Unit II Data Compression

1.1 Introduction :-
- Data Compression plays a very important role in Media Industry.
- Specially when the things comes in terms of sending, receiving and storing of data.
- There are various compression algorithms used to compress the Data, Video and audio etc.

Definition of Data Compression:-

" Data Compression is the process of coding that will effectively reduce the total number of bits needed to represent certain information."

- It is the process of reducing the amount of data required to represent a given quality of information.

- Data Compression is a technique used to reduce the size of data by removing number of bits. This technique uses various algorithm to do so. These compression algorithms are implemented according to type of data you want to compress.

1.2 Types of Data Compressions :-

There are mainly two types of Data Compression, these are –
1) **Lossless Compression**
2) **Lossy Compression**

1) **Lossless Compression :-**
- As its name indicate that this is loss-less compression. Because there is no data loss when the file is recovered or comes to its original state.
- Even a single bit is not deducted after restoration.
- We will get exact replica of the original file.
- Lossless Compression is used in Image, Text File Compression where smaller amount of data loss is not tolerated. Because smaller loss in bit changes its information which is not required.
- It is also used for Executable Files.

- This type of compression is done to get low space consumption file.
- For Ex – Runlength Encoding Compression , Huffman Compression.

2) Lossy Compression :-

- As its name indicates that lossy compression is done, where small losses in data is tolerable.
- Lossy Compression is used Video and Audio specially. Because small amount of data loss is tolerable in audio and video which do not affect the viewers.
- By doing so we can save lot of space in storage device and bandwidth consumption over internet.
- Sometimes this type of compression is also used in images where quality do not matters that much.
- This type of compression algorithm removes permanently repetitive content, Which saves lot of spacein video file.
- For Ex – JPEG Compression , MPEG Compression

Advantage of Compression :-

- Saves Disc Space
- Saves bandwidth consumption over internet while transmission.
- Improves file transfer speed

Disadvantages :-

- Needs high CPU Power while compressing and decompressing file.
- Needs high RAM power while compressing and decompressing file.
- Error effects in transmission.

Storage Space :-

- Uncompressed graphics, audio, and video data required substantial storage capacity and high bandwidth , which is not possible in the case of uncompressed video data, even given todays CD & DVD technology.
- The same is true for multimedia communication. That's why to maintain the storage space data will be compressed by compression techniques.

1.3 Coding Requirements :-

- To compare data storage and bandwidth requirements of different media such as, Text and Graphics.
- Following are the specification is based on typical window 640 X 480 pixels on the screen –

1. Text :-

- To representation of text, two by the are used for each character. - Each character is displayed by using 8 x 8 pixel.
- Which is sufficient on display the ASCII characters.
- ASCII stands for American Standard Code for Information Interchange - ASCII characters set includes the value of 128 characters.
- ASCII characters set are used to display. A to Z uppercase letter, a to z lowercase letters, Punctuation makes and mathematical symbols.

2. Graphics :-

- To representation of graphics that is vector graphics includes a typical still image is composed of 500 lines.
- Each line defines Horizontal Position, Vertical Position and 8 bit attributes field.
- The Horizontal axes are represented by using 10 bits and Vertical axes are coded with 9 bits.

3. Audio:-

- Following are the examples that specify the continuous media and drive, the amount of storage is required 1 second of playback.
- An uncompressed audio **signal telephone quality** is sampled at rate of 8 KHz and is quantized with 8 bits per sample.
- This need to bandwidth requirements is 64 kilobits to store 1 second of playback.
- An uncompressed sterol and audio is used with **CD quality**. These sampled rate of 44.1 KHz and contized with 16 bits per   sample.
- Hence, the storage requirements is, 44.1 KHz x 16 bits=705.6 bits x 103 bits to store 1 second of playback.

4. Video :-

- According to the European PAL (Phase Alternate Line) standard.
- A video is defined by 625 lines at the frame rate of 25 frames per second.
- To determine the storage requirements of PAL standard, we assume the image resolution as used before 640 x 480 pixels and 3 bytes coded with luminance and chrominance components.
- Hence, the storage requirement of 1 image [frame] is 640 x 480 pixels x 3 bytes that is 92, 16,000 bytes or 73, 72,800 bits To store 25 frames per second.

1.4 Basic Compression Techniques :-

- Following compression techniuques are used to compress the data -
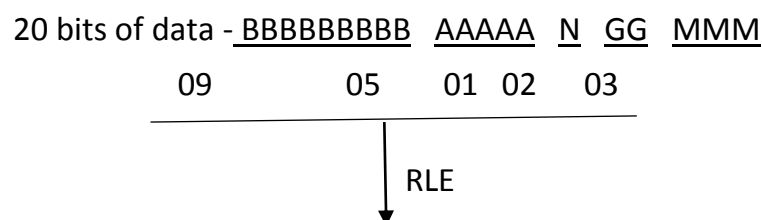
1. Runlength Encoding :-

- Run length Encoding is simplest and earliest of the data compression schema that is used to data compression.
- In this runs of data are stored as a single data value and count rather than original run.(Run means Sequence of same symbol/data values.)
- Run length encoding is applied on simple graphics, icons, and line drawing.
- Run length encoding is used to compresses the black and white images.

For Ex – 1)

BBBBBBBB   B8

Long sequence of 'B'   Data value   Count of the data value
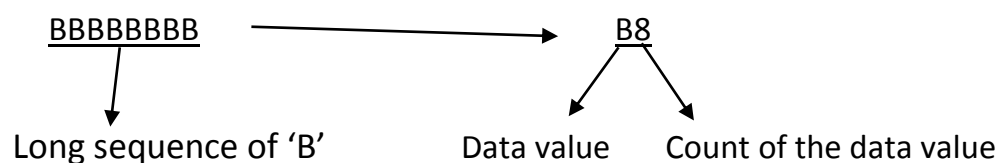
In above example it will reduce the data 8 bits into 2 bits.

For Ex – 2)

20 bits of data - BBBBBBBBB AAAAA N GG MMM
                  09   05   01 02   03

                               RLE

15 bits of data - <u>B09A05N01G02M03</u>

RLE

10 bits of data - <u>B9A5N1G2M3</u>

2. **Huffman Encoding:-**

- Huffman Algorithm was developed by David Huffman in 1951.
- This is a technique which is used in a data compression or it is a coding technique which is used for encoding the data.
- This technique is basically dependent upon the frequency i.e the frequency of the corresponding characters which needs to be compressed and by that frequency, only Huffman code will be generated.
- In case of Huffman coding the most generated character will get the small code and least generated character will get large code.
- Huffman tree is a specific method of representing each synmbol.
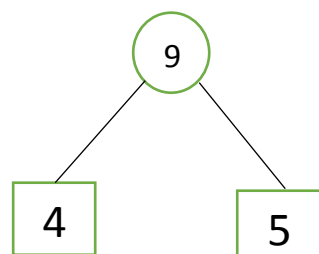- The modified version of Runlength encoding is known as Huffman Encoding.

For Ex – Let obtain a set of Huffman code for the message (m1……m7) with relative frequencies (q1…..q7) = (4,5,7,8,10,12,20).
Let us draw the Huffman tree for the given set of codes.

Step 1) Arrange the data in ascending order in a table.

4, 5, 7, 8, 10, 12, 20

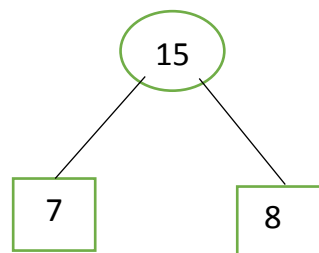**Step 2)** Combine first two entries of a table and create a parent node.

Step 3)

A) Remove the entries 4 and 5 from the table and insert 9 at its appropriate position.
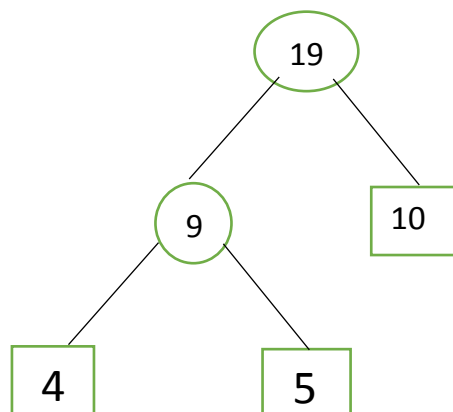
7 , 8 , 9 , 10 , 12 , 20

Combine minimum values of a table and create a parent node.

B) Remove the entries 7 and 8 from the table and insert 15 at its appropriate position.

9 , 10 , 12 , 15 , 20

Combine minimum values of a table and create a parent node.

C) Remove the entries 9 and 10 from the table and insert 19 at its appropriate position.

12 , 15 , 19 , 20

Combine minimum values of a table and create a parent node.