
TEMA 5

Optimización del rendimiento de un servidor mediante análisis operacional

PROBLEMA 5.1 Un servidor dedicado a la reserva de billetes de una compañía de ferrocarriles tiene un tiempo medio de respuesta de 15 milisegundos por petición. Si por término medio hay 345 peticiones activas en el servidor, ¿cuál es su productividad? Expresé el resultado en peticiones por segundo.

SOLUCIÓN: La productividad del servidor es de 23000 peticiones por segundo. ■

PROBLEMA 5.2 El servidor web de una compañía comercial ha servido 120 peticiones durante el último minuto. Si el número medio de peticiones activas ha sido de 5, calcule el tiempo medio de respuesta experimentado por una petición al servidor web.

SOLUCIÓN: El tiempo medio de respuesta es de 2,5 segundos. ■

PROBLEMA 5.3 El computador de una empresa conservera se utiliza de manera interactiva por los 45 empleados que tiene en plantilla. Si el tiempo medio de reflexión es de 17 segundos y su productividad de 2,5 peticiones por segundo, ¿cuál es el tiempo medio de respuesta de cada interacción?

SOLUCIÓN: El tiempo medio de respuesta de cada interacción es de 1 segundo. ■

PROBLEMA 5.4 Durante una sesión de medida de media hora un monitor software ha extraído las siguientes variables operacionales básicas de un disco duro:

Variable	Valor
A (arrivals)	364 peticiones
C (completions)	359 peticiones
B (busy time)	23 minutos

A partir de la información anterior calcule las siguientes variables operacionales deducidas del disco duro: tasa de llegada, productividad, utilización y tiempo medio de servicio. No olvide expresar las unidades de cada variable.

SOLUCIÓN: $\lambda = 12,13$ peticiones/minuto, $X = 11,97$ peticiones/minuto, $U = 0,77$ y $S = 3,84$ segundos/petición. ■

PROBLEMA 5.5 El sistema informático dedicado a la atención al cliente de una empresa de comestibles se puede modelar mediante una red de colas abierta con un procesador y dos unidades de disco. El comportamiento de cada petición se asemeja al modelo de servidor central. Se sabe que los tiempos de respuesta (expresados en milisegundos) y las razones de visita de estas estaciones son los siguientes:

Estación	V_i	R_i
Procesador	7	4,3
Disco 1	2	1,5
Disco 2	4	2,3

Determine el tiempo medio de respuesta de una petición a este sistema informático. Si el número medio de peticiones activas en el sistema es 80, ¿cuál es la tasa de llegadas que soporta?

SOLUCIÓN: El tiempo de respuesta del sistema es 42,3 ms y la tasa de llegadas es 1,89 peticiones/ms. ■

PROBLEMA 5.6 El subsistema de disco de un servidor dedicado a comercio electrónico se ha monitorizado durante 120 segundos. El monitor ha permitido saber que, de este tiempo, el subsistema ha estado activo durante 78 segundos; además, se han producido 84 peticiones de acceso, de las cuales se han servido 82. Según esta información determine:

1. La productividad y la utilización del subsistema de disco.

Por otro lado, se sabe que cada interacción con el servidor provoca un número medio de 5 visitas al subsistema de disco y que el número medio de peticiones activas en el servidor es de 13. A partir de estos datos calcule:

2. La productividad del servidor y el tiempo medio de respuesta de una interacción con el servidor.

SOLUCIÓN:

1. Respecto del subsistema de disco: el error es del 2,4 %, la productividad es 0,68 peticiones/segundo y la utilización 0,65.
 2. Respecto del servidor: la productividad es 0,137 peticiones/segundo y el tiempo medio de respuesta 95,1 segundos.
-

PROBLEMA 5.7 El sitio web de una empresa dedicada a productos de deporte recibe una media de 450 visitas por minuto. De todas estas visitas únicamente el 20% hace un pedido de material en firme. Cada uno de estos pedidos se procesa en un servidor dedicado mediante un script escrito en PHP y requiere, por término medio, una demanda de servicio del procesador de 0,6 segundos.

- a) Calcule la utilización media del procesador debida al procesamiento de pedidos.
- b) ¿Cuál sería la nueva utilización del procesador si un nuevo diseño del programa PHP permite mejorar su tiempo de ejecución (el de la CPU) 2,5 veces?
- c) ¿Cuál sería la nueva utilización del procesador si utilizáramos un viejo diseño del programa PHP con un tiempo de ejecución por parte del procesador 2 veces mayor? ¿Qué podríamos concluir en ese caso sobre el funcionamiento del servidor?

SOLUCIÓN: a) La utilización del procesador es 0,9. b) Con la mejora introducida la nueva utilización es de 0,36. c) En ese caso la utilización sería 1 (100%) ya que nunca puede ser mayor que ese valor y el servidor estaría saturado. ■

PROBLEMA 5.8 Un servidor web recibe, por término medio, 4 peticiones por segundo. El comportamiento de las peticiones se asemeja al modelo del servidor central. Los tiempos de servicio y de respuesta (expresados en segundos), así como las razones de visita a los dispositivos de este servidor se indican en la siguiente tabla:

Dispositivo	V_i	S_i	R_i
Procesador (1)	8	0,01	0,0147
Disco (2)	4	0,04	0,1111
Disco (3)	3	0,03	0,0469

A partir de la información anterior determine:

- La demanda de servicio de cada dispositivo (D_i).
- El tiempo de respuesta del servidor web (R_0).
- El número medio de peticiones en el servidor web (N_0).
- La productividad de cada dispositivo (X_i).
- La utilización de cada dispositivo (U_i).

SOLUCIÓN:

- $D_1 = 0,08$, $D_2 = 0,16$, $D_3 = 0,09$ (expresados en segundos o, más concretamente, segundos/petición).
- $R_0 = 0,7027$ segundos.
- $N_0 = 2,8108$ peticiones.
- $X_1 = 32$, $X_2 = 16$, $X_3 = 12$ (expresados en peticiones/s).
- $U_1 = 0,32$, $U_2 = 0,64$, $U_3 = 0,36$. ■

PROBLEMA 5.9 Un determinado servidor web que consta esencialmente de un procesador y un disco duro tiene una productividad máxima de 25 peticiones por segundo. Un monitor software instalado en el mismo ha permitido conocer que la demanda de servicio del procesador es de 0,02 segundos; sin embargo, un problema de compatibilidad binaria ha impedido medir la demanda de servicio del disco, el cual parece estar dando problemas de congestión. ¿Podría indicar cuánto vale esta demanda?

SOLUCIÓN: La demanda de servicio del disco es 0,04 segundos/petición por ser éste el cuello de botella. ■

PROBLEMA 5.10 Durante un tiempo T , se encuentran haciendo uso de un servidor de ficheros un total de 3000 clientes, cada uno asociado a un único fichero (1 cliente = 1 fichero = 1 trabajo de nuestro modelo). Suponiendo que el tiempo medio de reflexión de cada cliente es de 20 segundos y que el tiempo medio de respuesta del servidor es de 10 segundos por cada fichero:

- ¿Cuál es la productividad media del servidor y cuántos trabajos se encuentran, de media, en reflexión?
- Si se quiere conseguir una productividad de 125 trabajos por segundo, ¿qué tiempo de respuesta debería tener el servidor?
- ¿Qué habría que hacer para conseguir una productividad de 200 trabajos/s?

SOLUCIÓN: a) $X_0 = 100 \text{tr/s}$, $N_z = 2000 \text{tr}$. b) $R_0 = 4 \text{s}$. c) Tal y como está diseñado el servidor (con ese número de

clientes y ese tiempo medio de reflexión), no se puede alcanzar una productividad media de 200tr/s. En el caso límite de que $R_0=0s$, X_0 sería 150tr/s. ■

PROBLEMA 5.11 Consideremos un sistema informático interactivo (=servidor + clientes) con un procesador y tres unidades de disco. Los tiempos de servicio y razones de visita de estos dispositivos se muestran en la siguiente tabla:

Dispositivo	Razón de visita	Tiempo de servicio (s)
Procesador (1)	7	0,1
Disco (2)	3	0,025
Disco (3)	1	0,050
Disco (4)	2	0,035

Sabiendo que el número de clientes en todo el sistema informático es de 10, el tiempo medio de reflexión es de 6 segundos, y que la productividad del servidor es de 1,2 trabajos por segundo:

- Calcúlense las demandas de servicio de cada dispositivo.
- ¿Cuál es el número medio de clientes (=trabajos) que están en reflexión?
- ¿Cuántos clientes están conectados de media en el servidor?
- ¿Cuál es el tiempo de respuesta del servidor?
- Calcúlense, para cada dispositivo del servidor, la productividad y la utilización.

SOLUCIÓN:

- $D_1 = 0,7$, $D_2 = 0,075$, $D_3 = 0,05$, $D_4 = 0,07$ (expresados en segundos o segundos/trabajo).
- Hay 7,2 trabajos (clientes) en reflexión.
- En el servidor hay una media de 2,8 clientes conectados.
- El tiempo medio de respuesta es 2,3 segundos.
- Productividades: $X_1 = 8,4$, $X_2 = 3,6$, $X_3 = 1,2$, $X_4 = 2,4$ (expresados en trabajos/s). Utilizaciones: $U_1 = 0,84$, $U_2 = 0,09$, $U_3 = 0,06$, $U_4 = 0,08$. ■

PROBLEMA 5.12 Los parámetros del modelo de un sistema informático transaccional (red abierta) son los siguientes (los tiempos se expresan en milisegundos):

Dispositivo	S_i	V_i
Procesador (1)	0,4	9
Disco (2)	0,5	8

La tasa de llegadas al sistema es de 0,15 transacciones por milisegundo.

- Identifique el cuello de botella del sistema.
- ¿Cuál es la utilización del cuello de botella?
- Calcule la productividad máxima del sistema.
- Determine el tiempo mínimo de respuesta de una transacción.

SOLUCIÓN: El cuello de botella del sistema es el disco y su utilización es 0,6. La productividad máxima del sistema es 0,25 transacciones/ms y el tiempo mínimo de respuesta de una transacción es 7,6 ms. ■

PROBLEMA 5.13 Considere la siguiente parametrización del modelo de un sistema informático interactivo con 25 clientes (suponga un trabajo por cliente) y un tiempo medio de reflexión de 6 segundos (los tiempos de la tabla se expresan en segundos):

Dispositivo	S_i	V_i
Procesador (1)	0,5	4
Cinta (2)	0,75	3

- Identifique el cuello de botella.
- Determine el tiempo mínimo de respuesta del servidor.
- ¿Cuál es el punto teórico de saturación (*knee point*)? A la vista de su valor, ¿el servidor se encuentra sometido a baja o alta carga?
- Indique las ecuaciones de los límites optimistas del tiempo de respuesta y de la productividad.

SOLUCIÓN:

- El cuello de botella es la cinta.
- El tiempo mínimo de respuesta es 4,25 segundos.
- El punto teórico de saturación es $N_T^* = 4,5$. Por lo tanto, el servidor está en un régimen de alta carga.
- Los límites son $R_0 \geq \max\{4,25, 2,25 \times N_T - 6\}$ y $X_0 \leq \min\{N_T/10,25, 0,44\}$.

■

PROBLEMA 5.14 Considere la siguiente información referida al modelo de un sistema informático donde los tiempos se expresan en milisegundos:

Dispositivo	S_i	V_i
Procesador (1)	0,5	29
Disco (2)	0,3	13
Disco (3)	2,4	15

El sistema recibe una media de 18 peticiones por segundo durante el mediodía, que corresponde al segmento de horario con mayor actividad. Calcule:

- El tiempo mínimo de respuesta de una petición para cualquier valor de carga.
- El tiempo medio de respuesta de cada estación suponiendo que $W_i = N_i \times S_i$.
- El tiempo medio de respuesta del sistema.
- La mejora obtenida en el tiempo medio de respuesta del sistema si se substituye el disco más lento por uno idéntico al rápido.

SOLUCIÓN:

- El tiempo mínimo de respuesta absoluto es 54,4 ms.
- $R_1 = 0,68$, $R_2 = 0,32$, $R_3 = 6,82$ (expresado en ms).
- El tiempo medio de respuesta del sistema es $R_0 = 126,1$ ms.
- El tiempo medio de respuesta mejora 4,39 veces.

■

PROBLEMA 5.15 El equipo de informáticos de una gran empresa tiene dos alternativas para implementar el subsistema de discos de la base de datos a la que se accede a través de una página web: un único disco con tiempo de servicio de 0,03 segundos, o tres discos idénticos con tiempo de servicio de 0,09 segundos. Cada petición recibida en el servidor web genera, de media, 36 solicitudes al subsistema de discos.

- a) Demuestre numéricamente qué alternativa de las dos anteriores podrá conseguir una mayor productividad media del servidor suponiendo que:
- Suposición 1: Las visitas se reparten equitativamente entre los tres discos en la segunda configuración.
 - Suposición 2: El disco es el dispositivo cuello de botella en el caso de la primera configuración.
- b) ¿A qué conclusión podríamos llegar si no se cumpliera la segunda de las suposiciones?

SOLUCIÓN: a) En el primer caso la demanda del disco es $36 \times 0,03 = 1,08$ segundos; en el segundo, la demanda de cualquiera de los tres discos es $12 \times 0,09 = 1,08$ segundos. Como ningún otro dispositivo tiene una demanda de servicio mayor (ya que el resto de dispositivos es idéntico en ambas configuraciones) también los discos son los cuellos de botella en la segunda configuración. Por tanto, en ambos casos la productividad máxima del servidor viene determinada por la inversa de la demanda de servicio de los discos, es decir, 0,93 tr/s. Por tanto, las dos opciones ofrecen la misma productividad máxima.

b) En ese caso, otro dispositivo sería el cuello de botella, pero sería exactamente el mismo dispositivo en ambas configuraciones por lo que, nuevamente, las dos opciones ofrecerían la misma productividad máxima. ■

PROBLEMA 5.16 Considere que en el supuesto del problema anterior el procesador del servidor web tiene un tiempo de servicio de 0,01 segundos y una razón de visita de 37. Si el servidor web recibe una media de 0,5 peticiones por segundo determine, para cada configuración del sistema de discos, la siguiente información sobre las prestaciones del servidor web:

1. Cuello de botella.
2. Productividad máxima.
3. Tiempo mínimo de respuesta.
4. Tiempo medio de respuesta suponiendo que $W_i = N_i \times S_i$.

Atendiendo al tiempo medio de respuesta, ¿cuál es la mejor opción? ¿Qué mejora se consigue?

SOLUCIÓN:

- En el primer caso el cuello de botella es el disco. La productividad máxima es 0,926 peticiones por segundo. El tiempo mínimo de respuesta es 1,45 s, y su valor medio es 2,8 s.
 - En el segundo caso los tres discos actúan como cuellos de botella. La productividad máxima tiene el mismo valor que el caso anterior. El tiempo mínimo de respuesta es 3,61 s y su valor medio es 7,5 s.
 - Atendiendo al tiempo de respuesta es preferible disponer de un único disco tres veces más rápido que tres discos lentos; la mejora conseguida en este índice es 2,7. ■
-

PROBLEMA 5.17 Un sistema interactivo con 30 clientes (suponga un trabajo por cliente) y un tiempo medio de reflexión de 12 segundos se modela mediante los siguientes parámetros (los tiempos se expresan en segundos):

Dispositivo	S_i	V_i
CPU (1)	0,01	11
Disco (2)	0,05	3
Disco (3)	0,08	7

Determine:

- El cuello de botella del servidor.
- La productividad máxima y el punto teórico de saturación (*knee point*).
- Los límites optimistas del tiempo de respuesta y de la productividad.
- Utilice solvenet para calcular el número medio de trabajos en reflexión suponiendo que $R_i(n_T) = (N_i(n_T - 1) + 1) \times S_i$.

SOLUCIÓN:

- El cuello de botella es el segundo disco (dispositivo número 3).
- La productividad máxima es 1,79 peticiones/s y $N_T^* = 23$ clientes.
- Los límites son $R_0 \geq \max\{0,82, 0,56 \times N_T - 12\}$ y $X_0 \leq \min\{N_T/12,82, 1,79\}$.
- 20,98 trabajos.

■

PROBLEMA 5.18 Los límites optimistas del tiempo de respuesta y de la productividad de un sistema informático interactivo son:

$$R_0 \geq \max\{0,49, 0,22 \times N_T - 5\}$$

$$X_0 \leq \min\left\{\frac{N_T}{5,49}, 4,55\right\}$$

Considere que el tiempo se expresa en segundos. A partir de la información anterior indique:

- Tiempo de reflexión.
- Tiempo mínimo de respuesta del servidor.
- Punto teórico de saturación (*knee point*).
- ¿Cuál es el tiempo medio de respuesta del servidor que se podría esperar con 100 clientes (suponga un trabajo por cliente)? Justifique la respuesta.
- Si hubiera 35 clientes en el sistema, ¿sería posible obtener un tiempo de respuesta de aproximadamente 0,49 segundos? ¿Por qué?

SOLUCIÓN:

- El tiempo de reflexión es $Z = 5$ segundos.
- El tiempo mínimo de respuesta es 0,49 segundos.
- El punto teórico de saturación es $N_T^* = 25$ clientes.
- El tiempo medio de respuesta esperable es de poco más de 17 segundos.
- No, porque al haber más clientes que el límite teórico de saturación, ya seguro que hay colas en, al menos, el cuello de botella por lo que nunca se podría conseguir el tiempo de respuesta mínimo que precisamente es de 0,49 segundos.

■

PROBLEMA 5.19 Considere el sistema informático por lotes ($Z = 0$) con 15 trabajos modelado con los siguientes parámetros (los tiempos se expresan en segundos):

Dispositivo	S_i	V_i
Procesador (1)	0,016	5
Disco (2)	0,02	4

Utilizando SOLVENET, justifique numéricamente qué opción de las dos que se indican a continuación representa la mejor elección para mejorar el rendimiento del sistema: substituir el procesador por uno dos veces más rápido o substituir el disco por otro con un tiempo medio de servicio de 0,01 segundos.

SOLUCIÓN: Ambas opciones sobre el sistema ofrecen el mismo grado de mejora en el rendimiento del sistema. Se trata de un sistema equilibrado (todas los dispositivos tienen la misma demanda de servicio). Se debe comprobar mediante SOLVENET que tanto el tiempo de respuesta del servidor, como su productividad como los límites asintóticos son los mismos con ambas mejoras. Nótese que, sin embargo, el punto teórico de saturación (knee point) es menor en ambos casos que el del sistema original sin mejora.

■

PROBLEMA 5.20 Considere un servidor web que recibe una media de 0,3 peticiones por segundo y es modelado con los siguientes parámetros (los tiempos de la tabla se expresan en segundos):

Dispositivo	S_i	V_i
CPU (1)	0,2	15
DiscoA (2)	0,07	6
DiscoB (3)	0,02	8

Después de apurar su copa de vino, una informática avezada en temas de modelado y evaluación de rendimiento hace estas confesiones a sus compañeros de cena respecto del modelo anterior (suponga que $W_i = N_i \times S_i$):

- Si se sustituye el procesador por otro dos veces y media más rápido, el tiempo medio de respuesta del servidor web mejora más del 1100% (es decir, la mejora en velocidad es mayor del 1100%).
- Si se reestructura el contenido de los dos discos hasta conseguir igualar sus demandas de servicio (=equilibrar sus cargas), entonces el tiempo medio de respuesta del servidor web mejora menos del 1% (es decir, la mejora en velocidad es menor del 1%).

¿Ha afectado la ingesta de alcohol la mente despierta de nuestra protagonista? Justifique numéricamente la respuesta.

SOLUCIÓN: Las dos predicciones son correctas. En el primer caso la mejora en velocidad obtenida es 12,2 (1120%) (1114% si usamos todos los decimales en los cálculos) y en el segundo 1,006 (0,6%). Las nuevas razones de visita de los discos A y B serían, en el segundo caso, 3,1 y 10,9 respectivamente.

■

PROBLEMA 5.21 Los usuarios del sitio web del grupo Pink Floyd se han quejado formalmente al administrador (*webmaster*) debido a los altos tiempos de respuesta que experimentan al acceder a sus contenidos. Incluso, afirman, a veces el servidor se muestra incapaz de responder y no tienen más remedio que cerrar el navegador. El

administrador aduce en su defensa que, tras analizar los datos ofrecidos por *sar*, la utilización del disco duro, que resulta ser el cuello de botella, es del 80% para una productividad del disco de 38 accesos por segundo. Además, la razón de visita del disco duro es muy baja, aproximadamente 1,5. Según él, hay margen de sobra para poder llegar a más de 40 peticiones al sitio web por segundo, valor suficiente para atender a todos los usuarios del sitio; el problema, añade, está en la configuración de los navegadores de los usuarios. ¿Tiene base científica la afirmación del administrador? Justifique la respuesta.

SOLUCIÓN: El administrador no tiene razón porque la productividad máxima que soporta el sitio web es de 31,7 peticiones por segundo. ■

PROBLEMA 5.22 El informático responsable de la instalación de una empresa dedicada a la venta de billetes de avión de bajo coste ha modelado el servidor web que atiende a los clientes utilizando técnicas del análisis operacional. Este modelo comprende el procesador y dos discos; los parámetros relevantes del mismo se muestran a continuación:

Dispositivo	Tiempo de servicio (s)	Razón de visita
Procesador (1)	0,01	9
Disco (2)	0,02	4
Disco (3)	0,02	4

En las horas de máxima actividad el sitio web ha llegado a recibir una media de 11 peticiones por segundo. Responda a las siguientes cuestiones justificando numéricamente la respuesta.

- Desde el punto de vista del reparto de la carga entre los componentes del servidor web, ¿estamos ante un servidor equilibrado?
- Indique si el servidor está sometido a alta o baja carga.
- ¿Cuál es el tiempo mínimo de respuesta de este servidor web?
- ¿Podríamos aumentar la productividad máxima (capacidad) del servidor si sustituimos ambos discos por versiones más rápidas?
- Determine el tiempo medio de respuesta del servidor suponiendo que $W_i = N_i \times S_i$.
- ¿Cambiaría la localización del cuello de botella si la tasa de llegada bajara hasta las 5 peticiones por segundo?
- Dibuje una gráfica en la que se represente la evolución del tiempo medio de respuesta del servidor en función de la tasa de llegada de trabajos. Indique los puntos más representativos de la curva.

SOLUCIÓN:

- El servidor está cerca del equilibrio porque las demandas son parecidas: $D_1 = 0,09$ y $D_2 = D_3 = 0,08$ segundos/petición.
- La carga del servidor es elevada porque el cuello de botella está cerca de la saturación.
- El tiempo mínimo de respuesta es 0,25 segundos.
- No porque el cuello de botella es el procesador.
- El tiempo medio de respuesta es 10,3 segundos.
- No porque la disminución de la carga afecta a todo el sistema, no solamente al cuello de botella. ■

PROBLEMA 5.23 El sistema informático de una empresa, al que se conectan unos 32 clientes de media (suponga un trabajo por cliente), parece que tiene problemas para soportar la carga actual. El administrador ha calculado los siguientes límites optimistas del tiempo de respuesta y de la productividad:

$$R_0 \geq \max\{1,6, 1,1 \times N_T - 4\}$$

$$X_0 \leq \min\left\{\frac{N_T}{5,6}, 0,91\right\}$$

- El sistema, ¿está realmente soportando una carga elevada?
- Haga una estimación del tiempo medio de respuesta del servidor en las condiciones actuales.

SOLUCIÓN:

- La carga es elevada porque el punto teórico de saturación (knee point) está situado en torno a los 5 clientes y en el sistema hay 32 clientes conectados.
- El tiempo medio de respuesta del servidor estará situado ligeramente por encima de los 31,2 segundos (parte derecha del límite optimista).

■

PROBLEMA 5.24 El proceso de modelado de un servidor de base de datos mediante técnicas de análisis operacional ha dado los siguientes parámetros:

Dispositivo	Tiempo de servicio (s)	Razón de visita
Procesador (1)	0,15	6
Disco (2)	0,05	5

El servidor recibe una media de 1,05 peticiones por segundo. Suponiendo que $R_i = (N_i + 1) \times S_i$, responda a las siguientes cuestiones justificando numéricamente la respuesta.

- Indique si el servidor está sometido a alta o baja carga.
- ¿Cuál es el número medio de trabajos en la cola del procesador?
- Calcule el tiempo medio de respuesta del servidor.
- ¿Tendría algún efecto sobre las prestaciones sustituir el procesador por una versión más rápida?
- Determine cuál sería el cuello de botella del servidor si el procesador y el disco se sustituyen, respectivamente, por versiones 5 y 2 veces más rápidas.

SOLUCIÓN:

- La carga es alta porque la tasa de llegadas está cerca de su valor máximo de 1,11 peticiones por segundo.
- En la cola del procesador hay una media de 16,24 trabajos.
- El tiempo medio de respuesta es de 16,7 segundos.
- Sí porque el procesador es el cuello de botella del servidor.
- En este caso el cuello de botella seguirá siendo el procesador.

■

PROBLEMA 5.25 Un sistema interactivo con 3 clientes (suponga un trabajo por cliente) y un tiempo medio de reflexión de 5 segundos se modela mediante los siguientes parámetros (los tiempos se expresan en segundos):

Dispositivo	S_i	V_i
Procesador (1)	0,01	15
Disco (2)	0,04	8
Disco (3)	0,08	6

Determine, sabiendo que la productividad del servidor es 0,49 trabajos/s:

- El tiempo medio de respuesta del servidor.
- Las utilidades de cada dispositivo.

SOLUCIÓN:

- $R_0=1,1s$.
- $U_1=0,07$, $U_2=0,16$, $U_3=0,24$.

■

PROBLEMA 5.26 Un servidor web no saturado recibe, por término medio, 4 peticiones de páginas web por segundo. Los tiempos de servicio (expresados en segundos), así como las razones de visita a los dispositivos de este servidor web se indican en la siguiente tabla:

Dispositivo	Tiempo de servicio (s)	Razón de visita
Procesador (1)	0,01	8
Disco duro (2)	0,04	4
Red (3)	0,03	3

A partir de la información anterior:

- Calcule la demanda de servicio, la productividad y la utilización de cada dispositivo.
- Determine el tiempo mínimo posible de respuesta del servidor web. Justifique la respuesta.
- ¿Qué dispositivo es el cuello de botella del servidor y por qué? ¿Qué valor tendría que tener la tasa de llegadas para que el cuello de botella fuese otro dispositivo? Desde el punto de vista del reparto de la carga entre los componentes del servidor web, ¿estamos ante un sistema equilibrado?
- Calcule la productividad máxima del servidor web. ¿Qué tiempo de servicio debería tener el dispositivo cuello de botella para obtener el doble de esa productividad máxima? Razone la respuesta.
- Suponiendo que $R_i=(N_i+1) \times S_i$ para cada dispositivo, calcule el tiempo de respuesta del servidor web.
- Calcule el número medio de peticiones en el servidor web. ¿Cómo se llama la ley que ha utilizado?

SOLUCIÓN:

- $D_1=0,08$ s/petición al servidor, $X_1= 32$ trabajos/s, $U_1=0,32$; $D_2=0,16$ s/petición al servidor, $X_2= 16$ trabajos/s, $U_2=0,64$; $D_3=0,09$ s/petición al servidor, $X_3= 12$ trabajos/s, $U_3=0,36$;
- $R_{0min}=0,33s$. Se produce cuando no hay esperas en la cola de ningún dispositivo por lo que los tiempos

- de respuesta individuales coinciden con los tiempos de servicio.
- c) El disco ya que es el que tiene mayor demanda de servicio y, por tanto, mayor utilización. La tasa de llegadas no influye en cuál es el dispositivo cuello de botella. El sistema no está equilibrado ya que el disco presenta una utilización muy superior al del resto de dispositivos.
 - d) $X_{0\max}=6,25$ peticiones/s. Podría pensarse que el tiempo de servicio del disco debería ser la mitad, pero eso haría que el cuello de botella dejase de ser el disco para pasar a ser la red. Es la red la que entonces limitaría la productividad máxima y ésta nunca podría superar las 11,1 peticiones/s.
 - e) $R_0=0,7$ s/petición.
 - f) $N_0=2,81$ peticiones. Hemos usado la ley de Little.

■

PROBLEMA 5.27 Considere la siguiente parametrización del modelo de un servidor de apuestas deportivas interactivo con 25 clientes en total conectados (suponga un trabajo por cliente) y un tiempo medio de reflexión de 6 segundos (los tiempos de la tabla se expresan en segundos):

Dispositivo	Tiempo de Servicio	Razón de Visita
CPU	0,5	4
Red	0,75	3

- a) Explique el significado de cada una de las variables que aparecen en las siguientes expresiones y obtenga su valor (no olvide las unidades).

$$X_0 \leq \min \left\{ \frac{N_T}{D+Z}, \frac{1}{D_b} \right\}$$

- b) ¿Cuál es el punto teórico de saturación (*knee point*) del servidor? A la vista de su valor, ¿el sistema se encuentra sometido a baja o alta carga?

SOLUCIÓN:

$N_T = 25$ clientes; $D=4,25$ s/petición al servidor (suma de demandas de servicio = tiempo mínimo de respuesta); $Z=6$ s (tiempo medio de reflexión de los clientes); $D_b=2,25$ s/petición al servidor (demanda de servicio del cuello de botella, en nuestro caso: la red). $X_0 \leq 0,44$ peticiones al servidor por segundo (productividad del servidor). Si quisiéramos obtener un valor más exacto de X_0 deberíamos resolver la red de colas mediante Solvenet aunque se puede estimar su valor teniendo en cuenta que el sistema se encuentra en alta carga, ya que $N_T^*=4,6$ clientes. Como $N_T \gg N_T^*$ el servidor se encuentra sometido a alta carga. Por lo tanto, X_0 será un valor próximo, aunque inferior, a 0,44 peticiones/s.

■

PROBLEMA 5.28 Un ingeniero informático pretende modelar el servidor de base de datos que está administrando utilizando un modelo basado en redes de colas. Para ello, ha monitorizado el servidor durante 24 horas, contabilizando un total de 15000 peticiones externas al servidor. Durante ese tiempo, el monitor *sar* le ha indicado que el procesador ha estado ocupado un total de 800 minutos y ejecutado 60000 procesos, mientras que se han realizado un total de 135000 accesos al disco duro, habiendo éste trabajado un total de 1200 minutos. Suponiendo que el servidor no está saturado:

- a) Calcule la razón de visita, el tiempo de servicio, la productividad y la utilización tanto del procesador como del disco duro.
- b) ¿Cuál es la productividad máxima que puede alcanzar este servidor? ¿Y el tiempo de respuesta mínimo?

- c) Suponiendo que $W_i = N_i \times S_i$, ¿cuál es el tiempo de respuesta actual de los trabajos que llegan al servidor? ¿y el nº medio de trabajos en la cola de cada dispositivo?

SOLUCIÓN:

- a) $V_{cpu}=4$; $V_{disco}=9$; $S_{cpu}=0,8$ s; $S_{disco}=0,53$ s; $X_{cpu}=0,69$ trabajos/s; $X_{disco}=1,56$ trabajos/s ;
 $U_{cpu}=0,56$; $U_{disco}=0,83$;
- b) $X_{0max}=0,21$ trabajos/s; $R_{0min}=8,0$ s
- c) $R_0=36$ s; $Q_{cpu}=0,69$ trabajos; $Q_{disco}=4,17$ trabajos;

■

PROBLEMA 5.29 Los parámetros del modelo de un servidor de comercio electrónico (red abierta) son los siguientes:

Dispositivo	Tiempo de Servicio (s)	Razón de Visita
CPU (1)	0,025	8
HDD (2)	0,050	9

La tasa de llegada al servidor es de 1,5 transacciones por segundo.

- a) Identifique el cuello de botella y calcule la productividad máxima del servidor.
- b) ¿Cuál es la utilización de la CPU?
- c) ¿Cuál sería dicha utilización si la tasa de llegada fuese de 3 transacciones/s?
- d) ¿Cuál sería ahora la productividad máxima del servidor si añadiéramos dos discos duros idénticos al actual suponiendo que la carga se repartiera equitativamente entre los tres discos?

SOLUCIÓN:

- a) El HDD ya que su demanda de servicio (0,45s) es mayor que la de la CPU (0,2s). $X_{0max}=2,22$ tr/s.
- b) $U_{cpu}=0,3$ (30%)
- c) En ese caso el servidor estaría saturado. La utilización de la CPU jamás podría ser mayor que el 44% en este servidor.
- d) X_{0max} ahora = 5 tr/s (el cuello de botella pasaría a ser la CPU)

■

PROBLEMA 5.30 Durante las últimas 24 horas, se ha monitorizado un servidor de base de datos no saturado con el fin de obtener un modelo del mismo basado en redes de colas. Como resultado de dicha monitorización, se han obtenido las siguientes medidas:

- Se han contabilizado un total de 54000 consultas al servidor.
 - La utilización de la unidad SSD es del 60%.
 - Cada consulta al servidor requiere una media de 5 accesos a la unidad SSD.
- a) Calcule cuánto tiempo, de media, le dedica la unidad SSD a cada consulta que llega al servidor.
 - b) Calcule el tiempo medio de servicio de la unidad SSD.

SOLUCIÓN:

- a) $D_{ssd}=0,96$ s (nos piden la demanda de servicio)
- b) $S_{ssd}=0,192$ s

■

PROBLEMA 5.31 En una red interactiva formada por un servidor de impresión, durante un tiempo $T=2$ horas, se encuentran conectados un total de $NT=30$ clientes, cada uno imprimiendo un único fichero (1 cliente = 1 fichero). Durante esas dos horas, el tiempo medio entre que un cliente solicita la impresión de un fichero al servidor y éste termina de imprimir dicho fichero (es decir, se completa la tarea) es de 45s. Asimismo, el tiempo que transcurre entre que un cliente ve impreso su fichero y vuelve a pedirle al servidor la impresión de otro nuevo es, de media, 25s.

- Calcule la productividad media del servidor.
- ¿Cuántos clientes se encuentran, de media, en reflexión?

SOLUCIÓN:

- $X_0=0,43$ tr/s (nos dan NT , R_0 y Z). T no es necesario para calcular la solución.
- $N_z=10,7$ clientes.

■

PROBLEMA 5.32 Suponga que la estación de servicio i -ésima de una red de colas que simula el comportamiento de un servidor de base de datos tiene un tiempo de servicio constante igual a 2s. Suponga que los trabajos (jobs) llegan con la siguiente distribución temporal:

- Durante los primeros 2 segundos no llega ningún trabajo.
- En $t=2s$ llegan 2 trabajos: J_1 y J_2 (por ese orden).
- En $t=3s$ llega otro trabajo: J_3 .

- Calcule los tiempos de espera en la cola y los tiempos de respuesta que experimentan cada uno de los tres trabajos. Calcule finalmente sus valores medios.
- Para el intervalo de medida $[0, 10]s$, calcule la productividad de la estación de servicio, su utilización y el número medio de trabajos en la cola.

SOLUCIÓN:

- $W_i(J_1) = 0s$, $W_i(J_2) = 2s$, $W_i(J_3) = 3s$. Valor medio = $W_i = 5/3 = 1,67s$. $R_i(J_1) = 2s$, $R_i(J_2) = 4s$, $R_i(J_3) = 5s$. Valor medio = $R_i = 11/3 = 3,67s$.
- $X_i = 0,3$ tr/s, $U_i = 0,6$ (60%). $Q_i = 0,5$ trabajos.

■

PROBLEMA 5.33 Queremos diseñar un servidor de ayuda a la docencia al que se conectarán unos 30 estudiantes durante las 2 horas que duran las sesiones de prácticas de la asignatura. Este servidor consta de una CPU, un disco duro y una tarjeta de red. Tras la prueba de funcionamiento de 2 horas con 30 estudiantes, se han medido los siguientes valores:

Dispositivo	Tiempo de servicio (s)	Razón de visita
CPU (1)	0,01	80
DISCO (2)	0,5	20
RED (3)	0,24	5

¿Cuánto tiempo debería transcurrir, de media, entre que un estudiante recibe la respuesta de este servidor hasta que vuelve a realizar una nueva petición, para que 30 sea precisamente el número ideal de clientes de este servidor?

SOLUCIÓN: 288 s

PROBLEMA 5.34 Partiendo de la hipótesis de que $W_i = N_i \times S_i$ para cada estación de servicio de una red de colas que simula el comportamiento de un servidor, demuestre que el cuello de botella del mismo será aquel dispositivo con mayor número medio de trabajos en la cola.

SOLUCIÓN:

Si se cumple dicha hipótesis, usando las leyes operacionales se obtiene que: $Q_i = U_i^2 / (1 - U_i)$. Si $0 < U_i < U_j < 1$, se cumple que a) $U_i^2 < U_j^2$ y b) $(1 - U_i) > (1 - U_j)$. Por lo tanto: $U_i^2 / (1 - U_i) < U_j^2 / (1 - U_j)$ (ya que el numerador en el caso de la estación j-ésima será mayor y el denominador menor que en el caso de la estación i-ésima). Es decir, la estación de servicio con mayor utilización (es decir, el cuello de botella) también será la que tenga un mayor número de trabajos en la cola. Otra forma de razonar esto último consiste en ver que, en el intervalo $[0, 1[$ (que es el rango en el que se van a mover las utilidades), $f(x) = x^2 / (1 - x)$ es una función estrictamente creciente.
