



Comparativa de Eficiencia entre Virtualización Completa y Paravirtualización sobre Hardware Heterogéneo en Función del Rendimiento y el Consumo de Recursos

David Castro ⁽¹⁾, Patricio Villacís ⁽²⁾, Ignacio Marin-García ⁽³⁾

Facultad de Ingeniería en Electricidad y Computación.

Escuela Politécnica del Litoral

Campus Gustavo Galindo, Km 30.5 vía Perimetral

Apartado 09-01-5863, Guayaquil-Ecuador

davacast@espol.edu.ec ⁽¹⁾, pjvillacis@espol.edu.ec ⁽²⁾, imaringa@espol.edu.ec ⁽³⁾

Resumen

En el proyecto se realizó una comparativa de eficiencia entre las dos técnicas de virtualización más usadas en el mercado: la virtualización completa y la paravirtualización, en función del rendimiento y el consumo de recursos. Se utilizaron las plataformas más conocidas: VMware vSphere, Citrix XenServer y Xen sobre tres tipos diferentes de hardware: un servidor Sun Microsystems, un computador de escritorio y un servidor Hewlett-Packard. La hipótesis planteada tuvo dos niveles, el primer nivel consistió en determinar cuál de las técnicas de virtualización presentó un mejor rendimiento para cada escenario; y el segundo nivel permitió determinar la independencia del rendimiento de las técnicas sobre el hardware usado. Para llevar a cabo la comparativa se realizaron diferentes pruebas consistentes en medir el consumo de memoria RAM y procesador usando los diferentes softwares de medición para cada sistema operativo, Linux y Windows. Se crearon los escenarios para las pruebas planteadas, las mismas que se ejecutaron sobre los ambientes de trabajo individual y simultáneo, utilizando carga de uso de recursos y de tráfico con solicitudes web; siendo el número de repetición para cada una de ellas de 200. De los resultados obtenidos se realizó un análisis estadístico para determinar si la hipótesis planteada en cada caso fue cumplida.

Palabras Claves: comparativa, virtualización, paravirtualización, RAM, CPU, vSphere, XenServer, Xen..

Abstract

In this research project we conducted an efficient comparison between the two virtualization techniques more used in the market: full virtualization and paravirtualization, basing in the comparative performance and the resource consumption, on the most popular platforms: VMware vSphere, Citrix XenServer and Xen on three different types of hardware: a server Sun Microsystems, a desktop computer and a server Hewlett-Packard. The hypothesis had two levels, the first level was to determine which of the virtualization techniques presented better performance for each scenario, and the second level allowed us to determine the independence of the technical performance of the hardware used. We carried out various tests the experiments where we measured the consumption of RAM and processor using different measurement software for each operating system, Linux and Windows. Scenarios were created for testing posed, they were run on individual work environments and simultaneously, using resource usage load and traffic to web server with stress load, and the number of repetition for each 200 times. From the results obtained, we performed a statistical analysis to determine if the proposed hypothesis in each case was accomplished. From the results obtained, we performed a statistical analysis to determine if the proposed hypothesis in each case was accomplished.

Keywords: comparative; virtualization; paravirtualization; RAM; CPU; vSphere; XenServer; Xen

1. Introducción

En la actualidad el campo de la virtualización tiene un gran impacto en el diseño y funcionalidad de los centros de datos tanto empresarial como de investigación. La técnica de virtualización completa, que a lo largo de este artículo será referida como virtualización, y la técnica de paravirtualización son dos de las técnicas mayormente utilizadas [1]. Cada una de estas técnicas tiene aplicaciones específicas en función del hardware y de las necesidades de los centros de datos. En los últimos años, ambas técnicas han sido mejoradas con el fin de entregar una solución completa a los clientes, pero con eficiencias comparables entre ellas.

La principal diferencia de estas técnicas, es que la paravirtualización realiza modificaciones en los sistemas invitados, con el objetivo de alcanzar un mejor rendimiento. Esta diferencia de características influye en los sistemas que no son de código abierto, tal es el caso de Microsoft Windows o Mac OS, que no permiten dichas modificaciones. Debido a esta característica de la paravirtualización, múltiples empresas tales como VMware Inc. y Citrix Systems, Inc., han creado plataformas para la gestión de entornos virtualizados aprovechando la ventaja de la virtualización que trabaja con sistemas no modificados.

A la hora de seleccionar la técnica de virtualización adecuada, la infraestructura existente es fundamental. Debido que al tener diferentes sistemas virtualizados ocupando los mismos recursos de un computador puede provocar problemas, ya que los equipos de cómputo se ven sometidos a altas cargas de procesamiento y consumo de memoria, introduciendo penalizaciones. Para evitar estos problemas de rendimiento surge la idea de llevar a cabo un estudio comparativo de dichas técnicas de virtualización utilizando carga de uso de recursos y de tráfico con solicitudes web sobre hardware heterogéneo, para de esta forma poder correlacionar no sólo el software sino también la influencia del hardware sobre el rendimiento de la virtualización en cuanto a la memoria RAM, CPU y consumo de recursos.

Para llevar a cabo la comparativa de eficiencia entre virtualización y paravirtualización utilizamos diferentes hipervisores, tanto de distribución libre como propietaria, sobre hardware heterogéneo en

condiciones de laboratorio en función del rendimiento y el consumo de recursos. El estudio tuvo una hipótesis de dos niveles. El nivel uno, radicó en determinar cuál de las técnicas de virtualización presentó un mejor rendimiento para cada escenario, y el nivel dos, consistió en determinar la independencia del rendimiento de las técnicas sobre el hardware que se estuvo utilizando.

2. Implementación

Para el proyecto se escogieron las plataformas: VMware vSphere ESXI (virtualizador), Citrix XenServer (virtualizador) y Xen (paravirtualizador), basándonos al estudio de Tendencias de virtualización de escritorios para pequeñas y medianas empresas (PYMES) [2] realizado por Spiceworks Community, donde se indicaba que el 53% de las instalaciones de virtualización de escritorio en el sector de las PYMES se desplegaron usando soluciones de VMware Inc., y el otro 30% de instalaciones fueron soluciones Citrix Systems Inc., representando los otros proveedores únicamente el diecisiete por ciento (17%).

Para poder emitir con criterio cuál de estas dos técnicas fue la más recomendable en los diferentes entornos de trabajo que tenga una empresa, se instalaron las plataformas de virtualización en los diferentes entornos arquitectónicos como fueron: servidores y estaciones de trabajo, similares a las que se utilizan en los medios de producción, que son:

- Un servidor Sun Microsystems Sunfire X2200 que consta: de un procesador Dual-Core AMD Opteron™ con frecuencia 1.8 GHz, memoria RAM de 6GB, dos discos duros de 250 GB cada uno y una unidad óptica DVD RW.
- Un computador de escritorio que posee un procesador Intel® Core™ i7-2600 de 3.40GHz, memoria RAM de 8 GB, un disco duro de 1TB y un DVDRAM.
- Un servidor HP Proliant DL160 G6 que tiene un procesador Intel® Xeon® CPU X3430 de 2.40GHz, memoria RAM de 4 GB, un disco duro de 250 GB y una unidad óptica DVD RAM.

En el caso de los sistemas invitados se establecieron las características de hardware que fueron configuradas sobre cada plataforma en el momento de instalar las diferentes máquinas

virtuales. Se asignó para cada máquina virtual 1024 MB de memoria RAM, 16 GB de disco duro, un procesador y una tarjeta de red en modo puente.

Una vez definidos los recursos de cómputo de los sistemas anfitriones y huéspedes, tuvimos como resumen un total de nueve arquitecturas generadas por la combinación del hardware y las plataformas de virtualización, sobre cada combinación se instalaron: Fedora 15 con kernell vanille 3.6.3, CentOS 5.6 con kernell 2.6.18 o Windows Server 2008 Standard Edition, resultando un total de 27 sujetos de medición, que en los cuales se realizaron las diferentes pruebas de rendimiento.

Como sabemos la virtualización de servidores ofrece un enorme potencial de ahorro debido a la consolidación de las cargas de trabajo; esto reduce notablemente tanto el consumo de energía como la demanda de refrigeración. Para garantizar el alto desempeño y la alta disponibilidad de los servicios, la capacidad de procesamiento y la cantidad de memoria del equipo anfitrión son factores críticos al momento de adoptar esta tecnología; ya que todos estos servicios residen en un único equipo. Por lo anteriormente expuesto, las pruebas para la presente comparativa nos basamos en el rendimiento de CPU y de memoria RAM.

Las técnicas de medición de rendimiento, *benchmarks* en inglés, son pruebas que se realizan a través de un programa que evalúa las prestaciones de un sistema, con el objetivo de estimar el rendimiento de un elemento concreto para poder comparar los resultados con sistemas similares [3]. Existen paquetes especializados para cada sistema operativo, en el caso de Linux resalta *SYSTAT*, el cual consta de varias herramientas que ofrecen funcionalidades avanzadas de monitoreo del sistema [4]. Cada sistema además brinda herramientas de medición propias como en el caso de Windows que trae al *Administrador de tareas*, el mismo que proporciona información sobre los programas y procesos que se ejecutan mostrando indicadores claves de rendimiento [5].

Para el caso de Linux, tuvimos dos ambientes en los que trabajan los programas de medición, líneas de comandos e interfaz gráfica. En el ambiente de líneas de comandos se usó el programa *sar*, el cual es una de las herramientas que recoge, muestra y guarda la información de la actividad del sistema ya sea esta de CPU, memoria, discos, interrupciones, interfaces de red, tablas del núcleo, entre otras [6].

Se puede especificar parámetros para seleccionar que información acerca del sistema se desea mostrar.

En la interfaz gráfica de Linux se utilizó: el programa *hardinfo*, que permite obtener información sobre el sistema operativo, generar informes en formato HTML o texto plano de la utilización de hardware, y ejecutar ciertas pruebas benchmark [7]. Y el comando *top* que permite obtener una información en tiempo real del procesador además de proporcionar una lista de los procesos más intensos en lo referente a consumo de CPU [8].

Para la ejecución de estas herramientas se implementaron scripts personalizados, para estandarizar parámetros y automatizar la obtención de resultados.

En el caso de Windows se utilizaron: el *Administrador de tareas*, el cual nos permite ver los programas, procesos y servicios que está ejecutando el sistema. Además de supervisar el rendimiento del equipo a nivel de memoria y capacidad de procesador, ver el estado de la red y qué usuarios están conectados a la PC [5]. Y el *Monitor de rendimiento*, que nos permite obtener de forma visual contadores de rendimientos del sistema, ya sea en tiempo real o histórico [9]. Este presenta en forma más exacta el consumo de CPU, en comparación con el *Administrador de tareas*.

Con el fin de obtener datos reales a los que se obtienen en un ambiente de producción. Los diferentes sujetos de medición se sometieron a situaciones de carga de trabajo, los cuales hemos denominados: método de carga de estrés de uso de recursos y método de carga de estrés de tráfico utilizando solicitudes web, los mismos que fueron usados en dos ambientes diferentes: individual y simultaneo, como podemos observar en la figura 1.



Figura 1. Entorno de Trabajo

Una vez establecidos los requerimientos para realizar las pruebas, se planteó los diferentes escenarios que resultan de la combinación de los sujetos de medición y las pruebas. Dándonos un total de 360 escenarios de los cuales son 288 escenarios para Linux y 72 escenarios para Windows. Siendo teniendo un mínimo de repeticiones mínimo de 200 por cada prueba efectuada, según el estudio previo con los diferentes escenarios.

3. Resultados

A partir de la hipótesis planteada procedimos a realizar las comparativas sobre diferentes escenarios de pruebas. El lector encontrara a continuación los resultados promedios obtenidos. Cada figura consta de 24 barras que representan los resultados obtenidos de los escenarios presentes por sujeto de medición e interfaz utilizada. El grupo de barras en la sección izquierda representan los resultados obtenidos con el método de carga de estrés de uso de recursos y las barras en la sección derecha representan los resultados obtenidos con el método de carga estrés de tráfico utilizando solicitudes web.

Los valores que mostramos en la figura 2 a la figura 6 representan el valor promedio de consumo de memoria RAM sobre los diferentes sujetos de medición.

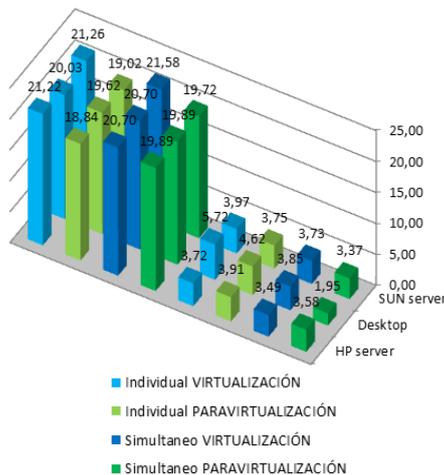


Figura 2: Porcentaje de consumo de memoria RAM obtenidos en CentOS utilizando la Interfaz de Línea de Comandos

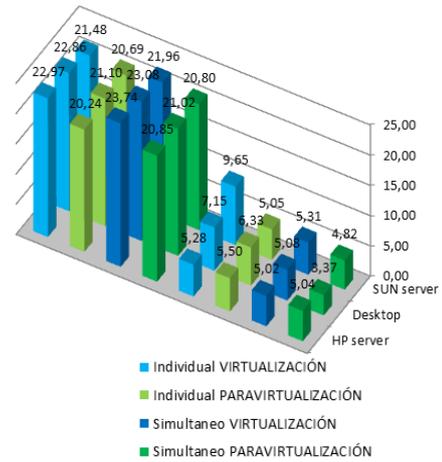


Figura 3: Porcentaje de consumo de memoria RAM obtenidos en CentOS utilizando la Interfaz de Grafica de Usuario

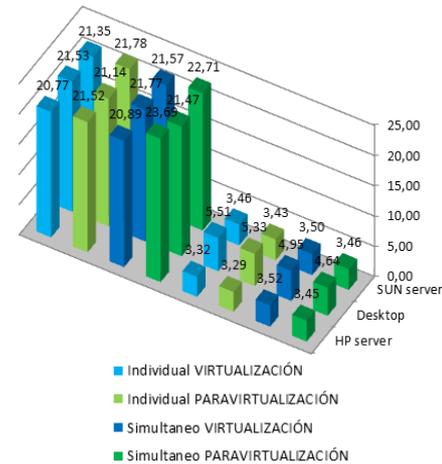


Figura 4: Porcentaje de consumo de memoria RAM obtenidos en Fedora utilizando la Interfaz de Línea de Comandos

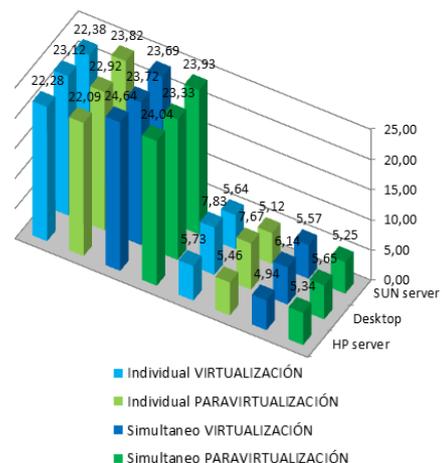


Figura 5: Porcentaje de consumo de memoria RAM obtenidos en Fedora utilizando la Interfaz de Grafica de Usuario

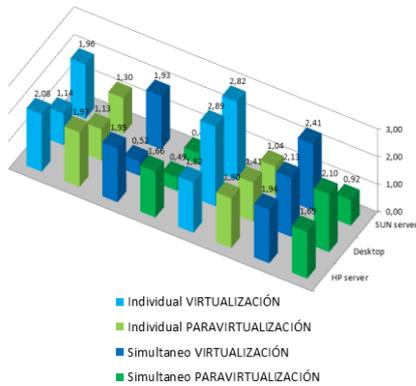


Figura 6: Porcentaje de consumo de memoria RAM obtenidos en Windows

A continuación los valores que mostramos de la figura 7 a la figura 11 representan el valor promedio de procesador sobre los diferentes sujetos de medición.

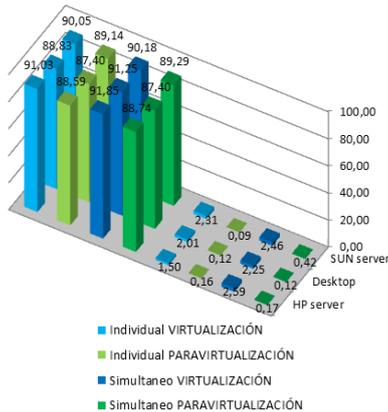


Figura 7: Porcentaje de consumo de procesador obtenidos en CentOS utilizando la Interfaz de Línea de Comandos

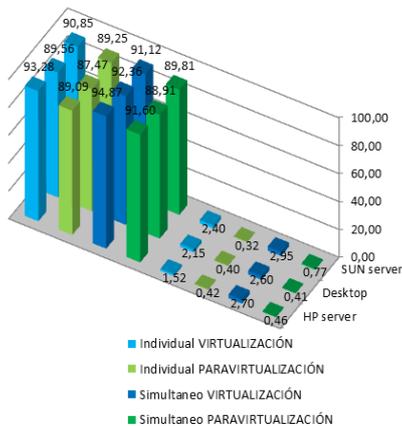


Figura 8: Porcentaje de consumo de procesador obtenidos en CentOS utilizando la Interfaz de Gráfica de Usuario

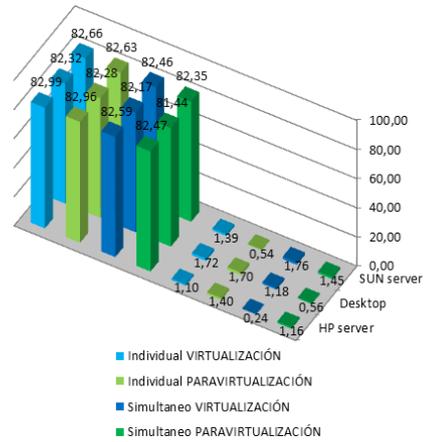


Figura 9: Porcentaje de consumo de procesador obtenidos en Fedora utilizando la Interfaz de Línea de Comandos

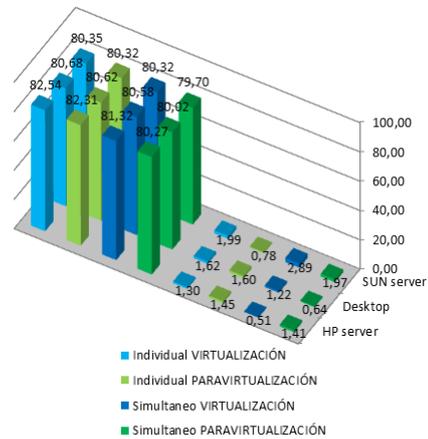


Figura 10: Porcentaje de consumo de memoria RAM obtenidos en Fedora utilizando la Interfaz de Gráfica de Usuario

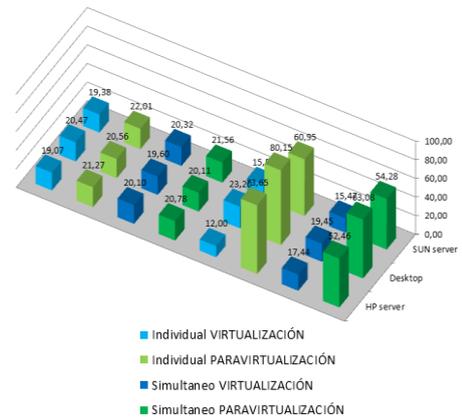


Figura 11: Porcentaje de consumo de memoria RAM obtenidos en Windows



4. Conclusiones.

La técnica de paravirtualización presenta un mejor consumo de recursos de cómputo y un mejor rendimiento cuando se usa un núcleo compatible con XEN, como en el caso de CentOS (figuras 2, 3, 7, 8). Esto indica que la paravirtualización debería ser la opción referente cuando la infraestructura de la empresa y las aplicaciones soportan este núcleo.

Como se observó en las figuras anteriores, el rendimiento y el consumo de memoria RAM en el sistema Linux tienen un comportamiento similar independientemente de la técnica utilizada. Sin embargo, el sistema Windows tiene un mayor consumo de memoria RAM y presenta un mejor desempeño para la técnica de paravirtualización,

El rendimiento de consumo de CPU depende del hardware. Según el equipo, la virtualización y la paravirtualización podrían utilizar más recursos.

5. Recomendaciones

Para futuras aplicaciones del presente proyecto se deberá tomar en cuenta nuevas variables de medición y escenarios acorde a las necesidades de la implementación.

Revisar los requerimientos mínimos para implementar los escenarios aquí expuestos, y de esta manera obtener los diferentes comportamientos esperados.

5. Agradecimiento

Este proyecto no sería posible sin el apoyo de nuestro director, Ignacio Marín-García y nuestra coordinadora de carrera, Patricia Chávez-Burbano, quien nos ofreció valiosa colaboración, apoyo y orientación en nuestra primera investigación real.

6. Referencias

- [1] Almería Eugenio Villar y Julio Gómez, “Virtualización de Servidores de Telefonía IP en GNU/Linux”, 2010.
- [2] VMware Inc, Trends around Desktop, “Virtualization for Small and Mid-sized Organizations”, 2012.
- [3] Caracterización de la Carga: Benchmarks. [Online]. Disponible en: <http://geneura.ugr.es>

- [4] SYSSTAT Howto: A Deployment and Configuration Guide for Linux Servers. [Online]. Disponible en: <http://www.linux.com>
- [5] Microsoft Inc. “Task Manager Overview”. [Online]. <http://www.microsoft.com>
- [6] SAR Manual Page. [Online]. <http://sebastien.godard.pagesperso-orange.fr>
- [7] HARDINFO. [Online]. <http://www.linuxlinks.com>
- [8] LINUX / UNIX Command: TOP. [Online]. <http://linux.about.com>
- [9] Overview of Windows Performance Monitor. [Online]. <http://technet.microsoft.com>