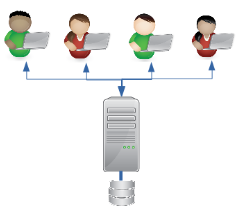
Clase del día - 05/03/2021

La clase de hoy vamos a ver los conceptos de sistema centralizado y sistema distribuido.

**Sistema centralizado**

Un sistema centralizado es aquel dónde el código y los datos residen en una sola computadora.



Un sistema centralizado tiene las siguientes ventajas:

* **Facilidad de programación**. Los sistemas centralizados son fáciles de programar, ya que no existe el problema de comunicar diferentes procesos en diferentes computadoras, tampoco es un problema la consistencia de los datos debido a que todos los procesos ejecutan en una misma computadora con una sola memoria.
* **Facilidad de instalación**. Es fácil instalar un sistema central. Basta con instalar un solo *site*el cual va a requerir una acometida de energía eléctrica, un sistema de enfriamiento (generalmente por agua), conexión a la red de datos y comunicación por voz. Más adelante en el curso veremos cómo el cómputo en la nube está cambiando la idea de instalación física en pos de sistemas virtuales en la nube.
* **Facilidad de operación**. Es fácil operar un sistema central, ya que la administración la realiza un solo equipo de operadores, incluyendo las tareas de respaldos, mantenimiento preventivo y correctivo, actualización de versiones, entre otras.
* **Seguridad**. Es fácil garantizar la seguridad física y lógica de un sistema centralizado. La seguridad física se implementa mediante sistemas CCTV, controles de cerraduras electrónicas, biométricos, etc. La seguridad lógica se implementa mediante un esquema de permisos a los diferentes recursos como son el sistema operativo, los archivos, las bases de datos.
* **Bajo costo**. Dados los factores anteriores, instalar un sistema centralizado resulta más barato que un sistema distribuido ya que solo se pagan licencias para un servidor, sólo se instala un *site*, se tiene un solo equipo de operadores.

Por otra parte, un sistema centralizado tiene las siguientes desventajas:

* **El procesamiento es limitado**. El sistema centralizado cuenta con un número limitado de procesadores, por tanto a medida que incrementamos el número de procesos en ejecución, cada proceso ejecutará más lentamente. Por ejemplo, en Windows podemos ejecutar el Administrador de Tareas para ver el porcentaje de CPU que utiliza cada proceso en ejecución, si la computadora ha llegado a su límite, entonces veremos que el porcentaje de uso del CPU es 100%.
* **El almacenamiento es limitado**. Un sistema centralizado cuenta con un número limitado de unidades de almacenamiento (discos duros). Cuando un sistema llega al límite del almacenamiento se detiene, ya que no es posible agregar datos a los archivos ni realizar *swap*.
* **El ancho de banda es limitado**. Un sistema centralizado puede llegar al límite en el ancho de banda de entrada y/o de salida, en estas condiciones la comunicación con los usuarios se va a alentar.
* **El número de usuarios es limitado**. Un sistema centralizado tiene un máximo de usuarios que se pueden conectar o que pueden consumir los servicios. Por ejemplo, por razones de licenciamiento los manejadores de bases de datos tienen un máximo de usuarios que pueden conectarse, así mismo, el sistema operativo tiene un límite en el número de *descriptores de archivos* que puede crear. Recordemos que cada vez que se abre un archivo y cada vez que se crea un socket se ocupa un descriptor de archivo.
* **Baja tolerancia a fallas**. En un sistema centralizada una falla suele ser catastrófica, ya que sólo se tiene una computadora y una memoria. Cualquier falla suele producir la inhabilitación del sistema completo.

**Ejemplos de sistemas centralizados**

Un servidor Web centralizado

Actualmente los servidores Web suelen ser distribuidos, ya que resulta muy sencillo redirigir las peticiones a múltiples servidores utilizando un balanceador de carga. Sin embargo, todavía los sitios Web pequeños utilizan un servidor centralizado debido a su bajo costo.

Un DBMS centralizado

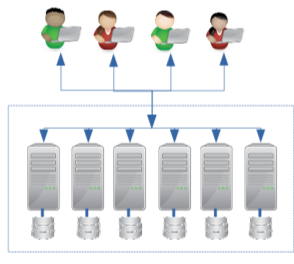
Generalmente los sistemas manejadores de bases de datos (DBMS) son centralizados debido a que resulta más fácil programar sistemas que accedan los datos que se encuentran en una base de datos central. Sin embargo, las plataformas de alcance mundial como Facebook, Twitter, Uber, etc. requieren distribuir los datos en diferentes localizaciones por razones de rendimiento.

Una computadora stand-alone

Una computadora *stand-alone* se refiere a un sistema único integrado. Generalmente entendemos una computadora personal como un sistema stand-alone ya que integra el CPU con un teclado, un monitor, una impresora, etc. Un sistema único es por antonomasia un sistema centralizado.

**Sistema distribuido**

“*Un sistema distribuido es una colección de computadoras independientes que dan al usuario la impresión de constituir un único sistema coherente.*” Andrew S. Tanenbaum



Esta definición de sistema distribuido implica que el usuario de un sistema distribuido tiene la impresión de estar utilizando un sistema central no obstante el sistema estaría compuesto de múltiples servidores interconectados.

La definición anterior tiene importantes implicaciones desde el punto de vista técnico. El hacer que una colección de computadoras se comporten como un sistema único requiere implementar mecanismos de memoria compartida distribuida, migración de procesos, sistemas de archivos distribuidos, entre muchas tecnologías.

De alguna forma, los sistemas distribuidos son antónimos de los sistemas centralizados, de manera que las desventajas de un sistema central son ventajas en un sistema distribuido y viceversa.

Las ventajas de un sistema distribuido son, entre otras:

* **El procesamiento es (casi) ilimitado**. Un sistema distribuido puede tener un número casi ilimitado de CPUs ya que siempre será posible agregar más servidores, por tanto a medida que incrementamos el número de CPUs podemos esperar que los procesos ejecuten más rápido debido a que los procesos ejecutarán en paralelo en diferentes CPUs. El límite del paralelismo queda definido por la **ley de Amdahl**.
* **El almacenamiento es (casi) ilimitado**. Un sistema distribuido cuenta con un número casi ilimitado de unidades de almacenamiento (discos duros). Siempre es posible conectar más servidores de almacenamiento.
* **El ancho de banda es (casi) ilimitado**. En un sistema distribuido cada computadora aporta su ancho de banda, esto es, en la medida que agregamos servidores podemos enviar y recibir una mayor cantidad de datos por unidad de tiempo (es decir, aumentamos el ancho de banda).
* **El número de usuarios es (casi) ilimitado**. El número de usuarios que pueden conectarse a un sistema distribuido aumenta en la medida que agregamos servidores. Si bien es cierto que cada servidor tiene un límite en el número de *descriptores de archivos*, y con ello un límite al número de conexiones que puede abrir, cada servidor en el sistema distribuido agrega descriptores (conexiones).
* **Alta tolerancia a fallas**. En un sistema distribuido la falla de un servidor no es catastrófica, ya que el sistema está diseñado para retomar el trabajo que realizaba el servidor que falla. Más adelante en el curso veremos las estrategias que se utilizan en los sistemas distribuidos para la replicación de datos y la replicación del sistema completo.

Las desventajas de los sistemas distribuidos son:

* **Dificultad de programación**. La definición que hace Tanenbaum de los sistemas distribuidos implica que los usuarios del sistema tienen la impresión de utilizar un sistema único, esto incluye a los programadores. Sin embargo, en la realidad actual los sistemas distribuidos son difíciles de programar ya que el programador es quienel que tiene que implementar la comunicación entre los diferentes componentes del sistema.
* **Dificultad de instalación**. Es complicado instalar un sistema distribuido. Es necesario interconectar múltiples computadoras, lo cual implica la necesidad de una red de alta velocidad.
* **Dificultad de operación**. Es complicado operar un sistema distribuido, ya que se requiere un equipo de administración por cada *site*. Los equipos deberán coordinarse para realizar las tareas de respaldos, mantenimiento preventivo y correctivo, actualización de versiones, entre otras.
* **Seguridad**. Es complicado garantizar la seguridad física y lógica de un sistema distribuido. Tanto la seguridad física como la seguridad lógica requieren la coordinación de múltiples equipos dedicados a la seguridad del sistema. La interconexión remota de los diferentes servidores implica el riesgo de ataques al sistema a través de los puertos de comunicación.
* **Alto costo**. Instalar un sistema distribuido resulta más costoso que un sistema centralizado ya que será necesario pagar licencias para cada servidor, para cada *site*se requiere un equipo de operadores, así mismo, cada *site* requiere su propia acometida de energía, un sistema de seguridad física, infraestructura de refrigeración, etc.

**Tipos de distribución**

Distribución del procesamiento

La distribución del procesamiento permite repartir el cómputo entre diferentes servidores. La distribución del procesamiento se utiliza para el cómputo de alto rendimiento (*HPC: High Performance Computing*), para la implementación de sistemas tolerantes a fallas y para el balance de carga

En el **cómputo de alto rendimiento** los programas se ejecutan en forma distribuida, dividiendo el problema en componentes los cuales se ejecutan en paralelo en diferentes servidores. La clave para obtener rendimientos superiores es que los servidores se conecten mediante una red de alta velocidad.

La distribución del procesamiento permite implementar **sistemas tolerantes a fallas**. Algunos ejemplos de sistemas tolerantes a fallas son los programas que ejecutan en un avión o en una central nuclear. En estos casos los procesos se replican en diferentes computadoras, si una computadora falla entonces el proceso sigue ejecutando en otra computadora.

En el caso de los servidores Web el procesamientos de las peticiones se distribuye con el propósito de **balancear la carga** y evitar que un servidor se sature.

Distribución de los datos

La distribución de los datos aumenta la **confiabilidad**del sistema, ya que si falla el acceso a una parte o copia de los datos es posible seguir trabajando con otra parte o copia de los datos.

La distribución de datos también mejora el **rendimiento**de un sistema distribuido que requiere escalar en tamaño y geografía. Es una buena práctica distribuir los catálogos de los sistemas (por ejemplo, los catálogos de clientes, de productos, de cuentas, etc.) ya que se trata de datos que se modifican poco, por tanto en un sistema distribuido resulta más rápido el acceso a estos datos si los tiene cerca.

**Ejemplos de sistemas distribuidos**

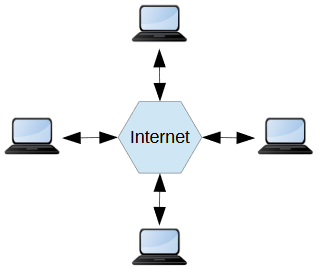
World Wide Web

La web es un sistema distribuido compuesto por servidores (web) y clientes (navegadores) que se conectan a los servidores.

La web permite la distribución a nivel mundial de documentos hipertexto (páginas web) escritos en lenguaje HTML (*Hypertext Markup Language*).

En la web un URL (*Uniform Resource Locator*) permite identificar de manera única a nivel mundial un recurso (página web, imagen, video, etc.).

El protocolo que utiliza el cliente y servidor para comunicarse es HTTP (*Hypertext Transfer Protocol*) el cual funciona sobre el protocolo TCP (*Transfer Control Protocol*).



Cómputo en la nube

En 2006 aparece en la revista Wired el artículo [**The Information Factories**](https://www.wired.com/2006/10/cloudware/) de George Gilder que describe un nuevo modelo de arquitectura basado en una infraestructura de cómputo ofrecida como servicios virtuales a nivel masivo, a este nuevo modelo se le llamó *cloud computing* (cómputo en la nube).

El concepto clave en el cómputo en la nube es el "servicio":

* Infrastructure as a Service (**IaaS**): infraestructura virtual, sistema operativo y red.
* Platform as a Service (**PaaS**): DBMS, plataformas de desarrollo y pruebas como servicio.
* Software as a Service (**SaaS**): aplicaciones de software como servicio

SETI

Seach for Extra-Terrestrial Intelligence ([SETI](https://setiathome.berkeley.edu/)) es un proyecto de la Universidad de Berkeley que integra al rededor de 290,000 (2009) computadoras buscando patrones "inteligentes" en señales obtenidas de radiotelescopios. El sistema alcanza los 617 TFlop/s (1 TFlop/s = 1012 operaciones de punto flotante por segundo).

TOP500

Las [500 computadoras](https://www.top500.org/lists/top500/2020/11/) más grandes del mundo.

Actualmente la computadora más grande del mundo tiene 7,630,848 procesadores con Linux Red Hat; no se trata de una computadora centralizada sino de un sistema distribuido, alcanzando un rendimiento pico de 537,212.0 TFlop/s

.

Actividades individuales a realizar

1. Investigar la ley de Amdahl.
2. En la tarea 1 utilizamos cuatro nodos (los nodos 1, 2, 3 y 4) para calcular una aproximación de PI (notar que el nodo 0 no realiza cálculo, solo actúa como "master"). Suponga que modifica el programa para ejecutar sobre un nodo, dos nodos, tres nodos y cuatro nodos. Si para cada uno de los casos se obtienen los siguientes tiempos: 20 segundos, 10 segundos, 7 segundos y 5 segundos, obtenga la gráfica de aceleración (speedup) de acuerdo a la ley de Amdahl.
3. Considere la lista del TOP500 ¿En qué lugar aparece la primera computadora con Windows? [TOP500 List Statistics](https://www.top500.org/statistics/list/)