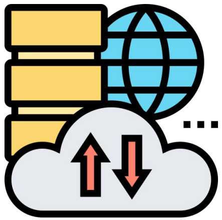
COEN 241
Introduction to Cloud Computing

Lecture 11 - Networking 101 & DC Networks



Lecture 10 Recap

- OpenFaaS Demo
- Sustainable Data Centers
- Midterm Review

Agenda for Today

- Networking 101
- Data Center Networking
- Switch Architecture
- Prepare for Software Defined Networking
- Readings
  - Recommended: None
  - Optional: https://engineering.fb.com/2019/03/14/data-center-engineering/f16-minipack/

Computer Networking 101

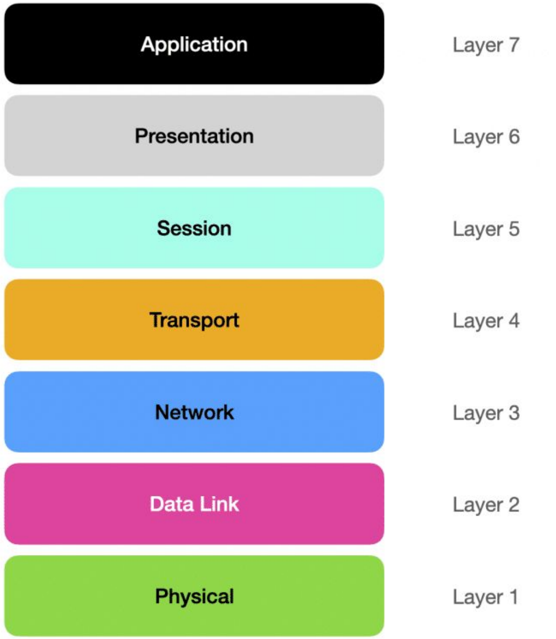
OSI Model : Building Block of Networking

- Open Systems Interconnection provides a standard for different computer systems to be able to communicate with each other
- Can be seen as a universal language for computer networking
- It’s based on the concept of splitting up a communication system into **seven abstract layers**, each one stacked upon the last
- Each layer has different **protocols** implementations
  - **Protocols:** An established set of rules that determine how data is transmitted between different devices in the same network
  - Example: HTTP Status Codes: 200, 401, 500
HTTP Commands: GET, POST

OSI Model : Building Block of Networking

OSI Model : Building Block of Networking

- Application Layer (Layer 7)
  - Only layer that directly interacts with data from the user
  - Software applications like web browsers and email clients rely on the application layer to initiate communications
  - Example Protocols:
    - HTTP
    - SMTP
    - FTP
    - DNS

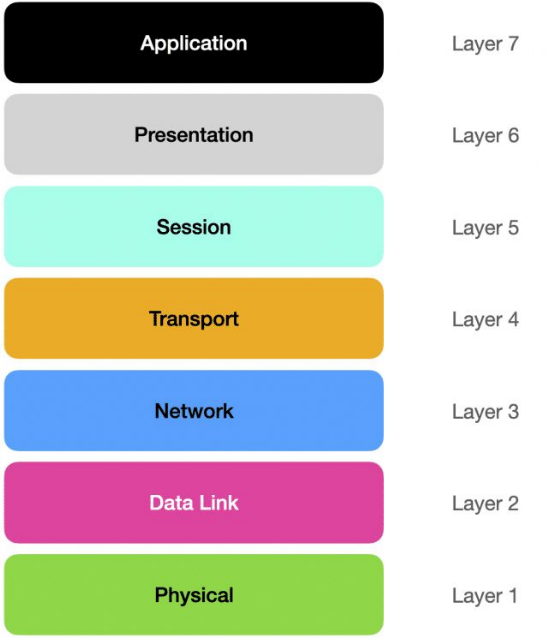


OSI Model : Building Block of Networking

- Presentation Layer (Layer 6)
  - Primarily responsible for preparing data so that it can be used by the application layer
  - Used for translation, encryption, and compression of data
  - Example Protocols:
    - SSL
    - TLS

OSI Model : Building Block of Networking

- Session Layer (Layer 5)
  - Responsible for opening and closing communication between the two devices.
  - Session: Time between when the communication is opened and closed
  - Example Protocols:
    - RPC
    - SMB



OSI Model : Building Block of Networking

- Transport Layer (Layer 4)
  - Responsible for end-to-end communication between the two devices
  - Takes data from the session layer and breaking it up into chunks called **segments** before sending it to layer 3
  - Example Protocols:
    - TCP
    - UDP

OSI Model : Building Block of Networking

- Network Layer (Layer 3)
  - Responsible for facilitating data transfer between two different networks
  - Breaks up **segments** from the transport layer into smaller units, called **packets**, on the sender's device, and reassembling these packets on the receiving device
  - Example Protocols:
    - **IPV4**
    - **IPV6**
    - ECN
    - **ICMP (ping)**

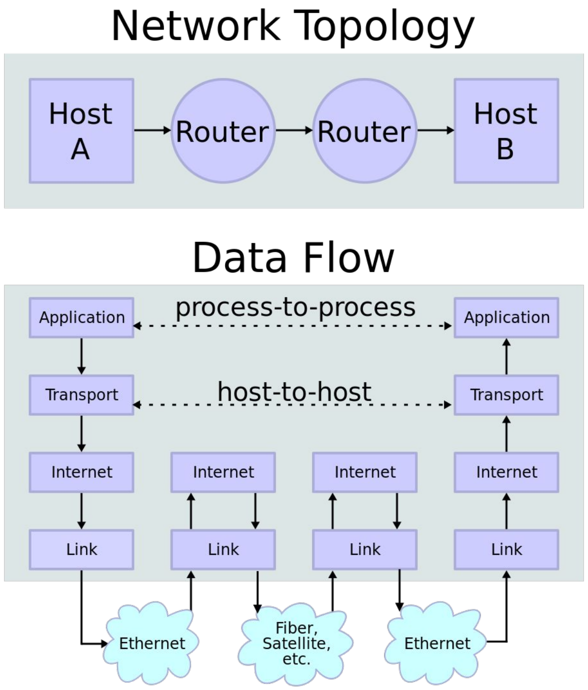
OSI Model : Building Block of Networking

- Data Link Layer (Layer 2)
  - Facilitates data transfer between two devices on the SAME network
  - Example Protocols:
    - ARP
    - MAC
    - OSPF

OSI Model : Building Block of Networking

- Physical Layer (Layer 1)
  - Includes the physical equipment involved in the data transfer, such as the cables and switches
  - Example Protocols:
    - Bluetooth
    - IEEE.802.11
    - IEEE.802.3

IP Stack Connections

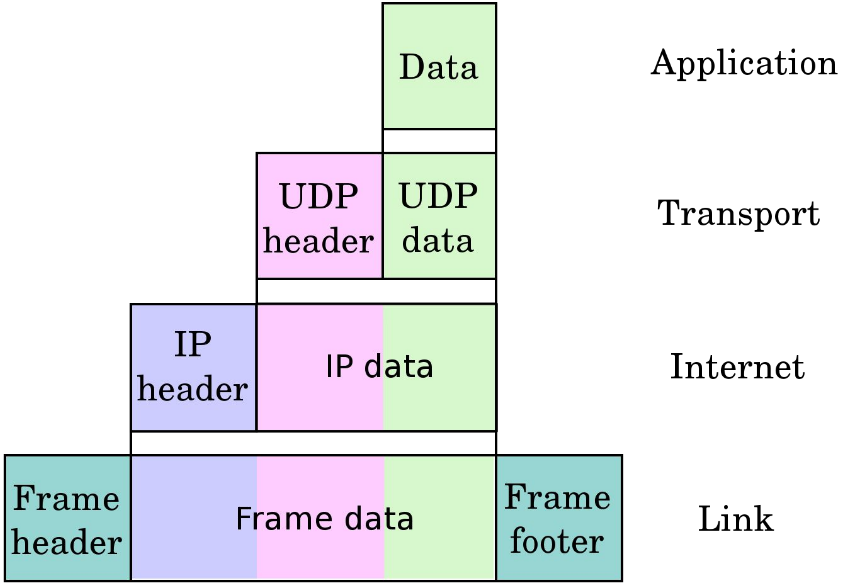


OSI Model & IP Networking

- Success of IP Networking attributed to OSI Model
  - Different layers can change technologies independently
    - e.g., Wi-Fi versus cable at link level
- Different devices handle different layers
  - Switches usually work at data link layer
  - Routers work at IP / transport layer
- Layering is not strictly enforced
  - E.g., cross-layer optimization, VPNs

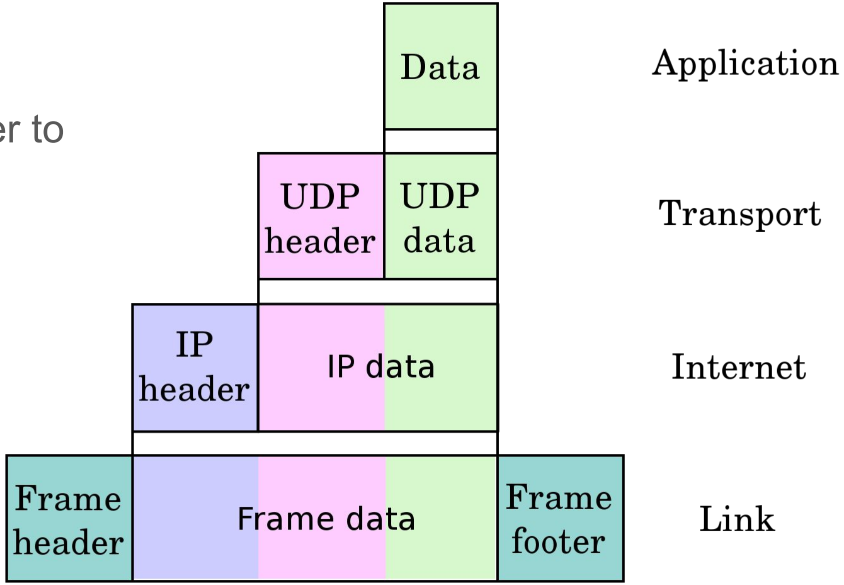
Data Carried at Different Layers

- Data for applications using IP is augmented at each layer below
  - Ethernet frames' total size larger than IP packets' total size for a given data
- UDP transports datagrams (chunks of data with no reliability)
- TCP transports streams of data
  - With retransmissions and congestion control
- UDP and TCP requires port used by the application

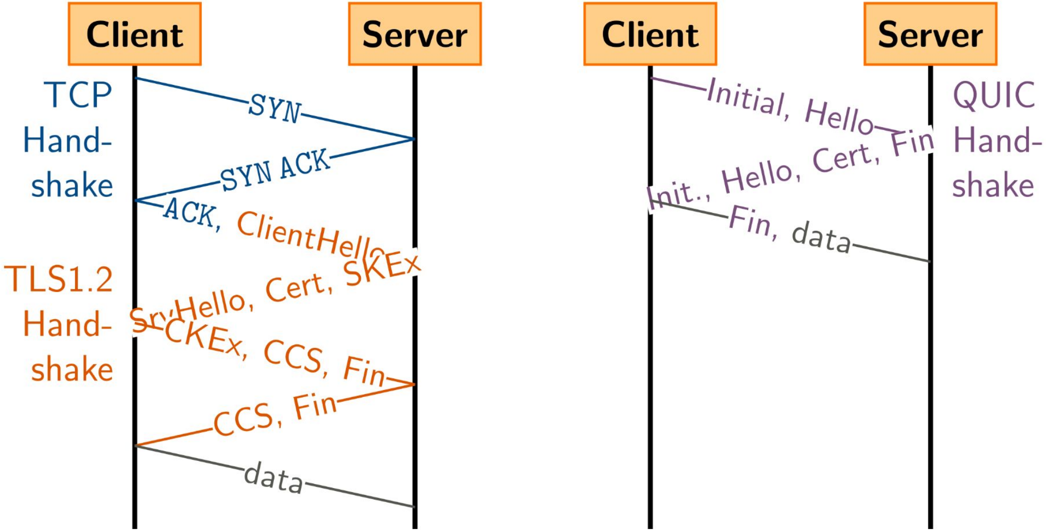


Data Encapsulated in Layers

- Often times people customize UDP or TCP headers for their own protocol
- Example
  - Google QUIC protocol uses UDP header to encapsulate QUIC headers
  - QUIC allows faster transport
  - Used in Chrome since 2012 FireFox in 2021
  - https://en.wikipedia.org/wiki/QUIC
- Header is usable, but body is usually encrypted



QUIC


Client
Server
TCP Hand-shake
SYN
SYN ACK
ACK, ClientHello
TLS1.2 Hand-shake
SrvHello, Cert, SKEx
CKEx, CCS, Fin
CCS, Fin
data
Client
Server
Initial, Hello, Fin
QUIC Hand-shake
Init., Hello, Cert, Fin
Fin, data

Ethernet Switch Hardware (Layer 2)

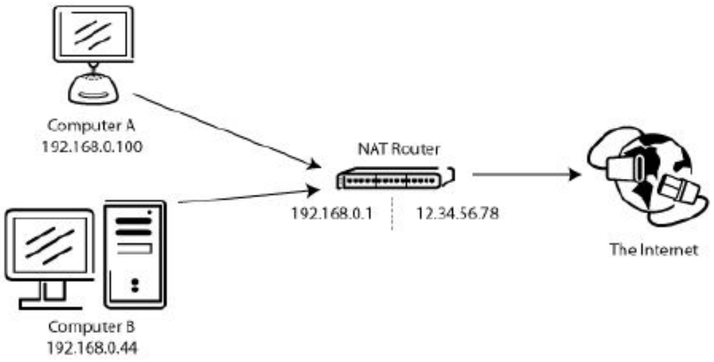
- Hosts connected to ports of a switch
- Switch examines the MAC addresses in Ethernet frames
  - Ternary content addressable memory (TCAM) used to quickly find MAC
  - Determines which switch port(s) to send Ethernet frame to
- Switch backplane has higher bandwidth than ports
  - Needs to allow pairs of ports to communicate in parallel
  - Uplink ports often higher speed than normal ports (why?)
- Switches run software for tasks like firmware upgrades, etc.
  - Done in the control plane or the microserver
  - Also, **virtual switches can be run by hypervisors**, e.g., VirtualBox

TCPDump

- How can we see each layer in practice?
- TCPDump or Wireshark
  - ICMP Echo request
    - tcpdump --interface eth0 “icmp[0] == 8” -XX -s 100 -n
  - SSH
    - tcpdump --interface eth0 tcp port 22
- Demo

Network Address Translation

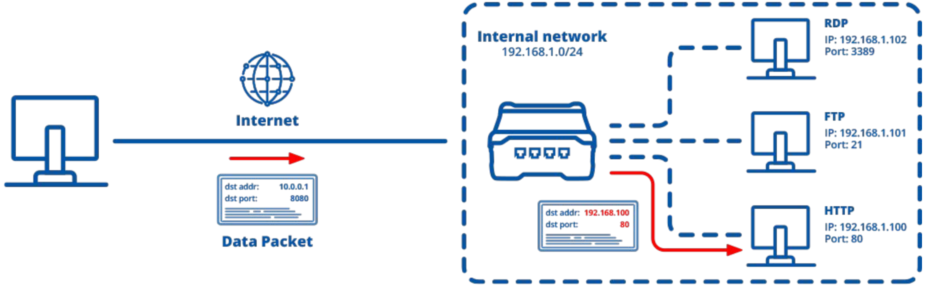
- Mapping an IP address to another by modifying the header of IP packets while in transit across a router
- Useful for
  - Saving addresses
  - Speed
  - Flexibility
  - VMs!



Port Forwarding

- Port forwarding or port mapping is an application of NAT that redirects a communication request from one address and port number combination to another

- Keeps unwanted traffic from entering



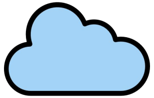
Data Center Networking

Internet Access to the Cloud

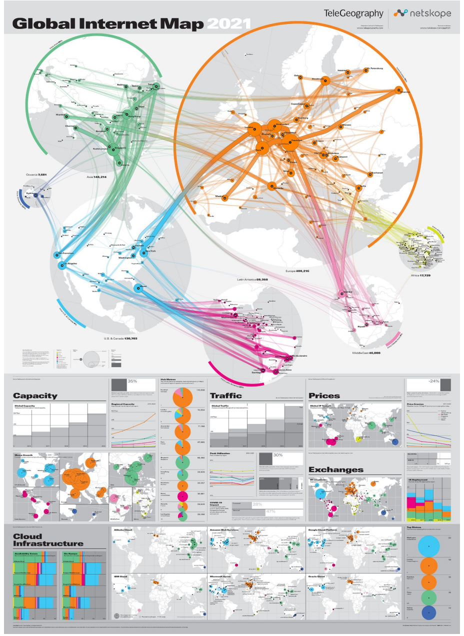
- Assume we send IP packets from a web browser to the Cloud

- First data makes its way to the edge of your Internet Service Provider:
  - Wi-Fi in your laptop to layer-2 wireless access points (APs)
  - Transition from layer-2 (MAC) to layer-3 (IP) occurs in the APs
  - Internet router aggregates traffic and sends it off to your ISP
  - Your ISP gathers traffic at the ISP data centers
  - Then traffic leaves the ISP network

- The data travels across the internet to reach a data center



Internet Access to the Cloud



https://global-internet-map-2021.telegeography.com/

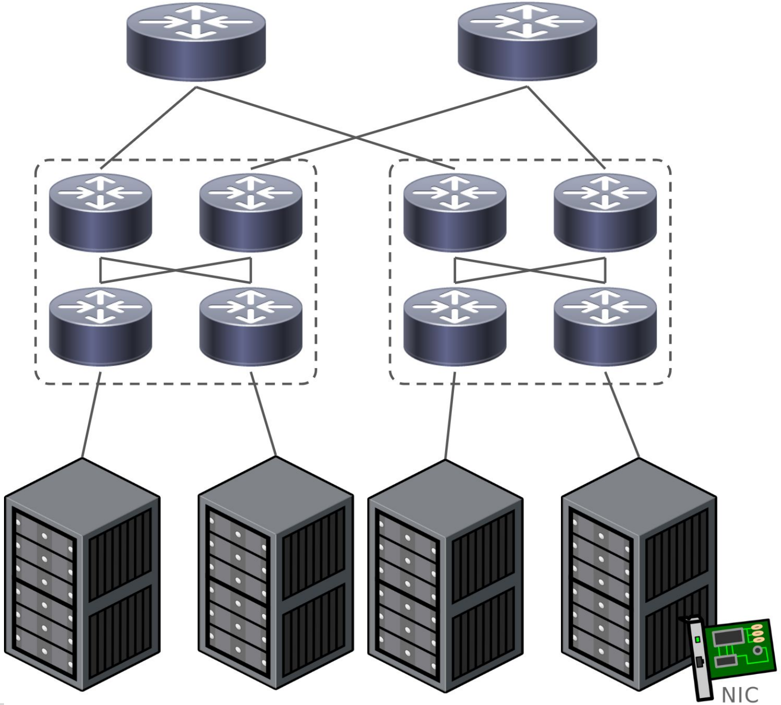
Type of Traffic and Workloads in a Data Center

- Traffic **to/from Internet**
  - The node-to-node internal network use is small
- Consider scale-out applications within a DC
  - Interacting servers do so with distributed effect & high volume
  - Application dependent
    - MapReduce, Storage Replication
- New ‘interesting’ traffic patterns: (i.e., non-unicast)
  - Anycast: traffic reaches any ‘nearby’ applicable host
  - Multicast or 1–N: traffic is being sent to multiple hosts simultaneously

Data Center Networks

- Once users’ packet reaches the data center, it enters the DC network
- Data centers have a structured physical layout
  - Physical servers are grouped into racks, which usually provide:
    - Intra-rack network switches; power distribution; wiring management
  - Racks grouped into aisles or zones—for maintenance / access
- DC networking is hierarchically organized:
  - Core router receiving data from the internet
  - Aggregation layer receiving data from the core
  - Top-of-rack switches installed within racks
  - **Virtual networking within physical servers themselves**

Data Center Networks



Core
Switch

Aggregation
Switch

Edge/ToR
Switch

Servers

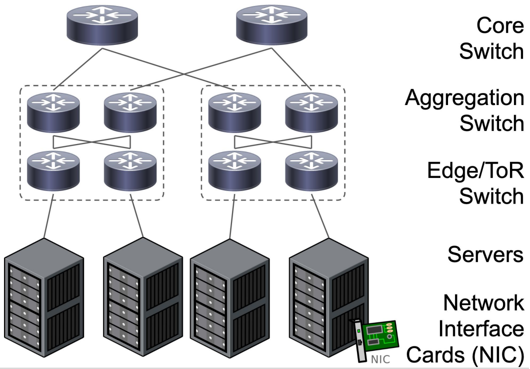
Network
Interface
Cards (NIC)

Data Center Network Topology

- DC Networks need to be connected in some topology
  - Topology: Arrangement of switches
  - Wiring up complete graph isn't practical O(n^2)
- Must provide a topology that's cheap but also effective
  - Must consider contention and congestion across some wires
- The topology must also consider the traffic type
  - Cloud workloads are likely to be more bursty
  - Cloud workloads also depend on application; time-of-day

Data Center Network Topology (Fat-Tree)

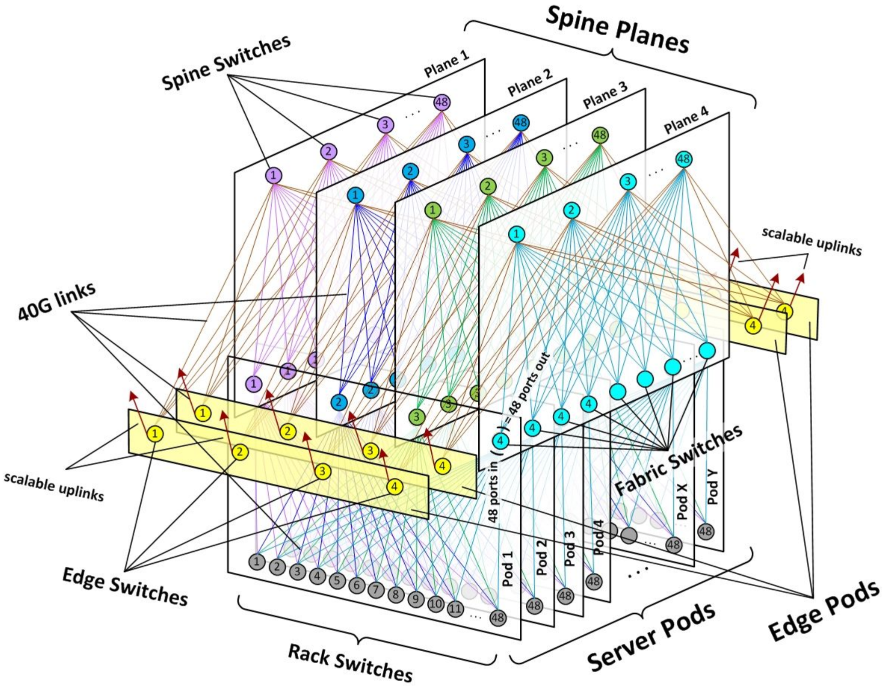
- The three-tier, fat-tree topology mentioned is now outdated
- Multiple issues with the three-tier design
  - Upper-level routers: highly specialized and expensive
    - Designed to sustain high bandwidth in all directions
    - Core routers' prices in the order of $100,000 a piece
  - High energy usage
  - Not well suited for intra-DC traffic (why?)



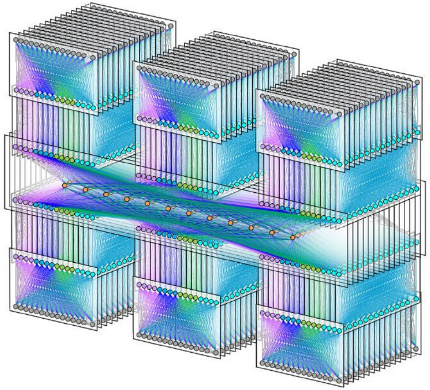
Data Center Network Topology (Clos)

- Modern trend is toward using commodity switching kit
  - Create an aggregation switch from a set of cheaper switches
  - Employs a particular addressing scheme and routing algorithm
  - To build and configure by the DC operators
- Often related to a topology known as a Clos network
  - Have ingress, middle and egress stages built from switches
  - The structure can be used recursively (expand middle)
  - Simpler and flatter

Data Center Network Topology (Clos)



Data Center Network Topology (Clos)



https://engineering.fb.com/2019/03/14/data-center-engineering/f16-minipack/

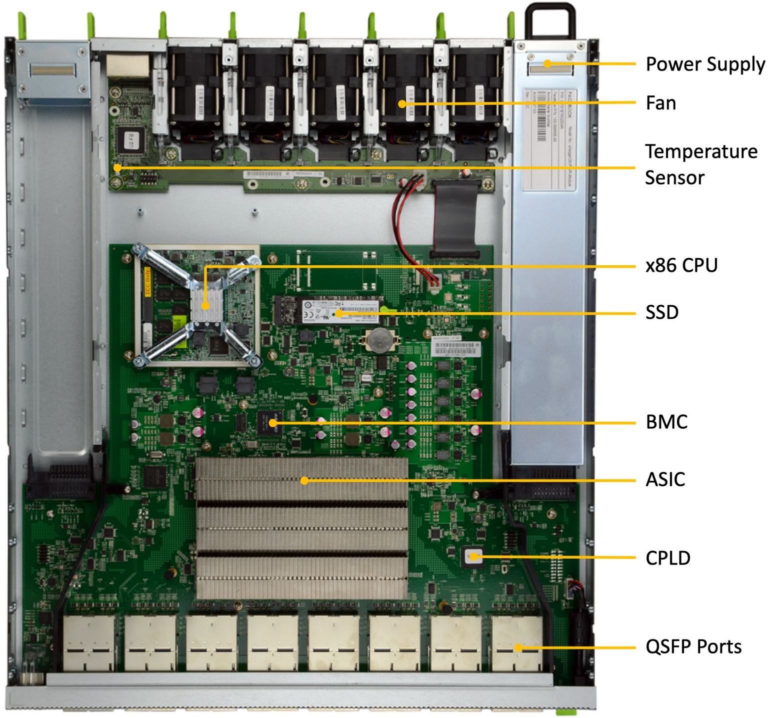
Open Compute Project

- Initiated by Facebook to incorporate commodity parts in custom, open designs for building a data center
  - Google had previously pioneered using commodity kit for server
  - Covers: DC facility; racks; power; networking; servers; storage; …
- OCP networking scope includes:
  - Disaggregated and open network hardware and software
  - Automated configuration management and provisioning
  - Switch motherboard hardware and form-factor mapping
  - Software Defined Networking (SDN)

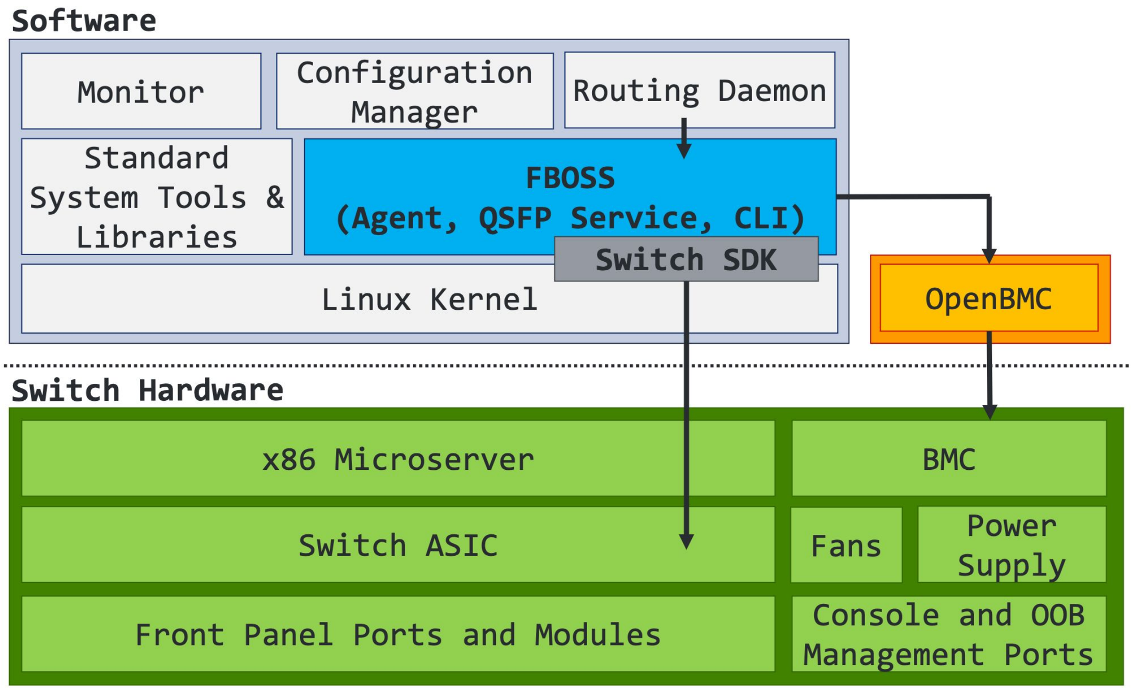
Data Center Switch

- Data Center Switch (Router) consists of multiple components
- **Data Plane**
  - Hardware that is responsible for actual packet switching
  - Switch ASIC
- **Control Plane**
  - Hardware that is responsible for making routing decisions
  - Usually a Linux software running on x86 CPU
- Quad Small Form factor Pluggable (QSFP) Ports
  - Allows 100+Gbps speed
- Miscellaneous Hardware
  - Baseboard Management Controller (BMC)

Data Center Switch Architecture



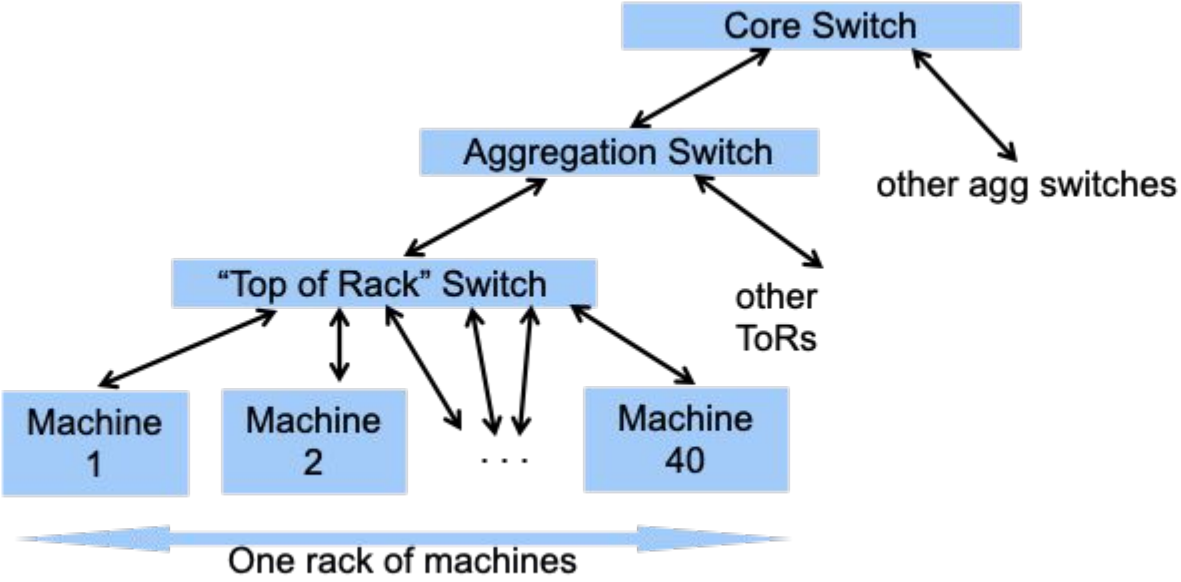
Data Center Switch Architecture



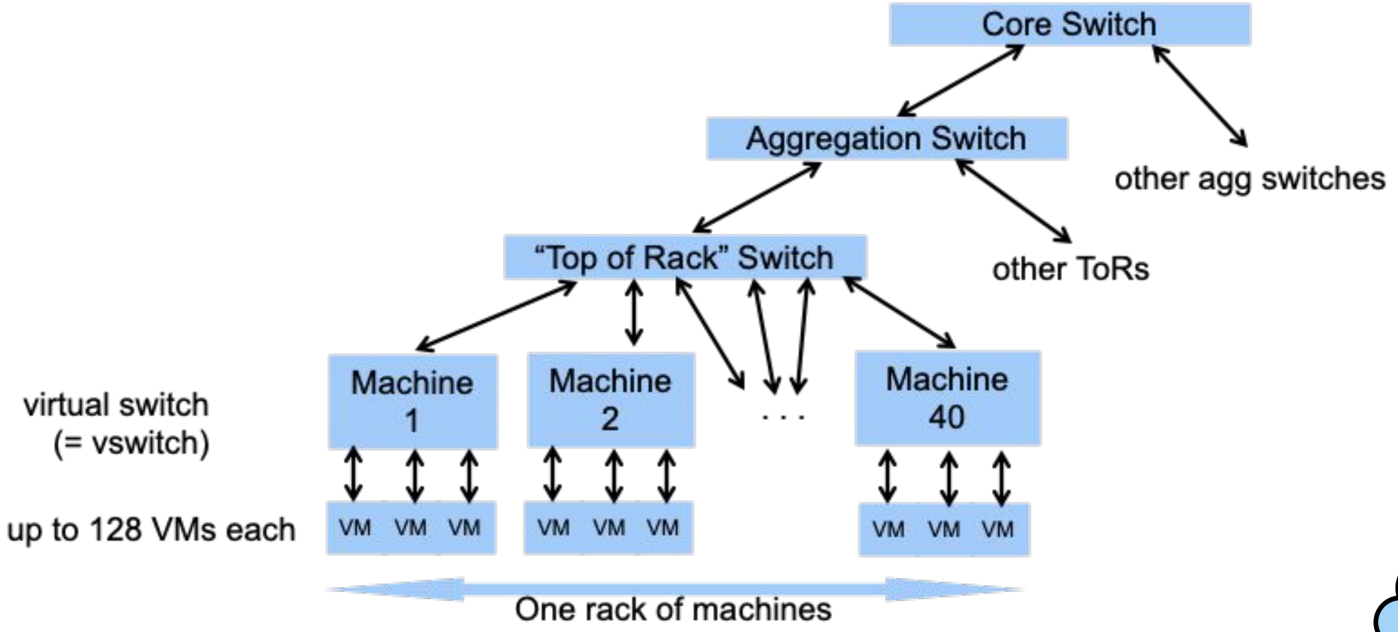
Virtual Switch

- Switch within a server that switches packets between VMS
  - Another big part of Data Center networking and **Software-Defined Networking**
- **Data Plane**
  - Software that is responsible for actual packet switching
  - Implemented in Intel DPDK and/or in Kernel modules
- **Control Plane**
  - Software that is responsible for making routing decisions
- Examples:
  - Open vSwitch (OvS)

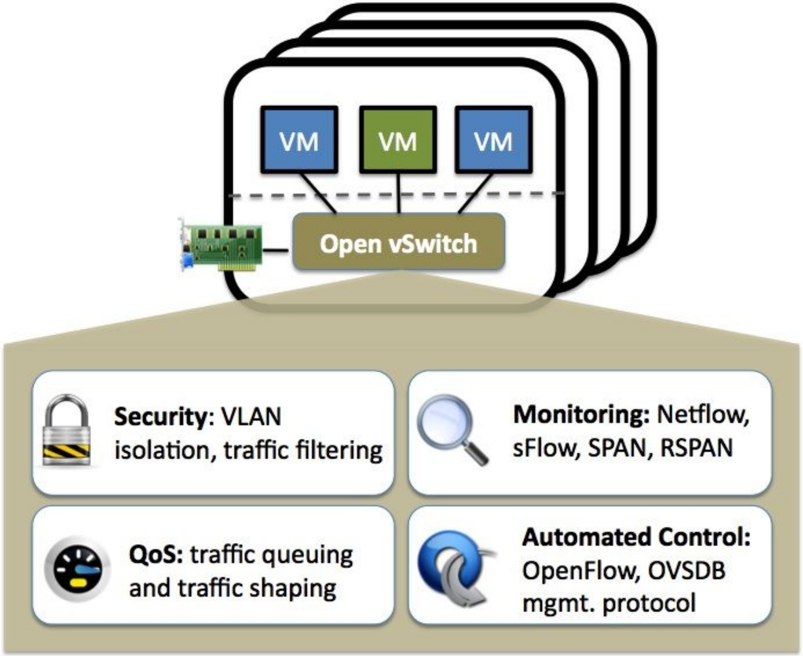
Virtual Switch: DC Network Before VMs



Virtual Switch: DC Network After VMs



Open vSwitch Features



Programmable Data Plane

- Latest advancement in networking

- **Data Plane**
  - Hardware that is responsible for actual packet switching
  - Switch ASIC
  - Cannot be programmed like a CPU

- Programmable switches have programmable ASICs
  - Can run custom programs on the ASIC
  - Similar to how we run our own programs on a CPU
  - Capable of understanding custom packet header types
  - Lot of exciting projects in this space

Fixed Set of Features

TCP

Ethernet

IPv4

UDP

HTTP

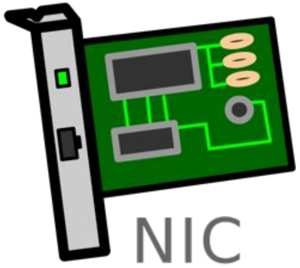
IPv6

BGP

TLS

Proprietary Software

Fixed-Function Hardware



What does it mean to be Programmable?

Behavior of the device is defined by **software** that operates **independently** from the **hardware**

vs.

Programmable Control Plane

- Control plane is the brain of the network
- Often times it must communicate with other software or hardware
    - Must expose an API
- OpenFlow!
    - Will come back to this next lecture!

Midterm Feedback

- Please finish a poll in the following link
  - https://forms.gle/WsperRbhkshFuXAY7

TODOs!

- Project!
- HW 3 will be available next class

Agenda for Today

- Networking 101
- Data Center Networking
- Switch Architecture
- Prepare for Software Defined Networking
- Readings
  - Recommended: None
  - Optional: https://engineering.fb.com/2019/03/14/data-center-engineering/f16-minipack/

Questions?