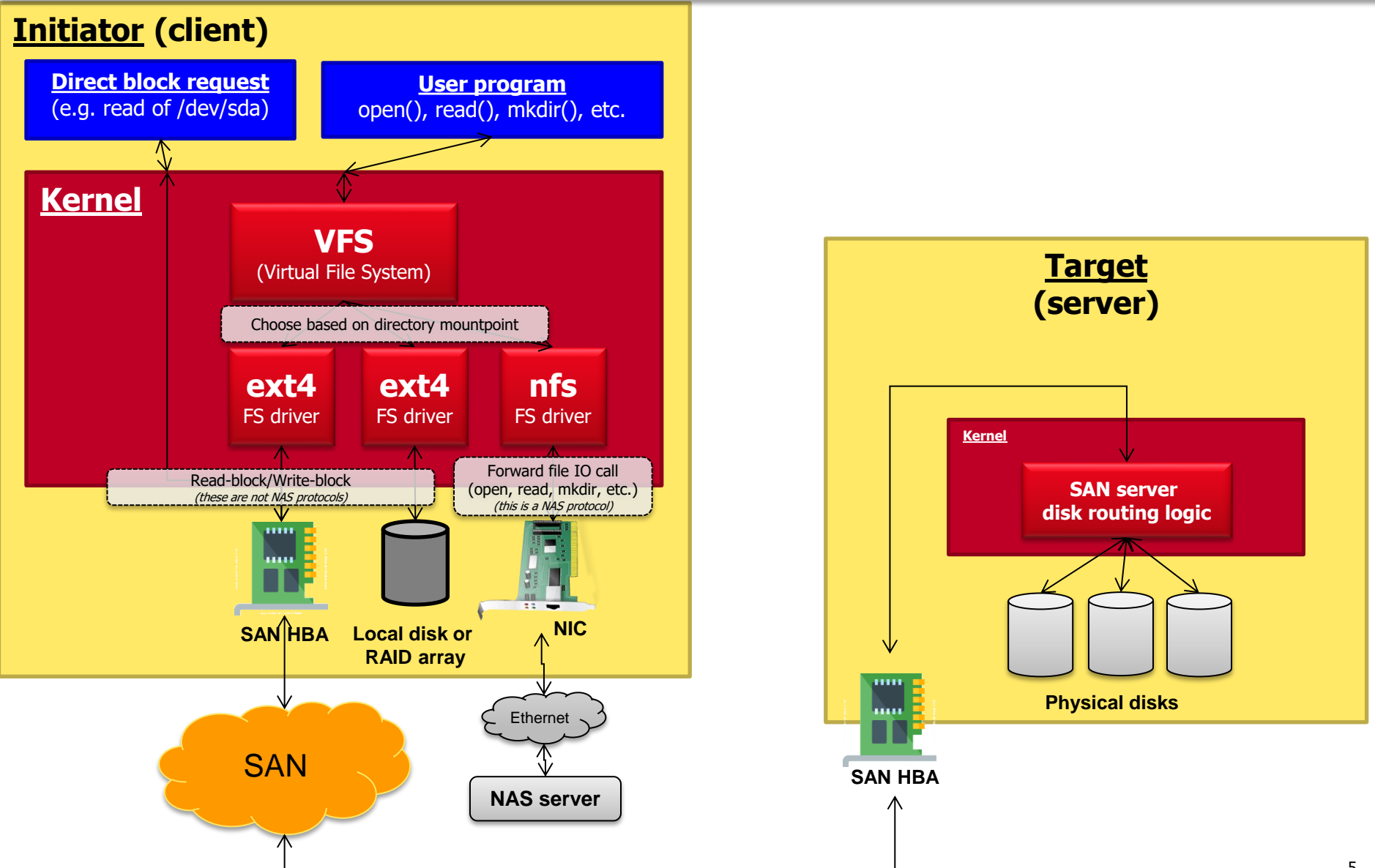- FC interconnectivity options
- FC port types

What is a SAN?

SAN

It is a high-speed, dedicated network of servers and shared storage devices.

- Centralizes storage and management
- Enables sharing of storage resources across multiple servers at block level
- Meets increasing storage demands efficiently with better economies of scale
- Common SAN deployments are:
  - Fibre Channel (FC) SAN: uses FC protocol for communication
  - IP SAN: uses IP-based protocols for communication

SAN block diagram

Initiator (client)

Direct block request
(e.g. read of /dev/sda)

User program
open(), read(), mkdir(), etc.

Kernel

VFS
(Virtual File System)

Choose based on directory mountpoint

ext4
FS driver

ext4
FS driver

nfs
FS driver

Read-block/Write-block
(these are not NAS protocols)

Forward file IO call
(open, read, mkdir, etc.)
(this is a NAS protocol)

SAN HBA

Local disk or RAID array

NIC

SAN

Ethernet

NAS server

Target (server)

Kernel

SAN server disk routing logic

Physical disks

SAN HBA

Understanding Fibre Channel

- High-speed network technology
  - ▸ Latest FC implementation supports speed up to 16 Gb/s
- Highly scalable
  - ▸ Theoretically, accommodate approximately 15 million devices

**Clients**　　　　**Application Servers**　　　　**Storage Arrays**
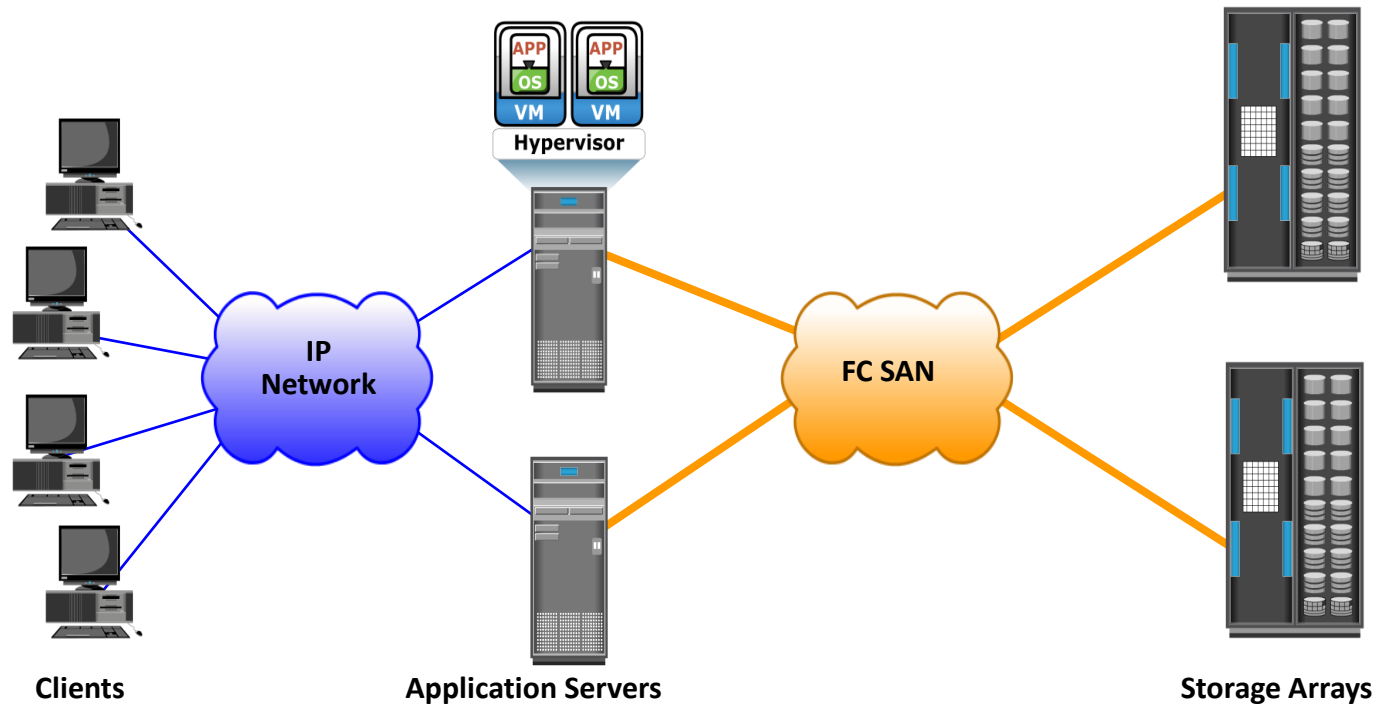
Components of FC SAN

- Node (server and storage) ports
- Cables
- Connectors
- Interconnecting devices such as FC switches and hubs
- SAN management software

Node Ports

- Provide physical interface for communicating with other nodes
- Exist on
  - HBA in server
  - Front-end adapters in storage
- Each port has a transmit (Tx) link and a receive (Rx) link

Node

Port 0

Port 1

Port n

Port 0

Tx

Rx

Link
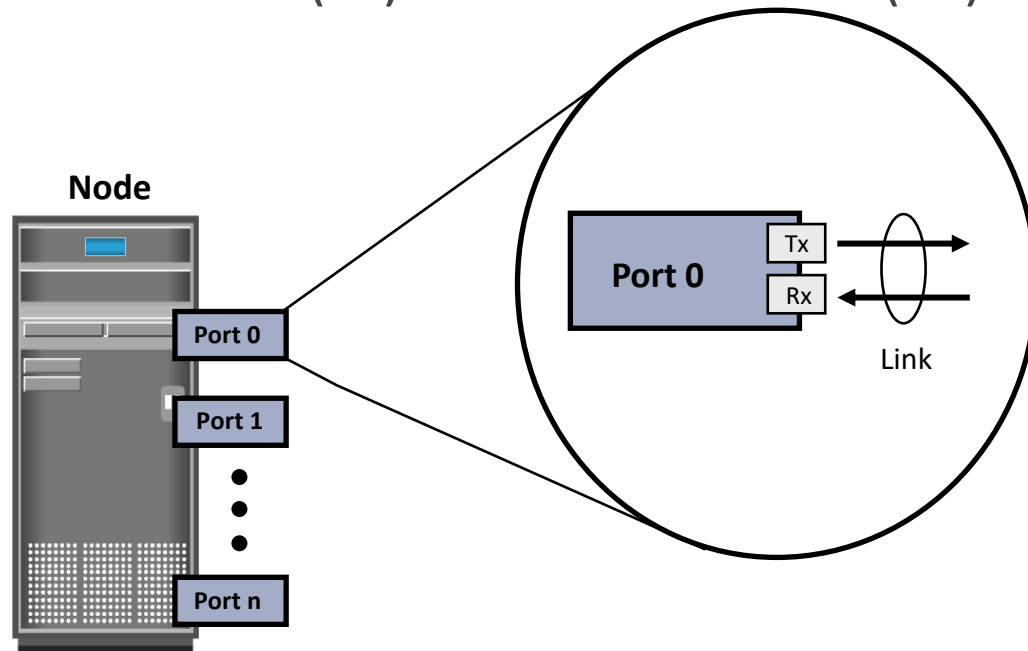
Cables

- SAN implementation uses
  - ▸ Copper cables for short distance
  - ▸ Optical fiber cables for long distance
- Two types of optical cables: single-mode and multimode

Single-mode	Multimode
Carries single beam of light	Can carry multiple beams of light simultaneously
Distance up to 10km	Used for short distance (Modal dispersion weakens signal strength after certain distance)

Cladding   **Core**

Light In →

Single-mode Fiber

Cladding   **Core**

Light In →

Multimode Fiber
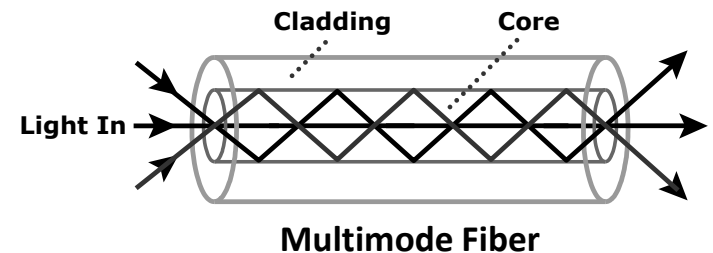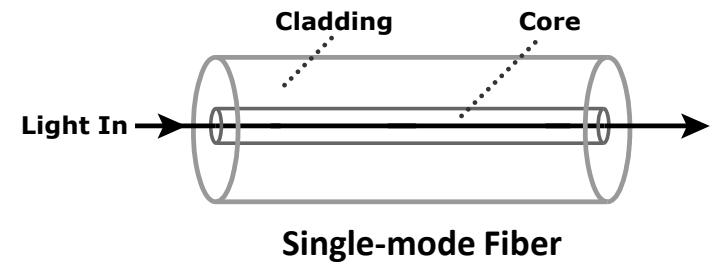
Connectors

- Attached at the end of a cable
- Enable swift connection and disconnection of the cable to and from a port
- Commonly used connectors for fiber optic cables are:
  - Standard Connector (SC)
    - Duplex connectors
  - Lucent Connector (LC)
    - Duplex connectors
  - Straight Tip (ST)
    - Patch panel connectors
    - Simplex connectors

Standard Connector

Lucent Connector

Straight Tip Connector

Interconnecting Devices

- Commonly used interconnecting devices in FC SAN are:
  - ▶ Hubs, switches, and directors
- Hubs provide limited connectivity and scalability
- Switches and directors are intelligent devices
  - ▶ Switches are available with fixed port count or modular design
  - ▶ Directors are always modular, and its port count can be increased by inserting additional 'line cards' or 'blades'
  - ▶ High-end switches and directors contain redundant components

SAN Management Software

- A suite of tools used in a SAN to manage interfaces between host and storage arrays

- Provides integrated management of SAN environment

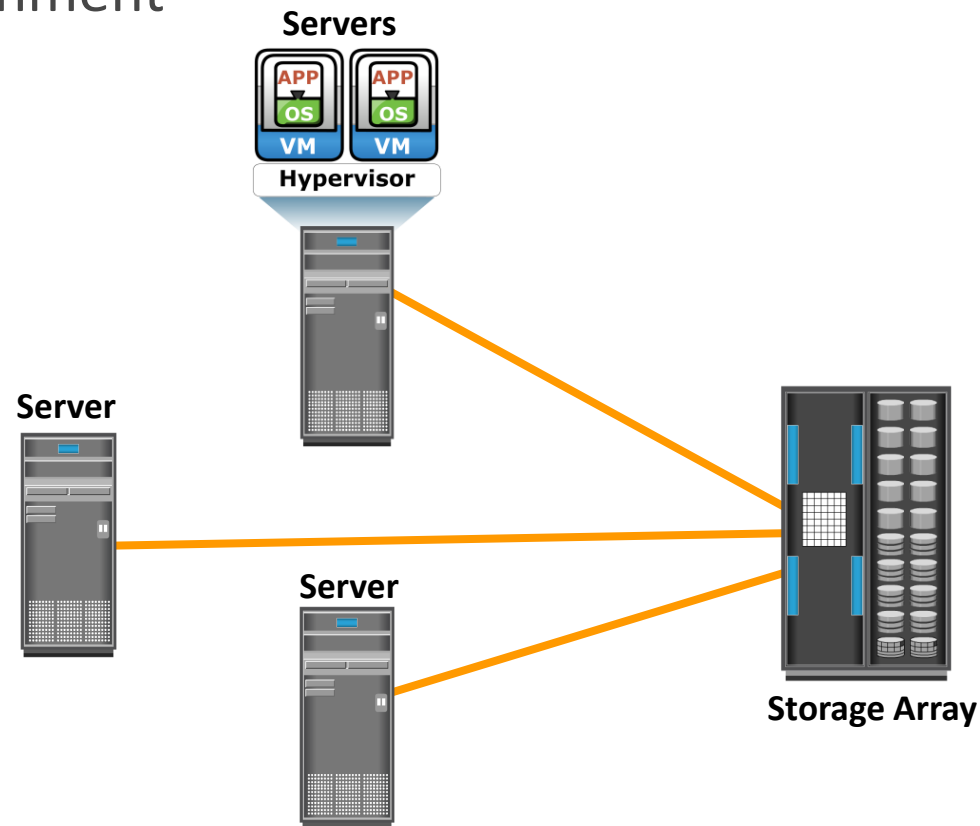- Enables web-based management using GUI or CLI
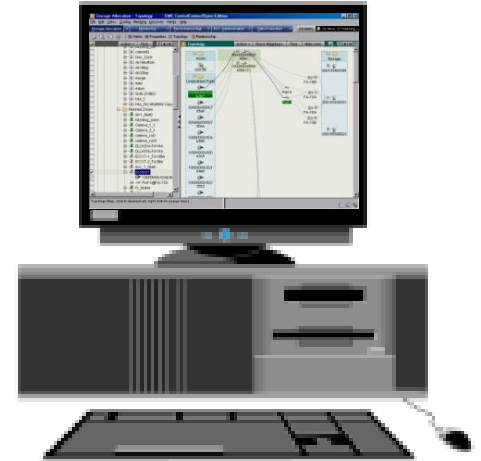
FC Interconnectivity Options

- Point-to-Point

- Fibre Channel Arbitrated Loop (FC-AL)

- Fibre Channel Switched Fabric (FC-SW)

Point-to-Point Connectivity

- Enables direct connection between nodes
- Offers limited connectivity and scalability
- Used in DAS environment

FC-AL Connectivity

- Provides shared loop to attached nodes
  - ▶ Nodes must arbitrate to gain control
- Implemented using ring or star topology
- Limitations of FC-AL
  - ▶ Only one device can perform I/O operation at a time
  - ▶ Supports up to 126 nodes
  - ▶ Addition or removal of a node causes momentary pause in loop traffic

Servers

Server

Hypervisor

FC Hub

Storage Array
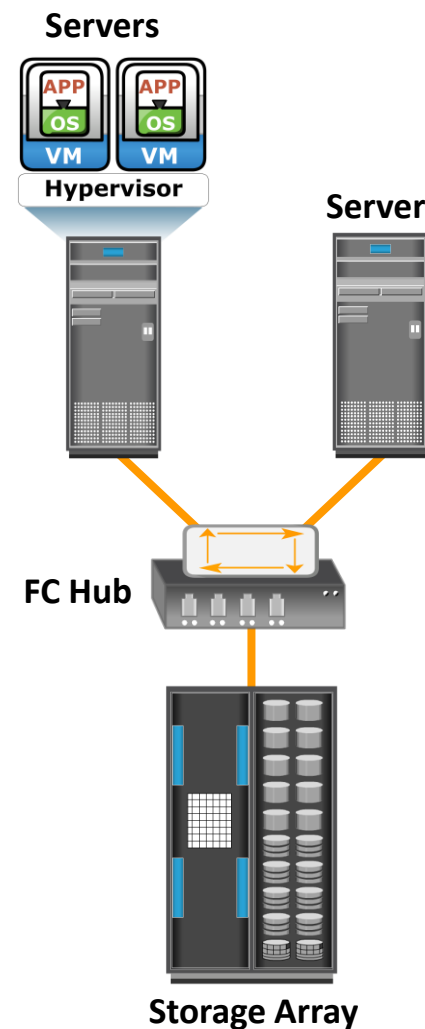
FC-SW Connectivity

- Creates a logical space (called fabric) in which all nodes communicate with one another using switches
  - ▸ Interswitch links (ISLs) enable switches to be connected together
- Provides dedicated path between nodes
- Addition/removal of node does not affect traffic of other nodes

Servers

Storage Array

FC Switch

FC Switch

Server

Interswitch Links
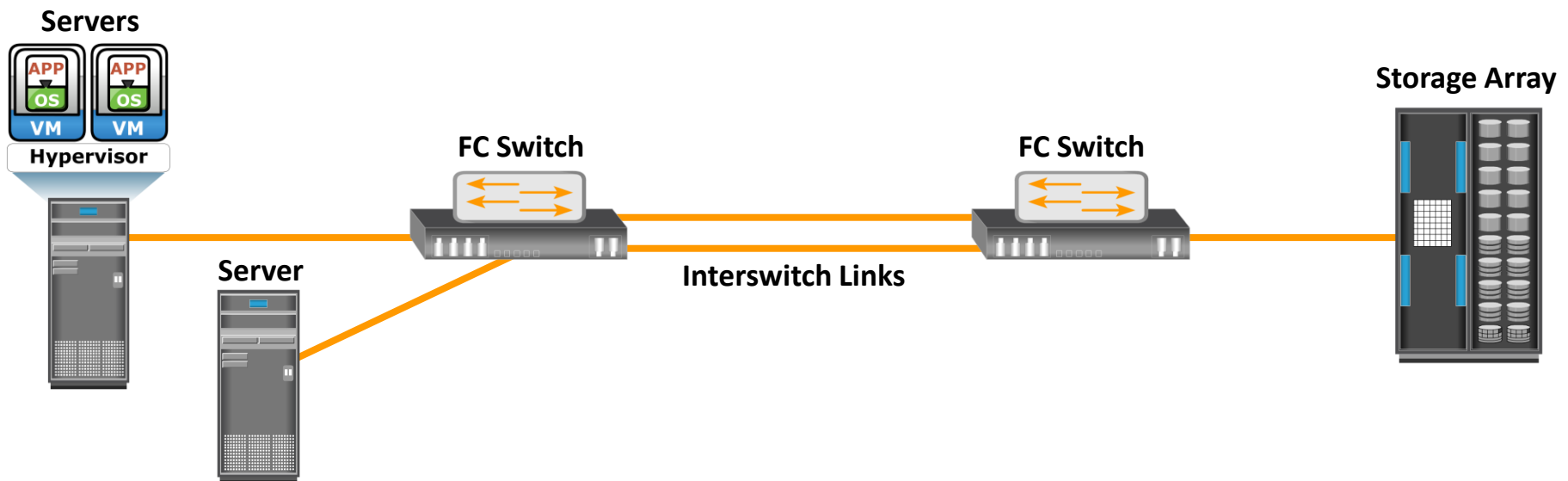
Terminology

- **Initiator**: Client using block devices
- **Target**: Storage controller with the data
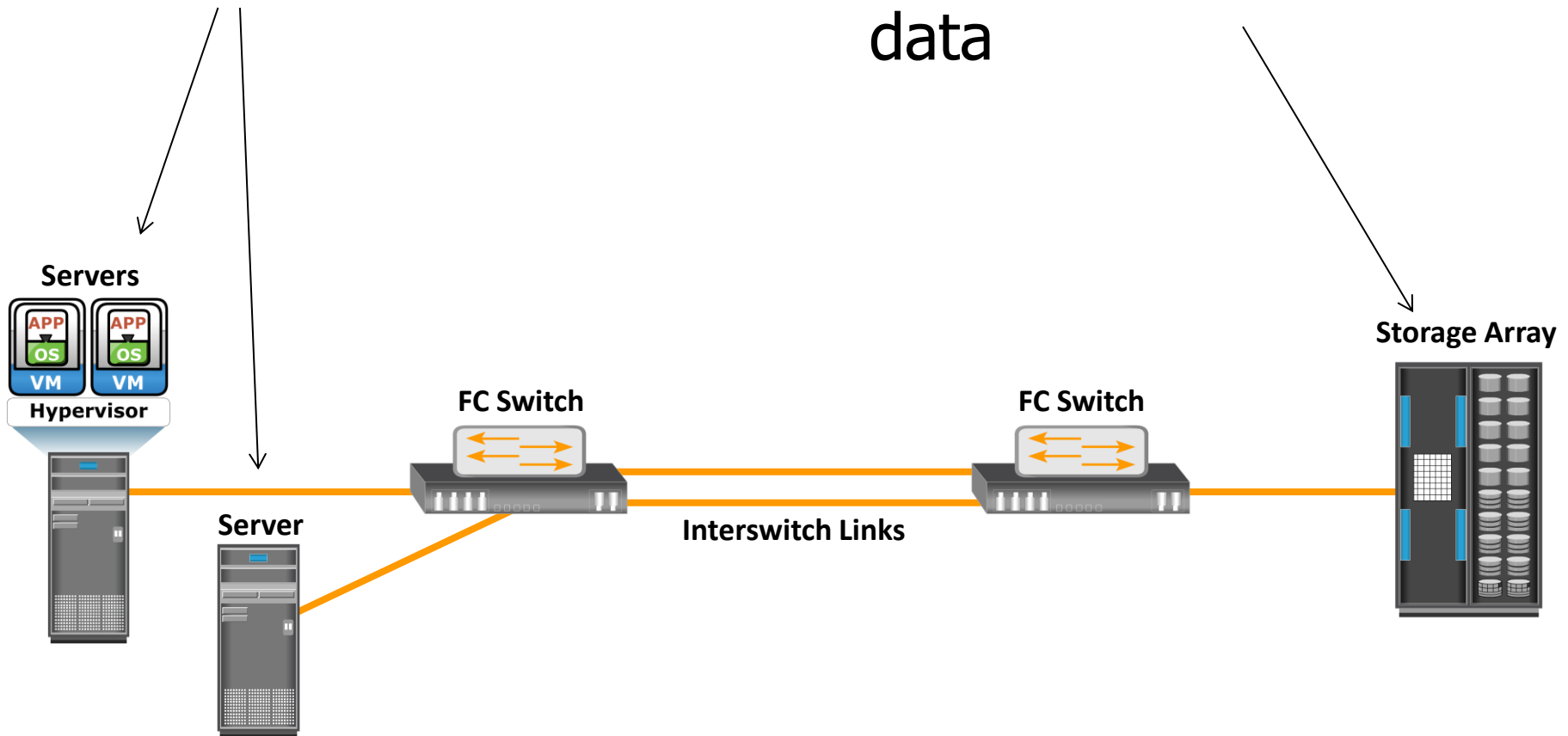
Module 5: Fibre Channel Storage Area Network (FC SAN)

## Lesson 2: Fibre Channel (FC) Architecture

During this lesson the following topics are covered:

- FC protocol stack
- FC addressing
- WWN addressing
- Structure and organization of FC data
- Fabric services
- Fabric login types

FC Architecture Overview

- Provides benefits of both channel and network technologies
  - ▸ Provides high performance with low protocol overheads
  - ▸ Provides high scalability with long distance capability
- Implements SCSI over FC network
  - ▸ Transports SCSI data through FC network
- Storage devices, attached to SAN, appear as local storage devices to host operating system

Fibre Channel Protocol Stack
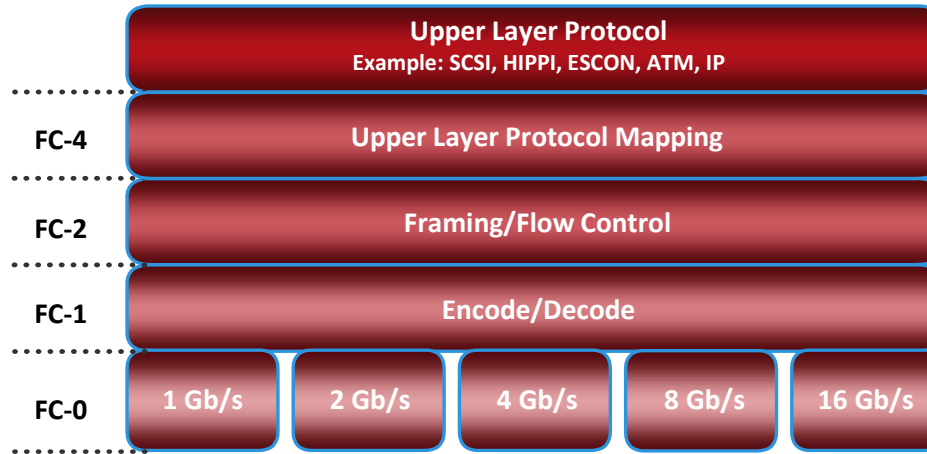
FC Layer	Function	Features Specified by FC Layer
FC-4	Mapping interface	Mapping upper layer protocol (e.g. SCSI) to lower FC layers
FC-3	Common services	Not implemented
FC-2	Routing, flow control	Frame structure, FC addressing, flow control
FC-1	Encode/decode	8b/10b or 64b/66b encoding, bit and frame synchronization
FC-0	Physical layer	Media, cables, connector

World Wide Name (WWN)

- Unique 64 bit identifier
- Static to node ports on an FC network
  - ▸ Similar to MAC address of NIC
  - ▸ World Wide Node Name (WWNN) and World Wide Port Name (WWPN) are used to uniquely identify nodes and ports respectively

World Wide Node Name (for the 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
Format Type	Company ID 24 bits						Port	Model Seed 32 bits							

World Wide Port Name (for the HBA port)															
1	0	0	0	0	0	0	0	c	9	2	0	d	c	4	0
Format Type	Reserved 12 bits			Company ID 24 bits						Company Specific 24 bits					

Fabric Services

- FC switches provide fabric services as defined in FC standards

Fabric Services	Description
Fabric Login Server	• Used during the initial part of the node's fabric login process • Located at pre-defined address of FFFFFE
Name Server	• Responsible for name registration and management of node ports • Located at pre-defined address FFFFFC
Fabric Controller	• Responsible for managing and distributing Registered State Change Notifications (RSCNs) to attached node ports • Responsible for distributing SW-RSCNs to every other switch – SW-RSCNs keep the name server up-to-date on all switches • Located at pre-defined address FFFFFD
Management Server	• Enables FC SAN management using fabric management software • Located at pre-defined address FFFFFA

Login Types in Switched Fabric

- Fabric login (FLOGI)

  ▸ Occurs between an endpoint and switch

  ▸ Node sends a FLOGI frame with WWN to Fabric Login Server on switch

  ▸ Node obtains FC address from switch

  ▸ Immediately after FLOGI, endpoint registers with Name Server on switch, indicating its WWN, port type, assigned FC address, etc.

  ▸ Endpoint queries name server about all other logged in ports

# Module 5: Fibre Channel Storage Area Network (FC SAN)

Lesson 3: FC SAN Topologies and Zoning

During this lesson the following topics are covered:

- Mesh and core-edge topologies
- Benefits of zoning
- Types of zoning

Mesh Topology

- Full mesh
  - ▸ Each switch is connected to every other switch
  - ▸ Maximum of one ISL or hop is required for host-to-storage traffic
  - ▸ Host and storage can be connected to any switch
- Partial mesh
  - ▸ Not all the switches are connected to every other switch

Full Mesh

FC Switches

Server

Storage Array

Partial Mesh

FC Switches

Server

Storage Array
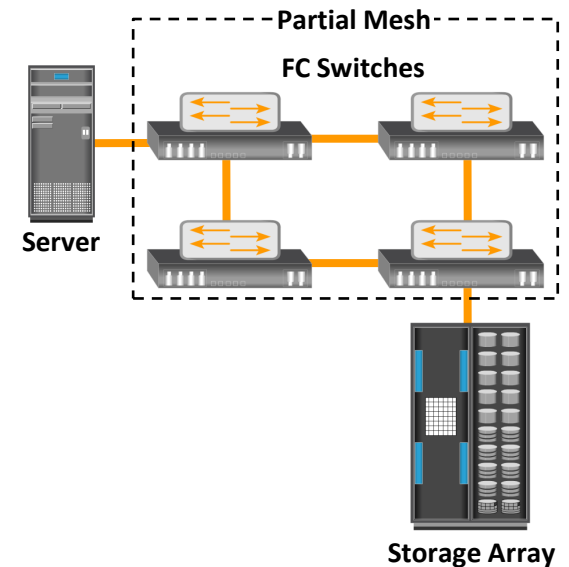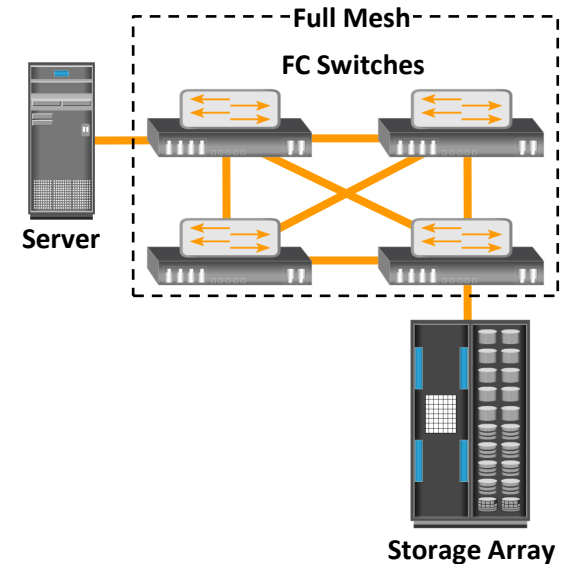
Core-edge Topology

- Consists of edge and core switch tiers
- Network traffic traverses core tier or terminate at core tier
- Storage is usually connected to the core tier
- Benefits
  - High availability
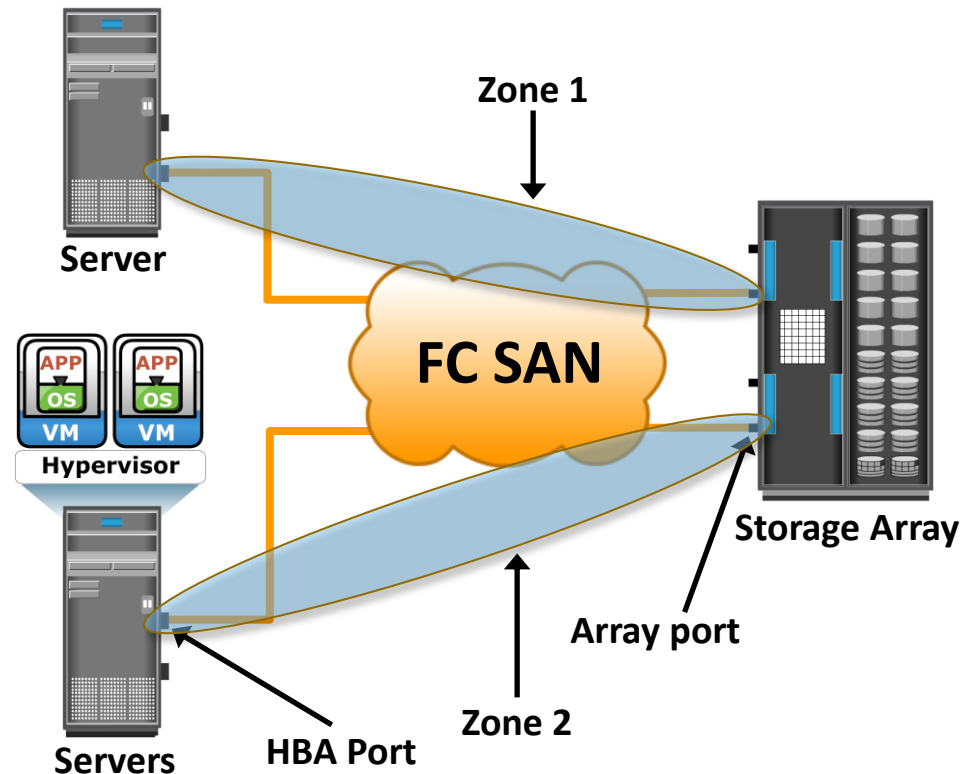 - Medium scalability
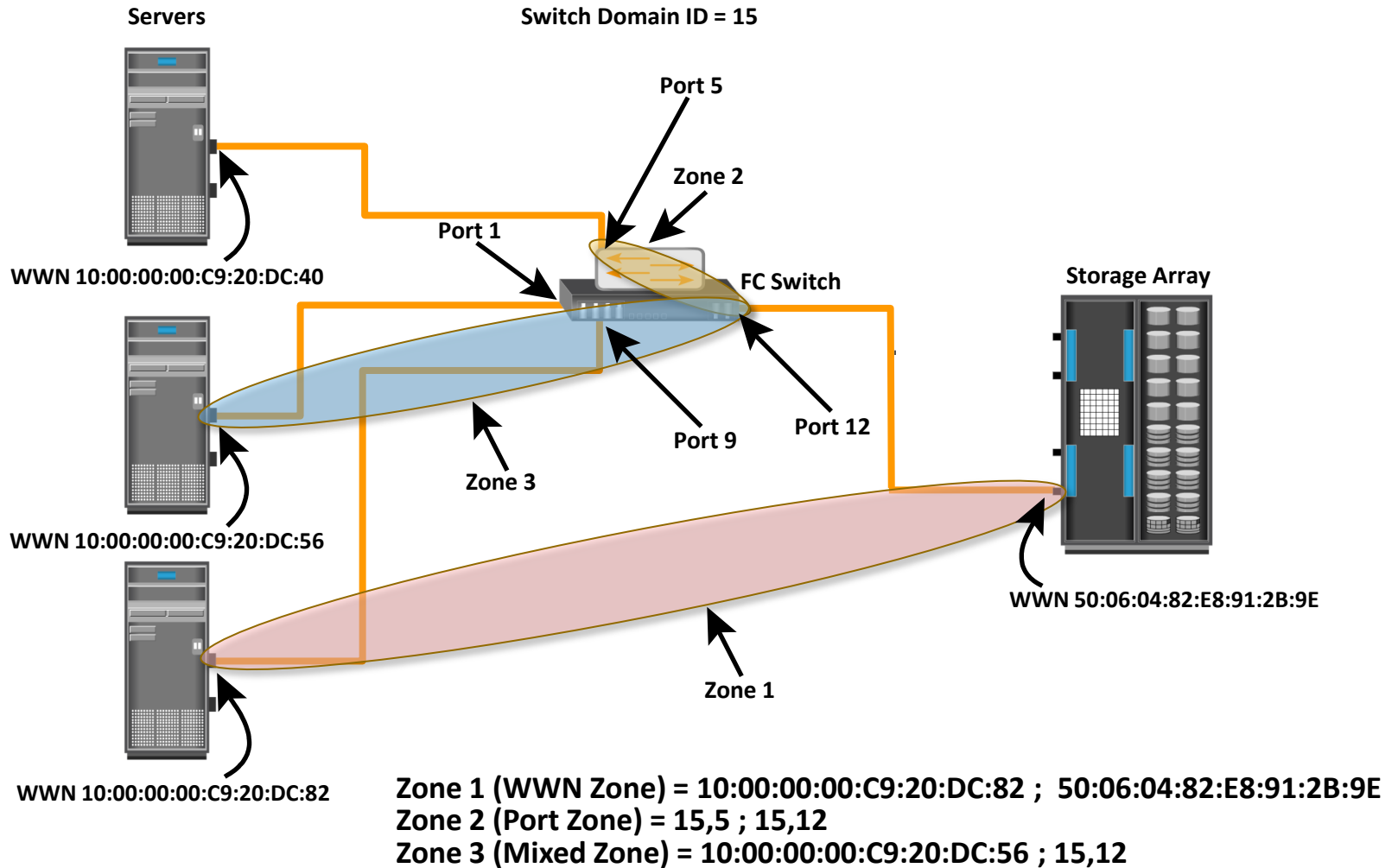 - Medium to maximum connectivity
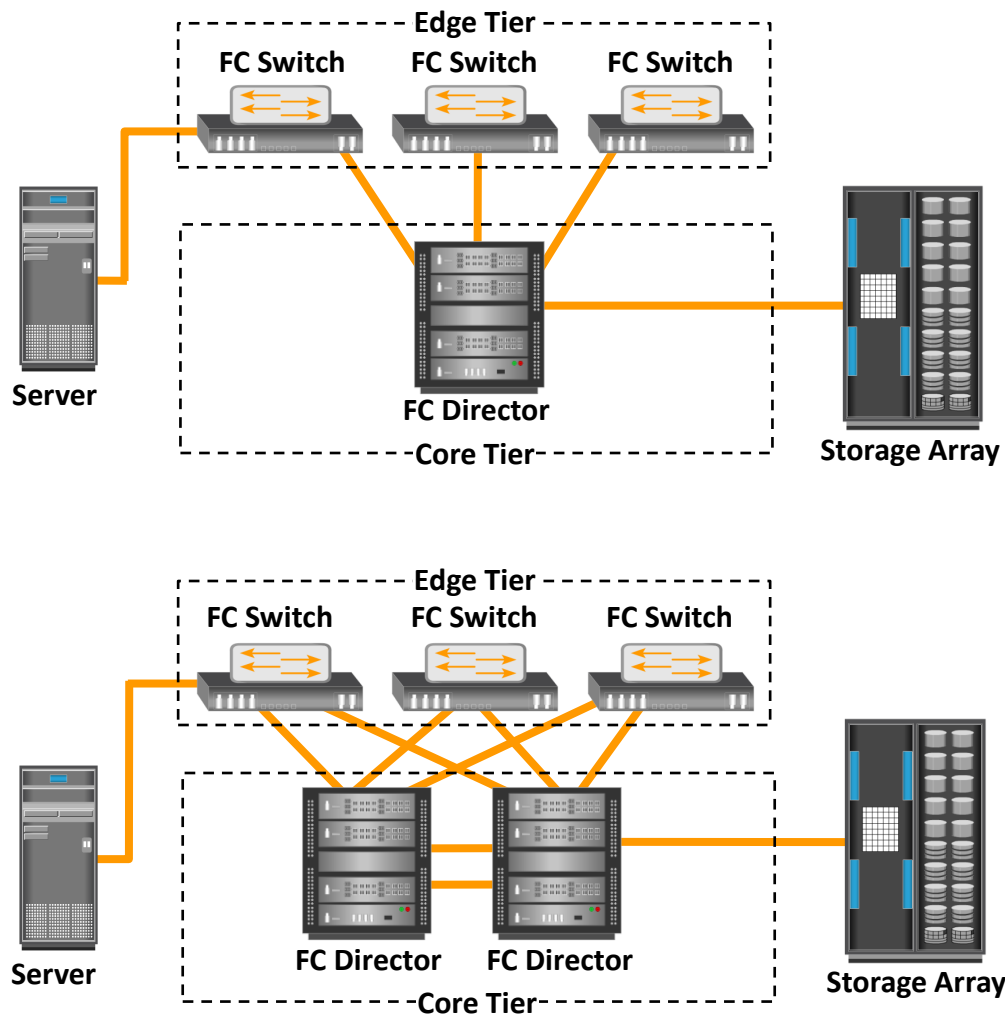
Zoning

> **Zoning**
>
> It is an FC switch function that enables node ports within the fabric to be logically segmented into groups, and communicate with each other within the group.

- Zone set comprises zones
- Each zone comprises zone members (HBA and array ports)
- Benefits
  - Restricts RSCN traffic
  - Provides access control

Server

Hypervisor

FC SAN

Zone 1

Zone 2

Storage Array

Array port

HBA Port

Servers

Types of Zoning

Servers

Switch Domain ID = 15

Port 5

Zone 2

Port 1

FC Switch

Storage Array

WWN 10:00:00:00:C9:20:DC:40

Port 9   **Port 12**

Zone 3

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 ; 15,12
Zone 3 (Mixed Zone) = 10:00:00:00:C9:20:DC:56 ; 15,12

Module 5: Fibre Channel Storage Area Network (FC SAN)

Lesson 4: Virtualization in SAN
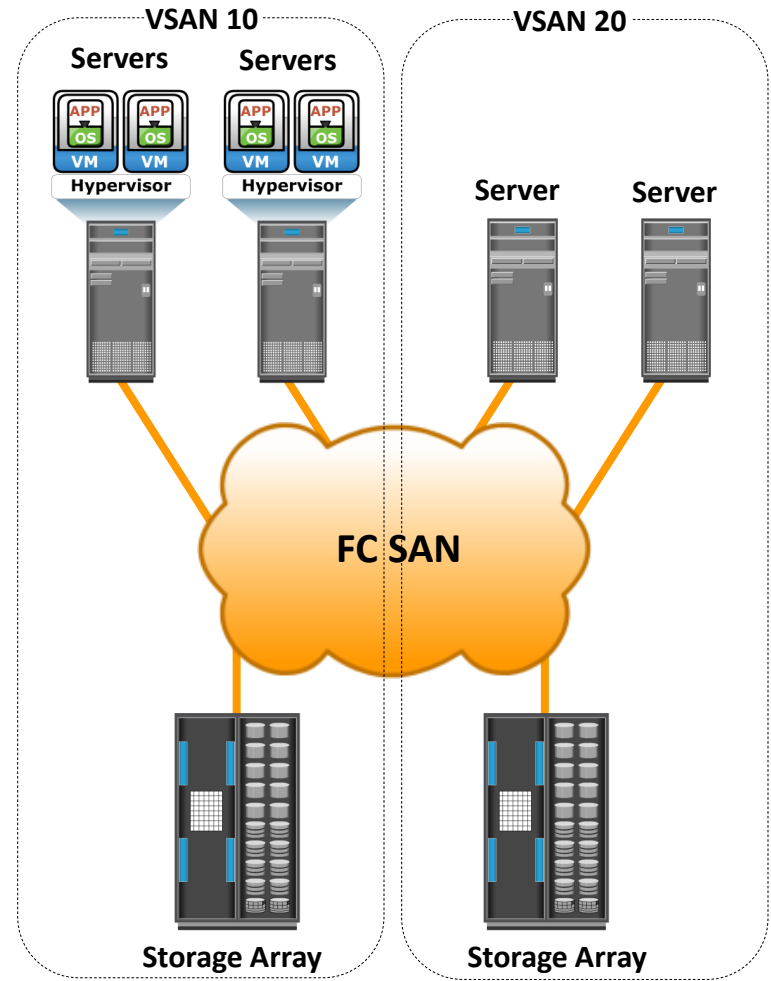
During this lesson the following topics are covered:

- Virtual SAN

Virtual SAN (VSAN)/Virtual Fabric

VSAN

It is a logical fabric on an FC SAN, enabling communication among a group of nodes, regardless of their physical location in the fabric.

- Each VSAN has its own fabric services (name server, zoning), configuration, and set of FC addresses

- VSANs improve SAN security, scalability, availability, and manageability

Ethernet SAN: iSCSI and FCoE

Module 6: IP SAN and FCoE

Lesson 1: IP SAN

During this lesson the following topics are covered:

- Drivers for IP SAN
- IP SAN Protocols: iSCSI and FCIP
- Components, topologies, and protocol stack for iSCSI and FCIP

Drivers for IP SAN

- IP SAN transports block-level data over IP network
- IP is being positioned as a storage networking option because:
  - ▸ Existing network infrastructure can be leveraged
  - ▸ Reduced cost compared to investing in new FC SAN hardware and software
  - ▸ Many long-distance disaster recovery solutions already leverage IP-based network
  - ▸ Many robust and mature security options are available for IP network

IP SAN Protocol: iSCSI

- IP based protocol that is used to connect host and storage
- Encapsulates SCSI commands and data into an IP packet and transports them using TCP/IP

Components of iSCSI

- iSCSI initiator
  - ▶ Example: iSCSI HBA (hardware, rare nowadays)
  - ▶ Example: iSCSI software initiator (pure software, common)
- iSCSI target
  - ▶ Storage array with iSCSI support
  - ▶ iSCSI gateway – enables communication with FC storage array
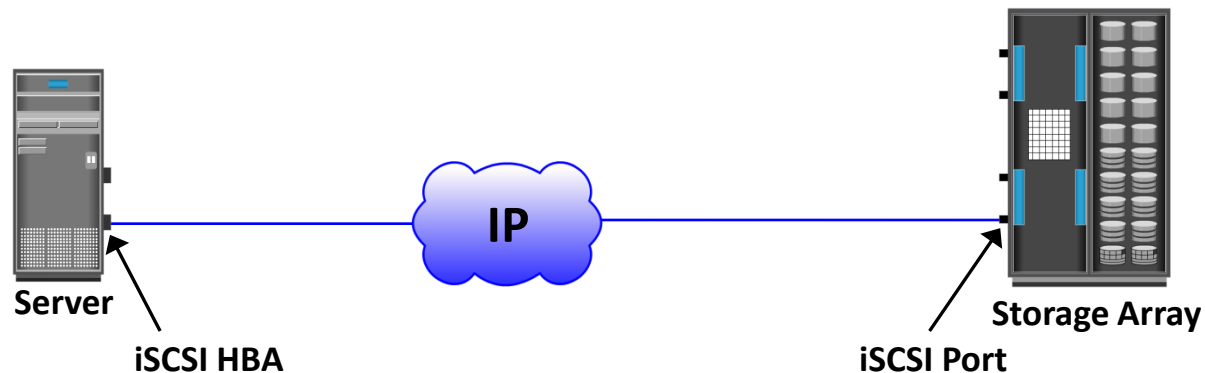- IP network
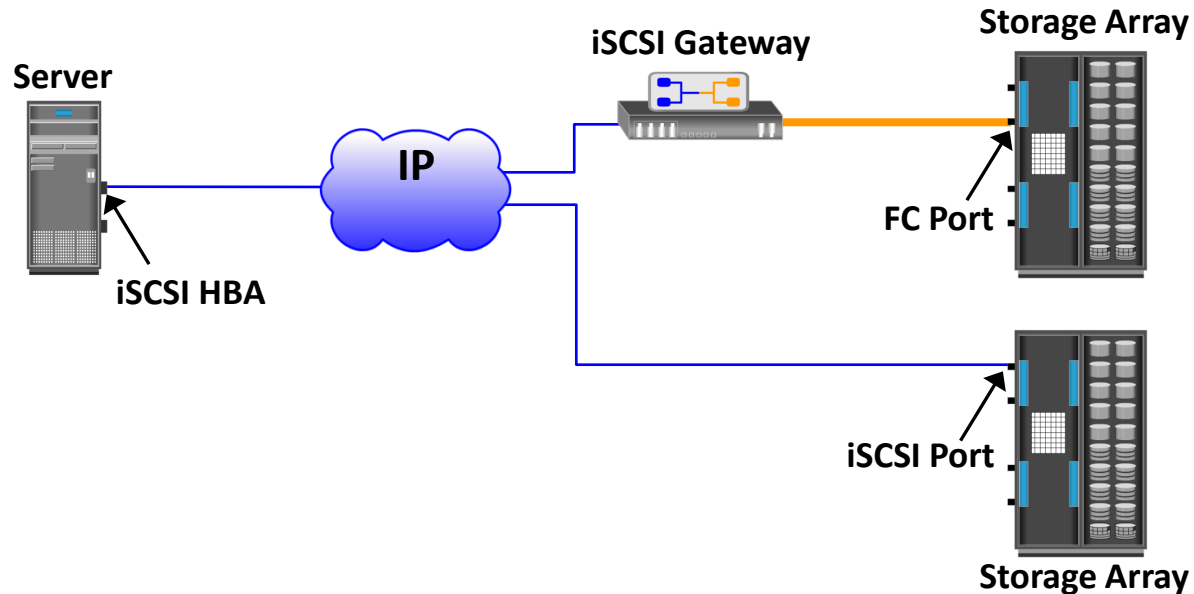
iSCSI Host Connectivity Options

- Standard NIC with software iSCSI initiator
  - ▸ NIC provides network interface
  - ▸ Software initiator provides iSCSI functionality
  - ▸ Requires host CPU cycles for iSCSI and TCP/IP processing
- TCP Offload Engine (TOE) NIC with software iSCSI initiator
  - ▸ Moves TCP processing load off the host CPU onto the NIC card
  - ▸ Software initiator provides iSCSI functionality
  - ▸ Requires host CPU cycles for iSCSI processing
- iSCSI HBA
  - ▸ Offloads both iSCSI and TCP/IP processing from host CPU
 - ▸ Simplest option for boot from SAN

iSCSI Topologies: Native iSCSI

- iSCSI initiators are either directly attached to storage array or connected through IP network
  - ▸ No FC component
- Storage array has iSCSI port
- Each iSCSI port is configured with an IP address

Server

IP

Storage Array
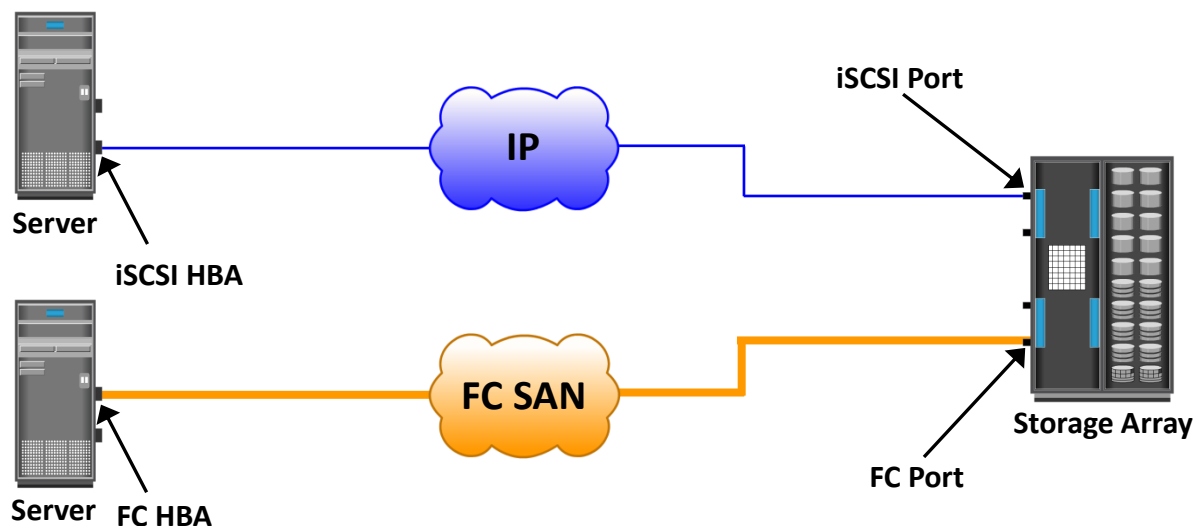
iSCSI HBA

iSCSI Port

iSCSI Topologies: Bridged iSCSI

- iSCSI gateway is used to enable communication between iSCSI host and FC storage
- iSCSI gateway works as bridge between FC and IP network
    - Converts IP packets to FC frames and vice versa
- iSCSI initiator is configured with gateway's IP address as its target
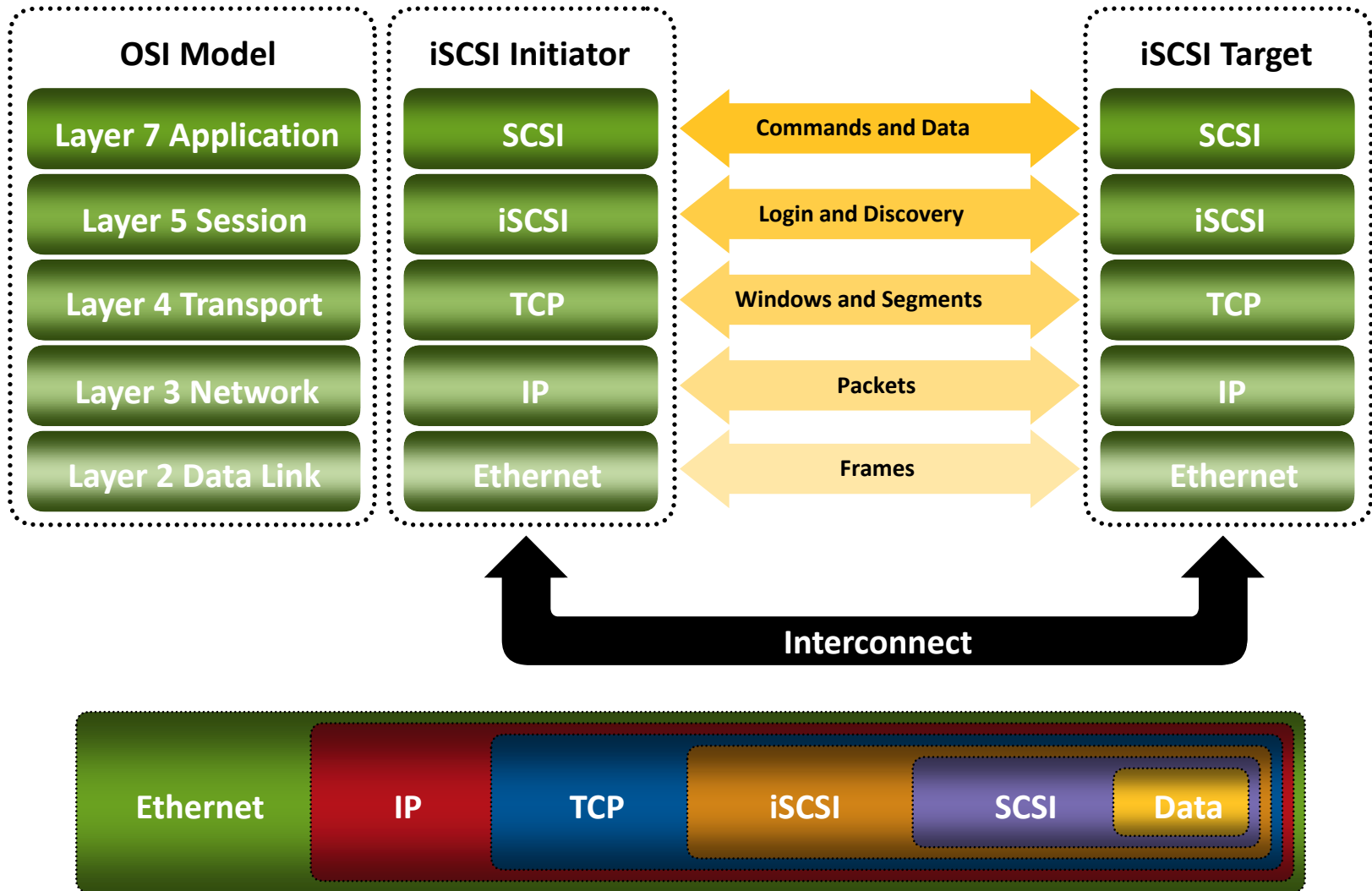- iSCSI gateway is configured as FC initiator to storage array
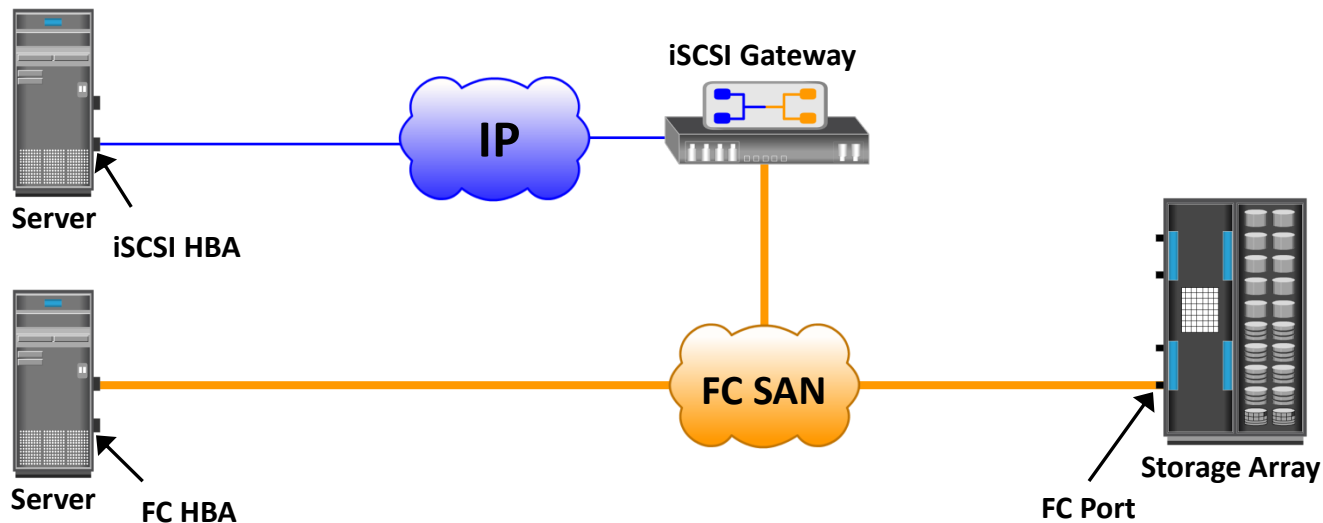
Combining FC and Native iSCSI Connectivity

- Array provides both FC and iSCSI ports
  - ▸ Enable iSCSI and FC connectivity in the same environment
  - ▸ No bridge devices needed

iSCSI Protocol Stack

OSI Model	iSCSI Initiator		iSCSI Target
Layer 7 Application	SCSI	Commands and Data	SCSI
Layer 5 Session	iSCSI	Login and Discovery	iSCSI
Layer 4 Transport	TCP	Windows and Segments	TCP
Layer 3 Network	IP	Packets	IP
Layer 2 Data Link	Ethernet	Frames	Ethernet

Interconnect

Ethernet	IP	TCP	iSCSI	SCSI	Data

iSCSI Discovery

- For iSCSI communication, initiator must discover location and name of target on a network

- iSCSI discovery takes place in two ways:

  ▸ SendTargets discovery

    ▸▸ Initiator is manually configured with the target's network portal

    ▸▸ Initiator issues SendTargets command; target responds with required parameters

  ▸ Internet Storage Name Service (iSNS)

    ▸▸ Initiators and targets register themselves with iSNS server

    ▸▸ Initiator can query iSNS server for a list of available targets

iSCSI Name

- iSCSI name is a unique iSCSI identifier that is used to identify initiators and targets within an iSCSI network

- Two common types of iSCSI names are:

  - iqn: iSCSI Qualified Name

    - iqn.2008-02.com.example:optional_string

 - eui: Extended Unique Identifier

    - eui.0300732A32598D26

IP SAN Protocol: FCIP

- IP-based protocol that is used to connect distributed FC SAN islands

- Creates virtual FC links over existing IP network that is used to transport FC data between different FC SANs

- Encapsulates FC frames onto IP packet
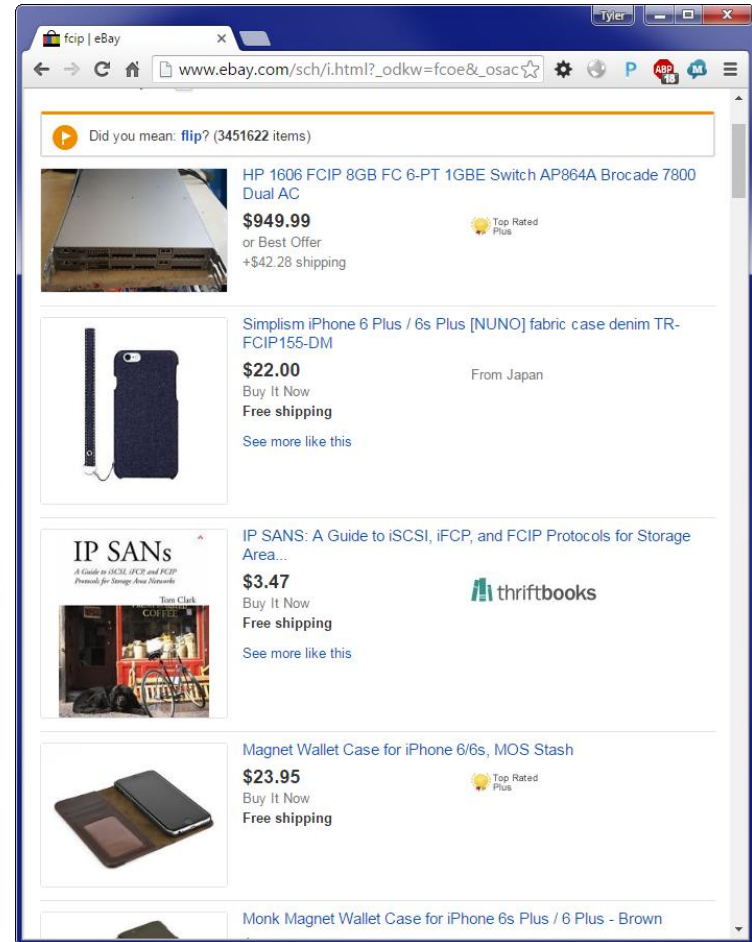
- Provides disaster recovery solution

Safe to ignore

FCIP is mostly dead

- FCIP is mostly dead, and here's how you can check. This is what you see when you search ebay for "FCOE" vs "FCIP":
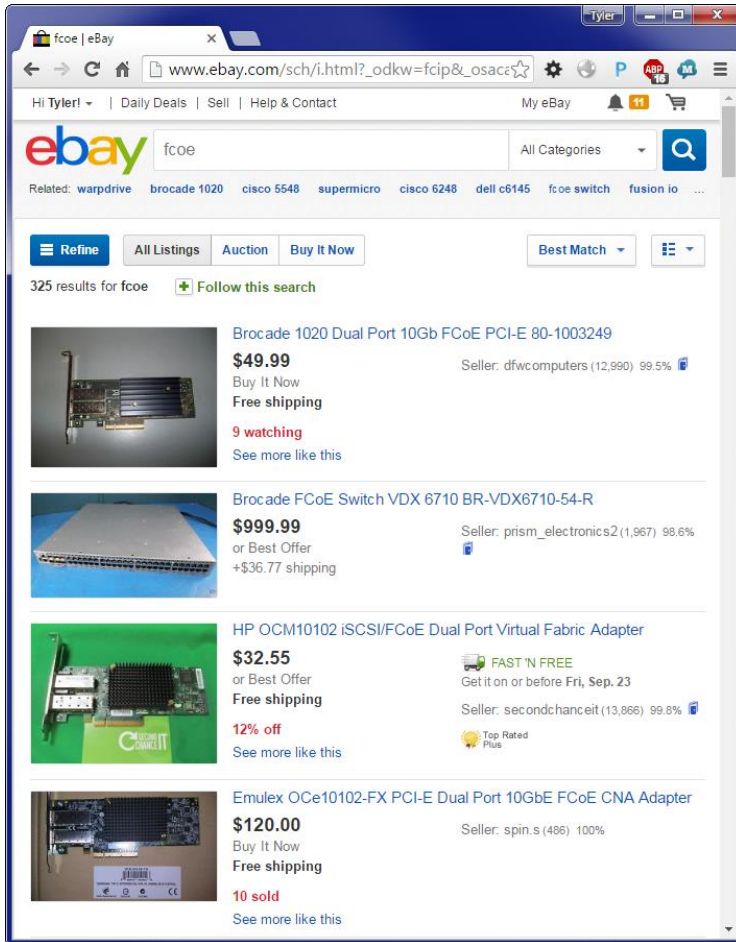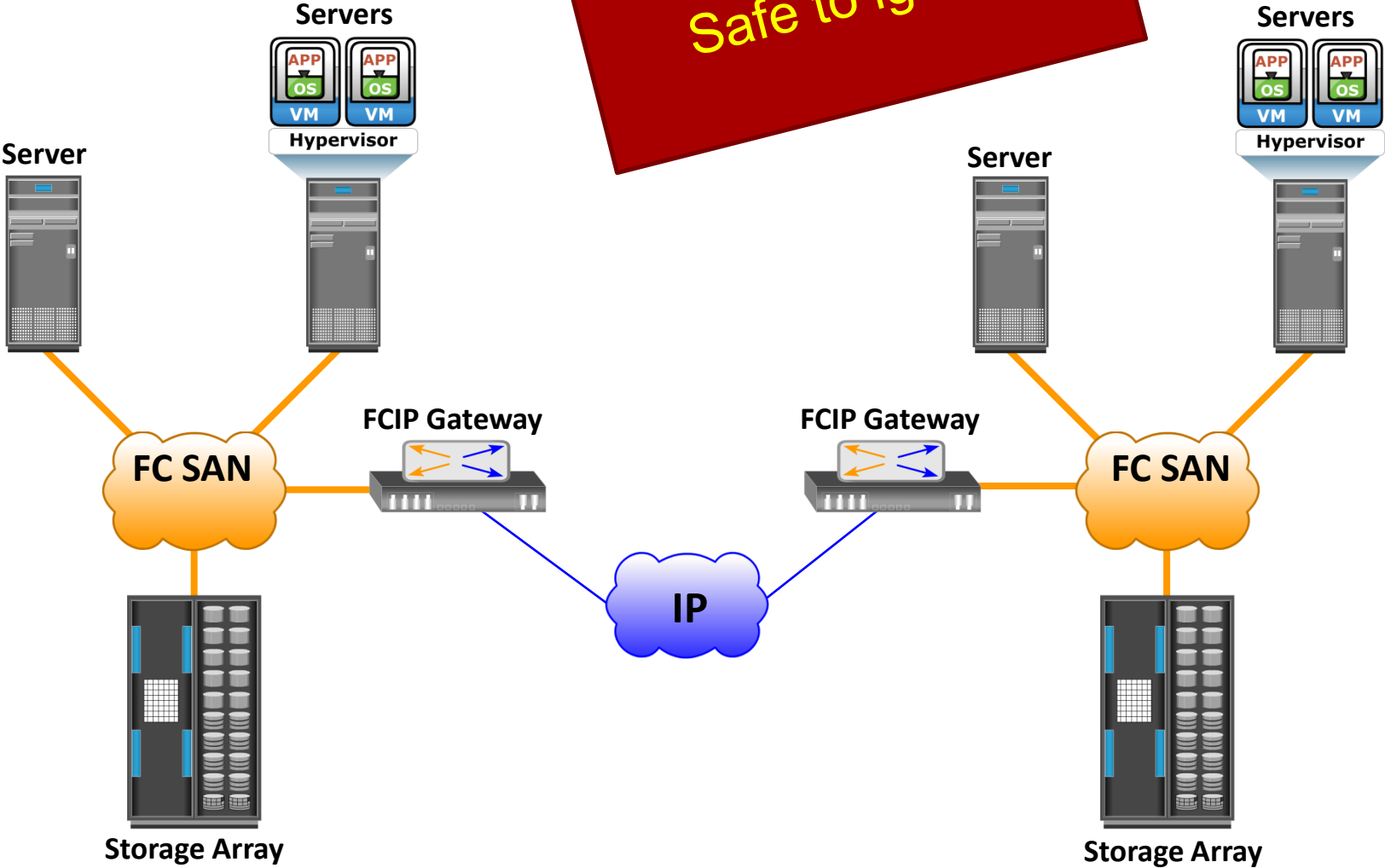

Lots of hardware

One switch, a $3 book, and iPhone accessories
(also a spellcheck warning)

FCIP Topology
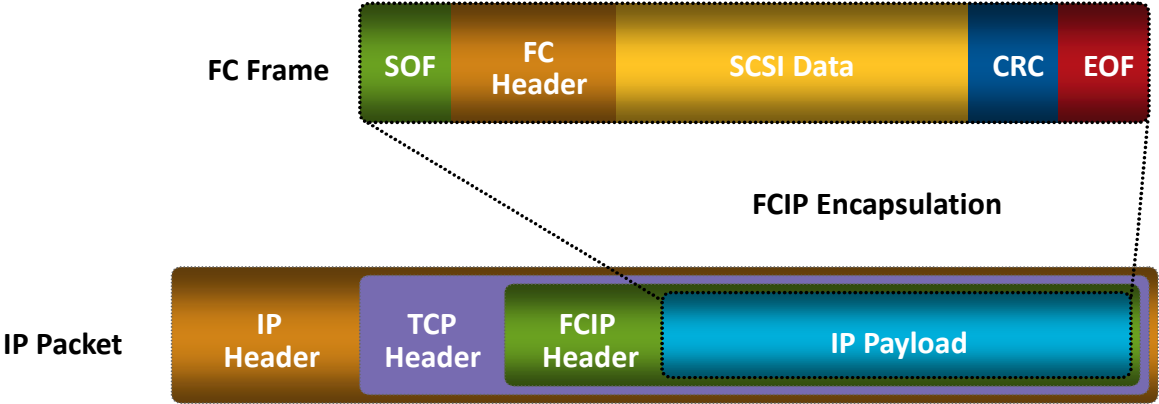
Safe to ignore

Servers

APP OS VM | APP OS VM
Hypervisor

Server

FC SAN

FCIP Gateway

FCIP Gateway

IP

Servers

APP OS VM | APP OS VM
Hypervisor

Server

FC SAN

Storage Array

Storage Array

FCIP Protocol Stack

Application	
SCSI Commands, Data, and Status	FC Frame
FCP (SCSI over FC)	
FCIP	
TCP	FC to IP Encapsulation
IP	
Physical Media	

Safe to ignore

FC Frame

SOF	FC Header	SCSI Data	CRC	EOF

FCIP Encapsulation

IP Packet

IP Header	TCP Header	FCIP Header	IP Payload
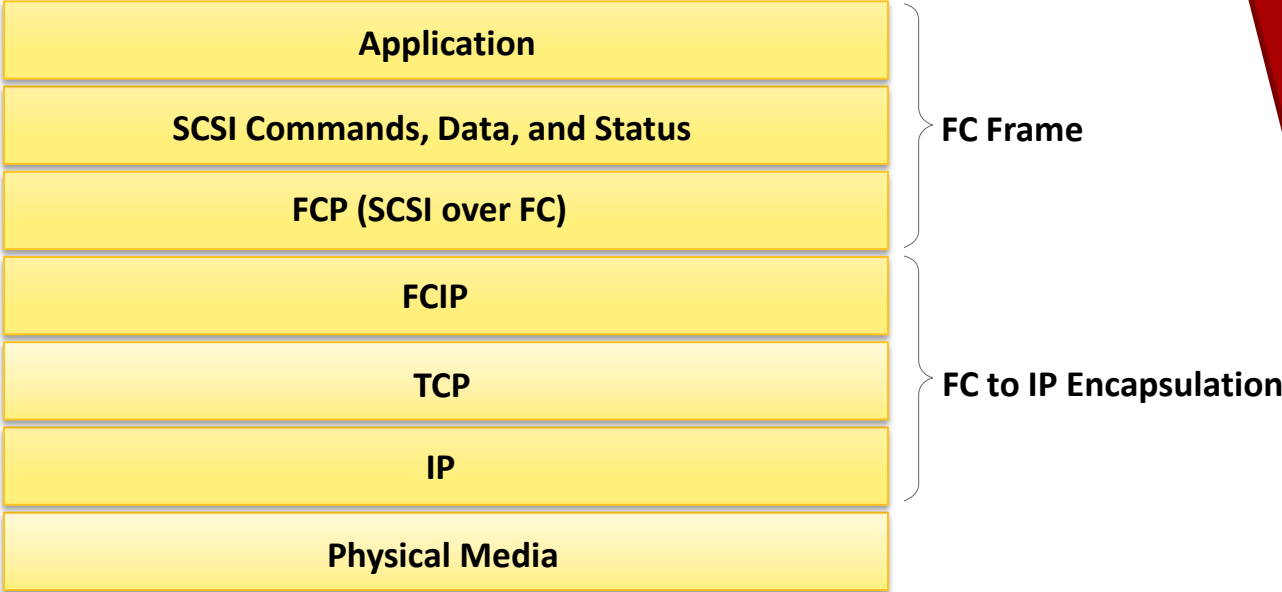
Module 6: IP SAN and FCoE

Lesson 2: Fibre Channel over Ethernet (FCoE)

During this lesson the following topics are covered:

- Drivers for FCoE
- Components of FCoE network
- FCoE frame mapping
- Converged Enhanced Ethernet (CEE)

Drivers for FCoE

- FCoE is a protocol that transports FC data over Ethernet network (Converged Enhanced Ethernet)

- FCoE is being positioned as a storage networking option because:

  ▸ Enables consolidation of FC SAN traffic and Ethernet traffic onto a common Ethernet infrastructure

  ▸ Reduces the number of adapters, switch ports, and cables

  ▸ Reduces cost and eases data center management

  ▸ Reduces power and cooling cost, and floor space
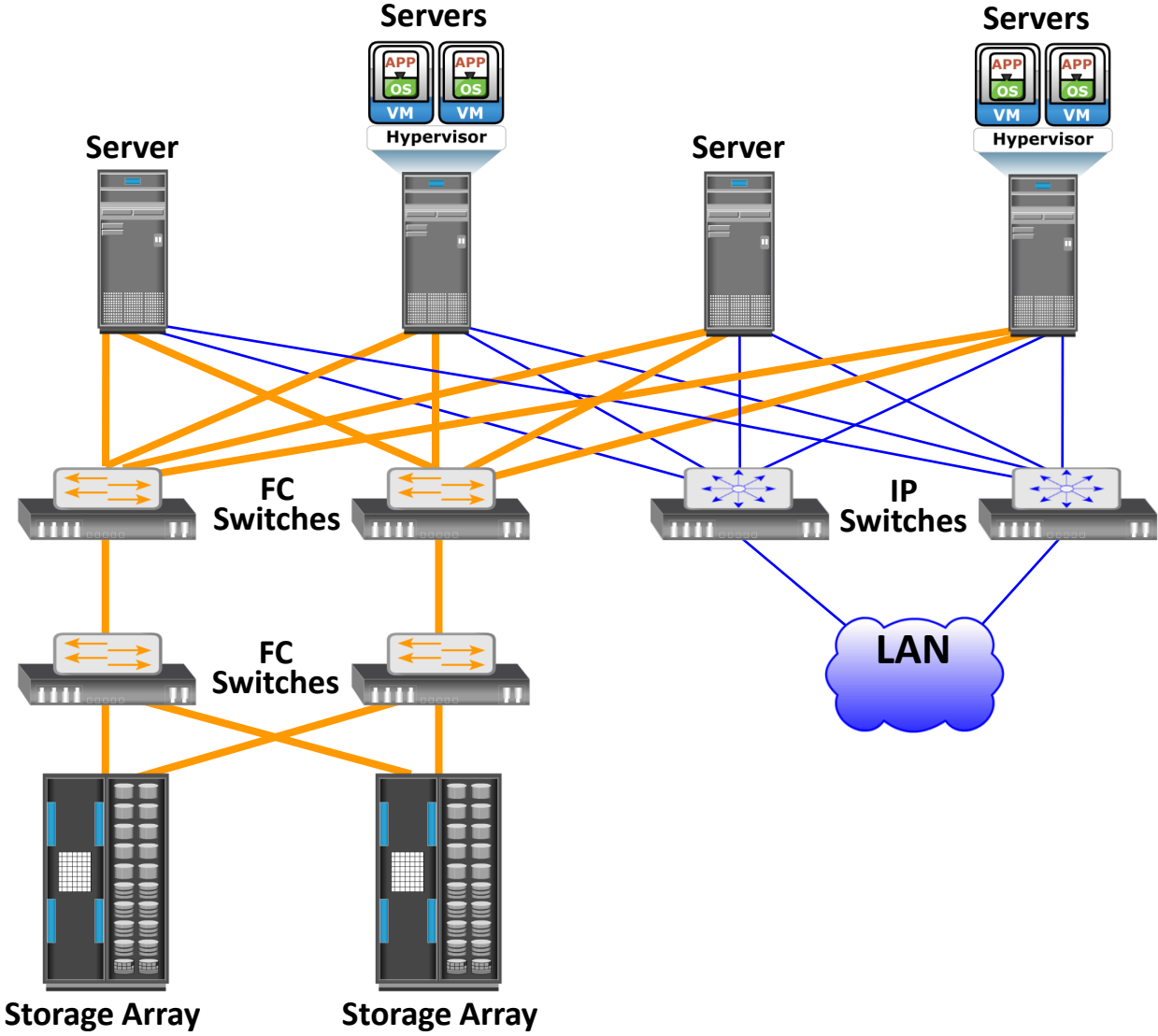
FCoE Motivation, translated

(My personal take on the evolution of FCoE)

Pro-FCoE: "Oh my god, can we stop running two networks with separate hardware, terminology, cables, and teams? It's just data!"

Anti-FCoE: "But-but-but– my job security!! I mean, uh, your packets will be haunted if they don't pass through this special network that costs 10 times a commodity Ethernet network!"

Compromise: We'll just make a variant of Ethernet so the hardware still costs 10 times as much, then everyone's happy!

Data Center Infrastructure – Before Using FCoE

Data Center Infrastructure – After Using FCoE

Servers

Servers

Server

Server

FCoE Switches

FC Switches

LAN

Storage Array

Storage Array
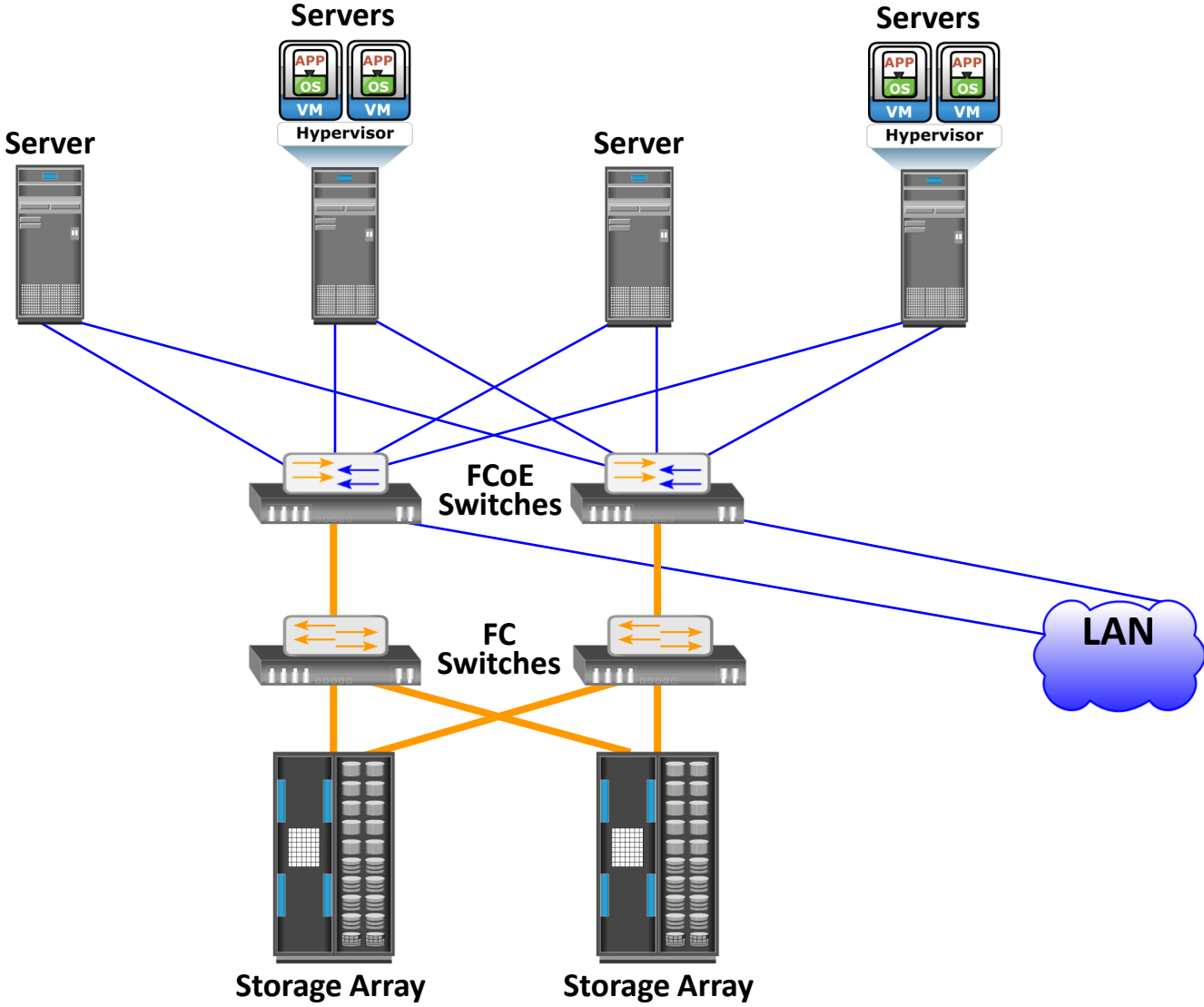
Components of an FCoE Network

- Converged Network Adapter (CNA)
- Cable
- FCoE switch

Converged Network Adapter (CNA)

- Provides functionality of both – a standard NIC and an FC HBA
  - ▸ Eliminates the need to deploy separate adapters and cables for FC and Ethernet communications
- Contains separate modules for 10 Gigabit Ethernet, FC, and FCoE ASICs
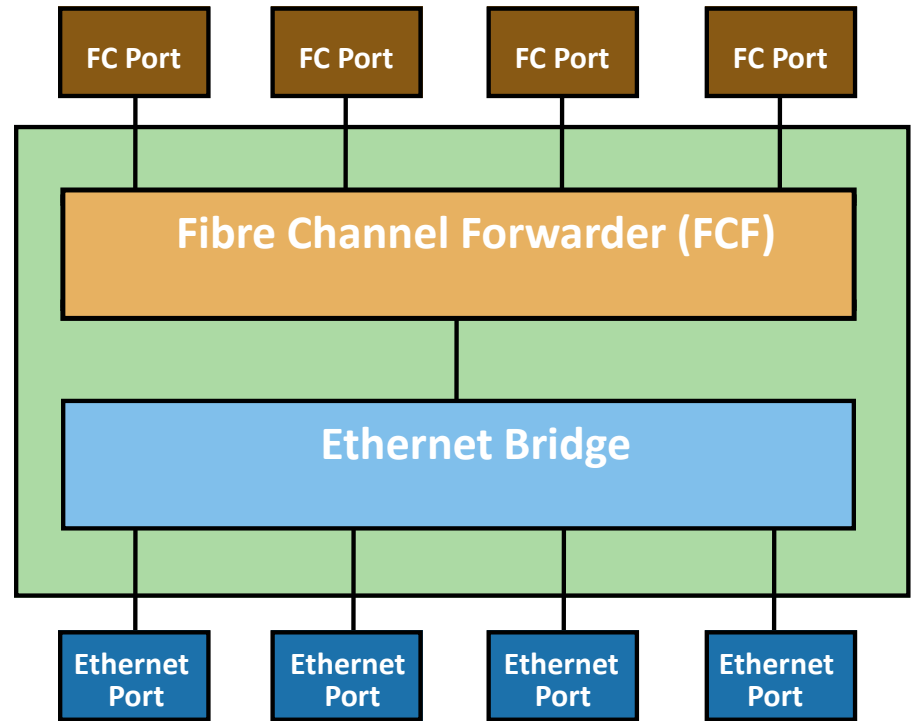 - ▸ FCoE ASIC encapsulates FC frames into Ethernet frames
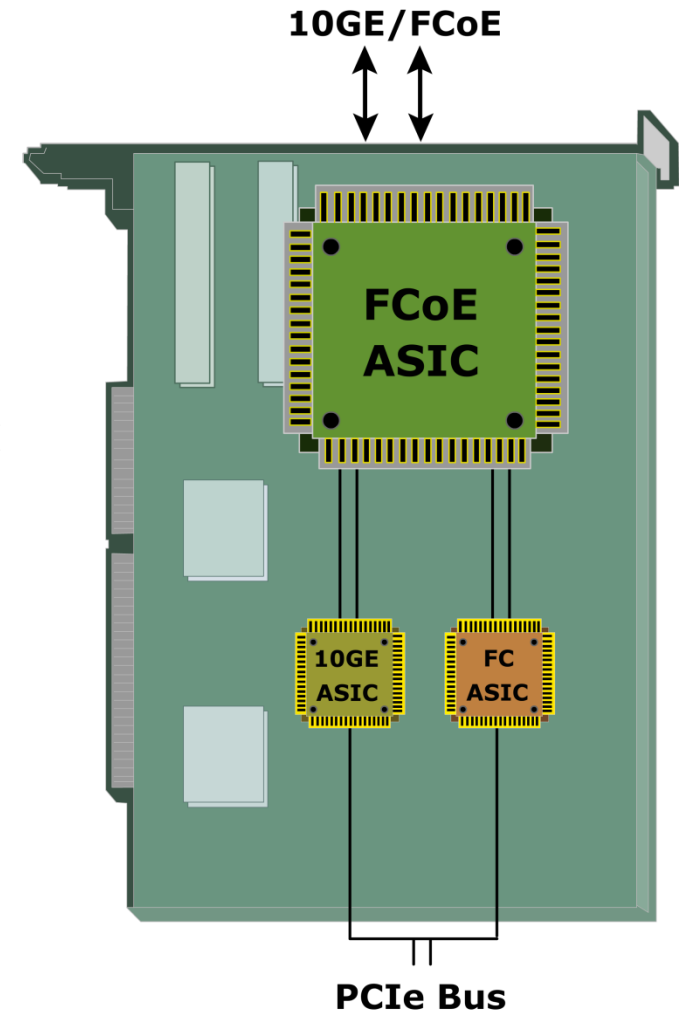
Cable

- Two options are available for FCoE cabling
  - ▸ Copper based Twinax cable
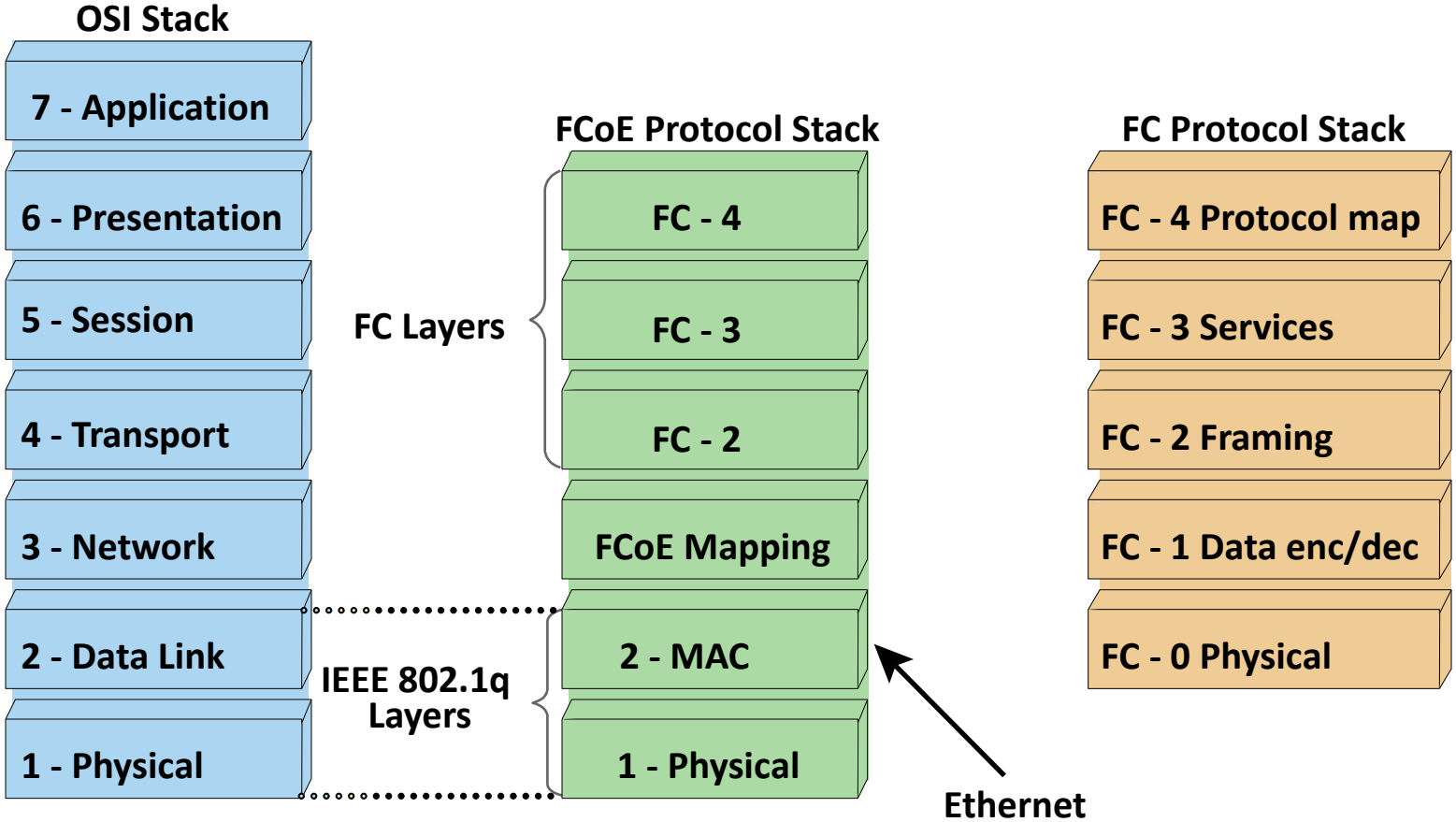 - ▸ Standard fiber optical cable

Twinax Cable	Fiber Optical Cable
Suitable for shorter distances (up to 10 meters)	Can run over longer distances
Requires less power and are less expensive than fiber optical cable	Relatively more expensive than Twinax cables
Uses Small Form Factor Pluggable Plus (SFP+) connector	Uses Small Form Factor Pluggable Plus (SFP+) connector

FCoE Switch

- Provides both Ethernet and FC switch functionalities
- Consists of FCF, Ethernet bridge, and set of CEE ports and FC ports (optional)
  - ▸ FCF encapsulates and de-encapsulates FC frames
- Forwards frames based on Ethertype

FCoE Frame Mapping

OSI Stack

| 7 - Application |
| 6 - Presentation |
| 5 - Session |
| 4 - Transport |
| 3 - Network |
| 2 - Data Link |
| 1 - Physical |

FC Layers

IEEE 802.1q Layers

FCoE Protocol Stack

| FC - 4 |
| FC - 3 |
| FC - 2 |
| FCoE Mapping |
| 2 - MAC |
| 1 - Physical |

Ethernet

FC Protocol Stack

| FC - 4 Protocol map |
| FC - 3 Services |
| FC - 2 Framing |
| FC - 1 Data enc/dec |
| FC - 0 Physical |

Converged Enhanced Ethernet

- Provides lossless Ethernet

- Lossless Ethernet requires following functionalities:

  ‣ Priority-based flow control (PFC)

  ‣ Enhanced transmission selection (ETS)

  ‣ Congestion notification (CN)

  ‣ Data center bridging exchange protocol (DCBX)

Reminder

- "Lossless Ethernet" is a contradiction
- Ethernet was fundamentally designed around packet dropping
- Higher-level protocols (e.g. TCP) interpret packet loss as a mechanism for flow control

Module 6: Summary

Key points covered in this module:

- IP SAN protocols, their components, and topologies
- FCoE protocol, its components, and topology