**Topic: Data compression**

Reading Time: 20 mins

**·        Note\* Highlight important/core points while reading**

·        Read the content and write the answers given in the document in your words, to get the solid grip on topic.

**Data Compression**

Data compression is the process of **reducing the size of a file** while maintaining as much of the original data as possible. It is widely used in **images, audio, video, and text files** to save storage space and improve transfer speeds.

**Why is Data Compression Needed?**

Compression is important for the following reasons:

1. **Saving Storage Space**
   * Reduces the amount of storage needed on devices such as **hard disk drives (HDDs) and solid-state drives (SSDs)**.
2. **Faster Streaming**
   * Compressed audio and video files require **less data to be loaded**, making streaming services faster.
3. **Faster File Transfer**
   * Smaller files take **less time** to upload, download, or send over a **network**.
4. **Bandwidth Efficiency**
   * The amount of **data transferred per second** (bandwidth) is **lower** for compressed files, improving performance in **limited network conditions**.
5. **Reduced Costs**
   * Many **cloud storage providers** charge based on storage usage.
   * **Internet service providers (ISPs)** may charge based on data consumption.

**Types of Data Compression**

There are **two main types** of data compression:

1. **Lossy Compression**
2. **Lossless Compression**

**1. Lossy Compression**

* **Removes some data permanently** to reduce file size.
* Used in **multimedia files** (images, audio, video).
* **Example formats:**
  + **JPEG** (for images)
  + **MP3** (for audio)
  + **MP4** (for video)
* **Advantages:**
  + Very high compression rates.
  + Significantly reduces file size.
* **Disadvantages:**
  + **Loss of quality** (data is removed and cannot be restored).

**Example of Lossy Compression**

* A **10 MB JPEG** image may be reduced to **1 MB**, but fine details might be lost.

**2. Lossless Compression**

* **No data is lost**; the file can be restored to its original state.
* Used for **text and software files** where data accuracy is crucial.
* **Example formats:**
  + **ZIP** (compressed folders)
  + **PNG** (for images)
  + **FLAC** (for audio)
* **Advantages:**
  + No loss of data.
  + Can restore the original file exactly.
* **Disadvantages:**
  + Compression rates are lower than lossy methods.

**Example of Lossless Compression**

* A **5 MB PNG image** may be reduced to **3 MB** but can be **fully restored**.

**How Does Compression Work?**

Compression techniques find **patterns in data** and store them **more efficiently**.

**1. Run-Length Encoding (RLE)**

* Used in **simple images and text files**.
* **Identifies repeating patterns** and replaces them with a **shorter representation**.
* **Example:**

AAAAABBBCCCCCC → 5A3B6C

* Instead of storing every letter, it stores the letter + count.

2. Huffman Coding

* Used in lossless compression.
* Replaces common characters with shorter binary codes.
* Example in text compression:

"hello" → 101 111 110 110 100

·

* Frequently used letters get **shorter codes**.

**3. Perceptual Encoding (Used in MP3 & JPEG)**

* Removes data that **humans are less likely to notice**.
* Example:
  + **MP3** removes **inaudible sound frequencies**.
  + **JPEG** removes **small color variations**.

**Examples of Compression in Real Life**

1. **Streaming Services (Netflix, YouTube, Spotify)**
   * Use **lossy compression (MP4, MP3)** to stream smoothly over the internet.
2. **ZIP Folders**
   * Used to **send multiple files** in a smaller size.
3. **Cloud Storage (Google Drive, Dropbox)**
   * Compressed files use **less space**, reducing storage costs.
4. **Web Pages (HTML, CSS, JavaScript Compression)**
   * Web developers **compress code** to **make websites load faster**.

**A-Rated Questions/Answers By Examiner**

**Q1: Why is data compression used?**

**Answer:** Data compression is used to:

* Save **storage space** on devices.
* Reduce **file transfer time** over networks.
* Decrease **bandwidth usage**.
* Lower **costs** for cloud storage and internet usage.

**Q2: What is the difference between lossy and lossless compression?**

**Answer:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Compression Type** | **Description** | **Example File Types** | **Data Loss?** |
| **Lossy** | Removes **unnecessary data** | JPEG, MP3, MP4 | **Yes** |
| **Lossless** | Reduces file size **without data loss** | PNG, ZIP, FLAC | **No** |

**Q3: Explain Run-Length Encoding (RLE) with an example.**

**Answer:**

* RLE replaces **repeating patterns** with a **shorter code**.
* Example:

WWWWWBBBBW → 5W4B1W

* Instead of storing **each letter**, it stores the **letter + count**.

**Q4: A 3-minute MP3 song has a file size of 3 MB. If stored as an uncompressed WAV file, the size is 30 MB. Explain why.**

**Answer:**

* **MP3 uses lossy compression** to remove **unnecessary audio data**, reducing file size.
* **WAV is uncompressed**, meaning it stores all sound data **without loss**, making it much larger.

**Q5: A 10 MB ZIP file contains text documents. If extracted, the total file size is 50 MB. Explain why.**

**Answer:**

* The **ZIP file uses lossless compression**, removing **redundant data**.
* When extracted, **all original data is restored**, increasing file size.

### Write your Answers on your Notebook and Verify it on Next Screen

**Q6: Why do video streaming services use lossy compression instead of lossless compression?**

**Q7: A 5-minute WAV file is 50 MB, while the same file as MP3 is 5 MB. What causes the size difference?**

**Q8: How does Huffman coding reduce file size?**

**Q9: A website loads faster after compressing HTML, CSS, and JavaScript files. Why?**

**Q10: What happens when you repeatedly compress a JPEG image?**

**6. Answer:**

* **Lossy compression significantly reduces file size**, allowing faster streaming.
* **Reduces bandwidth usage**, making it accessible on slower internet connections.
* **Most quality loss is unnoticeable** to human perception due to perceptual encoding.
* **Lossless compression would require too much storage and bandwidth**, making it impractical for large-scale streaming.

**Answer:** **Video streaming services use lossy compression (MP4) to reduce file size while maintaining good quality, ensuring smooth playback with less bandwidth.**

**7. Answer:**

* **WAV files are uncompressed**, storing all original sound data.
* **MP3 uses lossy compression**, removing inaudible sound frequencies and redundant data.
* **MP3 reduces bitrate**, lowering file size while keeping essential audio quality.

**Answer:** **MP3 files are smaller because they use lossy compression, removing unnecessary sound data, unlike uncompressed WAV files.**

**8. Answer:**

* **Assigns shorter binary codes to frequently used characters** and longer codes to less frequent ones.
* **Replaces original data with optimized binary codes**, reducing the overall file size.
* **Commonly used in text compression (ZIP) and image compression (JPEG).**

**Example:**  
If "A" appears 50 times, it gets a **shorter binary code** than "Z," which appears once.

**Answer:** **Huffman coding reduces file size by assigning shorter codes to frequently used characters, optimizing data storage.**

**9. Answer:**

* **Compressed files require less bandwidth**, making downloads quicker.
* **Redundant spaces, line breaks, and repeated code are removed** (minification).
* **Browsers can process and display web pages faster** with smaller files.

**Answer:** **Web compression reduces file size, leading to faster loading times and improved website performance.**

**10. Answer:**

* **Each compression removes more data**, leading to noticeable quality loss.
* **Image artifacts (blurring, pixelation) appear** due to discarded details.
* **JPEG is lossy, meaning lost data cannot be recovered.**

**Example:**

* **First save:** Minimal quality loss.
* **Multiple saves:** Image becomes blurry and distorted.

**Answer:** **Repeated JPEG compression degrades image quality because lossy compression removes more data each time.**