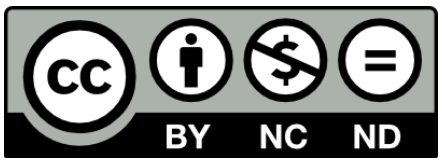
Ingeniería de Servidores

Los Del DGIIM, `losdeldgiim.github.io`

Doble Grado en Ingeniería Informática y Matemáticas
Universidad de Granada



Ingeniería de Servidores

Los Del DGIIM, losdeldgiim.github.io

Arturo Olivares Martos

Granada, 2025

Índice general

1. Resumen de fórmulas

Este apartado trata de ser un resumen de la mayoría de las fórmulas vistas en cada tema de la asignatura, está pensado para recordar de forma rápida las fórmulas vistas en la asignatura, y no debe usarse como medio de estudio.

1.1. Introducción a la Ingeniería de Servidores

Los dos conceptos más importantes del tema son:

- Tiempo de respuesta (latencia).
- Productividad (ancho de banda).

Si queremos comparar dos dispositivos A y B con tiempos t_A y t_B, la ganancia en velocidad de A respecto a B viene dada por:

$$S_B(A) = \frac{t_A}{t_B}$$

Si tenemos un dispositivo que tarda en ejecutar una tarea un tiempo T_0 y mejoramos dicho dispositivo sustituyendoo un compoente que se usa durante una fracción de tiempo $f \in [0, 1]$ de forma que mejoramos dicho componentes k veces, el nuevo tiempo vendrá dado por:

$$T_M = (1 - f)T_0 + \frac{fT_0}{k}$$

Bajo las mismas condiciones, la Ley de Amdahl nos dice que:

$$S = \frac{T_0}{T_M} = \frac{T_0}{(1 - f)T_0 + \frac{fT_0}{k}} = \frac{1}{1 - f + \frac{f}{k}}$$

1.4. Análisis comparativo del rendimiento

Si sabemos las instrucciones de un programa (NI), cuántas de ellas son en coma flotante (FL) y el tiempo que tarda un dispositivo en ejecutar el programa (t), podremos calcular:

$$\text{MIPS} = \frac{NI}{t \cdot 10^6}$$
$$\text{MFLOPS} = \frac{FL}{t \cdot 10^6}$$

Si para un cierto programa conocemos el número de instrucciones necesarias para su ejecución (NI), el número medio de ciclos por instrucción de la CPU (CPI) y la frecuencia del procesador (f), podremos calcular el tiempo que tarda la CPU en ejecutar el programa, mediante:

$$T_{CPU} = NI \cdot CPI \cdot \frac{1}{f}$$

Si al ejecutar un benchmark de n programas obtenemos las puntuaciones $t_1, t_2, \ldots, t_n$ y las puntuaciones de referencia eran $t_{REF_1}, t_{REF_2}, \ldots, t_{REF_n}$, podemos calcular el índice SPEC mediante:

$$SPEC = \sqrt[n]{\frac{t_{REF_1}}{t_1} \cdot \frac{t_{REF_2}}{t_2} \cdot \ldots \cdot \frac{t_{REF_n}}{t_n}}$$

1.4.1. Distribución t-Student

Si extraemos n muestras de ejecución de varios programas por dos dispositivos o programas distintos y consideramos la diferencia de los datos obtenidos: $d_1, d_2, \ldots, d_n$, estamos interesados en calcular si las dos muestras tienen diferencias significativas con un grado de significatividad mayor al 95 %. Para ello, lo que haremos será suponer la hipótesis nula H_0:

Las dos máquinas/programas tienen rendimientos equivalentes

Definimos:

- La media de las diferencias:
$$\overline{d} = \frac{1}{n} \sum_{i=1}^{n} d_i$$
- La desviación típica muestral:
$$s = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \overline{d})^2}{n-1}}$$
- El error estándar:
$$\frac{s}{\sqrt{n}}$$

Supuesta la hipótesis nula H_0, tras un estudio podremos rechazar la hipótesis nula (por lo que los datos no son significativamente distintos) o no podremos rechazarla. Este estudio se puede realizar de 3 formas distintas. Fijado un grado de significatividad usualmente de $\alpha = 0{,}5$ (para el 95 %):

1. Una vez calculados $\overline{d}, s$ y el error estándar, calculamos:
$$t_{exp} = \frac{\overline{d}}{s/\sqrt{n}}$$

Tras lo cual podemos calcular el $p-$value:

$$P(|t| \geqslant |t_{exp}|)$$

Donde consideramos la distribución de probabilidad de $t-$Student de $n-1$ grados de libertad. Si $P(|t| \geqslant |t_{exp}|) < \alpha$, entonces podremos rechazar H_0.

2. Conocidos α y n, si calculamos $t_{\frac{\alpha}{2},n-1}$ que cumple:

$$P(|t| \geqslant |t_{\frac{\alpha}{2},n-1}|) = \alpha$$

Para una distribución de $t-$Student de $n-1$ grados de libertad, si calculamos t_{exp} y:

$$t_{exp} \notin \left[-t_{\frac{\alpha}{2},n-1}, t_{\frac{\alpha}{2},n-1}\right]$$

Entonces podremos rechazar H_0.

3. Si calculamos $t_{\frac{\alpha}{2},n-1}$, si:

$$0 \notin \left[\overline{d} - \frac{s}{\sqrt{n}} \cdot t_{\frac{\alpha}{2},n-1},\ \overline{d} + \frac{s}{\sqrt{n}} \cdot t_{\frac{\alpha}{2},n-1}\right]$$

Entonces podremos rechazar H_0.

1.5. Optimización del rendimiento

En un servidor con K dispositivos o estaciones, para cada uno registraremos los siguientes parámetros (estarán marcados con subíndice i, ya que harán referencia al $i-$ésimo dispositivo, con $i \in \{1, \ldots, K\}$), donde T es el tiempo total durante el que observamos el servidor:

- A_i, número de trabajos solicitados a la estación.
- C_i, número de trabajos completados por la estación.
- B_i, tiempo que el dispositivo ha estado en uso.
- S_i, tiempo medio de servicio:

$$S_i = \frac{B_i}{C_i}$$

- W_i, tiempo medio de espera en cola.
- R_i, tiempo medio de respuesta del dispositivo:

$$R_i = W_i + S_i$$

- λ_i, tasa media de llegada (cuántos trabajos por segundo llegan):

$$\lambda_i = \frac{A_i}{T}$$

- X_i, productividad media (cuántos trabajos se completan por segundo):

$$X_i = \frac{C_i}{T}$$

- U_i, utilización media (porcentaje de tiempo durante el cual el dispositivo ha estado en uso):

$$U_i = \frac{B_i}{T}$$

- Q_i, número medio de trabajos en cola.
- N_i, número medio de trabajos siendo servidos por el dispositivo:

$$N_i = Q_i + U_i$$

Si vemos el servidor como un dispositivo, podemos generalizar estos parámetros para el mismo, los cuales denotaremos con subíndice 0, y nos importarán especialmente:

$$A_0, \quad C_0, \quad R_0, \quad \lambda_0 = \frac{A_0}{T}, \quad X_0 = \frac{C_0}{T}, \quad N_0 = \sum_i N_i$$

Estos parámetros nos permiten considerar un par más de parámetros de dispositivos:

- V_i, la razón media de visita (número medio de veces que un trabajo visita al dispositivo $i-$ésimo durante su visita al servidor):

$$V_i = \frac{C_i}{C_0}$$

- D_i, la demanda media del servicio (cantidad media de tiempo que el dispositivo dedica a cada trabajo que abandona el servidor):

$$D_i = \frac{B_i}{C_0} = V_i \cdot S_i$$

Los parámetros esenciales en un servidor son su productividad X_0 y tiempo de respuesta, R_0, y siempre tratraremos de aumentar la productividad y de reducir el tiempo de respuesta, algo que se pondrá de manifiesto más adelante.

1.5.1. Leyes operacionales

El valor de las variables anteriores depende obviamente del tiempo de observación T, pero existen algunas relaciones entre variables que son ciertas independientemente de dicho tiempo T, lo que se denominan leyes operacionales.

Definición 1.1. Supuesto que T es suficientemente grande (en relación con R_0), decimos que el servidor está en equilibrio de flujo si el número de trabajos que este completa coincide de forma aproximada con el número de trabajos solicitados al mismo, es decir, si:

$$A_0 \approx C_0$$

En dicho caso, también tendremos que $X_0 \approx \lambda_0$, lo que puede tomarse como una definición equivalente.

A continuación mostraremos las distintas leyes operacionales que manejamos en este curso. Muchas de ellas no necesitan la hipótesis de que el servidor esté en equilibrio de flujo. Sin embargo, si en estas aparece X_0, por cada fórmula podemos obtener otra equivalente, suponiendo que el servidor está en equilibrio de flujo y sustituyendo X_0 por λ_0.

Proposición 1.1 (Ley de la utilización)**.** *Podemos relacionar la utilización media de un dispositivo con el número de trabajos que completa por unidad de tiempo (es decir, con su productividad) y con el tiempo medio que le dedica a cada uno de ellos:*

$$U_i = X_i \cdot S_i$$

Y si está en equilibrio de flujo: $U_i = \lambda_i \cdot S_i$.

Demostración.

$$S_i = \frac{B_i}{C_i} = \frac{\frac{B_i}{T}}{\frac{C_i}{T}} = \frac{U_i}{X_i} \Longrightarrow U_i = X_i \cdot S_i$$

□

Proposición 1.2 (Ley del flujo forzado)**.** *Los flujos de salida de cada estación de servicio han de ser proporcionales a la productividad global del servidor:*

$$X_i = X_0 \cdot V_i$$

En caso de equilibrio de flujo: $X_i = X_0 \cdot V_i = \lambda_0 \cdot V_i = \lambda_i$.

Demostración.

$$V_i = \frac{C_i}{C_0} = \frac{\frac{C_i}{T}}{\frac{C_0}{T}} = \frac{X_i}{X_0} \Longrightarrow X_i = X_0 \cdot V_i$$

□

Proposición 1.3 (Relación utilización-demanda de servicio)**.** *Las utilizaciones de cada servicio son proporcionales a las demandas de servicio del mismo, siendo la constante de proporcionalidad la productividad del servidor:*

$$U_i = X_0 \cdot D_i$$

En caso de equilibrio de flujo: $U_i = \lambda_0 \cdot D_i$.

Demostración.

$$D_i = \frac{B_i}{C_0} = \frac{\frac{B_i}{T}}{\frac{C_0}{T}} = \frac{U_i}{X_0} \Longrightarrow U_i = X_0 \cdot D_i$$

□

Proposición 1.4 (Ley de Little)**.** *Si el sistema está en equilibrio de flujo*[1]*:*

$$N_i = \lambda_i \cdot R_i = X_i \cdot R_i$$

También puede aplicarse a la cola de un dispositivo:

$$Q_i = \lambda_i \cdot W_i = X_i \cdot W_i$$

Proposición 1.5 (Ley general del tiempo de respuesta)**.** *El tiempo medio de respuesta que experimenta una petición en un servidor en equilibrio de flujo teniendo en cuenta que cada una de ellas ha visitado V_i veces el dispositivo $i-$ésimo, donde ha permanecido R_i segundos de media:*

$$R_0 = \sum_{i=0}^{K} V_i \cdot R_i$$

Demostración. Tenemos que:

$$N_0 = \sum_{i=0}^{K} N_i$$

Y si aplicamos la Ley de Little:

$$X_0 \cdot R_0 = \sum_{i=0}^{K} X_i \cdot R_i$$

Si usamos la Ley del flujo forzado:

$$X_0 \cdot R_0 = \sum_{i=0}^{K} X_0 \cdot V_i \cdot R_i$$

De donde:

$$R_0 = \sum_{i=0}^{K} V_i \cdot R_i$$

□

Antes de ver la siguiente ley cabe mencionar que las redes cerradas siempre están en equilibrio de flujo, ya que supondremos que sus colas son suficientemente grandes como para alvergar la cantidad de trabajos N_T en estas redes.

Proposición 1.6 (Ley del tiempo de respuesta interactivo)**.** *En una red cerrada:*

$$R_0 = \frac{N_T}{X_0} - Z$$

Demostración. Sabemos que $N_T = N_Z + N_0$ y si aplicamos la Ley de Little a ambos:

$$N_T = N_Z + N_0 = X_0 \cdot Z + X_0 \cdot R_0 = X_0(Z + R_0) \Longrightarrow R_0 = \frac{N_T}{X_0} - Z$$

□

[1]En otro caso la Ley de Little no es cierta.

En resumen:

- Ley de la Utilización.
$$U_i = X_i \cdot S_i$$
- Ley del flujo forzado.
$$X_i = X_0 \cdot V_i$$
- Relación de la utilización-demanda.
$$U_i = X_0 \cdot D_i$$
- Ley de Little (solo en equilibrio de flujo).
$$N_i = \lambda_i \cdot R_i$$
- Ley general del tiempo de respuesta.
$$R_0 = \sum_{i=1}^{k} V_i \cdot R_i$$
- Ley del tiempo de respuesta interactivo (para redes cerradas).
$$R_0 = \frac{N_T}{X_0} - Z$$

1.5.2. Limites del tiempo de respuesta y de la productividad

Definición 1.2. Decimos que un dispositivo es el cuello de botella de un servidor si es el que primero llega a saturarse (es decir, a tener el porcentaje de utilización al 100 %). Si llamamos b al índice del dispositivo del cuello de botella, como:

$$U_b = X_0 \cdot D_b$$

Tenemos que una definición equivalente de dispositivo de cuello de botella es aquel con mayor tiempo de demanda.

Por esta definición, tenemos que:

$$U_b = \max_{i=1,\dots,k} \{U_i\} = V_b \cdot S_b$$
$$D_b = \max_{i=1,\dots,k} \{D_i\} = X_0 \cdot U_b$$

Definición 1.3. Diremos que el servidor está saturado si el dispositivo que es el cuello de botella del servidor lo está, es decir, si $U_b = 1$.

Estudiaremos ahora cómo determinar los límites de productividad máxima $X_0^{\max}$ y tiempo de respuesta mínimo $R_0^{\min}$ de un servidor, que dependerá si trabajamos en una red abierta o cerrada:

Redes abiertas

- Para el cálculo de la productividad máxima $X_0^{\max}$ parece lógico pensar que esta se dé cuando el servidor esté saturado, ya que en dicho caso estará atendiendo la máxima cantidad de trabajos por segundo, puesto que al terminar una tarea en un dispositivo este volverá nuevamente a estar ocupado por otra tarea, al estar su cola no vacía. Por tanto, si suponemos que $U_b = 1$ y aplicamos la ley de relación de utilización-demanda, vemos que:

$$U_b = X_0 \cdot D_b \Longrightarrow X_0 = \frac{U_b}{D_b}$$

  Y como estábamos suponiendo que $U_b = 1$ para conseguir la mayor productividad, tenemos que:

$$X_0^{\max} = \frac{1}{D_b}$$

- Para el cálculo del tiempo de respuesta mínimo $R_0^{\min}$, parece lógico pensar lo contrario: el tiempo de respuesta mínimo se dará cuando todas las colas estén vacías y llegue un proceso a la cola. En dicho caso, el tiempo de respuesta en cada cola será despreciable, quedando solo el tiempo que tarde el trabajo en cada dispositivo que visita. De esta forma:

$$R_0 = \sum_{i=1}^{k} V_i \cdot R_i = \sum_{i=1}^{k} V_i \cdot (W_i + U_i) \overset{(*)}{=} \sum_{i=1}^{k} V_i \cdot U_i = \sum_{i=1}^{k} D_i$$

  Donde en $(*)$ hemos usado que como todas las colas están vacías, $W_i = 0$ para todo $i = 1, \ldots, k$. Si definimos:

$$D := \sum_{i=1}^{k} D_i$$

  Tendremos que:

$$R_0^{\min} = D$$

Redes cerradas

Como en redeces cerradas siempre tenemos equilibrio de flujo, nunca podrá estar el servidor saturado, por lo que hemos de distinguir casos con el fin de calcular la productividad máxima y tiempo de respuesta mínimo:

- Suponiendo que hay muchos trabajos en la red (N_T grande), entonces el servidor no llegará a estar saturado, pero podemos pensar que no le falta mucho para ello. Intuitivamente, podemos tomar $U_b \to 1$, por lo que estamos en un caso similar a la saturación en redes abiertas, y tendremos que:

$$X_0 \to \frac{1}{D_b}$$

  Para el tiempo de respuesta mínimo, recordamos que teníamos en redes cerradas que:

$$R_0 = \frac{N_T}{X_0} - Z$$

Y como N_T y Z son fijos, vemos que el tiempo mínimo de respuesta se alcanza cuando se alcanza la productividad máxima del sistema, por lo que:

$$R_0 = \frac{N_T}{X_0} - Z \to D_b \cdot N_T - Z$$

- Suponiendo que hay pocos trabajos en la red (N_T pequeño), entonces las colas del servidor estarán relativamente vacías, por lo que podemos pensar en que $W_i \to 0$ para $i = 1, \ldots, k$. En dicho caso:

$$R_0 = \sum_{i=1}^{k} V_i \cdot (W_i + U_i) \to \sum_{i=1}^{k} V_i \cdot U_i = D$$

Y como es una red cerrada, seguimos teniendo que:

$$R_0 = \frac{N_T}{X_0} - Z \Longrightarrow X_0 = \frac{N_T}{R_0 + Z}$$

De donde al igual que antes, la productividad máxima se obtendrá cuando el tiempo de respuesta sea mínimo, por lo que:

$$X_0 = \frac{N_T}{R_0 + Z} \to \frac{N_T}{D + Z}$$

Finalmente, como buscábamos cotas inferiores y cotas superiores, tendremos que:

$$X_0^{\max} \leqslant \text{mín} \left\{ \frac{1}{D_b}, \frac{N_T}{D + Z} \right\}$$
$$R_0^{\min} \geqslant \text{máx}\{D_b \cdot N_T - Z, D\}$$

Además, podremos pensar en el punto de equilibrio entre todos estos valores, es decir, cuando:

$$D = D_b \cdot N_T - Z \Longrightarrow N_T = \frac{D + Z}{D_b}$$

Al que llamaremos punto teórico de saturación del servidor.

2. Relaciones de Ejercicios

2.1. Introducción a la Ingeniería de Servidores

Ejercicio 2.1.1. Un programa para la simulación de sistemas hidráulicos se ejecuta en 122 segundos. Si las operaciones de división con números reales consumen el 73 % de este tiempo, ¿en cuánto se tendría que mejorar la velocidad de estas operaciones si queremos conseguir que dicho programa se ejecute seis veces más rápidamente? ¿Cuál es la ganancia en velocidad máxima que podríamos conseguir si pudiésemos mejorar dichas operaciones tanto como quisiéramos?

Tenemos que, representando por k el número de veces que se mejora la velocidad de las operaciones de división con números reales, la ganancia en velocidad S es:

$$S = \frac{T_o}{T_m} = \frac{T_o}{(1-0{,}73)T_o + \dfrac{0{,}73T_o}{k}} = \frac{1}{0{,}27 + \dfrac{0{,}73}{k}} \leqslant \frac{1}{0{,}27} = 3,7$$

Por tanto, la ganancia máxima que vamos a poder conseguir mejorando tan solo las operaciones de división con números reales es de 3,7. Si quisiésemos que el programa se ejecutase seis veces más rápidamente, tendríamos que mejorar otros aspectos del sistema.

Ejercicio 2.1.2. Una mejora en un sitio web ha permitido rebajar de 17 a 9 segundos el tiempo medio de descarga de sus páginas. Si la mejora ha consistido en hacer 3 veces más rápido el subsistema de discos que almacena las páginas del servidor, ¿cuánto tiempo se dedicaba a acceder a los discos antes de realizar la mejora?

Sea T_o el tiempo medio de descarga de las páginas antes de la mejora, y T_m el tiempo medio de descarga de las páginas después de la mejora. Sea además f la fracción del tiempo de descarga de las páginas que se dedica a acceder a los discos. Entonces, se tiene:

$$T_m = (1-f)T_o + \frac{fT_o}{3} \Longrightarrow 3 \cdot \frac{T_o - T_m}{2T_o} = f = 3 \cdot \frac{17-9}{2 \cdot 17} = \frac{12}{17}$$

Por tanto, antes de la mejora, el tiempo medio de descarga de las páginas que se dedicaba a acceder a los discos era de:

$$fT_o = \frac{12}{17} \cdot 17 = 12\,\mathrm{s}$$

Ejercicio 2.1.3. Un computador tarda 100 segundos en ejecutar un programa de simulación de una red de interconexión para multicomputadores. El programa dedica

el 20 % en hacer operaciones de aritmética entera (AE), el 30 % en hacer operaciones de aritmética en coma flotante (CF), mientras que el resto se emplea en operaciones de entrada/salida (E/S). Calcule la ganancia en velocidad y el tiempo de ejecución si las operaciones aritméticas enteras y reales se mejoran de manera simultánea 2 y 3 veces, respectivamente.

Sea T_o el tiempo de ejecución del programa antes de la mejora, y T_m el tiempo de ejecución del programa después de la mejora. Sea además f_{AE} y f_{CF} las fracciones del tiempo de ejecución del programa que se dedican a las operaciones de aritmética entera y aritmética en coma flotante, respectivamente. Entonces, se tiene:

$$\begin{aligned} T_m &= (1 - f_{AE} - f_{CF})T_o + \frac{f_{AE}T_o}{2} + \frac{f_{CF}T_o}{3} = \\ &= T_o \left[(1 - 0{,}2 - 0{,}3) + \frac{0{,}2}{2} + \frac{0{,}3}{3}\right] = 0{,}7T_o = \\ &= 70\,\text{s} \end{aligned}$$

Por tanto, la ganancia en velocidad es:

$$S = \frac{T_o}{T_m} = \frac{T_o}{0{,}7T_o} = \frac{1}{0{,}7} \approx 1{,}43$$

Ejercicio 2.1.4. Una aplicación informática se ejecuta en un computador durante un total de 70 segundos. Mediante el uso de un monitor de actividad se ha podido saber que el 85 % del tiempo se utiliza la tarjeta de red, mientras que el resto del tiempo se hace uso del procesador. Se pide, considerando el sistema original como punto de partida:

1. Calcular el incremento de prestaciones si se mejora en 8 veces la velocidad de la tarjeta de red.

   Sea T_o el tiempo de ejecución del programa antes de la mejora, y T_m el tiempo de ejecución del programa después de la mejora. Sea además f la fracción del tiempo de ejecución del programa que se dedica a la tarjeta de red. Entonces, se tiene:

   $$\begin{aligned} T_m &= (1 - f)T_o + \frac{fT_o}{8} = \\ &= T_o \left[(1 - 0{,}85) + \frac{0{,}85}{8}\right] = 0{,}25625\ T_o = \\ &= 17{,}9375\,\text{s} \end{aligned}$$

   Por tanto, la ganancia en velocidad es:

   $$S = \frac{T_o}{T_m} = \frac{T_o}{0{,}25625T_o} = \frac{1}{0{,}25625} \approx 3{,}9$$

2. Determinar en cuánto hay que mejorar el rendimiento del procesador si se quiere ejecutar la aplicación en 25 segundos.

El tiempo que se emplea la tarjeta de red (sin hacer mejoras de esta) es de:

$$fT_o = 0{,}85 \cdot 70 = 59{,}5\,\mathrm{s}$$

Por tanto, mejorando únicamente el rendimiento del procesador, no es posible ejecutar la aplicación en 25 segundos.

Ejercicio 2.1.5. De acuerdo con la ley de Amdahl, deduzca una expresión para la fracción de tiempo f en función de S (el speedup) y k (el número de veces mejorado).

La ley de Amdahl establece que la ganancia en velocidad S es:

$$S = \frac{T_o}{T_m} = \frac{T_o}{(1-f)T_o + \dfrac{fT_o}{k}} = \frac{1}{1-f+\dfrac{f}{k}} \Longrightarrow \frac{1}{S} - 1 = f\left(-1 + \frac{1}{k}\right) \Longrightarrow$$

$$\Longrightarrow f = \frac{k(1-S)}{S(1-k)} = \frac{k(S-1)}{S(k-1)} \qquad k \neq 1$$

Ejercicio 2.1.6. El administrador de un sistema informático pretende aumentar el rendimiento para evitar que el director del centro lo cese en sus funciones (ha habido más de quince quejas de usuarios en el último mes por el excesivo tiempo de ejecución de los programas). Indíquese, teniendo en cuenta la relación entre prestaciones y coste, qué opción de actualización de un sistema informático, de las dos que se enumeran, resultará más ventajosa:

1. Cambio del procesador (250 €). Esta modificación permite que el 75 % de los programas se ejecuten dos veces más rápidamente.

   Sea T_o el tiempo total de ejecución de los programas antes de la actualización, y T_A el tiempo total de ejecución de los programas después de la actualización. Sea además f la fracción de los programas cuya velocidad se vería aumentada (75 % en este caso). Entonces, se tiene:

   $$T_A = (1-f)T_o + \frac{fT_o}{2} =$$
   $$= T_o\left[(1-0{,}75) + \frac{0{,}75}{2}\right] = 0{,}625\ T_o$$

2. Ampliación de la memoria principal (150 €). La capacidad extra de memoria mejora tres veces el tiempo de ejecución del 40 % de los programas.

   Sea T_o el tiempo total de ejecución de los programas antes de la actualización, y T_B el tiempo total de ejecución de los programas después de la actualización. Sea además f la fracción de los programas cuya velocidad se vería aumentada (40 % en este caso). Entonces, se tiene:

   $$T_B = (1-f)T_o + \frac{fT_o}{3} =$$
   $$= T_o\left[(1-0{,}4) + \frac{0{,}4}{3}\right] = \frac{11}{15} \cdot T_o$$

Aunque vemos que la primera alternativa es, de media, un 17,33 % más rápida que la segunda, hemos de comparar en función de la relación prestaciones/coste.

$$\frac{{}^{\text{Prestaciones}}/_{\text{Coste}_A}}{{}^{\text{Prestaciones}}/_{\text{Coste}_B}} = \frac{{}^{11}/_{15} \cdot T_o \cdot 150€}{0{,}625 \cdot T_o \cdot 250€} = \frac{88}{125} = 0{,}704$$

Por tanto, la primera alternativa (cambio del procesador) es 0,704 veces más ventajosa que la segunda. Equivalentemente, y para entendernos, la segunda alternativa es 1,42 veces (un 42 %) más ventajosa que la primera. Por tanto, el administrador debería optar por la segunda alternativa, ampliar la memoria principal.

Ejercicio 2.1.7. Un programa de predicción meteorológica tarda 84 minutos en ejecutarse en un supercomputador diseñado al efecto. Sin embargo, esta cantidad de tiempo origina muchos problemas para los estudios de los meteorólogos. El responsable del equipo informático quiere reducir este tiempo sustituyendo la memoria principal por una más rápida, para lo cual existen dos modelos alternativos:

1. Modelo Lupita (1100 €), que disminuye el tiempo de ejecución hasta los 71 minutos.

2. Modelo Lucho (1300 €), que rebaja este tiempo de ejecución hasta los 63 minutos.

Determine cuál de los dos modelos anteriores representa la mejor opción ateniéndonos a la relación prestaciones/coste. Exprese el resultado como "% de mejora en la relación prestaciones/coste".

La comparación de la relación prestaciones/coste de los dos modelos es:

$$\frac{{}^{\text{Prestaciones}}/_{\text{Coste}_{\text{Lupita}}}}{{}^{\text{Prestaciones}}/_{\text{Coste}_{\text{Lucho}}}} = \frac{63 \cdot 1300}{71 \cdot 1100} = \frac{819}{781} \approx 1{,}049$$

Por tanto, el modelo Lupita es 1,049 veces (un 4,9 %) más ventajoso que el modelo Lucho.

Ejercicio 2.1.8. El tiempo medio de respuesta de un sitio web es de 15 segundos. Mediante un monitor software se ha podido determinar que el 55 % de este tiempo es utilizado por el subsistema de discos, mientras que el resto se dedica a la ejecución de los scripts en el procesador de 2 GHz de que dispone el servidor. El administrador del sitio, después de soportar estoicamente las quejas de los usuarios, pretende reducir este tiempo por debajo de los 11 segundos. ¿Cuál de las dos opciones planteadas a continuación consigue este objetivo?

1. Adquirir un nuevo procesador que trabaja a 3 GHz.

   Sea T_o el tiempo de respuesta del sitio web antes de la actualización, y T_A el tiempo de respuesta del sitio web después de la actualización. Sea además f la fracción del tiempo de respuesta del sitio web que se dedica a la ejecución de los scripts en el procesador. Entonces, se tiene:

$$\begin{aligned} T_A &= (1-f)T_o + fT_o \cdot \frac{2}{3} = \\ &= T_o \left[(1 - 0{,}45) + 0{,}45 \cdot \frac{2}{3} \right] = 0{,}85\ T_o = \\ &= 12{,}75\,\text{s} \end{aligned}$$

2. Substituir el subsistema de discos por uno de segunda mano 2,5 veces más rápido que el actual.

   Sea T_o el tiempo de respuesta del sitio web antes de la actualización, y T_B el tiempo de respuesta del sitio web después de la actualización. Sea además f la fracción del tiempo de respuesta del sitio web que se dedica al subsistema de discos. Entonces, se tiene:

$$\begin{aligned} T_B &= (1-f)T_o + fT_o \cdot \frac{1}{2{,}5} = \\ &= T_o \left[(1-0{,}55) + 0{,}55 \cdot \frac{1}{2{,}5} \right] = 0{,}67\ T_o = \\ &= 10{,}05\,\text{s} \end{aligned}$$

Por tanto, la segunda opción es la que consigue el objetivo de reducir el tiempo de respuesta por debajo de los 11 segundos.

Ejercicio 2.1.9. Un programa de simulación de sistemas aerodinámicos de control se ejecuta en 280 segundos. El 70 % del tiempo de ejecución se utiliza el procesador; el resto se dedica a acceder al subsistema de discos. Un incremento del presupuesto aportado por el ministerio ha permitido adquirir un nuevo procesador tres veces más rápido.

1. Determine el tiempo de ejecución del simulador después de actualizar el procesador.

   Sea T_o el tiempo de ejecución del programa antes de la actualización, y T_m el tiempo de ejecución del programa después de la actualización. Sea además f la fracción del tiempo de ejecución del programa que se dedica al procesador. Entonces, se tiene:

$$\begin{aligned} T_m &= (1-f)T_o + fT_o \cdot \frac{1}{3} = \\ &= T_o \left[(1-0{,}7) + 0{,}7 \cdot \frac{1}{3} \right] = \frac{8}{15}\ T_o = \\ &= 149{,}33\,\text{s} \end{aligned}$$

2. Calcule ahora, esto es, después de haber hecho la actualización del procesador, cuál es la fracción del tiempo mejorado de ejecución durante el cual se utiliza el nuevo procesador. Haga un análisis del fenómeno observado.

   La fracción del tiempo de ejecución mejorado de ejecución durante el cual se utiliza el nuevo procesador es:

$$f_m = 1 - \frac{(1-0{,}7)T_o}{T_m} = 1 - \frac{0{,}3T_o}{{}^{8}\!/_{15}T_o} = 0{,}4375$$

   El tiempo dedicado a acceder al subsistema de discos se mantiene constante, pero el tiempo dedicado a la ejecución de los scripts en el procesador se reduce del 70 % al 43.75 %. Por tanto, anteriormente sí era vital mejorar dicho componente, pero ahora, con el nuevo procesador, es más importante mejorar el subsistema de discos.

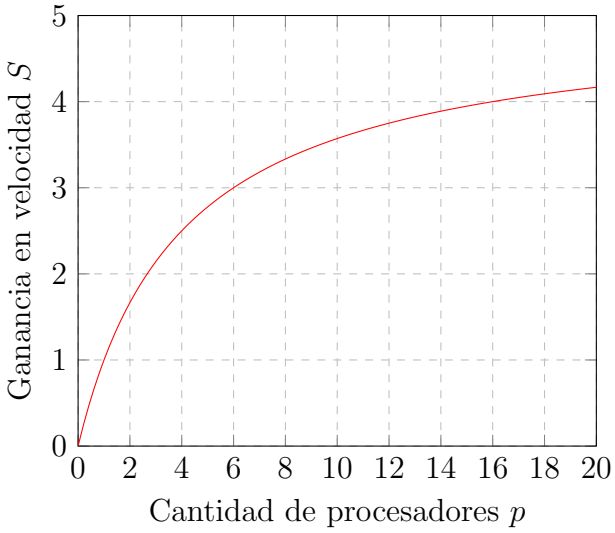


Figura 2.1: Ganancia en velocidad (S) en función del número de procesadores (p).

3. A raíz del resultado obtenido en el apartado anterior, si hubiéramos de mejorar este sistema actualizado, ¿sobre qué componente del mismo deberíamos incidir? Justifique numéricamente la respuesta.

   Como hemos indicado, y debido a que siempre hemos de priorizar mejorar el componente que más tiempo de ejecución consume, deberíamos incidir sobre el subsistema de discos.

Ejercicio 2.1.10. Un equipo de biólogos que investiga sobre clonación de células utiliza el multiprocesador ALLIANT para ejecutar un simulador que se puede paralelizar en una fracción f de su tiempo de ejecución. La Figura 2.1 presenta la ganancia en velocidad conseguida por la máquina paralela en la ejecución del simulador para diferentes valores del número de procesadores (p).

1. ¿Cuál es la fracción paralelizable f del programa de simulación?

   El tiempo mejorado, T_m, es:

   $$T_m = (1-f)T_o + \frac{fT_o}{p}$$

   Por tanto, la ganancia en velocidad S en función del número de procesadores p es:

   $$S = \frac{T_o}{T_m} = \frac{T_o}{(1-f)T_o + \frac{fT_o}{p}} = \frac{1}{1-f+\frac{f}{p}}$$

   Despejando f de forma análoga al Ejercicio 2.1.5, se obtiene:

   $$f = \frac{p(S-1)}{S(p-1)}$$

   Empleando tanto los puntos $(6,3)$ como $(16,4)$, se obtiene en ambos casos:

   $$f = \frac{4}{5} = 0{,}8$$

2. Si la parte secuencial (es decir, la no paralelizable) del simulador se ejecuta en 65s, ¿cuánto tiempo han de esperar los biólogos para obtener los resultados de la simulación con una configuración de 6 procesadores?

   Como el 20 % del tiempo total de ejecución (en el caso de un único procesador) se dedica a la parte secuencial, el tiempo de ejecución del simulador en dicho únco procesador es:

   $$0{,}2T_o = 65\,\mathrm{s} \Longrightarrow T_o = 325\,\mathrm{s}$$

   Por tanto, el tiempo de ejecución del simulador en un sistema con 6 procesadores es:

   $$T_m = (1-f)T_o + \frac{fT_o}{6} = 0{,}2T_o + \frac{0{,}8T_o}{6} = T_o\left(0{,}2 + \frac{0{,}8}{6}\right) = \frac{T_o}{3} = 108{,}3\,\mathrm{s}$$

3. Los científicos pretenden obtener resultados del simulador en un tiempo máximo de 70s sin modificar el código del programa. Si el sistema ALLIANT está preparado para ampliar el número de procesadores hasta $p = 30$, ¿podrán conseguir los biólogos su objetivo?

   El tiempo de ejecución del simulador en un sistema con 30 procesadores sería:

   $$T_m = (1-f)T_o + \frac{fT_o}{30} = 0{,}2T_o + \frac{0{,}8T_o}{30} = T_o\left(0{,}2 + \frac{0{,}8}{30}\right) = \frac{17}{75}T_o = 73{,}667\,\mathrm{s}$$

   Por tanto, los biólogos no podrán conseguir su objetivo, puesto que el tiempo de ejecución del simulador en un sistema con 30 procesadores es mayor que su objetivo.

   Veamos cuántos procesadores necesitarían para conseguir su objetivo:

   $$T_m = (1-f)T_o + \frac{fT_o}{p} = 0{,}2T_o + \frac{0{,}8T_o}{p} \leqslant 70 \iff 65 + \frac{260}{p} \leqslant 70 \iff p \geqslant \frac{260}{5} = 52$$

   Por tanto, necesitarían al menos 52 procesadores para conseguir su objetivo.

4. Un informático afirma que el sistema ALLIANT podría conseguir el objetivo anterior con $p = 6$ procesadores si se reduce a la mitad la fracción secuencial (es decir, no paralelizable) del simulador. ¿Es válida esta propuesta?

   Supongamos ahora que el tiempo secuencial es tan solo el 10 % del tiempo total de ejecución con un único procesador; es decir, $f = 0{,}9$. En este caso, el tiempo de ejecución del simulador en un sistema con 6 procesadores sería:

   $$T_m = (1-f)T_o + \frac{fT_o}{6} = 0{,}1T_o + \frac{0{,}9T_o}{6} = T_o\left(0{,}1 + \frac{0{,}9}{6}\right) = \frac{T_o}{4} = 81{,}25\,\mathrm{s}$$

   Por tanto, la propuesta no es válida, ya que el tiempo de ejecución del simulador en un sistema con 6 procesadores sería mayor que su objetivo.

Ejercicio 2.1.11. Ante la necesidad de reducir el tiempo de ejecución de un programa de cálculo de trayectorias espaciales, un equipo de arquitectos de computadores ha diseñado un nuevo procesador que mejora 3 veces la ejecución de las operaciones de coma flotante. El programa, cuando se ejecuta utilizando este nuevo procesador, emplea el 65 % del tiempo en la realización de operaciones de coma flotante.

1. Calcule qué tanto por ciento del tiempo de ejecución necesitaban las operaciones de coma flotante en el sistema con el procesador original.

   Sea f la fracción del tiempo de ejecución del programa original que se dedica a las operaciones de coma flotante. Entonces, se tiene:

   $$T_m = (1-f)T_o + \frac{fT_o}{3}$$
   $$0{,}65T_m = \frac{fT_o}{3}$$

   Por tanto, tenemos que $fT_o = 1{,}95T_m$. Por tanto:

   $$T_m = T_o - 1{,}95T_m + 0{,}65T_m \Longrightarrow 2{,}3T_m = T_o$$

   Por tanto, tenemos que:

   $$0{,}65 \cdot \frac{T_o}{2{,}3} = \frac{fT_o}{3} \Longrightarrow f = 3 \cdot \frac{0{,}65}{2{,}3} = \frac{39}{46} \approx 0{,}848$$

   Por tanto, el 84,8 % del tiempo de ejecución necesitaban las operaciones de coma flotante en el sistema con el procesador original.

2. Indique cuál es la ganancia en velocidad global conseguida por el nuevo procesador.

   Tenemos que:

   $$S = \frac{T_o}{T_m} = \frac{2{,}3T_m}{T_m} = 2{,}3$$

Ejercicio 2.1.12. La gráfica de la Figura 2.2 muestra la ganancia en velocidad (speedup), calculada mediante la ley de Amdahl, que se consigue en un computador después de reemplazar la vieja unidad de disco por una nueva, en función de la fracción del tiempo de ejecución en el que se usaba la antigua unidad.

1. Indique cuántas veces es más rápida la nueva unidad de disco respecto de la que se ha retirado del computador.

   Tenemos que:

   $$S = \frac{1}{1-f+\frac{f}{k}}$$

   Empleando el punto $(1, 4)$, se tiene:

   $$4 = \frac{1}{1-1+\frac{1}{k}} \Longrightarrow k = 4$$

   Por tanto, la nueva unidad de disco es 4 veces más rápida que la vieja.

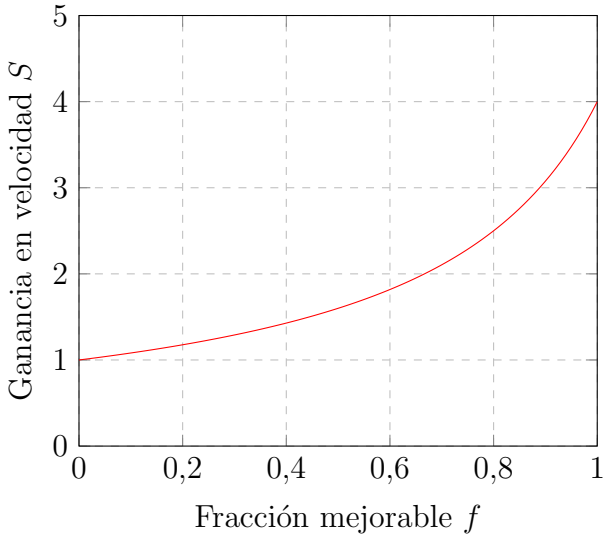


Figura 2.2: Ganancia en velocidad (S) en función de la fracción mejorable (f).

2. El computador, antes de hacer la actualización, tardaba 126 segundos en ejecutar la aplicación. Determine, en el mejor de los casos, cuál sería el tiempo de ejecución en el sistema actualizado. Justifique la respuesta.

   Tenemos que:

   $$T_m = (1-f)T_o + \frac{fT_o}{4} = T_o\left(1 - f + \frac{f}{4}\right)$$

   El mejor de los casos sería suponer que la fracción del tiempo de ejecución en el que se usaba la antigua unidad era la máxima posible, es decir, $f = 1$. Por tanto, el tiempo de ejecución en el sistema actualizado sería:

   $$T_m = \frac{T_o}{4} = 31{,}5\,\mathrm{s}$$

3. Dibuje sobre la misma gráfica la curva que se obtendría si la nueva unidad de disco fuera 2 veces más rápida que la vieja.

   La gráfica obtenida se encuentra en la Figura 2.3, y vendría dada por la ecuación:

   $$S = \frac{1}{1 - f + \frac{f}{2}}$$

Ejercicio 2.1.13. Una aplicación informática se ejecuta en un computador durante un total de 70s. Mediante el uso de un monitor de actividad se ha podido saber que durante el 85 % del tiempo de ejecución se utiliza la CPU (CPUo), mientras que el resto del tiempo se hace uso del disco duro (DD). Determine cuántas veces debe ser, como mínimo, más rápido un procesador (CPUm) que cuesta el doble que el procesador actual para que hubiese valido la pena comprarlo en lugar de éste ateniéndonos a la relación prestaciones del sistema/coste del procesador.

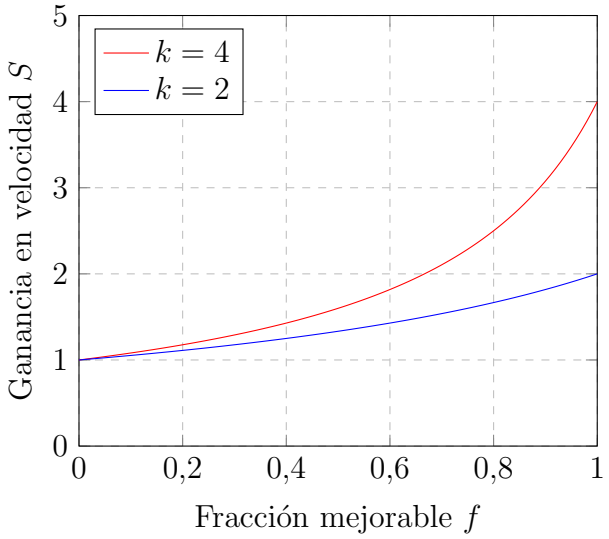


Figura 2.3: Ganancia en velocidad (S) en función de la fracción mejorable (f).

Sea k el número de veces más rápido que el procesador actual ha de ser el nuevo procesador. Entonces, se tiene:

$$T_m = (1-f)T_o + \frac{fT_o}{k} = \\ = T_o\left[(1-0{,}85) + 0{,}85\cdot\frac{1}{k}\right] = T_o\left[0{,}15 + \frac{0{,}85}{k}\right]$$

Tenemos por tanto que:

$$\frac{{}^{\text{Prestaciones}}/_{\text{Coste}_{\text{CPUm}}}}{{}^{\text{Prestaciones}}/_{\text{Coste}_{\text{CPUo}}}} = \frac{T_o}{T_m}\cdot\frac{\text{Coste}_{\text{CPUo}}}{\text{Coste}_{\text{CPUm}}} = \frac{1}{0{,}15+\frac{0{,}85}{k}}\cdot\frac{1}{2} \geqslant 1 \iff \\ \iff 1-0{,}3 \geqslant \frac{1{,}7}{k} \iff k \geqslant \frac{1{,}7}{0{,}7} \approx 2{,}4286$$

Por tanto, el nuevo procesador ha de ser al menos 2.43 veces más rápido que el procesador actual para que hubiese valido la pena comprarlo en lugar de éste.

Ejercicio 2.1.14. Se sabe que el tiempo de respuesta de una petición a un servidor de bases de datos es de 23 segundos, y que el 72 % de ese tiempo se emplea en acceder al subsistema de discos, cuyo coste es de 3500 €. Con el objetivo de mejorar las prestaciones del servidor, un ingeniero en informática está estudiando la posibilidad de adquirir, en su lugar, otro subsistema de discos tres veces más rápido pero con un coste de 4800 €.

1. Calcúlese el nuevo tiempo de respuesta del servidor con el subsistema de discos más caro.

   El tiempo de respuesta del servidor con el subsistema de discos más caro sería:

$$T_m = (1-f)T_o + \frac{fT_o}{3} = T_o\left(1-f+\frac{f}{3}\right) = \\ = \frac{13}{25}\,T_o = 11{,}96\,\text{s}$$

2. ¿Merece la pena comprar el sub-sistema de discos más caro ateniéndonos exclusivamente a la relación prestaciones/coste?

   Tenemos que:

$$\frac{{}^{\text{Prestaciones}}/_{\text{Coste}_{\text{caro}}}}{{}^{\text{Prestaciones}}/_{\text{Coste}_{\text{barato}}}} = \frac{T_o}{T_m} \cdot \frac{\text{Coste}_{\text{barato}}}{\text{Coste}_{\text{caro}}} = \frac{T_o}{{}^{13}/_{25}\ T_o} \cdot \frac{3500}{4800} \approx 1{,}4022$$

   Por tanto, sí merece la pena comprar el sub-sistema de discos más caro.

3. ¿Cuál es la mejora máxima teórica que se podría alcanzar en el tiempo de respuesta manteniendo el subsistema de discos más barato y mejorando el resto de componentes? Exprese el resultado en "número de segundos más rápido" y en "número de veces más rápido".

   Suponiendo ahora que $f = 1 - 0{,}72 = 0{,}28$ y que k es la mejora, se tiene:

$$T_m = (1-f)T_o + \frac{fT_o}{k} = 0{,}72T_o + \frac{0{,}28T_o}{k} \geqslant 0{,}72T_o = 16{,}56\,\text{s}$$

   Por tanto, tenemos que:

$$T_o - T_m \leqslant 23 - 16{,}56 = 6{,}44\,\text{s}$$

   Es decir, la mejora máxima teórica haría que el tiempo de respuesta fuese 6,44 s más rápido que el original. Por otro lado, se tiene:

$$\frac{T_o}{T_m} \leqslant \frac{T_o}{0{,}72T_o} = \frac{25}{18} \approx 1{,}39$$

   Por tanto, la mejora máxima teórica haría que el tiempo de respuesta fuese 1,39 veces más rápido que el original.

Ejercicio 2.1.15. Un computador tarda 1000 segundos en ejecutar un proceso de formateo y conversión de imágenes. De todo ese tiempo, el programa dedica un 30 % en hacer operaciones de aritmética en coma flotante, y 250s en accesos al subsistema de discos.

1. Calcule la ganancia en velocidad que se consigue si añadimos al equipo una GPU de 500€ capaz de ejecutar las operaciones en coma flotante 10 veces más rápido.

   Tenemos que:

$$\begin{aligned} T_m &= (1-f)T_o + \frac{fT_o}{10} = T_o\left(1 - f + \frac{f}{10}\right) = \\ &= 0{,}73\ T_o \end{aligned}$$

   Por tanto, la ganancia en velocidad que se consigue es de:

$$S = \frac{T_o}{T_m} = \frac{T_o}{0{,}73\ T_o} = \frac{100}{73} \approx 1{,}37$$

2. Calcule la ganancia en velocidad que se consigue con respecto al tiempo original si remplazamos el subsistema de discos por otro cuyo precio es de 400€ y consigue que los accesos al mismo sean 5 veces más rápidos.

   La francción del tiempo original que se dedica al subsistema de discos es:

$$f = \frac{250}{1000} = 0{,}25$$

   Por tanto, el tiempo de ejecución del proceso con el nuevo subsistema de discos sería:

$$T_m = (1-f)T_o + \frac{fT_o}{5} = T_o\left(1 - f + \frac{f}{5}\right) = \\ = 0{,}8\ T_o$$

   Por tanto, la ganancia en velocidad que se consigue es de:

$$S = \frac{T_o}{T_m} = \frac{T_o}{0{,}8\ T_o} = \frac{5}{4} = 1{,}25$$

3. Calcule la ganancia en velocidad que se consigue utilizando simultáneamente las dos mejoras de los apartados anteriores.

   El tiempo de ejecución del proceso con las dos mejoras sería:

$$T_m = (1 - f_{\text{CPU}} - f_{\text{DD}})T_o + \frac{f_{\text{CPU}}T_o}{10} + \frac{f_{\text{DD}}T_o}{5} = \\ = T_o\left(1 - f_{\text{CPU}} - f_{\text{DD}} + \frac{f_{\text{CPU}}}{10} + \frac{f_{\text{DD}}}{5}\right) = \\ = 0{,}53\ T_o$$

   Por tanto, la ganancia en velocidad que se consigue es de:

$$S = \frac{T_o}{T_m} = \frac{T_o}{0{,}53\ T_o} = \frac{100}{53} \approx 1{,}89$$

4. ¿Qué inversión es la más rentable ateniéndonos únicamente a la relación prestaciones/coste: comprar la GPU, el nuevo sub-sistema de discos o ambos a la vez?

   Calculamos la relación prestaciones/coste de cada una de las mejoras:

   - Para la mejora tan solo de la CPU:

$$\frac{{}^{\text{Prestaciones}}/_{\text{Coste}_{\text{CPUo}}}}{{}^{\text{Prestaciones}}/_{\text{Coste}_{\text{CPUm}}}} = \frac{T_o}{T_m} \cdot \frac{\text{Coste}_{\text{CPUo}}}{\text{Coste}_{\text{CPUm}}} = \frac{T_o}{0{,}73\ T_o} \cdot \frac{1}{500} = \frac{1}{365}$$

   - Para la mejora tan solo del disco:

$$\frac{{}^{\text{Prestaciones}}/_{\text{Coste}_{\text{DDo}}}}{{}^{\text{Prestaciones}}/_{\text{Coste}_{\text{DDm}}}} = \frac{T_o}{T_m} \cdot \frac{\text{Coste}_{\text{DDo}}}{\text{Coste}_{\text{DDm}}} = \frac{T_o}{0{,}8\ T_o} \cdot \frac{1}{400} = \frac{1}{320}$$

- Para la mejora de ambos:

$$\frac{{}^{\text{Prestaciones}}/_{\text{Coste}_{\text{CPUo+DDo}}}}{{}^{\text{Prestaciones}}/_{\text{Coste}_{\text{CPUm+DDm}}}} = \frac{T_o}{T_m} \cdot \frac{\text{Coste}_{\text{CPUo}} + \text{Coste}_{\text{DDo}}}{\text{Coste}_{\text{CPUm}} + \text{Coste}_{\text{DDm}}} = \frac{T_o}{0{,}53\ T_o} \cdot \frac{1}{900} = \frac{1}{477}$$

Por tanto, la mejora más rentable sería la del disco, con una ganancia en prestaciones por euro invertido de ${}^{1}/_{320} = 3{,}125 \cdot 10^{-3}$.

Ejercicio 2.1.16. Después de reemplazar el antiguo disco duro del servidor de base de datos de una pequeña compañía granadina por una nueva unidad SSD, se ha constatado experimentalmente que el proceso principal se ejecuta 1,5 veces más rápido que antes. También se ha medido que ahora dicho proceso consume el 50 % de su tiempo accediendo a esa nueva unidad SSD.

1. Calcule la fracción de tiempo que el proceso consumía antes accediendo al antiguo disco duro.

   Tenemos que:

$$\begin{aligned} 1{,}5 = S &= \frac{T_o}{T_m} \\ T_m &= (1-f)T_o + fT_o \cdot \frac{1}{k} \\ 0{,}5T_m &= fT_o \cdot \frac{1}{k} \end{aligned}$$

   Por tanto:

$$T_m = (1-f)T_o + 0{,}5T_m \Longrightarrow T_m = \frac{(1-f)T_o}{0{,}5}$$

   Sustituyendo en la expresión de la ganancia en velocidad:

$$1{,}5 = S = \frac{T_o}{\frac{(1-f)T_o}{0{,}5}} \Longrightarrow 1{,}5 = \frac{0{,}5}{1-f} \Longrightarrow f = 1 - \frac{0{,}5}{1{,}5} = \frac{2}{3} \approx 0{,}6667$$

   Por tanto, la fracción de tiempo que el proceso consumía antes accediendo al antiguo disco duro era del 66,67 %.

2. ¿Cuántas veces es más rápida la nueva unidad SSD que el antiguo disco duro?

   Tenemos que:

$$1{,}5 = S = \frac{1}{1 - f + \frac{f}{k}} \Longrightarrow \frac{1}{1{,}5} - 1 + f = \frac{f}{k} \Longrightarrow k = \frac{f}{{}^{1}/_{1{,}5} - 1 + f} = 2$$

   Por tanto, la nueva unidad SSD es 2 veces más rápida que el antiguo disco duro.

2.3. Monitorización

Ejercicio 2.3.1. En un sistema Linux se ha ejecutado la orden `uptime` tres veces en momentos diferentes. El resultado, de forma resumida, se muestra en el siguiente listado:

```
... load average: 6.85, 7.37, 7.83
... load average: 8.50, 10.93, 8.61
... load average: 37.34, 9.47, 3.30
```

Indique si la carga crece, decrece, se mantiene estacionaria o bien no puede decidir sobre ello.

No hay una tendencia clara en los valores de las medidas, por lo que no podemos concluir nada realmente. Sería necesario saber cuánto tiempo ha pasado entre cada ejecución para estudiar así en detalle la situación. Tan solo podríamos afirmar que, en el último instante, la carga ha crecido de forma notable respecto a las dos anteriores.

Ejercicio 2.3.2. En un sistema Linux se ha ejecutado la siguiente orden:

```
$ time quicksort
real 0m40.2s
user 0m17.1s
sys 0m3.2s
```

Indique si el sistema está soportando mucha o poca carga. Razone la respuesta.

En este caso, la carga del sistema es considerable, puesto que el tiempo de ejecución del programa `quicksort` es de 20,3 s (suma de los tiempos `user` y `sys`) y el tiempo real que ha tardado en ejecutarse es de 40,2 s. Esto da a entender dos posibilidades:

- El sistema está soportando una carga considerable, ya que el proceso ha estado esperando mucho tiempo para que el sistema le proporcionara los recursos necesarios para su ejecución.
- El proceso `quicksort` ha estado bloqueado por E/S (`I/O blocked`) durante una parte importante de su ejecución, lo que ha provocado que el tiempo real sea mucho mayor que el tiempo de CPU consumido.

Ejercicio 2.3.3. Se sabe que la sobrecarga (*overhead*) de CPU de un monitor software en un determinado servidor es del 4 %. Si el monitor se activa cada 2 segundos, ¿cuánto tiempo tarda el monitor en ejecutarse por cada activación?

Sea T el tiempo que tarda el monitor en ejecutarse por cada activación. Si la sobrecarga es del 4 %, entonces tenemos que:

$$4 = \frac{T}{2} \cdot 100 \Longrightarrow T = \frac{4 \cdot 2}{100} = 0{,}08\,\text{s} = 80\,\text{ms}$$

Ejercicio 2.3.4. A continuación se muestra el resultado obtenido tras ejecutar la orden `top` en un sistema informático que emplea Linux como sistema operativo:

```
2:52pm  up 17 days, 3:41, 1 user, load average: 0.15, 0.27, 0.32
54 processes: 51 sleeping, 3 running, 0 zombie, 0 stopped
%Cpu(s): 23.8% user, 14.0% system, 0.0% nice, 17.0% idle, 45.2% wa
Mem:  257124K av, 253052K used,  4072K free,  8960K shrd, 182972K buff
Swap: 261496K av,  21396K used,240100K free, 26344K cached

  PID USER      PRI  NI  VIRT  RSS  SHARE STAT LC %CPU %MEM   TIME COMMAND
  807 joan       0   0  5708 5708   532   R  N  0 23.0  2.2   6:16 p_exec
  809 joan       0   0  5708 5708   532   R  N  0 14.0  2.2   3:42 p_exec
  185 tomi       0   0   824  824   632   R     0  0.5  0.3   0:00 top
  201 xp         0   0  1272 1208   644   S     0  0.1  0.4   5:49 xp_stat
    1 root       0   0    60   56    36   S     0  0.0  0.0   0:03 init
    2 root       0   0     0    0     0   SW    0  0.0  0.0   0:13 kflushd
    7 root       0   0     0    0     0   SW    0  0.0  0.0   0:00 nfsiod
  194 root       0   0    72    4     4   S     0  0.0  0.0   0:00 migetty
  195 root       0   0    68    0     0   SW    0  0.0  0.0   0:00 migetty
  179 root       0   0   532  312   236   S     0  0.0  0.1   0:00 sndmail
```

1. ¿Cuánta memoria física tiene la máquina?

   Este es el parámetro `Mem: 257124K av`, que indica que la máquina tiene 257124 KiB ≈ 251 MiB de memoria física.

2. ¿Qué porcentaje de la memoria física está marcada como usada según el monitor?

   Este es el parámetro `Mem: 253052K used`, que indica que 253052 KiB ≈ 247 MiB de memoria física está usada. Por lo tanto, el porcentaje de memoria física usada es:

   $$\frac{253052}{257124} \cdot 100 \approx 98{,}41\,\%$$

3. ¿Cuál es la utilización media del procesador?

   Este parámetro se calcula como la totalidad (100 %) menos el tiempo que el procesador está inactivo (reflejado por los parámetros `idle` y `wa`):

   $$100\,\% - 17{,}0\,\% - 45{,}2\,\% = 37{,}8\,\%$$

4. ¿Cómo es la evolución de la carga media del sistema, ascendente o descendente?

   Esta es descendente, y se puede ver en el parámetro `load average: 0.15, 0.27, 0.32`, que indica que la carga media del sistema ha ido disminuyendo desde 0,32 a 0,15.

5. ¿Cuánta memoria física ocupa el monitor?

   La memoria física ocupada por el monitor es la que se muestra en el parámetro `RSS` (*Resident Set Size*) de la línea del proceso `top`, que es de 824 KiB.

Ejercicio 2.3.5. Considere las órdenes siguientes ejecutadas en un sistema Linux:

```
$ time simulador_original
real 0m24.2s
user 0m15.1s
sys 0m1.6s
$ time simulador_mejorado
real 0m32.8s
user 0m10.7s
sys 0m2.1s
```

1. ¿Cuál es el tiempo de ejecución de ambos simuladores?

   El tiempo de ejecución del simulador original es de 16,7 s (suma de los tiempos `user` y `sys`) y el del simulador mejorado es de 12,8 s. Notemos que el tiempo real representa el tiempo total que ha tardado en ejecutarse el programa, incluyendo el tiempo de espera por parte del sistema operativo para que le proporcione los recursos necesarios y el tiempo atendiendo interrupciones de otros procesos. Es por esto que el simulador mejorado tarda más tiempo en ejecutarse que el original, puesto que el sistema en ese momento tiene más carga de trabajo.

2. Calcule, si es el caso, la mejora en el tiempo de ejecución del simulador mejorado respecto del original.

   La mejora en el tiempo de ejecución del simulador mejorado respecto del original es:

   $$\frac{16{,}7}{12{,}8} \approx 1{,}30469$$

   Por lo tanto, el simulador mejorado es aproximadamente 1,3 veces más rápido que el original.

Ejercicio 2.3.6. El monitor `sar` *(system activity reporter)* de un computador se activa cada 15 minutos y tarda 750 ms en ejecutarse por cada activación. Se pide:

1. Calcular la sobrecarga que genera este monitor sobre el sistema informático.

   La sobrecarga que genera el monitor `sar` se calcula como:

   $$\frac{750}{15 \cdot 60 \cdot 10^3} \cdot 100 = \frac{1}{12} \approx 0{,}08333\,\%$$

2. Si la información generada en cada activación ocupa 8192 bytes, ¿la monitorización de cuántos días completos se pueden almacenar en el directorio `/var/log/sysstat` si se dispone únicamente de 200 MiB de capacidad libre?

   Suponemos que la totalidad de capacidad disponible se puede emplear para guardar recursos (en realidad, también ha de guardarse información como el nombre de los ficheros, por ejemplo). Calculemos por tanto en primer lugar los registros de cuántas activaciones pueden almacenarse:

   $$200\,\text{MiB} \cdot \frac{2^{20}\,\text{bytes}}{1\,\text{MiB}} \cdot \frac{1\,\text{registro}}{8192\,\text{bytes}} = 25600\,\text{registros}$$

Calculemos ahora la monitorización de cuántos días completos se pueden almacenar:

$$25600\,\text{registros} \cdot \frac{15\,\text{minutos}}{1\,\text{registro}} \cdot \frac{1\,\text{hora}}{60\,\text{minutos}} \cdot \frac{1\,\text{dia}}{24\,\text{horas}} = 266{,}67\,\text{dias}$$

Por tanto, se pueden almacenar la información de 266 dias completos.

Ejercicio 2.3.7. El día 8 de octubre se ha ejecutado la siguiente orden en un sistema Linux:

```
% ls /var/log/sysstat
-rw-r--r-- 1 root root 3049952 Oct  6 23:50 sa06
-rw-r--r-- 1 root root 3049952 Oct  7 23:50 sa07
-rw-r--r-- 1 root root 2372184 Oct  8 18:40 sa08
```

Suponiendo que la primera muestra se toma a las 0 : 00 de cada día y que `sadc` se ejecuta con un tiempo de muestreo constante, ¿cada cuánto tiempo se activa el monitor `sar`? ¿Cuál es la anchura de entrada del monitor?

Puesto que las últimas activaciones del día 6 y del día 7 se hacen a las 23 : 50 horas, podemos concluir que el monitor se activa cada 10 minutos. Por tanto, el número de veces que se activa el monitor en un día es:

$$\frac{24 \cdot 60}{10} = 144$$

La anchura de entrada del monitor se calcula como el tamaño del fichero dividido por el número de activaciones:

$$\frac{3049952\,\text{bytes}}{144\,\text{activaciones}} \approx 21180{,}22\,{}^{\text{B}}/_{\text{entrada}} \approx 20{,}68\,{}^{\text{KiB}}/_{\text{entrada}}$$

Por tanto, la anchura de entrada del monitor es de aproximadamente 20,68 KiB.

Ejercicio 2.3.8. Indique el resultado que produce la ejecución de las siguientes órdenes sobre un sistema Linux con el monitor `sar` instalado:

1. `sar`

   Informa sobre la utilización del procesador durante el día actual (opción por defecto, `-u`).

2. `sar -A`

   Informa sobre toda la información recogida durante el día actual.

3. `sar -u 1 30`

   Informa sobre la utilización del procesador en el momento actual, mostrando 30 medidas tomadas con un período de un segundo.

4. `sar -uB -f /var/log/sysstat/08`

   Informa sobre la utilización del procesador y paginación de la memoria virtual durante el día 8 del mes.

5. `sar -d -s 12:30:00 -e 18:15:00 -f /var/log/sysstat/08`

   Informa sobre las transferencias de disco desde las 12 : 30 hasta las 18 : 15 horas del día 8 del mes.

Ejercicio 2.3.9. Después de instrumentar un programa con la herramienta `gprof` el resultado obtenido ha sido el siguiente:

```
Flat profile:
Each sample counts as 0.01 seconds.
  %   cumulative   self              self     total
 time   seconds   seconds    calls  s/call   s/call  name
 59.36     27.72    27.72        3    9.24    14.39  reduce
 33.08     43.17    15.45        6    2.57     2.57  invierte
  7.56     46.70     3.53        2    1.76     1.76  calcula
```

El grafo de dependencias muestra que `invierte()` es llamado desde el procedimiento `reduce()`.

1. ¿Cuál es el procedimiento cuyo código propio sería más conveniente optimizar?

   Se debe optimizar la función con mayor tiempo propio (`self seconds`), que en este caso es `reduce()`, ya que consume casi el 60 % del tiempo de CPU del programa.

2. Si el código propio de `reduce()` se sustituye por otro tres veces más rápido, ¿cuánto tiempo tardará en ejecutarse el programa?

   El tiempo que tardaría sería:

   $$27{,}72 \cdot \frac{1}{3} + 15{,}45 + 3{,}53 = 28{,}22\,\text{s}$$

3. Si el procedimiento `invierte()` se sustituye por una nueva versión cuatro veces más rápida, ¿qué mejora se obtendrá en el tiempo de ejecución?

   La mejora sería:

   $$S = \frac{T_o}{T_m} = \frac{27{,}72 + 15{,}45 + 3{,}53}{27{,}72 + \frac{15{,}45}{4} + 3{,}53} \approx 1{,}33$$

4. Calcule cuál es la ganancia en velocidad máxima que se podría conseguir en el tiempo de ejecución mediante la optimización del código del procedimiento `invierte()`.

   Sea k el factor de mejora del procedimiento `invierte()`, entonces la ganancia en velocidad máxima sería:

   $$\lim_{k\to\infty} S = \lim_{k\to\infty} \frac{27{,}72 + 15{,}45 + 3{,}53}{27{,}72 + \frac{15{,}45}{k} + 3{,}53} \approx 1{,}49$$

   Por tanto, la máxima ganancia en velocidad que se podría conseguir es de aproximadamente 1,49.

Ejercicio 2.3.10. Un informático desea evaluar el rendimiento de un computador por medio del benchmark SPEC CPU 2017. Una vez compilados todos los programas del paquete y lanzado su ejecución monitoriza el sistema con la orden `vmstat 1 5`. El resultado de las medidas de este monitor es el siguiente:

```
procs  ----------memory---------  --swap--  ----io---- ---system---  ----cpu----
r  b   swpd   free   buff  cache   si   so    bi    bo   in   cs  us  sy  id  wa
0  0      8  14916  92292 833828    0    0     0     3    0    7   3   1  96   0
1  0      8  14916  92292 833828    0    0     0     0 1022   40 100   0   0   0
3  0      8  14916  92292 833828    2    1    16     3 1016   34  99   1   0   0
1  0      8  14916  92292 833828    0    4     0     8 1035   36  98   2   0   0
2  0      8  14916  92292 833828    1    5     4    28 1035   36  99   1   0   0
```

Indique si, a la vista de los datos anteriores, los resultados obtenidos en la prueba de evaluación serán correctos o no. Justifique la respuesta.

No aparentan ser correctos, puesto que este benchmark tiene como objetivo medir el rendimiento de la CPU, y no se producen entonces intercambio con el disco. Por tanto, deberá haber otros programas en ejecución que los provoquen, alterando los resultados de la medición.

Ejercicio 2.3.11. La monitorización de un programa de dibujo en tres dimensiones mediante la herramienta `gprof` ha proporcionado la siguiente información (por errores en la transmisión hay valores que no están disponibles):

```
Flat profile:
  %     cumulative   self              self     total
 time     seconds   seconds    calls   s/call   s/call  name
xxxxx       xxxxx    15.47        3     5.16     5.16  colorea
xxxxx       xxxxx     1.89        5     0.38     0.38  interpola
xxxxx       xxxxx     1.76        1     1.76     3.65  traza
xxxxx       xxxxx     0.46                            main

Call graph:
index % time    self  children    called     name
[1]    100.0    0.46    19.12               main [1]
               15.47     0.00      3/3          colorea [2]
                1.76     1.89      1/1          traza [3]

               15.47     0.00      3/3          main [1]
[2]     79.0   15.47     0.00      3          colorea [2]

                1.76     1.89      1/1          main [1]
[3]     18.6    1.76     1.89      1          traza [3]
                1.89     0.00      5/5          interpola [4]

                1.89     0.00      5/5          traza [3]
[4]      9.7    1.89     0.00      5          interpola [4]
```

1. ¿En cuánto tiempo se ejecuta el programa de dibujo?

   Hay varias formas de verlo, y evidentemente todas ellas coinciden.

- En el `flat profile`, podemos obtener el tiempo total de ejecución como la suma de los `self seconds`:

$$15{,}47 + 1{,}89 + 1{,}76 + 0{,}46 = 19{,}58\,\text{s}$$

- Desde el `call graph`, como el método `main` se ejecuta el 100 % del tiempo, podemos calcular este tiempo como la suma de su tiempo propio de ejecución más el de sus hijos:

$$0{,}46 + 19{,}12 = 19{,}58\,\text{s}$$

2. Indique cuánto tiempo tarda en ejecutarse el código propio de `main()`.

   El código propio de `main()` se ejecuta en 0,46 s, que es el valor del campo `self seconds` de la línea correspondiente al procedimiento `main()` en el `flat profile`.

3. Establezca la relación de llamadas entre los procedimientos del programa así como el número de veces que se ejecuta cada uno de ellos.

   El procedimiento `main()` llama 3 veces al procedimiento `colorea()` y una vez al procedimiento `traza()`; a su vez, el procedimiento `traza()` llama 5 veces al procedimiento `interpola()`.

4. Calcule el nuevo tiempo de ejecución del programa si se elimina el código propio de `main()` y se reduce a la mitad el tiempo de ejecución del código propio del procedimiento `traza()`.

   El nuevo tiempo de ejecución del programa sería:

$$15{,}47 + 1{,}89 + \frac{1{,}76}{2} + 0 = 18{,}24\,\text{s}$$

5. Proponga y justifique numéricamente una acción sobre el programa original que no afecte el procedimiento `colorea()` (ni su código ni el número de veces que es ejecutado) con el fin de conseguir que el programa se ejecute en 10 segundos.

   Esto no es posible, puesto que el procedimiento `colorea()` ya consume más de 10 segundos de tiempo de ejecución propio, y no se puede reducir su tiempo de ejecución sin afectar a su código o al número de veces que se ejecuta. Por tanto, cualquier acción que se proponga para reducir el tiempo de ejecución del programa afectará necesariamente al procedimiento `colorea()`.

2.4. Análisis comparativo del rendimiento

Ejercicio 2.4.1. En la Tabla 2.1 se muestran los tiempos de ejecución (en segundos) y el número de instrucciones ejecutadas (en millones) en el computador Cleopatra para cinco programas distintos.

Programa	Tiempo [s]	Instrucciones ($\cdot 10^6$)
`asterix`	56	543
`obelix`	59	346
`panoramix`	113	415
`idefix`	132	256
`abraracurcix`	120	235

Tabla 2.1: Tiempos de ejecución e instrucciones ejecutadas en Cleopatra.

1. Calcule el número medio de MIPS (millones de instrucciones por segundo) de este computador al ejecutar los 5 programas.

$$\frac{543 + 346 + 415 + 256 + 235}{56 + 59 + 113 + 132 + 120} = \frac{1795}{480} \approx 3{,}74\,\text{MIPS}$$

2. Determine el número medio de ciclos por instrucción (CPI) obtenidos por este computador. Considere que las instrucciones ejecutadas por los tres primeros programas duran 3 ciclos de media, mientras que las del resto duran 5 ciclos.

$$\frac{3 \cdot 543 + 3 \cdot 346 + 3 \cdot 415 + 5 \cdot 256 + 5 \cdot 235}{543 + 346 + 415 + 256 + 235} = \frac{6367}{1795} \approx 3{,}546\,\text{CPI}$$

Ejercicio 2.4.2. La Tabla 2.2 muestra el tipo y número de las operaciones de coma flotante ejecutadas por un programa de prueba en el computador MATES; la última columna representa el coste computacional en operaciones normalizadas. Se sabe que

Operación	Cantidad ($\cdot 10^9$)	Operaciones normalizadas
`add.s, sub.s`	456	1
`div.s, mul.s`	340	3
`sqrt.s`	180	12
`sqrt.d`	70	15
`log.d`	30	18

Tabla 2.2: Operaciones de coma flotante en MATES.

el programa tarda una hora en ejecutarse. Indique el rendimiento de este computador mediante el uso de MFLOPS (millones de operaciones de coma flotante por segundo) y MFLOPS normalizados. ¿Existe mucha diferencia entre ambos valores?

- MFLOPS:

$$\frac{(456 + 340 + 180 + 70 + 30) \cdot 10^9}{3600 \cdot 10^6} = \frac{1076}{3600} \cdot 10^3 \approx 299\,\text{MFLOPS}$$

- MFLOPS normalizados:

$$\frac{(456 + 3 \cdot 340 + 12 \cdot 180 + 15 \cdot 70 + 18 \cdot 30) \cdot 10^9}{3600 \cdot 10^6} = \frac{5226}{3600} \cdot 10^6 \approx 1451{,}67\,\text{MFLOPS}$$

Como vemos, la diferencia es significante.

Ejercicio 2.4.3. Considere la información (incompleta) obtenida por la orden siguiente en un computador sin más carga que la ejecución de esta orden y sin operaciones de E/S:

```
$ time simulador
real 0m130s
user  -----s
sys 0m5s
```

Se sabe que el número de instrucciones ejecutadas es de $32 \cdot 10^9$; de estas últimas, el 60 % se ejecuta en dos ciclos, mientras que el resto lo hace en cinco ciclos. Calcule el número medio de ciclos por instrucción (CPI) obtenidos por el programa, la frecuencia de funcionamiento del procesador y los MIPS alcanzados por el procesador.

En primer lugar, tenemos que:

$$\frac{32 \cdot 10^9 \cdot (2 \cdot 0{,}6 + 5 \cdot 0{,}4)}{32 \cdot 10^9} = 3{,}2\,\text{CPI}$$

Para calcular los MIPS, puesto que el computador no tiene más carga y no hay operaciones de E/S, sabemos que el tiempo de ejecución es de 130 s. Por tanto:

$$\frac{32 \cdot 10^9}{130 \cdot 10^6} = \approx 246{,}15\,\text{MIPS}$$

Respecto a la frecuencia de funcionamiento del procesador, sabemos que:

$$f = \frac{3{,}2 \cdot 32 \cdot 10^9}{130} \approx 787692307{,}7\,\text{Hz} \approx 0{,}7877\,\text{GHz}$$

Ejercicio 2.4.4. La Tabla 2.3 muestra los tiempos de ejecución de tres programas de un benchmark en tres máquinas diferentes: REF, A y B.

Programa	t_{REF} [s]	t_A [s]	t_B [s]
1	20	12	15
2	20	10	15
3	40	25	16

Tabla 2.3: Tiempos de ejecución de programas en diferentes máquinas.

Calcule, entre la máquina A y la máquina B, cuál presenta mejor rendimiento, según los siguientes criterios:

1. Media aritmética.

$$\bar{t}_A = \frac{12 + 10 + 25}{3} = \frac{47}{3} \approx 15{,}67\,\text{s}$$
$$\bar{t}_B = \frac{15 + 15 + 16}{3} = \frac{46}{3} \approx 15{,}33\,\text{s}$$

   Como $\bar{t}_B < \bar{t}_A$, la máquina B presenta mejor rendimiento para esta carga.

2. Media aritmética ponderada, donde los pesos se escogen de forma inversamente proporcional al tiempo de ejecución de la máquina de referencia REF.

   Los pesos son:

$$C = \frac{1}{1/20 + 1/20 + 1/40} = 8$$
$$w_1 = \frac{C}{20} = 0{,}4$$
$$w_2 = \frac{C}{20} = 0{,}4$$
$$w_3 = \frac{C}{40} = 0{,}2$$

   Las medias ponderadas son:

$$\overline{t_W}_A = 0{,}4 \cdot 12 + 0{,}4 \cdot 10 + 0{,}2 \cdot 25 = 13{,}8\,\text{s}$$
$$\overline{t_W}_B = 0{,}4 \cdot 15 + 0{,}4 \cdot 15 + 0{,}2 \cdot 16 = 15{,}2\,\text{s}$$

   Como $\overline{t_W}_A < \overline{t_W}_B$, la máquina A presenta mejor rendimiento para esta carga (según este criterio).

3. SPEC, usando nuevamente REF como máquina de referencia.

   Calculamos la media geométrica de los tiempos en cada caso:

$$\overline{t_g}_A = \sqrt[3]{12 \cdot 10 \cdot 25} \approx 14{,}42\,\text{s}$$
$$\overline{t_g}_B = \sqrt[3]{15 \cdot 15 \cdot 16} \approx 15{,}33\,\text{s}$$

   Aunque estos valores no son los índices SPEC, sabemos que este es inversamente proporcional a estos. Por tanto, según este criterio, la máquina A es mejor que la máquina B, puesto que la A tendrá mayor índice SPEC.

Ejercicio 2.4.5. La Tabla 2.4 muestra los tiempos de ejecución, en segundos, de los 14 programas de prueba que integran un determinado benchmark empleado para el cálculo del rendimiento en aritmética de coma flotante. En particular, los tiempos corresponden a la máquina de referencia y a una máquina que denominaremos A (columnas "Base" y "Peak", con el mismo significado que usa SPEC para sus comparaciones).

Programa	Referencia	A-Base	A-Peak
`168.wupwise`	1600	419	300
`171.swim`	3100	562	562
`172.mgrid`	1800	607	607
`173.applu`	2100	658	605
`177.mesa`	1400	273	242
`178.galgel`	2900	571	571
`179.art`	2600	1040	1038
`183.equake`	1300	501	387
`187.facerec`	1900	434	434
`188.ammp`	2200	705	697
`189.lucas`	2000	784	758
`191.fma3d`	2100	534	534
`200.sixtrack`	1100	395	336
`301.apsi`	2600	866	866

Tabla 2.4: Tiempos de ejecución de programas en diferentes máquinas.

1. Calcúlense los índices SPECfp_base y SPECfp de la máquina A según el criterio de SPEC.

   El índice SPECfp_base se calcula como la media geométrica de los tiempos de ejecución normalizados respecto a la máquina de referencia:

   $$SPECfp\_base = \sqrt[14]{\frac{1600}{419} \cdot \frac{3100}{562} \cdot \frac{1800}{607} \cdot \cdots \cdot \frac{2100}{534} \cdot \frac{1100}{395} \cdot \frac{2600}{866}} \approx 3{,}48$$

   El índice SPECfp se calcula de forma similar, pero usando los tiempos de ejecución en la máquina A-Peak:

   $$SPECfp = \sqrt[14]{\frac{1600}{300} \cdot \frac{3100}{562} \cdot \frac{1800}{607} \cdot \cdots \cdot \frac{1100}{336} \cdot \frac{2600}{866}} \approx 3{,}74$$

2. Para la columna A-Base, si se considera el tiempo total de ejecución, ¿cuántas veces es más rápida la máquina A que la máquina de referencia?

   Calculemos los tiempos totales:

   $$t_R = 1600 + 3100 + \cdots + 2600 = 28700\,\text{s}$$
   $$t_A = 419 + 562 + \cdots + 866 = 8349\,\text{s}$$

   Por tanto, tenemos que:

   $$\frac{t_R}{t_A} = \frac{28700}{8349} \approx 3{,}4375$$

   Por tanto, la máquina A es aproximadamente 3,44 veces más rápida que la máquina de referencia.

3. ¿Qué mejora del rendimiento se obtiene utilizando las opciones de optimización que ofrece el compilador?

   Para calcular la mejora del rendimiento, comparamos el tiempo total de ejecución de la máquina A-Base con el de la máquina A-Peak:

$$t_{A\_base} = 419 + 562 + \cdots + 866 = 8349\,\text{s}$$
$$t_{A\_peak} = 300 + 562 + \cdots + 866 = 7937\,\text{s}$$

   La mejora del rendimiento se calcula como:

$$\frac{t_{A\_base}}{t_{A\_peak}} = \frac{8349}{7937} \approx 1{,}05$$

   Por tanto, la optimización del compilador permite mejorar el tiempo de ejecución en un 5 % (1,05 veces).

Ejercicio 2.4.6. Considere los tiempos de ejecución, en segundos, obtenidos en los computadores R (referencia), A y B para un conjunto de cinco programas de prueba, mostrados en la Tabla 2.5.

Programa	R [s]	A [s]	B [s]
`tinky-winky`	2600	503	539
`dipsy`	2100	654	762
`laa-laa`	9800	707	716
`po`	2300	748	760
`noo-noo`	1800	363	235

Tabla 2.5: Tiempos de ejecución de programas en diferentes máquinas.

1. Compare el rendimiento de A y B utilizando el tiempo total de ejecución.

   Calculemos ambos tiempos totales de ejecución:

$$t_A = 503 + 654 + 707 + 748 + 363 = 2975\,\text{s}$$
$$t_B = 539 + 762 + 716 + 760 + 235 = 3012\,\text{s}$$

   Calculamos cuántas veces es más rápida la máquina A que la máquina B:

$$\frac{t_B}{t_A} = \frac{3012}{2975} \approx 1{,}0124$$

   Por tanto, la máquina A es aproximadamente 1,0124 veces más rápida que la máquina B.

2. Calcule, a la manera de SPEC, un índice de rendimiento para A y B, y compare el rendimiento de ambas máquinas con este índice. ¿Obtiene los mismos resultados que en el apartado anterior?

Para calcular el índice de rendimiento, usamos la media geométrica de los tiempos de ejecución normalizados respecto a la máquina R:

$$\text{SPEC\_A} = \sqrt[5]{\frac{2600}{503} \cdot \frac{2100}{654} \cdot \frac{9800}{707} \cdot \frac{2300}{748} \cdot \frac{1800}{363}} \approx 5{,}12$$
$$\text{SPEC\_B} = \sqrt[5]{\frac{2600}{539} \cdot \frac{2100}{762} \cdot \frac{9800}{716} \cdot \frac{2300}{760} \cdot \frac{1800}{235}} \approx 5{,}31$$

Ahora, comparamos los índices:

$$\frac{\text{SPEC\_B}}{\text{SPEC\_A}} = \frac{5{,}31}{5{,}12} \approx 1{,}037$$

Por tanto, según el índice SPEC, la máquina B es aproximadamente 1,04 veces mejor que la máquina A.

Ejercicio 2.4.7. La Tabla 2.6 muestra los tiempos de ejecución, en minutos, de una serie de programas de prueba en dos sistemas informáticos SI1 y SI2, junto con el número de instrucciones ejecutadas por cada programa.

Programa	SI1 [min]	SI2 [min]	Instrucciones
`charlie brown`	35	70	$3{,}55 \cdot 10^8$
`lucy`	101	78	$7{,}78 \cdot 10^{13}$
`linus`	57	55	$9{,}12 \cdot 10^7$
`patty`	76	83	$2{,}94 \cdot 10^{10}$

Tabla 2.6: Tiempos de ejecución e instrucciones ejecutadas en SI1 y SI2.

1. Suponiendo que todos los programas tienen la misma importancia en este estudio de evaluación, compárese el rendimiento de estos dos sistemas según:

   a) Media aritmética de los tiempos de ejecución.

$$\bar{t}_{SI1} = \frac{35 + 101 + 57 + 76}{4} = \frac{269}{4} = 67{,}25\,\text{min}$$
$$\bar{t}_{SI2} = \frac{70 + 78 + 55 + 83}{4} = \frac{286}{4} = 71{,}5\,\text{min}$$

   Como $\bar{t}_{SI1} < \bar{t}_{SI2}$, el sistema SI1 es más rápido que el SI2. Calculemos cuántas veces es más rápido:

$$\frac{\bar{t}_{SI2}}{\bar{t}_{SI1}} = \frac{71{,}5}{67{,}25} \approx 1{,}0632$$

   Por tanto, el sistema SI1 es aproximadamente 1,06 veces más rápido que el SI2.

b) MIPS (millones de instrucciones ejecutadas por segundo).

$$\text{MPIS}_{SI1} = \frac{3{,}55 \cdot 10^8 + 7{,}78 \cdot 10^{13} + 9{,}12 \cdot 10^7 + 2{,}94 \cdot 10^{10}}{(35 + 101 + 57 + 76) \cdot 60 \cdot 10^6} = \approx 4822{,}17\,\text{MIPS}$$
$$\text{MPIS}_{SI2} = \frac{3{,}55 \cdot 10^8 + 7{,}78 \cdot 10^{13} + 9{,}12 \cdot 10^7 + 2{,}94 \cdot 10^{10}}{(70 + 78 + 55 + 83) \cdot 60 \cdot 10^6} = \approx 4535{,}54\,\text{MIPS}$$

Por tanto, como $\text{MPIS}_{SI1} > \text{MPIS}_{SI2}$, el sistema SI1 es más rápido que el SI2. Calculemos cuántas veces es más rápido:

$$\frac{\text{MPIS}_{SI1}}{\text{MPIS}_{SI2}} = \frac{4822{,}17}{4535{,}54} \approx 1{,}0632$$

Notemos que, como es de esperar, hemos obtenido el mismo resultado, aunque la primera forma es más sencilla.

2. Repítase la primera parte del estudio suponiendo que los pesos atribuidos a cada programa de prueba son, respectivamente: 0,5, 0,1, 0,1, y 0,3.

   Calculemos la media ponderada de los tiempos de ejecución:

$$\overline{t_W}_{SI1} = 0{,}5 \cdot 35 + 0{,}1 \cdot 101 + 0{,}1 \cdot 57 + 0{,}3 \cdot 76 = 56{,}1\,\text{min}$$
$$\overline{t_W}_{SI2} = 0{,}5 \cdot 70 + 0{,}1 \cdot 78 + 0{,}1 \cdot 55 + 0{,}3 \cdot 83 = 73{,}2\,\text{min}$$

   Por tanto, como $\overline{t_W}_{SI1} < \overline{t_W}_{SI2}$, el sistema SI1 es más rápido que el SI2. Calculemos cuántas veces es más rápido:

$$\frac{\overline{t_W}_{SI2}}{\overline{t_W}_{SI1}} \approx 1{,}305$$

   Por tanto, el sistema SI1 es aproximadamente 1,3 veces más rápido que el SI2.

Ejercicio 2.4.8. La Tabla 2.7 muestra los resultados obtenidos tras la ejecución de tres programas de prueba en un computador con un procesador que dispone de un reloj de 2 GHz. Indique, a partir de los datos anteriores, los siguientes índices de

Programa	Instrucciones ($\cdot 10^9$)	Ciclos por instrucción	Operaciones coma flotante ($\cdot 10^6$)
1	150	3,5	50
2	35	2,8	20
3	250	5,2	175

Tabla 2.7: Resultados de la ejecución de programas de prueba.

rendimiento de este computador:

1. MIPS.

En primer lugar, hemos de calcular el tiempo que tarda en ejecutarse cada programa:

$$t_1 = 150 \cdot 10^9 \text{ instrucciones} \cdot \frac{3{,}5 \text{ ciclos}}{1 \text{ instruccion}} \cdot \frac{1 \text{ s}}{2 \cdot 10^9 \text{ ciclos}} = 262{,}5 \text{ s}$$
$$t_2 = 35 \cdot 10^9 \text{ instrucciones} \cdot \frac{2{,}8 \text{ ciclos}}{1 \text{ instruccion}} \cdot \frac{1 \text{ s}}{2 \cdot 10^9 \text{ ciclos}} = 49 \text{ s}$$
$$t_3 = 250 \cdot 10^9 \text{ instrucciones} \cdot \frac{5{,}2 \text{ ciclos}}{1 \text{ instruccion}} \cdot \frac{1 \text{ s}}{2 \cdot 10^9 \text{ ciclos}} = 650 \text{ s}$$

Ahora, calculamos el número de MIPS:

$$\text{MIPS} = \frac{(150 + 35 + 250) \cdot 10^9}{(262{,}5 + 49 + 650) \cdot 10^6} \approx 452{,}42$$

2. MFLOPS.

$$\text{MFLOPS} = \frac{(50 + 20 + 175) \cdot 10^6}{(262{,}5 + 49 + 650) \cdot 10^6} \approx 0{,}25$$

3. CPI.

   El número medio de ciclos por instrucción (CPI) se calcula como:

$$\text{CPI} = \frac{(150 \cdot 3{,}5 + 35 \cdot 2{,}8 + 250 \cdot 5{,}2) \cdot 10^9}{(150 + 35 + 250) \cdot 10^9} \approx 4{,}42$$

Ejercicio 2.4.9. La página oficial de SPEC muestra en la Tabla 2.8 los resultados de rendimiento para dos sistemas informáticos de la casa comercial ACER obtenidos mediante el conocido benchmark CPU2006.

Sistema	Modelo	SPECint_base2006	SPECint2006
A	Altos G5350 (AMD Opteron 246)	13,47	14,38
B	Altos G5350 (AMD Opteron 254)	17,88	19,18

Tabla 2.8: Resultados de rendimiento de sistemas informáticos ACER.

1. ¿Cuál de los dos sistemas presenta mejor rendimiento? Cuantifique numéricamente la mejora.

   El sistema B presenta mejor rendimiento que el A atendiendo a ambos criterios. Cuantifiquemos la mejora numéricamente:

$$\text{Mejora en SPECint\_base} = \frac{17{,}88}{13{,}47} \approx 1{,}3274$$
$$\text{Mejora en SPECint} = \frac{19{,}18}{14{,}38} \approx 1{,}3338$$

   Por tanto, el sistema B es aproximadamente 1,33 veces mejor que el A en ambos índices.

2. A la vista de los resultados anteriores, ¿afecta al rendimiento de ambos sistemas la optimización llevada a cabo por el compilador en las pruebas de evaluación?

   Sí, la optimización del compilador mejora el rendimiento de ambos sistemas. Calculémosla:

$$S_A = \frac{14{,}38}{13{,}47} \approx 1{,}0675$$
$$S_B = \frac{19{,}18}{17{,}88} \approx 1{,}0721$$

   Por tanto, vemos que hay mejora en ambos casos.

3. ¿En qué medida se reflejará en los resultados anteriores una mejora importante en la unidad de coma flotante (`FPU, floating point unit`) del procesador?

   En teoría no afectarán, puesto que estos índices se realizan empleando programas que trabajan con aritmética entera, como muestra el elemento `int` del nombre.

4. ¿Cuál de los dos sistemas ejecutará el benchmark Whetstone más rápidamente?

   No se puede saber, puesto que este benchmark es de aritmética en coma flotante y los índices mostrados afectan únicamente a la aritmética entera.

Ejercicio 2.4.10. Responda brevemente a las siguientes cuestiones sobre el benchmark CPU2017 que ha desarrollado el consorcio SPEC:

1. ¿Qué componentes del sistema informático evalúa?

   Se evalúa el procesador, pero también se evalúa el sistema de memoria y el compilador (puesto que se proporcionan tan solo los ficheros en código fuente).

2. ¿Cuáles son los lenguajes en que están programados los diferentes programas que lo integran?

   Los programas están escritos en C, C++ y Fortran.

3. ¿Cuál es la diferencia entre el índice CPU2017IntegerSpeed_peak y el índice CPU2017IntegerSpeed_base?

   El índice CPU2017IntegerSpeed_peak se obtiene en base a la ejecución de los programas, pero empleando parámetros de optimización a la hora de compilar específicos para cada uno de ellos, buscando el menor tiempo de ejecución posible en dicha máquina para ese programa.

   Por el contrario, el índice CPU2017IntegerSpeed_base se obtiene con los mismos programas, pero compilados con opciones de compilación genéricas y comunes a todos los programas, sin buscar la optimización específica para cada uno.

4. Indique cómo se calcula el índice CPU2017IntegerSpeed_peak. El método de cálculo empleado, ¿satisface todas las exigencias de un buen índice de prestaciones? Razone la respuesta.

Supongamos que se emplean n programas. Cada programa i se ejecuta 3 veces, y se toma como t_i el valor intermedio (descartándose los valores extremos). Asísmismo, se hace lo mismo en una máquina de referencia, y sea t_i^{REF} el valor de referencia. El valor del índice SPEC correspondiente es:

$$\text{SPEC} = \sqrt[n]{\prod_{i=1}^{n} \frac{t_i^{\text{REF}}}{t_i}}$$

Aunque es un buen índice de prestaciones, no satisface todas las exigencias de un buen índice de prestaciones, ya que no refleja de manera correcta la comparación basada en los tiempos de ejecución debido al uso de la media geométrica.

Ejercicio 2.4.11. En un computador se ha llevado a cabo un estudio para determinar si el tipo de memoria principal es un factor importante en su rendimiento. Para ello se ha medido el tiempo de ejecución de seis programas con dos tipos de memoria: MA (más rápida y más cara) y MB (más lenta y más barata). Las medidas de los tiempos de ejecución (en segundos) de los programas son los mostrados en la Tabla 2.9. Calcule si las diferencias observadas son significativas al 95 % de confianza

Programa	MA [s]	MB [s]
`lucho`	45	48
`lupita`	32	35
`lulila`	51	56
`lurdo`	43	49
`lutecio`	48	51

Tabla 2.9: Tiempos de ejecución de programas con diferentes tipos de memoria.

y, en caso afirmativo, determine la mejora de velocidad conseguida debido al uso del tipo de memoria más rápida.

Observación. Debe emplear la tabla de la distribución T-Student.

Supongamos la hipótesis nula H_0, es decir, que las diferencias no son significativas ($\overline{d}_{\text{real}} = 0$). Entonces, las diferencias d_i siguen una distribución normal de media 0. Calculemos su media y su desviación típica (sabiendo que tiene $5 - 1 = 4$ grados de libertad):

$$\overline{d} = \frac{3+3+5+6+3}{5} = \frac{20}{5} = 4$$

$$s = \sqrt{\frac{1}{4} \cdot \sum_{i=1}^{5} \left(d_i - \overline{d}\right)^2} = \sqrt{\frac{1}{4} \cdot (1+1+1+4+1)} = \sqrt{2}$$

Por tanto, el valor de t_{exp} es:

$$t_{exp} = \frac{\overline{d}}{s/\sqrt{5}} = \frac{4\sqrt{5}}{\sqrt{2}} = 2\sqrt{10} \approx 6{,}3245$$

Ahora, tomamos $\alpha = 0{,}05$ como nivel de significabilidad, y buscamos en la tabla de la distribución T-Student con 4 grados de libertad, y vemos que es 2,7764. Como $|t_{exp}| > 2{,}7764$, podemos rechazar la hipótesis nula H_0 y concluir que las diferencias son significativas al 95 % de confianza. La mejora de velocidad conseguida debido al uso del tipo de memoria más rápida es:

$$S = \frac{48+35+56+49+51}{45+32+51+43+48} = \frac{239}{219} \approx 1{,}091$$

Por tanto, la memoria MA permite obtener una mejora de velocidad del 9,1 % con respecto a la memoria MB.

Ejercicio 2.4.12. La empresa *Facebook* está estudiando dos grandes propuestas con el objetivo de actualizar los computadores personales de su oficina principal en Menlo Park, California. El precio de cada computador es de 1850 € para el Modelo A y 2200 € para el Modelo B. Los responsables informáticos de la empresa han ejecutado los ocho programas que utilizan habitualmente en un computador de cada propuesta, y han obtenido los tiempos de ejecución, expresados en segundos, que se muestran en la Tabla 2.10. Determínese, para un nivel de confianza del 95 %, si

Programa	Modelo A [s]	Modelo B [s]
1	23,6	24,0
2	33,7	41,6
3	10,1	8,7
4	12,9	13,5
5	67,8	66,4
6	9,3	15,2
7	47,4	50,5
8	54,9	52,3

Tabla 2.10: Tiempos de ejecución de programas en los Modelos A y B.

existen diferencias significativas en el rendimiento de los computadores personales de las dos propuestas y qué opción sería mejor.

Observación. Debe emplear la tabla de la distribución T-Student.

Supongamos la hipótesis nula H_0, es decir, que las diferencias no son significativas ($\overline{d}_{\text{real}} = 0$). Entonces, las diferencias d_i siguen una distribución normal de media 0. Calculemos su media y su desviación típica (sabiendo que tiene $8-1=7$ grados de libertad):

$$\overline{d} = \frac{-0{,}4-7{,}9+1{,}4-0{,}6+1{,}4-5{,}9-3{,}1+2{,}6}{8} \approx -1{,}5625$$

$$s = \sqrt{\frac{1}{7}\cdot\sum_{i=1}^{8}\left(d_i-\overline{d}\right)^2} \approx 3{,}751167$$

Por tanto, el valor de t_{exp} es:

$$t_{exp} = \frac{\overline{d}}{s/\sqrt{8}} \approx \frac{-1{,}5625}{3{,}751167/\sqrt{8}} \approx -1{,}178$$

Ahora, tomamos $\alpha = 0{,}05$ como nivel de significabilidad, y buscamos en la tabla de la distribución T-Student con 7 grados de libertad, y vemos que es 2,3646. Como $|t_{exp}| < 2{,}3646$, no podemos rechazar la hipótesis nula H_0 y concluimos que las diferencias no son significativas al 95 % de confianza. Por tanto, hemos de optar por la opción más barata, que es el Modelo A.

Ejercicio 2.4.13. La Tabla 2.11 muestra los tiempos de ejecución (en segundos) medidos en tres computadores, A, B y R, para un conjunto de cinco programas de prueba. Calcule el índice de prestaciones de las máquinas A y B según se hace en

Programa	A [s]	B [s]	R [s]
1	96,2	95,3	103,9
2	13,1	10,2	53,8
3	79,6	67,4	156,3
4	45,2	51,8	98,1
5	88,3	89,3	238,5

Tabla 2.11: Tiempos de ejecución de programas en diferentes máquinas.

el benchmark SPEC_CPU, tomando como referencia la máquina R. Compárese el rendimiento de estas máquinas atendiendo tanto a este índice como al tiempo total de ejecución. ¿Hay diferencias significativas con un grado de confianza del 95 %?

Observación. Debe usar la tabla de la distribución T-Student.

Para calcular el índice de prestaciones SPEC, tenemos que:

$$\text{SPEC\_A} = \sqrt[5]{\frac{103{,}9}{96{,}2} \cdot \frac{53{,}8}{13{,}1} \cdot \frac{156{,}3}{79{,}6} \cdot \frac{98{,}1}{45{,}2} \cdot \frac{238{,}5}{88{,}3}} \approx 2{,}196$$

$$\text{SPEC\_B} = \sqrt[5]{\frac{103{,}9}{95{,}3} \cdot \frac{53{,}8}{10{,}2} \cdot \frac{156{,}3}{67{,}4} \cdot \frac{98{,}1}{51{,}8} \cdot \frac{238{,}5}{89{,}3}} \approx 2{,}3216$$

Ahora, comparemos los índices:

$$\frac{\text{SPEC\_B}}{\text{SPEC\_A}} = \frac{2{,}3216}{2{,}196} \approx 1{,}057$$

Por tanto, según el índice SPEC, la máquina B es aproximadamente 1,06 veces mejor que la máquina A. Ahora, calculemos el tiempo total de ejecución:

$$T_A = 96{,}2 + 13{,}1 + 79{,}6 + 45{,}2 + 88{,}3 = 322{,}4\,\text{s}$$
$$T_B = 95{,}3 + 10{,}2 + 67{,}4 + 51{,}8 + 89{,}3 = 314{,}0\,\text{s}$$

Comparemos ahora los tiempos totales de ejecución:

$$\frac{T_A}{T_B} = \frac{322{,}4}{314{,}0} \approx 1{,}0267$$

Por tanto, la máquina B es aproximadamente 1,03 veces más rápida que la máquina A en términos de tiempo total de ejecución. Veamos ahora si las diferencias son

significativas al 95 % de confianza. Para ello, calculamos las diferencias entre los tiempos de ejecución de A y B:

$$d_1 = 96{,}2 - 95{,}3 = 0{,}9$$
$$d_2 = 13{,}1 - 10{,}2 = 2{,}9$$
$$d_3 = 79{,}6 - 67{,}4 = 12{,}2$$
$$d_4 = 45{,}2 - 51{,}8 = -6{,}6$$
$$d_5 = 88{,}3 - 89{,}3 = -1{,}0$$

Calculamos la media y la desviación típica de las diferencias, sabiendo que hay $5 - 1 = 4$ grados de libertad:

$$\overline{d} = \frac{0{,}9 + 2{,}9 + 12{,}2 - 6{,}6 - 1{,}0}{5} = \frac{8{,}4}{5} = 1{,}68$$
$$s = \sqrt{\frac{1}{4} \cdot \sum_{i=1}^{5} \left(d_i - \overline{d}\right)^2} \approx 6{,}86$$

Por tanto, el valor de t_{exp} es:

$$t_{exp} = \frac{\overline{d}}{s/\sqrt{5}} \approx 0{,}547216$$

Ahora, tomamos $\alpha = 0{,}05$ como nivel de significabilidad, y buscamos en la tabla de la distribución T-Student con 4 grados de libertad, y vemos que es 2,7764. Como $|t_{exp}| < 2{,}7764$, no podemos rechazar la hipótesis nula H_0 y concluimos que las diferencias no son significativas al 95 % de confianza.

Ejercicio 2.4.14. Una gran empresa de seguros está estudiando dos propuestas con el objetivo de actualizar los computadores de su instalación informática. El precio de cada computador es de 1300 € para los de tipo A y 1450 € para los de tipo B. Se estima que el número de computadores a reemplazar es de 75. El ingeniero informático jefe de la empresa ha mandado ejecutar cinco de los programas que utilizan habitualmente en un computador de cada propuesta y ha obtenido los tiempos de ejecución, expresados en segundos, que se muestran en la Tabla 2.12.

Programa	Propuesta A [s]	Propuesta B [s]
1	23,6	24,5
2	33,7	41,6
3	10,1	6,6
4	12,9	13,7
5	67,8	66,4

Tabla 2.12: Tiempos de ejecución de programas en las Propuestas A y B.

1. Calcule el índice de prestaciones de las máquinas A y B según se hace en el benchmark SPEC_CPU, tomando como referencia la máquina A. Según ese

índice, y suponiendo que no hay aleatoriedad en las medidas, ¿qué opción es la mejor? ¿qué opción sería la que compraría ateniéndonos a la relación prestaciones/coste?

Para calcular el índice de prestaciones SPEC, tenemos que:

$$\text{SPEC\_A} = 1$$
$$\text{SPEC\_B} = \sqrt[5]{\frac{23{,}6}{24{,}5} \cdot \frac{33{,}7}{41{,}6} \cdot \frac{10{,}1}{6{,}6} \cdot \frac{12{,}9}{13{,}7} \cdot \frac{67{,}8}{66{,}4}} \approx 1{,}028$$

Por tanto, según el índice SPEC, la máquina B es aproximadamente 1,028 veces mejor que la máquina A. Ahora, calculemos la relación prestaciones/coste, siendo las prestaciones el valor de SPEC:

$$\text{Relación prestaciones/coste}_A = \frac{1}{1300} \approx 7{,}69 \cdot 10^{-4} €^{-1}$$
$$\text{Relación prestaciones/coste}_B = \frac{1{,}028}{1450} \approx 7{,}08 \cdot 10^{-4} €^{-1}$$

Por tanto, la opción A es la mejor opción a comprar ateniéndonos a la relación prestaciones/coste, ya que es más barata y ofrece un mejor rendimiento por euro invertido.

2. Suponiendo que hay aleatoriedad en las medidas, determine si existen diferencias significativas (para un nivel de confianza del 95 %) en el rendimiento de los computadores de las dos propuestas y qué opción sería la que compraría según esa información ateniéndonos a la relación prestaciones/coste. Justifique la respuesta.

   Observación. Debe usar la tabla de la distribución T-Student.

   Supongamos la hipótesis nula H_0, es decir, que las diferencias no son significativas ($\overline{d}_{\text{real}} = 0$). Entonces, las diferencias d_i siguen una distribución normal de media 0. Calculemos su media y su desviación típica (sabiendo que tiene $5 - 1 = 4$ grados de libertad):

$$\overline{d} = \frac{-0{,}9 - 7{,}9 + 3{,}5 - 0{,}8 + 1{,}4}{5} = \frac{-4{,}7}{5} = -0{,}94$$
$$s = \sqrt{\frac{1}{4} \cdot \sum_{i=1}^{5} \left(d_i - \overline{d}\right)^2} \approx 4{,}2910371$$

   Por tanto, el valor de t_{exp} es:

$$t_{exp} = \frac{\overline{d}}{s/\sqrt{5}} \approx \frac{-0{,}94}{4{,}2910371/\sqrt{5}} \approx -0{,}48983$$

   Ahora, tomamos $\alpha = 0{,}05$ como nivel de significabilidad, y buscamos en la tabla de la distribución T-Student con 4 grados de libertad, y vemos que es 2,7764. Como $|t_{exp}| < 2{,}7764$, no podemos rechazar la hipótesis nula H_0 y concluimos que las diferencias no son significativas al 95 % de confianza. Por tanto, a falta de otra información, hemos de optar por la opción más barata, que es la Propuesta A.

Ejercicio 2.4.15. En la empresa KANDOR GRAPHICS están intentando mejorar la técnica de distribución de carga de su servidor principal de streaming de vídeo. Para ello, han realizado cinco medidas de la productividad media del servidor durante un número determinado, pero fijo, de horas para las 2 configuraciones principales de distribución de carga: *Least Connected* (LC) y *Round Robin* (RR). Los resultados, expresados como número medio de MB transmitidos por segundo son los que aparecen en la Tabla 2.13. Como existe aleatoriedad en los experimentos, para poder

Nº Experimento	X_{LC} (MB/s)	X_{RR} (MB/s)
1	157	165
2	125	123
3	172	185
4	152	158
5	165	172

Tabla 2.13: Productividad media del servidor en diferentes configuraciones de distribución de carga.

estar seguros de la decisión, los ingenieros informáticos de la empresa realizaron un test t sobre estos datos, obteniéndose los resultados de la Tabla 2.14.

Med. 1	Med. 2	t	df	p	Dif. media.	ES	IC 90 % para dif. media
LC	RR	−2,64	4	0,057	−6,4	2,42	[−11,56, −1,24]

Tabla 2.14: Resultados del test t para la comparación de configuraciones de distribución de carga.

1. A la vista de los resultados y para un 90 % de confianza, ¿qué configuración de distribución de carga utilizaría y por qué?

   Sea $\alpha = 0{,}1$ el nivel de significación. Como el valor del $p-$value es $0{,}057 < 0{,}1$, podemos rechazar la hipótesis nula H_0 de que las dos configuraciones son equivalentes y concluir que las diferencias son significativas al 90 % de confianza. Empleamos por tanto la que tenga mayor productividad media. Estas son:

   $$\overline{X}_{LC} = \frac{157 + 125 + 172 + 152 + 165}{5} = \frac{771}{5} = 154{,}2\,\text{MB/s}$$
   $$\overline{X}_{RR} = \frac{165 + 123 + 185 + 158 + 172}{5} = \frac{803}{5} = 160{,}6\,\text{MB/s}$$

   Por tanto, la configuración de distribución de carga que utilizaría sería la *Round Robin* (RR), ya que tiene una productividad media mayor. La ganancia en productividad media es de:

   $$\frac{\overline{X}_{RR}}{\overline{X}_{LC}} = \frac{160{,}6}{154{,}2} \approx 1{,}04$$

   Por tanto, la productividad media de la configuración *Round Robin* es aproximadamente un 4 % mayor que la del método *Least Connected*.

2. ¿Qué conclusiones podríamos extraer para un 99 % de confianza? Razone de forma general por qué y en qué medida puede afectar el % de confianza a la decisión a tomar.

   Sea ahora $\alpha = 0{,}01$ el nivel de significación. Como el valor del $p-$value es $0{,}057 > 0{,}01$, no podemos rechazar la hipótesis nula H_0 de que las dos configuraciones son equivalentes y concluimos que las diferencias no son significativas al 99 % de confianza. Por tanto, a este nivel de confianza, no podríamos decidir entre las dos configuraciones, ya que ambas serían equivalentes.

   En general, al aumentar el grado de confianza cada vez es más difícil descartar la hipótesis nula (llegando a ser imposible si se trata del 100 % de confianza).

Ejercicio 2.4.16. En la empresa *KINGSTON* están intentando comprobar la mejora en la latencia de los módulos de memoria RAM que introduce la nueva técnica *HyperX*. Para ello, han realizado 100 experimentos para calcular las latencias medias en múltiples diferentes contextos. Finalmente, para comprobar que las diferencias en las latencias entre el método tradicional y la técnica *HyperX* no se deben a efectos aleatorios, han realizado un test t, cuyos resultados son los que aparecen en la Tabla 2.15.

Med. 1	Med. 2	t	*df*	p	Dif. media.	ES	IC 95 % para dif. media
NORMAL	HYPERX	4,59	99	0,000	0,33	0,073	[0,188, 0,476]

Tabla 2.15: Resultados del test t para la comparación de latencias de módulos de memoria.

1. Mirando únicamente valores medios, ¿qué técnica parece ser mejor? Justifique la respuesta.

   Tenemos que la diferencia media es positiva, por lo que notando por $\overline{N}$ a la latencia media de la RAM normal y por $\overline{X}$ a la latencia media de la RAM mejorada, tenemos que:

   $$\overline{N - X} = \overline{N} - \overline{X} = 0{,}33 \Longrightarrow \overline{N} > \overline{X}$$

   Como una latencia media mayor supone un rendimiento peor, la técnica mejorada parece ser mejor.

2. Al 95 % de nivel de confianza, ¿son significativas esas diferencias? Justifique la respuesta.

   Como $0 \notin [0{,}188, 0{,}476]$, tenemos que podemos rechazar la hipótesis nula, luego las diferencias son significativas al 95 % de confianza.

Ejercicio 2.4.17. En la empresa *SERENDIPITY S.L.* están intentando mejorar el servidor web que alberga las páginas de la Universidad de Granada. Para ello, han ejecutado un conocido benchmark de servidores web para 5 configuraciones distintas del S.O. actualmente en uso. Como la fuente de variabilidad es alta debido a que las

pruebas han tenido que realizarlas en el equipo ya actualmente en uso (se ha elegido el intervalo entre las 4 y las 5 de la mañana en días sucesivos) los experimentos se han realizado 10 veces. Los resultados del número medio de páginas servidas por segundo se muestran en la Tabla 2.16, y los resultados del análisis ANOVA son los que aparecen en la Tabla 2.17 y Tabla 2.18.

Exp.	Conf. 1	Conf. 2	Conf. 3	Conf. 4	Conf. 5
1	152	155	178	162	178
2	162	152	185	157	179
3	165	163	182	153	181
4	159	162	189	158	182
5	148	154	190	162	189
6	152	152	186	158	183
7	156	158	195	152	188
8	160	160	185	149	178
9	163	152	194	149	182
10	153	155	197	150	181

Tabla 2.16: Número medio de páginas servidas por segundo en diferentes configuraciones del S.O.

Cases	Sum of Squares	*df*	Mean Square	*F*	*p*
Config.	10292,600	4	2573,150	102,299	$<$,001
Residuals	1131,900	45	25,153		

Tabla 2.17: ANOVA - Productividad (X)

1. Si atendiéramos exclusivamente a la media aritmética de los resultados, ¿qué configuración parecería la mejor?

   Tenemos que las medias aritméticas de número medio de páginas servidas por segundo es de:

   $$\overline{C}_1 = 157 \qquad \overline{C}_2 = 156{,}3 \qquad \overline{C}_3 = 188{,}1 \qquad \overline{C}_4 = 155 \qquad \overline{C}_5 = 182{,}1$$

   Por tanto, la mejor configuración sería la tercera.

2. Para un nivel de confianza del 95 %, ¿afecta la configuración del S.O. al rendimiento del equipo?

3. Para un nivel de confianza del 95 %, agrupe las configuraciones que afectan estadísticamente por igual. ¿Cuáles serían, en ese caso, las mejores configuraciones? ¿Y para un nivel de confianza del 90 %? Explique razonadamente los resultados.

		Mean Difference	Lower (95 % CI)	Upper (95 % CI)	*SE*	*t*	*p*
1	2	0,700	−5,673	7,073	2,243	0,312	0,998
	3	−31,100	−37,473	−24,727	2,243	−13,866	< ,001
	4	2,000	−4,373	8,373	2,243	0,892	0,898
	5	−25,100	−31,473	−18,727	2,243	−11,191	< ,001
2	3	−31,800	−38,173	−25,427	2,243	−14,178	< ,001
	4	1,300	−5,073	7,673	2,243	0,580	0,977
	5	−25,800	−32,173	−19,427	2,243	−11,503	< ,001
3	4	33,100	26,727	39,473	2,243	14,758	< ,001
	5	6,000	−0,373	12,373	2,243	2,675	0,074
4	5	−27,100	−33,473	−20,727	2,243	−12,082	< ,001

Tabla 2.18: Post Hoc Comparisons - Config.

Ejercicio 2.4.18. Un estudiante de Ingeniería de Servidores ha realizado un estudio sobre la influencia del parámetro *swappiness* del Sistema Operativo Linux sobre las prestaciones de su servidor Web. Para ello, ha realizado un total de 10 experimentos, calculando el número máximo de conexiones simultáneas que su servidor Apache es capaz de manejar, para dos valores concretos de dicho parámetro a los que ha llamado *BAJO* y *ALTO*. Para poder estar seguro de que la diferencia entre las medias de los valores medidos sea estadísticamente significativa, este estudiante ha realizado un test t, obteniéndose los resultados de la Tabla 2.19.

Med. 1	Med. 2	*t*	*df*	*p*	Mean Dif.	SE Dif.	95 % CI for Mean Difference
BAJO	ALTO	2,11	9	0,064	8,3	3,93	Lower: −0,6, Upper: 17,2

Tabla 2.19: Resultados del test t para el parámetro *swappiness*.

1. A la vista de los resultados y para un 90 % de confianza, ¿qué método utilizaría y por qué?

   Como el valor del p-value es $p = 0{,}064 < 0{,}1$, podemos rechazar la hipótesis nula H_0 de que las dos configuraciones son equivalentes y concluir que las diferencias son significativas al 90 % de confianza. Empleamos por tanto la que tenga mayor productividad media. Puesto que la diferencia media es positiva, tenemos que optamos por la configuración *swappiness="BAJO"* ya que permite, de media, 8,3 más conexiones que con el parámetro de *swappiness="ALTO"*.

2. ¿Qué conclusiones podríamos extraer para un 95 % de confianza? ¿Y para un 99 %?

   Como $0 \in [-0{,}6, 17{,}2]$, no podemos rechazar la hipótesis nula H_0 de que las dos configuraciones son equivalentes y concluimos que las diferencias no son significativas al 95 % de confianza, por lo que tampoco lo serían al 99 %. Por tanto, a este nivel de confianza, no podríamos decidir entre las dos configuraciones, ya que ambas serían equivalentes.

Ejercicio 2.4.19. Determine, al 95 % de nivel de confianza, un intervalo en el que se debe encontrar el tiempo medio de ejecución de un determinado programa escrito en Python. En la Tabla 2.20 se muestran los tiempos de ejecución obtenidos en 10 experimentos independientes.

Observación. Debe usar la tabla de la distribución T-Student.

Experimento	Tiempo ejecución (s)
1	15,2
2	16,2
3	16,5
4	15,9
5	14,8
6	15,2
7	15,6
8	16,0
9	16,3
10	15,3

Tabla 2.20: Tiempos de ejecución de un programa en Python.

2.5. Optimización del rendimiento

Ejercicio 2.5.1. 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? Exprese el resultado en peticiones por segundo.

Sabemos que $R_0 = 15\,\text{ms}$ y $N_0 = 345$ trabajos. Entonces, Por la Ley de Little la productividad del servidor es:

$$X_0 = \frac{N_0}{R_0} = \frac{345}{15 \cdot 10^{-3}} = 23000\,{}^{\text{peticiones}}/_{\text{s}}$$

Ejercicio 2.5.2. 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.

Tenemos que $C_0 = 120$ peticiones, $T = 60$ segundos y $N_0 = 5$ peticiones. En primer lugar, calculamos la productividad del servidor:

$$X_0 = \frac{C_0}{T} = \frac{120}{60} = 2\,{}^{\text{peticiones}}/_{\text{s}}$$

Por tanto, el tiempo medio de respuesta del servidor es:

$$R_0 = \frac{N_0}{X_0} = \frac{5}{2} = 2{,}5\,\text{s}$$

Ejercicio 2.5.3. 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?

Sabemos que $Z = 17\,\text{s}$, $X_0 = 2{,}5\,{}^{\text{peticiones}}/_{\text{s}}$ y $N_0 = 45$ trabajos. Por la Ley del Tiempo de Respuesta Interactivo, el tiempo medio de respuesta del computador es:

$$R_0 = \frac{N_0}{X_0} - Z = \frac{45}{2{,}5} - 17 = 18\,\text{s} - 17\,\text{s} = 1\,\text{s}$$

Ejercicio 2.5.4. Durante una sesión de medida de media hora un monitor software ha extraído las variables operacionales básicas de un disco duro mostradas en la Tabla 2.21. A partir de la información anterior calcule las siguientes variables

Variable	Valor
A_i	364 peticiones
C_i	359 peticiones
B_i	23 minutos

Tabla 2.21: Variables operacionales del disco duro.

operacionales deducidas del disco duro:

1. Tasa de llegada de peticiones al disco duro.

   Del enunciado, deducimos que $T = 30\,\text{min} = 1800\,\text{s}$. Por tanto, la tasa de llegada de peticiones al disco duro es:

   $$\lambda_i = \frac{A_i}{T} = \frac{364}{1800} = 0{,}2022\,{}^{\text{peticiones}}/_{\text{s}}$$

2. Productividad del disco duro.

   La productividad del disco duro es:

   $$X_i = \frac{C_i}{T} = \frac{359}{1800} = 0{,}1994\,{}^{\text{peticiones}}/_{\text{s}}$$

3. Utilización del disco duro.

   La utilización del disco duro es:

   $$U_i = \frac{B_i}{T} = \frac{23 \cdot 60}{1800} = \frac{23}{30} = 0{,}7667 = 76{,}67\,\%$$

4. Tiempo medio de servicio del disco duro.

   El tiempo medio de servicio del disco duro es:

   $$S_i = \frac{B_i}{C_i} = \frac{23 \cdot 60}{359} = 3{,}84\,{}^{\text{s}}/_{\text{trabajo}}$$

Ejercicio 2.5.5. 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 mostrados en la Tabla 2.22. Determine el tiempo medio de respuesta de una petición a este sistema

Estación	V_i	R_i
Procesador	7	4.3
Disco 1	2	1.5
Disco 2	4	2.3

Tabla 2.22: Tiempos de respuesta y razones de visita del sistema informático.

informático. Si el número medio de peticiones activas en el sistema es 80, ¿cuál es la tasa de llegadas que soporta?

Por la Ley General del Tiempo de Respuesta, el tiempo medio de respuesta del sistema es:

$$R_0 = \sum_{i=1}^{3} V_i \cdot R_i = 7 \cdot 4{,}3 + 2 \cdot 1{,}5 + 4 \cdot 2{,}3 = 42{,}3\,\text{ms}$$

Por la Ley de Little y suponiendo que $N_0 = 80$ peticiones y que el sistema está en estado estacionario, la tasa de llegadas al sistema es:

$$\lambda_0 = X_0 = \frac{N_0}{R_0} = \frac{80}{42{,}3} = 1{,}89\,{}^{\text{peticiones}}/_{\text{ms}}$$

Ejercicio 2.5.6. 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.

   Los datos dados son:

   $$T = 120\,\text{s}, \quad B_i = 78\,\text{s}, \quad A_i = 84\,\text{peticiones}, \quad C_i = 82\,\text{peticiones}$$

   La productividad del subsistema de disco es:

   $$X_i = \frac{C_i}{T} = \frac{82}{120} = 0{,}6833\,{}^{\text{peticiones}}/_{\text{s}}$$

   La utilización del subsistema de disco es:

   $$U_i = \frac{B_i}{T} = \frac{78}{120} = 0{,}65 = 65\,\%$$

2. 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 los datos anteriores, calcule la productividad del servidor y el tiempo medio de respuesta de una interacción con el servidor.

   En este caso, el número de peticiones activas en el servidor es $N_0 = 13$ y la razón de visita al subsistema de disco es $V_i = 5$. Por la Ley del Flujo Forzado, la productividad del servidor es:

   $$X_0 = \frac{X_i}{V_i} = \frac{0{,}6833}{5} = 0{,}1367\,{}^{\text{peticiones}}/_{\text{s}}$$

   El tiempo medio de respuesta del servidor, por la Ley de Little, es:

   $$R_0 = \frac{N_0}{X_0} = \frac{13}{0{,}1367} = 95{,}1\,\text{s}$$

Ejercicio 2.5.7. 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.

1. Calcule la utilización media del procesador debida al procesamiento de pedidos.

   El tiempo de ocupación del procesador en T minutos es de:

   $$B_i = 450 \cdot T \cdot 0{,}2 \cdot 0{,}6 = 54T\,\text{s}$$

   Por tanto, la utilización del procesador es:

   $$U_i = \frac{B_i}{T} = \frac{54T}{T \cdot 60} = \frac{54}{60} = 0{,}9 = 90\,\%$$

2. ¿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?

   El nuevo tiempo de ocupación del procesador en T minutos es de:

   $$B_i = 450 \cdot T \cdot 0{,}2 \cdot \frac{0{,}6}{2{,}5} = 21{,}6T\,\text{s}$$

   Por tanto, la nueva utilización del procesador es:

   $$U_i = \frac{B_i}{T} = \frac{21{,}6T}{T \cdot 60} = \frac{21{,}6}{60} = 0{,}36 = 36\,\%$$

3. ¿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?

   El nuevo tiempo de ocupación del procesador en T minutos es de:

   $$B_i = 450 \cdot T \cdot 0{,}2 \cdot 2 \cdot 0{,}6 = 108T\,\text{s}$$

   Esto no es realista, puesto que $B_i \leqslant 60T\,\text{s}$, por lo que el servidor estaría saturado. Por tanto, la utilización del procesador sería $U_i = 1$ (100 %).

Ejercicio 2.5.8. 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 Tabla 2.23. A partir

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

Tabla 2.23: Tiempos de servicio y razones de visita del servidor web.

de la información anterior determine:

1. La demanda de servicio de cada dispositivo (D_i).

   $$D_1 = V_1 \cdot S_1 = 8 \cdot 0{,}01 = 0{,}08\,{}^{\text{s}}/_{\text{peticion}}$$
   $$D_2 = V_2 \cdot S_2 = 4 \cdot 0{,}04 = 0{,}16\,{}^{\text{s}}/_{\text{peticion}}$$
   $$D_3 = V_3 \cdot S_3 = 3 \cdot 0{,}03 = 0{,}09\,{}^{\text{s}}/_{\text{peticion}}$$

2. El tiempo de respuesta del servidor web (R_0).

   Ley General del Tiempo de Respuesta:

   $$R_0 = \sum_{i=1}^{3} V_i \cdot R_i = 8 \cdot 0{,}0147 + 4 \cdot 0{,}1111 + 3 \cdot 0{,}0469 = 0{,}7027\,\text{s}$$

3. El número medio de peticiones en el servidor web (N_0).

   Por la Ley de Little, el número medio de peticiones en el servidor web es:

   $$N_0 = \lambda_0 \cdot R_0 = 4 \cdot 0{,}7027 = 2{,}8108 \,\text{peticiones}$$

4. La productividad de cada dispositivo (X_i).

   Como la tasa de llegadas al servidor web es $\lambda_0 = 4 \,{}^{\text{peticiones}}/_{\text{s}}$ y suponemos que el servidor está en estado estacionario, tenemos que:

   $$X_0 = \lambda_0 = 4 \,{}^{\text{peticiones}}/_{\text{s}}$$

   Por tanto, por la Ley del Flujo Forzado, la productividad de cada dispositivo es:

   $$X_1 = X_0 \cdot V_1 = 4 \cdot 8 = 32 \,{}^{\text{peticiones}}/_{\text{s}}$$
   $$X_2 = X_0 \cdot V_2 = 4 \cdot 4 = 16 \,{}^{\text{peticiones}}/_{\text{s}}$$
   $$X_3 = X_0 \cdot V_3 = 4 \cdot 3 = 12 \,{}^{\text{peticiones}}/_{\text{s}}$$

5. La utilización de cada dispositivo (U_i).

   Por la Relación Utilización-Demanda de Servicio, la utilización de cada dispositivo es:

   $$U_1 = X_0 \cdot D_1 = X_0 \cdot V_1 \cdot S_1 = 4 \cdot 8 \cdot 0{,}01 = 0{,}32$$
   $$U_2 = X_0 \cdot D_2 = X_0 \cdot V_2 \cdot S_2 = 4 \cdot 4 \cdot 0{,}04 = 0{,}64$$
   $$U_3 = X_0 \cdot D_3 = X_0 \cdot V_3 \cdot S_3 = 4 \cdot 3 \cdot 0{,}03 = 0{,}36$$

Ejercicio 2.5.9. 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?

Puesto que un servidor web es una red abierta, tenemos que la productividad máxima del servidor se produce cuando el cuello de botella tiene utilización 1. Por tanto:

$$1 = U_b = X_0^{\max} \cdot D_b \Longrightarrow D_b = \frac{1}{X_0^{\max}} = \frac{1}{25} = 0{,}04 \,{}^{\text{s}}/_{\text{peticion}}$$

Como la demanda de servicio del procesador es $D_p = 0{,}02$ segundos, este no es el cuello de botella del servidor, sino que lo es el disco duro. Por tanto, la demanda de servicio del disco duro es $D_d = 0{,}04$ segundos.

Ejercicio 2.5.10. 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:

1. ¿Cuál es la productividad media del servidor y cuántos trabajos se encuentran, de media, en reflexión?

   En este caso, contamos con $N_T = 3000$ trabajos, $Z = 20$ segundos y $R_0 = 10$ segundos. Por la Ley de Little, la productividad media del servidor es:

   $$X_0 = \frac{N_T}{R_0 + Z} = \frac{3000}{10 + 20} = 100\,{}^{\text{trabajos}}/_{\text{s}}$$

   El número medio de trabajos en reflexión es:

   $$N_z = Z \cdot X_0 = 20 \cdot 100 = 2000 \text{ trabajos}$$

2. Si se quiere conseguir una productividad de 125 trabajos por segundo, ¿qué tiempo de respuesta debería tener el servidor?

   Por la Ley de Little, si se quiere conseguir una productividad de 125 trabajos por segundo, el tiempo medio de respuesta del servidor debería ser:

   $$R_0 = \frac{N_T}{X_0} - Z = \frac{3000}{125} - 20 = 4\,\text{s}$$

3. ¿Qué habría que hacer para conseguir una productividad de $200\,{}^{\text{trabajos}}/_{\text{s}}$?

   El servidor está configurado con $N_T = 3000$ trabajos y $Z = 20$ segundos. Por la Ley de Little, si se quiere conseguir una productividad de $200\,{}^{\text{trabajos}}/_{\text{s}}$, el tiempo medio de respuesta del servidor debería ser:

   $$R_0 = \frac{N_T}{X_0} - Z = \frac{3000}{200} - 20 = -5\,\text{s}$$

   Como vemos, no tiene sentido alguno. En el caso límite de que $R_0 = 0\,\text{s}$, la productividad del servidor sería:

   $$X_0 = \frac{N_T}{Z} = \frac{3000}{20} = 150\,{}^{\text{trabajos}}/_{\text{s}}$$

Ejercicio 2.5.11. 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 Tabla 2.24. Sabiendo que

Dispositivo	V_i	S_i
Procesador (1)	7	0.1
Disco (2)	3	0.025
Disco (3)	1	0.050
Disco (4)	2	0.035

Tabla 2.24: Tiempos de servicio y razones de visita del sistema informático.

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:

1. Calcúlense las demandas de servicio de cada dispositivo.

   Tenemos que:

$$D_1 = V_1 \cdot S_1 = 7 \cdot 0{,}1 = 0{,}7\,{}^{\text{s}}/_{\text{trabajo}}$$
$$D_2 = V_2 \cdot S_2 = 3 \cdot 0{,}025 = 0{,}075\,{}^{\text{s}}/_{\text{trabajo}}$$
$$D_3 = V_3 \cdot S_3 = 1 \cdot 0{,}05 = 0{,}05\,{}^{\text{s}}/_{\text{trabajo}}$$
$$D_4 = V_4 \cdot S_4 = 2 \cdot 0{,}035 = 0{,}07\,{}^{\text{s}}/_{\text{trabajo}}$$

2. ¿Cuál es el número medio de clientes (=trabajos) que están en reflexión?

   Por la Ley de Little, el número medio de clientes en reflexión es:

$$N_z = Z \cdot X_0 = 6 \cdot 1{,}2 = 7{,}2\,\text{trabajos}$$

3. ¿Cuántos clientes están conectados de media en el servidor?

   Por la Ley de Little, el número medio de clientes conectados al servidor es:

$$N_0 = N_T - N_z = 10 - 7{,}2 = 2{,}8\,\text{trabajos}$$

4. ¿Cuál es el tiempo de respuesta del servidor?

   Por la Ley del Tiempo de Respuesta Interactivo, el tiempo de respuesta del servidor es:

$$N_T = X_0 \cdot (R_0 + Z) \Longrightarrow R_0 = \frac{N_T}{X_0} - Z = \frac{10}{1{,}2} - 6 = 2{,}3333\,\text{s}$$

5. Calcúlense, para cada dispositivo del servidor, la productividad y la utilización.

   Por la Ley del Flujo Forzado, la productividad de cada dispositivo es:

$$X_1 = X_0 \cdot V_1 = 1{,}2 \cdot 7 = 8{,}4\,{}^{\text{trabajos}}/_{\text{s}}$$
$$X_2 = X_0 \cdot V_2 = 1{,}2 \cdot 3 = 3{,}6\,{}^{\text{trabajos}}/_{\text{s}}$$
$$X_3 = X_0 \cdot V_3 = 1{,}2 \cdot 1 = 1{,}2\,{}^{\text{trabajos}}/_{\text{s}}$$
$$X_4 = X_0 \cdot V_4 = 1{,}2 \cdot 2 = 2{,}4\,{}^{\text{trabajos}}/_{\text{s}}$$

   Por la Relación Utilización-Demanda de Servicio, la utilización de cada dispositivo es:

$$U_1 = X_0 \cdot D_1 = X_0 \cdot V_1 \cdot S_1 = 1{,}2 \cdot 7 \cdot 0{,}1 = 0{,}84$$
$$U_2 = X_0 \cdot D_2 = X_0 \cdot V_2 \cdot S_2 = 1{,}2 \cdot 3 \cdot 0{,}025 = 0{,}09$$
$$U_3 = X_0 \cdot D_3 = X_0 \cdot V_3 \cdot S_3 = 1{,}2 \cdot 1 \cdot 0{,}05 = 0{,}06$$
$$U_4 = X_0 \cdot D_4 = X_0 \cdot V_4 \cdot S_4 = 1{,}2 \cdot 2 \cdot 0{,}035 = 0{,}08$$

Ejercicio 2.5.12. Los parámetros del modelo de un sistema informático transaccional (red abierta) son los que se muestran en la Tabla 2.25 (los tiempos se expresan en milisegundos). La tasa de llegadas al sistema es de 0.15 transacciones por milisegundo.

Dispositivo	S_i	V_i
Procesador (1)	0.4	9
Disco (2)	0.5	8

Tabla 2.25: Parámetros del sistema informático transaccional.

1. Identifique el cuello de botella del sistema.

   Calculamos la demanda de servicio de cada dispositivo:

   $$D_1 = V_1 \cdot S_1 = 9 \cdot 0{,}4 = 3{,}6\,{}^{\text{ms}}/_{\text{transaccion}}$$
   $$D_2 = V_2 \cdot S_2 = 8 \cdot 0{,}5 = 4\,{}^{\text{ms}}/_{\text{transaccion}}$$

   El cuello de botella del sistema es por tanto el disco, ya que su demanda de servicio es mayor que la del procesador.

2. ¿Cuál es la utilización del cuello de botella?

   Por la Relación Utilización-Demanda de Servicio, la utilización del cuello de botella es:

   $$U_b = X_0 \cdot D_b = \lambda_0 \cdot D_b = 0{,}15 \cdot 4 = 0{,}6 = 60\,\%$$

3. Calcule la productividad máxima del sistema.

   La productividad máxima del sistema se alcanza cuando el cuello de botella tiene una utilización de 1. Por la Relación Utilización-Demanda de Servicio:

   $$1 = U_b = X_0^{\max} \cdot D_b \Longrightarrow X_0^{\max} = \frac{1}{D_b} = \frac{1}{4} = 0{,}25\,{}^{\text{transacciones}}/_{\text{ms}}$$

4. Determine el tiempo mínimo de respuesta de una transacción.

   Por la Ley General del Tiempo de Respuesta, el tiempo mínimo de respuesta de una transacción es:

   $$R_0^{\min} = \sum_{i=1}^{2} V_i \cdot R_i = \sum_{i=1}^{2} V_i \cdot (S_i + \cancelto{0}{W_i}) = \sum_{i=1}^{2} V_i \cdot S_i$$
   $$= 9 \cdot 0{,}4 + 8 \cdot 0{,}5 = 3{,}6 + 4 = 7{,}6\,\text{ms}$$

   Por tanto, el tiempo mínimo de respuesta de una transacción es 7,6 ms.

Ejercicio 2.5.13. 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

Tabla 2.26: Parámetros del sistema informático interactivo.

1. Identifique el cuello de botella.

   La demanda de servicio de cada dispositivo es:

   $$D_1 = V_1 \cdot S_1 = 4 \cdot 0{,}5 = 2\,{}^{\text{s}}/_{\text{trabajo}}$$
   $$D_2 = V_2 \cdot S_2 = 3 \cdot 0{,}75 = 2{,}25\,{}^{\text{s}}/_{\text{trabajo}}$$

   El cuello de botella del sistema es la cinta, ya que su demanda de servicio es mayor que la del procesador.

2. Determine el tiempo mínimo de respuesta del servidor.

   Hay dos opciones, que el servidor esté sometido a una carga baja o alta.

   - Si el servidor está sometido a baja carga, el tiempo mínimo de respuesta viene determinado por la Ley General del Tiempo de Respuesta:

   $$R_0^{\min} = \sum_{i=1}^{2} V_i \cdot R_i = \sum_{i=1}^{2} V_i \cdot (S_i + \cancelto{0}{W_i}) = \sum_{i=1}^{2} V_i \cdot S_i = \sum_{i=1}^{2} D_i = 2 + 2{,}25 = 4{,}25\,\text{s}$$

   - Si el servidor está sometido a alta carga, la productividad máxima del servidor es:

   $$X_0^{\max} = \frac{1}{D_2} = \frac{1}{2{,}25} = 0{,}4444\,{}^{\text{trabajos}}/_{\text{s}}$$

   Por la Ley de Little, el tiempo medio de respuesta del servidor es:

   $$R_0^{\min} = \frac{N_T}{X_0^{\max}} - Z = \frac{25}{0{,}4444} - 6 = 50{,}25\,\text{s}$$

   Para ver si se encuentra en baja o alta carga sabiendo que $N_T = 25$ trabajos, calculamos el punto teórcico de saturación (knee point). Este se produce cuando:

   $$4{,}25 = \frac{N_T^*}{0{,}4444} - 6 \Longrightarrow N_T^* = 0{,}4444 \cdot (4{,}25 + 6) = 4{,}55555\,\text{trabajos}$$

   Como $N_T = 25 > N_T^* = 4{,}55555$, el servidor se encuentra sometido a alta carga, por lo que el tiempo mínimo de respuesta del servidor es $R_0^{\min} = 50{,}25$ segundos.

   Observación. Notemos que, si tuviésemos libertad con el número de clientes, podríamos reducir el número de clientes y obtener así que el tiempo mínimo de respuesta del servidor fuese $R_0^{\min} = 4{,}25$ segundos, que es el tiempo mínimo de respuesta absoluto del servidor.

3. ¿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?

   Como hemos visto en el apartado anterior, el punto teórico de saturación es $N_T^* = 4{,}55555$ trabajos. Como $N_T = 25 > N_T^*$, el servidor se encuentra sometido a alta carga.

4. Indique las ecuaciones de los límites optimistas del tiempo de respuesta y de la productividad.

   Como hemos visto antes, el tiempo mínimo de respuesta del servidor es:

   $$R_0^{\min} = \text{máx}\{4{,}25, 2{,}25 \cdot N_T - 6\}\text{segundos}$$

   Por otro lado, distinguimos casos para la productividad del servidor:

   - Si el servidor está en baja carga, la productividad del servidor viene determinada por la Ley de Little:

     $$X_0^{\max} = \frac{N_T}{R_0^{\min} + Z} = \frac{N_T}{4{,}25 + 6} = \frac{N_T}{10{,}25}$$

   - Si el servidor está en alta carga, la productividad del servidor viene determinada por la Relación Utilización-Demanda de Servicio:

     $$1 = U_b = X_0^{\max} \cdot D_b \Longrightarrow X_0^{\max} = \frac{1}{D_b} = \frac{1}{2{,}25} = 0{,}4444\,{}^{\text{trabajos}}/_{\text{s}}$$

   Por tanto, los límites optimistas de la productividad del servidor son:

   $$X_0^{\max} = \text{mín}\left\{\frac{N_T}{10{,}25}, 0{,}4444\right\}{}^{\text{trabajos}}/_{\text{s}}$$

Ejercicio 2.5.14. Considere la información referida al modelo de un sistema informático donde los tiempos se expresan en milisegundos reflejada en la Tabla 2.27. El sistema recibe una media de 18 peticiones por segundo durante el mediodía, que

Dispositivo	S_i	V_i
Procesador (1)	0.5	29
Disco (2)	0.3	13
Disco (3)	2.4	15

Tabla 2.27: Parámetros del sistema informático.

corresponde al segmento de horario con mayor actividad. Calcule:

1. El tiempo mínimo de respuesta de una petición para cualquier valor de carga.

   El tiempo mínimo de respuesta de una petición viene determinado por la Ley General del Tiempo de Respuesta:

   $$R_0^{\min} = \sum_{i=1}^{3} V_i{\cdot}R_i = \sum_{i=1}^{3} V_i{\cdot}(S_i + \cancelto{0}{W_i}) = \sum_{i=1}^{3} V_i{\cdot}S_i = 29{\cdot}0{,}5 + 13{\cdot}0{,}3 + 15{\cdot}2{,}4 = 54{,}4\,\text{ms}$$

2. El tiempo medio de respuesta de cada estación suponiendo que $W_i = N_i \times S_i$.

Por definición:

$$R_i = S_i + W_i = S_i + N_i \cdot S_i = S_i \cdot (1 + N_i)$$

Nuestro objetivo es por tanto calcular N_i para cada dispositivo. Por la Ley de Little, tenemos que:

$$N_i = X_i \cdot R_i$$

Sustituyendo en la ecuación anterior, tenemos que:

$$R_i = S_i \cdot (1 + X_i \cdot R_i) \Longrightarrow R_i = \frac{S_i}{1 - S_i \cdot X_i}$$

Por la Ley del Flujo Forzado, la productividad de cada dispositivo es:

$$X_i = X_0 \cdot V_i = 18 \cdot 10^{-3} \cdot V_i$$

Notemos que ha sido necesario multiplicar por 10^{-3} para pasar de peticiones por segundo a peticiones por milisegundo, para que las unidades sean coherentes con los tiempos de servicio.

Por tanto, la productividad de cada dispositivo es:

$$\begin{aligned} X_1 &= 18 \cdot 10^{-3} \cdot 29 = 522 \cdot 10^{-3} \,{}^{\text{peticiones}}/_{\text{s}} \\ X_2 &= 18 \cdot 10^{-3} \cdot 13 = 234 \cdot 10^{-3} \,{}^{\text{peticiones}}/_{\text{s}} \\ X_3 &= 18 \cdot 10^{-3} \cdot 15 = 270 \cdot 10^{-3} \,{}^{\text{peticiones}}/_{\text{s}} \end{aligned}$$

Por tanto, el tiempo medio de respuesta de cada dispositivo es:

$$\begin{aligned} R_1 &= \frac{0{,}5}{1 - 0{,}5 \cdot 522 \cdot 10^{-3}} \approx 0{,}68\,\text{ms} \\ R_2 &= \frac{0{,}3}{1 - 0{,}3 \cdot 234 \cdot 10^{-3}} = 0{,}32\,\text{ms} \\ R_3 &= \frac{2{,}4}{1 - 2{,}4 \cdot 270 \cdot 10^{-3}} = 6{,}82\,\text{ms} \end{aligned}$$

3. El tiempo medio de respuesta del sistema.

   Por la Ley General del Tiempo de Respuesta, el tiempo medio de respuesta del sistema es:

$$R_0 = \sum_{i=1}^{3} V_i \cdot R_i \approx 29 \cdot 0{,}68 + 13 \cdot 0{,}32 + 15 \cdot 6{,}82 = 126{,}18\,\text{ms}$$

4. La mejora obtenida en el tiempo medio de respuesta del sistema si se sustituye el disco más lento por uno idéntico al rápido.

En este caso, suponemos $S_3 = S_2 = 0{,}3\,\text{ms}$. Por tanto:

$$R'_3 = \frac{0{,}3}{1 - 0{,}3 \cdot 270 \cdot 10^{-3}} \approx 0{,}32\,\text{ms}$$

$$R'_0 = \sum_{i=1}^{3} V_i \cdot R'_i = 29 \cdot 0{,}68 + 13 \cdot 0{,}32 + 15 \cdot 0{,}32 = 28{,}68\,\text{ms}$$

La mejora obtenida en el tiempo medio de respuesta del sistema es:

$$\frac{R_0}{R'_0} = \frac{126{,}18}{28{,}68} \approx 4{,}39$$

Ejercicio 2.5.15. 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.

1. Demuestre numéricamente qué alternativa de las dos anteriores podrá conseguir una mayor productividad media del servidor suponiendo que:

   - Las visitas se reparten equitativamente entre los tres discos en la segunda configuración.
   - El disco es el dispositivo cuello de botella en el caso de la primera configuración.

   Como el disco es el dispositivo cuello de botella en la primera configuración, este determina la productividad máxima del servidor. Puesto que la razón de visita del disco es $V_i = 36$, la demanda de servicio del disco es:

   $$D_d = 36 \cdot 0{,}03 = 1{,}08\,{}^{\text{s}}/_{\text{peticion}}$$

   En la segunda configuración, como cada petición genera 36 solicitudes a los discos y hay tres discos que se reparten de forma equitativa las visitas, la razón de visita de cada disco es $V_i = 12$. Por tanto, la demanda de servicio de cada disco es:

   $$D_d = 12 \cdot 0{,}09 = 1{,}08\,{}^{\text{s}}/_{\text{peticion}}$$

   En ambos casos, la demanda de servicio del disco es la misma. Además, puesto que el resto de dispositivos es idéntico en ambas configuraciones, las demandas de servicio de los demás dispositivos también son idénticas. Como en la primera configuración el disco es el cuello de botella, deducimos que en la segunda configuración cualquiera de las tres unidades de disco es el cuello de botella. Por tanto, la productividad máxima del servidor en ambas configuraciones es:

   $$X_0^{\max} = \frac{1}{D_d} = \frac{1}{1{,}08} \approx 0{,}926\,{}^{\text{peticiones}}/_{\text{s}}$$

   Por tanto, ambas configuraciones ofrecen la misma productividad máxima.

2. ¿A qué conclusión podríamos llegar si no se cumpliera la segunda de las suposiciones?

   En el caso de que no se cumpla la segunda suposición, el disco no sería el cuello de botella en la primera configuración, sino que otro dispositivo lo sería. Como este no cambiaría en la segunda configuración, el cuello de botella sería el mismo en ambas configuraciones. Por tanto, la productividad máxima del servidor seguiría siendo la misma en ambos casos.

Ejercicio 2.5.16. Considere que en el supuesto del ejercicio 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.

   Sea el procesador el dispositivo número 1 y el disco el dispositivo número 2. Hemos visto que, en ambas configuraciones, la demanda de servicio del disco es $D_2 = 1{,}08\,{}^{\mathrm{s}}/_{\mathrm{peticion}}$. Por otro lado, la demanda de servicio del procesador es:

   $$D_1 = V_1 \cdot S_1 = 37 \cdot 0{,}01 = 0{,}37\,{}^{\mathrm{s}}/_{\mathrm{peticion}}$$

   Por tanto, sigue dándose el mismo caso que en el ejercicio anterior. En la primera configuración, el cuello de botella es el disco, mientras que en la segunda configuración cualquiera de los tres discos es el cuello de botella.

2. Productividad máxima.

   Se trata de una red abierta, por lo que la productividad máxima del servidor es:

   $$X_0^{\max} = \frac{1}{D_2} = \frac{1}{1{,}08} \approx 0{,}926\,{}^{\mathrm{peticiones}}/_{\mathrm{s}}$$

3. Tiempo mínimo de respuesta.

   El tiempo mínimo de respuesta se alcanza cuando las peticiones no tienen que esperar en ningún dispositivo. En este caso, puesto que el número de estaciones de servicio varía, el tiempo mínimo de respuesta es diferente en cada configuración. En la primera configuración, el tiempo mínimo de respuesta es:

   $$R_0^{\min} = \sum_{i=1}^{2} V_i \cdot S_i = D_1 + D_2 = 0{,}37 + 1{,}08 = 1{,}45\,\mathrm{s}$$

   En la segunda configuración, el tiempo mínimo de respuesta es:

   $$R_0^{\min} = D_1 + 3 \cdot D_2 = 0{,}37 + 3 \cdot 1{,}08 = 0{,}37 + 3{,}24 = 3{,}61\,\mathrm{s}$$

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?

Calculemos el tiempo medio de respuesta de cada uno de los dispositivos suponiendo que $W_i = N_i \cdot S_i$. Por definición:

$$R_i = S_i + W_i = S_i + N_i \cdot S_i = S_i \cdot (1 + N_i)$$

Por la Ley de Little, tenemos que:

$$N_i = X_i \cdot R_i$$

Sustituyendo en la ecuación anterior, tenemos que:

$$R_i = S_i \cdot (1 + X_i \cdot R_i) \Longrightarrow R_i = \frac{S_i}{1 - S_i \cdot X_i}$$

Por la Ley del Flujo Forzado, la productividad de cada dispositivo es:

$$X_i = X_0 \cdot V_i = 0{,}5 \cdot V_i$$

Por tanto, el tiempo medio de respuesta de cada dispositivo es:

$$R_i = \frac{S_i}{1 - S_i \cdot 0{,}5 \cdot V_i}$$

El tiempo de respuesta del procesador no varía en ninguna de las dos configuraciones, ya que su razón de visita es la misma. Por tanto, el tiempo medio de respuesta del procesador es:

$$R_1 = \frac{0{,}01}{1 - 0{,}01 \cdot 0{,}5 \cdot 37} \approx 0{,}0123\,\text{s}$$

En la primera configuración, el tiempo medio de respuesta del disco es:

$$R_2 = \frac{0{,}03}{1 - 0{,}03 \cdot 0{,}5 \cdot 36} \approx 0{,}0652\,\text{s}$$

En la segunda configuración, el tiempo medio de respuesta de cada disco es:

$$R_2' = \frac{0{,}09}{1 - 0{,}09 \cdot 0{,}5 \cdot 12} \approx 0{,}195652\,\text{s}$$

Por la Ley General del Tiempo de Respuesta, el tiempo medio de respuesta del servidor en la primera configuración es:

$$R_0 = V_1 \cdot R_1 + V_2 \cdot R_2 = 37 \cdot 0{,}0123 + 36 \cdot 0{,}0652 \approx 2{,}8\,\text{s}$$

En la segunda configuración, el tiempo medio de respuesta del servidor es:

$$R_0' = V_1 \cdot R_1 + 3 \cdot V_2 \cdot R_2' = 37 \cdot 0{,}0123 + 3 \cdot 12 \cdot 0{,}195652 \approx 7{,}5\,\text{s}$$

Por tanto, atendiendo al tiempo medio de respuesta, la mejor opción es la primera configuración, ya que el tiempo medio de respuesta es menor. La mejora conseguida es:

$$\frac{R_0'}{R_0} = \frac{7{,}5}{2{,}8} \approx 2{,}68$$

Dispositivo	S_i	V_i
CPU (1)	0.01	11
Disco (2)	0.05	3
Disco (3)	0.08	7

Tabla 2.28: Parámetros del sistema interactivo.

Ejercicio 2.5.17. Un sistema interactivo con 30 clientes (suponga un trabajo por cliente) y un tiempo medio de reflexión de 12 segundos se modela mediante los parametros de la Tabla 2.28 (los tiempos se expresan en segundos). Determine:

1. El cuello de botella del servidor.

   Tenemos que calcular la demanda de servicio de cada dispositivo:

   $$D_1 = V_1 \cdot S_1 = 11 \cdot 0{,}01 = 0{,}11\,{}^{\text{s}}/_{\text{trabajo}}$$
   $$D_2 = V_2 \cdot S_2 = 3 \cdot 0{,}05 = 0{,}15\,{}^{\text{s}}/_{\text{trabajo}}$$
   $$D_3 = V_3 \cdot S_3 = 7 \cdot 0{,}08 = 0{,}56\,{}^{\text{s}}/_{\text{trabajo}}$$

   El cuello de botella del servidor es el segundo disco (dispositivo número 3), ya que su demanda de servicio es la mayor de las tres.

2. La productividad máxima y el punto teórico de saturación (knee point).

   La productividad máxima del servidor depende de si el servidor está en baja o alta carga. Veamos ambos casos:

   - Si el servidor está en baja carga, el tiempo mínimo de respuesta del servidor viene determinado por la Ley General del Tiempo de Respuesta:

     $$R_0^{\min} = \sum_{i=1}^{3} V_i \cdot R_i = \sum_{i=1}^{3} V_i \cdot (S_i + \cancelto{0}{W_i}) = \sum_{i=1}^{3} V_i \cdot S_i = \sum_{i=1}^{3} D_i = 0{,}11 + 0{,}15 + 0{,}56 = 0{,}82\,\text{s}$$

     Por la Ley de Little, la productividad máxima del servidor es:

     $$X_0^{\max} = \frac{N_T}{R_0^{\min} + Z} = \frac{N_T}{0{,}82 + 12} = \frac{N_T}{12{,}82}$$

   - Si el servidor está en alta carga, la productividad máxima del servidor viene determinada por la Relación Utilización-Demanda de Servicio. Puesto que esta productividad máxima se alcanzará cuando se produzca saturación en el cuello de botella, tenemos que:

     $$1 = U_b = X_0^{\max} \cdot D_b \Longrightarrow X_0^{\max} = \frac{1}{D_b} = \frac{1}{0{,}56} \approx 1{,}79\,{}^{\text{trabajos}}/_{\text{s}}$$

   El punto teórico de saturación (knee point) se produce cuando se da siguiente igualdad:

   $$1{,}79 = \frac{N_T^*}{12{,}82} \Longrightarrow N_T^* = 1{,}79 \cdot 12{,}82 \approx 22{,}89$$

   Por tanto, en este caso el servidor se encuentra en un régimen de alta carga, ya que $N_T = 30 > N_T^* \approx 22{,}89$. Por tanto, la productividad máxima del servidor es $X_0^{\max} \approx 1{,}79\,{}^{\text{trabajos}}/_{\text{s}}$.

3. Los límites optimistas del tiempo de respuesta y de la productividad.

   Como hemos visto anteriormente, tenemos que:

$$X_0^{\max} = \text{mín}\left\{\frac{N_T}{12{,}82}, 1{,}79\right\} \text{ trabajos/s}$$

   Por otro lado, distinguimos casos para el tiempo mínimo de respuesta del servidor:

   - Si el servidor está en baja carga, hemos visto que:

$$R_0^{\min} = 0{,}82\,\text{s}$$

   - Si el servidor está en alta carga, el tiempo mínimo de respuesta viene dado por la Ley de Little:

$$R_0^{\min} = \frac{N_T}{X_0^{\max}} - Z = \frac{N_T}{1{,}79} - 12$$

   Por tanto, los límites optimistas del tiempo de respuesta son:

$$R_0^{\min} = \text{máx}\left\{0{,}82, 0{,}56 \cdot N_T - 12\right\} \text{ segundos}$$

4. Utilice `solvenet` para calcular el número medio de trabajos en reflexión suponiendo que $R_i(n_T) = (N_i(n_T - 1) + 1) \times S_i$.

Ejercicio 2.5.18. Los límites optimistas del tiempo de respuesta y de la productividad de un sistema informático interactivo son:

$$R_0 \geqslant \text{máx}\{0{,}49, 0{,}22 \cdot N_T - 5\}$$
$$X_0 \leqslant \text{mín}\left\{\frac{N_T}{5{,}49}, 4{,}55\right\}$$

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

1. Tiempo de reflexión.

   Identificando términos, tenemos que el tempo de reflexión es de $Z = 5\,\text{s}$.

2. Tiempo mínimo de respuesta del servidor.

   Este se da cuando el servidor está sometido a una baja carga, y es de $0{,}49\,\text{s}$.

3. Punto teórico de saturación (knee point).

   Este se da con la siguiente igualdad:

$$\frac{N_T^*}{5{,}49} = 4{,}55 \Longrightarrow N_T^* = 24{,}9795\,\text{trabajos}$$

4. ¿Cuál es el tiempo medio de respuesta del servidor que se podría esperar con 100 clientes (suponga un trabajo por cliente)? Justifique la respuesta.

   En este caso el servidor estaría sometido a una alta carga, luego:

   $$R_0 = 0{,}22 \cdot 100 - 5 = 15\,\text{s}$$

5. Si hubiera 35 clientes en el sistema, ¿sería posible obtener un tiempo de respuesta de aproximadamente 0,49 segundos? ¿Por qué?

   No, puesto que en este caso el sistema estaría saturado y, por tanto, al menos en el cuello de botella habría elementos en la cola de espera. Por tanto, no se podría alcanzar el tiempo mínimo de respuesta, que es de 0,49 s.

Ejercicio 2.5.19. Considere el sistema informático por lotes ($Z = 0$) con 15 trabajos modelado con los parámetros de la Tabla 2.29 (los tiempos se expresan en segundos). Utilizando `solvenet`, justifique numéricamente qué opción de las dos

Dispositivo	S_i	V_i
Procesador (1)	0.016	5
Disco (2)	0.02	4

Tabla 2.29: Parámetros del sistema informático por lotes.

que se indican a continuación representa la mejor elección para mejorar el rendimiento del sistema:

1. Sustituir el procesador por uno dos veces más rápido.

2. Sustituir el disco por otro con un tiempo medio de servicio de 0.01 segundos.

Ejercicio 2.5.20. Considere un servidor web que recibe una media de 0,3 peticiones por segundo y es modelado con los parámetros de la Tabla 2.30 (los tiempos de la tabla se expresan en segundos). Después de apurar su copa de vino, una informáti-

Dispositivo	S_i	V_i
CPU (1)	0.2	15
DiscoA (2)	0.07	6
DiscoB (3)	0.02	8

Tabla 2.30: Parámetros del servidor web.

ca avezada en temas de modelado y evaluación de rendimiento hace las siguientes confesiones a sus compañeros de cena respecto del modelo anterior (suponga que $W_i = N_i \times S_i$). ¿Ha afectado la ingesta de alcohol la mente despierta de nuestra protagonista? Justifique numéricamente la respuesta.

1. 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 %).

Calculamos en primer lugar los tiempos de respuesta de cada uno de los dispositivos del servidor web. Por definición:

$$R_i = S_i + W_i = S_i + N_i \cdot S_i = S_i \cdot (1 + N_i)$$

Por la Ley de Little, tenemos que:

$$N_i = X_i \cdot R_i$$

Por la Ley del Flujo Forzado, la productividad de cada dispositivo es:

$$X_i = X_0 \cdot V_i = 0{,}3 \cdot V_i$$

Por tanto, el tiempo medio de respuesta de cada dispositivo cumple:

$$R_i = S_i \cdot (1 + 0{,}3 \cdot V_i \cdot R_i) \Longrightarrow R_i = \frac{S_i}{1 - 0{,}3 \cdot S_i \cdot V_i}$$

Calculamos el tiempo medio de respuesta de cada dispositivo:

$$R_1 = \frac{0{,}2}{1 - 0{,}2 \cdot 0{,}3 \cdot 15} = 2\,\mathrm{s}$$
$$R_2 = \frac{0{,}07}{1 - 0{,}07 \cdot 0{,}3 \cdot 6} = 0{,}08\,\mathrm{s}$$
$$R_3 = \frac{0{,}02}{1 - 0{,}02 \cdot 0{,}3 \cdot 8} = 0{,}021\,\mathrm{s}$$

Por la Ley General del Tiempo de Respuesta, el tiempo medio de respuesta del servidor web es:

$$R_0 = V_1 \cdot R_1 + V_2 \cdot R_2 + V_3 \cdot R_3 = 30{,}648\,\mathrm{s}$$

Si sustituimos el procesador por otro dos veces y media más rápido, su nuevo tiempo de servicio es $S_1' = {}^{0{,}2}/_{2{,}5} = 0{,}08\,{}^{\mathrm{s}}/_{\mathrm{trabajo}}$. Por tanto, el nuevo tiempo medio de respuesta del procesador es:

$$R_1' = \frac{0{,}08}{1 - 0{,}08 \cdot 0{,}3 \cdot 15} = 0{,}125\,\mathrm{s}$$

El nuevo tiempo medio de respuesta del servidor web es:

$$R_0' = V_1 \cdot R_1' + V_2 \cdot R_2 + V_3 \cdot R_3 = 2{,}523\,\mathrm{s}$$

La mejora en velocidad es:

$$\frac{R_0}{R_0'} = \frac{30{,}648}{2{,}523} \approx 12{,}14$$

Por tanto, la afirmación de la informática es correcta, ya que la mejora en velocidad es del 1114 % (es decir, la mejora en velocidad es mayor del 1100 %).

2. 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 %).

   Calculamos en primer lugar las nuevas razones de visita de los discos. En primer lugar, como el número de visitas al subsistema de discos es constante, tenemos que:

   $$V_2' + V_3' = V_2 + V_3 = 6 + 8 = 14$$

   Si equilibramos las cargas de los discos, y sabiendo que los valores del tiempo de servicio de los discos no se ven alterados (puesto que no se ha cambiado el hardware), tenemos que:

   $$D_2' = D_3' \Longrightarrow V_2' \cdot 0{,}07 = V_3' \cdot 0{,}02$$

   Resolvemos el sistema de ecuaciones formado por las dos ecuaciones anteriores:

   $$\left.\begin{array}{l} V_2' + V_3' = 14 \\ V_2' \cdot 0{,}07 = V_3' \cdot 0{,}02 \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} V_2' = 3{,}1111 \\ V_3' = 10{,}889 \end{array}\right.$$

   Por tanto, los nuevos tiempos medios de respuesta de los discos son:

   $$R_2' = \frac{0{,}07}{1 - 0{,}07 \cdot 0{,}3 \cdot 3{,}1111} = 0{,}0748\,\mathrm{s}$$
   $$R_3' = \frac{0{,}02}{1 - 0{,}02 \cdot 0{,}3 \cdot 10{,}889} = 0{,}0213\,\mathrm{s}$$

   El nuevo tiempo medio de respuesta del servidor web es:

   $$R_0'' = V_1 \cdot R_1 + V_2' \cdot R_2' + V_3' \cdot R_3' \approx 30{,}466\,\mathrm{s}$$

   La mejora en velocidad es:

   $$\frac{R_0}{R_0''} = \frac{30{,}648}{30{,}466} \approx 1{,}006$$

   Por tanto, la afirmación de la informática es correcta, ya que la mejora en velocidad es del 0.6 % (es decir, la mejora en velocidad es menor del 1 %).

Ejercicio 2.5.21. 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 web; el problema, añade, está en la configuración de los navegadores de los usuarios. ¿Tiene base científica la afirmación del administrador? Justifique la respuesta.

Por la Ley de la Utilización, tenemos que:

$$U_b = 0{,}8 = 38 \cdot S_b \Longrightarrow S_b = \frac{0{,}8}{38} \approx 0{,}0211\,{}^{\mathrm{s}}/_{\mathrm{peticion}}$$

Por tanto, la demanda de servicio del disco duro es:

$$D_b = V_b \cdot S_b = 1{,}5 \cdot 0{,}0211 \approx 0{,}0316\,{}^{\mathrm{s}}/_{\mathrm{peticion}}$$

La productividad máxima del servidor web es:

$$X_0^{\max} = \frac{1}{D_b} = \frac{1}{0{,}0316} \approx 31{,}6666\,{}^{\mathrm{peticiones}}/_{\mathrm{s}}$$

Por tanto, no se puede alcanzar una productividad de 40 peticiones por segundo, ya que la productividad máxima del servidor web es de aproximadamente 31,7 peticiones por segundo.

Ejercicio 2.5.22. 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 en la Tabla 2.31. En las horas de máxima actividad el sitio web ha llegado a

Dispositivo	S_i	V_i
Procesador (1)	0.01	9
Disco (2)	0.02	4
Disco (3)	0.02	4

Tabla 2.31: Parámetros del servidor web.

recibir una media de 11 peticiones por segundo. Responda a las siguientes cuestiones justificando numéricamente la respuesta.

1. Desde el punto de vista del reparto de la carga entre los componentes del servidor web, ¿estamos ante un servidor equilibrado?

   Un servidor está perfectamente equilibrado si la demanda de servicio de todos sus dispositivos es la misma. En este caso, tenemos que calcular la demanda de servicio de cada dispositivo:

   $$D_1 = V_1 \cdot S_1 = 9 \cdot 0{,}01 = 0{,}09\,{}^{\mathrm{s}}/_{\mathrm{peticion}}$$
   $$D_2 = V_2 \cdot S_2 = 4 \cdot 0{,}02 = 0{,}08\,{}^{\mathrm{s}}/_{\mathrm{peticion}}$$
   $$D_3 = V_3 \cdot S_3 = 4 \cdot 0{,}02 = 0{,}08\,{}^{\mathrm{s}}/_{\mathrm{peticion}}$$

   La demanda de servicio del procesador es ligeramente mayor que la de los discos, por lo que el servidor no está perfectamente equilibrado, pero sí que está cerca del equilibrio, ya que las demandas son parecidas.

2. Indique si el servidor está sometido a alta o baja carga.

   Calculamos los límites optimistas del tiempo de respuesta:

   - Si el servidor está en baja carga, el tiempo mínimo de respuesta del servidor es:

   $$R_0^{\min} = \sum_{i=1}^{3} V_i \cdot S_i = 9 \cdot 0{,}01 + 4 \cdot 0{,}02 + 4 \cdot 0{,}02 = 0{,}25\,\text{s}$$

   - Si el servidor está en alta carga, entonces por la Ley de la Utilización tenemos que:

   $$1 = U_b = X_0^{\max} \cdot D_b \Longrightarrow X_0^{\max} = \frac{1}{D_b} = \frac{1}{0{,}09} = 11{,}1111\,{}^{\text{peticiones}}/_{\text{s}}$$

   Por tanto, el tiempo mínimo de respuesta del servidor es:

   $$R_0^{\min} = \frac{N_T}{X_0^{\max}} = \frac{N_T}{11{,}1111}$$

   Por tanto, el punto teórico de saturación (knee point) se produce cuando:

   $$11{,}1111 = \frac{N_T^*}{0{,}25} \Longrightarrow N_T^* = 2{,}7777$$

   Como $N_T = 11 > N_T^* \approx 2{,}78$, el servidor está sometido a alta carga.

3. ¿Cuál es el tiempo mínimo de respuesta de este servidor web?

   El tiempo mínimo de respuesta del servidor web es de 0,25 s, cuando este no estuviese sometido a alta carga.

4. ¿Podríamos aumentar la productividad máxima (capacidad) del servidor si sustituimos ambos discos por versiones más rápidas?

   Si sustituimos ambos discos por versiones más rápidas, la demanda de servicio de los discos disminuiría, pero la demanda de servicio del procesador seguiría siendo la misma. Por tanto, el cuello de botella del servidor seguiría siendo el procesador, y por tanto no se podría aumentar la productividad máxima del servidor.

5. Determine el tiempo medio de respuesta del servidor suponiendo que $W_i = N_i \cdot S_i$.

   Por definición, el tiempo medio de respuesta de cada dispositivo es:

   $$R_i = S_i + W_i = S_i + N_i \cdot S_i = S_i \cdot (1 + N_i)$$

   Por la Ley de Little, tenemos que:

   $$N_i = X_i \cdot R_i$$

Por la Ley del Flujo Forzado, la productividad de cada dispositivo es:

$$X_i = X_0 \cdot V_i = 11 \cdot V_i$$

Por tanto, el tiempo medio de respuesta de cada dispositivo cumple:

$$R_i = S_i \cdot (1 + 11 \cdot V_i \cdot R_i) \Longrightarrow R_i = \frac{S_i}{1 - 11 \cdot S_i \cdot V_i}$$

Calculamos el tiempo medio de respuesta de cada dispositivo:

$$R_1 = \frac{0{,}01}{1 - 0{,}01 \cdot 11 \cdot 9} = 1\,\text{s}$$
$$R_2 = \frac{0{,}02}{1 - 0{,}02 \cdot 11 \cdot 4} = 0{,}1666\,\text{s}$$
$$R_3 = \frac{0{,}02}{1 - 0{,}02 \cdot 11 \cdot 4} = 0{,}1666\,\text{s}$$

Por la Ley General del Tiempo de Respuesta, el tiempo medio de respuesta del servidor web es:

$$R_0 = V_1 \cdot R_1 + V_2 \cdot R_2 + V_3 \cdot R_3 = 10{,}3333\,\text{s}$$

6. ¿Cambiaría la localización del cuello de botella si la tasa de llegada bajara hasta las 5 peticiones por segundo?

   No, puesto que la demanda de servicio del procesador seguiría siendo mayor que la de los discos. Por tanto, el cuello de botella seguiría siendo el procesador. Esta disminución afecta a todo el sistema, no solamente al cuello de botella.

7. 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.

   Por el apartado 5, tenemos que:

$$R_i = \frac{S_i}{1 - \lambda_0 \cdot D_i}$$

   Por la Ley General del Tiempo de Respuesta, el tiempo medio de respuesta del servidor es:

$$\begin{aligned} R_0 &= \sum_{i=1}^{3} V_i \cdot R_i = \sum_{i=1}^{3} V_i \cdot \frac{S_i}{1 - \lambda_0 \cdot D_i} = \sum_{i=1}^{3} \frac{D_i}{1 - \lambda_0 \cdot D_i} = \\ &= \frac{0{,}09}{1 - \lambda_0 \cdot 0{,}09} + \frac{0{,}08}{1 - \lambda_0 \cdot 0{,}08} + \frac{0{,}08}{1 - \lambda_0 \cdot 0{,}08} = \\ &= \frac{0{,}09}{1 - \lambda_0 \cdot 0{,}09} + \frac{0{,}16}{1 - \lambda_0 \cdot 0{,}08} \end{aligned}$$

   La gráfica del tiempo medio de respuesta del servidor en función de la tasa de llegada de trabajos se muestra en la Figura 2.4.

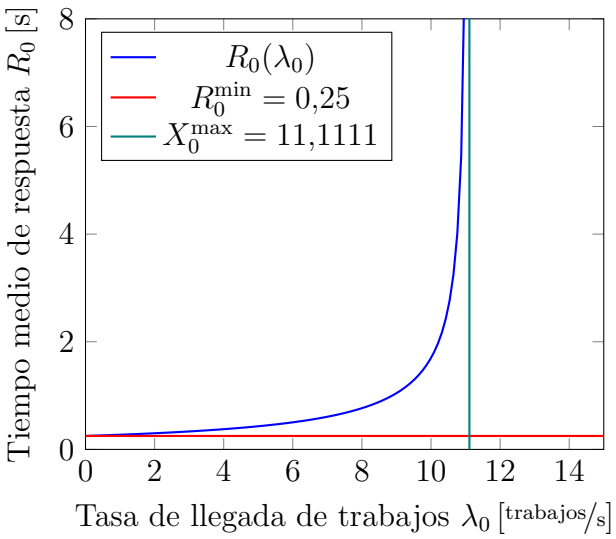


Figura 2.4: Límites del Tiempo de Respuesta R_0.

Ejercicio 2.5.23. 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:

$$R_0 \geqslant \text{máx}\{1{,}6, 1{,}1 \times N_T - 4\}$$
$$X_0 \leqslant \text{mín}\left\{\frac{N_T}{5{,}6}, 0{,}91\right\}$$

1. El sistema, ¿está realmente soportando una carga elevada?

   Para determinar si el sistema está soportando una carga elevada, debemos calcular el punto teórico de saturación (knee point) del sistema. Este se produce cuando:

   $$\frac{N_T^*}{5{,}6} = 0{,}91 \Longrightarrow N_T^* = 0{,}91 \cdot 5{,}6 \approx 5{,}096$$

   Como $N_T = 32 > N_T^* \approx 5{,}1$, el sistema está soportando una carga elevada.

2. Haga una estimación del tiempo medio de respuesta del servidor en las condiciones actuales.

   Para estimar el tiempo medio de respuesta del servidor, debemos considerar que el sistema está en alta carga. Por tanto, el tiempo mínimo de respuesta del servidor es:

   $$R_0^{\min} = 1{,}1 \cdot N_T - 4 = 1{,}1 \cdot 32 - 4 = 31{,}2\text{s}$$

Ejercicio 2.5.24. El proceso de modelado de un servidor de base de datos mediante técnicas de análisis operacional ha dado los parámetros que se muestran en la Tabla 2.32 (los tiempos se expresan en segundos). 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.

Dispositivo	S_i	V_i
Procesador (1)	0,15	6
Disco (2)	0,05	5

Tabla 2.32: Parámetros del servidor de base de datos.

1. Indique si el servidor está sometido a alta o baja carga.

   En este caso no tiene sentido preguntarse por el punto teórico de saturación (knee point) porque estamos ante una red abierta. Calculemos en primer lugar el cuello de botella, y posteriormente su utilización:

   $$D_1 = V_1 \cdot S_1 = 6 \cdot 0{,}15 = 0{,}9\,{}^{\mathrm{s}}/_{\mathrm{peticion}}$$
   $$D_2 = V_2 \cdot S_2 = 5 \cdot 0{,}05 = 0{,}25\,{}^{\mathrm{s}}/_{\mathrm{peticion}}$$

   El cuello de botella es el procesador porque su demanda de servicio es mayor que la del disco. Por la Ley de la Utilización, tenemos que:

   $$U_1 = X_0 \cdot D_1 = 1{,}05 \cdot 0{,}9 = 0{,}945 \Longrightarrow U_1 = 94{,}5\,\%$$

   Como el cuello de botella tiene una utilización del 94.5 %, el servidor está sometido a alta carga, aunque aún no está saturado.

2. ¿Cuál es el número medio de trabajos en la cola del procesador?

   Por la Ley de Little, el número medio de trabajos en la cola del procesador es:

   $$Q_1 = X_1 \cdot W_1$$

   Calculamos en primer lugar X_1 empleando para ello la Ley del Flujo Forzado:

   $$X_1 = X_0 \cdot V_1 = 1{,}05 \cdot 6 = 6{,}3^{\mathrm{trabajos}}/_{\mathrm{s}}$$

   A continuación, calculamos el tiempo medio de espera en la cola del procesador:

   $$W_1 = R_1 - S_1 = (N_1 + 1) \cdot S_1 - S_1 = N_1 \cdot S_1$$

   Por la Ley de Little, tenemos que:

   $$N_1 = X_1 \cdot R_1 = X_1 \cdot (N_1 + 1) \cdot S_1 \Longrightarrow N_1 = \frac{X_1 \cdot S_1}{1 - X_1 \cdot S_1} = 17{,}181818\mathrm{trabajos}$$

   Por tanto, el número medio de trabajos en la cola del procesador es:

   $$Q_1 = X_1 \cdot W_1 = 6{,}3 \cdot 17{,}181818 \cdot 0{,}15 = 16{,}24\mathrm{trabajos}$$

3. Calcule el tiempo medio de respuesta del servidor.

   Por la Ley General del Tiempo de Respuesta, el tiempo medio de respuesta del servidor es:

   $$R_0 = V_1 \cdot R_1 + V_2 \cdot R_2$$

Para calcular R_1 y R_2, necesitamos optener N_1 y N_2:

$$N_i = X_i \cdot R_i = X_i \cdot (N_i + 1) \cdot S_i \Longrightarrow N_i = \frac{X_i \cdot S_i}{1 - X_i \cdot S_i}$$

Calculamos ahora X_1 y X_2 empleando la Ley del Flujo Forzado:

$$X_i = X_0 \cdot V_i$$

Por tanto:

$$N_i = \frac{X_0 \cdot V_i \cdot S_i}{1 - X_0 \cdot V_i \cdot S_i}$$

Por tanto:

$$R_i = \left(\frac{X_0 \cdot V_i \cdot S_i}{1 - X_0 \cdot V_i \cdot S_i} + 1 \right) \cdot S_i$$
$$R_1 = 2{,}72727\text{s}$$
$$R_2 = 0{,}0677966\text{s}$$

Por tanto, el tiempo medio de respuesta del servidor es:

$$R_0 = V_1 \cdot R_1 + V_2 \cdot R_2 = 16{,}702619\text{s}$$

4. ¿Tendría algún efecto sobre las prestaciones sustituir el procesador por una versión más rápida?

   Sí, puesto que el procesador es el cuello de botella del servidor, por lo que la productividad máxima del servidor aumentaría.

5. 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.

   En este caso, la razón de visita no cambia, pero:

$$S_1' = \frac{0{,}15}{5} = 0{,}03$$
$$S_2' = \frac{0{,}05}{2} = 0{,}025$$

   Calculamos la demanda de servicio de cada dispositivo:

$$D_1' = V_1 \cdot S_1' = 6 \cdot 0{,}03 = 0{,}18\,{}^{\text{s}}/_{\text{peticion}}$$
$$D_2' = V_2 \cdot S_2' = 5 \cdot 0{,}025 = 0{,}125\,{}^{\text{s}}/_{\text{peticion}}$$

   En este caso, el cuello de botella seguiría siendo el procesador.

Ejercicio 2.5.25. Un sistema interactivo con 3 clientes (suponga un trabajo por cliente) y un tiempo medio de reflexión de 5 segundos se modela mediante los parámetros que se muestran en la Tabla 2.33 (los tiempos se expresan en segundos): Determine, sabiendo que la productividad del servidor es $0{,}49\,{}^{\text{trabajos}}/_{\text{s}}$, los siguientes valores:

Dispositivo	S_i	V_i
Procesador (1)	0.01	15
Disco (2)	0.04	8
Disco (3)	0.08	6

Tabla 2.33: Parámetros del sistema interactivo.

1. El tiempo medio de respuesta del servidor.

   Por la Ley del Tiempo de Respuesta Interactivo:

$$N_T = X_0 \cdot (R_0 + Z) \Longrightarrow R_0 = \frac{N_T}{X_0} - Z = 1{,}12\text{s}$$

2. Las utilizaciones de cada dispositivo.

   Por la Relación Utilización-Demanda de Servicio:

$$\begin{aligned} U_i &= X_0 \cdot D_i = X_0 \cdot V_i \cdot S_i \\ U_1 &= 0{,}49 \cdot 15 \cdot 0{,}01 = 0{,}0735 = 7{,}35\,\% \\ U_2 &= 0{,}49 \cdot 8 \cdot 0{,}04 = 0{,}1568 = 15{,}68\,\% \\ U_3 &= 0{,}49 \cdot 6 \cdot 0{,}08 = 0{,}2352 = 23{,}52\,\% \end{aligned}$$

Ejercicio 2.5.26. 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 Tabla 2.34. A partir de la información anterior:

Dispositivo	S_i	V_i
Procesador (1)	0.01	8
Disco duro (2)	0.04	4
Red (3)	0.03	3

Tabla 2.34: Parámetros del servidor web.

1. Calcule la demanda de servicio, la productividad y la utilización de cada dispositivo.

   La demanda de servicio de cada dispositivo se calcula como:

$$\begin{aligned} D_i &= V_i \cdot S_i \\ D_1 &= 8 \cdot 0{,}01 = 0{,}08\,{}^{\text{s}}\!/_{\text{peticion}} \\ D_2 &= 4 \cdot 0{,}04 = 0{,}16\,{}^{\text{s}}\!/_{\text{peticion}} \\ D_3 &= 3 \cdot 0{,}03 = 0{,}09\,{}^{\text{s}}\!/_{\text{peticion}} \end{aligned}$$

   La productividad de cada dispositivo se calcula por la Ley del Flujo Forzado como:

$$\begin{aligned} X_i &= X_0 \cdot V_i = 4 \cdot V_i \\ X_1 &= 4 \cdot 8 = 32\,{}^{\text{trabajos}}\!/_{\text{s}} \\ X_2 &= 4 \cdot 4 = 16\,{}^{\text{trabajos}}\!/_{\text{s}} \\ X_3 &= 4 \cdot 3 = 12\,{}^{\text{trabajos}}\!/_{\text{s}} \end{aligned}$$

La utilización de cada dispositivo se calcula por la Ley de la Utilización como:

$$\begin{aligned} U_i &= X_i \cdot S_i \\ U_1 &= 32 \cdot 0{,}01 = 0{,}32 = 32\,\% \\ U_2 &= 16 \cdot 0{,}04 = 0{,}64 = 64\,\% \\ U_3 &= 12 \cdot 0{,}03 = 0{,}36 = 36\,\% \end{aligned}$$

2. Determine el tiempo mínimo posible de respuesta del servidor web. Justifique la respuesta.

   Por la Ley General del Tiempo de Respuesta, el tiempo mínimo de respuesta del servidor web es:

$$\begin{aligned} R_0^{\min} &= \sum_{i=1}^{3} V_i \cdot S_i = \sum_{i=1}^{3} D_i = D_1 + D_2 + D_3 \\ &= 0{,}08 + 0{,}16 + 0{,}09 = 0{,}33\,\mathrm{s} \end{aligned}$$

3. ¿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?

   El cuello de botella del servidor es el disco duro, ya que es el dispositivo con mayor demanda de servicio ($D_2 = 0{,}16^{\mathrm{s}}/_{\mathrm{peticion}}$). La tasa de llegadas X_0 no afecta al dispositivo cuello de botella, ya que este se determina por la demanda de servicio de cada dispositivo. Por último, el sistema no está equilibrado porque la utilización del disco duro es muy superior a la del resto de dispositivos ($U_2 = 64\,\%$ frente a $U_1 = 32\,\%$ y $U_3 = 36\,\%$).

4. 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.

   La productividad máxima del servidor web se calcula como:

$$X_0^{\max} = \frac{1}{D_b} = \frac{1}{D_2} = \frac{1}{0{,}16} = 6{,}25^{\mathrm{peticiones}}/_{\mathrm{s}}$$

   Si quisiéramos obtener el doble de esa productividad máxima, tendríamos que tener:

$$X_0^{\max} = 2 \cdot 6{,}25 = 12{,}5^{\mathrm{peticiones}}/_{\mathrm{s}} = \frac{1}{D_b'} \Longrightarrow D_b' = \frac{1}{12{,}5} = 0{,}08^{\mathrm{s}}/_{\mathrm{peticion}}$$

   Por tanto, la demanda de servicio del dispositivo cuello de botella debería ser:

$$D_b' = 0{,}08^{\mathrm{s}}/_{\mathrm{peticion}}$$

   Por tanto, por un lado necesitaríamos que:

$$S_2' = \frac{0{,}08}{4} = 0{,}02^{\mathrm{s}}/_{\mathrm{peticion}}$$

No obstante, puesto que $D_3 = 0{,}09 > D'_b$, también tendríamos que reducir el tiempo de servicio del dispositivo red:

$$S'_3 = \frac{0{,}08}{3} = 0{,}0266667\,{}^{\text{s}}/_{\text{peticion}}$$

5. Suponiendo que $R_i = (N_i + 1) \cdot S_i$ para cada dispositivo, calcule el tiempo de respuesta del servidor web.

   Por la Ley General del Tiempo de Respuesta, el tiempo de respuesta del servidor web es:

   $$R_0 = \sum_{i=1}^{3} V_i \cdot R_i$$

   Para calcular R_i, necesitamos calcular N_i. Por la Ley de Little, tenemos que:

   $$N_i = X_i \cdot R_i = X_i \cdot (N_i + 1) \cdot S_i \Longrightarrow N_i = \frac{X_i \cdot S_i}{1 - X_i \cdot S_i}$$

   $$N_1 = \frac{32 \cdot 0{,}01}{1 - 32 \cdot 0{,}01} = 0{,}47058\,{}^{\text{trabajos}}/_{\text{s}}$$

   $$N_2 = \frac{16 \cdot 0{,}04}{1 - 16 \cdot 0{,}04} = 1{,}7777\,{}^{\text{trabajos}}/_{\text{s}}$$

   $$N_3 = \frac{12 \cdot 0{,}03}{1 - 12 \cdot 0{,}03} = 0{,}5625\,{}^{\text{trabajos}}/_{\text{s}}$$

   Por tanto, el tiempo de respuesta de cada dispositivo es:

   $$R_i = (N_i + 1) \cdot S_i$$
   $$R_1 = 0{,}0147\,\text{s}$$
   $$R_2 = 0{,}1111\,\text{s}$$
   $$R_3 = 0{,}0468\,\text{s}$$

   Por tanto, el tiempo de respuesta del servidor web es:

   $$R_0 = V_1 \cdot R_1 + V_2 \cdot R_2 + V_3 \cdot R_3 = 0{,}70271\,\text{s}$$

6. Calcule el número medio de peticiones en el servidor web. ¿Cómo se llama la ley que ha utilizado?

   Por la Ley de Little, el número medio de peticiones en el servidor web es:

   $$N_0 = X_0 \cdot R_0 = 4 \cdot 0{,}70271 = 2{,}81086\,\text{peticiones}$$

Ejercicio 2.5.27. 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): A partir de la información anterior:

Dispositivo	S_i	V_i
CPU (1)	0.5	4
Red (2)	0.75	3

Tabla 2.35: Parámetros del servidor de apuestas deportivas.

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

$$X_0 \leqslant \text{mín}\left\{\frac{N_T}{D+Z}, \frac{1}{D_b}\right\}$$

   - X_0: productividad del servidor, en $^{\text{peticiones}}/_{\text{s}}$.
   - N_T: número total de clientes conectados al servidor, en unidades. En este caso:

$$N_T = 25\,\text{peticiones}$$

   - D: demanda de servicio del servidor, en $^{\text{s}}/_{\text{peticion}}$. En este caso:

$$D = D_1 + D_2 = V_1 \cdot S_1 + V_2 \cdot S_2 = 4 \cdot 0{,}5 + 3 \cdot 0{,}75 = 4{,}25\,{}^{\text{s}}/_{\text{peticion}}$$

   - Z: tiempo medio de reflexión de los clientes, en segundos. En este caso:

$$Z = 6\,\text{s}$$

   - D_b: demanda de servicio del cuello de botella, en $^{\text{s}}/_{\text{peticion}}$. Esta es la mayor demanda de servicio entre los dispositivos del servidor. En este caso:

$$\begin{aligned} D_b &= \text{máx}\{D_1, D_2\} = \text{máx}\{V_1 \cdot S_1, V_2 \cdot S_2\} = \\ &= \text{máx}\{4 \cdot 0{,}5, 3 \cdot 0{,}75\} = \text{máx}\{2, 2{,}25\} = 2{,}25\,{}^{\text{s}}/_{\text{peticion}} \end{aligned}$$

2. ¿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?

   El punto teórico de saturación (knee point) del servidor se calcula como:

$$\frac{N_T^*}{D+Z} = \frac{1}{D_b} \Longrightarrow N_T^* = \frac{D+Z}{D_b} = \frac{4{,}25+6}{2{,}25} \approx 4{,}555\,\text{clientes}$$

   Como $N_T = 25 > N_T^* \approx 4{,}555$, el sistema se encuentra sometido a alta carga.

Ejercicio 2.5.28. 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:

1. Calcule la razón de visita, el tiempo de servicio, la productividad y la utilización tanto del procesador como del disco duro.

   Los datos proporcionados son:

$$X_0 = \frac{15000}{24 \cdot 3600} \approx 0{,}1736 \,{}^{\text{peticiones}}/_{\text{s}}$$
$$U_1 = \frac{800}{24 \cdot 60} \approx 0{,}5556 = 55{,}56\,\%$$
$$V_1 = \frac{60000}{15000} = 4$$
$$V_2 = \frac{135000}{15000} = 9$$
$$U_2 = \frac{1200}{24 \cdot 60} = 0{,}8333 = 83{,}33\,\%$$

   Por tanto, tan solo nos falta por calcular el tiempo de servicio y la productividad de cada dispositivo. Por la Ley del Flujo Forzado, tenemos que:

$$X_i = X_0 \cdot V_i$$
$$X_1 = 0{,}1736 \cdot 4 \approx 0{,}6944 \,{}^{\text{trabajos}}/_{\text{s}}$$
$$X_2 = 0{,}1736 \cdot 9 \approx 1{,}5624 \,{}^{\text{trabajos}}/_{\text{s}}$$

   Por la Ley de la Utilización, tenemos que:

$$U_i = X_i \cdot S_i \Longrightarrow S_i = \frac{U_i}{X_i}$$
$$S_1 \approx 0{,}800\,\text{s}$$
$$S_2 \approx 0{,}533\,\text{s}$$

2. ¿Cuál es la productividad máxima que puede alcanzar este servidor? ¿Y el tiempo de respuesta mínimo?

   En primer lugar, calculamos las demandas de servicio de cada dispositivo:

$$D_1 = V_1 \cdot S_1 = 4 \cdot 0{,}800 = 3{,}2 \,{}^{\text{s}}/_{\text{peticion}}$$
$$D_2 = V_2 \cdot S_2 = 9 \cdot 0{,}533 = 4{,}797 \,{}^{\text{s}}/_{\text{peticion}}$$

   Por tanto, el cuello de botella del servidor es el disco, ya que su demanda de servicio es mayor que la del procesador. Por tanto, la productividad máxima del servidor es:

$$X_0^{\text{máx}} = \frac{1}{D_b} = \frac{1}{D_2} \approx 0{,}2084 \,{}^{\text{peticiones}}/_{\text{s}}$$

   El tiempo de respuesta mínimo del servidor es:

$$R_0^{\text{mín}} = \sum_{i=1}^{2} V_i \cdot S_i = \sum_{i=1}^{2} D_i \approx 7{,}997\,\text{s}$$

3. Suponiendo que $W_i = N_i \cdot S_i$, ¿cuál es el tiempo de respuesta actual de los trabajos que llegan al servidor? ¿y el número medio de trabajos en la cola de cada dispositivo?

   Por la Ley General del Tiempo de Respuesta, el tiempo de respuesta del servidor es:

   $$R_0 = \sum_{i=1}^{2} V_i \cdot R_i$$

   Para calcular R_i, necesitamos calcular N_i. Por la Ley de Little, tenemos que:

   $$N_i = X_i \cdot R_i = X_i \cdot (W_i + S_i) = X_i \cdot (N_i \cdot S_i + S_i) \Longrightarrow N_i = \frac{X_i \cdot S_i}{1 - X_i \cdot S_i}$$

   Por tanto, tenemos que:

   $$N_1 = \frac{X_1 \cdot S_1}{1 - X_1 \cdot S_1} \approx 1{,}24982 \,\text{trabajos}$$

   $$N_2 = \frac{X_2 \cdot S_2}{1 - X_2 \cdot S_2} \approx 4{,}9794021 \,\text{trabajos}$$

   Por tanto, el tiempo de respuesta de cada dispositivo es:

   $$\begin{aligned} R_i &= W_i + S_i = N_i \cdot S_i + S_i = (N_i + 1) \cdot S_i \\ R_1 &\approx 1{,}79985 \,\text{s} \\ R_2 &\approx 3{,}18702 \,\text{s} \end{aligned}$$

   Por tanto, el tiempo de respuesta del servidor es:

   $$R_0 = V_1 \cdot R_1 + V_2 \cdot R_2 \approx 35{,}88261 \,\text{s}$$

   El número medio de trabajos en la cola de cada dispositivo es:

   $$\begin{aligned} Q_i &= X_i \cdot W_i = X_i \cdot N_i \cdot S_i \\ Q_1 &\approx 0{,}6943 \,\text{trabajos} \\ Q_2 &\approx 4{,}1466 \,\text{trabajos} \end{aligned}$$

Ejercicio 2.5.29. Los parámetros del modelo de un servidor de comercio electrónico (red abierta) son los reflejados en la Tabla 2.36 (los tiempos de la tabla se expresan en segundos). La tasa de llegada al servidor es de 1.5 transacciones por segundo.

1. Identifique el cuello de botella y calcule la productividad máxima del servidor.

   Calculamos la demanda de servicio de cada dispositivo:

   $$\begin{aligned} D_1 &= V_1 \cdot S_1 = 8 \cdot 0{,}025 = 0{,}2 \,{}^{\text{s}}/_{\text{transaccion}} \\ D_2 &= V_2 \cdot S_2 = 9 \cdot 0{,}050 = 0{,}45 \,{}^{\text{s}}/_{\text{transaccion}} \end{aligned}$$

Dispositivo	S_i	V_i
CPU (1)	0.025	8
HDD (2)	0.050	9

Tabla 2.36: Parámetros del servidor de comercio electrónico.

El cuello de botella del servidor es el HDD, ya que su demanda de servicio es mayor que la del procesador. Por tanto, la productividad máxima del servidor es:

$$X_0^{\text{máx}} = \frac{1}{D_b} = \frac{1}{D_2} = \frac{1}{0{,}45} \approx 2{,}22\,{}^{\text{transacciones}}/_{\text{s}}$$

2. ¿Cuál es la utilización de la CPU?

   Por la Relación Demanda-Utilización, tenemos que:

$$U_{cpu} = X_0 \cdot D_1 = 1{,}5 \cdot 0{,}2 = 0{,}3 = 30\,\%$$

3. ¿Cuál sería dicha utilización si la tasa de llegada fuese de $3\,{}^{\text{transacciones}}/_{\text{s}}$?

   En ese caso el servidor estaría saturado puesto que la tasa de llegada es mayor que la productividad máxima del servidor. La utilización de la CPU máxima que podría alcanzar sería:

$$U_{cpu}^{\text{máx}} = X_0^{\text{máx}} \cdot D_1 = 2{,}22 \cdot 0{,}2 = 0{,}444 = 44{,}4\,\%$$

4. ¿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?

   En este caso, la demanda de servicio de cada dispositivo HDD sería:

$$D_2' = S_2 \cdot \frac{V_2}{3} = 0{,}050 \cdot \frac{9}{3} = 0{,}15\,{}^{\text{s}}/_{\text{transaccion}}$$

   Por tanto, el cuello de botella del servidor pasaría a ser la CPU, ya que su demanda de servicio es mayor que la del HDD. Por tanto, la productividad máxima del servidor sería:

$$X_0^{\text{máx}} = \frac{1}{D_1} = \frac{1}{0{,}2} = 5\,{}^{\text{transacciones}}/_{\text{s}}$$

Ejercicio 2.5.30. 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 partir de la información anterior:

1. Calcule cuánto tiempo, de media, le dedica la unidad SSD a cada consulta que llega al servidor.

   Nos piden el valor de D_i, puesto que es el tiempo dedicado, de media, a cada petición que llega al *servidor*. Los datos que tenemos son:

   $$\begin{aligned} X_0 &= \frac{54000}{24 \cdot 3600} \approx 0{,}625 \,{}^{\text{consultas}}/_{\text{s}} \\ U_i &= 0{,}6 \\ V_i &= 5 \end{aligned}$$

   Por la Relación Demanda-Utilización, tenemos que:

   $$U_i = X_0 \cdot D_i \Longrightarrow D_i = \frac{U_i}{X_0} = \frac{0{,}6}{0{,}625} = 0{,}96 \,{}^{\text{s}}/_{\text{consulta}}$$

   Por tanto, la demanda de servicio de la unidad SSD es $D_i = 0{,}96 \,{}^{\text{s}}/_{\text{consulta}}$.

2. Calcule el tiempo medio de servicio de la unidad SSD.

   Por la definición de demanda de servicio, tenemos que:

   $$D_i = V_i \cdot S_i \Longrightarrow S_i \qquad = \frac{D_i}{V_i} = \frac{0{,}96}{5} = 0{,}192 \,{}^{\text{s}}/_{\text{consulta}}$$

Ejercicio 2.5.31. En una red interactiva formada por un servidor de impresión, durante un tiempo $T = 2$ horas, se encuentran conectados un total de $N_T = 30$ 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 45 s. 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, 25 s.

1. Calcule la productividad media del servidor.

   Como datos, nos proporcionan:

   $$\begin{aligned} N_T &= 30 \\ R_0 &= 45 \,\text{s} \\ Z &= 25 \,\text{s} \end{aligned}$$

   Por la Ley del Tiempo de Respuesta Interactivo, tenemos que:

   $$N_T = X_0 \cdot (R_0 + Z) \Longrightarrow X_0 = \frac{N_T}{R_0 + Z} = \frac{30}{45 + 25} = \frac{30}{70} = 0{,}42857 \,{}^{\text{transacciones}}/_{\text{s}}$$

2. ¿Cuántos clientes se encuentran, de media, en reflexión?

   Por la Ley de Little, el número medio de clientes en reflexión es:

   $$N_z = X_0 \cdot Z = 0{,}42857 \cdot 25 = 10{,}7143 \,\text{clientes}$$

Ejercicio 2.5.32. 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 2 s. Suponga que los trabajos (jobs) llegan con la siguiente distribución temporal:

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

A partir de la información anterior:

1. Calcule los tiempos de espera en la cola y los tiempos de respuesta que experimentan cada uno de los trabajos. Calcule finalmente sus valores medios.
2. 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.

Ejercicio 2.5.33. 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 valores reflejados en la Tabla 2.37 (los tiempos de la tabla se expresan en segundos): ¿Cuánto tiempo debería transcurrir, de media, entre que un

Dispositivo	S_i	V_i
CPU (1)	0,01	80
Disco (2)	0,5	20
Red (3)	0,24	5

Tabla 2.37: Parámetros del servidor de ayuda a la docencia.

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?

Notemos que, si 30 es el número ideal de clientes, entonces $N_T^* = 30$. Calculemos los límites optimistas de la productividad del servidor:

- Si $N_T < N_T^*$, entonces la carga es baja. El tiempo de respuesta mínimo del servidor es:

$$R_0^{\text{mín}} = \sum_{i=1}^{3} V_i \cdot S_i = 0{,}8 + 10 + 1{,}2 = 12\,\text{s}$$

  En este caso, por la Ley del Tiempo de Respuesta Interactivo, tenemos que:

$$N_T = X_0^{\text{máx}} \cdot (R_0^{\text{mín}} + Z) \Longrightarrow X_0^{\text{máx}} = \frac{N_T}{R_0^{\text{mín}} + Z} = \frac{N_T}{12 + Z}$$

- Si $N_T > N_T^*$, entonces la carga es alta. La productividad máxima del servidor es:

$$Z_0^{\text{máx}} = \frac{1}{D_b} = \frac{1}{10} = 0{,}1\,\text{s}$$

Como $N_T^* = 30$, tenemos que:

$$\frac{30}{12+Z} = 0{,}1 \Longrightarrow 12 + Z = 300 \Longrightarrow Z = 288$$

Por tanto, el tiempo que debería transcurrir, de media, entre que un estudiante recibe la respuesta de este servidor hasta que vuelve a realizar una nueva petición es de 288 s.

Ejercicio 2.5.34. Partiendo de la hipótesis de que $W_i = N_i \cdot S_i$ 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.

Calculamos el número medio de trabajos en la cola de cada estación de servicio, Q_i. Usando la Ley de Little, tenemos que:

$$Q_i = X_i \cdot W_i = X_i \cdot N_i \cdot S_i$$

Por la Ley de la Utilización, tenemos que:

$$U_i = X_i \cdot S_i$$

Por tanto, podemos expresar Q_i como:

$$Q_i = U_i \cdot N_i$$

Calculemos N_i:

$$N_i = X_i \cdot R_i = X_i \cdot (W_i + S_i) = X_i \cdot (N_i \cdot S_i + S_i) = X_i \cdot S_i \cdot (N_i + 1) \Longrightarrow$$
$$\Longrightarrow N_i = \frac{X_i \cdot S_i}{1 - X_i \cdot S_i} = \frac{U_i}{1 - U_i}$$

Por tanto, podemos expresar Q_i como:

$$Q_i = U_i \cdot \frac{U_i}{1 - U_i} = \frac{U_i^2}{1 - U_i}$$

Ahora queremos ver que, efectivamente, si $U_i < U_j$ entonces $Q_i < Q_j$. Para ello, bastará demostrar que la siguiente función es estrictamente creciente:

$$\begin{array}{rccc} f: & [0,1[& \longrightarrow & \mathbb{R} \\ & x & \longmapsto & \dfrac{x^2}{1-x} \end{array}$$

Calculamos su derivada:

$$f'(x) = \frac{(1-x)\cdot 2x + x^2}{(1-x)^2} = \frac{2x - 2x^2 + x^2}{(1-x)^2} = \frac{2x - x^2}{(1-x)^2} = \frac{x(2-x)}{(1-x)^2}$$

Como $x(2-x) > 0$ para $x \in]0,1[$, tenemos que $f'(x) > 0$ para $x \in]0,1[$. Por tanto, f es estrictamente creciente en el intervalo $]0,1[$, y por tanto si $0 < U_i < U_j < 1$, se cumple que $Q_i < Q_j$. 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.