SUSE Best Practices for SAP HANA on KVM

SUSE Linux Enterprise Server for SAP Applications 12 SP2

SUSE Linux Enterprise Server for SAP Applications 12 SP2

Matt Fleming, Senior Software Engineer (SUSE)

Lee Martin, SAP Architect & Technical Manager (SUSE)

SUSE Best Practices for SAP HANA on KVM

SUSE Linux Enterprise Server for SAP Applications 12 SP2

This best practice document describes how SUSE Linux Enterprise Server for SAP Applications 12 SP2 with KVM should be configured to run SAP HANA for use in production environments. Configurations which are not set up according to this best practice guide are considered as unsupported by SAP for production workloads.

While this document is not compulsory for non-production SAP HANA workloads, it may still be useful to help ensure optimal performance in such scenarios.

Disclaimer: Documents published as part of the SUSE Best Practices series have been contributed voluntarily by SUSE employees and third parties. They are meant to serve as examples of how particular actions can be performed. They have been compiled with utmost attention to detail. However, this does not guarantee complete accuracy. SUSE cannot verify that actions described in these documents do what is claimed or whether actions described have unintended consequences. SUSE LLC, its affiliates, the authors, and the translators may not be held liable for possible errors or the consequences thereof.

Contents

1 Introduction

This best practice document describes how SUSE Linux Enterprise Server for SAP Applications 12 SP2 with KVM should be configured to run SAP HANA for use in production environments. The setup of the SAP HANA system or other components like HA clusters are beyond the scope of this document.

The following sections describe how to set up and configure the three KVM components required to run SAP HANA on KVM:

- *Section 3, "Hypervisor"* - The host operating system running the Hypervisor directly on the server hardware

- *Section 4, "Guest VM XML Configuration"* - The libvirt domain XML description of the guest VM

- *Section 5, "Guest Operating System"* - The operating system inside the VM where SAP HANA is running

Follow *Section 2, "Supported Scenarios and Prerequisites"* and the respective SAP Notes to ensure a supported configuration. Most of the configuration options are specific to the libvirt package and therefore require modifying the VM guest's domain XML file.

1.1 Definitions

- Hypervisor: The software running directly on the physical sever to create and run VMs (Virtual Machines).

- Virtual Machine: is an emulation of a computer.

- Guest OS: The Operating System running inside the VM (Virtual Machine). This is the OS running SAP HANA and therefore the one that should be checked for SAP HANA support as per SAP Note 2235581 "SAP HANA: Supported Operating Systems" (https://launchpad.support.sap.com/#/notes/2235581 ⌐) and the "SAP HANA Hardware Directory" (https://www.sap.com/dmc/exp/2014-09-02-hana-hardware/enEN/appliances.html ⌐).

- Paravirtualization: Allows direct communication between the Hypervisor and the VM Guest resulting in a lower overhead and better performance.

- libvirt: A management interface for KVM.

SUSE Best Practices for SAP HANA on KVM

- qemu: The virtual machine emulator, also seen a process on the Hypervisor running the VM.

- SI units: Some commands and configurations use the decimal prefix (for example GB), while other use the binary prefix (for example GiB). In this document we use the binary prefix where possible.

For a general overview of the technical components of the KVM architecture, refer to https://documentation.suse.com/sles/12-SP4/html/SLES-all/cha-kvm-intro.html ↗

1.2 SAP HANA Virtualization Scenarios

SAP supports virtualization technologies for SAP HANA usage on a per scenario basis:

- Single-VM - One VM per Hypervisor/physical server for SAP HANA Scale-Up (NOTE: SAP does not allow any other VM or workload on the same server)

- Multi-VM - Multiple VM's per Hypervisor/physical server for SAP HANA Scale-Up

- Scale-Out - For SAP HANA Scale-Out

See SAP Notes:

- SAP Note 2284516 "SAP HANA virtualized on SUSE Linux Enterprise Hypervisors" (https://launchpad.support.sap.com/#/notes/2284516 ↗) for details

- SAP Note 2607144 "SAP HANA on KVM included with SLES for SAP Applications 12 SP2 in production" (https://launchpad.support.sap.com/#/notes/2607144 ↗)

2 Supported Scenarios and Prerequisites

Follow this SUSE Best Practices for SAP HANA on KVM - SUSE Linux Enterprise Server for SAP Applications 12 SP2 document which describes the steps necessary to create a supported SAP HANA on KVM configuration. SUSE Linux Enterprise Server for SAP Applications must be used for both Hypervisor and Guest.

Inquiries about scenarios not listed here should be directed to `saphana@suse.com`

2.1 Supported Scenarios

At the time of this publication the following configurations are supported for production use:

TABLE 1: **SUPPORTED COMBINATIONS**

CPU Architecture	SAP HANA scale-up (single VM)	SAP HANA scale-up (multi VM)	SAP HANA Scale-Out
Haswell (Intel v3)	- Hypervisor: SLES for SAP 12 SP2 - Guest: SLES for SAP 12 SP1 onward - Size: max. 4 socket[1], 2 TiB RAM	no	no

[1] Maximum 4 sockets using Intel standard chipsets on a single system board, for example Lenovo* x3850, HPE*/SGI* UV300 etc.

Check the following SAP Notes for the latest details of supported SAP HANA on KVM scenarios.

- SAP Note 2284516 "SAP HANA virtualized on SUSE Linux Enterprise Hypervisors" (https://launchpad.support.sap.com/#/notes/2284516 ↗)

- SAP Note 2607144 "SAP HANA on KVM included with SLES for SAP Applications 12 SP2 in production" (https://launchpad.support.sap.com/#/notes/2607144 ↗)

2.2 Sizing

It is recommended to reserve the following resources for the Hypervisor:

- 7% RAM

- 1x Physical CPU core (2x Logical CPU/Hyperthreads) per Socket

2.2.1 Memory Sizing

Since SAP HANA runs inside the VM, it is the RAM size of the VM which must be used as the basis for SAP HANA Memory sizing.

2.2.2 CPU Sizing

In addition to the above mentioned CPU core reservation for the Hypervisor (see *Section 4.3, "vCPU and vNUMA Topology"* section for details), some artificial workload tests on Intel Haswell CPUs have shown an approximately 20% overhead when running SAP HANA on KVM. Therefore a thorough test of the configuration for the required workload is highly recommended before "go live".

There are two main ways to deal with CPU sizing from a sizing perspective:

1. Follow the fixed core-to-memory ratios for SAP HANA as defined by SAP:

   - The certification of the SAP HANA Appliance hardware to be used for KVM prescribes a fixed maximum amount of memory (RAM) which is allowed for each CPU core, also known as "core-to-memory ratio". The specific ratio also depends on what workload the system will be used for, that is the Appliance Type: OLTP (Scale-up: SoH/S4H) or OLAP (Scale-up: BWoH/BW4H/DM/SoH/S4H).

   - The relevant core-to-memory ratio required to size a VM can be easily calculated as follows:

     - Go to the "SAP HANA Certified Hardware Directory" https://www.sap.com/dmc/exp/2014-09-02-hana-hardware/enEN/appliances.html ↗.

     - Select the required SAP HANA Appliance and Appliance Type (for example Haswell for BWoH).

     - Look for the largest certified RAM size for the number of CPU Sockets on the server (for example 2048 GiB on 4-Socket).

     - Look up the number of cores per CPU of this CPU Architecture used in SAP HANA Appliances. The CPU model numbers are listed at: https://www.sap.com/dmc/exp/2014-09-02-hana-hardware/enEN/index.html#details ↗ (for example 18).

     - Using the above values calculate the total number of cores on the certified Appliance by multiplying number of sockets by number of cores (for example 4x18 = 72).

     - Now divide the Appliance RAM by the total number of cores (not hyperthreads) to give you the "core-to-memory" ratio. (for example 2048 GiB/72 = approx. 28 GiB per core).

- Calculate the RAM size the VM needs to be compliant with the appropriate core-to-memory ratio defined by SAP:

  - Take the total number of CPU cores (not hyperthreads) on the Hypervisor and subtract one core per socket for the Hypervisor (for example 72-4 = 68).

  - Now take account of the Hypervisor overhead by multiplying the previous value by a factor of "1-overhead" (for example 1 - 0.20% = factor 0.8, so 68*0.8 = 55 effective cores).

  - Multiply the resulting number of CPU cores for the VM by the SAP HANA core-to-memory ratio to calculate to maximum VM RAM size limit by SAP for this amount of CPU power (for example 55 effective cores * 28 GiB per core = 1540 GiB Max VM RAM size for BWoH).

  - Now, calculate the maximum VM RAM size limit by SUSE by checking the *Table 1, "Supported Combinations"* table in this document for the maximum supported KVM Hypervisor RAM size for SAP HANA and then subtracting the 7% memory overhead (for example 2048 GiB * 0.93 (the 7% RAM overhead) = 1904 GiB Max VM RAM size).

- Finally, the actual RAM size of the VM to be configured must not exceed the LOWEST of the two above calculated SAP and SUSE "Max VM RAM size" limits.

- Conclusion:

  - Based on the example given above: From available CPU power in the VM, SAP would allow a maximum RAM size of up to 1540 GiB for a VM running OLAP/BWoH when following the predefined core-to-memory ratio.

  - Since OLTP/SoH has a much higher core-to-memory ratio (43 GiB/core) SAP would allow a maximum of 2611 GiB, which is well above the 1904 GiB limit for KVM in the example above.

- See the table *Table 2, "SAP HANA core-to-memory ratio examples"* below for some current examples of SAP HANA core-to-memory ratios.

2. Follow the SAP HANA TDI "Phase 5" rules as defined by SAP:

   - SAP HANA TDI Phase 5 rules allow customers to deviate from the above described SAP HANA core-to-memory sizing ratios in certain scenarios. The KVM implementation must still however adhere to the SUSE Best Practices for SAP HANA on KVM - SUSE Linux Enterprise Server for SAP Applications 12 SP2. Details on SAP HANA TDI Phase 5 can be found in the following blog from SAP: https://blogs.sap.com/2017/09/20/tdi-phase-5-new-opportunities-for-cost-optimization-of-sap-hana-hardware/ .

   - Since SAP HANA TDI Phase 5 rules use SAPS based sizing, SUSE recommends applying the same overhead as measured with SAP HANA on KVM for the respective KVM Version/CPU Architecture. SAPS values for servers can be requested from the respective hardware vendor.

   - The following SAP HANA sizing documentation should also be useful

     - SAP Best Practice "Sizing Approaches for SAP HANA": https://websmp203.s-ap-ag.de/~sapidb/011000358700000050632013E

     - Sizing SAP HANA on help.sap.com: https://help.sap.com/doc/eb3777d5495d46c5b2fa773206bbfb46/2.0.02/en-US/d4a122a7bb57101493e3f5ca08e6b039.html

     - SAP Sizing at: http://sap.com/sizing

TABLE 2: SAP HANA CORE-TO-MEMORY RATIO EXAMPLES

CPU Architecture	Appliance Type	Max Memory Size	Sockets	Cores per Socket	SAP HANA core-to-memory ratio
Haswell (Intel v3)	OLTP	3072 GiB	4	18	43 GiB/core
Haswell (Intel v3)	OLAP	2048 GiB	4	18	28 GiB/core

2.3   KVM Hypervisor Version

The Hypervisor must be configured according to this "SUSE Best Practices for SAP HANA on KVM - SUSE Linux Enterprise Server for SAP Applications 12 SP2" guide and fulfill the following minimal requirements:

- SUSE Linux Enterprise Server for SAP Applications 12 SP2 ("Unlimited Virtual Machines" subscription)

    - kernel (Only major version 4.4, minimum package version 4.4.49-92.11)

    - libvirt (Only major version 2.0, minimum package version 2.0.0-27.12.1)

    - qemu (Only major version 2.6, minimum package version 2.6.2-41.9.1)

2.4   Hypervisor Hardware

Use SAP HANA certified servers and storage as per SAP HANA Hardware Directory at: https://www.sap.com/dmc/exp/2014-09-02-hana-hardware/enEN/appliances.html ↗

2.5 Guest VM

The guest VM must:

- Run SUSE Linux Enterprise Server for SAP Applications 12 SP1 or later.

- Be a SUSE Linux Enterprise Server Supported VM Guest as per Section 7.1 "Supported VM Guests" of the SUSE Virtualization Guide (https://documentation.suse.com/sles/12-SP4/html/SLES-all/cha-virt-support.html#virt-support-guests ↗).

- Comply with KVM limits as per SUSE Linux Enterprise Server 12 SP2 release notes https://www.suse.com/releasenotes/x86_64/SUSE-SLES/12-SP2/#TechInfo.KVM ↗.

- Fulfill the SAP HANA HWCCT storage KPI's as per SAP Note 1943937 "Hardware Configuration Check Tool - Central Note" (https://launchpad.support.sap.com/ ↗ notes/1943937) and SAP Note 2501817 "HWCCT 1.0 (≥ 220)" (https://launchpad.support.sap.com/ ↗ notes/2501817). Refer to *Section 4.4, "Storage"* for storage configuration details.

- Be configured according to this SUSE Best Practices for SAP HANA on KVM - SUSE Linux Enterprise Server for SAP Applications 12 SP2 document.

3 Hypervisor

3.1 KVM Hypervisor Installation

For details refer to Section 6.4 Installation of Virtualization Components of the SUSE Virtualization Guide (https://documentation.suse.com/sles/12-SP4/html/SLES-all/cha-vt-installation.html#sec-vt-installation-patterns ↗)

Install the KVM packages using the following Zypper patterns:

```
zypper in -t pattern kvm_server kvm_tools
```

In addition, it is also useful to install the "lstopo" tool which is part of the "hwloc" package contained inside the "HPC Module" for SUSE Linux Enterprise Server.

3.2 Configure Networking on Hypervisor

To achieve maximum performance required for productive SAP HANA workloads the PCI address of the respective network port(s) must be assigned directly into the KVM Guest VM to ensure that the Guest VM has enough network device channels to accommodate the network traffic. Ideally the VM Guest should be able to access the same number of network device channels as the host, this can be checked and compared between host and guest VM with `ethtool -l <device>`, for example:

```
# ethtool -l eth1
Channel parameters for eth1:
Pre-set maximums:
RX:             0
TX:             0
Other:          1
Combined:       63
Current hardware settings:
RX:             0
TX:             0
Other:          1
Combined:       63
```

3.2.1 Assign Network Port at PCI NIC Level

The required network port(s) should be assigned to the Guest VM as described in section "14.10.2 Adding a PCI Device with virsh" in the SUSE Virtualization Guide for SUSE Linux Enterprise Server 12 SP2(https://documentation.suse.com/sles/12-SP2/ ↗)

Persist detach of PCI NIC port. Before starting the VM, the PCI NIC port must be detached from the Hypervisor OS, otherwise the VM will not start. The PCI NIC detach can be automated at boot time by creating a service file (after-local.service) pointing to /etc/init.d/after.local which contains the commands to detach the NIC.

Create the systemd unit file /etc/systemd/system/after.local.

```
[Unit]
Description=/etc/init.d/after.local Compatibility
After=libvirtd.service
Requires=libvirtd.service
[Service]
Type=oneshot
ExecStart=/etc/init.d/after.local
```

```
RemainAfterExit=true

[Install]
WantedBy=multi-user.target
```

Then create the script /etc/init.d/after.local which will detach the PCI device (where "pci_xxxx_xx_xx_0" must be replaced with the correct PCI address).

```
#! /bin/sh
#
# Copyright (c) 2010 SuSE LINUX Products GmbH, Germany.  All rights reserved.
# ...
virsh nodedev-detach pci_xxxx_xx_xx_0
```

3.3   Storage Configuration on Hypervisor

As with compute resources, the storage used for running SAP HANA must also be SAP certified. Therefore only the storage from SAP HANA Appliances or SAP HANA Certified Enterprise Storage (https://www.sap.com/dmc/exp/2014-09-02-hana-hardware/enEN/enterprise-storage.html ↗) is supported. In all cases the SAP HANA storage configuration recommendations from the respective hardware vendor and the SAP HANA Storage Requirements for TDI (hhttps://www.sap.com/documents/2016/05/e8705aae-717c-0010-82c7-eda71af511fa.html (https://www.s-ap.com/documents/2016/05/e8705aae-717c-0010-82c7-eda71af511fa.html) ↗) should be followed. The SUSE Best Practices for SAP HANA on KVM - SUSE Linux Enterprise Server for SAP Applications 12 SP2 has been designed and tested to map the block devices for SAP HANA on the Hypervisor directly into the VM. Therefore any LVM (Logical Volume Manager) configuration should be made inside the Guest VM only. Multipathing by contrast should be only configured on the Hypervisor.

Ultimately the storage for SAP HANA must be able to fulfill the SAP HANA HWCCT requirements from within the VM. For details on HWCCT and the required storage KPI's refer to SAP Note 1943937 "Hardware Configuration Check Tool - Central Note" (https://launchpad.support.sap.com/ ↗ notes/1943937) and SAP Note 2501817 - HWCCT 1.0 ($\geq$ 220) (https://launchpad.support.sap.com/ ↗ notes/2501817).

Network Attached Storage has not been tested with SAP HANA on KVM, if there is a requirement for this contact saphana@suse.com.

Most of the configuration steps to configure the storage are at the Guest VM XML level, see section *Section 4.4, "Storage"*. Nevertheless storage on the Hypervisor should:

- Follow the storage layout recommendations from the appropriate hardware vendor.

- Not use LVM (Logical Volume Manager) on the Hypervisor level for SAP HANA volumes since nested LVM is not supported.

- Configure Multipathing on the Hypervisor only, not inside the Guest VM.

3.4 Hypervisor Operating System Configuration

3.4.1 tuned

Install "tuned" and set the profile to "latency-performance". Do not use the "sap-hana profile" on the Hypervisor. This can be configured with the following commands:

```
zypper in tuned

systemctl enable tuned

systemctl start tuned

tuned-adm profile latency-performance
```

3.4.1.1 Verify "tuned" Has Set CPU Frequency Governor and Performance Bias

The CPU frequency governor should be set to "performance" to avoid latency issues because of ramping the CPU frequency up and down in response to changes in the system's load. The governor setting can be verified with the following command to check what is set under "current policy":

```
cpupower -c all  frequency-info
```

Additionally the performance bias setting should also be set to 0 (performance). The performance bias setting can be verified with the following command:

```
cpupower -c all info
```

3.4.2 irqbalance

The irqbalance service should be disabled because it can cause latency issues when the /proc/irq/* files are read. To disable irqbalance run the following command:

```
systemctl disable irqbalance.service

systemctl stop irqbalance.service
```

3.4.3 Customize the Linux Kernel Boot Options

To edit the boot options for the Linux kernel to the following:

1. Edit `/etc/defaults/grub` and add the following boot options to the line "GRUB_CMDLINE_LINUX_DEFAULT" (A detailed explanation of these options follows).

   ```
   numa_balancing=disable   kvm_intel.ple_gap=0  transparent_hugepage=never
     elevator=deadline  intel_idle.max_cstate=1  processor.max_cstate=1
    default_hugepagesz=1GB hugepagesz=1GB hugepages=<number of hugepages>
   ```

   Note: Calculating Value

   The value for "<number of hugepages>" should be calculated by taking the number GiB's of RAM minus approx. 7% for the Hypervisor OS. For example 2 TiB RAM (2048 GiB) minus 7% are approx. 1900 hugepages

2. Run the following command:

   ```
   grub2-mkconfig -o /boot/grub2/grub.cfg
   ```

3. Reboot the system:

   ```
   reboot
   ```

3.4.4 Technical Explanation of the Above Described Configuration Settings

Automatic NUMA Balancing (numa_balancing = disable)

Automatic NUMA balancing can result in increased system latency and should therefore be disabled.

KVM PLE-GAP (kvm_intel.ple_gap = 0)

Pause Loop Exit (PLE) is a feature whereby a spinning guest CPU releases the physical CPU until a lock is free. This is useful in cases where multiple virtual CPUs are using the same physical CPU but causes unnecessary delays when a guest is not overcommitted.

Transparent Hugepages (transparent_hugepage = never)

Because 1G pages are used for the virtual machine, then there is no additional benefit from having THP enabled. Disabling it will avoid khugepaged interfering with the virtual machine while it scans for pages to promote to hugepages.

I/O Scheduler (elevator = deadline)

The deadline I/O scheduler should be used for all disks/LUNs mapped into the KVM guest.

Processor C-states (intel_idle.max_cstate = 1 processor.max_cstate = 1)

The processor will attempt to save power when idle by switching to a lower power state. Unfortunately this incurs latency when switching in and out of these states. Optimal performance is achieved by limiting the processor to states C0 (normal running state) and C1 (first lower power state). Note that while there is an exit latency associated with C1 states, it is offset on hyperthread-enabled platforms by the fact sibling cores can borrow resources from sibling cores if they are in the C1 state and some CPUs can boost the CPU frequency higher if siblings are in the C1 state.

Hugepages (default_hugepagesz = 1 GB hugepagesz = 1 GB hugepages = <number of hugepages>)

The use of 1 GiB hugepages is to reduce overhead and contention when the guest is updating its page tables. This requires allocation of 1 GiB hugepages on the host. The number of pages to allocate depends on the memory size of the guest. 1 GiB pages are not pageable by the OS, so they always remain in RAM and therefore the "locked" definition in libvirt XML files is not required. It also important to ensure the order of the hugepage options, specifically the "number of hugepages" option must come after the 1 GiB hugepage size definitions.

The value for "<number of hugepages>" should be calculated by taking the number GiB's of RAM minus approx. 7% for the Hypervisor OS. For example 2 TiB RAM (2048 GiB) minus 7% are approx. 1900 hugepages.

4 Guest VM XML Configuration

This section describes the modifications required to the libvirt XML definition of the Guest VM. The libvirt XML may be edited using the following command:

```
virsh edit <Guest VM name>
```

4.1 Create an Initial Guest VM XML

Refer to section 9 "Guest Installation" of the SUSE Virtualization Guide (https://documenta-tion.suse.com/sles/12-SP4/html/SLES-all/cha-kvm-inst.html ↗).

4.2 Global vCPU Configuration

Ensure that the following XML elements are configured:

- domain XML supports "xmlns:qemu" to use qemu commands directly

- architecture and machine type are set to match the qemu version installed on the Hypervisor

    - for example "2.6" for qemu 2.6 on SUSE Linux Enterprise Server for SAP Applications 12 SP2

- cpu mode is set to "host-passthrough"

- the defined qemu CPU command lines necessary for SAP HANA support are used

The following XML example demonstrates how to configure this:

```
<domain type='kvm' xmlns:qemu='http://libvirt.org/schemas/domain/qemu/1.0'>
...
    <os>
       <type arch='x86_64' machine='pc-i440fx-2.6'>hvm</type>
    ...
    </os>
    ...
    <cpu mode='host-passthrough'>
    ...
    </cpu>
    ...
    <qemu:commandline>
```

```
        <qemu:arg value='-cpu'/>
        <qemu:arg value='host,migratable=off,+invtsc,l3-cache=on'/>
    </qemu:commandline>
</domain>
```

Explanation of the critical "l3-cache" option: If a KVM guest has multiple vNUMA nodes it is critical that any L3 CPU cache present on the host be mirrored in the KVM guest. When vCPUs share an L3 cache the Linux kernel scheduler can use an optimized mechanism for enqueuing tasks on vCPUs. Without L3 cache information the guest kernel will always use a more expensive mechanism that involves Inter-Processor Interrupts (IPIs).

Explanation of the "host,migratable-off,+invtsc" options: For best performance, SAP HANA requires the invtsc CPU feature in the KVM guest. However, KVM will remove any non-migratable CPU features from the virtual CPU presented to the KVM guest. This behavior can be overridden by passing the 'migratable = off' and ' + invtsc' values to the '-cpu' option.

4.3   vCPU and vNUMA Topology

To achieve maximum performance and be supported for use with SAP HANA the KVM guest's NUMA topology should exactly mirror the host's NUMA topology and not overcommit memory or CPU resources. This requires pinning virtual CPUs to unique physical CPUs (no virtual CPUs should share the same hyperthread/ physical CPU) and configuring virtual NUMA node relationships for those virtual CPUs.

Note: Physical CPU Core

One physical CPU core (that is 2 hyperthreads) per NUMA node should be left unused by KVM guests so that IOThreads can be pinned there.

Note: Hypervisor Topology

In many use cases it is advisable to map the Hyperthreading topology into the Guest VM as described below since this allows SAP HANA to spread workload threads across many vCPUs. However there maybe workloads which perform better without hyperthreading. In this case only the first physical hyperthread from each core should be mapped into the VM. In the simplified example below that would mean only mapping host processor 0-15 into the VM.

It is important to note that KVM/QEMU uses a static hyperthread sibling CPU APIC ID assignment for virtual CPUs irrespective of the actual physical CPU APIC ID values on the host. For example, assuming that the first hyperthread sibling pair is CPU 0 and CPU 16 on the Hypervisor host, you will need to pin that sibling pair to vCPU 0 and vCPU 1.

Below is a table of a hypothetical configuration for a "4-socket NUMA topology with 4 cores per socket and hyperthreading" server to help understand the above logic. In real world SAP HANA scenarios CPUs will typically have 18+ CPU cores, and will therefore have far more CPUs for the Guest compared to iothreads.

VM Guest vCPU #	Physical Server Numa node #	Physical Server "core id"	Physical Server processor #
emulator	0	0	0
emulator	0	0	16
0	0	1	1
1	0	1	17
2	0	2	2
3	0	2	18
4	0	3	3
5	0	3	19
emulpin 1	1	0	4
emulpin 4	1	0	20
6	1	1	5
7	1	1	21
8	1	2	6
9	1	2	22
10	1	3	7
11	1	3	23
iohtread 2	2	0	8
iothread 5	2	0	24
12	2	1	9
13	2	1	25
14	2	2	10
15	2	2	26
16	2	3	11
17	2	3	27
iothread 3	3	0	12
iothread 6	3	0	28
18	3	1	13
19	3	1	29
20	3	2	14
21	3	2	30
22	3	3	15
23	3	3	31

The following commands can be used to determine the CPU details on the Hypervisor host (see Appendix for an *Section 7.2, "Example "lscpu --extended=CPU,SOCKET,CORE" from a Lenovo x3850 x6"* and an *Section 7.3, "Example "lstopo-no-graphics" from a Lenovo x3850 x6"*):

```
lscpu --extended=CPU,SOCKET,CORE

lstopo-no-graphics
```

Using the above information the CPU and memory pinning section of the Guest VM XML can be created. Below is an example based on the hypothetical example above.

Make sure to take note of the following configuration points:

- The "vcpu placement" element lists the total number of vCPUs in the Guest.

- The "iothreads" element lists the total number of iothreads (6 in this example).

  - iothreads should be pinned to the Sockets where the respective storage is physically attached. This mapping can be found by looking for the "Block(Disk)" entries in output from "lstopo-no-graphics", see Appendix *Section 7.3, "Example "lstopo-no-graphics" from a Lenovo x3850 x6"*.

- The "cputune" element contains the attributes describing the mappings of vCPU, emulator and iothreads to physical CPUs.

- The "numatune" element contains the attributes to describe distribution of RAM across the virtual NUMA nodes (CPU sockets).

  - The "mode" attribute should be set to "strict".

  - The appropriate number of nodes should be entered in the "nodeset" and "memnode" attributes. In this example there are 4 sockets, therefore nodeset = 0-3 and cellid 0 to 3.

- The "cpu" element lists:

  - "mode" attribute which should be set to "host-passthrough" for SAP HANA.

  - "topology" attributes to describe the vCPU NUMA topology of the Guest. In this example, 4 sockets, each with 3 cores (see the cpu pinning table) and 2 hyperthreads per core. Set "threads = 1" if hyperthreading is not to be used.

- The attributes of the "numa" elements to describes which vCPU number ranges belong to which NUMA node/socket. Care should be taken since these number ranges are not the same as on the Hypervisor host.

- In addition, the attributes of the "numa" elements also describe how much RAM should be distributed per NUMA node. In this 4-node example enter 25% (or 1/4) of the entire Guest VM Memory. Also refer to *Section 4.5, "Memory Backing"* and *Section 2.2.1, "Memory Sizing"* Memory section of this paper for further details.

```
<vcpu placement='static'>24</vcpu>
<iothreads>6</iothreads>
  <cputune>
    <vcpupin vcpu='0' cpuset='1'/>
    <vcpupin vcpu='1' cpuset='17'/>
    <vcpupin vcpu='2' cpuset='2'/>
    <vcpupin vcpu='3' cpuset='18'/>
    <vcpupin vcpu='4' cpuset='3'/>
    <vcpupin vcpu='5' cpuset='19'/>

    <vcpupin vcpu='6' cpuset='5'/>
    <vcpupin vcpu='7' cpuset='21'/>
    <vcpupin vcpu='8' cpuset='6'/>
    <vcpupin vcpu='9' cpuset='22'/>
    <vcpupin vcpu='10' cpuset='7'/>
    <vcpupin vcpu='11' cpuset='23'/>

    <vcpupin vcpu='12' cpuset='9'/>
    <vcpupin vcpu='13' cpuset='25'/>
    <vcpupin vcpu='14' cpuset='10'/>
    <vcpupin vcpu='15' cpuset='26'/>
    <vcpupin vcpu='16' cpuset='11'/>
    <vcpupin vcpu='17' cpuset='27'/>

    <vcpupin vcpu='18' cpuset='13'/>
    <vcpupin vcpu='19' cpuset='29'/>
    <vcpupin vcpu='20' cpuset='14'/>
    <vcpupin vcpu='21' cpuset='30'/>
    <vcpupin vcpu='22' cpuset='15'/>
    <vcpupin vcpu='23' cpuset='31'/>

    <emulatorpin cpuset='0,16'/>

    <iothreadpin iothread='1' cpuset='4'/>
    <iothreadpin iothread='2' cpuset='8'/>
    <iothreadpin iothread='3' cpuset='12'/>
```

```
    <iothreadpin iothread='4' cpuset='20'/>
    <iothreadpin iothread='5' cpuset='24'/>
    <iothreadpin iothread='6' cpuset='28'/>
  </cputune>

  <numatune>
    <memory mode='strict' nodeset='0-3'/>
    <memnode cellid='0' mode='strict' nodeset='0'/>
    <memnode cellid='1' mode='strict' nodeset='1'/>
    <memnode cellid='2' mode='strict' nodeset='2'/>
    <memnode cellid='3' mode='strict' nodeset='3'/>
  </numatune>

  <cpu mode='host-passthrough'>
    <topology sockets='4' cores='3' threads='2'/>
    <numa>
      <cell id='0' cpus='0-5' memory='<Memory per NUMA node>' unit='KiB'/>
      <cell id='1' cpus='6-11' memory='<Memory per NUMA node>' unit='KiB'/>
      <cell id='2' cpus='12-17' memory='<Memory per NUMA node>' unit='KiB'/>
      <cell id='3' cpus='18-23' memory='<Memory per NUMA node>' unit='KiB'/>
    </numa>
  </cpu>
```

Note: Memory Unit

The memory unit can be set to GiB to ease the memory computations.

4.4 Storage

4.4.1 Storage Configuration for Operating System Volumes

The performance of storage where the Operating System is installed is not critical for the performance of SAP HANA, and therefore any KVM supported storage may be used to deploy the Operating system itself.

4.4.2 Storage Configuration for SAP HANA Volumes

The Guest VM XML configuration must be based on the underlying storage configuration on the Hypervisor, see section *Section 3.3, "Storage Configuration on Hypervisor"* for details and adhere the following recommendations:

- Follow the storage layout recommendations from the appropriate hardware vendors.

- Only use the KVM virtio "threads" driver

- Distribute block devices evenly across all available iothreads (see *Section 4.4.3, "IOThreads"*)

  - Avoid placing devices for SAP HANA data and log on the same iothreads.

  - Pin iothreads to the CPU sockets where the storage is attached, see section *Section 4.3, "vCPU and vNUMA Topology"* for details.

- Set the following virtio attributes: name='qemu' type='raw' cache='none' io='threads'.

- Use persistent device names in the Guest VM XML configuration (see example in *Section 4.4.3, "IOThreads"*).

4.4.3 IOThreads

As described in section *Section 4.3, "vCPU and vNUMA Topology"*, iothreads should be pinned to a set of physical CPUs which are not presented to the Guest VM OS.

Below is an example (device names and bus addresses are configuration dependent) of how to add the iothread options to a virtio device. Note that the iothread numbers should be distributed across the respective virtio devices.

```
<disk type='block' device='disk'>
   <driver name='qemu' type='raw' cache='none' io='threads' iothread='1'/>
   <source dev='/dev/disk/by-id/<source device path>'/>
   <target dev='vda' bus='virtio'/>
</disk>
```

For further details refer to section 12 "Managing Storage" in the SUSE Virtualization Guide (https://documentation.suse.com/sles/12-SP4/html/SLES-all/cha-libvirt-storage.html ↗)

4.5   Memory Backing

Configure the memory size of the Guest VM in KiB and in multiples of 1 GiB (because of the use of 1 GiB hugepages). The max VM size is determined by the total number of 1 GiB hugepages defined on the Hypervisor OS as described in section 4.3.

```
<memory unit='KiB'><enter memory size in KiB here></memory>
<currentMemory unit='KiB'><enter memory size in KiB here></currentMemory>
```

It is important to use 1 gigabyte hugepages for the guest VM memory backing to achieve optimal performance of the KVM guest. In addition, Kernel Same Page Merging (KSM) should be disabled.

```
<memoryBacking>
  <hugepages>
     <page size='1048576' unit='KiB'/>
  </hugepages>
  <nosharepages/>
</memoryBacking>
```

4.6   Virtio Random Number Generator (RNG) Device

The host /dev/random file should be passed through to QEMU as a source of entropy using the virtio RNG device:

```
<rng model='virtio'>
   <backend model='random'>/dev/random</backend>
   <alias name='rng0'/>
</rng>
```

5   Guest Operating System

5.1   Install SUSE Linux Enterprise Server for SAP Applications Inside the Guest VM

Refer to the SUSE Guide "SUSE Linux Enterprise Server for SAP Applications 12 SP2" (https://documentation.suse.com/sles-sap/12-SP2/ ⬀).

5.2   Guest Operating System Configuration for SAP HANA

Install and configure SUSE Linux Enterprise Server for SAP Applications 12 SP2 and SAP HANA as described in:

- SAP Note 1944799 "SAP HANA Guidelines for SLES Operating System Installation" (https://launchpad.support.sap.com/#/notes/1944799 ↗)

- SAP Note 2205917 "SAP HANA DB: Recommended OS settings for SLES 12 / SLES for SAP Applications 12" (https://launchpad.support.sap.com/#/notes/2205917 ↗)

irqbalance

The irqbalance service should be disabled because it can cause latency issues when the /proc/irq/* files are read. To disable irqbalance run the following command:

```
systemctl disable irqbalance.service
systemctl stop irqbalance.service
```

5.3   Guest Operating System Storage Configuration for SAP HANA Volumes

- Follow the storage layout recommendations from the appropriate hardware vendors.

- Only use LVM (Logical Volume Manager) inside the VM for SAP HANA. Nested LVM is not to be used.

- Do not configure Multipathing in the guest, but instead on the Hypervisor (see section *Section 3.3, "Storage Configuration on Hypervisor"*).

6   Administration

For a full explanation of administration commands, refer to official SUSE Virtualization documentation such as:

- Section 10 "Basic VM Guest Management" and others in the SUSE Virtualization Guide for SUSE Linux Enterprise Server 12 (https://documentation.suse.com/sles/12-SP4/html/SLES-all/cha-libvirt-managing.html ↗)

- SUSE Virtualization Best Practices for SUSE Linux Enterprise Server 12 (https://documentation.suse.com/sles/12-SP4/html/SLES-all/article-vt-best-practices.html ↗)

6.1   Useful Commands on the Hypervisor

Checking kernel boot options used

```
cat /proc/cmdline
```

Checking hugepage status (This command can also be used to monitor the progress of hugepage allocation during VM start)

```
cat /proc/meminfo |grep Huge
```

List all VM Guest domains configured on Hypervisor

```
virsh list --all
```

Start a VM (Note: VM start times can take some minutes on larger RAM systems, check progress with " /proc/meminfo | grep Huge "

```
virsh start <VM/Guest Domain name>
```

Shut down a VM

```
virsh shutdown <VM/Guest Domain name>
```

Location of VM Guest configuration files

```
/etc/libvirt/qemu
```

Location of VM Log files

```
/var/log/libvirt/qemu
```

6.2   Useful Commands Inside the VM Guest

Checking L3 cache has been enabled in the guest

```
lscpu | grep L3
```

Validating Guest and Host CPU Topology

```
lscpu
```

7 Examples

7.1 Example "lscpu" from a Lenovo x3850 x6

```
# lscpu
Architecture:          x86_64
CPU op-mode(s):        32-bit, 64-bit
Byte Order:            Little Endian
CPU(s):                144
On-line CPU(s) list:   0-143
Thread(s) per core:    2
Core(s) per socket:    18
Socket(s):             4
NUMA node(s):          4
Vendor ID:             GenuineIntel
CPU family:            6
Model:                 63
Model name:            Intel(R) Xeon(R) CPU E7-8880 v3 @ 2.30GHz
Stepping:              4
CPU MHz:               2700.000
CPU max MHz:           3100.0000
CPU min MHz:           1200.0000
BogoMIPS:              4589.07
Virtualization:        VT-x
L1d cache:             32K
L1i cache:             32K
L2 cache:              256K
L3 cache:              46080K
NUMA node0 CPU(s):     0-17,72-89
NUMA node1 CPU(s):     18-35,90-107
NUMA node2 CPU(s):     36-53,108-125
NUMA node3 CPU(s):     54-71,126-143
Flags:                 fpu vme de pse tsc msr pae mce cx8 apic sep mtrr pge mca cmov
 pat pse36 clflush dts acpi mmx fxsr sse sse2 ss ht tm pbe syscall nx pdpe1gb rdtscp
 lm constant_tsc arch_perfmon pebs bts rep_good nopl xtopology nonstop_tsc aperfmperf
 eagerfpu mce_recovery pni pclmulqdq dtes64 monitor ds_cpl vmx smx est tm2 ssse3 sdbg fma
 cx16 xtpr pdcm pcid dca sse4_1 sse4_2 x2apic movbe popcnt tsc_deadline_timer aes xsave
 avx f16c rdrand lahf_lm abm ida arat epb pln pts dtherm tpr_shadow vnmi flexpriority ept
 vpid fsgsbase tsc_adjust bmi1 hle avx2 smep bmi2 erms invpcid rtm cqm xsaveopt cqm_llc
 cqm_occup_llc
```

7.2 Example "lscpu --extended=CPU,SOCKET,CORE" from a Lenovo x3850 x6

```
#  lscpu --extended=CPU,SOCKET,CORE
CPU SOCKET CORE
0   0      0
1   0      1
2   0      2
3   0      3
4   0      4
5   0      5
6   0      6
7   0      7
8   0      8
9   0      9
10  0      10
11  0      11
12  0      12
13  0      13
14  0      14
15  0      15
16  0      16
17  0      17
18  1      18
19  1      19
20  1      20
21  1      21
22  1      22
23  1      23
24  1      24
25  1      25
26  1      26
27  1      27
28  1      28
29  1      29
30  1      30
31  1      31
32  1      32
33  1      33
34  1      34
35  1      35
36  2      36
37  2      37
38  2      38
39  2      39
40  2      40
```

```
41   2      41
42   2      42
43   2      43
44   2      44
45   2      45
46   2      46
47   2      47
48   2      48
49   2      49
50   2      50
51   2      51
52   2      52
53   2      53
54   3      54
55   3      55
56   3      56
57   3      57
58   3      58
59   3      59
60   3      60
61   3      61
62   3      62
63   3      63
64   3      64
65   3      65
66   3      66
67   3      67
68   3      68
69   3      69
70   3      70
71   3      71
72   0      0
73   0      1
74   0      2
75   0      3
76   0      4
77   0      5
78   0      6
79   0      7
80   0      8
81   0      9
82   0      10
83   0      11
84   0      12
85   0      13
86   0      14
87   0      15
```

```
88   0      16
89   0      17
90   1      18
91   1      19
92   1      20
93   1      21
94   1      22
95   1      23
96   1      24
97   1      25
98   1      26
99   1      27
100  1      28
101  1      29
102  1      30
103  1      31
104  1      32
105  1      33
106  1      34
107  1      35
108  2      36
109  2      37
110  2      38
111  2      39
112  2      40
113  2      41
114  2      42
115  2      43
116  2      44
117  2      45
118  2      46
119  2      47
120  2      48
121  2      49
122  2      50
123  2      51
124  2      52
125  2      53
126  3      54
127  3      55
128  3      56
129  3      57
130  3      58
131  3      59
132  3      60
133  3      61
134  3      62
```

SUSE Best Practices for SAP HANA on KVM

```
135 3      63
136 3      64
137 3      65
138 3      66
139 3      67
140 3      68
141 3      69
142 3      70
143 3      71
```

7.3   Example "lstopo-no-graphics" from a Lenovo x3850 x6

```
# lstopo-no-graphics
Machine (504GB total)
  NUMANode L#0 (P#0 126GB)
    Package L#0 + L3 L#0 (45MB)
      L2 L#0 (256KB) + L1d L#0 (32KB) + L1i L#0 (32KB) + Core L#0
        PU L#0 (P#0)
        PU L#1 (P#72)
      L2 L#1 (256KB) + L1d L#1 (32KB) + L1i L#1 (32KB) + Core L#1
        PU L#2 (P#1)
        PU L#3 (P#73)
      L2 L#2 (256KB) + L1d L#2 (32KB) + L1i L#2 (32KB) + Core L#2
        PU L#4 (P#2)
        PU L#5 (P#74)
      L2 L#3 (256KB) + L1d L#3 (32KB) + L1i L#3 (32KB) + Core L#3
        PU L#6 (P#3)
        PU L#7 (P#75)
      L2 L#4 (256KB) + L1d L#4 (32KB) + L1i L#4 (32KB) + Core L#4
        PU L#8 (P#4)
        PU L#9 (P#76)
      L2 L#5 (256KB) + L1d L#5 (32KB) + L1i L#5 (32KB) + Core L#5
        PU L#10 (P#5)
        PU L#11 (P#77)
      L2 L#6 (256KB) + L1d L#6 (32KB) + L1i L#6 (32KB) + Core L#6
        PU L#12 (P#6)
        PU L#13 (P#78)
      L2 L#7 (256KB) + L1d L#7 (32KB) + L1i L#7 (32KB) + Core L#7
        PU L#14 (P#7)
        PU L#15 (P#79)
      L2 L#8 (256KB) + L1d L#8 (32KB) + L1i L#8 (32KB) + Core L#8
        PU L#16 (P#8)
        PU L#17 (P#80)
      L2 L#9 (256KB) + L1d L#9 (32KB) + L1i L#9 (32KB) + Core L#9
        PU L#18 (P#9)
```

```
            PU L#19 (P#81)
        L2 L#10 (256KB) + L1d L#10 (32KB) + L1i L#10 (32KB) + Core L#10
            PU L#20 (P#10)
            PU L#21 (P#82)
        L2 L#11 (256KB) + L1d L#11 (32KB) + L1i L#11 (32KB) + Core L#11
            PU L#22 (P#11)
            PU L#23 (P#83)
        L2 L#12 (256KB) + L1d L#12 (32KB) + L1i L#12 (32KB) + Core L#12
            PU L#24 (P#12)
            PU L#25 (P#84)
        L2 L#13 (256KB) + L1d L#13 (32KB) + L1i L#13 (32KB) + Core L#13
            PU L#26 (P#13)
            PU L#27 (P#85)
        L2 L#14 (256KB) + L1d L#14 (32KB) + L1i L#14 (32KB) + Core L#14
            PU L#28 (P#14)
            PU L#29 (P#86)
        L2 L#15 (256KB) + L1d L#15 (32KB) + L1i L#15 (32KB) + Core L#15
            PU L#30 (P#15)
            PU L#31 (P#87)
        L2 L#16 (256KB) + L1d L#16 (32KB) + L1i L#16 (32KB) + Core L#16
            PU L#32 (P#16)
            PU L#33 (P#88)
        L2 L#17 (256KB) + L1d L#17 (32KB) + L1i L#17 (32KB) + Core L#17
            PU L#34 (P#17)
            PU L#35 (P#89)
    HostBridge L#0
        PCIBridge
            PCI 8086:1521
                Net L#0 "eth0"
            PCI 8086:1521
                Net L#1 "eth1"
            PCI 8086:1521
                Net L#2 "eth2"
            PCI 8086:1521
                Net L#3 "eth3"
NUMANode L#1 (P#1 126GB)
    Package L#1 + L3 L#1 (45MB)
        L2 L#18 (256KB) + L1d L#18 (32KB) + L1i L#18 (32KB) + Core L#18
            PU L#36 (P#18)
            PU L#37 (P#90)
        L2 L#19 (256KB) + L1d L#19 (32KB) + L1i L#19 (32KB) + Core L#19
            PU L#38 (P#19)
            PU L#39 (P#91)
        L2 L#20 (256KB) + L1d L#20 (32KB) + L1i L#20 (32KB) + Core L#20
            PU L#40 (P#20)
            PU L#41 (P#92)
        L2 L#21 (256KB) + L1d L#21 (32KB) + L1i L#21 (32KB) + Core L#21
```

```
        PU L#42 (P#21)
        PU L#43 (P#93)
      L2 L#22 (256KB) + L1d L#22 (32KB) + L1i L#22 (32KB) + Core L#22
        PU L#44 (P#22)
        PU L#45 (P#94)
      L2 L#23 (256KB) + L1d L#23 (32KB) + L1i L#23 (32KB) + Core L#23
        PU L#46 (P#23)
        PU L#47 (P#95)
      L2 L#24 (256KB) + L1d L#24 (32KB) + L1i L#24 (32KB) + Core L#24
        PU L#48 (P#24)
        PU L#49 (P#96)
      L2 L#25 (256KB) + L1d L#25 (32KB) + L1i L#25 (32KB) + Core L#25
        PU L#50 (P#25)
        PU L#51 (P#97)
      L2 L#26 (256KB) + L1d L#26 (32KB) + L1i L#26 (32KB) + Core L#26
        PU L#52 (P#26)
        PU L#53 (P#98)
      L2 L#27 (256KB) + L1d L#27 (32KB) + L1i L#27 (32KB) + Core L#27
        PU L#54 (P#27)
        PU L#55 (P#99)
      L2 L#28 (256KB) + L1d L#28 (32KB) + L1i L#28 (32KB) + Core L#28
        PU L#56 (P#28)
        PU L#57 (P#100)
      L2 L#29 (256KB) + L1d L#29 (32KB) + L1i L#29 (32KB) + Core L#29
        PU L#58 (P#29)
        PU L#59 (P#101)
      L2 L#30 (256KB) + L1d L#30 (32KB) + L1i L#30 (32KB) + Core L#30
        PU L#60 (P#30)
        PU L#61 (P#102)
      L2 L#31 (256KB) + L1d L#31 (32KB) + L1i L#31 (32KB) + Core L#31
        PU L#62 (P#31)
        PU L#63 (P#103)
      L2 L#32 (256KB) + L1d L#32 (32KB) + L1i L#32 (32KB) + Core L#32
        PU L#64 (P#32)
        PU L#65 (P#104)
      L2 L#33 (256KB) + L1d L#33 (32KB) + L1i L#33 (32KB) + Core L#33
        PU L#66 (P#33)
        PU L#67 (P#105)
      L2 L#34 (256KB) + L1d L#34 (32KB) + L1i L#34 (32KB) + Core L#34
        PU L#68 (P#34)
        PU L#69 (P#106)
      L2 L#35 (256KB) + L1d L#35 (32KB) + L1i L#35 (32KB) + Core L#35
        PU L#70 (P#35)
        PU L#71 (P#107)
  HostBridge L#7
    PCIBridge
      PCI 1000:005d
```

```
          Block(Disk) L#4 "sda"
          Block(Disk) L#5 "sdb"
          Block(Disk) L#6 "sdc"
          Block(Disk) L#7 "sdd"
          Block(Disk) L#8 "sde"
      NUMANode L#2 (P#2 126GB) + Package L#2 + L3 L#2 (45MB)
      L2 L#36 (256KB) + L1d L#36 (32KB) + L1i L#36 (32KB) + Core L#36
        PU L#72 (P#36)
        PU L#73 (P#108)
      L2 L#37 (256KB) + L1d L#37 (32KB) + L1i L#37 (32KB) + Core L#37
        PU L#74 (P#37)
        PU L#75 (P#109)
      L2 L#38 (256KB) + L1d L#38 (32KB) + L1i L#38 (32KB) + Core L#38
        PU L#76 (P#38)
        PU L#77 (P#110)
      L2 L#39 (256KB) + L1d L#39 (32KB) + L1i L#39 (32KB) + Core L#39
        PU L#78 (P#39)
        PU L#79 (P#111)
      L2 L#40 (256KB) + L1d L#40 (32KB) + L1i L#40 (32KB) + Core L#40
        PU L#80 (P#40)
        PU L#81 (P#112)
      L2 L#41 (256KB) + L1d L#41 (32KB) + L1i L#41 (32KB) + Core L#41
        PU L#82 (P#41)
        PU L#83 (P#113)
      L2 L#42 (256KB) + L1d L#42 (32KB) + L1i L#42 (32KB) + Core L#42
        PU L#84 (P#42)
        PU L#85 (P#114)
      L2 L#43 (256KB) + L1d L#43 (32KB) + L1i L#43 (32KB) + Core L#43
        PU L#86 (P#43)
        PU L#87 (P#115)
      L2 L#44 (256KB) + L1d L#44 (32KB) + L1i L#44 (32KB) + Core L#44
        PU L#88 (P#44)
        PU L#89 (P#116)
      L2 L#45 (256KB) + L1d L#45 (32KB) + L1i L#45 (32KB) + Core L#45
        PU L#90 (P#45)
        PU L#91 (P#117)
      L2 L#46 (256KB) + L1d L#46 (32KB) + L1i L#46 (32KB) + Core L#46
        PU L#92 (P#46)
        PU L#93 (P#118)
      L2 L#47 (256KB) + L1d L#47 (32KB) + L1i L#47 (32KB) + Core L#47
        PU L#94 (P#47)
        PU L#95 (P#119)
      L2 L#48 (256KB) + L1d L#48 (32KB) + L1i L#48 (32KB) + Core L#48
        PU L#96 (P#48)
        PU L#97 (P#120)
      L2 L#49 (256KB) + L1d L#49 (32KB) + L1i L#49 (32KB) + Core L#49
        PU L#98 (P#49)
```

```
      PU L#99 (P#121)
    L2 L#50 (256KB) + L1d L#50 (32KB) + L1i L#50 (32KB) + Core L#50
      PU L#100 (P#50)
      PU L#101 (P#122)
    L2 L#51 (256KB) + L1d L#51 (32KB) + L1i L#51 (32KB) + Core L#51
      PU L#102 (P#51)
      PU L#103 (P#123)
    L2 L#52 (256KB) + L1d L#52 (32KB) + L1i L#52 (32KB) + Core L#52
      PU L#104 (P#52)
      PU L#105 (P#124)
    L2 L#53 (256KB) + L1d L#53 (32KB) + L1i L#53 (32KB) + Core L#53
      PU L#106 (P#53)
      PU L#107 (P#125)
  PCIBridge
    PCI 1000:005d
      Block(Disk) L#9 "sdf"
      Block(Disk) L#10 "sdg"
      Block(Disk) L#11 "sdh"
      Block(Disk) L#12 "sdi"
  NUMANode L#3 (P#3 126GB) + Package L#3 + L3 L#3 (45MB)
    L2 L#54 (256KB) + L1d L#54 (32KB) + L1i L#54 (32KB) + Core L#54
      PU L#108 (P#54)
      PU L#109 (P#126)
    L2 L#55 (256KB) + L1d L#55 (32KB) + L1i L#55 (32KB) + Core L#55
      PU L#110 (P#55)
      PU L#111 (P#127)
    L2 L#56 (256KB) + L1d L#56 (32KB) + L1i L#56 (32KB) + Core L#56
      PU L#112 (P#56)
      PU L#113 (P#128)
    L2 L#57 (256KB) + L1d L#57 (32KB) + L1i L#57 (32KB) + Core L#57
      PU L#114 (P#57)
      PU L#115 (P#129)
    L2 L#58 (256KB) + L1d L#58 (32KB) + L1i L#58 (32KB) + Core L#58
      PU L#116 (P#58)
      PU L#117 (P#130)
    L2 L#59 (256KB) + L1d L#59 (32KB) + L1i L#59 (32KB) + Core L#59
      PU L#118 (P#59)
      PU L#119 (P#131)
    L2 L#60 (256KB) + L1d L#60 (32KB) + L1i L#60 (32KB) + Core L#60
      PU L#120 (P#60)
      PU L#121 (P#132)
    L2 L#61 (256KB) + L1d L#61 (32KB) + L1i L#61 (32KB) + Core L#61
      PU L#122 (P#61)
      PU L#123 (P#133)
    L2 L#62 (256KB) + L1d L#62 (32KB) + L1i L#62 (32KB) + Core L#62
      PU L#124 (P#62)
      PU L#125 (P#134)
```

```
            L2 L#63 (256KB) + L1d L#63 (32KB) + L1i L#63 (32KB) + Core L#63
              PU L#126 (P#63)
              PU L#127 (P#135)
            L2 L#64 (256KB) + L1d L#64 (32KB) + L1i L#64 (32KB) + Core L#64
              PU L#128 (P#64)
              PU L#129 (P#136)
            L2 L#65 (256KB) + L1d L#65 (32KB) + L1i L#65 (32KB) + Core L#65
              PU L#130 (P#65)
              PU L#131 (P#137)
            L2 L#66 (256KB) + L1d L#66 (32KB) + L1i L#66 (32KB) + Core L#66
              PU L#132 (P#66)
              PU L#133 (P#138)
            L2 L#67 (256KB) + L1d L#67 (32KB) + L1i L#67 (32KB) + Core L#67
              PU L#134 (P#67)
              PU L#135 (P#139)
            L2 L#68 (256KB) + L1d L#68 (32KB) + L1i L#68 (32KB) + Core L#68
              PU L#136 (P#68)
              PU L#137 (P#140)
            L2 L#69 (256KB) + L1d L#69 (32KB) + L1i L#69 (32KB) + Core L#69
              PU L#138 (P#69)
              PU L#139 (P#141)
            L2 L#70 (256KB) + L1d L#70 (32KB) + L1i L#70 (32KB) + Core L#70
              PU L#140 (P#70)
              PU L#141 (P#142)
            L2 L#71 (256KB) + L1d L#71 (32KB) + L1i L#71 (32KB) + Core L#71
              PU L#142 (P#71)
              PU L#143 (P#143)
```

7.4 Example Guest VM XML Based on the Example Lenovo x3850 x6 Above

✋ Warning: XML Configuration Example

The XML file below is only an **example** showing the key configurations based on the about command outputs to assist in understanding how to configure the XML. The actual XML configuration must be based on your respective hardware configuration and VM requirements.

Points of interest in this example (refer to the detailed sections of SUSE Best Practices for SAP HANA on KVM - SUSE Linux Enterprise Server for SAP Applications 12 SP2 for a full explanation):

- Memory

    - The Hypervisor has 512 GiB RAM, of which 488 GiB has been allocated as 1 GB Hugepages and therefore 488 GiB is the max VM size in this case

    - 488 GiB = 511705088 KiB

    - In the "numa" section memory is split evenly over the 4 NUMA nodes (CPU sockets)

- CPU Pinning

    - Note the alternating CPU pinning on the Hypervisor, see *Section 4.3, "vCPU and vNUMA Topology"* section for details

    - Note the topology difference between the Guest VM (4x17 CPU cores) the Hypervisor (4x18 CPU cores)

    - Referring to the *Section 7.3, "Example "lstopo-no-graphics" from a Lenovo x3850 x6"* output we know that the disks are attached via CPU sockets 1 and 2, therefore the iothreads are pinned (see vcpupin) to logical CPUs on those sockets

- Storage/IO

    - Storage is configured with virtio and the block devices for SAP HANA are spread evenly across the iothreads

    - Source devices use persistent multipath addresses

    - Type is set to raw, and cache = none

    - See *Section 4.4, "Storage"* section for details

- " rng model = 'virtio' ", for details see section *Section 4.6, "Virtio Random Number Generator (RNG) Device"*

- qemu:commandline elements to describe CPU attributes, for details see section *Section 4.2, "Global vCPU Configuration"*

```
# cat /etc/libvirt/qemu/SUSEKVM.xml
<!--
```

```
WARNING: THIS IS AN AUTO-GENERATED FILE. CHANGES TO IT ARE LIKELY TO BE
OVERWRITTEN AND LOST. Changes to this xml configuration should be made using:
  virsh edit SUSEKVM
or other application using the libvirt API.
-->

<domain type='kvm' xmlns:qemu='http://libvirt.org/schemas/domain/qemu/1.0'>
  <name>SUSEKVM</name>
  <uuid>39112135-9cee-4a5e-b36b-eba8757d666e</uuid>
  <memory unit='KiB'>511705088</memory>
  <currentMemory unit='KiB'>511705088</currentMemory>
  <memoryBacking>
    <hugepages/>
      <page size='1048576' unit='KiB'/>
    <nosharepages/>
  </memoryBacking>
  <vcpu placement='static'>136</vcpu>
  <iothreads>5</iothreads>
  <cputune>
    <vcpupin vcpu='0' cpuset='1'/>
    <vcpupin vcpu='1' cpuset='73'/>
    <vcpupin vcpu='2' cpuset='2'/>
    <vcpupin vcpu='3' cpuset='74'/>
    <vcpupin vcpu='4' cpuset='3'/>
    <vcpupin vcpu='5' cpuset='75'/>
    <vcpupin vcpu='6' cpuset='4'/>
    <vcpupin vcpu='7' cpuset='76'/>
    <vcpupin vcpu='8' cpuset='5'/>
    <vcpupin vcpu='9' cpuset='77'/>
    <vcpupin vcpu='10' cpuset='6'/>
    <vcpupin vcpu='11' cpuset='78'/>
    <vcpupin vcpu='12' cpuset='7'/>
    <vcpupin vcpu='13' cpuset='79'/>
    <vcpupin vcpu='14' cpuset='8'/>
    <vcpupin vcpu='15' cpuset='80'/>
    <vcpupin vcpu='16' cpuset='9'/>
    <vcpupin vcpu='17' cpuset='81'/>
    <vcpupin vcpu='18' cpuset='10'/>
    <vcpupin vcpu='19' cpuset='82'/>
    <vcpupin vcpu='20' cpuset='11'/>
    <vcpupin vcpu='21' cpuset='83'/>
    <vcpupin vcpu='22' cpuset='12'/>
    <vcpupin vcpu='23' cpuset='84'/>
    <vcpupin vcpu='24' cpuset='13'/>
    <vcpupin vcpu='25' cpuset='85'/>
    <vcpupin vcpu='26' cpuset='14'/>
    <vcpupin vcpu='27' cpuset='86'/>
```

```
<vcpupin vcpu='28' cpuset='15'/>
<vcpupin vcpu='29' cpuset='87'/>
<vcpupin vcpu='30' cpuset='16'/>
<vcpupin vcpu='31' cpuset='88'/>
<vcpupin vcpu='32' cpuset='17'/>
<vcpupin vcpu='33' cpuset='89'/>
<vcpupin vcpu='34' cpuset='19'/>
<vcpupin vcpu='35' cpuset='91'/>
<vcpupin vcpu='36' cpuset='20'/>
<vcpupin vcpu='37' cpuset='92'/>
<vcpupin vcpu='38' cpuset='21'/>
<vcpupin vcpu='39' cpuset='93'/>
<vcpupin vcpu='40' cpuset='22'/>
<vcpupin vcpu='41' cpuset='94'/>
<vcpupin vcpu='42' cpuset='23'/>
<vcpupin vcpu='43' cpuset='95'/>
<vcpupin vcpu='44' cpuset='24'/>
<vcpupin vcpu='45' cpuset='96'/>
<vcpupin vcpu='46' cpuset='25'/>
<vcpupin vcpu='47' cpuset='97'/>
<vcpupin vcpu='48' cpuset='26'/>
<vcpupin vcpu='49' cpuset='98'/>
<vcpupin vcpu='50' cpuset='27'/>
<vcpupin vcpu='51' cpuset='99'/>
<vcpupin vcpu='52' cpuset='28'/>
<vcpupin vcpu='53' cpuset='100'/>
<vcpupin vcpu='54' cpuset='29'/>
<vcpupin vcpu='55' cpuset='101'/>
<vcpupin vcpu='56' cpuset='30'/>
<vcpupin vcpu='57' cpuset='102'/>
<vcpupin vcpu='58' cpuset='31'/>
<vcpupin vcpu='59' cpuset='103'/>
<vcpupin vcpu='60' cpuset='32'/>
<vcpupin vcpu='61' cpuset='104'/>
<vcpupin vcpu='62' cpuset='33'/>
<vcpupin vcpu='63' cpuset='105'/>
<vcpupin vcpu='64' cpuset='34'/>
<vcpupin vcpu='65' cpuset='106'/>
<vcpupin vcpu='66' cpuset='35'/>
<vcpupin vcpu='67' cpuset='107'/>
<vcpupin vcpu='68' cpuset='37'/>
<vcpupin vcpu='69' cpuset='109'/>
<vcpupin vcpu='70' cpuset='38'/>
<vcpupin vcpu='71' cpuset='110'/>
<vcpupin vcpu='72' cpuset='39'/>
<vcpupin vcpu='73' cpuset='111'/>
<vcpupin vcpu='74' cpuset='40'/>
```

```
<vcpupin vcpu='75' cpuset='112'/>
<vcpupin vcpu='76' cpuset='41'/>
<vcpupin vcpu='77' cpuset='113'/>
<vcpupin vcpu='78' cpuset='42'/>
<vcpupin vcpu='79' cpuset='114'/>
<vcpupin vcpu='80' cpuset='43'/>
<vcpupin vcpu='81' cpuset='115'/>
<vcpupin vcpu='82' cpuset='44'/>
<vcpupin vcpu='83' cpuset='116'/>
<vcpupin vcpu='84' cpuset='45'/>
<vcpupin vcpu='85' cpuset='117'/>
<vcpupin vcpu='86' cpuset='46'/>
<vcpupin vcpu='87' cpuset='118'/>
<vcpupin vcpu='88' cpuset='47'/>
<vcpupin vcpu='89' cpuset='119'/>
<vcpupin vcpu='90' cpuset='48'/>
<vcpupin vcpu='91' cpuset='120'/>
<vcpupin vcpu='92' cpuset='49'/>
<vcpupin vcpu='93' cpuset='121'/>
<vcpupin vcpu='94' cpuset='50'/>
<vcpupin vcpu='95' cpuset='122'/>
<vcpupin vcpu='96' cpuset='51'/>
<vcpupin vcpu='97' cpuset='123'/>
<vcpupin vcpu='98' cpuset='52'/>
<vcpupin vcpu='99' cpuset='124'/>
<vcpupin vcpu='100' cpuset='53'/>
<vcpupin vcpu='101' cpuset='125'/>
<vcpupin vcpu='102' cpuset='55'/>
<vcpupin vcpu='103' cpuset='127'/>
<vcpupin vcpu='104' cpuset='56'/>
<vcpupin vcpu='105' cpuset='128'/>
<vcpupin vcpu='106' cpuset='57'/>
<vcpupin vcpu='107' cpuset='129'/>
<vcpupin vcpu='108' cpuset='58'/>
<vcpupin vcpu='109' cpuset='130'/>
<vcpupin vcpu='110' cpuset='59'/>
<vcpupin vcpu='111' cpuset='131'/>
<vcpupin vcpu='112' cpuset='60'/>
<vcpupin vcpu='113' cpuset='132'/>
<vcpupin vcpu='114' cpuset='61'/>
<vcpupin vcpu='115' cpuset='133'/>
<vcpupin vcpu='116' cpuset='62'/>
<vcpupin vcpu='117' cpuset='134'/>
<vcpupin vcpu='118' cpuset='63'/>
<vcpupin vcpu='119' cpuset='135'/>
<vcpupin vcpu='120' cpuset='64'/>
<vcpupin vcpu='121' cpuset='136'/>
```

```
    <vcpupin vcpu='122' cpuset='65'/>
    <vcpupin vcpu='123' cpuset='137'/>
    <vcpupin vcpu='124' cpuset='66'/>
    <vcpupin vcpu='125' cpuset='138'/>
    <vcpupin vcpu='126' cpuset='67'/>
    <vcpupin vcpu='127' cpuset='139'/>
    <vcpupin vcpu='128' cpuset='68'/>
    <vcpupin vcpu='129' cpuset='140'/>
    <vcpupin vcpu='130' cpuset='69'/>
    <vcpupin vcpu='131' cpuset='141'/>
    <vcpupin vcpu='132' cpuset='70'/>
    <vcpupin vcpu='133' cpuset='142'/>
    <vcpupin vcpu='134' cpuset='71'/>
    <vcpupin vcpu='135' cpuset='143'/>
    <emulatorpin cpuset='0,54'/>
    <iothreadpin iothread='1' cpuset='72'/>
    <iothreadpin iothread='2' cpuset='18'/>
    <iothreadpin iothread='3' cpuset='36'/>
    <iothreadpin iothread='4' cpuset='90'/>
    <iothreadpin iothread='5' cpuset='108'/>
  </cputune>
  <numatune>
    <memory mode='strict' nodeset='0-3'/>
    <memnode cellid='0' mode='strict' nodeset='0'/>
    <memnode cellid='1' mode='strict' nodeset='1'/>
    <memnode cellid='2' mode='strict' nodeset='2'/>
    <memnode cellid='3' mode='strict' nodeset='3'/>
  </numatune>
  <os>
    <type arch='x86_64' machine='pc-i440fx-2.6'>hvm</type>
    <boot dev='hd'/>
  </os>
  <features>
    <acpi/>
    <apic/>
    <vmport state='off'/>
  </features>
  <cpu mode='host-passthrough'>
    <topology sockets='4' cores='17' threads='2'/>
    <numa>
      <cell id='0' cpus='0-33' memory='127926272' unit='KiB'/>
      <cell id='1' cpus='34-66' memory='127926272' unit='KiB'/>
      <cell id='2' cpus='67-101' memory='127926272' unit='KiB'/>
      <cell id='3' cpus='102-135' memory='127926272' unit='KiB'/>
    </numa>
  </cpu>
  <clock offset='utc'>
```

```
    <timer name='rtc' tickpolicy='catchup'/>
    <timer name='pit' tickpolicy='delay'/>
    <timer name='hpet' present='no'/>
  </clock>
  <on_poweroff>destroy</on_poweroff>
  <on_reboot>restart</on_reboot>
  <on_crash>restart</on_crash>
  <pm>
    <suspend-to-mem enabled='no'/>
    <suspend-to-disk enabled='no'/>
  </pm>
  <devices>
    <emulator>/usr/bin/qemu-kvm</emulator>
...
    <disk type='block' device='disk'>
      <driver name='qemu' type='raw' cache='none' io='threads' iothread='1'/>
      <source dev='/dev/disk/by-id/dm-uuid-mpath-xxxxx...'/>
      <target dev='vda' bus='virtio'/>
    </disk>
    <disk type='block' device='disk'>
      <driver name='qemu' type='raw' cache='none' io='threads' iothread='2'/>
      <source dev='/dev/disk/by-id/dm-uuid-mpath-xxxxx-cd5e'/>
      <target dev='vdf' bus='virtio'/>
    </disk>
    <disk type='block' device='disk'>
      <driver name='qemu' type='raw' cache='none' io='threads' iothread='3'/>
      <source dev='/dev/disk/by-id/dm-uuid-mpath-xxxxx-cd89'/>
      <target dev='vdg' bus='virtio'/>
    </disk>
    <disk type='block' device='disk'>
      <driver name='qemu' type='raw' cache='none' io='threads' iothread='4'/>
      <source dev='/dev/disk/by-id/dm-uuid-mpath-xxxxx-c9bb'/>
      <target dev='vdh' bus='virtio'/>
    </disk>
    <disk type='block' device='disk'>
      <driver name='qemu' type='raw' cache='none' io='threads' iothread='5'/>
      <source dev='/dev/disk/by-id/dm-uuid-mpath-xxxxx-c9e5'/>
      <target dev='vdi' bus='virtio'/>
    </disk>

    <hostdev mode='subsystem' type='pci' managed='yes'>
      <source>
        <address domain='0x0003' bus='0x03' slot='0x00' function='0x0'/>
      </source>
    </hostdev>
...
    <memballoon model='virtio'>
```

```
      <address type='pci' domain='0x0000' bus='0x00' slot='0x07' function='0x0'/>
    </memballoon>
...
    <rng model='virtio'>
      <backend model='random'>/dev/random</backend>
    </rng>
  </devices>
  <qemu:commandline>
    <qemu:arg value='-cpu'/>
    <qemu:arg value='host,migratable=off,+invtsc,l3-cache=on'/>
  </qemu:commandline>
</domain>
```

8 Additional Information

8.1 Resources

- SUSE Best Practices (https://documentation.suse.com/sbp/all/ ↗)

- SUSE Virtualization Guide for SUSE Linux Enterprise Server 12 (https://documenta-tion.suse.com/sles/12-SP4/html/SLES-all/book-virt.html ↗)

8.2 Feedback

Several feedback channels are available:

Bugs and Enhancement Requests

For services and support options available for your product, refer to http://www.suse.com/support/ ↗.

To report bugs for a product component, go to https://scc.suse.com/support/ ↗ requests, log in, and select Submit New SR (Service Request).

Report Documentation Bug

To report errors or suggest enhancements for a certain document, use the `Report Doc-umentation Bug` feature at the right side of each section in the online documentation. Provide a concise description of the problem and refer to the respective section number and page (or URL).

SUSE Best Practices for SAP HANA on KVM

Mail

For feedback on the documentation of this product, you can also send a mail to `doc-team@suse.com`. Make sure to include the document title, the product version and the publication date of the documentation.

8.3   Version History

Version	Publication Date	Author	Comment
0.1	Oct 2017	Lee Martin	Initial version
0.2	Dec 2017	Lee Martin	Pilot Customers
0.3	Jan 2018	Lee Martin	Add storage section
0.4	Feb 2018	Lee Martin	Add sizing section
1.0	Feb 2018	Lee Martin	SAP GA Release for Haswell Single-VM

SUSE Best Practices for SAP HANA on KVM

9 Legal notice

GNU Free Documentation License

Copyright (C) 2000, 2001, 2002 Free Software Foundation, Inc. 51 Franklin St, Fifth Floor, Boston, MA 02110-1301 USA. Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or non-commercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/ ↗.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

ADDENDUM: How to use this License for your documents

```
Copyright (c) YEAR YOUR NAME.
Permission is granted to copy, distribute and/or modify this document
under the terms of the GNU Free Documentation License, Version 1.2
or any later version published by the Free Software Foundation;
with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.
A copy of the license is included in the section entitled "GNU
 Free Documentation License".
```

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts". line with this:

```
with the Invariant Sections being LIST THEIR TITLES, with the
Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.
```

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.