

Guía de presentación del tema 5

Jerarquía de memoria

El objetivo de este tema es presentar el concepto de jerarquía de memoria en un computador. En este tema se presentan los siguientes contenidos:

1. Introducción a la jerarquía de memoria
2. Memoria caché
3. Memoria virtual

1. Introducción a la jerarquía de memoria

En este apartado se describen los distintos tipos de dispositivos que se emplean en un computador como dispositivo de memoria: memorias de semiconductores, memorias magnéticas y memorias ópticas. Se analiza cómo obtener el número de accesos a memoria que genera un programa durante su ejecución y el impacto que tiene en el rendimiento del computador. A continuación, se describe el concepto de jerarquía de memoria y se analiza la estructura de las memorias RAM estáticas y dinámicas.

2. Memoria caché

La memoria caché es una pequeña memoria rápida de tipo estática (SRAM), que se sitúa entre el procesador y la memoria principal y permite acelerar la ejecución de los programas. En esta parte se estudia el funcionamiento básico de la memoria caché y se muestra cómo mejora el tiempo de ejecución de los programas. Se presenta el tiempo medio de acceso a un sistema de memoria que utiliza memoria caché.

A continuación, se estudia la estructura de la memoria caché y el diseño interno de la misma analizando su tamaño, la función de correspondencia, los algoritmos de sustitución de caché y la política de escritura. Se presta especial atención a la función de correspondencia, que es el algoritmo que determina en qué lugar de la memoria caché se puede almacenar un bloque concreto de la memoria principal. Es el mecanismo que permite saber qué bloque de memoria principal está ubicado en una línea de memoria caché. Se describen las funciones de correspondencia directa, asociativa y asociativa por conjuntos.

3. Memoria Virtual

Este tema comienza ilustrando el proceso de carga y ejecución de un programa. Se denomina proceso a un programa en ejecución e imagen de memoria al conjunto de direcciones de memoria que ocupa un programa durante su ejecución. También se describe el proceso de reubicación de un programa en memoria, tanto software como hardware.

En un sistema sin memoria virtual, la imagen de memoria de un programa debe residir de forma completa en la memoria principal del computador para que se pueda ejecutar. En un sistema con memoria virtual, las imágenes de memoria de los procesos residen en un mapa de memoria virtual o espacio de direcciones virtual que está físicamente soportado por dos niveles de la jerarquía de memoria: la memoria principal y el disco. Durante la ejecución de un programa, habrá partes de la imagen de memoria de un proceso que estarán en memoria principal y otras que podrán residir en disco.

En este tema se estudian todos los fundamentos básicos de la memoria virtual y con especial detalle el mecanismo conocido como paginación. En un sistema de memoria virtual paginada, el espacio de direcciones virtuales de un proceso se divide en trozos de igual tamaño denominado

páginas. La memoria principal se divide en trozos de igual tamaño, denominado marcos de página y la zona de disco que da soporte a la memoria virtual se divide en trozos de igual tamaño denominados páginas de intercambio o páginas de swap. Se describen las tareas que realiza la unidad de gestión de memoria (MMU) para traducir las direcciones virtuales que genera el procesador a las direcciones físicas donde se almacena realmente el dato al que está referenciando el procesador. También se describen las tareas que debe realizar el sistema operativo durante un fallo de página, interrupción que se produce cuando se intenta acceder a una dirección que no reside en memoria principal.

Se describen las tablas de páginas, que son las estructuras que utilizan los computadores para poder conocer en qué partes de memoria o disco se almacena una determinada parte de la imagen de memoria virtual de un proceso. A continuación, se describe el proceso de traducción de direcciones virtuales a físicas y se presentan diferentes estructuras de tablas de páginas:

- Tablas de página de un nivel.
- Tablas de página de dos niveles.
- Tablas de página invertida.

Este apartado también presenta el concepto de TLB (Translation Lookaside Buffer) que es una pequeña memoria caché de entradas de tablas de páginas usadas más recientemente y que se emplea para acelerar el proceso de búsqueda de una página en memoria.

Por último, se describe el funcionamiento general de un sistema que emplea memoria virtual y memoria caché.

Material asociado

Como material asociado a este tema se incluye el material de teoría y una colección de ejercicios propuestos y resueltos sobre los aspectos tratados en el tema. Se propone una práctica para comprender el efecto que tiene en el rendimiento de un computador la memoria caché. Esta práctica utiliza un programa, denominado calibrator, que permite determinar el número de niveles de memoria caché, el tamaño y el tamaño de la línea de la memoria caché instalada en un computador. En esta práctica se propone al alumno que determine la tasa de fallos que se produce durante la ejecución de un programa y analice cómo afecta la memoria caché al funcionamiento de los programas.

En otros recursos se proporciona un enlace a diversos simuladores de memoria caché que ilustran el funcionamiento básico de una memoria caché. También se presenta un enlace a diversos simuladores sobre memoria virtual y políticas de reemplazo de páginas.

Lecturas recomendadas

- Capítulo 6 y 8 del libro “Problemas resueltos de estructuras de computadores” (GARCIA CARBALLEIRA, Félix et al.).
- Capítulo 5 del libro “Computer organization and design. The hardware/software interface” (PATTERSON, David, et al).
- Capítulo 4 y 5 del libro “Computer Organization and Architecture” (STALLINGS, William).
- Capítulo 11 del libro de “Fundamentos de los sistemas digitales” (FLOYD, Thomas L.).

