

Tema 1: Introducción a la Ingeniería de Servidores

1.1. ¿Qué es un servidor?

1. Un sistema aislado es un sistema computacional en el que existe distribución de la información y que no interactúa con otros sistemas.
2. En una arquitectura cliente-servidor las tareas se reparten entre los servidores y los clientes.
3. En una arquitectura cliente-cola-cliente, los clientes actúan como una cola que va capturando las peticiones del servidor y sincronizando el funcionamiento del sistema.

1.2. Fundamentos de Ingeniería de Servidores

4. El rendimiento (performance) es una medida de tiempo con que se realiza una determinada carga.
5. El tiempo de respuesta o latencia es el tiempo total desde que se solicita una tarea al servidor o a un componente del mismo y la finalización de la misma.
6. La productividad es la cantidad de tiempo requerido por el servidor o por un componente del mismo para realizar una tarea.
7. Un servidor se encuentra disponible solo si se encuentra en estado operativo.
8. El tiempo de inactividad planificado hace referencia a las actualizaciones de hardware o software que requieren rearranques.
9. La confidencialidad hace referencia a la corrupción o alteración no autorizada de datos.
10. Todos los sistemas escalables son extensibles pero no a la inversa.
11. Todos los sistemas extensibles son escalables pero no a la inversa.

12. La escalabilidad hace referencia a la capacidad del sistema para aumentar sus características o recursos.

13. La reducción del uso de electricidad no se considera un factor importante a la hora de calcular el coste.

1.3 Introducción a la comparación de características entre servidores

14. Si $t_A=5s$ y $t_B=10s$, entonces B es el doble de rápido que A.

15. La velocidad de una máquina es inversamente proporcional a su tiempo de ejecución.

16. La ganancia de una máquina A con respecto a una máquina B viene dada por t_B/t_A .

1.4. Introducción a la Ley de Amdahl

17. Si $f=0$, no hay ninguna mejora en el sistema.

18. Si $f=0$, el sistema mejora tantas veces como el componente.

19. Si $f=1$, no hay ninguna mejora en el sistema.

20. Si $f=1$, el sistema mejora tantas veces como el componente.

21. El tiempo no utilizado por el recurso viene dado por $(1-f)*T_0$.

Soluciones Tema 1:

1.1: F, V, F

1.2: F, V, F, V, V, F, V, F, F, F

1.3: F, V, V

1.4: V, F, F, V, V

Tema 2: Componentes hardware de un ordenador

1. A la placa base se conectan los componentes software del ordenador.
2. La fuente de alimentación convierte corriente continua en alterna.
3. Los zócalos PGA-ZIF tienen las hendiduras donde entran los pines del CPU.
4. Es el propio microprocesador de muchas placas base actuales el que realiza la función de puente norte del chipset.
5. Las pistas de en una placa base están normalmente hechas de cobre rodeadas de láminas de un sustrato no conductor.
6. En los zócalos LGA se insertan los pines del CPU.
7. Un procesador dedicado a servidores tiene un mayor número de núcleos.
8. Un procesador dedicado a servidores no suele incorporar soporte para multiprocesamiento.
9. Los procesadores de Intel dedicados a servidores se denominan Xeon.
10. AMD fueron los primeros en sacar el repertorio de 64 bits en 1996.
11. Los procesadores de IBM fueron los primeros en implementar el repertorio de instrucciones de 64 para la arquitectura x64.
12. La DRAM es volátil y necesita refresco.
13. Las prestaciones de SRAM frente a DRAM son menores pero ofrecen mayor densidad y menor coste por bit.

14. El número de contactos en las RAM aumenta al evolucionar la generación.
15. El bus de datos de las RAM se mantiene al evolucionar la generación.
16. El voltaje aumenta al evolucionar la generación.
17. Las siglas ECC hacen referencia a “Error Correcting Code”.
18. Un microprocesador puede acceder simultáneamente a dos módulos del mismo banco de memoria.
19. A partir de las memorias SIPP, los pines coinciden por delante y por detrás.
20. La interfaz PCI es Half-Duplex.
21. La versión PCI para servidores se conoce como PCI-X.
22. El bus paralelo de AGP puede ser de 64 bits.
23. La versión PCIE es full-duplex.
24. Cada LANE del puerto PCIE está compuesta por 4 cables en el mismo sentido de la comunicación.
25. En la codificación 8B/10b, de cada 10 bits de información transmitida, 8 son de datos y 2 de redundancia.
26. Cada pista de un disco duro se subdivide en sectores de 512 bytes.
27. Las unidades SSD suelen ser menos rápidas que los discos HDD.
28. Las unidades SSD son capaces de alcanzar anchos de banda superiores a los que el protocolo SATA-3 puede proporcionar.
29. El puente sur del chipset se encarga de la comunicación con la RAM.
30. Los procesadores AMD Opteron están basados en microprocesadores de ARM.

31. Un módulo de DRAM con chips en ambas caras no tiene por qué ser de doble rango.
32. Tanto las SRAM como las DRAM son volátiles pero sólo las DRAM necesitan refresco.
33. Las primeras instrucciones que ejecuta un procesador en el arranque proceden de las primeras direcciones de la DRAM.
34. El almacenamiento en los SSD es volátil.
35. Existe RAID por software y por hardware.
36. Jugando con las distintas configuraciones de un RAID se puede conseguir más fiabilidad o más disponibilidad pero no mayores prestaciones.
37. Con “almacenamiento permanente” queremos indicar que el dispositivo es de solo lectura y que el dato se va a quedar escrito en el dispositivo de forma permanente.
38. Los protocolos de comunicación serie pueden ser half-duplex.
39. En RAID 1, el contenido total está almacenado en dos discos diferentes.
40. La interfaz PATA es half-duplex.
41. En la interfaz PATA los dispositivos del conector comparten el mismo bus.
42. La longitud del cable SATA es mayor que la del cable PATA.
43. La interfaz SCSI permite conectar varias unidades en cadena.
44. La interfaz SATA es half-duplex.
45. Los conectores mini-SAS permiten hasta 5 conectores SAS O SATA.

46. Con un conector SATA de la placa base puedo conectar 4 discos mini-SAS.
47. La interfaz SATA es compatible con SAS. Dicho de otra manera, si tengo una placa con conectores SATA voy a poder pinchar en ella unidades SAS.
48. Los paneles traseros de placas de servidores suelen tener conectores de audio y vídeo de altas prestaciones.
49. Un microprocesador puede acceder simultáneamente a dos módulos de memoria DRAM solo si están en canales diferentes.
50. Tanto PCI como PCIe definen un protocolo de comunicación serie punto a punto.
51. La versión serie del interfaz SCSI se llama M.2.
52. Todos los protocolos de comunicación serie son full-duplex.
53. En los paneles traseros de las placas base de servidores los conectores de red son de bajas prestaciones.
54. USB 2.0 ES FULL-duplex.
55. El puente norte se encarga de las transferencias de menor velocidad.

Soluciones Tema 2:

1.F	11.F	21.V	31.V	41.V	51.F
2.F	12.V	22.F	32.V	42.V	52.F
3.V	13.F	23.V	33.F	43.V	53.F
4.V	14.V	24.F	34.F	44.V	54.F
5.V	15.V	25.V	35.V	45.F	55.F
6.F	16.F	26.V	36.F	46.F	
7.V	17.V	27.F	37.F	47.F	
8.F	18.F	28.V	38.V	48.F	
9.V	19.F	29.F	39.F	49.V	
10.F	20.V	30.V	40.V	50.F	

Tema 3: Componentes hardware de un ordenador

3.1. Concepto de monitor de actividad

1. La carga es el conjunto de tareas que ha de realizar el servidor.
2. Algunas variables que reflejan la carga son los procesadores o la memoria.
3. Un monitor por eventos ofrece información estadística.
4. La realización de gráficas diversas o modificación de parámetros del propio monitor corresponden a la interacción con el administrador.
5. La precisión de un sensor está relacionada con la dispersión de las medidas que realiza.

3.2. Monitorización a nivel de sistema

6. A través de `/proc` podemos acceder a cada uno de los procesos del sistema usando `/proc/pid/[pid]`.
7. El primer valor que arroja la salida de `uptime` hace referencia a la carga media del sistema en el último minuto.
8. La cola de procesos “run queue” está formada por procesos en estado “runnable” y “I/O blocked”.
9. En el monitor `vmstat` aparecen procesos en estado “running” y “runnable”.
10. El monitor `sar` recopila información sobre la actividad de un componente.
11. “-f” Es un parámetro de `sar` e indica el fichero donde extraer la información.

12. Con carga del sistema, el S.O de Linux se refiere al número de procesos en modo “running”, “Runnable” o “I/O blocked”.
13. sar es un monitor software por eventos.
14. Con “sar -u” iremos obteniendo la información de la utilización del procesador desde el momento actual en adelante.
15. En Linux, la carga del sistema (system load) hace referencia al número de procesos en la cola de procesos del núcleo (run queue) más los bloqueados por E/S (I/O blocked).
16. En Linux, la información a la que se accede a través de /proc está almacenada realmente en la RAM.

3.3. Monitorización a nivel de aplicación

17. El objetivo de un profiler es monitorizar la actividad generada por un componente concreto.
18. El profiler gprof estima el tiempo de CPU que consume cada función de un proceso/hilo.

Soluciones Tema 3:

3.1: V, V, F, V, V

3.2: F, V, F, F, F, V, F, F, V, V, F

3.3: F, V

Tema 4: Análisis comparativo del rendimiento

4.1. Referenciación (Benchmarking)

1. La frecuencia de reloj es un buen indicador de rendimiento.
2. Los modelos de carga son representaciones exactas de la carga que recibe un sistema informático.
3. La carga de prueba es muy representativa de la carga real.
4. El modo “base” de SPEC permite a todos los programas escritos en el mismo lenguaje usar las mismas opciones de compilación.
5. Cuando nos referimos al índice de prestaciones que se calcula según el benchmark SPEC CPU2017, el SPEC pico nunca puede ser menor que el SPEC base.
6. El índice SPEC es la media geométrica de las generaciones en velocidad con respecto a una máquina de referencia.
7. El resultado de un benchmark siempre se expresa como el tiempo necesario para ejecutar una cantidad preestablecida de tareas.
8. Una máquina con un SPEC menor que 1 significa que es más rápida que la máquina de referencia.

4.2. Análisis de los resultados de un test de rendimiento

9. El speedup es un índice a minimizar.
10. La media aritmética es independiente de ninguna máquina de referencia.
11. Uno de los inconvenientes de utilizar la media aritmética de los tiempos de ejecución de los programas de un benchmark es que su valor dependerá de la máquina de referencia que se escoja.

4.3. Comparación de prestaciones en presencia de aleatoriedad

12. Cuando el p-valor sea mayor que alpha diremos que las máquinas A y B tienen rendimientos diferentes.
13. Cuando decimos que el rendimiento de dos máquinas es equivalente, eso implica que el valor de $d_{\text{real}} = 0$.
14. La hipótesis inicial de un test t es que los rendimientos de ambas alternativas a analizar son estadísticamente diferentes.
15. En un test t, el valor obtenido depende del nivel de confianza seleccionado.
16. En un test t, a mayor nivel de confianza es más fácil rechazar la hipótesis de que ambas alternativas son iguales.

4.4. Diseño de experimentos de comparación de rendimiento

17. El nivel es cada uno de los valores que puede aceptar un factor.
18. Una interacción ocurre cuando el efecto de un factor cambia para diferentes niveles de otro factor.
19. 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 sí que influye en la variable respuesta.

Soluciones Tema 4:

- 4.1: F, F, V, V, V, V, F, F
4.2: F, V, F
4.3: F, V, F, F, F
4.4: V, V, F

Tema 5: Optimización del rendimiento de un servidor mediante análisis operacional.

5.1. Introducción: Redes de Colas de Espera

1. Normalmente, un dispositivo o recurso sólo puede ser usado por un trabajo a la vez. El resto de trabajos tendrá que esperar.
2. Las estaciones de servicio pueden no tener una cola de espera.
3. Si un trabajo realiza un acceso a una unidad de E/S, después no volverá a pasar por el procesador.
4. Las estaciones con más de un servidor pueden atender a más de un trabajo en paralelo.
5. En una red de colas cerrada, el número de trabajos es variable.
6. Las redes de colas cerradas se dividen en redes de tipo batch y redes cerradas interactivas.

5.2. Variables y Leyes Operacionales

7. La demanda de servicio de un dispositivo es el tiempo medio que dedica el dispositivo en servir cada trabajo que realiza.
8. $Q_i = X_0 * W_i$.
9. $X_i = V_i * X_0$.
10. $R_0 = R_1 + R_2 + \dots + R_K$.
11. La expresión $N_i = X_i * R_i$ solo es válida si el servidor no está saturado.
12. $S_i = C_i / B_i$.

13. Si reemplazamos un dispositivo por otro el doble de rápido, su razón de visita será menor.
14. La expresión $U_i = X_i * S_i$ es válida si el servidor no está saturado.
15. $N_0 = N_1 + N_2 + .. + N_K$.
16. La expresión $Q_i = \lambda_i * W_i$ es válida, aunque el servidor no se encuentre saturado.
17. $Z = N_Z * R_0$.
18. $R_0 = V_1 R_1 + V_2 R_2 + ...$
19. Si aplicamos la Ley de Little a la cola de una estación de servicio obtenemos que $Q_i = \lambda_i * W_i$.
20. La expresión $W_i = N_i * S_i$ es una ley operacional.
21. El número medio de trabajos en un servidor es la suma de los números medios de trabajos en cada uno de sus componentes.
22. La razón media de visita de un dispositivo no tiene por qué ser un número entero.
23. $Q_i = \lambda_i * W_i$.
24. En un servidor modelado mediante una red de colas se cumple que $B_i = (N_i - Q_i) * T$.
25. En un servidor, si la demanda de servicio de un dispositivo es menor que la de otro, su utilización nunca podrá ser mayor que la de ese otro dispositivo.

5.3. Límites optimistas del rendimiento

26. Actuando sobre el elemento con mayor razón de visita nos garantizamos mejorar la productividad máxima de un servidor.
27. En un servidor equilibrado todos los dispositivos son cuellos de botella.

28. Si un servidor siempre responde a nuestras peticiones podemos decir que es un servidor de alta fidelidad.
29. El tiempo medio de respuesta de un servidor es la suma de los tiempos medios de respuesta de todos sus componentes.
30. Un mismo servidor, según su tipo de carga, puede tener distintos cuellos de botella.
31. En un servidor modelado mediante una red abierta la tasa media de llegadas no puede superar $1/D_b$.
32. Si $N_T \gg N_T^*$ el servidor está saturado (para el caso de redes de colas cerradas).
33. La productividad de un servidor nunca podrá ser superior a $1/D$.
34. Si la sumatoria de las utilizaciones es mayor a 1, el servidor está saturado.
35. Si el número total de usuarios en un servidor modelado mediante una red de colas cerradas interactiva es superior a N_T^* , entonces el servidor está saturado.

Soluciones Tema 5:

5.1: V, F, F, V, F, V

5.2: F, F, V, F, V, F, F, F, V, V, F, F, V, F, V, V, V, V, V

5.3: F, V, F, F, V, F, F, F, F

Tema 6: Pliegos de prescripciones técnicas

1. El contratante es el proponente que resulta adjudicatario de una licitación y quien finalmente se encargará de llevar a cabo los trabajos contratados.
2. El pliego de prescripciones técnicas debe contener, con carácter general, la información sobre el plazo de ejecución o duración del contrato.
3. La sección de garantías de un pliego de prescripciones se refiere esencialmente a la garantía en la realización del pago del contrato.
4. Bajo ningún concepto se puede hacer referencia a una fabricación o una procedencia determinada cuando se especifican los componentes a instalar en un pliego de prescripciones técnicas.
5. Los códigos CPV deben figurar obligatoriamente en todos los pliegos de prescripciones técnicas.
6. El licitante que finalmente llevará a cabo el trabajo recibe el nombre de contratista.
7. Cuando sólo se utilice un criterio de adjudicación, éste ha de ser el del precio más alto.
8. El documento de formalización será firmado por el adjudicatario y se unirá al mismo, como anexo, un ejemplar del pliego de cláusulas administrativas particulares y del pliego de prescripciones técnicas. V

Soluciones Tema 6:

F, F, F, F, F, V, F, V