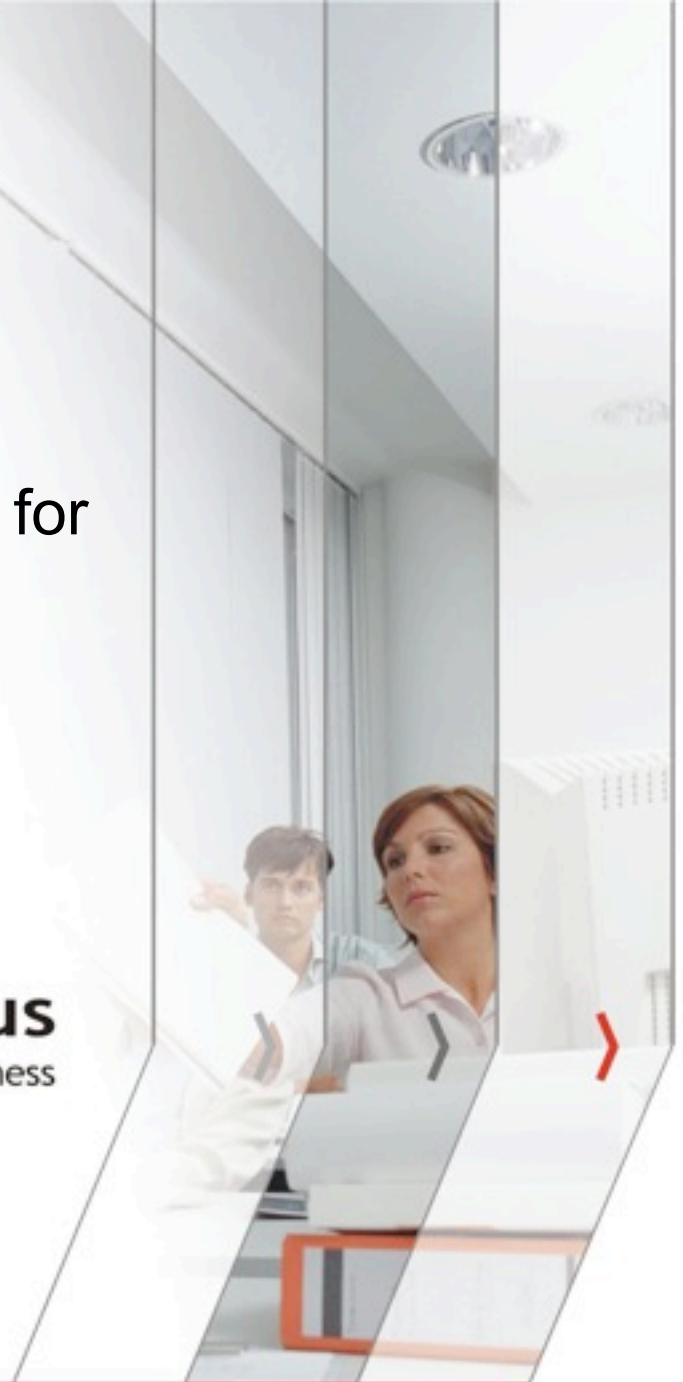


Best Practice Network Design for the Data Center

Mihai Dumitru, CCIE2 #16616



A Few Words about Cronus eBusiness:

- 39 employees, 3 national offices
- Focus on large enterprise customers from banking and retail, plus education
- Specializing in:
 - System integration (consulting, project management, network equipment sale and deployment, maintenance)
 - Managed services (operational support, network management, server hosting and business continuity)
- Cisco Gold Partner
 - One triple CCIE, one dual CCIE and more...
- Solarwinds Gold Partner

What We Will Cover In This Session:

- Classical Data Center Network Architecture
- Impact of new features and products on hierarchical design for data center networks
- Data center services insertion
- Layer 3 features and best practices
- Layer 2 features, enhancements and best practices

Hierarchical Design Network Layers:

Defining the Terms

- Data Center Core

Routed layer which is distinct from enterprise network core

Provides scalability to build multiple aggregation blocks

- Aggregation Layer

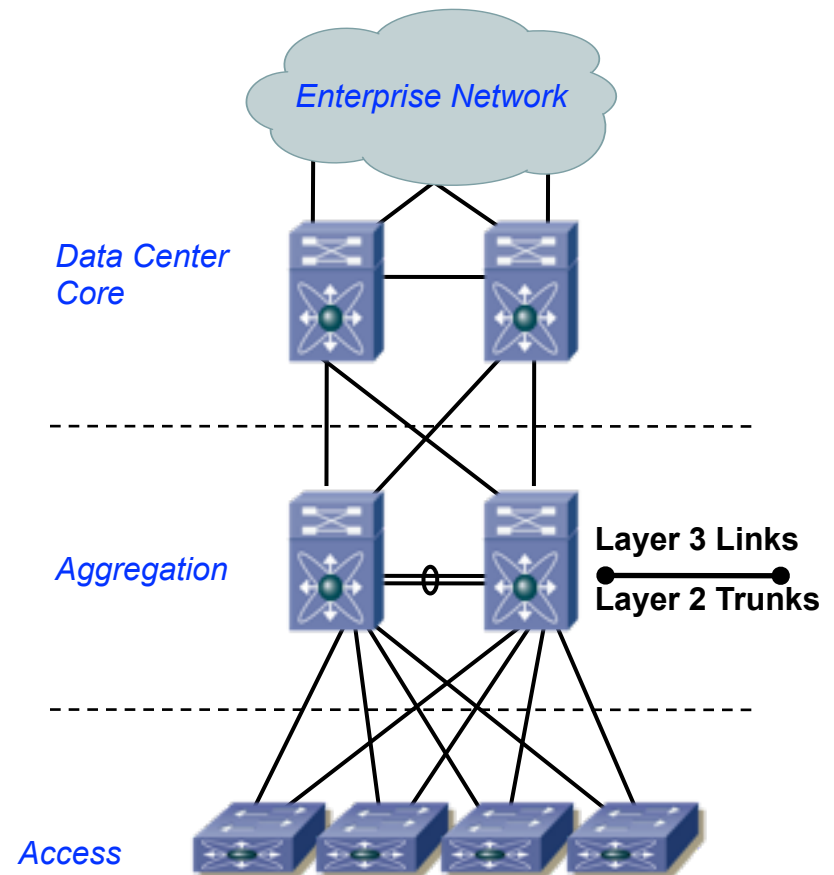
Provides the boundary between layer-3 routing and layer-2 switching

Point of connectivity for service devices (firewall, SLB, etc.)

- Access Layer

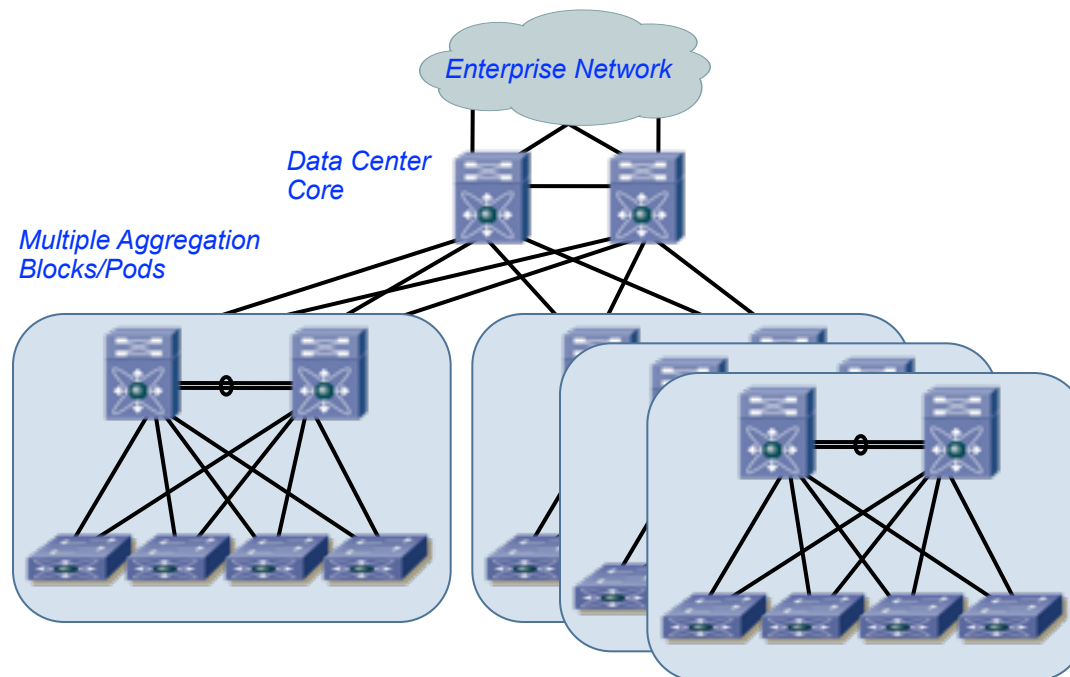
Provides point of connectivity for servers and shared resources

Typically layer-2 switching



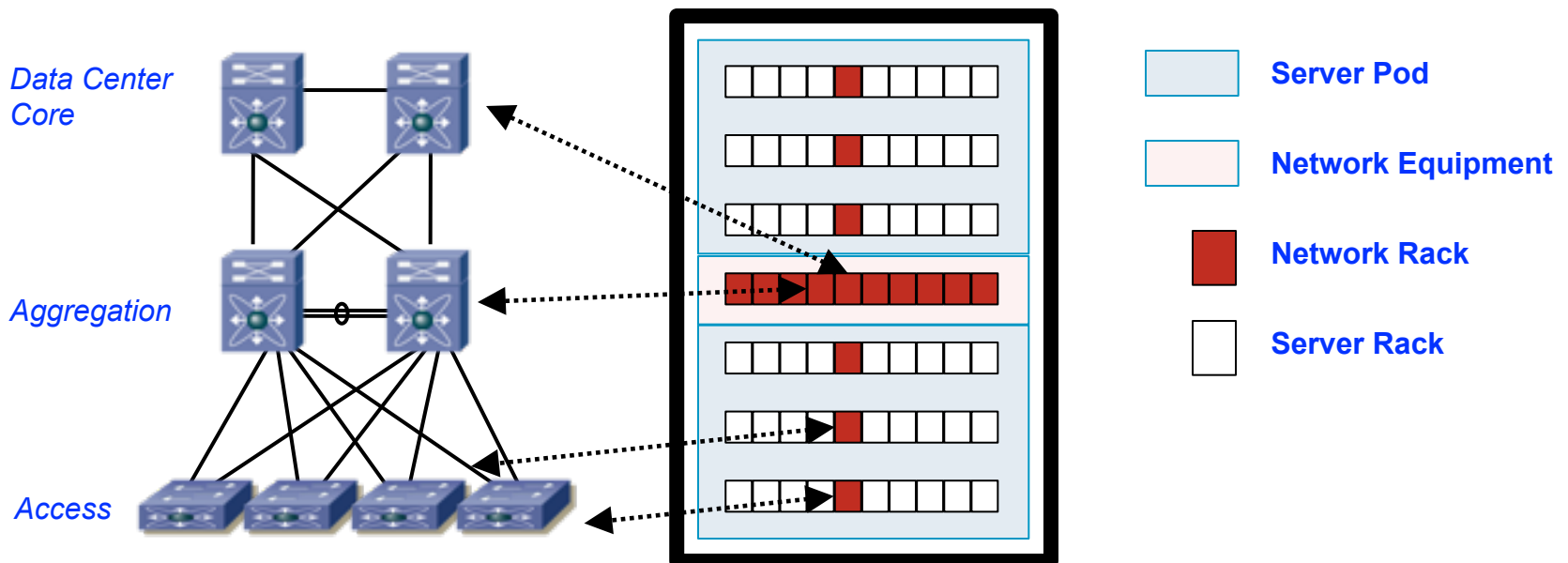
Scaling the Topology With a Dedicated Data Center Core

- A dedicated Data Center Core provides layer-3 insulation from the rest of the network
- Switch port density in the DC Core is reserved for scaling additional DC Aggregation blocks or pods
- Provides single point of DC route summarization



Mapping Network Topology to the Physical Design

- Design the Data Center topology in a consistent, modular fashion for ease of scalability, support, and troubleshooting
- Use a pod definition to map an aggregation block or other bounded unit of the network topology to a single pod
- The server access connectivity model can dictate port count requirements in the aggregation and affect the entire design



Traditional Data Center Server Access Models

- End-of-Row (EoR)

High density chassis switch at end or middle of a row of racks, fewer overall switches

Provides port scalability and local switching, may create cable management challenges

- Top-of-Rack (ToR)

Small fixed or modular switch at the top of each rack, more devices to manage

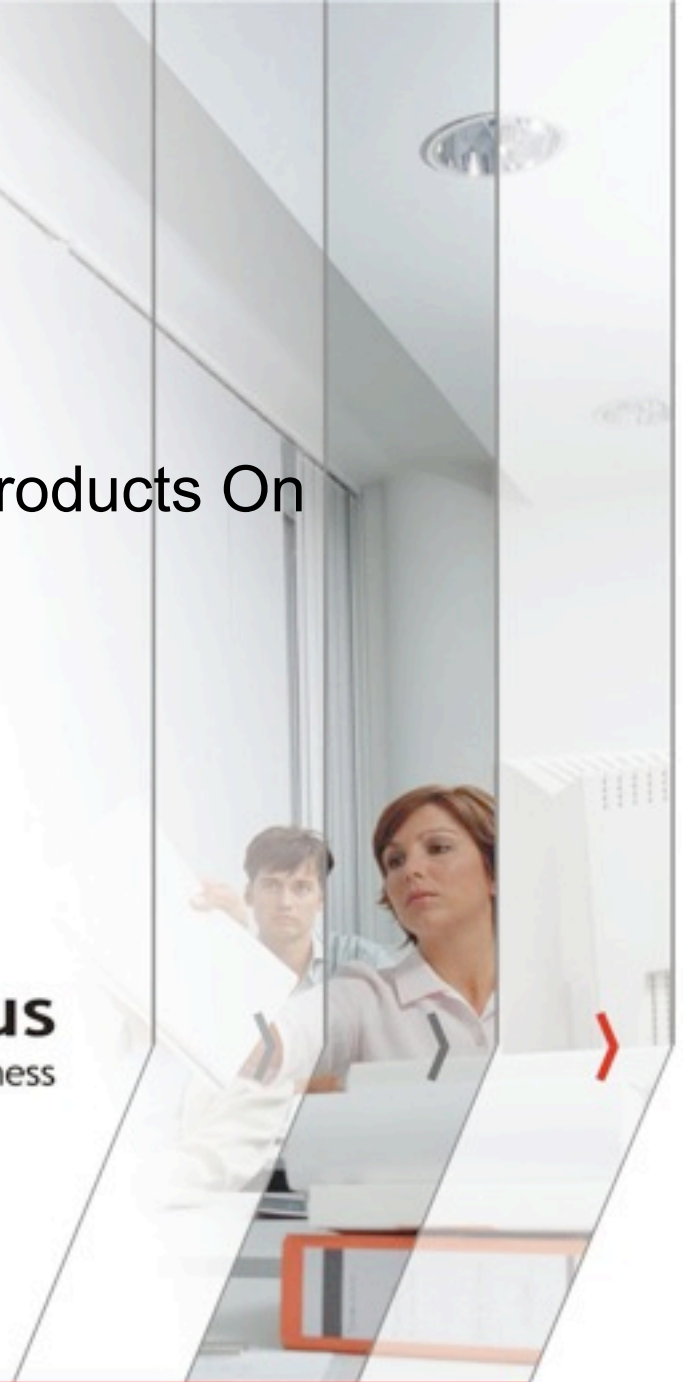
Significantly reduces bulk of cable by keeping connections local to rack or adjacent rack

- Integrated Switching

Switches integrated directly into blade server chassis enclosure

Maintaining feature consistency is critical to network management, sometimes pass-through modules are used

Impact of New Features and Products On Hierarchical Design for Data Center Networks



Building the Access Layer using Virtualized Switching

- Virtual Access Layer

Still a single logical tier of layer-2 switching

Common control plane with virtual hardware and software based I/O modules

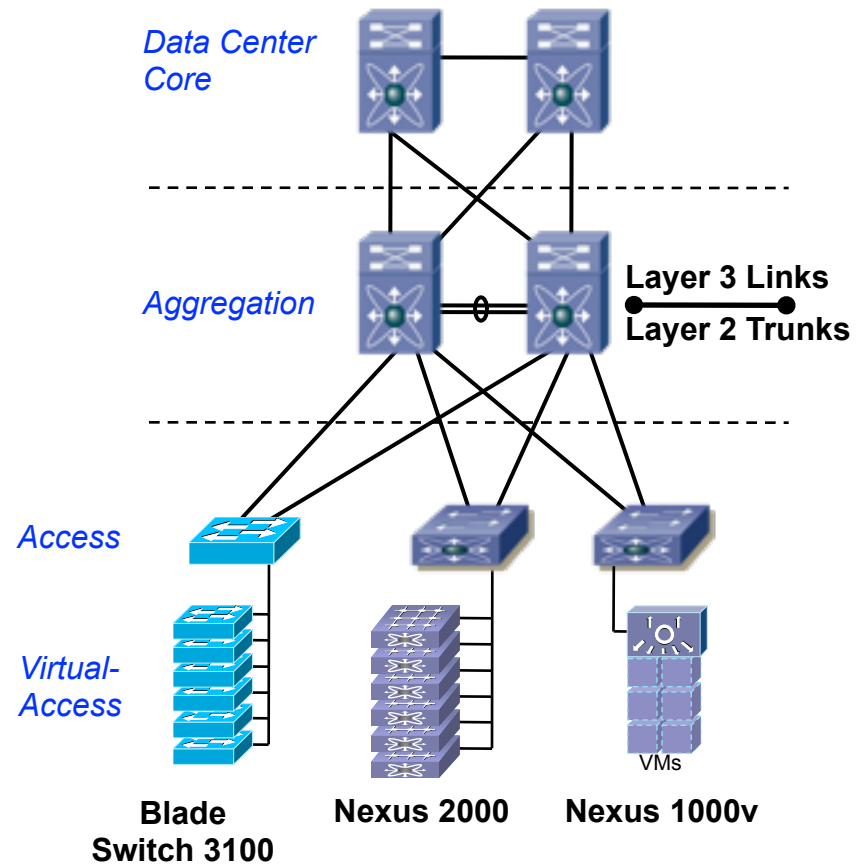
- Cisco Nexus 2000

Switching fabric extender module

Acts as a virtual I/O module supervised by Nexus 5000

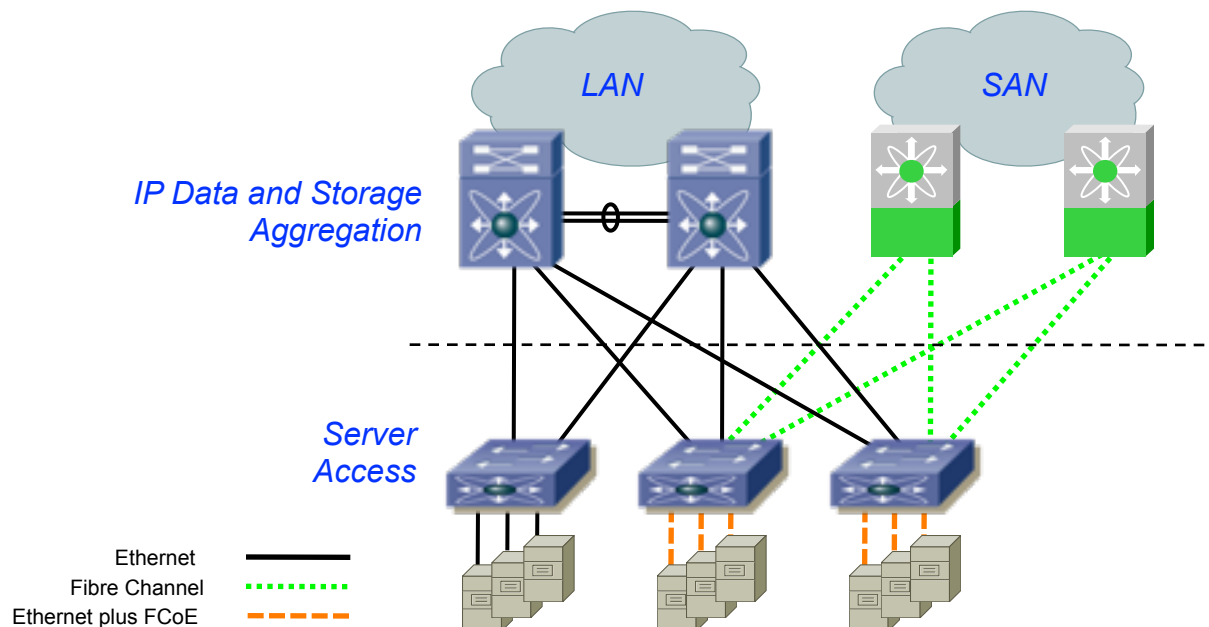
- Nexus 1000v

Software-based Virtual Distributed Switch for server virtualization environments.



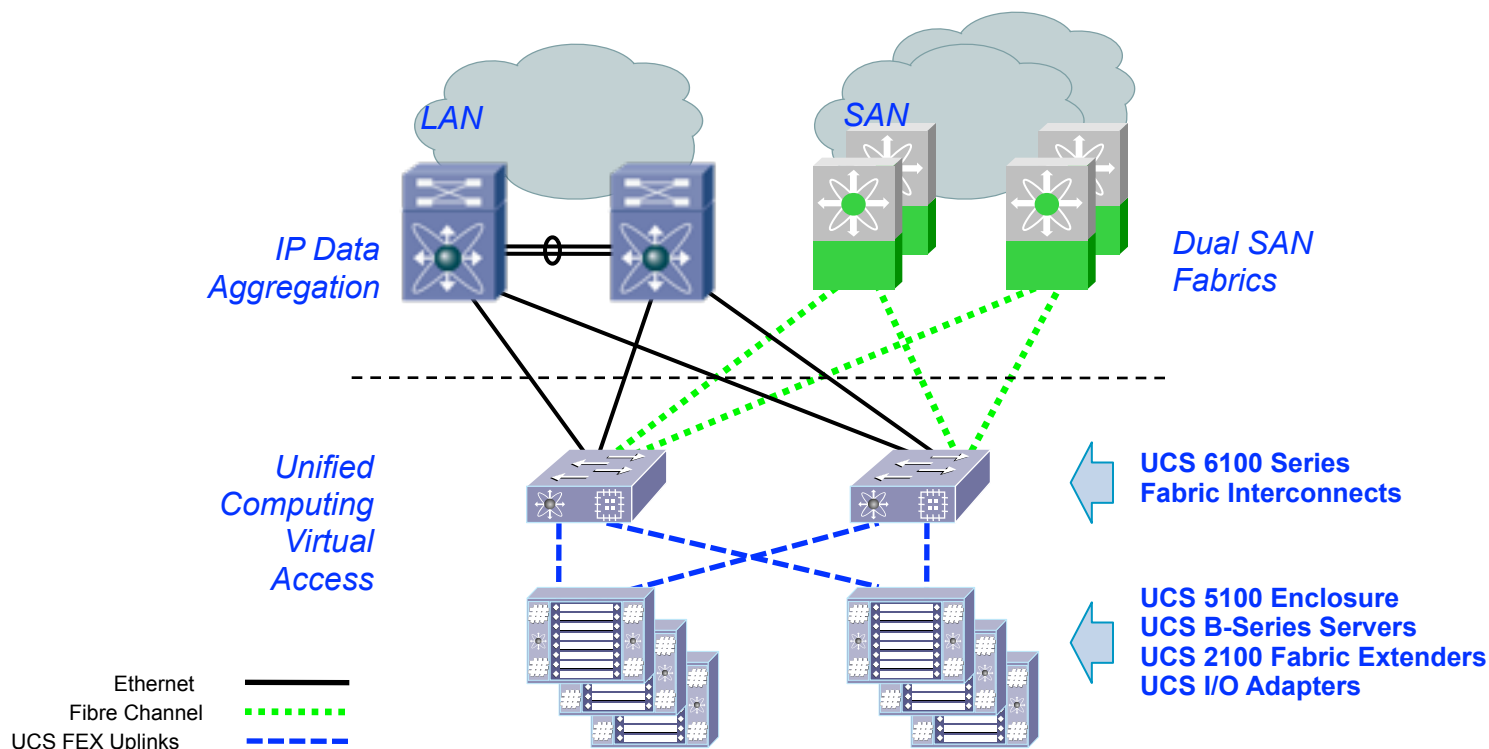
Migration to a Unified Fabric at the Access Supporting Data and Storage

- Nexus 5000 Series switches support integration of both IP data and Fibre Channel over Ethernet at the network edge
- FCoE traffic may be broken out on native Fibre Channel interfaces from the Nexus 5000 to connect to the Storage Area Network (SAN)
- Servers require Converged Network Adapters (CNAs) to consolidate this communication over one interface, saving on cabling and power



Cisco Unified Computing System (UCS)

- A cohesive system including a virtualized layer-2 access layer supporting unified fabric with central management and provisioning
- Optimized for greater flexibility and ease of rapid server deployment in a server virtualization environment
- From a topology perspective, similar to the Nexus 5000 and 2000 series



Nexus 7000 Series Virtual Device Contexts (VDCs)

- Virtualization of the Nexus 7000 Series Chassis

Up to 4 separate virtual switches from a single physical chassis with common supervisor module(s)

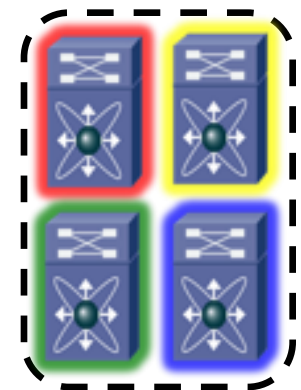
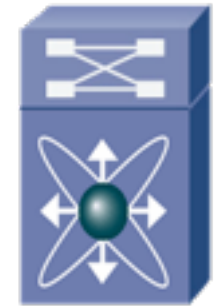
Separate control plane instances and management/CLI for each virtual switch

Interfaces only belong to one of the active VDCs in the chassis, external connectivity required to pass traffic between VDCs of the same switch

- Designing with VDCs

VDCs serve a “role” in the topology similar to a physical switch; core, aggregation, or access

Two VDCs from the same physical switch should not be used to build a redundant network layer – physical redundancy is more robust



Virtual Device Context Example:

Services VDC Sandwich

- Multiple VDCs used to “sandwich” services between switching layers

Allows services to remain transparent (layer-2) with routing provided by VDCs

Aggregation blocks only communicate through the core layer

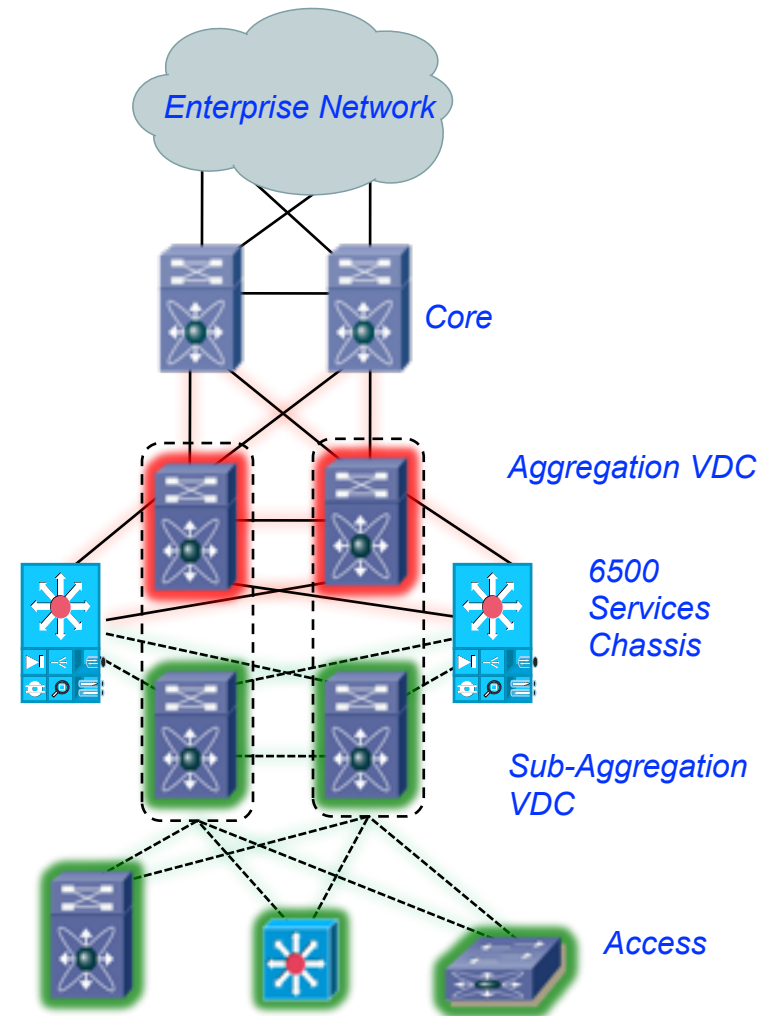
- Design considerations:

Access switches requiring services are connected to sub-aggregation VDC

Access switches not requiring services may be connected to aggregation VDC

Allows firewall implementations not to share interfaces for ingress and egress

Facilitates virtualized services by using multiple VRF instances in the sub-aggregation VDC



Data Center Service Insertion



Data Center Service Insertion:

Direct Services Appliances

- Appliances directly connected to the aggregation switches

Service device type and Routed or Transparent mode can affect physical cabling and traffic flows.

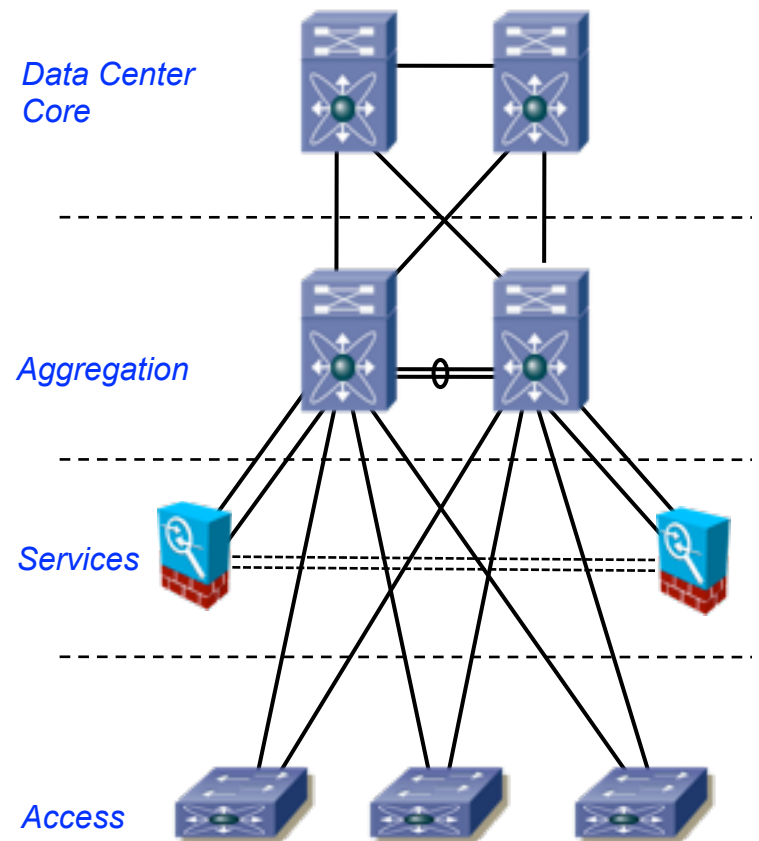
- Transparent mode
ASA example:

Each ASA dependant on one aggregation switch

Separate links for fault tolerance and state traffic either run through aggregation or directly

Dual-homed with interface redundancy feature is an option

Currently no EtherChannel supported on ASA



Data Center Service Insertion:

External Services Chassis

- Dual-homed Catalyst 6500

Services do not depend on a single aggregation switch

Direct link between chassis for fault-tolerance traffic, may alternatively trunk these VLANs through Aggregation

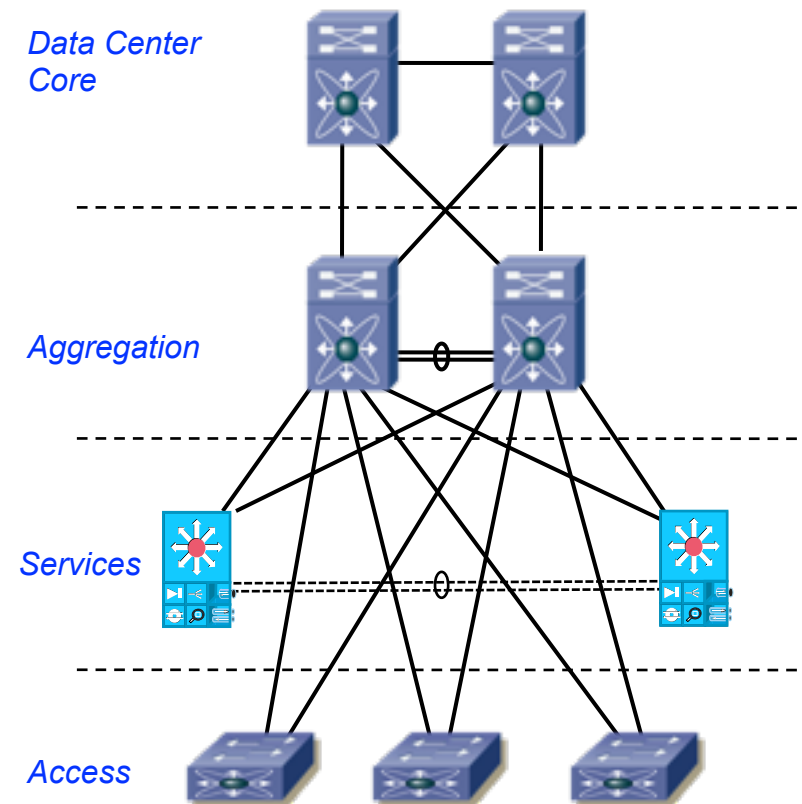
- Dedicated integration point for multiple data center service devices

Provides slot real estate for 6500 services modules

Firewall Services Module (FWSM)

Application Control Engine (ACE) Module

Other services modules, also beneficial for appliances



Using Virtualization and Service Insertion to Build Logical Topologies

- Logical topology example using services VDC sandwich physical model

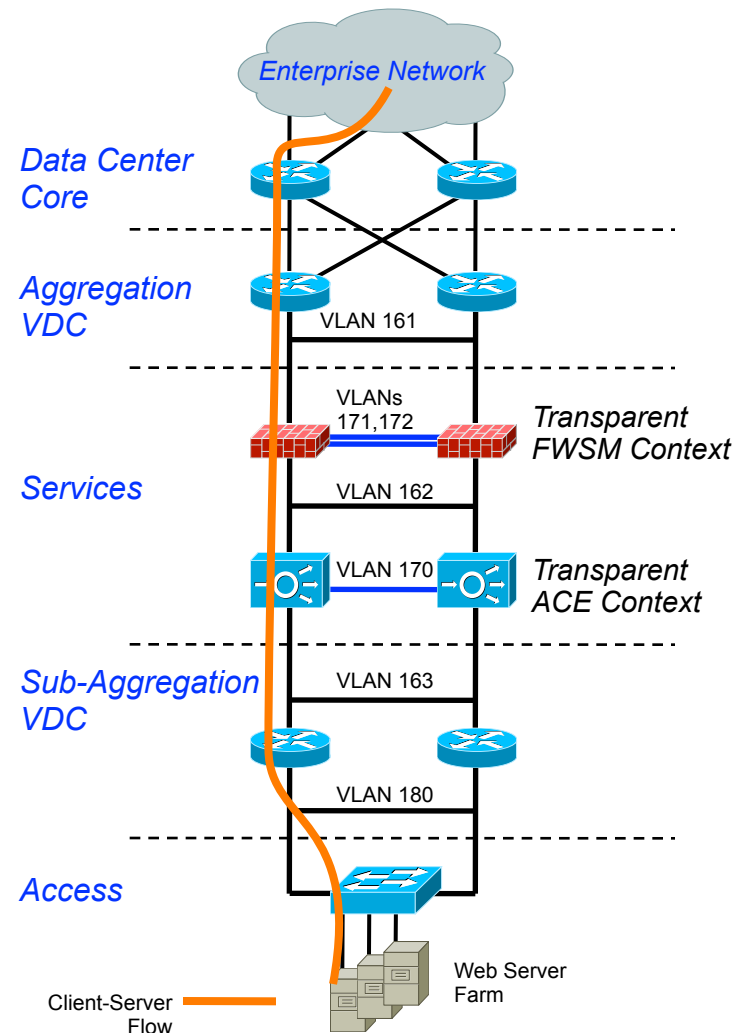
Layer-2 only services chassis with transparent service contexts

VLANs above, below, and between service modules are a single IP subnet

Sub-aggregation VDC is a layer-3 hop running HSRP providing default gateway to server farm subnets

Multiple server farm VLANs can be served by a single set of VLANs through the services modules

Traffic between server VLANs does not need to transit services device, but may be directed through services using virtualization



Using Virtualization and Service Insertion to Build Logical Topologies

- Logical Topology to support multi-tier application traffic flow

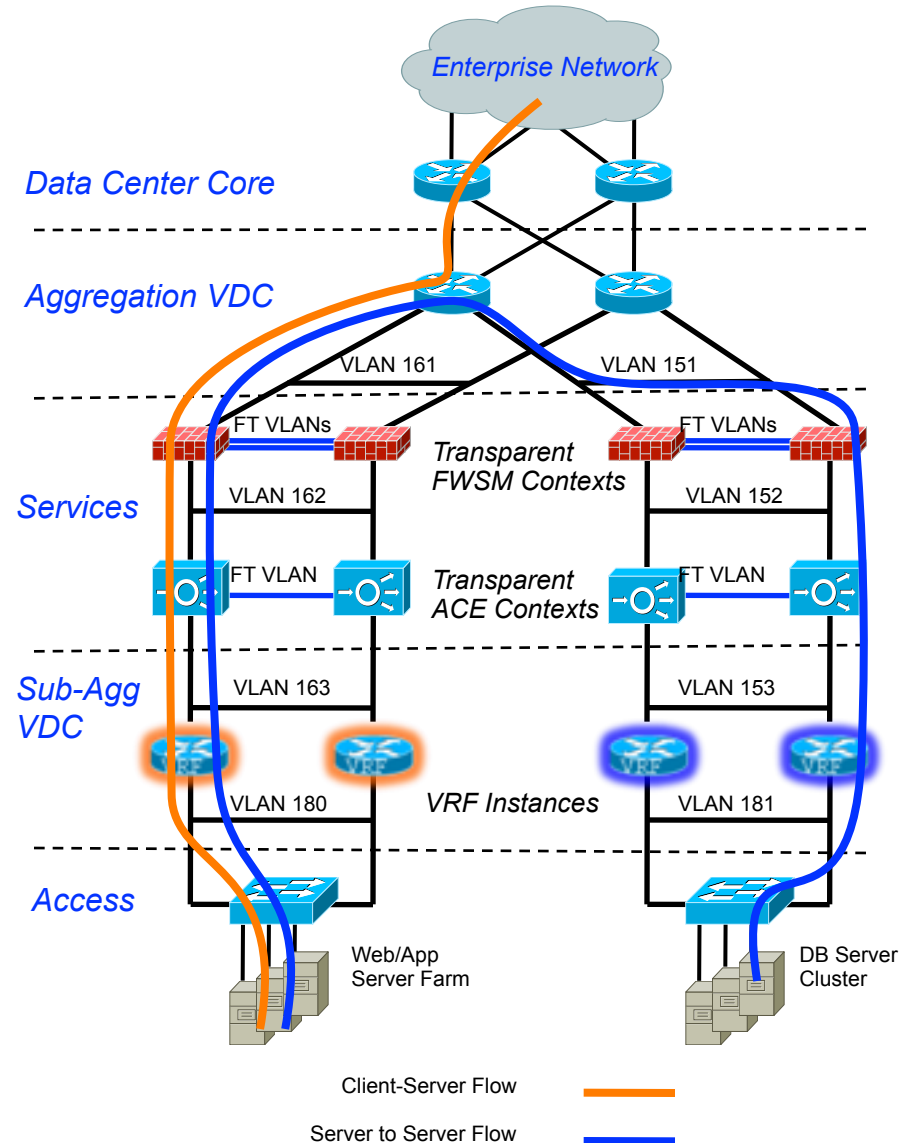
Same physical VDC services chassis sandwich model

Addition of multiple virtual contexts to the transparent services modules

Addition of VRF routing instances within the sub-aggregation VDC

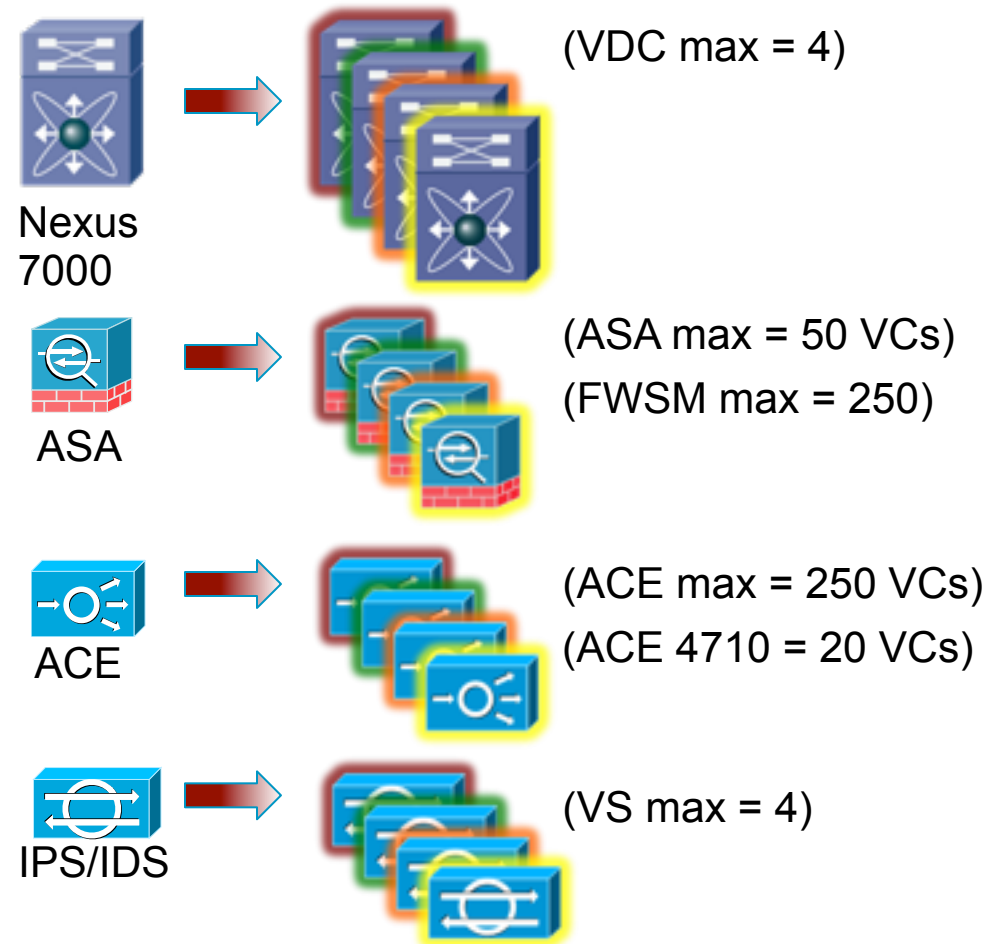
Service module contexts and VRFs are linked together by VLANs to form logical traffic paths

Example Web/App server farm and Database server cluster homed to separate VRFs to direct traffic through the services

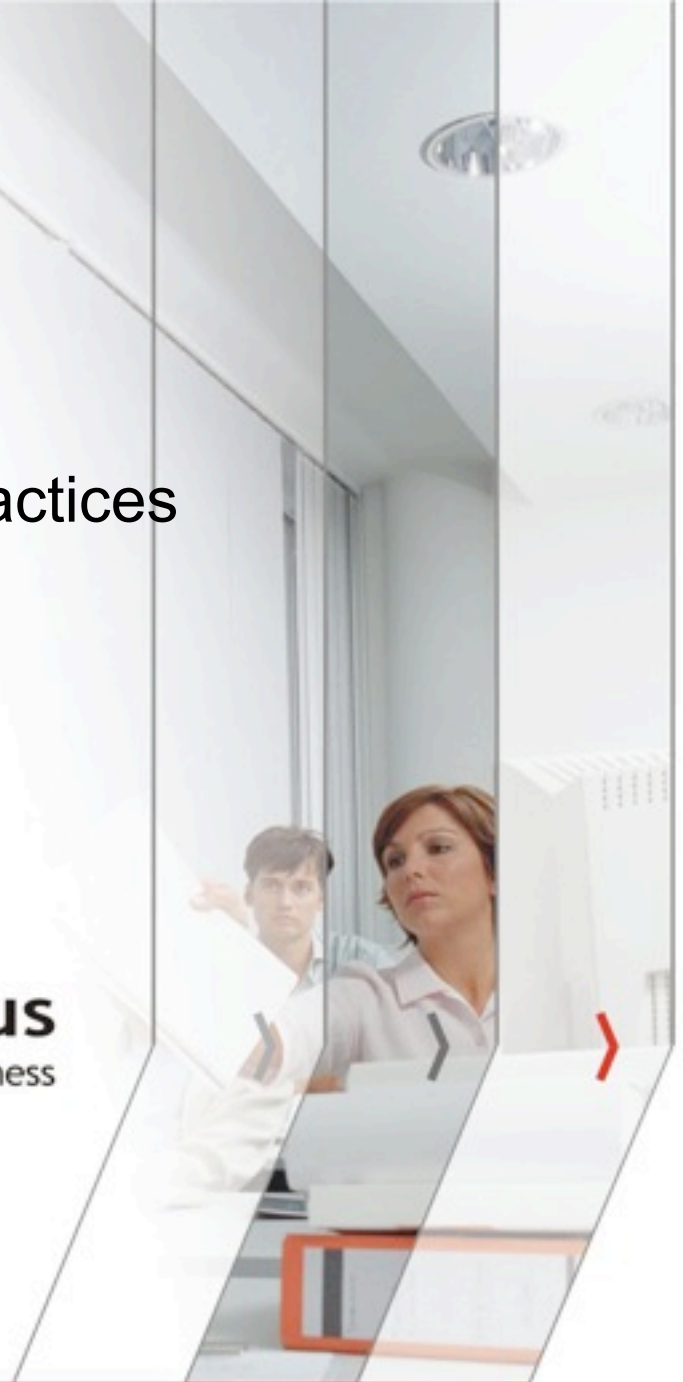


Active-Active Solution Virtual Components

- Nexus 7000
VDCs, VRFs, SVIs
- ASA 5580
Virtual Contexts
- ACE Service Module
Virtual Contexts, Virtual IPs (VIPs)
- IPS 4270
Virtual Sensors
- Virtual Access Layer
Virtual Switching System
Nexus 1000v
Virtual Blade Switching

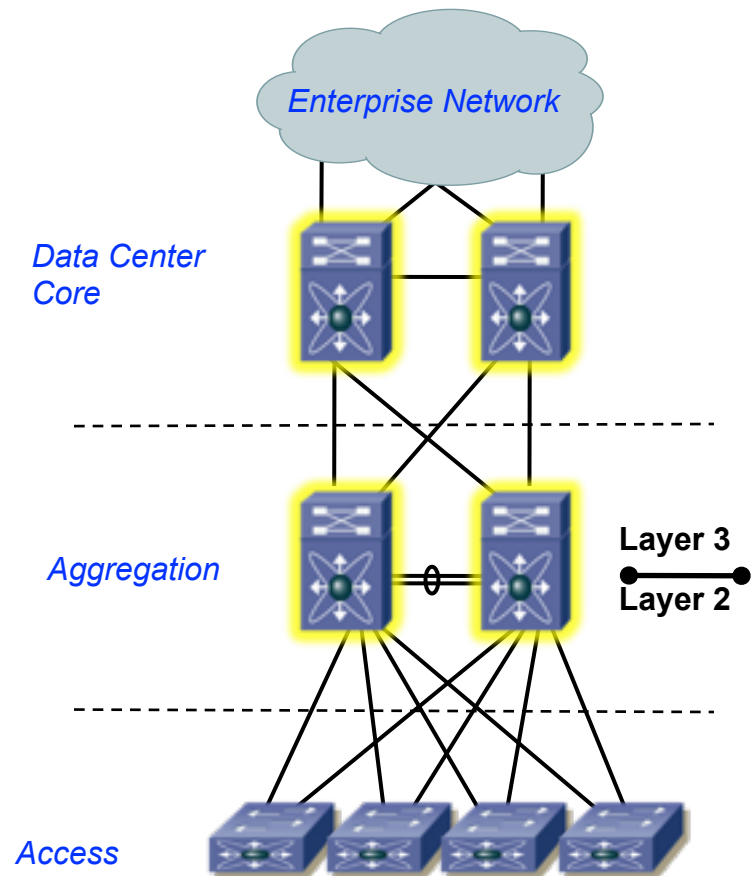


Layer 3 Features and Best Practices



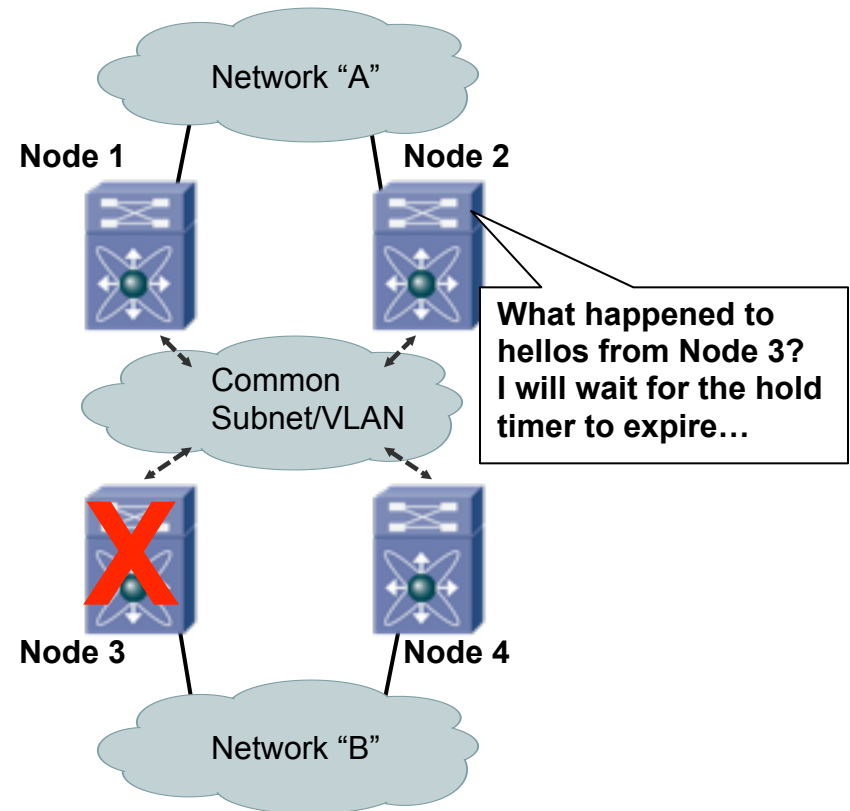
Layer-3 Feature Configuration in the Data Center

- Summarize IP routes at the DC Aggregation or Core to advertise fewer destinations to the enterprise core
- Avoid IGP peering of aggregation switches through the access layer by setting VLAN interfaces as passive
- Use routing protocol authentication to help prevent unintended peering
- If using OSPF, set consistent reference bandwidth at 10,000 or higher for support of 10 Gigabit Ethernet
- HSRP timers at hello-1 dead-3 provide a balance of fast failover without too much sensitivity to control plane load



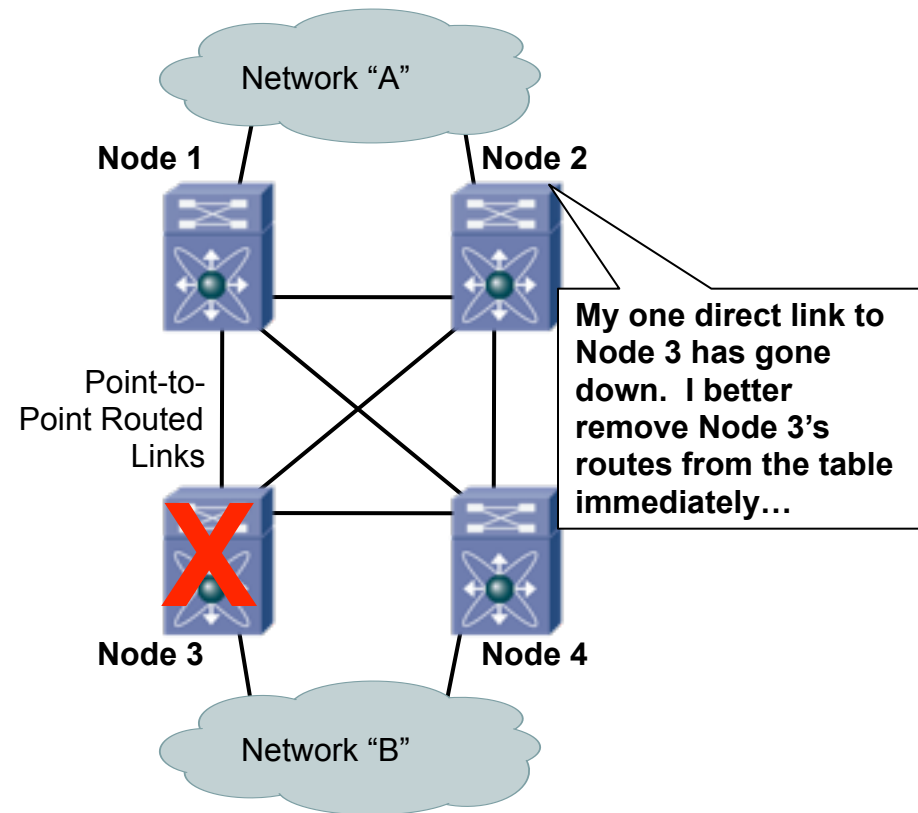
IGP Hello and Dead/Hold Timers Behavior Over Shared Layer 2 Domain

- Routing protocols insert destinations into the routing table and maintain peer state based on receipt of continuous Hello packets.
- Upon device or link failure, routing protocol only removes the failed peer's routes only after Dead/Hold timer is expired.
- Tuning Dead/Hold timers lower provides faster convergence over this type of topology.
- A firewall module running an IGP is an example of a Data Center device peering over a L2 domain, or any VLAN interface (SVI).



IGP Hello and Dead/Hold Timers Behavior Over Layer-3 Links

- Upon device or link failure, routing protocol immediately removes routes from failed peer based on interface down state.
- Tuning the IGP Hello and Dead/Hold timers lower is not required for convergence due to link or device failure.
- Transparent-mode services or using static routing with HSRP can help ensure all failover cases are based on point-to-point links.
- Note that static routing with HSRP is not a supported approach for IP multicast traffic.



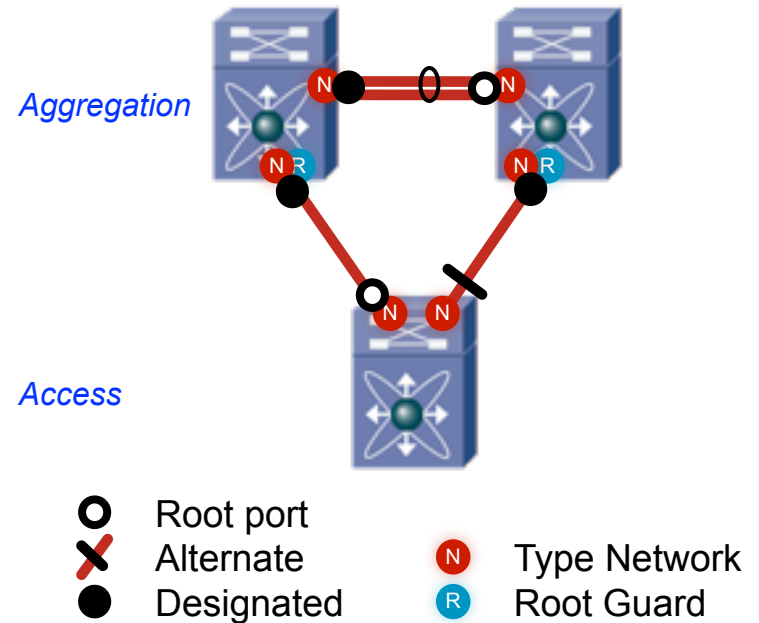
Layer 2 Features, Enhancements and Best Practices



Classic Spanning Tree Topology

“Looped Triangle” Access

- Layer-2 protocols are designed to be plug-and-play, and forward traffic without configuration
- Stability is enhanced by controlling the location of the STP root switch, and using consistent topologies
- Looped topologies are required to provide link redundancy and server mobility across access switches
- Using STP to break the network loop reduces available bandwidth in a VLAN due to blocked links
- Most STP issues result from undesired flooding due to link issues or software problems causing loss of BPDUs



Spanning Tree Configuration Features:

Rootguard, Loopguard, Portfast, BPDUguard

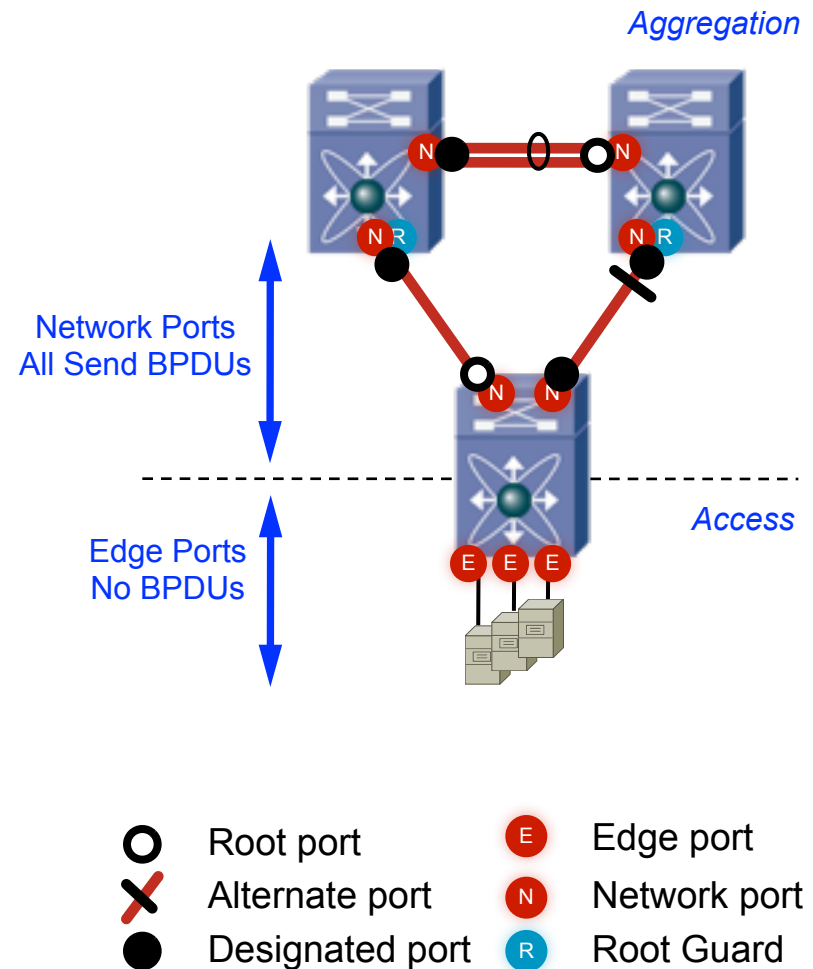
These Features Allow STP to Behave with More Intelligence, but Require Manual Configuration:

- Rootguard prevents a port from accepting a better path to root where this information should not be received
- Loopguard restricts the transition of a port to a designated forwarding role without receiving a BPDU with an inferior path to root
- Port fast (Edge Port) allows STP to skip the listening and learning stages on ports connected to end hosts
- BPDUguard shuts down a port that receives a BPDU where none should be found, typically also used on ports facing end hosts

Updated STP Features:

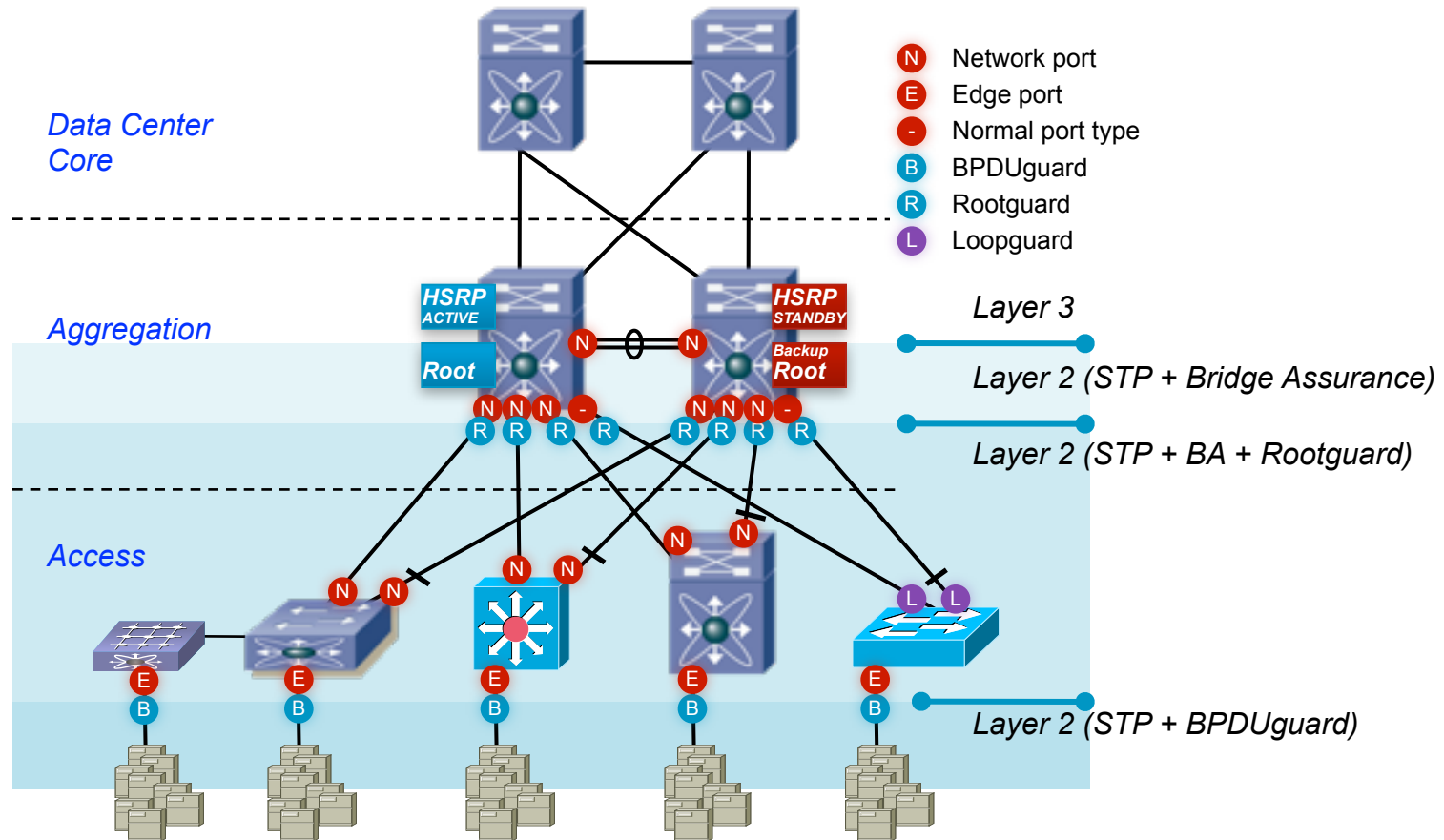
Bridge Assurance

- Specifies transmission of BPDUs on all ports of type “network”.
- Protects against unidirectional links and peer switch software issues (LoopGuard can only be enabled on root and alternate ports)
- Requires configuration, best practice is to set global default to type “network”, default is “normal”
- IOS Example:
spanning-tree portfast network default
- Caution: Both ends of the link must have Bridge Assurance enabled (otherwise the port is blocked)



STP Configuration Feature Placement In the Data Center

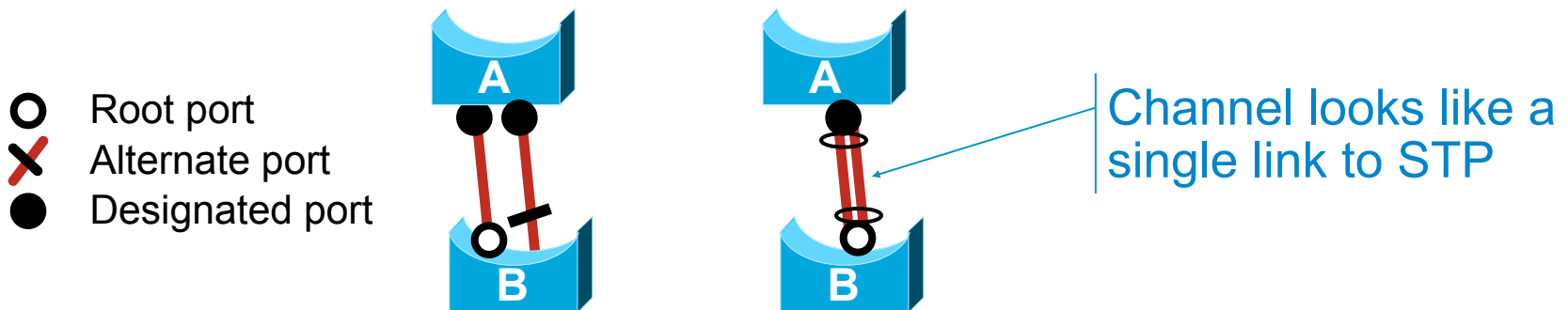
Bridge Assurance Replace the Requirement For Loopguard On Supported Switches



Redundant Paths Without STP Blocking:

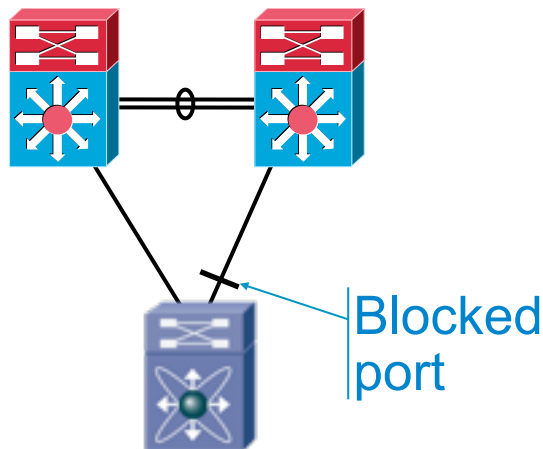
Basic EtherChannel

- Bundles several physical links into a logical one
 - No blocked ports (redundancy not handled by STP)
 - Per frame (not per-vlan) load balancing
- Control protocols like PAgP (Port Aggregation Protocol) and LACP (Link Aggregation Control Protocol) handle the bundling process and monitor the health of the link
- Limited to parallel links between two switches

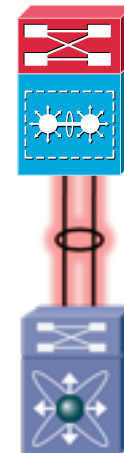


Designs Not Relying on STP: Virtual Switching System (VSS)

- Merges two bridges into one, allowing Multi-Chassis EtherChannels
- Also merges Layer-3 and overall switch management
- Does not rely on STP for redundancy
- Limited to pair of switches

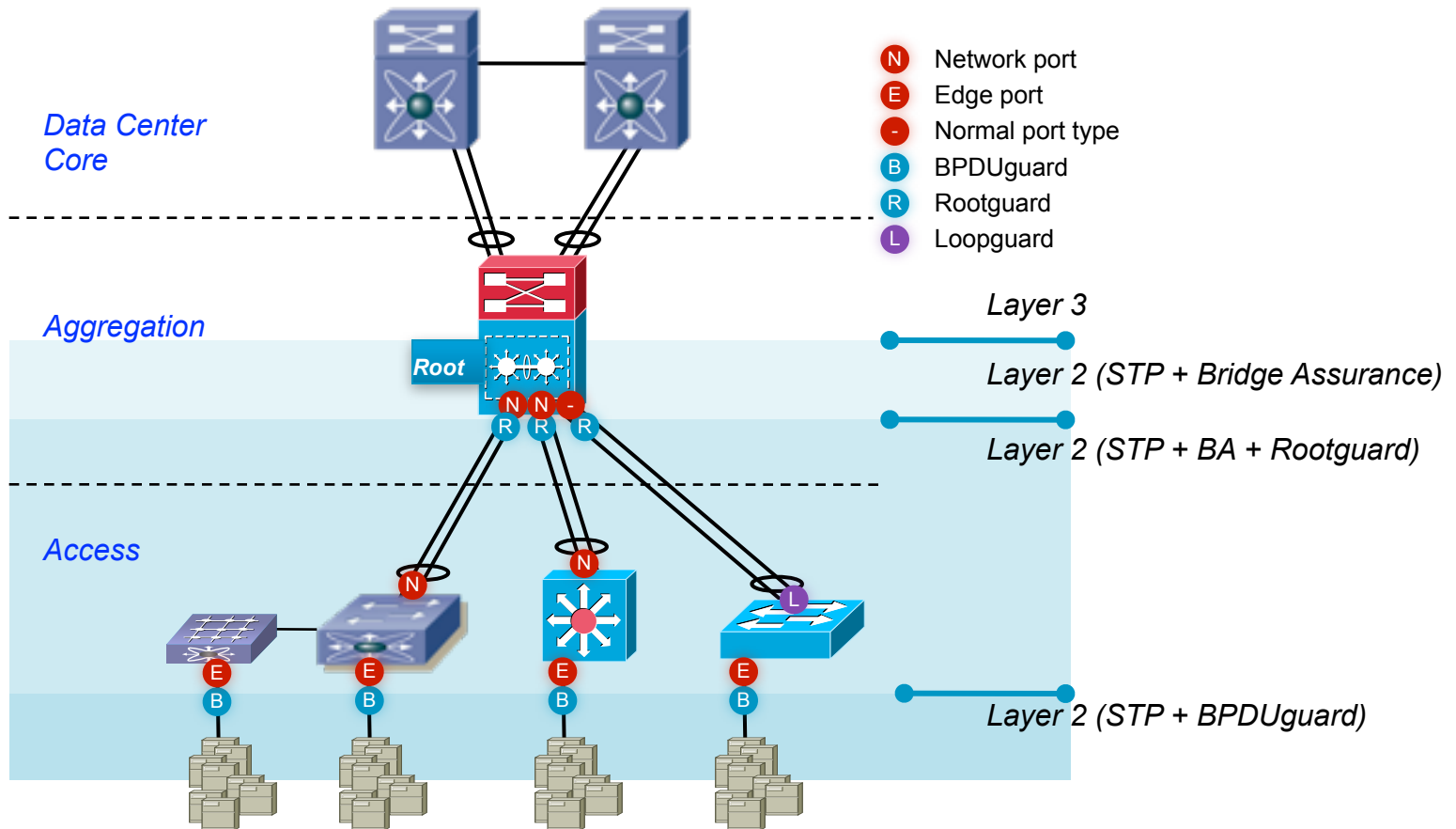


**Redundancy
handled by STP**



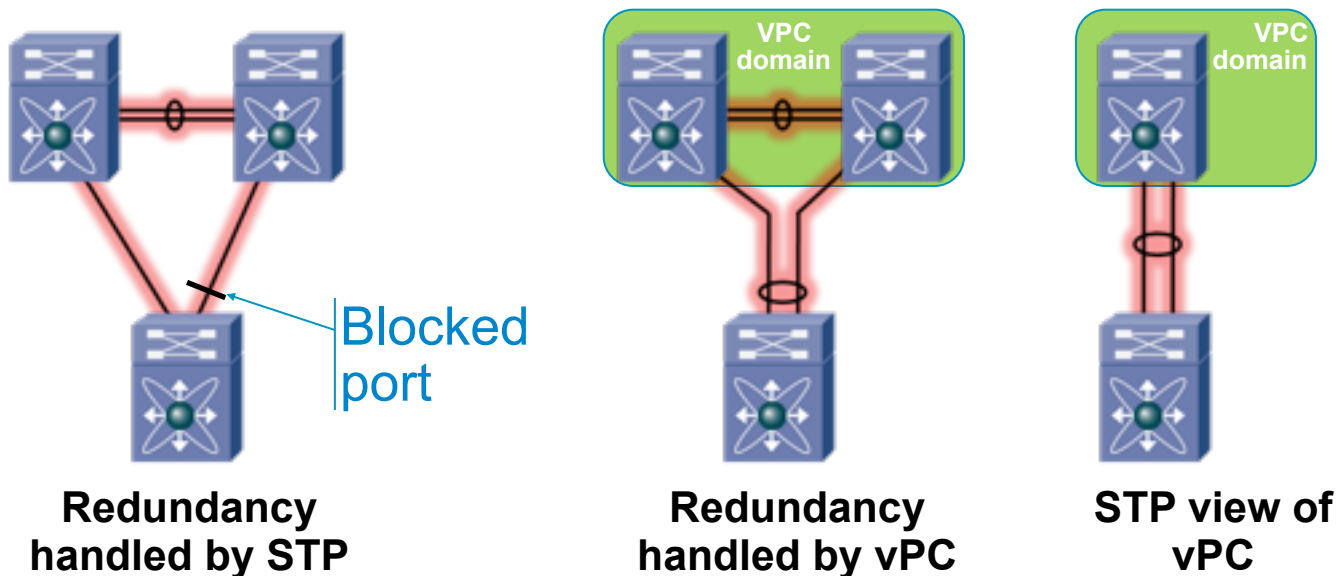
**Multi Chassis EtherChannel
(STP logical view)**

VSS Design

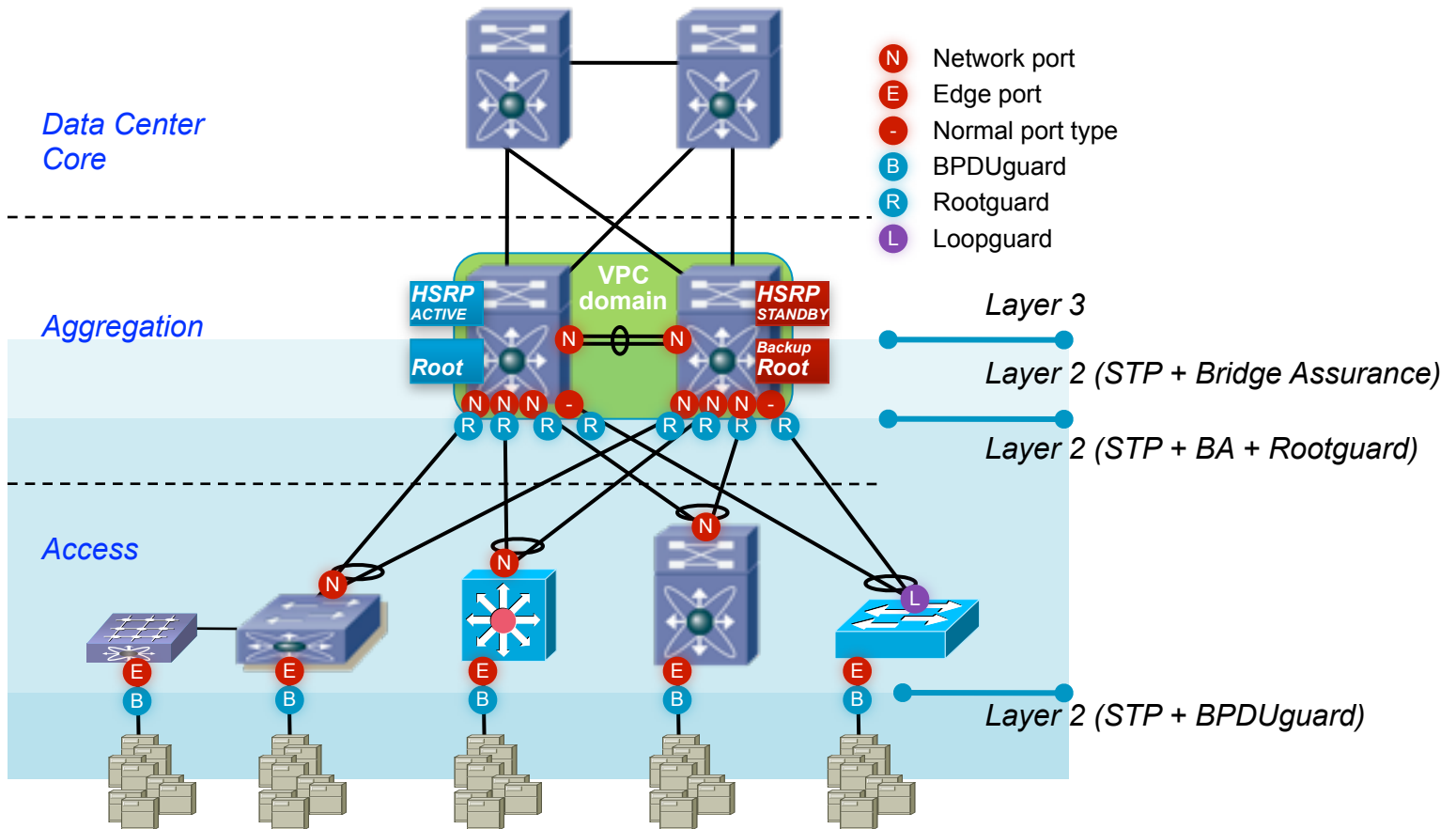


Designs Not Relying on STP: Virtual Port Channel (vPC)

- Appears as a single EtherChannel to the access layer
- Two independent control planes
- Active/active HSRP, separate Layer-3 and management
- Still no STP blocked ports



vPC Design



Summary

- Virtualization of the network infrastructure improves utilization and offers new deployment models in the data center
- Best practice, flexible models are needed for application requirements
- Case studies:
 - BRD - Groupe Société Générale (starting with 280 servers in 2007 at the new data center)
 - MOBIASBANCA - Groupe Société Générale
 - Carrefour Romania - Hiproma



65A Aron Cotruș St.
014131 Bucharest 1
Phone: +4(021) 204 3636
Fax: +4(021) 204 3635

www.cronus.ro
office@cronus.ro

