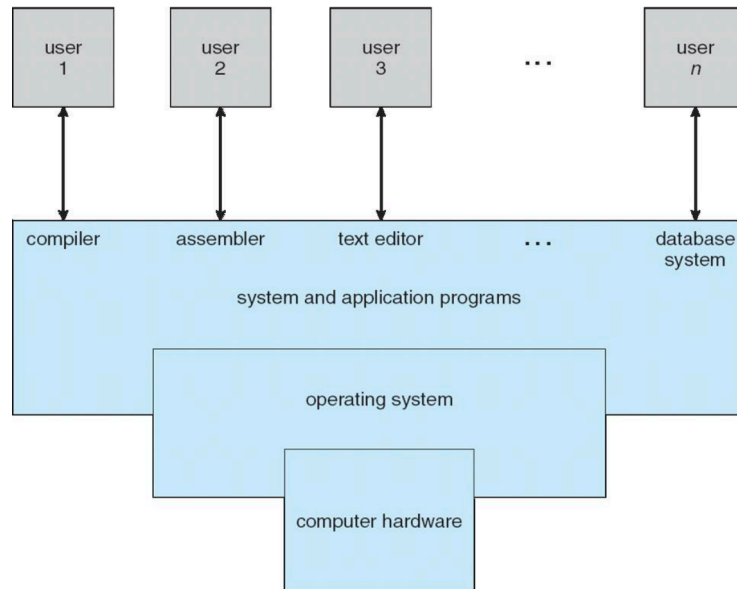# Introduction

An operating system is a program that acts as an indermediary between the user and the hardware. The main goals of an operating system are to:

- execute programs that solve problems
- make the computer easy to use
- use the hardware in an efficient manner

To do so it coordinates access to the hardware among the different applications and users.

The program that runs all the time is the **kernel**, everything else is either a system program (ships with the operating system) or an application program.

Computer startup

The first program that runs on sturtup is the **bootstrap program**. This is stored in ROM or EEPROM and is usually called **firmware**. It initializes all aspects of the system and loads the kernel.

Computer architecure

The general model of a computer architecture is the following: there are one or more CPUs and device controllers that are connected through a common bus with a shared memory.
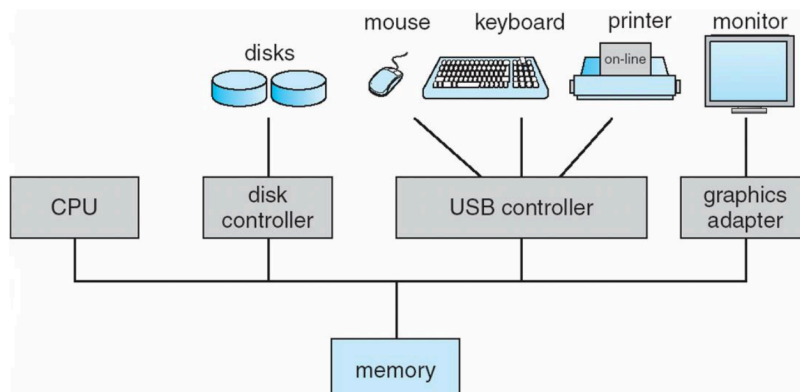
Interrupts

In interrupt driven operating systems, device controllers communicate with the CPU using **interrupts**. While waiting for the CPU device controllers store information in local buffers, that are then read/written by the CPU and the stored data is trasferred from/to the main memory.
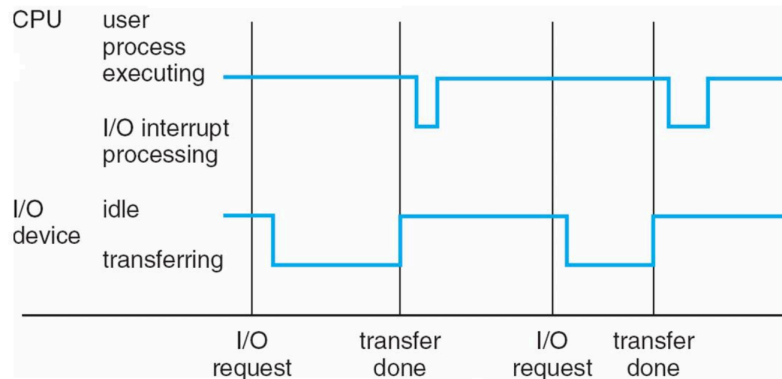
When an interrupt happens, the OS saves the current program counter (PC), jumps to the routine that handles the interrupt and then resumes normal execution by jumping back to where the PC was pointing to. The position to where to jump to when an interrupt is received is stored in the **interrupt vector**, which is a map where each interrupt is linked to an instruction memory address.

A **trap** or **exception** is a software-generated interrupt caused by an error or a user request.

Depending on the importance of the interrupt, some interrupts must be handeled imediately, while others can wait. The first ones are called **non-maskable**, while the laatter are **maskable**.

While transferring data, the CPU receives an interrupt when every chunk of data has been successfully received. This can be very very wasteful, therefore a feature called DMA (Direct Memory Access) has been introduced. In this tecnique the CPU receives an interrupt only after all the data has been transferred successfully.

Storage

Storage systems are categorized by speed, cost and volatility. Each storage systems has therefore its advantages and disadvantages, therefore there is no "best" storage device.

The main memory of a computer, which can be accessed directly by the CPU is DRAM (dynamic random access memory, based on charged capacitors) or SRAM (static random access memory, based on inverters), which is usually volatile. Secondary storage has much bigger capacity and is non-volatile (ex. hard disks, solid-state drives).

Each storage system has a device controller and a device driver. The driver provides an uniform interface between the controller and the kernel.

Caching

Caching is a very common tecninque for speeding up access to commonly used data. Information is copied from slower to faster storage temporarely and then every time that information is needed the OS will first check if it is present in cache. Due to cost constraints, cache is very often smaller than the storage being cached, therefore cache management must be properly optimized.