FUJIFILM

FLASH
TECHNICAL SUPPORT
MAGNETIC MEDIA, OPTICAL DISCS AND OTHER DATA STORAGE PRODUCTS

Tape Drive Data Compression Q & A

Question – What is data compression and how does compression work?

Answer – Data compression permits increased storage capacities by using a mathematical algorithm that reduces redundant strings of data. There are many different compression schemes that can be performed by software or hardware. For tape drives, the compression algorithm is typically implemented in hardware and will remove redundancy from data by encoding ***patterns*** of input characters more efficiently. The more repeat patterns in the data, the more the data can be compressed. Thus the amount of data compression obtained will vary depending upon the characteristics of the data.

Question – Can I assume that I will get the advertised 2:1 compressed capacity?

Answer – **No**; you may get more or less. A more definitive statement that could be used for advertising capacity, which "assumes 2:1 data compression," would be "Assuming ***Your Data*** Compresses 2:1."

Question – Why wouldn't my tape drive get the capacity indicated at the advertised compression ratio?

Answer – Your data is unique – not like anyone else's. Therefore when your data is compressed the ratio will also be unique. The amount of compressed data stored on your tape will be different; the capacity is unpredictable. Furthermore, data with random bit strings, encrypted or pre-compressed data are unlikely to show any capacity improvement from the tape drive compression. Capacity can even decrease slightly when highly random data is compressed (see page 5 for more on this subject).

Question – Do all tape drives feature hardware based data compression?

Answer – No; check your drive's manual for feature information relating to the *exact* model-number of your drive or contact the drive manufacturer.

Question – Some tape drive vendors claim **>** 2:1 data compression. Why?

Answer – The effects of data compression vary, because some types of data are more "compressible" than others. Different mathematical compression algorithms are better at compressing certain data types than others.

Some vendors may conclude that their compression scheme is more powerful; however, it may not be more powerful for your mix of data. When comparing technologies, it is best to compare native-to-native capacity figures. Compression claims may or may not be relevant to your application.

Question – My backup software features data compression and my tape drive also has a compression feature. Should I use software compression and/or the tape drive's hardware based compression?

Answer – When compression is available in the tape drive hardware/firmware, the compression feature of any software package should be turned off. This reduces the processing burden on your computer. Also, hardware based compression is typically much more effective. Do not double compress; compressed files could even be bigger after compression. This can even cause the tapes to hold less-than their native capacity.

Some tape drives prevent lowering capacity due to files that grow larger from compression by writing the uncompressed file when this happens. The drive tags each block (compressed/not compressed) for appropriate handling during playback – e.g., LTO Ultrium drives have this feature. See page 5 for more information.

Question – Does data compression affect the tape drive's data transfer rate?

Answer – Sustained data transfer rates will increase at the same ratio as data capacity when employing hardware based data compression (up to the speed of the compression hardware). The same may not be true for compression implemented by software running on your computer.

Software-based compression typically degrades performance, since the software or separate utility has to compress the data before saving it to tape. Retrieval time is also longer, since the software must decompress. Today's hardware-based compression chips are typically many times faster than the tape drives they work with. Therefore compression does not introduce any overhead into the writing or reading processes. Since compression reduces the number of bits written to or read from the tape, significant performance gains are realized.

Question – Does the tape media need to store more data bytes with compression?

Answer – No; the tape may appear to hold twice as much data, but the capacity (density) of bits/bytes written on the tape media will ***not*** be increased.

Question – Can data compressed on my drive be read on other [same type] drives?

Answer – Yes, if you used the drive's hardware based compression process ***and*** the other drive has the compression feature. All drives within a family of drives will employ the same data compression hardware. If another *compression capable* drive can read uncompressed data from your device, it will be able to decompress your compressed data too.

Of course if you are using a software program to compress your data, the same software program must be available to the other drive.

Question – What is decompression?

Answer – An inverse procedure of the compression software or the compression hardware that returns compressed data to its original size and condition.

Question – What is loss-free compression?

Answer – Loss-free compression schemes allow full recovery of the original data. Only loss-free (lossless) algorithms are used for computer data, insuring that decompressed data output is exactly the same as the uncompressed data input.

Question – What is lossy compression?

Answer – Lossy compression is ***not*** applicable to tape drives used by computers. Lossy compression allows some of the original data to be lost. Lossy compression can be used where the data is meant to be interpreted by humans (not computers). Employing lossy compression schemes can provide a much better compression ratio. Lossy compression can be used for data such as graphic images, video, pictures, photographic images and music.

Question – How do I compute my Compression Ratio?

Answer – Compression ratio is the ratio of a file's original size divided by its compressed size. A compression ratio of 2:1 means the compressed file is ½ as large as the original file.

Data Storage Tape Drive Hardware – Examples
Example Data Compression Features & Issues:

Important Considerations When Troubleshooting DLTtape™ Capacity Issues:

Troubleshooting DLTtape Capacity Issues

Question – Why do some of my DLTtape IV data cartridges hold less than 80 GB when used on my DLT8000 drive [or < 70 GB on a DLT7000, etc.]?

Answer – There are **three possible reasons**:

> **First**, those values assume 2:1 compression of your data – and your data (or some of your data) may **not be compressible** at the typical 2:1 ratio. See above questions and answers for more information on compression.

> **Second**, if the data cartridge was used previously on a DLT4000 drive, and your software isn't set up to change the 20/40 GB format on a <u>write from beginning of tape</u> (BOT), this will cause the DLT8000 drive to write in its 20 GB native (40 GB compressed) backward compatible mode. Not only will this cause capacity to be reduced to the DLT4000 value, but also the drive will write at the slower transfer-rate of the DLT4000, which is only 1.5 MB/second.

> Likewise for prior use on a DLT7000 drive, if your software isn't set up to change the existing DLT7000 format, this will cause the DLT8000 drive to write data in its DLT7000 backward compatible mode (lower capacity and lower speed).

> Refer to Fujifilm's Technical Support Document, "DLTtape Data Cartridge Initial Calibration" for more information concerning density format compatibility between the different DLT drive models.

> **Third**, native capacity (without compression turned on) is 20 GB, 35 GB or 40 GB for DLTtape IV on DLT4000, DLT7000 and DLT8000 drives, respectively. If you are writing without the drive's compression turned on, this will limit capacity to the native value.

> Perhaps you (or the software application) did not specify for the drive to use compression or maybe compression was automatically turned off – because the host computer system could not meet the drive's transfer-rate requirements. Consult with your drive hardware manufacturer or your application-software manufacturer concerning these possibilities.

DATA COMPRESSION 07-5

Question – I get less than the advertised Native capacity on my backup tapes; how could this be possible?

Answer – There are several common possibilities:

1) Compression could be expanding your data:

Data that is already highly compressed or random data such as encrypted data or seismic data (i.e. geoseismic data recorded during exploration for oil) can actually expand if you write with the tape drive's compression turned on. However, some tape drive families have a pass-through feature for incompressible data.

Some tape drive technologies (tape drive families), such as 34XX (mainframe tape) and LTO Ultrium, limit expansion of data by automatically switching compression off, if the data cannot be compressed. Data compression automatically switches back on, when data becomes compressible again. Both compressed and not-compressed data can be recorded on a tape; data will be marked accordingly for proper treatment during playback. Such intelligence prevents the expansion of previously compressed or otherwise incompressible data.

<u>Perhaps</u> your data is incompressible and you are trying to use the compression feature of a tape drive that does not have this intelligent compression feature. This can cause a 5 - 10 percent reduction in capacity capability.

2) You could have compression off *and* have the application software set to limit the data capacity to less than the full capability:

Many software applications will have a feature that allows a user to limit the capacity of their data cartridge tapes to less than full capacity. This feature is often employed when the user wants to make one-to-one copies of the recorded tapes. All tapes have an allowable +/– variation in tape media length. A tape recorded to its full capacity using a slightly longer tape, may not copy onto another slightly shorter tape.

By limiting capacity for all original recordings to something slightly less than the full capacity capability, copies will always fit on one tape when duplicating.

<u>Perhaps</u> you (or someone before you) set this software feature to limit the capacity of your data cartridges to some percentage of full capacity.

3) If the Read/Write head on your tape drive is dirty or worn to near its end-of-life, high error rewrite activity (skipping small sections of tape) can reduce capacity.

For More Data Storage Tape Technical Support Documents and Product Information, go to: <u>www.fujifilmusa.com/tapestorage</u> → Resource Center → Technical Center.
For questions or comments, go to "Ask the Expert" in the Technical Center.

Page 5 of 5
DATA COMPRESSION 07-5