

Informe sobre el desarrollo del Benchmark

Santiago de Diego

1 Planteamiento del problema

Se requiere de la implementación de un Benchmark que permita comprobar las diferencias de velocidad entre distintos sistemas de ficheros. En este informe se detalla el procedimiento seguido, así como los resultados, de comprobar la velocidad de los sistemas de ficheros EXT4, NTFS y FAT32. Para ello se ha desarrollado un programa en C para poner a prueba estos sistemas de ficheros y así establecer un veredicto.

1.1 Intereses del problema

He escogido este tipo de prueba a raíz de una búsqueda que efectué hace varios meses por internet, en la que pretendía documentarme sobre las diferencias de velocidad entre los tres sistemas de ficheros descritos anteriormente. En cuanto al resultado esperado encontré opiniones de todo tipo siendo la más extendida la que defendía que la velocidad de los tres sistemas de ficheros era la misma. En lo que sí que parecían ponerse de acuerdo todos era en que FAT32 era el más lento de los tres.

A raíz de estas opiniones me planteo la hipótesis de si existen diferencias significativas de velocidad entre los diferentes sistemas de ficheros anteriormente expuestos.

1.2 Herramientas utilizadas

Por un lado he desarrollado un programa muy sencillo en C que lo que hace es copiar 4 veces un fichero de 1.2GB y contabiliza el tiempo total que tarda. Esto lo hace tantas rondas como le indiquemos, en mi caso lo he hecho 15 veces. Como resultado, devuelve un archivo en formato csv para poder importarlo a un software de estadística y hacer el posterior análisis de los datos. El programa se llama *FS_Benchmark* y no requiere de ningún argumento para funcionar. Simplemente es necesario que haya al menos 6GB de espacio

en el disco donde queremos hacer la prueba. Podemos ver el diagrama del programa en el anexo de la última página.

Una vez recogidos los datos, he utilizado la herramienta R para hacer un Test Anova a fin de comprobar si la diferencia de medias es significativa.

El equipo utilizado para realizar las pruebas es un Ubuntu Server virtualizado al que he asignado 4000 MB de memoria RAM.

2 Hipótesis

Mi hipótesis inicial es que NTFS y EXT4 arrojarán resultados similares en cuanto a velocidad, mientras que FAT32 resultará ser un poco más lento que los anteriores. Se nos plantea por tanto un experimento con las siguientes hipótesis:

$H_0 \equiv$ Los tres sistemas de archivos presentan la misma velocidad

$H_1 \equiv$ Existen diferencias de velocidad entre los tres sistemas de ficheros

2.1 Conceptos

Es importante explicar como trabaja internamente R con este tipo de test. Como bien es sabido, el test Anova nos permite saber si dos medias pueden considerarse iguales, o en otras palabras, si existe una diferencia significativa entre ellas. Para definir formalmente la significación tenemos el término *p - valor* y α , conocido como el nivel de significación. R nos devolverá un número al final del experimento que hará referencia al p-valor. En nuestro caso, fijada $\alpha = 0.05$ tenemos que si $p\text{-valor} > \alpha$ entonces podemos afirmar con un 95% de confianza que no existen diferencias significativas entre las medias, y por tanto aceptaremos H_0 .

3 Resultados obtenidos

Una vez ejecutado el benchmark en tres particiones distintas del mismo tamaño, en iguales condiciones de consumo de recursos, hemos obtenido una serie de resultados:

Nos han salido tiempos muy similares en los tres sistemas de ficheros evaluados. Por un lado, NTFS y EXT4 tardan tiempos del orden de 68-71 segundos por ronda, mientras que FAT32 oscila entre los 70-72 segundos por ronda,

con excepción de una ronda que ha tardado 200 segundos, pero que probablemente se deba a ruido ambiental, a pesar de que el equipo no se ha tocado durante todo el experimento.

Una vez sacados los resultados de los tres sistemas de ficheros, juntamos los tres archivos .csv obtenidos en uno solo con 45 muestras y ahora vamos a analizar los resultados con R.

4 Interpretación de los resultados obtenidos

Vista la secuencia de resultados obtenidos, vamos a introducirlos en R para realizar el test Anova. Primero de todo vamos a comparar la primera hipótesis, es decir, que realmente no existen diferencias significativas entre los tres sistemas de ficheros.

Tenemos que aceptaremos H_0 si R nos devuelve un p-valor por encima de 0.05

En la siguiente imagen podemos ver el resultado que nos devuelve R al realizar el Test Anova:

```
> summary(aov.datos)
          Df    Sum Sq   Mean Sq F value    Pr(>F)
Tipo        2 2.340e+09  1.170e+09   1.015    0.386
Residuals   15 1.729e+10  1.152e+09
>
```

Como podemos ver, el p-valor que nos devuelve R es de 0.386, mayor que el nivel de significación que hemos fijado, y por tanto, no podemos afirmar que haya diferencias significativas de velocidad entre los tres sistemas de ficheros y por tanto aceptamos H_0 .

5 Conclusión

Al parecer los tres sistemas de ficheros presentan una velocidad similar, resultado distinto al que esperaba encontrar en cuanto al sistema de ficheros FAT32, en el que esperaba encontrar una velocidad inferior a los anteriores. Por otro lado si que he confirmado la teoría de que no existen diferencias significativas en los sistemas de ficheros EXT4 y NTFS, lo cual coincide con la opinión generalizada que circula por la red.

El entorno del experimento ha sido el mismo en los 3 casos y se han realizado bajo las mismas condiciones, por lo que el análisis resulta fiable.

Anexo: Diagrama del benchmark

