- Switching Architectures -

Network Traffic Models

Traffic flow is an important consideration when designing scalable, efficient networks. Fundamentally, this involves understanding two things:

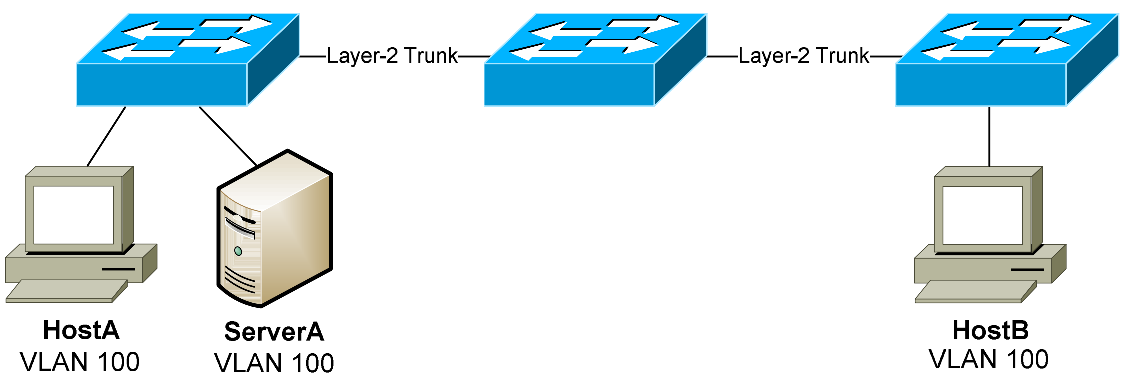
- Where do *resources* reside?
- Where do the *users* reside that access those resources?

Legacy networks adhered to the **80/20 design**, which dictated that**:**

- 80 percent of traffic should remain on the local network.
- 20 percent of traffic should be routed to a remote network.

To accommodate this design practice, resources were placed as close as possible to the users that required them. This allowed the majority of traffic to be *switched*, instead of *routed*, which reduced latency in legacy networks.

The 80/20 design allowed VLANs to be trunked across the entire campus network, a concept known as **end-to-end VLANs**:



End-to-end VLANs allow a host to exist *anywhere* on the campus network, while maintaining Layer-2 connectivity to its resources.

However, this *flat* design poses numerous challenges for scalability and performance:

- STP domains are very large, which may result in instability or convergence issues.
- Broadcasts proliferate throughout the entire campus network.
- Maintaining end-to-end VLANs adds administrative overhead.
- Troubleshooting issues can be difficult.

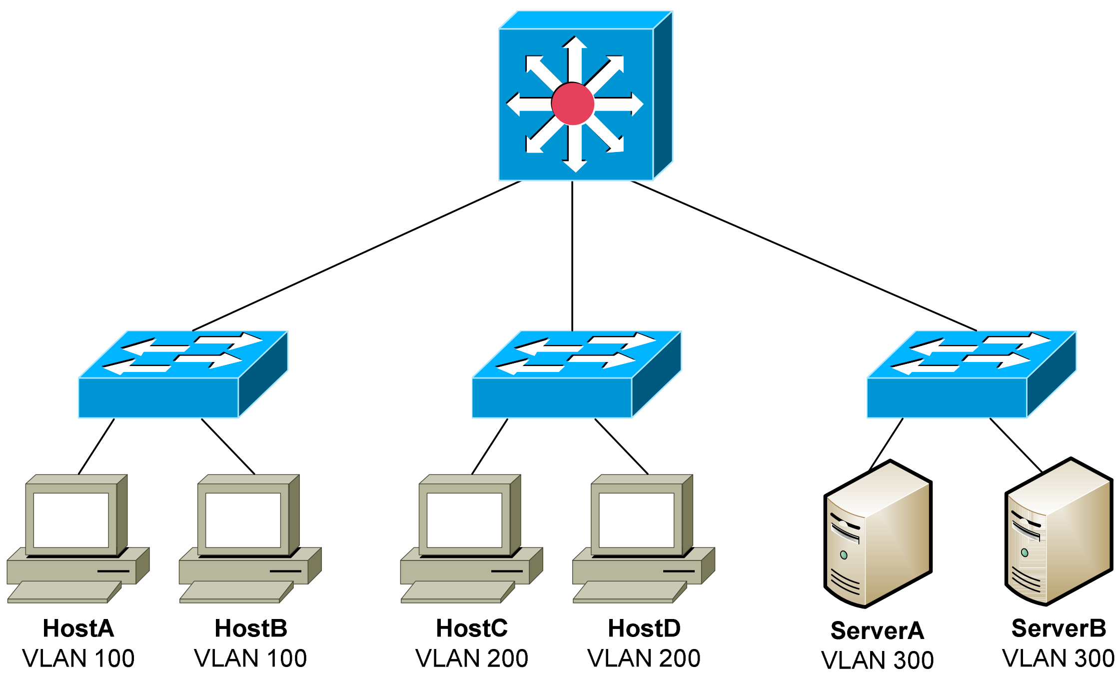
Network Traffic Models (continued)

As network technology improved, centralization of resources became the dominant trend. Modern networks adhere to the **20/80 design**:

- 20 percent of traffic should remain on the local network.
- 80 percent of traffic should be routed to a remote network.

Instead of placing *workgroup* resources in every local network, most organizations centralize resources into a datacenter environment. Layer-3 switching allows users to access these resources with minimal latency.

The 20/80 design encourages a **local VLAN** approach. VLANs should stay **localized** to a single switch or switch block:



This design provides several benefits:

- STP domains are limited, reducing the risk of convergence issues.
- Broadcast traffic is isolated within smaller broadcast domains.
- Simpler, *hierarchical* design improves scalability and performance.
- Troubleshooting issues is typically easier.

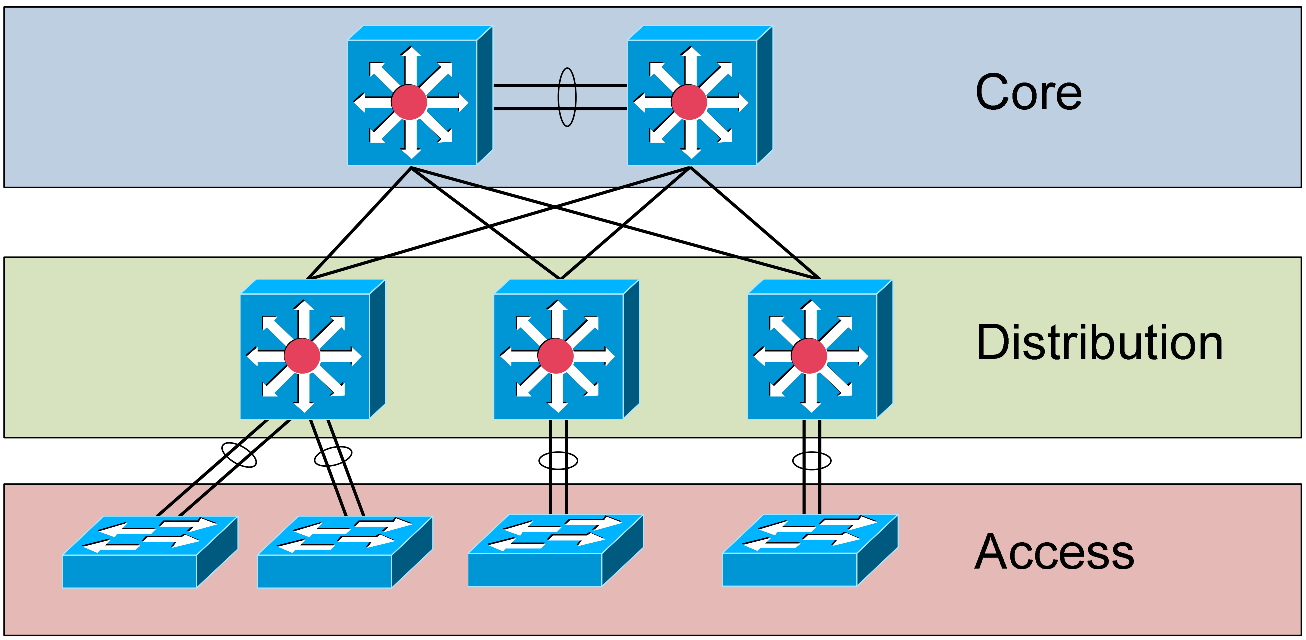
There are nearly no drawbacks to this design, outside of a legacy application requiring Layer-2 connectivity between users and resources. In that scenario, it's time to invest in a better application. ☺

The Cisco Hierarchical Network Model

To aid in designing scalable networks, Cisco developed a **hierarchical network model,** which consists of three layers:

- **Access** layer
- **Distribution** layer
- **Core** layer

Cisco Hierarchical Model – Access Layer



The **access layer** is where users and hosts connect into the network. Switches at the access layer typically have the following characteristics:

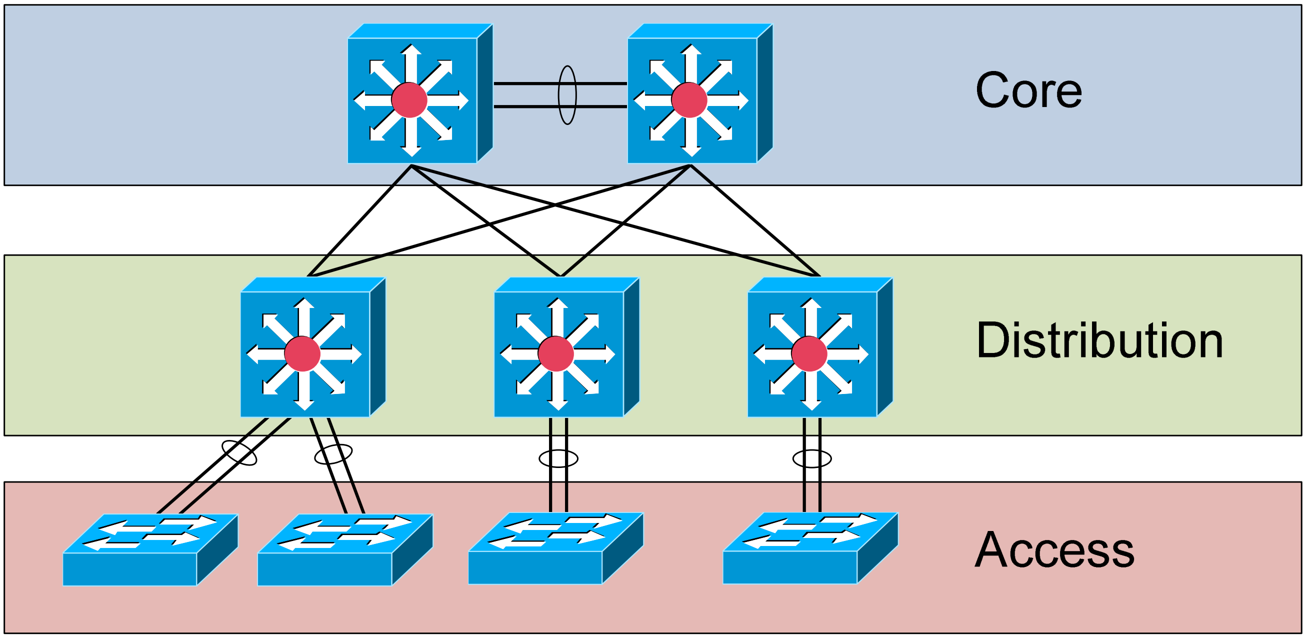
- High port density
- Low cost per port
- Scalable, redundant uplinks to higher layers
- Host-level functions such as VLANs, traffic filtering, and QoS

In an *80/20* design, resources are placed as close as possible to the users that require them. Thus, most traffic will never need to leave the access layer.

In a *20/80* design, traffic must be forwarded through higher layers to reach centralized resources.

(Reference: CCNP Switch 642-813 Official Certification Guide by David Hucaby. Cisco Press)

Cisco Hierarchical Model – Distribution Layer



The **distribution layer** is responsible for aggregating access layer switches, and connecting the access layer to the core layer. Switches at the distribution layer typically have the following characteristics:

- Layer-3 or multilayer forwarding
- Traffic filtering and QoS
- Scalable, redundant links to the core and access layers

Historically, the distribution layer was the Layer-3 boundary in a hierarchical network design:

- The connection between access and distribution layers was Layer-2.
- The distribution switches are configured with VLAN SVIs.
- Hosts in the access layer use the SVIs as their default gateway.

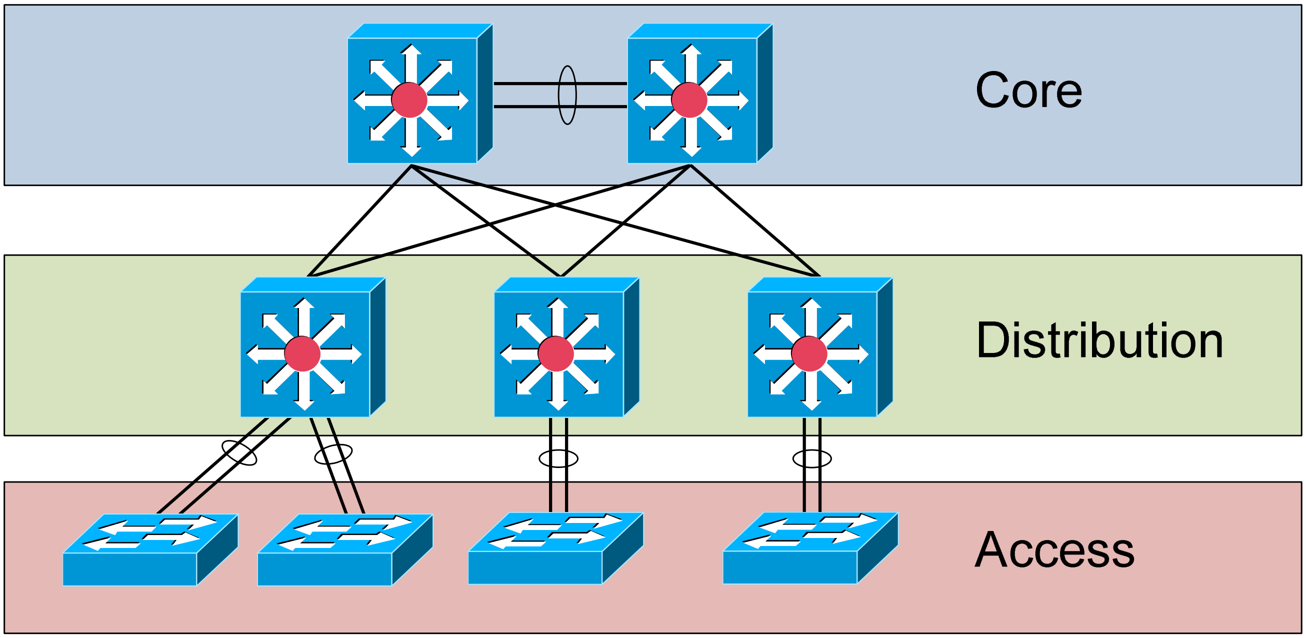
This remains a common design today.

However, **pushing Layer-3 to the access-layer** has become increasingly prevalent. VLAN SVIs are configured on the *access layer* switch, which hosts will use as their default gateway.

A routed connection is then used between access and distribution layers, further minimizing STP convergence issues and limiting broadcast traffic.

(Reference: CCNP Switch 642-813 Official Certification Guide by David Hucaby. Cisco Press)

<u>Cisco Hierarchical Model – Core Layer</u>



The **core** layer is responsible for connecting all distribution layer switches. The core is often referred to as the **network backbone**, as it forwards traffic from to every end of the network.

Switches at the core layer typically have the following characteristics:

- High-throughput Layer-3 or multilayer forwarding
- Absence of traffic filtering, to limit latency
- Scalable, redundant links to the distribution layer and other core switches
- Advanced QoS functions

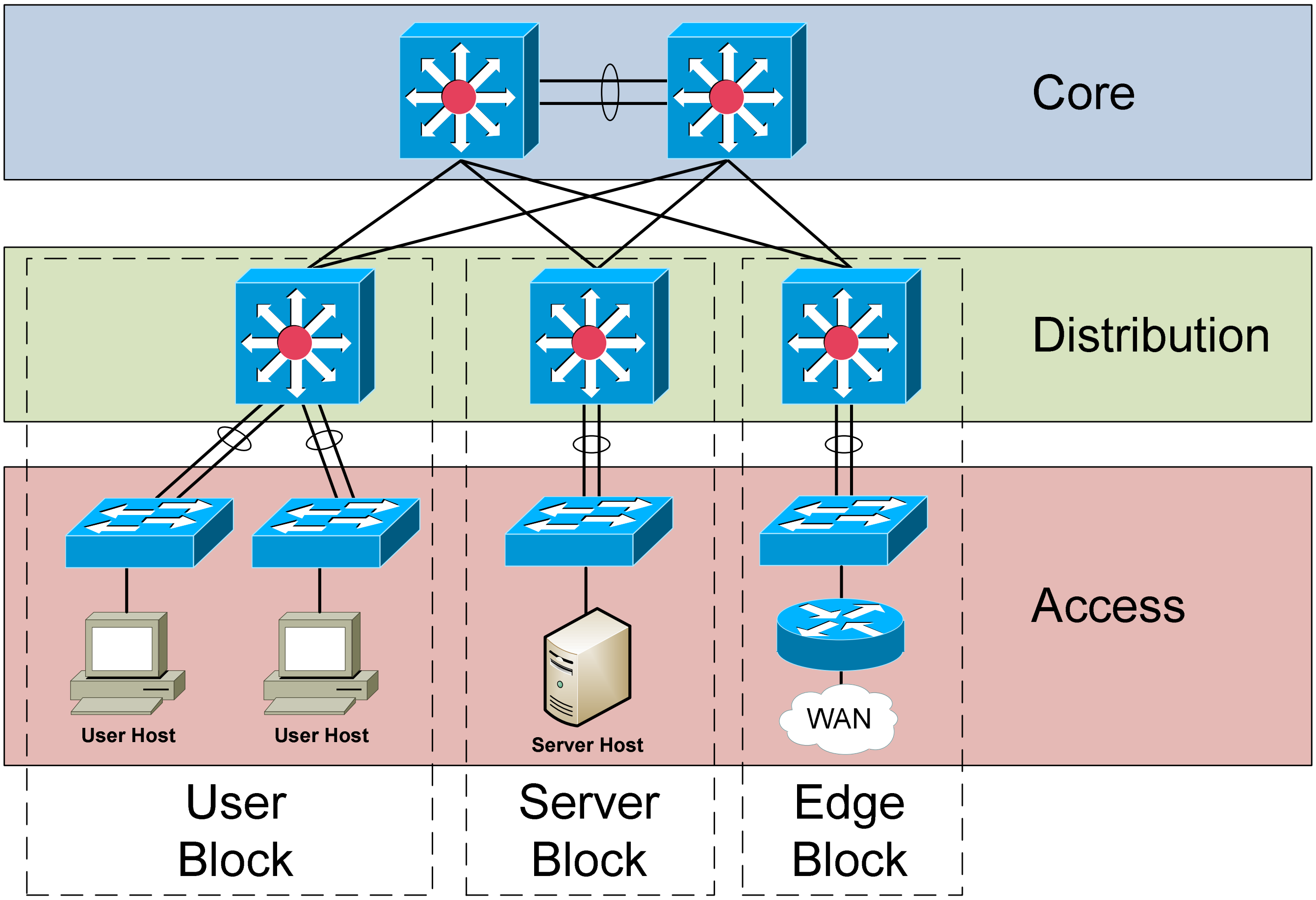
Proper core layer design is focused on **speed and efficiency**. In a 20/80 design, most traffic will traverse the core layer. Thus, core switches are often the highest-capacity switches in the campus environment.

Smaller campus environments may not require a clearly defined core layer separated from the distribution layer. Often, the functions of the core and distribution layers are combined into a single layer. This is referred to as a **collapsed core** design.

(Reference: CCNP Switch 642-813 Official Certification Guide by David Hucaby. Cisco Press)

Cisco Hierarchical Model – Practical Application

A hierarchical approach to network design enforces scalability and manageability. Within this framework, the network can be compartmentalized into modular **blocks**, based on function.



The above example illustrates common block types:

- **User block** – containing end users
- **Server block** – containing the resources accessed by users
- **Edge block** – containing the routers and firewalls that connect users to the WAN or Internet

Each block connects to each other through the core layer, which is often referred to as the **core block.** Connections from one layer to another should always be redundant.

A large campus environment may contain *multiple* user, server, or edge blocks. Limiting bottlenecks and broadcasts are key considerations when determining the size of a block.