



Automatic CPU and NUMA pinning

Liran Rotenberg
Software Engineer

02/2022



High Performance VM

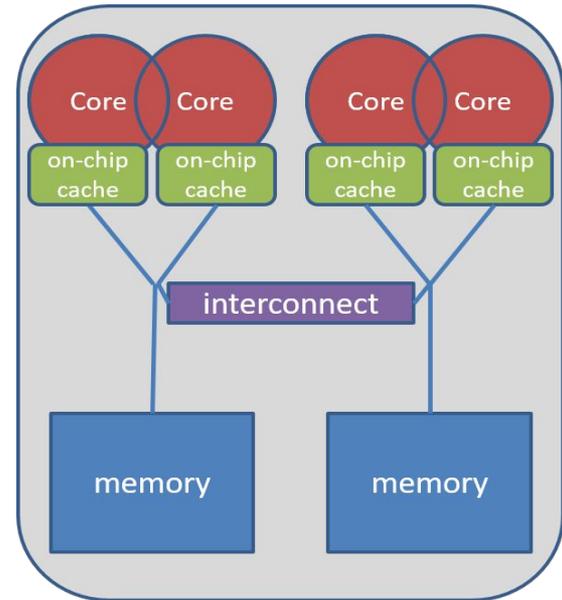
- FOSDEM '19: New VM type: High Performance
- Useful for CPU-intensive workloads, like SAP HANA
- Automatically configured VM properties, that might not be straightforward to the user.
 - Making it headless, without USB controller, etc
- Not everything is automated, manual modifications are needed

CPU and Topology

- Socket – physical connector on motherboard for CPU package
- Die – piece of semiconducting material on which cores are fabricated
(not configurable in oVirt)
- Core – a processor
- Thread – logical unit sharing resources with other threads on core

NUMA – Non-Uniform Memory Access

- Each NUMA node has separate:
 - CPUs
 - memory controller and memory
 - IO controllers and devices
- Locality matters
- Typically NUMA node = Socket, but this is not a rule



Source: HPC Wiki (CC BY-SA)

CPU Assignment in oVirt

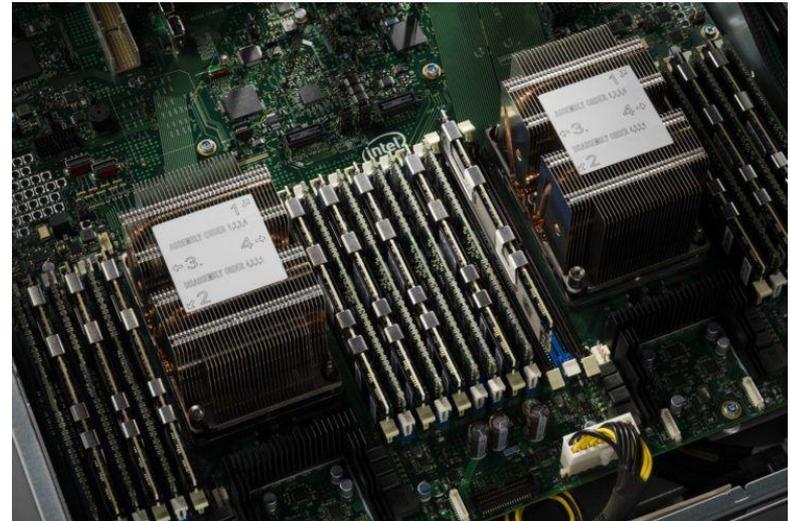
CPU Pinning

- Specified by pinning string
- Difficult to understand
- Difficult to write
- Can limit a vCPU to one or more pCPU
 - Reduces the movement of processes on the host

0#3_2#1-2,12_5#3,4,10,^10_6#6-9,^8_9#13-15

Limitations

- Static or evaluated on VM edit
- Requires host pinning
- CPUs are shared (!)
 - Other VMs and processes can run on the same CPUs
- Configuring meaningful pinning for high number of VMs that run on a host is a tedious task



Source: Optocrypto (CC BY-NC-SA)

Manual procedure defining pinning

Manual pinning (guidelines for SAP HANA)

- After selecting a host, get its topology:
 - CPU topology
 - NUMA topology
- Change the VM CPU topology to fit the host CPU topology and leave space for host processes (Resize)
- Change the vNUMA to fit the host pNUMA
- Run a script on the target host that generates the CPU pinning string based on the host topology
 - Pins according to the sockets and cores
 - Only some topologies supported
- Copy that CPU pinning string to the VM configuration
- Pin vNUMA to pNUMA accordingly

CPU and NUMA Auto Pinning (oVirt 4.4)

- Assigns CPUs based on host topology
- Only one policy “Resize and Pin” that resizes the CPU topology of VM based on our advice for SAP HANA users
- Effective on VM edit
- Part of the VM static configuration
- Does not change on VM start

CPU Pinning Policy

CPU Pinning Policy

- A new configuration added to the VM - CPU Pinning Policy
- The “Resize and Pin” option does the manual procedure automatically (oVirt 4.4)
- It has enhanced support (1 thread topology)
- Limitation:
 - You need to pin the VM to one host or more

CPU Pinning Policy - Pin

- We added an option to use Pin policy, that would not change the VM CPU topology, and try to create a pinning based on that
 - A major flaw was that we used the same pCPUs for multiple VMs
 - Alternative solution - TBD

CPU Pinning Policy - UI

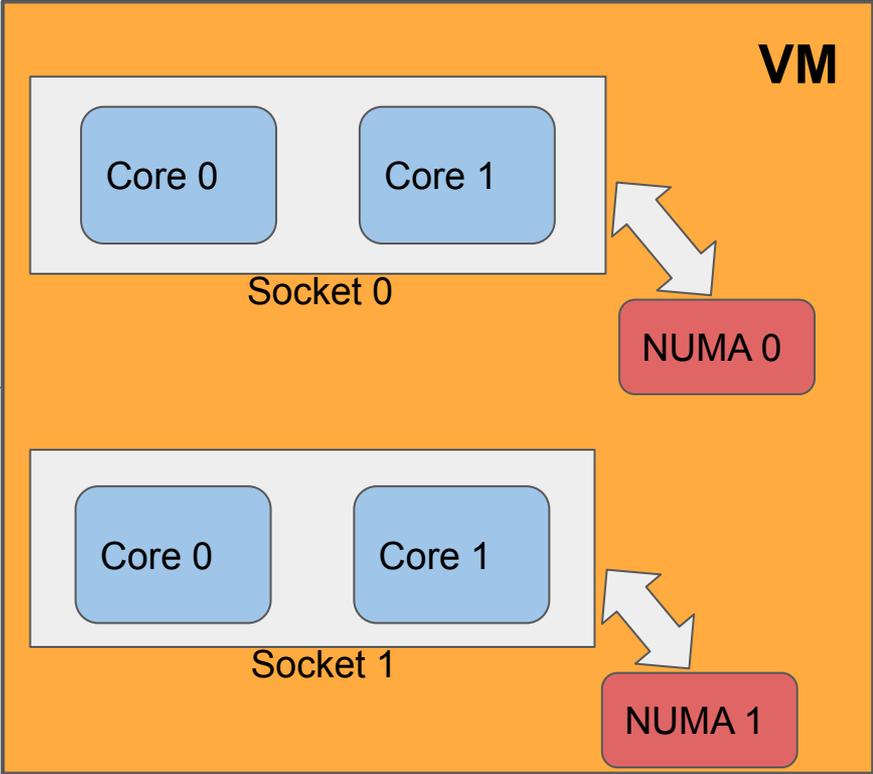
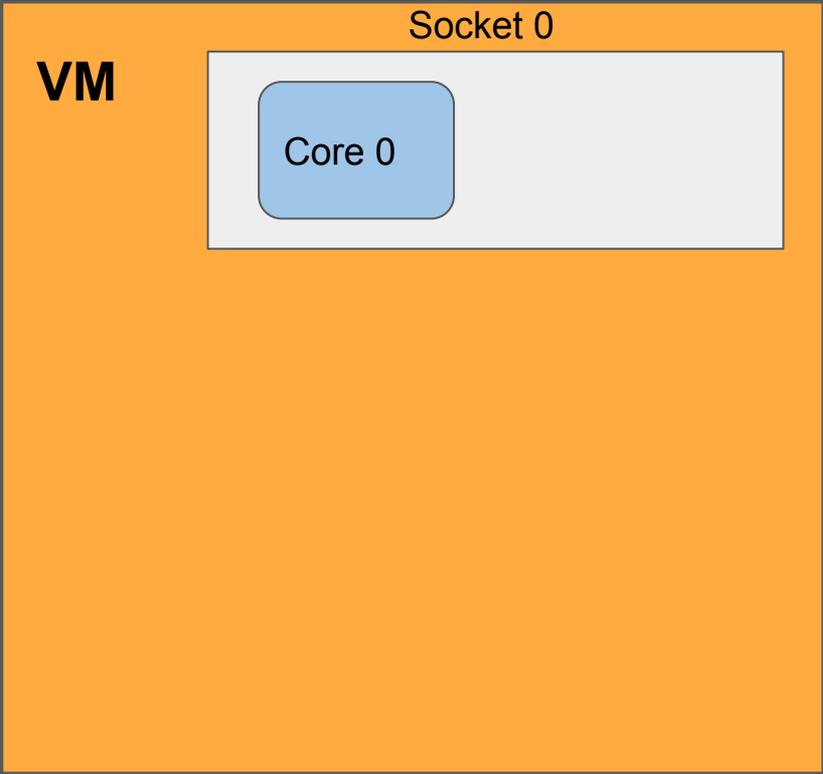
The screenshot shows the 'Edit Virtual Machine' dialog box with the 'Resource Allocation' tab selected. The 'CPU Pinning Policy' dropdown menu is highlighted with a red box and set to 'Resize and Pin NUMA'. Other settings include 'Cluster' (Default), 'Template' (Blank | (0)), 'Operating System' (Other OS), 'Chipset/Firmware Type' (Q35 Chipset with BIOS), 'Optimized for' (Server), 'CPU Profile' (Default), 'CPU Shares' (Disabled, 0), 'CPU Pinning topology' (empty), 'Memory Allocation' (Memory Ballooning Enabled), 'Trusted Platform Module' (TPM Device Enabled), 'I/O Threads' (I/O Threads Enabled, 1), and 'Queues' (Multi Queues Enabled, VirtIO-SCSI Enabled). The 'Hide Advanced Options' button is visible at the bottom left, and 'OK' and 'Cancel' buttons are at the bottom right.

Category	Setting	Value
General	Cluster	Default
	Data Center	Default
System	Template	Blank (0)
Initial Run	Operating System	Other OS
Console	Chipset/Firmware Type	Q35 Chipset with BIOS
Host	Optimized for	Server
High Availability		
Resource Allocation	CPU Allocation:	
	CPU Profile	Default
	CPU Shares	Disabled 0
	CPU Pinning Policy	Resize and Pin NUMA
	CPU Pinning topology	
	Memory Allocation:	
	Memory Ballooning Enabled	<input checked="" type="checkbox"/>
	Trusted Platform Module:	
	TPM Device Enabled	<input type="checkbox"/>
	I/O Threads:	
	I/O Threads Enabled	<input checked="" type="checkbox"/> 1
	Queues:	
	Multi Queues Enabled	<input checked="" type="checkbox"/>
	VirtIO-SCSI Enabled	<input checked="" type="checkbox"/>

An example of resizing

1 socket, 1 core, 1 thread

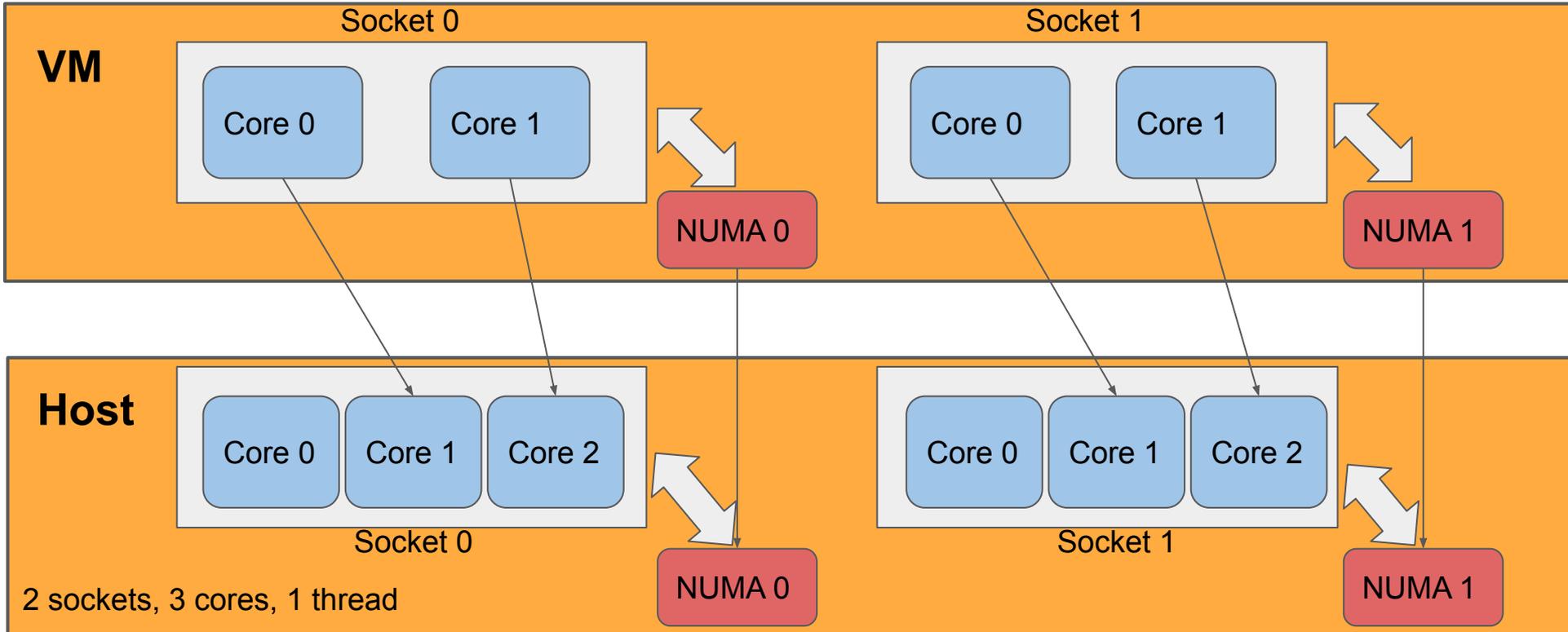
2 sockets, 2 cores, 1 thread



An example of auto pinning

2 sockets, 2 cores, 1 thread

CPU Pinning string: 0#1_1#2_2#4_3#5

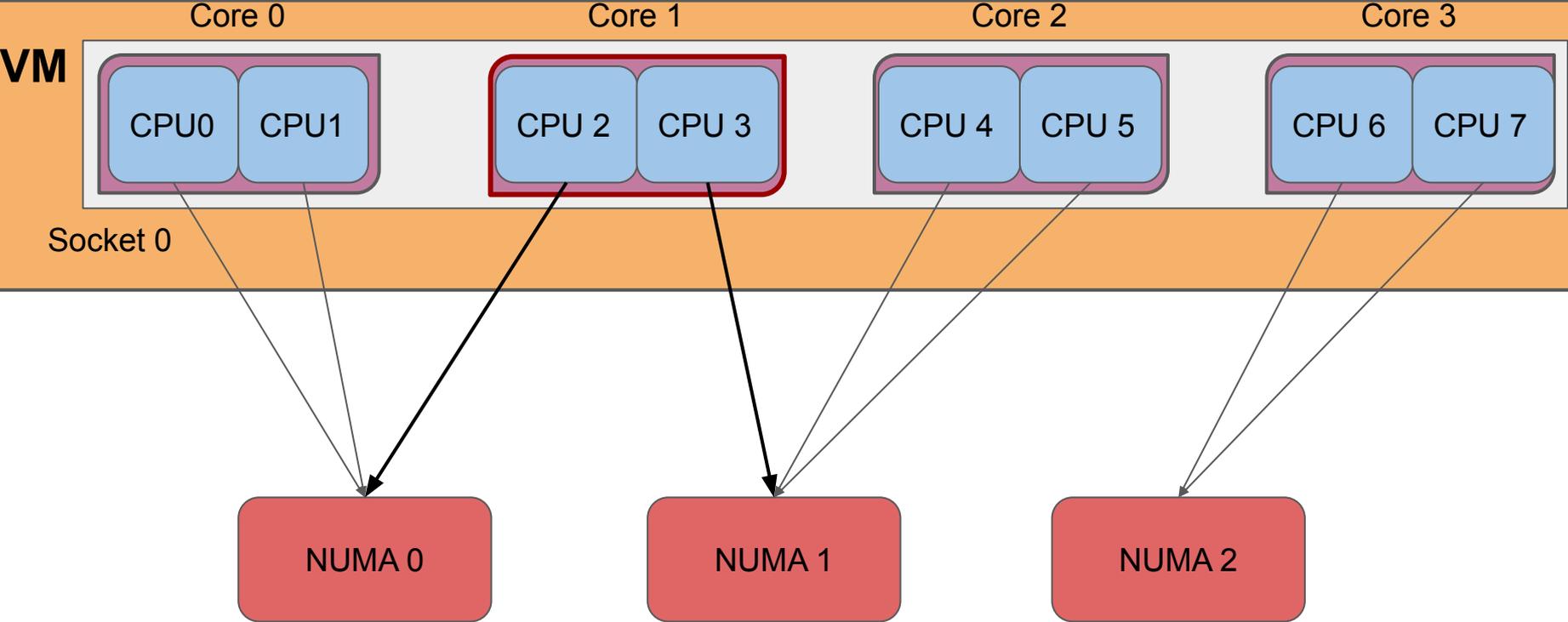


CPU Pinning Policy

- Fixes incorrect splitting of vCPUs to vNUMA nodes
 - oVirt generates the CPU set to the NUMA
 - A core can be divided into two different NUMA nodes
 - Within the guest the CPU topology won't match the VM configuration

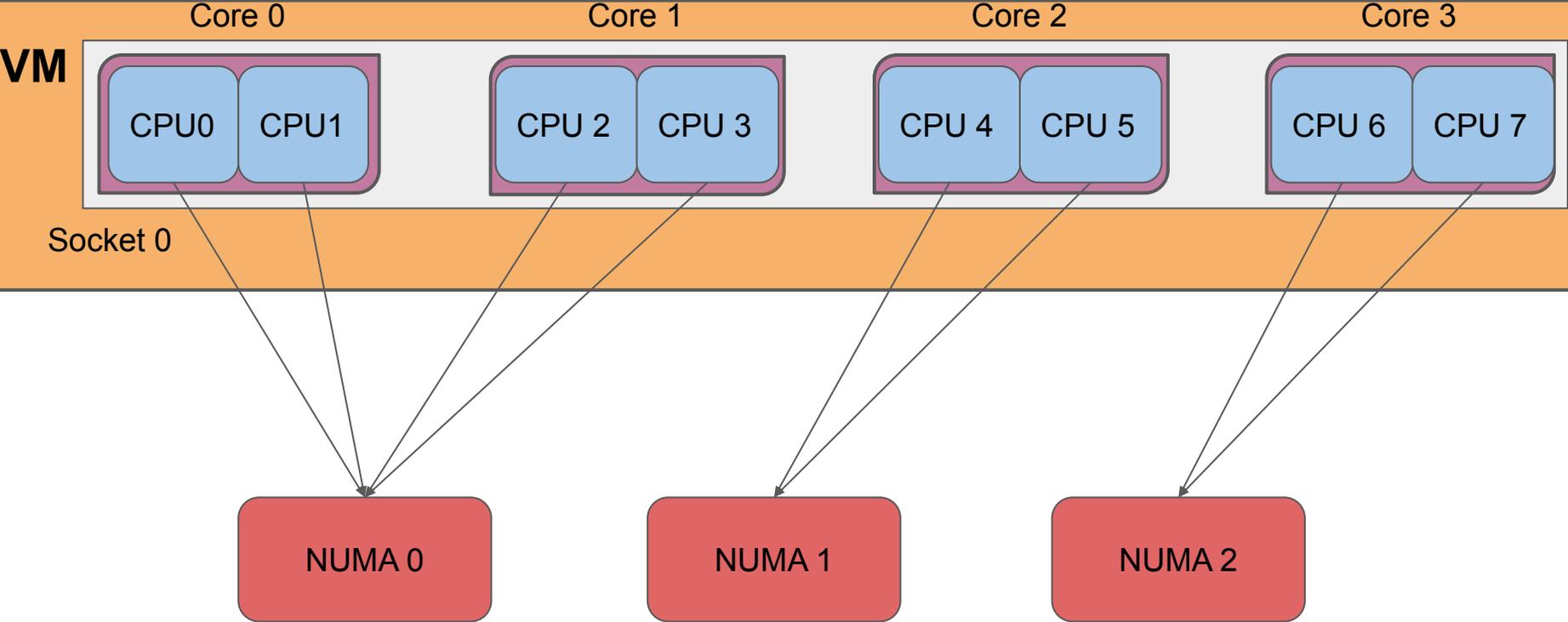
An example of wrong NUMA pinning

1 sockets, 4 cores, 2 thread



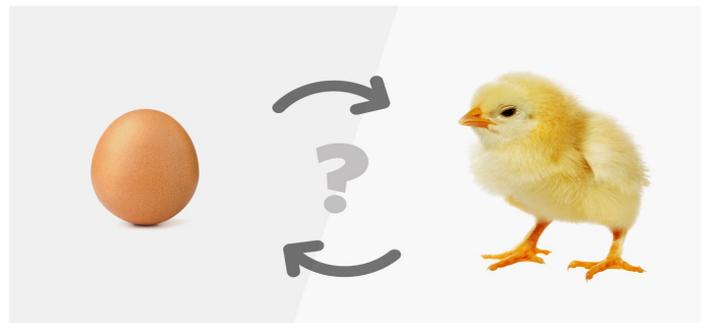
An example of a proper NUMA pinning

1 sockets, 4 cores, 2 thread



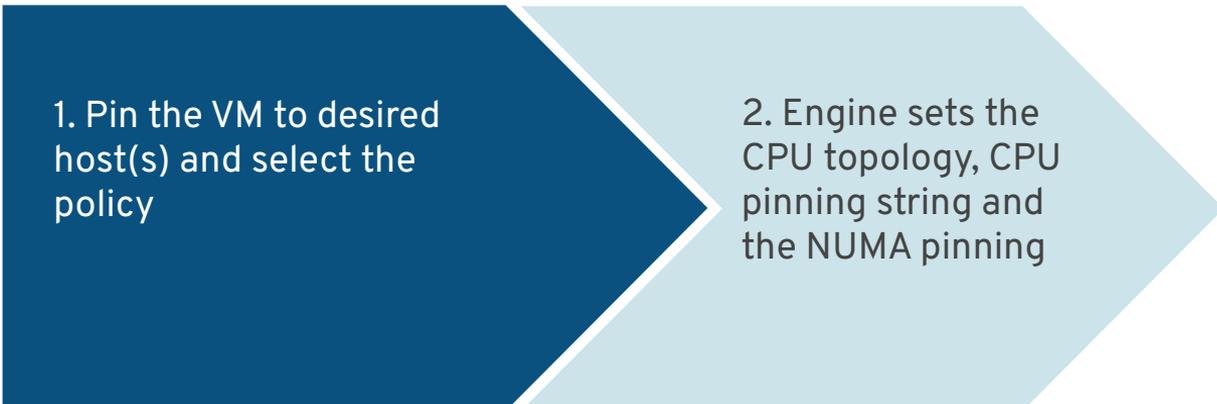
CPU Pinning Policy - Dedicated

- oVirt 4.5: Dedicated CPUs
- No host pinning is required
- The new policy will make CPU pinning exclusive (a vCPU will get exclusiveness over pCPUs, other vCPUs won't be able to use it)
- An effort to have the CPU pinning policies similar to that of OpenStack
 - Requires CPU assignments on runtime!



Source: clockwise.software (CC BY-NC-SA)

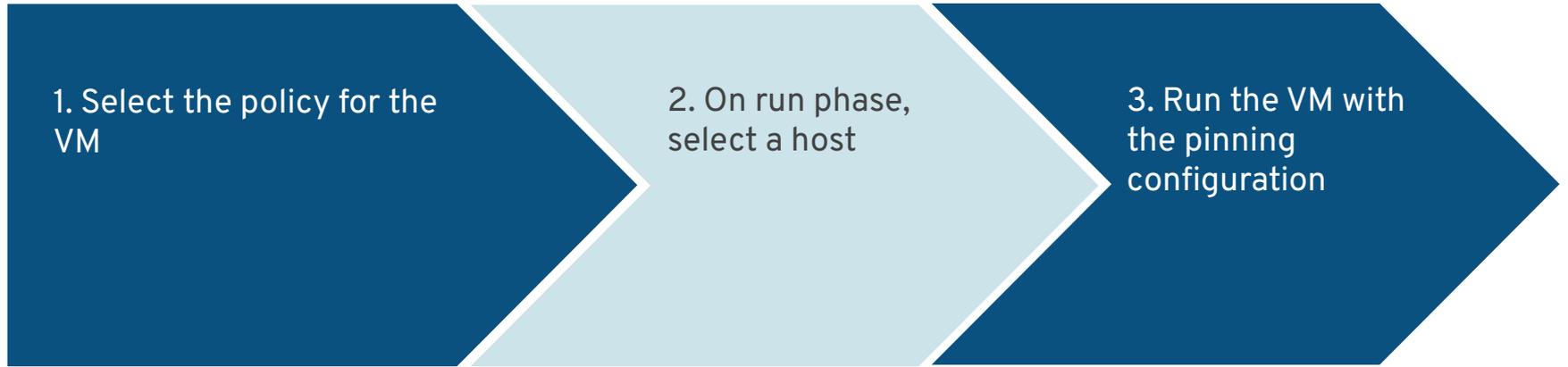
Old Resize and Pin flow



1. Pin the VM to desired host(s) and select the policy

2. Engine sets the CPU topology, CPU pinning string and the NUMA pinning

New Resize and Pin flow



What is next? What is left?

- Pin policy
- Hugepages configuration:
 - This very depends on the user requirements and the hosts
 - Requires preparations
 - Can fail to run a VM
 - 1 GB hugepages can harm migration flows (which size we need?)
- Dedicated CPUs policy is planned for oVirt 4.5
<https://ovirt.org/develop/release-management/features/virt/dedicated-cpu.html>
- FOSDEM'19 - High Performance VMs
https://archive.fosdem.org/2019/schedule/event/vai_high_performance_vms/

oVirt

Thank you!

<https://ovirt.org/>

users@ovirt.org

lrotenbe@redhat.com



@ovirt