

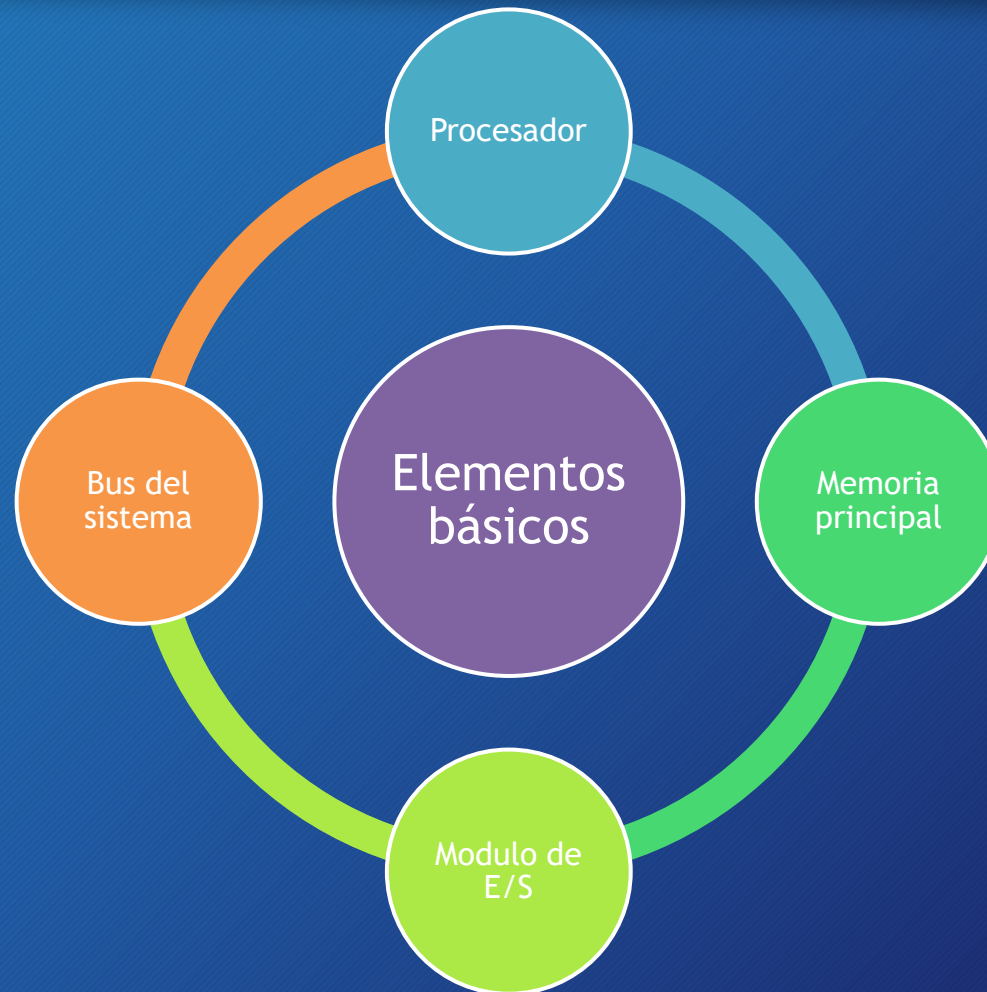
SISTEMAS OPERATIVOS

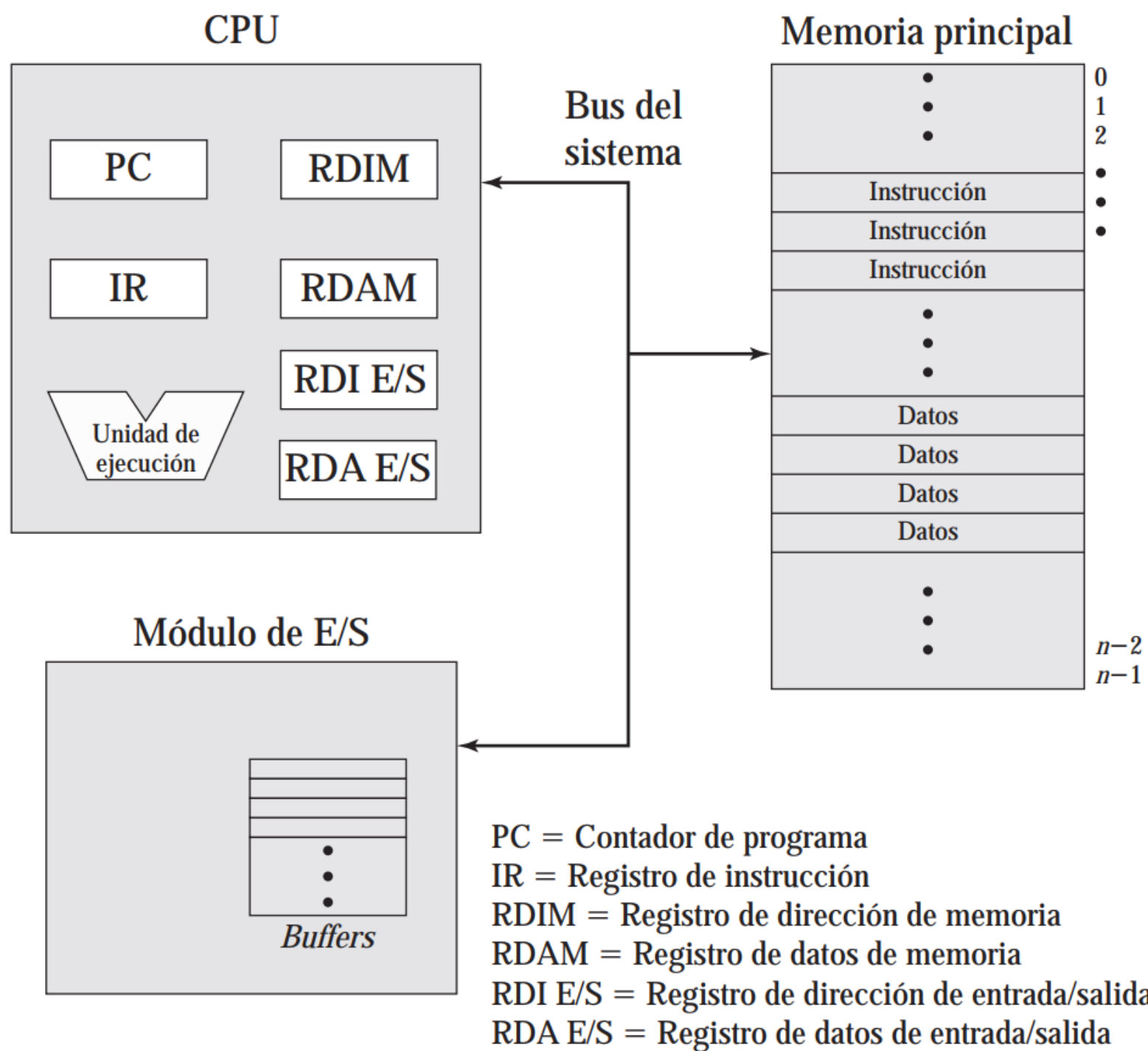
Ingeniería civil informática

Gonzalo Carreño

gonzalocarrenob@Gmail.com

Elementos básicos





MULTIHILO Y MULTINÚCLEO

- Con el veloz crecimiento en la capacidad de producción de chips pequeños y de mayor poder fue posible multiplicar las capacidades de las CPU.
- El paso inicial fue multiplicar no sólo las unidades funcionales, sino, parte del sistema de control.
- A partir del Pentium 4 nace la tecnología "multihilo" (multithreading). Esto permite que la CPU contenga el estado de 2 hilos de ejecución y alterne entre ellos, esto simula un paralelismo ya que la alternancia es en nanosegundos.

- Para el S.O. las capacidades multihilo tiene consecuencias directas ya que cada hilo es visto por el S.O. como una CPU individual, lo que implica que el S.O. debe tener capacidad de decidir que CPU activar para cada trabajo a procesar.
- Más allá del multihilo, están las CPU multinúcleo, chips con 2,4 o más procesadores completos. Para poder aprovechar estas capacidades el S.O. debe ser multinúcleo

MEMORIA

- Como segundo componente importante dentro de una computadora está la memoria.
- En teoría la memoria es muy rápida (más rápida que la ejecución de una instrucción) de gran tamaño y económica. (en teoría)

Tiempo de acceso típico		Capacidad típica
1 nseg	Registros	<1 KB
2 nseg	Caché	4 MB
10 nseg	Memoria principal	512-2048 MB
10 mseg	Disco magnético	200-1000 GB
100 seg	Cinta magnética	400-800 GB

DISPOSITIVOS E/S

- Otro recurso del sistema que deben ser administrados por el S.O. son los elementos de E/S. Estos constan de dos partes. El controlador y el dispositivo en sí.
- Cada dispositivo (controlador) requiere software para "entenderse" con el chip, esto es lo que conocemos como driver.
- El driver es un puente de software entre el controlador y el S.O. Cada fabricante de dispositivos tiene que suministrar un driver específico para cada S.O.
- El driver funciona en modo Kernel.

TIPOS DE S.O.

- Los S.O. llevan en funcionamiento, desarrollo más de 50 años, en ese tiempo se han creado diversos tipos de ellos. S.O. Mainframe: Enormes equipos computacionales, miles de discos y millones de Gb de datos (ej: Servidores web de alto rendimiento). Estos S.O. están orientados al procesamiento de muchos trabajos al unísono con muchas operaciones de E/S.

- S.O. para servidores: Utilizados en servidores, computadores personales grandes, estaciones de trabajo. Dan servicio a varios usuarios a la vez, permitiendo compartir los recursos de hardware y software. Algunos S.O: Solaris, FreeBSD, GNU/Linux, Windows Server 200x
- S.O. multiprocesadores: Un medio de obtener poder de computo es conectar varias cpu "en línea" como si de un sólo sistema se tratase (Sistemas en paralelo, multicomputadores, etc). Utilizan variaciones de los S.O. de servidores que permiten la comunicación y consistencia de trabajos y datos.

- S.O. para pc: Corresponden a los S.O. que usamos todos los días.
- Su misión es dar buen soporte al usuario, procesamiento de texto, cálculo y acceso a internet. ej: GNU/Linux, FreeBSD, Windows 7, OS X.
- S.O móviles: Son sistemas normalmente basados en los de pc pero optimizados para trabajar en dispositivos pequeños, de bajo consumo y con funciones restringidas.

- S.O. integrados: Los sistemas integrados (embedded) operan en computadoras que controlan dispositivos como: hornos microondas, televisores, PLC, etc Todo el software se encuentra en formato ROM. ej: QNX, VxWorks
- S.O. en tiempo real: Sistemas que tiene el tiempo como elemento clave: Sistemas de control de aviones, procesos industriales, etc.
- A menudo estos S.O. son básicamente una librería enlazadas con los programas de aplicación con gran acoplamiento, ej: e-Cos

CONCEPTOS DE LOS S.O.

- Algunos conceptos básicos y abstracciones utilizados en todo S.O. ayudan a comprender su funcionamiento.
- Procesos: Es en esencia un programa en ejecución. Cada procesos tiene asociado un espacio de direcciones y una lista de ubicaciones de memoria que el proceso puede leer/escribir.
- También se le asocian algunos recursos que comúnmente incluye registros, lista de archivos abiertos, alarmas, procesos relacionados, etc.
- Un proceso es un recipiente que guarda toda la información necesaria para ejecutar un programa.

Registros del procesador

Visibles para el usuario

Permiten al programador minimizar las referencias a memoria principal optimizando el uso de registros.

De control y estado

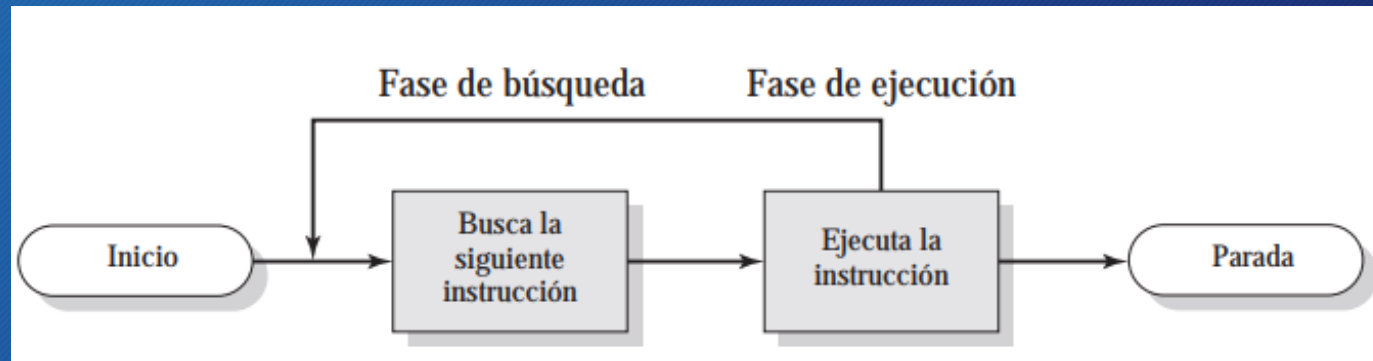
Usados por el procesador para controlar su operación y por rutinas privilegiadas del sistema operativo para controlar la ejecución de programas.

Ejecución de instrucciones

- Un programa que va a ejecutarse en un procesador consta de un conjunto de instrucciones almacenado en memoria. En su forma más simple, el procesamiento de una instrucción consta de dos pasos: el procesador lee (busca) instrucciones de la memoria, una cada vez, y ejecuta cada una de ellas. La ejecución del programa consiste en repetir el proceso de búsqueda y ejecución de instrucciones. La ejecución de la instrucción puede involucrar varias operaciones dependiendo de la naturaleza de la misma.

Ciclo de instrucción básico

- Se denomina ciclo de instrucción al procesamiento requerido por una única instrucción. En la Figura se describe el ciclo de instrucción utilizando la descripción simplificada de dos pasos. A estos dos pasos se les denomina fase de búsqueda y de ejecución. La ejecución del programa se detiene sólo si se apaga la máquina, se produce algún tipo de error irre recuperable o se ejecuta una instrucción del programa que para el procesador.



Búsqueda y ejecución de una instrucción

- La instrucción leída se carga dentro de un registro del procesador conocido como registro de instrucción (IR). La instrucción contiene bits que especifican la acción que debe realizar el procesador.
- El procesador interpreta la instrucción y lleva a cabo la acción requerida. En general, estas acciones se dividen en cuatro categorías:

Categorías

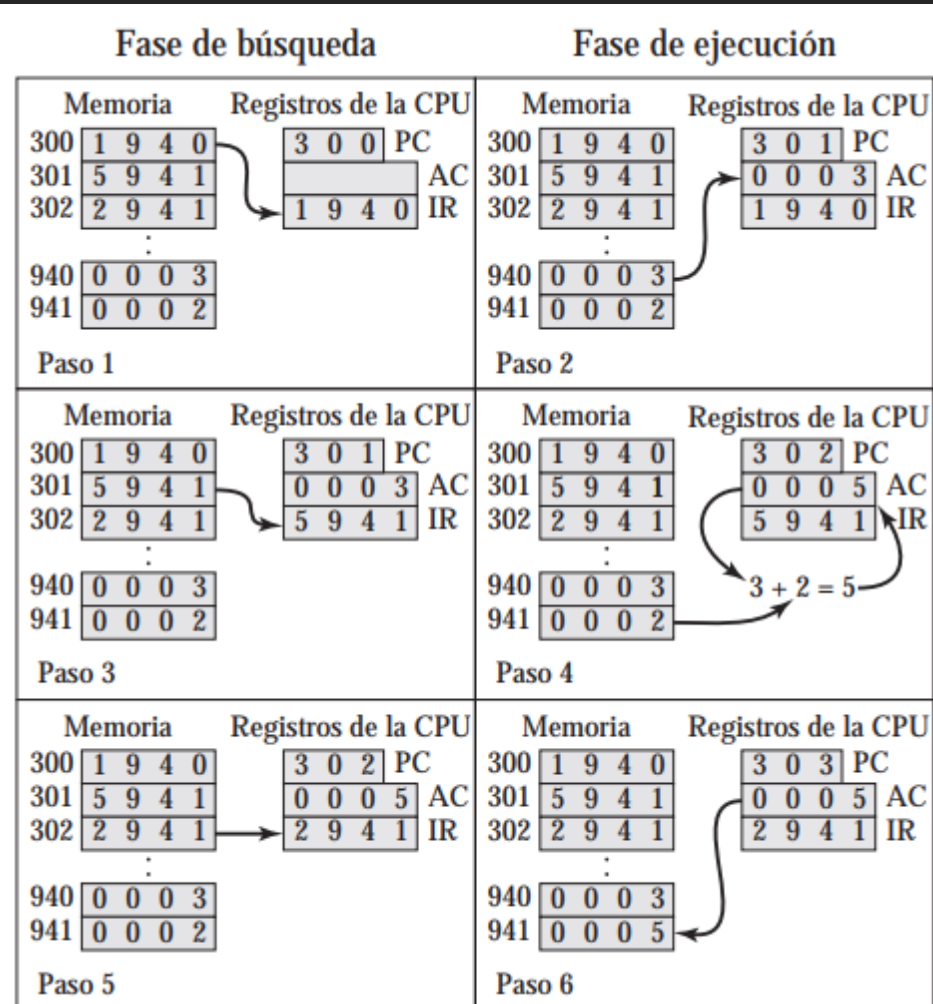
Procesador - Memoria

Procesador - E/S

Procesamiento de datos

Control

Ejemplo



Contador de programa (PC) = Dirección de la instrucción
 Registro de instrucción (IR) = Instrucción que se está ejecutando
 Acumulador (AC) = Almacenamiento temporal

Pasos

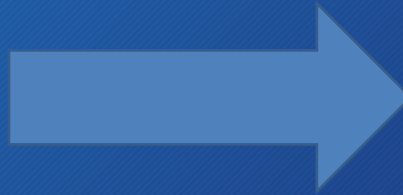
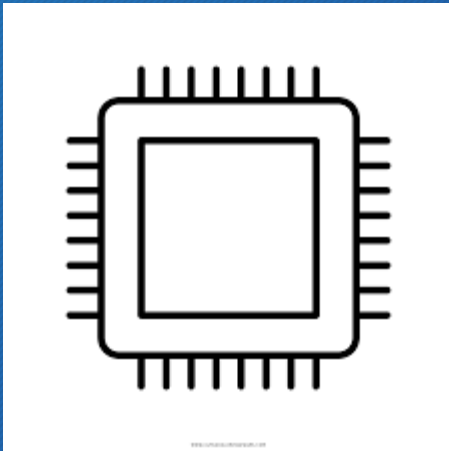
- 1. El PC contiene el valor 300, la dirección de la primera instrucción. Esta instrucción (el valor 1940 en hexadecimal) se carga dentro del registro de instrucción IR y se incrementa el PC. Nótese que este proceso involucra el uso del registro de dirección de memoria (RDIM) y el registro de datos de memoria (RDAM). Para simplificar, no se muestran estos registros intermedios.
- 2. Los primeros 4 bits (primer dígito hexadecimal) en el IR indican que en el AC se va a cargar un valor leído de la memoria. Los restantes 12 bits (tres dígitos hexadecimales) especifican la dirección de memoria, que es 940.
- 3. Se lee la siguiente instrucción (5941) de la posición 301 y se incrementa el PC.
- 4. El contenido previo del AC y el contenido de la posición 941 se suman y el resultado se almacena en el AC.
- 5. Se lee la siguiente instrucción (2941) de la posición 302 y se incrementa el PC.
- 6. Se almacena el contenido del AC en la posición 941.

SISTEMA DE E/S

- Se pueden intercambiar datos directamente entre un módulo de E/S (por ejemplo, un controlador de disco) y el procesador. Al igual que el procesador puede iniciar una lectura o una escritura en memoria, especificando la dirección de una posición de memoria, también puede leer o escribir datos en un módulo de E/S. En este caso, el procesador identifica un dispositivo específico que está controlado por un determinado módulo de E/S. Por tanto, podría producirse una secuencia de instrucciones similar a la del ejemplo anterior, con instrucciones de E/S en vez de instrucciones que hacen referencia a memoria.

INTERRUPCIONES

- Básicamente, las interrupciones constituyen una manera de mejorar la utilización del procesador.
- Por ejemplo, la mayoría de los dispositivos de E/S son mucho más lentos que el procesador.

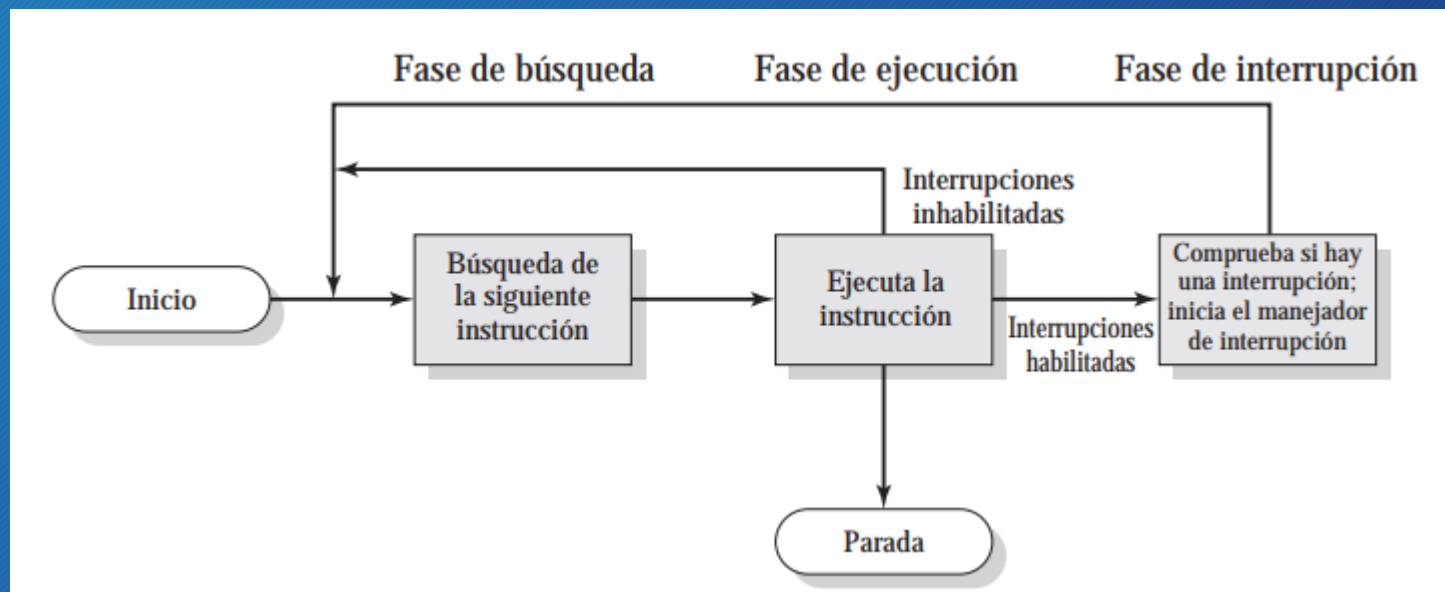


Clases de interrupciones

De programa	Generada por alguna condición que se produce como resultado de la ejecución de una instrucción, tales como un desbordamiento aritmético, una división por cero, un intento de ejecutar una instrucción de máquina ilegal, y las referencias fuera del espacio de la memoria permitido para un usuario.
Por temporizador	Generada por un temporizador del procesador. Permite al sistema operativo realizar ciertas funciones de forma regular.
De E/S	Generada por un controlador de E/S para señalar la conclusión normal de una operación o para indicar diversas condiciones de error.
Por fallo del hardware	Generada por un fallo, como un fallo en el suministro de energía o un error de paridad en la memoria.

Ciclos de interrupciones

- Cuando el dispositivo externo está listo para ser atendido, es decir, cuando está preparado para aceptar más datos del procesador, el módulo de E/S de este dispositivo externo manda una señal de petición de interrupción al procesador. El procesador responde suspendiendo la ejecución del programa actual, saltando a la rutina de servicio específica de este dispositivo de E/S, conocida como manejador de interrupción, y reanudando la ejecución original después de haber atendido al dispositivo.



En síntesis

- Las interrupciones son eventos que son atendidos por la CPU en un momento determinado. Puede tener múltiples orígenes:
 - A nivel de componente de hardware
 - A nivel de componente de software

¿Qué es lo que pasa?

La CPU detiene el programa que esta ejecutando en un momento determinado, por que necesita atender el evento que acaba de recibir.

Luego la CPU ejecuta una rutina de atención a interrupciones

Luego la CPU retorna al programa que ha sido detenido

MULTIPROGRAMACIÓN

- Incluso utilizando interrupciones, puede que el procesador siga sin utilizarse eficientemente. Si el tiempo requerido para completar una operación de E/S es mucho mayor que el código de usuario entre las llamadas de E/S (una situación habitual), el procesador estará parado la mayor parte del tiempo.
- Una solución a este problema es permitir que múltiples programas de usuario estén activos al mismo tiempo.

LA JERARQUÍA DE MEMORIA

- ¿cuál es su capacidad? ¿Cuál es su velocidad? ¿Cuál es su coste?
- Como se podría esperar, hay un compromiso entre las tres características fundamentales de la memoria: a saber, coste, capacidad y tiempo de acceso. En cualquier momento dado, se utilizan diversas tecnologías para implementar los sistemas de memoria. En todo este espectro de tecnologías, se cumplen las siguientes relaciones:
 - Cuanto menor tiempo de acceso, mayor coste por bit.
 - Cuanto mayor capacidad, menor coste por bit.
 - Cuanto mayor capacidad, menor velocidad de acceso.

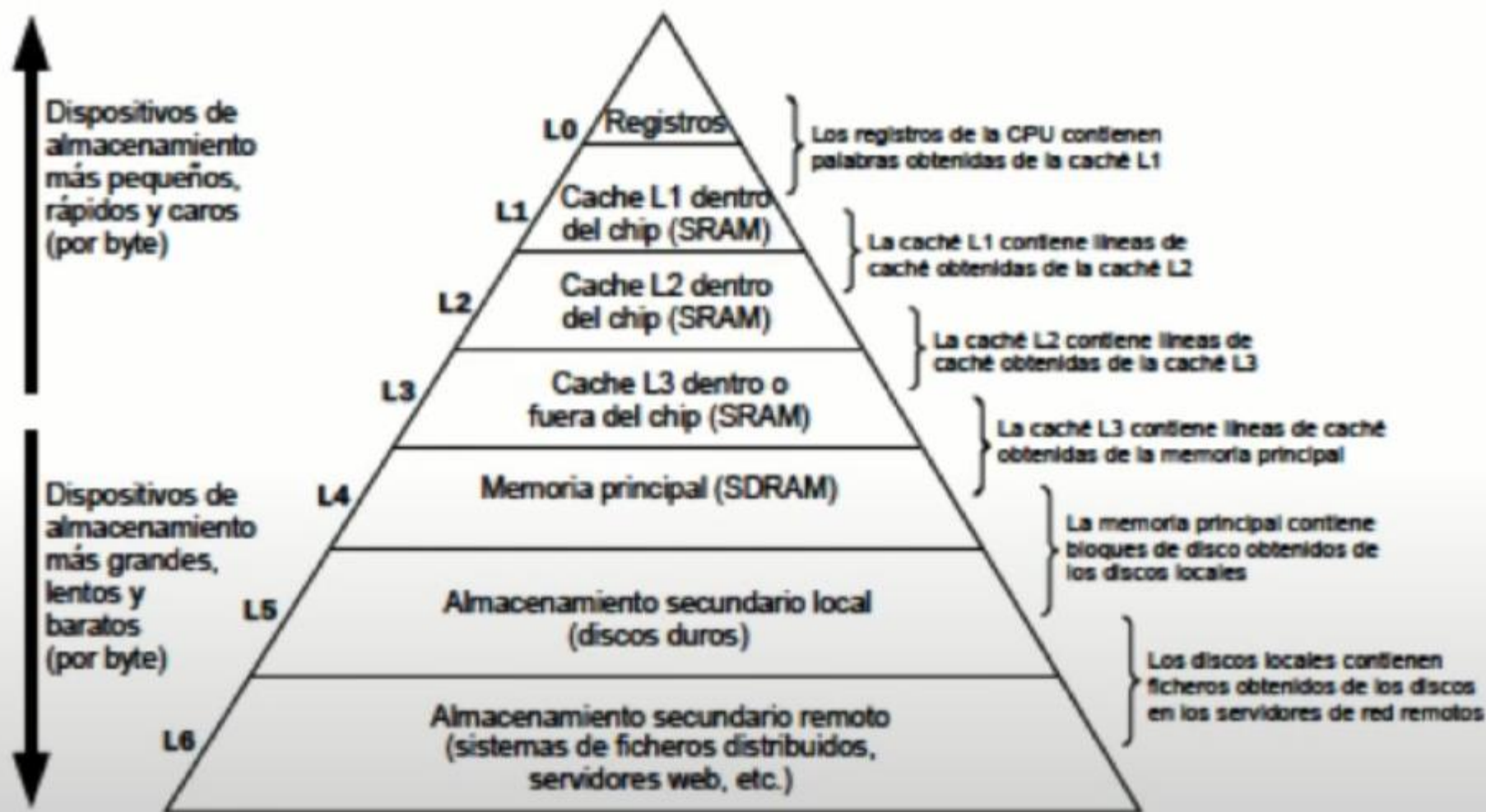
Tecnología	Tiempo de acceso típico	€ por MB
SRAM	1 ns	20 €
SDRAM	5 ns	0,01 €
Disco magnético	8.500.000 ns	0,0001 €

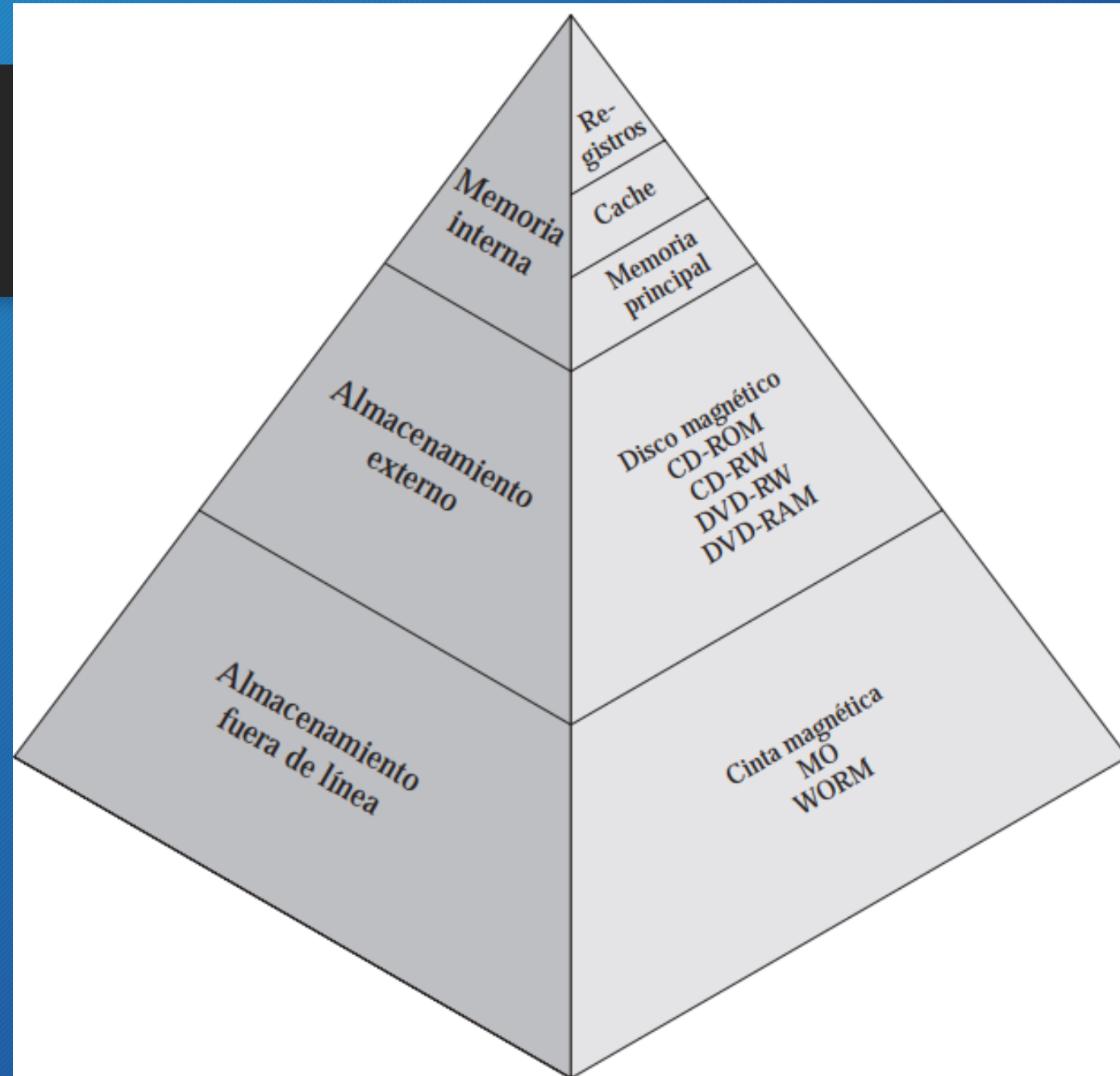
⇒ \$18,300
⇒ \$9
⇒ \$0,092

- Las 3 relaciones presentan un dilema.
- La solución a este dilema consiste en no basarse en un único componente de memoria o en una sola tecnología, sino emplear una jerarquía de memoria. Según se desciende en la jerarquía, ocurre lo siguiente:
 - a) Disminución del coste por bit.
 - b) Aumento de la capacidad.
 - c) Aumento del tiempo de acceso.
 - d) Disminución de la frecuencia de acceso a la memoria por parte del procesador.

Memoria grande y a costo razonable

La memoria del computador está organizada como una jerarquía de memorias

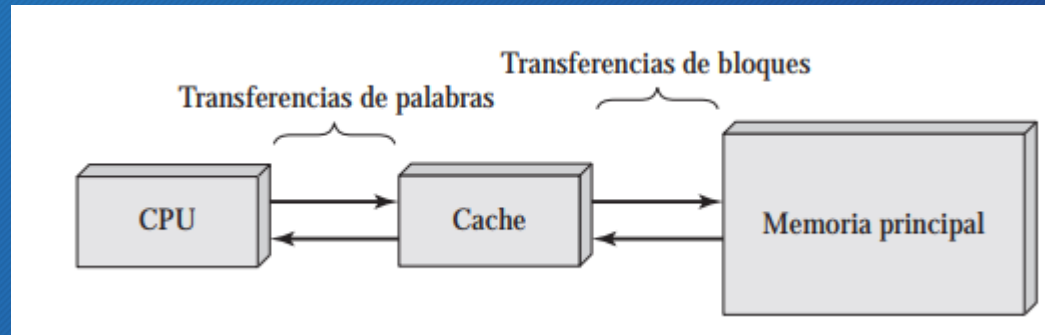




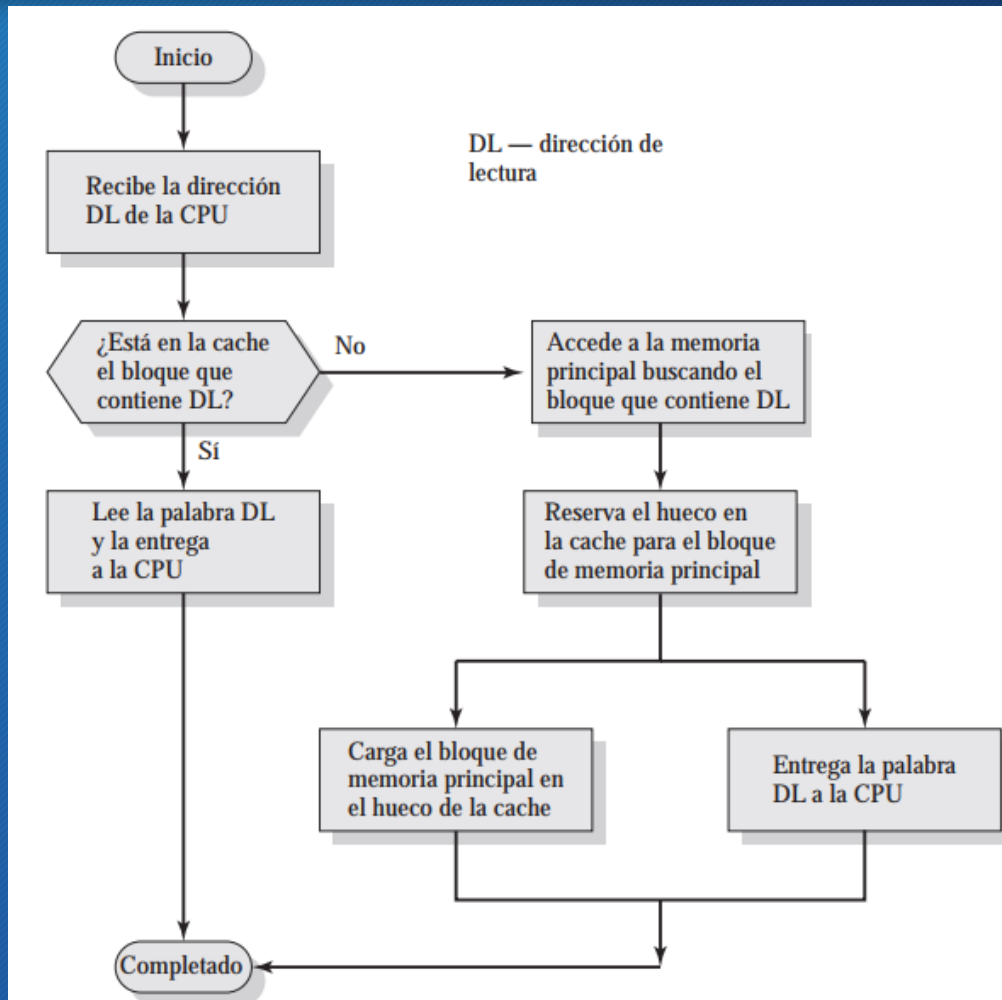
Memoria Cache

- El propósito de la memoria cache es proporcionar un tiempo de acceso a memoria próximo al de las memorias más rápidas disponibles y, al mismo tiempo, ofrecer un tamaño de memoria grande que tenga el precio de los tipos de memorias de semiconductores menos costosas. Hay una memoria principal relativamente grande y lenta junto con una memoria cache más pequeña y rápida. La cache contiene una copia de una parte de la memoria principal. Cuando el procesador intenta leer un byte de la memoria, se hace una comprobación para determinar si el byte está en la cache. Si es así, se le entrega el byte al procesador. En caso contrario, se lee e introduce dentro de la cache un bloque de memoria principal, que consta de un cierto número fijo de bytes, y, a continuación, se le entrega el byte pedido al procesador.

Diagrama cache y memoria principal



Operación de lectura de la cache



EL SISTEMA OPERATIVO COMO UNA INTERFAZ DE USUARIO/COMPUTADOR

- El hardware y software utilizados para proporcionar aplicaciones a los usuarios se pueden ver de forma jerárquica o en capas. El usuario de dichas aplicaciones, es decir, el usuario final, normalmente no se preocupa por los detalles del hardware del computador. Por tanto, el usuario final ve un sistema de computación en términos de un conjunto de aplicaciones. Una aplicación se puede expresar en un lenguaje de programación y normalmente es desarrollada por un programador de aplicaciones.

Servicios proporcionados por el S.O.

- De forma resumida, el sistema operativo proporciona normalmente servicios en las siguientes áreas:
- Desarrollo de programas: El sistema operativo proporciona una variedad de utilidades y servicios, tales como editores y depuradores, para asistir al programador en la creación de los programas. Normalmente, estos servicios se ofrecen en la forma de utilidades que, aunque no forman parte del núcleo del sistema operativo, se ofrecen con dicho sistema y se conocen como herramientas de desarrollo de programas de aplicación.

Servicios proporcionados por el S.O.

- Ejecución de programas: Se necesita realizar una serie de pasos para ejecutar un programa. Las instrucciones y los datos se deben cargar en memoria principal. Los dispositivos de E/S y los ficheros se deben inicializar, y otros recursos deben prepararse. Los sistemas operativos realizan estas labores de planificación en nombre del usuario.
- Acceso a dispositivos de E/S: Cada dispositivo de E/S requiere su propio conjunto peculiar de instrucciones o señales de control para cada operación. El sistema operativo proporciona una interfaz uniforme que esconde esos detalles de forma que los programadores puedan acceder a dichos dispositivos utilizando lecturas y escrituras sencillas

Servicios proporcionados por el S.O.

- Acceso controlado a los ficheros: Para el acceso a los ficheros, el sistema operativo debe reflejar una comprensión detallada no sólo de la naturaleza del dispositivo de E/S (disco, cinta), sino también de la estructura de los datos contenidos en los ficheros del sistema de almacenamiento. Adicionalmente, en el caso de un sistema con múltiples usuarios, el sistema operativo puede proporcionar mecanismos de protección para controlar el acceso a los ficheros.
- Acceso al sistema: Para sistemas compartidos o públicos, el sistema operativo controla el acceso al sistema completo y a recursos del sistema específicos. La función de acceso debe proporcionar protección a los recursos y a los datos, evitando el uso no autorizado de los usuarios y resolviendo conflictos en el caso de conflicto de recursos.

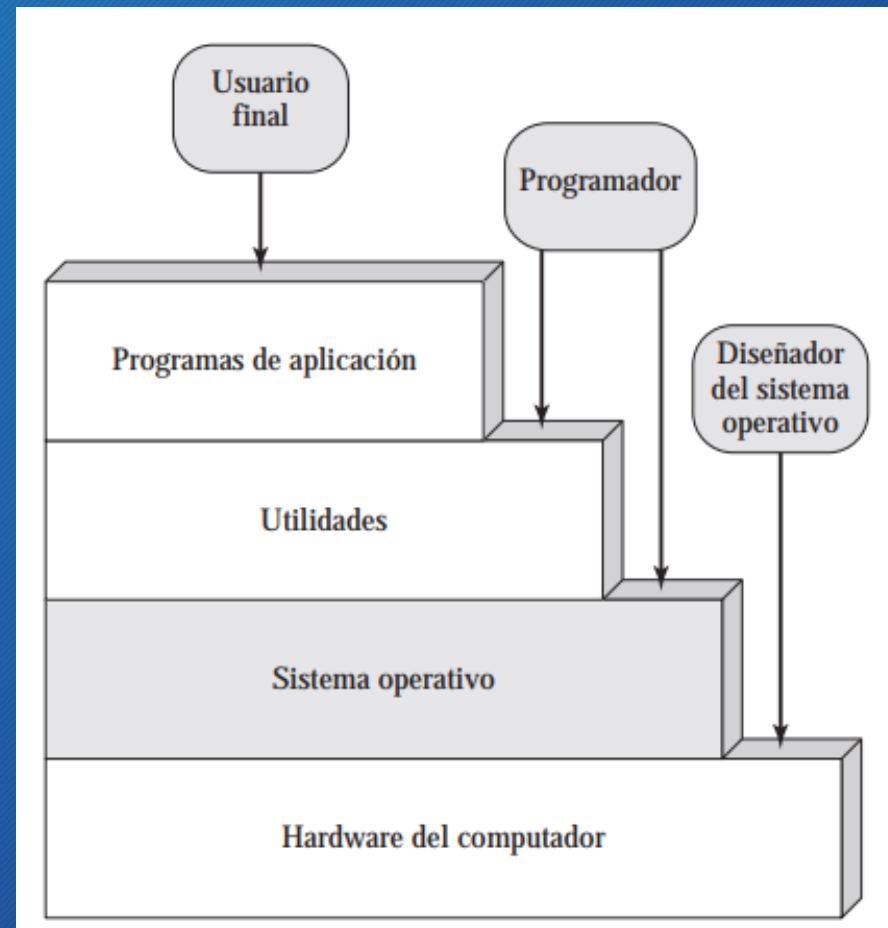
Servicios proporcionados por el S.O.

- Detección y respuesta a errores: Se pueden dar gran variedad de errores durante la ejecución de un sistema de computación. Éstos incluyen errores de hardware internos y externos, tales como un error de memoria, o un fallo en un dispositivo; y diferentes errores software, tales como la división por cero, el intento de acceder a una posición de memoria prohibida o la incapacidad del sistema operativo para conceder la solicitud de una aplicación. En cada caso, el sistema operativo debe proporcionar una respuesta que elimine la condición de error, suponiendo el menor impacto en las aplicaciones que están en ejecución. La respuesta puede oscilar entre finalizar el programa que causó el error hasta reintentar la operación o simplemente informar del error a la aplicación.

Servicios proporcionados por el S.O.

- Contabilidad. Un buen sistema operativo recogerá estadísticas de uso de los diferentes recursos y monitorizará parámetros de rendimiento tales como el tiempo de respuesta. En cualquier sistema, esta información es útil para anticipar las necesidades de mejoras futuras y para optimizar el sistema a fin de mejorar su rendimiento. En un sistema multiusuario, esta información se puede utilizar para facturar a los diferentes usuarios.

Capas y vistas de un sistema de computación



EL SISTEMA OPERATIVO COMO GESTOR DE RECURSOS

- Un computador es un conjunto de recursos que se utilizan para el transporte, almacenamiento y procesamiento de los datos, así como para llevar a cabo el control de estas funciones. El sistema operativo se encarga de gestionar estos recursos.
- ¿Se puede decir que es el sistema operativo quien controla el transporte, almacenamiento y procesamiento de los datos?

S.O. un mecanismo de control inusual en dos aspectos

- Las funciones del sistema operativo actúan de la misma forma que el resto del software; es decir, se trata de un programa o conjunto de programas ejecutados por el procesador.
- El sistema operativo frecuentemente cede el control y depende del procesador para volver a retomarlo.

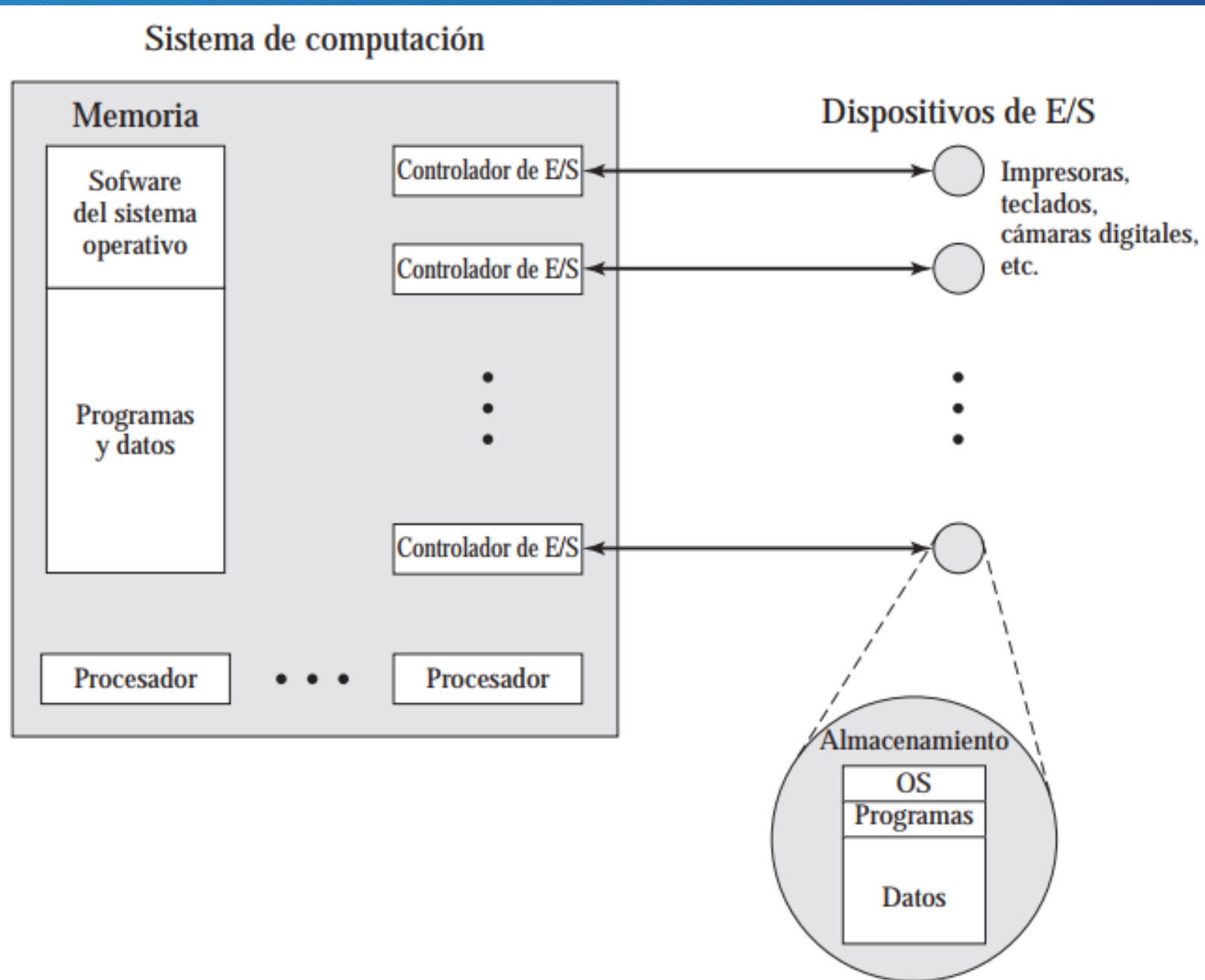


Figura 2.2. El sistema operativo como gestor de recursos.