

Fundamentos Tecnológicos para el Tratamiento y
Análisis de Datos

Tema 5. Infraestructura de la tecnología de la información

Índice

Esquema

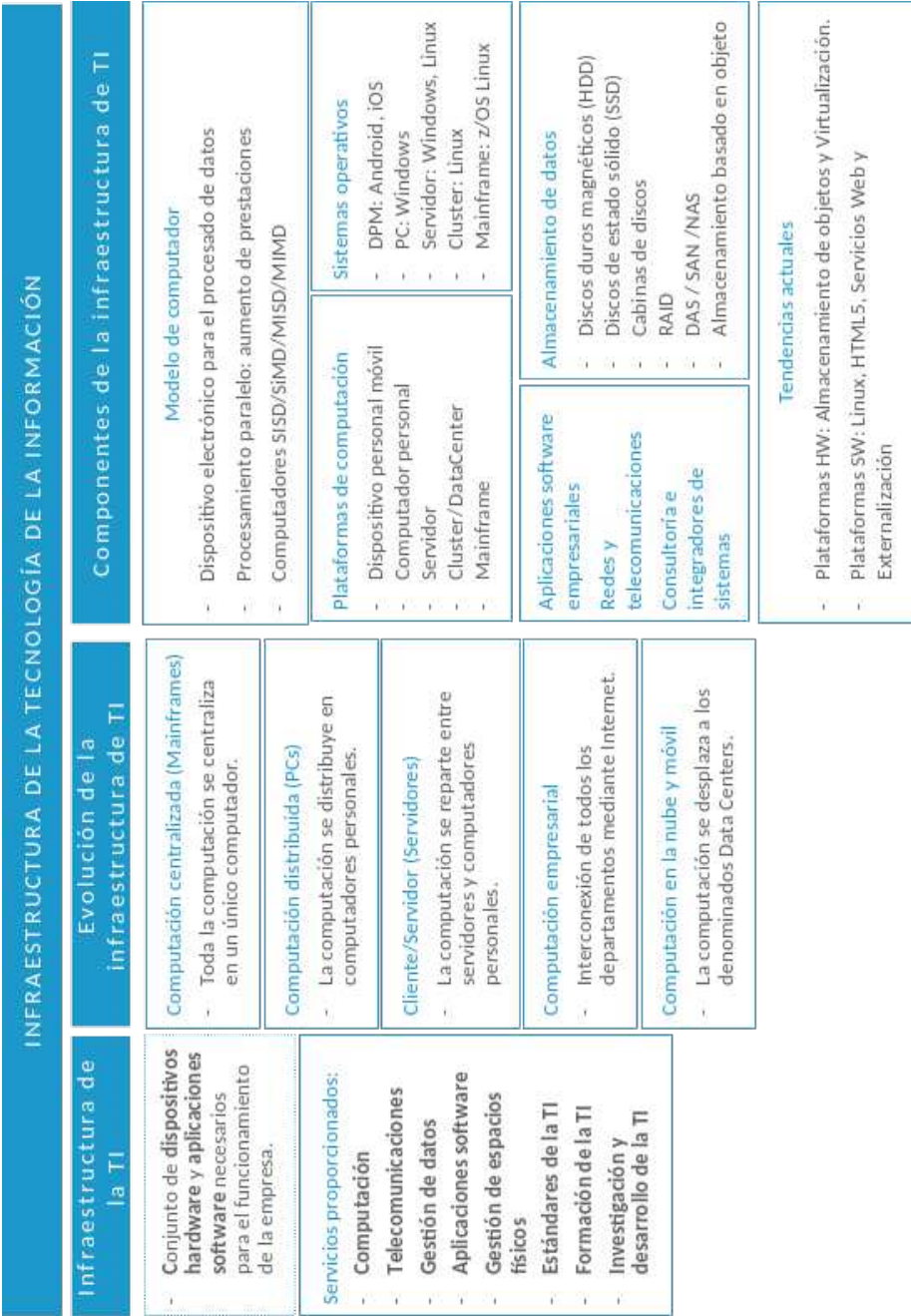
Ideas clave

- 5.1. Introducción y objetivos
- 5.2. Infraestructura de la tecnología de la información
- 5.3. Evolución de la infraestructura de la tecnología de la información
- 5.4. Componentes de la infraestructura de la tecnología de la información
- 5.5. Tendencias de las plataformas hardware y software
- 5.6. Referencias bibliográficas

A fondo

- Las 30 preguntas que debes hacerte antes de comprar un sistema de almacenamiento
- El mainframe vs. la granja de servidores

Test



5.1. Introducción y objetivos

Este tema constituye una introducción a las infraestructuras que componen la tecnología de la información (TI) de una empresa. El objetivo general es que el alumno se familiarice con la terminología usada y con los componentes tecnológicos principales. Se realizará una presentación de componentes, tanto *hardware* (sistemas computadores, sistemas de almacenamiento, etc.) como *software* (sistemas operativos, aplicaciones empresariales, bases de datos, etc.).

En el resto de temas de esta asignatura se desarrollará en profundidad alguna de las secciones del tema actual.

Objetivos del tema:

- ▶ Entender qué servicio proporciona la infraestructura de la TI.
- ▶ Conocer la evolución histórica de los sistemas de computación empresarial.
- ▶ Conocer los componentes de la infraestructura de la TI.
- ▶ Entender qué es un sistema computador y los tipos que hay.
- ▶ Entender la relación entre la computación paralela y la productividad.
- ▶ Conocer los tipos de sistemas de almacenamiento empresarial disponibles.
- ▶ Conocer las tendencias actuales de las plataformas HW y SW.

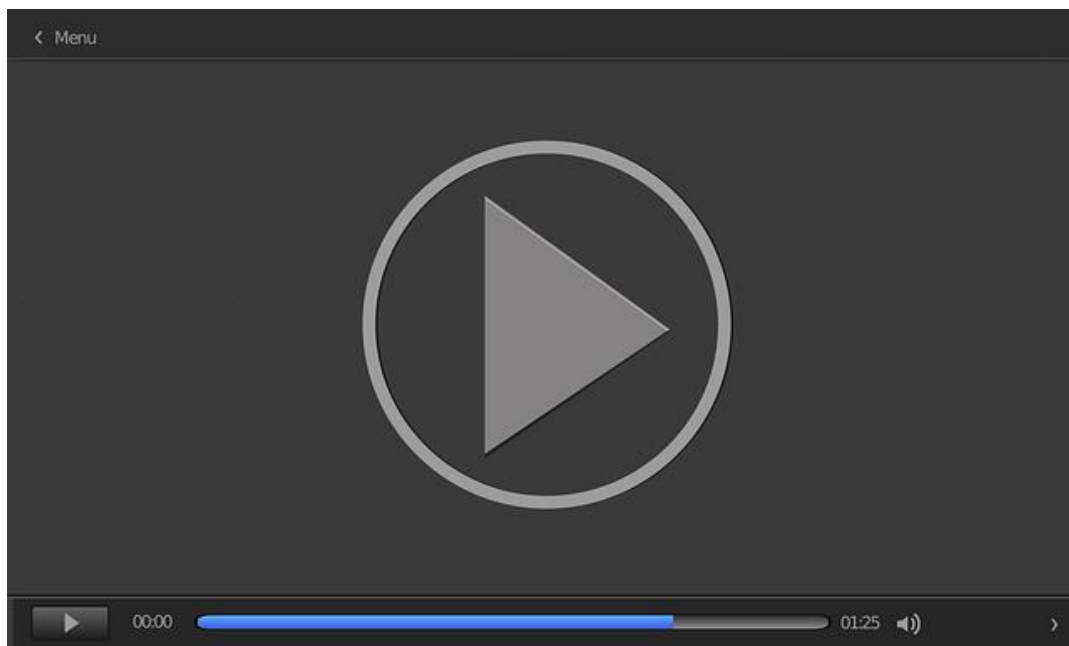
5.2. Infraestructura de la tecnología de la información

La infraestructura de la tecnología de la información está constituida por un conjunto de dispositivos *hardware* y aplicaciones *software* necesarios para el funcionamiento de la empresa. Esta infraestructura se utiliza para proporcionar un conjunto de servicios, entre los que destacan:

- ▶ Entender qué servicio proporciona la infraestructura de la TI.
- ▶ Servicios de computación basados en las diversas **plataformas de computación**. Estas plataformas pueden incluir grandes computadoras (*mainframes*), computadores de tamaño medio, computadores de escritorio, computadores portátiles, tabletas, teléfonos inteligentes y servicios de computación en la nube (*cloud computing*).
- ▶ Servicios de **telecomunicaciones** que proporcionan servicios de conectividad a los empleados, clientes y proveedores mediante datos, voz o vídeo.
- ▶ Servicios de gestión de datos usados para almacenar y gestionar el conjunto corporativo de datos y que pueden proporcionar capacidades de análisis de datos.
- ▶ Servicios basados en **aplicaciones software** que proporcionan a la empresa capacidades tales como la planificación de recursos empresariales (las ERP), la administración de relaciones con el cliente (las CRM), la gestión de la cadena de suministro (SCM) y los sistemas de administración del conocimiento.
- ▶ Servicios para la gestión de **espacios físicos** requeridos por los sistemas de computación, telecomunicaciones y servicios de gestión de datos.
- ▶ Servicios de **gestión de TI** que planeen y desarrollen la infraestructura, se coordinen con las unidades de negocios para los servicios de TI, administren la contabilidad

para los gastos de TI y proporcionen servicios de gestión de proyectos.

- ▶ Servicios de **estándares de TI** que proporcionen a la empresa y a sus unidades de negocios, políticas que determinen qué tecnología de información se utilizará, cuándo y cómo.
- ▶ Servicios de **educación de TI** que provean capacitación, en cuanto al uso del sistema, para los empleados y que ofrezcan a los gerentes instrucción, en cuanto a la forma de planear y gestionar las inversiones en TI.
- ▶ Servicios de **investigación y desarrollo** de TI que proporcionen a la empresa investigación sobre futuros proyectos e inversiones de TI, que podrían ayudar a la empresa a sobresalir en el mercado.



Infraestructura de la Tecnología de la Información

Accede al vídeo:

<https://unir.cloud.panopto.eu/Panopto/Pages/Embed.aspx?id=96daec15-8eaa-43d7-9355-b16c00959eab>

5.3. Evolución de la infraestructura de la tecnología de la información

La evolución de la infraestructura de la tecnología de la información en la empresa ha evolucionado de forma pareja a la evolución de las plataformas de computación. Es posible diferenciar cinco estados en esta evolución, correspondiendo cada uno a una configuración diferente en los elementos de la infraestructura y a una capacidad de computación (ver figura 1).

Las diferentes etapas se denominan en función de la plataforma de computación predominante o bien por la arquitectura u organización de computación utilizada. A continuación, se presentan las principales características de estas etapas.

Computación centralizada o departamental desde 1959 a la actualidad

También se la denomina la era de los *mainframes* y minicomputadores. Se caracteriza por la utilización en la empresa de uno o varios computadores con una gran potencia de cálculo para la época. El comienzo de esta etapa se establece con el desarrollo de los primeros computadores basados en transistores (IBM 1401 y 7090) hacia el año 1959.

El primer *mainframe* de éxito fue la serie IBM 360. Fue el primer computador que disponía de un sistema operativo avanzado que proporcionaba servicios de tiempo compartido, multitarea y memoria virtual.

La principal característica de la computación centralizada es la utilización de un gran ordenador central o *mainframe* al que se conectan hasta cientos de usuarios a través de terminales. Los **terminales** son simples **dispositivos de visualización**, no tienen capacidad de cómputo, ya que toda la computación se realiza en el computador central.

La tecnología de conexión entre los terminales y el computador central es propietaria de cada fabricante, dificultando la interconexión entre diferentes fabricantes. El computador central comparte el tiempo de computación entre todos los usuarios. Estos *mainframes* necesitan de personal operador de sistemas especializado.

Un primer paso hacia el reparto en la capacidad de computación fue la aparición de los denominados minicomputadores, que fueron introducidos en el mercado por la empresa Digital Equipment Corporation (DEC) hacia 1965 bajo el nombre comercial de PDP-11 y posteriormente computadores VAX. Eran computadores de menor precio y capacidad de cómputo que los IBM, pero con capacidades suficientes para adaptarse a las necesidades de departamentos individuales o unidades de negocio en la empresa.

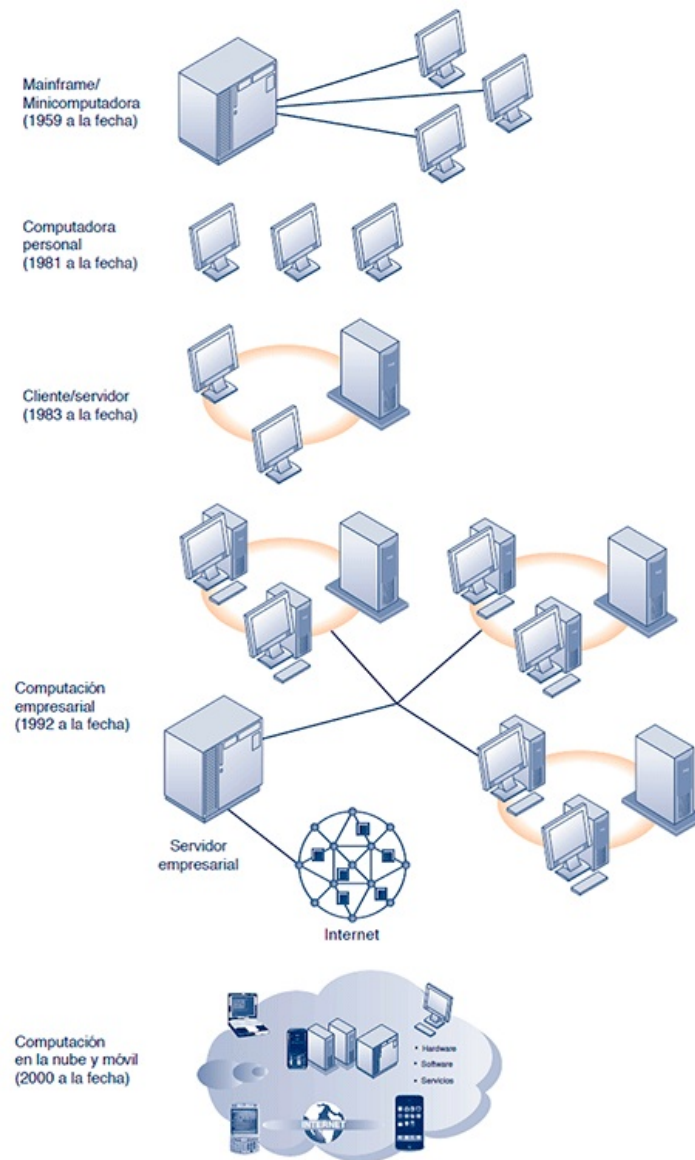


Figura 1. Evolución de la infraestructura de la TI. Fuente: Laudon y Laudon (2014).

Computación distribuida mediante computadores personales desde 1981 a la actualidad

La aparición del PC IBM en 1981 se considera el comienzo de la era de la computación personal, ya que fue el primer computador personal ampliamente aceptado por las empresas. Estos computadores comenzaron utilizando el sistema operativo, basado en comandos de texto, denominado DOS y progresaron a

sistemas operativos Windows, que incluía la utilización de ventanas gráficas y periféricos, como el ratón. Los computadores personales (PC) han sido la plataforma de computación dominante hasta la aparición de las tabletas y los teléfonos inteligentes o *smartphones*.

Un computador personal, también denominado computador de escritorio, proporciona una herramienta de computación individual para cada usuario y representa el paradigma de la computación distribuida no conectada (la computación se realiza en el propio computador del usuario sin necesidad de uno central). Este incluye un sistema operativo sobre el que se ejecutan aplicaciones *software* enfocadas a aumentar la productividad de los usuarios.

Estas aplicaciones incluyen procesadores de texto, hojas de cálculo, *software* de preparación de presentaciones electrónicas y pequeños sistemas de almacenamiento y gestión de datos. Los PC eran plataformas de trabajo aisladas, hasta la aparición de las redes de computadores en la década de los noventa, las cuales permitieron la comunicación entre ellos.

Paradigma cliente-servidor desde 1983 a la actualidad

La aparición de la tecnología de conexión entre computadores y, por tanto, de las redes de computadores facilitó la aparición de la denominada computación cliente-servidor. En la computación cliente-servidor, los PC ya sean de escritorio o portátiles (**clientes**) se conectan a potentes computadores, denominados **servidores**, a través de algún tipo de infraestructura de red de computadores.

Los servidores proporcionan a los clientes una gran variedad de servicios y capacidades. El proceso de computación se reparte entre ambos tipos de máquinas y constituye, por tanto, un punto intermedio entre la computación centralizada en un solo *mainframe* y la distribuida en los PC.

El computador cliente se utiliza como punto de acceso del usuario al sistema. Por su

parte, el servidor se dedica a procesar y almacenar datos compartidos, hospedar servicios como el de páginas web o a la gestión de la red de computadores. El término servidor incluye ambos, las aplicaciones *software* y el computador físico que las contiene.

Aunque el servidor podría ser un computador tipo *mainframe*, en la mayoría de los casos es un único PC de altas prestaciones, lo que incluye alta capacidad de memoria y varios procesadores; todo ello en una misma carcasa. Otra opción es que el servidor sea un conjunto de PC interconectados en lo que se denomina una **cabina o rack**, formando un **clúster** de computadores.

Una de las características importantes de la computación cliente/servidor es su organización o **arquitectura en capas**. La arquitectura más sencilla es la de dos capas. En este caso se tiene un conjunto de computadores cliente conectado, mediante algún tipo de red de computadores, a un único servidor. Este tipo de arquitectura ha sido muy utilizada en pequeñas y medianas empresas. Sin embargo, en las grandes corporaciones se tiende a utilizar organizaciones más complejas denominadas **arquitecturas multicapa o Ncapa**. En este tipo de arquitecturas la carga computacional es balanceada entre los diferentes niveles de servidores, dependiendo del tipo de servicio requerido.

Arquitectura cliente-servidor multicapa

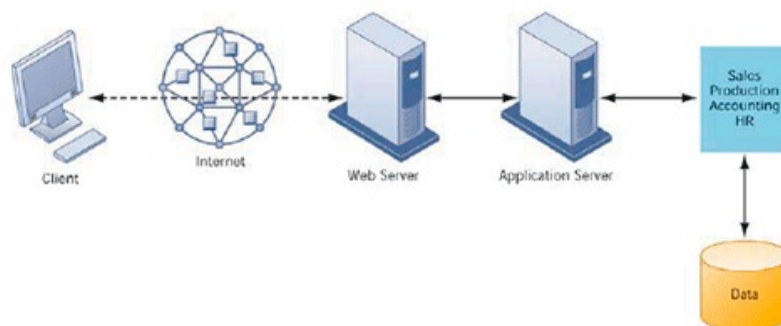


Figura 2. Arquitectura cliente-servidor multicapa. Fuente: Laudon y Laudon (2014).

En la figura 2 se observa una arquitectura multicapa. El servidor web es el encargado de suministrar páginas web al cliente que está ejecutando un navegador web. Por tanto, el servidor web es el encargado de localizar y gestionar el conjunto de páginas web almacenado. Si el cliente necesita acceder a la información incluida en el sistema de información corporativo, por ejemplo, una lista de clientes o precios de productos, dicha petición se realiza a través del servidor de aplicación. Este es el encargado de rescatar esos datos y hacerlos llegar al servidor web, el cual se los entrega al cliente en forma de página web.

En resumen, la arquitectura cliente-servidor posibilita a las empresas distribuir el esfuerzo computacional entre un conjunto de computadores más baratos y pequeños que la alternativa basada en un gran computador *mainframe*.

Desde un punto de vista tecnológico la mayoría de las arquitecturas cliente-servidor se implementan mediante la familia de sistemas operativos Windows, o bien, mediante el sistema operativo Linux y las herramientas *software* desarrolladas bajo ese sistema operativo.

Computación empresarial desde 1992 a la actualidad

El desarrollo del protocolo de comunicaciones TCP/IP (*Transmission Control Protocol/Internet Protocol*) permitió la interconexión de redes de computadores heterogéneas. Esto supuso un paso hacia delante en las infraestructuras de la tecnología de la información empresarial.

La infraestructura resultante enlaza diferentes tipos de *hardware* de computación y pequeñas redes de computadores en una gran red de tamaño empresarial donde la información puede fluir libremente por toda la empresa, o bien, entre la empresa y otras organizaciones. Se pueden conectar diferentes tipos de computadores incluidos

los *mainframes*, servidores, PC y dispositivos móviles. También se tiene acceso a infraestructuras públicas de comunicaciones, tales como el sistema de telefonía e Internet, la red de redes de computadores.

Computación en la nube y móvil desde 2000 hasta ahora

El incremento del ancho de banda en Internet, esto es, el incremento en la velocidad y capacidad de transferir información ha permitido llevar el paradigma cliente-servidor un paso más allá, hacia lo que se denomina el modelo de computación en la nube.

Computación en la nube es un modelo de computación que proporciona acceso a un conjunto compartido de recursos de computación a través de una red de computadores, normalmente Internet. Los recursos accesibles son, entre otros, computadores, almacenamiento, aplicaciones software y servicios. A estos recursos se puede acceder bajo demanda desde cualquier dispositivo conectado a la red. Este tipo de computación está actualmente en claro crecimiento y con las mismas expectativas con vistas al futuro.

Los **centros de datos** (*data centers*) son el corazón de la computación en la nube. Están compuestos por miles o incluso cientos de miles de computadores. Estos sistemas de computación son accesibles mediante los PC, tabletas, teléfonos inteligentes o cualquier otro dispositivo conectado a Internet.

Las empresas que prestan servicios en la nube ofrecen potencia y tiempo de computación, almacenamiento de datos y conexiones de alta velocidad a Internet a las empresas que quieran mantener sus infraestructuras de tecnología de la información de forma remota. Más adelante estudiaremos en profundidad la computación en la nube.

5.4. Componentes de la infraestructura de la tecnología de la información

Este tema está dedicado a los diferentes componentes que hoy en día conforman la infraestructura de la tecnología de la información. En la figura 3 se pueden consultar dichos componentes y algunas de sus principales empresas proveedoras.

Estos constituyen para la empresa una inversión en la que se debe tener en cuenta la coordinación entre ellos a la hora de construir una infraestructura coherente. En el pasado las empresas distribuidoras de tecnología competían unas contra otras ofreciendo sistemas propietarios que incompatibilizaban su interacción.

La tendencia actual es que se busque la colaboración entre distintos distribuidores tecnológicos, con el fin de facilitar la interoperabilidad entre los distintos componentes.

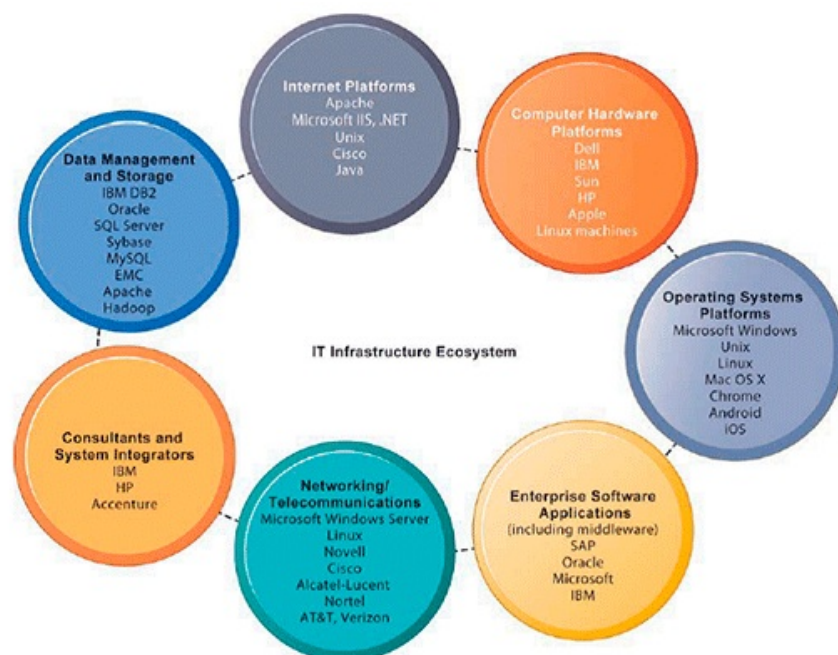


Figura 3. Componentes de la infraestructura de la TI junto con los principales proveedores. Fuente: Laudon y Laudon (2014).

Sistemas computadores y modelos básicos

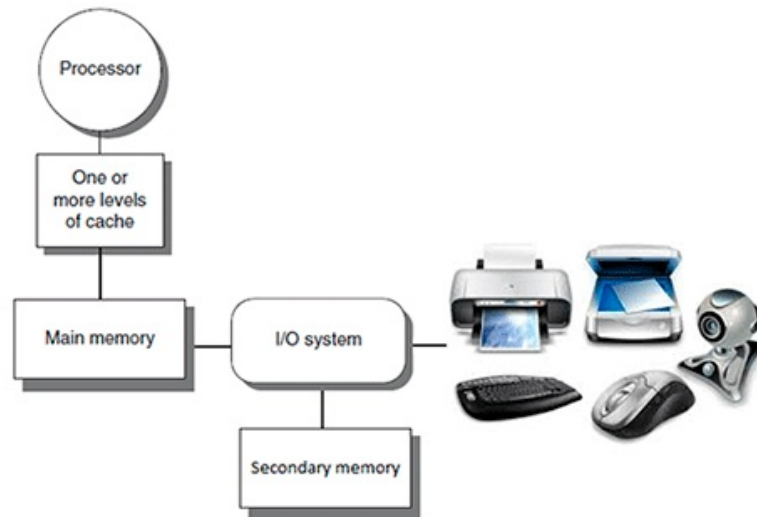


Figura 4. Modelo simple de sistema computador. Fuente: adaptado de Hennessy y Patterson (2012).

Un computador es un dispositivo electrónico digital cuya función principal es realizar **procesamientos de datos**. Dicho procesamiento se realiza siguiendo un conjunto de instrucciones que se denomina **programa**. Tanto el programa a ejecutar como el conjunto de datos se pueden definir por el usuario, por lo que los computadores son dispositivos programables. Un ejemplo de esto último es cuando un usuario termina de utilizar su procesador de texto y comienza a usar una hoja de cálculo, hay un cambio de programa y de datos.

Un computador básico está compuesto por los siguientes elementos: procesador, memoria, sistema de entrada/salida y periféricos (ver figura 4). A continuación, se realiza una breve descripción de dichos elementos:

Procesador

Dispositivo del computador encargado de leer el programa y los datos necesarios de la memoria, ejecutar instrucciones y guardar los resultados obtenidos en la memoria. Tiene asociado un reloj digital cuya frecuencia de referencia define la velocidad de ejecución del programa.

Las instrucciones básicas que realiza un procesador son: leer de la memoria un dato, aplicar una operación a un par de datos (por ejemplo, sumar dos números) y guardar el resultado en la memoria. Un programa de computador es un conjunto de instrucciones de ese tipo.

Memoria principal

Conjunto de circuitos donde se almacenan los programas en ejecución y los datos que están siendo utilizados por el computador. La memoria principal se caracteriza por su capacidad de cambiar la información que contiene y por su rapidez a la hora de proporcionar y almacenar información. Mantiene comunicación directa con el procesador.

Las memorias rápidas suelen ser volátiles, es decir, pierden la información si se corta el suministro eléctrico. También se pueden utilizar memorias no volátiles, cuyo contenido se mantiene a pesar de la pérdida de alimentación energética. Sin embargo, su velocidad de recuperación y almacenamiento suele ser menor.

Con el fin de aumentar el rendimiento de la memoria principal se suele organizar jerárquicamente. Los elementos de memoria más cercanos al procesador y que, por tanto, están en un lugar alto en la jerarquía (denominados **memorias caché**) son de pequeño tamaño, pero de gran velocidad.

Las memorias que ocupan un lugar más bajo en la jerarquía aumentan su tamaño, pero tienden a ser más lentas. La tecnología de las memorias caché es más cara y más rápida que la tecnología con la que se fabrica la memoria principal (más baratas, pero más lentas). Normalmente el computador guarda en la memoria caché las

instrucciones y datos que se utilizan mucho y deja en memoria principal las instrucciones y datos que se utilizan menos.

Memoria secundaria

Conjunto de dispositivos que permiten el almacenamiento masivo de datos y programas. Son sistemas de memoria no volátil de acceso lento, pero de gran capacidad. Tecnológicamente, se basan en discos duros magnéticos, o bien, discos de estado sólido. Los primeros mantienen un sistema electromecánico de almacenamiento, mientras que los segundos son puramente electrónicos (memorias *flash*).

Sistema de entrada/salida

Son el conjunto de circuitos que permiten al computador la comunicación con el exterior. El sistema de entrada/salida hace posible la comunicación con los denominados periféricos, o bien, con las redes de computadores. Los **periféricos** más utilizados hoy en día son el teclado, ratón, monitor, pantallas táctiles, micrófonos, cámaras, etc. Las redes de computadores se tratarán con mayor profundidad más adelante.

Es importante indicar que la memoria principal tiene comunicación directa (y por tanto rápida) con el procesador, mientras que la memoria secundaria lo hace a través del sistema de entrada/salida, lo que le lleva a ser una comunicación más lenta.

Por esta razón, los programas o conjunto de instrucciones que se estén ejecutando en el computador deben trasladarse desde la memoria secundaria a la memoria principal. Este proceso se puede observar en un computador personal cuando cliqueamos una nueva aplicación.

El computador tarda cierto tiempo en mostrarla en pantalla y la mayor parte de ese tiempo se debe a la transferencia de la aplicación desde la memoria secundaria a la principal. Una vez en la memoria principal, la ejecución de la aplicación es ágil desde el punto de vista del usuario.

Procesamiento paralelo y paralelismo en sistemas computadores

La aparición de Internet, el *big data* y el uso masivo de los computadores hace que la necesidad de computación sea mayor. El procesamiento paralelo ha sido la tecnología que ha permitido **aumentar la productividad** de los sistemas computadores.

Procesamiento paralelo

Para explicar qué es el procesamiento paralelo pongamos un simple ejemplo. Supongamos que una instrucción básica que queremos ejecutar en un computador es sumar dos listas de números por parejas. Desde el punto de vista matemático corresponde a sumar dos vectores, por lo que denominaremos a esta instrucción «suma vectorial».

Si disponemos de una única unidad de suma, hay que realizar las sumas de forma secuencial, pareja a pareja. Sin embargo, si disponemos de tantas unidades de suma como de parejas podríamos sumar todas las parejas de forma simultánea. Eso es una ejecución paralela.

La ventaja principal del paralelismo es el considerable **ahorro de tiempo** al ejecutar la instrucción y, por tanto, un aumento de la productividad al conseguir realizar más sumas, en este caso, por unidad de tiempo.

En las aplicaciones informáticas podemos encontrarnos dos tipos de paralelismos:

- ▶ Paralelismo **a nivel de dato** (DLP), que ocurre cuando tenemos muchos datos que pueden ser procesados al mismo tiempo. El ejemplo de la suma anterior (suma vectorial) es de este tipo.
- ▶ Paralelismo **a nivel de tarea** (TLP), que ocurre cuando varias tareas se pueden realizar de forma independiente. Por ejemplo:
 - Una compra por Internet se puede considerar una tarea. Cuando de forma más o menos simultánea varias personas realizan compras, tenemos tareas a procesar en paralelo.
 - La ejecución de varias transacciones bancarias de varios clientes diferentes, situación que se puede dar cuando varios clientes acceden a sus cuentas desde sus respectivos teléfonos móviles.

Desde el punto de vista del procesamiento paralelo podemos clasificar a los computadores en los siguientes grupos:

SISD (Single instruction stream-single data stream)

Este tipo de computadores también son conocidos como de **uniprocador**: ejecutan secuencialmente las instrucciones y las operaciones con los datos, por lo que no permite paralelización ni de datos (DLP) ni de tareas (TLP). En nuestro ejemplo de instrucción de sumar dos listas de números, las sumas se realizan de forma secuencial. Hoy en día únicamente los sistemas computadores de menor capacidad pertenecen a esta categoría.

SIMD (Single instruction stream-multiple data stream)

La misma instrucción es ejecutada por múltiples unidades operativas que procesan en paralelo múltiples datos. Corresponde a la paralelización de nuestro ejemplo de instrucción de suma vectorial. Este tipo de tecnología paralela se explota mucho en las GPU, *hardware* especializado en procesamiento de gráficos, por ejemplo, para

videojuegos. Este tipo de procesamiento corresponde al paralelismo a nivel de dato (DLP).

MISD (Multiple instruction stream- single data stream)

Este tipo de computador no se comercializa.

MIMD (Multiple instruction stream-multiple data stream)

Este tipo de computadores ejecuta en paralelo varias instrucciones, cada una de ellas con sus propios datos. Con un computador MIMD se podrían **ejecutar de forma simultánea varias instrucciones** de suma vectorial de nuestro ejemplo.

Los computadores MIMD consiguen paralelismo a nivel de tarea (TLP), dotándole de mayor flexibilidad y aplicabilidad que los computadores SIMD. Por supuesto, esto se consigue incluyendo en el computador varios procesadores (computador multiprocesador) o un procesador con varios núcleos (multinúcleo o *multicore*), lo que encarece el precio del computador.

Este tipo de computadores es el que domina el mercado hoy en día. Podemos encontrar computadores MIMD tanto en dispositivos personales como las tabletas o en computadores de altas prestaciones.

Los computadores MIMD se pueden clasificar de la siguiente forma:

- ▶ Multiprocesadores con memoria principal compartida.
 - Multiprocesadores con memoria-compartida centralizada (SMP).
 - Multiprocesadores con memoria-compartida distribuida (DSM).
- ▶ Multiprocesadores con memoria principal independiente.

En la figura 5 se puede ver un esquema de un computador con multiprocesadores de memoria-compartida centralizada. Este tipo de computadores tienen pocos

procesadores (hasta 8 suele ser un número típico) y todos estos comparten, de forma simétrica, el acceso a la memoria principal.

Un chip multinúcleo, un chip que incluye varios procesadores en su interior, normalmente es SMP, todos los núcleos comparten la misma memoria principal. Al estar centralizada la memoria, el tiempo de acceso de cualquier procesador al contenido de la memoria es el mismo.

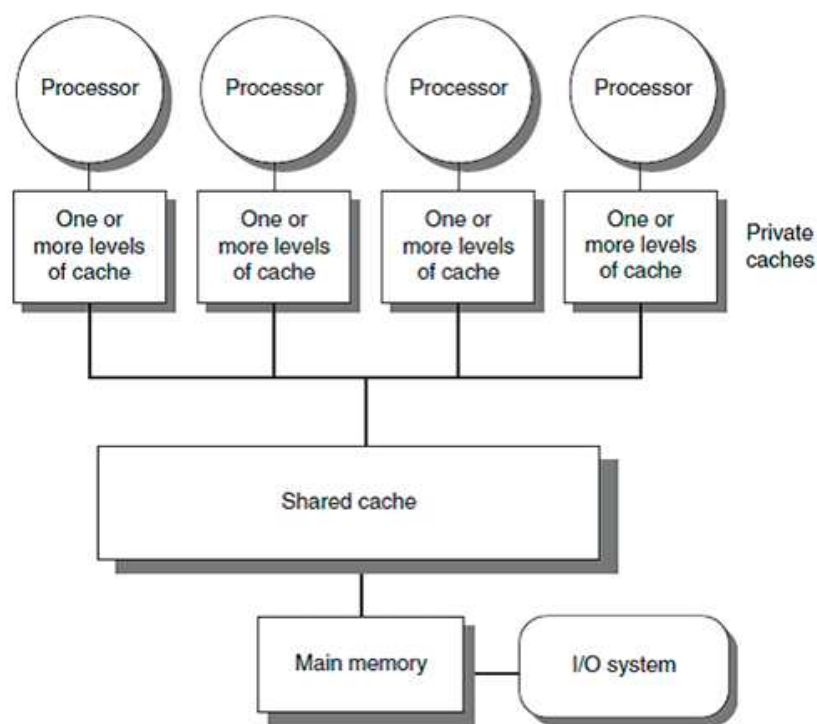


Figura 5. Modelo de computador multiprocesador con memoria-compartida centralizada. Fuente: Hennessy y Patterson (2012).

La figura 6 muestra un computador multiprocesador con memoria principal compartida-distribuida (DSM). Cada procesador tiene su propia parte o trocito de la memoria principal, la cual no está centralizada, sino distribuida en torno a los múltiples procesadores.

Los computadores con dos o más procesadores multinúcleo optan por una estructura

DSM. Al estar distribuida la memoria, el tiempo de acceso de cualquier procesador al contenido de la memoria depende de a qué bloque de memoria desea acceder. Un acceso a «su» bloque es más rápido que un acceso al bloque de otro procesador.

Nótese que la red de interconexión entre procesadores es una red especializada en este tipo de comunicaciones, no es una red de transferencia de datos estándar como la que nos podemos encontrar en una oficina para conectar computadores o realizar una conexión a Internet.

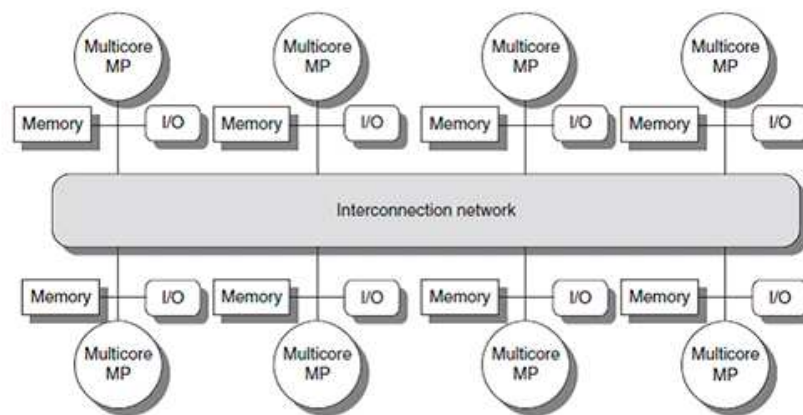


Figura 6. Modelo de computador multiprocesador con memoria-compartida distribuida. Fuente: Hennessy y Patterson (2012).

Por último, la figura 7 representa un tipo de computador multiprocesador con memoria principal independiente. En este caso, cada procesador tiene su propia memoria principal. En la imagen se puede observar que un computador de este tipo es la agrupación de varios computadores interconectados mediante una red de computadores de área local.

En este caso, la red de conexión sí puede ser del mismo tipo económico de red de computadores que nos podemos encontrar en una oficina. Esto significa que para comunicar dos procesadores o nodos en este tipo de computadores es necesario utilizar el sistema de entrada/salida y los protocolos de comunicación de la red de computadores utilizada, con la consecuente pérdida de tiempo.

En cambio, en los multiprocesadores de memoria compartida, dos procesadores se pueden comunicar directamente mediante la memoria principal que comparten, sin usar el sistema de entrada/salida.

La siguiente sección presenta las plataformas de computación usadas hoy en día. Indicar que la mayoría pertenecen a la familia MIMD (normalmente los computadores personales, servidores y *mainframes* computadores MIMD de memoria compartida), mientras que los clústeres son MIMD de memoria independiente.

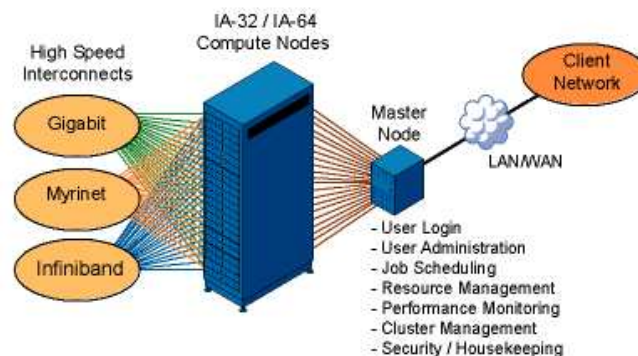


Figura 7. Modelo de computador multiprocesador con memoria independiente. Fuente:

http://www1.ap.dell.com/content/topics/topic.aspx/ap/topics/pub_solutions/hied/high_performance_computing?c=ap&l=en&s=gen&~page=6

Tipos de sistemas computadores

Dispositivo personal móvil

En este tipo de computador se incluyen dispositivos inalámbricos como los teléfonos inteligentes, tabletas, etc. El coste de estos dispositivos ronda los cientos de euros. Su diseño pone especial énfasis en la eficiencia energética, con el objetivo de alargar la duración de las baterías que les proporcionan energía y la optimización del uso de la memoria.

Utilizan memorias electrónicas perdurables, en vez de discos magnéticos, por lo que su capacidad es limitada. Las aplicaciones incluidas en este tipo de dispositivos

están orientadas a la web y al procesamiento de audio y vídeo.

Computador personal

Computador de uso personal con capacidades de computación, almacenamiento y conexión a la red. El coste de este tipo de computadores varía según lo que cueste un ultraportátil hasta lo que puede alcanzar una estación de trabajo bien equipada que dispone de alta capacidad de computación y almacenamiento.

El diseño de este tipo de computadores tiende a optimizar la pareja rendimiento-precio. Esto se traduce en que los nuevos procesadores de alto rendimiento y coste reducido suelen instalarse primero en este tipo de computador.

Servidores

Computadores dedicados a proporcionar servicios de computación y almacenamiento a gran escala y fiabilidad.

En los servidores varias son las características a tener en consideración:

- ▶ Primero la **disponibilidad**, ya que los servidores deben operar siete días a la semana, veinticuatro horas al día. La alta disponibilidad de un servidor se consigue mediante la duplicación o redundancia del *hardware* interno, lo que obviamente encarece el producto.
- ▶ La segunda característica es la **escalabilidad**. Los servidores deben ser capaces de responder adecuadamente a un incremento en la demanda de los servicios que soportan. Por tanto, la habilidad para incrementar la capacidad de cómputo, la memoria principal, la de almacenamiento y el ancho de banda en las operaciones de E/S es crucial.
- ▶ Finalmente, los servidores deben tener una **alta productividad**. Deben alcanzar un gran rendimiento en términos de transacciones por unidad de tiempo o páginas web servidas por unidad de tiempo. La respuesta a una petición individual es importante,

pero la eficiencia global se mide por el número de peticiones por unidad de tiempo que son capaces de atender. La mayoría de los servidores se basan en la arquitectura x86 que fue diseñada por la empresa Intel.

Clústeres

Un clúster es una colección de computadores personales o servidores conectados mediante redes de área local para actuar como un solo gran computador. La idea básica es construir un computador potente, uniendo la fuerza computacional de un conjunto de computadores de potencia media baja y, por tanto, de poco coste.

Cada nodo-computador ejecuta su propio sistema operativo y las comunicaciones se realizan a través de un protocolo de comunicación. Por supuesto, un clúster debe disponer de *software* especializado, que atienda las peticiones de servicio que se realizan al clúster y reparta el trabajo entre todos los nodos-computadores.

Un clúster se puede configurar para diversas funcionalidades. A continuación, se exponen tres de ellas.

- ▶ Clúster para **alta disponibilidad**. Como en el caso de los servidores, la disponibilidad es crítica en este tipo de computadores. A diferencia de los primeros, en un clúster la disponibilidad se consigue por la combinación de redundancia en componentes baratos (los propios nodos-computadores) y un *software* especializado que localiza los fallos y activa o desactiva nodos en función de las necesidades.

N-Redundancia

Por ejemplo, un clúster puede tener 10 nodos activos y 4 no activos. Los nodos no activos son de respaldo que se activan ante fallos de los activos. Este caso se llama 4-redundancia pues pueden fallar 4 nodos, pero el clúster sigue dando el mismo servicio.

- ▶ Clúster para **balanceo de carga**. En este caso el objetivo del clúster es prestar un servicio, realizando un reparto del trabajo entre todos los nodos disponibles. Es un concepto muy cercano al de granja de servidores. Una granja de servidores es un conjunto de servidores que proporcionan un servicio (por ejemplo, servir página web) y una aplicación *software* que reparte trabajo entre ellos.
- ▶ Clúster para **computación de altas prestaciones**. Este tipo de clúster se utiliza en aplicaciones que necesitan de gran potencia de cómputo. Son aplicaciones científicas, de modelado, de procesamiento gráfico, etc.

Una de las principales características de un clúster es que la mayoría de las funcionalidades que proporciona las realiza el *software*, debido a la simplicidad *hardware* de los computadores que componen el clúster. El no disponer de *hardware* especializado puede implicar un límite en la productividad alcanzada por el clúster.

A un *data center* se le puede considerar un clúster compuesto por decenas de miles de computadores. Grandes empresas como Google, Amazon, Facebook o Twitter disponen de *data centers* que proporcionan servicio a millones de usuarios.

Mainframes

Aunque fueron una de las principales estructuras de computación, los *mainframes* no han desaparecido. Básicamente, un *mainframe* es un computador centralizado con **múltiples capas de redundancia** en los procesadores, módulos de E/S, sistemas de alimentación, refrigeración y baterías de respaldo. Su principal característica es que está diseñado con *hardware* especializado en incrementar la productividad, esto es, carga de trabajo realizada por unidad de tiempo.

Los *mainframes* están especializados en realizar un alto volumen de trabajos por lote y procesamiento de transacciones en línea. Un ejemplo de trabajo por lotes típico en un banco es la actualización de las cuentas bancarias de sus clientes tras el pago de recibos, nóminas, intereses, tarjetas, etc.

Una transacción es una operación con datos que consta de varios pasos. Lo importante es que se debe realizar de una sola vez (no se puede interrumpir), se debe acabar una vez empezada y varias transacciones simultáneas no se pueden interferir.

Un ejemplo clásico sería un traspaso bancario, que implica los siguientes pasos:

- ▶ Comprobar si hay fondos.
- ▶ Restar el traspaso de una cuenta.
- ▶ Sumar el traspaso a otra.

Cuando un cliente accede *online* a su cuenta y realiza operaciones, está realizando transacciones *online*. Los *mainframes* más potentes están diseñados para realizar billones de transacciones y operaciones por lote en un día. Esta es la razón de por qué muchas de las mayores compañías del mundo y, entre ellas, los principales bancos siguen confiando en este tipo de computador.

Desde el punto de vista comercial, el IBM z Systems es la familia de *mainframes* que dominan el mercado. Es importante recordar que el mantenimiento de un *mainframe* necesita de personal y servicios especializados. El principal negocio de la empresa IBM no es la venta de *mainframes*, sino los servicios que realiza a las empresas que los adquieren.

Actualmente existe cierta competencia entre los *mainframes* y los clústeres de computadores. Mientras los *mainframes* aportan una solvencia muy contrastada en ciertas áreas empresariales y basan sus servicios en *hardware* especializado y, por tanto, más costoso, los clústeres han tenido éxito en recientes aplicaciones (por ejemplo, redes sociales o buscadores web) y basan su estrategia en *hardware* económico, compensado con más cantidad de *software* de servicio.

Sistemas operativos

El sistema operativo es el conjunto de programas informáticos que gestiona los recursos *hardware* del computador y proporciona una interfaz de comunicación con los usuarios. El resto de las aplicaciones o programas *software* se ejecutan sobre el correspondiente sistema operativo. Dependiendo del tipo de computador existen varias opciones a la hora de seleccionar un sistema operativo.

A continuación, se realiza un breve repaso de las principales opciones:

Dispositivo personal móvil

Android es un sistema operativo de gran uso en tabletas y teléfono móviles inteligentes desarrollado por la Open Handset Alliance, liderada por Google. Compite con el iOS que es el sistema operativo de Apple, incluido en sus teléfonos móviles y tabletas.

Estos sistemas operativos están diseñados para interactuar con el usuario mediante la tecnología táctil, la pantalla sirve de periférico de salida al mostrar imágenes, pero también de entrada al permitir al usuario interactuar con los dedos, incluso a través de los denominados teclados virtuales. Otras empresas como Microsoft o BlackBerry también disponen de sistemas operativos para este tipo de dispositivos.

Computador personal

La mayoría de los computadores personales incluyen alguna versión del sistema operativo Windows de Microsoft. Tradicionalmente se han diseñado para interactuar con periféricos del tipo teclado, ratón y pantalla. Apple también desarrolla el macOS para sus computadores Mac.

Servidores

En la gama de los servidores, el mercado se reparte entre la familia de sistemas operativos de Microsoft Windows orientados a servidor, o bien, sistemas operativos

de la familia Unix o Linux que es una versión de Unix de fuente abierta, económica y robusta. Unix y Linux son sistemas operativos escalables, fiables y mucho más económicos que la familia Windows o un sistema operativo de *mainframe*.

Mainframes

En el caso de IBM, sus *mainframes* soportan la familia z/OS un sistema operativo propietario diseñado para proporcionar seguridad, estabilidad y alta disponibilidad. IBM también ofrece las versiones z/VSE y z/TPF de su sistema operativo orientadas al procesamiento seguro de grandes volúmenes de transacciones y cargas de trabajo por lotes.

Así mismo, IBM también incluye Linux on Z y z/VM que, mediante virtualización, permite la convivencia de miles de Linux y el hospedaje simultáneo con z/OS, z/VSE y z/TPF. De esta forma se tiene la potencia en servicios de un data center concentrada en un computador *mainframe*.

Clúster

Tanto Microsoft Windows como GNU/Linux ofrecen soluciones para conectar un conjunto de computadores económicos en clúster.

Aplicaciones software empresariales

Esta sección es un pequeño resumen de los sistemas de información para la gestión.

El primer tipo son los **sistemas para el procesamiento de transacciones (TPS)**. Son las primeras aplicaciones en aparecer en los *mainframes* de la década de los sesenta y son ampliamente utilizadas hoy. Se encargan de realizar y grabar las transacciones comerciales que se realizan en la empresa y están en contacto directo con los sistemas de almacenamiento de datos.

Ejemplos de su aplicación son: la gestión de las cuentas de los clientes en los

bancos, el mantenimiento y seguimiento de ventas y *stocks* en supermercados. Los gerentes los utilizan para tareas como facturar a clientes o pagar a proveedores.

Otro tipo de aplicaciones *software* empresariales son los denominados **sistemas de información gerencia (MIS)**. MIS proporciona a los gerentes, de nivel medio en la empresa, informes del actual rendimiento de la organización. Prácticamente, las aplicaciones MIS resumen e informan de las operaciones básicas de la empresa, utilizando datos proporcionados por los TPS. Por lo tanto, la mayoría de los MIS realizan informes basados en operaciones simples como sumas y comparaciones de cifras generadas en los TPS y responden a preguntas simples, pero muy recurrentes en gerencia.

Otro tipo de aplicaciones informáticas son los **sistemas de apoyo a la decisión (DSS)**. Van más allá de los tradicionales MIS. Intentan dar respuesta a problemas o preguntas que no se pueden anticipar. Básicamente, un DSS toma como entrada datos aportados por TPS y MIS e información externa que sea de interés. El DSS también dispone de un modelo matemático, más o menos complejo, que proporciona respuestas. Otros DSS utilizan modelos analíticos de datos históricos de la compañía para proporcionar las respuestas.

En el más alto nivel se tienen las **aplicaciones de apoyo a los ejecutivos (ESS)**. Son aplicaciones informáticas que presentan gráficas y datos de diversas fuentes, en una interfaz sencilla de usar para gerentes de alto nivel. Los ESS utilizan información resumida de la proporcionada por los MIS y DSS e información externa sobre competidores, impuestos, etc. Estos sistemas pueden incluir analítica de inteligencia de negocio para analizar posibles tendencias o predicciones.

Por otro lado, se tienen las aplicaciones necesarias para enlazarlas diferentes facetas que tiene una empresa. Estas aplicaciones incluyen:

- Planificación de recursos empresariales (los ERP).

- ▶ La administración de relaciones con el cliente (los CRM).
- ▶ La gestión de la cadena de suministro (SCM).
- ▶ Los sistemas de administración del conocimiento.

Infraestructura de redes y telecomunicaciones

Más adelante, se profundizará en las plataformas de redes y telecomunicaciones empresariales. En el mundo de las redes de computadores de área local, el sistema operativo predominante es Windows es su versión Server, seguido de Linux. Los dispositivos que implementan las redes empresariales de área extensa utilizan alguna variante de Unix como, por ejemplo, Linux. Desde el punto de vista de los protocolos de comunicación, la familia TCP/IP es la mayoritaria debido a la conectividad que proporciona Internet.

Parte del *hardware* necesario para crear redes de computadores son los *routers*, *switches*, tarjetas de red, dispositivos inalámbricos, etc. Algunos de los fabricantes o distribuidores actualmente líderes en ventas de estos dispositivos son Cisco, HPE, Juniper, Dell o Aerohive.

Por otro lado, las plataformas de telecomunicaciones son proporcionadas principalmente por compañías de telefonía o telecomunicaciones. Dichas compañías ofrecen servicios como transmisión de voz, datos, Internet de alta velocidad, telefonía inalámbrica, etc. Ejemplo de empresas de este tipo son: Telefónica, Vodafone, etc.

Consultoría e integradores de sistemas

Muchas empresas no tienen el personal, las herramientas, el presupuesto o la necesaria experiencia para desarrollar y mantener su infraestructura de tecnología de la información. Implementar una nueva infraestructura requiere cambios significativos en los procedimientos y procesos del negocio, formación e integración del *software*. Las empresas de consultoría proporcionan ese conocimiento. Accenture, IBM, HP, Infosys o Wipro son ejemplos de este tipo de empresas.

La integración de sistemas *software* asegura que la nueva infraestructura instalada en la empresa trabaja adecuadamente con infraestructuras previas, los denominados **sistemas heredados**. Estos normalmente permanecen en una organización para evitar el alto coste de reemplazarlos o de rediseñarlos y pueden integrarse en la nueva infraestructura.

Gestión de datos y almacenamiento

Soportes para el almacenamiento de datos empresarial

Una cuestión importante, con los datos creados o recopilados por individuos o empresas, es que deben ser almacenados de forma que sean accesibles para posteriores procesos. En los sistemas computadores, los dispositivos diseñados para almacenar datos se denominan **dispositivos de almacenamiento** o simplemente almacén de datos.

La clase de almacenamiento usado varía según el tipo de dato y la velocidad a la que es creado y utilizado. Por ejemplo, dispositivos como tarjetas de memoria en los teléfonos móviles o cámaras digitales, DVD, CD-ROM y discos duros en los computadores personales. En el mundo de la empresa las principales opciones de almacenamiento son los discos duros magnéticos (HDD), discos de estado sólido (SSD), cabinas o *arrays* de discos inteligentes o cintas magnéticas.

Discos duros magnéticos

Un disco duro magnético (HDD) es un dispositivo de almacenamiento de información que utiliza un sistema de almacenamiento magnético para guardar la información. Es un sistema de memoria no volátil, ya que no pierde la información cuando está apagado.

En esencia, se compone de uno o más discos o platos magnéticos, donde reside la información, y un cabezal mecánico, que realiza las operaciones de lectura y

escritura (figura 8). Los discos giran a varios miles de revoluciones por minuto alrededor de un eje central.

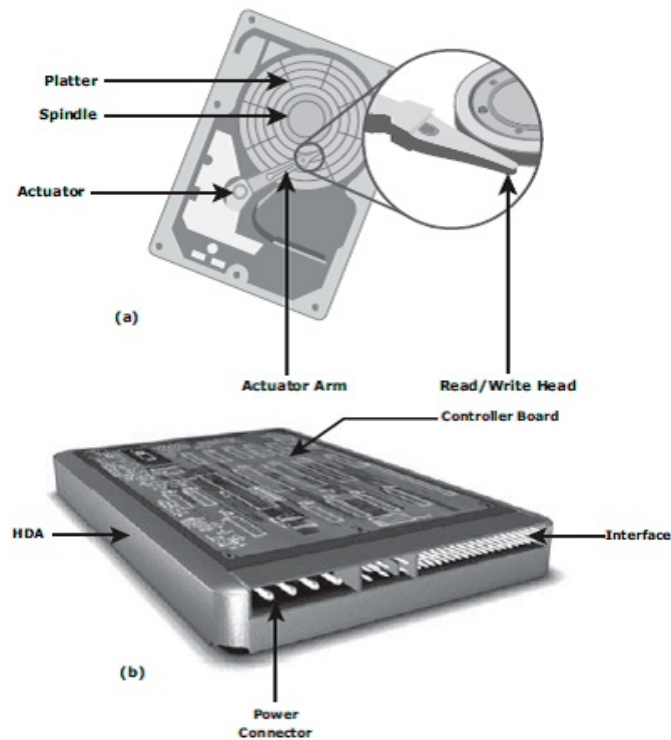


Figura 8. Disco duro magnético (HDD).Fuente: Gnanasundaram y Shrivastava (2012).

Discos de estado sólido

Un disco de estado sólido (HDD) es un dispositivo de almacenamiento de información que utiliza un sistema de almacenamiento electrónico (*flash* NAND) para guardar la información (figura 9). Es un sistema de memoria no volátil, ya que no pierde la información cuando está apagado.

En general los SSD, al no tener elementos mecánicos, son dispositivos de almacenamiento más rápidos que los HDD, pero actualmente tienen un precio por unidad de almacenamiento mayor que un HDD. Así mismo, los HDD tienen una vida media mayor que los SSD. La tendencia actual es combinar las dos tecnologías, de forma que la información más usada se almacene en los SSD y la menos usada en

los HDD (figura 10). Sin embargo, no es descartable que la tecnología SSD acabe sustituyendo a los HDD. Ya existen en el mercado soluciones de almacenamiento empresarial basadas totalmente en SSD y memoria *flash*.

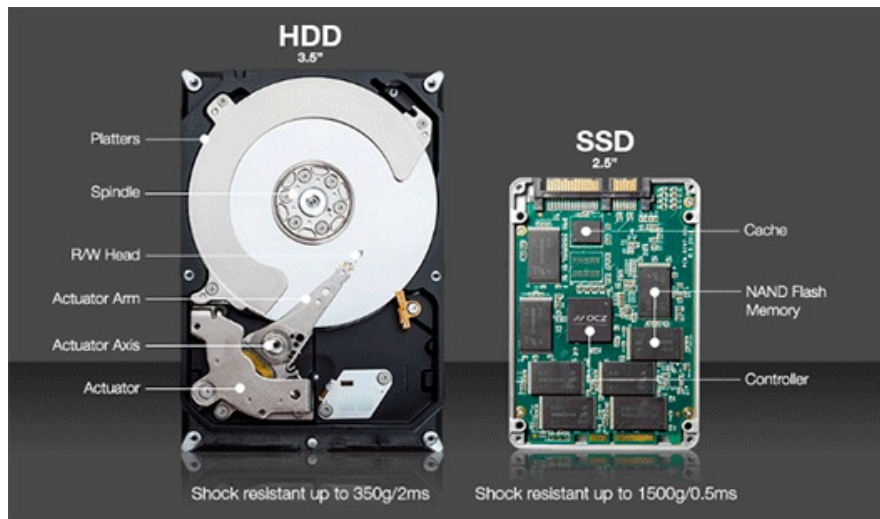


Figura 9. Diferencia entre un disco duro magnético (HDD) y un disco de estado sólido (SSD).Fuente:

<https://www.enterprisestorageforum.com/storage-hardware/ssd-vs-hdd.html>

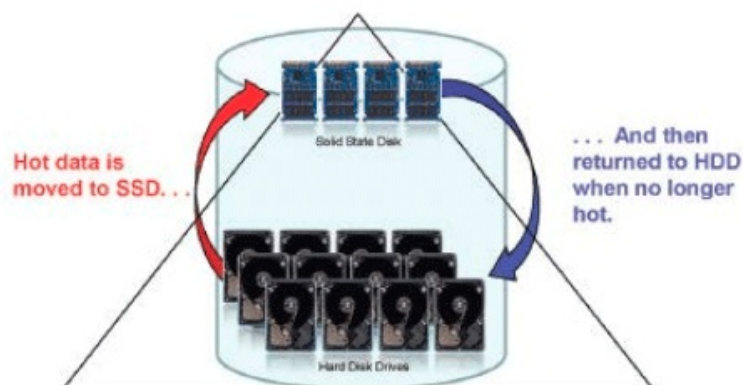


Figura 10. Esquema de una cabina compuesta por varios SSD y varios discos HDD. Fuente:

https://www.dell.com/downloads/global/products/pvaul/en/ssd_vs_hdd_price_and_performance_study.pdf

Cabina o array de discos

Un servidor puede tener conectado directamente uno o varios discos duros. Sin

embargo, si se necesitan cientos de discos duros, hay que optar por una solución diferente, como usar cabinas o *arrays* de discos, que es un dispositivo de almacenamiento de datos. Físicamente, los *arrays* o cabinas tienen el aspecto de un cajón lleno de discos.

El objetivo de la utilización de *arrays* es conseguir la integridad de los datos, disponibilidad, rendimiento y eficiencia.

Como todos los dispositivos electromecánicos, los discos de almacenamiento se pueden estropear. El fallo de un disco puede significar la pérdida de información importante. **Redundant array of inexpensive disks (RAID)** es una tecnología presente en la gran mayoría de las cabinas de discos que proporciona protección ante fallos en los discos. Existen diferentes niveles de RAID que se definen mediante un número.

La figura 11 muestra ejemplos de diferentes niveles de RAID y sus características principales:

- ▶ La primera columna indica la **resistencia ante fallos** que tiene ese nivel de RAID. Un nivel bajo o nulo indica que un fallo de disco implica pérdida de información. Un nivel alto indica que, a pesar de posibles fallos en los discos, debido a la redundancia de la información, esta no se pierde.
- ▶ RAID tiene un impacto en la rapidez con la que la cabina es capaz de leer o escribir información en los discos. El **rendimiento** de esas operaciones en función del tipo de RAID se puede apreciar en las columnas 2 y 3.
- ▶ Finalmente, si ante un fallo en un disco la información no se pierde, es porque esa información tiene copia en otro. La última columna muestra el «gasto» en espacio de disco que implica el nivel de RAID aplicado.

Nótese que el espacio de disco gastado en seguridad se pierde para datos nuevos.

Por supuesto, los niveles RAID con mejores ratios en seguridad ante fallos y rendimiento en lectura/escritura requieren mucho espacio de disco para proporcionar ese servicio.

RAID level	Fault-tolerance	Read performance	Write performance	Space requirement
RAID 0	none	good	very good	minimal
RAID 1	high	poor	poor	high
RAID 10	very high	very good	good	high
RAID 4	high	good	very very poor	low
RAID 5	high	good	very poor	low

Figura 11. Distintos tipos de RAID y sus características. Fuente: Troppens, Erkens y Muller (2004).

El nivel de servicios que proporciona una cabina de discos depende de la gama a la que pertenezca. La gama define, entre otras características, el rendimiento en operaciones de entrada/salida de la cabina, el número de servidores que se pueden conectar a la cabina, las posibilidades de aplicación de su capacidad de almacenamiento, etc. En la figura 12 se muestra la arquitectura interna de una cabina de discos inteligente.

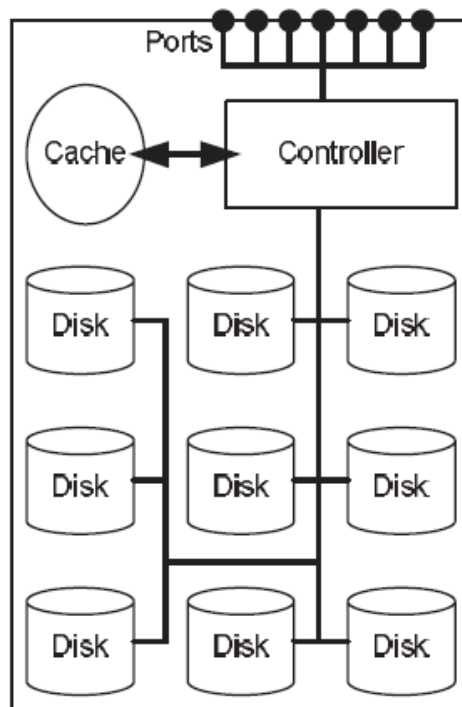


Figura 12. Arquitectura interna de una cabina de discos inteligente. Fuente: Troppens, Erkens y Muller (2004).

Métodos de conexión

Una de las características más importantes de una cabina de discos es qué tipo de conexión admite. Básicamente, podemos diferenciar entre tres modelos: almacenamiento de conexión-directa (DAS), red de área de almacenamiento (SAN) y almacenamiento conectado en red (NAS).

Almacenamiento de conexión-directa (DAS)

Consiste en conectar el dispositivo de almacenamiento directamente al servidor o estación de trabajo, físicamente conectado al dispositivo que hace uso de él. Es el método más sencillo y tradicional. Sin embargo, no es un método que escale bien a nivel empresarial, ya que crea silos de datos aislados de difícil acceso para todos los sistemas de la empresa. Además, los servidores y cabinas tienen un número acotado y limitado de puntos de acceso que impide la conexión de nuevos elementos cuando se ocupan todos.

Red de área de almacenamiento (SAN)

La cabina de discos se conecta a los servidores mediante una red dedicada de alta velocidad (generalmente *fibre channel* o iSCSI). Esto permite acceso a la cabina de todos los servidores que estén conectados a esa red de alta velocidad. Es importante aclarar que esa red dedicada solo sirve para conectar sistemas de almacenamiento con computadores y no para conectar computadores entre ellos, ya que este tipo de conexión está orientada a la transferencia de bloques de datos.

Es posible conectar un conjunto de computadores a un conjunto de cabinas de almacenamiento de forma sencilla. SAN es una solución empresarial muy escalable, pues permite añadir nuevas cabinas a la red de almacenamiento de forma sencilla y

amplia.

Almacenamiento conectado en red (NAS)

En este caso la cabina se une a la red de computadores convencional, como cualquier otro dispositivo y se le asigna una dirección IP, como un miembro más de la red de computadores. En otras palabras, la cabina es para la red de computadores otro computador más, cuya única finalidad es proporcionar ficheros. Por eso se entiende que esta conexión está orientada a la transferencia a nivel de fichero. NAS también es una solución escalable, ya que su ampliación equivale a incorporar nuevos computadores a la red.

A nivel general, la conexión SAN permite accesos a los datos de muy alta velocidad, pero el coste de la instalación es alto. Por otro lado, NAS tiene un rendimiento algo menor, debido al *software* de servicio no productivo usado, pero su coste es menor. Por tanto y, de nuevo, a nivel general:

- ▶ Si se necesita un uso muy intensivo del **sistema de almacenamiento** (muchas lecturas y escrituras), SAN puede ser una buena solución.
- ▶ Sin embargo, si el sistema de almacenamiento se utiliza como **sistema de respaldo** (copias de seguridad, por ejemplo) con pocos accesos, NAS puede ser una mejor solución.

Sistemas de gestión de bases de datos

Son las aplicaciones *software* encargadas de almacenar y gestionar los datos de la empresa. Están ubicadas en alguno de los soportes *hardware* vistos con anterioridad. Este tipo de sistemas se tratarán posteriormente en profundidad.

Plataformas de Internet

Otro de los componentes importantes de infraestructura de la tecnología de la información de una empresa es su plataforma de Internet. Este componente se

solapa y relaciona directamente con otros como las infraestructuras de redes y telecomunicaciones o las plataformas *software* y *hardware*.

Esta plataforma incluye el *hardware*, *software* y servicios que dan soporte al portal web de la empresa. Por lo tanto, estamos hablando de los servicios de hospedaje web, de los dispositivos de conexión tipo cableado, *routers*, *switches* o equipos inalámbricos.

Desde el punto de vista del *software*, la creación de un portal web necesita un conjunto de aplicaciones que implementen dichos servicios. Como posibilidades de *software* libre o abierto, una es la utilización de la familia de aplicaciones Apache HTTP, desarrolladas por la Apache Software Foundation, y otra es Nginx. Otra opción es utilizar las herramientas propietarias de Microsoft, proporcionadas con el nombre de Microsoft Internet Information Server.

Desde el punto de vista de la plataforma *hardware*, es necesario la utilización de un servidor con gran capacidad de computación, o bien, varios más económicos acoplados en clúster o granja. Dependiendo de la disponibilidad y productividad, el *hardware* tenderá a ser más o menos complejo.

Una opción interesante para muchas empresas es externalizar estos servicios, contratando un **servicio de hospedaje**. Las empresas de hospedaje mantienen el *hardware* y *software* necesarios para proporcionar servicios web y ofrecen suscripciones por pago a terceras personas, para mantener portales web y almacenamiento de datos.

5.5. Tendencias de las plataformas hardware y software

Esta sección es un complemento de la anterior. Se proporciona conocimiento sobre algunas tecnologías que actualmente constituyen tendencia en las infraestructuras de la tecnología de la información.

Plataformas hardware

Dispositivos de almacenamiento basados en objeto (OSD)

Muchos de los datos manejados por una empresa (*mails*, documentos de oficina, audio, vídeos, etc.) no son estructurados. Un OSD es un dispositivo de almacenamiento diseñado para almacenar datos no estructurados como objetos, también proporciona una opción de almacenamiento compartido escalable, autogestionada y protegida.

Almacena los datos en forma de objetos con direccionamiento plano, es decir, sin jerarquía de directorios o carpetas y ficheros (figura 13). Además, los objetos guardados permiten incluir metadatos definidos por el usuario, lo que facilita operaciones posteriores de análisis o agrupamiento de esos datos no estructurados.

A nivel general, el almacenamiento de objetos es altamente escalable (por eso se utiliza en algunos *data centers*), pero puede ser más lento que otro tipo de conexión, por ejemplo, NAS.

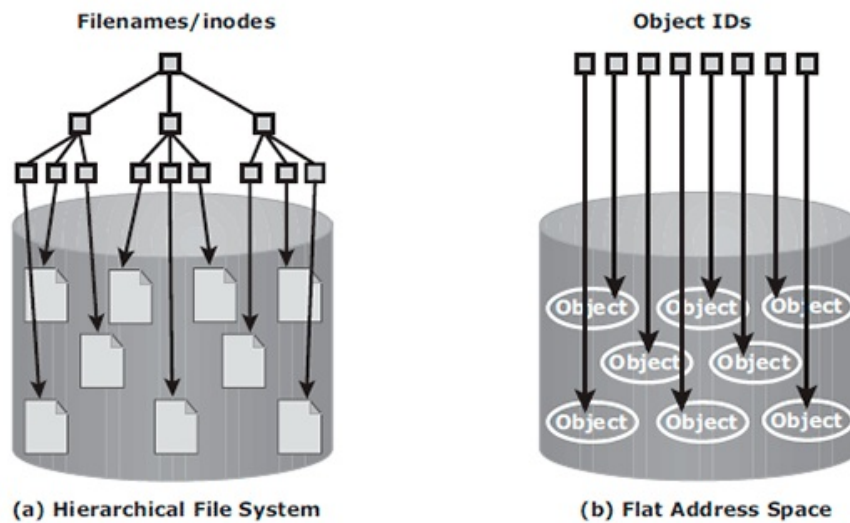


Figura 13. Diferencia entre un sistema de almacenamiento basado en archivos y directorios y otro basado en objeto. Fuente: Gnanasundaram y Shrivastava (2012).

Virtualización

Una de las tendencias que más se ha consolidado en los últimos años es la virtualización. Se puede ver como una técnica que permite la replicación o agregación de recursos *hardware* mediante el *software*.

En el primer caso, la virtualización es la **replicación de un recurso *hardware*** mediante una aplicación *software*. La virtualización permite que un solo recurso *hardware* (físico) mediante *software* se pueda ver como un conjunto de recursos replicados (lógicos). Por ejemplo, un computador con alto poder de computación puede ejecutar un *software* de virtualización para crear varias réplicas lógicas de sí mismo, de forma que, una vez virtualizado, se dispone de un conjunto de computadores replicados por *software*.

La aplicación práctica es que en cada uno de esos computadores replicados se puede instalar un sistema operativo diferente con aplicaciones *software* específicas. De esta forma se parte de un computador único que, mediante virtualización, se ha convertido en un conjunto de computadores, cada uno de ellos configurado para una

tarea o servicio específico.

En la figura 14, en la parte superior se presenta un computador no virtualizado y, en la parte inferior, un computador que se ha virtualizado en tres computadores lógicos.

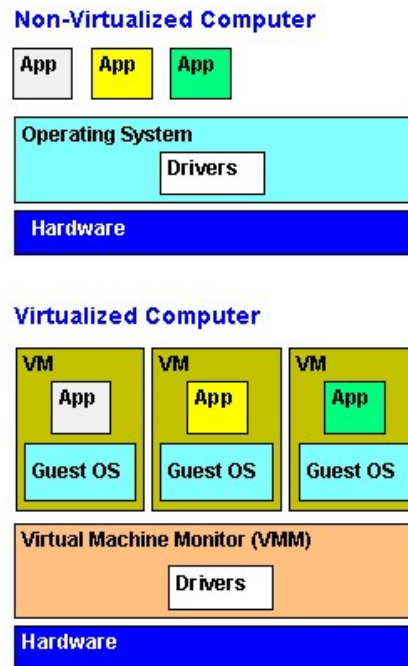


Figura 14. Virtualización de un computador. Fuente:

<https://encyclopedia2.thefreedictionary.com/Virtual+platforms>

La virtualización también se puede utilizar para conseguir el efecto contrario, esto es, a partir de un conjunto de recursos *hardware* crear, mediante *software*, un único recurso agregado (recurso lógico). Por ejemplo, un *array* de discos de almacenamiento de mediana capacidad se puede virtualizar para convertirlo en un único disco (lógico) de gran capacidad.

El usuario, para su comodidad, percibe un solo disco de almacenamiento y es el *software* de virtualización el encargado de repartir la información por todo el *array* de forma transparente para el usuario (ver figura 15). Actualmente, VMware es el

software líder de virtualización para servidores Windows y Linux.

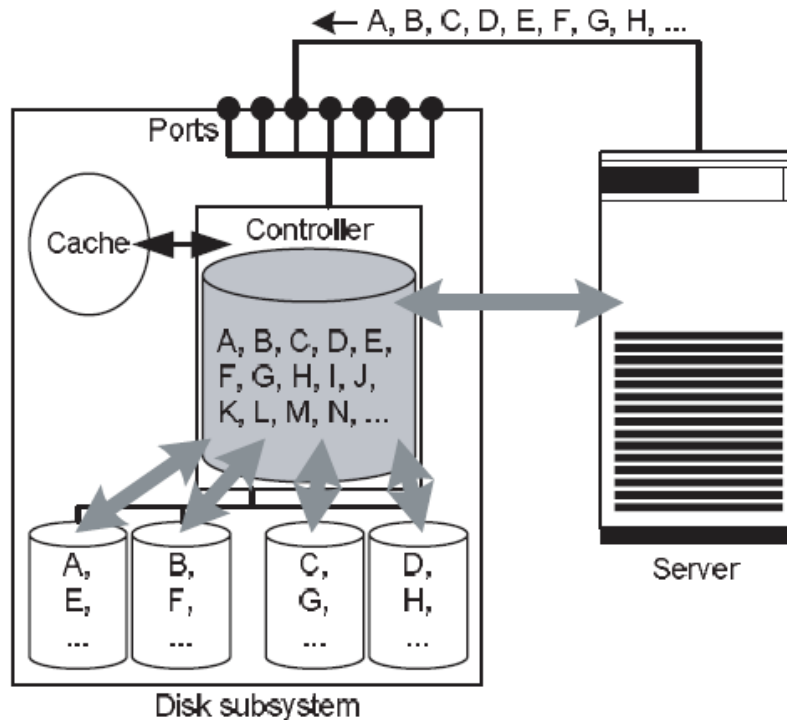


Figura 15. Virtualización de un array de discos. Fuente: Troppens, Erkens y Muller (2004).

La principal ventaja que aporta la virtualización es el mejor aprovechamiento del *hardware*. Servidores que tenían una tasa de uso del 20-30 % del tiempo disponible, al aumentar su cartera de servicios, aumentan su usabilidad y pueden llegar a ratios de 70-80 % de uso de su tiempo.

Una mayor tasa de uso llega a una menor necesidad de compra de *hardware*, ya que las tareas que antes realizaban, por ejemplo, dos servidores físicos de distinta configuración lo puede hacer uno, virtualizando en él dos servidores lógicos. Así mismo, la virtualización centraliza la administración del *hardware* y, por tanto, la facilita.

Plataformas software

Linux y el software de código abierto

El *software* de código abierto es uno creado por una comunidad de cientos de miles de programadores. De acuerdo con la asociación líder de código abierto opensource.org (<https://opensource.org/>):

El *software* de código abierto es gratuito y puede ser modificado por los usuarios.

El *software* modificado también debe ser gratuito y puede ser redistribuido por el usuario sin ninguna licencia adicional. Así, por definición, este tipo de *software* no está restringido a ningún sistema operativo o tecnología *hardware* específica. Si bien es verdad que la mayoría del *software* abierto se basa en el sistema operativo Linux o Unix.

El movimiento de código abierto ha evolucionado a lo largo de más de 30 años y ha demostrado que puede producir *software* de alta calidad y, por tanto, comercialmente aceptable. Ejemplos de productos *software* desarrollados por este movimiento son: el sistema operativo Linux, el servidor web Apache HTTP, el navegador Mozilla Firefox y las herramientas de escritorio Apache OpenOffice.

La mayoría de los grandes proveedores de *hardware* y *software*, incluyendo a IBM, HP, Dell, Oracle y SAP ofrecen versiones de sus productos compatibles con Linux.

Linux es quizás el *software* de código abierto más conocido, es un sistema operativo emparentado con Unix. Fue creado por el programador Linus Torvalds y publicado en Internet en 1991. Está disponible de forma gratuita en Internet o en versiones comerciales de bajo coste, que incluyen herramientas y soporte de proveedores como Red Hat.

Aunque Linux no se utiliza apenas en computadores personales, sí es muy popular en redes de área local, servidores web y computadores de alto rendimiento. IBM, HP,

Intel, Dell y Oracle tienen a Linux como parte central de sus ofertas a corporaciones. La llegada de este *software* y las aplicaciones que soporta ha tenido profundas implicaciones en las plataformas de *software* corporativo: reducción de costes, fiabilidad e integración, ya que Linux ha sido adaptado a la mayoría de las plataformas *hardware*: *mainframes*, servidores, computadores personales o móviles.

Software para dispositivos y la web: Java, HTML yHTML5

Java es un lenguaje de programación independiente de sistema operativo o procesador que es ampliamente utilizado para crear entornos interactivos en la web, por ejemplo, páginas con las que se puede interaccionar mediante un formulario. Fue creado en 1992 y en noviembre de 2006 fue liberado como *software* de código abierto.

Es compatible con navegadores web como Internet Explorer, Firefox, Chrome u otros. A nivel empresarial, se está usando para desarrollar aplicaciones complejas de *e-commerce* y *e-business*, que requieren comunicación con los sistemas de procesamiento de transacciones de la organización.

La plataforma Java ha migrado a teléfonos móviles e inteligentes, automóviles, tabletas e incluso sistemas de televisión por cable. El *software* creado con Java está diseñado para ejecutarse en cualquier computador o dispositivo de computación, independientemente del procesador o sistema operativo que utilice. Actualmente, es la principal plataforma de desarrollo para dispositivos móviles que utilicen Android como sistema operativo.

HTML (*Hypertext Markup Language*) es un lenguaje de descripción de páginas que define cómo se colocan el texto, gráficos, vídeo y sonido en páginas web. Así mismo, incluye la descripción en enlaces que permiten desplazarse por distintas páginas, simplemente activándolos con el puntero del ratón, o bien, táctilmente.

HTML5 es una evolución HTML que hace posible incluir imágenes, audio, vídeo y

otros elementos directamente en páginas web sin necesidad de *software* adicional, además de facilitar la visualización de páginas web en las pantallas de diferentes dispositivos.

Servicios web y arquitectura orientada al servicio

El servicio web es una tecnología que permite la interconexión de herramientas *software* de diferentes naturalezas, de diferentes fabricantes o creadores, corriendo en diferentes sistemas operativos y siendo programadas en diferentes lenguajes de programación.

Esto siempre ha sido un problema que se incrementó con la creciente necesidad de conexión al aparecer Internet. Los servicios web intentan dar una solución a dicho problema. Estos se refieren a un conjunto de componentes *software* que permiten el intercambio de información unos con otros, utilizando estándares de comunicación web. Los servicios web no están ligados a ningún sistema operativo o lenguaje de programación, por lo que diferentes aplicaciones *software* los pueden usar para comunicarse entre ellas de forma estándar y sin crear código particularizado.

Esta comunicación es posible gracias a la utilización de ficheros estándar **Extensible Markup Language (XML)**. Se trata de un fichero de texto que contiene etiquetas y datos. Por tanto, XML proporciona un formato estándar para intercambio de datos, posibilitando que los servicios web pasen datos desde un proceso a otro.

Un ejemplo práctico de servicio web lo tenemos cuando desde un navegador un cliente reserva un avión, coche, hotel y otras posibilidades desde un mismo portal. Para los usuarios es una experiencia satisfactoria pues les evita tener que acceder a cuatro o cinco portales diferentes (compañía aérea, alquiler de coche, web del hotel, etc.).

Internamente, los sistemas informáticos de esas compañías, a pesar de ser bastante heterogéneos, se comunican mediante servicios web usando ficheros XML.

Externalización o subcontratación del software empresarial

Existen tres formas de externalizar los servicios de *software* en una empresa:

- ▶ La primera es comprar paquetes de aplicaciones *software* desarrollados por **proveedores externos**. Esto evita a la empresa tener que desarrollar su propio *software* empresarial. Algunos proveedores (tales como SAP y Oracle-PeopleSoft) han desarrollado paquetes de *software* empresarial que pueden dar soporte al proceso de negocio de una empresa internacional. Abarcan todos los aspectos que puedan ser interesantes para la empresa. Sin embargo, son paquetes generalistas que deben ser parametrizados o ajustados a la empresa interesada.
- ▶ La segunda opción es **subcontratar** los servicios a otra empresa externa. Se puede contratar el desarrollo de *software* empresarial a medida, o bien, el mantenimiento de los sistemas heredados.

Los dos métodos anteriores, aunque diferentes tienen algo en común y es que los datos y las aplicaciones empresariales están ubicados en instalaciones de la propia empresa.

- ▶ La tercera opción es utilizar **computación en la nube**. En este caso, el *software* empresarial basado en la nube y los datos utilizados están físicamente hospedados en servidores de *data centers* externos. Se puede acceder al *software* y a los datos mediante una conexión a Internet y un navegador web estándar.

5.6. Referencias bibliográficas

Gnanasundaram, S. y Shrivastava, A. (2012). *Information Storage and Management: Storing, Managing, and Protecting Digital Information in Classic, Virtualized, and Cloud Environments* (2ª ed.). Hoboken: John Wiley & Sons.

Hennessy, J. L. y Patterson, D. A. (2012). *Computer architecture: a quantitative approach*. Massachusetts: Morgan Kauffman Publishers.

Laudon, K. C. y Laudon, J. P. (2014). *Management Information Systems: managing the Digital Firm* (13ª ed.). Essex: Editorial Pearson.

Troppens, U., Erkens, R. y Muller, W. (2004). *Storage Networks Explained: Basics and Application of Fibre Channel SAN, NAS, iSCSI and InfiniBand*. Hoboken: John Wiley & Sons.

Las 30 preguntas que debes hacerte antes de comprar un sistema de almacenamiento

Cinalli, F. (6 de mayo de 2013). Las 30 preguntas que debes hacerte antes de comprar un sistema de almacenamiento [Blog post]. Recuperado de <https://federicocinalli.com/blog/item/158-las-30-preguntas-que-debes-hacerte-antes-de-comprar-un-sistema-de-almacenamiento>

El artículo se presta a la reflexión sobre la problemática de los sistemas de almacenamiento empresariales desde un punto de vista muy práctico.

El mainframe vs. la granja de servidores

Chabowski. (24 de abril de 2017). The Mainframe versus the Server Farm – A Comparison [Blog post]. Recuperado de <https://www.suse.com/c/mainframe-versus-server-farm-comparison/>

Este artículo realiza una comparación entre la computación basada en *mainframe* y la realizada mediante clúster o granja de computadores. Es interesante para apreciar la diferencia entre ambas.

1. Si en nuestra infraestructura de TI no tenemos espacio para almacenar los datos de nuevos clientes, hemos de incrementar los servicios de:
 - A. Computación.
 - B. Diseño estratégico.
 - C. Telecomunicaciones.
 - D. Gestión de datos.

2. Si nuestra infraestructura de TI no permite acceder a los datos almacenados en otra provincia remota, hemos de incrementar los servicios de:
 - A. Computación.
 - B. Diseño estratégico.
 - C. Telecomunicaciones.
 - D. Gestión de datos.

3. Un *mainframe* es un sistema:
 - A. De computación distribuida, ya que pueden acceder a él muchos usuarios.
 - B. De computación centralizada, ya que realiza todo el esfuerzo de computación.
 - C. De almacenamiento adecuado para el *big data*.
 - D. Computador compuesto de muchos computadores económicos.

4. En el paradigma cliente-servidor:
 - A. No necesita infraestructura de red de computadores.
 - B. Los computadores clientes deben tener grandes prestaciones.
 - C. Los servidores tienen que ser computadores tipo *mainframe*.
 - D. Los computadores servidores proporcionan a los computadores clientes servicios.

5. La computación empresarial permitió:
 - A. Enlazar diferentes tipos de *hardware* de computación.
 - B. Almacenar conjuntos de datos en bases de datos.
 - C. La utilización de computadores personales en la empresa.
 - D. Centralizar toda la computación en un computador de altas prestaciones.

6. La computación paralela:
 - A. Consigue aumentar la productividad de las organizaciones.
 - B. Consigue decrementar la productividad de las organizaciones.
 - C. Necesita de grandes inversiones económicas.
 - D. Siempre decrementa el tiempo de respuesta ante una petición del sistema.

7. El paralelismo a nivel de tarea se consigue con computadores:
 - A. SISD.
 - B. SIMD.
 - C. MISD.
 - D. MIMD.

8. Un clúster es un computador:
 - A. Multiprocesador con memoria-compartida centralizada.
 - B. Multiprocesador con memoria-compartida distribuida.
 - C. De computadores con memoria principal independiente.
 - D. De tipo *mainframe*.

9. Apache HTTP es:

- A. Una herramienta propietaria para crear una plataforma de Internet.
- B. Una herramienta de *software* de código abierto para crear una plataforma de Internet.
- C. Un navegador web de libre distribución.
- D. Un sistema de gestión de base de datos de código abierto.

10. Linux es:

- A. Un sistema operativo de código abierto mayoritario en los computadores personales.
- B. Un sistema operativo de código abierto utilizado en servidores.
- C. Un sistema operativo propietario mayoritario en los computadores personales.
- D. Un sistema operativo propietario utilizado en servidores.