Database Performance in a Virtualized World

NoCOUG Winter Conference 2012

Eric Jenkinson

Agenda

- Organizational Challenges
- Types of virtualization
- CPU Scheduling and Resource Allocation
- What you as a DBA need to do to thrive in a virtualized environment

Organizational Issues in Virtualized Environments

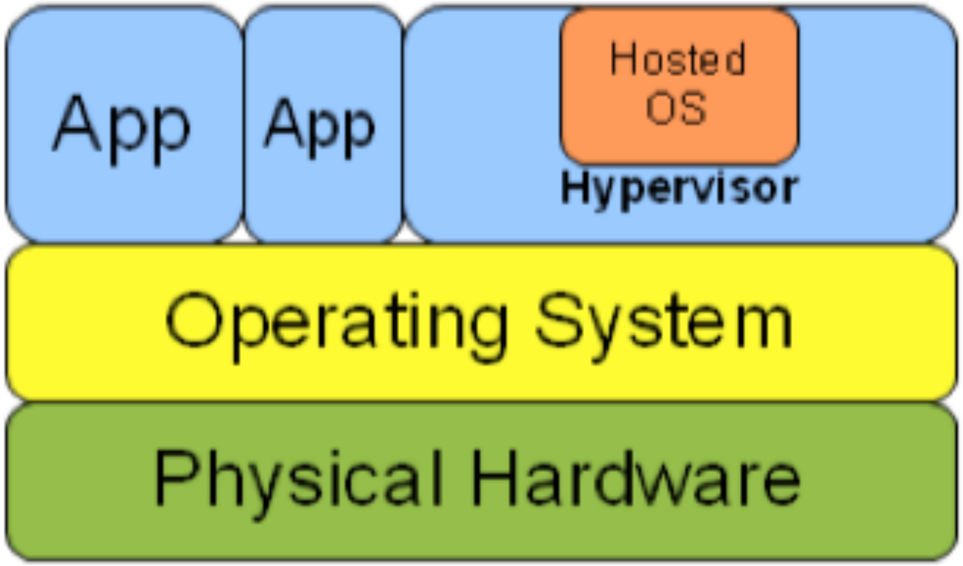
- Knowledge of virtualization technology is not well understood outside of the server group
- Limited visibility into the virtualization technology stack
- “Throwing Hardware at the Problem” is easier.
- VM Sprawl
- DBAs, especially Oracle DBAs, are less likely to adopt virtualization

Hypervisor

- Provides an abstraction of the physical hardware to the guests
- Manages the execution of the guest OS
- Manages the physical hardware resources
- Two hypervisor types
  - Type 2
  - Type 1

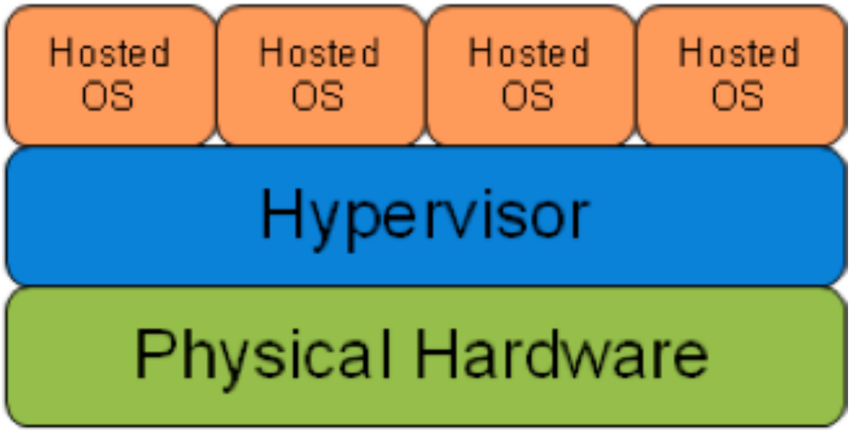
Type 2 Hypervisor

- Hosted
- Hypervisor runs as an application
- Hypervisor does not have direct control of hardware
- Examples
  - VMWare Server
  - Oracle VirtualBox



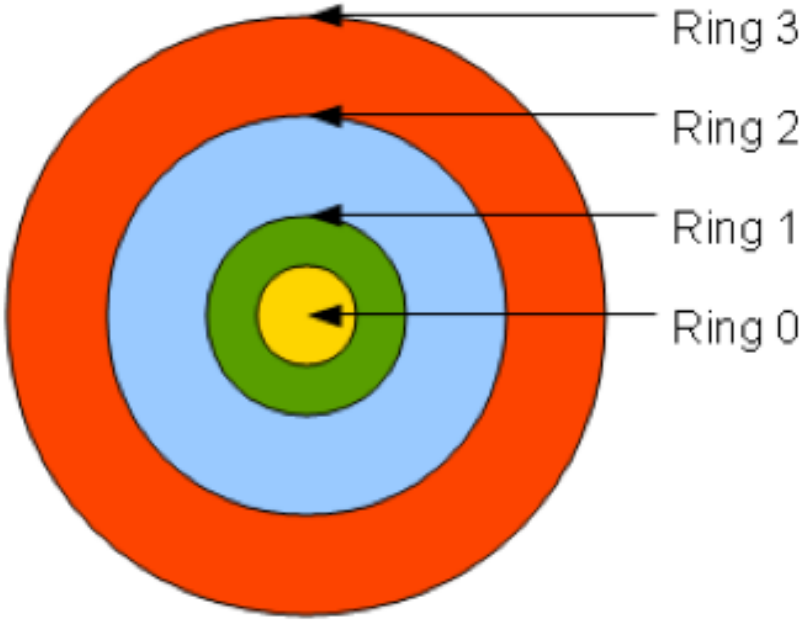
Type1 Hypervisor

- Bare metal
- Has full control of hardware
- Examples
  - VMWare ESX, ESXI
  - Oracle VM



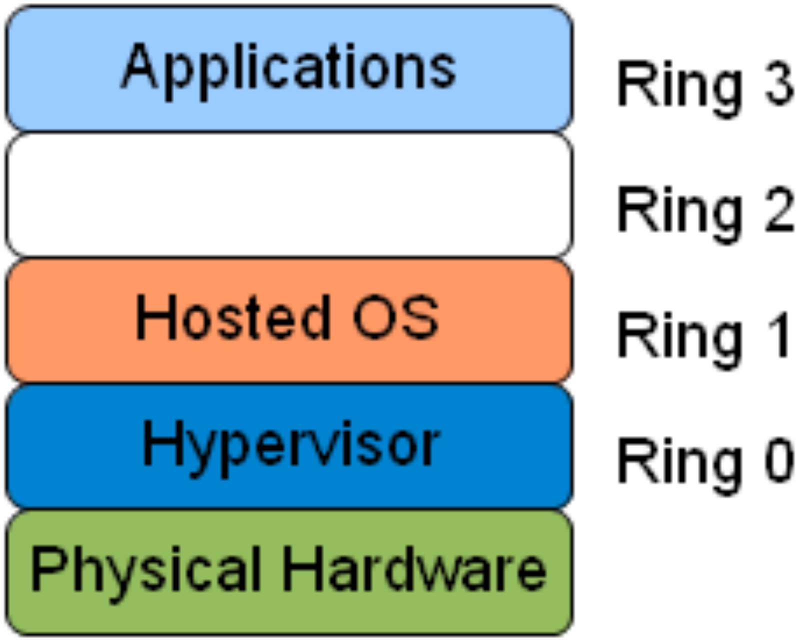
CPU Rings

- Ordered from most privileged (ring 0) to least privileged (ring 3)
- OS and device drivers operate in ring 0
- Applications run in ring 3



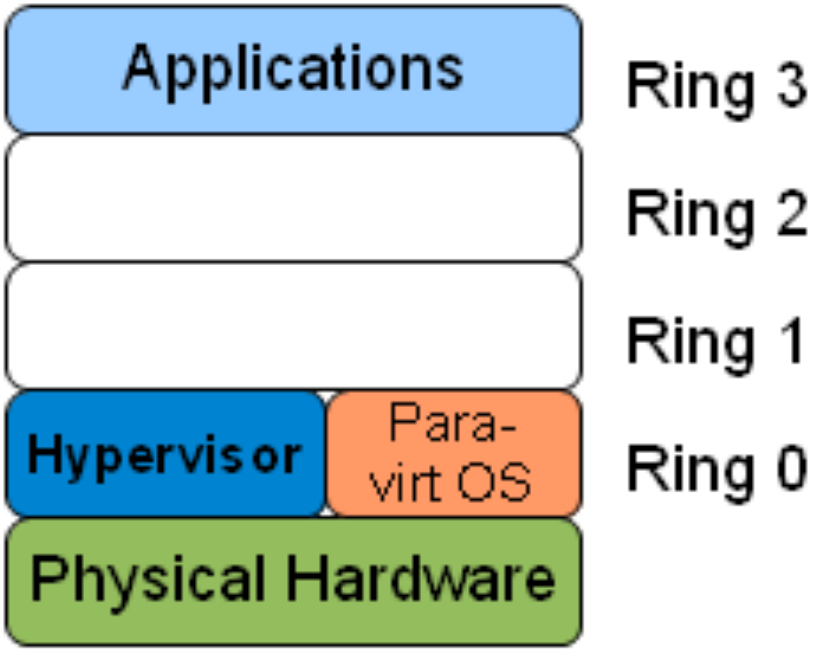
Full Virtualization

- Guest OS is unaware of virtualization
- Hypervisor traps privileged OS calls and reprocesses them
- Guest OS kernel is not modified
- Support
  - VMWare on old hardware



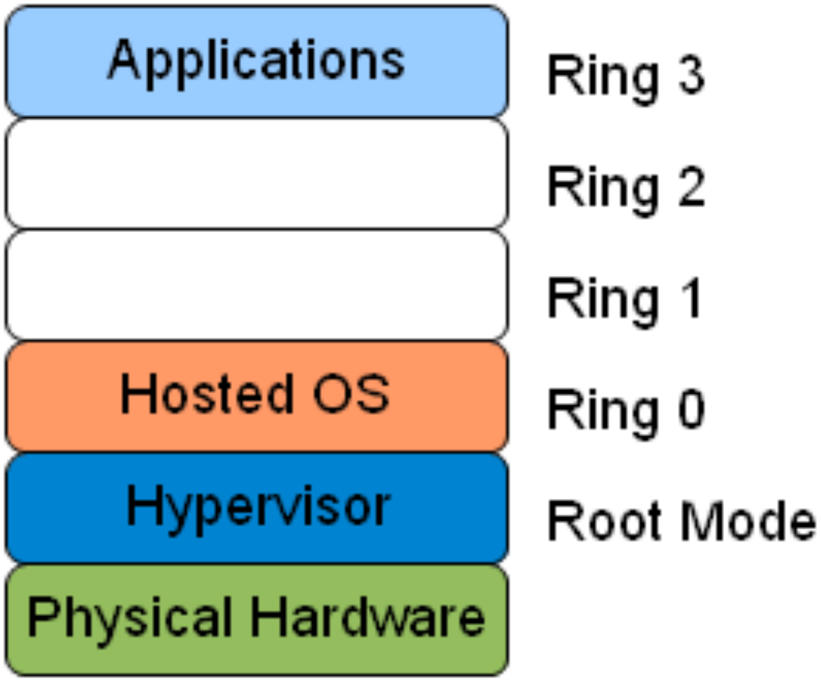
Paravirtualization

- Guest OS is aware of virtualization
- Guest OS Kernel modified to make hyper-calls instead of privileged calls
- Paravirtualization support in Linux Kernel 2.6.23 and higher



Hardware Assisted Virtualization

- CPU support for Virtualization
- Root Mode ring below ring 0
- Privileged calls trapped and sent hypervisor
- Guest OS does not need to be modified
- Support
  - VMWare and Oracle VM

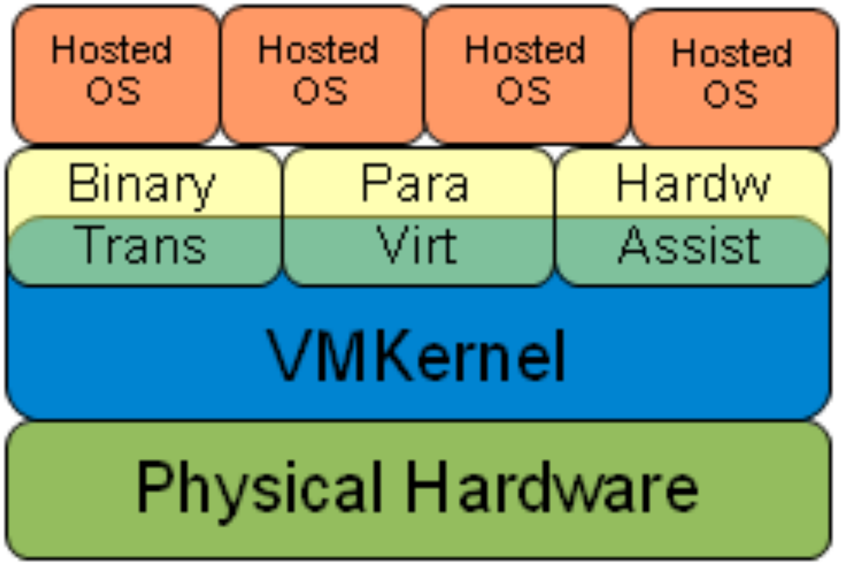


Hardware Assisted Virtualization with Paravirtualized Drivers

- Hybrid approach
- Guest OS does not have to be modified
- CPU virtualization support is required
- Paravirtualized drivers are required
- Support
  - VMWare and Oracle VM

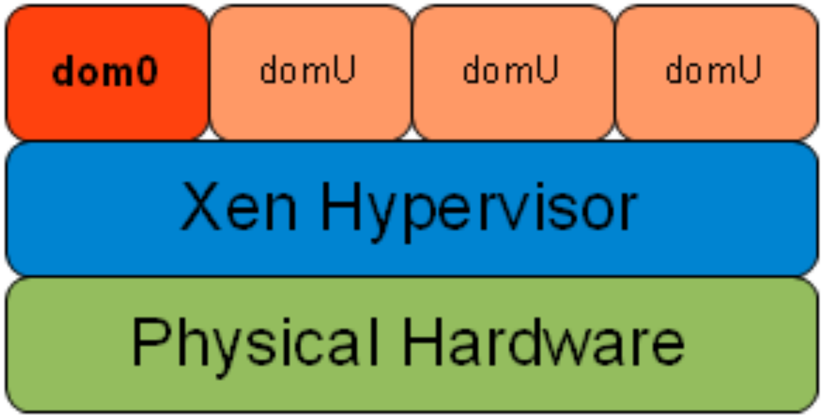
VMWare ESX Architecture

- VMKernel
  - CPU scheduler
  - Memory
  - Device Drivers
- Virtualization types
  - Binary Translation
  - Paravirtualized Drivers
  - Hardware Assisted Virtualization



Xen / Oracle VM Architecture

- Three components
  - Xen Hypervisor
  - Domain 0 (dom0)
  - Domain U (domU)
- Two domain types
  - Privileged
    - dom0
  - Unprivileged
    - domU (Guest VMs)



Resource Over Commitment

- Number of vCPUs can exceed the number of physical processors
- Sum of memory allocated to VM can exceed the amount of memory of the host
  - Oracle VM only through Max Memory

ESX CPU Relaxed Co-Scheduling

- A vCPU can be in one of three states
  - Waiting for a CPU to become available
  - Has a CPU and executing
  - Has a CPU and idle
- Relaxed CPU co-scheduling
  - Per vCPU
  - vCPUs that advance too much are individually stopped

ESX CPU Relaxed Co-Scheduling

- A vCPU is making progress if running or idle at the guest level
- Progress of each vCPU is tracked individually
  - Skew is measures as the difference between the slowest vCPU and other vCPUs
  - Skew does not grow if the vCPUs make equal progress during the co-scheduling period
- Skew enforcement
  - vCPUs that advance too much are stopped once the skew is reduced the stopped vCPUs may start individually

Oracle VM – Xen Credit Scheduler

- Proportional fair CPU scheduler
- Each domain is assigned a weight and cap
  - Weight: a domain with 256 received twice as much CPU as a domain with 128
  - Cap maximum amount of CPU a domain can consume even if there are idle CPUs
- Automatically load balances vCPU across all available CPUs on SMP host

Oracle VM – Xen Credit Scheduler

- Each physical CPU manages a run queue of vCPUs sorted by priority
  - Priority UNDER and OVER
- When inserting a vCPU to a queue it is put after all vCPU of equal priority
- When vCPU runs it consume credits. Until all credits are consumed its priority is UNDER
- Fair CPU scheduling, I/O can be skewed

Memory Management

- Transparent Page Sharing
- Memory Ballooning
  - Requires Guest Additions to be installed
- Memory Compression
  - Compress memory pages that need to be swapped to disk
- ESX Swapping – Demand Paging
- Oracle VM only has memory ballooning at this time
  - Page sharing and demand paging is in Xen unstable

Distributed Resource Scheduling (DRS)

- Both VMWare and Oracle VM have the ability to move VMs across physical servers
- The goal is to provide consistent resources to running VMs
- Moves VMs from heavily loaded servers to servers with a lighter load
- With out rules or affinity groups in place, DRS can be a source of “random” performance issues

So what am I as as a DBA supposed to do with this?

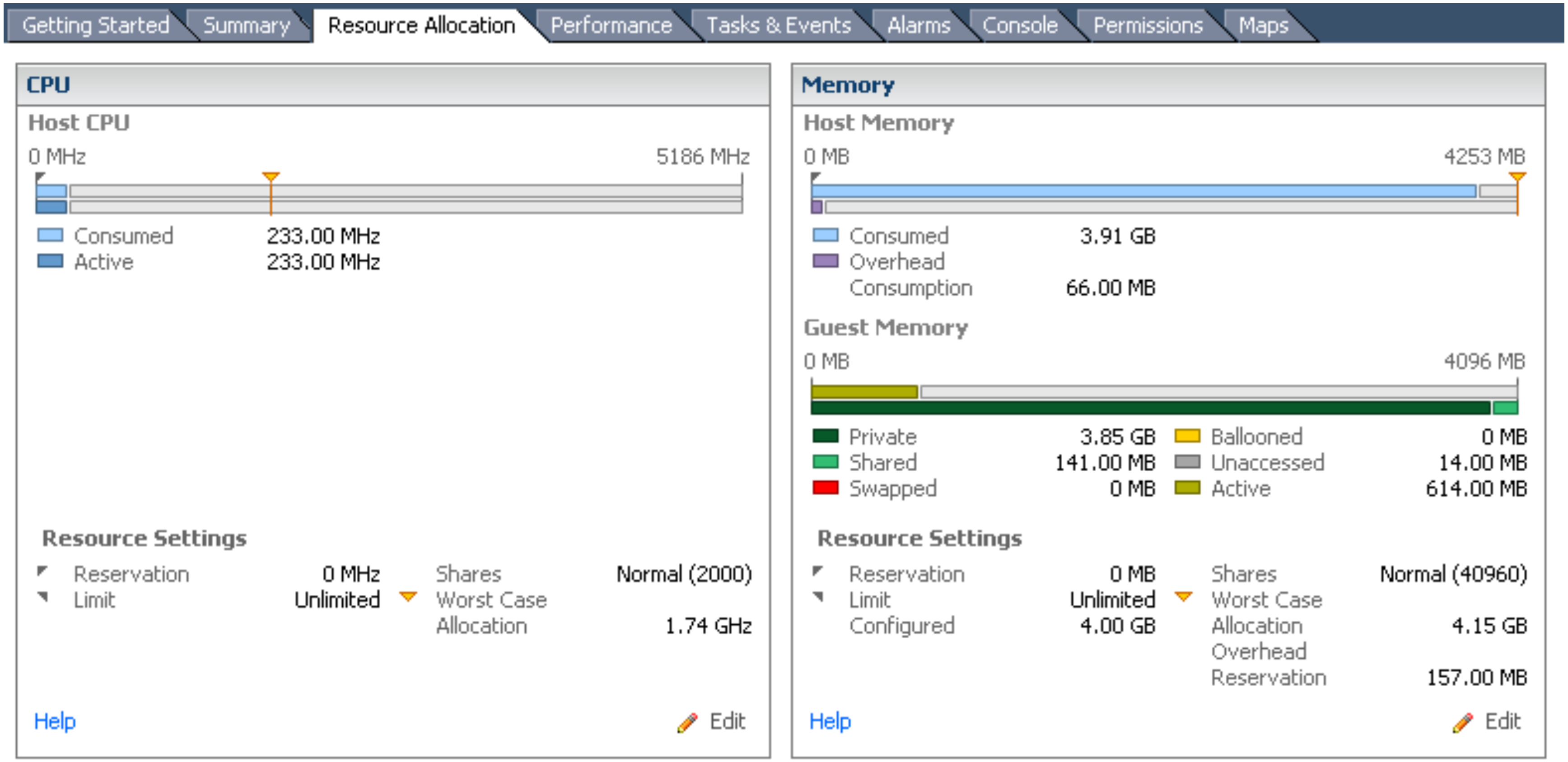
My Experience

- Most performance problems fall under these areas
  - Poor knowledge of the virtualization stack
  - Poor or no VM placement policies
  - Poor or no resource prioritization
  - Little to no visibility into the virtualization stack
- The rest are the same problems that can exist in a purely physical environment

Visibility is a Must

- VMWare: Request a read only account in vSphere
- Oracle VM: Oracle Enterprise Manager
  - Can be a problem with Oracle VM 2
- VMWare: esxtop OVM: top, vmstat, sar with paravirtualized kernel
- Third Party tools
  - Quest Spotlight for Oracle
  - Confio Ignite for VM

Recognize Default Settings



CPU

Host CPU

0 MHz — 5186 MHz

Consumed	233.00 MHz
Active	233.00 MHz

Resource Settings

Reservation	0 MHz	Shares	Normal (2000)
Limit	Unlimited	Worst Case Allocation	1.74 GHz

Memory

Host Memory

0 MB — 4253 MB

Consumed	3.91 GB
Overhead Consumption	66.00 MB

Guest Memory

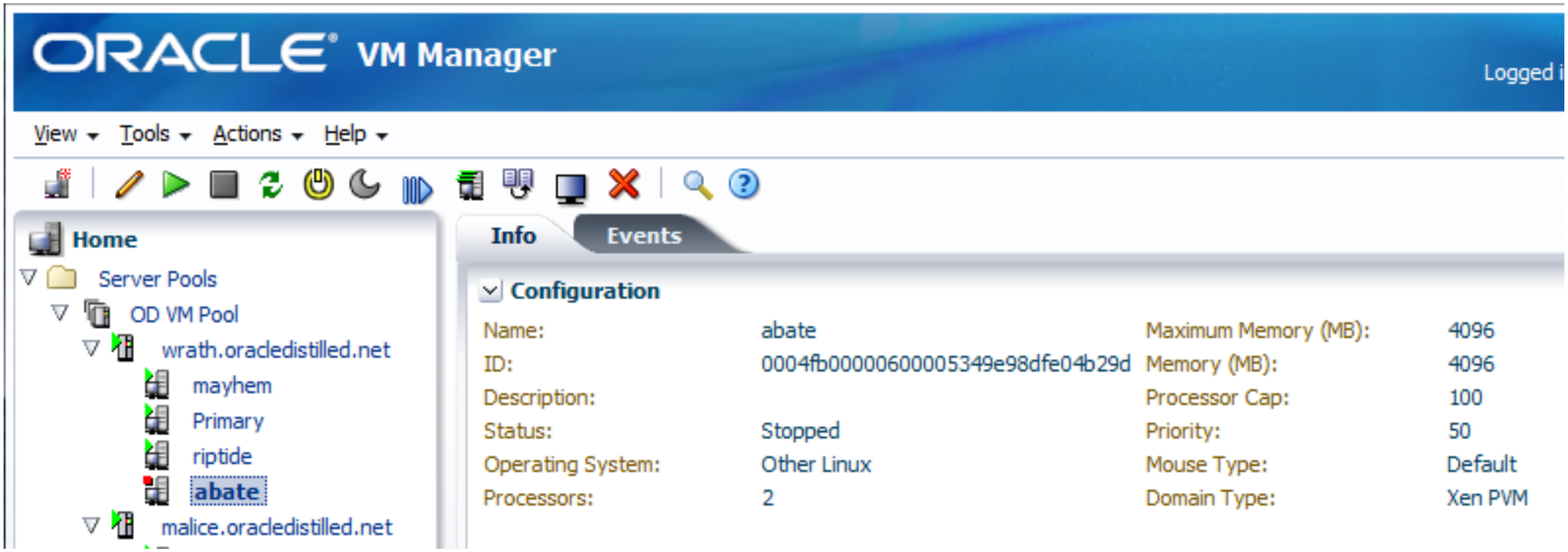
0 MB — 4096 MB

Private	3.85 GB	Ballooned	0 MB
Shared	141.00 MB	Unaccessed	14.00 MB
Swapped	0 MB	Active	614.00 MB

Resource Settings

Reservation	0 MB	Shares	Normal (40960)
Limit	Unlimited	Worst Case Allocation	4.15 GB
Configured	4.00 GB	Overhead Reservation	157.00 MB

Recognize Default Settings



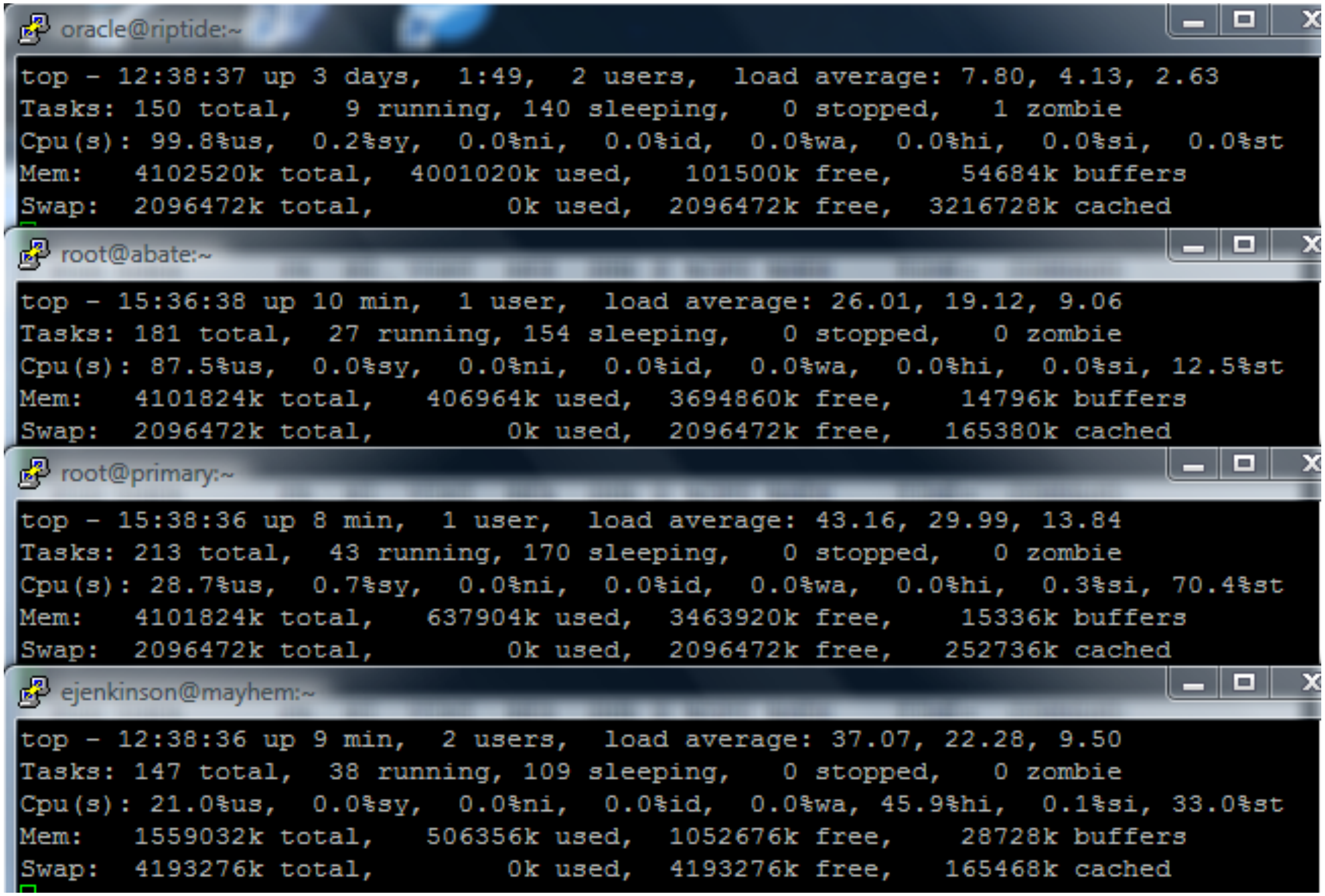
Configuration

Name:	abate	Maximum Memory (MB):	4096
ID:	0004fb00000600005349e98dfe04b29d	Memory (MB):	4096
Description:		Processor Cap:	100
Status:	Stopped	Priority:	50
Operating System:	Other Linux	Mouse Type:	Default
Processors:	2	Domain Type:	Xen PVM

Recognize Default Settings

- VMWare (CPU and Memory)
  - Reservation: minimum amount allocated/available at VM power on
  - Limit: maximum amount of the resource
  - Shares: Priority in acquiring the resource
- Oracle VM (CPU only)
  - Priority: The higher the priority, the more physical CPU cycles given to the VM
  - Processor Cap: The maximum amount of CPU a VM can consume

The Problem with Defaults/Unlimited



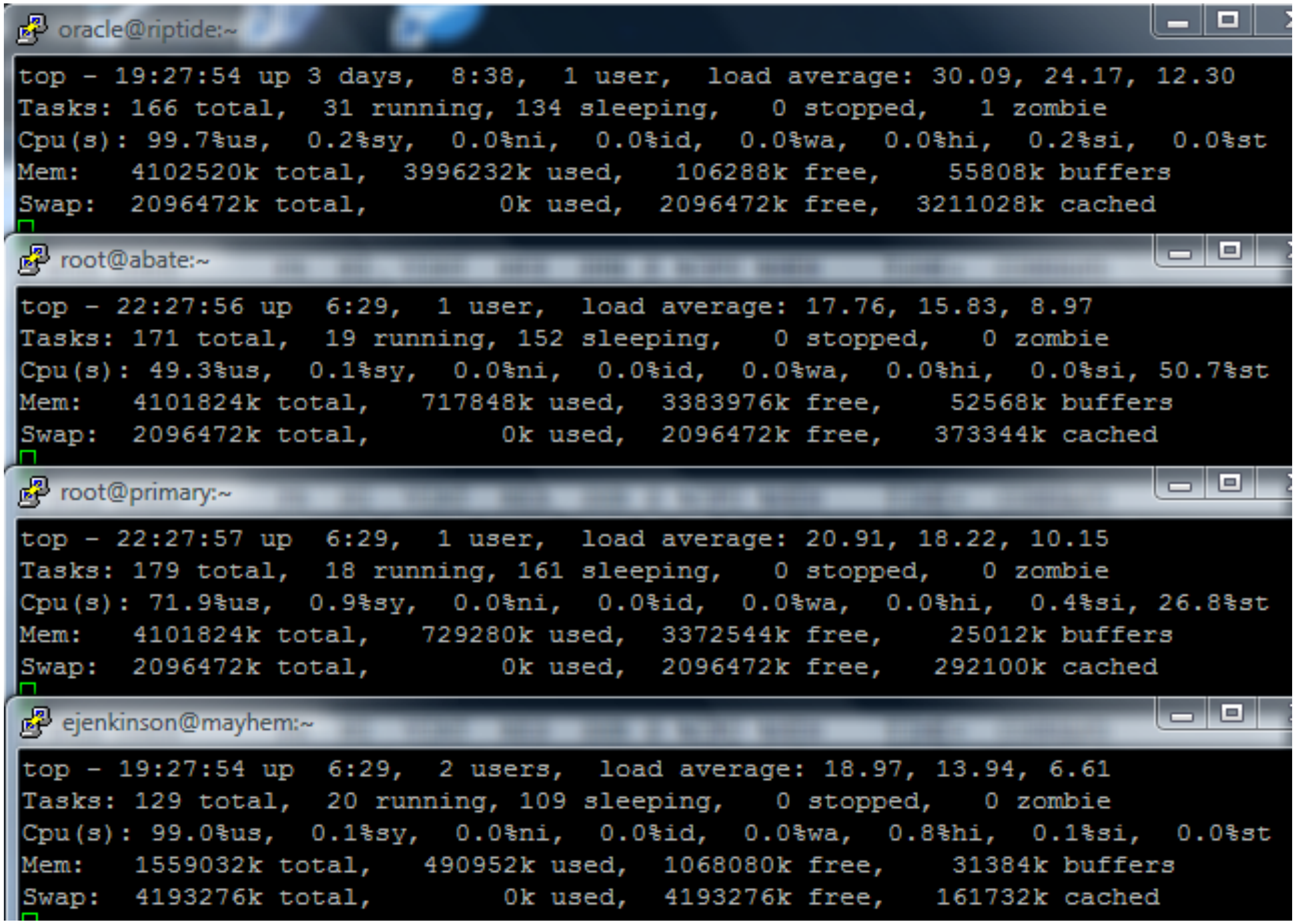
```
top - 12:38:37 up 3 days,  1:49,  2 users,  load average: 7.80, 4.13, 2.63
Tasks: 150 total,   9 running, 140 sleeping,   0 stopped,   1 zombie
Cpu(s): 99.8%us,  0.2%sy,  0.0%ni,  0.0%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Mem:   4102520k total,  4001020k used,   101500k free,    54684k buffers
Swap:  2096472k total,        0k used,  2096472k free,  3216728k cached
```

```
top - 15:36:38 up 10 min,  1 user,  load average: 26.01, 19.12, 9.06
Tasks: 181 total,  27 running, 154 sleeping,   0 stopped,   0 zombie
Cpu(s): 87.5%us,  0.0%sy,  0.0%ni,  0.0%id,  0.0%wa,  0.0%hi,  0.0%si, 12.5%st
Mem:   4101824k total,   406964k used,  3694860k free,    14796k buffers
Swap:  2096472k total,        0k used,  2096472k free,   165380k cached
```

```
top - 15:38:36 up 8 min,  1 user,  load average: 43.16, 29.99, 13.84
Tasks: 213 total,  43 running, 170 sleeping,   0 stopped,   0 zombie
Cpu(s): 28.7%us,  0.7%sy,  0.0%ni,  0.0%id,  0.0%wa,  0.0%hi,  0.3%si, 70.4%st
Mem:   4101824k total,   637904k used,  3463920k free,    15336k buffers
Swap:  2096472k total,        0k used,  2096472k free,   252736k cached
```

```
top - 12:38:36 up 9 min,  2 users,  load average: 37.07, 22.28, 9.50
Tasks: 147 total,  38 running, 109 sleeping,   0 stopped,   0 zombie
Cpu(s): 21.0%us,  0.0%sy,  0.0%ni,  0.0%id,  0.0%wa, 45.9%hi,  0.1%si, 33.0%st
Mem:   1559032k total,   506356k used,  1052676k free,    28728k buffers
Swap:  4193276k total,        0k used,  4193276k free,   165468k cached
```

The Problem with Defaults/Unlimited



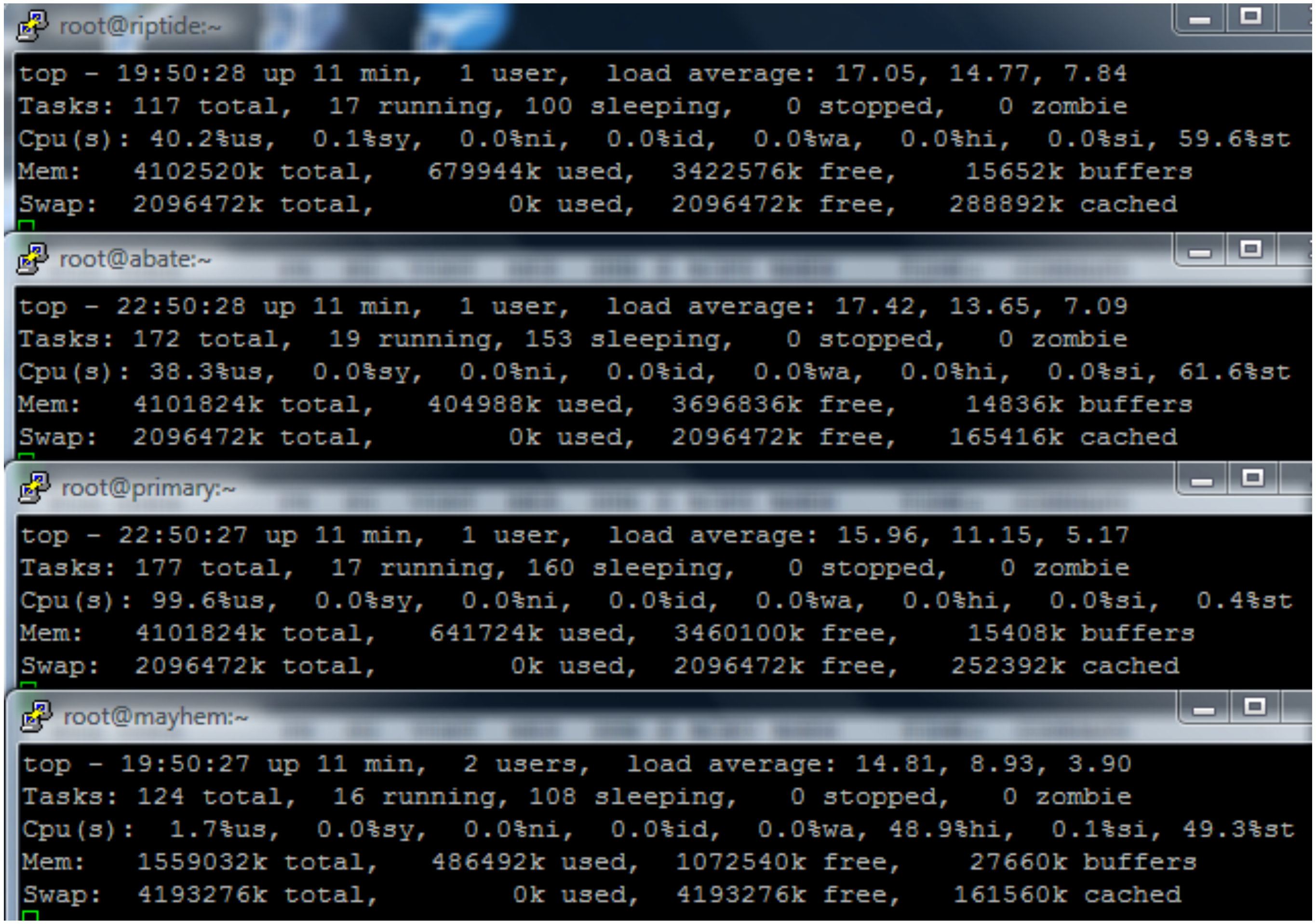
```
top - 19:27:54 up 3 days,  8:38,  1 user,  load average: 30.09, 24.17, 12.30
Tasks: 166 total,  31 running, 134 sleeping,   0 stopped,   1 zombie
Cpu(s): 99.7%us,  0.2%sy,  0.0%ni,  0.0%id,  0.0%wa,  0.0%hi,  0.2%si,  0.0%st
Mem:   4102520k total,  3996232k used,   106288k free,    55808k buffers
Swap:  2096472k total,        0k used,  2096472k free,  3211028k cached
```

```
top - 22:27:56 up  6:29,  1 user,  load average: 17.76, 15.83, 8.97
Tasks: 171 total,  19 running, 152 sleeping,   0 stopped,   0 zombie
Cpu(s): 49.3%us,  0.1%sy,  0.0%ni,  0.0%id,  0.0%wa,  0.0%hi,  0.0%si, 50.7%st
Mem:   4101824k total,   717848k used,  3383976k free,    52568k buffers
Swap:  2096472k total,        0k used,  2096472k free,   373344k cached
```

```
top - 22:27:57 up  6:29,  1 user,  load average: 20.91, 18.22, 10.15
Tasks: 179 total,  18 running, 161 sleeping,   0 stopped,   0 zombie
Cpu(s): 71.9%us,  0.9%sy,  0.0%ni,  0.0%id,  0.0%wa,  0.0%hi,  0.4%si, 26.8%st
Mem:   4101824k total,   729280k used,  3372544k free,    25012k buffers
Swap:  2096472k total,        0k used,  2096472k free,   292100k cached
```

```
top - 19:27:54 up  6:29,  2 users,  load average: 18.97, 13.94, 6.61
Tasks: 129 total,  20 running, 109 sleeping,   0 stopped,   0 zombie
Cpu(s): 99.0%us,  0.1%sy,  0.0%ni,  0.0%id,  0.0%wa,  0.8%hi,  0.1%si,  0.0%st
Mem:   1559032k total,   490952k used,  1068080k free,    31384k buffers
Swap:  4193276k total,        0k used,  4193276k free,   161732k cached
```

The Problem with Defaults/Unlimited



```
top - 19:50:28 up 11 min,  1 user,  load average: 17.05, 14.77, 7.84
Tasks: 117 total,  17 running, 100 sleeping,   0 stopped,   0 zombie
Cpu(s): 40.2%us,  0.1%sy,  0.0%ni,  0.0%id,  0.0%wa,  0.0%hi,  0.0%si, 59.6%st
Mem:   4102520k total,   679944k used,  3422576k free,    15652k buffers
Swap:  2096472k total,        0k used,  2096472k free,   288892k cached
```

```
top - 22:50:28 up 11 min,  1 user,  load average: 17.42, 13.65, 7.09
Tasks: 172 total,  19 running, 153 sleeping,   0 stopped,   0 zombie
Cpu(s): 38.3%us,  0.0%sy,  0.0%ni,  0.0%id,  0.0%wa,  0.0%hi,  0.0%si, 61.6%st
Mem:   4101824k total,   404988k used,  3696836k free,    14836k buffers
Swap:  2096472k total,        0k used,  2096472k free,   165416k cached
```

```
top - 22:50:27 up 11 min,  1 user,  load average: 15.96, 11.15, 5.17
Tasks: 177 total,  17 running, 160 sleeping,   0 stopped,   0 zombie
Cpu(s): 99.6%us,  0.0%sy,  0.0%ni,  0.0%id,  0.0%wa,  0.0%hi,  0.0%si,  0.4%st
Mem:   4101824k total,   641724k used,  3460100k free,    15408k buffers
Swap:  2096472k total,        0k used,  2096472k free,   252392k cached
```

```
top - 19:50:27 up 11 min,  2 users,  load average: 14.81, 8.93, 3.90
Tasks: 124 total,  16 running, 108 sleeping,   0 stopped,   0 zombie
Cpu(s):  1.7%us,  0.0%sy,  0.0%ni,  0.0%id,  0.0%wa, 48.9%hi,  0.1%si, 49.3%st
Mem:   1559032k total,   486492k used,  1072540k free,    27660k buffers
Swap:  4193276k total,        0k used,  4193276k free,   161560k cached
```

Virtual CPU Recommendations

- Set values for Limit/Shares or Priority/Cap to match business value
- Use only the vCPUs required and no more
- Monitor stolen time (OVM) and ESX Ready time to ascertain competition between VMs
- Watch out for CPU over commitment with VM that have many vCPUs

Virtual Memory Recommendations

- VMWare
  - Use Reservation to avoid ballooning and swapping
    - SGA + PGA + processes overhead
  - Ensure VMWare Tools are installed (and up to date) to provide ballooning
- Oracle VM
  - Set Memory = SGA + PGA + process overhead

Storage I/O

- Avoid sparse or dynamic growth virtual disks
- Follow Oracle and Storage vendor’s best practices for Oracle Databases
- Use Storage IO Control (VMWare) to prioritize VM access to datastore
- Use dedicated datastores (VMWare) to avoid sharing disk workloads

Network I/O

- Avoid having multiple high storage I/O VMs on the same physical host
- Insure paravirtualized drivers are installed
- Host server should have 1Gb min 10Gb recommended network adapter

When Requesting a VM

- Request paravirtulaized drivers / VMWare tools to be installed
- Plan CPU and memory requirements and priority
  - Avoid “Cookie Cutter VMs”
- Know the business importance of this database
- Find where the VM is going to be placed and who its neighbors are

Questions/Answers

Thank you!