UNIT –IV
DATA FORMATS

1	**a**	Summarize various types of common data that can be represented in a computer.	[L2][CO5]	**[6M]**

Answer:

EXPLANATION-3M
DATA STANDARDS-3M

- All computers and computer-based devices the binary number system is the system of choice, both for all forms of data storage and for all internal processing of operations.
- As human beings, we normally don't choose to do our work in binary form. Our communications are made up of language, images, and sounds.
- Today, multimedia, consisting of images and sounds in the form of video conferencing, PowerPoint presentations, VoIP telephony, Web advertising, YouTube, smartphone-based news clips and photos on TV, and more is of at least equal importance.
- Since data within the computer is limited to binary numbers, it is almost always necessary to convert our words, numbers, images, and sounds into a different form in order to store and process them in the computer.
- Hence standardization is an important consideration in a modern world where the ability to share data is assumed. There are many different standards in use for different types of data. A few of the common ones are

Some Common Data Representations

Type of data	Standard(s)
Alphanumeric	Unicode, ASCII, EBCDIC
Image (bitmap)	GIF (graphical image format), TIFF (tagged image file format), PNG (portable network graphics), JPEG,
Image (object)	PostScript, SWF (Adobe Flash), SVG
Outline graphics and fonts	PostScript, TrueType
Sound	WAV, AVI, MP3 or 4, MIDI, WMA, AAC
Page description and markup	pdf (Adobe Portable Document Format), HTML, XML
Video	Quicktime, MPEG-2 or -4, H.264, WMV, DivX, WebM

	b	Briefly explain the three standards that are used in common for alphanumeric characters.	[L2][CO5]	**[6M]**

Answer:

EXPLANATION FOR EACH STANDARD -2M X 3-6M

Much of the data that will be used in a computer are originally provided in human-readable form, specifically in the form of letters of an alphabet, symbols representing a word, syllable, or sound element, numbers, and punctuation, whether English or some other language.

The data entered as characters, symbols, number digits, and punctuation are known as **alphanumeric data**. Since alphanumeric data must be stored and processed within the computer in binary form, three alphanumeric codes are in common use.

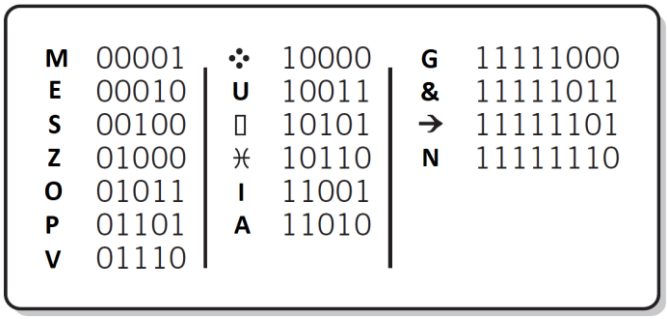
1. **EBCDIC** (Extended Binary Coded Decimal Interchange Code):
   - EBCDIC was developed by IBM. Its use is restricted mostly to IBM and IBM-compatible mainframe computers and terminals.
   - The EBCDIC code is somewhat less standardized; the punctuation symbols have changed over the years.
   - The codes for each symbol are given in hexadecimal, with the most significant digit across the top and the least significant digit down the side.

2. **ASCII** (which stands for American Standard Code for Information Interchange):
   - The ASCII code was originally developed as a standard by the American National Standards Institute **(ANSI)**.
   - ANSI also has defined 8-bit extensions to the original ASCII codes that provide various symbols, line shapes, and accented foreign letters for the additional 128 entries.
   - Together, the 8-bit code is known as Latin-1. Latin-1 is an ISO (International Standards Organization) standard.

3. **Unicode**:
   - Unicode supports approximately a million characters, using a combination of 8-bit, 16-bit, and 32-bit words.
   - The ASCII Latin-1 code set is a subset of Unicode, occupying the values 0–255 in the Unicode table,
   - Unicode divides its character encodings into sixteen 16-bit code pages, called planes.
   - Unicode defines three encoding methods, UTF-8, UTF-16, and UTF-32.
   - It defines codes for the characters of nearly every character-based alphabet of the world in modern use

2 **a** A secret message is transmitted from the other planet to earth in the form of binary and each binary code has a unique character. By using the table given below, analyze the given message and determine the secret code. [L4][CO5] **[8M]**

Binary code:
1100111010000011111100000010011011111111011111100000000100100



M	00001	✣	10000	G	11111000
E	00010	U	10011	&	11111011
S	00100	▯	10101	→	11111101
Z	01000	⚹	10110	N	11111110
O	01011	I	11001		
P	01101	A	11010		
V	01110				

Answer:

DECODING THE MESSAGE-8M

As the message is encoded in binary form, we need to decode the binary message from **right to left**.

	By decoding the above binary number into characters as follows: 11001 11010 00001 11111000 00010 01101 11111110 11111000 00001 00100 **11001 - I** **11010 - A** **00001 - M** **11111000 - G** **00010 - E** **01101 - P** **11111110 - N** **11111000 - G** **00001 - M** **00100 - S** Finally, the message is "**IAMGEPNGMS**"			
	b	Write a short note on PNG & JPEG image formats.	[L2][CO5]	**[4M]**
	Answer: **EXPLANATION OF PNG-2M** **EXPLANATION OF JPEG-2M** **PNG (Portable Network Graphics) format** is the best-known losslessly compressed alternative to GIF. PNG can store up to 48 bits of color per pixel, and additionally can store a transparency percentage value and a correction factor for the color in a monitor or printer. Its compression algorithm is often more efficient than that used with GIF. Unlike GIF, PNG stores only a single image in a file. **JPEG (Joint Photographers Expert Group) format**, employs a lossy compression algorithm to reduce the amount of data stored and transmitted, but the algorithm used reduces the image resolution under certain circumstances, particularly for sharp edges and lines. This makes JPEG more suitable for the representation of highly detailed photographs and paintings, but GIF and PNG are preferable for line drawings and simple images.			
3	**a**	Write the characteristics of a bitmap image.	[L2][CO5]	**[4M]**
	Answer: **EXPLANATION-4M** Bitmap images (also known as **raster images**) are **made up of pixels** in a grid; each pixel containing a color value. These **pixels** (short for **picture elements**) are tiny, individual squares of color that are arranged in a grid to form an image. Bitmap images are characterized by two parameters: the **number of pixels (resolution)** and the **color depth per pixel**. • **Color depth** refers to the information contained within the image. For example, a 1-bit image means that a pixel could either be black or white. An 8-bit image means that each pixel can consist of any one out of 256 color values (2 to the power of 8). "True color" images use 24-bit RGB, in which there are 8 bits allocated to each red, green and blue component.			

- **The number of pixels** is a measure of how many pixels there are in a given area, or pixel density. This is measured in pixels per inch (PPI) or dots per inch (DPI). The resolution of an image affects the quality of the bitmap being displayed or printed. The higher the resolution, the less apparent the pixel nature of the bitmap will be.

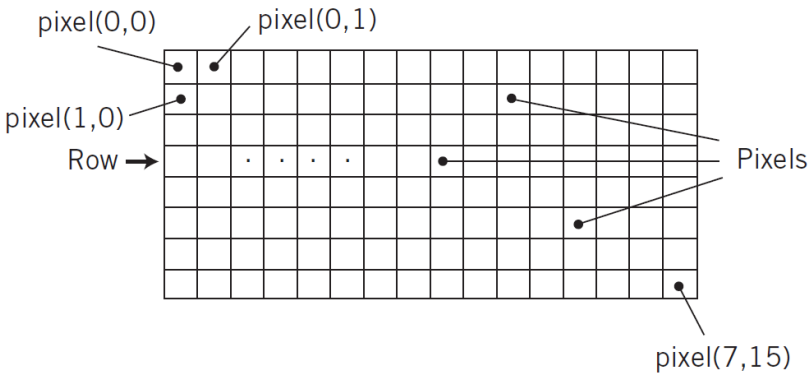
b	Why images must be stored and manipulated as bitmap images? Justify your answer.	[L5][CO5]	**[8M]**

Answer:

EXPLANATION-6M
DIAGRAM-2M

- Most images—photographs, graphical images, and the like—are described most easily using a bitmap image format.
- The basic principle for representing an image as a digital bitmap is simple. A rectangular image is divided into rows and columns, as shown in Figure.
- The junction of each row and column is a point (actually a small area) in the image known as a **pixel**, for *pi[x]*cture *el*ement.

A 16 × 8 Bitmap Image Format



Pixels stored in the order p(0,0), p(0,1),... p(0,15), p(1,0),... p(1,15),... p(7,15)

- The image represented within the computer is really only an approximation to the original image, since the original image presents a continual range of intensity, and color.
- The faithfulness of the computer representation depends on the size of the pixels and the number of levels representing each pixel.
- Reducing the size of each pixel improves the possible **resolution**, or detail level, of the representation by increasing the number of pixels per inch used to represent a given area of the image.
- It also reduces the "stepping" effects seen on diagonal lines.
- Increasing the range of values available to describe each pixel increases the number of different gray levels or colors available, which improves the overall accuracy of the colors or gray tones in the image.
- Bitmap representations are particularly useful when there is a great amount of detail within an image, and for which the processing requirements are fairly simple.
- If you want to apply effects such as lighten, darken, sharpen, blur, crop, color correct, colorize, paint, and other operations along those lines, bitmap software will work best.

4	**a**	Define image metadata. Give at least three examples of metadata that would be required for a bitmap image.	[L2][CO5]	**[5M]**

Answer:

DEFINITION OF METADATA – 2M
EXAMPLES – 3M

Image Metadata

- Image/Photo metadata is a set of data describing and providing information about rights and administration of an image. It allows information to be transported with an image file, in a way that can be understood by other software and human users.

The pixels of image files are created by automated capture from cameras or scanners. Metadata is stored in two main places:

- Internally – embedded in the image file, in formats such as JPEG, DNG, PNG, TIFF …
- Externally – outside the image file in a digital asset management system (DAM) or in a "sidecar" file (such as for XMP data) or an external news exchange format document as specified by the IPTC.

There are 3 main categories of data:

- **Descriptive** – information about the visual content. This may include headline, caption, keywords. Further persons, locations, companies, artwork or products shown in the image .
- **Rights** – identification of the creator, copyright information, credits and underlying rights in the visual content including model and property rights. Further rights usage terms and other data for licensing the use of the image.
- **Administrative** – creation date and location, instructions for the users, job identifiers, and other details.

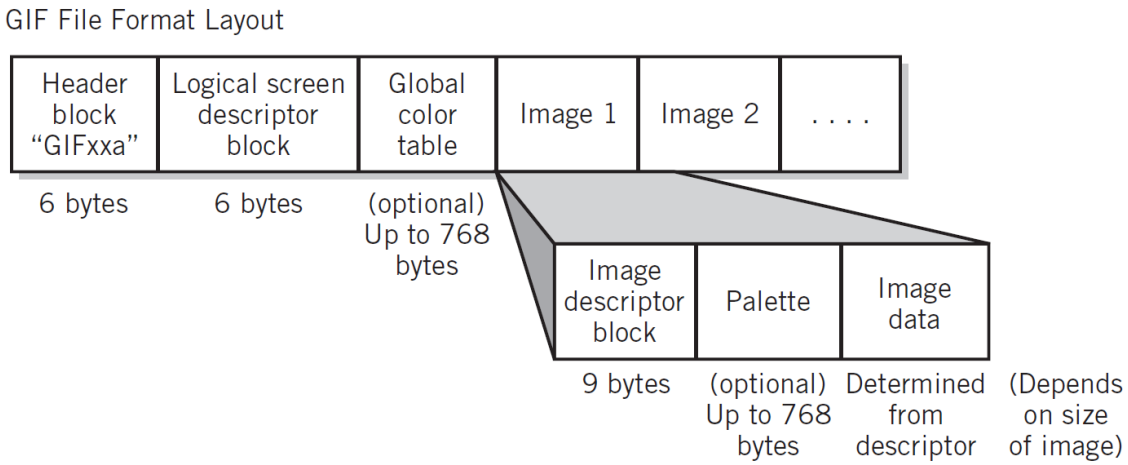
	b	With a neat sketch, explain the bitmap image storing format GIF.	[L2][CO5]	**[7M]**

Answer:

DIAGRAM – 3M
EXPLANATION – 4M

Graphics Interchange Format

- GIF was first developed by CompuServe in 1987 as a proprietary format that would allow users of the online service to store and exchange bitmap images on a variety of different computing platforms.



- The later version, GIF89a, also allows a series of GIF images to be displayed sequentially at fixed time intervals to create "animated GIF images". The GIF format is used extensively on the Web.
- Figure illustrates the layout of the screen and its images. The format divides the picture information and data into a number of blocks, each of which describes different aspects of the image.
- The first block, called the header block, identifies the file as a GIF file and specifies the version of GIF that is being used.
- Following the header block is a logical screen–descriptor block, which identifies the width and height of the screen, describes an optional color table for the images on the screen (the palette), indicates the number of bits per color available, identifies the background screen color, and specifies the pixel aspect ratio.

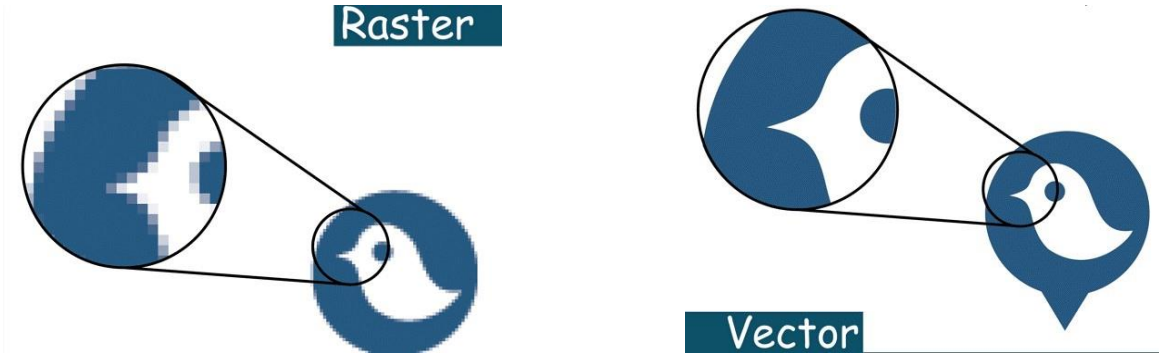
5	With an example, explain about the object image.	[L2][CO5]	**[12M]**

Answer:

DIAGRAM – 4M
EXPLANATION – 8M

Object Image:

- Object images/Vector images/Vector graphics are computer images created using a sequence of commands or mathematical statements that place lines and shapes in a two-dimensional or three-dimensional space.
- In vector graphics, a graphic artist's work, or file, is created and saved as a sequence of vector statements. A vector graphic file describes a series of points to be connected. These files are sometimes called *geometric files*.

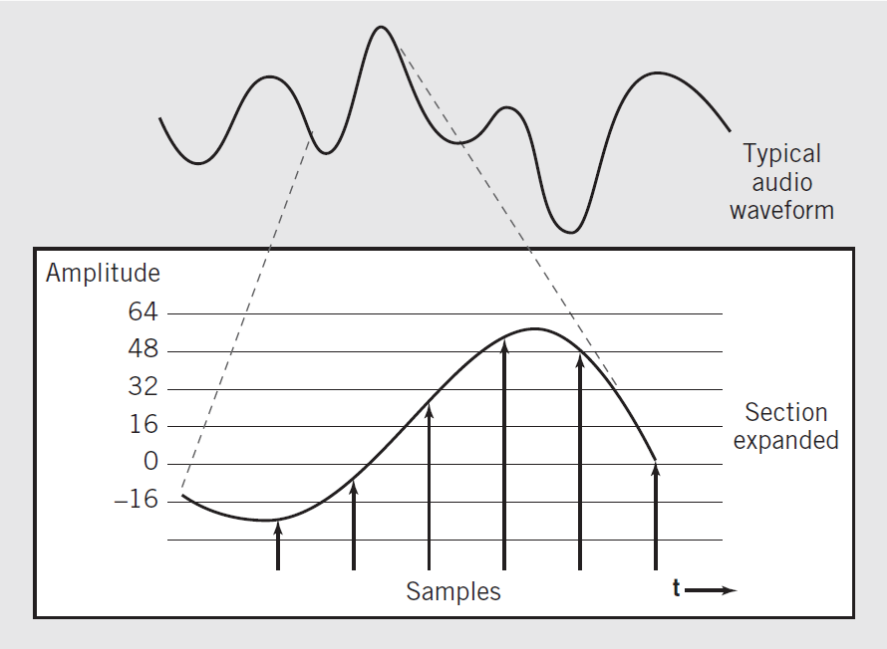


Graphic artists, illustrators and designers use vector graphics for a variety of reasons, including the following:

- **Scalability.** Vector formats are good for projects that require scalable graphics, including scalable type and text. For example, A logo created with vector graphics can be scaled up or down without loss of quality or creating a large file.
- **App and web development.** Vector graphics are useful in application and web development because web apps and the graphics they contain must work with various screen sizes and device types.
- **Animation.** Animated images are also usually created as vector files, which provide for cleaner and smoother images.

	• **Computer-aided design (CAD).** CAD programs frequently use vector files for manufacturing, engineering and design because of their scalability and ease when it comes to editing the mathematical formulas. Advantages • **Scalability** - vector graphics are derived from mathematical vector relationships, or relationships between points that create lines and curves, they appear clean and exact at any size. • **Small file size** - Vector graphics generally have a small file size because they only store a small number of points and the mathematical relationships between them. Those relationships are expressed in code, which is less memory-intensive compared to storing pixels. • **Easy to edit** - This is useful in an iterative process, like graphic design, that requires a lot of editing. • **Easy to load** - Because file sizes are smaller, it is easy to port and load vector files to different devices and programs. • **Easy to duplicate -** It is also easy to create clones of a vector image and copy certain features of one graphic to another. • **Precision** - The ability to scale vector graphics up or down means they have a precise look and feel.		
6	With a neat sketch, describe how an A-to-D converter converts audio data into binary data.	[L2][CO5]	**[12M]**
	Answer: **DIAGRAM – 4M** **EXPLANATION – 8M** • Sound is used as an instructional tool, as an element of multimedia presentations, for computer-based telephony—voice over IP (VoIP) tools, Skype, and the like, to signal events within the computer, and to enhance the enjoyment of games. • Since the original sound wave is analog in nature, it is necessary to convert it to digital form for use in the computer. • The analog waveform is sampled electronically at regular time intervals. • Each time a sample is taken, the amplitude of the sample is measured by an electronic circuit that converts the analog value to a binary equivalent. • The circuit that performs this function is known as an **A-to-D converter**. • The largest possible sample, which represents the positive peak of the loudest possible sound, is set to the maximum positive binary number being used, and the most negative peak is set to the largest negative number. • Binary 0 falls in the middle. The amplitude scale is divided uniformly between the two limits. • The sampling rate is chosen to be high enough to capture every nuance in the signal being converted. • For audio signals, the sampling rate is normally around 50 kilohertz, or fifty thousand times a second.		

Digitizing an Audio Waveform



- In this diagram shown above, the signal is allowed to fall between −64 and 64.
- The consecutive values for the signal in this diagram will be the binary equivalents to −22, −7, +26, 52, 49, and 2.

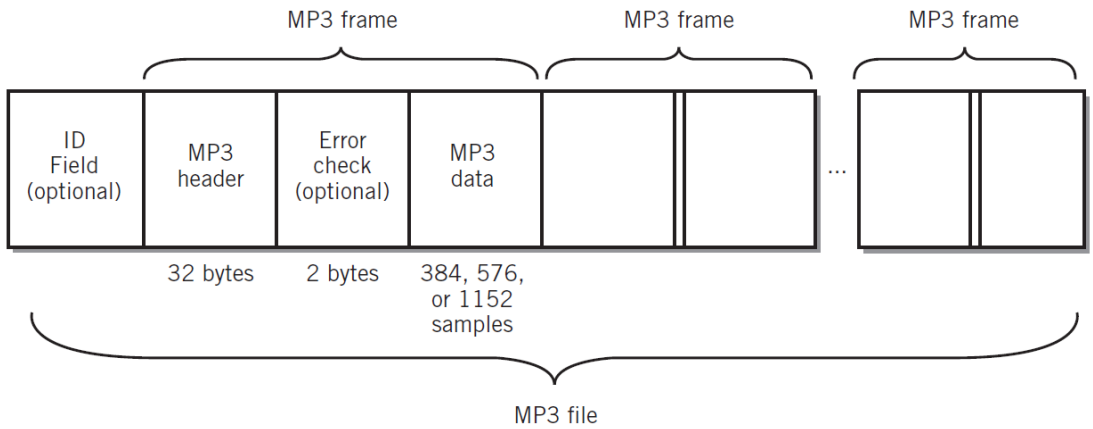
7	Describe the most important characteristics and features of the following audio file formats: (i) .MP3 (ii) .WAV	[L2][CO5]	**[12M]**

Answer:

MP3 FORMAT DIAGRAM – 3M
MP3 FORMAT EXPLANATION – 3M
WAV FORMAT DIAGRAM – 3M
WAV FORMAT EXPLANATION – 3M

(i) MP3:

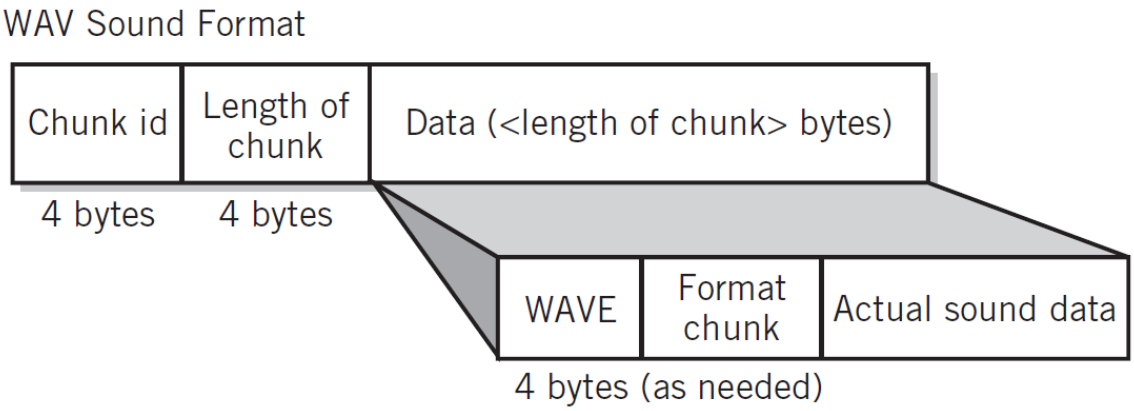
MP3 Audio Data Format



- **MP3** is the predominant digital audio data format for the storage and transmission of music.
- It is characterized by reasonable audio quality and small file size.

- MP3 uses a number of different tactics and options to achieve its small file sizes. Because of psychoacoustic lossy compression. The size of an MP3 file is typically about 1/10th the size of an equivalent uncompressed .WAV file.
- Figure shows the structure of MP3 file. The file consists of an optional ID field that contains such information as song title and artist, followed by multiple data frames.
- Each frame has a 32-byte header that describes the frame data, followed by an optional 2-byte error-checking code, followed by the data itself.
- The header contains 2 bytes of synchronization and MP3 version data followed by the bit rate, the audio sample rate, the type of data (for example, stereo or monaural audio), copy protection, and other information.
- The MP3 standard requires that each frame contains 384, 576, or 1152 audio samples of data.

(ii) WAV:



- The .WAV format was designed by Microsoft as part of its multimedia specification.
- The format supports 8- or 16-bit sound samples, sampled at 11.025 kHz, 22.05 kHz, or 44.1 kHz in mono or stereo..WAV data is not compressed.
- The format consists of a general header that identifies a “chunk” of data and specifies the
- length of a data block within the chunk.
- The general header is used for a number of different multimedia data types.
- The data block is itself broken into three parts. First, a 4-byte header identifies a sound file with the ASCII word “WAVE”.
- A format chunk follows. This chunk contains such information as the method used to digitize the sound, the sampling rate in samples per second, the data transfer rate in average number of bytes per second, the number of bits per sample, and whether the sound is recorded in mono or stereo.The actual data follows.

8	**a**	Write the advantages of data compression.	[L3][CO5]	**[2M]**

Answer:

FOUR ADVANTAGES – 2M

Advantages of Data Compression:

- Less disk space (more data in reality) (*)
- Faster writing and reading (*)
- Faster file transfer
- Byte order independent

b	Distinguish lossless and lossy data compressions algorithms.	[L4][CO5]	**[10M]**

Answer:

TEN DIFEERENCES – 10M

Parameters	**Lossy Data Compression**	**Lossless Data Compression**
Definition	The lossy data compression technique removes a specified amount and quality of data from the intended original file (data loss).	The lossless data compression leads to a reduction in file size while still maintaining the original amount and quality of the data that it carries.
Non-noticeable Data Elimination	This technique leads to data loss but eliminates the non-noticeable form of data.	This technique does not lead to the elimination of non-noticeable data forms.
Restoration of files	This technique cannot restore the original amount and quality of the data contained in the original file.	This technique can easily restore the original quality and amount of data in an available file after decompression.
Data Quality	It compromises the original quality of data.	It does not lead to the compromise of the original data quality.
Data Size	The original size of data reduces after lossy data compression.	The original size of data stays intact after lossless data compression.
Algorithms	This technique uses algorithms like Fractal Compression, DWT (Discrete Wavelet Transform), DCT (Discrete Cosine Transform), Transform Coding, etc.	This technique uses algorithms like Arithmetic Coding, Huffman Coding, LZW (Lempel Ziv – Welch), RLE (Run Length Encoding), etc.
Uses	Mainly MP3 audio, MPEG video, and JPEG image formats make use of the technique of lossy data technique.	Mainly confidential information and sensitive documents make use of lossless data compression. File formats like BMP, GIF, RAW, ONG, etc., also use this technique.
Data Holding Capacity	This technique can hold more data because it does not need to restore them back to their original form.	This technique can hold less data because it restores the original quality and amount of data without any hassle.
Commonly Used Term	We can also call it irreversible compression.	We can also call it reversible compression.
Compression Rate	Compression rate is high	Compression rate is low

9	**a**	Define page description language and enumerate various page description languages.	[L2][CO5]	**[4M]**

Answer:

DEFINITION – 1M
EXAMPLES – 3M

- A **page description language** is a language that describes the layout of objects on a displayed or printed page.
- Page description languages incorporate various types of objects in various data formats, including, usually, text, object images, and bitmap images.
- **HTML** (HyperText Markup Language), provides, except for text, most objects are stored in separate files, the details of layout are left mostly to the Web browser that is recreating the page, and programming language capability and other features are provided as extensions.
- **PDF** (Portable Document Format) and **PostScript** offer the ability to recreate sophisticated pages with surprising faithfulness to the intentions of the original page designer.

	b	List the five simple data types that are provided in most high-level programming languages and write a short note on each datatype.	[L1][CO5]	**[8M]**

Answer:

EXPLANATION – 8M

The raw binary numbers stored in a computer can easily be interpreted to represent data of a variety of different types and formats. C, Java, Visual Basic, and other languages all provide a programmer with the capability to identify binary data with a particular data type. Typically, there are five different simple data types:

- *Boolean*: two-valued variables or constants with values of true or false.
- *char*: the character data type. Each variable or constant holds a single alphanumeric character code representing, for example, the single strike of a key.
- *String*: It is also common to process groups of characters together as strings. Strings are simply arrays of individual characters.
- *enumerated* data types: user-defined simple data types, in which each possible value is listed in the definition, for example,

type DayOfWeek = Mon, Tues, Wed, Thurs, Fri, Sat

- *integer*: positive or negative whole numbers. The string of characters representing a number is converted internally by a conversion routine built into the program by the compiler and stored and manipulated as a numerical value.
- *real* or *float*: numbers with a decimal portion, or numbers whose magnitude, either small or large, exceeds the capability of the computer to process and store as an integer. Again, the routine to convert a string of characters into a real number is built into the program.

10	Write a pseudocode procedure that performs string conversion to number.	[L3][CO4]	**[12M]**

Answer:

PSEUDOCODE – 12M

A Pseudocode Procedure that Performs String Conversion

```
//variables used
char key;
int number = 0;
boolean error, stop;

void main()
{
stop = false;
error = false;
ReadAKey;
while (NOT stop && NOT error) {
number = 10 * number + (ASCIIVALUE(key) – 48);
ReadAKey;
} //end while

if (error == true) {
printout('Illegal Character in Input');
else printout('input number is ' number);
} //end if

} //end procedure

function ReadAKey(); {
read(key);
if (ASCIIVALUE(key) == 13 or ASCIIVALUE(key) == 32 or ASCIIVALUE(key) == 44)
stop = true;
else if ((key < '0' ) or (key > '9' )) error = true;
} //end function ReadAKey
```