Enterprise Networking and Networking at Duke

Will Brockelsby, PhD
Chief Network Architect
Duke University and Health System
2025-04-17

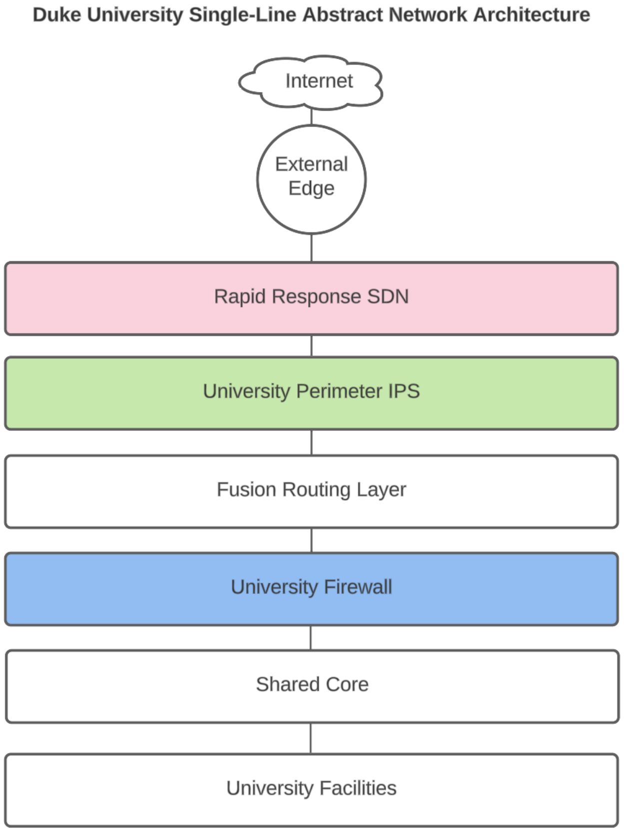
Agenda

- Review
- Enterprise Network Architectures
- Networking at Duke

Context – Duke Network

Duke University Network Strategy

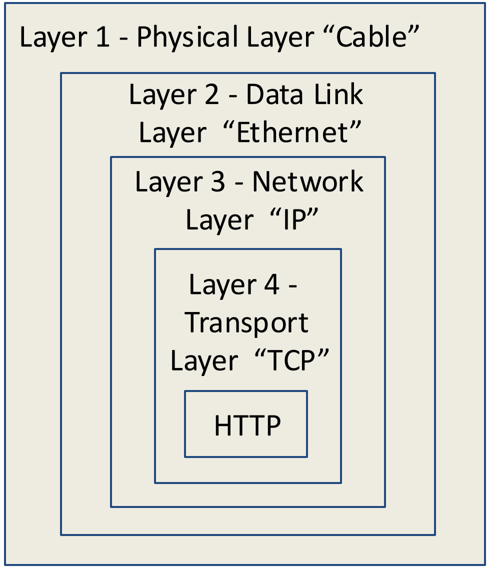
- Leverage horizontal scalability
- Virtualize physical network infrastructure to support diverse use cases
- Leverage standard protocols
- Extensively test within the lab



Network Review – Layering

- Layer 7 - Application
- Layer 6 - Presentation
- Layer 5 - Session
- Layer 4 - Transport (TCP/UDP)
- Layer 3 - Network (IP)
- Layer 2 - Data Link (Ethernet)
- Layer 1 - Physical (Fiber, Copper, RF)

- Encapsulation!

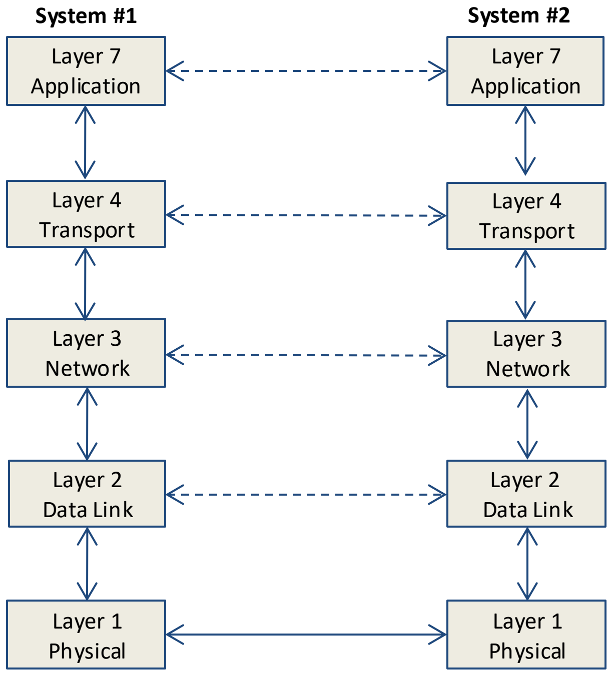


Network Review – Layering

- A layer consumes services from the layer below
- A layer provides services to the layer above
- A layer communicates with peer layers

- Systems are physically connected at layer 1

- Protocols can be swapped in/out as needed!

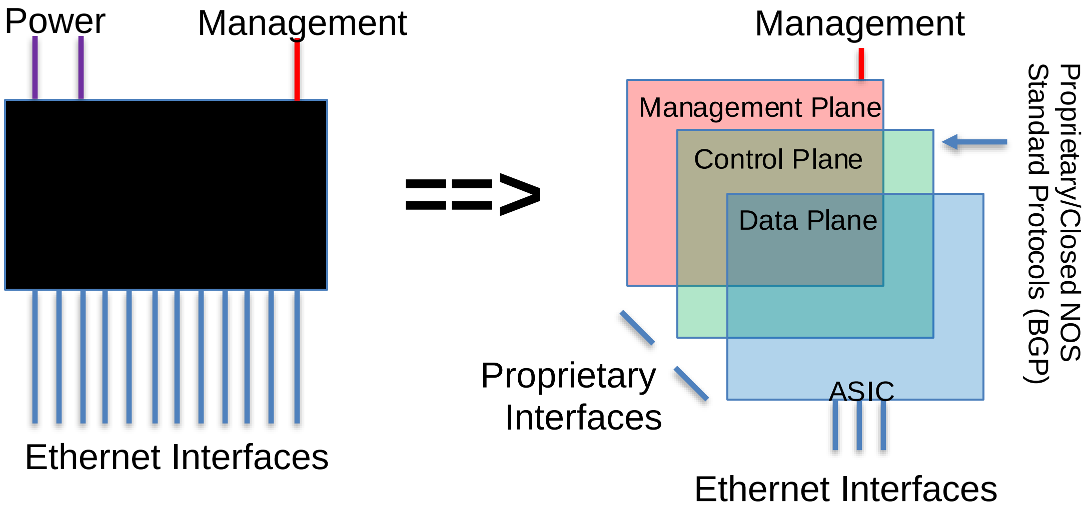


Network Review – Forwarding Devices

- Physical Box:
  Case, Fans, Power Supplies,
  Circuit Board, Interface Ports

- Control Plane
- Forwarding Plane

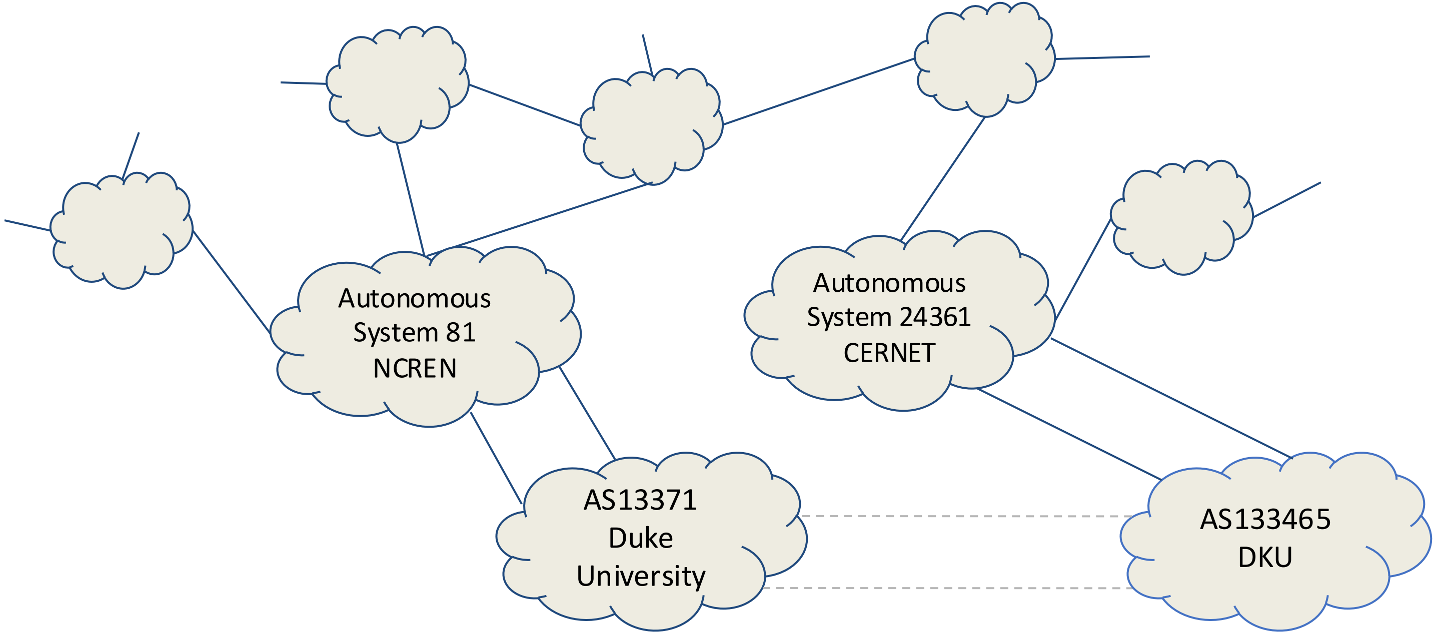
- Layer 2 devices - "switch"
- Layer 3 devices - "router"



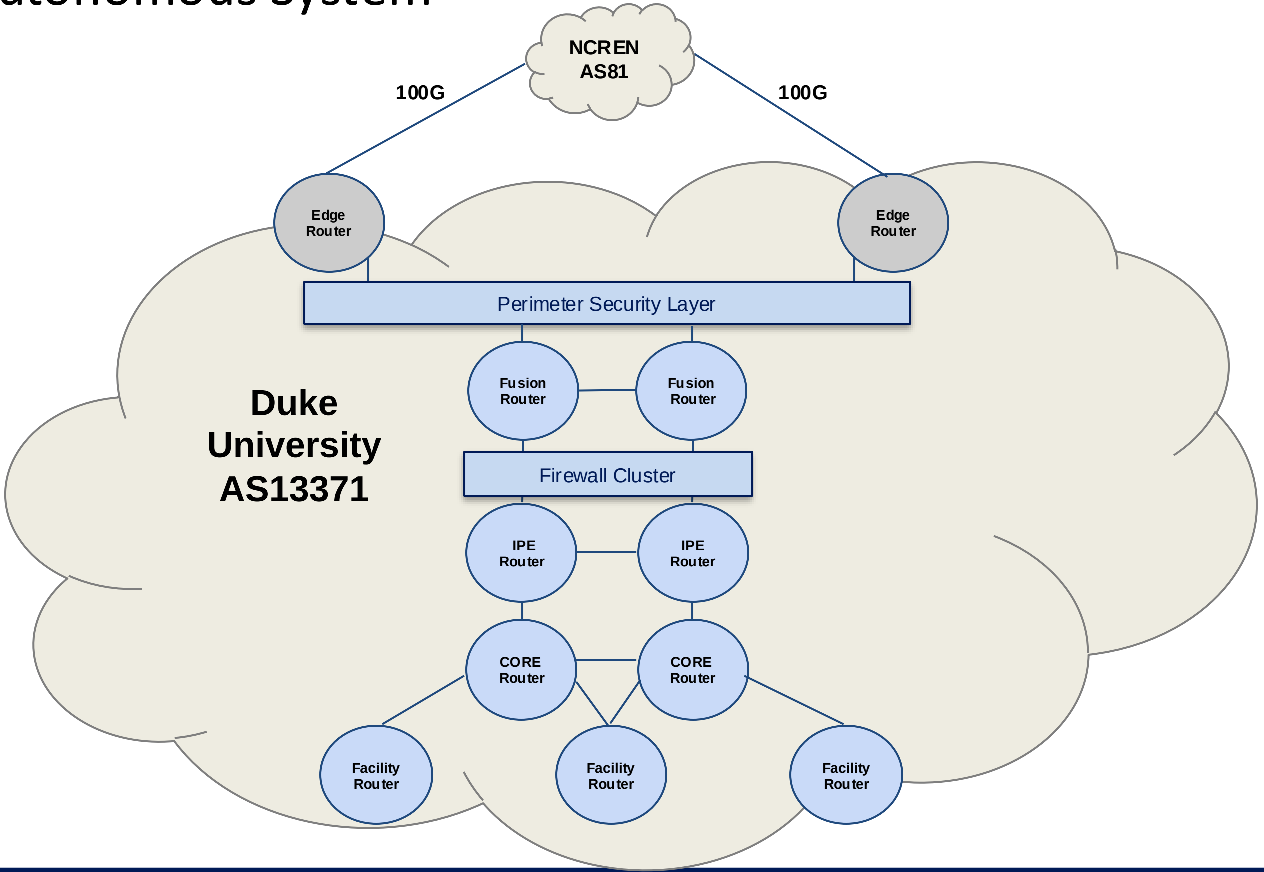
- Software Defined Networking (SDN) changes some of this...

Global Network Architecture

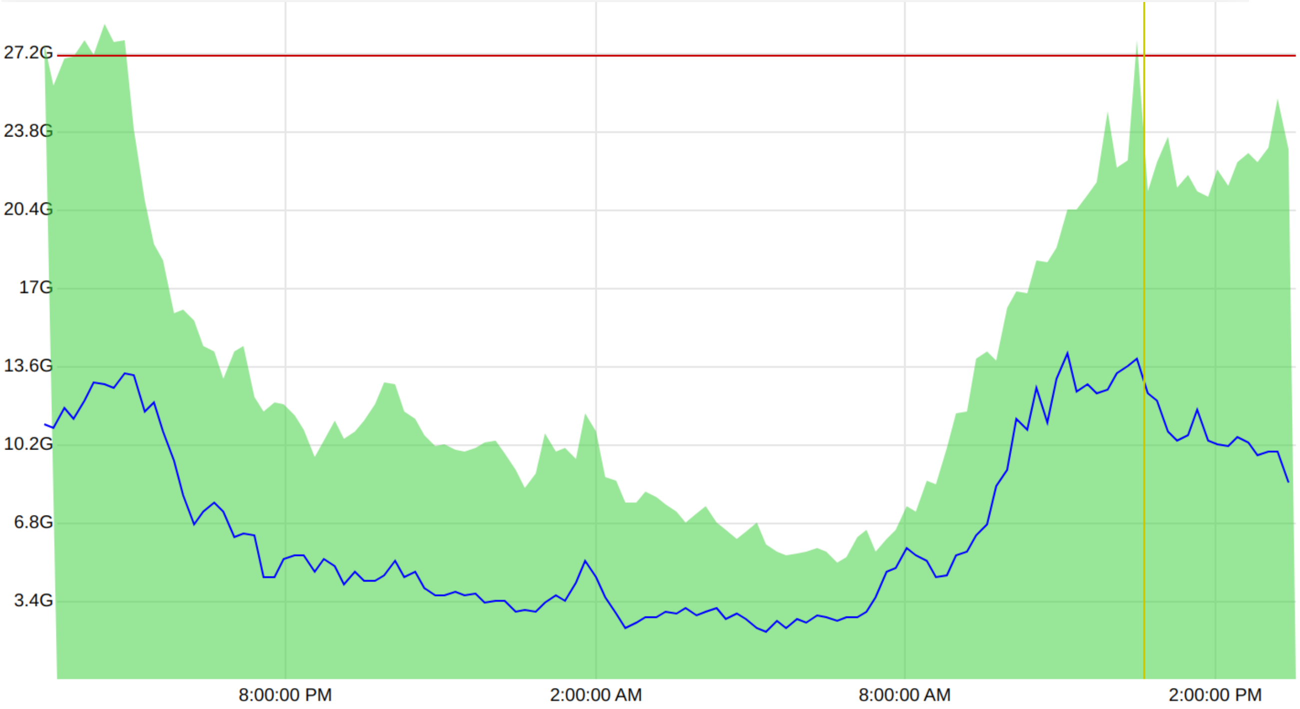
- The Internet is a network of networks.
- Independent networks are Autonomous Systems (AS)...



Inside an Autonomous System



Duke Upstream Aggregate Throughput

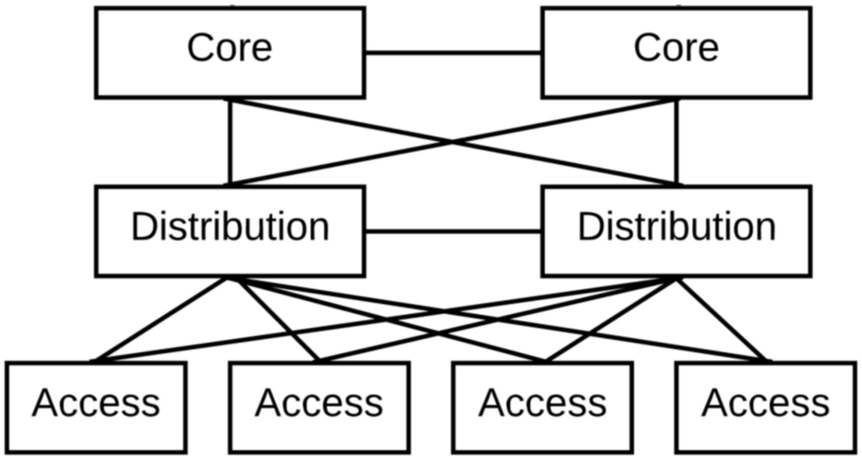


	Avg	Max	11/1/2024 11:40:00 AM	95th Percentile
Input	14.02G	28.57G	21.3G	**27.23G**
Output	6.72G	14.23G	12.44G	

Traditional Enterprise/Campus Network Architectures

Three-Tier Architecture

- Popular enterprise/campus network architecture introduced by Cisco in the late 1990s
- Permits evolutionary growth as the needs of the organization change and can provide highly available network connectivity to critical resources

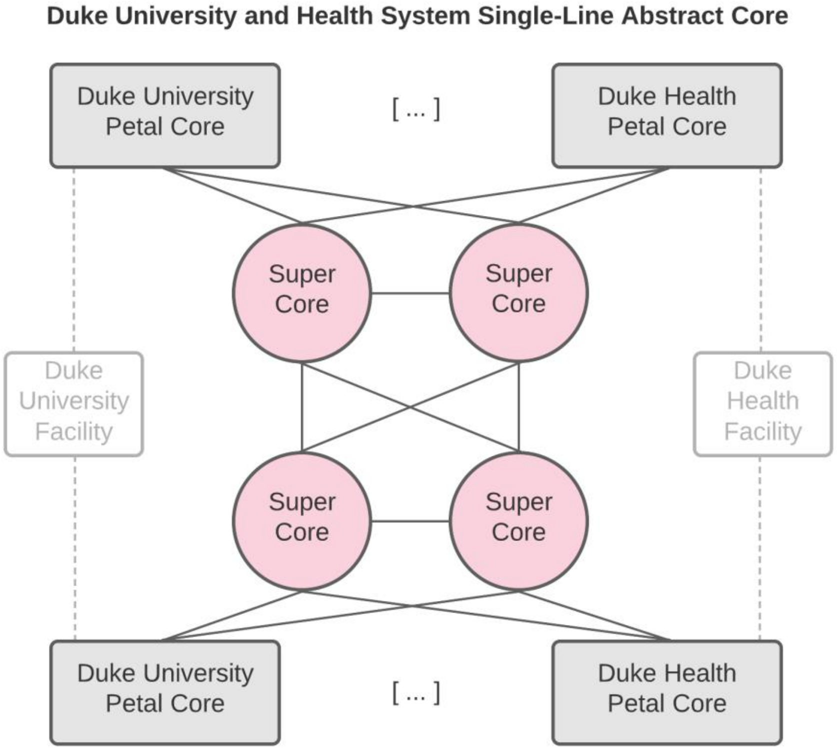


Duke Network – Core Architecture

Duke University and Health System Partnership:

- Distributed Nx100Gb/s hierarchical architecture
- Shared by university and health system
- Highly available, horizontally scalable
- Label based forwarding

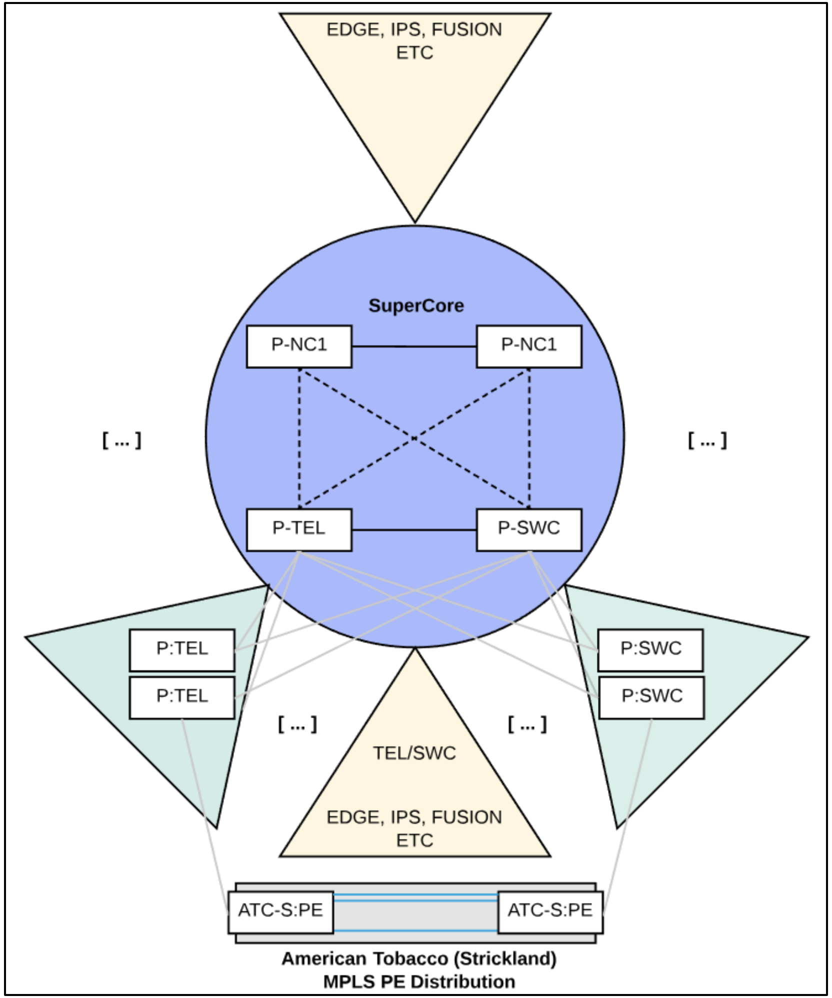
- 4x Super core label switching routers
- 6x University petal core label switching routers
- 6x Health petal core label switching routers



Duke Network – Facility Attachment

Highly Resilient Connectivity:

- Facility routers attach to geographically diverse core locations as shown
- Two options for resiliency within a given facility
  - University: 2x routers as active/active virtual router
  - Health: 2x routers as active/standby independent routers



Duke Network – Wide Area Network (WAN) - Edge Routers

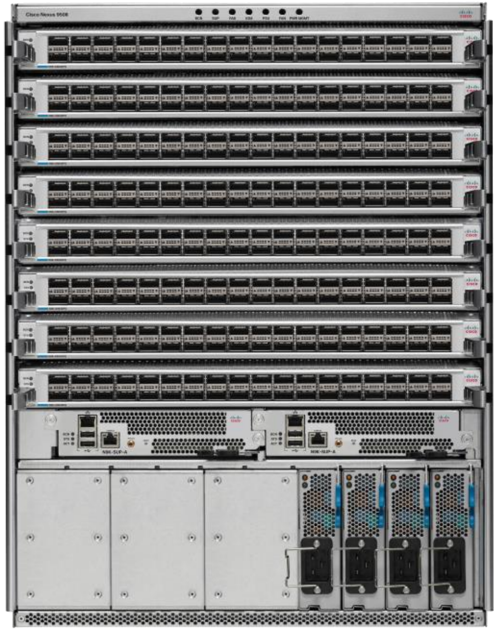
Juniper MX10003

- Carrier grade
- Fully modular and resilient
- 2x MX10003-LC2103 Cards
  - Modular line cards
  - 6x 40G QSFP+ internal ports
  - 12x JNP-MIC1-MACSEC 12-port 100G modular interface card
  - Deep buffer - 6GB
- Up to 10M routes in FIB!
- 100G WAN circuits virtualized for Internet connectivity, cloud and research connectivity (AL2S)

Images Courtesy of Juniper Networks

Duke Network – SuperCore Routers

Nexus 9508



- Fully modular and resilient
- Hardware similar to NCS-5508 carrier-grade router
- 2x N9K-X9636C-RX 36-port 100G cards
  - Deep Buffers – 16GB
  - Expanded TCAM
- Cisco Enhancements for Duke as of NX-OS 9.3(1)
  - LDP support and scale parity with Nexus 7700 (200 sessions)
  - 100G ER4 transceiver support
  - 40G and 100G BiDi support

$88,710.47

Images Courtesy of Cisco Systems

Duke Network – PetalCore Routers

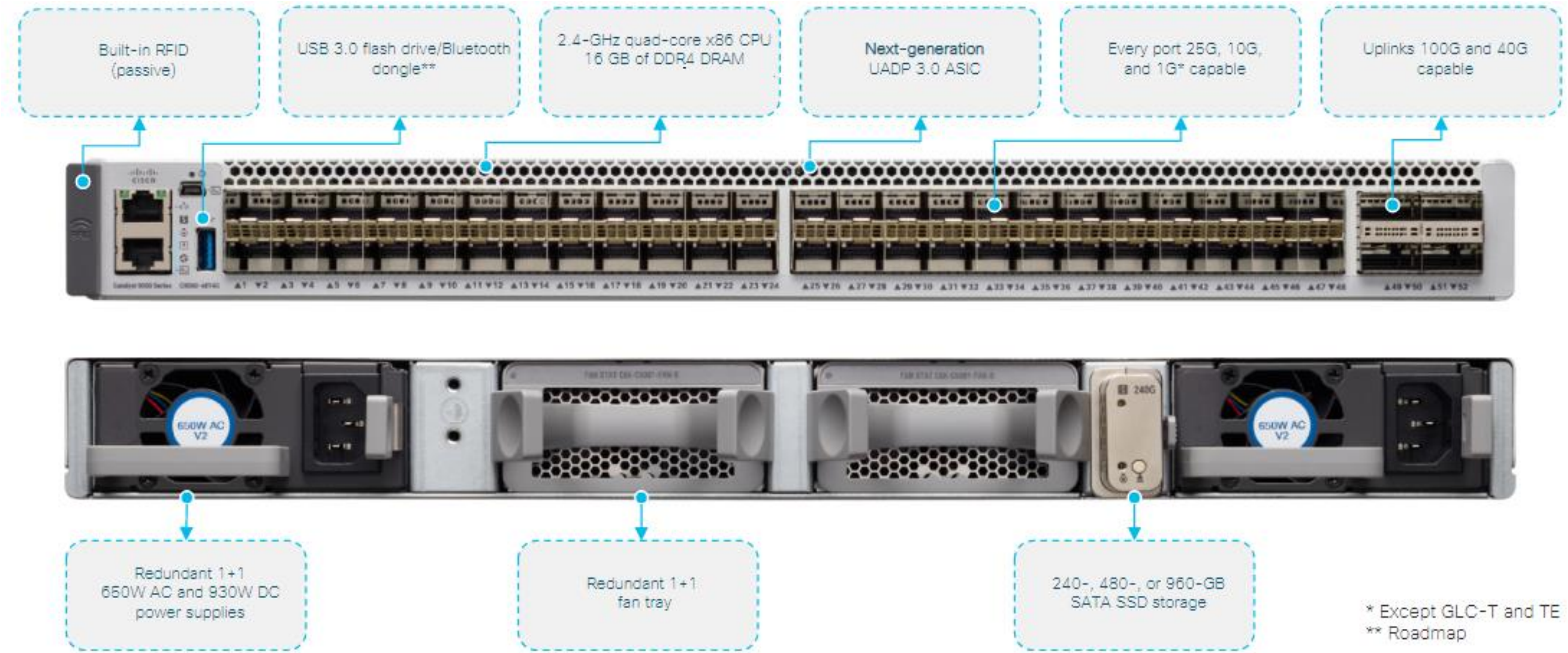
Nexus 3636C-R

- High density, fixed configuration 1RU chassis with resilient power and cooling
- Used in PetalCore, Fusion and IPE roles for standardization and economies of scale
- Same ASIC as N9K-X9636C-RX
  - Deep Buffers – 16GB
  - Standard TCAM
- Cisco Enhancements for Duke as of NX-OS 9.3(1)
  - LDP support and scale parity with Nexus 7700 (200 sessions)
  - 100G ER4 transceiver support
  - 40G and 100G BiDi support
  - DraftRosen MVPN (IPv4)

Images Courtesy of Cisco Systems

Duke Network – Campus Building Multi-Layer Distribution Switches

Catalyst 9500: 48x 10/25G downlink/uplink interfaces; 4x40/100G uplinks



Images Courtesy of Cisco Systems

Duke Network – Campus Building Access Layer

Catalyst 9300: 48x 10/100/1000 downlink interfaces; 2x10/25G uplinks

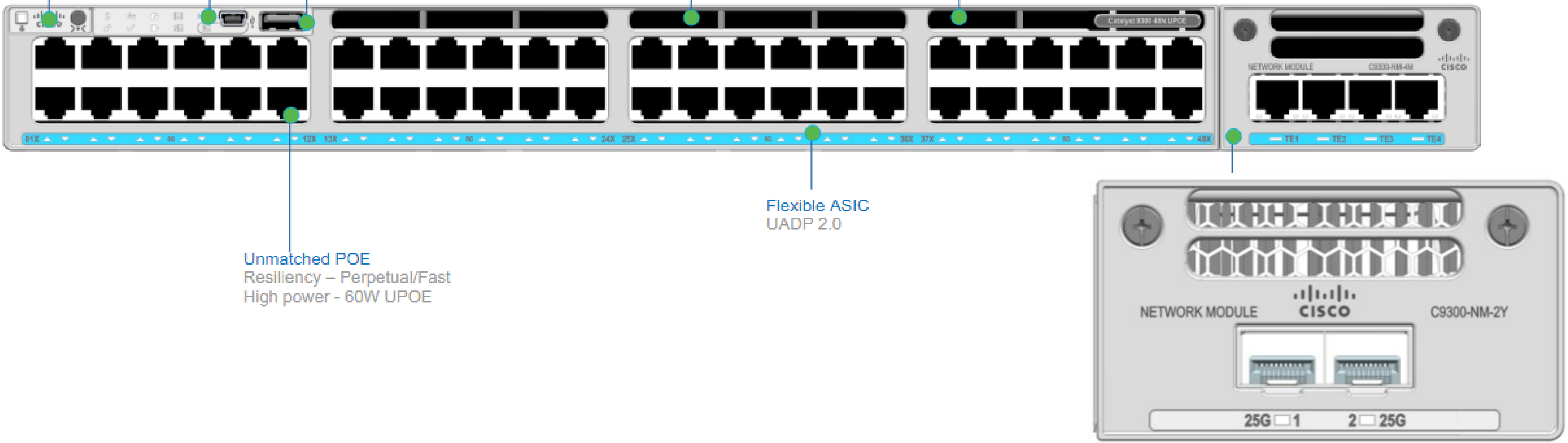
Built-in RFID
Passive

USB Console
Mini-USB type B

Bluetooth Dongle Support/External Storage
(USB 2.0)

In-built Memory
8GB Memory
16GB Flash

Powerful CPU complex
x86 CPU
4-core 1.8GHz



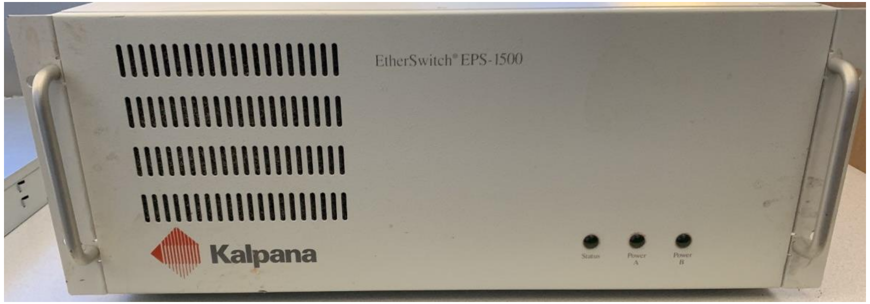
Flexible ASIC
UADP 2.0

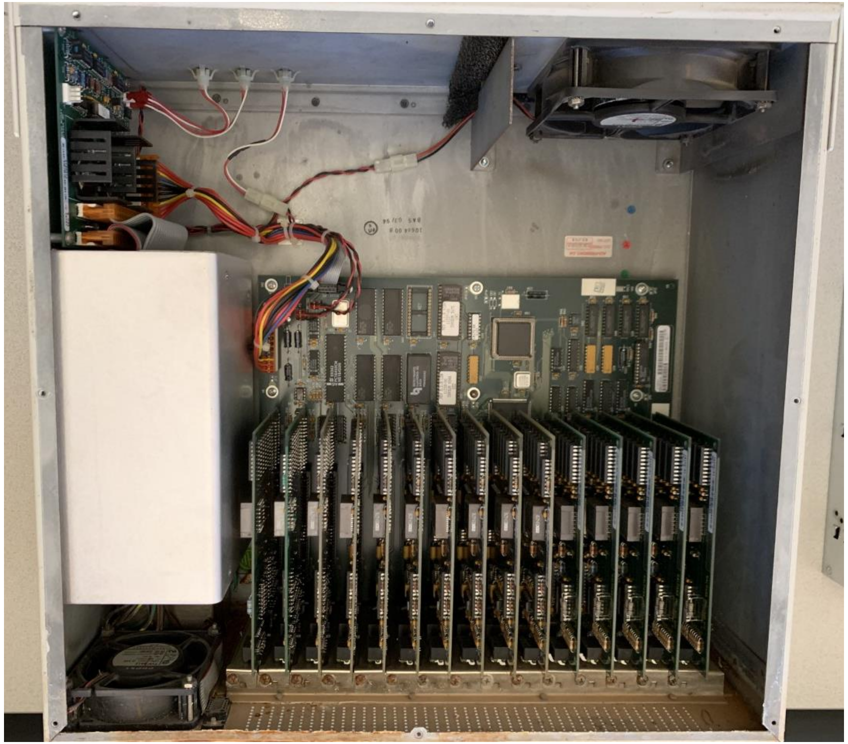
Unmatched POE
Resiliency – Perpetual/Fast
High power - 60W UPOE

Up to 9 switches can be stacked with a 480G proprietary ring topology

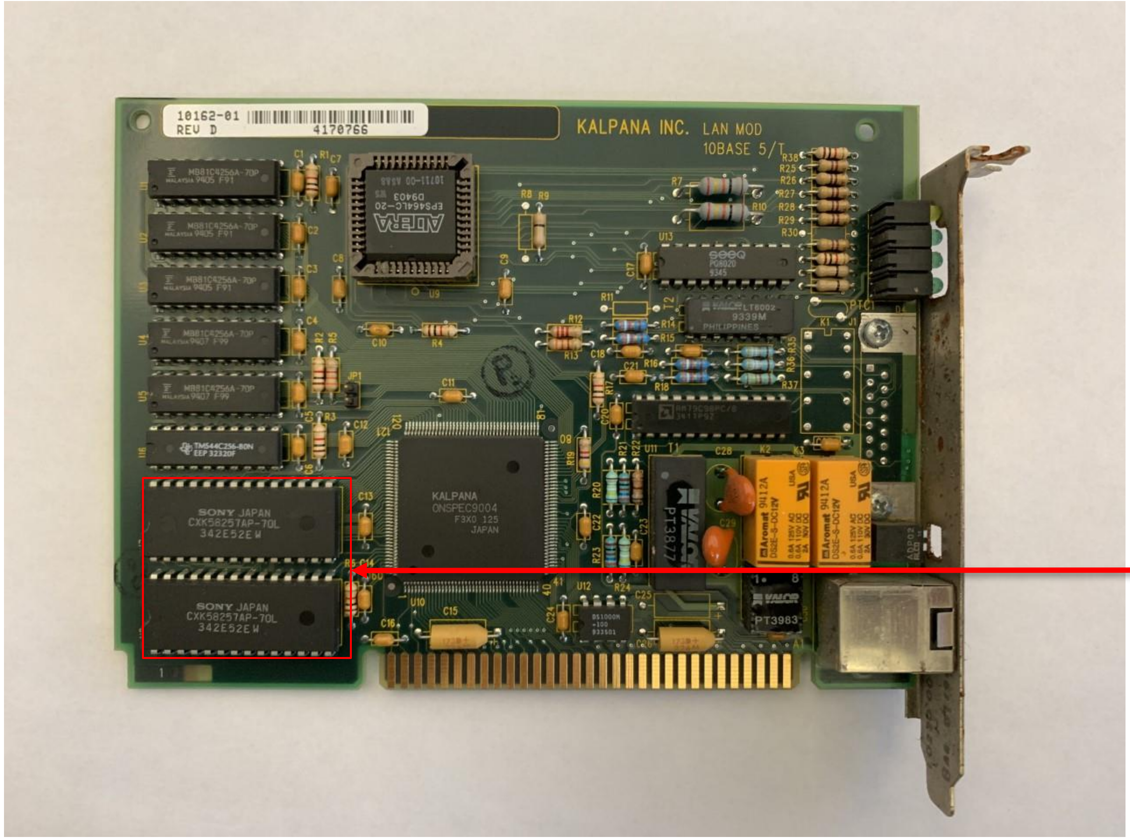
Images Courtesy of Cisco Systems

Historical Perspective – Kalpana Etherswitch





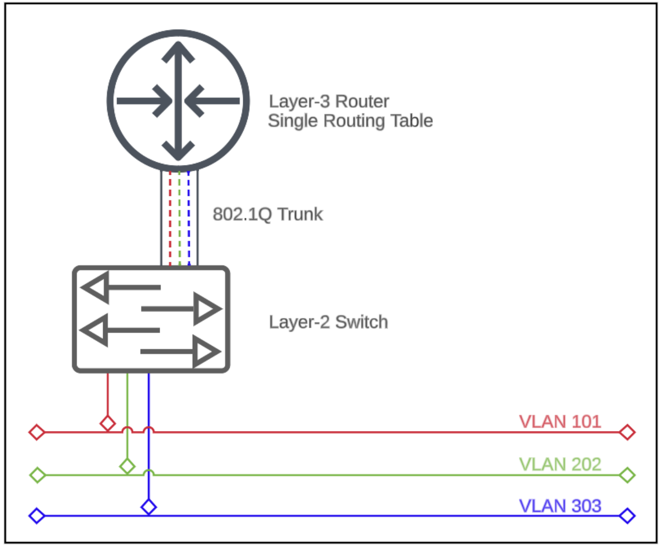
Historical Perspective – Kalpana – Network Interface Card (NIC)

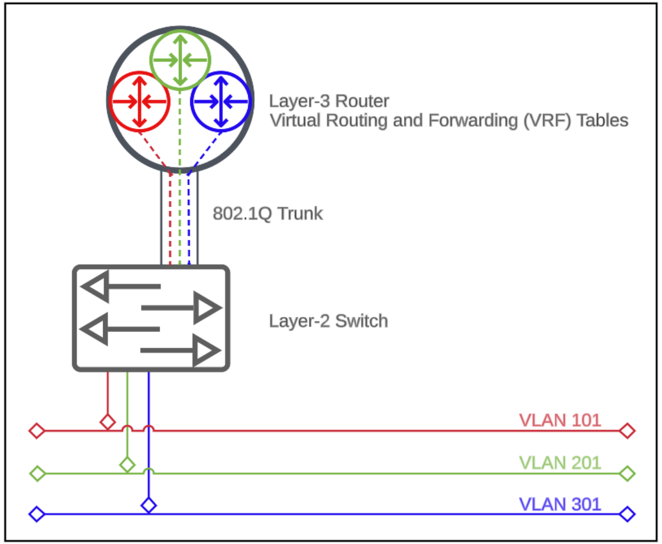


- 8-bit ISA Bus
- Qty 2: 32768-word X 8-bit High Speed CMOS Static RAM

Network Virtualization

- Segmentation supports multi-tenancy ex: sales, marketing, engineering
- Typically uses tags or labels for segmentation
- Support has evolved over time...





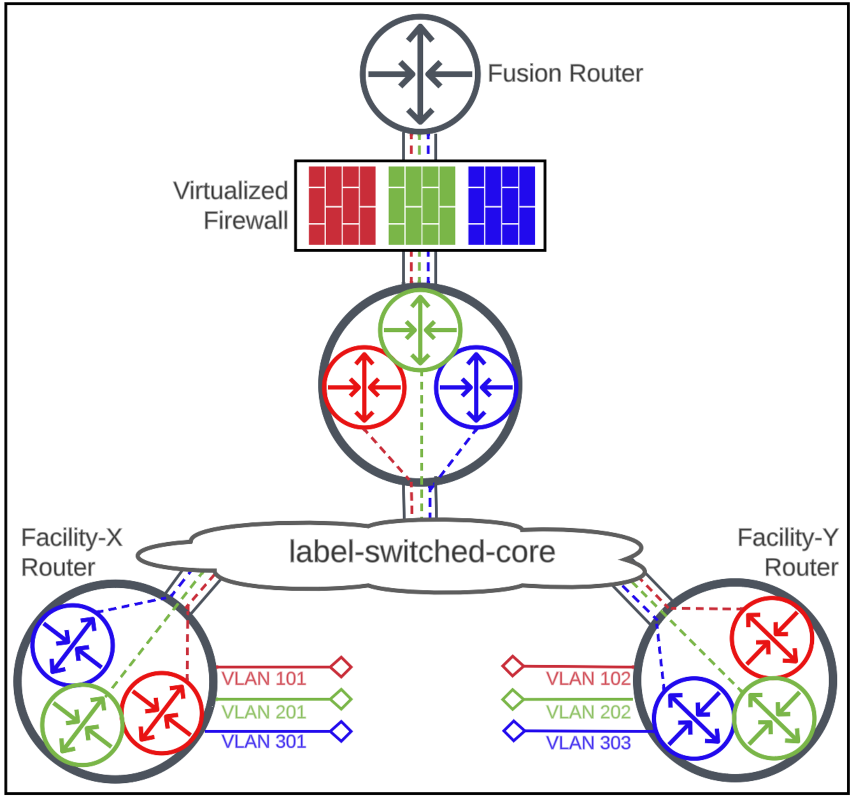
Duke Network Virtualization

- Objects in a virtual network can communicate with other objects within the same virtual network without restriction

- Objects in different virtual networks can communicate if allowed by virtualized firewalls

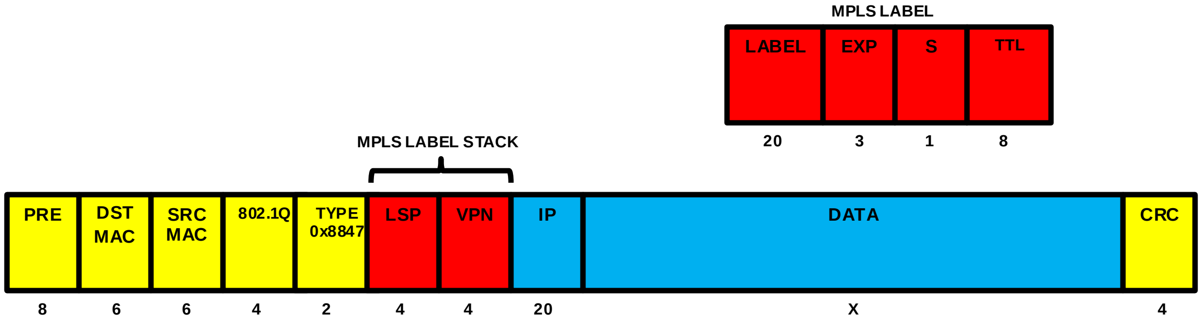
- Duke University has ~50 routed virtual networks
- Duke Health has ~45 routed virtual networks
- DKU has ~15 routed virtual networks

- These form a "coarse" form of virtualization given the number of hosts (computers)

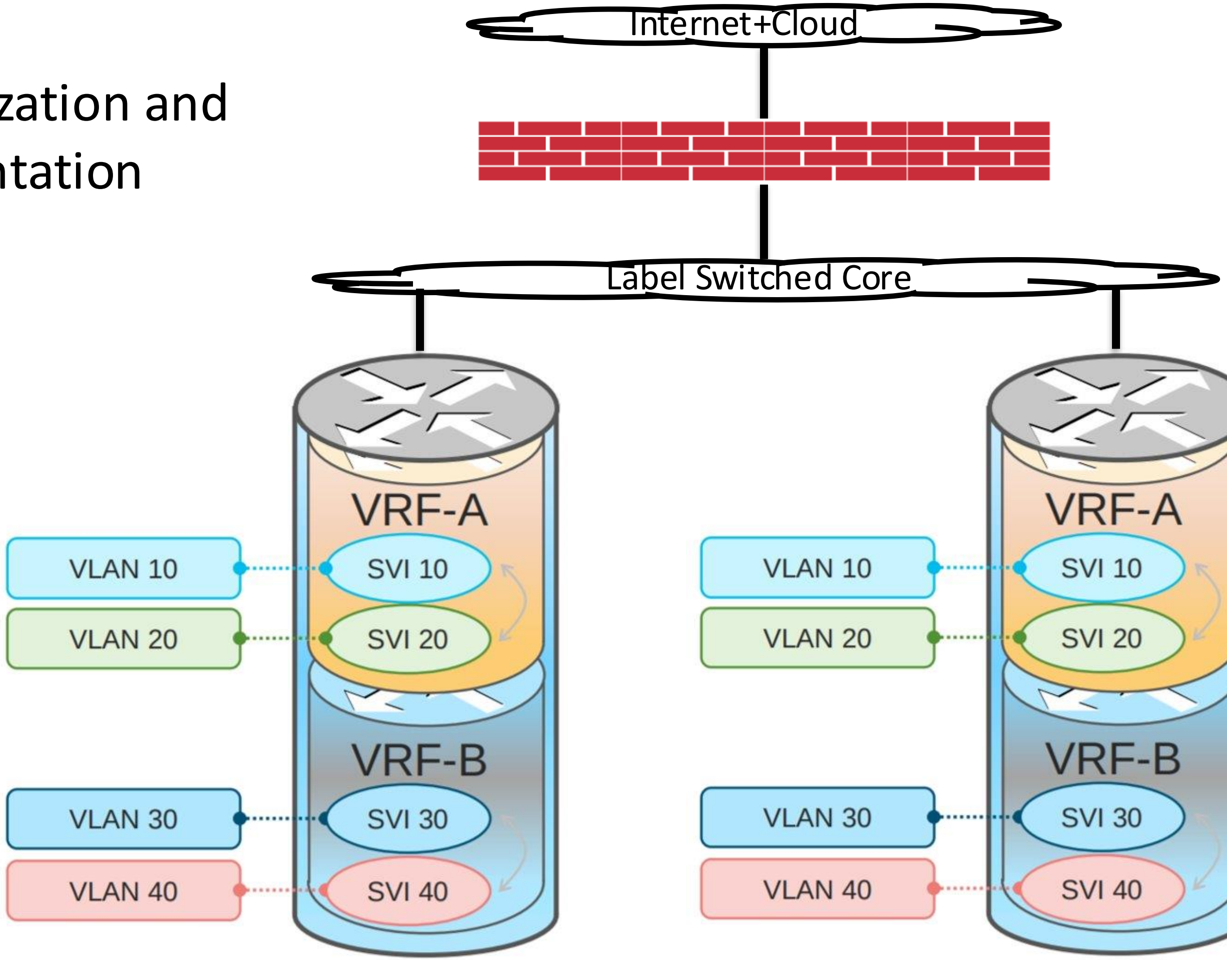


Multi-Protocol Label Switching (MPLS)

- In use by service providers for nearly two decades
- Many “enterprise” core/distribution platforms already have support
- MPLS resides at Layer 2.5:
  - Ethernet Frame – Layer 2
  - MPLS Frame – Layer 2.5
  - Payload (IP, etc) – Layer 3



Virtualization and Segmentation



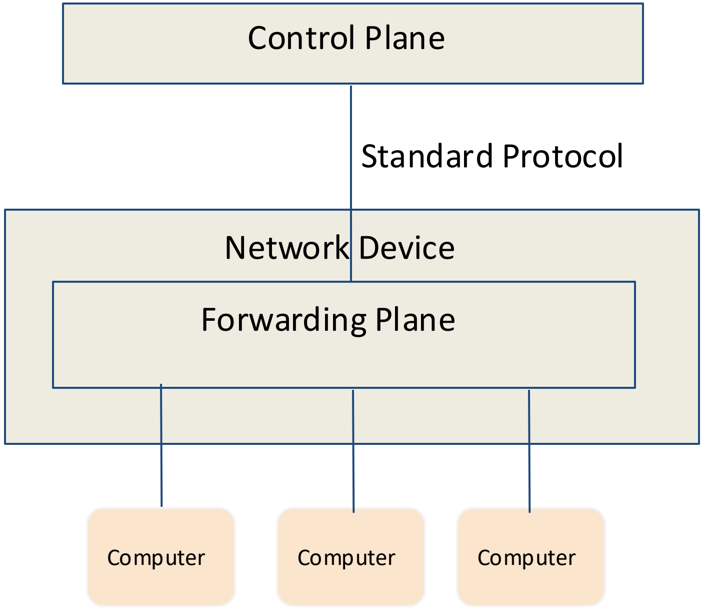
Images Courtesy of Cisco Systems

Software-Defined Networking (SDN)

- What is SDN?

  One answer: Separation of control plane from forwarding plane*

- Benefits: Rapid feature/protocol deployment; custom forwarding policies
  - Traffic engineering
  - Policy driven networking
  - Facilitates automation
  - Reduced complexity
  - Improve reliability by minimizing human error



SDN Review: Approaches

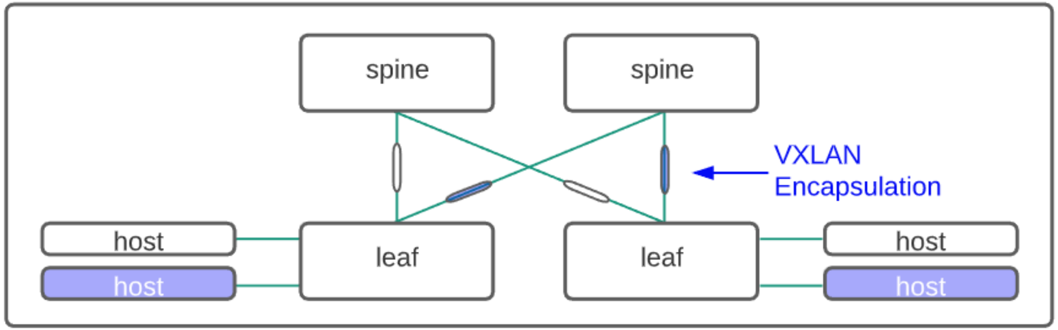
- Flow table/pipeline abstraction
  - Forwarding plane: table(s) with match+action
  - Control plane: Controller(s) populate tables with match+action
  - OpenFlow
  - P4: Programming Protocol-Independent Packet Processors
  - Switch Abstraction Interface (SAI)

- Overlay
  - Data plane: Tunnel Encap/Decap (VXLAN, GRE, etc)
  - Control Plane: EVPN (RFC7209), OpenContrail, etc

- Alternate: I2RS, PCEP, etc

EVPN/VXLAN - How Does It Work?

- VXLAN Encapsulation
  - Tenants, projects, customers are split into virtual networks called overlays
  - Traffic is encapsulated within VXLAN headers and transmitted on a common underlay network
  - Additional Questions:
    - What does the VXLAN encapsulation look like?
    - How are VXLAN datagrams directed to the right place?
    - What about segmentation after de-encapsulation?



EVPN/VXLAN - Encapsulation Format

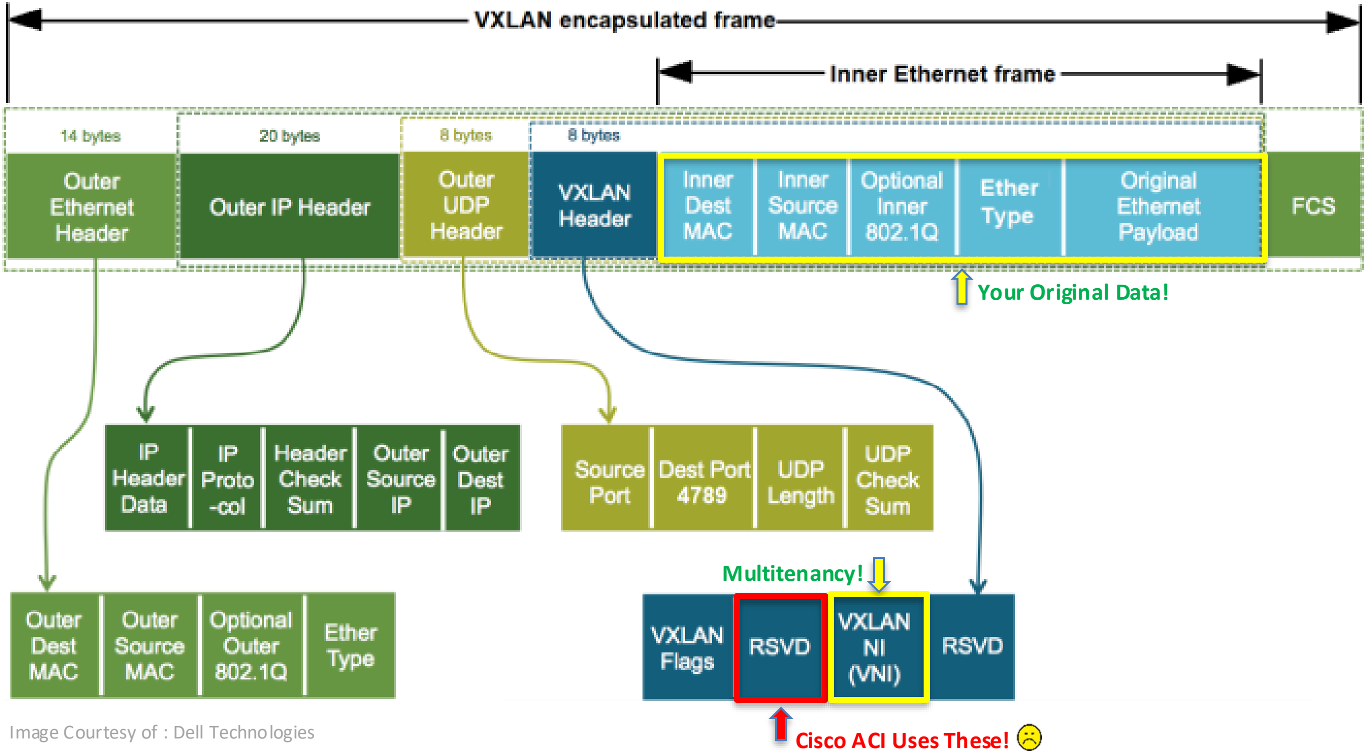
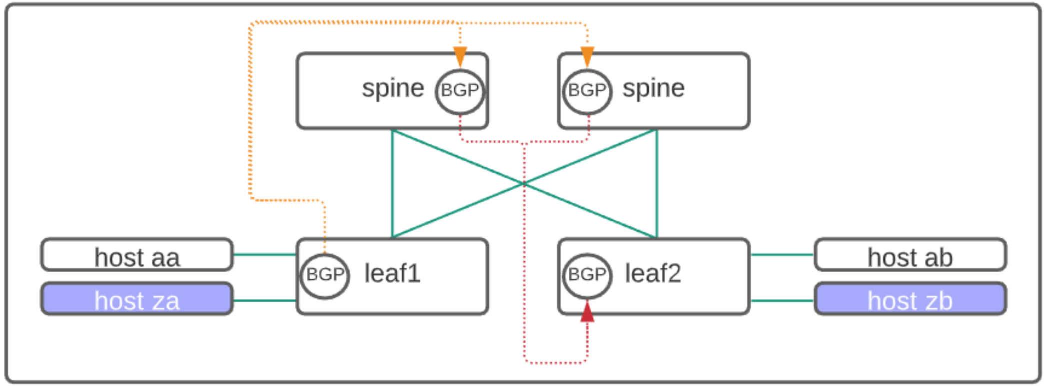


Image Courtesy of : Dell Technologies

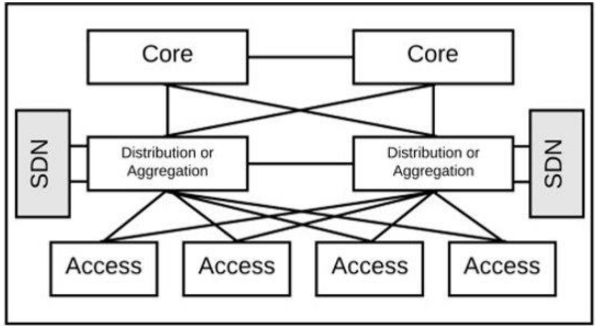
EVPN/VXLAN - VXLAN Datagram "Steering"

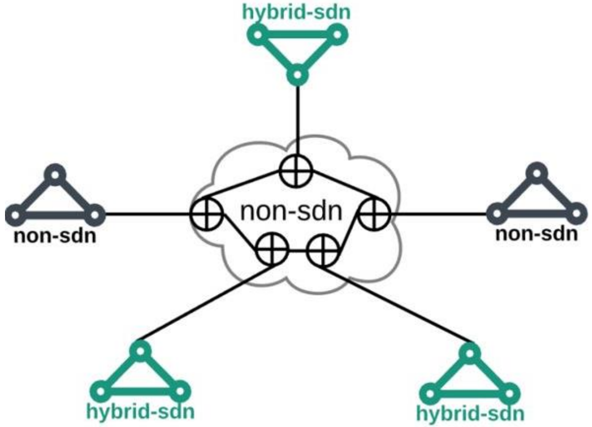
- Multi-Protocol Border Gateway Protocol (MP-BGP) Control Plane
  - There is no centralized server or expensive software!
  - How scalable is BGP? BGP routes the entire Internet!
  - How does it work? Host reachability is advertised between leaves via BGP. A given leaf then knows which leaf it should direct outbound traffic to!



SDN Research at Duke: Archipelago

- Leverages software to implement policy-driven forwarding
- Friction free policy enforcement WITHIN virtual networks
- Allow authorized flows, such large research data transfers, to bypass existing sources of friction
- Use low-cost appliances with commodity components and Smart Network Interface Cards (NICs) to improve likelihood of adoption
- **Islands of SDN nodes forms an Archipelago**

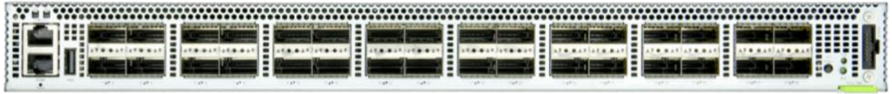




SDN Research at Duke: Data Plane Testing

Noviflow NS-2122 2x100G 20x 10G
NVIDIA/Mellanox NP-5 NPU

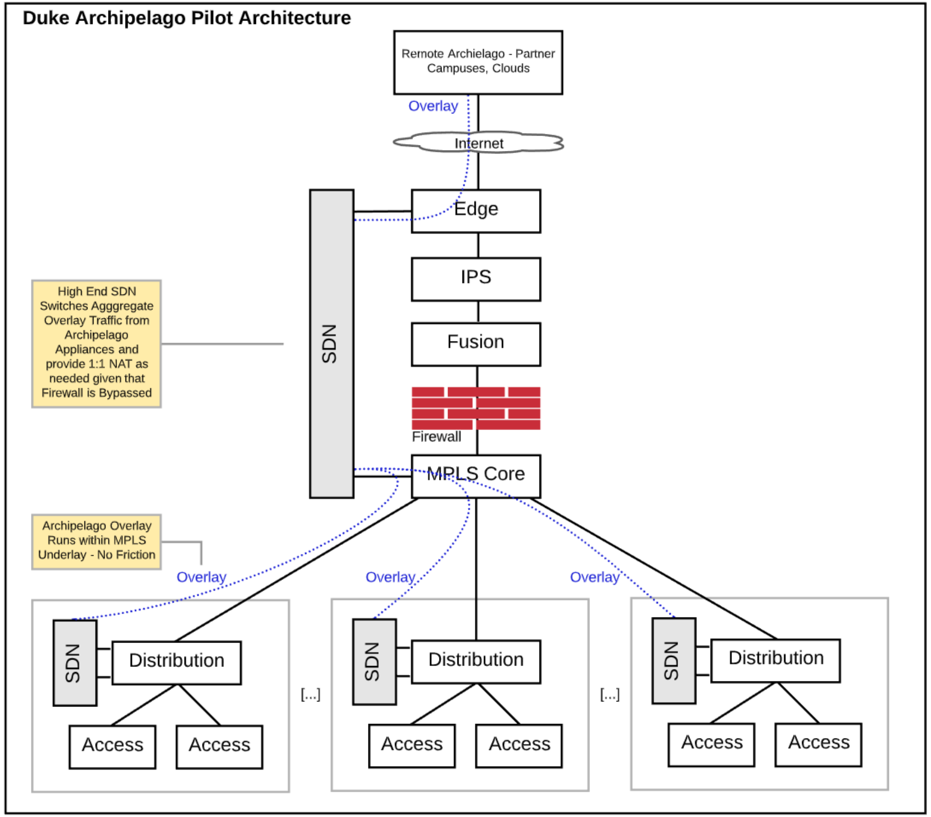
Noviflow Edge-Core WEDGE 100BF-32X
Intel/Barefoot Networks Tofino 32x 100G



3rd Generation Archipelago SmartNIC Appliance

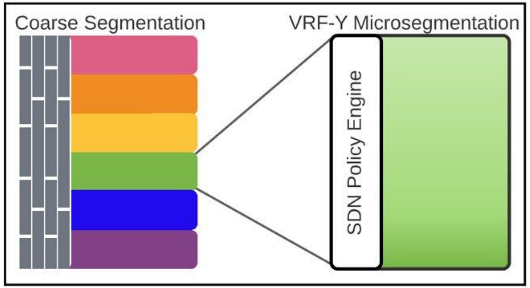
Images Courtesy of Noviflow, Supermicro, Netronome

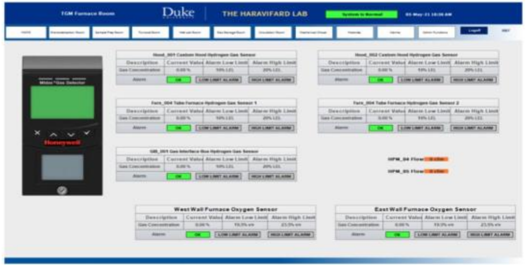
SDN Research at Duke: Archipelago Architecture



Archipelago Deployment: French Science TGMS

- Toxic Gas Monitoring System – False alarms cause weekly building evacuations
- Undesirable IP traffic was causing the gas sensors to become overwhelmed
- Traffic was sourced from WITHIN the virtual network – our firewall couldn't help – we need more segmentation!
- Friction-free policy-driven Archipelago SDN node deployed in the building, no more false alarms for several months!





Archipelago Deployment: Duke SMIF

- Thermo Fisher Cryo-Transmission Electron Microscope
- Outfitted with multiple specialized sensors
- Sound attenuating cabinet houses local servers
- Multiple 10Gb/s Ethernet connections
- Subnet in Duke OIT Research Computing Cluster VRF

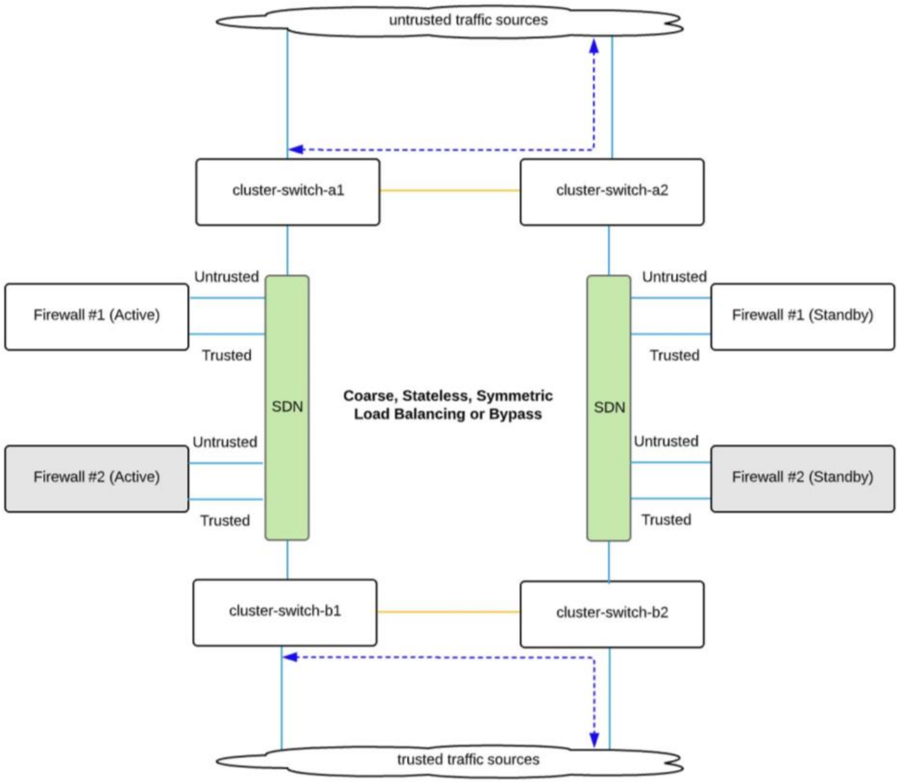
- **Protected via an Archipelago SDN appliance**

SDN at Duke: Research To Production

Migration to a single, unified network substrate in support of diverse academic, research and administrative use cases

Research-to-production:

- SDN perimeter rapid response layer
- SDN driven perimeter IPS
- SDN driven scalable firewall
- SDN driven network monitoring fabric*



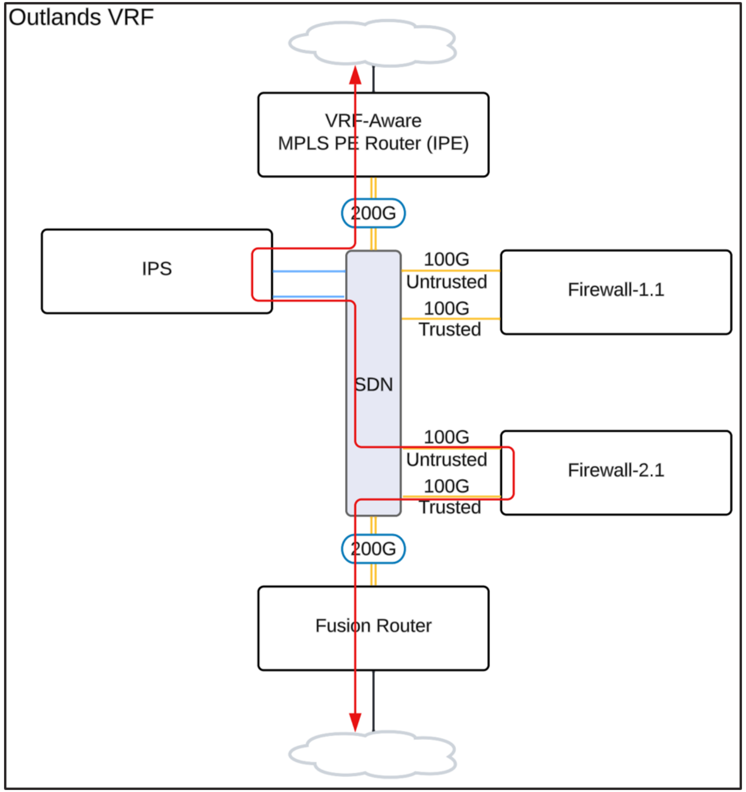
* Installation in progress

SDN at Duke: Research To Production – Service Chaining

Research-to-production:

- We leverage service chaining with SDN to transparently insert additional forms of inspection to the architecture described previously

- Virtual Computing Manager (VCM) traffic in this example... but also for visitor and DKU traffic arriving at Duke



SDN at Duke: Research To Production – Monitoring Fabr

- Monitoring Fabric: collects traffic at strategic vantage points using passive optical taps and/or port mirroring

- Slices and intelligently load-balances collected traffic and presents this to low-cost Archipelago-style appliances for traffic sensing via containerized Zeek, Suricata, and Yaf instances as developed by our partners in the IT Security Office (ITSO)

- A data source for MISTRAL - Massive Internal System Traffic Research Analysis and Logging – an NSF funded project at Duke

Futures: Explore Commodity Network Infrastructure

- The "opaque" network forwarding elements described previously are starting to become more open and commoditized
- Software for Open Networking In the Cloud (SONIC) - initially developed by Microsoft for Azure – now a Linux Foundation Project
- In 2025 we will explore some merchant silicon-based switches for campus use that support a variety of open network operating systems including SONIC

Images Courtesy of Edge-Core

Futures: DPU Analysis and Testing

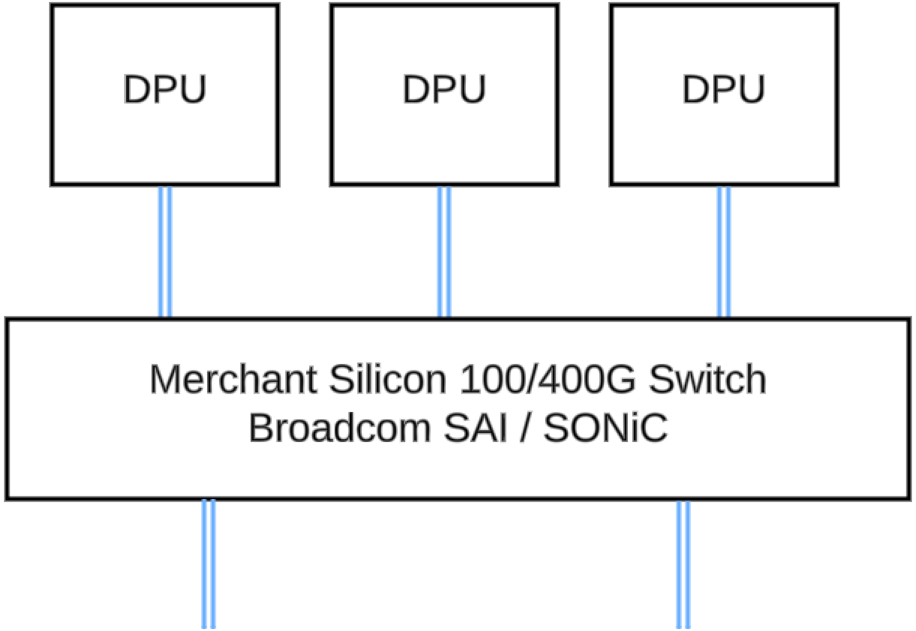
AMD/Pensando DPU

Nvidia Bluefield-3 DPU

Archipelago SDN/NFV/Control Appliances

Images Courtesy of AMD, NVIDIA, Supermicro

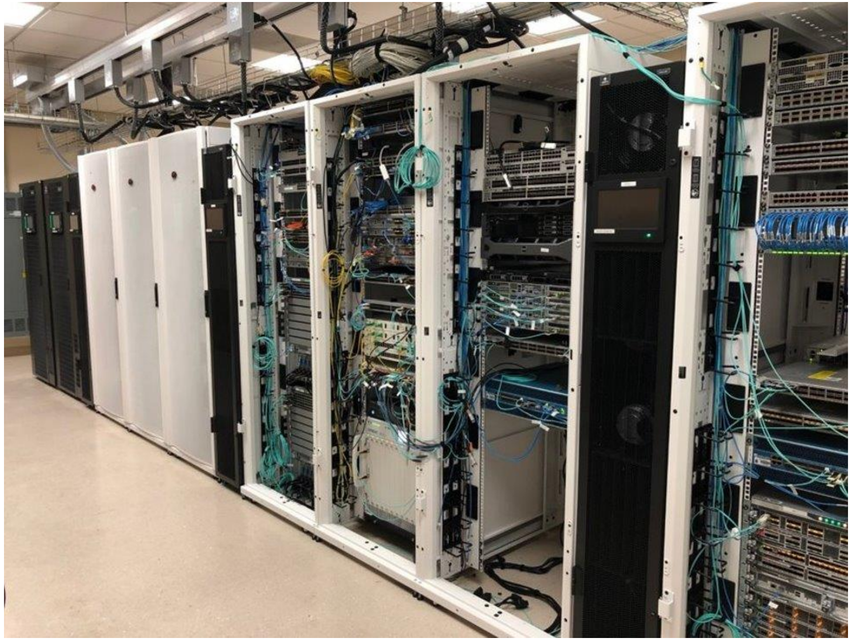
Futures: Commodity Open Networking for Expedited Research (CONIFER)



Network Infrastructure Lab

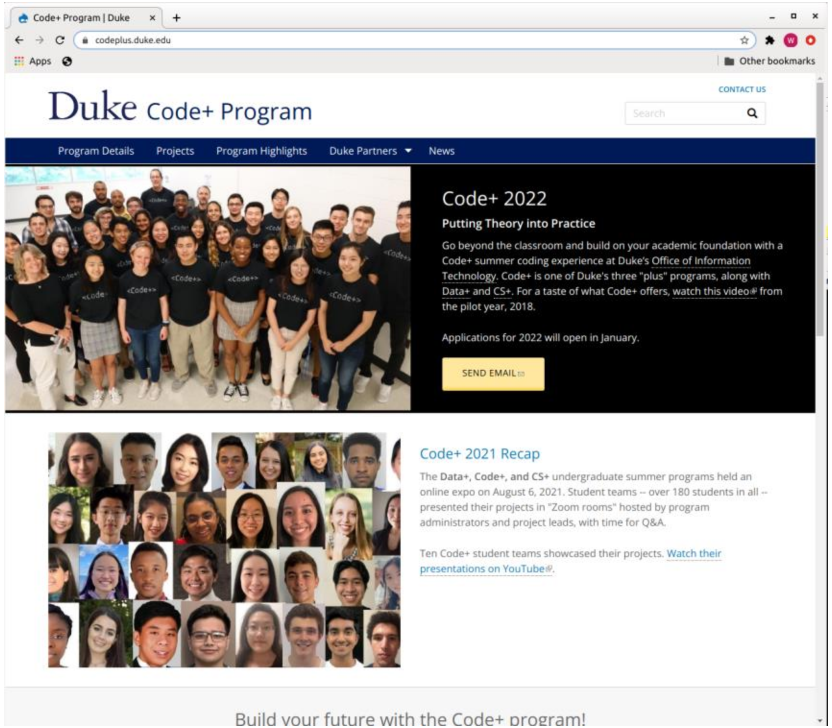
Makes all of this possible:

- Increases operational excellence
- Facilitates testing of hardware/software; a variety of interconnections and network topologies
- Part of our strategy to minimize equipment makes/models - a repository of on-site spares
- Spirent traffic generator and network emulator – essential to apply load+stress to infrastructure
- Savings in qualification of 3rd party optics helped to fund many components within the lab
- Supports research too!

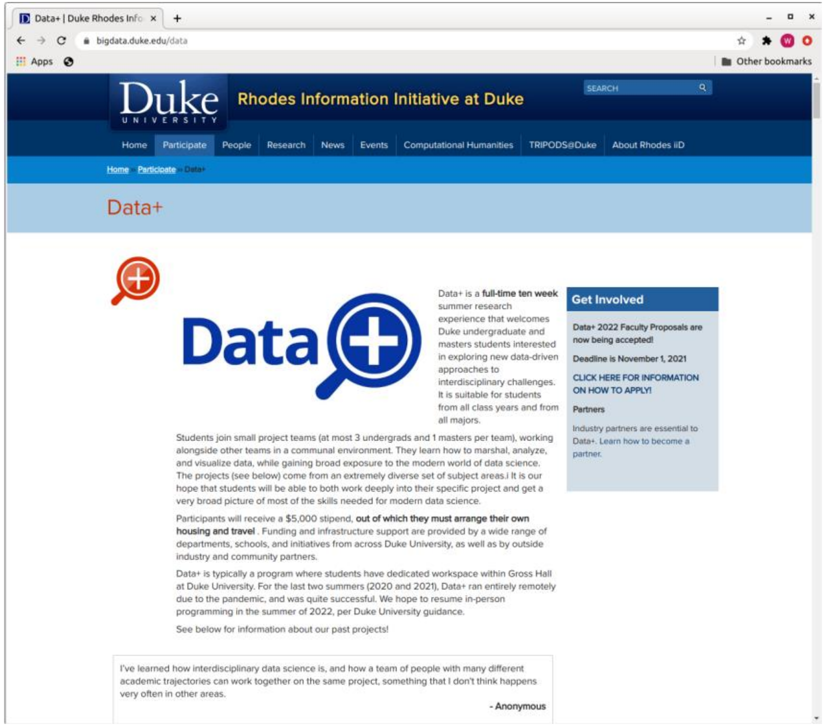


Get Involved!

Duke Code+



Duke Data+



OIT Research

Active Awards:

- CC* Data Storage: Flexible Affordable Scalable Technology for Research Storage – OAC-2232810

- CC* Compute: NCShare Compute as a Service – OAC-2201105
- CC* Regional: NCShare Science DMZ – OAC-2201525
- CC* Regional: NCShare GPU-as-a-Service (proposed)

- CICI: RSSD: Massive Internal System Traffic Research Analysis and Logging – OAC-2232819
- ECE+OIT: EAGER: An Integrated Fiber Sensing and Communication Living Lab in the Research Triangle (CNS-2330333)

Q&A