Administración de servidores

PID_00275601

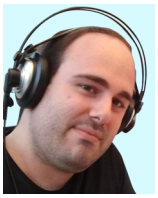
Eduard Marco Galindo
Javier Panadero Martínez

Tiempo mínimo de dedicación recomendado: 6 horas

Universitat Oberta de Catalunya

Eduard Marco Galindo

Ingeniero superior informático por la UPC. Desde 2003, colabora con la UOC como tutor y profesor colaborador en el grado de Informática y en el máster de Seguridad. Especializado en el ámbito de la Administración de Sistemas, ha formado parte del equipo de redacción del temario de la asignatura Administración de redes y sistemas operativos (ARSO) y ha ejercido de consultor y tribunal en su TFG. En el ámbito profesional, trabaja desde hace más de veinte años en el mundo de los sistemas informáticos, especialmente en la capa *middleware* de empresas y Gobiernos. Especializado en la arquitectura de sistemas en el ámbito empresarial, y también en la gestión de equipos de proyecto y de servicios gestionados. Recientemente, ha iniciado una nueva etapa profesional y de investigación en proyectos de inteligencia artificial en el ámbito empresarial. Forma parte del equipo técnico de diseño e implementación de soluciones.



Javier Panadero Martínez

Ingeniero informático y doctor en Computación de Altas Prestaciones por la Universidad Autónoma de Barcelona (UAB). Desde 2019, es profesor de los Estudios de Informática, Multimedia y Telecomunicación de la Universitat Oberta de Catalunya (UOC). Director del máster universitario de Ingeniería Computacional y Matemática. Ha elaborado varios materiales sobre administración de sistemas y programación. Sus intereses de investigación incluyen la computación paralela y distribuida, la optimización y simulación de sistemas complejos y los algoritmos inteligentes.

Primera edición: septiembre 2020

Av. Tibidabo, 39-43, 08035 Barcelona
Autoría: Eduard Marco Galindo, Javier Panadero Martínez
Producción: FUOC

Índice

Introducción

Hoy en día los servidores ya no son ordenadores «de película» que ocupan habitaciones enteras, sino que son ordenadores, parecidos en su aspecto a los que podemos tener en casa, con características especiales de hardware y de software. Incluso, actualmente, un servidor puede llegar a ser una instancia virtual situada a miles de kilómetros, que gestionaremos a distancia.

Si hemos de administrar los servidores, necesitamos saber qué tienen de especial, qué podemos esperar de ellos y qué les podemos pedir que hagan. También es importante tener presente todo lo que debemos hacer para administrarlos y protegerlos, además de elegir, configurar y mantener sus componentes y el sistema operativo.

Adicionalmente, hemos de decidir cuál es la configuración más adecuada dependiendo de la función a la que lo queramos destinar, sea física o virtualmente, en agrupaciones, clústeres o individualmente, y habrá que tener presente que el servidor estará conectado a la red y esto afectará a su configuración y seguridad.

Finalmente, hay que recordar que como administradores de servidores hay un conjunto de tareas y de responsabilidades que debemos conocer.

Objetivos

En los materiales didácticos de este módulo, presentamos los contenidos y las herramientas para alcanzar los objetivos siguientes:

1. Conocer las características que deben tener los servidores, los cuales han de cumplir unos requisitos de funcionamiento bastante estrictos.

2. Conocer las características que deben tener los sistemas operativos de los servidores, dado que tienen que cumplir diferentes funciones siguiendo unos requisitos de seguridad bastante estrictos.

3. Entender las posibles configuraciones de los servidores, sean físicos o virtuales, para obtener sistemas con un rendimiento óptimo.

4. Comprender las diferentes combinaciones de servidores, ya sean físicos o virtuales, con objetivos comunes o específicos.

5. Comprender los distintos tipos de almacenamiento, interno y externo, sus componentes, configuraciones y variedades para garantizar el rendimiento y la seguridad de los datos.

6. Conocer los diferentes dispositivos y las políticas a la hora de hacer copias de seguridad.

7. Conocer los diversos componentes de hardware que se instalan en un servidor físico para poder obtener un buen rendimiento. También los componentes virtuales para los servidores virtualizados.

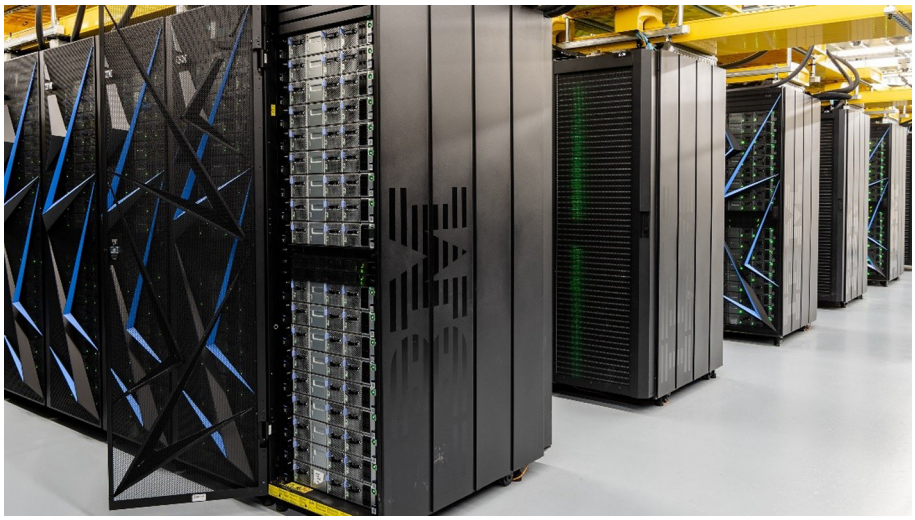
8. Comprender cómo debe aplicarse a los servidores el concepto de seguridad, y también los aspectos legales que hay que tener en cuenta.

9. Conocer las responsabilidades de un administrador de servidores.

1. Desmitificando el servidor

Cuando se habla de servidores, existe una tendencia generalizada a pensar en máquinas físicas enormes que ocupan salas enteras y que están protegidas en ambientes especiales y con una seguridad de película. En un principio, los servidores sí que ocupaban grandes espacios y tenían ambientes especiales. Hoy en día, estamos acostumbrados a ver grupos de servidores dispuestos físicamente en centros de procesamiento de datos (también denominados CPD), de manera que ofrecen este aspecto.

Figura 1. Aspecto de un CPD



Fuente: IBM Summit Supercomputer.

Sin embargo, la realidad de los servidores ha cambiado radicalmente en los últimos años. Los servidores ya no solo son físicos, también pueden ser virtuales, lo que implica un cambio de paradigma en esta rama de la informática.

Así pues, a pesar de que los servidores no son iguales a la imagen que tenemos predefinida, sí que son ampliamente diferentes en funcionalidad y servicio a cualquier ordenador personal.

Hoy en día los **servidores físicos** tienen una apariencia similar a cualquier ordenador personal o estación de trabajo. Lo que varía sustancialmente es el software y el hardware instalado dentro de la carcasa externa.

Los servidores físicos

Como veremos en este apartado, los servidores físicos solo son una parte de los servidores existentes hoy en día.

Un servidor es una máquina (sea física o virtual) que funciona 24 × 7 × 365 (veinticuatro horas × siete días × todos los días del año), y esto significa que ha de estar preparado para no parar nunca y soportar reparaciones y manteni-

mientos técnicos «en caliente» (sin dejar de dar el servicio que se le requiera). También debe poder aguantar múltiples peticiones concurrentes de servicio con tiempo de respuesta dentro de los parámetros definidos.

Los servidores, finalmente, disponen de sistemas de gestión especializada y mecanismos de seguridad que permiten gestionar la disponibilidad, la confidencialidad y la integridad de los datos y de los servicios gestionados.

1.1. Funciones del servidor

Hay bastante unanimidad en la definición de servidor informático.

Un **servidor** es un sistema físico o virtual que pone recursos propios a disposición de otros ordenadores clientes.

Recursos y ordenadores clientes

Los recursos pueden ser datos, ficheros, aplicaciones, servicios, etc.

Al mismo tiempo, ordenadores clientes también pueden ser servidores de otros recursos, lo que crea una cadena de servicio entre diferentes ordenadores.

1.1.1. Servidores funcionales

La cantidad de tareas que realizan los servidores es muy elevada. Conceptualmente, un servidor proporciona recursos y, por lo tanto, ya sea físico o virtual, puede servir a muchas necesidades. Del mismo modo, un ordenador puede no estar dedicado a hacer de servidor, pero sí a servir alguna demanda (**dar un servicio**).

Ved también

Podéis ver el módulo «Administración web» para repasar los conceptos de un servidor de aplicaciones.

Así pues, podemos encontrar servidores de diferentes recursos o servicios. Vamos a definir algunos de los más conocidos, a pesar de que hay infinidad según las necesidades del servicio:

Tabla 1. Clasificación de servidores funcionales

Tipo de servidor	Características
Servidores de dominio	• Definen el dominio de la infraestructura TI de la organización. • Permiten estructurar la infraestructura, los usuarios y los permisos. • Permiten aplicar políticas de gestión, seguridad, etc.
Servidor de ficheros	• Se pueden definir grupos de usuarios. • Permiten compartir ficheros entre todos los usuarios. • Permiten compartir ficheros entre los grupos de usuarios. • Permiten que cada usuario tenga espacio personal para guardar la información. El hecho de que esté en el servidor facilita la movilidad y las copias de seguridad.
Servidor de aplicaciones	• Permiten compartir programas entre todos los usuarios. • Permiten compartir programas entre los grupos de usuarios. • Servidor avanzado que permite gestionar aplicaciones y todos los recursos necesarios asociados, como por ejemplo el acceso a bases de datos, seguridad, mantenimiento, etc. Un servidor de aplicaciones se relaciona normalmente con un sistema de tres capas: – Primera capa (*front-end*): capa de interacción con el usuario, basada en navegadores gráficos. Servidores web. – Capa intermedia (*middle-tier*): servidor de aplicaciones en red local. – Tercera capa (*back-end*): servidor de base de datos.
Servidor de bases de datos	• Servidores con SGBD de gestión de BD relacionales. • Servidores con BD NoSQL (bases de datos *Not only SQL*).

Tipo de servidor	Características
Servidores de *backup*	• Servidores que gestionan íntegramente la gestión del *backup* de la organización. • Permiten definir agrupaciones de servidores, políticas, rotaciones e informes. • Centralizan toda la gestión de las copias de seguridad.
Servidores de gestión documental	• Servidores que permiten gestionar ordenada y estructuradamente la documentación de las organizaciones. • Gran capacidad de almacenamiento y de gestión de archivado para documentación histórica.
Servidor de impresión	• Permiten compartir las impresoras.
Servidor de correo	• Servidores que permiten enviar y recibir mensajes.
Servidores de seguridad	• Servidores especializados en la gestión de la seguridad. Desde servidores de análisis de las vulnerabilidades, SIEM (*security information and event management*), hasta cortafuegos (en inglés, *firewall*), *proxy*, etc. • Ayudan a la organización a garantizar la seguridad de la infraestructura TI.

Y otros muchos, como por ejemplo:

- Servidores de ML y DL (*machine learning* y *deep learning*) en el entrenamiento de modelos de inteligencia artificial.
- Servidores de gestión de sistemas de fabricación.
- Servidores de sistemas de laboratorio.
- Etc.

1.1.2. Requisitos de los sistemas operativos de red

La instalación del sistema operativo (SO) en el servidor debe permitirnos conseguir la funcionalidad necesaria. No todos los SO son compatibles con todos los servicios y, por lo tanto, hay que planificar y prever con antelación cuál es el SO que mejor se adecua a las necesidades.

Una de las ventajas de los SO de los servidores actuales es que se pueden desplegar en todo tipo de servidores, sean físicos, virtuales e, incluso, en servidores en la nube (Cloud OS) especializados para cada tipo de servidor.

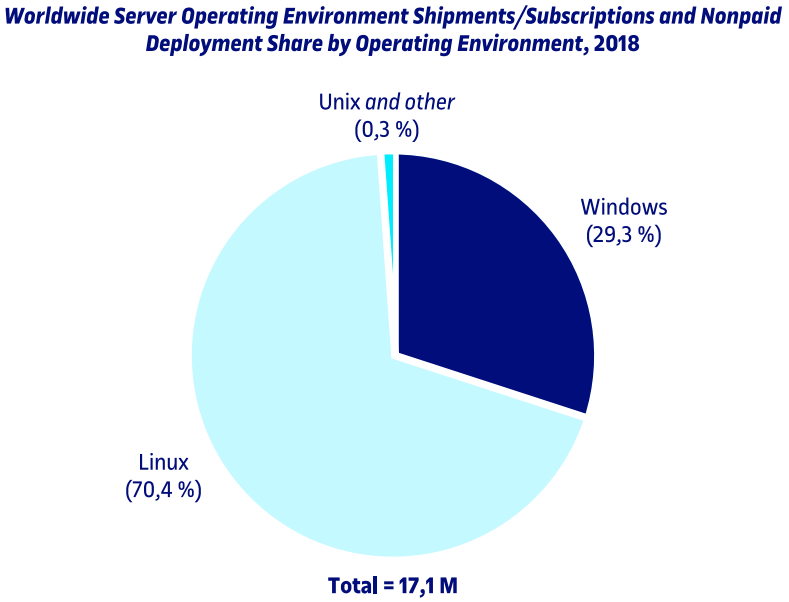
Así pues, todos los fabricantes de SO, sean de pago o de libre distribución, tienen versiones compatibles con diversidad de servidores. La compatibilidad vendrá dada, por un lado, por la correcta comunicación del *kernel* del SO con el hardware del servidor (sea físico o virtual), y por otro, por la certificación de los *drivers* para las diferentes tarjetas y dispositivos asociados (también podrán ser físicos o virtuales).

Compatibilidad de los SO con los servidores

A pesar de que, por norma, son compatibles, hay que consultar para cada SO la información del fabricante para certificarlo.

Según publica la *International Data Corporation* (IDC), como principal proveedor de inteligencia de mercados, en su resumen anual de 2018 y tal y como muestra la figura 2, actualmente hay dos grandes familias de SO para servidores: Windows y Linux, cada una con sus diferentes versiones y distribuciones.

Figura 2. Distribución de sistemas operativos en el mercado



Fuente: IDC.

1.2. Configuraciones de los servidores

Las diversas necesidades de una organización hacen que a menudo un servidor no sea suficiente, de manera que es habitual que las organizaciones tengan más de uno, sean físicos o virtuales, para alcanzar sus objetivos. Por lo tanto, podemos encontrar un servidor que lleve a cabo una o muchas tareas, o muchos servidores trabajando por un propósito común. También es posible encontrar servidores muy diferentes entre sí agrupados en un mismo espacio que llevan a cabo tareas diversas.

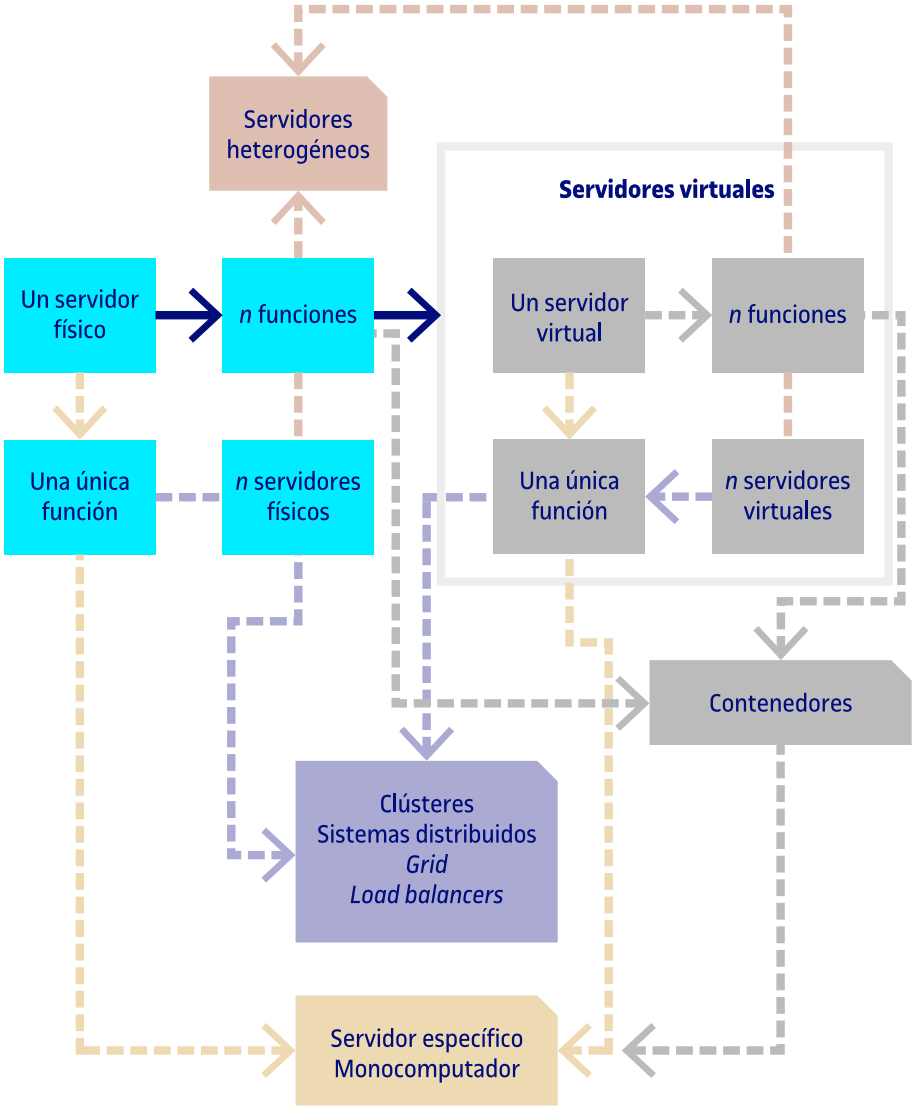
Estas combinaciones, muchas veces heterogéneas de los servidores, se basan en la funcionalidad. Así, si por ejemplo queremos un servicio de correo que difícilmente falle, pondremos un clúster de correo en alta disponibilidad (como veremos más adelante). Esto representa al menos dos servidores exclusivamente dedicados al correo. Si, además, necesitamos un servicio de ficheros muy grande, entonces pondremos un servidor dedicado a NAS con una librería de *backup*.

Como podemos ver, es la necesidad de la organización la que configura la estructura de los servidores. Debido a su entorno dinámico, se debería hacer una planificación inicial para prever, en la medida de lo posible, las ampliaciones que pueda haber a medio plazo para no hacer gastos y tareas de organización del sistema informático que sean insuficientes en poco tiempo. En este sentido, la virtualización de los servidores y los servicios que se ofrecen en la nube facilitan una adaptación dinámica de las necesidades de las organizaciones.

Esta gran variedad de necesidades, atendiendo a la configuración y la función, hace necesaria una clasificación de los servidores. La clasificación que mostramos a continuación en la figura 3 no pretende ser exhaustiva, sino orientativa y didáctica, sabiendo que puede haber otras alternativas.

La configuración de los servidores debe cubrir las necesidades específicas de la organización.

Figura 3. Configuraciones de los servidores



Este diagrama de cruces define conceptualmente los diferentes tipos de servidores y los servicios que estos pueden ofrecer a sus clientes conectados.

Un aspecto destacado, como se puede ver, es que la **virtualización** de los servidores crea un nuevo nivel de configuración de servicios y provee recursos virtuales del mismo modo que lo harían los servidores físicos.

También hay que tener en cuenta una nueva capa de servicio de recursos, que son los **contenedores**. Estos servidores virtuales suelen tener, por definición, un solo servicio y, por lo tanto, se podrían definir como servidor específico.

Así pues, independientemente de si son servidores físicos, virtuales o, incluso, contenedores, los servidores pueden tener los comportamientos siguientes:

- **Un servidor, una función**: es el nivel más sencillo de servidor, un solo sistema dedicado a una sola función. Por ejemplo, un sistema que realiza tareas de gestión de correo.

- **Un servidor, *n* funciones**: si disponemos de un servidor poco utilizado en cuanto a recursos, podemos aprovechar este remanente para ofrecer otros servicios a los clientes. Así pues, tenemos un sistema que optimiza los recursos con varias funciones de servicio. El ejemplo más claro sería la virtualización de un servidor físico para dar servicio a diferentes servidores virtuales.

Ved también

Podéis ver el apartado «Tipos de servidores: servidores físicos y servidores virtuales».

- ***n* servidores, una función**: en nuestra organización podemos tener servicios críticos, ya sea por necesidad de servicio, seguridad o rendimiento, que hacen necesario un número de recursos muy importante y escalable. Esta necesidad se gestiona con arquitecturas en las que una sola tarea es tratada por más de un sistema.

Ved también

Podéis ver el apartado «Agregación de servidores».

- ***n* servidores, *n* funciones**: cuando varias funciones son tratadas por diferentes ordenadores, tenemos un sistema de servidores heterogéneo en el que pueden aparecer un gran número de combinaciones posibles.

2. Tipos de servidores

En el apartado anterior se han comentado las funciones del servidor, y también las diferentes configuraciones que podemos tener según su tipo: físicos o virtuales. Ahora, debemos definir qué familias de servidores hay y sus principales características.

2.1. Servidores físicos

Actualmente, las organizaciones tienden a implementar la mayoría de los servicios en servidores virtuales. Este hecho implica que la mayoría de los servidores físicos se dediquen a la virtualización de recursos. Aun así, todavía hay muchos servidores físicos dedicados a diferentes tareas, ya sea por requisitos de software, rendimiento, seguridad u otros aspectos, como por ejemplo la compatibilidad con los dispositivos externos.

De entrada, los componentes de un servidor físico son los mismos que para un portátil u ordenador de sobremesa. Así, en un servidor podremos encontrar la memoria RAM, la placa de comunicaciones, las unidades de almacenamiento (discos duros), la CPU, la fuente de alimentación, la placa gráfica o GPU, el lector óptico (DVD) y la placa base.

CPU y GPU

CPU: Unidad de procesamiento central. Es la encargada de gestionar todas las operaciones de procesamiento y almacenamiento de la información, al menos, en la memoria principal.

GPU: Unidad de procesamiento de gráficos. Normalmente es la encargada de gestionar las funciones gráficas, pero últimamente, visto su potencial de cálculo, se utiliza para mejorar los tiempos de respuesta en operaciones complejas. Actualmente, es una pieza básica en el entrenamiento de los modelos de inteligencia artificial.

Algunos de los componentes que tenemos en los ordenadores físicos con los que estamos acostumbrados a trabajar, como por ejemplo el monitor, el teclado y el ratón, ya no se utilizan, dado que los servidores permiten ser gestionados simplemente teniendo acceso a la red.

2.1.1. RAM

Todos los usuarios o sistemas cliente piden (hacen peticiones) a los servidores. Por lo tanto, es importante que estos nos puedan responder lo antes posible. Por este motivo, es muy importante una buena cantidad de RAM, y cuanto más rápida sea la RAM que se instale, mejor. Si, por ejemplo, se trata de un servidor físico dedicado a la virtualización, entonces la cuestión es mucho más crítica y debemos instalar la cantidad de RAM que sea necesaria para que todos los sistemas virtuales puedan compartirla con garantías.

Hay que conocer las necesidades de la organización para proveer correctamente la cantidad de RAM necesaria.

Es muy necesaria una gran cantidad de RAM en servidores físicos de virtualización, SGBD, etc.

2.1.2. CPU y GPU

Tipos de CPU

Actualmente, hay distintos tipos de procesadores o CPU. En este documento nos centraremos en comentar mayoritariamente las CPU comerciales actuales, a pesar de que hay otros tipos de CPU, como las que podemos encontrar en procesadores vectoriales o en procesadores cuánticos.

No es el objetivo de esta asignatura explicar qué son la CPU o la GPU, ya que en otras asignaturas de los grados de Informática se explican adecuadamente. Lo que sí que es necesario es explicar la importancia actual de estos elementos centrales de los servidores físicos actuales.

En los últimos años ha habido un cambio muy importante en cuanto a la virtualización de servidores, y también en cuanto al nuevo paradigma de computación que implica la inteligencia artificial (IA o AI, en inglés) o el *blockchain*. Si hace unos años la CPU y la GPU de un sistema físico tenían una importancia relativa en la mayoría de los sistemas (exceptuando los servidores dedicados a cálculos o BD), en la actualidad son un factor diferenciador.

CPU

Si nos centramos en la definición de CPU en los sistemas comerciales actuales (ved la nota de CPU), observamos que la evolución de este componente ha sido muy significativa, teniendo la posibilidad de proveer varios procesadores multinúcleo en el mismo servidor.

Cada núcleo actúa como una CPU (unidad de control y ALU); así, los servidores comerciales más potentes actualmente disponen de diferentes *sockets* (en inglés) con múltiples núcleos o *cores* (en inglés).

Los *sockets* permiten conectar los microprocesadores con la placa base, y en cada uno puede haber varios núcleos según el fabricante y el modelo de servidor.

Los sistemas actuales disponen de S (sockets) × N (núcleo) procesadores o CPU.

GPU

Antiguamente la GPU (*graphics processing unit*) básicamente se empleaba para la gestión gráfica del sistema. Hoy en día, gracias a su estructura de computación en paralelo con valores en coma flotante, se utiliza para el cálculo matemático y ha creado una revolución proveyendo a los sistemas de servidores actuales la potencia de cálculo necesaria para afrontar retos de computación como, por ejemplo, la inteligencia artificial o el *blockchain*, acelerando los procesos de aprendizaje de ML y DL (*machine learning* y *deep learning*), y reduciendo el tiempo de implementación de las soluciones.

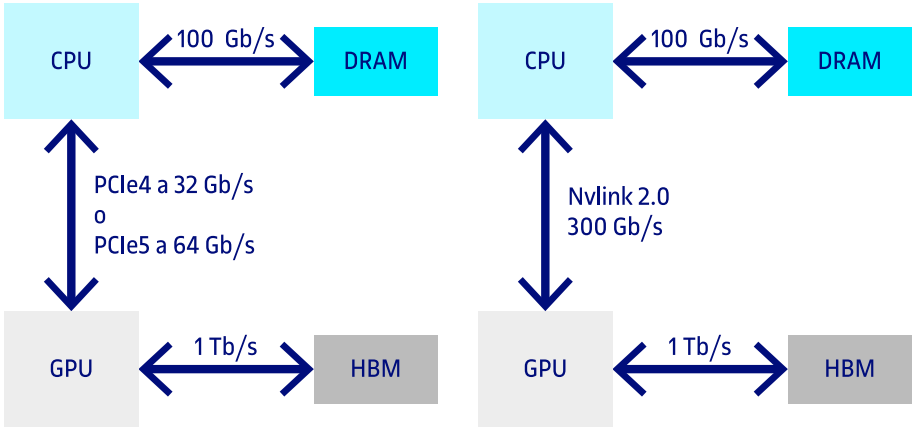
Las GPU se utilizan siempre como un procesador adicional y, por lo tanto, se combinan con las CPU tradicionales para proveer la capacidad de cálculo requerida.

En los sistemas industriales actuales, podemos tener para cada CPU diversas GPU asociadas (normalmente entre 2 y 3), y es un factor diferenciador el canal de comunicaciones entre la CPU y la GPU y entre las diferentes GPU.

La comunicación entre las diferentes GPU suele ser muy efectiva, ya que los fabricantes crean canales con el ancho de banda suficiente; en cambio, entre la CPU y la GPU, el ancho de banda de la comunicación varía dependiendo del tipo de canal empleado. Este ancho de banda del canal es muy importante, dado que las GPU no tienen acceso directo a la memoria RAM del sistema y su memoria suele ser muy limitada (entre 16 y 32 GB).

Hay básicamente dos canales de comunicación entre la CPU y la GPU actualmente en el mercado: PCie y NVLink, que se muestran en la figura 4.

Figura 4. Canales de comunicación entre la CPU y la GPU



Dependiendo de la versión de PCie, podemos llegar a tener ratios de comunicación de hasta 64 Gb/s en ambas direcciones (32 Gb/s en cada sentido). En cambio, con el conector comercial NVlink 2.0 podemos llegar a 300 Gb/s (150 Gb/s en cada sentido).

FPGA

Las tarjetas FPGA (*field-programmable gate array*) permiten proveer a los sistemas servidores de capacidad de computación adicional especializada. Estas tarjetas contienen bloques lógicos que en algunos casos se pueden programar y que permiten realizar operaciones concretas de manera muy eficiente y rápida, como por ejemplo la inferencia de datos de modelos de inteligencia artificial.

Últimamente, este tipo de tarjetas también han evolucionado gracias a conectores como el CAPI 2.0, que permiten acceder a la RAM del sistema de manera óptima, sin tener que contar con la colaboración de la CPU.

2.1.3. Placa base

Es esencial que este componente sea de muy buena calidad para asegurar que hay una buena velocidad de transmisión entre todos los componentes del servidor. El bus del sistema forma parte de la placa base[1] y es el componente que permite la comunicación entre la CPU, la RAM y todos los dispositivos dentro del ordenador.

(1)La placa base también se denomina *placa madre* o, su equivalente en inglés, *motherboard*.

Como el resto de los componentes de los servidores, las placas base han evolucionado mucho en los últimos años y se han adaptado al modelo del sistema elegido. Por ejemplo:

- Tenemos placas base de servidores *blade* pensadas para interaccionar con chasis de servidores (en inglés, *blade center*).

- Hay placas base con espacio para varios *sockets* de procesadores, dependiendo de las necesidades de computación. Sucede lo mismo con los circuitos integrados de memoria principal.

- También podemos encontrar placas base con varias interfaces integradas: pueden ser comunicaciones, almacenamiento y dispositivos diversos.

Uno de los aspectos más importantes de las placas base es el bus de comunicación interno (bus de datos interno). Este bus interno suele estar formado por *n* carriles PCIe (actualmente PCIe G4) que proporcionan un ancho de banda de (*n* × 16 GBps) *full-duplex* entre los elementos del sistema.

La **placa base** es vital para el servidor. Hay diferentes tipos según las necesidades del servidor.

2.1.4. Tarjetas I/O

Es el **punto de comunicación** entre el servidor y «todo el mundo». Por lo tanto, su calidad y velocidad determinan el comportamiento del servidor hacia los periféricos, el almacenamiento y la red. Es un componente crítico.

Las tarjetas las podemos definir por tipos o por funcionalidad. En cuanto al tipo de tarjetas, podemos tener PCIe3, PCIe4, PCIe5, SAS, NVMe, etc., dependiendo de los conectores con la placa base.

En cuanto a la funcionalidad, hay multitud de tarjetas I/O, pero sin duda, actualmente, las más importantes son las tarjetas de comunicación Ethernet (con tarjetas desde 1 GbE de cobre o fibra óptica hasta tarjetas de 100 GbE de fibra óptica) y las tarjetas *fibre channel* (FC) para acceso al almacenamiento de bloque (tarjetas de 8 Gbps, 16 Gbps, 32 Gbps y hasta 128 Gbps).

Cuanto más rápida sea la conexión del servidor con la red y el acceso al disco externo, antes podrá atender las peticiones de los clientes e irá más descargado (o podrá soportar más carga sin colapsarse).

La placa de comunicaciones determina la capacidad de transmitir la información en la red del servidor. La placa FC (o HBA, *host bus adapter*) determina la capacidad de escribir y leer datos de los discos externos a nivel de bloque.

2.1.5. Disposición física

La disposición física de los servidores es variada. Desde cajas especiales para soportar el calentamiento (sobre todo, si tienen muchas unidades de disco) hasta los sistemas *rack* (los más comunes), que pueden estar refrigerados con aire o, incluso, con agua.

Finalmente, tal y como podemos ver en la figura 5, encontramos el sistema *blade*, en el que cada servidor se integra como una lámina dentro de una estructura (*blade center*) donde se comparten recursos, como por ejemplo el acceso a la red de comunicaciones y de almacenamiento, las fuentes de alimentación, los ventiladores, etc.

Figura 5. Servidores *blade* UCS dentro de un *blade center* UCS de Cisco

2.2. Servidores virtuales

Los servidores virtuales basan su funcionamiento en la tecnología de la virtualización.

Esencialmente, la virtualización es dar a una computadora la posibilidad de **realizar el trabajo de múltiples computadoras**, compartiendo los recursos entre los diversos entornos. Típicamente, se ha referido a una sola computadora capaz de hacer trabajar, al mismo tiempo, diferentes sistemas operativos y servicios de manera independiente y segura.

Por lo tanto, podemos afirmar que un servidor virtualizador es aquel servidor capaz de realizar el trabajo de varios servidores compartiendo los recursos del sistema, mediante uno o más sistemas operativos de manera segura. La virtualización y los servidores virtuales tienen bastantes ventajas que los hacen atractivos, como por ejemplo:

- Reducción del número de servidores físicos.
- Reducción del espacio dentro del centro de procesamiento de los datos.
- Reducción del consumo de energía.
- Compartición de recursos y eficiencia de utilización.
- Centralización y simplificación de la gestión.

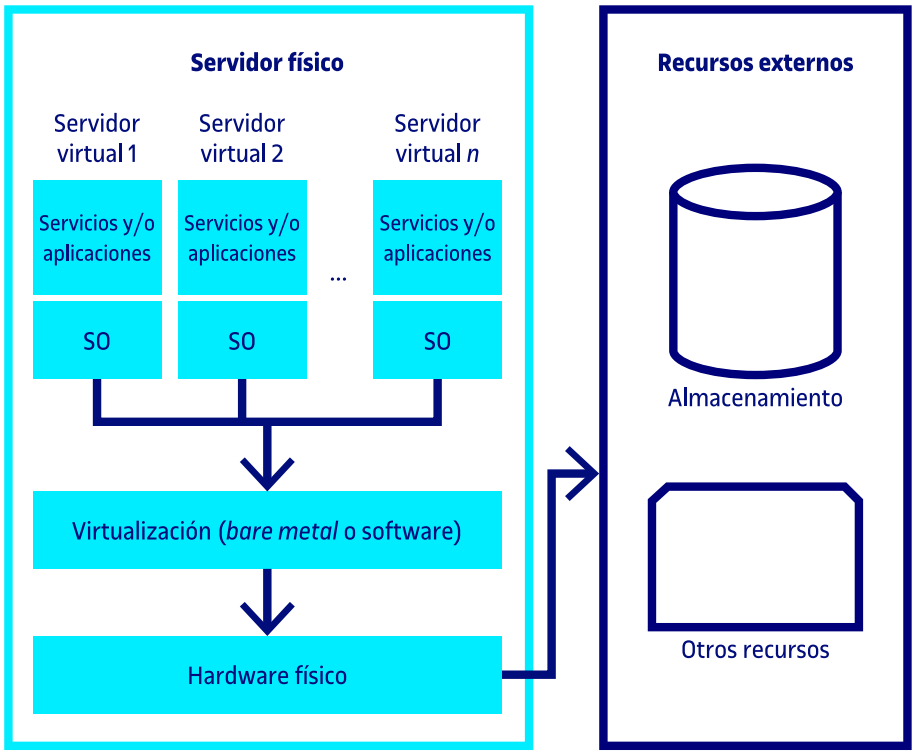
Hay varios sistemas de virtualización, dependiendo de la plataforma tecnológica del servidor físico y del tipo de virtualizadores:

- ***Bare metal*[2] *hypervisor***: donde el software del virtualizador interactúa con el hardware del servidor físico sin necesidad de un sistema operativo adicional (ved la figura 6).

- **Hipervisores de software basados en aplicaciones**: los cuales por medio del sistema operativo *host* comparten los recursos de la máquina física.

(2) *Bare metal* significa servidor físico dedicado.

Figura 6. Estructura de un hipervisor



Virtualizadores comerciales

Algunos virtualizadores comerciales conocidos son VMWARE, Hyper-V de Microsoft, Oracle Virtual Box, etc.

Cada virtualizador comercial tiene sus propias herramientas de virtualización de los recursos físicos.

Del mismo modo que los componentes de los servidores físicos son importantes para el correcto rendimiento, los componentes virtuales asignados a una máquina virtual por los hipervisores también lo son:

- **RAM**: los hipervisores comparten la RAM necesaria en cada máquina virtual que gestionan.

- **CPU y GPU**: los hipervisores comparten los núcleos del procesador físico, asignando a cada máquina virtual *sockets* y núcleos virtuales según se requiera.

- **Tarjetas I/O**: los hipervisores comparten recursos, ya sean de almacenamiento, acceso al adaptador de la red virtual u otras, como SCSI, USB, etc., en las máquinas virtuales que lo necesiten.

2.3. Contenedores

Tecnología de contenedores

La tecnología de contenedores se ha convertido en un referente, ya que permite una gestión dinámica muy eficiente y versátil de los recursos dedicados a las aplicaciones. Actualmente, las organizaciones tienen una fuerte dependencia de las aplicaciones en servicios críticos, ya sean locales o como servicio en la nube y, esto hace que la solución que ofrecen los contenedores sea la más adecuada.

La última capa de virtualización, que actualmente está tomando mucha fuerza en el ámbito de las TI, son los contenedores.

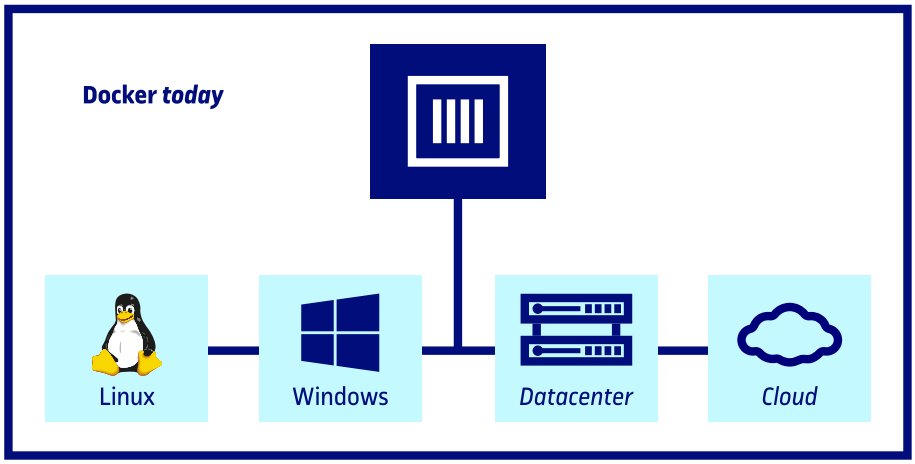
Una posible definición de contenedor, extraída de la página web de Docker, podría ser:

> «Unidad estándar de software que empaqueta código y todas sus dependencias, de modo que la aplicación se ejecuta de manera rápida y fiable de un entorno informático a otro. Una imagen de un contenedor es un paquete de software ejecutable ligero y autónomo que incluye todo lo necesario para ejecutar una aplicación: código, herramientas del sistema, bibliotecas del sistema y configuración».
>
> **Docker**. «What is a container» [en línea]. *Docker*. Disponible en: https://www.docker.com/resources/what-container.

A nivel de servidor, un contenedor es una máquina virtual parecida a las que hemos comentado en el apartado «Servidores virtuales», pero que comparte partes del SO de la máquina que la virtualiza. No obstante, tiene sus propios recursos, por ejemplo su propio sistema de ficheros, CPU, RAM, etc.

Inicialmente, los contenedores se crearon para entornos Linux, pero hoy en día esta última capa de virtualización se puede encontrar en la mayoría de los SO e hipervisores. Así pues, los contenedores se pueden implementar tanto en servidores físicos como virtuales, y también directamente en los hipervisores que los soportan.

Figura 7. Plataformas de contenedores



Fuente: www.docker.com.

Una de las ventajas claras del uso de los contenedores es su facilidad de creación y clonación, lo que permite crear grupos de contenedores para diferentes finalidades, según sea necesario. Imaginemos la versatilidad y potencia de esta tecnología si conseguimos que los contenedores trabajen en grupo para dar servicio a una única aplicación. Podremos tener capas de contenedores

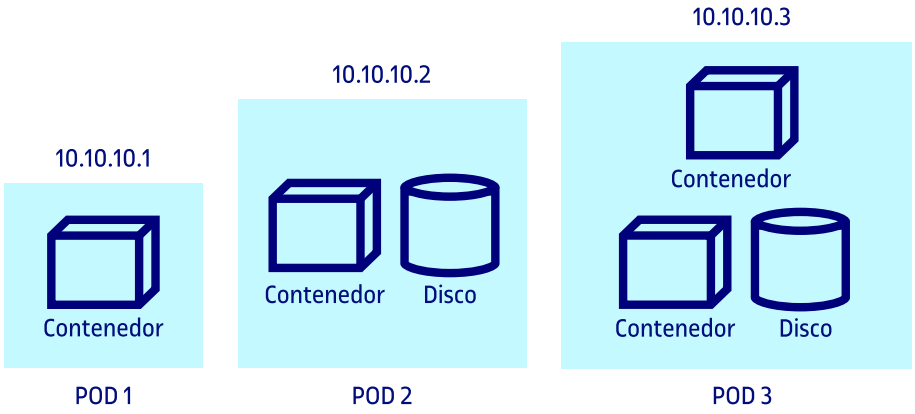
especializados en la misma aplicación, en la gestión del propio grupo de contenedores, en la gestión integrada de comunicación o en el almacenamiento permanente. Veamos algunas **estructuras de contenedores**:

- **POD**. Puede ser un solo contenedor, un contenedor único con almacenamiento persistente o varios contenedores que comparten el almacenamiento persistente. En cualquier caso, el acceso a nivel de direccionamiento de servicio es único, como se puede ver en la figura 8.

- ***Deployment***. Gestiona múltiples réplicas de POD para conseguir una alta disponibilidad.

- **CSI (*container storage interface*)**. Partiendo de la estandarización de la API de acceso al almacenamiento persistente, es un tipo de contenedor especializado que se añade a un POD para gestionar el acceso al disco.

Figura 8. Ejemplos de POD



Pero ¿cómo se gestionan los contenedores?

Para responder esta pregunta aparecen dos soluciones que se complementan:

- Los orquestadores de contenedores (CO, *container orchestration*), tal como dice su nombre, permiten gestionar el ciclo de vida de los contenedores. Algunos ejemplos actuales pueden ser Docker o Cloud Foundry.

- Los gestores de automatismos de despliegue, escalado y gestión de contenedores, como por ejemplo Kubernetes, un producto *open-source* líder en este sector (creado por un consorcio de fabricantes llamado Cloud Native Computing Foundation).

A nivel comercial, la mayoría de los fabricantes importantes y proveedores de servicios en la nube aportan sus propias soluciones (Red HAt OpenShift, Amazon ECS, Azure ACS, Google Container Engine, etc.).

2.4. Servidores en la nube

Aunque realmente los servidores en la nube no son un tipo específico de servidor, teniendo en cuenta su importancia, se describe brevemente las posibles configuraciones según el servicio ofrecido.

Los actuales servicios en la nube:

- **IaaS (*infrastructure as a service*)**: se pone directamente a disposición de los administradores la infraestructura física o *bare metal*.

- **PaaS (*platform as a service*)**: plataforma como servicio, donde podemos tener una infraestructura plataformada según las necesidades de la organización.

- **SaaS (*software as a service*)**: software como servicio, donde podemos tener servicios especializados como: correo, almacenamiento, ofimática, etc.

- **Otros servicios**: BaaS (*backup as a service*), SECaaS (*security as a service*), etc.

Permiten ofrecer desde servidores físicos dedicados (o *bare metal*) a servicios específicos a nivel de aplicación. Así pues, los servidores en la nube se podrán considerar de tipo físico, virtual o contenedor dependiendo del servicio contratado.

Es una tarea del administrador de servidores elegir los servicios en la nube teniendo en cuenta las necesidades de la organización. Dependiendo de estas necesidades, los tipos de servidores que habrá que administrar serán físicos, virtuales o contenedores.

3. Agregación de servidores

La agregación de servidores es una herramienta que permite a los administradores de sistemas disponer de una infraestructura preparada y dimensionada para hacer frente a los retos de la organización.

3.1. Balanceo de carga

Balancear una carga significa **dividir el total de trabajo** que un sistema debe realizar entre dos o más sistemas, sean físicos o virtuales. Las características generales del balanceo de carga son:

- El balanceo de carga se puede implementar por hardware, software o una combinación de los dos.

- El balanceo de carga está especialmente indicado para entornos en los que es muy difícil prever el volumen de carga de trabajo.

- El factor de división de la carga se puede definir dando más o menos carga a cada uno de los sistemas implicados. Esta característica es la carga asimétrica.

3.2. Sistemas clúster

Un clúster es un grupo de **computadoras interconectadas** que trabajan conjuntamente en la solución de un problema. Este sistema constituye una solución flexible, de bajo coste y de gran escalabilidad para aplicaciones que requieren una elevada capacidad de cómputo y memoria. Los clústeres aparecen ante los clientes y aplicaciones como un solo sistema.

Si nos fijamos en la historia de los clústeres, encontramos que si bien no se sabe la fecha exacta del primer clúster, se considera que la base científica del concepto del procesamiento en paralelo la estableció Gene Amdahl, que trabajaba en IBM, hacia el año 1967. El desarrollo de los clústeres ha estado siempre unido al de las redes de computadoras y al de las propias supercomputadoras, ya que desde el comienzo se buscó la unión de los sistemas informáticos para obtener más rendimiento y capacidades.

3.2.1. Características

- Un clúster consta de dos o más nodos conectados entre sí por un canal de comunicación.

- Cada nodo solo necesita un elemento de proceso, memoria y una interfaz para comunicarse con la red del clúster.

- Los clústeres necesitan software especializado, sea a nivel de aplicación o a nivel de núcleo.

- Todos los elementos del clúster trabajan para cumplir una funcionalidad conjunta, sea esta la que sea. Es la funcionalidad la que caracteriza al sistema.

3.2.2. Ventajas

- **Económicas**: es una razón importante para la construcción de clústeres. Reduce costes en el gasto inicial tanto de planificación como de instalación, y también en los asociados al mantenimiento (el TCO, *total cost of ownership* o coste total), si se compara con un servidor único de prestaciones equivalentes.

- **Sencillez**: la tecnología que hace funcionar un clúster se basa en la unión de elementos sencillos (suelen ser servidores específicos o monocomputadoras). Y esta sencillez aporta beneficios adicionales cuando hablamos de disponibilidad de cada uno de los nodos del clúster.

- **Disponibilidad**: la interconexión de dos o más computadoras, trabajando conjuntamente en la solución de un problema, permite incrementar la disponibilidad del servicio, ya que se divide aproximadamente el número de puntos críticos del servicio entre el número de nodos del clúster.

- **Escalabilidad**: si el SO del clúster lo permite, solo hay que conectar más equipos a la red del clúster, configurarlos correctamente y ya tenemos un clúster ampliado y mejorado. Incluso mejorando alguno de los elementos que forman parte de cada nodo (CPU, GPU, memoria RAM o disco por ejemplo) se obtiene una mejora del rendimiento o la disponibilidad.

- **Rendimiento**: el incremento de recursos asignados para resolver la misma carga de trabajo permite aumentar el rendimiento del sistema como conjunto.

- **Balanceo de carga**: la tecnología de clúster de los servidores por balanceo de carga mejora la respuesta a las peticiones, conmutándolas entre los distintos nodos del clúster.

3.2.3. Componentes

- **Nodos**: pueden ser servidores físicos, virtuales o incluso contenedores de aplicaciones que formen un clúster.

- **Sistemas operativos**: deben ser de fácil uso y acceso, y también permitir múltiples procesos y usuarios.

- **Conexiones de red**: los nodos de un clúster pueden conectarse mediante una simple red Ethernet o se pueden utilizar tecnologías especiales de alta velocidad, como por ejemplo Fast Ethernet, Gigabit Ethernet, Myrinet, Infiniband o SCI.

- ***Middleware***: es un software que generalmente actúa entre el sistema operativo y las aplicaciones con el fin de proveer una interfaz única de acceso al sistema, denominada SSI (*single system image*), que genera la sensación al usuario de que utiliza un único ordenador muy potente.

- **Herramientas para la optimización y el mantenimiento del sistema**: migración de procesos, *checkpoint-restart* (detener uno o varios procesos, llevarlos a otro nodo y continuar su funcionamiento), balanceo de carga, tolerancia a fallos, etc.

- **Ambientes de programación paralela**: los ambientes de programación paralela permiten implementar algoritmos que utilizan recursos compartidos: CPU (*central processing unit*), memoria, datos y servicios.

3.2.4. Tipos

Los clústeres pueden clasificarse de acuerdo con sus características. Se pueden tener clústeres de alto rendimiento (HPC, *high performance clusters*), clústeres de alta disponibilidad (HA, *high availability*) o clústeres de alta eficiencia (HT, *high throughput*).

- ***High performance***: se trata de clústeres que ejecutan tareas que requieren una gran capacidad computacional. Dentro de este tipo de clústeres, podemos encontrar todas las supercomputadoras actuales, como por ejemplo Summit, Sierra o nuestro conocido Mare Nostrum.

- ***High availability***: son clústeres diseñados para proporcionar disponibilidad y confiabilidad, se provee mediante software que detecta fallos del sistema y permite recuperarse frente a estos, mientras que en hardware se evita tener un único punto de fallo.

- ***High throughput***: son clústeres que están diseñados con el objetivo de ejecutar la mayor cantidad de tareas concretas en el menor tiempo posible.

3.3. Sistemas *grid*

La computación en *grid* o malla es un sistema de computación distribuida en el que todos los recursos de un número indeterminado de computadoras son englobados como un único superordenador de manera transparente.

Estas computadoras englobadas no están conectadas o enlazadas rígidamente, es decir, no deben estar necesariamente en el mismo punto geográfico. Se conectan entre sí mediante la red global de internet.

Los orígenes de la computación en *grid* se deben a la idea de la compartición de recursos. La práctica conocida como «computación distribuida» nos lleva a los inicios de la informática. En las postrimerías de los años cincuenta y a comienzos de los años sesenta, los investigadores se dieron cuenta de que necesitaban hacer más eficientes los sistemas que habían costado una fortuna: «Los sistemas pierden mucho tiempo esperando que los usuarios introduzcan datos». Los investigadores razonaron, entonces, que varios usuarios podrían compartir el sistema aprovechando el tiempo de procesamiento no empleado.

> Un ejemplo de proyecto de computación distribuida en *grid* que surgió con el patrocinio de la NASA en los años setenta y que todavía hoy continúa vivo es el SETI (*search for extra terrestrial intelligence*). Este proyecto intenta encontrar vida extraterrestre inteligente analizando las señales capturadas por radiotelescopios y satélites. Debido al gran volumen de datos que se capturan diariamente, utilizando solo un ordenador se tardarían miles de años en analizarlos. Por ello, el proyecto SETI distribuye los datos entre los participantes del proyecto, que son analizados cuando los ordenadores de los participantes están parados, y devuelven información en el caso de que encuentren algo destacado. Un punto interesante, que ha hecho perdurar este proyecto tantos años, es que cualquier persona puede unirse al proyecto cediendo su ordenador personal.

Las **características principales** de la computación en *grid* se detallan a continuación:

- **Sus recursos coordinados no están sujetos a un control central**. Un *grid* integra y coordina recursos y usuarios que trabajan con diferentes dominios –por ejemplo, estaciones de trabajo de usuarios frente a computadoras centrales, unidades administrativas diferenciadas de la misma organización o diferentes organizaciones–.

- **Utiliza un estándar abierto, protocolos e interfaces genéricas**. Un *grid* está formado por protocolos genéricos e interfaces que tienen como principales inconvenientes la autenticación, la autorización, el descubrimiento y el acceso a los recursos. Es importante que estos protocolos sean estándares y abiertos.

- **Entrega las calidades no triviales de servicio**. Un *grid* permite a los recursos que lo constituyen ser empleados de una manera coordinada, en-

tregando diferentes calidades de servicio, relacionadas por ejemplo con el tiempo de respuesta, el rendimiento, la disponibilidad y la seguridad, o la asignación de múltiples recursos.

Grid ofrece nuevas y más potentes vías de trabajo. A continuación, se detallan algunos ejemplos:

- **Portales científicos**: aprovechar la computación *grid* para resolver problemas científicos de gran complejidad.

- **Computación distribuida**: aprovechar la mayor capacidad que tienen las estaciones de trabajo para conseguir recursos sustanciales de computación. Por ejemplo, para hacer minería de bloques, como por ejemplo minar bitcoins o cualquier otro servicio basado en cadenas de bloques (*blockchain*).

- **Trabajo colaborativo**: *grid* permite trabajar en equipo compartiendo recursos, pero también los resultados de los diferentes estudios para su análisis. Desgraciadamente, también puede servir para atacar con *bots* que se comportan como ejércitos de ordenadores infectados haciendo un ataque común.

3.4. Agregaciones de servidores en la nube

¿En la computación en la nube, el actual paradigma de los servicios informáticos, sabemos o conocemos cómo se trabaja a nivel de servidores? La respuesta la encontramos, sin duda, en la clusterización de los servidores.

Del mismo modo que en el apartado «Servidores en la nube» se ha comentado la relación entre los tipos de servidores y los servicios ofrecidos, en este apartado asociamos la computación de servidores en la nube, en su mayor parte, con sistemas de balanceo de carga. Se tratará, pues, de virtualización sobre sistemas físicos en clústeres de balanceo de carga con alta disponibilidad.

La gestión de una infraestructura completa en la nube puede llegar a ser extremadamente compleja para un administrador, y por este motivo se han creado las herramientas que permiten facilitarla y centralizarla.

3.4.1. Gestión de la nube

La gestión de la nube se basa en técnicas, estándares y softwares especializados que permiten gobernar entornos complejos en la nube. Hay un consenso en las características principales que ha de cumplir una buena gestión de sistemas en la nube:

- **Provisión y gestión de recursos**: crear, modificar y borrar recursos, y también orquestar recursos controlando las cargas de trabajo.

- **Escalabilidad**: permitir ampliar y reducir recursos en la propia nube.

- **Automatización**: permitir la automatización de la gestión y la provisión de recursos.

- **Monitorización y seguimiento**: gestión de la monitorización del uso de los recursos asignados, y también el seguimiento y el control de los *logs*.

- **Seguridad**: gestión de todos los aspectos relacionados con la seguridad en una nube multicapa.

- **Migración, *backup* y recuperación de desastres**: ofrecer herramientas de migración en la nube, y también herramientas de *backup* y de gestión de desastres o de alta disponibilidad.

- **Optimización y reducción de costes**: gestionar la provisión de recursos eficientes para conseguir un óptimo consumo.

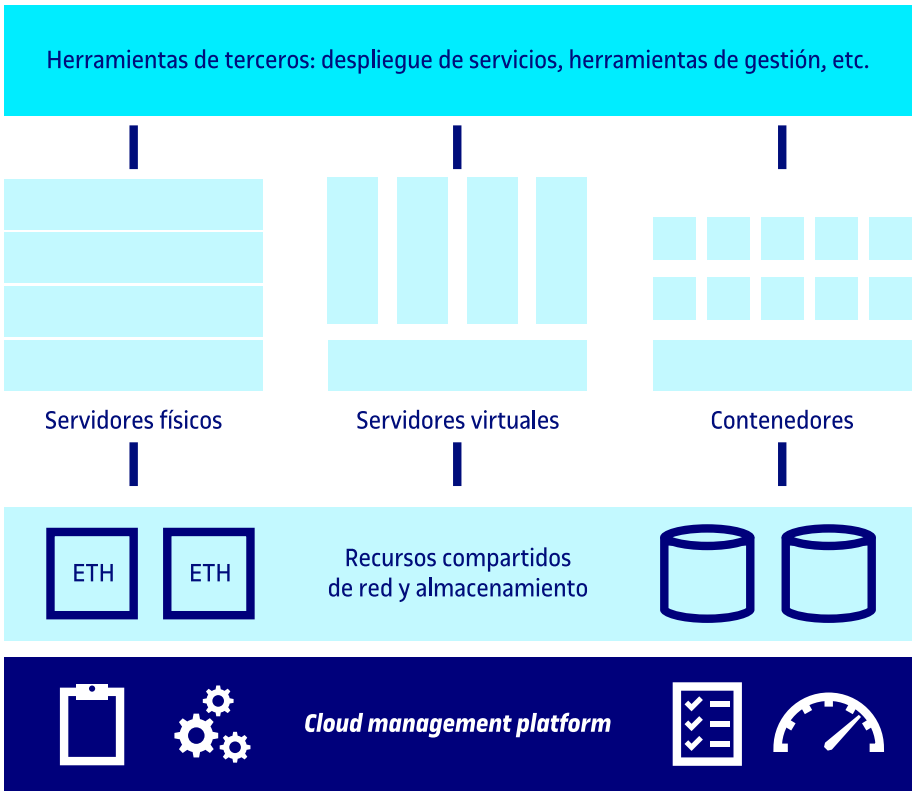
- **Multinube**: la gestión debe permitir trabajar con diferentes nubes, ya sean públicas, híbridas o privadas.

3.4.2. Software de gestión de las nubes

Hoy en día, existen múltiples sistemas comerciales de gestión de la nube denominados *CMP* (*cloud management platform*). Algunos de los CMP más conocidos actualmente son: OpenStack, Morpheus, Cloud Stack, Amazon Ec2, Open Nebula, Google Comput Engine, Rack Space Managed Cloud, IBM Cloud Private y Azure Virtual Machines.

La estructura básica de gestión de un CMP sería la que mostramos en la figura 9.

Figura 9. Estructura básica de un CMP



No todos los CMP tienen el mismo alcance en cuanto a la gestión de plataformas en la nube, ni se basan en los mismos principios. Por esta razón, podemos encontrar CMP que son literalmente sistemas operativos especializados, así como CMP que son software específico.

Definimos, a pesar de la diversidad de posibilidades, los elementos que hay que gestionar mediante los CMP:

- **Almacenamiento**: un CMP gestiona de manera común el almacenamiento que permitirá compartirlo con todos los servidores que lo necesiten.

- **Comunicaciones**: un CMP gestiona de manera común el acceso a las comunicaciones, segmentando las diferentes redes y validando la seguridad según sea necesario.

- **Servidores físicos**: un CMP gestiona los servidores *bare metal* de la nube para asignarlos según las necesidades de las organizaciones.

- **Servidores virtuales**: un CMP gestiona los hipervisores que permiten compartir servidores virtuales en las organizaciones.

- **Contenedores**: un CMP permite la gestión de los entornos de contenedores según las necesidades de las organizaciones en cada momento.

4. Almacenamiento

El **disco** es el componente del servidor que almacena los datos. Es un componente crítico porque contiene toda la información de la organización.

Nota

Si falla el disco físico, todas las particiones quedan inaccesibles y no se puede acceder a la información que contienen.

La capacidad, la velocidad, la seguridad y la ubicación de los discos son los aspectos básicos y más importantes que hay que tener en cuenta a la hora de elegir los discos que se quieren asignar a los servidores. Para esta elección habrá que tener muy en cuenta si se tratan de servidores físicos o virtuales.

¿Cuántos discos debe tener nuestro servidor? ¿Para qué los queremos?

Un disco es un espacio para guardar información que se divide en partes denominadas **particiones**. Pero, si las particiones pueden ser de muchos GB, ¿de qué sirve crear particiones?

Crear una partición de un disco tiene dos utilidades básicas. La primera, y la más importante, es que divide el disco en zonas independientes. Dado que está formateada independientemente, cada partición del disco es un disco lógico (no físico) diferente para el SO. Por lo tanto, en el caso de que por algún problema el sistema de ficheros quede corrompido y la información de dentro sea inaccesible, el contenido se pierde y la partición se debe reformatear. El resto de las particiones son accesibles y la información se mantiene intacta. Incluso se puede recuperar toda la partición de la copia de seguridad.

La otra utilidad es que, dado que son independientes, pueden estar formateadas en sistemas de ficheros diferentes. Por lo tanto, incluso podemos iniciar el ordenador desde diferentes particiones a partir de sistemas operativos distintos. Se utiliza mucho en la preparación de máquinas.

4.1. Particiones del disco

Las particiones estándares, que pueden necesitar los administradores de sistemas, son las siguientes:

- **Sistema**: la partición de sistema es necesaria para arrancar el servidor y para que funcione. Siempre se deja una única partición para el sistema operativo del servidor.

- **Usuarios**: la partición de usuarios contiene los directorios de los usuarios (las carpetas personales y, si las hay, las carpetas de grupo).

- **Datos**: en la partición de datos normalmente hay directorios con datos de programas que deben estar instalados localmente en las estaciones de

trabajo, datos compartidos por grupos de usuarios, y también puede haber un lugar para poner el «disco común», que es una carpeta común a toda la organización para transferir datos.

- **Aplicaciones básicas**: son las aplicaciones que usan todos los usuarios. El software base al que necesitan acceder todos los usuarios y que debe estar en la red. El permiso debe ser de lectura y ejecución para todo el mundo.

- **Aplicaciones**: esta partición contiene las aplicaciones que no son comunes a todo el mundo, por eso están separadas. Hay personas que las usan y otras que no. Se aplican permisos por grupos de usuarios. Además de las aplicaciones, muy posiblemente encontraremos datos asociados a las personas que acceden a ellas.

- **Otros**: teniendo en cuenta las necesidades reales de la organización, pueden ser necesarias otras particiones. Servidores de bases de datos, particiones por desarrollo, etc.

4.2. Sistemas de ficheros

Una vez que hemos creado las particiones necesarias en los discos, hay que hacer una operación adicional para que nuestro sistema pueda trabajar: se debe dar formato a cada partición. Este paso es imprescindible porque informa al sistema operativo y al disco de cómo se reparte el espacio (tamaño del sector) y de cómo se distribuirá lógicamente el disco. Por lo tanto, el resultado de este formateado genera un sistema de ficheros preparado para poder almacenar información.

Existen diferentes tipos de sistemas de ficheros y hay que conocer sus principales características para poder elegir el más adecuado según las necesidades de la organización.

4.2.1. Sistemas de ficheros locales

Son los más usuales y se refieren a los sistemas de ficheros de un servidor en concreto.

- **NTFS**. El sistema NTFS (*new technology file system*) de Microsoft se introdujo con la aparición de Windows NT y hoy continúa siendo el sistema de ficheros referente de las plataformas Windows. Tiene un tamaño de sector y de clúster muy pequeño, de modo que se aprovecha muy bien el espacio del disco. Posee firma de la partición, por lo que el disco no se puede leer en otro ordenador. Cuenta con seguridad en el sistema de ficheros y con múltiples funciones adicionales, como por ejemplo la compresión, la encriptación, etc. Por lo tanto, el conjunto hace que sea muy robusto.

- **ReFS**. Para sistemas Windows, Microsoft introduce ReFS (*resilient file system*) con funcionalidades nuevas que pueden, de manera precisa, detectar y resolver corrupciones mientras permanecen en línea, lo que ayuda a mejorar la integridad y la disponibilidad para sus datos. Sin embargo, las cualidades más apreciadas de este sistema de ficheros actualmente son las operaciones de aceleración con máquinas virtuales, gracias a su capacidad de gestionar eficientemente grandes cantidades de datos, lo que permite, por ejemplo, copiar de manera muy rápida discos de VM (*virtual machine*) o crear discos de VM nuevos en segundos en lugar de minutos utilizando particiones de NTFS. Por el contrario, este sistema de ficheros pierde una decena de funcionalidades respecto a NTFS, como por ejemplo la compresión, la encriptación y otras.

- **ext *family* (ext2, ext3 y ext4), XFS**. Son diferentes sistemas de ficheros empleados en los sistemas Unix y Linux. El tamaño de sector es de 256 bytes (muy pequeño). Tiene una estructura de inodos (en inglés, *index nodes*) para gestionar los ficheros y la seguridad de Unix Standard. Los más recientes y más utilizados actualmente (ext4 y XFS) ofrecen la posibilidad de trabajar con una gran cantidad de datos y con sistema de *journaling*, sistema mediante el cual se guardan periódicamente los archivos abiertos para evitar la pérdida de información o la corrupción de los datos si se produce una desconexión no planificada. Este sistema de ficheros aporta más seguridad, a pesar de que, por el contrario, hace perder recursos de máquina, asignados precisamente a la tarea de *journaling*.

- **APFS (Apple *file system*)**. Dada la importancia actual de las computadoras con plataforma Mac, hay que comentar el sistema de archivo de los sistemas Apple (APFS), que es el sistema de archivo por defecto para ordenadores de Mac con funcionalidades de encriptación, compartición de espacio, *snapshots* y otras. APFS está optimizado para el almacenamiento con dispositivos Flash/SSD.

Unionfs

Unionfs es un servicio de sistema de ficheros para entornos Linux. Permite disponer de diferentes sistemas de ficheros, denominados ramas. Unionfs realiza una superposición (*overlay*) de formato en un único y coherente sistema de ficheros. Este tipo de servicios se utiliza principalmente en contenedores, ya que necesita muchas veces acceder a datos comunes en el nivel de *host* para hacer más eficiente la creación del mismo contenedor y para almacenar datos permanentes. Algunos de los sistemas más conocidos son: AUFS, OverlayFS, btrfs, vfs y DeviceMapper.

4.2.2. Sistemas de ficheros distribuidos

Un sistema de ficheros distribuido permite almacenar ficheros en uno o más ordenadores (servidores) y permite que sean accesibles a otros, denominados *clientes*, utilizando la misma semántica de acceso que los sistemas de ficheros locales.

- **NFS**: el ejemplo más claro de los sistemas de ficheros distribuidos es el NFS (*network file systems*).

- **Sistemas de ficheros basados en objetos**: algunos sistemas de ficheros distribuidos utilizan una arquitectura basada en objetos, donde los metadatos se almacenan en servidores de metadatos y los datos se almacenan en servidores de almacenamiento de objetos. El software del cliente de fi-

cheros interacciona con los diferentes servidores para presentar un sistema de ficheros completo a los usuarios y las aplicaciones. De este nuevo sistema de ficheros surgen servicios en la nube tan conocidos como el S3 de Amazon o el COS de IBM.

4.2.3. Sistemas de ficheros en clúster

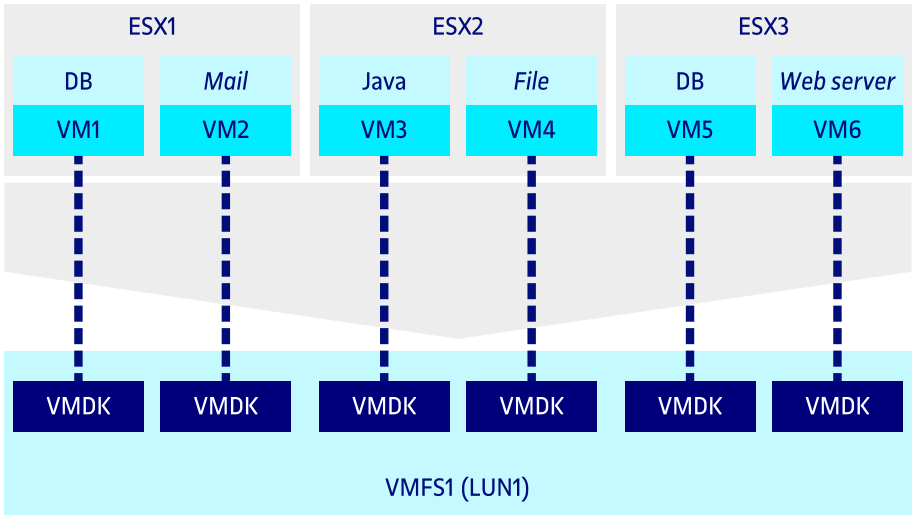
Sistemas de ficheros que se comparten en diferentes servidores. Todos los servidores del clúster gestionan de manera compartida el mismo sistema de ficheros, consiguiendo funcionalidades como por ejemplo la deslocalización, la redundancia, la fiabilidad y el aumento del rendimiento. Estos sistemas de ficheros se utilizan en la actualidad para la virtualización y el tratamiento de enormes cantidades de datos en el *big data analytics*, *blockchain* o los modelos de inteligencia artificial. Algunos ejemplos de sistemas de ficheros en clúster pueden ser: VMFS (VMWare *file system*), IBM Spectrum Scale o GPFS (IBM *general parallel file system*), OCFS (Oráculo *cluster file system*), BeeGFS.

Nota

Para sistemas virtuales en plataforma de virtualización Microsoft (Hyper-V), los sistemas de ficheros son los mismos que para servidores con sistemas de ficheros locales.

VMFS (VMWare *file system*) es, quizá, el sistema de ficheros en clúster más conocido actualmente. Diseñado, construido y optimizado por entornos de virtualización, incrementa los recursos de utilización mediante el acceso compartido y consolidado a un *pool* del clúster. Permite servicios como los *snapshots* y otros servicios propios de VMWare como: VMware vSphere Thin Provisioning, VMware vSphere vMotion®, VMware vSphere Distributed Resource Scheduler™ (vSphere DRES), VMware vSphere High Availability (vSphere HA), VMware vSphere Storage DRES™ y VMware vSphere Storage vMotion®.

Figura 10. Sistema de ficheros VMFS



Fuente: www.vmware.como.

4.3. Tipos de discos

Una vez que tenemos el sistema de fichero que se requiere para las necesidades de nuestra organización, hay que decidir en qué almacenamiento físico lo dispondremos. Por lo tanto, es necesario conocer en primera instancia qué tipos de discos existen y cuáles son sus características. Así, describiremos los discos según su tipo y su interfaz de transferencia de datos.

4.3.1. Discos físicos

En la actualidad, hay básicamente dos tipos de discos físicos:

1) HDD (*hard disk drive*): dispositivos de almacenamiento persistente formados por platos o discos superpuestos giratorios. Los datos se modifican magnéticamente mediante un cabezal.

2) SDD (*solid state drive*): dispositivos de almacenamiento persistente que utiliza la memoria Flash en celdas de memoria que consisten en puertas lógicas NAND o NOR.

4.3.2. Interfaz de transferencia de datos

En cuanto a la interfaz de transferencia de datos entre los discos y el sistema, podemos especificar los tipos siguientes:

1) SATA (serial ATA):[3] sustituye la tradicional *parallel* ATA o P-ATA (estándar que también es conocido como IDE o ATA). El S-ATA proporciona velocidades más altas, mayor aprovechamiento cuando hay varios discos y capacidad para conectar discos en caliente.

(3)Acrónimo de *serial advanced technology attachment*.

2) SAS: la interfaz SAS (*serial attached* SCSI) sustituye a la antigua SCSI. Implementa la transmisión en serie entre el controlador y los dispositivos, lo que permite obtener mejoras significativas:

- Incrementa la velocidad de transmisión: en las actuales versiones, esta velocidad se ha situado en 12 Gb/s (SAS 3.0) y 22,5 Gb/s (SAS 4.0).

- Incrementa el número de dispositivos SAS conectados y permite la conexión de discos «en caliente»: permite añadir discos a la configuración, mientras el sistema está funcionando con normalidad.

3) NLSAS: los discos Nearline SAS combinan discos SATA con interfaz de comunicación SAS.

4) NVMe: la interfaz de comunicación NVMe (*non-volatile memory host controller interface specification*) establece un nuevo protocolo de comunicación con dispositivos Flash, aprovechando todos los niveles de paralelismo de este tipo

de discos y evitando utilizar un protocolo pensado para discos rotatorios. Esta nueva interfaz está cambiando el paradigma de la gestión de acceso al disco mejorando enormemente los rendimientos de transferencia de los datos. Se utiliza normalmente en canales PCie, a pesar de que ya podemos disponer de encapsulamientos en SAN: el NVMe-oF (NVMe *over Fabric*) permite el acceso y la gestión NVMe de discos en la SAN, mientras que NVMe/TCP permite accesos vía Ethernet del protocolo NVMe.

4.3.3. Combinaciones de discos e interfaces

Una vez especificados los diferentes tipos de discos e interfaces de comunicación, comentamos algunos de los modelos de discos más extendidos:

- **SAS HDD**: discos rotatorios con interfaz de comunicación SAS. Son discos económicos que pueden ser de gran capacidad, y hasta hace pocos años eran el referente a nivel de servidores.

- **SAS SSD**: discos de estado sólido con interfaz SAS. Son la primera generación de discos SSD, con una mejora de los tiempos de acceso al disco muy importante, pero con carencias en la interfaz de comunicación, dado que esta no está específicamente diseñada para este tipo de discos.

- **NL-SAS**: combinación de discos SATA con interfaz SAS. Permiten gestionar grandes capacidades de datos con un rendimiento óptimo y un precio muy económico. Suelen servir para almacenar datos históricos, de archivado o no críticos.

- **NVME SSD**: discos de estado sólido con una interfaz que aprovecha el paralelismo de acceso a los datos. Es el presente y el futuro de los discos a nivel de servidor. Son discos con un precio elevado, pero con un rendimiento muy alto.

4.4. Agrupaciones de discos en el servidor

Los discos siempre son una de las piezas clave en los servidores. Esto ha provocado diferentes aproximaciones tecnológicas para mejorar su capacidad y rendimiento. A continuación se explican las aproximaciones más importantes.

4.4.1. Multivolumen

¿Qué pasa si creemos que tendremos una base de datos que ocupará 2 TB y solo tenemos discos de 500 GB? Hay una solución mediante el SO que consiste en convertir cuatro discos de 500 GB en uno de 2 TB. Es la gestión multivolumen. En general, se trata de juntar varias particiones físicas en una sola partición lógica de un tamaño equivalente a la suma de los tamaños de las particiones.

Partición lógica total = 500 GB + 500 GB + 500 GB + 500 GB = 2.000 GB

La principal ventaja es que se puede obtener una partición del tamaño que se quiera juntando particiones de discos de otros tamaños.

El principal inconveniente es que si un disco falla físicamente (se estropea), no podremos acceder a ninguna de las particiones físicas que integran la partición multivolumen creada. Por este motivo, surge la protección por paridad de discos de multivolumen.

Pero **¿cómo funciona la protección por paridad de un disco?**

La paridad de disco trabaja a nivel de «bit» y se calcula mediante operaciones lógicas (normalmente XOR).

Veamos un ejemplo: supongamos un sistema con dos discos con los bits siguientes: Disco1: 1011 y Disco2: 0110. Si aplicamos el operador lógico XOR(1001, 0110) = 1101, tendríamos como bits de paridad: 1101, que se almacenarían en un disco adicional Disco3. Ahora podemos suponer que el Disco2 falla y deja de ser accesible; los datos de este disco se podrían calcular fácilmente con la misma operación XOR con los Disco1 y Disco3, ya que XOR(1011, 1101) = 0110 es precisamente el valor de los bits del Disco2.

4.4.2. Sistemas de redundancia de datos

El RAID (*redundant array of inexpensive disks*) permite gestionar el cálculo y la gestión de los bits de paridad de un multivolumen de discos. Incluso, en caso de una avería en uno de los discos que formen parte de este, permite su cambio en caliente, es decir, sin detener el servidor se puede sustituir el disco que ha dejado de funcionar por uno nuevo, que el propio RAID vuelve a poner en funcionamiento, y reconstruye los datos a partir del resto de los discos y de los bits de paridad.

El RAID se puede ejecutar de varias maneras, según el grado de velocidad y seguridad que se necesite. Se clasifican en niveles:

1) RAID 0. La información se distribuye en varias unidades, pero no hay redundancia. Por lo tanto, no hay protección en caso de fallo del disco.

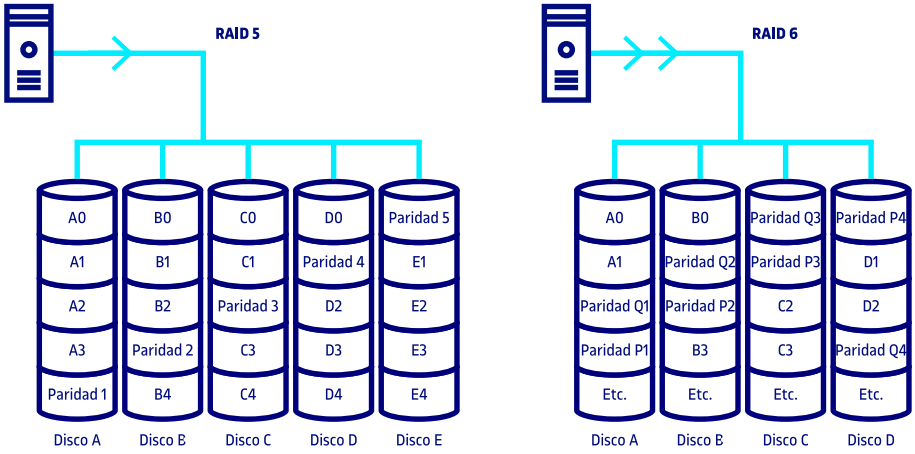
2) RAID 1. También denominado *espejo*. Cada unidad está duplicada con una unidad de apoyo. Por lo tanto, con seis unidades de disco, tres son de copia. La información se distribuye entre parejas de discos.

3) RAID 2 a RAID 4. Niveles de RAID que ya no se utilizan, pues hay niveles mucho más óptimos.

4) RAID 5. Se escriben en todos los sectores de todas las unidades y se añaden códigos correctores en cada sector. Este nivel de RAID ofrece una escritura más rápida porque la información de redundancia se distribuye a todas las unidades. Las lecturas del disco también tienen unos tiempos de acceso muy buenos.

5) RAID 6. Este nivel de RAID es similar al 5, pero utiliza dos códigos correctores para cada sector y un grupo de RAID. Se puede ver en el ejemplo de la figura 11. Las informaciones de paridad se distribuyen entre todos los discos del grupo.

Figura 11. Ejemplos de RAID 5 y 6

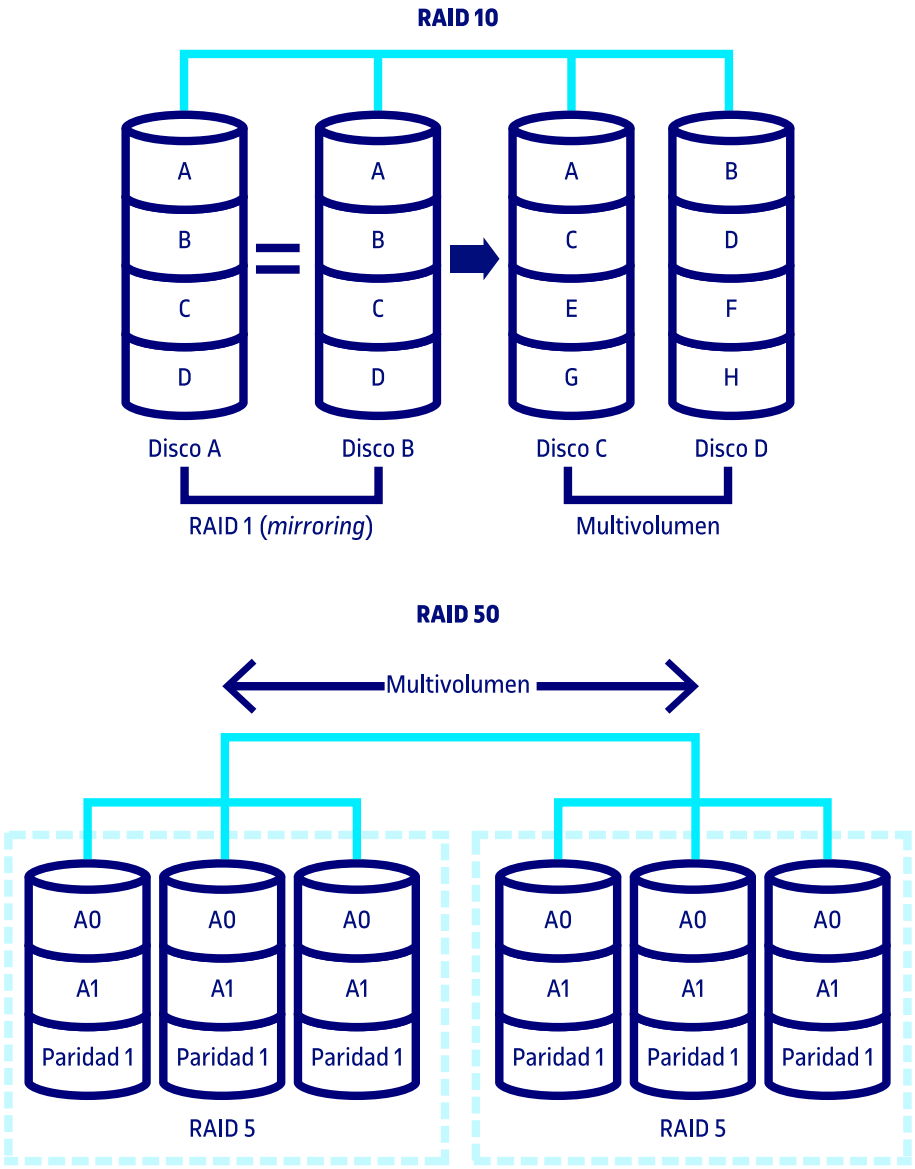


6) RAID 5e y 6e. Estos dos niveles de paridad se basan en sus predecesores (5 y 6), pero añaden un elemento más de seguridad: el *hot spare*. Este es un nuevo disco en espera que entra a formar parte del grupo de RAID activamente cuando uno de los discos de este deja el grupo.

7) RAID 10. Aparecen varias combinaciones de niveles de seguridad, a partir de los niveles básicos comentados. Uno de estos es el nivel 10 = 1 + 0, que replicaría un grupo de RAID 1 en un grupo de discos con RAID 0.

8) RAID 50. Como se puede ver en la figura 12, un grupo de nivel 50 = 5 + 0 distribuiría la información por multivolumen entre dos grupos de RAID 5.

Figura 12. Ejemplos de RAID 10 y 50



Actualmente la tendencia es utilizar los *distributed* RAID. Esta técnica permite distribuir los bits de paridad de los diferentes niveles de RAID en todos los discos que forman parte de él. De este modo, el *distributed* RAID permite reconstruir un disco, en caso de fallo, de manera mucho más eficiente, reduciendo el riesgo de perder discos adicionales en este periodo.

La técnica del RAID mejora el rendimiento, ya que distribuye la información entre las diversas unidades y puede ofrecer redundancia para aumentar la seguridad.

Una vez más, el RAID puede ser por software o por hardware. Si es por software es más lento, y si es por hardware es transparente al SO.

Hay una gran cantidad de sistemas de RAID comerciales internos y externos, pero citaremos algunos fabricantes que se pueden encontrar en la web: Dell (PowerVault), Compaq, StorageTeK, Clarion, Hewlett Packard, IBM, RaidTec, etc.

4.5. Sistemas de almacenamiento

Dentro del apartado del almacenamiento, solo queda detallar los diferentes sistemas de gestión de los discos, según su utilidad empresarial.

4.5.1. Disco interno

Las tecnologías «tradicionales» de almacenamiento se basan en la conexión directa (física) del dispositivo al servidor. Como consecuencia, las aplicaciones y los usuarios hacen las peticiones directamente al sistema de ficheros. Así pues, hay un controlador de discos que implementará el RAID en los discos, llamados *internos*, conectados al servidor.

El problema principal de esta tecnología es que crea islas de información, en las que cada servidor controla su almacenamiento independientemente del resto.

Para resolver este problema hay que emplear una nueva técnica que nos permita la compartición global de los recursos de almacenamiento. Existen diferentes soluciones dependiendo de las necesidades.

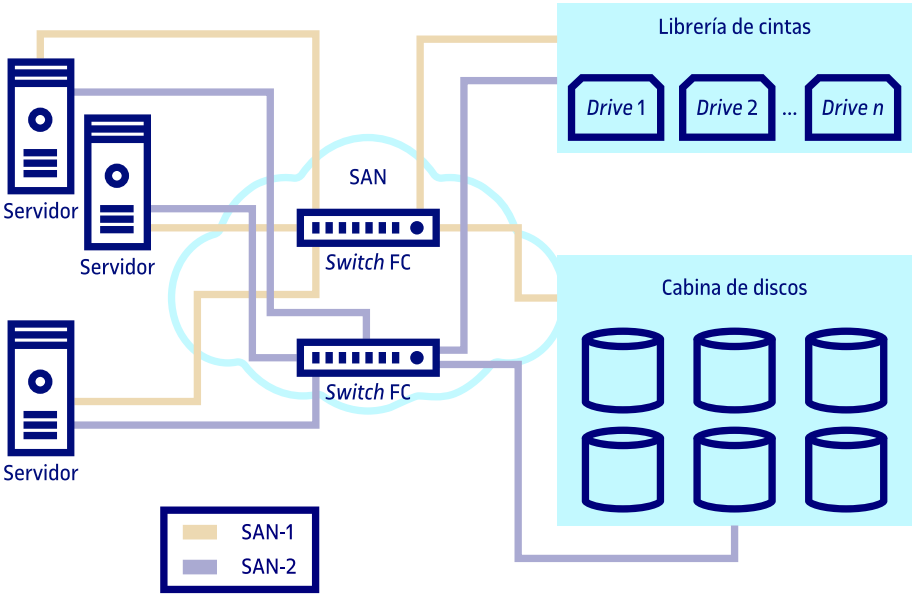
4.5.2. Redes de área de almacenamiento

Principalmente, se utilizan las redes SAN (*storage area network*). A continuación, se definen las principales características de estos tipos de redes.

La red SAN es una red especializada de alta velocidad que comunica los servidores y los dispositivos de almacenamiento. Una SAN también puede ser un sistema de almacenamiento formado por elementos y dispositivos de almacenamiento, computadoras, aplicaciones o software de control, y todos estos elementos se comunican mediante una red. El acceso a los datos se lleva a cabo a nivel de bloque I/O.

Los elementos de una red SAN, tal y como se puede ver en la figura 13, se pueden dividir en tres grandes grupos: servidores, elementos de conexión y almacenamiento.

Figura 13. Elementos de una red SAN



1) Servidores: forman parte de una red SAN todos aquellos servidores que disponen de tarjetas específicas HBA (*host bus adapter*).

Ved también

Podéis ver el subapartado «Tarjetas I/O» para establecer la comunicación con los elementos de conexión.

2) Elementos de conexión: forman parte de este grupo:

a) Cableado: específico para las redes SAN, suele ser cable de fibra óptica. Hay dos tipos, cableado multimodo de fibra de 50 micrones para distancias cortas y monomodo para distancias largas (menos de 10 micrones).

b) Conmutador: conmutadores (denominados *switch* FC) especializados en comunicación en redes SAN. Cada conmutador FC puede formar parte de un *fabric* o *switched fabric*, que es uno o más conmutadores formando una única red SAN.

c) Directores: conmutador principal. Punto central de gobierno de las redes SAN. Gobiernan una red SAN formada por diferentes conmutadores.

3) Almacenamiento

a) Sistemas de discos (llamadas *«cabinas» de discos*): dispositivos especializados en servir almacenamiento virtual de disco. Las cabinas actuales trabajan con *pools* de almacenamiento que están formados por agrupaciones de discos que pueden ser de diferentes tipos, como hemos visto en los subapartados «Tipos de discos» y «Agrupaciones de discos en el servidor».

Todos los discos de una agrupación deben ser del mismo tipo y las cabinas clasifican estos tipos en capas o *tiers*. También suelen tener mecanismos para mover de manera automática los datos entre las capas según el uso que se

haga de ellos (esta tarea se denomina *tiering*). Las capas o *tier* 0 son las capas más rápidas, formadas por discos SSD SAS o NVMe y son las capas donde se moverán los datos más accesibles.

Las cabinas de discos permiten gestionar inteligentemente los volúmenes virtuales que se presentan en los servidores por medio de la red SAN. Así, ofrecen servicios como:

- Crear volúmenes virtuales *thin*, donde la capacidad del volumen definida no se reserva directamente. El volumen va ampliando su capacidad a medida que la necesita.

- Crear volúmenes encriptados: permiten la encriptación de los datos con varios sistemas, según el fabricante.

- Creación de volúmenes virtuales comprimidos para reducir el espacio necesario para el almacenamiento de los datos.

- Deduplicación de datos en volúmenes virtuales. Permite reducir el espacio de almacenamiento empleando la técnica de deduplicación.

- Creación de *snapshots* de volúmenes virtuales, duplicando en un tiempo muy reducido grandes volúmenes de datos para su uso inmediato.

- Replicación remota de volúmenes virtuales en otras cabinas compatibles para disponer de un sistema de recuperación de desastres.

Deduplicación

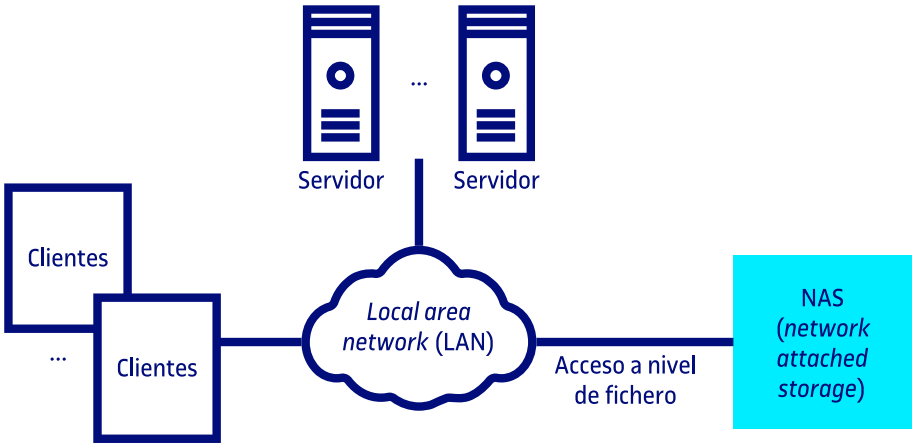
La deduplicación es una técnica que permite la compresión de datos mediante la eliminación de bloques de memoria repetidos, que son sustituidos por apuntadores a un único bloque original.

b) Sistemas de cintas, básicamente **librerías de cinta**, como elementos de gestión de grandes volúmenes de datos para *backup* y dispositivos de cinta.

4.5.3. Almacenamiento conectado en red

Principalmente, se utilizan las redes de tipo NAS (*network attached storage*). Este tipo de redes son básicamente un hardware especializado o un servidor de ficheros conectado a una red que sirve ficheros utilizando un protocolo. Un elemento NAS consiste en una máquina que implementa los servicios de ficheros (empleando protocolos de acceso como, por ejemplo, NFS o CIFS) y uno o más dispositivos, donde los datos están almacenados. Se puede observar en la figura 14.

Figura 14. Esquema de un NAS



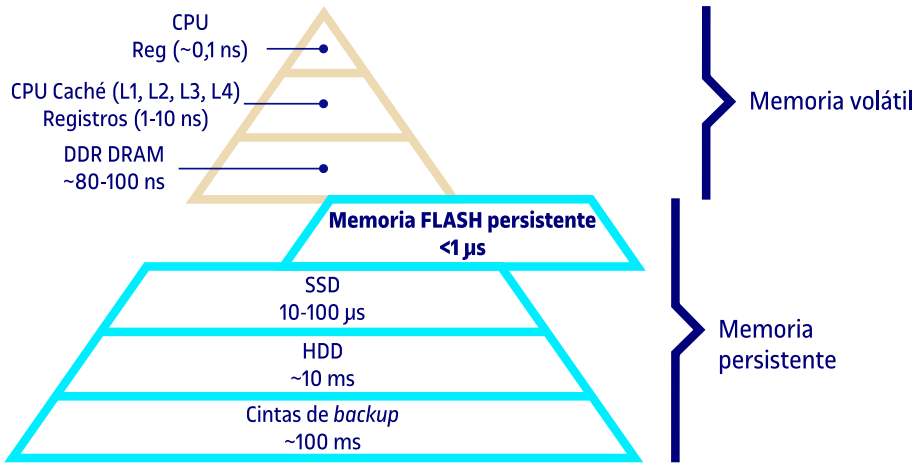
Un NAS proporciona capacidad de almacenamiento empleando la misma red de comunicaciones o una adicional de bajo coste. Así como los accesos que se realizan en una SAN son a nivel de bloque I/O, en un NAS se realizan a nivel de sistema de ficheros. Es decir, las aplicaciones acceden al sistema de ficheros que proporciona el propio dispositivo NAS, mientras que en una SAN el sistema de ficheros pertenece al mismo servidor.

A menudo se debe tomar una decisión de diseño y elegir entre una SAN o un NAS. Esta decisión ha de estar reforzada por dos premisas básicas: la velocidad de comunicación (rendimiento de acceso al disco) y el sistema de acceso (acceso a nivel de bloque o de ficheros).

4.5.4. Sistemas de memoria persistente Flash

Dada la creciente necesidad de mejora de los tiempos de respuesta de las peticiones de entrada/salida en servidores críticos, ha aparecido recientemente una nueva capa de almacenamiento intermedio que agiliza los accesos. Esta capa está basada en dispositivos Flash (normalmente con puertas lógicas NAND o NOR).

Figura 15. Pirámide de memoria



Por lo tanto, tenemos soluciones comerciales que cubren estas necesidades. Quizá las más conocidas sean actualmente Intel Optane e IBM Flash Storage.

4.5.5. Hiperconvergencia

Actualmente, en el mercado hay otras soluciones comerciales que ofrecen herramientas de compartición del almacenamiento. La mayoría pasa por la **virtualización de los recursos** o, dicho de otro modo, la hiperconvergencia.

La hiperconvergencia de infraestructura (en inglés, HCI) combina, por un lado, hardware común dentro de un centro de proceso de datos, como por ejemplo los recursos de almacenamiento locales, y por otro, un software inteligente. La hiperconvergencia proporciona recursos compartidos a toda la infraestructura provenientes de elementos locales aislados.

Las **ventajas** de la HCI son varias y facilitan a las organizaciones un nuevo método de gestión:

- Optimización de costes en infraestructura, eliminando el aislamiento de los componentes y permitiendo una integración de recursos bajo el paraguas de una gestión eficiente, centralizada y basada en software.

- Simplificación de la gestión de recursos, pues su gestión es centralizada y gestionada por software. Se evita una gestión dedicada para cada elemento de la infraestructura.

- Facilidad en la escalabilidad de los recursos porque se trata de recursos compartidos entre todas las necesidades de la organización.

- Mejora de la eficiencia y el rendimiento de la infraestructura, permitiendo reajustar los recursos existentes a las necesidades.

Algunos de los **productos comerciales** más conocidos son:

- **VSAN de VMWare**: es un almacenamiento definido por software (SDS), producto desarrollado por Vmware, que virtualiza los discos DAS asignados a los servidores ESX que forman parte de la granja del clúster. Este *pool* puede ser utilizado por todas las máquinas virtuales de la granja.

- **NUTANIX**: líder en el mercado para productos de hiperconvergencia, permite reaprovechar todo el almacenamiento de servidores físicos de la infraestructura para crear un *pool* virtual totalmente accesible desde cualquier servidor físico o virtual.

- **CISCO HyperFlex HX-Series**: solución de Cisco para la hiperconvergencia que permite compartir recursos de computación, red y también almacenamiento.

4.5.6. Soluciones híbridas

Los sistemas NAS y SAN no son excluyentes, todo lo contrario, se pueden combinar para dar todavía una mayor flexibilidad y servicio de almacenamiento a los servidores. Un ejemplo sería tener un servidor conectado a un NAS que tiene los discos en un sistema de almacenamiento conectado a una SAN. Otro ejemplo podría ser un sistema de alto rendimiento con una capa de memoria persistente Flash, una capa de negocio con discos SSD con interfaz NVME y, finalmente, una capa de almacenamiento histórico gestionada por hiperconvergencia sobre discos NL-SAS.

5. Copia de seguridad

Ante el problema de copiar la información de la organización para evitar su pérdida, hay muchos dispositivos (aparatos físicos) y técnicas. Debemos buscar los mejores para cada caso.

Los dispositivos de copia de seguridad son los aparatos físicos que se utilizan para hacer copias de seguridad de la información de los servidores. Normalmente, las copias son procesos que tardan horas en terminarse, y también se tarda un tiempo similar o superior en recuperar los ficheros del dispositivo en el que se ha almacenado la información.

5.1. Políticas de copia de seguridad

Una buena política de copias de seguridad es la clave para tener segura la información de la organización.

Algunos de los motivos para hacer copias de seguridad son los siguientes:

- Proteger la información contra un fallo del sistema o algún desastre natural.
- Proteger la información de los usuarios (los ficheros) contra borrados accidentales.
- Proteger la información de los usuarios y de la organización contra ataques por parte de terceros.
- Duplicar la información de los usuarios por seguridad, ya que se pueden dar casos de usos incorrectos que la dejen inconsistente o la modifiquen incorrectamente.
- Posibilitar un traspaso de la información cuando se actualiza o se reinstala el sistema.

5.1.1. Tipos de copias de seguridad

Según la necesidad definida por la organización, el administrador de sistemas deberá definir las copias según los tipos que se detallan a continuación:

1) Copia de seguridad completa. También se conoce con el nombre de copia de seguridad total o copia *full backup*. Se hace una copia de toda la partición del disco en el dispositivo de copia elegido. A menudo la copia se hace con-

siderando el formato del disco y sin tener en cuenta el sistema de ficheros, ya que solo hay que conocer la tabla de particiones del disco y en qué parte está la partición para duplicarla en un dispositivo de cinta. En estos casos, la restauración no puede ser selectiva, se ha de restaurar toda la partición y no se puede seleccionar solo un fichero. También encontramos la copia de seguridad completa del sistema de ficheros, en la que sí que es posible una restauración selectiva o granular.

2) Copia de seguridad incremental. En este caso, se guardan solo los ficheros que se han modificado desde la última copia de seguridad que se ha realizado. Las copias de seguridad incrementales se utilizan junto con las copias de seguridad completas en lo que se denominan políticas de copias de seguridad.

3) Copia de seguridad selectiva. También es posible hacer una copia de unos ficheros determinados. Normalmente esta acción se lleva a cabo con ficheros de comandos.

4) Copia de seguridad diferencial. Este nuevo tipo de copia realiza una copia de todos los ficheros que se han modificado desde la última copia total. Así, si hacemos una copia total cada sábado y diferencial el resto de los días, la copia del viernes contendrá todos los ficheros modificados desde el sábado. Tenemos varias ventajas de la copia diferencial respecto a la copia total. La primera, y como es natural, es que requiere menos espacio, y la segunda, asociada a la primera, es que reduce el tiempo o ventana de copia. Respecto a la copia incremental, aporta a su vez la ventaja de que en el proceso de recuperación solo necesitaremos la última copia total y la última copia diferencial. No obstante, la copia diferencial, a partir del segundo día, requerirá más espacio y más tiempo o ventana de copia.

5.1.2. Políticas de copias de seguridad

La estrategia sobre cómo se deben realizar las copias de seguridad es crítica para asegurar que se haga todo correctamente y que se pueda restaurar la información cuando sea necesario.

La necesidad de crear **estrategias de copias de seguridad** proviene del hecho de que actualmente los servidores disponen de mucha capacidad y, por lo tanto, hay mucha información (tanto de usuarios como de sistema), y toda esta información puede no caber en un solo dispositivo de salida (en una sola cinta, por ejemplo). Finalmente, la transferencia dura horas y, por lo tanto, hay que buscar soluciones para optimizar su uso.

Analizamos la variabilidad de la información. Con un vistazo rápido, nos podemos dar cuenta de lo siguiente:

- Hay información que varía diariamente.
- Hay información que se modifica muy poco a lo largo del tiempo.

- Hay información que no es necesario guardar en copias de seguridad (los ficheros temporales, por ejemplo).

Por lo tanto, una estrategia de copia que lo copie todo diariamente no parece muy acertada.

Sí que parece evidente que debemos realizar una copia diaria de la información que varía cada día (suele ser la información de la organización). Se puede encontrar en los servidores o distribuida por toda la organización. En cualquier caso, es necesario que hagamos una copia diaria de estos datos.

Con la información sobre la cantidad de datos que hay que copiar (el volumen) y sabiendo el dispositivo en el que queremos hacer la copia, tenemos una idea aproximada de los dispositivos de copia que necesitamos.

5.2. Dispositivos

5.2.1. Unidades de cinta

Uno de los dispositivos de copias más utilizado actualmente son los LTO (*linear tape open*), que fueron desarrollados por Hewlett Packard, IBM y Seagate.

Estos tipos de cintas han ido evolucionando rápidamente. Mientras que en el año 2000 hablábamos de LTO1, que permitía hasta 100 GB de copia por cinta, la capacidad de las cintas LTO8 actuales puede llegar hasta 12 TB sin compresión (30 TB con compresión). Ya se trabaja en nuevas versiones para los próximos años. La versión LTO12 podrá almacenar entre 192 y 480 TB de datos.

Su velocidad de copia puede llegar a 900 Mb/s en la LTO8 y ya está planificada la versión LTO10, que permitirá una velocidad de 2.750 Mb/s.

Hay otras cintas, como por ejemplo la Storagetek de Oracle, o los discos extraíbles RDX para sistemas con poco almacenamiento, que no son muy utilizados actualmente.

5.2.2. Disco duro o cintas virtuales

Hoy en día, teniendo en cuenta la continua evolución de las tecnologías, no se ha de descartar nunca la posibilidad de hacer una copia de seguridad (o, incluso, de copiar toda la información) en otro disco duro solo dedicado a esta función o, incluso, en la nube.

La estrategia es hacer una primera copia de seguridad en el disco (se puede realizar con un procedimiento automático y varias veces al día, si es necesario), y de este disco, posteriormente, se hará una copia de seguridad en otro dispositivo (que puede ser una cinta, un almacenamiento en la nube, etc.).

A veces, esta estrategia es necesaria si el procedimiento de copia necesita bloquear la información a la que accede y es, por ejemplo, una gran base de datos de la que depende toda la organización. La copia de disco a disco, ya sea gestionada por las cabinas de almacenamiento con *snapshots* de máquinas virtuales, o gestionada internamente por medio de los buses del sistema y con velocidades de transferencia muy elevadas, necesita bloquear muy poco tiempo la información para realizar la copia. Por lo tanto, la interrupción para hacer esta tarea es prácticamente imperceptible.

5.2.3. Tendencias

Gracias a la proliferación de las redes SAN o los dispositivos NAS que permiten una gran cantidad de espacio para almacenar, se utilizan cada vez más los discos como dispositivo de copia. Los propios proveedores ofrecen herramientas específicas que permiten realizar estas copias transparentes en el mismo sistema.

5.3. Librerías de copia

Se puede dar el caso de que nuestra organización manipule cantidades de datos que ocupen varias cintas de copia al día. En este caso, una sola persona se pasaría el día haciendo copias de seguridad y no acabaría nunca. ¿Cuál es la solución para estos volúmenes de información tan grandes? Existen unos dispositivos denominados *librerías de copia* (o de cintas). Son externos, con unos brazos articulados y contienen desde veinte hasta dos mil cintas de copias de seguridad (son como robots). Con el software adecuado, esto se ve, por ejemplo, como una unidad de 400 PB para guardar información. El software sabe en qué cinta está almacenada la copia y qué cintas están llenas, y gestiona la política de sustitución de las cintas. Las librerías de copia solo tienen sentido para organizaciones de grandes dimensiones o que gestionan cantidades de información muy grandes.

5.3.1. Librerías de cintas físicas

Hay varias marcas que fabrican librerías en colaboración con marcas de software para que puedan funcionar correctamente con los servidores en los que se instalen.

Algunas de estas marcas, con webs para poder ver los aparatos, son Hewlett Packard, Oracle, IBM, etc.

5.3.2. Librerías de cintas virtuales (VTL, *virtual tape library*)

Ved también

Podéis ver el subapartado «Unidades de cinta».

Partiendo de la base que disponemos de cintas virtuales, también hay librerías de cintas virtuales.

Las VTL o librerías virtuales virtualizan el almacenamiento interno o externo mediante una SAN y ofrecen dispositivos y cintas virtuales a los sistemas por medio de la propia SAN o mediante softwares específicos de copias. Los fabricantes de cabinas de disco también suelen ofrecer sistemas VTL asociados que tienen propiedades parecidas a las cabinas de discos comentadas anteriormente.

Las VTL permiten una gestión más dinámica de las unidades de cinta necesarias para la realización de las copias de seguridad. Algunas de las mejoras que ofrecen respecto a las librerías de cintas físicas pueden ser las siguientes:

- Reducen la gestión manual de las cintas.
- Ofrecen sistemas de compresión, deduplicación y replicación, al igual que las cabinas de discos.
- Permiten una mayor velocidad de escritura que las cintas físicas.
- La replicación permite un sistema de recuperación de desastres más amplio.
- Permite crear tantos dispositivos virtuales como sea necesario (según las limitaciones de cada fabricante), lo que a su vez permite la ejecución de copias y restauraciones en paralelo.
- Las duplicaciones en cintas físicas son más rápidas.

5.3.3. Copias de seguridad en la nube

Ved también

Podéis ver el apartado «Sistemas de ficheros distribuidos».

Actualmente, existe un nuevo sistema de almacenamiento donde nuestra organización puede guardar las copias realizadas. Esta nueva ubicación es la nube. La mayoría de los grandes proveedores de servicios en la nube ofrecen un almacenamiento económico donde podemos guardar las copias de seguridad de nuestros servidores. Por lo tanto, aparece un concepto nuevo, que son los servicios de ***cloud object storage*** (AWA S3 de Amazon, GCS de Google, IBM COS, Alibaba OSS, Azure Blob Storage y Oracle Storage son algunos de los proveedores más destacados).

Hay protocolos estándares de almacenamiento que aprovechan los softwares de gestión de copias de seguridad para acceder a los contenedores de espacio o *buckets* de la nube y gestionar los *backups*.

Aprovechando, pues, este nuevo servicio de almacenamiento económico en la nube, podemos optar por esta solución y evitar los dispositivos dedicados, ya sean físicos o virtuales.

También existen multitud de servicios BaaS (*backup as a service*) que permiten gestionar directamente nuestras copias de seguridad sencillamente teniendo una buena conexión de los servidores de la organización en la nube.

5.3.4. Tendencias

Las organizaciones necesitan aumentar constantemente su almacenamiento para poder gestionar las ingentes cantidades de datos que se generan. Así pues, las copias de seguridad han de ser dinámicas y se deben adaptar a las nuevas necesidades.

Por lo tanto, la tendencia actual es la utilización de dispositivos y librerías de cintas virtuales, y cada vez adquiere más relevancia la copia en la nube mediante el almacenamiento en COS o algún servicio de BaaS.

Otro punto importante que hay que comentar es la necesidad de centralizar toda la gestión de copias de seguridad mediante un software que lo haga posible. Hoy en día no se puede plantear un sistema de copias en el que cada servidor actúe independientemente del resto.

Hay multitud de **softwares de gestión de copias** según las necesidades. Los más comunes actualmente son Veeam Backup, NetBackup, IBM Spectrum Protect, etc. Estos nos ofrecen:

- Gestión centralizada del servicio.
- Configuración de políticas y dispositivos según las necesidades.
- Encriptación, compresión y deduplicación de los datos.
- Informes de gestión y seguimiento.
- *Backups* y recuperaciones especializados por tecnología.
- Etc.

5.4. ¿Dónde deben estar los dispositivos de copia?

Como ya hemos visto, los dispositivos podrán estar ubicados según las necesidades de nuestra organización, desde directamente conectados al servidor del cual se quiere hacer el *backup* (para organizaciones pequeñas), en librerías virtuales en centros de proceso de datos deslocalizados o, incluso, en la nube mediante COS o BaaS.

5.5. ¿Dónde se pueden guardar las copias de seguridad?

Las copias de seguridad tienen dos finalidades:

- Protegernos de fallos de los servidores.
- Proteger la información de la organización.

Antiguamente, los administradores tenían las copias con los servidores para poderlos recuperar rápidamente en caso de fallo. Hoy no es así, porque la virtualización de los dispositivos y el uso de los servicios en la nube han cambiado el paradigma de la gestión de las cintas.

En todo caso, las **recomendaciones** de seguridad y externalización de las cintas continúan vigentes.

- Siempre debería haber una copia de seguridad lo más actualizada posible fuera de la organización. Podría estar en una caja fuerte, en una empresa especializada en custodia de cintas o en algún servicio de almacenamiento o BaaS en la nube.

- Por otro lado, dadas las herramientas que existen hoy en día, se recomienda la encriptación de las copias de seguridad para todas aquellas que salgan de nuestra organización, ya sea física o virtualmente en la nube. En este último caso, también se recomienda la encriptación de los canales de comunicación.

6. Impresoras

Son otra familia de dispositivos que se conectan al sistema informático y están controladas por los servidores de la organización. Actualmente, las estaciones de trabajo no suelen tener impresoras conectadas físicamente y la organización tiene muy pocas en relación con el número de estaciones de trabajo, por lo que es un recurso compartido, gestionado por el servidor mediante una cola de impresión.

La **cola de impresión** es un recurso de software para conseguir que una impresora (inherentemente no compartible) pueda ser compartida.

Por lo tanto, en el servidor se deben crear tantas colas de impresión como impresoras haya que gestionar. Se ha de configurar la impresora para que se comporte como un dispositivo de red y, después, se debe configurar correctamente el servidor. Básicamente los pasos son los siguientes:

1) Conectar la impresora a la red: la conexión física se reduce a conectar la impresora a la red. Todas las impresoras actuales lo permiten nativamente.

2) Configurar el dispositivo de red de la impresora: siempre funcionan en el protocolo TCP/IP, por lo que tienen una dirección IP. La configuración del dispositivo, una vez conectado a la red y puesto en marcha, se puede hacer vía web o por medio del propio panel de la impresora. A partir de aquí se configuran todos los parámetros.

3) Declarar al servidor la impresora física (modelo, etc.): se ha de informar al servidor de que hay una impresora remota, la dirección IP que tiene, el tipo y el modelo de impresora y sus características relevantes.

4) Asociar una cola de impresión a esta impresora declarada: finalmente, hay que asociar una cola de impresión a la impresora remota que se ha creado y ponerla en marcha.

5) Compartir la cola de impresión con los usuarios de la organización: se realiza la compartición dependiendo del dominio o del tipo de sistemas de los usuarios. Con todo esto, los usuarios ya podrán enviar trabajos –que el servidor gestionará sin problemas– por la red a la impresora.

6.1. Tipos de impresoras

A continuación, se describen los diferentes tipos de impresoras que podemos encontrar en una organización. Evidentemente, no todas las organizaciones tendrán todos los tipos de impresoras.

1) Impresoras de chorro de tinta. Tienen una alta utilidad. Su coste es bajo y muchas veces están instaladas en mesas de despacho. Todas son de color (es inherente a estas impresoras). Funcionan según el principio de lanzar una gota de tinta electroestáticamente sobre el papel. Hoy en día se utilizan mucho en el ámbito doméstico pero poco en el empresarial.

2) Impresoras láser. Son las más extendidas y funcionan según el principio de dibujar la página en un tambor especial con un rayo láser y después transferirlo al papel con un polvo que se fija con calor. En la actualidad hay multitud de modelos, pero últimamente en las organizaciones se prefiere centralizar las impresiones en impresoras multifunción (escáner, impresora y fotocopiadora) con gestión de acceso y envío de correos o acceso directo al almacenamiento. Actualmente, su uso también está muy extendido en ámbitos domésticos.

3) Impresoras 3D. En los últimos años, el mundo de la impresión ha recibido una nueva familia de impresoras, las llamadas 3D, que permiten crear réplicas de diseños en formato 3D. Hay diferentes modelos según los materiales y el método de creación empleado. Es evidente que no todas las organizaciones necesitan impresoras de este tipo, pero hay que remarcar que su uso es cada vez más frecuente, dado que permiten crear piezas o maquetas totalmente adaptables de manera unitaria.

6.2. Protocolo de impresión en internet

Teniendo en cuenta que actualmente podemos tener servidores en la nube, adquiere fuerza el protocolo de impresión remoto. *Internet printing protocol* (IPP) define un método estándar de envío de trabajos de impresión empleando internet. Fue desarrollado por el consorcio de compañías del sector Printer Working Group.

IPP provee un estándar único y simple para gestionar los procesos de impresión. Como trabajan con TCP/IP, se pueden dirigir a una red local, a una intranet o a internet.

7. La corriente eléctrica

La corriente eléctrica es uno de los grandes olvidados en el momento de diseñar la disposición de los equipamientos. A pesar de ello, resulta que los servidores, las estaciones de trabajo, la electrónica de la red, las impresoras, los monitores, todos los dispositivos y toda la electrónica asociada a la informática están conectados.

Si disponemos de centros de procesamiento de datos propios (no en la nube) y nos limitamos a enchufar los equipos suponiendo que tendremos una corriente perfecta de 220 V, 60 Hz, 24 horas al día, 7 días la semana, 365 días el año, estamos gestionando de manera errónea nuestra infraestructura TI. Debemos pensar en la corriente eléctrica desde una perspectiva mucho más realista.

Centro de procesamiento de datos propio

Los servidores situados en un centro de proceso de datos local de la organización se denomina *on premises* u *on prem* en el argot TI.

Empecemos por estudiar la corriente eléctrica que pasa por la organización. ¿Cuáles son los problemas más habituales que nos puede dar?

- Picos de tensión.
- Caídas de corriente o microcortes.
- Proximidad con otras líneas. Las señales de otras líneas cercanas (de tensión o de datos) influyen en la calidad global de la tensión.
- Ruido, que es la suma de picos de tensión y caídas de corriente.

¿Qué **consecuencias** puede tener?

- **Pérdida o corrupción de datos**. Si afecta al equipo, puede ocasionar caídas no controladas que corrompan los datos con los que se estaba trabajando.

- **Daños en el equipamiento**. Si hay grandes sobretensiones, pueden destruir las fuentes de alimentación e, incluso, los chips de las placas. También puede estropear los controladores de disco (con la consiguiente pérdida de información), las memorias, las placas base, etc., por lo que el equipo ya no funcionará.

- **Desgaste prematuro**. Si un equipo está alimentado con corriente eléctrica de mala calidad (ruido), los circuitos electrónicos se desgastan antes de lo normal y el equipo falla sin motivo y de una manera aleatoria. Los chips degeneran de una manera desconocida y los resultados son imprevisibles. Entonces puede suceder que se produzcan errores de paridad pocos minutos después de haber arrancado el ordenador, cuando en principio ha superado correctamente los diagnósticos.

7.1. La toma de tierra

Según el informe *Power and Ground for Distributed Computing*, de David Fencl y Larry Fish, de ONEACH Corporation: los edificios tienen una toma de tierra de baja resistencia para proteger a la gente de choques eléctricos. La finalidad de la toma de tierra es que la corriente la siga porque hay menos resistencia y, por lo tanto, en caso de tocar algún aparato electrificado, la descarga no pase a través de la persona. Por este motivo, hay que seguir las normas generales a nivel eléctrico y asegurar que todos los *racks* de la organización tengan su propia toma de tierra.

7.2. Sistema de alimentación ininterrumpida

El SAI (sistema de alimentación ininterrumpida) protege a los servidores de cortes de corriente y otros problemas con la tensión.

La importancia de una buena corriente para los servidores se debe a que una falta de corriente repentina (corte) no le permitirá pararse correctamente. Esto provocará que las memorias caché se pierdan y no se actualicen en el disco (si no tienen baterías internas), con lo que quedarán sin guardar las transacciones que no se hayan completado. Es posible que, cuando se vuelva a poner en marcha el sistema, no se pueda poner en marcha completamente y se pierda información o ficheros. Si algún fichero es una base de datos, las consecuencias pueden ser desastrosas (se debe recuperar de la copia de seguridad, pero desde que se ha hecho hasta que se ha producido el corte se ha perdido la información y el tiempo invertido en generarla).

Un SAI suministra corriente cuando la red eléctrica no la da, de modo que el ordenador continúa funcionando correctamente, sin verse afectado por el hecho de que no haya suministro eléctrico general. Esto permite apagar los sistemas con total normalidad.

Las **características** más relevantes de un SAI son las siguientes:

- **Potencia que hay que suministrar**: son los vatios de potencia que puede dar el SAI cuando no hay corriente de entrada. Determina el número de servidores que le podremos conectar.

- **Tiempo de duración de las baterías**: los SAI tienen baterías que se cargan con la corriente eléctrica y son las que después dan electricidad cuando falla la corriente general. El número de baterías determina el tiempo que podrán suministrar corriente antes de agotarse.

- **Estabilizador**: esta característica significa que el SAI es capaz de suprimir el ruido. A pesar de ello, necesita una toma de tierra para desviar este exceso de corriente.

- **Tiempo de vida de las baterías**: un SAI sirve de poco si falla cuando debe funcionar. Las baterías tienen una vida útil determinada. Agotado este tiempo, no hay garantías de que funcionen y que respondan correctamente cuando sea necesario. Es el fabricante del SAI quien dice cada cuántos años se han de cambiar estas baterías.

- **Aviso al servidor**: actualmente los SAI cuentan con una línea (USB, serie o Ethernet) que llega al ordenador o a un sistema de control. De este modo, cuando entra en funcionamiento, es capaz de enviar una señal al servidor, que con el software adecuado (suministrado con el SAI) mantiene un diálogo en el que informa del estado de la alimentación y de las baterías. Cuando falta poco para agotar la carga de las baterías, el SAI informa al servidor y puede proceder a enviar mensajes a los usuarios y hacer una parada correcta, ordenada y automática del ordenador. Los servidores suelen estar preparados para arrancar solos, sin intervención del administrador, por lo que cuando se restablezca el suministro eléctrico normal el servidor se pondrá en marcha y todo volverá a funcionar correctamente.

8. Seguridad de los servidores

La seguridad es un tema muy amplio y está perfectamente cubierto en el módulo que trata de la seguridad informática. No obstante, aquí comentamos dos aspectos genéricos referidos a la seguridad de los servidores. Esta seguridad debe conocerla, aplicarla y tenerla en cuenta el administrador de servidores, y afecta básicamente al buen funcionamiento de los servidores corporativos.

8.1. Física

Todo sistema de seguridad, a pesar de que parezca muy evidente, empieza por la seguridad física, siempre que tengamos los servidores localmente en los centros de proceso de datos de la organización (*on premises*). No sirve de nada proteger todo el sistema informático contra todo tipo de ataques por red si es muy sencillo llegar a los servidores físicamente.

Si podemos acceder físicamente a un ordenador, podremos acceder a la información que contiene.

Esta premisa indica que la información está segura en la medida en que el servidor está físicamente seguro. Estas son algunas de las **precauciones** que se pueden tomar:

- Cerrar el recinto (o centro de proceso de datos: CPD) donde están los servidores.
- Control del acceso al CPD mediante tarjetas o elementos biométricos.
- Encriptación de los datos para evitar su lectura en caso de robo o, incluso, en caso de reparación por cambio de discos u otros elementos de los sistemas.

8.2. Software

También hay unas precauciones genéricas que se pueden aplicar a todos los sistemas operativos. Esta **seguridad de software** se orienta a dar unas indicaciones sobre las medidas generales que hay que tomar para tener un sistema más seguro, ya sea físico o virtual.

Ved también

En el módulo «Administración de la seguridad», podréis encontrar detallados todos los aspectos que hay que tener en cuenta respecto a la seguridad informática.

- Es necesario que las cuentas de administrador o superusuario tengan contraseñas bien construidas, con una política de cambio periódica. Las cuentas con privilegios especiales no deberían tener los nombres por defecto. Esto significa que, si es posible, en un ordenador Unix la cuenta de super-

usuario no debería ser *root*, y en una máquina Windows Server, la cuenta de máximos privilegios no tendría que ser *administrator* o administrador, porque de alguna manera se está dando pistas a los posibles atacantes.

- No ejecutar ni instalar software no necesario en el servidor, pues existe el peligro de instalar un virus o programas maliciosos.

- Disponer del último nivel de actualización de los sistemas para evitar «agujeros» de seguridad conocidos y explotables.

- Disponer de un sistema de antivirus, *malware* y otros tipos de ataques informáticos.

8.3. Alta disponibilidad

La alta disponibilidad es la capacidad de **mantener operativas las aplicaciones** de la organización, eliminando las paradas de los sistemas de información. Los sistemas informáticos se deben haber configurado para reducir al mínimo porcentaje el tiempo de inactividad o de falta de disponibilidad, para conseguir la máxima cuota de utilidad. La alta disponibilidad de un sistema se consigue cuando se reduce al mínimo la posibilidad de que un error de hardware o un defecto de software implique la interrupción de uso del sistema o la pérdida de datos del sistema.

8.3.1. Mito de los 9

Es el tiempo que un sistema está activo al año. Se buscan los cinco 9, un 99,999 % en el que el sistema debe estar disponible. Esto significa que en un año puede no estar activo durante cinco minutos, no necesariamente consecutivos.

99 %	3 días y 15 horas
99,9 %	8 horas y 15 minutos
99,99 %	53 minutos
99,999 %	5 minutos
99,9999 %	32 segundos

Cada 9 que se añade representa un incremento de costes muy considerable. Para conseguirlo, se utilizan componentes redundantes y aislados como, por ejemplo, fuentes de alimentación redundantes, controladores redundantes, buses dobles, dispositivos de E/S y copias dobles de los datos.

El objetivo es eliminar los periodos de falta de servicio al usuario. Estas paradas pueden ser dos tipos:

1) Paradas planificadas: aquellas debidas a actualizaciones de software o hardware.

2) Paradas no planificadas: son las causadas por un mal funcionamiento del hardware o por un desastre de tipo natural (como inundaciones o incendios) o de tipo no natural (sabotaje, error humano, etc.).

Hay organizaciones en las que no es imprescindible un servicio ininterrumpido del sistema informático. En ellas, es necesario un plan de recuperación de los datos para garantizar que el tiempo y el coste de la interrupción serán mínimos. En caso contrario, debemos disponer de una solución de alta disponibilidad, teniendo en cuenta las necesidades reales de la compañía.

Podemos conseguir una alta disponibilidad por medio de sistemas tolerantes a fallos o mediante técnicas de *clustering*. Los sistemas tolerantes a fallos son sistemas muy costosos porque hay que asegurar la redundancia de los componentes de su hardware y esto implica un alto coste. Los sistemas que usan técnicas de *clustering* son más económicos, pues no hay que utilizar hardware específico. Además, estos sistemas ofrecen balanceo de carga, por lo que obtenemos doble provecho con un coste menor.

La alta disponibilidad se puede aplicar a cualquier servicio. Los más comunes son:

- Servidor de dominio, DNS, DHCP, etc.
- Servidor web.
- Servidor de bases de datos.
- Servidor de ficheros.
- Servidor de correo.
- Servidor de aplicaciones.

Detallamos las características de los sistemas tolerantes a fallos y, por otro lado, de los clústeres de alta disponibilidad.

8.3.2. Sistemas tolerantes a fallos

Estas son algunas de las cuestiones que hay que tener en cuenta en un sistema tolerante a fallos:

- **Redundancia en el suministro eléctrico**. Un corte en el suministro eléctrico, aunque sea de pocos segundos, provocará que durante un tiempo nuestra máquina esté fuera de servicio. Por lo tanto, resulta vital conseguir que nunca falte el suministro eléctrico. Hay que valorar la instalación de sistemas de alimentación ininterrumpida (SAI), grupos electrógenos, fuen-

tes de alimentación redundantes en el mismo equipo (intercambiables en caliente) o, incluso, contratos con dos compañías eléctricas.

- **Discos duros redundantes o en grupos de paridad**.

- **Conexiones de red**. La red se ha convertido en un elemento indispensable para las aplicaciones actuales, por lo que hay que garantizar que la red estará disponible en todo momento. Para conseguir una red tolerante a fallos, hay que emplear dispositivos de red tolerantes a fallos.

Ved también

Podéis ver el subapartado «Sistemas de redundancia de datos».

8.3.3. Clústeres de alta disponibilidad

Los clústeres de alta disponibilidad y tolerancia a fallos están destinados a proporcionar **disponibilidad ininterrumpida** de recursos y servicios mediante la redundancia. Si un nodo del clúster falla, las aplicaciones y los servicios que se ejecutan pasarán a ejecutarse en uno de los nodos disponibles.

Algunas de las **ventajas** de este tipo de configuraciones son:

- **Escalabilidad**: puede aumentar la capacidad de cálculo del clúster si se añaden más procesadores o equipos.

- **Alta disponibilidad**: el clúster está diseñado para evitar un único punto de error. Las aplicaciones pueden distribuirse en más de un equipo, lo que aporta un grado de paralelismo y recuperación de errores, y proporciona más disponibilidad.

- **Balanceo de carga**: los nodos del clúster se pueden repartir las tareas del servicio para aumentar el rendimiento.

9. Aspectos legales

El administrador de servidores es una figura que tiene a su cargo, de una manera directa o indirecta, una gran cantidad de información de la organización. Toda esta información es sensible, por lo que, además de velar por que esté disponible y al alcance de las personas que deben usarla, hay que conocer los límites en su gestión y manipulación. ¿Dónde están las fronteras legales de todo esto? ¿Qué debe hacer si le piden que extraiga información de un cierto lugar? ¿O que la mire? ¿Y si le dicen que instale un programa que controle la actividad de los usuarios sobre cierta información? ¿Qué puede hacer y qué no un administrador de servidores con toda esta responsabilidad?

Actualmente la cuestión varía mucho y la legislación se mueve en un panorama muy cambiante.

Ved también

Intentaremos hacer un repaso de estas cuestiones en el módulo «Administración de la seguridad».

Somos conscientes de que en el momento en el que aparece el problema uno mismo debe buscar asesoramiento legal para resolverlo, pero consideramos que una de las cuestiones más importantes es saber reconocer, en materia legal, cuándo hay un problema real y cuándo no.

10. Tareas y responsabilidades

Por lo tanto, con todo lo que hemos visto, una posible relación de las tareas y responsabilidades del administrador de servidores podría ser la siguiente:

- Velar por el funcionamiento correcto de los servidores.
- Atender a la protección física y lógica de los servidores.
- Vigilar la copia de seguridad de los servidores.
- Procurar el buen funcionamiento de los subsistemas asociados a los servidores (colas de impresión, correo electrónico, etc.).
- Asegurar la disponibilidad de espacio para el trabajo de las aplicaciones y los usuarios.
- Velar por unos tiempos de respuesta correctos de los sistemas.
- Mantener el sistema operativo actualizado.
- Mantener las aplicaciones de las que es responsable actualizadas.
- Garantizar que la información del sistema esté protegida contra fallos, desastres naturales y eliminaciones accidentales.
- Proteger los datos y el contenido de los servidores.
- Asegurar la disponibilidad y la integridad de la información que contiene.
- Configurar los servidores corporativos según los requisitos de la organización, sean físicos, virtuales o, incluso, en la nube.

Resumen

Hemos visto cómo debe ser un servidor y sus características. Hemos analizado los diferentes tipos de servidores, ya sean físicos o virtuales, y hemos profundizado en las diferentes configuraciones que nos permiten obtener funciones y rendimientos mucho mejores que con un servidor aislado. Nos hemos dado cuenta de la importancia del almacenamiento y de cómo se puede configurar y ajustar a las necesidades de la organización, dado que es una de las cuestiones clave.

Hemos remarcado mucho los dispositivos o repositorios de copia de seguridad y las políticas posibles para llevarlas a cabo, dependiendo del tamaño y de las necesidades de la organización. Hemos hecho mención de la importancia de otros factores relacionados, como la corriente eléctrica, para asegurar el funcionamiento y la vida de los servidores.

Tampoco hemos descuidado la seguridad de nuestros servidores, pues contienen toda la información de la organización.

Finalmente, hemos comentado aspectos de los sistemas operativos y las responsabilidades del administrador de servidores.

Actividades

1. Buscad un software gratuito que os permita realizar copias de seguridad de vuestra estación de trabajo y que sea compatible con el *cloud object storage*. Daos de alta en uno de los servicios del COS (elegid el que queráis) y haced pruebas de copia y restauración. Comparad los tiempos, por ejemplo, con una copia en el disco local y sacad conclusiones sobre qué tipo de sistemas son los más indicados para poder almacenar las copias en la nube.

2. Investigad en la red algún software de virtualización de pruebas y cread una máquina virtual. Intentad instalar en esta máquina virtual un sistema Linux y verificad cómo se pueden aprovechar los recursos físicos del sistema, como por ejemplo la tarjeta de red, el disco, la memoria, etc. Una vez que tengáis el servidor virtual Linux, intentad crear un contenedor (buscad la documentación y los paquetes necesarios).

Ejercicios de autoevaluación

1. Suponiendo que en una organización tienen un servidor con un almacenamiento muy grande que dispone de una protección de discos RAID-6e distribuida (*distributed*), con tarjetas HBA de Fibra FC a 16 GBps, aparte de una gran cantidad de memoria RAM, ¿qué tipo de servidor creéis que puede ser? ¿Qué tipo de datos sería lógico que almacenara? Y, finalmente, ¿creéis que los discos son internos o externos?

2. ¿Qué sentencia describe mejor las diferencias entre la copia de seguridad diferencial frente a una copia de seguridad incremental?

a) La copia se realiza en menos tiempo, pero ocupa más espacio.
b) La copia se realiza en más tiempo a partir del segundo día de copia, pero ocupa menos espacio.
c) Guarda todos los objetos modificados desde la última copia diferencial y tarda menos que una copia incremental, pero ocupa más espacio desde el segundo día.
d) Guarda todos los objetos modificados desde la última copia total, pero ocupa más espacio desde el segundo día y tarda más tiempo en hacer la copia.

3. ¿Qué tipo de discos creéis que se utilizarán, en el futuro cercano, en los servidores productivos? ¿Y qué interfaz de datos creéis que se utilizará con estos tipos de discos?

4. Os proponen implementar un sistema de servidores de aplicaciones para servir una aplicación crítica dentro de vuestra organización que tiene una fluctuación de accesos muy elevada. La aplicación en cuestión se dedica a la venta de entradas en línea para espectáculos de gran formato. Por lo tanto, dependiendo del espectáculo y de las fechas de las ventas requiere o un incremento de recursos muy importante o una disminución de estos a prácticamente residuales.

¿Cuál podría ser una buena solución?

a) Virtualización de un servidor físico en servidores de aplicación.
b) Clúster de servidores físicos, todos ellos dedicados a dar el servicio de servidores de aplicaciones.
c) Encapsulamiento de la aplicación en un contenedor y gestión de los contenedores necesarios para el servicio con un orquestador como, por ejemplo, Kubernetes.
d) Las tres respuestas anteriores son correctas.
e) Las respuestas b) y c) son correctas.

5. Escribid un pequeño texto en el que se relacionen los diferentes elementos: NAS, estación de trabajo, acceso a nivel de bloque, LAN, SAN, datos de usuario y elemento de almacenamiento.

Solucionario

Ejercicios de autoevaluación

1. Según los datos que nos han suministrado, podemos deducir diferentes aspectos funcionales del servidor.

Una gran cantidad de almacenamiento y una gran cantidad de RAM nos dirigen a un servidor destinado a servir datos o a la virtualización.

La utilización de un sistema de seguridad de discos RAID-6e distribuido utiliza doble bit de paridad para cada grupo de RAID, distribuyendo la paridad entre los distintos discos que forman el grupo. Además, dispone de una zona *hot spare* (en espera) por si uno de los discos falla, lo que permite la reconstrucción del disco afectado de manera rápida, al tener la paridad distribuida. Esto significa que los datos almacenados son muy importantes y hay que protegerlos.

En cuanto a la respuesta de las preguntas, estamos hablando de un servidor de almacenamiento, posiblemente de un servidor de bases de datos con datos muy importantes para la organización. También se podría tratar de un servidor dedicado a la virtualización de servicios productivos, a pesar de que no se nos ha especificado información sobre la capacidad de procesamiento (imprescindible para servidores especializados en virtualización).

Finalmente, en cuanto a la pregunta de si los discos son locales o externos al sistema, viendo que el sistema dispone de tarjetas de acceso a la SAN, suponemos que los discos son externos en cabina. También nos lo indica la gran cantidad de almacenamiento, que hoy en día siempre se suele disponer en este tipo de cabinas.

2. d) La sentencia d) nos detalla la mayor ventaja que tenemos: guardamos todos los objetos modificados desde la última copia total y a la hora de restaurar el sistema solo deberemos restaurar la copia total y la última diferencial. No obstante, ocupamos más espacio a partir del segundo día e incrementamos el tiempo de copia.

3. Tal y como evoluciona el mercado, en un futuro cercano, todos los servidores productivos dispondrán de discos SSD, ya sean en local o en cabinas por medio de la SAN. Los discos HDD quedarán restringidos a sistemas de almacenamiento de datos históricos, de archivado, etc.

En cuanto a la interfaz de acceso a los datos, claramente se está imponiendo el NVMe para discos SSD, y más teniendo en cuenta que no solo está restringido al ámbito de los canales PCie, sino que ya se puede encapsular en FC (para accesos a discos de la red SAN) y también se puede encapsular en tramas Ethernet, siempre dentro de accesos de bloque FC.

4. c) Teniendo en cuenta que se trata de una aplicación muy definida y que el entorno debe ser muy dinámico y escalable de manera horizontal, actualmente la mejor solución pasaría por un sistema de contenedores con el despliegue de la aplicación y una gestión orquestada.

5. Un usuario, que trabaja con su estación de trabajo, debería tener sus datos más importantes remotamente en un servidor de almacenamiento como, por ejemplo, un NAS. Para acceder a este servidor, usará la red LAN de comunicaciones.

A pesar de que el usuario acceda al NAS a buscar sus datos, puede que estos realmente estén en un elemento de almacenamiento al que accede el servidor NAS a nivel de bloque mediante una red SAN.

Glosario

alta disponibilidad *f* Instalación que intenta conseguir la máxima disponibilidad de un sistema (24 × 7).

APFS (Apple *file system*) *m* Sistema de ficheros de computadoras Apple.

BaaS (*software as a service*) *m* *Backup* como servicio ofrecido en la nube.

backup Véase **copia de seguridad**.

bare metal Servidor físico que ejecuta instrucciones directamente en el hardware, es un servidor *single tenant*.

bitcoin *m* Criptomoneda basada en cadenas de bloques.

blade Hoja o lámina. Se aplica a servidores en una tarjeta o lámina.

blade center Cabina específica para gestionar *blades*.

blockchain Sistema de seguridad y confianza informática formado por una cadena de bloques descentralizada.

CMP (*cloud management platform*) *f* Sistema de gestión de plataformas en la nube.

contenedor *m* Máquina virtual independiente, ligera y portable para aplicaciones.

copia de seguridad *f* Método para duplicar la información de la organización en otro soporte que sea más seguro.

cortafuegos *m* Sistema de seguridad para control de accesos a la red.
en.: *firewall*

COS (*cloud object storage*) *m* Espacio de almacenamiento de objetos en la nube.

CPD *m* Centro de procesamiento de datos. Ubicación de los servidores de una organización.

CPU *f* Véase **unidad de control de proceso**.

deduplicación *f* Sistema que permite ahorrar el almacenamiento de los datos redundantes.

descarga completa *f* Copia de seguridad completa de una partición de disco.
en.: *full dump*

directorio *m* Espacio lógico dentro de un disco, en el que se guardan ficheros y directorios.

disco duro *m* Dispositivo físico que sirve para guardar información.

DL (*deep learning*) *m* Véase **ML**. Parte de ML que se basa en el uso de redes neuronales.

docker Véase **contenedores**.

FC (*fibre channel*) *m* Canal de fibra empleado para la transferencia de almacenamiento en una red SAN.

full dump Véase **descarga completa**.

FPGA (*field-programmable gate array*) *f* Tarjeta programable que se puede emplear para acelerar procesos en sistemas integrados como *core* independiente.

grid Computación en malla que permite interconectar ordenadores dispersos por la red para aprovechar su potencia de cálculo.

GPU *f* Unidad de computación gráfica. Actualmente no solo se utiliza para acelerar los cálculos matemáticos de las tarjetas gráficas, sino que también se emplea para acelerar los cálculos en *machine learning* y *deep learning*.

HA (*high availability*) *f* Alta disponibilidad.

HBA (*host bus adapter*) *f* Tarjeta de entrada/salida que permite enlazar un servidor a una red SAN.

HDD (*hard disk drive*) *m* Disco duro (también llamado disco giratorio).

HPC (*high performance cluster*) *m* Clúster de alta eficiencia. Permite ejecutar un gran número de tareas.

HT (*high throughput*) *m* Clúster de alto rendimiento. Permite ejecutar las tareas asignadas en el menor tiempo posible.

hipervisor *m* Monitor de máquinas virtuales que permite la creación y la gestión de máquinas virtuales.

IA (inteligencia artificial) *f* Capacidad de los sistemas de aprender y aplicar soluciones basadas en algoritmos y modelos estadísticos.

IaaS (*infrastructure as a service*) *m* Hardware como servicio ofrecido en la nube.

impresora remota *f* Impresora que está conectada directamente a la red informática en lugar de estarlo a un ordenador. El servidor la gestiona por medio de la red, no localmente, porque no hay cable.

IPP (*internet printer protocol*) *m* Protocolo de acceso a impresoras por medio de internet.

J2EE (Java *to enterprise edition*) *f* Estándar Java orientado a arquitecturas de empresa.

JVM (Java *virtual machine*) *f* Máquina virtual de Java. Interpreta los comandos Java en un sistema.

LTO (*linear tape open*) *f* Tecnología de cinta magnética empleada para almacenamiento de *backup*.

memoria de acceso aleatorio *f* Memoria volátil que usan todos los ordenadores.

middleware Software que actúa entre el sistema operativo y las aplicaciones con el fin de proveer una interfaz única de acceso al sistema.

ML (*machine learning*) *m* Estudio que realiza un sistema informático, basado en algoritmos y modelos estadísticos que se utilizarán para una tarea específica sin unas instrucciones predefinidas.

motherboard Véase **placa base**.

NAS *m* Servidor de ficheros. Acceso a nivel de fichero.

NFS (*network file system*) *m* Sistema de ficheros distribuido en la red.

NLSAS (*near-line* SAS) *m* Discos SATA con interfaz de comunicación SAS.

NoSQL (*not only* SQL) *f* Bases de datos no solo SQL o relacionales, normalmente aceptan parejas clave, valor.

NTFS *m* Sistema de ficheros creado por Microsoft para los sistemas Windows.

NVMe (*non-volatile memory express*) *m* Protocolo de interfaz con discos de estado sólido.

NVMe-oF (NVMe *over fabric*) *m* Encapsulamiento del protocolo NVME sobre *fabric* en redes SAN.

partición *f* División del espacio interno del disco duro.

PCIe o PCI Express (*peripheral component interconnect express*) *f* Expansión de bus estándar para conexión de periféricos de alta velocidad.

PaaS (*platform as a service*) *f* Plataforma dedicada como servicio ofrecido en la nube.

placa base *f* Componente del ordenador que tiene los buses de sistema y el árbitro del bus. Controla toda la comunicación entre los diferentes componentes. Contiene la BIOS, el

espacio para montar la CPU, la RAM y las ranuras de expansión (*slots*) para la placa gráfica, la placa de red, etc.

plan de contingencia *m* Estudio del impacto de posibles contingencias y su tratamiento para recuperar la normalidad funcional.

placa de red *f* Componente del ordenador que permite la comunicación entre la red y los buses internos.

pod Agrupación de uno o más contenedores con o sin almacenamiento.

toma de tierra *f* Conductor que se pone en contacto directo con el suelo.
(Diccionario enciclopédico, Enciclopedia Catalana, edición 1984)

RAID (*redundant array of inexpensive disks*) *m* Se trata de distribuir la información entre varias unidades de disco con posibilidad de gestión de la paridad de los datos.

RAM *f* Véase **memoria de acceso aleatorio**.

ReFS (*resilient file system*) *m* Sistema de ficheros resiliente creado por Microsoft.

SaaS (*software as a service*) *m* Software como servicio ofrecido en la nube.

SAI *m* Véase **sistema de alimentación ininterrumpida**.

SAN (*store area network*) *m* Red especializada en comunicación entre servidores y elementos de almacenamiento.

SAS (*serial attached* SCSI) *m* Protocolo de acceso en serie a discos SCSI. Véase **SCSI**.

SATA (*serial-ATA*) *m* Protocolo de acceso en serie a discos ATA. Véase **IDE o P-ATA**.

SCSI (*small computer system interface*) *f* Tipo de controlador de dispositivos de altas prestaciones. Se pueden conectar a muchos dispositivos diferentes y las distintas revisiones permiten conectar hasta dieciséis dispositivos en el mismo controlador.

servidor institucional *m* Ordenador que consta de aplicaciones con tecnología cliente/servidor y que sirve peticiones por la red, bajo demanda de los clientes (estaciones de trabajo).

servidor virtual *m* Servidor que se ejecuta como una máquina virtual sobre recursos de un servidor físico gestionado por un hipervisor.

SGBD (sistema gestor de BD) *m* Software que gestiona una BD.

SIEM (*security information and event management*) *m* Sistema de información de seguridad y gestión de eventos o amenazas.

sistema de alimentación ininterrumpida *m* Componente que evita la caída de los servidores por falta de corriente eléctrica, ya que se encarga de suministrarla cuando no la hay.
sigla: SAI

sistema de ficheros *m* Configuración consistente en una partición para poner en ella los ficheros.

SSD (*solid state disk*) *m* Disco de estado sólido, basado en persistencia de datos NAND Flash.

proxy *m* Sistema de control intermedio entre los clientes y las fuentes solicitadas. Gestiona permisos de acceso.

velocidad de transferencia *f* Velocidad en Mb/segundo a la que viaja la información entre dos dispositivos o componentes.

VMFS (*virtual machine file system*) *m* Sistema de ficheros para máquinas virtuales (creado por VMWare).

VSAN (*virtual* SAN) *f* Red SAN virtual creada por VMWare.

VTL (*virtual tape library*) *m* Librería de cinta virtual. Simula una librería de cinta física en el disco.

Bibliografía

Blokdyk, G. (2019). *Server virtualization A Complete Guide*. Estados Unidos: 5starcooks.

Blokdyk, G. (2019). *Data Backup A Complete Guide*. Estados Unidos: 5starcooks.

Bolton, J. (2019). *What is Cloud Computing? All about Cloud Technology*. Publicación independiente.

DiSario, D. P. (2015). *Backup Fanatic: How to Ensure Business Continuity by Delivering Continuous Protection, Secured Storage, Data Compliance, and Instant Data Recovery*. Estados Unidos: CreateSpace.

Liebel, O. (2019). *Scalable Container Infrastructures with Docker, Kubernetes and OpenShift: The Compendium on Container Clusters for Enterprise Administrators and DevOps Teams*. Estados Unidos: Kindle Direct Publishing.

Mohammadabadi, A. A. (2017). *Comparing FPGA and GPU Performance for an Image Watermarking Algorithm*. Lambert Academic Publishing.

Moore, J. D. (2019). *Kubernetes: The Complet Guide To Master Kubernetes*. Publicación independiente.

Pethuru Raj, C.; Surianarayanan, C. (2019). *Essentials of Cloud Computing: A Holistic Perspective*. Estados Unidos: Springer.

Portnoy, M. (2016). *Virtualization Essentials*. Estados Unidos: Sybex.

Preston, W. C. (2002). *Using Sans and Nas*. Estados Unidos: O'Reilly Media.

Scholl, B. (2019). *Cloud Native: Using Containers, Functions, and Data to Build Next-Generation Applications*. Reino Unido: O'Reilly UK Ltd.

Takefuji, Y. (2017). *GPU parallel computing for machine learning in Python: how to build a parallel computer*. Publicación independiente.

Tate, J.; Kumaravel, S.; Miklas, L.; Hugo Ibarra, H.; Beck, P. (2003). *Introduction to Storage Area Networks*. Estados Unidos: IBM Redbooks.

Young, N. (2019). *Cloud Computing: A to Z of Cloud Computing*. Publicación independiente.

Webgrafía

NVM Express

Microsoft: «Resilient File System (ReFS) overview»

Microsoft: «NTFS overview»

Apple: «File system formats available in Disk Utility on Mac»

Intel: «Revolutionizing Memory and Storage»

Nutanix: «What is hyper-converged infrastructure?»

Flash Memory Summit: «Persistent Memory»

Docker: «Docker overview»

Knóldus: «Tale of a Container's File System»

Docker: «What is a Container?»

IBM: «Servidores dedicados: Capacidad de Intel, suministrada en minutos»

OpenStack: «Software»

OpenStack: «OpenStack Alternatives & Competitors»

Gartner: «Cloud Management Platforms»