Operating System

- Software that manages computer resources and provides an interface for system interaction.
- Starts running after getting loaded into memory by the bootloader.
- Can have multiple operating systems from which the user can select one when booting up.
  - Dual-boot or multi-boot

Operating System

- Operating Systems can be classified into various types:
  - Real-time OS
  - Multi-user vs Single-user
  - Multi-tasking vs Single-tasking
  - Distributed
  - Embedded

Operating System

- Multiprogramming – The technique of keeping multiple programs in main memory at the same time, competing for the CPU.
- Memory Management – The act of keeping track of how and where programs are loaded in main memory.
- Process – The dynamic representation of a program during execution.

Operating System

- Process Management – The act of keeping track of information for active processes.
- CPU Scheduling – The act of determining which process in memory is given access to the CPU so that it may execute.
- Kernel – The core part of the OS which controls every other part of the OS.
- Interrupts – An efficient way for the OS to interact with and react to its environment.

Operating System

- Modes – States in which the CPU operates.
- Filesystem – Data is stored in disks as files and accessed in a particular way.
- Device Drivers – Interacting with the hardware using device drivers.
- Networking – Connect to other machines.
- Security – OS provides controlled access to resources.
- User Interface – The interface to the computer.

Kernels

- Main component of most operating systems.
- Responsible for managing the system's resources.
- Interacts with applications using system calls.
- Monolithic Kernels – execute all OS code in the same address space to increase performance.
- Microkernels – execute most of the OS in user space as servers to increase modularity.
- Linux is a monolithic kernel.

Interrupts

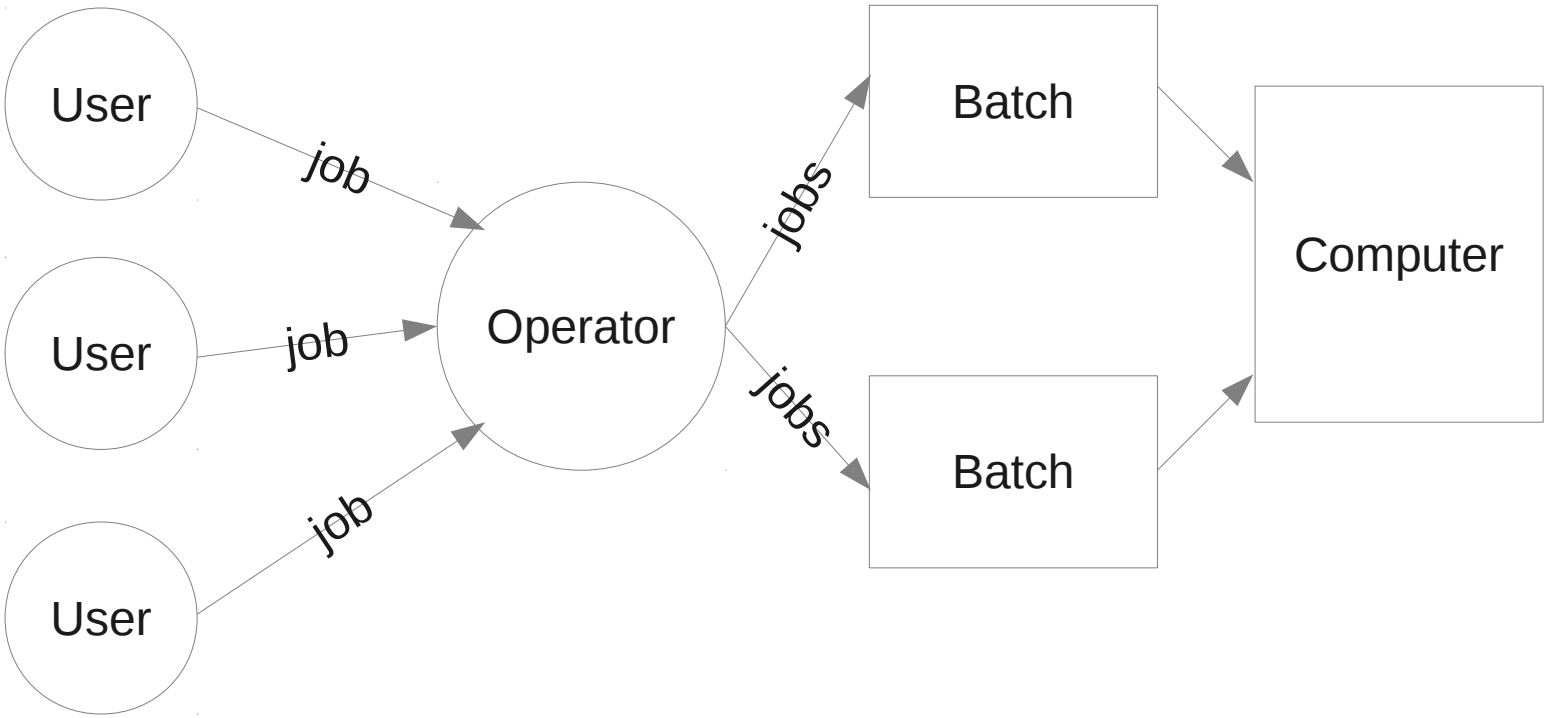
- An interrupt is an asynchronous signal indicating the need for attention or a synchronous event in software indicating the need for a change in execution.
- Hardware Interrupt – processor saves state and begins executing interrupt handler.
- Software Interrupt – causes context switch to an interrupt handler.
- Act of interrupting is interrupt request (IRQ).

Operating Modes

- Modern CPUs support multiple modes of operation. At least:
  - Supervisor mode – used by OS kernel for low level tasks that need unrestricted access to hardware.
  - Protected mode – for everything else.
- Application live only in protected mode and hardware access is through the kernel in supervisor mode.
- Virtual modes to emulate older processor types.

Operating System

- Batch Processing



Operating System

- Timesharing – A system in which CPU time is shared among multiple interactive users at the same time.
- Each user is represented by a login process that runs in the machine.
- Each program a user runs causes a new process to be created – spawned by the user's login process.
- The CPU time is shared by all of the processes created by all of the users.

Operating System

- Real-time system -- A system in which response time is crucial given the nature of the application domain.
- Response Time – The time delay between receiving a stimulus and producing a response.
  - Real-time systems strive to optimize the response time.
- Real-time responses are crucial in software that controls a robot or a nuclear reactor.
- Hard and Soft real-time.

Memory Management

- Multiple programs and their data are stored simultaneously in the main memory at the same time.
- The Operating System must manage memory to:
  - Track where and how a program resides in memory.
  - Convert logical program addresses into actual memory addresses.

Memory Management

- Programs often refer to memory locations.
- There is no way to know beforehand the actual locations the program will access.
- Solved by using two addresses:
  - Logical Address (Virtual or Relative address) – a value that specifies a generic location, relative to the program but not to the reality of the main memory.
  - Physical Address – an actual address in the main memory.
- Mapping of a logical address to a physical address is called address binding.

Memory Management

- Three schemes for memory management on the order of increasing complexity:
  - Single contiguous memory management
  - Partition memory management
  - Paged memory management
- Single contiguous memory management
  - A program is loaded into one continuous area of memory.
  - Only one program other than the OS can be in memory.

Memory Management

- Partition memory Management
  - More than one application is loaded into memory.
  - Memory is divided into partitions.
    - Fixed partitions – main memory is divided into a particular number of partitions.
    - Dynamic partitions – partitions are created as required to fit the unique needs of the program.
  - Base Register – A register that holds the beginning address of the current partition.
  - Bounds Register – A register that holds the length of the current partition.

Memory Management

- Paged Memory Management
  - Processes are divided into fixed size pages and stored in memory when loaded.
    - Frame : A fixed size portion of main memory that holds a process page.
    - Page : A fixed size portion of a process that is stored into a memory frame
  - Page-map table (PMT) – The table used by the operating system to keep track of page/frame relationships.

Memory Management

P1	PMT
Page	Frame
0	5
1	12
2	15
3	7
4	22

P2	PMT
Page	Frame
0	10
1	18
2	1
3	11

Memory	
Frame	Contents
0	P2/Page2
1	
2	
3	
4	
5	P1/Page0
6	
7	P1/Page3
8	
9	
10	P2/Page0
11	P2/Page3
12	P1/Page1
13	
14	
15	P1/Page2

Memory management

- Demand Paging – An extension to paged memory management in which pages are brought into memory only when referenced (on demand).
- Page swap – Bringing one page from secondary memory, possibly causing another to be removed.

Memory Management

- Virtual Memory – The illusion that there are no restrictions on the program size because the entire process need not be in memory at the same time.
- Thrashing – Inefficient processing caused by constant page swapping.