

# Microsoft Exchange Server 2010 Performance on Red Hat Enterprise Linux 6.3 KVM Hypervisor

**Mike Sullivan - Linux Technology Center, IBM**

**Andrew Theurer - Linux Technology Center, IBM**

Version 1.0

March 2013





IBM, the IBM logo, ibm.com, System x, and X-Architecture are trademarks of **IBM Corporation** in the United States, **other countries, or both. For a list of additional IBM trademarks, please see <http://ibm.com/legal/copytrade.shtml>.**

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

## Table of Contents

<u>Introduction.....</u>	<u>4</u>
<u>1.1 Overview.....</u>	<u>4</u>
<u>1.2 The KVM hypervisor.....</u>	<u>4</u>
<u>1.3 Red Hat Enterprise Linux 6 Virtualization.....</u>	<u>5</u>
<u>1.4 IBM System x Servers.....</u>	<u>5</u>
<u>Scale-up Performance Evaluation.....</u>	<u>6</u>
<u>2.1 Scale-up Hardware Configuration.....</u>	<u>6</u>
<u>2.2 Scale-up Test Description.....</u>	<u>7</u>
<u>2.3 Scale-up Test Case Permutations.....</u>	<u>8</u>
<u>2.4 Scale-up Test Case Results.....</u>	<u>8</u>
<u>Scale-out Performance Evaluation.....</u>	<u>10</u>
<u>3.1 Scale-out Hardware Configuration.....</u>	<u>10</u>
<u>3.2 Scale-out Test Description.....</u>	<u>11</u>
<u>3.3 Scale-out Test Case Permutations.....</u>	<u>12</u>
<u>3.4 Scale-out Test Case Results.....</u>	<u>12</u>
<u>Summary.....</u>	<u>14</u>
<u>References.....</u>	<u>15</u>

# Introduction

## 1.1 Overview

Microsoft Exchange Server 2010 provides email, calendar, and contacts on your PC, phone, and web browser. To more efficiently use hardware resources, Microsoft Exchange Server is often set up to run in a virtualized environment managed by a hypervisor.

Because a mail server is a critical business application, mail server performance in a virtualized environment is an important consideration. This paper examines two key performance characteristics of running Microsoft Exchange Server 2010 under the Red Hat Enterprise Linux 6.3 KVM hypervisor:

- The performance effects of scaling up a Microsoft Exchange Server 2010 mailbox server within a single virtual machine. Scaling up entails adding users and mailboxes, while increasing the number of virtual CPUs and the amount of memory assigned to the single virtual machine created to handle them.
- The performance effects of scaling out a Microsoft Exchange Server 2010 mailbox server by adding additional virtual machines. Scaling out entails adding users while also adding additional virtual machines and resources to service them.

## 1.2 The KVM Hypervisor

A hypervisor is a specialized program that allows multiple operating systems to concurrently share a single hardware host. Each operating system is run by the hypervisor in a virtual machine that is assigned a portion of the host's physical resources. Linux's Kernel-based Virtual Machine (KVM) project represents the next generation in open-source virtualization. KVM is fully integrated into the Linux operating system.

Linux's KVM's design principles include:

- Leveraging all hardware-assisted virtualization capabilities provided by Intel Virtualization Technology (VT) and AMD Secure Virtual Machine (SVM).
- Exploiting hardware capabilities while keeping the KVM virtualization overhead to the absolute minimum
- Full integration into the Linux operating system kernel

The Linux kernel, with its 20 years of development, is the industry leader in terms of performance and availability. The Linux process scheduler, for example, provides completely fair scheduling (CFS) that is optimized to manage complex workloads and NUMA systems, while offering low latency, high-performance determinism, and fine-grained Service Level Agreement (SLA) for applications. By placing the KVM hypervisor directly into the Linux kernel, all of these services and advantages have a direct impact on hypervisor performance.

## 1.3 Red Hat Enterprise Linux 6 Virtualization

Red Hat's unique approach to virtualization is easy to adopt because it is delivered as an integral part of the Red Hat Enterprise Linux platform. Based on KVM technology, Red Hat's virtualization capabilities are integrated into Red Hat Enterprise Linux and leverage the latest in hardware virtualization that Intel and AMD processor platforms can provide. The modular design of Red Hat Enterprise Linux allows customers to choose when and where to use virtualization. For additional flexibility, customers can deploy both Red Hat Enterprise Linux and Microsoft Windows as fully supported guests within a Red Hat Enterprise Linux virtualized environment. Red Hat Enterprise Linux also supports multiple virtualization use cases, from hardware abstraction for existing software stacks and data center consolidation to virtualized clusters and private clouds.

The Red Hat Enterprise Linux 6.3 release supports up to 160 virtual CPUs (vCPUs) per virtual machine, allowing even the largest workloads to be virtualized.

## 1.4 IBM System x Servers

IBM System x servers are intelligent systems, designed to reduce costs and complexity for enterprise workloads. With the introduction of eX5,—IBM's 5th-generation industry-leading enterprise X-Architecture servers,—IBM engineers have redefined x86 servers by expanding their capabilities. A member of the eX5 server family, the x3850 X5 is a scalable, 4-socket, 4U, rack-optimized enterprise server that delivers the following benefits to enterprise customers:

- High memory capacity (up to 3TB, 3 times the memory capacity of other 4-socket x86 servers, using the industry-unique IBM MAX5 memory expansion unit)
- Processor scalability up to 8 sockets (up to 80 processor cores) by connecting two 4-socket x3850 X5 systems together and doubling all system resources (including up to 6TB of memory, using two MAX5 memory expansion units)
- The broadest range of network and storage support in the industry for ultimate flexibility and choice
- Support for IBM eXFlash solid-state storage technology for extreme storage I/O performance
- Integrated Emulex 10 GbE Virtual Fabric Adapter with capability for upgrades to Fiber Channel over Ethernet (FCoE)
- Fifth-generation eX5 chipset design enhancements, built on the latest X-Architecture blueprint
- Balanced systems for virtualization, database and enterprise workloads
- Workload-optimized systems with customizable configurations for target workloads
- Greater performance and utilization at a lower total cost
- Mainframe-inspired reliability
- Simplified power and systems management with an energy-smart design and remote access

# Scale-up Performance Evaluation

## 2.1 Scale-up Hardware Configuration

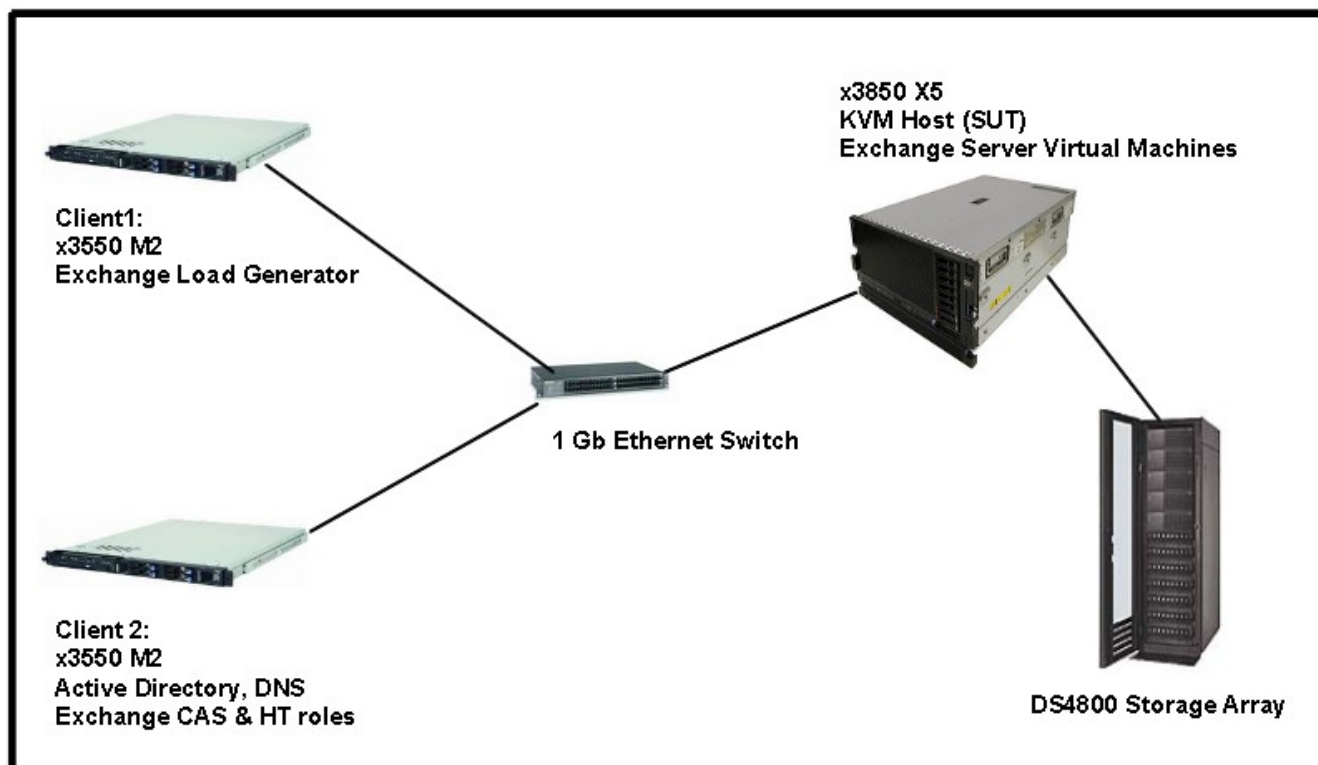


Figure 1. Scale-up test configuration

### KVM Host System Under Test (SUT)

- IBM System x3850 X5
- 4 x Intel Xeon E7-8870 10c 2.40GHz CPUs
- 1.0 TB memory
- QLogic QLE2462 Dual-Port, 4Gbps Fibre Channel-to-PCI Express Host Bus Adapter
- A single KVM virtual machine running the Microsoft Exchange Mailbox server role

### Storage

- IBM TotalStorage® DS4800
- 12 x IBM TotalStorage EXP710 storage expansion units
- 192 x 73GB 15K RPM FC HDDs

## Client 1

- IBM System x3550 M2
- 2 x Intel Xeon E5530 4c 2.40GHz CPUs
- 48GB memory
- Microsoft Windows Server® 2008 R2 Datacenter Edition
- Microsoft Exchange Load Generator 2010

## Client 2

- IBM System x3550 M2
- 2 x Intel Xeon E5530 4c 2.40GHz CPUs
- 96GB memory
- Microsoft Windows Server 2008 R2 Datacenter Edition
- Microsoft Active Directory, DNS feature
- Microsoft Exchange 2010 Client Access Server (CAS) and Hub Transport (HT) roles

## ***2.2 Scale-up Test Description***

The Microsoft Exchange Load Generator 2010 (LoadGen) is a benchmarking, stress testing tool that generates a wide range of Exchange workloads. Running on a client system, LoadGen creates a messaging load across a variety of Exchange transaction types. For the scale-up test, LoadGen was configured to simulate Outlook 2007 online clients using the “VeryHeavy” user profile. The “VeryHeavy” profile generates 150 messages per user sent and received per day. Additionally, each user mailbox was initialized with 100MB of mailbox data. The duration of the test was set at 8 hours to simulate a normal work day.

Performance metrics were collected using the Performance Monitor tool included in Windows Server 2008 R2. The key performance metric measured was the 95th-percentile sendmail transaction latency. This data represents the maximum amount of time required for sending an email for 95% of all sendmail transactions. Industry consensus is that the 95th-percentile sendmail latency should be less than 500 milliseconds to provide a satisfactory user experience. The reported value was collected for the last hour of each run after allowing the benchmark to achieve steady state.

Additionally, the mpstat tool included in RHEL 6.3 was used to collect and report CPU utilization metrics.

## 2.3 Scale-up Test Case Permutations

The mail server virtual machine was configured with one virtual CPU for each 1000 users with the “VeryHeavy” user profile within a test case permutation. Memory assigned was calculated as a base of 4GB plus 6GB for each 1000 “VeryHeavy” users supported.

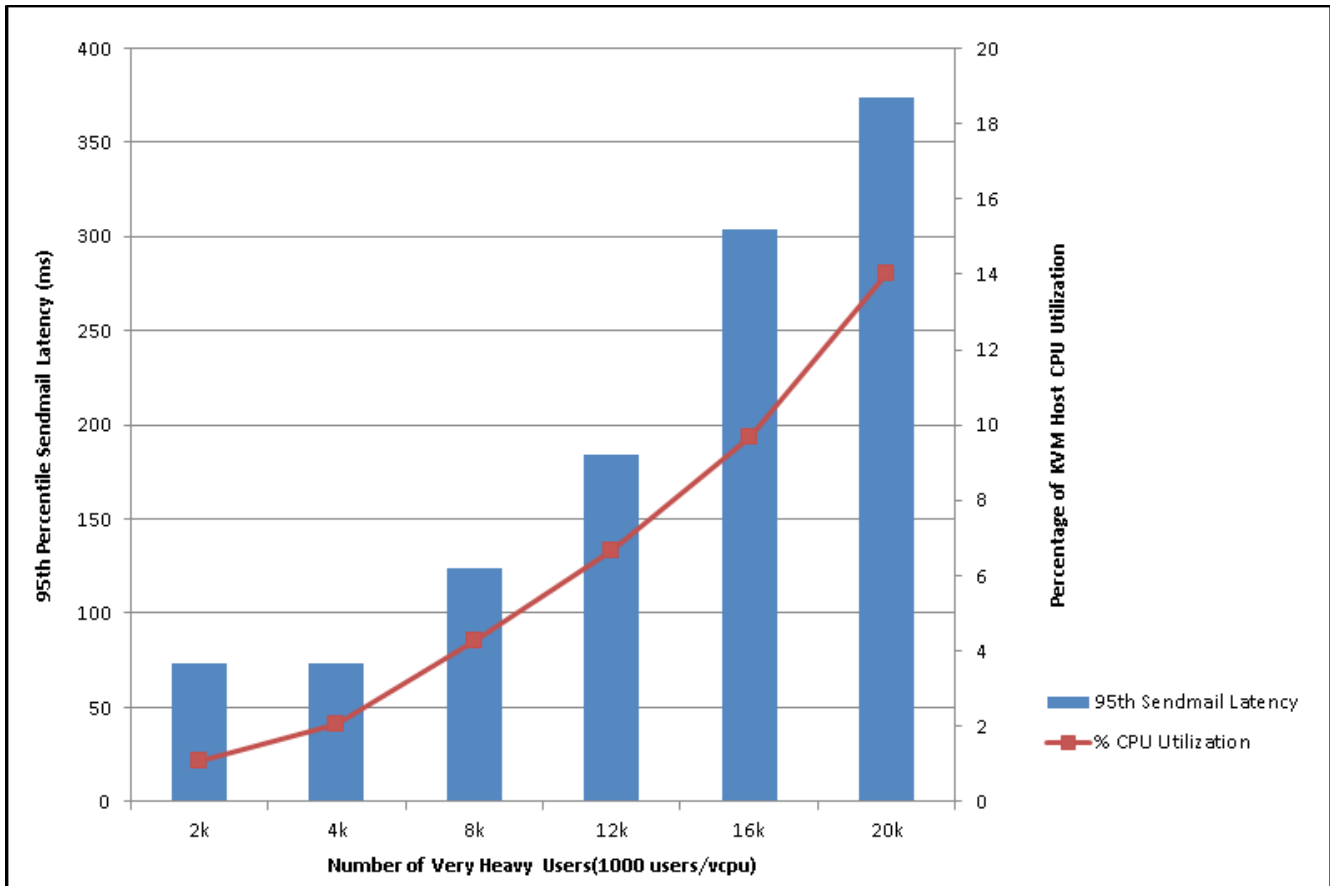
Number of “VeryHeavy” users in test	Virtual CPUs	Memory (GB)
2000	2	16
4000	4	28
8000	8	52
12000	12	76
16000	16	100
20000	20	124

Table 1. Test case permutations

## 2.4 Scale-up Test Case Results

The following results show the scale up of a single virtual machine running the Exchange Mailbox server role. Test case permutations supporting 2000 “VeryHeavy” users scaling up to 20000 “VeryHeavy” users were run. PerfMon data was collected tracking the 95<sup>th</sup> percentile for sendmail latencies. A 95th-percentile sendmail latency at or below 500 milliseconds is considered an acceptable response time to the user. CPU consumption on the KVM host as a percentage of overall available CPU available was also collected using mpstat. The chart that follows contains sendmail latency and CPU consumption information for all listed test case permutations.





**Figure 2. 95<sup>th</sup> Percentile Sendmail Latency & CPU Consumption Scale-up**

All test case permutations from 2000 up to 20000 users had 95th-percentile latencies well below the 500 millisecond threshold needed to provide a satisfactory user experience. In the permutation with the highest load, the 95th-percentile sendmail latency for 20000 users was 374 milliseconds, well below the 500 millisecond usability threshold. Additionally, this 20000 user test consumed only 14% of the x3850 x5 KVM host's available CPU hardware, leaving additional room for expansion.

# Scale-out Performance Evaluation

## 3.1 Scale-out Hardware Configuration

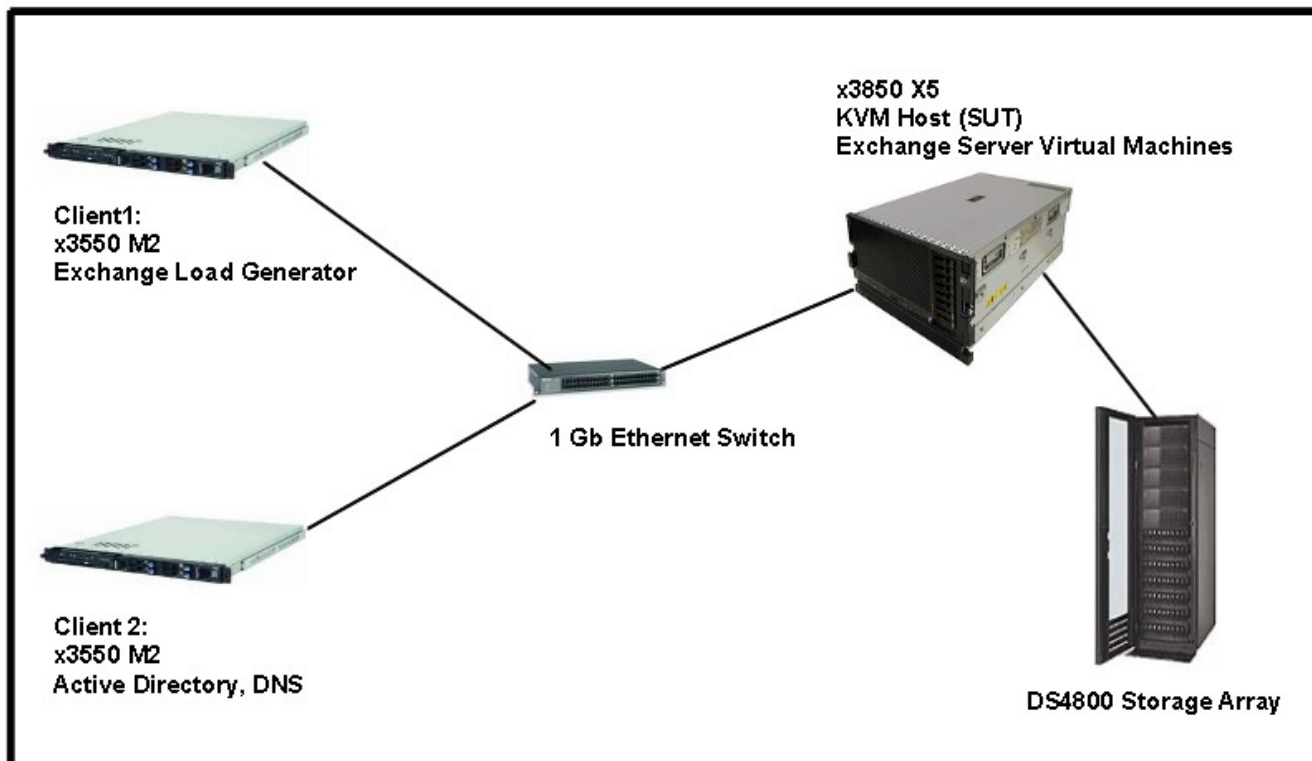


Figure 3. Scale-out test configuration

### KVM Host System Under Test (SUT)

- IBM System x3850 X5
- 4 x Intel Xeon E7-8870 10c 2.40GHz CPUs
- 1.0 TB Memory
- QLogic QLE2462 Dual-Port, 4Gbps Fibre Channel-to-PCI Express Host Bus Adapter
- Multiple KVM virtual machine pairs running Exchange Mailbox Server and CAS/HT Roles

### Storage

- IBM TotalStorage DS4800
- 12 x IBM TotalStorage EXP710 storage expansion units
- 192 x 73GB 15K RPM FC HDDs

### Client 1

- IBM System x3550 M2
- 2 x Intel Xeon E5530 4c 2.40GHz CPUs
- 48GB memory
- Microsoft Windows Server 2008 R2 Datacenter Edition
- Microsoft Load Generator 2010

#### **Client 2**

- IBM System x3550 M2
- 2 x Intel Xeon E5530 4c 2.40GHz CPUs
- 96GB memory
- Microsoft Windows Server 2008 R2 Datacenter Edition
- Microsoft Active Directory, DNS feature

## ***3.2 Scale-out Test Description***

Again, the Microsoft Exchange Load Generator 2010 was used to create the load for each test case, simulating Outlook 2007 online clients and using the “VeryHeavy” user profile. Each user's mailbox was initialized with 100MB of mailbox data. The duration of the test was set at 8 hours to simulate a normal work day.

The Performance Monitor tool was used to collect the 95th-percentile sendmail transaction latency to ensure that the maximum amount of time required for sending an email for 95% of all sendmail transactions was less than or equal to 500 milliseconds. The latency value was collected for the last hour of each run after allowing the benchmark to achieve steady state. The mpstat tool was used to collect and report CPU utilization metrics.

## ***3.3 Scale-out Test Case Permutations***

The scale-out test increased the number of virtual machines running on the physical x3850 x5 KVM host hardware. Adding virtual machines increased hardware load on CPU, memory, networking, and storage subsystems. The number of virtual machines running on the KVM host was incrementally scaled out to handle 24000 “VeryHeavy” users. For each set of 4000 users, an additional pair of virtual machines was run:

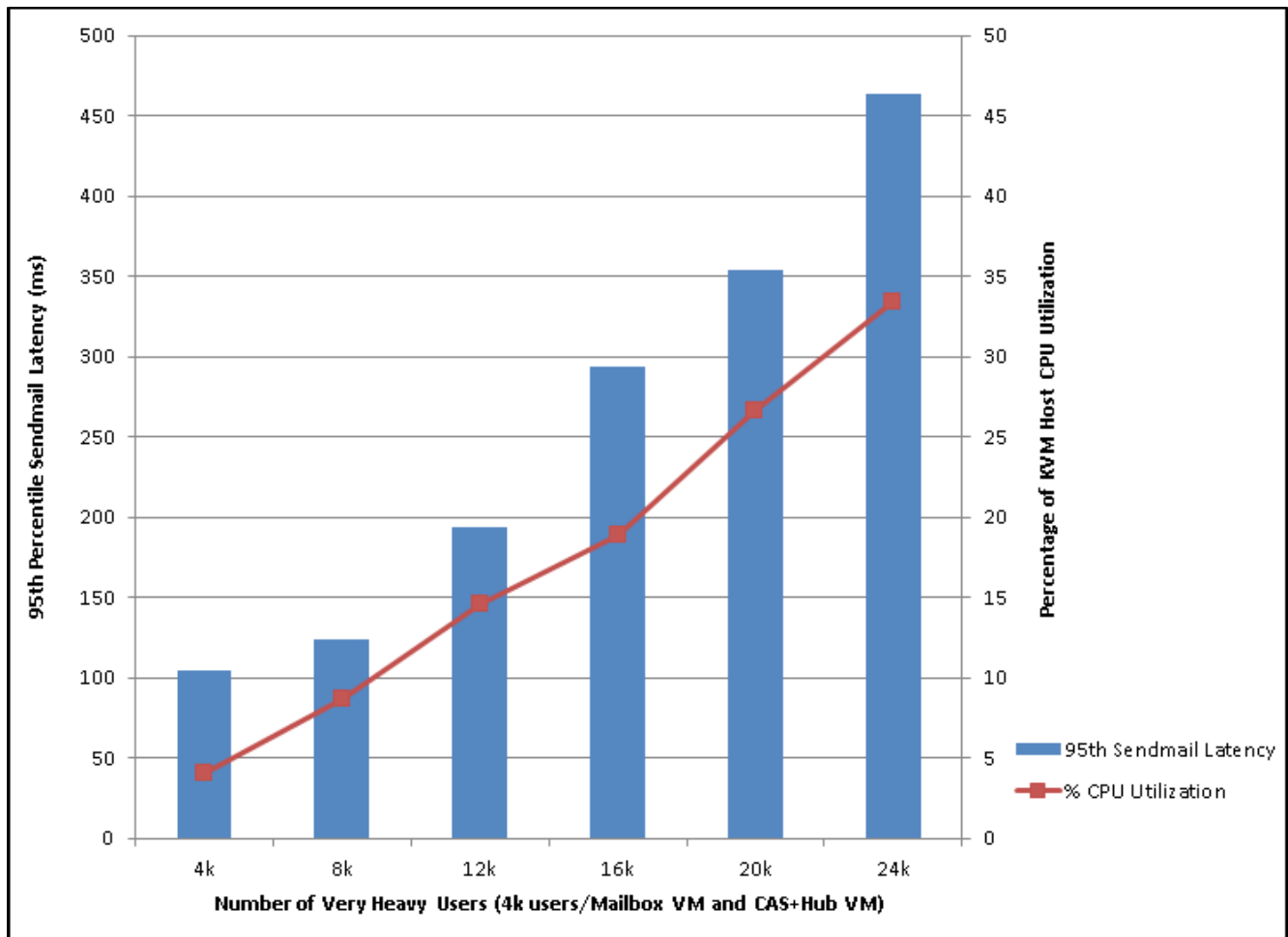
- One additional virtual machine was added to handle the Exchange Mailbox Server role. This virtual machine was configured with 4 virtual CPUs and 28GB memory.
- The second virtual machine was configured to perform both the Exchange CAS and HT roles. This virtual machine was configured with 4 virtual CPUs and 8GB memory.

Number of mailbox and CAS/HT virtual machines	Number of "VeryHeavy" users supported	Total number of virtual CPUs	Total memory (GB)
1+1	4000	8	36
2+2	8000	16	72
3+3	12000	24	108
4+4	16000	32	144
5+5	20000	40	180
6+6	24000	48	216

**Table 2. Test Case permutations**

### ***3.4 Scale-out Test Case Results***

The following chart shows the results of the scale-out tests. The 95<sup>th</sup> percentile sendmail latency and the percentage of KVM host CPU used is provided.



**Figure 4. 95<sup>th</sup> Percentile Sendmail Latency & CPU Consumption Scale-out**

The KVM host was able to scale-out to 24000 users running across 12 virtual machines, while providing a satisfactory user experience, by keeping the 95<sup>th</sup> percentile of sendmail latency to less than 500 milliseconds. Even running 12 virtual machines supporting 24000 users consumes only 1/3 of total server hardware CPU capacity.

# Summary

The virtualization technologies built into Intel processors and the integration of the KVM hypervisor into the Linux kernel have combined to provide a first class hypervisor offering. The results detailed in this paper demonstrate how Microsoft Exchange Server 2010 running virtualized under the Red Hat Enterprise Linux 6.3 KVM hypervisor provides excellent scalability while still maintaining performance:

- In the scale-up configuration, up to 20000 user mailboxes were supported by a single virtual machine. 95<sup>th</sup> percentile latencies of 374 milliseconds were maintained, well below the 500 millisecond usability threshold, while only 14% of the physical CPU hardware was used.
- For the scale-out test, 24000 users were supported by 12 virtual machines, while excellent sendmail latencies were maintained.

These results compare favorably to the results that VMware documented for similar scale-up and scale-out configurations on VMware vSphere 5 [1]. In both configurations KVM demonstrated it's ability to support more mailbox users while still staying below the latency requirement. VMware's white paper provides scale-up measurements up to a limit of only 12000 users and scale-out results up to only 16000 users and 8 virtual machines.

The performance proof points provided by the Red Hat Enterprise Linux 6.3 KVM tests demonstrate that it is an ideal virtualization platform for Microsoft Exchange Server 2010 deployments. These results demonstrate that the latest hypervisor technology available in KVM provides an opportunity for increased cost savings through mail server consolidation without sacrificing efficiency or performance.

# References

[1] VMware Inc., "Microsoft Exchange Server 2010 Performance on VMware vSphere 5.0", <http://www.VMware.com/resources/techresources/10204>, August 25<sup>th</sup>, 2011