Computer and Data, Data processing

Ravindra Singh
Asst.Prof.
Computer Department
Durga College, Raipur (C.G)

A computer is a programmable electronic device that accepts raw data as input and processes it with a set of instructions to produce the result as output. It renders output just after performing mathematical and logical operations and can save the output for future use. It can process numerical as well as non-numerical calculations. The term "computer" is derived from the Latin word "computare" which means to calculate.

A computer is designed to execute applications and provides a variety of solutions through integrated hardware and software components. It works with the help of programs and represents the decimal numbers through a string of binary digits. It also has a memory that stores the data, programs, and result of processing. The components of a computer such as machinery that includes wires, transistors, circuits, hard disk are called hardware. Whereas, the programs and data are called software.

It is believed that the Analytical Engine was the first computer which was invented by Charles Babbage in 1837. It used punch cards as read-only memory. Charles Babbage is also known as the father of the computer.

The basic parts without which a computer cannot work are as follows:

- **Processor:** It executes instructions from software and hardware.
- **Memory:** It is the primary memory for data transfer between the CPU and storage.
- **Motherboard:** It is the part that connects all other parts or components of a computer.
- **Storage Device:** It permanently stores the data, e.g., hard drive.
- **Input Device:** It allows you to communicate with the computer or to input data, e.g., a keyboard.
- **Output Device:** It enables you to see the output, e.g., monitor.

The characteristics of the computer system are as follows −

Speed

A computer works with much higher speed and accuracy compared to humans while performing mathematical calculations. Computers can process millions (1,000,000) of instructions per second. The time taken by computers for their operations is microseconds and nanoseconds.

Accuracy

Computers perform calculations with 100% accuracy. Errors may occur due to data inconsistency or inaccuracy.

Diligence

A computer can perform millions of tasks or calculations with the same consistency and accuracy. It doesn't feel any fatigue or lack of concentration. Its memory also makes it superior to that of human beings.

Versatility

Versatility refers to the capability of a computer to perform different kinds of works with same accuracy and efficiency.

Reliability

A computer is reliable as it gives consistent result for similar set of data i.e., if we give same set of input any number of times, we will get the same result.

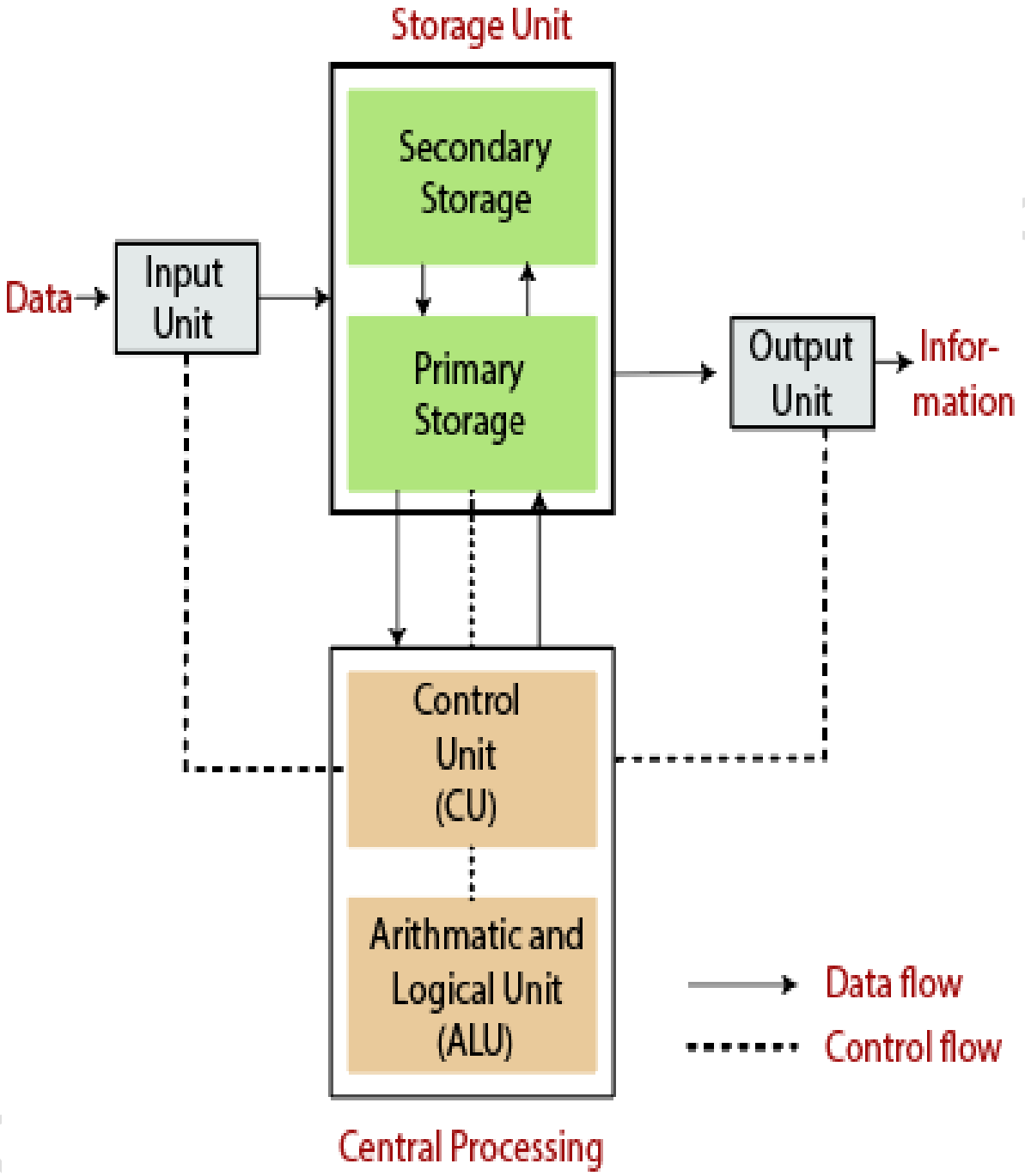
Automation

Automation is the use of technology to accomplish a task with as little human interaction as possible. In computing, automation is usually accomplished by a program, a script, or batch processing..

Memory

A computer has built-in memory called primary memory where it stores data. Secondary storage are removable devices such as CDs, pen drives, etc., which are also used to store data.

Block diagram of Computer



Data

In computing, data is information that has been translated into a form that is efficient for movement or processing. Relative to today's computers and transmission media, data is information converted into binary digital form. It is acceptable for data to be used as a singular subject or a plural subject

Data processing is the collection and manipulation of digital data to produce meaningful information. It is a form of information processing, which is the modification of information in any manner detectable by an observer.

The data processing cycle typically consists of the following steps:

1. Data collection: This is the process of gathering data from a variety of sources, such as sensors, databases, and surveys.
2. Data cleaning: This is the process of removing errors and inconsistencies from the data.
3. Data preparation: This is the process of formatting the data so that it can be processed by a computer.
4. Data processing: This is the process of performing operations on the data, such as sorting, filtering, and summarizing.
5. Data analysis: This is the process of extracting insights from the data.
6. Data presentation: This is the process of communicating the insights to decision-makers.

Types of data processing

- Batch processing: Batch processing is a type of data processing in which data is collected and processed in batches. This type of processing is typically used for large volumes of data that do not need to be processed in real time.
- Online processing: Online processing is a type of data processing in which data is processed as soon as it is entered. This type of processing is typically used for applications that require real-time response, such as customer service applications.
- Real-time processing: Real-time processing is a type of data processing in which data is processed as it is generated. This type of processing is typically used for applications that require immediate response, such as financial trading applications.
- Distributed processing: Distributed processing is a type of data processing in which data is processed across multiple computers. This type of processing is typically used for applications that require high performance or scalability.
- Parallel processing: Parallel processing is a type of data processing in which multiple processors are used to process data simultaneously. This type of processing is typically used for applications that require high performance.

Type of data processing	Description	Advantages	Disadvantages
Batch processing	Data is collected and processed in batches.	- Efficient for large volumes of data. - Can be used for historical analysis.	- Not suitable for real-time applications. - Can be slow for small batches.
Online processing	Data is processed as soon as it is entered.	- Suitable for real-time applications. - Can be used for ad hoc queries.	- Not suitable for large volumes of data. - Can be more expensive than batch processing.
Real-time processing	Data is processed as it is generated.	- Suitable for applications that require immediate response. - Can be used for fraud detection.	- Not suitable for large volumes of data. - Can be very expensive.
Distributed processing	Data is processed across multiple computers.	- High performance. - Scalable.	- Complex to manage. - Can be more expensive than centralized processing.
Parallel processing	Multiple processors are used to process data simultaneously.	- High performance. - Can be used for large volumes of data.	- Complex to manage. - Can be more expensive than serial processing.

Shorting Data

Data processing, "shorting data" refers to the process of removing or filtering out data points that are outliers or that do not fit the expected distribution. This can be done to improve the accuracy of models or to make the data more manageable.

There are a number of different ways to short data, including:

- Outlier detection: This involves identifying data points that are significantly different from the rest of the data set. These data points can then be removed or filtered out.
- Data normalization: This involves transforming the data so that it has a standard distribution. This can make the data more manageable and can improve the accuracy of models.
- Feature selection: This involves selecting only the most important features for a model. This can help to improve the accuracy of the model and can make it more efficient.

Shorting data can be a valuable tool for improving the accuracy and efficiency of models. However, it is important to use caution when shorting data, as it can also remove important information.

Here are some of the benefits of shorting data:

- Improved accuracy: Shorting data can help to improve the accuracy of models by removing outliers and other noise from the data.
- Increased efficiency: Shorting data can help to increase the efficiency of models by reducing the number of data points that need to be processed.
- Improved interpretability: Shorting data can help to improve the interpretability of models by making the relationships between the features and the target variable more clear.

Here are some of the challenges of shorting data:

- Loss of information: Shorting data can result in the loss of important information. This can make it difficult to build accurate models or to interpret the results of models.

- Overfitting: Shorting data can lead to overfitting, which is when a model is too closely fit to the training data and does not generalize well to new data.
- Data bias: Shorting data can introduce bias into the model, which can lead to inaccurate results.

Overall, shorting data can be a valuable tool for improving the accuracy and efficiency of models. However, it is important to use caution when shorting data, as it can also remove important information and introduce bias into the model.

Reference

- https://www.javatpoint.com/what-is-computer
- https://www.tutorialspoint.com/basics_of_computers/basics_of_computers_quick_guide.ht
- https://bard.google.com/
- https://www.tutorialandexample.com/block-diagram-of-a-computer