CORD: FABRIC

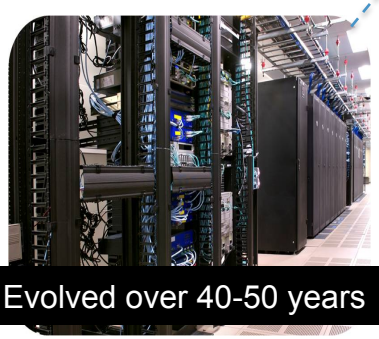
An Open-Source Leaf-Spine L3 Clos Fabric

Saurav Das
Principal System Architect, ONF

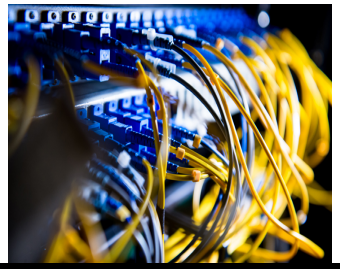
In collaboration with:

Problem: Today's Telco Central Offices (COs)

Large number of COs



Evolved over 40-50 years



Huge source of CAPEX/OPEX

Message Router

CDN

Session Border Controller

DPI

Firewall

Carrier Grade NAT

SGSN/GGSN/ PDN-GW

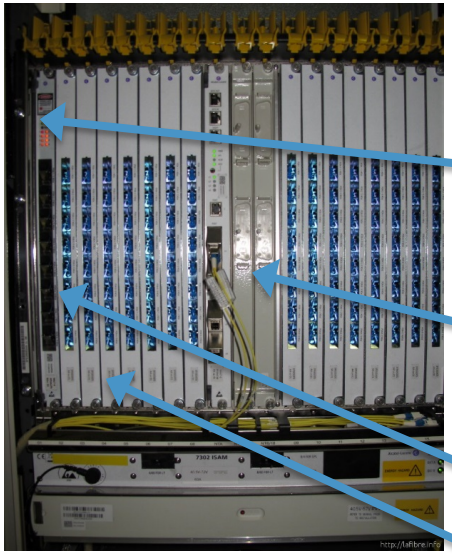
PE Router

BNG

- Fragmented non-commodity hardware.
- Physical install per appliance per site
- Nearly **300+** unique deployed appliances.

Huge source of CAPEX/OPEX
Not geared for Agility/ Programmability
Does not benefit from Commodity Hardware

CORD: Central Office Re-architected as Datacenter



GPON OLT

Applications

SDN Control Plane- ONOS

NFVI Orch- XOS

Spine Switches

9 10 11 12

Fabric

Leaf Switches

1 2 3 4 5 6 I O I O

Control

PMC

PON OLT MACs

vOLT vBNG LDAP vCPE DHCP RADIUS

Metro Core Link

Access Link

Commodity hardware

Data

Simple CPE

ONT

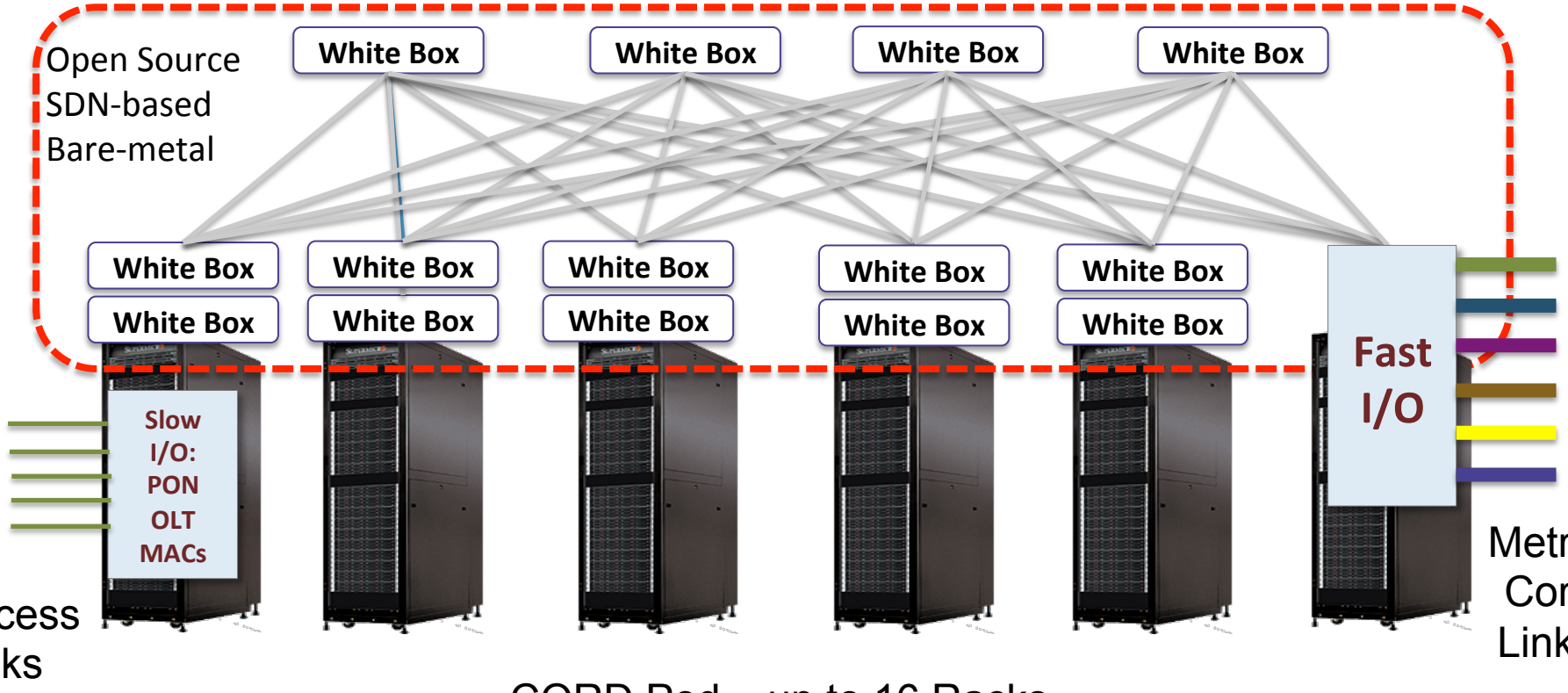
GPON

Open-Source Leaf-Spine Fabric

Fabric Control Application: Addressing, ECMP Routing, Recovery, Interoperability, API support

HA, scales to 16 racks, OF 1.3, Topo-Discovery, Configuration, GUI, CLI, Troubleshooting, ISSU

ONOS Controller Cluster



CORD Pod – up to 16 Racks

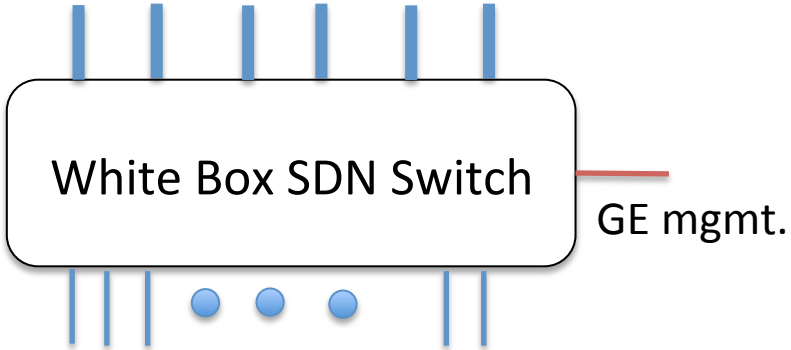
Open-Source Leaf-Spine Fabric

Spine Switch

32 x 40G ports downlink to leaf switches
40G QSFP+

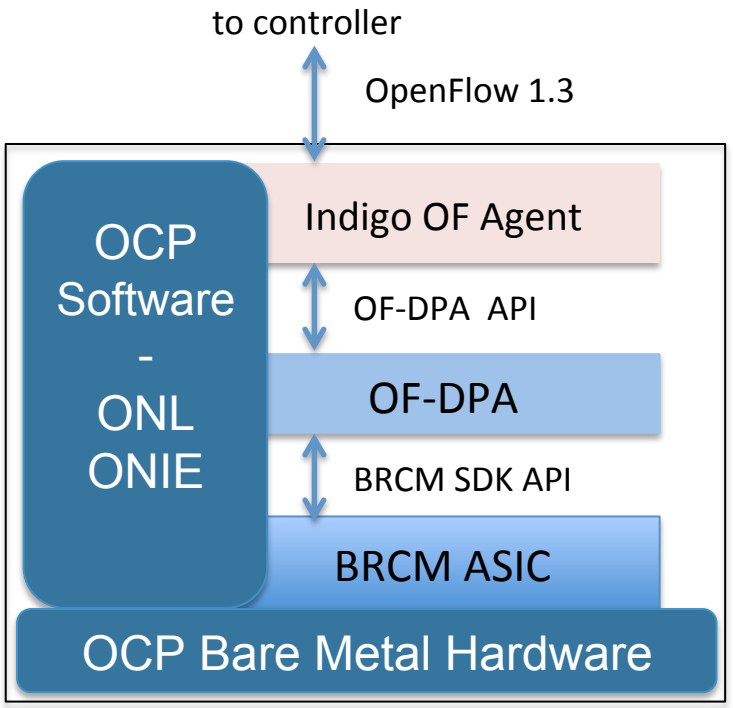
Leaf Switch

6 -12 x 40G ports uplink to different spine switches
ECMP across all uplink ports



48 x 10G ports downlink to servers
in the same rack (subnet)
10G Base T or 10G SFP+

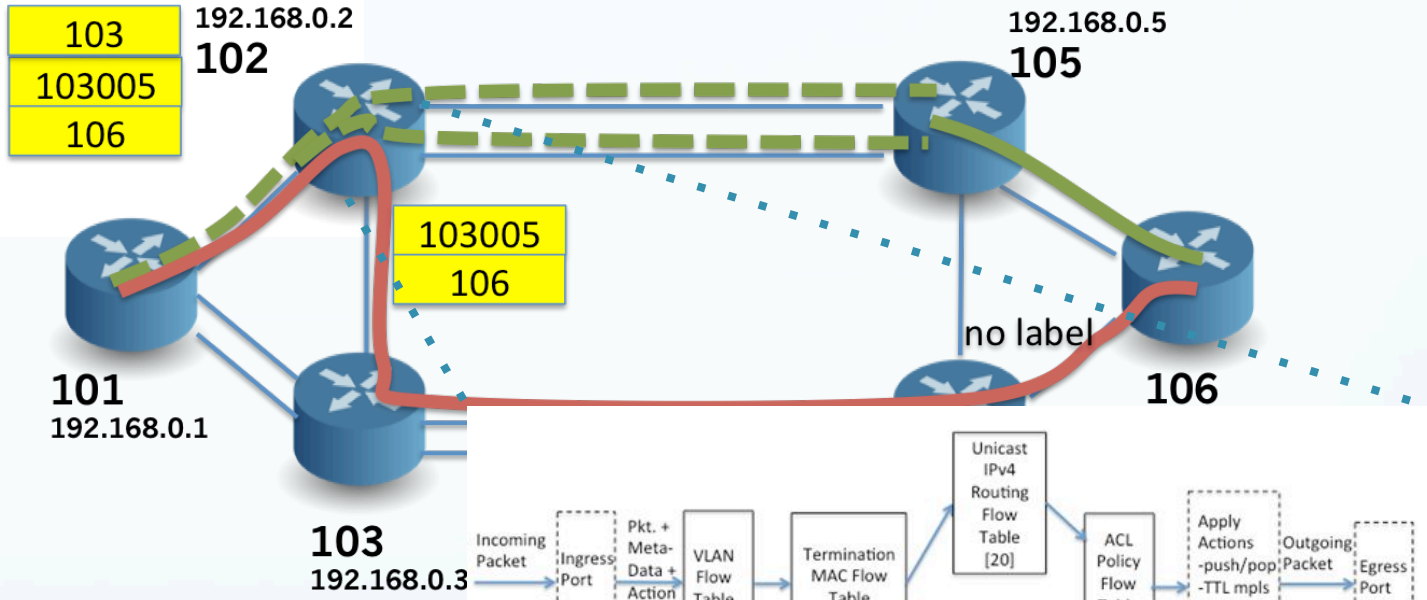
Leaf/Spine Switch Software Stack

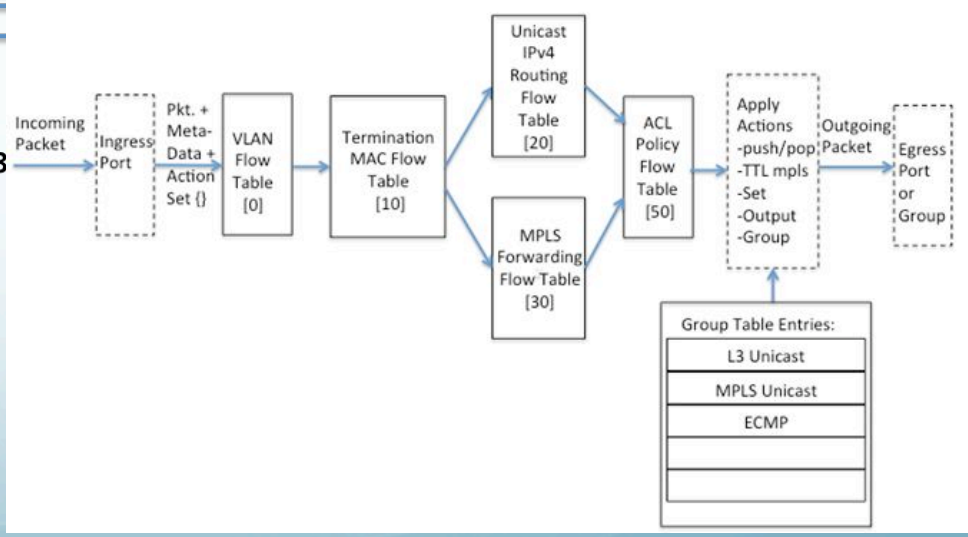


OCP: Open Compute Project
ONL: Open Network Linux
ONIE: Open Network Install Environment
BRCM: Broadcom Merchant Silicon ASICs
OF-DPA: OpenFlow Datapath Abstraction

SPRING-OPEN

Segment Routing on Bare Metal Hardware

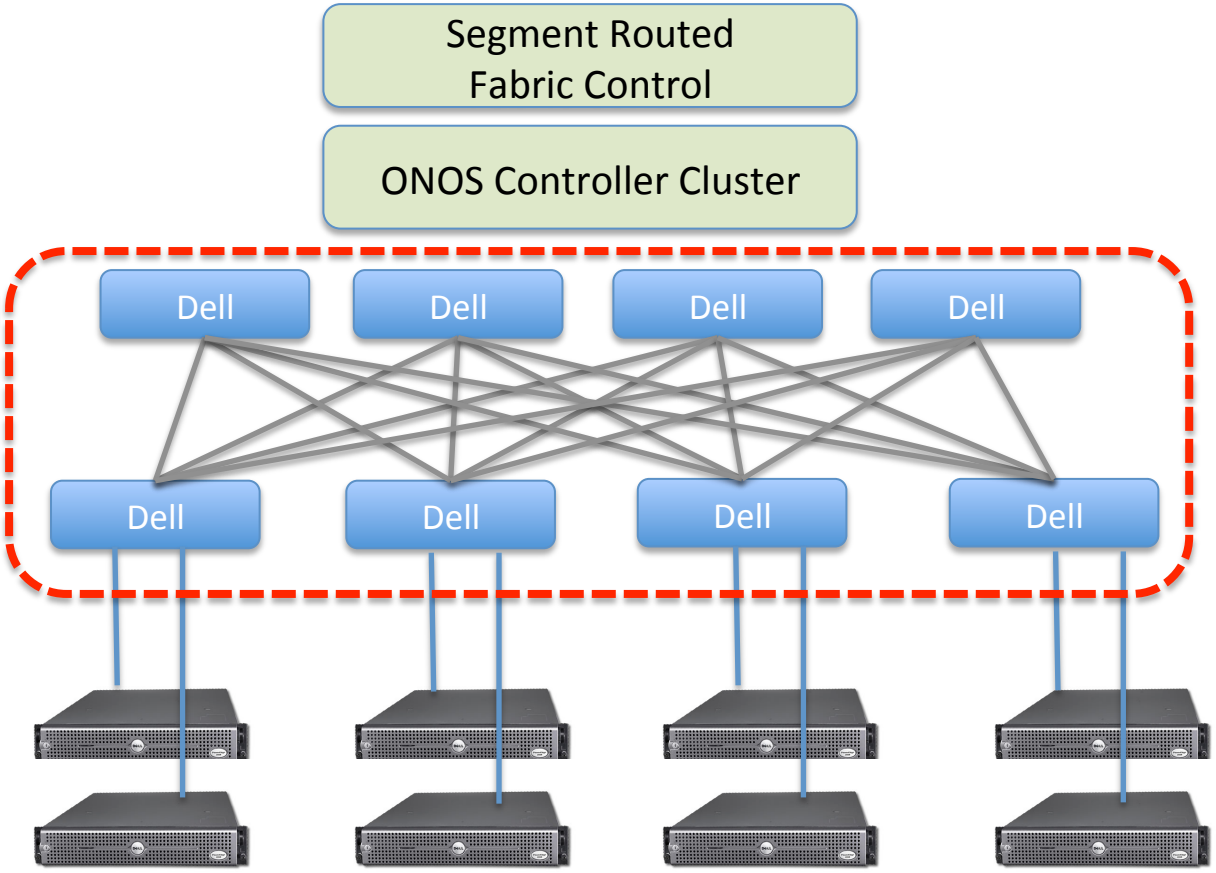




Learn more:
https://wiki.onosproject.org/display/ONOS/Segment+Routing

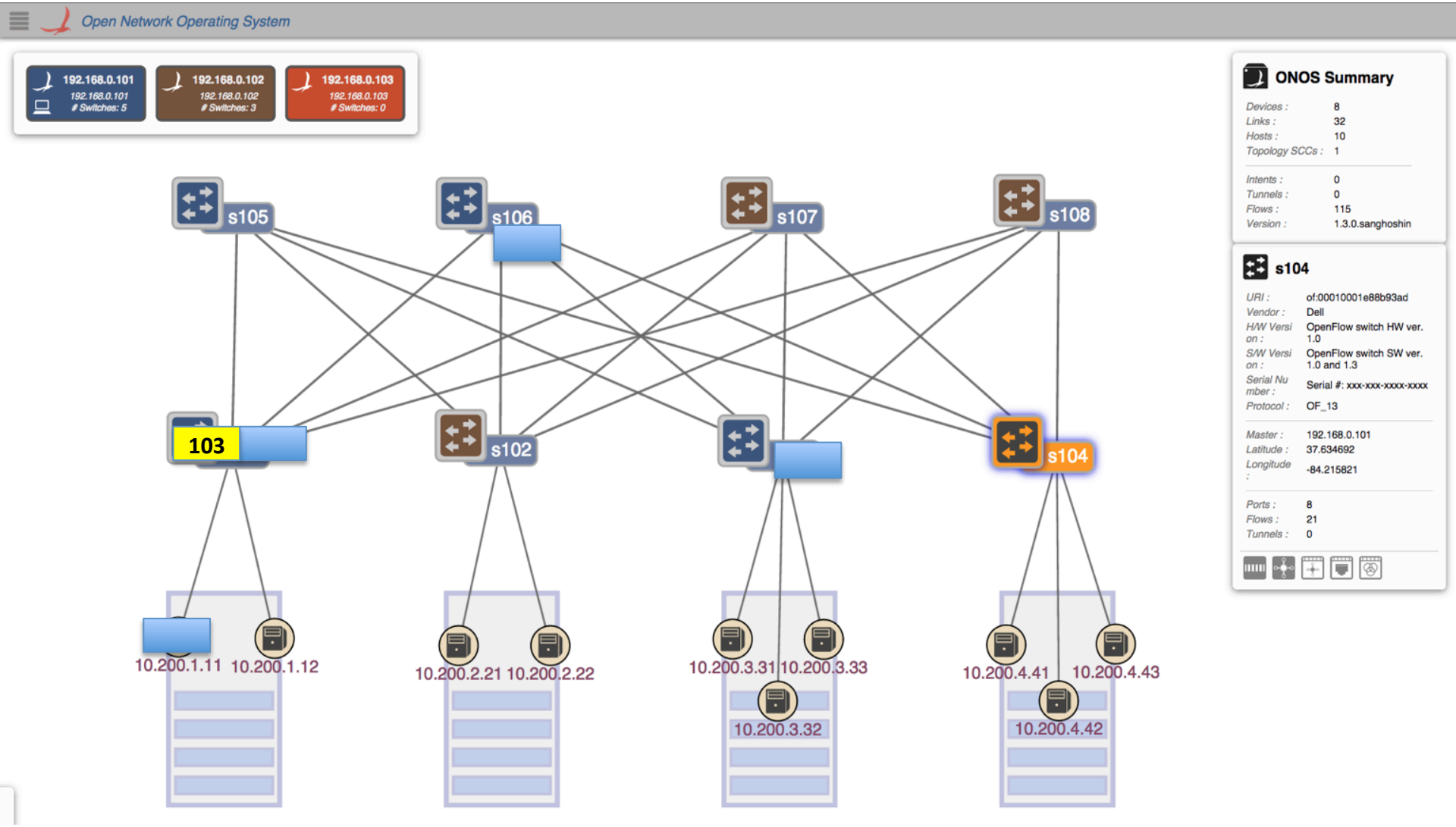
First Step: PoC Demo at Solution Showcase

SDN controlled
L3 Leaf-Spine
Clos fabric.

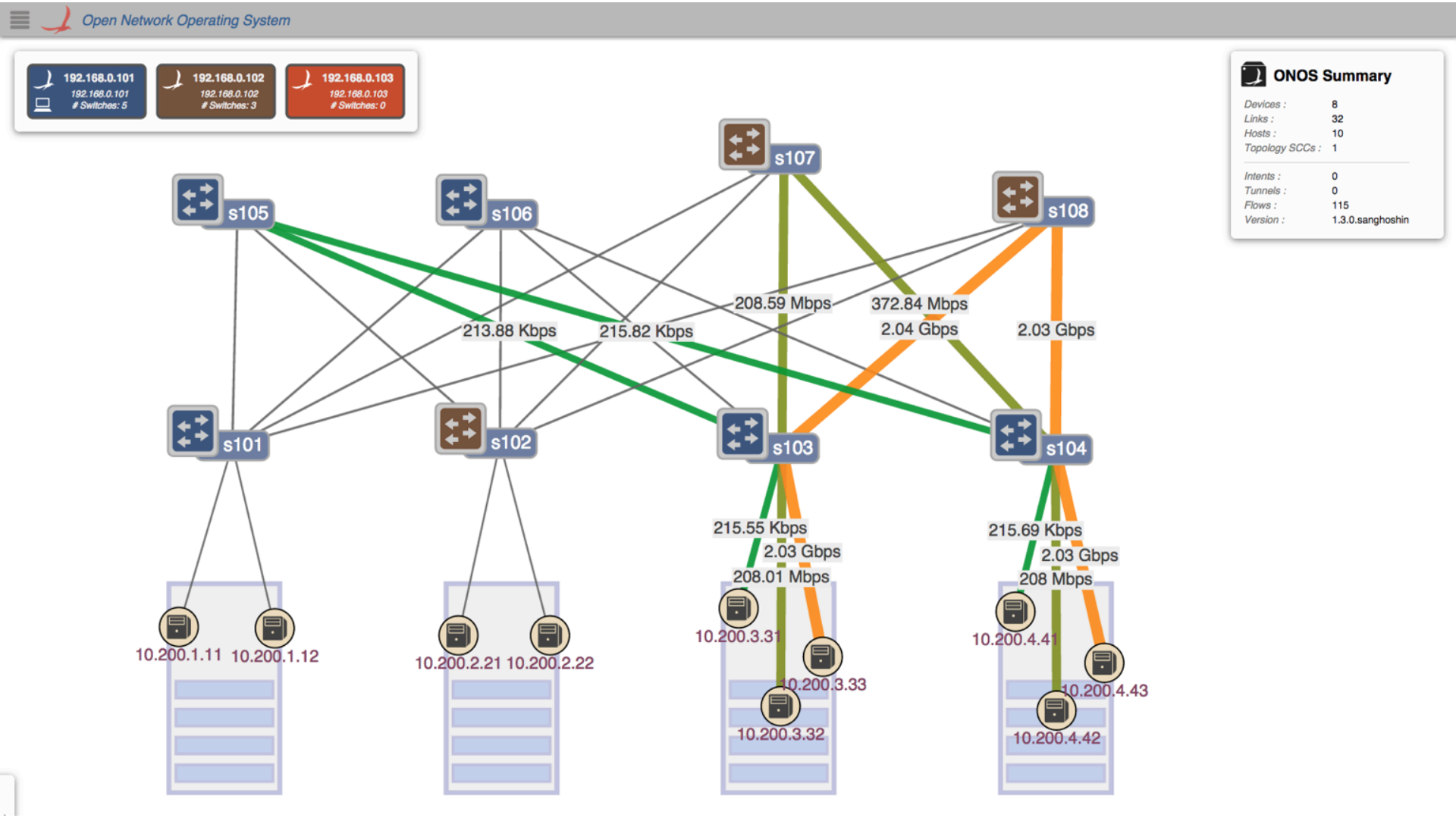


4 racks, 2 servers/rack, Dell 4810 bare metal,
ONOS Cardinal controller cluster

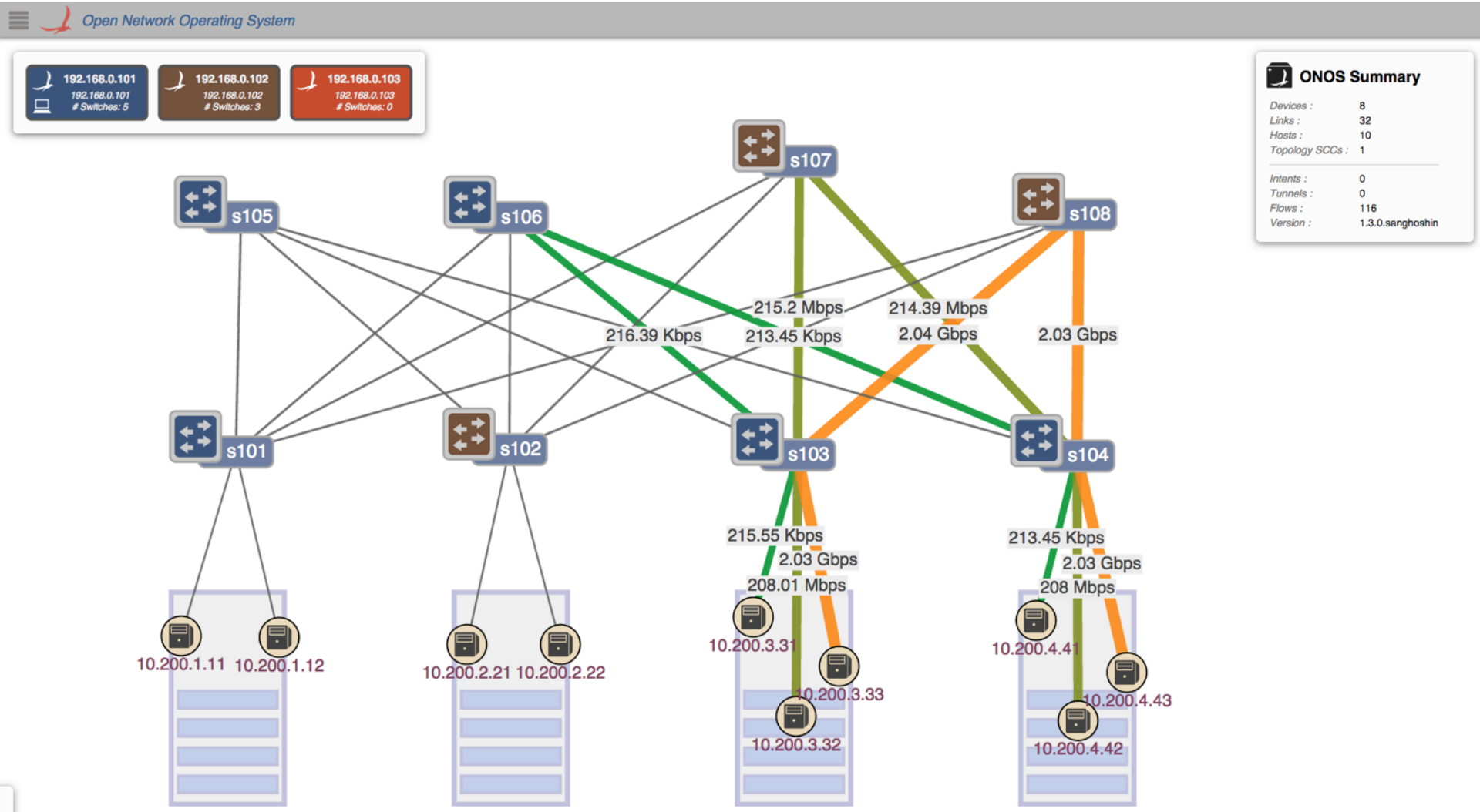
CORD: Fabric


Open Network Operating System
192.168.0.101
192.168.0.101
Switches: 5
192.168.0.102
192.168.0.102
Switches: 3
192.168.0.103
192.168.0.103
Switches: 0
ONOS Summary
Devices : 8
Links : 32
Hosts : 10
Topology SCCs : 1
Intents : 0
Tunnels : 0
Flows : 115
Version : 1.3.0.sanghoshin
s104
URI : of:00010001e88b93ad
Vendor : Dell
H/W Version : OpenFlow switch HW ver. 1.0
S/W Version : OpenFlow switch SW ver. 1.0 and 1.3
Serial Number : Serial #: xxx-xxx-xxxx-xxxx
Protocol : OF_13
Master : 192.168.0.101
Latitude : 37.634692
Longitude : -84.215821
Ports : 8
Flows : 21
Tunnels : 0
s105
s106
s107
s108
103
s102
s104
10.200.1.11
10.200.1.12
10.200.2.21
10.200.2.22
10.200.3.31
10.200.3.33
10.200.3.32
10.200.4.41
10.200.4.43
10.200.4.42

ECMP Routing



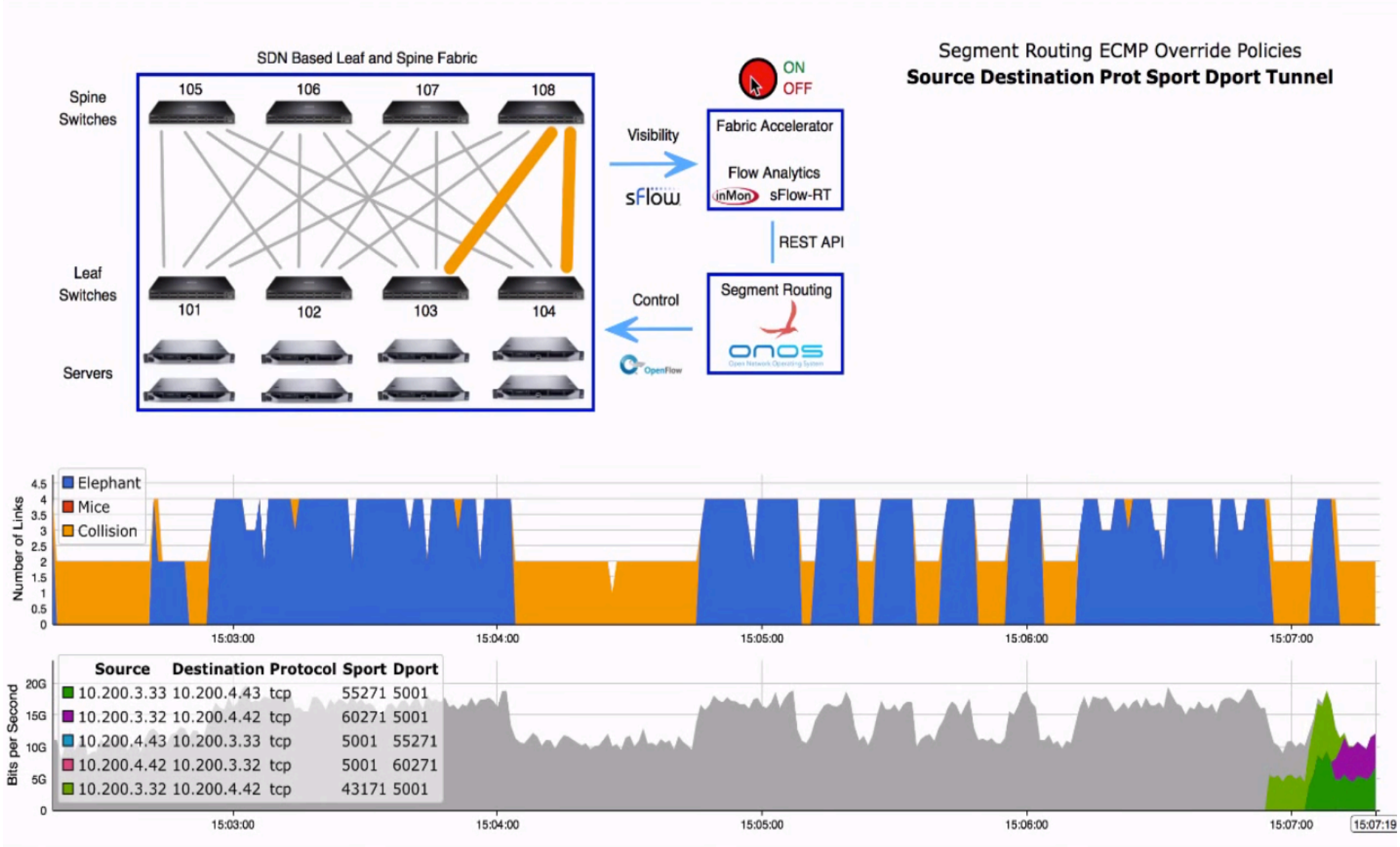
Policy Driven Traffic Engineering



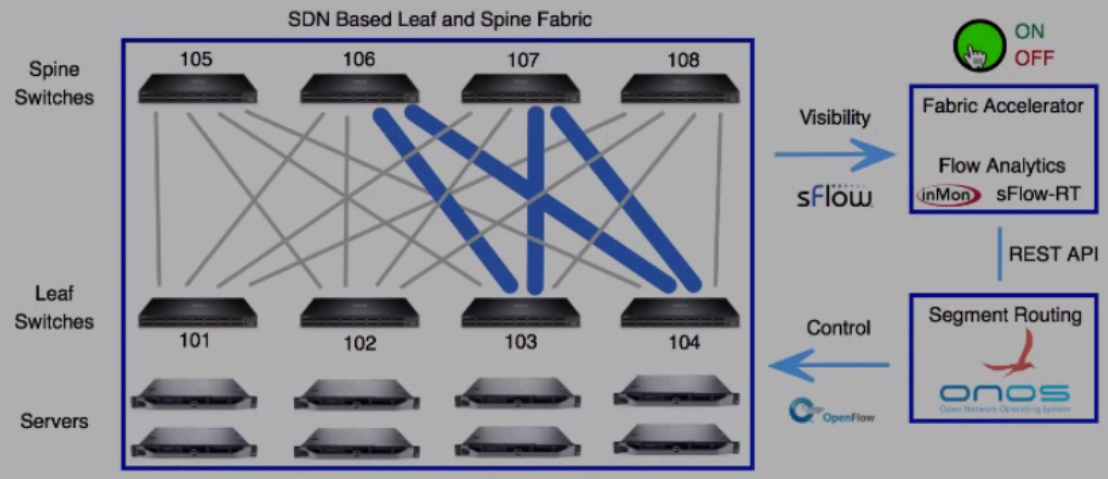

```
onos> srpolicy-add p1 1000 10.200.3.31/32 10.200.4.41/32 TUNNEL_FLOW sr34
onos> srpolicy-list
  ID    TYPE          PRIORITY  SRC_IP           DST_IP           TUNNEL_ID
  p1 _  TUNNEL_FLOW   1000      10.200.3.31/32   10.200.4.41/32   sr34
```

```
onos> srtunnel-list
   ID  GROUP    LABELS
sr34   175      [103, 106, 104]
```

Analytics Driven Traffic Engineering

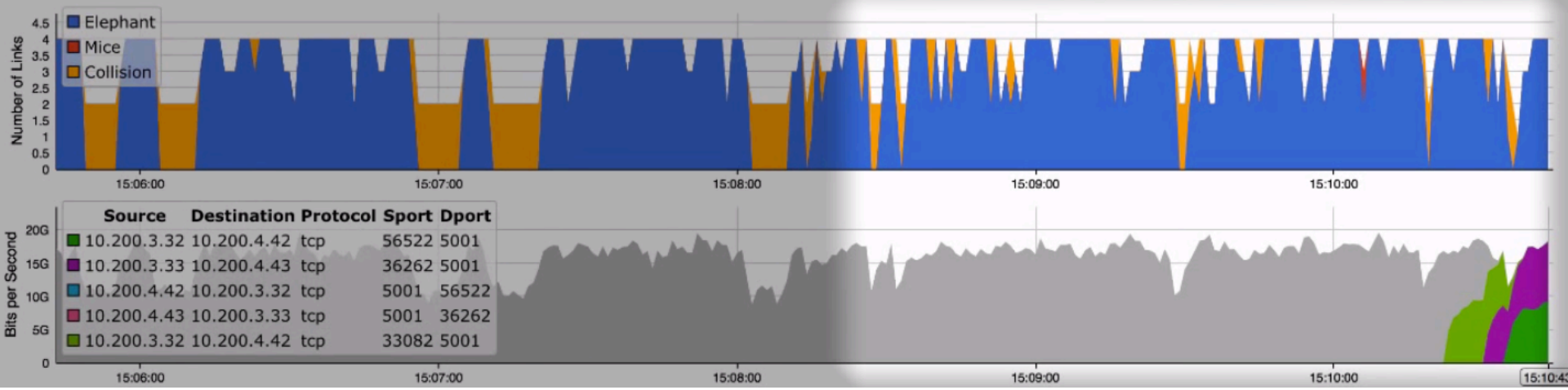


Analytics Driven Traffic Engineering

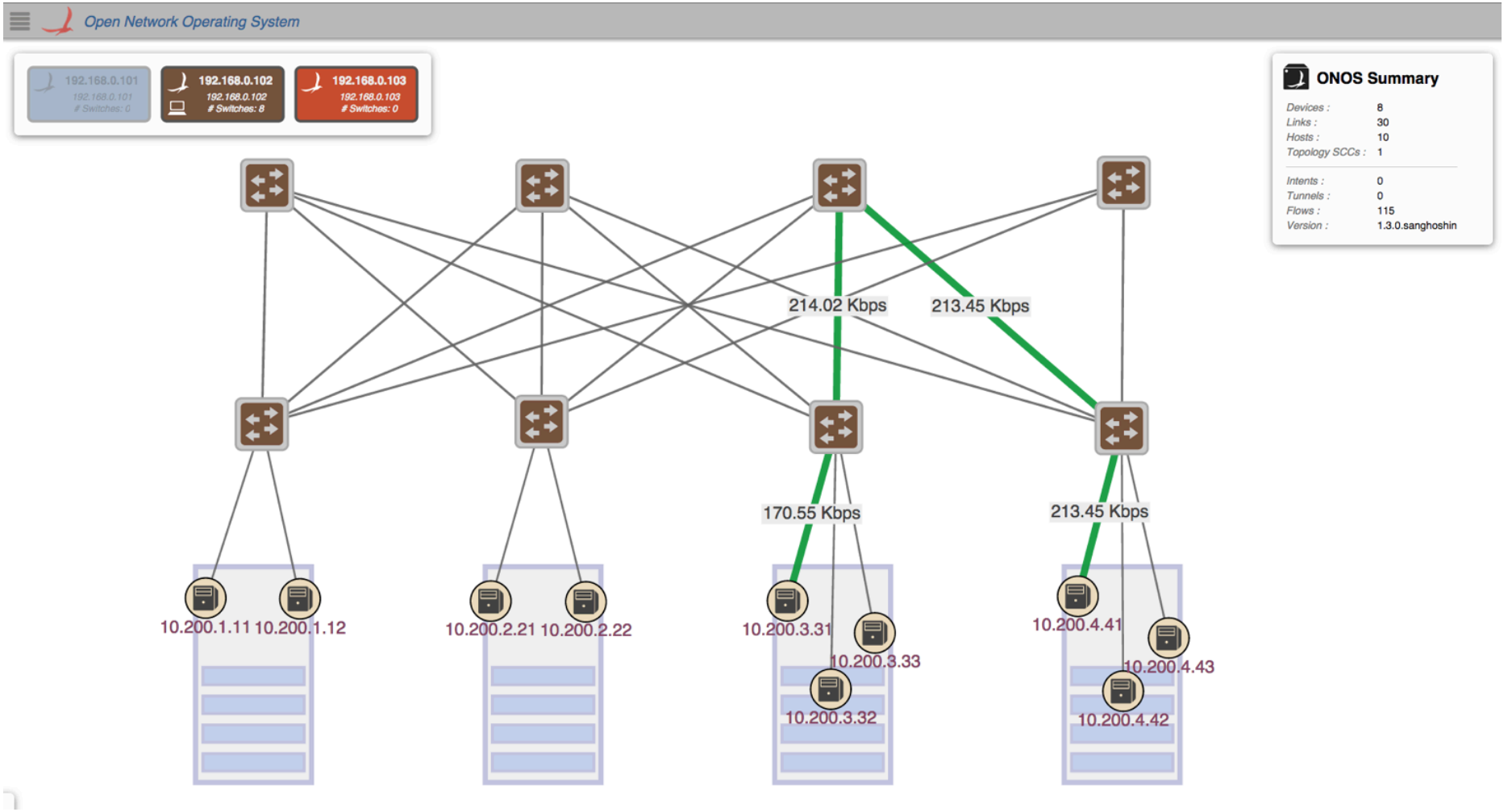


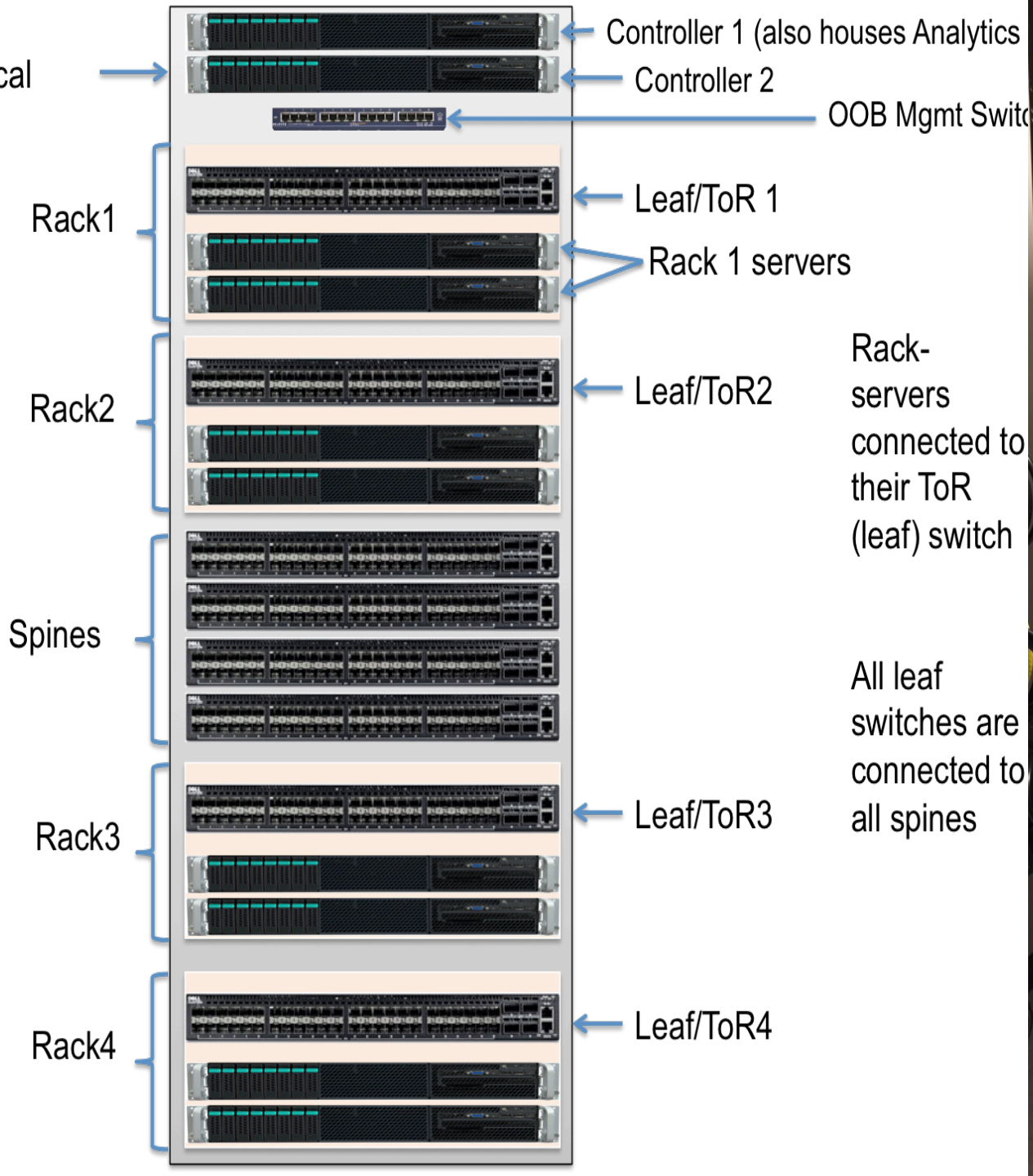
Segment Routing ECMP Override Policies

Source	Destination	Prot	Sport	Dport	Tunnel
10.200.3.33	10.200.4.43	tcp	36262	5001	103,107,104
10.200.3.32	10.200.4.42	tcp	56522	5001	103,106,104

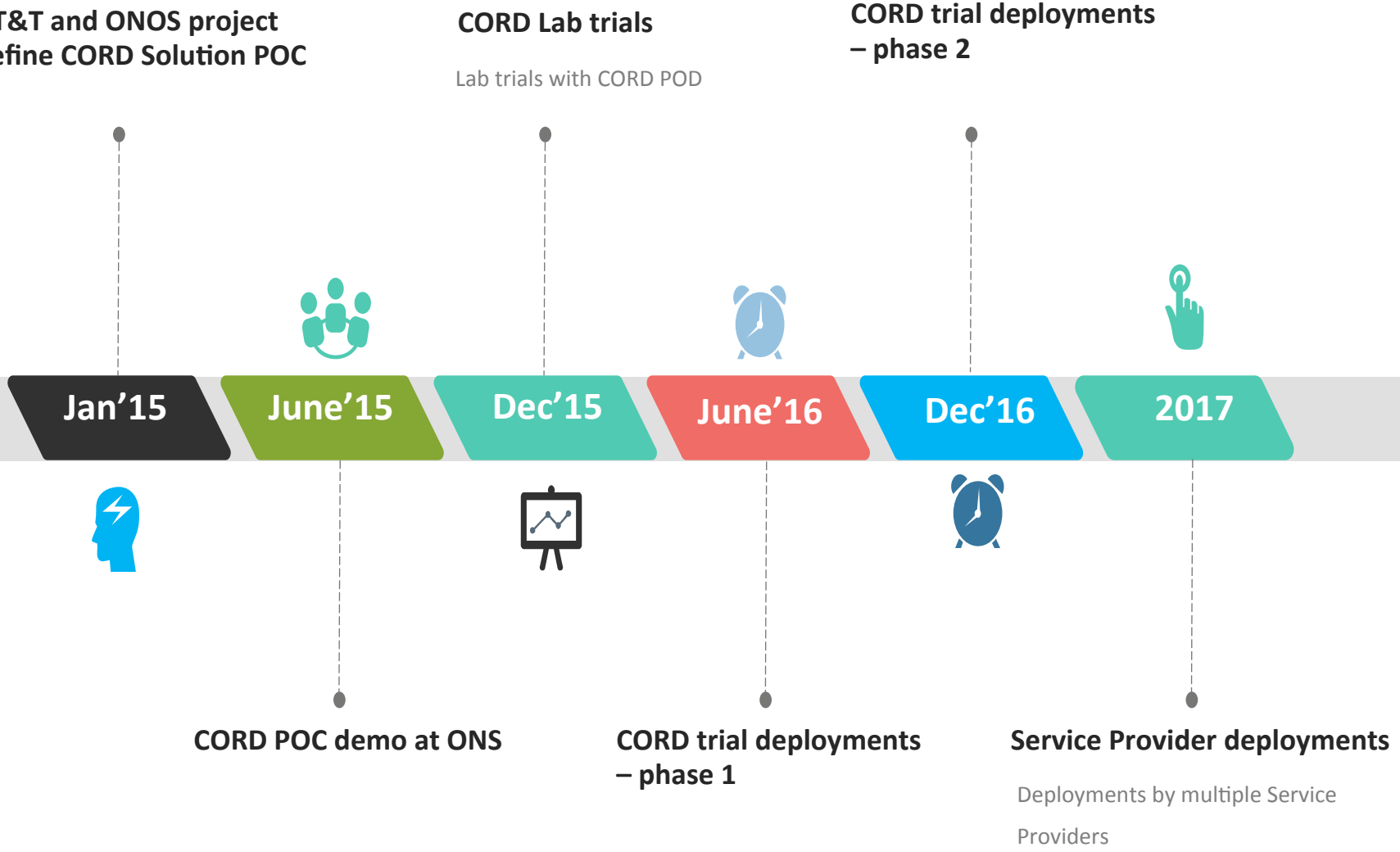


Control Plane Failure Recovery




cal
Controller 1 (also houses Analytics
Controller 2
OOB Mgmt Switc
Rack1
Leaf/ToR 1
Rack 1 servers
Rack2
Leaf/ToR2
Rack-servers connected to their ToR (leaf) switch
Spines
All leaf switches are connected to all spines
Rack3
Leaf/ToR3
Rack4
Leaf/ToR4

CORD Roadmap – From demo to deployment



Note- these timelines are ON.Lab's projections and forward looking

Summary

- CORD Fabric
  - Open-source
  - Spine-leaf architecture: L3 Clos
  - Bare metal hardware
  - SDN based – no use of distributed protocols
  - OF 1.3 multi-tables & ECMP groups
  - ONOS cluster controllers
  - IP/MPLS network using Segment Routing
  - sFlow based analytics for TE of elephant flows

- Next?
  - Integration with vCPE-vOLT-NFaaS
  - Special CORD requirements eg. QinQ
  - Pod based deployment requirements eg. BGP peering
  - Move to open source hardware i.e OCP/ONL/ONIE/OF-DPA