Chapter 5

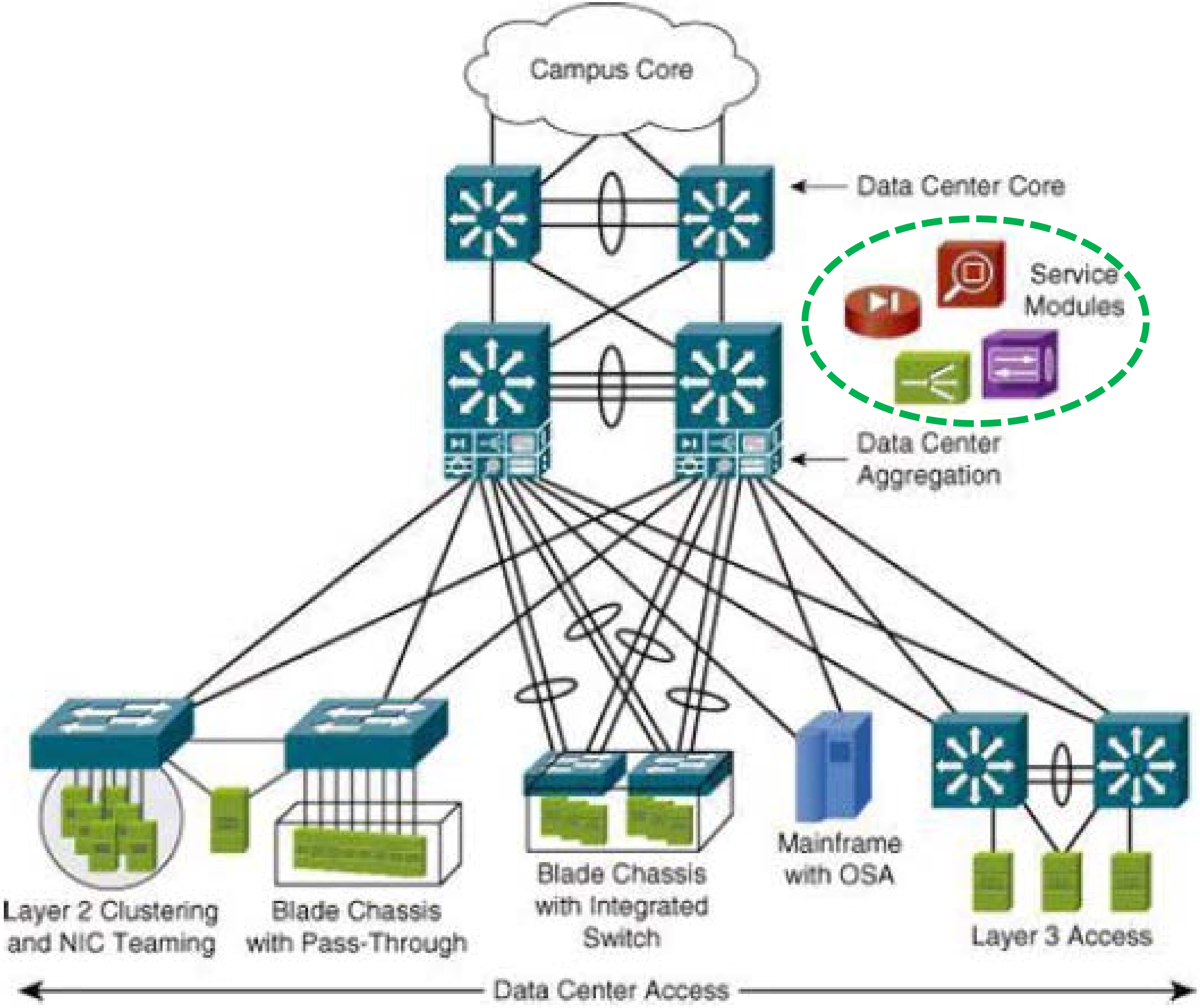
Enterprise Data Center Design

Enterprise Data Center

- The data center is home to the computational power, storage, and applications necessary to support an enterprise business.
  - Performance
  - Resiliency
  - Scalability

- The layered approach is the basic foundation of the data center design that seeks to improve scalability, performance, flexibility, resiliency, and maintenance.

Data Center Architectural Overview



Open Systems Adapter (OSA)

Enterprise Data Center

- Virtualization allows optimization of the data center provides many services.
  - It allows resources on demand, and optimization of computer and network resources.

- Virtualization also lowers **operating expenses (OPEX)** by optimizing power; heating, ventilation, air conditioning (HVAC); and data center floor space.

Enterprise Data Center

- The data center architecture is based on a three-layer approach.
  - The core layer provides a high-speed Layer 3 fabric for packet switching.
  - The aggregation layer extends spanning-tree or Layer 3 routing protocols into the access layer, depending on which access layer model is used.
  - The access layer provides physical connectivity for the servers.

Access Layer

- The access layer supports both Layer 2 and Layer 3 topologies.
  - Layer 2 adjacency requirements that fulfill the various server broadcast domain or administrative requirements.
- The server components can consist of single and dual-attached one-rack unit (1RU) servers, blade servers with integral switches.
  - blade servers with pass-through cabling, clustered servers, and mainframes with a mix of oversubscription requirements.

Aggregation Layer

- The aggregation layer supports **integrated service modules** that provide services such as
  - security,
  - load balancing,
  - content switching,
  - firewall,
  - Secure Sockets Layer (SSL) offload,
  - intrusion detection,
  - network analysis.

Note

- Small and medium data centers have a two-tier design, with the Layer 3 access layer connected to the backbone database core (collapsed core and aggregation layers).

- Three-tier designs allow for greater scalability in the number of access ports, but a two-tier design is ideal for small server farms.

The Benefits for Separation Layer

1. Layer 2 domain sizing:
    - When a requirement exists to extend a VLAN from one switch to another, the domain size is determined at the aggregation layer.
    - If the access layer is absent, the Layer 2 domain must be configured across the core for extension to occur.
        - Extending Layer 2 through a core causes path blocking by the spanning tree
            - It might cause uncontrollable broadcast issues related to extending Layer 2 domains and should be avoided.

Benefits

2. Service module support:
    - The aggregation layer with the access layer and enables services to be shared across the entire access layer of switches.
    - This lowers the total cost of ownership (TCO) and lowers complexity by reducing the number of components to configure and manage.
3. Support for a mix of access layer models:
    - The three-layer approach permits a mix of both Layer 2 and Layer 3 access models with 1RU and modular platforms, permitting a more flexible solution and allowing application environments to be optimally positioned.

Benefits

4. Support for network interface card (NIC) teaming and high-availability clustering:
    - Supporting NIC teaming with switch fault tolerance and high-availability clustering requires Layer 2 adjacency between NICs, resulting in Layer 2 VLAN extension between switches.
    - VLAN extension can also require extending the Layer 2 domain through the core, which is not recommended.

Note

- The extension of VLANs across the data center core and other layers is not best practice.
  - Customer requirements such as clustering or virtual machine mobility across multiple data centers may require VLANs to be extended across the data center core.
- These cases significantly change the data center core design and introduce the risk of failures in one area of the network affecting the entire network.
- Therefore, network designers should investigate technologies such as Cisco **Overlay Transport Virtualization (OTV)** to overcome the drawbacks and mitigate the risks of stretching Layer 2 domains across a data center core.

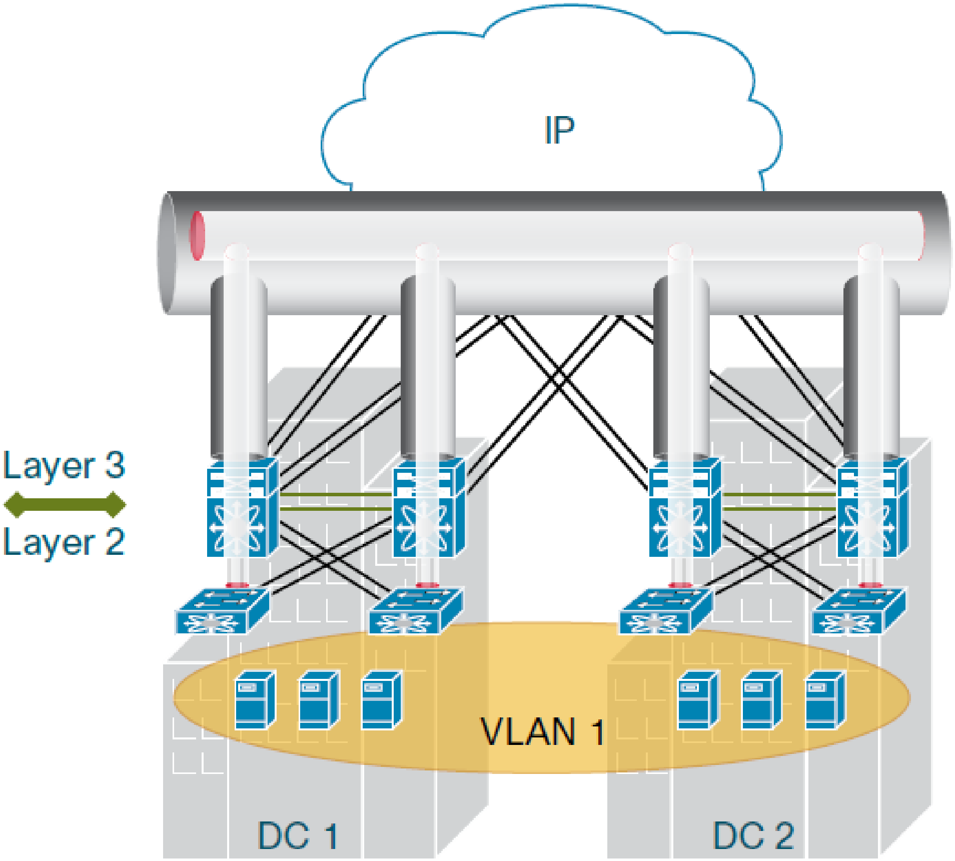
Cisco Overlay Transport Virtualization

- A critical network design requirement for deployment of distributed virtualization and cluster technologies is having all servers in the same Layer 2 VLAN.

- Meeting this requirement means extending VLANs over Layer 3 networks, but current solutions introduce operational and resiliency challenges.

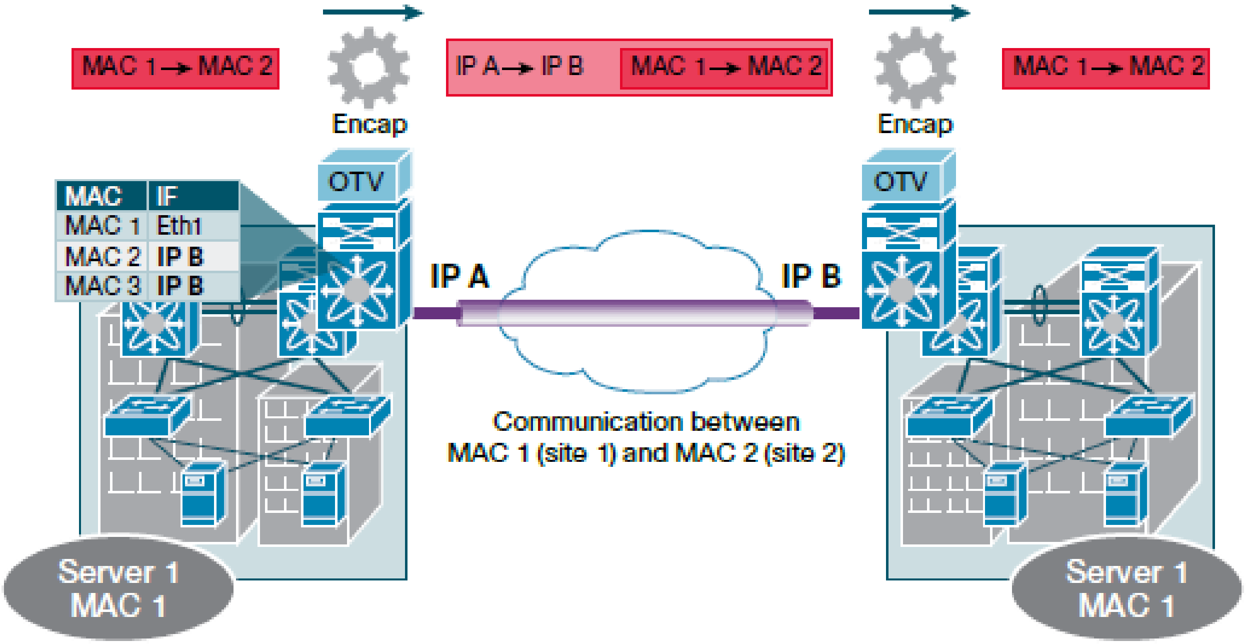
Cisco Overlay Transport Virtualization

- OTV is a new industry solution, extending Layer 2 networks over Layer 3 networks for both intra– and inter–data center applications without the operational complexities of existing interconnect solutions



Cisco Overlay Transport Virtualization

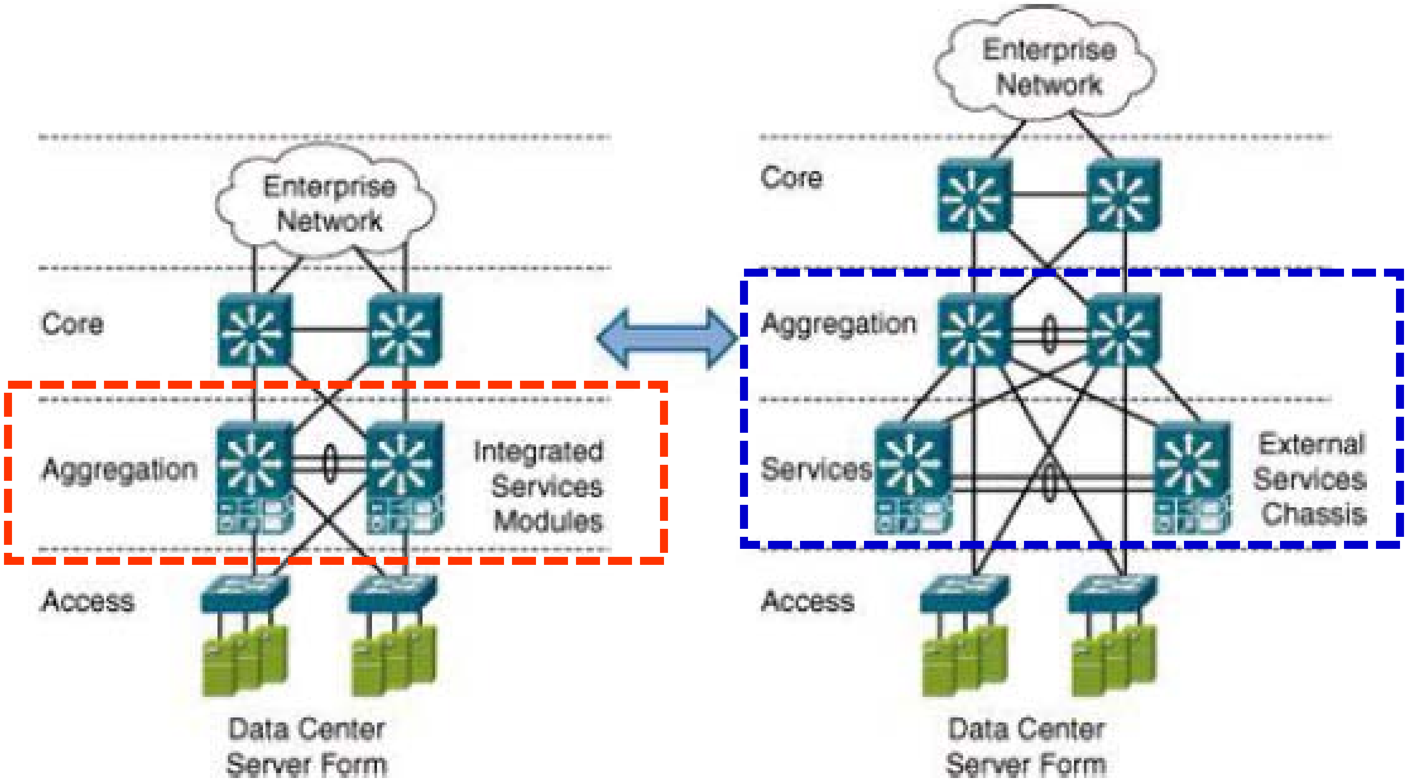
- OTV can be thought of as MAC-address routing, in which destinations are MAC addresses, and next hops are IP addresses.
- Traffic destined for a particular MAC address is encapsulated in IP and carried through the IP cloud to its MAC-address routing next hop.



MAC	IF
MAC 1	Eth1
MAC 2	IP B
MAC 3	IP B

The Services Layer

- Load balancing and security services can be integrated in the aggregation layer of the data center or designed as a separate layer.



The Services Layer

- Load balancing, firewall services, and other network services are commonly integrated in the aggregation layer in the aggregation switches.
  - It allows for a compact design, saving on power, rack space, and cabling.

- When you run out of ports on the aggregation switches, you must add a new pair of aggregation switches, including additional service modules.
  - Using this approach effectively links service scaling and aggregation layer scaling.

Dedicated Service Appliances

- Network services can be implemented on external, standalone devices that are commonly referred to as **appliances**.

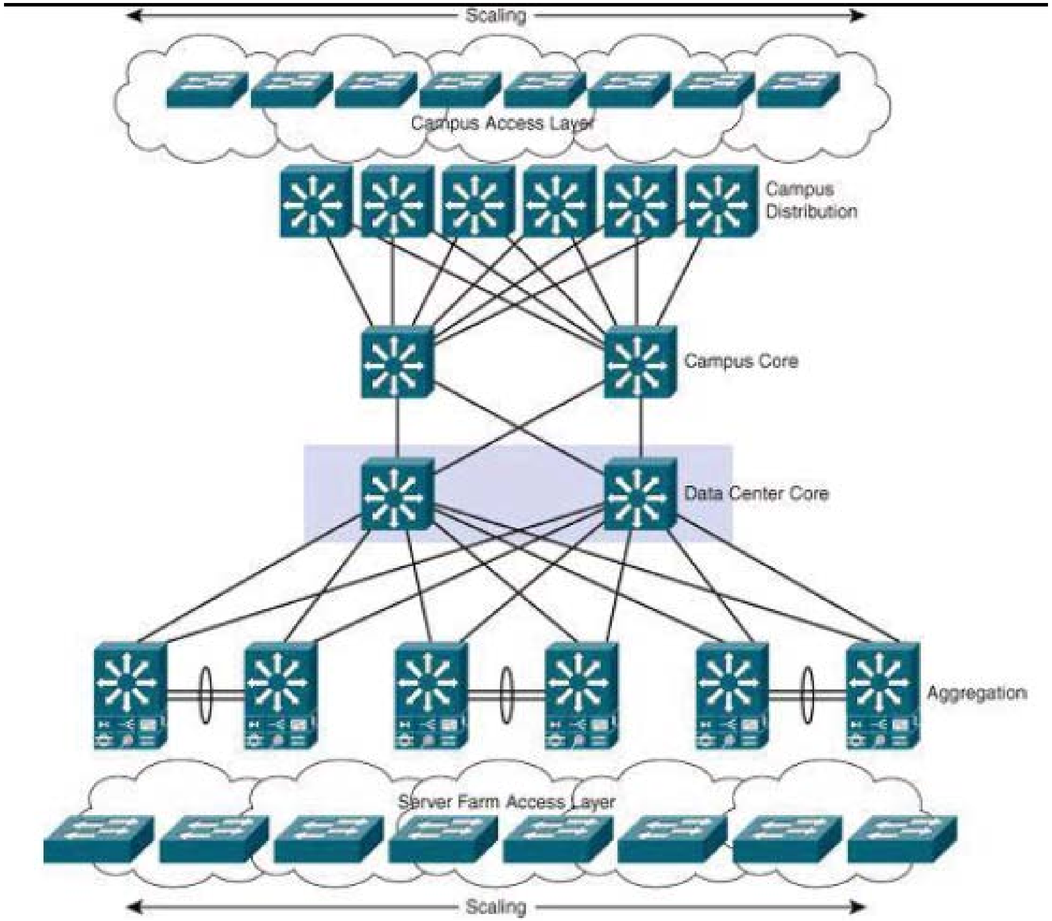
- When choosing between a design that is based on appliances or service chassis, you should consider the following design aspects:
    - Power and rack space: Combining several service modules in a single Catalyst 6500 series chassis may require less rack space and reduce the power requirements compared with using several external appliances.

Dedicated Service Appliances

- Performance and throughput:
  - The performance of some of the individual appliances can be higher than the corresponding service module.
- Fault tolerance:
  - Appliances are connected to only one of the aggregation layer switches.
    - When that aggregation switch fails, any directly attached appliances are also lost.
  - A service chassis can be dual-homed to two different aggregation switches.
    - The loss of an aggregation layer switch does not cause the associated service chassis to be lost.
- Required services:
  - Some services are supported on an appliance but not on the comparable service module.
    - However, there is now a Cisco ASA service module that no longer has this limitation

Core Layer Design

- The core layer provides a fabric for high-speed packet switching between multiple aggregation modules.
- The core layer is not necessarily required but is recommended when **multiple aggregation modules** are used for scalability.

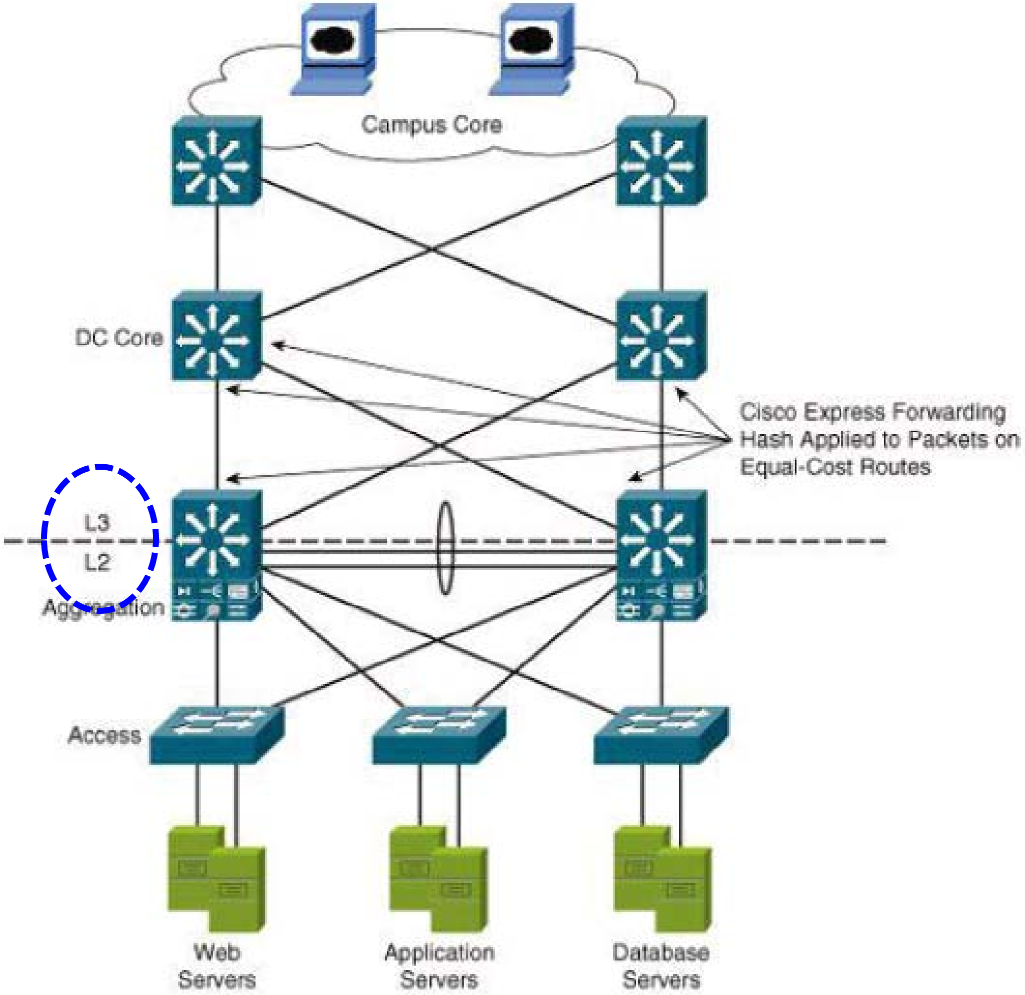


Core Layer Design

- Whether to implement a data center core, consider the following:
  - 10 Gigabit Ethernet port density
  - Administrative domains and policies:
    - Separate cores help isolate campus distribution layers and data center aggregation layers in terms of administration and policies, such as quality of service (QoS), access lists, troubleshooting, and maintenance.
  - Future growth:
    - The impact of implementing a separate data center core layer at a later date might make it worthwhile to implement it during the initial implementation stage. (系統implement的明確性)
- The core layer serves as the gateway to the campus core, where other campus modules connect, including the enterprise edge and WAN modules.
- Links connecting the data center core are connected at Layer 3 (use Layer 3) and use a distributed, low-latency forwarding architecture and 10 Gigabit Ethernet interfaces for a high level of throughput and performance.

Layer 3 Characteristics for the Data Center Core

- When designing the enterprise data center, consider where in the infrastructure to place the Layer 2 to Layer 3 boundary



```
CORE1(config)#mls ip cef load full
```

Layer 3 Characteristics for the Data Center Core

- The recommended practice is
  - the core infrastructure to be implemented at Layer 3
  - the Layer 2 to Layer 3 boundary to be implemented either within or below the aggregation layer modules.

- Layer 3 links allow the core to achieve bandwidth scalability and quick convergence, and to avoid path blocking or the risk of uncontrollable broadcast issues related to extending Layer 2 domains.
  - Layer 2 should be avoided in the core because a STP loop could cause a full data center outage.

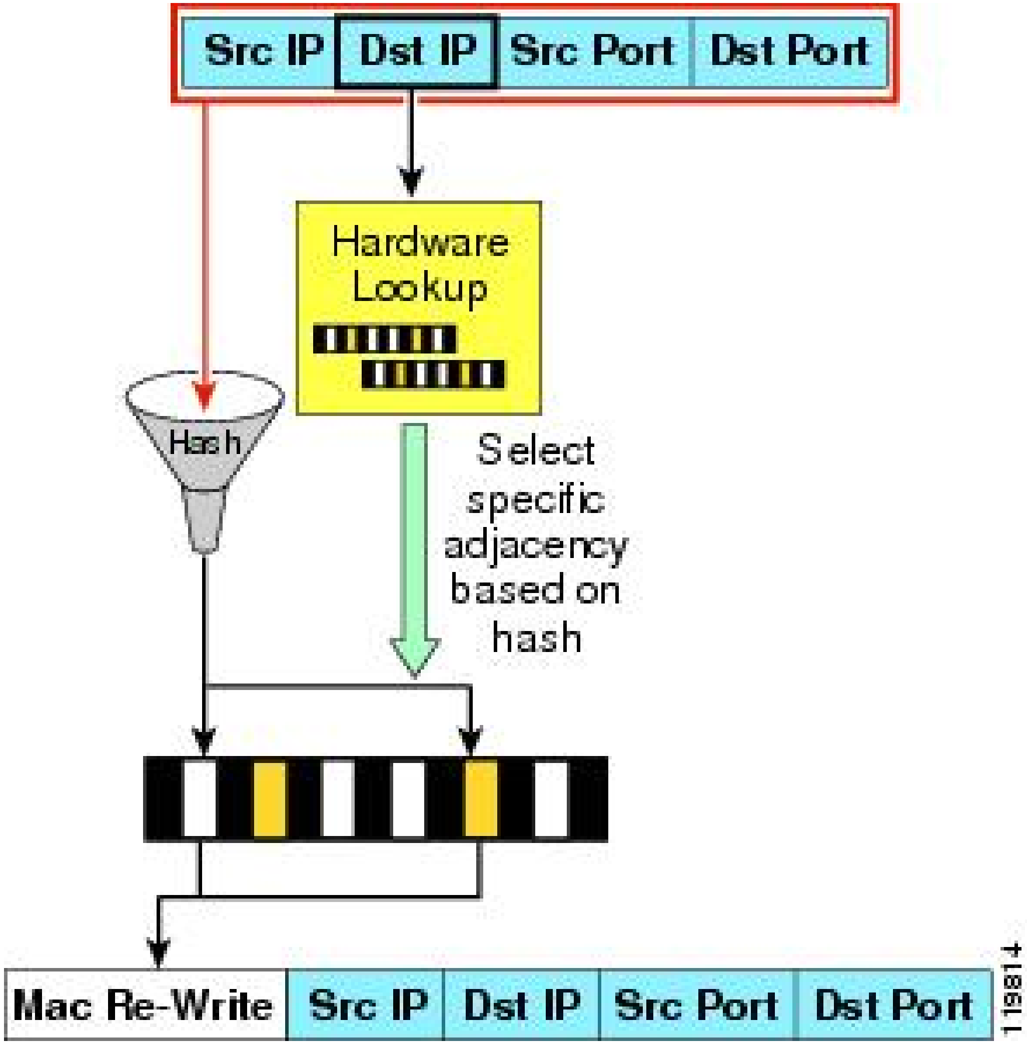
Layer 3 Characteristics for the Data Center Core

- The core layer should run an interior routing protocol such as Open Shortest Path First (OSPF) or Enhanced Interior Gateway Routing Protocol (EIGRP).
- Load balance traffic between the campus core and core aggregation layers using Cisco Express Forwarding (CEF)-based hashing algorithms.

- From a campus core perspective, at least two equal-cost routes to the server subnets permit the core to load balance flows to each aggregation switch in a particular module (p. 20).
    - Load balancing is performed using CEF-based load balancing on Layer 3 source and destination IP address hashing.
    - An option is to use Layer 3 IP plus Layer 4 port-based CEF load-balance hashing algorithms.
        - This usually improves load distribution because it presents more unique values to the hashing algorithm in the client TCP stack.

Cisco Express Forwarding

- CEF is a deterministic algorithm.

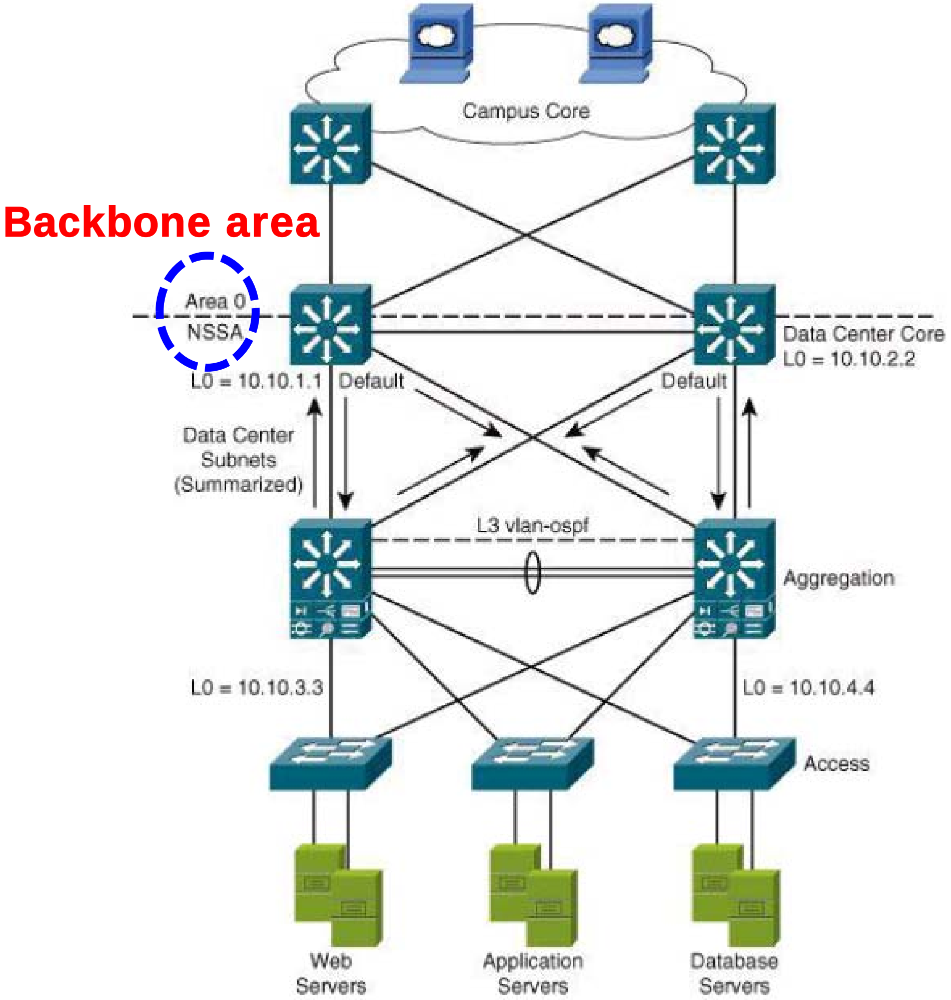


Cisco Express Forwarding

- CEF determines the longest path match for the destination address using a hardware lookup.
- Each specific index is associated with a next-hop adjacencies table.
  - By default, one of the possible adjacencies is selected by a hardware hash where the packet source and destination IP address are used.
  - One of the possible adjacencies can also be selected by a hardware hash using L4 port information in addition to the packet source and destination IP address.
- The new MAC address is attached and the packet is forwarded.

OSPF Routing Protocol Design Recommendations

- The OSPF routing protocol design should be tuned for the data center core layer.



OSPF Routing Protocol Design Recommendations

- The OSPF routing protocol suggested configuration is as follows:
- Use a not-so-stubby area (NSSA) (不是那麼STUB的區域) from the core down. It limits link-state advertisement (LSA) propagation but permits route redistribution.
- You can advertise the default route into the aggregation layer and summarize the routes coming out of the NSSA.
- Use the **auto-cost reference-bandwidth 10000** command to set the bandwidth to a 10 Gigabit Ethernet value and allow OSPF to differentiate the cost on higher-speed links, such as 10 Gigabit Ethernet trunk links.
  - The OSPF default reference bandwidth is 100 Mbps.

OSPF

- OSPF relies on several types of Link State Advertisements (LSAs) to communicate link state information between neighbors.
- A brief review of the most applicable **LSA types**:
  - **Type 1** - Represents a router
  - **Type 2** - Represents the pseudonode (designated router) for a multiaccess link
  - **Type 3** - A network link summary (**internal** route)
  - **Type 4** - Represents an ASBR
  - **Type 5** - A route **external** to the OSPF domain
  - **Type 7** - Used in stub areas in place of a type 5 LSA

OSPF

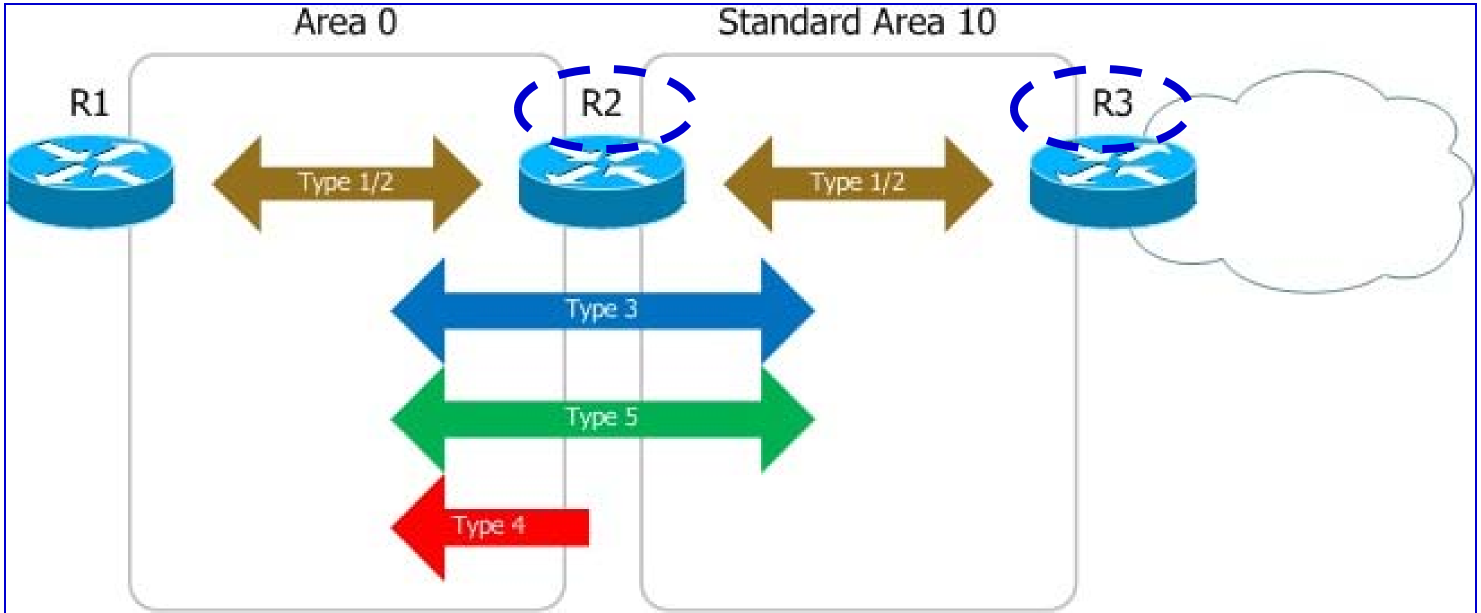
- LSA types 1 and 2 are found in all areas, and are never flooded outside of an area.
- Whether the other types of LSAs are advertised within an area depends on the area type, and there are many:
- **Backbone area** (area 0)
- **Standard area**
- **Stub area**
- **Totally stubby area**
- **Not-so-stubby area** (NSSA)

OSPF Router

- Backbone router: the router of area 0
- Internal router: the inner routers of same area
- ABR (area broder router): each interface connects to different areas, but at least one interface connects with area 0.
- ASBR (autonomous system border router): connect with other as (Autonomous System), and imports other as’s routing information into own OSPF.

Standard Areas

- Backbone area is essentially a standard area which has been designated as the central point to which all other areas connect, so a discussion of standard area behavior largely applies to the backbone area as well.
- Router 2 acts as the area border router (ABR) between a standard area and the backbone.
- R3 is redistributing routes from an external domain, and is therefore designated as an autonomous system boundary router (ASBR).



Standard Areas

- Type 1 and 2 LSAs are being flooded between routers sharing a common area.
  - This applies to all area types, as these LSAs are used to build an area's shortest-path tree, and consequently only relevant to a single area.

- Type 3 and 5 LSAs, which describe **internal** and **external** IP routes, respectively, are flooded throughout the backbone and all standard areas.
  - External routes are generated by an **ASBR**, while internal routes can be generated by any **OSPF router**.

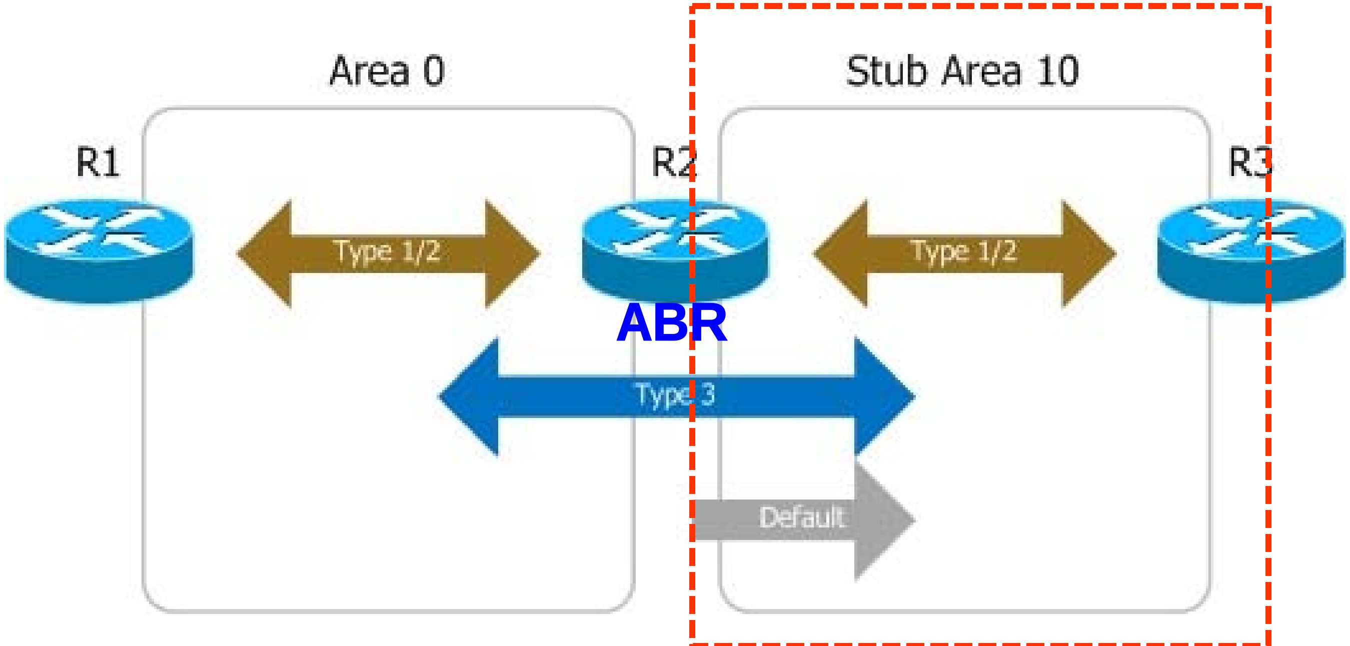
- Note the peculiar case of **type 4** LSAs. These LSAs are injected into the backbone by the ABR of an area which contains an ASBR.
- This is to ensure all other routers in the OSPF domain can reach the ASBR.

Standard Areas

- Standard areas work fine and ensure optimal routing since all routers know about all routes.
- However, there are often situations when an area has limited access to the rest of the network, and maintaining a full link state database is unnecessary.
- Additionally, an area may contain low-end routers incapable of maintaining a full database for a large OSPF network.
  - Such areas can be configured to block certain LSA types and become **lightweight stub areas**.

Stub Areas

- R2 and R3 share a common stub area. Instead of propagating external routes (type 5 LSAs) into the area, the ABR injects a type 3 LSA containing a default route (0.0.0.0) into the stub area.
- This ensures that routers in the stub area will be able to route traffic to external destinations without having to maintain all of the individual external routes.
- Because external routes are not received by the stub area, ABRs also do not forward type 4 LSAs from other areas into the stub.



Stub Areas

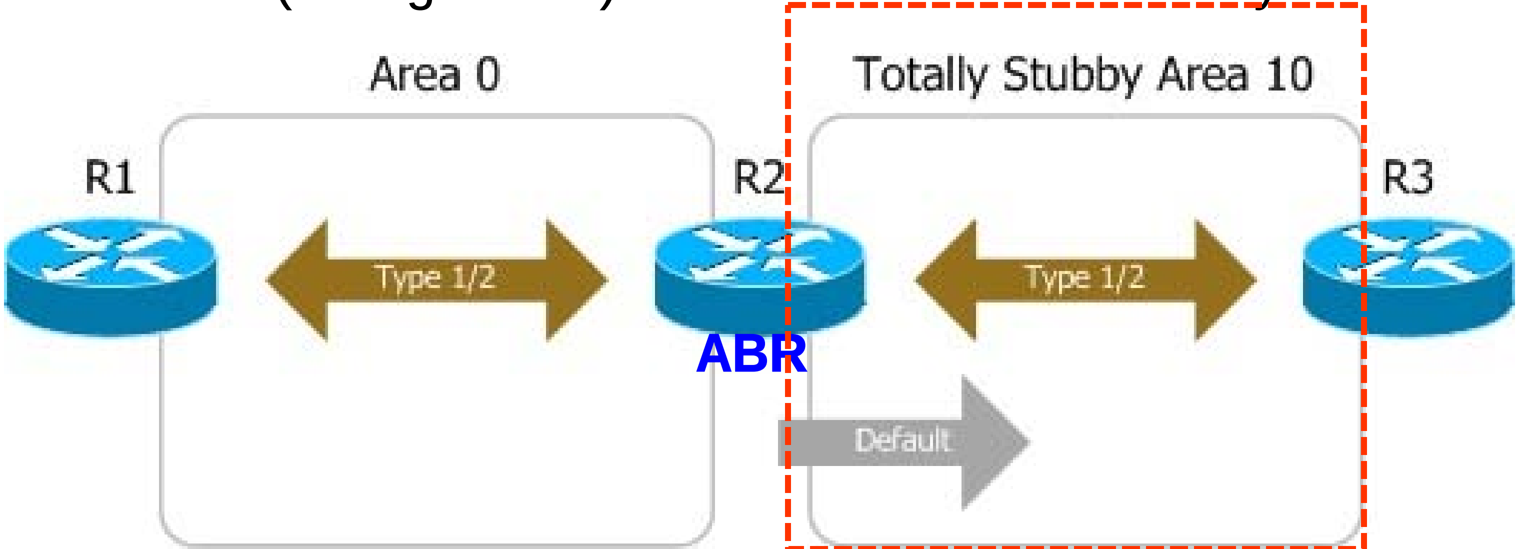
- For an area to become a stub, all routers belonging to it must be configured to operate as such.
  - Router(config-router)# area 10 stub

- Stub routers and non-stub routers will not form adjacencies.

Totally Stubby Areas

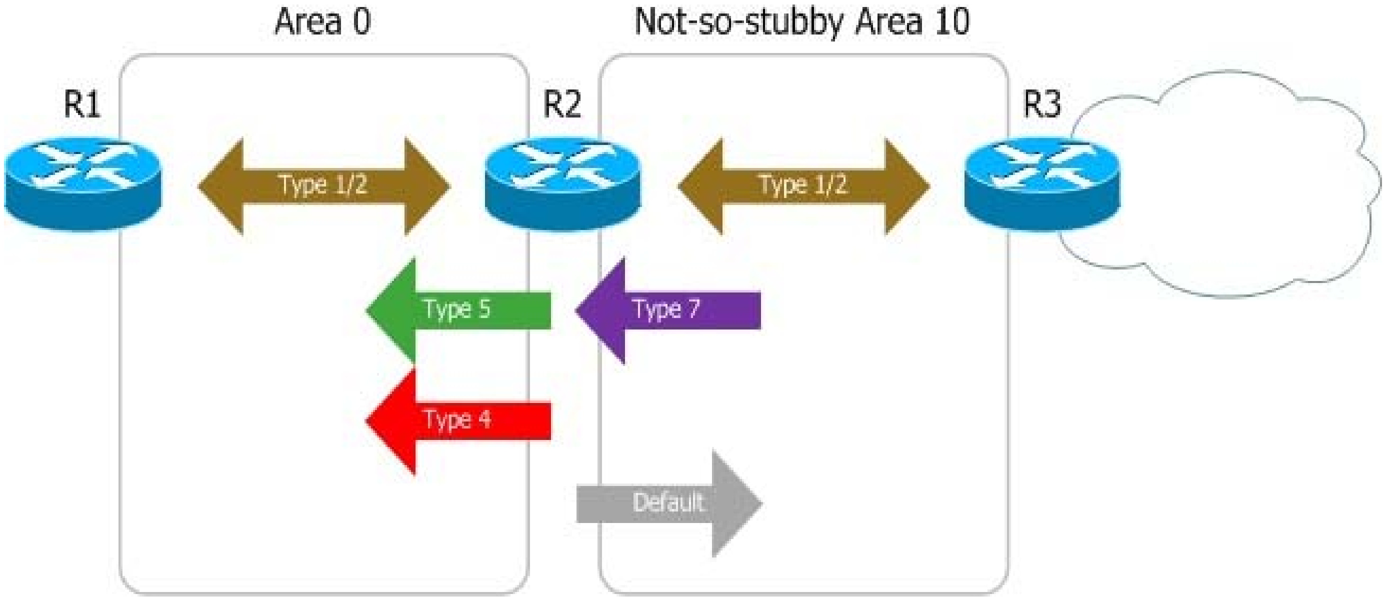
Totally stubby areas, this idea of substituting a single default route for many specific routes can be applied to internal routes as well.

- Like stub areas, totally stubby areas do not receive type 4 or 5 LSAs from their ABRs.
- However, they **also** do not receive type 3 LSAs; all routing out of the area relies on the single default route (0.0.0.0) injected by the ABR.
- A stub area is extended to a totally stubby area by configuring all of its ABRs with the no-summary parameter:
  - Router(config-router)# area 10 stub no-summary



Not-so-stubby Areas

- Stub and totally stubby areas can certainly be convenient to reduce the resource utilization of routers in portions of the network not requiring full routing knowledge.
- However, neither type can contain an ASBR, as type 4 and 5 LSAs are not permitted inside the area.
- To solve this problem, and in what is arguably the worst naming decision ever made, Cisco introduced the concept of a **not-so-stubby area (NSSA)**.



Not-so-stubby Areas

- An NSSA makes use of type 7 LSAs, which are essentially type 5 LSAs in disguise.
    - This allows an ASBR to advertise external links to an ABR, which converts the type 7 LSAs into type 5 before flooding them to the rest of the OSPF domain.
- An NSSA can function as either a stub or totally stubby area.
    - To designate a normal (stub) NSSA, all routers in the area must be so configured: Router(config-router)# area 10 nssa
- Type 3 LSAs will pass into and out of the area. Unlike a normal stub area, the ABR will *not* inject a default route into an NSSA unless explicitly configured to do so.
- As traffic cannot be routed to external destinations without a default route, you'll probably want to include one by appending default-information-originate.
    - Router(config-router)# area 10 nssa default-information-originate

Not-so-stubby Areas

- To expand an NSSA to function as a totally stubby area, eliminating type 3 LSAs, all of its ABRs must be configured with the no-summary parameter:
  - Router(config-router)# area 10 nssa no-summary

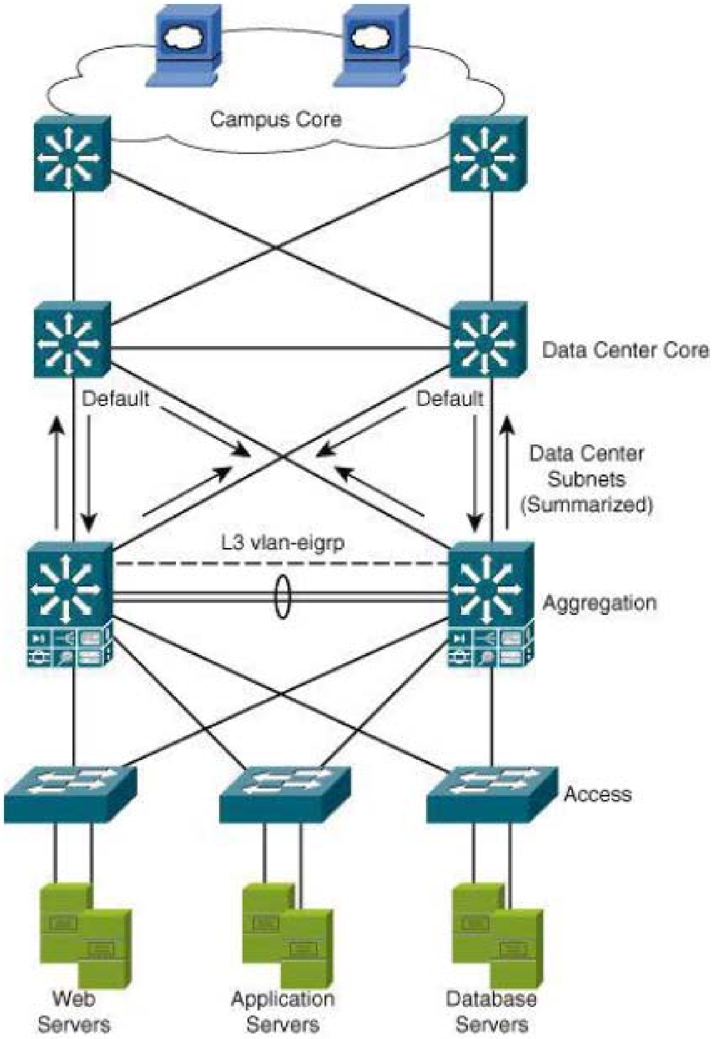
- The ABR of a totally stubby NSSA (or not-so-totally-stubby area, if you prefer) injects a default route without any further configuration.

Summary

- **Standard areas** can contain LSAs of type 1, 2, 3, 4, and 5, and may contain an ASBR.
  - The backbone is considered a standard area.
- **Stub areas** can contain type 1, 2, and 3 LSAs.
  - A default route is substituted for external routes.
- **Totally stubby areas** can only contain type 1 and 2 LSAs, and a single type 3 LSA.
  - The type 3 LSA describes a default route, substituted for all external and inter-area routes.
- **Not-so-stubby areas** implement stub or totally stubby functionality yet contain an ASBR.
  - Type 7 LSAs generated by the ASBR are converted to type 5 by ABRs to be flooded to the rest of the OSPF domain.

EIGRP Routing Protocol Design Recommendations

- The EIGRP routing protocol design should be tuned for the data center core layer



EIGRP Routing Protocol Design Recommendations

- Here are some recommendations on EIGRP design for the data center core layer:

- Advertise a default summary route into the data center access layer with the **ip summary-address eigrp** interface command on the aggregation layer.

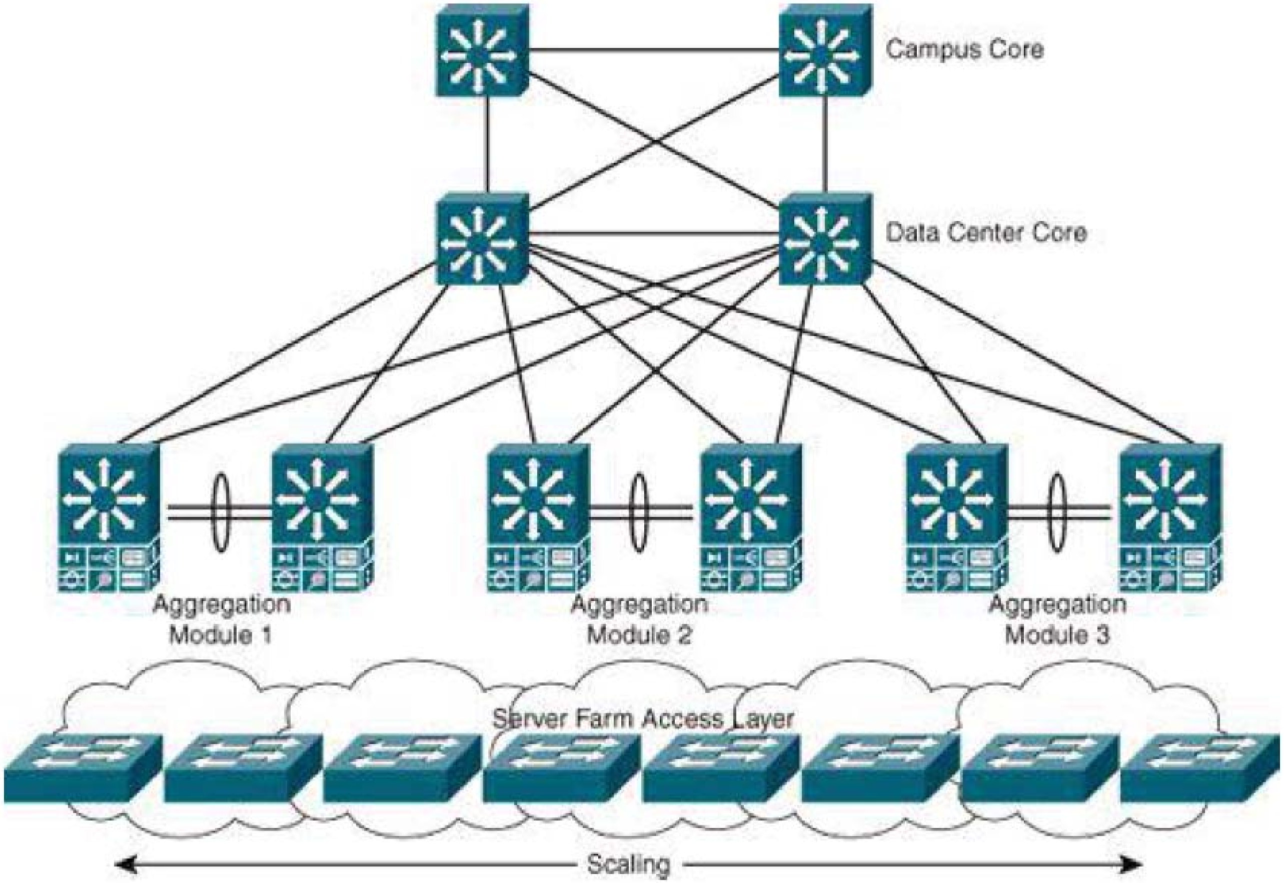
- If other default routes exist in the network, such as from the Internet edge, you might need to filter them using distribute lists.

Aggregation Layer Design

- The aggregation layer design is critical to the stability and scalability in data center architecture.
- The following aggregation layer design topics are discussed in this section:
  - Scaling the aggregation layer
  - STP design
  - Integrated services support
  - Service module placement considerations
  - STP, Hot Standby Router Protocol (HSRP), and service context alignment
  - Active/standby service design
  - Active/active service design
  - Establishing path preference
  - Using virtual routing and forwarding (VRF) instances in the data center

Scaling the Aggregation Layer

- Multiple aggregation modules allow the data center architecture to scale as additional servers are added.



Scaling the Aggregation Layer

- Multiple aggregation modules are used to scale the aggregation layer:
  - Spanning-tree scaling: By using multiple aggregation modules, you can limit Layer 2 domain size and can limit failure exposure to a smaller domain.
  - Access layer density scaling: This trend can create a density challenge in existing or new aggregation layer designs.
    - Currently, the maximum number of 10 Gigabit Ethernet ports that can be placed in the aggregation layer switch is 64 (the WS-X6708-10G-3C line card in the Cisco Catalyst 6509 switch, $4,050.00).

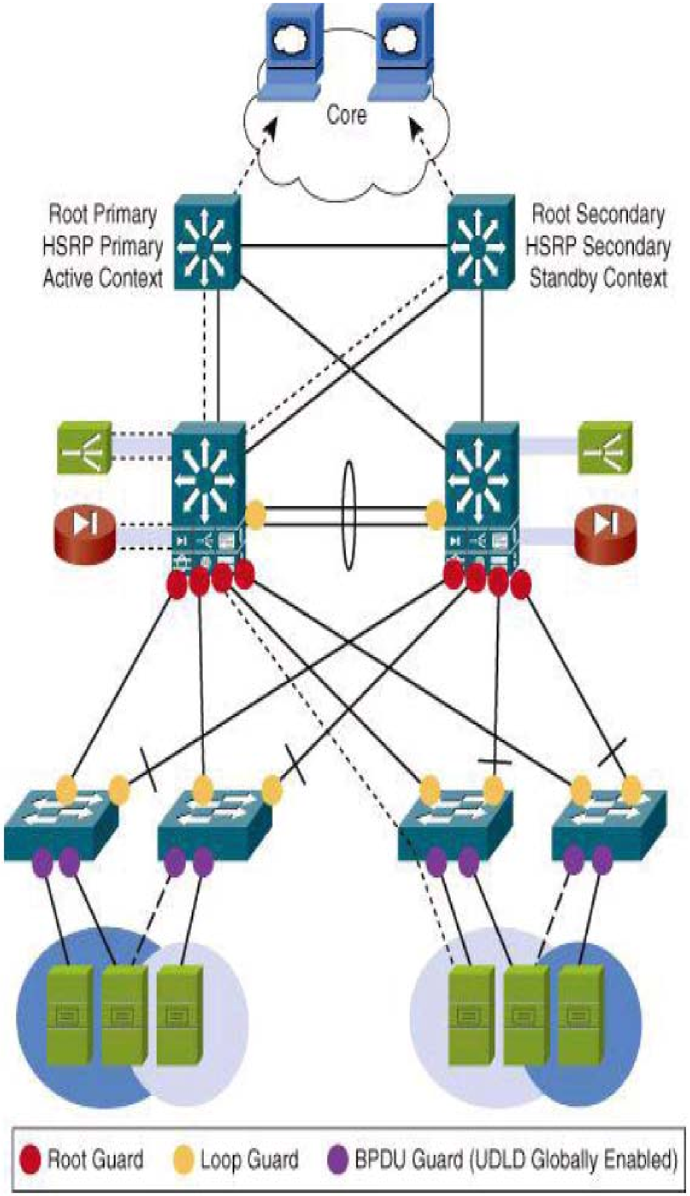
  - HSRP scaling: HSRP is the most widely used protocol for default gateway redundancy.
    - The aggregation layer provides a primary and secondary router “default gateway” address for all servers on a Layer 2 access topology across the entire access layer.

Scaling the Aggregation Layer

- Application services scaling: The aggregation layer supports applications across multiple access layer switches, scaling the ability of the network to provide application services.
  - Some examples of supported applications are Server Load Balancing (SLB) and firewalls.

STP Design

- Layer 2 in the aggregation layer, the STP design should be your first concern.
- The aggregation layer carries the largest burden with Layer 2 scaling because the aggregation layer establishes the Layer 2 domain size and manages it with a spanning tree protocol.
  - Such as Rapid Per-VLAN Spanning Tree (RPVST+) or Multiple Spanning Tree (MST)
- MST requires careful and consistent configuration to avoid "**regionalization**" and reversion to a single global spanning tree.



Understanding Bridge Assurance

- Bridge assurance can be used as protection against certain problems that can cause bridging loops in the network.
- Specifically, bridge assurance is used to protect against a unidirectional link failure or other software failure and a device that continues to forward data traffic when it is no longer running the spanning-tree algorithm.
- If the device on one side of the link has bridge assurance enabled and the device on the other side either does not support bridge assurance or does not have this feature enabled, the connecting port is blocked.
- **Note**
  - Bridge assurance is preferred over loop guard.
  - If an access switch does not support bridge assurance, loop guard can be implemented between that access switch and the aggregation switch.
  - Do not enable both bridge assurance and loop guard at the same time.

Integrated Service Modules

- Integrated service modules provide services such as content switching, firewall, SSL offload, intrusion detection, and network analysis.
- For redundancy, the integrated services may be deployed in one of two scenarios:
  - Active/standby pairs, where one appliance is active and the other appliance is in standby mode.
  - Active/active pairs, where both appliances are active and providing services.
- Integrated service modules or blades can provide flexibility and economies of scale by optimizing rack space, cabling, and management.

Service Modules and the Services Layer

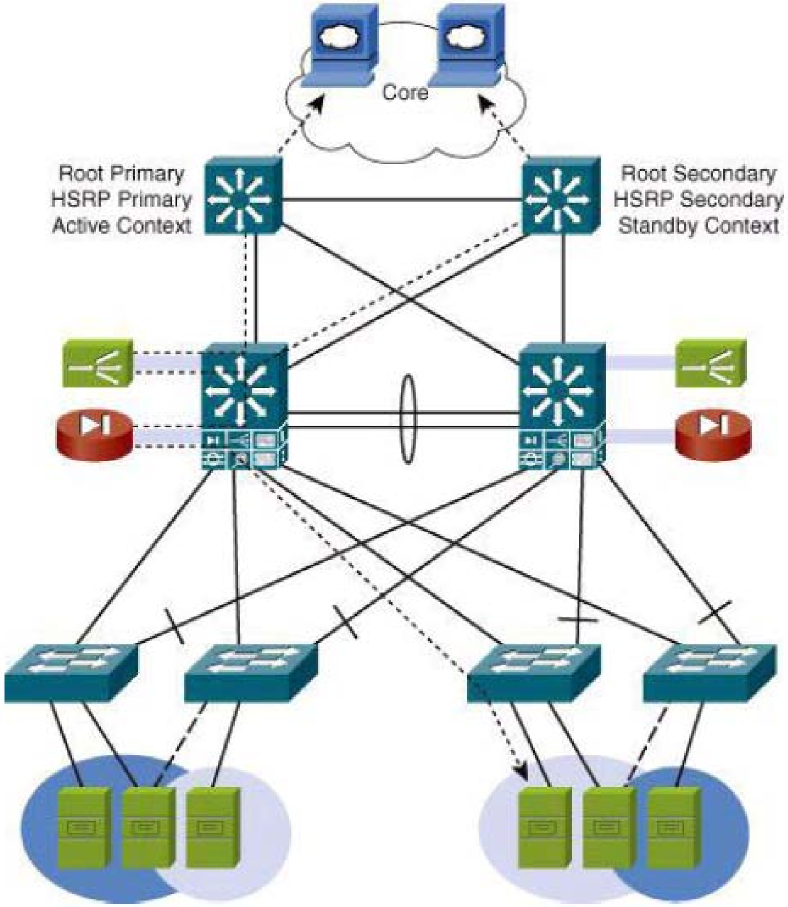
- The best choice for a particular data center deployment depends on the specific requirements of the applications in use.
- When designing data center services include the following:
  - Determining the default gateway for the servers.
  - Service modules can be integrated in the aggregation layer switches or implemented as a separate services layer.
  - Service modules are efficient with regard to rack space, power, and cabling.
  - Dedicated appliances may offer higher throughput or features that are unavailable on a service module.

Service Modules and the Services Layer

- active/standby design
  - All traffic flows through a single service chassis or chain of appliances.
  - A second service chassis or appliance chain is provisioned and kept in a standby state, to take over only if components in the primary service chain fail.
- active/active design
  - Leverages the fact that a physical service module or appliance can be divided into virtual contexts, such as firewall contexts.
  - The active/active model allows all available hardware resources to be used and is more complex.

Active STP, HSRP, and Service Context Alignment

- A recommended practice aligns the active STP, HSRP, and service context in the aggregation layer to provide a more deterministic environment.



Active STP, HSRP, and Service Context Alignment

- The active service context can be aligned by connecting the service module on the aggregation switch supporting the primary STP root and primary HSRP instance.

- Active component alignment prevents session flow from entering one aggregation switch and then hopping to a second aggregation switch to reach a service context.

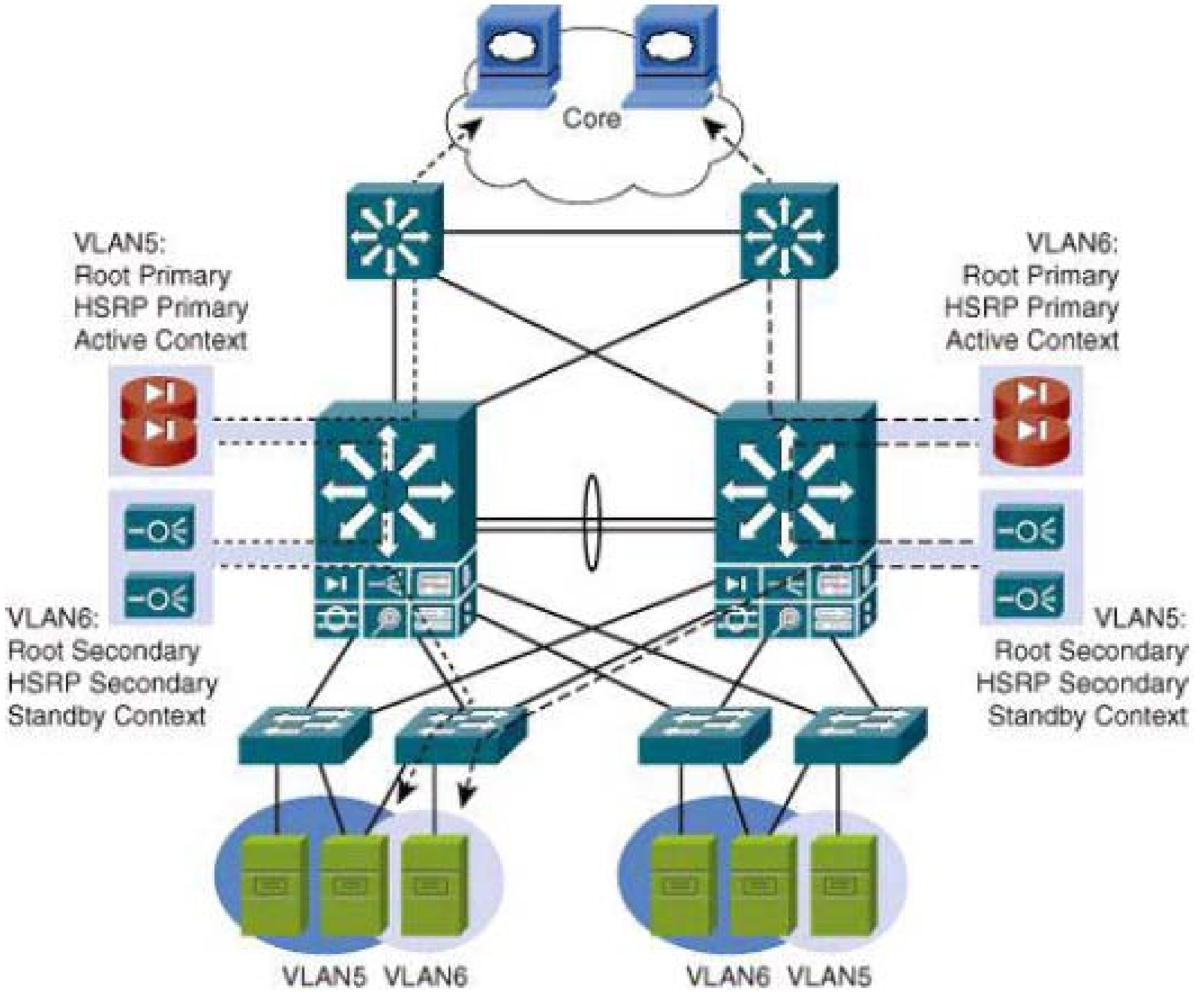
- In other words, when the traffic enters the aggregation switch that is connected to the active service context, the traffic is forwarded to the service module directly and avoids the Inter-Switch Link (ISL).

Active/Standby Service Module Design

- The active/standby mode of operation is used by service modules that require Layer 2 adjacency with the servers.
- Advantages:
  - It is a predictable deployment model.
  - This traditional model simplifies troubleshooting.
  - You know in the primary situation what service modules are active and where the data flows should occur.
- Disadvantages:
  - It underutilizes the access layer uplinks because it may not use both uplinks.
  - It underutilizes service modules and switch fabrics because it does not use both modules.
- This model uses the aligned spanning tree root, the primary HSRP, and the active service module.

Active/Active Service Module Design

- The active/active mode of operation is used by service modules that support multiple contexts or multiple active/standby groups.



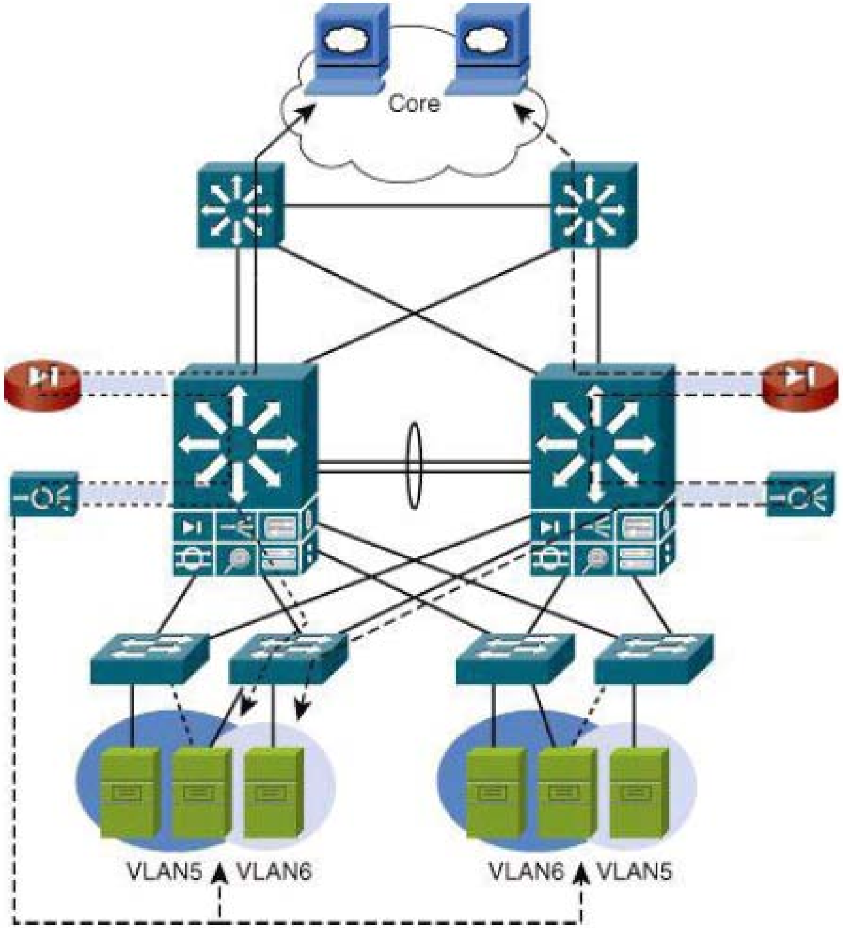
Active/Active Service Module Design

- Advantages:
  - It distributes the services and processing and increases the overall service performance.
  - It supports uplink load balancing by VLAN, so that the uplinks can be used more efficiently.

- This model aligns the spanning-tree root, the primary HSRP, and the service module per active context on each VLAN.

Establishing Inbound Path Preference

- Active/standby service module pairs become important to align traffic flows so that the active primary service modules are the preferred path to a particular server application.



Establishing Inbound Path Preference

- When a client initiates a connection to the virtual server, the Cisco Content Switching Module (CSM) chooses a real physical server in the server farm for the connection based on configured load-balancing algorithms and policies such as access rules.
- The Route Health Injection (RHI) feature allows a Cisco switch to install a host route in the Multilayer Switch Feature Card (MSFC) if the virtual server is in the operational state.
- By using RHI with specific route map attributes to set the desired metric, a /32 route for the virtual IP address is injected into the routing table.

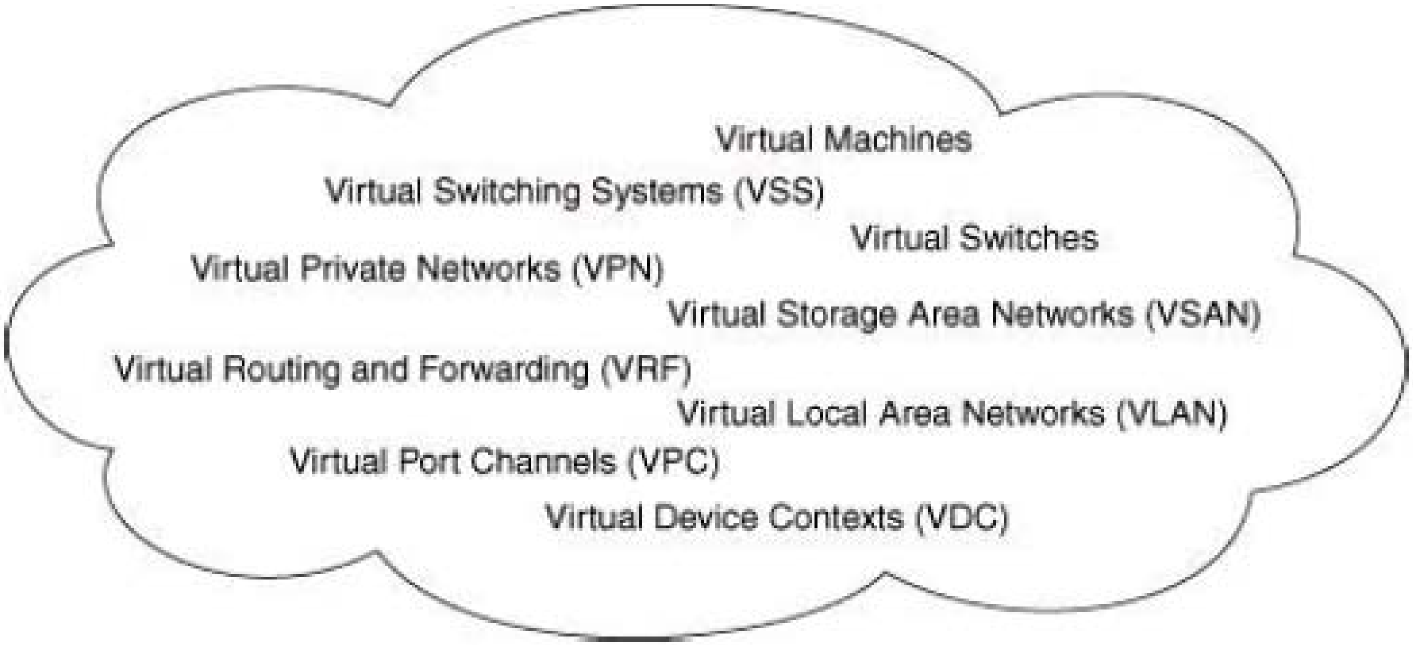
Establishing Inbound Path Preference

- This establishes a path preference with the enterprise core so that all sessions to a particular virtual IP address go to the aggregation layer switch where the primary service module is located.

- If context failover occurs, the Route Health Injection (RHI) and path preference point to the new active server.

Definition of Virtualization

- Virtualization is a term that is used in the field of IT for the concept of creating abstract entities from a pool of physical resources, while hiding these physical resources from the users or systems that interact with these abstract entities.


Virtual Machines
Virtual Switching Systems (VSS)
Virtual Switches
Virtual Private Networks (VPN)
Virtual Storage Area Networks (VSAN)
Virtual Routing and Forwarding (VRF)
Virtual Local Area Networks (VLAN)
Virtual Port Channels (VPC)
Virtual Device Contexts (VDC)

Virtualization

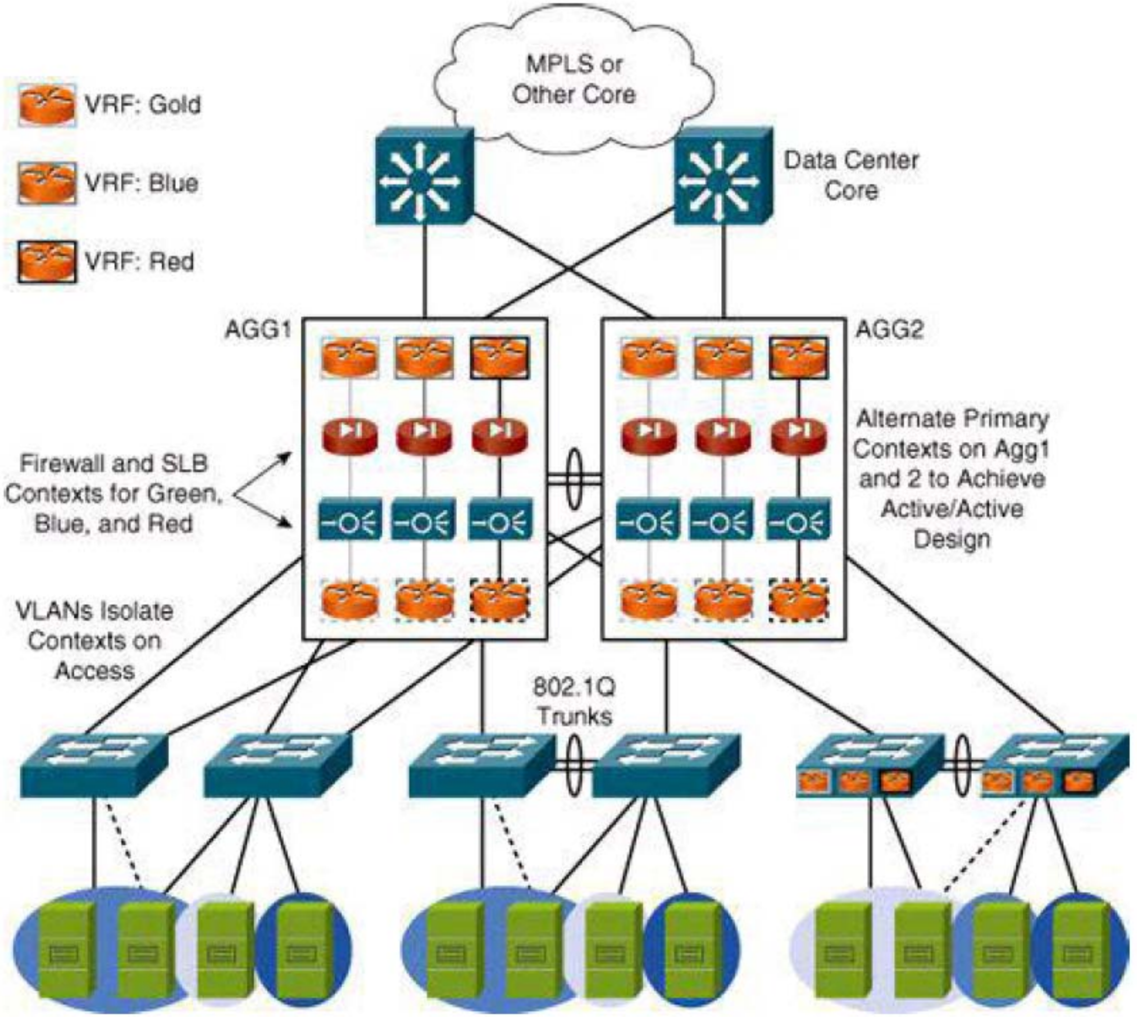
- Virtualization covers many different technologies:
  - Virtual machine
  - VLAN
  - Virtual SAN (VSAN)
  - VPN
  - VDC
  - VRF: VRF creates multiple, logical Layer 3 routing and forwarding instances inside a single physical router.
  - VSS
  - vPC
  - vSwitch

Virtualization Categories

- There are several different categories of virtualization:
- **Network virtualization:**
  - Network virtualization typically consists of two components: separation of control and data plane functions for the virtual networks inside the network nodes
  - For example, VLANs inside a switch, and separation of traffic on the links between the nodes (for example, the use of IEEE 802.1Q tagging on a trunk between switches).
- **Device virtualization:**
  - A major benefit of device virtualization is that several low-performance devices can be replaced by one high-performance device.
  - This replacement typically yields a better price-to-performance ratio.
- **Device clustering:**
  - Device clustering allows multiple physical devices to be combined into a larger logical device.
  - They allow systems to scale beyond the size of a single system and attain a higher availability than what could be achieved by a single system.

Using VRFs in the Data Center

- Finally, separate VRFs can be used as a virtualization and management approach in data center designs.


MPLS or Other Core
VRF: Gold
VRF: Blue
VRF: Red
Data Center Core
AGG1
AGG2
Firewall and SLB Contexts for Green, Blue, and Red
Alternate Primary Contexts on Agg1 and 2 to Achieve Active/Active Design
VLANs Isolate Contexts on Access
802.1Q Trunks

Virtual Routing and Forwarding

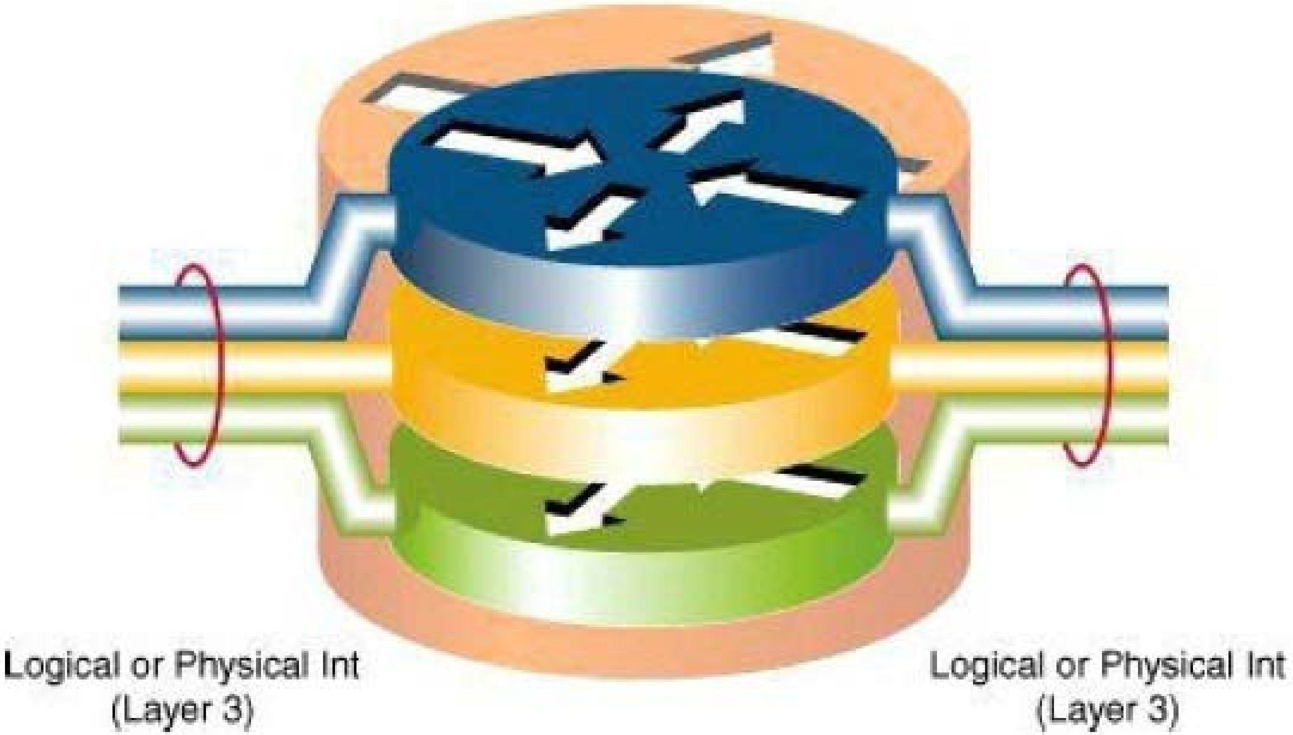
- VLANs (Virtual Local Area Network) are virtual Layer 2 networks.
  - Traffic is separated at Layer 2, and hosts in different VLANs cannot communicate at Layer 2.
- To communicate between VLANs, a router is needed, and network policies and traffic control can be implemented on the router.
- In most campus network designs, the routing function is implemented on the distribution layer, which is the Layer 2 to Layer 3 boundary.
- Therefore, VLAN separation alone does not automatically imply complete network separation.
  - For example, voice and data traffic can be separated into two VLANs.

Virtual Routing and Forwarding

- This separation will not stop devices on the data VLAN from communicating with devices on the voice VLAN on Layer 3 via the distribution switches unless specific measures are taken to restrict the flow of traffic between the VLANs.
- To create multiple separated Layer 3 networks on a shared Layer 3 infrastructure, an additional layer of virtualization is necessary.
- To achieve the separation on Layer 3, the data plane and control plane functions of the router or Layer 3 switch need to be segmented into different Layer 3 VPNs (Virtual Private Network).

Virtual Routing and Forwarding

- This process is similar to the way that a Layer 2 switch segments the Layer 2 control and data plane into different VLANs.



Virtual Routing and Forwarding

- The core concept in Layer 3 VPNs is a virtual routing and forwarding (VRF) instance.
- VRF consists of all the data plane and control plane data structures and processes that together define the Layer 3 VPN.
- The virtualized network consists of Layer 2 VLANs and Layer 3 VRFs to provide logical, end-to-end isolation across the network.
- The number of VRFs and VLANs match the number separate paths needed and are mapped to each other.

Virtual Routing and Forwarding

- A VRF includes the following **four** components:

1. A subset of the router interfaces:
   - This component includes software interfaces, such as subinterfaces, tunnel interfaces, loopback interfaces, and SVIs (switch virtual interface).
   - The VRF holds its own separate routing and forwarding tables.
   - Interfaces are either associated with global routing or a particular VRF
2. A routing table or RIB (Routing Information Base):
   - Because traffic between Layer 3 interfaces that are in different VRFs should remain separated, a separate routing table is necessary for each VRF.
   - The separate routing table ensures that traffic from an interface in one VRF cannot be routed to an interface in a different VRF.

Virtual Routing and Forwarding

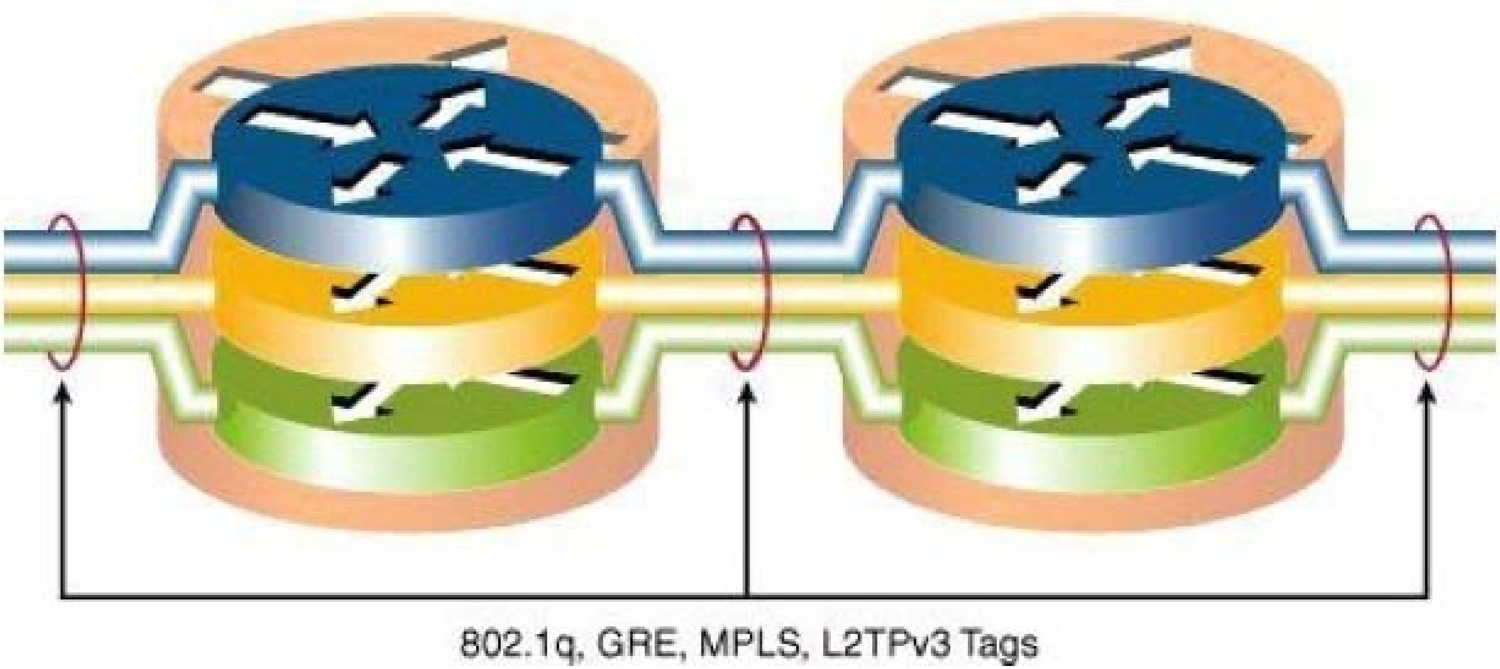
3. A FIB (Forwarding Information Base):
   - The routing table or RIB is a control plane data structure, and an associated FIB is calculated, which is used in the actual packet forwarding.
   - This also needs to be separated by VRF
4. Routing protocol instances:
   - To ensure control plane separation between the different Layer 3 VPNs, it is necessary to implement routing protocols on a per-VRF basis.
   - You can run an entirely separate process for the routing protocol in the VRF.
   - You can use a sub-process or routing protocol instance in a global process that is in charge of the routing information exchange for the VRF.

Layer 3 VPNs and Network Virtualization

- Layer 3 must be separated inside the router or Layer 3 switch VPNs for the routing and forwarding of traffic from the different VPNs.
  - You can accomplish the implementation of different VRFs for each VPN inside the router.
- It must be possible to identify the traffic that belongs to each VPN as it travels from router to router.

Layer 3 VPNs and Network Virtualization

- The sending router should mark the traffic on egress in such a way that the receiving router can identify the originating VPN on ingress.



Four Major Mechanisms

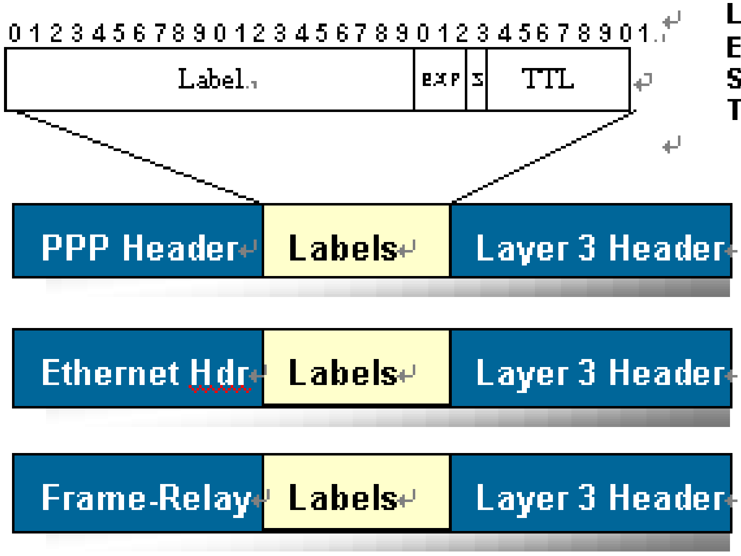
- There are four major mechanisms that you can use to accomplish this action:
- **802.1Q VLAN tagging:**
  - Multiple sub-interfaces or SVIs (switch virtual interface) are created for each physical link between two VPN-enabled routers.
  - These sub-interfaces or SVIs are then associated with the different VRFs.
  - Packets that are sent from a specific VRF are sent out a particular sub-interface or SVI and use the associated VLAN tag for that sub-interface or SVI.

Virtual Routing and Forwarding

- **GRE tunnels:** GRE tunnels are logical point-to-point links between two routers that encapsulate packets in a generic routing encapsulation (GRE) header.
- **MPLS:** This method uses Multiprotocol Border Gateway Protocol (MP-BGP) to exchange MPLS (Multiprotocol Label Switching) labels for each of the different VPNs.
- **L2TPv3:**
  - L2TPv3 (Layer 2 Tunneling Protocol Version 3) is a technology that allows Layer 2 frames, such as Ethernet, PPP, and Frame Relay to be transported across an IP network.
  - L2TPv3 tunnels can be used to establish logical point-to-point links between two routers.

MPLS

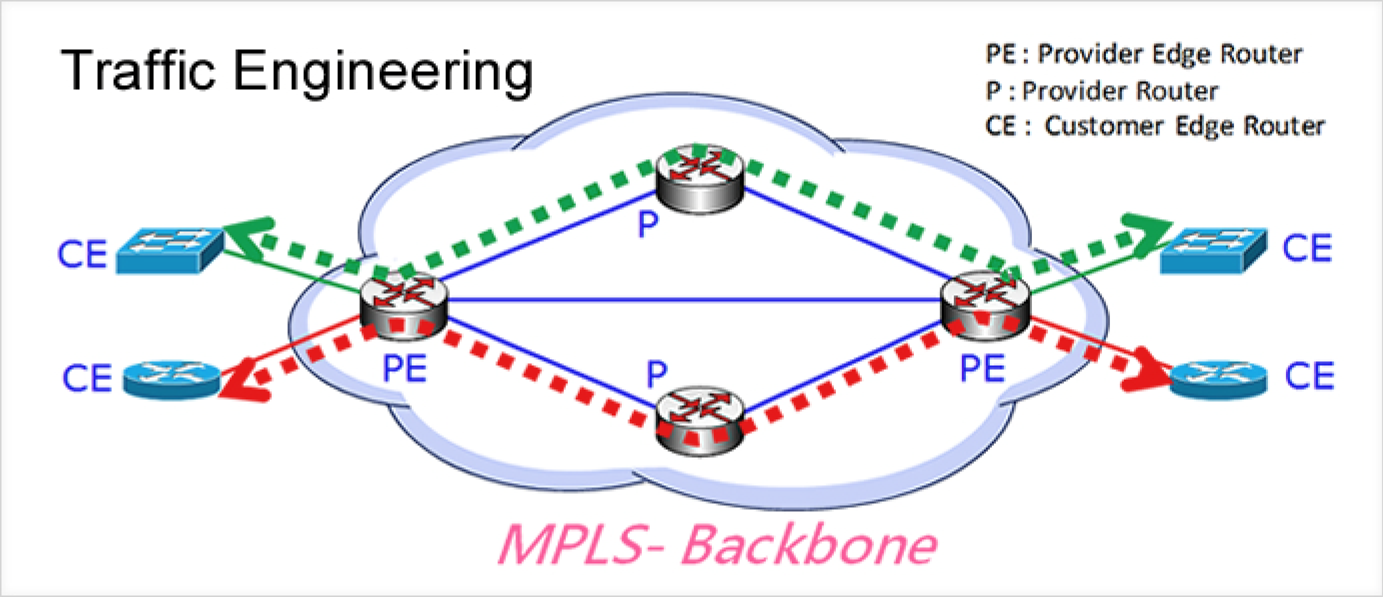
- The labels identify virtual links (*paths*) between distant nodes rather than endpoints.
- MPLS can encapsulate packets of various network protocols. MPLS supports a range of access technologies, including T1/E1, ATM, Frame Relay, and DSL.



Label = 20 bits
Exp = Experimental, 3 bits
S = Bottom of stack, 1bit
TTL = Time to live, 8 bits

MPLS

- Data packets are assigned labels.
- Packet-forwarding decisions are made solely on the contents of this label, without the need to examine the packet itself.
- This allows one to create end-to-end circuits across any type of transport medium, using any protocol.


Traffic Engineering
PE : Provider Edge Router
P : Provider Router
CE : Customer Edge Router
CE
CE
P
PE
P
PE
CE
CE
MPLS- Backbone