Supercomputadoras, ... una visión general

IEC – UTM
Moisés E. Ramírez G.

Introducción

- Una supercomputadora es una computadora que es considerada en su momento de introducción, como lo máximo en capacidad de procesamiento y cálculo, con capacidades muy superiores a la de una computadora de uso común.
- Una supercomputadora es un tipo de computadora muy potente y rápida, diseñada para procesar enormes cantidades de información en poco tiempo y dedicada a una tarea específica.
- La primera vez que apareció el término fue en 1929 en el periodico "New York World" refiriéndose a una computadora construida en la Universidad de Columbia.

- Son las computadoras más caras, su precio alcanza los 30 millones de dólares o más; y cuentan con un control de temperatura especial, esto para disipar el calor que algunos componentes alcanzan a tener.
- Por su alto costo, su uso verdadero está limitado a organismos gubernamentales, militares y grandes centros de investigación, en donde tienen aplicaciones científicas, como en la simulación de procesos naturales (previsión del tiempo, análisis de cambios climáticos, entre otros procesos), modelaje molecular, simulaciones físicas como túneles de viento, criptoanálisis, etc.

Historia

- En 1960, Seymour Cray trabajaba en la CDC (Control Data Corporation) fue quien introdujo la primera súpercomputadora, fue líder durante toda la década hasta que en 1970 Cray formó su propia compañía, **Cray Research**
- Durante 5 años (1985-90), Cray lideréo el mercado de supercomputadoras con sus nuevos diseños.
- Actualmente el mercado de las supercomputadoras está dominado por las empresas IBM y HP que han absorvido a muchas de las empresas que figuraron en la década de los 80's, con la intensión de ganar experiencia.

Seymour Cray

- En 1957, junto con otros ingenieros fundó una nueva compañía denominada Control Data Corporation, en abreviatura CDC, para la cual construyó el CDC 1604, que fue uno de los primeros ordenadores comerciales que utilizaron transistores en lugar de tubos de vacío.
- En 1963, el CDC 6600, batió ampliamente en capacidad de cálculo y en coste al ordenador más potente de que disponía IBM en aquella época.
- En el año 1972 fundó Cray Research, con el compromiso de dedicarse a construir exclusivamente supercomputadores y además de uno en uno, por encargo.
- CRAY-1 (1976) en el Laboratorio Nacional Los Álamos, incorporaba el primer ejemplo práctico en funcionamiento de procesador vectorial, junto con el procesador escalar más rápido del momento, con una capacidad de 1 millón de palabras de 64 bits y un ciclo de 12,5 nanosegundos. Su coste se situaba en torno a los 10 millones de dólares.

- El CRAY-2, entre 6 y 12 veces más rápido que su predecesor. Disponible en 1985, disponía de 256 millones de palabras y 240.000 chips. Su interior se encontraba inundado con líquido refrigerante.
- En el año 1986, existían en todo el mundo unos 130 sistemas de este tipo, de los cuales más de 90 llevaban la marca Cray.
- A mediados de los 80 controlaba el 70% del mercado de la supercomputación. Sin embargo Seymour Cray se encontraba incómodo, pues la problemática empresarial le resultaba escasamente interesante y difícil de soportar. Cedió la presidencia, y dejó la responsabilidad del desarrollo tecnológico de la línea CRAY-2 a Steve Chen, que concibió y construyó los primeros multiprocesadores de la firma, conocidos como serie **X-MP**.

Primeros prototipos de investigación

- Eran equipos que variaban desde arquitecturas SIMD, SIMD/MIMD o completamente MIMD. Los principios usados para la construcción de estos equipos fueron útiles para las siguientes generaciones.

Máquina	Control	Lugar y Fecha
Illiac IV	SIMD	Universidad de Illinois, 1968
MPP	SIMD	Goodyear AeroSpace, 1980
HEP	MIMD	Finales de los 70's
PASM	SIMD/MIMD	Universidad Pardue, 1980
TRAC	SIMD/MIMD	U.T. Austin, 1977
NYU Ultra	MIMD	NYU, 1983
RP3	MIMD	IBM, 1983
Cosmic Cube	MIMD	Caltech, 1980

Tabla 1. Primeras computadoras paralelas

Primera generación

- Estas computadoras fueron en principio proyectos de empresas comerciales.

Máquina	Control	Lugar y Fecha
NCUBE-1	MIMD	NCUBE, 1988
GP1000	MIMD	BBN, 1985
Balance	MIMD	Sequent, 1985
FX/8	Propio	Alliant, 1985
iPSC/1	MIMD	Intel, 1985
SUPRENUM	MIMD	GMD FIRST, 1985
MP-1	SIMD	MasPar, 1985
CM-2	SIMD	TMC, 1985
GF11	SIMD	IBM, 1986

Tabla 2. Primera generación de computadoras paralelas

Segunda generación

- Gracias a la VLSI se pudieron construir computadoras cada vez más veloces y pequeñas.

Máquina	Control	Lugar y Fecha
AP1000	MIMD	Fujitsu, 1991
Cedar	MIMD	Universidad de Illinois
IPSC/2 iPSC/860	MIMD	Intel, 1988/89
NCUBE-2	MIMD	NCUBE, 1992
CM-5	MIMD	TMC 1992
TC2000	MIMD	BBN, 1989
Symmetry	MIMD	Sequent, 1990
FX/2800	MIMD	Alliant, 1990
MP-2	SIMD	MasPar, 1992
KSR1	MIMD	Kendall Square 1989

Tabla 3. Segunda generación de computadoras paralelas

Tercera generación

- Existe una variedad mayor de equipos cada vez más complejos y son cada vez más comerciales. Los sistemas con memoria compartida prácticamente han desaparecido para optar por la memoria distribuida, dispuesta para cada unidad de procesamiento.

Máquina	Procesador	Fabricante	MHz	# elementos de procesamiento	Memoria (MB)
T3D	Dec Alpha	Cray	150	128-2048	16-64
VPP500	Propio GAs, Vector	Fujitsu	100	4-222	128-256
SP1 (SP2)	RS/6000	IBM	62.5	8-64	64-256 (64M – 2GB)
Paragon XP/S	i860XP	Intel	50	4-2048	64-128
Cenju-3	NEC/MIPS CR4400SC	NEC	75	8-256	32-64
GC	Motorola Pwr PC 601	Parsytec	50	16-128	2-32
CS-2	Sparc	Meiko Limited	40	4-1024	32-128

Tabla 4. Tercera generación de computadoras paralelas

Las computadoras más rápidas del mundo

(www.top500.org)

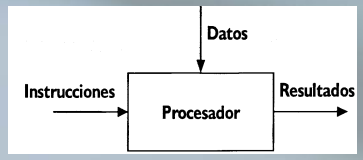
Rank	Site	Computer	# CPU's	Year	R_{max}	R_{peak}
1	DOE/NNSA/LLNL United States	BlueGene/L - eServer Blue Gene Solution IBM	131,072	2005	280,600	367,000
2	Oak Ridge National Laboratory United States	Jaguar - Cray XT4/XT3 Cray Inc.	23,016	2006	101,700	119,350
3	NNSA/Sandia National Laboratories United States	Red Storm - Sandia/ Cray Red Storm, Opteron 2.4 GHz dual core Cray Inc.	26,544	2006	101,400	127,411
4	IBM Thomas J. Watson Research Center United States	BGW - eServer Blue Gene Solution IBM	40,960	2005	91,290	114,688
5	Stony Brook/BNL, New York Center for Computional Sciences United States	New York Blue - eServer Blue Gene Solution IBM	36,864	2007	82,161	103,219
6	DOE/NNSA/LLNL United States	ASC Purple - eServer pSeries p5 575 1.9 GHz IBM	12,208	2006	75,760	92,781
7	Rensselaer Polytechnic Institute, Computional Center for Nanotechnology Innovations United States	eServer Blue Gene Solution IBM	32,768	2007	73,032	91,750
8	NCSA United States	Abe - PowerEdge 1955, 2.33 GHz, Infiniband Dell	9,600	2007	62,680	89,587
9	Barcelona Supercomputing Center Spain	MareNostrum - BladeCenter JS21 Cluster, PPC 970, 2.3 GHz, Myrinet IBM	10,240	2006	62,630	94,208
10	Leibniz Rechenzentrum Germany	HLRB-II - Altix 4700 1.6 GHz SGI	9,728	2007	56,520	62,259.2

Clasificación de Flynn

- Flujo único de instrucciones, flujo único de datos (SISD).
- Flujo único de instrucciones, flujo múltiple de datos (SIMD).
- Flujos múltiples de instrucciones, flujo único de datos (MISD).
- Flujos múltiples de instrucciones, flujos múltiples de datos (MIMD).

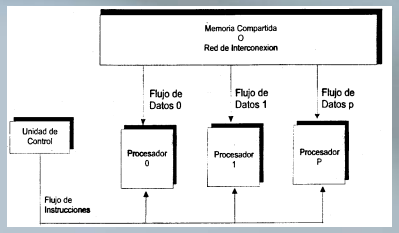
SISD (single Instruction, Single Data)

- Las computadoras que entran en la clasificación SISD tienen un único procesador y ejecutan una sola instrucción a la vez.



SIMD (Single Instruction, Multiple Data)

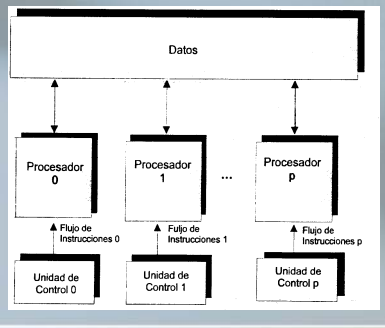
- En esta arquitectura se tienen **p** procesadores idénticos, los cuales poseen una memoria local. Trabajan bajo un solo flujo de instrucciones originado por una unidad central de control, por lo que se tienen **p** flujos de datos .



Processor board of a CRAY YMP vector computer (operational ca. 1992-2000). The board was liquid cooled and is one vector processor with shared memory (access to one central memory)

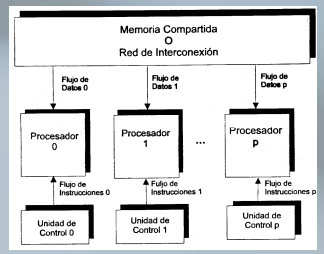
MISD (Multiple Instructions, Single Data)

- Realizan múltiples instrucciones para un solo conjunto de datos. Estas máquinas no han sido construidas por que no es práctico su uso



MIMD (Multiple Instruction, Multiple Data)

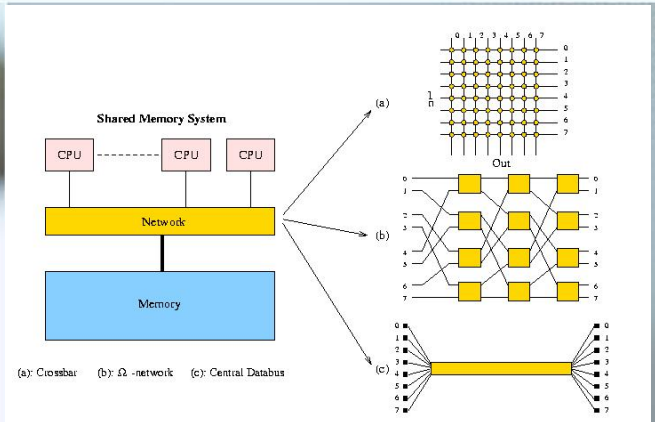
- La arquitectura MIMD está conformada por p procesadores, p flujos de instrucciones y p flujos de datos. Cada procesador trabaja de modo asíncrono bajo el control de un flujo de instrucciones proveniente de su propia unidad de control



Shared Memory Systems

- cada nodo tiene acceso a una amplia memoria compartida que se añade a la memoria limitada privada, no compartida, propia de cada nodo.
- Los sistemas con memoria compartida, tienen multiples CPU's que comparten las mismas direcciones de memoria. Esto significa que que existe una única memoria que es accesada por todas las unidades de procesamiento.
- Los sistemas con memoria compartida pueden ser SIMD o MIMD, en dichos casos se pueden abreviar como SM-SIMD y SM-MIMD respectivamente.
- Para desarrollar programas usando este paradigma, se utiliza OpenMP disponible para C(++) y Fortran

Ejemplos de redes de interconexión para sistemas con memoria compartida.

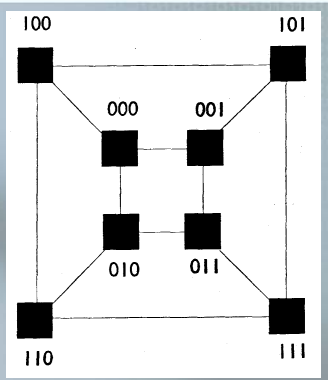


Distributed memory systems

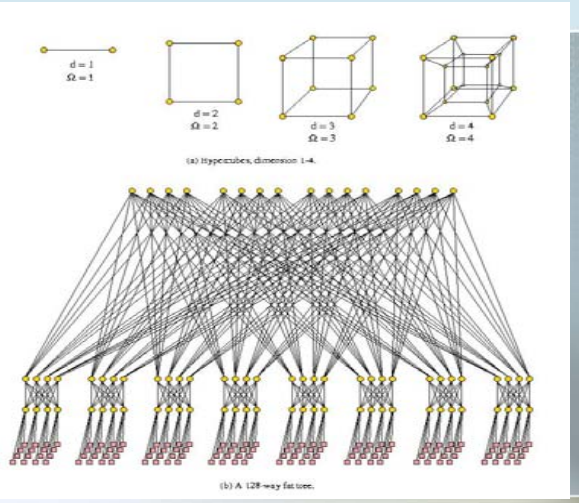
- Cada CPU tiene su propia memoria, los CPU's están interconectados a través de una "red de interconexión" que permite intercambiar datos entre los diferentes nodos cuando se requiere.
- En este tipo de implementaciones, el desarrollador debe controlar donde está cada parte de la información, es decir debe controlar y administrar cada parte de los datos que son distribuidos en todos los nodos.
- Los sistemas distribuidos pueden ser SIMD o MIMD (DM-SIMD o DM-MIMD)
- Los sistemas de interconexión para MIMD, tienen una gran cantidad de topologías de interconexión. En general estas topologías son transparentes para el usuario de tal forma que permita la portabilidad de aplicaciones.

Redes de interconexion

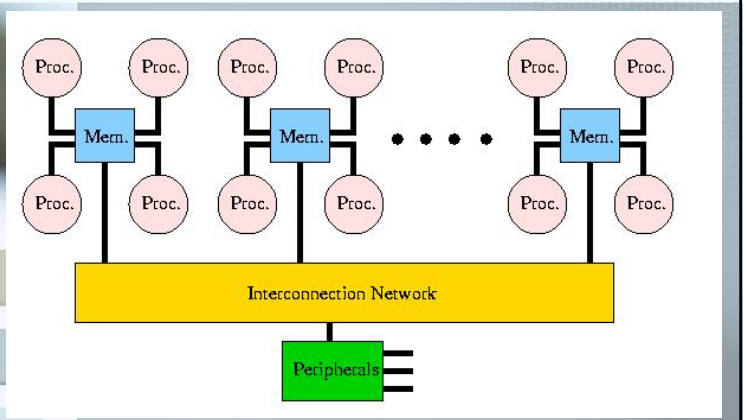
- Arreglo lineal
- Anillo
- Anillo con cuerda
- Árbol binario
- Malla
- Toroide
- Hipercubo (img)



Redes de interconexión para Sistemas con memoria distribuida



NUMA (Non Uniform Memory access)



Clústers

- Simplemente, cluster es un grupo de múltiples computadoras unidas mediante una red de alta velocidad, de tal forma que el conjunto es visto como un único ordenador, más potente que los comunes de escritorio.
- De un cluster se espera que presente combinaciones de los siguientes servicios:
  - Alto rendimiento (*High Performance*)
  - Alta disponibilidad (*High Availability)*
  - Equilibrio de carga (*Load Balancing*)
  - Escalabilidad (*Scalability*)

- La construcción de los ordenadores del cluster es más fácil y económica debido a su flexibilidad: pueden tener todos la misma configuración de hardware y sistema operativo (cluster homogéneo), diferente rendimiento pero con arquitecturas y sistemas operativos similares (cluster semi-homogéneo), o tener diferente hardware y sistema operativo (cluster heterogéneo).
- Para que un cluster funcione como tal, no basta sólo con conectar entre sí los ordenadores, sino que es necesario proveer un sistema de manejo del cluster, el cual se encargue de interactuar con el usuario y los procesos que corren en él para optimizar el funcionamiento

Componentes de un cluster

- En general, un cluster necesita de varios componentes de software y hardware para poder funcionar. A saber:
  - **Nodos** (los ordenadores o servidores)
  - **Conexiones de Red** (Ethernet, Fast Ethernet, Gigabit Ethernet, Myrinet, Infiniband, SCI,...)
  - **Middleware** (capa de abstracción entre el usuario y los sistemas operativos) genera la sensación al usuario de que utiliza un único ordenador muy potente (MOSIX, OpenMOSIX)
  - **Protocolos de Comunicación y servicios**.
  - **Aplicaciones** (pueden ser paralelas o no)

El middleware

- El middleware también debe poder migrar procesos entre servidores con distintas finalidades:
  - **balancear la carga**: si un servidor está muy cargado de procesos y otro está ocioso, pueden transferirse procesos a este último para liberar de carga al primero y optimizar el funcionamiento;
  - **Mantenimiento de servidores**: si hay procesos corriendo en un servidor que necesita mantenimiento o una actualización, es posible migrar los procesos a otro servidor y proceder a desconectar del cluster al primero
  - **priorización de trabajos:** en caso de tener varios procesos corriendo en el cluster, pero uno de ellos de mayor importancia que los demás, puede migrarse este proceso a los servidores que posean más o mejores recursos para acelerar su procesamiento.

Grid

- Es un tipo especial de Clústers que permite utilizar todo tipo de recursos (cómputo, almacenamiento, ...) sin un control centralizado. En general son sistemas muy heterogéneos. Generalmente interconectados por redes de área extensa como Internet.
- Cómputo Grid o Clústers Grid son tecnologías muy relacionadas al cómputo Grid. Las principales diferencias son:
  - Los grid conectan colecciones de computadoras que están geográficamente dispersas o que no son del todo fiables entre sí.
  - Los grids soportan colecciones heterogeneas de equipos (más que los clústers)
  - Los gris son más como una herramienta de cómputo, que como una computadora única.
- El cómputo Grid está optimizado para muchos trabajos independientes que no tienen que compartir datos durante el proceso de cómputo. El trabajo se realiza en cada nodo independientemente del resto del grid.
- Los recursos tales como almacenamiento pueden ser compartidos por todos los nodos, sin embargo resultados inmediatos son almacenados temporalmente y no afectan a otros trabajos que se están ejecutando en otros nodos del grid.
- Ejemplos de Grids:
  - Folding@home project que busca analizar datos usados por investigadores para encontrar curas de enfermedades tales com Alzheimer y cáncer.
  - SETI@home es el grid distribuido más grande en existencia. Usa aproximadamente 3 millones de computadoras personales para analizar datos del radiotelescopio ubicado en el observatorio Arecibo para la búsqueda de vida extraterrestre.

SMP (Multiprocesadores simétricos)

- **Características:**

1. Hay 2 o más procesadores de similares capacidades.
2. Estos procesadores comparten la memoria principal y la E/S y están interconectados por un bus u otro tipo de sistema de interconexión, de tal forma que el tiempo de acceso a memoria para las UP es aproximadamente el mismo
3. Todos los Procesadores comparten la E/S
4. Todos los procesadores pueden desempeñar la misma tarea
5. El sistema está controlado por un sistema operativo integrado, que proporciona la interacción entre los procesadores y sus programas en los niveles de programa, tarea, archivos y datos.

   El S.O. del SMP planifica la distribución de procesos o hilos (threads) entre los procesadores.

Tecnologías relevantes

- Hoy en día el diseño de Supercomputadoras se sustenta en 4 importantes tecnologías:
  - La tecnología de **registros vectoriales**, creada por Seymour Cray, considerado el padre de la Supercomputación, quien inventó y patentó diversas tecnologías que condujeron a la creación de máquinas de computación ultra-rápidas. Esta tecnología permite la ejecución de innumerables operaciones aritméticas en paralelo.
  - El sistema conocido como M.P.P. por las siglas de **Massively Parallel Processors** o Procesadores Masivamente Paralelos, que consiste en la utilización de cientos y a veces miles de microprocesadores estrechamente coordinados.

Tecnologías relevantes (cont)

- La tecnología de **computación distribuida**: los clusters de computadoras de uso general y relativo bajo costo, interconectados por redes locales de baja latencia y el gran ancho de banda.
- **Cuasi-Supercómputo**: Recientemente, con la popularización de la Internet, han surgido proyectos de computación distribuida en los que software especiales aprovechan el tiempo ocioso de miles de ordenadores personales para realizar grandes tareas por un bajo costo.

  A diferencia de las tres últimas categorías, el software que corre en estas plataformas debe ser capaz de dividir las tareas en bloques de cálculo independientes que no se ensamblaran ni comunicarán por varias horas. En esta categoría destacan BOINC y Folding@home, **SETI@home**.

Earth Simulator

- Desarrollado por las agencias japonesas NASDA, JAERI y JAMSTEC y en operación desde finales del año 2001, para aplicaciones de carácter científico siendo utilizado principalmente en simulaciones climáticas y de convección en el interior terrestre.
- Hasta finales del año 2003, ostentó el título de superordenador más rápido del mundo, con una capacitad computacional de más de 35 Teraflops.
  - 5120 CPUs especiales de 500 MHz fabricados por NEC Corporation
  - 640 nodos, con 8 procesadores cada uno
  - 8 GFLOPS por CPU (41 TFLOPS total)
  - 2 GB (4 módulos de 512 MB FPLRAM) por CPU (10 TB total)
  - memoria compartida en cada nodo
  - Switch crossbar 640 × 640 entre los nodos
  - Ancho de banda de 16 GB/s entre los nodos
  - Consumo de energía de 20 kVA por nodo
  - Sistema operativo Super-UX, basado en Unix

MareNostrum

- El superordenador más potente de Europa y el noveno en todo el mundo. Cuando fue puesto en marcha en 2005, el superordenador constaba de 2406 nodos de computación, cada uno de los cuales cuenta con procesadores duales IBM PowerPC 970FX de 64 bits a una velocidad de reloj de 2.2 GHz, 4812 CPUs en total. Actualmente cuenta con con **10,240 procesadores.**
- Los nodos del ordenador se comunican entre sí a través de una red Myrinet de gran ancho de banda y baja latencia. El sistema cuenta con **20 terabytes de memoria central, 280 terabytes de disco**. Utiliza el sistema operativo **Suse Linux**. Capacidad de cálculo de 62,63 teraflops (94,208 teraflops pico). Ocupa una instalación de 160 m² y pesa 40.000 kg.
- El MareNostrum será utilizado en la investigación del genoma humano, la estructura de las proteínas y en el diseño de nuevos medicamentos.
- **Myrinet** es una red de interconexión de altas prestaciones. Desarrollado por Myricom. Una de sus principales características, además de su rendimiento, es que el procesamiento de las comunicaciones de red se hace a través de chips integrados en las tarjetas de red, descargando a la CPU de parte del procesamiento de las comunicaciones.

Kan Balam (super.unam.mx)

- El sistema **HP Cluster Platform 4000**, **"KanBalam"** es la supercomputadora paralela más poderosa de México y América Latina.
  - Capacidad de procesamiento de 7.113 Teraflops (7.113 billones de operaciones aritméticas por segundo).
  - Cuenta con 1,368 procesadores (core AMD Opteron de 2.6 GHz)
  - Memoria RAM total de 3,000 Gbytes y un sistema de almacenamiento masivo de 160 Terabytes.