

# NIELIT GORAKHPUR

**Course Name:** O Level (2nd Sem)  
**Topic:** Data Compression

**Subject:** Introduction to ICT Resources  
**Date:** 21-04-2020

## Data Compression

### Data Compression

The process of reducing data by applying a compression technique is called compression, and the resulting data is called compressed data. The reverse process of reproducing the original data from compressed data is called decompression, while the resulting data is called decompressed data.

The disk compression method works by compressing the data stored on the disk drive at the time of writing the data and when this data is retrieved, the data is again expanded to its original form. This compression is possible because most of the files stored on the disk drive contain a lot of repeating information. The data is somewhat like a sponge with a lot of holes in it. When the sponge is squeezed, the holes disappear and the sponge size shrinks. Later, when you release the sponge, it returns to its normal shape and size by filling the holes.

In the same manner, a file that contains a lot of repeating data is stored on the disk after replacing the repeated data with some type of token value. Later, when the file is retrieved from the disk, this token value is replaced with the original value.

Compression is the process of reducing the size of a file by encoding its data information more efficiently. By doing this, the result is a reduction in the number of bits and bytes used to store the information. In effect, a small file size is generated in order to achieve faster transmission of electronic files and a small space required for its downloading.

### Concept of Compression

When you have a file containing text, there can be repetitive single words, combination of words and phrases that use up storage space unproductively. Or there can be media such as high-tech graphical images in it whose data information occupies too much space. To reduce this inefficiency electronically, you can compress the document. Compression is done by using compression algorithms (formulae) that rearrange and reorganize data information so that it can be stored more economically. By encoding information, data can be stored using fewer bits. This is done by using a **compression/ decompression** program that alters the structure of the data temporarily for transporting, reformatting, archiving, saving, etc. Compression reduces information by using different and more efficient ways of representing the information. Methods may include simply removing space characters, using an angle character to identify a string of repeated characters, or substituting smaller bit sequences for recurring characters. Some compression algorithms delete information altogether to achieve a smaller file size. Depending on the algorithm used, files can be adequately or greatly reduced from their original size.

Compression is the reduction in size of data in order to save space or transmission time. For data transmission, compression can be performed on just the data content or on the entire transmission unit (including header data) depending on a number of factors.

## Type of Compression

There are two type of compression.

- Lossless Compression
- Lossy Comprassion

## Lossless Compression

Lossless compression can reduce files without loss of information, in the process. The original file can be recreated exactly when uncompressed. This is important for the data and program files. This type of compression method provides an average 2:1 compression ratio. When a 10 KB file is compressed, you get a tile of around 5 KB of size. To achieve this, algorithms create reference points (substitution characters) for things such as textual patterns, store them in a catalogue send them along with the smaller encoded file. When uncompressed, the file is "re-generated" by using those documented reference points to re-substitute the original information.

Lossless compression is ideal for documents con- taming text and numerical data, where any loss of textual information cannot be tolerated. ZIP compression, for instance, is a lossless compression that detects patterns replaces them with a single character.

## Lossy Compression

Lossy compression (some information is permanently lost), on the (other hand, reduces the file size be eliminating bits of information. It permanently deletes any unnecessary data. This compression is usually used with images, audio and graphics, or instance, in an image containing a green landscape with a blue sky; all the different and slight shades of blue and green are eliminated with compression. The essential nature of the data is not lost - the essential colours are still there. One popular example of lossy compression is JPEG compression (Joint Photographic Experts Group), which is suitable for grayscale or colour images.

Lossy compression technique usually has a higher compression ratio (larger reduction in the volume of data), but may distort the original quality of data when decompressed. As this type of files takes a lot of space in original shape, they are often stored after compressing. This type of compression provides a high level of compression as much as 25:1 or more compression ratio. This means that a 1000 KB file can be stored in just 40 KB or less storage space.

If the decompressed data is the same as the original data, it is referred to as lossless compression; otherwise the compression is lossy because there is loss or originality in the reproduced data.

## Exercise:

- 1: What is data compression?**
- 2: What is compress and decompress program in compression?**
- 3: How many type of compression technique?**
- 4: Write the differences between Lossy and Lossless compression.**