PAC094
Performance Tips for New Features in Workstation 5

Anne Holler
Irfan Ahmad
Aravind Pavuluri
Overview of Talk

- Virtual machine teams
- 64-bit guests
- SMP guests
- e1000 NIC support
- Fast snapshots
- Virtual machine disk performance characteristics
- Full and linked clones
- Virtual machine performance measurement tips
This presentation may contain VMware confidential information.

Copyright © 2005 VMware, Inc. All rights reserved. All other marks and names mentioned herein may be trademarks of their respective companies.
New: Virtual Machine Teams

- Feature definition:
  - A set of virtual machines that are managed together
  - Development and test of multi-system workloads
- Key performance areas:
  - CPU, memory, networking, and disk
- Example workload: SPECjAppServer2004
Virtual Machine Teams

CPU

- Key concepts
  - CPU is allocated across virtual machines fairly (within constraints of host scheduler policies)
  - Need enough CPU to run all virtual machines in team
  - Use host CPU utilization tools to check CPU usage
  - If CPU pegged at 100% usage, consider
    - Using faster machine
    - Reducing or offloading some virtual machine load
    - Checking whether overhead from memory or networking issues
Virtual Machine Teams

Memory

- Key concepts
  - Trimming: pages removed from guest for system reuse
  - New: Page Sharing: 1 COW page for same content in different virtual machines or the same virtual machine
    - Transparent; Isolation is maintained
  - Memory for each virtual machine includes guest RAM + overhead memory for virtualization
  - Over-commit: memory used by virtual machines exceeds VMware reserved memory
Virtual Machine Teams

Memory

- Performance guidelines
  - Avoid swapping
  - Ensure adequate physical memory on host, more is better
  - Right-size guest memory
    - Not too little, but also, not too much
  - Guest workloads with high disk I/O: disable page sharing, trimming
    - Refer to Workstation 5 User’s Manual for details
Virtual Machine Teams
Networking

- Features
  - New: Virtual machine NIC morphing
    - (vlance turns into vmxnet)
  - New: Network bandwidth limiting
Virtual Machine Teams
Networking

- Performance guidelines
  - Gigabit networking is CPU-intensive
    - Native: can consume 30%+ 2.4GHz P4 to saturate link
    - Virtualization adds CPU overhead
  - For heavy virtual machine intra-team networking
    - Consider virtual machine team bandwidth limiting, if CPU used for networking is too high relative to other parts of workload
  - For heavy virtual machine bridged networking workloads
    - Consider tuning host NIC interrupt coalescing policies
      [ref: WS5 User’s Manual]
Virtual Machine Team Case
SPECjAppSrvr04

- SPECjAppServer2004 is end-to-end workload comprising all major J2EE components
  - Ambitious workload to run on a desktop system
  - Public disclosures for clustered systems of 6+ CPUs
- Setup:
  - Host was dual 2.4GHz P4 HToff w/4GB running WS2003; Team was comprised of the following guests:
    - WS2003 1.5GB virtual machine w/Weblogic App Server 8.1
    - WS2003 1.5GB virtual machine w/Oracle 9i DB Server
    - External client: 2.4GHz P4 WS2003 w/2GB
Virtual Machine Team Case
SPECjAppSrvr04

Performance vs. native:
- For InjRate=1, CPU is not oversubscribed
- For InjRate=3, CPU is oversubscribed
**New: 64-bit Guests**

- First commercial virtualization vendor to support 64-bit guests
- Full support on AMD Athlon 64 Rev “D” and Opteron Rev “E” or later
- Experimental support for VT
- Most applications should run right out of the box with optimal performance
- Competitive VT performance

**x64 vs x32 Applications Performance on 64bit Guest** (higher is better)

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>% Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>getpid</td>
<td>15%</td>
</tr>
<tr>
<td>forkwait</td>
<td>20%</td>
</tr>
<tr>
<td>nativeswitch</td>
<td>10%</td>
</tr>
<tr>
<td>divzero</td>
<td>35%</td>
</tr>
</tbody>
</table>

**TIP:** For optimal performance, recompile applications for 64-bit
New: SMP Guests

- Workstation fully supports two-way virtual SMP on all supported 32- and 64-bit configurations. Host must be at least two-way.
- To get good scaling in an SMP virtual machine, an application should be multi-threaded, with more than 1 worker thread. [Single-threaded applications should be run UP.] No SMP performance tweaks needed.
- CPU over-commit scenarios perform better on ESX Server than WS
New: e1000 Guest Device

- Widely-supported Intel Pro/1000 MT Single-Port Server Adapter is the default NIC emulation for 64-bit guests
- e1000 incorporates clustering, and checksum and TCP segmentation offloading for better performance
- Performance on par with the vmxnet device

*vmxnet vs. e1000 NetPerf comparison*
New: Fast Snapshots

- A snapshot creates a copy of a running virtual machine’s context, fast snapshots seem instantaneous.
- A fast snapshot marks the virtual machine’s context COW, saves it to disk in background.
- Performance guidelines:
  - Need sufficiently fast local [non-network] disk.
  - Need spare host memory.
  - Need adequate free host disk space; in general, disk operations are slow on an almost full disk.
Overview of Talk

- Virtual machine teams
- 64Bit guests
- SMP guests
- e1000 NIC support
- Fast snapshots
- Virtual machine disk performance characteristics
- Full and linked clones
- Virtual machine performance measurement tips
Disk Basics
Buffered Access

- Normal filesystem operations can use buffered access
  - e.g. a write of 64KB on NTFS gets absorbed by a buffer cache
- A system crash would result in data loss
- Example is an application I wrote which writes a 2GB file one 64KB IO at a time
- The perf difference can be huge
- Unsafe for data integrity sensitive apps to use buffered access
  - e.g. databases, VMware Workstation

Test performed on a single disk SCSI Drive
Seagate 36GB 10K RPM
Windows 2003 Enterprise Edition
Disk Basics

Unbuffered IO

- How to optimize IO performance while still being safe?
- Issue many IOs at a time
- Having just two outstanding IOs compensates for the lack of buffering
- Multiple IOs allow the kernel or disk itself to optimize the IO
- Data integrity sensitive apps use Unbuffered IO

Test performed on a single disk SCSI Drive
Seagate 36GB 10K RPM
Windows 2003 Enterprise Edition
Iometer - 64K Sequential Writes
Disk Basics
Issue Multiple IOs (contd.)

- Trend from previous slide also exists for RAID arrays
  - Not just an artifact of my disk
  - Iometer with 64K sequential writes but on a 14-disk RAID0 SAN
- More outstanding IOs needed to match buffered due to higher level of available parallelism on the underlying device
- VMware Workstation 5 can safely issue multiple IOs from a guest OS simultaneously (unbuffered)
- All disk data presented so far has been native

Test performed on an EMC CX700 SAN (14-disk RAID0 LUN)
Windows 2003 Enterprise Edition
Iometer - 64K Sequential Writes
Workstation 5 Disk IO

- VMware Workstation issues IOs to a disk on behalf of a guest operating system
- Issue multiple IOs concurrently if the guest does
- Workstation IO throughput is comparable to native (92%-95%)

Test performed on an EMC CX700 SAN (14-disk RAID0 LUN)
Windows 2003 Enterprise Edition Guest; same host OS
Iometer 64K Sequential Writes
Workstation 5 Pre-allocated virtual disk
Virtual Disks Concepts

- Virtual Machines do their IO to a virtual disk
  - IO is not directly to the real disk
  - Virtual disk is backed by a File on the Host computer
- Different Virtual Disk types based on file backing them
  - Pre-Allocated Virtual Disk
    - All space in the host file allocated up front
  - Sparse Virtual Disk
    - Space allocated on demand
    - Host file is Empty when first created

What are the performance differences ....
Virtual Disks Types

- In Workstation 5, Sparse Virtual disks are created by default
- User has option to create Pre-allocated disks
- Use Pre-allocated Virtual Disks for disk performance sensitive workloads
Sparse vs. Pre-Allocated

- In an empty sparse disk, every write to the virtual disk will grow the corresponding host file.
- Let’s look at 64K sequential writes in an empty versus pre-allocated virtual disk:
  - more like a worst case
- File growth should be a fast operation but NTFS has a known performance problem in this area!
- Actual performance can vary significantly depending upon workload.

Test performed on an EMC CX700 SAN (14-disk RAID0 LUN)
Windows 2003 Enterprise Edition Guest; same host OS
Iometer 64K Sequential Writes
Snapshots and Sparse disks

- Whenever you take a snapshot, an empty sparse disk is created
  - Delta Disks are sparse
- Score hardly changes for Business Winstone 2002 in a newly snapshotted virtual machine.
- Performance drop with multiple snapshots is only with the first snapshot
Dealing with Sparse Disks

- Empty Sparse disks do eventually grow 😊
- Workloads tend to speed up once the virtual disk is grown

Test performed on an EMC CX700 SAN (14-disk RAID0 LUN)
Windows 2003 Enterprise Edition Guest; same host OS

64K sequential writes in an empty sparse virtual disk
Defragging

- Defrag in an inside-out fashion
  1. Defragment the virtual disk in the booted up Virtual Machine (e.g., using Windows XP defragment tool)
  2. Then use the VMware Defrag option (or use the cmd line utility)
  3. Defragment the host disk (e.g., using Windows 2003 Enterprise defragment tool)
New: Full and Linked Clones

- Full clone: An independent copy of existing virtual machine
  - No need to access parent virtual machine
  - Slower to create than linked clone (may be minutes)
  - Outperforms linked clone for disk workloads
- Linked clone: A dependent snapshot of existing virtual machine
  - Needs access to parent virtual machine version from which linked clone created
  - Could be slower for disk workloads due to sparse disks
Local vs. Network Disks

- Putting a virtual disk file on a network share has a performance cost.
- Here we ran a realistic workload on a network share across a 100Mbit network.

Business Winstone 2002 Running Time(s) (lower is better)

<table>
<thead>
<tr>
<th># Snapshots</th>
<th>Local Virtual Disk</th>
<th>Virtual Disk on Network Share (100Mbit)</th>
<th>Lightweight Clone - Base on Network Share (100Mbit)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1200</td>
<td>1400</td>
<td>1600</td>
</tr>
</tbody>
</table>

Business Winstone 2002 Score (higher is better)

<table>
<thead>
<tr>
<th># Snapshots</th>
<th>Local Virtual Disk</th>
<th>Virtual Disk on Network Share (100Mbit)</th>
<th>Lightweight Clone - Base on Network Share (100Mbit)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>16</td>
<td>14</td>
<td>12</td>
</tr>
</tbody>
</table>
Disk Performance Tips

- Pre-allocated disks are faster than sparse disks
- Access to local disks is faster than access to disks over a network
- Defrag your disks: inside-out order
- SCSI guest disks perform better than IDE guest disks
  - Also use SCSI or SATA disks on the host if disk performance is critical
Virtual Machine Perf. Measurement Tips

- Performance tools and timings run in the guest may yield unexpected results
  - CPU time in guest may be inaccurate
  - Processor Performance counters not virtualized
- Host performance tools may not reflect VMware or guest activities; e.g.:
  - Host may not show VMware memory usage
  - Virtual machine to virtual machine networking may not involve host
- On Windows host, VMware exports perfmon objects
Summary

- VMware Workstation 5 delivers great new features and performance
- We want to hear your feedback on performance
  - post to forums!
This presentation covers the current versions of our products. Details about future releases of our products are available in select sessions at VMworld, including:

**PAC879:** The Next Phase of Virtual Infrastructure: Introducing ESX Server 3.0 and VirtualCenter 2.0

**PAC177:** Distributed Availability Services Architecture

**PAC484:** Consolidated Backup with ESX Server: In-Depth Review

**PAC485:** Managing Data Center Resources Using the VirtualCenter Distributed Resource Scheduler

**PAC532:** iSCSI and NAS in ESX Server 3