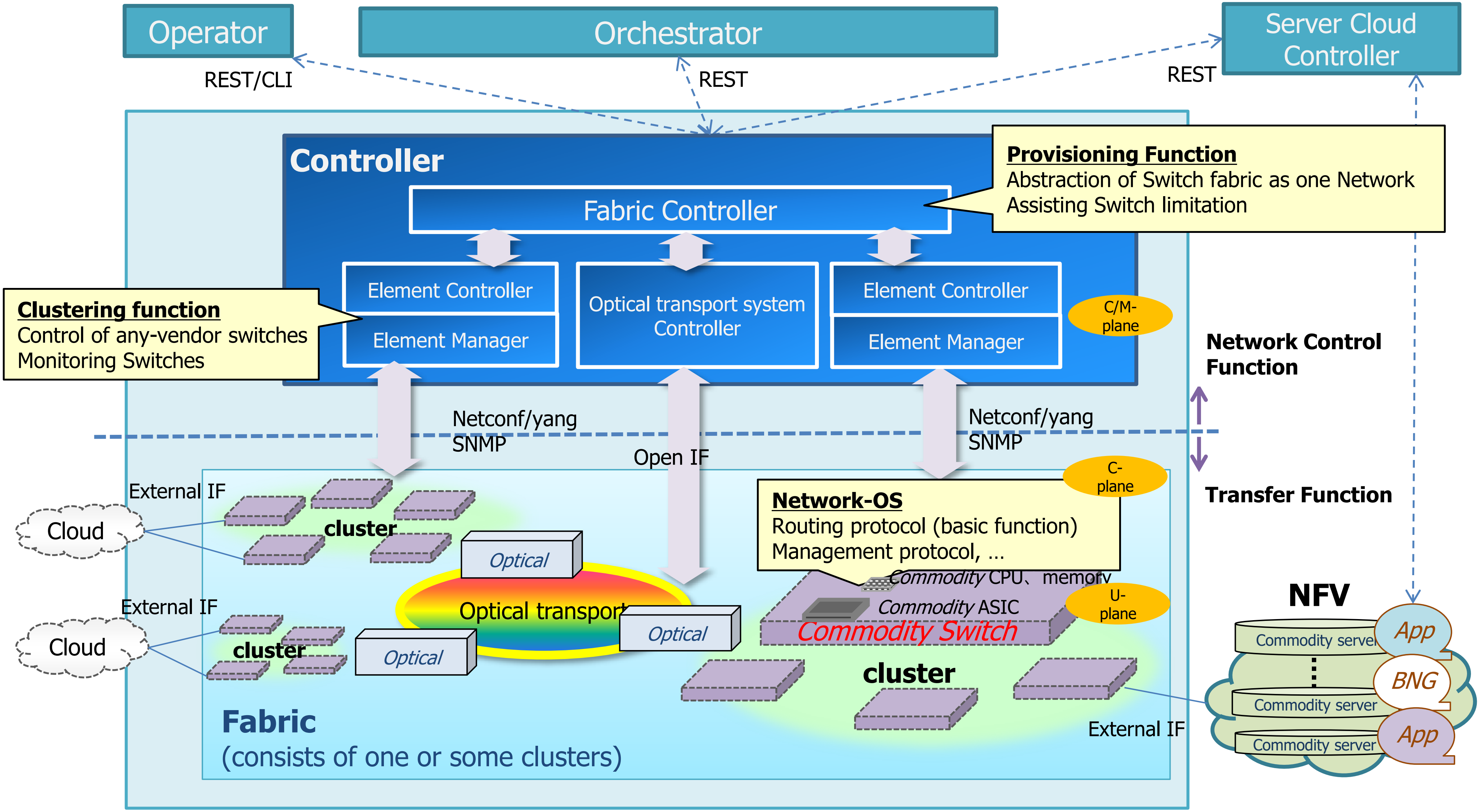
Technical Details

Jun. 2019

Architecture outline



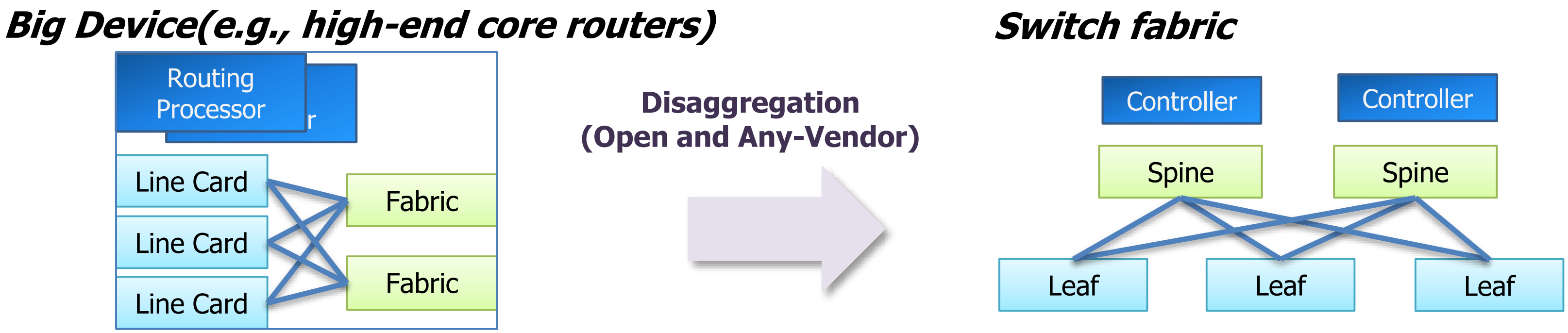
Provisioning Function
Abstraction of Switch fabric as one Network
Assisting Switch limitation

Clustering function
Control of any-vendor switches
Monitoring Switches

Network-OS
Routing protocol (basic function)
Management protocol, …

Disaggregation concept

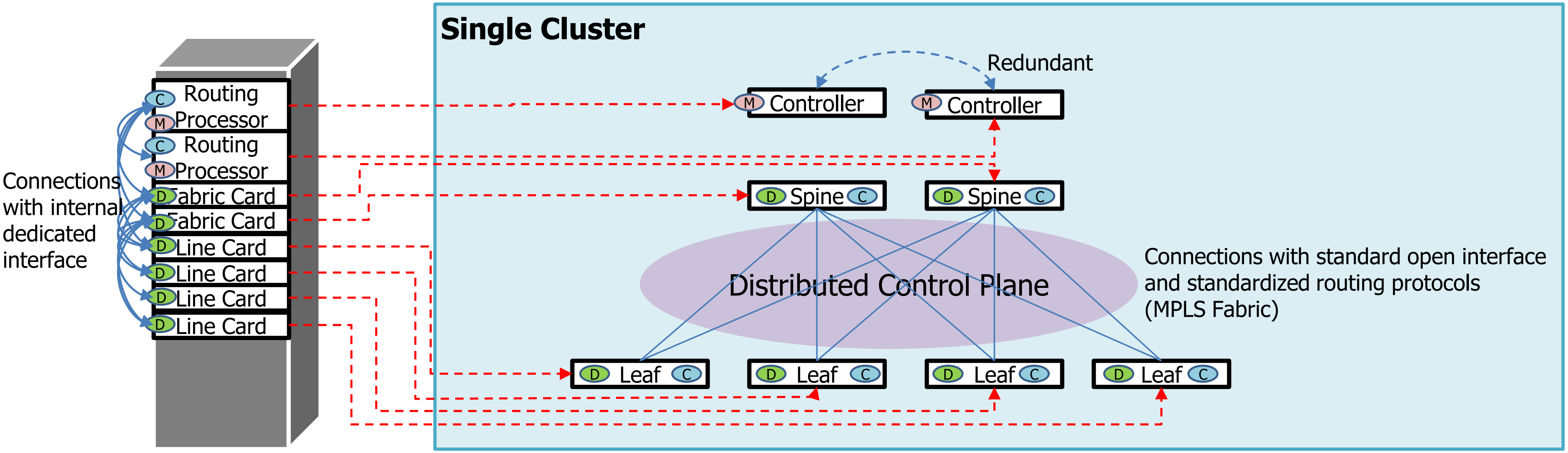
- Disaggregation for big device (e.g., high-end core routers) can be come due to technical progression for merchant silicon.
- Multi Service Fabric is a research for disaggregating router with distributed control plane which consists of standardized routing protocols and standard physical interface.
- Each device has autonomously control plane basically.
- SDN controller is centralized management system for numerous network nodes.
- Controller uses the same cluster service model to manage multi vendor switches.



Function	Big Device		Component	
Management Plane	Routing Processor	Processing Module	Controller	IA Server (VM)
Control Plane			Spine Leaf	Datacenter SW
Data Plane (Total Switching)	Fabric	Switch Fabric Module	Spine	Datacenter SW
Data Plane (Service Scalability)	Line Card	Line Card Module with Distributed ASIC	Leaf	Datacenter SW

Fabric architecture

- ◆ Controller manages nodes’ configuration and status.
- ◆ OSPF and LDP are used in underlay network configuration. VXLAN (L2VPN) and MP-BGP (L3VPN) are used in overlay service configuration.
- ◆ Cluster Scalability depends on mainly Switch hardware.



C Control plane

M Management plane

D User (data) plane

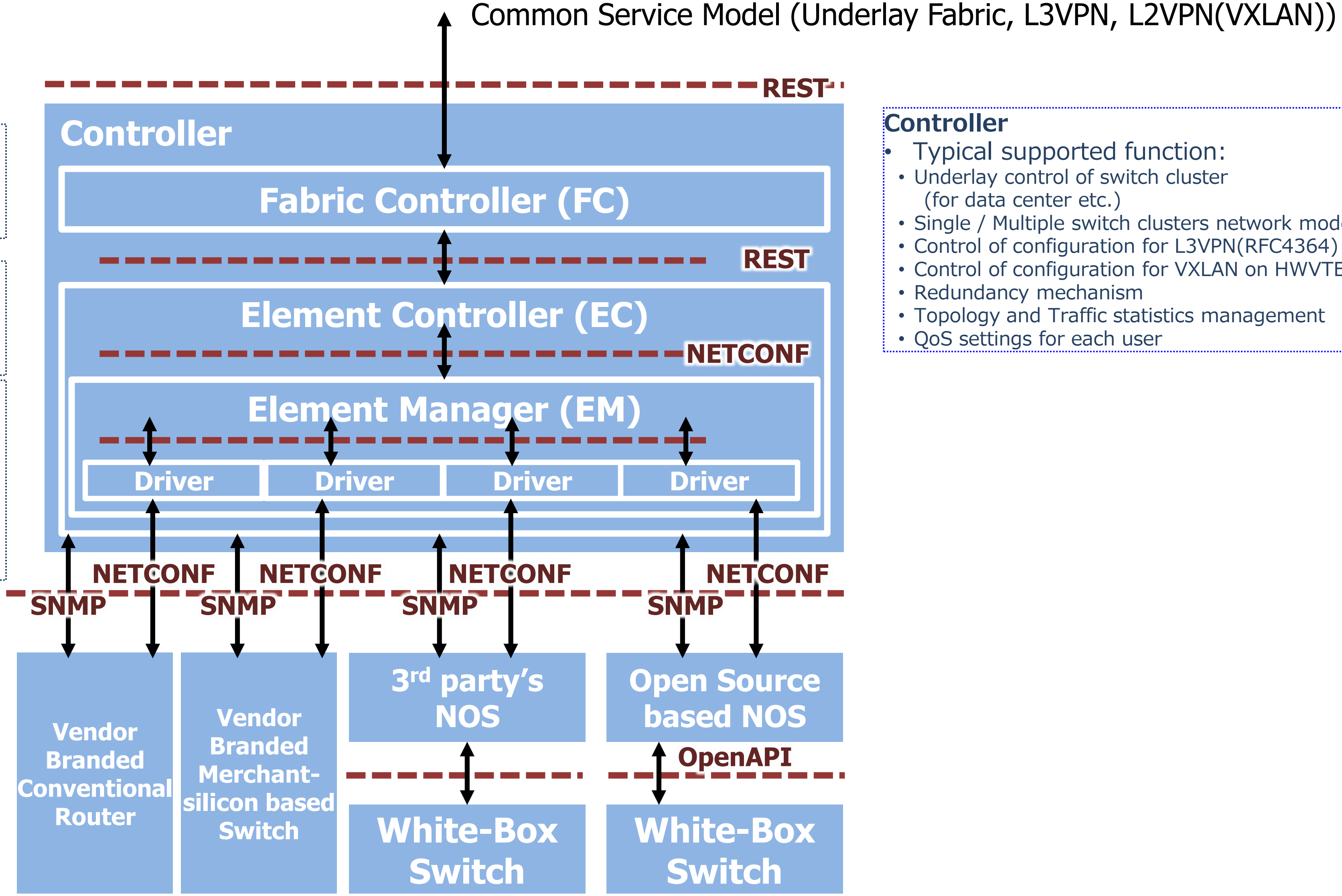
- Underlay network : OSPF
- VPN (overlay) route exchange: iBGP
- Supporting LAG between Spine and Leaf, Leaf and CE
- Supporting Redundancy for CE (Dual-home Device)
  - L2 : VXLAN (EVPN multi-homing is available)
  - L3 : VRRP, eBGP, OSPF

Controller architecture

FC
- Managing switch fabric model and network model

EC
- Conversion between logical and physical information

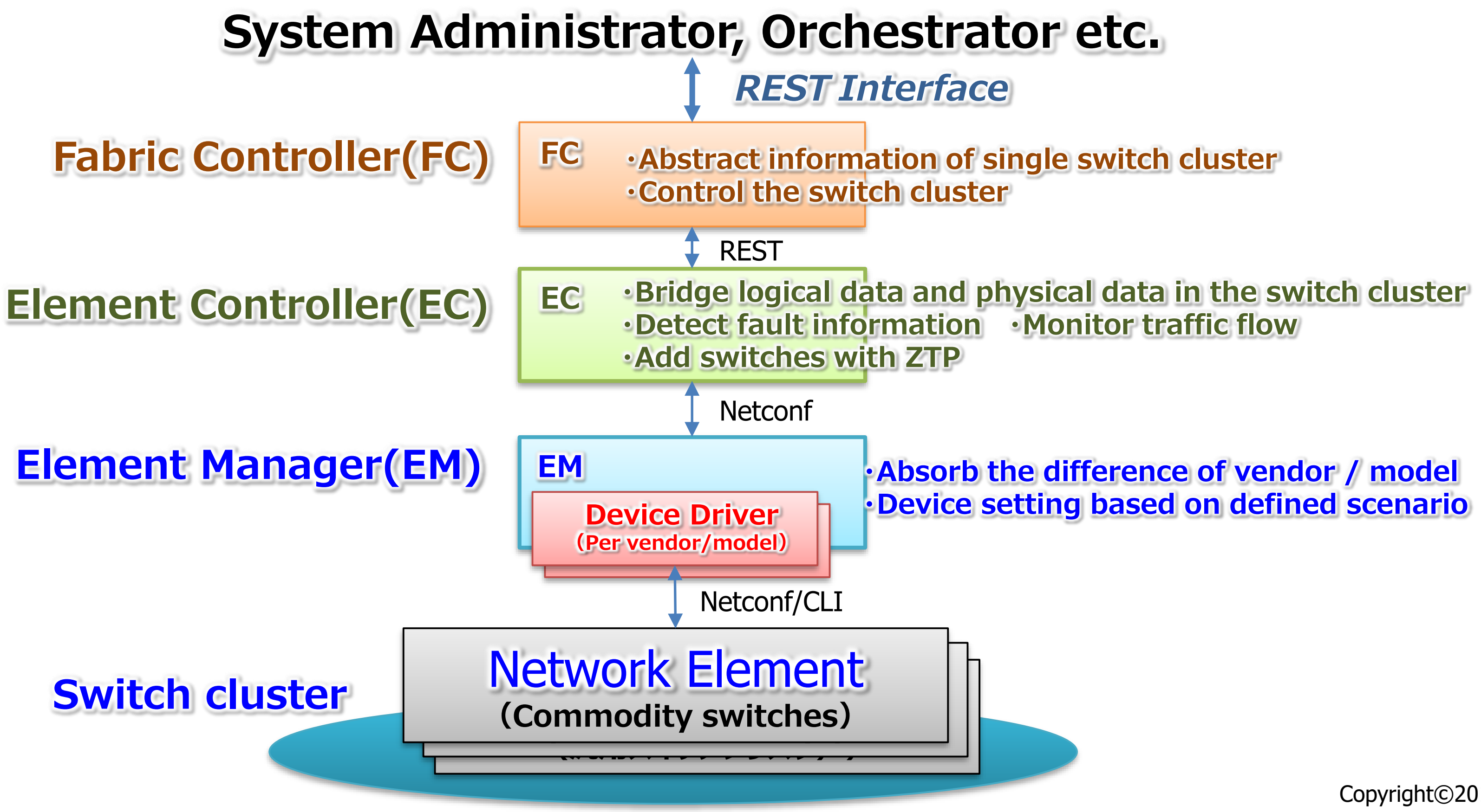
EM
- Absorb difference of vender-specific
- Provide switch driver to support device models for each device type



Controller
- Typical supported function:
  - Underlay control of switch cluster (for data center etc.)
  - Single / Multiple switch clusters network model
  - Control of configuration for L3VPN(RFC4364)
  - Control of configuration for VXLAN on HWVTEP
  - Redundancy mechanism
  - Topology and Traffic statistics management
  - QoS settings for each user

Controller Configuration for Single-location

- ◆ System administrators or upper systems can command the controller using the standard protocol REST.
- ◆ Each controller process the abstracted information, and finally set up the switches.
- ◆ The controllers not only process instructions from upper system, but also can notify the state change of the switches.



System Administrator, Orchestrator etc.

REST Interface

Fabric Controller(FC)

FC

・Abstract information of single switch cluster
・Control the switch cluster

REST

Element Controller(EC)

EC

・Bridge logical data and physical data in the switch cluster
・Detect fault information　・Monitor traffic flow
・Add switches with ZTP

Netconf

Element Manager(EM)

EM

Device Driver
(Per vendor/model)

・Absorb the difference of vendor / model
・Device setting based on defined scenario

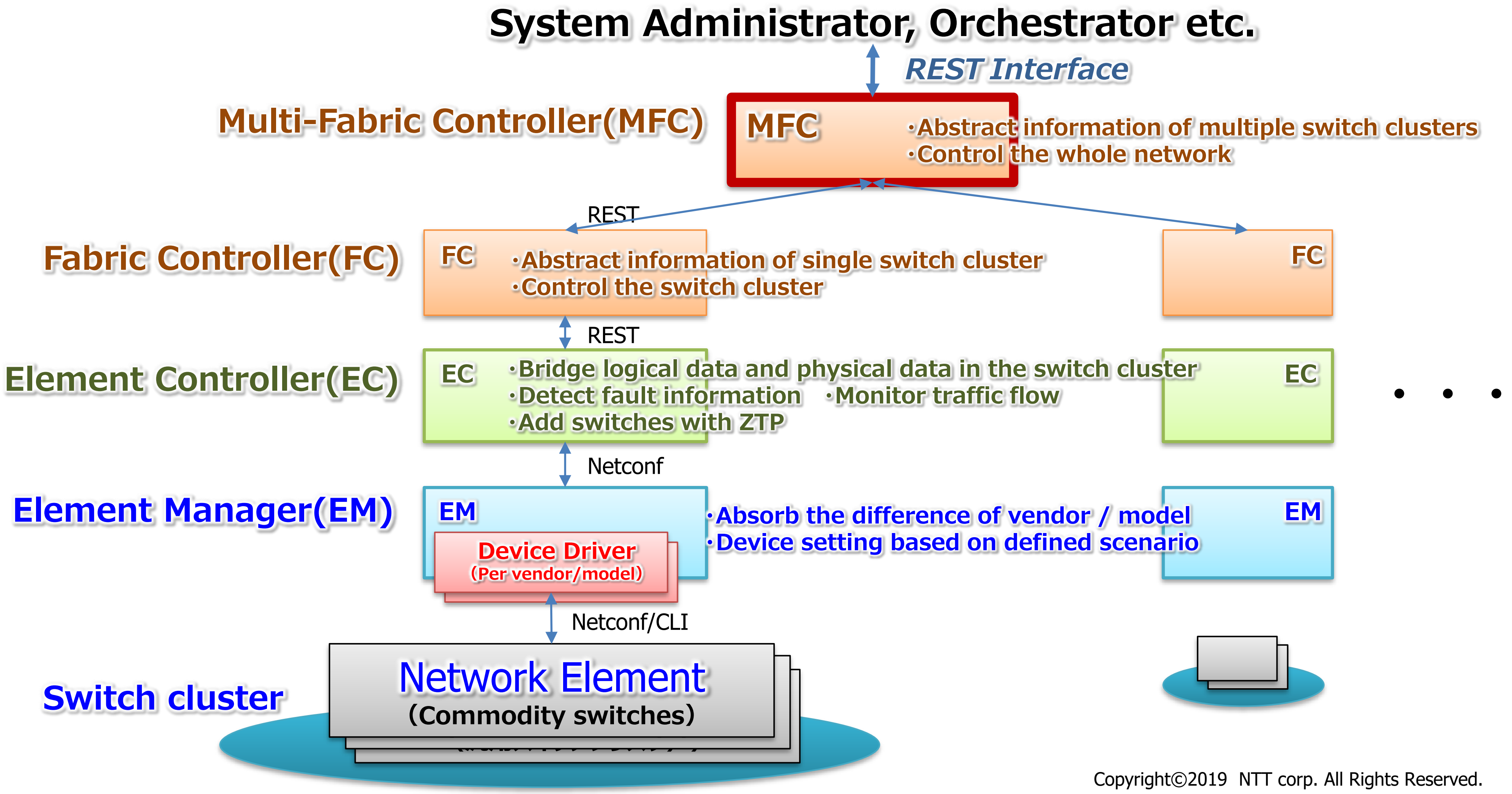
Netconf/CLI

Switch cluster

Network Element
(Commodity switches)

Controller Configuration for Multi-location

- ◆ The upper system can control the multiple switch clusters through Multi-Fabric Controller (MFC).
- ◆ It is possible to build VPNs between multiple switch clusters.



System Administrator, Orchestrator etc.

REST Interface

Multi-Fabric Controller(MFC)

MFC

- ·Abstract information of multiple switch clusters
- ·Control the whole network

REST

Fabric Controller(FC)

FC

- ·Abstract information of single switch cluster
- ·Control the switch cluster

FC

REST

Element Controller(EC)

EC

- ·Bridge logical data and physical data in the switch cluster
- ·Detect fault information ·Monitor traffic flow
- ·Add switches with ZTP

EC

・ ・ ・

Netconf

Element Manager(EM)

EM

Device Driver
(Per vendor/model)

- ·Absorb the difference of vendor / model
- ·Device setting based on defined scenario

EM

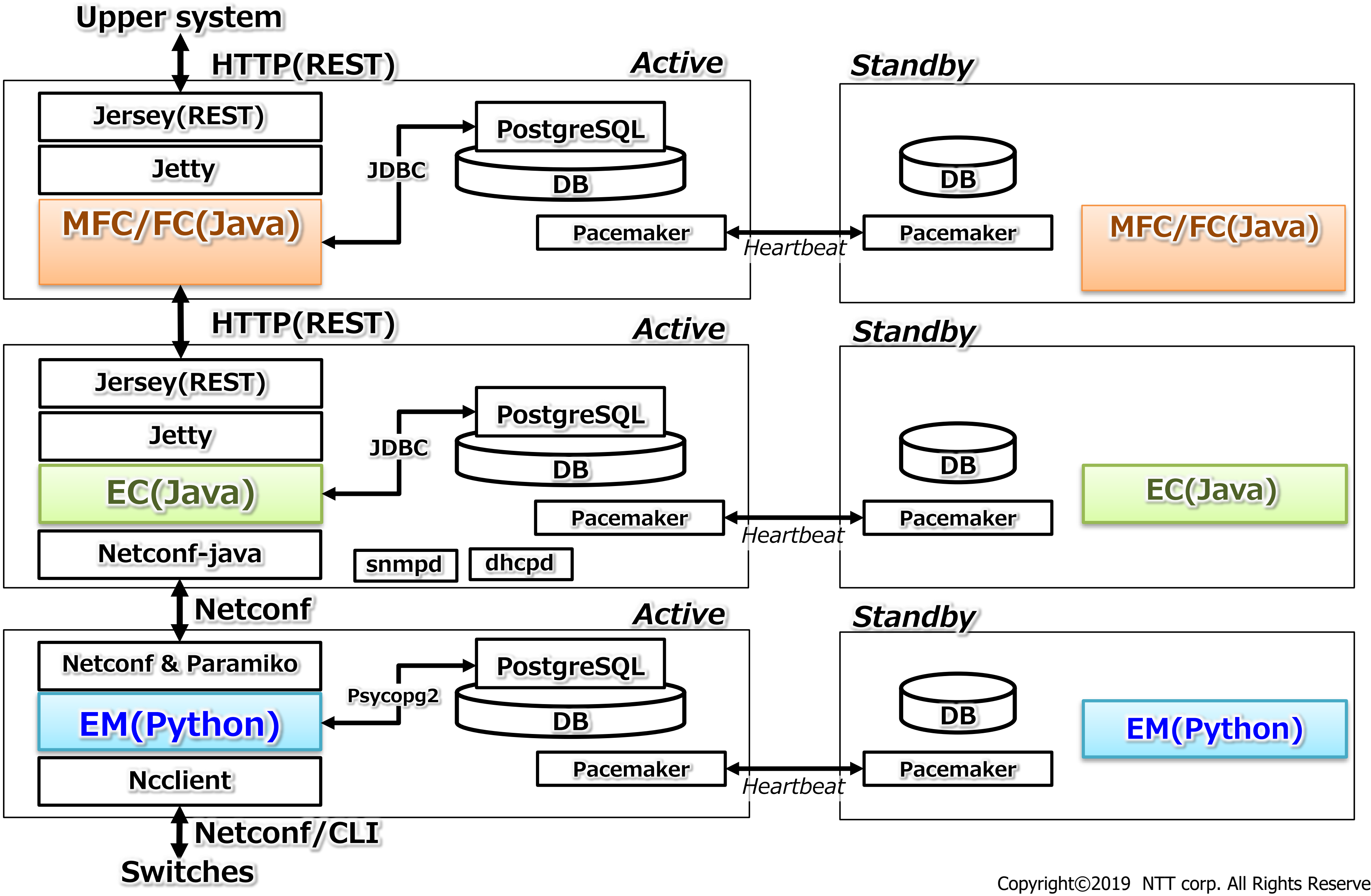
Netconf/CLI

Switch cluster

Network Element
(Commodity switches)

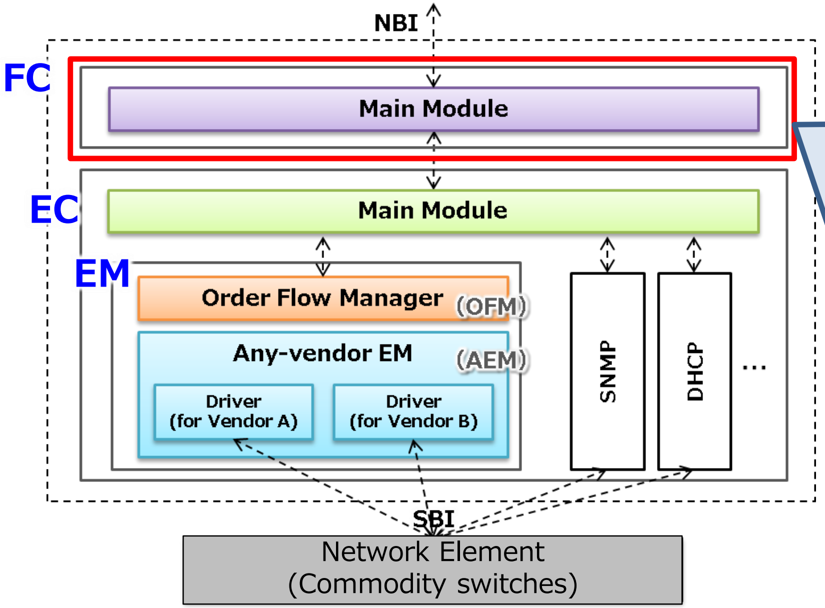
Controller Software component

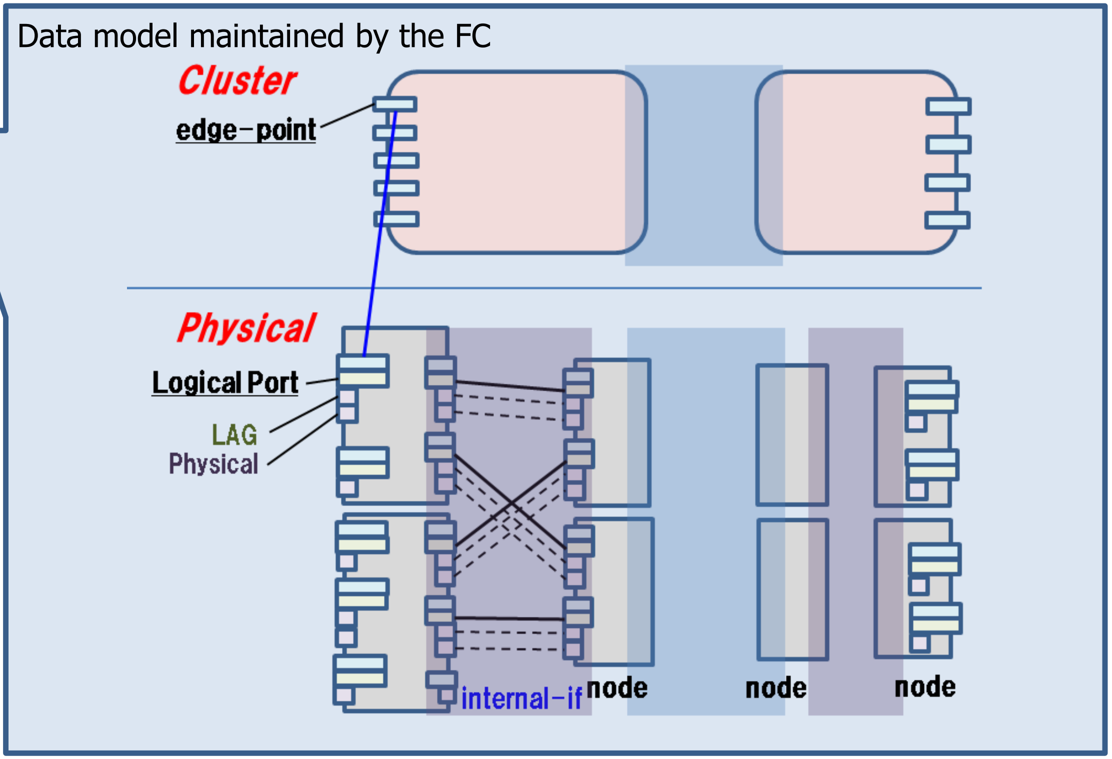
- ◆ Controller uses several OSS software for functions NOT included in the controller main module.



Fabric Controller (FC)

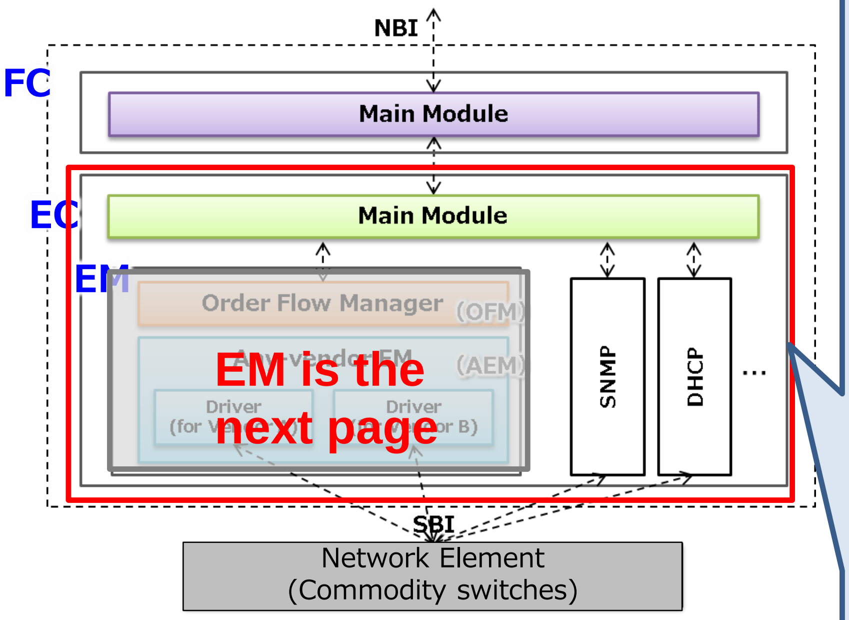
- ◆ **FC provides Network abstraction, management and the interface for northbound systems.**
  - □ **Network Abstraction : Show multiple switches as one logical switch**. “Edge-point” is defined in order to indicate the port (unique ID among a single cluster). This enables to hide the physical information to northbound.
  - □ Network Management : Maintain the network topology with logical information (edge-points).
    - ➢ Interface types and vendor-specific information is hidden by the EC.
    - ➢ EC does the mapping of physical port to logical port, FC does the mapping of logical port to edge-point.





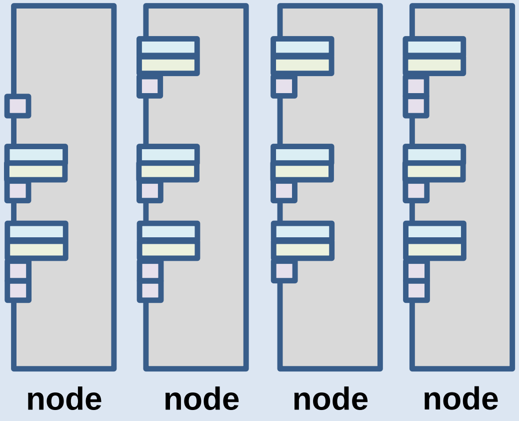
Element Controller (EC)

- **EC provides physical-logical mapping, concealment of vendor-specific information**. It also provides control interface to switch devices.
  - □ physical-logical mapping : Maintain mapping of physical ports to logical ports.
  - □ Concealment of vendor-specific information : Consolidate the difference between vendors MIB. This is injected via REST IF.



SBI protocols: SNMP Trap/MIB, Netconf, BGP

Data model maintained in the EC



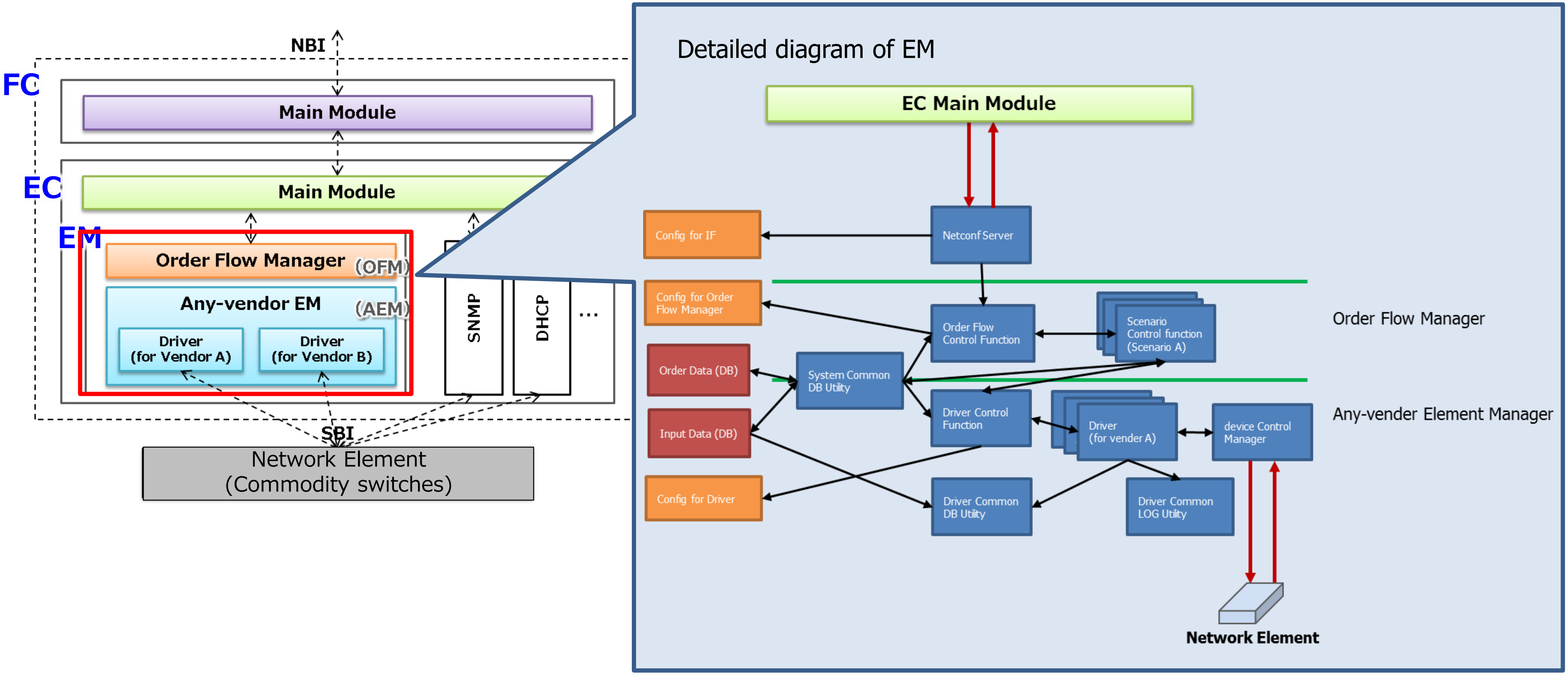
Does not maintain topology info

Physical-logical mapping table

Device	Logical Port	IF name
A	Port1	xe-0/0/1
	Port2	xe-0/0/2
B	port1	TenGigE0/0/0/1
	port2	TenGigE0/0/0/2
	⋮	

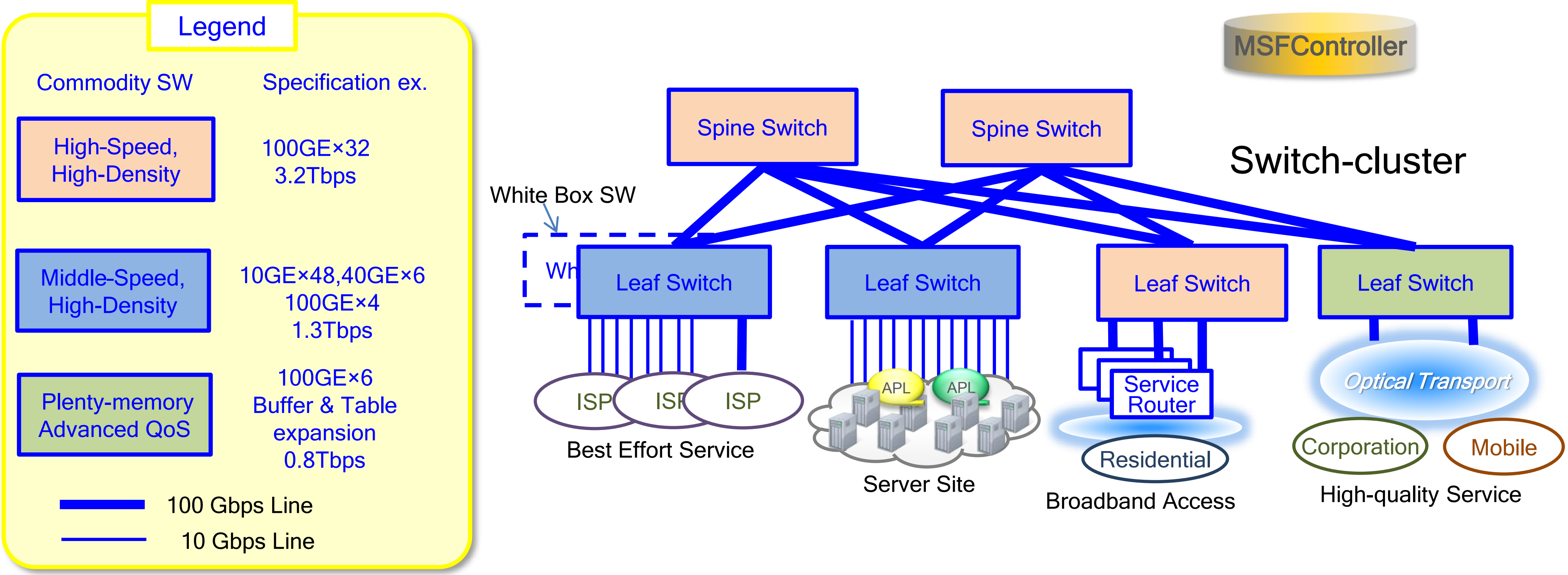
Element Manager (EM)

- ◆ **Concealment of vendor-specific configurations** and order flow management. It also provides control interface to switch devices.
  - □ **Concealment of vendor-specific configurations : Enabled by drivers implemented for each vendor products.**
  - □ Order Flow Management : Manage the configuration to multiple devices with one single transaction. Executes roll-back in case of error.



Physical network design (physical topology and applicable switches)

- For supporting flexible scale(easy scale up/down), MSF adopts CLOS(Leaf & Spine) topology.
- Various types of commodity switches can be applicable for MSF according to functions and port scale.



Legend

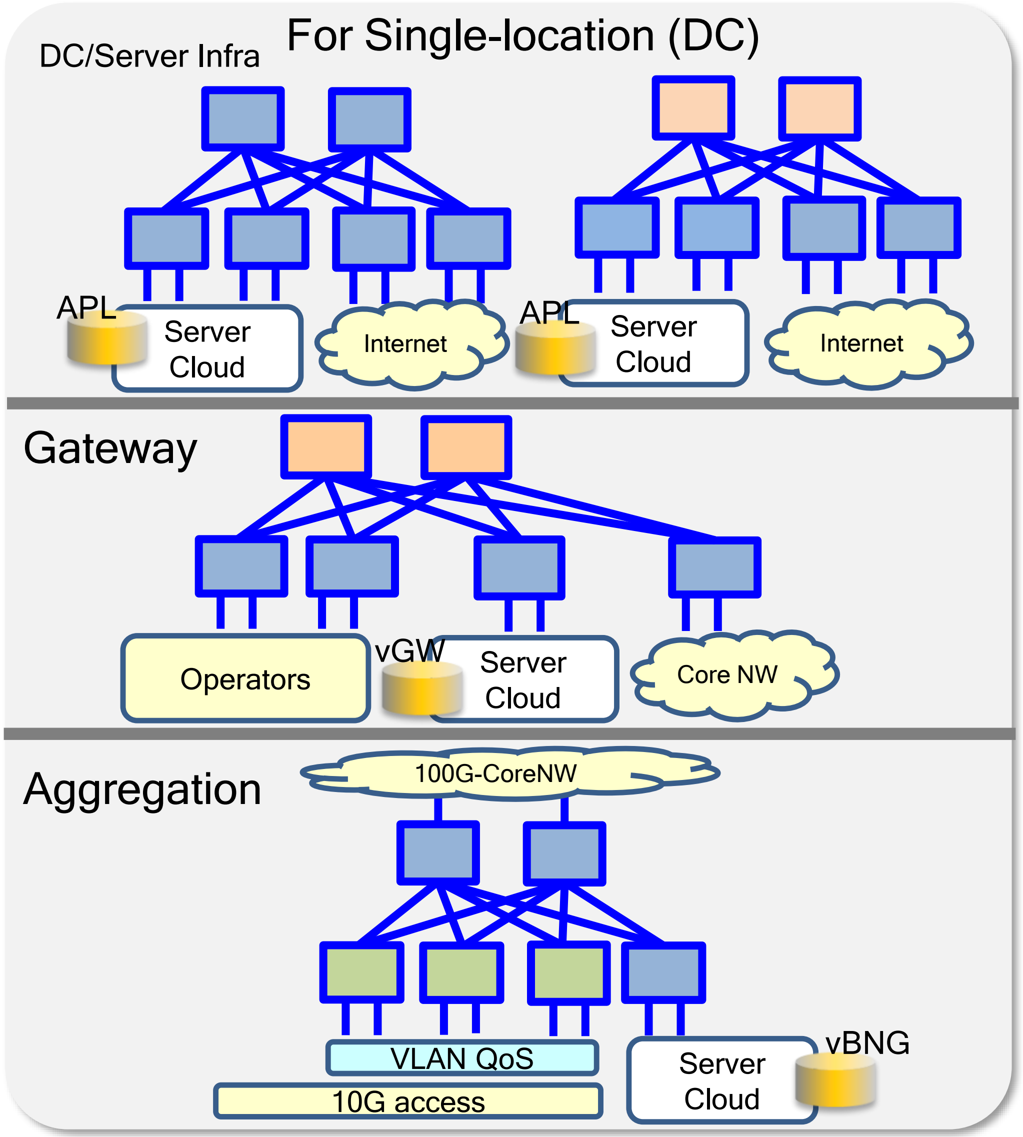
Commodity SW	Specification ex.
High-Speed, High-Density	100GE×32 3.2Tbps
Middle-Speed, High-Density	10GE×48,40GE×6 100GE×4 1.3Tbps
Plenty-memory Advanced QoS	100GE×6 Buffer & Table expansion 0.8Tbps

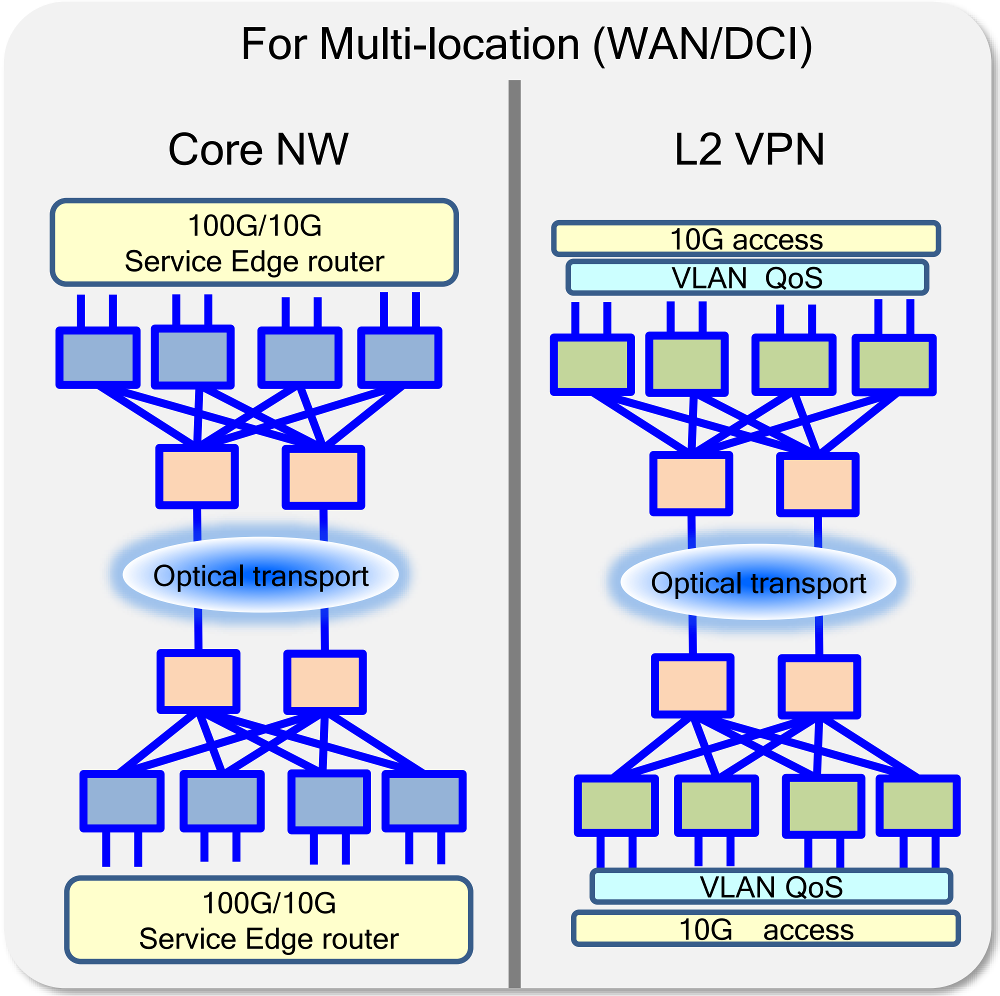
100 Gbps Line

10 Gbps Line

Image of multi-types of switches and offered services

Network configuration patterns for some domains




For Multi-location (WAN/DCI)
Core NW
100G/10G Service Edge router
Optical transport
100G/10G Service Edge router
L2 VPN
10G access
VLAN QoS
Optical transport
VLAN QoS
10G access

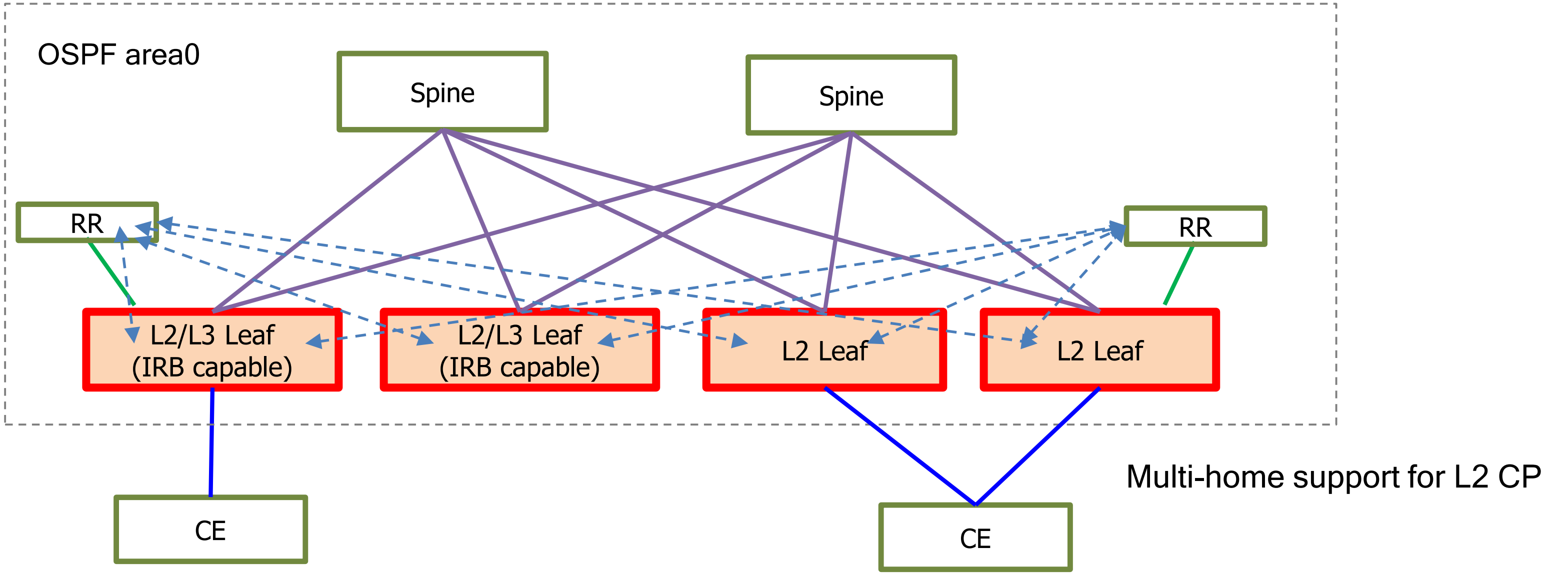
High-Speed, High-Density

Middle-Speed, High-Density

Plenty-memory Advanced QoS

Logical network design (for single location)

- ◆ Basically MSF makes use of standard based technology for multi-vender support.
- ◆ For underlay configuration;
  - □ Only OSPF is used for L2/L3 Leaf and L2 Leaf pattern, data-plane is used VXLAN.
  - □ OSPF and LDP is used for L3 Leaf pattern , data-plane is used MPLS.
- ◆ For overlay configuration;
  - □ EVPN is used for L2/L3 Leaf and L2 Leaf pattern (EVPN service)
  - □ MP-BGP is used for L3 Leaf pattern (BGP/MPLS VPN service)



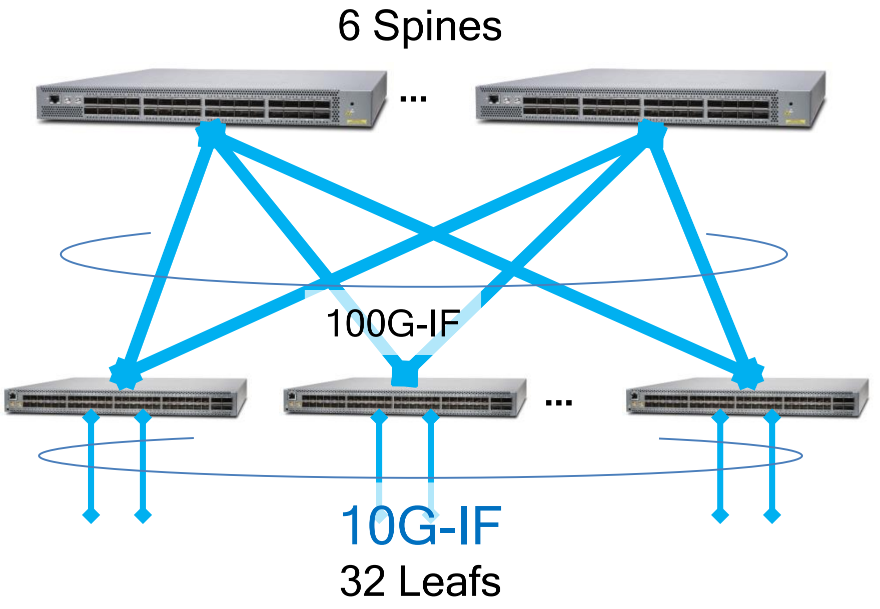
<L2/L3(IRB) and L2 Leaf pattern>

Example of maximum number of ports

◆The maximum number of 10G/100G ports that can be used as downlink in 1 cluster is as follows.

(no oversubscription, and no redundancy)

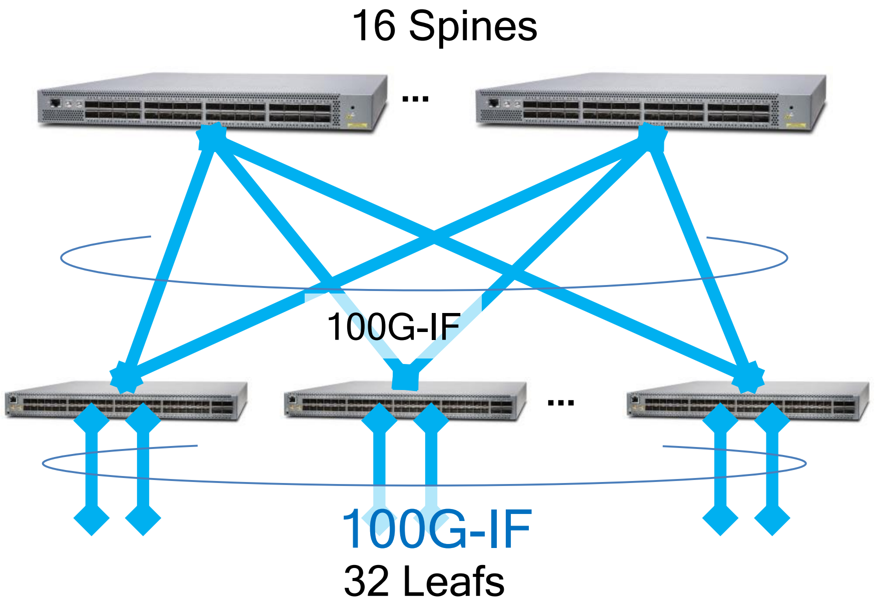
When using the 10G interface



Number of downlink ports: 48 ports/Leaf

<u>48 ports * 32 leafs = 1,536 ports</u>

When using the 100G interface



Number of downlink ports: 16 ports/Leaf

<u>16 ports * 32 leafs = 512 ports</u>

Controller functions

Category	Function	Description
Model information management	Register model information	**Register the switch model to be used in controller** You can put the model name, OS version, ID of each physical interface, and capability ..etc.
Underlay management	Switch cluster control	**Add and Delete Switch cluster** To add multiple switch clusters, add the switch cluster to the controller.
	Leaf/Spine control	**Add and Delete Leaf-node or Spine-node** You can automatically add the nodes into the network by registering the link connection information in the controller, and turning on the node.
	Interface control	**Physical<->Link-Aggregation or Breakout-IF** You can convert physical-IF to LAG-IF or Breakout-IF, or check the information on each interface
	edge-point control	**Specify any interfaces as edge-point** The end point to which end-user connects to slice, called connection-point; CP, can be created on the edge-point.
	Switch software upgrade	**Upgrading switch software** You can upgrade switch software automatically and planned.
Overlay management	Slice control	**Create/delete L2/L3 slice, also called VPN**
	CP control	**Create/delete end point to which end-user connects to slice**
	QoS control	**Control of traffic flow rate, QoS value** - Limit traffic flow by CP. - Schedule the traffic flow out in each CP. - Remark the QoS value (AF3/AF2/AF1/BE) for each slice.

Controller functions

Category	Function	Description
Network Operation	Fault detection	**Detection of link or switch failure** You can detect failures at physical, link-aggregation, or breakout-Ifs.
	Reachability visualization	**Reachability monitoring for each CP-CP pair** Monitor all reachability between CPs belonging each slice (VPN). Notify the information when the reachability between a specific CP-CP pair changes.
	Traffic measurement	**Measurement the traffic volume of each interface** You can specify the traffic on each interface.
	Traffic notification	**Notification of traffic volume using threshold** When you set the traffic threshold in any interfaces, notify the interface exceeding the threshold value.
	Simplified switch-exchange operation	**Controller supports the switch exchange operation** When replacing a failed switch, restore the setting of the switch before failure to the replacing one.
	Traffic route control	**Changing traffic route** You can control traffic route by changing routing costs.
Controller Operation	Controller state	**The state of the controller** You can acquire CPU and memory utilization.
	Controller log	**Get the controller log** You can get the controller processing log.
	Controller software Upgrade	**Upgrading controller software**

Controller API

For details, refer to multi-service-fabric/fabric-controller/API

Class	Group	Interface description	Method
Common	Processing request	Getting list of operational state	GET
		Getting information of detailed operation state	GET
	Controller status confirmation	Getting controller state	GET
	Controller log	Getting controller log	GET
Underlay management	Equipment-type information management	Registering equipment information	POST
		Getting equipment list in switch cluster	GET
		Getting equipment information	GET
		Deleting equipment information	DELETE
	Switch-cluster management	Adding Switch-cluster	POST
		Getting list of Switch-cluster	GET
		Getting information of Switch-cluster	GET
		Deleting Switch-cluster	DELETE
	Node information	Getting list of nodes	GET
	Leaf management	Adding Leaf-node	POST
		Getting list of Leaf-nodes	GET
		Getting information of Leaf-node	GET
		Deleting Leaf-node	DELETE
		Updating Leaf-node	PUT
	Spine management	Adding Spine-node	POST
		Getting list of Spine-nodes	GET
		Getting information of Spine-node	GET
		Deleting Spine-node	DELETE

Controller API

For details, refer to multi-service-fabric/fabric-controller/API

Class	Group	Interface description	Method
Underlay management	RR (BGP Route Reflector) management	Getting list of RR-node	GET
		Getting information of RR-node	GET
	Interface information	Getting list of interfaces	GET
	Interface management (Physical interface)	Getting list of physical interfaces	GET
		Getting information of physical interface	GET
		Updating physical interface	PUT
	Interface management (Breakout interface)	Creating or deleting breakout interface	PATCH
		Getting list of breakout interfaces	GET
		Getting information of breakout interface	GET
	Interface management (Internal-link interface)	Getting list of internal-link interfaces	GET
		Getting information of internal-link interface	GET
	Interface management (Link aggregation interface)	Creating Link-aggregation interface	POST
		Getting list of Link-aggregation interfaces	GET
		Getting information of Link-aggregation interface	GET
		Deleting information of Link-aggregation interface	DELETE
	Interface management (Inter-cluster link interface)	Creating inter-cluster link interface	POST
		Getting list of inter-cluster link interfaces	GET
		Getting information of inter-cluster link interface	GET
		Deleting inter-cluster link interface	DELETE
	Edge point management	Creating edge-point	POST
		Getting list of edge-points	GET
		Getting information of edge-point	GET
		Deleting edge-point	DELETE

Controller API

For details, refer to multi-service-fabric/fabric-controller/API

Class	Group	Interface description	Method
Overlay management	Slice	Creating Slice	POST
		Changing Slice	PUT
		Deleting Slice	DELETE
		Getting information of Slice	GET
		Getting list of Slices	GET
	CP	Creating or deleting CP	PATCH
		Creating CP	POST
		Changing CP	PUT
		Deleting CP	DELETE
		Getting information of CP	GET
		Getting lists of CP	GET
		Creating or deleting static route	PATCH
Traffic information	Traffic information	Getting list of IF traffic	GET
		Getting IF traffic	GET
		Getting list of CP traffic	GET
		Getting CP traffic	GET
Fault detection	Failure detection	Getting list of failures	GET
Priority route Management	Internal-link priority	Changing internal-link interface priority	PUT
		Getting information of internal-link interface priority	GET
		Getting list of internal-link interface priority	GET
	Switch priority	Adding or deleting switch to preferred switch group	PATCH
		Getting information of preferred switch group	GET

New API

Controller API

For details, refer to multi-service-fabric/fabric-controller/API

Class	Group	Interface description	Method
Filter management	Filter information	Creating or Deleting filter by physical interface	PATCH
		Getting list of filter by physical interface	GET
		Getting information of filter by physical interface	GET
		Creating or Deleting filter by Link-aggregation interface	PATCH
		Getting list of filter by Link-aggregation interface	GET
		Getting information of filter by Link-aggregation interface	GET
Switch software Upgrade	Bypass traffic to switch	Bypass traffic to switch	PUT
		Getting information of switch whose traffic is diverted	GET
	Execute Switch OS upgrade	Execute Switch OS upgrade	POST
		Cancel Switch OS upgrade	PUT
Controller software Upgrade	Software upgrade status check	Execute controller construction	PUT
		Controller construction exe status acquisition	GET
		Controller system switchover	POST

New API

Controller API(Notification)

For details, refer to multi-service-fabric/fabric-controller/API

Group	Interface description	Method
Common	Processing result	PUT
	controller status	PUT
Traffic information	Traffic information	PUT
Failure detection	Failure information	PUT

Supported products and type

model	Tested OS version	type			
		Spine	L3-Leaf	Border-Leaf	L2-Leaf
QFX5100-48S	Junos14.1X53-D46 flex	Yes	Yes	Yes	Yes
QFX5100-24Q	Junos14.1X53-D46 flex	Yes	-	-	-
QFX5200-32C	Junos15.1X53-D30Flex	Yes	Yes	Yes	-
NCS5001	IOS-XR6.1.2	-	Yes	-	-
NCS5011	IOS-XR6.1.2	Yes	-	-	-
NCS5501-SE	IOS-XR6.3.1	-	Yes	Yes	-
AS5812	OcNOS IPBASE-1.3.1	-	-	-	Yes
AS7712	OcNOS IPBASE-1.3.1	Yes	-	-	-
QFX5110-48S	Junos17.4R1	-	-	-	Yes
AS5812	Cumulus 3.6.0	-	-	-	Yes
Dell S6000	Cumulus 3.6.0	Yes	-	-	-
AS5812	Beluganos 0.3	Yes	-	-	-

Yes : Supported by MSF
- : Not Supported by MSF

Supported functions(Spine)

model		QFX5100-48S	QFX5100-24Q	QFX5200-32C	NCS5011	AS7712	S6000	AS5812
Tested OS Version		Junos14.1X53-D46 flex	Junos14.1X53-D35	Junos15.1X53-D30 Flex-image	IOS-XR 6.1.2	OcNOS IPBASE-1.3.1	Cumulus	Beluganos
Interface	1G-LX	Yes	-	-	-	-	-	-
	10G-LR (Non Breakout)	Yes	-	-	-	-	-	Yes
	40G-SR4	Yes	Yes	Yes	Yes	Yes	Yes	-
	100G-SR4	-	-	Yes	Yes	-	Yes	-
	10G-SR*4 (Breakout-IF)	-	Yes	Yes	Yes	-	-	-
Control	Netconf	Yes	Yes	Yes	Yes	-	-	Yes
	CLI	-	-	-	-	-	Yes	-
MIB	L3 physical	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	L3VLAN	-	-	Yes	-	-	-	-
	L2 physical	Yes	-	-	-	Yes	-	-
how to configure	ZTP	Yes	Yes	Yes	Yes	-	Yes	-
	manual	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Traffic Engineering	Priority node	Yes	Yes	Yes	-	-	Yes	Yes

New function

Yes : Supported by MSF
- : Not Supported by MSF

Supported functions(L3-Leaf/Border-Leaf)

model		QFX5100-48S	QFX5200-32C	NCS5001	NCS5501-SE
Tested OS Version		Junos14.1X53-D46 flex	Junos15.1X53-D30Flex	IOS-XR6.1.2	IOS-XR6.3.1
L3CP type	direct(v4)	Yes	Yes	-	Yes
	direct(v6)	Yes	-	-	Yes
	VRRP(v4)	Yes	Yes	-	Yes
	VRRP(v6)	Yes	-	-	Yes
	BGP(v4)	Yes	Yes	Yes	Yes
	BGP(v6)	Yes	-	-	Yes
	static(v4)	Yes	Yes	Yes	Yes
	static(v6)	Yes	-	-	Yes
Interface	1G-LX	Yes	-	Yes	Yes
	10G-LR (Non Breakout)	Yes	-	Yes	Yes
	40G-SR4	Yes	Yes	Yes	Yes
	100G-SR4	-	Yes	Yes	Yes
	10G-SR*4 (Breakout-IF)	-	Yes	-	-
Control	Netconf	Yes	Yes	Yes	Yes
	CLI	-	-	-	-
MIB	L3 physical	Yes	Yes	Yes	Yes
	L3VLAN	-	Yes	-	-
	L2 physical	Yes	-	-	-
how to configure	ZTP	Yes	Yes	Yes	Yes
	manual	Yes	Yes	Yes	Yes

Yes : Supported by MSF
 - : Not Supported by MSF

Supported functions(L2-Leaf)

model		QFX5100-48S	AS5812	QFX5110-48S	AS5812
Tested OS Version		Junos14.1X53-D46 flex	OcNOS IPBASE-1.3.1	Junos17.4R1	Cumulus 3.6.0
L2CP type	EVPN (multi-home)	Yes	-	Yes	-
	EVPN (MC-LAG)	-	-	-	Yes
	EVPN (single)	Yes	Yes	Yes	Yes
	IRB (L2/L3 combination)	-	-	Yes	Yes
Interface	1G-LX	Yes	Yes	Yes	Yes
	10G-LR (Non Breakout)	Yes	Yes	Yes	Yes
	40G-SR4	Yes	Yes	Yes	Yes
	100G-SR4	-	-	Yes	-
	10G-SR*4 (Breakout-IF)	-	-	-	-
Control	Netconf	Yes		Yes	-
	CLI	-	Yes	-	Yes
MIB	L3 physical	Yes	Yes	Yes	Yes
	L3VLAN	-	-	-	-
	L2 physical	Yes	Yes	Yes	-
	L2VLAN	-	-	Yes	Yes
how to configure	ZTP	Yes	-	Yes	Yes
	manual	Yes	Yes	Yes	Yes
IRB type	direct(v4)	-	-	Yes	Yes
	direct(v6)	-	-	-	-
ACL	MAC address	-	-	Yes	Yes (Can not be combined)
	IP address	-	-	Yes	Yes (Can not be combined)
	L4 port	-	-	Yes	Yes (Can not be combined)
Traffic Engineering	Priority node	Yes	-	Yes	Yes

New function

Yes : Supported by MSF
- : Not Supported by MSF

Notice

- All company names and product names mentioned in this document are registered trademarks or trademarks of their respective companies.
- This document is not sponsored by, endorsed by or affiliated with Cisco Systems, Inc. Cisco, the Cisco logo, Cisco Systems and Cisco IOS are registered trademarks or trademarks of Cisco Systems, Inc. and/or its affiliates in the United States and certain other countries.