



# OpenStack NFV insights

Franck Baudin, Principal Product Manager - OpenStack NFV

July 6th, 2017 - OpenStackFR meetup



# Agenda

## OpenStack NFV insights

- OpenStack NFV: under the hood
- Installing OpenStack for NFV via TripleO
- OpenSource communities involved around NFV
- Key takeaways



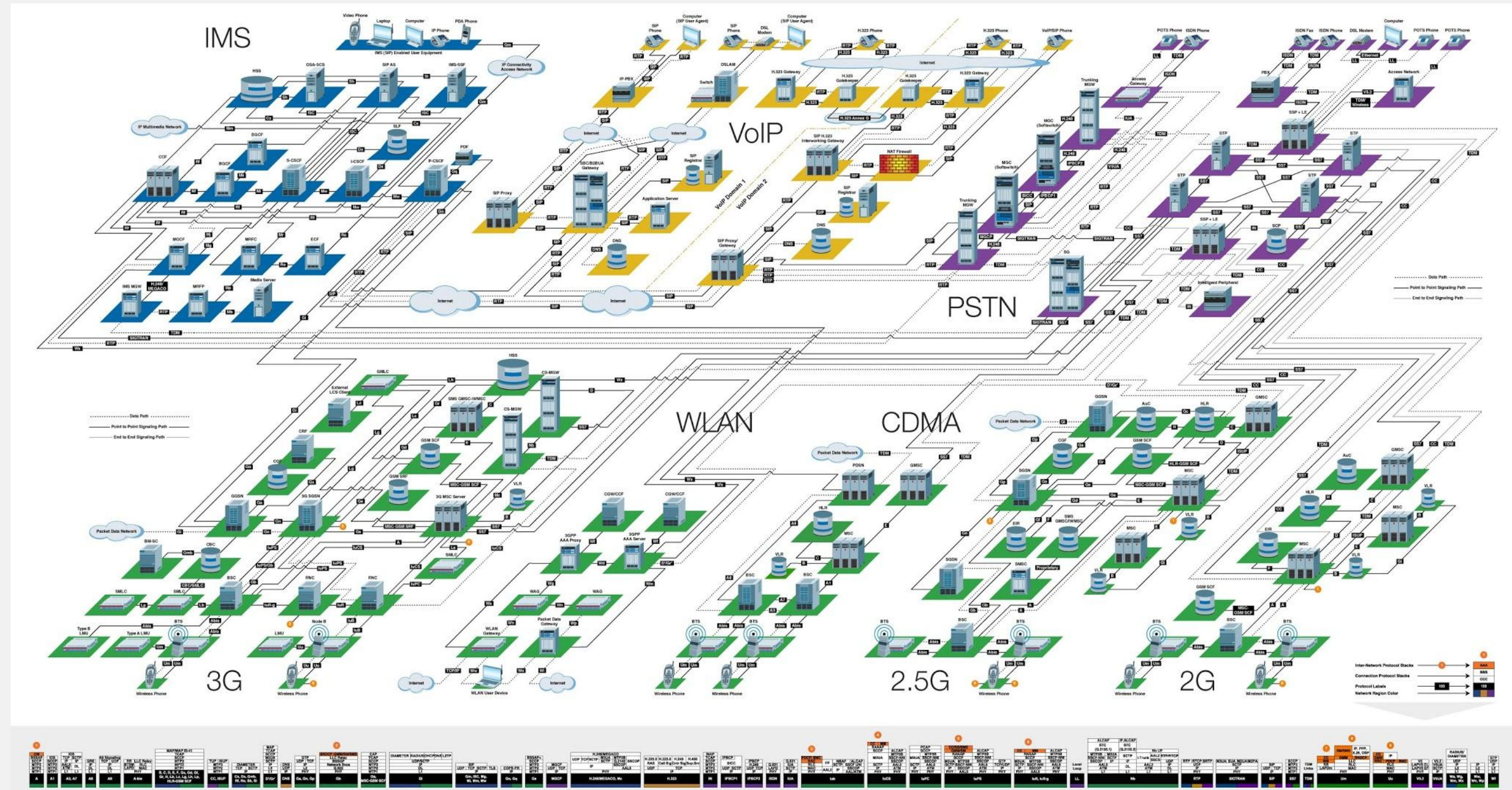
# OpenStack NFV: under the hood



# Mobile networks deployment today/yesterday



Picture credits: [wikipedia](https://en.wikipedia.org/wiki/File:Mobile_phone.jpg)





# SR-IOV NFV deployment

1 ATCA chassis  
~ 1 VNF

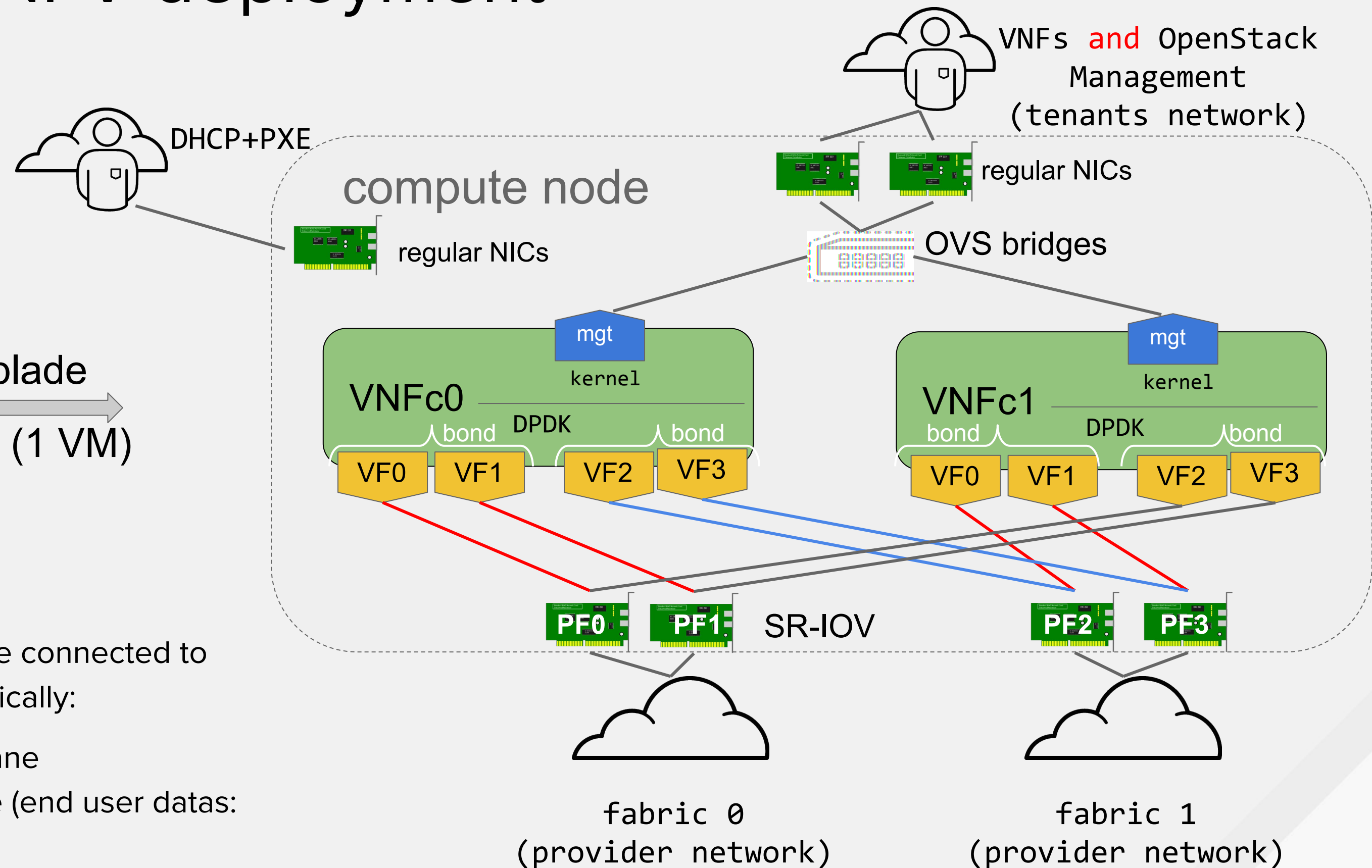


Picture credits: [wikipedia](https://en.wikipedia.org/wiki/ATCA)

1 ATCA blade  
~ 1 VNFC (1 VM)

Depending on the VNF, it will be connected to one or more fabric network, typically:

- One fabric for control plane
- One fabric for user plane (end user datas: mobile traffic, web, ...)



# SR-IOV

## Host/VNFs guests resources partitioning

Typical 18 cores per node dual socket compute node ([E5-2599 v3](#))

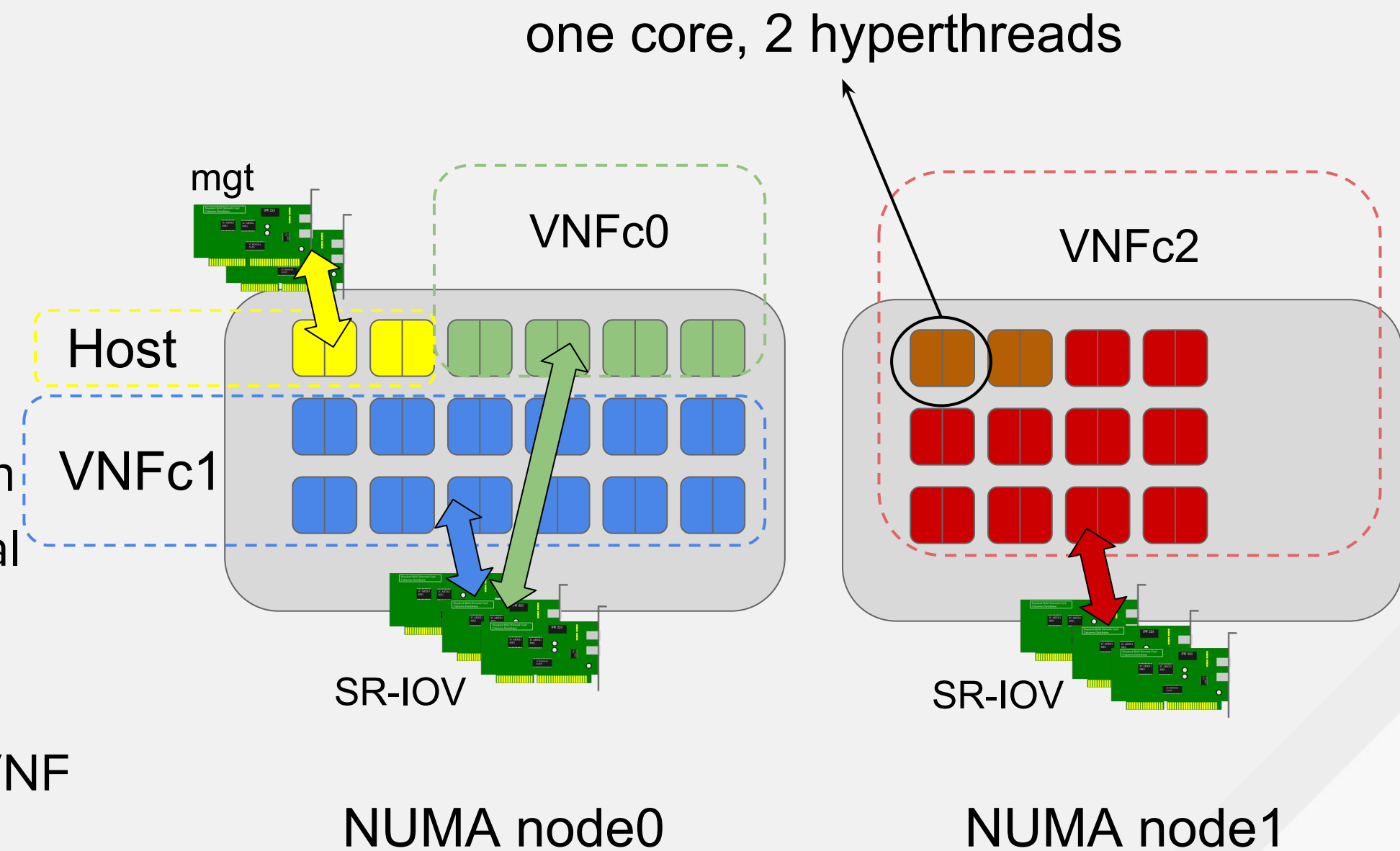
This looks like RT but is not RT, just partitioning

SR-IOV interfaces bonding handled by the VNF

All host IRQs routed on host cores

All VNFx cores dedicated to VNFx

- Isolation from others VNFs
- Isolation from the host
- 21 Mpps/core with zero frame loss, 12 hours run (I/O bound due to Niantic PCIe x4 switch): equal to bare-metal performances
- The number above show that **virtualization/SR-IOV cost is null**, and that the VNF is not preempted/interrupted.



# Tuned example: CPU0 for Linux CPU1-31 for OVS-DPDK and VMs

```
[root]# yum install -y tuned-profiles-cpu-partitioning
```

```
# The cores to be isolated from Linux scheduler and from interrupts: 1-31
```

```
[root]# echo isolated_cores=1\ -31 >> /etc/tuned/cpu-partitioning-variables.conf
```

```
[root]# reboot
```

```
[root]# cat /proc/cmdline =>
```

```
.../... nohz=on nohz_full=1-31 rcu_nocbs=1-31 intel_pstate=disable nosoftlockup
```

```
[root]# cat /proc/irq/default_smp_affinity => 1 # mask 0x1 == CPU0
```

```
[root]# cat /proc/irq/*/smp_affinity => 1
```

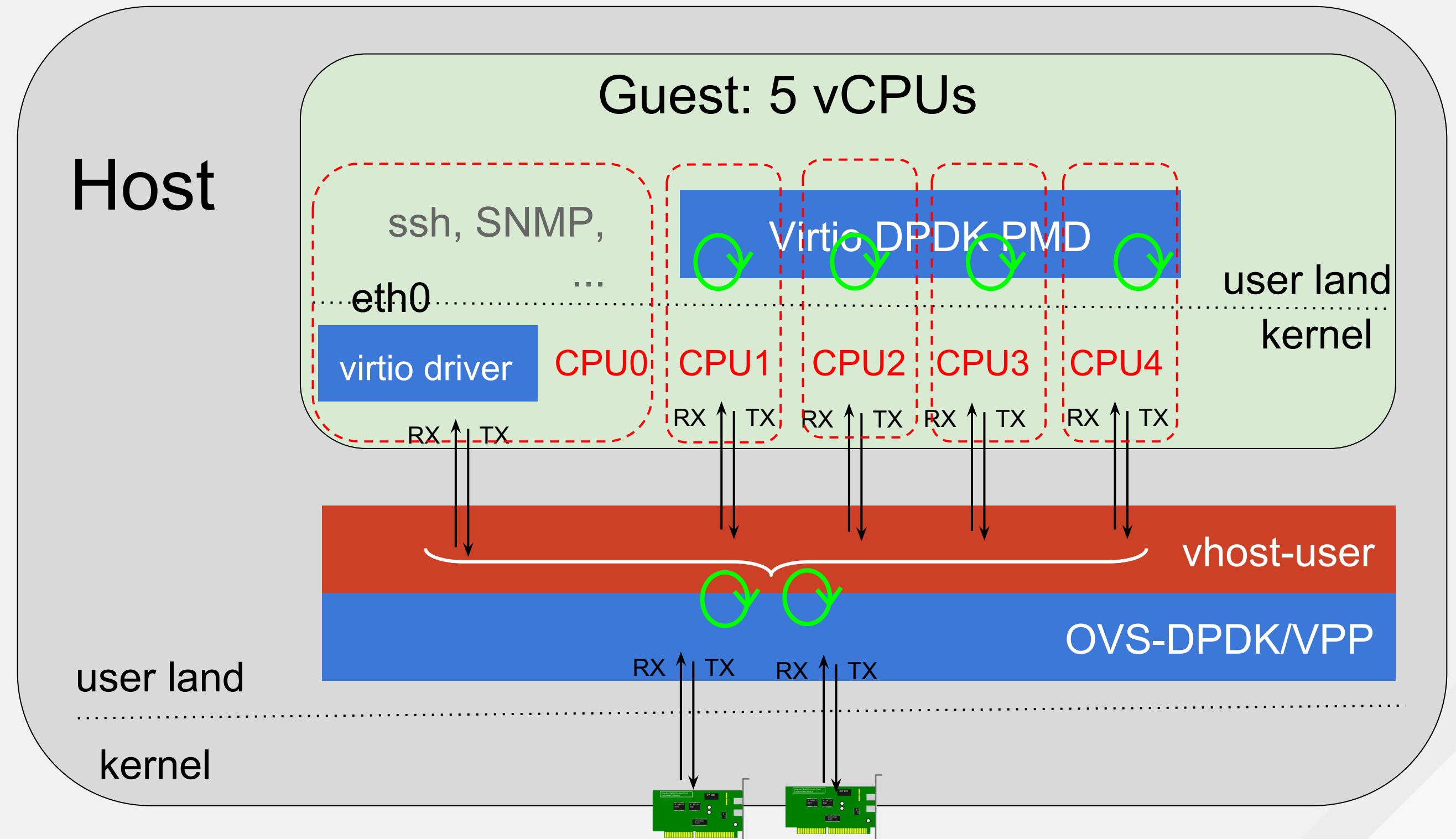
# OVS-DPDK: virtio, vhost-user, virtio PMD



ACTIVE LOOP



```
while (1) {  
    RX-packet()  
    forward-packet()  
}
```



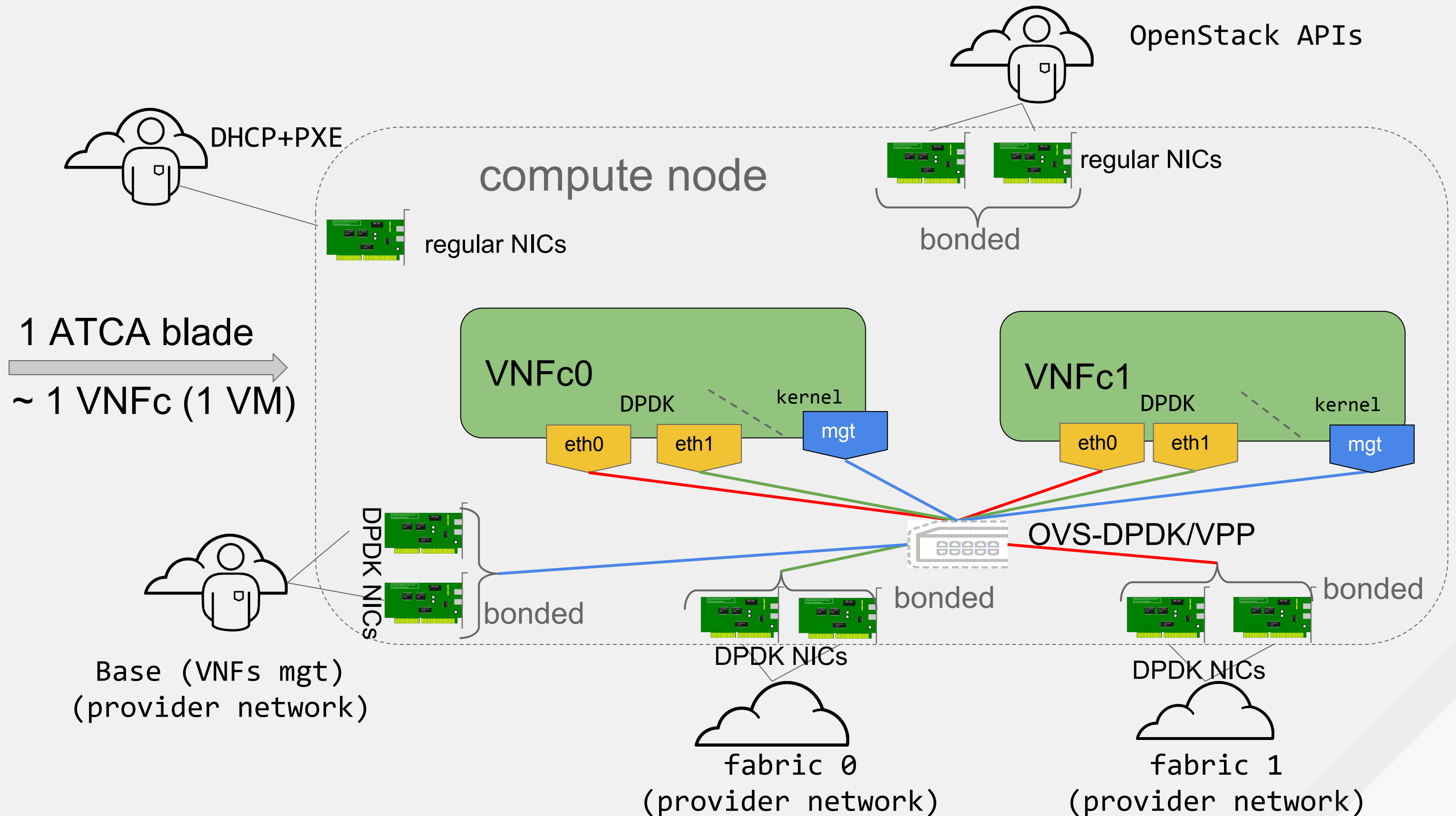


# OVS-DPDK NFV deployment

1 ATCA chassis  
~ 1 VNF



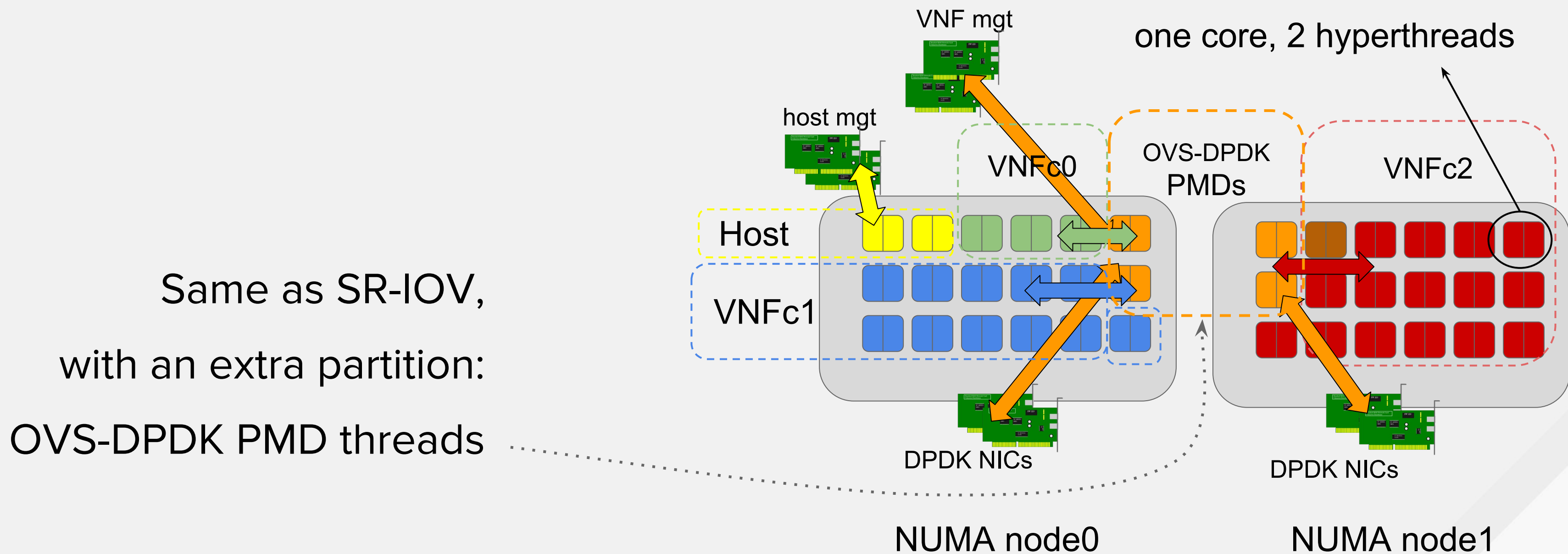
Picture credits: [wikipedia](https://en.wikipedia.org/wiki/ATCA)





# OVS-DPDK Host/VNFs guests resources partitioning

Typical 18 cores per node dual socket compute node ([E5-2599 v3](#))





# Configuring OpenStack for NFV via TripleO



# EPA & partitioning parameters

network-environment.yaml

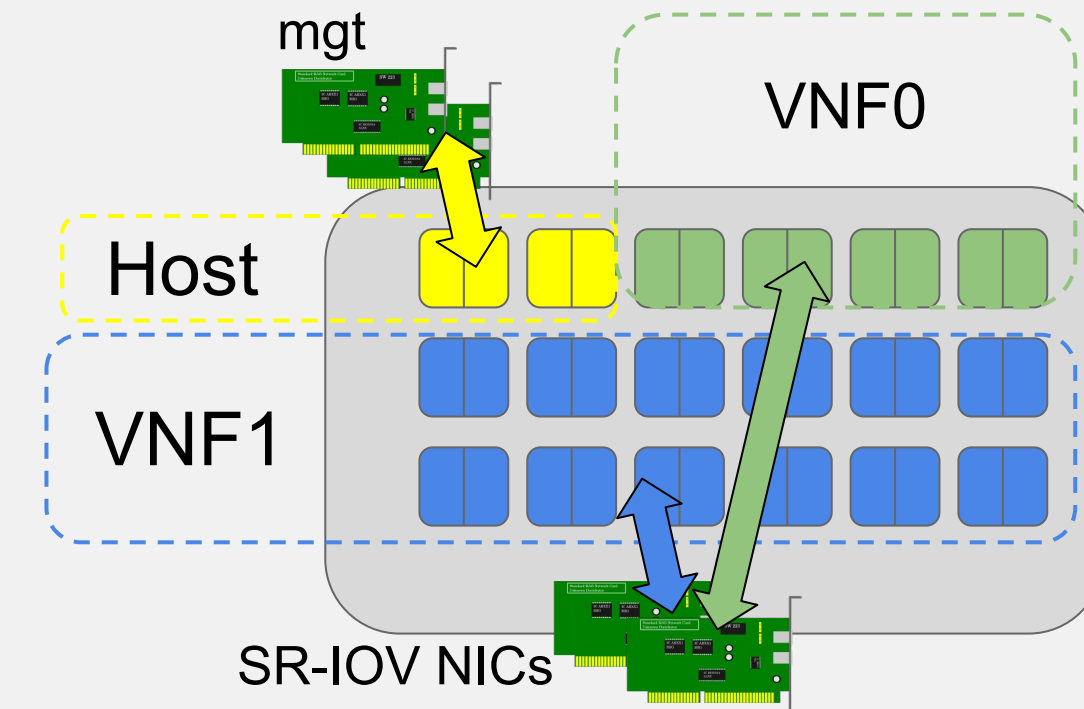
Linux host and Nova required configuration parameters  
exposed at TripleO level"

- **NovaReservedHostMemory**: 4096
- **NovaVcpuPinSet**: "4,20,5,21,6,22,7,23"
- **NovaSchedulerDefaultFilters**: "RamFilter, ComputeFilter, AvailabilityZoneFilter, ComputeCapabilitiesFilter, ImagePropertiesFilter, PciPassthroughFilter, NUMATopologyFilter, AggregateInstanceExtraSpecsFilter"
- **ComputeKernelArgs**: "intel\_iommu=on default\_hugepagesz=1GB hugepagesz=1G hugepages=12"
- **HostCpusList**: "'0-1'"

**nova.conf**

**grub2.cfg**

**tuned.conf**





# Example of Heat environment for SR-IOV

network-environment.yaml

resource\_registry:

OS::TripleO::Services::NeutronSriovAgent: ../puppet/services/neutron-sriov-agent.yaml

parameter\_defaults:

NeutronMechanismDrivers: ['openvswitch','sriovnicswitch']

# Add PciPassthroughFilter to the scheduler default filters *(previous slide)*

# Provide the vendorid:productid of the VFs

NeutronSupportedPCIVendorDevs: ['8086:154c','8086:10ca','8086:1520']

NeutronPhysicalDevMappings: "datacentre:ens20f2"

# Number of VFs that needs to be configured for a physical interface

NeutronSriovNumVFs: "ens20f2:5"

NovaPCIPassthrough:

- devname: "ens20f2"

physical\_network: "datacentre"



# OVS-DPDK TripleO Director (new) parameters

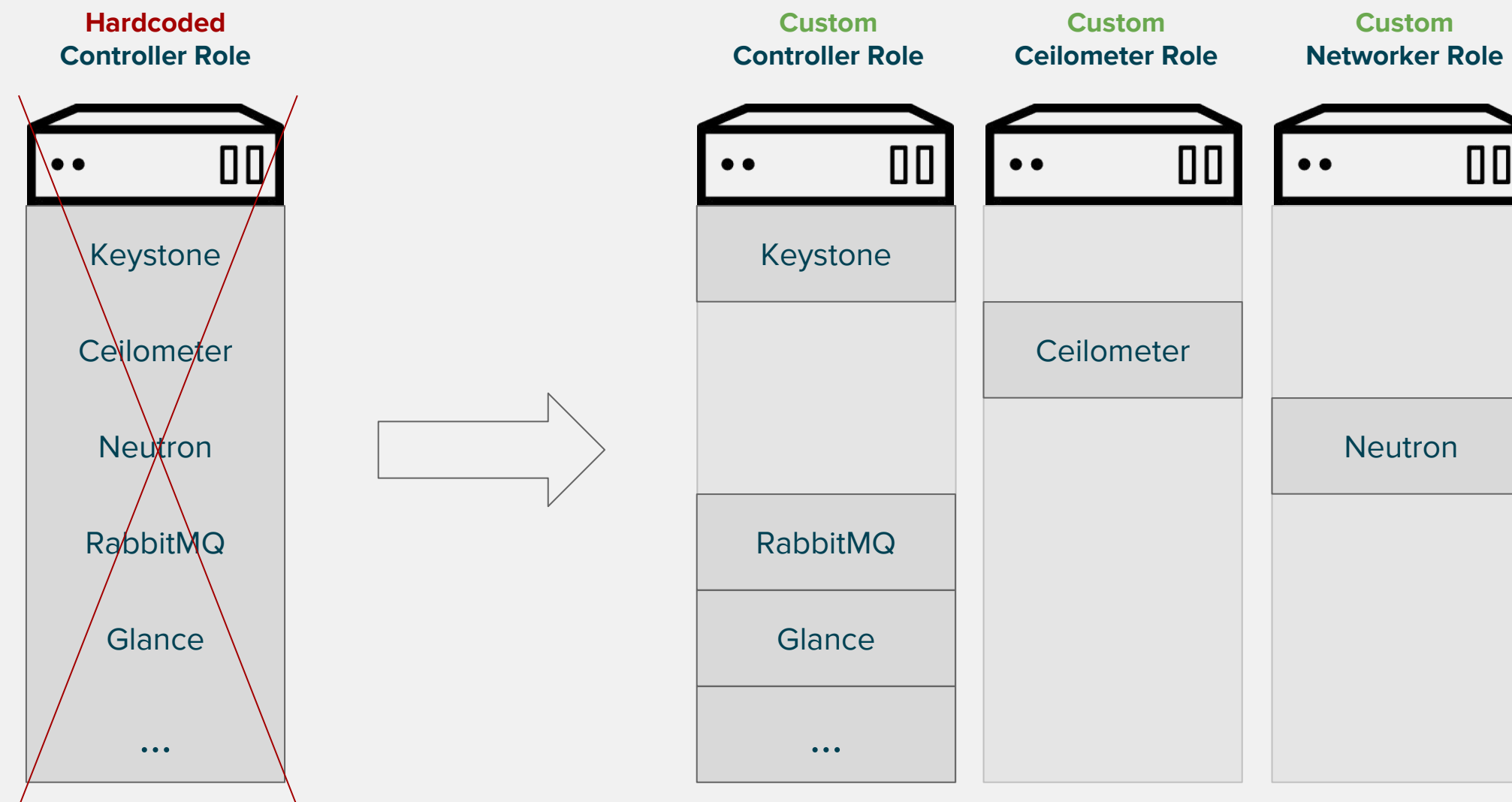
## network-environment.yaml

```
parameter_defaults:
  NeutronDpdkCoreList: "'1,2,18,19'"
  NeutronDpdkMemoryChannels: "4"
  NeutronDatapathType: "netdev"
  NeutronVhostuserSocketDir: "/var/lib/vhost_sockets"
  NeutronDpdkSocketMemory: "2048,2048"
  NeutronDpdkDriverType: "vfio-pci"
```

## compute.yaml

```
Type: ovs_user_bridge
  name br-link
  members:
    - ovs_dpdk_port
      name: dpdk0
      members:
        -
          Type: interface
          name: ens20f3
```

# Composable Services and Custom Roles



- OpenStack Services are now defined via a smaller set of templates
- Services now deployable standalone on separate nodes or combined with other services into Custom Role(s).



# Compute node default role

based on regular OVS and SR-IOV agent

- name: Compute
- CountDefault: 1
- HostnameFormatDefault: '%-compute-%'
- ServicesDefault:
  - OS::TripleO::Services::CACerts
  - OS::TripleO::Services::CephClient
  - OS::TripleO::Services::CephExternal
  - OS::TripleO::Services::Timezone
  - OS::TripleO::Services::Ntp
  - OS::TripleO::Services::Snmp
  - OS::TripleO::Services::NovaCompute
  - OS::TripleO::Services::NovaLibvirt

- OS::TripleO::Services::Kernel
- OS::TripleO::Services::ComputeNeutronCorePlugin
- OS::TripleO::Services::ComputeNeutronOvsAgent
- OS::TripleO::Services::ComputeCeilometerAgent
- OS::TripleO::Services::ComputeNeutronL3Agent
- OS::TripleO::Services::ComputeNeutronMetadataAgent
- OS::TripleO::Services::TripleoPackages
- OS::TripleO::Services::TripleoFirewall
- OS::TripleO::Services::NeutronSriovAgent
- OS::TripleO::Services::OpenDaylightOvs
- OS::TripleO::Services::SensuClient
- OS::TripleO::Services::FluentdClient
- OS::TripleO::Services::VipHosts

# NFV Composable roles

Compose your own NFV compute node role

Based on newly added services in Newton, to be added to the existing/cloned roles

- neutron-sriov-agent, already included in the default role
- neutron-ovs-dpdk-agent

To create an OVS-DPDK role, just duplicate the default compute node role and replace **ComputeNeutronOvsAgent** by **ComputeNeutronOvsDpdkAgent**, defined as follow in your environment.yaml

```
resource_registry:
    OS::TripleO::Services::ComputeNeutronOvsDpdkAgent:
        <path>/openstack-tripleo-heat-templates/puppet/services/neutron-ovs-dpdk-agent.yaml
```



# Open Source Communities



# Queens NFV feature example: OVS offload/vSwitch offload

Goal: replace OVS-DPDK by a SmartNIC

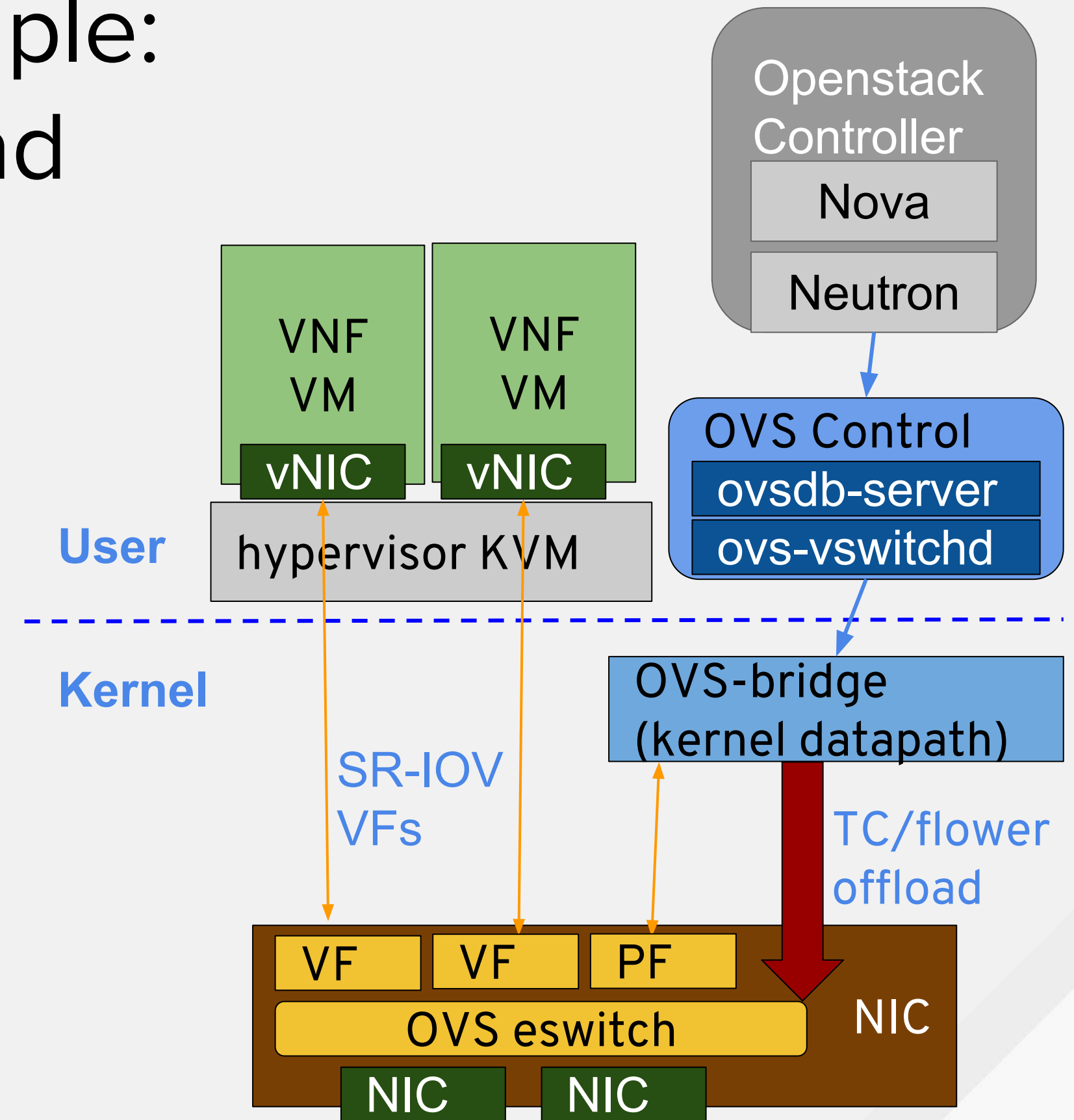
- OVS offload
- OVS-DPDK offload

Same principle apply to any vSwitch offload

- VPP FD.io
- Contrail vRouter

OpenStack impact:

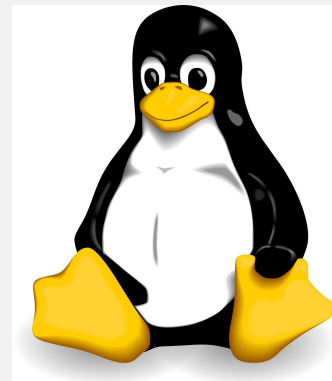
- [Nova](#)
- Neutron
- TripleO





# Communities involved

*simplified view ;-)*



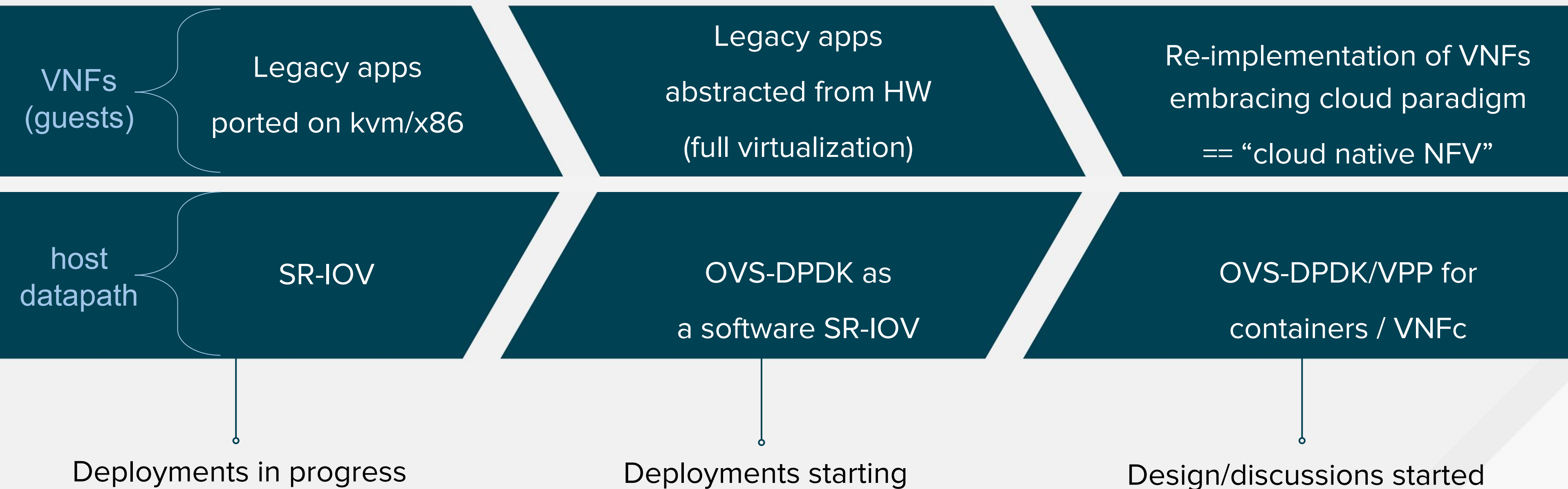


# Key Takeaway



# VNFs evolution driving host datapath evolution

Production vs Innovation





# Thank you!

*[fbaudin@redhat.com](mailto:fbaudin@redhat.com)*



[plus.google.com/+RedHat](https://plus.google.com/+RedHat)



[facebook.com/redhatinc](https://facebook.com/redhatinc)



[linkedin.com/company/red-hat](https://linkedin.com/company/red-hat)



[twitter.com/RedHatNews](https://twitter.com/RedHatNews)



[youtube.com/user/RedHatVideos](https://youtube.com/user/RedHatVideos)



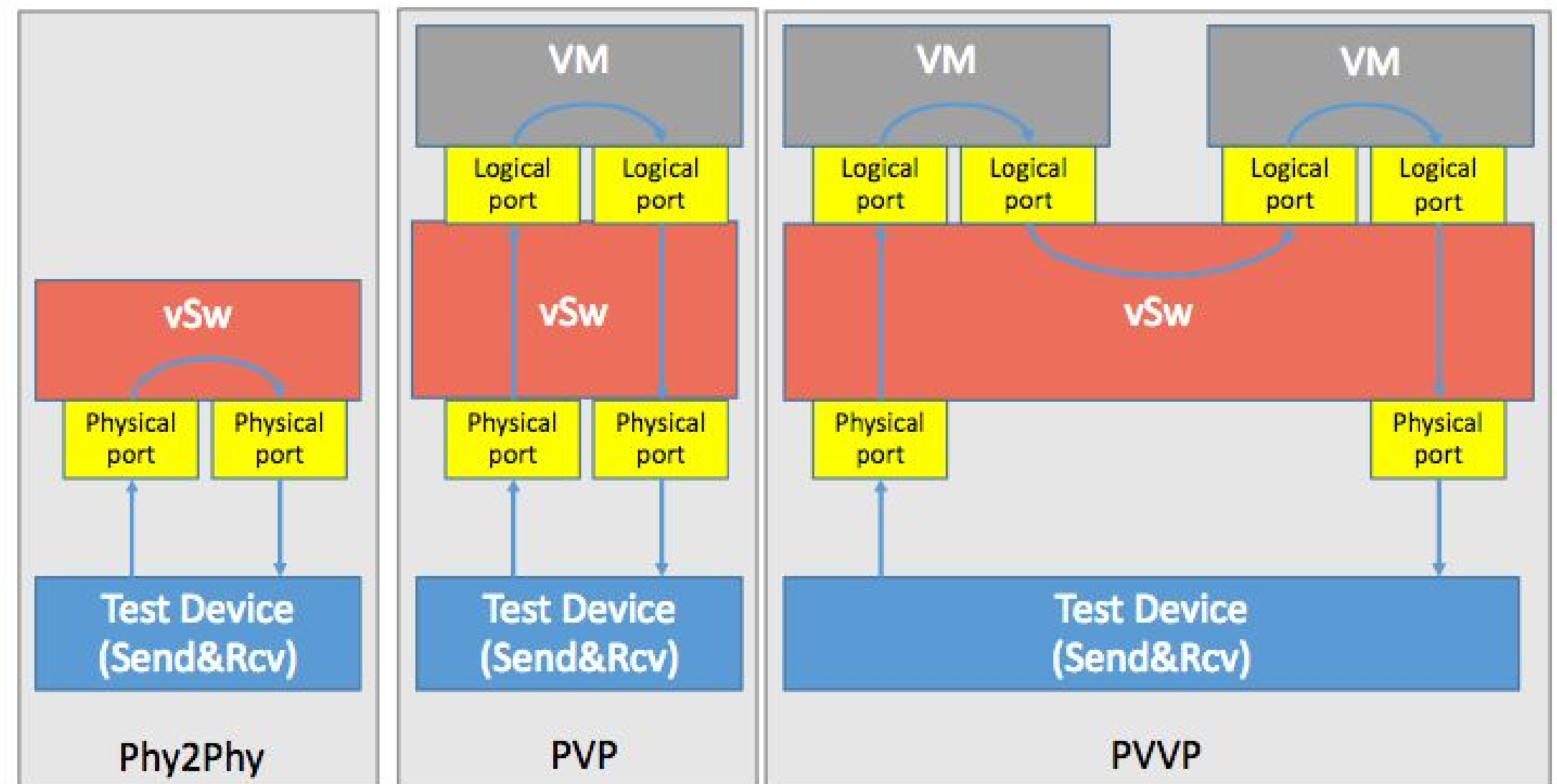
# vSwitches and VNFs end to end performance

PVP and PVVP, with zero frame loss

- 0% loss rates with multiple Queue/PMD configurations
- [VSPerf tests](#) done by RHEL QE and sampled end to end by NFV-QE
- Few established flow-rules, RFC2544
- Testpmd bare-metal and as a vSwitch used to calibrate the platform
- VPP support WIP

## VSPERF LTD Supported Deployment Scenarios

Credit: Maryam Tahhan - Al Morton  
*Benchmarking Virtual Switches in OPNFV  
draft-vsperf-bmwg-vswitch-opnfv-01*



# PVP Testing scenario OVS-DPDK

- DUT Dell R730 18 Core E5-2699 v3 Haswell
- Bi-directional traffic with 1k, 4k and 1 million flows/multi-stream at layer 3
- With and without VLAN tagging

