

## Compresión de datos

### Idea general

El propósito de los programas compresores es muy sencillo: comprimir archivos para que ocupen menos espacio y descomprimirlos cuando haya que volver a usarlos. Para aprovechar mejor el espacio, estos programas pueden unir varios archivos en un solo archivo comprimido. Por tanto, es necesario que los programas compresores ofrezcan opciones para añadir, actualizar y eliminar archivos a un archivo comprimido, y para consultar qué archivos se encuentran en un archivo comprimido.

Las tecnologías que permiten realizar la compresión de datos son muy variadas, y normalmente se utilizan aquéllas más apropiadas al tipo de datos que haya que manejar, aunque existen compresores de uso general. Estas tecnologías no sólo se usan en programas compresores, sino integradas en muchos otros programas y en algunos instrumentos de uso común.



### Tipos de compresión

Existen dos tipos de compresión, basadas en principios diferentes y con utilizaciones completamente distintas: compresión sin pérdidas y compresión con pérdidas.

#### Compresión sin pérdidas

Es aquella en la que al comprimir un archivo y descomprimirlo se obtiene exactamente el archivo original.

El modo de actuación del compresor consiste en buscar repeticiones de símbolos en el archivo original y escribirlos de otra forma, que sea más corta; el programa descompresor reconoce la nueva manera de escribir los datos y puede ir regenerando los originales. Exponemos un método extremadamente simple y que no se usa en la vida real: imaginemos un archivo compuesto de letras que se repiten; para comprimirlo, escribimos cada letra una vez y luego el número de veces que se repite (si una letra sólo aparece una vez, no escribimos el número). En el siguiente cuadro se muestra un ejemplo:



Archivo original (32 símbolos)	Archivo comprimido (16 símbolos)
bbbbCCcGHHHHHHaZZZQQQrjbbbbbb	b4C3gH7aZ3Q5rjb6

La compresión sin pérdidas se utiliza, entre otros casos, en los siguientes:

- ◆ **Copias de seguridad.** Al ocupar menos los ficheros, se utilizan mejor los dispositivos de almacenamiento: en el mismo espacio disponible se almacenará más información y se tardará menos en grabarla.
- ◆ **Transferencia de información.** El tiempo de conexión entre ordenadores se reduce considerablemente si la información se envía comprimida.
- ◆ **Almacenamiento de imágenes artificiales.** Las imágenes que se crean artificialmente admiten muy bien esta compresión; existen gran cantidad de formatos gráficos que la incluyen, como por ejemplo, TIFF, GIF y PNG. Los faxes también se incluyen en esta categoría.

#### Compresión con pérdidas

En este tipo de compresión no se obtiene el mismo archivo que el original al comprimir y descomprimir un archivo, parte de la información original se pierde. Puede parecer un problema, pero se supone que la información que se pierde no es imprescindible para comprender el mensaje del original. Se utiliza principalmente en estos campos:

- ◆ **Almacenamiento de imágenes naturales.** Las imágenes obtenidas del natural mediante escáneres y cámaras fotográficas digitales suelen ser muy grandes y con gran cantidad de colores; sin embargo, el ojo no tiene tanta capacidad de discernimiento, por lo que es posible eliminar parte de la información y que aún así el resultado sea aceptable. Por ejemplo, el formato JPEG.



- ◆ **Almacenamiento de vídeo.** Los vídeos digitales ocupan muchísimo, de modo que siempre se manejan comprimidos. El formato más conocido es el MPEG, utilizado (entre otros lugares) en los DVD de vídeo.
- ◆ **Almacenamiento de sonido natural.** La mayor calidad de sonido se obtiene en el CD audio, que no tiene ningún tipo de compresión, pero ocupa mucho espacio. Algunas frecuencias de sonido se pueden eliminar y pocos oídos lo notan. Utilizan esta compresión el formato mp3, los mini-disc y el iPod.



## Formatos

Existen muchísimos formatos de archivos comprimidos sin pérdidas de uso general, como por ejemplo ZIP, LHA, ZOO, ARJ, CAB, RAR y GZ. Probablemente el formato ZIP sea el más popular; existen muchísimos programas capaces de manejarlo, así como muchos programas que crean y descomprimen ficheros ZIP, para casi todos los sistemas operativos conocidos; esto hace que el formato ZIP sea muy bueno para llevar datos de un sistema operativo a otro: se comprimen los datos en el origen, se traslada un único archivo ZIP y se descomprime en el destino.



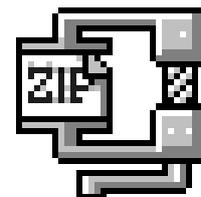
## InfoZIP

Esta entidad ofrece programas compresores y descompresores de formato ZIP para gran cantidad de sistemas operativos, así como código fuente para usar este formato en programas que lo deseen, todo ello de modo gratuito. Se accede a la entidad en <http://www.cdrom.com/pub/infzip/>. Existen otras empresas que ofrecen programas de este tipo, pero no de modo gratuito.



### Uso de “zip” y “unzip”

La entidad InfoZIP ofrece estos dos programas (para MS-DOS, Windows, GNU/Linux y muchos otros sistemas operativos), que son de fácil manejo y libre distribución. Los ficheros que crea son totalmente estándar y se pueden manejar con cualquier otro programa que maneje ficheros ZIP. Estos programas se invocan desde el indicador del sistema y ofrecen una pantalla con instrucciones de uso, que pueden ser ampliadas consultando los manuales que acompañan al producto.



Los dos programas ofrecen muchas opciones, pero las que más se usan son unas pocas, de modo que es muy sencillo usarlos. Baste señalar unos ejemplos:

- ◆ Para crear el archivo **MiPagina.zip** y comprimir en él los archivos **Imagen.png**, **Texto.html** y **Foto.jpg** basta teclear **zip MiPagina Imagen.png Texto.html Foto.jpg**.
- ◆ Para ver los ficheros contenidos es **MiPagina.zip** se usa **unzip -l MiPagina**.
- ◆ Para descomprimir **MiPagina.zip** se tecllea **unzip MiPagina**.