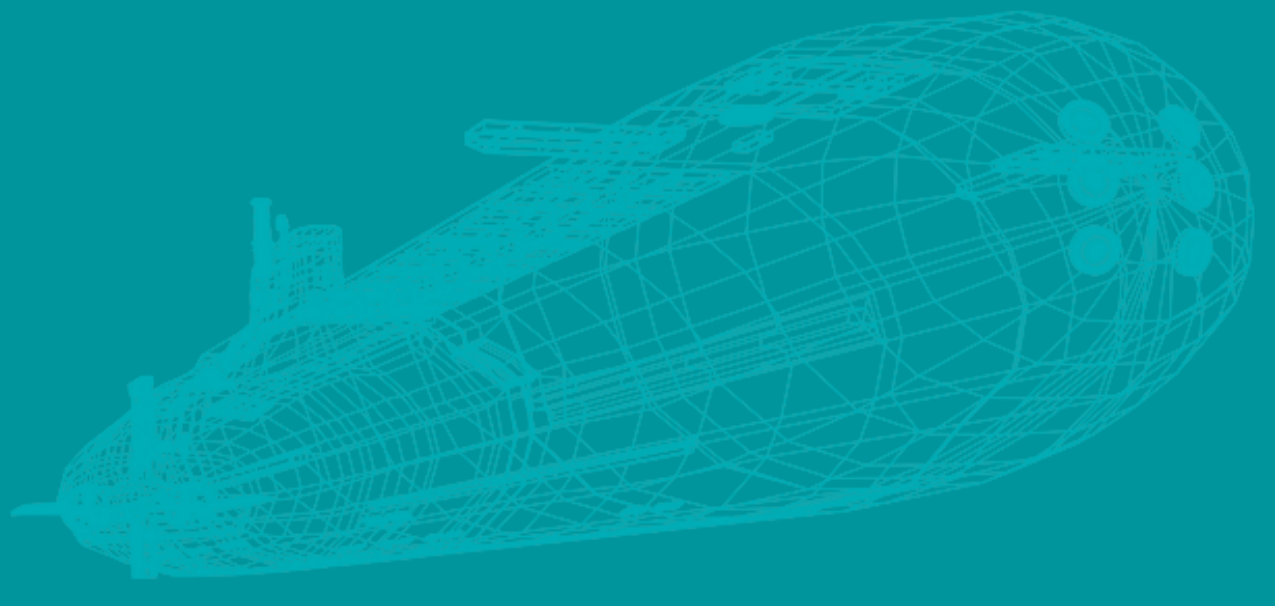
# Election of the Pope, Saint Peter's Squere

ALEF

Trust the Strong

# FabricPath Use Cases

Jiří Cihlář
CCIE#24609
jiri.cihlar@alefnula.com



# Agenda

- Introduction to FabricPath
- Use Case 1 - FabricPath in DC
- Use Case 2 - FabricPath for DCI
- Testing results
- Use Case 3

# FabricPath

# TRILL (Transparent Interconnection of Lots Links)

- IETF standard for Layer 2 multipathing

Drawbacks of STP
- Local STP problems have network-wide impact – low scalability
- Troubleshooting is difficult
- Slower convergence
- STP provides limited bandwidth

Benefits of Routing
- Fast Convergence
- Highly Scalable
- Multi – pathing (ECMP)

Layer 3 routing benefits to Layer 2 bridged Ethernet networks

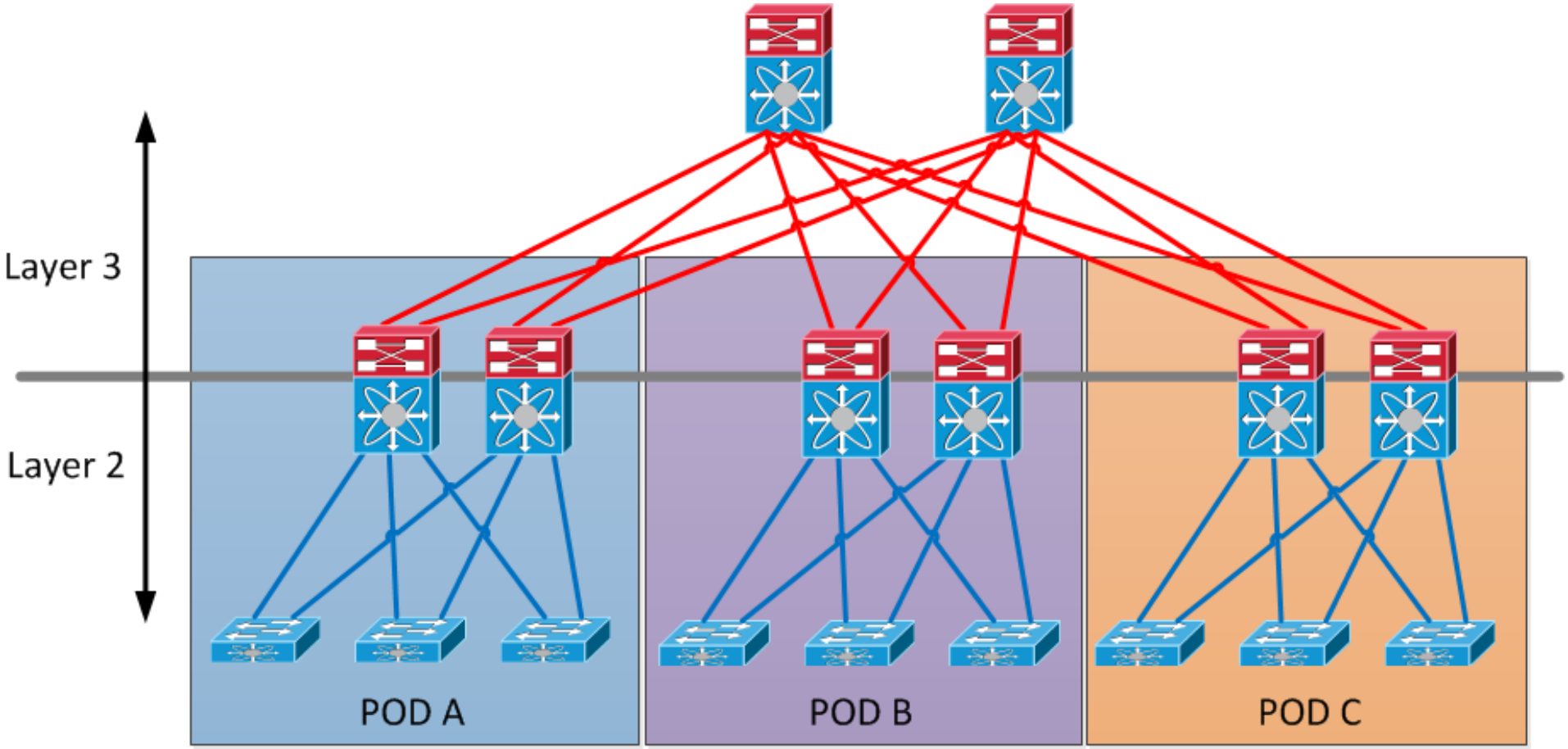
- Driven by multiple vendors, including Cisco
- TRILL now officially moved from Draft to Proposed Standard in IETF
- Proposed Standard status means vendors can confidently begin developing TRILL-compliant software implementations
- Cisco FabricPath is pre-standard implementation of TRILL, available from 2010 on Cisco Nexus switches

# FabricPath vs. TRILL

| | Fabricpath | TRILL |
|---|---|---|
| Frame routing (ECMP, TTL, RPFC etc…) | YES | YES |
| VPC+ | YES | NO |
| FHRP active/active | YES | NO |
| Multiple Topologies | YES | NO |
| Conversational learning | YES | NO |
| Inter-switch links | Point-to-point only | Point-to-point OR shared |
| STP interaction | Fabric is root to CE region, no BPDUs forwarded accross Fabric | CE region part of TRILL fabric, snoop BPDUs for prevent loops |
| Encapsulation | end to end | end to end + hop by hop |

- FabricPath will provide a TRILL mode with a software upgrade (hardware is already TRILL capable)
- Cisco will push FabricPath specific enhancements to TRILL

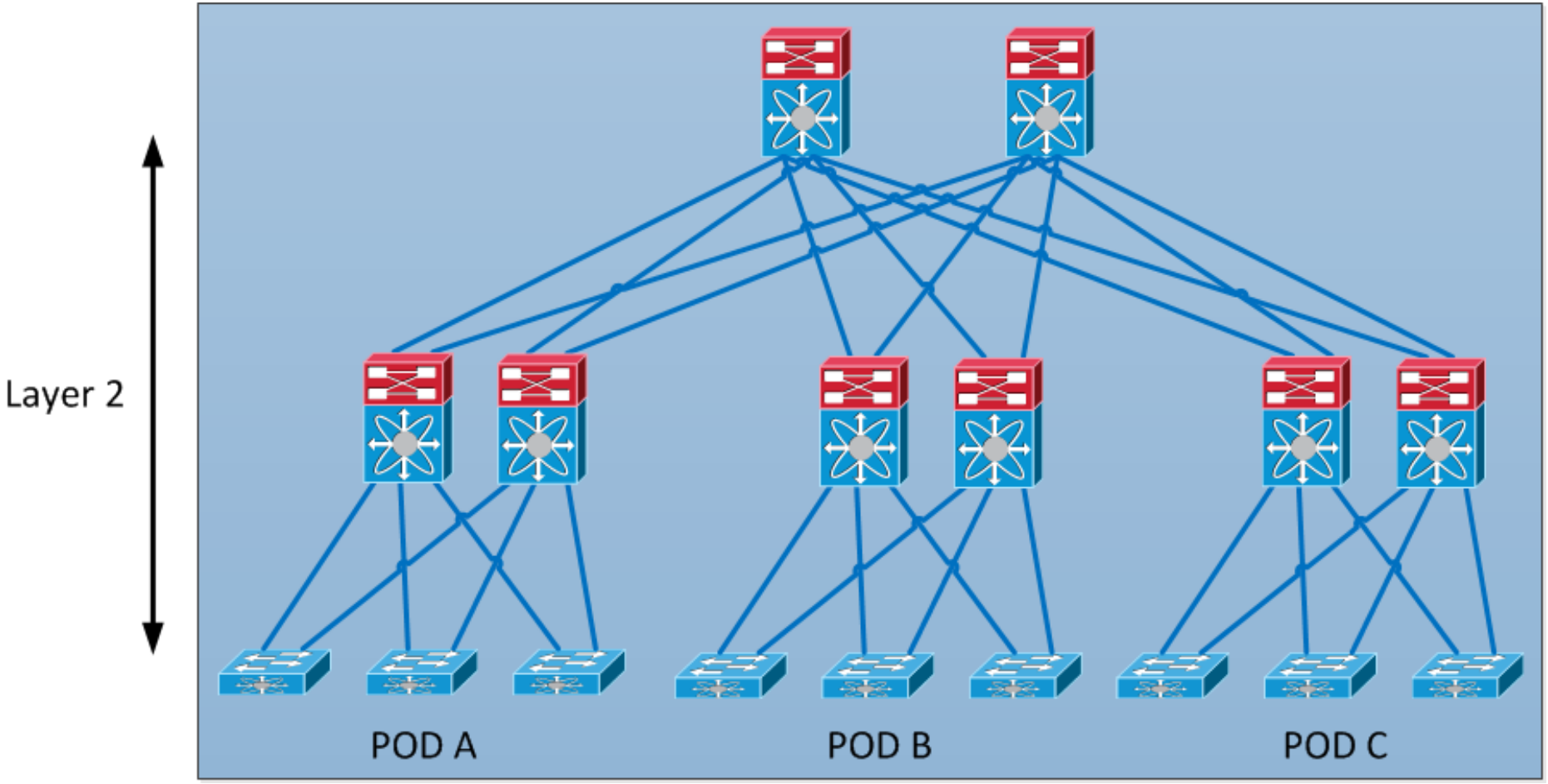
# Data Center Design




L2 limited only to POD

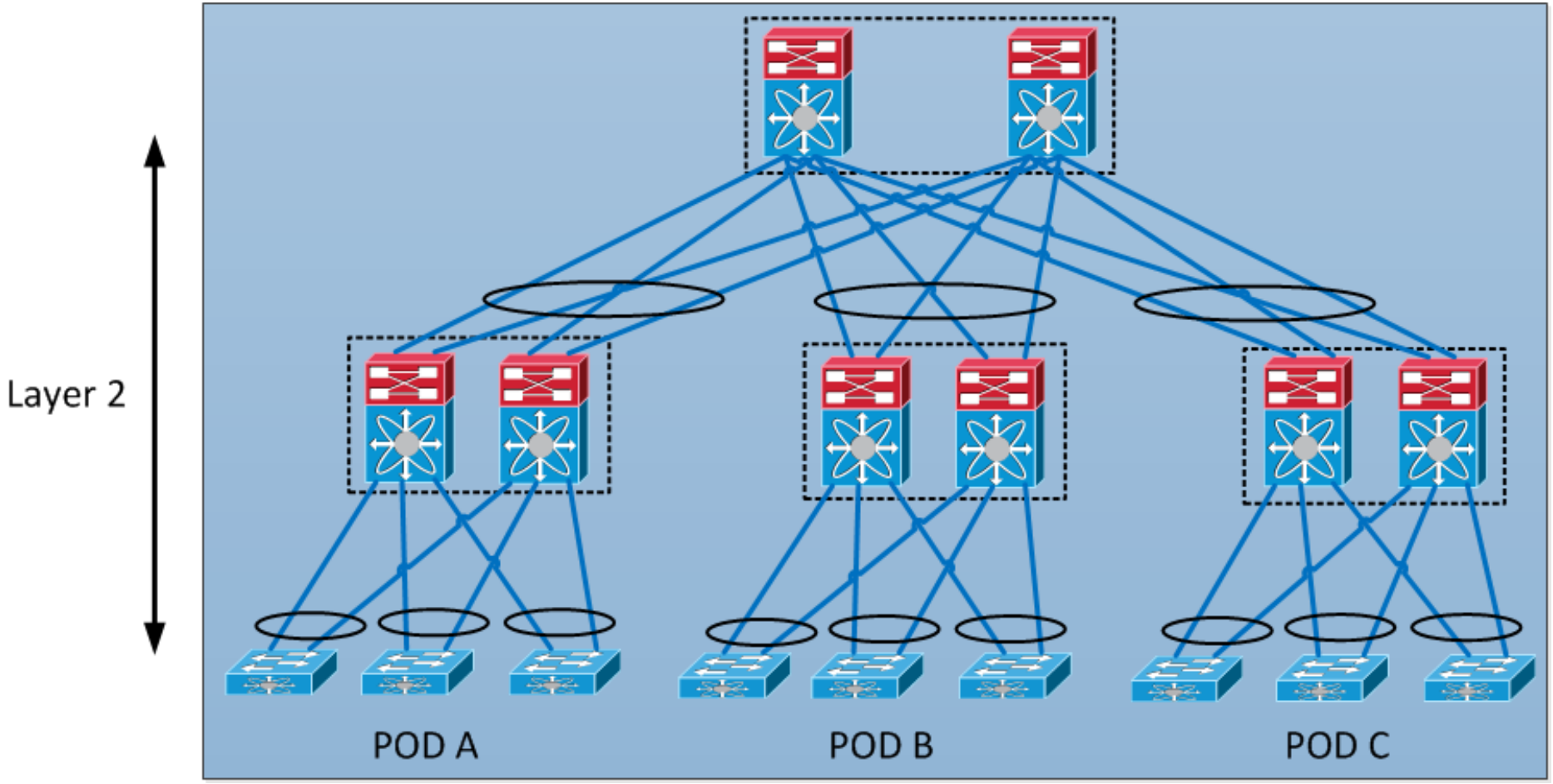
- Requirement to extend L2 to other PODs because of VM mobility, clusters/heart-beat

# Data Center Design



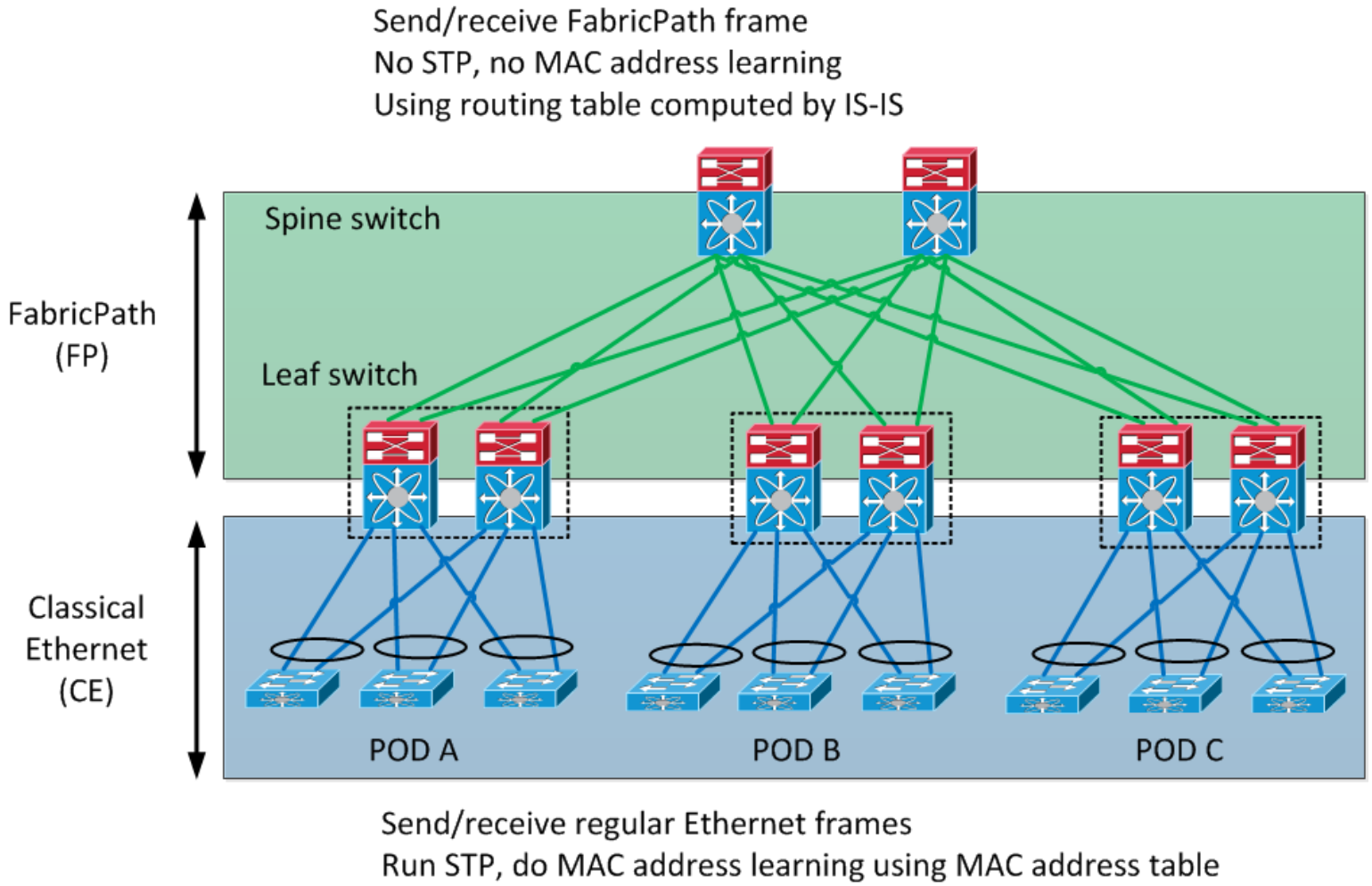

L2 over all PODs with STP just extended to the whole network

# Data Center Design



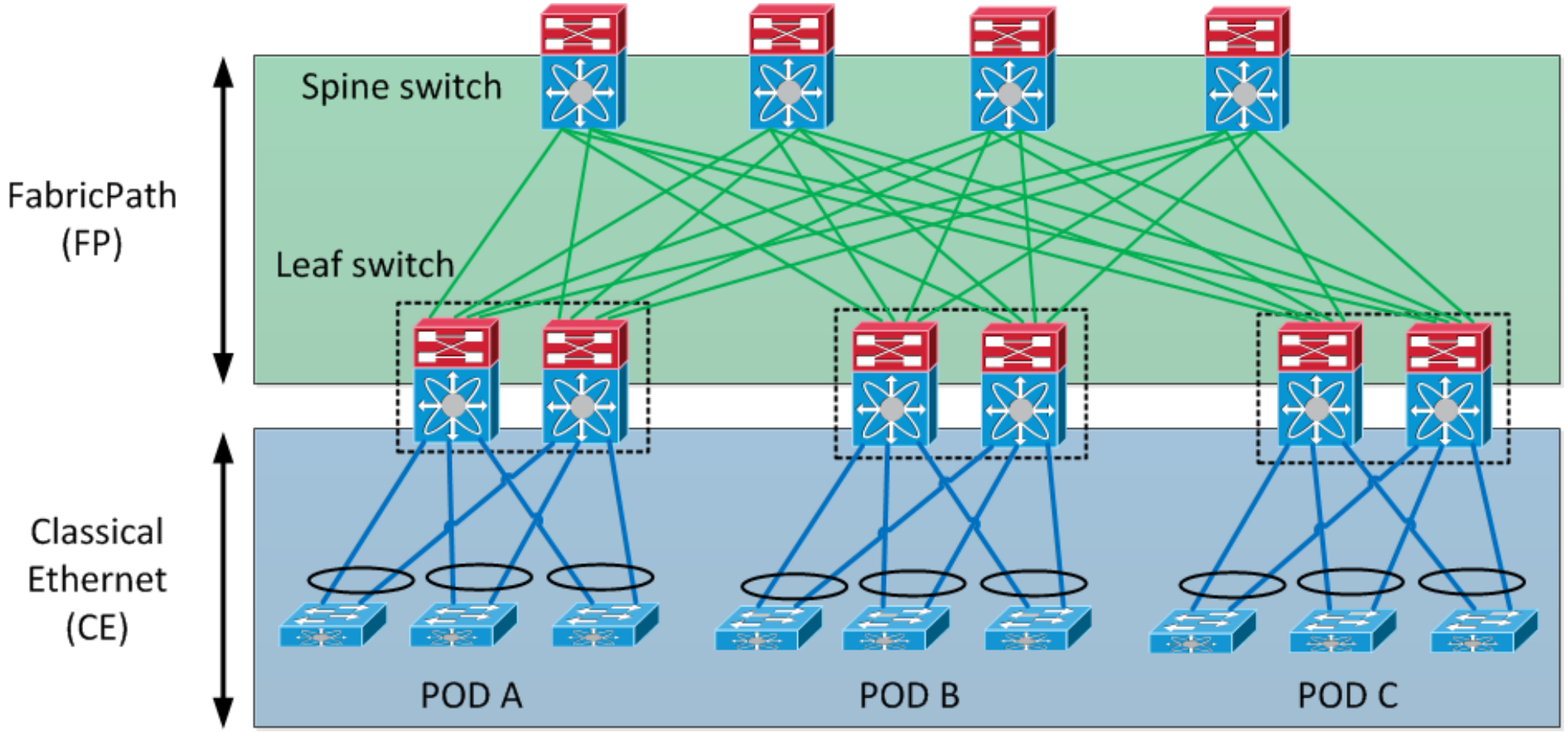

L2 over all PODs with vPC/VSS technology and STP as a last resort

# Use case 1 – FabricPath inside Data Center



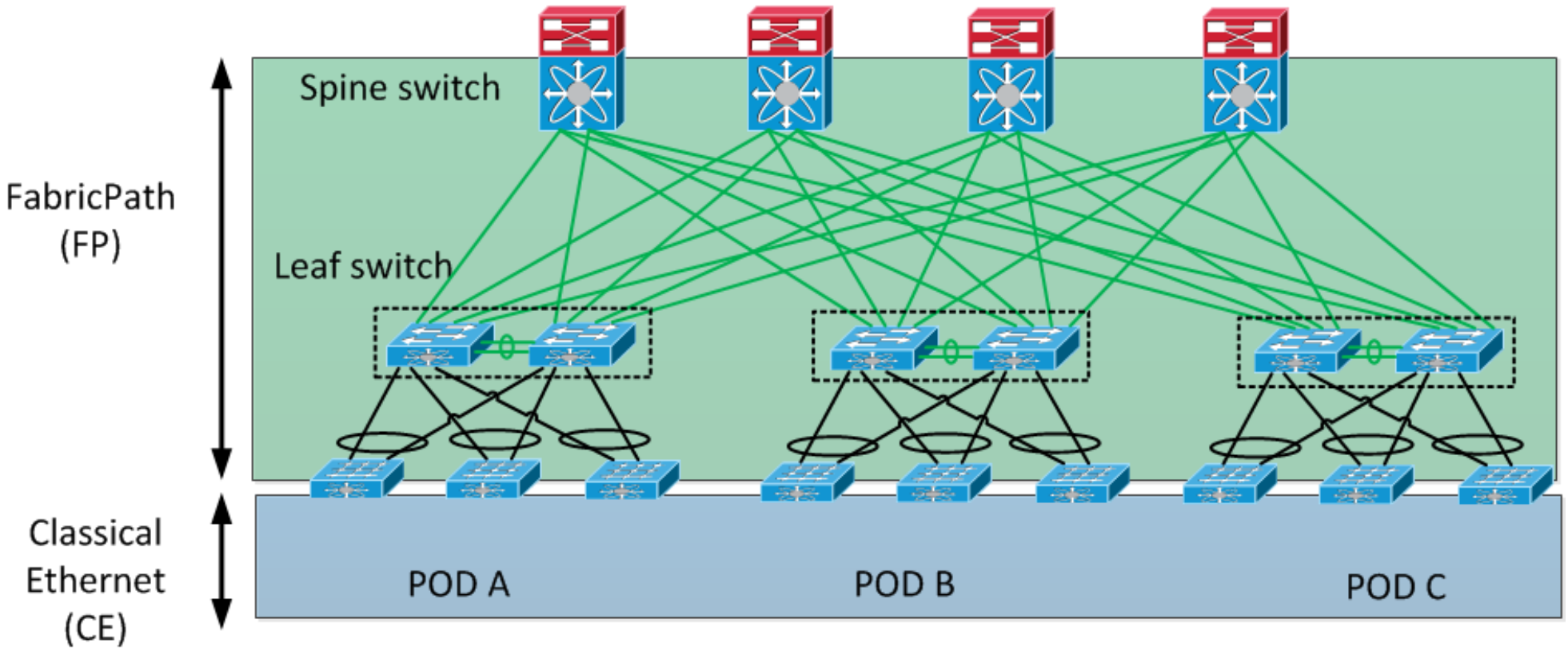

L2 over all PODs with FabricPath

# Use case 1 – FabricPath inside Data Center




Adding Spine switch easily increases the bandwith between PODs

# Use case 1 – FabricPath inside Data Center




Using FEXes we can extend FabricPath to the whole DC

# Data Center Interconnect

# Requirements for DCI

- LAN extension between DC
- Why do the customers request LAN extension?
  - Addressing constraints
  - Allows easy server provisioning
  - Some Applications/Protocols rely on L2
  - **High-availability clusters**
  - **Allows virtual machines mobility**
- Low latency
- Redundancy with fast end-to-end convergence
- Loopfree network
- **Multi Path load balancing**
- **Avoid End-to-End STP, STP isolation between DCs**

# Requirements for DCI

What is used for loop prevention?

Spanning Tree Protocol

Typical limitations of L2 network based on STP:

- Local STP problems have network-wide impact, troubleshooting is difficult
- Flooding impacts the whole network
- STP provides limited bandwidth (no load balancing)
- MAC address tables don’t scale

We can add other requirements:

- avoid STP as much as possible, STP isolation between DCs
- avoid unicast flooding

# How to choose the right DCI design?

Who owns the core infrastructure?

- Enterprise
- ISP

Which type of service is possible over core infrastructure?

- Dark fiber or Layer 1 services – DWDM, CWDM
- Layer 2 services offered by ISP – EoMPLS, VPLS
- Layer 3 services offered by ISP – IP, L3 MPLS VPN

How many datacenters?

- Two DCs
- More than two DCs

How many VLANs, MAC addresses should be exended?

What is the distance between DC?

# Interconnection of two DCs

Two datacenters

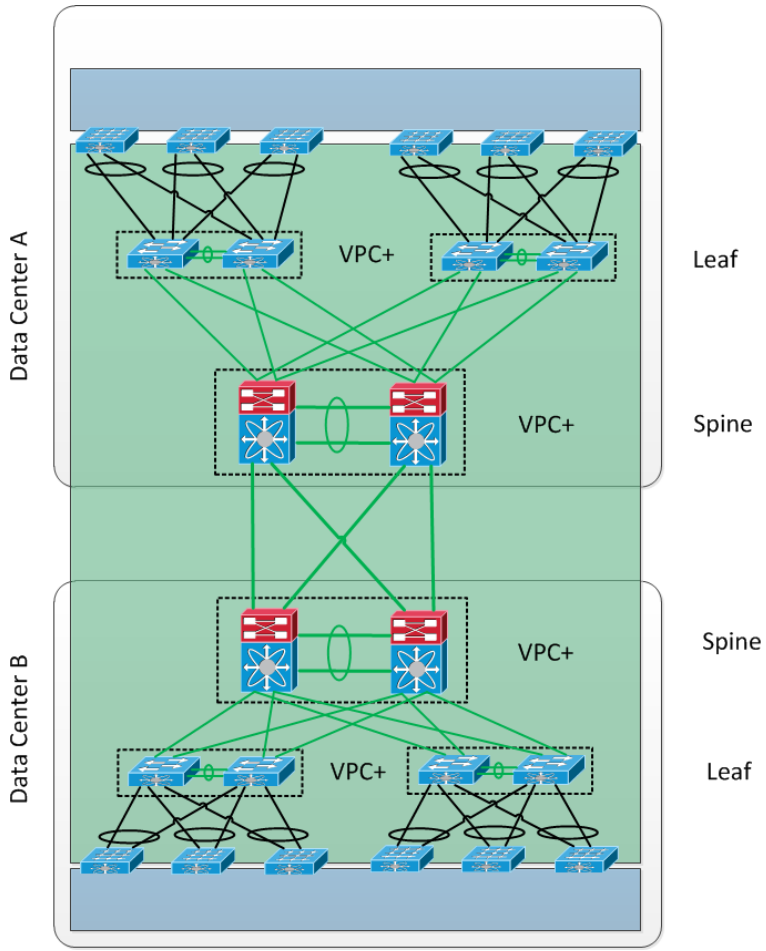
| | Core | Technology used by enterprise | Mechanism for loops prevention |
|---|---|---|---|
| Native Layer 2 | dark fiber or L1 services | **MEC over DWDM based on VSS, VPC** | no loops, STP as a last resort or BPDU filtering |
| L2 over L3 | dark fiber or L3 services | **OTV** | OTV, STP isolation between DCs |
| L2 over L3 | L2 services | **EoMPLS** | STP end-to-end, VSS |
| L2 over L3 | L3 services | **EoMPLSoGRE, L2TPv3** | STP end-to-end, EEM scripts solution, VSS |

More than two DCs

| | Core | Technology used by enterprise | Mechanism for loops prevention |
|---|---|---|---|
| Native Layer 2 | dark fiber or L1 services | **MEC over DWDM based on VSS, VPC** | no loops, STP as a last resort or BPDU filtering |
| L2 over L3 | dark fiber or L3 services | **OTV** | OTV, STP isolation between DCs |
| L2 over L3 | L2 services | **VPLS** | STP end-to-end, EEM scripts, PE clustering = VSS with A-VPLS |
| L2 over L3 | L3 services | **VPLSoGRE** | STP end-to-end, EEM scripts, PE clustering = VSS with A-VPLS |

# Use case 2 - Interconnection of two DCs

| | Core | Technology used by enterprise | Mechanism for loops prevention |
|---|---|---|---|
| Native Layer 2 | dark fiber or L1 services | **FabricPath** | FabricPath, no STP |




Pros:

- No loop in topology
- Multi Path load balancing
- No STP, FabricPath in DCs
- Fast convergence (IS-IS)

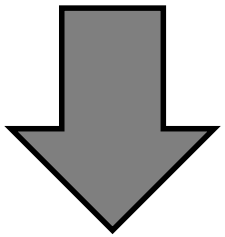
Cons:

- Dark fiber/L1 needed
- Cisco proprietary

# Testing of Fabricpath in DCs

# Requirements

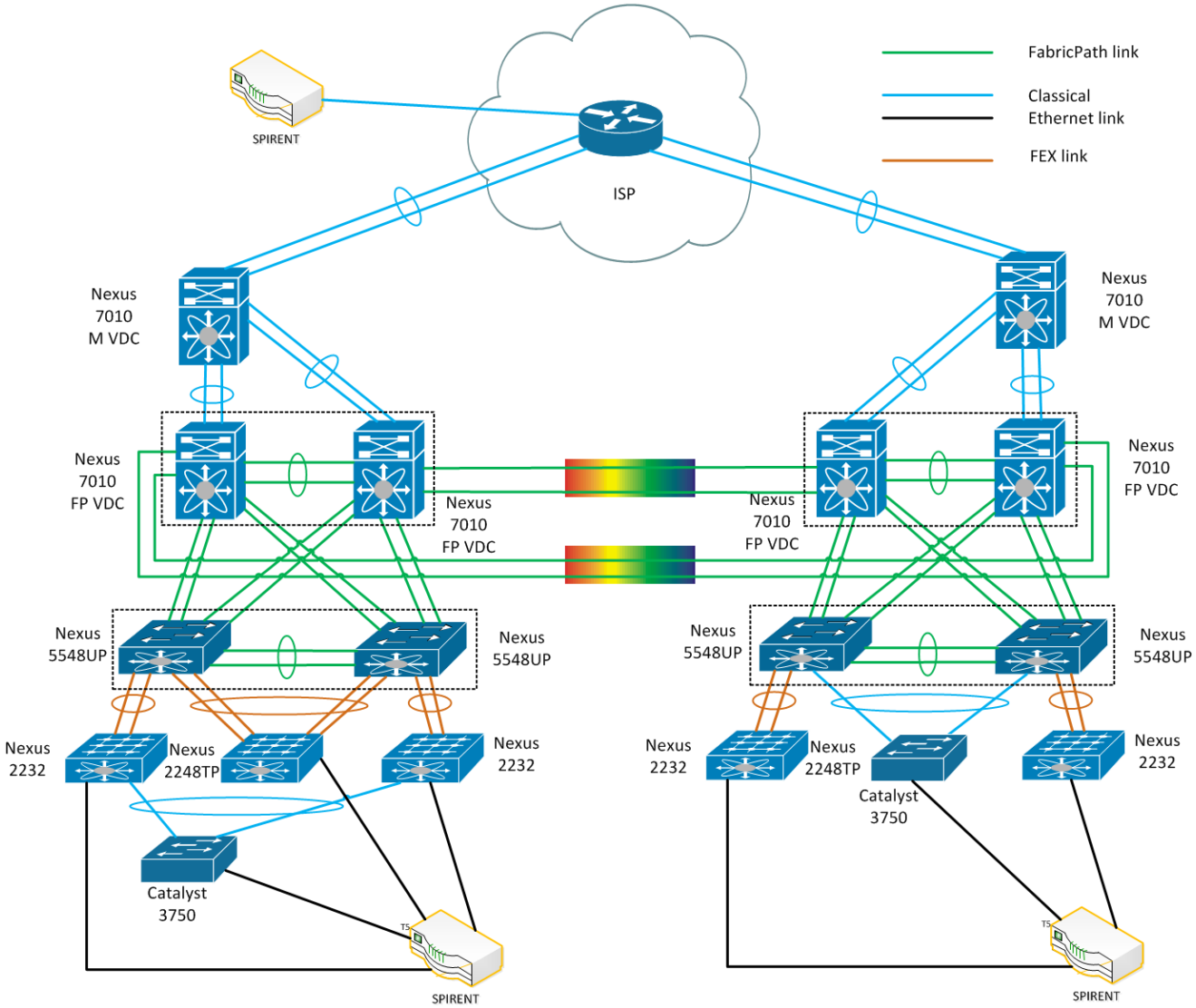
- L2/L3 design verification of new geographically dispersed datacenter  across two sites (distance around 20 km)
- Sites interconneted by dark fibers (DWDM)
- LAN extension between sites
- STP prevention between sites
- IPv4 and IPv6



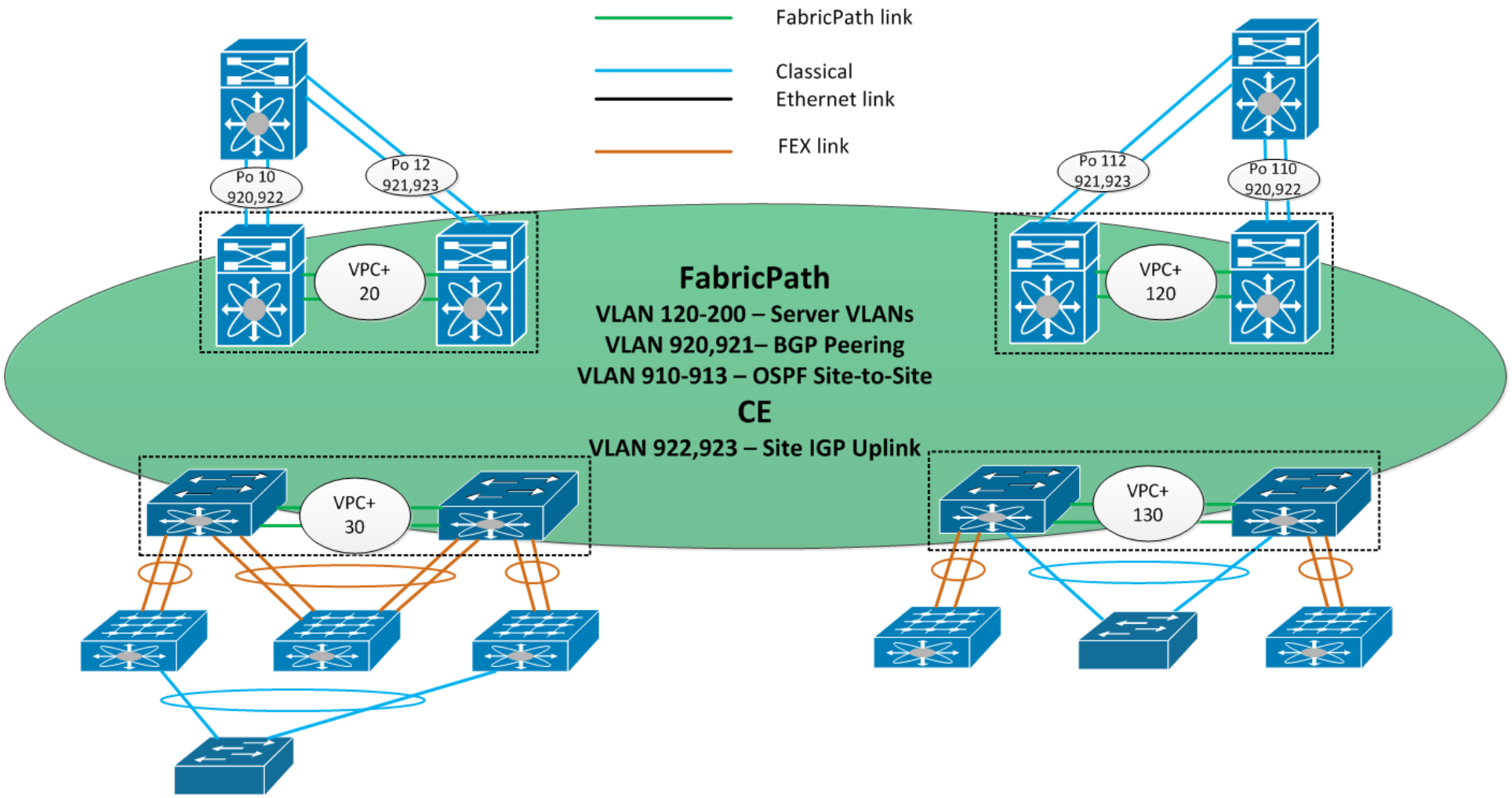
Use case 1 + Use case 2

Design based on Cisco Nexus 7000, 5000, 2000 and Fabricpath

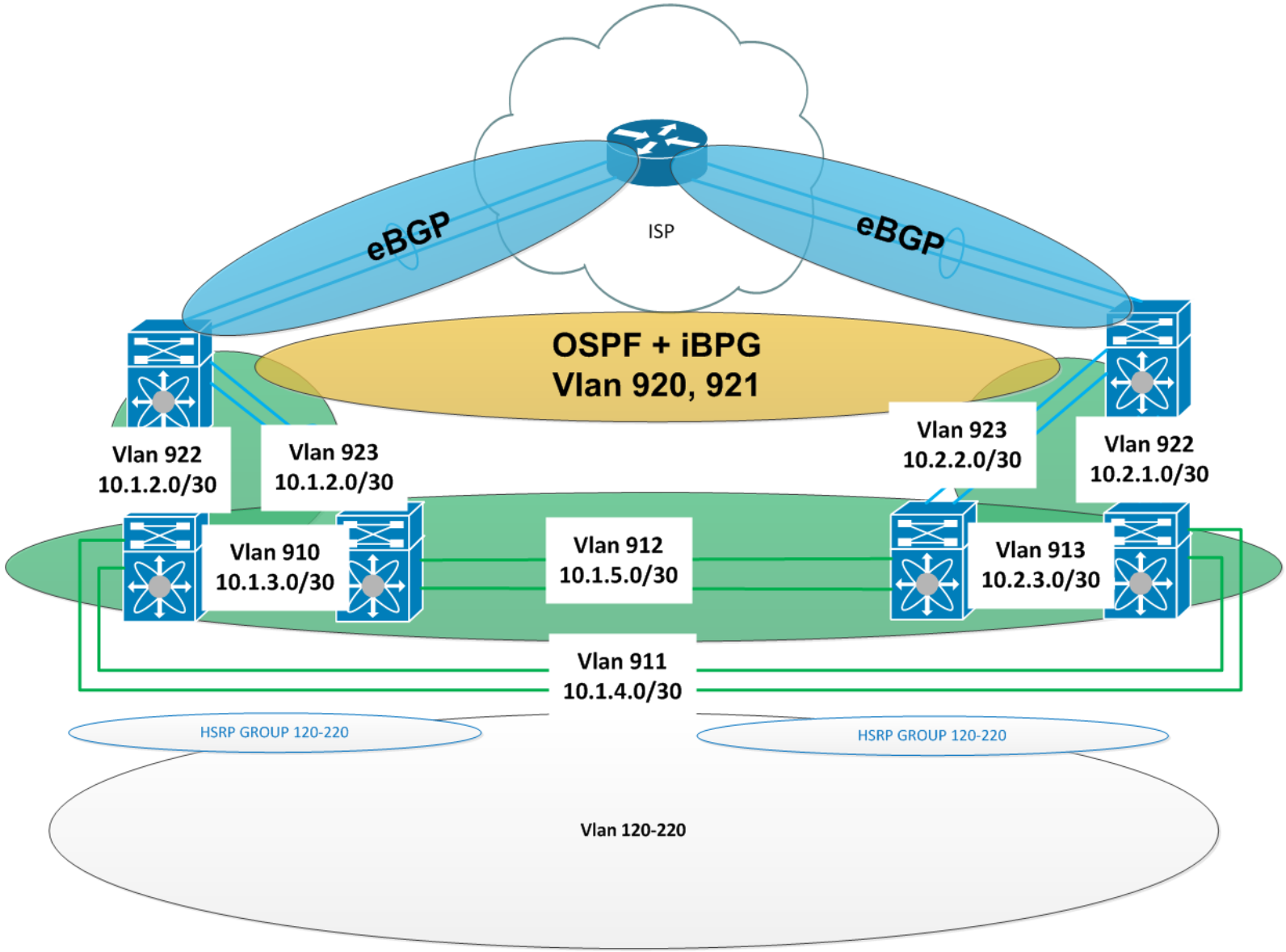
# L1 Topology


FabricPath link
Classical Ethernet link
FEX link
SPIRENT
ISP
Nexus 7010 M VDC
Nexus 7010 M VDC
Nexus 7010 FP VDC
Nexus 7010 FP VDC
Nexus 7010 FP VDC
Nexus 7010 FP VDC
Nexus 5548UP
Nexus 5548UP
Nexus 5548UP
Nexus 5548UP
Nexus 2232
Nexus 2248TP
Nexus 2232
Nexus 2232
Nexus 2248TP
Nexus 2232
Catalyst 3750
Catalyst 3750
SPIRENT
SPIRENT

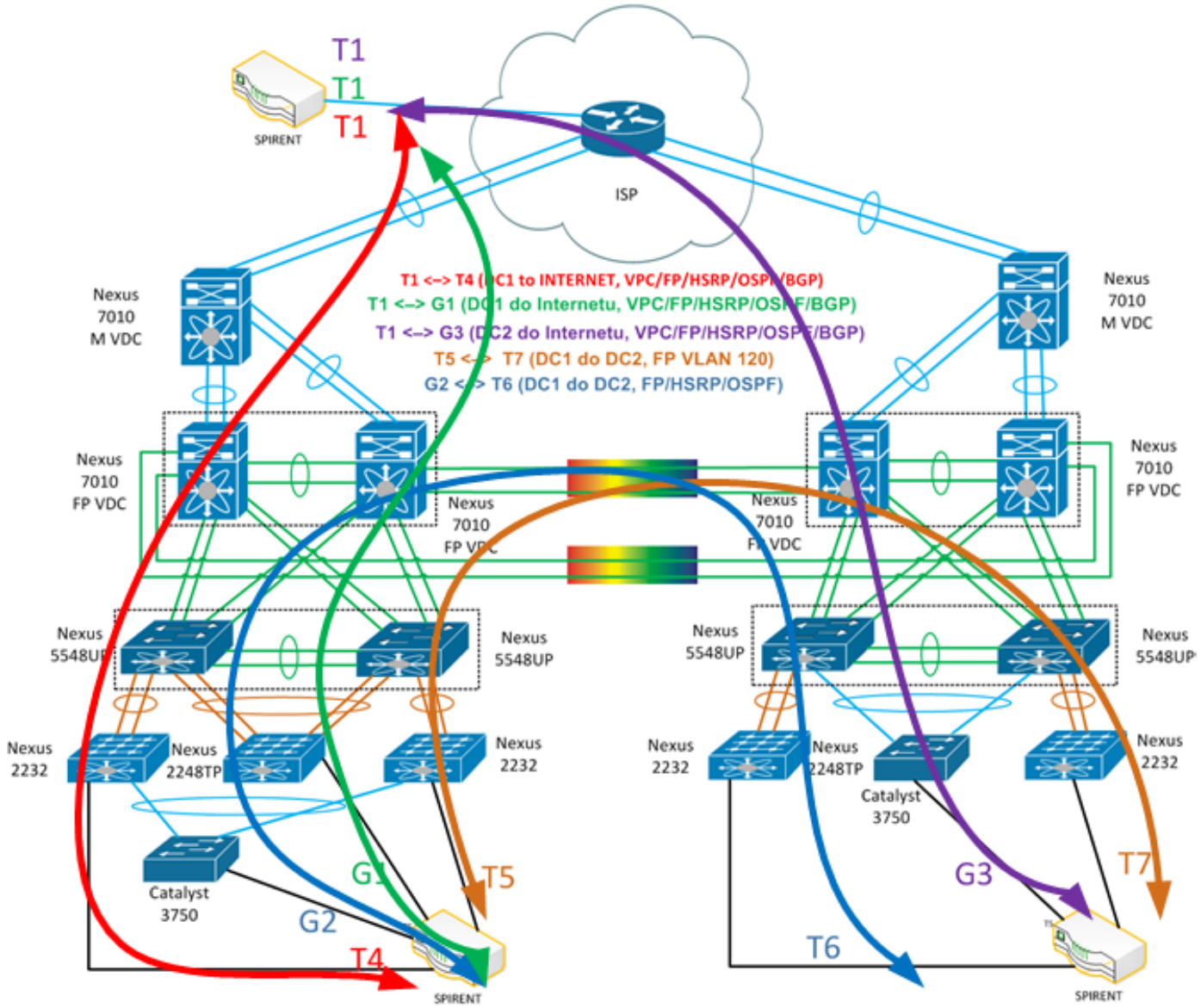
# L2 Topology


FabricPath link
Classical Ethernet link
FEX link
Po 10
920,922
Po 12
921,923
VPC+
20
Po 112
921,923
Po 110
920,922
VPC+
120
FabricPath
VLAN 120-200 – Server VLANs
VLAN 920,921– BGP Peering
VLAN 910-913 – OSPF Site-to-Site
CE
VLAN 922,923 – Site IGP Uplink
VPC+
30
VPC+
130

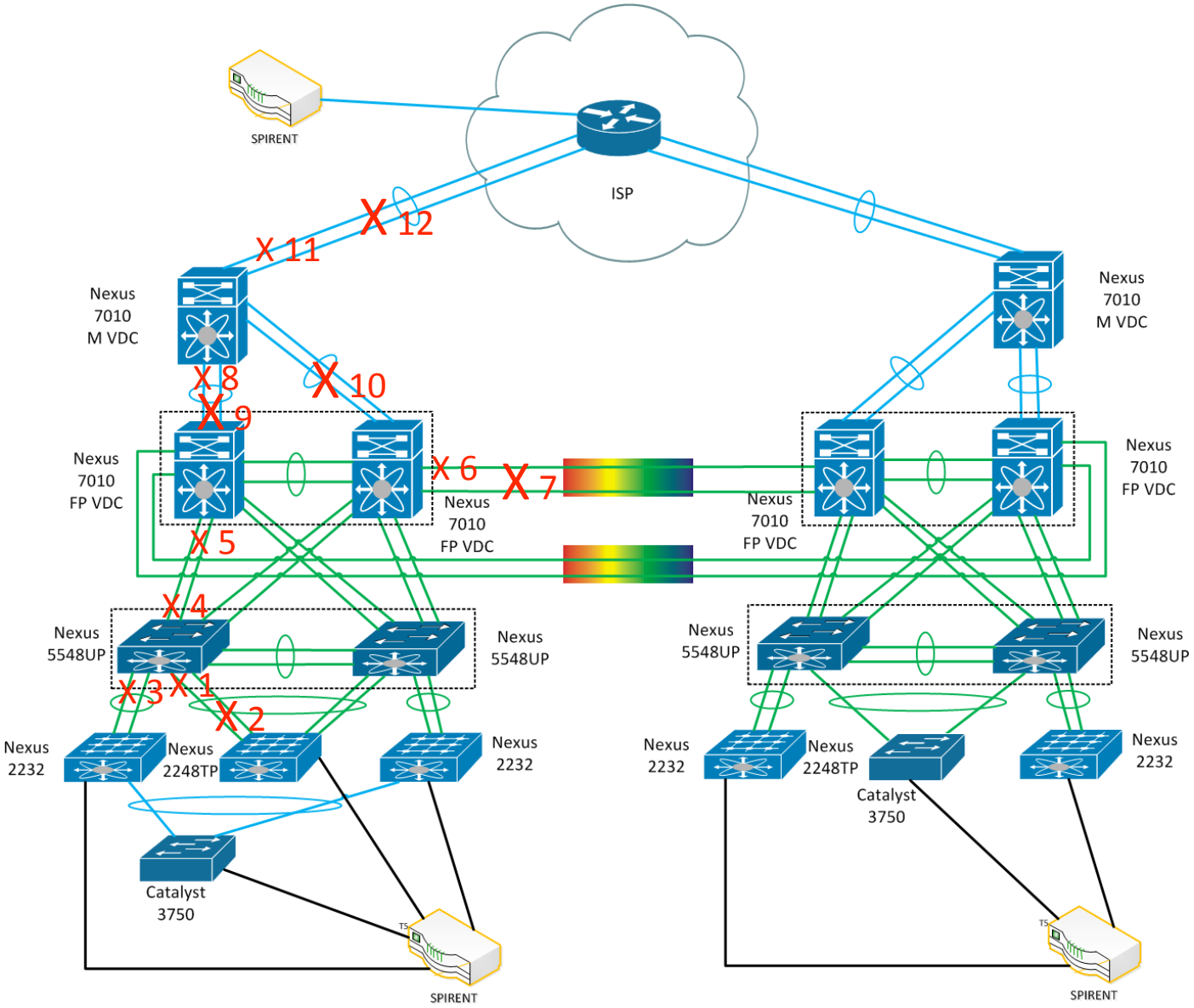
# L3 Topology


ISP
eBGP
eBGP
OSPF + iBPG
Vlan 920, 921
Vlan 922
10.1.2.0/30
Vlan 923
10.1.2.0/30
Vlan 923
10.2.2.0/30
Vlan 922
10.2.1.0/30
Vlan 910
10.1.3.0/30
Vlan 912
10.1.5.0/30
Vlan 913
10.2.3.0/30
Vlan 911
10.1.4.0/30
HSRP GROUP 120-220
HSRP GROUP 120-220
Vlan 120-220

# Streams


T1
T1
T1
SPIRENT
ISP
T1 <–> T4 (DC1 to INTERNET, VPC/FP/HSRP/OSPF/BGP)
T1 <–> G1 (DC1 do Internetu, VPC/FP/HSRP/OSPF/BGP)
T1 <–> G3 (DC2 do Internetu, VPC/FP/HSRP/OSPF/BGP)
T5 <–> T7 (DC1 do DC2, FP VLAN 120)
G2 <–> T6 (DC1 do DC2, FP/HSRP/OSPF)
Nexus 7010 M VDC
Nexus 7010 M VDC
Nexus 7010 FP VDC
Nexus 7010 FP VDC
Nexus 7010 FP VDC
Nexus 7010 FP VDC
Nexus 5548UP
Nexus 5548UP
Nexus 5548UP
Nexus 5548UP
Nexus 2232
Nexus 2248TP
Nexus 2232
Nexus 2232
Nexus 2248TP
Nexus 2232
Catalyst 3750
Catalyst 3750
G1
G2
T4
T5
G3
T6
T7
SPIRENT
SPIRENT

# Failures


SPIRENT
ISP
X 12
X 11
Nexus
7010
M VDC
Nexus
7010
M VDC
X 8
X 10
X 9
Nexus
7010
FP VDC
X 6
X 7
Nexus
7010
FP VDC
Nexus
7010
FP VDC
Nexus
7010
FP VDC
X 5
X 4
Nexus
5548UP
Nexus
5548UP
Nexus
5548UP
Nexus
5548UP
X 3
X 1
X 2
Nexus
2232
Nexus
2248TP
Nexus
2232
Nexus
2232
Nexus
2248TP
Nexus
2232
Catalyst
3750
Catalyst
3750
SPIRENT
SPIRENT

# Convergence tests

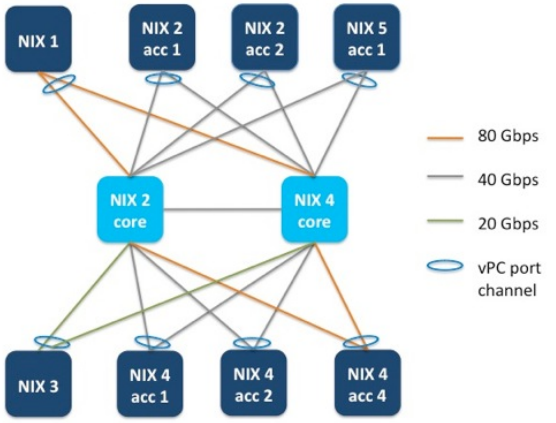
| Failure number/Failure | Failure | Restore |
|---|---|---|
| 1 | 90 ms G1 – T1 | 292 ms G1 – T1 |
| 2 | 418 ms G1 – T1 | 0 |
| 3 | 58 ms T5 – T7 | 0 |
| 4 | 340 ms T1 – T4, T6 – G2 | 0 |
| 5 | 216 ms T1 – T4 | 0 |
| 6 | 236 ms G3 – T1, T6 – G2 | 0 |
| 7 | 171 ms T5 – T7 | 42 ms T4 – T6 |
| 8 | 4 ms T1 – T4 | 0 |
| 9 | 597 ms G1 – T1 | 0 |
| 10 | 222 ms G1 – T1 | 0 |
| 11 | 3 ms T1 – T4 | 0 |
| 12 | 140 ms T1 – T4, G1 – T1 | 0 |
| Chassis Pwr-Off | 2s N7k, 680 N5k, 0 N2k | 4s* N7K, 173 N5K, 417 N2K |
| Fabric release N7K | 0 | 0 |
| Sup release N7K, VDC FP | 1,3s G2-T6 | 0 |

* Only for IPv4 traffic, for IPv6 traffic required max-metric router-lsa for ospfv3 in NX-OS 6.2, BFD for OSPFv3 and FP

# Use case 3

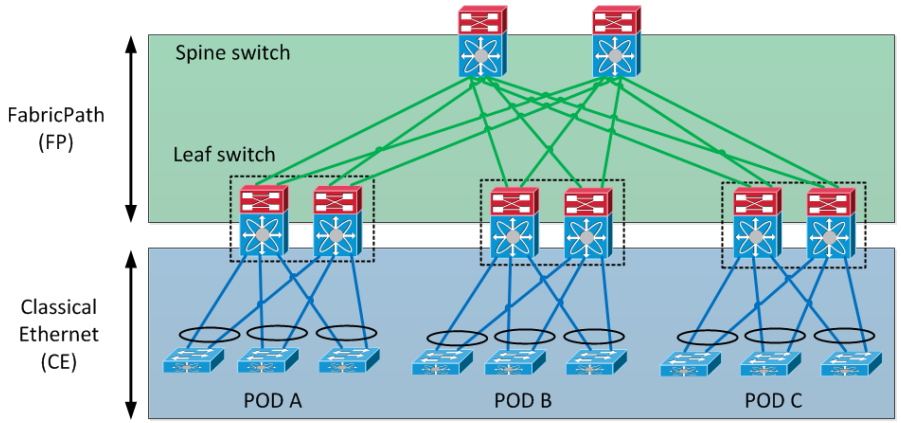
# Use case 3 - Internet eXchange

## Current NIX topology

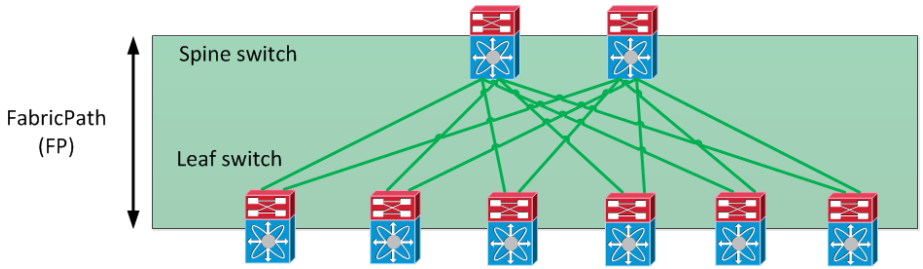



2.4.2013

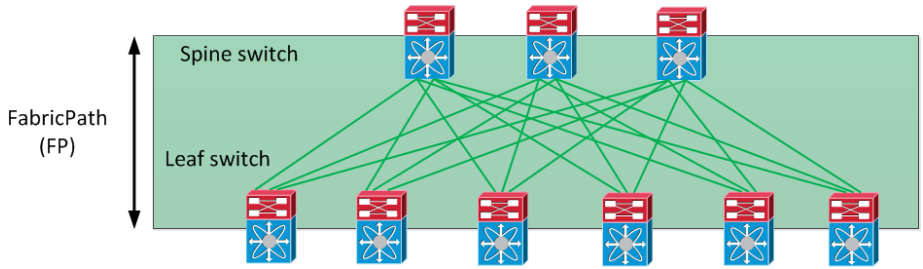
## Have you seen this before?




## NIX design with FabricPath




## Increasing the bandwith and redundancy

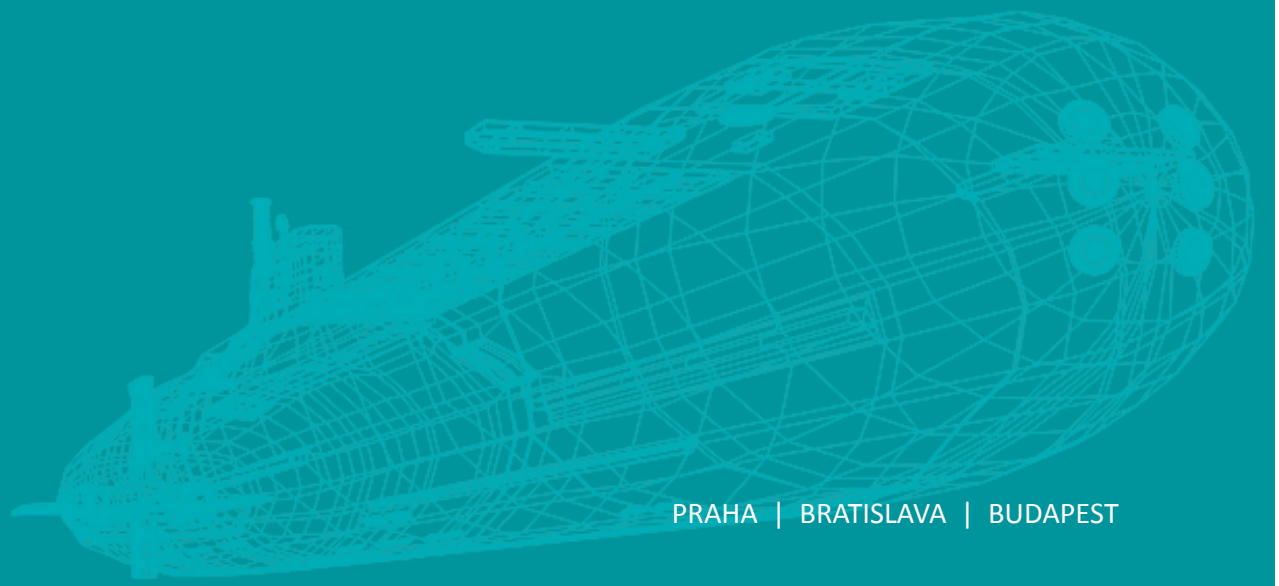



FabricPath can be suitable technology for Internet eXchange

ALEF

Trust the Strong

# Thank you for your attention



www.alef.com

PRAHA | BRATISLAVA | BUDAPEST