



Universitat
de les Illes Balears

Evaluación de Comportamiento de Sistemas Informáticos

Práctica 2 - Evaluación y modelado del rendimiento de los sistemas informáticos

Alberto Pérez Ancín

Professores:

Dr. Carlos Juíz García

Dra. Belén Bermejo González



APLICACIÓN PRÁCTICA

1.1 Objetivo

El objetivo principal es la evaluación práctica, real y completa de un sistema informático. Para ello se aplicarán todos los conceptos, metodologías y técnicas vistos a lo largo de todo el curso. Se tratarán desde los aspectos relacionados con la monitorización y el benchmarking, pasando por el modelado y llegando finalmente a la predicción de la carga.

Un ejemplo real del camino expuesto se ve en el artículo “The Performance Evaluation Journey of a Flight Seats Availability Service: A Real-world Business Case Study of Transactional Workload Running in Virtual Machines”, el cual se recomienda encarecidamente que se tenga a disposición y se comprenda en su totalidad.

1.2 Segunda parte

Requisitos previos

El alumno ya contará con el entorno Ubuntu (o el que haya considerado) instalado, siendo totalmente funcional. Para asegurarnos de que el entorno está totalmente listo, se deberá poder acceder al directorio /proc. Además, se contará ya con una herramienta para filtrar y limpiar los ficheros de datos como para realizar representaciones gráficas.

En esta parte no se tendrá en cuenta qué tipo de actividad está realizando el sistema mientras se realiza la monitorización de este. Antes de empezar a responder las diferentes partes, se recomienda probar los monitores, sus filtros, el volcado de ficheros y su tratamiento.

Monitorización de la CPU

En esta primera parte, se pide monitorizar la CPU durante 2 horas haciendo uso del monitor TOP. El intervalo de muestreo será de 5 segundos. Los datos obtenidos (ÚTILES) deberán ser guardados en un fichero de salida para posteriormente tratarlos y responder a las siguientes preguntas.

El fichero de salida será en formato Excel (o csv) el cual tendrá el siguiente formato:

Timestamp	% CPU (global)	% CPU (user)	%CPU (system)
...
...

1. **¿Cuántas CPUs tiene el sistema que se ha monitorizado? ¿De dónde se ha obtenido esa información?**

El sistema monitorizado tiene 8 CPU(s), ya que al ejecutar en la terminal el comando `lscpu | grep 'CPU(s):'` como se puede comprobar en la imagen inferior nos muestra cuantas CPUs tiene el sistema que se ha monitorizado.

```
alberto@aperez:~$ lscpu | grep 'CPU(s):'
CPU(s): 8
```

2. **¿Cuál es la utilización media de la CPU en modo usuario, sistema y en global?**

Para monitorizar la CPU durante 2 horas con el monitor TOP y un intervalo de muestreo de 5 segundos, he usado el siguiente comando en la terminal:

```
alberto@aperez:~$ timeout 2h top -d 5 -b > TOP.csv
```

Este comando ejecutará el monitor TOP en modo batch (-b), con un intervalo de muestreo de 5 segundos (-d 5) y se ejecutará durante 2h (timeout 2h), lo que equivale a 1440 iteraciones. El comando timeout detendrá la ejecución del monitor TOP después de 2 horas y los datos se guardarán en un archivo llamado TOP.csv.

Luego con el comando `grep 'Cpu(s)' top.csv >MonitorizacionCPUSegundoEjercicio.txt` busco todas las líneas en el archivo TOP.csv que contengan la cadena de texto "Cpu(s)" y las guarda en un archivo llamado MonitorizacionCPUSegundoEjercicio.txt para poder calcular la utilización media de la CPU en modo usuario (%us), sistema (%sy) y en global (100 - %id).

CPU en modo usuario

⇒ Utilizando en Excel la fórmula: =PROMEDIO a la columna %us

CPU en modo sistema

⇒ Utilizando en Excel la fórmula: =PROMEDIO a la columna %sy

CPU en modo global

⇒ Primero hago 100 - el resultado de cada valor de la columna %id

⇒ Obtengo una nueva columna con el valor de CPU en modo global.

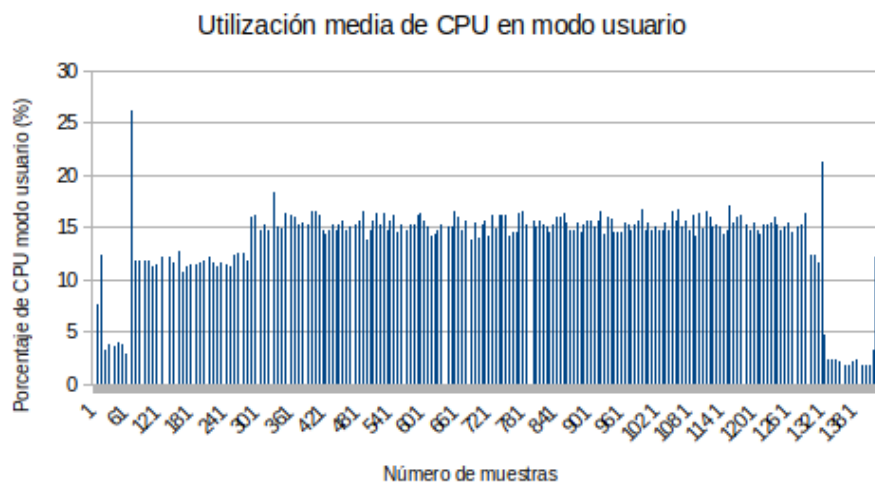
⇒ Utilizando en Excel la fórmula: =PROMEDIO a la columna nueva generada

Con todos estos datos obtenidos, después de haber realizado los procedimientos mencionados anteriormente, los resultados de la utilización media de la CPU en modo usuario, sistema y en global se muestran en la siguiente tabla:

UTILIZACIÓN MEDIA EN PORCENTAJES (%)		
CPU en modo usuario	CPU en modo sistema	CPU en modo global
13,37%	1,6%	15,29%

3. **¿Cómo se comportan las medidas anteriores a lo largo del tiempo de observación? Muestra las tres métricas de forma gráfica.**

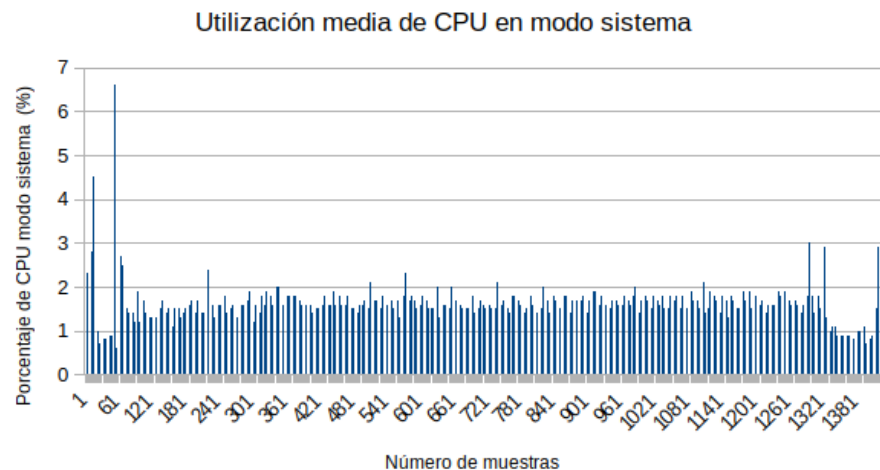
Respecto a como se comportan las medidas anteriores, nos encontramos en primer lugar la utilización media de CPU en modo usuario, en segundo lugar la utilización media de CPU en modo sistema y por último la utilización media de CPU en modo global.

Utilización media de CPU en modo usuario

1. APLICACIÓN PRÁCTICA

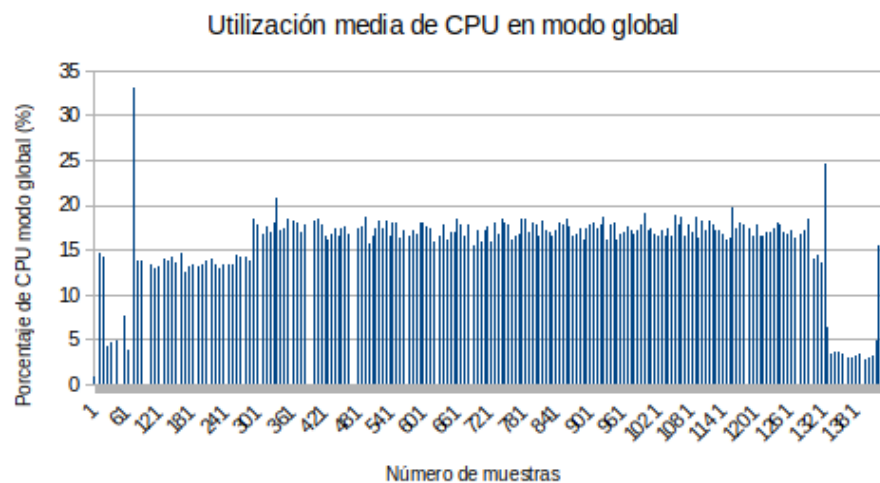
Como podemos observar, el porcentaje de CPU en modo usuario por lo general se mantiene bastante constante en el tiempo sobre el 15%, menos al principio y al final del muestreo que disminuye considerablemente el porcentaje de uso respecto al 15%. Además, añadir que el valor máximo de una muestra supera el 25% mientras que el mínimo ronda los 2,5% de utilización media de CPU en modo usuario.

Utilización media de CPU en modo sistema



Como podemos contemplar, el porcentaje de CPU en modo sistema por lo general se mantiene bastante constante en el tiempo sobre el 2%, menos al principio y al final del muestreo que disminuye llegando a valores incluso inferiores al 1%. Además, añadir que el valor máximo de una muestra llega a un 4% de utilización media de CPU en modo sistema.

Utilización media de CPU en modo global



Como podemos apreciar, el porcentaje de CPU en modo global por lo general se mantiene bastante constante en el tiempo sobre el 17 %, menos al principio y al final del muestreo que disminuye llegando a valores incluso inferiores o rondando al 5 %. Además, añadir que el valor máximo de una muestra sobrepasa el 30 % de utilización media de CPU en modo global.

4. ¿Cuál es la sobrecarga provocada por el monitor TOP?

Podemos calcular la sobrecarga del monitor utilizando la información proporcionada por el monitor TIME al ejecutar el monitor TOP para una sola muestra.

```
alberto@aperez:~$ time top -b -d 1 -n 1 | grep -i "Cpu(s)"
%Cpu(s):  0,8 us,  1,5 sy,  0,0 ni, 97,7 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
real    0m0,165s
user    0m0,003s
sys     0m0,013s
```

La sobrecarga se determina dividiendo el tiempo que tarda el monitor en realizar una muestra por el tiempo de muestreo.

$$\text{Sobrecarga} = \frac{0,165s}{5s} \times 100 = 3,3 \%$$

Monitorización de la memoria principal

En esta segunda parte, se pide monitorizar la memoria principal del sistema durante 2 horas haciendo uso del monitor VMSTAT con un intervalo de muestreo de 15 segundos. Los datos obtenidos (ÚTILES) deberán ser guardados en un fichero de salida para posteriormente tratarlos y responder a las siguientes preguntas.

El fichero de salida será en formato Excel (o csv) el cual tendrá el siguiente formato:

Timestamp	Capacidad disponible	Capacidad utilizada	% Memoria utilizada
...
...

1. ¿Qué capacidad total tiene la memoria principal del sistema? ¿De dónde se ha obtenido ese dato?

La memoria principal del sistema tiene una capacidad total de 7760228 kB como se puede observar en la imagen inferior, he podido obtener esta información gracias a la primera línea resultante al ejecutar el comando `more /proc/meminfo`.

```
alberto@aperez:~$ more /proc/meminfo
MemTotal:      7760228 kB
MemFree:       341500 kB
MemAvailable:  1893952 kB
```

2. ¿Cuál es la utilización media de la memoria? ¿Y la capacidad media utilizada?

El comando `vmstat 15 480 > VMSTAT.csv` ejecuta el comando `vmstat` con un intervalo de 15 segundos entre cada actualización y un total de 480 actualizaciones. Los resultados de cada actualización se redirigen al archivo `VMSTAT.csv` utilizando el operador `>`.

```
alberto@aperez:~$ vmstat 15 480 > VMSTAT.csv
```

Esto significa que el comando `vmstat` se ejecutará durante un total de $480 * 15 = 7200$ segundos (2 horas) y registrará información sobre el uso de la memoria, la CPU y otros recursos del sistema en el archivo `VMSTAT.csv`.

Ahora que ya tengo los datos procedo a obtener la información que me interesa mediante el uso del siguiente comando:

```
alberto@aperez:~$ cat VMSTAT.csv | awk '{print $4}' > MemoriaLibre.csv
```

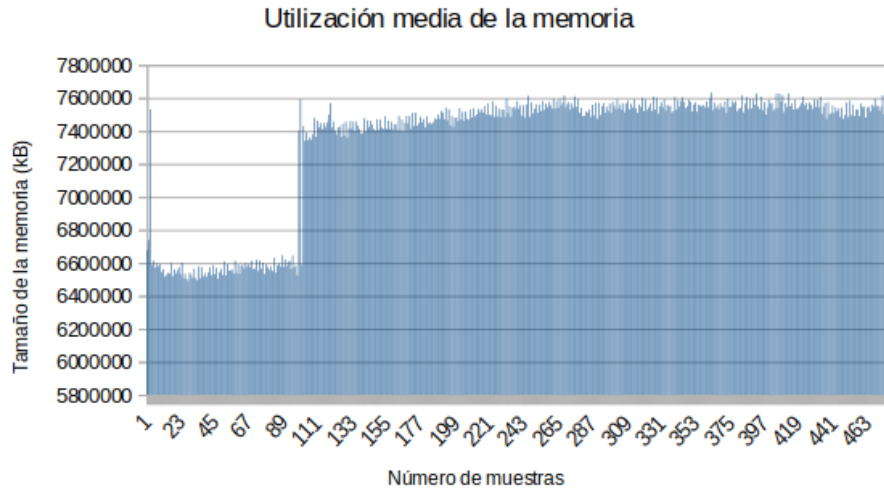
En el último archivo generado por el comando `vmstat`, se registran todos los valores de la memoria libre que ha medido el monitor. Para calcular la cantidad de memoria utilizada en cada momento, debemos restar el valor de la memoria libre al total de la memoria principal del sistema. Con este valor podemos calcular tanto la utilización media como la capacidad media utilizada de la memoria a lo largo del tiempo durante el que se ejecutó el monitor. Aclarar que la capacidad la calculo diviendo cada valor de memoria libre entre la capacidad total de la memoria principal y lo multiplico por 100 para obtener su correspondiente porcentaje.

Utilización media de la memoria (kB)	Capacidad media utilizada (%)
7324660,84 kB	94,39 %

3. ¿Cómo se comporta la utilización de la memoria y la capacidad utilizada? Representa estas métricas gráficamente.

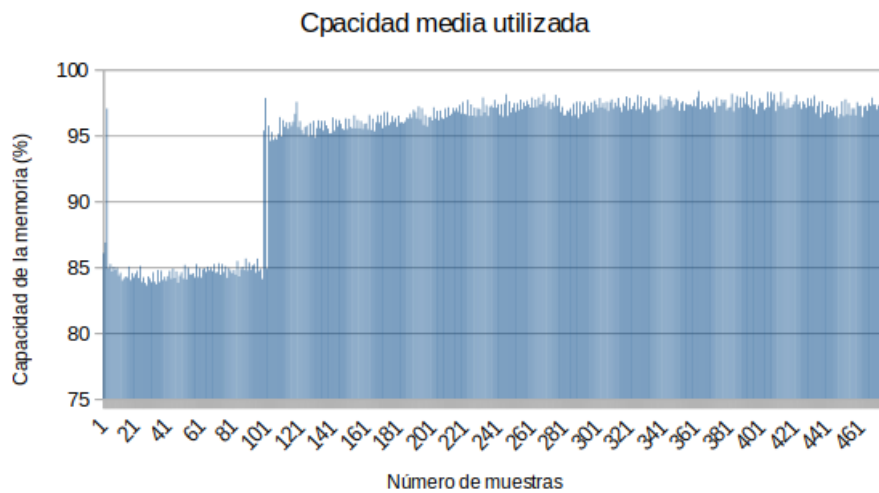
Respecto a como se comportan las medidas anteriores, nos encontramos en primer lugar la utilización media de la memoria y en segundo lugar la capacidad media utilizada.

Utilización media de la memoria



Como podemos contemplar, el tamaño de la memoria utilizado por lo general se mantiene bastante constante en el tiempo sobre 7400000 kB, menos al principio del muestreo que disminuye llegando a valores incluso inferiores al 6600000 kB. Además, añadir que el valor máximo de una muestra llega a 7600000 kB de utilización media de la memoria.

Capacidad media utilizada



Como podemos contemplar, el porcentaje de capacidad media utilizada por lo general se mantiene bastante constante en el tiempo sobre el 95 %, menos al principio del muestreo que disminuye llegando a valores incluso inferiores al 85%. Además, añadir que el valor máximo de una muestra llega a un 98 % de capacidad media utilizada.

4. ¿Cuál es la sobrecarga provocada por el monitor VMSTAT?

Podemos calcular la sobrecarga del monitor utilizando la información proporcionada por el monitor TIME al ejecutar el monitor VMSTAT para una sola muestra.

```
alberto@aperez:~$ time vmstat 1 1 -n
procs -----memoria----- --swap-- -----io---- -sistema-- -----cpu-----
r  b   swpd  libre búfer caché  si   so    bi    bo   in   cs us sy id wa st
1  0 196864 541792 121492 3622548    0    1    19    47   223 111 3  1 96  0  0

real    0m0,010s
user    0m0,002s
sys     0m0,000s
```

La sobrecarga se determina dividiendo el tiempo que tarda el monitor en realizar una muestra por el tiempo de muestreo.

$$\text{Sobrecarga} = \frac{0,010s}{15s} \times 100 = 0,07\%$$

Pregunta final

¿Qué hubiera pasado si los dos monitores (TOP y VMSTAT) se hubieran lanzado en paralelo?

Si los dos monitores (TOP y VMSTAT) se hubieran lanzado en paralelo, es posible que se hubiera producido una sobrecarga en la CPU y en la memoria principal del sistema.

Esto se debe a que ambos monitores están diseñados para recopilar información sobre el uso de recursos del sistema, lo que significa que ambos estarían compitiendo por los mismos recursos.

Por lo tanto, es probable que si se lanzan en paralelo, la CPU y la memoria principal del sistema puedan estar sobrecargadas, lo que puede ralentizar el sistema y hacer que sea menos eficiente. En general, es mejor lanzar los monitores de forma secuencial o en diferentes momentos, para evitar la sobrecarga del sistema y obtener mediciones más precisas.

¿Cómo variaría el cálculo de la sobrecarga? Exprésalo de forma matemática

Si los dos monitores (TOP y VMSTAT) se ejecutan en paralelo para monitorizar la memoria principal y la CPU, es posible que la sobrecarga del sistema aumente debido a la necesidad de ejecutar ambos monitores al mismo tiempo. En este caso, la sobrecarga del sistema sería el resultado de tener ambos monitores consumiendo recursos simultáneamente. Por lo tanto, se podría decir que la sobrecarga final sería el resultado de sumar las sobrecargas individuales de cada monitor.

$$\text{Sobrecarga} = \text{Sobrecarga}_{TOP} + \text{Sobrecarga}_{VMSTAT}$$