

Juniper QFXで実践！ EVPN/VXLANの実力値

日商エレクトロニクス株式会社

目次

1. EVPN/VXLANの概要
2. 検証で見えてきたEVPN/VXLANの実力値
3. EVPN/VXLANを構成するJuniper製品

1. EVPN/VXLANの概要

EVPN/VXLANについて

❖ EVPN/VXLANとは

L3ネットワーク上にオーバーレイでL2ネットワークを構築するL2VPN技術

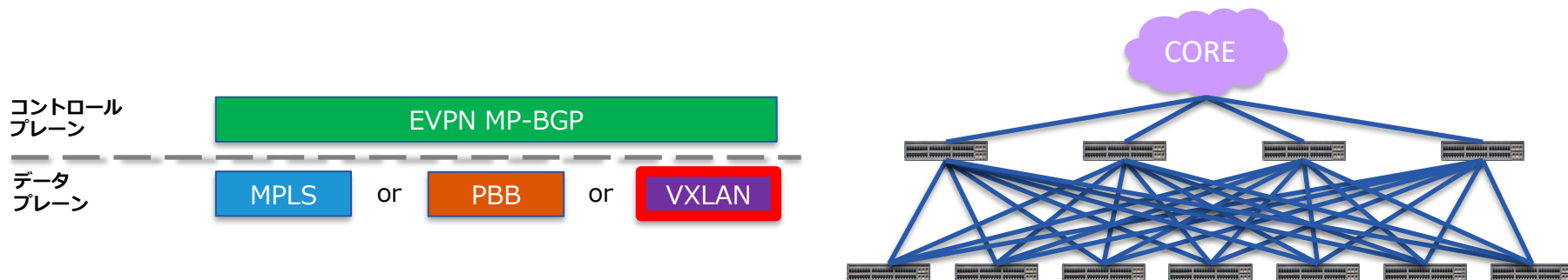
コントロールプレーンとデータプレーンより構成されており、

コントロールプレーンではBGPを利用しテナントごとのMAC/IP情報の取り交わしを実施

データプレーンでは任意のプロトコルを用いてパケットのカプセル化を実施

その両者の組み合わせによって、仮想的なL2/L3ネットワークを実現

昨今ではデータプレーンのプロトコルにおいてVXLANを用いるEVPN/VXLAN構成が主流



EVPN/VXLANの活用シーン

- ❖ EVPN/VXLANは、主にIP CLOS構成のDCネットワークを前提として様々な用途での効果が期待できる

【EVPN/VXLANの用途例】

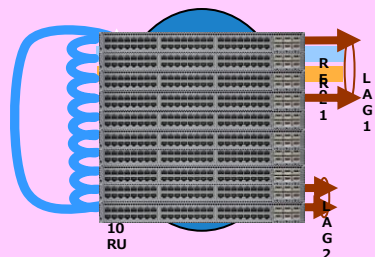
- ① VC構成、VCF構成からの置き換え
- ② DC内ネットワークのループフリー化
- ③ DC間におけるL2延伸

ユースケース① VC構成、VCF構成からの置き換え

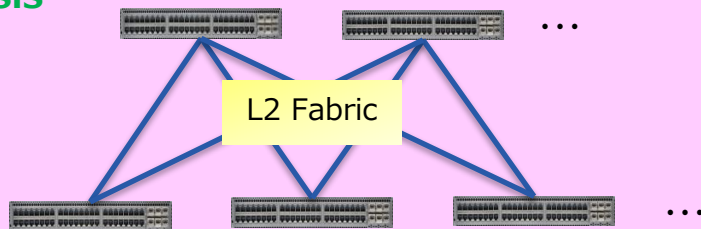
- ❖ これまでのL2ファブリック構成では台数やトポロジーの制約あり
EVPN/VXLANへの移行により、台数・トポロジーの制約から解放

Legacy Solution

Virtual Chassis
(最大10台)

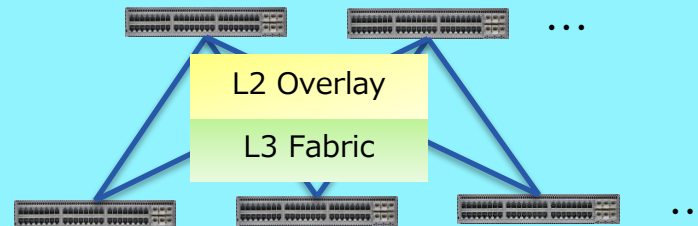


Virtual Chassis Fabric
(最大20台)



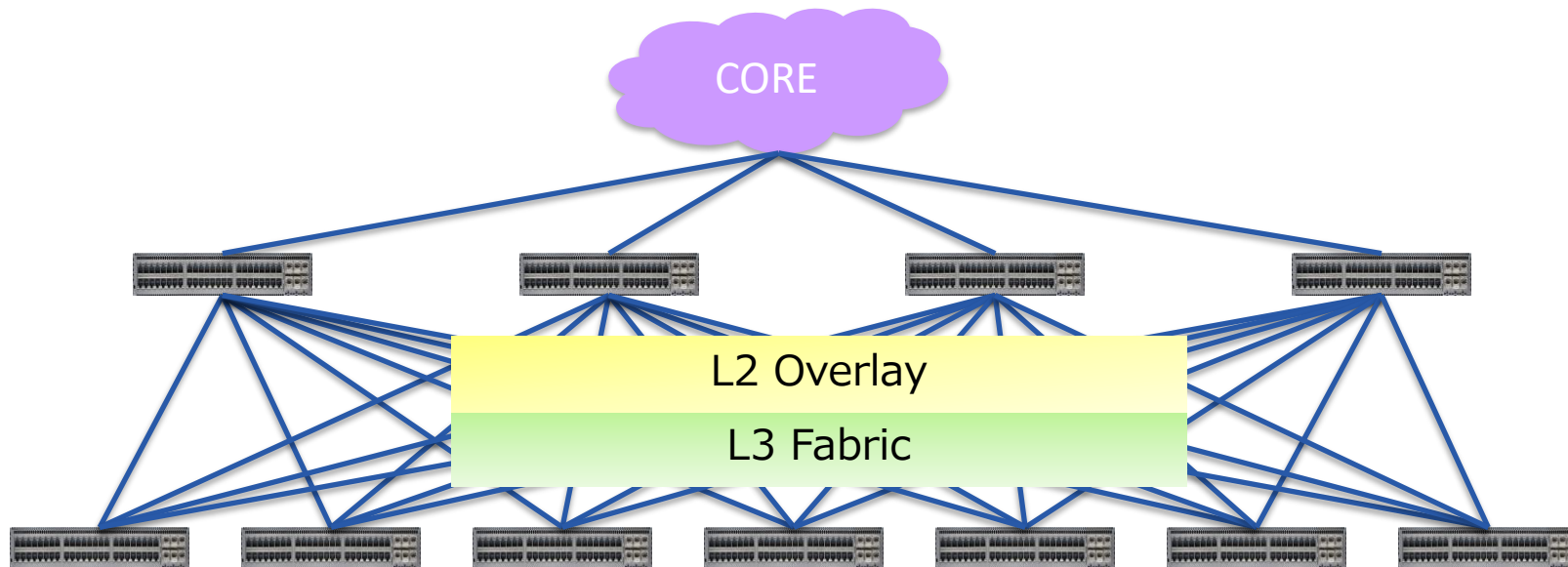
Newer Solution

EVPN/VXLAN
(台数制限なし)



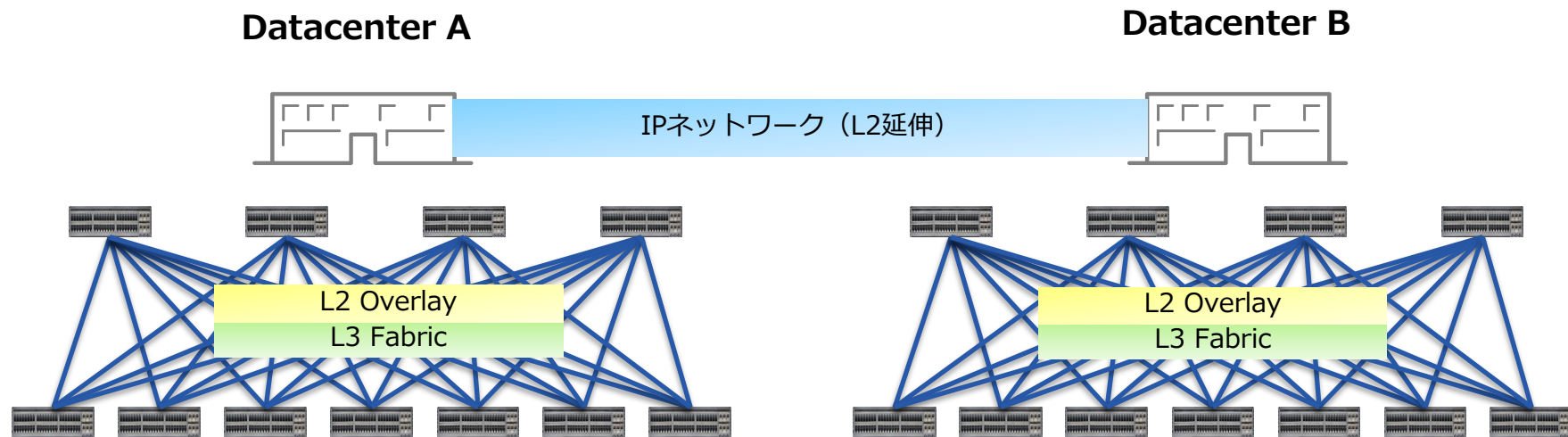
ユースケース② DC内ネットワークのループフリー化

- ❖ OSPFやBGPなどを用いる以上、L3ファブリックはループの問題を内包
VXLANのOverlay構成で、Underlayに依存しないL2フラットなNWへ



ユースケース③ DC間におけるL2延伸

- ❖ 複数DC間の接続も、IPネットワーク上にトンネルを張りL2延伸でフラットなデータ転送を実現



JUNOS 20.4におけるEVPN/VXLAN機能のアップデート

❖ 複数のVLAN-Serviceに対応

従来から対応していたVLAN-Awareのほか、**VLAN-Based,VLAN-Bundleに対応**
使用環境に応じ、柔軟な選択が可能に

❖ EVI設定可能数の拡張

従来QFXではEVI (EVPN Instance) を一つしか作れず、一つのインスタンスの中で
広報するVLAN(VXLAN)を指定する形式であったが、MAC-VRFのサポートによって
EVIを複数作成できるようになり、**EVI単位でのユーザ分割案をサポート**

MAC-VRF形式コンフィグのサポート

❖ JUNOS 20.4以降では、従来形式とMAC-VRF形式の2つのコンフィグ形式をサポート（※混在不可）

従来形式

```
set protocols bgp group ibgp type internal
set protocols bgp group ibgp local-address 10.0.0.6
set protocols bgp group ibgp family evpn signaling
set protocols bgp group ibgp neighbor 10.0.0.1
set protocols bgp group ibgp neighbor 10.0.0.2
set protocols evpn encapsulation vxlan
set protocols evpn extended-vni-list 10002
set protocols evpn extended-vni-list 10003
set routing-options router-id 10.0.0.6
set routing-options autonomous-system 65000
set vlans VLAN0002 vlan-id 2
set vlans VLAN0002 vxlan vni 10002
set vlans VLAN0003 vlan-id 3
set vlans VLAN0003 vxlan vni 10003
set switch-options vtep-source-interface lo0.0
set switch-options route-distinguisher 10.0.0.6:1
set switch-options vrf-target vrf-target target:65000:1
```

MAC-VRF形式

```
set protocols bgp group ibgp type internal
set protocols bgp group ibgp local-address 10.0.0.6
set protocols bgp group ibgp family evpn signaling
set protocols bgp group ibgp neighbor 10.0.0.1
set protocols bgp group ibgp neighbor 10.0.0.2
set forwarding-options evpn-vxlan shared-tunnels
set routing-options router-id 10.0.0.6
set routing-options autonomous-system 65000
set routing-instances EVPN-1 protocols evpn encapsulation vxlan
set routing-instances EVPN-1 vtep-source-interface lo0.0
set routing-instances EVPN-1 instance-type mac-vrf
set routing-instances EVPN-1 service-type vlan-aware
set routing-instances EVPN-1 interface ae0.0
set routing-instances EVPN-1 route-distinguisher 10.0.0.6:1
set routing-instances EVPN-1 vrf-target target:65000:1
set routing-instances EVPN-1 vlans VLAN0002 vlan-id 2
set routing-instances EVPN-1 vlans VLAN0002 vxlan vni 10002
set routing-instances EVPN-1 vlans VLAN0003 vlan-id 3
set routing-instances EVPN-1 vlans VLAN0003 vxlan vni 10003
```

2. 検証で見えてきた EVPN/VXLANの実力値

❖ Juniper QFXシリーズによるEVPN/VXLAN検証

20.4R1でEVPN/VXLANを構成した際のスケーラビリティ
及び障害時の通信影響を確認する検証を実施

1 MACスケーラビリティ検証

Juniper QFXでEVPN/VXLANを構成した際のMACスケーリングを確認

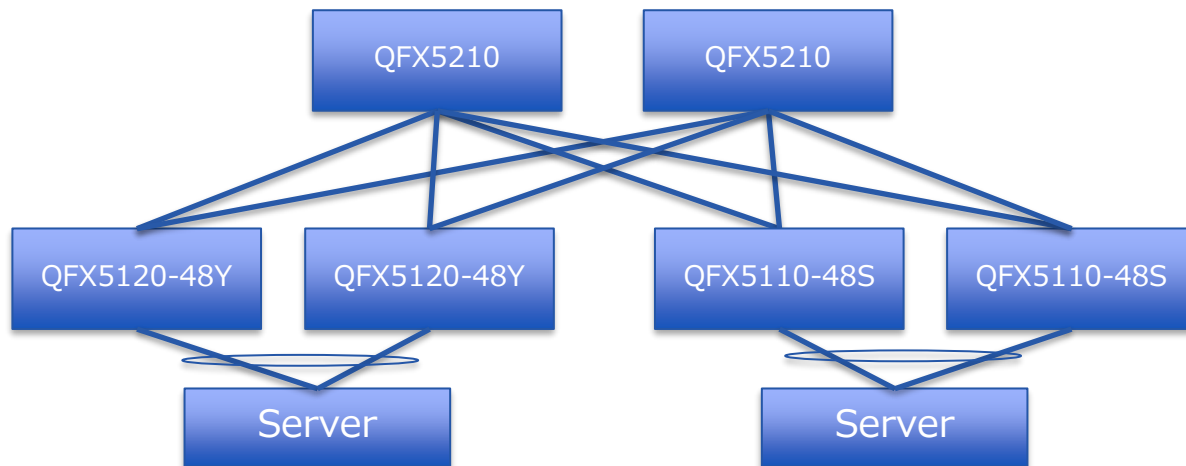
2 障害検証

EVPN/VXLAN網でリンク・ノード障害が発生した際の通信影響を確認

検証構成

❖ テスト環境の構成条件

- Juniper QFX5000シリーズを使用してEVPN/VXLAN網を構成
- VTEPにはL3Gateway機能を持たせず、L2延伸のみの設定を投入

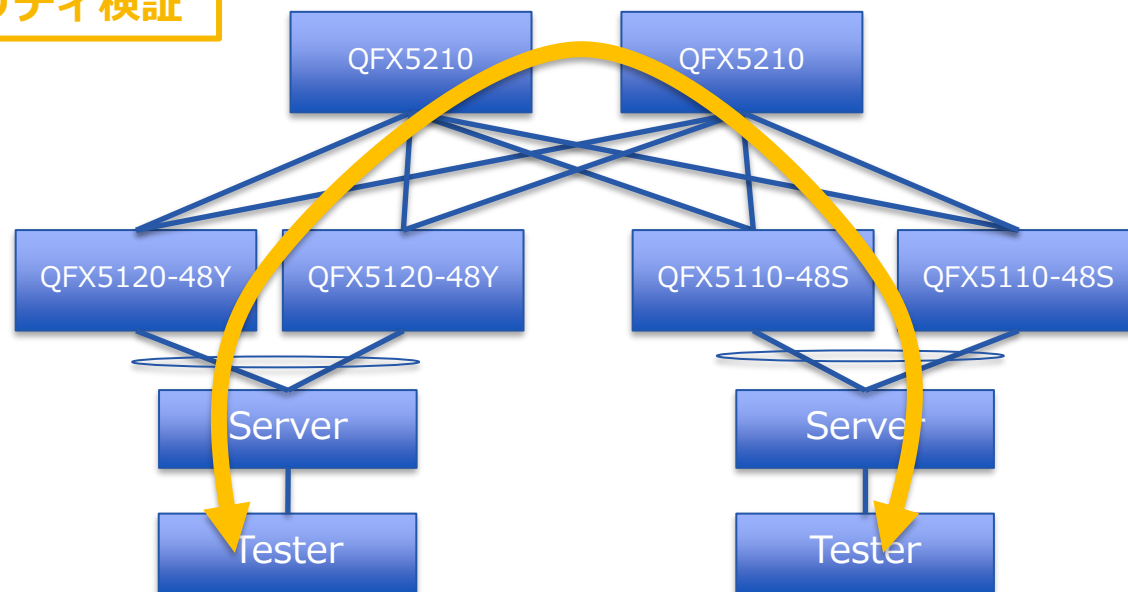


検証① スケーラビリティ検証 検証条件

❖ MACアドレスを大量に学習させスケールを確認

各Testerにて同一のVNI間で8万MACを生成し、トラフィックを印加（合計16万MAC）

1 MACスケーラビリティ検証



検証① スケーラビリティ検証 結果

❖ MACの学習結果をコマンドにて出力

#MACテーブルに登録されているMACのカウントを表示

```
nelco@QFX5120> show ethernet-switching table summary  
Total dynamic and static MAC addresses learned globally : 160001  
Configured static MAC addresses learned globally       : 0
```

#EVPNデータベースに登録されているMAC/IPのカウント

```
nelco@QFX5120> show evpn database | count  
Count: 160003 lines
```

【結果】

同一のVNI間で16万MACを学習でき
正常に通信できることを確認

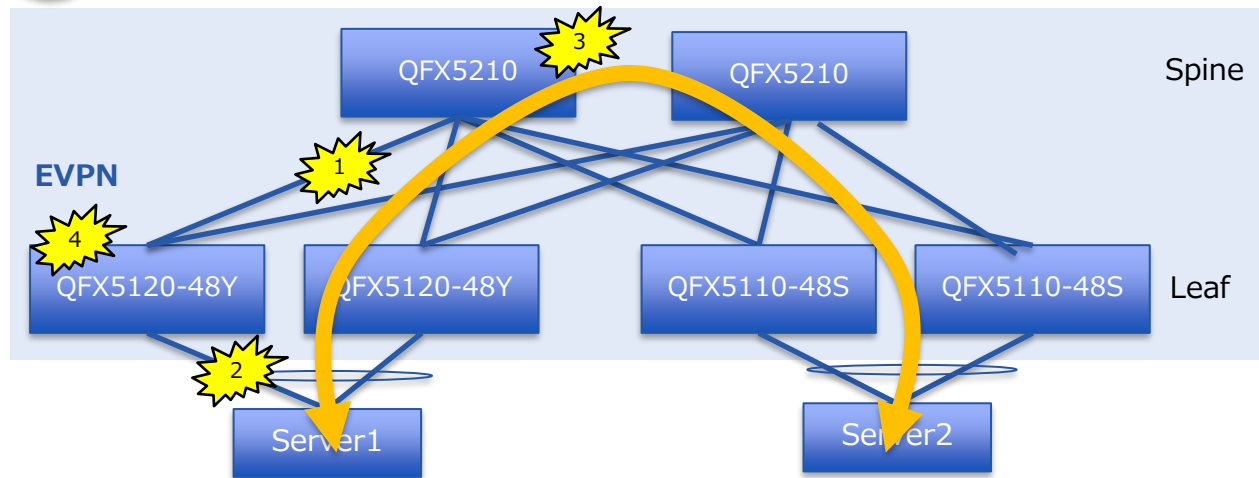
検証② 障害検証 検証条件

❖ 4か所で障害が発生した際の通信影響について確認

各Testerにて同一のVNI間で5000MACを生成して通信（合計1万MAC）

パケットは1万ppsで印加

2 障害検証



【検証パターン】

- ① Leaf – Spine間のリンク障害
- ② Leaf – Server間のリンク障害
- ③ Spineのノード障害
- ④ Leafのノード障害

検証② 障害検証 ロス計測結果

障害パターン	Frames Delta	Loss time (ms)
①Spine – Leaf間リンクダウン	Server1 → Server2 : 684 Server2 → Server1 : 777	Server1 → Server2 : 68ms Server2 → Server1 : 77ms
①Spine – Leaf間リンクアップ	Server1 → Server2 : 0 Server2 → Server1 : 0	Server1 → Server2 : 影響なし Server2 → Server1 : 影響なし
②Leaf – Server間リンクダウン	Server1 → Server2 : 44 Server2 → Server1 : 16,954	Server1 → Server2 : 4ms Server2 → Server1 : 1,695ms ※1
②Leaf – Server間リンクアップ	Server1 → Server2 : 3010 Server2 → Server1 : 0	Server1 → Server2 : 301ms Server2 → Server1 : 影響なし
③Spineノードダウン	Server1 → Server2 : 1,600 Server2 → Server1 : 2,017	Server1 → Server2 : 160ms Server2 → Server1 : 201ms
③Spineノードアップ	Server1 → Server2 : 0 Server2 → Server1 : 7,570	Server1 → Server2 : 影響なし Server2 → Server1 : 757ms
④Leafノードダウン	Server1 → Server2 : 675 Server2 → Server1 : 2,748	Server1 → Server2 : 67ms Server2 → Server1 : 274ms
④Leafノードアップ	Server1 → Server2 : 11,844 Server2 → Server1 : 0	Server1 → Server2 : 1,184ms ※2 Server2 → Server1 : 影響なし

ロスタイム1,000ms以下を基準とし、**多くのパターンで通信への影響がないことを確認**
1,000msを超える箇所については、下記の理由に拠るものと推測

- ※ 1 : Leafが障害検知し、該当MAC/IP経路をSpineに対してBGP Withdrawするまでに時間を要したものと推測
Server間通信は冗長Leafを介して継続
- ※ 2 : スケールによって影響度が変わる可能性あり

検証② 障害検証 結果

【結果】

- ・それぞれのパターンにおいて障害時には経路が適切に切り替わっている事を確認
- ・復旧後もトラフィックが偏り続けることはなく、バランシングされている事を確認

ロスが発生している部分も実際の断時間としては1秒程度
想定の範囲内の動作であったものと評価

参考：パターン③Spineのノード障害時におけるmonitor interface trafficの結果

ノードダウン時 monitor interface traffic 結果

QFX5120-48Y		Seconds: 224		Time: 02:48:55	
Interface	Link	Input packets	(pps)	Output packets	(pps)
xe-0/0/0	Up	163846994	(5109)	130412375	(5128)
et-0/0/48	Down	69314564	(0)	97550535	(0)
et-0/0/49	Up	64594038	(5290)	89740384	(5397)
ae0	Up	163846994	(5109)	130412375	(5128)



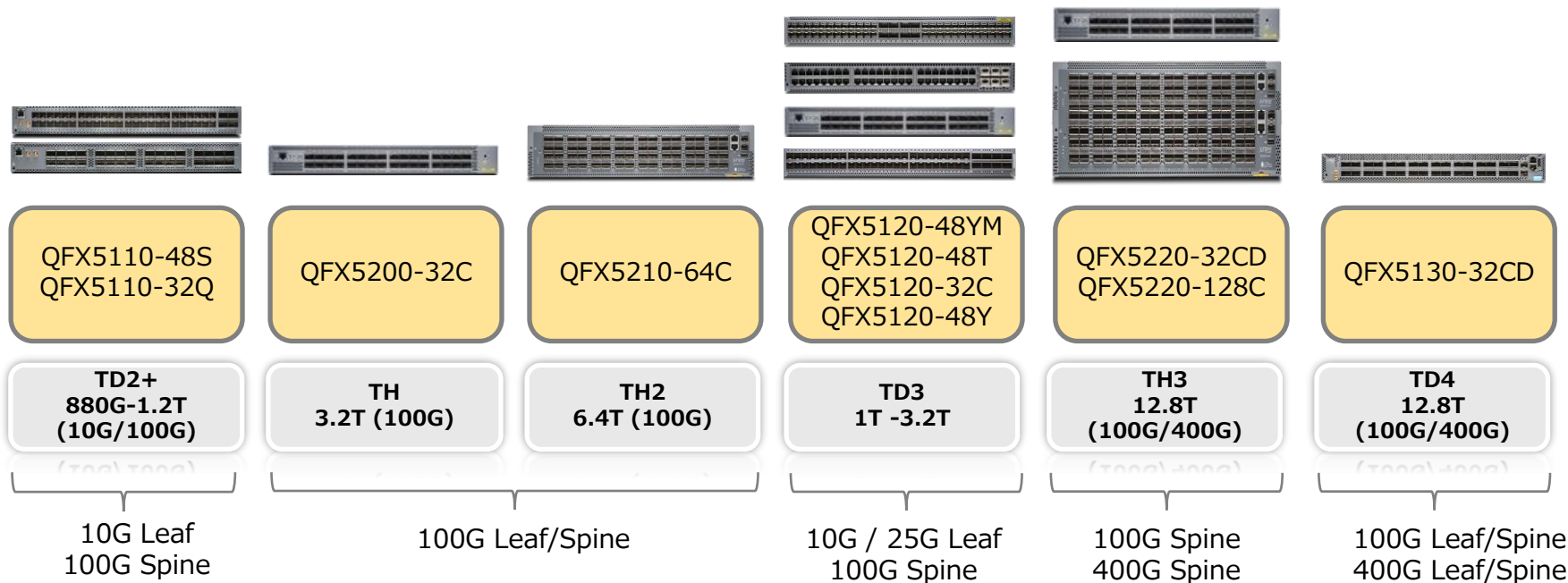
ノードアップ時 monitor interface traffic 結果

QFX5120-48Y		Seconds: 901		Time: 03:00:12	
Interface	Link	Input packets	(pps)	Output packets	(pps)
xe-0/0/0	Up	166948127	(5104)	133474151	(4907)
et-0/0/48	Up	69928886	(2416)	98251985	(2755)
et-0/0/49	Up	67159414	(2648)	92335417	(2628)
ae0	Up	166948127	(5104)	133474151	(4907)

3. EVPN/VXLANを構成する JUNIPER製品

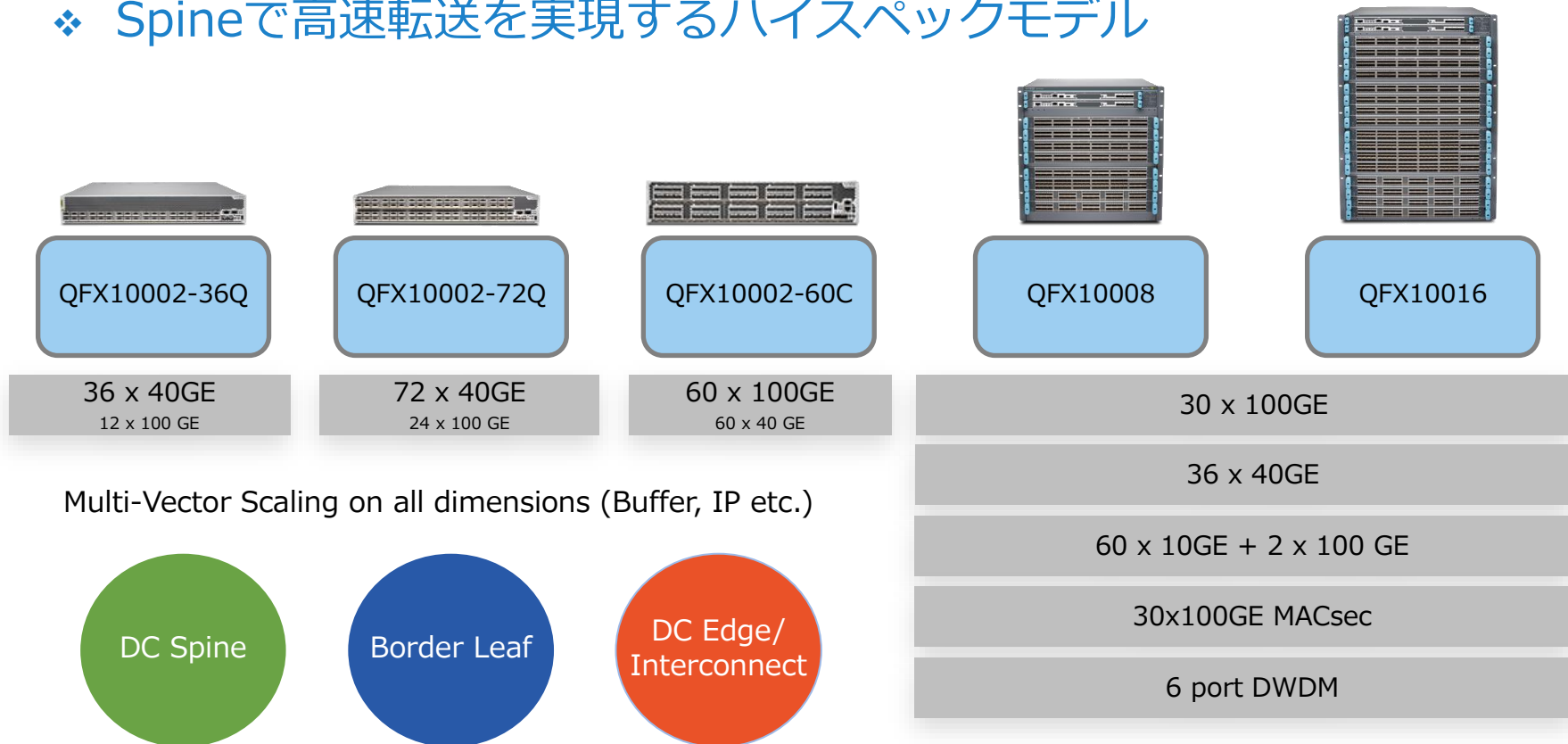
QFX5000シリーズラインナップ

- ❖ Leaf, Lean-Spine, モデルによってはSpineまでをカバーする
多用途ボックス型モデル

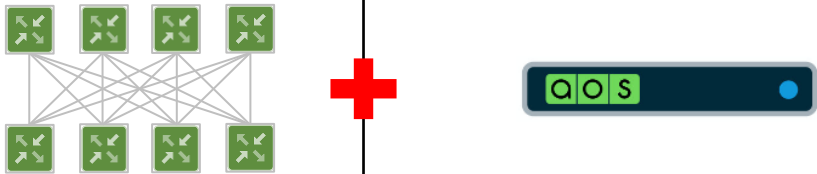


QFX10000シリーズラインナップ

❖ Spineで高速転送を実現するハイスペックモデル



EVPN/VXLANを自動構築 : Apstra Intent-Base System (AIS)

	IP Fabric (VXLAN)		Apstra Intent-Base System	
				
設計	△	設定パラメータが多い	◎	WebUIで自動コンフィグ作成
	○	複数レイヤを使いLB	○	対応
	○	端末接続スイッチがDG	○	対応
	○	マルチベンダー接続可	◎	複数ベンダー間の接続をテスト済み
運用	○	ループ排除で安定	○	スイッチの実装で可
	○	トラブルシュート容易 (オープン技術)	◎	豊富な運用・監視ツールを提供
拡張	○	コアスイッチ3つ以上可	○	対応
	○	3階層以上に対応	○	5-Stage CLOS (3階層)対応
人材	△	技術者はまだ少ない	◎	クラウドラボでトレーニング

AISはさらに**自動化**を提供

ネットワーク設計



パラメータ定義



コンフィグ作成



移行作業



運用監視



全体のまとめ

❖ EVPN/VXLAN構成の採用が本格化

VC構成、VCF構成からの置き換え、DC内ネットワークのループフリー化、DC間におけるL2延伸などをお考えのお客様にご検討いただいています

❖ Juniper製品における機能拡張

JUNOS 20.4以降で大幅な機能拡張が図られ、VLAN-Based,VLAN-Bundleに対応
また、MAC-VRF対応によるEVI単位でのユーザ分割案も可能に

❖ 検証におけるパフォーマンスも良好

MACスケラビリティの検証、障害検証の両者ともに、実環境への導入を前提として
申し分のないパフォーマンス結果を発揮

❖ 自動化、Intent-base Networksへの道のり

Apstra Intente-base System (AIS)の導入で運用を含めた自動化を実現

「日商エレクトロニクス、日商エレクトロニクスのロゴは、日商エレクトロニクス株式会社の登録商標です。

Juniper Networks、Juniper Networksのロゴは、米国及びその他の国のJuniper Networks, Inc.の登録商標です。
その他記載されている全ての商標、サービスマーク、登録商標、登録サービスマークは、各所有者に所有権があります。」

「本書記載の仕様はすべて予告なく変更される場合があります。本書の記載内容に誤りがあった場合、あるいは記載内容を更新する義務が生じた場合も、日商エレクトロニクスは一切責任を負いません。日商エレクトロニクスは、本書を予告なく変更、修正、または改訂する権利を有します。」「本書に記載されております内容を無断で複写、複製することを禁じます。」

jg@nissho-ele.co.jp

ご質問、ご用命はお気軽にお問い合わせください

