Track 2: Operations: Data Center Architectures and Technologies

Tutorials: 1 August 2006

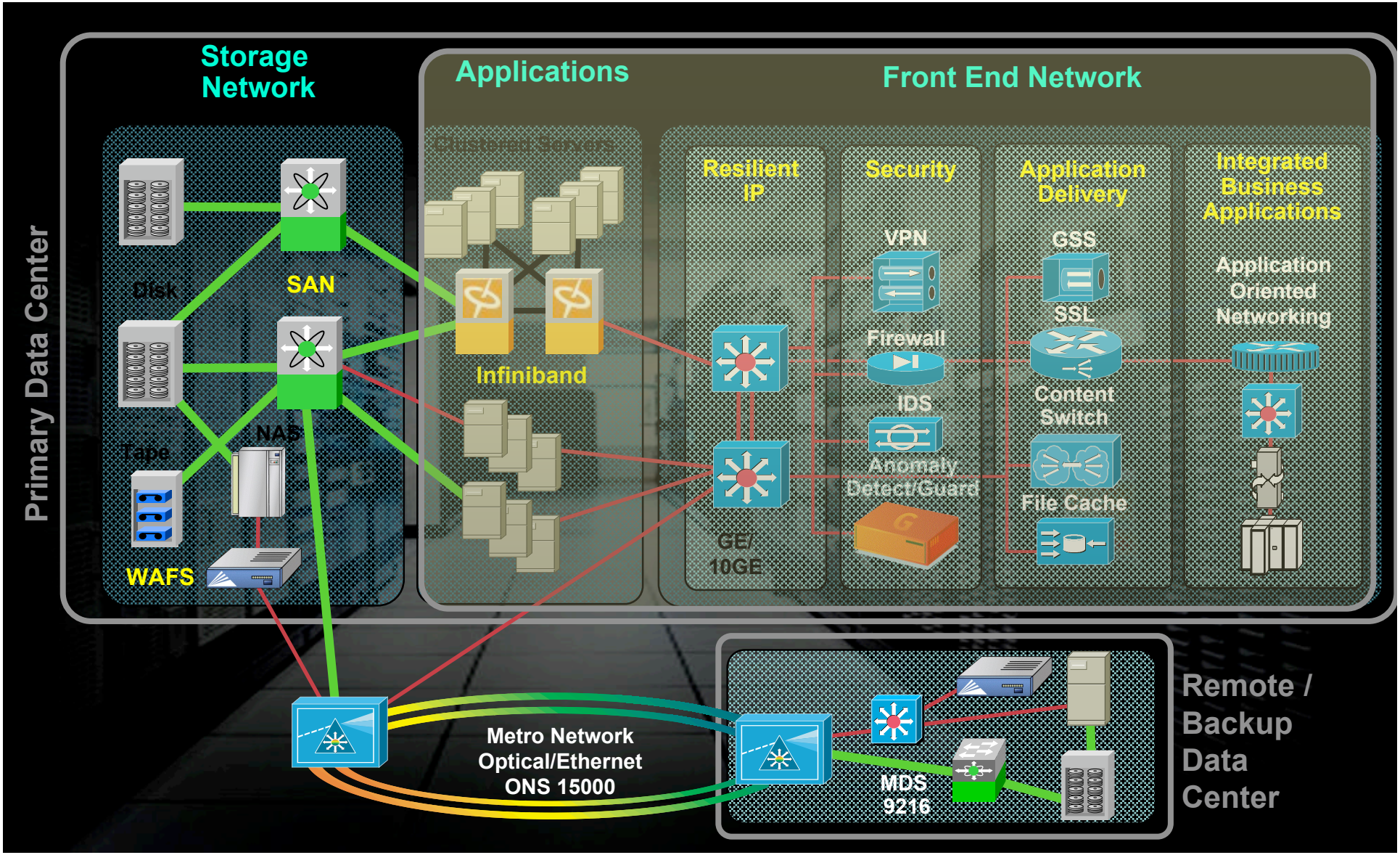
Asim Khan

Agenda

- Introduction to Storage Area Networking
- SAN Components and Topologies
- SAN Applications
- Fibre Channel Primer

Introduction to SAN

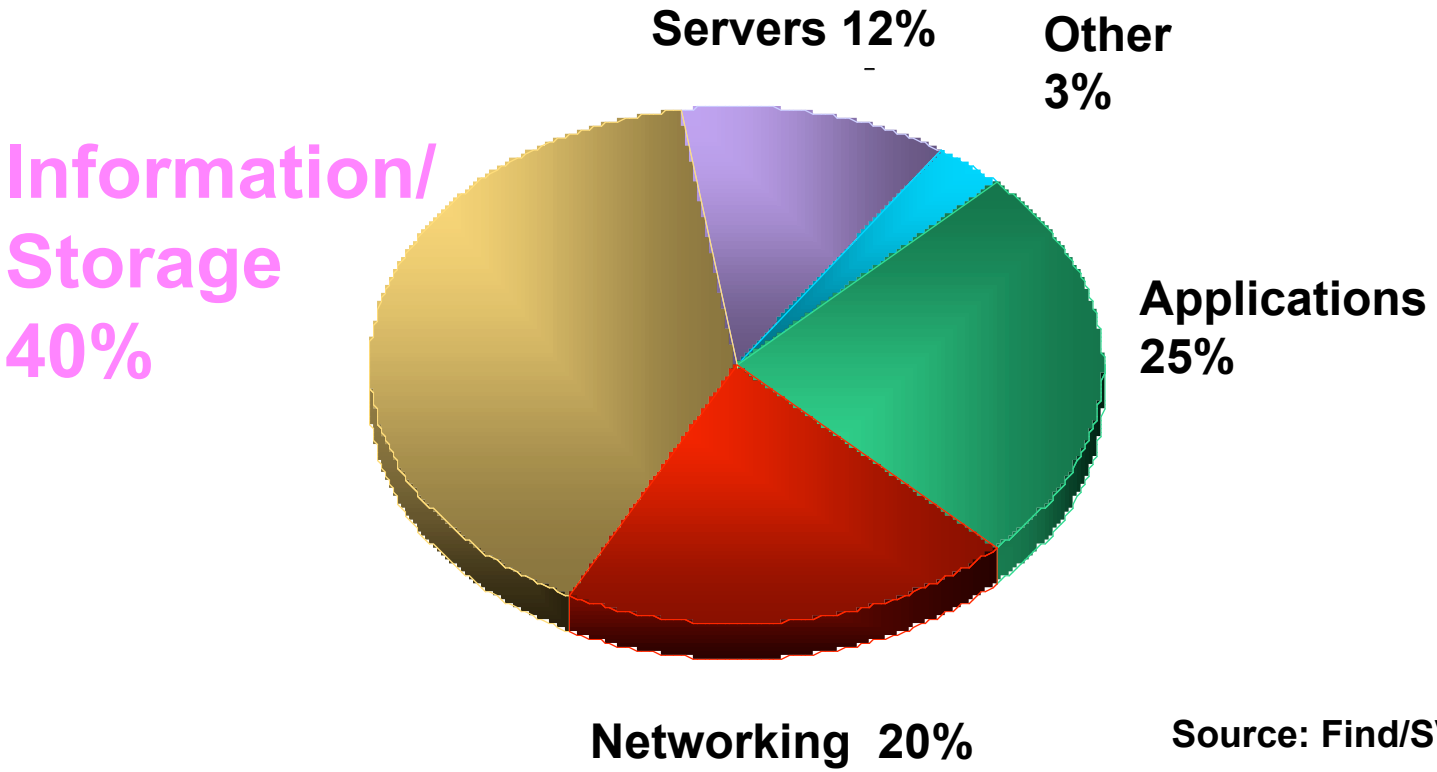
The Data Center Network



Why Are We Being Asked About Storage Networking?

Increasingly Enterprises are defining and building their infrastructure around their Enterprise Storage strategy

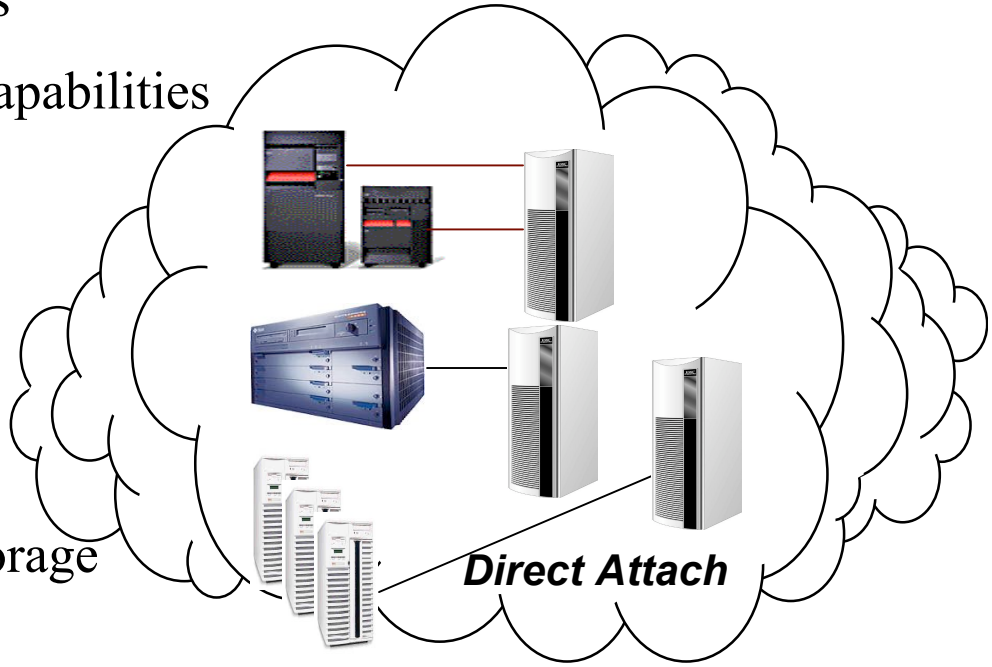
"The foundation of my companies infrastructure is….."



Source: Find/SVP 1999

Direct Attached Storage

- SCSI or FC attached. SCSI dominates
- Limited scalability & management capabilities
- Distance Limitations
- Lack of device sharing
- Limited ability to scale servers and storage independently
- 80% of Market is Direct Attached Storage



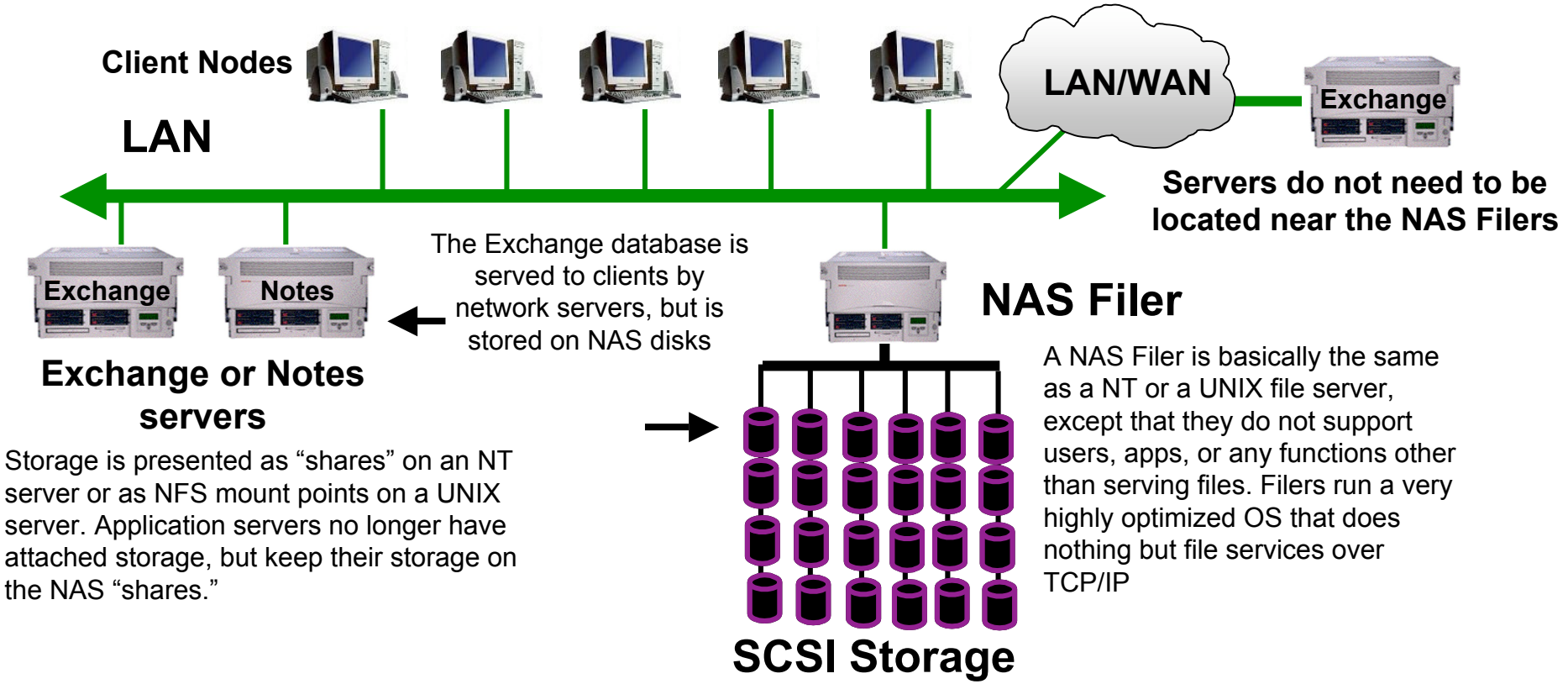
Network Attached Storage

– NAS is an open system technology introduced by SUN in 1985 as NFS or Network File System, where it allowed Unix systems to shared files over TCP/IP network.

– NFS protocol is a distributed protocol that provides transparent access to remote files or drives.

– It is designed as a client-server application, where client is the machine that accesses the file systems and the server makes file systems available to the network.

NAS

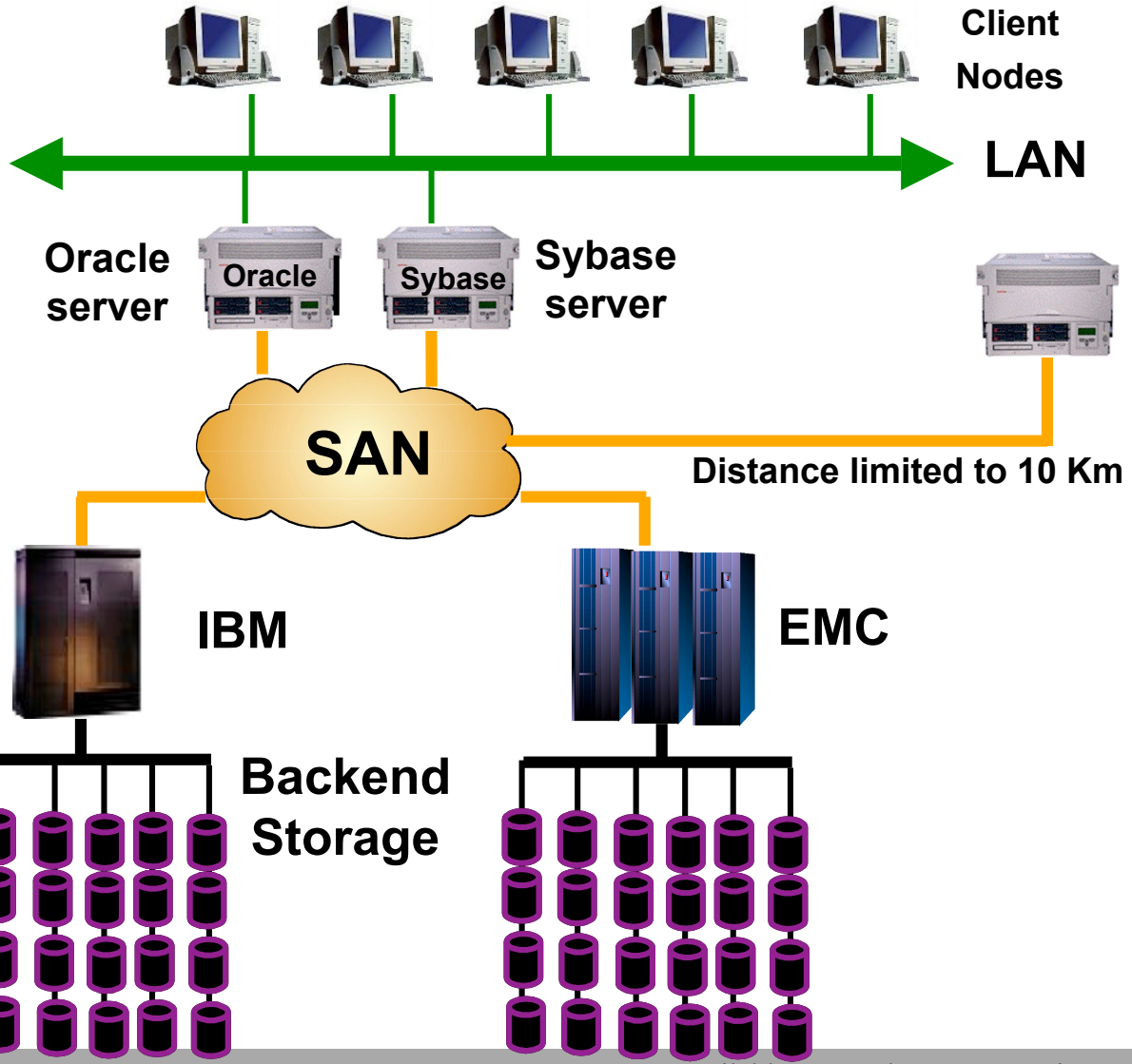


Storage Area Network

- The Fibre Channel Protocol was evolved, to overcome the distance limitation of SCSI technology which in turn gave birth to Storage Area Networking.

- Storage is presented as block level Storage where each server thinks that the Storage is locally attached.

- Relieves the LAN of storage traffic.



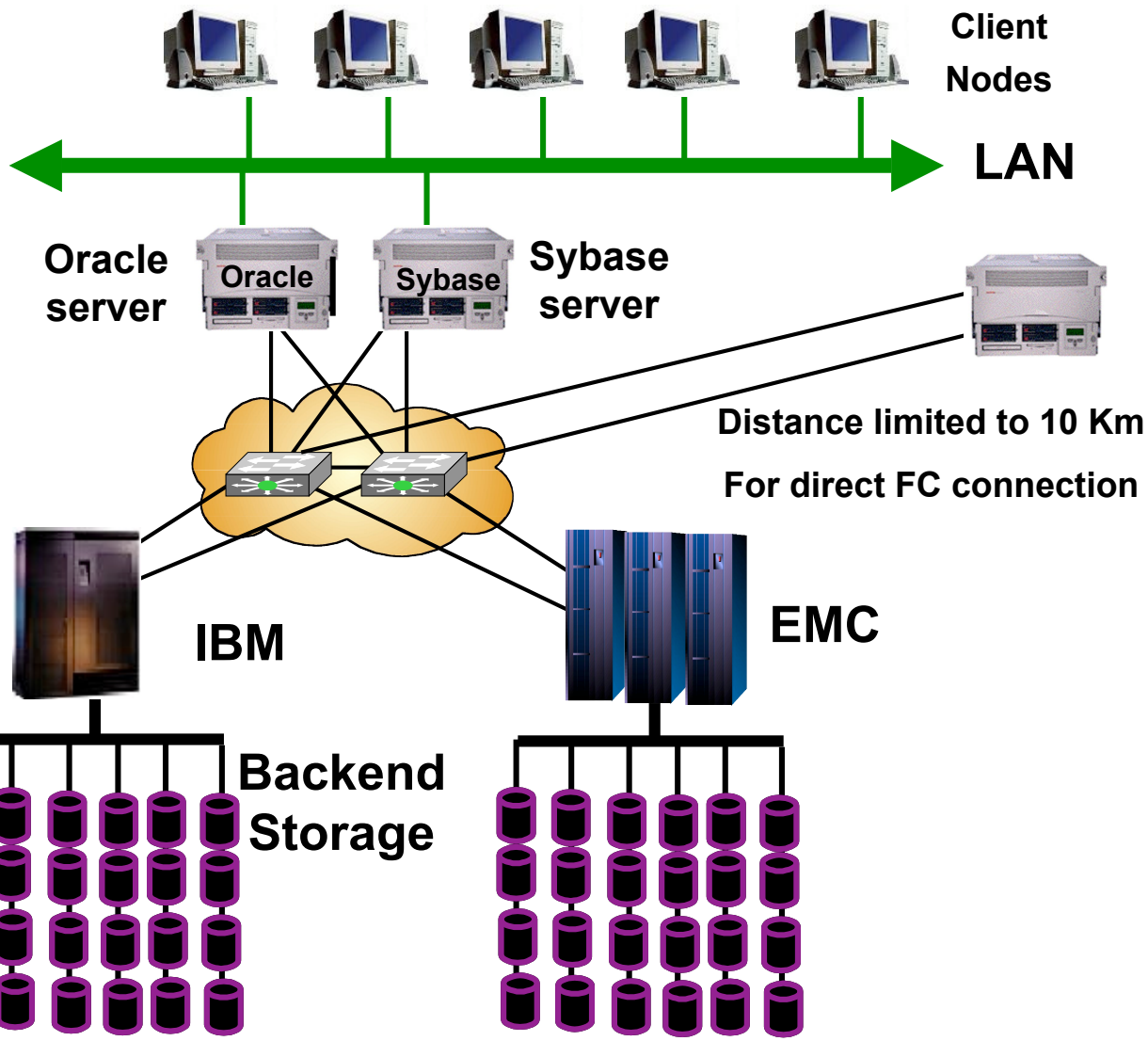
Storage Area Network

- Consolidation of Storage Resources and centralized management. Allows all Data to become an Enterprise Resource allowing much easier storage resource management.

- Any-to-any connections with multiple paths to each resource. Flexible storage allocation, movement and sharing.

- Open structure using industry standard protocol.

- Scales up with no performance loss, lowering TCO.



SAN's Defined

- **A SAN consists of 2 tiers…(according to Gartner)**

Plumbing

The first tier, the storage plumbing tier, provides connectivity between nodes in a network fashion and transports device-oriented commands and status. At least one storage node must be connected to this network.

Application

The second tier, the software tier, uses software to provide value-added services that operate over the first tier.

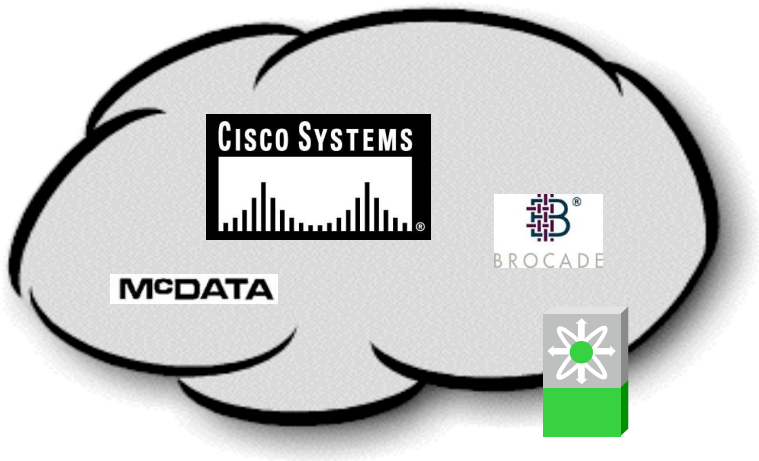
Remote mirrors, zoning, data backup, disk pooling / virtualization, Storage Resource Management, SAN file system

"Vendors of SAN offerings that do not conform to this two-configuration definition should not be considered credible SAN suppliers"

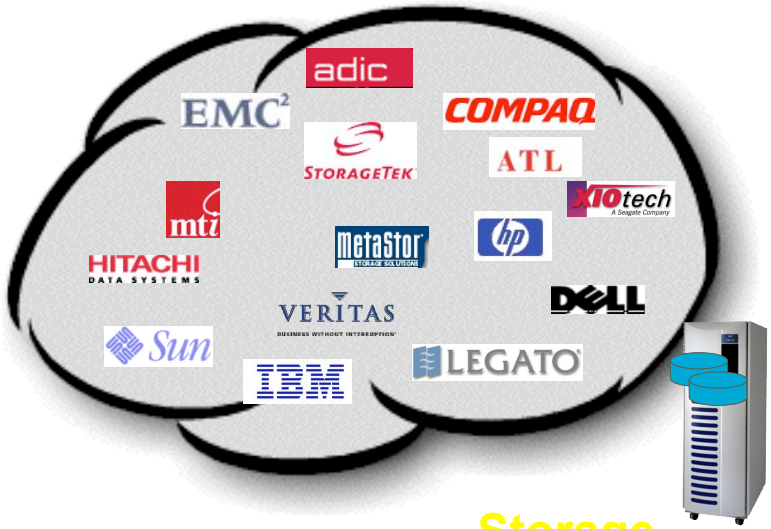
(Gartner Group 4/7/99)

SAN Components and Topologies

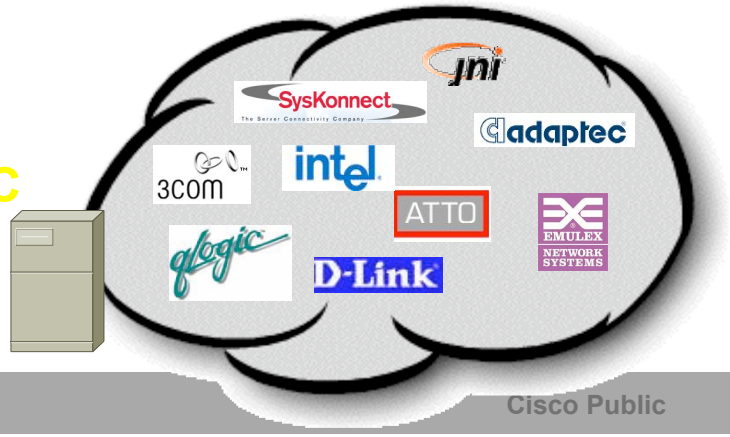
SAN Components



FC Switch



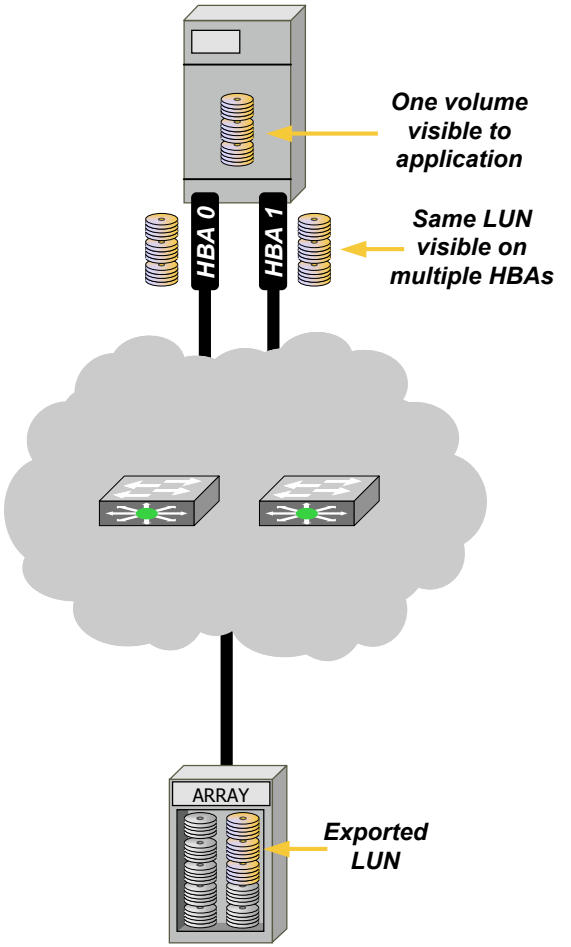
Storage



HBA/NIC

HBA Multipathing

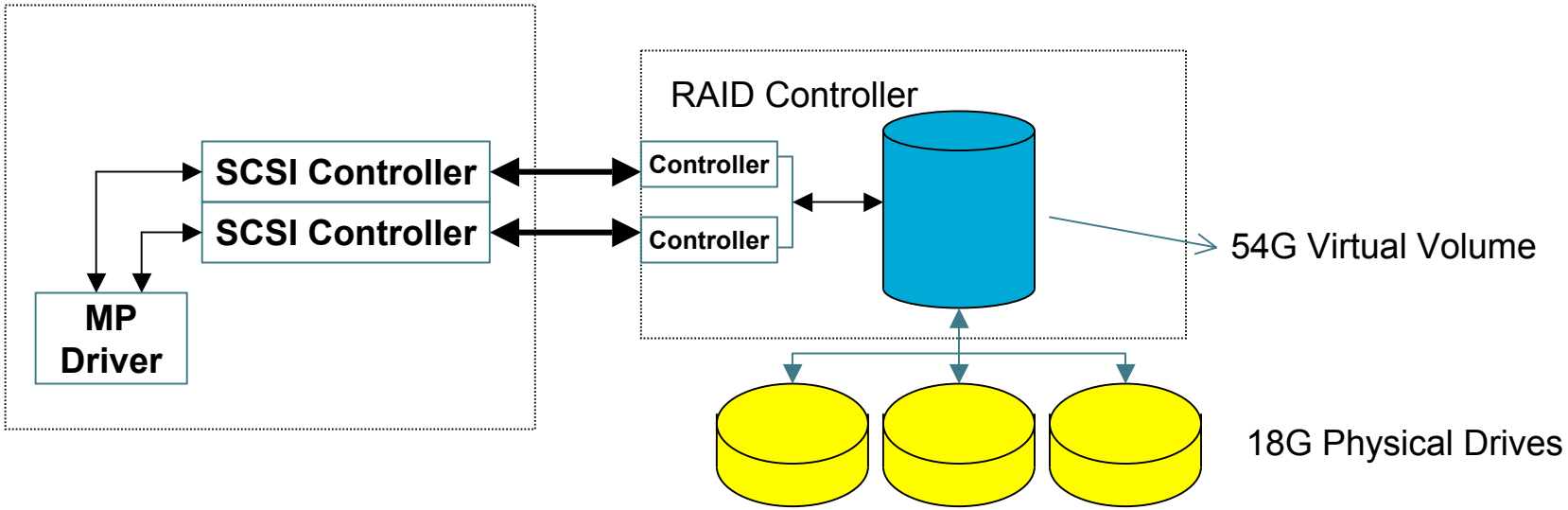
Host-Based Dedicated Software Solution



- **Multipathing provides 3 services:**
  - **Fault Detection and Isolation**
  - **HBA Fail-Over and Fail-Back**
  - **Dynamic I/O Load Balancing/Distribution**
- **Multiple vendors offer solutions**
- **All solutions require a host-based shim**
- **Some solutions are array-specific**
- **Each HBA has visibility to same LUNs**
- **Manual intervention sometimes required to re-establish failed path**
- **Host Application/OS ensures no conflict in access to common LUN**

Multi Pathing

- **DMP by Veritas or Powerpath from EMC are examples of multipathing software driver that provides High availability by managing access to the same storage array through more than one path (Intercepts the I/O and redirects it).**
- **Dynamically provides multiple paths and can be used for load balancing (active-active) or backup mode (active-passive)**



Large-Scale Core-Edge Design

8 Switches

1456 total deployed ports

192 ISL ports

1264 usable ports (86.8%)

---

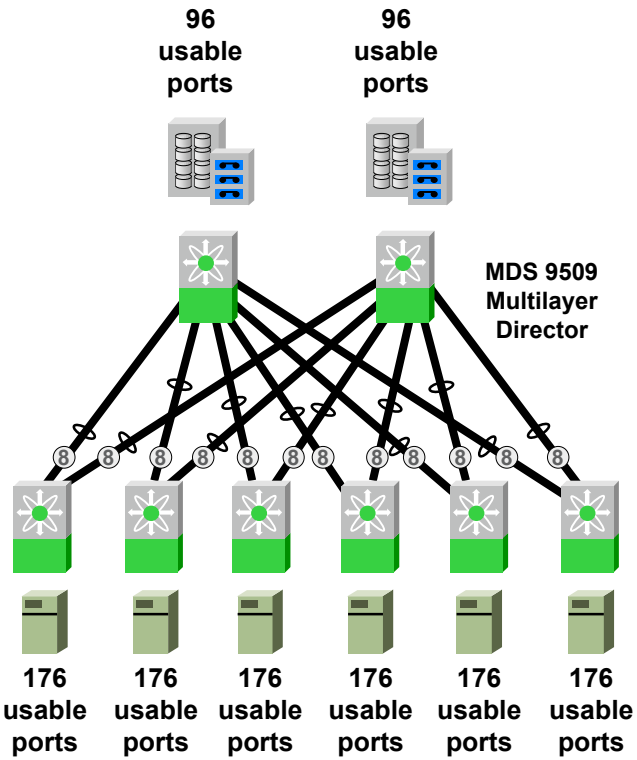
16 ports of IP storage providing iSCSI interconnect for midrange servers and Fibre Channel over IP (FCIP) services for remote data center/DR site interconnect

Virtual SANs (VSAN) support

PortChanneling with HA

10:1 max design OS* @ 2Gbps

(design oversubscription renders 32-port module oversubscription irrelevant)



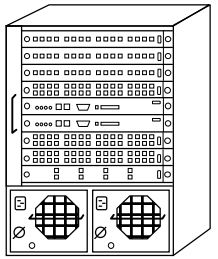
Note: 3.2:1 OS* on 32-port modules is a non-factor due to OS* of the design itself

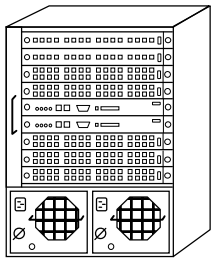
Core directors equipped with:

3 x 16 FC Port Modules
3 x 32 FC Port Modules
1 x 8 IPS (Ethernet) Module

---

144 FC ports per Director
8 IPS ports per Director





Edge directors equipped with:

2 x 16 FC Port Modules
5 x 32 FC Port Modules

---

192 FC ports per Director

Very Large-Scale Core-Edge Design

Object of Design

Keep 80%+ of traffic local which is subjected to small 3.2:1 O.S. and use ISLs only for cross-switch traffic (20%)

24 Edge Switches

4864 total deployed ports

3840 host ports

576 storage ports

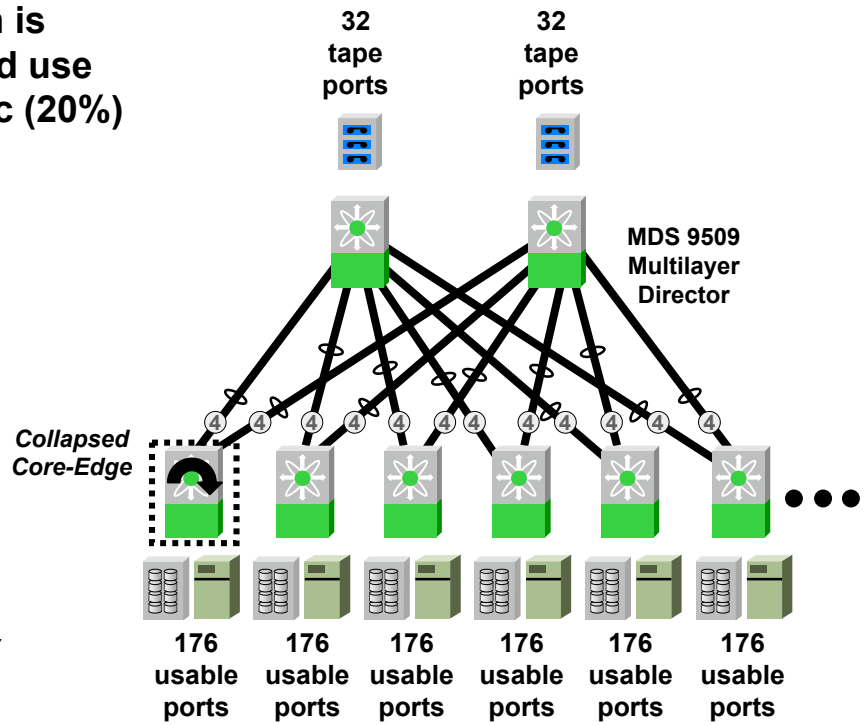
Virtual SANs (VSAN) support

Port Channeling with HA

Can add IP storage if necessary

Maximum Capacity (assuming 8 ISLs only from edge-to-cores)

Effective fan-out of ~7:1



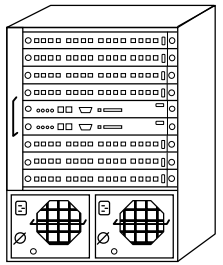
Note: 3.2:1 OS* on 32-port modules is a non-factor due to OS* of the design itself

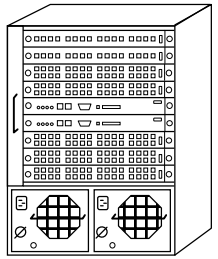
Core directors equipped with:

6 x 16 FC Port Modules
1 x 32 FC Port Modules

---

128 FC ports per Director
96 ISL Ports
32 Tape Ports





Edge directors equipped with:

2 x 16 FC Port Modules
5 x 32 FC Port Modules

---

192 FC ports per Director
24 Storage Ports
160 Host Ports
8 Potential ISL ports

SAN Applications

SAN Applications

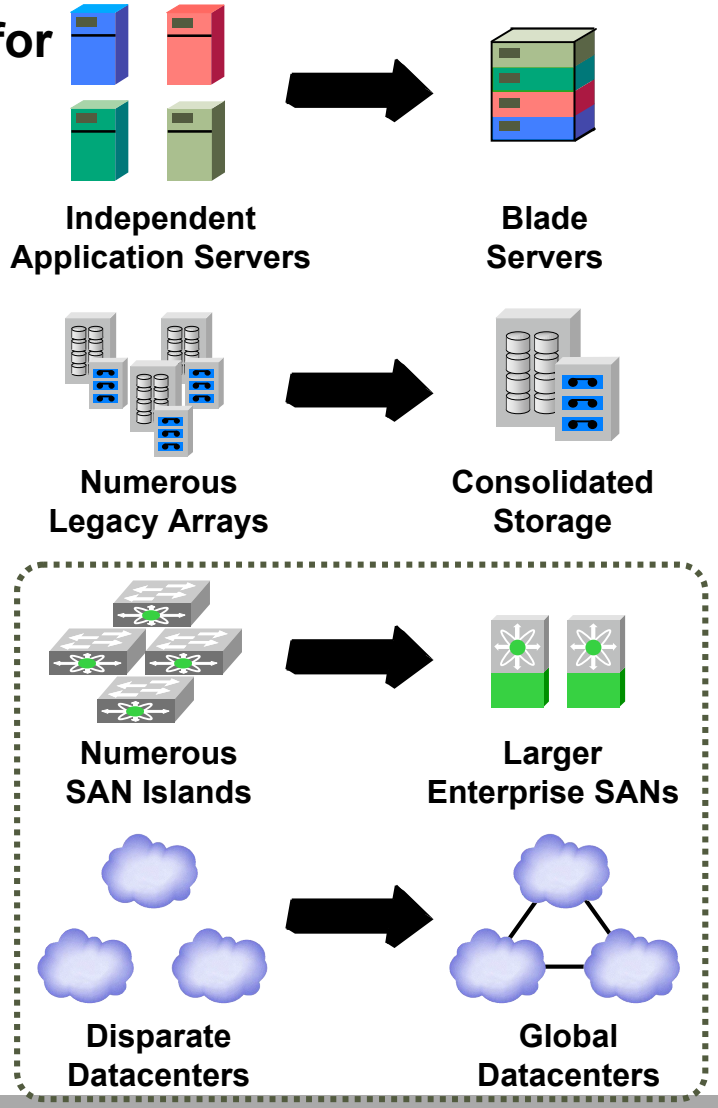
- Consolidation/Optimization
- Network Virtualization
- Storage Virtualization
- Server-less Backup
- Network Extension
- Network Security

SAN Applications

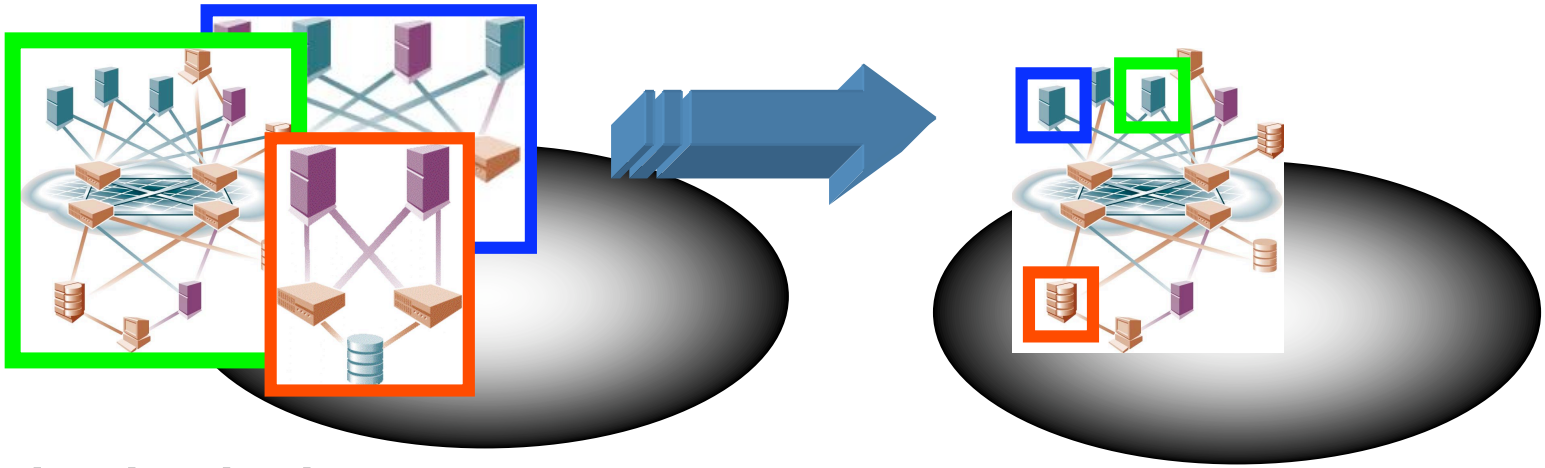
- **Consolidation/Optimization**
- Network Virtualization
- Storage Virtualization
- Server-less Backup
- Network Extension
- Network Security

Consolidation, Consolidation, Consolidation

- **Consolidation offers many new opportunities for optimization**
  - **Server consolidation**
    - **Higher powered servers**
    - **Blade server technology**
  - **Storage consolidation**
    - **Higher density arrays**
    - **Higher aggregate performance**
  - **SAN consolidation**
    - **Drive to reduce SAN *island* explosion**
    - **Increase effective port densities**
    - **Still a challenge for legacy switches**
  - **Data center consolidation**
    - **Physical moves**
    - **Interconnect existing data centers**

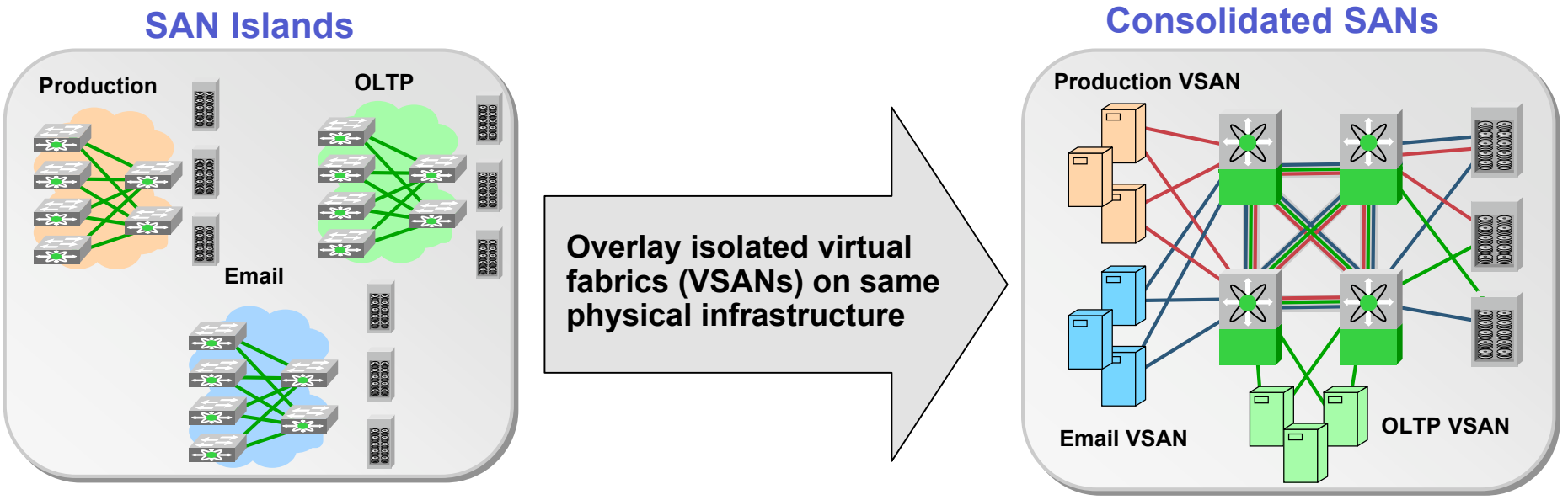


Storage Consolidation



- **Reduction in data management costs**
  (from 55% to 15% of total storage budget - IDC)
- **Ability to share, manage and protect data across the enterprise**
- **Improved storage scalability**

The Case for SAN Consolidation



	Attribute	
More	Number of SAN Switches	Fewer
No	Share Disk/Tape	Yes
No	Share DR Facilities	Yes
Complex	SAN Management	Simple
High	Overall TCO	Low

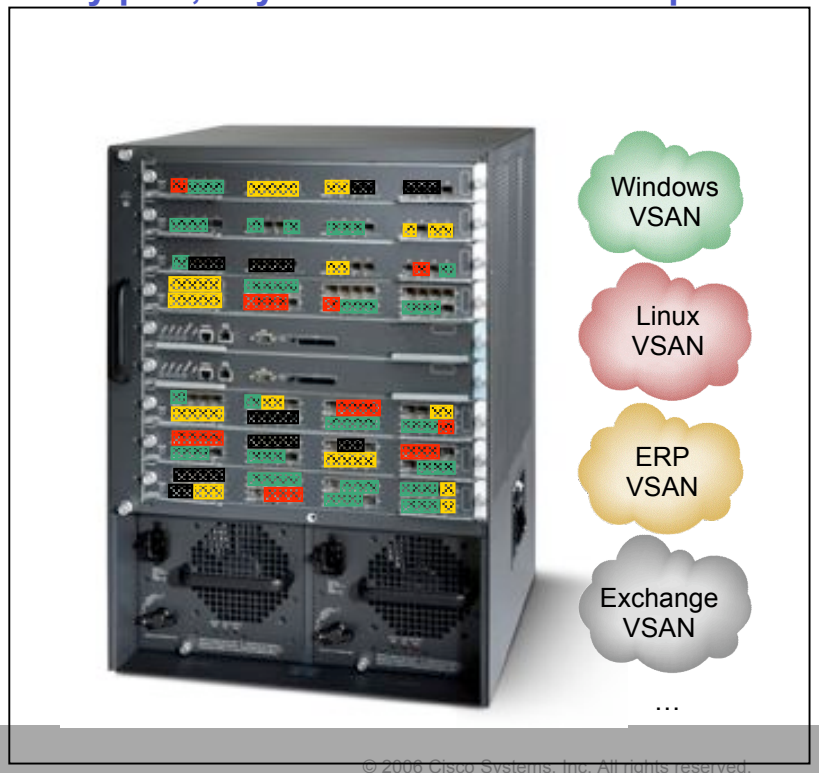
SAN Applications

- Consolidation/Optimization
- **Network Virtualization**
- Storage Virtualization
- Server-less Backup
- Network Extension
- Network Security

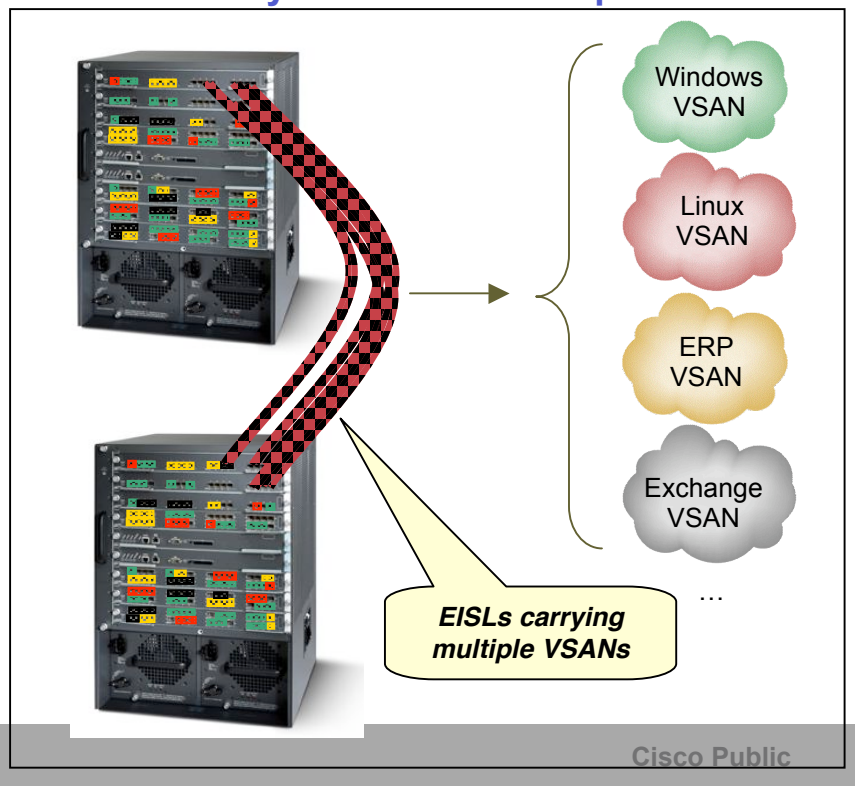
Introducing Virtual SANs (VSANs) for Network Consolidation

Virtual SANs (VSANs) offers the ability to overlay (or deploy) multiple SANs over a consolidated physical infrastructure as virtual fabrics (VSANs), where each VSAN is completed isolated with independent set of fabric services, quality of service (QoS), security, and management functions.

Any port, any module. No stranded ports.

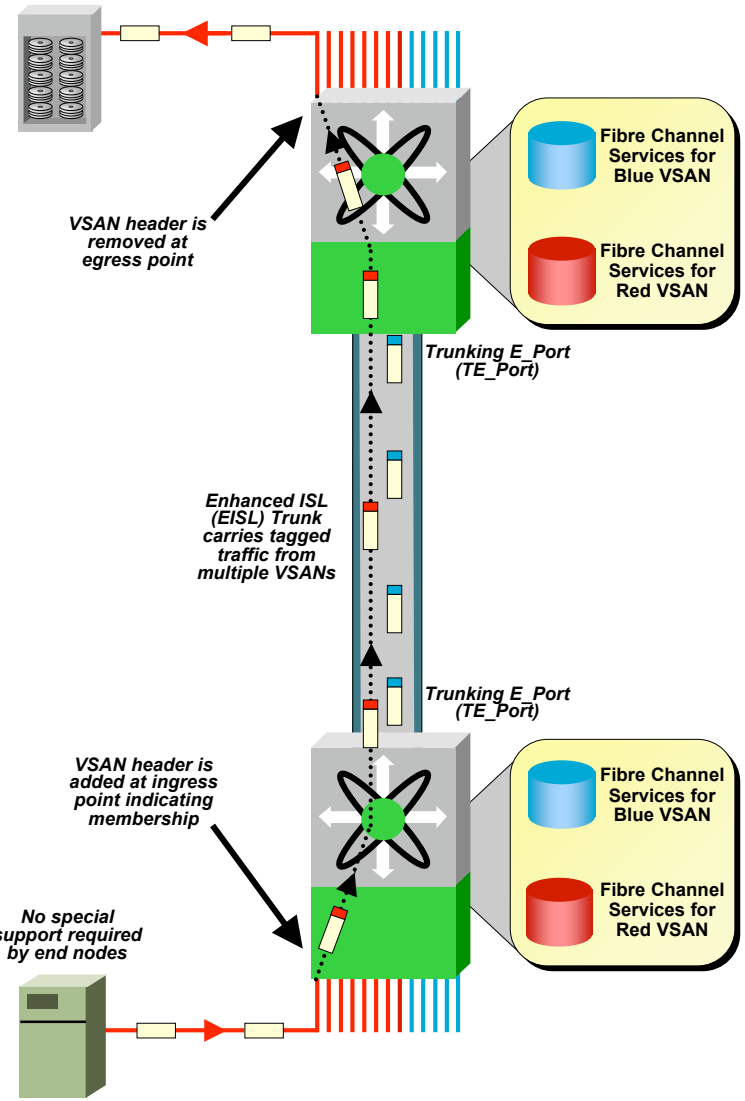


EISLs carry traffic from multiple VSANs



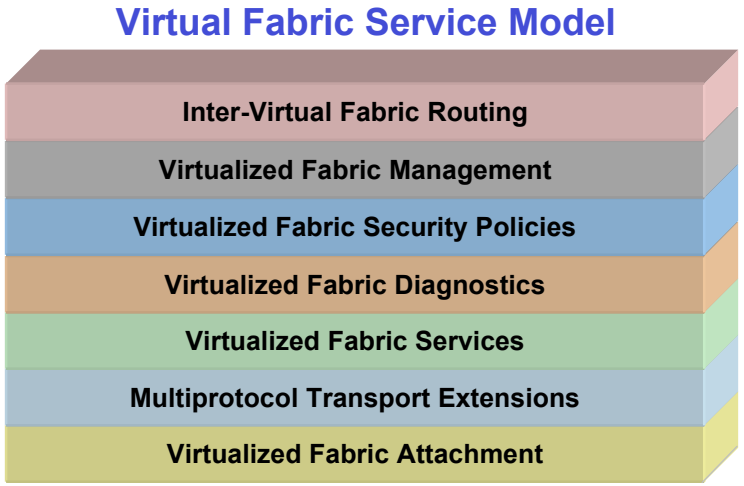
VSAN Benefit:
Build Very Large-Scale SANs

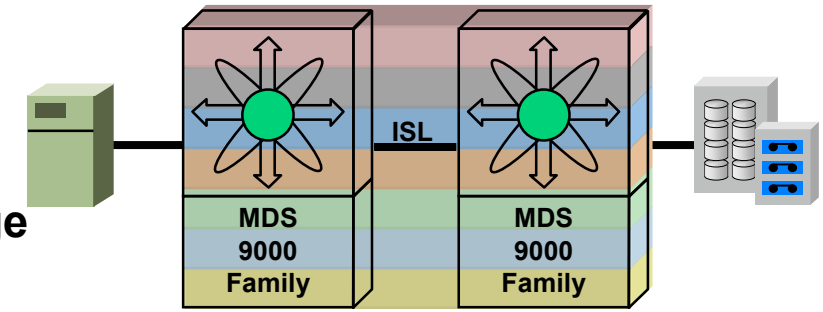
- **Independent instances of Fibre Channel services for each VSAN**
  - **Name server, management server, principal switch selection, FSPF, Zoning, etc.**
  - **Each service runs and is managed independently**
  - **Localized fault-isolation**
- **Hardware-based isolation of tagged traffic belonging to different VSANs**
  - **No special drivers or configuration required for hosts and targets**
  - **Traffic tagged at Fx_Port ingress and carried across EISL links between Switches**



Cisco Virtual SANs (VSANs):
Reduce TCO By Virtualization of The Fabric

- **To build a cost saving fabric virtualization solution, 7 key services are required:**
  - **Virtual Fabric Attachment – assign virtual fabric membership at the port level**
  - **Multiprotocol Extensions – extend virtual fabric service to iSCSI, FCIP, FICON, etc.**
  - **Virtual Fabric Services – create fabric services per virtual fabric (routing, zones, RSCNs, QoS, etc.)**
  - **Virtual Fabric Diagnostics – troubleshoot per virtual fabric problems**
  - **Virtual Fabric Security – define separate security policies per virtual fabric**
  - **Virtual Fabric Management – map and manage virtual fabrics independently**
  - **Inter-Fabric Routing – provide connectivity across virtual fabrics – without merging the fabrics**





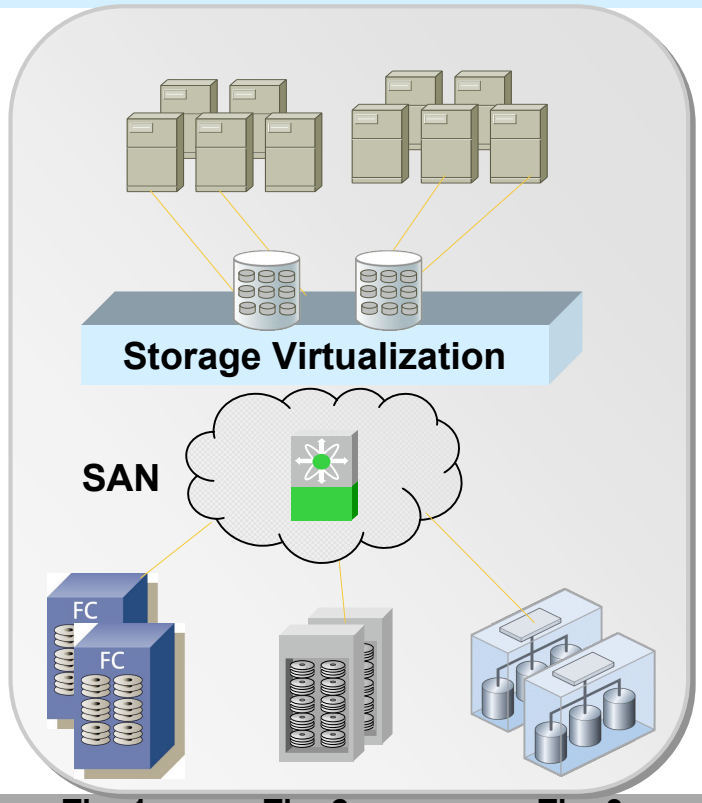
Full Service End-to-End Virtual Fabric Implementation

SAN Applications

- Consolidation/Optimization
- Network Virtualization
- **Storage Virtualization**
- Server-less Backup
- Network Extension
- Network Security

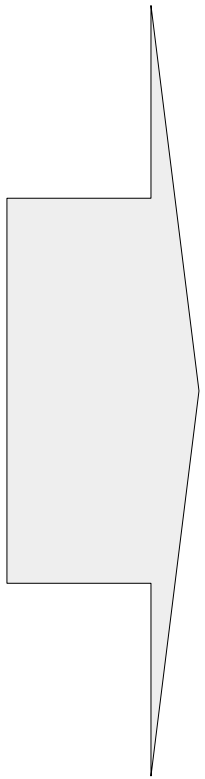
Storage Virtualization

Storage virtualization is layer of abstraction that hides servers from the physical aspects of the storage arrays and represents Logical Units that could be distributed across multiple physical disks which results in simplified management, rapid provisioning, and increased utilization of storage assets, thereby lowering the overall TCO



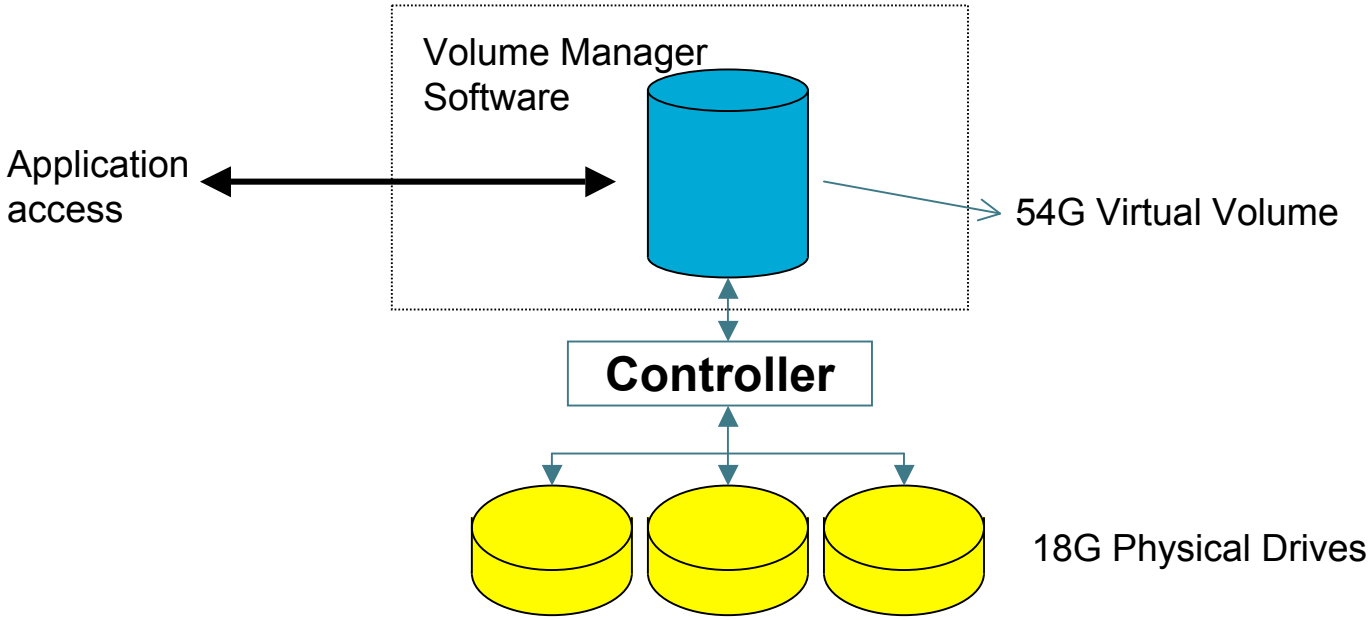
Customer Benefits of Storage Virtualization

Capability	Benefit
Simplify Storage Management	• A unified provisioning and management interface for heterogeneous storage pool
Optimize Storage Utilization	• Storage capacity pooling • Flexible storage allocation to reduce gap between “Allocated” vs. “Utilized” capacity
Non-Disruptive Data Migration	• No-impact to applications when data is being migrated across arrays due to lease rollover and workload re-balancing etc.
Enable Tiered Storage Model	• Allocate class-of-storage based on application needs i.e. policy-based allocation • Enable Information Lifecycle Management (ILM)



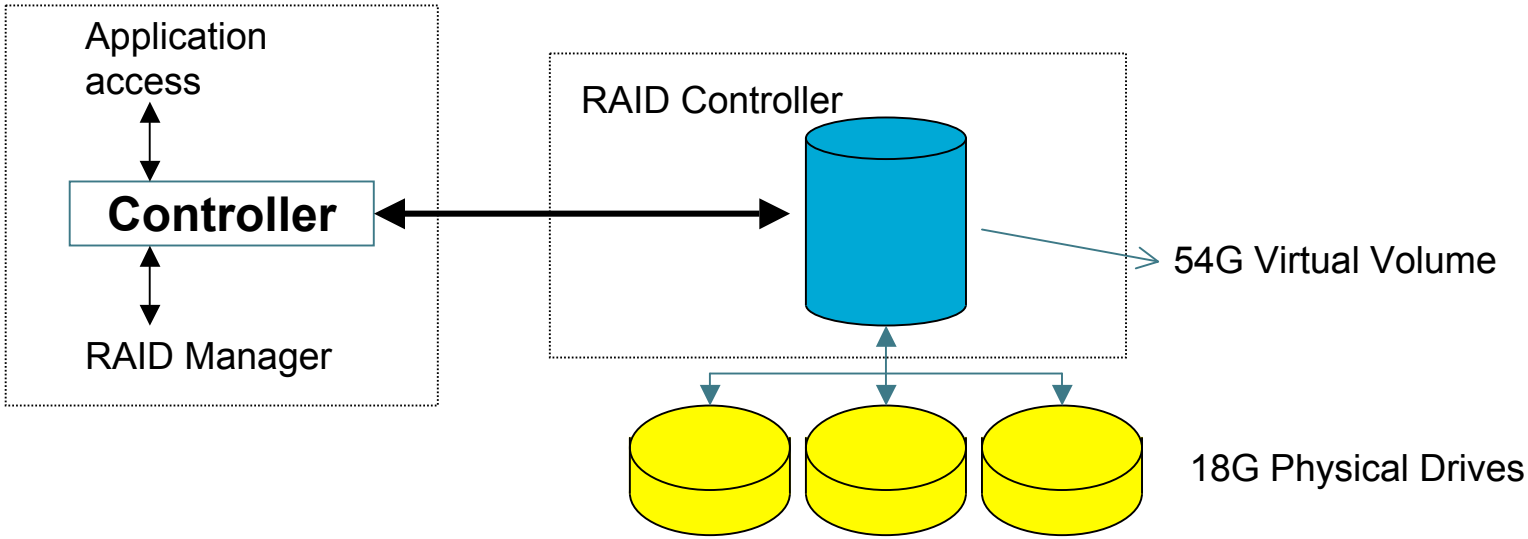
Host Based Virtualization

- **Veritas Volume Manager is an example of host based RAID which creates virtual structures and manage them on the host. Software RAID is layered on top of the operating system**



Storage Controller Based Virtualization

- **Controller based RAID maintains virtual structures on an external controller board.**
- **Dedicated controller for RAID provides better performance.**



RAID levels

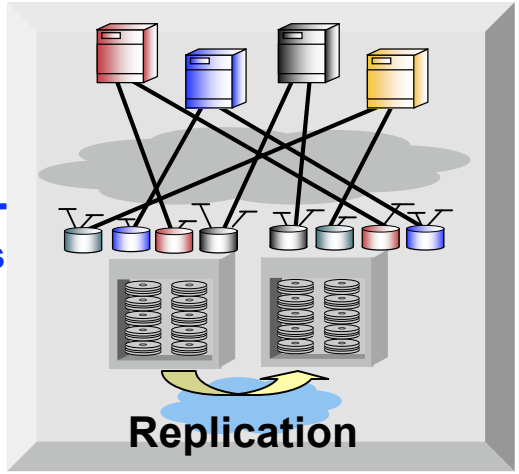
RAID Level	Description	Min Disks
0	Striping/Concatenation	2/1
1	Mirror	2
0+1	Striping/Concatenation then Mirror	4
1+0	Mirror then Striping/Concatenation	4
2	Hamming Code	N/A
3	Fix parity with concert I/O	N/A
4	Fix parity with Random I/O	N/A
5	Stripe with distributed parity with Random I/O	3 without log 4 with log

Network-hosted Storage Applications
Storage Virtualization

Storage Virtualization - Today

Host-based Apps
- App Integration
- Multi-pathing
- Volume Mgmt

Array-based Apps
- RAID/Volume Mgmt
- HA Upgrades
- Multiple Paths
- Snapshot
- Replication

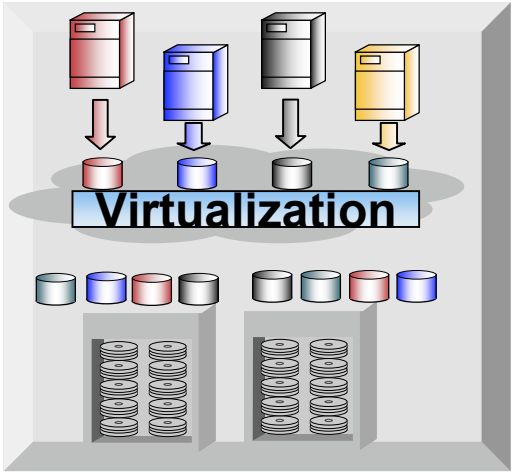


Network-hosted Virtualization

Host-based Apps
- App Integration
- Multi-pathing

Network-hosted Apps
- Volume Mgmt
- Snapshot
- Replication

Array-based Apps
- RAID
- HA Upgrades
- Multiple Paths

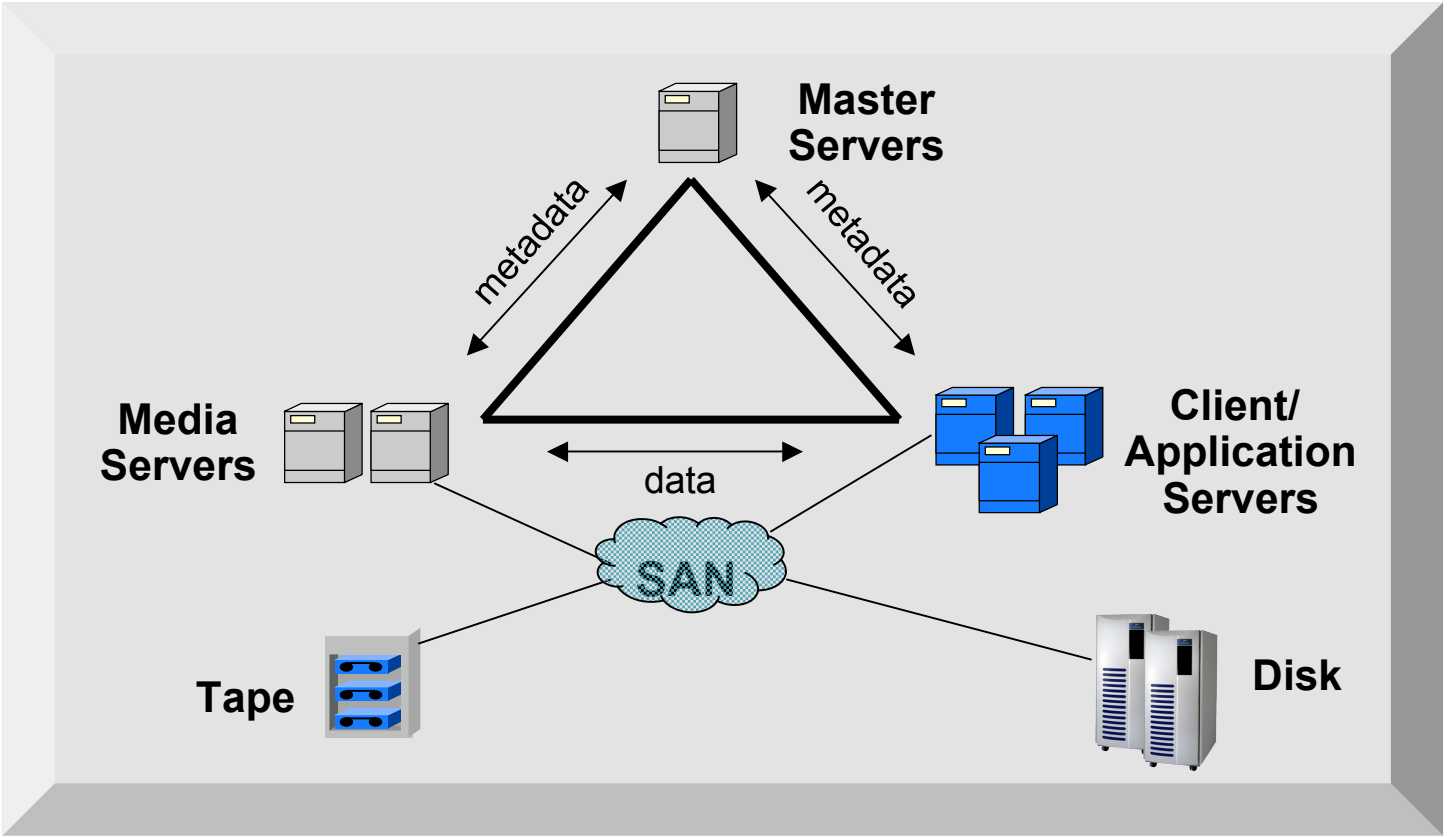


Benefits	Proof Points
Information Lifecycle Management	• Simplified management • Non-disruptive data migration across tiered storage
Increased Storage Utilization	• Heterogeneous storage pooling • Flexible storage provisioning
Improved Business Continuance	• Supports point-in-time copy, replication • Flexible data protection services

SAN Applications

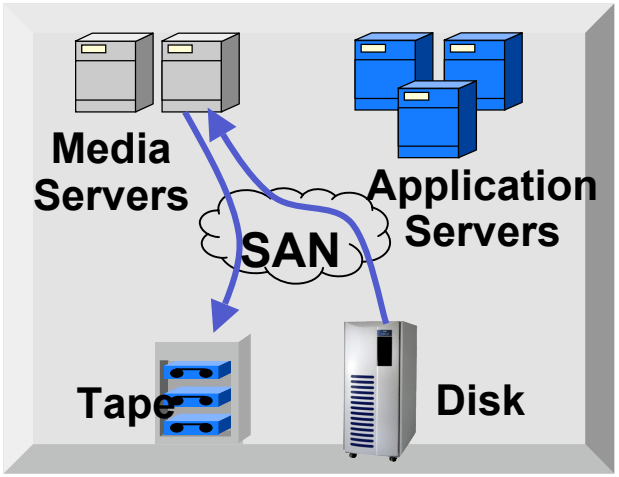
- Consolidation/Optimization
- Network Virtualization
- Storage Virtualization
- **Server-less Backup**
- Network Extension
- Network Security

Backup Components



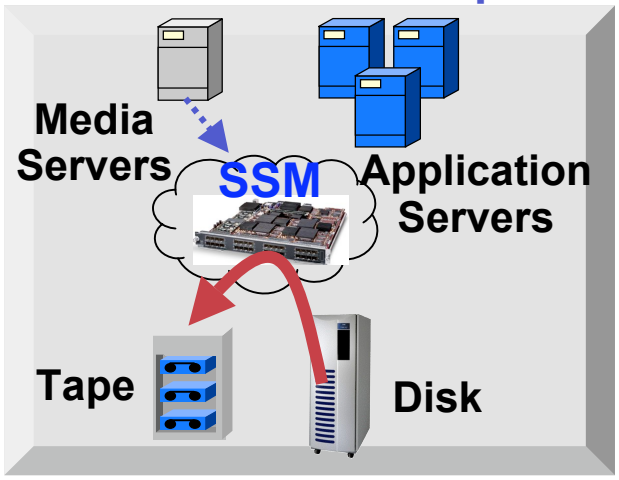
Serverless Backup

Server Backup - Today



Network Accelerated Serverless Backup

Computer Associates
CommVault® Systems
Unified Data Management
VERITAS™
EMC² legato



Customer Benefit	Proof Points
Lower TCO	• Offload I/O & CPU work from Media Servers to SSM • Reduce server administration & management tasks
Higher Performance & Reliability	• Each SSM delivers up to 20 Gbps throughput • SSM integrated into a high availability MDS platform

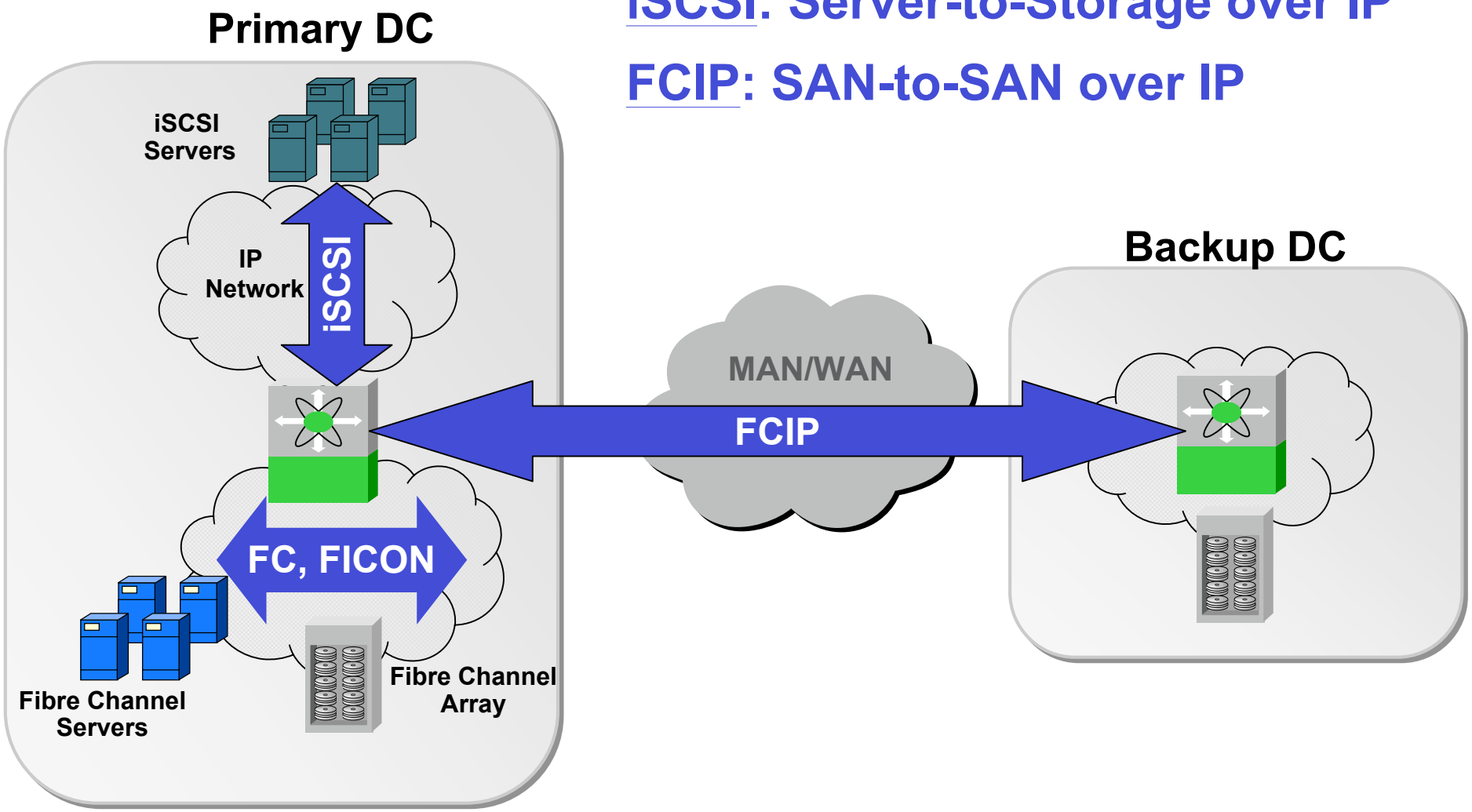
SAN Applications

- Consolidation/Optimization
- Network Virtualization
- Storage Virtualization
- Server-less Backup
- **Network Extension**
- Network Security

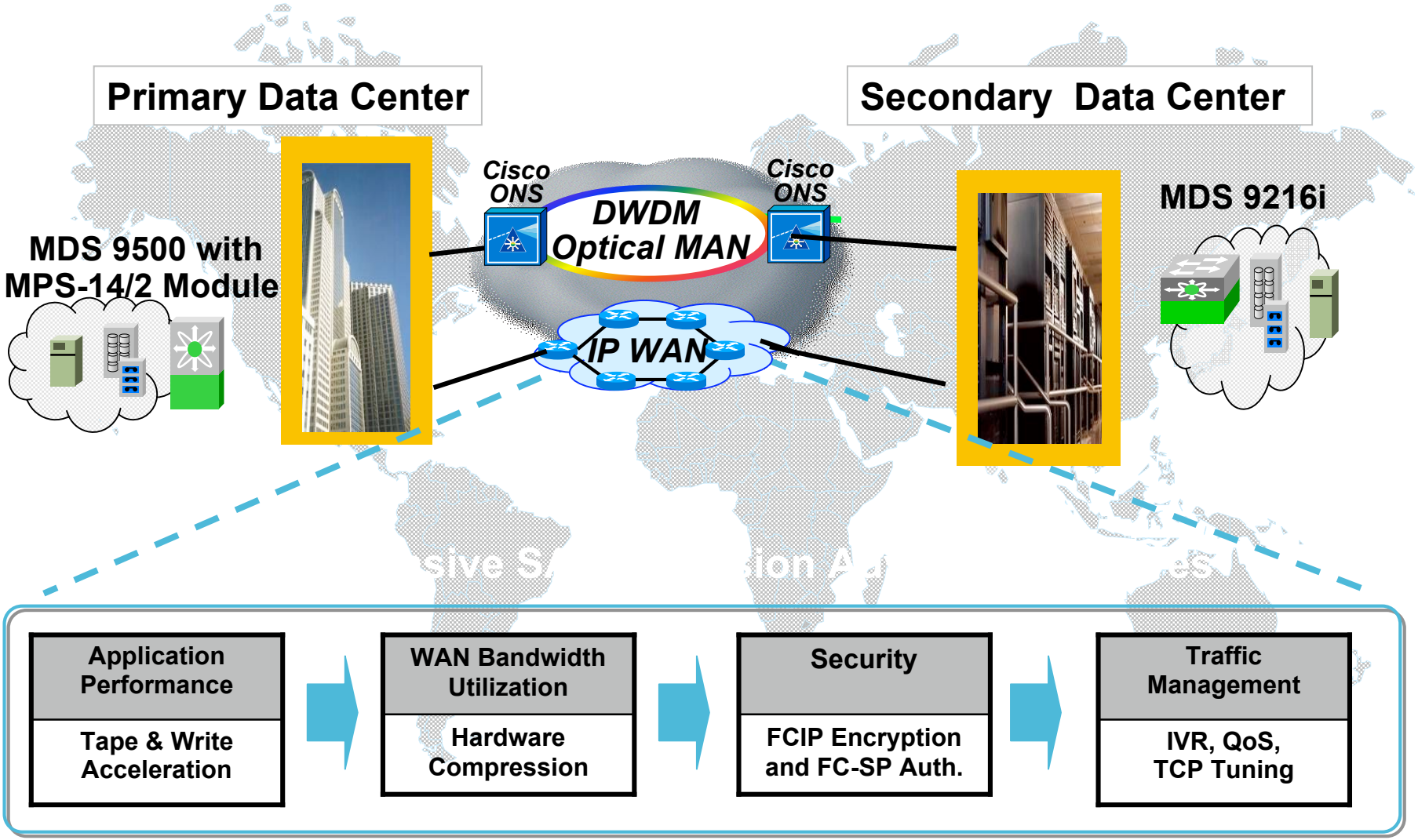
Integrated Multi-Protocol Connectivity:
Connectivity Options for Cost Effective SANs

iSCSI: Server-to-Storage over IP

FCIP: SAN-to-SAN over IP



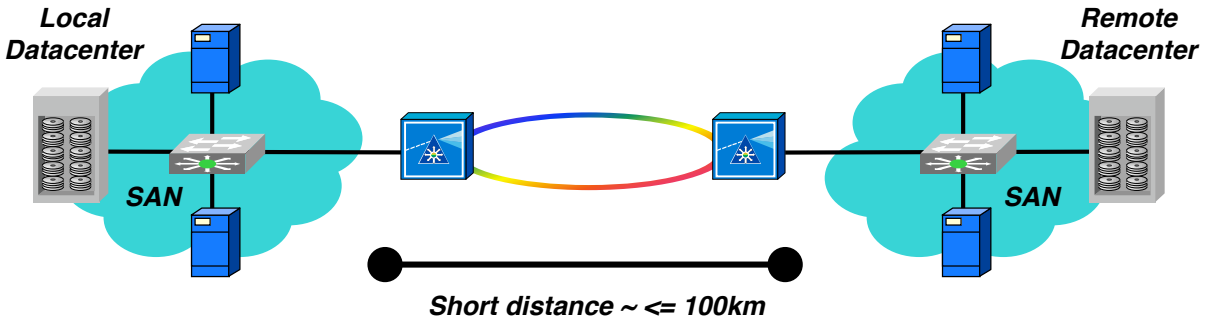
Integrated Business Continuity Solution:
Comprehensive SAN Extension Solution



Multiprotocol support: SAN Extension / IP SAN Extension

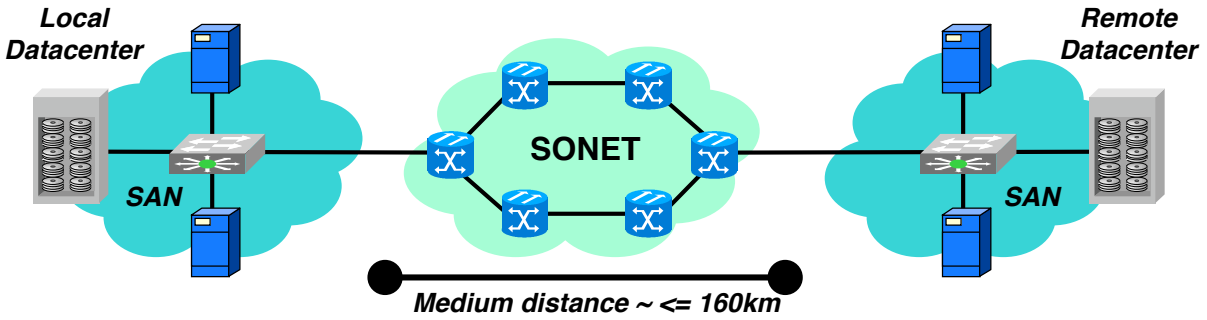
FC over DWDM/CWDM

- short distance
- dark fiber available
- dedicated links
- lowest latency - suitable for sync apps



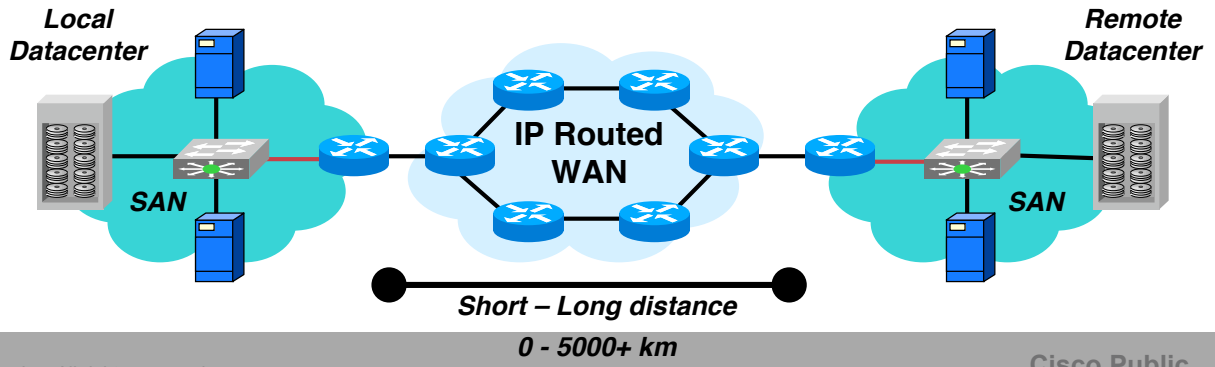
FC over SONET/SDH

- short – intermediate distance
- dark fiber not avail. – distance, cost, exhaust
- links may be shared
- suitable for most synchronous apps

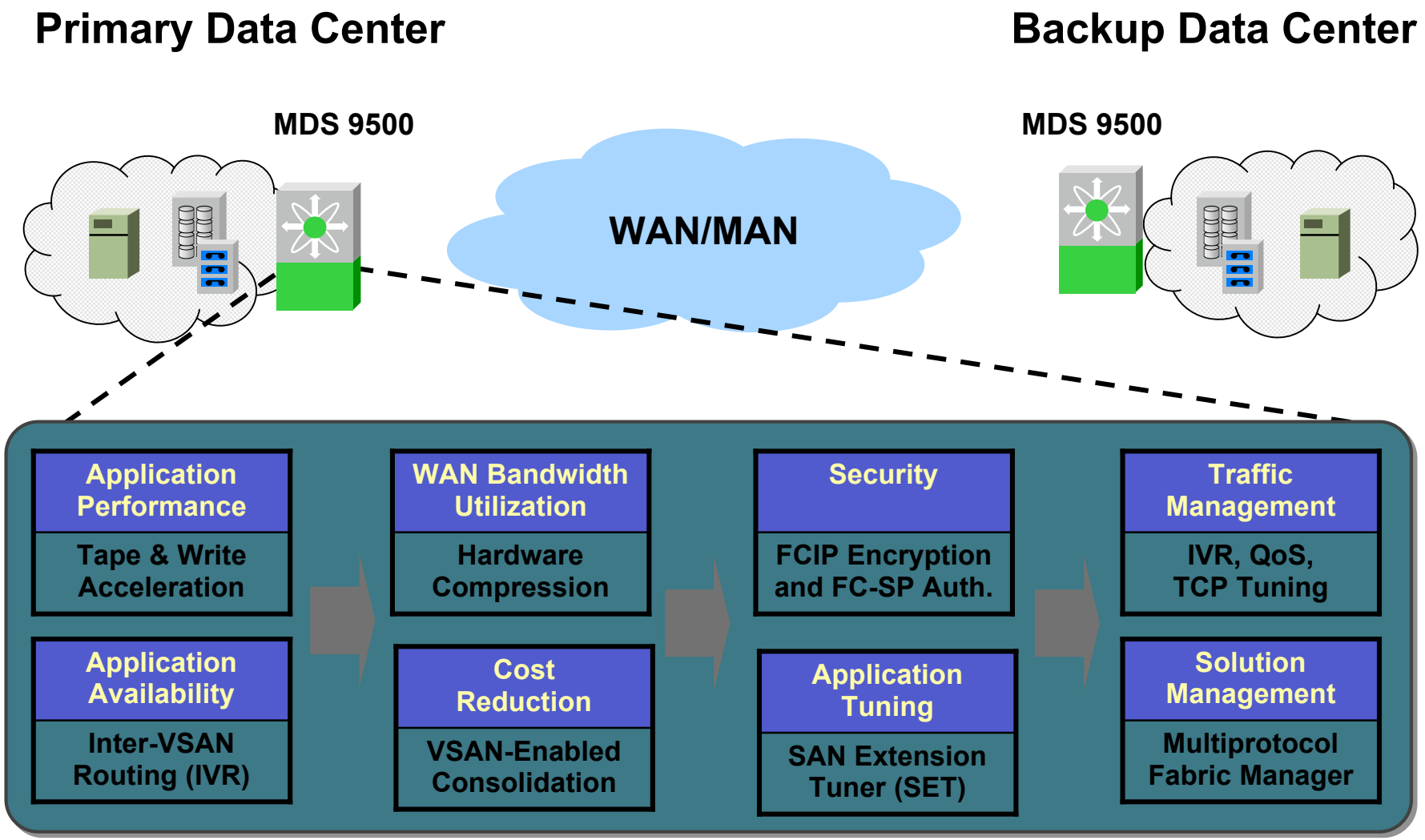


FC and FICON over IP

- short – long distance
- dark fiber not available
- links may be shared
- suitable for sync apps across metro ethernet
- suitable for async applications across WAN



SAN Extension Solution for Business Continuity:



SAN Applications

- Consolidation/Optimization
- Network Virtualization
- Storage Virtualization
- Server-less Backup
- Network Extension
- **Network Security**

Comprehensive SAN Security:
Address Compliance and Regulation Requirements

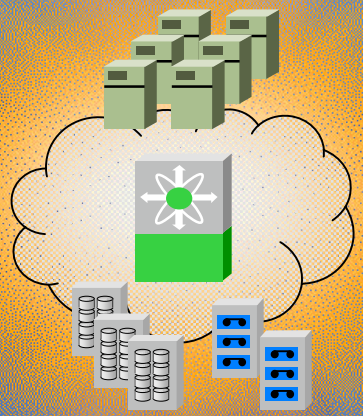
Data Integrity and Encryption

Security for Data-in-Motion

-IPSec for FCIP & iSCSI

Management Access Controls

- SSHv2, SNMPv3, SSL
- Centralized AAA
  -RADIUS, TACACS+
- Role Based Access Controls
- VSAN based RBAC
- IP ACLs



Device Authorization and Authentication

- Port Security
- Fabric Binding
- Host/Switch Authentication
- iSCSI CHAP Authentication

Server and Target Access Controls

- VSANs
- Hardware Zoning
- LUN Zoning
- Read-only Zones

Cisco
Driving Standards and Innovation

- **T10 and T11 (SCSI, FC)**
  - Technical editors
- **Security**
  - Co-authored FC-SP standard
- **Virtualization**
  - Chair of T11.5 working group on FAIS API and co-author
- **VSANs**
  - Technical editor of VSAN standard
  - Inter-VSAN routing submitted to T11
- **Network management**
  - Defined CIM MIBs
  - Implemented SMI-S, FDMI
- **Storage Networking Industry Association Europe**
  - Board Member, Chair of Education Committee
- **IETF (iSCSI)**
  - Co-authored iSCSI standard

Storage Networking Industry Association

IP Storage (IPS) Working Group

Technical Committee T10 (SCSI)
Technical Committee T11 (Fibre Channel)

FC Primer

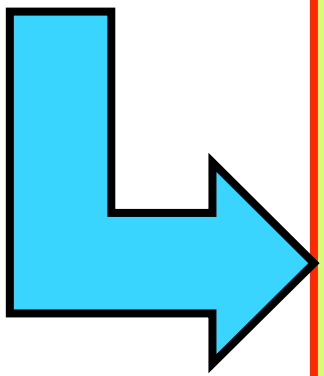
Structure & Concepts

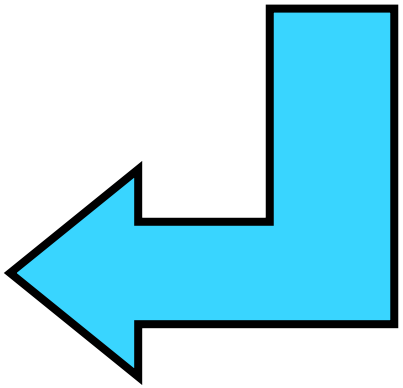
Channels

- Connection Service
- Physical Circuits
- Hardware Reliable Transfers
- High Speed
- Low Latency
- Short Distance
- Hardware intense

Networks

- Connectionless
- Logical Circuits
- Unreliable Transfers
- High Connectivity
- Higher Latency
- Longer Distance
- Software intense





Fibre Channel

- **Circuit & Packet Switched**
- **Reliable Transfers** - BER<10^{-12}
- **High Data Integrity** – error detect
- **High Data Rates** – 100 and 200MB/s
- **Low Latency** – short delay 0.6-3 us
- **High Connectivity** – 24 bit address
- **Long Distance** – 10Km to 100Km

Structure & Concepts

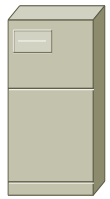
Fibre Channel Strength

- Longer distance – Serial
- Higher sustained throughput
- Encapsulate multi-protocol (SCSI, IP or ATM)
- Scalable
- Lower Hardware cost then SCSI
- Open Standards

Structure & Concepts

Fibre Channel Node

- A Fibre Channel network has a minimum of one link between two nodes. A node could be a computer, disk or a tape drive which is a either source or destination of information.
- Each node has a unique 64bit Node_Name (WWN).



Structure & Concepts

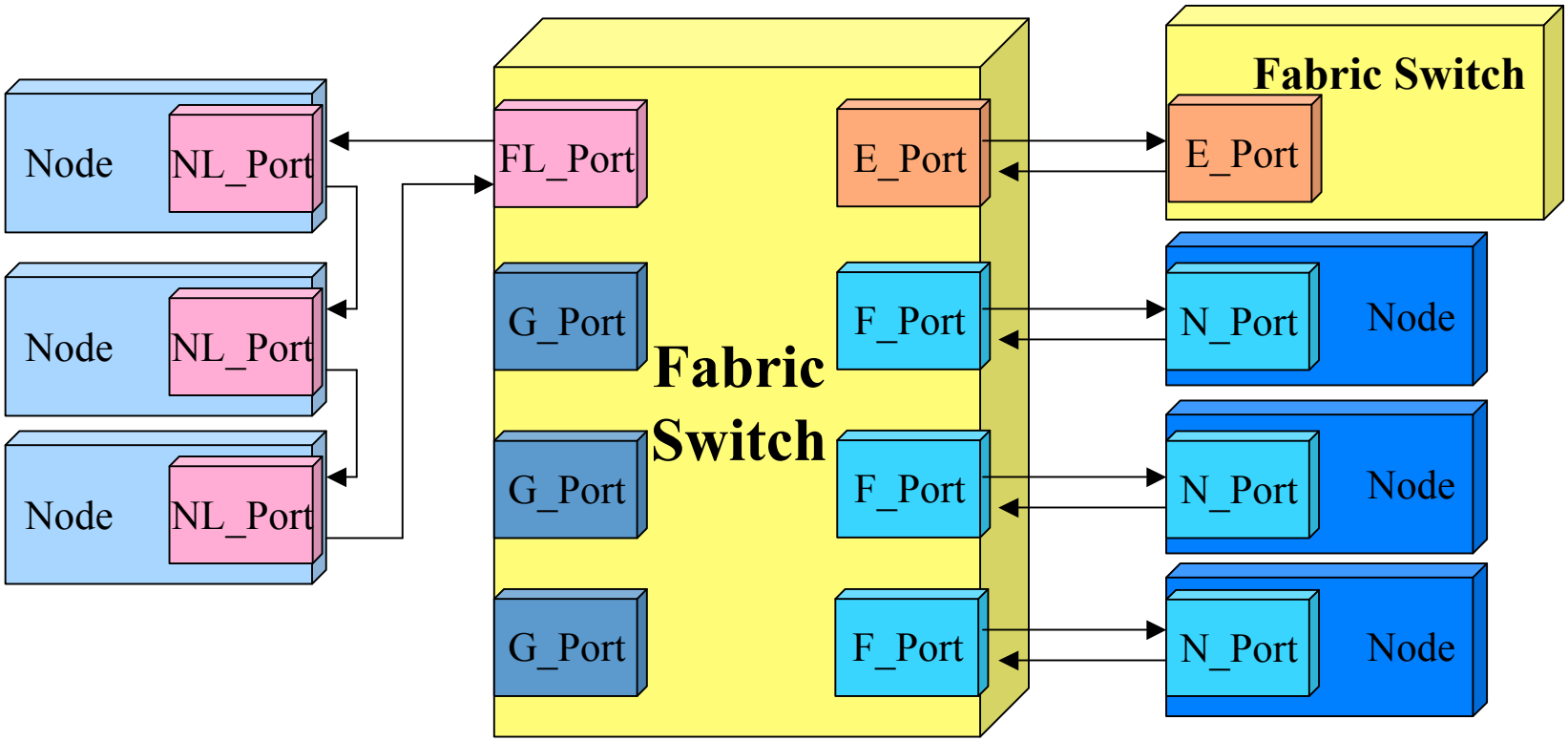
Fibre Channel Node Port

- Each node requires a physical (Hardware) port to communicate with other nodes via their ports over the Fibre Channel medium.
- The port is responsible of creating, controlling and managing the Fibre Channel functions.
- Each port has a unique 64-bit Port_Name.

Structure & Concepts

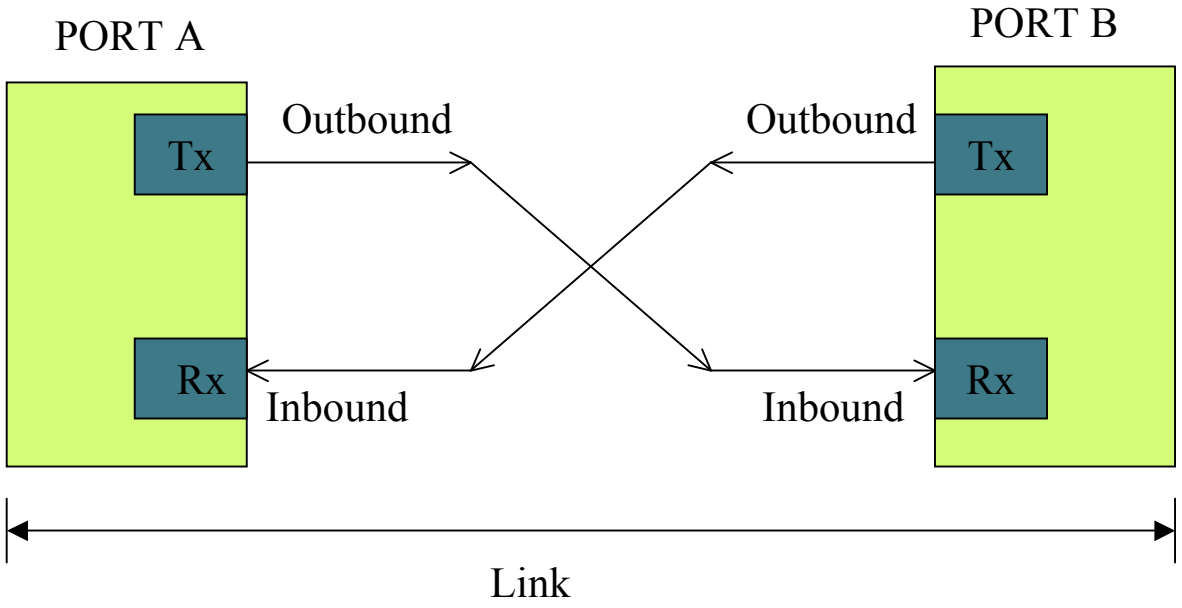
- ‘N’ port: Node ports used for connecting peripheral storage devices to switch fabric or for point to point configurations. Can be considered the end port.
- ‘F’ port: Fabric ports reside on switches and allow connection of storage peripherals (‘N’ port devices)
- ‘L’ port: Loop ports are used in arbitrated loop configurations to build storage peripheral networks without FC switches. These ports often also have ‘N’ port capabilities and are called ‘NL’ ports.
- ‘E’ port: Expansion ports are essentially trunk ports used to connect two Fibre Channel switches
- ‘G’ port: A generic port capable of operating as either an ‘E’ or ‘F’ port. Its also capable of acting in an ‘L’ port capacity. Auto Discovery.

Fabric Switch



FC Link

- The port connects to a topology via a link
- FC-0 level of FC-PH describes the FC link
- Port A & B could be a N_Port or F_Port



Fabric

- **Fabric**

  The entity which interconnects N_Ports

  Provides routing based on Destination address. Internal addressing system for passing data frames through multiple frames.

  Fabric may be:

  Point to point – No routing required.

  Switched - Routing provided by switch.

  Arbitrated Loop – Routing is distributed throughout attached L_Ports.

Fibre Channel Names

- The Fibre Channel name is a 64-bit unique identifier to identify nodes, ports and fabrics used whenever a login session is established .
  - When a Fibre Channel name is associated with a port it is called Port_Name or World Wide Port_Name (WWPN)
  - When a Fibre Channel name is associated with a node it is called Node_Name or World Wide Node_Name (WWNN)
  - When a Fibre Channel name is associated with a fabric, it is called Fabric_Name or World Wide Fabric Name (WWFN)
- The first four bit of the name identify the Name Assignment Authority (NAA) and the remaining 60 bits are determined by the format in NAA field.
- The 48 LSB are based on IEEE 802.1A MAC address
- For a complete list of NAA identifiers, refer to pg 113 of FC-PH rev 4.3

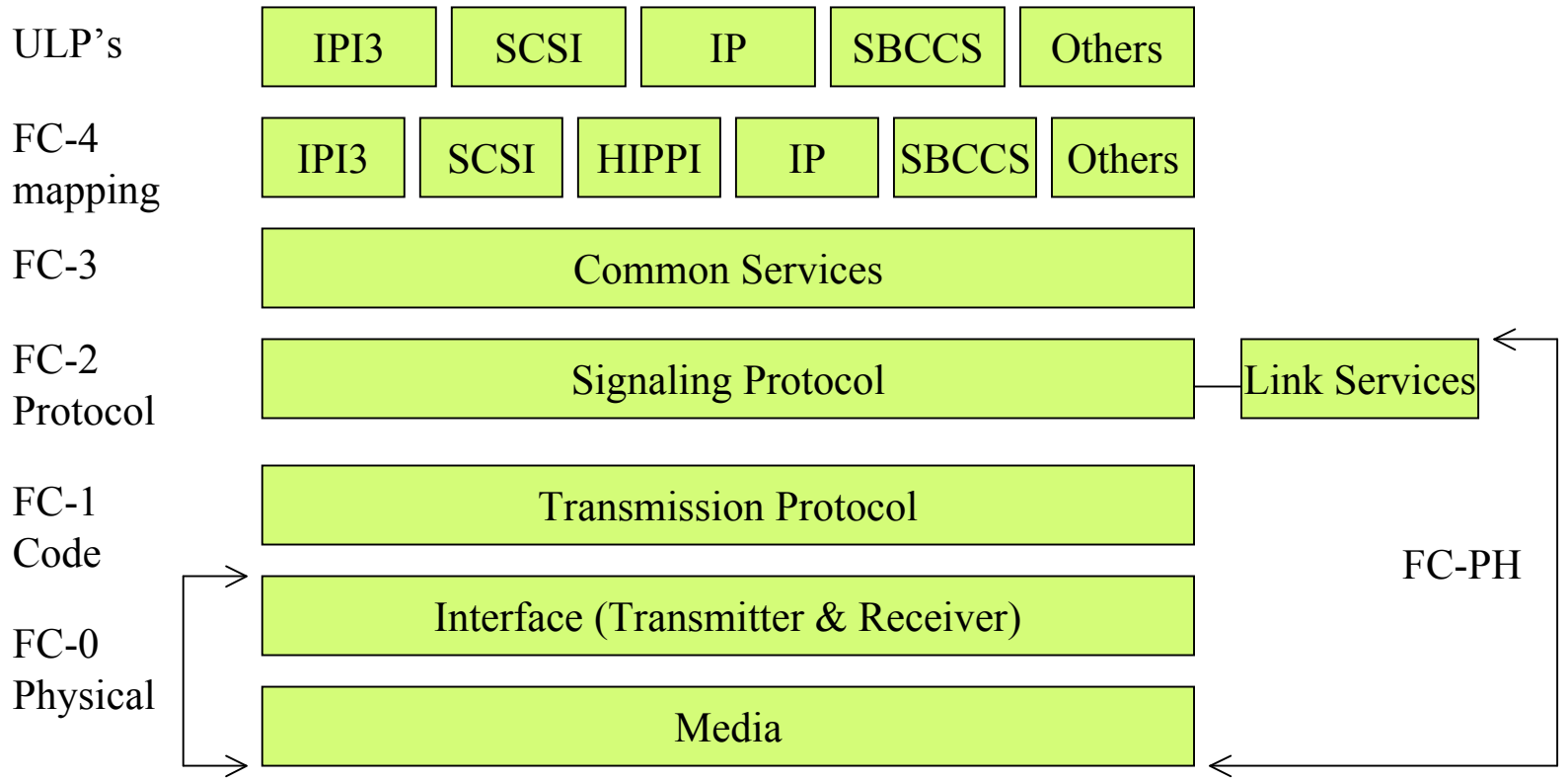
NAA Identifiers				
63	62	61	60	NAA
0	0	0	0	Ignored
0	0	0	1	IEEE
0	0	1	0	IEEE Extended
0	1	0	0	IP

Fibre Channel Names

Name Identifier Formats			
NAA	Name_Identifier (64 bits)		
	NAA ID 4 bits	60 bit field	
		12 bits	48 bits
IEEE	0 0 0 1	zeros	IEEE address for node
IEEE extended	0 0 1 0	N-Port	IEEE address for node
IEEE extended	0 0 1 0	F-Port	IEEE address for fabric
IP	0 1 0 0	Zeros (28 bit)	IP address (32 bits)

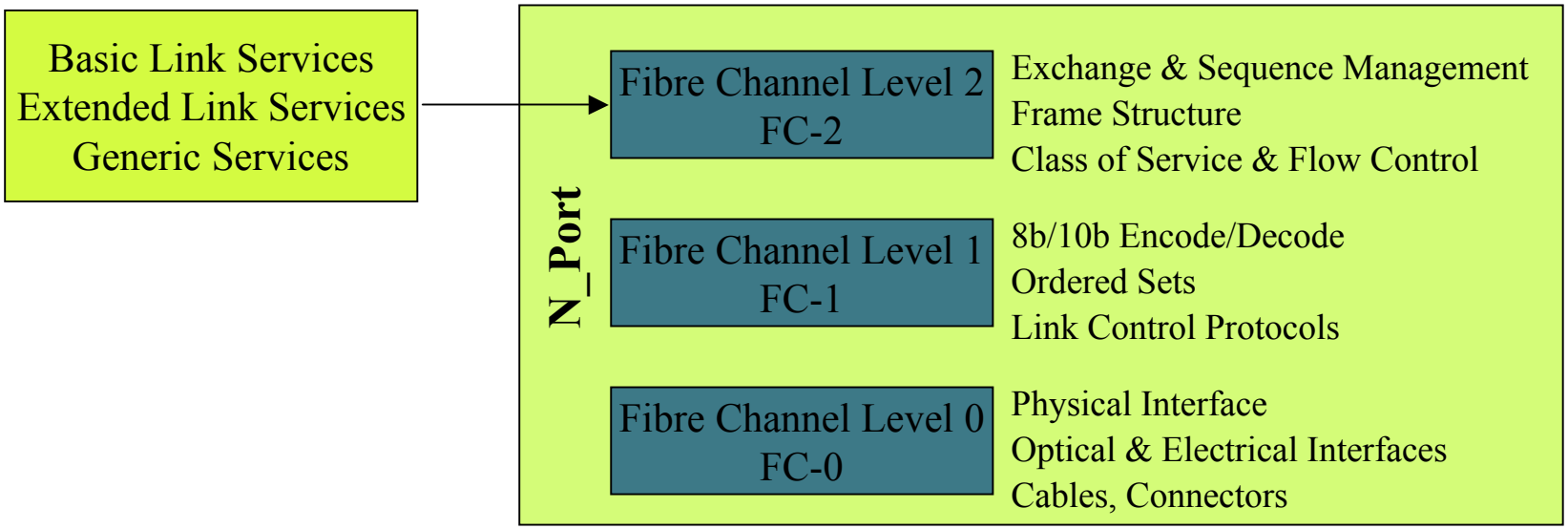
Structure & Concepts

- Fibre Channel is logically a bi-directional point-point serial data channel, structured for high performance capability.



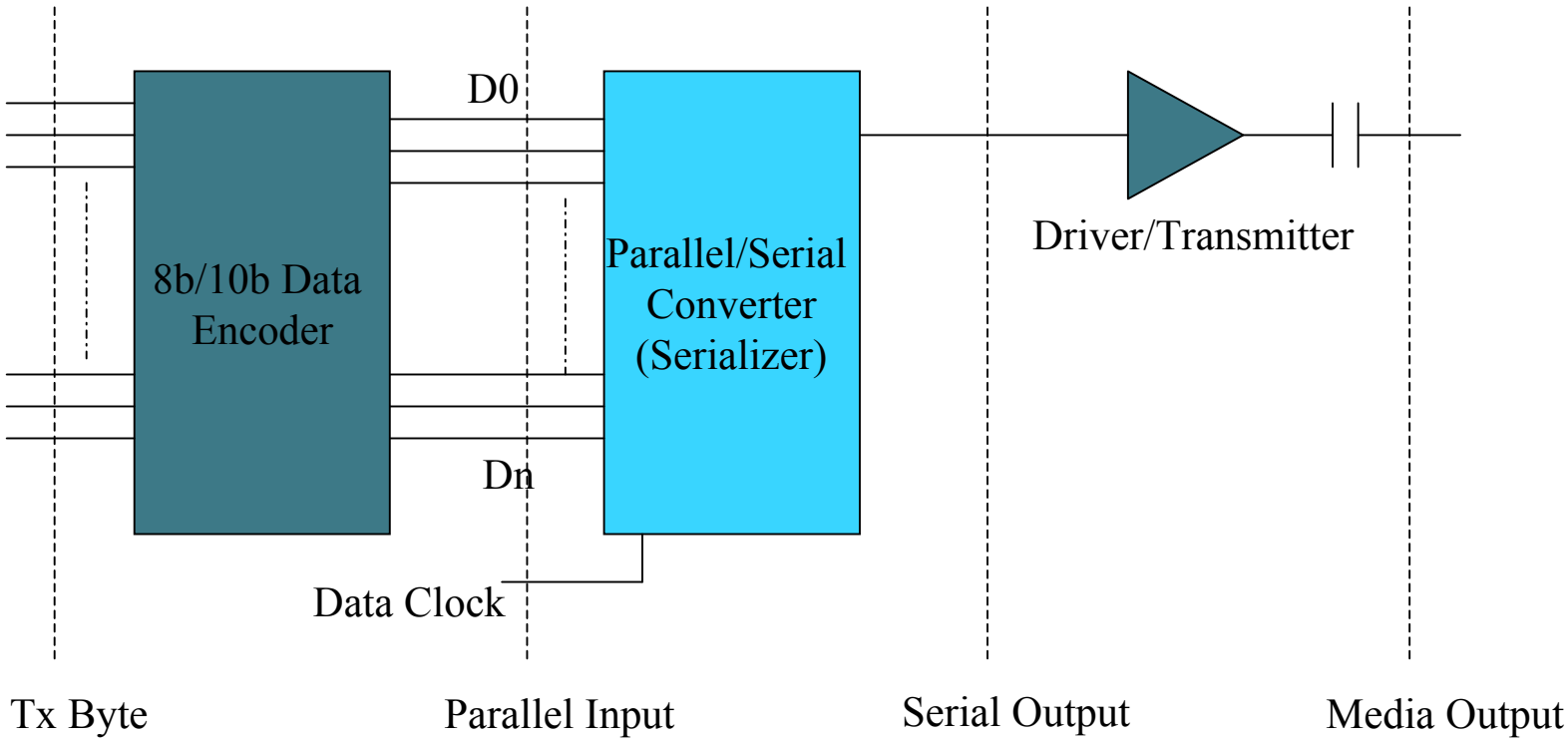
Structure & Concepts

- FC-PH is the Fibre Channel Physical & Signaling Interface.
- FC-0, Physical interface specifies variety of media and associated drivers and receivers capable of operating at various speeds.
- Transmission Protocol, FC-1, uses 8b/10b transmission code.
- Signaling Protocol, FC-2, specifies rules and mechanism needed to transfer blocks of data end to end.
- FC-4 provides upper layer protocol mapping to FC-PH



8b/10b Transmission Code

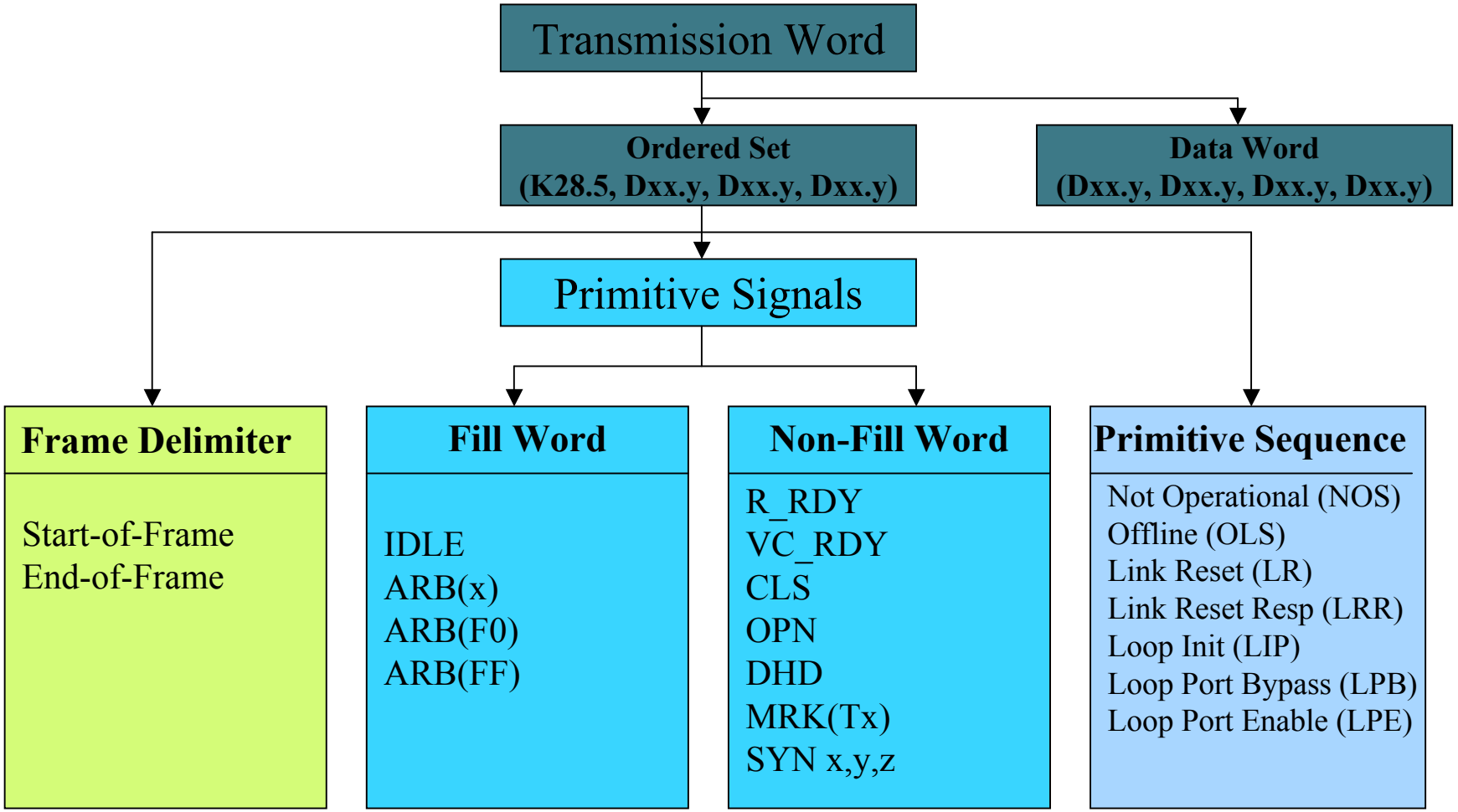
- 8b/10b is a data encoding and transmission scheme patented by IBM.
- Info transmitted over a fibre shall be encoded 8 bits at a time in to a 10 bit transmission character and then sent serially by bit.



8b/10b Transmission Code

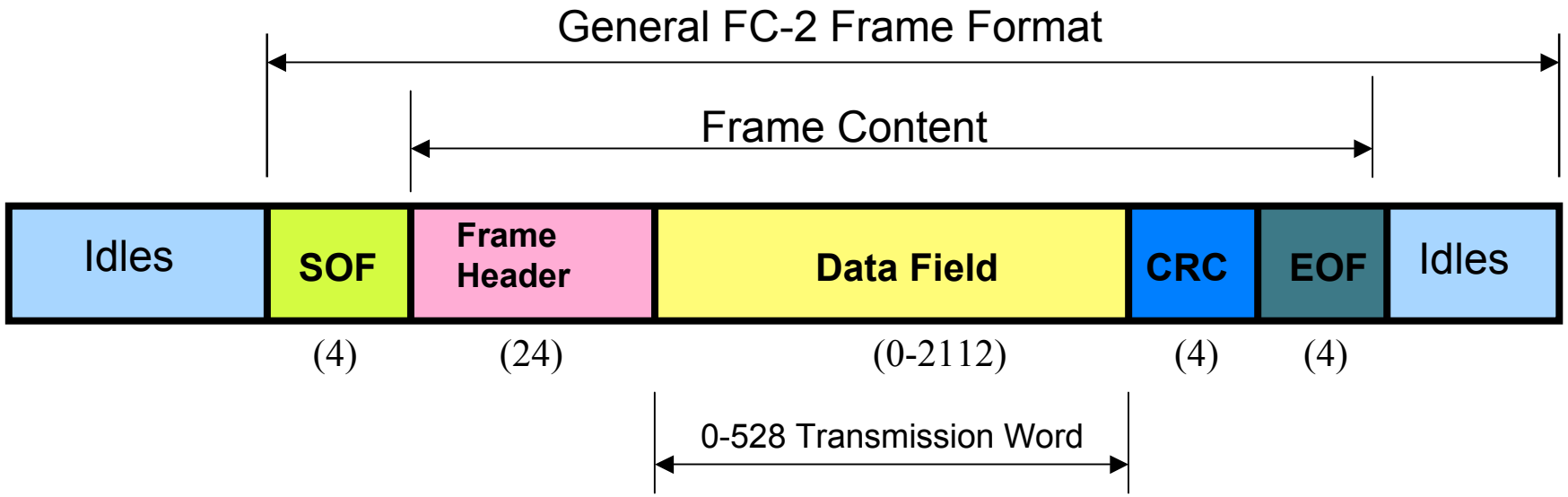
- The 10 bit transmission code supports all 256 eight combinations. There are 1024 10 bit patterns. 8b/10b encoding defines a number of unique special characters.
- Primary reason to encode is to improve the transmission characteristic of information.
  - Ensures that sufficient transitions are present in the serial bit stream to make clock recovery possible.
  - Increases the likelihood of detecting single or multiple bit errors that may occur.
  - Some of the special character patterns assist in word alignment.
  - Maintain a balance between one's and zero's transmitted, ensuring that the receiving signal is free of any DC-component (Running Disparity)

Ordered Sets



Frame Structure

- All FC-2 frames follow the general frame format as shown below.



Address Identifiers

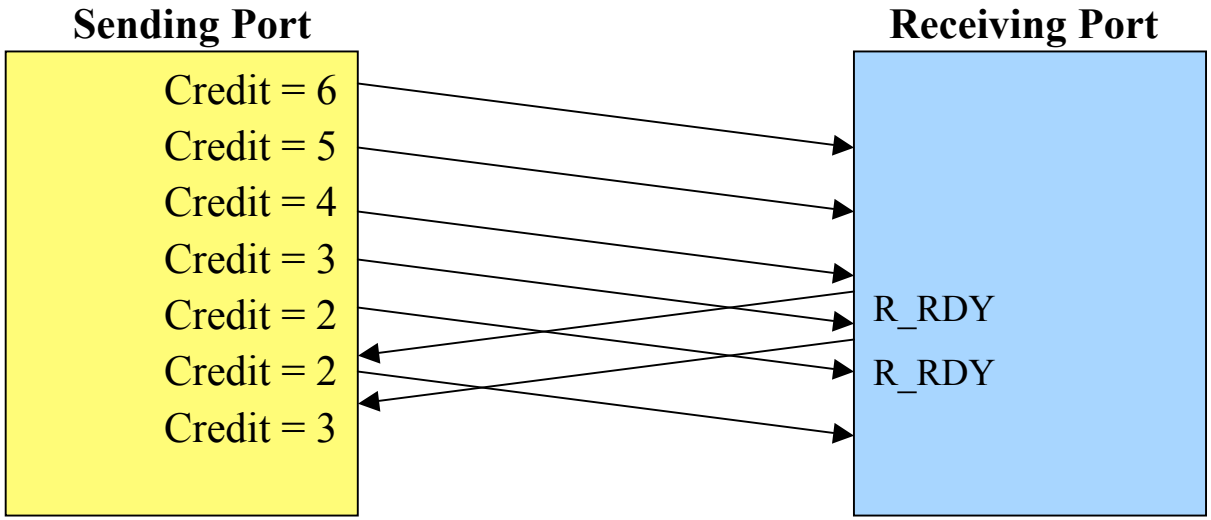
	8 bits	8 bits	8 bits
Switch Topology Model	Switch Domain	Area	Device
Private Loop (not connected to a switch)	00	00	**Arbitrated Loop Physical Address (AL_PA)**
Public Loop (connected to switch)	Domain	Area	AL_PA

Flow Control

- Flow control ensures that frames only move when there is buffer to store them
- Using Flow Control, the receiving port indicates to the transmitting port that it can accept the next frame.
- To pace the rate at which a source is allowed to transmit frames, Fibre Channel implements two different levels of flow control
  - Port to Port (Buffer-Buffer) is a Link Level protocol.
  - Source to Destination (End-End)

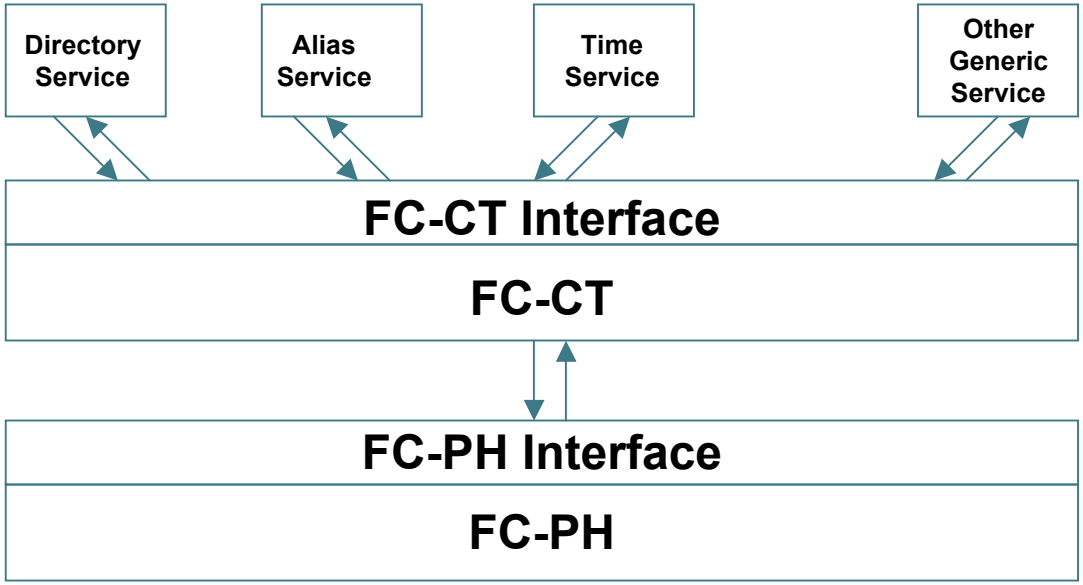
Link Level Flow Control

- Flow control protocols use a Credit concept. Credit is the permission granted by the receiving port to the sending port to send a specified number of frames. Suspend transmission when credit =0
- Buffer to Buffer Credit (BB_Credit) is granted during the login process and depends on interconnection topology.
- When R_RDY is received, BB_Credit is incremented



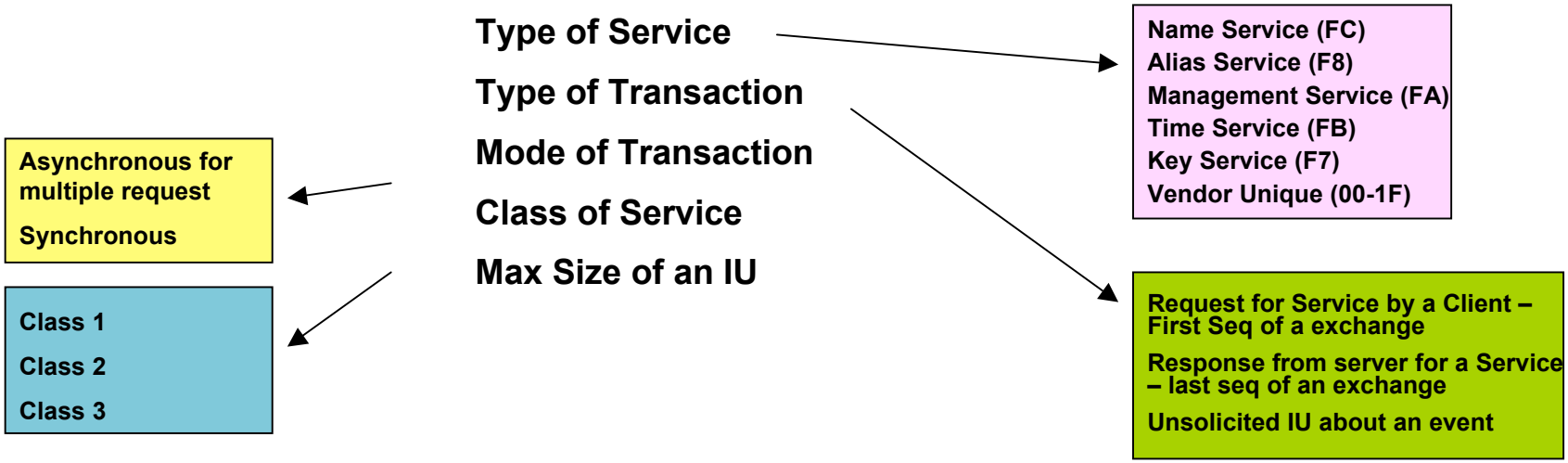
FC-CT

- FC-GS shares a Common Transport (CT) at the FC-4 level.
- The CT provides access to service (accessed via WKA) and set of Service parameters



FC-CT

- **Information that a CT delivers to a Service**

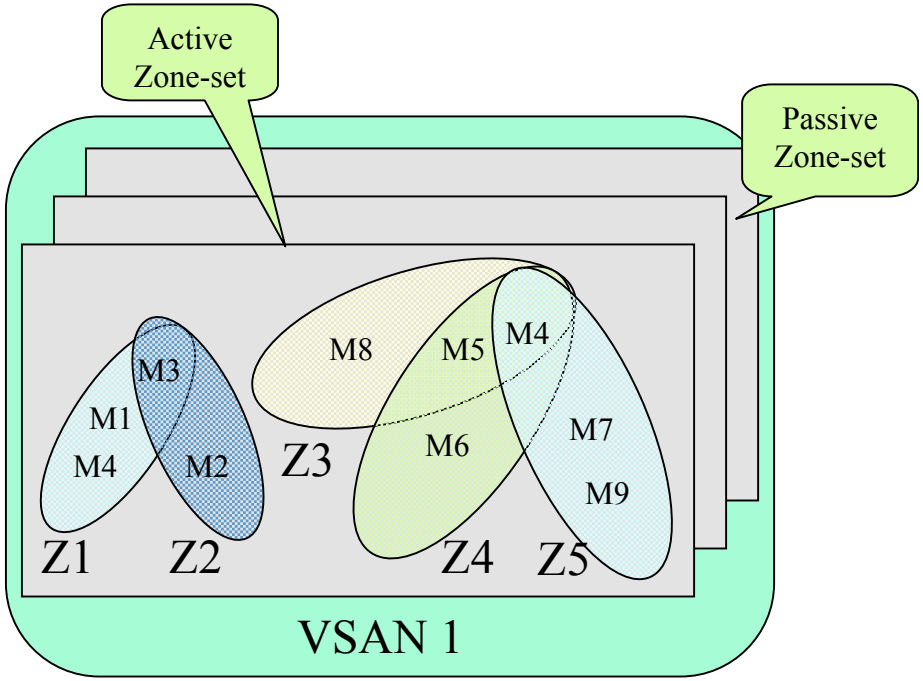


Zoning

- Zoning is used to control access in a SAN
- Zoning is enforced within a VSAN
- MDS supports both hard zoning and soft zoning
  - Soft zoning enforced by Name Server queries
  - Hard zoning enforced on every frame by the ASIC
- Fully compliant with FC-GS3, FC-SW2 and FC-MI

Zoning

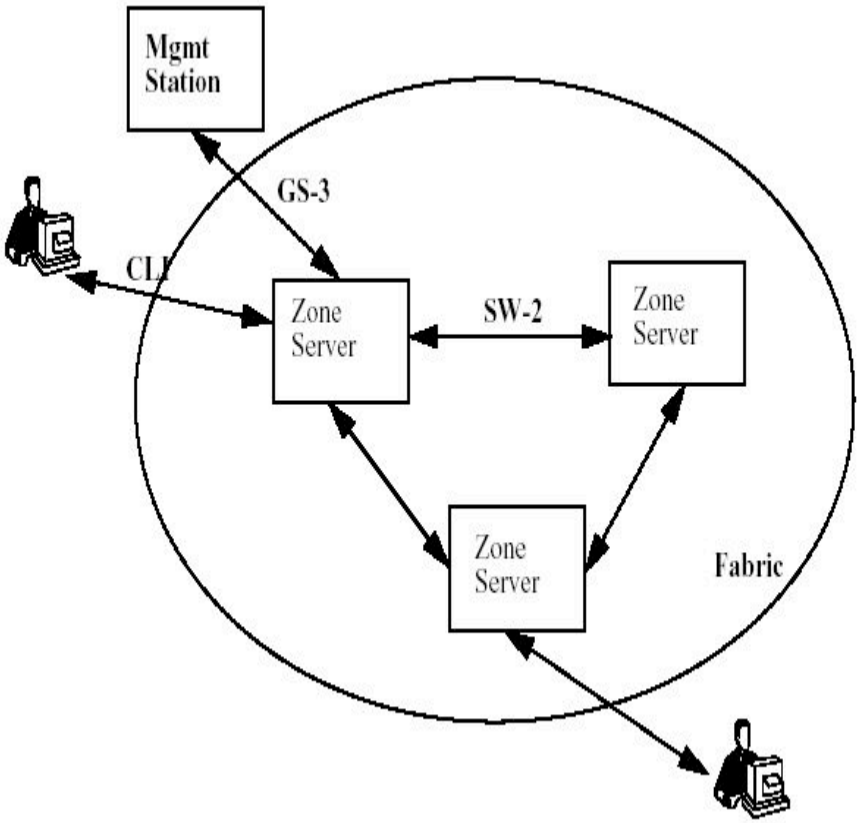
- **A fabric can have multiple VSAN's**
- **Each VSAN can have multiple zones where zones can not span across multiple VSANs.**
- **A zone consists of multiple zone members**
- **Each zone member can belong to multiple zones.**
- **Multiple zones in a VSAN constitutes a zone set.**
- **Each VSAN can have multiple zone set but there can be only one active zone set.**



Zoning DataBase (ZDB)

- **Each Zone Server interacts with other zone servers in the fabric to maintain a consistent zoning information of the fabric. Change and merge Protocol is used to synchronize the database among zone servers.**
- **Change Protocol is used to communicate any changes in the database.**
- **Merge Protocol is used whenever an ISL between two switches become operational. Merge protocol will merge the two databases to create a new Zoning DataBase (ZDB).**
- **If the members of a zone between two switches are not identical, the ISL becomes isolated. The import/Export feature is used to update ZDB.**
- **Vegas supports both active and full zone sets**

  Active zoneset is same across fabric



Switch Flow