ESXi
Tuning Guide
Eli-v6.5.391

Bridgeworks
Unit 1, Aero Centre, Ampress Lane,
Ampress Park, Lymington,
Hampshire SO41 8QF
Tel: +44 (0) 1590 615 444
Email: support@4bridgeworks.com

Table of Contents

1 Introduction **3**

1.1 When to use this guide . . . 3

1.2 When not to use this guide . . . 3

1.3 Further reading . . . 4

2 Summary **5**

2.1 CPU Core Allocation: . . . 5

2.2 Core Affinity: . . . 5

2.3 Memory Allocation: . . . 5

2.4 Memory (NUMA - Non-Uniform Memory Access) Alignment: . . . 6

3 Requirements **7**

3.1 ESXi Tools . . . 7

3.1.1 ESXi built-in Web Client . . . 7

3.1.1.1 Accessing ESXi built-in Web Client . . . 7

3.1.2 ESXi vSphere Web Client . . . 9

3.1.2.1 Accessing ESXi vSphere Web Client . . . 9

4 CPU Allocation **12**

4.1 Adjusting Core Allocation . . . 12

4.1.1 ESXi Built-in Web Client . . . 12

4.1.2 vSphere Web Client . . . 14

5 Core Affinity **18**

5.1 Assessing Viability . . . 18

5.1.1 ESXi built-in Web Client . . . 18

5.1.2 vSphere Web Client . . . 20

5.2 Applying Affinity . . . 22

5.2.1 ESXi built-in Web Client . . . 22

5.2.2 vSphere Web Client . . . 23

6 Memory Allocation **26**

6.1 Adjusting Allocation 27

6.1.1 ESXi Built-in Web Client 27

6.1.2 vSphere Web Client 28

7 Memory Alignment **29**

8 Useful Links **30**

1 Introduction

This document details the prerequisites for tweaking a deployed virtual Bridgeworks Node to improve resource use and maintain peak performance.

The suggested modifications included in this guide are given as a best practice approach to reduce the chance of having a sub-optimal virtual installation.

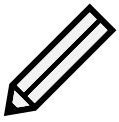
Most of the optimisation is targeted at ensuring that the virtual Bridgeworks Node can keep its CPU cores and RAM access aligned. Without alignment the RAM speeds can suffer, which can affect acceleration performance.

1.1 When to use this guide

Generally ESXi tuning is not required. The system does a good job of moving resources around to keep everything aligned and provided with enough memory and CPU time.

Key symptoms that could be improved by tuning the ESXi guest settings are:

- Reduced performance, these can be:
  - Reduced throughput. If it is proven that the Bridgeworks Node is not reaching the full speed of the network connection it has been given.
  - Inconsistent throughput. If the graphing in the *Transfer Statistics* page show that throughput is not consistent during a full bandwidth data a transfer in your typical network environment, and no other environmental variables are causing this.
  - Very high CPU usage. If the System info page is showing more than 80% CPU utilisation during a full bandwidth data transfer in your typical network environment.
  - Very high memory usage. If the System info page is showing more than 80% memory utilisation at any time.
- Conserving ESXi resources when performance is not being reduced*. Such as:
  - Very low CPU usage.
    If the System info page is showing less than 25% CPU utilisation during a full bandwidth data transfer in your typical network environment.
  - Very low memory usage.
    If the System info page is showing less than 25% memory utilisation during a full bandwidth data transfer in your typical network environment.



Note: *Reducing resources to the Bridgeworks Node should always take into account any resource headroom required to sustain performance if network quality deteriorates. See Chapter 4: CPU Allocation and Chapter 6: Memory Allocation sections for minimum supported resource settings.

1.2 When not to use this guide

Other factors can result in lower utilisation of network connection between Bridgework Nodes.

Possible other factors include:

- Onload/Offload storage I/O

  If the systems on either side of the Bridgeworks Nodes cannot provide or receive the data quickly enough a bottleneck will occur and the Nodes will throttle the connection appropriately.

- Shared network resources

  If either Bridgeworks Node shares the network connection with other systems then competition for network time can reduce throughput depending on the demand for the network bandwidth.

- Product bandwidth limits Each series of Bridgeworks node has a different network bandwidth capacity.

  - Series 100 = 125MB/s.
  - Series 200 = 250MB/s.
  - Series 400 = 1GB up to 10GBe dependant on hardware specification.

1.3 Further reading

For more detailed information regarding ESXi tuning, see VMwares own ”Performance Best Practices” documentation for your relevant version of ESXi.

- ESXi 7.0 Best Practices (2020)
- ESXi 8.0 Best Practices (2022)

2 Summary

2.1 CPU Core Allocation:

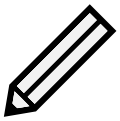
It is beneficial to keep core allocation conservative. Increasing the number of assigned cores will improve performance only when the existing setup is consistently experiencing CPU usage above 75%. The OVAs for Bridgeworks Nodes have a preset core allocation that is set to accommodate the throughput intended for that model. Scenarios that could result in needing more CPU power are; running encryption on top of full network speeds, or having a very large packet loss environment. Conversely, if the Node never sees high core usage, then reducing the allocation may help ensure that the host never places the Bridgeworks Node into a sub-optimal state.

In ESXi the core allocation is dictated by vCPUs. These cores are also presented in a settable number of sockets. This setting is for operating system compatibility only and does not provide a performance increase. The Bridgeworks Node should be given all vCPUs as individual sockets. Running the system with a 'multiple cores per socket' setup can have adverse effects on its performance.

2.2 Core Affinity:

Core affinity can have an immediate effect on acceleration performance, this is evident when the core affinity forces execution to be spread across the physical sockets of the host. In ESXi, core affinities can be set on a per guest basis. The only requirement is that the affinity gives at least the number of vCPUs given to the guest. The recommended practice is to give an affinity setting that ensures all execution of the Bridgeworks Node will happen on the same physical socket, this ensures that all memory access will be using the same caching and RAM modules.

In addition, when using a processor feature such as Intel Hyperthreading the underlying hardware will present 2 CPUs per physical processor core. ESXi will handle these as *logical cores*, which results in the ESXi scheduler load balancing the 2 logical cores across the physical core evenly*. This load balancing means that the effective performance of 1 logical core can be less than the actual performance of the physical core it is running on.



Note: *When ESXi CPU scheduling results in only 1 of the 2 available logical cores on a given physical core being utilised, the guest running on that logical core is given the full performance of the physical core. Logical core performance is reduced when both logical cores on the same physical core are being utilised by guests, at which time the performance of the physical core is split equally between the two logical cores presented by ESXi.

2.3 Memory Allocation:

Just like vCPUs, assigning the minimum necessary memory will reduce any chance of a sub-optimal setup. A smaller RAM footprint makes it a lot easier for the ESXi host to ensure that the memory is aligned to the same socket that the vCPUs reside on.

2.4 Memory (NUMA - Non-Uniform Memory Access) Alignment:

ESXi can split memory into NUMA pools, this is due to separate RAM access across separate physical sockets on a motherboard. In a worst case scenario a virtual guest could end up with all of its CPU resource being scheduled on physical CPU cores that are not directly attached to the RAM that the guest is using. Bandwidth testing has shown that memory performance can degrade significantly, slowdowns of over 20% were frequently seen. In addition, running a badly aligned Bridgeworks Node will result in a decrease to the total throughput attained on a like-for-like network environment. Memory alignment through NUMA affinity is similar to running CPU affinity. Manipulating CPU affinity is preferred as it can be adjusted on a live system.

3 Requirements

3.1 ESXi Tools

There are several methods for configuring virtual machines in an ESXi system:

- Basic settings via a built-in web client.
- More advanced settings using the ESXi vCenter Server system and its vSphere web client.

The following sections list what each method can and cannot achieve with regards to the tuning solutions included in this guide.

3.1.1 ESXi built-in Web Client

ESXi installation software deploys a web based control GUI automatically.

This method can control:

- Core allocation and socket/core distribution.
- RAM amount and priority/reservation values.
- Core affinity.

This method cannot control:

- RAM NUMA affinity.

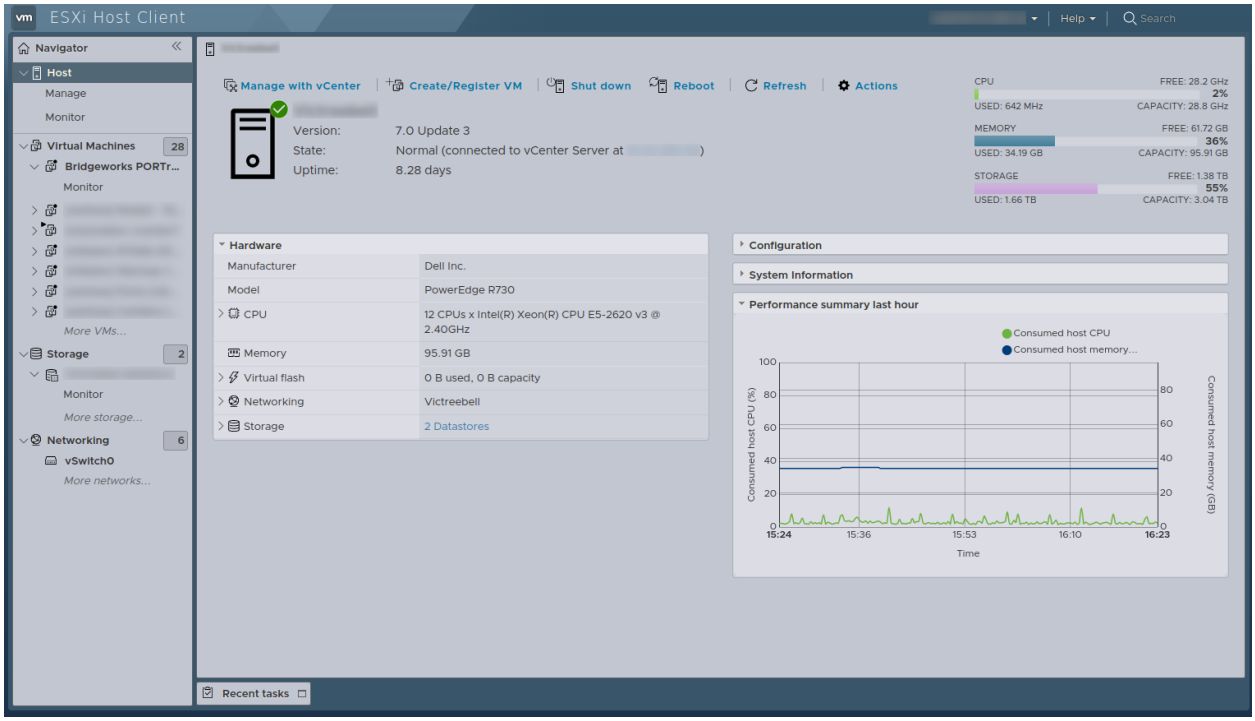
3.1.1.1 Accessing ESXi built-in Web Client

To connect to the built-in web client, open your browser of choice and navigate to the IP address or website name of the ESXi host machine.

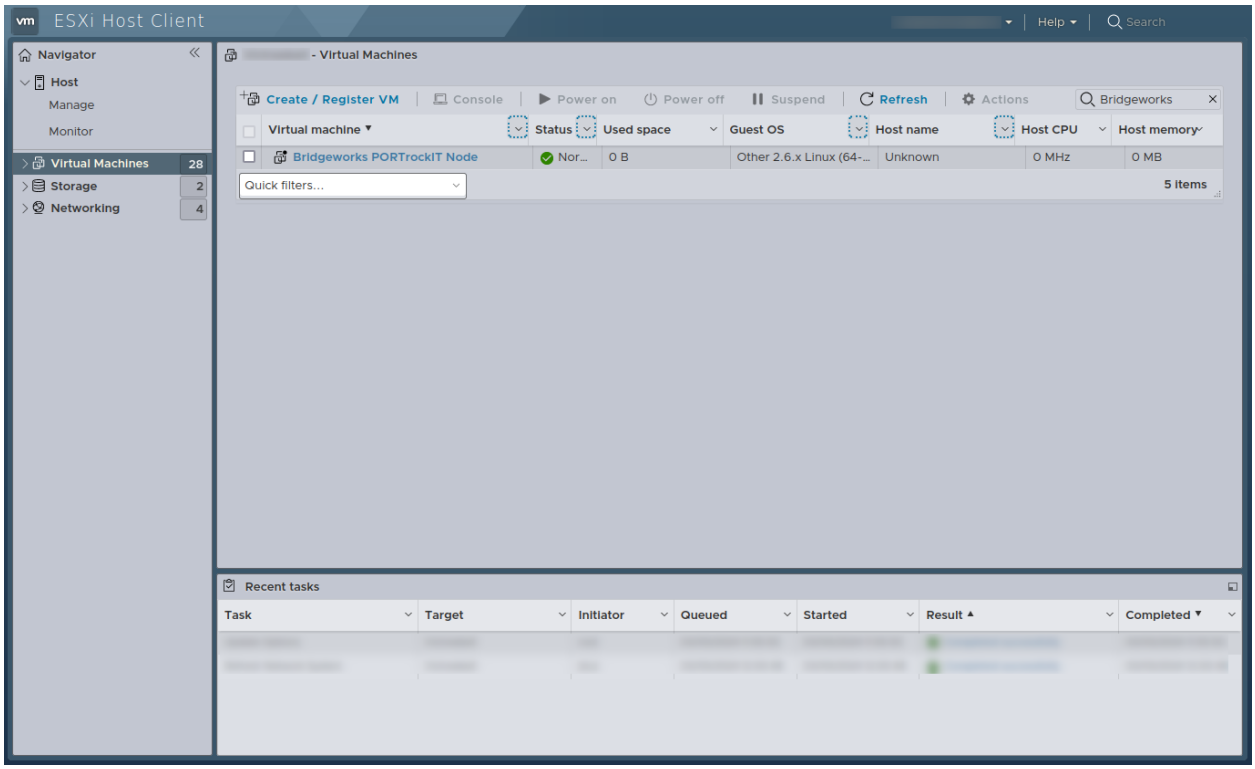
This should present a login screen.

Enter the correct user and password and click ”Log in”.

At this stage the web client will show the host machine information page.



In the left hand pane there is a list of options. Click the ”Virtual Machines” entry. For tuning, only the Bridgeworks Node will be used.



3.1.2 ESXi vSphere Web Client

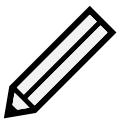
Optionally, a ”vCenter Server” system can be deployed and connected to your ESXi host. This allows far greater control than the basic ESXi web client.

This method can control:

- Core allocation and socket/core distribution.
- RAM amount and priority/reservation values.
- Core affinity.

This method cannot control:

- RAM NUMA affinity.*

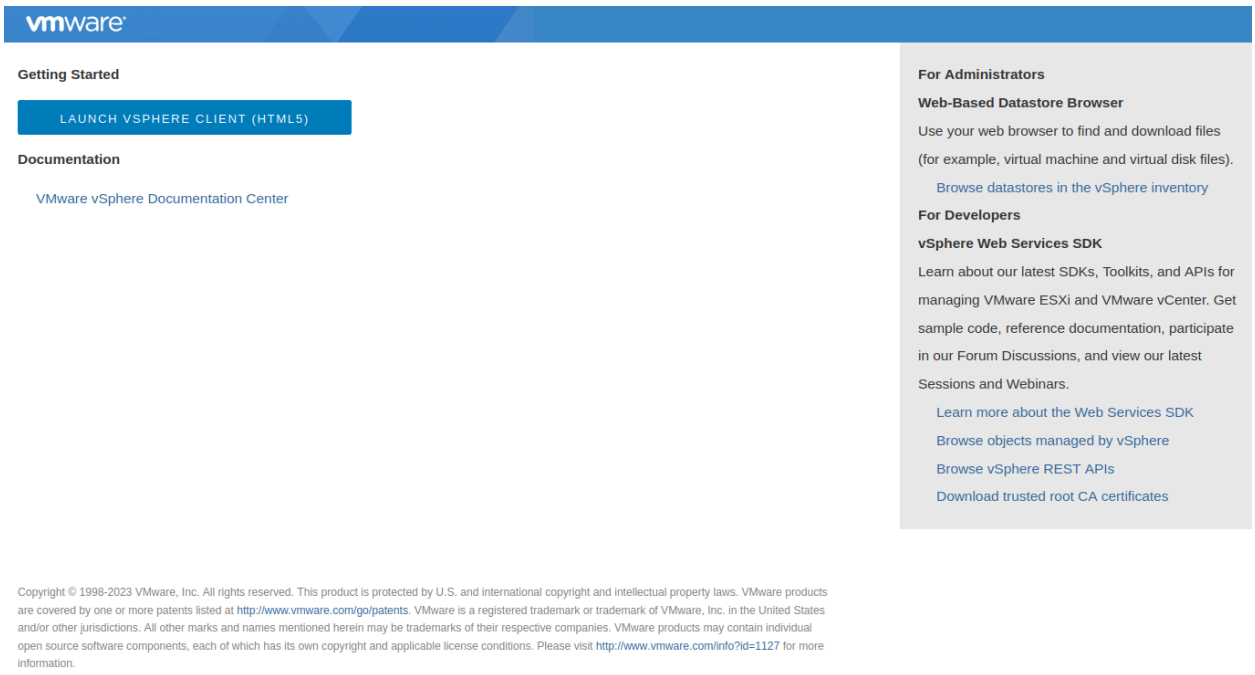
	Note: *NUMA can be set by using custom parameters options, this is outside the scope of this document.

3.1.2.1 Accessing ESXi vSphere Web Client

To connect to the web client, open your browser of choice and navigate to the IP address or website name of the vCenter Server.

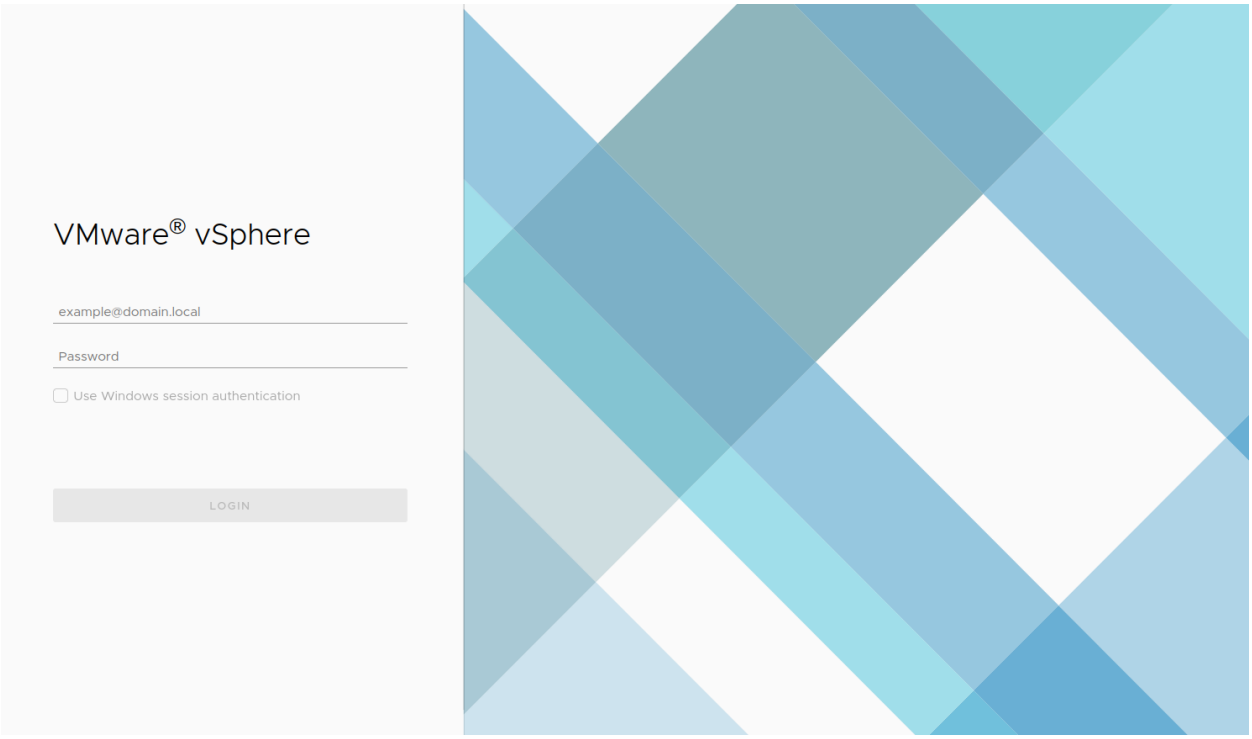
If you do not know the address for the vCenter Server you may be able to get it from the Host info page on the ESXi built-in web client. See above for instructions on accessing it. On the Host info page presented at login, look for ”State” in the main pane, to the right there should be a status, along with notification of ”connected to vCenter Server” and an IP address or website URL to the vCenter Server.

Upon accessing the vCenter Server there should be a page listing the access.

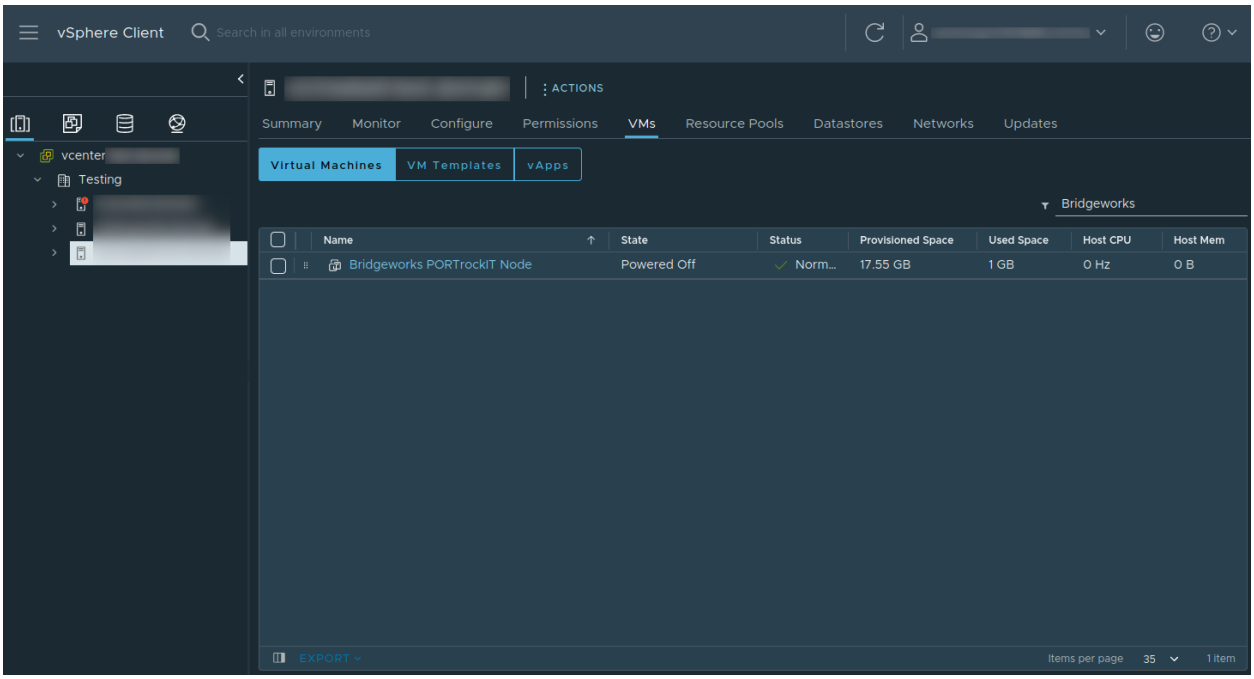


Select the ”Launch vSphere Client” link.

This should present a login screen. Enter the correct user and password and click ”Log in”.



In the left hand pane there should be an entry for the target ESXi host. By clicking on it, then on VMs and searching for ”Bridgeworks”, it should reveal the target Bridgeworks Node.



For tuning, only the target Bridgeworks Node will be used.

4 CPU Allocation

ESXi allows provisioning of a huge number of vCPUs to a virtual system. Across multiple virtual guests, this core allocation can exceed the physical CPU count of the host.

In addition, vCPUs can be presented to a virtual guest's operating system as either multiple sockets, or as a multiple-cores-per-socket system. Generally the user should always choose multiple sockets, with multiple cores being a compatibility setting for operating systems that do not support multiple sockets.

Initial CPU cores for each system are:

- 100 Series = 2
- 200 Series = 3
- 400 Series = 4

These values can be changed to suit the individual applications. The number of cores can be reduced if core utilisation is consistently low, though less than 2 cores is not supported. Likewise CPU allocation can be increased if CPU utilisation is consistently high, this scenario could be due to overheads from excessively detrimental network conditions or enabling features that require CPU time, such as IPSec.

In most cases no more than 8 cores should be needed, with core clock speed being more important for performance than further increasing core allocation.

Core allocation should always try to be no more than that of the underlying hardware. Once vCPUs start sharing the resources of a hardware core, the scheduling for CPU resources can result in high latency and reduced performance. If an ESXi host needs to have oversaturation of the CPU cores, then consider using CPU affinity to ensure that the Bridgeworks Node has cores reserved. This ensures that network acceleration is not affected by other virtual guests competing for CPU time.

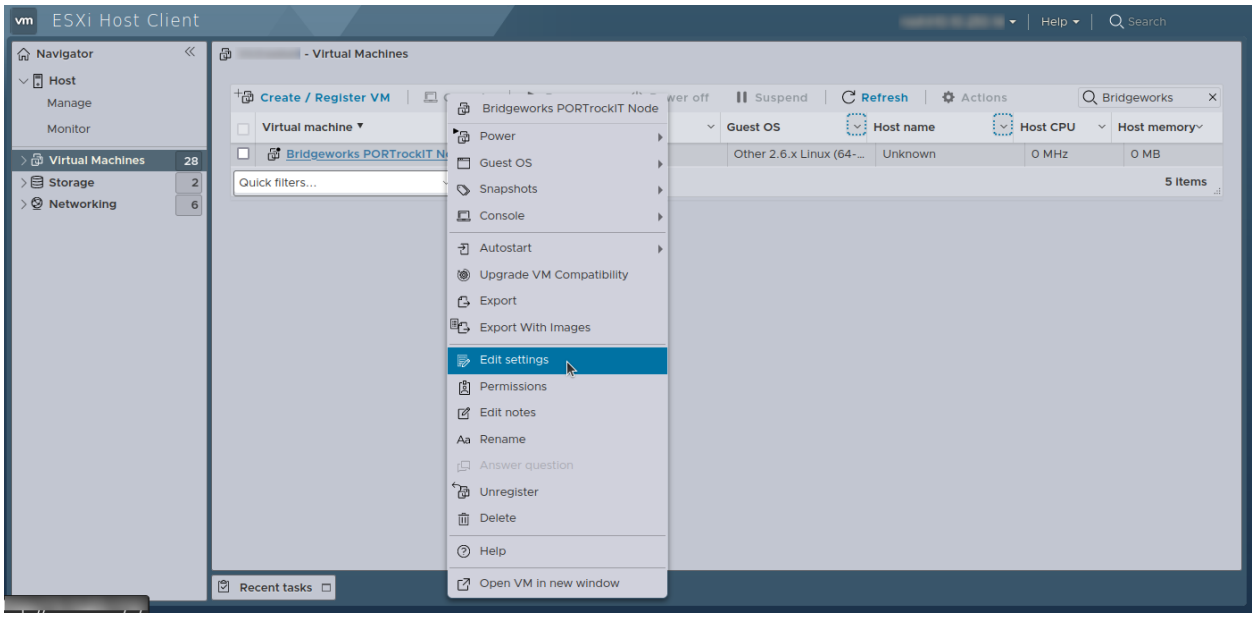
An ideal setup would allow the Bridgeworks Node to run entirely on a single physical socket, so when possible, the core allocation should be no more than the core count of one of the hosts CPU sockets.

When assessing core allocation requirements it is important to include the performance deficit from running on logical cores. This is not to be confused with the vCPUs terminology used in ESXi. On the Intel platform this logical core feature is known as ”Intel Hyper-Threading Technology”, this can be enabled/disabled in BIOS settings, and will double the detected cores of the underlying hardware. A logical core will not perform as well as a physical core; e.g. a dual core guest running on two logical (hyper-threaded) cores will not perform as well as a dual core guest running on two physical cores.

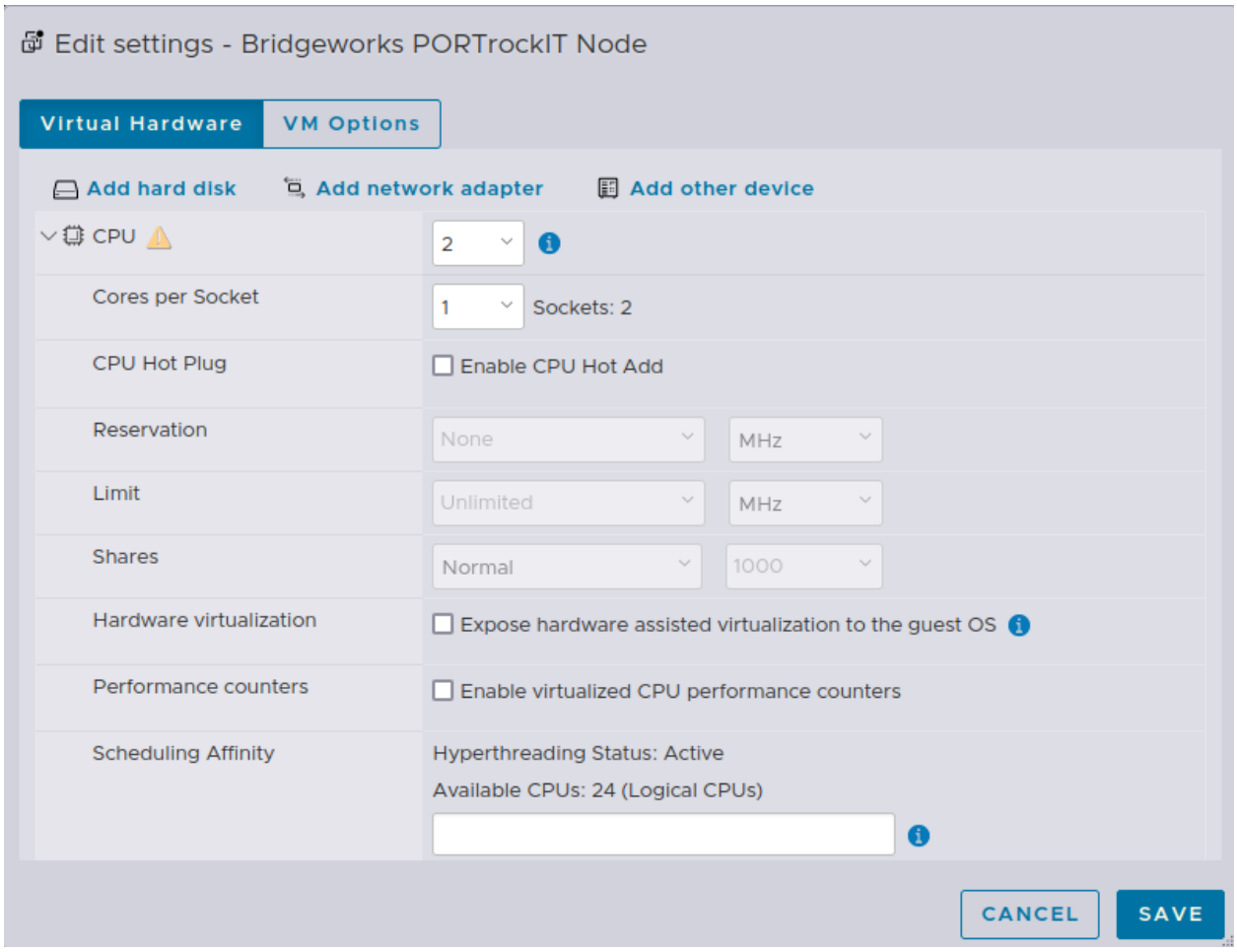
4.1 Adjusting Core Allocation

4.1.1 ESXi Built-in Web Client

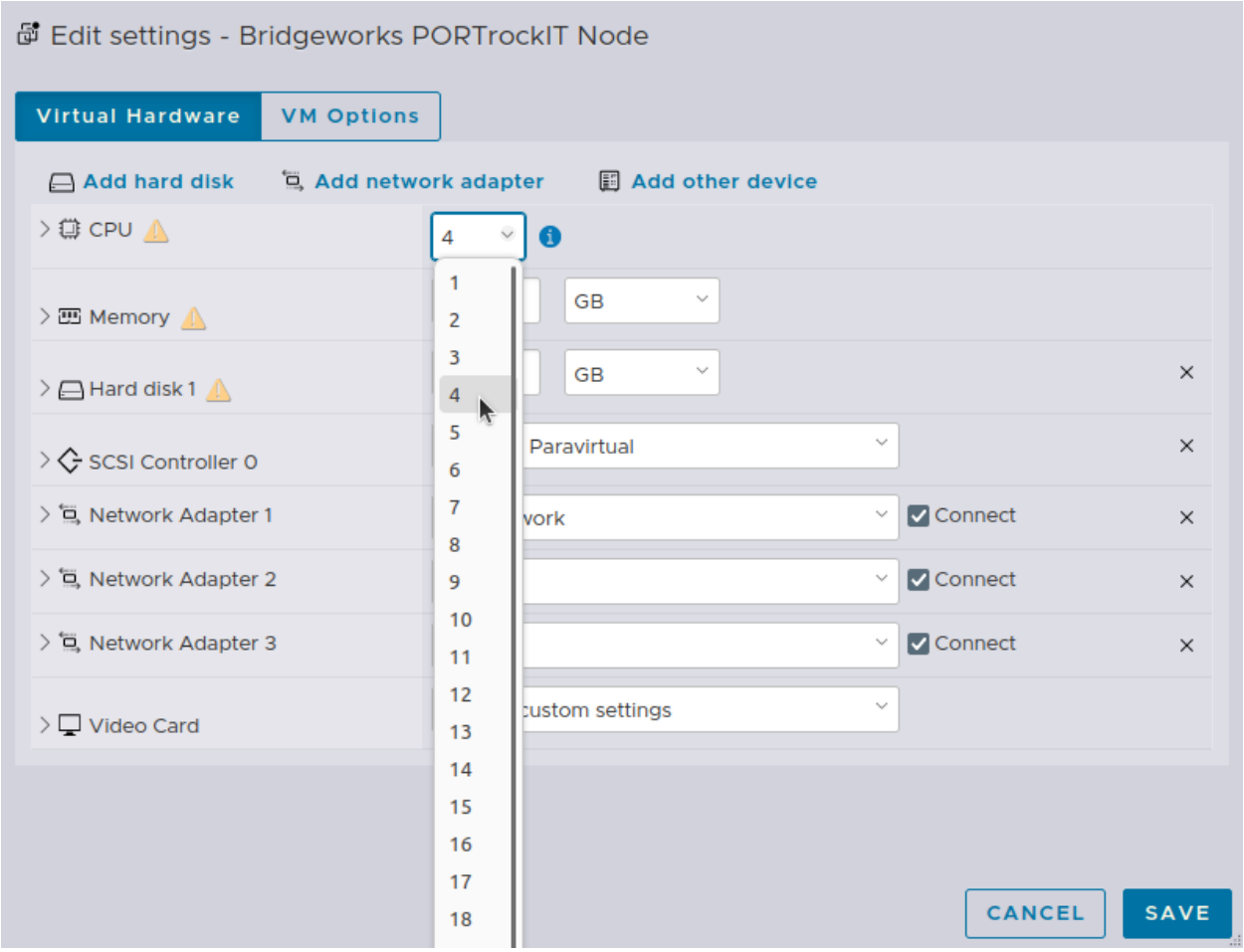
After logging in to the Web Client, navigate to the *Virtual Machines* page and bring up the *Edit settings* pane for the Bridgeworks Node.



In this section, click the arrow left of the *CPU* entry to expand the settings pane and ensure that the *Cores per Socket* is set to 1.



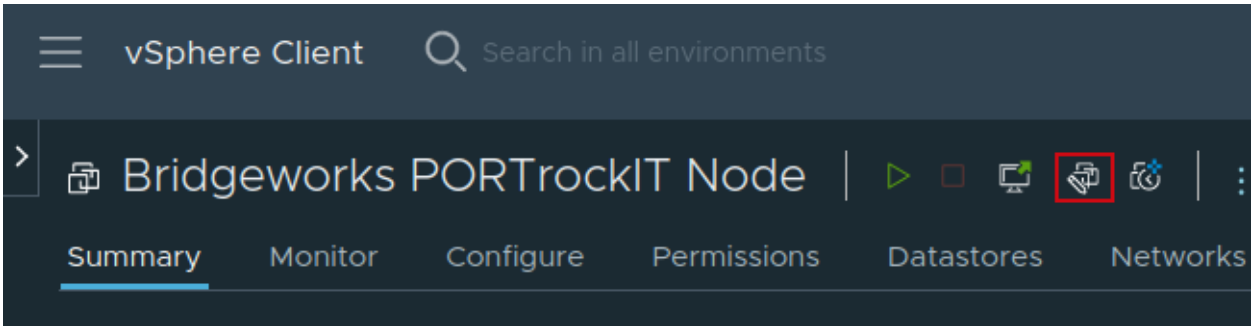
Click the number right of *CPU* and set the desired number of cores.



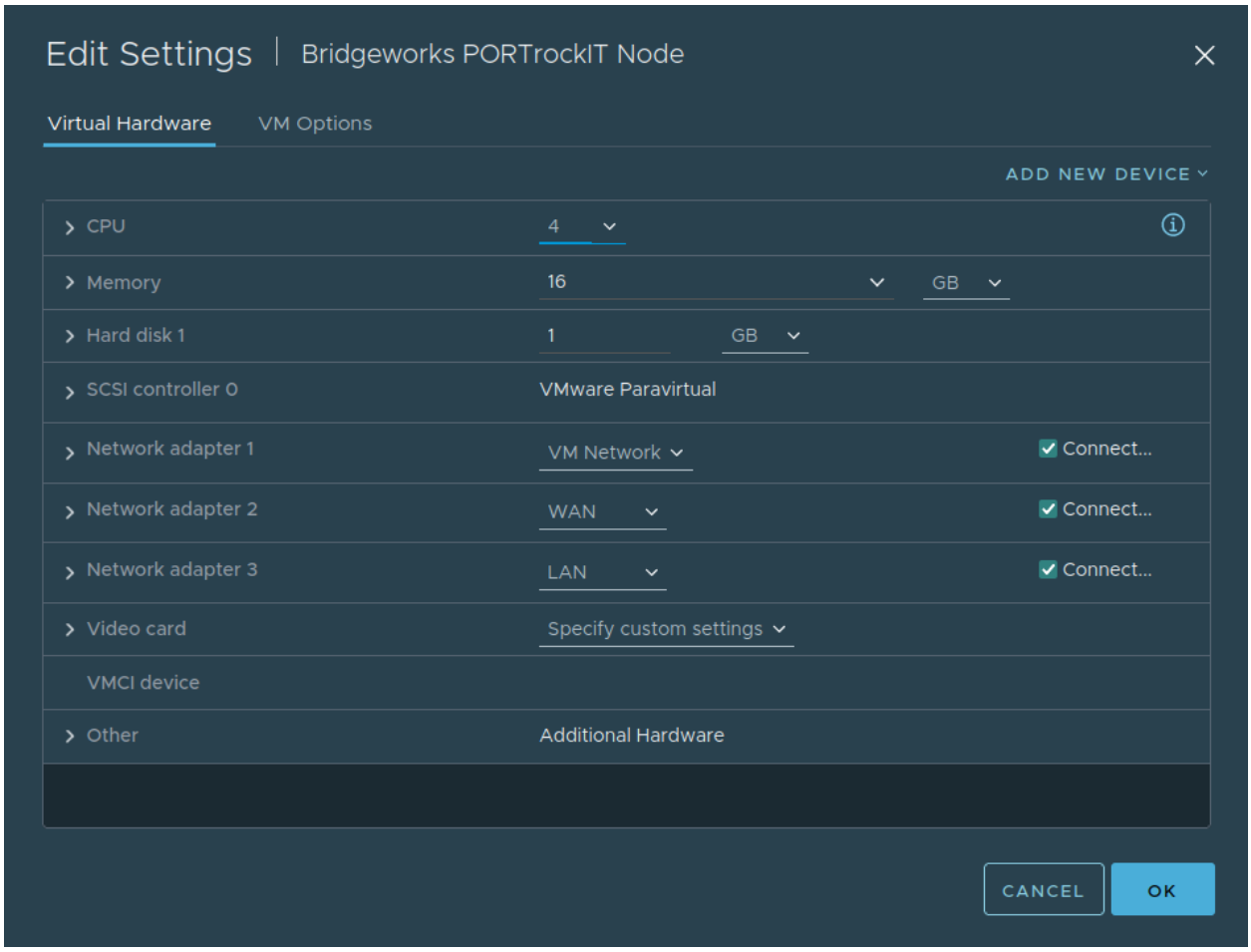
Click *Save* at the bottom of the pane.

4.1.2 vSphere Web Client

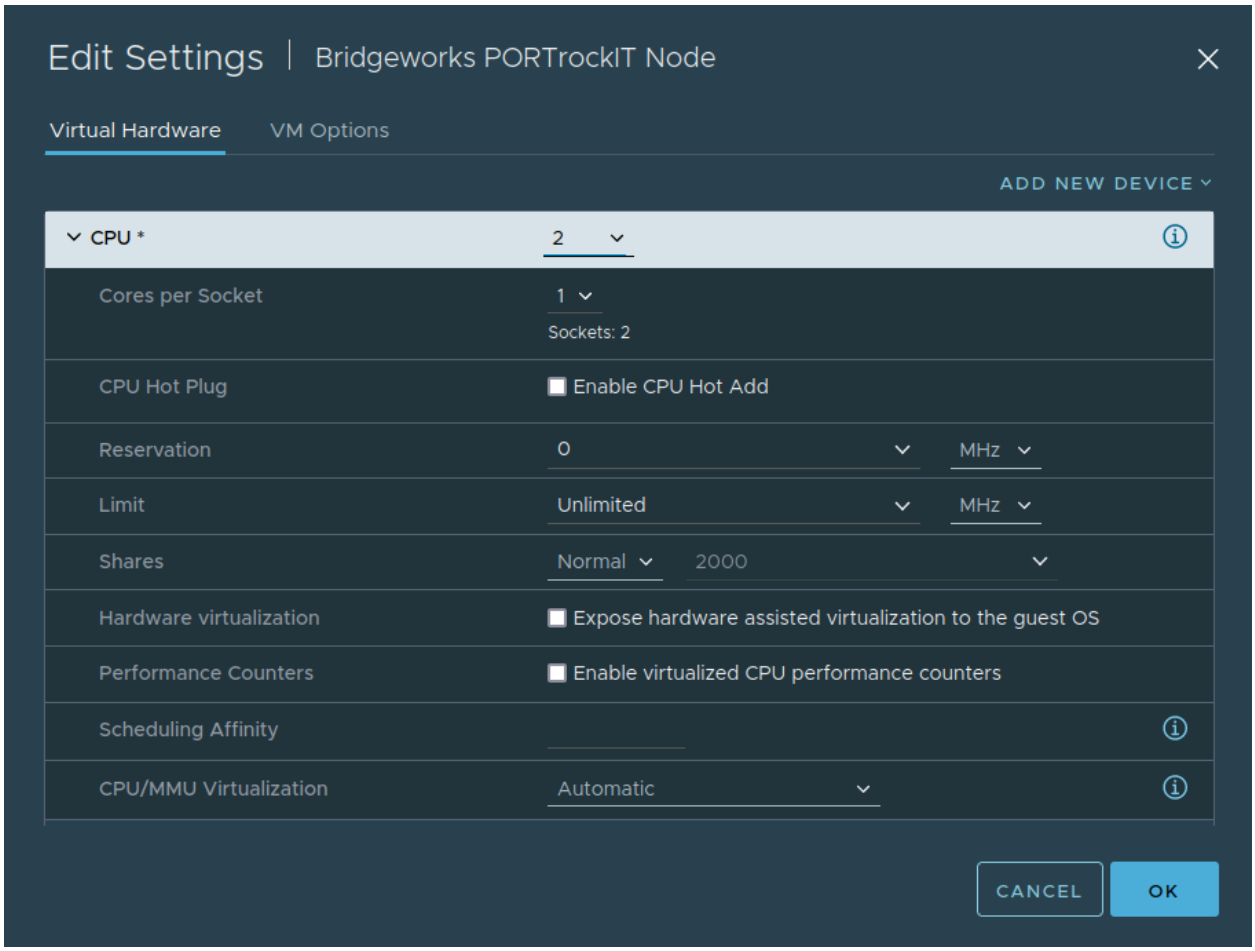
After logging in to the Web Client, navigate to the Bridgeworks Node and access the *Edit Settings* page by clicking on the Bridgeworks Node edit settings icon shown in the next image *Edit Settings*.



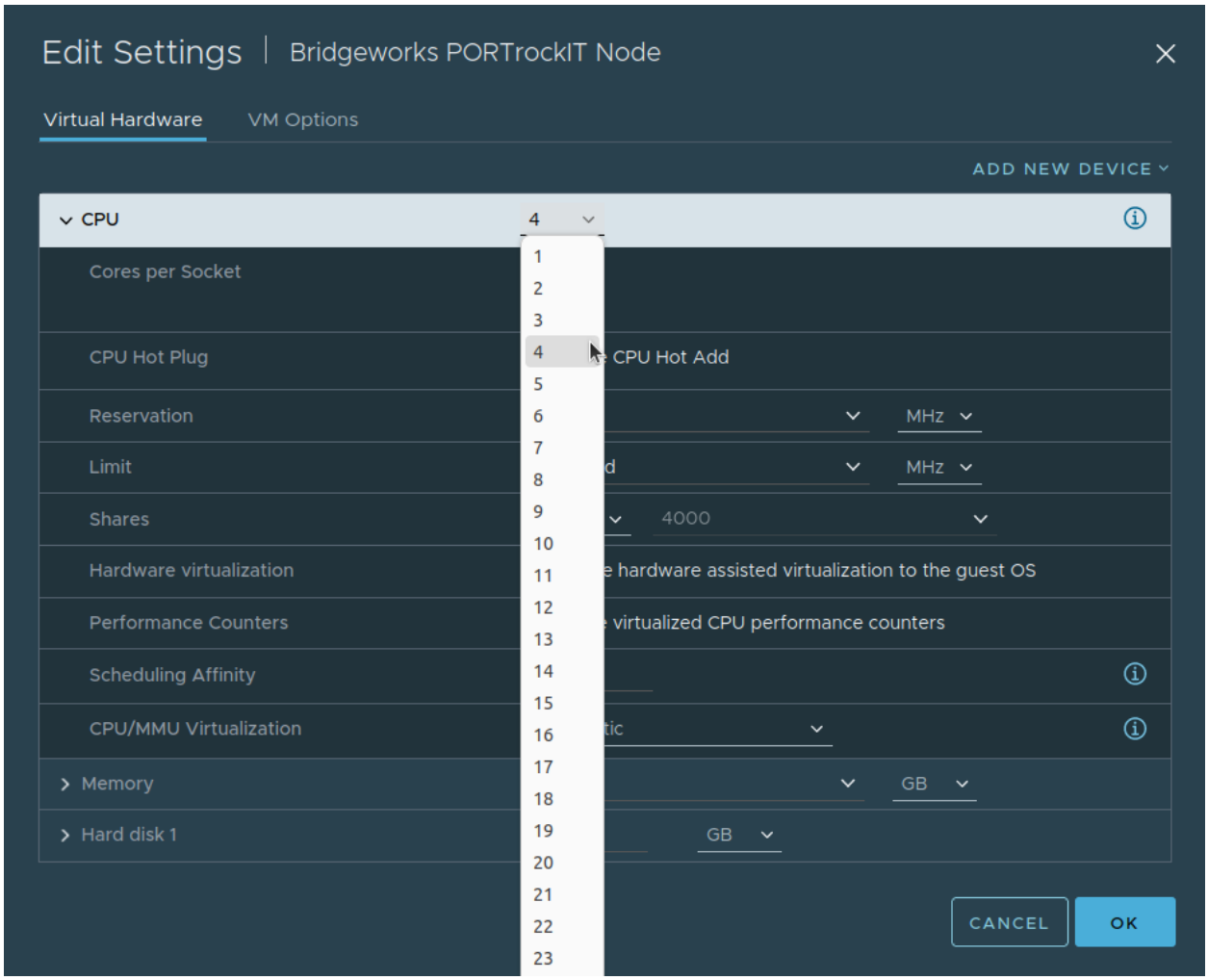
Expand the *CPU* section by clicking the arrow left of the entry.



Ensure that the *Core per Socket* entry is set to 1.



Click the number right of *CPU* and select the desired core count from the dropdown selection.



Click *OK* at the bottom of the pane.

5 Core Affinity

Affinity settings can be changed without having to bring down the guest system, so this can be applied during operation for immediate effect.

ESXi attempts to manage its guests to ensure that competition for any given core is reduced to a minimum, along with arranging memory resources to try to keep guest systems in an aligned setup. An aligned state is when the resources being used by the virtual guest are all locally addressable. For example, memory is allocated on RAM local to the CPU cores that the guest vCPUs are scheduled on. Generally ESXi can keep everything running well.

Issues can arise when resources on the host are overcommitted; multiple guest systems may queue for scheduling on CPU cores, or the ESXi host may have to re-arrange memory usage when host memory availability is running low. The result of resource overcommitment is that the memory allocation may fall over to RAM that is not local to the active vCPUs.

The Bridgeworks Node may suffer when physical CPUs become overcommitted, possibly resulting in reduced network acceleration capabilities.

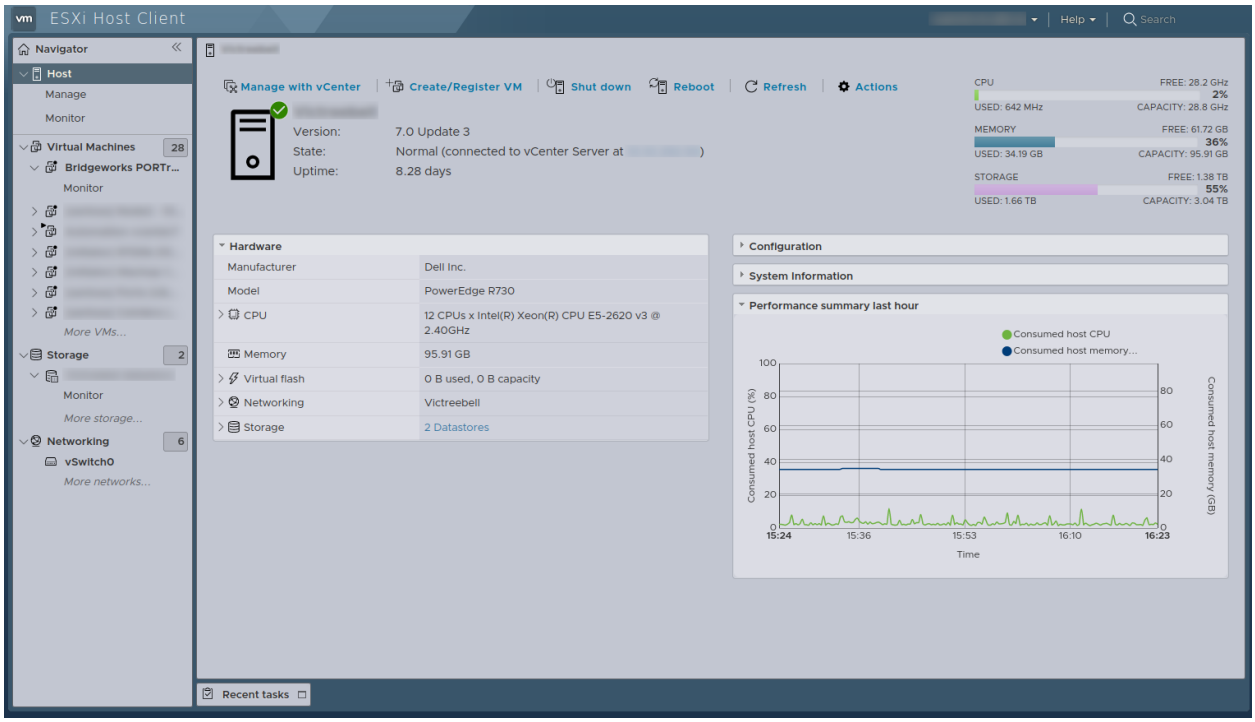
If there is the possibility of a CPU overcommit, or execution getting moved over to a sub-optimal socket, then giving the Bridgeworks Node a dedicated affinity will help to mitigate slowdowns. Increasing reserve priority will also help if CPU cores cannot be 100% dedicated to the Bridgeworks Node.

5.1 Assessing Viability

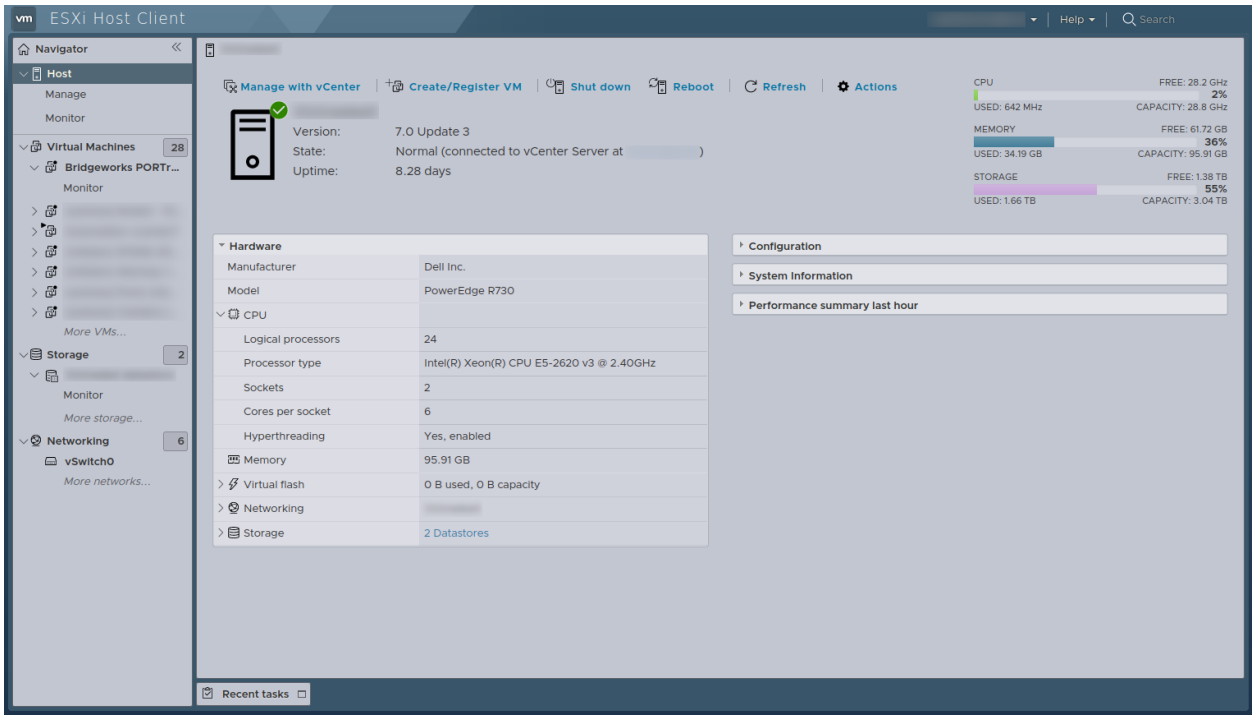
To setup an optimal CPU Affinity some information regarding the ESXi host hardware is required.

5.1.1 ESXi built-in Web Client

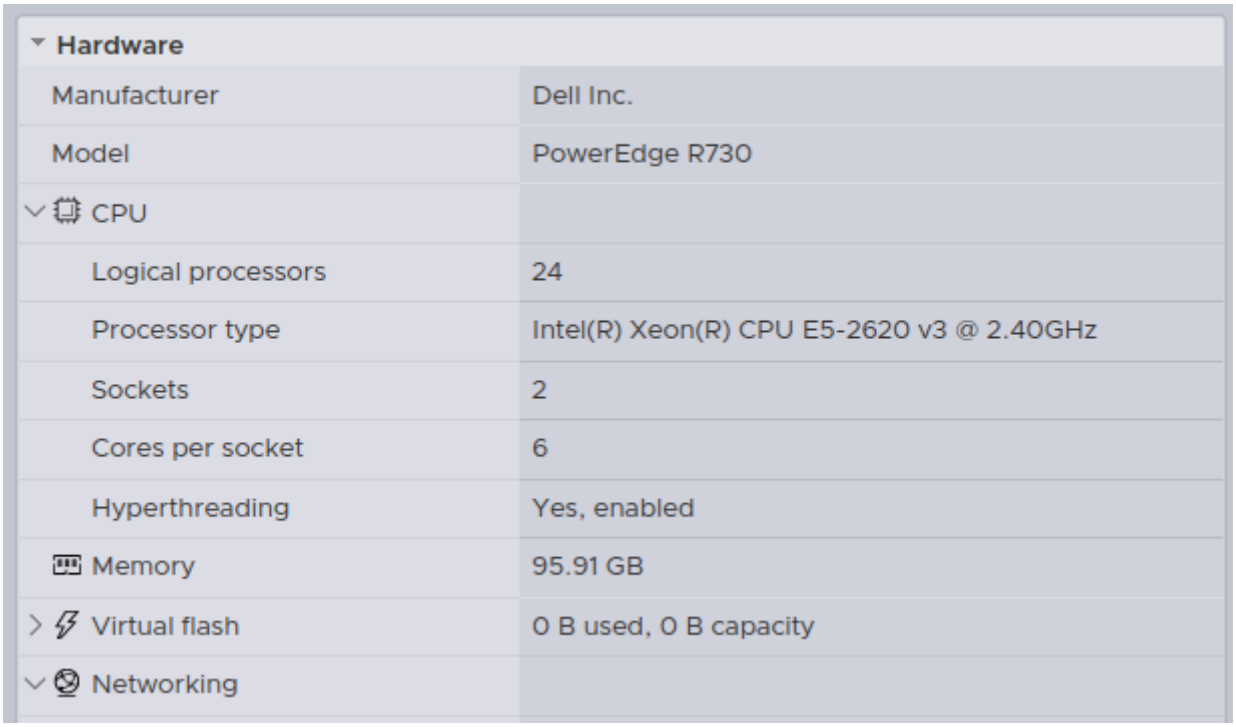
Upon successful login to the ESXi Web Client the page should navigate automatically to the *Host* page, otherwise select it from the pane at the left side of the browser.



On this page the CPU statistics can be found in the *CPU* section of the *Hardware* box in the main pane.



Left clicking on the arrow next to the *CPU* entry will expand the form and present the information needed.

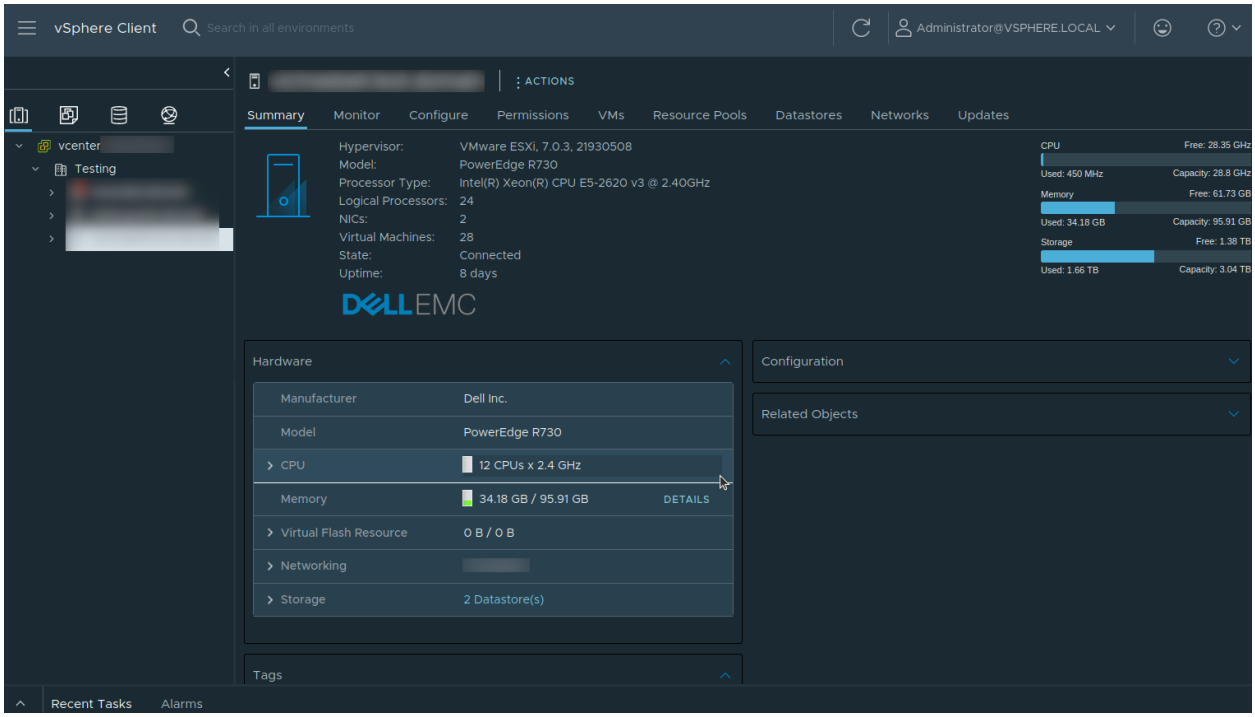


From this screen note the statistics about the processors. The most important entries for this guide are the *Cores per Socket* and *Hyperthreading*.

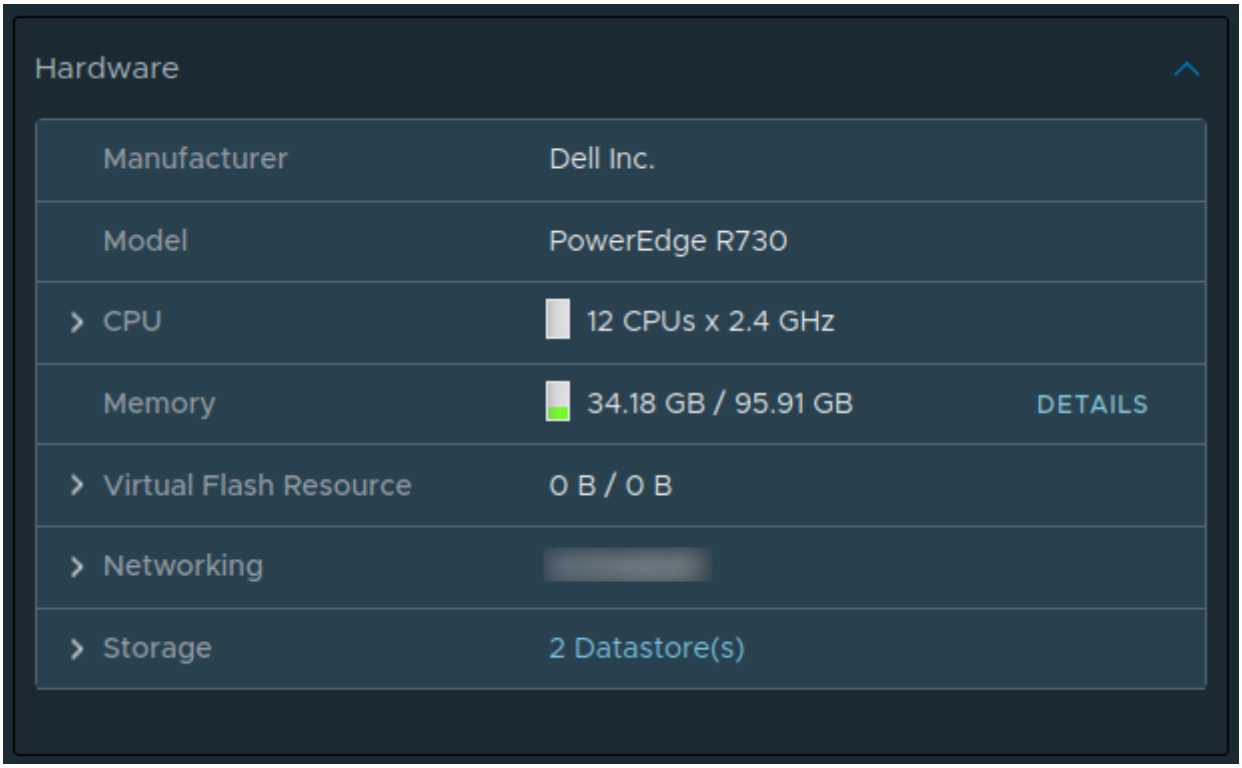
5.1.2 vSphere Web Client

After successful login, use the left pane to navigate to the parent ESXi host of the Bridgeworks Node.

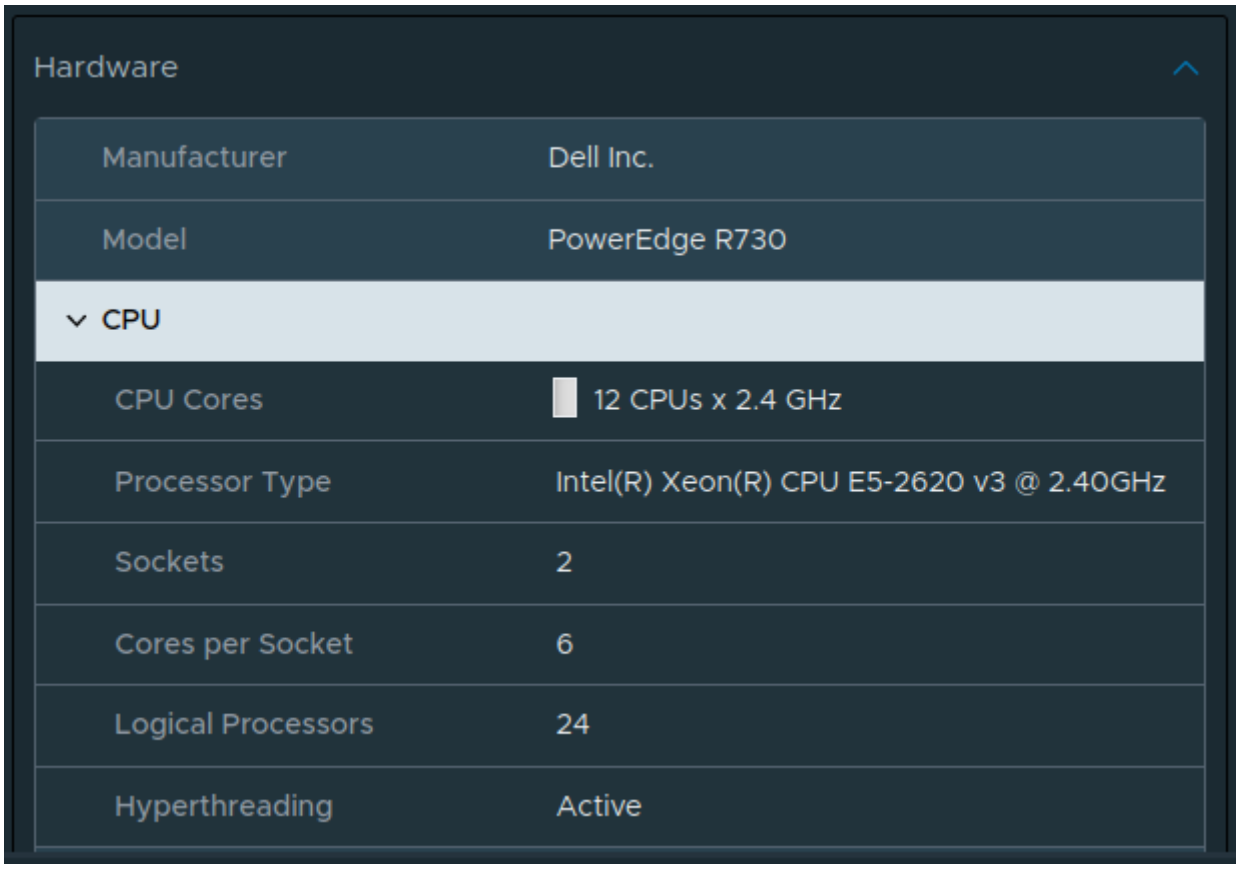
In the main pain, navigate to the *Summary* tab.



Expand the *Hardware* section in the main pane.



Further expand the *CPU* section.



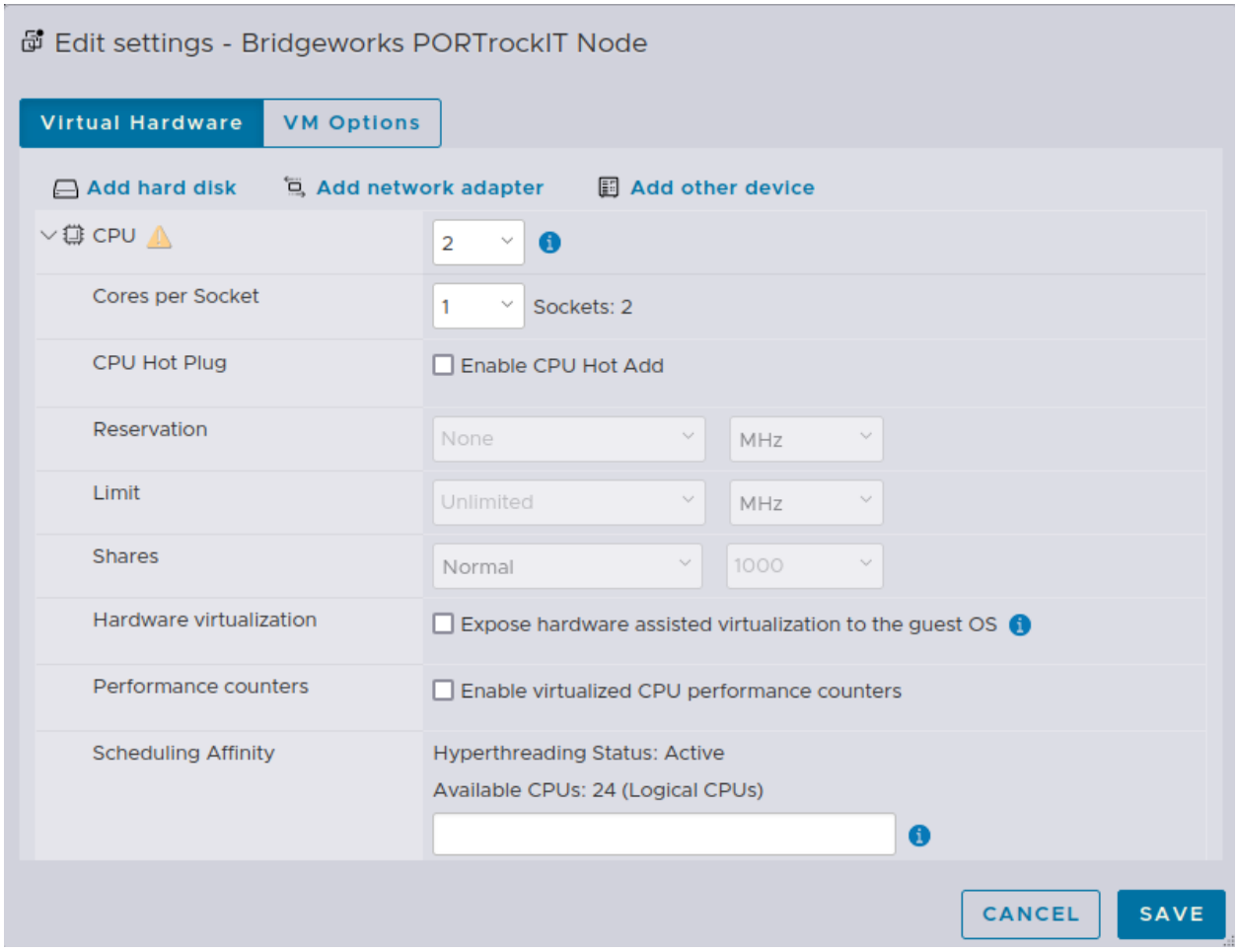
From this screen note the statistics about the processors. The most important entries for this guide

are the *Cores per Socket* and *Hyperthreading*.

5.2 Applying Affinity

5.2.1 ESXi built-in Web Client

In the *Edit Settings* pane, use the arrow to the left of the *CPU* entry to expand that section. Scroll down to locate the *Scheduling Affinity* section.



For setting CPU affinity the text box in the *Scheduling Affinity* section will be used. The affinity input is formatted as a series of numbers representing the cores. The range is from 0 to one less than the total number of cores.

For example, an ESXi host running 12 cores will have a range of 0 to 11.

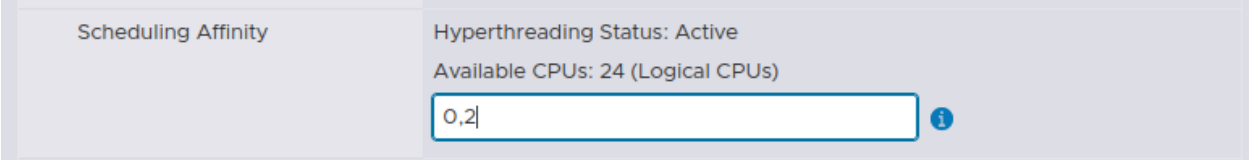
The numbers are separated by '-' or ',' where:

- '0-3' represents cores: 0,1,2,3.
- '0,3' represents only cores 0 and 3.
- The formatting can be mixed. E.g. '0,2-4' represents cores: 0,2,3,4.

At this stage the information from the Host processor setup becomes useful. The best affinity

settings would have a CPU affinity that only selects physical cores that all exist on the same socket.

The image below shows a 2 core Bridgeworks Node being given an affinity of cores 0 and 2. The Host information states that Hyperthreading is not running on this ESXi system so all available cores are real physical CPU cores.

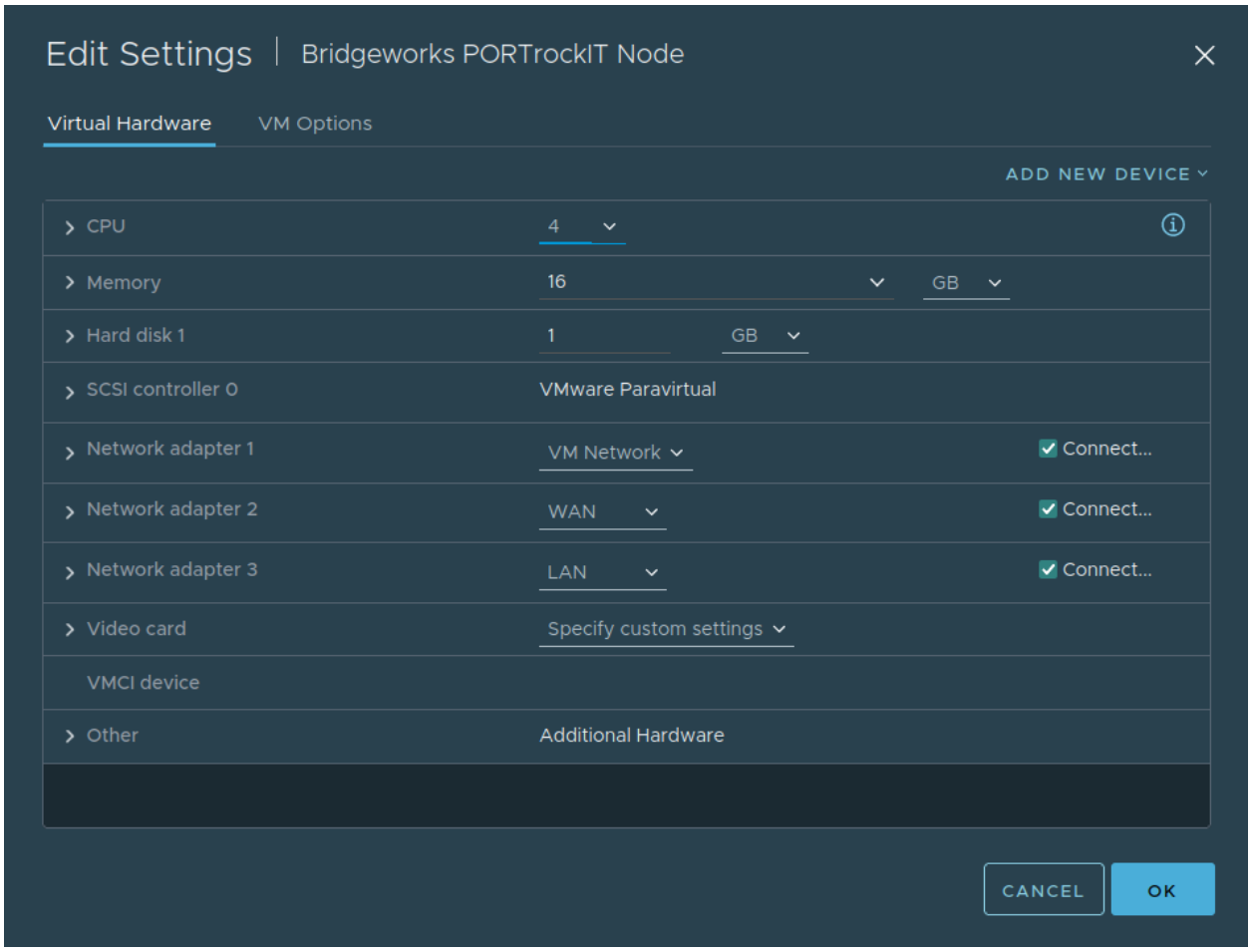


If Hyperthreading were enabled, it would be important to note that 0 and 1 would exist on the first physical core of the first socket. Likewise 2 and 3 would be the second physical core of the same socket.

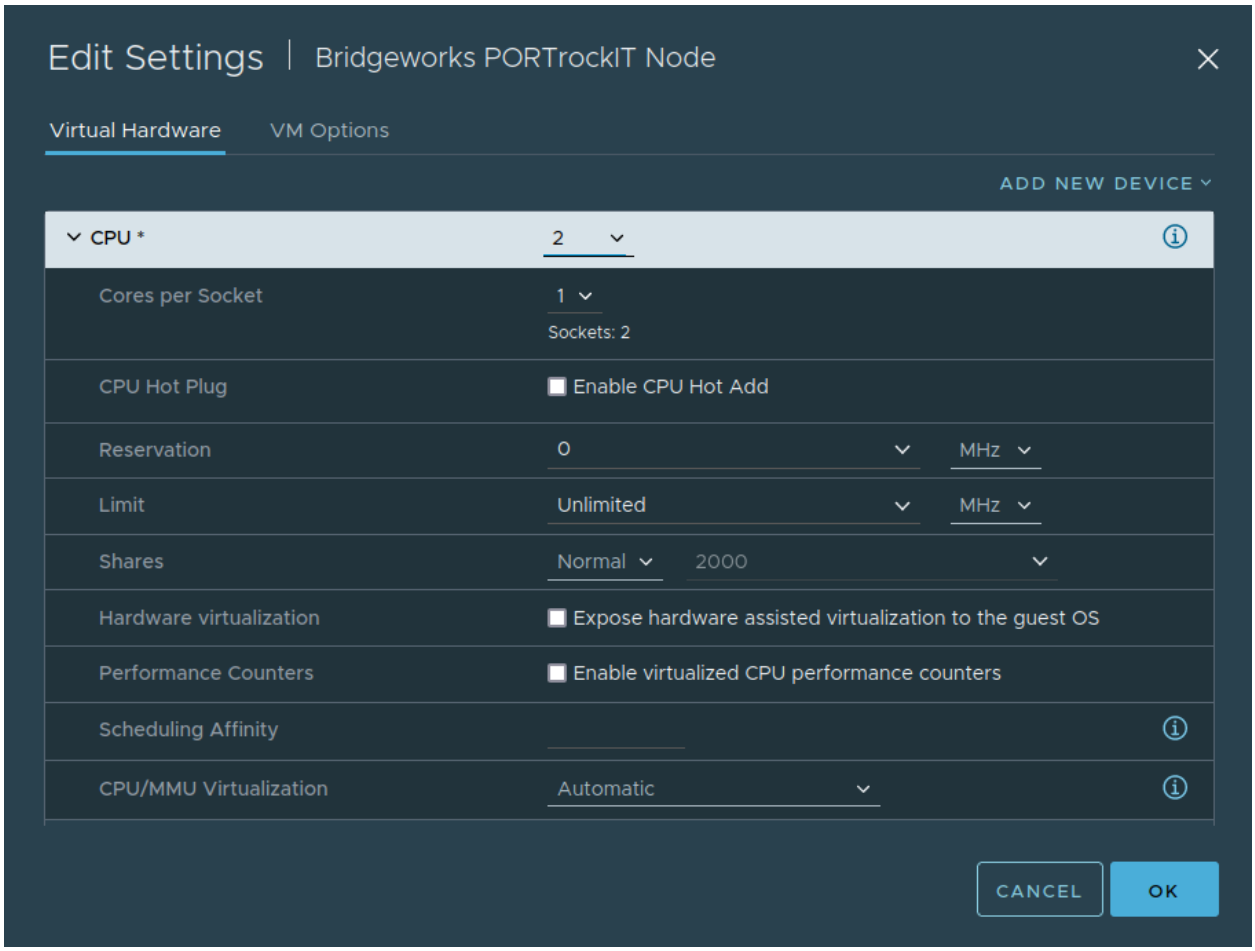
Click *Save* at the bottom of the pane to apply the new memory settings.

5.2.2 vSphere Web Client

In the *Edit Settings* pane, use the arrow to the left of the *CPU* entry to expand that section.



Scheduling Affinity is located at the bottom of the expanded section.



For setting CPU affinity the text box in the *Scheduling Affinity* section will be used. The affinity input is formatted as a series of numbers representing the cores. The range is from 0 to one less than the total number of cores.

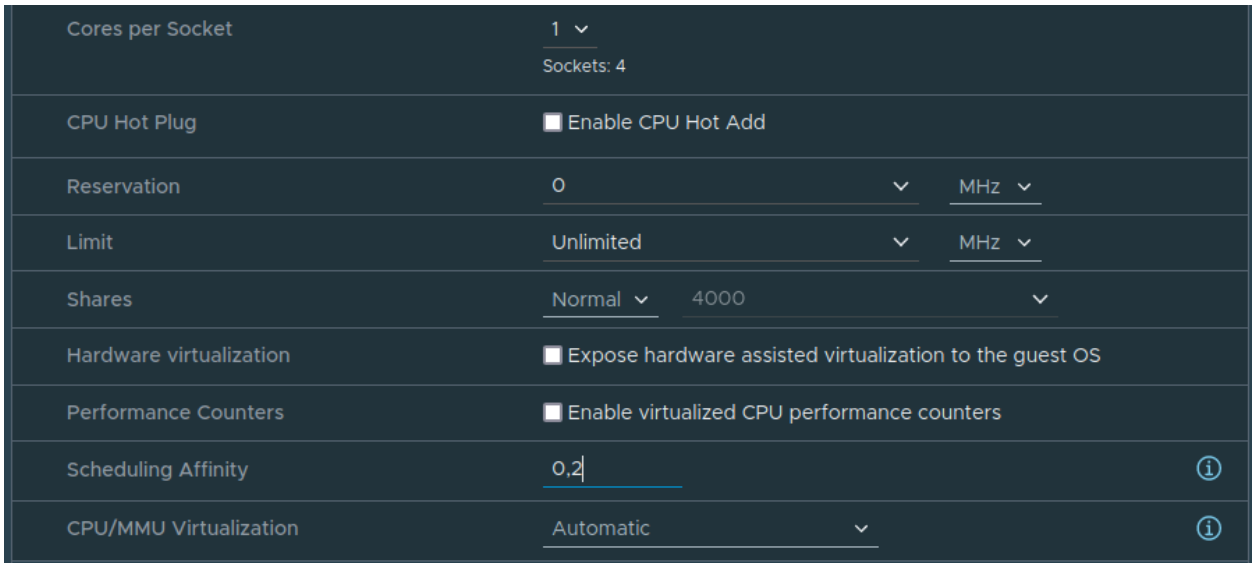
For example, an ESXi host running 12 cores will have a range of 0 to 11.

The numbers are separated by '-' or ',' where:

- '0-3' represents cores: 0,1,2,3.
- '0,3' represents only cores 0 and 3.
- The formatting can be mixed. E.g. '0,2-4' represents cores: 0,2,3,4.

At this stage the information from the Host processor setup becomes useful. The best affinity settings would have a CPU affinity that only selects physical cores that all exist on the same socket.

The image below shows a 2 core Bridgeworks Node being given an affinity of cores 0 and 2. The Host information states that Hyperthreading is not running on this ESXi system so all available cores are real physical CPU cores.



If Hyperthreading were enabled, it would be important to note that 0 and 1 would exist on the first physical core of the first socket. Likewise 2 and 3 would be the second physical core of the same socket.

Click *OK* at the bottom of the pane to apply the new memory settings.

6 Memory Allocation

Initial memory allocations for each system are:

- 100 Series = 2GB
- 200 Series = 4GB
- 400 Series = 16GB

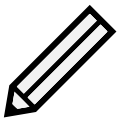
These values are calculated on the basis of how much temporary pool data is utilised during a full bandwidth transfer across a potentially highly utilised network environment.

Ideally, the ESXi host will be able to accommodate the entire memory allocation on the physical RAM slots for a single socket without any competition associated with other guest systems.

If the memory allocation results in RAM starvation of the ESXi host or memory being split across the physical sockets (see Chapter 7: Memory Alignment), then the options are:

- Assess whether the RAM allocation of the Bridgeworks Node can be reduced.

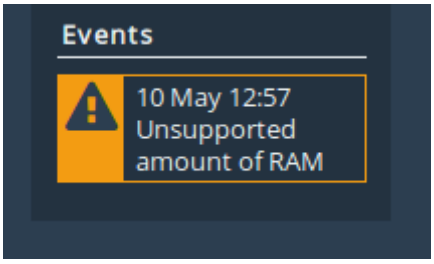
  The System Information page on the web GUI for the node shows the current memory usage, during typical data acceleration this can give a good indication of the memory allocation requirements.
- Consider reordering other virtual guests on the ESXi host to redistribute memory usage. This could free up enough memory on a single socket to fit the Bridgeworks node.
- If the previous options aren't possible then either move the Bridgeworks Node to another ESXi host entirely, or leave the memory in the sub-optimal setup.



Note: The minimum supported memory allocation for a Bridgeworks node is:

- 100 Series = 2GB
- 200 Series = 4GB
- 400 Series = 12GB

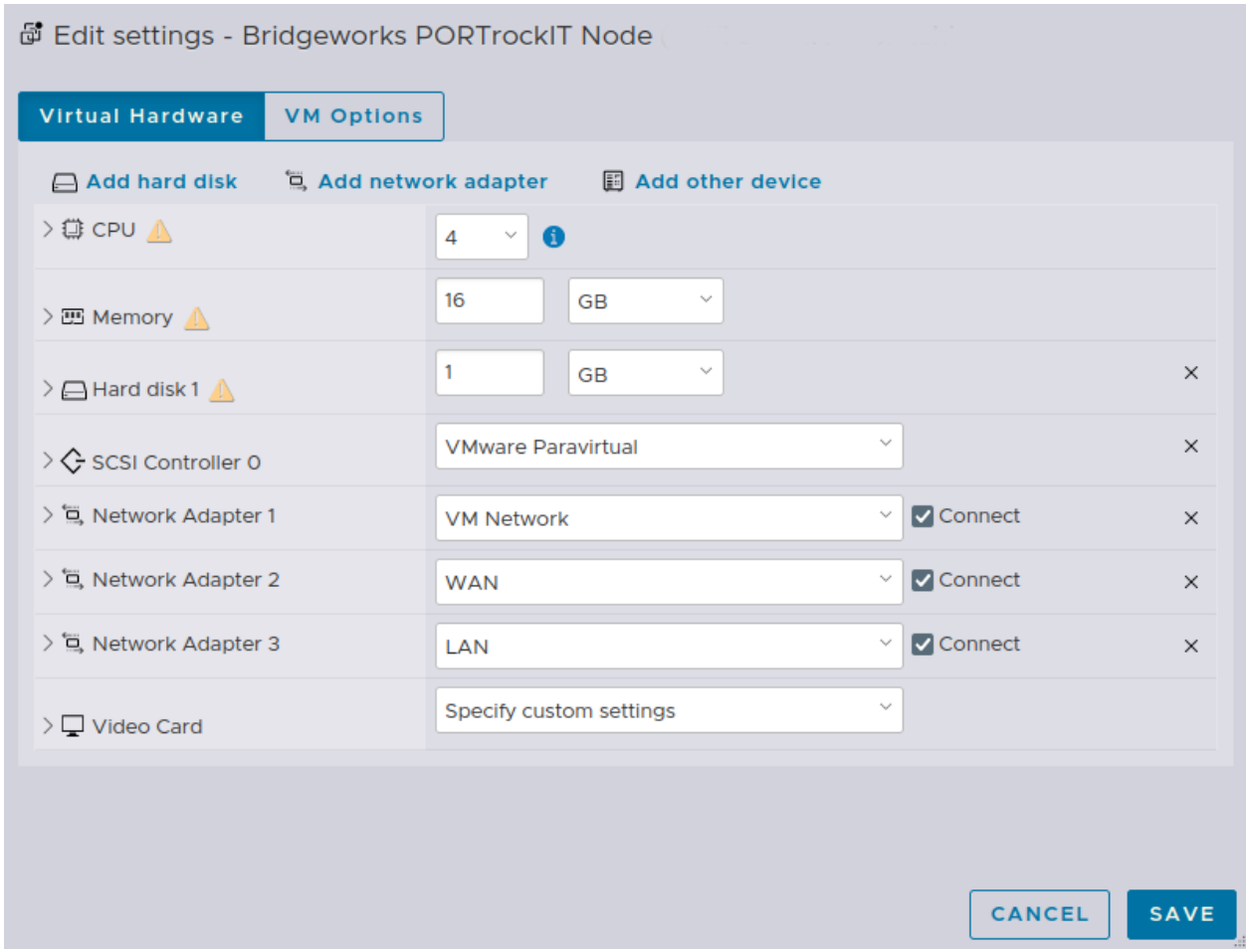
If the Bridgeworks Node is set below the minimum supported memory allocation the Web GUI will display a warning to the user:



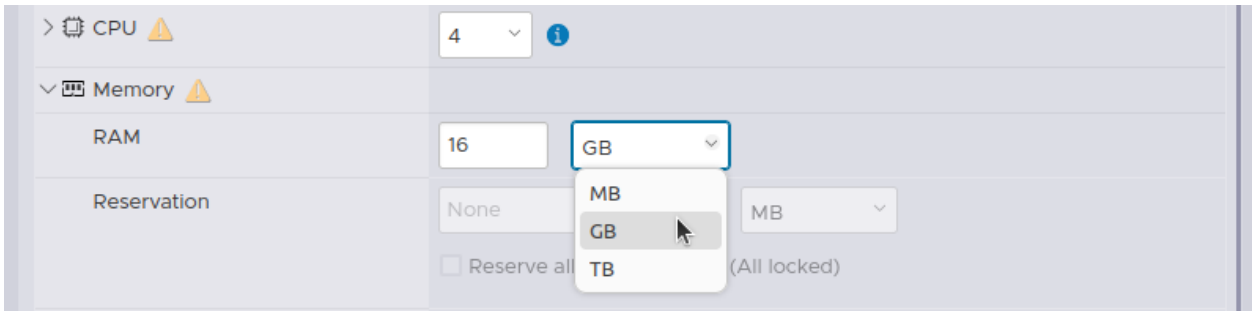
6.1 Adjusting Allocation

6.1.1 ESXi Built-in Web Client

In the *Edit Settings* pane, navigate to the *Virtual Hardware* tab and select the value for the *Memory* entry.



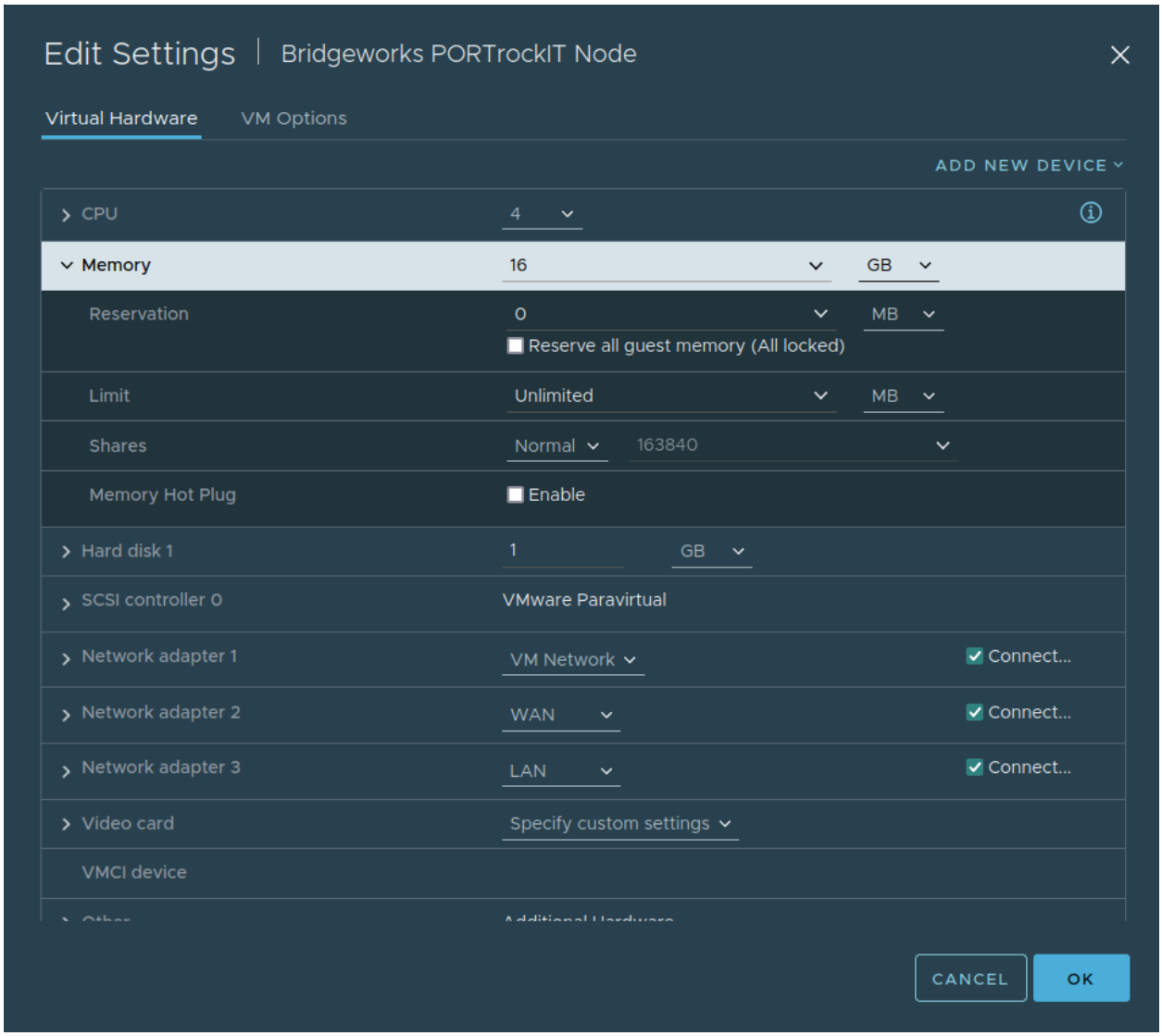
Enter the desired memory allocation. Changing the value to the right from *MB* to *GB* will be easier. Scale the number to suit the change.



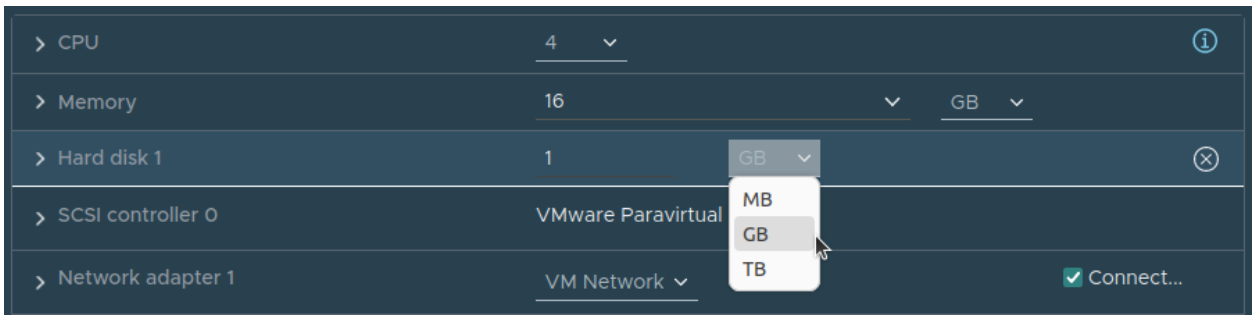
Click *Save* at the bottom of the pane to apply the new memory settings.

6.1.2 vSphere Web Client

In the *Edit Settings* pane, navigate to the *Virtual Hardware* tab and select the value for the *Memory* entry.



Changing the value to the right from *MB* to *GB* will be easier.



Enter the desired memory allocation.

Click *Save* at the bottom of the pane to apply the new memory settings.

7 Memory Alignment

A lack of memory available can cause a drop in performance. In some cases an operating system can become unresponsive while it reverts to pagefile or swap space to maintain operation.

Unlike a memory deficit, memory alignment will not be as catastrophic. Although performance can be reduced by as much as 20-30%.

Ensuring memory is aligned will reduce: CPU usage, memory usage, and increase overall performance.

The same considerations as the memory allocation section apply here. If the ESXi host does not have enough memory in a single NUMA node (see Section 2.4: Memory (NUMA - Non-Uniform Memory Access) Alignment:) for the Bridgeworks Node,then memory misalignment can occur.

Consider shrinking the memory allocation, or reordering the virtual guests on the host, in order to ensure that all of the Bridgeworks memory allocation exists on physical RAM that is local to the CPU cores used by the Bridgeworks Node.

8 Useful Links

Further documentation and support is available through our website: https://support.4bridgeworks.com/

If your question is not answered in our documentation, please submit a ticket: https://support.4bridgeworks.com/contact/