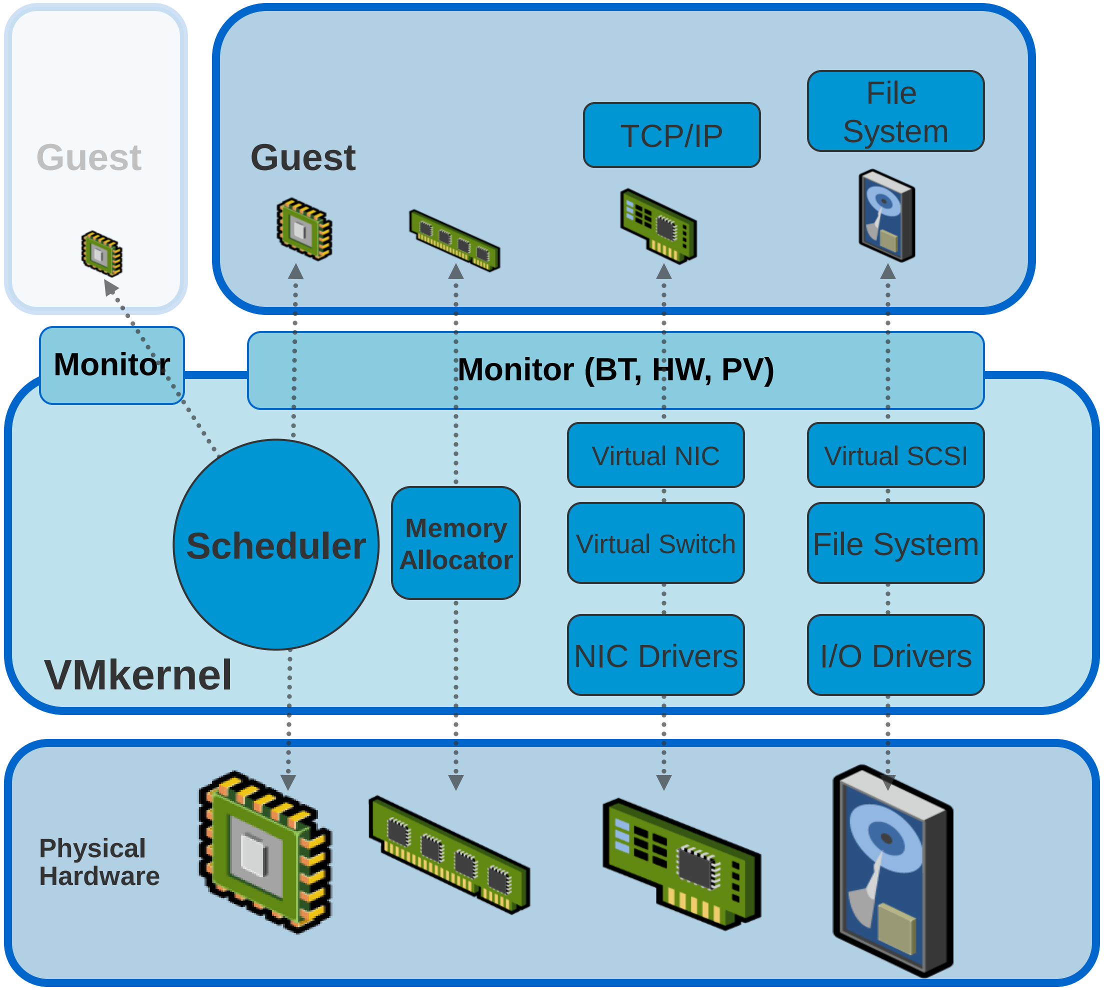
VMware Performance Overview

Confidential

VMware ESXESXi Architecture



CPU is controlled by scheduler and virtualized by monitor

Monitor supports:
- •BT (Binary Translation)
- •HW (Hardware assist)
- •PV (Paravirtualization)

Memory is allocated by the VMkernel and virtualized by the monitor

Network and I/O devices are emulated and proxied though native device drivers

Performance Factors in a vSphere Environment

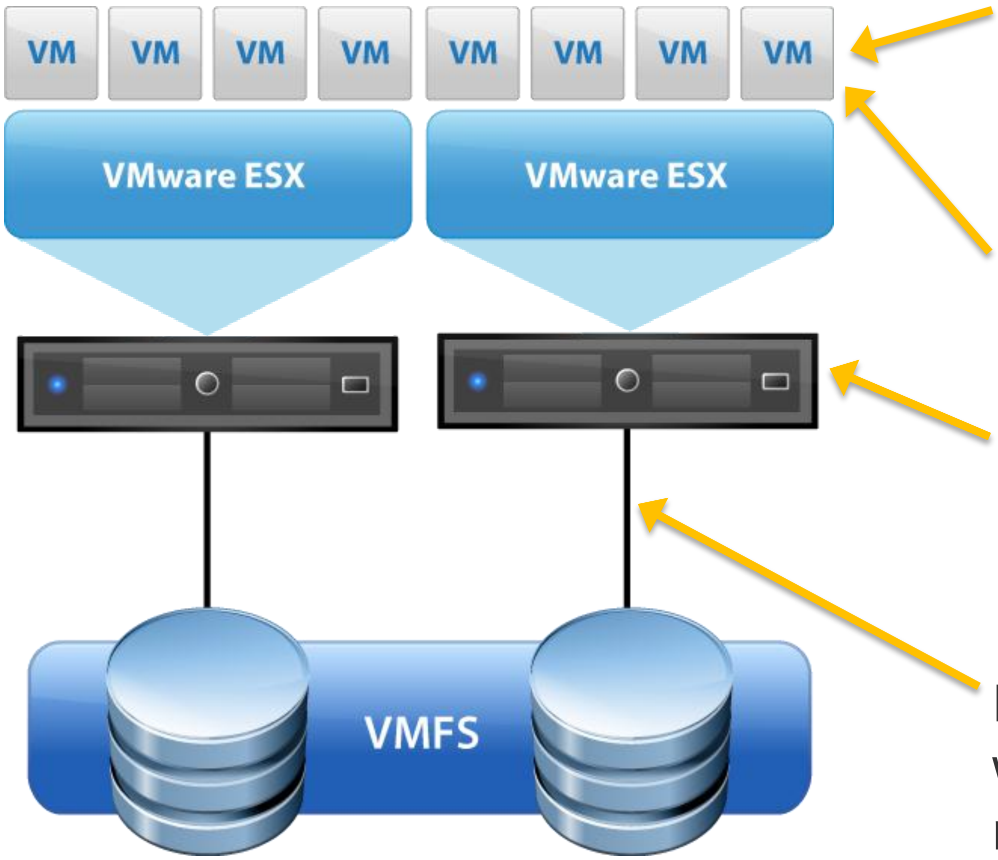
Hardware:

- CPU
- Memory
- Storage
- Network

Software:

- VMM
- Virtual machine settings
- Applications

Traditional Best Practices for Performance



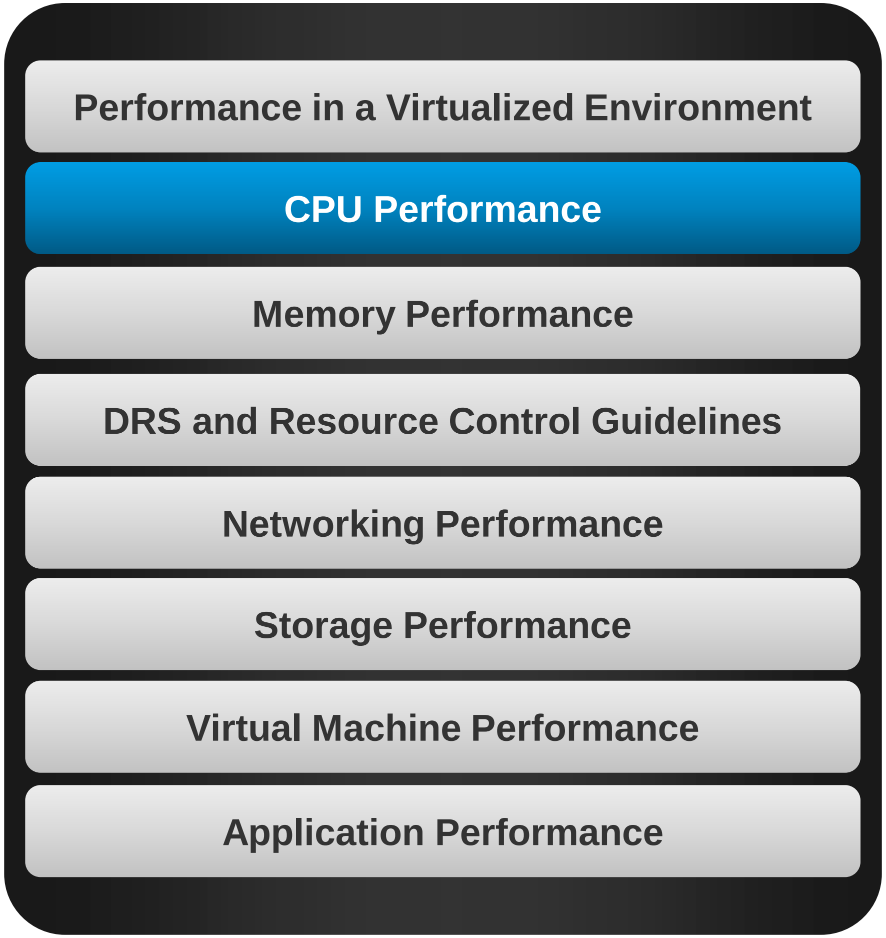
Understand and account for the target application's profile.

Optimally configure virtual machines.

Improve virtualization performance by using newer platforms when available.

Monitor consolidated workloads closely to discover new bottlenecks quickly.

Performance Factors



Performance in a Virtualized Environment

CPU Performance

Memory Performance

DRS and Resource Control Guidelines

Networking Performance

Storage Performance

Virtual Machine Performance

Application Performance

CPU Scheduler Overview

- **The CPU scheduler is crucial to providing good performance in a consolidated environment.**
- **The CPU scheduler has the following features:**
  - Schedules virtual CPUs (vCPUs) on physical CPUs
  - Enforces the proportional-share algorithm for CPU usage
  - Supports SMP virtual machines
  - Uses relaxed co-scheduling for SMP virtual machines
  - Is NUMA/processor/cache topology–aware

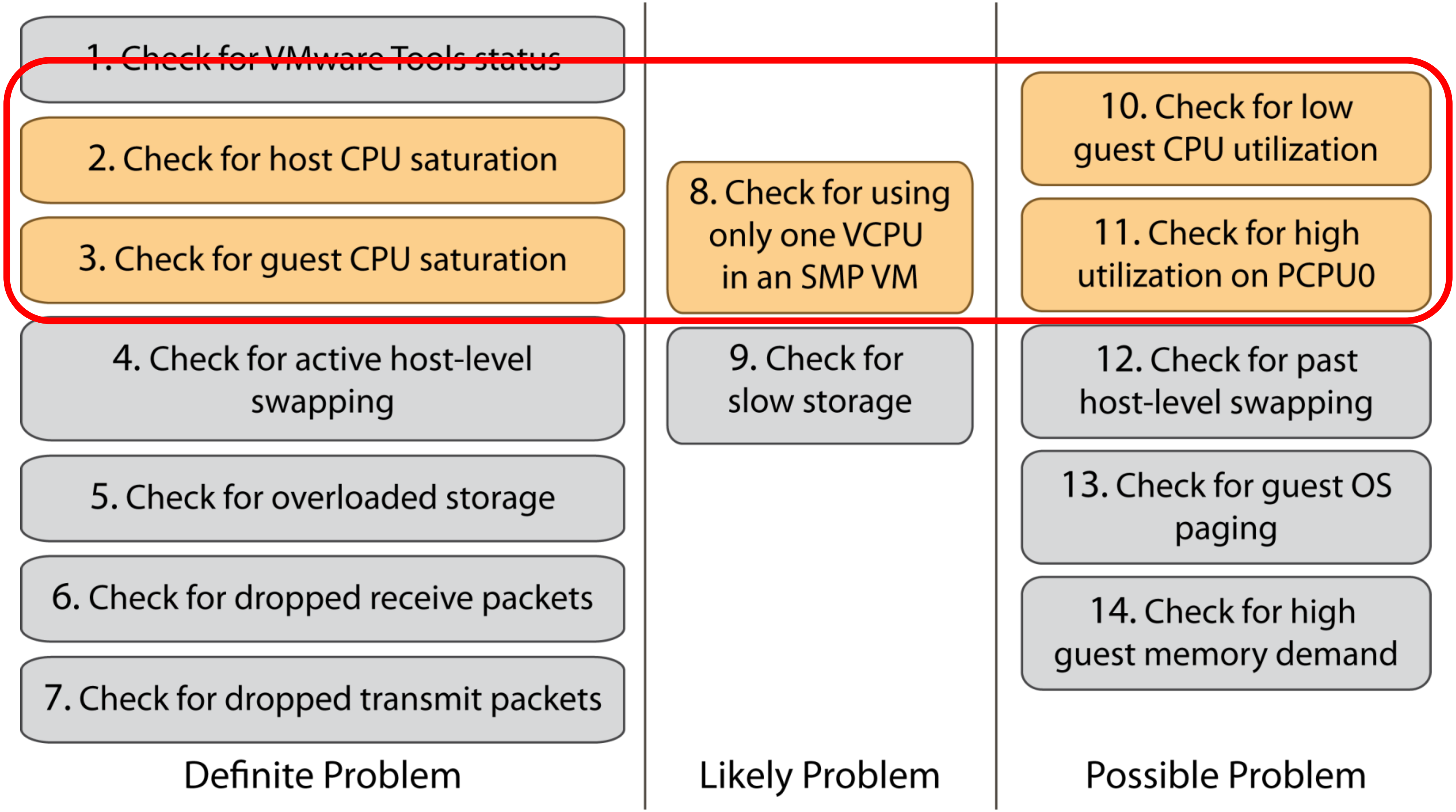
What Affects CPU Performance?

- **Idling virtual machines:**
  - Consider the overhead of delivering guest timer interrupts.
- **CPU affinity:**
  - This constrains the scheduler and can cause an imbalanced load.
- **SMP virtual machines:**
  - Some co-scheduling overhead is incurred.
- **Insufficient CPU resources to satisfy demand:**
  - If CPU contention exists, the scheduler forces vCPUs of lower-priority virtual machines to queue their CPU requests in deference to higher-priority virtual machines.
- **Mixing large SMPs with many small SMPs**

Warning Sign: Ready Time

- **vCPUs are allocated CPU cycles on an assigned physical CPU based on the proportional-share algorithm enforced by the CPU scheduler.**
  - If a vCPU tries to execute a CPU instruction while no cycles are available on the physical CPU, the request is queued.
  - A physical CPU with no cycles could be due to high load on the physical CPU or a higher-priority vCPU receiving preference.
- **The amount of time that the vCPU waits for the physical CPU to become available is called ready time.**
  - This latency can affect performance of the guest operating system and its applications within a virtual machine.

Review: Basic Troubleshooting Flow



1. Check for VMware Tools status

2. Check for host CPU saturation

3. Check for guest CPU saturation

4. Check for active host-level swapping

5. Check for overloaded storage

6. Check for dropped receive packets

7. Check for dropped transmit packets

Definite Problem

8. Check for using only one VCPU in an SMP VM

9. Check for slow storage

Likely Problem

10. Check for low guest CPU utilization

11. Check for high utilization on PCPU0

12. Check for past host-level swapping

13. Check for guest OS paging

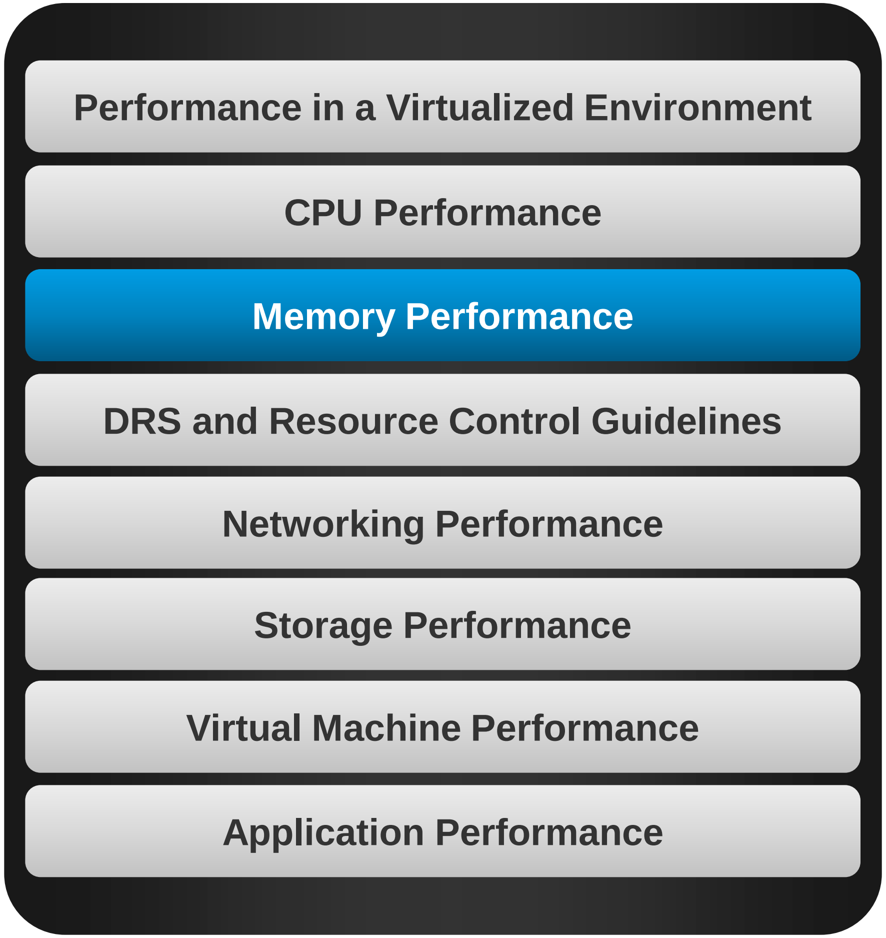
14. Check for high guest memory demand

Possible Problem

CPU Performance Best Practices

- **Avoid using SMP unless specifically required by the application running in the guest operating system.**
- **Prioritize virtual machine CPU usage with the proportional-share algorithm.**
- **Use vMotion and DRS to redistribute virtual machines and reduce contention.**
- **Increase the efficiency of virtual machine usage by:**
  - Leveraging application tuning guides
  - Tuning the guest operating system
  - Optimizing the virtual hardware

Performance Factors



Performance in a Virtualized Environment

CPU Performance

Memory Performance

DRS and Resource Control Guidelines

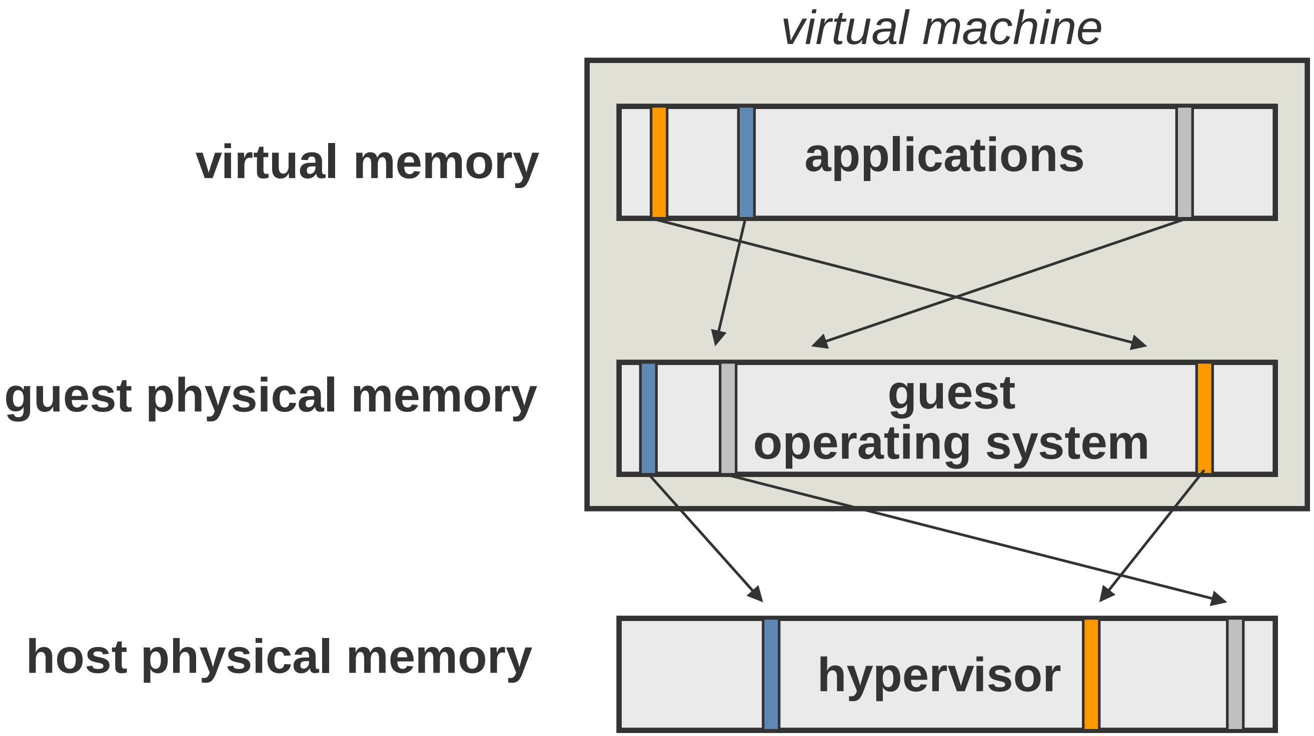
Networking Performance

Storage Performance

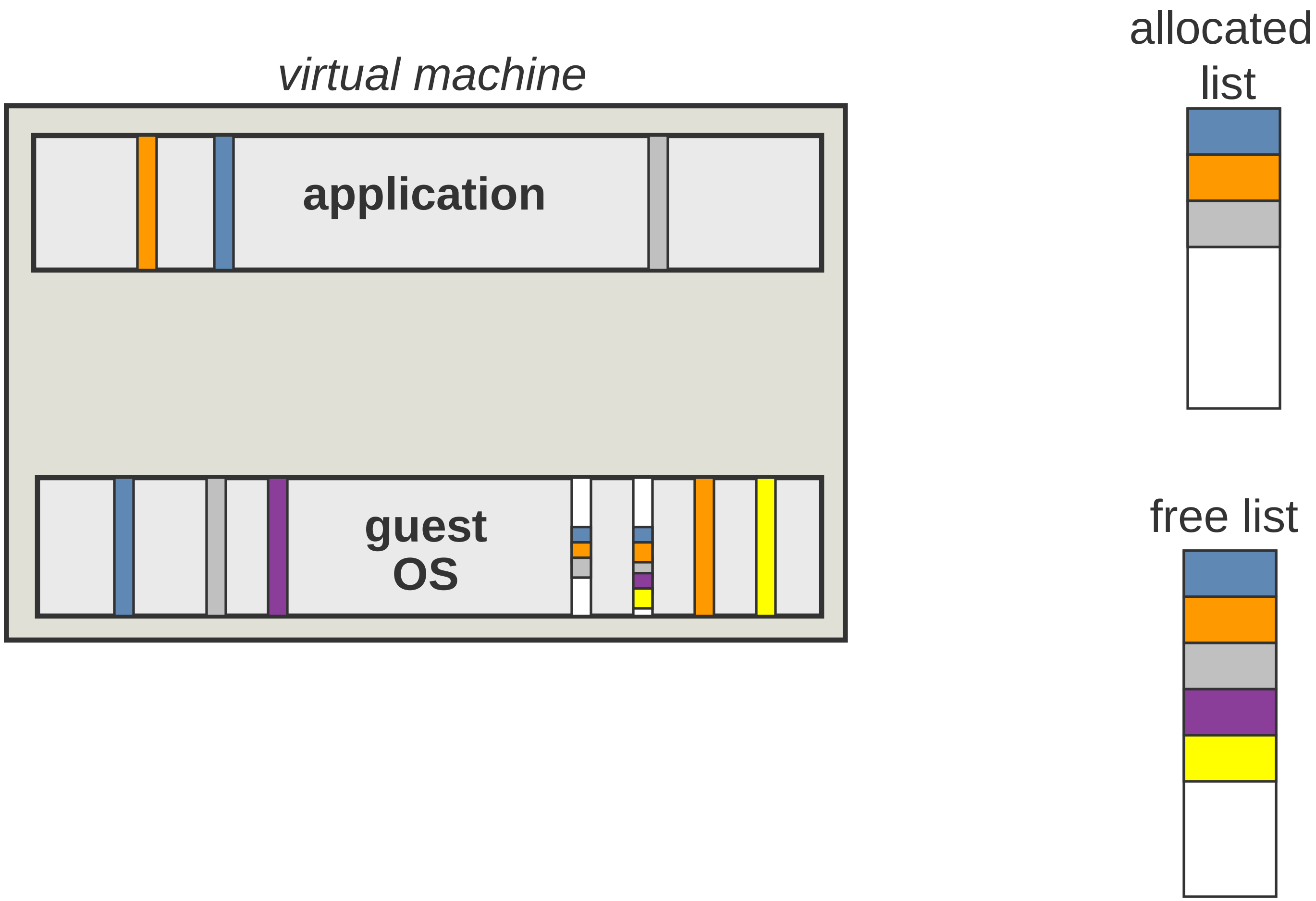
Virtual Machine Performance

Application Performance

Virtual Memory Overview

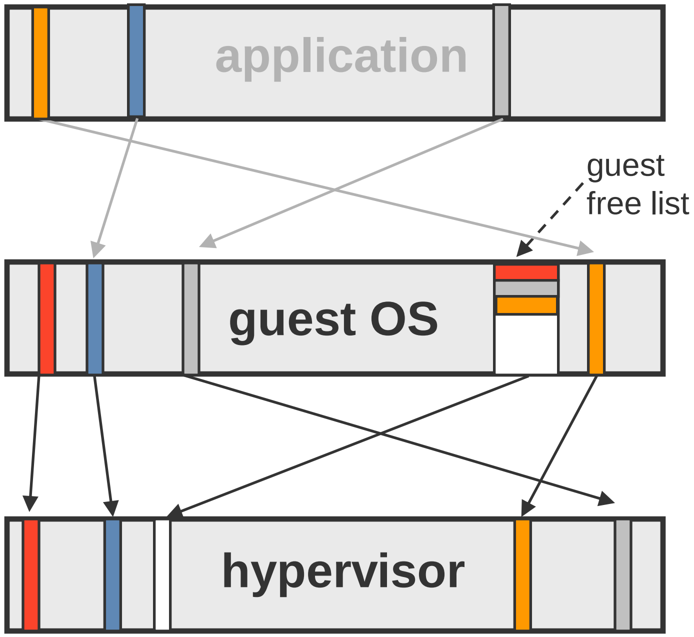


Application and Guest OS Memory Management



Memory Reclamation

- **Guest physical memory is not “freed” in the typical sense.**
  - Memory is moved to the “free” list.
- **The hypervisor is not aware of when the guest frees memory.**
  - It has no access to the guest’s “free” list.
  - The virtual machine can accrue lots of host physical memory.
- **The hypervisor cannot “reclaim” the host physical memory freed up by the guest.**

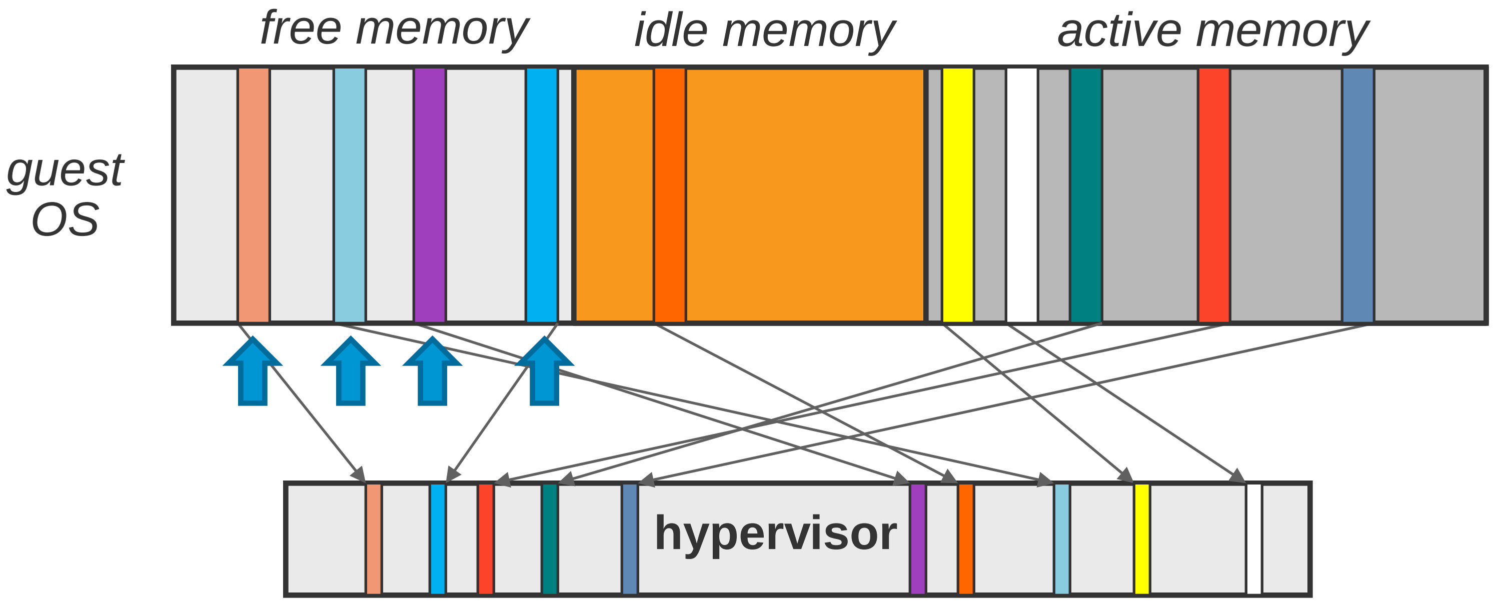


Virtual Machine Memory-Reclamation Techniques

- **The hypervisor relies on memory-reclamation techniques to free host physical memory:**
  - Transparent page sharing:
    - Enabled by default
  - Ballooning
  - Host-level, or hypervisor, swapping

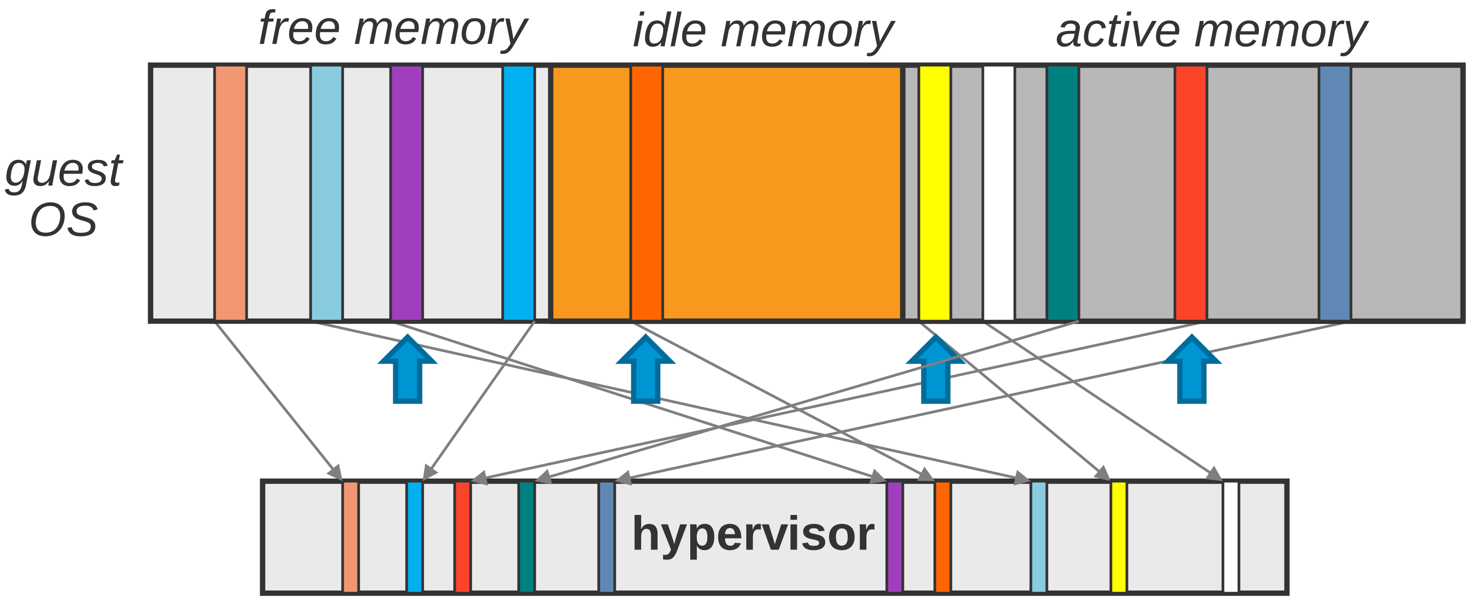
Reclaiming Memory with Ballooning

- **Ballooning preferentially selects free or idle VM memory.**
  - This is because the guest OS allocates from free memory.
- **But if asked to reclaim too much, ballooning will eventually start reclaiming active memory.**



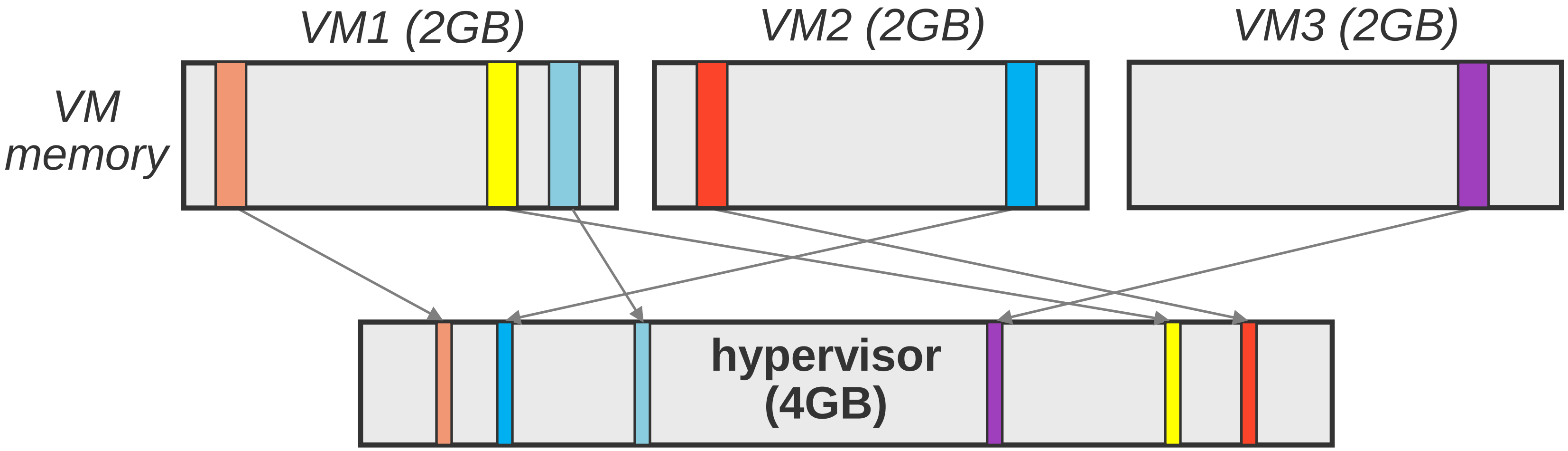
Reclaiming Memory with Host Swapping

- **Host-level swapping randomly selects guest physical memory to reclaim, potentially including a virtual machine's active memory.**

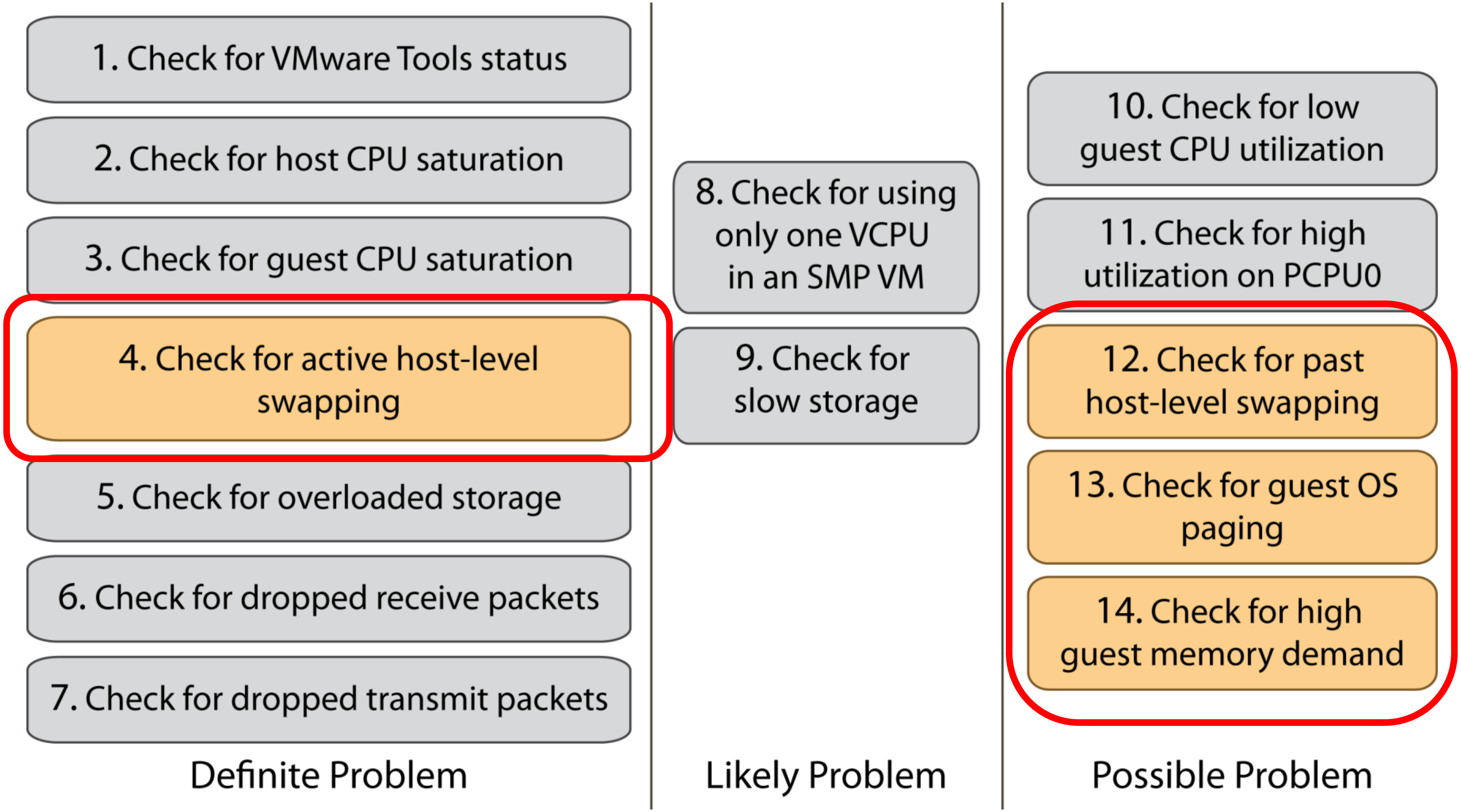


Why Does the Hypervisor Reclaim Memory?

- The hypervisor reclaims memory to support VMware® ESX™/ESXi memory overcommitment.
- A host’s memory is overcommitted when total amount of guest physical memory is greater than the amount of host physical memory.



Review: Basic Troubleshooting Flow



1. Check for VMware Tools status

2. Check for host CPU saturation

3. Check for guest CPU saturation

4. Check for active host-level swapping

5. Check for overloaded storage

6. Check for dropped receive packets

7. Check for dropped transmit packets

Definite Problem

8. Check for using only one VCPU in an SMP VM

9. Check for slow storage

Likely Problem

10. Check for low guest CPU utilization

11. Check for high utilization on PCPU0

12. Check for past host-level swapping

13. Check for guest OS paging

14. Check for high guest memory demand

Possible Problem

Causes of Active Host-Level Swapping

- **The basic cause of host-level swapping:**
  - Memory overcommitment from using memory-intensive virtual machines whose combined configured memory is greater than the amount of host physical memory available
- **Causes of active host-level swapping:**
  - Excessive memory overcommitment
  - Memory overcommitment with memory reservations
  - Balloon drivers in virtual machines not running or disabled

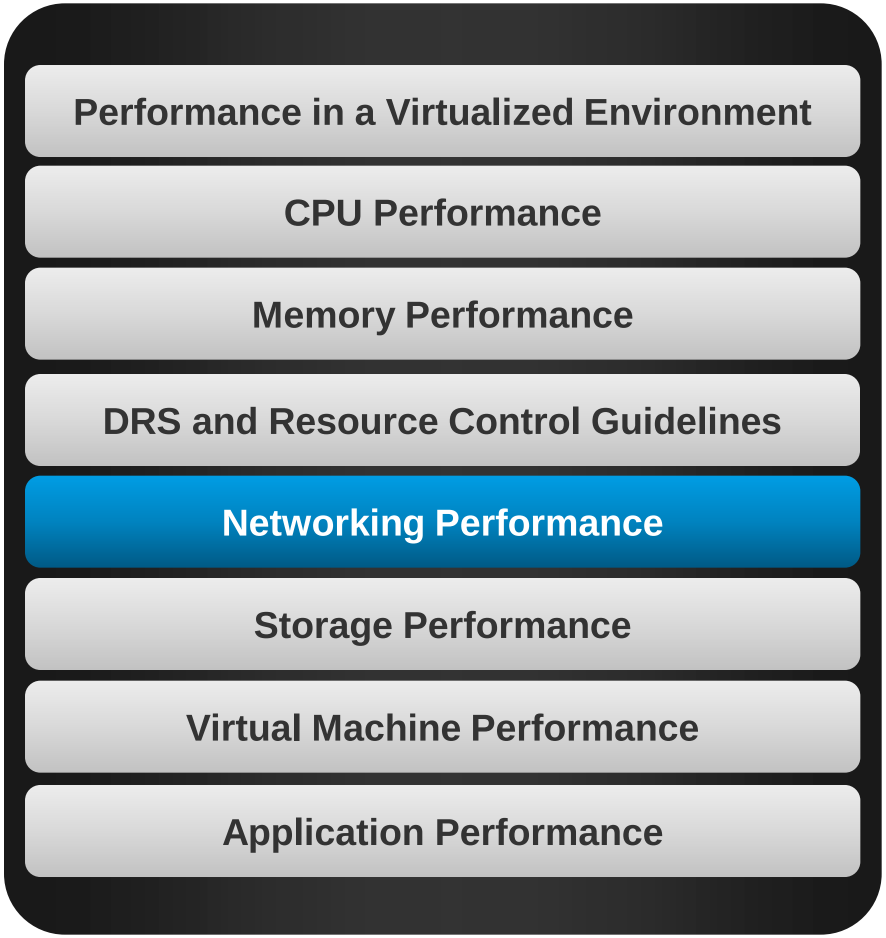
Resolving Host-Level Swapping

- **To resolve this problem:**
  - Reduce the level of memory overcommitment.
  - Enable the balloon driver in all virtual machines.
  - Reduce memory reservations.
  - Use resource controls to dedicate memory to critical virtual machines.

Memory Performance Best Practices

- Allocate enough memory to hold the working set of applications you will run in the virtual machine, thus minimizing swapping.
- Never disable the balloon driver. Always keep it enabled.
- Keep transparent page sharing enabled
- Avoid overcommitting memory to the point that it results in heavy memory reclamation.

Performance Factors



Performance in a Virtualized Environment

CPU Performance

Memory Performance

DRS and Resource Control Guidelines

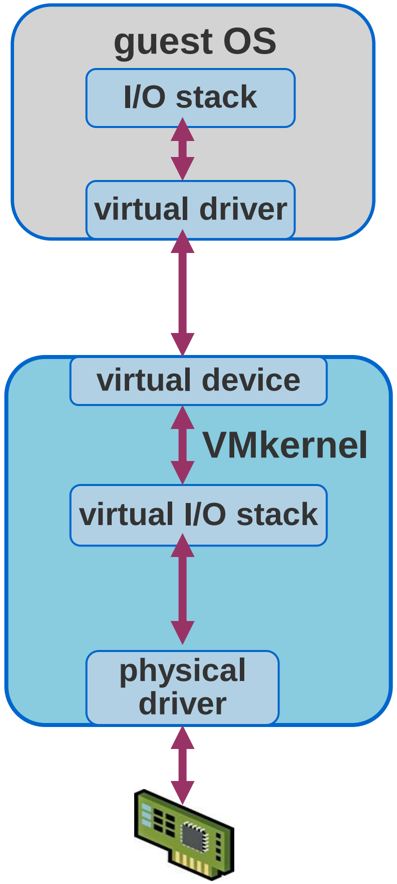
Networking Performance

Storage Performance

Virtual Machine Performance

Application Performance

I/O Virtualization Overhead



- **CPU cycles are used to move the data from the virtual device to the physical device:**
  - Virtual interrupts
  - Address space switches
- **Overhead is higher on the Rx side than the Tx side.**
- **Slower I/O does not get penalized by CPU overhead.**

Network Adapter Features

- **VMware vSphere™ takes advantage of many of the performance features of modern network adapters, including:**
  - Checksum off-load
  - TCP segmentation off-load
  - Jumbo frames
  - Capability of handling high memory
  - 10Gb Ethernet
  - NetQueue
  - VMDirectPath I/O

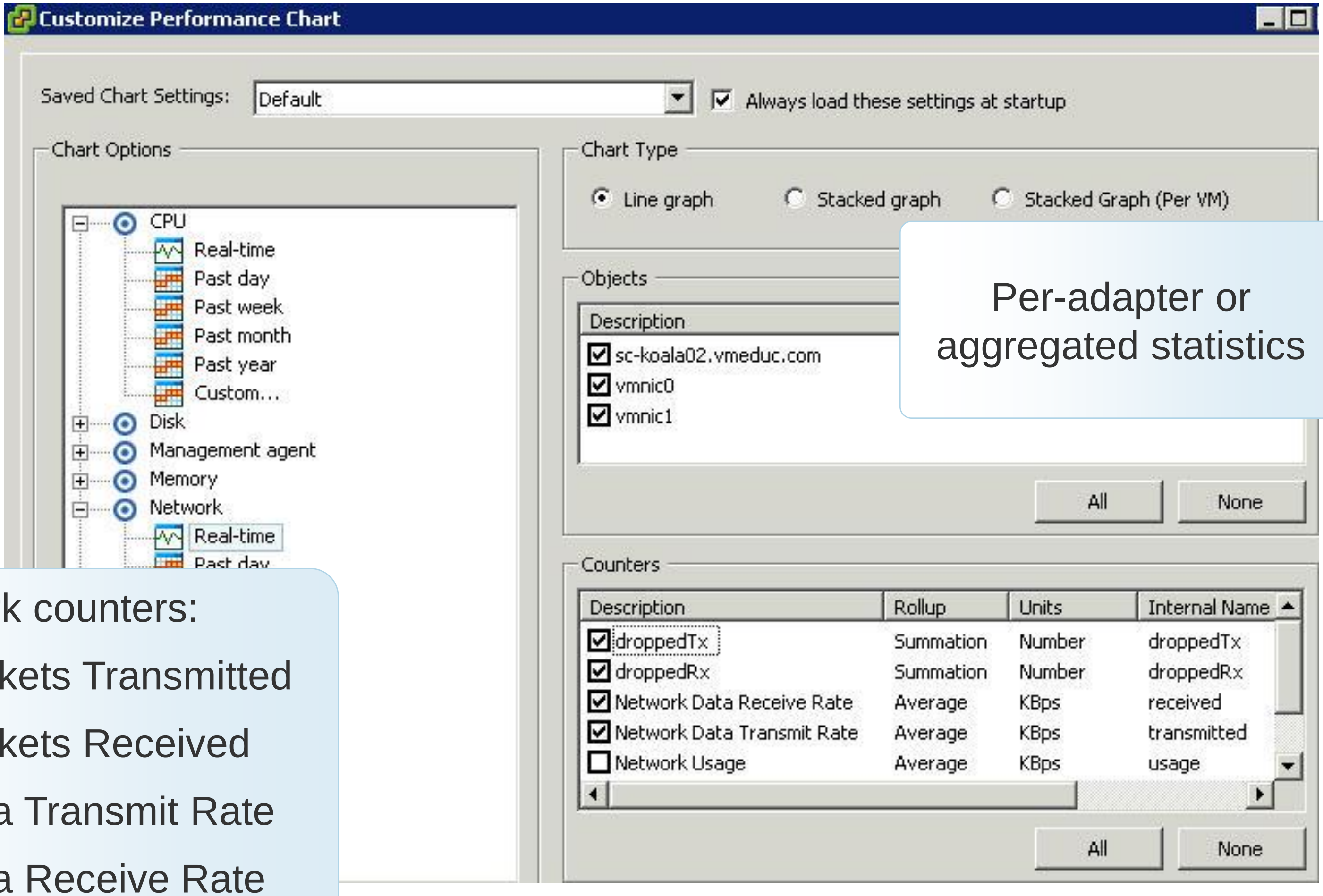
Virtual Machine–to–Virtual Machine Networking

- **Virtual machine communication interface (VMCI):**
  - For virtual machine–to–virtual machine communication on the same host:
    - Shared memory region with queue abstraction on top
  - VMCI sockets: An API library that supports fast and efficient communication between a virtual machine and its host, or between guest virtual machines on the same host
- **Performance:**
  - Virtual machine–to–virtual machine throughput can reach up to 29Gbps.
  - Throughput is specific to hardware and scheduling behavior.

Network Capacity Metrics

- **Identify network problems.**
  - Determine available bandwidth and compare with expectations.
- **What do I do?**
  - Check key metrics. Significant network statistics in a vSphere environment are:
    - Network usage
    - Host droppedRx (received packets dropped)
    - Host droppedTx (transmitted packets dropped)
    - Net packets received
    - Net packets transmitted

vSphere Client Networking Statistics



Customize Performance Chart

Saved Chart Settings: Default ☑ Always load these settings at startup

Chart Options

- CPU
  - Real-time
  - Past day
  - Past week
  - Past month
  - Past year
  - Custom...
- Disk
- Management agent
- Memory
- Network
  - Real-time
  - Past day

Chart Type

● Line graph ○ Stacked graph ○ Stacked Graph (Per VM)

Objects

Description
☑ sc-koala02.vmeduc.com
☑ vmnic0
☑ vmnic1

Counters

Description	Rollup	Units	Internal Name
☑ droppedTx	Summation	Number	droppedTx
☑ droppedRx	Summation	Number	droppedRx
☑ Network Data Receive Rate	Average	KBps	received
☑ Network Data Transmit Rate	Average	KBps	transmitted
□ Network Usage	Average	KBps	usage

Per-adapter or aggregated statistics

Useful network counters:

-Network Packets Transmitted

-Network Packets Received

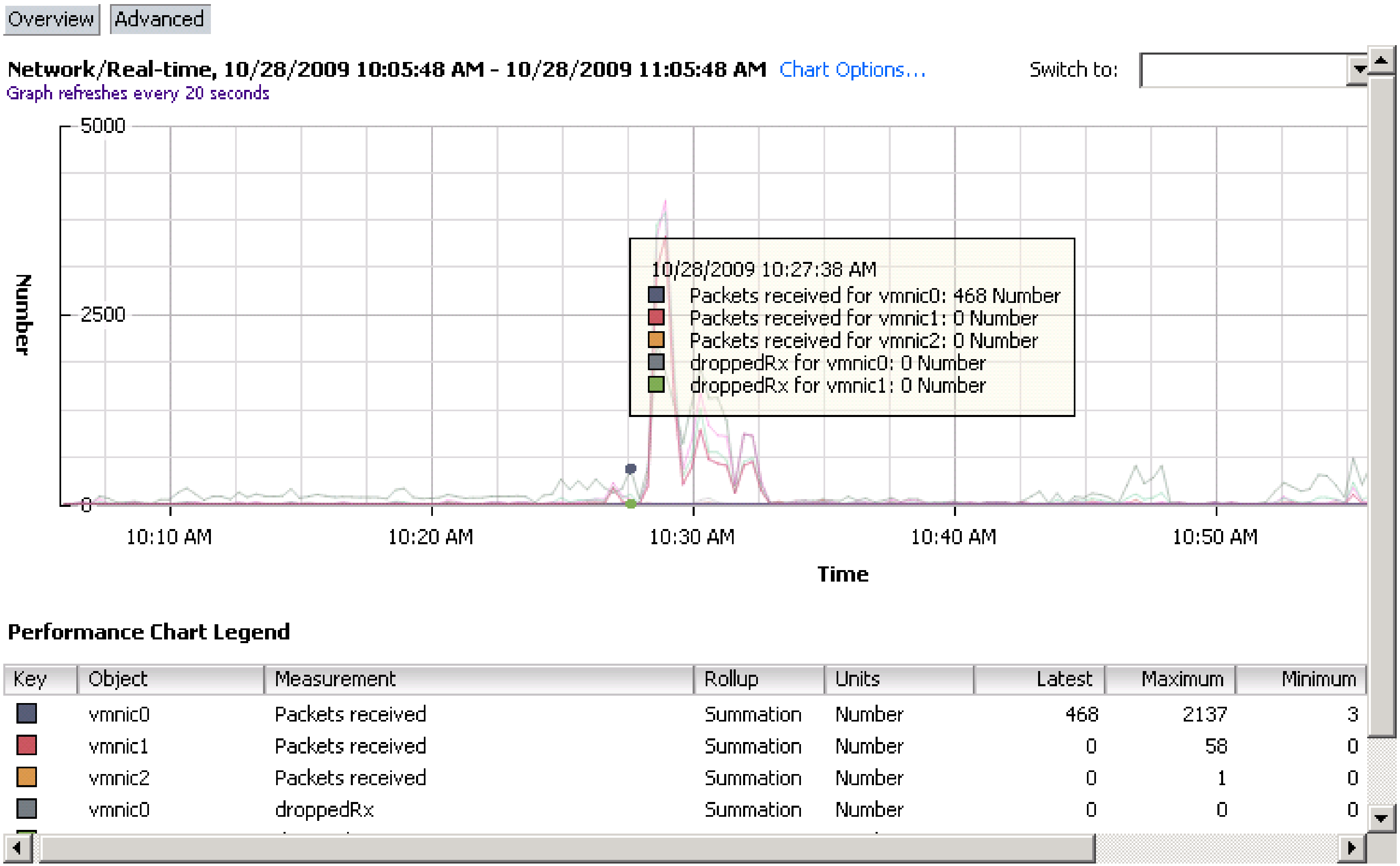
-Network Data Transmit Rate

-Network Data Receive Rate

-droppedTx

-droppedRx

vSphere Client Network Performance Graph



Network/Real-time, 10/28/2009 10:05:48 AM - 10/28/2009 11:05:48 AM

Graph refreshes every 20 seconds

10/28/2009 10:27:38 AM
- ■ Packets received for vmnic0: 468 Number
- 🟥 Packets received for vmnic1: 0 Number
- 🟧 Packets received for vmnic2: 0 Number
- ⬛ droppedRx for vmnic0: 0 Number
- 🟩 droppedRx for vmnic1: 0 Number

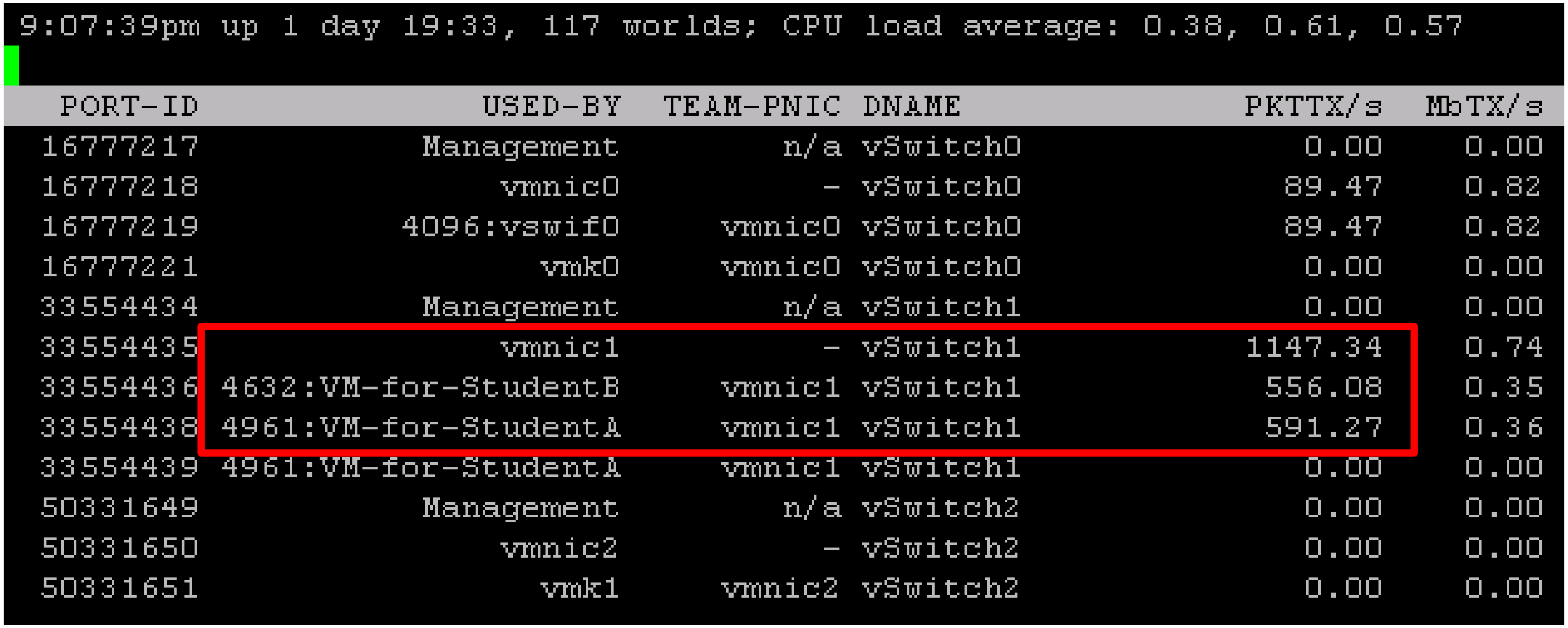
Performance Chart Legend

Key	Object	Measurement	Rollup	Units	Latest	Maximum	Minimum
■	vmnic0	Packets received	Summation	Number	468	2137	3
🟥	vmnic1	Packets received	Summation	Number	0	58	0
🟧	vmnic2	Packets received	Summation	Number	0	1	0
⬛	vmnic0	droppedRx	Summation	Number	0	0	0

resxtop/esxtop Networking Statistics

- **Bandwidth:**
  - Receive (MbRX/s), Transmit (MbTX/s)
- **Operations/s:**
  - Receive (PKTRX/s), Transmit (PKTTX/s)
- **Errors:**
  - Packets dropped during transmit (%DRPTX), receive (%DRPRX)

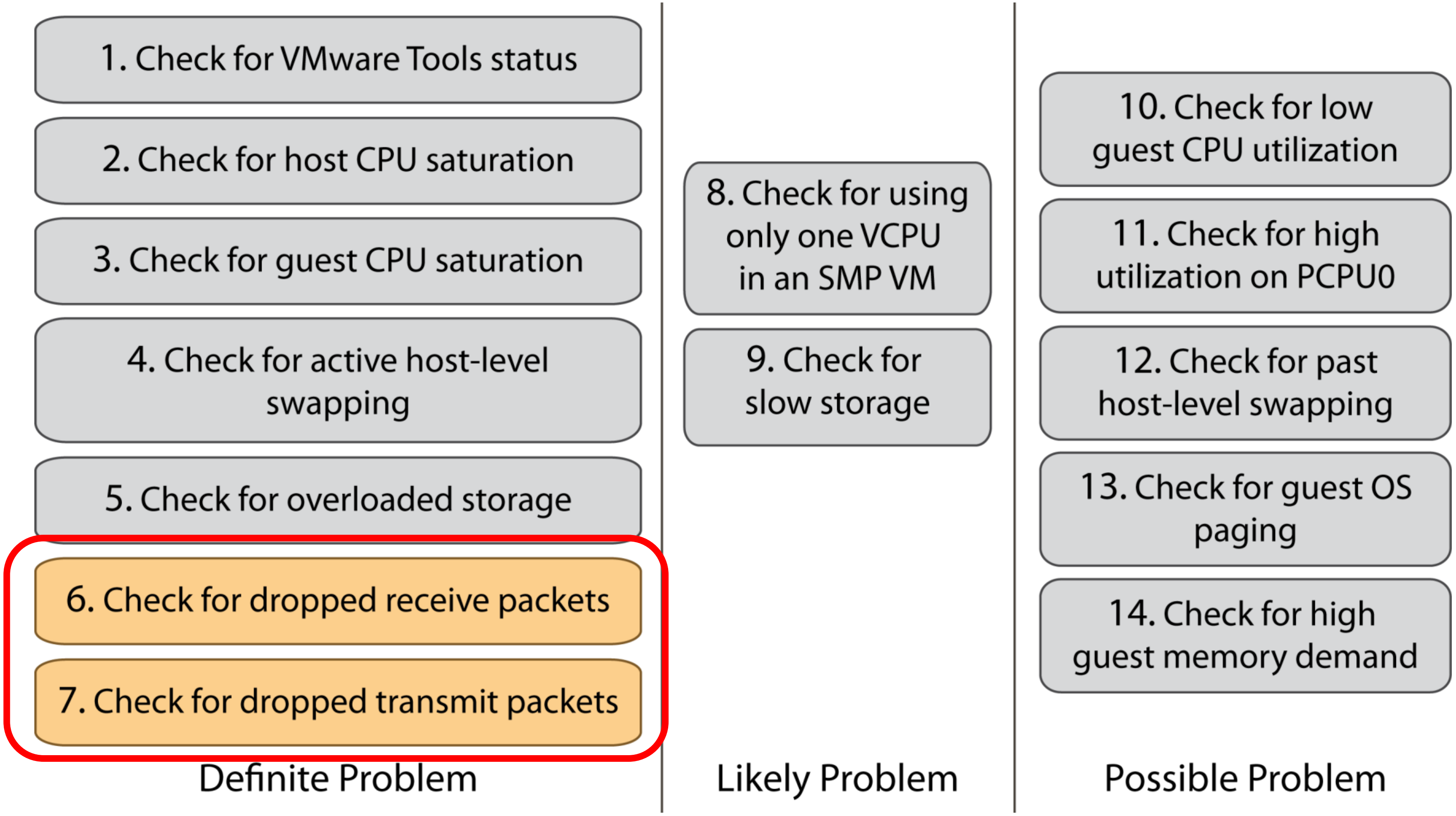
vSphere Client or resxtop/esxtop?



```
9:07:39pm up 1 day 19:33, 117 worlds; CPU load average: 0.38, 0.61, 0.57

  PORT-ID              USED-BY TEAM-PNIC DNAME                 PKTTX/s  MbTX/s
 16777217           Management       n/a vSwitch0                 0.00    0.00
 16777218               vmnic0         - vSwitch0                89.47    0.82
 16777219          4096:vswif0    vmnic0 vSwitch0                89.47    0.82
 16777221                 vmk0    vmnic0 vSwitch0                 0.00    0.00
 33554434           Management       n/a vSwitch1                 0.00    0.00
 33554435               vmnic1         - vSwitch1              1147.34    0.74
 33554436 4632:VM-for-StudentB    vmnic1 vSwitch1               556.08    0.35
 33554438 4961:VM-for-StudentA    vmnic1 vSwitch1               591.27    0.36
 33554439 4961:VM-for-StudentA    vmnic1 vSwitch1                 0.00    0.00
 50331649           Management       n/a vSwitch2                 0.00    0.00
 50331650               vmnic2         - vSwitch2                 0.00    0.00
 50331651                 vmk1    vmnic2 vSwitch2                 0.00    0.00
```

Review: Basic Troubleshooting Flow



1. Check for VMware Tools status
2. Check for host CPU saturation
3. Check for guest CPU saturation
4. Check for active host-level swapping
5. Check for overloaded storage
6. Check for dropped receive packets
7. Check for dropped transmit packets

Definite Problem

8. Check for using only one VCPU in an SMP VM
9. Check for slow storage

Likely Problem

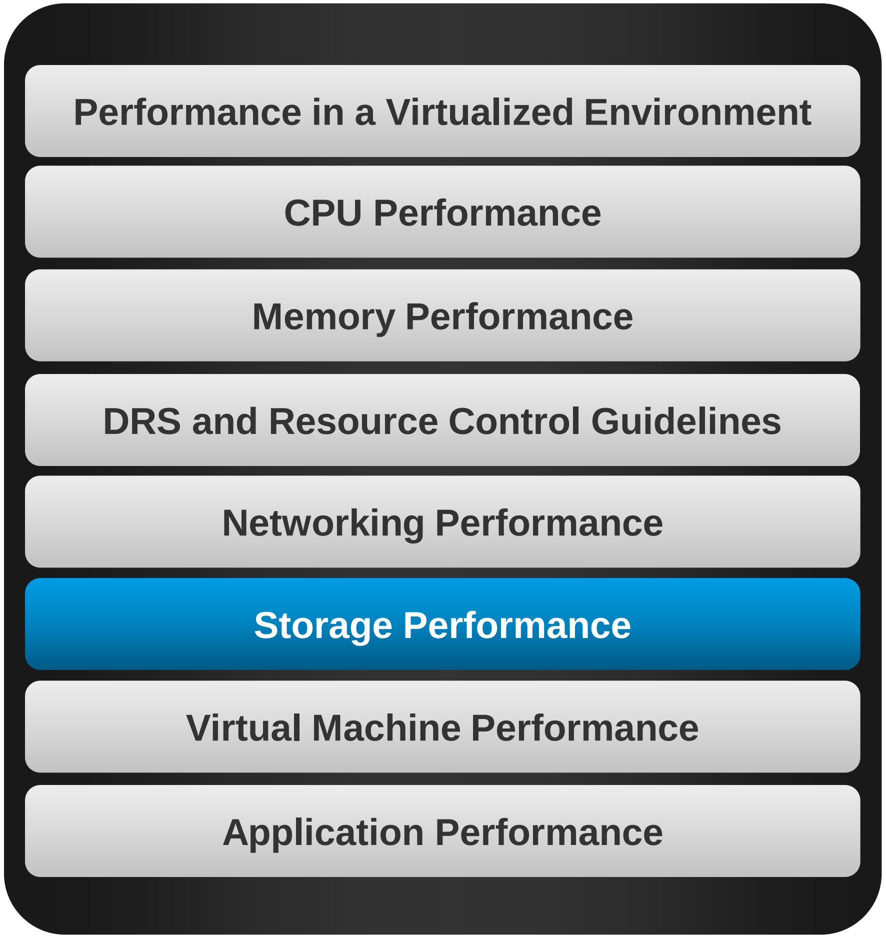
10. Check for low guest CPU utilization
11. Check for high utilization on PCPU0
12. Check for past host-level swapping
13. Check for guest OS paging
14. Check for high guest memory demand

Possible Problem

Networking Best Practices

- **Use the vmxnet3 network adapter where possible.**
  - Use vmxnet or vmxnet2 if vmxnet3 is not supported by the guest operating system.
- **Use a physical network adapter that supports high-performance features.**
- **Ensure that network adapters are running with full duplex and the highest supported speed.**
- **Team NICs for load balancing.**
- **Use separate NICs to avoid traffic contention.**
- **Run co-dependent virtual machines on the same host to take advantage of VMCI.**

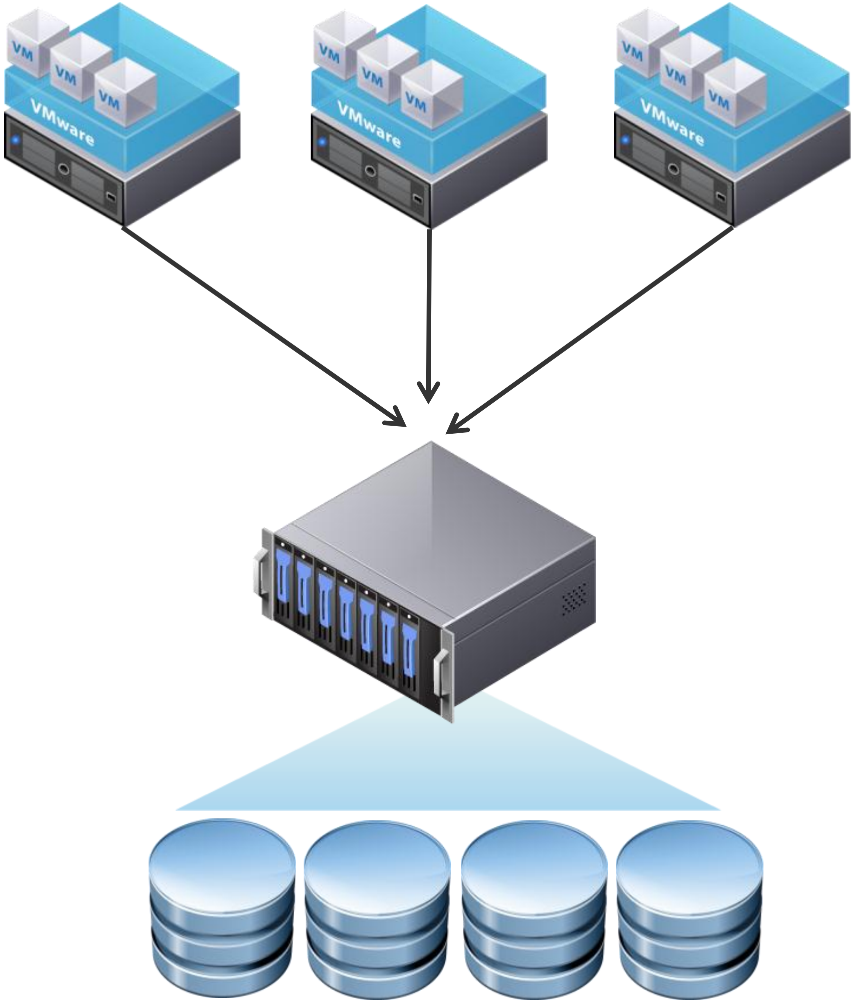
Performance Factors



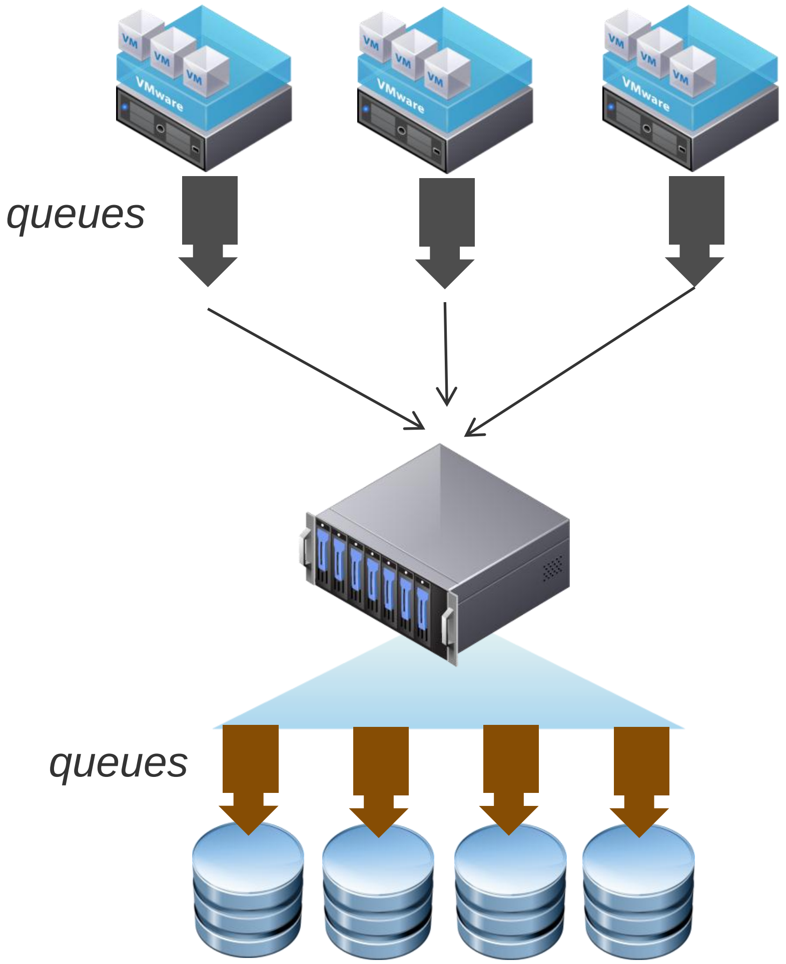
- Performance in a Virtualized Environment
- CPU Performance
- Memory Performance
- DRS and Resource Control Guidelines
- Networking Performance
- **Storage Performance**
- Virtual Machine Performance
- Application Performance

Storage Performance Overview

- **What affects storage performance?**
  - Storage protocols:
    - Fibre Channel, hardware iSCSI, software iSCSI, NFS
  - Proper storage configuration
  - Load balancing
  - Queuing and LUN queue depth
  - VMware® vStorage VMFS configuration:
    - Choosing between VMFS and RDMs
    - SCSI reservations
  - Virtual disk types

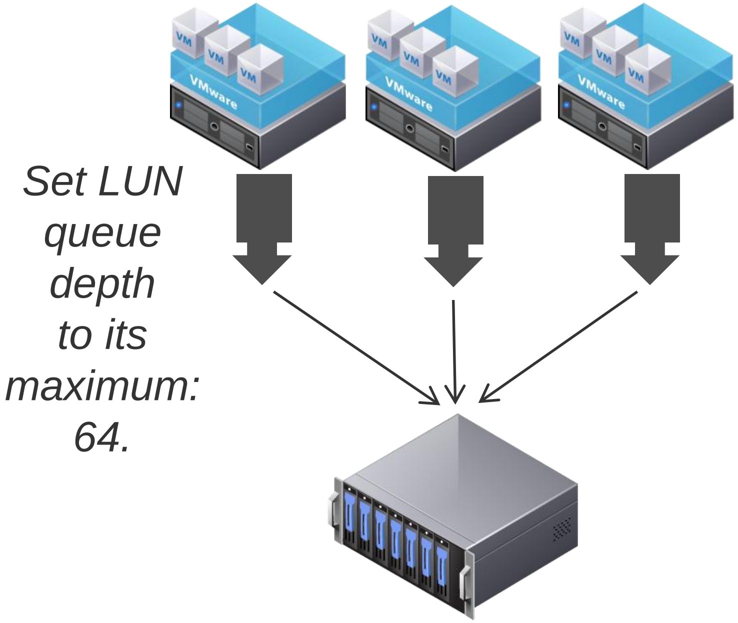


Storage Queues



- **Queuing at the host:**
  - Device queue controls number of active commands on LUN at any time.
    - Depth of queue is 32 (default).
  - VMkernel queue is an overflow queue for device driver queue.
- **Queuing at the storage array:**
  - Queuing occurs when the number of active commands to a LUN is too high for the storage array to handle.
- **Latency increases with excessive queuing at host or storage array.**

LUN Queue Depth



- **LUN queue depth determines how many commands to a given LUN can be active at one time.**
- **Set LUN queue depth size properly to decrease disk latency.**
  - Depth of queue is 32 (default).
  - Maximum recommended queue depth is 64.
- **Set Disk.SchedNumReqOutstanding to the same value as the queue depth.**

Network Storage: iSCSI and NFS

- **Avoid oversubscribing your links.**
  - Using VLANs does not solve the problem of oversubscription.
- **Isolate iSCSI traffic and NFS traffic.**
- **Applications that write a lot of data to storage should not share Ethernet links to a storage device.**
- **For software iSCSI and NFS, protocol processing uses CPU resources on the host.**

What Affects VMFS Performance?

- **VMFS partition alignment:**
  - VMware vSphere™ Client properly aligns a VMFS partition along the 64KB boundary.
  - Performance improvement is dependent on workloads and array types.
- **Spanning VMFS volumes:**
  - This is a great feature for increasing VMFS size dynamically.
  - Predicting performance is not straightforward.

SCSI Reservations

- **A SCSI reservation:**
  - Causes a LUN to be used exclusively by a single host for a brief period
  - Is used by a VMFS instance to lock the file system while the VMFS metadata is updated
- **Operations that result in metadata updates:**
  - Creating or deleting a virtual disk
  - Increasing the size of a VMFS volume
  - Creating or deleting snapshots
- **To minimize the impact on virtual machine performance:**
  - Postpone major maintenance/configuration until off-peak hours.

VMFS Versus RDMs

- **VMFS is the preferred option for most enterprise applications. Examples:**
  - Databases, ERP, CRM, VCB, Web servers, and file servers
- **RDM is preferred when raw disk access is necessary.**

I/O characteristic	Which yields better performance?
Random reads/writes	VMFS and RDM yield similar I/O operations/second
Sequential reads/writes at small I/O block sizes	VMFS and RDM yield similar performance
Sequential reads/writes at larger I/O block sizes	VMFS

Disk Capacity Metrics

- **Identify disk problems.**
  - Determine available bandwidth and compare with expectations.
- **What do I do?**
  - Check key metrics. In a VMware vSphere environment, the most significant statistics are:
    - Disk throughput
    - Latency (device, kernel)
    - Number of aborted disk commands
    - Number of active disk commands
    - Number of active commands queued

vscsiStats

- **`vscsiStats` collects and reports counters on storage activity:**
  - Data is collected at the virtual SCSI device level in the kernel.
  - Results are reported per VMDK, regardless of the underlying storage protocol.
- **Reports data in histogram form, which includes:**
  - I/O size
  - Seek distance
  - Outstanding I/Os
  - Latency in microseconds
- **Can be run in the ESX service console and in ESXi:**
  - ESXi: http://vpivot.com/2009/10/21/vscsistats-for-esxi/

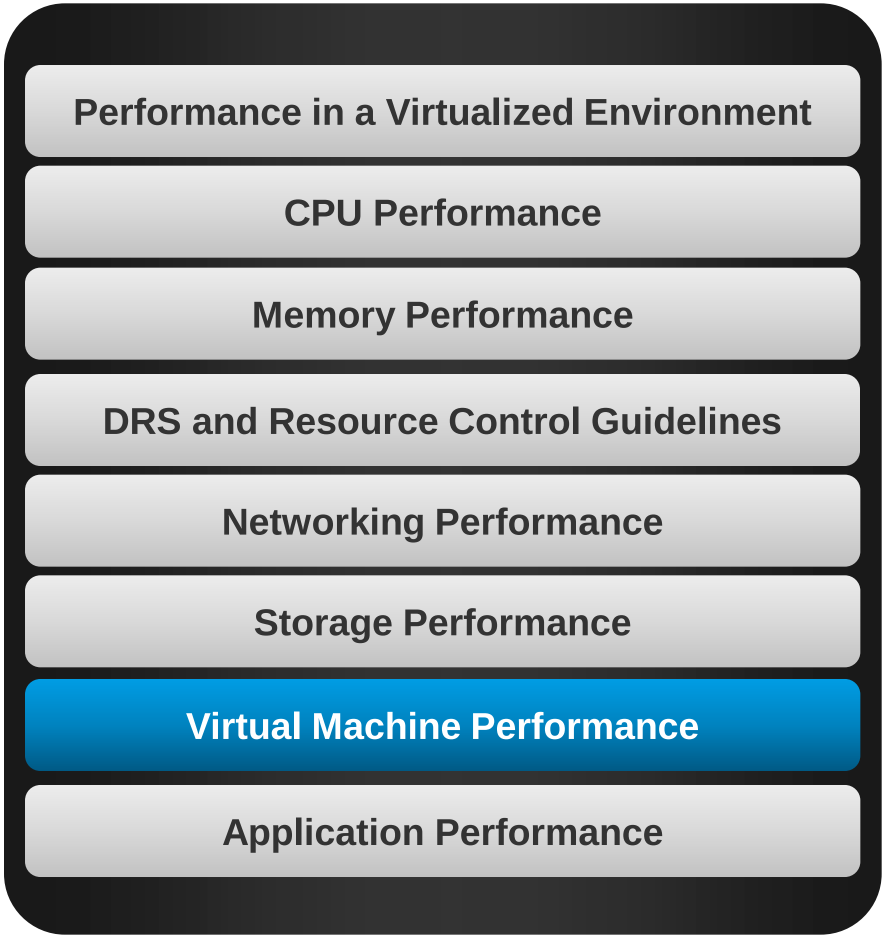
Why Use vscsiStats?

- **Provides NFS latency statistics for ESX 4.0 and earlier**
- **Reports data in histogram form instead of averages:**
  - Histograms of observed data values can be much more informative than single numbers like mean, median, and standard deviations from the mean.
- **Exposes sequentiality of I/O:**
  - This tool shows sequential versus random access patterns.
  - Sequentiality can help with storage sizing, LUN architecture, and identification of application-specific behavior.
- **Provides a breakdown of I/O sizes:**
  - Knowledge of expected I/O sizes can be used to optimize performance of the storage architecture.

Storage Performance Best Practices

- Configure each LUN with the right RAID level and storage characteristics for the applications in virtual machines that will use it.
- Use VMFS file systems for your virtual machines.
- Avoid oversubscribing paths (SAN) and links (iSCSI and NFS).
- Isolate iSCSI and NFS traffic.
- Applications that write a lot of data to storage should not share Ethernet links to a storage device.
- Postpone major storage maintenance until off-peak hours.
- Eliminate all possible swapping to reduce the burden on the storage subsystem.
- In SAN configurations, spread I/O loads over the available paths to the storage devices.
- Strive for complementary workloads.

Performance Factors



Performance in a Virtualized Environment

CPU Performance

Memory Performance

DRS and Resource Control Guidelines

Networking Performance

Storage Performance

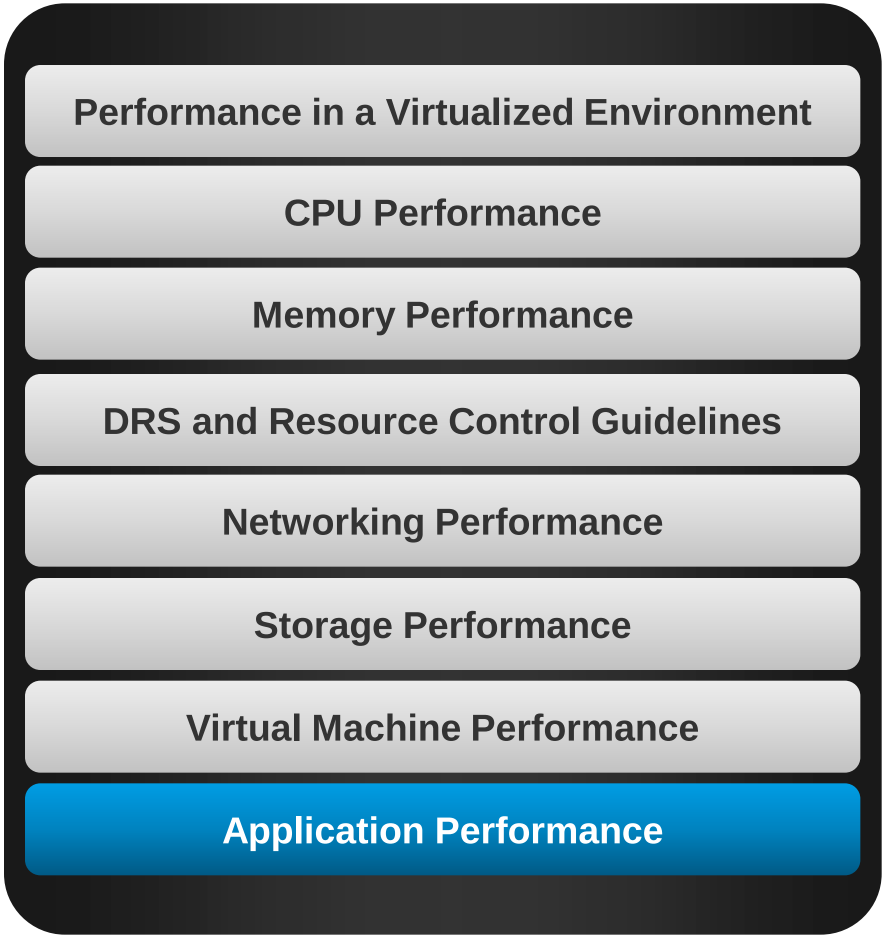
Virtual Machine Performance

Application Performance

Virtual Machine Performance Best Practices

- Select the right guest operating system type during virtual machine creation.
- Use 64-bit operating systems only when necessary.
- Do not deploy single-threaded applications in SMP virtual machines.
- Configure proper guest time synchronization.
- Install VMware Tools in the guest operating system and keep it up to date.
- Use the correct virtual hardware.
- Size the guest operating system queue depth appropriately.
- Be aware that large I/O requests split up by the guest might affect I/O performance.
- Align the guest operating system partitions.
- Use vmxnet3 where possible.
- Disable unused devices, for example, USB, CD-ROM, and floppy devices.
- NUMA locality

Performance Factors



Performance in a Virtualized Environment

CPU Performance

Memory Performance

DRS and Resource Control Guidelines

Networking Performance

Storage Performance

Virtual Machine Performance

Application Performance

Key Points

- Application performance problems are not always the result of resource capacity constraints.
- With AppSpeed, you can view performance from the perspective of the application and the end user: You can view transactions, dependencies, and response times.
- The building-block approach can help improve an application's scalability.
- The mix-and-match approach for building blocks can help improve an application's performance.