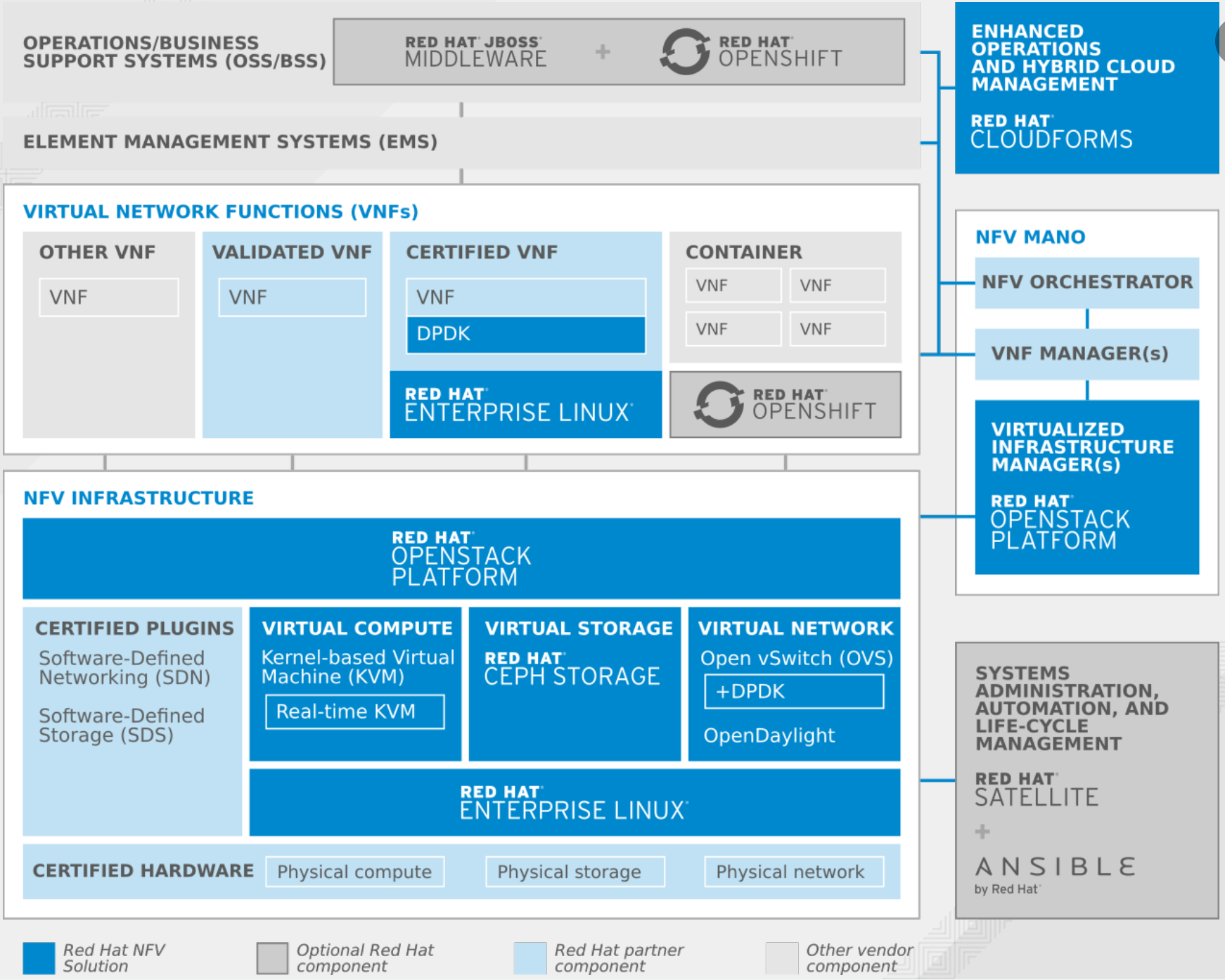
FAST DATAPATH & HYPER CONVERGED INFRASTRUCTURE IN RED HAT OPENSTACK PLATFORM

Won Young Choi

Senior Solution Architect

Red Hat, Inc.


OPERATIONS/BUSINESS SUPPORT SYSTEMS (OSS/BSS)
RED HAT JBOSS MIDDLEWARE
+
RED HAT OPENSHIFT
ELEMENT MANAGEMENT SYSTEMS (EMS)
VIRTUAL NETWORK FUNCTIONS (VNFs)
OTHER VNF
VNF
VALIDATED VNF
VNF
CERTIFIED VNF
VNF
DPDK
RED HAT ENTERPRISE LINUX
CONTAINER
VNF
VNF
VNF
VNF
RED HAT OPENSHIFT
NFV INFRASTRUCTURE
RED HAT OPENSTACK PLATFORM
CERTIFIED PLUGINS
Software-Defined Networking (SDN)
Software-Defined Storage (SDS)
VIRTUAL COMPUTE
Kernel-based Virtual Machine (KVM)
Real-time KVM
VIRTUAL STORAGE
RED HAT CEPH STORAGE
VIRTUAL NETWORK
Open vSwitch (OVS)
+DPDK
OpenDaylight
RED HAT ENTERPRISE LINUX
CERTIFIED HARDWARE
Physical compute
Physical storage
Physical network
ENHANCED OPERATIONS AND HYBRID CLOUD MANAGEMENT
RED HAT CLOUDFORMS
NFV MANO
NFV ORCHESTRATOR
VNF MANAGER(s)
VIRTUALIZED INFRASTRUCTURE MANAGER(s)
RED HAT OPENSTACK PLATFORM
SYSTEMS ADMINISTRATION, AUTOMATION, AND LIFE-CYCLE MANAGEMENT
RED HAT SATELLITE
+
ANSIBLE
by Red Hat
Red Hat NFV Solution
Optional Red Hat component
Red Hat partner component
Other vendor component

NFV & NETWORK GUIDE

https://access.redhat.com/documentation/en/red-hat-openstack-platform/11/

Network Functions Virtualization

Network Functions Virtualization Product Guide

Overview of the Network Functions Virtualization (NFV)

Network Functions Virtualization Planning and Prerequisites Guide

Planning for NFV in Red Hat OpenStack Platform

Network Functions Virtualization Configuration Guide

Configuring the Network Functions Virtualization (NFV) OpenStack deployment

Networking

Networking Guide

An advanced guide to OpenStack Networking

Red Hat OpenDaylight Product Guide

Overview of Red Hat OpenDaylight

OpenDaylight and Red Hat OpenStack Installation and Configuration Guide

NEUTRON NETWOKING

오픈스택 NEUTRON을 통해 네트워크 관리

VIRTIO

PCI-PASSTHROUGH

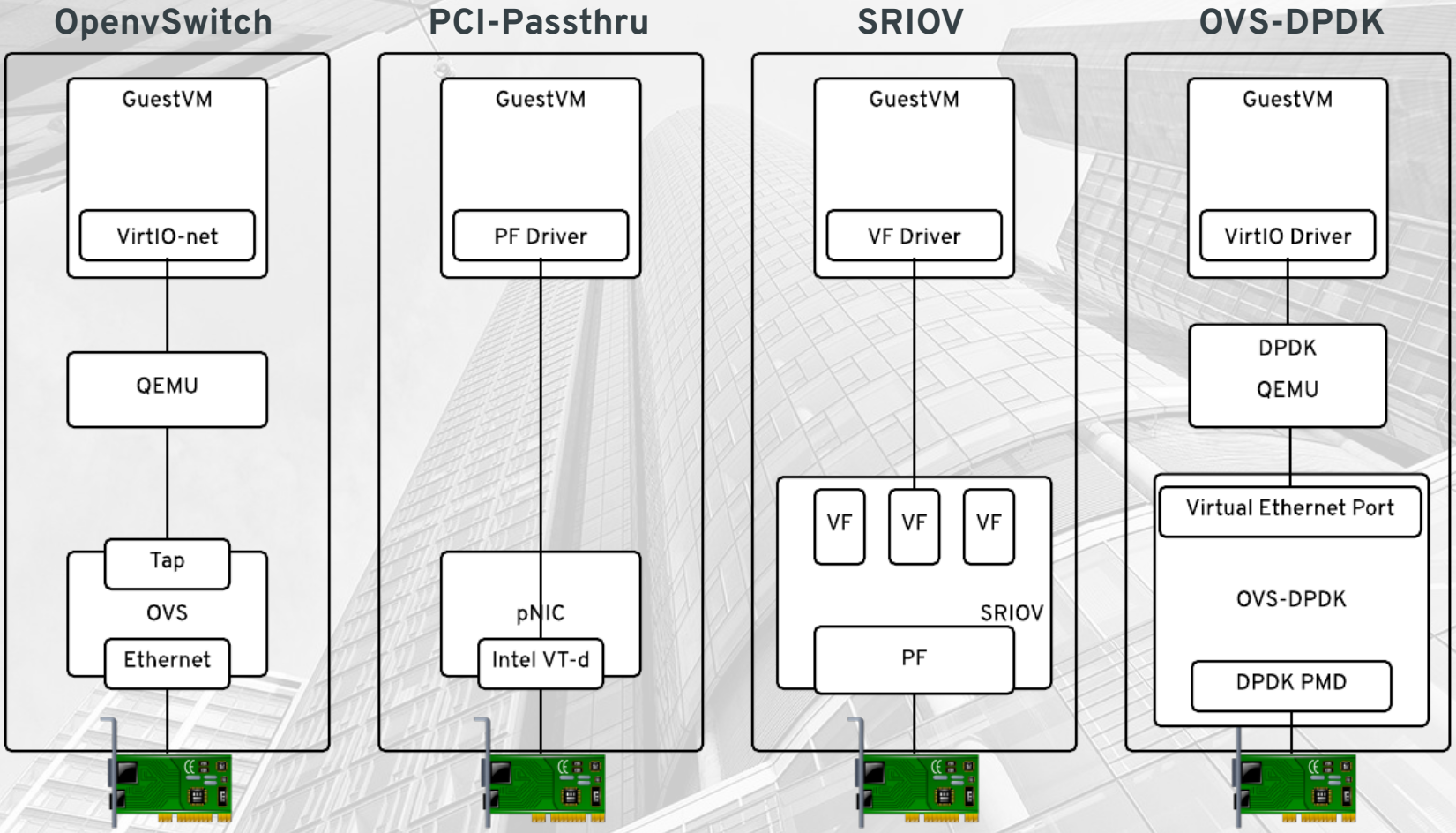
SRIOV

OVS-DPDK

Standard vSwitch	Hypervisor Bypass	DPDK accelerated vSwitch
virtio	PCI - Passthrough	OVS-DPDK
	SRIOV	

FAST DATAPATH

VNF 데이터플래인의 다양한 네트워크 연결 옵션



OSP DIRECTOR SUPPORTS

- OpenvSwitch, OVN, OVS-DPDK
- SRIOV
- Cisco n1kv, nexus-ucsm
- Fujitsu cfab, fossw
- Nokia Nuage
- BigSwitch
- OpenDaylight
- Midonet
- Contrail
- Plumgrid

TESTED NICS

SRIOV

- 10G Mellanox & Qlogic
- Intel

  82598, 82599, X520, X540, X550, X710, XL710, X722

OVS-DPDK

- Intel

  82598, 82599, X520, X540, X550, X710, XL710, X722

OVS-DPDK

DPDK-16.11 + OPENVSWITCH-2.7.2

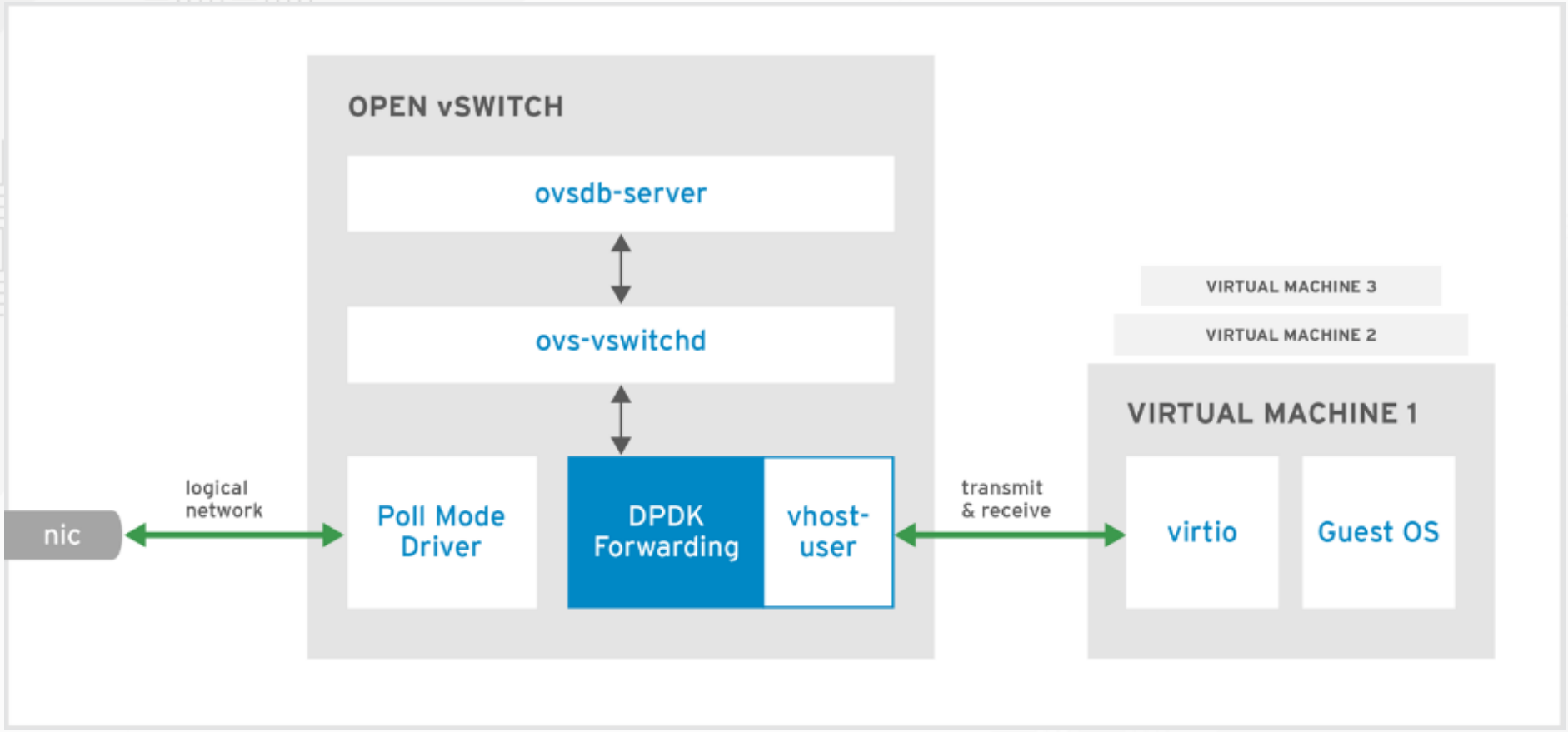
RHCDN channel : rhel-7-fast-datapath-rpms

```
dpdk-16.11-3.el7fdp.x86_64.rpm
dpdk-devel-16.11-3.el7fdp.x86_64.rpm
dpdk-doc-16.11-3.el7fdp.noarch.rpm
dpdk-tools-16.11-3.el7fdp.x86_64.rpm
driverctl-0.95-1.el7fdp.noarch.rpm
openvswitch-2.7.2-1.git20170719.el7fdp.x86_64.rpm
openvswitch-devel-2.7.2-1.git20170719.el7fdp.x86_64.rpm
openvswitch-ovn-central-2.7.2-1.git20170719.el7fdp.x86_64.rpm
openvswitch-ovn-common-2.7.2-1.git20170719.el7fdp.x86_64.rpm
openvswitch-ovn-docker-2.7.2-1.git20170719.el7fdp.x86_64.rpm
openvswitch-ovn-host-2.7.2-1.git20170719.el7fdp.x86_64.rpm
openvswitch-ovn-vtep-2.7.2-1.git20170719.el7fdp.x86_64.rpm
openvswitch-test-2.7.2-1.git20170719.el7fdp.noarch.rpm
python-openvswitch-2.7.2-1.git20170719.el7fdp.noarch.rpm
tuned-2.8.0-2.el7fdp.noarch.rpm
tuned-profiles-cpu-partitioning-2.8.0-2.el7fdp.noarch.rpm
```

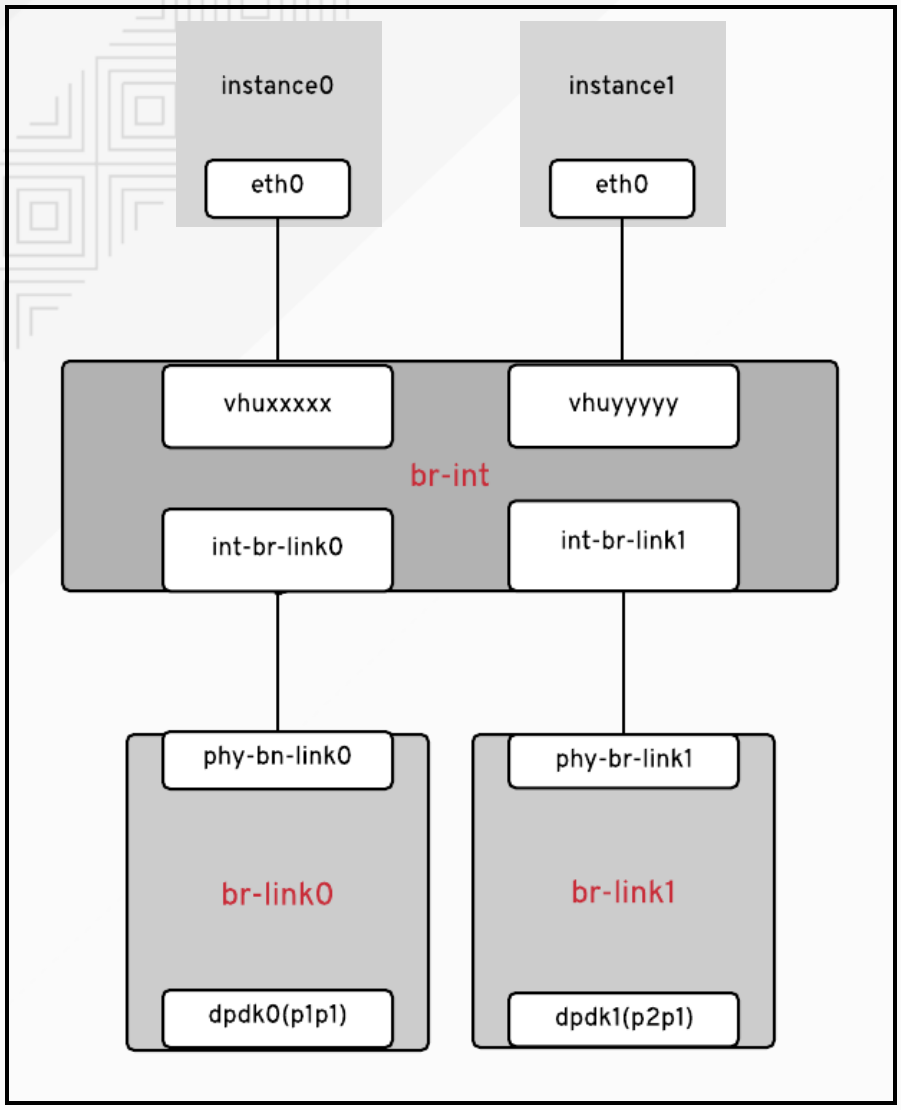
RHOSP 11 : openvswitch-**2.6.1**-13.git20161206.el7ost.x86_64.rpm

RHOSP10 : openvswitch-**2.6.1**-15.git20161206.el7ost.x86_64.rpm

OVS-DPDK 구조

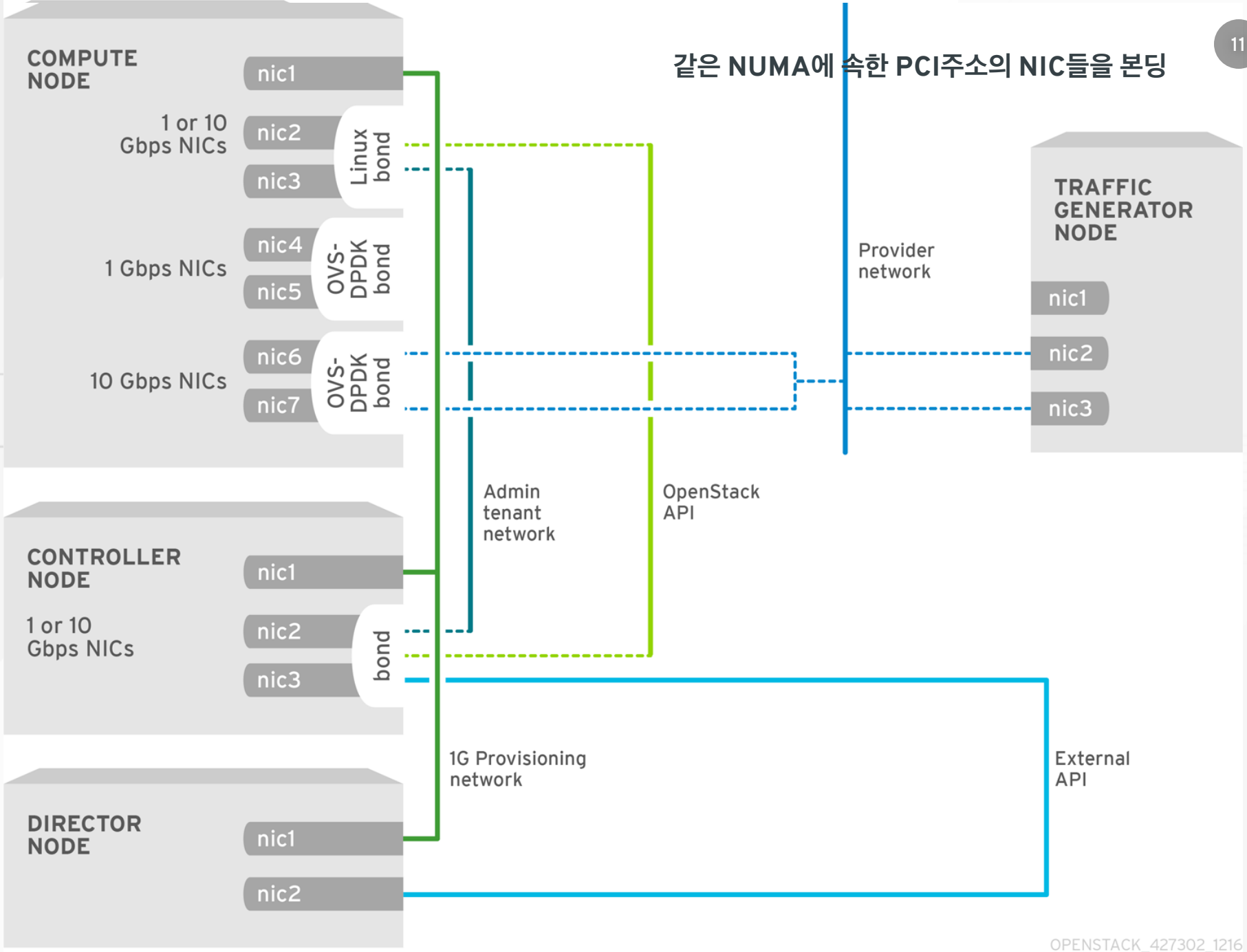

OPEN vSWITCH
ovsdb-server
ovs-vswitchd
Poll Mode Driver
DPDK Forwarding
vhost-user
nic
logical network
transmit & receive
VIRTUAL MACHINE 3
VIRTUAL MACHINE 2
VIRTUAL MACHINE 1
virtio
Guest OS

OVS-DPDK 브릿지 및 포트 구조

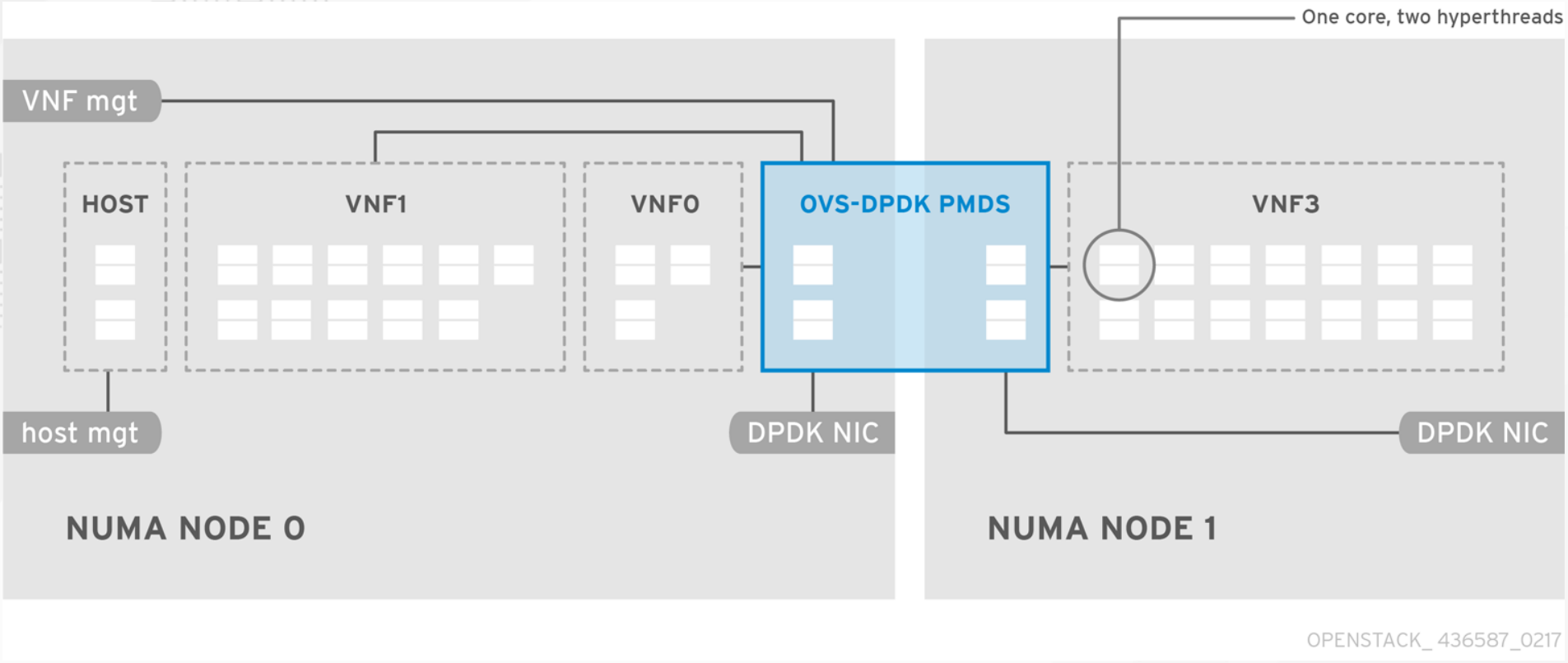



```
Bridge br-int
    fail_mode: secure
    Port "vhua115043e-e2"
        tag: 1
        Interface "vhua115043e-e2"
            type: dpdkvhostuser
    Port "vhuc5079cb0-2c"
        tag: 1
        Interface "vhuc5079cb0-2c"
            type: dpdkvhostuser
    Port br-int
        Interface br-int
            type: internal
    Port "int-br-link2"
        Interface "int-br-link2"
            type: patch
            options: {peer="phy-br-link2"}
    Port "int-br-link4"
        Interface "int-br-link4"
            type: patch
            options: {peer="phy-br-link4"}
    Port "int-br-link1"
        Interface "int-br-link1"
            type: patch
            options: {peer="phy-br-link1"}
    Port "vhu9521d78e-f1"
        tag: 2
        Interface "vhu9521d78e-f1"
            type: dpdkvhostuser
    Port "int-br-link3"
        Interface "int-br-link3"
            type: patch
            options: {peer="phy-br-link3"}
    Port "vhue2fbbf4c-3a"
        tag: 3
        Interface "vhue2fbbf4c-3a"
            type: dpdkvhostuser
Bridge "br-link1"
    Port "br-link1"
        Interface "br-link1"
            type: internal
    Port "dpdk0"
```

같은 NUMA에 속한 PCI주소의 NIC들을 본딩



VNF NUMA PARTITION 예시



OVS-DPDK CPU CORE 할당 예시

AllCores =
HostCpusList + NeutronDpdkCoreList + NovaVcpuPinSet

HostCpusList = Hypervisor OS + OVS processes (+ Ceph OSD)

HostIsolatedCoreList = NeutronDpdkCoreList + NovaVcpuPinSet

NIC1 FOR DPDK, 1 PHYSICAL CORE FOR PMD

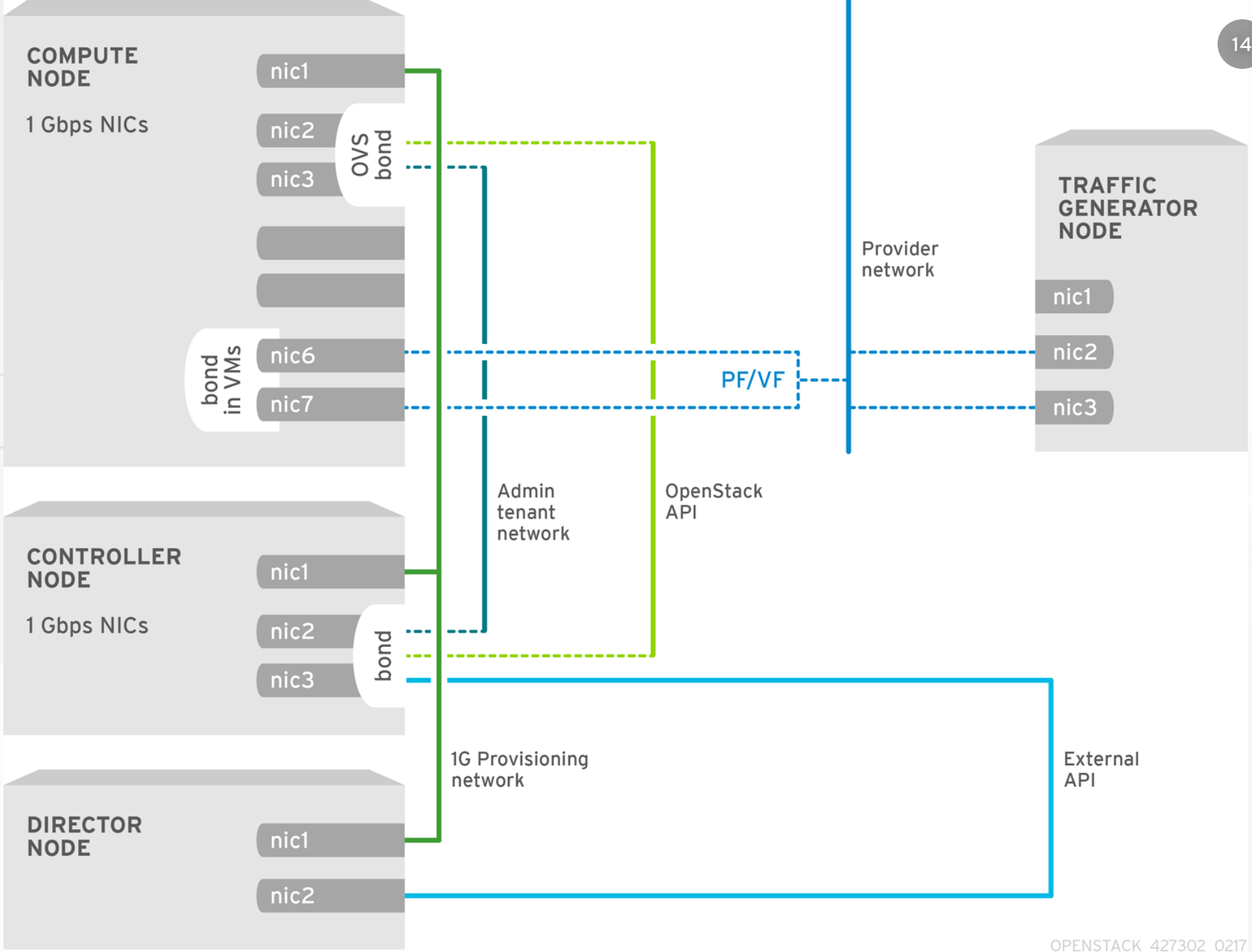
NeutronDpdkCoreList: "**2,3**,10,11"

NovaVcpuPinSet: "4,5,6,7,12,13,14,15"

NIC2 FOR DPDK, 2 PHYSICAL CORES FOR PMD

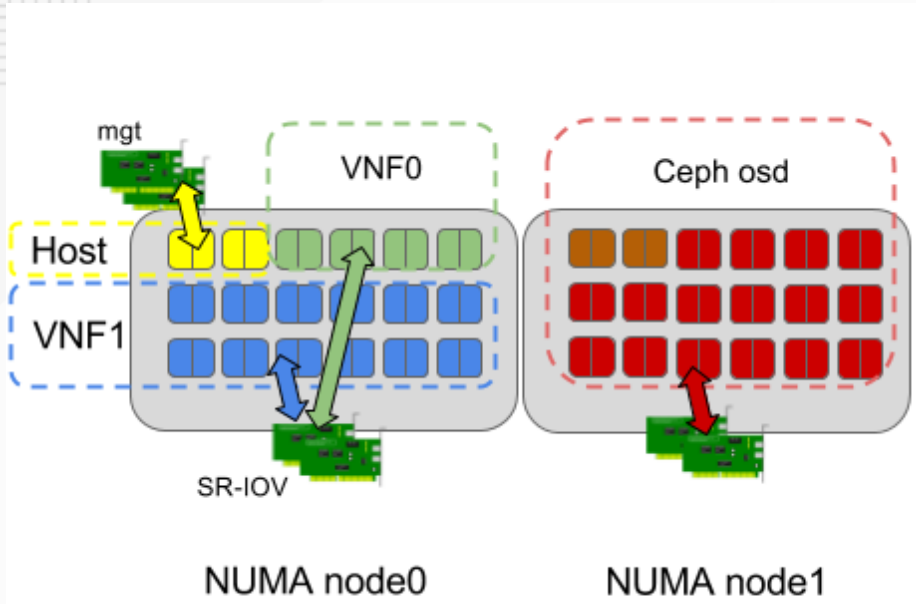
NeutronDpdkCoreList: "2,3,**10,11,12,13**"

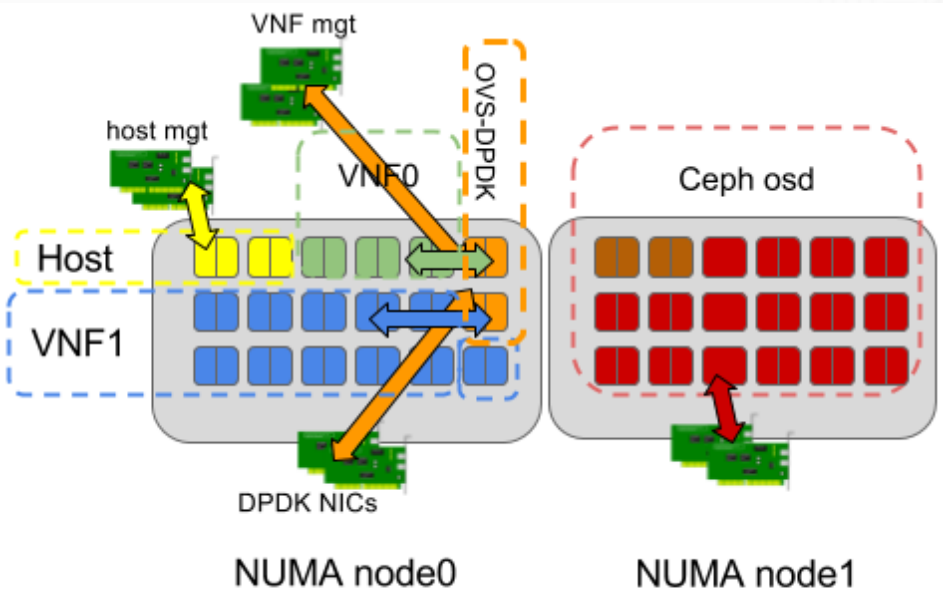
NovaVcpuPinSet: "4,5,6,7,14,15"


COMPUTE NODE
1 Gbps NICs
nic1
nic2
nic3
OVS bond
bond in VMs
nic6
nic7
CONTROLLER NODE
1 Gbps NICs
nic1
nic2
nic3
bond
DIRECTOR NODE
nic1
nic2
Admin tenant network
OpenStack API
1G Provisioning network
Provider network
PF/VF
TRAFFIC GENERATOR NODE
nic1
nic2
nic3
External API

HYPER CONVERGED INFRASTRUCTURE

- Compute 노드 = Storage 노드
- NUMA 아키텍쳐 활용
  - 스토리지 전용 NUMA 노드 할당 및 NIC 맵핑
  - CPU 코어의 분산 배치 설계 고려(vcpu_pin_set & isolcpus)
- 비용효율적인 아키텍쳐
- 디렉터를 통해 구성





하드웨어 토폴로지 확인

```
[stack@c10-h01-r730xd ~]$ lstopo-no-graphics
Machine (128GB total)
  NUMANode L#0 (P#0 64GB)
      ...
    HostBridge L#0
      PCIBridge
        PCI 1000:005d
          Block(Disk) L#0 "sda"
          Block(Disk) L#1 "sdb"
          ...
      PCIBridge
        PCI 8086:1572
          Net L#17 "em1"
        PCI 8086:1572
          Net L#18 "em2"
      ...
      ...
  NUMANode L#1 (P#1 64GB)
     ...
    HostBridge L#11
      PCIBridge
        PCI 8086:1572
          Net L#23 "p4p1"
        PCI 8086:1572
          Net L#24 "p4p2"
[stack@c10-h01-r730xd ~]$
```

RED HAT OPENSTACK PLATFORM

2 sockets 기반의 서브스크립션 정책 , RHEL7.x 호스트 OS 기본 포함

RED HAT CLOUDFORMS
RHOSP 관리 용도로만
외부 PostgreSQL 연결 가능

RED HAT OPENSTACK PLATFORM
controller
(director, object storage,compute **without RHEL guest**)
compute
(unlimited RHEL guest)

RED HAT CEPH STORAGE
서브스크립션 소진없이
3 OSD nodes, raw 64TB

FD.io Integrations