

PROCESAMIENTO CENTRALIZADO Y DISTRIBUIDO DE DATOS

Procesamiento Centralizado:

En la década de los años 50's las computadoras eran máquinas del tamaño de todo un cuarto con las siguientes características:

- Un CPU
- Pequeña cantidad de RAM
- Dispositivos DC almacenamiento secundario (cintas)
- Dispositivos d salida (perforadoras de tarjetas)
- Dispositivos de entrada (lectores de tarjeta perforada)

Con el paso del tiempo, las computadoras fueron reduciendo su tamaño y creciendo en sofisticación.. Aunque la industria continuaba siendo dominada por las computadoras grandes "mainframes". A medida que la computación evolucionaba, las computadoras, fueron capaces de manejar aplicaciones múltiples simultáneamente, convirtiéndose en procesadores centrales "hosts" a los que se les conectaban muchos periféricos y terminales tontas que consistían solamente de dispositivos de entrada/salida (monitor y teclado) y quizá poco espacio de almacenamiento, pero que no podían procesar por sí mismas. Las terminales locales se conectaban con el procesador central a través de interfaces seriales ordinarias de baja velocidad, mientras que las terminales remotas se enlazaban con el "host" usando módems y líneas telefónicas conmutadas. En este ambiente, se ofrecían velocidades de transmisión de 1200, 2400, o 9600 bps. Un ambiente como el descrito es lo que se conoce como procesamiento centralizado en su forma más pura "host/terminal".

Aplicaciones características de este tipo de ambiente son:

- Administración de grandes tuses de datos integradas
- Algoritmos científicos de alta velocidad
- Control de inventarios centralizado

Al continuar la evolución de los "mainframes", estos se comenzaron a conectar a enlaces de alta velocidad donde algunas tareas relacionadas con las comunicaciones se delegaban a otros dispositivos llamados procesadores comunicaciones "Front End Procesos" (I7EP's) y controladores de grupo "Cluster Controllers" (CC's).

El procesamiento centralizado tenía varios inconvenientes, entre los que podemos mencionar:

- Que un número limitado de personas controlaba el acceso a la información y a los reportes.
- Se requería un grupo muy caro de desarrolladores de sistemas para crear las aplicaciones.
- Los costos de mantenimiento y soporte eran extremadamente altos.

La evolución natural de la computación fue en el sentido del procesamiento distribuido, así las minicomputadoras (a pesar de su nombre siguen siendo máquinas potentes) empezaron a tomar parte del procesamiento que tenían los "mainframes".

Procesamiento distribuido

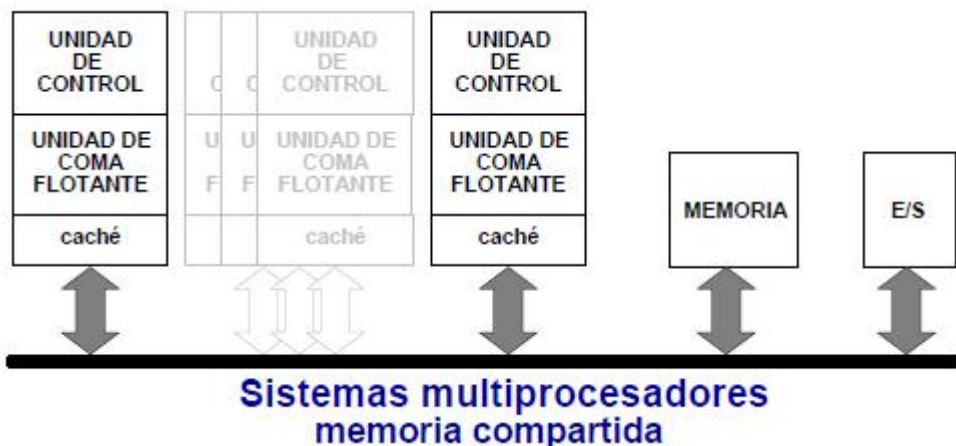
El procesamiento distribuido se define como, la forma en que es posible conectar distintas máquinas, en cierto tipo de red de comunicaciones, generalmente una LAN o una red de área amplia o una red como Internet, logrando así, que una sola tarea de procesamiento de datos pueda ser procesada o ejecutada entre varias máquinas de la red, es decir que un solo proceso se pueda realizar entre varias máquinas diferentes y conectadas a una red.

Un error común es confundir procesamiento distribuido con procesamiento paralelo; el término "procesamiento paralelo", básicamente es el mismo, con excepción que las máquinas distintas tienden a estar físicamente muy cerca en un sistema "paralelo", lo que no es necesario en un sistema "distribuido".

DESCRIPCIÓN

Proceso distribuido, varios procesos ejecutándose en paralelo, en la misma máquina o distribuidos entre computadoras interconectados a través de una red de comunicaciones, colaboran en la realización de una tarea, esta colaboración puede ser tan sencilla como distribuir la carga de trabajo entre procesos idénticos, en el caso de una red de cajeros automáticos, o tan compleja como multitud de procesos distintos, interdependientes, controlando el vuelo de una nave espacial.

El procesamiento distribuido permite una mejor utilización de equipos y mejora el balanceo del procesamiento dentro de una aplicación, este último tiene una gran importancia ya que en algunas aplicaciones simplemente no hay una máquina que sea capaz de realizar todo el procesamiento.



Existe una tendencia inevitable al desarrollo de aplicaciones distribuidas, el procesamiento distribuido permite dispersar los procesadores, datos y otros elementos de una aplicación, la dispersión ofrece un sistema más sensible a las necesidades de los usuarios, capaz de ofrecer tiempos de respuesta mejores y minimizar los costes de comunicación, un sistema distribuido consiste de un gran número de CPU's conectados por medio de una red, un sistema distribuido se encarga del procesamiento cooperativo de solicitudes mediante una colección de computadoras independientes que aparecen ante los usuarios del sistema como una única computadora.

La principal meta es facilitar a los usuarios el acceso a recursos remotos y compartir dichos recursos entre los usuarios, ejemplos de recursos son: computadoras, impresoras, archivos, dispositivos de almacenamiento, páginas web, redes; compartir recursos facilita la colaboración y el intercambio de información.

Transparencia de Distribución

Es la capacidad de un sistema de ocultar el hecho de que sus recursos están físicamente distribuidos en muchos computadores, un sistema distribuido transparente es aquel que se presenta a los usuarios y a las aplicaciones como si fuera un único sistema de computación, este concepto puede ser aplicado a varios aspectos de un sistema distribuido.

Grado de Transparencia

Aspirar a una transparencia total puede ser demasiado: Los usuarios pueden estar en continentes diferentes; con frecuencia esto no es algo que conviene ocultar, ocultar totalmente las fallas de redes y

nodos es imposible ya que no hay manera de distinguir entre un recurso excesivamente lento y uno que ha fallado, mantener la consistencia de múltiples copias de un recurso replicado puede implicar un tiempo significativo, en conclusión, la transparencia es algo que se desea lograr, pero debe ser considerada junto con otros aspectos como el rendimiento.

Escalabilidad

La escalabilidad se refiere a varios aspectos:

- Capacidad del sistema de incrementar el rendimiento ante un incremento de la carga si se agregan recursos adicionales
- Un sistema distribuido "x" es más escalable que un sistema distribuido "y" si utilizando el mismo hardware, "x" puede atender una carga más alta.
-

La escalabilidad puede medirse en tres dimensiones:

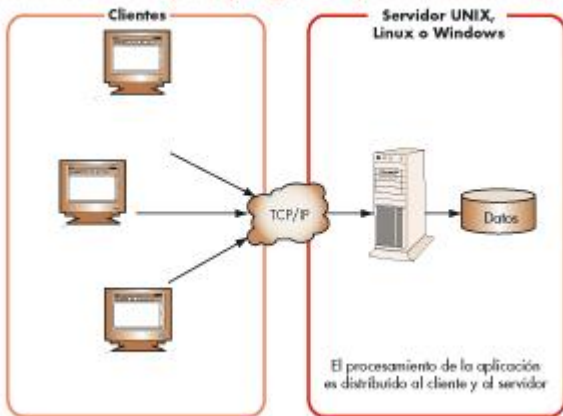
- Con respecto a su tamaño: permite incrementar fácilmente el número de usuarios o procesos soportados.
- Geográficamente: permite que sus usuarios y/o recursos estén alejados geográficamente entre sí.
- Administrativamente: permite cubrir muchas organizaciones administrativas independientes.

La escalabilidad geográfica se ve afectada por los siguientes aspectos: uso de comunicación síncrona y la comunicación en WANs es poco confiable y casi siempre punto a punto, algunos problemas de escalabilidad que se presentan son:

- Servicios Centralizados: Uso de un solo servidor para atender solicitudes de los usuarios
- Datos Centralizados: Tener todos los datos en un único servidor de base de datos.

En ambos casos el servidor se puede volver un cuello de botella si aumenta el número de usuarios, Incluso si el servidor tiene capacidad ilimitada de procesamiento y almacenamiento, las líneas de comunicación desde y hacia el servidor se saturarán y harán imposible el crecimiento.

Modelo de desarrollo y despliegue del procesamiento distribuido



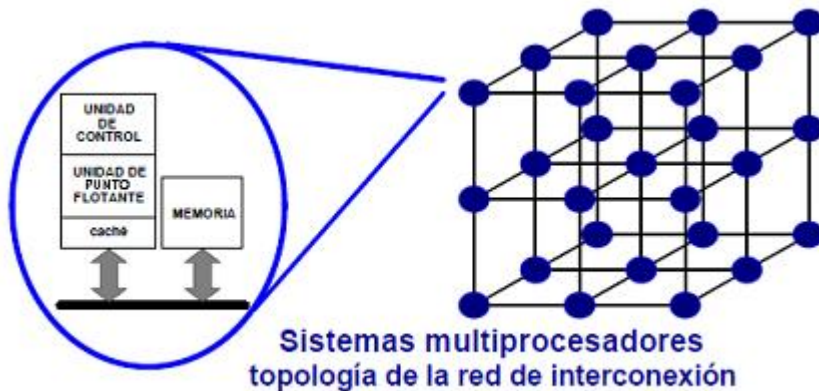
Dentro de la escalabilidad se encuentra el uso de algoritmos que requieran centralizar información proveniente de todos los nodos del Sistema de Distribución, algunos problemas en este caso son las siguientes:

- El transporte de toda la información al punto centralizado haría que una parte de la red se sobrecargue.
- Sólo deben usarse algoritmos descentralizados.
- El uso de algoritmos centralizados está especialmente relacionado con problemas de escalabilidad geográfica.

Por el contrario también se encuentran los algoritmos descentralizados; tienen las siguientes características:

- Ninguna máquina tiene la información completa sobre el estado del sistema.
- Las máquinas toman decisiones basándose únicamente en información local.
- La falla de una máquina no arruina el algoritmo.
- No se asume que existe un reloj global.
-

Para manejar este tipo de procesamiento en las aplicaciones existen diversas maneras, siendo la arquitectura “cliente-servidor” la tendencia actual, e tanto el uso actual de esta arquitectura que por diversas razones, el término “cliente-servidor” ha llegado a aplicarse casi exclusivamente al caso en el que el cliente y el servidor están, en efecto en máquinas distintas.



Una aplicación muy común del procesamiento distribuido es en las bases de datos, donde el procesamiento distribuido podría realizar la entrada/salida, la selección y la validación de los datos en una computadora, y luego crear un reporte basado en esos datos o una consulta en otra computadora.

Estrategias básicas de procesamiento distribuido

Cuando una organización necesita utilizar dos o más sistemas de computación, se puede seguir una de tres estrategias básicas de procesamiento de datos: centralizada, descentralizada o distribuida, con el procesamiento centralizado, todo el procesamiento se lleva a cabo en una sola ubicación o instalación, este enfoque ofrece el más alto grado de control; mediante el procesamiento descentralizado, los dispositivos de procesamiento se colocan en diversas ubicaciones lejanas, los sistemas de computación individuales están aislados y no se comunican entre sí; los sistemas descentralizados son convenientes para compañías que tienen divisiones operativas independientes; en el procesamiento distribuido, las computadoras se ubican en sitios lejanos pero se conectan entre sí a través de dispositivos de telecomunicaciones.

El beneficio del procesamiento distribuido es que la actividad de procesamiento se puede asignar a la ubicación o ubicaciones donde sea posible realizarlo con más eficiencia, mediante el procesamiento distribuido, cada oficina puede organizar y manipular los datos para satisfacer sus necesidades específicas, así como compartir el producto de su trabajo con el resto de la organización.

Sistemas Distribuidos Abiertos

Un sistema distribuido es abierto, ofrece servicios de acuerdo a reglas estandarizadas que describen la

sintaxis y semántica de dichos servicios, debe permitir la interoperabilidad y la portabilidad, la interoperabilidad se refiere a la capacidad de dos sistemas o componentes de desarrolladores distintos para funcionar en conjunto basándose únicamente en los servicios ofrecidos por cada sistema según se especifica en un estándar común, la portabilidad se refiere a la capacidad de una aplicación desarrollada para un sistema distribuido "x", de ser ejecutada en un sistema distribuido "y" que implemente las mismas interfaces de "x".

Un sistema distribuido es abierto, es independiente de la heterogeneidad del entorno subyacente:

- Hardware
- Sistemas Operativos
- Lenguajes de Programación

El clúster es un tipo de procesamiento distribuido "paralelo", es decir, es un conjunto de computadoras o máquinas que pueden trabajar de manera coordinada en la solución de un mismo problema o en la ejecución de un mismo proceso.

Los inicios del clúster fueron cuando los pioneros de la supercomputación trataban de implementar diferentes procesos entre varias computadoras, para luego poder recoger los resultados que dichos procesos debían producir.

Un clúster entonces es un conjunto de computadoras interconectadas con dispositivos de alta velocidad que actúan en conjunto usando el poder de cómputo combinado de varias CPU's para resolver ciertos procesos dados de forma conjunta, y aparecen ante clientes y aplicaciones como un solo sistema, los clúster permiten aumentar la escalabilidad, disponibilidad y fiabilidad de múltiples niveles de red.

Definamos que es un sistema distribuido; este sistema trata a los usuarios como un Sistema Operativo convencional, se ejecuta sobre múltiples CPU's, es transparente: el usuario no ve las múltiples CPU's, Es tolerante a fallos, este tipo de sistema se centra en sistemas de propósito general formados por computadoras autónomas unidos por una red.

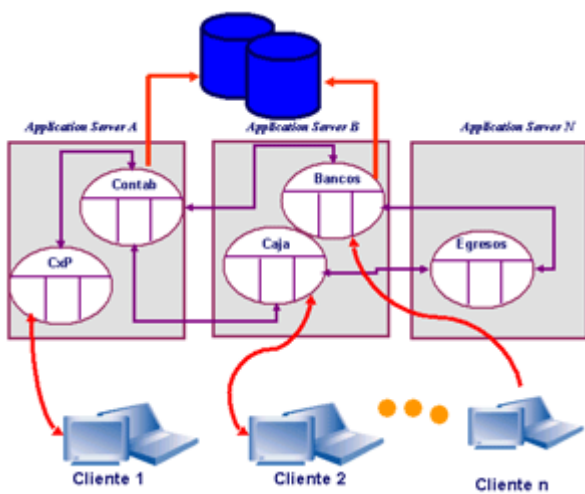
EJEMPLOS DE SISTEMAS DISTRIBUIDOS

Son sistemas a cuyo avance están contribuyendo los sistemas distribuidos, y cuya búsqueda está haciendo avanzar a los sistemas distribuidos.

Sus aplicaciones se dividen en tres categorías importantes con son:

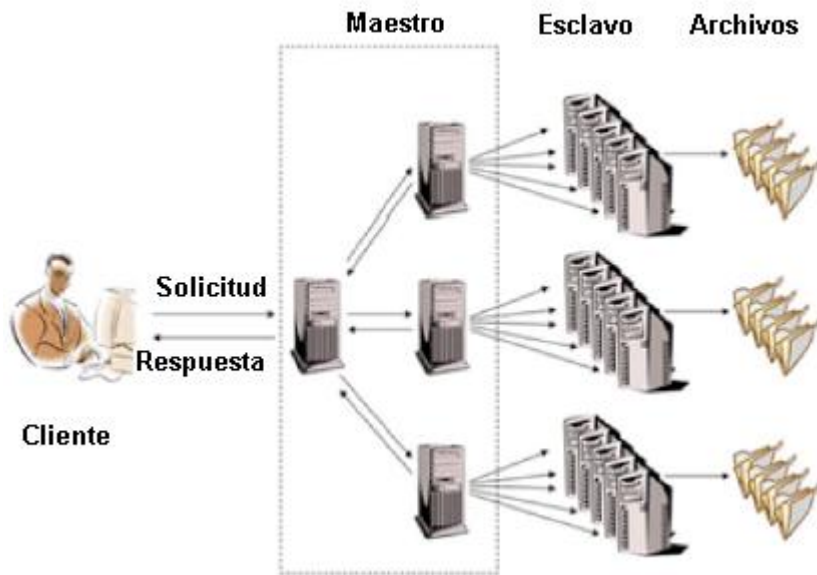
- Aplicaciones comerciales, casos muy variados son:
 - Sistemas de reservas de líneas aéreas.

- Redes bancarias: cajeros automáticos.
 - Hipermercados: suministros, almacén, ventas.
 - Actualmente usan hw, sw y redes dedicadas con estructuras centralizadas.
- Aplicaciones de área global, la más importante en esta área es la Internet que ha cumplido con creces el requisito de la extensibilidad.
 - Aplicación Video-conferencia y acceso a información multimedia
 - Datos basados en tiempo: mantenimiento de tasa, sincronización, compresión.
 - Aplicaciones interactivas: retardo.
 - Requisitos de tiempo-real.



4.5.2. EVOLUCIÓN

En un sistema de bases de datos se almacena la base de datos en varias computadoras, varios medios de comunicación, como las redes de alta velocidad o las líneas telefónicas, son los que pueden poner en contacto las distintas computadoras de un sistema distribuido, no comparten ni memoria ni discos, las computadoras de un sistema distribuido pueden variar en tamaño y función, pudiendo abarcar desde las estaciones de trabajo a los grandes sistemas, dependiendo del contexto en el que se mencionen, existen diferentes nombres para referirse a las computadoras que forman parte de un sistema distribuido, tales como emplazamiento o nodos, para enfatizar la distribución física de estos sistemas se usa principalmente el término emplazamiento.



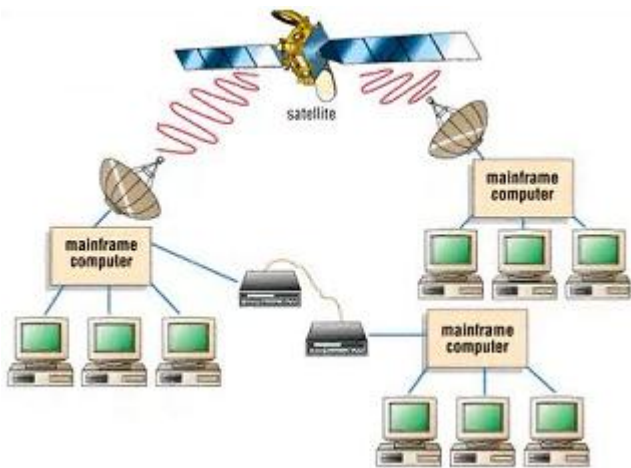
Modelos de Sistema Distribuido

Estaciones y servidores

Es el más común en la actualidad, es aquella en la que a cada usuario, se le asigna una estación; y por otro lado, las estaciones ejecutan las aplicaciones, dan soporte a la interfaz de usuario (GUI), acceden a los servicios compartidos, mediante software de comunicaciones en este modelo los servidores dan acceso a información compartida, dispositivos de hardware compartidos (impresoras, scanner, etc.), funciones de Sistema Operativo (autenticación).

Banco de procesadores

En este modelo los procesadores del banco (PB) tienen una CPU, más suficiente memoria, además no tienen ni discos ni terminales, los usuarios acceden al sistema desde terminales conectados a PAD's; cuenta con un gestor de recursos que controla el acceso a los bancos de procesadores, mediante este gestor el usuario especifica sus requerimientos (CPU, memoria, programa), el gestor de recursos le asigna un procesador como computadora personal.



Minicomputadoras integradas

Este modelo está basado en máquinas multiusuario; el usuario se conecta a una máquina específica,, cada computadora mantiene su autonomía, corre un conjunto completo de software estándar, maneja sus propias aplicaciones y servicios, en este modelo se comparte un esquema de nombramiento global; es decir, el acceso es independiente de la ubicación y el sistema operativo puede migrar y replicar archivos.

Modelos híbridos

Consiste en: Un sistema de Estaciones y Servidores más un Banco de Procesadores, sus características son: un núcleo pequeño, el Sistema Operativo corre como procesos de usuario, servicio de pasarelas a WAN, sus ventajas principales son: recursos de procesamiento ajustados a las necesidades del usuario, ejecución concurrente, acceso a través de terminales.

4.5.3 CARACTERISTICAS

Las características del procesamiento distribuido son:

- Economía: Los microprocesadores ofrecen una mejor proporción precio/rendimiento.
- Velocidad: Un sistema distribuido puede tener un mayor poder de cómputo que un mainframe.
- Distribución Inherente: Algunas aplicaciones utilizan máquinas que están distantes.
- Confiabilidad: Si una máquina se descompone, el sistema sobrevive como un todo.
- Crecimiento por incrementos: Se puede añadir poder de cómputo en pequeños incrementos.
- Datos compartidos: Permiten que varios usuarios tengan acceso a una base de datos común.

- Dispositivos compartidos: Permiten que varios usuarios compartan periféricos caros, como impresoras a color.
- Comunicación: Facilita la comunicación de persona a persona; por ejemplo, mediante correo electrónico .
- Flexibilidad: Difunde la carga de trabajo entre las máquinas disponibles en la forma más eficaz en cuanto a costos.
-

4.5.4 VENTAJAS Y DESVENTAJAS

Existen varias razones que justifican la construcción de sistemas distribuidos de bases de datos, incluyendo el comportamiento de datos, a la autonomía y la disponibilidad.

- **Compartimiento de datos.** La mayor ventaja de los sistemas distribuidos de bases de datos es que proporcionan un entorno en el que los usuarios de un emplazamiento pueden ser capaces de acceder a los datos que residen en otros equipos de cómputo o sistemas.
- **Autonomía.** La ventaja principal del comportamiento de datos, es que cada equipo de cómputo puede conservar un cierto grado de control sobre los datos que tiene almacenado localmente.
- **Disponibilidad.** Si en un sistema distribuido falla un equipo de cómputo, los equipos de cómputo restantes pueden continuar funcionando, en particular, si se duplican los elementos de datos en varios equipos, una transacción que necesite un determinado elemento de datos puede encontrarlo en cualquiera de dichos equipos, de esta manera, el fallo de un equipo no implica necesariamente el cierre del sistema.
-

El sistema debe ser capaz de detectar un fallo en uno de los equipos, de modo que pueda decidir si es necesario realizar alguna acción de recuperación.

El principal inconveniente de los sistemas distribuidos de bases de datos es la complejidad añadida que es necesaria para garantizar la coordinación apropiada entre los equipos. Esta creciente complejidad tiene varias facetas:

- **Coste de desarrollo del software.** La implementación de un sistema distribuido de bases de datos es más difícil y por tanto más costoso.
- **Mayor probabilidad de errores.** Como los equipos que constituyen el sistema distribuido operan en paralelo, es más difícil asegurarse de la corrección de los algoritmos, del funcionamiento

especial durante los fallos de parte del sistema, así como de la recuperación.

- **Mayor sobrecarga de procesamiento.** El intercambio de mensajes y el cómputo adicional necesario para conseguir la coordinación entre los distintos equipos constituyen una forma de sobrecarga que no surge en los sistemas centralizados.
- **Seguridad débil:** Un acceso sencillo también a datos secretos.