3 Basic Software

Chapter Objectives

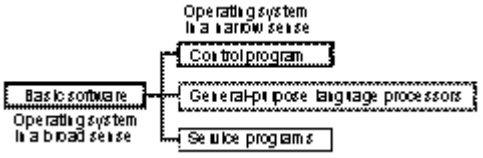
In order to use a computer, we need basic software.
In order to efficiently operate the hardware composing the computer system as well as the application software, it is necessary to understand the mechanism and functions of the operating system (control program), which performs different kinds of control/management. Therefore, we need to:

① Understand well the software names and classifications, functions and roles, including the relation with the hardware and the user.
② Understand the reasons why the operating system is necessary, its roles, structure, functions, etc.
③ Understand the types and characteristics of the major operating systems.

Introduction

Software that helps users to make effective use of the hardware functions is generically called system software. Systems software is roughly classified into basic software and middleware.
A basic software is a set of programs aimed at efficient use and control of the different types of resources provided by the hardware. It can be thought as the operating system in a broad sense.
As it is shown in Figure 3-1-1, the basic software itself is classified into control program, language processors and service programs. Software has a great number of complex functions, which will be explained in detail.

Figure 3-1-1
Basic software



3.1 Operating system

Software corresponding to its purpose is incorporated into computers utilized on various field. Such software is called "Application Software."
On the other hand, the software acting as a bridge between the application software and the hardware is the operating system (OS), which will be studied hereafter.
Here, the objectives and functions of the operating system, which is the closest system software to the hardware, will be studied.

Figure 3-1-2
Operating system
position

3.1.1 OS configuration and functions

The fully-fledged operating system was born in the 1960s. Since computers in that era were extremely expensive, the main consideration of the users was how to operate them efficiently. For example, if the computer was laying idle while the data to be processed by the computer was being prepared or while the results processed by the computer were being processed manually by humans, it can not be said that these expensive computers were being efficiently used.
Therefore, the operating system was born for the purpose of having the computer prepare the data to be processed and control the execution process by itself.

(1) OS role

The following are the purposes of the operating system that is incorporated in today's general-purpose computers, and is used in different fields:

- Efficient use of resources
- Consecutive job processing
- Multiple programming
- Reduction of the response time
- Improvement of reliability.

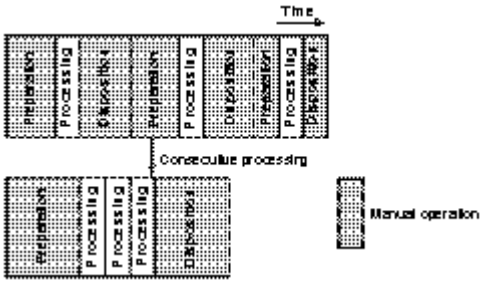
① Efficient use of the resources

From the operating system point of view, the processor, main memory unit, auxiliary storage devices, input devices, output devices, application software, and other components of the computer system are all resources for computer use. The purpose of the operating system is to efficiently use these resources without relying on humans and without waste.

② Consecutive job processing

The work done by a computer is called job. If human manual operation is required between jobs that process data at electronic speeds, the processor use efficiency would drop dramatically.
For that reason, by eliminating as much human intervention as possible, the operating system implements automatic consecutive processing and enhances the processing efficiency of the whole computer system.

Figure 3-1-3
Consecutive job processing



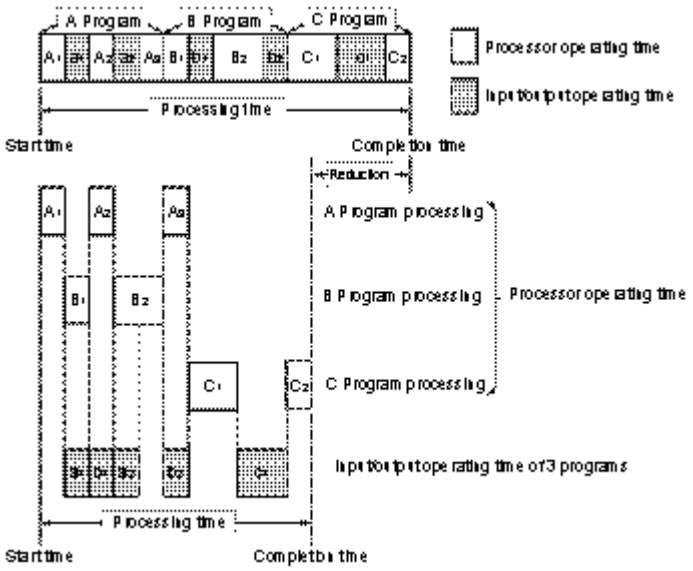
③ Multi-programming

Multi-programming consists of an approach to simultaneously processing multiple jobs with the same processor. If multiple jobs can be processed simultaneously, the computer processing efficiency will, of course, improve. The operating time of a job processed in a computer can be divided into the following:

- Time during which the data to be processed is entered
- Time during which calculations and other processes are performed using the processor
- Time during which process results are outputted

Most of the data input and process result's output are mechanical operations. Compared to them, the calculations and other processes performed using the processor are electronic operations. Therefore most of the time, the processor has to wait for the input/output operations.
For that reason, multi-programming was born from the need to separate the input/output operations and the processing operations so that the processor idle time can be used to process other job computing, etc.

Figure 3-1-4 Multi-programming approach



④ Reduction of the response time

The response time is the time elapsed since input from the terminal, etc. is completed until the system output resulting from that input is started. For example, when a reserved-seat ticket is bought via an automatic dispenser if the time elapsed since the necessary information is entered in the terminal device until the ticket is issued is too long, a long queue would be formed. For online transaction processing systems of this kind, the reduction of the response time is a factor of great importance.

⑤ Improvement of reliability

The improvement of the reliability of the different computer system components is also an important role played by the OS of general-purpose computers.

⑥ Other roles

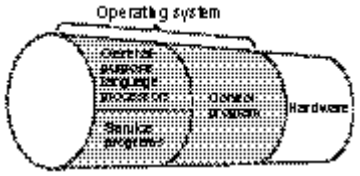
Information processing engineers do not have a thorough knowledge of every single detail of the computers. Therefore, one important function of the operating system is enabling "user friendliness" so that software development can be performed without having to keep in mind the hardware functions.
Likewise, "extensibility" of the operating system itself as well as the resources in order to support the increase of the information to be processed.

(2) OS configuration

The operating system with its complex and wide range of functions is composed of diverse programs. Figure 3-1-5 shows the relation among the following components of the operating system:

- Control program
- General-purpose language processors
- Service programs

Figure 3-1-5
Operating system configuration



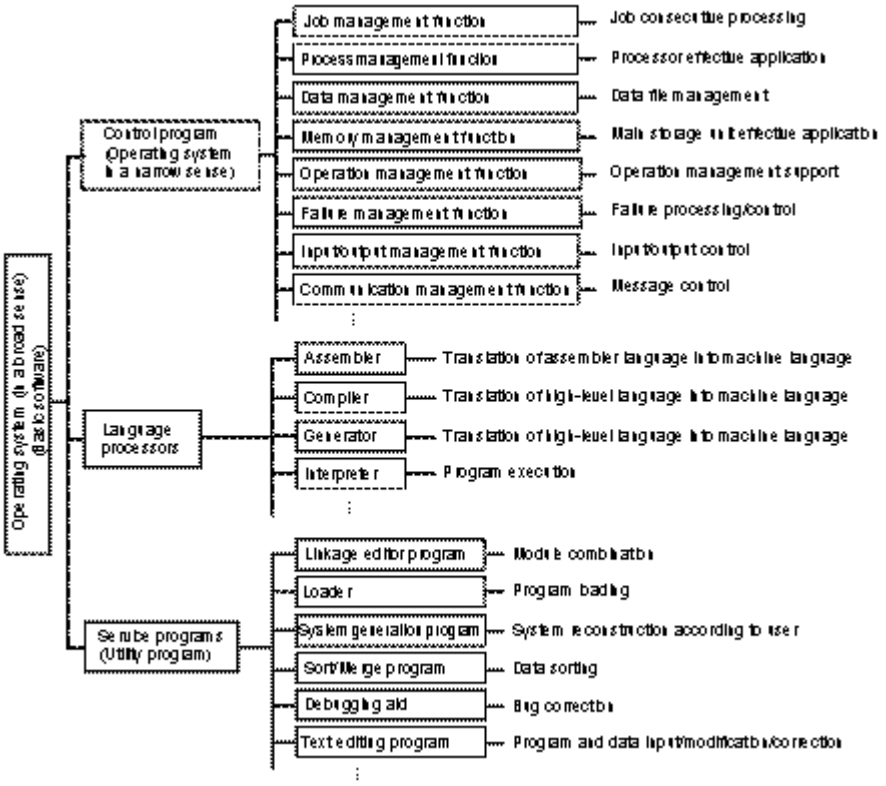
(3) OS functions

The control program, which is the nucleus of the operating system, is equipped with diverse functions such as the ones mentioned below:

- Job management function
- Process management function
- Data management function
- Memory management function
- Operation management function
- Failure management function
- Input/output management function
- Communication management function

The outline of the operating system functions is shown in Figure 3-1-6.
Here, among the different functions, mainly the control program functions, which are functions aimed at enabling efficient use of the hardware, will be explained. The language processors and service programs will be studied in Section 3.4.

Figure 3-1-6
Operating system functions



3.1.2 Job management

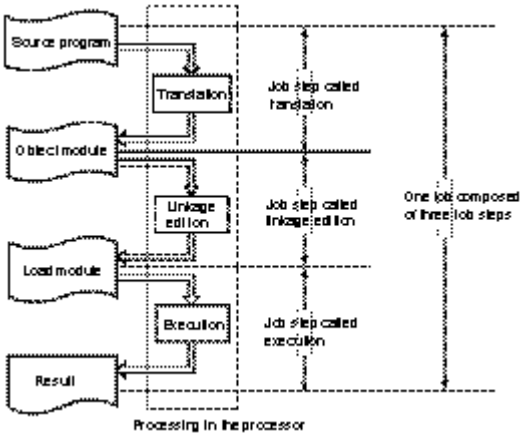
The main purpose of job management is to improve the computer system processing capacity by performing consecutive processing of the job.
In order to implement the consecutive job processing, the following are indispensable:

- Job control language
- SPOOL

(1) Job control language (JCL)

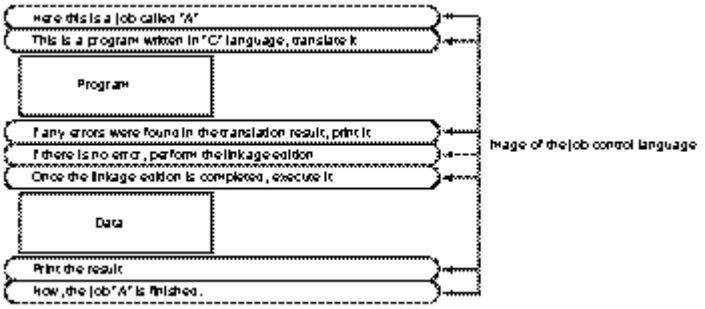
The unit of the works given to the computer by humans is called the job. Commonly, one job is composed of multiple job steps.
Figure 3-1-7, shows one job, that is, the process in which, after a given application software is submitted into the computer, data is entered and processed. The source programs written by humans become processable only after they are translated into the machine language (object module) by the language processor and edited by the linkage editor.

Figure 3-1-7
Job and job steps



In order to move from one job step to another, detailed instructions have to be provided to the computer system. The job control language is used to provide these instructions.
Since the job control language provides the instructions "Translation," "Linkage editing," "Execution," etc. to the job submitted to the computer, processing is conducted without having to rely on humans (Figure 3-1-8). The function of the job management is to decode and execute these detailed instructions written in the job control language.

Figure 3-1-8
Job control language functions



The syntax of the job control language differs depending on the operating system, but the main statements are as follows:

① JOB statement

The job to be submitted to the computer system is given a name and the job start is declared using the JOB statement.

② EXEC statement

Control information such as the order of execution of the programs performing the processing is indicated using the EXEC statement.

③ DD statement

The location where the files required for the process are located, etc. is indicated using the DD statement.

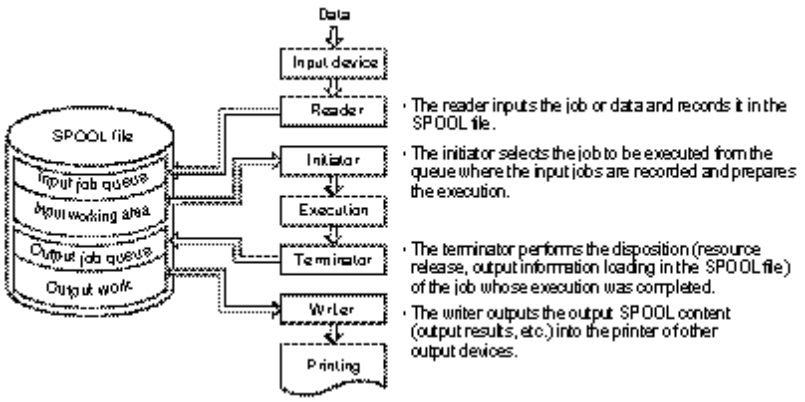
(2) SPOOL (Simultaneous Peripheral Operations Online)

SPOOL is an indispensable function in multi-programming environments.
If a given program occupies the printer to print the process result, even if the processor is free, it can not process other programs scheduled to use the printer. In order to solve this problem, for the execution of all the programs, the process result is once written out onto an auxiliary device before proceeding to print. In other words, the processor and the printer are physically separated. This is the SPOOL approach.

(3) Job scheduling

The series of controls performed after a job described in JCL, etc. is entered into the computer until the result outputted is called job scheduling (Figure 3-1-9). In practice, this processing is executed by the job scheduler using a dedicated program incorporated in the OS.

Figure 3-1-9
Job scheduling



3.1.3 Process management

For the operating system, a process (task) is the control unit of a job. The main purpose of the process management is to efficiently use the processor. In order to achieve this purpose, the operating system performs this process management.

(1) Execution control

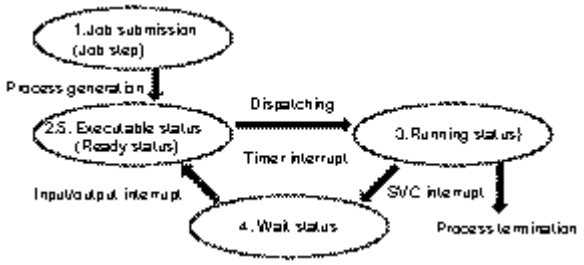
① State transition

After being subdivided into various job steps by the process management program, the jobs submitted to the computer are processed following a procedure such as the one mentioned below:

1. A job step is generated as a process that can be performed by the computer.
2. Immediately after being generated, the process turns into the executable status.
3. When the processor becomes vacant, the process in the executable status is immediately executed.
4. If input/output operations are generated while the process is being executed, the process turns into wait status.
5. When these input/output operations are completed, the process turns into the executable status again.

This procedure is called the processor state transition. The job step is converted into a processing unit called process and is processed while repeating the state transition through the process management function of the operating system.

Figure 3-1-10
Process state transition



Since the processor becomes free when the process being executed turns into the wait status, the approach of the process management consists of executing another process in the executable status during this time. The act of enhancing the efficiency of the processor use by controlling the status of multiple processes is called multiprocessing (multitasking).

② Dispatcher

The act of selecting the process to be executed from among the processes in the executable status for the processor allotment is called dispatching. The program that performs this operation is called the dispatcher. The following are the two main methods by which the dispatcher grants processor use rights:

a. Preemption

The preemption is the method by which an order of priority is given to each process and the processor is always assigned to the processes with high priority. In this method, when a process with a higher priority than the process being executed is generated, the execution of the process with lower priority is halted and the processor use is switched to the process with higher priority.

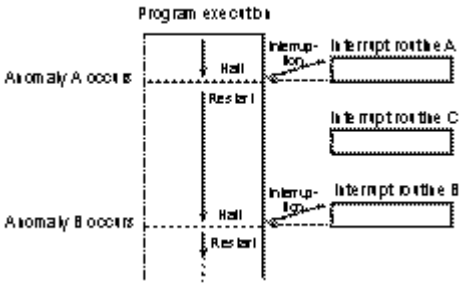
b. Round robin

The round robin is the method by which the processor use time is minutely divided (time slicing) and equally assigned to each process. In this method, once a process has used the processor for a specific time (time slice), its execution is halted and the process is sent to the last position in the process queue.

③ Kernel and interruption control

When the process state transition is performed, an interruption is performed in order to control the process execution. Interruption is the act of halting a program being executed to switch to the prevention program when a process transition or anomaly occurs (Figure 3-1-11).
This prevention program is called the interrupt routine and is stored in the computer in advance. Once the interrupt routine has been executed and the processing, etc. to fix the anomaly has been completed, the execution of the former program is restarted. The central part of the OS performing the interruption control is called the kernel.

Figure 3-1-11
Interruption



According to the location where the anomaly occurs, the interruption is divided into the following:

- Internal interrupt
- External interrupt

a. Internal interrupt

Internal interrupt is the general term for the interruptions that occur due to errors of the program itself. The internal interrupts that occur are as follows:

- Program interrupt
- Supervisor call interrupt

● Program interrupt

Program interrupt is the interruption that occurs due to an error generated during the execution of a program. For example, when the denominator of a division is zero, or when the number of digits of the result of an operation exceeds the acceptable limits, etc.

● Supervisor call (SVC) interrupt

This interruption occurs in cases where unless the operating system functions are used, a correct result can not be obtained, for example, when data input is requested during the execution of a program, etc.

b. External interrupt

External interrupt is the interruption that occurs due to external factors and not due to the program. The following external interrupts exist:

- Input/output interrupt
- Machine check interrupt
- Timer interrupt
- Console interrupt

● Input/output interrupt

Input/output interrupt occurs when an anomaly occurs in the input/output process completion report or in an input device or output device during processing.

● Machine check interrupt

Machine check interrupt occurs when a malfunction of the processor or the main storage unit or an anomaly in the power supply, etc. happen. The failure occurrence is reported to the operating system by the processor.

● Timer interrupt

Timer interrupt is an interruption generated by the timer contained inside the processor. Programs exceeding the execution time specified with the time sharing process, etc. are subject to forced termination by this interruption. Likewise, timer interrupt occurs when an abortion of programs of

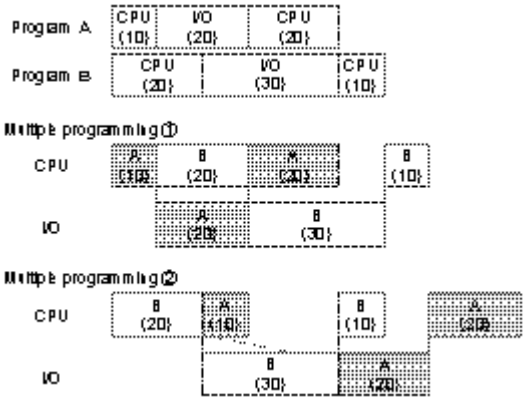
routines that never end, called infinite loops, is to be performed.

- Console interrupt
  Console interrupt is an interruption that occurs when a special process request was indicated from the operator console during the execution of a program.

(2) Multi-programming

Multi-programming is implemented by the process management to efficiently use the processor. This function enables the simultaneous execution of multiple programs through the execution of other processes while the process being executed turns into the wait status due to an input/output request.
This function is explained in Figure 3-1-12 using programs A and B as examples whereby an execution time of 50 seconds and 60 seconds for each program, respectively, is given. If program B is executed after the execution of program A, it will take 110 seconds to finish the execution of both programs (this is called the simple execution time).
However, using the multi-programming approach, which consists of the execution of one program while the input/output processing of another program is performed, the time needed to finish both programs will be 70 seconds (Multi-programming ①). It should be noted that in the event that program B is executed first, even if the same multi-programming approach is applied, the time needed to finish both programs will be 90 seconds (Multi-programming ②). From this, we can see that the execution order of the programs is very important for the processing efficiency.

Figure 3-1-12 Multiple programming



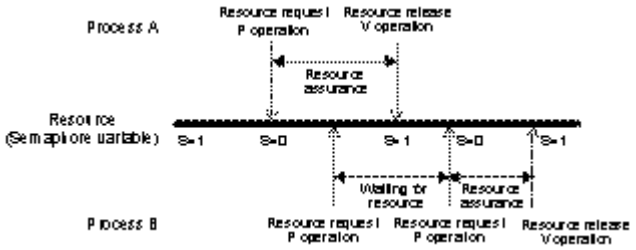
(3) TSS (Time sharing system)

By assigning equal CPU time to all the processes according to the round robin method, multiple users can simultaneously use one computer. This user format in which "one feels like the only user of the computer" is called TSS. TSS is one of the main interactive-type processing and it is used by a great number of centralized processing systems of the conventional host computer use.

(4) Exclusive control

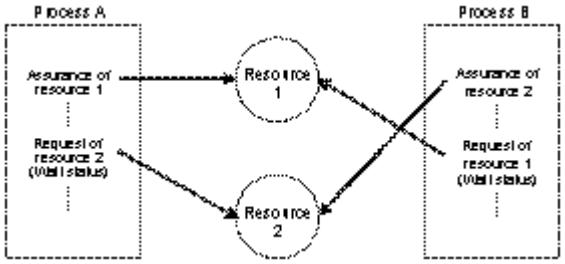
In the process generation stage, besides the processor, resources are assigned to each process. Here, the same resource can be shared by multiple processes; however, the same resource can not be used by all the processes at the same time. Therefore a semaphore is used to limit the resource use (exclusive control).
The semaphore, which in the broad sense is a word that means signal, is composed of semaphore variables and 2 operation instructions (P operation and V operation). The semaphore variables hold integer values according to the condition of each resource, and according to the integer value, the synchronization among the processes is conducted. In the binary semaphore, which is a typical semaphore, the semaphore values are 0 and 1. P operation is the operation for resource use that reduces the semaphore variable value. On the other hand, V operation is the operation for resource release that increases the semaphore variable value. Figure 3-1-13 shows an example of exclusive control using a binary semaphore.

Figure 3-1-13
Semaphore



Through the exclusive control performed using the semaphore, synchronization among processes is conducted and resource sharing is implemented. However, due to this control an event called deadlock can occur. Deadlock is the status in which two or more processes wait for the resource release of each other. Since the processes in this status are unable to assure a resource, processing is halted.

Figure 3-1-14
Deadlock



3.1.4 Main memory management

The main memory management controls the storage area of the main storage unit.
The following main memory management techniques exist:

- Partition method
- Swapping
- Overlay
- Memory protection

(1) Partition method

In the program storage method (or program built-in method) it is necessary to store programs and data in the main storage unit in advance. When programs are to be stored in the main storage unit, the method that divides the main storage unit into several parts, and stores programs in each of these parts is called the partition method, because these parts are called partitions.
The partition method can be roughly divided into the following three methods:

① Single-partition method

In the single-partition method, the main storage unit is controlled by dividing it into the area to store the control program and the area to store only one program. This method was applied in the early computers, but since it is not suitable for the efficient use of the main storage unit and other resources, it is practically obsolete today.

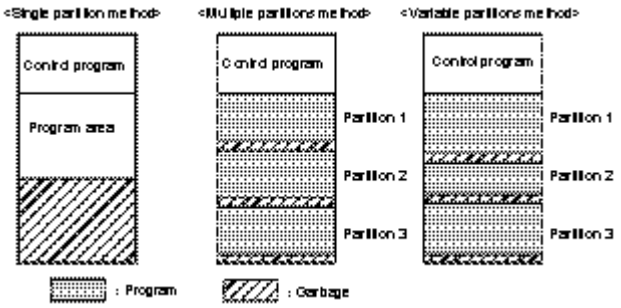
② Multiple partitions method

In the multiple partitions method, the program storage area is divided and multiple programs are stored in each of the partitions divided. This method was conceived for the implementation of multi-programming, however, since the main storage unit is subdivided, its results are inadequate for performing processing on programs that exceed the partition capacity.

③ Variable partitions method

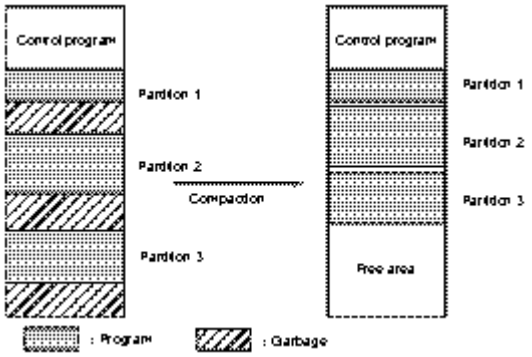
The variable partitions method is the method that sequentially assigns the area required by application programs in the program storage area. The main memory management program of the control program performs the area allotment.

Figure 3-1-15
Partition method



However, in these methods, areas that are not used (garbage) are generated in each partition of the main storage unit. This phenomenon is called fragmentation.
In order to solve this fragmentation, it is necessary to reset each partition at specific times or at specific intervals. This operation is called compaction (Figure 3-1-16).

Figure 3-1-16
Compaction

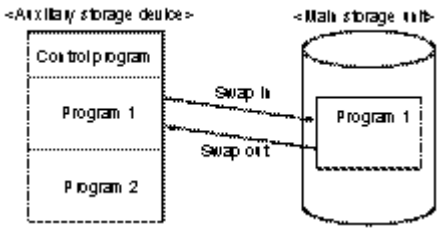


However, when compaction is performed, the address of each program instruction stored in the main memory unit changes. In order to solve this problem, it is necessary to reset and update the address of each instruction. This operation is called program relocation.

(2) Swapping

When multi-programming is performed in multiple partitions or other methods, if a job with high priority is generated, the job with low priority being executed has to be interrupted. In that case, in order to avoid letting the interrupted processing come to nothing, it is saved in an auxiliary storage device as it was when interrupted. This operation is called swap out (or roll out). On the other hand, the operation by which a job with high priority is transferred from an auxiliary storage device to the main storage unit is called swap in (or roll in). This kind of exchange of jobs between the main storage unit and the auxiliary storage devices is generically called swapping.

Figure 3-1-17
Swapping



(3) Overlay

Overlay is used for the execution of programs that are larger than the storage capacity of the partitions of the main storage unit.
Through the overlay technique, the application program is divided beforehand into units called segments, and after each of the segments is stored in the main storage unit, the program is executed.

(4) Memory protection

In order to avoid the misappropriation or destruction of the contents of the main storage unit, memory protection becomes necessary. Among the main memory protection methods, the following three methods can be mentioned.

① Boundary address method

The boundary address method is the method by which the address range that can be accessed is specified for each of the programs to be executed. The memory is protected by checking whether or not the access to the main storage unit to be executed is within the address range.

② Ring protection method

The ring protection method is the method by which a ring number is assigned to each program and the access is controlled according to the number size. A small number is assigned to important programs (OS, etc.) and a large number is assigned to user programs, etc. In this method, access from small numbers to large numbers can be performed, but in the opposite case, access can only be performed through service provision.

③ Keylock method

The keylock method is the method by which the main storage unit is divided into multiple partitions and each partition is locked for memory protection. Each program to be executed has its respective memory protection key(s) and access is authorized when the memory can be unlocked (the key and the lock match).

(5) Other main memory management

① Dynamic allocation

The dynamic allocation is the technique by which the main storage unit is dynamically assigned during the program execution.

② Memory leak

A memory leak occurs due to the failure to release the area that should have been released by a program that used the main storage unit, reducing as a consequence, the area of the main memory that is available for use. However, due to the volatility of the main storage unit, if the power is turned off, all the storage area is released. These kinds of events especially occur in servers, etc. that remain operational 24 hours a day.
Since the memory leak is not an event that occurs in all the OS, it is necessary to check the OS product information.

3.1.5 Virtual storage management

In the main storage unit, operations such as swapping and overlay become necessary in order to execute programs that are larger than the partition size of the storage area or to change the processing order. The development of programs under this kind of restriction can not be considered productive. For that reason, the approach of the virtual storage, which enables the execution of programs without worrying about the storage capacity of the main storage unit, was born.
The basic approach to implement virtual storage is as follows.

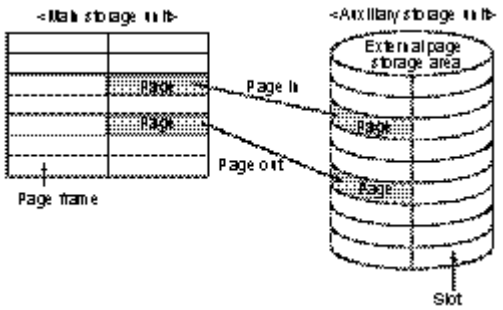
- The main storage unit is divided into partitions of a specific size. These partitions are called page frames.
- The program is temporarily stored in an area called the external page storage area of an external storage device.
- The external page storage area is divided into partitions called slots, which have the same size as the page frame. Therefore, the programs stored in the external page storage area are automatically divided into parts the size of a slot.
- The programs stored in the page frames or the slots are called pages. Generally the size of one page is 2 kilo bytes. Of course, the size of a page frame and the size of a slot are also 2 kilo bytes.
- The external page storage area of the main storage unit and the auxiliary storage device is called the logical address space.
- Among the programs stored in the external page storage area, the pages of the slots needed for execution are transferred to empty page frames of the main storage unit to be executed.

In this way, in the virtual storage method, execution is repeatedly performed by transferring the programs that are stored by page unit in the external page storage area to the page frames of the main storage unit. The act of transferring a program to the main storage unit is called load.

(1) Paging

The exchange of programs between the main storage unit and an auxiliary storage device is generically known as paging.
The transference of a slot from the external page storage area of the auxiliary storage device to the page frame of the main storage unit is called page in The transference in the opposite direction, when a page whose execution has been completed is transferred to the slot, is called page out.

Figure 3-1-18
Paging

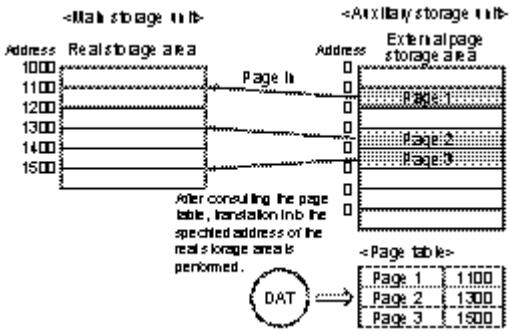


In the multi-programming method, there are cases where paging occurs frequently. This condition is called slashing.

(2) Address translation

An issue that often arises when paging is performed is that the page-in address of the main storage unit is unknown. Since in the virtual storage method, when a page frame becomes vacant, the next page to be executed is "paged in" to this page frame, it is necessary to translate the instruction address according to the address of the page frame. This conversion is called address translation.
The address assigned to each instruction of the programs stored in the external page storage area is called the static address and the address stored in the page frame of the main storage unit after the address translation is performed is called the dynamic address. The main address translation method is called the dynamic address translation (DAT), which is a method performed using the hardware.
The DAT performs address translation at the time the instruction paged in is executed. The addresses of the external page storage area start from address 0 and increase by page unit, while the addresses of the pages to be "paged in" are converted into dynamic addresses after consulting the page table.

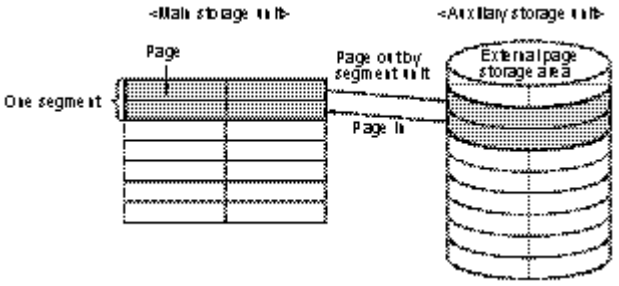
Figure 3-1-19
Address translation



(3) Segmentation paging

A group of pages logically related is called the segment. In segmentation paging, page in and page out are performed by these segments. Compared to the method in which paging is performed by page units, in this method paging occurs less frequently.

Figure 3-1-20
Segmentation paging



(4) Page replacement

In the page replacement, in order to achieve system processing efficiency, pages with a high application frequency are permanently stored in the main storage unit while pages with a low application frequency are stored in the external page storage area and are transferred to the main storage unit only when they are needed.
In this case, the following two methods are used to send out the pages from the main storage unit (paging algorithm).

① LRU (Least recently used) method

In the LRU method, among the pages of the page frame of the main storage unit, the page for which the time elapsed since it was used the last time is the longest is sent out.

② FIFO (First-in first-out) method

In the FIFO method, the page that was the first to be stored among the pages of the main storage unit, is sent out to the external page storage area.

3.1.6 File management

The data processed by the computer is controlled by the data management function of the operating system. Since most of the data is stored in auxiliary storage devices, file handling plays a central role in this data management. Therefore, it is also called file management.

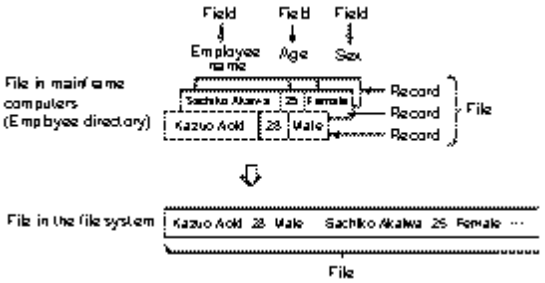
(1) File system

The concept of a file in personal computers and workstations differs from that of mainframe computers. The concept of low-end computer files, and the file system controlling those files, will be explained below.

① File concept and configuration in low-end computers

In personal computers or workstations, there are no concepts of records or fields as in the files handled in mainframe computers. Files simply record character strings, and there is no difference between data and programs.

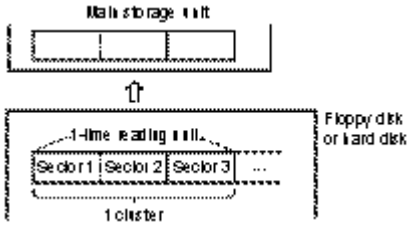
Figure 3-1-21 File concept in low-end computers



a. Cluster

The data sets composing a file are written on floppy disks (flexible disks) or hard disks in units called clusters. A cluster, which is a set of several sectors, is the input/output unit between these auxiliary storage devices and the main storage unit.

Figure 3-1-22
Cluster

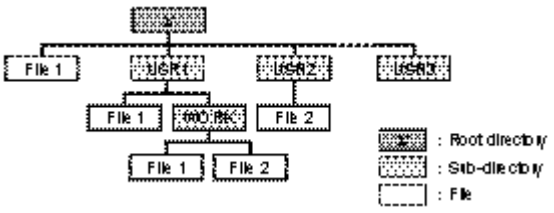


When data cannot be filed in one cluster, an unused cluster is coupled and the remaining data written in this cluster. When this operation is performed, it is not necessary for the coupled cluster to be a sequential cluster in the storage area.

b. Directory

The file system is composed of the directory and the file. The directory is the register where the file management information is recorded and stored. It is possible to have files and directories beneath a directory. The file system organizes these directories and files in a hierarchical structure to control them (Figure 3-1-23).
The highest directory of the hierarchical structure is called a root directory and it is an important directory in the aspect of volume. The directories positioned beneath the root directory are called sub-directories.

Figure 3-1-23 Hierarchical structure of the file system



② File operation

When personal computers and workstations start up, the directory is automatically set by the operating system. Normally, it is a root directory, but the user can freely set it.
It should be noted that the user has to move to the target directory in order to access a directory or file.

a. Directory

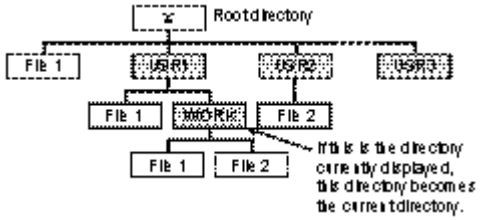
1) Home directory

The directory that can be freely used by the user is called the home directory. The user can freely create and access the sub-directories and files registered in the home directory.

2) Current directory

The current directory is the directory in current use. If the home directory is being used, the home directory will be the current directory.

Figure 3-1-24 Current directory



b. Path

When a file is sought inside the file system, the route along which to search for that file is specified. This route is called a path.

Depending on the specification method, paths are classified as follows:

- Absolute path
- Relative path

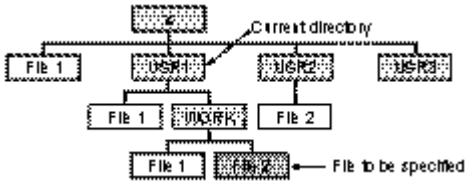
1) Absolute path

The absolute path is the path to the target directory or file from the root directory, which is at the highest position of the file system. In this specification method, all of the directories and files, from the root directory to the target directory or file, are written, using the ¥ sign or the / sign to separate them. In Figure 3-1-25, the absolute path to specify file 2 is as follows:

¥USR1¥WORK¥FILE2

The "¥" at the top represents the root directory.

Figure 3-1-25
Path specification



2) Relative path

The relative path is the path to the target directory or file from the current directory. In Figure 3-1-25, if the current directory is USR1, the relative path to specify file 2 would be as follows:

WORK¥FILE2

If the current directory is WORK, the relative path would simply be the following:

FILE2

Likewise, if the current directory is USR3, it is necessary to go up to the root directory once. Since ".." is used to specify a directory that is one level higher, the relative path would be as follows:

..¥USR1¥WORK¥FILE2

c. Command

In the operating systems of personal computers and workstations, programs are executed as a result of the input of commands. The commands related to file operations in the case of MS-DOS are shown in Figure 3-1-26.

Figure 3-1-26
Commands used to perform file operations

Commands	Functions
DIR	Display the file name(s) contained in the specified directory
DEL	Delete the specified file(s)
REN	Change the specified file name
TYPE	Display the file(s) content on the screen
COPY	Copy the file(s) in other directory or volume
PRINT	Print the specified file(s) content
MKDIR	Create a sub-directory (or subdirectories) beneath the current directory
RMDIR	Delete the specified sub-directory
CHDIR	Transfer the current directory to the specified directory

For example, when the file name "File 1" is changed to "File 2," the change command REN is entered after the input prompt (A>) as follows:

A>REN FILE 1 FILE 2

The "A" written before the prompt (>) indicates the location of the device subject to the operation. It is called the current drive.

d. Extension and wild cards

The file names used in the file system are expressed using a file name of 8 or less alphanumeric characters (○○○) and an extension of 3 or less alphanumeric characters (ΔΔΔ) separated by a period. "○○○. ΔΔΔ"
Among the extensions, there are some that are given a special meaning by the operating system, as well as some that are freely set by the users, or uniquely set by application software.
When a file is specified, the following wild cards (? or *) can take the place of file names and extensions:

- ?: Any character can be placed in the ? position. That means any single character.
- *: Any character(s) can be placed in the * (and subsequent) position(s). The character types or character string lengths are not specified.

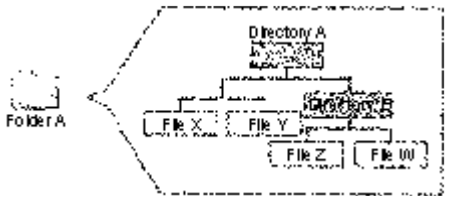
For example, the command, "Show all the files whose extension is BAK," would be specified as follows:

DIR*.BAK

e. File operations in GUI environments

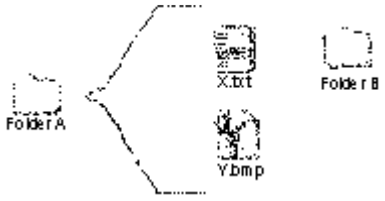
In the case of MS-DOS, files are operated using command operations consisting of character input. However, in recent years, GUI (Graphical User Interface) environments, in which icons on window screens are manipulated using a mouse, have become the mainstream.
In GUI environments such as MacOS and Windows, directories are represented by folder icons, and files are represented using icons set according to the extension. It should be noted that a user is free to design his/her own icons (Figure 3-1-27).

Figure 3-1-27
Folder and icon



Transfers to the current directory and access to files can be performed by clicking or dragging folders and icons with the mouse. To open a file means to display the contents of a folder on the screen. The screen displayed when Folder A of Figure 3-1-27 is opened is shown in Figure 3-1-28.

Figure 3-1-28
Opening Folder A



3.1.7 Security management

The protection of a computer and its resources from diverse menaces (natural disasters, failures, human errors and intentional malice) is called information security or computer security.
Security management aims at the achievement of the following three specific characteristics:

- Confidentiality
  Prevents the leakage of information contained in a computer due to illegal access, etc.
- Integrity
  Prevents the modification of information contained in a computer due to illegal access, etc.
- Availability
  Prevents the obstruction of the use (information reference or modification) by a legitimate user.

In general, the OS performs security control through access control and flow control.

① Access control

Access control is that which limits direct access to computer resources to legitimate users only.

② Flow control

Flow control is that which prevents the leakage of information to users that are not authorized as legitimate users.

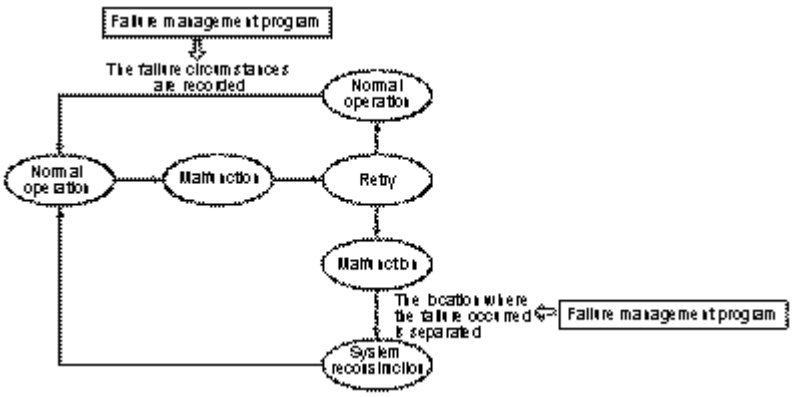
3.1.8 Failure management

As the system becomes larger, the impact of a failure becomes larger too.
Since a computer system is an extremely complex device, it is not easy to find what is wrong. In order to cope with these problems, the operating system is equipped with the following functions:

- Instruction retry
- Failure management program

The instruction retry is a function that automatically retries the execution of an instruction when a malfunction occurs, as is shown in Figure 3-1-29. If the computer operates normally after an instruction retry is performed, the failure management program records the failure circumstances. If a malfunction occurs again after an instruction retry is performed, the program isolates the location where the failure occurred and reconstructs the system. The failure circumstances record helps in selecting the check points to be emphasized at routine inspections, contributing to the MTTR (Mean time to repair) reduction as well.

Figure 3-1-29 Instruction retry and failure management program



3.1.9 Supervisor

The supervisor is a monitoring program functioning as the central part of the OS. It performs resource distribution and program control in order to implement the TSS, multi-programming, etc. The processing program sends an interrupt instruction called a supervisor call (SVC) or system call in order to request a special service from the supervisor. As a result of this instruction, the SVC interrupt is generated, the program (process) execution is temporarily interrupted and control is transferred to the supervisor.

3.2 Types of OS

Until the previous section, explanations were given with the general operating systems of mainframe computers in mind. Here, based on that knowledge, the operating systems that are actually widely used will be explained.

3.2.1 General-purpose OS

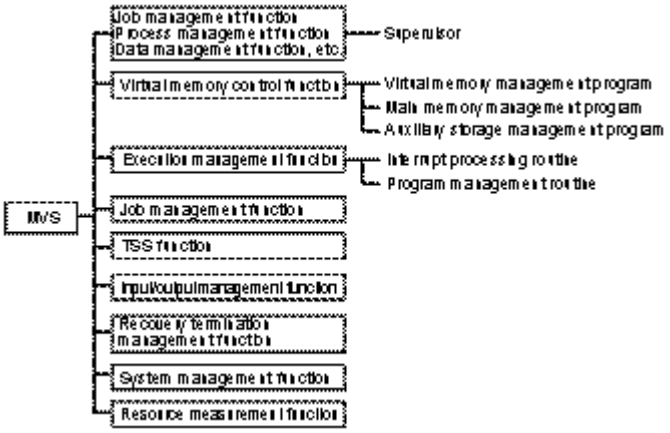
Diverse OSs are used in today's computers. As characteristics of OSs of recent years, the improvement of human interfaces using GUI, etc., the fulfillment of multimedia processing that enables easy use of audio and image data, etc., can be mentioned. Likewise, regarding the OS used in Japan, Japanese language processing functions have seen extreme improvement.
The main computer OS will be explained below.

(1) MVS (Multiple Virtual Storage)

The MVS is the most representative operating system adopted in high-end mainframe computers. Since this OS was developed by IBM for its own computers, it was introduced into the market as "MVS/370" in the 1970s, but was repeatedly upgraded thereafter. In MVS, 32-bit words are the foundation for everything.

① MVS configuration

Figure 3-2-1
MVS configuration



In MVS, besides the conventional logical address space, a data space and a hyper space, where only data is stored, are also provided. This was set up in order to reduce input/output frequency. In this system, data used frequently is stored in advance in the data space and hyper space, and the programs of the logical address space directly check the data space.

② MVS characteristics

The following points can be mentioned as MVS characteristics:

- It is the operating system for high-end general-purpose computers.
- It provides a multi-user function, which enables simultaneous MVS use by multiple users.
- It provides a multi-task function, which enables simultaneous processing of multiple tasks.
- It adopts the approach of multiple address space.
- One logical address space reaches up to 2 gigabytes.
- It has all the file organization functions.

(2) UNIX

UNIX is an operating system developed by AT&T Bell Laboratories, widely used in computer network systems that have personal computers, workstations, etc. connected by telecommunication lines. Since Version 1.0 was launched in 1969, it has since been upgraded in diverse ways.
The most distinctive characteristic of UNIX is that, unlike other operating systems, detailed contents of the operating system written in C language have been released to the public. For that reason, a great number of computer manufacturers, besides AT&T Bell Laboratories, can easily port it to the hardware of their own products. As a result, users are able to operate UNIX in all computer manufacturer products.
The following can be mentioned as the most representative examples of UNIX upgraded editions:

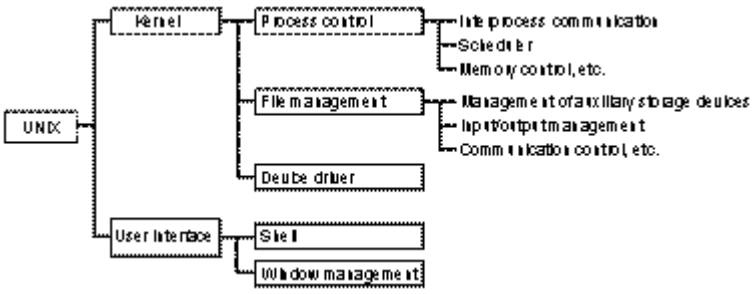
- XENIX (Microsoft)
- AIX (IBM)
- Ultlix (DEC)
- SunOS, Solaris (Sun Microsystems)

Even though their names differ, they are basically UNIX operating systems.

① UNIX configuration

UNIX is an operating system which can be simultaneously used by multiple users, and in which each user can simultaneously perform multiple job processing. Its configuration is shown in Figure 3-2-2.

Figure 3-2-2
UNIX configuration



UNIX has a control program called Kernel, which has the following functions:

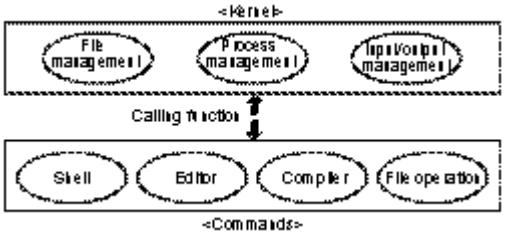
- It is the central part of the operating system, and controls the system resources.
- It performs the process management. (Since UNIX is distributed-processing oriented, jobs are called processes.)

Likewise, there are command sets that directly instruct jobs to UNIX. These command sets are composed of shells and commands. A shell has the following functions:

- It interprets the commands input by the users and calls the kernel function.
- It executes a program called a shell script, which combines commands.

Likewise, the command uses the devices connected to the system, and can call the kernel function to use the command sets.

Figure 3-2-3
Kernel and command



② UNIX characteristics

UNIX is an operating system which can perform distributed processing in computer network environments. Its characteristics are mentioned below.

- Distributed processing is presupposed.
- It was developed assuming that it would be used in workstations.
- It provides the multi-user function, which enables simultaneous use by multiple users. In operating systems that provide the multi-user function, a procedure called log-in, to receive the service, and a procedure called log-out, to report completion, are required.
- Through multi-programming, it can simultaneously process multiple jobs. In UNIX, this is called the multiprocessing function.
- As a technology to connect computers of different manufacturers, the communication protocol called TCP/IP has been established.
- It has instruction rights called commands that enable the user to use UNIX in an interactive mode. This function is called human interface, and has been implemented by X-Window.
- The program development tools are abundant.

(3) Windows

The operating system of more than half of the personal computers around the world is Windows.
There are the following Windows versions:

- Windows
- Windows 95
- Windows 98
- Windows NT
- Windows 2000
- Windows ME

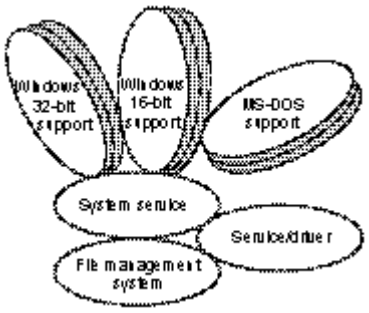
① Windows history

The first personal computers had 16-bit words, and IBM personal computers, called PCs, were the mainstream. The operating system adopted for these PCs was MS-DOS with single-task functions, developed by Microsoft.
Afterwards, with the appearance of 32-bit-word personal computers, Windows was born.
Inheriting the MS-DOS functions without changing them, Windows fulfilled the GUI environment and had outstanding operability, therefore becoming a worldwide best seller. (It is said that in the U.S., alone, 2 million sets were adopted.) However, since it inherited the basic concept of MS-DOS, it was not able to master the hardware functions of 32-bit words.
As a result, in 1995 Microsoft introduced into the market a new operating system that fulfilled multimedia functions, communication functions and network functions, while inheriting the unchanged concept of Windows. This operating system is Windows 95.
On the other hand, Windows NT was developed completely independently, without inheriting the restrictions of past operating systems. It is used as the server operation system in client/server systems, etc. (NT are the initial letters of New Technology). It should be noted that Windows NT upgrades, up to Version 4.0, have been put on the market, and Version 5.0, launched in 1999, was named Windows 2000.

② Windows 95 configuration

Windows 95 is an operating system that capsules MS-DOS. In the time since the file system was newly created, constraints have been substantially reduced.
Figure 3-2-4 shows an image of the configuration of Windows 95. However, since the functions of Windows are included, it still maintains the 16-bit-word MS-DOS environment.

Figure 3-2-4
Windows 95 configuration



③ Windows 95 characteristics

Since an outstanding GUI environment is provided, and it is widely used around the world, in practice Windows 95 has become a *de facto* standard.
GUI characteristics are mentioned below.

- Desktop approach
  Considering the display screen as one desk, screens can be used as though several documents were spread over the desk. These screens are called windows.
- Multi-task function
  Not only can multiple windows be simultaneously displayed on the display screen, a multi-task function is also provided.
- Mouse pull down menu/dialog box manipulation
  With one mouse, a great variety of menus can be selected/operated.

④ Windows 98

As the expanded version of Windows 95, Windows 98 is an operating system that incorporates the following:

- Fulfillment of Internet software
- Support of new interfaces such as USB and IEEE 1394

After the first version was launched in 1998, a second version was launched in 1999. It allows for possible use of the same peripheral devices in order to share the device driver with Windows 2000.

(4) MacOS

MacOS is the OS developed by Apple Computer for its own product (Macintosh)
and:

- Almost all the operations can be performed with the mouse
- The operation method of application software is almost the same, etc.

The user interfaces are abundant. For that reason, it is said that MacOS is an OS that is easy for beginners to use.
The OS "MacOS X," for the client, and the OS "MacOS X Server," for the server, launched in 1999, integrate the former "MacOS 8" and "Rhapsody," and was announced as a new OS.

(5) Linux

Linux is the UNIX-based OS launched in 1991 by Linus Torvalds. The main characteristic is that the software is free. Since the source code has been released to the public, and redistribution and changes can be freely performed, a great number of people around the world have participated to make Linux a better OS. As a background factor, it should not be forgotten that the Internet expansion that enabled people around the world to communicate with each other allowed this participation.
It should be noted that the copyright is protected by GPL (GNU Public License).

3.2.2 Network OS (NOS)

The network OS is the OS used to construct LANs, in which computers are connected and used through a network. Besides providing the same services as a computer OS, based on the SNMP (Simple Network Management Protocol), it provides network management functions.
As the main network OS, NetWare and LAN manager will be explained.

(1) NetWare

NetWare is a network OS developed by Novell. It is the most common NOS, with file sharing and printer sharing functions.

(2) LAN manager

The LAN manager is a network OS developed jointly by Microsoft and 3 Com. The functions of this network OS were inherited from Microsoft's OS "Windows NT."

3.3 Middleware

Middleware is positioned between basic software and application software. This software provides the basic processing functions that are used in common by users.
Among main middleware, the following, whose applications are diverse, can be mentioned:

- DBMS (Database management system)
- Communication management system
- Software development support tool
- Operation management tool
- ORB
- Japanese word processor
- Spreadsheet software
- Graphic processing system

3.3.1 DBMS

DBMS (Database Management System) is dedicated software aimed at efficient database creation/maintenance/operation. The following are the three main characteristics:

① Integrity

Even when the database is simultaneously used by multiple users, it prevents the generation of data inconsistency.

② Security

It protects data secrecy by setting database access rights, etc.

③ Failure recovery

In the event that a failure occurs in a database, it promptly recovers that database.

3.3.2 Communication management system

The communication management system is software aimed at supporting computer network construction/operation. A recent tendency in software of this kind is to emphasize LAN control. The following are the three main characteristics of the communication management system.

① Network independence

In order to facilitate network construction, lines, communication equipment, and other network environments are separated from user programs.

② Network flexibility

Through the provision of flexibility to the devices and network mechanism that make up the network, the construction of network systems with high expandability is enabled.

③ Network transparency

This provides an environment in which network users can use the system without being aware of the network.

3.3.3 Software development support tool

A software development support tool is software that supports computer-aided software development. As software development support tools aimed at achieving development labor saving as well as quality improvement, CASE (Computer Aided Software Engineering) tools exist. Depending on the content supported, CASE tools are classified as follows:

① Upstream CASE tools

Upstream CASE tools support the high-end process (analysis, design, etc.) of software development.

② Downstream CASE tools

Downstream CASE tools support the lower-end process (programming, testing, etc.) of software development.

③ Maintenance CASE tools

Maintenance CASE tools support the operation and maintenance of the developed software.

④ Integrated CASE tools

Integrated CASE tools support overall functions from upstream CASE tools to maintenance CASE tools.

3.3.4 Operation management tool

An operation management tool is software aimed at supporting the operation duties of system operation managers. Among the operation management characteristics, the following can be mentioned:

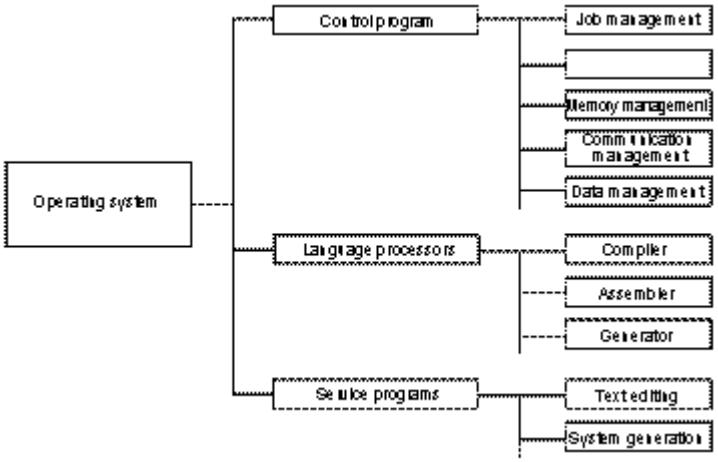
- Multiplicity optimization of multi-programming
- Allocation of system resources in order to reduce response time at peak periods
- Grasp of the operating conditions of system resources
- Recording of accounting information and creation of summaries
- Logging of operation records

3.3.5 ORB

ORB (Object Request Broker) is software used for the creation of object requests and responses, as well as for communication between objects in object environments. CORBA (Common Object Request Broker Architecture), which was completed as a standard specification by the Object Oriented Management Group (OMG), is among the most representative software of this kind.

Exercises

Q1 **The following diagram shows the relation of some of the functions of the operating system. Which is the appropriate function to fill in the blank?**

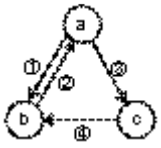


A. Overlay management B. Catalog management C. Process management
D. Project management E. Message management

Q2 **Which of the following is the most appropriate description of spooling?**

A. Provides a standard communication procedure regardless of the other devices and the communication network.
B. Using an external storage device, provides a virtual memory larger than the main storage unit.
C. Records the information related to the computer system operation process.
D. The operation of peripheral devices is separated and performed in parallel to the processor operation.
E. Enables processing on a logical record basis without having to worry about the physical record.

Q3 **The following diagram shows the process state and transition. Which of the following is the correct combination of a, b, and c states?**



The causes of status transition:
① **The CPU use right was transferred to a process with a high execution priority.**
② **The CPU use right was provided.**
③ **Wait for the completion of the input/output operation.**
④ **The input/output operation has been completed.**

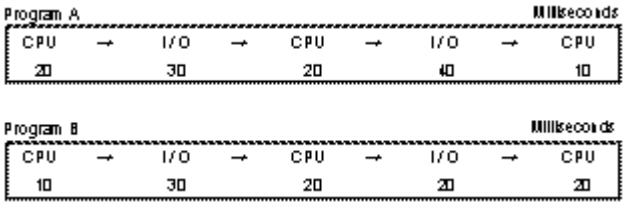
	a	b	c
A	Executable status	Execution status	Wait status
B	Executable status	Wait status	Execution status
C	Execution status	Executable status	Wait status
D	Execution status	Wait status	Executable status

Q4 **Which of the following corresponds to the cause of internal interrupt?**

A. Occurrence of anomalies in the computer power-supply unit
B. The counter that measures clock time inside the processor has exceeded the preset value
C. Input/output device operation completion or failure occurrence
D. Occurrence of overflow in floating point operations

Q5 **Given the two programs, A and B, the occupancy time of the processor (CPU) and the input/output devices (I/O) when each program is executed separately is shown in the diagram. Considering that programs A and B are started simultaneously in the same CPU, how many milliseconds after the startup will program B be completed? The program execution conditions, etc., are as follows:**

① **A has a higher program execution priority than B.**
② **Programs A and B use the same input/output devices.**
③ **The execution of programs in the CPU is not interrupted until the input/output processing is started.**
④ **The execution of input/output processing in input/output devices is not interrupted until it is completed.**
⑤ **The time needed for the CPU task switching can be ignored.**



A. 120 B. 140 C. 160 D. 180

Q6 **Which of the following is used in process mutual exclusion (exclusive control)?**

A. Contention B. Semaphore C. Check point D. Hash

Q7 **In the operating system, a large number of small unused portions in the memory result from the repetition of the allocation and release of the memory space. What is the name of this phenomenon?**

A. Compaction B. Swapping C. Fragmentation D. Paging

Q8 **Which is the processing that transfers a program being executed to an auxiliary storage device in order to load and execute a program with a higher priority level?**

A. Overlay B. Swapping C. Paging D. Relocation

Q9 **Which is the method that divides the storage space into specific sizes, manages it, and implements virtual storage?**

A. Slashing B. Swapping C. Blocking D. Paging

Q10 **Which of the following is the most suitable explanation of dynamic address translation?**

A. It is the translation of virtual addresses into real addresses in the virtual storage system.
B. It is the act of changing the base address of a program being executed in order to transfer and execute it in a new location.
C. It is the vicarious execution of the main memory reading and writing by the cache memory.
D. It is the act of resolving address references between modules in order to add a module during the execution of a program.

Q11 **Which of the following is the explanation of the LRU, which is one of the page replacement algorithms of the virtual memory?**

A. The page with the lower priority according to a priority level established in advance is expelled.
B. The page whose period of existence in the main storage unit is the longest is expelled.
C. The page whose period of existence in the main storage unit is the shortest is expelled.
D. The page that has not been referenced for the longest period is expelled.

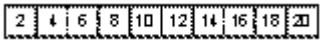
Q12 **Which is the most suitable explanation of indexed sequential organization, which is one the file organization methods?**

A. Direct access to the records can be performed using the address of each record. Sometimes, the efficiency of the medium use is low.
B. The records are recorded in the order in which physical writing is performed. Only sequential access can be performed.
C. It is composed of a data area called member and a directory area that controls the member information. It is suitable for storing programs.
D. It is composed of an area to store the records and an area to store the record key information.

Q13 **Considering that 10 records, whose keys are the numbers shown in the figure, are to be stored in direct organization files, if a division method in which 7 is the divisor is used as the hashing (address translation) method, how many records would be synonym records? It should be noted that in hashing using a division method,**

Key value ¸ Divisor = X with the remainder Y

Y is the record address.



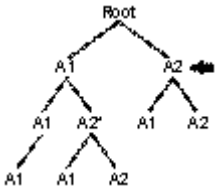
A. 1 B. 2 C. 3 D. 4 E. 5

Q14 **Given an operating system performing file management using a directory with a hierarchical structure, which of the following is specified to indicate the directory where the file is located?**

A. Extension B. Sub-directory C. Path
D. Root directory E. Wild card

Q15 **Directories A1 and A2 are managed with the structure shown in the diagram. In each directory a file, f, exists. Which is the method to specify the file, f, located beneath the directory pointed with an arrow, from the directory with the asterisk (current directory)? Here, the file specification method is based on the following:**

① **The directories on the route are sequentially specified, separating them with "¥," specifying the file in the following way.**
"**DIRECTORY NAME¥…¥ DIRECTORY NAME ¥FILE NAME**"
② **The current directory is represented with ".".**
③ **The directory that is one level higher is represented with ".. ".**



A. .¥A2¥f B. ..¥..¥A2¥f C. ..¥A1¥..¥A2¥f D. ..¥A2¥f

Q16 **Considering a character string composed of multiple alphabetic characters and 1 delimiter ". ", if "*" represents any character string larger than 0, and "?" represents 1 character, which of the character strings corresponds to the representation shown below?**
X*.Y??

A. XY.XYY B. XXX.YY C. XYX.YXY D. YXYX.YXY

Q17 **Which of the following is <u>not</u> a correct explanation of UNIX which is one of the operation systems (OS)?**

A. Provides an interactive human interface that uses character-based commands.
B. Since its specifications have been released to the public and it has a high portability, it has been adopted in a wide range of devices.
C. It is a single-user and multi-task OS.
D. Provides network functions that easily implement distributed processing.
E. It is the most representative workstation OS.