White Paper

Cisco HyperFlex Solutions

Running SAP Nonproduction Landscapes in a Cisco HyperFlex 2.0 All-Flash System

Author: Brian Everitt

4/10/2017

Executive Summary

SAP landscapes frequently are deployed in a fully virtualized system, often including a virtualized back-end database server. In recent years, SAP has been encouraging its customers to migrate to its own database platform of the future: SAP HANA. SAP HANA databases can be deployed as virtual servers. They can also be run on lower-cost platforms, such as Intel® Xeon® processor E5 family–based hardware, through the SAP HANA Tailored Datacenter Integration (TDI) program for entry-level systems. With the launch of the Cisco HyperFlex™ system, Cisco offers a low-cost, easy-to-deploy, high-performance hyperconverged virtual server platform that is an excellent solution for SAP landscapes. Although the Cisco HyperFlex system is fully capable of running the entire SAP landscape, and the Cisco HyperFlex server models are listed in the SAP HANA TDI Product Availability Matrix (PAM), official support for production-level SAP HANA databases on Cisco HyperFlex is not offered by SAP at this time. As a result, this combination of fully virtualized SAP application servers and SAP HANA within a Cisco HyperFlex system is an excellent choice for nonproduction landscapes, such as development or testing systems. Full production support is offered for Cisco HyperFlex systems used for the deployment of SAP application servers, in combination with other SAP HANA deployment strategies, which are described in separate white papers published by Cisco.

Solution Overview

This section defines current IT Infrastructure challenges and presents a high-level view of the Cisco HyperFlex and SAP HANA solution and its benefits.

Document Purpose

This document describes deployment strategies for a fully virtualized SAP landscape, using SAP NetWeaver application servers and an SAP HANA back-end database, running on a Cisco HyperFlex 2.0 all-flash cluster.

Solution Purpose

Cisco HyperFlex HX Data Platform solutions are built on the Cisco Unified Computing System™ (Cisco UCS®) platform. They offer faster deployment and greater flexibility and efficiency at a competitive price, while lowering risk for the customer. Proven components from Cisco are integrated to form a software-defined storage (SDS) platform. This approach eliminates or reduces the need for planning and configuration decisions, while allowing the levels of customization needed to address customer workload needs. The platform and management model adopted is an extension of established Cisco UCS data center strategy, in which familiar components are managed in a consistent manner through a policy-based framework with Cisco UCS Manager.

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

Business Challenge

Deployment of an SAP landscape in a typical data center environment presents several challenges:

- **Mixed hardware platforms:** Many landscapes are deployed in such a way that the NetWeaver application tier runs in a separate environment from the other components, such as the database tier, often on completely different hardware. This arrangement can also apply to storage; in this case, multiple storage systems may contain the application virtual machines, while the database data is stored elsewhere.
- **Multiple management silos:** Use of multiple hardware platforms and storage systems can lead to a disaggregated management model, in which only certain members of the IT staff can perform tasks on systems for which they have specialized knowledge.
- **Speed of deployment:** Deploying all the layers of a typical SAP environment can be time consuming, requiring setup of networking resources, computing hardware, storage arrays, SANs, and virtualization software.

The Solution

Cisco HyperFlex 2.0 all-flash systems can be used to rapidly deploy an SAP landscape as a fully virtualized system.

Solution Benefits

This solution provides the following benefits to customers:

1. **Single hardware platform:** Cisco UCS is the base platform for Cisco HyperFlex systems, which provide a fully contained hyperconverged environment, combining networking, storage, and virtualization resources in a single system.
2. **Simplified management:** A single administrator can manage all aspects of Cisco UCS and the Cisco HyperFlex system through Cisco UCS Manager and the VMware vCenter Web Client, making tasks much easier and faster to complete.
3. **Rapid deployment:** The programmability and ease of use of Cisco UCS Manager allow Cisco HyperFlex systems to be deployed quickly

Customers who have already invested in Cisco® products and technologies have the opportunity to mitigate their risk further by deploying familiar and tested Cisco UCS technology.

Recommendations Summary

The following list provides a summary of the recommendations for setting up your system:

- Consult the SAP Quick Sizer to help determine the memory, storage, and processor requirements for the SAP HANA database and the application servers to be used in your specific deployment.
- For your server, consider deploying the Cisco HyperFlex HXAF240c M4SX model nodes because it offers a larger number of capacity SSDs, with a higher performance potential.
- Select an Intel® processor model for your Cisco HyperFlex converged nodes that offers an optimal balance of CPU cores, clock speed, and price for your deployment and budget.
- Deploy servers with enough RAM to hold the entire SAP HANA virtual machine in memory. Do not overprovision memory for SAP HANA systems.
- Deploy the maximum number of servers in the Cisco HyperFlex cluster, with as many disks as possible, to benefit from the collective storage performance of all the nodes and disks.

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

- Make sure that the amount of RAM in the SAP HANA virtual machine is large enough to cache the entire database.
- Set **numa.vcpu.preferHT** to true for SAP HANA virtual machines and SAP application server virtual machines, to help ensure that they use hyperthreading properly.
- Configure SAP application server virtual machines with a number of virtual CPUs (vCPUs) that fit within a single non-uniform memory access (NUMA) node. Review the virtual machine's vmware.log file to help ensure that the configuration is correct.
- If the Hana database requires more RAM than is available to a single CPU socket, configure SAP HANA virtual machines with the maximum amount of vCPU resources possible, and specify "wide" virtual machines, spanning multiple NUMA nodes. Review the vmware.log file of the virtual machine to verify that the configuration is correct.
- Give consideration to running a "wide" SAP Hana virtual machine on a HyperFlex "compute-only" node, to avoid resource contention with the storage controller VM (SCVM).
- Reserve all guest virtual machine memory for every SAP HANA virtual machine.
- Use VMware Distributed Resource Scheduler (DRS) groups and affinity rules to keep SAP HANA virtual machines separate from other virtual machines.
- Configure SAP system monitoring at the guest virtual machine level.
- Deploy the solution on a Cisco HyperFlex cluster configured with a replication factor of 3 (RF3), to provide the best possible level of data protection, or alternatively consider replication factor of 2 (RF2), if the additional redundancy is not required, in exchange for lower write latency.

Detailed descriptions of these recommendations follow in the solution guide.

Solution Guide

The following sections describe the system architecture used for the solution, which provides a fully virtualized SAP landscape, using SAP HANA as the database platform and running within a Cisco HyperFlex 2.0 all-flash system. Additional sections present sizing guidelines for the CPU and memory resources, high-availability options, tuning options for virtualized SAP systems, and a reference system design tested by Cisco. The Cisco HyperFlex system provides a hyperconverged architecture, containing all the processor, memory, networking, and storage resources necessary to run a virtual server platform.

This document does not describe the design, installation, or configuration of the Cisco HyperFlex system; those details are covered in other Cisco Validated Design documents, which are listed in the references section at the end of this paper. In addition, this document does not provide detailed instructions about the installation of the SAP software or OS-level tuning required by SAP, although references to this information can also be found at the end of this paper.

Requirements

To configure an SAP landscape running on a Cisco HyperFlex 2.0 all-flash cluster you need:

- A functional and healthy running Cisco HyperFlex cluster
  - Cisco HyperFlex Release 2.0.1.a or later
  - Cisco UCS Firmware Release 3.1(2f) or later
  - VMware ESXi 6.0 Update 2 or later

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

- VMware vCenter Server Appliance or Microsoft Windows–based vCenter Server
- Appropriate software and licensing from SAP for SAP HANA and NetWeaver

Limitations

Based on information and support stances from SAP, VMware, and Cisco, a virtual SAP landscape with a virtualized SAP HANA database on Cisco HyperFlex is subject to the following limitations:

- Only single-node **scale-up** SAP HANA virtual machines can be deployed. Scale-out deployments are not allowed.
- SAP HANA virtual machines require a full memory reservation. Therefore, the maximum size of a single SAP HANA virtual machine cannot exceed the physical memory of a single server.
- SAP HANA virtual machines cannot be configured as fault-tolerant virtual machines in vCenter.

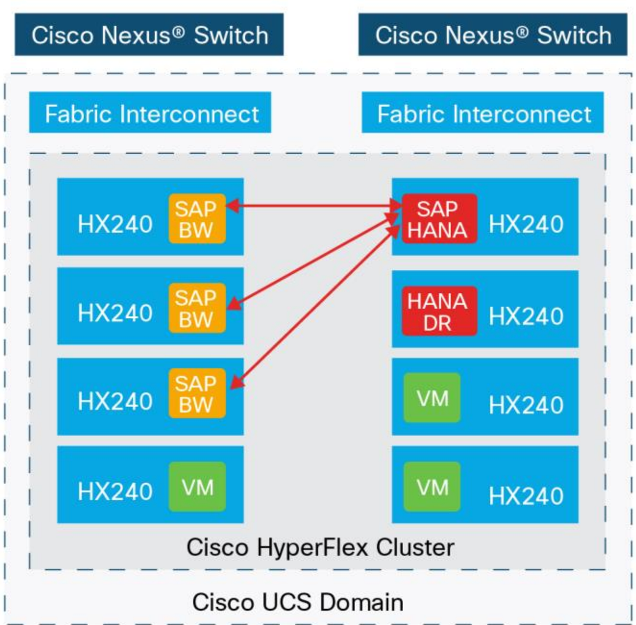
Solution Overview

For a fully virtualized SAP landscape, all SAP application server instances, central services instances, and database instances are constructed as virtual machines within the Cisco HyperFlex system. Therefore, various support and sizing guidelines from SAP, Cisco, and VMware need to be followed to help ensure that the virtual applications and database function at their maximum potential.

SAP Landscape Design

The recommended virtualized SAP landscape on Cisco HyperFlex design (Figure 1) includes one or more NetWeaver application servers. The tested design ran SAP Business Warehouse (BW) and a single-node SAP HANA server. This design is also known as a Scale-Up SAP HANA system. An optional component is an additional SAP HANA virtual machine, which can receive a replicated copy of the SAP HANA database through the SAP HANA System Replication feature.

Figure 1. SAP Landscape Design



© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

SAP HANA Deployment Options

In a nonproduction environment, SAP allows additional flexibility of virtualized HANA deployments, including scale-out, or multi-node SAP HANA database deployments. However, SAP does not support SAP HANA scale-out deployments on Intel Xeon processor E5 family–based systems. Because all Cisco HyperFlex server models are based on Intel Xeon processor E5 family processors, only single-node scale-up SAP HANA virtual machines can be deployed. However, multiple independent SAP HANA virtual machines can be deployed. These can operate separate database instances, or they can be part of an active-passive partnership through SAP HANA system replication. Alternatively, if multiple HANA databases are needed, consider using the SAP HANA Multi-Tenant Database Containers (MDC) feature, which allows multiple independent databases to operate under the management of a single installation of SAP HANA.

SAP System Landscape Sizing

An important first step in designing any SAP system is to properly size the system based on a number of factors:

- Applications to be used
- Number of users
- Relative performance requirements and resource utilization of the users
- Amount of data in the system

To facilitate this process, SAP provides various sizing resources at https://service.sap.com/sizing. One of the primary resources is the SAP Quick Sizer, which helps determine the memory and processing requirements of both the application servers and the database. To achieve accurate results using the Quick Sizer, you need:

- Clear definitions of the various software components that will be used in the SAP system
- Detailed estimates of the number of users and the characteristics of their daily use of the system
- Information about the size and type of data already stored in the database, if one exists

Given the potential complexity of this information, for the purposes of this document Cisco has defined a base workload of an SAP system using SAP BW powered by SAP HANA, a community of about 700 users, and a database with 250 GB of columnar data. For assistance with the proper sizing of your SAP system, please contact your Cisco account manager to discuss the services offered for SAP system landscape planning and sizing.

Tables 1 and 2 show sample Quick Sizer input and output.

Table 1. Sample SAP Quick Sizer BW on HANA Input

Sample SAP Quick Sizer BW on HANA Input	
Business Planning and Simulation 1 User	100
Business Planning and Simulation 2 Users	50
Business Planning and Simulation 3 Users	25
Information Consumers	350
Business Users	100
Expert Users	50
Database Column Store Footprint	250 GB

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

Table 2. Sample SAP Quick Sizer BW on HANA Output

Sample SAP Quick Sizer BW on HANA Output	
SAP BW Server SAP Application Performance Standard (SAPS) Required	11,000
SAP BW Server Advanced Business Application Programming (ABAP) Memory	73 GB
SAP BW Server Single Computing Unit (SCU) Class	A
SAP HANA Database Server Memory	235 GB
SAP HANA Database Server Storage Required	972 GB

Cisco HyperFlex Node Models

Cisco HyperFlex all-flash systems are available in two base models: the Cisco HyperFlex HXAF220c M4 Node and the Cisco HyperFlex HXAF240c M4 Node. The primary difference between the two models lies in the number of disks that can be installed in the converged storage nodes. The HXAF220c server is limited to a maximum of 6 solid state disks (SSDs) to provide storage for the overall capacity of the cluster, whereas the HXAF240c can scale up to 10 SSDs per server, as of HXDP release 2.0.1.a. The caching solid-state disk (SSD) capacity of the two models is identical; each server is built with an 800 GB high endurance SSD for write caching.

The larger number of SSDs available to service read I/O in the HXAF240 server provides increased performance potential, making the HXAF240 model a better choice for a system that will run a database workload. In addition, future releases of the HXDP software will enable the HXAF240 server to utilize the full complement of 23 SSDs, further improving performance. For these reasons, Cisco recommends that any virtualized SAP landscape on a Cisco HyperFlex system that also has a virtualized SAP HANA (vHANA) database, in which the primary storage of the database is provided by the Cisco HyperFlex distributed storage file system, should use the HXAF240 model servers (Figure 2).

Figure 2. Cisco HyperFlex HXAF240 M4SX Server

Physical CPU Selection

Processor performance results provided by the Quick Sizer require special attention. In the design described here, a vHANA system is running on the same servers, so you need to choose an appropriate processor model for the servers. For the Cisco HyperFlex server nodes, any model of Intel Xeon processor E5-2600 v4 CPU is supported, allowing for a large amount of flexibility. However, the designs of officially supported SAP HANA appliance systems specify that certain guidelines should be followed. Certified SAP HANA appliances always use the highest tiers of Intel Xeon processor E7 CPUs, which combine high core counts, relatively high-frequency speeds, and greater amounts of cache memory. The goal of the SAP HANA TDI entry-level program, which allows the use of less expensive Intel Xeon processor E5 CPUs, is to provide customers with lower-priced options for running their SAP HANA systems. Although the program does allow the use of the lowest-priced Intel Xeon processor E5 CPU models, these models do not have the optimum combination of core count, clock speed, and cache memory needed to most effectively run an SAP HANA database.

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

For example, the highest-level CPU model, the Intel Xeon processor E5-2699 v4, has 22 cores per socket, but they run at a clock speed of only 2.2 GHz, and this model is nearly as expensive as the Intel Xeon processor E7-8880 v4. At the other end of the spectrum, the Intel Xeon processor E5-2637 v4 runs at a clock speed of 3.5 GHz, but it has only 4 cores per socket. SAP HANA generally benefits more from additional CPU cores than from a higher clock speed for each individual core. Another factor to consider is the SAP Application Performance Standard, or SAPS, requirement of the application servers reported in the Quick Sizer output. SAPS is a server performance benchmark standard measurement from SAP. In the example described here, a value of 11,000 SAPS is required by the application servers. All current Intel Xeon processor E5 v4 CPUs are capable of delivering that performance.

To find the best fit for the SAP HANA workload, you need to find the right balance in the Intel Xeon processor E5 product lineup that offers the best combination of cores, clock speed, cache, and price. Table 3 provides four recommendations of excellent combinations of the attributes needed for a high-performance, cost-effective Cisco HyperFlex cluster for hosting a vHANA system.

Table 3. Sample CPU Model Combinations for SAP HANA

Sample CPU Model Combinations for SAP HANA				
Intel Xeon Processor Model	E5-2660 v4	E5-2680 v4	E5-2683 v4	E5-2690 v4
Number of Cores per Socket	14	14	16	14
Clock Speed	2.0 GHz	2.4 GHz	2.1 GHz	2.6 GHz
Cache Size	35 MB	35 MB	40 MB	35 MB
Estimated SAPS value per CPU	36,450	42,250	43,350	43,950

Note: All processors recommended support the highest DDR4 DIMM clock speed of 2400 MHz.

In addition to the processor choice for the SAP HANA database workload, you need to consider the model used for the NetWeaver application servers. The output of the Quick Sizer provides a value for BW Server SCU Class. SAP defines SCU as a single computing unit, or as a somewhat generic representation of the relative CPU single-thread performance. A value of “A” indicates good single-thread performance, “AA” indicates very good performance, and “AAA” indicates excellent performance. In this value, SAP is attempting to distinguish between a CPU model with a high core count and a model with a high clock speed. Models with higher clock speeds will fall into the “AA” or “AAA” category: for instance, an E5-2637 v4 CPU running at 3.5 GHz would obviously be able to finish computing a single thread faster than would an E5-2609 v4 CPU running at 1.7 GHz. Certain SAP applications will benefit from higher SCU ratings, so the output of the Quick Sizer regarding the application server SCU performance must be viewed in combination with the needs of the SAP HANA database.

In the sample system discussed here, the recommended SCU rating for SAP BW application servers is only “A”. Therefore, processors with the highest clock speeds are not necessary. In a Cisco HyperFlex system, all processor models should be the same across all host servers. Therefore, you need to achieve a balance between the needs of the SAP HANA database virtual machines and NetWeaver virtual machines. The reference design created for this document includes E5-2690 v4 processors, which are an excellent combination of core count, clock speed, and cache size while remaining highly affordable compared to the Intel Xeon processor E7 models.

Table 4 summarizes the system used in this reference design.

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

Table 4. Cisco HyperFlex System and SAP Landscape Reference Design

Cisco HyperFlex System and SAP Landscape Reference Design	
Server Model	Cisco HyperFlex HXAF240 M4SX
Processor Model	Intel Xeon processor E5-2690 v4
RAM per Server	512 GB
Caching SSD	One 800 GB SSD per server
Capacity Disks	Ten 960 GB SSDs per server
Cisco HyperFlex Cluster Size	8 nodes
Cisco UCS Firmware Release	Release 3.1(2f)
Cisco HyperFlex Software Release	Release 2.0.1.a
VMware ESXi Hypervisor Release	Release 6.0 Update 2 Patch 4

Note: HyperFlex 2.0 all-flash systems can be ordered with 960-GB or 3.8-TB capacity SSDs. Choose the part that provides the capacity needed in the cluster. Performance results of the two models of disks is nearly identical.

Virtual Machine CPU Configuration

The selection of the physical CPUs in the Cisco HyperFlex nodes is related to the virtual CPU (vCPU) configuration of the SAP HANA virtual machines and the SAP application server virtual machines. In particular, you need to consider the number of physical cores per CPU socket when sizing the virtual machines and configuring the number of vCPUs they have.

Hyperthreading

Intel Xeon processor E5 family CPUs use Intel Hyper-Threading Technology, which enables each CPU core to accept a new thread for processing before the processing of the previous thread is complete. The hyperthreading feature commonly improves overall CPU processing throughput by up to 20 to 25 percent and therefore should be enabled in most circumstances. To enable the hypervisors or operating systems to submit these additional threads to the CPU cores, the underlying hardware presents each core as two cores. For example, to an OS or hypervisor, a CPU that has 10 cores and uses hyperthreading will look as though it is a CPU with 20 cores. Because the processor does not actually have 20 physical cores, these are commonly referred to as logical processors. This approach allows the OS or hypervisor to submit the additional threads for processing, and the server hardware will handle the processing of these threads and return the results.

Non-Uniform Memory Access

Modern Intel processor–based servers are designed using the non-uniform memory access (NUMA) architecture, in which each physical CPU socket has its own set of physical RAM resources to which that CPU has a direct physical connection. Memory that is controlled by another CPU socket is accessible from all other CPU sockets, however, the memory calls have to traverse the server’s front-side CPU bus. Although these buses have extremely high bandwidth, this additional hop can add to the latency of the memory calls. A single CPU socket and the memory it directly controls are known as a NUMA node. In most circumstances, a virtual machine will achieve the best performance if all of its vCPUs are scheduled within a single NUMA node, because this approach will make calls to remote memory on other NUMA nodes unnecessary. If the vCPU settings of a virtual machine cannot fit within a single NUMA node, this virtual machine is called a wide virtual machine.

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

vCPU Settings

To optimize the vCPU settings of the SAP HANA virtual machine and SAP application servers, you need to consider the number of physical CPU cores, the use of hyperthreading, and the size of the NUMA nodes. The settings are different for the SAP HANA database virtual machines and the SAP application server virtual machines, primarily because SAP HANA virtual machines will benefit significantly from the configuration of more vCPUs.

SAP Application Server Virtual Machines

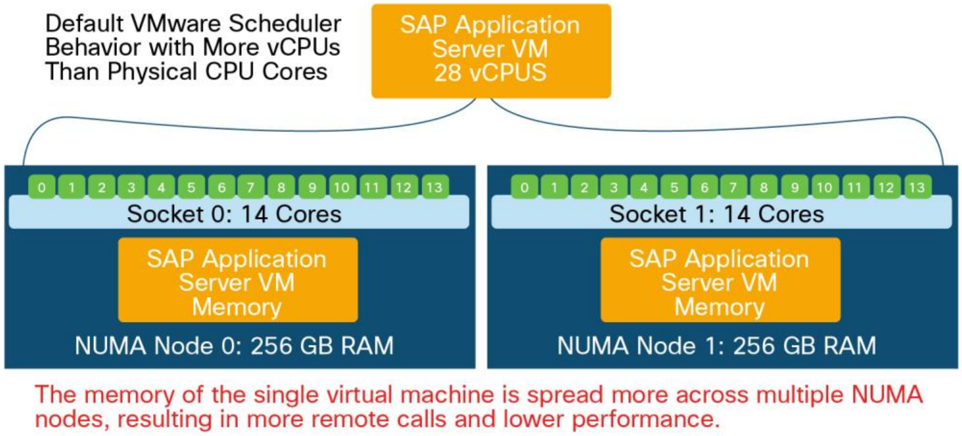
The CPU resource requirements for NetWeaver Advanced Business Application Programming, or ABAP, application servers are lower than those for SAP HANA database servers in most deployments. Because fewer vCPUs are required by the application server virtual machines, the most important setting is to specify that all the vCPUs configured for these virtual machines are contained in a single NUMA node. A basic sizing guideline is to configure an application server virtual machine with as many vCPUs as there are cores in a single physical CPU socket. For example, the reference server described here with Intel Xeon processor E5-2690 v4 CPUs has 14 cores per socket. Configuring the application server virtual machines with 14 vCPUs would be a safe initial sizing.

Configuring a virtual machine to have no more vCPUs than the number of cores in a single socket will help ensure that all the virtual machines' vCPUs and memory will be scheduled within one NUMA node. However, this configuration does not provide the maximum potential performance for the system because it does not take advantage of hyperthreading. Another guideline for sizing the number of vCPUs for an SAP application server virtual machine is to configure the virtual machine with as many vCPUs as there are logical processors of a single physical CPU socket. For example, the reference server described here with Intel Xeon processor E5-2690 v4 CPUs has 14 cores per socket with hyperthreading, so there are 28 logical processors per CPU socket. Configuring the application server virtual machines with 28 vCPUs would provide better performance and offers the option to create additional ABAP dialog processes on the server to take advantage of the additional vCPUs.

When a virtual machine is configured with more vCPUs than the number of physical cores of a single NUMA node, the default CPU scheduler in VMware ESXi will schedule those vCPUs across multiple NUMA nodes. The scheduler does not take into account the hyperthreading feature, which allows each core to process two threads. As a result, the vCPUs are not all contained within a single NUMA node, and therefore the memory of the virtual machine is also spread across multiple NUMA nodes. In the example described here, a virtual machine with 28 vCPUs on a server with 14-core physical processors would have the vCPUs scheduled across both physical sockets. This configuration does not result in the best performance for that virtual machine, because it can result in a much higher percentage of remote memory calls across NUMA nodes: for example, a vCPU scheduled on a core of socket 0 may have to access memory pages in the RAM of socket 1 (Figure 3).

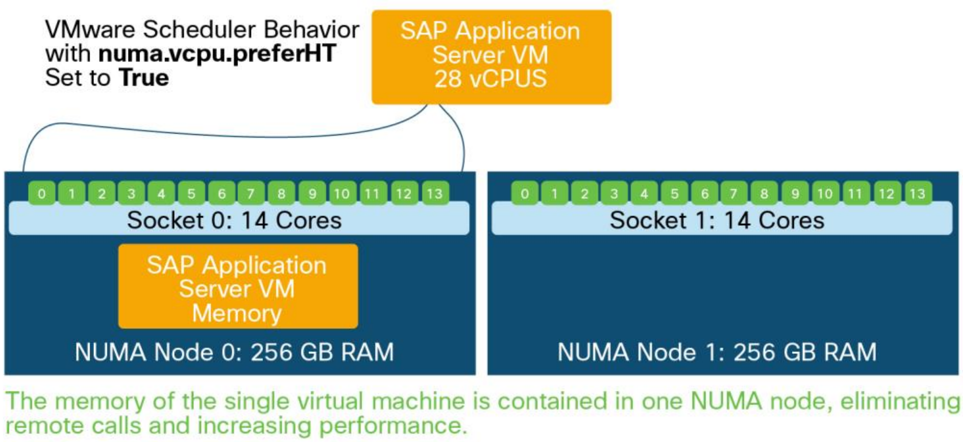
© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

Figure 3. Default CPU Scheduler



To change the CPU scheduler behavior so that the scheduler will account for the additional threads possible per core with hyperthreading, set the advanced virtual machine parameter **numa.vcpu.preferHT** to **true** to override the default scheduling behavior. Changing the parameter will force the scheduler to schedule all 28 vCPUs on the same NUMA node, therefore avoiding remote memory calls and increasing performance (Figure 4).

Figure 4. Virtual Machine Setting to Use Hyperthreading



For a virtual machine, you can configure the vCPU settings in a combination of virtual sockets and the number of virtual cores per virtual socket. You must ensure that the total number of vCPUs for an application server virtual machine does not exceed the number of logical processors in a single NUMA node. If the requirements of your SAP NetWeaver application servers do exceed the number of logical processors in a single NUMA node, refer to the section SAP HANA Virtual Machines about configuring SAP HANA virtual machine settings to deploy a wide virtual machine. Alternatively, consider choosing for the server a model of processor with a higher physical core count, or simply deploy additional SAP NetWeaver application servers to spread the load across multiple virtual machines.

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

Even with the **numa.vcpu.preferHT** setting configured as true and the total number of vCPUs set properly, it is possible to configure a virtual machine with a combination of virtual sockets and virtual cores per socket that would result in the vCPUs being scheduled across NUMA nodes. The recommended approach is to configure only one core per virtual socket, and then configure the number of virtual sockets to match the total number of vCPUs needed by the virtual machine. To help ensure that the NUMA configuration is correct, you can search the vmware.log file of the virtual machine for entries containing “numa.” The following example shows the vmware.log file of a virtual machine configured with the desired 28 vCPUs on the reference HXAF240c server, but with two virtual sockets and 14 cores per socket.

```
[root@hx240-01:/vmfs/volumes/ded01cd5-852ae5d5/NW75APP2] cat vmware.log | grep numa
2016-10-18T08:29:20.487Z| vmx| I125: DICT        numa.vcpu.preferHT = "true"
2016-10-18T08:29:20.544Z| vmx| I125: numa: Setting.vcpu.maxPerVirtualNode=14 to
match cpuid.coresPerSocket
2016-10-18T08:29:20.544Z| vmx| I125: numa: VCPU 0: VPD 0 (PPD 0)
2016-10-18T08:29:20.544Z| vmx| I125: numa: VCPU 1: VPD 0 (PPD 0)
2016-10-18T08:29:20.544Z| vmx| I125: numa: VCPU 2: VPD 0 (PPD 0)
2016-10-18T08:29:20.544Z| vmx| I125: numa: VCPU 3: VPD 0 (PPD 0)
2016-10-18T08:29:20.544Z| vmx| I125: numa: VCPU 4: VPD 0 (PPD 0)
2016-10-18T08:29:20.544Z| vmx| I125: numa: VCPU 5: VPD 0 (PPD 0)
2016-10-18T08:29:20.544Z| vmx| I125: numa: VCPU 6: VPD 0 (PPD 0)
2016-10-18T08:29:20.544Z| vmx| I125: numa: VCPU 7: VPD 0 (PPD 0)
2016-10-18T08:29:20.544Z| vmx| I125: numa: VCPU 8: VPD 0 (PPD 0)
2016-10-18T08:29:20.544Z| vmx| I125: numa: VCPU 9: VPD 0 (PPD 0)
2016-10-18T08:29:20.544Z| vmx| I125: numa: VCPU 10: VPD 0 (PPD 0)
2016-10-18T08:29:20.545Z| vmx| I125: numa: VCPU 11: VPD 0 (PPD 0)
2016-10-18T08:29:20.545Z| vmx| I125: numa: VCPU 12: VPD 0 (PPD 0)
2016-10-18T08:29:20.545Z| vmx| I125: numa: VCPU 13: VPD 0 (PPD 0)
2016-10-18T08:29:20.545Z| vmx| I125: numa: VCPU 14: VPD 1 (PPD 1)
2016-10-18T08:29:20.545Z| vmx| I125: numa: VCPU 15: VPD 1 (PPD 1)
2016-10-18T08:29:20.545Z| vmx| I125: numa: VCPU 16: VPD 1 (PPD 1)
2016-10-18T08:29:20.545Z| vmx| I125: numa: VCPU 17: VPD 1 (PPD 1)
2016-10-18T08:29:20.545Z| vmx| I125: numa: VCPU 18: VPD 1 (PPD 1)
2016-10-18T08:29:20.545Z| vmx| I125: numa: VCPU 19: VPD 1 (PPD 1)
2016-10-18T08:29:20.545Z| vmx| I125: numa: VCPU 20: VPD 1 (PPD 1)
2016-10-18T08:29:20.545Z| vmx| I125: numa: VCPU 21: VPD 1 (PPD 1)
2016-10-18T08:29:20.545Z| vmx| I125: numa: VCPU 22: VPD 1 (PPD 1)
2016-10-18T08:29:20.545Z| vmx| I125: numa: VCPU 23: VPD 1 (PPD 1)
2016-10-18T08:29:20.545Z| vmx| I125: numa: VCPU 24: VPD 1 (PPD 1)
```

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

```
2016-10-18T08:29:20.545Z| vmx| I125: numa: VCPU 25: VPD 1 (PPD 1)
2016-10-18T08:29:20.545Z| vmx| I125: numa: VCPU 26: VPD 1 (PPD 1)
2016-10-18T08:29:20.545Z| vmx| I125: numa: VCPU 27: VPD 1 (PPD 1)
2016-10-18T08:29:20.545Z| vmx| I125: numaHost: 2 virtual nodes, 2 virtual sockets, 2 physical domains
```

As can be seen in the log, the configuration resulted in 14 vCPUs being scheduled on one physical NUMA node, and the other 14 vCPUs scheduled on the second physical NUMA node, so this is not an ideal configuration. Changing the settings of this virtual machine to one core per virtual socket and 28 virtual sockets yields the following vmware.log file entries:

```
[root@hx240-01:/vmfs/volumes/ded01cd5-852ae5d5/NW75APP2] cat vmware.log | grep numa
2016-11-01T19:13:34.895Z| vmx| I125: DICT        numa.vcpu.preferHT = "true"
2016-11-01T19:13:34.954Z| vmx| A100: ConfigDB: Setting numa.autosize.vcpu.maxPerVirtualNode = "28"
2016-11-01T19:13:34.954Z| vmx| A100: ConfigDB: Setting numa.autosize.cookie = "280001"
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 0: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 1: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 2: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 3: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 4: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 5: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 6: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 7: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 8: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 9: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 10: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 11: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 12: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 13: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 14: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 15: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 16: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 17: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 18: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 19: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 20: VPD 0 (PPD 0)
```

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

```
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 21: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 22: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 23: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 24: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 25: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 26: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numa: VCPU 27: VPD 0 (PPD 0)
2016-11-01T19:13:34.959Z| vmx| I125: numaHost: 1 virtual nodes, 28 virtual sockets,
1 physical domains
```

This configuration would provide the best possible performance for the CPU and RAM for this virtual machine, because the 28 vCPUs are scheduled on a single physical NUMA node.

SAP HANA Virtual Machines

SAP HANA operates with awareness of the underlying NUMA architecture on the physical server, and it therefore attempts to consolidate memory cells to reduce the number of remote NUMA memory calls as much as it can. For the operating system in the SAP HANA virtual machine to be made aware of the physical NUMA architecture, the VMware virtual NUMA (vNUMA) feature must be enabled. vNUMA is turned on for a virtual machine by default whenever it is configured with more than eight vCPUs.

SAP HANA also benefits from being given as many CPU resources as possible. Therefore, you should configure the SAP HANA virtual machines with as many vCPUs as there are possible threads or logical processors on the host server. For example, the reference server has E5-2690 v4 CPUs that have 14 cores each and hyperthreading. Each CPU therefore has 28 logical processors, and with two sockets 56 total logical processors are available. For the best performance for the SAP HANA virtual machine on these servers, the virtual machine would be configured with 56 vCPUs. Because this number of vCPUs exceeds the size of a single physical NUMA node, these virtual machines would be considered as wide virtual machines.

Again, the recommended approach is to configure the virtual machine setting **numa.vcpu.preferHT** as **true**, and to configure only one core per virtual socket and to configure the number of virtual sockets to match the total number of vCPUs needed by the virtual machine as the initial setting. However, in the case of a wide virtual machine, which spans NUMA nodes, this setting will not expose the true physical underlying NUMA architecture to the OS running on the virtual machine. Therefore, it will not be the best-performing configuration. Following is an example of the vmware.log file of an SAP HANA virtual machine configured with 56 virtual sockets and one core per socket:

```
[root@hx240-02:/vmfs/volumes/ded01cd5-852ae5d5/HANASPS12] cat vmware.log | grep numa
2016-10-31T20:49:23.429Z| vmx| I125: DICT        numa.vcpu.preferHT = "true"
2016-10-31T20:49:23.507Z| vmx| A100: ConfigDB: Setting
numa.autosize.vcpu.maxPerVirtualNode = "28"
2016-10-31T20:49:23.507Z| vmx| A100: ConfigDB: Setting numa.autosize.cookie =
"560001"
2016-10-31T20:49:23.512Z| vmx| I125: numa: Exposing multicore topology with
cpuid.coresPerSocket = 28 is suggested for best performance
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 0: VPD 0 (PPD 0)
```

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

```
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 1: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 2: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 3: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 4: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 5: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 6: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 7: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 8: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 9: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 10: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 11: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 12: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 13: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 14: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 15: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 16: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 17: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 18: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 19: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 20: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 21: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 22: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 23: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 24: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 25: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 26: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 27: VPD 0 (PPD 0)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 28: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 29: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 30: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 31: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 32: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 33: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 34: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 35: VPD 1 (PPD 1)
```

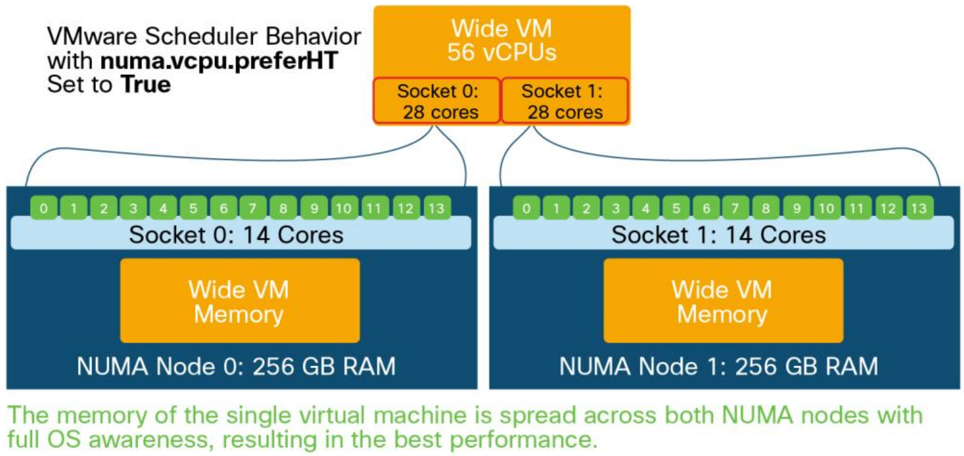
© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

```
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 36: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 37: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 38: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 39: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 40: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 41: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 42: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 43: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 44: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 45: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 46: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 47: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 48: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 49: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 50: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 51: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 52: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 53: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 54: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numa: VCPU 55: VPD 1 (PPD 1)
2016-10-31T20:49:23.512Z| vmx| I125: numaHost: 2 virtual nodes, 56 virtual sockets,
2 physical domains
```

As highlighted in the log, while the number of physical NUMA nodes in use is correct at two, ESXi recommends that the number of cores per virtual socket be changed to 28 to expose the true NUMA architecture to the OS and achieve the best performance. This change means that for the best performance of a wide virtual machine, you need to set the number of virtual sockets to match the number of physical CPU sockets in the server, and to set the number of cores per socket to match the number of logical processors in each physical NUMA node (Figure 5).

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

Figure 5. vHANA CPU Scheduling



The following example shows a vmware.log file after the virtual machine is changed to two virtual sockets with 28 cores per socket.

```
[root@hx240-02:/vmfs/volumes/ded01cd5-852ae5d5/HANASPS12] cat vmware.log | grep numa
2016-10-31T20:52:38.731Z| vmx| I125: DICT        numa.vcpu.preferHT = "true"
2016-10-31T20:52:38.731Z| vmx| I125: DICT numa.autosize.vcpu.maxPerVirtualNode =
"28"
2016-10-31T20:52:38.731Z| vmx| I125: DICT      numa.autosize.cookie = "560001"
2016-10-31T20:52:38.812Z| vmx| I125: numa: Setting.vcpu.maxPerVirtualNode=28 to
match cpuid.coresPerSocket
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 0: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 1: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 2: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 3: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 4: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 5: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 6: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 7: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 8: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 9: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 10: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 11: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 12: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 13: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 14: VPD 0 (PPD 0)
```

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

```
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 15: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 16: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 17: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 18: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 19: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 20: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 21: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 22: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 23: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 24: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 25: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 26: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 27: VPD 0 (PPD 0)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 28: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 29: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 30: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 31: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 32: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 33: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 34: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 35: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 36: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 37: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 38: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 39: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 40: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 41: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 42: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 43: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 44: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 45: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 46: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 47: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 48: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 49: VPD 1 (PPD 1)
```

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

```
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 50: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 51: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 52: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 53: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 54: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numa: VCPU 55: VPD 1 (PPD 1)
2016-10-31T20:52:38.812Z| vmx| I125: numaHost: 2 virtual nodes, 2 virtual sockets, 2 physical domains
```

Now the virtual machine has two virtual NUMA nodes, each with 28 virtual CPUs, scheduled on the two physical NUMA nodes, and there are no further recommendations from ESXi. The OS running in the virtual machine will see an accurate mapping of the virtual CPU resources that corresponds with the true underlying physical NUMA architecture.

Storage Controller Virtual Machines

The storage controller virtual machines (SCVMs) running on each Cisco HyperFlex node are each configured with a mandatory CPU reservation of 10,800 MHz (10.8 GHz). During normal operation of a Cisco HyperFlex system, the processor demands of the SCVMs are typically less than 50 percent of this value. The ESXi hosts will not perform any throttling of CPU use unless the system nears 100 percent overall CPU use. Only in those circumstances would the host set aside the 10.8 GHz of potential CPU performance, because this performance is otherwise guaranteed to the SCVMs. In most circumstances, the CPU time of the hosts is not being taxed to the limit, and therefore the reservation and CPU use of the SCVMs should have little overall impact on the performance of the virtual machines.

Memory Configuration

Memory sizing also plays a critical role in the overall sizing of an SAP system. In particular, SAP HANA is an in-memory database platforms, in which the entire contents of the database remain cached in memory while the system is operating. The SAP Quick Sizer output will provide details about the amount of memory recommended to operate the database without exhausting the system. The amount given in the example system is 235 GB, which should be considered a minimum amount for the virtual machine that will run the SAP HANA database, because additional RAM can be consumed for scratch space and query processing. The following items provide guidance in sizing the memory of the virtual machine that will run the SAP HANA database:

- Although the maximum amount of RAM that can be configured for a virtual machine in ESX 6.0 is 4 TB, the maximum amount of RAM that can be installed in an HXAF240 server is 1.5 TB.
- As noted in VMware's best practices guide for scale-up SAP HANA, all virtual machine guest memory should be reserved, and memory should not be overcommitted. Therefore, the theoretical maximum size of a vHANA virtual machine cannot exceed 1.5 TB of RAM, which is the maximum amount of physical RAM in a single Cisco HyperFlex node.
- As also noted by VMware, a portion of RAM should be set aside for the proper operation of ESXi itself. VMware recommends that approximately 3 to 4 percent be set aside. On a 1.5-TB server, that would further reduce the amount of available physical memory by up to 62 GB, reducing the maximum virtual machine RAM to 1474 GB.

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

- All Cisco HyperFlex converged storage nodes run an SCVM, which requires its own reserved virtual RAM. On the HXAF240 model server, the Cisco HyperFlex SCVM has a memory reservation of 72 GB. This further reduces the maximum virtual machine RAM for a vHANA virtual machine to about 1400 GB.

For the reference design discussed here, the HXAF240 servers have 512 GB of RAM each, so to accommodate the 235 GB of memory needed as specified by the Quick Sizer, the SAP HANA virtual machine has been built with 400 GB of RAM. This design allows plenty of space for the entire SAP HANA database to be cached in RAM. It also allows additional overhead space and leaves aside enough RAM for ESXi to operate and for the SCVM reservation.

For the NetWeaver application servers, the output of the SAP Quick Sizer provides the primary guidance regarding the amount of RAM that should be configured for the application server virtual machines. In the example here, the Quick Sizer specified only 73 GB, so for this reference design the application server virtual machines are configured with 128 GB of RAM. For the best performance, the amount of RAM configured for a single application server virtual machine should not exceed the amount of RAM in a single NUMA node. With the proper amount of memory and the proper vCPU configuration, all memory and CPU processing will take place inside a single NUMA node, helping ensure the best possible application server performance.

HyperFlex Compute-Only Nodes

Cisco HyperFlex systems can also contain what are called “compute-only” nodes, which are Cisco UCS B200 M4, C220 M4 or C240 M4 servers that are managed as part of the HyperFlex cluster, but they do not provide storage resources to the cluster as the “converged” HXAF220 or HXAF240 model servers do. Virtual machines operate on the “compute-only” nodes, using that node’s CPU and memory resources, but all storage for the VM is provided by the HyperFlex distributed filesystem running on the “converged” nodes. Because the “compute-only” nodes do not provide storage resources, they have much lower CPU and memory reservations configured for the SCVMs that run on them; only 1.4 GHz of CPU reservation and 512MB of RAM reservation. In a scenario where an SAP HANA VM must be configured as a “wide” VM, and it would need to consume almost all of the available RAM on the host, it may be advisable to run the SAP HANA VM on a “compute-only” node, in order to avoid the potential resource conflicts with the larger CPU and RAM reservations of the SCVM on the “converged” nodes.

Monitoring NUMA Memory Statistics

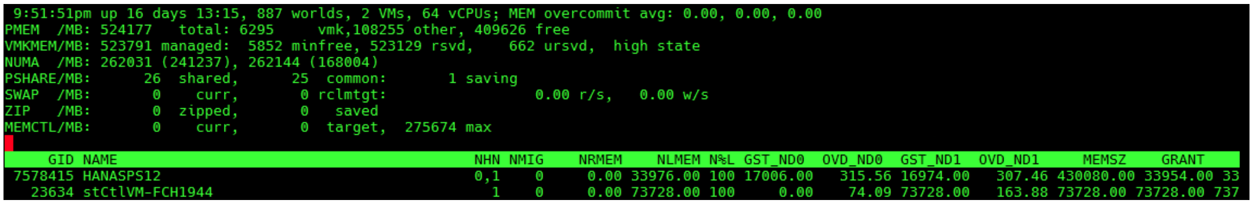
The esxtop monitoring tool on the Cisco HyperFlex hosts can be used to monitor the NUMA statistics of the virtual machines. To monitor the NUMA statistics, do the following:

1. Log on through Secure Shell (SSH) to the ESXi host that is running the virtual machine whose statistics you want to monitor.
2. Type **esxtop** and press Enter.
3. Type the letter **M**, then the letter F, and then the letter **G**. Then press Enter.
4. Monitor the statistics. When you are done, type the letter **Q**.

In the statistics, the field N%L indicates the percentage of NUMA memory that is operating local to the scheduled CPUs of the virtual machine (Figure 6). In general, this value should be 95 percent or higher, indicating that the virtual machine does not need to make many calls to a remote physical NUMA node for memory contents.

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

Figure 6. Esxtop NUMA Statistics



Storage Capacity Sizing

The overall storage capacity within a Cisco HyperFlex cluster is determined by the number of converged storage nodes in the cluster, and the number and size of the storage capacity disks in each node. The HXAF240 model server can be ordered with as few as 6 capacity disks, but can be configured with up to 10 capacity disks per node, using either 960-GB or 3.8-TB SSDs. Table 5 shows the overall cluster capacity in several scenarios.

Table 5. Cisco HyperFlex Cluster Usable Capacity

Cisco HyperFlex Cluster Capacity (tebibits TiB)			
Number of Nodes Disks per Node	4	6	8
6 (960 GB/3.8 TB)	6.4 TiB / 25.7 TiB	9.6 TiB / 38.6 TiB	12.8 TiB / 51.4 TiB
10 (960 GB/3.8 TB)	10.7 TiB / 42.8 TiB	16.1 TiB / 64.3 TiB	21.4 TiB / 85.7 TiB

Note: All capacities given at replication factor 3.

Storage Performance Considerations

As stated earlier, the HXAF240 model server is the model recommended for running a vHANA system, primarily because of its larger number of storage disks, which improves read performance potential. In the original release of Cisco HyperFlex systems, the server nodes offered a combination of a caching SSD and traditional hard disk drives (HDDs) which is known as a hybrid configuration. The caching SSD serviced all write operations along with caching recently read blocks, in order to improve performance by mitigating the frequency that data must be retrieved from the HDDs. With HyperFlex 2.0 all-flash systems, the capacity HDDs have been replaced with SSDs, and all read requests are serviced by those SSDs directly. The caching SSD no longer maintains a read cache and now only acts as a high endurance write caching device, however some read caching does occur within the RAM of the SCVM.

SAP HANA is referred to as an in-memory database, however, this term is slightly misleading. In reality, SAP HANA operates with an extremely large amount of RAM, and it caches the data in RAM until either the entire database is cached or the amount of RAM is exhausted. Upon a start or restart of the SAP HANA database, only small portions of the overall database are cached in RAM by default, along with additional tables that the SAP database administrator (DBA) or Basis administrator designates to automatically load into memory. The remainder of the data is cached in the RAM of the SAP HANA server when a client requests to read the data. Therefore, as a result of this initial load sequence, upon a start or restart of the database, the initial performance of the database will be slightly lower.

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

To optimize this initial ramp-up time, each node in the Cisco HyperFlex cluster should be configured with the maximum number of capacity disks as possible. All the data written to the Cisco HyperFlex distributed data platform is spread across all available nodes and disks in the cluster. Concurrently, when data read requests are issued, the data is read from all capacity disks throughout the cluster. If the data exists in the read cache area of the SCVM RAM, then that source will be used to provide the fastest response. If the data must be retrieved from the capacity disk layer, then all the disks in the cluster can be used to fulfill the request. Thus, read request response times can improve when more capacity disks exist in the cluster, similar to the way performance increases in any disk array when there are more members in the array.

As the SAP HANA database remains online, data which is read will be cached in the RAM of the SAP HANA virtual machine, and performance will quickly increase to its maximum potential. Data that is written to the SAP HANA database will be cached in the RAM of the SAP HANA virtual machine, and a smaller portion will be cached in the RAM of the SCVMs. Recently written data is expected to be immediately available at extremely high speeds. Data that was written or read a long time ago may have been purged from the RAM of the SAP HANA virtual machine if it experienced memory pressure. If the SAP HANA database is small enough to be cached entirely in the RAM of the SAP HANA virtual machine, then SAP HANA will generally generate only a small amount of capacity-disk-tier activity after the system has ramped up. Most capacity-disk-tier activity will occur during operations that write data to the database, because all data written will eventually be committed all the way to the capacity disk tier.

VMware Configuration and Recommendations

This section presents recommendations for configuring VMware.

Memory Reservation

SAP HANA virtual machines require full memory reservation to be configured in the settings of the virtual machine. Therefore, the physical RAM in the server cannot be overcommitted, limiting the size of SAP HANA virtual machines to the amount of RAM in the server.

Virtual Machine Placement

Given the typically large size of an SAP HANA virtual machine, often consuming the entire RAM and CPU resources of a full node, you should use VMware DRS anti-affinity rules to keep the SAP HANA virtual machine separate from the other virtual machines in the cluster. Normally, DRS would recognize the large size of an SAP HANA virtual machine and place it on a host by itself, but the use of DRS rules can help ensure that this placement always occurs. This action helps ensure that a wide SAP HANA virtual machine does not experience any contention for CPU or memory resources. A DRS rule can also help reduce the potential for loss, due to a single host failure, of both the primary and secondary SAP HANA database virtual machines that are performing replication using the SAP HANA system replication feature. Similarly, DRS anti-affinity rules should be used to prevent large SAP Netweaver virtual machines from conflicting with each other, or other VMs that have large resource requirements.

Multiple SAP HANA Virtual Machines

If a use case requires multiple SAP HANA virtual machines, the placement of those virtual machines should be considered to provide each virtual machine with the best possible performance. For the best performance of SAP HANA virtual machines, follow these guidelines:

- For wide SAP HANA virtual machines, which span multiple physical NUMA nodes, create a DRS anti-affinity rule to help ensure that the SAP HANA virtual machines stay separated from one another and do not attempt to use the same CPU resources on the host server.

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

- Smaller SAP HANA virtual machines can be co-located on a single physical server. However, for the best performance, you should prevent multiple SAP HANA virtual machines from competing for the same physical CPU resources. As with SAP application servers, a safe minimum guideline is to size smaller SAP HANA virtual machines with as many vCPUs as there are logical processors in a single physical CPU. This sizing would contain all the CPU and memory resources of a single SAP HANA virtual machine within a single physical NUMA node, leaving the second physical NUMA node available for another SAP HANA virtual machine to operate.

SAP System Monitoring

The SAP Host Agent, a software component that is installed by default on all SAP HANA and NetWeaver application servers, can also view extended system information about the virtualization hosts from within the virtual machines. This feature is called SAP System Monitoring, and is detailed in SAP note 1606643. SAP System Monitoring is a mandatory feature for SAP support in virtualized environments. To enable the feature, configure the following advanced settings in the vSphere client or vSphere Web Client:

- Verify that the virtual machines are running VMware Tools.

  Note: If generic Linux open-vm-tools are used, some libraries have different names. This issue can be fixed by entering the following Linux command to link the correct libraries:

  ```
  ln -s /usr/lib64/libguestlib.so.0 /usr/lib64/libvmGuestLib.so
  ```

- On each ESXi host, change the advanced setting **Misc.GuestLibAllowHostInfo** to **1**.
- On each SAP virtual machine, insert the advanced setting tools.guestlib.enableHostInfo set to true.

Additional Settings

On the SAP HANA virtual machines, additionally set the following virtual machine–level advanced settings:

- Set **halt_in_monitor = true**.
- Set **idleLoopSpinBeforeHalt = true**.
- Contrary to guidance from VMware which is now outdated, the latency sensitivity setting should be left at **Normal**.

Cisco HyperFlex Configuration and Recommendations

For the best level of redundancy for the data in the Cisco HyperFlex distributed file system, you should install the Cisco HyperFlex cluster using a replication factor of 3. This setting creates three copies of all data written to the cluster and allows simultaneous failures of multiple components throughout the cluster, even failure of entire nodes, without data loss. One of the most frequently failed components of a server or storage array is the legacy spinning hard drive (HDD). Studies have shown that solid state disks (SSDs) are anywhere from 2-10 times as reliable as HDDs on an annual replacement basis. Due to this, configuring the HyperFlex cluster to use replication factor 2 is a much more viable option in all-flash configurations. Because the focus of this paper is on non-production landscapes, it is worth the consideration to use RF2 if the extra redundancy is not considered to be necessary, because RF2 does benefit from 15-40% reduced write latency performance, plus it increases the usable capacity of the HyperFlex cluster.

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

As an option during the installation of a Cisco HyperFlex cluster, you can optimize the cluster for a virtual desktop infrastructure (VDI) deployment. However, this feature modifies the caching amounts and behaviors of the cluster and is not ideal for an environment with virtualized server application workloads. Therefore, optimizing Cisco HyperFlex for VDI deployment is not recommended.

Conclusion

This Cisco solution addresses the needs of many companies that operate SAP in their data centers. It presents the unique considerations and recommendations for running an SAP landscape in a Cisco HyperFlex cluster while staying aligned with established practices. By following the guidelines in this solution, your SAP landscape can achieve its best possible performance in a Cisco HyperFlex system and take advantage of the ease of use and rapid deployment provided by the Cisco HyperFlex solution.

For More Information

The following documents, available online, provide additional relevant information. If you do not have access to a document, please contact your Cisco representative.

- Cisco HyperFlex technical support documentation
- VMware Best Practices and Recommendation for Scale-Up Deployments of SAP HANA on VMware vSphere
- SAP on VMware Best Practices
- Performance Best Practices for VMware vSphere 6.0
- SAP Certified SAP HANA Hardware Directory—Entry-level systems
- SAP on VMware vSphere wiki
- SAP HANA on VMware vSphere wiki
- SAP HANA TDI FAQ
- SAP HANA Help Portal

SAP Notes

See the SAP Notes in Table 6 for the latest up-to-date information about SAP HANA settings and requirements and OS-specific settings to optimize the system. A valid logon with SAP support is required to view the content of the notes.

Table 6. SAP Notes

Note Number	Description
2240716	Recommended OS settings for SUSE Linux Enterprise Server (SLES) 11 and SLES for SAP Applications 11 SP4
2247020	Recommended OS settings for Red Hat Enterprise Linux (RHEL) 6.7
2205917	Recommended OS settings for SLES 12 and SLES for SAP Applications 12
2292690	Recommended OS settings for RHEL 7.2
2235581	SAP HANA: Supported operating systems
1492000	General support statement for virtual environments
1788665	SAP HANA support for virtualized and partitioned (multitenant) environments
2315348	Single SAP HANA virtual machine on VMware vSphere 6 in production
2024433	Multiple SAP HANA virtual machines on VMware vSphere in production

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com/go/trademarks. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)

Printed in USA

C11-738385-01 04/17

© 2017 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public.