Lecture 02: Data representation in computer systems

System

A system is a set of integrated devices that input, output, process, and store data and information.

Information System

An information system (IS) is a set of interrelated components that collect, process, store, and distribute data and information to support decision making and control in an organization.

1. **Data & Information**

   a. **Data**

   - Data is a collection of facts or observations.
   - It is raw and unorganized, and it does not have any meaning on its own.

   Eg: A list of names, A set of numbers, Collection of images

   b. **Information**

   - Information is data that has been processed and organized in a way that gives it meaning. It is the result of analyzing and interpreting data.

     Eg: The average temperature in a city over a period of time. It is derived from the data of individual temperature readings, The number of people who visited a website, The most popular product in a store.

Data and information are essential for decision-making.

However, it is important to note that not all data and information is created equal. Some data is more accurate and reliable than other data.

It is important to carefully evaluate the quality of the data before using it to make decisions.

Tips for evaluating the quality of data

- Look for data that is from a reliable source.
- Check for accuracy and consistency in the data.
- Make sure the data is relevant to the decision you are making.

Characteristics of Information

- **Accuracy**: Information must be accurate, meaning that it is free from errors and bias.
- **Completeness**: Information must be complete, meaning that it provides all of the necessary details for a user to make a decision.
- **Relevance**: Information must be relevant to the user's needs, meaning that it is useful and meaningful to them.
- **Timeliness**: Information must be timely, meaning that it is delivered when the user needs it.
- **Accessibility**: Information must be accessible to the user, meaning that they can find it and understand it.

2. Data Types

Qualitative & Quantitative

- Qualitative data is descriptive and cannot be measured. It is often expressed in words or phrases.

- Quantitative data is numerical and can be measured. It is often expressed in numbers or statistics.

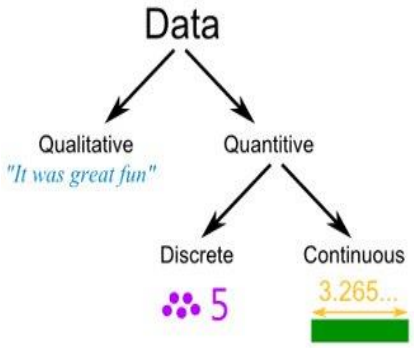


Figure 1

Here is a table that summarizes the key differences between qualitative and quantitative data:

Table 1

Feature	Qualitative data	Quantitative data
Nature	Descriptive	Numerical
Expression	Words or phrases	Numbers or statistics
Examples	Colours of cars, names of people, opinions of customers	Number of cars sold, average age of customers, amount of money spent by customers

Another classification of data types is shown below:

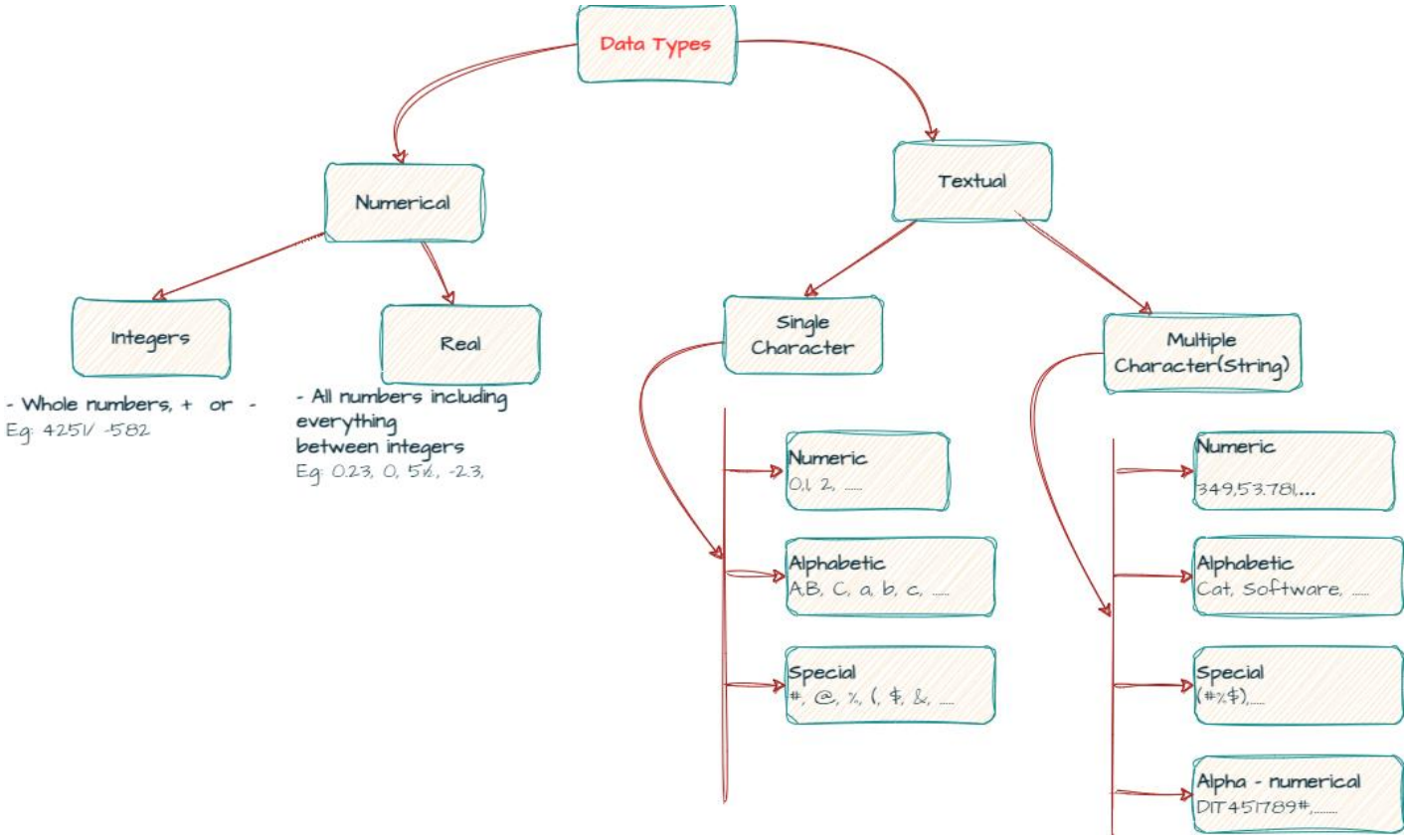


Figure 2

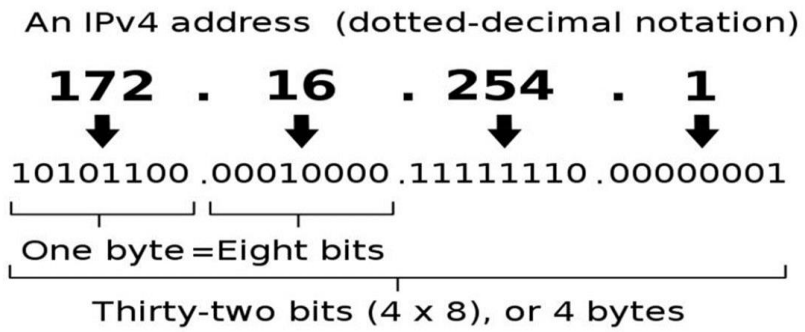
3. Data Representation

- Data representation is the process of converting data into a form that can be stored, processed, and transmitted.
- Data representation is an important part of data processing.
- By representing data in a way that is efficient and easy to understand, data can be processed more effectively and used to solve problems.
- Digitization is the process of converting information, such as text, numbers, photo, or music, into digital data that can be manipulated by electronic devices.

How do computers represent data?

- Computers represent data as binary digits, or bits.
- A bit is a single piece of information that can be either a 0 or a 1.
- Computers are electronic devices powered by electricity, which have only two states, on or off.

Example: IP addressing

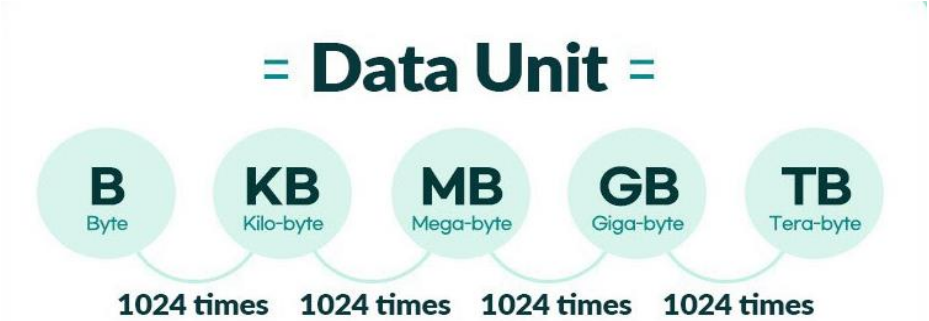


- There are many advantages to using the binary number system. Here are some of the most important ones:
  - **Simplicity**: The binary number system is very simple. It only uses two digits, 0 and 1. This makes it easy to understand and use.

- **Efficiency**: The binary number system is very efficient. It can represent any number with a finite number of digits. This makes it ideal for use in computers and other digital devices.
- **Accuracy**: The binary number system is very accurate. It is very unlikely that there will be any errors when representing numbers in binary. This is because there are only two possible values for each digit.
- **Reliability**: The binary number system is very reliable. It is very unlikely that there will be any errors when storing or transmitting binary data. This is because binary data is very easy to check for errors.

Binary Number System

- A bit is the smallest unit of data in a computer.
- The memory is made up of **BITS** and **BYTES.**
- Bits and bytes are extensively used to describe storage capacity and network access speed.
- The number of bits used to represent an address determines the total number of addresses that can be represented.



Unit	Definition	Storage space size
Bit	0 or 1	Yes/No
1 Byte	8 bit	Alphabets and one number
1 kilobyte (KB)	1,024 Byte	A few paragraphs
1 megabyte (MB)	1,024 KB	One minute-long MP3 song
1 gigabyte (GB)	1,024 MB	30 minute-long HD movie
1 terabyte (TB)	1,024 GB	About 200 FHD movies

$2^{10} = 1024$

Figure 3

For example, in an 8-bit address, each bit can be either a 0 or a 1, which means that there is 2^8 = 256 possible addresses. This means that if you have an 8-bit address bus, you can access 256 different memory locations.

A 32-bit address can represent 2^32 = 4,294,967,296 possible addresses, which corresponds to 4GB of memory.

ASCII and Unicode

- ASCII and Unicode are both character encoding schemes.
- A character encoding scheme is a way of representing characters using a sequence of bits.
- ASCII stands for **American Standard Code for Information Interchange**. It is a 7-bit character encoding scheme that represents 128 characters, including letters, numbers, and symbols.
- Unicode is a 16-bit character encoding scheme that represents over 130,000 characters, including those from a wide range of languages.

Table 2

Feature	ASCII	Unicode
Number of bits	7	16
Number of characters	128	Over 130,000
Range of characters	English and Western European languages	A wide range of languages

4. Error Checking – Parity

- Error checking is a general term for any method of detecting errors in data.
- Parity is a specific type of error checking that uses a parity bit to detect errors.

- A parity bit is an extra bit that is added to a data word to make the number of 1 bit in the data word even or odd.

- There are two modes of parity
  - **Odd parity**: The number of 1-bits (including the parity bit) must be an odd number
  - **Even parity**: The number of 1-bits (including the parity bit) must be an even number

- A single bit is appended to each data chunk (either as the least or most significant bit)

  Table 3

Original Data	Even Parity	Odd Parity
0 0 0 0 0 0 0 0	0	1
0 1 0 1 1 0 1 1	1	0
0 1 0 1 0 1 0 1	0	1
1 1 1 1 1 1 1 1	0	1
1 0 0 0 0 0 0 0	1	0
0 1 0 0 1 0 0 1	1	0

  *For example, if a **data word has 7 bits**, the parity bit would be set to **1** if the number of 1 bit in the data word is **odd**, and it would be set to **0** if the number of 1 bit in the data word is **even**.*

  When data is transmitted, the parity bit is also transmitted. If the receiver receives the data word and the parity bit does not match, it knows that an error has occurred.

Here is another example of how parity can be used to detect errors:

Data word: 1010100

Parity bit: 1

*If the data word is transmitted and the receiver receives **1010101**, the receiver knows that an error has occurred because the parity bit does not match the data word.*

5. Data Compression

- Data compression is the process of reducing the size of a data file by removing unnecessary or redundant information.

- This can be done without losing any of the essential data, or with some loss of data, depending on the type of compression used.

- Compression is commonly referred to as "zipping".

- The process of reconstituting zipped files is called extracting or unzipping. Compressed files may end with a .zip, .gz, .pkg, or .tar.gz