Juniper DC Trends

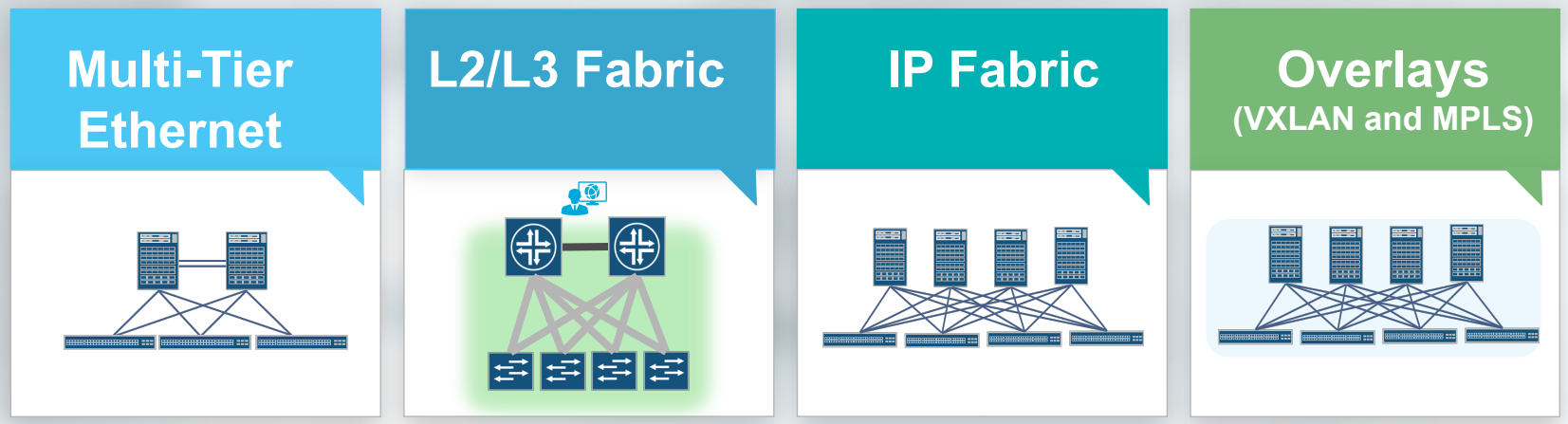
Michael Pergament, Senior Technical Marketing Engineer (JNCIE-ENT, JNCIE-SP)

Diverse Network Architectures

Network Ops

DevOps

IT/Cloud Ops

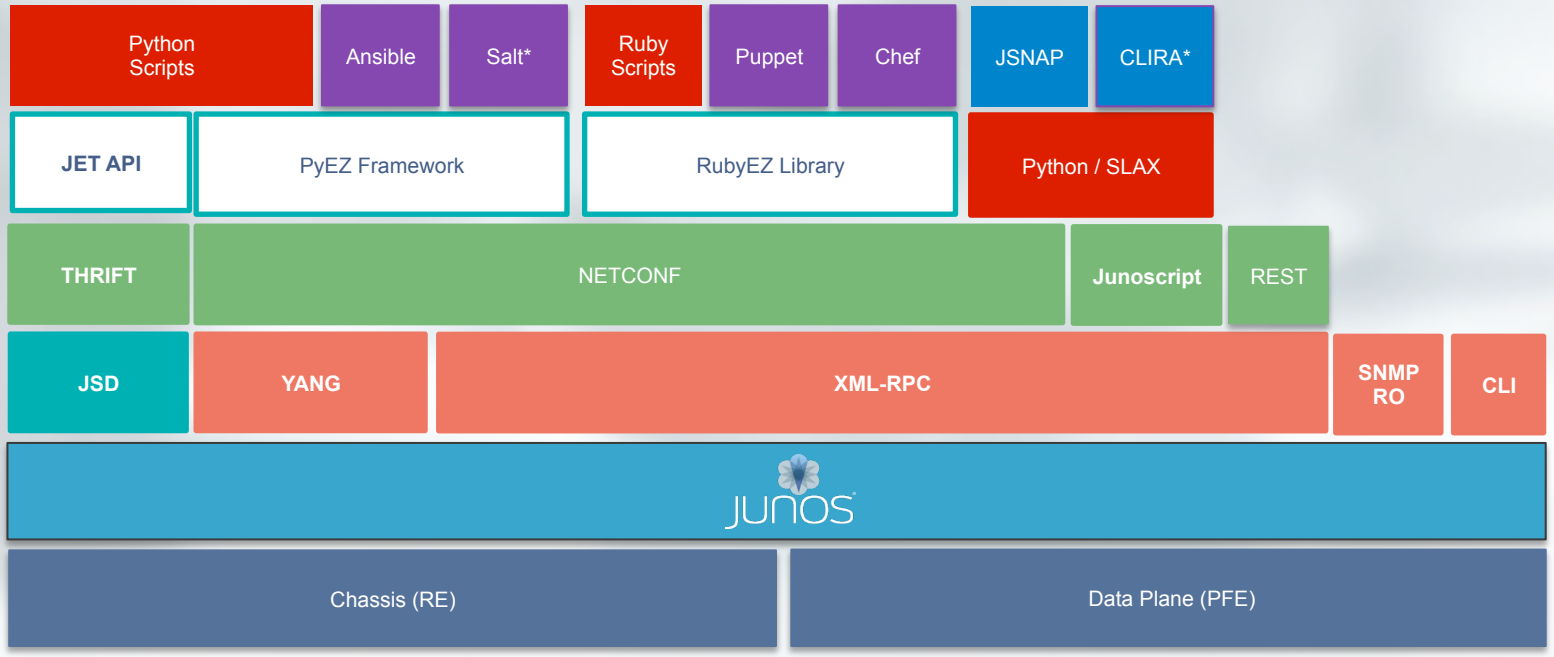


Common building block

QFX10000 spine switch

QFX5xxx Leaf switch

Junos Automation Stack



Junos Platform Automation Stack

Juniper's Contribution to NETCONF and YANG

Juniper has been a leader in programmability

- Supported XML APIs from day-one
- Well defined data models for configuration and operational commands

Took Juniper inventions to IETF and these are now industry standards:

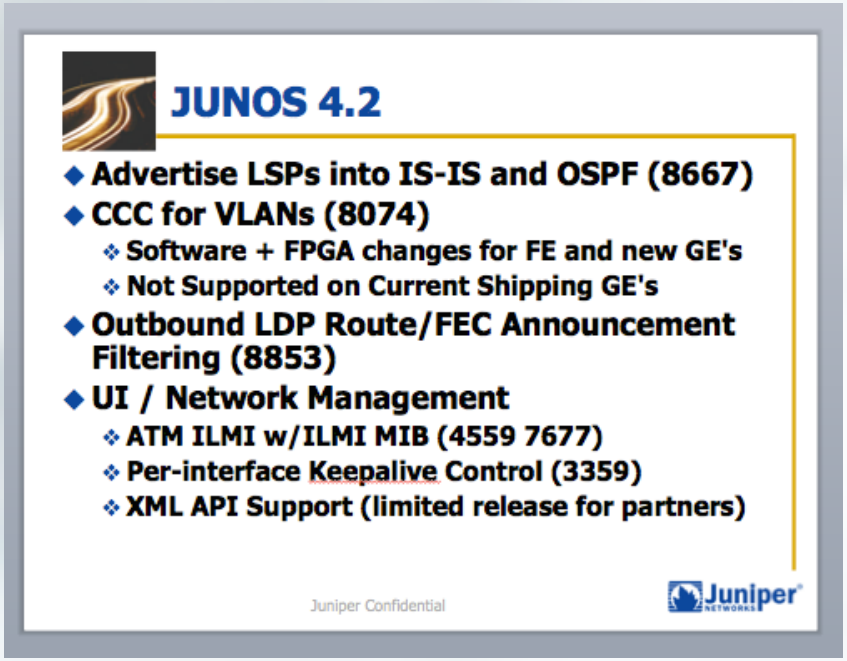
- NETCONF is the IETF standard protocol for managing devices and is derived from *Junoscript*
- YANG is the data modeling language and is based of Juniper *Data Definition Language*

Junos is XML-based from Day One

XML since 1996

Public programmatic XML interface since 13 October 2000

CLI code never has changed since then (it is still 200 lines!!!)



Junos and Data Modeling

Junos DDL 1998

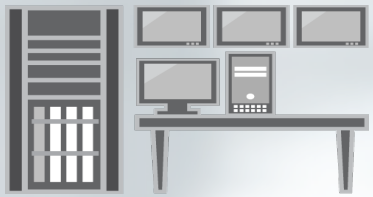
```
object user {
    flags list;
    attribute name {
        help "User name";
        flag identifier;
        type string 1 .. 32;
        match "[a-z]";
        match-message "must be lower case
        characters";
    }
}
```

Junos Yang 2015

```
list user {
    key name;
    leaf name {
        description "User name";
        type string {
            length "1 .. 32";
            pattern "[a-z]" {
                description "must be lower case
                character";
            }
        }
    }
}
```

Junos at run-time

Custom Translation script

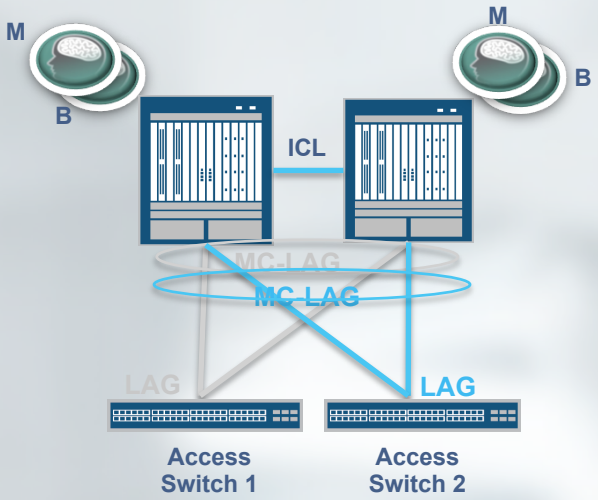


NETCONF

```
usr@rt> request system yang add package my-yang-pkg module
[opencfg_bgp.yang] translation-script opencfg_bgp.slax
Building schema and reloading /config/
. . .
. . .

usr@rt> configure
[edit]
user@router# set openconfig bgp local-as 12345 peer-as 8769
user@router# commit
```

MULTI-CHASSIS LAG - SUMMARY



MC-LAG provides a single (virtual) LAG interface towards LAN

- LAG interface spread to 2 MX Series chassis
- Eliminates STP – Reverse L2 Gateway Protocol Support
- Active-Active and Active-Standby modes
- HA/load-balancing solution

Integrated Routing and Bridging (IRB)

- Same gateway MAC address across 2 MX/QFX Series switches → eliminates need for VRRP
- Essential for VM mobility

State replication between 2 independent MX/QFX Series platforms (MC-LAG Active-Active)

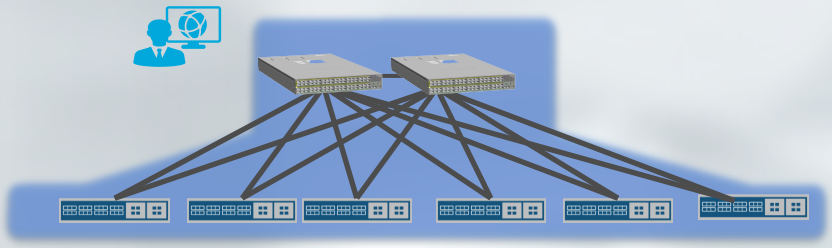
- L2, ARP, IGMP Snooping information synchronization
- LACP system-id coordination

Support for bridged and routed setups

- L2 only, L2 & L3, L3 only options

* No license required

Most Simplified and Coherent Technology Junos Fusion



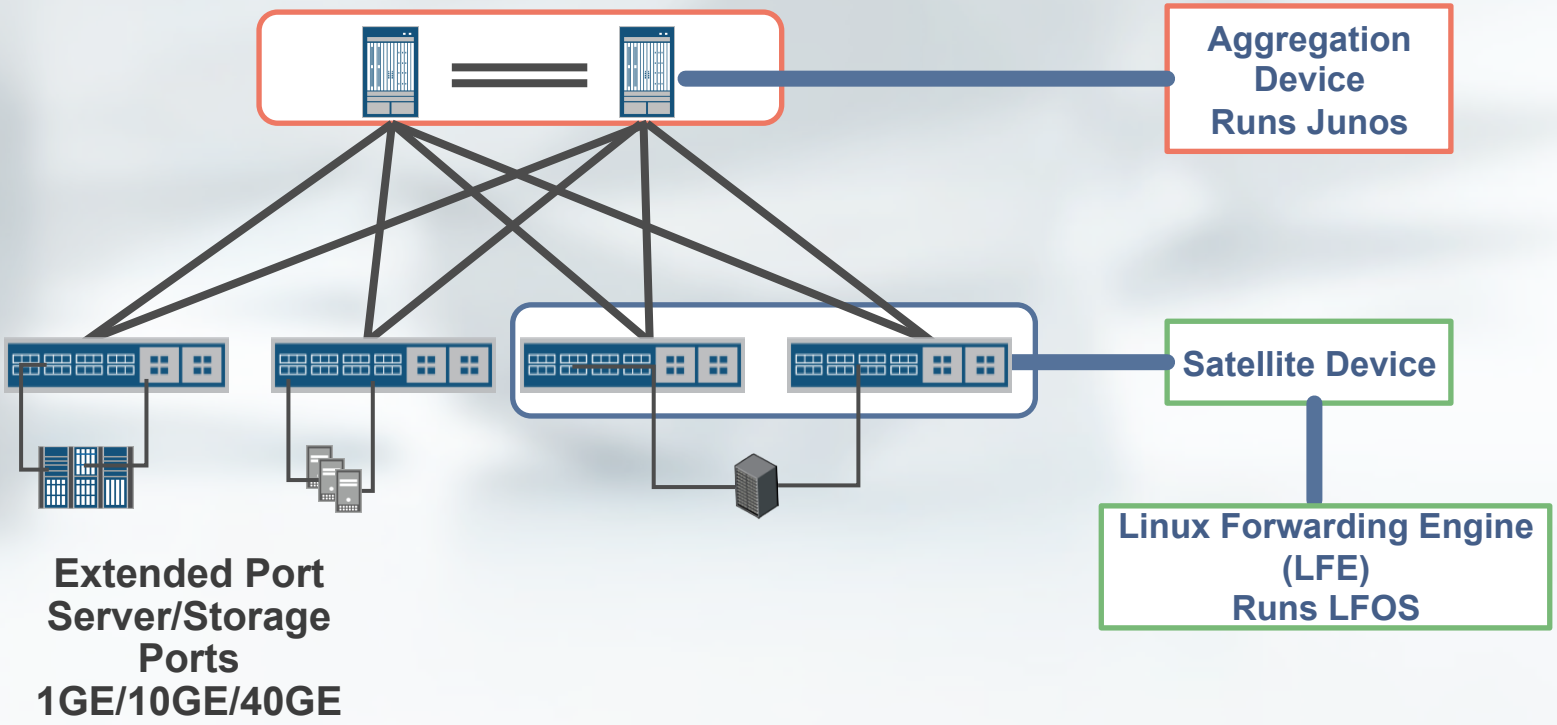
Aggregation Devices	QFX10000,MX
Satellite Devices	QFX5100, EX4300

SIMPLE
- Embedded automation
  - Plug-n-Play
  - VLAN Autoconf
- Centralized Management

SMART
- Dense 100G,40G for growing needs
- Universal SDN gateway (overlay routing)
- Deep virtual Buffers

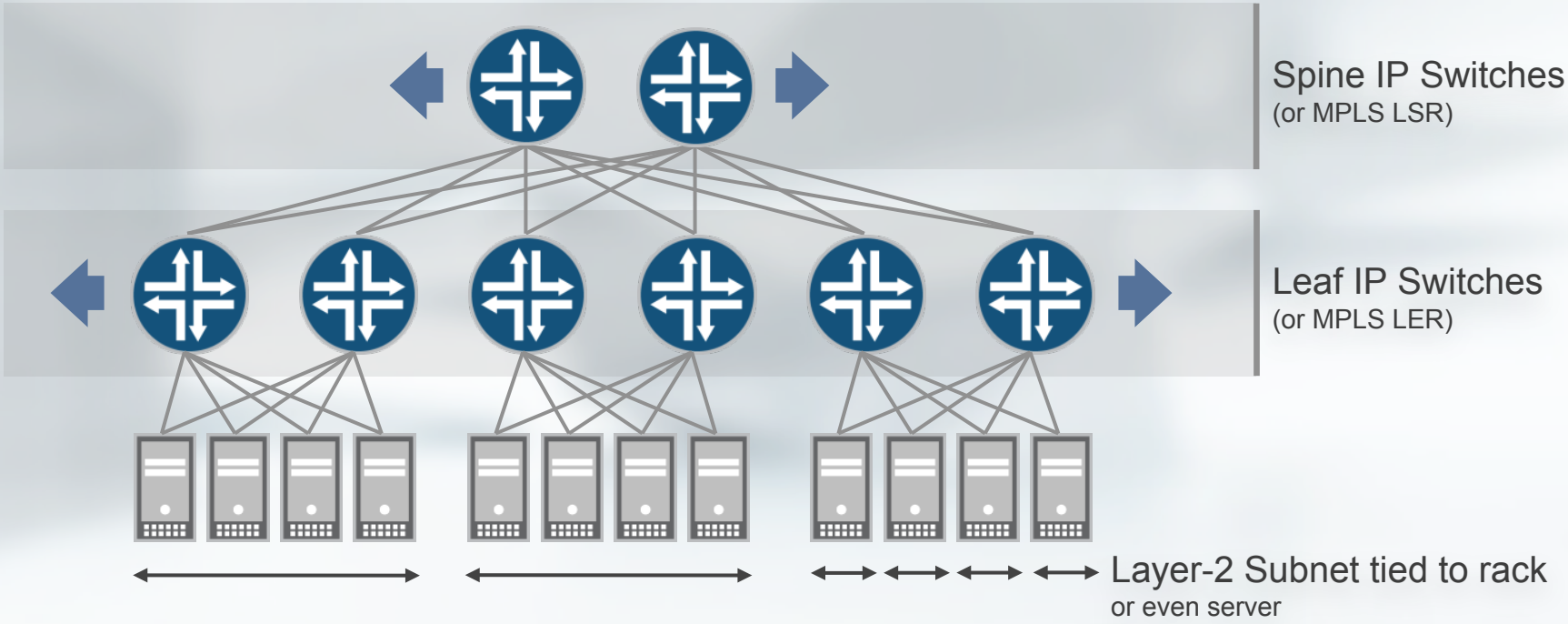
FLEXIBLE
- 1G, 10G, 25G, 40G standard TORs
- Choice of POD size up to 128 Racks
- Open & Programmable (802.1Br,EVPN,JSON)

Junos Fusion Terminology



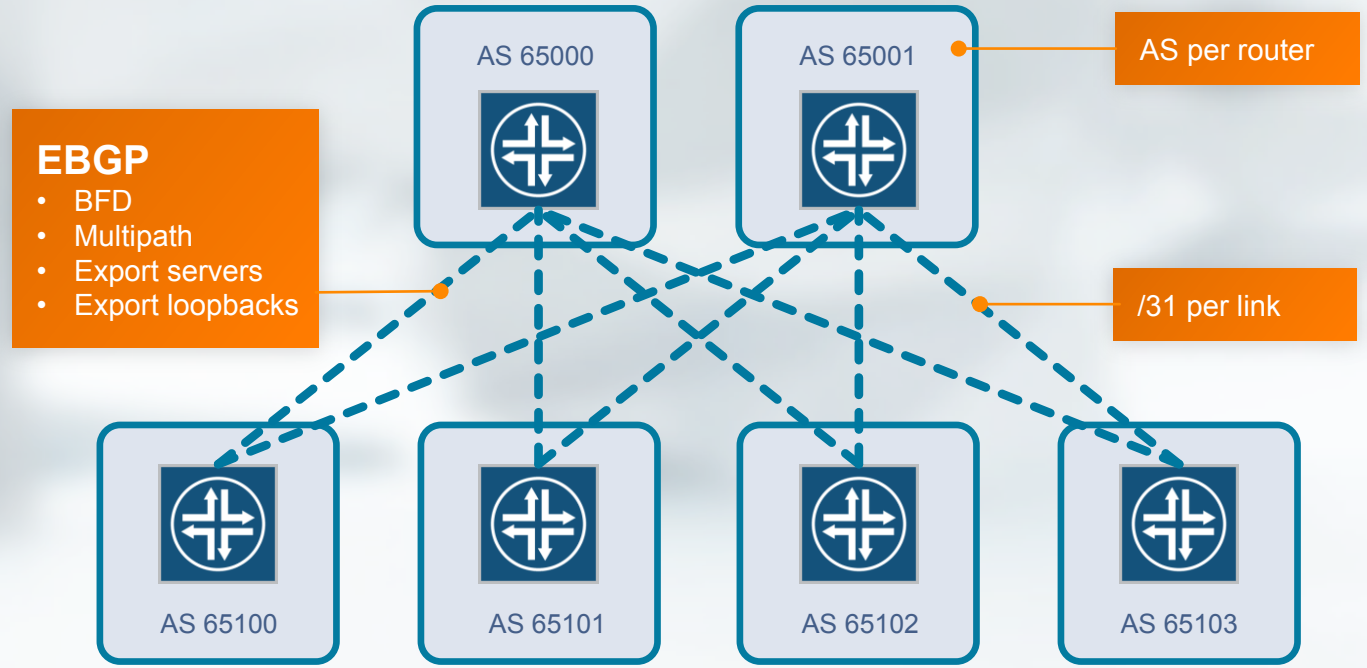
Layer 3 fabric underlay

For massive scale-out



BGP routing in the underlay

Massive scalability, rich policies, multi-protocol



OpenClos and Space Network Director

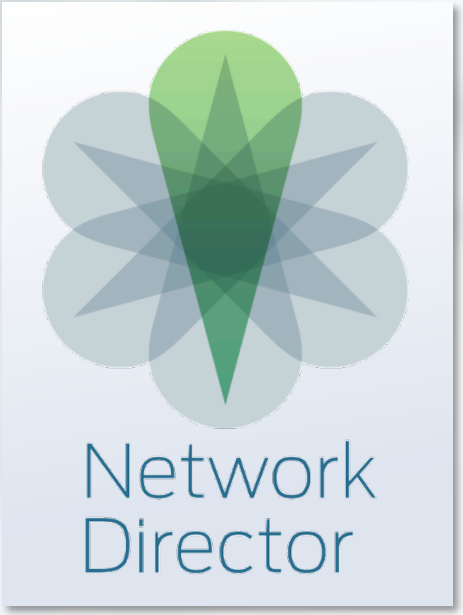
Automation for the underlay

Build
- 3-stage Clos topology
- 5-stage Clos topolog

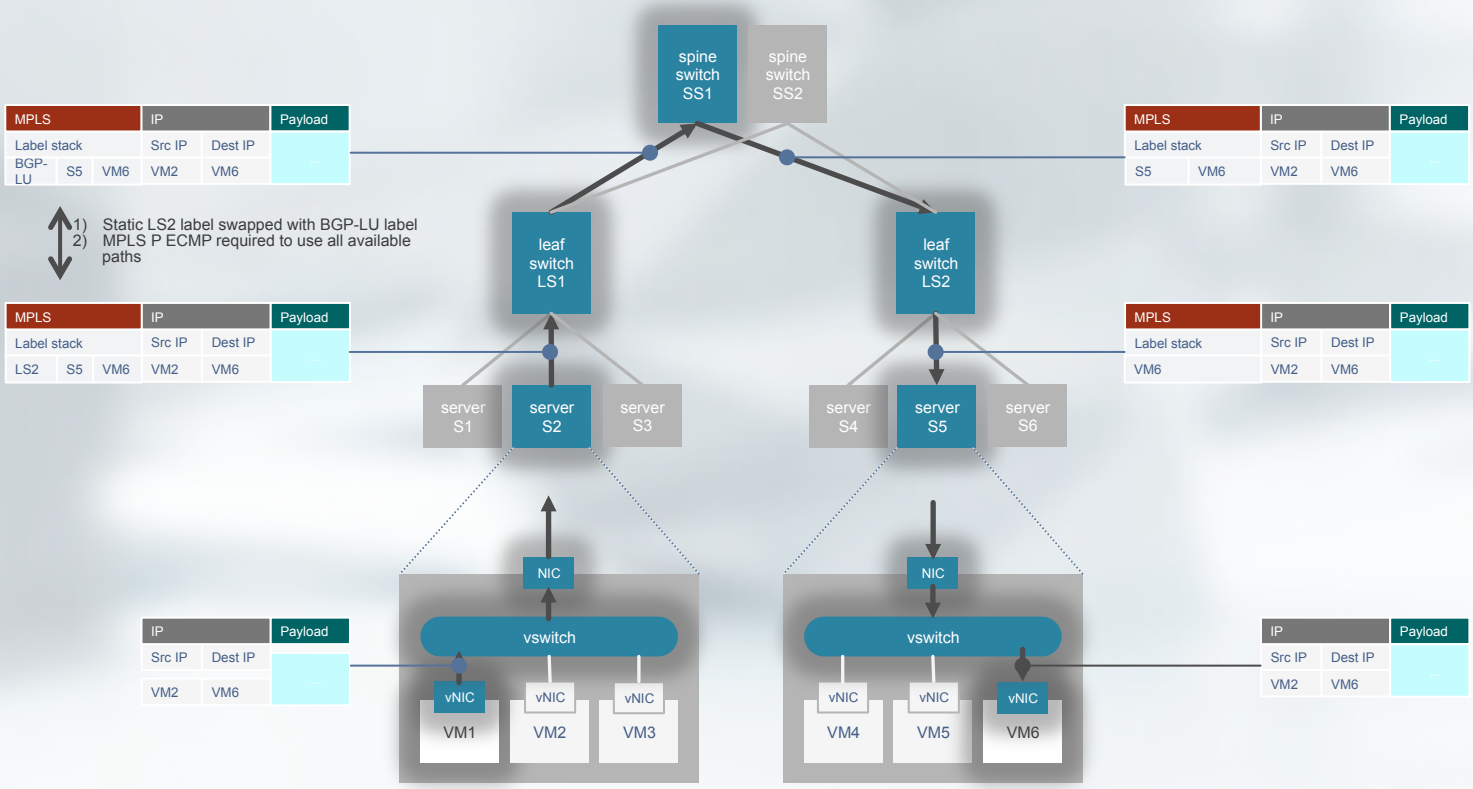
Maintain
- Add switches
- Replace switches
- Remove switches

Monitor
- Optics
- BGP sessions
- RIB and FIB
- Queues

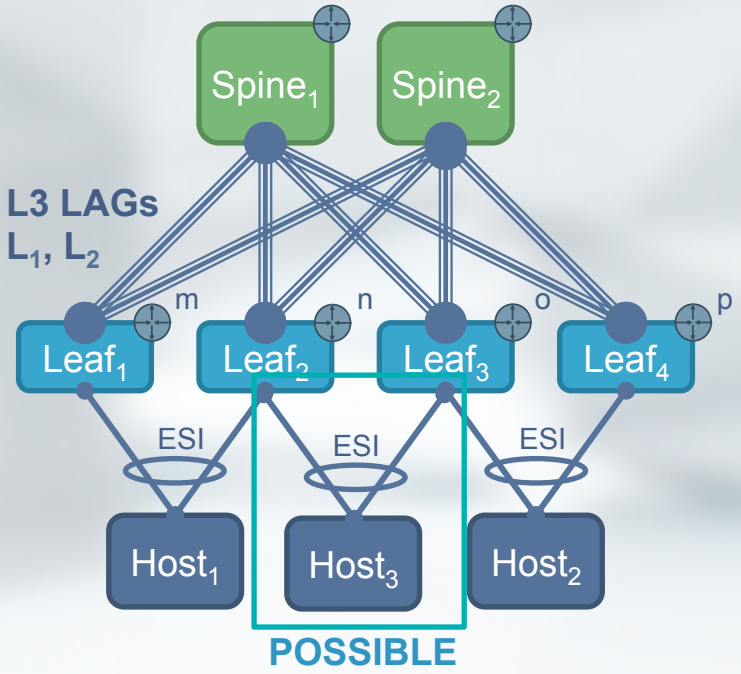
+



E2E MPLS DC



CONTROLLER-LESS EVPN/VXLAN



Requirements

- BMS data center
- L2 still required between all servers
- Take advantage of L3 underlay

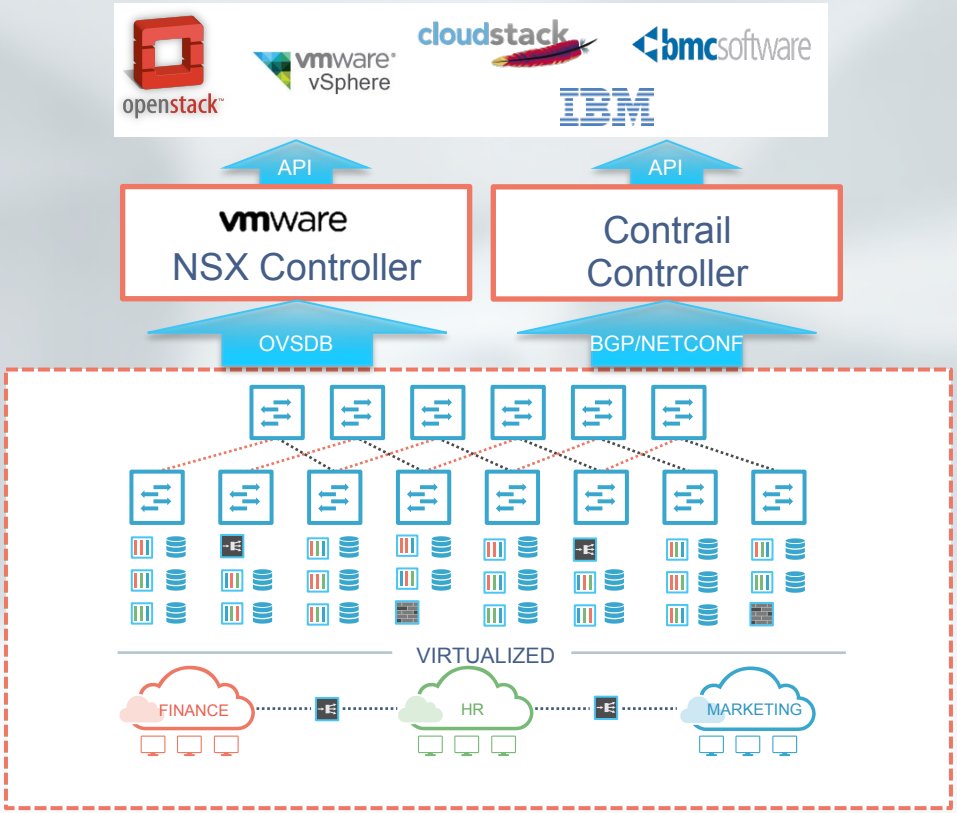
Solution

- Use distributed, controller-less EVPN/VXLAN
- Juniper uses unique VTEP addressing
- Different VTEP IP used per switch
- Uses VP-LAG to forward traffic to VTEPs

Advantages

- Traffic isn't blackholed during link failure
- Support ESI for multi-homing
- ICL between leaf switches not required

CONTROLLER-BASED OVERLAYS



Any Orchestration Systems

MX Series, QFX Series, EX Series and Contrail integration with cloud orchestration and automation systems

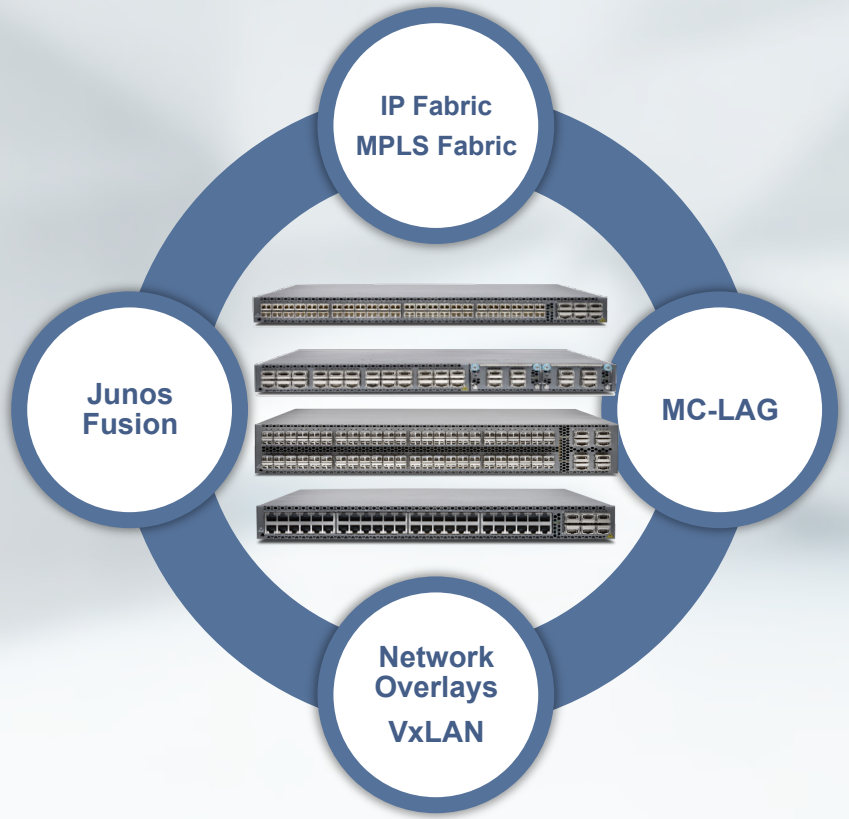
Any Network Virtualization Overlay

- Hardware integration with NSX, Contrail
- L2/L3 gateway services on MX Series (USG), QFX Series, EX Series and Contrail

Any Hardware Underlay

QFX Series, EX Series and MX Series—simple, universal building blocks to build networks designed for reliability, seamless scale and investment protection

QFX5100



Problem

- Coping up with change in server access technology
- Investment protection & SDN adoption
- Applications driving architecture diversity
- Increasing operations complexity & cost

Solution

- Choice of 10GE, 40GE, FCoE
- VXLAN L2 gateway , OVSDB and EVPN
- ISSU & automation integration

Benefits

- Future proof and investment protection
- Open & standards based for multi-vendor network
- ZTP for simplified operation
- ISSU with less than one second traffic impact during network software upgrades, upgrade time changed from 5-15 minutes to seconds
- OpenFlow

Why 25/50 GE matters?

- Reduces DC CapEx by providing migration pass from 10GE to 25 GE by leveraging existing (single port/lane) 10 GE infrastructure
- Maximizes ports and bandwidth in TOR switch faceplate (supports high server density in a rack)
- 2.5 speed increase at almost same cost as 10 GE
- Transition path from 25GE to 50GE to 100GE
- On the TOR same QSFP28 can be used for 25,50 GE (breakout) and 100GE

QFX5200 40G / 100G Data Center Leaf/ Spine Switches Summary

QFX5200-32C

QFX5200-64Q

	QFX5200-32C	QFX5200-64Q
Size, RU	1	2
Switch Throughput	3.2Tbps	3.2Tbps
25 GbE (Breakout Cable, QSFP28)	128	128
10 GbE* (Breakout Cable, QFSP +)	128	128
40 GbE (QSFP+)	32	64
50 GbE (Breakout Cable, QSFP28)	64	64
100 GbE (QSFP28)	32	32
PTP	Built-in	Built-in
Power Supplies	850W each	1600W each

*Port speeds < 10G are not supported

QFX10000 Product Summary

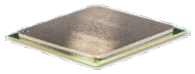
2.88 Tbps
1 Bpps

5.76 Tbps
2 Bpps

48 Tbps
16 Bpps

96 Tbps
32 Bpps

Juniper Q5 ASIC



Hybrid Memory Cube

Virtual Queuing | Virtual Network | Junos DevOps | Insight & CAE

Junos Fusion

Problem

- Physical & logical scale for cloud networks
- Transition to 100GbE
- Transition to SDN infrastructure
- Increasing operational complexity & cost

Solution

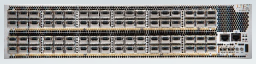
- Most scalable spine & core switch
- 10GbE, 40GbE & 100GbE (400GbE ready)
- Integrated high scale physical and virtual networking
- Integrated network & systems automation (Junos Fusion & Devops)

Benefits

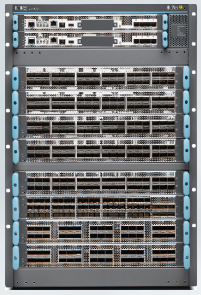
- Multi dimensional scale (bandwidth, hosts, buffer, filters)
- Overlay networking with standard protocols : VXLAN , EVPN,OVSDB
- Integrated precision telemetry & network analytics
- Carrier grade reliability & in service software upgrade

QFX10000 Product Summary

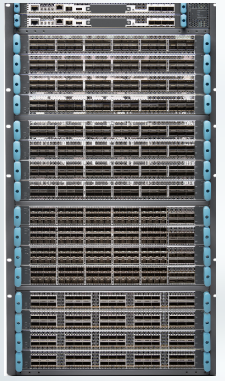
36X40GE
12X100GE
144X10GE



72X40GE
24X100GE
288X10GE



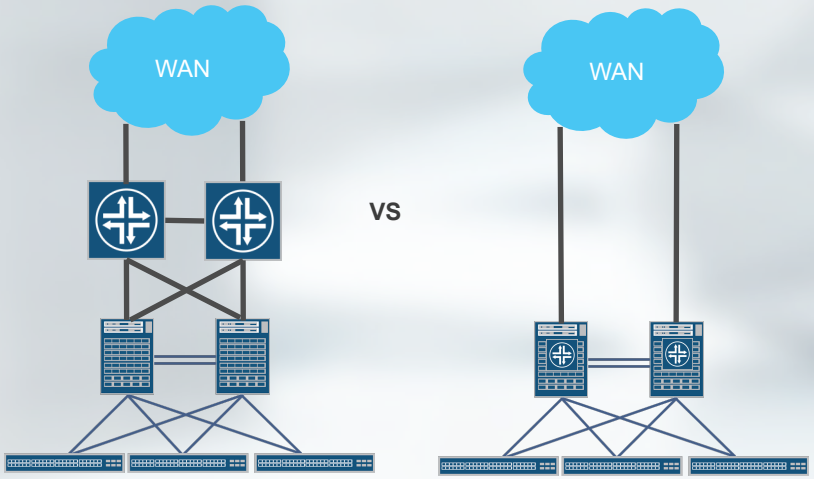
288X40GE
240X100GE
1152X10GE



576X40GE
480X100GE
2304X10GE

	QFX10002-36Q	QFX10002-72Q	QFX10008	QFX10016
MAC	256K	512K	1M	1M
FIB	512 LPM Routes (max 256K IPv4 and max 256K IPv6), XL License for QFX1002: 1M LPM Routes 2M Host Routes			
ACL	64K			
Latency	2.5us-5us			
Delay bandwidth buffer/packet buffer	Up to 100 ms/port	Up to 100 ms/port	Up to 100 ms/port	Up to 100 ms/port

QFX1000 Performance-Optimized Collapsed DC Design



Requirements on core/DCI collapsed component

- High throughput/port density
- High logical scale
- High feature richness (Universal GW nature)
- Deep buffers (speed difference between LAN and WAN ports)

Benefits

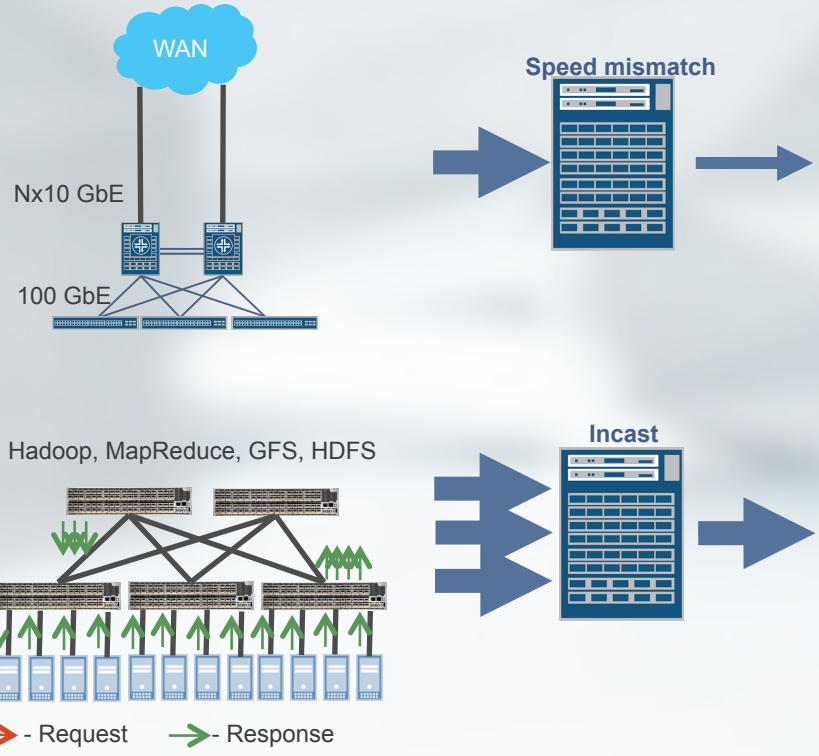
- For multi-tier DC architecture clean demarcation point between L2/L3 for L2 stretch across multiple DCs
- Better End-to-End latency
- Further simplification of management
- Saving physical links

Memory technology for 1Tbps silicon

HMC vs DDR

	DDR3/4	HMC
Number of Memory Devices	90 and Up	2
Total number of pins between asic and memory	More than 2400	422
Power	61W	49W
Memory surface area	12750mm^2 or more	1922mm^2

Virtual Queueing & high speed delay bandwidth buffer



Problem

- Incast challenges with Hadoop workloads
- DCI speed impedance
- Storage convergence

Solution

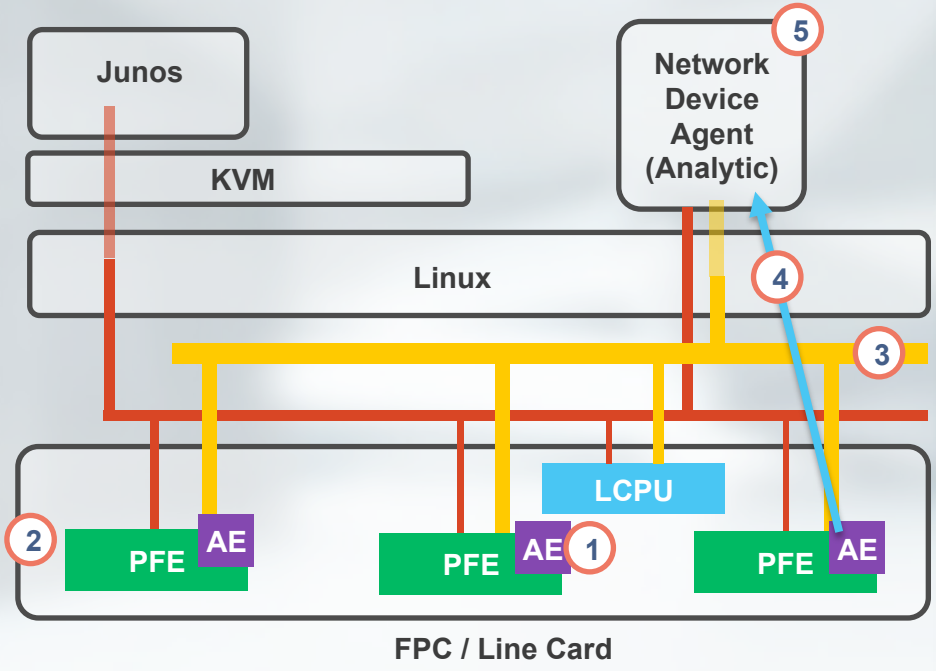
- Virtual output queue
- Deep buffers for speed impedance
- Elephant & mice flow separation

Benefits

- High speed memory with 100 msec buffer
- Sustain microbursts & unpredictable traffic flows
- Collapse multiple layers of networking saving Capex and OPex

Embedded Analytics Architecture (FRS+)

Junos Insight



Problem

Extract raw statistics from the data plane

- At Scale
- With reasonable resources
- With micro sec accuracy
- In a programmatic way

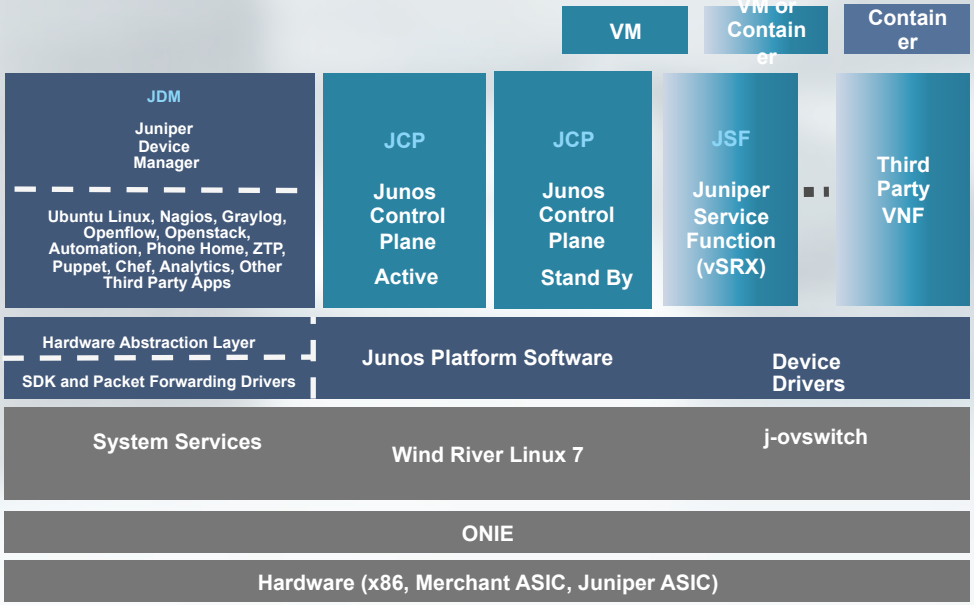
Solution

1. Have dedicated resource (Analytics Engine/AE) on each PFE to monitor all counters and resources
2. Implement PTP precision timing at PFE level
3. Create a dedicated high speed internal network
4. Implement a Push vs Pull model from Data plane to Control plane
5. Dedicated daemon on x86 platform

Benefits

- Micro burst monitoring with micro sec accuracy
- Very efficient resources utilization

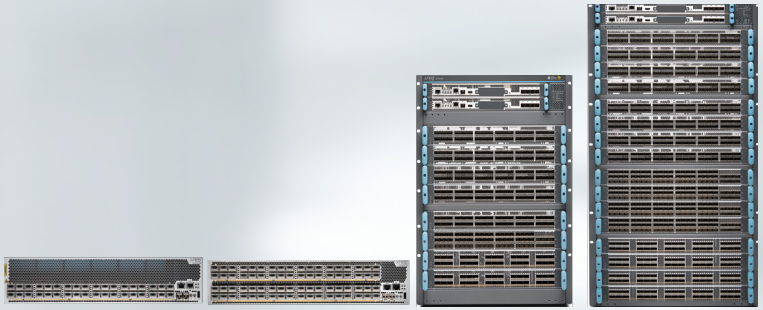
Software Architecture



Software Architecture Highlights

- Increase platform velocity (i.e. Time To Market)
- Versatile container and VM support
- Platform and PFE functions will be independent of Junos
- Flexible architecture support (TORs, Security, Chassis)
- Improve performance – Multicore CPU
- Allow multiple toolchain types via JDM
- Support programming via multiple APIs

Summary



① **IP, MPLS & virtual networking**

② **Robust network operating system with programmatic API**

③ **Network telemetry and analytics**

④ **High scale & performance through software & ASIC innovation**

⑤ **Carrier grade reliability & high availability**