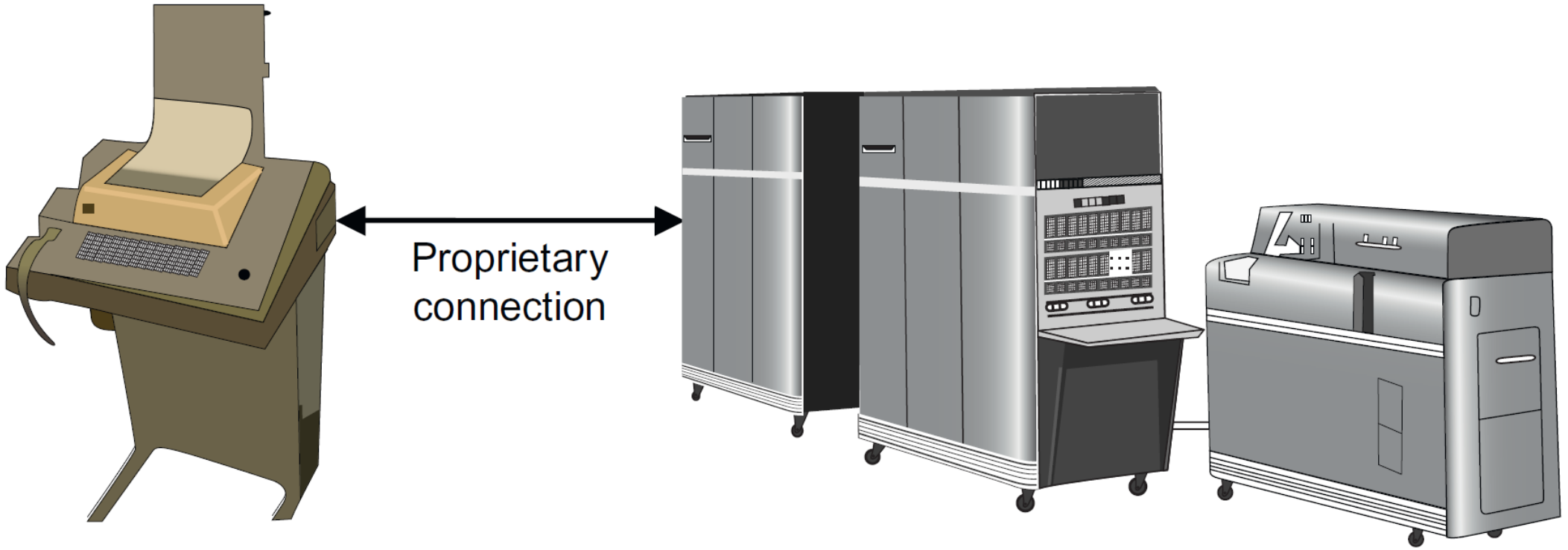
Days before Cloud

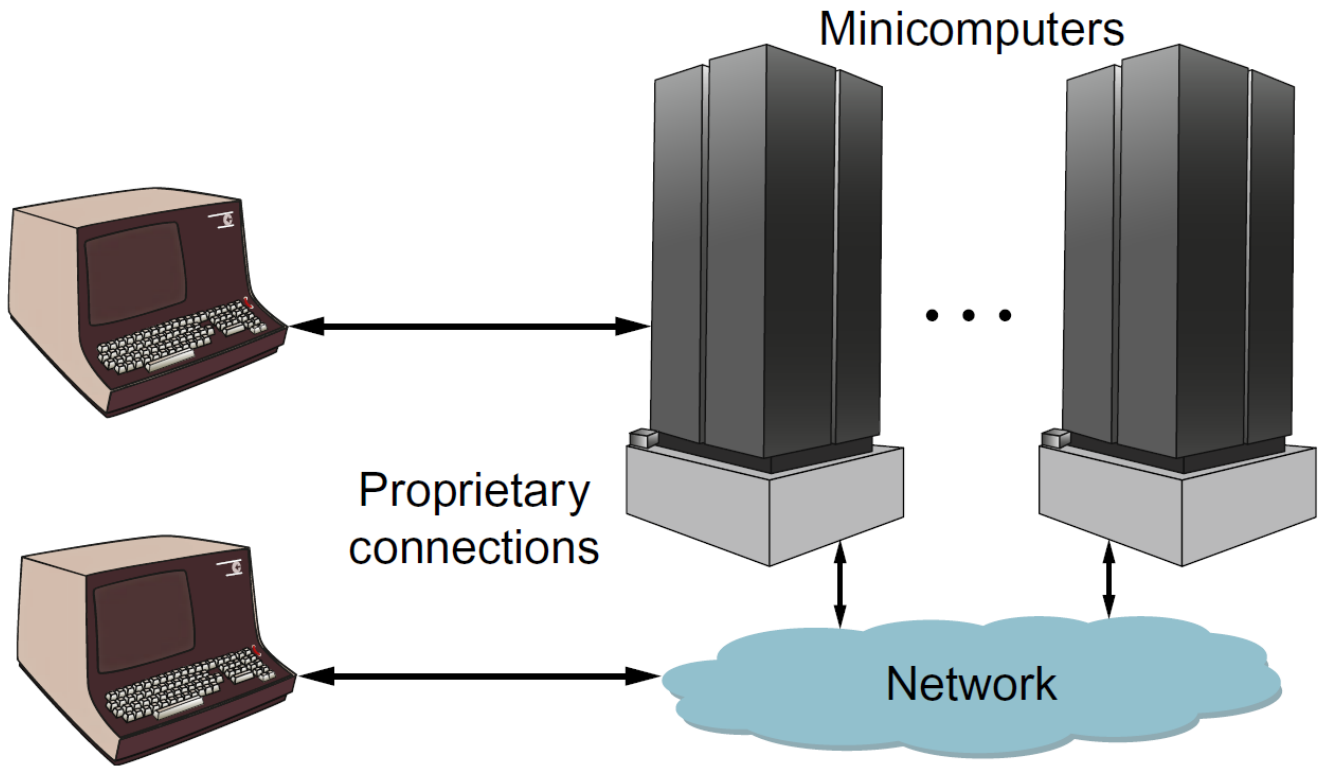
THE COMPONENTS

Data Center Evoluation

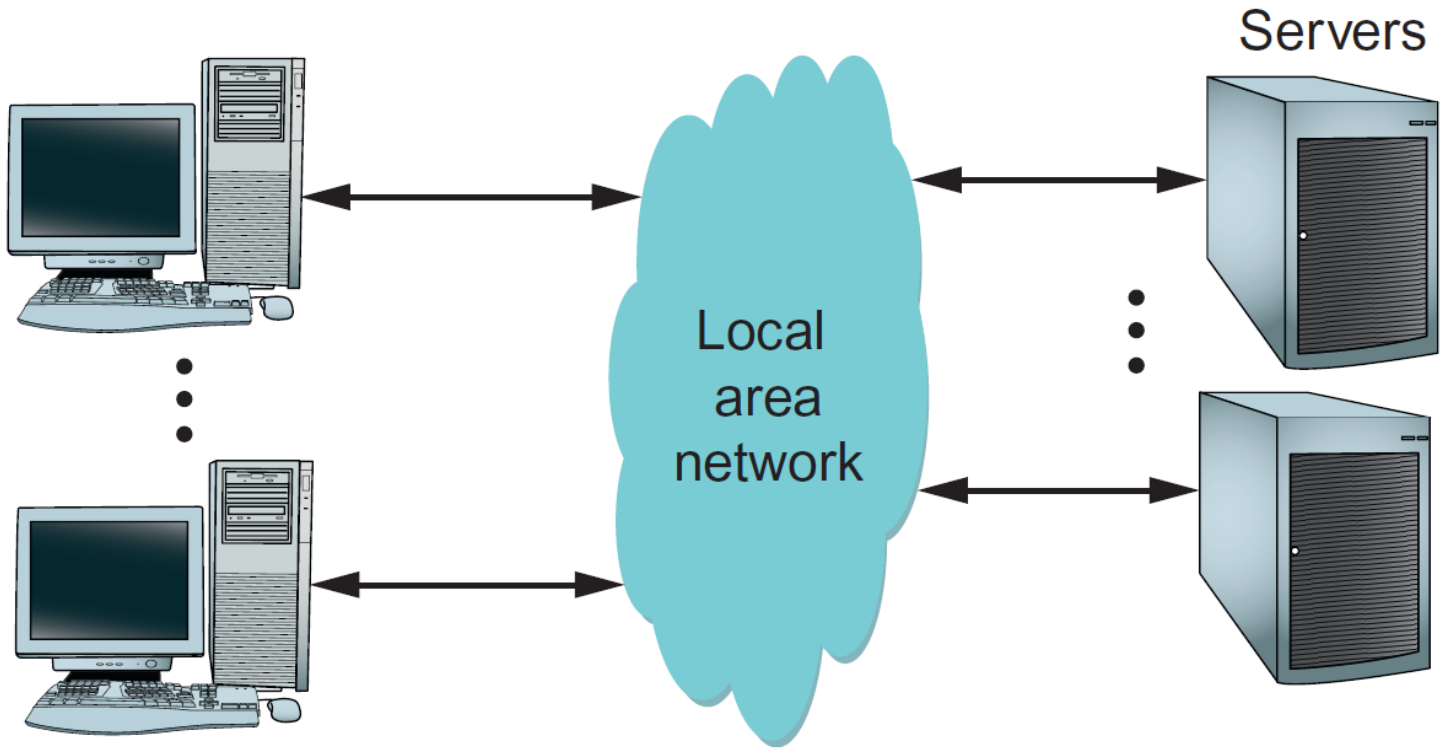


Minicomputers (late 1970s)

- Minic

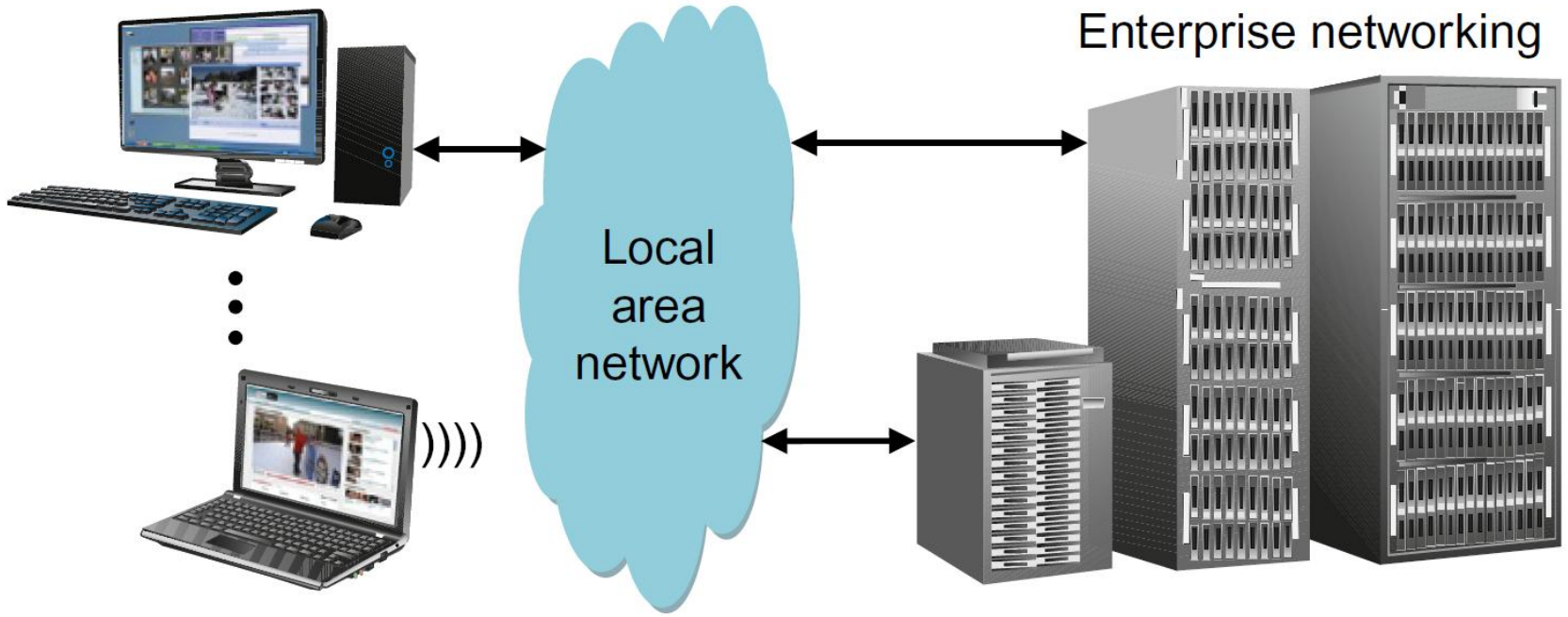


Servers (Late 1980s)



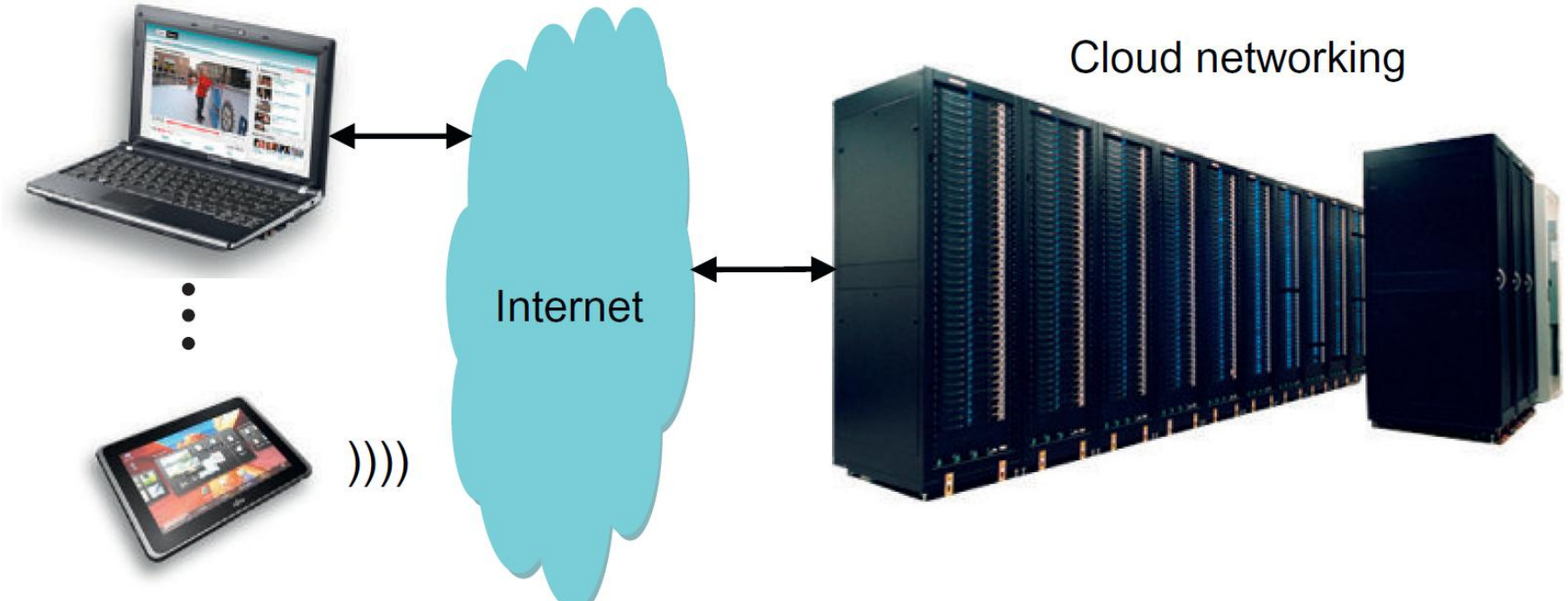
Enterprise data centers

- Around 2006, the first networking gear specifically designed for the data cente



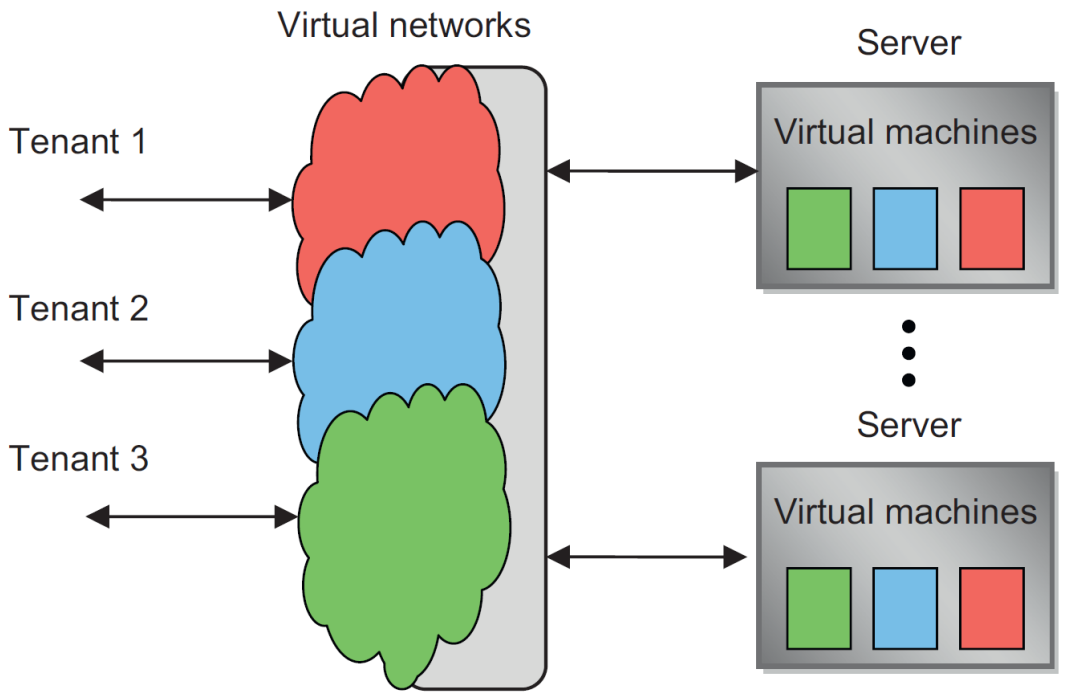
Cloud data centers

- Cloud data centers can contain tens of thousands of servers that must be connected to each other, to storage, and to the outside world.



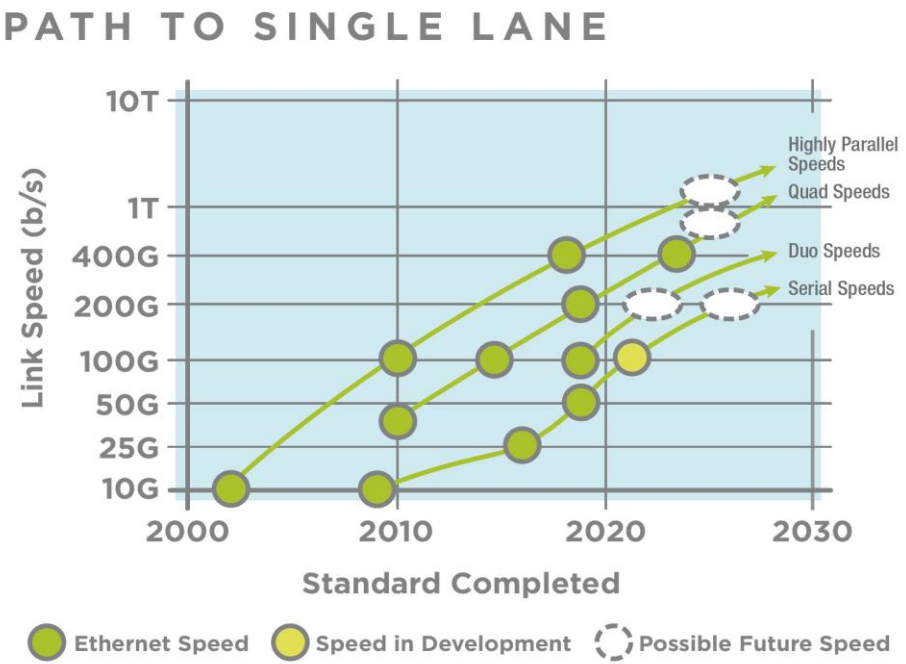
Virtualized data centers

- Multiple virtual data centers within their physical data centers.



CHARACTERISTICS of Cloud Networking

- Ethernet usage Ethernet has become the standard network technology.
  - Cloud data center server racks are interconnected using what is called direct attach copper cabling
  - Support 10GbE and 40GbE connections for distances up to a few meters



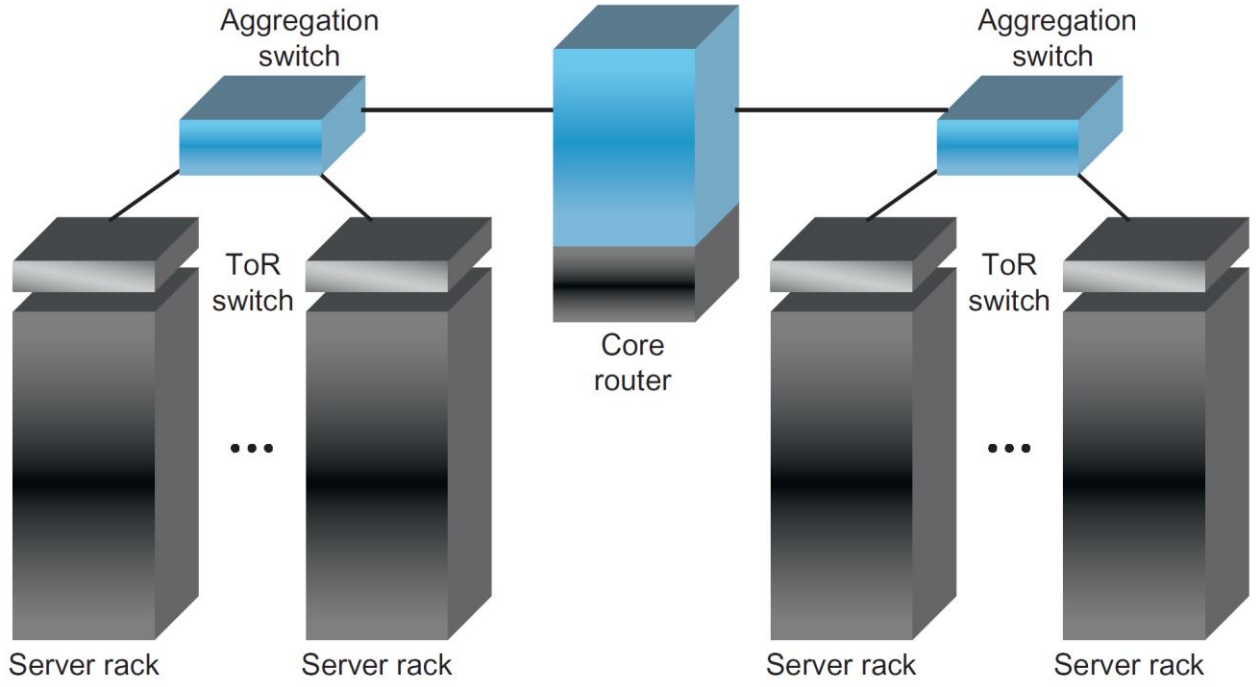
CHARACTERISTICS of Cloud Networking

- Virtualization
  - server virtualization can help improve resource utilization and, therefore, reduce operating costs. -> Multiple VM on a single physical server
- Convergence (Multi-Vendor Support)
  - data center bridging standards
  - Provide lossless operation and minimum bandwidth guarantees
- Scalability
  - Must interconnect tens of thousands of servers
- Software
  - Large cloud data center networks are set up, configured, and monitored using software
  - facilitated by software defined networking (SDN) initiatives such as OpenFlow.

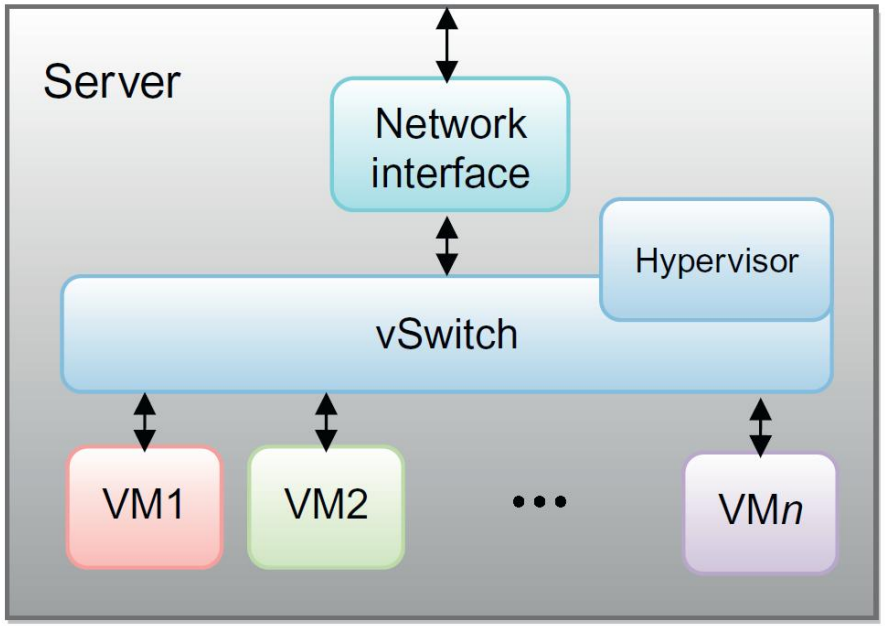
Intro to Data Center Networks (DCN)

FROM BGP IN THE DATA CENTER, CLOUD NETOWRKING)

Data Center Switch Type

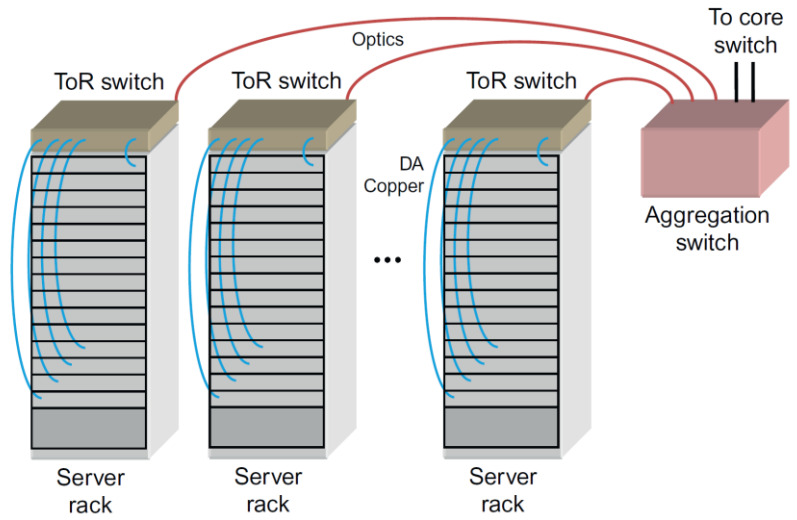


Virtual Switch



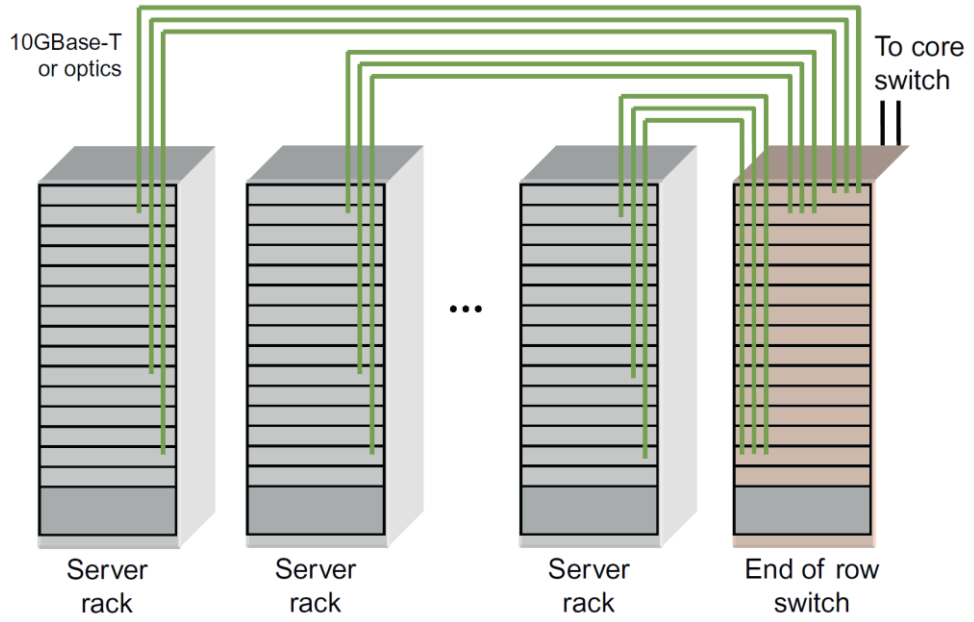
ToR Switch

- The servers within the rack connect to a ToR switch using a star topology
- Every server has a dedicated link to the ToR switch and the ToR switch may forward data to other servers in the rack, or out of the rack through high bandwidth uplink ports.



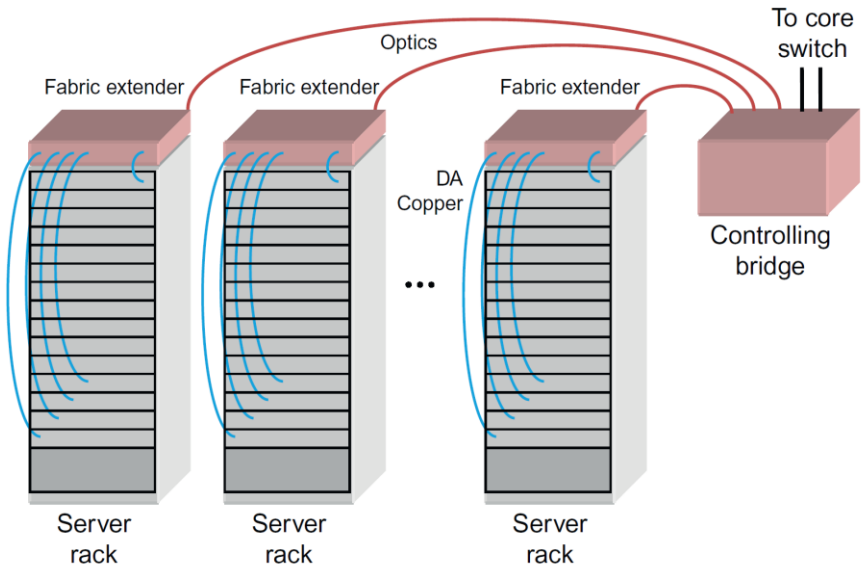
End of Row Switch

- EoR switches were developed in an attempt to reduce cost by sharing power, cooling, and management infrastructure across a large number of switch components.



Fabric extenders

- These are sometimes referred to as spine switches and are similar to ToR switches with the exception that they are controlled and monitored through a central controlling bridge. In some ways, they are like a distributed EoR switch.

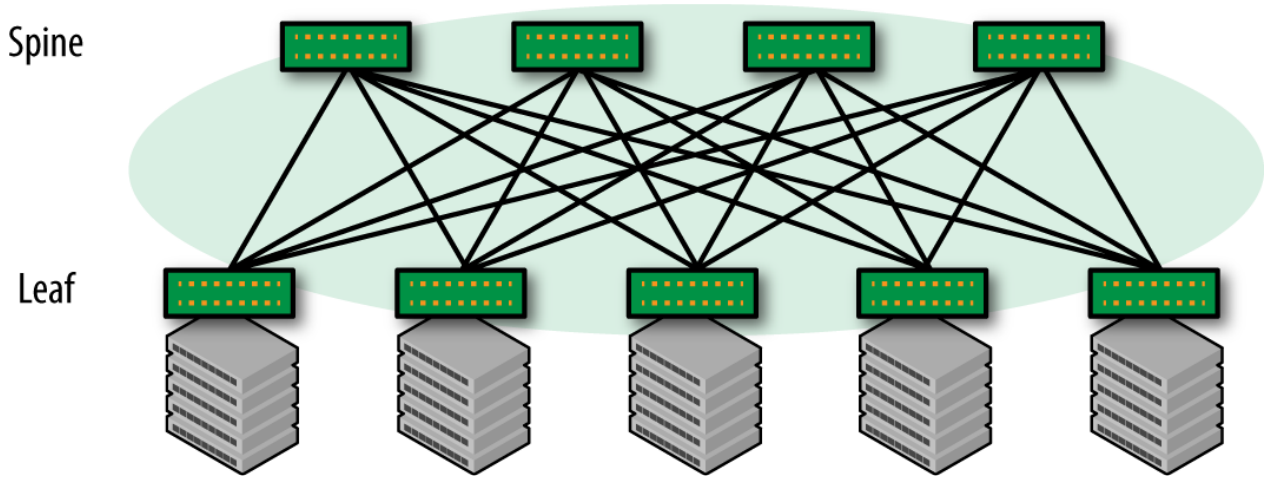


Requirement of DCN

- *Increased server-to-server communication*
  - Server-to-server communications is often called *East-West traffic*, because diagrams typically portray servers side-by-side. In contrast, traffic exchanged between local networks and external networks is called *North-South traffic*
- *SCALE*
  - Combined with increased server-to-server communication, the connectivity requirements at such scales force a rethink of how such networks are constructed.
- *Resilience*
  - The goal is an end-user experience mostly unaffected by network or server failures

Clos Network Topology

- Invertor by Charles Clos in 1950s
  - a telephony networking engineer
  - to solve a problem similar to the one faced by the web-scale pioneers: how to deal with the explosive growth of telephone networks.
- *A simple two-tier Clos network*

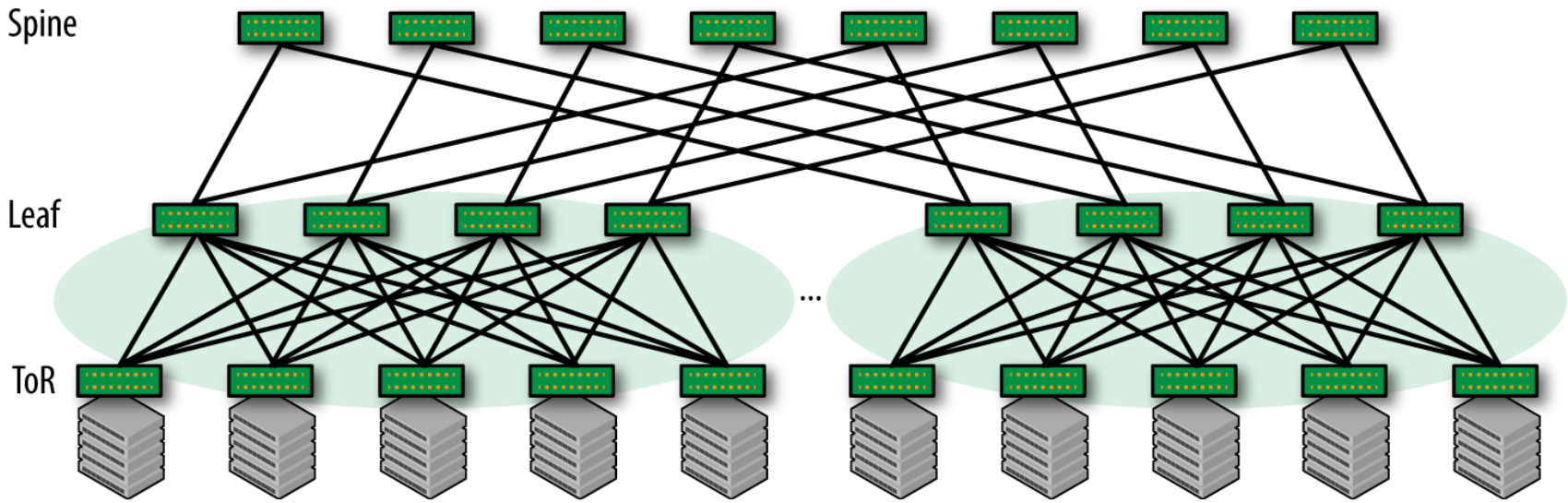


Some Properties

- First, servers are typically three network hops away from any other server.
- Next, the nodes are quite homogeneous: the servers look alike, as do the switches.
- The connectivity matrix is quite rich, which allows it to deal gracefully with failures. A single failure, or even multiple link failures, do not result in complete connectivity loss.
- the bandwidth between any two nodes is quite substantial. The bandwidth between nodes can be increased by adding more spines
- This model of scaling is called a *scale-out*

Three-Tier Clos Networks

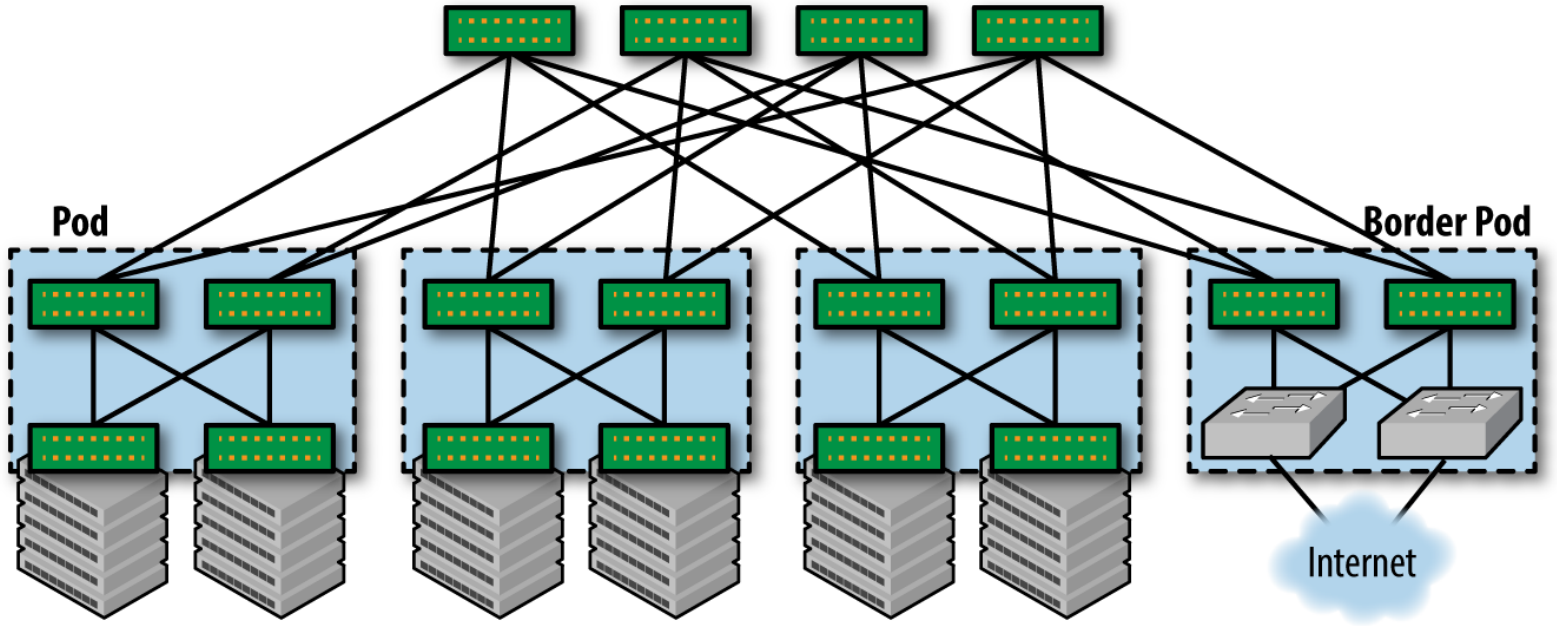
- Each two-tier network is called a pod or cluster, and the third tier of spines
- connecting all the pods is called an interpod spine or intercluster
- spine layer. Quite often, the first tier of switches, the ones servers connect to, are called top-of-rack (ToR)



- the total number of servers that you can connect is *n*^3 / 4.
  - Assuming 64-port switches, for example, we get 64^3 / 4 = 65,536 servers. Assuming the more realistic switch port numbers and servers per rack from the previous section, we can build 40 * 16 * 16 = 10,240 servers.

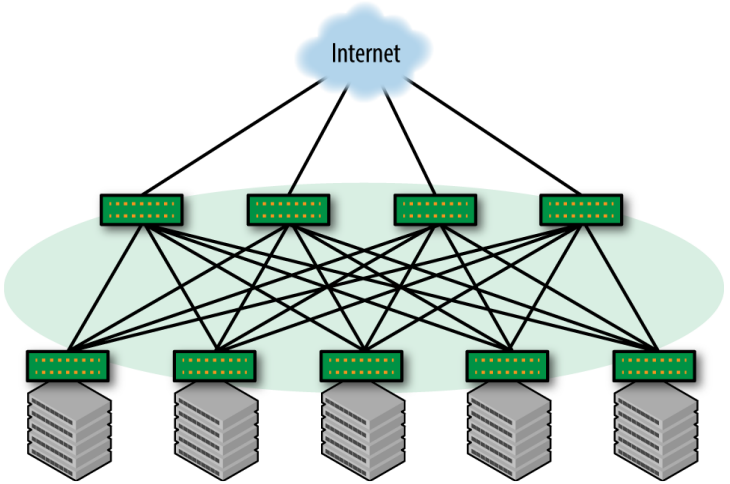
Connectivity to the External World

- this connectivity happens through what are called *border ToRs* or *border pods*.



Advantage of Border Pods

- they isolate the inside of the data center from the outside. The routing protocols that are inside the data center never interact with the external world, providing a measure of stability and security.
- Smaller networks might not be able to dedicate separate switches just to connect to the external world. Such networks might connect to the outside world via the spines



Support for Multitenancy (or Cloud)

- The additional goals of a cloud architecture are as follows:
- *Agility*
  - Given the typical use of the cloud, whereby customers spin up and tear down networks rapidly.
- *Isolation*
  - One customer's traffic must not be seen by another customer.
- *Scale*
  - Large numbers of customers, or tenants, must be supported.

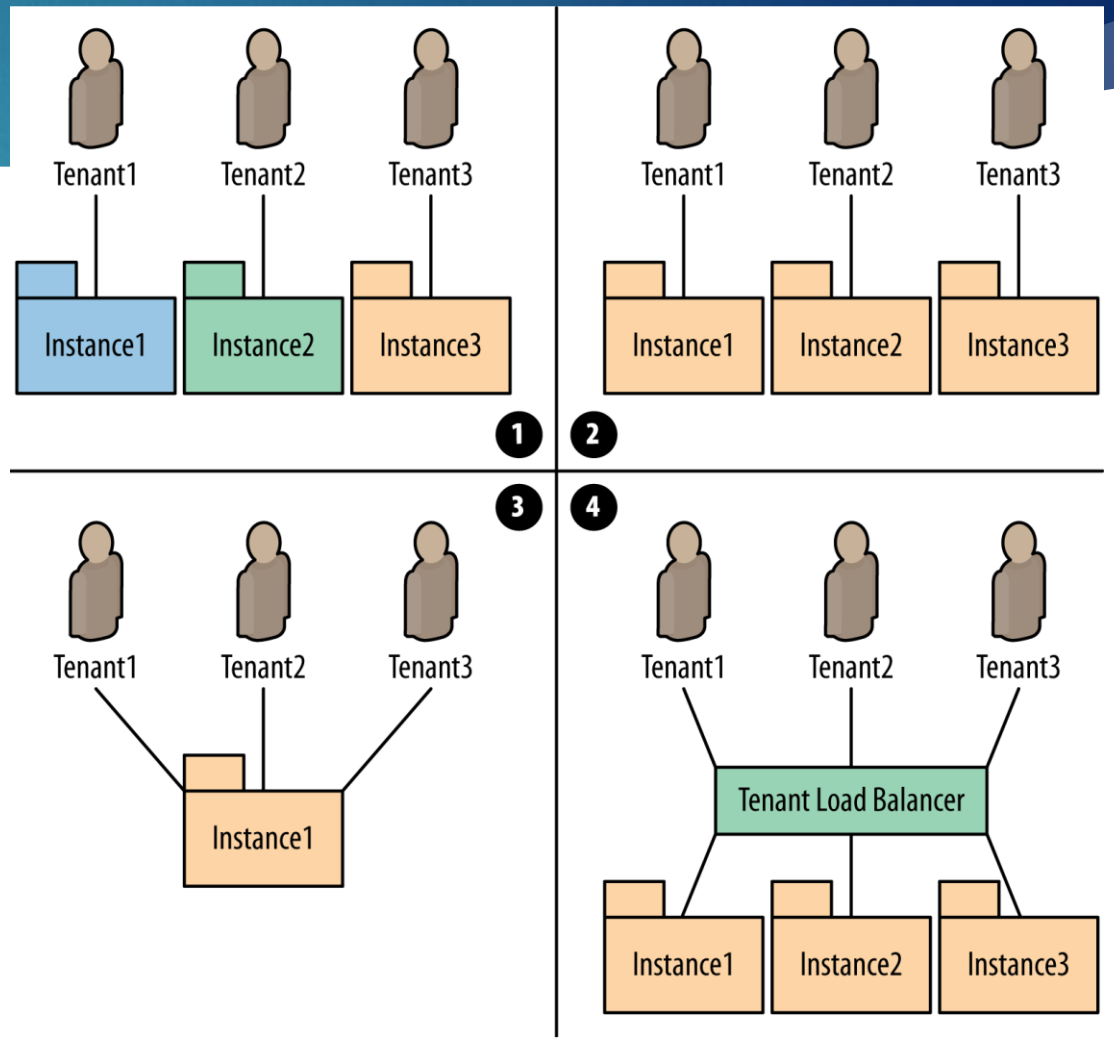
Choice of Routing Protocol

- Open Shortest Path First (OSPF)
  - OSPF required two separate protocols, similar mostly in name and basic function, to support both IPv4 and IPv6 networks.
- Intermediate System–to–Intermediate System (IS-IS)
  - IS-IS is a far better regarded protocol that can route both IPv4 and IPv6 stacks. However, good IS-IS implementations are few, limiting the administrator's choices.
- BGP stepped into such a situation and promised something that the other two couldn't offer.
  - mature, powers the internet, and is fundamentally simple to understand
  - Many mature and robust implementations of BGP exist, including open source.
  - supports multiprotocols (i.e., it supports advertising IPv4, IPv6, Multiprotocol Label Switching (MPLS), and VPNs natively).

The Multitenant Data CENTER

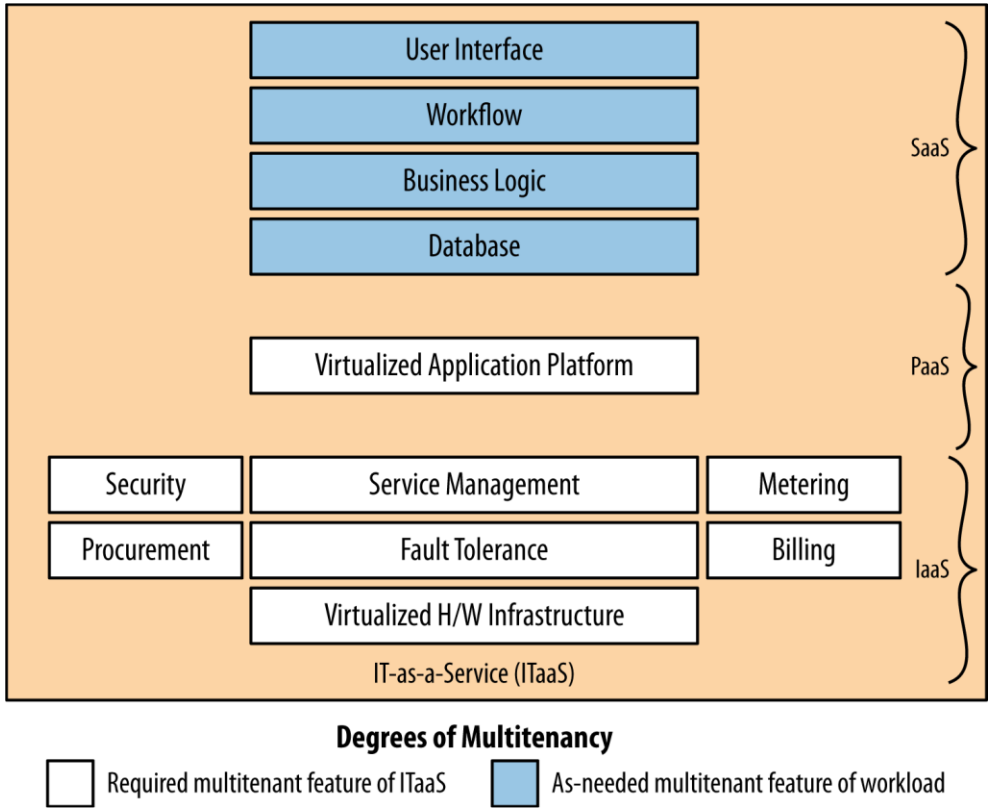
Multitenant Data Center

- Multitenancy is different than multiuser or multi-enterprise
- Tenancy occurs above the user or enterprise boundary.
- Multitenancy is common in both public and private clouds and not limited solely to Infrastructure as a Service (IaaS) data center offerings.



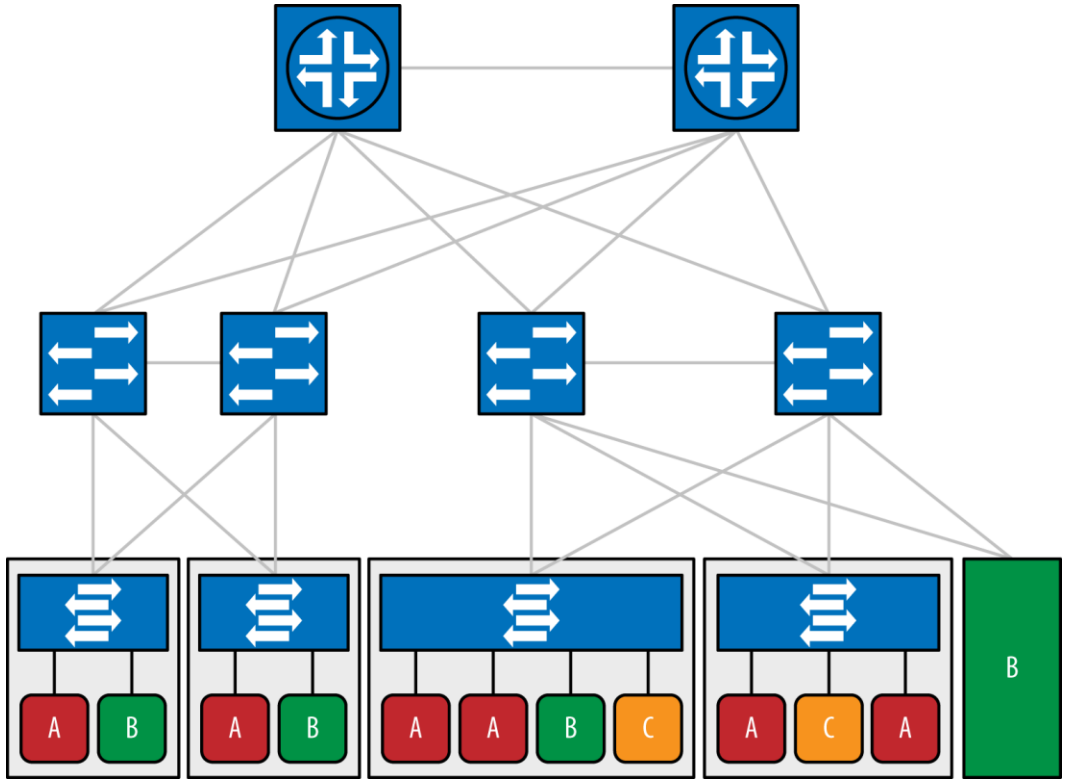
Cloud Service and Multitency

- *Infrastructure as a service (IaaS)*
  - Infrastructure (compute, storage, and network) are shared. Exemplified by Amazon.
- *Platform as a service (PaaS)*
  - An application development environment is shared. Exemplified by Google Apps.
- *Software as a service (SaaS)*
  - An application is shared. Exemplified by Salesforce.com.



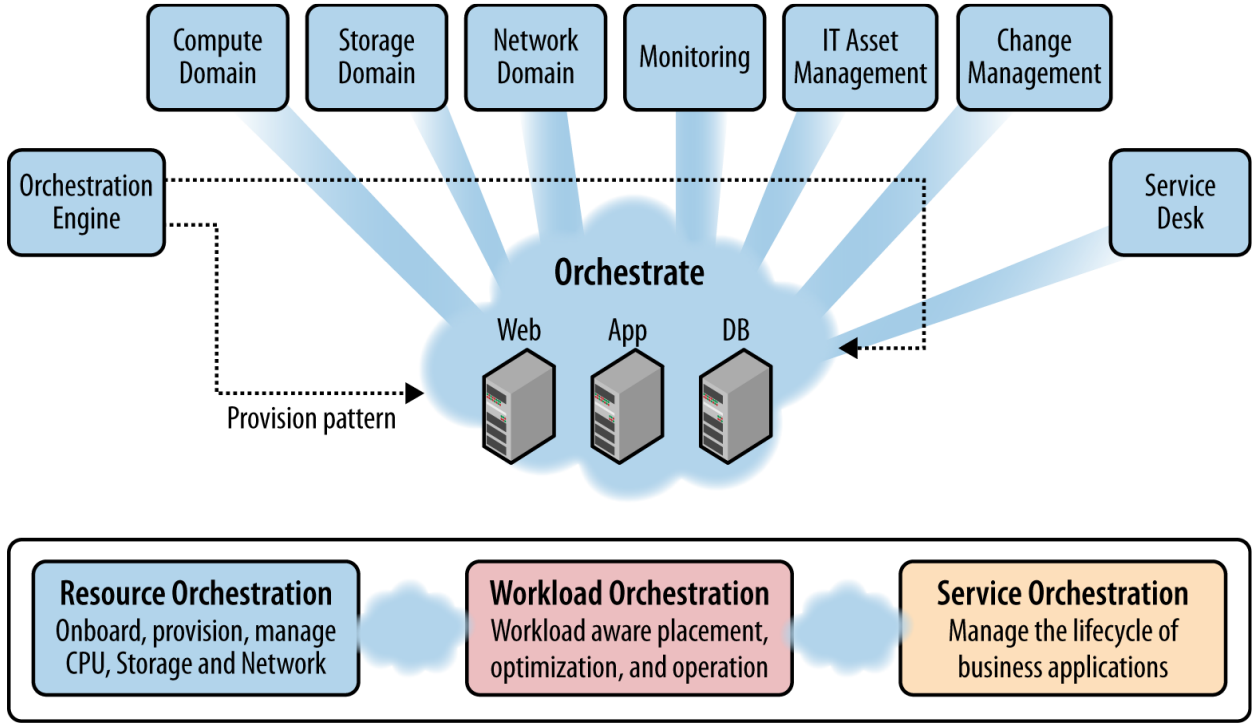
The Virtualized Multitenant Data Center

- each tenant corresponds to a set of virtual machines, which are hosted on servers running hypervisors.
- The hypervisors contain virtual switches (vSwitches) to connect the virtual machines to the physical network and to one another.



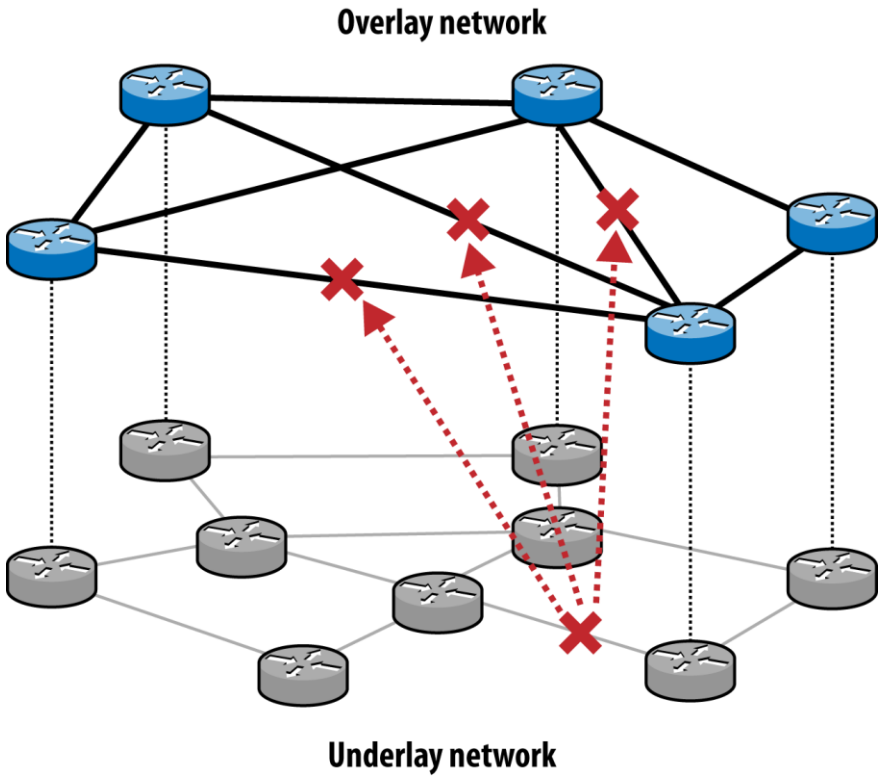
Orchestration

- Orchestration in a data center provides the logically centralized control and interaction point for network operators and is a central point of control of other network controllers.

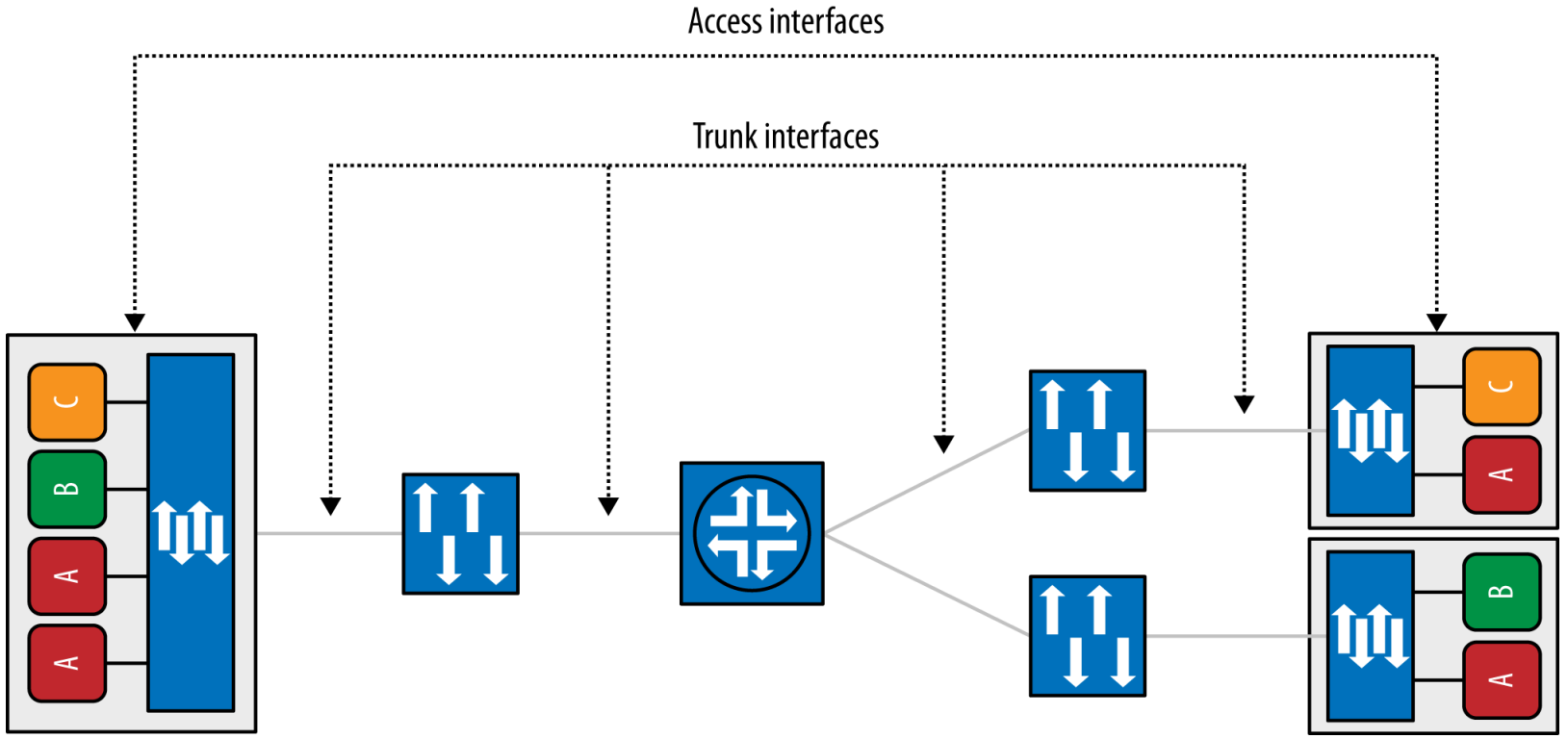


SDN Solutions for the Data Center Network

- The Network Underlay/Overlay
- Typically, the overlay network is created using virtual network elements such as virtual routers, switches, or logical tunnels.
- The underlay network is a typical network, such as an Ethernet, MPLS, or IP network

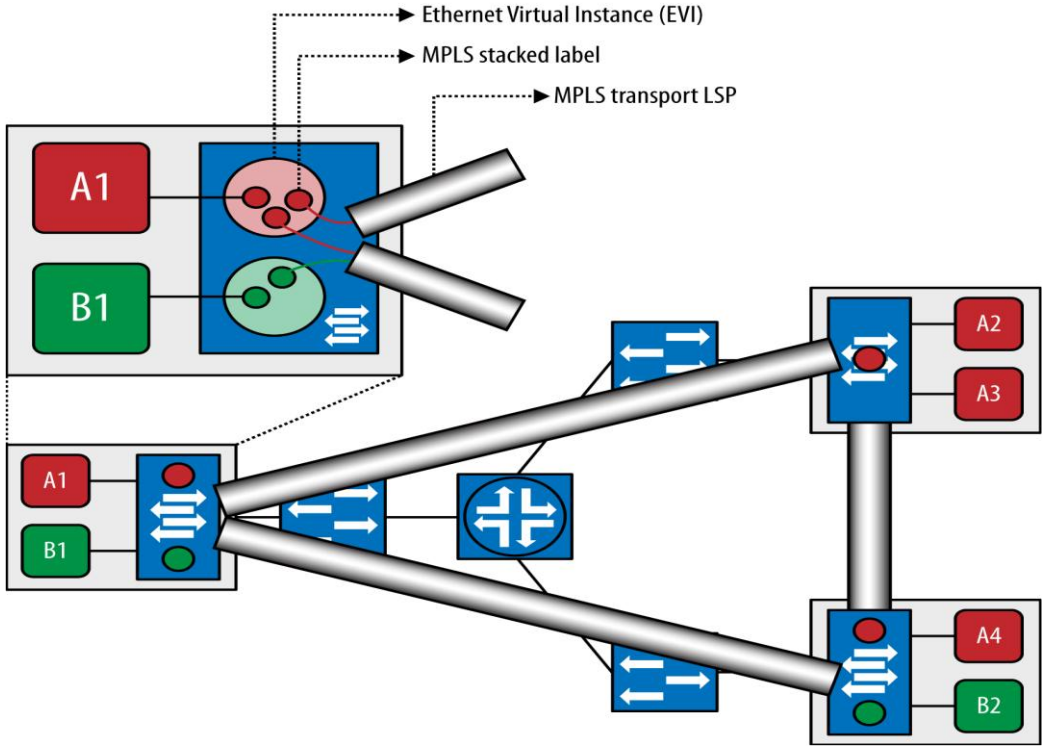


VLAN

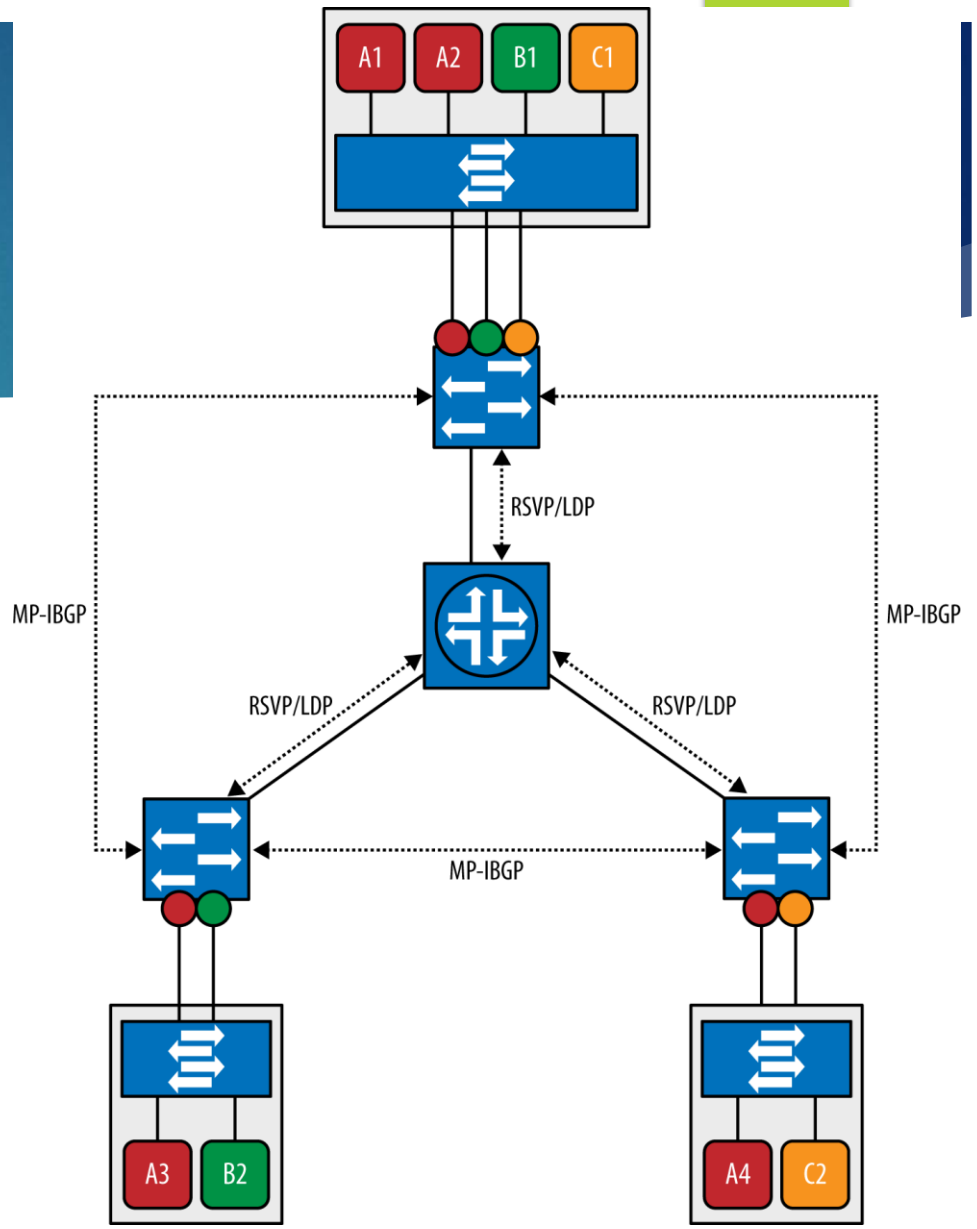


Ethernet VPN (EVPN)

- Can support beyond the 4000 VLAN tag limit

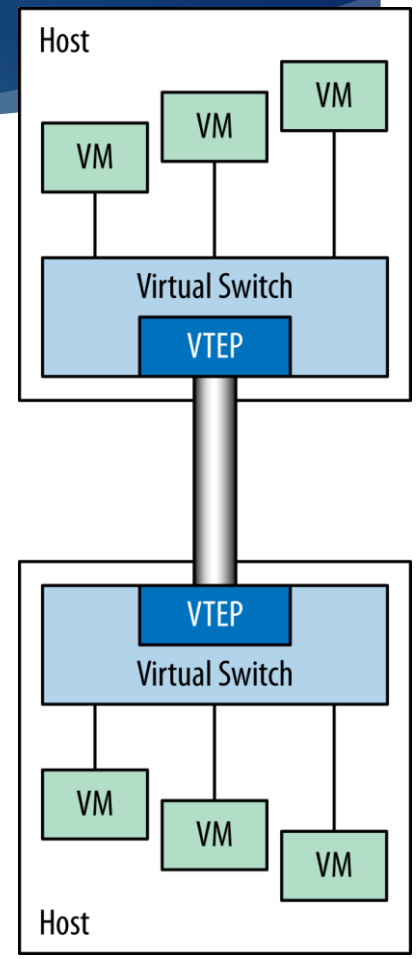


EVPN (data plane)



EVPN (control plane)

VxLAN



- Virtual Extensible LAN (VxLAN) is a network virtualization technology that attempts to ameliorate the scalability problems encountered with large cloud computing deployments when using existing VLAN technology.

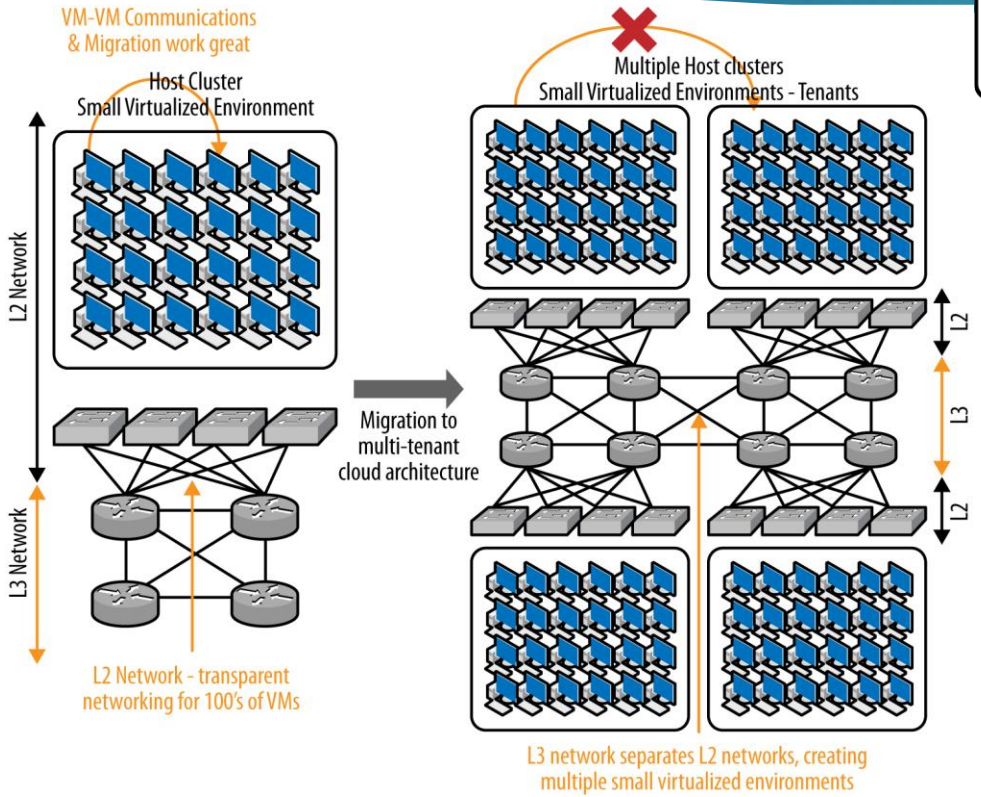
Outer MAC	Outer IP	Outer UDP	VXLAN	Inner MAC	Inner Payload	Outer CRC

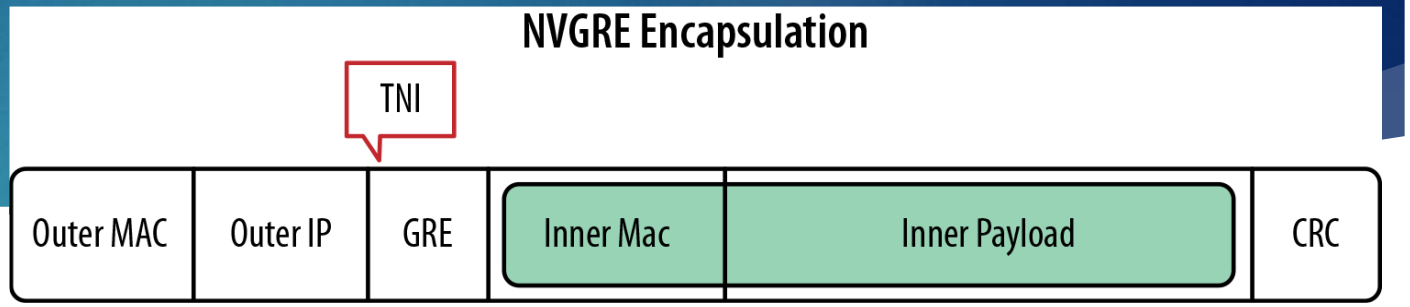
- The VxLAN Identifier space is 24 bits, allowing the VxLAN Id space
- to increase by over 400,000 percent to handle over 16 million unique identifiers.
- The normal operation of VxLAN relies on Virtual Tunnel Endpoints (VTEPs) that contain all the functionality needed to provide Ethernet layer 2 services to connected end systems.

The Network Virtualization using Generic Routing Encapsulation (NVGRE)

- A network virtualization technology that was invented in order to overcome the scalability problems associated with large data center environments that suffer from the issues described earlier in the VLAN underlay option.
  - Similar to VxLAN, it employs a packet tunneling scheme that encapsulates layer 2 information inside of a layer 3 packet.
  - NVGRE enables the connection between two or more L3 networks and makes it appear to end hosts as if they share the same L2 subnet

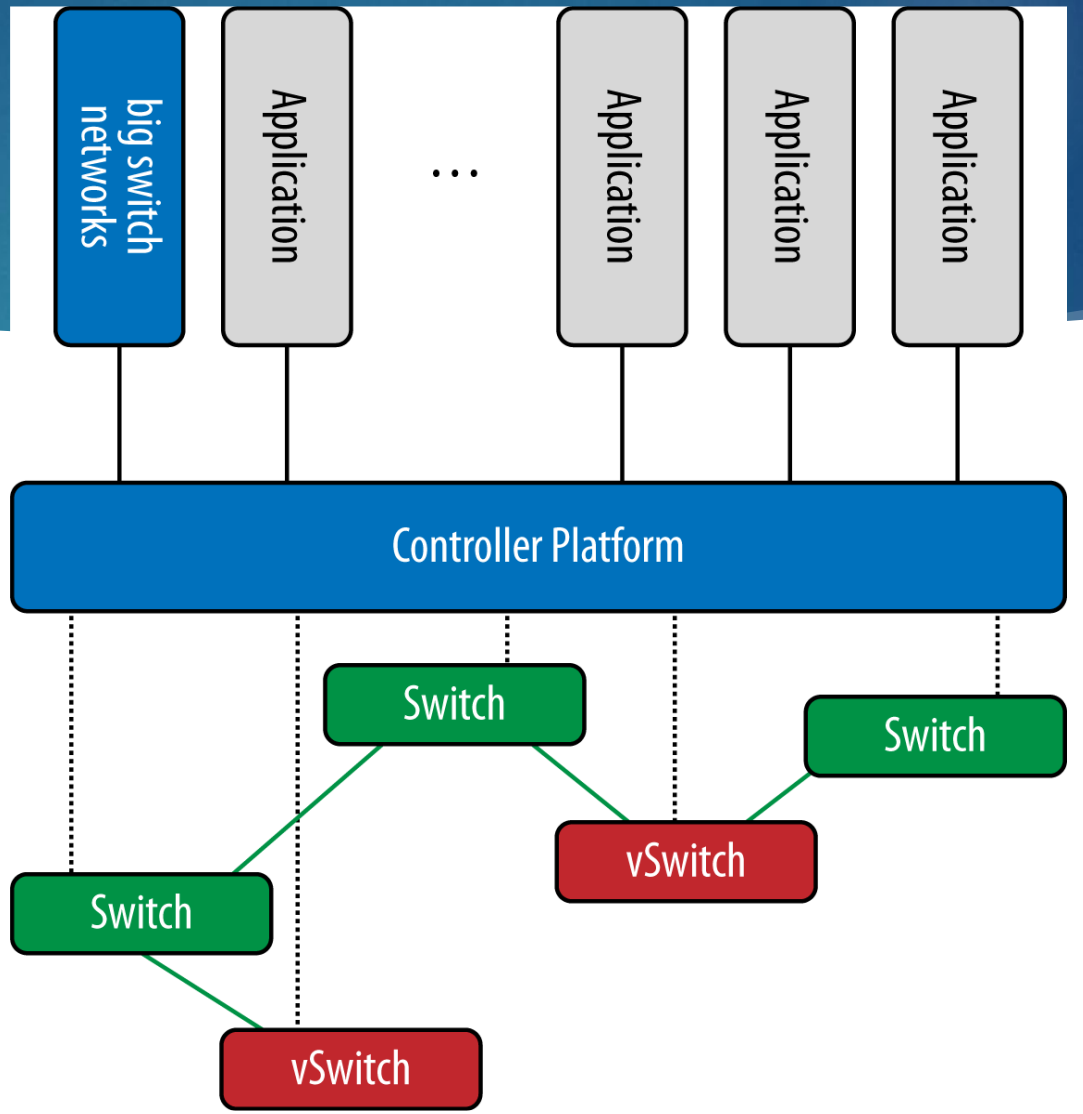
NVGRE





- NVGRE uses a unique 24-bit ID called a Tenant Network Identifier (TNI) that is added to the L2 Ethernet frame.
- The TNI is mapped on top of the lower 24 bits of the GRE Key field. T
- This new 24-bit TNI now enables more than 16 million L2 (logical) networks to operate within the same administrative domain,

OpenFlow


big switch networks
Application
...
Application
Application
Application
Controller Platform
Switch
Switch
Switch
vSwitch
vSwitch

Server Virtualization

- Server virtualization is the process of using software to create multiple independent virtual servers (virtual machines) or multiple independent containerized operating systems (containers) on a physical x86 server.
- Network functions virtualization (NFV) is the process of virtualizing specific network functions, such as a firewall function, into a virtual machine (VM) so that they can be run in common x86 hardware instead of a dedicated appliance.
- VMs and containers increase the overall efficiency and cost-effectiveness of a server by maximizing the use of the available resources.

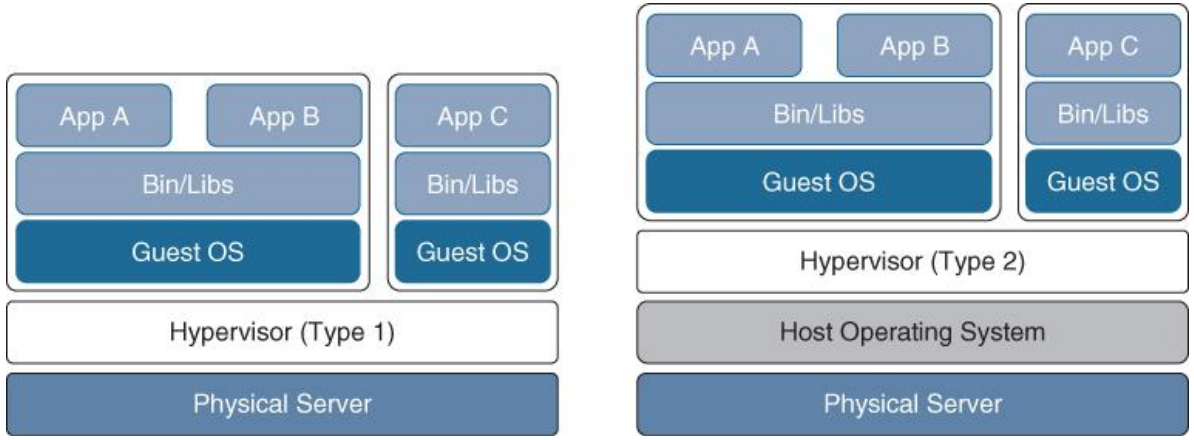
Server Virtualization

Virtual Machines

A virtual machine (VM) is a software emulation of a physical server with an operating system. The virtualization software that creates VMs and performs the hardware abstraction to allow multiple VMs to run concurrently is known as a **hypervisor**.

Type 1 vs Type 2

- **Type 1:** This type of hypervisor runs directly on the system hardware. It is commonly referred to as "bare metal" or "native." Examples include: VMware vSphere, Microsoft Hyper-V, Citrix XenServer, and Red Hat Kernel-based Virtual Machine (KVM).
- **Type 2:** This type of hypervisor (for example, VMware Fusion) requires a host OS to run. This is the type of hypervisor that is typically used by client devices.



Type 1 and Type 2 Hypervisors

Server Virtualization

Virtual Machines Advantages

One key capability of VMs is that they can be migrated from one server to another while preserving transactional integrity during movement. This has many advantages. For example, if a physical server needs a memory upgrade, the VMs can be migrated to other servers with no downtime. Another advantage is that it provides high availability. For example, if a server fails, the VMs can be spun up on other servers in the network, as illustrated in Figure 27-3.

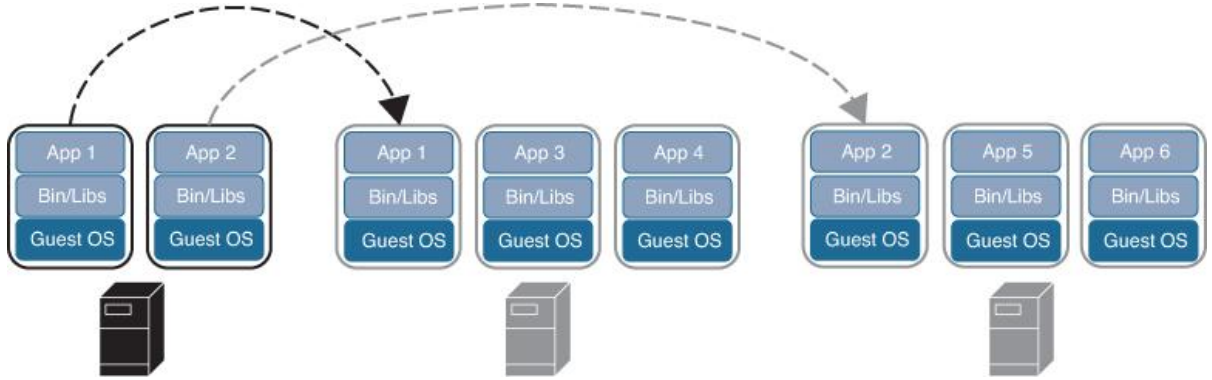


Figure 27-3 *VM Migration*

Server Virtualization

Containers

A container is an isolated environment where containerized applications run. It contains the application, along with the dependencies that the application needs to run. Though they have similarities to VMs, containers are not the same as VMs.

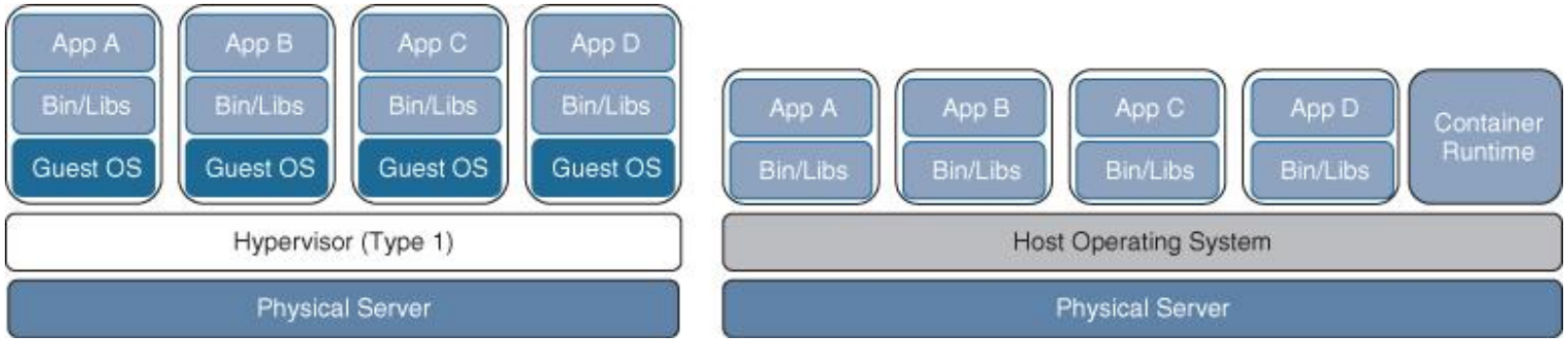


Figure 27-4 *Side-by-Side Comparison of VMs and Containers*

Server Virtualization

Virtual Switching

- A virtual switch (vSwitch) is a software-based Layer 2 switch that operates like a physical Ethernet switch.
- A vSwitch enables VMs to communicate with each other within a virtualized server and with external physical networks through the physical network interface cards (pNICs).
- Multiple vSwitches can be created under a virtualized server, but network traffic cannot flow directly from one vSwitch to another vSwitch within the same host, and the vSwitches cannot share the same pNIC.

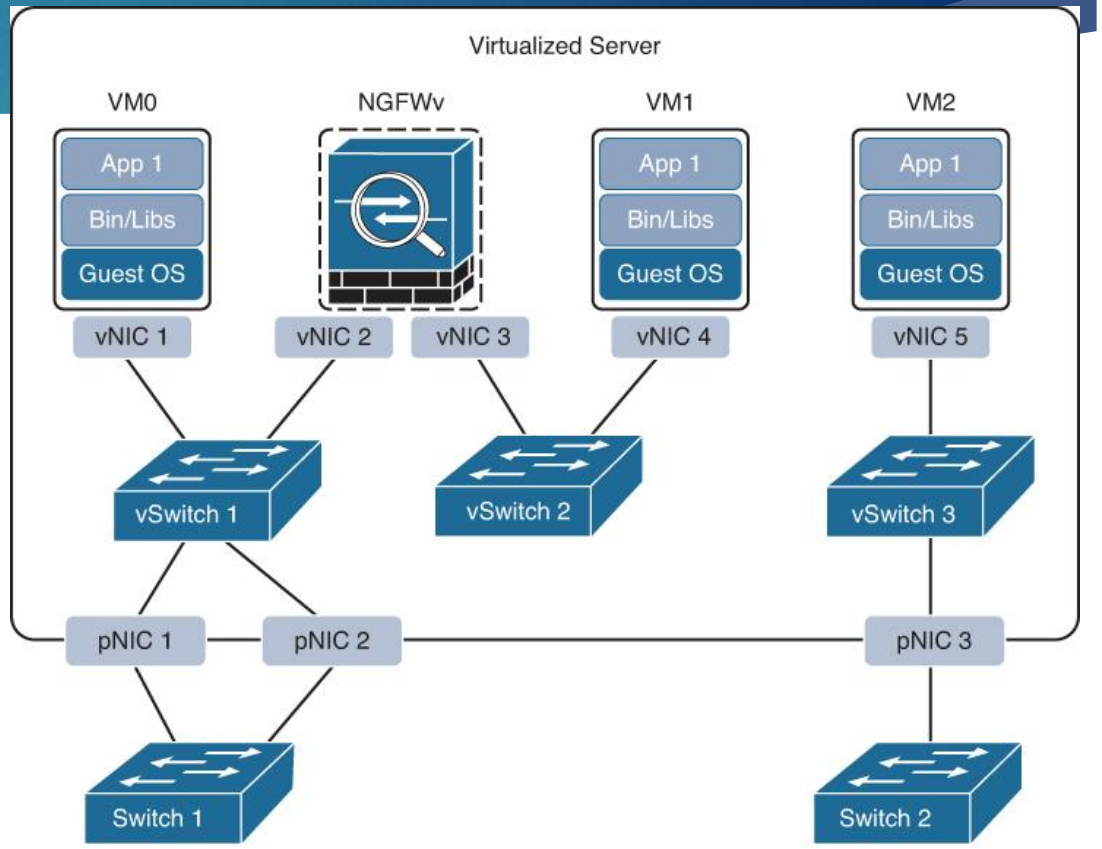


Figure 27-5 *Virtualized Server with vSwitches*

Server Virtualization

Distributed Virtual Switching Benefits

One of the downsides of standard vSwitches is that every vSwitch that is part of a cluster of virtualized servers needs to be configured individually in every virtual host. This problem is solved by using distributed virtual switching, a feature that aggregates vSwitches together from a cluster of virtualized servers and treats them as a single distributed virtual switch. Benefits include:

- Centralized management of vSwitch configuration for multiple hosts in a cluster
- Migration of networking statistics and policies with virtual machines during a live VM migration
- Configuration consistency across all the hosts that are part of the distributed switch

Like VMs, containers rely on vSwitches (also known as virtual bridges) for communication within a node (server) or the outside world.

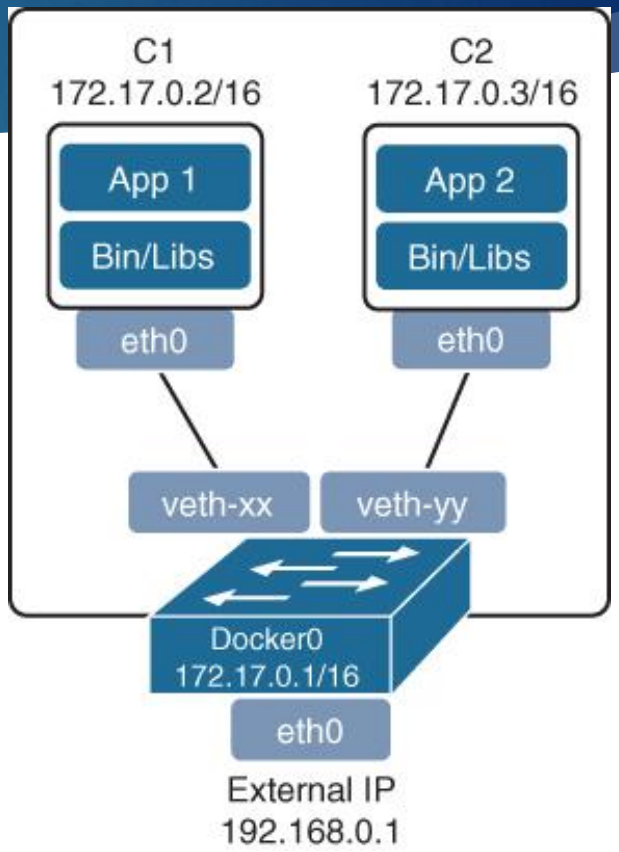


Figure 27-6 *Container Bridging*