

Contenido

Clusters versión 1.0
© David Abilleira Freijeiro 1996

Condiciones de Uso:

El programa que tienes en tus manos lo puedes usar cuanto quieras y pasarlo a quien quieras o a quien creas que le será de utilidad. **Este programa como cualquier otro ha requerido muchas horas de trabajo, por lo que si deseas colaborar a que este tipo de software siga existiendo se agradecerá un ingreso o transferencia de **500pts** (si quieres ingresar más no pasa nada ;)) en la cuenta del Banco Central Hispano nº **0049-2545-2194360652**.**

Qué tienes en las manos

Este es un programa realizado en C++ para calcular el espacio desperdiciado en el disco debido al tamaño de los clusters. Para más información consulta los siguientes apartados

Información

[¿Qué son los clusters?](#)

[¿Para qué sirve este programa?](#)

[Algunas fórmulas](#)

Manejo del programa

[Pantalla principal](#)

[Opciones](#)

¿Qué son los Clusters ?

El Sistema de Archivos

La parte un Sistema Operativo que se encarga de almacenar los ficheros en el disco, se llama *Sistema de Archivos*, en el caso del Ms Dos, el único sistema de archivos que es capaz de utilizar es conocido como **Sistema de Archivos FAT** (realmente FAT, o File Allocation Table, es el nombre de una de las partes del sistema de archivos de Ms Dos). Windows 95 también depende del sistema FAT, ya que es el único sistema de archivos que sabe manejar. Otros sistemas operativos como Windows NT, OS/2 y Linux son capaces de utilizar FAT así como otros Sistemas de Archivos más avanzados.

Los Bloques o Clusters

La mayoría de los sistemas de archivos, dividen el espacio de disco en bloques de tamaño fijo, de forma que, cuando guardamos un fichero de size bytes en el disco, no le asigna size bytes de espacio de disco, sino k bloques de disco tal que $k * \text{tamaño_bloque} \geq \text{size}$.

Esto se hace para facilitar la gestión del espacio libre del disco y para permitir que un fichero pueda guardarse en el disco a pesar de que no haya una zona contigua de disco lo suficientemente grande para almacenarlo. Se le asignarán los bloques que sean necesarios, sin importar donde estén y luego al leer el fichero se irá recorriendo la cadena de bloques asignados saltando de uno a otro.

Trabajando en Ms Dos a los bloques se les llama *Clusters*.

¿Cómo Funciona?

Para entender el mecanismo, con un tamaño de cluster o bloque de 8.192 bytes, un fichero de 37.543 bytes ocuparía:

$$\frac{37543}{4775} = \frac{8192}{4}$$

37.543 / 8.192 = 4 clusters llenos,
pero al quedar un resto de 4775 bytes, tenemos que usar un bloque más aunque no lo llenemos de todo

Vemos que es inevitable que se desperdicie una parte del espacio del último cluster de cada fichero (8192-4775=3417 bytes en el ejemplo anterior), y es obvio que cuanto mayor sea el tamaño de cluster, mayor será la pérdida. En general, sin tener en cuenta el tipo de ficheros que hay en el disco el espacio desperdiciado estimado es:

$$\text{Espacio Perdido} = \text{Nº Ficheros} * (\text{Tamaño Cluster} / 2)$$

es decir, se puede esperar que se pierda la mitad del último cluster de cada fichero. Logicamente el valor real de la pérdida depende del tipo de ficheros que haya en el disco, lo que puedes comprobar utilizando el programa en diferentes particiones o directorios y viendo que el porcentaje de pérdida es diferente (para algo tiene que servir el programa ¿no?).

Aunque esto se ha dicho respecto a sistemas de archivos basados en FAT, la mayoría de los sistemas de archivos también dividen el espacio de disco en bloques y también desperdician parte del espacio del disco.

La Importante Diferencia

En muchos sistemas de archivos, es el usuario quien decide el tamaño del cluster, sin embargo, este no es el caso del sistema FAT en el que el sistema de archivos escoge el tamaño del cluster.

El problema que tiene el antiguo sistema de archivos FAT es que debido a la estructura interna del las zonas especiales del disco, el número máximo de clusters es de 65520. Con discos de hasta 120 o 240 MB esto no es demasiado problema, pero cuando aumentamos el tamaño en MB del disco o de la partición, nos vemos obligados a aumentar el tamaño de cada cluster. Por ejemplo, con clusters de 4KB el tamaño máximo de la partición es de $4 * 65520 = 262080KB$ o casi 256MB.

La tabla siguiente muestra los tamaños de cluster que usa el Ms Dos al formatear particiones dependiendo del tamaño de la partición.

Tamaño de

Tamaño máximo de

cluster	partición
2KB	128MB
4KB	256MB
8KB	512MB
16KB	1024MB
32KB	2048MB

Debemos tener en cuenta de que si, por ejemplo, creamos una partición de 513MB, el sistema operativo se verá obligado a emplear clusters de 16KB en vez de los 8K que usaría con 511MB. El resultado es que perderíamos más espacio y **¡en una partición más grande podríamos guardar menos información!**

Por ejemplo, «scaneando» (perdónenme los puristas de la lengua por no poner rastreando o algo similar, pero me parece claro lo que quiero decir) el directorio de Windows 3.11 vemos que con clusters de 4Kb ocupa 21Mb, con 8KB ocupa 22 MB y con 16KB ocupa 24MB. Esto puede parecer poco pero supone en el caso de clusters de 16KB que con ficheros de este tipo perderás un 17.45% del espacio de tu disco, si tienes un disco o partición de 800MB, perderás **¡139MB!** de disco. Es interesante que tu uses el programa para ver el porcentaje de pérdida con los ficheros que tu tienes en el disco.

Si quieres ver algunas ecuaciones interesantes para ver exactamente cuanto tendría que ser más grande, por ejemplo, una partición con clusters de 16KB que una con clusters de 8KB para guardar lo mismo, mira [Algunas Formulas](#)

¿Para qué sirve este programa?

En la sección ¿Que son los Clusters? se vio una forma de estimar cuanto espacio se perderá debido a que el último bloque de cada fichero no se llena, pero el valor exacto de la pérdida depende de la *mezcla de ficheros* que tengas en la partición.

Este programa lo que hace es recorrer un árbol o subárbol de directorios y ver la diferencia entre la suma de los tamaños de los ficheros, y el número de clusters que ocuparía con diferentes tamaños de cluster. De esta forma podemos saber exactamente cuanto perdemos en nuestro disco duro por el espacio que se desperdicia del último bloque de cada fichero.

¿Es esto práctico?

Pues sí. Cuando una partición pasa de 512MB, el tamaño de cada cluster es de 16KB (a partir de 1Giga, 32KB), con este tamaño de cluster, se puede perder una gran cantidad de espacio de disco duro, especialmente si tenemos muchos ficheros pequeños (cuantos más ficheros, más últimos cluster de fichero hay, y más se pierde).

En estas condiciones, con una mezcla de ficheros de lo más habitual es posible que una partición de 510 MB, puedas meter más información que en una de 600MB ya que la partición de 600MB al usar bloques el doble de grandes, pierde más espacio (normalmente, más de del doble). Incluso puede ser interesante tener una partición de menos de 256 MB, con bloques de 4KB para aquellos programas que tengan muchos ficheros pequeños (por ejemplo: cliparts, ejemplos de un compilador, etc.).

¿Qué se puede hacer?

En mi opinión, a menos que vayas a almacenar, pocos ficheros y muy grandes, no es bueno crear particiones de más de 512MB, es un tamaño aceptable ya que cualquier programa actual (1996) cabe perfectamente en mucho menos, y no nos tenemos que pasar a clusters de 16KB con lo que perderíamos mucho espacio.

Si optas por tener una partición para aquellos ficheros que desperdicien mucho espacio, puedes utilizar una partición pequeña (menos 256MB), o también una unidad comprimida que no sufren esta pérdida al no usar FAT.

Puedes usar este programa para examinar los directorios de los diferentes programas y ver que porcentaje del espacio se pierde, para cambiar aquellos que más pierdan a las unidades más pequeñas (las que menos desperdician).

Podrías por ejemplo, si instalas un compilador de C, instalar los ficheros binarios, librerías, etc. en tu unidad principal (la grande) y los ficheros de cabecera y los ejemplos en una unidad comprimida o con clusters pequeños. Con clusters de 8KB, en el caso de uno de los compiladores de C++ de Borland, los ficheros de cabecera y de ejemplos usan un **52%** más de lo necesario, mientras que el resto del paquete solo usa un **4,19%** más.

Todo esto es una opinión muy personal, pero yo creo que se puede ahorrar bastante espacio en el disco sin por ello hacer las particiones tan pequeñas que nos cueste instalar un programa grande.

Un último consejo, fijate que la partición que crees no este justo en el borde entre un tamaño de cluster y otro, ya que si te pasas aunque sea por poco, el tamaño de cluster se duplicará. (Ver ¿Que son los Clusters?, para la tabla de tamaños de partición)

Algunas ecuaciones

En esta página se desarrollan algunas ecuaciones que te ayudarán a ver la importancia de elegir bien el tamaño de partición.

Tenemos:

Ocu *Espacio Ocupado*

Usa *Espacio Usado*

Per *Espacio Perdido*

Se utilizarán subíndices para indicar el tamaño de la partición

Para empezar es obvio que el espacio ocupado es igual al usado más el perdido:

$$\text{Ocu} = \text{Usa} + \text{Per}$$

Utilizando este programa puedes conocer el porcentaje de pérdida con la mezcla de ficheros que tienes en tu disco. Vamos a emplear el *porcentaje de pérdida con respecto al espacio usado* (la segunda columna empezando por la derecha de la tabla) para ver de qué tamaño tiene que ser una partición con clusters de 16KB para que pueda almacenar la misma información que una partición de 511MB. Llamaremos a los porcentajes de pérdida p8 y p16 para 8KB y 16KB respectivamente (variando de 0-1, no de 0 - 100).

Partiendo de la ecuación anterior, lo que se pierde es igual a lo usado * porcentaje de pérdida

$$\text{Ocu} = \text{Usa} + \text{Usa} * p$$

$$\text{Ocu} = (1 + p) * \text{Usa}$$

Tanto en la partición de 16KB de cluster como en la de clusters de 8KB (511MB) el espacio usado es el mismo ya que queremos guardar la misma información ya que podríamos igualar la ecuación correspondiente a clusters de 8KB y la de 16KB y reescribirla de la siguiente manera:

$$\text{Usa} = \text{Ocu8} / (1 + p8)$$

$$\text{Usa} = \text{Ocu16} / (1 + p16)$$

igualando

$$\frac{\text{Ocu8}}{(1 + p8)} = \frac{\text{Ocu16}}{(1 + p16)} \Rightarrow \frac{\text{Ocu8} * (1 + p16)}{(1 + p8)} = \text{Ocu16}$$

Bueno, pues aquí tenemos que sabiendo el porcentaje de pérdida (para algo debe servir este programa), podemos saber qué tamaño debe tener una partición con clusters de 16KB (Ocu16) para almacenar lo mismo que en 511 (Ocu8)

Este te puede parecer un poco confuso si no has seguido bien las operaciones así que vamos a poner un ejemplo supón:

$$p8 = 0.22 \quad (\text{Un } 22\%)$$

$$p16 = 0.50 \quad (\text{Un } 50\%)$$

$$\text{Ocu16} = \frac{\text{Ocu8} * (1 + p16)}{(1 + p8)} = 511 * \frac{(1 + 0.50)}{(1 + 0.22)} = 511 * 1.23 = 628 \text{ MB}$$

Pantalla principal

Para ver las condiciones de uso mira [Condiciones de uso](#)

Aquí solo se describe el manejo del programa, si no entiendes para que sirve mira [¿Qué son los clusters?](#) y [¿Para qué sirve este programa?](#)

Unidad y directorio

Lo primero que tendrás que hacer es seleccionar la unidad y el directorio que quieres scanear o rastrear, para lo que tienes el combo de selección de unidad y la lista de directorios. Al principio es muy probable que te interese examinar toda la unidad, de forma que basta con que selecciones la unidad que quieres y automáticamente la búsqueda comenzará en la raíz.

Scan

Cuando tengas seleccionada la unidad y el directorio, al pulsar el botón Scan, el programa recorrerá todo el árbol calculando el espacio ocupado, usado y desperdiciado para cada tamaño de cluster.

Opciones

Accede a la pantalla de [Opciones](#)

Ayuda

Su utilidad es bastante obvia si estás leyendo esto

Salir

Abandona el programa

La tabla de resultados

En la tabla se muestra en cada fila un tamaño de cluster (el tamaño real de la partición aparece resaltado), y para cada tamaño se muestra el espacio libre, el ocupado, el usado realmente, el desperdiciado y dos porcentajes:

El porcentaje de lo perdido con respecto a lo usado. Si aquí aparece por ejemplo un 50%, quiere decir que si queremos guardar 100MB nos va a ocupar $100 + 50 = 150$ MB.

El porcentaje de lo perdido con respecto a lo ocupado. Si aquí aparece por ejemplo un 50%, quiere decir que si hemos gastado 100MB de disco, 50MB se han desperdiciado y solo el resto ha sido realmente usado.

Como puedes ver estos dos porcentajes son dos caras de la misma moneda. Si quieres saber cuanto espacio vas a perder al meter N megas, usa el %usa, si quieres saber cuanto espacio perderás en tu partición de T megas usa %ocu.

Los valores reales de pérdida, son lo que están en la fila marcada, ya que indica el tamaño de cluster de la unidad. Los otros valores nos sirven de referencia para saber cuanto se perdería si estos ficheros los guardásemos en una partición de diferente tamaño y por lo tanto un cluster también de diferente tamaño.

Como ves el espacio usado siempre es el mismo, (es la suma de los tamaños de los archivos), sin embargo el espacio ocupado crece a medida que usamos clusters más grandes.

También puedes ver que $\text{Ocupado} = \text{Usado} + \text{Perdido}$.

En cuanto al espacio libre, no es muy relevante, ya que el tamaño de cluster viene dado por el tamaño de la partición, de forma que para, por ejemplo, reducir el tamaño de cluster debemos reducir el tamaño de la partición con lo que quedará menos espacio libre.

La gráfica

A medida que se rastrea el disco sumando los tamaños de los ficheros, y calculando el número de clusters que usan o usarían, se va dibujando una gráfica, en la que se muestra para cada tamaño de cluster, las proporciones entre el espacio usado, el perdido y el libre.

Modificando las [Opciones](#) hay dos tipos de gráficas:

- Usado, perdido y libre. La altura de la gráfica coincide con el tamaño del disco y vemos abajo el espacio usado por los ficheros, inmediatamente encima el espacio perdido, y arriba de todo el espacio

que queda libre. Esto nos permite ver las proporciones de estos valores con respecto al total del espacio en disco. Es fácil ver que el espacio ocupado del disco es la suma de las dos primeras barras.

.- Usado y perdido. En este caso no es muestra el espacio libre, sino que solo se muestra la proporción entre el espacio usado y el desperdiciado.

Opciones

En esta pantalla podrás configurar tres cosas:

Incluir Espacio

Aquí se especifica si al recorrer el árbol de directorios se deberá tener en cuenta también archivos ocultos, del sistema y directorios. Para que los cambios tengan efecto habrá que volver a pulsar Scan

Ver Como

Aquí podrás especificar si quieres que los resultados se presenten en bytes, Kbytes o MegaBytes.

Gráfica

En esta parte se indica que datos debe reflejar la gráfica habiendo dos opciones:

.- Usado, perdido y libre. La altura de la gráfica coincide con el tamaño del disco y vemos abajo el espacio usado por los ficheros, inmediatamente encima el espacio perdido, y arriba de todo el espacio que queda libre. Esto nos permite ver las proporciones de estos valores con respecto al total del espacio en disco. Es fácil ver que el espacio ocupado del disco es la suma de las dos primeras barras.

.- Usado y perdido. En este caso no se muestra el espacio libre, sino que solo se muestra la proporción entre el espacio usado y el desperdiciado.

Normalmente cuando se examina todo el disco es interesante ver también el espacio libre, pero cuando solo examinamos un subdirectorio, es preferible ver solo lo usado y lo perdido ya que, si el subdirectorio ocupa una parte pequeña del disco, casi toda la gráfica representaría el espacio libre.

