

RED HAT :: CHICAGO :: 2009

SUMMIT

FOLLOW US:

TWITTER.COM/REDHATSUMMIT

TWEET ABOUT US:

ADD #SUMMIT AND/OR #JBOSSWORLD TO THE END
OF YOUR EVENT-RELATED TWEET

presented by



RED HAT :: CHICAGO :: 2009

SUMMIT

Red Hat Virtualization: Breaking Scalability and Performance Barriers

John Shakshober
Senior Consulting Engineer, Red Hat

Vijay Trehan
Manager, Solutions Architecture, Red Hat

Sanjay Rao
Sr Software Engineer

September, 2009

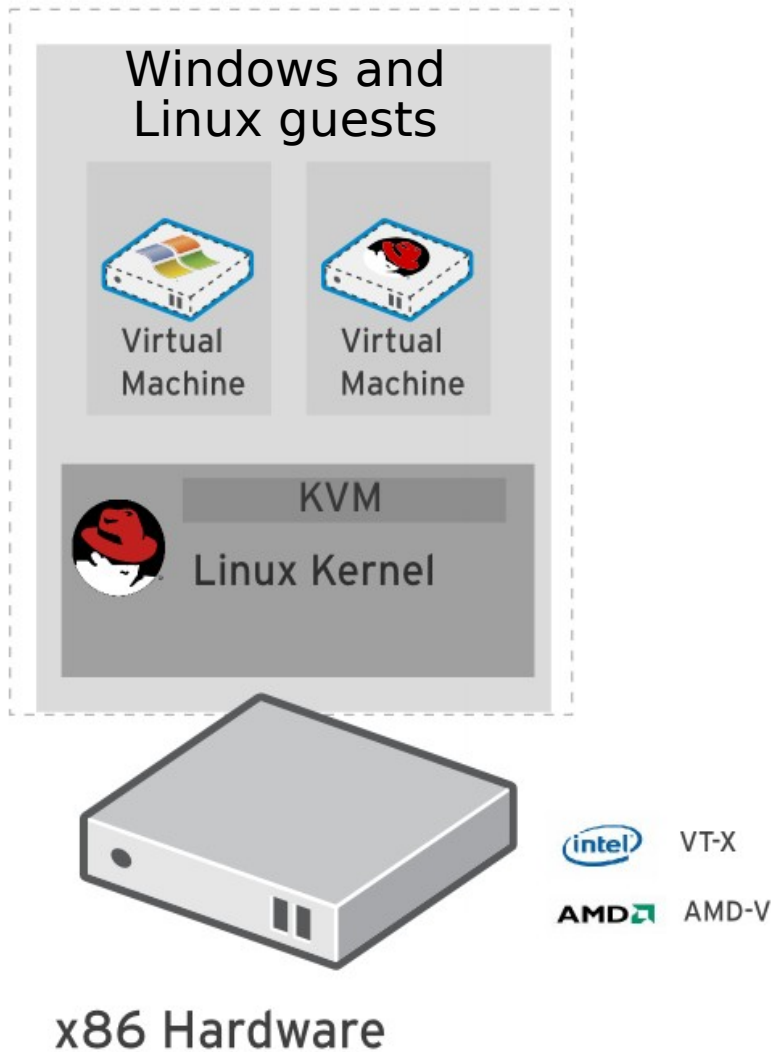
presented by



Agenda

- **Tuning Red Hat Virtualization Environment:**
 - CPU, VM, Disk and Network Performance
 - Using “normal” Linux tuning w/ KVM guest.
- **Performance & Scaling of Production Application stacks in a Red Hat Virtualization Environment:**
 - Oracle 10g – Linux guests
 - SAP – Linux guests
 - LAMP Stack – MySQL, Apache, Linux guests
 - Java Applications – Windows guests
 - Microsoft Exchange – Windows guests

RED HAT ENTERPRISE VIRTUALIZATION



- **Scalability**

- Host: 96 cores, 1 TB RAM
- Guest: 16vCPU, 64 GB RAM

- **Industry Standards**

- Trusted RHEL kernel + KVM
- High performance VirtIO drivers
- Libvirt management interface

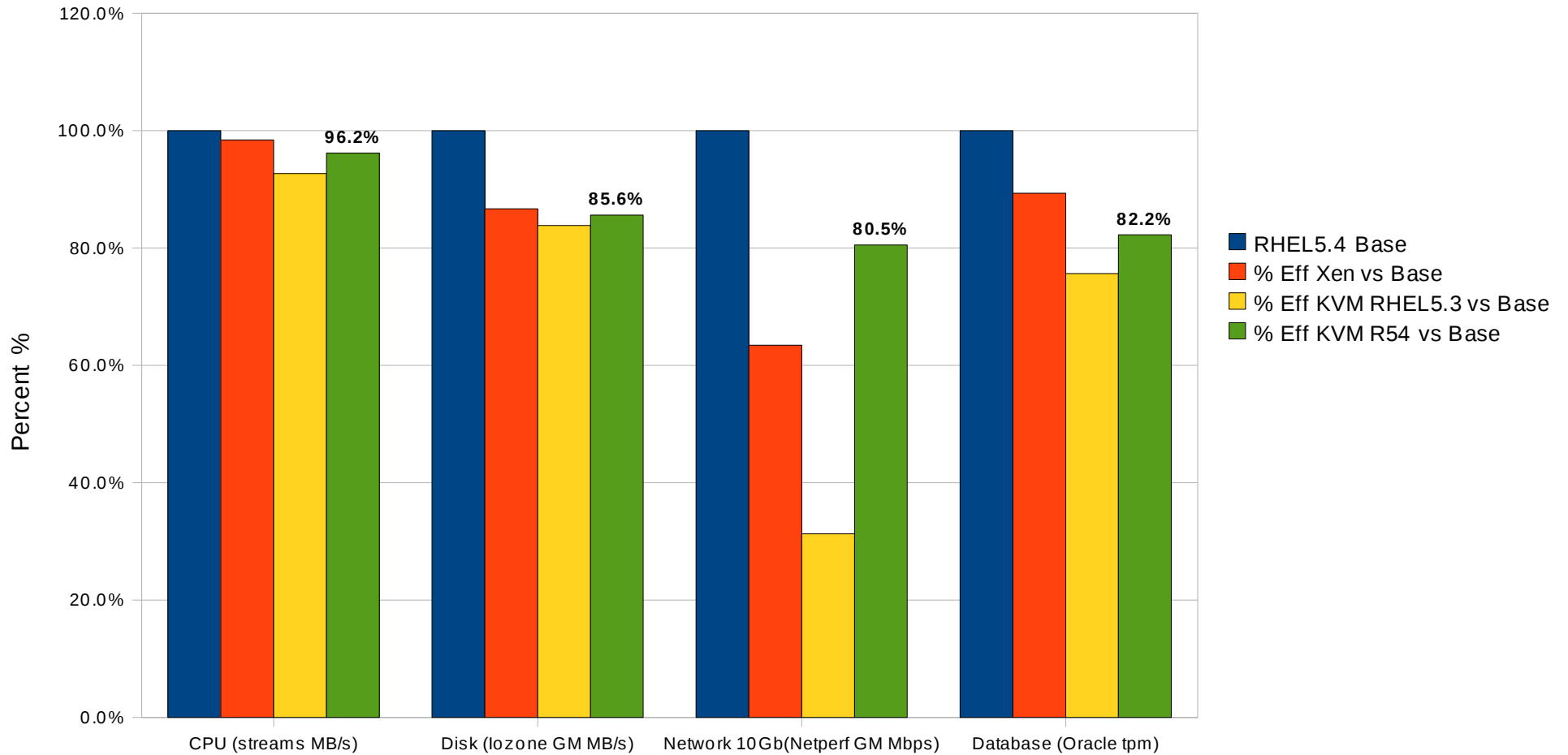
- **Advanced Features**

- KSM Kernel Shared Memory
- SELinux for high security and isolation
- Live migration
- Snapshots
- Thin provisioning

KVM Performance – CPU, Disk, Network, DB

RHEL5.4, KVM Performance Summary

Intel 16-cpu, 16GB, 2FC, 10Gbit Iguest 8-cpu Guest



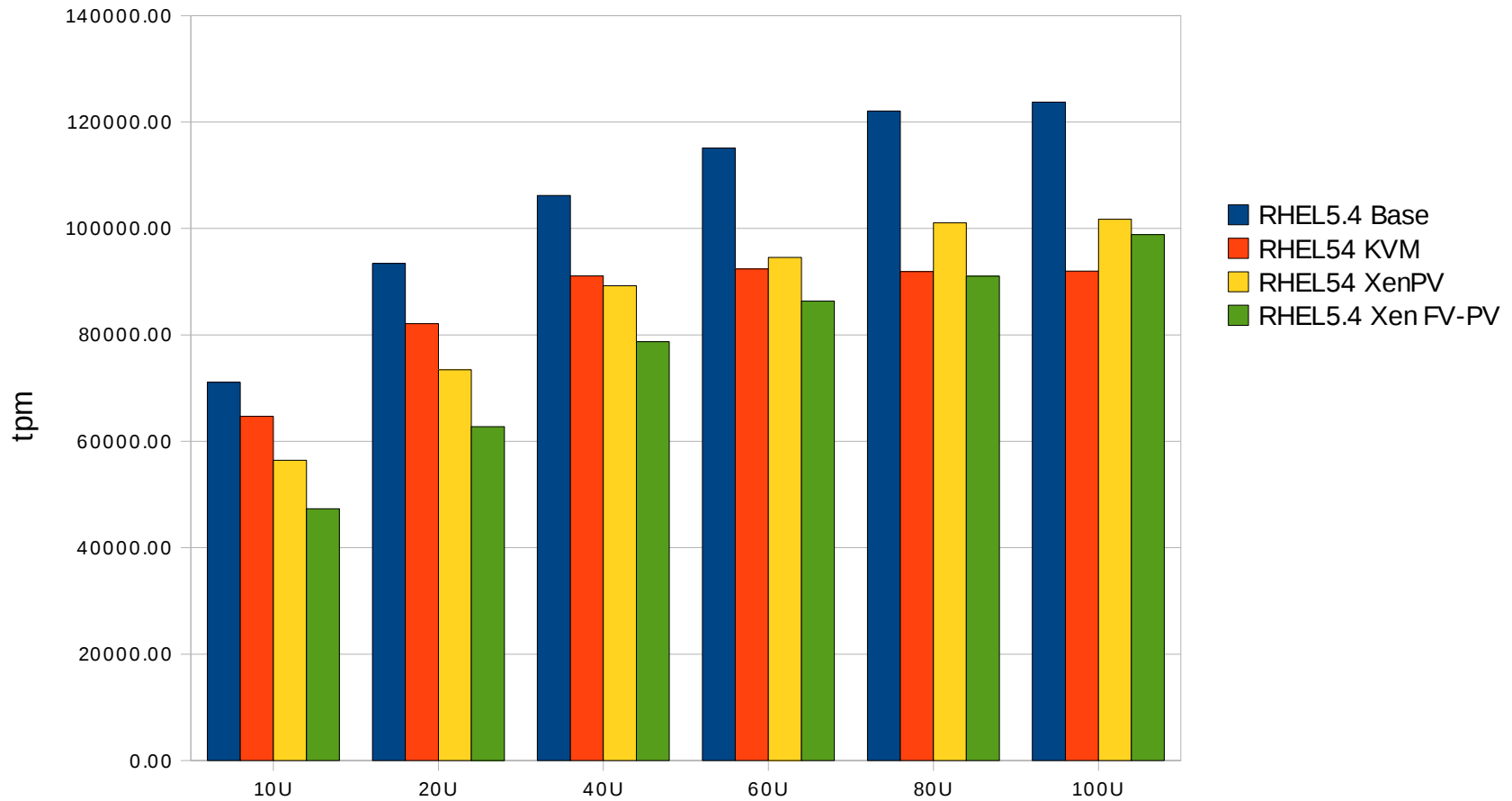
Tuning RHEL5.4 KVM - normal Linux tuning

- CPU
 - CPU features, extended/nested page tables EPT/NPT
 - CPU taskset affinity(vcpupin), alter priority (nice), policy (chrt)
- DISK I/O – host uses mpio used in host, iSCSI admin, NFS.
 - KVM cache-off default, VirtIO drivers, elevator=deadline.
- VM
 - NUMA – control with numactl
 - Huge Pages – same linux controls hugetlbfs
 - KSM – kernel shared memory
- Network
 - VirtIO, IRQ aff , General Receive Offload (GRO), VT-d
- POWER - default cpuspeed, prevent drift
- Tools - Oprofile, Systemtap, valgrind

RHEL5.4 KVM Performance vs Xen

RHEL5.4 KVM Oracle OLTP Performance

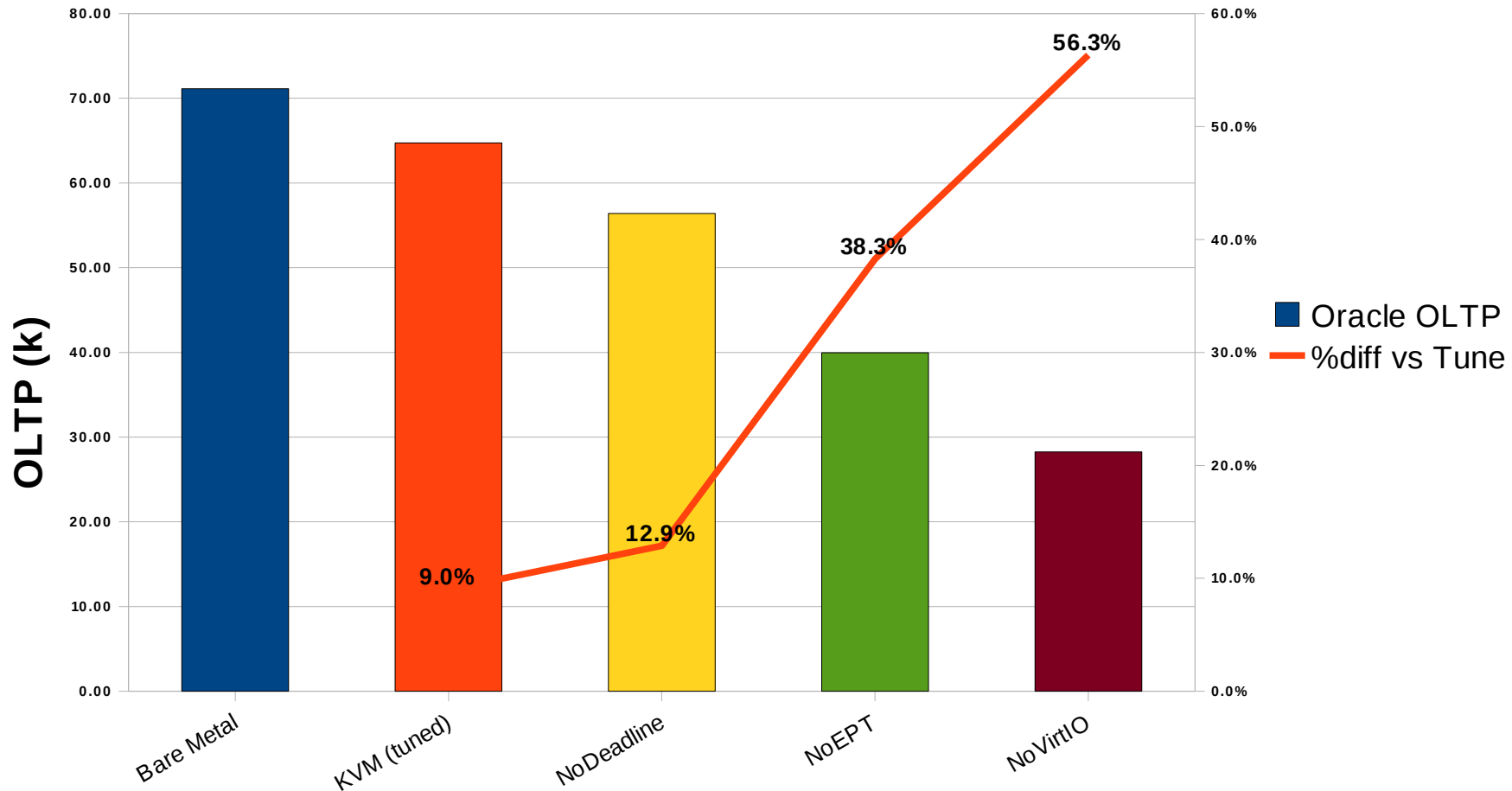
Intel Nehalem 8-cpu, 24 GB mem, 2FC



KVM Performance – CPU/Disk Effects

RHEL5.4 KVM Oracle OLTP Comparison

Intel Nehalem 8cpu, 24GB mem, 2 FC

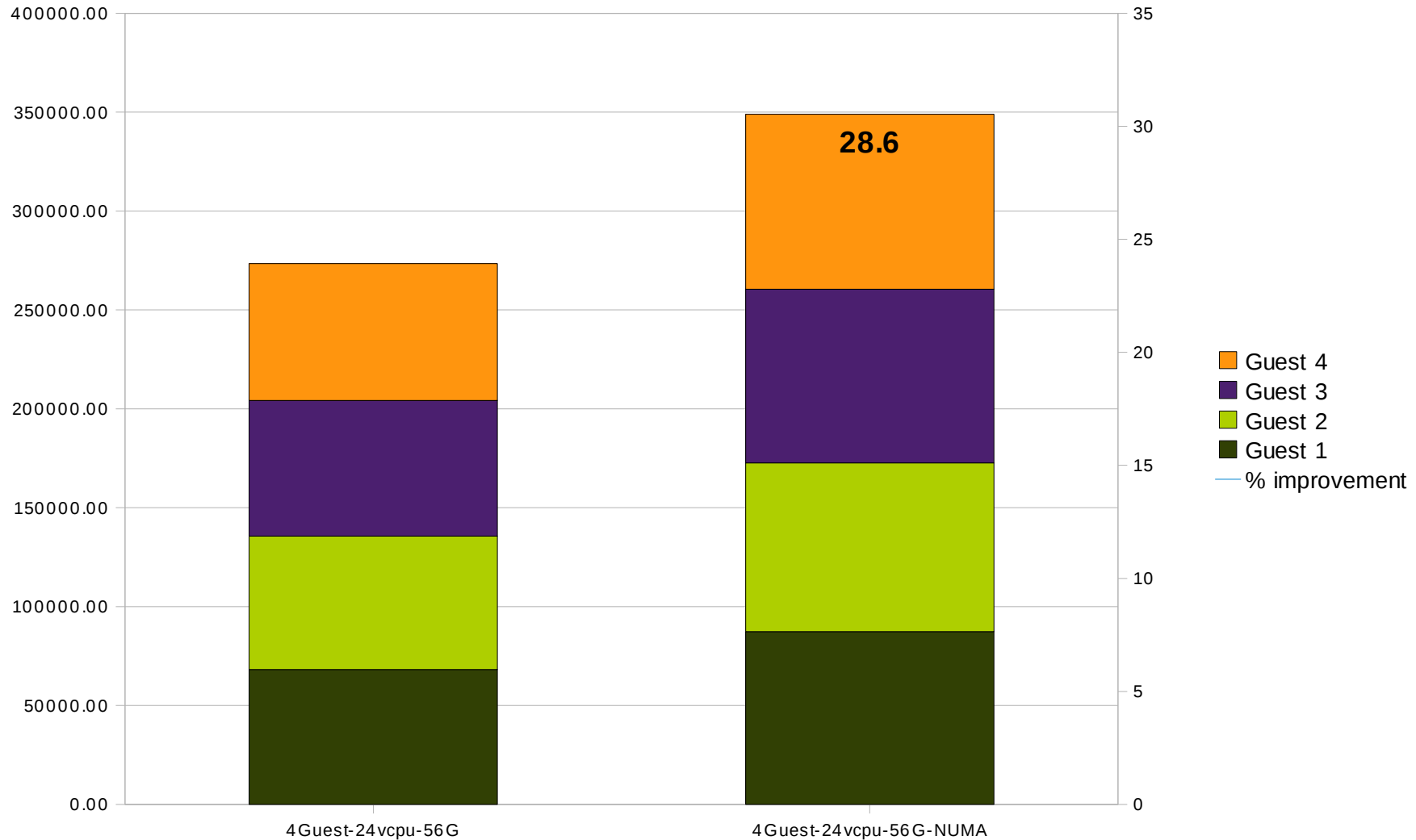


Process Control w/ KVM - taskset/numactl

- Taskset / numactl w/ qemu
 - `taskset -c 0,8 /usr/libexec/qemu-kvm -smp 2 -m 4192`
 - `numactl -m 0 --physcpubind=0,8 /usr/libexec/qemu-kvm`
- creates a 2-vCPU guest (-smp 2)
- binds to two threads in a single core
(--physcpubind=1,9)
- uses 4 GB of memory (-m 4192) from NUMA node 0 (-m 0)

KVM Performance – AMD Istanbul - 24 cpu

Effect of Numa on multiple Oracle Databases

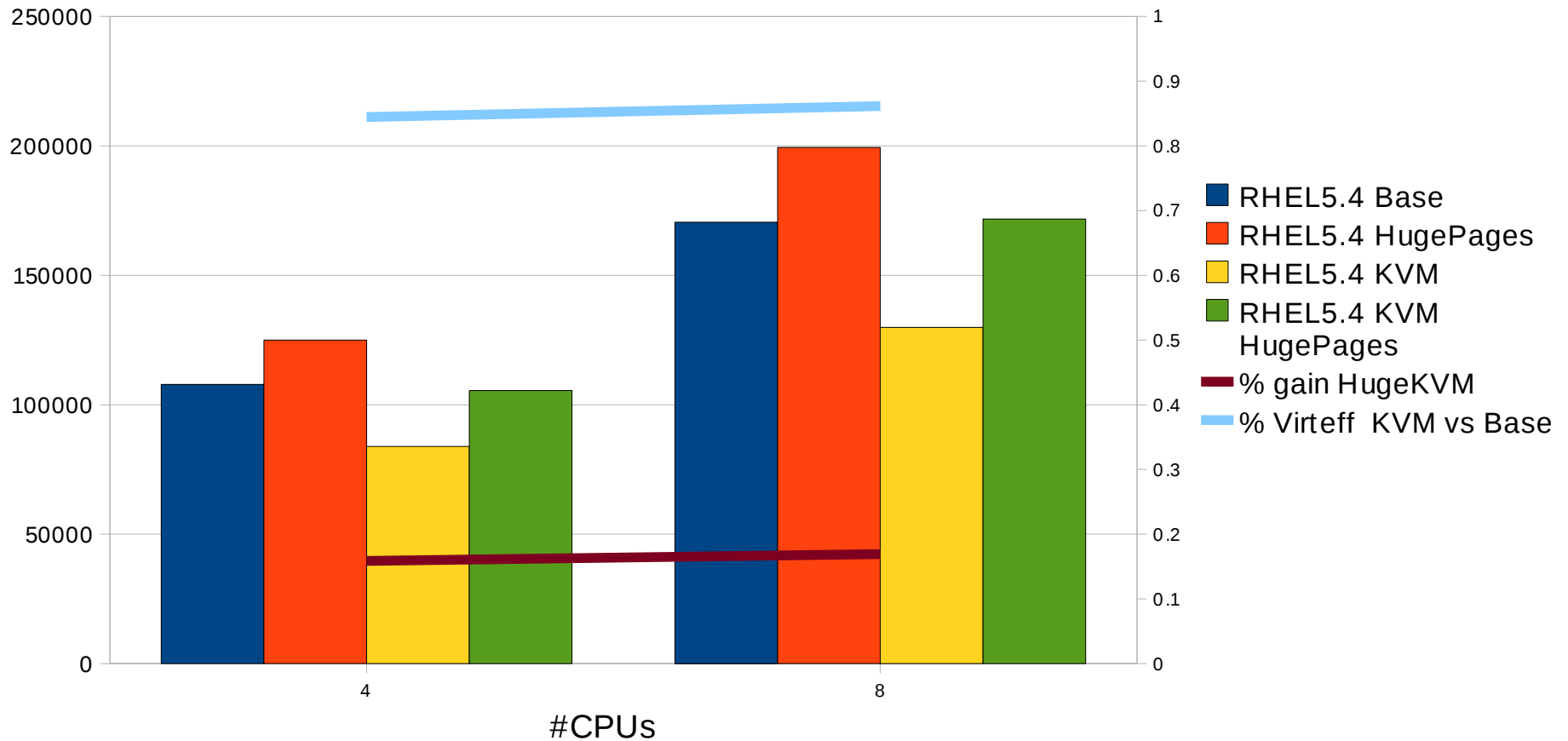


VM Tuning

- HugePages – hugetlbfs
 - For 20 GB of use 10k @ 2 MB/page
 - `echo 10000 > /proc/sys/vm/nr_hugepages`
 - Or enter in `/etc/sysctl.conf` `vm.nr_hugepages=10000`
 - `mount -t hugetlbfs hugetlbfs /mnt/libhugetlbfs`
- Pass to `qemu-kvm`
 - `/usr/libexec/qemu-kvm --mem-path /mnt/libhugetlbfs`
- Add to application – `openjdk` Java;
 - `java -Xms3900m -Xmx3900m -Xmn3300m -XX:+UseLargePages spec.jbb.JBBmain -propfile SPECjbb.props`

KVM Performance – Linux Intel EPT Effect of HugePages on Java op/s

RHEL5.4 KVM Java Performance
Intel Nehalem 2.4 GHz, 24 GB mem

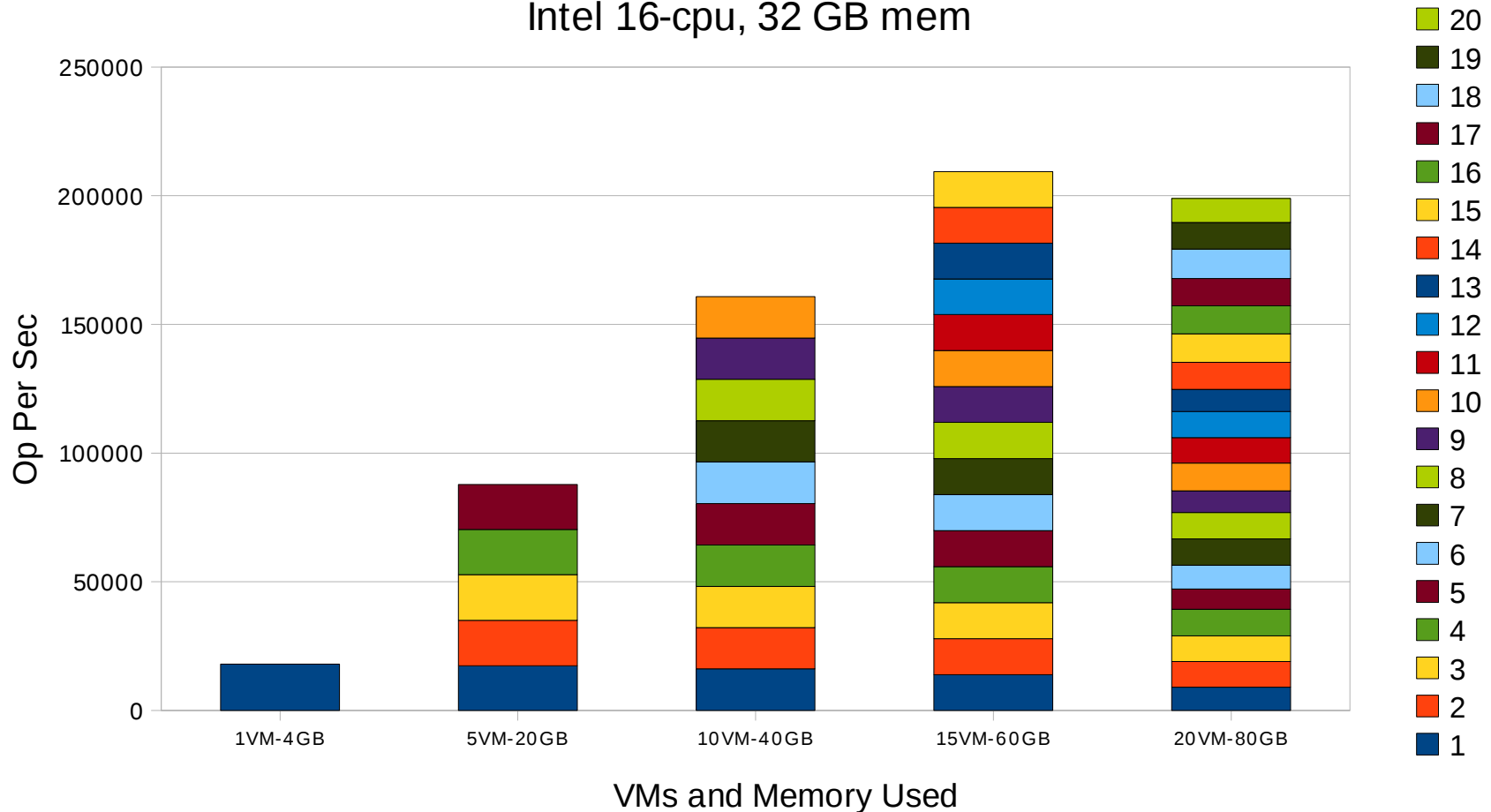


KVM Performance – Intel Windows 2003 guest

Effect of KSM on Java op/s – 300% over-commit

RHEL 5.4 KVM KSM w/ Win2003

Intel 16-cpu, 32 GB mem



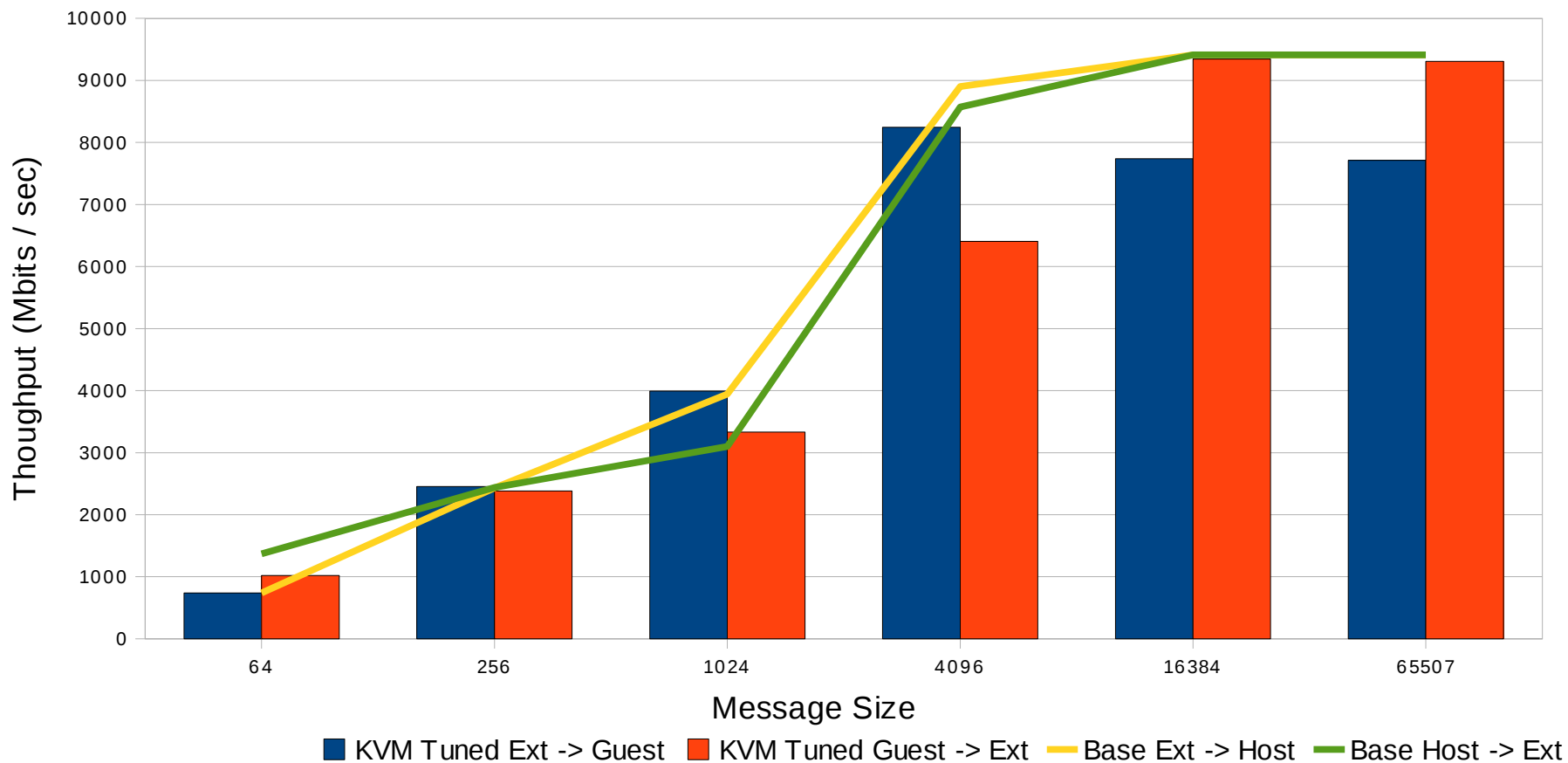
Network KVM Performance

- RHEL5.4 added General Receive Offload (GRO)
 - GRO hits 9 Gbps base and 8Gbps in KVM guest
- Tuning, IRQ pinning to same CPU/Socket w/ app
 - Netperf – various packet sizes -Gbps
 - AMQP Messaging max Msg/sec > 1Million Msg/sec
- PCI passthrough VT-d
 - Intel 10Gbit on Nehalem
 - AMQP Latency < 200 usec in KVM guest

Network Workloads – KVM IRQ Pinning

Intel Nehalem – Intel 10Gbit non-VT-d

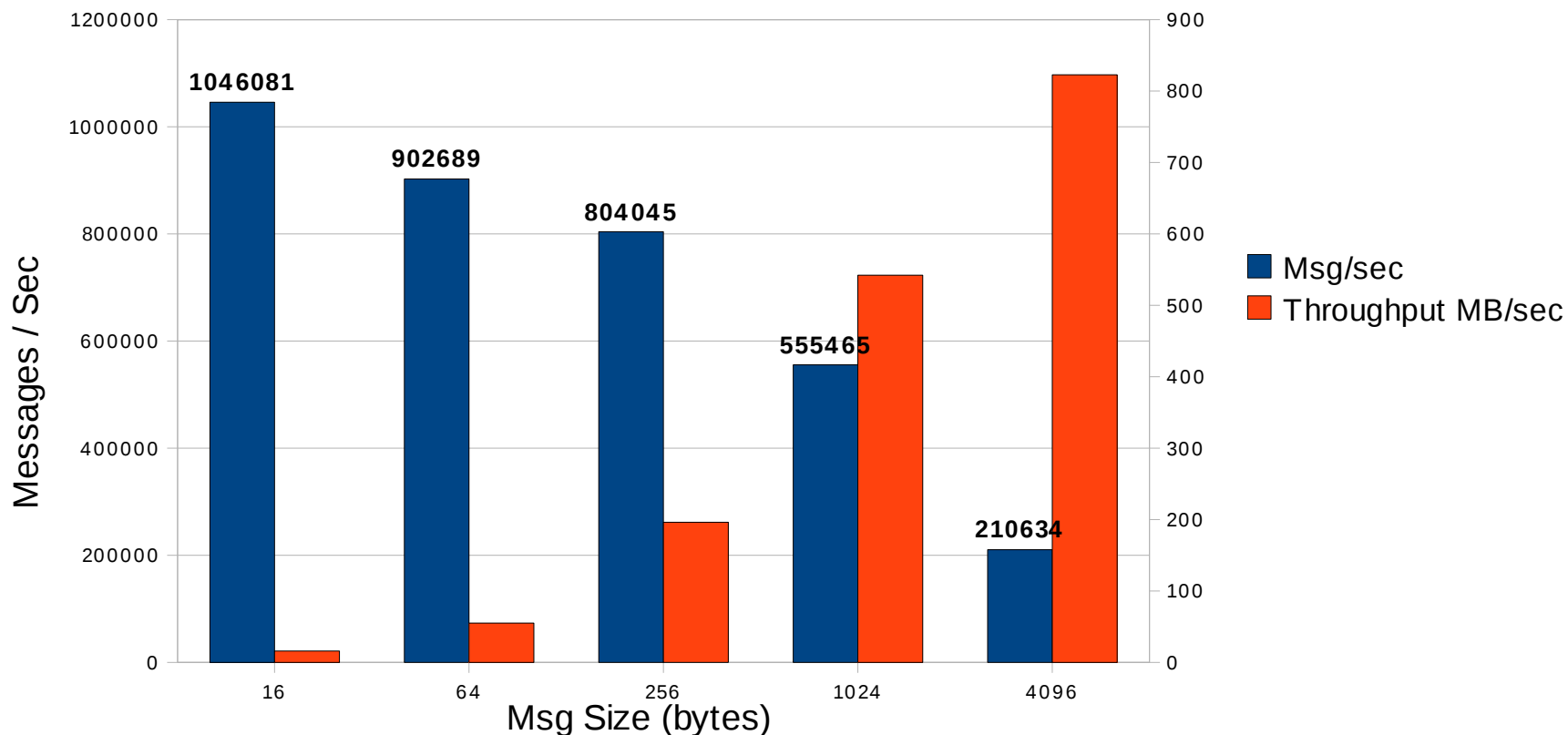
TCP Throughput KVM vs Bare Metal
single stream netperf, pinned



RHEL5.4 KVM on Intel Nehalem 2guest 1st > 1M msg/sec, 2 - Intel 10Gbit non-VT-d

RHEL 5.4 KVM AMQP 2-Guest

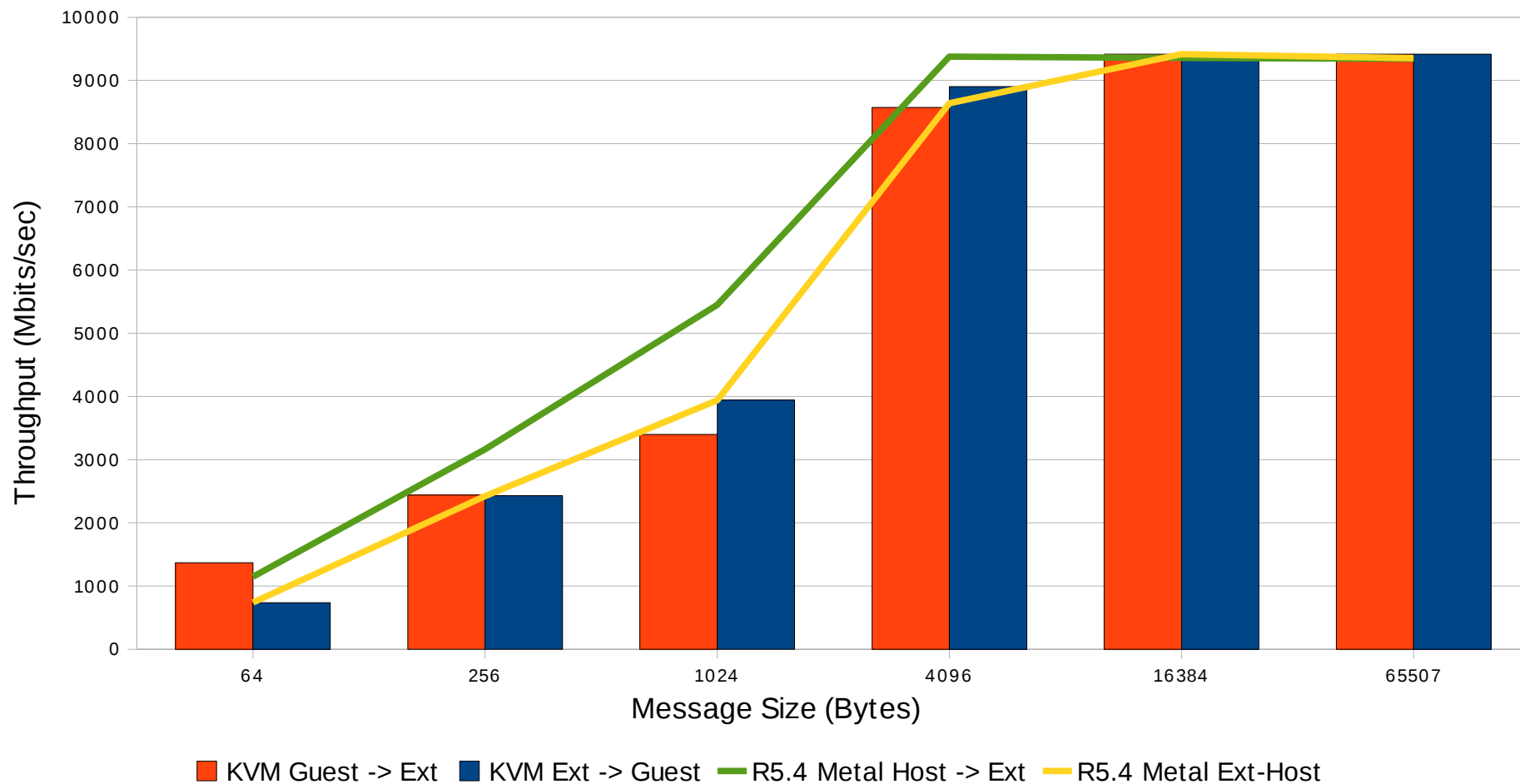
Dell Intel Nehalem, 2-10Gbit



Network Workloads – KVM w/ VT-d pci pass-through Intel Nehalem, Intel 10Gbit adapter

KVM Guest w/Device Assignment TCP Performance

single stream netperf, BareMetal vs KVM

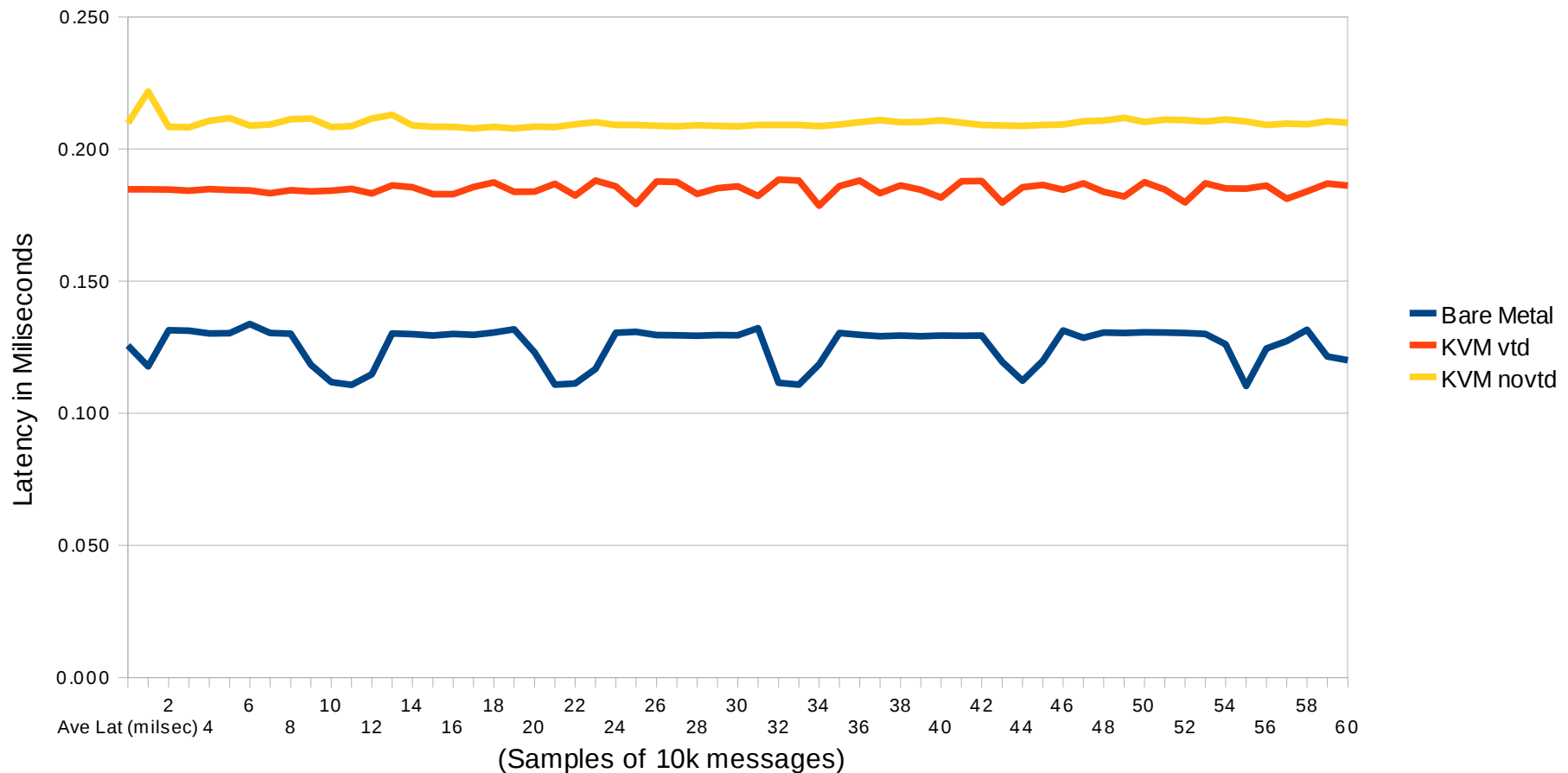


KVM Network Performance – AMQP Latency

Intel Nehalem 2guest 10Gbit Vt-d < 200 us lat

RHEL5.4 KVM AMQP Messaging Perf

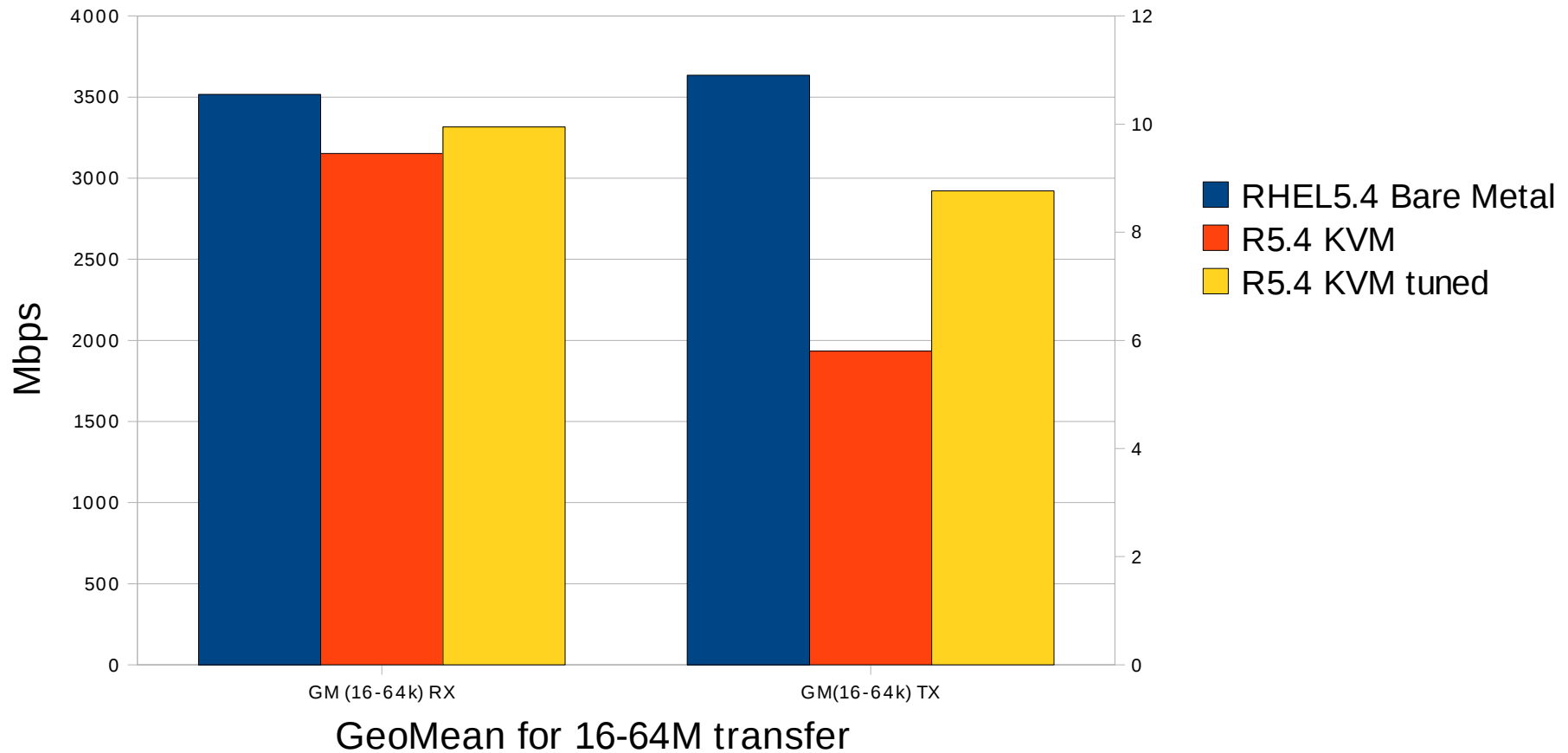
Dell R710 Intel Nahalem, 10Gbit w/ VT-d



Network Workloads – IRQ Tuning Effect on 10Gbit Intel Nehalem, Intel 10Gbit non-VT-d

RHEL5.4 KVM Network Tuning (IRQ)

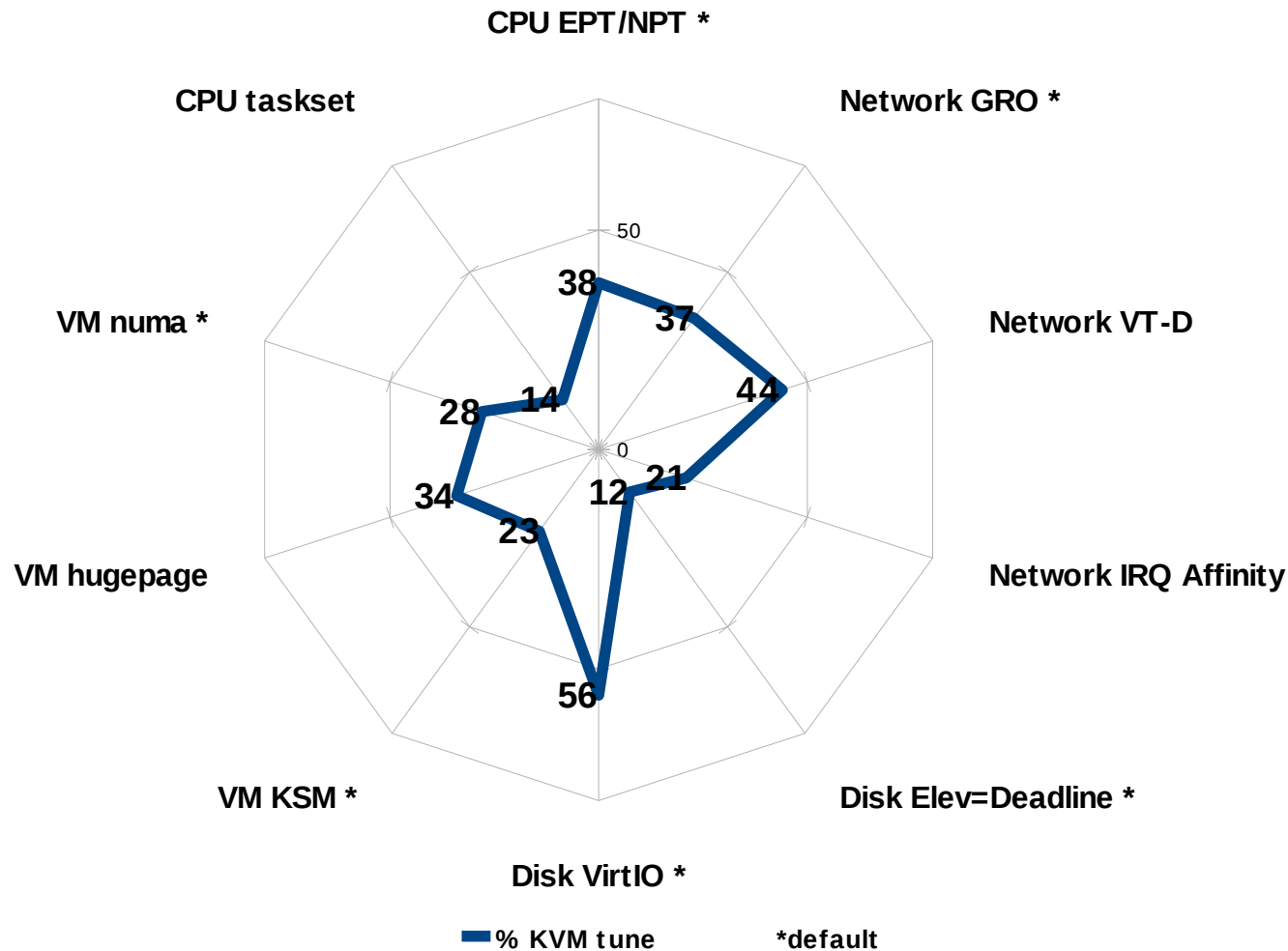
Intel Nehalem/10Gb enet



Summary KVM Tuning using “normal” Linux tools

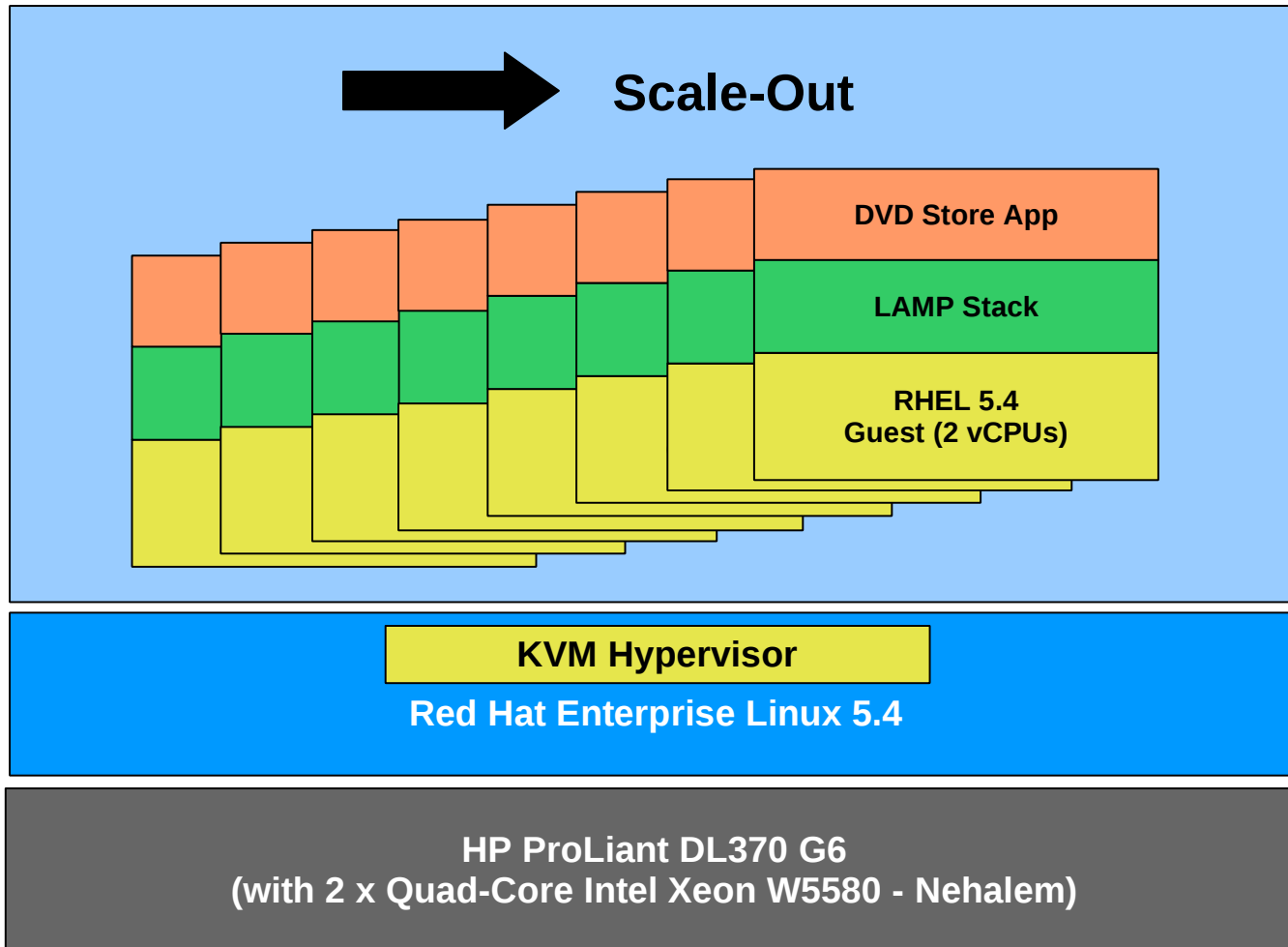
RHEL5.4 KVM Effect of Performance Tuning

Intel/AMD 16core machine, Fiber Channel

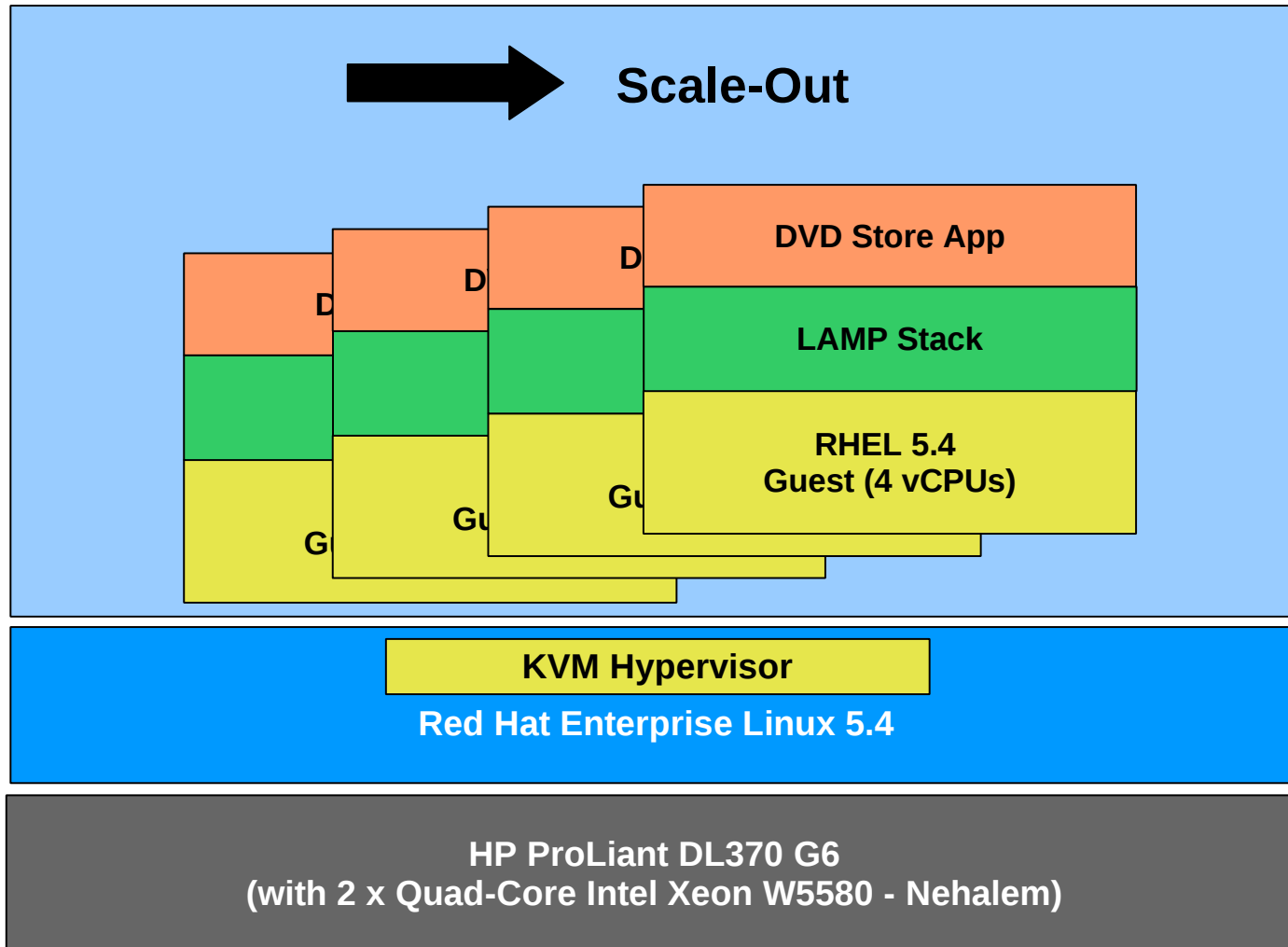


Scaling of Production Applications in a Red Hat Enterprise Virtualization Environment

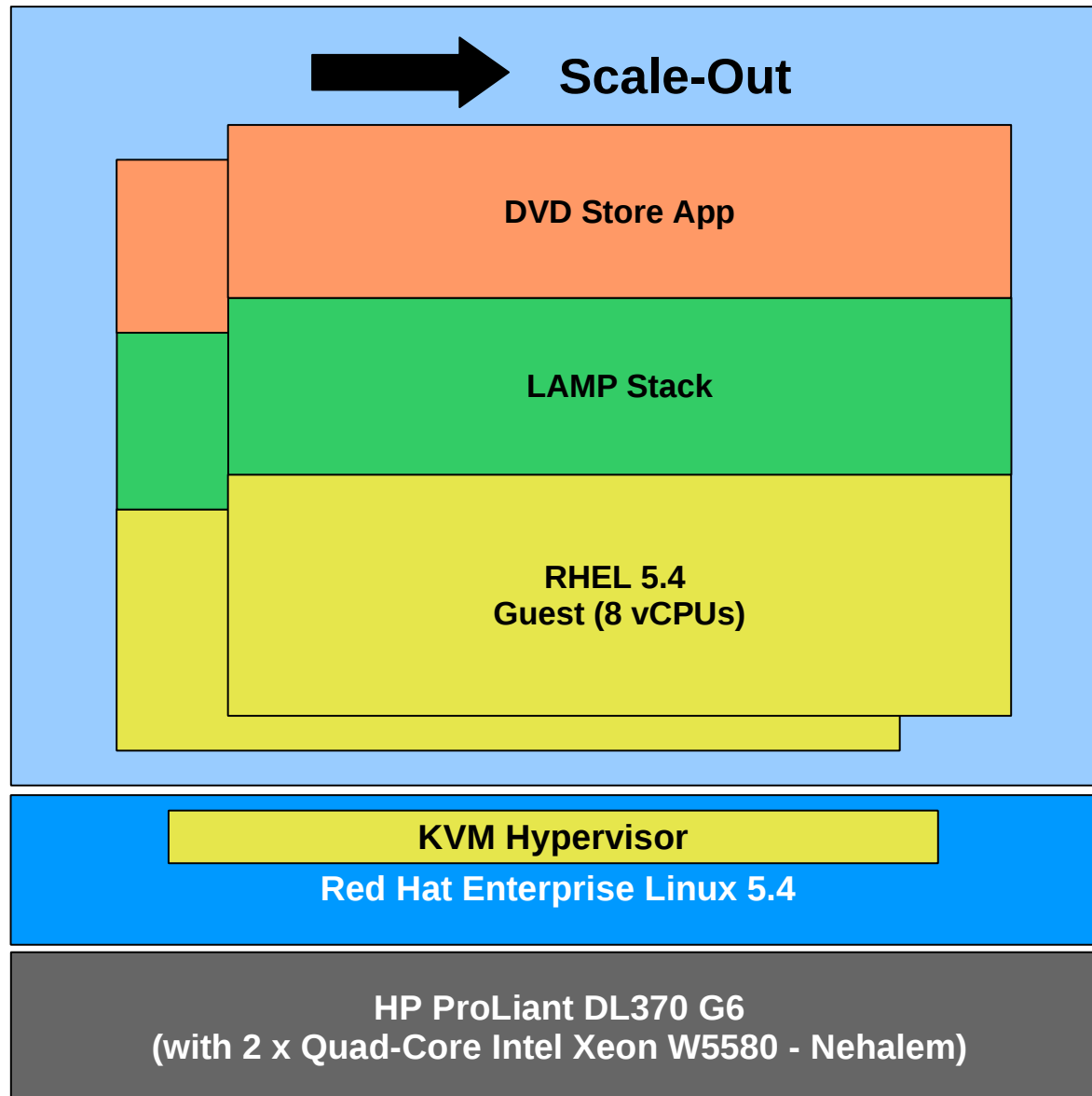
Scaling Multiple 2-vCPU Guests



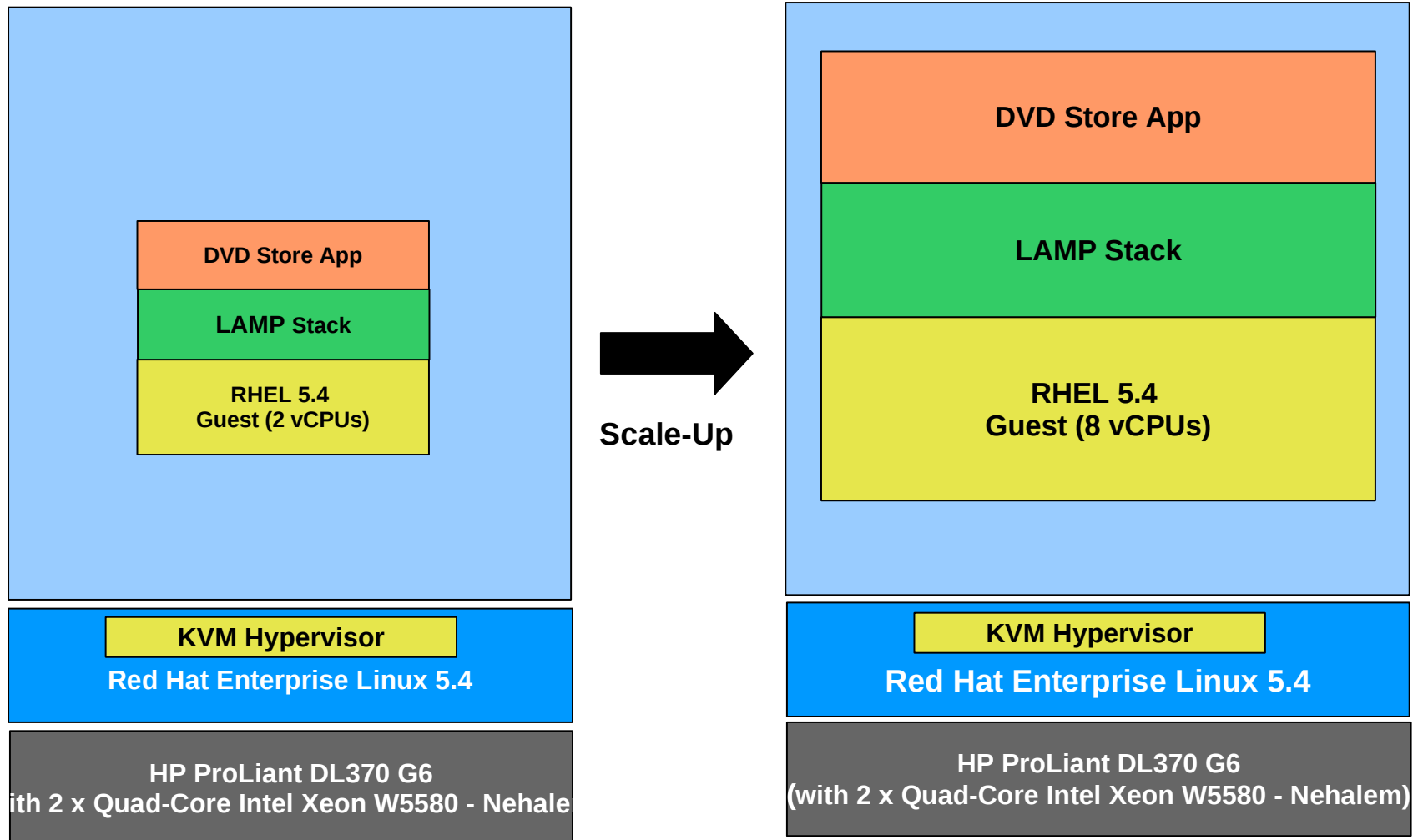
Scaling Multiple 4-vCPU Guests



Scaling Multiple 8-vCPU Guests



Scaling-up a Single Guest



Factors That Limit Scalability

- Hardware
 - Number cores/hyper-threads per socket – Socket Bandwidth
 - Caches and memory speeds, NUMA
 - I/O subsystems, e.g., storage bandwidth, network bandwidth
- Application
 - Locking and shared data
 - Memory reference, I/O
- Virtualization
 - Virt I/O efficiencies
 - Virtualization overhead



Red Hat Reference Architecture Series

Scaling Oracle 10g in a Red Hat® Enterprise Virtualization Environment

OLTP Workload
Oracle 10g
Red Hat® Enterprise Linux 5.3 Guest
Red Hat® Enterprise Linux 5.4 (with Integrated KVM Hypervisor)
HP ProLiant DL370 G6 (Intel Xeon W5580 - Nehalem)

Version v1.0
August 2009



Oracle – Configuration

HP ProLiant 370 G6	Dual Socket, Quad Core, Hyper Threading (Total of 16 processing threads) Intel(R) Xeon(R) CPU W5580 @ 3.20GHz
	12 x 4 GB DIMMs - 48 GB total
	6 x 146 GB SAS 15K dual port disk drives

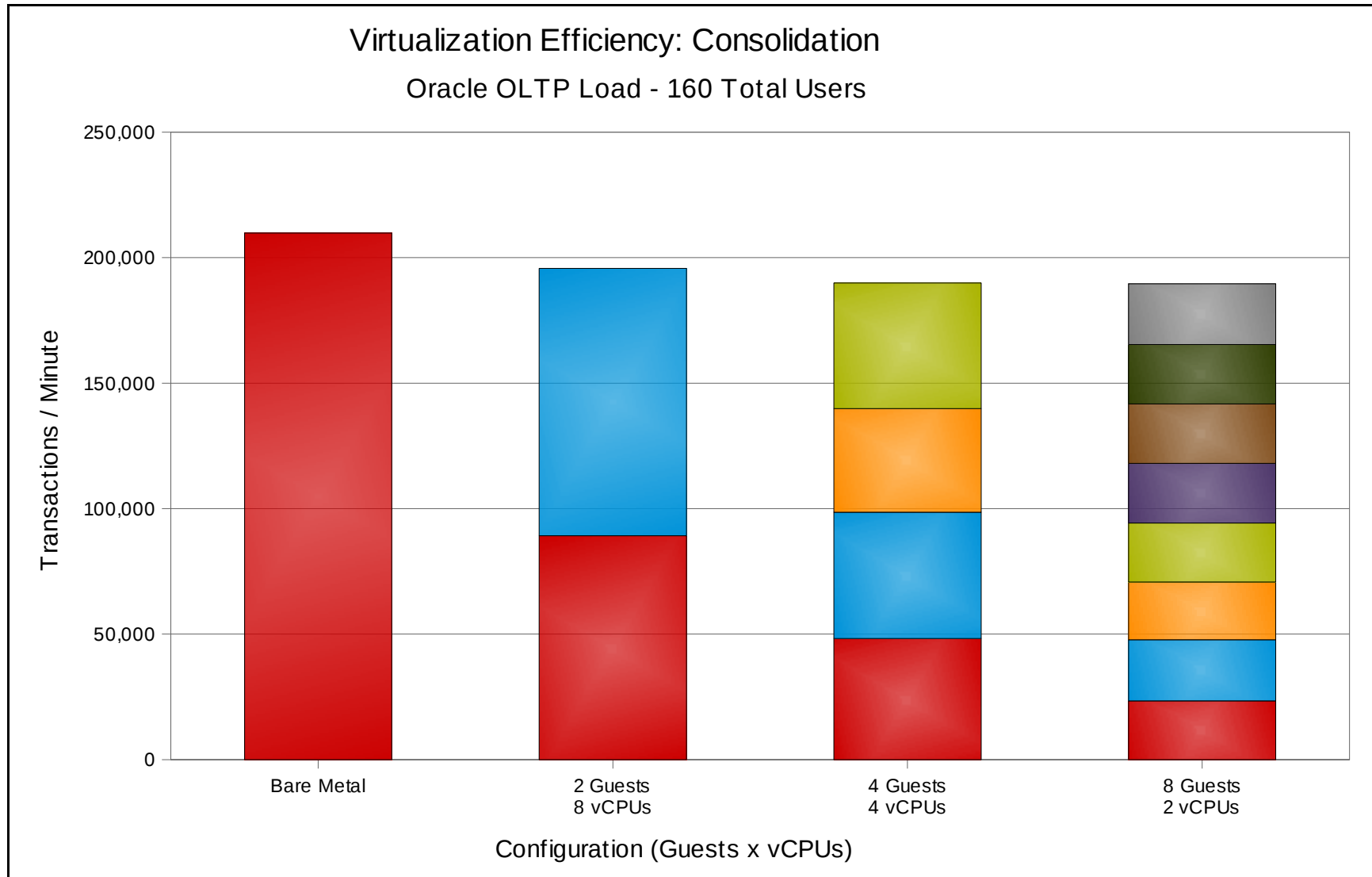
Red Hat® Enterprise Linux 5.4b	2.6.18-155.el5 kernel
KVM	kvm-83-80.el5
Oracle	v9.2.0.4

Oracle – Workload

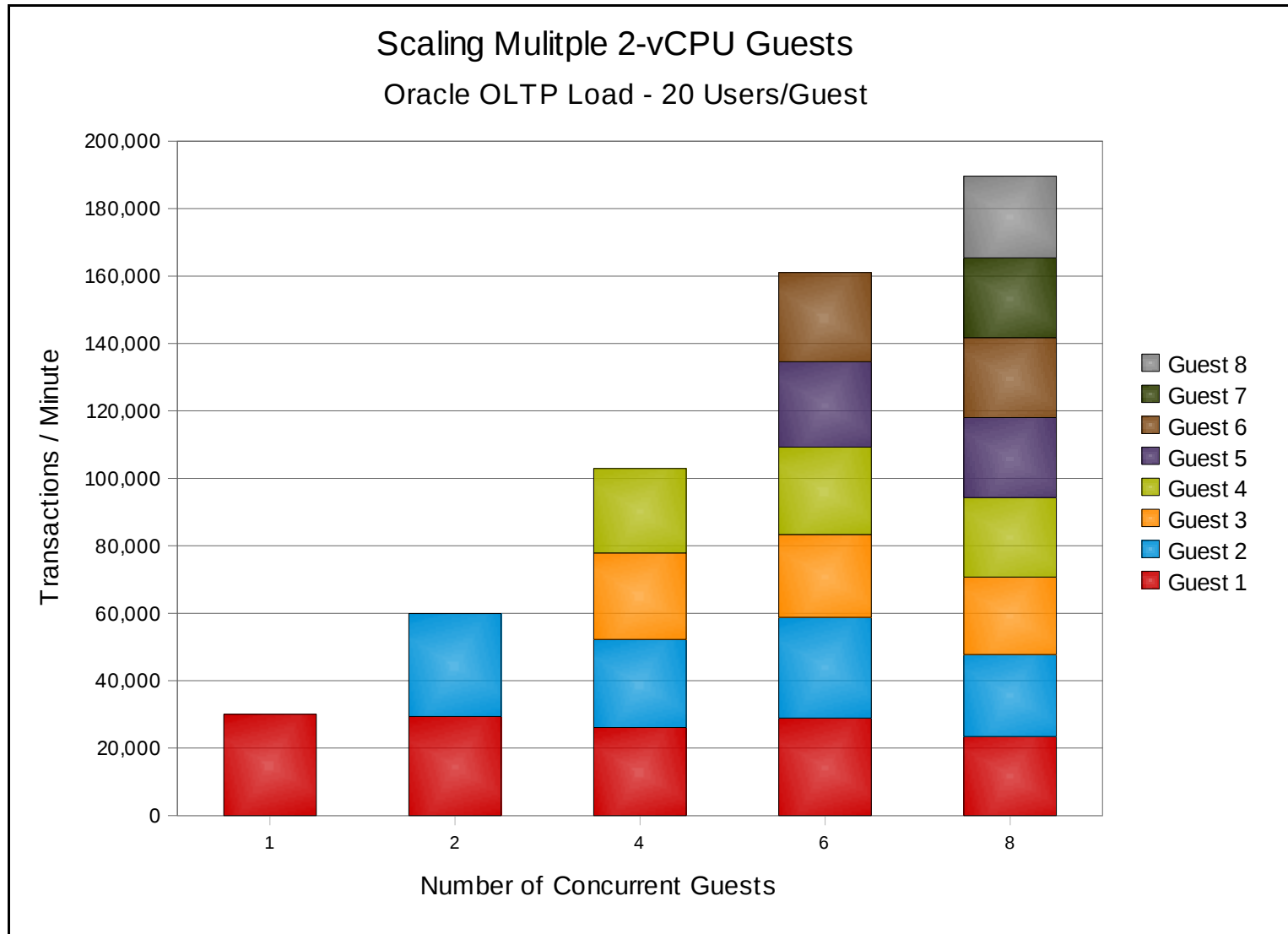
“TPC-C-like” OLTP Workload

VCPUs per Guest	Guest Memory	Oracle Users	Oracle SGA
1	2.5 GB	10	2 GB
2	5 GB	20	4 GB
4	10 GB	40	8 GB
6	15 GB	60	12 GB
8	20 GB	80	16 GB

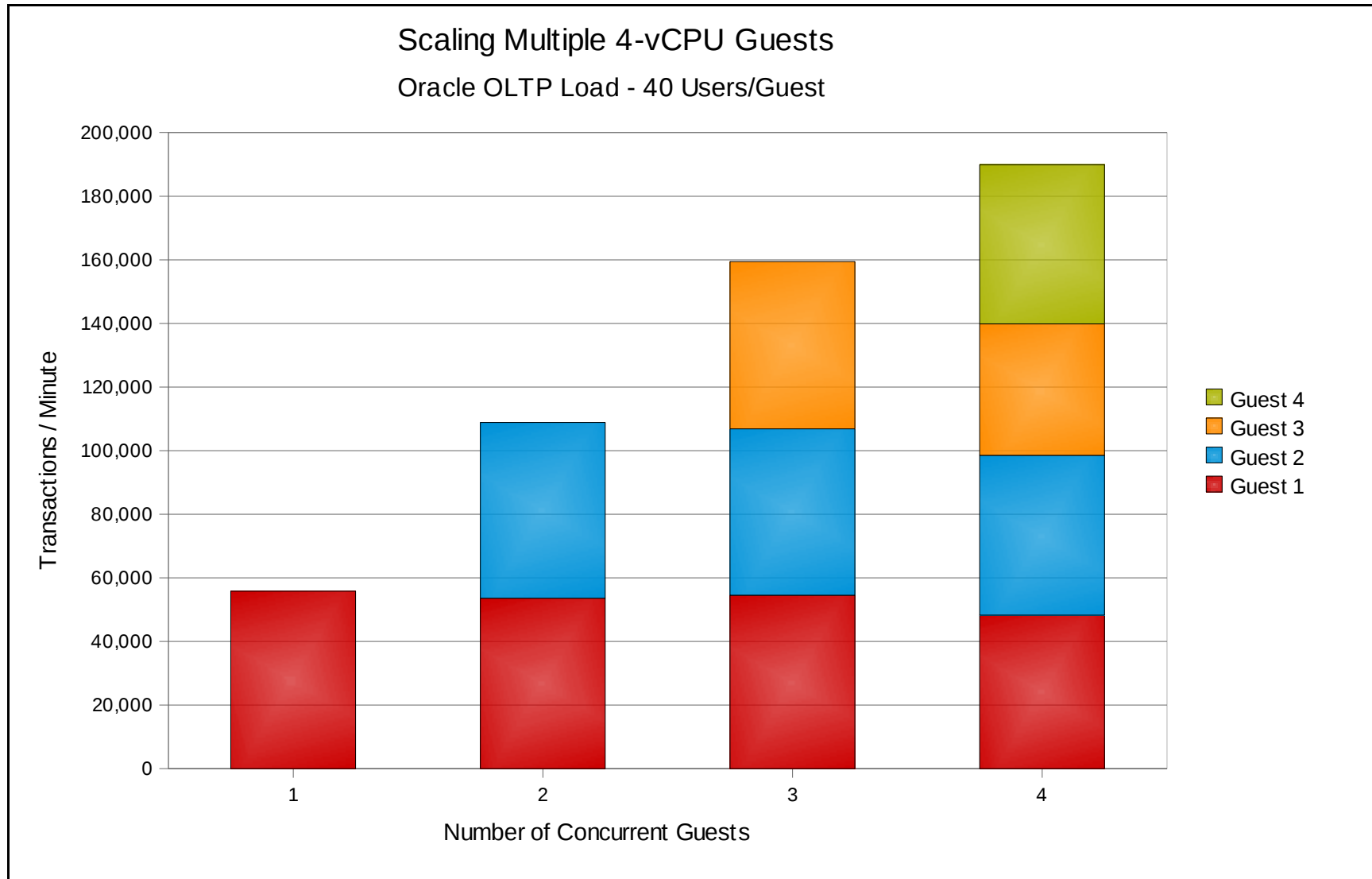
Oracle - Virtualization Efficiency: Consolidation



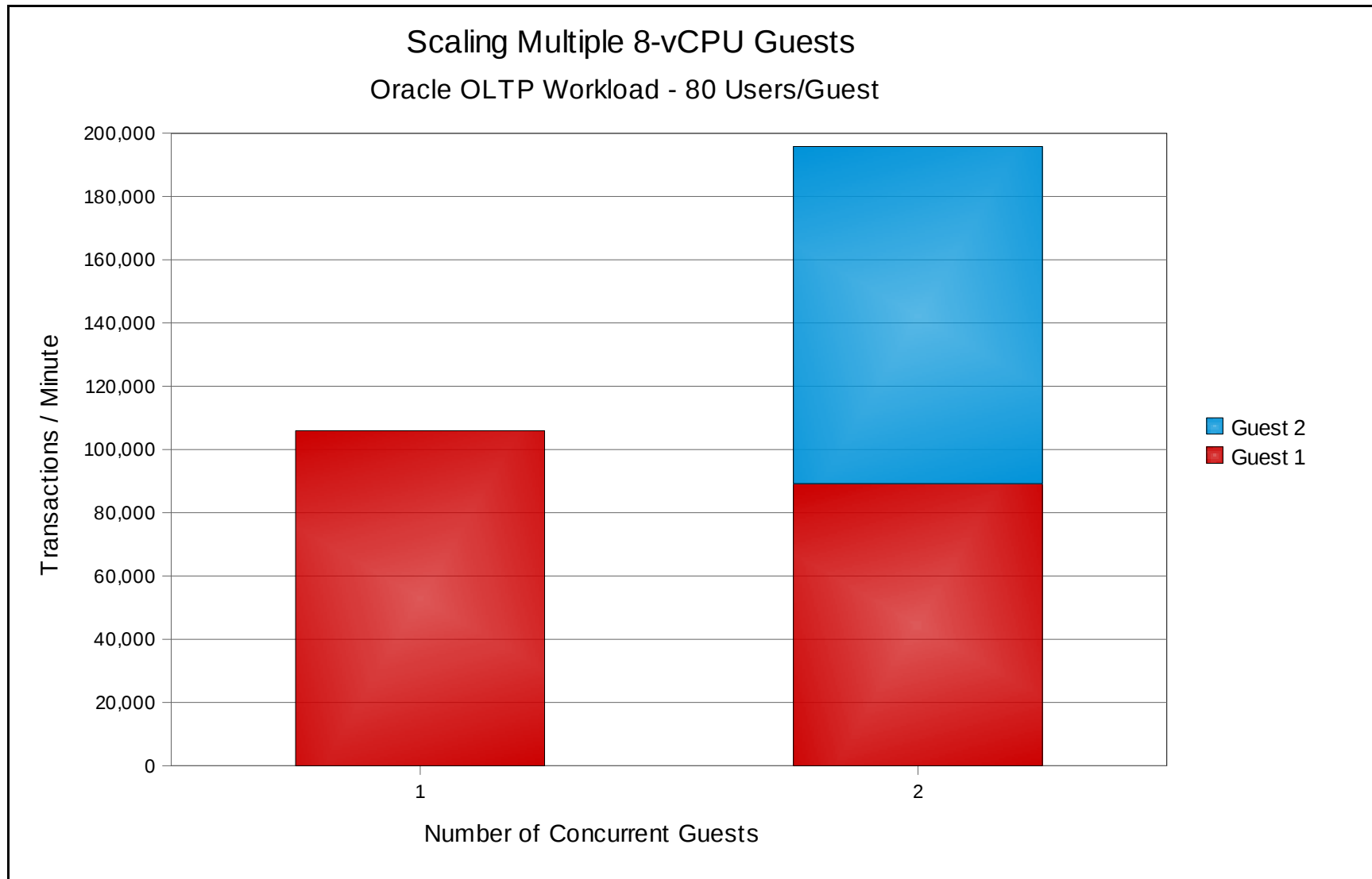
Oracle - Scaling Multiple 2-vCPU Guests



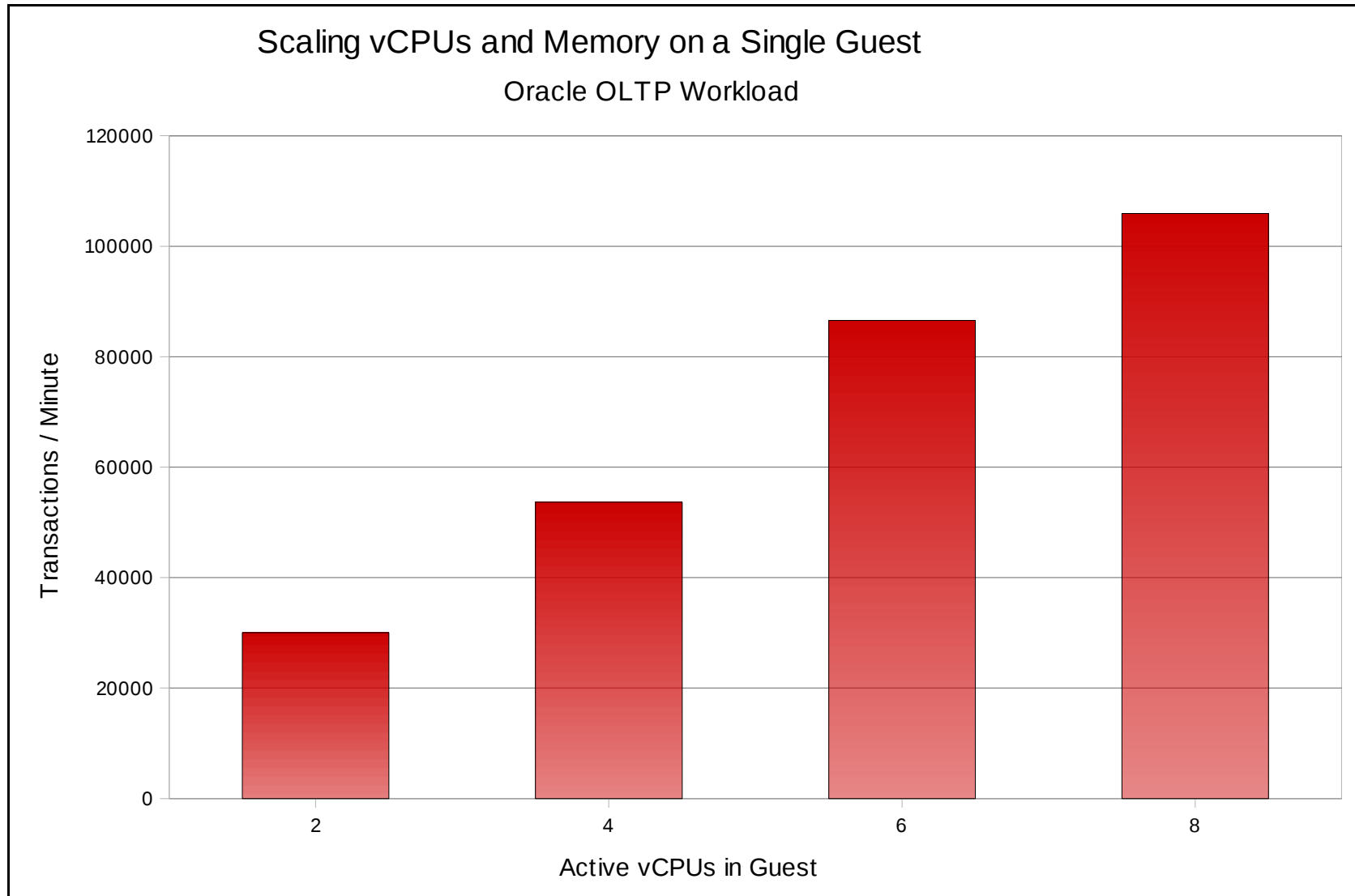
Oracle - Scaling Multiple 4-vCPU Guests



Oracle - Scaling Multiple 8-vCPU Guests



Oracle - Scaling-up a Single Guest





Red Hat Reference Architecture Series

Scaling the LAMP Stack in a Red Hat Enterprise Virtualization Environment

"DVD Store" LAMP Application		
Apache HTTP Server	PHP	MySQL
Red Hat Enterprise Linux 5.4 Guest		
Red Hat Enterprise Linux 5.4 (with integrated KVM Hypervisor)		
HP ProLiant DL370 G6 (Intel Xeon W5580 - Nehalem)		

Version v1.0
August 2009



LAMP Stack - Configuration

HP ProLiant DL370 G6	Dual Socket, Quad Core, Hyper Threading Technology (Total of 16 processing threads) Intel(R) Xeon(R) CPU W5580 @ 3.20GHz
	12 x 4 GB DIMMs- total: 48 GB
	6 x 146 GB SAS 15K dual port drives

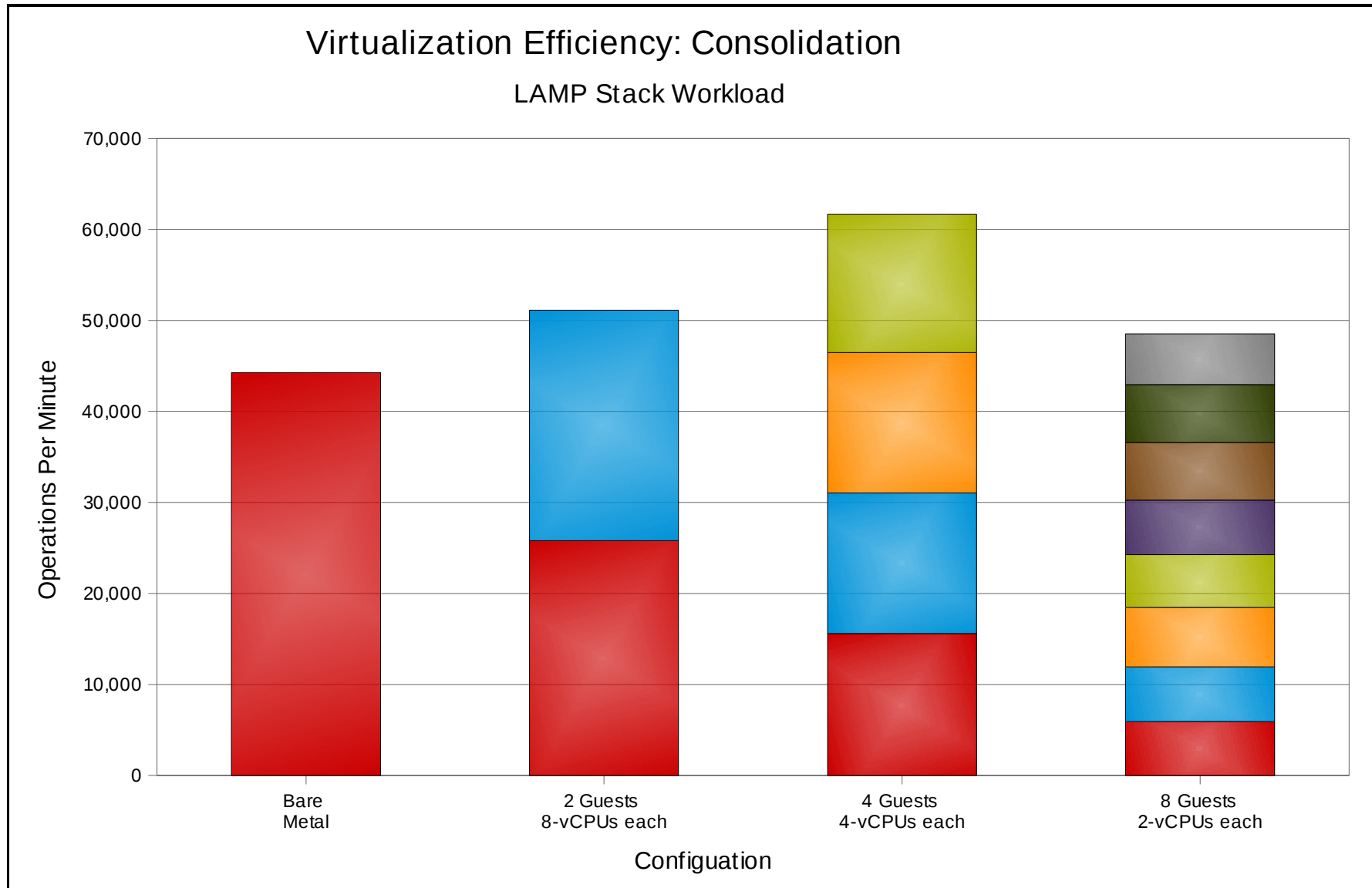
Red Hat Enterprise Linux 5.4 - Beta	2.6.18-155.el5
KVM	kvm-83-80.el5
Apache Httpd	v2.2.3
MySQL	v5.0.77
PHP	v5.1.6
APC	v3.0.19

LAMP Stack – Workload

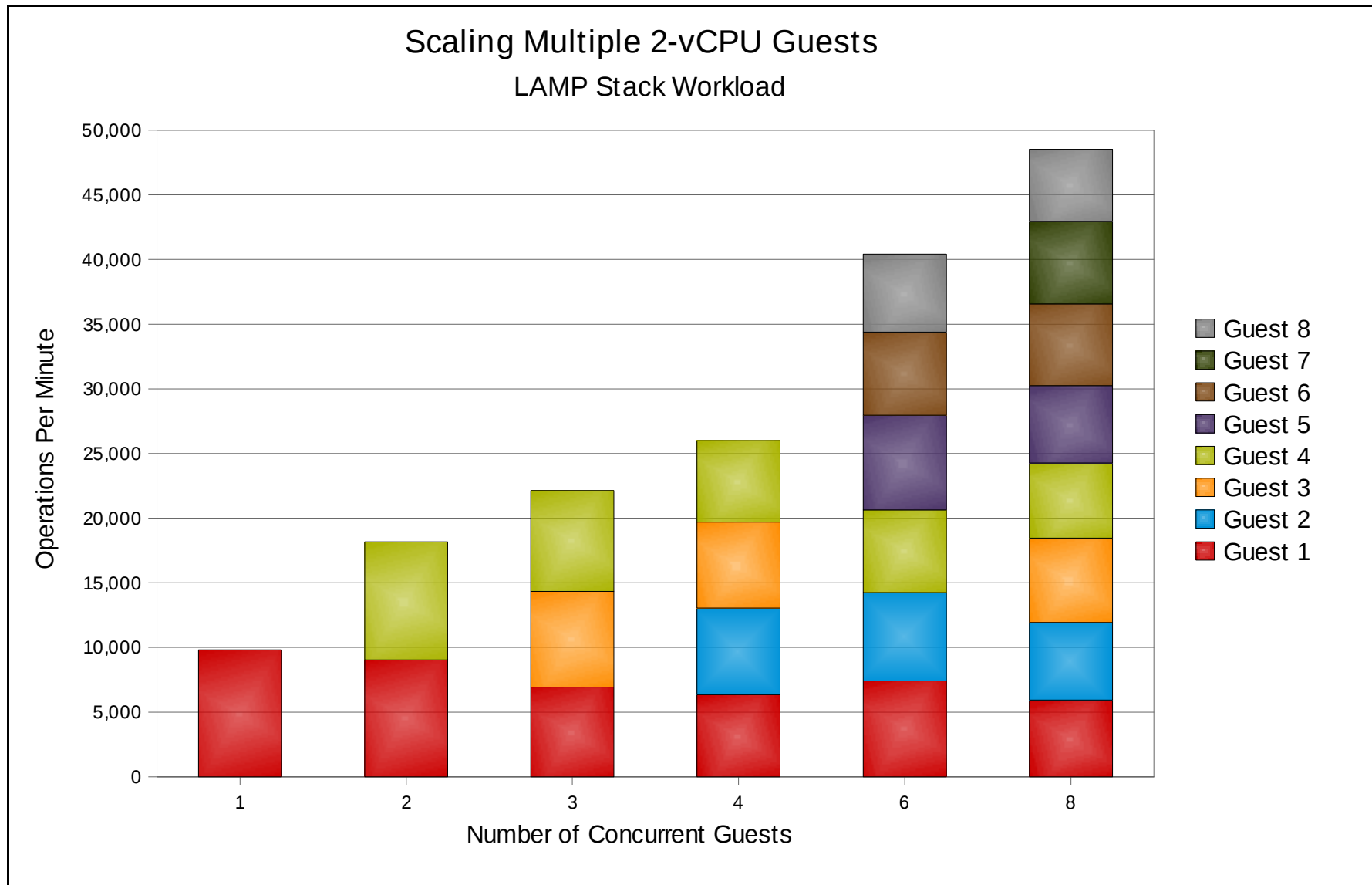
“DVD Store” LAMP Workload

vCPUs per Guest	Guest Memory	User Load	InnoDB Buffer Pool
1	2.5 GB	35	2 GB
2	5 GB	70	4 GB
4	10 GB	140	8 GB
6	15 GB	210	12 GB
8	20 GB	280	16 GB

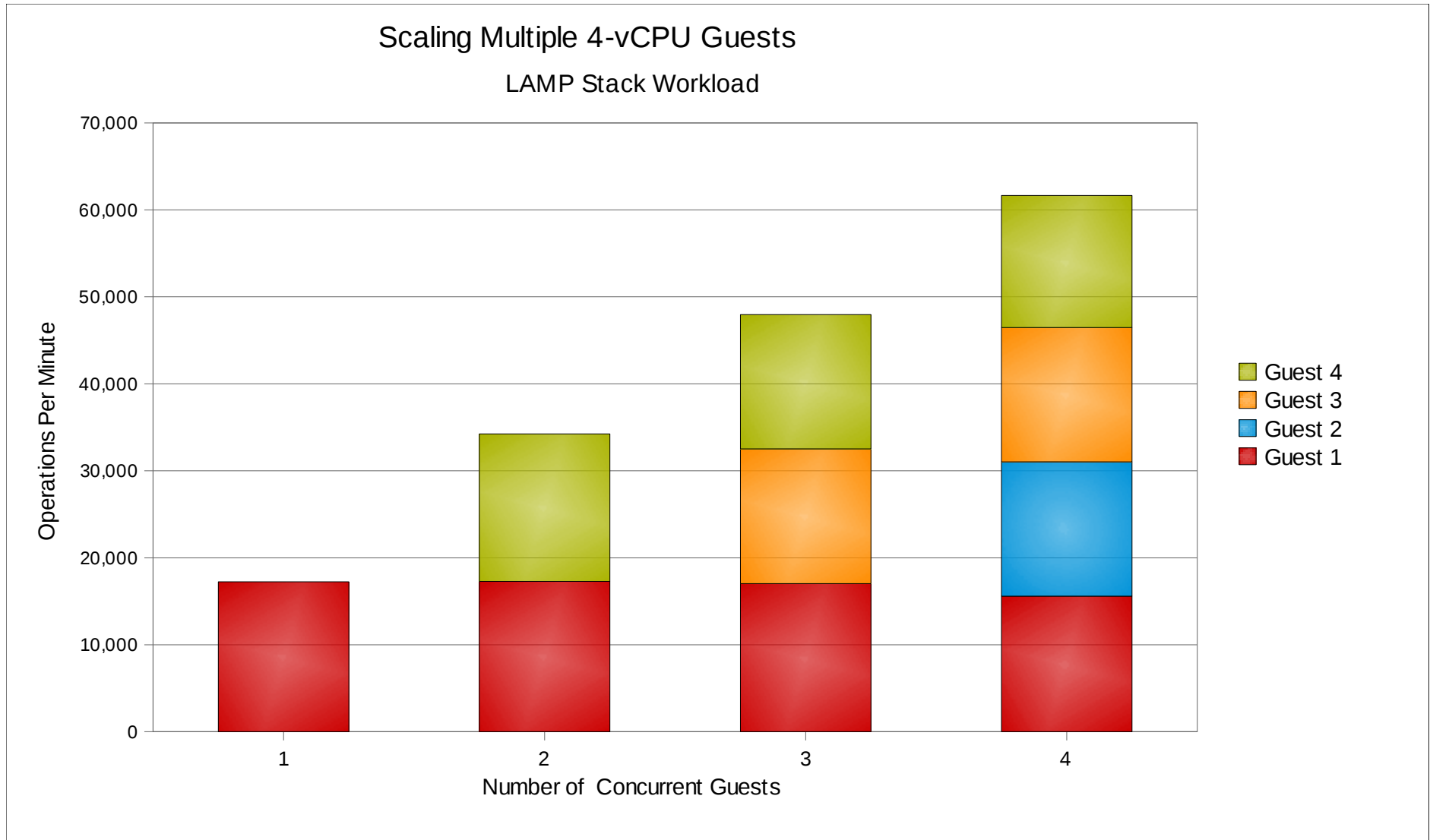
LAMP - Virtualization Efficiency: Consolidation



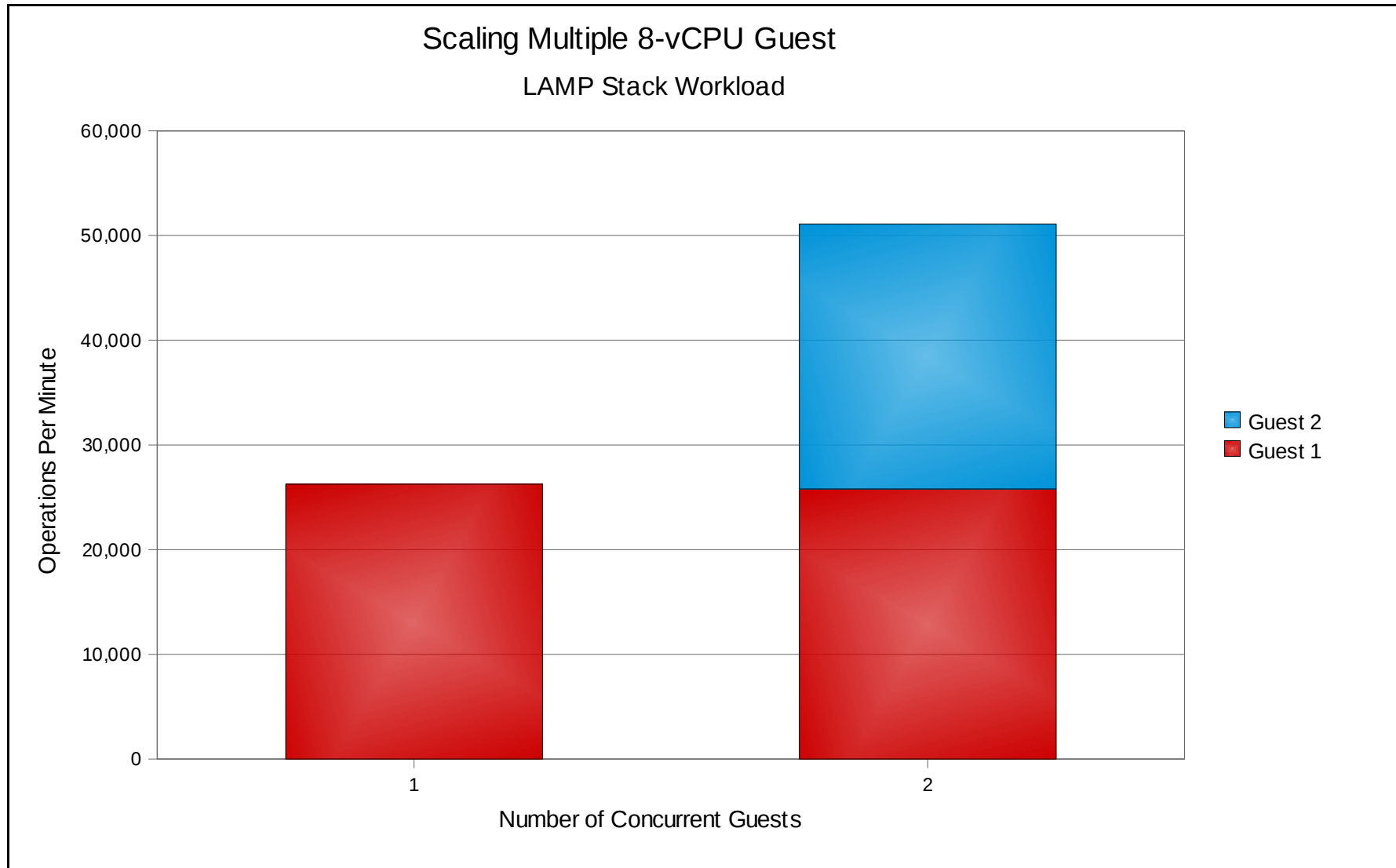
LAMP - Scaling Multiple 2-vCPU Guests



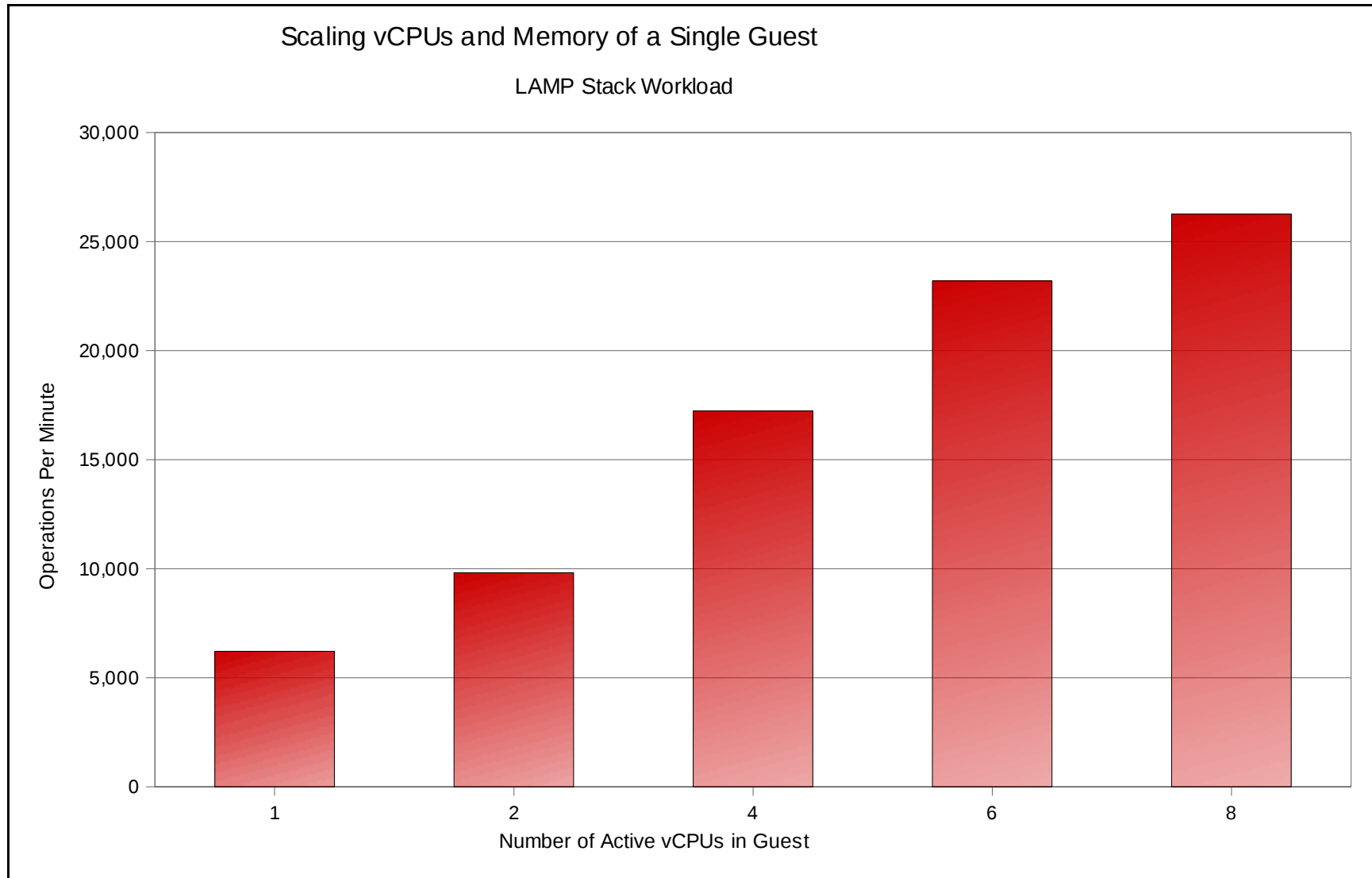
LAMP - Scaling Multiple 4-vCPU Guests



LAMP - Scaling Multiple 8-vCPU Guests



LAMP - Scaling-up a Single Guest





Red Hat Reference Architecture Series

Scaling SAP in a Red Hat® Enterprise Virtualization Environment

SAP LinuxLab Certification Suite (SLCS) 2.3
SAP ECC 6
MaxDB Database 7.6.06_09
Red Hat® Enterprise Linux 5.4 Guest
Red Hat® Enterprise Linux 5.4 (with Integrated KVM Hypervisor)
Intel Xeon X5570 (Nehalem)

Version v1.0
August 2009



SAP – Configuration & Workload

“SAP LinuxLab Certification Suite (SLCS)” Workload

Quad Core, Hyper Threading (Total of 16 processing threads)

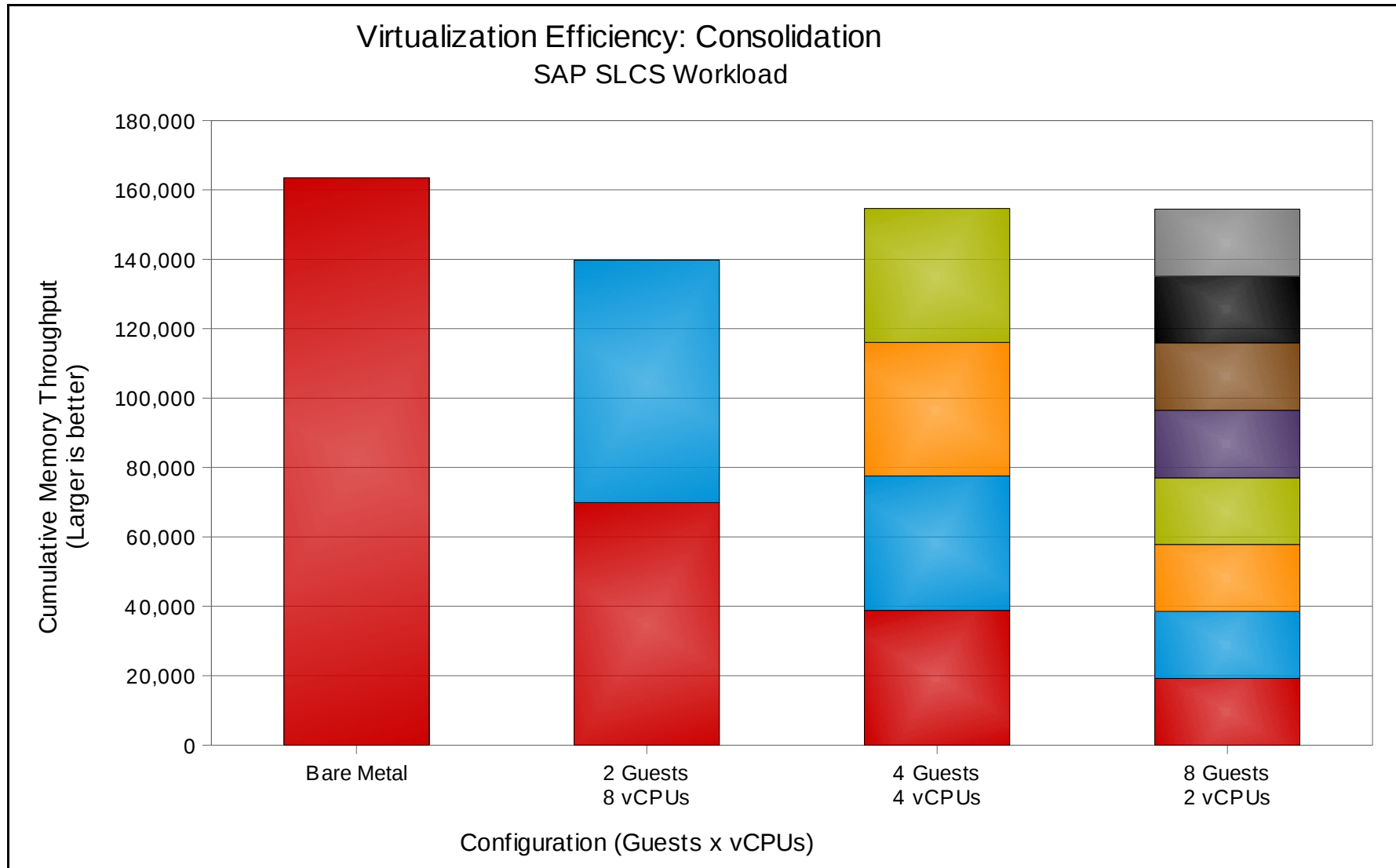
Intel(R) Xeon(R) CPU X5570 @ 2.93GHz

12 x 4 GB 1066MHz DDR3 RDIMMs - 48 GB total

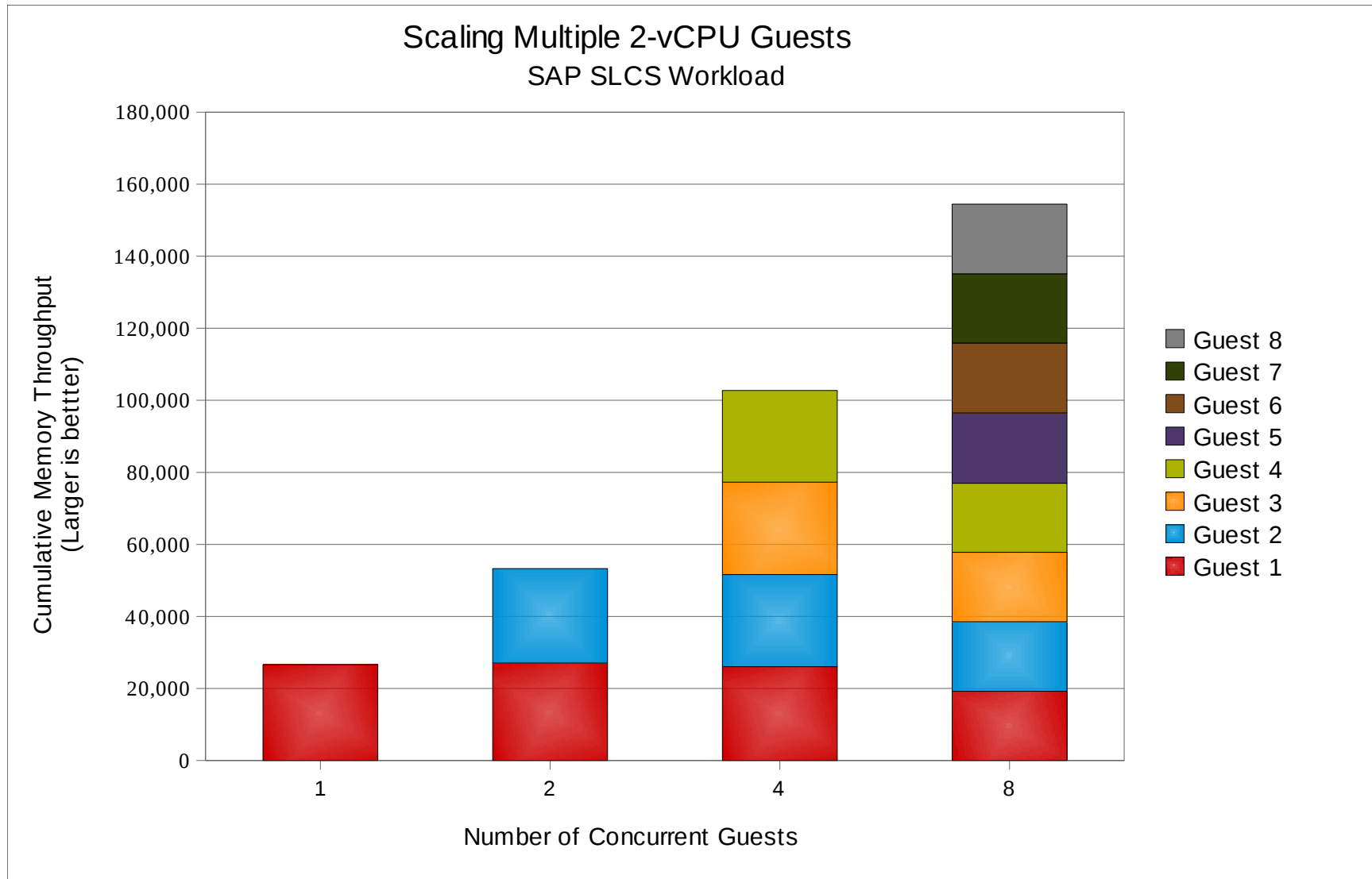
1 x 146 GB SAS 15K dual port disk drive

Red Hat® Enterprise Linux 5.4b	2.6.18-155.el5 kernel
KVM	kvm-83-80.el5
SAP	ECC 6
SLCS	2.3
MaxDB	7.6.06_09

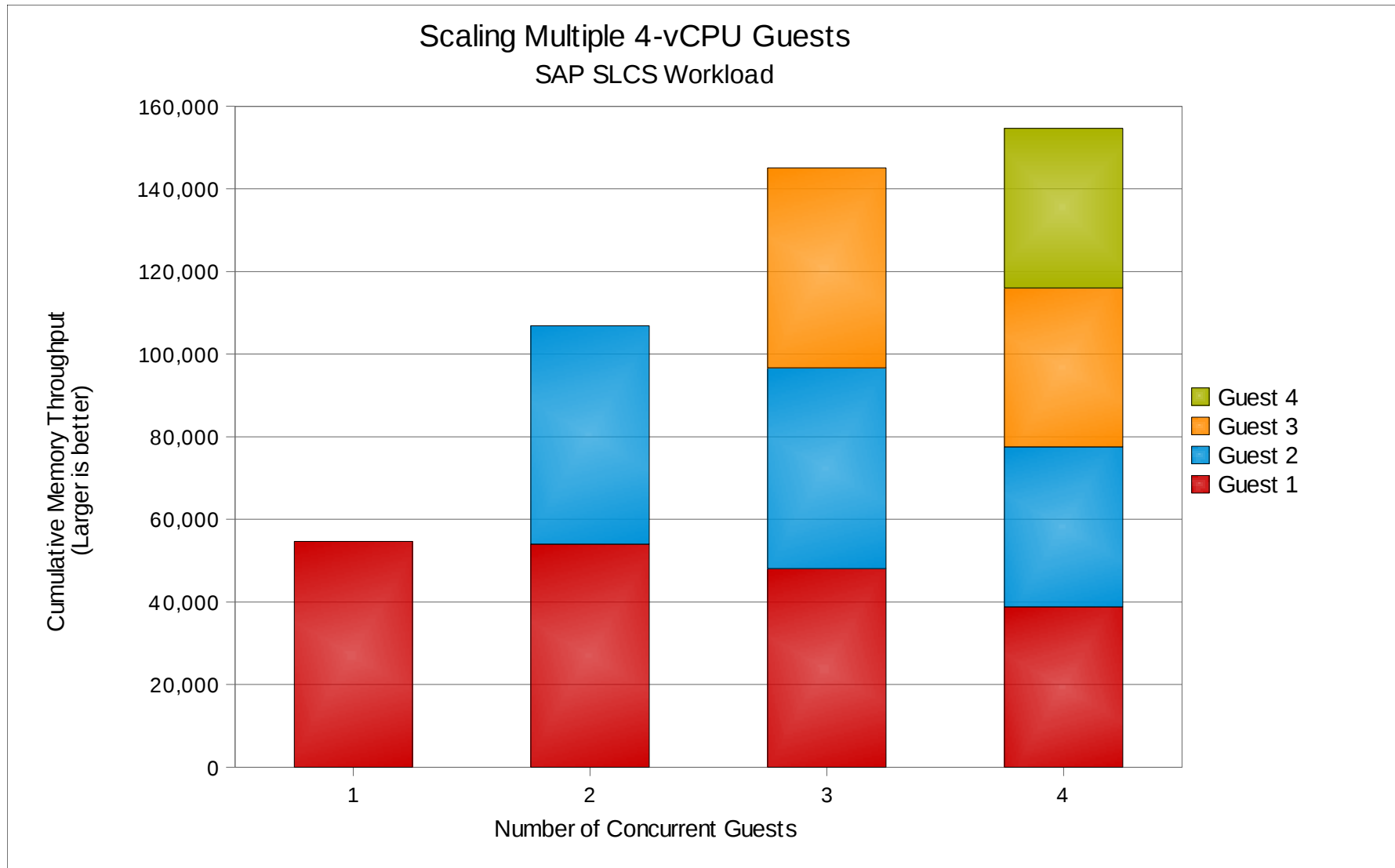
SAP - Virtualization Efficiency: Consolidation



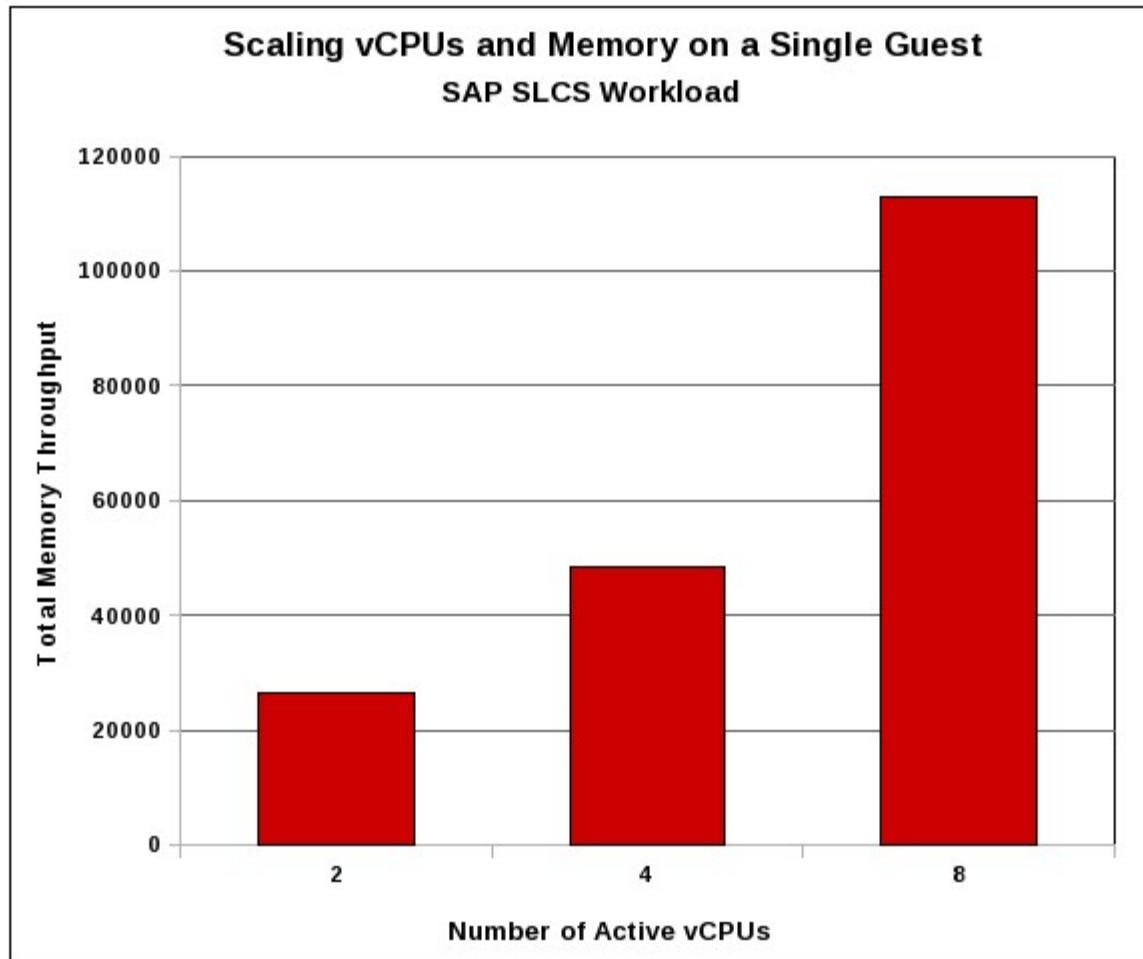
SAP - Scaling Multiple 2-vCPU Guests



SAP - Scaling Multiple 4-vCPU Guests



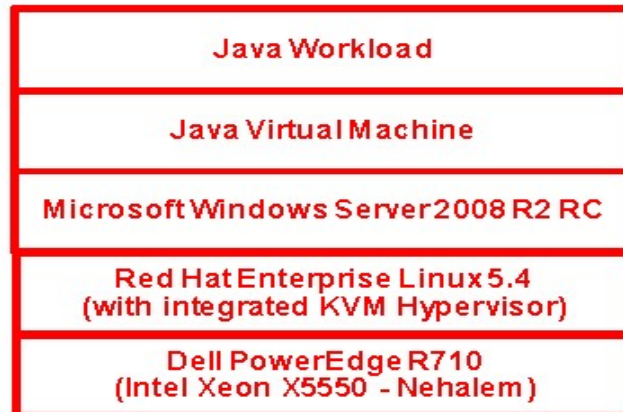
SAP - Scaling-up a Single Guest





Red Hat Reference Architecture Series

Scaling Java applications in a Red Hat Enterprise Virtualization Environment



Version 1.0
August 2009



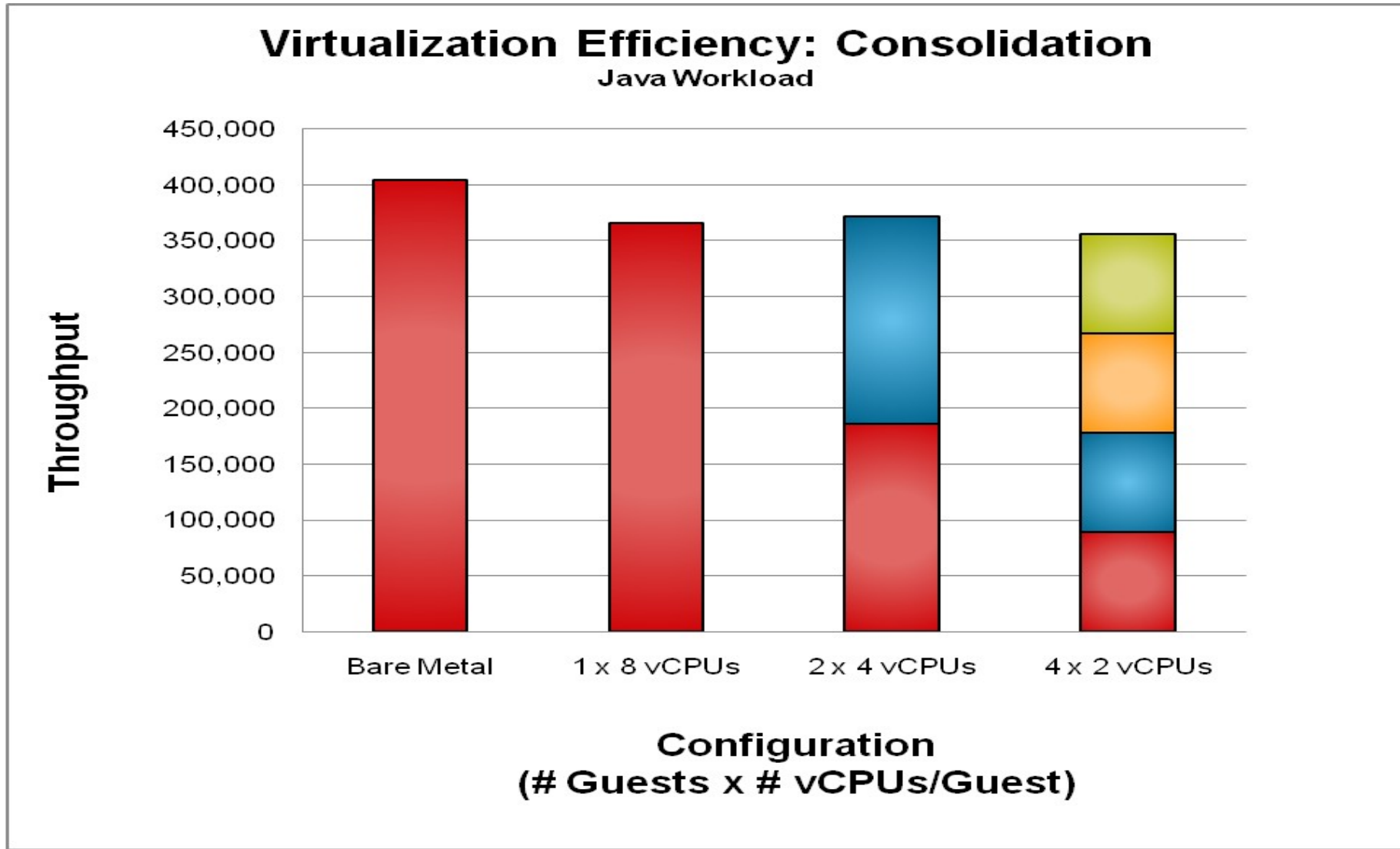
Java – Configuration & Workload

“SPECjbb-like” Workload

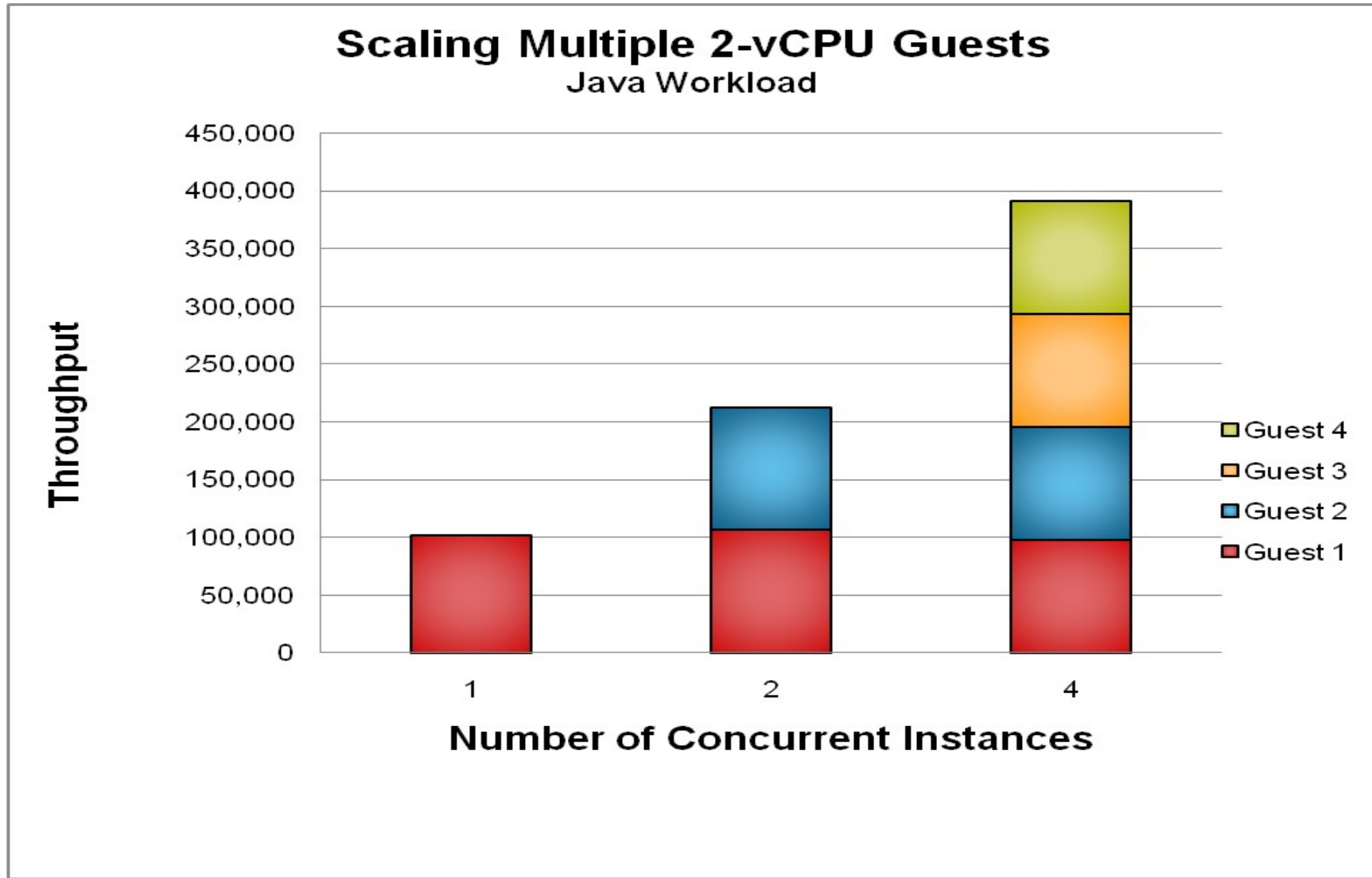
Dell™ PowerEdge™ R710	Dual Socket, Quad Core, Hyper-threading (Total of 16 processing threads) Intel® Xeon® CPU X5550 @ 2.66GHz
	18 x 4 GB DIMMs - total: 72 GB
	2 x 146 GB SAS 15K dual port drives

Red Hat Enterprise Linux 5.4 - Beta	2.6.18-159.el5
KVM	kvm-83-80.el5
Microsoft Windows Server 2008 R2	RC (Build 7100)

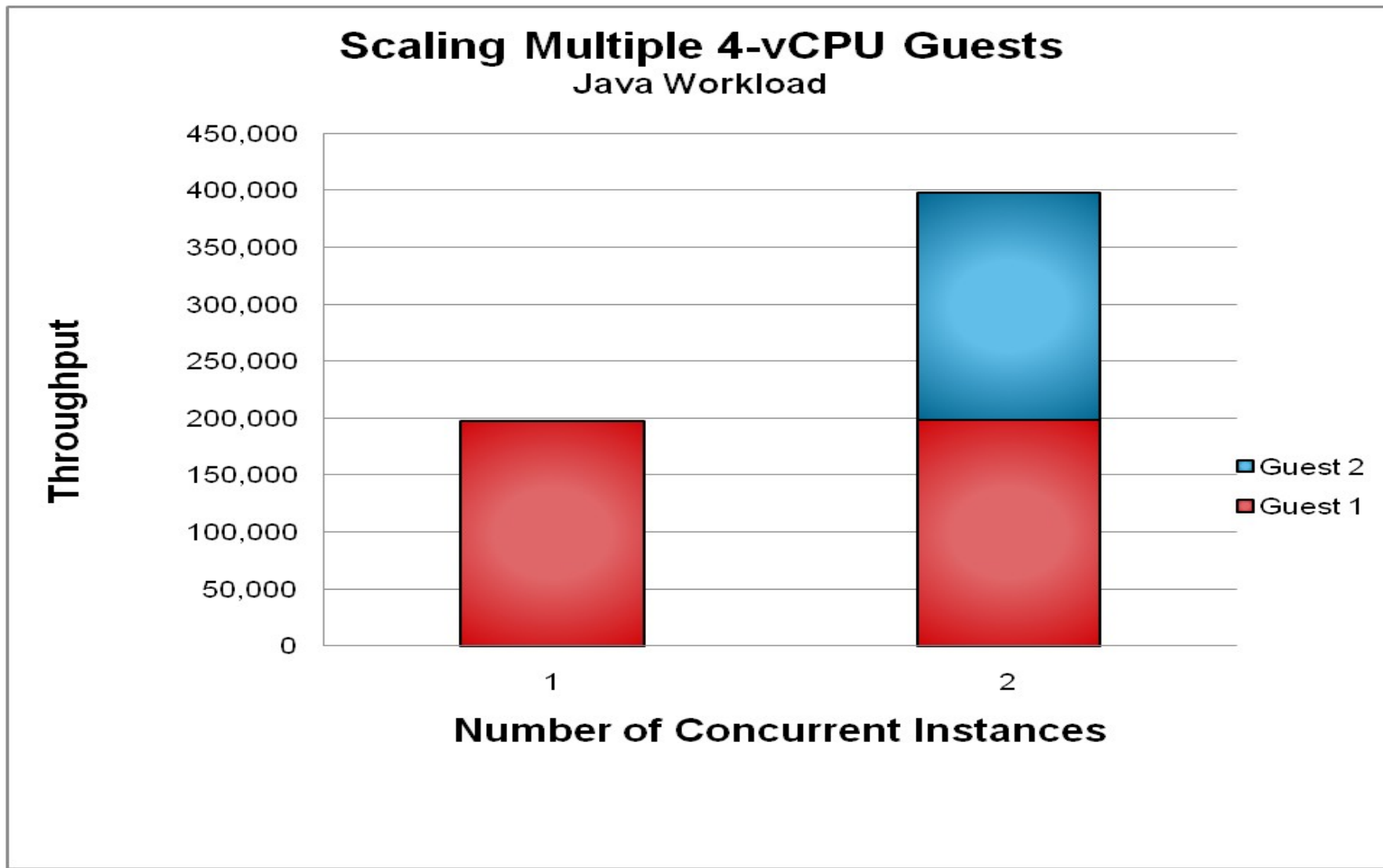
Java - Virtualization Efficiency: Consolidation



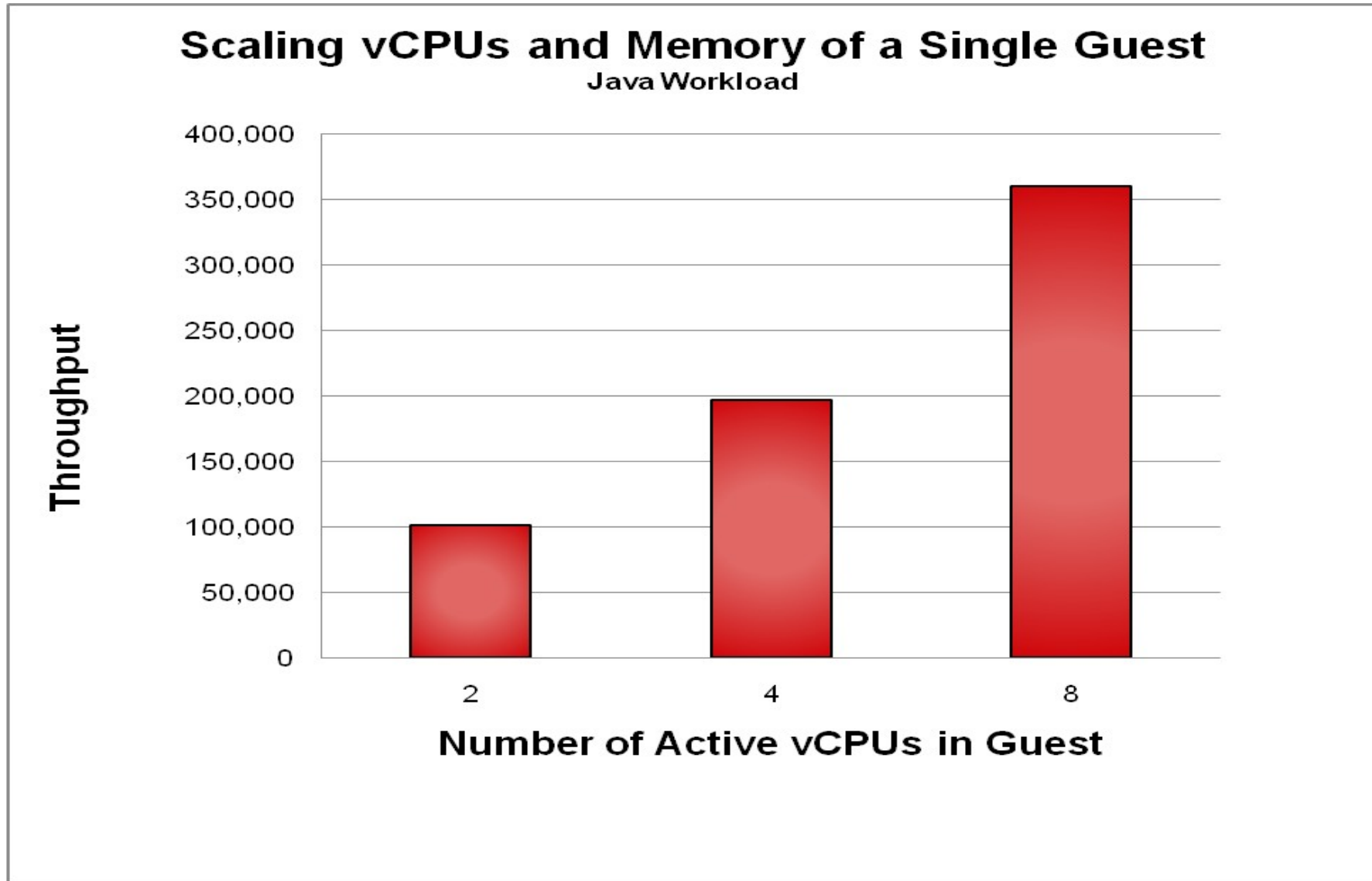
Java - Scaling Multiple 2-vCPU Guests



Java - Scaling Multiple 4-vCPU Guests



Java - Scaling-up a Single Guest





Red Hat Reference Architecture Series

Scaling Microsoft Exchange in a Red Hat Enterprise Virtualization Environment

LoadGen Workload
Microsoft Exchange Server2007
Microsoft Windows Server2008 R2RC
Red Hat Enterprise Linux 5.4 (with integrated KVM Hypervisor)
Intel Xeon Processor E5540 (Nehalem)

Version 1.0
August 2009



Exchange - Configuration

Dell™ PowerEdge™ R710	Dual Socket, Quad Core, Hyper-threading (Total of 16 processing threads) Intel® Xeon® CPU E5540 @ 2.53GHz
	18 x 4 GB DIMMs - total: 72 GB
	2 x 146 GB SAS 15K drives

Red Hat Enterprise Linux 5.4 - Beta	2.6.18-159.el5
KVM	kvm-83-80.el5
Microsoft Windows Server 2008 R2	RC (Build 7100)
Microsoft Exchange 2007 SP1	08.01.0240.006

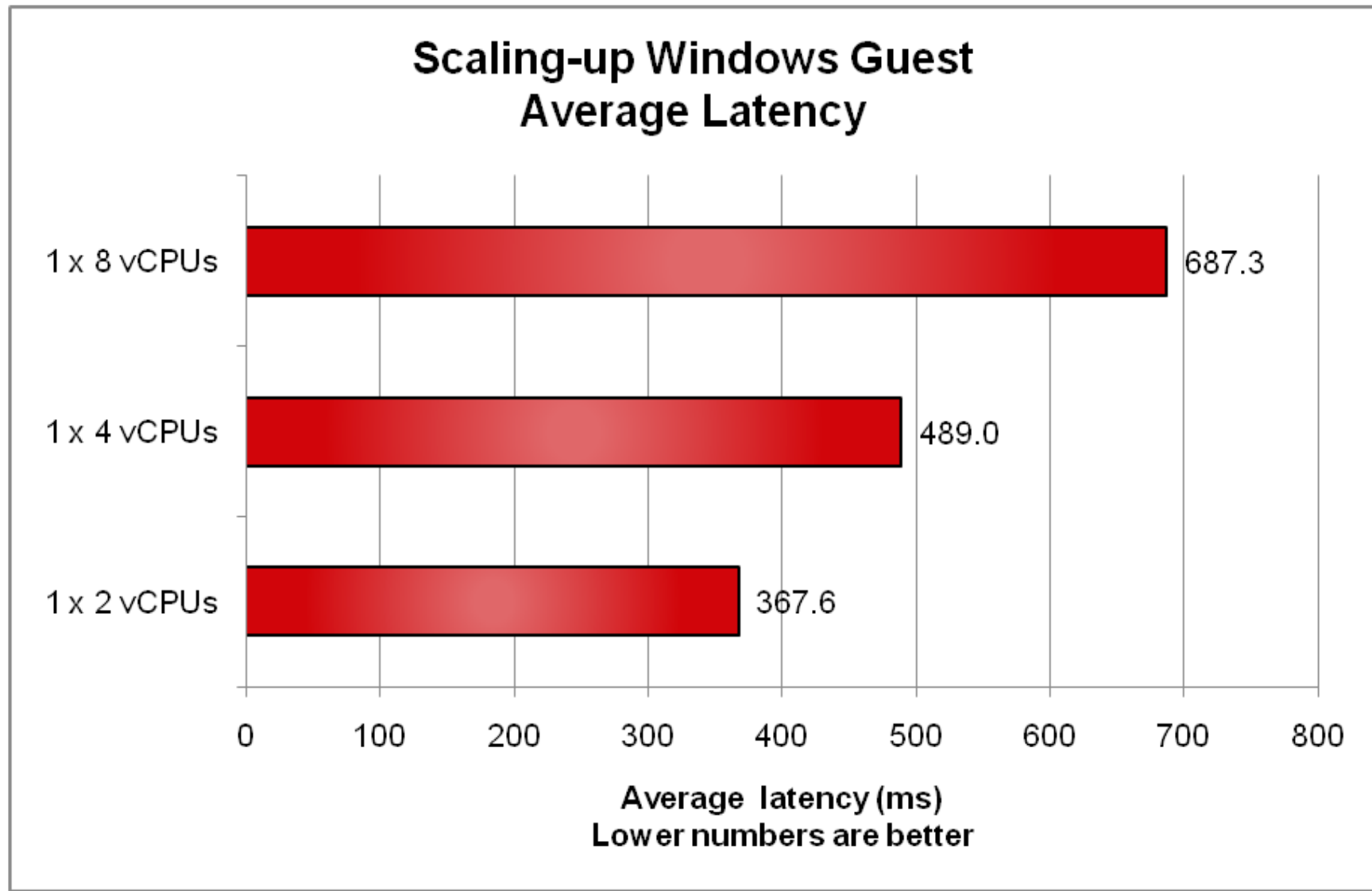
MS Exchange – Workload

“MS LoadGen” Workload

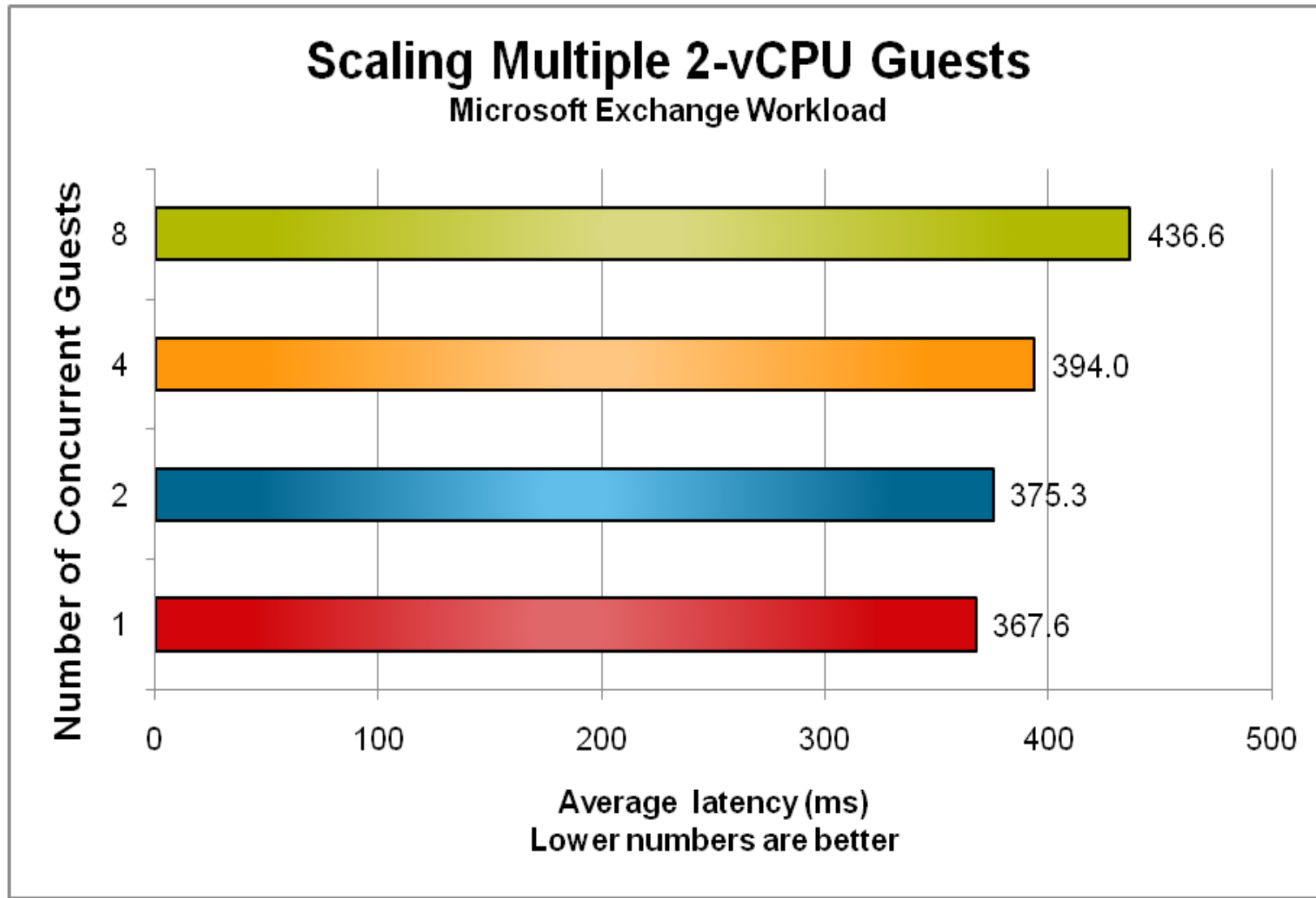
VCPUs per Guest	Guest Memory	Exchange Users
2	5 GB	2,000
4	10 GB	4,000
8	20 GB	8,000

95th percentile Latency < 750 msec.

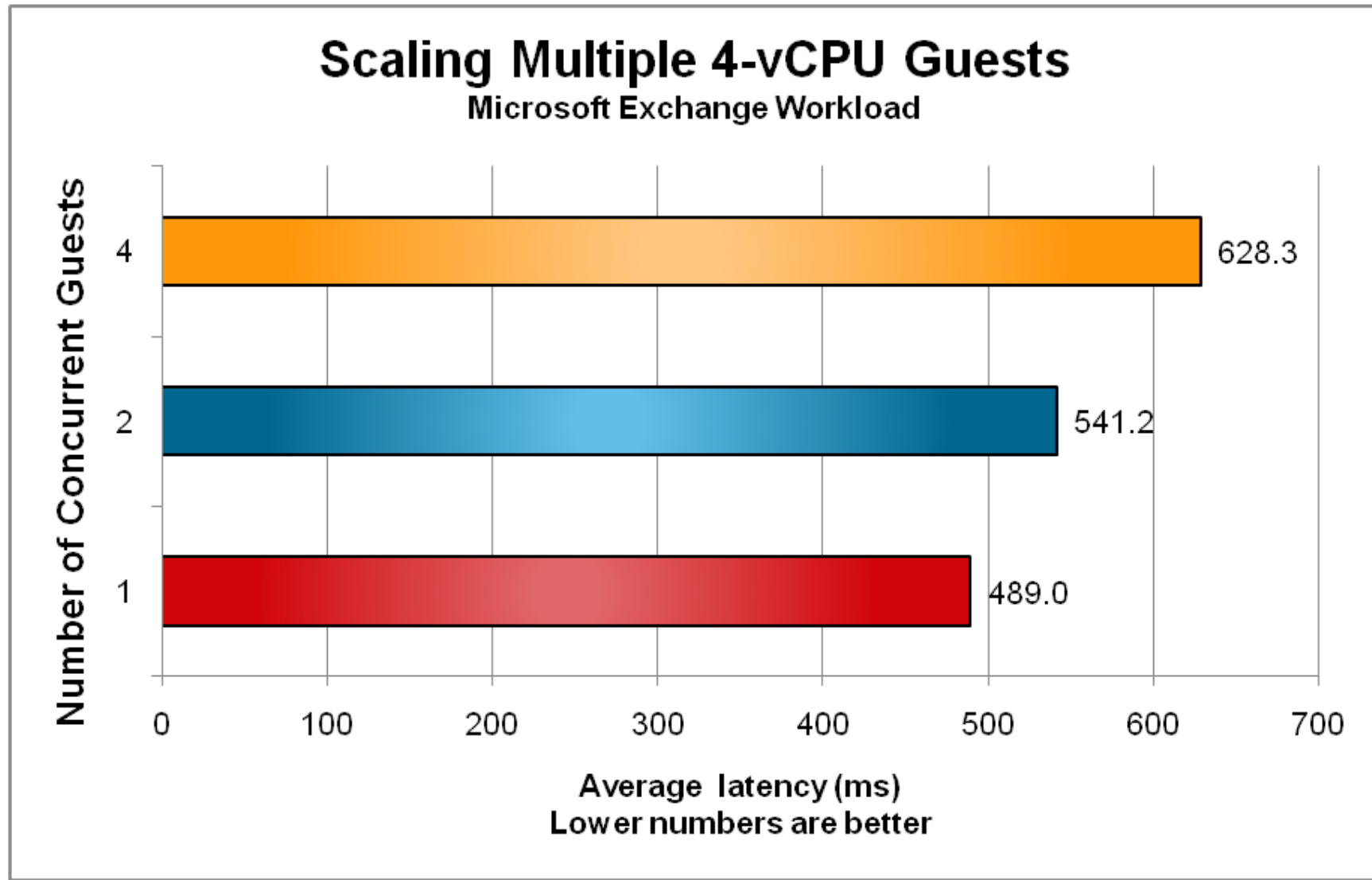
Exchange - Scaling-up a Single Guest



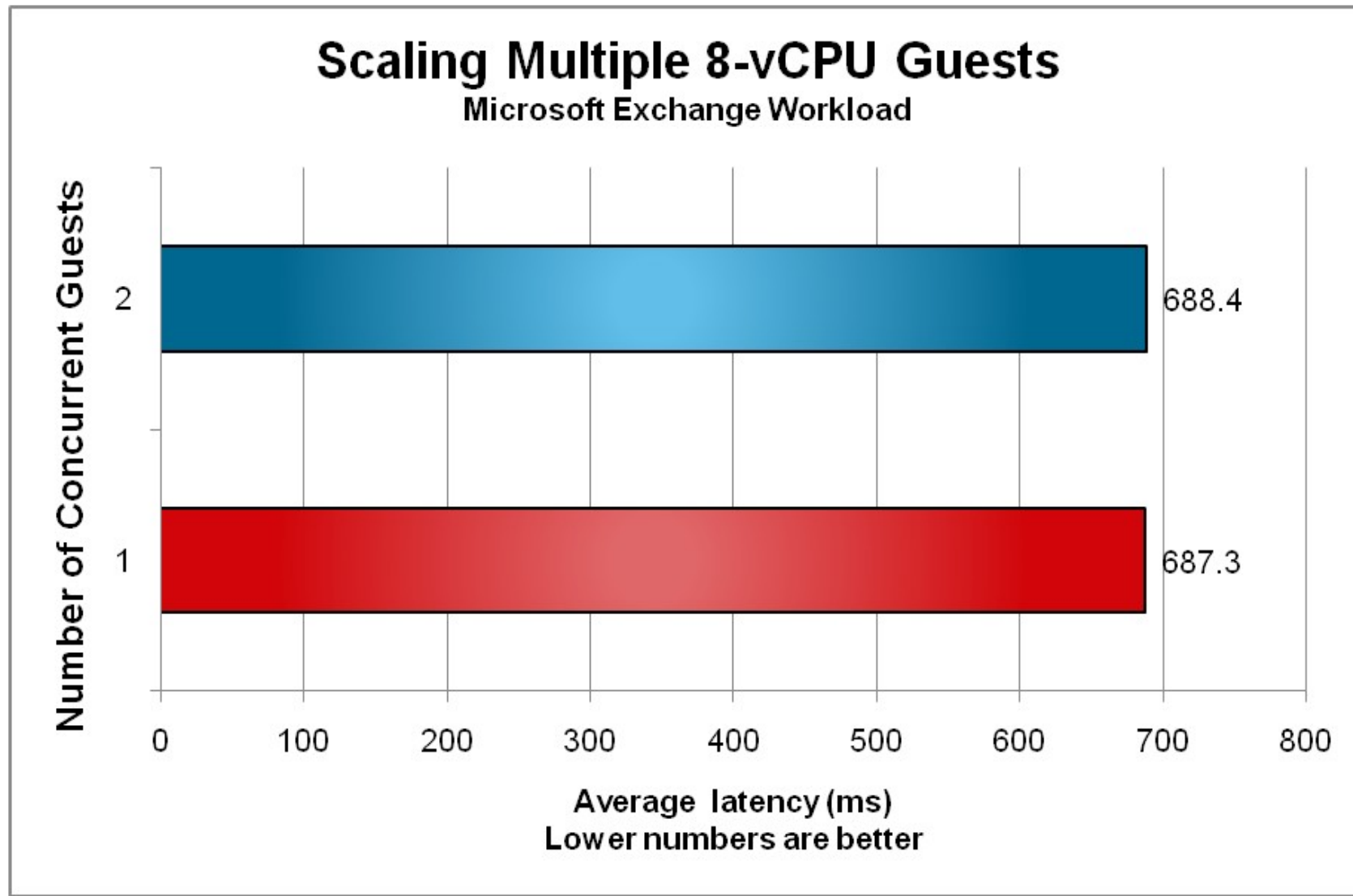
Exchange - Scaling Multiple 2-vCPU Guests



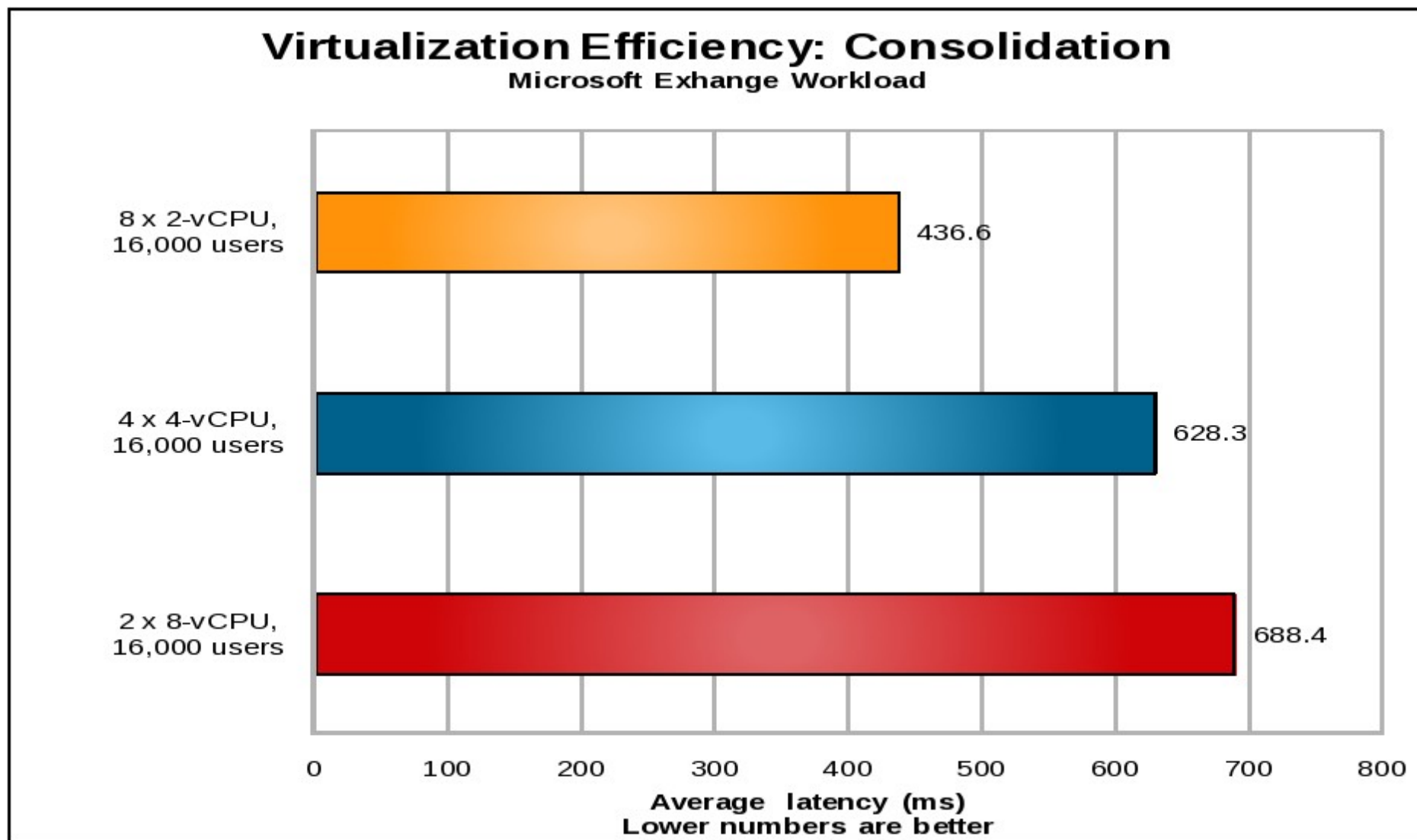
Exchange - Scaling Multiple 4-vCPU Guests



Exchange - Scaling Multiple 8-vCPU Guests



Exchange - Virtualization Efficiency: Consolidation



Conclusions & Wrap-up

KVM-based Red Hat Virtualization clearly demonstrates:

- KVM allows for “normal” Linux tuning
- Demonstrates competitive virtualization overhead
- In configurations using either horizontal-scaling (scale-out) of guests or vertical-scaling (scale-up) of guests
- Supports both Red Hat Enterprise Linux guests and Windows guests
- For a wide variety of production workloads

The continuing performance, scaling & tuning work will help Red Hat customers design virtualized environments to meet their SLA needs.

QUESTIONS?

**TELL US WHAT YOU THINK:
[REDHAT.COM/SUMMIT-SURVEY](https://redhat.com/summit-survey)**