

Building Massive Cloud Networks





Image from Microsoft Azure

Microsoft and Facebook just laid a 160-terabits-per-second cable 4,100 miles across the Atlantic 47

Enough bandwidth to stream 71 million HD videos at the same time

By [Thuy Ong](#) | [@ThuyOng](#) | Sep 25, 2017, 7:56am EDT

<https://www.nytimes.com/interactive/2019/03/10/technology/internet-cables-oceans.html>

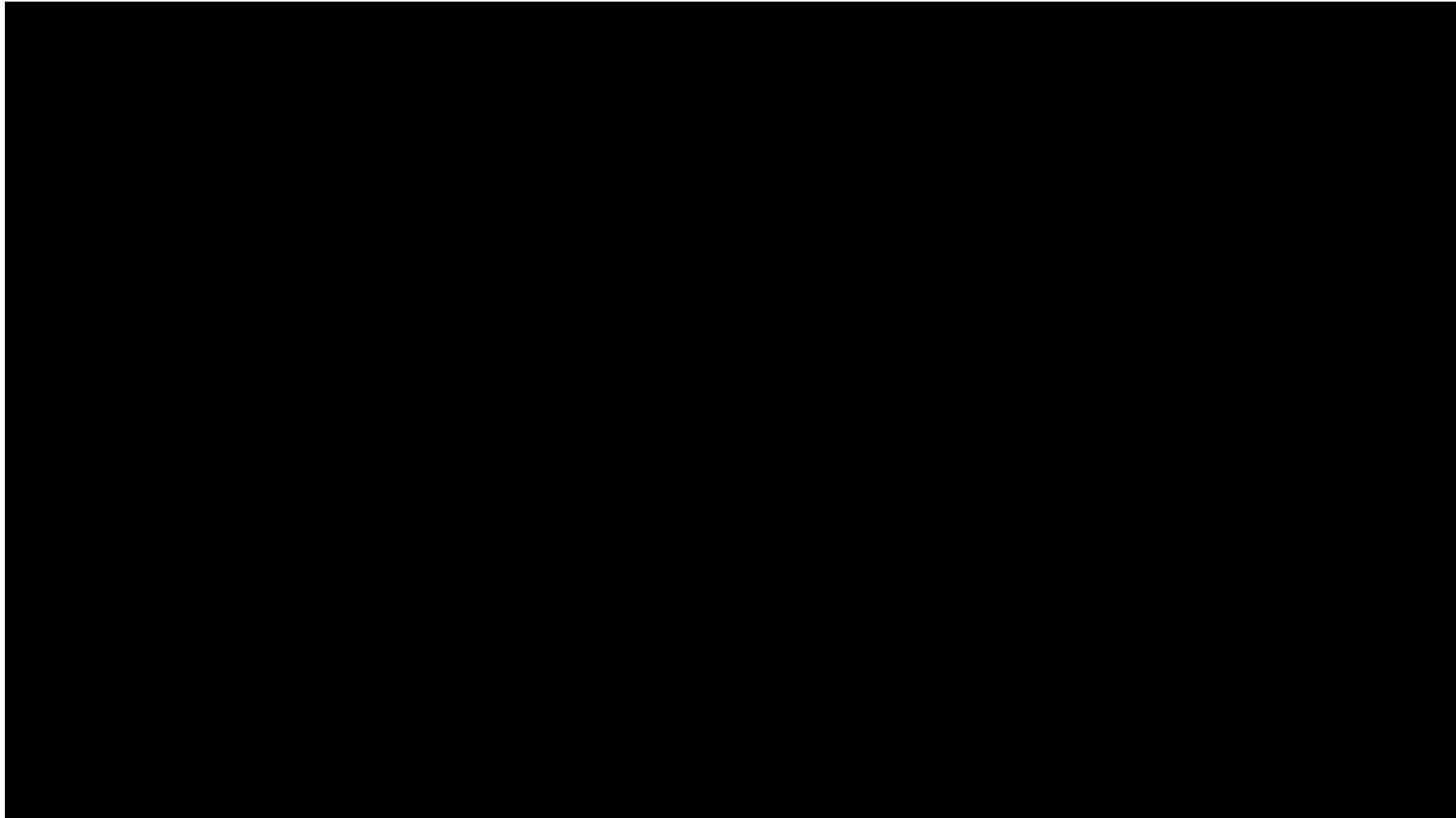
HUGE data center networks (DCN)

- Thousands of routers
- Hundreds of thousands of servers

Google's Oregon DC

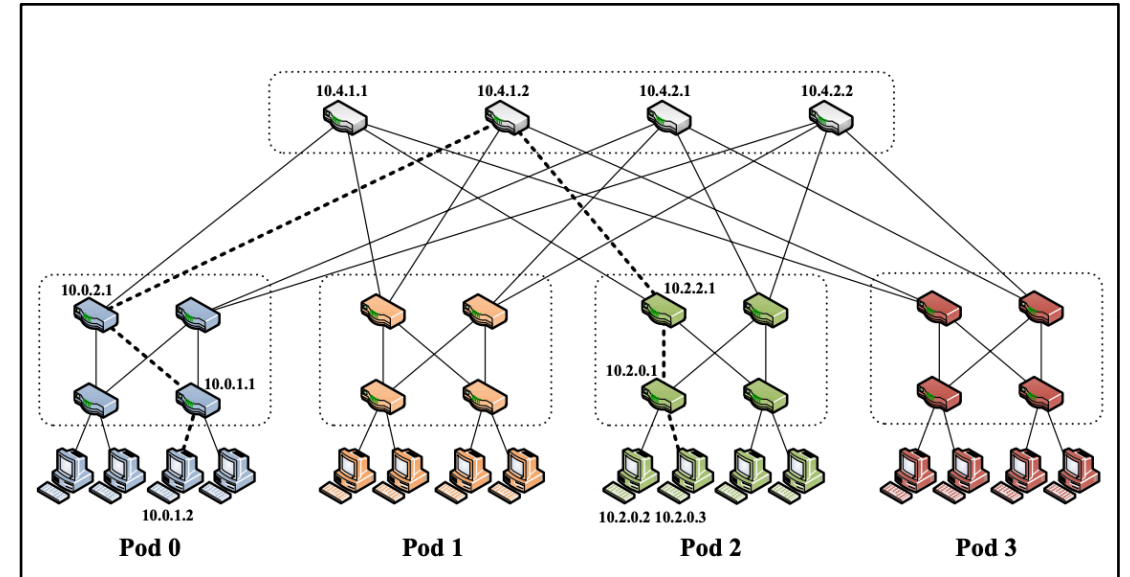
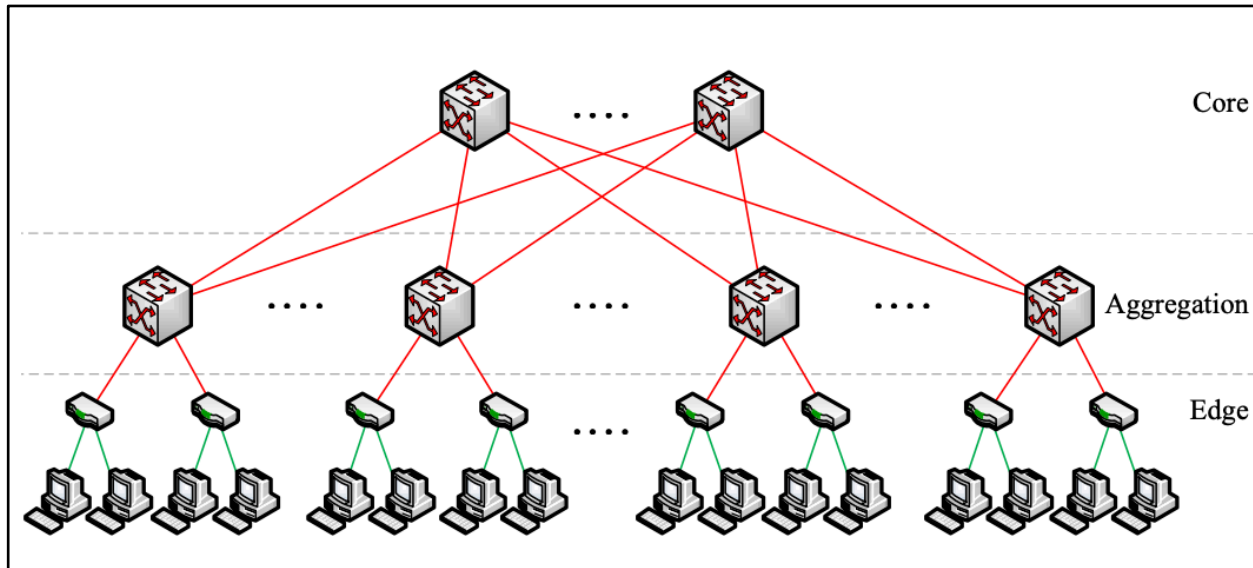


Inside a Google DC



DCN topologies

- Big iron → Commodity switches

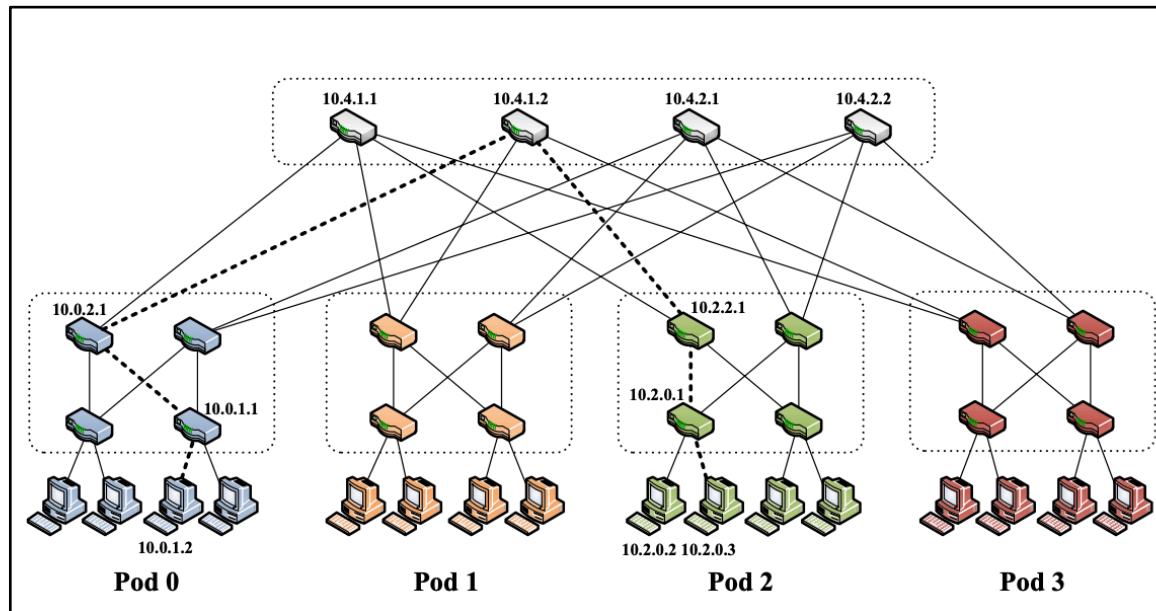


DCN topologies

- Big iron → Commodity switches
- 1 Gbps → 10 Gbps → 40 Gbps → 100 Gbps (soon)
- Copper → Fiber

Oversubscription ratio

- Ratio of bisection bandwidth across layers of hierarchy
- Key design parameter that trades-off cost and performance
 - Higher oversubscription = lower cost but higher chance of congestion

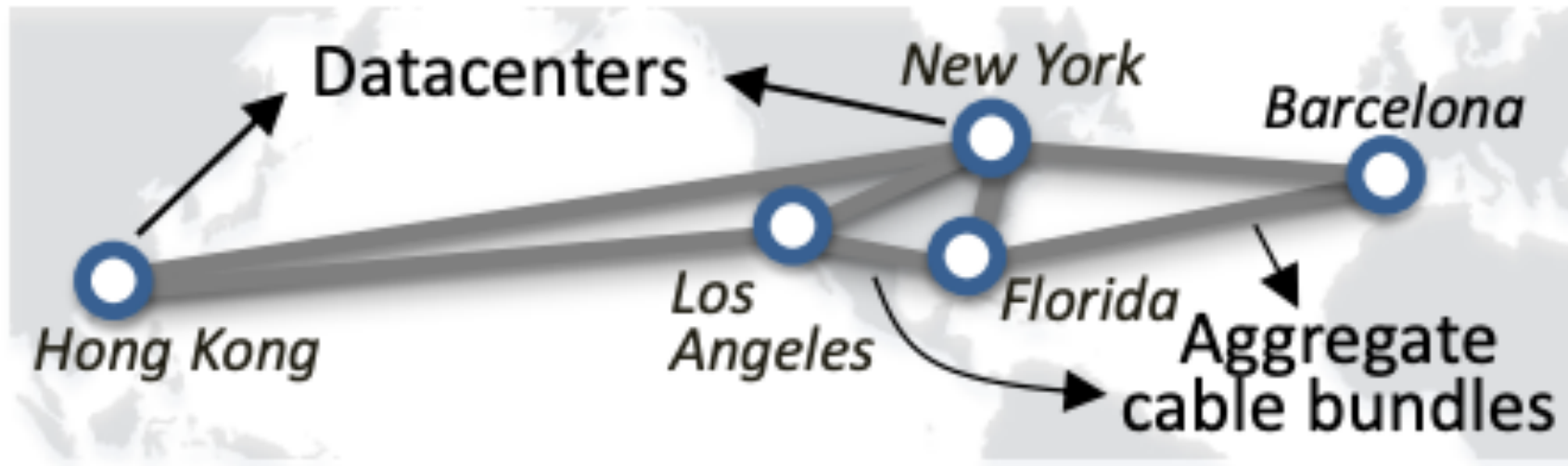


DCN routing

- Spanning tree (L2) → OSPF/ISIS → BGP
- Each routers acts as its own autonomous system (AS)

Backbone

- Provides global connectivity to DCs



Backbone

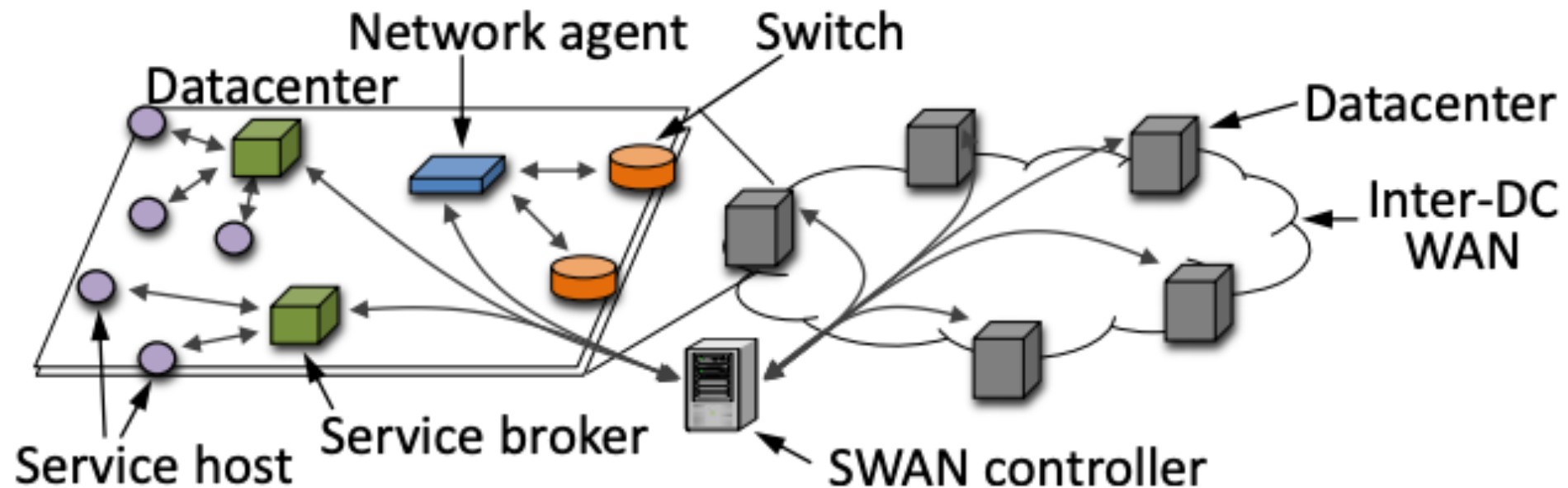
- Provides global connectivity to DCs
- May also have two backbones
 - A “public” backbone to connect to the outside world
 - A “private” backbone for inter-DC connectivity
- Uses transcontinental and transoceanic fiber cables
- Routing: ISIS/OSPF → MPLS → Centralized control (SDN)

MPLS – Multi Protocol Label Switching

- Can explicitly program paths -- tunnels
 - Allows taking non-shortest paths
- Auto-bandwidth: Constrained-shortest paths first (CSPF)
 - Fully distributed computation
 - Estimate demand
 - Find shortest path(s) that can fulfill the demand

SDN – Software Defined Networking

- Centralized computation of forwarding tables
 - Compute “optimal” paths outside of the network
 - Based on estimated load; also factor in application priorities



Using the cloud

- Use a software service (e.g., email) -- SaaS
- Use application building blocks -- PaaS
- Launch VMs – IaaS

- Build virtual networks
 - Provides the same abstraction as physical networks but with virtual devices

Connecting to the cloud

- Public Internet
- VPN from your physical resources to the cloud
- BGP peering
 - E.g., Amazon Direct Connect

The last ten years of the cloud

Scale, scale, scale ... (mostly)

Relatively small conceptual shifts

- Lot of automation – minimize “snowflakes” and “fat fingers”
- Troubleshooting: Find needles in haystack
 - E.g., Everflow [SIGCOMM '15], CorrOpt [SIGCOMM '17]
- Centralized control of resources
 - E.g., SWAN [SIGCOMM '13], Footprint [NSDI '16]
- Low-latency technologies, e.g., RDMA

Bigger shifts are coming

Verification

- E.g., Batfish [NSDI '15], Minesweeper [SIGCOMM '17]

High-level synthesis

- E.g., Propane [SIGCOMM '16, PLDI '17]

Programmable NICs and switches

New physical layers

- E.g., ProjecToR [SIGCOMM '16], RAIL [NSDI '17]

Edge computing

Tighter coupling with applications

....

