

Examen Febrero 2019

1.- Responda a las siguientes afirmaciones indicando V (Verdadero) o F (Falso) en la tabla que aparece al final del examen.

- a) Las unidades SSD tienen mayor latencia que los discos duros debido a que no tienen que esperar a que el cabezal se posicione sobre la pista a leer/escribir.
- b) Bajo ningún concepto se puede hacer referencia a una fabricación o una procedencia determinada cuando se especifican los componentes a instalar o suministrar en un pliego de prescripciones técnicas.
- c) Las pistas en una placa base están normalmente hechas de cobre rodeadas de láminas de un sustrato no conductor.
- d) El número medio de trabajos en un servidor es la suma de los números medios de trabajos en cada uno de sus componentes.
- e) Las unidades SSD son capaces de alcanzar anchos de banda superiores a los que el protocolo SATA-3 puede proporcionar.
- f) Los códigos CPV deben figurar obligatoriamente en todos los pliegos de prescripciones técnicas.
- g) En los paneles traseros de las placas de servidores los conectores de red son de bajas prestaciones.
- h) En un test ANOVA, si F_{exp} es menor que el grado de significatividad rechazamos la hipótesis nula y concluiremos que el factor a considerar si que influye en la variable respuesta.
- i) Con "sar -u" iremos obteniendo la información de la utilización del procesador desde el momento actual en adelante.
- j) El puente sur del chipset se encarga de la comunicación con la RAM.
- k) La productividad de un servidor nunca podrá ser superior a $\frac{1}{D}$.
- l) Intel Xeon es la familia de microprocesadores de Intel especializada en servidores.
- m) Los procesadores AMD Opteron Serie A están basados en microprocesadores de ARM.
- n) $\lambda_i = \frac{Q_i}{W_i}$
- o) Si $\sum_{i=1}^k U_i > 1$ el servidor está saturado.

a) F	b) F	c) V	d) V	e) V
f) F	g) F	h) F	i) F	j) F
k) F	l) V	m) V	n) V	o) F

2.- Una mejora en un sitio web ha permitido rebajar de 25 a 21 segundos el tiempo medio de descarga de sus páginas. Si la mejora ha consistido en reemplazar el subsistema de discos que almacena las páginas del servidor por otro un 40% más rápido, ¿Qué tanto por ciento del tiempo medio de descarga mejorado dedica ahora (= tras la mejora) el sitio web a acceder el subsistema de discos?

$$T_0 = 25s \rightarrow T_m = 21s \rightarrow k = 1,4$$

$$T_m = (1 - f) * T_m + f * T_m \rightarrow T_0 = (1 - f) * T_m + f * T_m * k$$

$$25 = (1 - f) * 21 + f * 21 * 1,4 \rightarrow 25 = 21 - 21 * f + 29,4 * f$$

$$25 - 21 = -21f + 29,4f \rightarrow 4 = 8,4f \rightarrow f = \frac{4}{8,4} = 0,4762$$

3.- Después de instrumentar un programa escrito en C con la herramienta gprof el resultado obtenido ha sido el siguiente:

%time	cumulative seconds	self seconds	calls	self s/call	total s/call	name
xxx	xxx	xxx	2	8	xxx	reduce
xxx	xxx	12	xxx	2	xxx	invierte
xxx	xxx	xxx	24	0,2	0,2	calcula

Teniendo en cuenta que el grafo de dependencias muestra que “invierte” es llamado únicamente desde el procedimiento “reduce”, que “calcula” es llamado únicamente desde “invierte” y que podemos despreciar el tiempo propio de “main”, dibuje un diagrama en donde se vea claramente que procedimiento llama a que otro y el número de veces que lo hace de media. A partir de ese diagrama, calcule de forma razonada cuanto se tarda, de media, en ejecutar una llamada al procedimiento reduce incluyendo los procedimientos a los que este llama.

“Main” llama, de media, 2 veces a reduce.

“Reduce” llama, de media, 3 veces a invierte.

“Invierte” llama, de media, 4 veces a calcula.

% tiempo	Cumulative Seconds	Self Seconds	Calls	Self s/call	Total s/call	Name
X	X	X	2	8	$8+4*0,2 = 2,8$	Reduce
X	X	12	X	2	$2+3*2,8 = 16,4$	Invierte
X	X	X	24	0,2	0,2	Calcula

4.- Considere los tiempos de ejecución, en segundos, obtenidos en los computadores Ref (referencia), A y B para un conjunto de cuatro programas de un benchmark:

Programa	Ref(s)	A (s)	B (s)
P1	1200	300	250
P2	2400	900	1200
P3	3600	700	1200
P4	1200	600	1000

- a) Calcule, a la manera de SPEC, un índice de rendimiento para A y B utilizando Ref como máquina de referencia, y compare el rendimiento de ambas máquinas usando este índice. ¿Qué máquina es más rápida según ese índice? Justifique la respuesta.

Programa	Ref/A	Ref/B
P1	1200/300 = 4	1200/250 = 4,8
P2	2400/900 = 2,67	2400/1200 = 2
P3	3600/700 = 5,14	3600/1200 = 3
P4	1200/600 = 2	1200/1000 = 1,2

$$SPEC_A = (4 * 2,67 * 5,14 * 2)^{\frac{1}{4}} = 3,24$$

$$SPEC_B = (4,8 * 2 * 3 * 1,2)^{\frac{1}{4}} = 2,42$$

El índice SPEC es un índice que cuanto mayor sea (media geométrica) es mejor. Según ese índice, la máquina A es más rápida que la B.

- b) Determine ahora si existen diferencias significativas, para un nivel de confianza del 95%, en el rendimiento de los computadores A y B. Justifique la respuesta indicando claramente cuál es la hipótesis de partida. Datos: $|t_{0,025, 3}| = 3,18$; $|t_{0,025, 4}| = 2,78$; $|t_{0,05, 3}| = 2,35$; $|t_{0,05, 4}| = 2,13$.

$$H_0: \text{Rendimiento A} \equiv \text{Rendimiento B} (d_{rea} = 0)$$

$$d = \frac{\sum_{i=1}^n d_i}{n} = \frac{(300 - 250) + (900 - 1200) + (700 - 1200) + (600 - 1000)}{4} = \frac{50 - 300 - 500 - 400}{4} = -\frac{1150}{4} = -287,5$$

$$s = \sqrt{\frac{\sum_{i=1}^n (d_i - d)^2}{n - 1}} = \sqrt{\frac{(50 + 287,5)^2 + (-300 + 287,5)^2 + (-500 + 287,5)^2 + (-400 + 287,5)^2}{4 - 1}} = \sqrt{\frac{113906,25 + 156,25 + 45156,25 + 12656,25}{3}} = \sqrt{\frac{171875}{3}} = \sqrt{57291,67} = 239,36$$

$$t_{exp} = \frac{d}{s/\sqrt{n}} = \frac{-287,5}{\frac{239,36}{\sqrt{4}}} = -\frac{287,5}{119,68} = -2,41$$

Para descartar H_0 para un nivel de confianza del 95% ($\alpha = 0,05$), t_{exp} debería estar fuera del rango $[-3,18, 3,18]$ (ya que $\frac{\alpha}{2} = 0,025$ y $n - 1 = 3$). Como esta dentro de dicho rango, no podemos rechazar la hipótesis.

5.- Un servidor web recibe, por término medio, 4 peticiones de páginas web por segundo. Los valores medios de los tiempos de servicio y de las utilizaciones de los dispositivos que más influyen en el rendimiento de este servidor web se indican en la siguiente tabla:

Dispositivo	Tiempo de servicio (ms)	Utilización (%)
CPU	10	32
Disco A	40	64
Disco B	30	36

A partir de la información anterior:

- a) Calcule los valores medios de las razones de visita y de las demandas de servicio de cada dispositivo.

$$U_i = X_0 * D_i \rightarrow D_i = V_i * S_i$$

Dispositivo	S_i	U_i	D_i	V_i
CPU	0,01	0,32	$0,32/4 = 0,08$	$0,08/0,01 = 8$
Disco A	0,04	0,64	$0,64/4 = 0,16$	$0,16/0,04 = 4$
Disco B	0,03	0,36	$0,36/4 = 0,09$	$0,09/0,03 = 3$

- b) ¿Qué valor tendría que tener la tasa de llegadas para que el cuello de botella fuese otro dispositivo diferente del actual?

El dispositivo cuello de botella es el que tiene mayor utilización o, como λ_0 se multiplica a todos por igual, la mayor demanda de servicio de forma independiente a la tasa de llegadas. Esta no afecta a la determinación del cuello de botella.

- c) ¿Qué tiempo de servicio debería tener el dispositivo cuello de botella para multiplicar por dos la productividad máxima del servidor?

El cuello de botella es el Disco A porque tiene mayor utilización.

$$D_b = D_{DiscoA} = 0,16s$$

$$X_0^{max} = \frac{1}{D_{DiscoA}} = \frac{1}{0,16} = 6,25 \frac{tr}{s}$$

Podría pensarse que se lograría doblar X_0^{max} actuando únicamente sobre el tiempo de servicio del Disco A (haciendo que valga la mitad), pero eso haría que el cuello de botella dejase de ser el Disco A para pasar a ser el Disco B. Es el Disco B el que entonces limitaría la productividad máxima y esta nunca podría superar los 11,1 tr/s.

- d) Calcule la pérdida en prestaciones en R_0^{min} y en X_0^{max} del servidor si eliminásemos el Disco A e hiciéramos, por tanto, que todos los accesos a disco se tuvieran que hacer al Disco B (basta con que indique los valores de R_0^{min} y X_0^{max} antes y después del cambio).

Antes del cambio: $R_0^{min} = 0,08 + 0,16 + 0,09 = 0,33s \rightarrow X_0^{max} = 6,25 \frac{tr}{s}$

La razón de visita del Disco B será igual a la suma de la razón de visita de ambos discos, 7.

Dispositivo	S_i	V_i
CPU	0,01	8
Disco B	0,03	7

$$R_0^{mi} = 0,08 + 0,03 * 7 = 0,08 + 0,21 = 0,29s \rightarrow \text{Mejora de } 0,04s$$

$$X_0^{max} = \frac{1}{0,21} = 4,76 \frac{tr}{s} \rightarrow \text{Empeoramiento}$$

- e) Suponiendo que $W_i = N_i * S_i$, encuentre, indicando de forma explícita el nombre de las leyes operacionales que necesite, una expresión que permita calcular el tiempo medio de respuesta de una estación de servicio en función de su tiempo de servicio y de su utilización.

$$\text{Ley de Little} \rightarrow N_i = \lambda_i * R_i$$

$$\text{Ley de Flujo Forzado} \rightarrow X_i = X_0 * V_i = \lambda_0 * V_i = \lambda_i$$

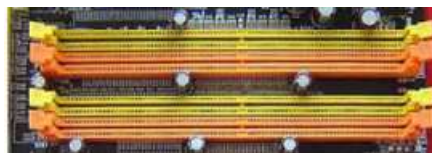
$$U_i = X_0 * D_i = \lambda_0 * D_i$$

$$R_i = W_i + S_i = N_i * S_i + S_i = \lambda_i * R_i * S_i + S_i$$

$$R_i = \frac{S_i}{1 - \lambda_i * S_i} = \frac{S_i}{1 - \lambda_0 * V_i * S_i} = \frac{S_i}{1 - \lambda_0 * D_i} = \frac{S_i}{1 - U_i}$$

6.- Cuestiones cortas.

- a) ¿Qué tipo de conectores son los que aparecen en la figura de la derecha? ¿Proporcionan una conexión full-duplex con el procesador o half-duplex?



Son ranuras para conectar módulos de DRAM. Proporcionan conexión half-duplex con la CPU.

- b) En el contexto del análisis de rendimiento basado en experimentos, ¿Qué diferencia hay entre los conceptos de "factor" y de "nivel"?

Factor es cada una de las variables que pueden afectar a la variable respuesta. Nivel es cada uno de los valores que puede asumir un factor.

c) ¿Por qué si PCIe 3.0 alcanza un ancho de banda de 8GT/s por cada LANE y cada sentido, dicho ancho de banda medido en bytes de información transmitida es de aproximadamente 1GB/s?

En el caso de PCIe 3.0, la codificación es 128b/130b por lo que apenas hay overhead en la transmisión y prácticamente todos los bits transmitidos son de información.

d) Indique las principales características de perf.

Perf es un conjunto de herramientas para el análisis de rendimiento en Linux basadas en eventos software y hardware (hacen uso de contadores hardware disponibles en los últimos microprocesadores de Intel y AMD). Permiten analizar el rendimiento de a) un hilo individual, b) un proceso + sus hijos, c) todos los procesos que se ejecutan en una CPU concreta, d) todos los procesos que se ejecutan en el sistema.

e) ¿Qué diferencia hay entre un banco de memoria y un rango de memoria?

Un banco de memoria es una agrupación de módulos de memoria que se comunican con la CPU a través de un mismo canal de memoria. Cada módulo de memoria está, a su vez, distribuido en rangos de memoria que no son más que agrupaciones de chips que me proporcionan una palabra completa de 64 bits (72 bits en caso de memorias ECC). Por lo tanto, un rango de memoria es como si tuviéramos un “sub-banco de memoria” dentro del propio módulo de memoria.

f) ¿Por qué hot-swapping mejora la disponibilidad en un servidor?

El hot-swapping consiste en el intercambio de componentes sin necesidad de apagar el computador o tener que reiniciarlo. Esto reduce el tiempo en el que el computador debe estar no operativo debido a ese intercambio y, por tanto, hace que mejore su disponibilidad.

g) ¿A que nos referimos como free-cooling cuando hablamos de ingeniería de servidores?

Utilización de bajas temperaturas exteriores para refrigeración gratuita.

h) ¿A que nos referimos cuando hablamos de memorias con ECC?

ECC viene de Error Correcting Code. Son módulos de memoria DRAM que incorporan bits redundantes (8 bits por cada 64) 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.

i) Cite al menos tres diferencias entre SRAM y DRAM.

SRAM tiene menor latencia que DRAM, mayor coste por bit, no necesita refresco. SRAM se suele encontrar formando las caches dentro del micro y la DRAM suele estar fuera, pinchada en la placa.