# Agenda

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Conquering Performance</td>
</tr>
<tr>
<td>2</td>
<td>New &amp; Important Performance Features</td>
</tr>
<tr>
<td>3</td>
<td>Recommended Practices</td>
</tr>
<tr>
<td>4</td>
<td>Performance Troubleshooting</td>
</tr>
<tr>
<td>5</td>
<td>Resources</td>
</tr>
</tbody>
</table>
**What Do You Want to Run on vSphere 6.0? ➩EVERTHING!**

### Overview
- vSphere is the platform for existing and new applications
- Scalability enhancements (VMs and Clusters) for all application workloads*
- Big Data Extensions and Pivotal CF (PaaS) Support
- Desktop Virtualization – 2D/3D Graphics, Instant Clone*
- OpenStack on vSphere = Success*
- Linux Container Support

### Benefits / Use Cases
- Increased scalability and performance
  - SAP Hana – 400% performance gains over RDBMS and 9x gains in planning load times
  - Rapid deployment of desktop virtual machines in seconds
    - 10x faster than in previous releases
  - Productivity and portability for application developers
  - Deliver Choice of Architecture

*New with vSphere 6.0
Conquering Performance
Virtual Machine Scalability

- Virtualize 99.99% of workloads today
## Platform and Host Scalability

<table>
<thead>
<tr>
<th>vSphere 5.5</th>
<th>vSphere 6.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>32 Hosts per Cluster</td>
<td>64 Hosts per Cluster</td>
</tr>
<tr>
<td>4000 Virtual Machines per Cluster</td>
<td>8000 Virtual Machines per Cluster</td>
</tr>
<tr>
<td>320 CPUs</td>
<td>480 CPUs</td>
</tr>
<tr>
<td>4 TB RAM</td>
<td>12 TB RAM</td>
</tr>
<tr>
<td>512 Virtual Machines per Host</td>
<td>1000 Virtual Machines Per Host</td>
</tr>
</tbody>
</table>
Virtual Machine Virtual Hardware 11

ESXi 6 Supports:

• 128 vCPUs
• 4 TB RAM
• Hot-add RAM now vNUMA aware
• WDDM 1.1 GDI acceleration features
• xHCI 1.0 controller compatible with OS X 10.8+ xHCI driver
• Serial and parallel port enhancements
  – A virtual machine can now have a maximum of 32 serial ports
  – Serial and parallel ports can now be removed

Low Latency Storage IO

- 1mm IOPS, >2ms latency, 8kb block, 32 OIO's

Low Latency Network IO

- Latency features reduce overhead to near native

The Worlds First TPC-VMS Benchmark Result

• Compliant and audited by a 3rd party.
• While not a direct comparison, you can see how database consolidation scenarios could achieve near native capabilities on the same hardware (>99%).

Virtualized Hadoop

• 12% better performance than native for TeraSort

Super Monster VM Performance

- Linear Performance gains as VM is scaled up to Monster Size

Scale Up Oracle 12c Database Performance on vSphere 6.0 using DVDStore 2.1

Reference: https://www.vmware.com/resources/techresources/10455
Super Monster VM Performance

- Linear Performance gains as VM is scaled up to Monster Size


Reference: https://www.vmware.com/resources/techresources/
New & Important Features
vCenter Server Features - Enhanced Capabilities – Parity

**Overview**

- Scalability supported by both Windows Install and vCenter Server appliance.
- Windows install supports Postgres and External SQL and Oracle DBs.
- vCSA supports embedded Postgres and external Oracle DBs.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Windows</th>
<th>Appliance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hosts per VC</td>
<td>1,000</td>
<td>1,000</td>
</tr>
<tr>
<td>Powered-On VMs per VC</td>
<td>10,000</td>
<td>10,000</td>
</tr>
<tr>
<td>Hosts per Cluster</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>VMs per Cluster</td>
<td>6,000</td>
<td>6,000</td>
</tr>
<tr>
<td>Linked Mode</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>
vCenter Server 6.0 - vSphere Web Client Features

Major Performance Improvements:

- UI
  - Screen by screen code optimization
  - Login now 13x faster
  - Right click menu now 4x faster
  - Most tasks end to end are 50+% faster

- Performance charts
  - Charts are available and usable in less then half the time

- VMRC integration
  - Advanced virtual machine operations
vCenter Throughput

Reference: What's New in VMware vSphere 6 – Performance
vCenter Latency

Reference: What's New in VMware vSphere 6 – Performance
vCenter Latency

Reference: What's New in VMware vSphere 6 – Performance
Performance: Caution Using FireFox

<table>
<thead>
<tr>
<th>Browser</th>
<th>e2eTime (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FireFox</td>
<td>36.3</td>
</tr>
<tr>
<td>IE9</td>
<td>21.37</td>
</tr>
<tr>
<td>IE11</td>
<td>20.06</td>
</tr>
<tr>
<td>Chrome</td>
<td>19.29</td>
</tr>
</tbody>
</table>
NVIDIA GRID vGPU
Superior 2D and 3D Graphics for Virtual Environments

**Overview**
- Superior 2D and 3D graphics on hosts with NVIDIA GRID GPU
- Enables shared access to physical GPUs
- Uses native NVIDIA drivers
- Best for high performance needs like desktop virtualization with graphic intensive applications in the fields of healthcare, manufacturing, higher education, and oil & gas

**Benefits**
- Superior 2D and 3D graphics
- Full capabilities of physical NVIDIA GPUs
- High Density, Low Cost Graphics workstation
Instant Clone
Rapidly Clone and Provision Virtual Machines

Overview

- Technology that enables the ability to rapidly clone and provision thousands of VMs in minutes.
- Part of vSphere 6.0 but will be enabled by other applications such as BDE and Horizon View in later releases.

Benefits

- Clone VMs 10X faster than what is currently possible today
- Deploy thousands of virtual desktops in a matter of minutes versus what would normally take hours
Network – New Features

• NetlOC v3
  Reserve bandwidth to guarantee service levels at vnic and portgroup

• Host-Wide Performance Tuning Engine
  10% higher consolidation ratios with web farm use case

• vmxnet3 LRO Support
  15-20% improvement in receive throughput and efficiency for Windows

• 40GbE Improvements
  35Gbps and 50% reduced compute cost
Storage – New Features

- **Storage Stack Optimizations**
  Effort spent reducing overhead and increasing capabilities to best leverage flash storage
  Examples: Samsung NVMe 240k -> 710k IOPS, EMC XtremSF 200k -> 670k IOPS

- **VSAN 6.0**
  7 Million IOPS, <2 ms Latency

- **VVOLs**
  Performance the same or better as previous forms of storage integration

Recommended Practices
vCenter – Recommendations

• Web Browser Selection Important
  Firefox slowest, Chrome fastest, IE11 very close 2nd

• Database Performance Critical
  vCenter experience most impacted by database performance, ensure proximity and speed

• Place vCenter on Tier 1 Storage
  Placing the vCenter virtual machine on low latency storage will improve performance and experience

• Don’t Change Statistics Levels
  Change only as necessary, short intervals, as it places a large demand on vCenter and the DB

• Check JVM Sizing
  Adding RAM to VM doesn’t automatically adjust JVM service sizing, see KB2021302
Compute & Memory – Recommendations (1/2)

• **Rightsize, Rightsize, Rightsize**  
  Spend effort on rightsizing workloads for vCPU count and assigned memory

• **Size VM into pNUMA Node if Possible**  
  Doing this will reduce the potential for remote memory access and/or thread migration

• **Don’t use vCPU Hot-Add**  
  As it disables vNUMA and presents the virtual machine with UMA topology

• **Select High Performance in BIOS or vSphere**  
  Selecting anything else will save power but does potentially induce compute latency

• **Enable Hyper-Threading**  
  vSphere understands and uses Hyper-Threading to its advantage
• Watch Memory Overcommit
  Overcommit provides consolidation value at risk of performance during shortages

• Do NOT Use ‘Active Memory’ in a Vacuum
  Active Memory is more a ‘rate’ counter than a ‘capacity’ counter, temper it with other counters like ‘Consumed’ or use vROPs
Network – Recommendations

• Use vmxnet3 Guest Network Driver
  Very efficient and required for maximum performance

• Evaluate Disabling Interrupt Coalescing
  Default mechanism may induce small amounts of latency in favor of throughout, evaluate disabling it as cost today is negligible

• Jumbo Frames Provide Value
  While challenging to enable end-to-end sometimes they provide value to high throughput functions like VSAN, vMotion and NAS

• It’s a 10Gb World
  1Gb saturation is real, more bandwidth required today, especially in light of VSAN, MonsterVM vMotion

• Use Latency Sensitivity ‘Cautiously’
  While it can reduce latency and jitter in the 10us use case, it comes at a cost with core reservations, etc
Storage – Recommendations

- Use Multiple vSCSI Adapters
  Allows for more queues and I/O’s in flight

- Use pvscsi vSCSI Adapter
  More efficient I/O’s per cycle

- Don’t Use RDM’s
  Unless needed for shared disk clustering, no longer a performance advantage

- Leverage Your Storage OEM’s Integration Guide
  They provide necessary guidance around items like multi-pathing
Performance Troubleshooting
Define the Performance Issue

• Understand Application Function & Architecture
  At a minimum know what your application does and what it's dependent on.

• Select Application KPIs
  Application performance must be measured using an application counters (tps, response time, etc) and not virtual resource consumption.

• Define Success Criteria
  With your app owner, define at what level the application KPI's must be to consider it performant.

• Comparisons must be Apples-to-Apples
  Any changes to infrastructure (physical or virtual) create comparison challenges.

• Now the Gap is Identified, Begin Troubleshooting
  With an understanding of the requirements and current deficiency, you can now begin to investigate and/or tune.
Use the Right Tool

• esxtop
  2 sec data points, VERY granular, not scalable across hosts

• vRealize Operations
  5 min data points, very scalable, best starting view

• vCenter Performance Charts
  20 sec data points, okay real-time data, poor history, recommend vROPs

• VSAN Observer
  Most detailed tool to troubleshoot VSAN related performance

• 3rd Party
  Ensure you know what the counters mean and their sample rate
Suggested Methodology

1. Storage
   - What are my config/tuning options?
   - Change & re-test
   - Acceptable?
     - yes
     - no

2. Network
3. Guest
4. Memory
5. CPU
Storage: What’s important in the stack

VMkernel
- Virtual SCSI
- File System

Guest
- File System
- I/O Drivers
- Application

Windows Device Queue

K = ESX Kernel
G = Guest Latency
S = Windows Physical Disk Service Time
R = Perfmon Physical Disk “Disk Secs/transfer”
A = Application Latency
D = Device Latency

K
G
S
R
A
Storage: Key Indicators

• Kernel Latency Average (KAVG)
  This counter tracks the latencies of IO passing thru the Kernel
  Investigation Threshold: 1ms

• Device Latency Average (DAVG)
  This is the latency seen at the device driver level. It includes the round-trip time between the HBA and the storage.
  Investigation Threshold: 10-15ms, lower is better, some spikes okay

• Device Latency Average (GAVG)
  This is the latency seen at the guest level. It is effectively DAVG + KAVG. Needed for network attached storage.
  Investigation Threshold: 10-15ms, lower is better, some spikes okay
CPU: Key Indicators

• Ready (%RDY)
  % time a vCPU was ready to be scheduled on a physical processor but couldn’t due to processor contention
  Investigation Threshold: 10% per vCPU

• Co-Stop (%CSTP)
  % time a vCPU in an SMP virtual machine is “stopped” from executing, so that another vCPU in the same virtual machine could be run to “catch-up” and make sure the skew between the two virtual processors doesn’t grow too large
  Investigation Threshold: 3%

• Used (%USED)
  Make sure the VM is not oversized.
Memory: Key Indicators

- **Balloon driver size (MCTLSZ)**
  - the total amount of guest physical memory reclaimed by the balloon driver
  - Investigation Threshold: 1

- **Swapping (SWCUR)**
  - the current amount of guest physical memory that is swapped out to the ESX kernel VM swap file.
  - Investigation Threshold: 1

- **Swap Reads/sec (SWR/s)**
  - the rate at which machine memory is swapped in from disk.
  - Investigation Threshold: 1
Network: Key Indicators

- Transmit Dropped Packets (%DRPTX)
  The percentage of transmit packets dropped.
  Investigation Threshold: 1

- Receive Dropped Packets (%DRPRX)
  The percentage of receive packets dropped.
  Investigation Threshold: 1
Resources
Resources

VMware’s Performance – Technical Whitepapers
http://www.vmware.com/resources/techresources/cat/91,96
VMware’s Tech-Marketing Performance Blog
VMware’s Perf-Eng Blog (VROOM!)
http://blogs.vmware.com/performance
Performance Community Forum
http://communities.vmware.com/community/vmtn/general/performance
VMware Performance Links – Master List
https://communities.vmware.com/docs/DOC-25253
Virtualizing Business Critical Applications
Resources

Performance Best Practices
http://www.vmware.com/pdf/Perf_Best_Practices_vSphere5.5.pdf
https://www.vmware.com/resources/techresources/10480 <-vSphere 6

Troubleshooting Performance Related Problems in vSphere Environments
http://communities.vmware.com/docs/DOC-14905 (vSphere 4.1)
http://communities.vmware.com/docs/DOC-19166 (vSphere 5)
http://communities.vmware.com/docs/DOC-23094 (vSphere 5.x with vCOps)
Resources

Virtualizing Microsoft Business Critical Applications on VMware vSphere
   by: Matt Liebowitz, Alexander Fontana

vSphere High Performance Cookbook
   by: Prasenjit Sarkar

Troubleshooting Storage Performance
   By: Mike Preston

VMware vSphere Performance: Designing CPU, Memory, Storage, and Networking for Performance-Intensive Workloads
   By: Matt Liebowitz, Christopher Kusek, Rynardt Spies

Virtualizing SQL Server with VMware: Doing IT Right
   By: Jeff Szastak, Michael Corey, Michael Webster

Virtualizing Oracle Databases on vSphere
   By: Don Sullivan, Kannan Mani

VMware vRealize Operations Performance and Capacity Management
   By: Ewan ‘e1’ Rahabok
Resources

VMware Hands-On-Labs
http://labs.hol.vmware.com/

HOL-SDC-1404:
vSphere Performance Optimization – This has always been one of the most popular labs and has content for both the beginner and the advanced vSphere Administrator. You can learn more about the basics of vSphere Performance or delve into esxtop, or vNUMA.
http://labs.hol.vmware.com/HOL/#lab/1474
Thank You