

PL4	01	Marques Ramos Mier Montoto	Marcel Juan Francisco
Nº PL	Equipo	Apellidos	Nombre

35625337-Q 71777658-V	UO289464@uniovi.es UO283319@uniovi.es
DNI	e-mail

3	Medición y análisis del rendimiento de un servidor	
Nº Práctica	Título	Calificación

Comentarios sobre la corrección

Asignatura de



Área de Arquitectura y Tecnología de Computadores
Departamento de Informática de la Universidad de Oviedo

CONFIGURACIÓN Y EVALUACIÓN DE SISTEMAS
Curso 2022-2023

Índice

1. Introducción
2. Preguntas
3. Distribución de las mediciones

Introducción

Este documento versa sobre el análisis hecho a un servidor, aclarando como se calcularon algunos valores, la realización de predicciones sobre su rendimiento teniendo en cuenta un número de usuarios y el tiempo de respuesta y el análisis de la utilización de recursos.

Preguntas

1. ¿Cómo has calculado el % de uso de memoria?

La memoria total del ordenador en el que se realizaron los análisis es de 16GB, que convertido a bytes da 17179869184. El uso de la memoria se calcula con la siguiente fórmula:

$$\% \text{ de uso} = \frac{\text{Bytes totales} - \text{Bytes libres} - \text{Bytes caché}}{\text{Bytes totales}} \times 100$$

2. Tomando como referencia la evolución de la productividad con el número de usuarios, ¿qué zonas de trabajo puedes diferenciar en el funcionamiento del servidor, y dónde están aproximadamente sus fronteras? ¿Se pueden apreciar claramente las fases de comportamiento lineal, rodilla de productividad y saturación?

Podemos diferenciar 3 zonas: lineal, rodilla y saturación.

La zona lineal empieza con 5 usuarios y termina con los 110 usuarios. A partir de ahí y hasta los 150 usuarios estaríamos en la zona de la rodilla y de ahí en adelante sería la zona de saturación. Sin embargo, al estar basado en datos experimentales, estos puntos pueden variar, aunque se realizan otras mediciones con las mismas condiciones.

3. ¿Qué tiempo de respuesta medio se puede garantizar con el servidor de forma que los recursos del servidor no estén ni infrautilizados ni saturados? Usar como referencia una utilización del 70% del recurso que primero se satura. Compara este valor con el obtenido para el punto de 5 usuarios.

El primer recurso en saturarse es la CPU. Vemos que se alcanza una utilización del 70 % entre los 140 y 300 usuarios. Como no tiene por qué darse una relación lineal entre el porcentaje de uso e la CPU y el número de usuarios, tomaremos como referencia el valor de 140 usuarios, ya que es el más más cercano al 70%. Para 140 usuarios vemos que el tiempo de respuesta es de 0,243386. Se puede observar que a partir de este punto el tiempo de respuesta empieza a incrementarse significativamente, ya que con 125 usuarios el tiempo de respuesta era casi la mitad, 0,15935 seg.

- 4. Si se desea asegurar un tiempo medio de respuesta inferior al doble del tiempo para 5 usuarios, ¿cuántos usuarios simultáneos soportaría el servidor?**

Para este caso estaríamos trabajando con un tiempo de respuesta máximo de 0,0541 seg. Luego tendríamos que limitar el número de usuarios a 20, ya que con 80 ya habríamos superado ligeramente ese tiempo de respuesta y como no podemos suponer que existe una relación lineal entre el número de usuarios y el tiempo de respuesta, aproximamos al valor más cercano que tenemos.

- 5. Si se desea asegurar que el 90% de las peticiones tengan un tiempo de respuesta inferior al doble del tiempo para 5 usuarios, ¿cuántos usuarios simultáneos soporta el servidor? Comenta la diferencia en el número de usuarios soportados, entre usar la media y el 90-percentil como métrica de la calidad del servicio.**

Con 5 usuarios, el 90% de los usuarios tuvo un tiempo de respuesta menor o igual a 0,035 seg. Luego para hallar que número de usuarios cumple ese tiempo de respuesta utilizaremos el punto de una recta que vamos a construir conociendo el 90 percentil y el número de usuarios de 2 puntos de la recta de los 90-percentiles. En este caso tomamos los puntos P1(20,0.0391) y el P2(80,0.118). Con estos puntos y despejando la x, nos queda una ecuación: $x = \frac{60y - 0.768}{0.0789}$. De ahí obtenemos el valor de 43,498 usuarios \approx 43 usuarios.

Luego si se desea asegurar que el 90% de las peticiones sea atendida con un tiempo de respuesta máximo 0.07 seg, podríamos tener como máximo 43 usuarios.

- 6. ¿Cuál es la máxima productividad absoluta que se puede obtener de este servidor y en que punto se alcanza? ¿Cuáles son los valores de tiempos de respuesta y las utilizaciones para ese punto? Compáralos con los valores correspondientes al punto de 5 usuarios, ¿son admisibles? ¿Por qué?**

La productividad máxima absoluta se obtiene con 300 usuarios, teniendo el valor de 118,24 peticiones por segundo.

Para alcanzar esa productividad, la CPU tiene un uso del 76,0754%, el disco 63,2306%, red 0,039% y memoria a 7,198%

Si los comparamos con el punto 5(productividad: 4,9 pet/seg, CPU 3,17%, disco 4,57%, red 0,02%, memoria 6,60%) vemos que, aunque el consumo de recursos aumenta bastante, al igual que la productividad no existe una relación lineal entre ambos, ya que el número de usuarios aumento en 60 veces, mientras que la productividad solo aumento en 24,1 veces.

En general podemos considerar válidos estos valores, ya que vemos que para conseguir valores de productividad el consumo de los recursos aumenta. Si este no fuera el caso podríamos pensar que existen anomalías en las mediciones, ya que, aunque no hay una relación lineal, sí que existe una relación entre productividad y el uso de los recursos.

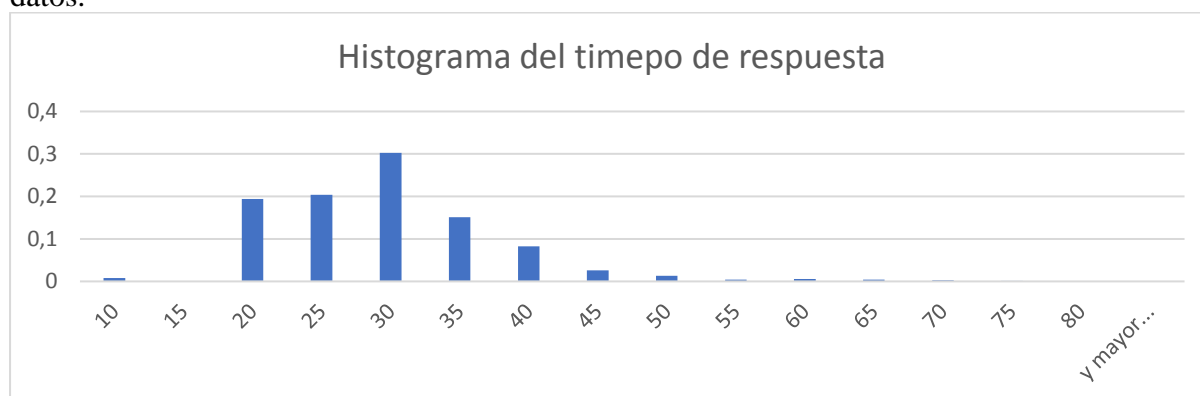
7. ¿Cuál es el recurso que actúa como cuello de botella? ¿Cuál es su valor máximo de utilización? Si el sistema está en zona de saturación, y el valor de la utilización del dispositivo cuello de botella no alcanza niveles iguales o superiores al 90% como predice la teoría, ¿Cuál podría ser la causa?

El recurso que actúa como cuello de botella en este caso es la CPU. Su valor de utilización máximo de media es de 76,0754%.

La causa por la cual el recurso CPU no alcanza valores altos es porque estamos haciendo una media de todos los valores de utilización, y luego trabajamos con ese valor medio. El problema es que la media es muy afectada por valores extremos, y hay ciertos momentos donde la CPU tiene un uso del 91,31% pero también momentos donde el uso es 0%. Por ende, al trabajar con valores medios, por culpa de algunos valores muy bajos, el valor de la media baja.

Distribución de las mediciones

Después de obtener el histograma sobre el punto de 5 usuarios, se puede observar que puede seguir una distribución normal, basándose en la amplitud de la curva y en el tiempo de decadencia, especialmente observable al ensanchar el histograma para poder apreciar mejor los datos.



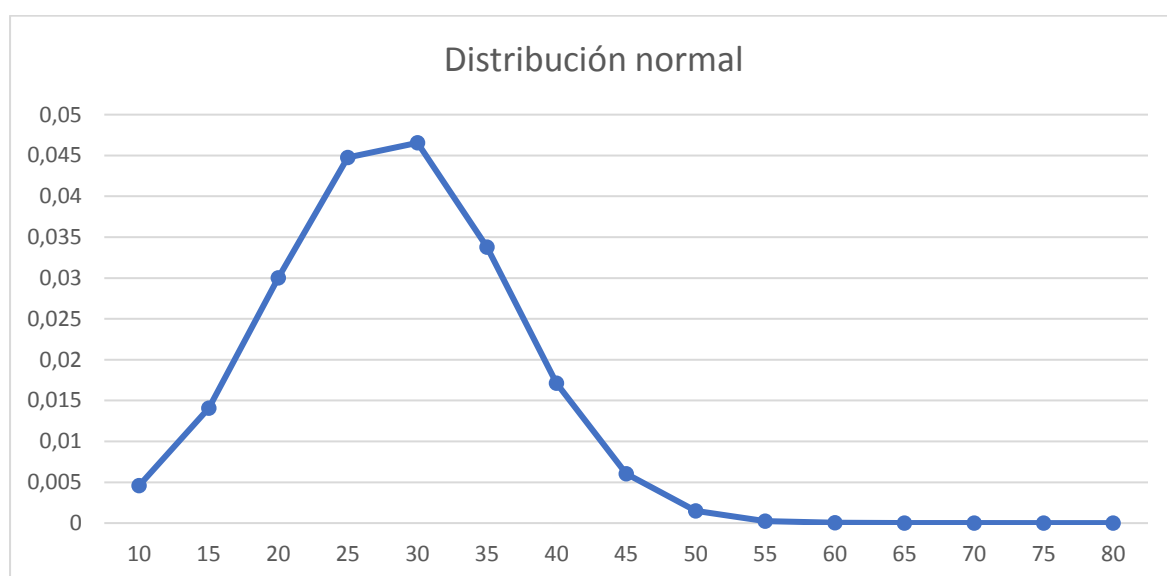
Obviamente, debido al poco número de datos con el que se trabaja, el histograma no representa una distribución claramente normal, pero es lo que más se acerca y lo que más fácil sería de explicar.

Representatividad del ajuste de la distribución escogida

A partir de lo calculado en la hoja de cálculo vemos que sale un resultado algo ambiguo. Como vemos, sí que podemos llegar a percibir una campana de Gauss, sin embargo, la forma de esta es algo irregular y presenta picos.

Esto en parte se puede deber a que se necesitan más mediciones en la zona donde podemos apreciar una campana de Gauss, pero también debido a que la naturaleza de los datos es experimental y por tanto pueden variar bastante.

Con base a esto sería recomendable realizar más mediciones en la zona de la campana para comprobar que la distribución normal representa el tiempo de respuesta para un número pequeño de usuarios (5 usuarios), pero aparentemente vemos que la gráfica sigue el comportamiento de una normal.



Al ajustar un poco la amplitud de la distribución, se puede apreciar la similitud entre los valores normales de la distribución escogida y los valores obtenidos, pese a las irregularidades previamente mencionadas.

