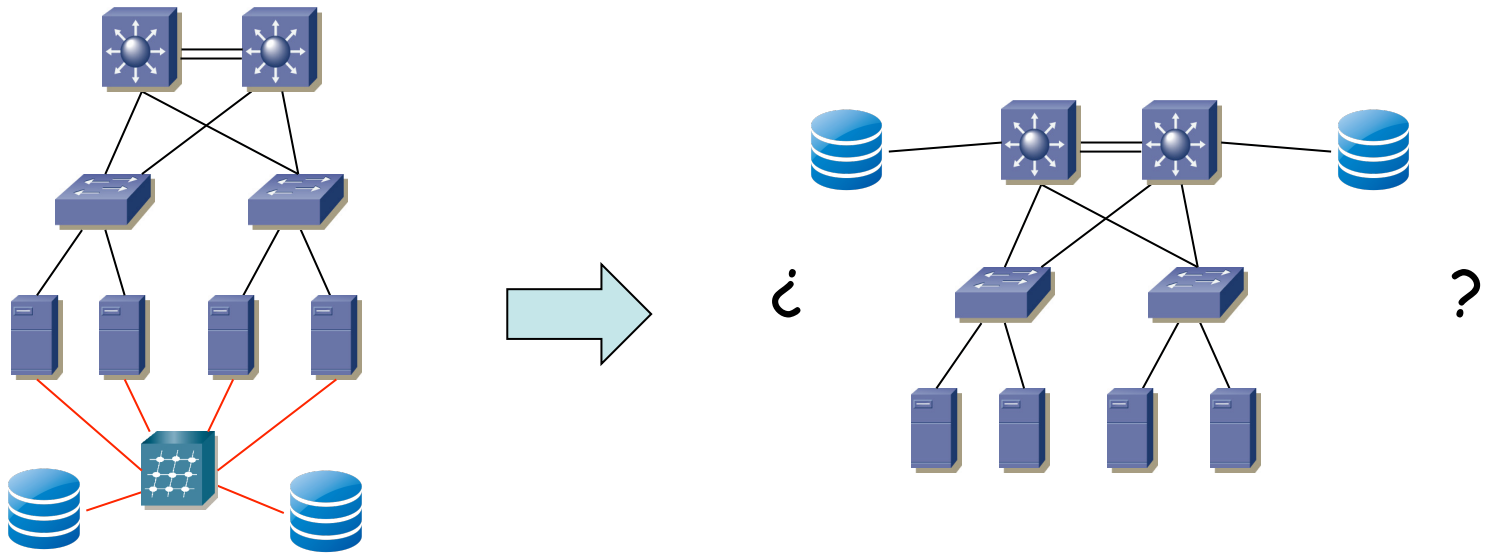
I/O Consolidation

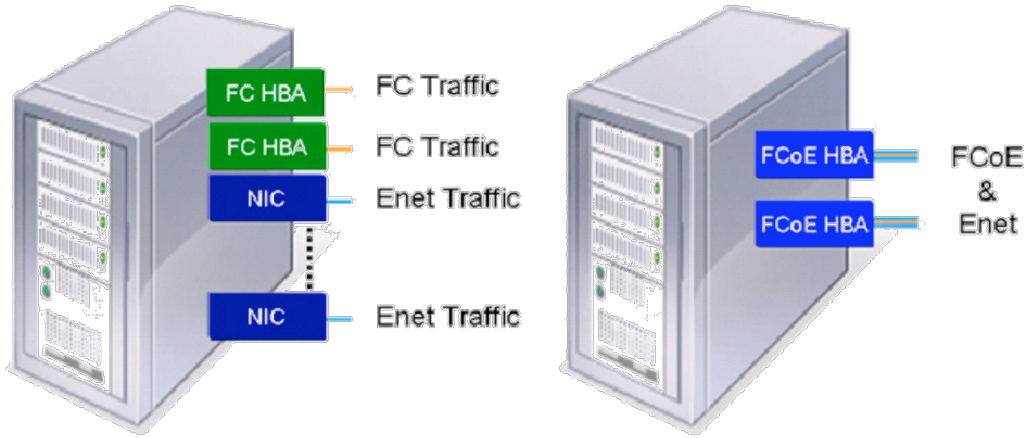
I/O Consolidation

- Consolidación: Emplear la misma infraestructura física para transportar múltiples tipos de tráfico
- En su día se produjo consolidación entre las redes de voz y datos
- Ahora entre la de datos (+voz) y la de almacenamiento
- Esto ya lo hacía Infiniband pero no ha sido popular
- Parece que finalmente van a converger sobre Ethernet (¿ o sobre IP ?)



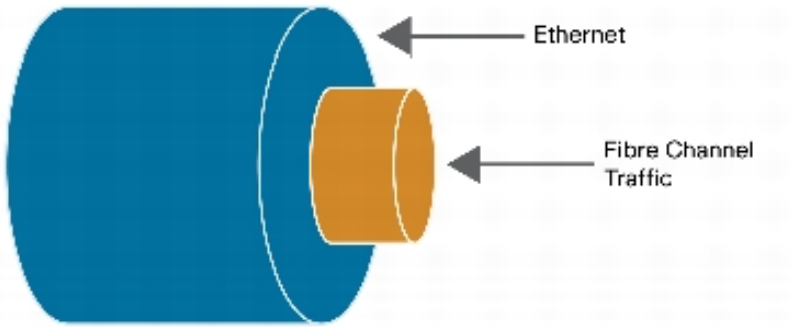
I/O Consolidation

- En lugar de NICs y HBAs tendremos solo CNAs
- CNAs (Converged Network Adapters) son NICs Ethernet que implementan por hardware parte de los protocolos
- Beneficios:
  - Simplifica el cableado y la infraestructura de red (pasamos de 2 redes a solo 1)
  - El equipo retirado nos ahorra costes de alimentación, refrigeración y espacio
- Desventajas
  - Una nueva tecnología a dominar y gestionar
  - Y hay que cambiar el hardware
  - Posibles acoplamientos
  - Mayor fragilidad



Tecnología

- Ethernet es la solución de LAN
- Fibre Channel la solución de almacenamiento en red
- Transportar los datos de LAN sobre FC no ha sido interesante por la falta de buen soporte de multicast/broadcast en FC
- SCSI requiere transporte fiable porque se diseñó para cables paralelos fiables y su recuperación ante pérdidas es lenta
- iSCSI es una opción pero tampoco se ha desplegado masivamente por el posible sobrecoste de TCP y porque no mantiene la gestión de FC
- En general los transportes sobre IP son más costosos en el hardware
- Finalmente Ethernet parece la solución ganadora, pero hay que modificarlo para evitar las pérdidas, que no afectan a SCSI



DCB

DCB

- *Data Center Bridging*
- Modificaciones a Ethernet por parte del IEEE 802.1 para permitir la convergencia con I/O en el data center
- Entornos con un producto retardo-ancho de banda limitado, así como un limitado número de saltos

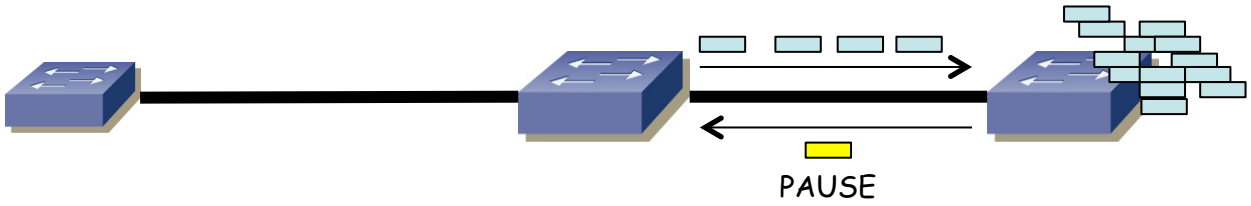
Feature	Benefit
Priority-based Flow control (PFC; IEEE 802.1 Qbb)	Provides capability to manage bursty, single traffic source on a multiprotocol link
Enhanced transmission selection (ETS; IEEE 802.1 Qaz)	Enables bandwidth management between traffic types for multiprotocol links
Congestion notification (IEEE 802.1 Qau)	Addresses the problem of sustained congestion by moving corrective action to the network edge
Data Center Bridging Exchange (DCBX) Protocol	Allows autoexchange of Ethernet parameters between switches and endpoints

Control de flujo en FC

- Fibre Channel ofrece una red sin pérdidas por congestión
- Mediante control de flujo salto a salto
- Esto le obliga a topologías simples
- El problema es que la congestión en un switch propaga el control *upstream*
- Puede afectar a flujos que no son responsables de la congestión
- En topologías complejas puede llevar a interbloqueos que degraden el rendimiento
- Así que requiere topologías simples

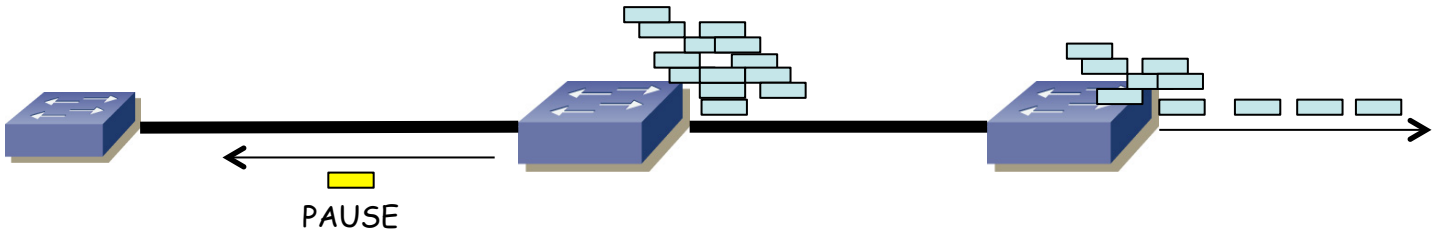
Control de flujo en Ethernet

- Inicialmente no tenía sentido salto a salto pues no había “saltos”
- Con la introducción de puentes/switches, se añade en 802.3x
- El receptor que sufre congestión envía tramas de PAUSE a la fuente
  - Son tramas de control MAC (Ethertype 0x8808)
  - Un campo indica el tiempo de pausa en tiempos de tx de 512 bits
  - Tramas enviadas a MAC multicast reservada (01:80:C2:00:00:01) que no son reenviadas
- ¿Qué sucederá a continuación?
- (…)



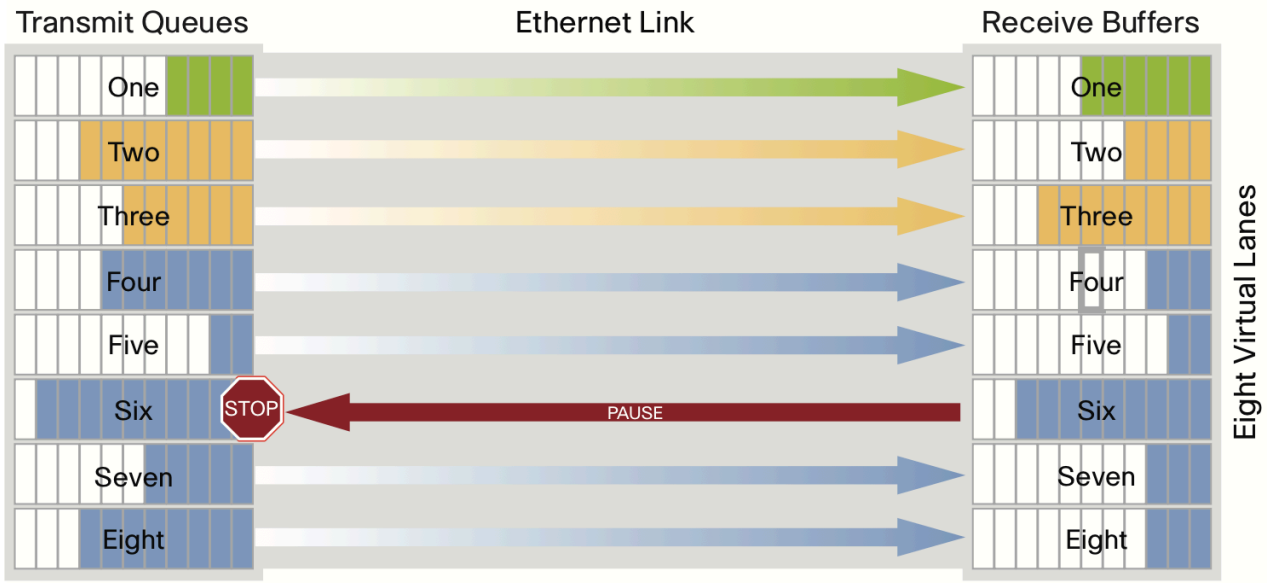
Control de flujo en Ethernet

- ¿Qué sucederá a continuación?
- El conmutador anterior dejará de enviar y con ello probablemente empiece a acumular tramas
- Con ello probablemente envíe una trama de PAUSE al conmutador anterior
- Esto no es un reenvío de la trama anterior
- El mecanismo es para cada salto
- Poco empleado, implementaciones inconsistentes



PFC

- *Priority-based Flow Control*
- IEEE 802.1Qbb (ya recogido en 802.1Q-2012)
- Permite que no haya pérdidas por congestión para los protocolos que así lo requieran
- Se hace control de flujo de forma independiente para cada clase de servicio de 802.1p



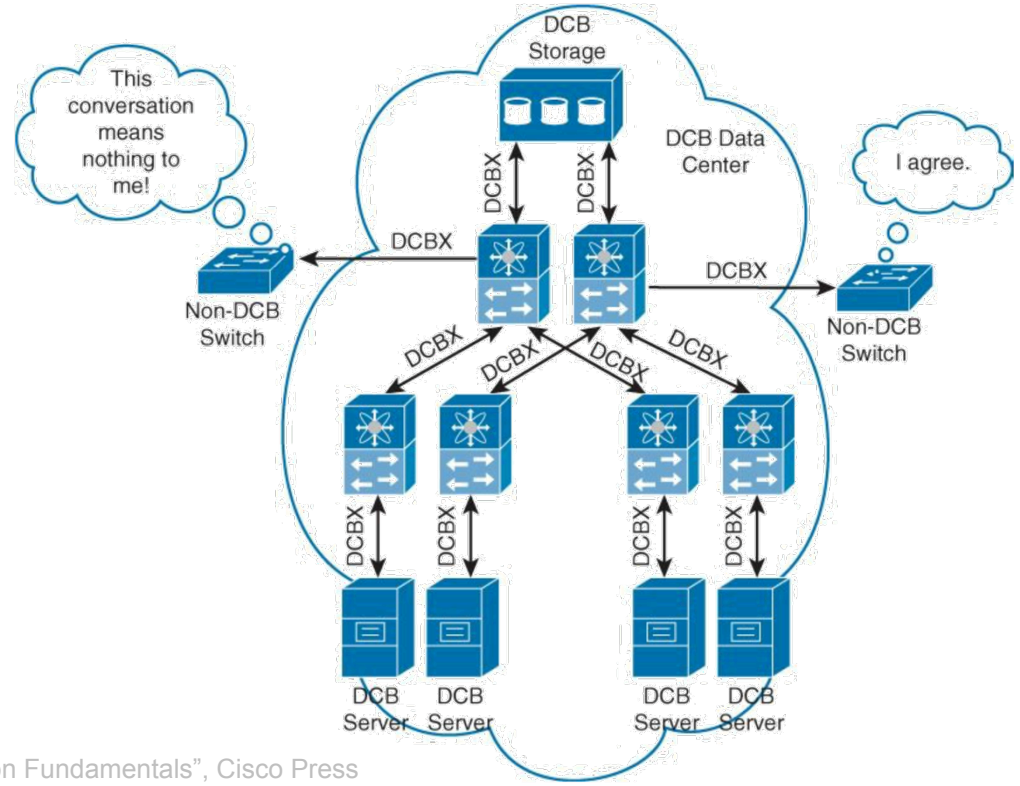
http://www.cisco.com/c/dam/en/us/solutions/collateral/data-center-virtualization/ieee-802-1-data-center-bridging/at_a_glance_c45-460907.pdf

ETS

- *Enhanced Transmission Selection*
- IEEE 802.1Qaz (ya recogido en 802.1Q-2012)
- 802.1Q definía 8 prioridades pero no cómo hacer la planificación entre ellas
- ETS no concreta el planificador a usar pero sí los requerimientos que debe cumplir
- Por defecto deberían soportar un planificador con prioridades estrictas
- También puede soportar lo que se conoce como un “credit-based shaper”, que es un token bucket para cada cola
- La tercera opción es el algoritmo ETS
  - Para cuando no hay tramas en las colas de prioridad o con “credit-based shaper”
  - Dice 802.1Q: “transmission selection is performed based on the allocation of bandwidth to that traffic class. Bandwidth is distributed among ETS traffic classes tat support ETS algoritm such that each traffic class is allocated available bandwidth in proportion to is TCBandwidth”
  - Se menciona WRR como una opción, pero no se especifica más detalle

DCBX

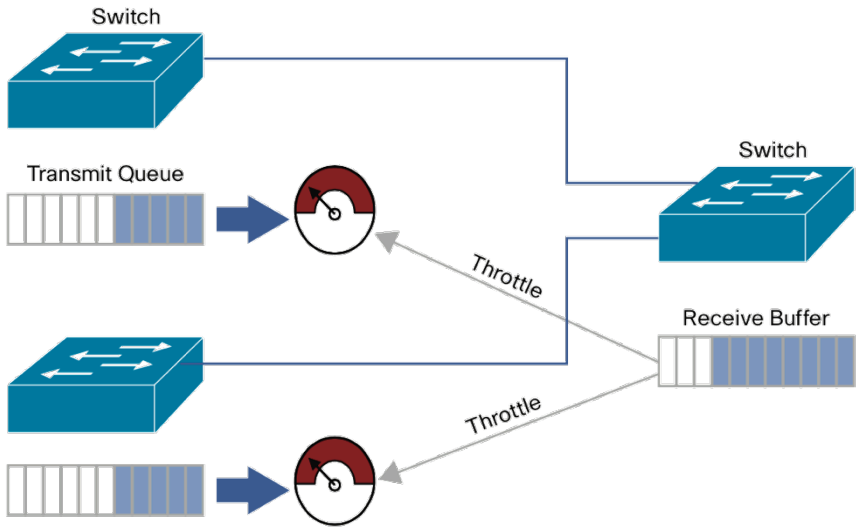
- *Data Center Bridging Exchange Protocol*
- IEEE 802.1Qaz (ya recogido en 802.1Q-2012), extensión de LLDP
- DCBX permite descubrir las capacidades de los extremos de un enlace
- Permite detectar y resolver conflictos en la configuración
- Permite la configuración de parámetros del otro extremo
- Empleado principalmente para parámetros de PFC y ETC



"Data Center Virtualization Fundamentals", Cisco Press

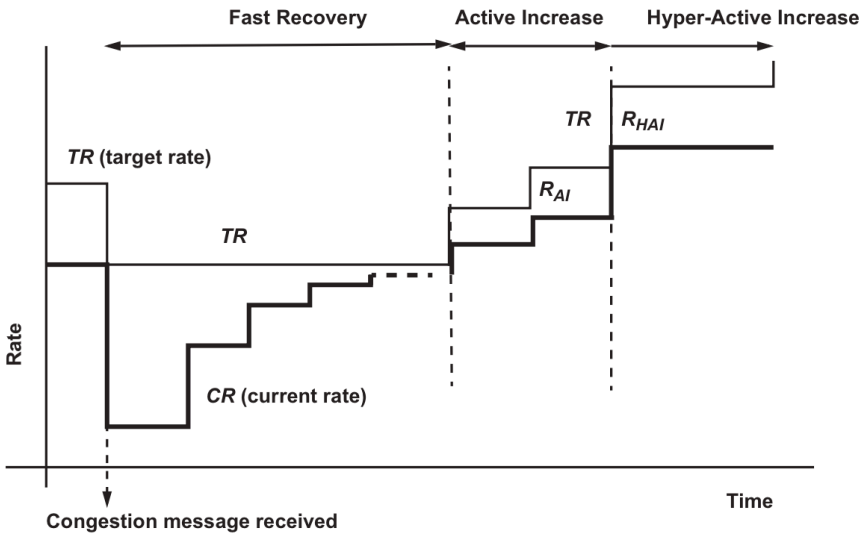
QCN

- *Quantized Congestion Notification* protocol
- IEEE 802.1Qau (ya integrado en 802.1Q-2011)
- Para dominios con BWxRTT < 5 Mbits
- Con enlaces de 10Gbps quiere decir un RTT < 0.5ms
- Es decir, data centers, backplanes, computing clusters, SANs
- Habilita la capacidad en los puentes (y hosts) de enviar señales de congestión a las estaciones finales para que limiten la tasa
- La fuente puede etiquetar las tramas con un CN-Tag
- Se le devuelve en el mensaje de CN
- Le permite identificar el flujo cuya tasa debe reducir



QCN

- Al que detecta la congestión se le llama *Congestion Point (CP)*
- Debe enviar el mensaje de notificación antes de llenar el buffer pues en lo que llega el mensaje seguirá recibiendo paquetes (estimación)
- Lo envía a la dirección origen del paquete con el que toma la decisión
- Quien recibe la notificación se llama el *Reaction Point (RP)*
- Al recibir la notificación el RP reduce la tasa del flujo mediante un *rate limiter*
- No recibe indicación de que pueda aumentar la tasa de nuevo así que lo hace unilateralmente



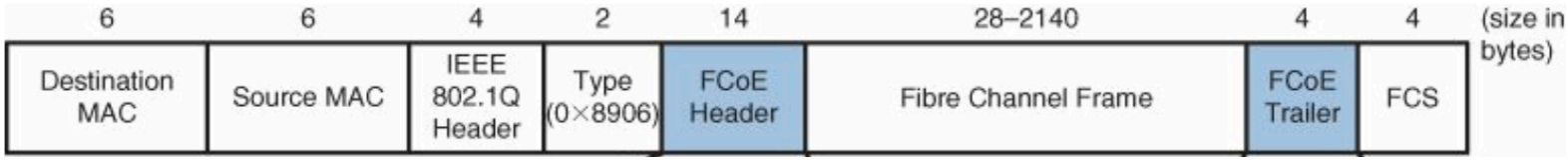
802.1Q

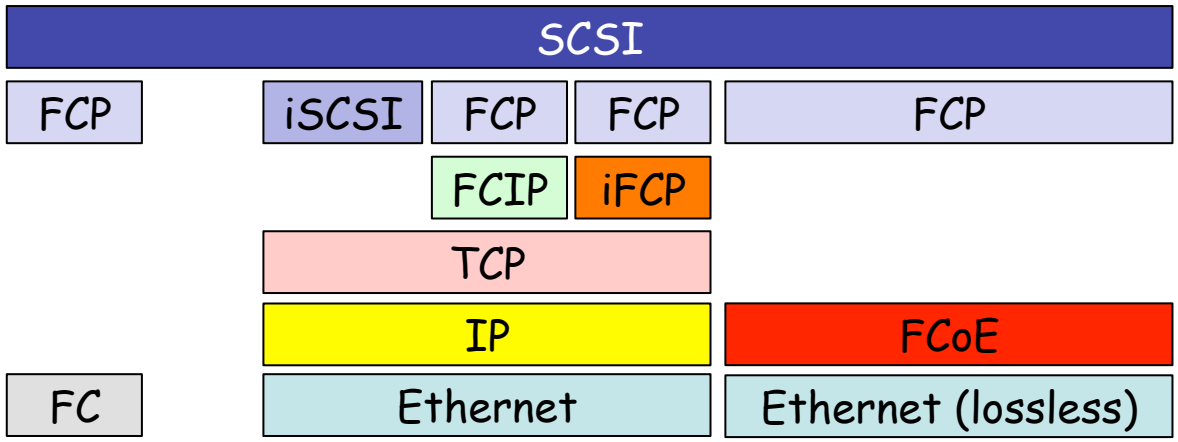
Figure 30-3—QCN RP operation

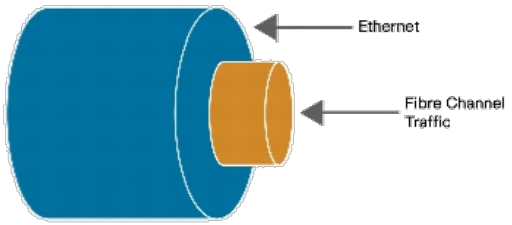
FCoE

FCoE

- *Fibre Channel over Ethernet*
- Mantiene el funcionamiento y la gestión de FC
- Estándar FC-BB-5 del grupo T11 del INCITS
- Tiene sentido a partir de 10GE, cuando está a la par con velocidades en Fibre Channel
- Requiere una Ethernet sin pérdidas (PFC)
- Requiere soporte de *jumbo frames* para transportar en una sola trama Ethernet toda la trama FC (36 bytes cabecera + hasta 2112 de datos)

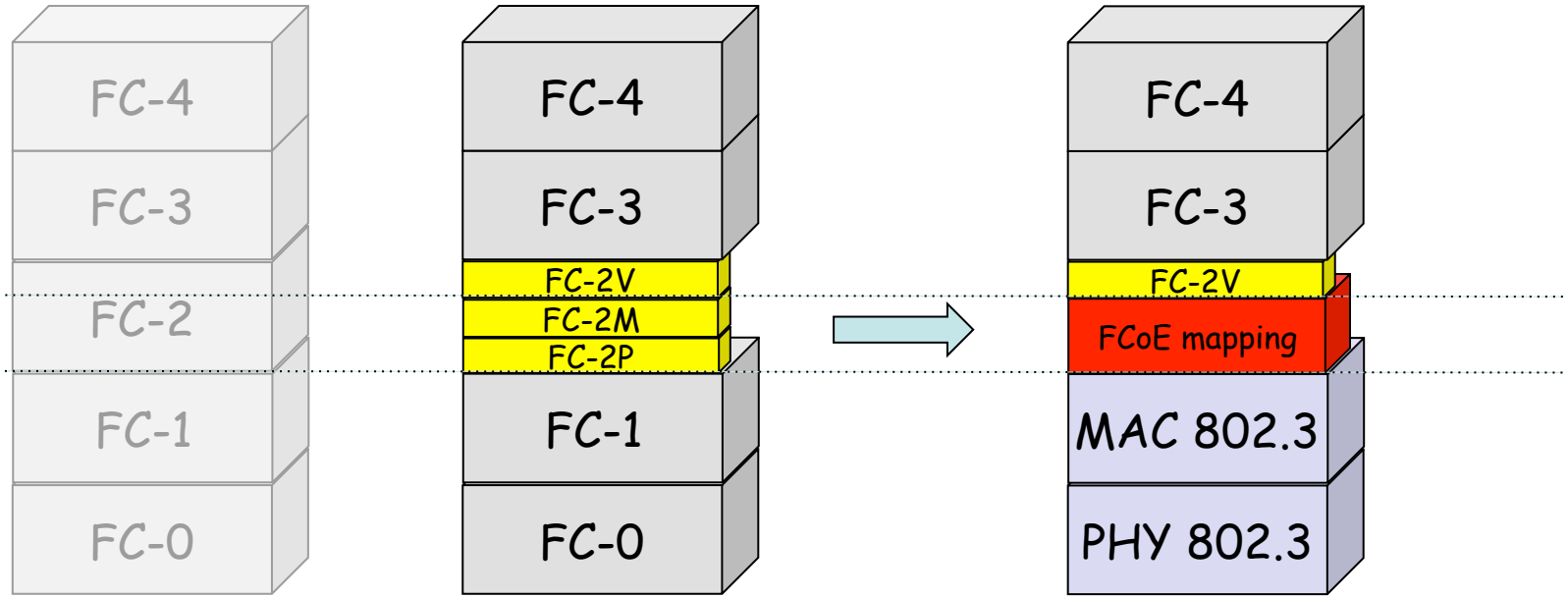






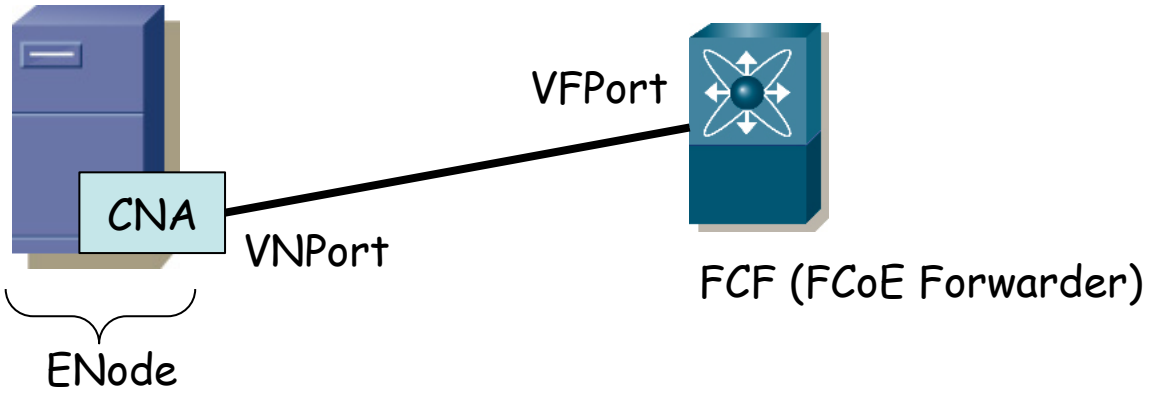
FCoE Mapping

- FC-0 y FC-1 son la capa física, sincronización, errores
- FC-2 se encarga del formato de la trama, señalización y gestión
  - FC-2V define el interfaz con FC-3 y las funciones que se le ofrecen, independientemente del FC-1
  - FC-2M define la multiplexación para cuando hay múltiples FC-2V sobre un FC-2P o viceversa
  - FC-2P implementa el control de flujo, se sustituye por el de PFC
- Mantener FC-2V hace FCoE transparente para el sistema operativo



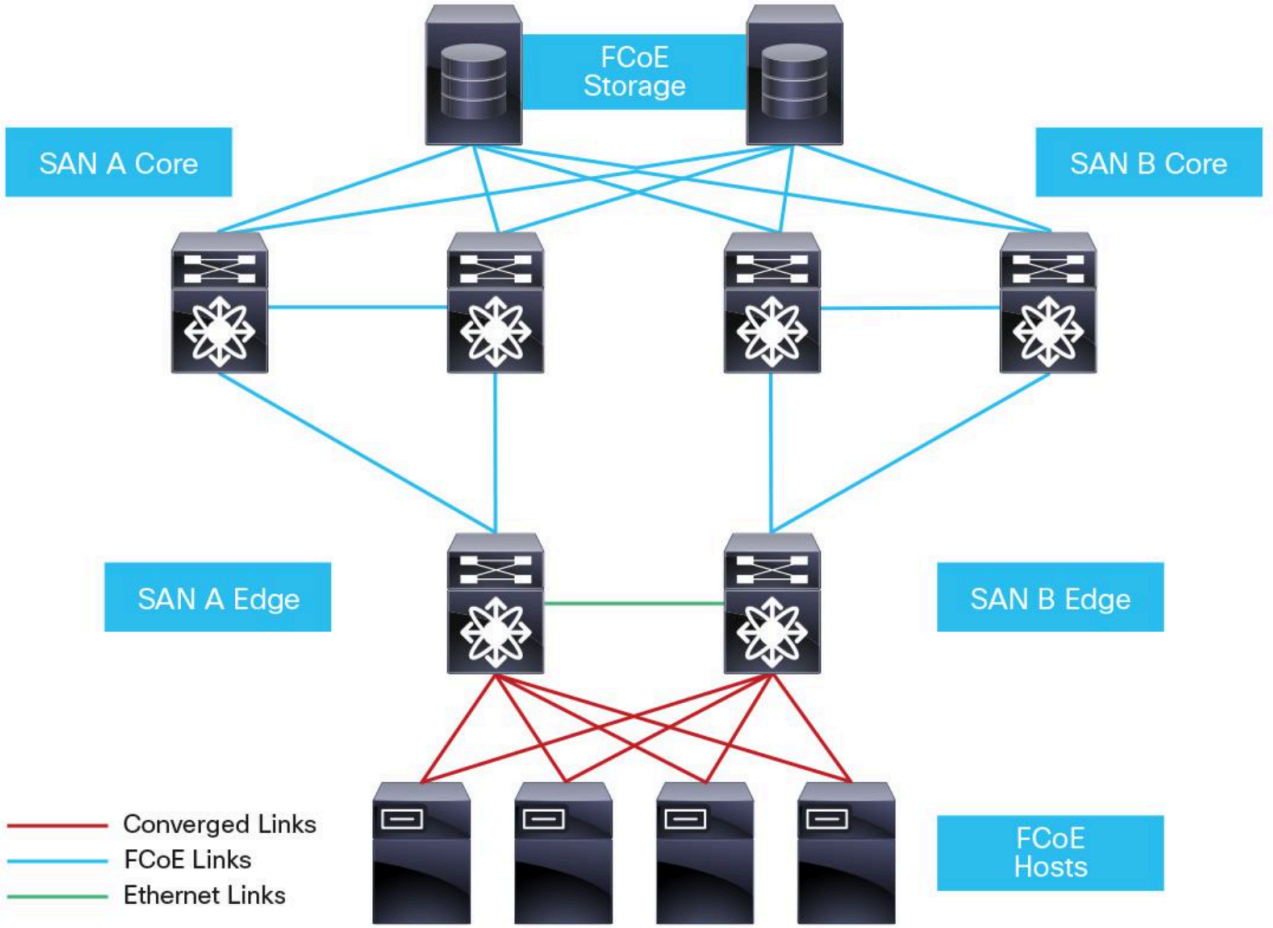
FCoE y FIP

- Un enlace nativo FC es punto-a-punto
- Ethernet es multi-acceso
- FCoE emplea el FCoE Initialization Protocol (FIP) para convertir la Ethernet multi-acceso en un conjunto de enlaces punto-a-punto virtuales
- Transporta también los servicios de descubrimiento y login de Fibre Channel
- FCoE emplea el Ethertype 0x8906 mientras que FIP el 0x8914
- FCoE es el plano de datos, FIP el de control



Ejemplo de topologías

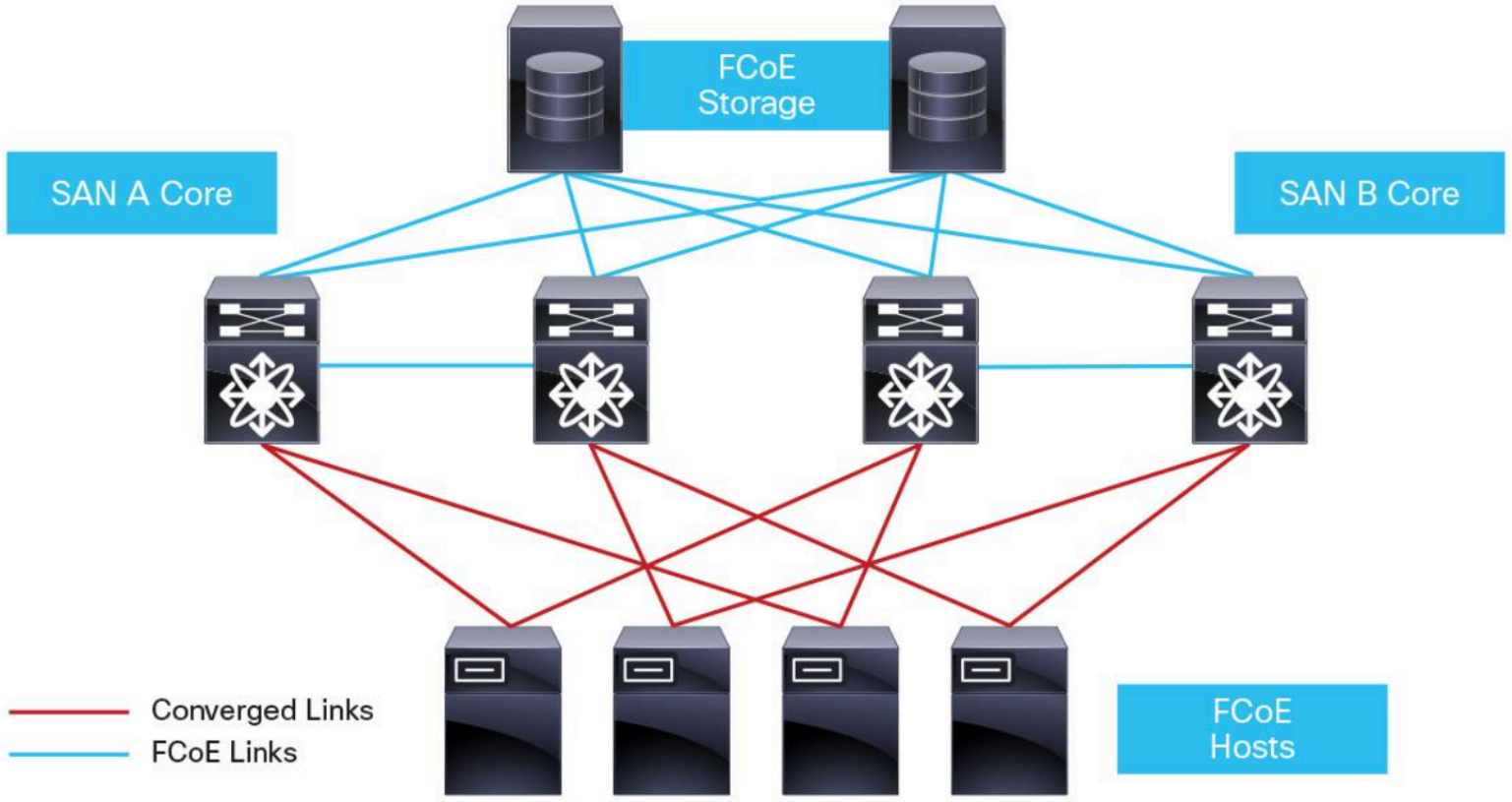
- Core-Edge



http://www.cisco.com/c/en/us/products/collateral/storage-networking/mds-9700-series-multilayer-directors/guide-c07-732733.pdf

Ejemplo de topologías

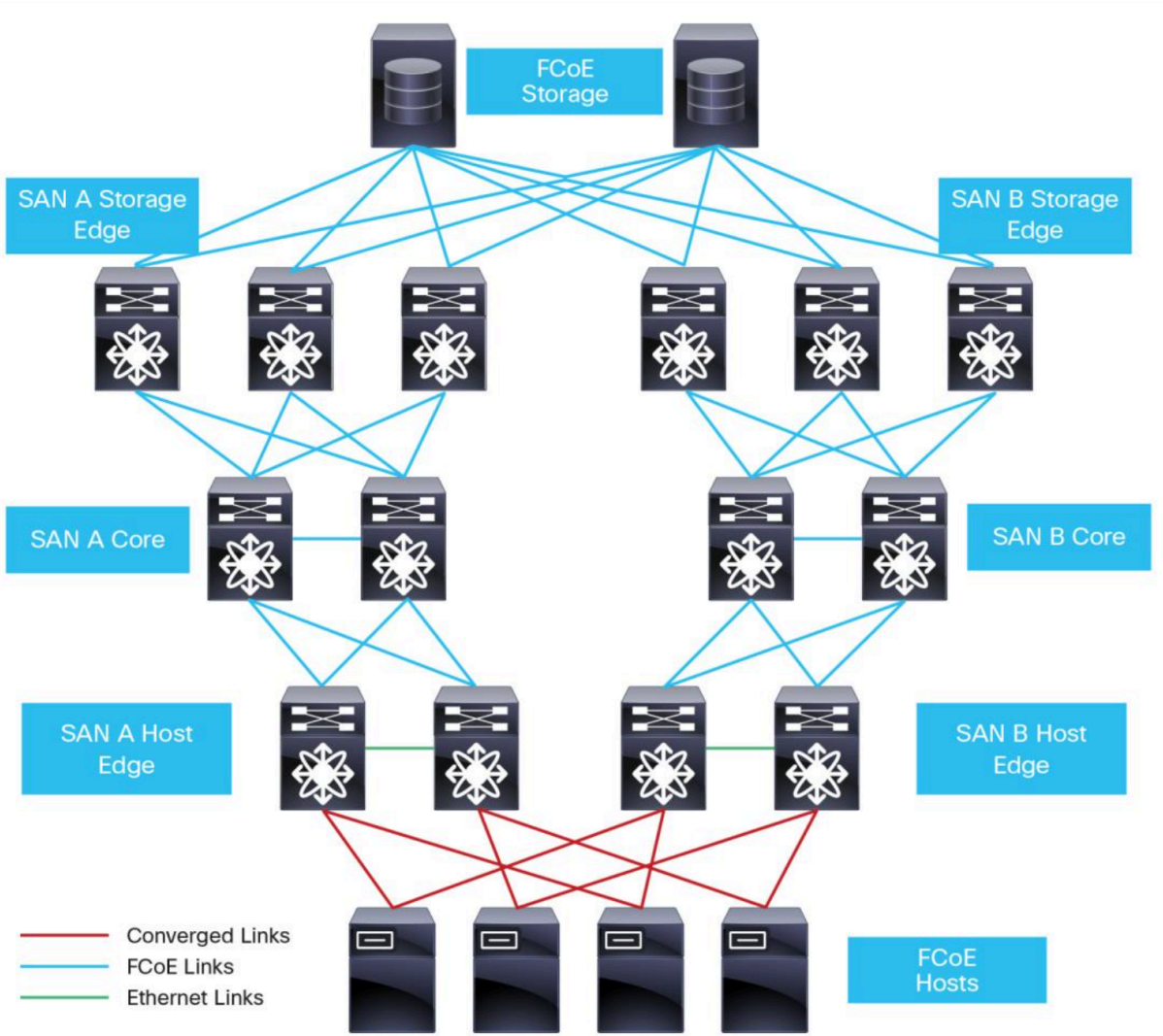
- Collapsed Core



http://www.cisco.com/c/en/us/products/collateral/storage-networking/mds-9700-series-multilayer-directors/guide-c07-732733.pdf

Ejemplo de topologías

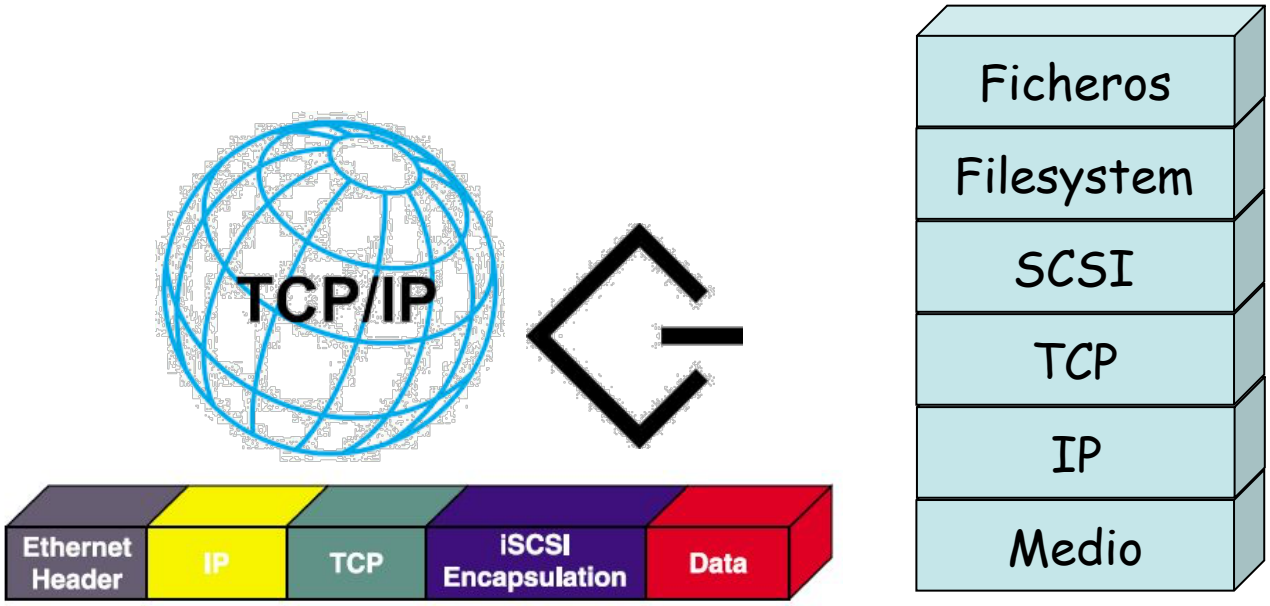
- Edge-Core-Edge



http://www.cisco.com/c/en/us/products/collateral/storage-networking/mds-9700-series-multilayer-directors/guide-c07-732733.pdf

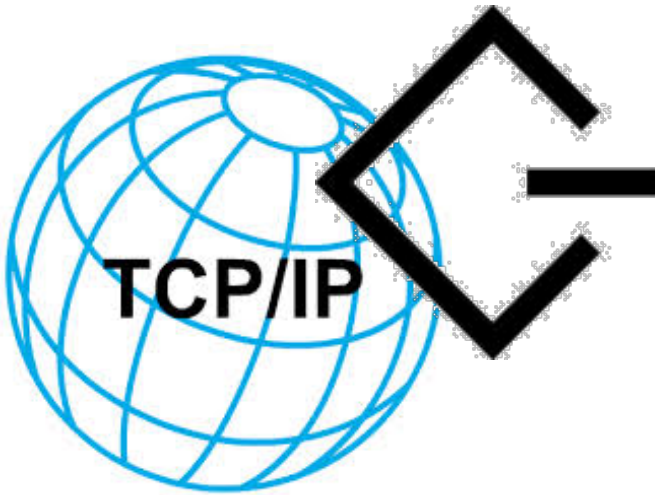
iSCSI

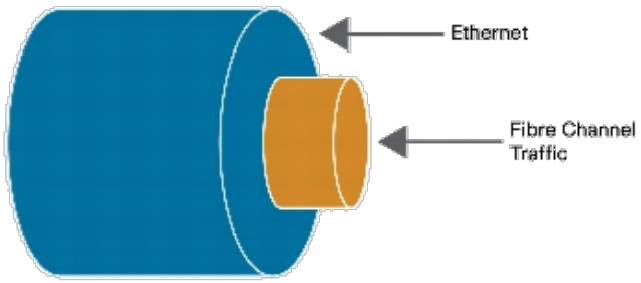
- Última versión RFC 7143 *"Internet Small Computer System Interface (iSCSI) Protocol (Consolidated)"*
- Transporte de comandos SCSI sobre una (o varias) conexiones TCP
- Permite crear una SAN atravesando una red IP
- Es una alternativa de bajo coste a Fibre Channel
- El coste puede estar en el rendimiento (retardo, pérdidas, congestión)
- Si por debajo empleamos Ethernet no ha sido interesante hasta que Ethernet ha alcanzado las velocidades de FC



iSCSI vs FCoE

- Ethernet ya tiene velocidades comparables a FC
- Pero hay otra alternativa que es FCoE
- FCoE permite una integración nativa con equipamiento FC pues transporta sus tramas
- iSCSI no transporta las tramas FC sino comandos SCSI
- Hace falta una pasarela para integrar interfaces iSCSI con FC, con sus problemas de escalabilidad y puntos de fallo
- La gestión cambia entre FC e iSCSI
- iSCSI puede ser más apropiado en un escenario *greenfield*





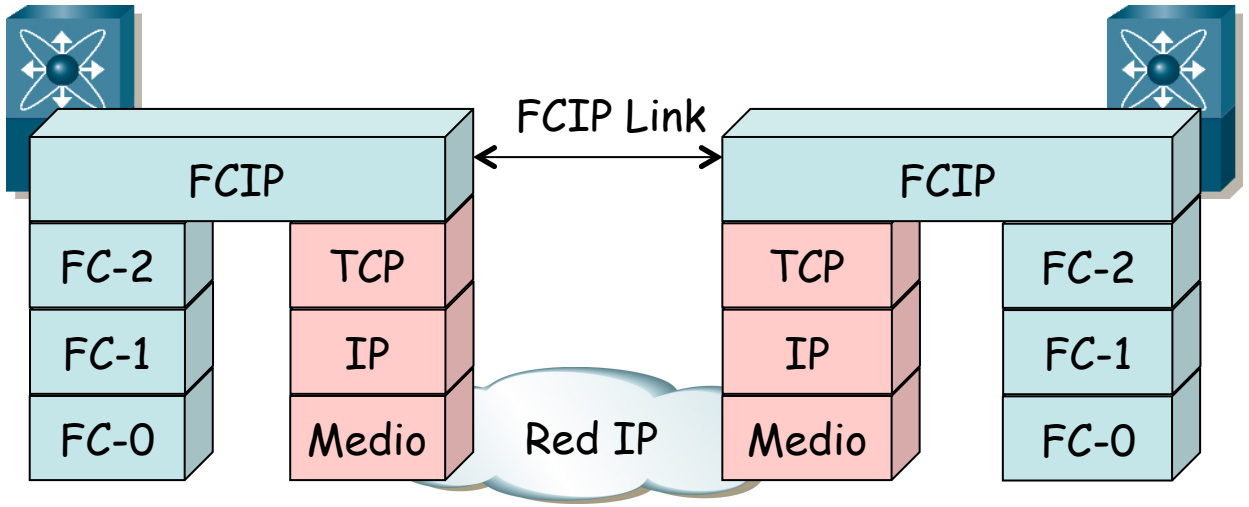
Otras soluciones

FCIP

- RFC 3821 “Fibre Channel Over TCP/IP (FCIP)”
- Es transparente para los equipos FC, solo clases 2, 3 y 4
- Interconexión de islas SAN a través de una red IP; una (o más) conexiones TCP
- Por ejemplo interconexión de SANs en diferentes DCs para replicación de datos
- Los equipos suelen permitir ajustar los parámetros de IP y TCP (timer de retransmisiones, control de flujo, control de congestión, etc)
- También diversos “hacks” para acelerar el protocolo FC

iFCP

- RFC 4172 “iFCP – A Protocol for Internet Fibre Channel Storage Networking”
- Solo clases 2 y 3



FC vs FCoE vs iSCSI

- ¿Retardo en la red por las cabeceras?
  - Despreciable en comparación con los tiempos de acceso al disco
- ¿Gestión y configuración en los hosts?
  - Similar en FC y FCoE
  - Superior en iSCSI pues muchas veces hay que activarlo en el HBA y además hay que configurar IP
- ¿Gestión y configuración en la red?
  - Similar entre FC e iSCSI
  - Superior en FCoE por configurar DCB

Ejemplo de equipos

Cisco Nexus 6001 Switch

The Cisco® Nexus® 6001 Switch is part of a robust fabric that can scale significantly to meet data center growth, delivering architectural flexibility that enables any data center to be a more agile, cost-effective, and efficient environment. The Cisco Nexus 6001 Switch is an important component of the Cisco Unified Data Center architecture, complementing existing Cisco Nexus switches. The Cisco Nexus 6001 is an energy-efficient switch that offers high port density in one rack unit (1RU), supports 10 and 40 Gigabit Ethernet and Fibre Channel over Ethernet (FCoE), provides integrated Layer 2 and 3 features at wire speed, and offers low latency of approximately 1 microsecond for any packet size. With a choice of front-to-back (port side exhaust) and back-to-front (port side intake) airflow options, the switch is designed for a broad range of traditional data center and large-scale virtualized cloud deployments.

The Nexus 6001 Switch runs the industry-leading Cisco NX-OS Software operating system, providing customers with features and capabilities that are widely deployed around the world. Cisco NX-OS Software is a purpose-built data center operating system designed for performance, resiliency, scalability, manageability, and programmability. It meets Ethernet and storage networking requirements, providing a robust and comprehensive feature set that can meet the demanding requirements of virtualization and automation in present and future data centers. The Cisco enhanced fabric solution allows the transparent integration of the virtual and physical devices on a unified network. In addition, the comprehensive Cisco NX-OS service set can be used to create unique innovations for customized solutions.

Cisco Nexus 6001 Switch

Performance

- Cisco Nexus 6001: Layer 2 and 3 hardware forwarding at 1.28 Tbps
- Support for up to 256,000 combined entries of MAC addresses and APR entries
- Low-latency of approximately 1 microsecond using cut-through forwarding for predictable, consistent traffic latency regardless of packet size, traffic pattern, or features enabled on 40 and 10 Gigabit Ethernet interfaces
- 25-MB buffer per 3x 40 Gigabit Ethernet QSFP interfaces
- Line-rate traffic throughput on all ports

Interfaces

- Cisco Nexus 6001P: 48 fixed 1/10 Gigabit Ethernet SFP+ and 4 fixed 40 Gigabit Ethernet QSFP+ports, with 10 and 40 Gigabit Ethernet FCoE support on all respective ports
- Cisco Nexus 6001T: 48 fixed 1/10 Gigabit BASE-T and 4 fixed 40 Gigabit Ethernet QSFP+ ports, with 10 and 40 Gigabit Ethernet FCoE support on all respective ports
- 40 Gigabit Ethernet ports can be converted to 10 Gigabit Ethernet interfaces through QSFP+breakout cable
- Fabric extension through the Cisco Nexus 2200

Layer 2 Features

- Layer 2 switch ports and VLAN trunks
- IEEE 802.1Q VLAN encapsulation
- Support for up to 4000 VLANs
- Support for up to 4000 access control list (ACL) entries
- Rapid Per-VLAN Spanning Tree Plus (PVRST+) (IEEE 802.1w compatible)
- Multiple Spanning Tree Protocol (MSTP) (IEEE 802.1s): 64 instances
- Spanning Tree PortFast
- Spanning Tree root guard
- Spanning Tree Bridge Assurance
- Cisco EtherChannel technology (up to 16 ports per EtherChannel)
- Cisco vPC technology
- vPC configuration synchronization

Cisco Nexus 6001 Switch

- vPC Shutdown
- Link Aggregation Control Protocol (LACP): IEEE 802.3ad
- Advanced PortChannel hashing based on Layer 2, 3, and 4 information
- Jumbo frames on all ports (up to 9216 bytes)
- Pause frames (IEEE 802.3x)
- Storm control (unicast, multicast, and broadcast)
- Private VLANs
- Private VLAN over trunks (isolated and promiscuous)
- Private VLANs over vPC and EtherChannels
- VLAN remapping
- Cisco FabricPath
- EvPC and vPC+ with Cisco FabricPath
- Cisco Adapter FEX
- Cisco Data Center VM-FEX
- Support for up to 24 fabric extenders (Layer 2) with each switch

Layer 3 Features

- Layer 3 interfaces: Routed ports, switch virtual interface (SVI), PortChannels, subinterfaces, and PortChannel subinterfaces
- Support for up to 32,000 IPv4 and 8000 IPv6 host prefixes
- Support for up to 8000 multicast routes (IPv4)
- Support for up to 8000 IGMP snooping groups
- Support for 4000 Virtual Routing and Forwarding (VRF) entries
- Support for up to 4096 VLANs
- Equal-Cost Multipathing (ECMP) up to 64 ways
- 4000 flexible ACL entries
- Routing protocols: Static, Routing Information Protocol Version 2 (RIPv2), Enhanced Interior Gateway Routing Protocol (EIGRP), Open Shortest Path First Version 2 (OSPFv2), Border Gateway Protocol (BGP) and Intermediate System-to-Intermediate System (IS-IS)
- IPv6 routing protocols: Static, OPFv3, BGPv6, and EIGRPv6
- IPv6 VRF-lite
- BFD support: OSPFv2, BGPv4, EIGRP, VRFs
- Policy Based Routing (IPv4 and IPv6)
- Hot-Standby Router Protocol (HSRP) and Virtual Router Redundancy Protocol (VRRP)
- IPdirect Broadcast
- vPC+ Routing Protocol Peering
- ACL: Routed ACL with Layer 3 and 4 options to match ingress and egress ACL
- Multicast: Protocol Independent Multicast Version 2 (PIMv2) sparse mode, Source-Specific Multicast (SSM), Bidir-PIM, Multicast Source Discovery Protocol (MSDP), IGMPv2 and v3, and Multicast VLAN Registration (MVR)
- VRF: VRF-lite (IP VPN); VRF-aware unicast; and BGP-, OSPF-, RIP-, and VRF-aware multicast
- Unicast Reverse Path Forwarding (uRFP) with ACL; strict and loose modes
- Jumbo frame support (up to 9216 bytes)
- Support for up to 24 fabric extenders on each Cisco Nexus 6001

Cisco Nexus 6001 Switch

High-Availability Features

- ISSU for Layer 2
- Hot-swappable field-replaceable power supplies and fan modules
- N+1 and N+N power redundancy
- N:1 fan module redundancy
- N+1 fan module redundancy

Data Center Bridging

- CEE- and IEEE-compliant priority flow control (PFC; per-priority Pause frame support)
- PFC link distance support: 300m
- CEE-compliant Data Center Bridging Exchange (DCBX) Protocol
- CEE- and IEEE-compliant enhanced transmission selection

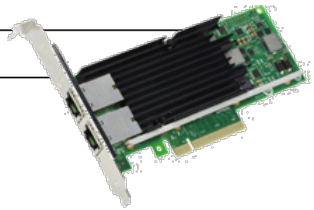
Cisco Nexus 6001 Switch

FCoE Features (Require Storage Services License)

- T11 standards-compliant FCoE (FC-BB-5)
- T11 FCoE Initialization Protocol (FIP) (FC-BB-5)
- Any 10 or 40 Gigabit Ethernet port configurable as FCoE
- SAN administration separate from LAN administration
- Fibre Channel forwarding (FCF)
- Fibre Channel enhanced port types: VE, and VF
- Direct attachment of FCoE targets
- Fabric Device Management Interface (FDMI)
- Fibre Channel ID (FCID) persistence
- Distributed device alias services
- In-order delivery
- Port tracking
- Cisco FCoE_NPV technology
- N-port identifier virtualization (NPIV)
- Fabric services: Name server, registered state change notification (RSCN), login services, and name-server zoning
- Per-VSAN fabric services
- Cisco Fabric Services
- Distributed device alias services
- Host-to-switch and switch-to-switch FC-SP authentication
- Fabric Shortest Path First (FSPF)
- Standard zoning
- Enhanced zoning
- Cisco Fabric Analyzer
- Cisco Data Center Network Manager - SAN
- Storage Management Initiative Specification (SMI-S)
- Boot from SAN over vPC and Enhanced vPC (EvPC)
- FCP
- VSAN trunking
- Fabric Device Management Interface (FDMI)
- Fibre Channel ID (FCID) persistence
- Distributed device alias services
- In-order delivery
- Port tracking
- Cisco NPV technology

Intel Ethernet CNA X540

Features	Benefits
Intel® Ethernet Converged Network Adapter X540	▪ Industry's first integrated MAC+PHY reducing cost and power
Low-profile	▪ Enables higher bandwidth and throughput from standard and low-profile PCIe slots and servers
Load balancing on multiple CPUs	▪ Increases performance on multi-processor systems by efficiently balancing network loads across CPU cores when used with Receive-Side Scaling from Microsoft* or Scalable I/O on Linux*
Remote boot iSCSI and FCoE	▪ Support for remote booting to an iSCSI or FCoE drive.
Support for most Network Operating Systems (NOS)	▪ Enables widespread deployment
RoHS compliant, lead-free technology	▪ Compliant with the European Union directive (July 2006) to reduce hazardous materials
I/O Features for Multi-core Processor Servers	
MSI-X support	▪ Minimizes the overhead of interrupts ▪ Allows load balancing of interrupt handling between different cores/CPUs
Low latency	▪ Based on the sensitivity of the incoming data, the adapter can bypass the automatic moderation of time intervals between the interrupts
Header Splits and Replication in Receive	▪ Helps the driver focus on the relevant part of the packet without the need to parse it
Multiple Queues	▪ Packet handling without waiting/buffer overflow provides efficient packet prioritization
Tx/Rx IP, SCTP, TCP, and UDP checksum offloading (IPv4, IPv6) capabilities	▪ Lower processor usage ▪ Checksum and segmentation capability extended to new standard packet type
Tx TCP segmentation offload (IPv4, IPv6)	▪ Increased throughput and lower processor usage ▪ Compatible with large-send offload feature (in Microsoft Windows* Server operating systems)
IPsec Offload	▪ Offloads IPsec capability to adapter instead of software to significantly improve through-put and CPU usage (for Windows* 7, Windows* 2008 Server R2, Windows* 2008 Server, and Windows Vista*)
Compatible with x8 and x16 standard and low-profile PCI Express* slots	▪ Enables each PCI Express* slot port to operate without interfering or competing with other ports
Receive/Transmit Side Scaling for Windows* and Scalable I/O for Linux* (IPv4, IPv6, TCP/ UDP)	▪ Enables direction of the interrupts to the processor cores in order to improve the CPU usage rate
RJ45 connections over CAT-6a cabling	▪ Ensures compatibility with cable lengths up to 100 meters

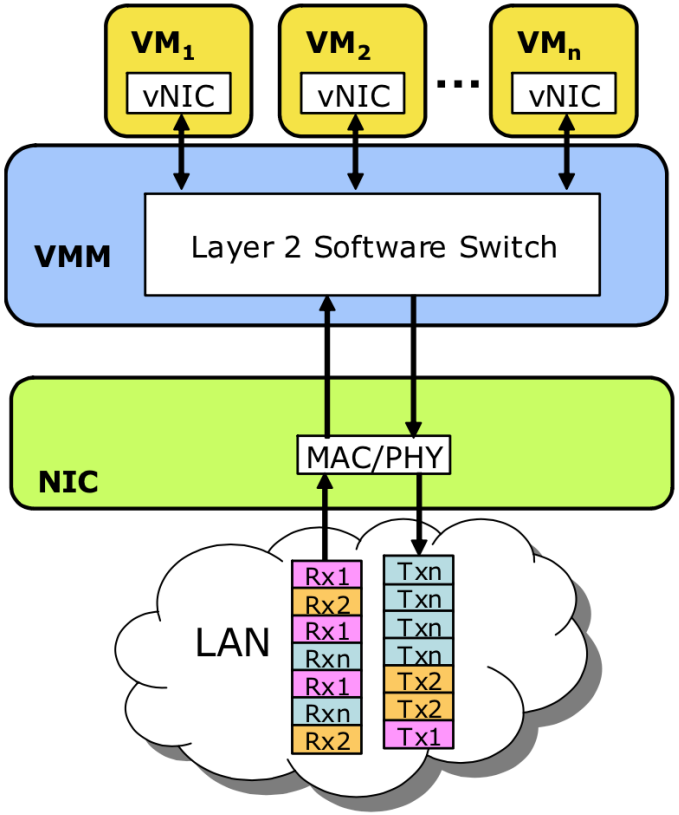


Intel Ethernet CNA X540

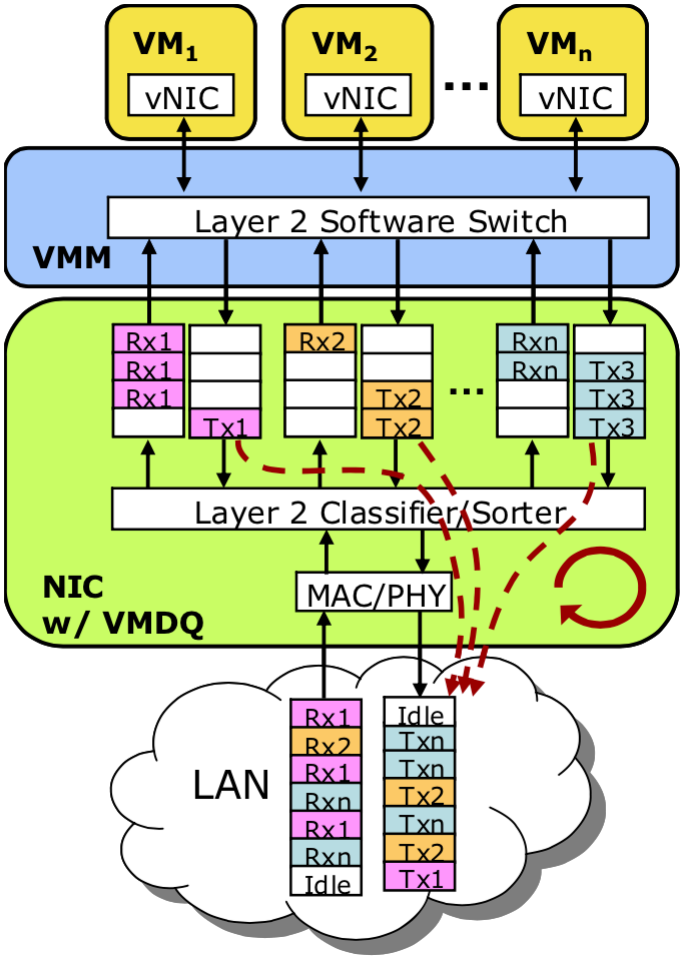
Virtualization Features

Virtual Machine Device queues (VMDq)	▪ Offloads data-sorting from the Hypervisor to silicon, improving data throughput and CPU usage ▪ QoS feature for Tx data by providing round-robin servicing and preventing head-of-line blocking ▪ Sorting based on MAC addresses and VLAN tags
Next-Generation VMDq (64 queues per port)	▪ Enhanced QoS feature by providing weighted round-robin servicing for the Tx data ▪ Provides loopback functionality, data transfer between the virtual machines within the same physical server don't go out to the wire and back in, improving throughput and CPU usage ▪ Supports replication of multicast and broadcast data
PCI-SIG SR-IOV Implementation (64 virtual functions per port)	▪ Implementation of I/O Virtualization. The physical configuration of each port is divided into multiple virtual ports. Each virtual port is assigned to an individual virtual machine directly by by-passing the virtual switch in the Hypervisor, resulting in near-native performance ▪ Iintegrated with Intel® Virtualization Technology for Directed I/O (Intel® VT-d) to provide data protection between virtual machines by assigning separate physical addresses in the memory to each virtual machine
IPv6 Offloading	▪ Checksum and segmentation capability extended to the new standard packet type
Advanced Packet Filtering	▪ 24 exact-matched addresses (unicast or multicast) ▪ 4096-bit hash filter for unicast and multicast frames ▪ Lower processor usage ▪ Promiscuous (unicast and multicast) transfer mode support ▪ Optional filtering of invalid frames
VLAN support with VLAN tag insertion, stripping and packet filtering for up to 4096 VLAN tags	▪ Ability to create multiple VLAN segments

Intel Virtual Machine Device Queues



VMDq



http://www.intel.com/content/www/us/en/virtualization/vmdq-technology-paper.html

Intel Ethernet CNA X540

General

Product codes	Dual Port: X540T2 Single Port: X540T1
Connector	RJ45 Copper
Cabling	See listing below

Adapter Product Features

Intel® PROSet Utility for easy configuration and management	
Intel® Lead-free technology	
Plug and play specification support	Standard
Full-height bracket installed; low-profile bracket included in package	
RoHS compliant	
Cabling Distances:	
10GBASE-T	▪ 100 m on Cat 6A ▪ 55 m on Cat 6
1000BASE-T	100 m on Cat 5e, Cat 6 or Cat 6A
100BASE-T	100 m on Cat 5e, Cat 6 or Cat 6A
Receive-Side Scaling	

Power Usage

Speed/LFM	Watts (typical)	
	X540T2	X540T1
10 Gbps	17.4	10.8
1 Gbps	9.5	7.7
100 Mbps	6.6	6.4
LFM	200	0

Advanced Software Features

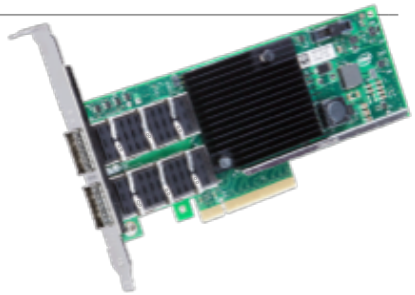
Adapter Fault Tolerance (AFT) Switch Fault Tolerance (SFT) Adaptive Load Balancing (ALB) Virtual Machine Load Balancing (VMLB)	Teaming support
IEEE 802.3ad (link aggregation control protocol)	
PCIe Hot Plug/Active Peripheral	
Component Interconnect (PCI)	
IEEE 802.1Q VLANs	
IEEE 802.3 2005 flow control support	
Tx/Rx IP, TCP, & UDP checksum offloading (IPv4, IPv6) capabilities Transmission control protocol (TCP), User Datagram Protocol (UDP), Internet Protocol (IP))	
IEEE 802.1p	
TCP segmentation/large send offload	
MSI -X: Multiple Independent Queues	
Interrupt moderation	
IPv6 offloading	

Hardware Features

Data rate(s) supported per port	100 Mbps, 1 Gbps, 10 Gbps
Bus Type	PCI Express 2.1 (5.0 GT/s)
Bus width	x8 lane PCI Express, operable in x8 and x16 slots
Bus speed (x8, encoded rate)	20 Gbps uni-directional; 40 Gbps bi-directional
Interrupt levels	INTA, MSI, MSI-X
Hardware certifications	FCC A, UL, CE, VCCI, BSMI, CTICK, KCC
Controller-processor	Intel® Ethernet Controller X540

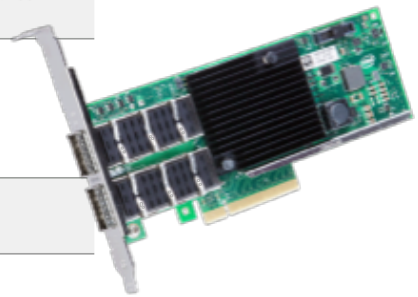
Intel Ethernet CNA XL710 10/40 GbE

GENERAL	
FEATURES	**BENEFITS**
Intel® XL710 10/40 Gigabit Ethernet Controller	• Industry-leading, energy-efficient design for next-generation 10/40 Gigabit performance and multi-core processors
QSFP+ Connectivity	• XL710 adapters with QSFP+ connections support 40GBASE-SR4, 40GBASE-LR4[1] and QSFP+ CR4 Copper Direct Attach physical media.
Low-profile	• Enables higher bandwidth and throughput from standard and low-profile PCIe slots and servers
Load balancing on multiple CPUs	• Increases performance on multi-processor systems by efficiently balancing network loads across CPU cores when used with Receive-Side Scaling (RSS) from Microsoft or Scalable I/O on Linux*
iSCSI remote boot support	• Provides centralized storage area network (SAN) management at a lower cost than other iSCSI solutions • No additional cost for iSCSI support, included in standard adapter
Fibre Channel over Ethernet (FCoE) Support[1]	• Includes FCoE Boot and Data Center Bridging • No additional cost for FCoE support, included in standard adapter
Support for most network operating systems	• Enables widespread deployment
RoHS-compliant	• Complies with the European Union directive 2011/65/EU to reduce the use of hazardous materials
Intel® PROSet Utility for Windows* Device Manager	• Provides point-and-click management of individual adapters, advanced adapter features, connection teaming, and virtual local area network (VLAN) configuration
Time Sync (IEEE 1588*, 802.1as)	• Enables networked Ethernet equipment to synchronize internal clocks according to a network master clock; endpoint can then acquire an accurate estimate of the master time by compensating for link latency



Intel Ethernet CNA XL710 10/40 GbE

I/O FEATURES FOR MULTI-CORE PROCESSOR SERVERS	
FEATURES	**BENEFITS**
Intel® Flow Director	• An advanced traffic steering capability increases the number of transactions per second and reduces latency for cloud applications like Memcached
MSI-X support	• Minimizes the overhead of interrupts • Load-balancing of interrupt handling between multiple cores/CPUs
Multiple Queues: 1,536 Tx and Rx queues per port	• Network packet handling without waiting for buffer overflow providing efficient packet prioritization • Actual number of queues will vary depending upon software implementation
Tx/Rx IP, SCTP, TCP, and UDP checksum offloading (IPv4, IPv6) capabilities	• Lower processor usage • Checksum and segmentation capability extended to new standard packet type
VIRTUALIZATION FEATURES	
FEATURES	**BENEFITS**
Next-Generation VMDq	• Up to 256 maximum VMDq VMs supported • Enhanced QoS feature by providing weighted round-robin servicing for the Tx data • Offloads the data-sorting functionality from the Hypervisor to the network silicon, improving data throughput and CPU usage • Provides QoS feature on the Tx data by providing round-robin servicing and preventing head-of-line blocking • Sorting based on MAC addresses and VLAN tags • Provides loopback functionality, where data transfer between the virtual machines within the same physical server need not go out to the wire and come back in, improving throughput and CPU usage
PC-SIG SR-IOV Implementation (128 per device)	• Provides an implementation of the PCI-SIG standard for I/O Virtualization. The physical configuration of each port is divided into multiple virtual ports. Each virtual port is assigned to an individual virtual machine directly by bypassing the virtual switch in the Hypervisor, resulting in near-native performance. • Integrated with Intel® VTI for Directed I/O (VT-d) to provide data protection between virtual machines by assigning separate physical addresses in the memory to each virtual machine • 64/port for dual port • 32/port for quad port
Virtual Machine Load Balancing (VLMB)	• Virtual Machines Load Balancing (VMLB) provides traffic load balancing (Tx and Rx) across Virtual Machines bound to the team interface, as well as fault tolerance in the event of switch, port, cable, or adapter failure
Advanced Packet Filtering	• 1536 exact matched packets (unicast or multicast) • 512 hash entries each for unicast and multicast • Lower processor usage • Promiscuous (unicast and multicast) transfer mode support • Optional filtering of invalid frames
VLAN support with VLAN tag insertion, stripping, and packet filtering for up to 4096 VLAN tags	• Ability to create multiple VLAN segments
VXLAN and NVGRE Support	• Preserves application performance in network virtualized environments



Intel Ethernet CNA XL710 10/40 GbE

MANAGEABILITY FEATURES	
FEATURES	**BENEFITS**
Preboot eXecution Environment (PXE) Support	• Enables system boot up via the LAN (32-bit and 64-bit) • Flash interface for PXE image
Simple Network Management Protocol (SNMP) and Remote Network Monitoring (RMON) Statistic Counters	• Easy system monitoring with industry-standard consoles
iSCSI Boot[1]	• Enables system boot up via iSCSI • Provides additional network management capability
Watchdog Timer	• Gives an indication to the manageability firmware or external devices that the chip or the driver is not functioning

SPECIFICATIONS	
GENERAL	
Connections	Single or Dual QSFP+ cages for: • QSFP+ SR4 fiber-optic transceivers • QSFP+ LR4 fiber-optic transceivers • QSFP+ CR4 Direct Attach Cables
Intel® Ethernet QSFP+ Optics required for a fiber configuration	
Network Standards Physical Layer Interface	IEEE 802.3: • 40GBASE-SR4 • 40GBASE-LR4 SFF-8431: • 40GSFP+ CR4 (Direct Attach Copper) • 40GSFP+ CR4 to 4x SFP+ (Breakout Cable)

ADVANCED SOFTWARE FEATURES – ALL ADAPTERS
Adapter fault tolerance (AFT)
Switch fault tolerance (SFT)
Adaptive load balancing (ALB)
Teaming Support
IEEE 802.3ad (link aggregation control protocol)
PCIe Hot Plug*/Active peripheral component interconnect (PCI)
IEEE 802.1Q* VLANs
IEEE 802.3 2005* flow control support
Tx/Rx IP, TCP, & UDP checksum offloading (IPv4, IPv6) capabilities (Transmission control protocol (TCP), user datagram protocol (UDP), Internet protocol (IP))
IEEE 802.1p*
TCP segmentation/large send offload
MSI-X supports Multiple Independent Queues
Interrupt moderation
IPv6 offloading—Checksum and segmentation capability extended to