Unidad 1

1.1.1 Arquitecturas Cl&aacutesicas

Arquitectura de computadoras. Es el dise&ntildeo y la organizaci&oacuten de un sistema para un equipo de c&oacutemputo. Es un modelo y descripci&oacuten de cada funci&oacuten, as&iacute como los requerimientos y las implementaciones de dise&ntildeo para varias partes del equipo de c&oacutemputo.

Principalmente enfocamos en la Unidad Central de Procesamiento lo conocemos como (CPU) el cual trabaja internamente y accede a las direcciones de memoria y a los sistemas de entrada salida, perif&eacutericos. Tambi&eacuten suele definirse como la selecci&oacuten e interconexi&oacuten de los componentes de hardware para crear computadoras seg&uacuten los requerimientos de funcionalidad, rendimiento y costo.

El equipo de c&oacutemputo recibe y env&iacutea la informaci&oacuten a trav&eacutes de los perif&eacutericos por medio de los canales de distribuci&oacuten o buses (cableado f&iacutesico de interconexi&oacuten). El CPU se encarga de procesar la informaci&oacuten que le llega al equipo de c&oacutemputo.

El intercambio de informaci&oacuten se tiene que hacer con los perif&eacutericos y el CPU.

Todas aquellas unidades de un sistema exceptuando el CPU se denomina perif&eacuterico, por lo que el equipo de c&oacutemputo tiene dos partes bien diferenciadas, que son: el CPU (se encarga de ejecutar programas y que est&aacute compuesta por la memoria principal, la (ALU) Unidad Aritm&eacutetica L&oacutegica y la (UC) unidad de control) y los perif&eacutericos (que pueden ser de entrada, salida, entrada-salida y las interconexiones).

Modelos de Arquitectura Cl&aacutesicas.

Estas arquitecturas se desarrollaron en las primeras computadoras electromec&aacutenicas y de tubos de vac&iacuteo.

Hay dos arquitecturas distintas relacionadas con el uso y distribuci&oacuten de la memoria: Arquitectura de Jonh Von Neumman y Arquitectura Harvard.

Arquitectura Von Neumann

Tradicionalmente los sistemas con microprocesadores se basan en esta arquitectura, en la cual la unidad central de proceso (CPU), est&aacute conectada a una memoria principal &uacutenica (casi siempre s&oacutelo RAM) donde se guardan las instrucciones del programa y los datos. A dicha memoria se accede a trav&eacutes de un sistema interconexi&oacuten de buses &uacutenico (control, direcciones y datos).

En un sistema con arquitectura Von Neumann el tama&ntildeo de la unidad de datos o instrucciones est&aacute fijado por el ancho del bus que comunica la memoria con la CPU.

As&iacute un microprocesador de 8 bits con un bus de 8 bits, tendr&aacute que manejar datos e instrucciones de una o m&aacutes unidades de 8 bits (bytes) de longitud.

Si tiene que acceder a una instrucci&oacuten o dato de m&aacutes de un byte de longitud, tendr&aacute que realizar m&aacutes de un acceso a la memoria.

El tener un &uacutenico bus hace que el microprocesador sea m&aacutes lento en su respuesta, ya que no puede buscar en memoria una nueva instrucci&oacuten mientras no finalicen las transferencias de datos de la instrucci&oacuten anterior.

Las computadoras digitales convencionales presentan un aspecto Von Neumann.

Este modelo consta de cinco componentes principales.

Unidad de Memoria.

Unidad de Entrada/Salida.

Unidad de Control. Incluido en el CPU.

Unidad Aritm&eacutetica L&oacutegica. Incluida en CPU.

Registros de Programas. Incluidos en CPU.

Las instrucciones provenientes del sistema de entrada, son almacenados por la memoria, procesados por la ALU bajo la direcci&oacuten de la unidad de control. Los resultados obtenidos son enviados a la unidad de salida.

El programa almacenado es lo m&aacutes importante en el modelo.

Los programas se almacenan en la memoria del equipo junto con los datos a procesar.

En las computadoras de programa almacenado, el programa puede manipularse como si se tratara de datos. Este concepto da origen a los compiladores, sistemas operativos y es la base de la gran versatilidad de las computadoras modernas.

Limitaciones Von Neumann

La longitud de las instrucciones por el bus de datos, que hace que el microprocesador tenga que realizar varios accesos a memoria para buscar instrucciones complejas.

La velocidad de operaci&oacuten a causa del bus &uacutenico para datos e instrucciones que no deja acceder simult&aacuteneamente a unos y otras, lo cual impide superponer ambos tiempos de acceso.

Modelo Harvard

Esta arquitectura utiliza los Micro controladores, tiene la unidad central de proceso (CPU) conectada a dos memorias (una con las instrucciones y otra con los datos) por medio de dos buses diferentes.

Una de las memorias contiene solamente las instrucciones del programa (Memoria de Programa), y la otra s&oacutelo almacena datos (Memoria de Datos).

Ambos buses son totalmente independientes lo que permite que la CPU pueda acceder de forma independiente y simult&aacutenea a la memoria de datos y a la de instrucciones.

Como los buses son independientes estos pueden tener distintos contenidos en la misma direcci&oacuten y tambi&eacuten distinta longitud.

Tambi&eacuten la longitud de los datos y las instrucciones puede ser distinta, lo que optimiza el uso de la memoria en general.

Para un procesador de Set de Instrucciones Reducido, o RISC (Reduced Instrucci&oacuten Set Computer), el set de instrucciones y el bus de memoria de programa pueden dise&ntildearse de tal manera que todas las instrucciones tengan una sola posici&oacuten de memoria de programa de longitud.

Adem&aacutes, al ser los buses independientes, el CPU puede acceder a los datos para completar la ejecuci&oacuten de una instrucci&oacuten, y al mismo tiempo leer la siguiente instrucci&oacuten a ejecutar.

El tama&ntildeo de las instrucciones no est&aacute relacionado con el de los datos, y por lo tanto puede ser optimizado para que cualquier instrucci&oacuten ocupe una sola posici&oacuten de memoria de programa, logrando as&iacute mayor velocidad y menor longitud de programa.

El tiempo de acceso a las instrucciones puede superponerse con el de los datos, logrando una mayor velocidad en cada operaci&oacuten.

La principal desventaja de esta arquitectura; el bus de datos y direcciones &uacutenico se convierte en un cuello de botella por el cual debe pasar toda la informaci&oacuten que se lee de o se escribe a la memoria, obligando a que todos los accesos a esta sean secuenciales.

Limita el grado de paralelismo (acciones que se pueden realizar al mismo tiempo) y por lo tanto, el desempe&ntildeo de la computadora.

1.1.2

Arquitectura Segmentada.

Es una de las tecnolog&iacuteas utilizadas para realizar la segmentaci&oacuten paralelismo.

Divide el procesador, en etapas, procesa una instrucci&oacuten diferente en cada una y trabaja con varias a la vez.

Pueden trabajar de forma paralela, en diferentes instrucciones, utilizando una cola de instrucciones para su comunicaci&oacuten, denominado entubamiento.

La dependencia de datos y de control, que tiene como efecto la disminuci&oacuten del rendimiento del pipelining.

La segmentaci&oacuten de cauce (pipelining) es una forma efectiva de organizar el hardware del CPU para realizar m&aacutes de una operaci&oacuten al mismo tiempo.

Consiste en descomponer el proceso de ejecuci&oacuten de las instrucciones en fases o etapas que permitan una ejecuci&oacuten simult&aacutenea.

Explota el paralelismo entre las instrucciones de un flujo secuencial.

La segmentaci&oacuten es una t&eacutecnica de implementaci&oacuten por la cual se solapa la ejecuci&oacuten de m&uacuteltiples instrucciones.

La t&eacutecnica de implementaci&oacuten clave utilizada para hacer CPU r&aacutepidas.

La segmentaci&oacuten es como una l&iacutenea de ensamblaje: cada etapa de la segmentaci&oacuten completa una parte de la instrucci&oacuten.

Como en una l&iacutenea de ensamblaje, la acci&oacuten a realizar en una instrucci&oacuten se descompone en partes m&aacutes peque&ntildeas, cada una de las cuales necesita una fracci&oacuten del tiempo necesario para completar la instrucci&oacuten completa.

Cada uno de estos pasos se define como etapa de la segmentaci&oacuten o segmento.

Las etapas est&aacuten conectadas, cada una a la siguiente, para formar una especie de cauce las instrucciones se entran por un extremo, son procesadas a trav&eacutes de las etapas y salen por el otro.

La productividad de la segmentaci&oacuten est&aacute determinada por la frecuencia con que una instrucci&oacuten salga del cauce.

Como las etapas est&aacuten conectadas entre s&iacute, todas las etapas deben estar listas para proceder al mismo tiempo. El tiempo requerido para desplazar una instrucci&oacuten, un paso, a lo largo del cauce es un ciclo m&aacutequina.

La duraci&oacuten de un ciclo m&aacutequina est&aacute determinada por el tiempo que necesita la etapa m&aacutes lenta (porque todas las etapas progresan a la vez).

Con frecuencia, el ciclo m&aacutequina es un ciclo de reloj (a veces dos, o raramente m&aacutes), aunque el reloj puede tener m&uacuteltiples fases.

La mejora de velocidad debida a la segmentaci&oacuten es igual al n&uacutemero de etapas.

La segmentaci&oacuten consigue una reducci&oacuten en el tiempo de ejecuci&oacuten medio por instrucci&oacuten.

Esta reducci&oacuten se puede obtener decrementando la duraci&oacuten del ciclo de reloj de la m&aacutequina segmentada o disminuyendo el n&uacutemero de ciclos de reloj por instrucci&oacuten, o haciendo ambas cosas.

El mayor impacto est&aacute en el n&uacutemero de ciclos de reloj por instrucci&oacuten, aunque el ciclo de reloj es, con frecuencia, m&aacutes corto en una m&aacutequina segmentada (especialmente en supercomputadores segmentados).

La segmentaci&oacuten es una t&eacutecnica de implementaci&oacuten, que explota el paralelismo entre las instrucciones de un flujo secuencial. Tiene la ventaja sustancial que, de forma distinta a algunas t&eacutecnicas de aumento de velocidad, no es visible al programador.

Tipos de cauces:

Unifunci&oacuten: ejecutan un &uacutenico proceso.

Multifunci&oacuten: pueden ejecutar varios procesos:

Est&aacuteticos: en un instante determinado s&oacutelo pueden ejecutar

uno.

Din&aacutemicos: pueden ejecutar simult&aacuteneamente varios

procesos.

Lineal: a cada etapa s&oacutelo le puede seguir otra etapa concreta.

No lineal: se pueden establecer recorridos complejos de

las etapas.

Ciclos de Instrucci&oacuten.

Ejecuci&oacuten de instrucciones en 5 ciclos

Etapas de cauce.

Donde:

IF: B&uacutesqueda de una instrucci&oacuten de la memoria.

ID: B&uacutesqueda de registros y decodificaci&oacuten de instrucciones

EXE: Ejecuci&oacuten o c&aacutelculo de direcci&oacuten

MEM: Acceso a la memoria de datos.

WB: Escribir datos en el archivo de registros

La segmentaci&oacuten ayuda en la realizaci&oacuten de una carga de

trabajo.

Realiza m&uacuteltiples tareas simult&aacuteneamente utilizando

diferentes recursos.

La velocidad se incrementa si se aumenta el n&uacutemero de

segmentos.

La raz&oacuten de segmentaci&oacuten est&aacute dada por el segmento m&aacutes lento.

El tiempo en llenar y vaciar los segmentos reduce la velocidad.

Dedica unidades espec&iacuteficas del procesador a cada una de las partes del ciclo de instrucci&oacuten y ejecutarlas paralelamente.

Mejora dram&aacuteticamente el rendimiento debido a que en los procesadores tienen etapas en el ciclo de instrucci&oacuten.

La unidad de control se subdivide en dos o m&aacutes unidades funcionales, cada una encargada de llevar a cabo una parte del ciclo de instrucci&oacuten.

Se comunican a trav&eacutes de una cola de instrucciones (Pipeline).

Arquitectura de Multiprocesamiento.

Se denomina multiprocesador a un computador que cuenta con dos o m&aacutes microprocesadores (CPUs).

El multiprocesador puede ejecutar simult&aacuteneamente varios hilos pertenecientes a un mismo proceso o bien a procesos diferentes.

La arquitectura NUMA, donde cada procesador tiene acceso y control exclusivo a una parte de la memoria.

La arquitectura SMP, donde todos los procesadores comparten toda la memoria.

Para que un multiprocesador opere correctamente necesita un sistema operativo especialmente dise&ntildeado para ello.

La mayor&iacutea de los sistemas operativos actuales poseen esta capacidad.

Cuando se desea incrementar el desempe&ntildeo m&aacutes de lo que permite la t&eacutecnica de segmentaci&oacuten (pipeline), se requiere utilizar m&aacutes de un procesador para la ejecuci&oacuten del programa de aplicaci&oacuten.

Los CPU de multiprocesamiento seg&uacuten Flynn se clasifican de la siguiente manera:

SISO - (Single Instruction, Single Operand) computadoras Monoprocesador

SIMO - (Single Instruction, Multiple Operand) procesadores vectoriales, Exenciones MMX

MISO - (Multiple Instruction, Single Operand) No implementado

MIMO - (Multiple Instruction, Multiple Operand) sistemas SMP, Clusters, GPUs

Los procesadores vectoriales, son computadoras pensadas para aplicar un mismo algoritmo num&eacuterico a una serie de datos matriciales, en especial en la simulaci&oacuten de sistemas f&iacutesicos complejos, tales como simuladores de clima, explosiones at&oacutemicas, reacciones qu&iacutemicas, etc.

Donde los datos son representados como grandes n&uacutemeros de datos en forma matricial sobre los que se deben se aplicar el mismo algoritmo num&eacuterico.

La mayor&iacutea de los procesadores modernos incluye algunas instrucciones de tipo vectorial, tales como las extensiones al conjunto de instrucciones tales como MMX y SSE. Estas instrucciones les permiten procesar flujos multimedia m&aacutes eficientemente.

Los Procesadores Digitales de Se&ntildeales (DSP), son procesadores especializados en el procesamiento de se&ntildeales tales como audio, v&iacutedeo, radar, sonar, radio, etc.

Cuentan con instrucciones tipo vectorial que los hace muy aptos para dicha aplicaci&oacuten. Suelen utilizarse en conjunto con un microcontrolador en dispositivos como reproductores de audio, reproductores de DVD y Blueray, tel&eacutefonos celulares, sistemas de entretenimiento, sistemas de adquisici&oacuten de datos, instrumentos m&eacutedicos, controles industriales, etc.

En los sistemas SMP (Simetric Multiprocessing), varios procesadores comparten la misma memoria principal y perif&eacutericos de Entrada /Salida, normalmente conectados por un bus com&uacuten.

Se conocen como sim&eacutetricos, ya que ning&uacuten procesador toma el papel de maestro y los dem&aacutes de esclavos, sino que todos tienen derechos similares en cuanto al acceso a la memoria y perif&eacutericos y ambos son administrados por el sistema operativo.

Pueden formarse con varios n&uacutecleos en un solo circuito integrado o con varios circuitos integrados en una misma tarjeta madre. La primera opci&oacuten ha sido popularizada al hacerse m&aacutes econ&oacutemicos los procesadores multin&uacutecleo de los principales fabricantes y con su uso en sistemas de gama media y baja, e inclusive en tel&eacutefonos celulares y tabletas.

La segunda opci&oacuten fue la que se uso en un principio y sigue siendo usada en en estaciones de trabajo y en servidores de alto rendimiento debido a que incrementa el poder computacional del sistema, pero tambi&eacuten incrementa considerablemente el costo del sistema.

Clusters

Conjuntos de computadoras independientes conectadas en una red de &aacuterea local o por un bus de interconexi&oacuten y que trabajan cooperativamente. Con un sistema de procesamiento paralelo o distribuido. Consta de un conjunto de computadoras independientes, interconectadas entre s&iacute, de tal manera que funcionan como un solo recurso computacional.

Es clave en su funcionamiento contar con un sistema operativo y programas de aplicaci&oacuten capaces de distribuir el trabajo entre las computadoras de la red.

Este tipo de computadora paralela se ha vuelto muy popular por que permite usar los avances en los procesadores comerciales que tienen una muy buena relaci&oacuten costo rendimiento y se puede incorporar r&aacutepidamente los avances que proporciona las nuevas tecnolog&iacuteas en cuanto es econ&oacutemicamente viable.

Sin embargo, se debe tener cuidado al implementar la aplicaci&oacuten, ya que si los datos que hay que pasar de un procesador a otro son demasiados, el tiempo empleado en pasar informaci&oacuten de un nodo a otro puede sobrepasar a la ganancia que se tiene al dividir el trabajo entre varios procesadores.

Procesadores Gr&aacuteficos

(Graphics Processing Unit GPU)

Sistemas dise&ntildeados originalmente para el procesamiento de gr&aacuteficos, con m&uacuteltiples procesadores vectoriales sencillos compartiendo la misma memoria.

La cual tambi&eacuten puede ser accedida por el CPU. Por la gran cantidad de n&uacutecleos con los que cuenta, logran un excelente desempe&ntildeo al ejecutar algoritmos que se adaptan a ser paralelizados, a tal grado que muchas de las supercomputadoras m&aacutes r&aacutepidas de la actualidad utilizan estos procesadores, y los fabricantes de tarjetas gr&aacuteficas producen versiones de sus productos especializadas en acelerar los c&aacutelculos de prop&oacutesito general.