

Isolating Functions in the System Under Test

Al Morton

Maryam Tahhan

02 July 2015

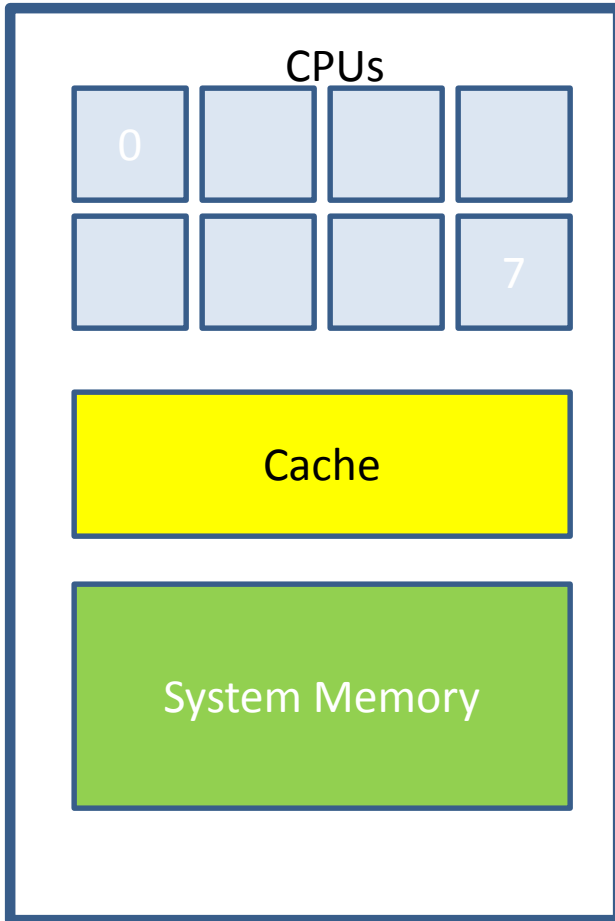
vswitchperf

Proposed Approach

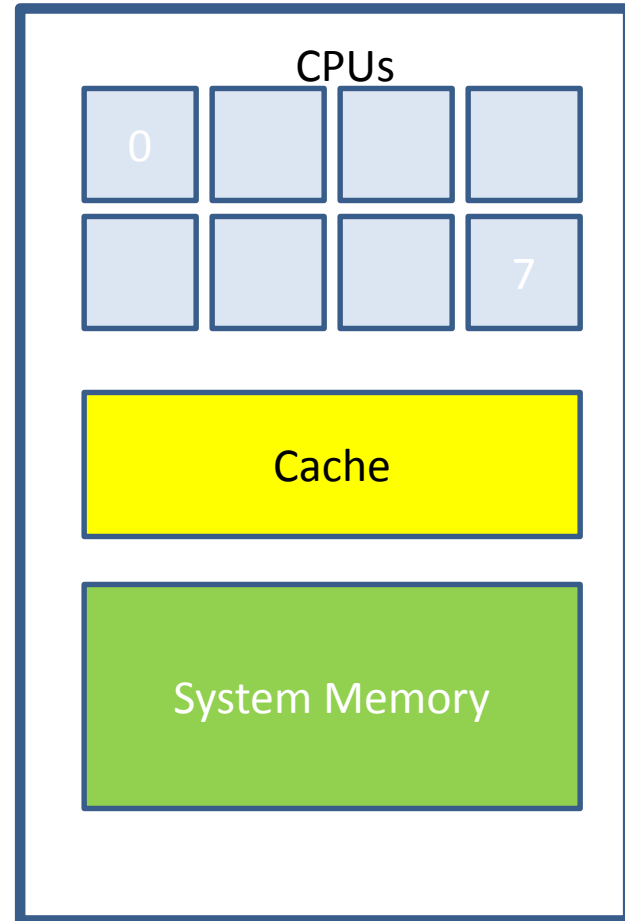
- **Baseline:**
 - A. Optional: Benchmark platform forwarding capability.
 - B. Benchmark VNF forwarding capability.
 - C. benchmarking with isolated resources alone, with other resources (both HW&SW) disabled
 - A. Example, vSw and VM are SUT
 - D. benchmarking with isolated resources alone, leaving some resources unused
 - E. Isolated resources and all resources occupied
- **Next Steps:**
 - Limited sharing
 - Production scenarios
 - Stressful scenarios

NUMA Nodes

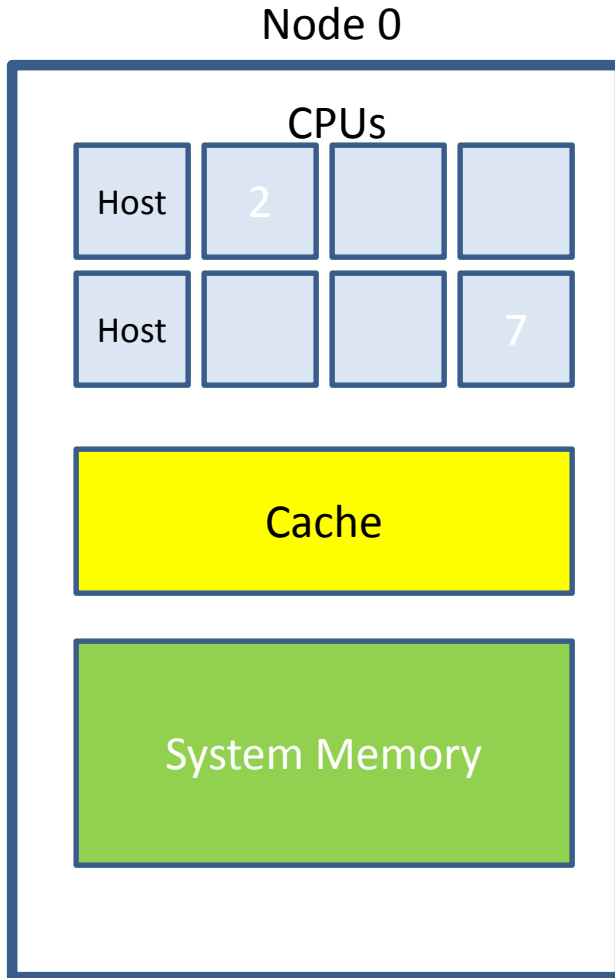
Node 0



Node 1

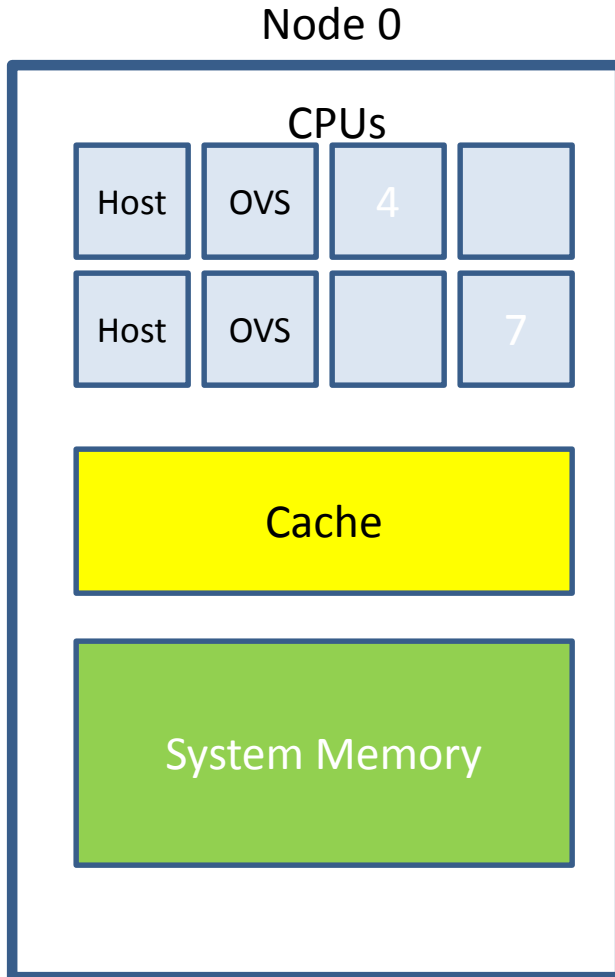


CPU Pinning



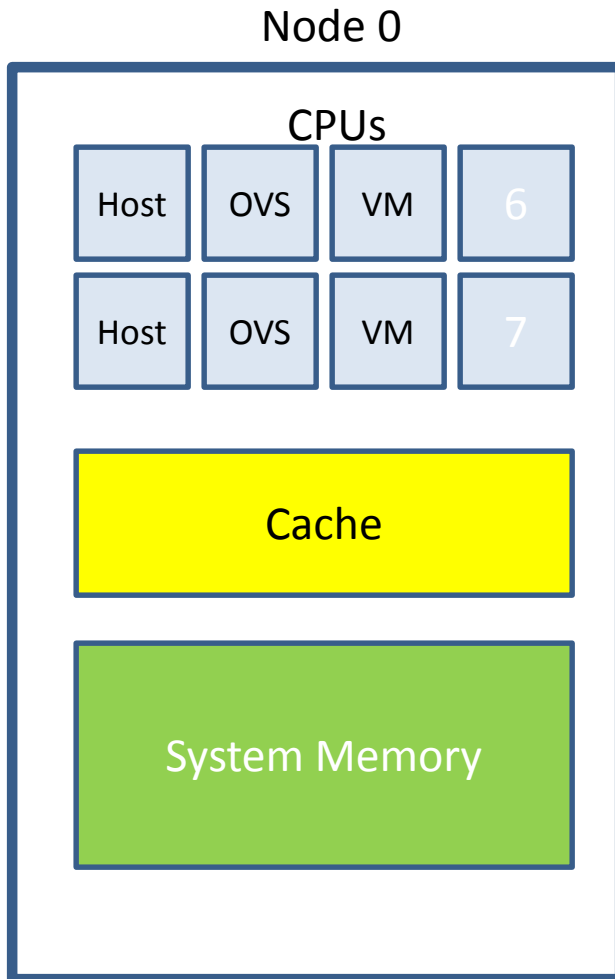
- Isolcpus = 2-7
- OpenStack Compute Scheduler

CPU Pinning



- Isolcpus = 2-7
- OpenStack Compute Scheduler
- `taskset <COREMASK> <EXECUTABLE>`
- `taskset 0x0C ovs-vswitchd`
- (limit NIC interrupt to cores)

CPU Pinning



- `Isolcpus = 2-7`
- OpenStack Compute Scheduler
- `taskset <COREMASK> <EXECUTABLE>`
- `taskset 0x0C ovs-vsitchd`
- (limit NIC interrupt to cores)
- `nova aggregate create/set`
- `nova flavor-create <sizes>`
- `nova aggregate-add-host`

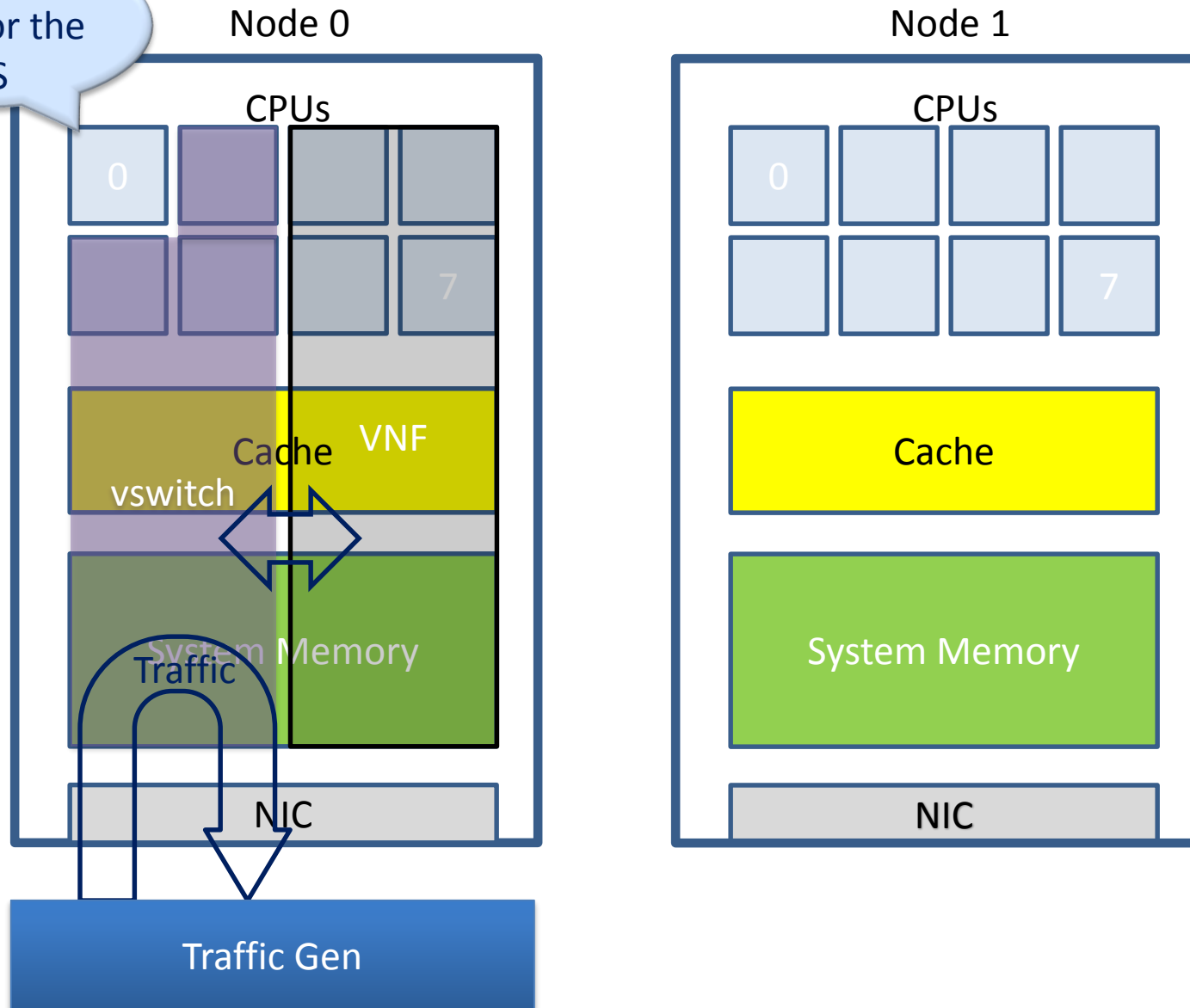
What Else can we Isolate?

- PMD (ovs config)
- Threads
- PCIe Lanes
- Verification methods (show that isolation configs worked)

BASELINE BENCHMARKING APPROACH

SUT (vSwitch + VNF)

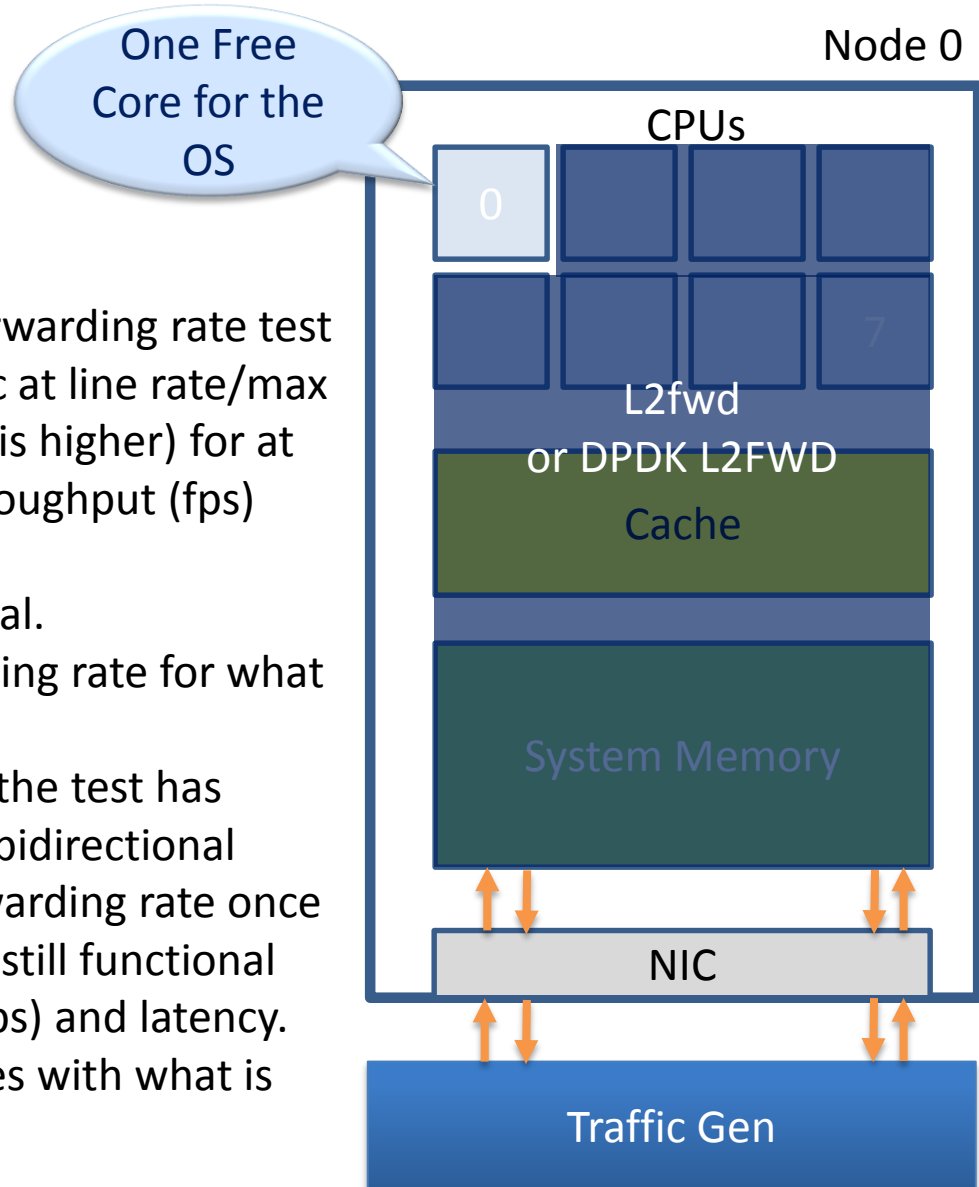
One Free Core for the OS



Benchmark platform forwarding capability

OPTIONAL

- Run RFC 2889 Maximum forwarding rate test
- Transmit bidirectional traffic at line rate/max forwarding rate (whichever is higher) for at least 72 hours, measure throughput (fps) and latency.
- Traffic should be bidirectional.
- Establish a baseline forwarding rate for what the platform can achieve.
- Additional validation: After the test has completed for 72hours run bidirectional traffic at the maximum forwarding rate once more to see if the system is still functional and measure throughput (fps) and latency. Compare the measure values with what is expected.



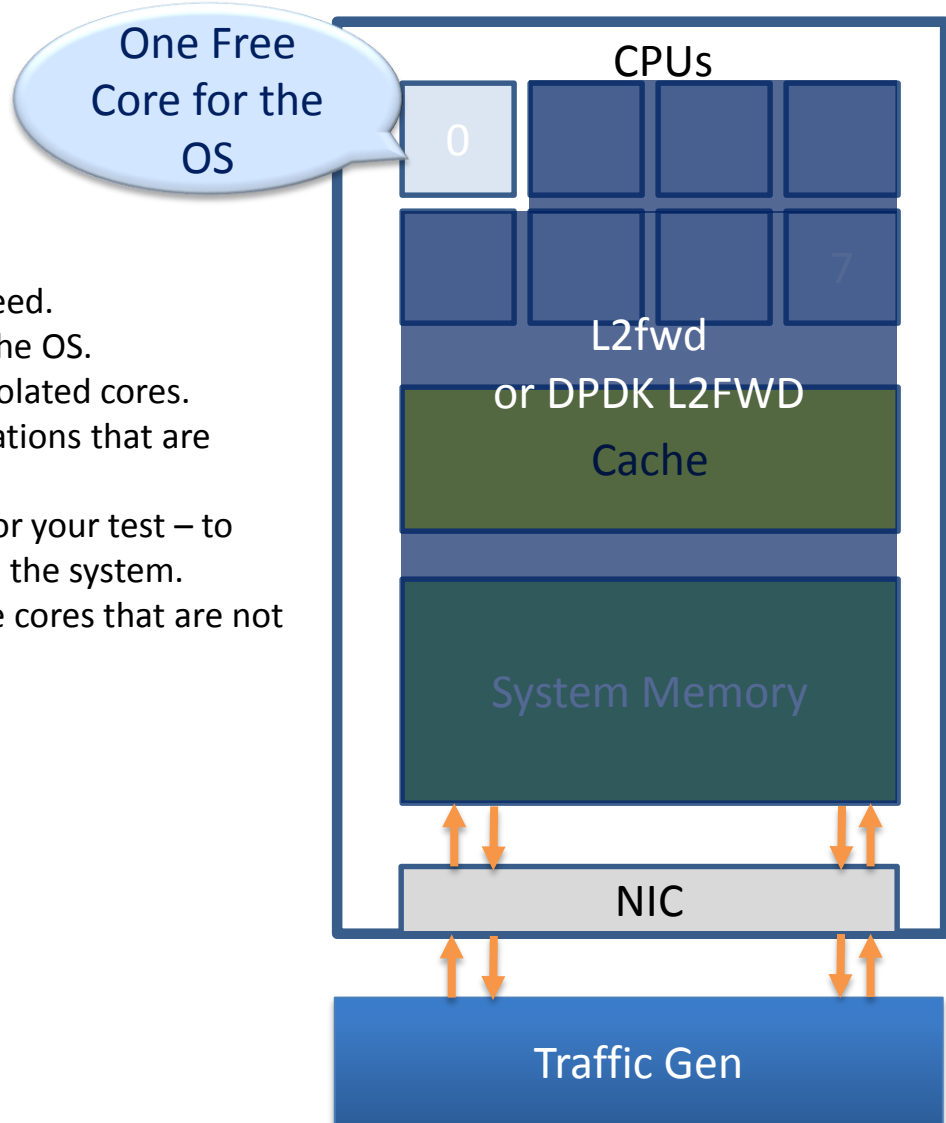
Benchmark platform forwarding capability

Settings

OPTIONAL

Settings:

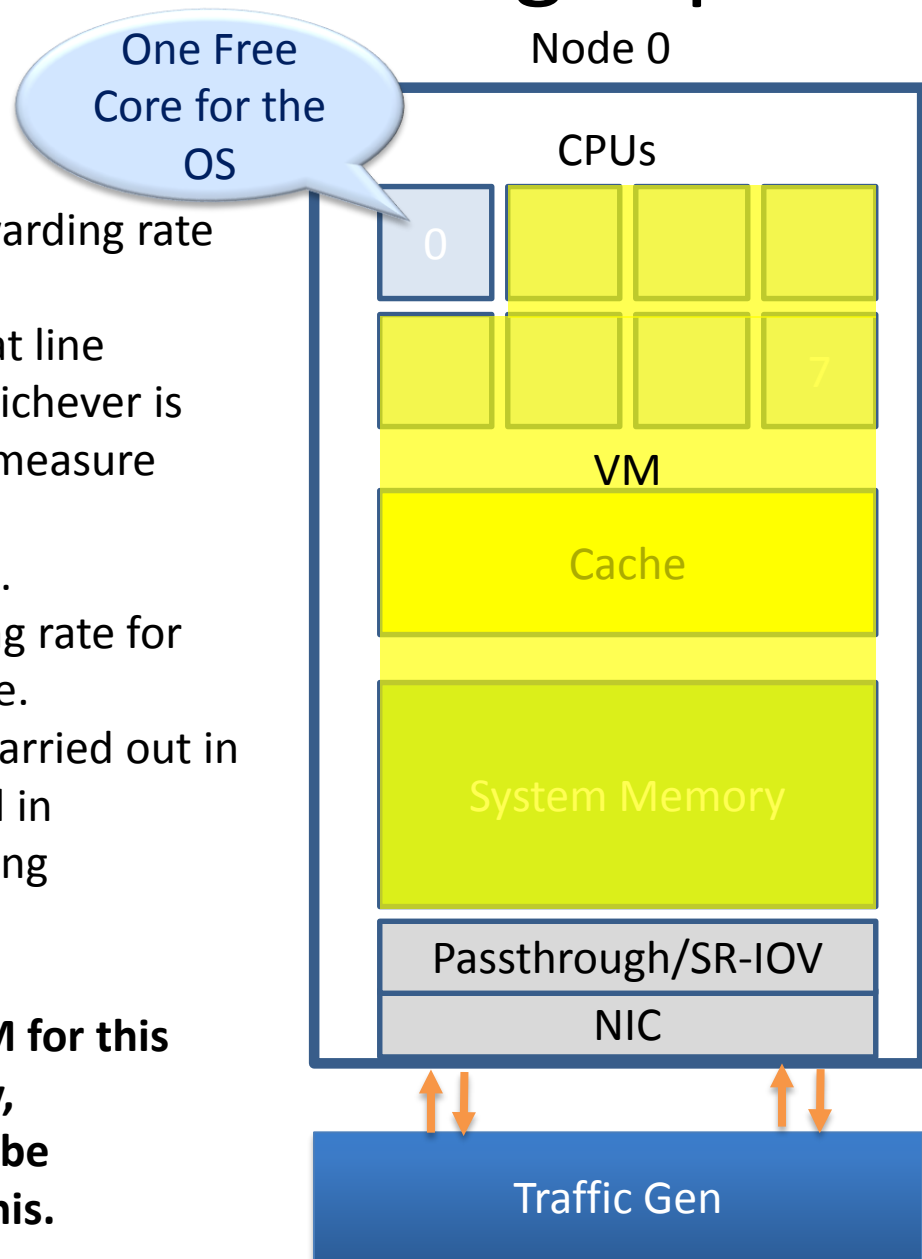
- maxcpus = only use exactly what you need.
- Isolcpus = everything bar one core for the OS.
- Taskset the forwarding application to isolated cores.
- Limit the amount of background applications that are running.
- Only enable HW that you need to use for your test – to ensure there are no other interrupts on the system.
- Configure NIC interrupts to only use the cores that are not allocated to any other process.
- Set OS to boot to runlevel 3
- **Disable other sockets in BIOS**



Benchmark VNF forwarding capability

- Run RFC 2889 Maximum forwarding rate test
- Transmit bidirectional traffic at line rate/max forwarding rate (whichever is higher) for at least 72 hours, measure throughput (fps) and latency.
- Traffic should be bidirectional.
- Establish a baseline forwarding rate for what the platform can achieve.
- Additional validation can be carried out in the same manner as specified in Benchmark platform forwarding capability.

NOTE: How we configure the VM for this test (# of vCPUs, vNICs, Memory, affinity...) is how it should be configured for every test after this.



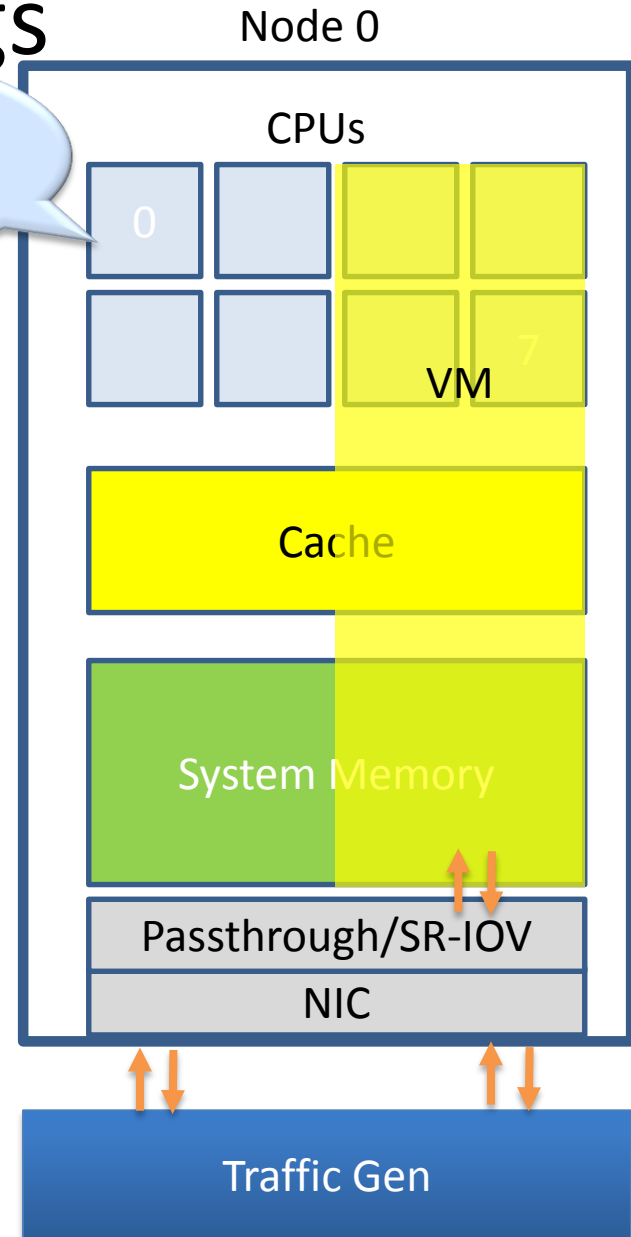
Benchmark VNF forwarding capability

Settings

Settings:

- maxcpus = only use exactly what you need.
- Isolcpus = Leave one core for the OS.
- Taskset the VM to isolated cores.
- Limit the amount of background applications that are running.
- Only enable HW that you need to use for your test – to ensure there are no other interrupts on the system.
- Configure NIC interrupts to only use the cores that are not allocated to any other process.
- Affinitize vCPUs to separate cores.
- Set OS to boot to runlevel 3
- **Disable other sockets in BIOS**

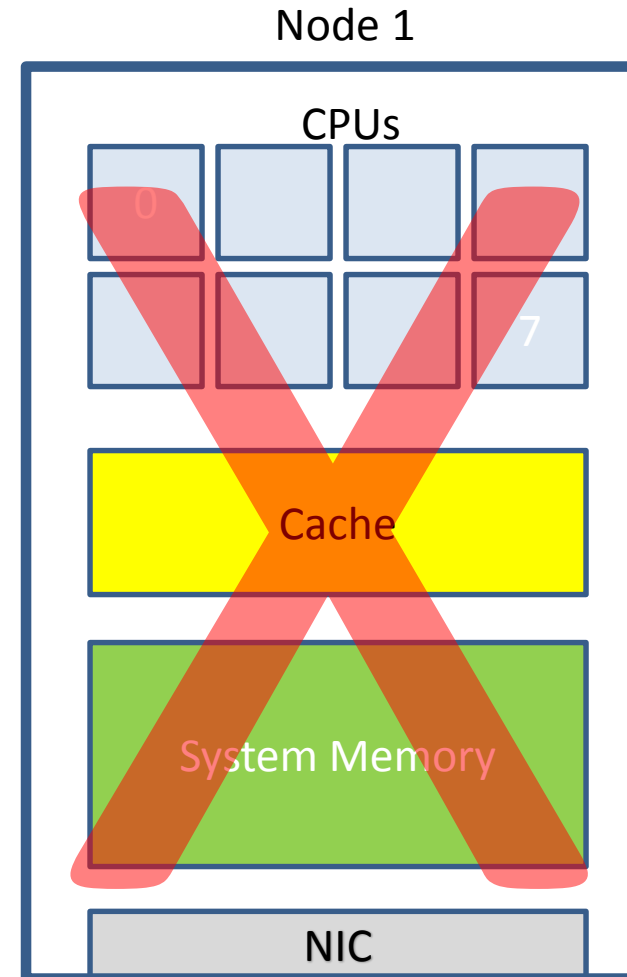
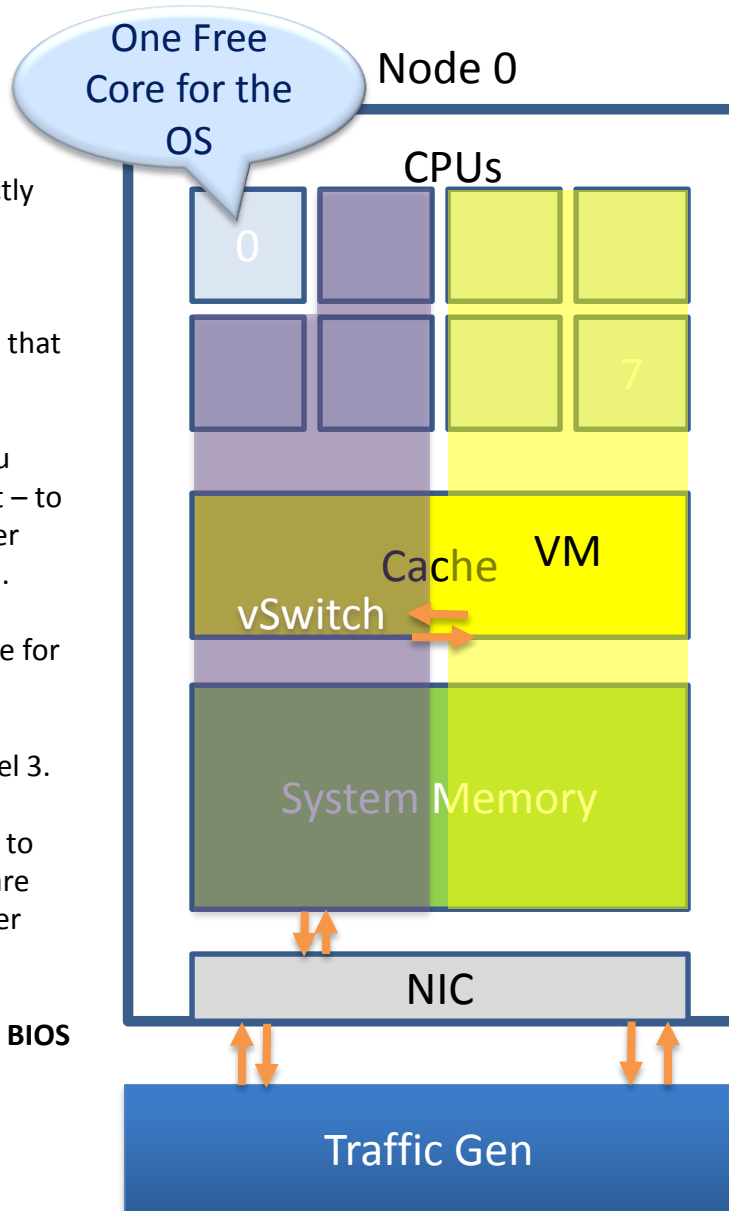
One Free Core for the OS



Benchmark with isolated resources **alone**

Settings:

- maxcpus = only use exactly what you need.
- Limit the amount of background applications that are running
- Only enable HW that you need to use for your test – to ensure there are no other interrupts on the system.
- Isolcpus = Leave one core for the OS.
- Set OS to boot to runlevel 3.
- Configure NIC interrupts to only use the cores that are not allocated to any other process (vSwitch/VNF).
- **Disable other sockets in BIOS**

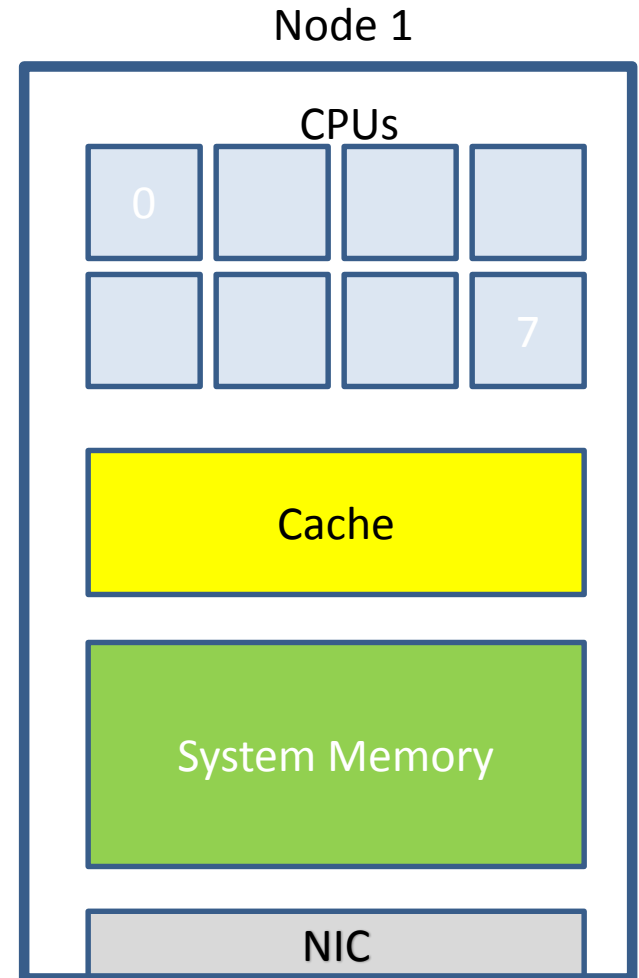
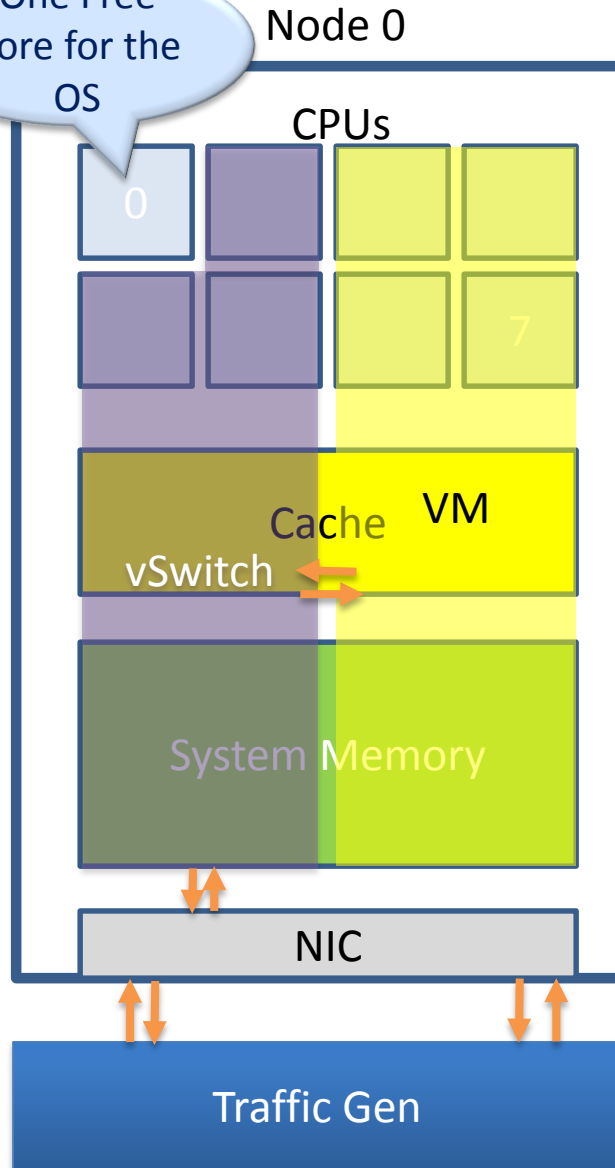


Benchmark with isolated resources

One Free Core for the OS

Settings:

- maxcpus = only use exactly what you need.
- Limit the amount of background applications that are running.
- Only enable HW that you need to use for your test – to ensure there are no other interrupts on the system.
- Isolcpus = Leave one core for the OS.
- Set OS to boot to runlevel 3.
- Configure NIC interrupts to only use the cores that are not allocated to any other process (vSwitch/VNF).
- **Enable the 2nd socket in BIOS.**



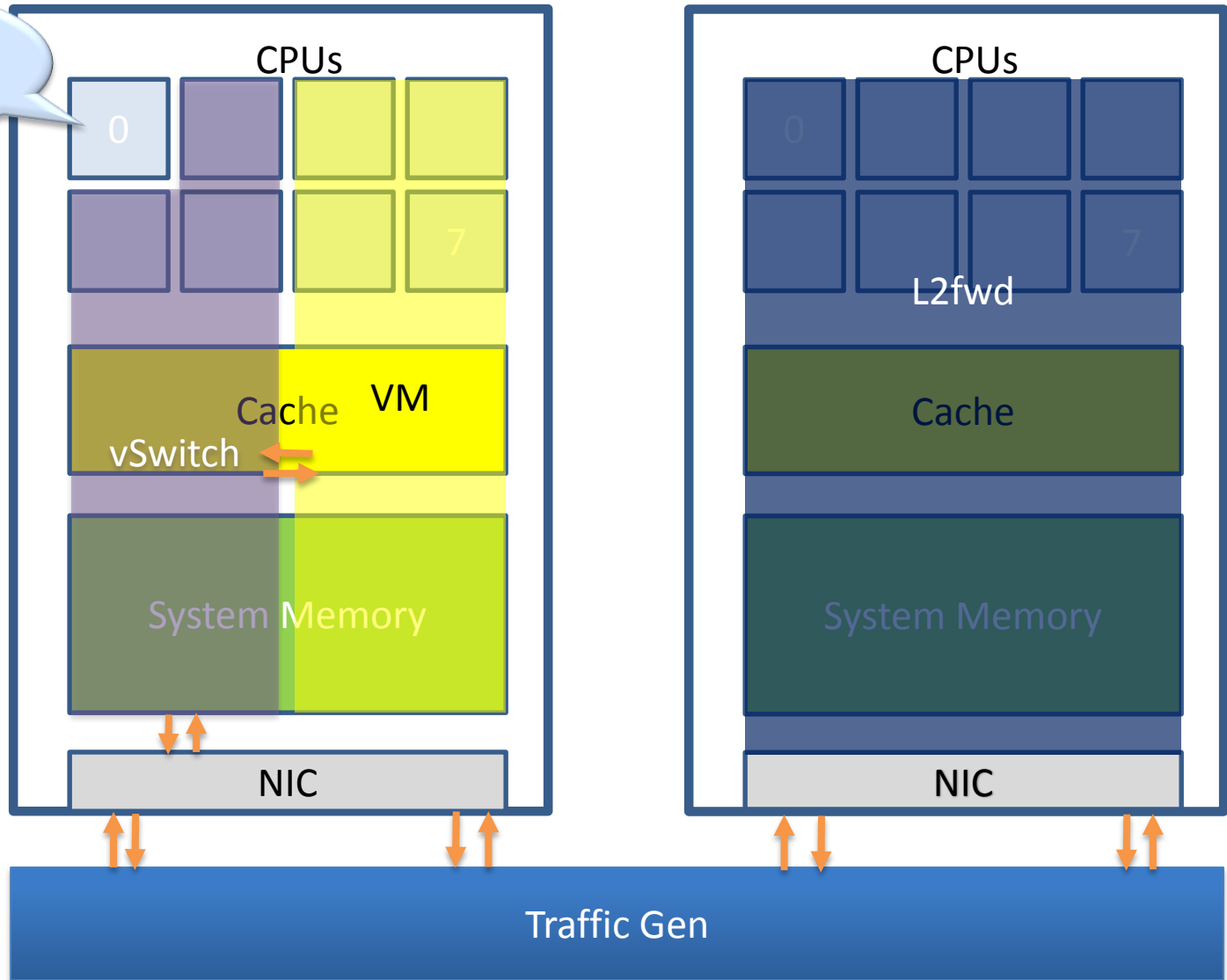
Benchmark with isolated resources and all resources occupied

Node 0

Node 1

One Free Core for the OS

- Enable and use all CPUs.



Backup

Parameters that can affect the repeatability of a test

There is a bunch of information related to the system that can affect the repeatability of a test and should be recorded:

- Platform details: CPU, memory, # of enabled cores and # of cores used for the test.
- OS version (for host and VNF), kernel version (for host and VNF), GRUB boot parameters (for host and VNF) and platform BIOS settings.
- Selected vSwitch, version # or commit id used.
- vSwitch launch command line if it's been parameterised.
 - Memory allocation to the vSwitch – which NUMA node it's using, how many memory channels...
- DPDK or any other SW dependency version # or commit id used.
- Memory allocation to a VM - if it's from Hupages/elsewhere.
- VM storage type: snapshot/independent persistent/ independent non-persistent.
- # of phy NICs, versions, type and the PCI slot they are plugged into.
- NIC interrupt configuration.
- # of VMs, hypervisor, and version of the hypervisor.
- # of Virtual NICs, versions, type and driver.
- # of virtual CPUs and their core affinity on the host.
- # vNIC interrupt configuration.
- Test duration.
- # of flows.
- Thread affinitization for the applications on the host.

Please note: Virtual memory acceleration technologies + hyper threading should be enabled by default on the SUT.