Clase 8: Procesos.

Instrucción, cpu, plan de acción, SO.

Un proceso sucede cuando la información que tiene un programa se carga en la memoria y comienza su ejecución. Todos los softwares ejecutables se organizan en procesos que quieren utilizar la CPU y es el sistema operativo quién organiza el orden en que se van ejecutando esos procesos. A este cambio de proceso lo llamamos cambio de contexto.

TODOS los softwares ejecutables se organizan en procesos que quieren utilizar el CPU, el SO es quien organiza esto. Los procesos se realizan de a uno a la vez. Todos los procesos son efímeros, se crean y se terminan.

Se crean de diferentes formas:

* De manera interactiva con el usuario.
* En segundo plano, como llamados al SO.

Para que un proceso termine tiene que pasar por algún estado que determine su condición. Estos son:

1. Nuevo.
2. Listo: Cuando el SO lo carga.
3. Ejecución.
4. (Solo si hubo un error) Bloqueado: Cuando está esperando que un proceso o recurso pueda ser utilizado
5. Salida: Cuando cumple su objetivo.

Los mecanismos de comunicación entre procesos o IPC:

* Señales: Avisos que puede enviar un proceso a otro, luego sistema operativo se encarga de que el proceso que reciba la señal tome una acción para gestionarla
* Memoria compartida: Recurso compartido a disposición de los softwares para que puedan intercambiar información.

Cuando se da un proceso que no puede resolverse instantáneamente, como por ejemplo cuando ocurre una llamada al sistema, se van a crear otros procesos que se denominan hijos, su función es realizar sus tareas para lograr que el proceso padre pueda cumplir su objetivo. Los procesos padres pueden tener varios procesos hijos, pero los procesos hijos sólo pueden tener un proceso padre.

Comunicación de procesos.

Existen dos tipos de procesos que se ejecutan de manera concurrente:

* Independientes: Autónomos. No pueden ser afectados ni tampoco afectar a otros procesos que se están ejecutando en el sistema
* Cooperativos: Pueden afectar y ser afectados. Cualquier proceso que comparta cualquier tipo de datos o recursos con otros procesos es considerado cooperativo.

¿Por qué cooperativos?

Algunos procesos carecen de información y deben consultarla para ejecutarse, para esto debe ser compartida.

El CPU trabaja más eficientemente. Da como resultado la modularidad,  es decir que cuando una tarea contiene varios pasos el CPU puede ejecutarlos de manera independiente y simultánea.

Puede traer problemas si se ejecuta una tarea errónea o si no hay planificación.

Existen dos métodos de comunicación entre procesos: Memoria compartida (se establece un espacio en memoria que será compartido por los proceso) y paso de mensajes (con el kernel como intermediario en los pedidos de información).

**Ventajas y desventajas de los métodos:**

* La memoria compartida es generalmente más económica que usar un multiprocesador.
* En el paso de mensajes no existen los errores como exclusión mutua y son compatibles con cualquier tipo de arquitectura de computadoras.

Sincronización de procesos.

Los semáforos: Son una herramienta que sincroniza los procesos en base a sus estados. Si un proceso se está ejecutando y aparece una llamada de espera este pasa a una lista de bloqueados y permanece allí hasta que se le envía la señal de avance. Entonces el proceso que permanecía bloqueado, se coloca una fila de espera para utilizar el CPU.

Es importante llevar una buena planificación del uso del CPU porque de lo contrario esto lleva a que la cola de procesos colapse o tenga una inanición. Esto se produce cuando los recursos necesarios para el procesamiento no están disponibles entonces se le niega la ejecución a los procesos. Para combatir esto existen diferentes **técnicas de planificación**.

Hilos de ejecución.

Los procesadores son un conjunto de transistores configurados de tal forma que realizan operaciones binarias con impulsos energía eléctrica. Estos pueden contener uno o más núcleos, a mayor cantidad de núcleos mayores son los procesos que se pueden ejecutar en paralelo.

Se denomina proceso al conjunto de operaciones que componen un programa. Estas a la hora de ejecutarse se reparten el uso del procesador.

Un proceso puede dividirse en secuencias de tareas también denominadas hilos. Los hilos son porciones de código que pueden ejecutarse de forma simultánea en cooperación con otros subprocesos. Múltiples hilos pueden existir dentro de un proceso ejecutándose de forma concurrente, compartiendo recursos y memoria.

Hasta la década de los 2000 los procesadores eran monolíticos, por lo cual, sólo podían trabajar con un solo hilo a la vez, luego aparecieron los procesadores multinúcleo los cuales comenzaron con esta metodología de trabajo de varios hilos de ejecución aumentando así la velocidad de procesamiento.

Los sistemas de un solo núcleo tienen una capacidad de respuesta menor, su comportamiento es más predecible, no presentan los errores que podrían llegar a presentar los multihilos y los problemas de bloqueo de recursos bajan considerablemente. En comparación, los sistemas multihilos, poseen ventajas como una excelente capacidad de respuesta a operaciones con un buen trabajo en paralelo de sus tareas, pero como contraparte la sincronización es compleja de planificar y su comportamiento es difícil de predecir ya que puede presentar errores pasados por alto en la etapa de prueba y desarrollo.

Planificación de procesos.

Cuando hablamos de planificación hacemos referencia a las políticas y mecanismos que poseen los sistemas operativos actuales para realizar la gestión del procesador. Su objetivo es dar un buen servicio a todos los procesos que existan en un momento dado en el sistema.

El planificador del procesador tiene como misión la asignación del mismo a los procesos que están en la cola de procesos preparados

Criterios que se deben tener en cuenta en torno a un algoritmo de planificación.

* Rendimiento: Es el número de trabajos o procesos realizados por unidad de tiempo, que debe ser lo mayor posible.
* Tiempo de respuesta: Es la velocidad con que el ordenador da la respuesta a una petición. Depende mucho de la velocidad de los dispositivos de entrada y salida.
* Tiempo de servicio: Es el tiempo que tarda en ejecutarse un proceso, donde se incluye el tiempo de carga del programa en memoria, el tiempo de espera en la cola de procesos separados, el tiempo de ejecución en el procesador y el tiempo consumido en operaciones de entrada/salida.
* Tiempo de ejecución: Es idéntico al tiempo de servicio menos el tiempo de espera en la cola de procesos separados; es decir, es el tiempo teórico que necesitaría el proceso para ser ejecutado si fuera el único presente en el sistema.
* Tiempo de procesador: Es el tiempo que un proceso está utilizando el procesador sin contar el tiempo que se encuentra bloqueado por operaciones de entrada/salida.
* Eficiencia: Se refiere a la utilización del recurso más caro en un sistema, el procesador, que debe estar el mayor tiempo posible ocupado para lograr así un gran rendimiento.
* Tiempo de espera: Es el tiempo que los procesos están activos, pero sin ser ejecutados, es decir, los tiempos de espera en las distintas colas.

Políticas de planificación

FIFO (primero en entrar, primero en salir)

El procesador ejecuta cada proceso hasta que termina; por tanto, los procesos entren en cola de procesos preparados permanecerán allí en el orden en que lleguen hasta que les toque su ejecución.

Round-Robin (RR)

Consiste en conceder a cada proceso de ejecución un determinado período de tiempo Q (quantum), transcurrido el cual, si el proceso no ha terminado, se le devuelve al final de la cola de procesos preparados, concediéndose el procesador al siguiente proceso por su correspondiente quantum.

En el caso de que aún falte tiempo de ejecución, vuelve a la cola ubicándose al final, hasta que nuevamente es su turno y de esta forma se establece que todos los procesos llevan un tiempo de ejecución equitativo y todos serán ejecutados de la misma manera.

SJF (tiempo más corto)

Esta política toma de la cola de procesos preparados el que necesite menos tiempo de ejecución para realizar su trabajo. Para ello, debe saber el tiempo de ejecución que necesita cada proceso, lo cual no es tarea fácil, pero es posible a través de diversos métodos como puede ser la información suministrada por el propio usuario o por el propio programa, basándose en la historia anterior.

SRTF (tiempo restante más corto)

Esta técnica cambia el proceso que está ejecutando cuando se ejecuta un proceso con una exigencia de tiempo de ejecución total menor que el que se está ejecutando en el procesador.

**Colas Múltiples**

Cuando los procesos que van a ser ejecutados en una computadora se pueden agrupar en distintos grupos, podemos asignarlos a diferentes colas, cada una con una planificación distinta, para darle a cada una de ellas lo que realmente necesite.

Esta política divide la cola en procesos preparados en varias colas separadas, de manera que los procesos se asignan a una determinad cola según sus necesidades y tipo.

Existen otras planificaciones, las cuales combinan de forma híbrida las anteriores o presentan algoritmos diferentes, como ser la retroalimentación multinivel o planificación por comportamiento estableciendo prioridades entre los procesos o cambiando el tiempo de ejecución, pero como sabemos estos procesos también viven en constante innovación gracias a los avances de la tecnología.

Un proceso puede abrirse en x hilos para ejecutarse paralelamente. Los hilos son subprocesos. Un proceso, independientemente de quien sea su padre y quienes sus hijos, puede ejecutarse en 1 o más hilos. Un proceso puede ejecutarse en varios hilos dependiendo de si las tareas pueden realizarse al mismo tiempo o no.