



Hálózati megoldások Szerverkonszolidációs környezetben

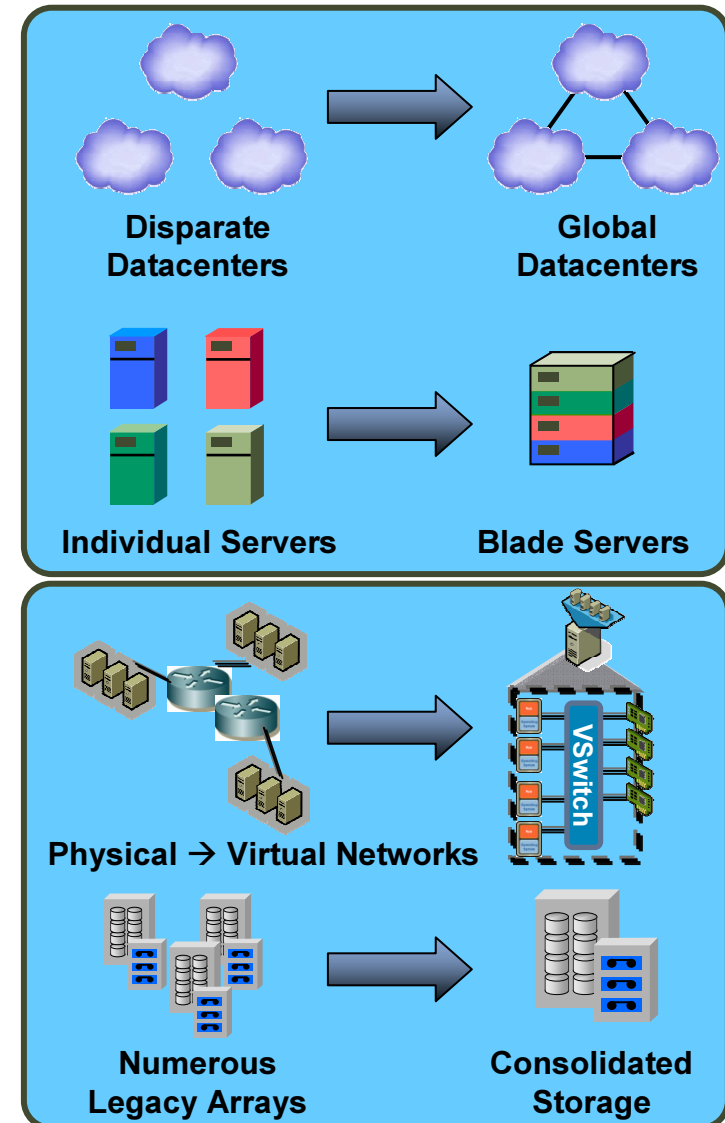


HBOne2008

Zeisel Tamás
Cisco Magyarország
tzeisel@cisco.com

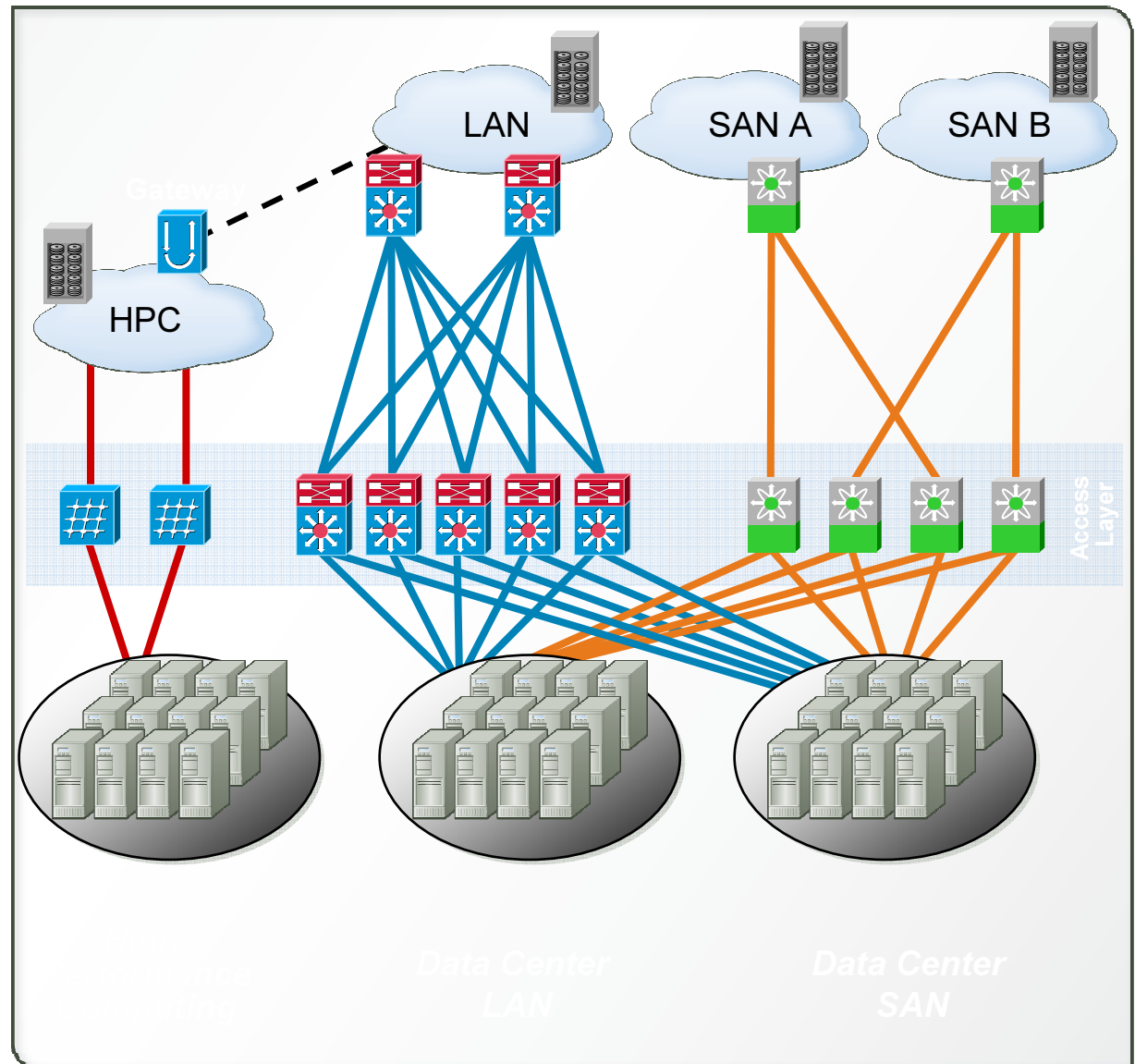
Infrastruktúra trendek

- Adatközpont konszolidáció
 - Fizikailag egy helyszín
 - Adatközpontok összekapcsolása
- Szerver konszolidáció
 - Nagyobb teljesítményű szerverek
 - Blade szerver technologia
 - Szerver virtualizáció (Vmware)**
- Hálózat konszolidáció
 - 10Gb aggregáció
 - Access hálózati réteg virtualizáció
 - Unified Fabric**
- Storage konszolidáció
 - Nagyobb társűrűség
 - Nagyobb aggregált performancia

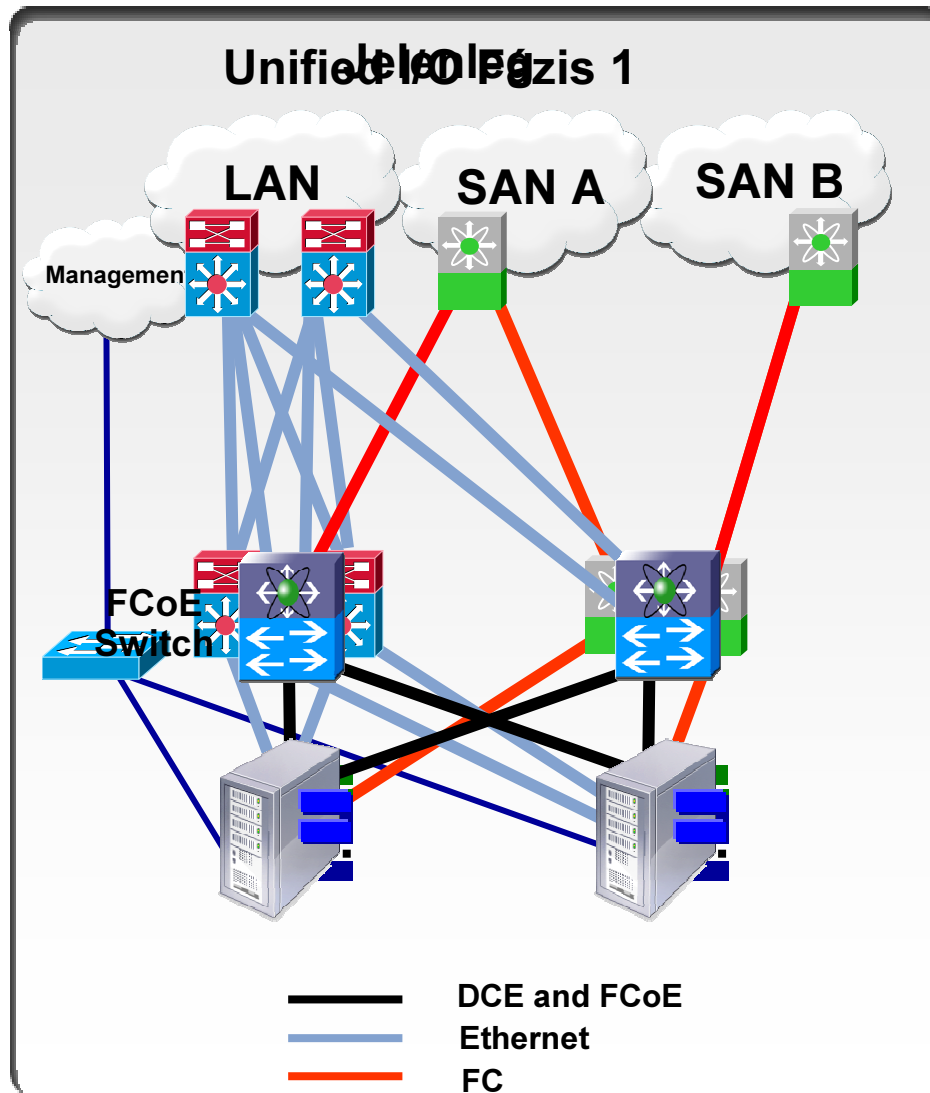


Tipikus Adatközpont/szerver konszolidáció

- Több független hálózat
- Szerver konszolidáció igénye nő
- Egyre több szerver
- Egyre több Virtuális szerver
- Növekvő teljesítmény felvétel és hűtési igény
- Hely szükséglet!!!!



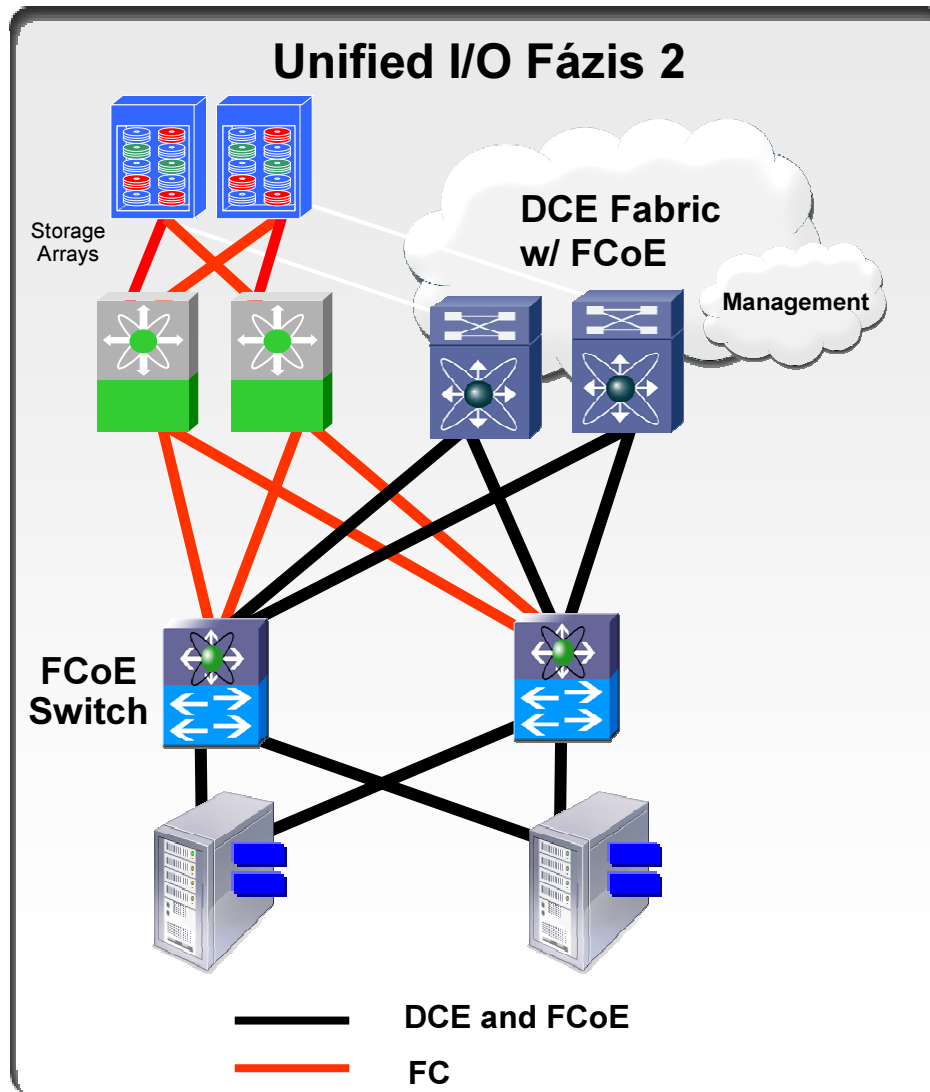
I/O Konzolidáció



Jelenlegi Unified I/O Fázis 1

- Párhuzamos LAN/SAN Infrastruktúra
- Szerver adapterek csökkenése
- Gazdaságtalan Hálózati Infrastruktúra
- Az access réteg és a kábelezés egyszerűsítése
- Gateway mentes implementálás
- 5+ kapcsolati szerverenként, magasabb adapter, és kábelezési költség
- Port ár növekedés
- L2 Multipathing támogatás
- Több Hibaforrás
- Beruházásvedelem (LANs és SANs)
- Folyamatos zökkenés
- Több hibalehetőség – bonyolultabb hibakeresés
- Menedzsment – firmware, driver-patch komplikáltabb

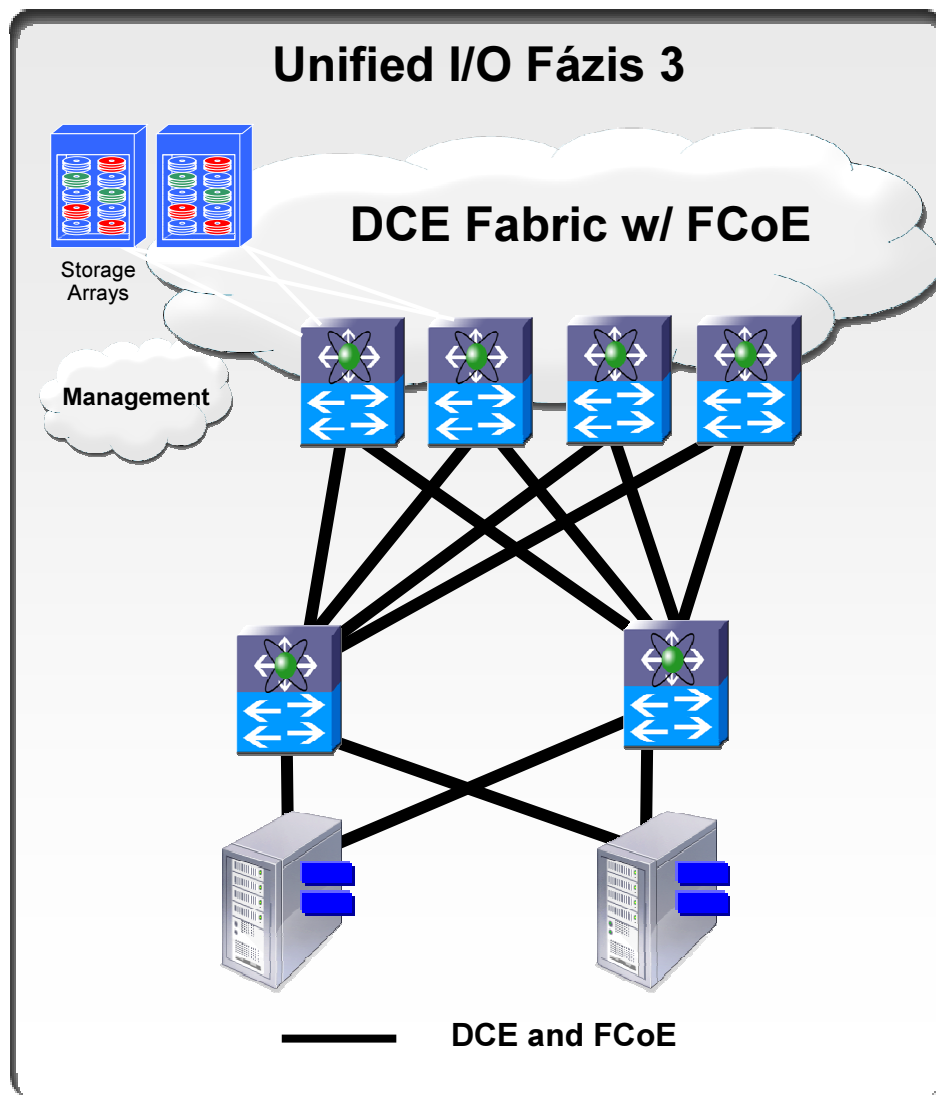
Unified Fabric



Unified I/O Fázis 2

- Párhuzamos hálózati infrastruktúra megszüntetése
- L2/L3 Multipathing
- Gyorsabb infrastruktúra változtatás
- Disk-ek DCE vagy FC protokollon is elérhetőek

Unified Fabric



Unified I/O Fázis 3

- LAN/SAN Unified Fabric a teljes Adatközpontra
- L2/L3 Multipathing
- Konzisztens hálózati szabályok a teljes adatközpontban

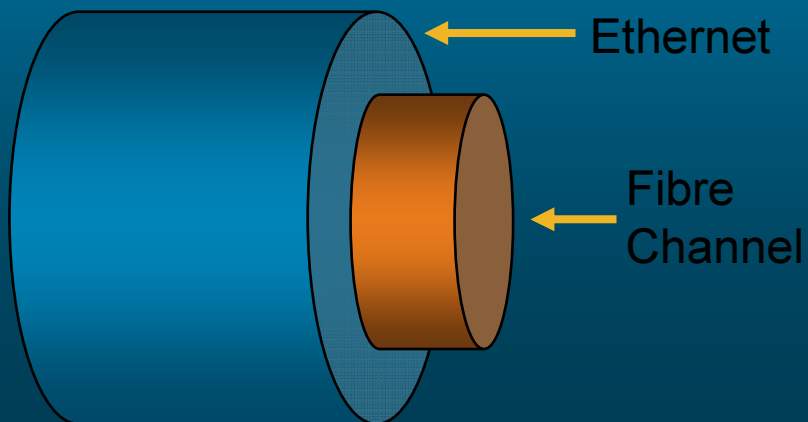
FC over Ethernet (FCoE)



FCoE

FC csomagokat Ethernet fölött visszük át

- FC átviteléhez egy Veszteségmentes Data Center Ethernet szükséges



Előnyök

Egyetlen közös un. Unified Fabric hálózat (LAN+SAN)

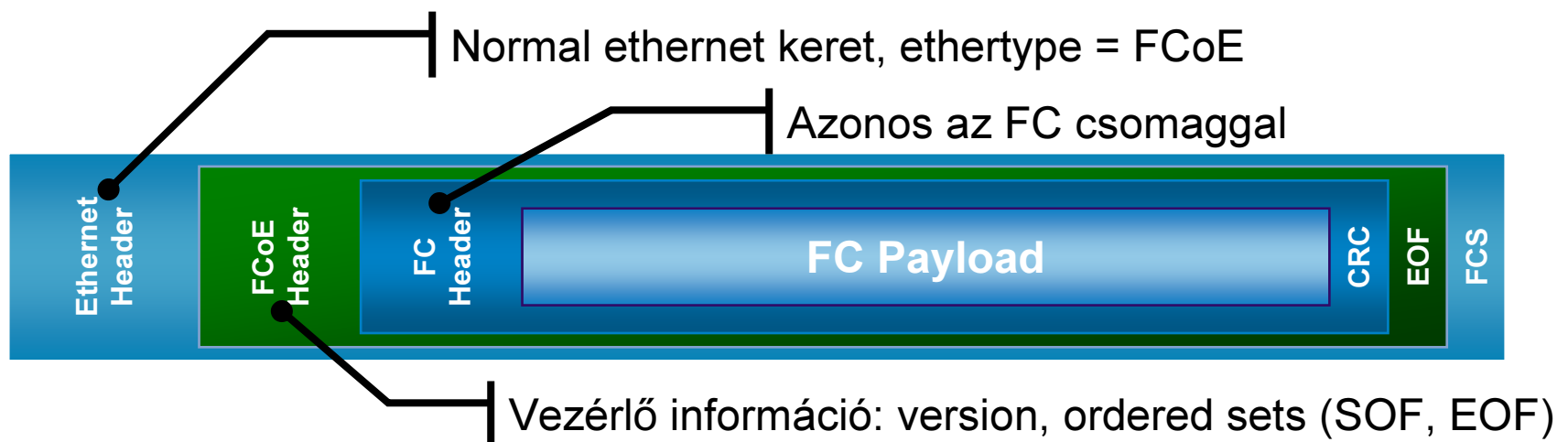
- Egyszerűbb menedzsment
- Kisebb hálózati fogyasztás
- Hatékonyabb helykihasználás
- Egyszerűsített kábelezés

■ Szerverek

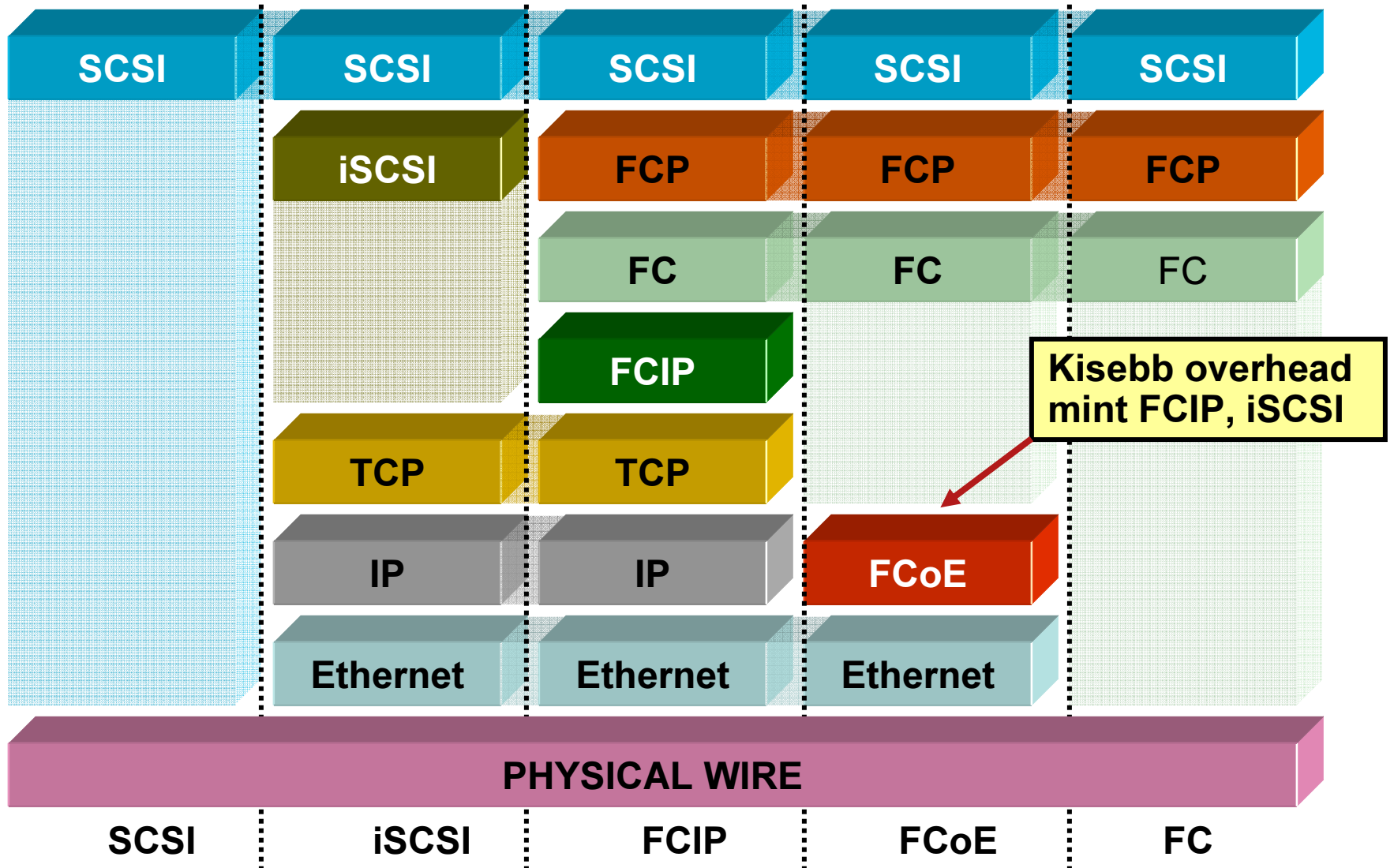
- Kisebb interface beruházás, kevesebb fogyasztás
- Egyszerűbb kábelezés

FCoE Jellemzők

- 10Gbps Ethernet
- Veszteségmentes Ethernet (DCE)
 - FC B2B credit menedzsment megvalósítása
- Ethernet jumbo frames
 - Max FC frame payload = 2112 bytes



Hálózati protokol stack



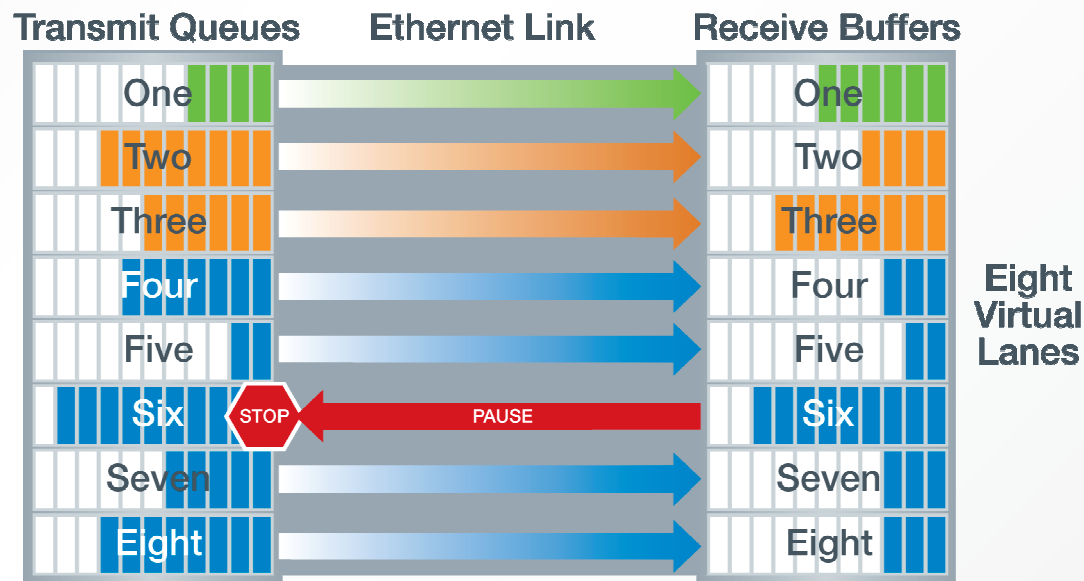
Data Center Ethernet jellemzők



Priority Flow Control	IEEE 802.1Qbb –biztosítja a class of service alapú flow controlt PAUSE mechanizmus révén IEEE 802.1p osztályokon
Sávszélesség Management	IEEE 802.1Qaz Enhanced Transmission Selection - sávszélesség és prioritás allokálás IEEE 802.1p alapú forgalom osztályokhoz
Torlódás Menedzsment	IEEE 802.1Qau szabványos torlódás menedzsment a hálózaton (BCN/QCN)
Data Center Bridging Exchange	Veszteségmentes (DCE) Ethernet Auto-negotiation (switch - NIC között DCBX üzenetek)
L2 Multipathing	Layer-2 multipathing biztosít loadbalnacig lehetőséget STP nélkül, aktív standby link lehetőséggel: Ethernet Host Virtualizer Mode.

Data Center Ethernet Jellemzők

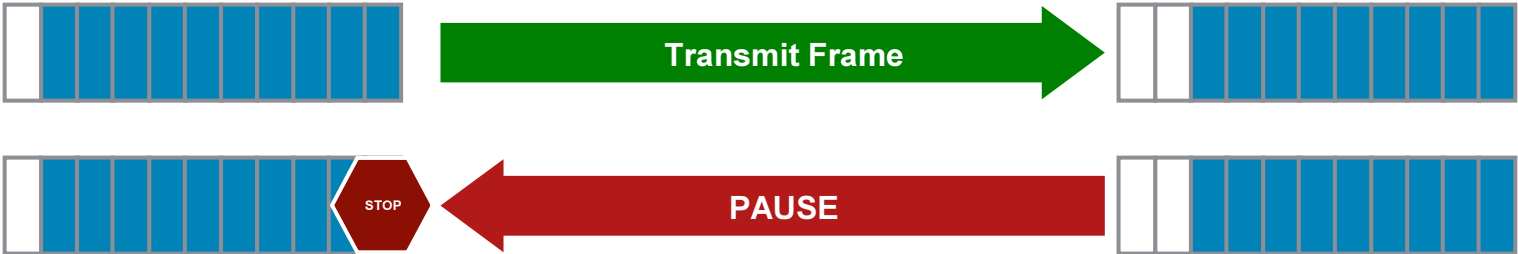
Priority-Based Flow Control (PFC)



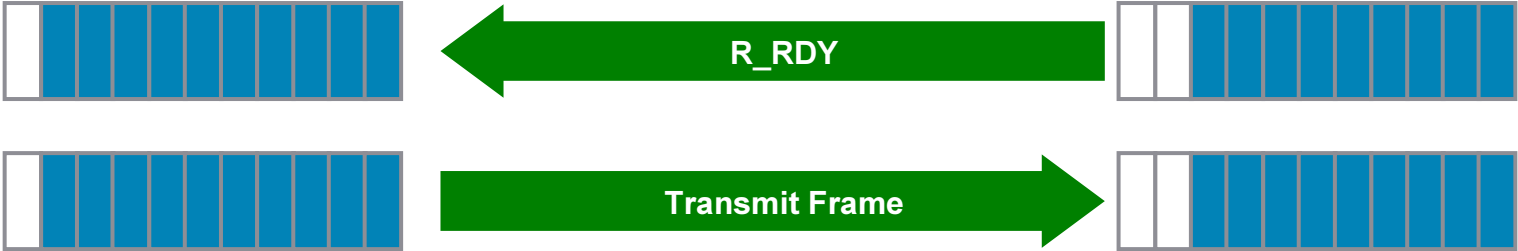
- Biztosítja a csomagvesztés mentességet minden 802.1p class of service osztályra
- Buffer limit elérésekor PAUSE üzenetet küld a megfelelő virtuális linken
- Hálózati erőforrások VL-enként vannak felosztva (input/output buffer queue)

Link Level Flow Control

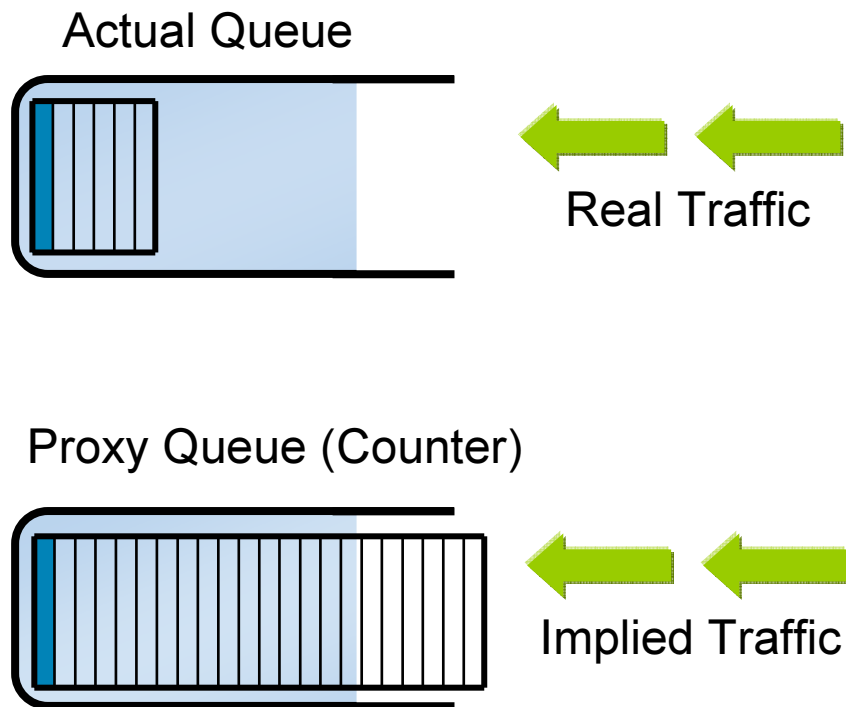
Ethernet PAUSE:



Fibre Channel Buffer-to-buffer Credits:

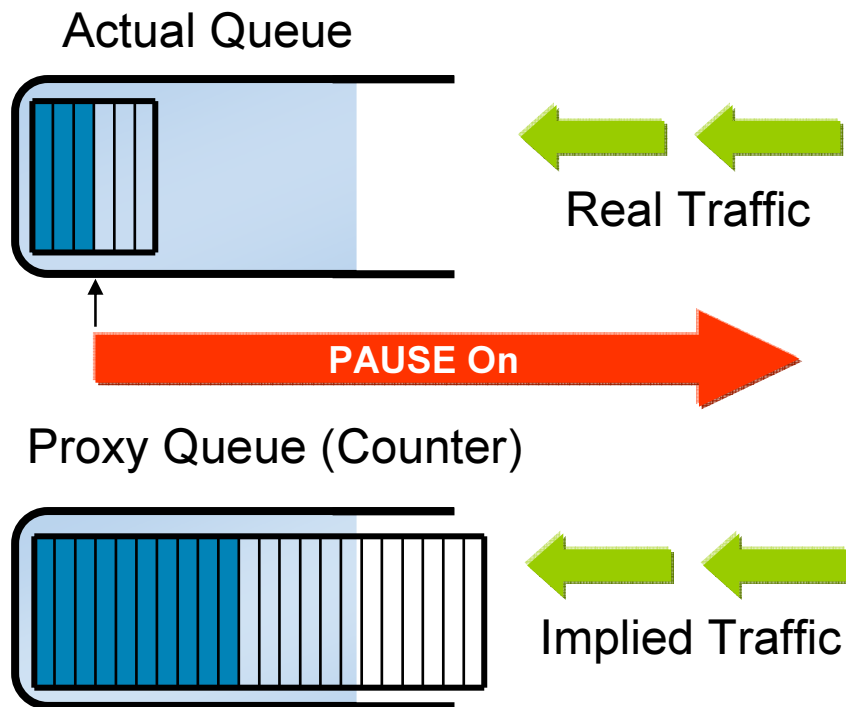


Delayed Drop: Normál forgalom



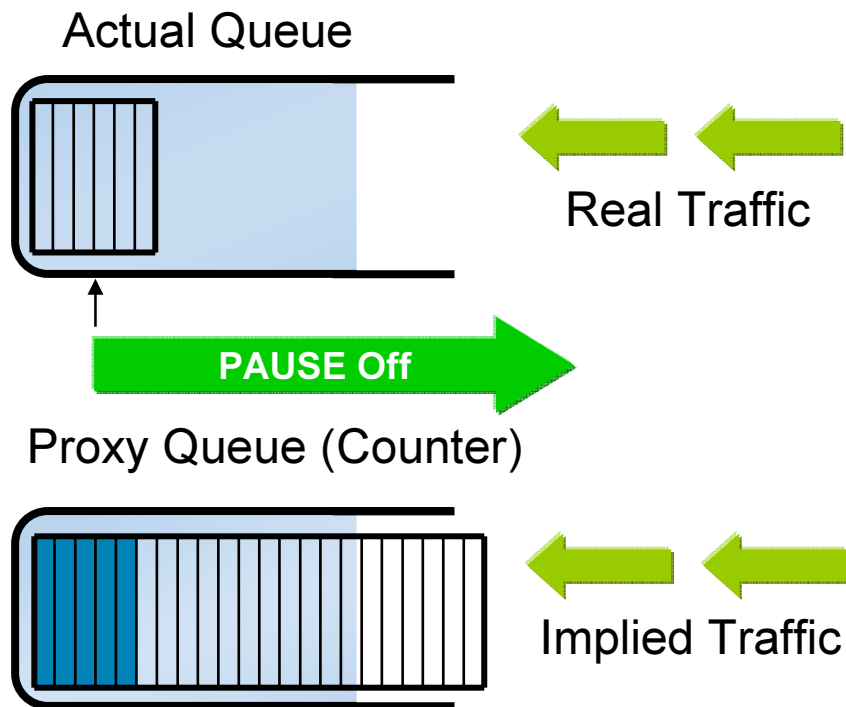
Actual Queue	Proxy Queue
Csomagot ad	Csomagot ad
Csomagot elvesz	Csomagot elvesz

Delayed Drop: Torlódás lép fel



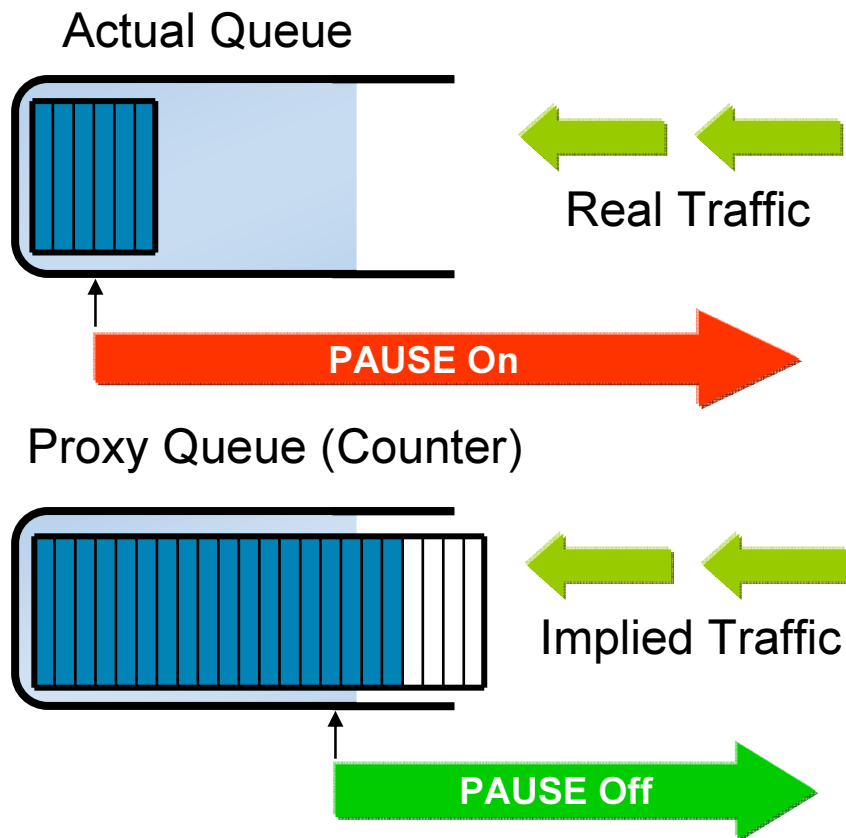
Actual Queue	Proxy Queue
Csomagot ad	Csomagot ad
Csomagot elvesz	Csomagot elvesz
Elküld PAUSE-t	Csomagot ad vonali sebességgel

Delayed Drop: Tranziens toródás



Actual Queue	Proxy Queue
Csomagot ad	Csomagot ad
Csomagot elvesz	Csomagot elvesz
Elküld PAUSE-t	Csomagot ad vonali sebességgel
Üres	Csomagot elvesz vonali sebességgel

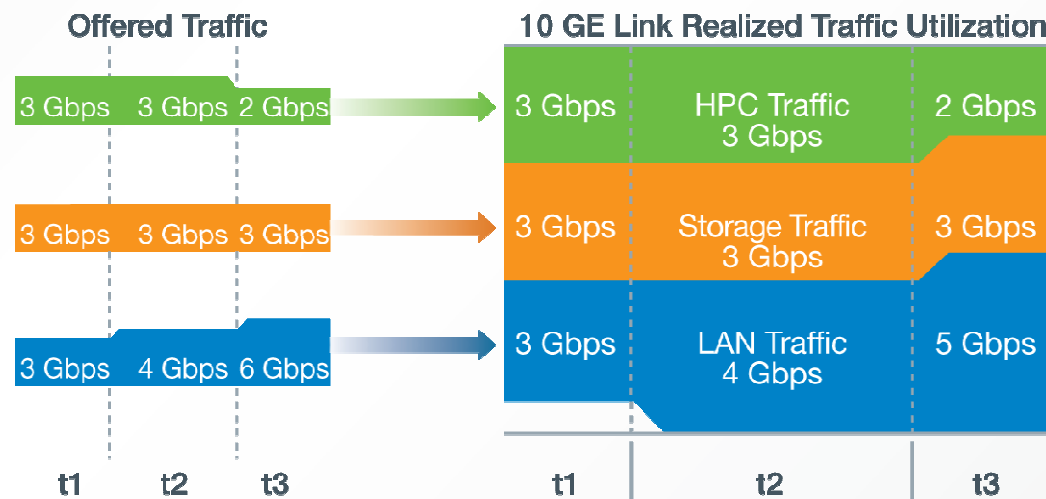
Delayed Drop: Folyamatos torlódás



Actual Queue	Proxy Queue
Csomagot ad	Csomagot ad
Csomagot elvesz	Csomagot elvesz
Elküld PAUSE-t	Csomagot ad vonali sebességgel
Üres	Csomagot elvesz vonali sebességgel

Data Center Ethernet Jellemzők

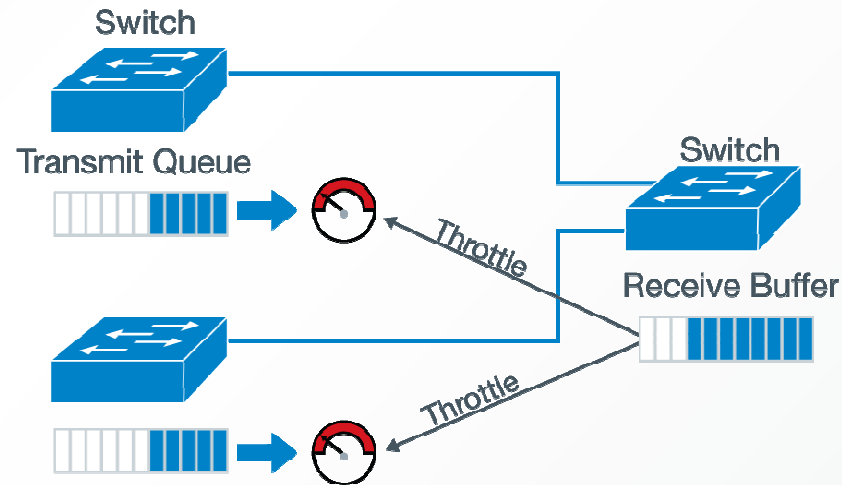
Sávszélesség Menedzsment



- **Intelligens sávszélesség megosztás az egyes forgalmi osztályok között**
- **IEEE 802.1Qaz szabványosítás alatt**
- **Priority Grouping néven is ismert**

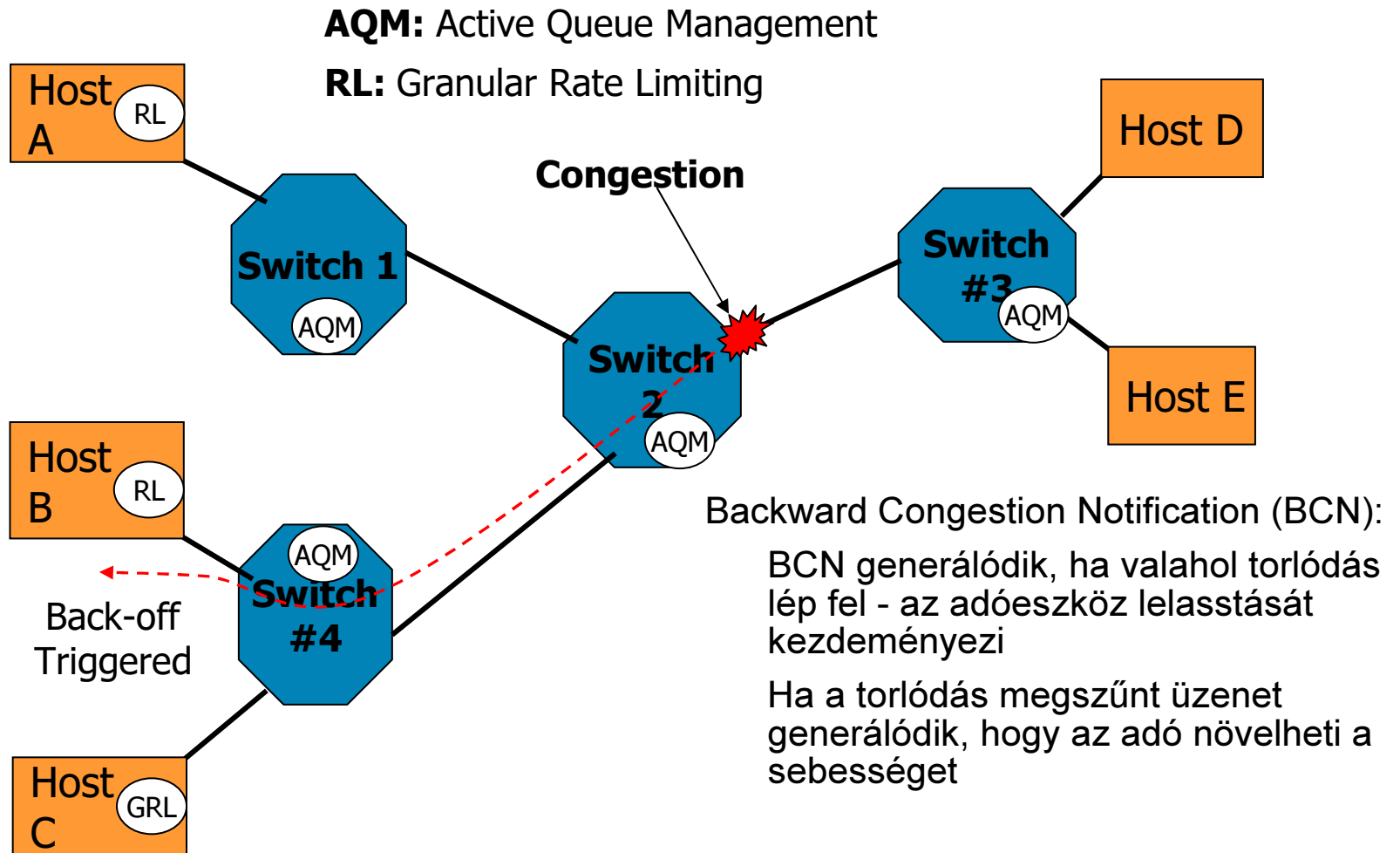
Data Center Ethernet Jellemzők

Torlódás Menedzsment



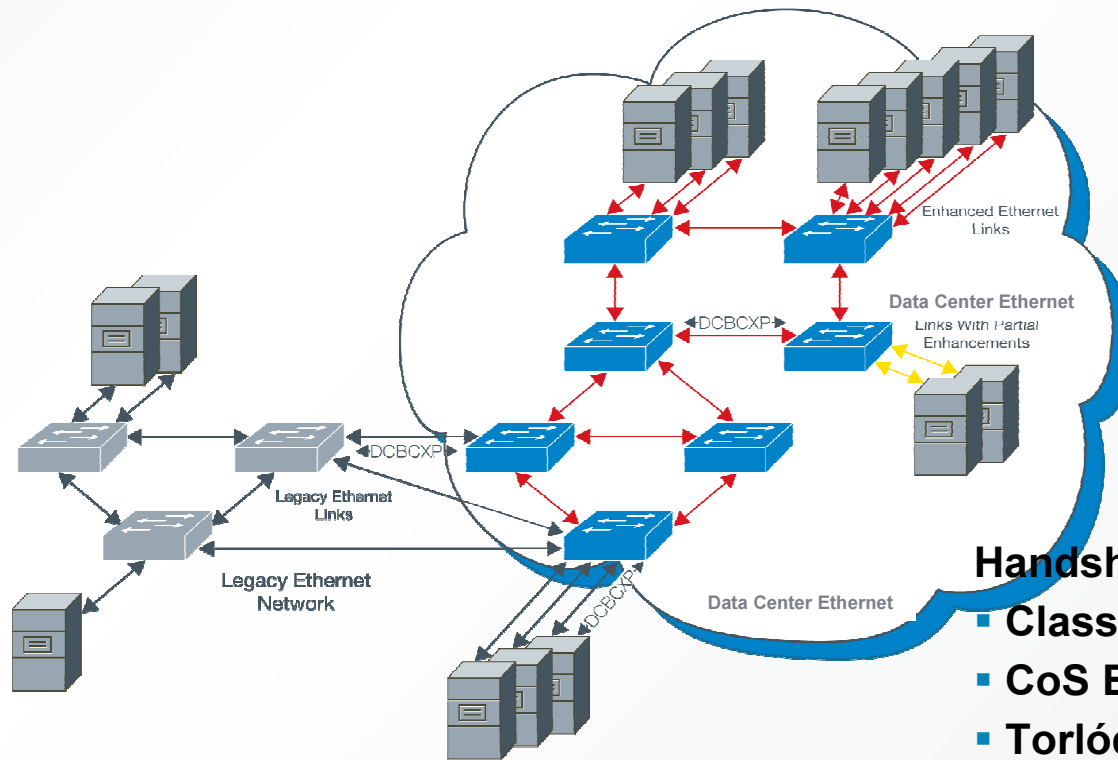
- Gerinc Torlódás mentesítésre szolgál
- End-to-End congestion menedzsment
- 802.1Qau szabványosítás alatt

End-to-End Torlódás menedzszment



Data Center Ethernet Jellemzők

Data Center Bridging Capability eXchange Protocol



Handshaking mechanismus:

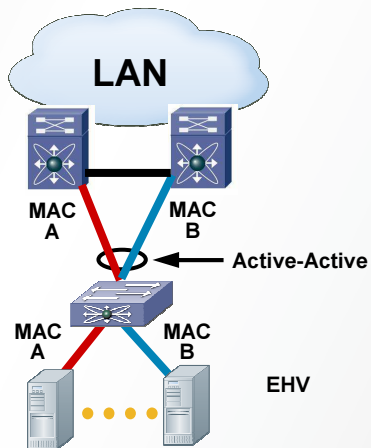
- Class Based Flow Control
- CoS BW menedzsment
- Torlódás menedzsment
- Alkalmazások (user_priority)
- Logical Link Down

<http://www.intel.com/technology/eedc/index.htm>

Data Center Ethernet Jellemzők

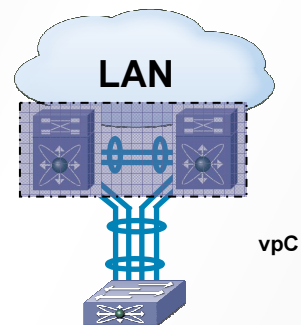
Layer 2 Multi-Pathing

Fázis 1



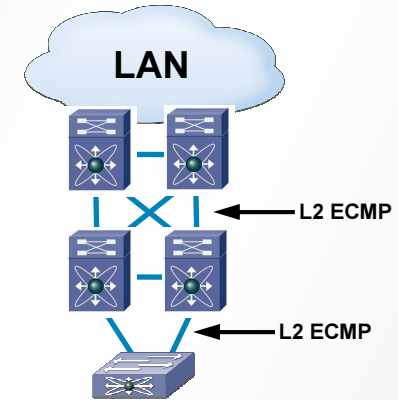
- Feleslegessé teszi az STP-t az uplink portokon
- Több egyidejűleg aktív uplinket biztosít a Switch és a hálózat között
- Hurok mentesítés MAC Pinning technológiával - a forgalom egyszerre csak egy porton történik az adott eszközre
- Transzparens

Fázis 2



- MCEC megvalósítás
- Virtual port channel mechanizmus transzparens a hálózat felé
- STP csak biztonsági megoldás controll plane hiba esetén

Fázis 3



- ISIS szerű L2 topológia
- L2 domén hurokmentesítése STP nélkül
- Elsődleges út választás, loadbalancing

Ethernet Host Virtualizer NPV

- Szükségtelemé teszi a Spanning Tree Protocolt az Uplink Portokon

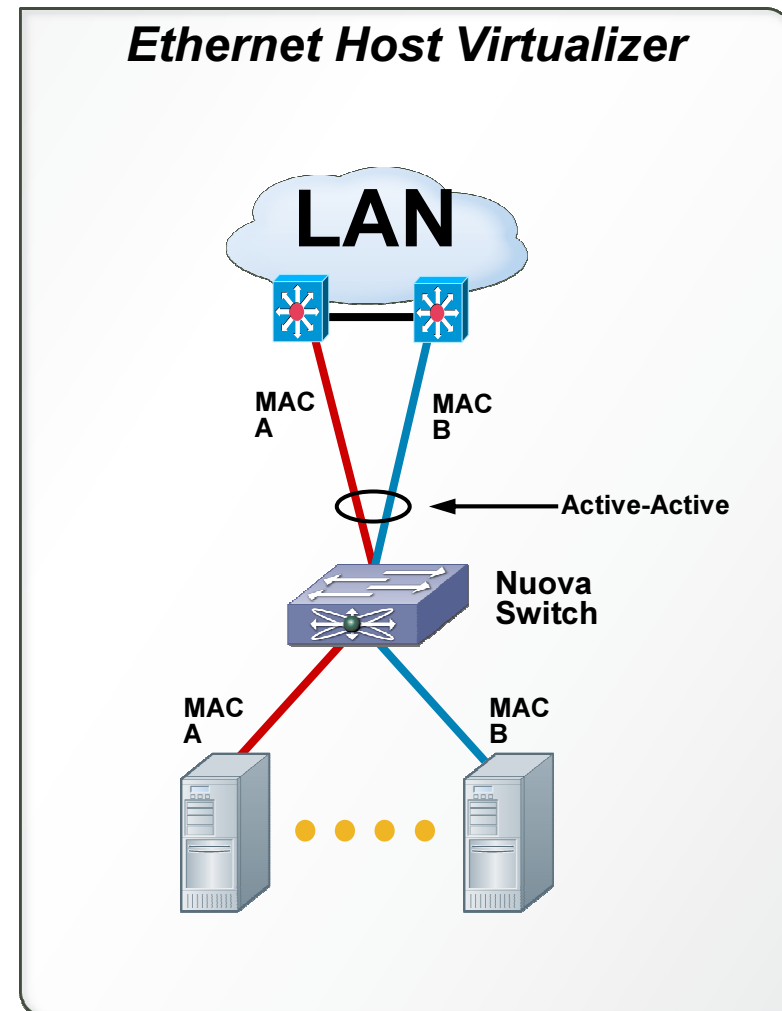
Gyorsabb, megbízhatóbb

- Aktív-Aktív kapcsolatot biztosít

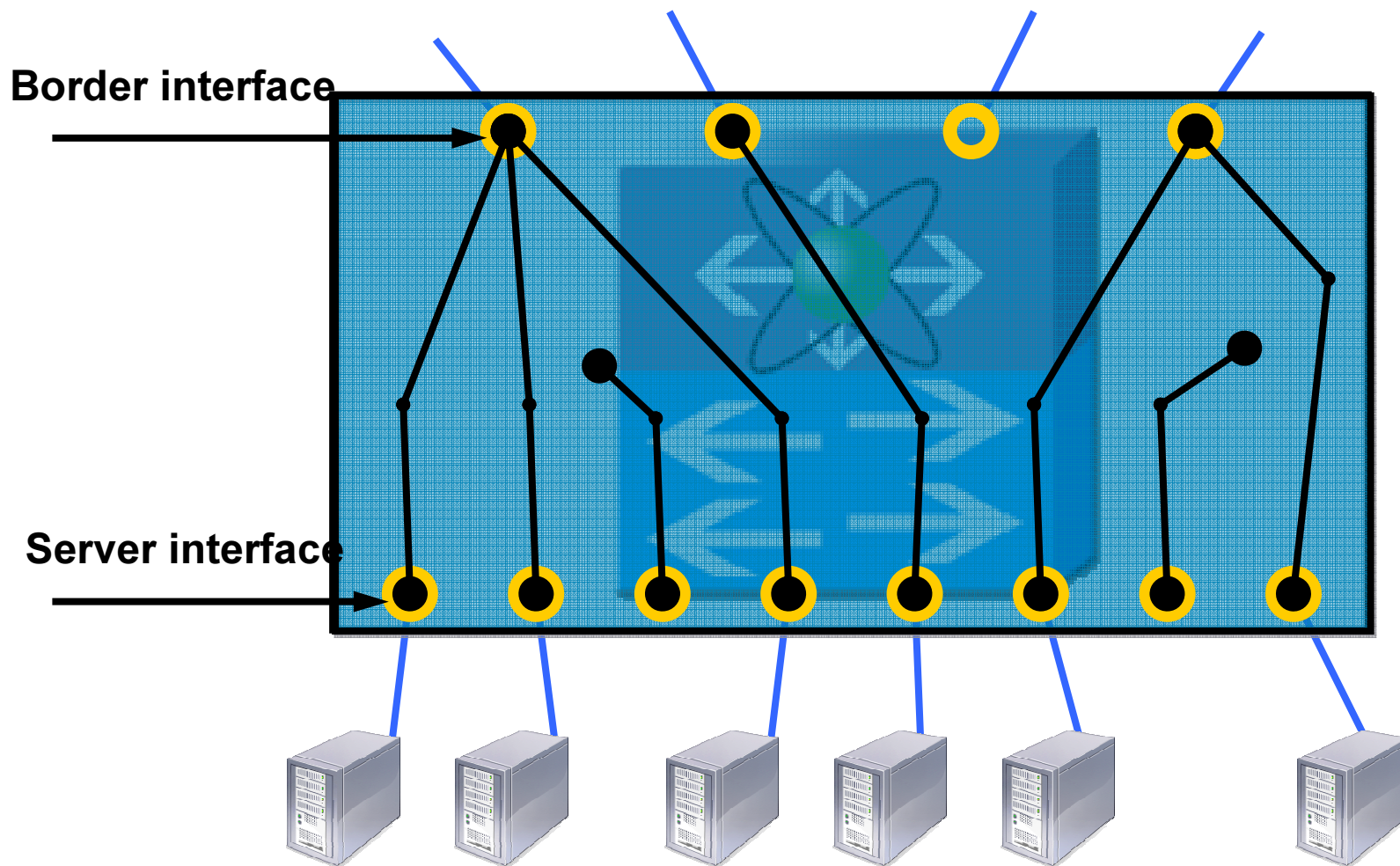
Dupla sávszélesség az STP-hez képest

- Hurok mentesítés MAC Address Pinning technológiával

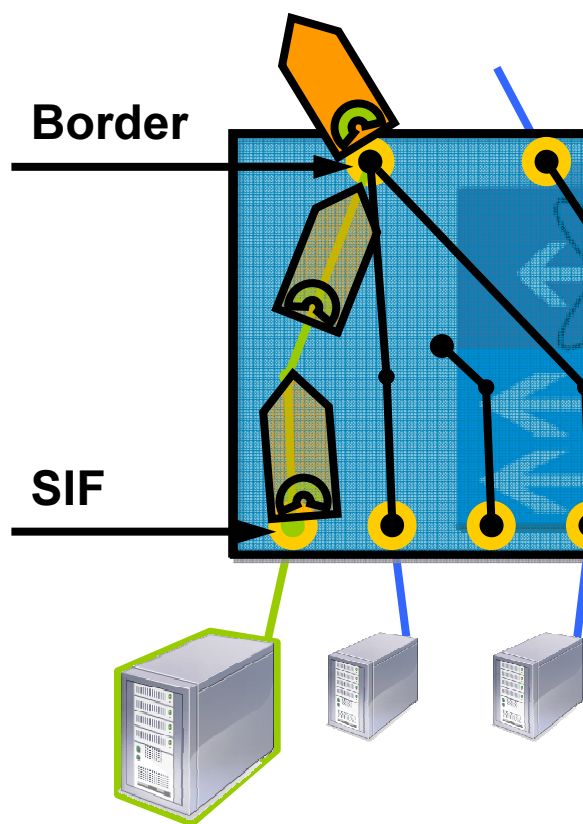
- Teljes transzparencia



Pinning

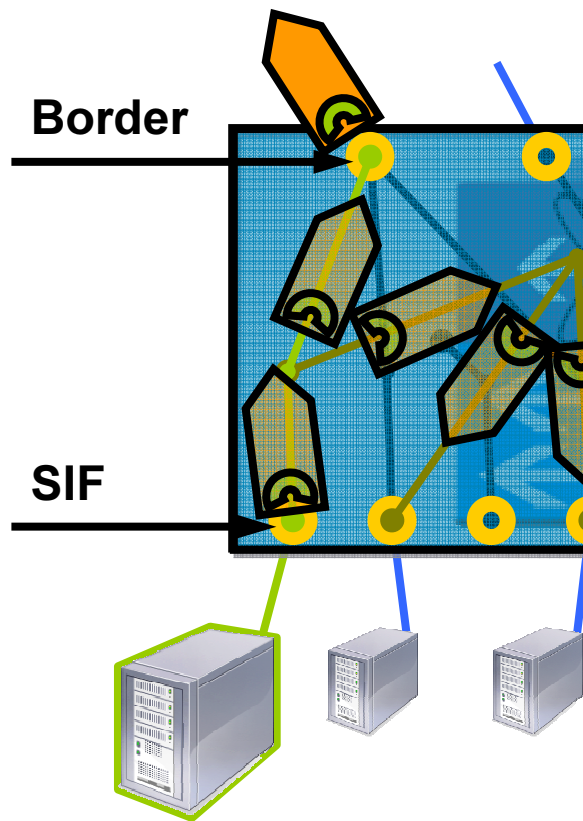


Kimenő forgalom: ismert unicast



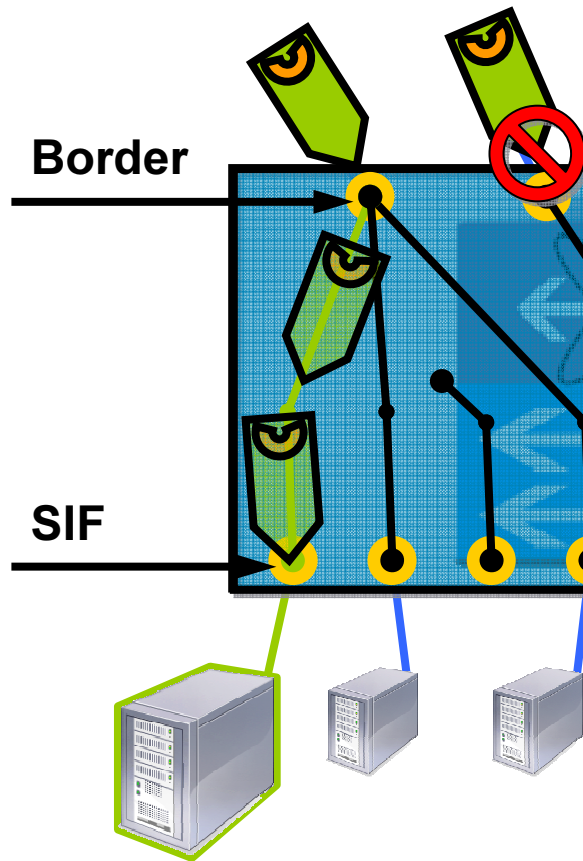
- A SIF irányból érkező forgalom a megfelelő „pinning” porton lép ki a switchből

Kimenő forgalom: Multicast/broadcast



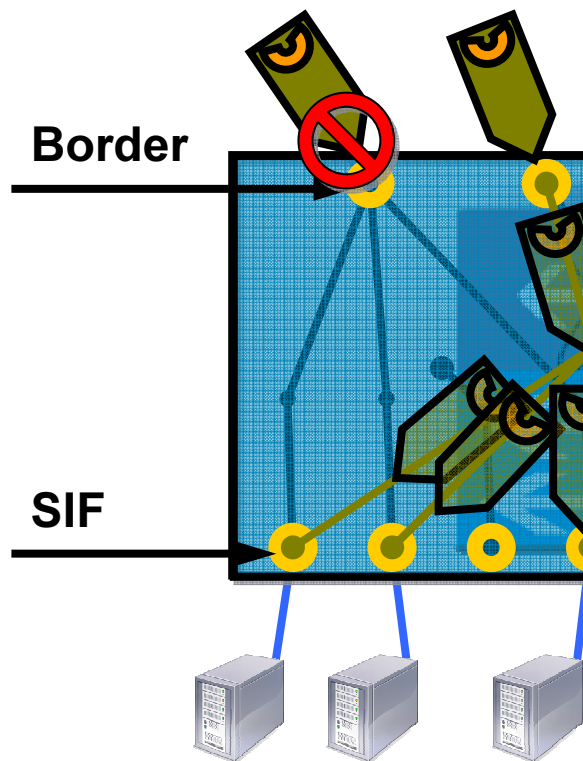
- A lokális forgalom replikálást a switch végzi a SIF portok felé
- Egyetlen mcast/bcast csomagot küld felfelé a megfelelő „pinning” porton

Bejövő forgalom: Reverse Path Forwarding



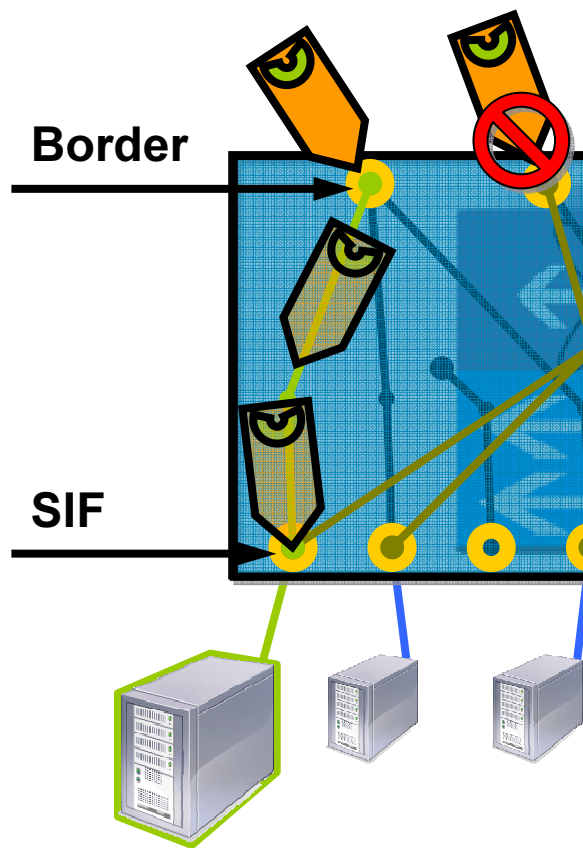
- Adott SIF porton lévő eszköznek szóló csomagot kizárólag a megfelelő „pinning” porton fogadja el
- (A SIF-en lévő MAC címeknek propagálódni kell a teljes switch adatbázisába)

Bejövő forgalom: Multicast/broadcast



- Egyetlen Border interface-t választ, ami minden SIF számára fogadja a mcast/bcast üzeneteket

Bejövő forgalom: Deja-vu check



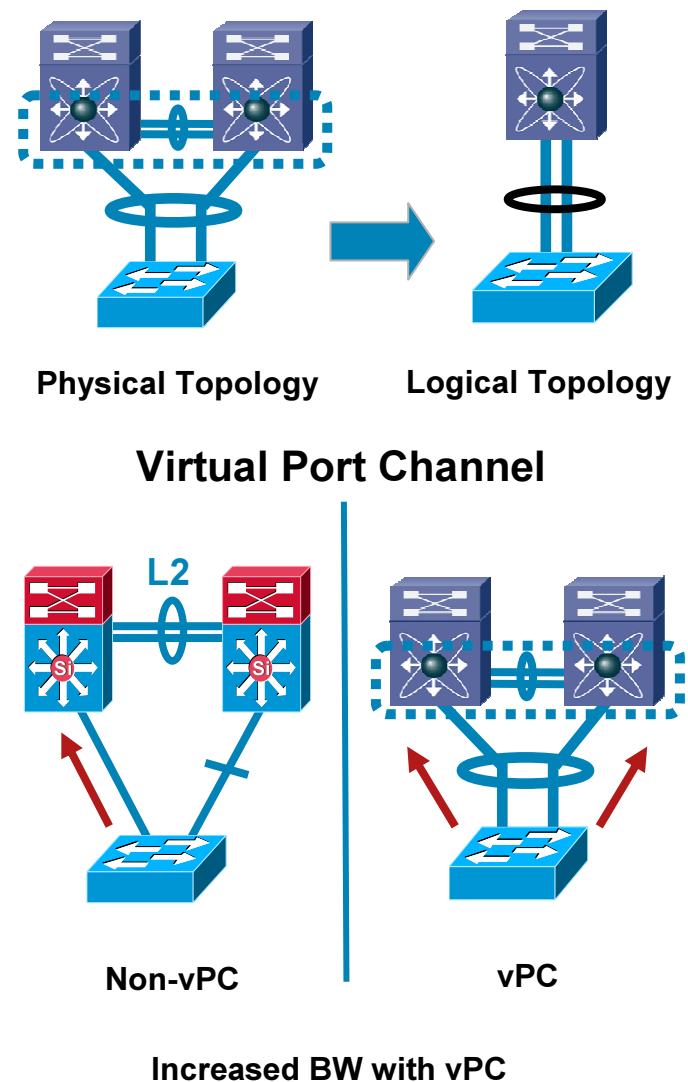
- Ha a bejövő csomag forrás lokális SIF MAC címről érkezik a Border porton

Multicast/broadcast logika eldobja a csomagot

A megfelelő „pinning” port elfogadja a csomagot de cél címtől függetlenül eldobja azt

VPC – MultiChassis EtherChannel

- vPC egy Port-channeling koncepció, ami a link aggregációt két külön fizikai switchen valósítja meg
- Redundáns L2 topológia alapú Link Aggregáció
- Szükségtelemmé teszi az STP-t az access-distribúciós réteg között
- Provides increased bandwidth
All links are actively forwarding
- MCEC is available in two implementations:
VSS on the Cat6k
vPC as a standalone feature in NX-OS on the Nexus 5000 and 7000 (Q1CY09)



“TRILL” WG az IETF-ben

- “Transparent Interconnection of Lots of Links” Internet Drafts (“Routing Bridges” vagy “RBridges”)

IETF törekvés az L2 STP limitációk kivédésére

TRILL célja STP nélküli megoldás a végberendezések és switchek vonatkozásában

Cél az STP helyettesítése link-state routing protocoollal

- Benefits –

Shortest-Path Frame routing multi-hop 802.1-L2 környezetben

Biztosítja a terhelés megosztást több link esetén (aktív-aktív)

Adás a cél bridge-id alapul

- Kompatibilis kell legyen a klasszikus 802.1d –vel

- Referenciák:

<http://www.ietf.org/html.charters/trill-charter.html>

<http://www.ietf.org/internet-drafts/draft-ietf-trill-rbridge-protocol-03.txt>

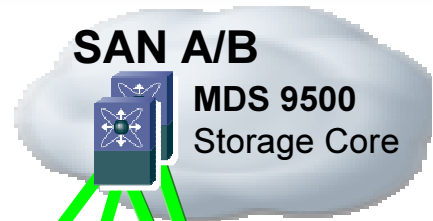
Unified Fabric Fejlődése

DC Core

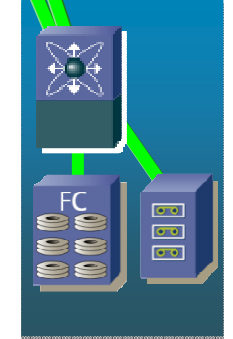
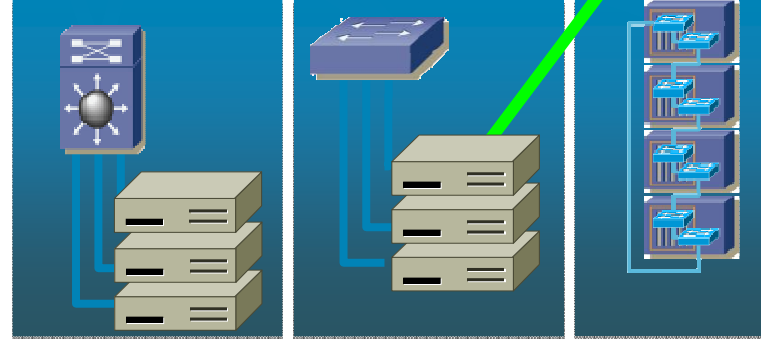
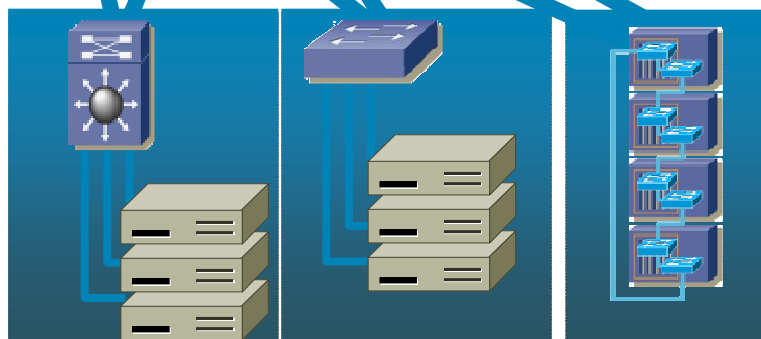


- Gigabit Ethernet
- 10 Gigabit Ethernet
- 10 Gigabit DCE
- 4Gb Fibre Channel

DC Aggregation



DC Access



Catalyst 6500 End-of-Row Catalyst 49xx Rack CBS 3100 Blade

Catalyst 6500 End-of-Row Catalyst 49xx Rack CBS 3100 MDS 9124e Blade

MDS 9500 Storage

1GbE Server Access

1GbE and 4Gb FC Server Access

Storage

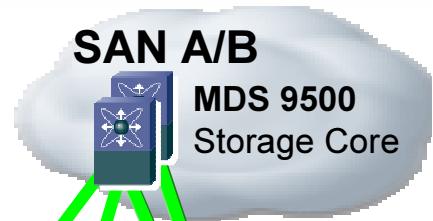
Unified Fabric Fejlődése

DC Core

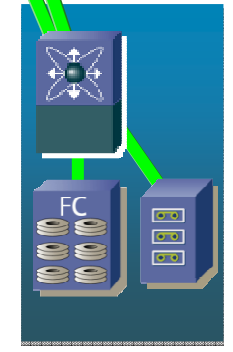
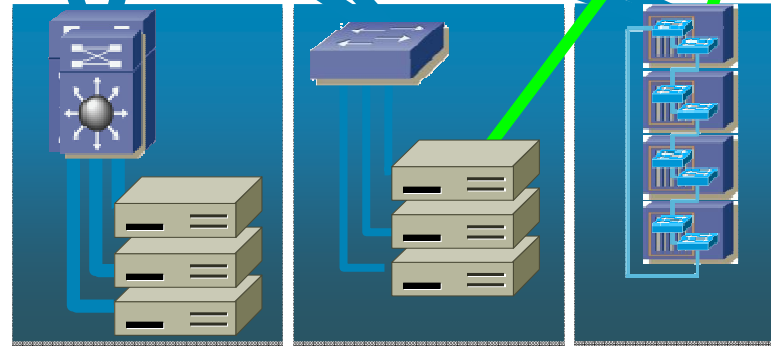
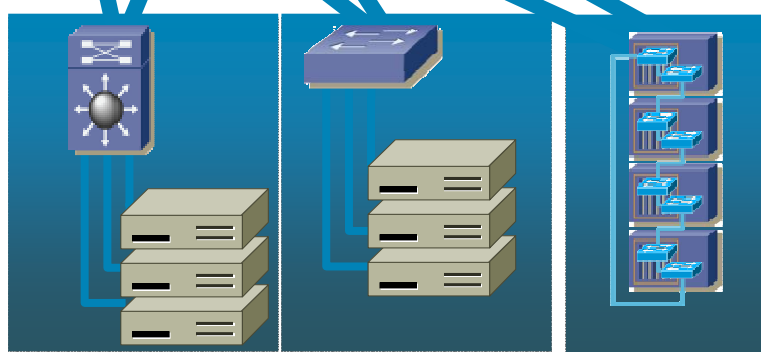


- Gigabit Ethernet
- 10 Gigabit Ethernet
- 10 Gigabit DCE
- 4Gb Fibre Channel

DC Aggregation



DC Access



1GbE Server Access

1/10GbE and 4Gb FC Server Access

Storage

1GbE Server Access

1/10GbE and 4Gb FC Server Access

Storage

Unified Fabric Fejlődése

DC Core

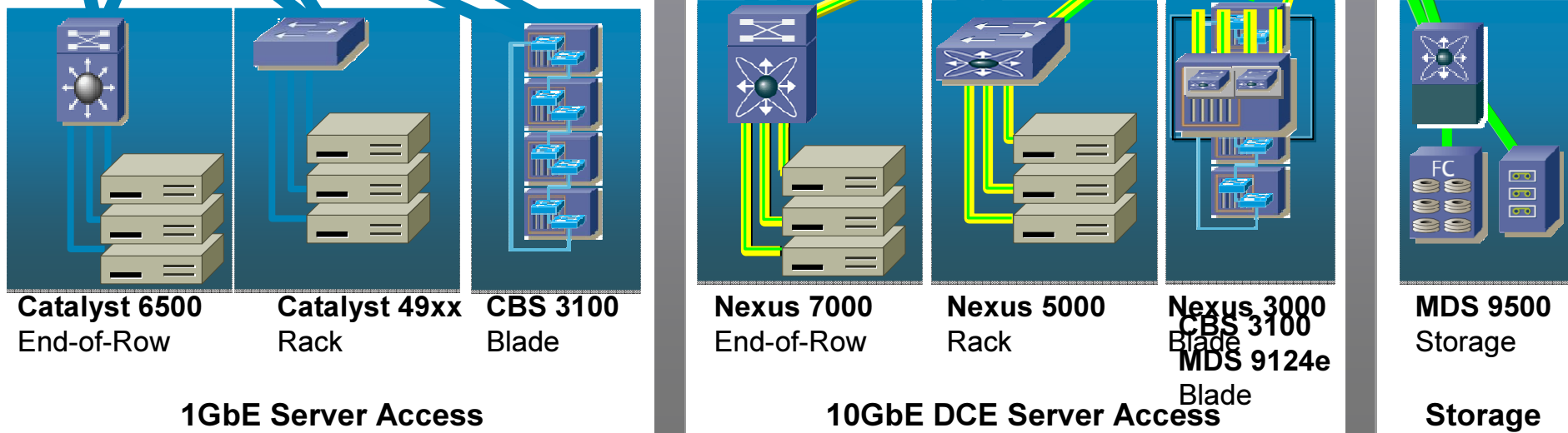


- Gigabit Ethernet
- 10 Gigabit Ethernet
- 10 Gigabit DCE
- 4Gb Fibre Channel
- 10 Gigabit FCoE/DCE

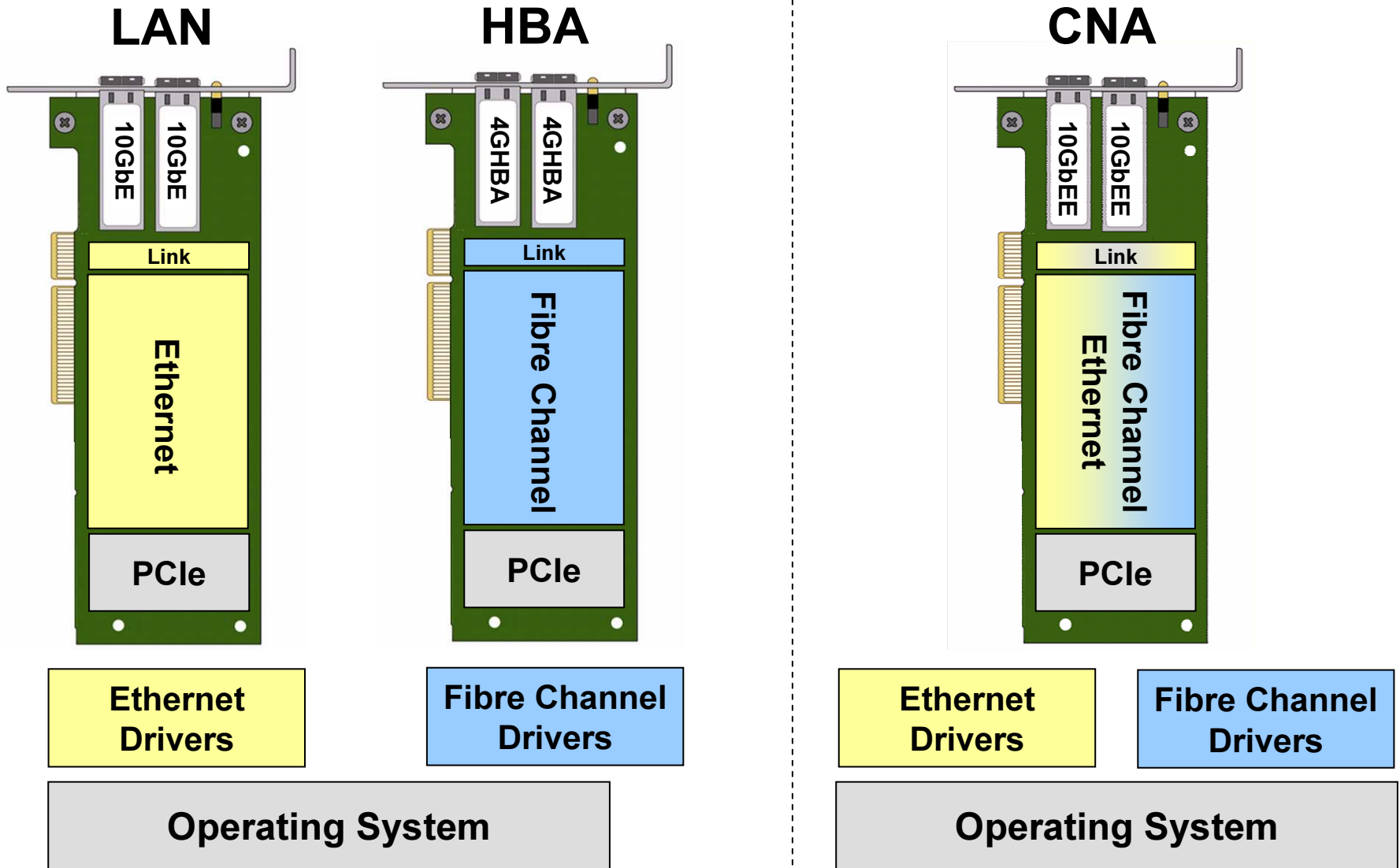
DC Aggregation



DC Access



Converged Network Adapter (CNA)

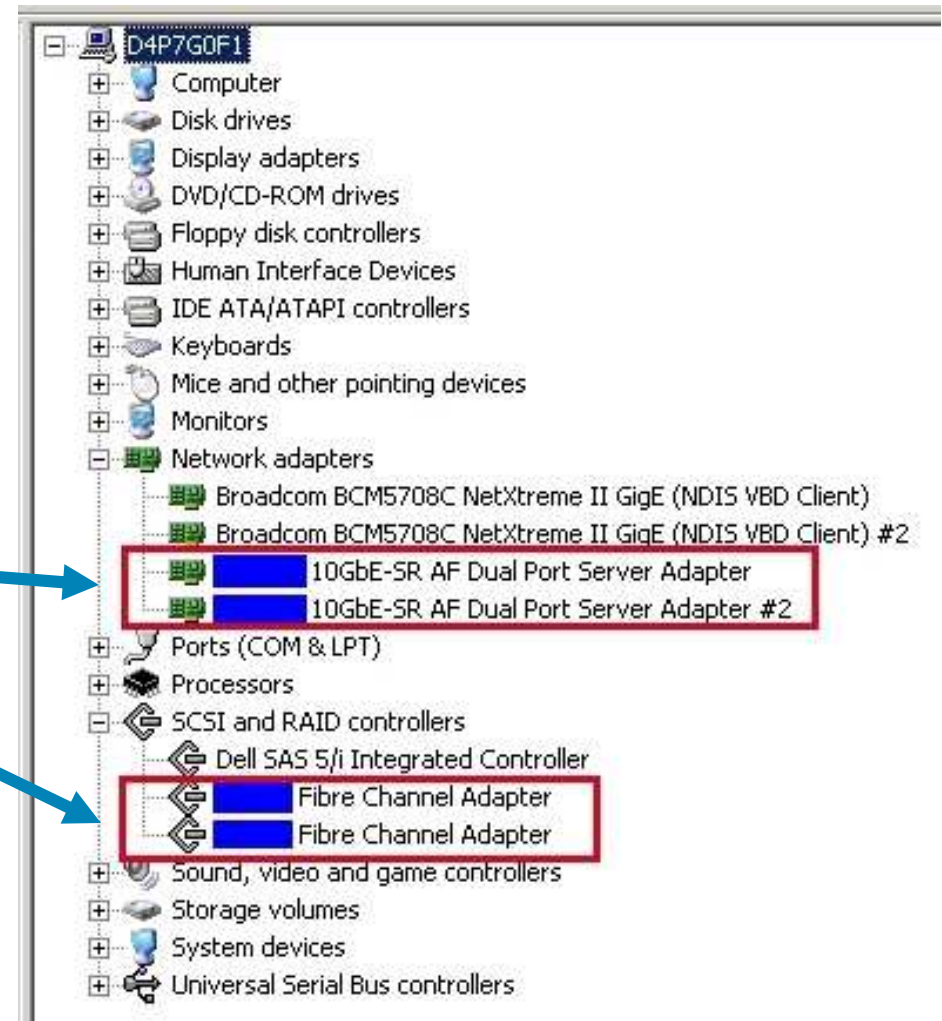


FCoE megjelenés az Operációs rendszerben

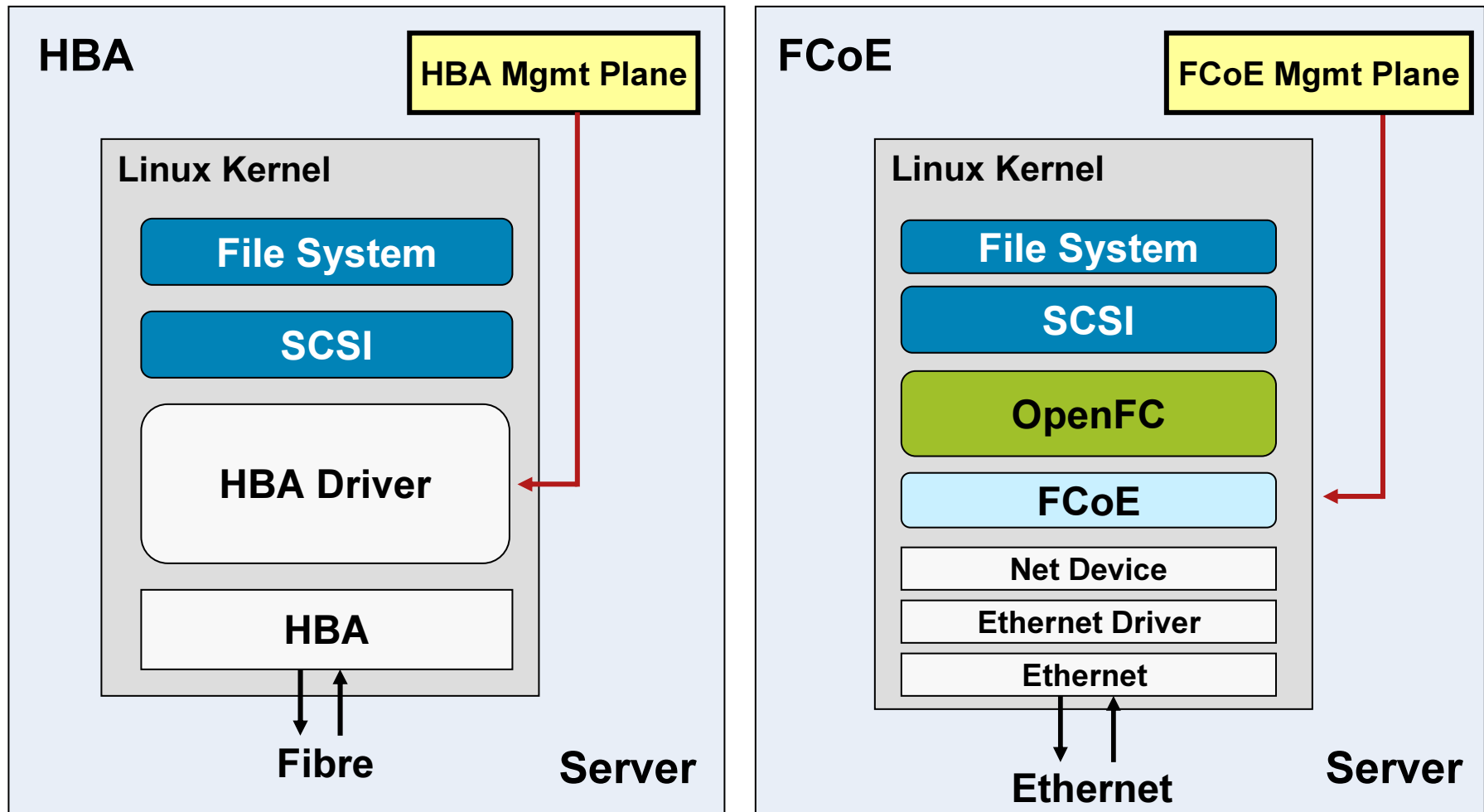
- Standard driverek
- Azonos menedzsment
- Operációs rendszer megjelenés:

Dual port 10 Gigabit Ethernet adapter

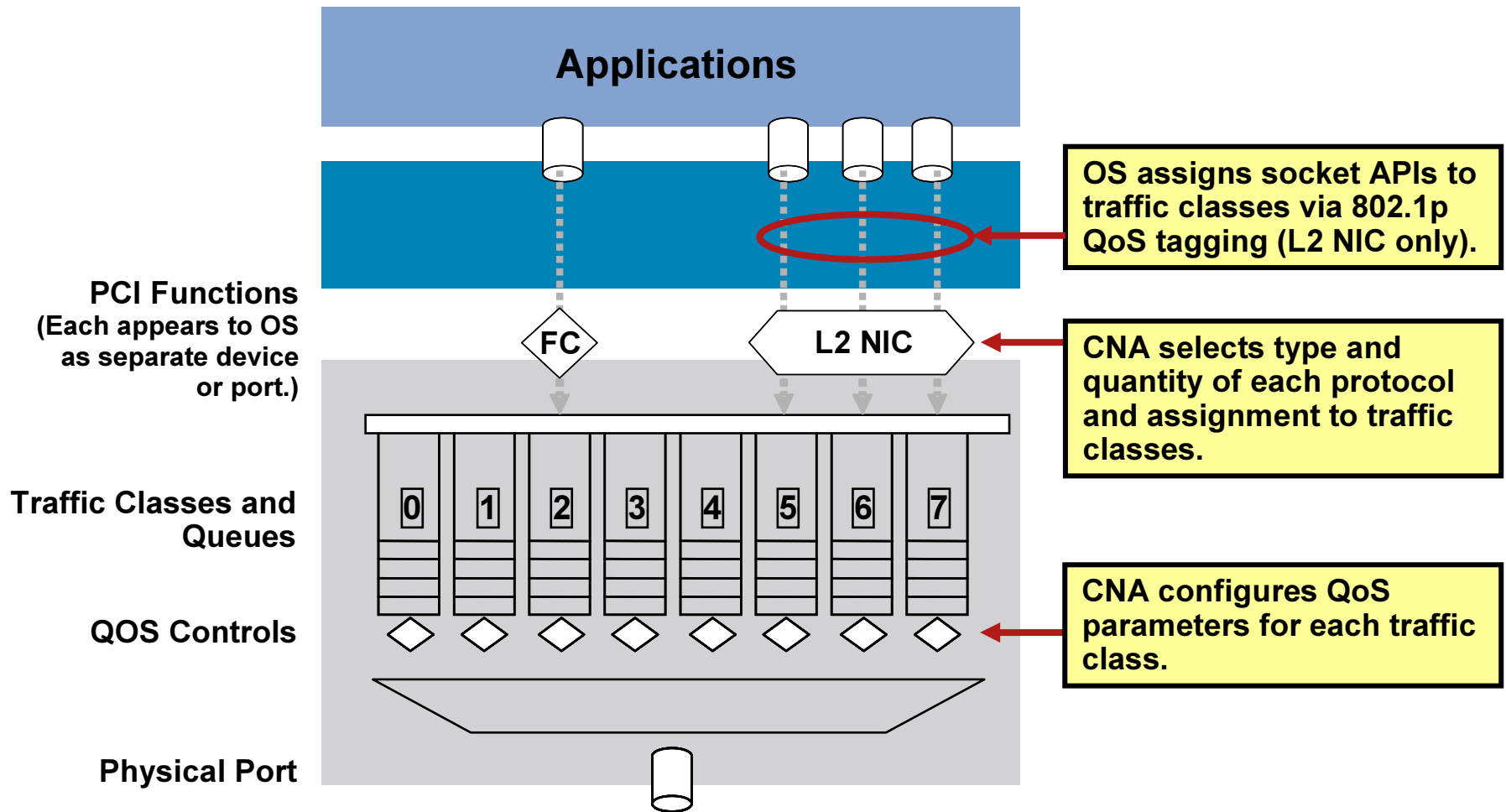
Dual Port 4 Gbps Fibre Channel HBA



FCoE software megvalósítása Intel Ethernet NIC-ben



Forgalom Priorizálás



Multi-Protocol Server Switch

Az első Unified I/O-t támogató switch

NX5020

Switch Family



56-Port L2 Switch

- 40 Ports 10GE fixed
- 2 Expansion Modules

Expansion Modules



Fibre Channel

- 8 Ports 1/2/4G FC



FC + Ethernet

- 4 Ports 10GE
- 4 Ports 1/2/4G FC



Ethernet

- 6 Ports 10GE

Partners



CNA

- 2 Port 10GE/FCoE



Server Adapter

- FCoE SW stack

OS

Cisco NX-OS

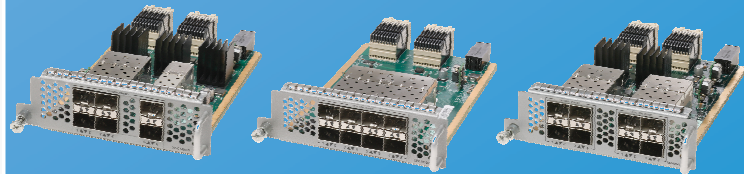
Mgmt

Cisco Fabric Manager and Cisco Data Center Network Manager

Valamennyi 10GE switch/module port FCoE/DCE képes

Nexus 5000 jellemzők

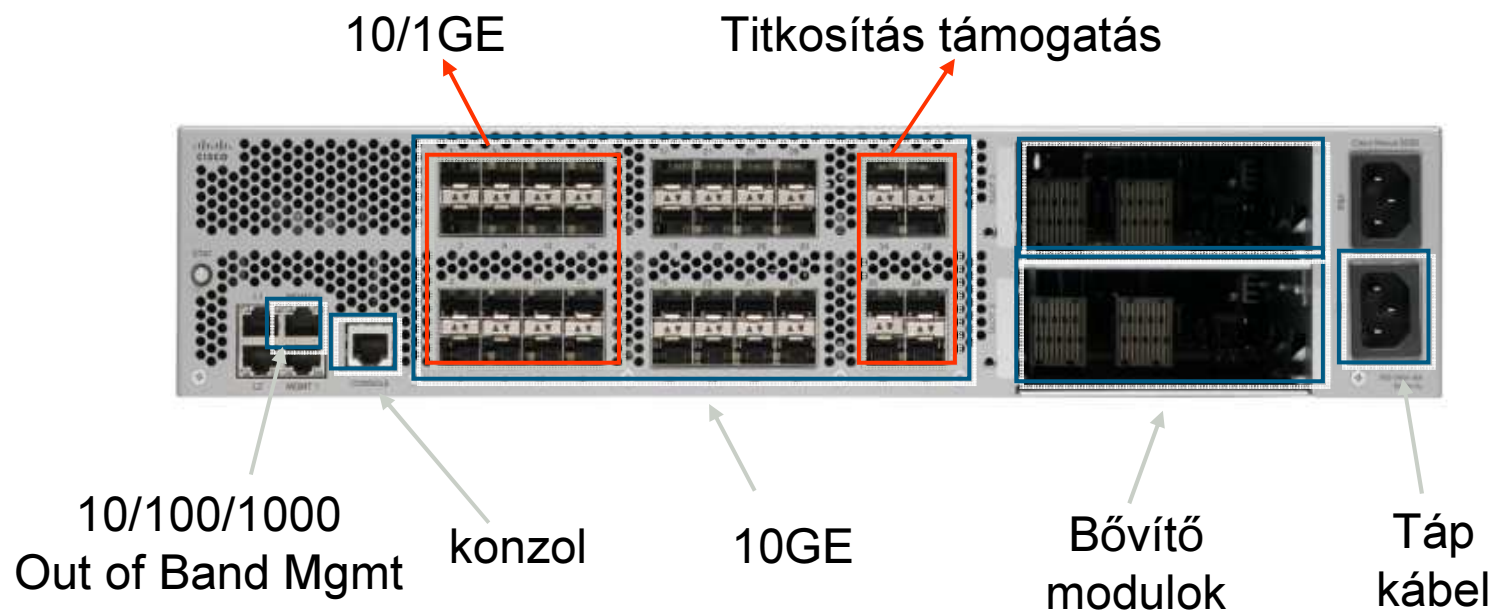
Max Port sűrűség
Max 52X10GE Line Rate
Max 16X1/2/4G FC + 40X10GE



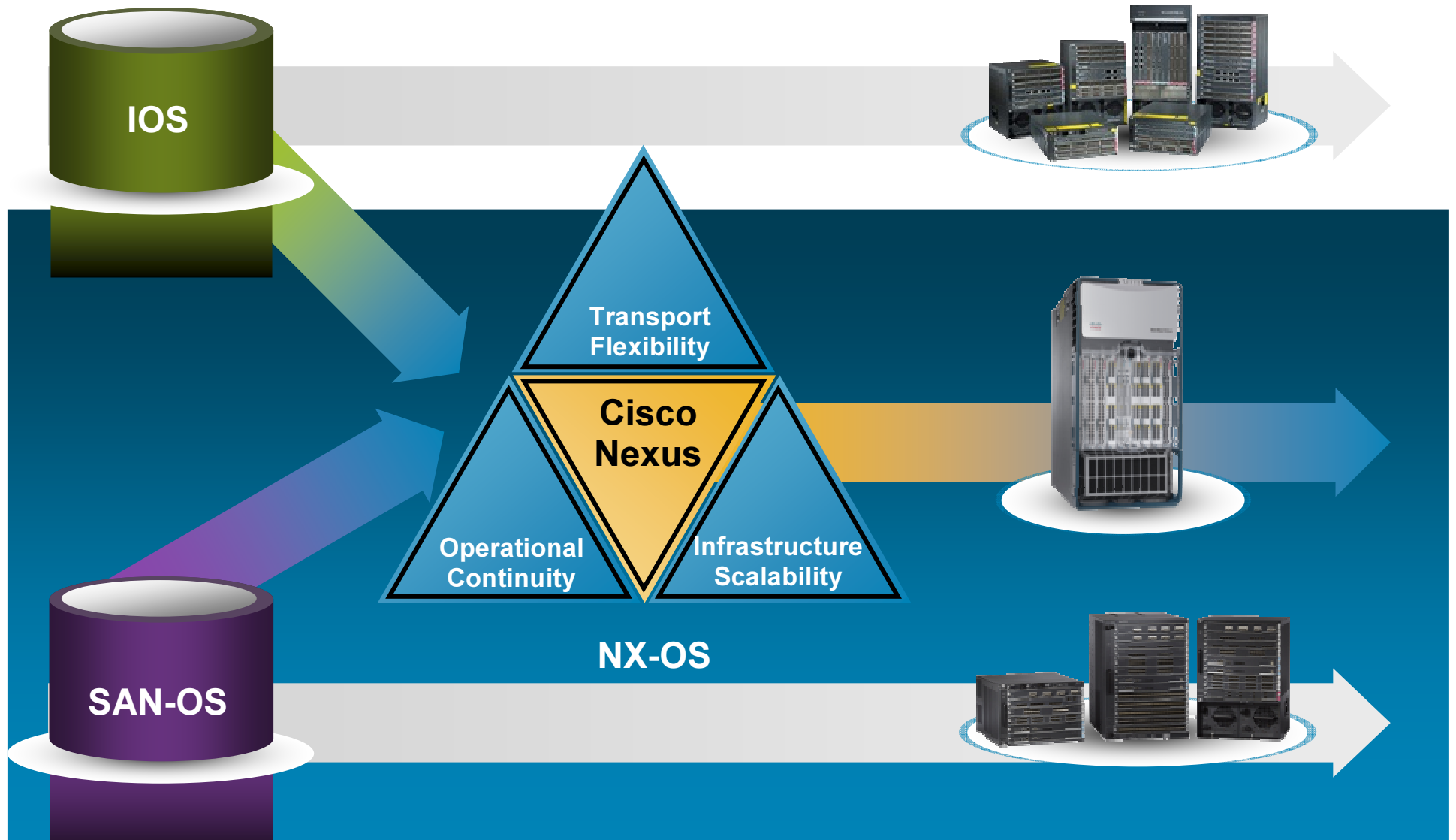
- **Jellemzők**
 - FCoE alapú Unified Fabric támogatás
 - Virtualizációra optimalizált
 - Ethernet, DCE, FCoE és FC támogatás
 - L2 támogatás
 - 56 teljes sebességű 10GbE port
- **Architektúra**
 - Max 1.2Tbps kapacitás
 - Cut-through architektúra
 - Kis 3.2µs port-port késleltetés
- **Nagy megbízhatóság – DC kategória**
 - Redundáns tápegység
 - Redundáns ventilátor

Hátlap

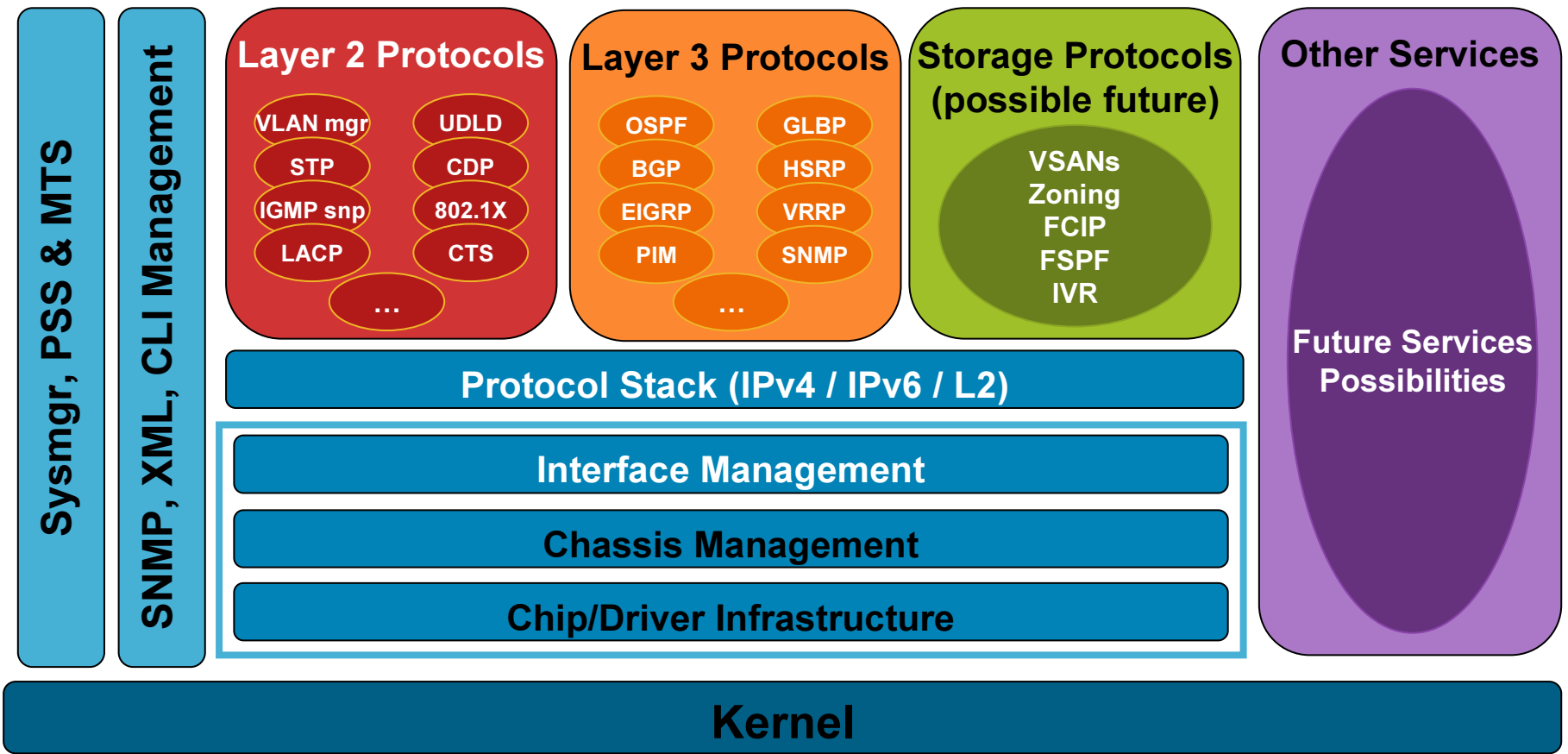
Összes 10GE port FCoE képes!



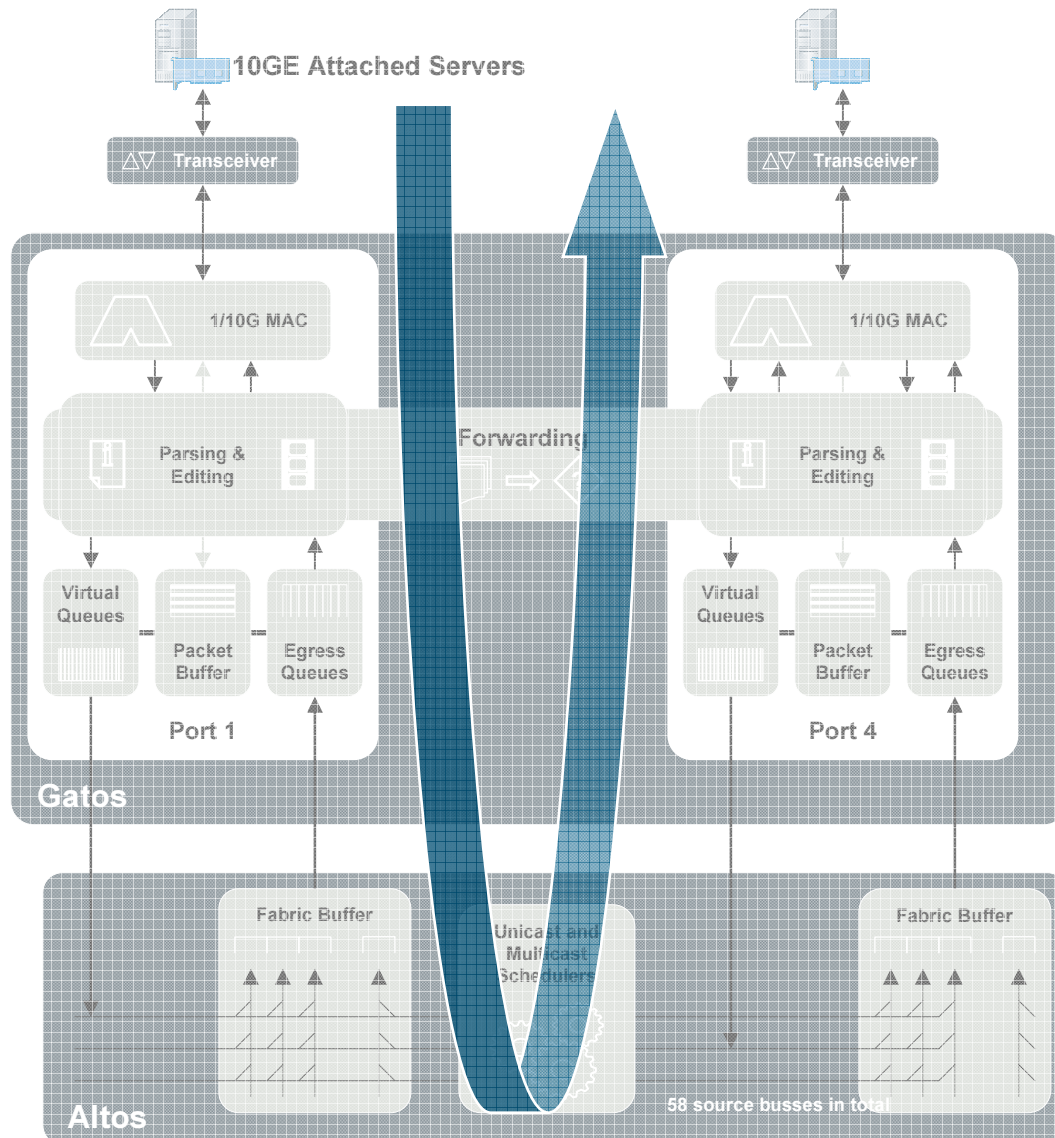
NX-OS: Adatközponti operációs rendszer



NX-OS – SAN OS alapú operációs rendszer



Késleltetés



- 3.2 μ sec port-port
Első byte be-ki alapon
Funkciók bekapcsolva
- 6.7 μ sec kernel - kernel
1.4 μ sec host adás
2.1 μ sec host vétel
OS, interrupt, és átviteli
késleltetés rakódik rá
- Socket layer app-app
Linux 2.6
Raw – 10.1 μ sec
UDP – 11.2 μ sec
TCP – 11.8 μ sec

SFP+ Ethernet meghajtók

- **Legkisebb 10GE** meghajtó
- Menetközben cserélhető
- Optikai SFP+ kompatibilis más 10GE modullal
 - XFP
 - XENPAK
 - X2
- Nexus 5000:
 - SFP+ Twinax rézkábel “direct connect”,
 - Kábel tartalmazza a meghajtót
 - 1M, 3M, & 5M
 - 10M később
 - Optikai meghajtó
 - SR
 - LR később



SFP+ Optical Module



SFP+ Copper

SFP+ Média lehetőségek

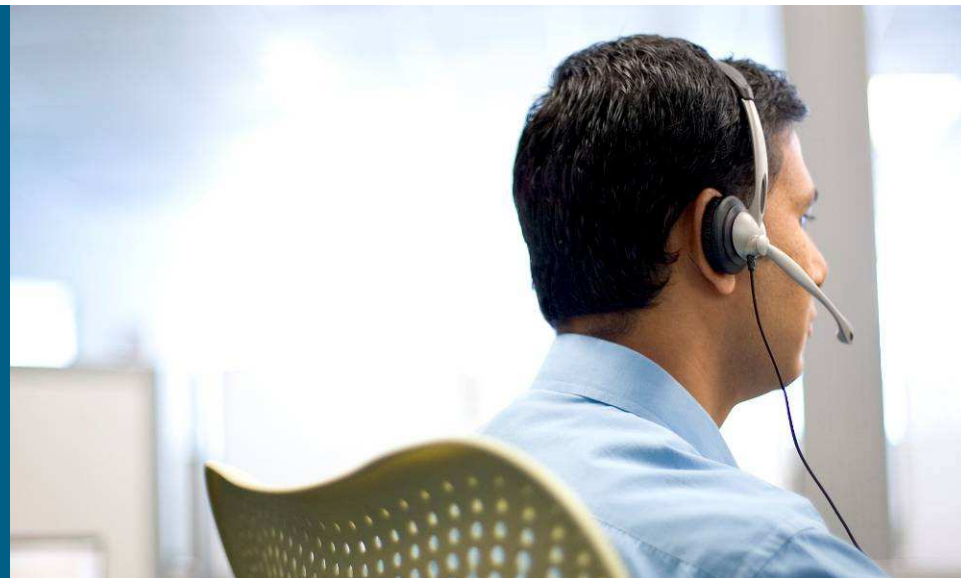


- Alacsony fogyasztás
- Alacsony ár
- Kis késleltetés
- Alacsony hiba arány(10 exp-17)

Technology	Cable	Distance	Power (each side)	Transceiver Latency (link)
SFP+ CU Copper	Twinax	10m	~0.1W	~0.25 μ s
SFP+ USR ultra short reach	MM OM2 MM OM3	10m 100m	1W	~0.1 μ s
SFP+ SR short reach	MM OM2 MM OM3	82m 300m	1W	~0.1 μ s
10GBASE-T	Cat6 Cat6a/7 Cat6a/7	55m 100m 30m	~8W ~8W ~4W	2.5 μ s 2.5 μ s 1.5 μ s



Hálózati megoldások Szervervirtualizáció esetén



Kihívások Vmware esetén

Cisco Virtual Network Link – VN-Link

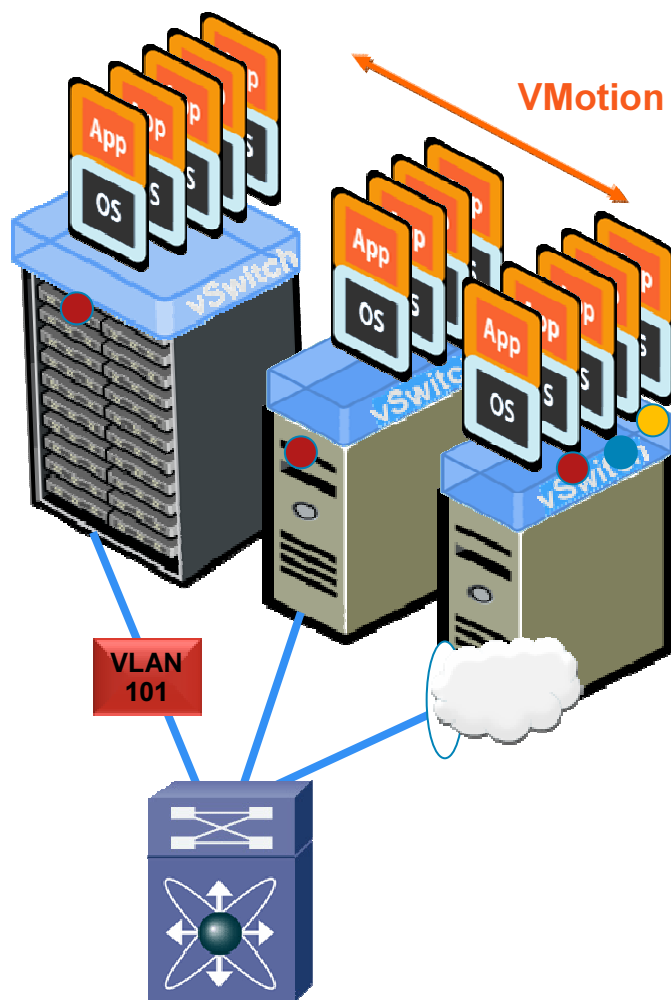
Hálózat alapú virtualizálás

Virtuális gépek az új építőkövek

Virtuális gép alapú hálózati (LAN, SAN) szolgáltatások

Absztrakt fizikai és logikai infrastruktúra

VM szintű hálózati skálázhatóság kérdései



Probléma:

- VMotion fizikailag más portra helyezheti át a Virtuális gépet—hálózati policynak ezt követnie kell
- Nem megoldható a gépen belüli lokális forgalom vizsgálata, policy beállítása
- Fizikai linkeken az egyes VM-ekre más hálózati szabályok vonatkoznak

VN-Link:

- Kiterjeszti a hálózatot VM-ekre
- Konzisztens hálózati szolgáltatás
- Egységes koherens menedzsment

Mi a VN-Link?

- VN egy speciális link a Virtuális gép virtuális NIC interface és a Cisco switch között
- VN-Link, (Virtual Network Link) a Virtuális gépek interface-eire kiterjesztett fogalom, ami egyedileg biztosítja a megszokott hálózati szolgáltatásokat, diagnosztika, policy(ACL, VLAN, QoS)

Cisco VN-Link

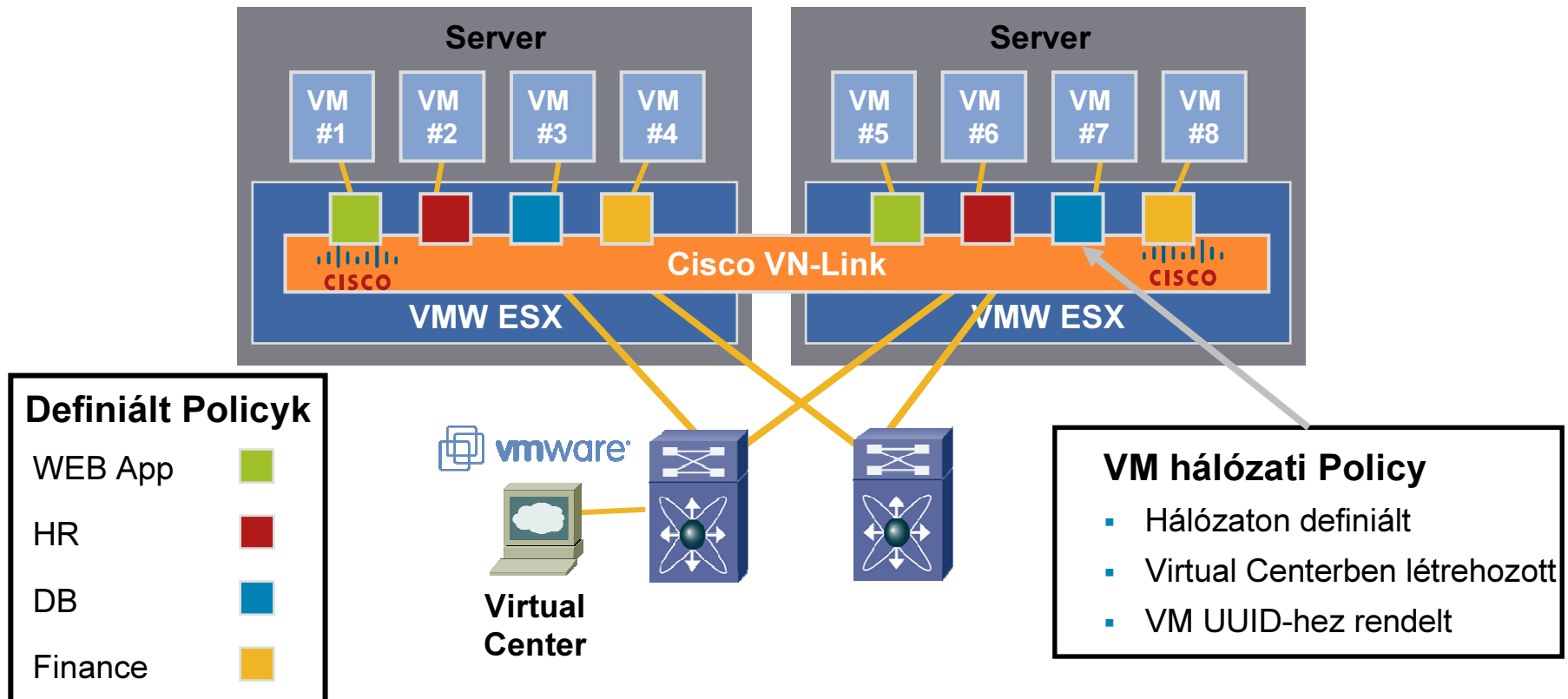
Gyorsabb VM létrehozás

Cisco VN-Link—Virtual Network Link

Policy-Based
VM Connectivity

Mobility of Network
& Security Properties

Non-Disruptive
Operational Model



Cisco VN-Link

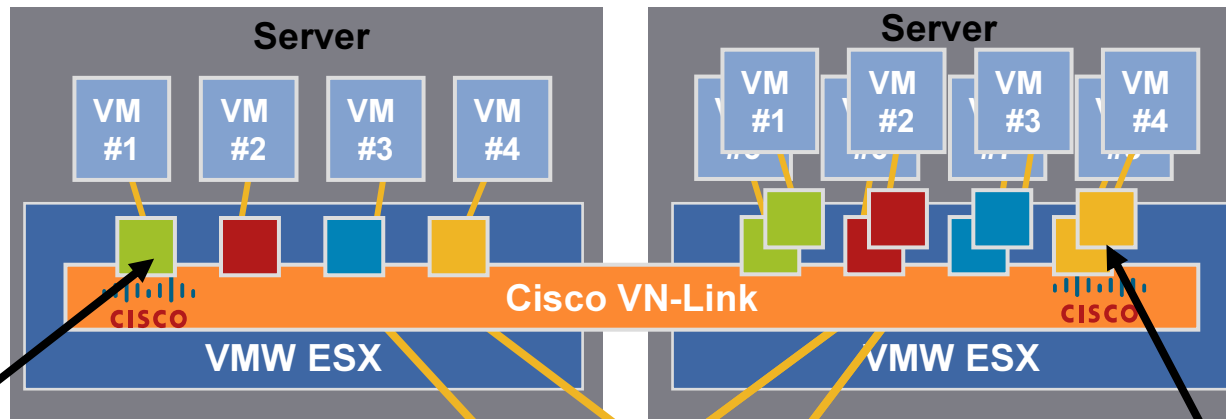
Hálózati szolgáltatások

Cisco VN-Link—Virtual Network Link

Policy-Based VM Connectivity

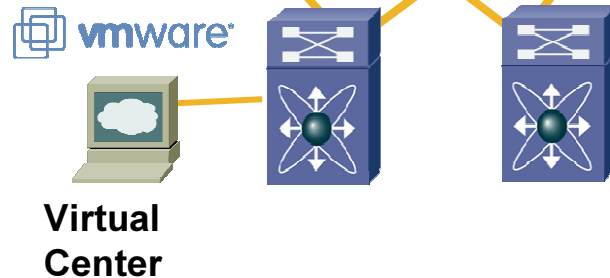
Mobility of Network & Security Properties

Non-Disruptive Operational Model



VMt át kell mozgatni

- VMotion
- DRS
- SW Upgrade/Patch
- Hardware hiba



VN-Link Mobilitás

- Vmotion hálózatra
- VM security biztosítása
- Connection táblát kezeli

Cisco VN-Link

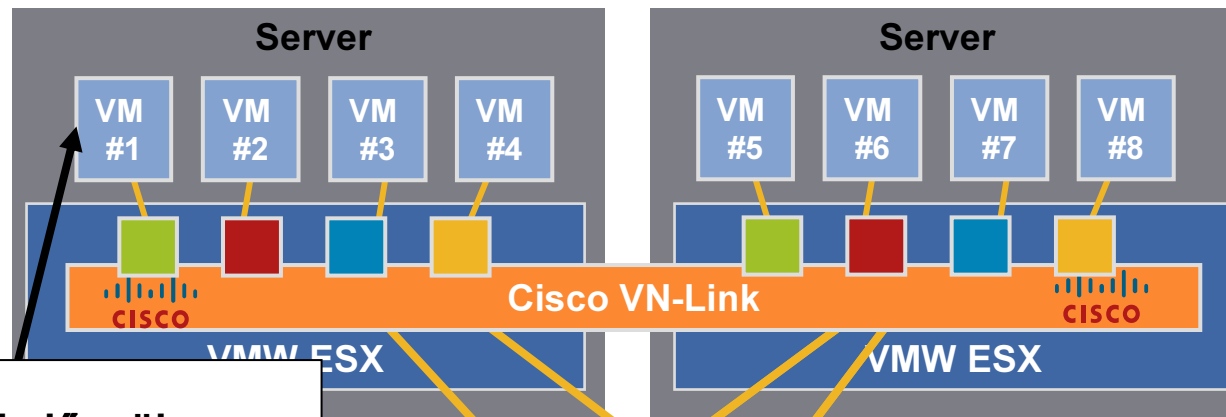
Működési hatékonyság

Cisco VN-Link—Virtual Network Link

Policy-Based
VM Connectivity

Mobility of Network
& Security Properties

Non-Disruptive
Operational Model



Szerver oldali előnyök

- VM mgmt változatlan
- Csökkenti az üzembeállítást
- Növeli a skálázhatóságot
- Növeli VM-szintű láthatóságot

Center

Hálózat oldali előnyök

- Egységes hálózat mgmt. és műk.
- Növeli a biztonságot
- VM hálózati funkciók növelése
- Perzisztens Policy
- Növeli a VM-szintű láthatóságot

Cisco Virtual Network Link – VN-Link Implementáció

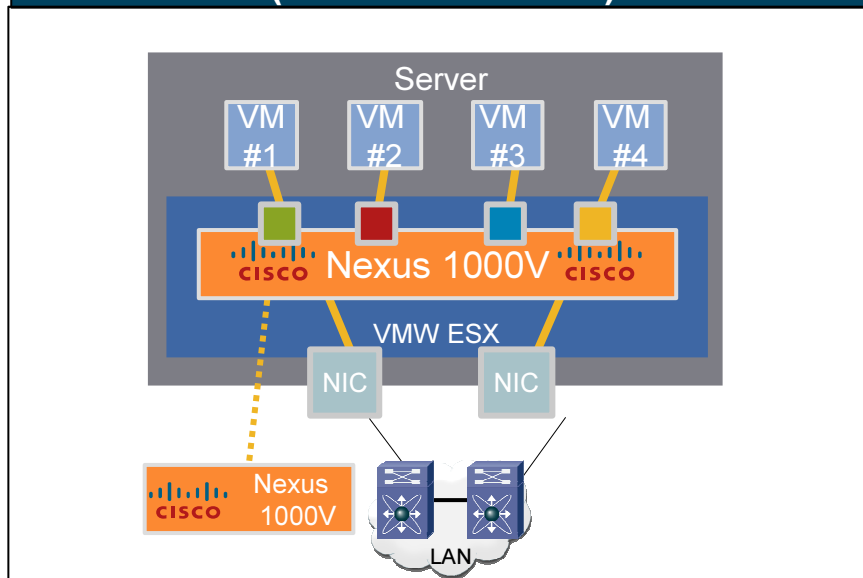
Virtualizing the Network Domain

Policy-Based
VM Connectivity

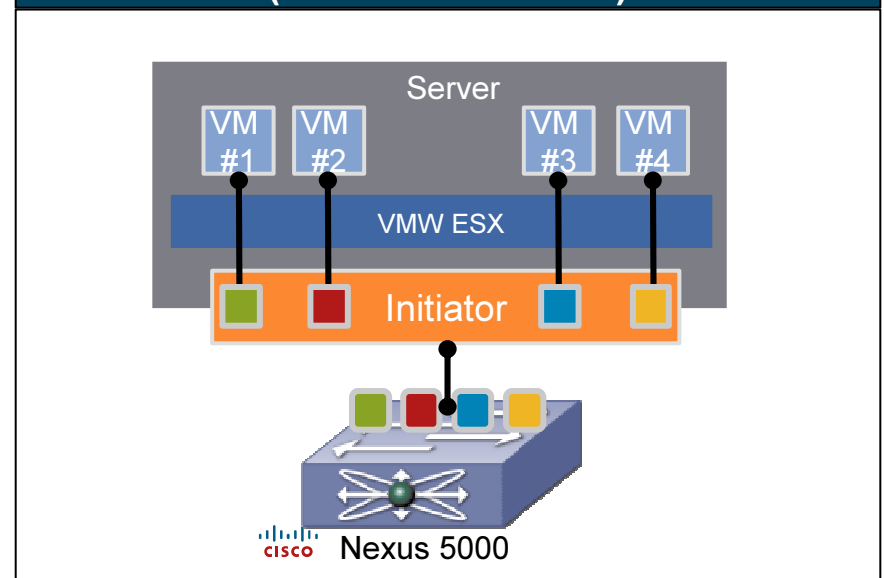
Mobility of Network
& Security Properties

Non-Disruptive
Operational Model

Cisco Nexus 1000V (Software Based)



Nexus 5000 with VN-Link (Hardware Based)

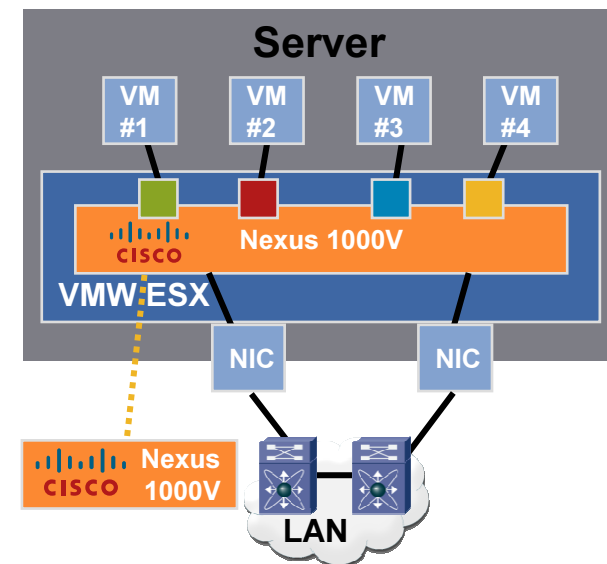


VN-Link a Cisco Nexus 1000V

Cisco Nexus 1000V Software alapú

- Az iparág első software alapú ESX switch eszköze
- Cisco NX-OS operációs rendszer alapú
- Teljes kompatibilitás a Cisco switchekhez
- Virtual Center alapú szolgáltatás mode változatlan server administráció
- Cisco NX-OS CLI alapú switch menedzsment a Nexus 1000V-re

BEST OF
vmworld 2008

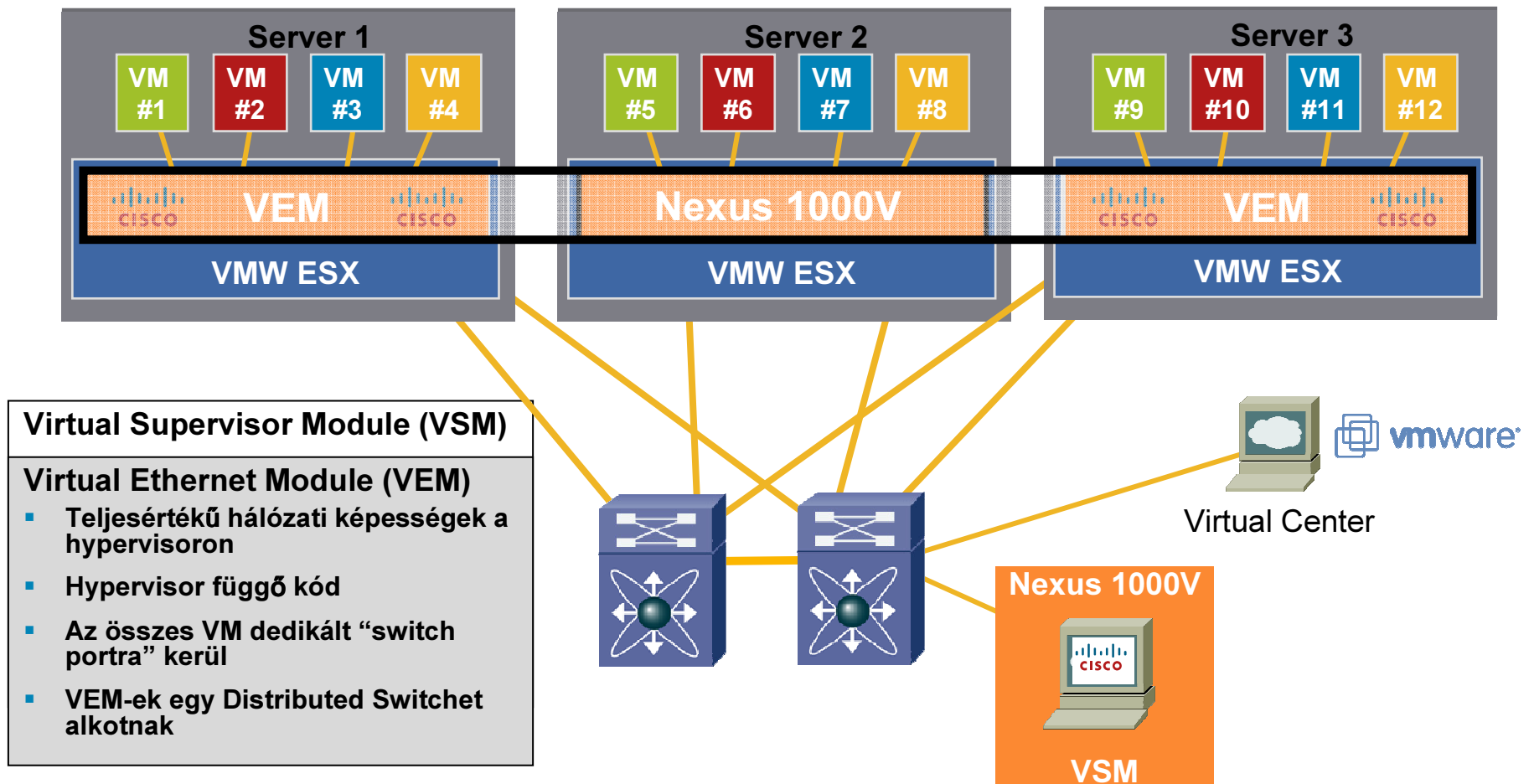


Policy-Based
VM Connectivity

Mobility of Network
and Security Properties

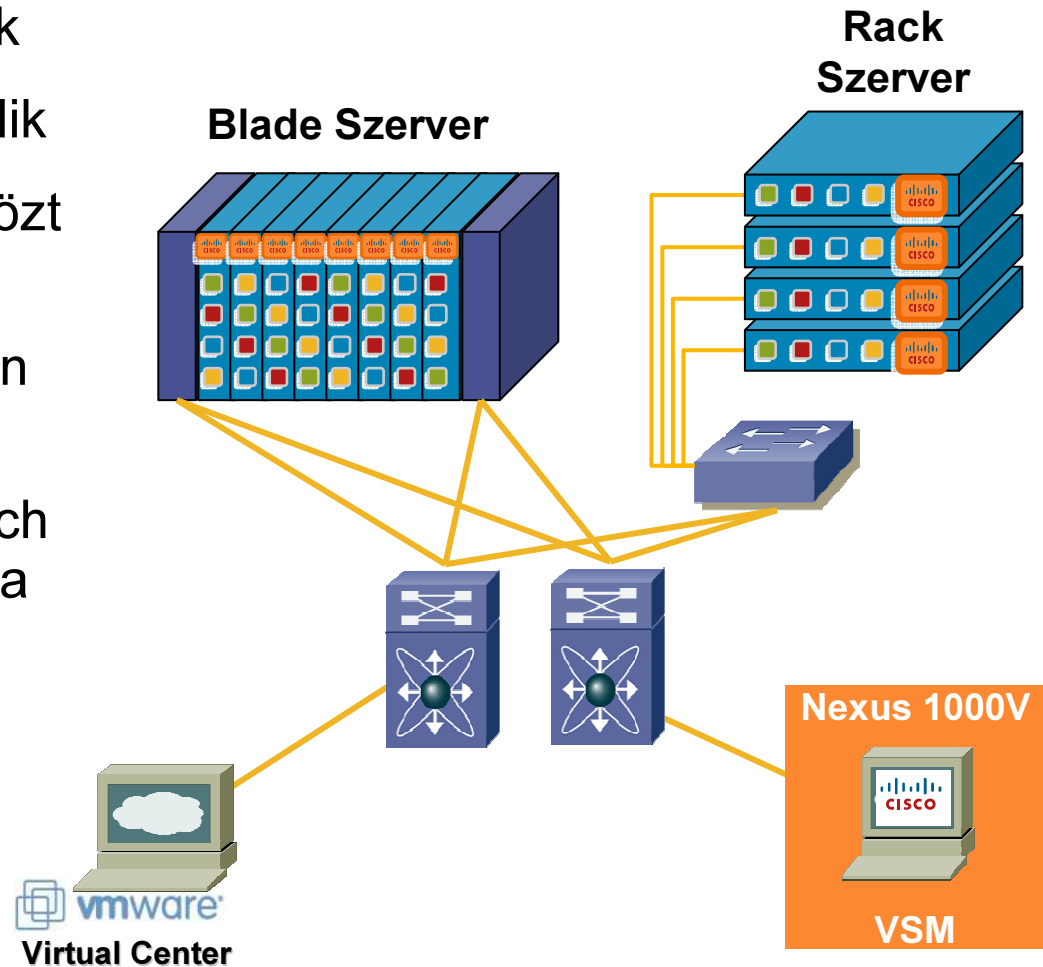
Non-Disruptive
Operational Model

Cisco Nexus 1000V Architektúra



VN-Link működés: Nexus 1000V

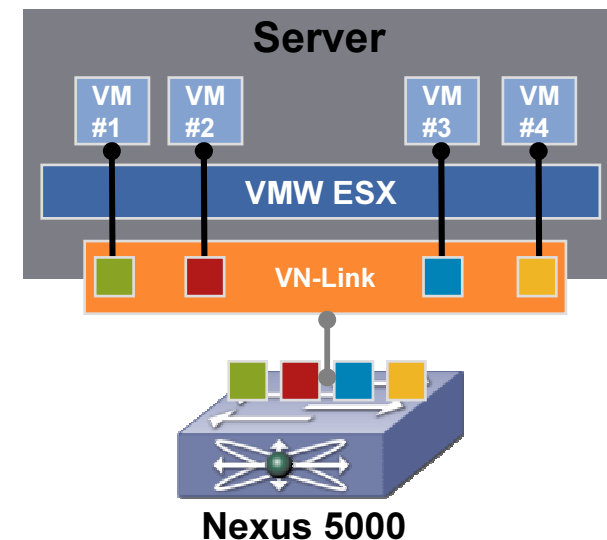
1. Minden szerverrel működik
2. Minden switch-csel működik
3. Külső Menedzsment eszközt igényel (VSM)
4. Cisco SW kell, hogy fusson a hypervisoron (VEM)
5. A VSM által beállított Switch jellemzők jutnak érvényre a VEM segítségével



VN-Link a Nexus 5000-en

Nexus Switch VN-Linkkel Hardware alapú

- Biztosítja a skálázható alapú implementálást a switcheken
- Szabványos megoldás: Cisco & VMware beadvány a IEEE 802 testülethez “Network Interface Virtualization”
- Egyesíti a VM & fizikai hálózati működést egyetlen menedzselhető eszközre



Policy-Based
VM Connectivity

Mobility of Network
and Security Properties

Non-Disruptive
Operational Model

Cisco Nexus 5000 VN-Link Architektúra

