Database Virtualization With vSphere 6.7 Doing IT Right

Dean Bolton

Dean Bolton launched his IT career in 2000 after completing his bachelor's degree in computer science and engineering at MIT. He started working with Oracle Databases from the beginning at an internet systems and applications design firm. Since then, he has continued working as an Oracle developer, database administrator, architect, consultant, and evangelist. Dean is the managing partner of LicenseFortress, the first and only Oracle software license management service with a guarantee, and VLSS, a premier Oracle and VMware consulting firm. Dean is recognized as one of the top three experts on licensing Oracle on VMware.

Michael Corey

Cloud: #42 – Top 100 Cloud Influencers and Brands 2017 & 2015

LICENSEFORTRESS®
ORACLE COMPLIANCE MANAGEMENT + AUDIT PROTECTION WITH A GUARANTEE

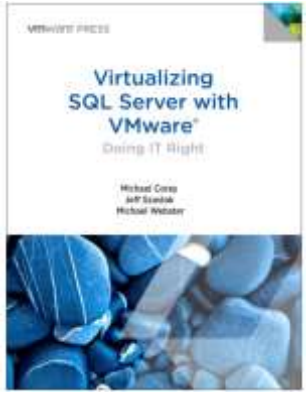
Started Working with **Oracle Version 3.0** Beta Tested Oracle 5,6,**6.2**,7,8.X,9.X….. Presented on Technology & Business Topics from Brazil to Australia Worked with Oracle on UNIX, Linux, Windows, MVS,VM, VMS,..

ORACLE ACE | vmware vEXPERT | Microsoft MVP Most Valuable Professional | BDQ BIG DATA QUARTERLY

@Michael_Corey

Books Include:

Virtualizing SQL Server with VMware Doing IT Right
Oracle Database 12c: Install, Configure & Maintain like a Professional
Oracle 11g A Beginner's Guide
Oracle 10g A Beginner's Guide
Oracle 9i - A Beginner's Guide
SQL Server 7 Data Warehousing
Oracle8i - Data Warehousing
Oracle8i - A Beginner's Guide
Oracle8 - Data Warehousing
Oracle8 – Tuning
Oracle8 - A Beginner's Guide
Oracle - Data Warehousing
Oracle - A Beginner's Guide
Tuning Oracle



Community Activities…
Past President Independent Oracle Users Group (IOUG)
President IOUG VMware SIG www.vmsig.org
Founding Board IOUG Virtualization SIG & VMware SIG
Founding Board Professional Association of SQL Server
Past Member IOUG Board of Directors
Talkin'Cloud Top 200 Channel Partner Experts Cloud
Past Member Microsoft Data Warehouse Council
Past Member Oracle Educational Advisory Council
Past Director of Conferences IOUG Alive
Executive Board Massachusetts Robert H. Goddard Council on Science, Technology, Engineering & Mathematics

BLOG: http://michaelcorey.com/
Regular Columnist Big Data Quarterly

Legal Disclaimer

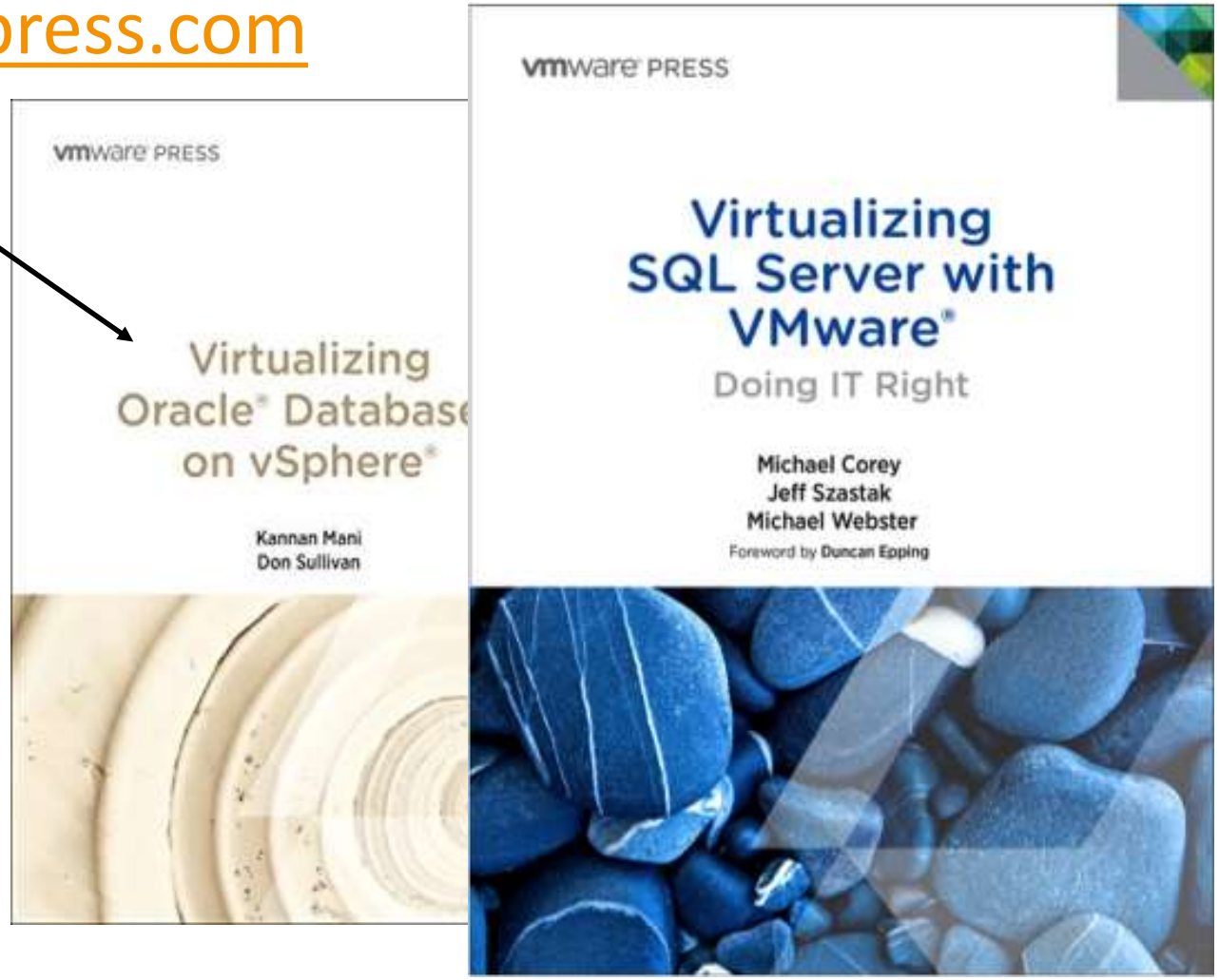
*The **views expressed anywhere on this presentation are strictly ours** and **not the opinions of VLSS Technologies, LicenseFortress** or anyone else. All content is provided without any form or warranty explicit or implied, for informational purposes and for use at your own risk. You should not rely on anything on this site as legal, financial, accounting, investment, tax, or any other kind of regulated advice*

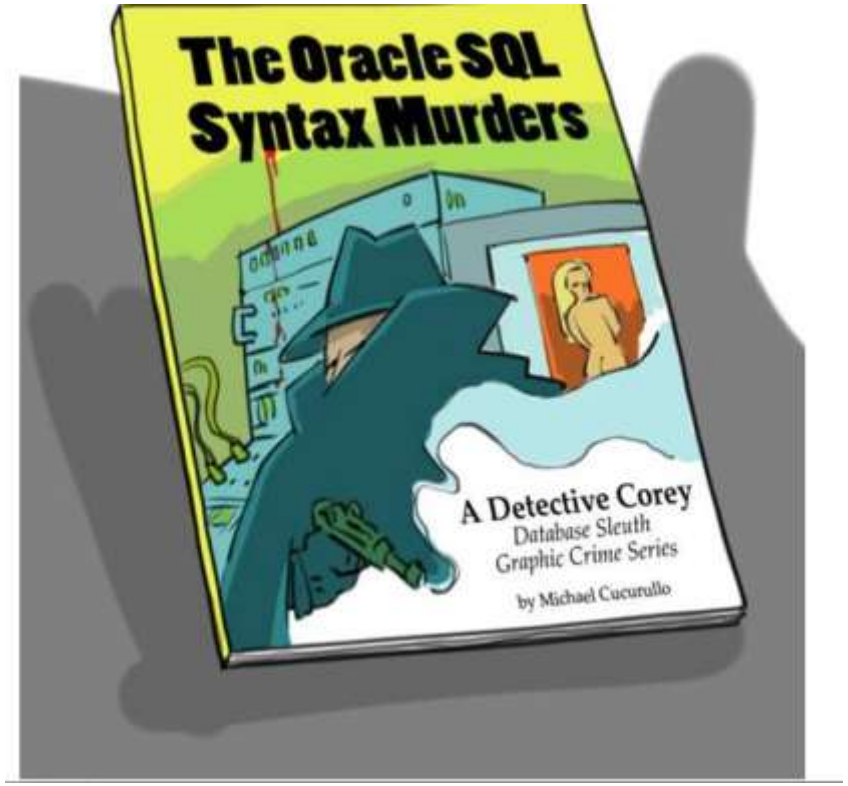
All product and company names are trademarks™ or registered® trademarks of their respective holders. Use of them does not imply any affiliation with or endorsement by them.

Authors Pride/And or Plug

vmwarepress.com

"Great Paper Weight





Source: http://www.cucurullo.com/

http://www.pearsonitcertification.com/store/virtualizing-oracle-databases-on-vsphere-9780133570182

http://www.pearsonitcertification.com/store/virtualizing-sql-server-with-vmware-doing-it-right-9780321927750

VMware Experts Program SQL Server Edition

VMware Experts Program Oracle Edition

Monster VM's (Database Virtualization) with vSphere 6.7: Doing IT Right

What is a Monster VM

A **monster virtual machine** is a **virtual machine** (**VM**) that typically has **more than eight** virtual CPUs (vCPUs) and more than **255 GB** of virtual RAM. **Monster VMs** are used to **virtualize applications with large resource needs**, such as **Microsoft Exchange**, **Microsoft SQL Server** or an **Oracle** database.

Term originated in 2011 when VMware increased the virtual hardware limits on virtual machines from vSphere 4 to vSphere 5.

Doing Something a Little Different

Oracle, Microsoft SQL Server & Monster VM's

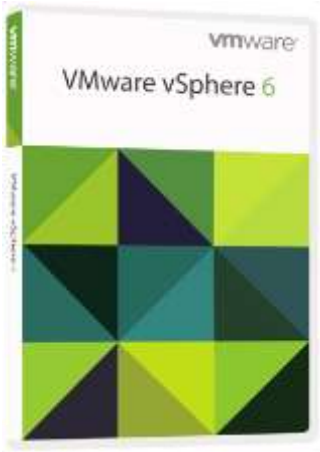
Principals Apply All Databases & Monster VM's

Don't Forget About Me

"This is a Database/Monster VM on Virtualized Infrastructure Session"

Project Capstone – VMworld 2015 (Story is only Better today)

A Collaboration of VMware, HP and IBM



VMware vSphere 6

+

HP Superdome X
(15 X 16 Cores, 480 Threads, 12TB)

+

IBM FlashSystem

Various Monster VMs running Oracle Databases
(120 vCPU, 60 vCPU and others)

Workload Used for testing: DVD Store Version 3
https://github.com/dvdstore/ds3

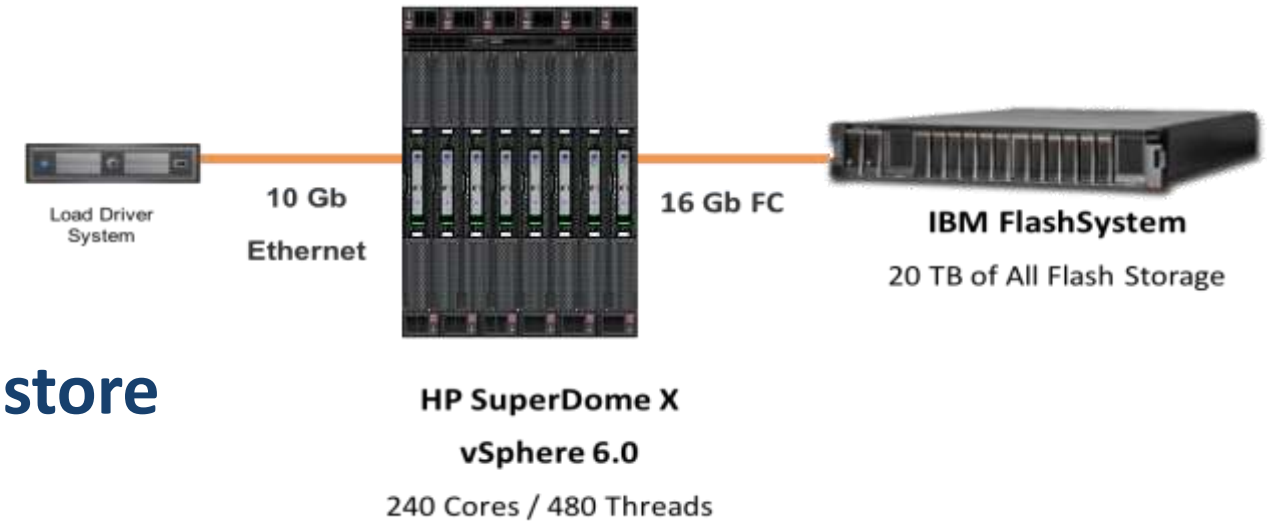
Store simulates a real online store with customers logging onto the site, browsing products and product reviews, rating products, and ultimately purchasing those products

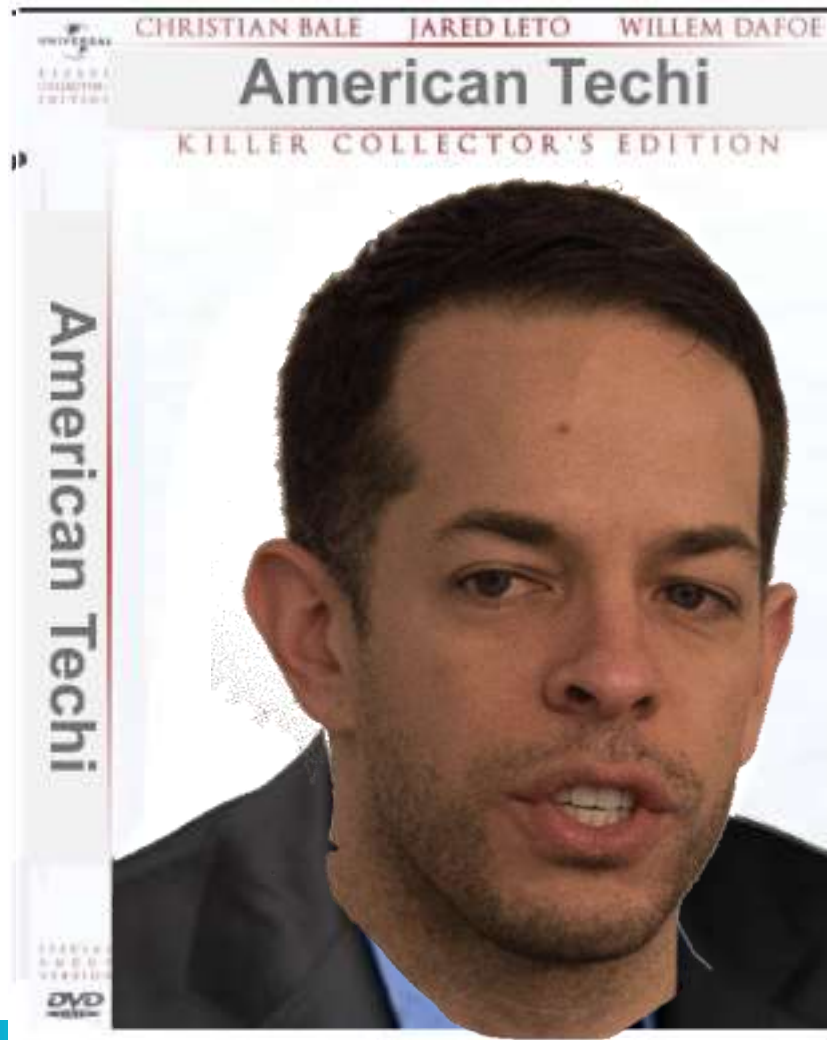
Benchmark: **Orders Per Minute**
Each order representing a complete login, browsing & purchasing process that includes many individual SQL operations against the database.

- Workload was run at increasing levels of load to find the highest performing test configuration
- **All 480 threads on server were near saturation during each max config test**

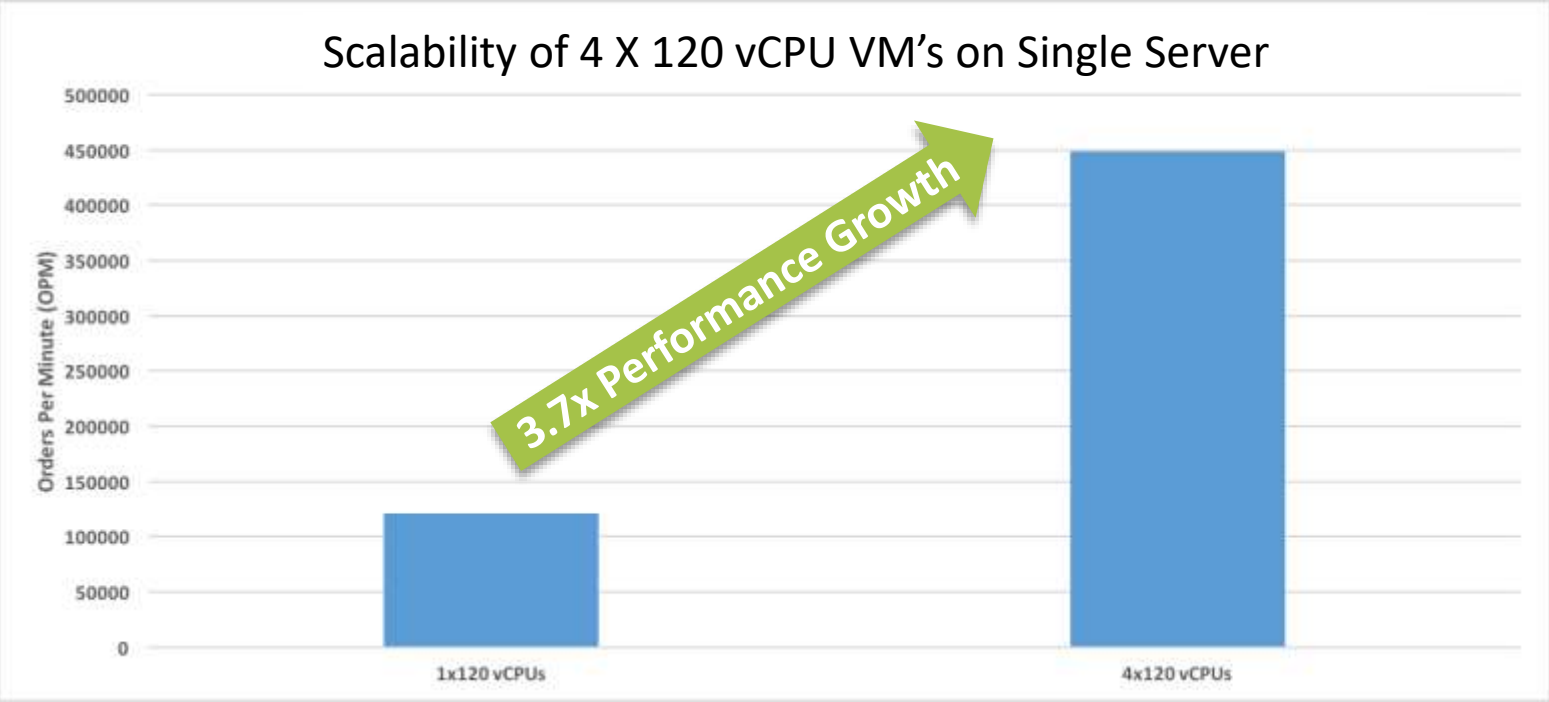
HOW TO GUIDE

Hfxte.ch/dvdstore





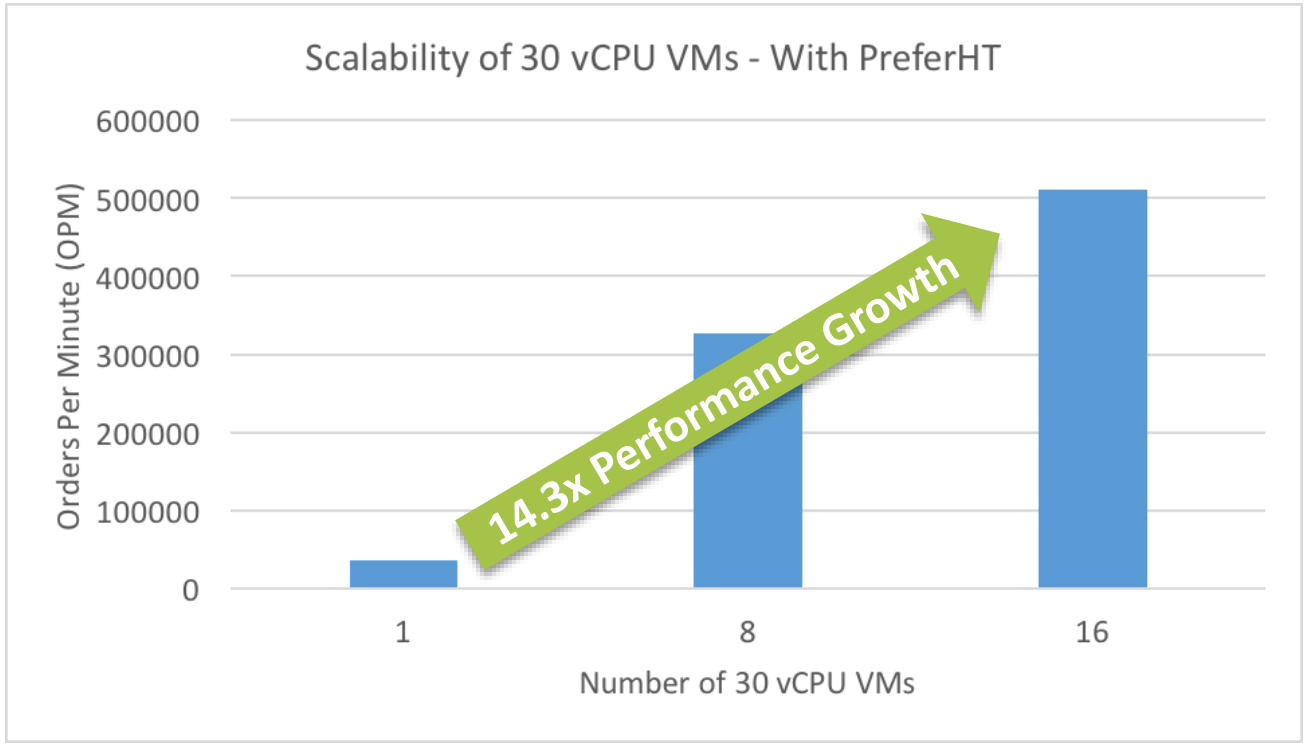
Project Capstone Test Results of 120 vCPU VMs



Reference

- 4 VMs achieve 3.7x performance of a single VM (92% of linear)
- Each VM uses 4 sockets / 60 cores / 120 Threads
- Average 20k IOPS at .3ms response time / Peak of ~50K IOPS

Project Capstone Test Results with 30 vCPU VMs



Reference

- Scalability from 1 VM to 16 VMs is 14.3x (89% of linear)
- Each VM uses 1 socket / 15 cores / 30 Threads
- Average 13K IOPS at .3 ms response time
- **To drive CPU usage so high all disk IO must be very fast. System is not waiting for a response**

Project Capstone – VM Configuration

Oracle 12c Database

Red Hat Enterprise Linux 6.5

120 vCPUs OR 60 vCPUs OR 30 vCPUs

256 GB of RAM

2 VMXNET3 NICs

pvscsi disk adapters

No pinning (but using **PreferHT)**

- ✓ **pvscsi virtual Disk Adapters (Talk More Later)**
- ✓ **VMXNET3 virtual NICs (Talk More Later)**

PreferHT:

Informs vSphere you'd **rather have access to processor cache and NUMA memory locality as priority, over** the **additional compute cycles.**

DBEXT01 - Edit Settings

Virtual Hardware | VM Options | SDRS Rules | vApp Options

- Power management — *Expand for power management settings*
- Boot Options — *Expand for boot options*
- Encryption — *Expand for encryption settings*
- Advanced
  - Settings: ☐ Disable acceleration; ☐ Enable logging
  - Debugging and statistics: Run normally
  - Swap file location: ◉ Default — Use the settings of the cluster or host containing the virtual machine. ○ Virtual machine directory — Store the swap files in the same directory as the virtual machine. ○ Datastore specified by host — Store the swap files in the datastore specified by the host to used for swap files. If not possible, store the swap files in the same directory as the virtual machine. Using a datastore that not visible to both hosts during vMotion might affect the vMot performance for the affected virtual machines.
  - Configuration Parameters: Edit Configuration...

Configuration Parameters

⚠ Modify or add configuration parameters as needed for experimental features or as instructed by technical support. Empty values will be removed (supported on ESXi 6.0 and later).

Filter

Name	Value
migrate.hostLogState	none
migrate.migrationId	0
monitor.phys_bits_used	42
numa.vcpu.min	8
nvram	DBEXT01.nvram
pciBridge0.pciSlotNumber	17
pciBridge0.present	true
pciBridge4.functions	8
pciBridge4.pciSlotNumber	21
pciBridge4.present	true
pciBridge4.virtualDev	pcieRootPo

Name: numa.PreferHT Value: 1 Add

OK Cancel

NUMA Is Really Important

Crossing NUMA Boundaries Results in NUMA hit.

More cores artificially inflated by hyper-threading not necessarily better

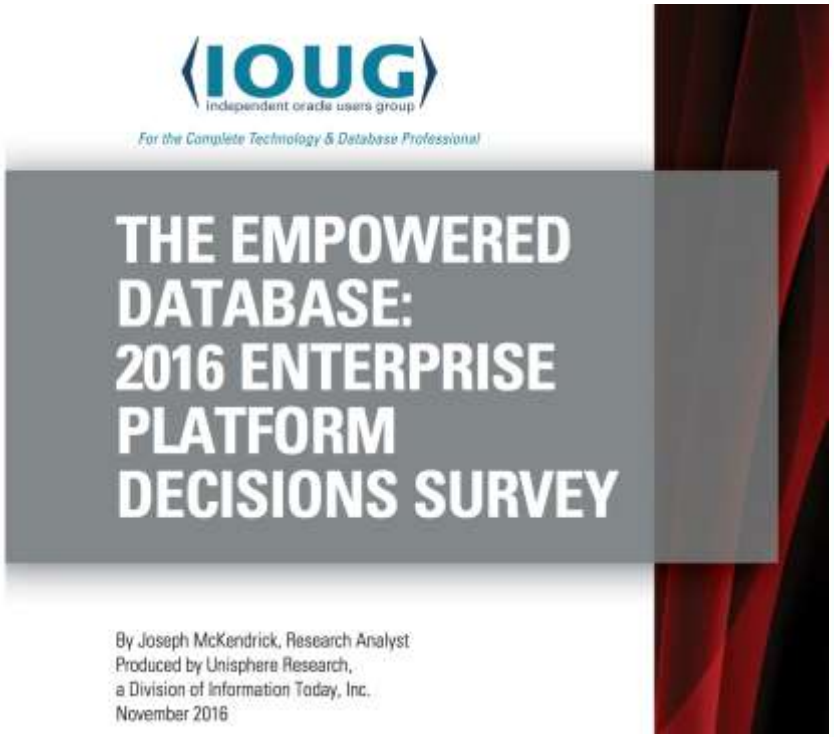
VMware Tools – Install It, Use It

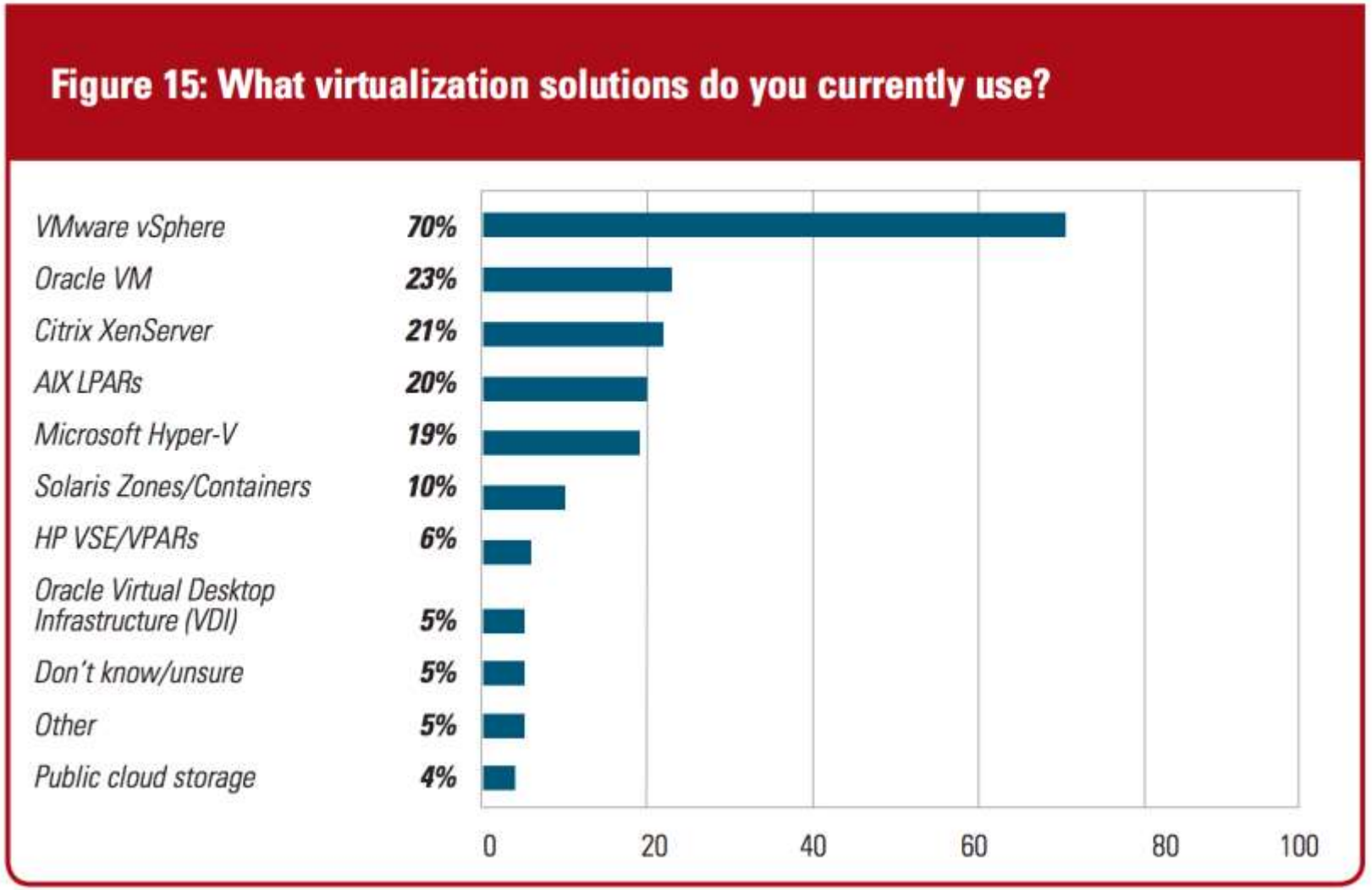
VMware Tools is a suite of **utilities that enhances the performance of the virtual machine's** guest operating system and improves management of the virtual machine. Although the guest operating system **can** run without **VMware Tools**, **you would** lose important functionality and convenience.

Includes VMXNET networking driver
Includes PVSCSI driver
Increased Disk Time Outs
Includes Balloon Driver
Ability to Issue In-Guest VSS
(ability to do crash consistent backups)

"Keep VMware Tools Up To Date"

November: THE EMPOWERED DATABASE: 2016 ENTERPRISE PLATFORM DECISIONS SURVEY



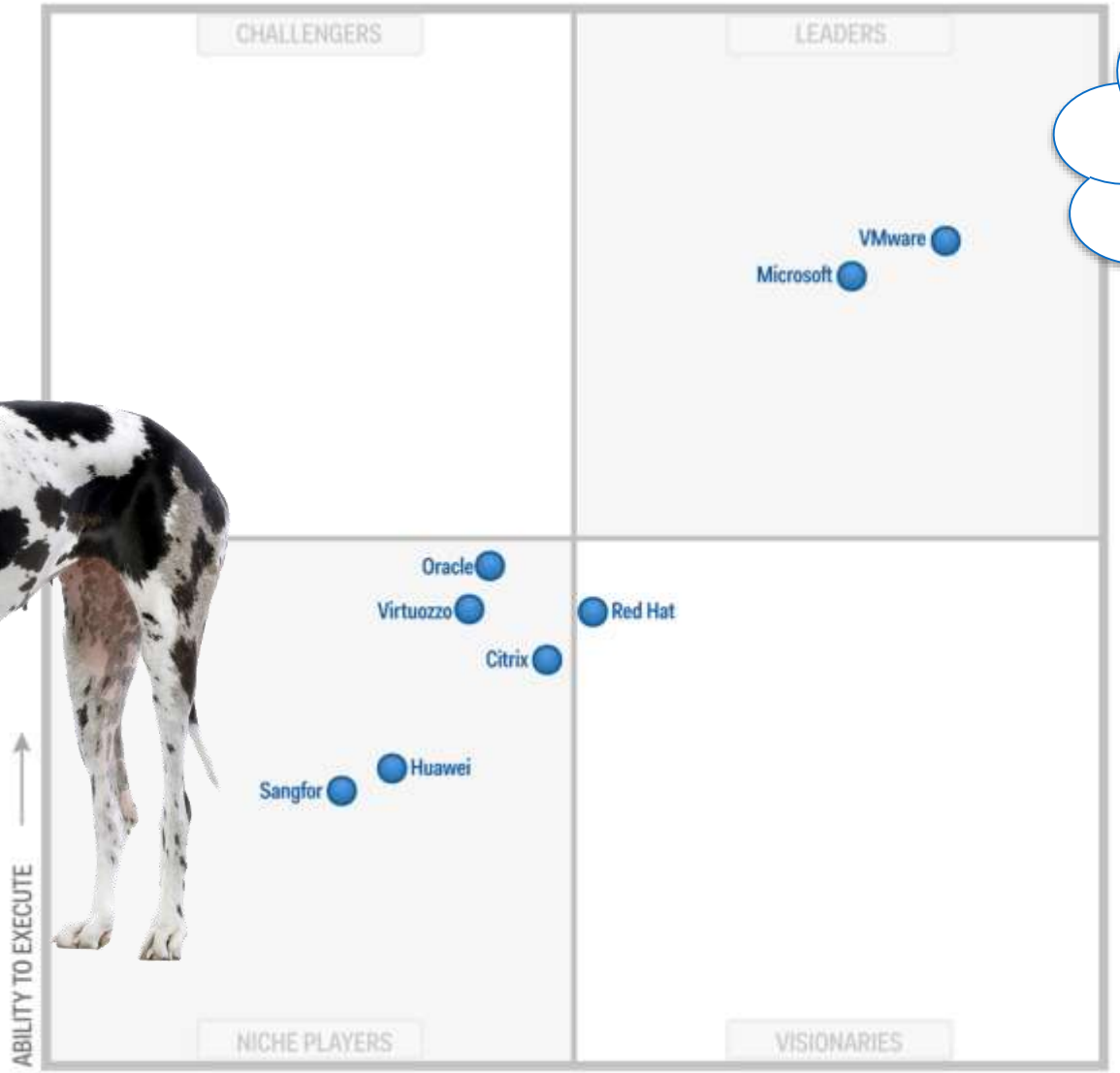


Source: http://www.dbta.com/DBTA-Downloads/ResearchReports/THE-EMPOWERED-DATABASE-2016-ENTERPRISE-PLATFORM-DECISIONS-SURVEY-6662.aspx

Gartner July 2018

Slow Down I want to Catch Up

Figure 1. Magic Quadrant for x86 Server Virtualization Infrastructure



Source: Gartner (August 2016)

I feel at home on vSphere

Source: https://www.redhat.com/de/blog/fundamental-shifts-virtualization-market#

Virtualization is the new Norm
The beer of Choice is VCDX Ale

2018 Gartner Report On Operational Database Management Systems

Why Customers Are Virtualizing Business Critical Applications

Your PC ran into a problem and needs to restart. We're just collecting some error info, and then we'll restart for you.

25% complete

For more information about this issue and possible fixes, visit
http://windows.com/stopcode

If you call a support person, give them this info:
Stop code: CRITICAL_PROCESS_DIED

The Number 1 Reason an O/S Crashes is Bad Drivers

Hardware Resource

vmware

Concise Set

Very Efficient Drivers

Focused Set Drivers Well Vetted

O/S Du Jour

Many Drivers

Many Versions

New Driver's Can Cause Issues

Picture Source: Harry Potter and The Chamber of Secrets 2002

Business Case: Why Virtualize Databases

Licensing	Reduce Licensing Cost
DB Consolidation	Reduce Hardware Cost
DB on Demand/DBaaS	Provision Databases On Demand
Quality of Service	Built in High Availability & Simple DR/Scale Dynamically
Security	Complete Isolation between systems on same Host

Winner Winner Chicken Dinner

"Database Consolidation > 50% Attainable"

"Do you even need Oracle enterprise edition when you are on Pure Storage?" Attendee VMware Experts Program Ireland 2017"

Reference

vmware

Resource Hot Plug – CPU Hot Plug Disables vNUMA

CPU Hot Plug
Dude Bad Idea

Edit Settings | DBEXT01

Virtual Hardware | VM Options

ADD NEW DEVICE

CPU *	2	
Cores per Socket	2	Sockets: 1
CPU Hot Plug	Enable CPU Hot Add	
Reservation	0	MHz
Limit	Unlimited	MHz
Shares	Normal	2000

David Klee Says,
Really Bad Idea

Oracle – Hot Add Memory

Oracle database memory parameters defined at instance startup. You will have to **restart** the database to take advantage of added memory.

Only Useful if you have set SGA_MAX_SIZE to Big (Bloating the SGA)

Recommend…

SGA_TARGET_SIZE <= SGA_MAX_SIZE

“AVOID Bloating the SGA”

“Virtualization is A Shared Resource Environment”

Architecting for Performance

The Right Hypervisor

vSphere 6.X Hypervisor Overhead – Very Low

"Dell recently published (two **TPCx-HS** (Transaction **Processing Performance Council (www.tpc.org)**) results that demonstrate that **Big Data technology on vSphere** is actually **'faster' than bare metal** March 2015"

Placing the big data application tier on vSphere 6, with everything else being equal, yielded an ***8% performance benefit over bare metal***.

The Configuration Details:

The test bed consisted of the following configuration:

Virtual Machine Workloads

- 128x Cloudera CDH 5.3.0 virtual machines
- 10 vCPU, 60GB RAM each
- SUSE SLES 11 SP3

Hosting Infrastructure

- Dell PowerEdge R720xd Servers
- Intel Xeon E5-2680v2 – 2.8Ghz, 256GB RAM
- Local DAS
- VMware vSphere 6

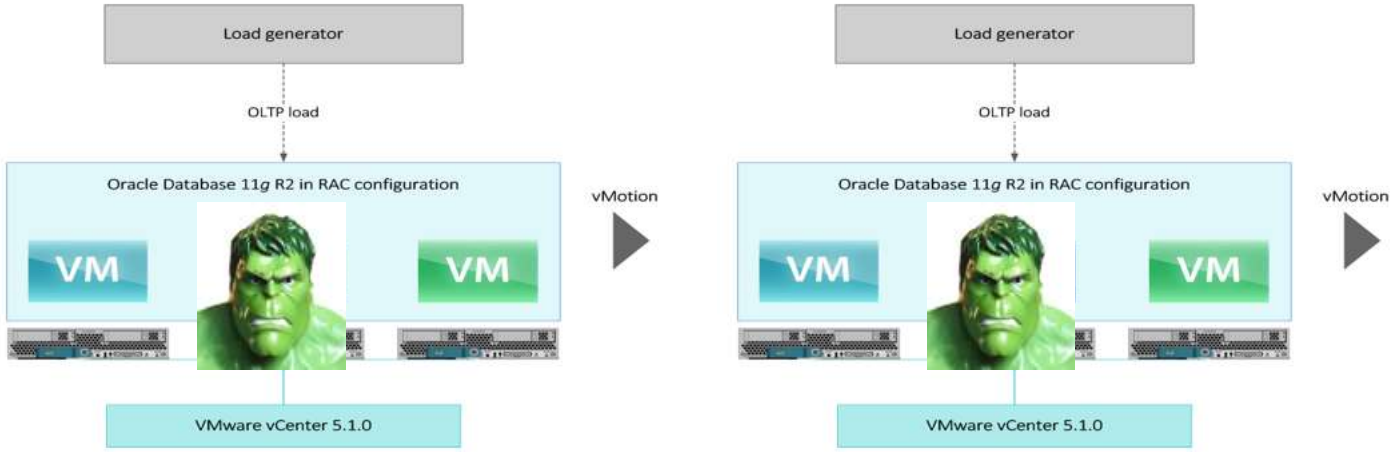
Source: https://blogs.vmware.com/vsphere/2015/03/virtualized-big-data-faster-bare-metal.html

Virtualized Hadoop Performance with VMware vSphere 6 on High-Performance Servers
Source: https://www.vmware.com/techpapers/2015/virtualized-hadoop-performance-with-vmware-vsphere-10452.html

vMotion Oracle RAC on vSphere 5.1 Functional Stress Test

VMware, EMC, Cisco Executed by "Principled Technologies" 2013

vSphere 6.5+ NOW Able to Encrypt vMotions



Move me on other Hypervisors at your own Risk

3 RAC Node, vMotion on all 3 Nodes Simultaneously – Without any network disruption

Proves 100% Abstraction between hardware and OS

vMotion Launched 2003

http://www.principledtechnologies.com/VMware/vMotion_Oracle_RAC_1013.pdf

vSphere 6.5+ Predictive DRS/Proactive HA

"Predictive DRS using a combination of DRS and vRealize Operations Manager to predict future demand and determine when and where hot spots will occur. When future hot spots are found, Predictive DRS moves the workloads long before any contention can occur".

Proactive HA – Works with OEM information

IMPORTANT Points

- **Make Sure Your Have Affinity Rules Set Up**
  - **For Performance Reasons**
  - **For License Reasons**
  - **For Availability (AG, Clustering, Chassis…)**

"With vROPS running, this feature Auto-Load Balance Ahead of Planned Consumption Spikes"

Microsoft SQL Server

ORACLE

Service Level Agreement & The DBA

Situation: **Monitors Critical Medical Equipment** within a Hospital. **SQL Server** only able to **to use 50% available CPU** - performance problems. This could have been…..

Solution: Take Server Down. Adjust BIOS Setting causing issue

Customer: **Can't take Server down for 5 minutes**

Stand Alone Instance – Had it been virtualized DBA would **have** had **options**

But Failure Was Not an Option

Situation points to a bigger issue….

“Management’s” expectations concerning the **availability** of the **database/Monster VM** and the **physical infrastructures ability** to **actually support** those Expectations

Source: https://www.youtube.com/watch?v=vOcsm5VnXLU

The Elephant is in the Room – Reset Expectations

Get the Resources You need to meet the expectations OR – Reset Expectations concerning Database/Monster VM Uptime

Have The Conversation, Set Proper Expectations

Avoid Good Intention Bios Settings

- Default lot of Servers is “Green” Friendly Setting
  - Many Times Does Not Ramp UP CPU Quickly/Some Cases Completely

✓**Set CPU to Hi-Performance/Enterprise in BIOS**

✓**Enable hyper-threading in BIOS**

✓**Enable hardware virtualization features (VT-x, AMD-V, EPT, RVI…)**

✓
- **Set CPU Snoop to Early Snoop (One NUMA node)**
- **Set CPU Snoop to Home Snoop (vNUMA spans physical NUMA)**

Power management also **reduces the voltage to your PCIe slots,** which can **affect things like PCIe flash storage cards, HBAs**, etc.

“Physical Resources of Host are hard Limits”

Is an Unplanned Outage of 3 Days In A Row Ok?

Availability Percentage	Downtime Year	Downtime Month*	Downtime Week
"Two Nines"-99%	3.65 Days	7.2 Hours	1.69 Hours
"Three Nines"-99.9%	8.76 Hours	43.2 Minutes	10.1 Minutes
"Four Nines" - 99.99%	52.56 Minutes	4.32 Minutes	1.01 Minutes
"Five Nines" - 99.999%	5.26 Minutes	25.9 Seconds	6.06 Seconds
		* Using 30 Day Month	

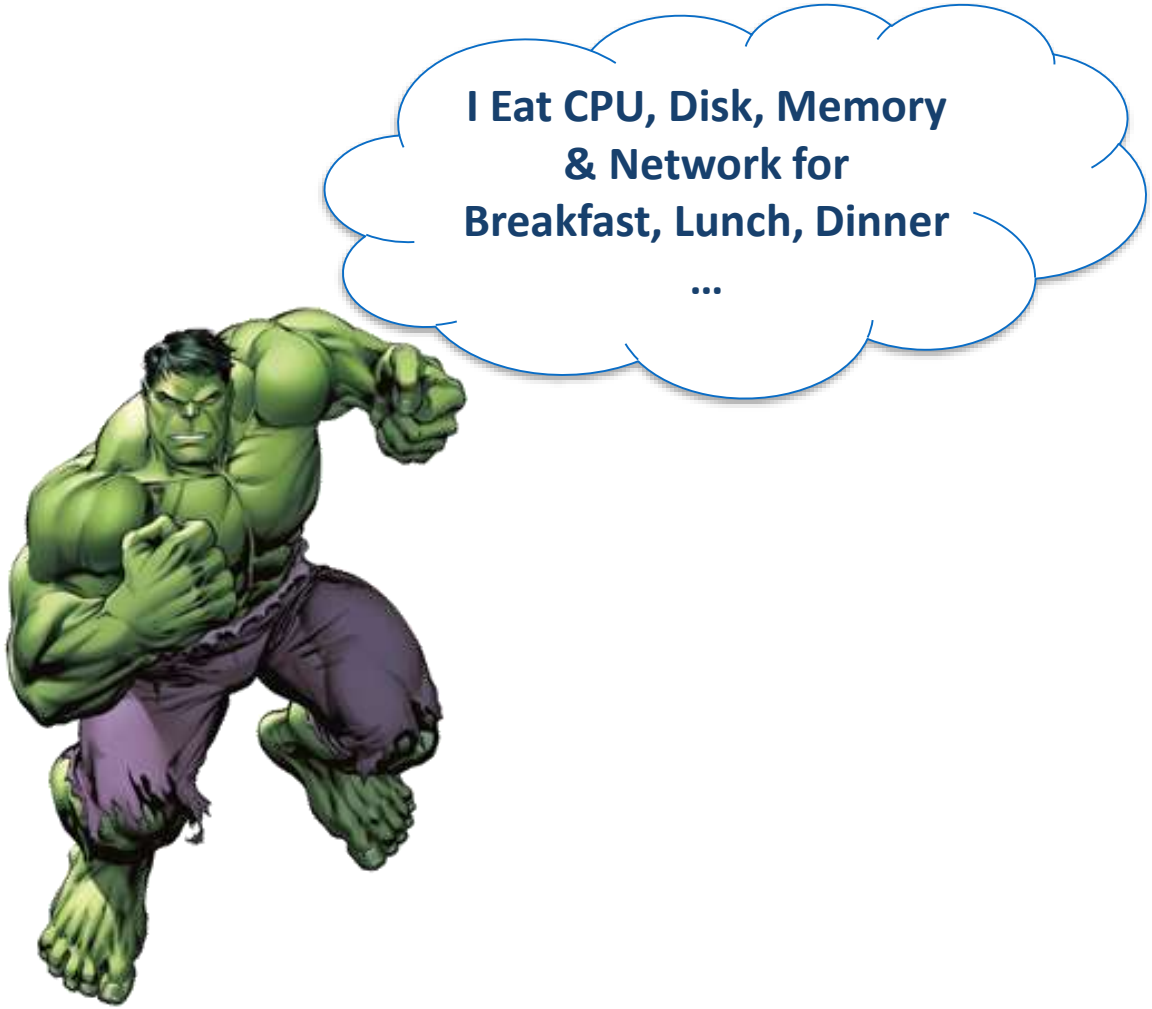
You Had 99% Availability !

Virtualizing Databases With vSphere 6.7

Doing IT Right

Lessons Learned in Non-Production & Third Party Applications

"What Works in Tier-2 (non-production), will not always work with Tier-1 (production)/Monster VMs"



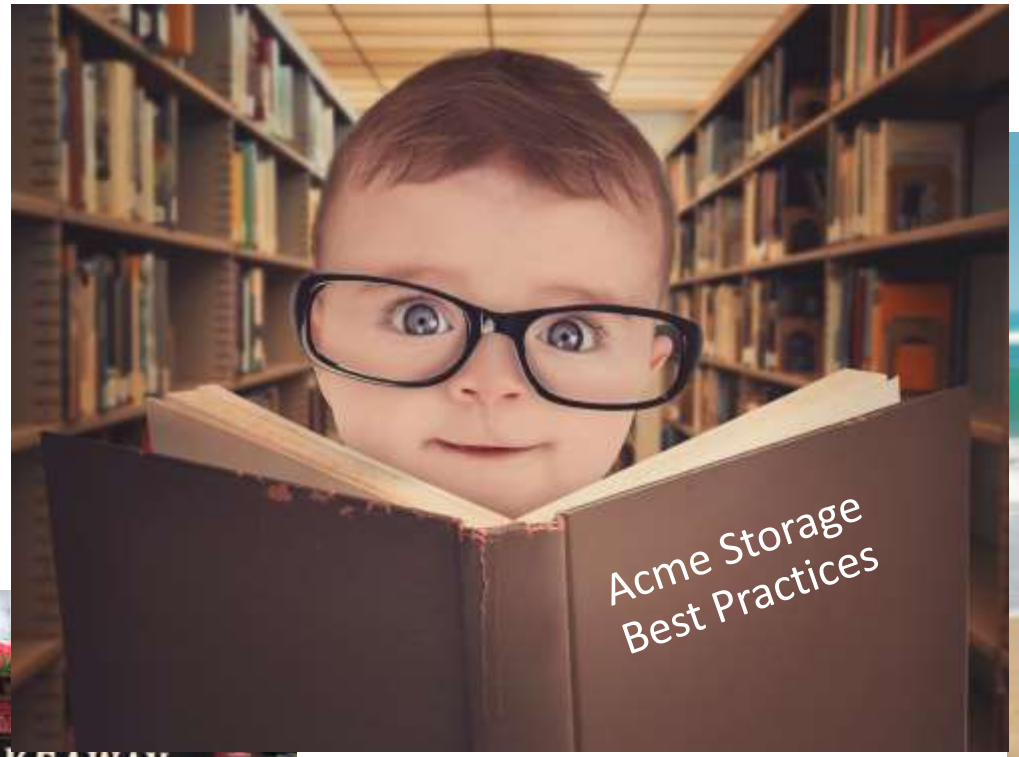
On Path To Database Virtualization/Monster VMs With a Little Help from our Friends

vmware

Doing It Right: Read Best Practices Guides

Read The Documentation

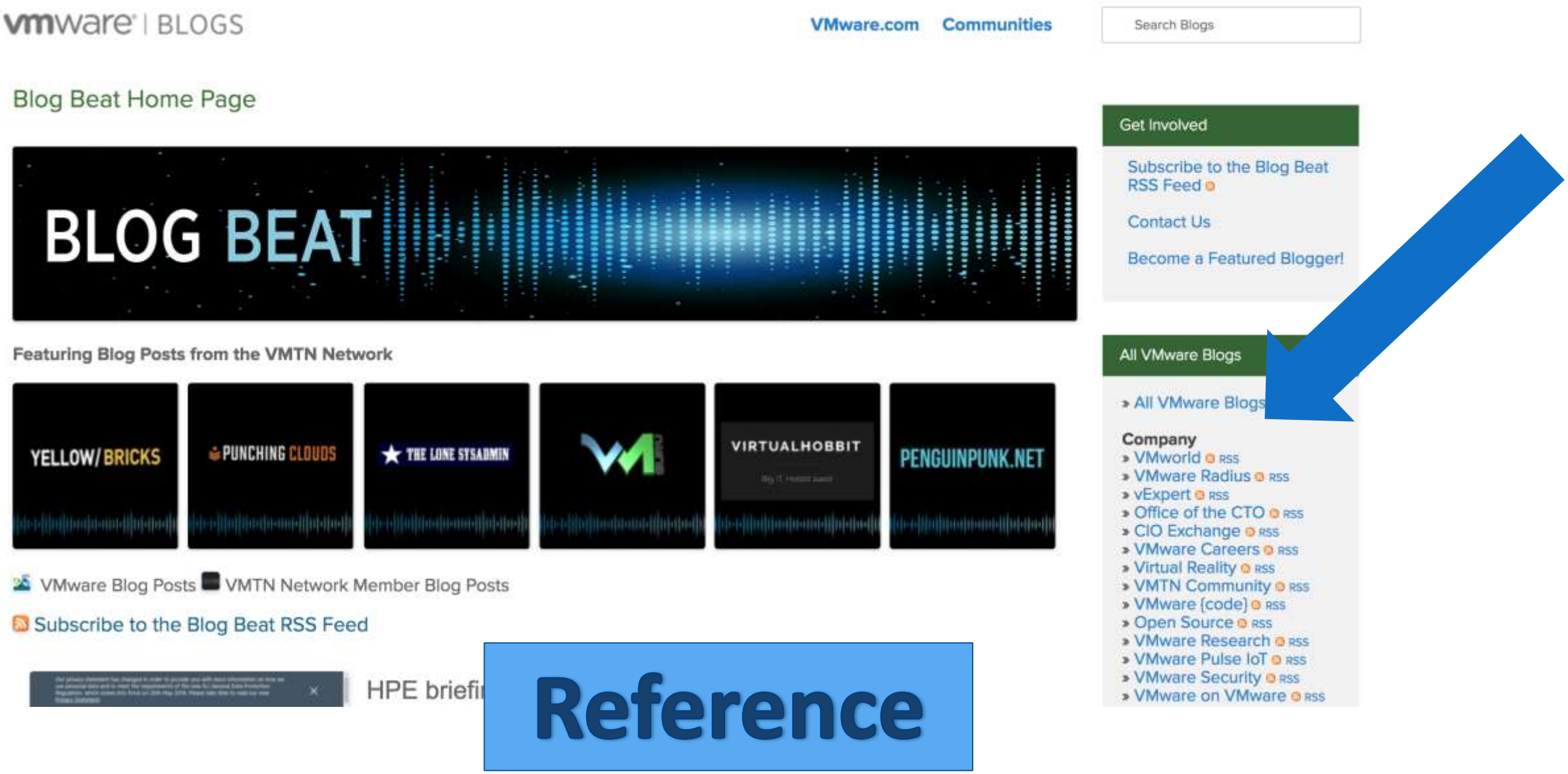
From **All** Your Vendors…… VMware, Microsoft, Oracle, **Storage Vendor**, Network Vendor….



Performance Best Practices vSphere 6.5

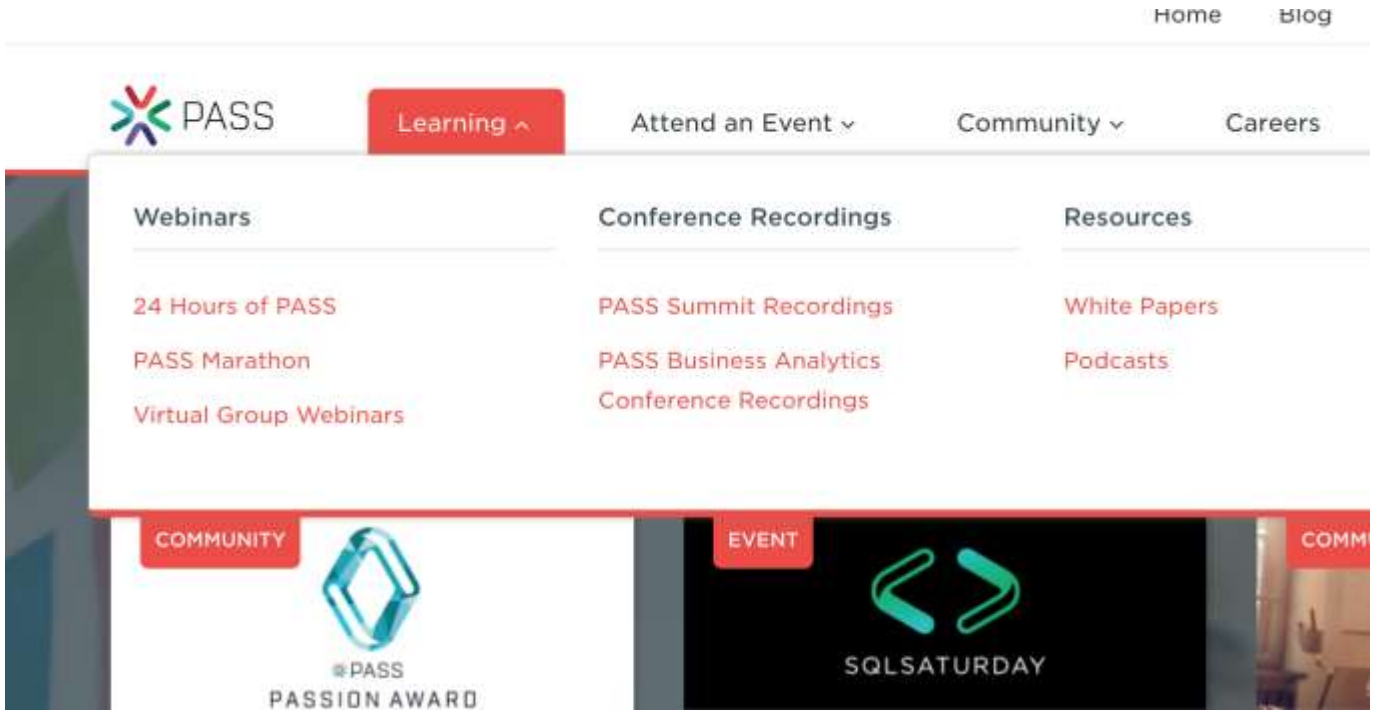
https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/solutions/sql-server-on-vmware-best-practices-guide.pdf

Useful Web Sites: blogs.vmware.com



http://www.vmsig.org/

http://www.pass.org/



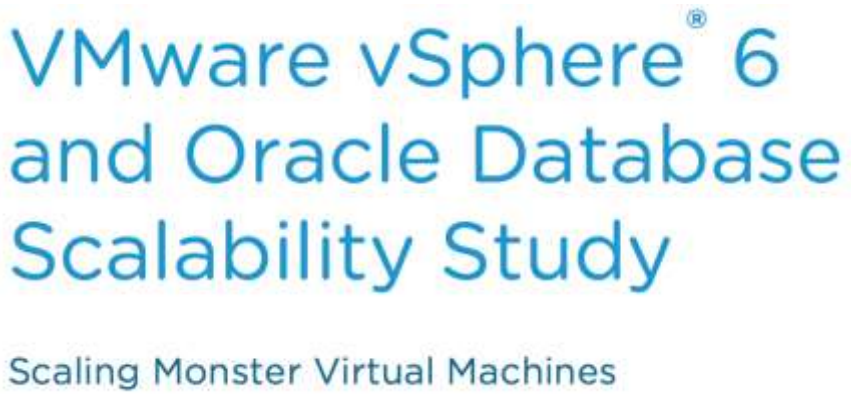
Latest Posts

Reference

Performance White Papers

Microsoft SQL Server

https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/techpaper/performance/sql-server-vsphere65-perf.pdf

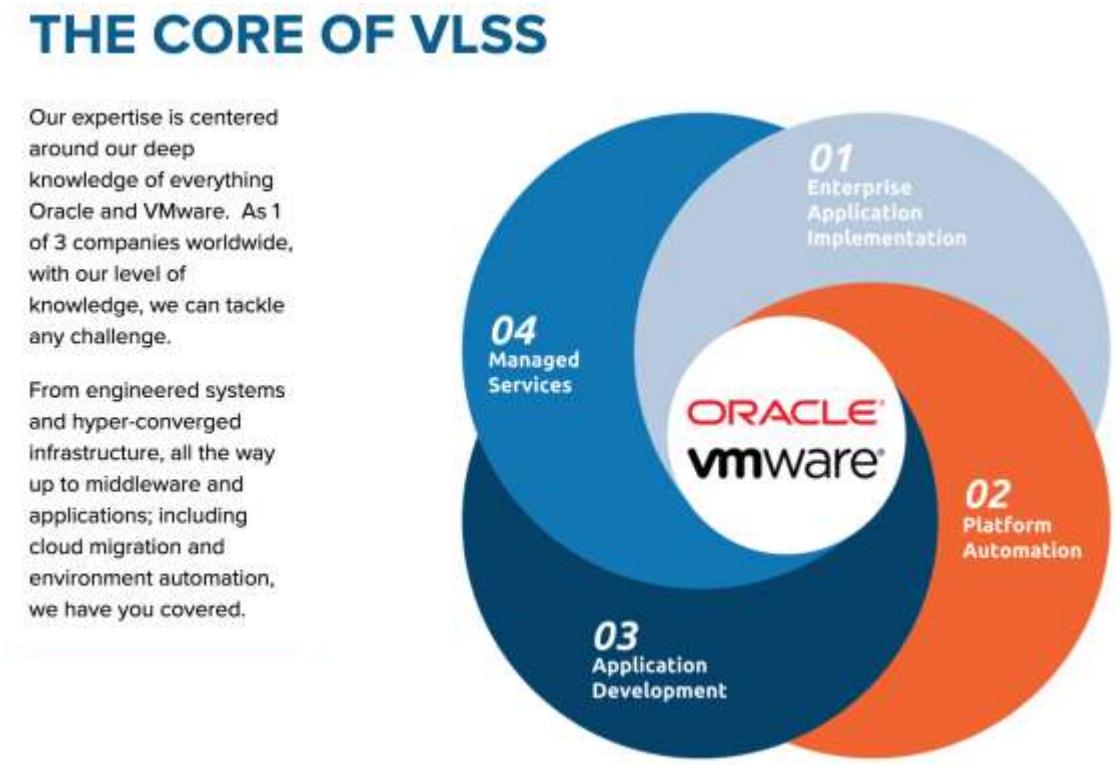


ORACLE

https://www.vmware.com/techpapers/2015/vmware-vsphere-6-and-oracle-database-scalability-s-10455.html

http://michaelcorey.com/

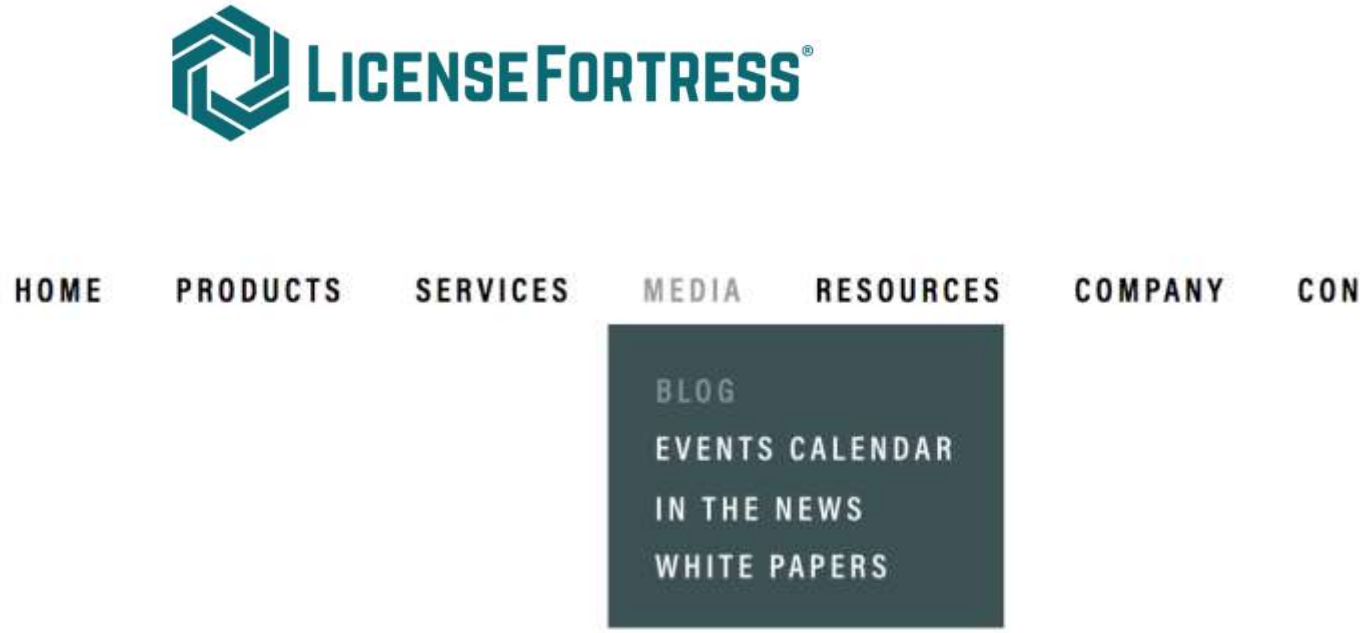
http://www.vlss-llc.com/



Reference

http://longwhiteclouds.com/

https://www.licensefortress.com/blog/



Licensing Oracle on AWS? A Word of Caution

MAY 4, 2018 · COMMENT

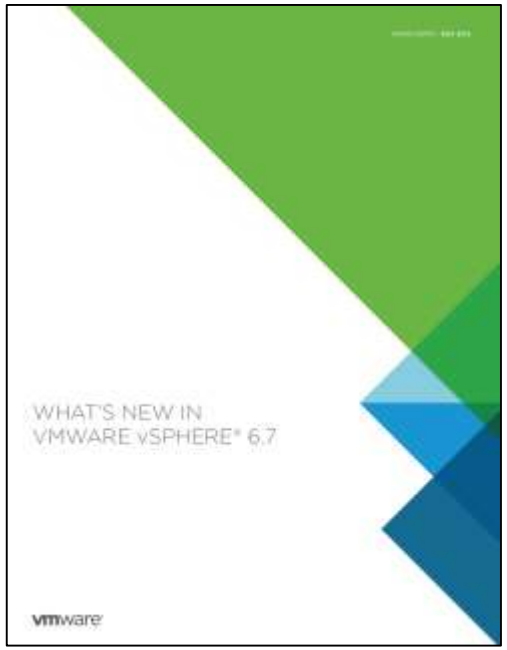
Reference

https://www.vmware.com/solutions/business-critical-apps/oracle-virtualization.html

https://www.vmware.com/solutions/business-critical-apps/sql-virtualization.html

Reference

Newest Documents on vSphere 6.7



https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/products/vsphere/vmware-whats-new-in-vsphere-whitepaper.pdf

Reference

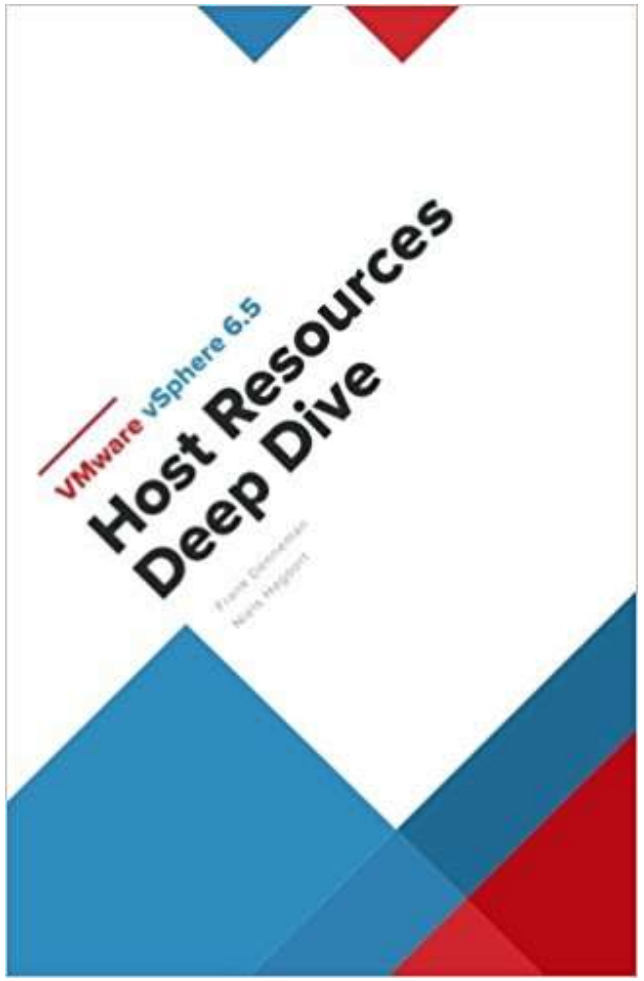
https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/techpaper/performance/vsphere-esxi-vcenter-server-67-performance-best-practices.pdf

More Useful Documentation/Resources

VMware vSphere 6.5 Host Resource Deep Dive

https://www.amazon.com/VMware-vSphere-Host-Resources-Deep/dp/1540873064/ref=sr_1_2?ie=UTF8&qid=1502235237&sr=8-2&keywords=VMware+vSphere+6.5+Host+Resource+Deep+Dive

“Also Good book for Other Hypervisors that shall remain Nameless”



Architecting Microsoft SQL Server on VMware vSphere (March 2017)
http://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/solutions/sql-server-on-vmware-best-practices-guide.pdf

Microsoft SQL Server

Installation – Plan Your SQL Server/Oracle Installation

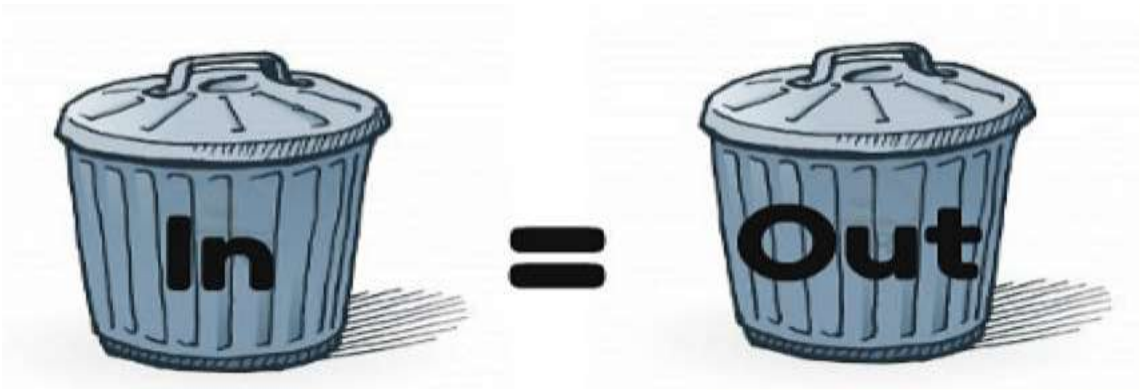
- ❑SLAs, RPOs, RTOs
- ❑**Baseline current workload**, at least 1 business cycle
- ❑**Baseline existing (workload) vSphere implementation**
- ❑Estimated growth rates
- ❑**I/O requirements** (I/O per sec, throughput, latency)
- ❑Storage (**Disk type/speed**, RAID, flash cache solution, etc)
- ❑Software versions (vSphere, Windows, SQL)
- ❑Substandard default options
- ❑Product Keys/**Licensing** (may determine architecture -)
- ❑Workload type (OLTP, Batch, Warehouse)
- ❑**High Availability strategy**
- ❑**Backup & Recovery strategy**

Ask Yourself the Question:
"What is the VM doing when its Running Well !"

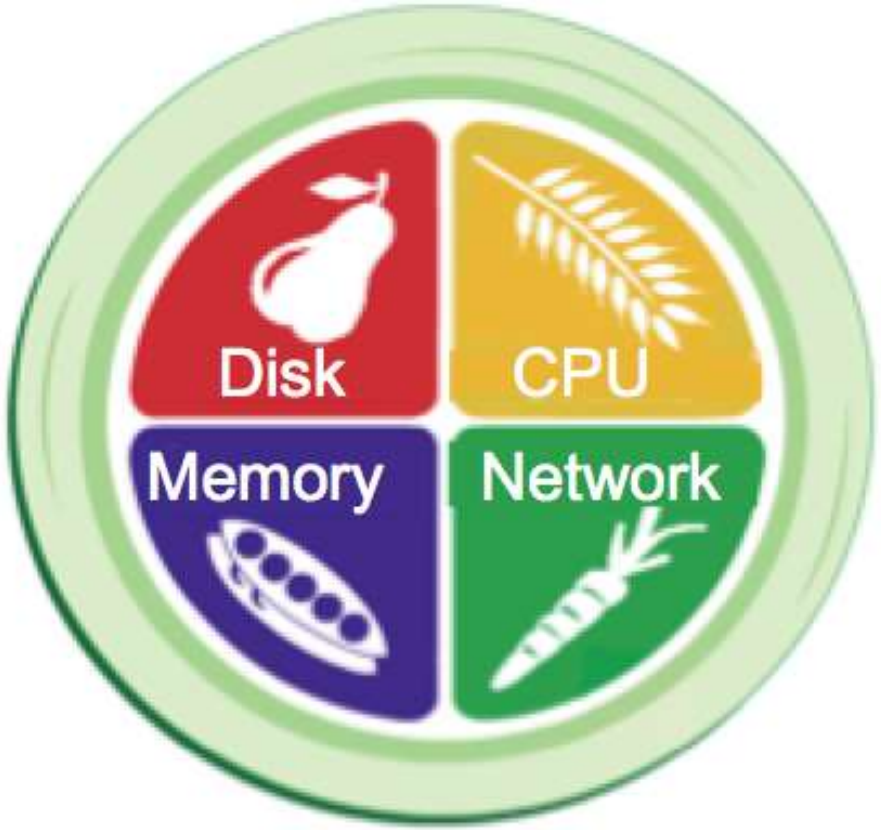
Baseline is critical for Monster VM's or any critical workload

"If you aim at nothing, you will hit it every time" – Zig Ziglar

If your Application Code S_KS…



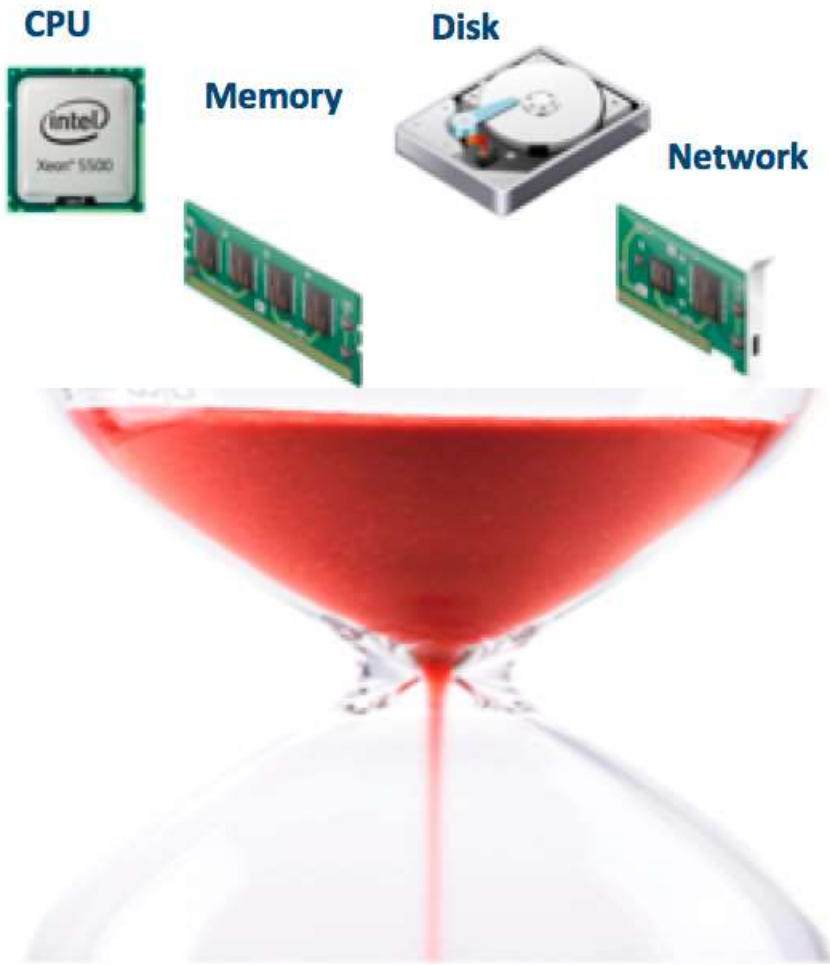
What to Base Line = “IT” Food Groups



- Existing Physical Database/Monster VM Infrastructure
- Existing/Proposed vSphere Infrastructure

Monster VM’s Consume Lots of DISK, CPU, Memory & Network
For Breakfast, Lunch, Dinner & Desert

Monster VM Choke Points



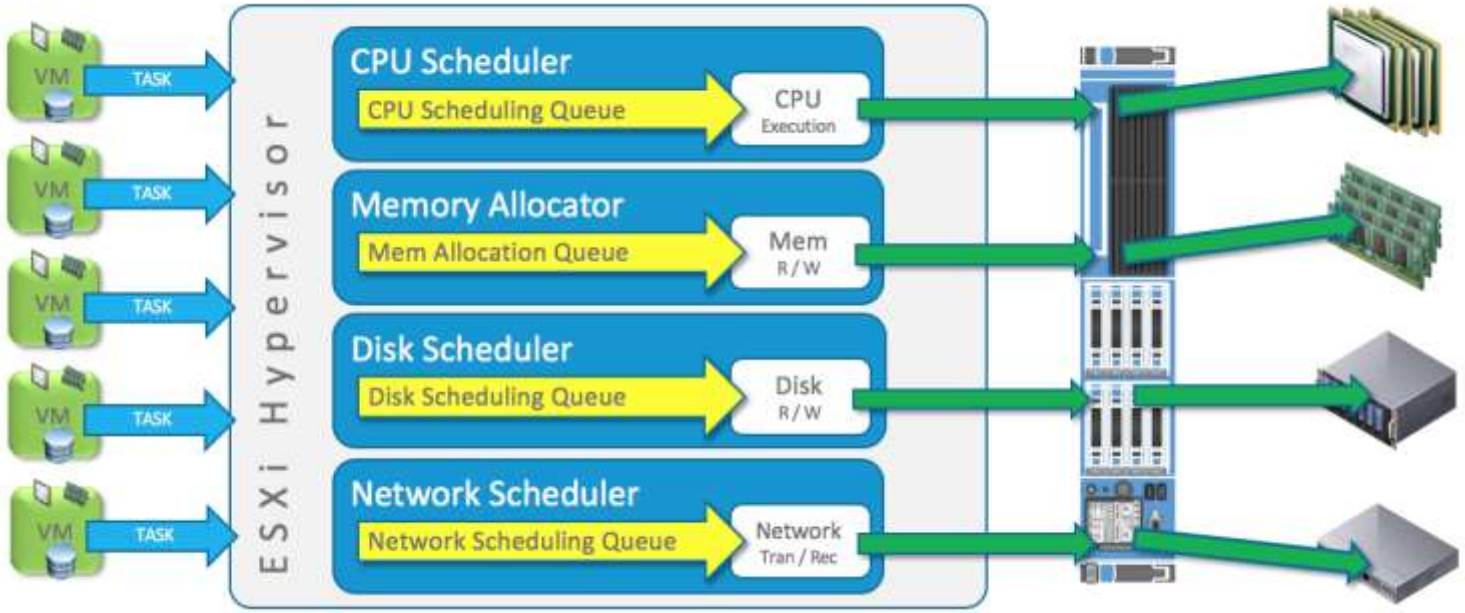


Image borrowed from VAP1452 Performance Deep Dive

"Good Baseline So Important"

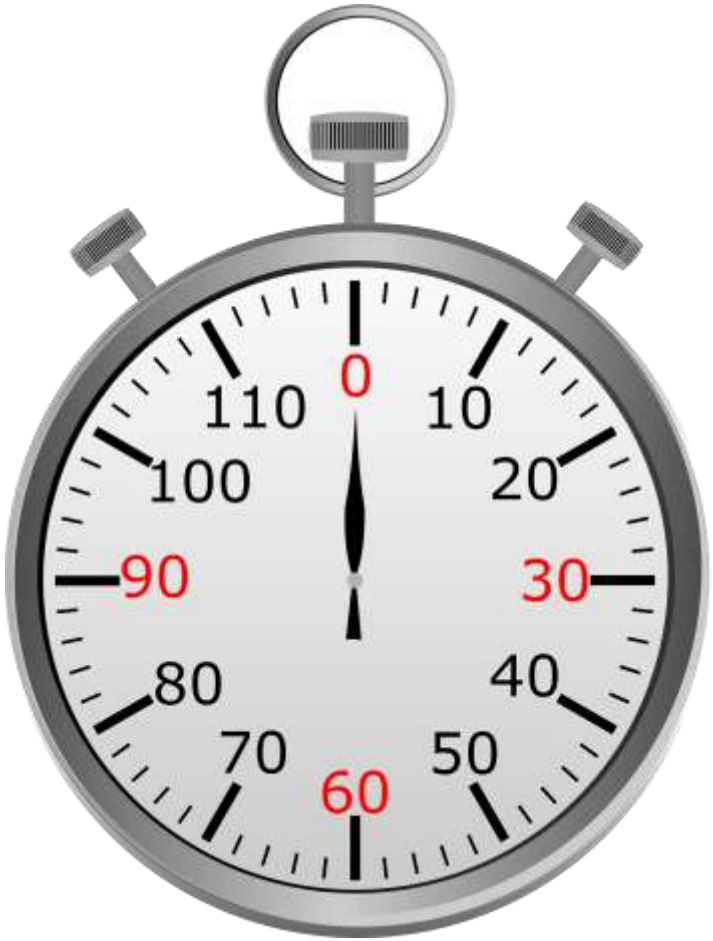
When you Baseline a Database/Monster VM

Make Sure The **Sample Interval Is frequent**

CPU, Memory, Disk (15 Seconds or less)

SQL Server TSQL (1 Minute)

“A Lot can happen in a short amount of time”



Databases, Monster VM’s, require frequent Sample Intervals.

Baseline to Lowest Level Possible – Especially with Monster VM's

By Core

By NUMA Node

By Drive

Disk Controller

Lowest is Best for Monster VM's

Source: https://www.youtube.com/watch?v=BAREgqZvHWg

When Baselining a Database/Monster VM

- **(1) When High Performance Requirements**
  - **Size the VM to the most sustained Peak**
- **(2) When Consolidation is higher priority than Performance**
  - **Size VM based on the average baseline**

Understanding Workload Resource Requirements→ Critical For Monster VM's

Basic performance characteristics (CPU, memory, IO, Network)

- **Daily average resource usage/peak resource usage**
- **Daily peak hours/Month-end, quarter-end, year-end peaks**
- Windows Perfmon (Example)
  - Processor(*) → %Processor Time
  - Process(sqlservr) → %Processor Time
  - SQLServer:Memory Manager → Total Server Memory (KB)
  - PhysicalDisk(*) → Disk Reads/Sec, Disk Writes/Sec
  - PhysicalDisk(*) → Disk Reads Bytes/Sec, Disk Write Bytes/Sec
  - Network Interface(*) → Bytes Received/Sec, Bytes Sent/Sec

REFERENCE

How to setup ongoing Perfmon Collection

http://www.davidklee.net/articles/sql-server-articles/perfmon

vSphere Environment

Use ESX TOP	http://kb.vmware.com/kb/1006797			
Resource	**Metric**	**Host / VM**	**Description**	**Threshold**
CPU	% USED	Both	CPU used over the collection interval (%)	
	%RDY	VM	CPU time spent in ready state	10
	%SYS	Both	Percentage of time spent in the ESX Server VMKernel	20
	%CSTP	Host	Percentage of time the world spend in ready, co-descheduled state (make sure largers SMP VM's are effective and not under contention)	< 3%
Memory	Swapin, Swapout	Both	Memory ESX host swaps in/out from/to disk (per VM) or cumulative over host	
	MCTLSZ (MB)	Both	Amount of memory reclaimed from resource pool by way of ballooning	0
	%SWPWT	Host	VM waiting swapped pages to be read from disk	
	CACHEUSED	Host	Compressed memory	0
Disk	READs/s, Writes/s	Both	Reads & writes issued in the collection interval	
	CMDS/s	Both	Number of IOPS being sent to or coming from the device or virtual machine being monitored	
	DAVG/cmd	Both	Average latency (ms) of the device (LUN)	Target is 10ms for ESX Hosts running DBs
	KAVG/cmd	Both	Average latency (ms) in the Vmkernel (aka queuing time). For databases we want this at or below 1 ms; other workloads OK at 2ms	1
Network	MbRX/s, MbTx/s	Both	Amount of data transmitted per second	
	PKTRX/s PKTTX,s	Both	Packets transmitted per second	1
	%DRPPX, %DRPTX	Both	Dropped packets per second	1

vSphere 6 Resource Management Guide ← Ties it All Together

http://pubs.vmware.com/vsphere-60/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-60-resource-management-guide.pdf

Half Way Point

Determine IOPS & Overall Throughput

SLOB (Silly Little Oracle Benchmark)

Calibrate I/O – Native to Oracle starting in 11.1

```
SQL> declare
  2    l_latency   integer;
  3    l_iops      integer;
  4    l_mbps      integer;
  5  begin
  6    dbms_resource_manager.calibrate_io
  7    (5,10,l_iops,l_mbps,l_latency);
  8    dbms_output.put_line ('max_iops = '||l_iops);
  9    dbms_output.put_line ('latency = '||l_latency);
 10    dbms_output.put_line ('max_mbps = '||l_mbps);
 11  end;
 12  /
max_iops = 5348
latency  = 10
max_mbps = 641
```

Other Free Tools:

- Swingbench
- TPC Benchmarks
- Custom scripts

How do you know for sure?

Oracle's - $$$:

Database Replay

ORACLE

Google "Oracle SLOB" - Wealth SLOB community information.

http://kevinclosson.net/slob/

Diskspd: Robust Storage Testing Tool (Replaces SQLIO)

More granular storage testing methodology

Sub-microsecond latency values **very important with today's all-flash** and **hybrid storage devices**

Diskspd.exe -b8K -d60 -h -L -o2 -t4 -r -w30 -c50M c:\io.dat

Example: Set the block size to 8K, run the test for 60 seconds, disable all hardware and software caching, measure and display latency statistics, leverage 2 overlapped IOs and 4 threads per target, random 30% writes and 70% reads and create a 50MB test file at c:\io.dat

Use SQL Server Distributed Replay

Diskspd Utility: A Robust Storage Testing Tool (superseding SQLIO)

https://gallery.technet.microsoft.com/DiskSpd-a-robust-storage-6cd2f223

Shared Environment – Don't Keep Your Needs a Secret

DBA's – tell vSphere, Storage, and Network Admins your needs

- Storage: (IOPS / throughput)
- CPU: (MHz)
- Memory: (Total GB)
- Network: Bandwidth
- Features (i.e.: Windows clustering)
- Anticipated Growth Rates
- Anticipated Activity

"Proper HA Requires Both Sides to Work Together"

For Monster VM's Critical to Communicate Resource Requirements

"They Flunked Mind Reading"

SQL Server HA,
Host Based Affinity Rules
Common Mistake

I told you I needed 12 vCPU's Not 8

Before You Install a Database/Monster VM

Do basic throughput testing of the IO subsystem prior to deploying a Database

Tools you can use

- **DISKSPD (Replaces SQLIO)**
- **Slob**
- **IOMETER**…..
- SQL Server Distributed Replay

ORACLE

Microsoft SQL Server

“Check It Before You Wreck it”

-- Jeff Szastak

P → V (Via Converter)/ Physical to Virtual

Production Environment's Build
"New" From Scratch – GI/GO

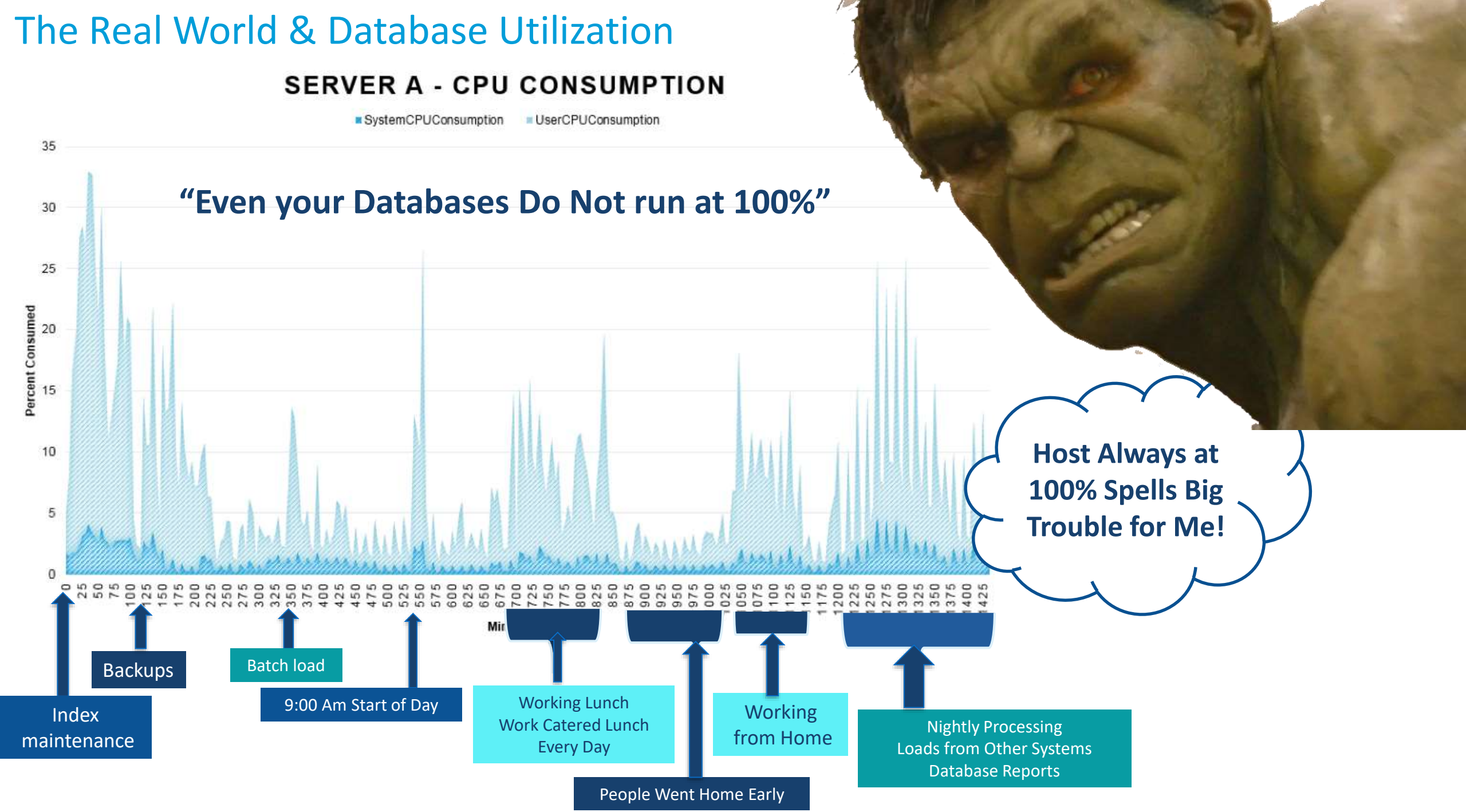
Installation: Same As Physical

Use SQL Server/Oracle recommended installation guidelines for respective operating system – same as physical !

ORACLE

vSphere Does Not the Change O/S Stack

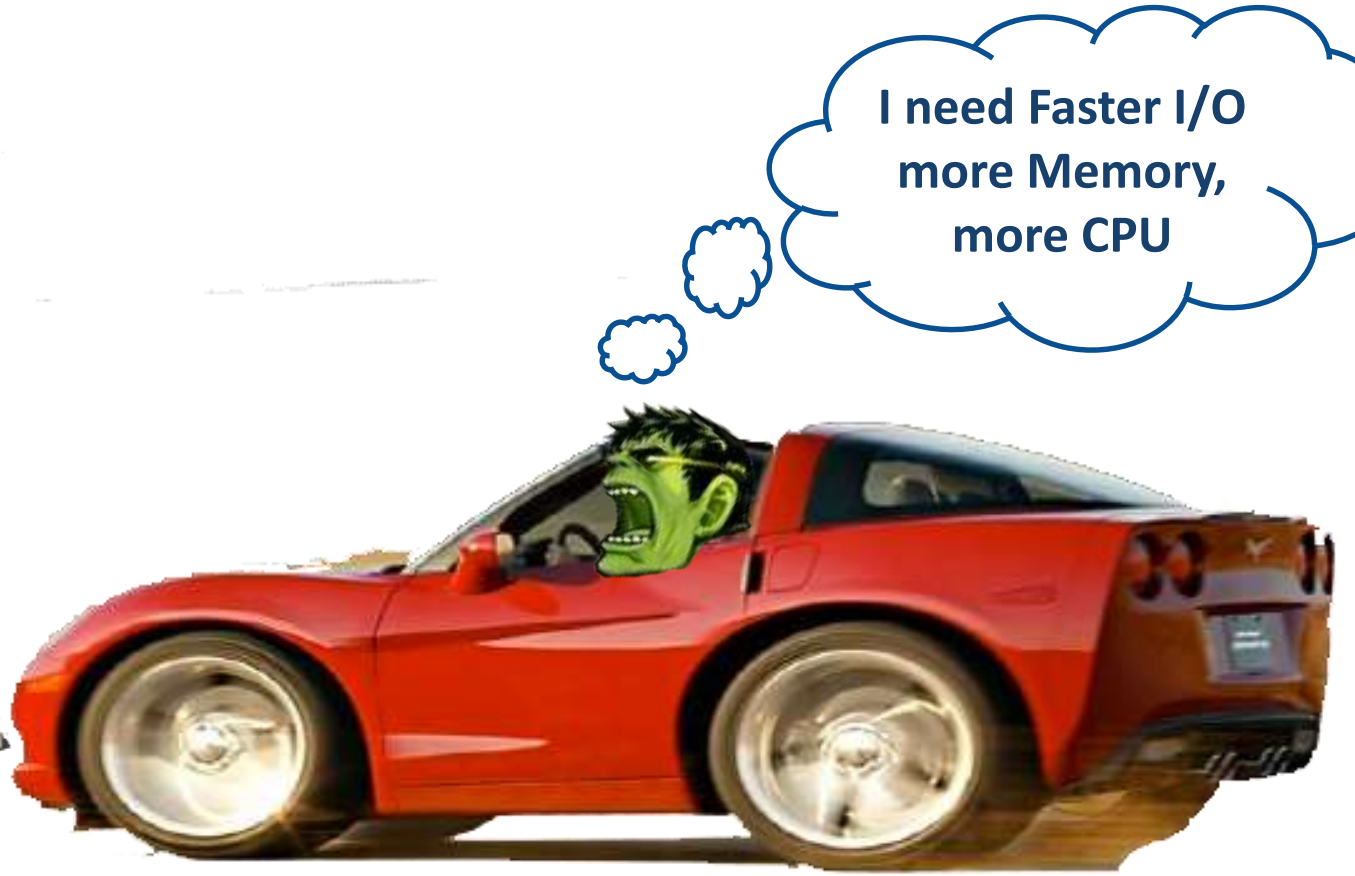
The Real World & Database Utilization



Architecting For Performance

Design

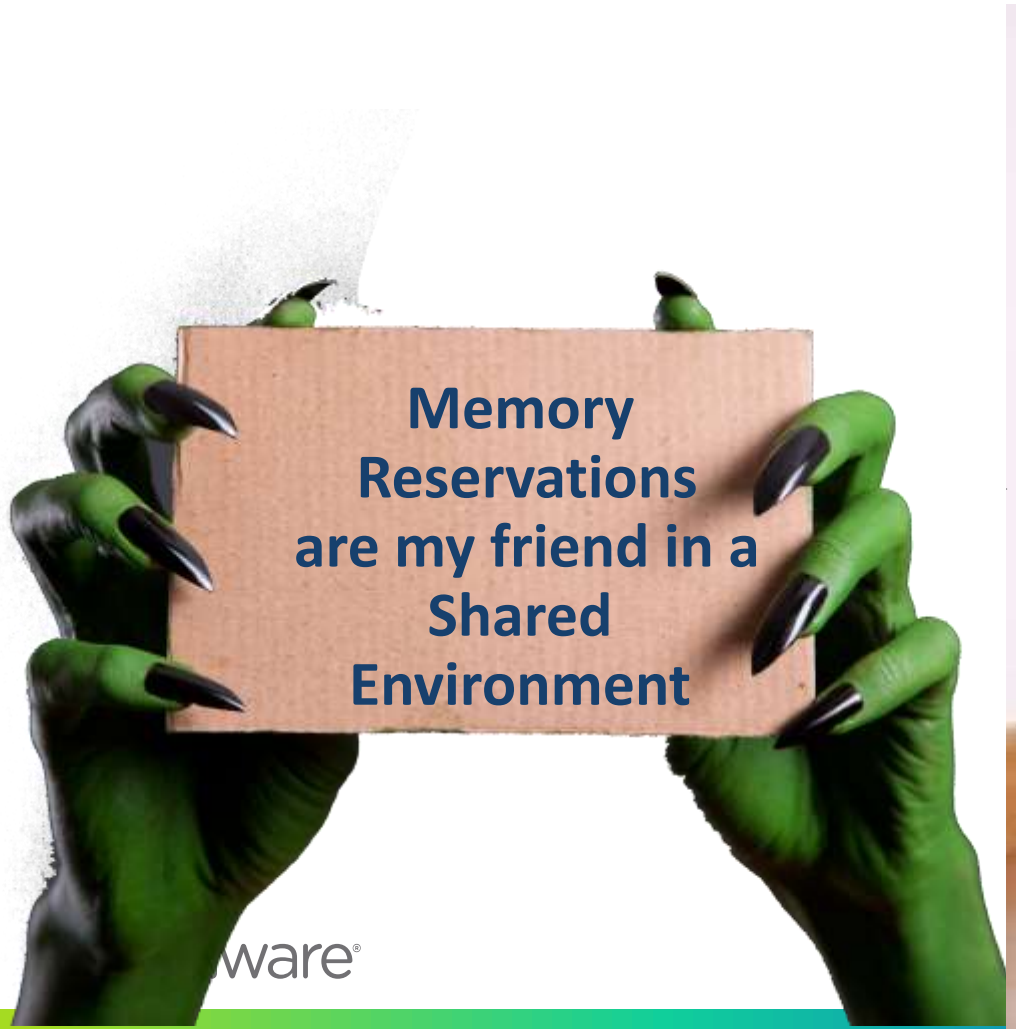
ier-2 Was Built for Consolidation/Capacity Not Performance



Monster VM's Don't Always Play Well with Others

Separate development, test from production environments into different host clusters in the beginning !

Maximize Your Licensing –
(Consider the Cost of DB License/Don't Let Dev Dilute your investment)



Architecting For Performance

Storage

Storage: Latency Matters More Than IOPS

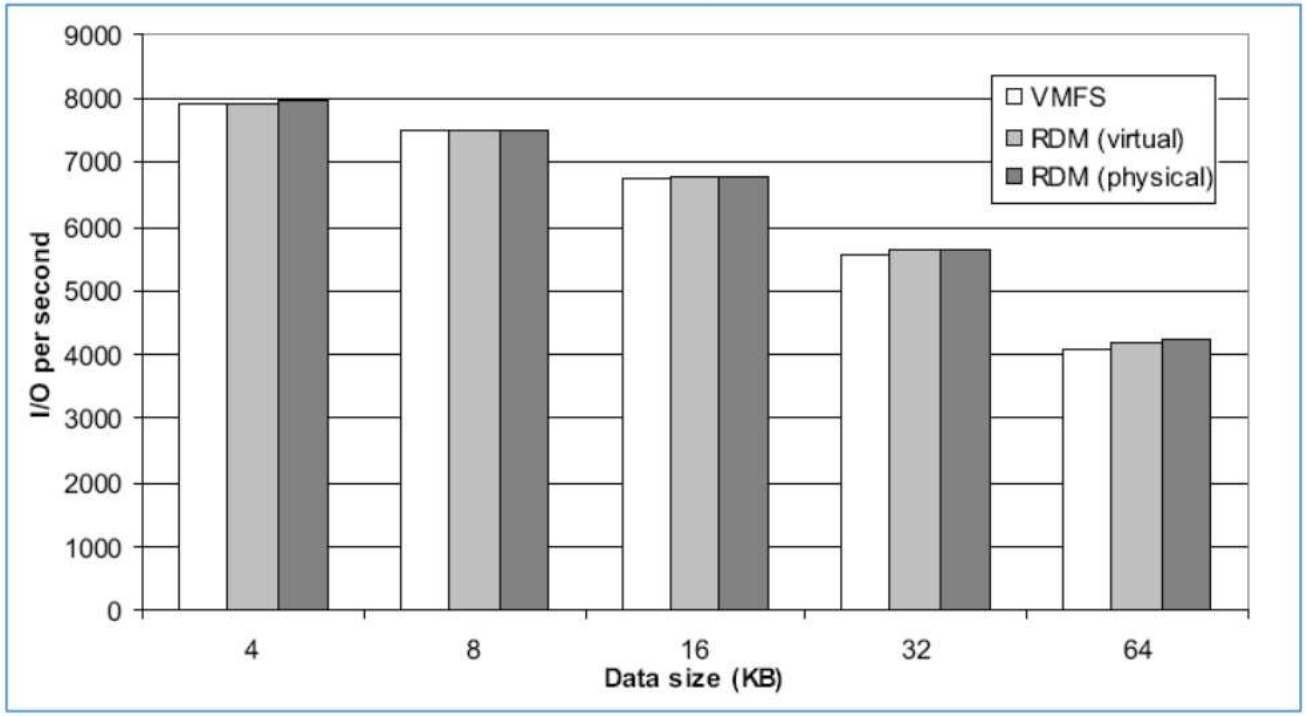
"When was the last time you hit 200 MPH on your Jaguar?"

(Maxed IOPS Claims from vendor)

"When was the last time you had to go 0-60 in 2.8 Seconds"

VMFS vs RDM

Figure 19. Random Mixed (50% Read/50% Write) I/O Operations per Second (Higher is Better)



- **VMFS Advantages**
  - Negligible performance cost and superior functionality
  - Ability take full advantage of future functionality enhancements
- VVol is worth considering

Source: Architecting Microsoft SQL Server on VMware vSphere - Best Practices Guide March 2017

Microsoft® SQL Server®

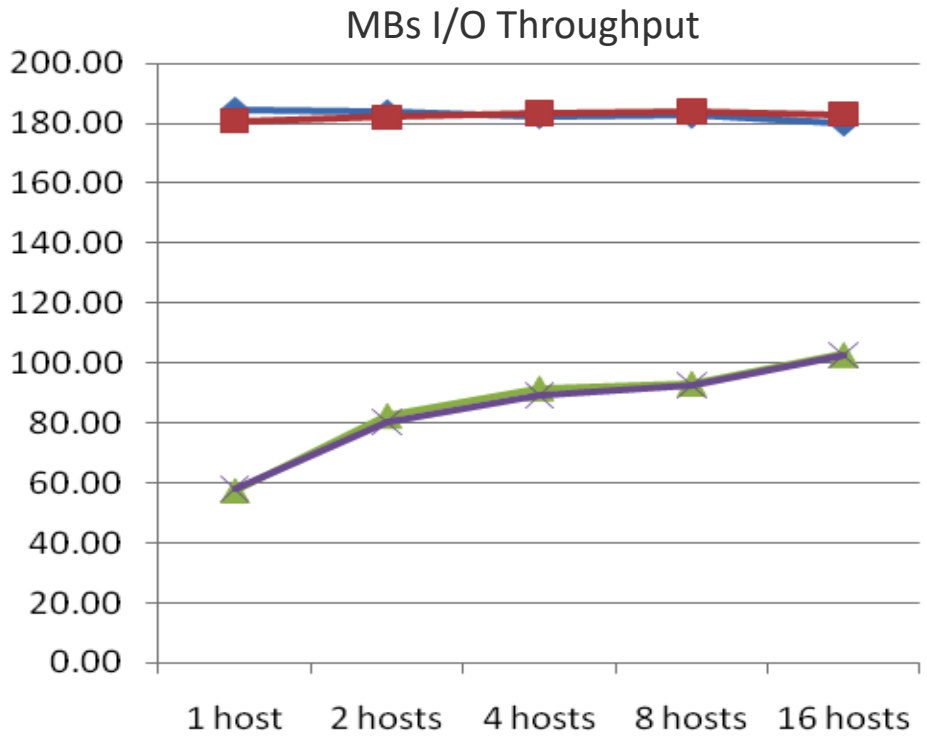
When Using SQL Server FCI*, you should use VVols (or RDM pre 6.7)

*FCI – Failover Cluster Instances

Thin Provisioning Performance/ Block Zeroing (Traditional Arrays)

USE **Thick Eager Zeroed Disk for best performance** (Don't use "Quick Format" option for database/Log Volumes)

- Maximum Performance happens eventually, but when using lazy zeroing, **zeroing needs to occur before you can get maximum performance**
- **At minimum Databases, LOGS, TEMPDB**
- **Check with Storage Vendor to see how they handle Thin Provisioning. Your Mileage may vary**
- **VAAI capable array can alter configuration**



ORACLE

Microsoft SQL Server 2016

"First Write Penalty"

vmware

Thin Provisioning - can lead to oversubscription

Thin Provisioning lets you overcommit the datastores

Not a fan of Thin Provisioning Except…

allow a VM to run with just the storage it needs, and to avoid giving a VM storage that it *might* use sometime in the future.

"Context of Mission Critical Workloads"

Cormac Hogan Blog on:
http://blogs.vmware.com/vsphere/2012/03/thin-provisioning-whats-the-scoop.html

Virtual Volumes (VVols)

Direct exposure of SAN LUN(s) from SAN

Management via APIs

Treated as virtual disks from VMware

Bypass VMFS

Useful for performance

Useful for WSFC shared storage

Can convert VMDK <-> VVol

- Don't do with SQL Server FCI

"The best of VMFS and RDM Combined", Dean Bolton

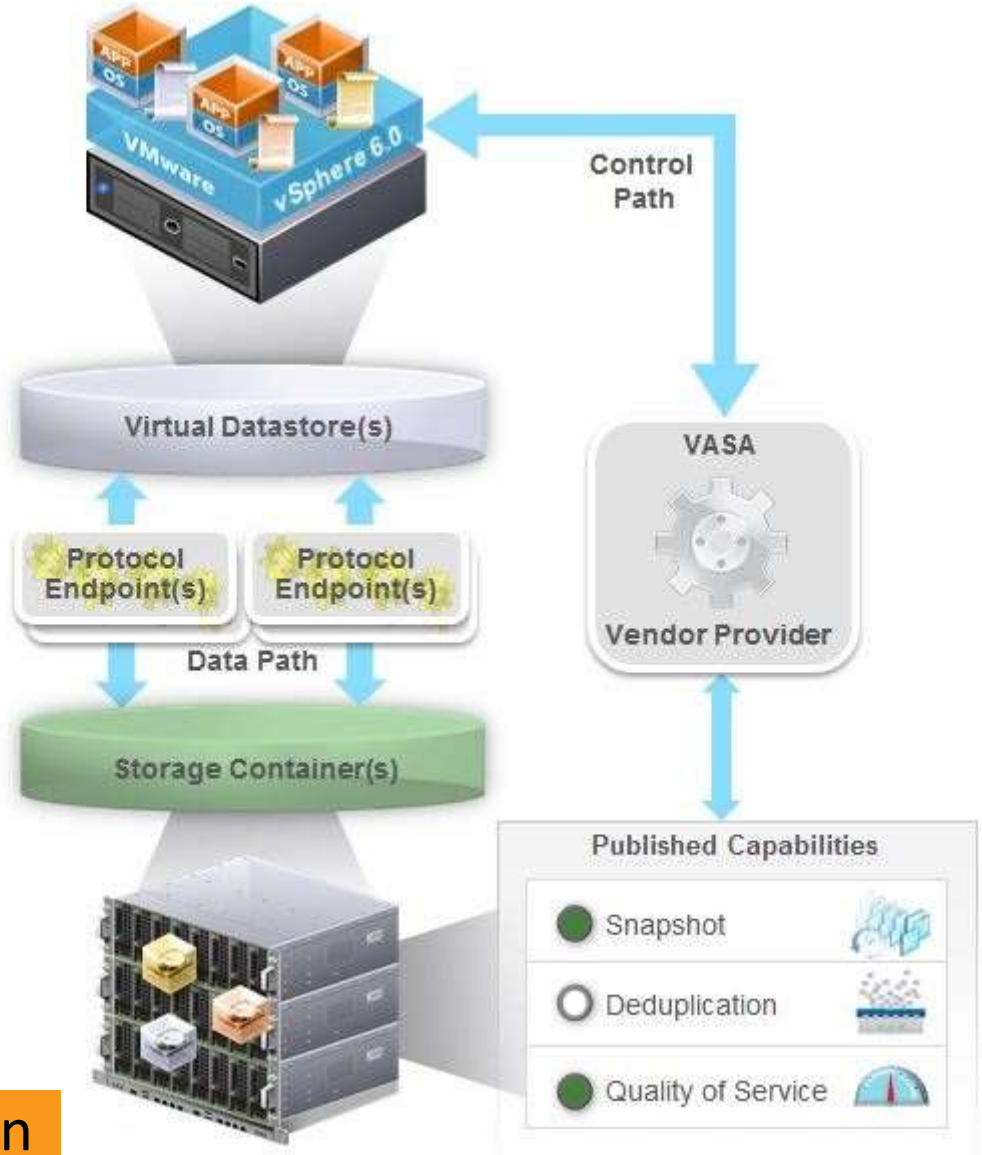


Image source: https://kb.vmware.com/s/article/2113013

Storage Best Practices:

Queue depths should be left to default

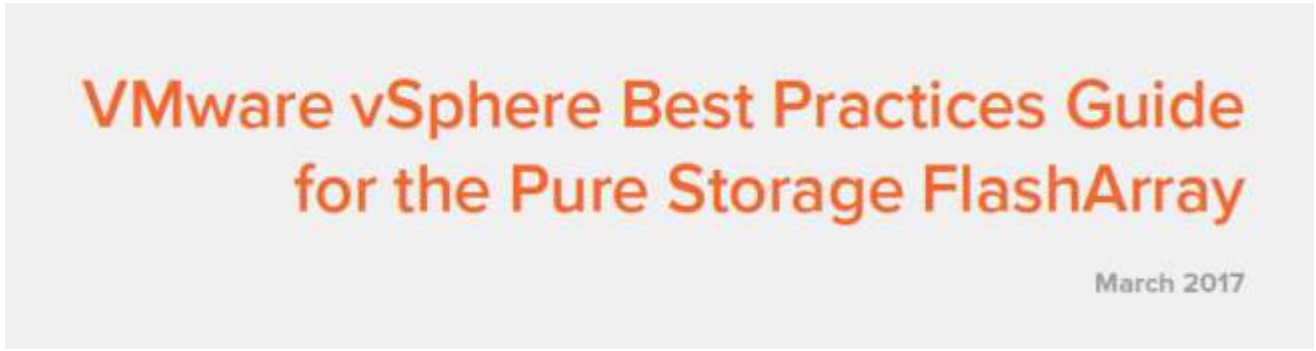
- Only adjust if there is a performance problem

Use Paravirtual SCSI Adapters (Need VMware Tools)

Use Active Multi-pathing (Configure server with multiple paths to storage)

- For Array Volumes Change Round Robin I/O from 1000 to 1

Use Latest Version of VMFS (Note VMFS-6 Not Default for vSphere 6.5+)



https://support.purestorage.com/Solutions/VMware_Platform_Guide/001VMwareBestPractices/PDF_Guide%3A_VMware_vSphere_Best_Practices_for_the_Pure_Storage%C2%AE_FlashArray

Storage: Paravirtual SCSI (PVSCSI) Adapters

PVSCSI adapters are high-performance storage adapters that can result in **greater throughput** and **lower CPU** utilization.

Up to 30% CPU Savings

Up to 12%-30% I/O Improvement (Faster storage better IO improvement)

"Use for all I/O intensive virtual machines"

4 Virtual SCSI adapters per VM*

NVMe adapter?

*Configuration Maximums vSphere 6.0+

Microsoft SQL Server 2016

ORACLE

4 Virtual SCSI adapters per VM*

Just as 4 Straws are faster than 1
So are 4 Virtual SCSI Adapters

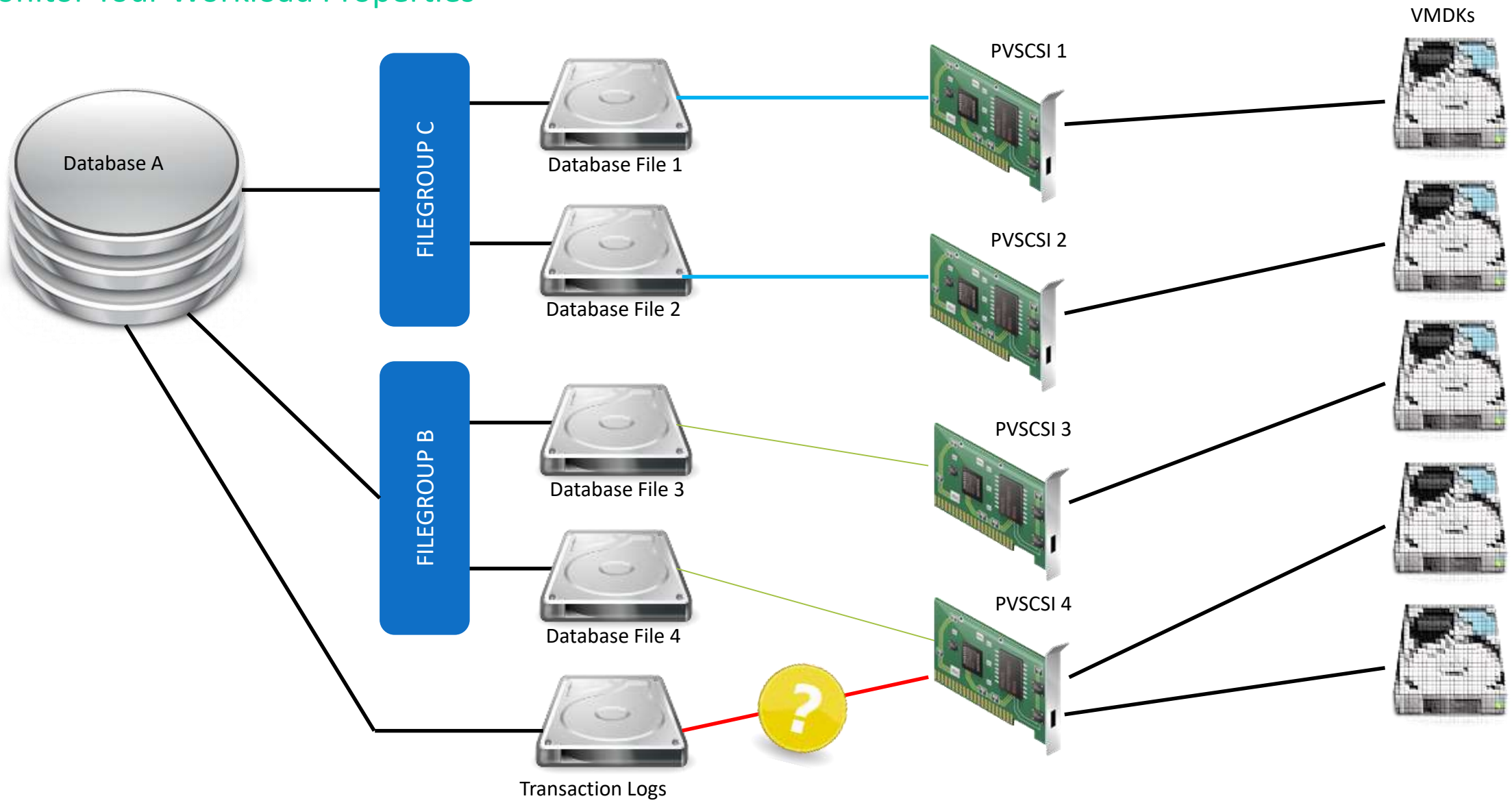
Follow KB 2053145 for large-scale I/O Intensive database deployments

Large-scale workloads with intensive I/O patterns might require queue depths significantly greater than Paravirtual SCSI default values

*Configuration Maximums vSphere 6.7

SQL Server Object Placement

Monitor Your Workload Properties



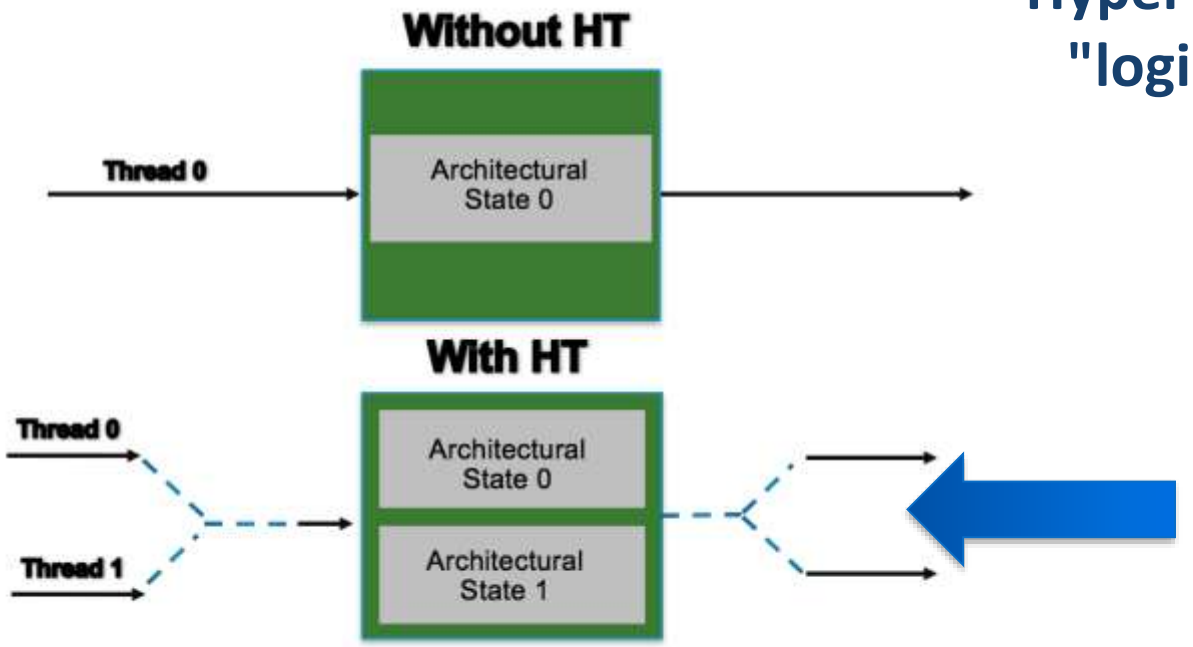
Always Check Storage Vendors Best Practices

“>80% of the issues in a virtualized Environment have to do with Storage misconfigurations”

Architecting For Performance

Processor

vCPU’s & Hyper-Threading



Hyper-Threading processor to appear as two "logical" processors to the host operating system

“CPU Intensive Workloads Could Get Slowed down by Hyper-Threading

vCPU's – Don't Over Commit (Out of the Gate)

- 1-1 Ratio **Physical Cores to vCPU's**
  - **Out of the gate !**
  - **Over Commit & Monitor Afterwards 3-1 Attainable**

"+-20% Uplift from a Non-Hyper-Threaded CPU"
5 CPU's = 6 vCPU's

Common Knowledge
VMware and Microsoft Both Support 2 to 1 Over commit

Over Commit Carefully if you care about me

Larry Ellison Announced – He Would Make Oracle Cloud Cheaper than All Other Clouds *

Approved Vendors

This policy applies to cloud computing environments from the following vendors: **Amazon Web Services – Amazon Elastic Compute Cloud (EC2), Amazon Relational Database Service (RDS)** and **Microsoft Azure Platform (collectively, the 'Authorized Cloud Environments')**. This policy applies to these Oracle programs.

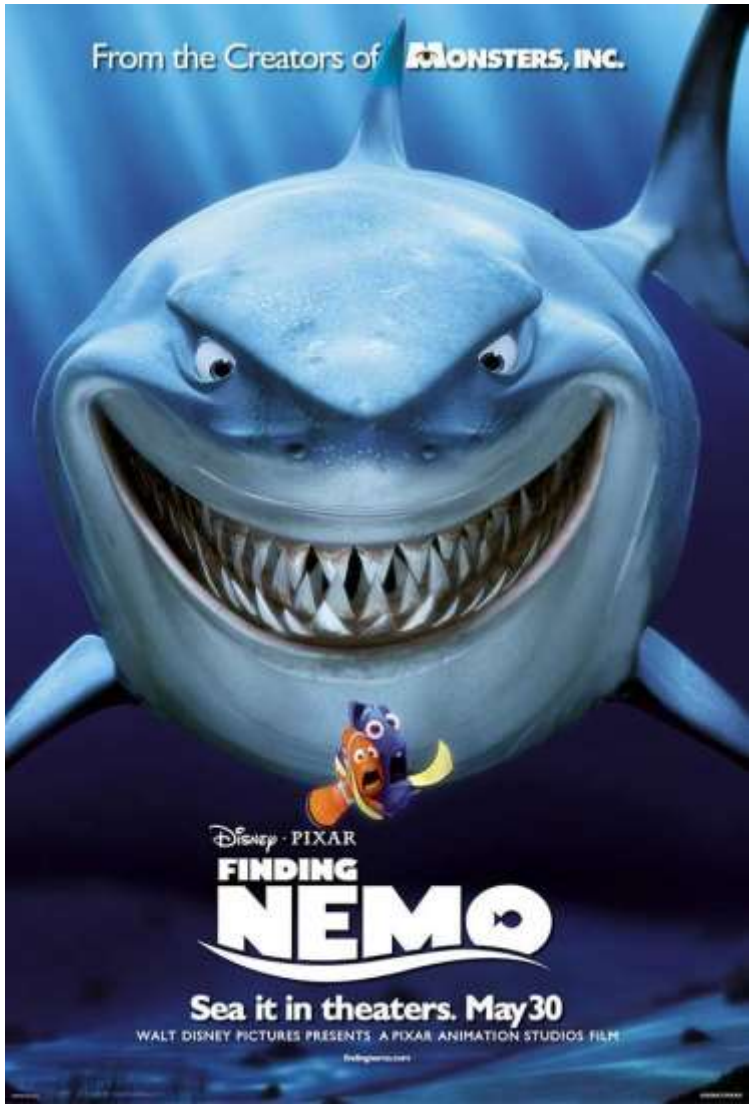
For the purposes of licensing Oracle programs in an Authorized Cloud Environment, customers are required to count as follows:

- **Amazon EC2 and RDS** - count two vCPUs as equivalent to one Oracle Processor license if hyper-threading is enabled, and one vCPU as equivalent to one Oracle Processor license if hyper-threading is not enabled.
- **Microsoft Azure** – count one Azure CPU Core as equivalent to one Oracle Processor license.

Straight Talk on Oracle on vSphere Licensing
http://www.dbta.com/emc/

New Approach to Dealing with Oracle Licensing
http://licensefortress.com/

"Oracle Hyper-threading – Tax"

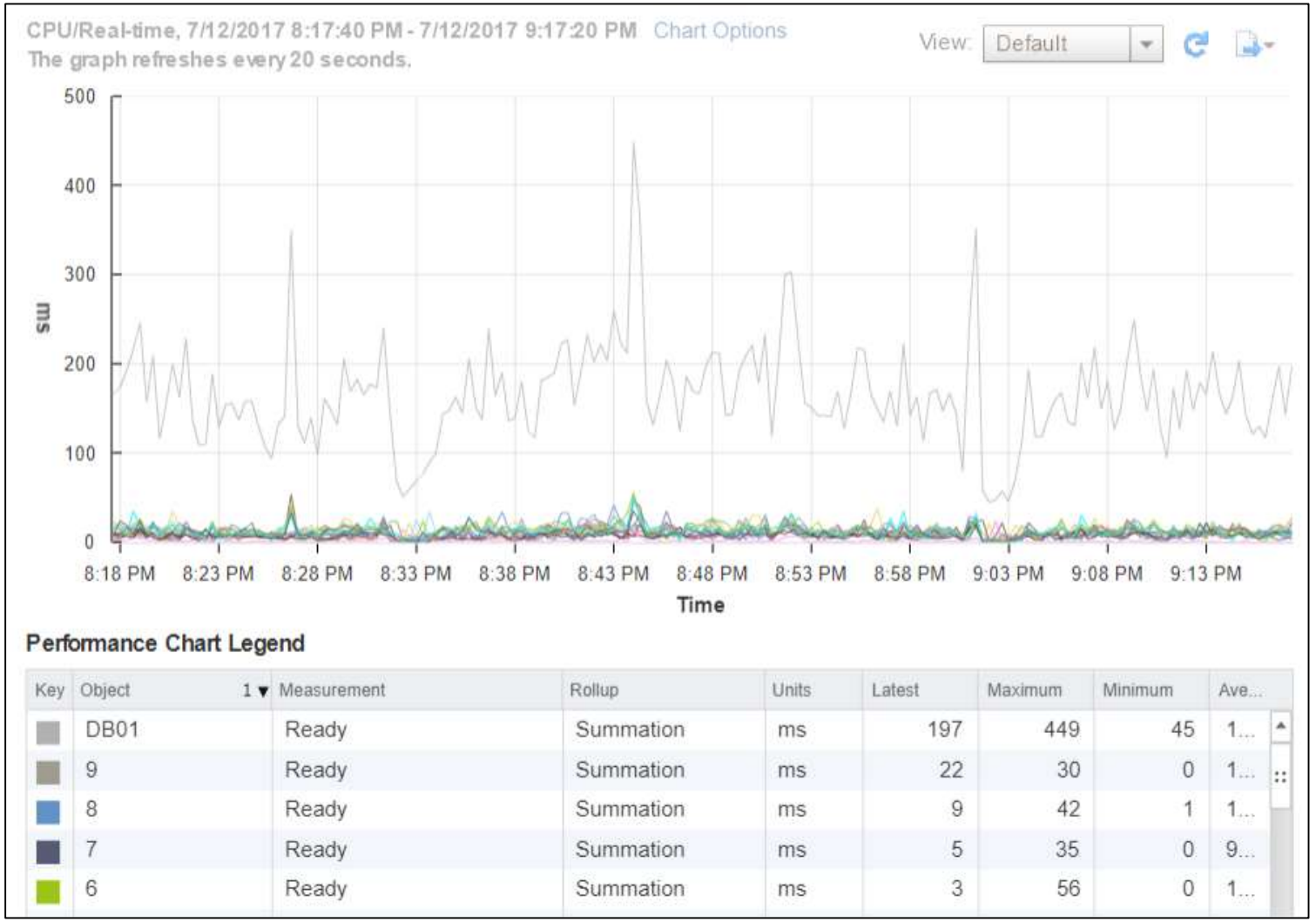


VMware CPU Ready Time – Is My Host Overloaded

CPU Ready metric is used to see a **percentage of time** that the **virtual machine was ready,** but **could not get scheduled to run on** the **physical CPU**.

VMware – CPU Ready Time

- Measured in milliseconds
- Sum total value or individual core values
- Fixed 20-second sample interval
- (Sum total / # cores / 20000ms) * 100%
- (Per core total / 20000ms) * 100%
- = Avg. percent perf loss



Key	Object	Measurement	Rollup	Units	Latest	Maximum	Minimum	Ave...
	DB01	Ready	Summation	ms	197	449	45	1...
	9	Ready	Summation	ms	22	30	0	1...
	8	Ready	Summation	ms	9	42	1	1...
	7	Ready	Summation	ms	5	35	0	9...
	6	Ready	Summation	ms	3	56	0	1...

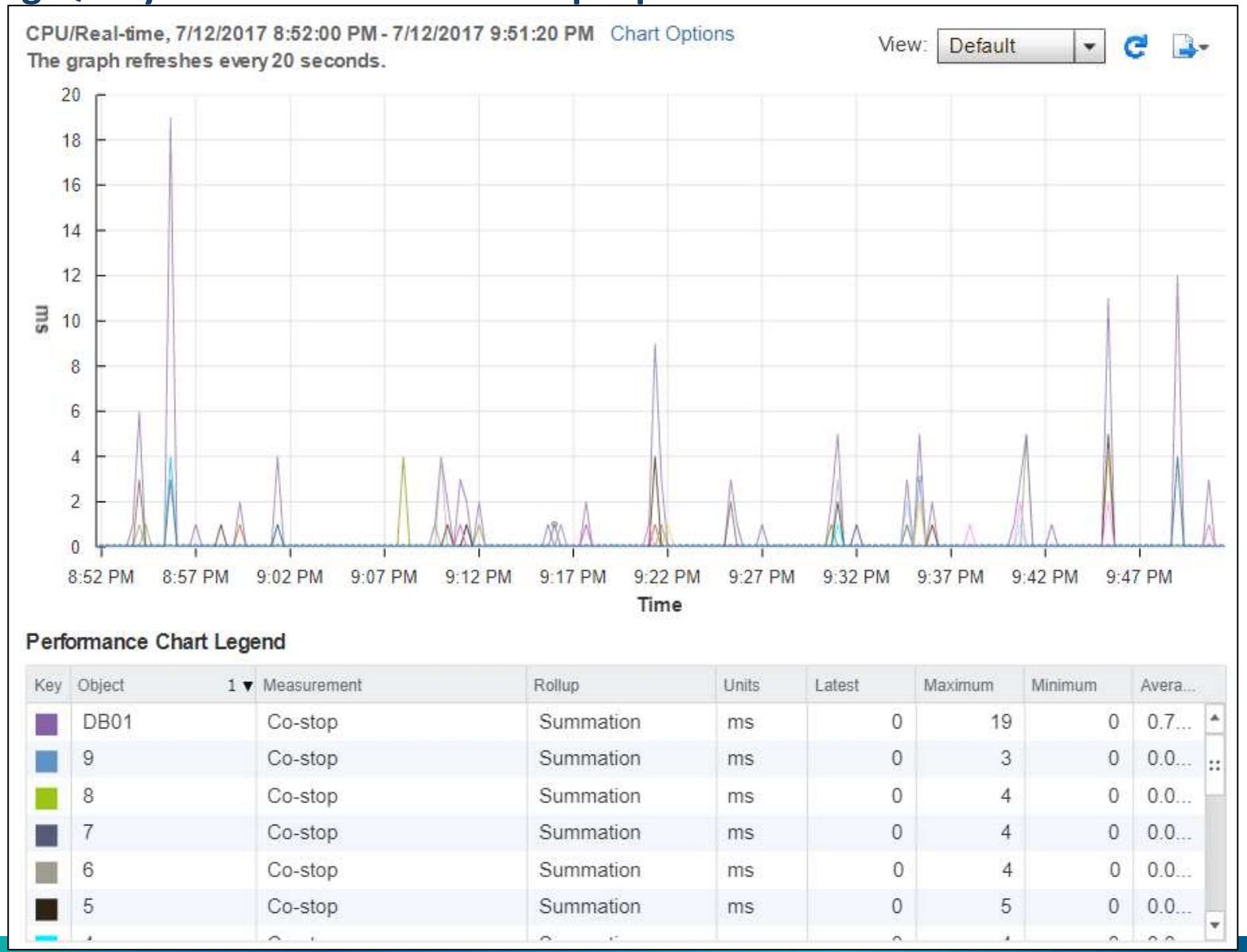
SMP vCPU Schedule Balancing (See with Co-Stop)

VMware CPU Co-Stop (VM's with Large vCPU Counts)

Amount time a vCPU is suspended waiting for the others in a parallel –
Any Good Reporting Query would break into multiple paths

VMware – Co-Stop

- Measured in milliseconds
- Sum total value or individual core values
- Fixed 20-second sample interval
- Look for sustained stretches



Performance Chart Legend

Key	Object	Measurement	Rollup	Units	Latest	Maximum	Minimum	Avera...
	DB01	Co-stop	Summation	ms	0	19	0	0.7...
	9	Co-stop	Summation	ms	0	3	0	0.0...
	8	Co-stop	Summation	ms	0	4	0	0.0...
	7	Co-stop	Summation	ms	0	4	0	0.0...
	6	Co-stop	Summation	ms	0	4	0	0.0...
	5	Co-stop	Summation	ms	0	5	0	0.0...

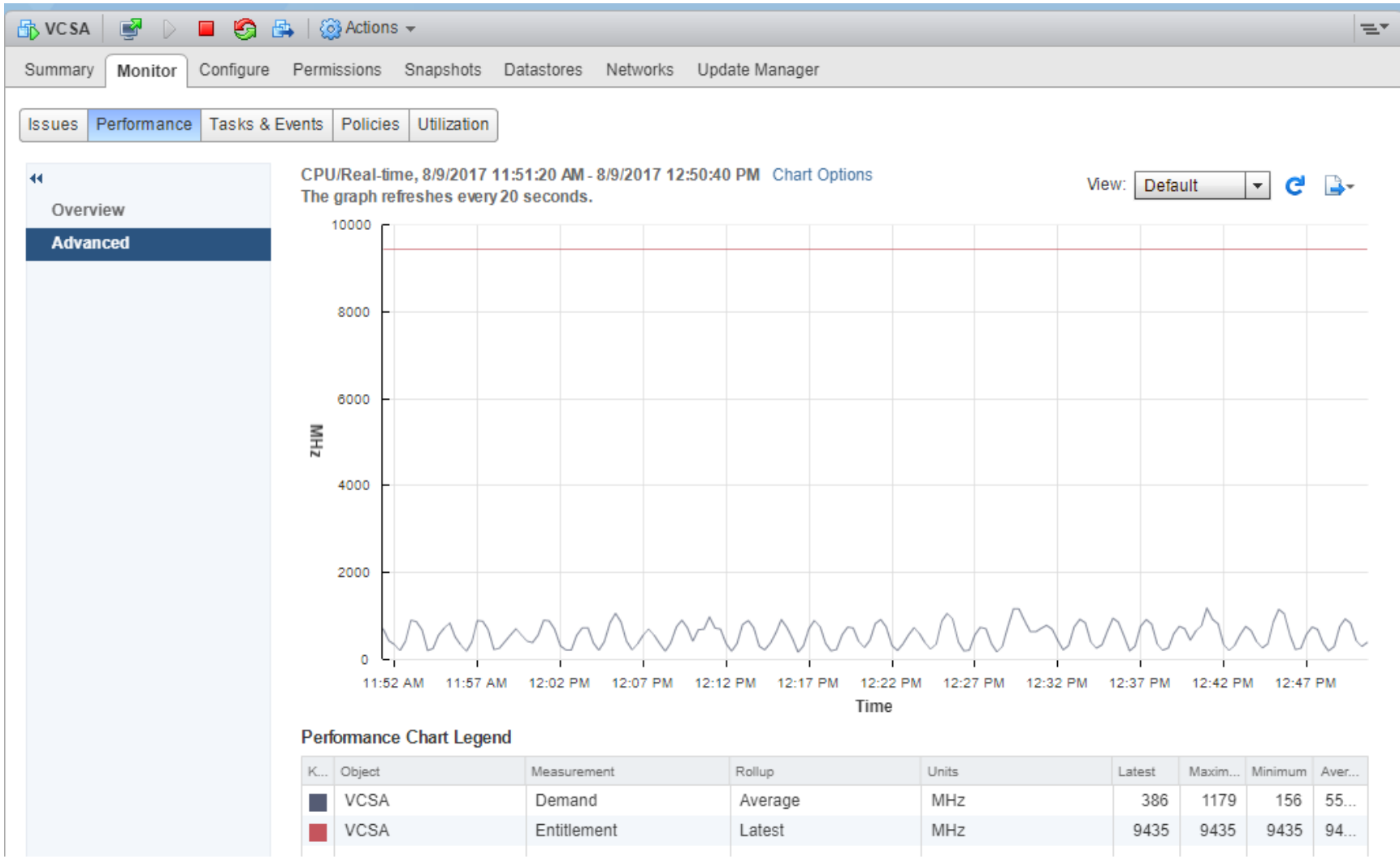
Demand Versus Entitlement

CPU Demand

- **What VM Wants**
- **What VM would Consume**
  - **If NO CPU Contention**
  - **If NO CPU Limit Set**

CPU Entitlement

- CPU Resources Available to VM's
- CPU Resources Available to Resource Pools

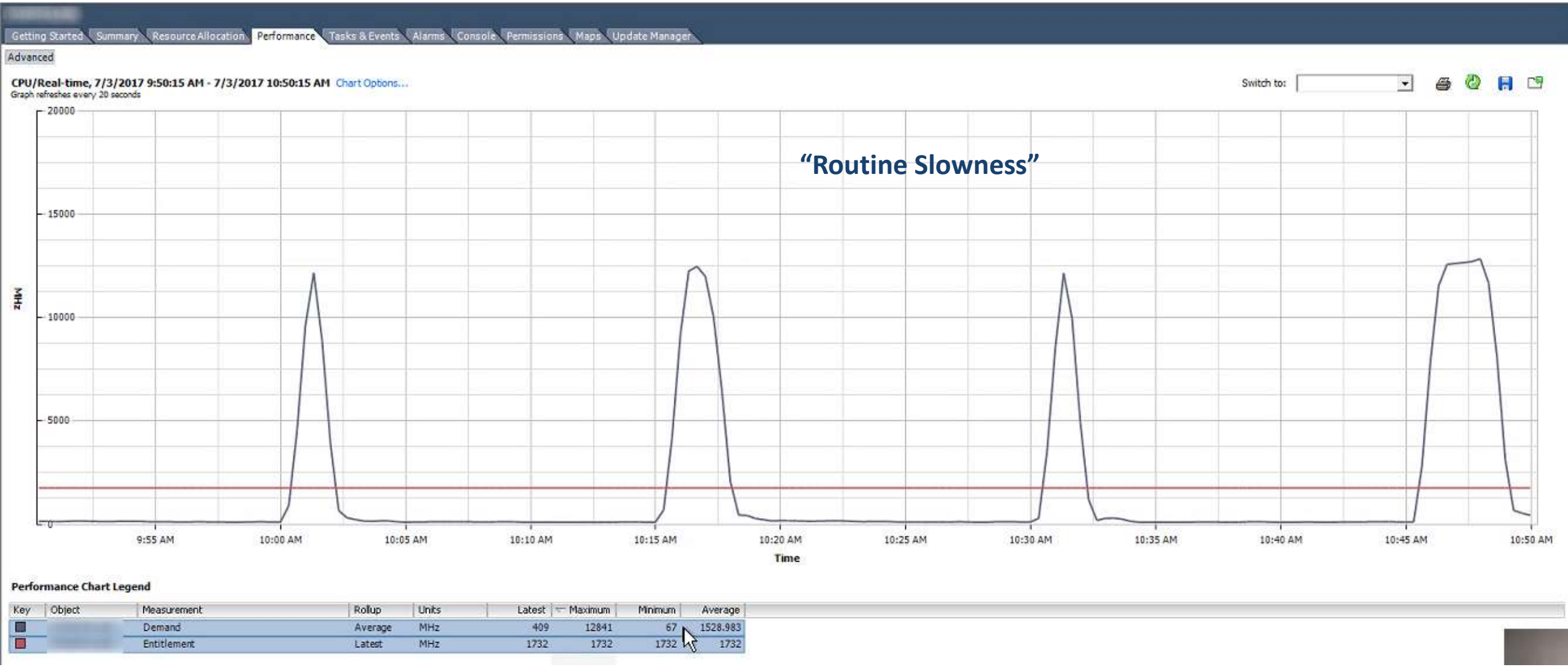


Demand Versus Entitlement

Demand Versus Entitlement

“Random Slowdowns (PITA)”

Demand Versus Entitlement



Architecting For Performance

Memory

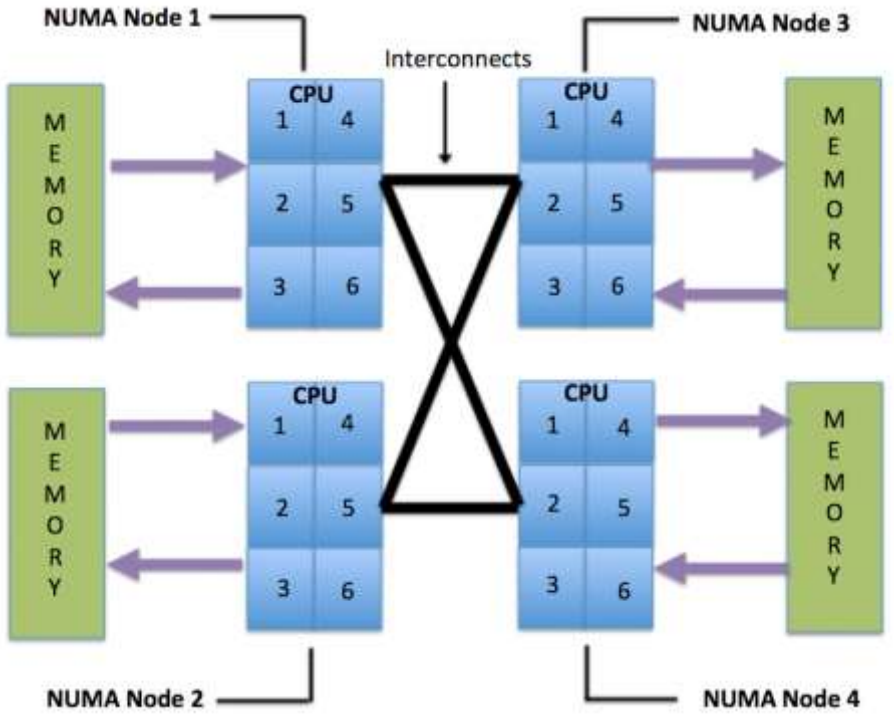
Non-Uniform Memory Access (NUMA)

NUMA, avoiding the performance hit when several processors attempt to address the same memory by **providing separate memory** for **each NUMA Node**.

NUMA Nodes Specific to Each Processor Model

Speeds up Processing (30-40%)

Project Capstone: PreferHT – Telling vSphere you'd **rather have access to processor cache and NUMA memory locality as priority**, **over** the **additional compute cycles**.



(NUMA) “All Processors Can Use All Memory”

4 Sockets, 6 cores

4 NUMA Nodes

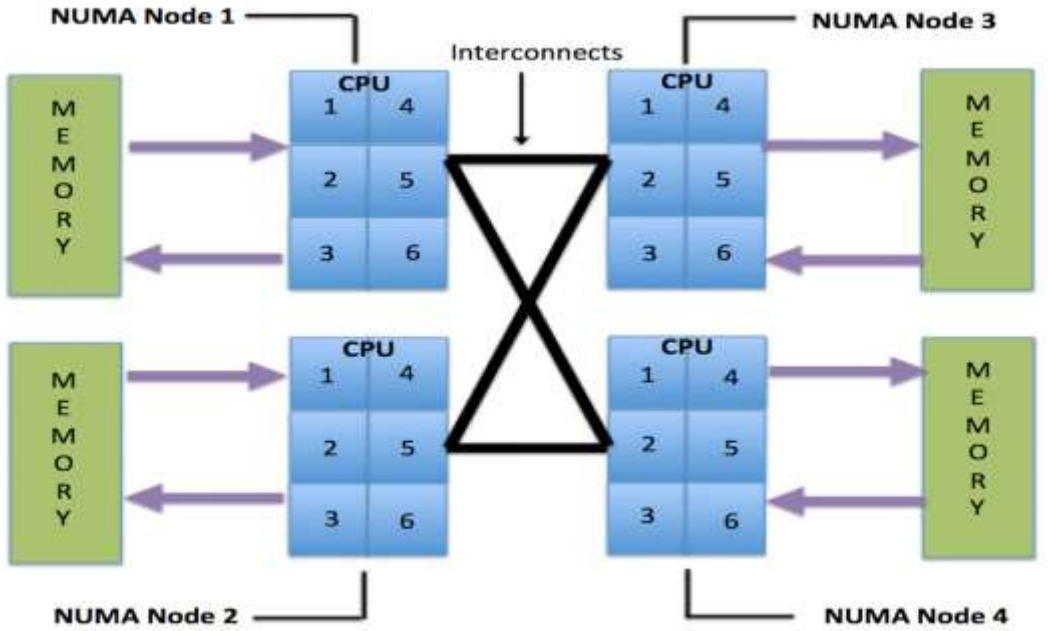
128 Gig RAM

Each NUMA Node = 32 Gig RAM

“In this example **Optimal Performance**: **Each VM <32GB*”**

*CPU Overhead Needs to be accounted for. Minimal

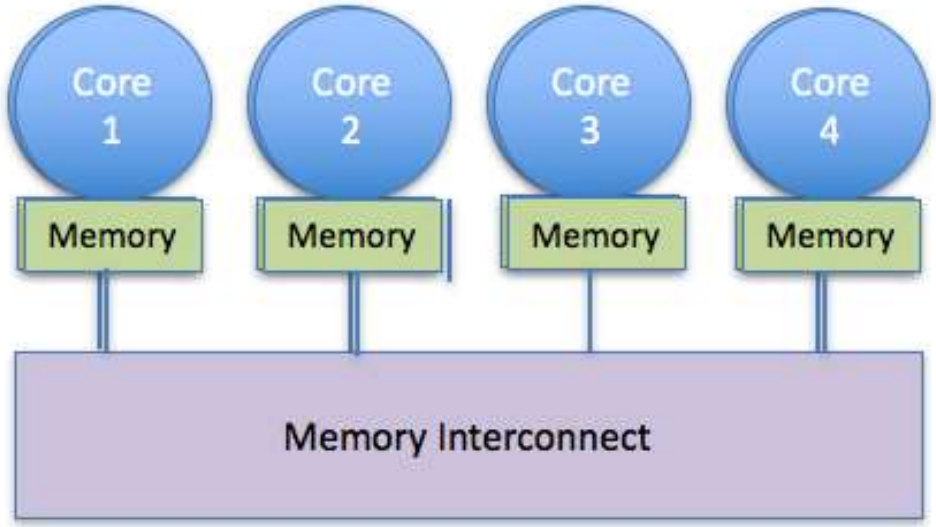
*vNUMA – Minimizes Impact when VM> 32GB happens



99

Keep **VM Footprint as small as Possible**: NUMA, **Shared Resource Pool**

vSphere 6.5 Decouples Cores per Socket From Virtual NUMA Topology



Blog on Changes in ESXi 6.5+ That Could Effect the Deployment of a Database

http://www.davidklee.net/2016/11/29/vmware-vsphere-6-5-breaks-your-sql-server-vnuma-settings

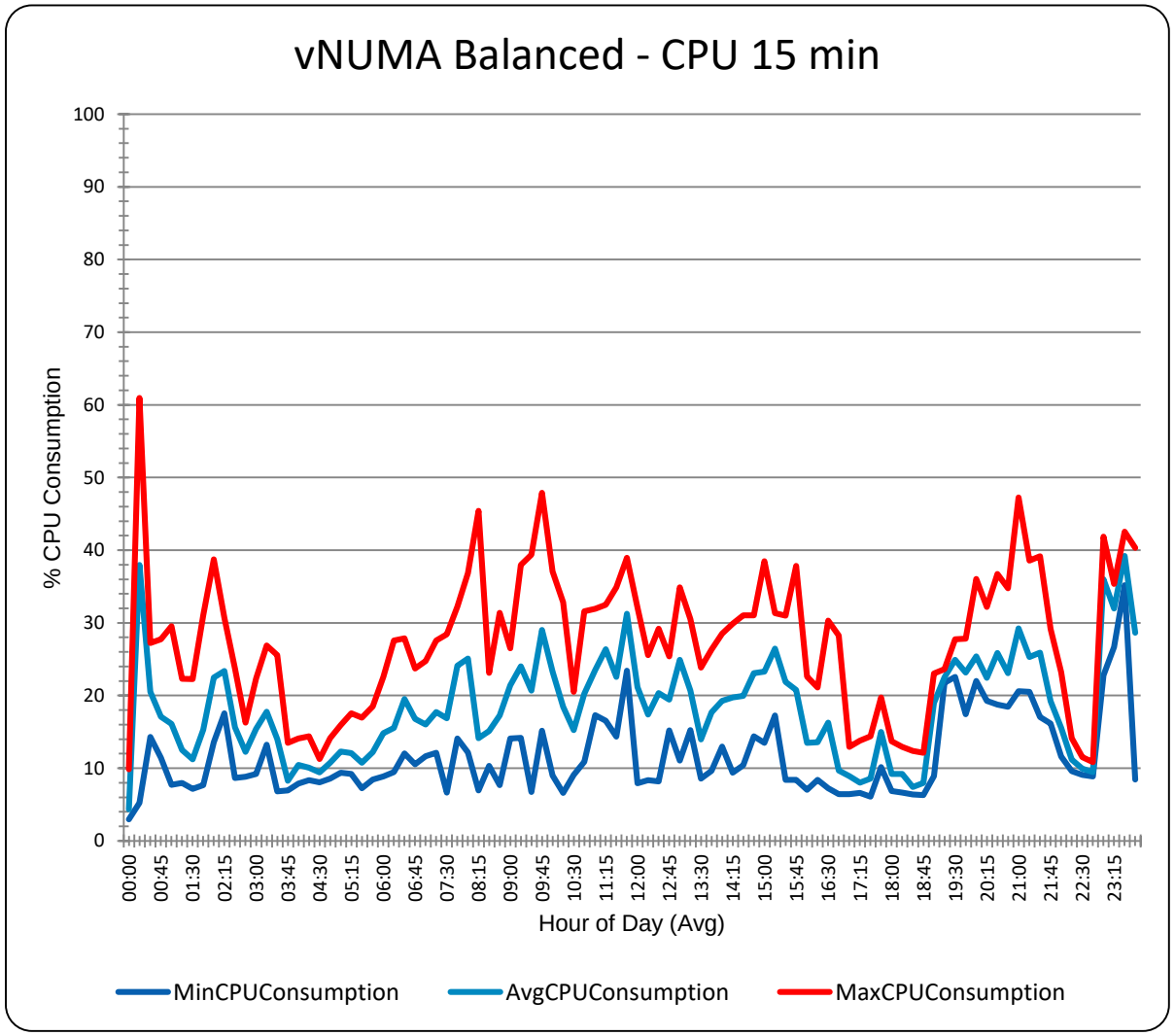
VMware Does Not *Always* Choose Best for SQL Server *(but you can correct it)*

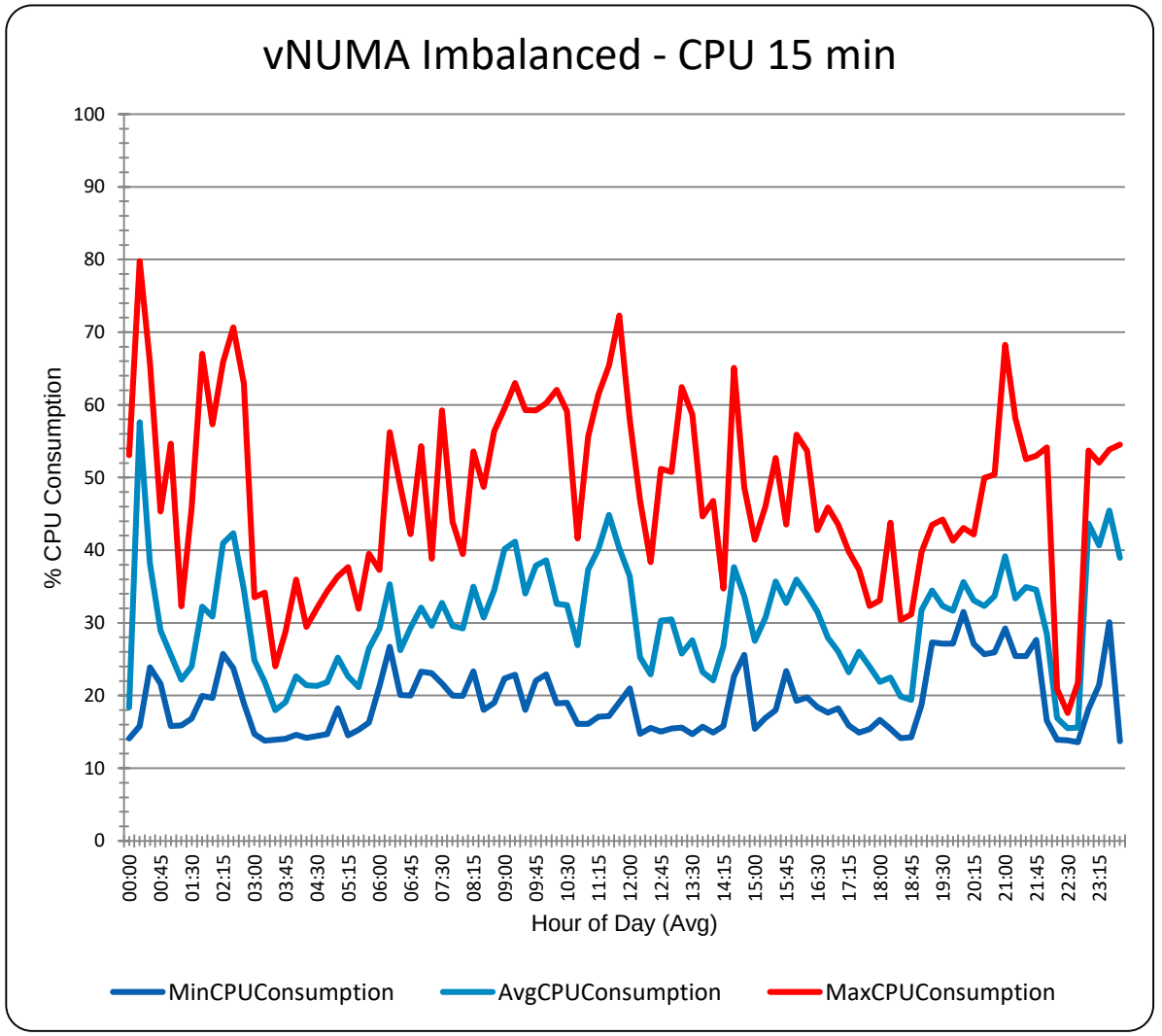
Great Blog on ESXi 6.5 Changes in vNUMA

http://frankdenneman.nl/2016/12/12/decoupling-cores-per-socket-virtual-numa-topology-vsphere-6-5/

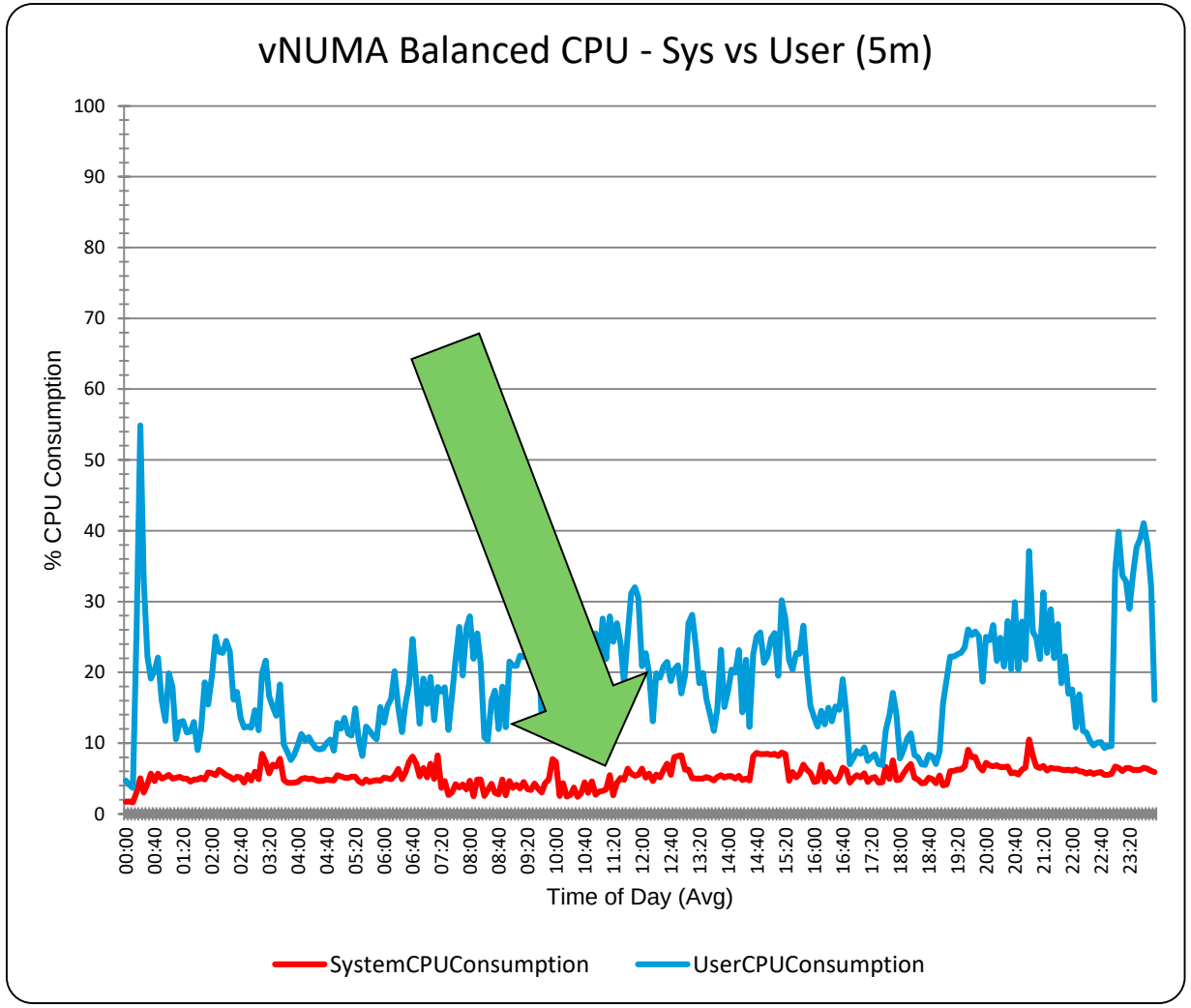
vNUMA Imbalance

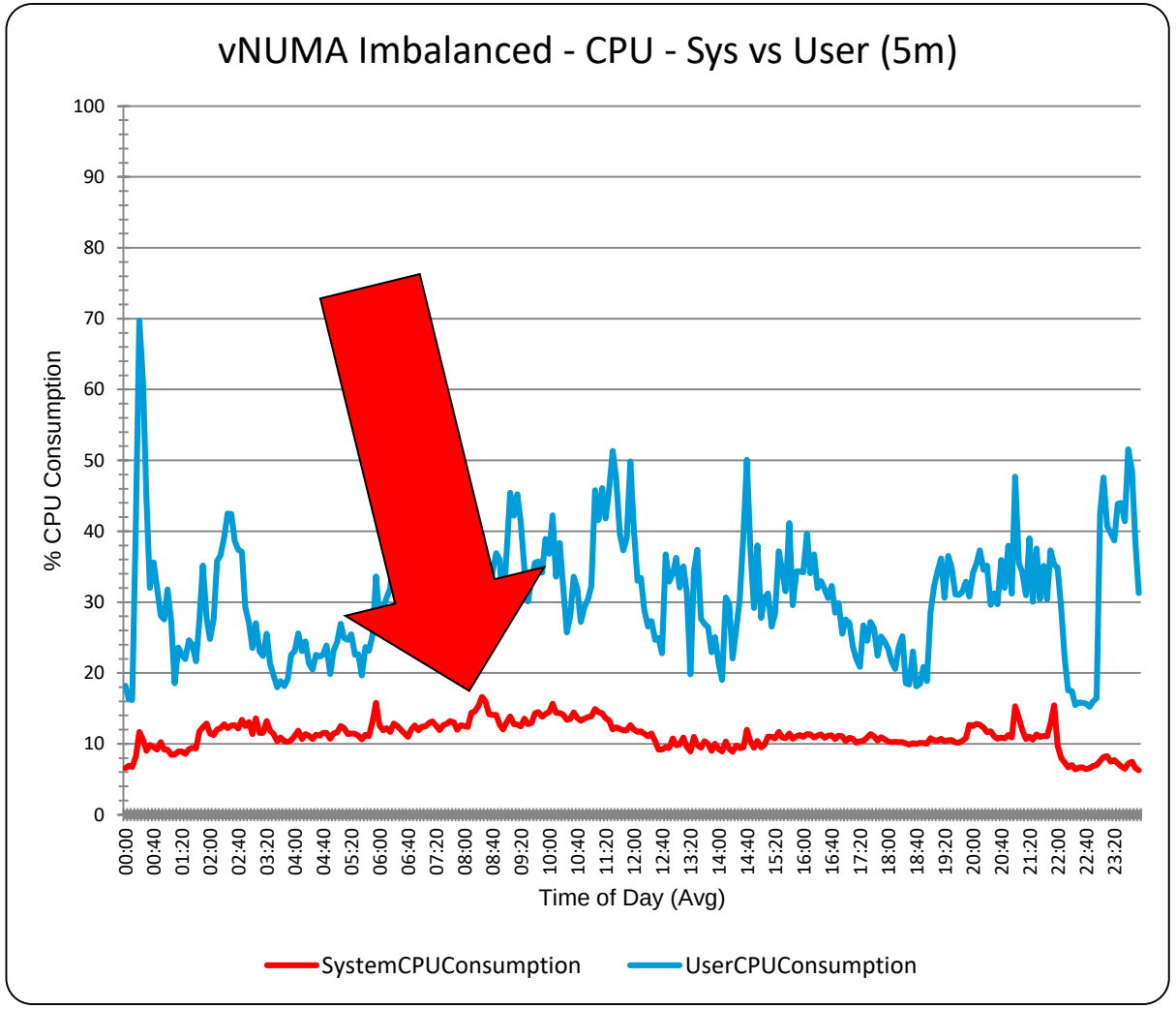
(2x6 core host, 1x8 core VM → 2x4 core + vNUMA override)





vNUMA Imbalance – OS Kernel Time





Ballooning, Memory Compression, Swapping Slow You Down

Stating the Obvious
That's a Rock

Ballooning

Kicks in – When **Physical Host experiencing memory contention**

Balloon Driver **Runs** on each **individual VM**

Communicates with guest O/S to determine what is happening with memory

Works with the server to reclaim pages that are considered least valuable by the guest OS

Ballooning
is Your First Line of Defense

Important: Right Size – VM'S (Shared Resource Environment)

"Keep in Mind Business Cycles:
– Baseline, Baseline…"

Allocating **too many resources** can actually **slow down** the **VM and hurt performance** of **other VM's**

Memory Reservations

VM is only allowed to power on if the CPU & memory reservation is available (Strict admission)

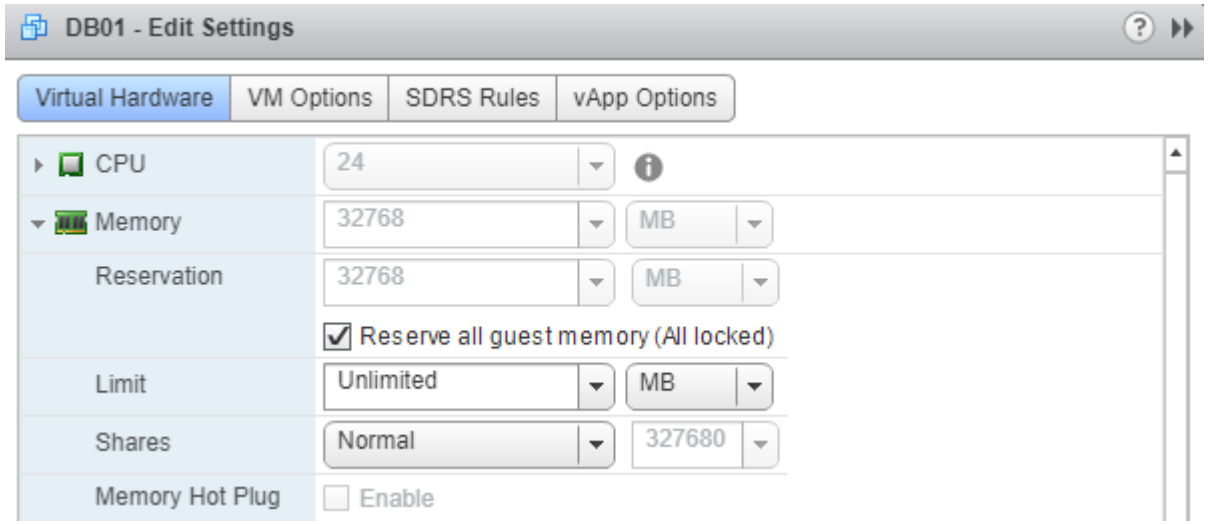
The amount **of memory can be guaranteed** even **under heavy loads**.

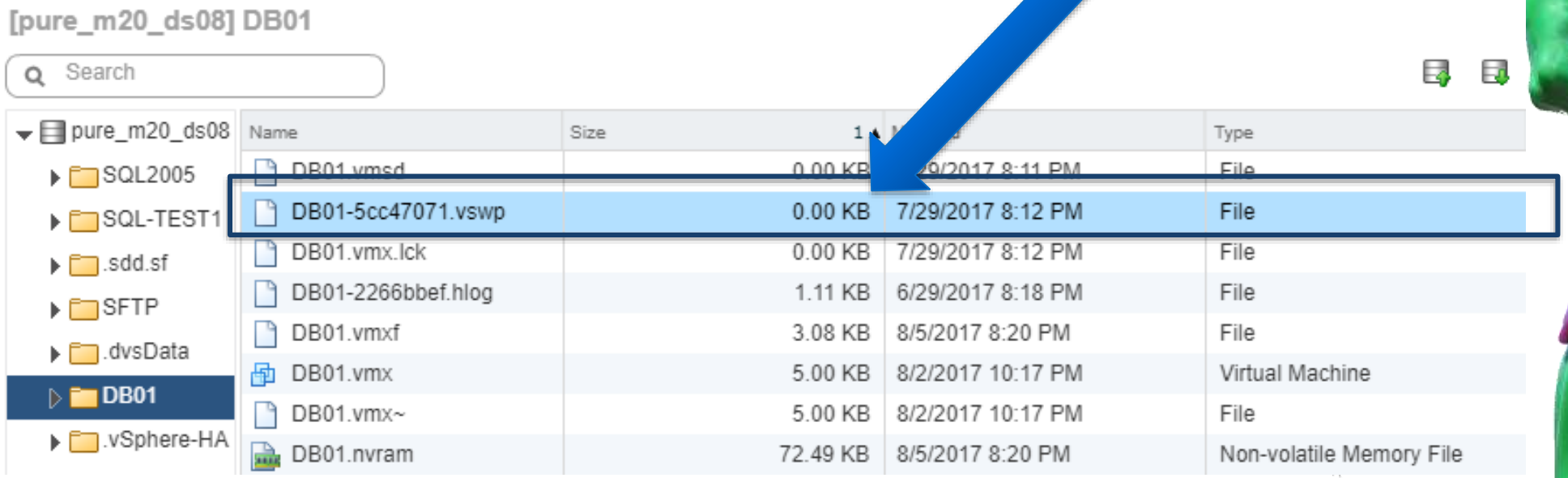
SET CPU/Not Guaranteed

Control – Settings Can Override this behavior

Reservation keep other VM's from taking my resources away

Reservations and the vswp



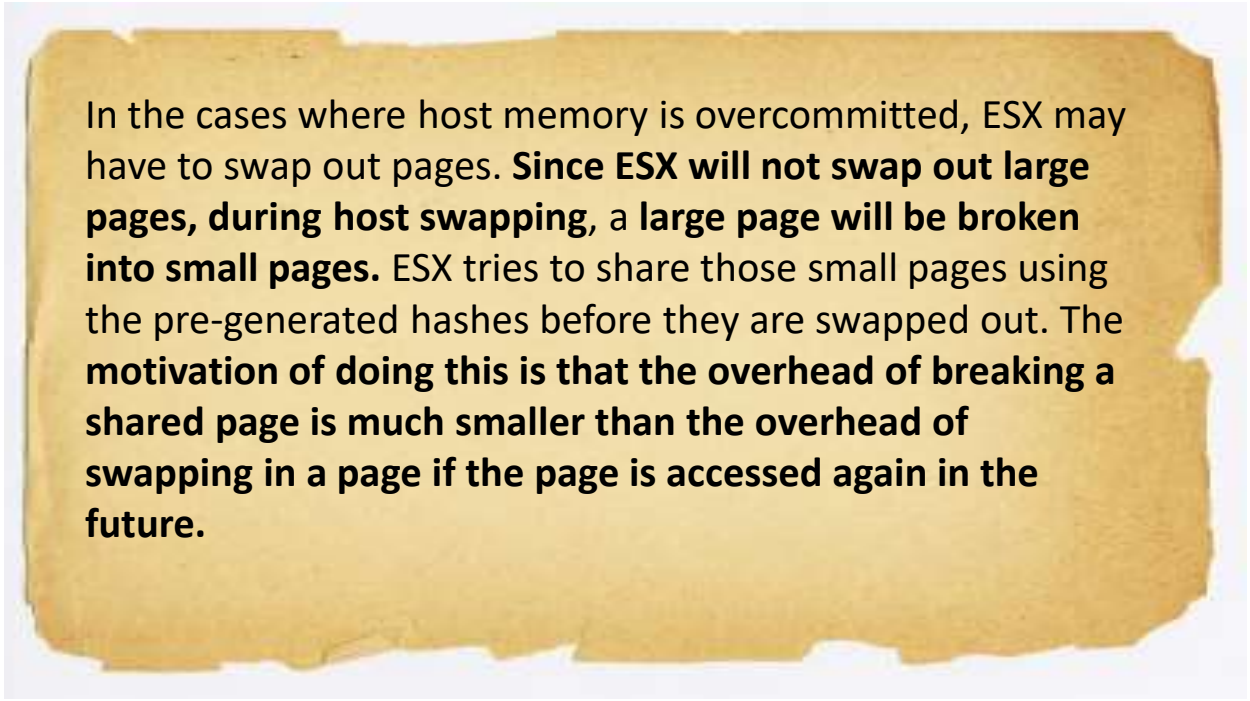


"0.00 KB VSWP File"

Memory Reservations Guarantee's no Swapping

ESX 6.5+ Large Pages/Huge Pages by Default

"For Years have Talked Benefits of Large Pages – **Now the Default in vSphere 6.5"**

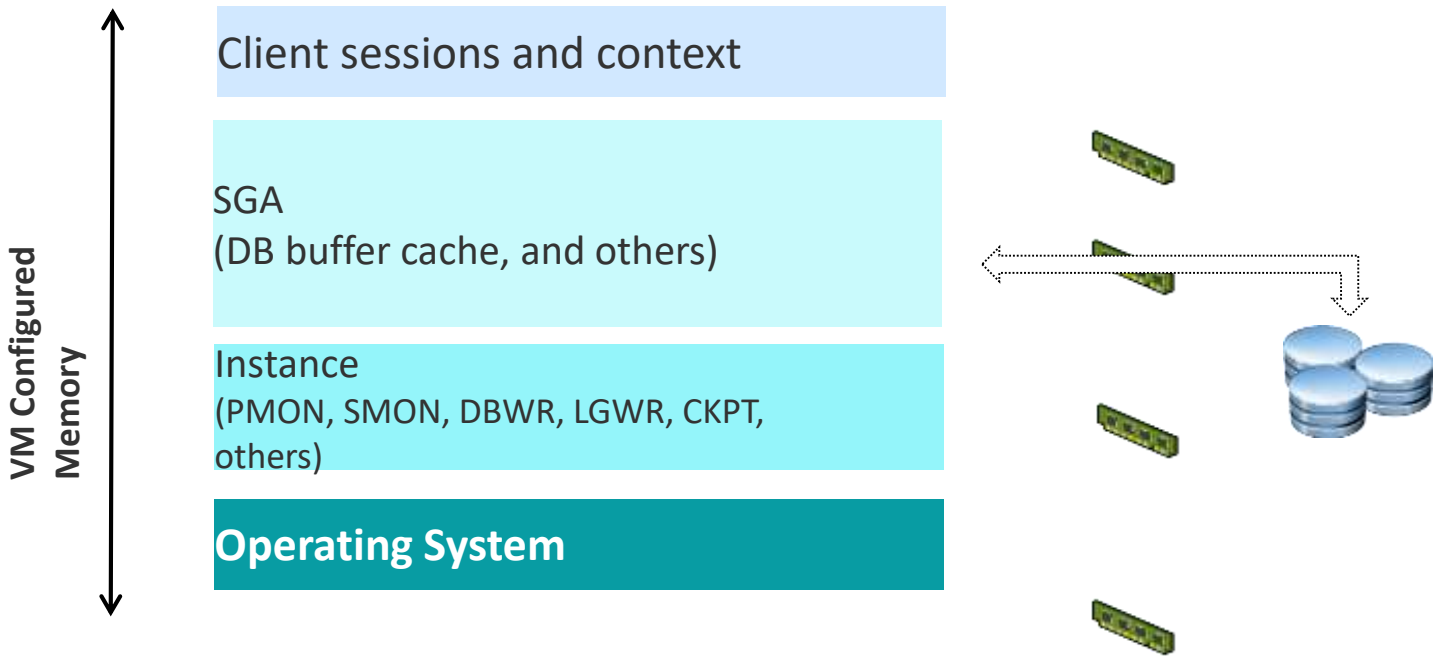


"Large/Huge PAGES Do Not Normally SWAP"

"HUGE PAGES Do Not Normally SWAP"

http://kb.vmware.com/kb/1021095

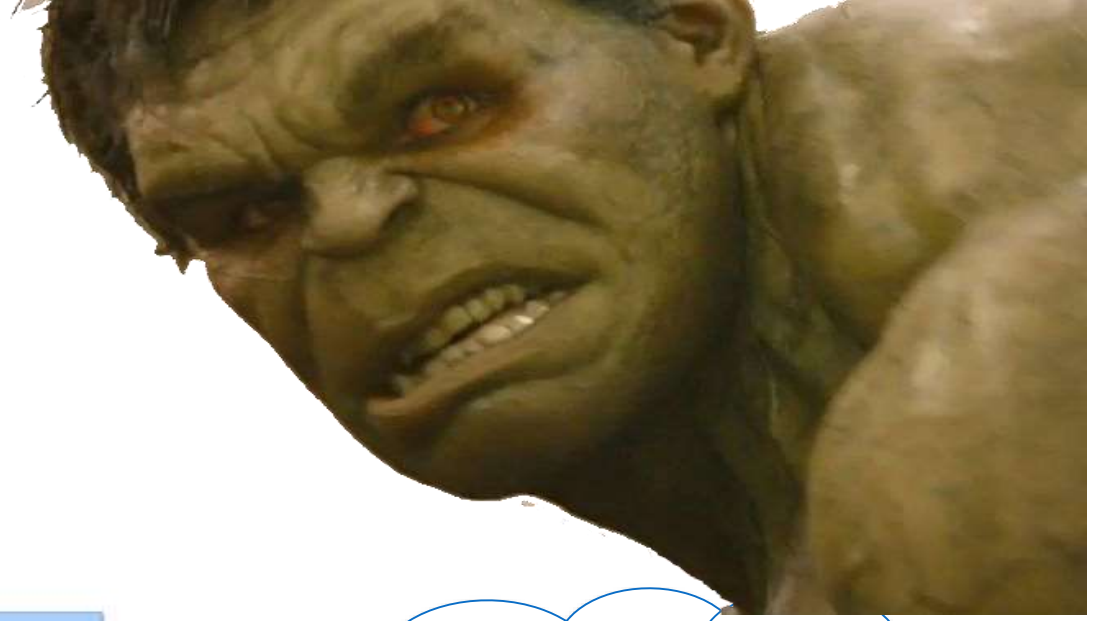
Oracle Approximate Memory Architecture

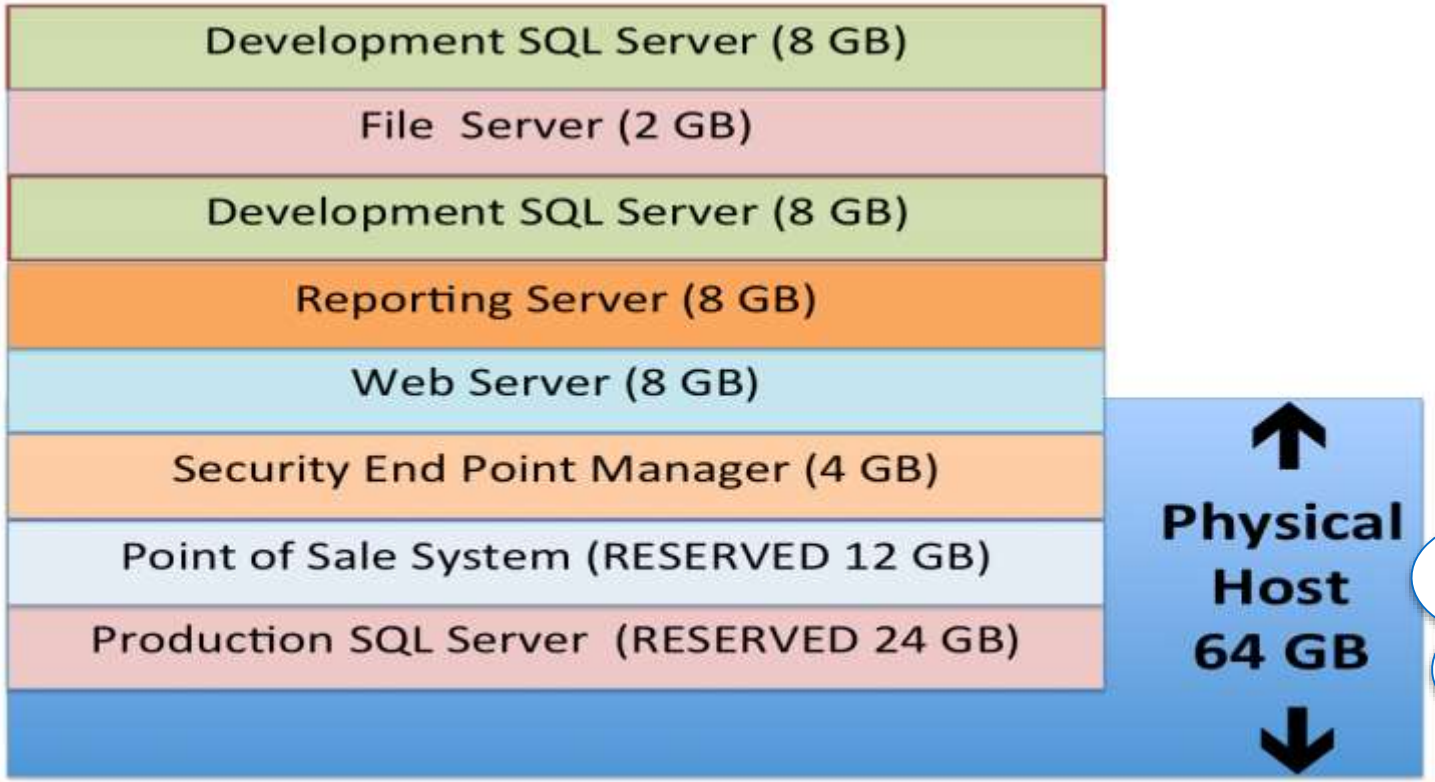


Set the memory reservation to SGA size plus OS.

(Reservation & configured memory might be the same.)

Don’t Over Subscribe Memory*





40 GB Out Of 64GB is Reserved, Total Virtual Memory Demand = 74 GB

Friends don’t let this happen to their Monster VM’s

“Physical Resources are Hard Limits”

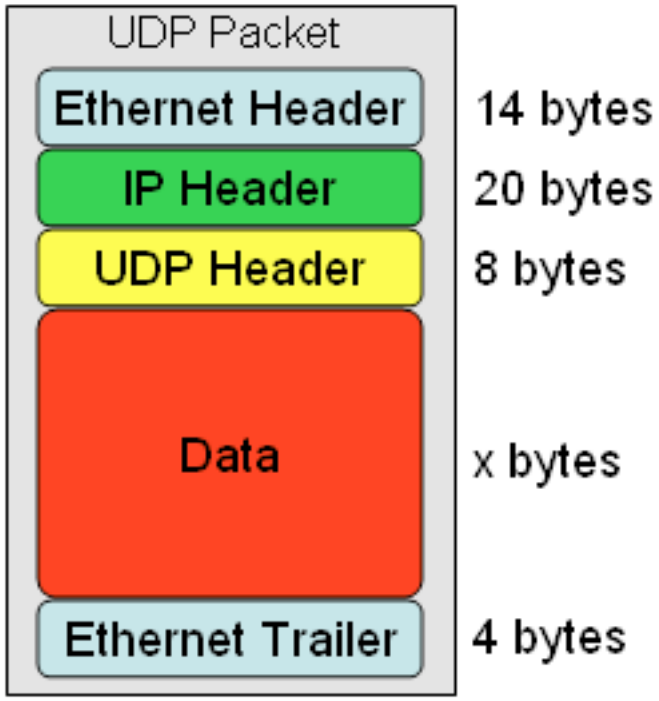
*Until You know Exactly How Memory is Utilized

Architecting For Performance

Network

Jumbo Frames

Jumbo frames are Ethernet Frames with more than **1500 bytes** of payload. Conventionally, **jumbo frames** can carry up to **9000 bytes** of payload



Jumbo Frames

The original 1500-byte payload size for Ethernet frames was used because of the **high error rates** and **low speed** of communications.

"Why The Picture Of A Typewriter Here?"

Network Putting it all Together

Use the **VMXNET3** Driver (Reduces CPU Utilization, More Throughput)

Separate Traffic for vMotion, SQL Server AGS, FCI Heartbeat

- **Easier to monitor**
- NIOC / QOS as needed

Set Database Packet Size 8192

- Only if Network can support Larger size end to end

Isolate Database Workloads from Chatty Network Traffic

Up to 4 NICS per Host (Redundancy & Performance)

Security Isolation using Non-Routable vLANS

Michael Corey
Michael.Corey@LicenseFortress.com.com

@Michael_Corey

Dean Bolton
Dean.Bolton@vlss-llc.com

LICENSEFORTRESS®
ORACLE COMPLIANCE MANAGEMENT + AUDIT PROTECTION WITH A GUARANTEE

VLSS