

República Bolivariana de Venezuela.

Ministerio del Poder Popular para la Educación.

Instituto Universitario Mario Briceño Iragorry "IUTEMBI".

Área: Arquitectura del Computador.

ARQUITECTURAS PARALELAS.

Alumna:

Karen Natacha Angel Valero.

C.I: 27.557.272.

Valera, 26 de julio de 2022.

ARQUITECTURA PARALELA.

Es un tipo de arquitectura de computación que permite procesar instrucciones y datos simultáneamente, siempre que el problema lo permita. Esto implica que la construcción de un programa, con instrucciones en paralelo, alterará el problema en tareas de menor complejidad que se ejecutarán simultáneamente.

El primer paso hacia la paralelización de las arquitecturas de los computadores, se da con la aparición de los procesadores o sistemas vectoriales. Los procesadores vectoriales extienden el concepto de paralelismo por segmentación al tratamiento de grandes cadenas de datos. El hecho de que los procesadores segmentados hayan venido asociados a los supercomputadores paralelos, los pone en la entrada a lo que son los sistemas paralelos, si bien siguen siendo una extensión del concepto de segmentación.

Cabe destacar que en el mundo de la computación el paralelismo se puede aplicar a otras áreas. En general, existen tres tipos de paralelismo:

- Paralelismo a nivel de bit: referido al tamaño de los datos con los que puede trabajar el procesador. Por ejemplo, un procesador con un tamaño de palabra (32 bits), puede efectuar 4 sumas independientes de 1 byte de tamaño al mismo tiempo.
- Paralelismo a nivel de instrucción: las instrucciones de un programa se reordenan y agrupan para ser ejecutadas en paralelo. Los procesadores modernos poseen unas estructuras conocidas como pipelines (tuberías), que permiten segmentar la ejecución de las instrucciones. De este modo, es posible ejecutar diferentes etapas de varias instrucciones al mismo tiempo.
- Paralelismo a nivel de tarea: un problema concreto se puede descomponer en tareas más pequeñas que pueden ser resueltas de manera concurrente por los diferentes elementos de proceso (procesadores, hilos...).

CLASIFICACIÓN DE LOS SISTEMAS PARALELOS.

Clasificación de Flynn.

Esta taxonomía de las arquitecturas está basada en la clasificación atendiendo al flujo de datos e instrucciones en un sistema. Un flujo de instrucciones es el conjunto de instrucciones secuenciales que son ejecutadas por un único procesador, y un flujo de datos es el flujo secuencial de datos requeridos por el flujo de instrucciones. Con estas consideraciones, Flynn clasifica los sistemas en cuatro categorías:

- SISD (Single Instruction stream, Single Data stream): Flujo único de instrucciones y flujo único de datos.
- MISD (Multiple Instruction stream, Single Data stream): Flujo múltiple de instrucciones y único flujo de datos. Esto significa que varias instrucciones actúan sobre el mismo y único trozo de datos.
- SIMD (Single Instruction stream, Multiple Data stream): Flujo de instrucción simple y flujo de datos múltiple. Esto significa que una única instrucción es aplicada sobre diferentes datos al mismo tiempo.
- MIMD (Multiple Instruction stream, Multiple Data stream): Flujo de instrucciones múltiples y flujo de datos múltiples. Son máquinas que poseen varias unidades procesadoras en las cuales se pueden realizar múltiples instrucciones sobre datos diferentes de forma simultánea.

Además de la clasificación de Flynn, existen otro tipo de clasificaciones, algunas de ellas son:

- UMA: acceso uniforme a memoria.
- COMA: acceso solo a memoria cache.
- NUMA: acceso no uniforme a memoria. La memoria es compartida pero el tiempo de acceso varía.
- COW: grupo de estaciones de trabajo.
- MPP: procesadores en paralelo.

- CC-NUMA: memoria temporal.
- NC-NUMA: sin memoria temporal.

IMPORTANCIA DE LA ARQUITECTURA PARALELA.

La arquitectura paralela es una técnica fundamental en el ámbito de investigación científica, especialmente en el campo de la simulación, donde se llevan a cabo cálculos y operaciones complejas que requieren de una gran capacidad de procesamiento. También tiene aplicaciones en la creación de modelos de todo tipo: matemáticos, estadísticos, climáticos e incluso en imágenes médicas.

Otros ejemplos relevantes serían los sistemas en tiempo real, la inteligencia artificial, el procesamiento de gráficos o los servidores. En estos últimos los procesadores multinúcleo son ideales porque permiten a muchos usuarios conectarse simultáneamente a un mismo servicio (por ejemplo, en el caso de un servidor web).

TÉRMINOS GENERALES DE ARQUITECTURA PARALELA.

1. Potencia de Computo:

Los superordenadores pueden procesar cálculos complejos muy rápidamente. Este es el secreto básico de la potencia de un ordenador. Todo se reduce a lo rápido que puede un ordenador realizar una operación. Todo lo que un ordenador puede hacer se puede simplificar en matemáticas. El procesador de tu ordenador interpreta cualquier comando que ejecutas como una serie de problemas de matemáticas. Los procesadores más rápidos pueden manejar más cálculos por segundo que los equipos más lentos, y son mejores para hacer cálculos mucho más complicados.

2. Concurrencia:

Es la propiedad de los sistemas que permiten que múltiples procesos sean ejecutados al mismo tiempo, y que potencialmente puedan interactuar entre sí.

Además es la habilidad de distintas partes de un programa, algoritmo, o problema de ser ejecutado en desorden o en orden parcial, sin afectar el resultado final

3. Paradigma Paralelo:

Es el método mediante el cual una serie de tareas e instrucciones se ejecutan de forma simultánea. Como cualquier trabajo en paralelo, se trata de dividir el trabajo en trozos más simples, que actualmente solemos llamar hilos o subprocesos.

4. Balanceo de carga:

Es una técnica usada para compartir el trabajo a realizar entre varios ordenadores, procesos, discos u otros recursos. Existen varios métodos para realizar el balanceo de carga. Desde el simple "Round Robin", hasta los equipos que reciben las peticiones, recogen información, en tiempo real, de la capacidad operativa de los equipos y la utilizan para enrutar dichas peticiones individualmente al servidor que se encuentre en mejor disposición de prestar el servicio adecuado.

5. Métricas para evaluar sistemas paralelos:

Cuando se utiliza un algoritmo paralelo para la resolución de un problema, interesa saber cuál es la ganancia en la performance obtenida y resulta importante estudiar el rendimiento de los programas paralelos con el fin de determinar el mejor algoritmo, evaluar las plataformas de hardware, y examinar los beneficios del paralelismo.

6. Clusters:

Es un sistema de procesamiento paralelo o distribuido. Consta de un conjunto de computadoras independientes, interconectadas entre sí, de tal manera que funcionan como un solo recurso computacional. A cada uno de los elementos del cluster se le conoce como nodo. Estos son aparatos o torres que pueden tener uno o varios procesadores, memoria RAM, interfaces de red, dispositivos de entrada y salida, y sistema operativo.

7. Speedup.

Es un proceso realizado para mejorar el rendimiento de un sistema que procesa un problema determinado. Es decir, es la mejora en la velocidad de ejecución de una tarea ejecutada en dos arquitecturas similares con diferentes recursos.

8. Política de asignación dinámica:

Se pueden cambiar orígenes, destinos y la lógica de transformación en tiempo de ejecución, en función de los parámetros y las reglas que se hayan definido. Puede configurar asignaciones dinámicas para permitir cambios de metadatos en orígenes y destinos. Puede determinar qué puertos recibe una transformación, qué puertos desea utilizar en la lógica de transformación y qué vínculos quiere establecer entre los grupos de transformación.

9. Overhead:

Tiempo de procesamiento empleado por el software del sistema; es el exceso de tiempo de computación, memoria, ancho de banda u otros recursos, que son necesarios para realizar una tarea específica.

10. Debugging y Tuning:

Debugging es un análisis exhaustivo del código fuente de un programa informático que tiene como objetivo identificar posibles errores. Tuning es afinar la configuración de hardware y software para optimizar su rendimiento.

11. Tiempo Efectivo:

Es el promedio de tiempo que representa la rapidez con que un equipo aborda las fallas.

12. Eficiencia:

Es el uso racional de los recursos con que se cuenta para alcanzar un objetivo predeterminado. A mayor eficiencia menor la cantidad de recursos que se emplearán, logrando mejor optimización y rendimiento.

13. Pipeline:

Consiste en una cadena de procesos conectados de forma tal, que la salida de cada elemento de la cadena es la entrada del próximo. Permiten la comunicación y sincronización entre procesos. Es común el uso de búfer de datos entre elementos consecutivos.

14. CPI:

Los ciclos por instrucción o CPI son una medida del rendimiento de un procesador: el número medio de ciclos de reloj por instrucción en un programa o fragmento. Es el inverso multiplicativo de las instrucciones por ciclo.

15. IPC:

El concepto instrucciones por ciclo hace referencia a la cantidad de instrucciones que un procesador ejecuta de manera simultánea por lo que está principalmente limitado a la cantidad de unidades de ejecución que hay en el procesador.

16. Compilador:

Es el proceso de transformar un programa informático escrito en un lenguaje, en un programa equivalente en otro formato; es decir, se encarga de traducir el código fuente de cualquier aplicación que se esté desarrollando.

DIFERENCIAS ENTRE LA ARQUITECTURA DE JOHN VON NEUMANN Y LA ARQUITECTURA PARALELA

Arquitectura de John von Neumann	Arquitectura paralela
Consistió en una sola memoria	Sistema con dos o más procesadores conectados. Un espacio de memoria para los datos y un espacio de memoria para las instrucciones
El conjunto de datos están almacenados en la memoria principal.	El procesador puede acceder separada y simultáneamente a las dos memorias.
Compuesto por un procesador, una memoria, unidades de E/S y un sistema de interconexión.	El procesador tiene unidades funcionales múltiples.
Los programas están formados por instrucciones de máquina.	Trabaja con un mapa de direcciones de instrucciones y un mapa de direcciones de datos separados.
La ejecución de las instrucciones se produce de manera secuencial: después de ejecutar una instrucción se ejecuta la instrucción siguiente.	Comparte la ejecución de una determinada tarea.
El sistema de interconexión es el bus.	Dispone de un sistema de conexión independiente para acceder a la memoria de instrucciones y a la memoria de datos.