Operating System (OS) is an intermediary between users and computer hardware. It provides users an environment in which a user can execute programs conveniently and efficiently. In technical terms, it is software which manages hardware. An operating System controls the allocation of resources and services such as memory, processors, devices and information.

**OS Types:**

* Batch operating system
  + To speed up processing, jobs with similar needs are batched together and run as a group.
* Time-sharing or multitasking
  + Processor's time is shared among multiple users simultaneously
* Distributed (loosely coupled systems) operating System
  + Distributed systems use multiple central processors to serve multiple real time application and multiple users.
  + Data processing jobs are distributed among the processors accordingly to which one can perform each job most efficiently.
  + The processors communicate with one another through various communication lines (such as high-speed buses or telephone lines).
* Network Operating System
  + Runs on a server and provides server the capability to manage data, users, groups, security, applications, and other networking functions.
* Real Time operating System
  + Response time is very less as compared to the online processing.

Time-sharing is a logical extension of multiprogramming.

In case of multiprogrammed batch systems, objective is to maximize processor use,

Whereas in Time-Sharing Systems objective is to minimize response time.

Batch processing is a technique in which Operating System collects one programs and data together as a single unit in a batch before processing starts and processed them in FCFS fashion.

**Process**

Process is a program in execution.

* Program contains the instructions to be executed by processor.
* Process is a static entity made up of program statement while process is a dynamic entity.

**Component of Process**

* Code to be executed.
* Data to be used for executing the program.
* Process requires some resources for execution.

**Process States**



**Process Control Block, PCB**

PCB is the data structure **used** by the operating system to **represent** each process.

PCB contains many pieces of information associated with a specific process

* **Pointer** points to another PCB. Pointer is used for maintaining the scheduling list.
* **Process State**
* **Program Counter** indicates the address of the next instruction to be executed for this process.
* **CPU registers** include general purpose register, stack pointers, index registers and accumulators
* **Memory management information** - the value of base and limit registers, the page tables, or the segment tables
* **Accounting information** This information includes the amount of CPU and real time used, time limits, job or process numbers, account numbers etc.

When process gets suspended, the contents of the registers are saved on a stack and the pointer to the particular stack frame is stored in the PCB. By this technique, the hardware state can be restored so that the process can be scheduled to run again.

**context switch:**

In multitasking environment, a context switch is the mechanism to **store** and restore the state or context of a CPU in PCB so that a process execution can be resumed from the same point at a later time.

The context of a process is represented in the PCB of a process. Context switch time is pure overhead.

When the process is switched, the following information is stored. ? Program Counter ? Scheduling Information ? Base and limit register value ? Currently used register ? Changed State ? I/O State ? Accounting

Process scheduling is an essential part of a Multiprogramming operating system. where more than one process to be loaded into the executable memory at a time and loaded process shares the CPU using time multiplexing.

When the process enters into the system, then this process is put into a job queue. This queue consists of all processes in the system. The operating system also maintains other queues such as device queue. Device queue is a queue for which multiple processes are waiting for a particular I/O device. Each device has its own device queue.

**Schedulers**

Schedulers are special system software which handles process scheduling.

Their main task is to select the jobs to be submitted into the system and to decide which process to run.

Schedulers are of three types

* Long Term Scheduler- job scheduler Job scheduler selects processes from the queue and loads them into memory for execution.
* Short Term Scheduler - CPU scheduler selects process among the processes that are ready to execute and allocates CPU to one of them.
  + execute most frequently and makes the fine grained decision of which process to execute next.
* Medium Term Scheduler - Medium term scheduling is part of the swapping. It removes the processes from the memory. It reduces the degree of multiprogramming.

**Process scheduling algorithms:**

* First Come First Serve (FCFS) Scheduling
* Shortest-Job-First (SJF) Scheduling
  + Best approach to minimize waiting time. ?
  + Impossible to implement ?
  + Processer should know in advance how much time process will take.
* Priority Scheduling ?
* Round Robin(RR) Scheduling ?
* Multilevel Queue Scheduling

Parameters used to evaluate scheduling algorithms:

* Average wait time
* First Response time
* Wait Time : Service Time - Arrival Time

Memory management

* Handles or manages primary memory.
* Keeps track of each and every memory location either it is allocated to some process or it is free. checks how much memory is to be allocated to processes.
* Decides which process will get memory at what time.
* Tracks whenever some memory gets freed or unallocated and correspondingly it updates the status.

Dynamic Loading - a routine of a program is not loaded on main memory until it is called by the program.

Dynamic loading makes better memory space utilization and unused routines are never loaded.

Linking is the process of collecting and combining various modules of code and data into a executable file that can be loaded into memory and executed.

Operating system can link system level libraries to a program.

When it combines the libraries at load time (program code size becomes bigger), the linking is called static linking and when this linking is done at the time of execution, it is called as dynamic linking.

**Page Replacement Algorithm**

Page replacement algorithms are the techniques using which Operating System decides which memory pages to swap out, write to disk when a page of memory needs to be allocated.

* First In First Out (FIFO) algorithm
* Optimal Page algorithm - Replace the page that will not be used for the longest period of time
* Least Recently Used (LRU) algorithm - not been used for the longest time
* Least frequently Used (LFU) algorithm
* Most frequently Used (LFU) algorithm

**concurrency-control problems.**

**readers–writers problem**

database is to be shared among several concurrent processes.

Some of these processes may want only to read the database,

Whereas others may want to update (that is, to read and write i.e writers) the database.

If two readers access the shared data simultaneously, no adverse effects will result.

However, if a writer and some other process (either a reader or a writer) access the database simultaneously, chaos may ensue.

We require that the writers have exclusive access to the shared database while writing to the database.

Different variations:

The first readers–writers problem,

requires that no reader be kept waiting unless a writer has already obtained permission to use the shared object. i.e. no reader should wait for other readers to finish simply because a writer is waiting. If readers are there write must wait, writer can access only when reader count is zero.

The second readers–writers problem requires that, once a writer is ready, that writer perform its write as soon as possible.

In other words, if a writer is waiting to access the object, no new readers may start reading.

A solution to either problem may result in starvation.

File System

File is a named collection of related information recorded on secondary storage

File Structure File structure is a required format that operating system can understand.

A file has a certain defined structure according to its type. text file, source file, object file

When operating system defines different file structures, it also contains the code to support these file structure. UNIX, MS-DOS support minimum number of file structure.

File Access Mechanisms refers to the manner in which the records of a file may be accessed.

There are several ways to access files

Sequential access

Direct/Random access

Indexed sequential access

Space Allocation: Files are allocated disk spaces by operating system in following three main ways.

* Contiguous Allocation
  + Each file occupies a contiguous address space on disk.
  + Suffers with External fragmentation
* Linked Allocation
  + Each file carries a list of links to disk blocks.
  + Inefficient in case of direct access file.
* Indexed Allocation
  + For every file index block is created.
  + Index block stores the addresses of disk space occupied by the file.