Data Center Front End Architecture
Solution for Business Continuance

Yves LOUIS

Cisco Networkers 2007

HOUSEKEEPING

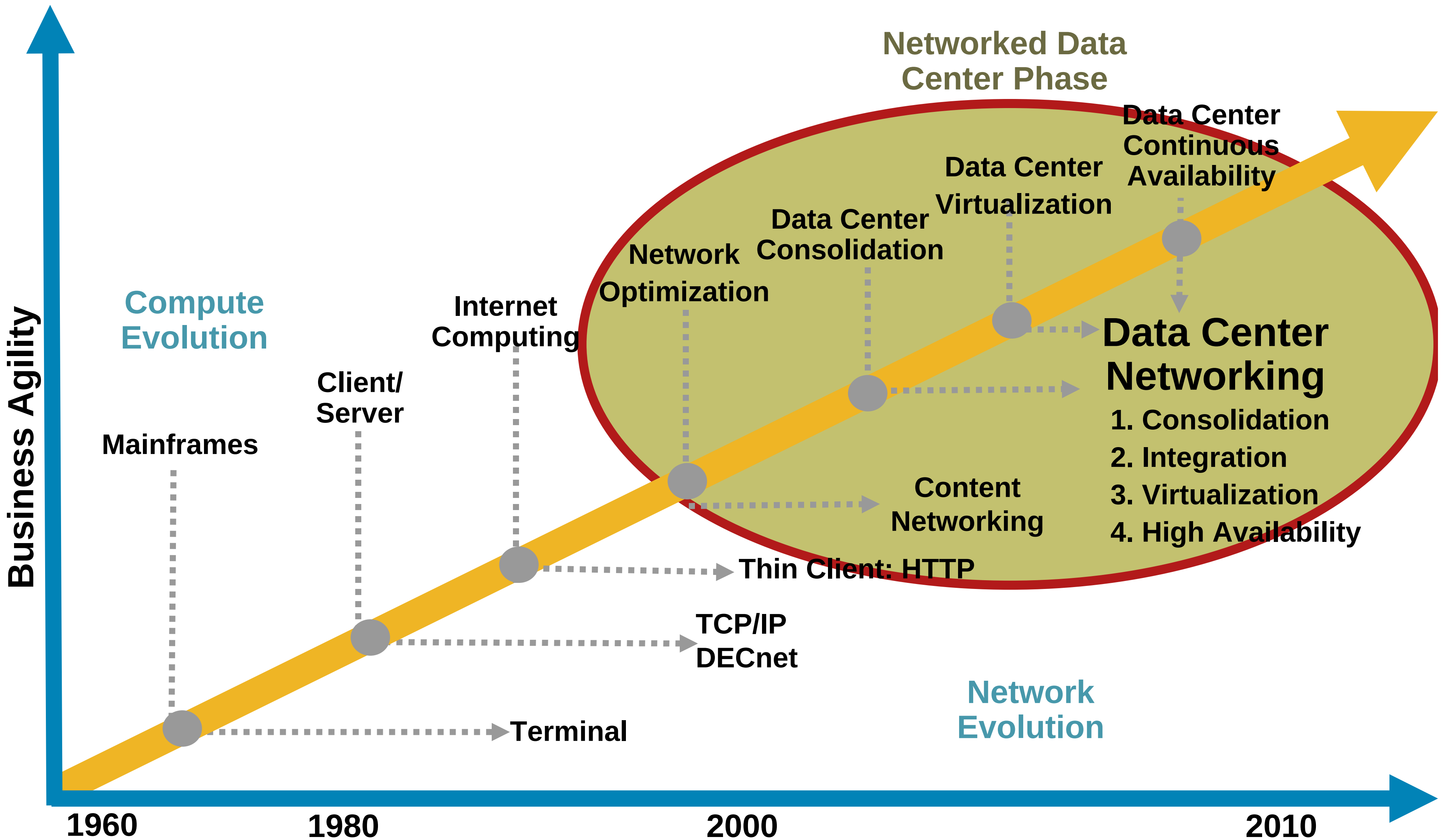
- We value your feedback, don’t forget to complete your online session evaluations after each session and complete the Overall Conference Evaluation which will be available online from Friday.
- Visit the World of Solutions on Level -01!
- Please remember this is a ‘No Smoking’ venue!
- Please switch off your mobile phones!
- Please remember to wear your badge at all times including the Party!
- Do you have a question? Feel free to ask them during the Q&A section or write your question on the Question form given to you and hand it to the Room Monitor when you see them holding up the Q&A sign.

Agenda

- Introduction to Data Center—The Evolution
- Application and Business Continuance
    - Increasing HA in the Data Center
    - HA with Virtualisation
- Data Center Disaster Recovery
    - Failure Scenarios
    - Design Options
- Components of Disaster Recovery
    - Site Selection—Front End GSLB
    - Server High Availability—Clustering

The Evolution of Data Center

Data Center Evolution



Data Center Elements

Application Solution

Linux/HP, Solaris/SunFire, WebLogic, J2EE Custom App, Etc.

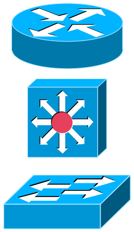
Database Solution

Linux/HP, Solaris/ SunFire, Oracle 10G RAC, Etc.

Storage Solution

MDS9000

Data Center Elements

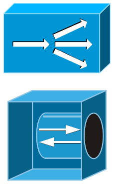


Network Infrastructure Solution

Cisco GSRs, Cisco Catalyst 6500, Cisco Catalyst Cat4000

Application Solution

Linux/HP, Solaris/SunFire, WebLogic, J2EE Custom App, Etc.

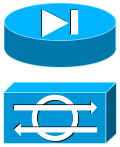


Layers 4–7 Services Solution

ACE, CSM, SSLM, CSS, CE, GSS

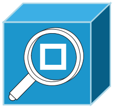
Database Solution

Linux/HP, Solaris/ SunFire, Oracle 10G RAC, Etc.



Network Security Solution

PIX®, FWSM, IDSM, VPNSM, CSA



Management and Instrumentation Solution

Terminal Servers, NAM, Cisco Works LMS/VMS, ANM, ASDM

Storage Solution

MDS9000

Data Center Elements
Redundancy

Network Infrastructure Solution

GSRs, Cat6500, Cat4000

HSRP, RPR, SSO, NSF, RPVST, MST

Application Solution

Linux/HP, Solaris/SunFire, WebLogic, J2EE Custom App, Etc.

Layers 4–7 Services Solution

ACE, CSM, SSLM, CSS, CE, GSS

Database Solution

Linux/HP, Solaris/ SunFire, Oracle 10G RAC, Etc.

99.999% Availability Desired

State Full Redundancy on ACE, CSM and FWSM

Network Security Solution

PIX, FWSM, IDSM, VPNSM, CSA

Management and Instrumentation Solution

Terminal Servers, NAM, Cisco Works LMS/VMS, ANM, ASDM

Storage Solution

MDS9000

Data Center Elements
Scalability

Network Infrastructure Solution

GSRs, Cat6500, Cat4000

Core, Aggregation/Distribution/ Services, Access Model

Application Solution

Linux/HP, Solaris/SunFire, WebLogic, J2EE Custom App, etc.

Layers 4–7 Services Solution

ACE, CSM, SSLM, CSS, CE, GSS

Database Solution

Linux/HP, Solaris/ SunFire, Oracle 10G RAC, etc.

Flexible and Simple Growth Capabilities Desired

Ability to Scale to Multiple Services Modules (ACE, SSLM, etc.)

Network Security Solution

PIX, FWSM, IDSM, VPNSM, CSA

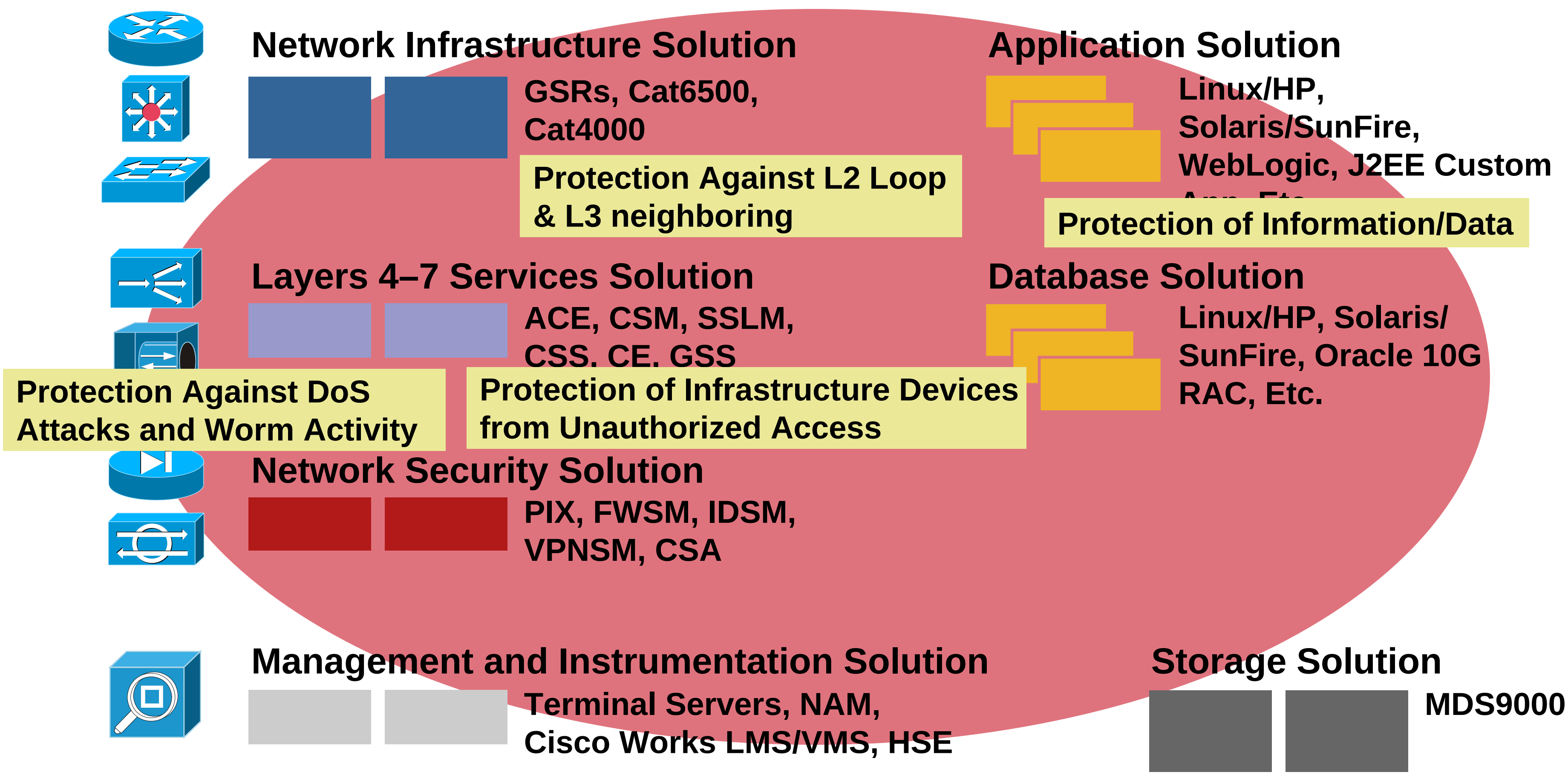
Management and Instrumentation Solution

Terminal Servers, NAM, Cisco Works LMS/VMS, ANM, ASDM

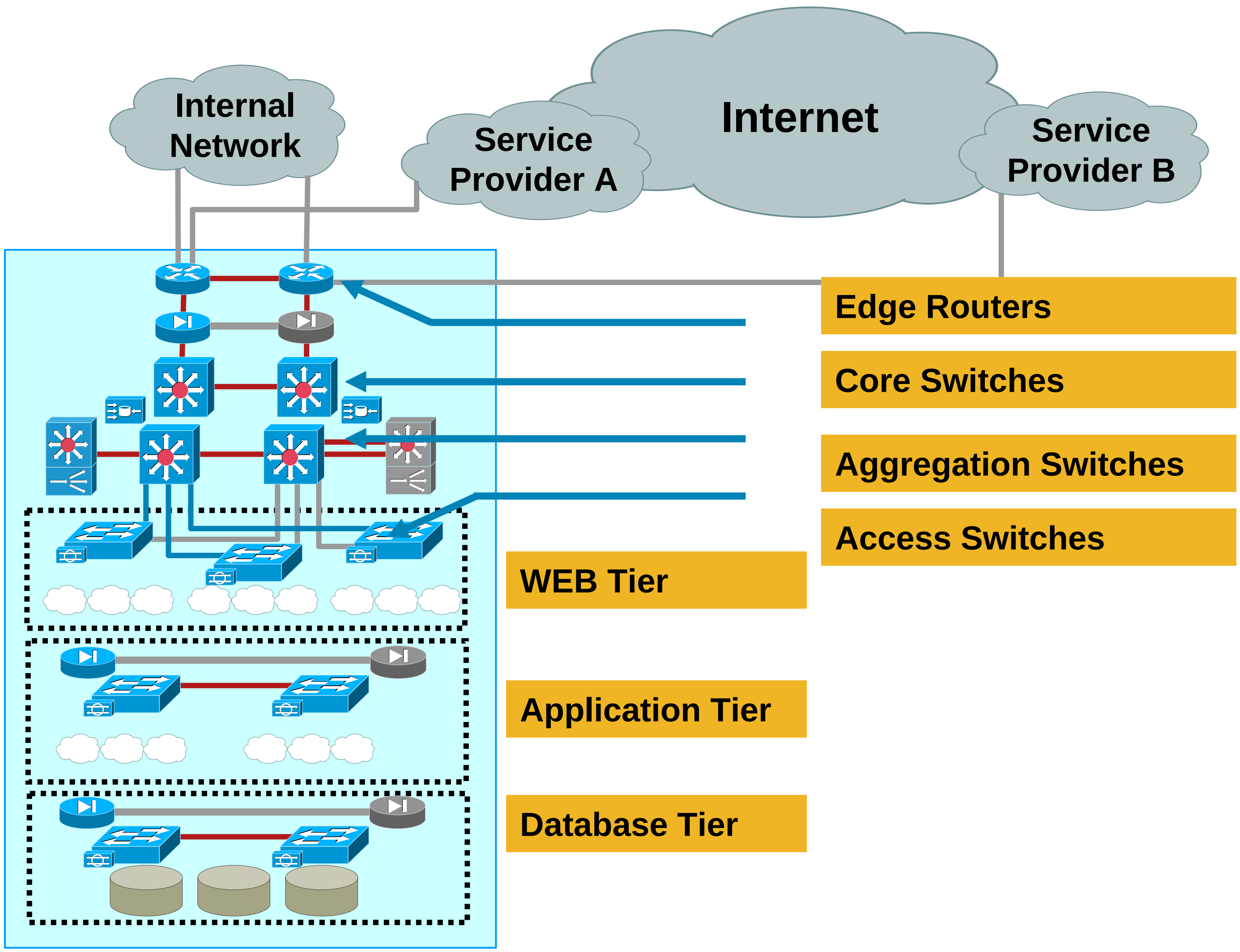
Storage Solution

MDS9000

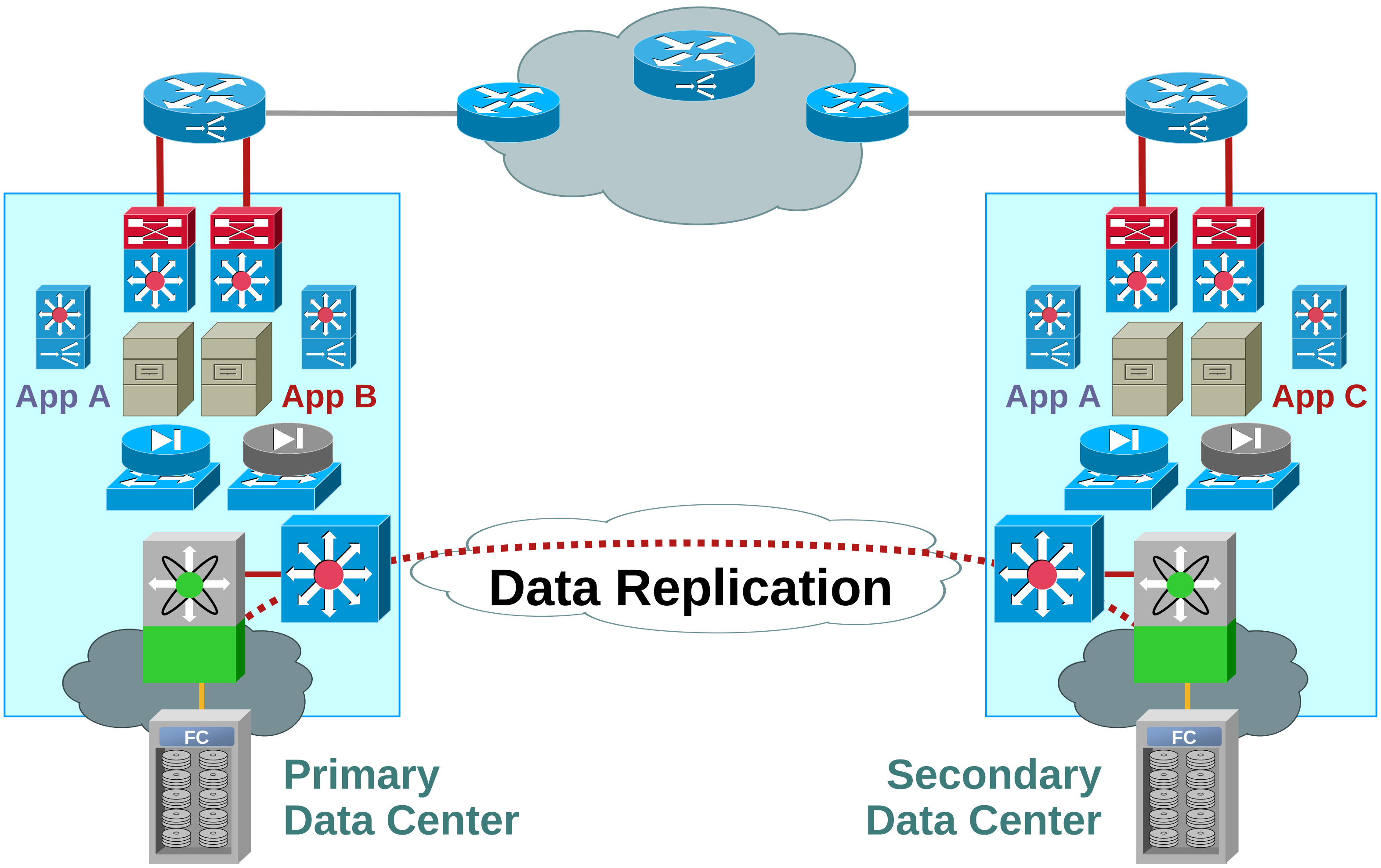
Data Center Elements
Security



Typical Data Center Topology



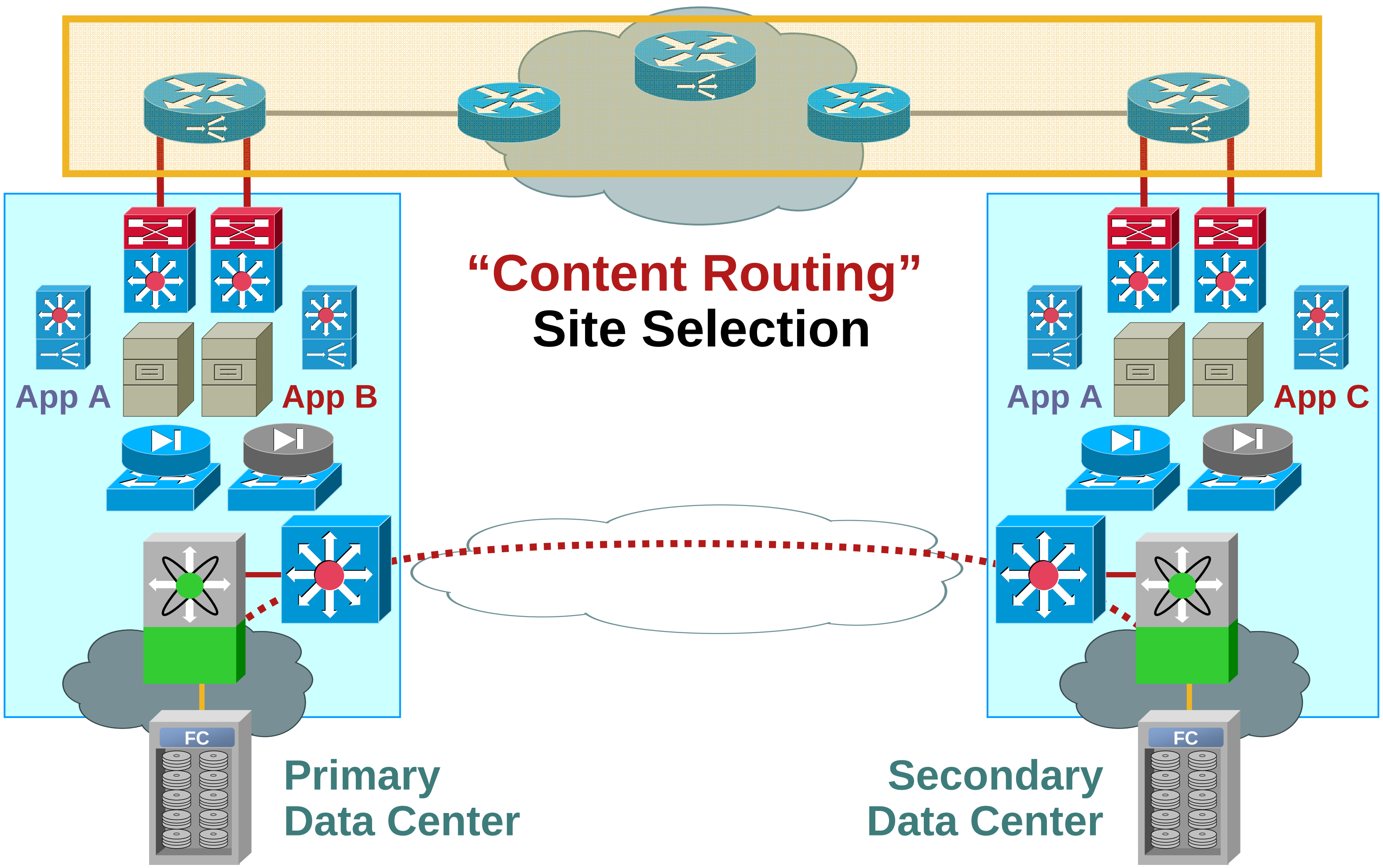
Distributed Data Center



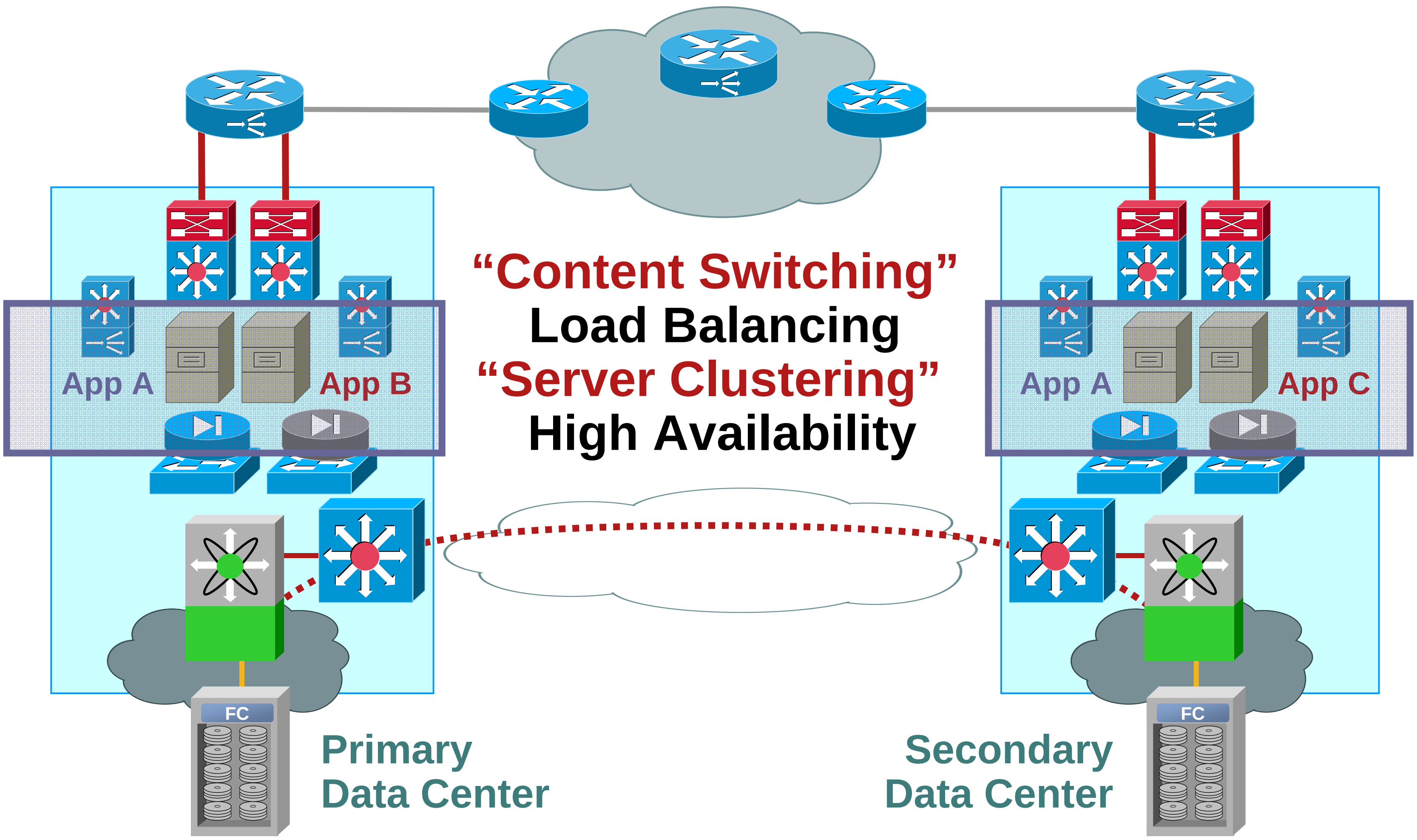
Why Distributed Data Centers?

- Required by disaster recovery and business continuance
- Avoid single, concentrated data depositary
- High availability of applications and data access
- Load balancing together with performance scalability
- Better response and optimal content routing: proximity to clients

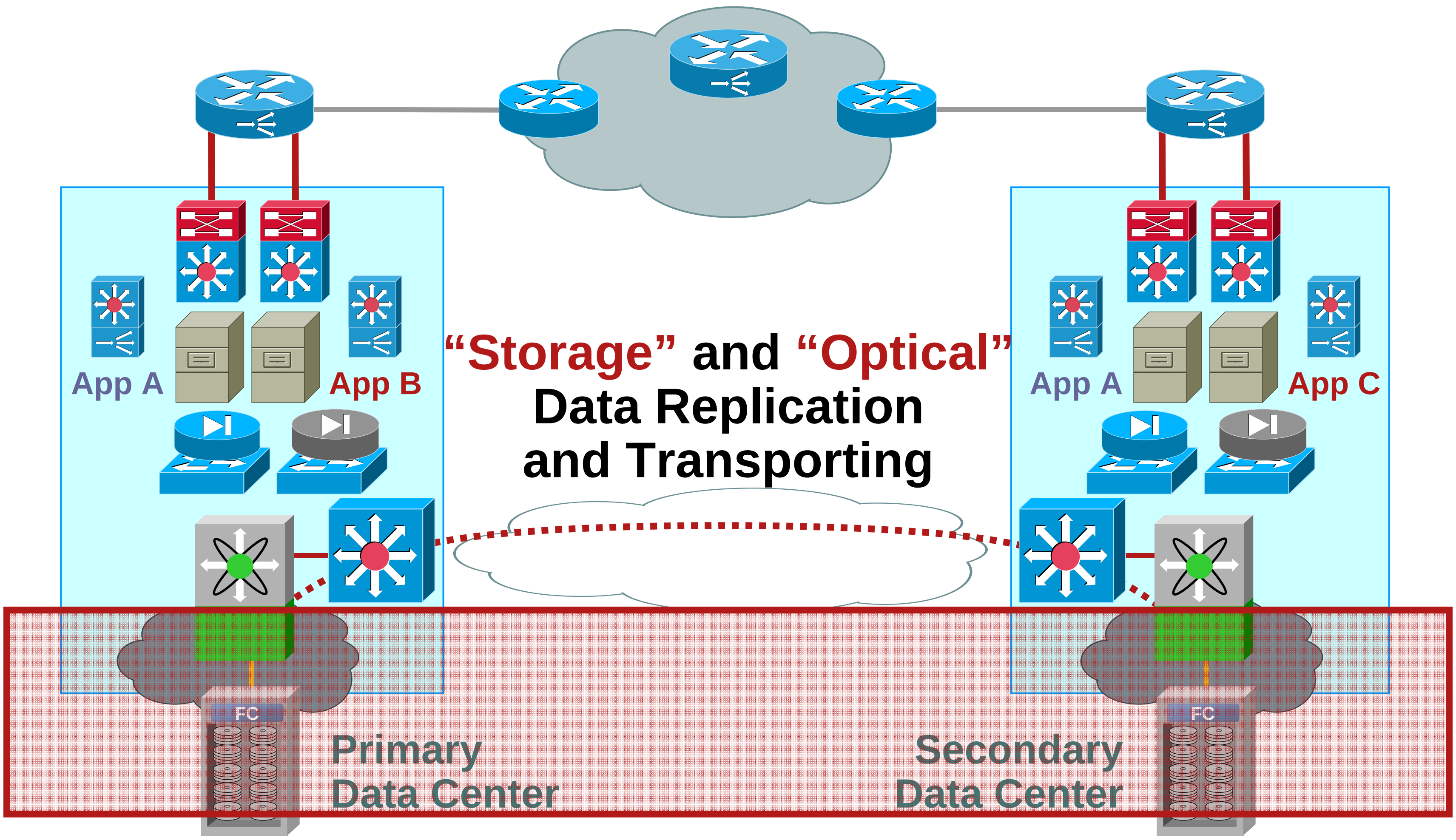
Front-End IP Access Layer



Application and Database Layer



Backend SAN Extension



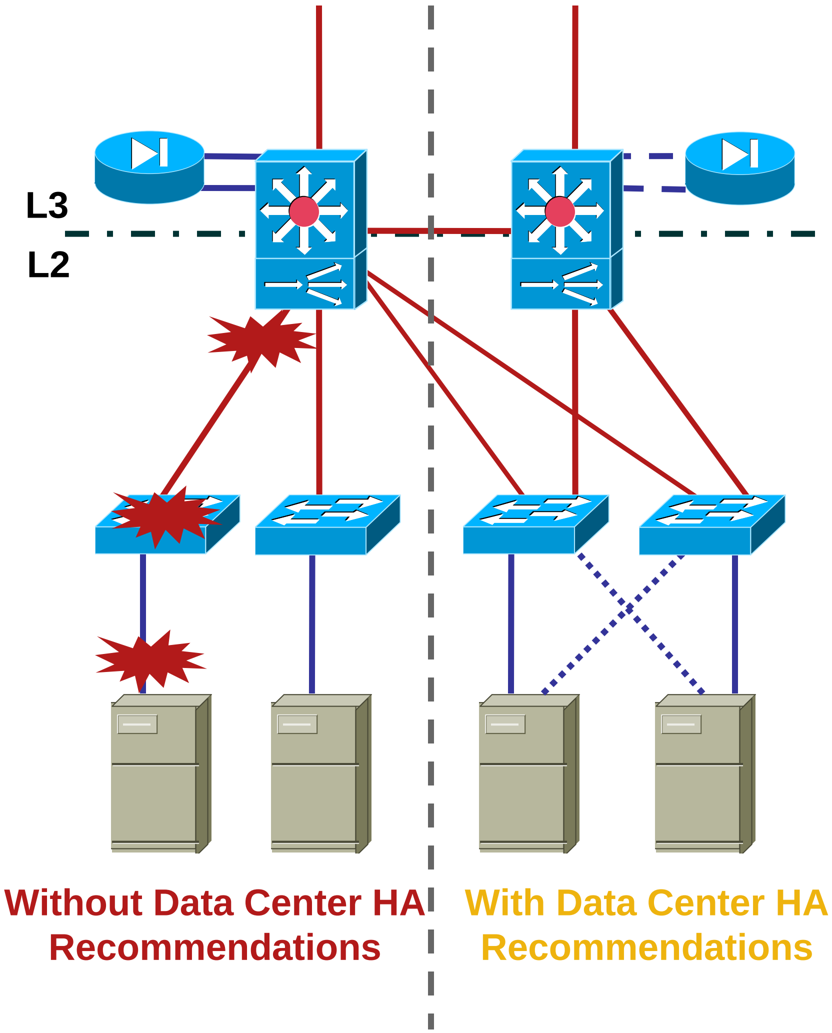
Data Center Application & Business Continuance

Agenda

- Introduction to Data Center—The Evolution
- Application and Business Continuance
  - Increasing HA in the Data Center
  - HA with Virtualisation
- Data Center Disaster Recovery
  - Failure Scenarios
  - Design Options
- Components of Disaster Recovery
  - Site Selection—Front End GSLB
  - Server High Availability—Clustering

High Availability in the Data Center
Server High Availability

Common Points of Failure



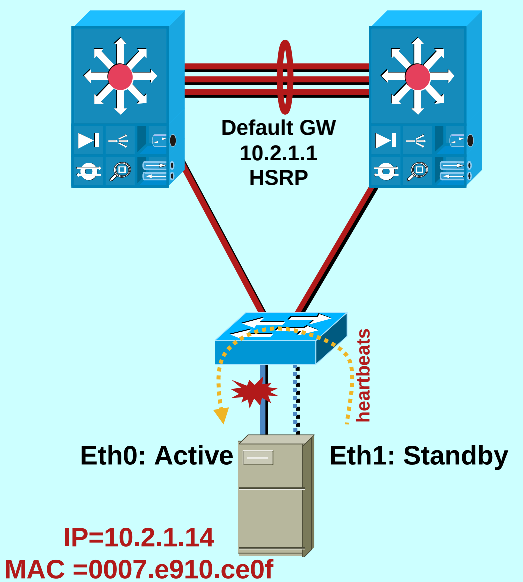
1. Server network adapter
2. Port on a multi-port server adapter
3. Network media (server access)
4. Network media (uplink)
5. Access switch port
6. Access switch module
7. Access switch
8. Aggregation switch port

These Network Failure Issues Can Be Addressed by Deployment of Dual Attached Servers Using Network Adapter Teaming Software

High Availability in the Data Center

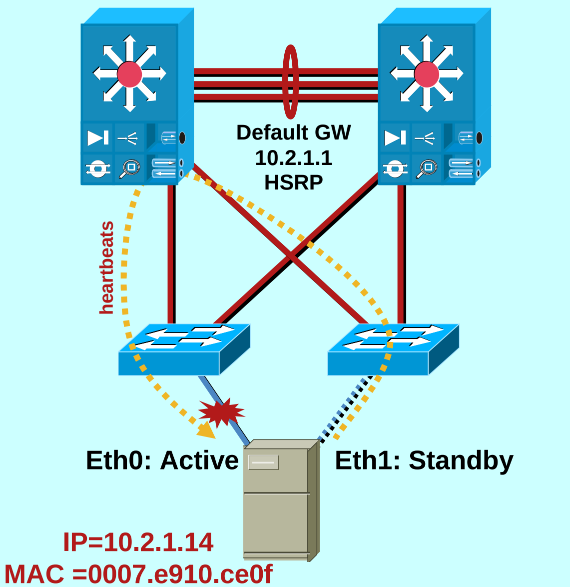
Common NIC Teaming Configurations

AFT—Adapter Fault Tolerance



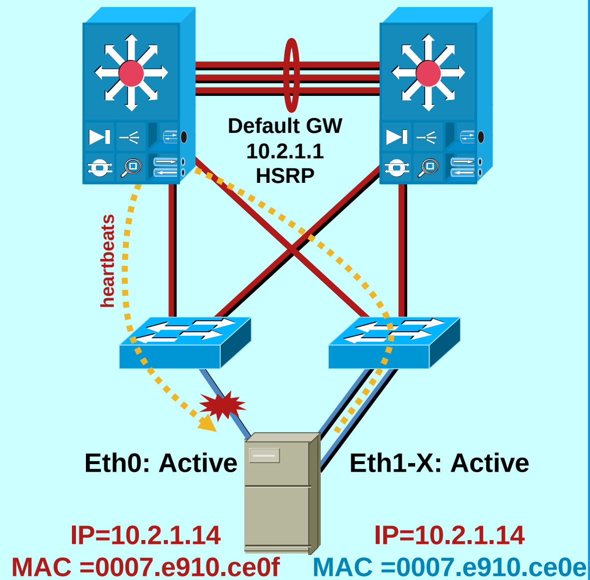
On failover, Src MAC Eth1 = Src MAC Eth0
IP address Eth1 = IP address Eth0

SFT—Switch Fault Tolerance



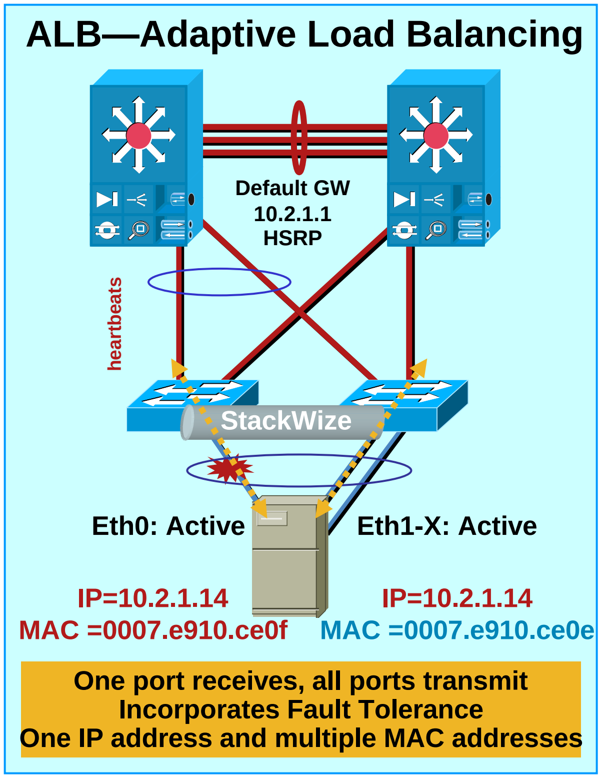
On failover, Src MAC Eth1 = Src MAC Eth0
IP address Eth1 = IP address Eth0

ALB—Adaptive Load Balancing



One port receives, all ports transmit
Incorporates Fault Tolerance
One IP address and multiple MAC addresses

High Availability in the Data Center
LACP – 802.1ad



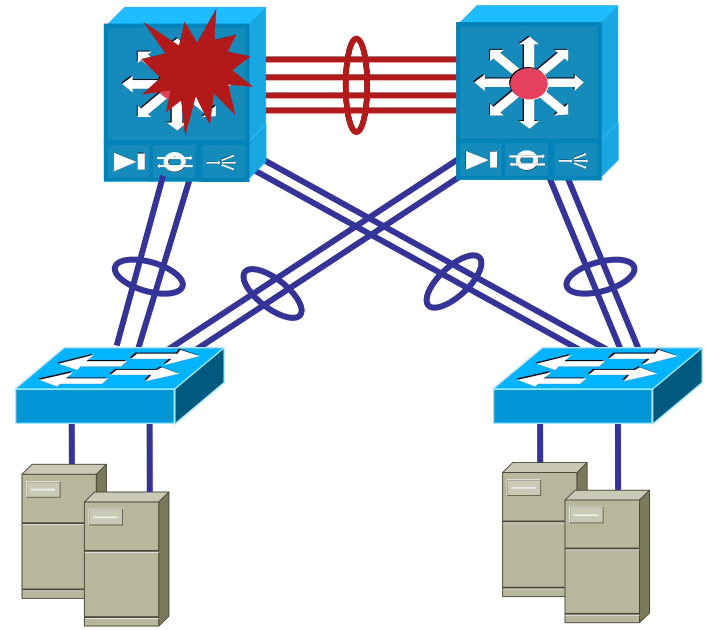
EtherChannel splitted between multiple Access Switches (Cat3750 StackWize) provides :

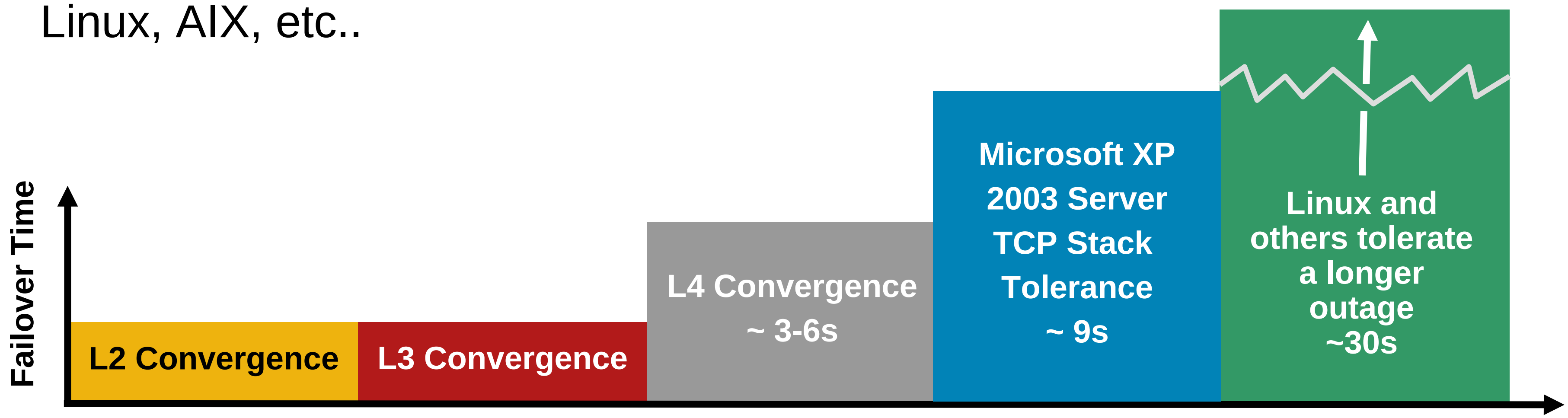
- **Higher Throughput**
- **Higher Availability**

at both Server and Switch sides….

High Availability in the Data Center
Failover Times

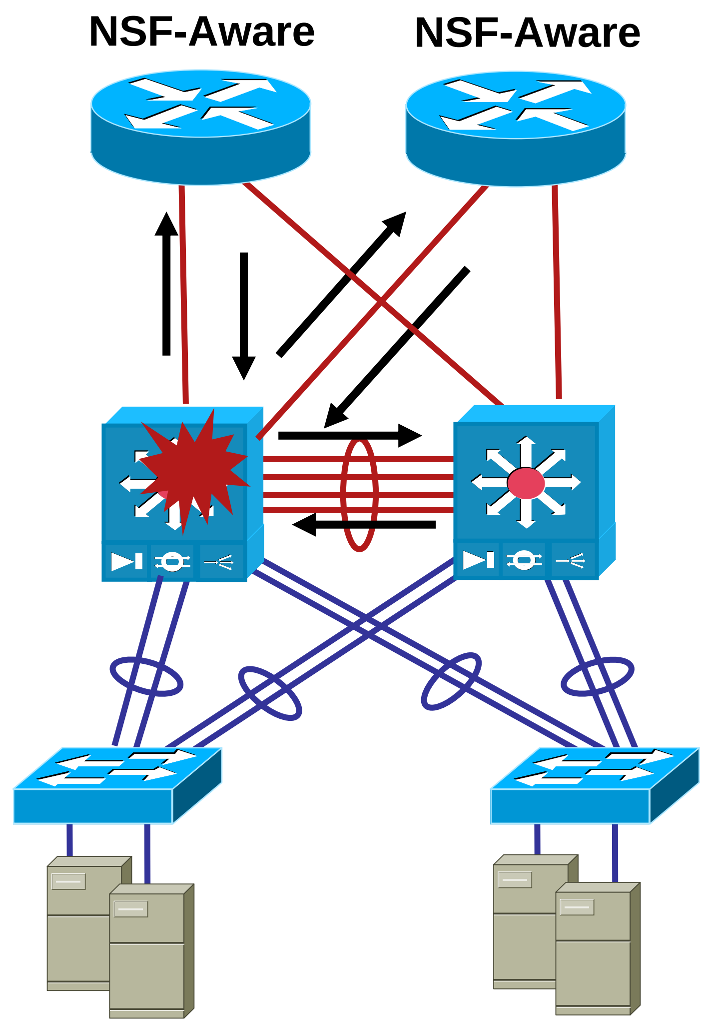
- The overall failover time is the combination of convergence at L2, L3, + L4 components
  - –Stateful devices can replicate connection information and typically failover within 3-5sec
  - –EtherChannels < 1sec
  - –STP converges in ~1 sec
  - –HSRP can be tuned to <1s , but why
- Where does TCP break? Microsoft, Linux, AIX, etc..





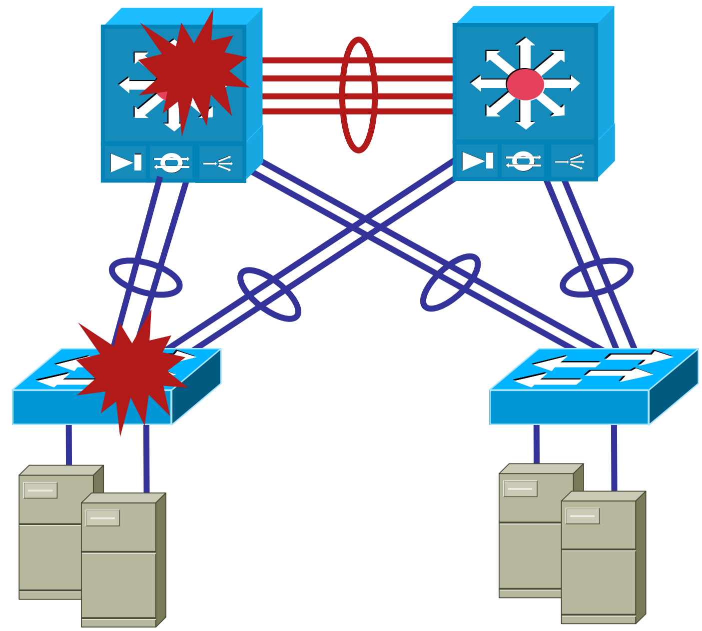
High Availability in the Data Center
NSF/SSO

- NSF/SSO is a supervisor redundancy mechanism for intra-chassis supervisor failover
- SSO synchronizes layer 2 protocol state, hardware L2/L3 tables (MAC, FIB, adjacency table), ACL and QoS tables
- SSO synchronizes state for: trunks, interfaces, EtherChannels, port security, SPAN/RSPAN, STP, UDLD, VTP
- SSO prevents line cards and service modules from reset.
- NSF with EIGRP, OSPF, IS-IS, BGP makes it possible to have no route flapping during the recovery



High Availability in the Data Center
NSF/SSO in the Data Center

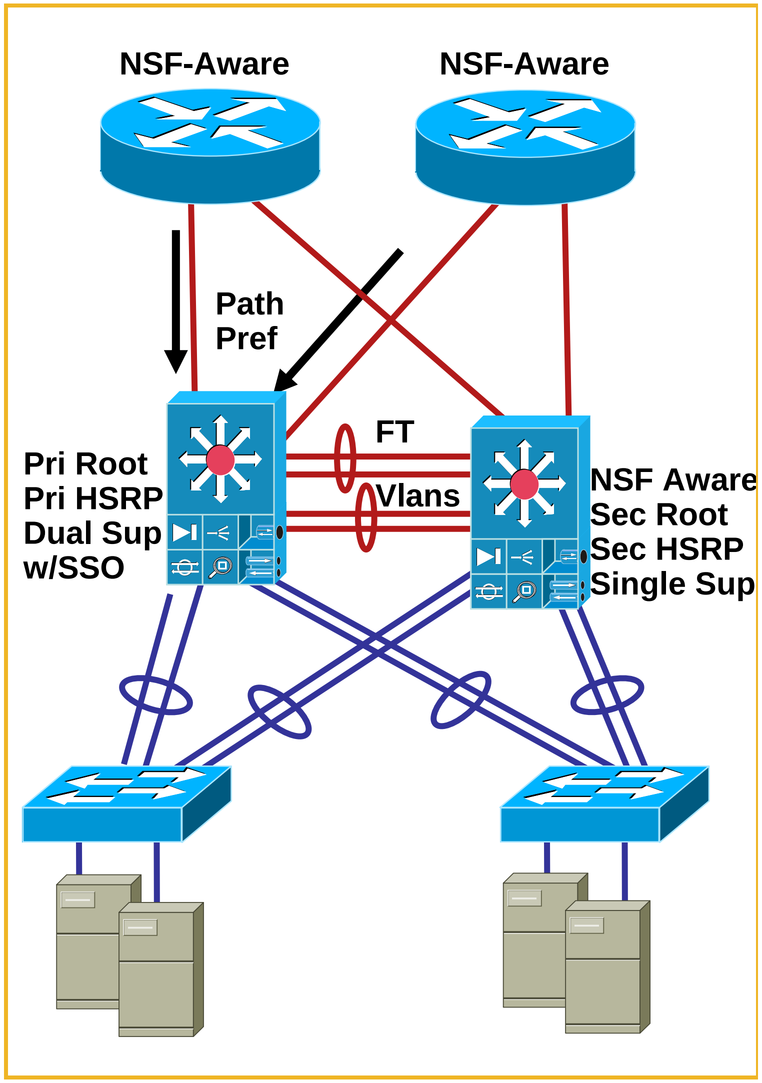
- SSO in the Access Layer:
  - –Improves availability for single attached servers
  - –Under 2s convergence
- SSO in the Aggregation Layer:
  - –Consider in primary agg layer switch
  - –Avoids rebuild of arp, igp, stp tables
  - –Prevents service module switchover (~6sec or greater)
  - –SSO switchover time less than 2sec
  - –12.2.18SXD3 or higher
- Possible Implications
  - –* HSRP state between Agg switches is not tracked and will show switchover, existing sessions resume with Agg1 as default gateway
  - –* RHI is not SSO aware: must extend failed and retry timers
  - –IGP Timers cannot be aggressive (tradeoff)



* Note: RHI and HSRP will be NSF/SSO aware in next release

High Availability in the Data Center
Hardening the Aggregation Layer

- FT Path for Service Modules
  - Consider second channel/link for FT vlans. Helps to prevent active/active scenario in congested or certain failure or mis-configuration conditions
  - FWSM checks for mate on 2 interfaces before switching to active (FT or data vlans)
- Establish Path Preference: Align primary service modules on Agg1 as preferred path – leverage Route Health Injection on CSM
  - Use Probes to monitor health of server farm
  - Use “Advertise Active” to dynamically install host route
  - Adjust Route-Map metric such that Agg1 is preferred route advertised to core. If active-active occurs, Agg1 will be preferred path reducing change of asymmetric connections.



High Availability in the Data Center
Hardening the Aggregation Layer

- Spanning Tree

  STP primary/secondary root alignment with HSRP primary/secondary

  Avoid getting close to STP watermarks

  Avoid 2 tier- Layer 2 “Super Aggregation” Designs

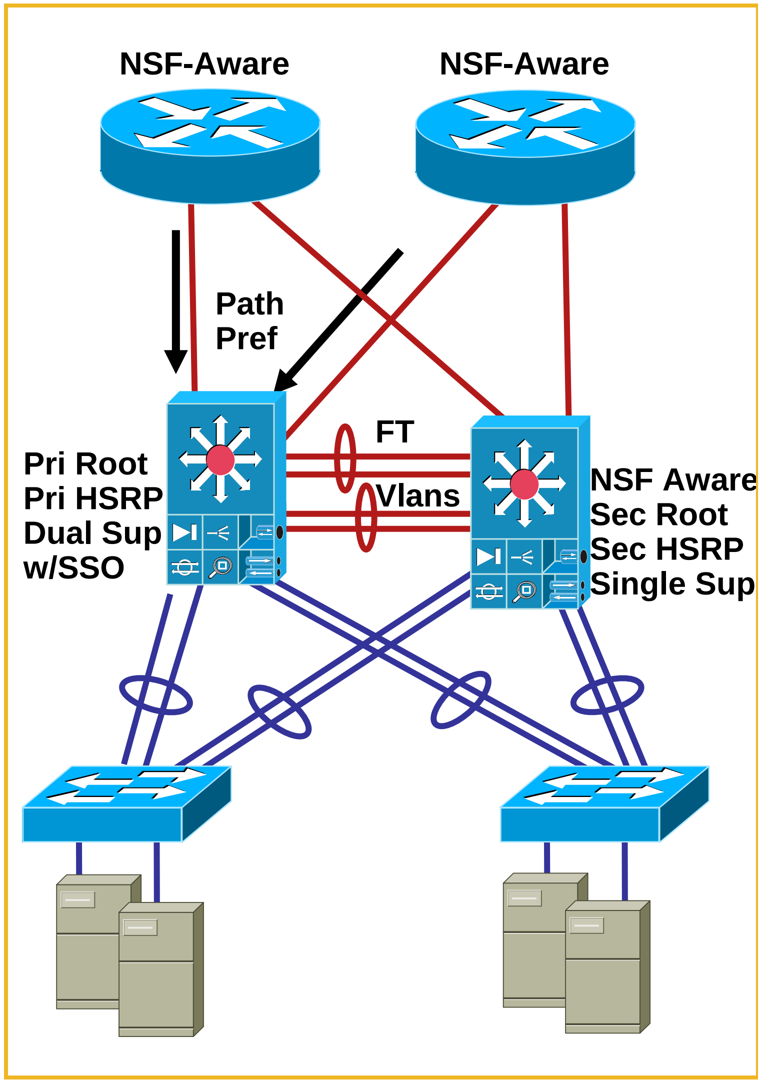
  Remove unused vlans from CSM EtherChannel interface (int range port-ch 255-259, no vlan xx)

- HSRP

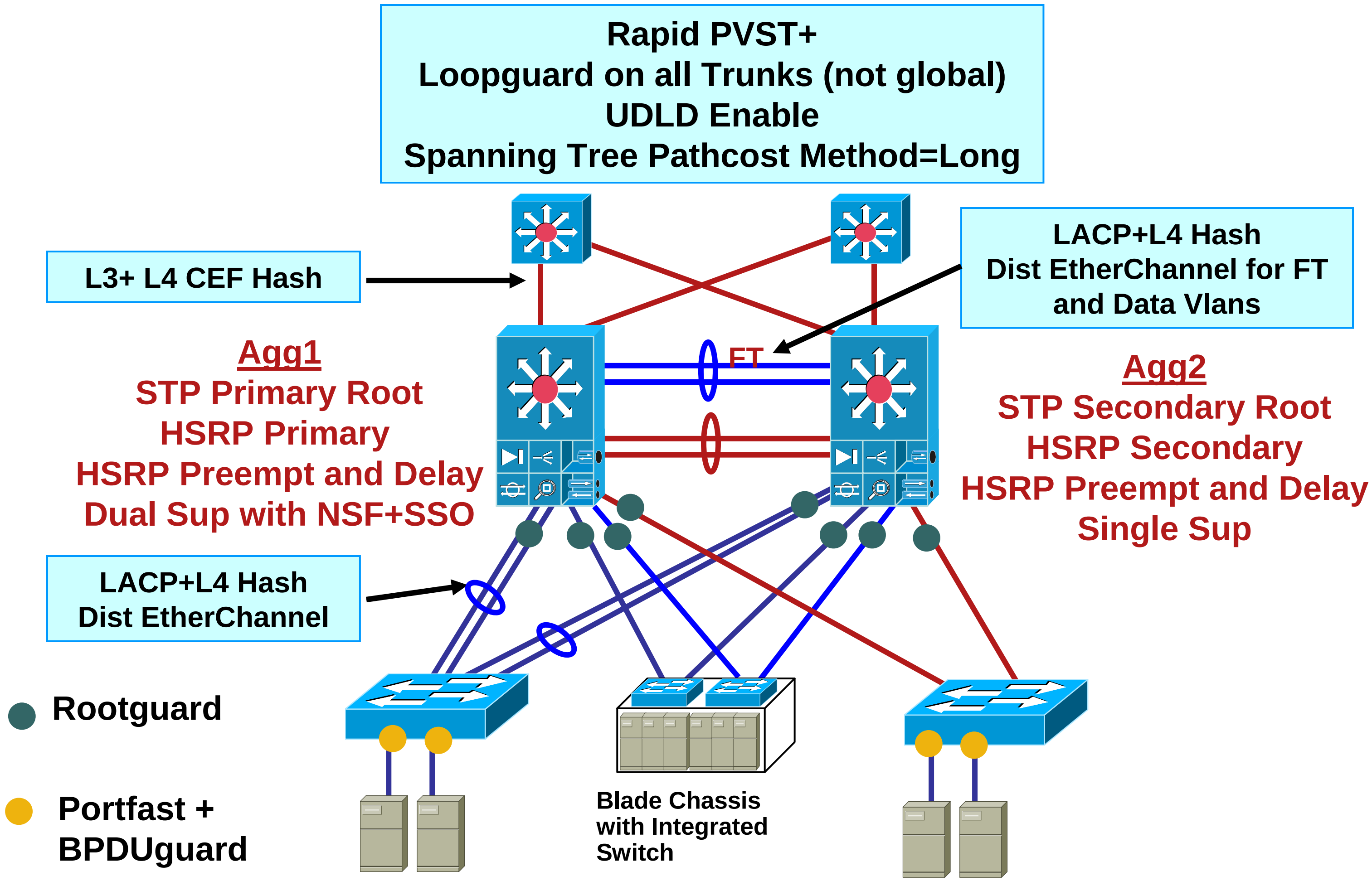
  Stay under 500 HSRP Instances per Agg module

  Recommend HSRP Hello-1, Holdown=3 timers

  Other CPU driven processes may reduce max number of instances, or increase timers



High Availability in the Data Center
Best Practices- STP, HSRP, Other



Rapid PVST+: Maximum Number of STP Active Logical Ports- 8000 and Virtual Ports Per Linecard-1800

Server Load Balancing & integrated Firewall Design

Context Switching design Approach

- Transparent & Routed approaches

Firewall design Approach

- Transparent approaches

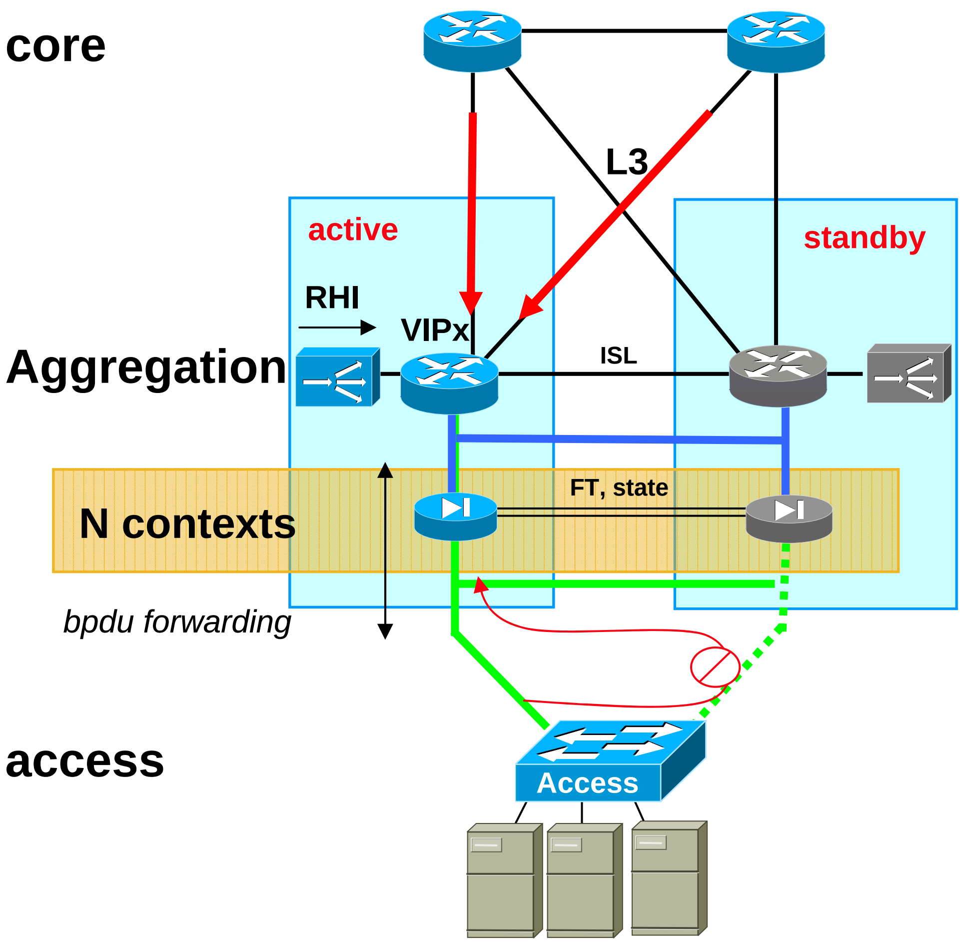
BRKAAP-1002: Introduction to ACE

BRKAPP-2005: Server Load Balancing Design

Agenda

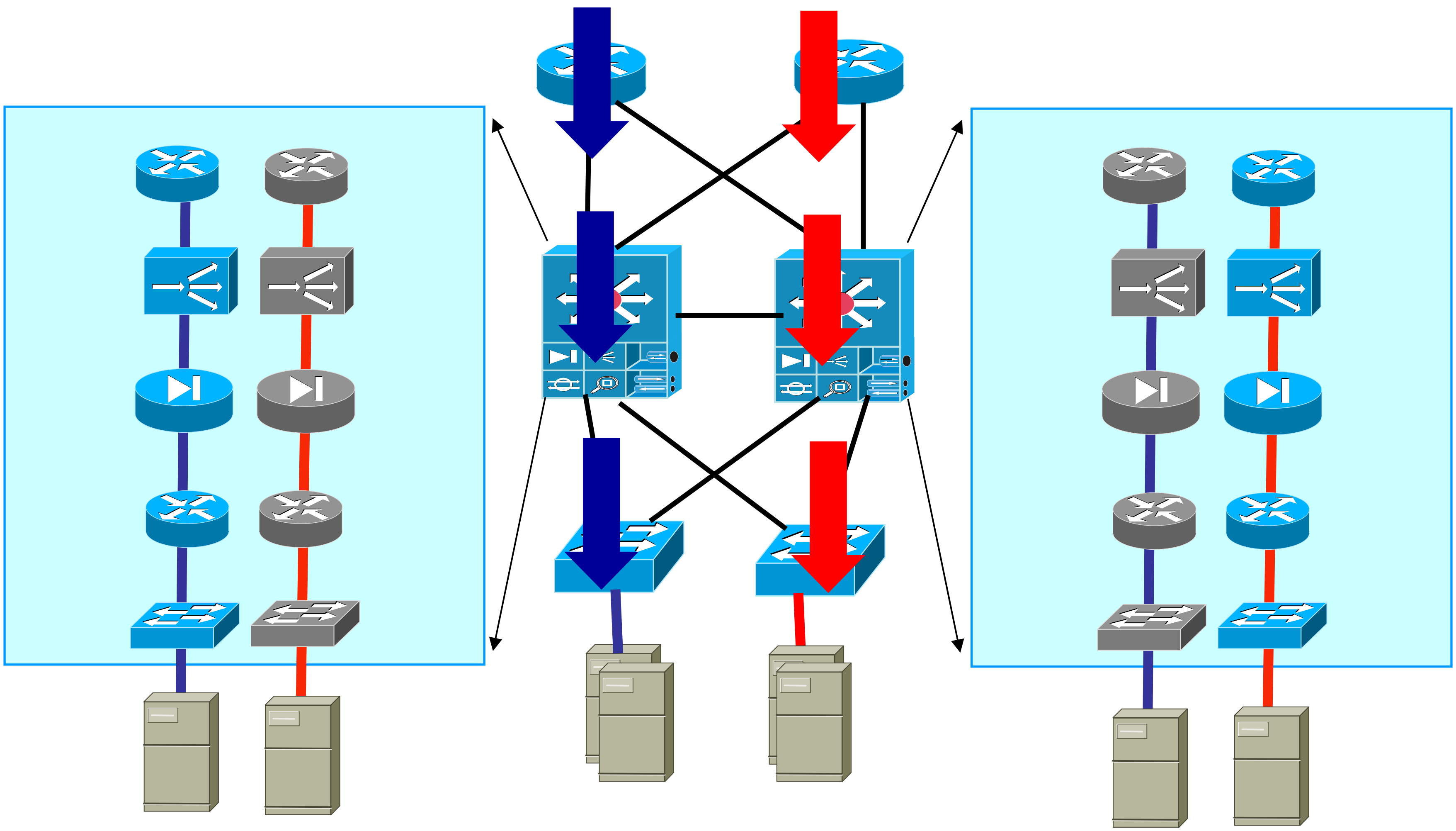
- Introduction to Data Center—The Evolution
- Application and Business Continuance
    - Increasing HA in the Data Center
    - HA with Virtualisation
- Data Center Disaster Recovery
    - Failure Scenarios
    - Design Options
- Components of Disaster Recovery
    - Site Selection—Front End GSLB
    - Server High Availability—Clustering

Typical Today's Data Center Design

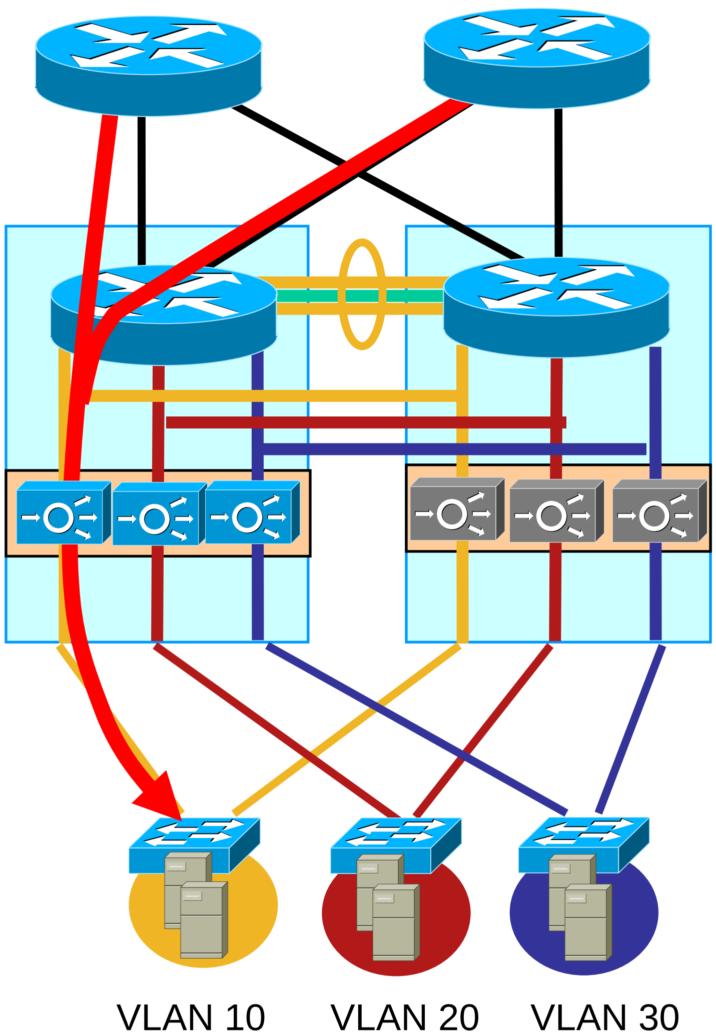


- active / standby configuration
- point-to-point L3 links to core
- CSM 1-arm routed
  - server-to-server offload
  - PBR / source NAT
- RHI attracts traffic to active switch
  - minimize ISL requirement
- segmentation
  - transparent firewalls
  - multiple contexts
  - single failover group
- Looped access with rapid PVST+
  - dual links + trunk

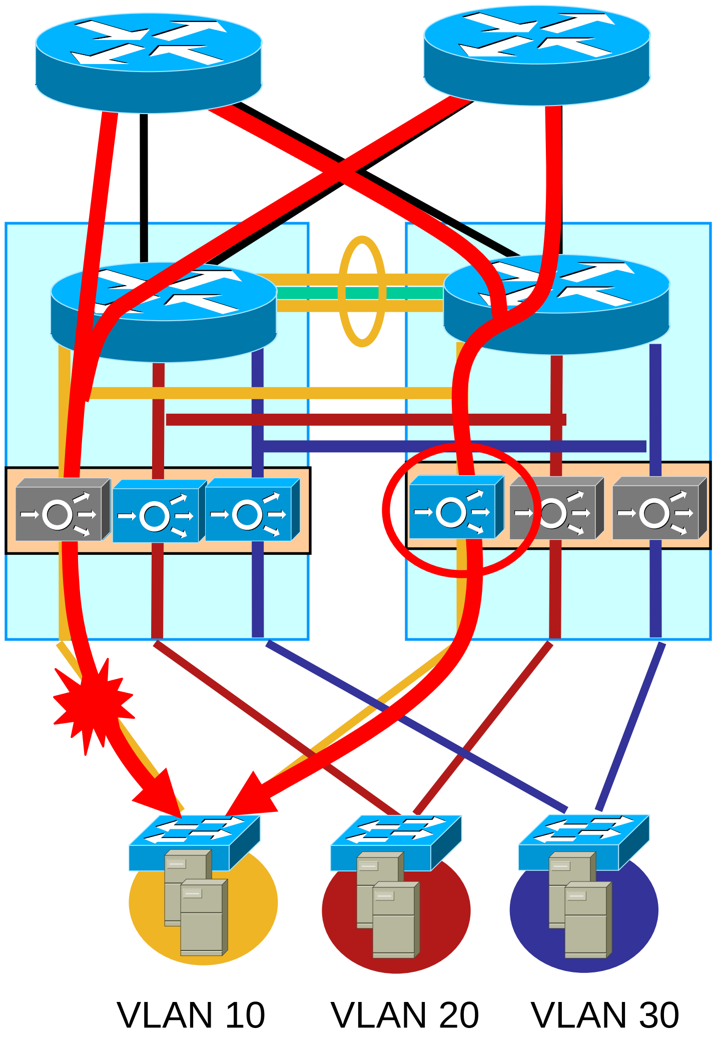
Chaining of L4-L7 Services



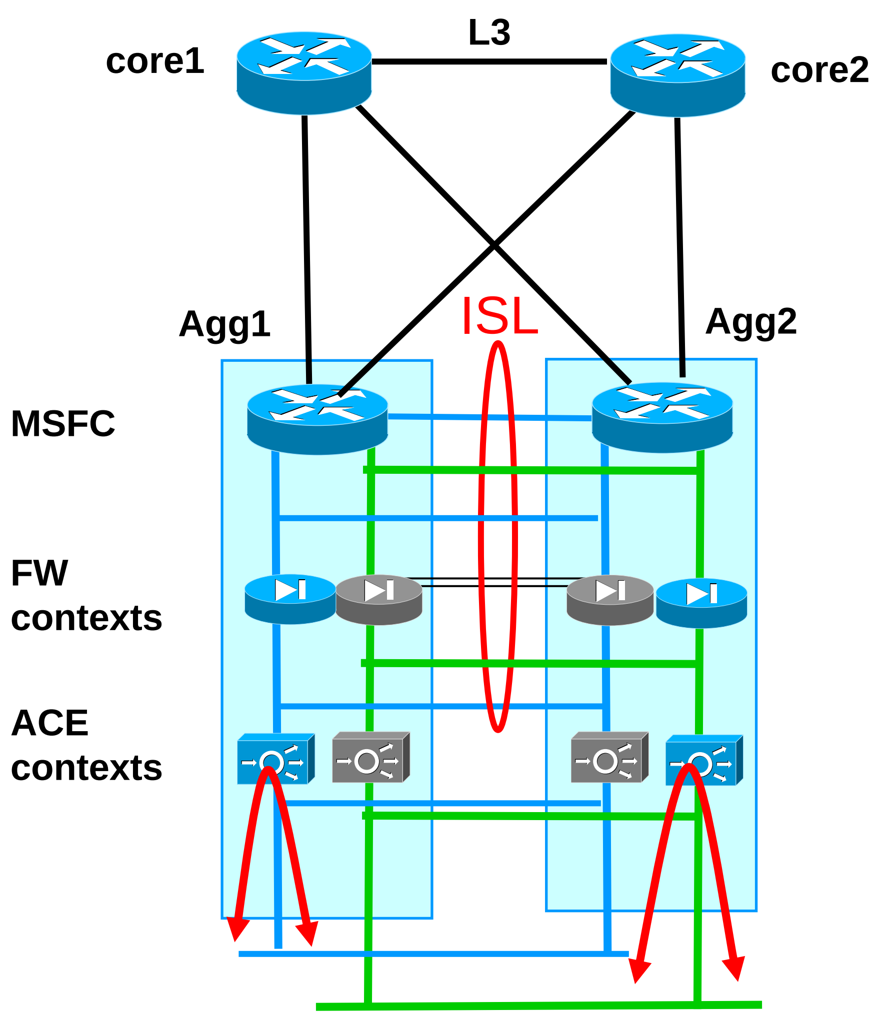
Failure Isolation with Virtualization



Failure Isolation with Virtualization



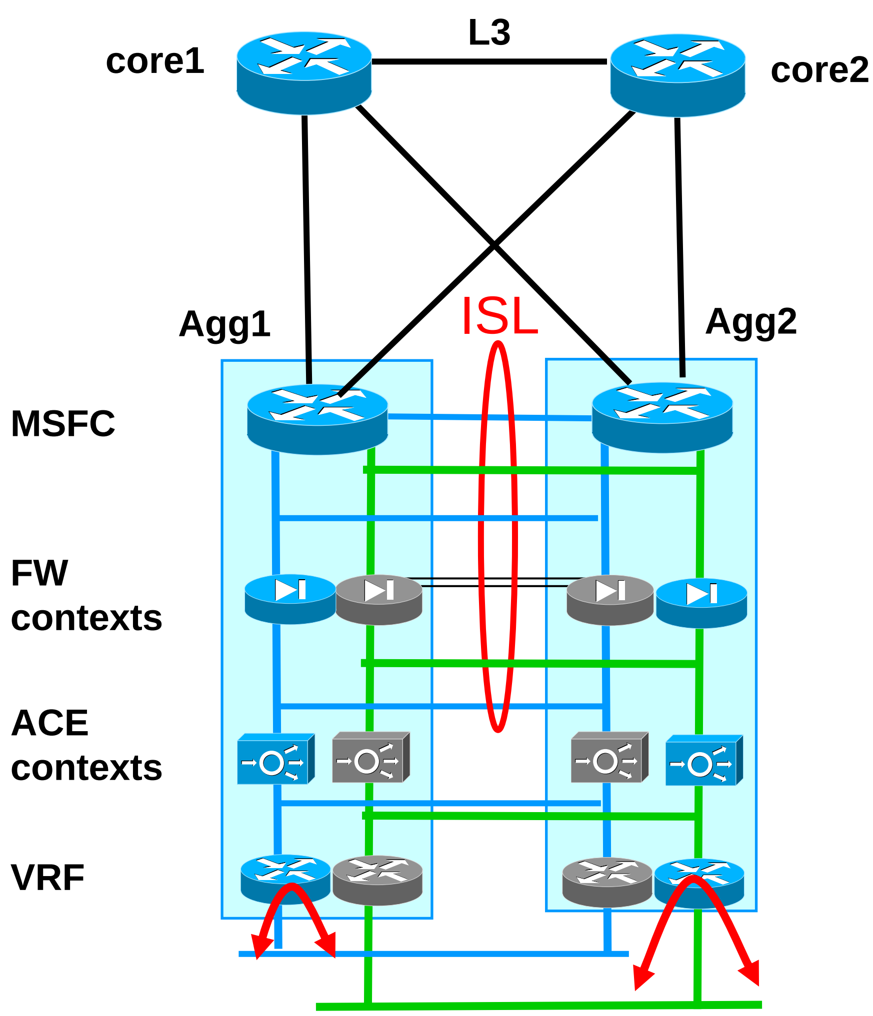
Adding Virtualized Firewalls



- You could place a Virtualized Firewall closer to the servers, or between the MSFC and the ACE

- Considering that ACE provides higher throughput you may want to keep the server-to-server with LB traffic off of the FWSM

Adding VRF-lite



- A Virtual Routing Instance closer to the server can provide higher server-to-server forwarding throughput with basic security mechanisms, such as ACLs

- You would add a VRF with multiple SVIs behind of the ACE

- Server-to-server traffic with load balancing would go up to the ACE and back to the VRF

Note: Static ARPs map to static mac-addresses

arp vrf red 13.20.80.2 0000.0000.0080 ARPA

mac-address 0000.0000.0208 (int vlan 208-vrf red)

arp 13.20.80.252 0000.0000.0208 ARPA

mac-address 0000.0000.0080 (int vlan 80)

Asymmetric routing support

- FWSM support differs slightly from PIX and ASA
- There are 2 flavors of asymmetric routing support on FWSM:
  - Single FWSM (or within a virtual firewall)
    - independently of redundancy
  - When running in active/active mode
- PIX/ASA support the latter only
- Option #1 is achieved using a new concept called “ASR group”
- Option #2 is automatically enabled when configuring active/active redundancy

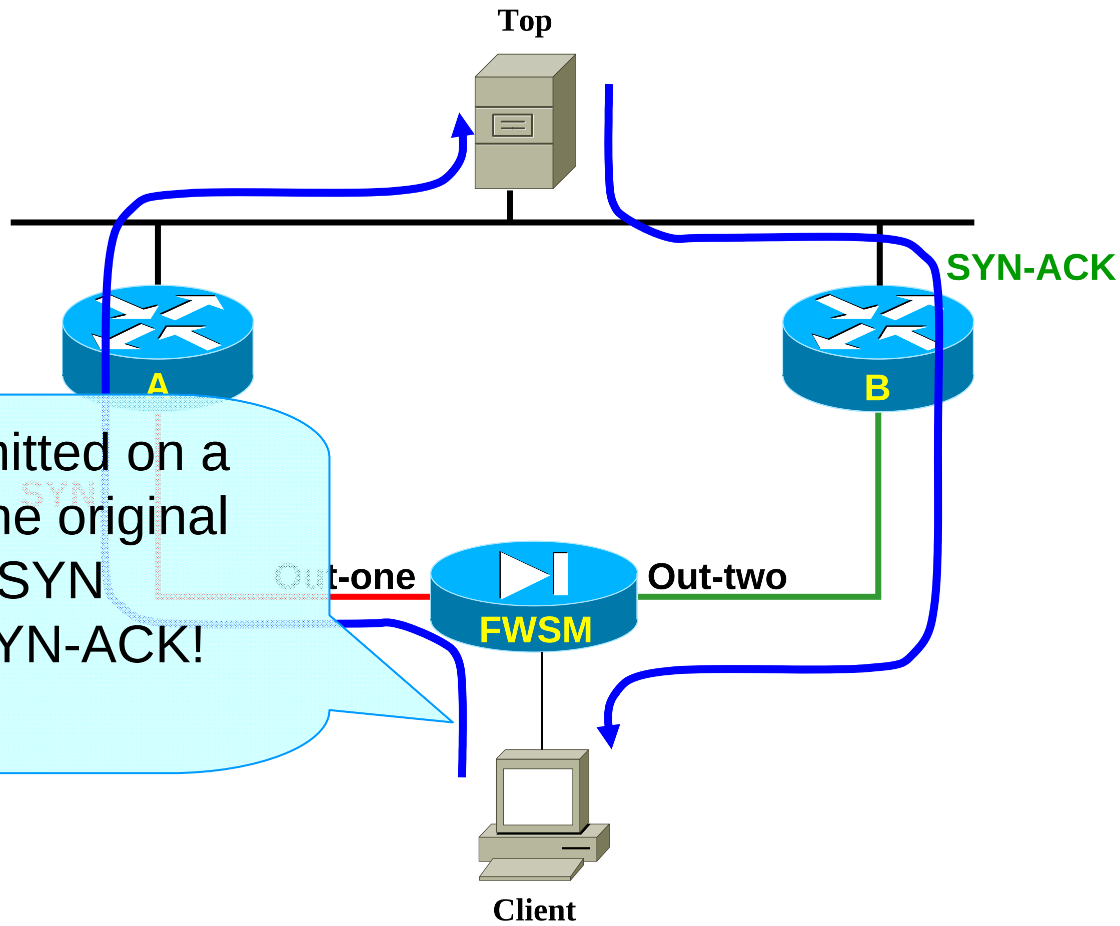
Asymmetric support using ASR groups

Client initiates a connection to Top.

The initial SYN could take the Router A route, while the SYN-ACK could come back via the Router B route.

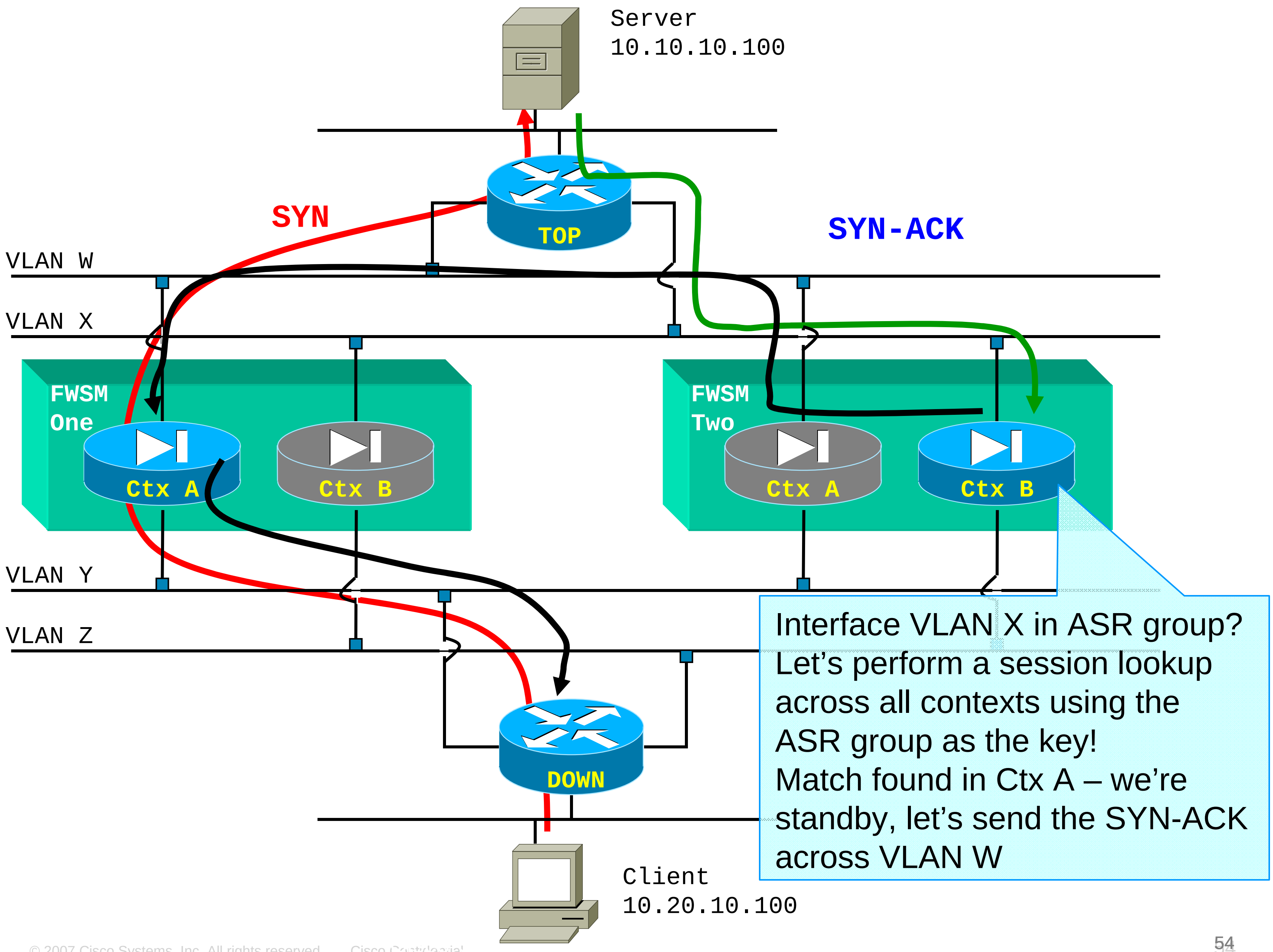
A SYN-ACK will be permitted on a different interface than the original one *if and only if* a valid SYN matches the returning SYN-ACK!

With ASR group concept it is possible for the FWSM to accept the returning SYN-ACK segment even though no corresponding SYN was ever seen on that interface.



- Up to 8 interfaces per ASR group
- Up to 32 groups per FWSM.
- Packets belonging to a given session can enter and leave from any interface within the ASR group.

Asymmetric routing support with Act/Act



Interface VLAN X in ASR group?
Let's perform a session lookup across all contexts using the ASR group as the key!
Match found in Ctx A – we're standby, let's send the SYN-ACK across VLAN W

Data Center Disaster Recovery

Agenda

- Introduction to Data Center—The Evolution
- Application and Business Continuance
  - Increasing HA in the Data Center
  - HA with Virtualisation
- Data Center Disaster Recovery
  - Failure Scenarios
  - Design Options
- Components of Disaster Recovery
  - Site Selection—Front End GSLB
  - Server High Availability—Clustering

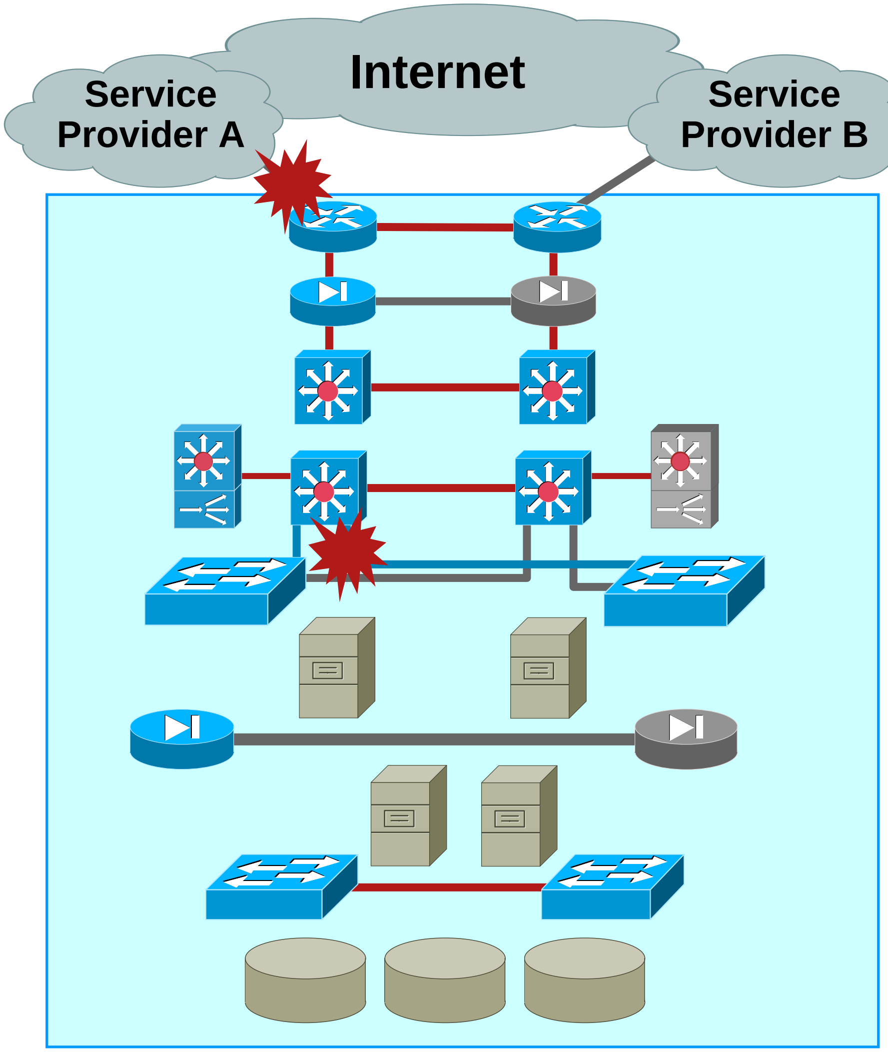
Failure Scenarios

Disaster Could Mean Many Types of Failure

- Network failure
- Device failure
- Storage failure
- Site failure

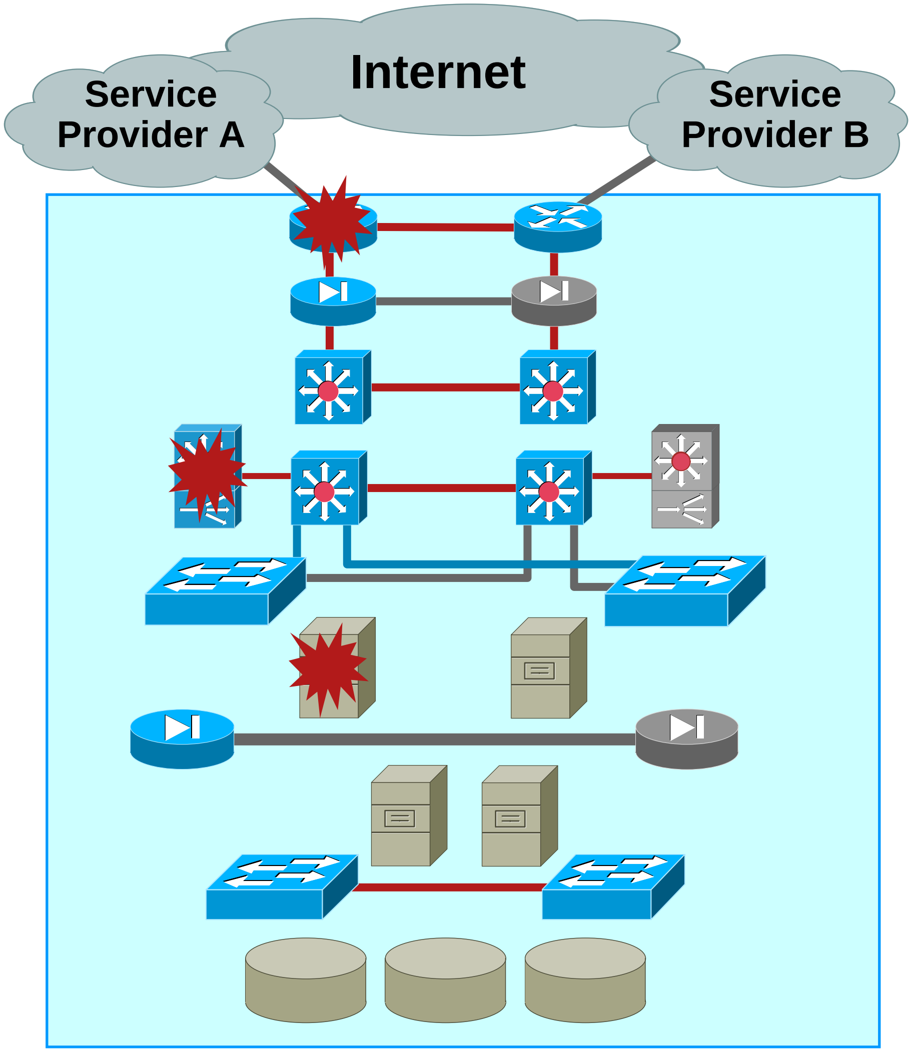
Network Failures

- ISP failure
  - ✓ Dual ISP connections
  - ✓ Multiple ISP
- Connection failure within the network
  - ✓ EtherChannel®
  - ✓ Multiple route paths



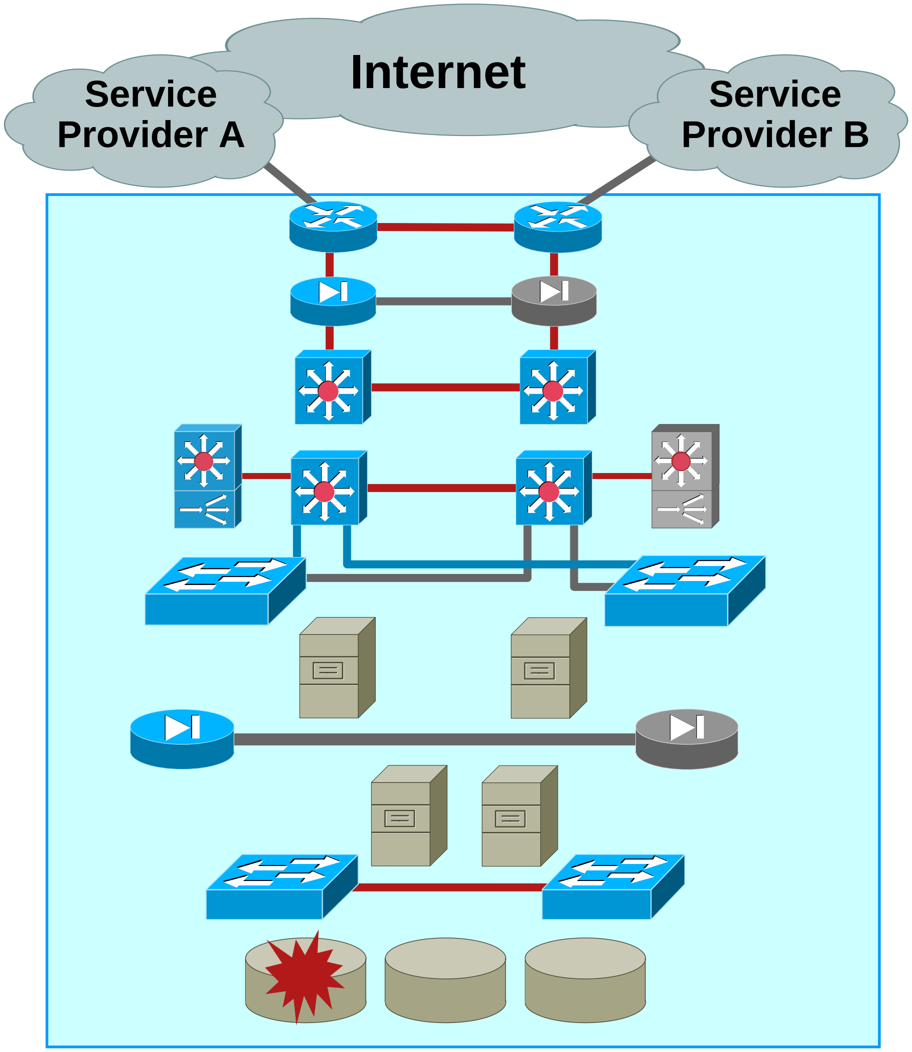
Device Failures

- Routers, switches, FWs
  - ✓ NSF/SSO
  - ✓ FT
  - ✓ HSRP
  - ✓ VRRP
- Hosts
  - ✓ HA cluster
  - ✓ LB server farm
  - ✓ NIC teaming



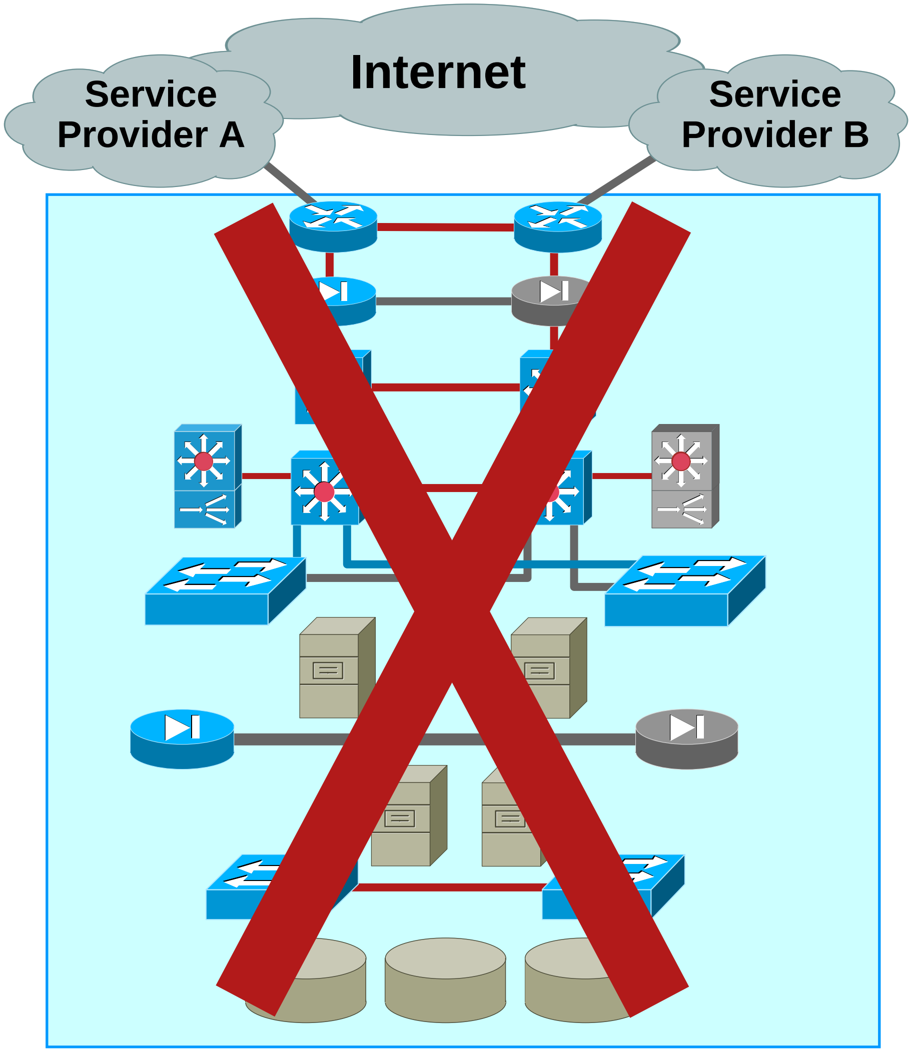
Storage Failures

- Disk arrays
  - ✓ RAID
- Disk controllers



Site Failures

- Partial site failure
  - ✓ Application maintenance
  - ✓ Application migration
  - ✓ Application scheduled DR exercise
- Complete site failure
  - ✓ Disaster



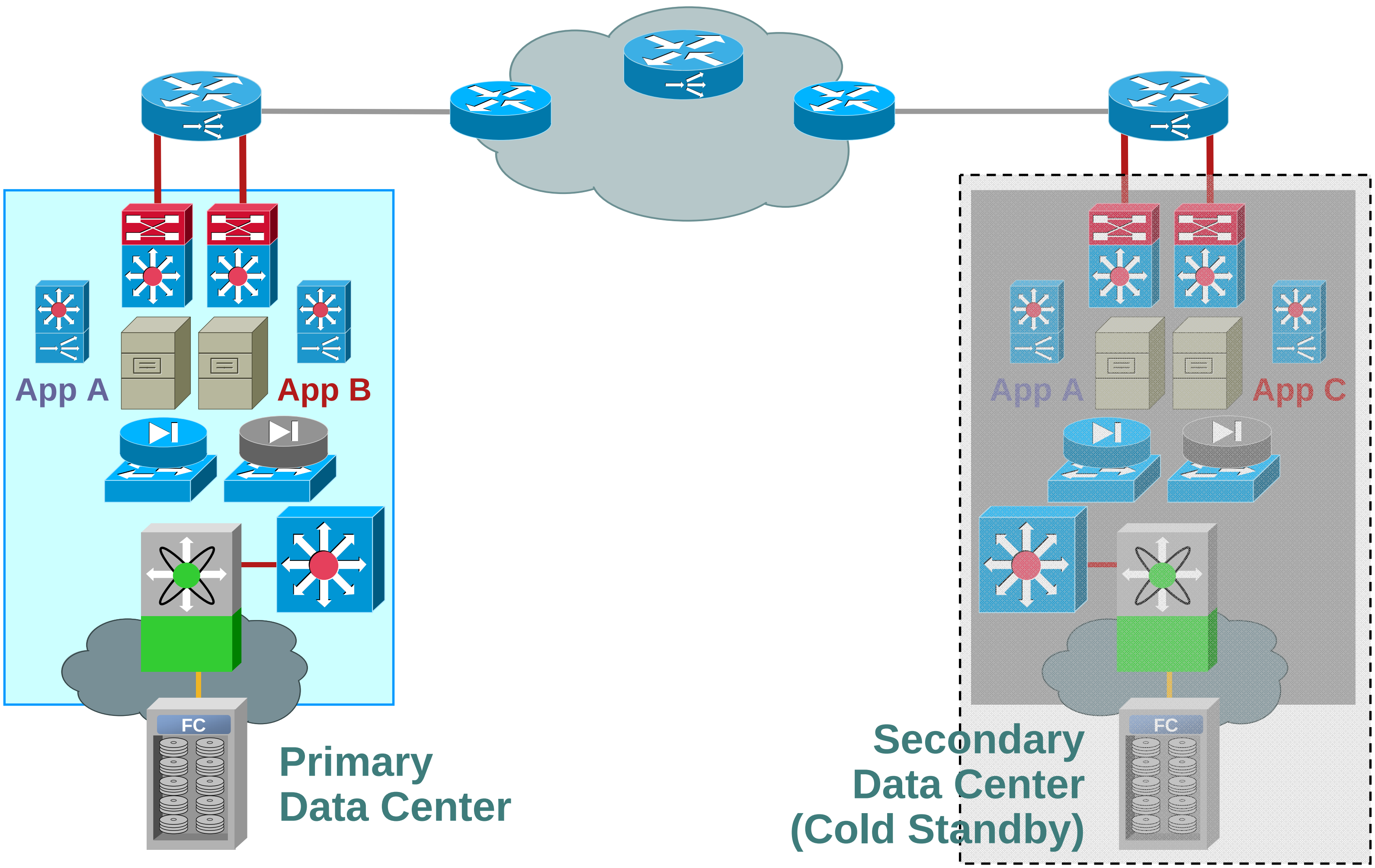
Agenda

- Introduction to Data Center—The Evolution
- Application and Business Continuance
  - Increasing HA in the Data Center
  - HA with Virtualisation
- Data Center Disaster Recovery
  - Failure Scenarios
  - Design Options
- Components of Disaster Recovery
  - Site Selection—Front End GSLB
  - Server High Availability—Clustering

Cold Standby

- One or more data center with appropriately configured space equipped with pre-qualified environmental, electrical, and communication conditioning
- Hardware and software installation, network access, and data restoration all need manual intervention
- Least expensive to implement and maintain
- Substantial delay from standby to full operation

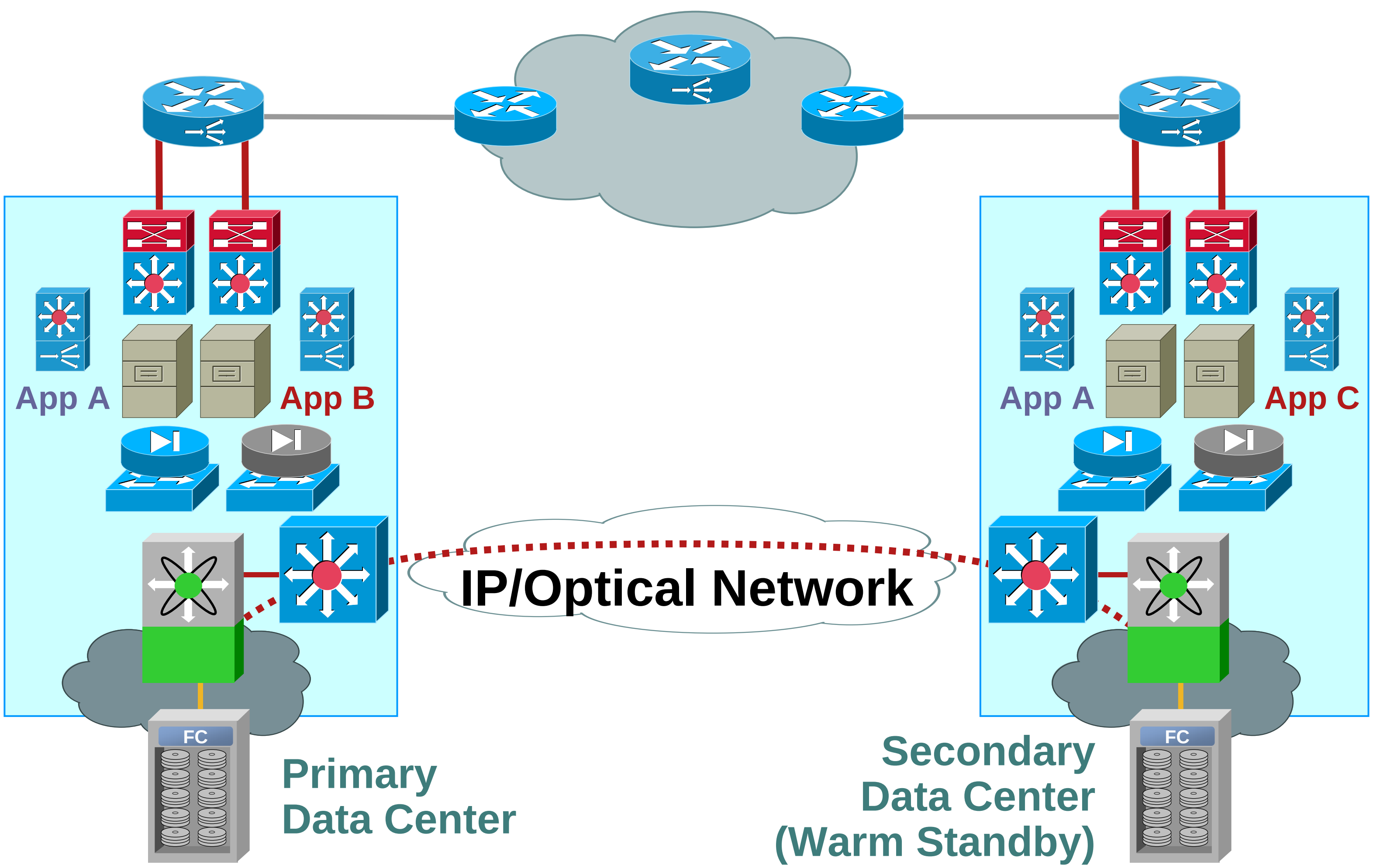
Disaster Recovery—Active/Standby



Warm Standby

- A data center that is equipped with hardware and communications interfaces capable of providing backup operating support
- Latest backups from the production data center must be delivered
- Network access needs to be activated
- Application needs to be manually started
- Provides better RTO and RPO than cold standby backup

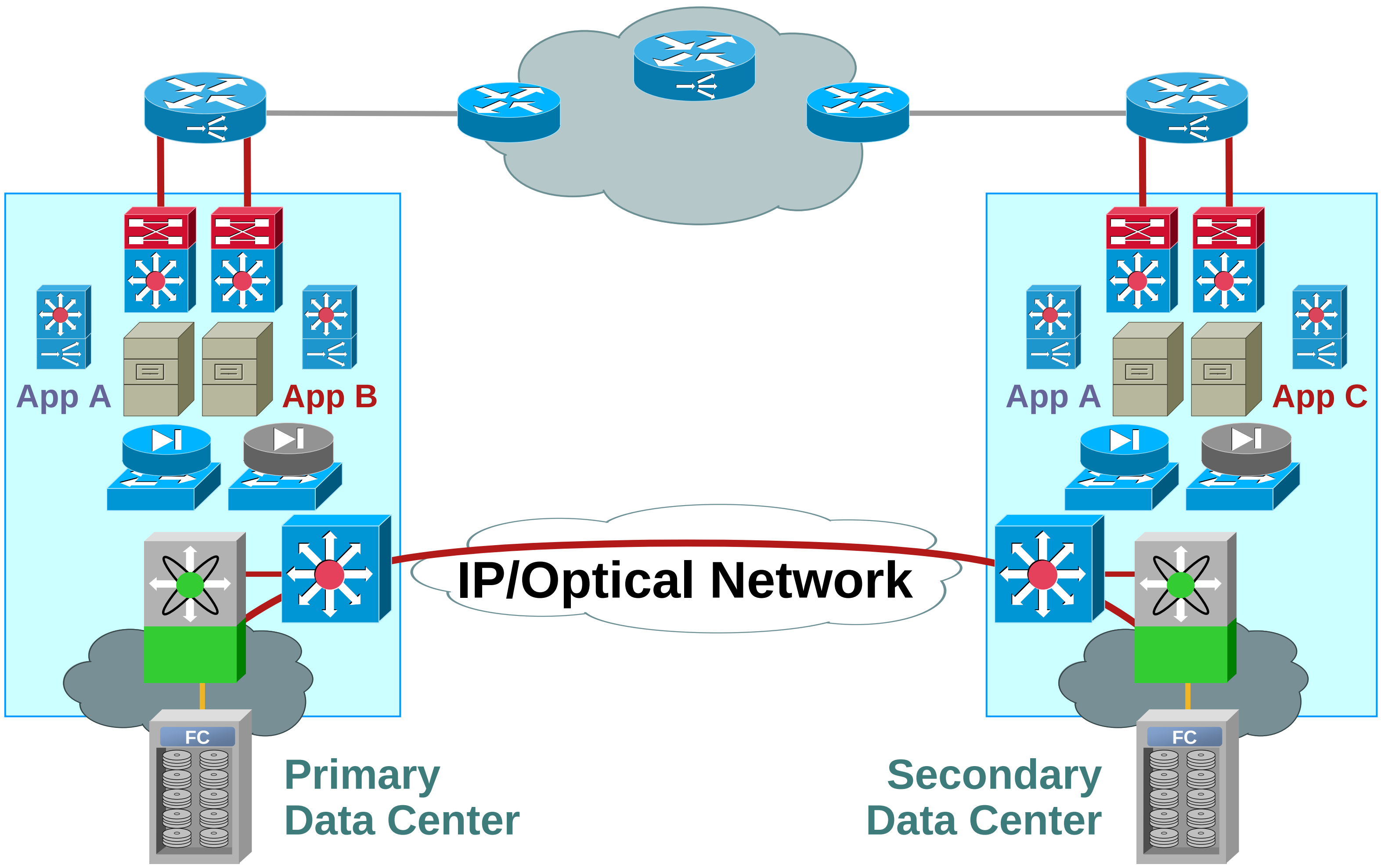
Disaster Recovery—Active/Standby



Hot Standby

- A data center that is environmentally ready and has sufficient hardware, software to provide data processing service with little down time
- Hot backup offers disaster recovery, with little or no human intervention
- Application data is replicated from the primary site
- A hot backup site provides better RTO/RPO than warm standby but cost more to implement
- Business continuance

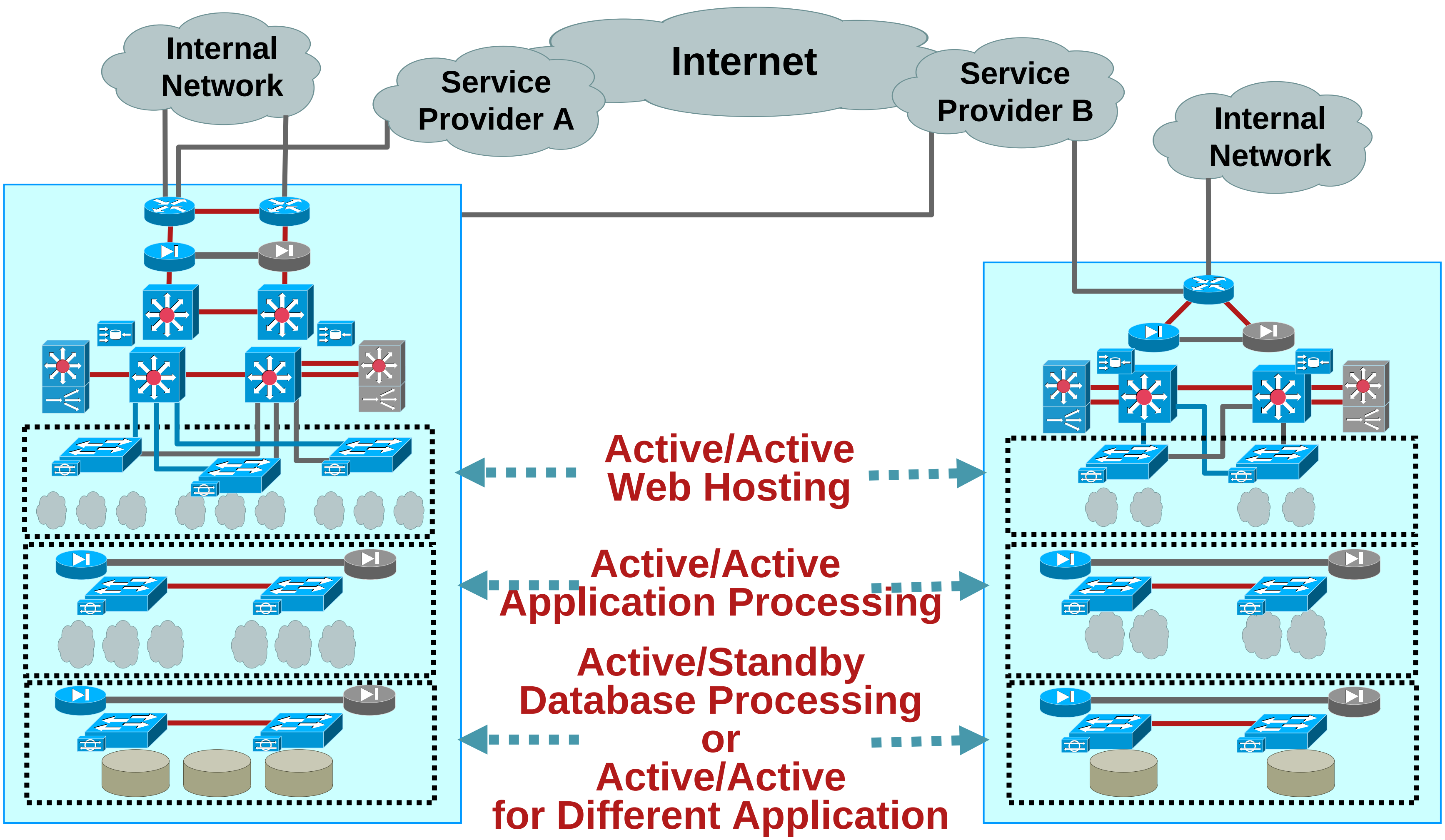
Disaster Recovery—Active/Standby



Disaster Recovery—Active/Active

What Does Active/Active Mean?

Active/Active Data Centers



Components of Disaster Recovery

Agenda

- Introduction to Data Center—The Evolution
- Application and Business Continuance
  - Increasing HA in the Data Center
  - HA with Virtualisation
- Data Center Disaster Recovery
  - Failure Scenarios
  - Design Options
- Components of Disaster Recovery
  - Site Selection—Front End GSLB
  - Server High Availability—Clustering

Site Selection Mechanisms

- Site selection mechanisms depend on the technology or mix of technologies adopted for request routing:

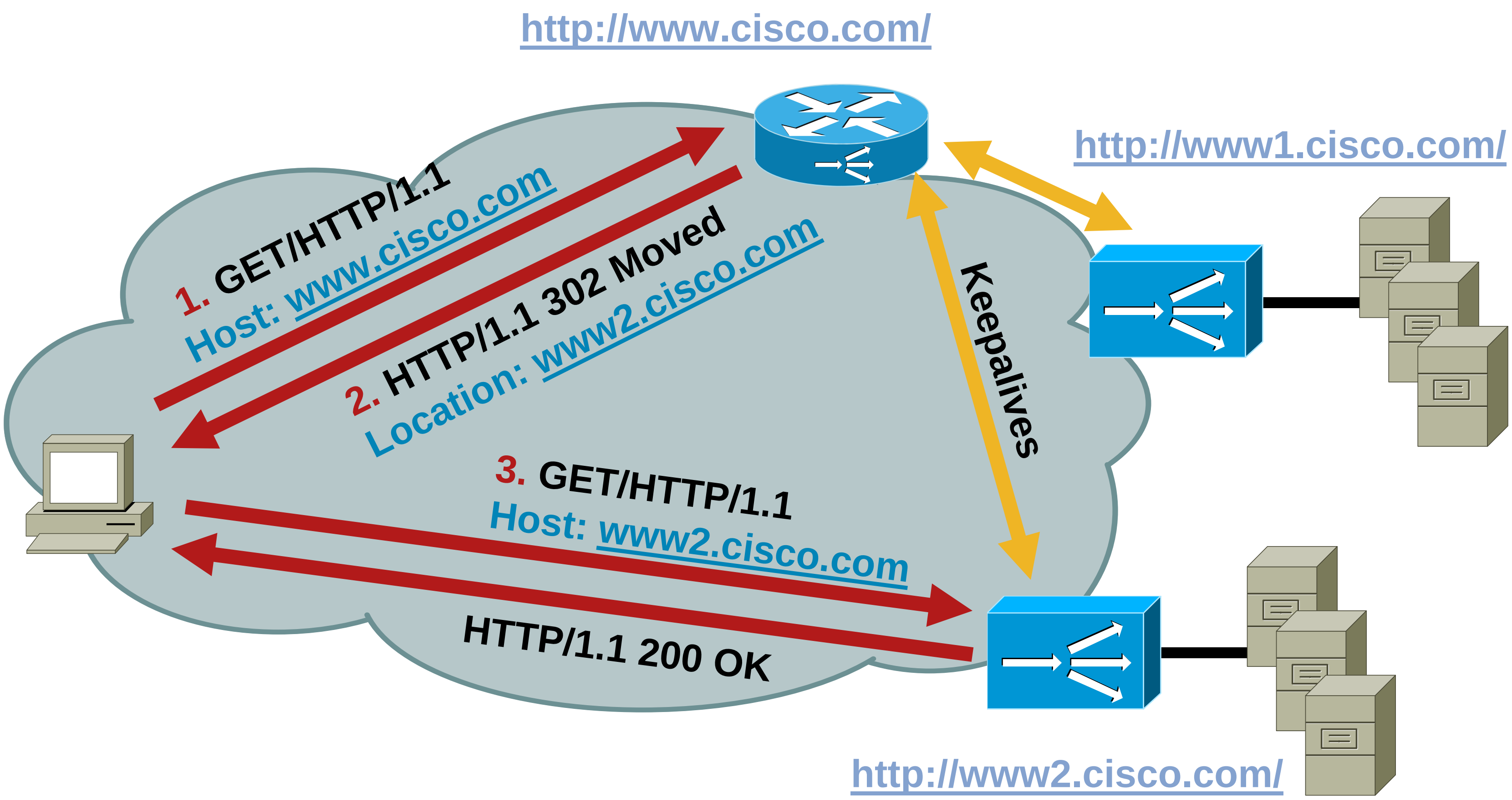
  1. HTTP redirect
  2. DNS-based
  3. L3 Routing with Route Health Injection (RHI)

- Health of servers and/or applications needs to be taken into account
- Optionally, other metrics (like load) can be measured and utilized for a better selection

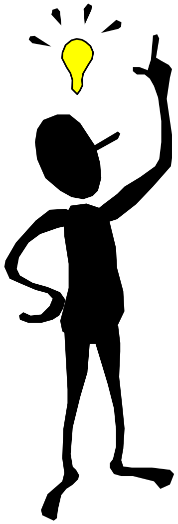
HTTP Redirection—The Idea

- Leveraging the HTTP redirect function: HTTP return code 302
- Proper site selection made after the initial DNS request has been resolved, via redirection
- Mainly as a method of providing site persistence while providing local server farm failure recovery

HTTP Redirection—Traffic Flow



Advantages of the HTTP Redirection Approach



- Can be implemented without any other GSLB devices or mechanisms
- Inherent persistence to the selected location
- Can be used in conjunction with other methods to provide more sophisticated site selection

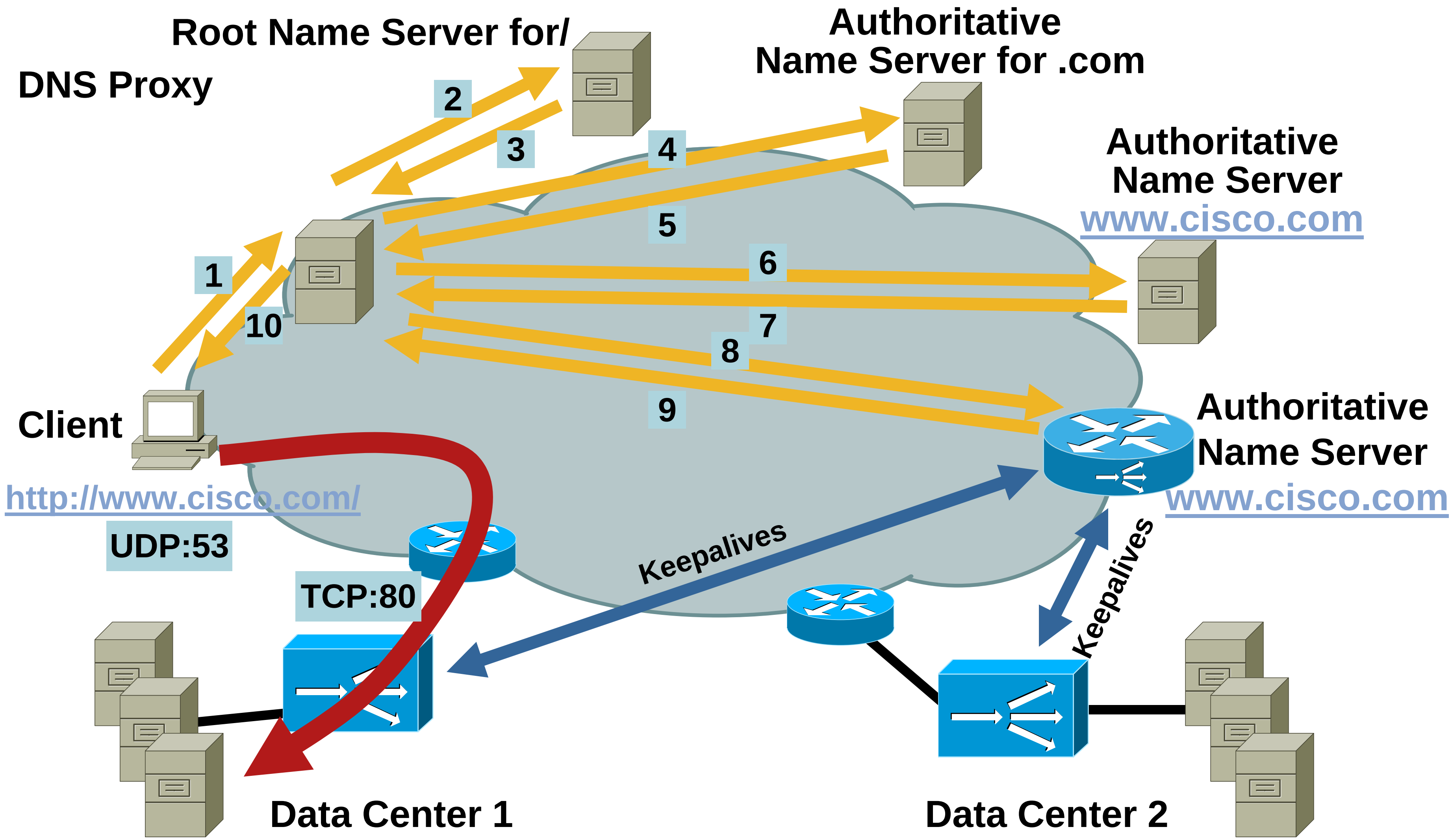
Limitations of the HTTP Redirection Approach

- It is protocol specific—relies on HTTP
- Requires redirection to fully qualified additional names—additional DNS records
- Users may bookmark a specific location—loosing automatic failover
- HTTPS redirect requires full SSL hand shake to be completed first

DNS-Based Site Selection—The Idea

- The client D-proxy (local name server) performs iterative queries
- The device which acts as “site selector” is the authoritative name server for the domain(s) distributed in multiple locations
- The “site selector” sends keepalives to servers or server load balancer in the local and remote locations
- The “site selector” selects a site for the name resolution, according to the pre-defined answers and site load balance method
- The user traffic is sent to the selected location

DNS-Based Site Selection—Traffic Flow



Advantages of the DNS Approach

- Protocol independent: works with any application that uses name resolution
- Minimal configuration changes in the current IP and DNS infrastructure (DNS authoritative server)
- Implementation can be different for specific host names
- A-records can be changed on the fly
- Can take load or data center size into account
- Can provide proximity

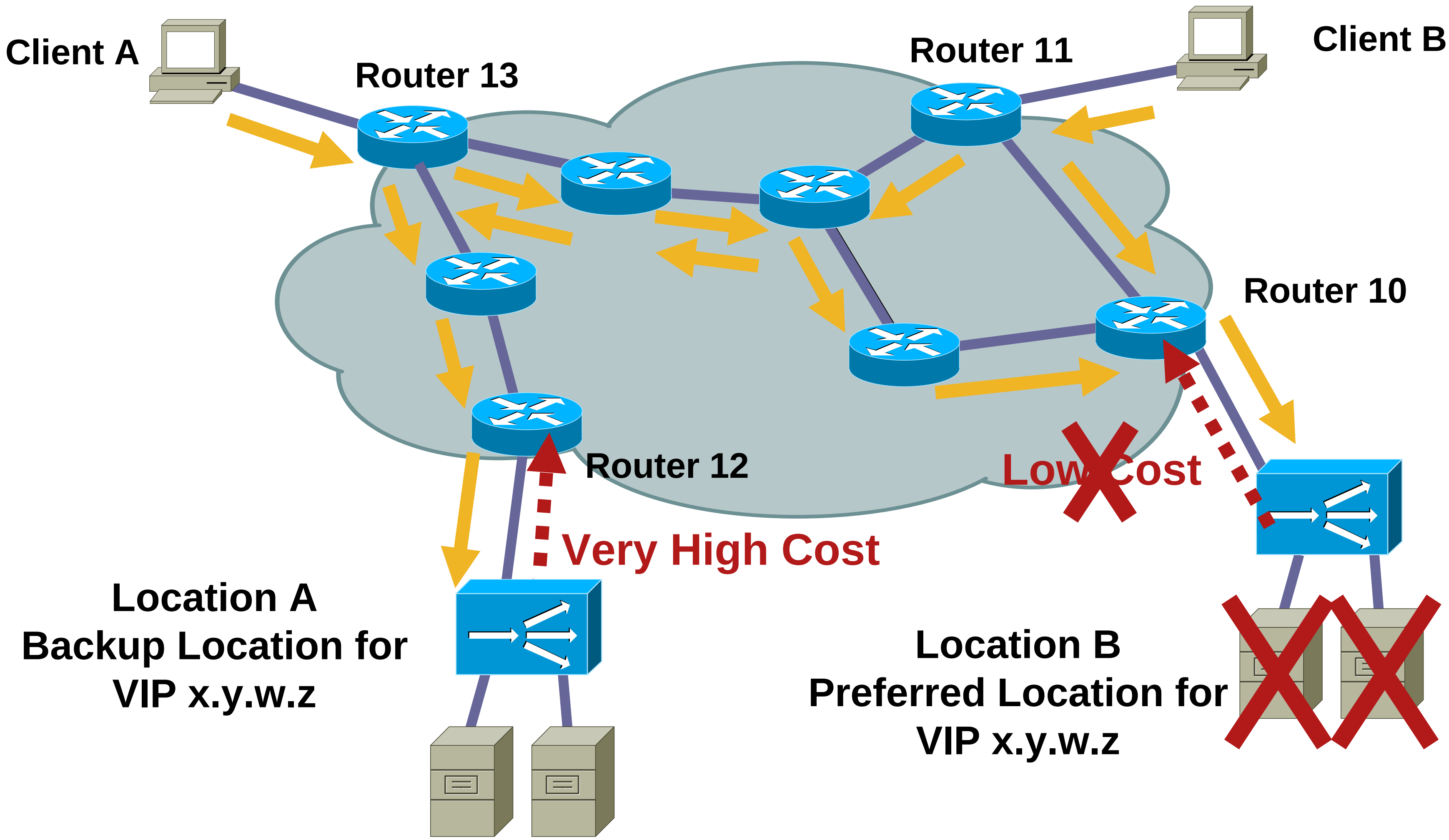
Limitations of the DNS-Based Approach

- Visibility limited to the D-proxy (not the client)
- Can not guarantee 100% session persistency
- DNS caching in the D-proxy
- DNS caching in the client application
- Order of multiple A-record answers can be altered by D-proxies

Route Health Injection—The Idea

- Server and application health monitoring provided by local server load balancers
- SLB can advertise or with draw VIP address to upstream routing devices depending on the availability of the local server farm
- Same VIP addresses can be advertised from multiple data centers—IP Anycast
- Relying on L3 routing protocols for route propagating and content request routing
- Disaster Recovery provided by network convergence

Route Health Injection—Implementation



Advantages of the RHI Approach

- Supports legacy application and does not rely on a DNS infrastructure
- Very good re-convergence time, especially in Intranets where L3 protocols can be fine tuned appropriately
- Protocol-independent: works with any application
- Robust protocols and proven features

Limitations of the RHI Approach

- Relies on host routes (32 bits), which cannot be propagated all over the internet
- Requires tight integration between the application-aware devices and the L3 routers
- Inability to intelligently load balance among the data centers

Agenda

- Introduction to Data Center—The Evolution
- Application and Business Continuance
  - Increasing HA in the Data Center
  - HA with Virtualisation
- Data Center Disaster Recovery
  - Failure Scenarios
  - Design Options
- Components of Disaster Recovery
  - Site Selection—Front End GSLB
  - Server High Availability—Clustering & L2 extension

Extended L2 deployment scenarios

- Migration purposes:
  1. Legacy Applications where the IP parameters can not be easily modified.
  2. Move a portion of the farm
- Geoclusters or geo dispersed HA Clusters
  1. Heartbeat
  2. VIP
- Geographically dispersed Network Services
  1. Statefull Failover
  2. Conns and Sticky Replication

Migration Purposes
Legacy Applications

- Many applications have been written for mainframes
- They have been used for many years ago
  - Airline reservation systems
  - Trading systems
  - ATMs, etc…
- Moving systems running such applications to a new facility lead to avoid to "readdress" the machines due to:
  - Complexity
  - Business continuance
  - Lack of knowledge for such Application changes (Hardcoded IP address)
- For these legacy Applications it is necessary to extend the layer 2 network between the original Data Center to the new one for the time it takes to migrate those servers to the new location.

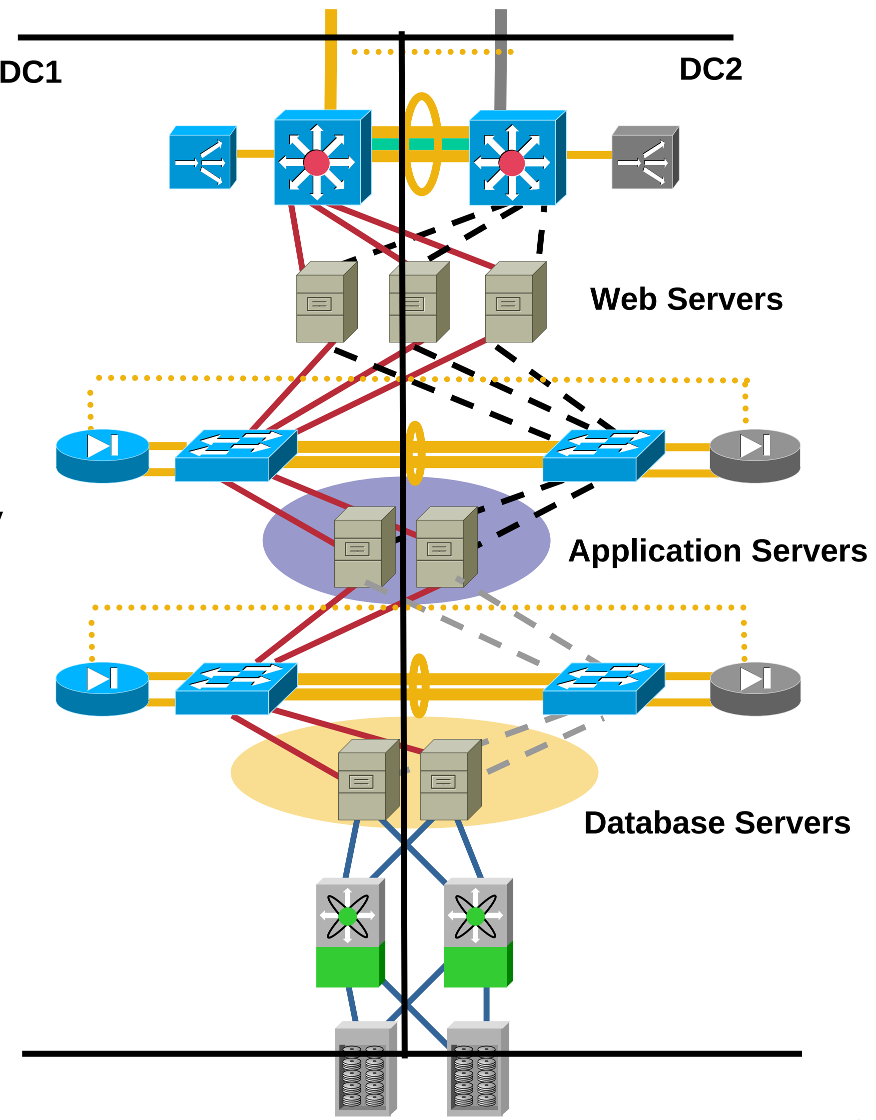
Distributed HA Data Center

Redundant Hardware Across Data Centers

- The picture shows the logical view of the Cisco multi-tier design Data Center divided in two, and run redundant hardware in separate floors, buildings or geographical regions.

- Deploy redundant network devices by placing one network device in one site and its peer network device in a remote site.

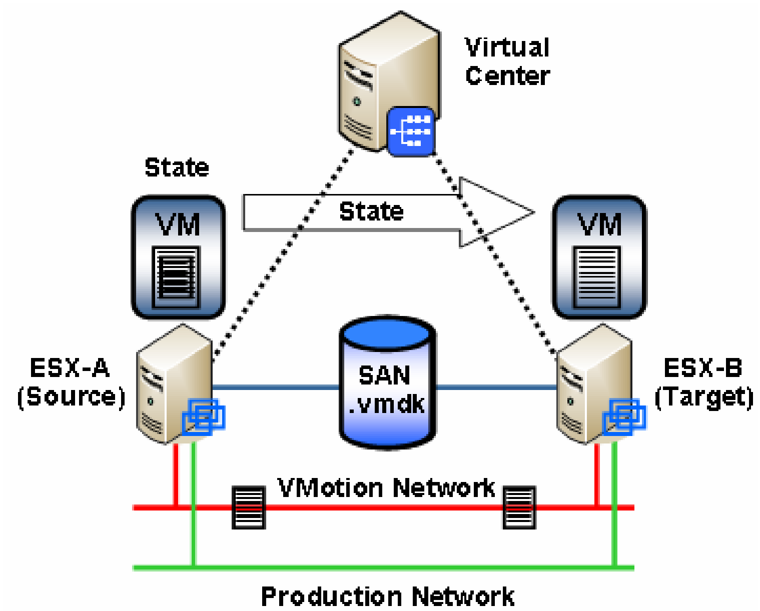
- The HSRP, OSPF, Firewalls, Load Balancers… hearbeats need to be carried across.



Virtual Machines
VMWARE and VMotion Requirements

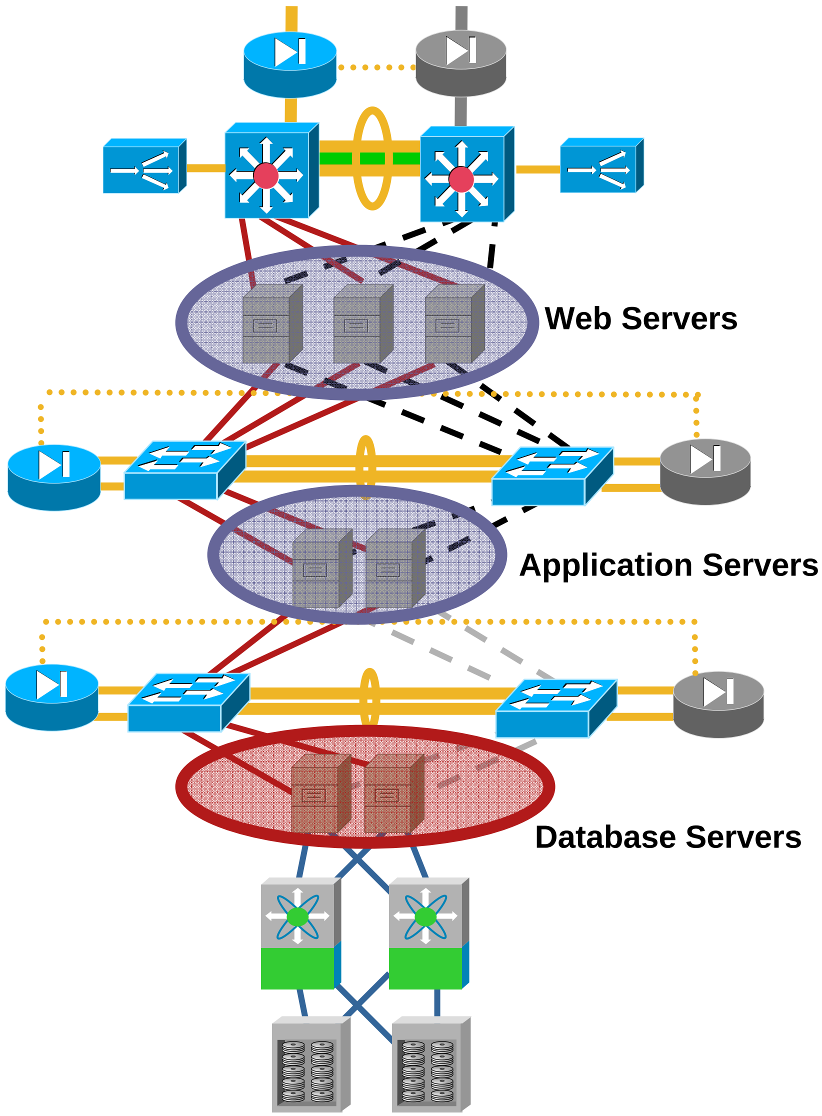
VMotion

➢ Method used by VMWares's ESX Server to migrate active virtual machines (VMs) within an ESX server farm from one physical ESX host to another.
➢This is the foundation of several high availability features provided in VMWare's Virtual Insfrastructure product.
➢ Allows the movement of active VMs with minimal downtime.
➢ Server administrators may schedule or initiate the VMotion process manually through the VMware VirtualCenter management tool.



Cluster Overview

- A cluster is two or more servers configured to appear as one
- Two types of clustering: Load balancing (LB) and High Availability (HA)
- Clustering provides benefits for availability, reliability, scalability, and manageability
- LB clustering: multiple copies of the same application against the same data set, usually read only
- HA clustering: multiple copies of long running application that requires access to a common data depository, usually read and write, running on same hardware and OS



HA Cluster/GeoCluster
Requirements

- * Microsoft MSCS
- * Veritas Cluster Server (Local)
- Solaris Sun Cluster Enterprise
- VMware Cluster (Local)
- Oracle RAC (Real Appl.Cluster)
- HP MC/ServiceGuard
- HP NonStop
- HP Open VMS/TruCluster
- IBM HACMP
- EMS/Legato Automated Availability Mgr

Common Functions

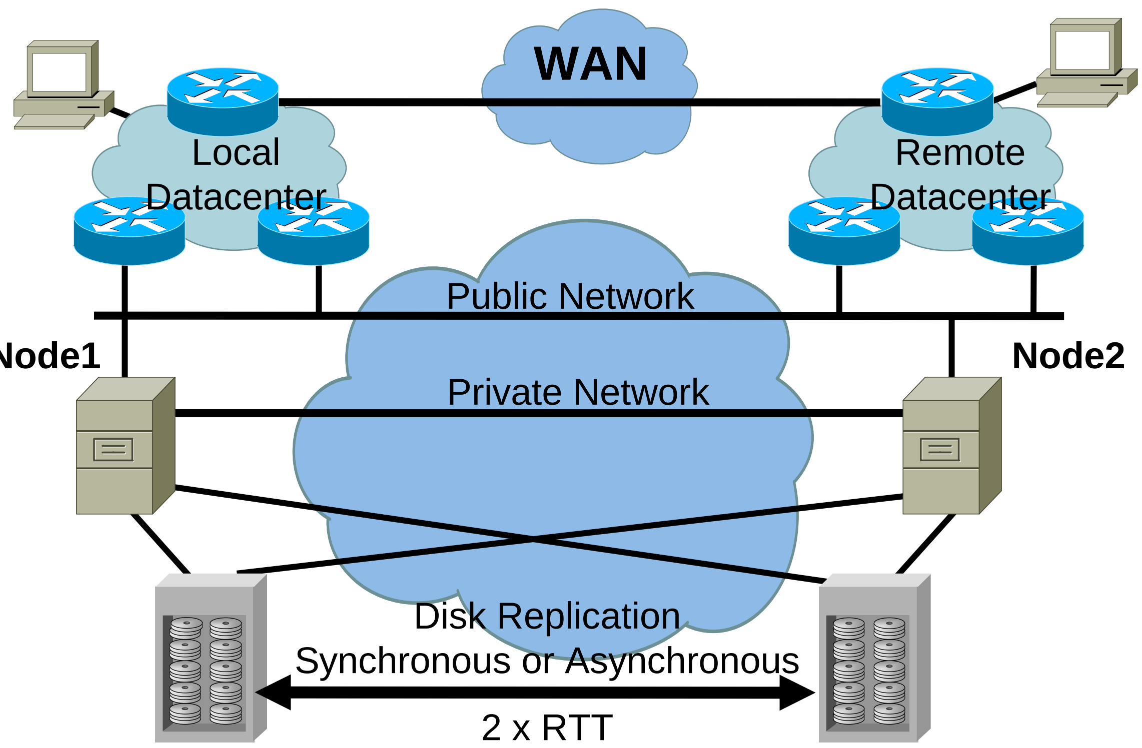
- VIP address on both nodes
- Extended L2 VLAN
- Dedicated L2 used for heartbeat & performance control
- Quorum Disk
- Software is unaware of extended members of cluster

*** Veritas VCS offers an extended Cluster solution using L3 for inter-site connectivity**
*** Next release of MS Longhorn to support L3 site to site.**

Geo-Clusters

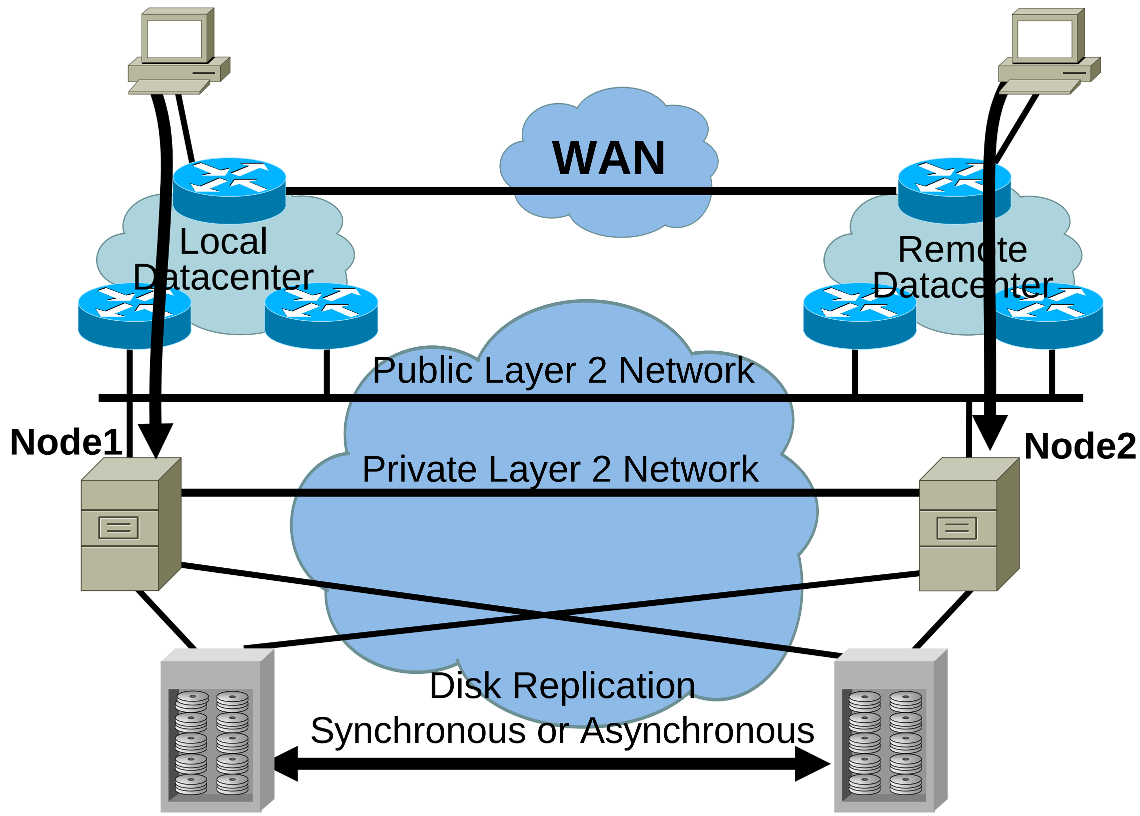
Geo-Cluster: Cluster That Span Multiple Data Centers

- Public Network (typically Ethernet) for client /Application requests
- Private Network (typically Ethernet) for interconnection between nodes; could be direct connect, or optionally going through the public network
- Storage Disk (typically Fiber) shared storage array, NAS or SAN

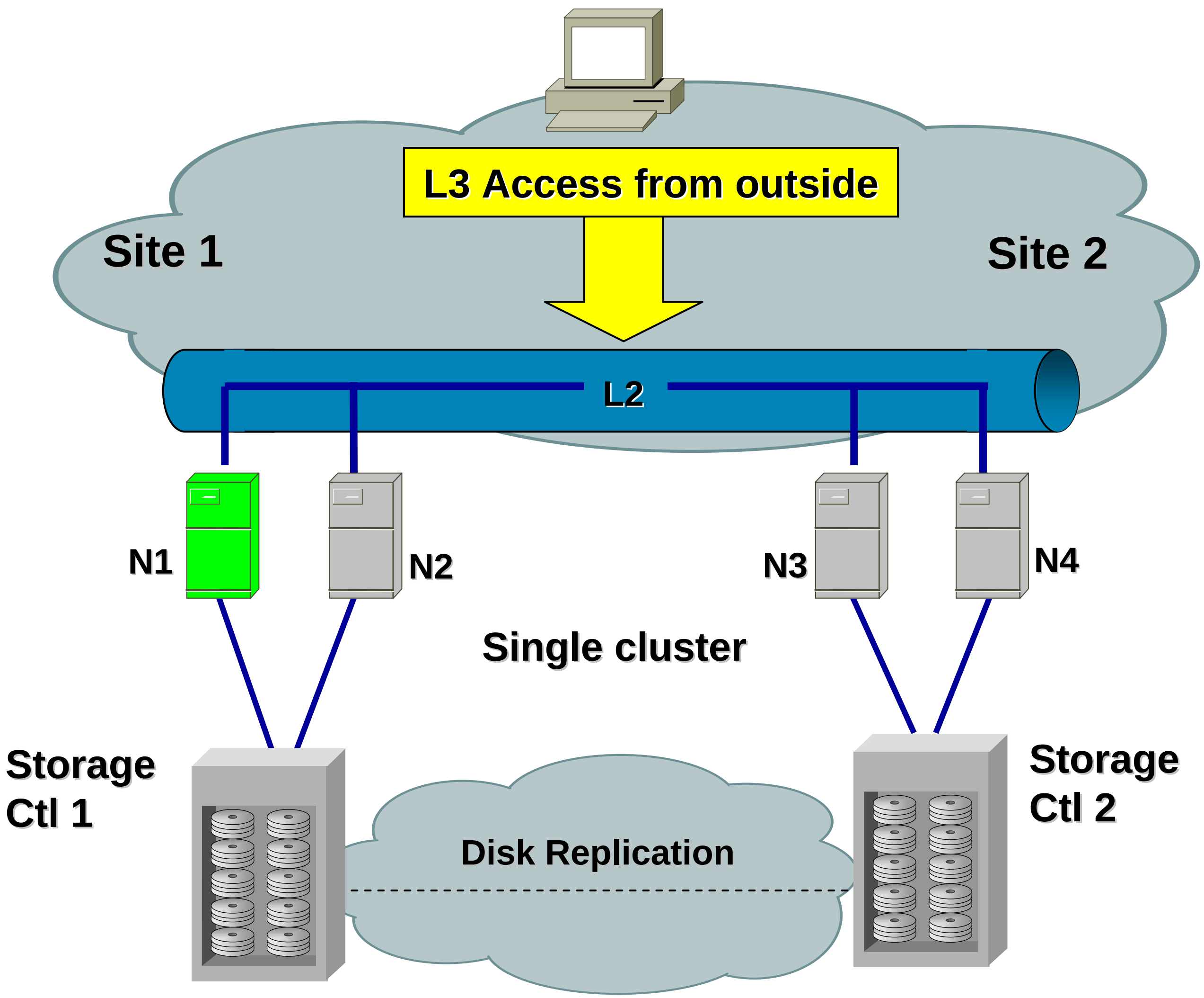


Extended Layer 2 Network

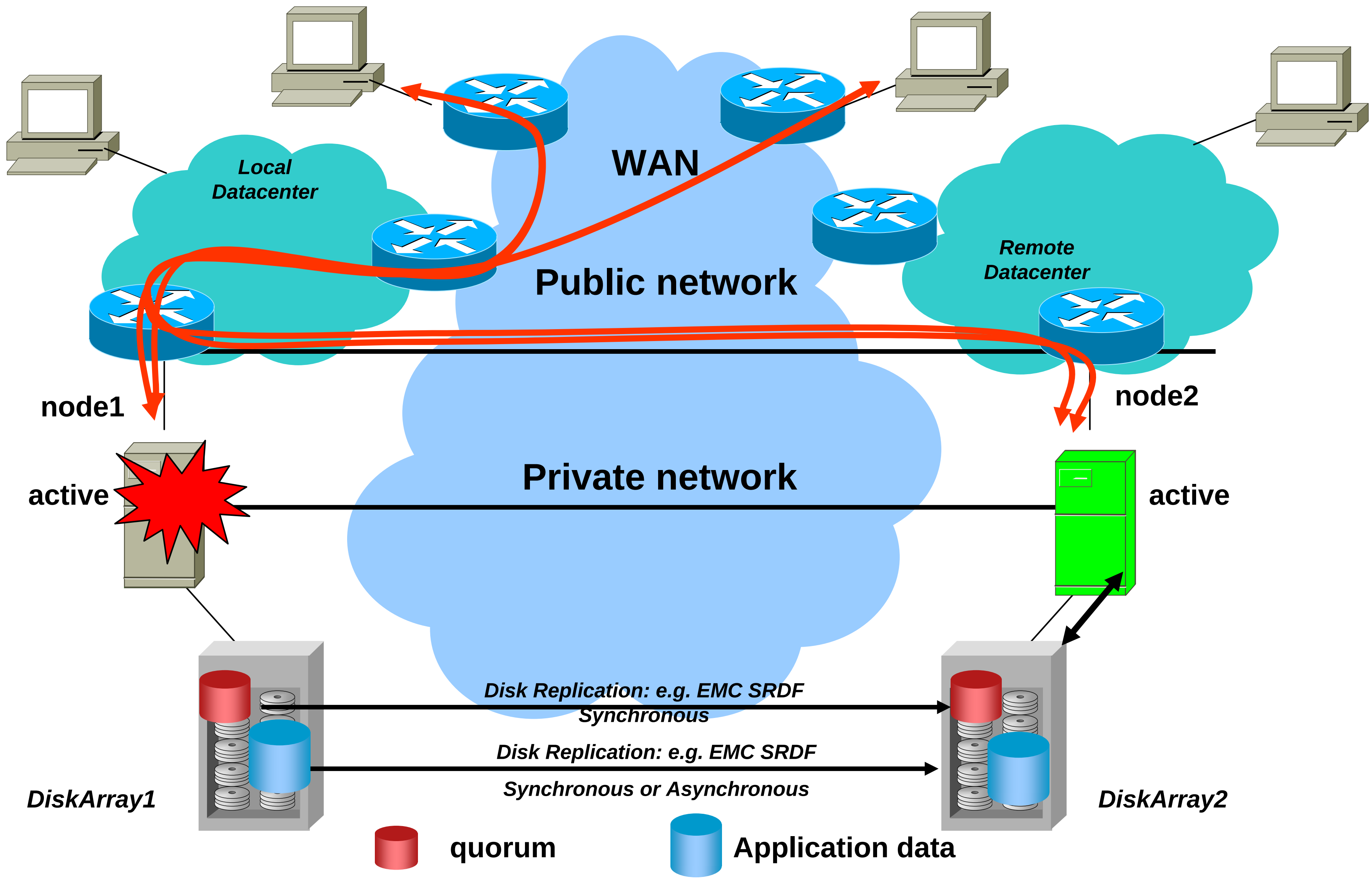
- In most implementation, a common L2 network is needed for the heartbeat between the nodes, as well as public client access (Cluster VIP)
- Extending VLAN on a geographical basis is not considered best practice because of the impact of broadcasts, multicast, flooding and Spanning-Tree integration issues



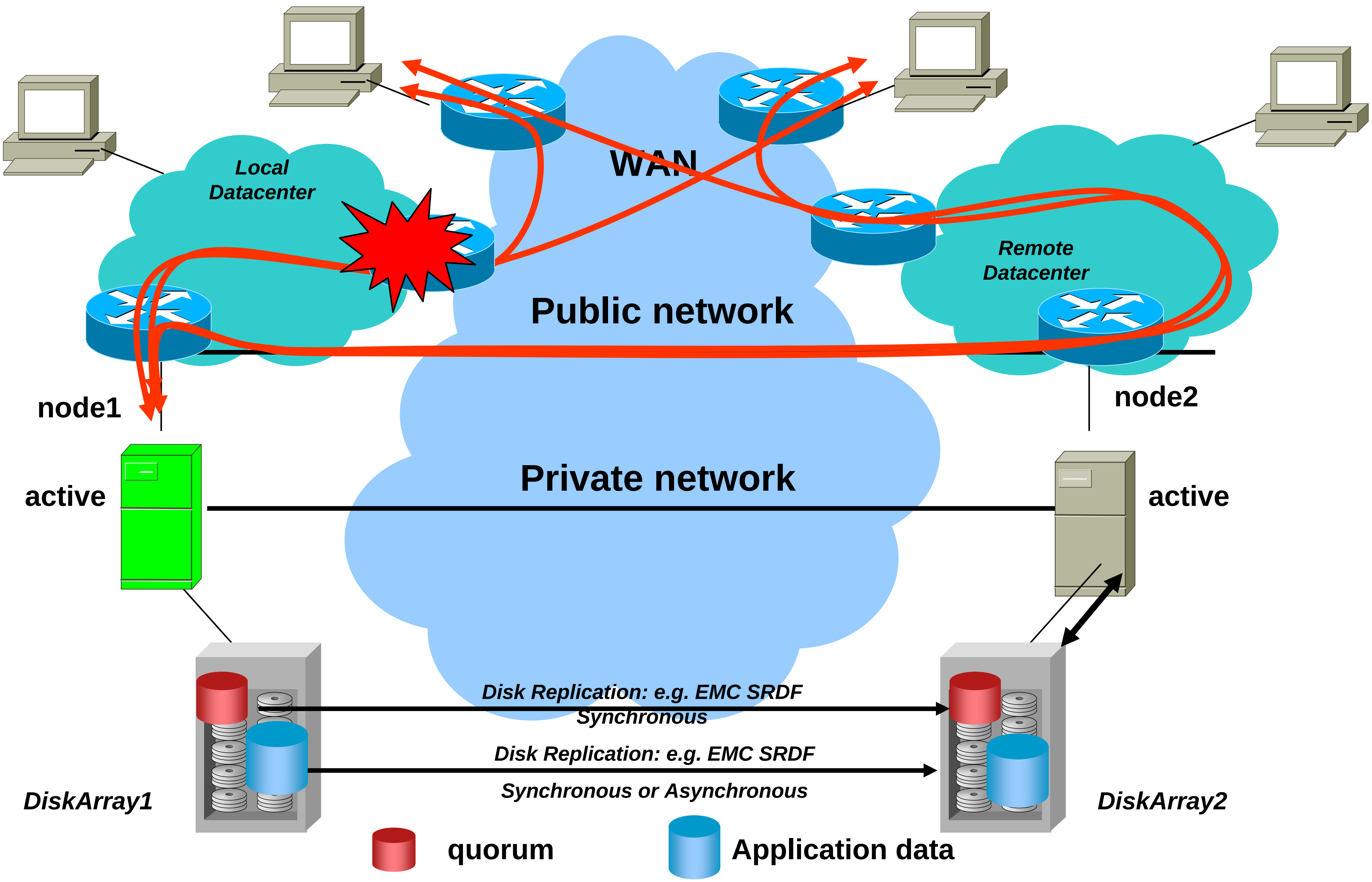
Geographically Dispersed Cluster
Logical Architecture



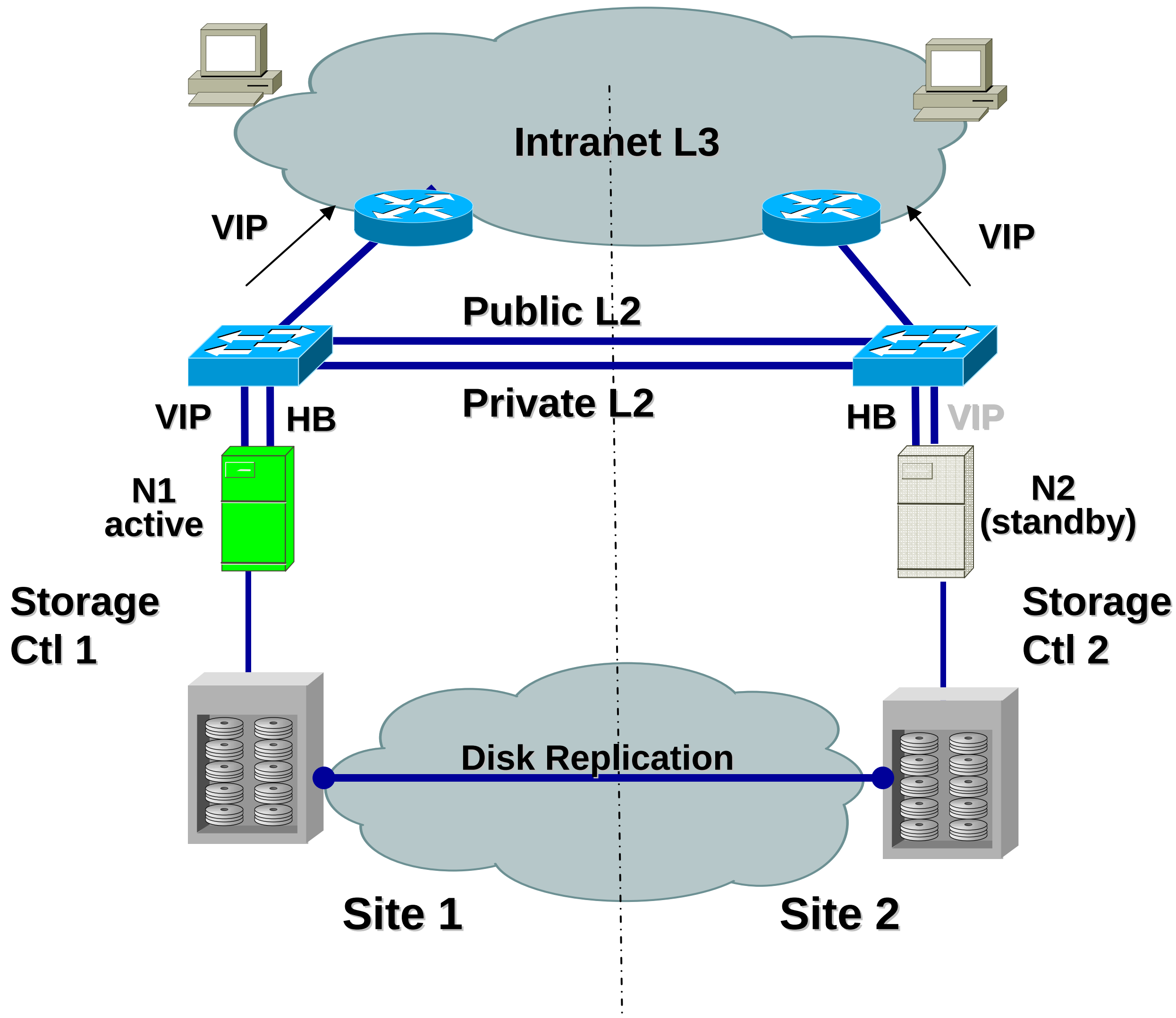
Routing in Presence of Failures
Node Failure



Routing in Presence of Failures
Wan Access *Failure*



Geographically Dispersed Cluster
Problematic



Requirements:

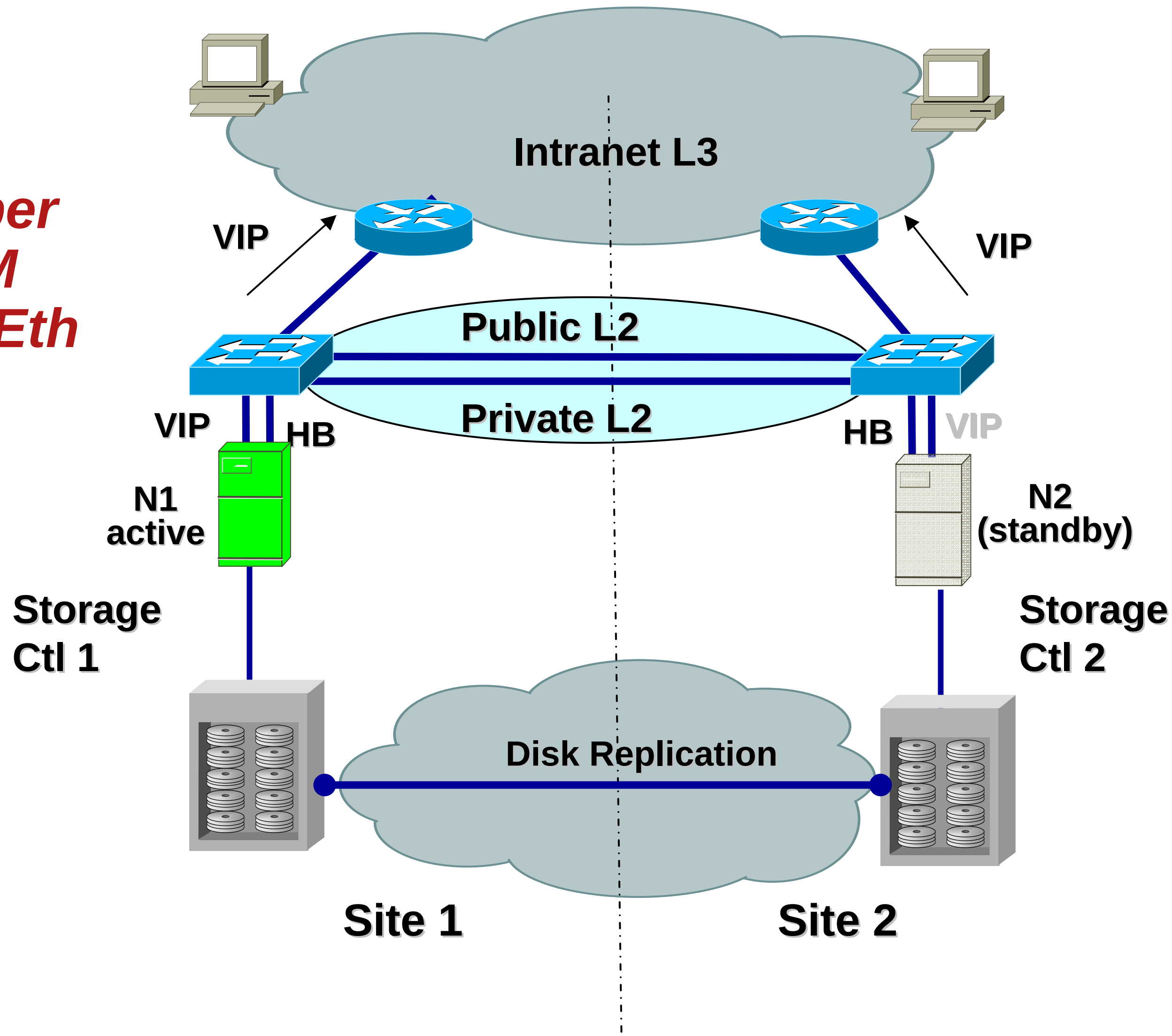
L2 node to node
(VIP + Bck-up HB)
100% resilient

Extended L2 options:

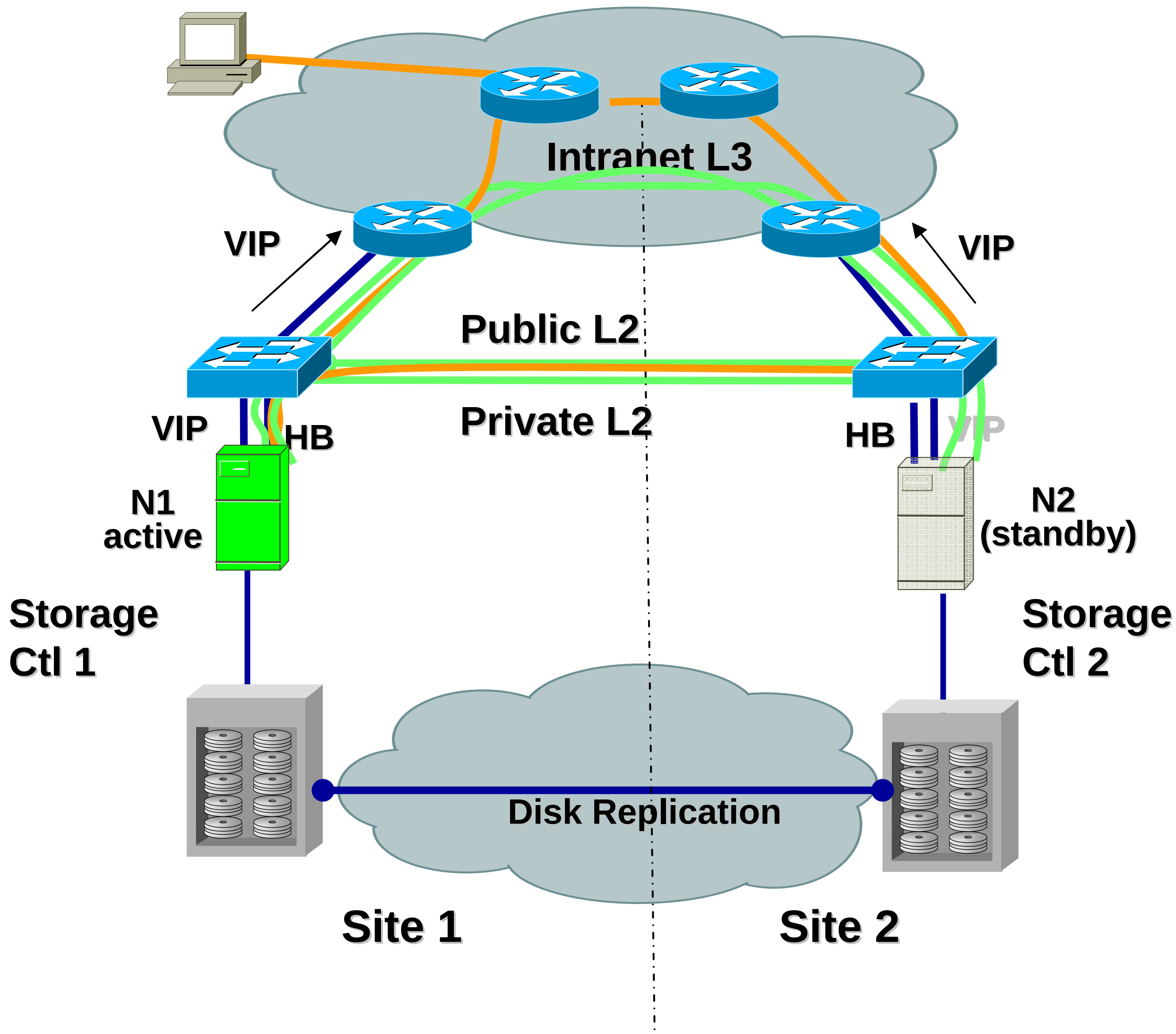
1. Dedicated Fiber
   - Dark Fiber
   - Gig Ethernet
   - WDM + STP
   - WDM + MPLS
2. Mix L2/L3 trunk
   - STP
3. Intranet MPLS DIY
   - L2/L3 VPN
   - VPLS
4. Intranet SP
   - L2VPN
   - CsC
   - L2TPv3

1 – Geographically Dispersed Cluster
Dedicated Fiber

- ***Dark Fiber***
- ***D/CWDM***
- ***Gigabit Eth***



2 – Geographically Dispersed Cluster
Mixed *L2/L3 trunk*

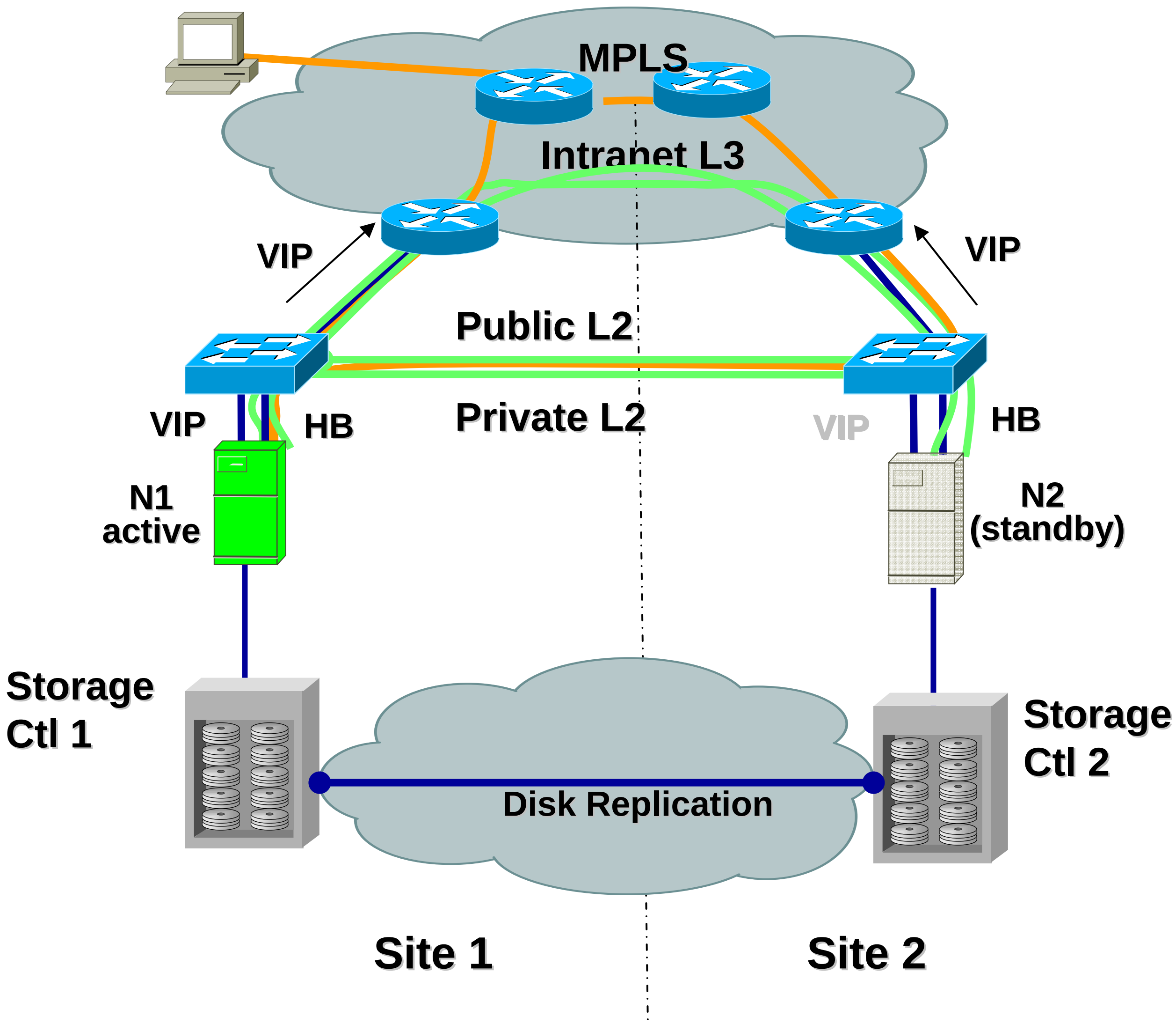


2 – Geographically Dispersed Cluster (cont)
Mixed L2/L3 trunk - Pros & Cons

- Pros:
  - If point to point only ➔ it may be as stable as WDM
  - If limited to a single Cluster, it should be ok (otherwize use QinQ to limit the number of STP instances)
  - No need for extra cost
- Cons:
  - Extended STP: Historically hasn't proved to be a stable solution (RSTP doesn't bring much added value)
  - Could be difficult to deploy - depends on number of multi-hops (L3)
  - Requires Multilayer switches from end to end
  - Open the door to other extended L2 applications
  - Implies the customer owns the Intranet L3

3 - Geographically Dispersed Cluster (cont)

Intranet MPLS DIY (self-deployed)



3 - Geographically Dispersed Cluster (cont)
Intranet MPLS DIY – several options

- EoMPLS **(Ethernet over MPLS)**
  - Port xconnect:

    Aka ELS (Ethernet Line Service)

    point to point port emulation accross MPLS
  - Internal-VLAN xconnect

    Mix any VRF, VLAN accross MPLS
- VPLS **(Virtual Private LAN Services)**

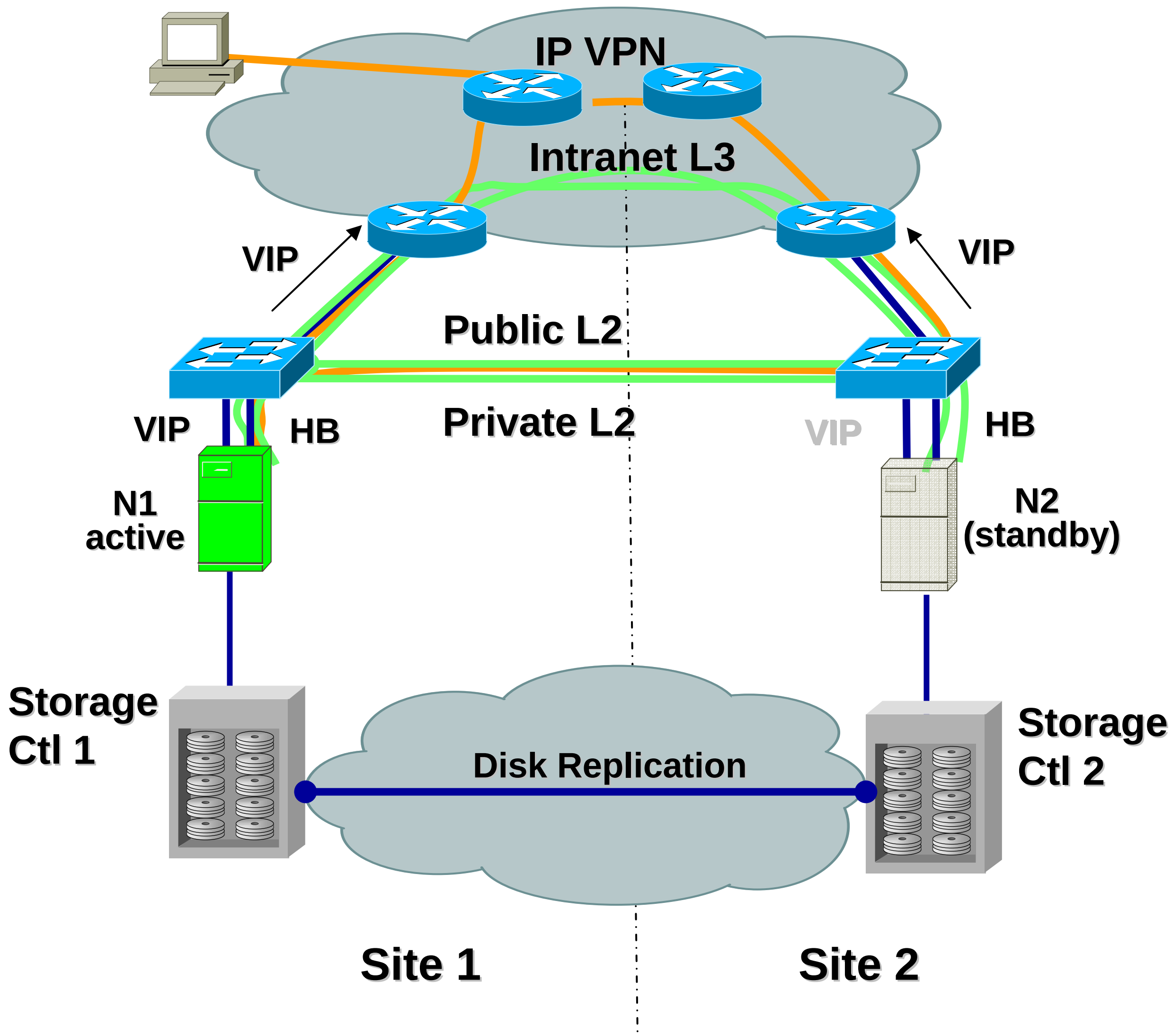
  Aka EMS (Ethernet MultiPoint Service)

  Multi-sites, dynamic Mp2Mp (L2 VLAN like)

3 - Geographically Dispersed Cluster (cont)
MPLS advantages in datacenter interconnection

- Core is any type of links (GE / POS)
- Core Links are MPLS L3 Fast Convergence protected
  - Stable
  - can be Fast Rerouting
  - no need for Optical Protection (cost reduction)
  - no STP , loop free on the core
- Same Core can be shared for Storage / Application / User traffic
- MPLS L3 VPN allows dynamic extension of VRF between Data Center
- QoS
  - per Classes of Services or per VLAN rate limiting
  - CoS transparency (keeps original CoS from end to end)
  - Redistribute unused bandwidth
- Traffic-engineering
  - Reserve Bandwidth
  - Load repartition (RSPAN, per VLAN repartition)

4 – Geographically Dispersed Cluster (cont) *SP owned Intranet*



4 – Geographically Dispersed Cluster
Intranet thru MPLS IP-VPN (SP owned) - L2 transit

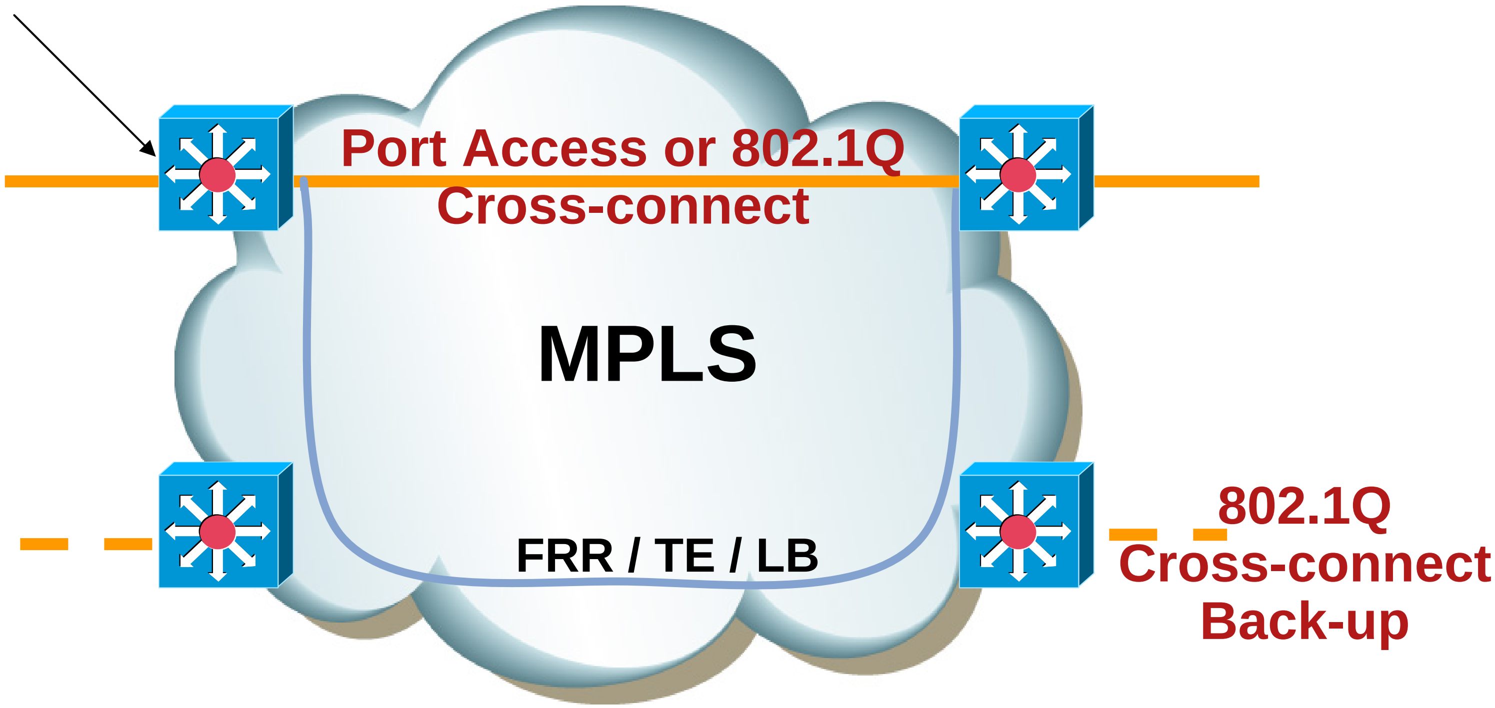
➔ SP is offering a L2 site to site transport

- Still quite rare today onto market
- Emerging and growing

1. SP provides L2VPN Ethernet
   - Xconnect VPWS or VPLS
2. SP provides MPLS access
   - Technology is CsC (Carrier supporting Carrier)
     - » Multi-points virtualized labels (hierarchy)
     - » Edge build L3 / L2 VPN over SP-labels

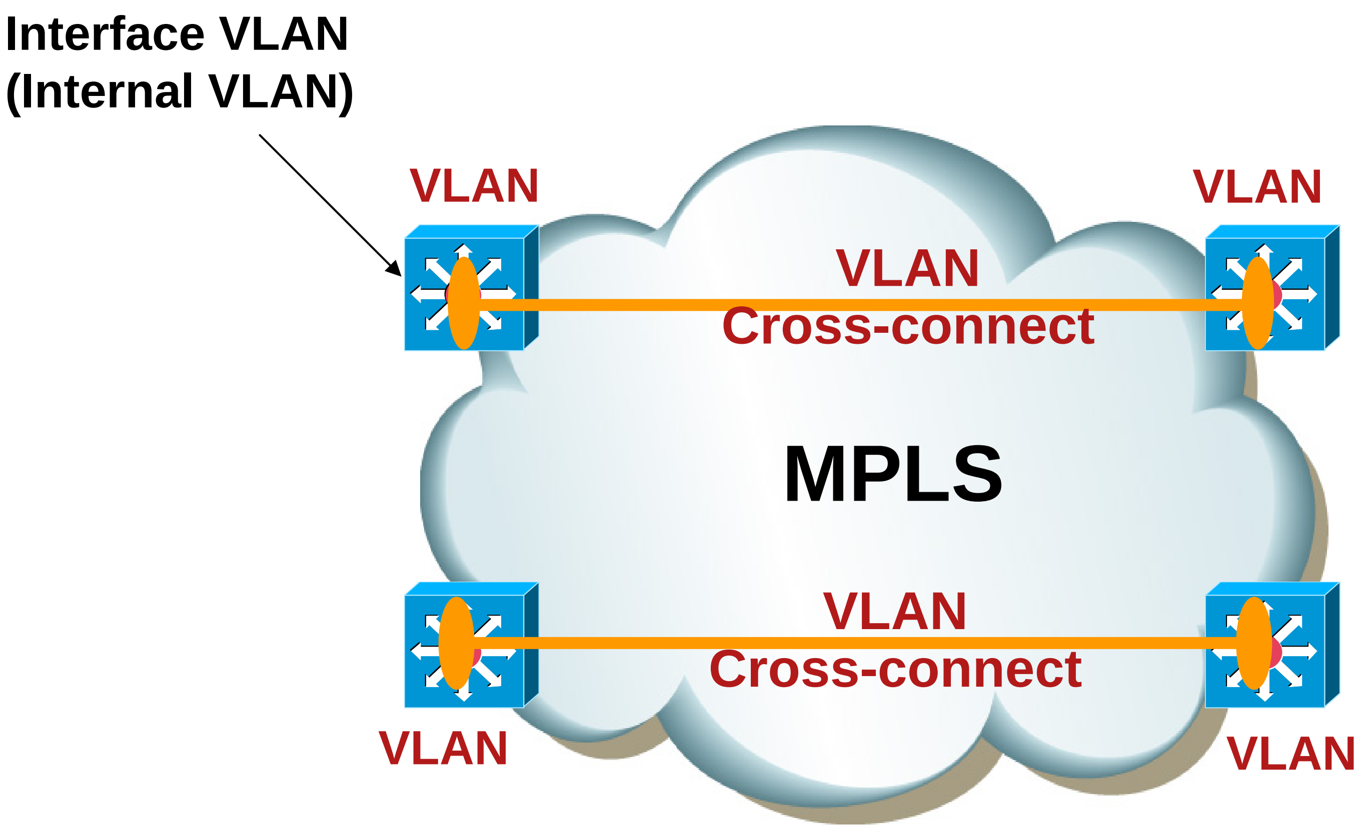
EoMPLS design model 1
Port Mode

Interface Giga n/n switch mode type
Access or Trunk



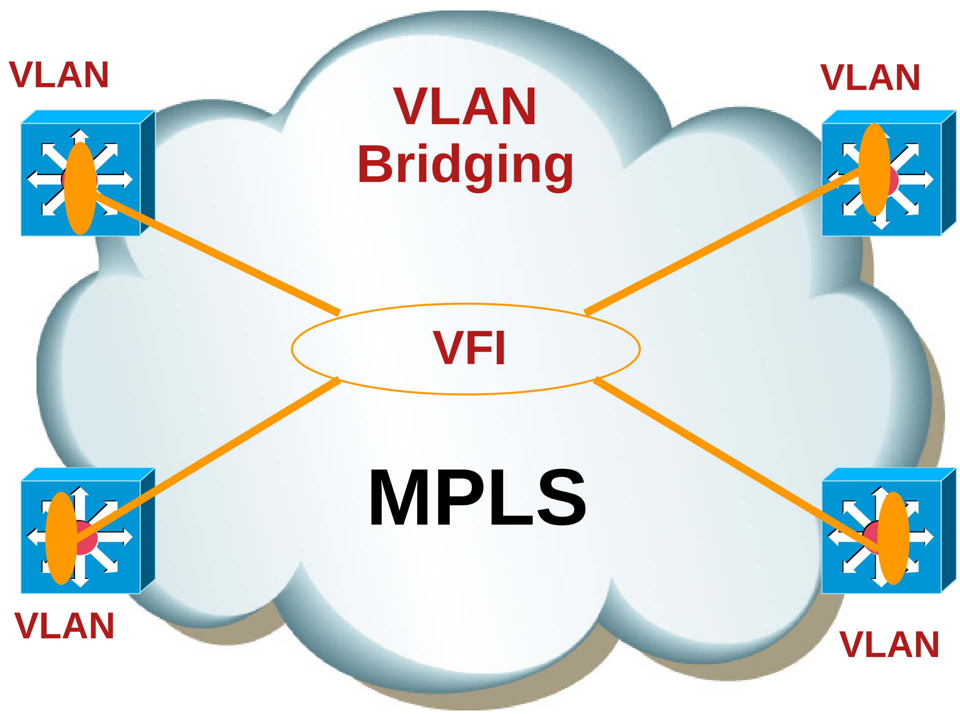
Transparent to Edge bridging
(BPDU, STP, VLAN, CoS)

EoMPLS design model 3
VLAN Mode (internal)



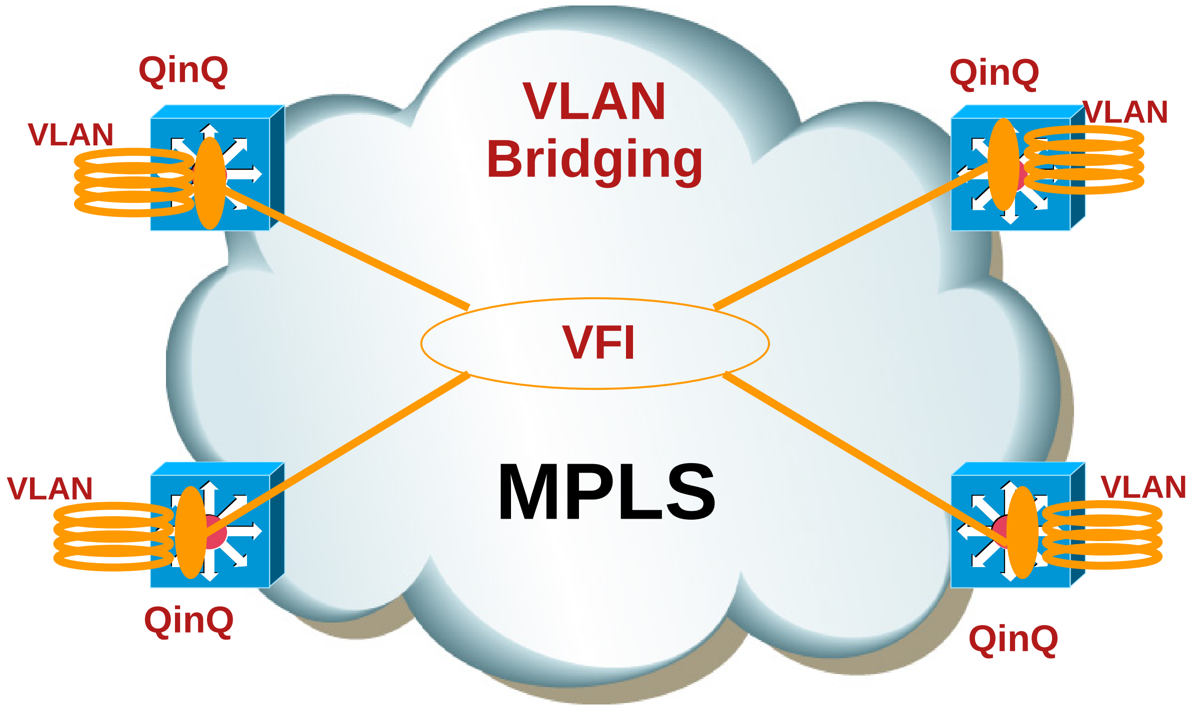
PE Can Participate to Edge bridging (BPDU, STP)
Supports CoS
Requires SIP facing Core

VPLS design model 1
Native VPLS



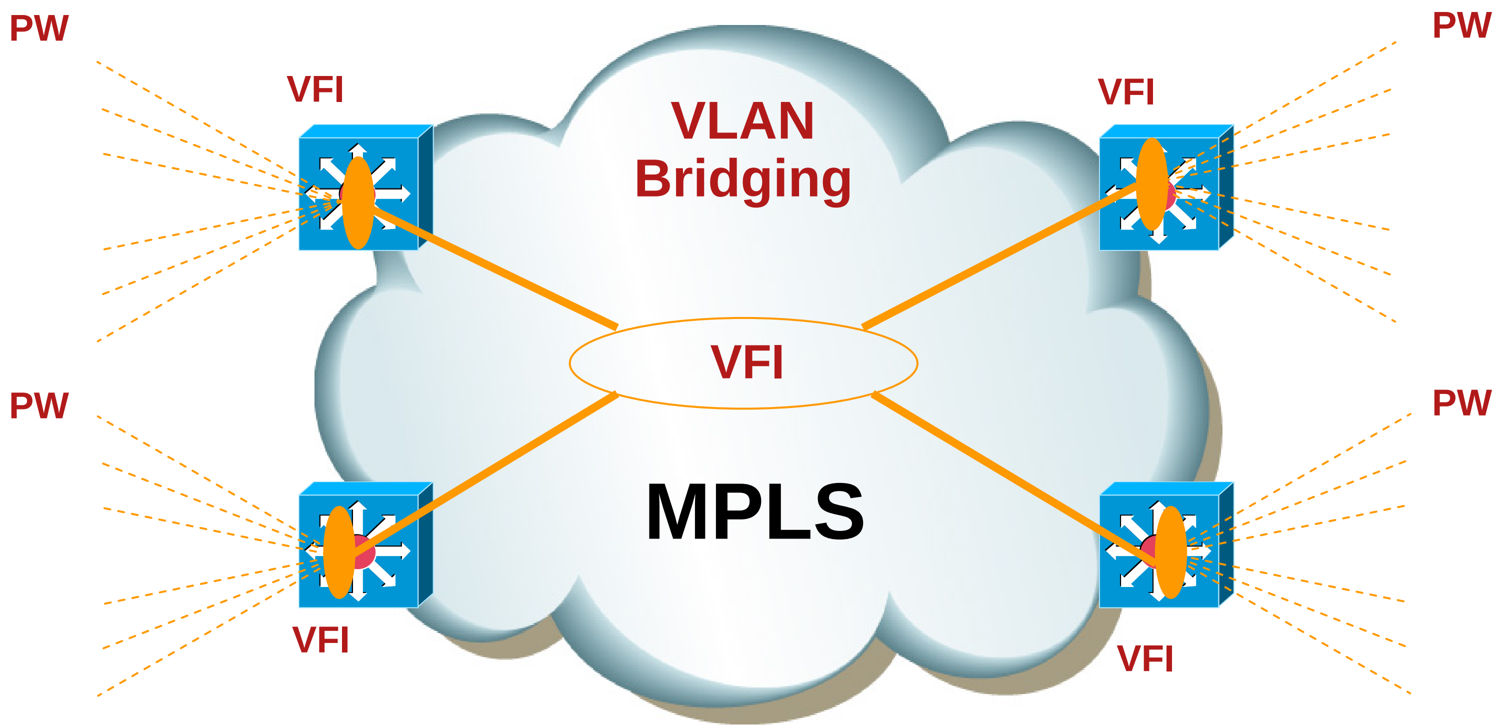
Can Participate to Edge bridging (BPDU, STP)
Supports CoS
Requires SIP facing Core

VPLS design model 2
Hierarchical-VPLS with QinQ edge



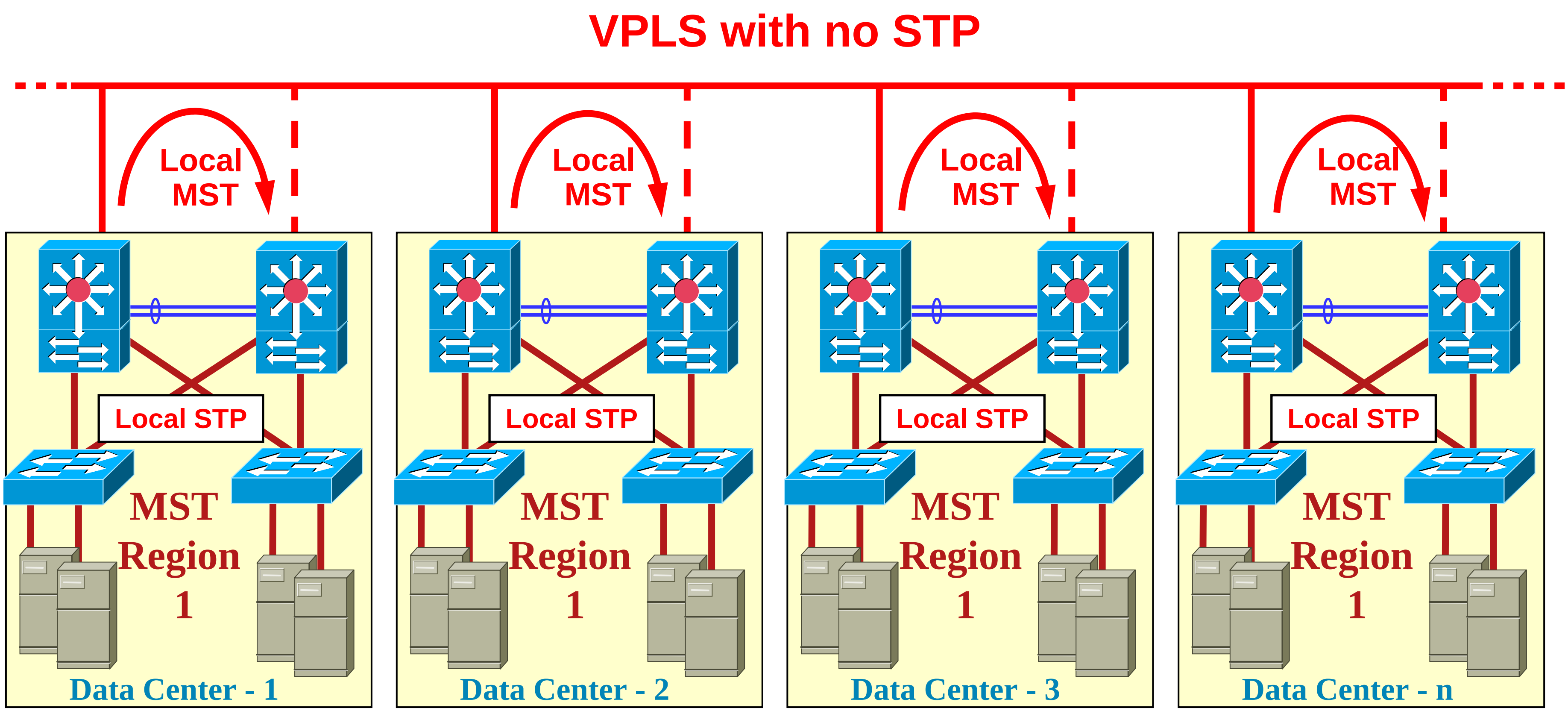
Can Participate to Edge bridging (BPDU, STP)
Supports CoS
Requires SIP facing Core
Complex support for Core multicast, QoS today…

VPLS design model 3
Hierarchical-VPLS with EoMPLS edge



Participate to Edge bridging (BPDU, STP)
Supports CoS
Requires SIP facing Core and SIP facing Edge
Quite new, still drawback to be solved (Core multicast, QoS, ..)

Loop-free thru Split-Horizon
STP design



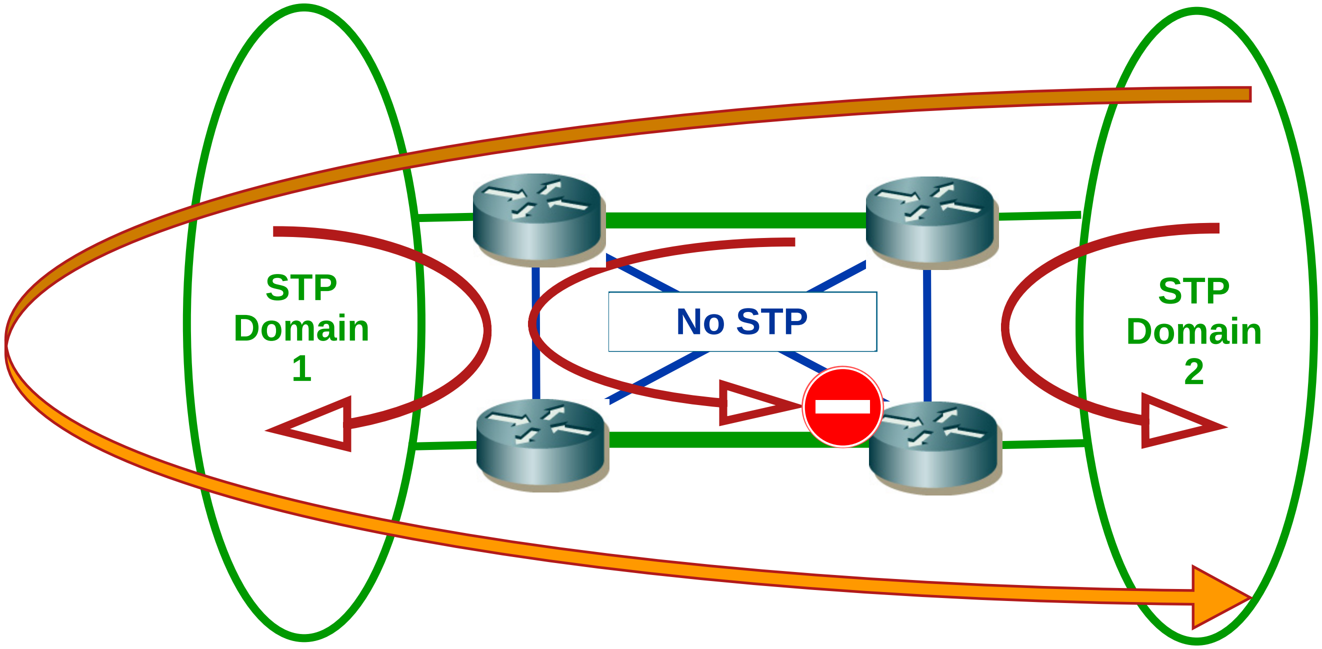
- Any intra-DC STP convergence will not force any STP convergence into other DC
- Any PE failure may lead to local STP convergence, but will not be extended to other DC

VPLS Split-Horizon

- A packet will never be bridged from a PW to an other PW in the VFI
- Assuming PW full-mesh in a VFI:
  - Full reachability
  - Core link back-up
  - No core L2 loop
  - ➔ No need for a loop prevention core STP

Remark:
Split-Horizon does not protect against loops on L2 parallel networks built for edge N-PE protection

VPLS with Split-Horizon
Loop-free interconnection with STP isolation



Split-Horizon prevent from Loop – no need to enable STP in the Core

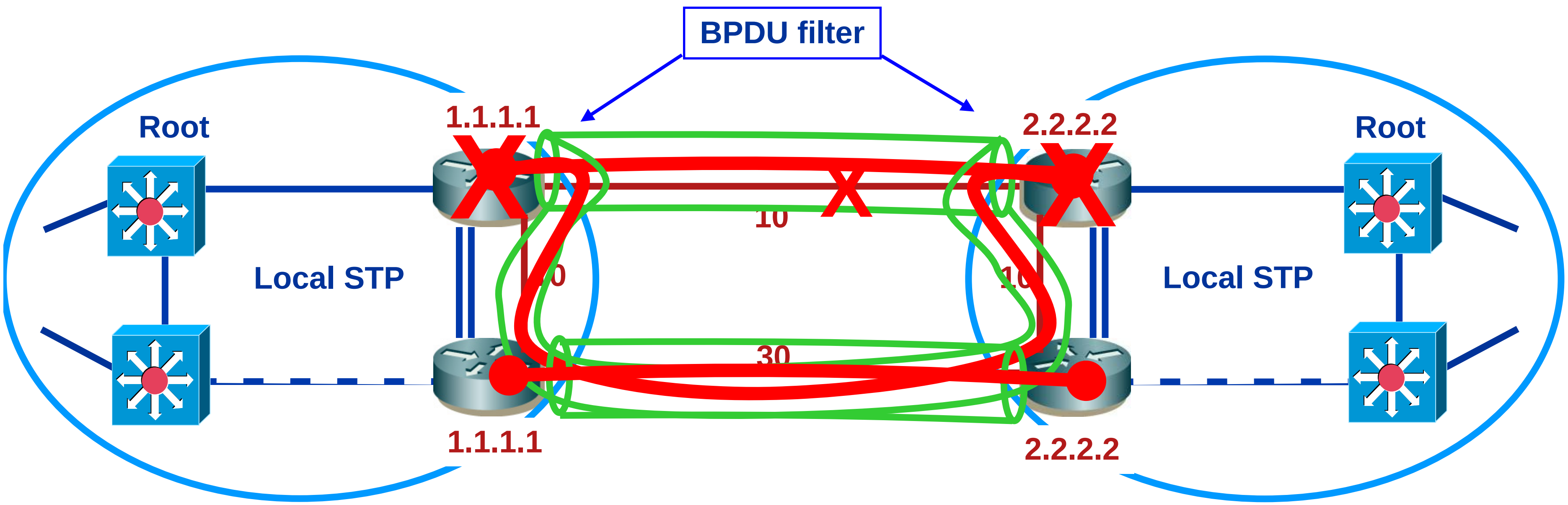
From end-to-end loop may exist and need to be understood

Loops means risks of permanent broadcast storms !!

VPLS implementation versus STP

- VPLS may work in two modes:
  1. STP transparency with extension
     - Core is tunneling BPDU (plain or QinQ)
     - Core is not L2 loop-free
     - End to End STP is preventing loops
  2. STP isolation
     - Core is filtering BPDU
     - Core & DC to DC must be L2 loop-free
     - DC independance / Small STP size
     - This is one important goal for customers
     - Mandatory VPLS-PE ! cannot be the aggregation switch
     - More complex with QinQ

Anycast PW with Traffic-Engineering



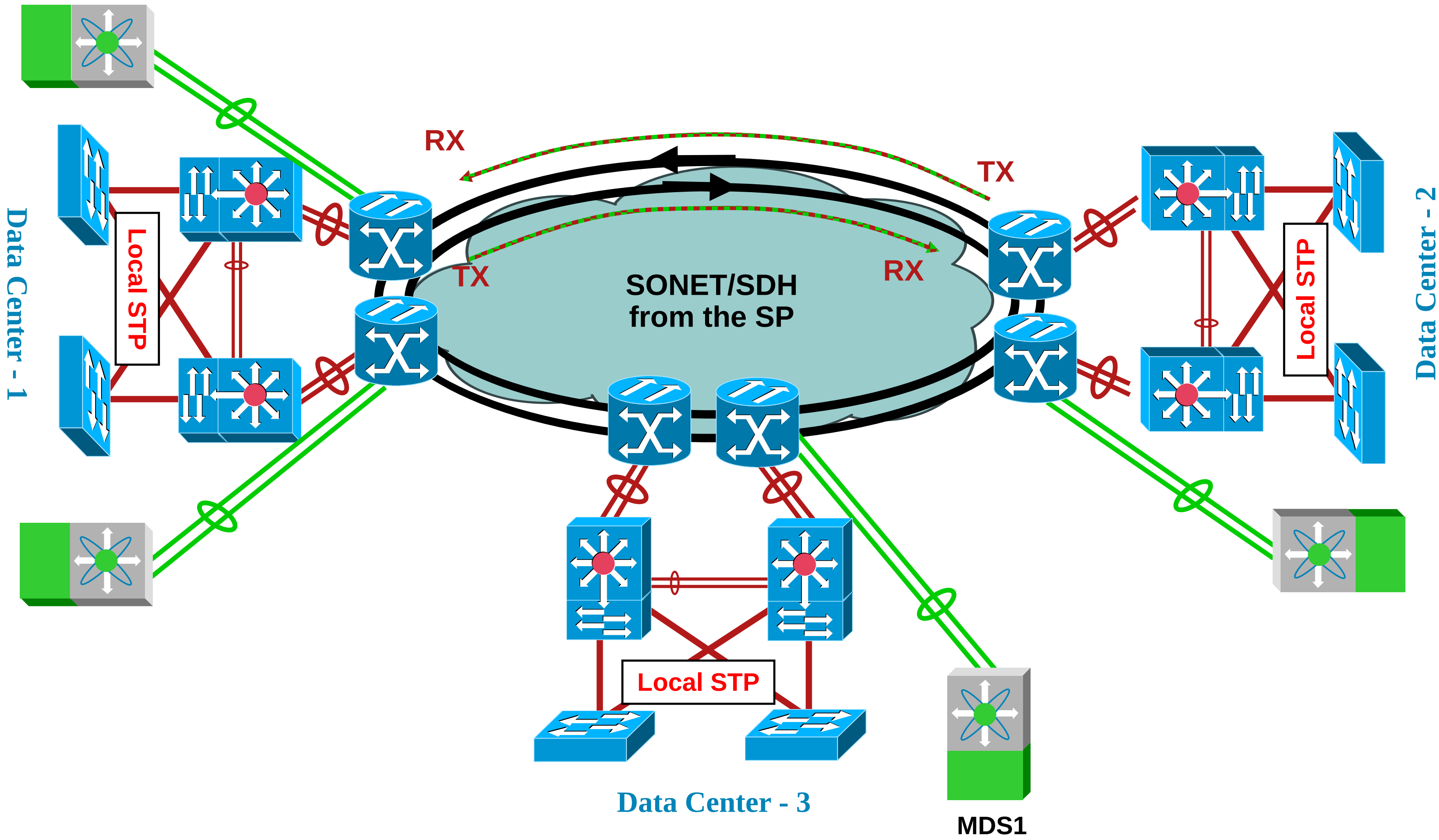
Anycast concept:

- **LDP Router-ID is duplicated into back-up N-PE**
- **TE is assuring the back-up thru alternate path**

Notes:

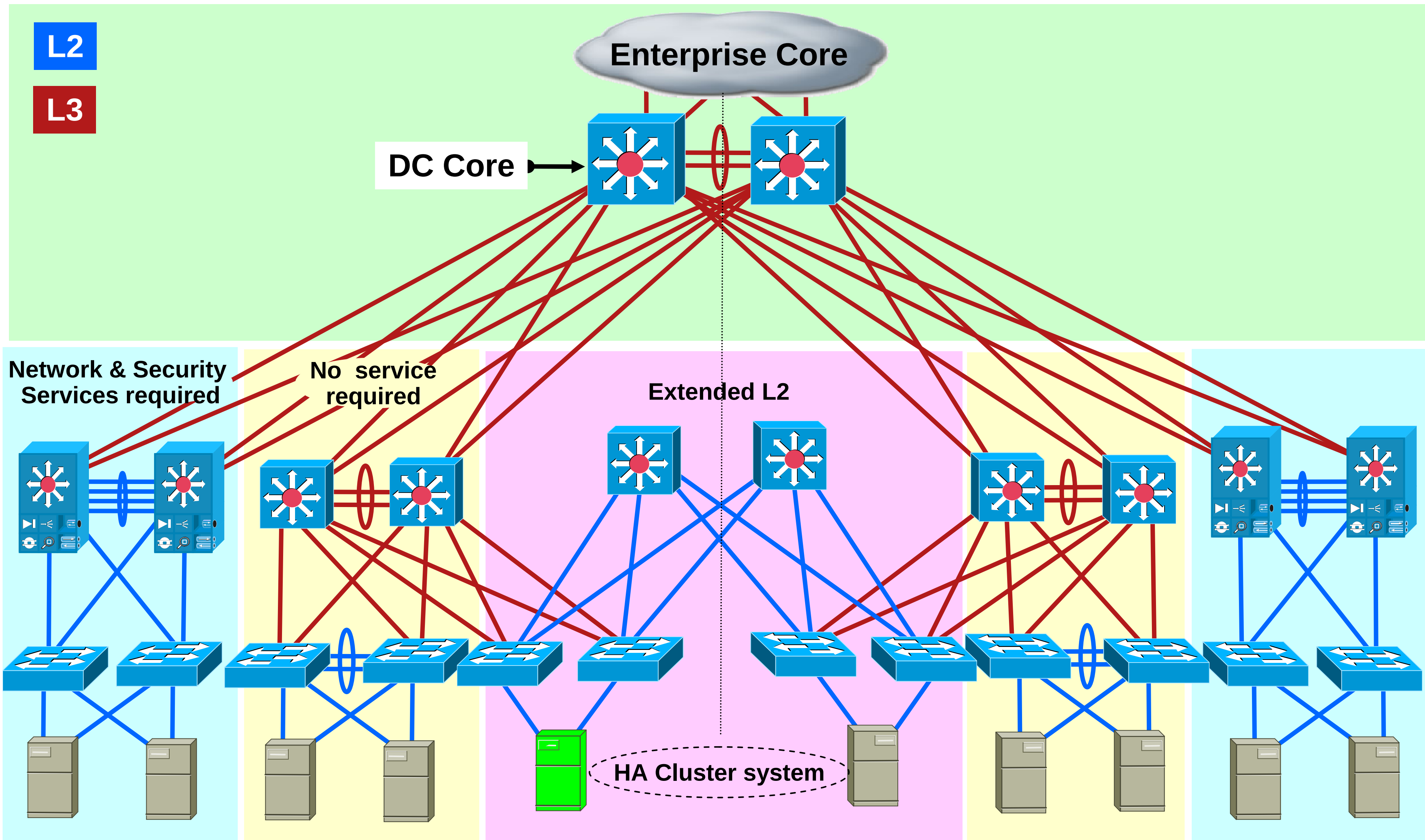
- **PW do not need to be stitched to physical topology**
- **Link core back-up is RSVP-TE protected**
- **MAC-@ flushing problem is occuring only on node up**
- **Edge links are RSTP protected**

SONET Topology with RPR
(802.17b for L2 bridging)



Extended L2 VLAN

Segmentation between distinct Applications



Conclusion

- For Business Continuance, HA includes the Network, the Devices, the Storage and the site.
- Segregate the different Applications using Layer 3
- Avoid Extending L2 VLAN if it’s not required by the Application
- If Extended L2 VLAN is required:
  - dedicate the L2 for the specific Application and keep it isolated from other Application via L3 Network
  - Avoid propagating the same STP outside your local DC
  - Police & Rate limit the traffic per VLAN ➔ prevent broadcast storm
- For long distance prefer L3 Fast-Convergence and MPLS FRR with TE to make a single L2-VPN Pseudowire
  - VPLS and Split-Horizon assure a fully resilient Loop-Free Network without the need to deploy STP.
  - The Physical layer becomes logically fully resilient
  - Physical link failure becomes transparent for L2

MPLS and GeoCluster More Information

For More Information, Please Refer to Sessions

BRKIPM-3014: Advanced MPLS deployment in Enterprise

BRKDCT-2004: Back-end Solution for Disaster Recovery

BRKDCT-2005: Design and Deployment of Layer 2 Clusters Geoclusters

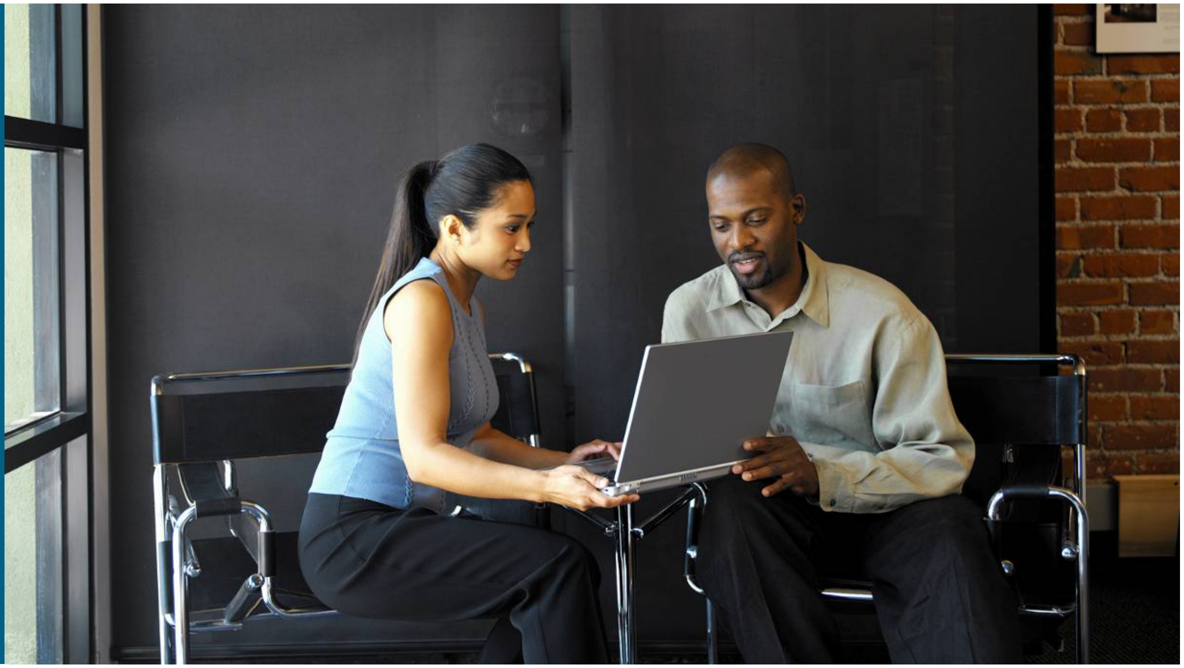
Recommended Reading

BRKDCT -2002

- Building Resilient IP Networks
- Data Center Fundamentals
- Storage Networking
- DNS and BIND
- Designing Content Switching

Available in the Cisco Company Store

Q and A



Complete Your Online Session Evaluation