

En SWAD he subido uno de los modelos de examen que se pusieron el pasado 19 de enero. La diferencia entre los diferentes modelos es que los datos concretos de los ejercicios son diferentes (pero la resolución es idéntica) y las preguntas tipo test pueden variar en orden y en sentido (pero con igual complejidad). Para facilitar que podáis auto-evaluaros, os pongo, a continuación, las soluciones en formato breve junto con algunas pistas de cómo llegar a dicha solución y algunos comentarios adicionales sobre su corrección.

1.-

a)F (al revés sí es cierto)

b)V

c)V

d)F (suelen tener un conector VGA básico).

e)V

f)F (PCI es paralelo)

g)V

h)F (es válida en general, es la ley de la utilización)

i)F (por ejemplo, los TPC-C tienen como resultado una productividad)

j)F (es el contratista)

2.- Se quiere ejecutar la hebra un 15% más rápido. Por tanto, queremos un speedup de 1,15 (el concepto "% más rápido" es uno de los más importantes de la asignatura y en el que más se ha incidido por lo que no darse cuenta de esto se considera un error muy grave en este ejercicio y el ejercicio completo está mal). Ahora bastaría aplicar la ley de Amdahl con  $f=0,65$  y despejar  $k$ .

$S=1,15=To/Tm=1/(1-0,65+0,65/k)$  - -  $\rightarrow k=1,25$ . El nuevo disco debe ser 1,25 veces más rápido que el anterior.

3.- La hipótesis de partida del test-t (también llamada hipótesis nula o  $H_0$ ) es que los rendimientos de ambas alternativas son equivalentes, es decir, que no hay diferencias significativas entre los rendimientos de ambas alternativas.

El p-value es la probabilidad de que dicha hipótesis sea cierta.

A un nivel de confianza del 95% el grado de significatividad,  $\alpha$ , es 0,05.

Como el p-value = 0,022 es menor que el grado de significatividad, concluimos que la probabilidad de que la hipótesis nula sea cierta es baja por lo que la descartamos. En otras palabras, los rendimientos de ambas alternativas sí que son estadísticamente diferentes para un 95% del nivel de confianza.

4.- Este ejercicio es idéntico al 3.18 de la relación (salvo que se han cambiado los valores de algunas celdas de la tabla).

a) Del grafo de llamadas, debemos deducir que:

- deriva es llamado 4 veces en todo el programa.
- redondea es llamado 25 veces.
- integra es llamado solo una vez.
- normaliza es llamado 2 veces.
- main llama una vez a integra, 2 a normaliza y 1 a deriva.

- integra llama 3 veces a deriva y 3 a redondea.
- normaliza llama 5 veces a redondea (de media). (Un fallo muy frecuente ha sido decir que llama a redondea 10 veces. Hay que tener en cuenta que normaliza se llama dos veces en total y no una).
- deriva llama 3 veces a redondea (de media). (Y no 12 veces ya que deriva se llama 4 veces en total)
- redondea no llama a ninguna otra función.

Tabla solución:

self_sec	calls	self_s/call	total_s/call	name
4	4	$4/4=1$	$1+3*0,2=1,6$	deriva
5	25	0,2	0,2	redondea
2,1	1	$7,5-3*1,6-3*0,2=2,1$	7,5	integra
2	2	1	$1+5*0,2=2$	normaliza

Otro fallo muy común ha sido calcular los self\_s/call de integra usando los self\_s/call de deriva y redondea en lugar de los total\_s/call. Por cada error en una celda se quitan 0,25 puntos por lo que tres errores hacen que el ejercicio completo esté mal.

Tiempo de CPU que consume el programa:  $4+5+2,1+2 = 13,1s$ .

Se debe optimizar el código propio de redondea que es el que más self\_secs tiene, es decir, es la función cuyo código propio consume más tiempo de CPU a lo largo de toda la ejecución del programa.

b) gprof usa temporizadores de profiling en Linux que interrumpen cada cierto tiempo la ejecución del programa para ver qué parte del código se está ejecutando. Estos temporizadores solo avanzan la cuenta cuando el procesador está ejecutando código del programa. Esa línea nos dice que cada 0,01s de tiempo de CPU real del programa se está interrumpiendo dicha ejecución para muestrear qué parte del código se está ejecutando.

5.-

En el enunciado me dan T, lambda0 y los valores de Xi, Ui(%) y Ri para los tres dispositivos. No saber identificar estos valores se considera un error muy grave. En concreto:

- Confundir el tiempo de respuesta con el tiempo de servicio se considera un error muy grave.
- Usar los valores de la utilización en tanto por ciento y no en tanto por uno en alguno de los apartados se considera un error muy grave.

a)

Debemos primero identificar el cuello de botella: Es el disco B por ser el de mayor utilización. Además, vemos que ninguna utilización se acerca al 100% por lo que el servidor no está saturado. Ya sabemos, por tanto, que  $X0=lambda0$ .

Ahora debemos calcular la demanda de servicio del cuello de botella (disco B).

$D_{discoB} = B_{discoB} / C0 = U_{discoB} / X0 = 0,81 / 15 = 0,054s$ .

$$X0_{max}=1/0,054=18,52\text{tr/s.}$$

Por tanto, cuando el valor medio de la tasa de llegada de este servidor supere las 18,52 páginas web por segundo el servidor estará saturado.

Realizar cualquier mejora en un dispositivo que no sea el cuello de botella (en este caso me dicen que hay una mejora en el disco A) no va a afectar a la productividad máxima del servidor.

b) Me piden  $N0$ .

Para simplificar la notación, llamo "1" a la CPU, "2" al disco A y "3" al disco B.

$$N0=N1+N2+N3=X1*R1+X2*R2+X3*R3=5,49 \text{ usuarios.}$$

donde he aplicado la ley de Little a cada estación de servicio (sabiendo que el sistema no está saturado).

Otra alternativa:

(Ley de Little aplicada al servidor no saturado):  $N0=X0*R0$  y se calcula  $R0$  a partir de la Ley General del Tiempo de Respuesta:  $R0=V1*R1+V2*R2+V3*R3$

En ese caso, habría que calcular primero las razones de visita:  $Vi=Ci/C0=Xi/X0$ .

c) Me piden  $Q_{discoB}=Q3$ .

Aplicando la ley de Little a la cola de la estación de servicio:

$$Q3=\lambda_3*W3 = X3*W3 \text{ (por no estar saturado) } = X3*(R3-S3).$$

Debería calcular  $S3$  (el resto ya lo conocemos).

$$S3=B3/C3=U3/X3=0,81/90=0,009\text{s}$$

$$\text{Por lo tanto: } Q3=90*(0,047-0,009)=3,42 \text{ peticiones.}$$

d) Está claro que solo nos conviene actuar sobre el cuello de botella: en este caso, el disco B. Como hay otro disco en el sistema, lo lógico es distribuir mejor la carga entre ambos discos de tal forma que sus demandas de servicio se igualen (y así también lo harán sus utilizaciones). Pero eso no se puede hacer alegremente ya que el número medio de visitas a los discos debe mantenerse. Por tanto, habría que repartir la carga entre los discos (nuevos valores de  $V2$  y  $V3$  a los que llamaré  $V2'$  y  $V3'$ ) de tal forma que:

$$V2'+V3'=V2+V3$$

$$S2*V2'=S3*V3'$$

Como dice el ejercicio, no hace falta llegar a la solución final. Con llegar hasta este planteamiento el ejercicio estaría correcto.

6.-

a) Diapositiva 42 del tema 2. Proporciona varios conectores a través de los cuales la placa base se puede comunicar con el chasis. Entre ellos, está el del botón de encendido del computador, el botón de reset, el altavoz del chasis, el LED de encendido y el LED acceso a las unidades de almacenamiento.

b) Diapositiva 22 del tema 2. Son módulos de memoria DRAM en los que hay registros donde se almacena tanto las señales de control (entre ellas, la dirección de memoria a la que se quiere acceder) como los datos. Esto permite que cada módulo pueda tener

mayor capacidad de memoria por lo que se favorece la escalabilidad. De hecho, es la tecnología de memoria actual que permite la mayor cantidad de memoria RAM dinámica en una placa base. Las siglas "LR" vienen de "Load Reduced" (carga reducida) y DIMM viene de Dual In-line Memory Module.

c) Latencia es sinónimo de tiempo de respuesta (tema 1). Las unidades SSD son chips de memoria flash en los que el tiempo de respuesta no depende de la localización del dato a buscar. Sin embargo, los HDD son discos magnéticos en los que la información está distribuida en una serie de pistas y existe un cabezal que debe primero buscar la pista donde se encuentra el dato a buscar y después esperar a que la rotación del disco sitúe el sector concreto a buscar debajo del cabezal. Por lo tanto, es razonable pensar que las unidades SSD tendrán menor latencia que los discos duros.

Algunos confunden latencia con ancho de banda o incluso hablan de "velocidad". Son conceptos diferentes y, por supuesto, esa respuesta sería incorrecta. Otros simplemente dicen qué tecnología es la más rápida sin aportar ningún razonamiento (también sería incorrecto).

d) Diapositiva 45 del tema 2: El puente sur del juego de chips se encarga de las transferencias entre el puente norte del chipset y el resto de periféricos con menores exigencias de velocidad de la placa.

e) Dispositiva 22 del tema 2: ECC viene de "Error Correcting Code". Son módulos de memoria DRAM que incorporan bits redundantes para poder detectar y corregir errores de lectura. Eso hace que sean más fiables ya que disminuye la probabilidad de que haya un error en la lectura del dato de memoria. Como se dice en la diapositiva 19 del tema 1: "Un sistema es fiable cuando desarrolla su actividad sin presencia de errores." Un matiz importante: no es lo mismo la tolerancia a fallos (diapositiva 18 del tema 1) que la fiabilidad.

f) En esta pregunta muchos han contestado definiendo lo que es un monitor por eventos y lo que es un monitor por muestreo. Eso NO se pide. Solo se pide la ventaja y la desventaja de un monitor por eventos frente a uno por muestreo. La respuesta está en la diapositiva 9 del tema 3:

Ventaja: Información exacta. Cuando ocurre el evento que queremos medir lo detectamos y lo medimos mientras que un monitor por muestreo mide cada cierto tiempo y la información es solo estadística.

Desventaja: El volumen de información recogida depende de la frecuencia de los eventos. Si los eventos ocurren con mucha frecuencia es posible que el volumen sea excesivo (o la sobrecarga sea excesiva en el caso de que el monitor consuma recursos del sistema). Esto no pasa con los monitores por muestreo en los que dicho volumen depende solo del periodo de muestreo y se puede controlar fácilmente.

g) Diapositiva 28 del tema 5: La demanda de servicio de un dispositivo es la cantidad de tiempo que dicho dispositivo le dedica a cada trabajo que realiza el servidor.

Cualquier otra respuesta que se base en una fórmula concreta se considera incorrecta ya que se pide el SIGNIFICADO de un concepto y no las operaciones que hay que

realizar para su cálculo.

h) Diapositiva 14 del tema 5: Es un parámetro que representa el tiempo que requiere el usuario antes de volver a lanzar una petición al servidor. Se usa en redes de colas cerradas de tipo interactivo.

7.-

Supongamos que tenemos un benchmark con n programas. Sean:

- $t_{Ri}$  lo que tarda la máquina de referencia en ejecutar el programa de benchmark i-ésimo.
- $t_{Ai}$  lo que tarda la máquina A en ejecutar el programa de benchmark i-ésimo.
- $t_{Bi}$  lo que tarda la máquina B en ejecutar el programa de benchmark i-ésimo.

Supongamos que la máquina B es el doble de rápida que la máquina A en todos los programas del benchmark:  $t_{Bi} = t_{Ai}/2 \rightarrow t_{Ai} = 2 * t_{Bi}$

Diapositiva 38 del tema 4 (aunque es difícil escribir las fórmulas en texto plano):

$$SPECB = (t_{R1}/t_{B1} * t_{R2}/t_{B2} * \dots * t_{Rn}/t_{Bn})^{(1/n)} = (t_{R1} * t_{R2} * \dots * t_{Rn})^{(1/n)} / (t_{B1} * t_{B2} * \dots * t_{Bn})^{(1/n)}$$

$$SPEC A = (t_{R1}/t_{A1} * t_{R2}/t_{A2} * \dots * t_{Rn}/t_{An})^{(1/n)} = (t_{R1} * t_{R2} * \dots * t_{Rn})^{(1/n)} / (t_{A1} * t_{A2} * \dots * t_{An})^{(1/n)}$$

Por tanto:

$$SPECB/SPEC A = (t_{A1} * t_{A2} * \dots * t_{An})^{(1/n)} / (t_{B1} * t_{B2} * \dots * t_{Bn})^{(1/n)} = (t_{A1}/t_{B1} * t_{A2}/t_{B2} * \dots * t_{An}/t_{Bn})^{(1/n)} = (2 * 2 * \dots * 2)^{(1/n)} = (2^n)^{(1/n)} = 2.$$

Por lo tanto la relación es que  $SPECB = 2 * SPEC A$

8.- Diapositiva 20 del tema 6 particularizándola al caso concreto que se indica en el enunciado.

-----