SUSE Best Practices for SAP HANA on KVM

SUSE Linux Enterprise Server for SAP Applications 15 SP2

Gereon Vey, SAP Solution Architect (SUSE)

Dario Faggioli, Software Engineer Virtualization Specialist (SUSE)

SUSE Best Practices for SAP HANA on KVM

Date: 2024-12-12

SUSE® Linux Enterprise Server for SAP Applications is optimized in various ways for SAP* applications. This best practice document describes how SUSE Linux Enterprise Server for SAP Applications 15 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.

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 15 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, "Setting up and configuring the hypervisor"* - The host operating system running the hypervisor directly on the server hardware

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

- *Section 5, "Installing the 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

Virtual Machine

is an emulation of a computer.

Hypervisor

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

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.

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 section "Introduction to KVM Virtualization" (https://documentation.suse.com/sles/15-SP2/html/SLES-all/cha-kvm-intro.html) of the Virtualization Guide.

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. No other VM or workload is allowed on the same server.

Multi-VM

Multiple VM's per hypervisor/physical server for SAP HANA Scale-Up.

Scale-Out

For an SAP HANA Scale-Out deployment, distributed over multiple VMs on multiple hosts.

2 Supported scenarios and prerequisites

Follow the **SUSE Best Practices for SAP HANA on KVM - SUSE Linux Enterprise Server for SAP Applications 15 SP2** document at hand 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 (mailto:s-aphana@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
1st Generation Intel Xeon Scalable Processor (Skylake)	*Hypervisor:* SLES for SAP 15 SP2 *Guest:* SLES for SAP 15 SP2 onwards *Size:* max. 4 sockets [a], 3 TiB RAM	no	no

a Maximum 4 sockets using Intel standard chipsets on a single system board, for example Lenovo* x3850, Fujitsu* rx4770 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 3120786 - "SAP HANA on SUSE KVM Virtualization with SLES 15 SP2" (https://launchpad.support.sap.com/#/notes/3120786) ⬈

2.2 Sizing

When sizing for a virtualized SAP HANA system, some additional factors need to be taken into account.

2.2.1 Resources for the hypervisor

It is recommended to reserve a minimum of about 8% of the hosts's main memory for the hypervisor.

The hypervisor will consume CPU capacity, approximately 5% to 10% of the SAPS capacity, depending on the workload characteristics:

- 5% of the SAPS capacity for mainly analytical workloads

- 10% of the SAPS capacity for mainly transactional workloads

It is however **not** required to dedicate CPUs to the hypervisor.

2.2.2 Memory sizing

Since SAP HANA runs inside the VM, it is the RAM size of the VM which needs to satisfy the memory requirements from the SAP HANA Memory sizing.

The memory used by the VM must be smaller than the physical memory of the machine. It is recommended to reserve at least 8% of the total memory reported by "/proc/meminfo" (in the "MemTotal" field) for the hypervisor. This leaves ~ 92% to the VM.

See *Section 4.3, "Backing memory"* for more details.

2.2.3 CPU sizing

Some artificial workload tests on 1st Generation Intel Xeon Scalable Processor (Skylake) CPUs have shown an approximately of up to 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 memory-to-core ratios for SAP HANA as defined by SAP

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

Both ways are described in the following sections.

2.2.3.1 Following the fixed memory-to-core ratios for SAP HANA

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 **memory-to-core 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 memory-to-core 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 CPU Architecture "Intel Skylake SP" for Appliance Type "Scale-up: BWoH").

- Look for the largest certified RAM size for the number of CPU Sockets on the server (for example 3 TiB/3072 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 28).

- 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 4x28 = 112).

- Now divide the Appliance RAM by the total number of cores (not hyperthreads) to give you the **memory-to-core** ratio (for example 3072 GiB/112 = approx. 28 GiB per core).

Table 2, "SAP HANA memory-to-core ratio examples" below has some current examples of SAP HANA memory-to-core ratios.

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

CPU Architecture	Appliance Type	Max Memory Size	Sockets	Cores per Socket	SAP HANA memory-to-core ratio
1st Generation Intel Xeon Scalable Processor (Skylake)	OLTP	6 TiB / 6144 GiB	4	28	54.86 GiB/core

CPU Architecture	Appliance Type	Max Memory Size	Sockets	Cores per Socket	SAP HANA memory-to-core ratio
1st Generation Intel Xeon Scalable Processor (Skylake)	OLAP	3 TiB / 3072 GiB	4	28	27.43 GiB/core

From your memory requirement, calculate the RAM size the VM needs to be compliant with the appropriate memory-to-core ratio defined by SAP.

- To get the memory per socket, multiply the memory-to-core ratio by the number of cores (not threads) of a single socket in your host

- Divide the memory requirement by the memory per socket, and round the result up to the next full number, and multiply that number by the memory per socket again

EXAMPLE 1: CALCULATION EXAMPLE

- From an S/4HANA sizing you get a memory requirement for SAP HANA of 2000 GiB.

- Your CPUs have 28 cores per socket. The memory per socket is `28 cores * 54.86 GiB/core = 1536 GiB`.

- Divide your memory requirement `2000 GiB / 1536 GiB = 1.2987` and round this result up to 2. Then multiply `2 * 1536 GiB = 3072 GiB`

- 3072 GiB is now the memory size to use in the VM configuration as described in *Section 4.3, "Backing memory"*

2.2.3.2   Following the SAP HANA TDI "Phase 5" rules

- SAP HANA TDI "Phase 5" rules allow customers to deviate from the above described SAP HANA memory-to-core sizing ratios in certain scenarios. The KVM implementation however must still adhere to the **SUSE Best Practices for SAP HANA on KVM - SUSE Linux Enterprise Server for SAP Applications 15 SP2** document at hand. Details on SAP

HANA TDI Phase 5 can be found in the following blog "TDI Phase 5: New Opportunities for Cost Optimization of SAP HANA Hardware" (https://blogs.sap.com/2017/09/20/tdi-phase-5-new-opportunities-for-cost-optimization-of-sap-hana-hardware/) ↗ from SAP.

- 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 HANA Master Guide: Sizing SAP HANA" (https://help.sap.com/viewer/eb3777d5495d46c5b2fa773206bbfb46/2.0.03/en-US/d4a122a7bb57101493e3f5ca08e6b039.html) ↗

- "General SAP Sizing information" (http://sap.com/sizing) ↗

2.3   Configuring the KVM hypervisor version

The hypervisor must be configured according to the **SUSE Best Practices for SAP HANA on KVM - SUSE Linux Enterprise Server for SAP Applications 15 SP2** guide at hand and fulfill the following minimal requirements:

- SUSE Linux Enterprise Server for SAP Applications 15 SP2 ("Unlimited Virtual Machines" subscription)

  - kernel (Only major version 5.3, minimum package version 5.3.18-24.24.1)

  - libvirt (Only major version 6.0, minimum package version 6.0.0-13.3.1)

  - qemu (Only major version 4.2, minimum package version 4.2.1-11.10.1)

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/ ↗.

2.5 Guest VM

The guest VM must:

- run SUSE Linux Enterprise Server for SAP Applications 15 SP2 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/15-SP2/html/SLES-all/cha-virt-support.html#virt-support-guests) ↗.

- comply with KVM limits as per "SUSE Linux Enterprise Server 15 SP2 release notes (https://www.suse.com/releasenotes/x86_64/SUSE-SLES/15-SP2/#allArch-virtualization-kvm-limits) ↗".

- fulfill the SAP HANA Hardware and Cloud Measurent Tools (HCMT) storage KPI's as per SAP Note 2493172 "SAP HANA Hardware and Cloud Measurement Tools" (https://launchpad.support.sap.com/#/notes/2493172) ↗. Refer to *Section 4.6, "Configuring storage"* for storage configuration details.

- be configured according to the **SUSE Best Practices for SAP HANA on KVM - SUSE Linux Enterprise Server for SAP Applications 15 SP2** document at hand.

3 Setting up and configuring the hypervisor

The following sections describe how to set up and configure the hypervisor for a virtualized SAP HANA scenario.

3.1 Installing the KVM hypervisor

For details refer to section 6.4 "Installation of Virtualization Components" of the SUSE Virtualization Guide (https://documentation.suse.com/sles/15-SP2/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 Configuring networking on the hypervisor

To achieve maximum performance required for productive SAP HANA workloads, one of the host networking devices must be assigned directly to the KVM guest VM. A Network Interface Card (NIC) including support for the technology that goes under the name of Single Root I/O Virtualization (SR-IOV) is required. This guarantees that the overhead in which we would have incurred if using IO Virtualization is avoided.

To check whether such technology is available, assuming that `17:00.0` is the address of the NIC on the PCI bus (as visible in the output of the `lspci` tool), the following command can be issued:

```
lspci -vs 17:00.0
17:00.0 Ethernet controller: Intel Corporation Ethernet Controller X710 for 10GbE SFP+
 (rev 01)
        Subsystem: Intel Corporation Ethernet Converged Network Adapter X710-2
        Flags: bus master, fast devsel, latency 0, IRQ 247, NUMA node 0
        Memory at 9c000000 (64-bit, prefetchable) [size=8M]
        Memory at 9d008000 (64-bit, prefetchable) [size=32K]
        Expansion ROM at 9d680000 [disabled] [size=512K]
        Capabilities: [40] Power Management version 3
        Capabilities: [50] MSI: Enable- Count=1/1 Maskable+ 64bit+
        Capabilities: [70] MSI-X: Enable+ Count=129 Masked-
        Capabilities: [a0] Express Endpoint, MSI 00
        Capabilities: [e0] Vital Product Data
        Capabilities: [100] Advanced Error Reporting
        Capabilities: [140] Device Serial Number d8-ef-c3-ff-ff-fe-fd-3c
        Capabilities: [150] Alternative Routing-ID Interpretation (ARI)
        Capabilities: [160] Single Root I/O Virtualization (SR-IOV)
        Capabilities: [1a0] Transaction Processing Hints
        Capabilities: [1b0] Access Control Services
        Capabilities: [1d0] #19
        Kernel driver in use: i40e
        Kernel modules: i40e
```

The output should contain a line similar to the following: `Single Root I/O Virtualization (SR-IOV)`. If such line is not present, it might be the case that SR-IOV needs to be explicitly enabled in the BIOS.

3.2.1 Preparing a Virtual Function (VF) for a guest VM

After checking that the NIC is SR-IOV capable, the host and the guest VM should be configured to use one of the available Virtual Functions (VFs) as (one of) the guest VM's network device(s). More information about SR-IOV as a technology and how to properly configure everything that

is necessary for it to work well in the general case can be found in the SUSE Virtualization Guide for SUSE Linux Enterprise Server 15 SP2 (https://documentation.suse.com/sles/15-SP2/single-html/SLES-virtualization ↗), and specifically in section "Adding SR-IOV Devices" (https://documentation.suse.com/sles/15-SP2/single-html/SLES-virtualization/#sec-libvirt-config-io ↗).

Enabling PCI passthrough for the host kernel

Make sure that the host kernel boot command line contains these two parameters: `intel_iommu=on iommu=pt`. This is done by editing */etc/default/grub*:

- Append `intel_iommu=on iommu=pt` to the string that is assigned to the variable `GRUB_CMDLINE_LINUX_DEFAULT`.

- Then run `update-bootloader` (more detailed information is provided later in the document).

Loading and configuring SR-IOV host drivers

Before starting the VM, SR-IOV must be enabled on the desired NIC, and the VFs must be created.

Always make sure that the proper SR-IOV-capable driver is loaded. For example, for an **Intel Corporation Ethernet Controller X710** NIC, the driver resides in the `i40e` kernel module. It can be loaded with the `modprobe` command, but chances are high that it is aldready loaded by default.

If the SR-IOV-capable module is not in use by default and it also fails to load with `modprobe`, this might mean that another driver, potentially one that is not SR-IOV-capable, is the one that is currently loaded. In which case, it should be removed with the `rmmod` command.

When the proper module is loaded, creating at least one VF happens with the following command (which creates four of them):

```
echo 4 > /sys/bus/pci/devices/0000\:17\:00.0/sriov_numvfs
```

Or, assuming that the designated NIC corresponds to the symbolic name of `eth10`, use the following command:

```
echo 4 > /sys/class/net/eth10/device/sriov_numvfs
```

The procedure can be automated to run at boot time: Create the following `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
```

After that, create the script */etc/init.d/after.local*:

```
#! /bin/sh
#
# Copyright (c) 2010 SuSE LINUX Products GmbH, Germany.  All rights reserved.
# ...
echo 4 > /sys/class/net/eth10/device/sriov_numvfs
```

3.3   Configuring storage on the hypervisor

As with compute resources, the storage used for running SAP HANA must also be SAP cer-
tified. Therefore only the storage from SAP HANA Appliances or SAP HANA Certified Enter-
prise Storage (https://www.sap.com/dmc/exp/2014-09-02-hana-hardware/enEN/#/solutions?filter-
s=v:deCertified;storage ↗) is supported. In all cases the SAP HANA storage configuration recom-
mendations from the respective hardware vendor and the SAP HANA Storage Requirements
for TDI (https://archive.sap.com/kmuuid2/70c8e423-c8aa-3210-3fae-e043f5c1ca92/SAP%20HANA
%20TDI%20-%20Storage%20Requirements.pdf ↗) should be followed.

There are two supported storage options to use for the SAP HANA database: Fibre Channel (FC)
storage and Network Attached Storage (NAS).

3.3.1   Network attached Storage

The SAP HANA storage is attached via the NFSv4 protocol. In this case, nothing needs to be
configured on the hypervisor. Do make sure though that the VM has access to one or more
dedicated 10 Gbit Ethernet interfaces for the network traffic to the network-attached storage.

3.3.2 Fibre Channel storage

As described in *Section 3.2, "Configuring networking on the hypervisor"*, to reach the adequate level of performance, the storage drives for actual SAP HANA data are attached to the guest VM via directly assigning the SAN HBA controller to it. One difference, though, is that there is no counterpart of SR-IOV commonly available for storage controllers. Therefore, a full SAN HBA controller must be dedicated and directly assigned to the guest VM.

To figure out which SAN HBA should be used, check the available ones, for example with the `lspci` command:

```
lspci | grep -i "Fibre Channel"
85:00.0 Fibre Channel: QLogic Corp. ISP2722-based 16/32Gb Fibre Channel to PCIe Adapter
 (rev 01)
85:00.1 Fibre Channel: QLogic Corp. ISP2722-based 16/32Gb Fibre Channel to PCIe Adapter
 (rev 01)
ad:00.0 Fibre Channel: QLogic Corp. ISP2722-based 16/32Gb Fibre Channel to PCIe Adapter
 (rev 01)
ad:00.1 Fibre Channel: QLogic Corp. ISP2722-based 16/32Gb Fibre Channel to PCIe Adapter
 (rev 01)
```

The HBAs that are assigned to the guest VM must not be in use on the host.

The remaining storage configuration details, such as how to add the disks and the HBA controllers to the guest VM configuration file, and what to do with them from inside the guest VM itself, are available in *Section 4.6, "Configuring storage"*.

3.4 Configuring the hypervisor operating system

The hypervisor host operating system needs to be configured to assure compatibility and maxized performance for an SAP HANA VM.

3.4.1 Installing `vhostmd`

The hypervisor needs to have the `vhostmd` package installed and the corresponding `vhostmd` service enabled and started. This is described in SAP Note 1522993 - "Linux: SAP on SUSE KVM - Kernel-based Virtual Machine" (https://launchpad.support.sap.com/#/notes/1522993) ↗.

3.4.2 Tuning the generic host with `tuned`

To apply some less specific, but nevertheless effective, tuning to the host, the **TuneD** tool (https://tuned-project.org/↗) can be used.

When installed (the package name is `tuned`), one of the preconfigured profiles can be selected, or a custom one created. Specifically, the `virtual-host` profile should be chosen. Do not use the `sap-hana profile` on the hypervisor. This can be achieved with the following commands:

```
zypper in tuned

systemctl enable tuned

systemctl start tuned

tuned-adm profile virtual-host
```

The `tuned` daemon should now start automatically at boot time, and it should always load the `virtual-host` profile, so there is no need to add any of the above commands in any custom start-up script. If in doubt, it is possible to check with the following command whether `tuned` is running and what the current profile is :

```
tuned-adm profile

Available profiles:
- balanced                      - General non-specialized tuned profile
...
- virtual-guest                 - Optimize for running inside a virtual guest
- virtual-host                  - Optimize for running KVM guests
Current active profile: virtual-host
```

3.4.2.1 Power management considerations

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 selected `tuned` profile should have done this already, and with the following command, it is possible to verify that it actually did:

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

The governor setting can be verified by looking at the **current policy**.

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
```

Modern processors also attempt to save power when they are idle, by switching to a lower power state. Unfortunately this incurs latency when switching in and out of these states.

To avoid that, and achieve better and more consistent the performance, the CPUs should not be allowed to go into too aggressive power saving modes (known as C-states). It therefore is recommended that only C0 and C1 are used.

This can be enforced by adding the following parameters to the kernel boot command line: `intel_idle.max_cstate=1`.

To double check that only the desired C-states are actually available, the following command can be used:

```
cpupower idle-info
```

The idle state settings can be verified by looking at the line containing`Available idle states:`.

3.4.3 `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
```

3.4.4 Kernel Samepage Merging (ksm)

Kernel Samepage Merging (KSM, https://www.kernel.org/doc/html/latest/admin-guide/mm/ ksm.html ↗) is of no use, because there is only one single VM. Thus it should be disabled. The following command makes sure that it is tuned off and that any sharing and de-duplication activity that may have happened, in case it was enabled, is reverted:

```
echo 2 >  /sys/kernel/mm/ksm/run
```

3.4.5 Customizing the Linux kernel boot options

To edit the boot options for the Linux kernel, perform the following steps:

1. Edit */etc/defaults/grub* and add the following boot options to the line **GRUB_CMDLINE_LINUX_DEFAULT** (a detailed explanation of these options will follow).

```
mitigations=auto kvm.nx_huge_pages=off numa_balancing=disable kvm_intel.ple_gap=0
 transparent_hugepage=never intel_idle.max_cstate=1 default_hugepagesz=1GB
 hugepagesz=1GB hugepages=<number of hugepages> intel_iommu=on iommu=pt
 intremap=no_x2apic_optout
```

2. Run the following command:

```
update-bootloader
```

3. Reboot the system:

```
reboot
```

3.4.6 Technical explanation of the above described configuration settings

Hardware vulnerabilities mitigations (mitigations = auto kvm.nx_huge_pages = off)

Recently, a class of side channel attacks exploiting the branch prediction and the speculative execution capabilities of modern CPUs appeared. On an affected CPU, these problems cannot be fixed, but their effect and their actual exploitability can be mitigated in software. However, this sometimes has a non-negligible impact on the performance.

For achieving the best possible security, the software mitigations for these vulnerabilities are being enabled (`mitigations=auto`) with the only exception of the one that deals with "Machine Check Error Avoidance on Page Size Change (CVE-2018-12207, also known as "iTLB Multiht").

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 the system is not overcommitted.

Transparent huge pages (transparent_hugepage = never)

Because 1 GiB 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.

Processor C-states (intel_idle.max_cstate = 1)

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.

Huge pages (default_hugepagesz = 1 GiB hugepagesz = <1 GiB hugepages = number of hugepages>)

The use of 1 GiB huge pages is to reduce overhead and contention when the guest is updating its page tables. This requires allocation of 1 GiB huge pages 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. Thus 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 huge page options. Specifically the `<number of hugepages>` option must be placed **after** the 1 GiB huge page size definitions.

> ✎ Note: Calculating value
>
> The value for `<number of hugepages>` should be calculated by taking the number GiB's of RAM minus approx. 8% for the hypervisor OS. For example, 3 TiB RAM (3072 GiB) minus 8% are approximately 2770 huge pages.

PCI Passthrough (intel_iommu = on iommu = pt)

For being able to directly assign host devices (like storage controllers and NIC Virtual Functions), with PCI Passthrough and SR-IOV, the IOMMU must be enabled. On top of that, `iommu=pt` makes sure that you set up the devices for the best performance (that is, passthrough mode).

Interrupt remapping (intremap = no_x2apic_optout)

Interrupt remapping interrupts from devices to be intercepted, validated and routed to a specific CPU (for example, one where a virtual CPU of the guest VM that has the device assigned is running). This parameter makes sure that such feature is always enabled.

4 Configuring the guest VM

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 Creating an initial guest VM XML

Refer to section 9 "Guest Installation" of the SUSE Virtualization Guide (https://documentation.suse.com/sles/15-SP2/html/SLES-all/cha-kvm-inst.html ⬈).

4.2 Configuring global vCPU

The virtual CPU configuration of the VM guest should reflect the host CPU configuration as close as possible. There cannot be any overcommitting of memory or CPU resources.

The CPU model should be set to `host-passthrough`, and any `check` should be disabled. In addition, the `rdtscp`, `invtsc` and `x2apic` features are required.

4.3 Backing memory

Huge pages, sized 1 GiB (that is, 1048576 KiB), must be used for all the guest VM memory. This guarantees optimal performance for the guest VM.

It is necessary that each NUMA cell of the guest VM have a whole number of huge pages assigned to them (that is, no fractions of huge pages). All the NUMA cells should also have the same number of huge pages assigned to them (that is, the guest VM memory configuration must be balanced).

Therefore the number of huge pages needs to be dividable by the number of NUMA cells.

For example, if the host has 3169956100 KiB (that is, 3 TiB) of memory and we want to leave 91.75% of it to the hypervisor (see *Section 2.2.2, "Memory sizing"*), and there are 4 NUMA cells, each NUMA cell will have the following number of huge pages:

- (3169956100 * (91.75/100)) / 1048576 / 4 = 693

This means that, in total, there will need to be the following number of huge pages:

- 693 * 4 = 2772

Such number must be passed to the host kernel command line parameter on boot (that is `hugepages=2772`, see *Section 3.4.6, "Technical explanation of the above described configuration settings"*).

Both the total amount of memory the guest VM should use and the fact that such memory must come from 1 GiB huge pages need to be specified in the guest VM configuration file.

It must also be ensured that the `memory` and the `currentMemory` element have the same value, to disable memory ballooning, which, if enabled, would cause unacceptable latency:

```
<domain type='kvm'>
  <!-- ... -->
  <memory unit='KiB'>2906652672</memory>
  <currentMemory unit='KiB'>2906652672</currentMemory>
  <memoryBacking>
    <hugepages>
      <page size='1048576' unit='KiB'/>
    </hugepages>
    <nosharepages/>
  </memoryBacking>
  <!-- ... -->
</domain>
```

Note: Memory Unit

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

4.4 Mapping vCPU and vNUMA topology and pinning

It is important to map the host topology into the guest VM, as described below. This allows HANA to spread its own workload threads across many virtual CPUs and NUMA nodes.

For example, for a 4-socket system, with 28 cores per socket and hyperthreading enabled, the virtual CPU configuration will also have 4 sockets, 28 cores, 2 threads.

Always make sure that, in the guest VM configuration file:

- the `cpu mode` attribute is set to `host-passthrough`.

- the `cpu topology` attribute describes the vCPU NUMA topology of the guest, as discussed above.

- the attributes of the `numa` elements describe which vCPU number ranges belong to which NUMA cell. Care should be taken since these number ranges are not the same as on the host. Additionally:

  - the `cell` elements 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 4.3, "Backing memory"* and *Section 2.2.2, "Memory sizing"* of this paper for further details.

  - each NUMA cell of the guest VM has 56 vCPUs.

  - the distances between the cells are identical to those of the physical hardware (as per the output of the command`numactl --hardware`).

```
<domain type='kvm'>
  <!-- ... -->
  <cpu mode='host-passthrough' check='none'>
    <topology sockets='4' cores='28' threads='2'/>
    <feature policy='require' name='rdtscp'/>
    <feature policy='require' name='invtsc'/>
    <feature policy='require' name='x2apic'/>
    <numa>
      <cell id='0' cpus='0-55' memory='726663168' unit='KiB'>
        <distances>
          <sibling id='0' value='10'/>
          <sibling id='1' value='21'/>
          <sibling id='2' value='21'/>
          <sibling id='3' value='21'/>
        </distances>
      </cell>
      <cell id='1' cpus='56-111' memory='726663168' unit='KiB'>
        <distances>
          <sibling id='0' value='21'/>
          <sibling id='1' value='10'/>
          <sibling id='2' value='21'/>
```

```
          <sibling id='3' value='21'/>
        </distances>
      </cell>
      <cell id='2' cpus='112-167' memory='726663168' unit='KiB'>
        <distances>
          <sibling id='0' value='21'/>
          <sibling id='1' value='21'/>
          <sibling id='2' value='10'/>
          <sibling id='3' value='21'/>
        </distances>
      </cell>
      <cell id='3' cpus='168-223' memory='726663168' unit='KiB'>
        <distances>
          <sibling id='0' value='21'/>
          <sibling id='1' value='21'/>
          <sibling id='2' value='21'/>
          <sibling id='3' value='10'/>
        </distances>
      </cell>
    </numa>
  </cpu>
  <!-- ... -->
</domain>
```

It is also necessary to pin virtual CPUs to physical CPUs, to limit the overhead caused by virtual CPUs being moved around physical CPUs by the host scheduler. Similarly, the memory for each NUMA cell of the guest VM must be allocated only on the corresponding host NUMA node.

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 112 on the host, you will need to pin that sibling pair to vCPU 0 and vCPU 1.

It is recommended to pin both the various sibling pairs of vCPUs to (the corresponding) sibling pairs of host CPUs. For example, vCPU 0 should be pinned to pCPU 0 and 112, and the same applies to vCPU 1. As far as both the vCPUs always run on the same physical core, the host scheduler is allowed to execute them on either thread, for example in case only one is free while the other is busy executing host or hypervisor activities.

Using the above information, the CPU and memory pinning section of the guest VM XML can be created. Below find a practical example based on the hypothetical example above.

SUSE Best Practices for SAP HANA on KVM

Make sure to take note of the following configuration components:

- The `vcpu placement` element lists the total number of vCPUs in the guest.

- The `cputune` element contains the attributes describing the mappings of vCPUs 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 the values are `nodeset=0-3` and `cellid` 0 to 3.

```
<domain type='kvm'>
  <vcpu placement='static'>224</vcpu>
  <cputune>
    <vcpupin vcpu='0' cpuset='0,112'/>
    <vcpupin vcpu='1' cpuset='0,112'/>
    <vcpupin vcpu='2' cpuset='1,113'/>
    <vcpupin vcpu='3' cpuset='1,113'/>
    <vcpupin vcpu='4' cpuset='2,114'/>
    <vcpupin vcpu='5' cpuset='2,114'/>
    <vcpupin vcpu='6' cpuset='3,115'/>
    <vcpupin vcpu='7' cpuset='3,115'/>
    <vcpupin vcpu='8' cpuset='4,116'/>
    <vcpupin vcpu='9' cpuset='4,116'/>
    <vcpupin vcpu='10' cpuset='5,117'/>
    <vcpupin vcpu='11' cpuset='5,117'/>
    <!-- output abbreviated -->
    <vcpupin vcpu='218' cpuset='109,221'/>
    <vcpupin vcpu='219' cpuset='109,221'/>
    <vcpupin vcpu='220' cpuset='110,222'/>
    <vcpupin vcpu='221' cpuset='110,222'/>
    <vcpupin vcpu='222' cpuset='111,223'/>
    <vcpupin vcpu='223' cpuset='111,223'/>
  </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>
```

```
    <!-- ... -->
</domain>
```

The following script generates a section of the domain configuration according to the described specifications:

```
#!/usr/bin/env bash
NUM_VCPU=$(ls -d /sys/devices/system/cpu/cpu[0-9]* | wc -l)
echo "  <vcpu placement='static'>${NUM_VCPU}</vcpu>"
echo "  <cputune>"
THREAD_PAIRS="$(cat /sys/devices/system/cpu/cpu*/topology/core_cpus_list | sort -n |
 uniq )"
VCPU=0
for THREAD_PAIR in ${THREAD_PAIRS}; do
  for i in 1 2; do
    echo "    <vcpupin vcpu='${VCPU}' cpuset='${THREAD_PAIR}'/>"
    VCPU=$(( VCPU + 1 ))
  done
done
echo "  </cputune>"
```

The following commands can be used to determine the CPU details on the hypervisor host:

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

lstopo-no-graphics
```

It is not necessary to isolate the guest VM's `iothreads`, nor to statically reserve any host CPU to either them or any other kind of host activity.

4.5   Configuring networking

One of the Virtual Functions prepared in *Section 3.2, "Configuring networking on the hypervisor"* must be added to the guest VM as (one of) its network adapter(s). This can be done by putting the following details in the guest VM configuration file:

```
 <domain type='kvm'>
  <!-- ... -->
  <devices>
    <!-- ... -->
    <interface type='hostdev' managed='yes'>
      <mac address='52:54:00:7f:12:fb'/>
      <driver name='vfio'/>
      <source>
```

```
        <address type='pci' domain='0x0000' bus='0x17' slot='0x02' function='0x0'/>
      </source>
    </interface>
    <!-- ... -->
  </devices>
  <!-- ... -->
</domain>
```

The various properties (for example `domain`, `bus`, etc.) of the `address` element should contain the proper values for pointing at the desired device (check with `lspci`).

4.6   Configuring storage

The storage configuration is critical, as in plays an important role in terms of performance.

4.6.1   Configuring storage for operating system volumes

The performance of storage where the operating system is installed is not critical for the performance of SAP HANA. Therefore any KVM supported storage may be used to deploy the operating system itself. See an example below:

```
<domain type='kvm'>
  <!-- ... -->
  <devices>
    <!-- ... -->
    <disk type='block' device='disk'>
      <driver name='qemu' type='raw' cache='none' io='native'/>
      <source dev='/dev/disk/by-id/wwn-0x600000e00d29000000293db000520000'/>
      <target dev='vda' bus='virtio'/>
    </disk>
    <!-- ... -->
  </devices>
  <!-- ... -->
</domain>
```

The `dev` attribute of the `source` element should contain the appropriate path.

4.6.2   Configuring storage for SAP HANA volumes

The configuration depends on the type of storage used for the SAP HANA Database.

In any case, the storage for SAP HANA must be able to fulfill the storage requirements for SAP HANA from within the VM. The SAP HANA Hardware and Cloud Measurement Tools (HCMT) can be used to assess if the storage meets the requirements. For details on HCMT refer to SAP Note 2493172 - "SAP HANA Cloud and Hardware Measurement Tools" (https://launchpad.support.s-ap.com/#/notes/2493172) ↗ .

4.6.2.1 Network attached storage

Follow the SAP HANA specific best practices of the storage system vendor. Make sure though that the VM has access to one or more dedicated 10 GiB Ethernet interfaces for the network traffic to the network attached storage.

4.6.2.2 Fibre Channel storage

Since storage controller passthrough is used (see *Section 3.3, "Configuring storage on the hypervisor"*), any LVM (Logical Volume Manager) and Multipathing configuration should, if wanted, be made inside the guest VM, always following the storage layout recommendations from the appropriate hardware vendor.

The guest VM XML configuration must be based on the underlying storage configuration on the hypervisor (see *Section 3.3, "Configuring storage on the hypervisor"*)

Since the storage for HANA (`/data`, `/log/` and `/shared` volumes) is performance critical, it is recommended to take advantage of an SAN HBA that is passed through to the guest VM.

Note that it is not possible to only use one function of the adapter, and both must always be attached to the guest VM. An example guest VM configuration with storage passthrough configured would look like the below (adjust the domain, bus, slot and function attributes of the `address` elements to match the adapter you chose):

```
<domain type='kvm'>
  <!-- ... -->
  <devices>
    <!-- ... -->
    <hostdev mode='subsystem' type='pci' managed='yes'>
      <source>
        <address domain='0x0000' bus='0x85' slot='0x00' function='0x0'/>
      </source>
    </hostdev>
    <hostdev mode='subsystem' type='pci' managed='yes'>
      <source>
```

```
        <address domain='0x0000' bus='0x85' slot='0x00' function='0x1'/>
      </source>
    </hostdev>
    <!-- ... -->
  </devices>
  <!-- ... -->
</domain>
```

More details about how to directly assign PCI devices to a guest VM are described in section 14.7 "Adding a PCI Device" of the Virtualization Guide (https://documentation.suse.com/sles/15-SP2/html/SLES-all/cha-libvirt-config-virsh.html#sec-libvirt-config-pci-virsh ↗).

4.7 Setting up a `vhostmd` device

The `vhostmd` device is passed to the VM so that the `vm-dump-metrics command` can retrieve metrics about the hypervisor provided by `vhostmd`. You can use either a vbd disk or a virtio-serial device (preferred) to set this up (see SAP Note 1522993 - "Linux: SAP on SUSE KVM - Kernel-based Virtual Machine" (https://launchpad.support.sap.com/#/notes/1522993) ↗ for details).

4.8 Setting up clocks and timers

Make sure that the clock timers are set up as follows, in the guest VM configuration file:

```
<domain type='kvm'>
  <!-- ... -->
  <clock offset='utc'>
    <timer name='rtc' tickpolicy='catchup'/>
    <timer name='pit' tickpolicy='delay'/>
    <timer name='hpet' present='no'/>
  </clock>
  <!-- ... -->
</domain>
```

4.9 Setting up the 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:

```
  <domain type='kvm'>
```

```
  <!-- ... -->
  <devices>
    <!-- ... -->
    <rng model='virtio'>
      <backend model='random'>/dev/urandom</backend>
    </rng>
    <!-- ... -->
  </devices>
  <!-- ... -->
</domain>
```

4.10 Configuring special features

It is necessary to enable for the guest VM a set of optimizations that are specific for the cases when the vCPUs are pinned and have (semi-)dedicated pCPUs all for themselves. This is done by having the following in the guest VM configuration file:

```
<domain type='kvm'>
  <!-- ... -->
  <features>
    <!-- ... -->
    <kvm>
      <hint-dedicated state='on'/>
    </kvm>
  </features>
  <!-- ... -->
</domain>
```

Note that this is a requirement for making it possible to load and use the "cpuidle-haltpoll" kernel module inside of the guest VM OS (see *Section 5.2.3.3, "Activating and configuring `haltpoll`"*).

5 Installing the guest operating system

5.1 Installing SUSE Linux Enterprise Server for SAP Applications inside the Guest VM

Refer to the SUSE Guide "`SUSE Linux Enterprise Server for SAP Applications 15 (https://documen-tation.suse.com/sles-sap/15-SP2/)` ↗.

5.2 Configuring the guest operating system for SAP HANA

Install and configure SUSE Linux Enterprise Server for SAP Applications 15 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 2684254 - "SAP HANA DB: Recommended OS settings for SLES 15 / SLES for SAP Applications 15" (https://launchpad.support.sap.com/#/notes/2205917) ↗

5.2.1 Customizing the Linux kernel parameters of the guest

Like the hypervisor host, the VM also needs special kernel parameters to be set. 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".

```
mitigations=auto kvm.nx_huge_pages=off intremap=no_x2apic_optout
```

A detailed explanation of these parameters has been given in *Section 3.4.6, "Technical explanation of the above described configuration settings"*.

5.2.2 Enabling host monitoring

The VM needs to have the `vm-dump-metrics` package installed, which dumps the metrics provided by the `vhostmd` service running on the hypervisor. This enables SAP HANA can collect data about the hypervisor. SAP Note 1522993 - "Linux: SAP on SUSE KVM - Kernel-based Virtual Machine" (https://launchpad.support.sap.com/#/notes/1522993) ↗ describes how to set up the virtual devices for `vhostmd` and how to configure it. When using a virtual disk for `vhostmd`, the virtual disk device must be world-readable, which is ensured with the boot time configuration below.

5.2.3 Configuring the Guest at boot time

The folling settings need to be configured at boot time of the VM. To persist these configurations it is recommended to put the commands provided below into a script which is executed as part of the boot process.

5.2.3.1 Disabling `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
```

5.2.3.2 Activating and configuring `sapconf` or `saptune`

The following parameters need to be set in `sapconf` version 5. Edit the file `/etc/sysconfig/sapconf` to reflect the settings below, and then restart the `sapconf` service.

```
GOVERNOR=performance
PERF_BIAS=performance
MIN_PERF_PCT=100
FORCE_LATENCY=5
```

> **Note**
>
> When using `sapconf` version 5, stop and disable the `tuned` service and instead enable and start the `sapconf` service.

If you use `saptune`, configure it accordingly:

- Apply the `HANA` solution: `saptune solution apply HANA`

- Create the file `/etc/saptune/override/2684254` with the following content.

```
[cpu]
force_latency=5
```

- Re-apply the recommendations for SAP Note 2684254: `saptune note apply 2684254`

Detailed documentation on `saptune` is available in chapter Tuning systems with `saptune` (https://documentation.suse.com/sles-sap/15-SP2/html/SLES-SAP-guide/cha-tune.html) ↗ of the SUSE Linux Enterprise Server for SAP Applications Guide.

5.2.3.3 Activating and configuring `haltpoll`

```
POLL_NS=800000
```

```
GROW_START=200000
modprobe cpuidle-haltpoll
echo $POLL_NS > /sys/module/haltpoll/parameters/guest_halt_poll_ns
echo $GROW_START > /sys/module/haltpoll/parameters/guest_halt_poll_grow_start
```

5.2.3.4　Setting the clock source

The clock source needs to be set to `tsc`.

```
echo tsc > /sys/devices/system/clocksource/clocksource0/current_clocksource
```

5.2.3.5　Disabling Kernel Same Page Merging

Kernel Same Page Merging (KSM) needs to be disabled, like on the hypervisor (see *Section 3.4.4, "Kernel Samepage Merging (ksm)"*).

```
echo 2 >/sys/kernel/mm/ksm/run
```

5.2.3.6　Implementing automatic configuration at boot time

The following script is provided as an example for a script implementing above recommendations, to be executed at boot time of the VM.

EXAMPLE 2: SCRIPT

```
#!/usr/bin/env bash
#
# Configure KVM guest for SAP HANA
#

POLL_NS=800000
GROW_START=200000

# disable irqbalance
systemctl disable --now irqbalance

modprobe cpuidle-haltpoll
echo $POLL_NS > /sys/module/haltpoll/parameters/guest_halt_poll_ns
echo $GROW_START > /sys/module/haltpoll/parameters/guest_halt_poll_grow_start

# Set clocksource to tsc
```

SUSE Best Practices for SAP HANA on KVM

```
echo tsc > /sys/devices/system/clocksource/clocksource0/current_clocksource

# disable Kernel Samepage Merging
echo 2 >/sys/kernel/mm/ksm/run
# 2: disable it, but make sure you also purify everything with fire!

# fix access to vhostmd device, so that SIDadm can read it
# see function setup_vhostmd_guest_device() in qacss-schwifty-common

# the vhostmd device has exactly 256 blocks, try to catch that from /proc/partitions
VHOSTMD_DEVICE=$(grep "   256 " /proc/partitions | awk '{print $4}' )
if [ -n "$VHOSTMD_DEVICE" ]; then
  chmod o+r /dev/"$VHOSTMD_DEVICE"
else
  echo "Missing vhostmd device, please check you XML file."
fi
```

Both `sapconf` and `saptune` apply their settings at boot time automatically and do not need to be included in the script above.

5.3   Configuring the guest operating system storage 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.

6   Performance considerations

The Linux kernel has code to mitigate existing vulnerabilities of the 1st Generation Intel Xeon Scalable Processor (Skylake) CPUs. Our testing showed no visible impact of those mitigations with regard to SAP HANA performance, except for the iTLB Multihit (https://www.kernel.org/doc/html/latest/admin-guide/hw-vuln/multihit.html) ↗ mitigation. This mitigation can be controlled by the kernel parameter `kvm.nx_huge_pages` (see SUSE support document 7023735 (https://www.suse.com/support/kb/doc/?id=000019411) ↗).

In general, the setting of parameter `kvm.nx_huge_pages` has an impact on performance. The implications on performance need to be considered as laid out in the Skylake example below.

Performance deviations for virtualization as measured on Intel Skylake (Bare Metal to single VM):

- Setting `kvm.nx_huge_pages=off`

  - The measured performance deviation for OLTP or mixed OLTP/OLAP workload is below 10%.

  - The measured performance deviation for OLAP workload is below 5%.

- Setting `kvm.nx_huge_pages=auto`

  - The measured performance deviation for OLTP or mixed OLTP/OLAP was impacted by this setting. For S/4HANA standard workload, OLTP transactional request times show an overhead of up to 30 ms. This overhead leads to an additional transactional throughput loss, but did not exceed 10%, running at a very high system load, when compared to the underlying bare metal environment.

  - The measured performance deviation for OLAP workload is below 5%.

  - During performance analysis with standard workload, most of the test cases stayed within the defined KPI of 10% performance degradation compared to bare metal. However, there are low-level performance tests in the test suite exercising various HANA kernel components that exhibit a performance degradation of more than 10%. This also indicates that there are particular scenarios which might not be suited for SAP HANA on SUSE KVM with kvm.nx_huge_pages = AUTO; especially those workloads generating high resource utilization, which must be considered when sizing SAP HANA instance in a SUSE KVM virtual machine. Thorough test of configuration for all workload conditions are highly recommended.

SUSE Best Practices for SAP HANA on KVM

7 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 15 (https://documentation.suse.com/sles/15-SP2/html/SLES-all/cha-libvirt-managing.html ↗)

- SUSE Virtualization Best Practices for SUSE Linux Enterprise Server 15 (https://documentation.suse.com/sles/15-SP2/html/SLES-all/article-vt-best-practices.html ↗)

7.1 Useful commands on the hypervisor

Check kernel boot options used:

```
cat /proc/cmdline
```

Check huge page status (This command can also be used to monitor the progress of huge page allocation during VM start):

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

List all VM guest domains configured on the hypervisor:

```
virsh list --all
```

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

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

Shut down a VM:

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

This is the location of VM guest configuration files:

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

This is the location of VM Log files:

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

7.2 Useful commands inside the VM guest

Checking L3 cache has been enabled in the guest:

```
lscpu | grep L3
```

Validate guest and host CPU topology:

```
lscpu
```

8 Examples

8.1 Example guest VM XML

✋ 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 the **SUSE Best Practices for SAP HANA on KVM - SUSE Linux Enterprise Server for SAP Applications 15 SP2** document at hand for a full explanation):

- Memory

  - The hypervisor has 3 TiB RAM (or 3072 GiB), of which 2772 GiB has been allocated as 1 GB huge pages and therefore 2772 GiB is the max VM size in this case

  - 2772 GiB = 2906652672 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 4.4, "Mapping vCPU and vNUMA topology and pinning"* for details

  - Note the topology of the guest VM mirrors the one of the hypervisor (4x28 CPU cores)

- Network I/O

    - Virtual functions of the physical network interface card have been added as PCI devices

- Storage I/O

    - A single SAN HBA is passed through to the VM as `hostdev` device (one for each function/port)

    - See *Section 4.6, "Configuring storage"* for details

- `rng model='virtio'`, for details see *Section 4.9, "Setting up the Virtio Random Number Generator (RNG) device"*

- `qemu:commandline` elements to describe CPU attributes, for details see *Section 4.2, "Configuring global vCPU"*

The following VM definition is an example for a VM configured to consume a 4-socket server with 3 TiB of main memory. It is taken from our actual validation machine. Note that this file is abridged for clarity; the cut is denoted by a `[…]` mark.

```
# 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'>
  <name>kvmvm11</name>
  <uuid>f529e0b0-93cc-4e83-87dc-65cb9922336d</uuid>
  <description>kvmvm11</description>
  <metadata>
    <libosinfo:libosinfo xmlns:libosinfo="http://libosinfo.org/xmlns/libvirt/domain/1.0">
      <libosinfo:os id="http://suse.com/sle/15.2"/>
    </libosinfo:libosinfo>
  </metadata>
  <memory unit='KiB'>2906652672</memory>
  <currentMemory unit='KiB'>2906652672</currentMemory>
  <memoryBacking>
    <hugepages>
      <page size='1048576' unit='KiB'/>
    </hugepages>
    <nosharepages/>
```

```
    </memoryBacking>
  <vcpu placement='static'>224</vcpu>
  <cputune>
    <vcpupin vcpu='0' cpuset='0,112'/>
    <vcpupin vcpu='1' cpuset='0,112'/>
    <vcpupin vcpu='2' cpuset='1,113'/>
    <vcpupin vcpu='3' cpuset='1,113'/>
    <vcpupin vcpu='4' cpuset='2,114'/>
    <vcpupin vcpu='5' cpuset='2,114'/>
    <vcpupin vcpu='6' cpuset='3,115'/>
    <vcpupin vcpu='7' cpuset='3,115'/>
    <vcpupin vcpu='8' cpuset='4,116'/>
    <vcpupin vcpu='9' cpuset='4,116'/>
    <vcpupin vcpu='10' cpuset='5,117'/>
    <vcpupin vcpu='11' cpuset='5,117'/>
[...]
    <vcpupin vcpu='214' cpuset='107,219'/>
    <vcpupin vcpu='215' cpuset='107,219'/>
    <vcpupin vcpu='216' cpuset='108,220'/>
    <vcpupin vcpu='217' cpuset='108,220'/>
    <vcpupin vcpu='218' cpuset='109,221'/>
    <vcpupin vcpu='219' cpuset='109,221'/>
    <vcpupin vcpu='220' cpuset='110,222'/>
    <vcpupin vcpu='221' cpuset='110,222'/>
    <vcpupin vcpu='222' cpuset='111,223'/>
    <vcpupin vcpu='223' cpuset='111,223'/>
  </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>
  <resource>
    <partition>/machine</partition>
  </resource>
  <os>
    <type arch='x86_64' machine='pc-q35-4.2'>hvm</type>
    <loader readonly='yes' type='pflash'>/usr/share/qemu/ovmf-x86_64-smm-ms-code.bin</
loader>
    <nvram>/var/lib/libvirt/qemu/nvram/kvmvm12_VARS.fd</nvram>
    <boot dev='hd'/>
  </os>
  <features>
    <acpi/>
    <apic/>
```

```
    <pae/>
    <kvm>
      <hint-dedicated state='on'/>
    </kvm>
    <vmport state='off'/>
  </features>
  <cpu mode='host-passthrough' check='none'>
    <topology sockets='4' cores='28' threads='2'/>
    <feature policy='require' name='rdtscp'/>
    <feature policy='require' name='invtsc'/>
    <feature policy='require' name='x2apic'/>
    <numa>
      <cell id='0' cpus='0-55' memory='726663168' unit='KiB'>
        <distances>
          <sibling id='0' value='10'/>
          <sibling id='1' value='21'/>
          <sibling id='2' value='21'/>
          <sibling id='3' value='21'/>
        </distances>
      </cell>
      <cell id='1' cpus='56-111' memory='726663168' unit='KiB'>
        <distances>
          <sibling id='0' value='21'/>
          <sibling id='1' value='10'/>
          <sibling id='2' value='21'/>
          <sibling id='3' value='21'/>
        </distances>
      </cell>
      <cell id='2' cpus='112-167' memory='726663168' unit='KiB'>
        <distances>
          <sibling id='0' value='21'/>
          <sibling id='1' value='21'/>
          <sibling id='2' value='10'/>
          <sibling id='3' value='21'/>
        </distances>
      </cell>
      <cell id='3' cpus='168-223' memory='726663168' unit='KiB'>
        <distances>
          <sibling id='0' value='21'/>
          <sibling id='1' value='21'/>
          <sibling id='2' value='21'/>
          <sibling id='3' value='10'/>
        </distances>
      </cell>
    </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>destroy</on_crash>
    <pm>
      <suspend-to-mem enabled='no'/>
      <suspend-to-disk enabled='no'/>
    </pm>
    <devices>
      <emulator>/usr/bin/qemu-system-x86_64</emulator>
      <disk type='block' device='disk'>
        <driver name='qemu' type='raw' cache='none' io='native'/>
        <source dev='/dev/disk/by-id/wwn-0x600000e00d29000000293db000520000'/>
        <target dev='vda' bus='virtio'/>
        <address type='pci' domain='0x0000' bus='0x04' slot='0x00' function='0x0'/>
      </disk>
      <disk type='file' device='disk'>
        <driver name='qemu' type='raw'/>
        <source file='/dev/shm/vhostmd0'/>
        <target dev='vdx' bus='virtio'/>
        <readonly/>
        <address type='pci' domain='0x0000' bus='0x0b' slot='0x00' function='0x0'/>
      </disk>
      <controller type='usb' index='0' model='qemu-xhci' ports='15'>
        <address type='pci' domain='0x0000' bus='0x02' slot='0x00' function='0x0'/>
      </controller>
      <controller type='sata' index='0'>
        <address type='pci' domain='0x0000' bus='0x00' slot='0x1f' function='0x2'/>
      </controller>
      <controller type='pci' index='0' model='pcie-root'/>
      <controller type='pci' index='1' model='pcie-root-port'>
        <model name='pcie-root-port'/>
        <target chassis='1' port='0x10'/>
        <address type='pci' domain='0x0000' bus='0x00' slot='0x02' function='0x0'
multifunction='on'/>
      </controller>
      <controller type='pci' index='2' model='pcie-root-port'>
        <model name='pcie-root-port'/>
        <target chassis='2' port='0x11'/>
        <address type='pci' domain='0x0000' bus='0x00' slot='0x02' function='0x1'/>
      </controller>
      <controller type='pci' index='3' model='pcie-root-port'>
        <model name='pcie-root-port'/>
        <target chassis='3' port='0x12'/>
```

```
      <address type='pci' domain='0x0000' bus='0x00' slot='0x02' function='0x2'/>
    </controller>
    <controller type='pci' index='4' model='pcie-root-port'>
      <model name='pcie-root-port'/>
      <target chassis='4' port='0x13'/>
      <address type='pci' domain='0x0000' bus='0x00' slot='0x02' function='0x3'/>
    </controller>
    <controller type='pci' index='5' model='pcie-root-port'>
      <model name='pcie-root-port'/>
      <target chassis='5' port='0x14'/>
      <address type='pci' domain='0x0000' bus='0x00' slot='0x02' function='0x4'/>
    </controller>
    <controller type='pci' index='6' model='pcie-root-port'>
      <model name='pcie-root-port'/>
      <target chassis='6' port='0x15'/>
      <address type='pci' domain='0x0000' bus='0x00' slot='0x02' function='0x5'/>
    </controller>
    <controller type='pci' index='7' model='pcie-root-port'>
      <model name='pcie-root-port'/>
      <target chassis='7' port='0x16'/>
      <address type='pci' domain='0x0000' bus='0x00' slot='0x02' function='0x6'/>
    </controller>
    <controller type='pci' index='8' model='pcie-root-port'>
      <model name='pcie-root-port'/>
      <target chassis='8' port='0x17'/>
      <address type='pci' domain='0x0000' bus='0x00' slot='0x02' function='0x7'/>
    </controller>
    <controller type='pci' index='9' model='pcie-root-port'>
      <model name='pcie-root-port'/>
      <target chassis='9' port='0x18'/>
      <address type='pci' domain='0x0000' bus='0x00' slot='0x03' function='0x0'
 multifunction='on'/>
    </controller>
    <controller type='pci' index='10' model='pcie-root-port'>
      <model name='pcie-root-port'/>
      <target chassis='10' port='0x19'/>
      <address type='pci' domain='0x0000' bus='0x00' slot='0x03' function='0x1'/>
    </controller>
    <controller type='pci' index='11' model='pcie-root-port'>
      <model name='pcie-root-port'/>
      <target chassis='11' port='0x1a'/>
      <address type='pci' domain='0x0000' bus='0x00' slot='0x03' function='0x2'/>
    </controller>
    <controller type='pci' index='12' model='pcie-root-port'>
      <model name='pcie-root-port'/>
      <target chassis='12' port='0x1b'/>
      <address type='pci' domain='0x0000' bus='0x00' slot='0x03' function='0x3'/>
```

```
  </controller>
  <controller type='pci' index='13' model='pcie-root-port'>
    <model name='pcie-root-port'/>
    <target chassis='13' port='0x1c'/>
    <address type='pci' domain='0x0000' bus='0x00' slot='0x03' function='0x4'/>
  </controller>
  <controller type='pci' index='14' model='pcie-root-port'>
    <model name='pcie-root-port'/>
    <target chassis='14' port='0x1d'/>
    <address type='pci' domain='0x0000' bus='0x00' slot='0x03' function='0x5'/>
  </controller>
  <controller type='virtio-serial' index='0'>
    <address type='pci' domain='0x0000' bus='0x03' slot='0x00' function='0x0'/>
  </controller>
  <interface type='direct'>
    <mac address='0c:fd:37:92:dc:99'/>
    <source dev='eth11' mode='vepa'/>
    <model type='virtio'/>
    <address type='pci' domain='0x0000' bus='0x01' slot='0x00' function='0x0'/>
  </interface>
  <interface type='hostdev' managed='yes'>
    <mac address='52:54:00:7f:12:fb'/>
    <driver name='vfio'/>
    <source>
      <address type='pci' domain='0x0000' bus='0x17' slot='0x02' function='0x0'/>
    </source>
    <address type='pci' domain='0x0000' bus='0x0c' slot='0x00' function='0x0'/>
  </interface>
  <serial type='pty'>
    <target type='isa-serial' port='0'>
      <model name='isa-serial'/>
    </target>
  </serial>
  <console type='pty'>
    <target type='serial' port='0'/>
  </console>
  <channel type='unix'>
    <target type='virtio' name='org.qemu.guest_agent.0'/>
    <address type='virtio-serial' controller='0' bus='0' port='1'/>
  </channel>
  <channel type='spicevmc'>
    <target type='virtio' name='com.redhat.spice.0'/>
    <address type='virtio-serial' controller='0' bus='0' port='2'/>
  </channel>
  <input type='tablet' bus='usb'>
    <address type='usb' bus='0' port='1'/>
  </input>
```

```
    <input type='mouse' bus='ps2'/>
    <input type='keyboard' bus='ps2'/>
    <graphics type='spice' autoport='yes'>
      <listen type='address'/>
      <image compression='off'/>
    </graphics>
    <sound model='ich9'>
      <address type='pci' domain='0x0000' bus='0x00' slot='0x1b' function='0x0'/>
    </sound>
    <video>
      <model type='qxl' ram='65536' vram='65536' vgamem='16384' heads='1' primary='yes'/>
      <address type='pci' domain='0x0000' bus='0x00' slot='0x01' function='0x0'/>
    </video>
    <!-- SAN 2-port HBA passthrough configuration -->
    <hostdev mode='subsystem' type='pci' managed='yes'>
      <source>
        <address domain='0x0000' bus='0x85' slot='0x00' function='0x0'/>
      </source>
      <address type='pci' domain='0x0000' bus='0x0d' slot='0x00' function='0x0'/>
    </hostdev>
    <hostdev mode='subsystem' type='pci' managed='yes'>
      <source>
        <address domain='0x0000' bus='0x85' slot='0x00' function='0x1'/>
      </source>
      <address type='pci' domain='0x0000' bus='0x0e' slot='0x00' function='0x0'/>
    </hostdev>
    <redirdev bus='usb' type='spicevmc'>
      <address type='usb' bus='0' port='2'/>
    </redirdev>
    <redirdev bus='usb' type='spicevmc'>
      <address type='usb' bus='0' port='3'/>
    </redirdev>
    <memballoon model='virtio'>
      <address type='pci' domain='0x0000' bus='0x05' slot='0x00' function='0x0'/>
    </memballoon>
    <rng model='virtio'>
      <backend model='random'>/dev/urandom</backend>
      <address type='pci' domain='0x0000' bus='0x06' slot='0x00' function='0x0'/>
    </rng>
  </devices>
</domain>
```

9 Additional information

9.1 Resources

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

- SUSE Virtualization Guide for SUSE Linux Enterprise Server 15 (https://documenta-tion.suse.com/sles/15-SP2/html/SLES-all/book-virt.html) ↗

- SAP Note 3120786 - "SAP HANA on SUSE KVM Virtualization with SLES 15 SP2" (https://launch-pad.support.sap.com/#/notes/3120786) ↗

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

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

- SAP Note 2684254 - "SAP HANA DB: Recommended OS settings for SLES 15 / SLES for SAP" (https://launchpad.support.sap.com/#/notes/2205917) ↗

- SAP Note 1522993 - "Linux: SAP on SUSE KVM - Kernel-based Virtual Machine" (https://launch-pad.support.sap.com/#/notes/1522993) ↗

9.2 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 mailto:Report Documentation 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).

Mail

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

10 Legal notice

11 GNU Free Documentation License

Copyright © 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 noncommercially. 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 in order 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.

SUSE Best Practices for SAP HANA on KVM

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.