XUNTA DE GALICIA

ESCOLA GALEGA DE ADMINISTRACIÓN PÚBLICA

46. ADMINISTRACIÓN Y GESTIÓN DE SISTEMAS Y ALMACENAMIENTO. VIRTUALIZACIÓN DE SERVIDORES. VIRTUALIZACIÓN DEL PUESTO CLIENTE. COMPUTACIÓN BASADA EN SERVIDOR (SBC). GRID COMPUTING. CLOUD COMPUTING. GREEN IT Y EFICIENCIA ENERGÉTICA. REDES SAN Y ELEMENTOS DE UM SAN. VIRTUALIZACIÓN DEL ALMACENAMIENTO. GESTIÓN DEL CICLO DE VIDA DE LA INFORMACIÓN (ILM). SISTEMAS DE BACKUP: HARDWARE Y SOFTWARE DE BACKUP. ESTRATEGIAS DE BACKUP A DISCO. REPLICACIÓN LOCAL Y REMOTA, ESTRATEGIAS DE RECUPERACIÓN.

Fecha de actualización 26 de julio de 2011

Tema 46: Administración y gestión de sistemas y almacenamiento. Virtualización de servidores. Virtualización del puesto cliente. Computación basada en servidor (SBC). Grid Computing. Cloud computing. Green IT y eficiencia energética. Redes SAN y elementos de un SAN. Virtualización del almacenamiento. Gestión del ciclo de vida de la información (ILM). Sistemas de backup: hardware y software de backup. Estrategias de backup a disco. Replicación local y remota, estrategias de recuperación.

---

461 ADMINISTRACIÓN Y GESTIÓN DE SISTEMAS Y ALMACENAMIENTO

461.1 Administrador de red

Un administrador de red como su nombre indica "administra" una red, es decir, se encarga de:

- Instalación y configuración de la red.
- Hardware de red, conexiones físicas, hubs, switches, routers, servidores y clientes.
- Software de red, como los sistemas operativos de red, servidores de correo electrónico, software para la realización de copias de seguridad, base de datos servidores y software de aplicación.

Lo más importante, el administrador tiene cuidado de los usuarios de la red, respondiendo a sus preguntas, escuchar sus problemas, y resolver sus problemas.

Cuando las tareas de administración se realizan en una red grande y compleja, este conjunto de tareas han de abarcarse de manera dedicada, es decir, poseer una o varias personas realizando únicamente las tareas de administración de la red. Esto debe ser así debido a que las redes tiende a ser volátiles en el sentido de:

- Los usuarios de la red cambian constantemente.
- Los equipos fallan.
- Se producen conflictos entre las distintas aplicaciones.
- En general, una red compleja sufre continuos estados de crisis.

Por el contrario, las redes de menor tamaño y por ello menos complejas son generalmente mucho más estables. Suele ser habitual que una vez puesta en funcionamiento una red sencilla no tenga que sufrir continuas y complejas tareas de administración ya sean de hardware o software.

En este tipo de redes pequeñas los problemas también aparecen, pero como intervienen un reducido número de equipos es normal que sean sencillos, pocos y distantes entre sí.

Independientemente del tamaño de una red, un administrador debe cubrir las siguientes tareas que son comunes a cualquier tipo de red:

- Involucrarse y formar parte en la toma de decisiones para la adquisición de nuevo equipamiento, servidores, equipos, impresoras, etc.
- Establecer las acciones necesarias para el correcto funcionamiento cada vez que se añada un nuevo equipo, es decir, un administrador de red cuando se integra un nuevo elemento a la red se encarga de introducir cambios en la configuración del cableado, de asignar un nombre de red al nuevo equipo, integrar a un nuevo usuario en el sistema de seguridad garantizando además sus privilegios.
- Estar al corriente de las actualizaciones de software que publiquen los proveedores y considerar si sus nuevas características son suficientes para justificar una posible actualización.

En la mayoría de los casos, la parte más difícil de un proceso de actualización de software es la determinación del camino a seguir, como llevar a cabo la actualización de toda la red afectando lo menos posible al funcionamiento de los usuarios. Esto suele ser aún más crucial si el software a actualizar es el sistema operativo de red, puesto que cualquier cambio en el puede afectar a toda la red.

Dentro de este procesos de actualización también intervienen afectando en menor medida a la estabilidad del sistema los parches y Service Packs que publican los proveedores para actualizar sus soluciones y que solventan problemas menores.

- Realizar tareas rutinarias como la realización de copias de seguridad de los servidores, la administración del historial de datos o la liberación de espacio en los discos duros. Gran parte de la tareas de administración de una red consisten en asegurarse de que todo funcione correctamente, buscando y corrigiendo los problemas que puedan tener los usuarios.

- Recopilar, organizar y controlar el inventariado de toda la red, para poder solventar en el menor tiempo posible cualquier imprevisto.

461.2 Administración y gestión de redes

El concepto de administración tiene asociado muchos significados. Desde un punto de vista informal, la gestión de redes se refiere a las actividades relacionadas con el funcionamiento de una red, junto con la tecnología necesaria para apoyar estas actividades. Otro aspecto de importancia en la gestión de una red es la monitorización de la misma, es decir, entender en todo momento que es lo que está sucediendo en la red.

Desde un enfoque software, la gestión de redes hace referencia al conjunto de actividades, métodos, procedimientos y herramientas que intervienen en las operaciones de administración, mantenimiento y aprovisionamiento de los sistemas existentes dentro de la red.

Supone además garantizar toda la oferta operativa de servicios manteniendo la red en marcha y funcionando sin problemas. Para conseguir esto se hace imprescindible la utilización de herramientas para la monitorización de la red, que ofrezcan la detección de problemas tan pronto como sea posible, incluso antes de que algún usuario se vea afectado.

La administración abarca a su vez las tareas de seguimiento de los recursos en la red y de cómo estos se asignan, haciendo uso de todos los procesos o acciones de limpieza del la red que sean necesarias para

mantener todo bajo el control del administrador o administradores.

El proceso de *mantenimiento*, que se ocupa de realizar las operaciones de reparación y mejora, ha de llevar a cabo tareas como el reemplazo de una tarjeta de red, actualización del sistema operativo de un router, añadir un nuevo switch al entramado de red. El mantenimiento también implica medidas para la corrección y prevención, como por ejemplo, el ajuste de los parámetros necesarios de un dispositivo en función de las necesidades que se soliciten o intervenir cuando sea necesario para mejorar el rendimiento de la red en momentos puntuales.

Otro aspecto de la administración de una red es el *aprovisionamiento*, tarea que concierne a la configuración y adaptación de los recursos de red para dar soporte a los servicios ofertados. Un ejemplo de aprovisionamiento es el hecho de añadir las configuraciones necesarias en los sistemas para proporcionar el servicio de voz a un nuevo usuario.

461.2.1 Tareas de la gestión de red

Las tareas de gestión de una red se pueden caracterizar de la siguiente manera:

- QoS y Gestión del Rendimiento: un administrador de red debe supervisar y analizar periódicamente los routers, hosts y el funcionamiento de los enlaces y luego en función de los resultados obtenidos realizar una redirección del flujo de datos para evitar la sobrecarga de ciertos puntos de la red. Para realizar esta tarea de seguimiento de la red, existen herramientas que detectan rápidamente los cambios que se producen en el tráfico de una red.

- Gestión de fallos por la red: cualquier fallo en la red, enlaces, nodos, routers, fallos de hardware o software, debe ser detectado, localizado y respondido por la propia red, es decir, la propia red debe

poseer mecanismos para intentar solventar por sí sola el mayor número de contingencias que se puedan producir.

- Gestión de la configuración: esta tares implica el seguimiento de todos los dispositivos bajo gestión y la confirmación de que todos los dispositivos están conectados y funcionan correctamente. Si se produce un cambio inesperado en las tablas de enrutamiento, el administrador ha de descubrir el problema de configuración y solucionarlo lo antes posible para que ningún servicio ni usuario se vea afectado.

- Gestión de la seguridad: el administrador de red es el responsable de la seguridad de la red. Para poder manejar esta tarea se utilizan principalmente los firewalls, puesto que un firewall puede monitorizar y controlar los puntos de acceso a la red informando sobre cualquier intento de intrusión.

- Gestión de facturación y contabilidad: el administrador especifica a los usuarios de la red los accesos o restricciones sobre los recursos y se encarga de la facturación y de los cargos a los usuarios por el uso de los mismos.

461.2.2 Elementos de la gestión de red

La gestión de red está compuesta por tres componentes principales:

- ***Centro de gestión:*** compuesto por el administrador de red y sus oficinas o centros de trabajo. Normalmente el centro de gestión está compuesto por un grupo humano importante.

- ***Dispositivos a gestionar:*** conformado por el equipamiento de la red, incluido su software, que es controlado mediante el centro de gestión. Cualquier hub, bridge, router, servidor, impresora o módem es considerado un dispositivo que ha de ser gestionado.

- ***Protocolo de gestión de la red:*** es el conjunto de políticas que adopta el centro de gestión para controlar y manejar todos los dispositivos que conforman la red. El protocolo de gestión de red permite al centro de gestión conocer el estado de los dispositivos.

461.2.2.1 Estructura de Gestión de la Información (SMI, Structure of Management Information):

Define las reglas para nombrar los objetos y para codificarlos en un centro de gestión de una red, es decir, es un lenguaje mediante el cual se definen las instancias dentro de un centro de gestión de red.

El lenguaje SMI también ofrece construcciones del lenguaje de mayor nivel que, habitualmente, especifican los tipo de datos, el estado y la semántica de los objetos que contienen la información necesaria para realizar las tareas de gestión. Por ejemplo, la cláusula STATUS especifica si la definición del objeto es actual o está obsoleta.

Trabaja bajo el protocolo SNMP (Simple Network Management Protocol) definiendo los conjuntos de objetos dentro la gestión de información base (MIB).

461.2.2.2 La Gestión de la Información Base (MIB, Management Information Base)

Es un medio de almacenamiento de información que contiene los objetos que muestran el estado actual de una red. Debido a que los objetos tienen asociado información que se almacena en el MIB, este forma colecciones de objeto, en las que incluye las relaciones entre ellos, en el centro de gestión.

Los objetos se organizan de una forma jerárquica y se identifican por la notación abstracta ASN.1, lenguaje de definición de objetos. La jerarquía, conocida como ASN.1, es un árbol de identificadores de objeto en el cual cada rama tiene un nombre y un número, permitiendo así a la gestión de red identificar objetos por una secuencia de nombres o números desde la raíz al objeto.

461.2.2.3 Protocolo SNMP (Simple Network Management Protocol)

El Simple Network Management Protocol (SNMP) está diseñado para monitorear el rendimiento de los protocolos de red y de los dispositivos. Las unidades de datos del protocolo SNMP (PDUs) pueden ser transportadas en un datagrama UDP, por lo que su entrega en destino no está garantizada. Los dispositivos que se administran como los routers o hosts, son objetos y cada uno tiene una definición formal y MIB adapta una base de datos de información que describe sus características. Con este protocolo un gestor de red puede encontrar donde se localizan los problemas.

Se ejecuta sobre UDP y utiliza una configuración cliente-servidor. Sus comandos definen como realizar las consultas sobre la información de un servidor o como enviar esta hacia un cliente o hacia otro servidor.

La tarea principal del protocolo SNMP es la de transportar información entre los centro de gestión y los agentes que se ejecutan en representación de los centros de gestión. Para cada objeto MIB que se gestiona se utiliza una petición SNMP para obtener su valor o para modificarla. Si un agente recibe un mensaje no solicitado o si una interfaz o dispositivo deja de funcionar, entonces el protocolo puede informar al centro de gestión del fallo que se está produciendo.

La segunda versión de este protocolo, SNMPv2, corre por encima de varios protocolos y tiene más opciones de mensajería, lo que resulta en una gestión más eficaz de la red. Tiene siete unidades de PDU, o mensajes:

1. ***GetRequest.*** Se utiliza para obtener un valor de objeto MIB.

2. ***GetNextRequest***. Se utiliza para obtener el siguiente valor de un objeto MIB.

3. ***GetBulkRequest***. Recibe múltiples valores, lo que equivale a GetRequests múltiples, pero sin necesidad de utilizar múltiples peticiones.

4. ***InformRequest***. Es un mensaje de director a director de comunicación que se envían entre sí dos centros de gestión a distancia el uno del otro.

5. ***SetRequest***. Es utilizado por un centro de gestión para inicializar el valor de un objeto MIB.

6. ***Response***. Es un mensaje de respuesta a una petición de tipo PDU.

7. ***Trap***. Notifica a un centro de gestión de un evento inesperado.

Hay dos tipos de representación de PDUs, Get o Set y Trap.

- El formato de PDU de Get o Set es el siguiente:

  - *PDU type,* indica uno de los siete tipos de PDU.

  - *Request ID,* es un ID que se utiliza para verificar la respuesta de una solicitud. Por lo tanto un centro de gestión puede detectar peticiones perdidas o duplicadas.

  - *Error status,* sólo es usado por PDUs *Response* para indicar tipos de errores reportados por un agente.

- *Error index*, es un parámetro que indica a un administrador el nombre del objeto que ha causado el error.

Si las solicitudes o respuestas se pierden, el protocolo no realiza un reenvío. Los campos Error status and Error index son todo ceros excepto para las PDUs *GetBulkRequest*

- El formato de PDU de Trap es:

- *Enterprise*, para usar en múltiples redes.

- *Timestamp*, para realizar las mediciones de tiempo.

- *Agentadress*, para indicar que la dirección del agente gestor está incluida en la cabecera PDU.

462 VIRTUALIZACIÓN DE SERVIDORES

Podemos definir virtualización como la técnica que consiste básicamente en agrupar diferentes aplicaciones y servicios de sistemas heterogéneos dentro de un mismo hardware, de forma que los usuarios y el propio sistema los vean como máquinas independientes dedicadas. Para ello, el sistema operativo virtualizado debe ver el hardware de la máquina real como un conjunto normalizado de recursos independientemente de los componentes reales que lo formen.

De esta forma, para virtualizar un sistema de servidores, los administradores deben, básicamente, optimizar los recursos disponibles, incluyendo el número y la identidad de los servidores físicos individuales, procesadores, y sistemas operativos, con el objetivo de producir una mejora tanto en la gestión como en el manejo de sistemas informáticos

complejos. El administrador del sistema virtual utilizará un software para la división del servidor físico en entornos virtuales aislados. Estos entornos son lo que se conoce técnicamente como servidores privados virtuales, pero también se pueden encontrar referencias como particiones, instancias, contenedores o emulaciones de sistemas.

En concreto, podemos decir que un servidor privado virtual es un término de marketing utilizado por los servicios de hosting para referirse a una máquina virtual para el uso exclusivo de un cliente individual del servicio. El término se utiliza para enfatizar que la máquina virtual, a pesar de ejecutarse en el mismo equipo físico que las máquinas virtuales de otros clientes, es funcionalmente equivalente a un equipo físico independiente, está dedicado a las necesidades individuales del cliente y puede ser configurado para ejecutarse como un servidor de internet (es decir, para ejecutar software de servidor). El término VDS o Virtual Dedicated Server (Servidor Virtual Dedicado) para el mismo concepto.

Cada servidor virtual puede ejecutar su propio sistema operativo y ser reiniciado de modo independiente.

462.1 Funcionamiento

El servidor físico realiza una abstracción de los recursos que se denomina Hypervisor o VMM (Virtual Machine Monitor), elemento software que se instala en la máquina donde se va a llevar a cabo la virtualización y sobre la que se configuran las máquinas virtuales que es donde van a residir las aplicaciones. Es el encargado de gestionar los recursos de los sistemas operativos “alojados” (guest) o máquinas virtuales.

Desde un punto de vista lógico, el usuario percibe que son máquinas independientes y aisladas entre sí, pero desde una perspectiva física, todas las máquinas virtuales residen en un único servidor. A estas máquinas virtuales se les asigna un porcentaje de los recursos del servidor físico, que serán los únicos que el cliente conozca.

Se pueden encontrar tres modelos de virtualización: el modelo de máquina virtual o virtualización completa, el modelo paravirtual o virtualización parcial; y la virtualización a nivel de sistema operativo.

462.1.1 Virtualización completa

El modelo de máquina virtual está basado en la arquitectura cliente/servidor, donde cada cliente funciona como una imagen virtual de la capa hardware. Este modelo permite que el sistema operativo cliente funcione sin modificaciones. Además permite al administrador crear diferentes sistemas cliente con sistemas operativos independientes entre sí. La ventaja principal de este modelo radica en el desconocimiento por parte de los sistemas huésped del sistema hardware real sobre el que está instalado. Sin embargo, realmente todos los sistemas virtuales hacen uso de recursos hardware físicos. Estos recursos son administrados por un el hypervisor que coordina las instrucciones CPU, convirtiendo las peticiones del sistema invitado en las solicitudes de recursos apropiados en el host, lo que implica una sobrecarga considerable. Casi todos los sistemas pueden ser virtualizados utilizando este método, ya que no requiere ninguna modificación del sistema operativo. A pesar de esto, es necesaria una virtualización de la CPU como apoyo para la mayoría de los hypervisores que llevan a cabo la virtualización completa.

Ejemplos típicos de sistemas de servidores virtuales son VMware Workstation, VMware Server, VirtualBox, Parallels Desktop, Virtual Iron, Adeos, Mac-on-Linux, Win4BSD, Win4Lin Pro, y z/VM, openvz, Oracle VM, XenServer, Microsoft Virtual, PC 2007 y Hyper-V.

462.1.2 Paravirtualización

El modelo de máquina paravirtual (PVM) o virtualización parcial se basa, como el modelo anterior, en la arquitectura cliente/servidor, incluyendo también la necesidad de contar con un sistema monitor. Sin embargo, en este caso, el VMM accede y modifica el código del sistema operativo del

sistema huésped. Esta modificación se conoce como porting. El porting sirve de soporte al VMM para que pueda realizar llamadas al sistema directamente. Al igual que las máquinas virtuales, los sistemas paravirtuales son capaces de soportar diferentes sistemas operativos instalados en el hardware real. Esta técnica se utiliza con intención de reducir la porción de tiempo de ejecución empleada por el huésped empleado en realizar las operaciones que son mucho más difíciles de ejecutar en un entorno virtual en comparación con un entorno no virtualizado. Así se permite que el invitado(s) y el huésped soliciten y reconozcan estas tareas, que de otro modo serían ejecutados en el dominio virtual (donde el rendimiento de ejecución es peor). Una plataforma paravirtualizada exitosamente puede permitir que el VMM sea menos complejo (por la reubicación de la ejecución de las tareas críticas del dominio virtual en el dominio del servidor), y/o reducir la degradación del rendimiento global de la máquina virtual durante la ejecución de invitado.

UML, XEN, Xen, Virtuozzo , Vserver y OpenVZ (que es el código abierto y la versión de desarrollo de Parallels Virtuozzo Containers) son modelos de máquinas paravirtuales.

462.1.3 Virtualización por S.O.

La virtualización a nivel de sistema operativo se diferencia de las anteriores en que, en este caso, no existe un sistema cliente/servidor propiamente dicho. En este modelo el sistema principal exporta la funcionalidad del sistema operativo desde su propio núcleo. Por esta razón, los sistemas virtuales usan el mismo sistema operativo que el nativo (aunque en la mayoría de los casos pueden instalar distintas distribuciones). Esta arquitectura elimina las llamadas del sistema entre capas, lo que favorece una reducción importante en el uso de CPU. Además, al compartir los ficheros binarios y librerías comunes del sistema en la misma máquina, la posibilidad de escalado es mucho mayor, permitiendo que un mismo servidor virtual sea capaz de dar servicio a un gran número de clientes al

mismo tiempo.

La Virtualización de SO mejora el rendimiento, gestión y eficiencia. Podemos entenderlo como un sistema en capas. En la base reside un sistema operativo huésped estándar. A continuación encontramos la capa de virtualización, con un sistema de archivos propietario y una capa de abstracción de servicio de kernel que garantiza el aislamiento y seguridad de los recursos entre distintos contenedores. La capa de virtualización hace que cada uno de los contenedores aparezca como servidor autónomo. Finalmente, el contenedor aloja la aplicación o carga de trabajo.

Ejemplos de sistemas que usan virtualización a nivel de sistema operativo son Virtuozzo y Solaris.

463 VIRTUALIZACIÓN DEL PUESTO CLIENTE

Esta técnica consiste en la separación del entorno de usuario de un ordenador personal de la máquina física con el modelo cliente-servidor. El modelo que sigue un servidor para implementar dicha característica se denomina VDI (Virtual Desktop Infrastructure, Infraestructura de Escritorio Virtual), también llamada Interfaz de Escritorio Virtual.

La mayoría de implementaciones comerciales de esta tecnología usan un servidor central remoto para llevar a cabo la “virtualización” del escritorio del cliente, en lugar de usar el almacenamiento local del cliente remoto. Esto implica que todas las aplicaciones, procesos, configuraciones y datos del cliente están almacenadas en el servidor y se ejecutan de forma centralizada.

El sistema cliente puede utilizar una arquitectura de hardware completamente diferente de la utilizada por el entorno de escritorio proyectado, y también puede estar basada en un sistema operativo

completamente diferente.

El modelo de virtualización del puesto cliente permite el uso de máquinas virtuales para que múltiples suscriptores de red puedan mantener escritorios individuales en un solo ordenador, el servidor central. Este servidor central puede operar en una residencia, negocio o centro de datos. Los usuarios pueden estar geográficamente dispersos, pero todos están conectados a la máquina central por una red de área local, una red de área amplia, o Internet.

463.1 Modos de Operación de VDI

Básicamente existen cuatro modelos de operación VDI:

- Alojado (como servicio). Suelen contratarse a proveedores comerciales y normalmente proporciona una configuración del sistema operativo del puesto cliente administrado. Los principales suministradores son CITRIX, VMware y Microsoft.

- Centralizado. En este caso todas las instancias VDI están alojadas en uno o más servidores centralizados, los datos están en sistemas de almacenamiento conectados a estos. Este modelo a su vez puede distinguir dos tipos:

  o VDI estático o persistente. Existe una única imagen de escritorio asignado por cliente y estos deben ser gestionados y mantenidos.

  o VDI dinámico o no persistente. Existe una imagen maestra común para todos los clientes que se clona y personaliza en el momento de la petición con los datos y aplicaciones particulares de cada cliente.

- Remoto (o sin ataduras). Tiene como base el concepto de VDI centralizado pero permite trabajar sin la conexión a un servidor central o a Internet. Se copia una imagen al sistema local y se ejecuta sin

necesidad de más conexión. Las imágenes tienen un cierto periodo de vida y se actualizan periódicamente. Esta imagen se ejecuta en el sistema local que necesita un sistema operativo y un hipervisor (que ejecuta la instancia VDI). Esto implica que el dispositivo cliente tenga mayores necesidades de memoria, espacio en disco, CPU... La ventaja es la menor dependencia de conexión.

Los modelos alojado y centralizado necesitan de una red que conecte con el servidor donde se ejecuta la instancia VDI. El concepto base de este modelo es similar al de clientes ligeros debido a que el cliente sólo tiene que mostrar el escritorio virtual.

En el caso del modelo remoto, se permite a los usuarios copiar la instancia VDI en el sistema y luego se ejecutará el escritorio virtual sin necesidad de ningún tipo de conexión.

464 COMPUTACIÓN BASADA EN SERVIDOR

También conocida como SBC del inglés Server Based Computing, consiste en la separación del procesamiento de ciertas tareas como la gestión de datos que será realizado en un servidor central y otras tareas de procesamiento, como la presentación de aplicaciones de usuario e impresión de datos en el cliente. Lo único transmitido entre servidor y cliente son las pantallas de información. Esta arquitectura puede dar solución a los principales problemas que aparecen cuando se ejecutan aplicaciones en los clientes. Además simplifica procesos como pueden ser los entornos hardware, actualizaciones de software, despliegue de aplicaciones, soporte técnico, almacenamiento y respaldo de datos. Se centraliza la gestión de todos estos procesos en un único servidor.

Los clientes que actúan en esta arquitectura suelen llamarse thin clients, o clientes ligeros, este es un término general para dispositivos que se basan en un servidor para operar. El thin client proporciona pantalla, teclado, ratón y un procesador básico que interactúa con el servidor. Los thin client

no almacenan ningún dato localmente y requiere de pocos recursos de procesamiento. La característica más destacada de estos terminales es la reducción de costes asociados con el mantenimiento, administración, soporte, seguridad e instalación de aplicaciones comparándolo con un PC tradicional.

Esta tecnología está compuesta por tres componentes principales:

- Sistemas operativos multi-usuario que permiten el acceso y ejecución de modo concurrente, usando aplicaciones diferentes y con sesiones de usuario protegidas. Ejemplos de algunas terminales de servicio son: 2x Terminal Server para Linux, Microsoft Windows Terminal Server (Windows NT/2000), Microsoft Windows Terminal Services (Windows 2003), Citrix Presentation Server, Citrix XenApp Server, AppliDis Fusion, 2X Application Server, HOBlink, Propalms TSE (antes Tarantella), Jethro cabina, GraphOn GO-Global, VMware View..

- El thin client se puede ejecutar con una cantidad mínima de software pero necesita al menos un programa de conexión a servicios de terminal. El thin client y el programa de servicios de terminal pueden ser ejecutados en sistemas operativos completamente diferentes.

- Un protocolo que permita al programa de servicios de terminal y al thin client comunicarse y enviar las pulsaciones de teclado, de ratón y las actualizaciones de pantalla a través de la red. Los protocolos más populares son RDP3 (Remote Desktop protocol), ICA y NX.

Entre las ventajas de la computación basada en servidor se puede nombrar:

- Reducción de los costes de administración. La gestión de clientes ligeros está casi en su totalidad centralizada en el servidor.

- Reducción de costes de hardware. El hardware en los clientes ligeros es generalmente más barato porque no es necesario tener memoria

para las aplicaciones o un procesador de gran alcance.

- Seguridad. Puede ser controlada centralmente.

- Menor consumo de energía. El hardware especializado en el cliente ligero tiene un consumo mucho menor de energía que los tradicionales.

- Reducción de la carga de red. El tráfico de red que generan los terminales ligeros sólo es el de los movimientos del ratón, teclado e información de pantalla desde / hacia el usuario. En el caso de que un cliente pesado abriese y guardase un documento ya implicaría el paso de este dos veces por la red. Usando protocolos eficientes de red tales como ICA y NX ya es posible usar esta tecnología en un ancho de banda de 28,8 Kbps.

- Actualización de hardware simple. Si el uso está por encima de un límite predefinido, es relativamente sencillo solucionar el problema, bastaría con un disco nuevo en un rack de servidores, aumentando así el número de recursos, exactamente la cantidad necesaria. Si ocurriese esto con clientes pesados habría que reemplazar un PC completo, lo que acarrearía tanto costes económicos como de recursos humanos.

A pesar de lo anterior, esta tecnología también presenta ciertos inconvenientes:

- Altos requerimientos de servidor. Al centrarse la carga de trabajo en el servidor, el sistema de clientes ligeros implica mayor consumo de recursos en los servidores, incluso es habitual que se use un gran número de servidores, lo que se denomina “granja de servidores”.

- Pobre rendimiento multimedia. El envío de datos de audio y video requieren mucho ancho de banda, por lo que estos sistemas son menos útiles para aplicaciones multimedia.

- Menos flexibilidad. No todos los productos software del mercado

pueden funcionar correctamente en un cliente ligero.

465 GRID COMPUTING

Arquitectura distribuida y paralela, de ámbito extenso geográficamente, en la que se premia la distribución, y a continuación la paralelización. Sus creadores fueron Ian Foster y Carl Kesselman. Su nombre proviene del paradigma de la red eléctrica (power grid).

Se basa en la compartición, selección y agregación de forma dinámica y en tiempo de ejecución de recursos autónomos, distribuidos geográficamente, dependiendo de criterios como la disponibilidad del hardware, la capacidad transaccional, el rendimiento que se pueda aportar a la solución final, el coste y los criterios de calidad del servicio que el demandante pueda proporcionar y exigir.

La red está formada por un conjunto de ordenadores independientes e interconectados que ponen a disposición del grid los excedentes de su procesamiento individual, es decir, los ciclos de reloj de sus CPUs no aprovechados por ellos mismos, sin poder superar un determinado porcentaje de dedicación configurado individualmente en cada nodo. A partir del porcentaje proporcionado por cada nodo, se virtualiza un recurso computacional único.

Los sistemas basados en grid computing están indicados para atender productividades sostenidas y sostenibles, sin poder nunca superar un determinado umbral. En estos sistemas se garantiza la escalabilidad como un criterio parametrizable. Es posible definir con qué criterio añadimos cada nuevo nodo a la solución final.

Actualmente, el único criterio que se tiene en cuenta es la capacidad de procesamiento (transaccionalidad), pero en el futuro, será posible tener en cuenta criterios más finos, referidos a la calidad del servicio.

Además, estos sistemas están dotados de un comportamiento dinámico,

según el cual, un determinado programa en ejecución en el sistema, puede modificar en tiempo real el dimensionamiento de la grid para adaptarlo a sus necesidades.

465.1 Características:

- Podemos conseguir un máximo aprovechamiento de los nodos (100% de utilización de la CPU).
- Los nodos no tienen que estar dedicados. Además, al contrario que en el caso del cluster, nos aseguramos que la aportación al Grid no va a sobrepasar un determinado porcentaje de tiempo de procesamiento en cada nodo.
- Son sistemas heterogéneos, en los que podemos encontrar diversos HW y SW.
- La escalabilidad parametrizable es la característica más potente de esta arquitectura.

465.2 Funcionalidades:

- Localización dinámica de recursos (máquinas con excedente).
- Optimización del acceso a datos, mapeando las estructuras de datos en cachés temporales locales (directorios).
- Autenticación del usuario (usr/pwd, certificados…).
- Monitorización de tareas y procesos desde cualquier nodo de la red, siempre que el usuario tenga permisos.
- Las máquinas se encuentran en situación paritaria.
- Si es posible, se paraleliza. Lo fundamental es la distribución de procesos débilmente acoplados.

465.3 Arquitectura Grid

Habitualmente se describe la arquitectura del Grid en términos de "capas", ejecutando cada una de ellas una determinada función. Como es habitual en este tipo de enfoque, las capas más altas están más cerca del usuario, en tanto que las capas inferiores lo están de las redes de comunicación.

Empezando por los cimientos, nos encontramos con la capa de red, responsable de asegurar la conexión entre los recursos que forman el Grid.

En la parte más alta está la capa de recursos, constituida por los dispositivos que forman parte del Grid: ordenadores, sistemas de almacenamiento, catálogos electrónicos de datos e incluso sensores que se conecten directamente a la red.

En la zona intermedia está la capa "middleware", encargada de proporcionar las herramientas que permiten que los distintos elementos (servidores, almacenes de datos, redes, etc.) participen de forma coordinada en un entorno Grid unificado. Esta capa es la encargada de las siguientes funciones:

Encontrar el lugar conveniente para ejecutar la tarea solicitada por el usuario.

- Optimiza el uso de recursos, que pueden estar muy dispersos.
- Organiza el acceso eficiente a los datos.
- Se encarga de la autenticación de los diferentes elementos.
- Se ocupa de las políticas de asignación de recursos.
- Ejecuta las tareas.
- Monitoriza el progreso de los trabajos en ejecución.
- Gestiona la recuperación frente a fallos.

- Avisa cuando se haya terminado la tarea y devuelve los resultados.

El ingrediente fundamental del middleware son los metadatos (datos sobre los datos), que contienen, entre otras cosas, toda la información sobre el formato de los datos y dónde se almacenan (a veces en varios sitios distintos).

El middleware está formado por muchos programas software. Algunos de esos programas actúan como agentes y otros como intermediarios, negociando entre sí, de forma automática, en representación de los usuarios del Grid y de los proveedores de recursos. Los agentes individuales presentan los metadatos referidos a los usuarios, datos y recursos. Los intermediarios se encargan de las negociaciones entre máquinas (M2M) para la autenticación y autorización de los usuarios y se encargan de definir los acuerdos de acceso a los datos y recursos y, en su caso, el pago por los mismos. Cuando queda establecido el acuerdo, un intermediario planifica las tareas de cómputo y supervisa las transferencias de datos necesarias para acometer cada trabajo concreto. Al mismo tiempo, una serie de agentes supervisores especiales optimizan las rutas a través de la red y monitorizan la calidad del servicio.

En la capa superior de este esquema está la capa de aplicación donde se incluyen todas las aplicaciones de los usuarios, portales y herramientas de desarrollo que soportan esas aplicaciones. Esta es la capa que ve el usuario.

Además, en las arquitecturas más comunes del Grid, la capa de aplicación proporciona el llamado "serviceware", que recoge las funciones generales de gestión tales como la contabilidad del uso del Grid que hace cada usuario.

Para poder hacer todo lo anterior, las aplicaciones que se desarrollen para ser ejecutadas en un PC concreto, tendrán que adaptarse para poder

invocar los servicios adecuados y utilizar los protocolos correctos. Igual que las aplicaciones que inicialmente se crearon para funcionar aisladamente se adaptan para poder ser ejecutadas en un navegador Web, el Grid requerirá que los usuarios dediquen cierto esfuerzo a "GRIDizar" sus aplicaciones.

Sin embargo, una vez adaptadas al Grid, miles de usuarios podrán usar las mismas aplicaciones, utilizando las capas de middleware para adaptarse a los posibles cambios en el tejido del Grid.

466 CLOUD COMPUTING

Modelo que permite acceso a un conjunto compartido de recursos informáticos configurables a través de la red (por ejemplo, redes, servidores, almacenamiento, aplicaciones y servicios) que pueden ser desarrollados y desplegados rápidamente con mínimo esfuerzo de gestión o interacción con el proveedor de servicios.

Este término se refiere a la utilización y el acceso de múltiples recursos basados en servidores a través de una red. Los usuarios de la “nube” pueden acceder a los recursos del servidor utilizando un ordenador, netbook, pad computer, smart phone u otro dispositivo. En el cloud computing, el servidor presenta y gestiona las aplicaciones; los datos también se almacenan de forma remota en la configuración de la nube. Los usuarios no descargan ni instalan aplicaciones en su sistema, todo el procesamiento y almacenamiento se mantiene por el servidor. Los servicios on-line pueden ser ofrecidos a partir de un “proveedor de la nube” o por una organización privada.

466.1 Arquitectura

Normalmente la arquitectura de los sistemas software implicados en el desarrollo de cloud computing incluyen múltiples componentes denominados “componentes cloud” que se comunican mediante

mecanismos de bajo acoplamiento, tales como las colas de mensajes.

Los dos componentes más significativos de la arquitectura cloud computing se conocen como el front-end y el back-end. El front-end es la parte vista por el cliente, es decir, el usuario del PC. Esto incluye la red del cliente y las aplicaciones utilizadas para acceder a la nube a través de una interfaz de usuario, como un navegador web. El back-end de la arquitectura es la propia nube, que comprende varios ordenadores, servidores y dispositivos de almacenamiento de datos.

Dentro de esta arquitectura se pueden distinguir las siguientes capas:

- Proveedor: Empresa responsable de proporcionar el servicio en la "nube".

- Cliente: Serán el hardware y software diseñados para cloud computing, que permiten interactuar con los servicios remotos.

- Aplicación: Son los servicios en la "nube" o "Software as a Service" (SaaS), el software se proporciona a través de internet como si de un servicio se tratase. De este modo se evita la necesidad de instalar y ejecutar en el equipo del cliente la aplicación. Se reducen así el mantenimiento y el apoyo.

- Plataforma: Son los servicios de plataforma en la "nube", también conocidos como "Platform as Service" (PaaS), proporcionan una plataforma de procesamiento y una pila de soluciones como un servicio, constituyen la base e infraestructura de las aplicaciones de la nube. Facilita el desarrollo de aplicaciones evitando el coste y la complejidad de comprar y mantener el hardware y las capas de software de base.

- Infraestructura. Servicios de infraestructura, también conocidos como "Infrastructure as a Service" (IaaS), proporciona la infraestructura como un servicio, suele ser una plataforma virtualizada. En lugar de comprar servidores, software, centro de datos especiales o equipos de red, los

clientes adquieren dichos recursos de servicios externos. La IaaS ha evolucionado a partir de las ofertas de servidores virtuales privados.

466.2 Modelos de implementación

- Nube pública o external cloud: Es el concepto tradicional donde los recursos se presentan a través de internet en función de la demanda, a través de aplicaciones o servicios web.

- Nube de la comunidad: Se da cuando varias organizaciones con las mismas necesidades comparten recursos. En este caso existen menos usuarios que en la nube pública y se ofrece mayor privacidad y seguridad. Un ejemplo puede ser el Google`s “Gov Cloud”.

- Nube híbrida. Es común que una empresa use tanto la nube pública como desarrollos privados para satisfacer sus necesidades con respecto a las TI. Existen varias empresas como HP, IBM, Oracle and VMware que ofrecen tecnologías para manejar la complejidad de mantenimiento, seguridad y privacidad consecuencia del uso del conjunto de estos servicios.

- Nube combinada. Se denomina al conjunto formado varios servicios cloud de distintos proveedores.

- Nube privada. Es trasladar el concepto de nube pública a una red de uso privado. Es decir, el uso de la nube única y exclusivamente dentro de la red de una empresa.

467 GREEN IT E EFICIENCIA ENERGÉTICA

El término Green Computing se acuñó posiblemente por primera vez tras el inicio del programa Energy Star en 1992, promocionado por el gobierno estadounidense.

Tenía por objetivo etiquetar monitores y equipamiento electrónico

caracterizados por su eficiencia energética. El término quedó registrado ya en 1992 en un grupo de noticias. Hoy en día el programa Energy Star es el motor de la eficiencia energética en los sistemas electrónicos (no sólo de procesamiento de la información, sino también del equipamiento electrónico doméstico).

La adopción de productos y aproximaciones más eficientes pueden permitir más equipamiento dentro del mismo gasto energético, lo que se denomina huella energética, o energy footprint. Las regulaciones se están multiplicando y podrían limitar seriamente a las empresas a la hora de construir centros de procesamiento datos, ya que el efecto de las redes de suministro eléctrico, las emisiones de carbono por el incremento de uso y otros impactos medioambientales están siendo investigadas. Por tanto, las organizaciones deben considerar las regulaciones y tener planes alternativos para el crecimiento de sus centros de procesamiento de datos y de su capacidad.

Con el paso de los años, el número de servidores existentes en todo el mundo crece de forma casi exponencial. Consecuencia de esto es el creciente gasto energético para la refrigeración y gestión de los equipos. Hoy en día ya se están empezando a plantear soluciones que optimicen este gasto energético.

Este consumo energético no es el único problema ambiental relacionado con las TI. La etapa de fabricación de equipos presenta serios problemas relacionados con el medio ambiente: materiales de desecho tóxicos, producción de gases contaminantes, etc. La tendencia actual es la de minimizar el impacto contaminante (carbon footprint) presente en las tecnologías de fabricación de los sistemas electrónicos.

Finalmente, también tiene un impacto inmediato la eliminación de equipos para las TI, caracterizados por un tiempo de vida increíblemente breve de unos dos o tres años. Si no se reciclan de forma eficiente, terminan tirados en vertederos, y debido a la presencia de componentes tóxicos, son una

fuente de contaminación terrestre y de las aguas. Todos estos aspectos deben ser considerados de manera global por los fabricantes y usuarios de equipos TI. La concienciación de la existencia de este problema ha llevado a la elaboración de numerosas y rígidas normativas a todos los niveles, lo que empieza a obtener algunos resultados.

GreenPeace Internacional realiza un ranking con los 18 principales fabricantes del sector electrónico (ordenadores personales, teléfonos móviles etc.) de acuerdo con sus políticas de reducción de emisiones tóxicas, reciclado o minimización de impacto en el cambio climático, y lo publica en su Guía para la Electrónica Verde (Guide to Greener Electronics), de publicación trimestral. Como se puede ver en los resultados de Diciembre de 2010, las empresas del sector obtienen unas calificaciones realmente bajas, siendo la mejor Nokia con un 7,5 sobre 10.

La mitad de estas 18 empresas suspenden un estudio que busca que las empresas analizadas:

- Limpien sus productos al eliminar sustancias peligrosas. Los productos químicos peligrosos con riesgo impiden el posterior reciclado de los equipos.
- Reciclen de equipos/productos bajo su responsabilidad una vez quedan obsoletos.
- Reduzcan el impacto climático debido a sus operaciones y productos.

Por todo lo expuesto, la resolución efectiva del impacto ambiental de las tecnologías TI requiere un enfoque holístico del problema que englobe las cuatro vías:

- Utilización ecológica: principalmente a través de la reducción del consumo energético. La producción de energía eléctrica es la principal fuente de generación de gases de efecto invernadero.

- Diseño ecológico o eco-diseño: incluye diseño de equipos más eficientes energéticamente y respetuosos con el medio ambiente.

- Fabricación ecológica: eliminando completamente o minimizando el impacto del proceso de fabricación en el medio ambiente (emisiones, materiales de desecho, etc.).

- Eliminación ecológica: una vez finalizado el período de utilización de un equipo se deben poner en marcha las estrategias denominadas tres R: reutilización y renovación de equipos y, si no son aprovechables, reciclado.

La idea principal del enfoque holístico es que se cierre el ciclo de vida de los equipos TI de forma que no se perjudique el medio ambiente, lo que permitiría conseguir una mejora sustancial de cara al desarrollo sostenible.

467.1 Tecnologías verdes

Hoy en día existen distintos enfoques tecnológicos que se acercan a un desarrollo sostenible de las TI.

- Monitores LCD. Con el paso de los años los monitores pasaron de ser CRT a LCD, este cambio no es sólo estético o de tamaño, sino que los niveles de consumo han disminuido notablemente. Un monitor CRT medio requiere 85W si está activo, frente a los 15W de uno LCD, 5W en modo bajo consumo para un CRT mientras que un LCD consumiría 1,5W. Apagados ambos consumirían 0,5W. En los últimos años se ha revolucionado el mercado de las pantallas de ordenador con la aparición de la tecnología OLED (Organic Light Emitting Diode), basadas en la utilización de diodos LED cuya capa electro-luminiscente se hace con un compuesto orgánico (un polímero que se ilumina al aplicarle un voltaje). La ventaja principal de este tipo de pantallas frente a las tradicionales de cristal líquido (LCD) es que los diodos OLED no necesitan retro-iluminación, por lo que el consumo de energía que requieren es muy

inferior.

- Discos duros. El consumo de los discos duros no es para nada despreciable, sobre todo en el arranque del sistema. Por ejemplo, el disco Seagate Barracuda 7200.8 requiere hasta 2,5 A de la línea de alimentación de 12 V. Si a esto le sumamos 3W que extrae desde la línea de +5 V se puede llegar a un consumo de pico en el arranque de 33 W. Si en lugar de sólo un disco duro hablamos de un equipo con dos o más empezamos a hablar de cifras muy comprometidas. Esto ha hecho que los fabricantes de discos duros comiencen a tener en cuenta el consumo en sus productos, creando casi todos una nueva gama denominada “verde” o “ecológica”. Por ejemplo Así, Western Digital con “Caviar Green”, Samsung con Eco Green, o Hitachi con eco-friendly Deskstar y Travelstar. Como alternativa a los discos tradicionales aparecen los discos en estado sólido (SSD), que presentan menores consumos de energía y es la tecnología a la que se espera evolucionen los sistemas de almacenamiento.

- CPDs. Aquí es donde se aloja toda la infraestructura de soporte a los diversos servicios computacionales, y una estructura adecuada permitirá buenos ahorros de energía, de espacio y de costos a mediano y/o largo plazo. Buscando la reducción de energía se puede empezar por la acción más simple que es apagar el equipo que no se esté utilizando, la reducción del hardware estudiando necesidades reales, o actuaciones específicas en función de la actividad de la empresa.

- Virtualización. La virtualización de servidores permite el funcionamiento de múltiples servidores en un único servidor físico. Esto ayuda a reducir la huella de carbono del centro de datos al disminuir el número de servidores físicos y consolidar múltiples aplicaciones en un único servidor con lo cual se consume menos energía y se requiere menos enfriamiento. Además se logra un mayor índice de utilización de recursos y ahorro de espacio.

- Cliente/Servidor. Estos sistemas mantienen el software, las aplicaciones y los datos en el servidor. Se puede tener acceso a la información desde cualquier ubicación y el cliente no requiere mucha memoria o almacenamiento. Este ambiente consume menos energía y enfriamiento.

- Cloud computing. Esto proporciona a sus usuarios la posibilidad de utilizar una amplia gama de recursos en red para completar su trabajo. Al utilizar computación en nube las empresas se vuelven más ecológicas porque disminuyen su consumo de energía al incrementar su capacidad sin necesidad de invertir en más infraestructura.

- Tele trabajo. Definido por Merrian-Webster como el trabajo en casa con el uso de un enlace electrónico con la oficina central. Al no desplazarse el empleado la contaminación es menor.

467.2 Actividades relacionadas con Green IT

Existen varias actividades que promocionan e intentan solventar las cuestiones expuestas anteriormente. Estas actividades están patrocinadas bien desde administraciones públicas, bien desde empresas, que están entendiendo que Green IT, además de una necesidad, puede ser un negocio, desde el punto de vista de consultoría y servicios, o bien por consorcios de empresas.

The Green Grid (http://www.thegreengrid.org) es un consorcio global dedicado a avanzar en la eficiencia energética de los centros de procesamiento de datos y en ecosistemas de computación de negocio. En cumplimiento de su misión, The Green Grid se centra en:

- Definir métricas y modelos significativos y centrados en el usuario.

- Desarrollar estándares, métodos de medida, procesos y nuevas tecnologías para mejorar el rendimiento de los centros de procesamiento de datos frente a las métricas definidas.

- Promocionar la adopción de estándares, procesos, medidas y tecnologías energéticamente eficientes.

El comité de directores de The Green Grid está compuesto por las siguientes compañías miembros: AMD, APC, Dell, HP, IBM, Intel, Microsoft, Rackable Systems, Sun Microsystems y VMware.

Climate Savers. Iniciada por Google e Intel en 2007, Climate Savers Computing Initiative (www.climatesaverscomputing.org) es un grupo sin ánimo de lucro de consumidores y negocios con conciencia ecológica y organizaciones conservacionistas. La iniciativa se inició bajo el espíritu del programa Climate Savers de WWF (http://www.worldwildlife.org/climate/projects/climateSavers.cfm), que ha movilizado a una docena de compañías desde 1999 a recortar las emisiones de dióxido de carbono, demostrando que reducir las emisiones es bueno para el negocio. Su objetivo es promover el desarrollo, despliegue y adopción de tecnologías inteligentes que puedan mejorar la eficiencia de uso de la energía del computador y reducir su consumo cuando el computador se encuentra inactivo.

SNIA (Storage Networking Industry Association, http://www.snia.org) es una organización global sin ánimo de lucro compuesta por unas compañías de la industria del almacenamiento. SNIA Green Storage Initiative (http://www.snia.org/green) está llevando a cabo una iniciativa para avanzar en el desarrollo de soluciones energéticamente eficientes para el almacenamiento en red, incluyendo la promoción de métricas estándares, la formación y el desarrollo de buenas prácticas energéticas o el establecimiento de alianzas con organizaciones como The Green Grid.

Energy Star. En 1992 la Agencia de Protección Medioambiental de EEUU (U.S. Environmental Protection Agency) lanzó el programa Energy Star, que se planificó para promocionar y reconocer eficiencia energética en monitores, equipos de climatización y otras tecnologías. Aunque de carácter voluntario inicialmente, resultó pronto de amplia aceptación,

pasando a ser un hecho la presencia de un modo de descanso (sleep mode) en la electrónica de consumo.

Directiva Europea de Eco-Diseño. Siguiendo la misma línea que la iniciativa Energy Star de EEUU, la Unión Europea aprobó la directiva 2005/32/EC para el eco-diseño, nuevo concepto creado para reducir el consumo de energía de productos que la requieren, tales como los dispositivos eléctricos y electrónicos o electrodomésticos. La información relacionada con las prestaciones medioambientales de un producto debe ser visible de forma que el consumidor pueda comparar antes de comprar, lo cual está regulado por la Directiva de Etiquetado de la Energía (Energy Labelling Directive). Los productos a los que se conceda la Eco-etiqueta serán considerados como cumplidores con la implementación de las medidas, de forma muy similar a la etiqueta de Energy Star.

El Código de Conducta de la Unión Europea para Centros de Datos está siendo creado como respuesta al creciente consumo de energía en centros de datos y a la necesidad de reducir el impacto ambiental, económico y de seguridad de abastecimiento energético relacionado. El objetivo es informar y estimular a los operadores o propietarios de los centros de datos a que reduzcan el consumo de energía de una forma rentable sin dificultar su funcionamiento. Este código de conducta quiere conseguir esto mediante la mejora de la comprensión de la demanda de energía dentro del centro de datos, aumentando la concienciación, y mediante la recomendación de prácticas y objetivos energéticamente eficientes.

Grupo de trabajo sobre Green IT de la plataforma INES (Iniciativa Española de Software y Servicios, http://www.ines.org.es) es la Plataforma Tecnológica Española en el área de los Sistemas y Servicios Software y constituye una red de cooperación científico-tecnológica integrada por los agentes tecnológicos relevantes de este ámbito (empresas, universidades, centros tecnológicos, etc.).

Según la Agenda Estratégica de Investigación de INES, el plan de

dinamización para el Grupo de Trabajo de Green IT consiste en las siguientes acciones:

- Análisis de la influencia e importancia de las soluciones de Green IT.
- Difusión de las informaciones, noticias y existencia de este grupo de trabajo por Internet.
- Fomentar el interés y apoyar el desarrollo bajo Green IT.

Big Green Innovations (http://www.ibm.com/technology/greeninnovations/), programa de IBM. Dentro de este programa, y con fines educativos, IBM ha presentado un centro de datos virtual ecológico denominado Virtual Green Data Center.

La lista Green500 (http://www.green500.org) proporciona una clasificación de los supercomputadores más eficientes energéticamente del mundo, sirviendo como una visión complementaria a la lista Top500 (http://www.top500.org).

Otras empresas, como Google, Dell o Symantec, están desarrollando programas de eficiencia energética, tanto para sus propios procesos de TI como para los de sus clientes.

468 REDES SAN Y ELEMENTOS DE UNA SAN

Como resumen una SAN es una red donde se realiza el almacenamiento y se gestiona la seguridad de los datos. Las SAN (Storage Area Network, Redes de Almacenamiento) son redes en las que se conectan servidores de almacenamiento (especialmente arrays de discos). También hay que considerar como parte de las SAN las librerías necesarias para el uso de los arrays y los accesos a las redes. De forma contraria a las redes tradicionales, en las SAN se emplean protocolos orientados a la

recuperación de la información de los arrays de disco e inspirados en los propios estándares de comunicación con discos tradicionales (SCSI y SATA).

Normalmente los equipos diseñados para participar en estas redes suele ser especialmente caro aunque su precio depende, en una grande medida, de las tecnologías y protocolos empleados para la transmisión de los datos. Entre las tecnologías disponibles en la actualidad se encuentran: iSCSI (Internet Small Computer Storage Interconnect), Fibre Channel y AOE (ATA Over Ethernet, Advanced Technology Attachment Over Ethernet).

Entre las ventajas de la interconexión de redes de almacenamiento se resaltan las siguientes:

- Elimina los límites de distancia de discos introducidos por SCSI o ATA
- Consigue un mayor caudal de datos ya que los protocolos están específicamente diseñados para la trasferencia de datos de dispositivos de almacenamiento.
- Permite un aprovechamiento mayor de los discos permitiendo que más de un servidor acceda al mismo disco.
- Capacidad para el uso de múltiples discos de forma trasparente desde uno o varios servidores.
- Adquisición de discos diferida debido al mayor aprovechamiento
- Capacidades de recuperación ante desastres. Los arrays de discos empleados en las SAN suelen disponer de discos de reserva (para fallos de otros discos) y permitir distintos esquemas de RAID.
- Recuperación en caliente ante desastres
- Mejor capacidad de administración. La administración es más sencilla y está más centralizada.
- Reducción de los costos de administración y de almacenamiento de

datos

- Mejora de disponibilidad global ya que las SAN tienen menos fallos que los discos internos de los equipos.
- Reducción de servidores eliminando servidores de arquitecturas antiguas (NFS, SMB, etc).
- Reducción del caudal de las redes convencionales, pues las copias de seguridad se pueden hacer desde las SAN.
- Incremento de la rapidez de las operaciones de Entrada/Salida
- Reducción de los costes de administración de backups
- Protección de datos críticos
- Incremento de la capacidad de forma trasparente
- Desarrollo y prueba de aplicacíones de forma más eficiente mediante el uso de copias de los datos de producción realizadas en la SAN.
- Facilita el empleo de clusters de servidores que tienen que disponer de un almacenamiento común.
- Permiten el almacenamiento bajo demanda de forma que cualquiera servidor puede solicitar espacio de almacenamiento segundo sus necesidades.

Dentro de una organización, se debería incluir en una SAN la siguiente información:

- La información almacenada por SGBDs (Sistemas Gestores de Bases de Datos). De hecho, algunos sistemas gestores como Oracle, Sybase, SQLServer, DB2, Informix o Adabase recomiendan esta alternativa
- La información almacenada por servidores de archivos. Los servidores de archivos funcionarán mejor y con menos recursos si los archivos están almacenados en una SAN.
- Servidores de backup. Si los servidores de backup están conectados a una SAN se conseguirá reducir los tiempos de copia de seguridad con

respecto a hacerlos en una LAN (Local Area Network, Red de Área Local) y reducir el tráfico de la LAN.

- Archivos de servidores de voz y video para streamming. Debido a que este tipo de servicios requiere grandes cantidades de disco, una SAN puede reducir los costes asociados al almacenamiento y desplazar el máximo posible el coste (incluir nuevos discos en los arrays cuando sean necesarios).
- Buzones de usuario (mailboxes) de servidores de correo permitiendo que los servidores de correo funcionen mas rápido y que se pueda realizar una restauración rápida en caso de que algún archivo se corrompa.
- Servidores de aplicaciones de alto rendimiento. Las SAN pueden mejorar el rendimiento de cualquier aplicación incluyendo gestores documentales, aplicaciones científicas, aplicaciones de datawarehouse y cuadros de mando integrales, aplicaciones para gestionar las relaciones con los clientes (CRM), etc.
- Soluciones de Virtualización.

Asi mismo, no es te conveniente usar una SAN para:

- Servidores web que no requieran grandes necesidades de almacenamiento (la mayoría)
- Servidores con servicios de red básicos como DNS, DHCP, WINS (Windows Internet Name Servers) y controladores de dominio de Windows (DC). Este tipo de servidores no requieren de las capacidades de almacenamiento permitidas por las SAN
- PCs de escritorio
- Servidores que necesitan menos de 10Gb de almacenamiento
- Servidores que no necesitan un acceso rápido a la información
- Servidores que no comparten archivos

468.1 Estructura de las SAN

Habitualmente las SAN se conciben y se estructuran en tres capas:

1. La capa de hosts: Constituida en su mayoría por los servidores, los drivers y software necesarios para la conexión a la red y los HBAs (Host Bus Adapters) que son dispositivos (tarjetas) que se conectan la cada servidor para acceder al almacenamiento (en algunas soluciones concretas son adaptadores Ethernet simples y en el caso Fibre Channel llevan un conector GBIC-Gigabit Interface Connector).

2. La capa de estructura (fabric layer): Constituida por HUBs, Switches, Gateways y Routers se fuera necesario. Si se emplea la tecnología Fibre Channel, todos estos dispositivos emplean GBICs (Gigabit Interface Conectors) para la interconexión de los dispositivos de las capas superiores e inferiores.

3. La capa de almacenamiento (storage layer): Constítuída por todo tipo de dispositivos de almacenamiento.

Un conjunto de discos situados en el mismo sitio y sin funcionalidades adicionales se conoce como JBOD (Just a Bunch Of Disks). Dentro de la capa de almacenamiento, los arrays no son simplemente JBODs, sino que incluyen ciertas funcionalidades interesantes implementadas en el firmware de la controladora como el RAID.

469 VIRTUALIZACIÓN DEL ALMACENAMIENTO

Este tipo de virtualización permite una mayor funcionalidad y características avanzadas en el sistema de almacenamiento. Consiste en abstraer el almacenamiento lógico del almacenamiento físico y suele

usarse en SANs (Storage Area Network, Red de área de almacenamiento).

Este sistema de almacenamiento también se conoce como “storage pool”, matriz de almacenamiento, matriz de disco o servidor de archivos. Estos sistemas suelen usar hardware y software especializado, junto con unidades de disco con el fin de proporcionar un almacenamiento muy rápido y fiable para el acceso a datos. Son sistemas complejos, y pueden ser considerados como un ordenador de propósito especial diseñado para proporcionar capacidad de almacenamiento junto con funciones avanzadas de protección de datos. Las unidades de disco son sólo un elemento dentro del sistema de almacenamiento, junto con el hardware y el software de propósito especial incorporado en el sistema.

Los sistemas de almacenamiento pueden ser de acceso a nivel de bloque, o acceso a nivel de ficheros. El acceso por bloques suele llevarse a cabo por medio de Fibre Channel , iSCSI , SAS , FICON u otros protocolos. Para el acceso a nivel de archivo se usan los protocolos NFS o CIFS.

Dentro de este contexto nos podemos encontrar con dos tipos principales de virtualización: la virtualización por bloques y la virtualización por archivos.

469.1 Virtualización por bloques

Este tipo de virtualización se basa en la abstracción (diferenciación) entre el almacenamiento lógico y el almacenamiento físico, consiguiendo que el acceso no tenga en cuenta el almacenamiento físico o estructura heterogénea.

Existen tres tipos de virtualización por bloques: basada en host, basada en dispositivos de almacenamiento, basada en red.

469.1.1 Virtualización basada en host

Esta virtualización requiere software adicional que se ejecuta en el host. En algunos casos la administración de volúmenes está integrada en el sistema

operativo, y en otros casos se ofrece como un producto separado. Los volúmenes (LUN) disponibles en el sistema son manejados por un controlador de dispositivos físicos tradicional. Por encima de este controlador se encuentra una capa software (el gestor de volúmenes) que intercepta las peticiones de E / S, y proporciona la búsqueda de meta-datos y mapeos de E / S.

Los sistemas operativos más modernos tienen algún tipo de gestor de volúmenes lógicos integrado (MVI en UNIX / Linux, o Administrador de discos lógicos o LDM en Windows), que realiza tareas de virtualización.

Existen varias tecnologías que implementan este tipo de virtualización, como pueden ser la gestión de volúmenes lógicos (Logical Volume Management, LVM), los sistemas de archivos (CIFS, NFS) o el montaje automático (autofs)

469.1.2 Virtualización basada en dispositivos de almacenamiento

Se puede llevar a cabo la virtualización basada en medios de almacenamiento masivo utilizando un controlador de almacenamiento primario que proporcione los servicios de virtualización y permita conexión directa de los controladores de almacenamiento. En función de la implementación es posible usar modelos de distintos fabricantes.

El controlador primario proporcionará la puesta en común y los meta-datos de servicio de gestión. También puede ofrecer servicios de replicación y migración a través de los controladores que se virtualizan.

Una nueva generación de controladores de serie del disco permite la inserción posterior de los dispositivos de almacenamiento.

Los sistemas RAID pueden ser un ejemplo de esta técnica. Estos sistemas combinan varios discos en una sola matriz.

Las matrices avanzadas de disco, cuentan a menudo con clonación, instantáneas y replicación remota. En general, estos dispositivos no ofrecen

los beneficios de la migración de datos o de replicación a través de almacenamiento heterogéneo, ya que cada fabricante tiende a utilizar sus propios protocolos propietarios.

469.1.3 Virtualización basada en red

Esta es una virtualización de almacenamiento operando en un dispositivo basado en red (por lo general un servidor estándar o un smart switch) y el uso de redes iSCSI o FC de Fibre Channel para conectar como SAN (Storage Area Network). Este es el tipo de virtualización de almacenamiento más común.

El dispositivo de virtualización se encuentra en la SAN y proporciona la capa de abstracción entre los host, que permiten la entrada/salida, y los controladores de almacenamiento, que proporcionan capacidad de almacenamiento.

Hoy en día existen dos implementaciones distintas, la basada en el **dispositivo** y la basada en **conmutación**. Ambos modelos proporcionan los mismos servicios: gestión de discos, búsqueda de meta-datos, migración y replicación de datos. Igualmente, ambos modelos necesitan de un hardware específico que permita ofrecer dichos servicios.

La basada en dispositivos consiste en establecer el hardware especializado entre los hosts y la parte de almacenamiento. Las solicitudes de entrada/salida se redirigen al dispositivo, que realiza la asignación de meta-datos, mediante el envío de sus propias órdenes de E/S a la solicitud de almacenamiento subyacente. El hardware usado también puede proporcionar almacenamiento de datos en caché, y la mayoría de las implementaciones proporcionan algún tipo de agrupación de cada uno de los dispositivos para mantener un punto de vista atómico tanto de los meta-datos como de los datos de la caché.

Este tipo de almacenamiento también puede clasificarse en in-band (simétrica) o out-of-band (asimétrica).

469.1.3.1 In-band (simétrica)

En este caso los dispositivos de virtualización se asientan entre el host y el almacenamiento. Todas las peticiones de E/S y datos pasan a través del dispositivo. Los host nunca interactúan con el dispositivo de almacenamiento sino con el dispositivo de virtualización.

469.1.3.2 Out-of-band (asimétrica)

Los dispositivos usados en este tipo de virtualización también son llamados servidores de meta-datos. La única finalidad de estos dispositivos es proporcionar la asignación de meta-datos. Esto implica el uso de software adicional en el host, que es conocedor de la ubicación real de los datos. De este modo, se intercepta la petición antes de que salga del host, se solicita una búsqueda de meta-datos en el servidor (puede ser a través de una interfaz que no sea SAN) y se devuelve la ubicación real de los datos solicitados por el host. Finalmente se recupera la información a través de una solicitud de E/S común al dispositivo de almacenamiento. No se puede dar un almacenamiento en caché ya que los datos nunca pasan a través del dispositivo de virtualización.

469.2 Virtualización a nivel de archivo

Con este tipo de virtualización se pretende eliminar las dependencias entre el acceso a datos a nivel de archivo y la ubicación física de los mismos. Esta técnica, conocida como NAS (Network-Attached Storage) o almacenamiento conectado a red, suele ser un equipo especializado pensado exclusivamente para almacenar y servir ficheros. Los equipos que funcionan como dispositivo NAS suelen incluir un sistema operativo específico para el propósito, como puede ser FreeNAS o FreeBSD.

Estos sistemas pueden contener uno o más discos duros, dispuestos a menudo en contenedores lógicos redundantes o arrays RAID.

NAS utiliza protocolos basados en archivos como NFS (sistemas UNIX), SMB / CIFS (Server Message Block/Common Internet File System) (sistemas MS Windows), o AFP (Apple Filing Protocol, sistemas Apple Macintosh). Las unidades NAS no suelen limitar a los clientes a un único protocolo. FTP, SFTP, HTTP, UPnP, rsync y AFS (Andrew File System) también lo soportan.

De este modo se consigue optimizar la utilización del almacenamiento y las migraciones de archivos sin interrupciones.

469.3 Diferencias entre NAS y SAN

NAS proporciona almacenamiento y un sistema de archivos, lo que suele contrastar con SAN, que solamente proporciona almacenamiento basado en bloques y deja del lado del cliente la gestión del sistema de archivos.

NAS aparece en el sistema cliente como un servidor de archivos (se pueden asignar unidades de red a las acciones del servidor) mientras que un disco a través de una SAN se presenta al cliente como un disco más del sistema operativo, que podemos montar, desmontar, formatear...

	NAS	**SAN**
Tipo de datos	Archivos compartidos	Datos a nivel de bloque, por ejemplo, bases de datos.
Cableado utilizado	Ethernet LAN	Fibre Channel dedicado
Clientes principales	Usuarios finales	Servidores de aplicaciones
Acceso a disco	A través del dispositivo NAS (IP propia)	Acceso directo

4610 GESTIÓN DEL CICLO DE VIDA DE LA INFORMACIÓN (ILM)

4610.1 Gestión del Ciclo de Vida de los Datos

La gestión del ciclo de vida de los datos o DLM (Data Lifecyle Management) es un enfoque de la gestión de la información desde el punto de vista del manejo del flujo de los datos de un sistema de información durante todo su ciclo de vida, desde que se crean y se produce su primer almacenamiento, hasta que son declarados obsoletos y eliminados del sistema.

Los productos para la gestión del ciclo de vida de los datos tratan de automatizar los procesos que forman parte de este ciclo de vida. Organizan los datos en distintos niveles siguiendo unas políticas especificadas, y automatizan la migración o intercambio de los datos entre unos niveles y otros basándose para ello en los criterios especificados de cada uno.

Como norma general, los datos más recientes y aquellos a los que se accede con más frecuencia se tienden a almacenar en medios de almacenamiento más rápidos, pero también más caros, mientras que los datos de un nivel menos critico se almacenan en los dispositivos más baratos y más lentos.

Las arquitecturas que gestionan el ciclo de vida de los datos suelen incluir un sistema de archivos que indexa toda aquella información critica y aquella considerada no tan crítica pero que guarda relevancia o relación que esta. Con esta información crea copias de respaldo, los almacena en ubicaciones seguras para evitar manipulaciones pero que puedan ser accesibles de una manera segura y confiable.

Estas arquitecturas también se encargan de las posibles duplicaciones de datos y de la comprensión de los mismos para garantizar un correcto y eficiente uso del espacio de almacenamiento disponible.

Desafortunadamente, muchas implementaciones de DLM de negocios se han estancado, principalmente porque las empresas no han logrado definir ni las políticas de migración adecuadas ni el archivado de datos. Dado que esas políticas necesitan reflejar las prioridades de regulación y de negocio,

en sus definiciones es necesario un a colaboración que involucre no solo a miembros del departamento de tecnologías de la información, sino también a miembros de diferentes departamentos del negocio.

Por otro lado, el criterio más sencillo para realizar una migración de la información a un sistema de almacenamiento más económico es el temporal, es decir, los datos más antiguos en los sistemas más lentos y baratos. Sin embargo, las empresas en industrias altamente reguladas a menudo quieren ir más lejos, estableciendo la clasificación de los datos en función de la rapidez con la que se precisen, o la frecuencia con la que se accede a ellos, o en base a quien los ha enviado o recibido, o en base a un conjunto de palabras clave o cadenas numéricas, etc. Entonces el reto está en conseguir definirlos de tal manera que sea viable realizarlo en el tiempo y mediante la menor intervención humana.

4610.2 Gestión del Ciclo de Vida de la Información

La gestión del ciclo de vida de la información o ILM (Information Lifecycle Management) es un enfoque integral para el manejo del flujo de los datos de un sistema de información y los metadatos asociados desde su creación y almacenamiento inicial hasta el momento en que estos se vuelven obsoletos y son borrados.

A diferencia de anteriores enfoques para la gestión de almacenamiento de datos, ILM abarca todos los aspectos en los que se tratan los datos, partiendo de las prácticas de los usuarios, en lugar de la automatización de los procedimientos de almacenamiento y en contraste con los sistemas más antiguos, ILM permite criterios mucho más complejos para la realización de la gestión del almacenamiento que la antigüedad de los datos o la frecuencia de acceso a ellos.

Es importante destacar que ILM no es sólo una tecnología sino que integra los procesos de negocio y TI con el fin de determinar cómo fluyen los datos

a través de una organización, permitiendo a los usuarios y administradores gestionarlos datos desde el momento que se crean hasta el instante en el que ya no son necesarios.

Aunque los términos gestión del ciclo de vida de los datos (DLM) y gestión del ciclo de vida de la información (ILM) a veces se utilizan indistintamente, ILM a se considera un proceso más complejo.

La clasificación de los datos en función de valores del negocio es una parte integral y muy importante del proceso ILM. Esto quiere decir que ILM reconoce que la importancia de los datos no se basa únicamente en su antigüedad o en su frecuencia de acceso, sino que ILM espera que sean los usuario y los administradores los que especifiquen distintas directivas para que los datos vayan variando de una manera decreciente su relevancia o grado de importancia para la organización, o que puedan conservar su importancia durante todo su ciclo de vida, etc.

Para una exitosa y eficiente implementación de IML se necesita que la organización identifique requisitos de seguridad de los datos críticos e incluirlos en sus procesos de clasificación. Los usuarios de los datos, tanto los individuos como las aplicaciones, deben de ser identificados y categorizados en función de las necesidades asociadas con sus tareas.

Algunas de las mejores prácticas relacionadas con la implementación de IML comparten enfoques como:

- Se centran en la productividad del usuario con el fin de obtener una ventaja estratégica a través del acceso a los datos necesarios.
- Proteger los datos contra el robo, la mutilación, la divulgación involuntaria, o la eliminación.
- Crear múltiples capas de seguridad, sin crear una gestión excesivamente compleja.

- Asegurarse que los procesos de seguridad están incorporados en los procesos generales del negocio y en los procesos de TI.
- Utilizar estándares y modelos de referencias con el fin de satisfacer únicamente las necesidades de seguridad de la organización.

Por supuesto, cada organización deberá desarrollar e implementar su propia solución de seguridad de almacenamiento, que debe seguir evolucionando, adaptándose a las nuevas oportunidades, amenazas y capacidades.

4610.3 Alguna soluciones para la gestión

4610.3.1 Microsoft

Microsoft Identity Lifecycle Manager ofrece una solución integrada y completa para la gestión del ciclo de vida de las identidades de usuario y sus credenciales asociadas. Esta solución aporta la sincronización de identidades, los certificados y administración de contraseñas y suministro de usuarios. La solución funciona bajo plataformas Windows y otros sistemas organizacionales.

4610.3.2 IBM

Las soluciones de IBM para la gestión del ciclo de vida de la información se han agrupado en cinco categorías (IBM, 2008):

- *Archivo de correo electrónico* (IBM DB2 CommonStore, VERITAS Enterprise Vault, OpenText-IXOS Livelink)
- *Aplicación y base de datos de archivo* (Archivo Activo de Princeton Softech),

- *Gestión del ciclo de vida de los datos* (TotalStorage de IBM SAN File System)
- *Gestión de contenidos* (repositorio de administración de contenido, DB2 Content Manager)
- *Gestión de registros* (IBM DB2 Records Manager).

4610.3.3 Oracle

Oracle ILM Assistant es una herramienta que se basa en una interfaz gráfica de usuario para la gestión de entorno de ILM. Ofrece la posibilidad de crear definiciones de ciclo de vida, que se asignan a las tablas en la base de datos. Posteriormente basándose en las políticas establecidas sobre el ciclo de vida, ILM Assistant informa cuando es el momento para mover, archivar o suprimir los datos. También muestra las necesidades de almacenamiento y el ahorro de costes asociados con el cambio de ubicación de los datos.

Otras capacidades de Oracle ILM Assistant incluyen la habilidad de mostrar cómo particionar una tabla basada en una definición del ciclo de vida, y poder simular los eventos para comparar el resultado en caso de que la tabla fuera particionada.

4611 SISTEMAS DE BACKUP: HARDWARE Y SOFTWARE DE BACKUP

Un factor importante en todo sistema de backup es la elección de los sistemas hardware y software que lo componen.

4611.1 Hardware de Backup

En la categoría de elementos hardware de backup tenemos:

4611.1.1 Cintas

Tradicionalmente, los cartuchos de cinta magnética son los medios de comunicación más habituales en los sistemas de backup. Como soporte de almacenamiento de los respaldos de datos, la cinta magnética tiene una larga historia de uso y es el medio de copia de seguridad con mayor nivel de madurez. La cinta magnética, o de una forma más abreviada, la cinta, es un componente basado en cartuchos que se hace típicamente de algún tipo de plástico rígido. Contiene uno o más bobinas de plástico flexible que se han impregnado con un material con comportamiento magnético.

Los cartuchos de cinta están fabricados en varios formatos. Cada formato tiene unas características diferentes que responden a las diferentes necesidades de almacenamiento físico y de tiempo de preservación de la copia de seguridad, tanto en términos de la cantidad de datos almacenados, como de vida útil de los medios de almacenamiento o su coste. Los formatos de cinta de uso común son los siguientes:

- DLT/ SDLT
- LTO
- AIT
- STK 9840/9940/T10000

Según el tipo de cada cartucho este posee distintas capacidades o características como la velocidad de funcionamiento. El mercado está renovando continuamente este tipo de dispositivos con el fin de mejorar ambos aspectos. Sin embargo, existen tres formatos que podemos considerar de los más comunes y tienen características particulares que se describen aquí como ejemplos de elementos arquitectónicos de diseño:

DLT, LTO, T10000 y STK.

El resto de formatos, aunque sean formatos comunes, se utilizan normalmente para entornos especializados, como el archivado y almacenamiento intermedio (nearline storage) empleado entre el almacenamiento online y el almacenamiento de backups.

4611.1.1.1 Digital Linear Tape (DLT)

Digital Linear Tape (DLT) es el formato de cinta más antiguo y por lo tanto uno de los productos más maduros del mercado. Originalmente fue diseñado e implementado por DEC en 1984, para posteriormente ser adquirida por Quantum y redistribuido en 1994.

DLT es el primer cartucho de cinta compacta para copias de seguridad de sistemas abiertos en la empresa. Mientras que otros tipos de medios se encontraban en uso (como la cinta media pulgada, 4mm/8mm, y otros), DLT proporciona el mejor compromiso entre todos los factores debido a su tamaño, la fiabilidad de su almacenamiento, la capacidad, y disponibilidad relativa.

La conectividad de DLT, se limita a los tradicionales de SCSI, y está limitado a 300 GB de capacidad nativa de almacenamiento y 160 MB /seg velocidad de transferencia (SDLT600). Existían otras variantes disponibles, pero nunca llegaron a popularizarse con carácter general. Hoy en día, DLT se encuentra normalmente como copia de seguridad de larga duración en entornos pequeños que no requieren mayor capacidad.

4611.1.1.2 Linear Tape Open (LTO)

Linear Tape Open (LTO) fue diseñado y concebido como una evolución y alternativa a los formatos DLT y otros ya existentes, y estaba destinado a proporcionar una plataforma común para los backups en cinta.

Seagate, HP e IBM fueron los iniciadores originales del consorcio LTO, encargado de realizar el desarrollo inicial y el cuál mantiene la licencia de la tecnología y la certificación del proceso. En teoría, se debería de haber producido un formato estándar de cinta, con el cual los fabricantes podrían seguir trabajando con el estándar en el mercado e incorporando sus propias características y funciones adicionales.

Sin embargo, entre el original LTO-1 y los formatos de LTO-2 hubo problemas de compatibilidad. Estos problemas abarcaban desde bloqueos en las cintas cuando se utilizan medios adquiridos a dos proveedores distintos hasta la incapacidad de una unidad LTO de un fabricante a leer los datos escritos en un cartucho de otra.

El LTO-1 inicial proporcionaba 100 GB de almacenamiento nativo y 15 MB /seg; con los actuales sistemas de LTO-4 se proporcionan 400 GB de almacenamiento nativo de 160 MB / seg. Por su parte, el LTO-5 proporciona 800 GB de capacidad de almacenamiento nativo a 160 MB / seg.

4611.1.1.3 Sun StorageTek T10000 (T10k)

El T10000 / StorageTek (T10k) de Sun representa uno de las tecnologías de almacenamiento en cinta que mejor se ha comportado en términos de capacidad. El T10k es un formato propietario producido únicamente por

StorageTek y se encuentra normalmente en entornos en los que se empleaban las tecnologías anteriores de Sun como el STK (9840/9940). También se han utilizado en sistemas abiertos de servidores o mainframe. El T10k está diseñado para 500 GB de almacenamiento nativo de 120 MB / seg.

4611.1.1.4 Características de almacenamiento en cinta

Aunque todos los datos anteriores indican un valor interesante en cuanto al rendimiento, todos los dispositivos de cinta con características similares de rendimiento deben tenerse en cuenta a la hora de diseñar entornos de backup.

La primera y más importante de ellas es el hecho de que todas las unidades de cinta son entornos serie. A diferencia de los dispositivos de disco, los dispositivos de cinta escriben los bloques de datos de forma lineal, uno tras otro. Las unidades de cinta sólo tienen una cabeza de escritura que escribe un bloque de datos de cada vez en la cinta, a medida que ésta se mueve por ella. Los dispositivos de disco tienen una serie de dispositivos de escritura, o cabezas, que se mueven a varios puntos del disco giratorio para situar los datos de una manera óptima. Esto permite que los dispositivos de disco puedan leer cualquier trozo de información solicitada. Dado que los discos tienen varias cabezas para obtener bloques de datos en paralelo, varios sistemas pueden acceder al disco al mismo tiempo

La lectura de los datos de una cinta, se realiza mediante el proceso inverso: La cinta debe rebobinarse hasta el principio, hacia adelante hasta bloque que se necesita, y leer así el bloque de datos. Al poder devolverse únicamente un segmento de datos con cada lectura, los dispositivos de

cinta no se pueden compartir de forma paralela entre sistemas sin un mecanismo para transferir el control entre los sistemas que usan dicho dispositivo.

El tipo de conectividad también tiene influencia sobre la utilización de dispositivos de cinta. Las unidades de cinta dependen de una conexión directa con el host para el transporte de los datos. Una vez más, esto se debe al hecho de que las unidades de cinta son dispositivos de serie que sólo aceptan una sola conexión a la vez.

4611.1.2 Disco

La cinta proporciona un método muy maduro, muy conocido, y de bajo coste para almacenar copias de seguridad. Sin embargo, las debilidades, tales como la naturaleza secuencial de la cinta, la complejidad mecánica, y la gran variabilidad del rendimiento de los dispositivos de cinta están rápidamente relegando a la cinta a medio de almacenamiento secundario o terciario en muchos entornos.

Con todos los problemas con la cinta, los administradores buscaban un medio que permitiera un rápido acceso a las copias de seguridad y que proporcionase una forma de tener un almacenamiento rápido y fiable: *el disco.*

Los backup a disco son simples sistemas de archivos que han sido situados aparte para que el software de backup los use. Aunque esto parece sencillo, la implementación y gestión de las soluciones basadas en disco pueden ser muy complejas.

El almacenamiento en disco supera algunas de las desventajas propias de las cintas. Por la capacidad de recibir datos de forma rápida, tiene múltiples flujos para almacenar copias de seguridad al mismo tiempo, y tiene la capacidad de presentar el almacenamiento de un número de maneras diferentes, dependiendo de la necesidad del sistema, por eso, el disco es muy empleado como un medio de almacenamiento de copia de seguridad

primario.

Pero el disco también tiene sus debilidades, el coste de los medios de comunicación, la falta de portabilidad, y la dificultad de asegurar la plena utilización de los medios de comunicación hacen que el disco no sea tan satisfactorio como parece a priori.

4611.1.3 Medios Virtuales

Los medios virtuales emulan el hardware físico de cinta con el objetivo de reducir o eliminar los problemas de gestión asociados a los medios físicos. Mediante la eliminación del hardware con una alta complejidad mecánica y de gestión y la eliminación de sus sistemas asociados y reemplazándolos por unidades de disco, los medios virtuales también tiene la ventaja de aumentar la fiabilidad general del entorno de backup. Los medios virtuales ofrecen estas ventajas sin cambiar los procedimientos operativos o exigir modificaciones del software de copia de seguridad. Además, en algunos casos, el rendimiento puede aumentarse a través de un mejor uso del ancho de banda en los medios de comunicación utilizados para conectar los medios virtualizados con los servidores de backup.

Los Medios virtuales de copia de seguridad se asocian tradicionalmente de forma exclusiva con bibliotecas de cintas virtuales (VTL) pero recientemente se han realizado nuevas implementaciones a través de protocolos que permiten la virtualización de otros tipos de sistemas de almacenamiento.

4611.1.4 Medios Ópticos

Los medios ópticos se sitúan entre las ventajas de las cintas y las del disco.

Sobresalen en las áreas de fiabilidad, flexibilidad, ciclo de trabajo e inamovilidad, mientras que sus retos los encontramos en las áreas de rendimiento, capacidad y coste.

4611.1.4.1 CD

CD, o compact disk, es un soporte digital óptico que se utiliza para el almacenamiento de prácticamente cualquier tipo de datos. En la actualidad el uso del CD está decayendo a favor del aumento del uso de un nuevo medio de similares características como el DVD.

El CD ha servido y sigue sirviendo como medio de almacenamiento de copias de seguridad gracias a su fiabilidad e inamovilidad. Proporciona en comparación con otro medios como la cinta magnética, mayor seguridad y protección de los datos, dado que el propio medio es mucho más robusto frente a interacciones físicas externas (por ejemplo los campos magnéticos).

Además de ser un medio habitual para el almacenamiento de pistas de audio, los CDs se utilizan habitualmente para la generación de copias de seguridad relacionadas con la recuperación de los sistemas.

Los sistemas de CD utilizan un dispositivo hardware específico para grabar información, conocido como grabadora/regrabadora de CD. Existen también dispositivos hardware similares que solamente permiten la lectura de este medio.

Las capacidades habituales de los CD estándar abarcan desde los 650MB hasta los 900MB.

4611.1.4.2 DVD

Los DVDs vienen a ser la evolución de la tecnología digital óptica de los CDs.

Al igual que los CDs, existen dos tipos de dispositivos para el uso de los DVDs que son las grabadoras y los lectores. Existen diferentes tipos de DVDs y diferentes categorizaciones, siendo la más importante la relativa al número de capas, factor que determina la capacidad final del dispositivo.

Las capacidades actuales abarcan desde los 4,3Gb hasta los 17Gb. Los DVD utilizan dos tipos de sistemas de ficheros que reemplazas el antiguo ISO 9660 de los CDs, y que son el UDF y el Joliet.

4611.2 Software de Backup

En la categoría de elementos software de backup tenemos herramientas de código abierto o software libre y software privativo o comercial. Las herramientas más comunes a nivel de software son:

4611.2.1 Herramientas de código abierto – AMANDA

Amanda (Advanced Maryland Automated Network Disk Archiver), es el software de código abierto de copia de seguridad más conocido. Amanda se desarrolló inicialmente en la Universidad de Maryland en 1991 con el objetivo de proteger los archivos de un gran número de estaciones de trabajo cliente con un servidor de copia de seguridad único. James da Silva fue uno de sus desarrolladores originales.

El proyecto Amanda se registró en SourceForge.net en 1999. Jean-Louis Martineau, de la Universidad de Montreal ha sido el líder del desarrollo de Amanda en los últimos años. Durante años, más de 250 desarrolladores han contribuido al código fuente de Amanda, y miles de usuarios aportan pruebas y comentarios, lo que lo convierte en un paquete robusto y estable. Amanda se incluye con la mayor parte de las distribuciones Linux.

En un principio, Amanda fue utilizado mayoritariamente en las universidades, laboratorios técnicos, y departamentos de investigación. Hoy, con la amplia adopción de Linux en los departamentos de informática, Amanda se encuentra en muchos otros lugares, sobre todo cuando la atención se centra en aplicaciones LAMP (Linux+Apache+MySQL+PHP). Con los años, Amanda ha recibido múltiples premios de los usuarios.

Amanda permite configurar un único servidor backup maestro para realizar múltiples copias de seguridad de equipos Linux, Unix, Mac OS X, y Windows en una amplia variedad de dispositivos: cintas, discos, dispositivos ópticos, bibliotecas de cintas, sistemas RAID, dispositivos NAS, y muchos otros.

Las principales razones para la adopción generalizada de Amanda son:

- Se puede configurar un único servidor de copia de seguridad de varios clientes en red con cualquier dispositivo de almacenamiento: una cinta, disco o sistema de almacenamiento óptico.
- Está optimizado para el backup en disco y cinta, permitiendo escribir simultáneamente backup a cinta y disco.
- No utiliza drivers propietarios, cualquier dispositivo soportado por un sistema operativo también podrá funcionar en Amanda.
- Utiliza herramientas estándar, como dump y tar. Puesto que no son formatos propietarios, los datos se pueden recuperar con esas mismas herramientas.
- Se utiliza un planificador que optimiza niveles de seguridad para los diferentes clientes, de tal manera que el tiempo total del backup es aproximadamente el mismo para cada ejecución.
- Existe una amplia y activa comunidad de usuarios que crece día a día.

- El coste total de propiedad (TCO) de una solución de backup basada en Amanda es significativamente menor que el TCO de cualquier solución que utilice software privativo.

4611.2.2 Herramientas de código abierto – BackupPC

BackupPC es un sistema de alto rendimiento que permite realizar copias de seguridad de sistemas Unix, Linux, Windows y MacOS en un disco. Es por tanto una herramienta basada totalmente en disco.

Ofrece una serie de ventajas como son:

- **Soporta cualquier sistema operativo cliente.** Esto se debe a que se utilizan herramientas estándar que o vienen con el SO o se pueden añadir al SO, sin necesidad de instalar cliente. Así resulta más fácil integrar un nuevo cliente.

- **Interfaz Web** con control de usuario para acceder a copias de seguridad. La mayoría de los SO trae un navegador web, así que usar una interfaz web es otra manera de acelerar el proceso de incorporación de nuevos clientes con diferentes sistemas operativos. La interfaz web está diseñada para dar el máximo control posible al cliente de forma segura. El usuario puede solicitar restauraciones, y navegar fácilmente y restaurar archivos individuales. Sin embargo, el usuario no podrá ver las máquinas de otro usuario.

- **Soporte de clientes DHCP**. Mediante el uso de servicios estándar, BackupPC soporta clientes DHCP, siempre y cuando el cliente esté registrado con un servicio de nombres como DNS, Active Directory o LDAP.

Funcionamiento de BackupPC

El modelo de BackupPC tiene un usuario por cliente. Esto es así porque

BackupPC fue específicamente diseñado para realizar copias de seguridad de PCs de varios usuarios (de ahí el nombre).

Normalmente, el usuario es el propietario de los datos de la máquina. Si se trabaja con un servidor de ficheros, el usuario deberá ser un administrador.

BackupPC envía mensajes de correo electrónico al propietario si no puede realizar la copia de seguridad después de un tiempo configurable; el propietario puede gestionar las restauraciones de las copias a través de una interfaz web.

En los siguientes puntos se describen algunas de las características proporcionadas por BackupPC:

- **Directo al disco**. BackupPC almacena todas sus copias de seguridad directamente en el disco. Los archivos idénticos en cualquier directorio o cliente se guardan sólo una vez, lo que reduce drásticamente los requisitos de almacenamiento del servidor. Estos archivos se almacenan en un conjunto de discos. Además del conjunto de discos, las copias de seguridad están en un árbol de directorios organizados por host.

  BackupPC también tiene un proceso (que se lanza por las noches) que recupera espacio del conjunto de discos que no está referenciado por ningún backup, lo que evita un uso inadecuado del espacio en disco. Este es un proceso automático que el administrador no tiene que configurar.

- **Sistema operativo del servidor** La parte del servidor de BackupPC está diseñada para ejecutarse en un sistema tipo Unix con Perl y mod_perl. Ofrece el mejor rendimiento con Apache, pero se puede ejecutar en cualquier servidor web que soporte Perl (se requiere mod_perl o Perl setuid.) El servidor debe tener un disco con gran capacidad o RAID para almacenar los backups.

- **Sistema operativo del cliente.** Como se comentó anteriormente, soporta cualquier SO. Las versiones más modernas de las variantes

comerciales de Unix (Solaris, AIX, IRIX,HP-UX) traen en la propia distribución las herramientas tar, compress, gzip, rsync, y rsh y / o ssh. Otros sistemas operativos tipo Unix (Linux, FreeBSD, OpenBSD, NetBSD, Mac OS X) también cuentan con estas herramientas.

Los clientes de Windows pueden hacer copias de seguridad de diferentes formas dependiendo de si las políticas locales permiten o no la instalación de software. Si no se permite, BackupPC utilizar parte de la suite Samba (http://www.samba.org) para hacer backup de la información compartida mediante SMB o CFIS. Si se permite instalar software, se utiliza rsync junto con el conjunto de herramientas Cygwin (http://www.cygwin.com).

- **Soporte para herramientas nativas**. BackupPC utiliza las herramientas estándar de Unix para su funcionamiento interno. Esto incluye programas como Perl, tar, rsync comprime, gzip, bzip2, zip, apache y samba.

  BackupPC no utiliza una base de datos o catálogo para almacenar la información de respaldo. En su lugar, utiliza el árbol de directorios para almacenar esta información. Esto simplifica las actualizaciones del sistema operativo del servidor de BackupPC o de la propia aplicación BackupPC.

- **Control de los backups y restauraciones a través de interfaz web**. La Web es la interfaz principal de BackupPC. Tras la configuración inicial, no es necesario acceder al servidor mediante línea de comandos para administrar BackupPC. La interfaz web está escrita en Perl y fue diseñada para funcionar tanto con mod_perl como con CGIs o con Perl setuid.

  La interfaz permite a los usuarios identificarse, acceder y controlar los respaldos y las restauraciones.

  El usuario puede solicitar copias de seguridad de tipo one-time, de tipo

completa, o de tipo incremental.

Se pueden utilizar varias opciones para recuperar ficheros:

- Los archivos individuales se recuperan mediante selección.
- Los grupos de archivos o directorios se pueden restaurar a su ubicación original.
- El usuario puede descargar los archivos como un archivo tar o zip.

El usuario tiene control absoluto sobre qué archivos o directorios se restauran y donde hay que restaurarlos. Un histórico muestra que archivos se han modificado durante cada copia de seguridad en cada directorio.

- **Soporte para clientes DHCP**. Los clientes BackupPC se referencian por nombre de host. Si la red de la copia de seguridad utiliza DHCP y se permite la resolución de nombres dinámica, no hay que hacer nada más para que el servidor BackupPC respalde a los clientes DHCP. Si este no es el caso, y los clientes son máquinas Windows, BackupPC se puede configurar para buscar un conjunto de direcciones de los clientes, localizándolos mediante SMB.

Si el cliente no está en línea durante el período de copia de seguridad normal, el servidor BackupPC no genera un error a menos que haya transcurrido un período de tiempo establecido desde la última copia de seguridad. En este punto, el servidor envía un email al propietario del cliente y le recuerda que se asegure que la máquina está en la red para hacer una copia de seguridad. (El servidor también puede enviar cualquier error al administrador.)

Los clientes que residen en otra LAN pueden ser gestionados a nivel local asumiendo que hay conectividad de entre las redes. Esto significa que se puede hacer backup de los clientes conectados a través de una red

privada virtual (VPN).

Si el usuario no desea realizar copias de seguridad en un momento dado, se conectaría a través de la interfaz web para cancelar la copia de seguridad.

- **Pool de Backups**. Cuando los clientes utilizan el mismo sistema operativo se duplican los archivos respaldados. Si se quiere mantener múltiples copias de seguridad completas aumenta el número de archivos duplicados, lo que aumenta los requisitos de capacidad de almacenamiento para el servidor. BackupPC almacena un árbol de directorios por cliente respaldado, pero comprueba si los archivos se han almacenado antes. Si es así, BackupPC utiliza un enlace que apunta al fichero existente en el conjunto de discos común, ahorrando una gran cantidad de espacio. Además, BackupPC puede comprimir opcionalmente para ahorrar más espacio.

- **Fácil configuración por cliente.** Una vez que el administrador haya definido cuáles deberían de ser las políticas de backup del sitio, es muy fácil anular cualquier opción de configuración en base a un cliente. Esto permite una gran flexibilidad sobre qué, cuándo, y cómo hacer copia de seguridad de un cliente. No hay clases de clientes por sí mismo.

4611.2.3 Herramientas de código abierto – Bacula

Bacula es un conjunto de programas Open Source, listos para ser utilizados en un entorno doméstico y profesional, que permiten administrar los backups, restauración y verificación de datos en una red heterogénea. Bacula es relativamente fácil de usar y eficiente, a la vez que ofrece muchas funcionalidades avanzadas para la administración de los datos almacenados, lo cual facilita hallar y recuperar archivos perdidos o

dañados. En términos técnicos, Bacula es un sistema de backups Open Source, orientado a la red y listo para la empresa.

Es capaz de realizar copias de seguridad en disco, cinta o medios ópticos. Bacula fue escrita originalmente por John Walker y Kern Sibbald en el año 2000. John dejó el proyecto no mucho tiempo después de su creación, y Kern, trabajó en él desde mediados del 2000 hasta el primer lanzamiento público de Bacula en abril de 2002. Desde entonces, otros desarrolladores han contribuido a su desarrollo.

Bacula está disponible bajo licencia AGPL versión 3. La página web del proyecto se encuentra en http://www.bacula.org, y los archivos descargables y un repositorio CVS se alojan en SourceForge.

Bacula Arquitectura

Bacula es una solución distribuida de backups. Esto significa que Bacula está compuesto por varios elementos, que pueden o no residir en el mismo host. Por ejemplo, se puede tener un host con el catálogo y en otro el storage.

Se basa en una arquitectura Cliente-servidor que resulta eficaz y fácil de manejar, dada la amplia gama de funciones y características que brinda: copiar y restaurar ficheros dañados o perdidos. Además, debido a su desarrollo y estructura modular, Bacula se adapta tanto al uso personal como profesional.

Se puede utilizar TLS (Transport Layer Security) para proteger los datos durante la transmisión.

Los componentes principales de esta arquitectura son:

- **Director (DIR)** es el encargado de gestionar de forma centralizada la lógica de los procesos de backup y los demás servicios. Trabaja en base a una unidad básica denominada JOB (un cliente, un conjunto de archivos, ...)

de tal forma que el Director planifica, inicia y supervisa todos los jobs.

También es el encargado de mantener el catálogo, por lo que el servidor de la base de datos debe estar accesible desde la máquina que ejecuta el Director.

- **Storage** es el encargado de gestionar los dispositivos de almacenamiento; esto exige que esté instalado en la máquina que posea la conexión física a los dispositivos de almacenamiento, tales como: discos locales, grabadoras, unidades de cinta, volúmenes NAS o SAN, autocargadores o librerías de cinta.

- **File Daemon** es el agente que corre del lado del cliente, es decir, en la máquina cuyos datos se van a respaldar, y que tiene como objetivo empaquetar los datos y enviarlos al Storage, donde serán almacenados.

- **Consola** es la herramienta que permite al usuario o administrador controlar Bacula. Se comunica con el director vía red, iniciando los jobs, revisando la salida del job, haciendo consultas y modificaciones en el catálogo.

  Existen consolas en modo texto, modo GUI para Windows y Linux/UNIX e interfaces web.

- **Catálogo** es una base de datos donde se guarda información sobre los jobs y sobre los datos respaldados. El catalogo permite dos cosas:

  - Por un lado, como guarda información de los jobs, pools y volúmenes, Bacula lo usa para saber si hay un backup completo para un job, y si no lo hay, realizará para ese backup una copia completa.

  - Por otro lado, el catálogo tiene todos los nombres de archivo (y sus atributos, como fecha de última modificación, etc.) que se respaldaron, y eso es lo que permite hacer una recuperación selectiva, es decir, seleccionar (marcar, en la jerga de Bacula) individualmente qué

archivos y/o directorios restaurar.

4611.2.4 Software Propietario CommVault Simpana

Simpana comenzó como un proyecto dentro de AT & T Labs en 1987 y posteriormente fue adquirido por la empresa CommVault.

Simpania es un software de backup que realiza copias de seguridad de entornos Unix, Windows, Linux, servidores de correo Exchange, Lotus Notes, bases de datos Oracle, MySQL, SQLServer y máquinas virtuales VMware. Además permite funciones avanzadas como puede ser el archivado, la deduplicación y la replicación de ficheros.

El funcionamiento de la aplicación se basa en el uso de los bloques de disco, por lo que todos los módulos no utilizan la información del archivo, si no que trabaja a más bajo nivel. Con ello consigue mejores ratios de compresión y una importante reducción de la ventana de backup, al utilizar únicamente los bloques modificados y no el fichero entero para realizar estas operativas.

Otra característica que incide en el uso de almacenamiento de bajo coste es la capacidad de generar políticas como las de archivado, mediante las que automáticamente permite mover ficheros de un almacenamiento a otro con mayor capacidad a menor coste. De esta forma, por ejemplo se podrían pasar los datos de una cabina de fibra a otra con discos SATA, pudiendo llegar a un tercer nivel a cinta, en base a unos requisitos (fecha del archivo, último acceso al archivo, etc.). Todos los movimientos se realizan de forma transparente para el usuario, tanto en el archivado como en su recuperación (si fuese necesario).

A estas funcionalidades hay que sumar la capacidad de deduplicación, que realiza una compresión de los datos aprovechando las duplicidades de los datos a nivel de bloque, consiguiendo alcanzar ratios de hasta el 50% de ahorro en el uso de almacenamientos en datos de segundo nivel y hasta el 90% en los de tercer nivel.

Para terminar el repaso a las principales funcionalidades, la replicación, permite la utilización de snapshots a nivel de cabina permitiendo volver el almacenamiento replicado a un estado anterior o montar la imagen snapshot como un recurso compartido.

Todo se administra desde una única consola centralizada, que simplifica toda la administración de la plataforma. Adicionalmente el motor de búsqueda ofrece la opción de buscar rápidamente y recuperar datos sin necesidad de saber donde se ubican.

4611.2.5 Software Propietario Symantec NetBackup

Symantec NetBackup es actualmente el titular de la mayor cuota de mercado del entorno de software de copia de seguridad.

Netbackup 7 es la nueva versión de la solución de copia de seguridad y recuperación de datos orientada a grandes corporaciones. Esta herramienta trata de simplificar la gestión de la información reduciendo el volumen de almacenamiento de datos con técnicas de deduplicación en los ordenadores cliente de la red además del propio servidor, ofreciendo protección para entornos virtualizados. Todo ello con el único propósito de agilizar los procesos de backup y recuperación de datos.

La nueva herramienta incluye eliminación de datos duplicados nativos dentro del cliente NetBackup y permite a los clientes multiplicar por diez la velocidad de las copias de seguridad en oficinas remotas, el propio centro de datos y los entornos virtuales. Esta eliminación de datos duplicados en el cliente y en el destino ofrece una mayor cobertura con menos herramientas.

El proceso de deduplicación se contempla para todos los sistemas físicos y virtuales, independientemente del método de copia de seguridad. De este modo se integra una mayor protección para los cada vez más extendidos entornos virtualizados bajo las plataformas Hyper-V y VMware. Es en el caso de esta última en la que se ha podido observar un incremento de

velocidad de hasta el 50% a la vez que disminuye el volumen de almacenamiento necesario en un 40%.

Otro de los aspectos notablemente mejorados en Netbackup 7 es la velocidad de recuperación de datos ante desastres. Permitiendo la restauración de grandes volúmenes de información en pocos segundos desde cualquier lugar y punto en el tiempo. Esta gestión se facilita al administrador de TI mediante un sistema centralizado de supervisión y alerta, que integra la administración de varios dominios de archivos con sus respectivas políticas de salvaguarda de datos.

La tecnología incluida en NetBackup acelerará la transición a un entorno virtual para las organizaciones empresariales que instalen un gran número de máquinas virtuales o que decidan crear una infraestructura de nube privada.

La solución NetBackup también ofrece una elaboración de informes simplificada y un mayor soporte a las aplicaciones de bases de datos de Oracle y MySQL.

Algunas de las prestaciones y beneficios incluidos en la última versión de la herramienta son:

- La tecnología Virtual Machine Intelligent Policy incorpora la automatización a la localización y la protección de máquinas virtuales y minimiza los esfuerzos de administración necesarios para realizar copias de seguridad de máquinas virtuales VMware de alto rendimiento.

- Un 50% más de rapidez en copias de seguridad de máquinas virtuales gracias a que la tecnología Granular Recovery Technology (GRT) se encuentra ahora disponible para sistemas Linux en entornos VMware. Esto permite a los clientes reducir los tiempos comparables de copias de seguridad de máquinas virtuales en un 50%, además de simplificar la administración y mejorar la velocidad de recuperación de

archivos individuales.

- Recuperación “a la carta” desde cualquier lugar con la nueva tecnología de replicación de imagen que permite a los clientes que replican datos entre múltiples sitios o dominios de NetBackup realizar backup de datos en un sitio alternativo.

- Recuperación acelerada: NetBackup RealTime ofrece soporte a entornos VMware para eliminar el espacio de tiempo entre copias de seguridad, además de reducir el impacto para grandes hosts de VMware y permitir la recuperación casi instantánea de sistemas completos.

- Satisfacer los requisitos normativos y de cumplimiento para seguimiento de auditorías.

- Incorpora informes mejorados de las políticas del ciclo de vida del almacenamiento, del seguimiento de las auditorías y del estado de las licencias.

- Deduplicación para Oracle mejorando el rendimiento de las copias de seguridad.

- Se añade un nuevo agente que presta soporte a MySQL para centralizar y automatizar las copias de seguridad y la recuperación de datos de las bases de datos de MySQL.

- Actualización simplificada de clientes con LiveUpdate que permite mejoras en equipos cliente para UNIX, Linux y Windows respecto a la versión NetBackup 6.5 y posterior desde una política única controlada por el administrador de NetBackup.

4612 ESTRATEGIAS DE BACKUP A DISCO

Las estrategias de backup definen el plan que se ha de seguir para garantizar la integridad de la información. Los motivos por los que se debe establecer una correcta estrategia antes de comenzar a realizar las copias de seguridad pueden ser muy diversos, pero en esencia se trata de determinar la mejor manera para asegurar la información teniendo en cuenta las posibles dificultades de recuperación de parte de los datos, el coste de los medios que se emplearan y el tiempo que se necesitara.

Como no todos los sistemas son iguales, no todas las estrategias de backup son adecuadas para todos los sistemas. Partiendo de unas características comunes, algunas de las propiedades básicas de una estrategia backup son:

- Tiempo de almacenamiento. Define el tiempo máximo que una copia permanece almacenada en un dispositivo. Al finalizar este tiempo la copia puede cambiar de dispositivo o ser borrada para liberar espacio en el medio de almacenamiento y poder hacer uso del mismo.

- Almacenamiento alternativo. Posibilita realizar una o varias copias de seguridad en una ubicación externa al sistema y a la localización geográfica del mismo, manteniéndola durante un elevado período de tiempo, aumentando la seguridad ante cualquier catástrofe, ya sea a nivel de software o de hardware.

- Protección ante fallo de los dispositivos. Establece el número de medios que se emplean. Cuanto mayor es el número de medios utilizados, mayor es la seguridad contra posibles pérdidas de información producidas por un fallo en el dispositivo de almacenamiento.

- Tiempo de restauración. Esta característica especifica el tiempo de regeneración del sistema en caso de producirse algún fallo.
- El coste. Suele ser un factor determinante a la hora de seleccionar la estrategia a realizar.

Las estrategias para la realización de copias de seguridad pueden ser muy distintas, dependiendo del sistema en cuestión sobre el cual se realizan.

En algunos casos, solamente se efectúa un backup de todo el contenido. Esto se produce cuando por algún motivo especial y muy específico o por algún motivo técnico, cuestiones de tiempo o por que existe un elevado riesgo para los datos. Alguno de estos casos especiales pueden ser:

- No disponer del software original.
- Desconocimiento de la ubicación de los ficheros de configuración.
- Cambiar un disco de almacenamiento rígido.
- Realizar cambios en las particiones de uno o más discos de almacenamiento rígidos.

Es habitual que este tipo de situaciones concretas se produzcan a la hora de llevar a cabo tareas de reparación o actualización sobre sistemas no controlados.

Cuando se trata de cubrir alguno de estos casos la estrategia de backup a seguir es sencilla, realizar un resguardo o copia de seguridad de todo el contenido de las unidades involucradas para así garantizar que no se perderá ninguna información y que será posible realizar la restauración completa del sistema.

Por otro lado, cuando realmente se ha de diseñar un plan estratégico para la realización de las copias de seguridad de un sistema propio o de una organización externa, se deben tener en cuenta una serie de pautas que

ayudan a que el plan estratégico de backups sea el más conveniente y conseguir la mejor relación coste/beneficio posible.

Estas pautas aportan una reducción en el tiempo de respuesta a la hora de realizar una recuperación en caso de que se producirse cualquier tipo de contingencia.

Al intentar definir un plan de backups, surgen una serie de dudas:

¿Qué datos se deberían resguardar en cada backup? Datos a resguardar.

Es un factor determinante para una estrategia de backup que se determine el grado o grados de importancia de la información, es decir, establecer que información resulta de mayor valor para la organización. No tienen la misma transcendencia un documento de trabajo que una copia de respaldo de la configuración de una aplicación.

¿Cada cuánto se debería efectuar un backup de los datos? Frecuencia del backup.

Para determinar la periodicidad con la que se deben realizar las copias de seguridad no existe un criterio claramente definido. Sin embargo si se tienen en cuenta factores como:

- Tiempo empleado en la creación de la información.
- Coste invertido en la creación de la información.
- Posibles consecuencias derivadas de su pérdida.

¿Cuánto tiempo deberían permanecer guardadas las copias de seguridad? Tiempo de Almacenamiento.

El período máximo de tiempo de estancia de una copia de seguridad en un dispositivo, es decir, el tiempo de retención, está directamente relacionado con los medios de almacenamiento disponibles, y por consiguiente por el

presupuesto de la estrategia de backup.

Otra de las decisiones importantes a tomar durante la elaboración de una estrategia para la realización de copias de seguridad es la de seleccionar y planificar los distintos tipos de copias de seguridad.

Los backups son copias exactas de la información. Se pueden definir como instantáneas de los datos en un momento determinado, almacenados en un formato estándar, se puede realizar un seguimiento a lo largo de su periodo de utilidad y con cada nueva copia se mantiene la independencia con copia inicial. Se pueden crear múltiples niveles de backups, siendo los principales:

- ***Copias de seguridad completas (Full backups):*** representan una copia exacta en un momento dado, de los datos que se pretende proteger. Proporcionan la base para todos los demás niveles de backup.

- Por otro lado, están dos niveles de backup que capturan únicamente los cambios realizados sobre una copia de seguridad completa.

o ***Copia de seguridad diferencial***, también conocida como la *copia de seguridad incremental acumulativa*, captura copias de seguridad que se han producido desde el último backup completo y suele utilizarse en entornos en los que no se produce un elevado número de cambios. La copia de seguridad diferencial se debe utilizar con cuidado debido a que puede crecer con rapidez e igualar e incluso superar el tamaño de la copia de seguridad completa.

La ventaja de utilizar las copias de seguridad diferenciales viene dada en el momento de la restauración puesto que en el momento de restaurar una copia de seguridad diferencial sólo se necesita el backup completo y la última copia diferencial realizada. Debido a que únicamente se precisan dos imágenes para la restauración,

la probabilidad de que ambas imágenes sufran algún percance, perdida, corrupción, etc., se reduce significativamente.

- ***Copia de seguridad incremental***, es capaz de capturar los cambios que se han producido desde la última copia de seguridad realizada, independientemente del tipo que sea. Es la forma más utilizada para la realización de copias de seguridad, evidentemente combinada con una copia de seguridad completa.

  Este tipo de copia de seguridad contiene la menor cantidad de datos necesarios durante cada ciclo de backup, reduciendo la cantidad de datos que se transfieren y el tiempo que se necesita para la creación de una copia de seguridad.

  Sin embargo las copias de seguridad incrementales tienen aspectos negativos. Si se está recuperando un grupo de archivos de un conjunto de copias de seguridad completas e incrementales, es probable que se requieran más de dos imágenes de copias de seguridad diferentes para completar la restauración, lo que aumenta la probabilidad de que alguna de estas partes sufra algún tipo de problema y no se pueda completar la restauración.

4613 REPLICACIÓN LOCAL Y REMOTA, ESTRATEGIAS DE RECUPERACIÓN

La replicación es el proceso de creación de una copia exacta de los datos. La creación de una o varias réplicas de los datos de producción es una de las maneras de proporcionar continuidad al del negocio (BC).

Estos modelos pueden ser utilizados para operaciones de recuperación y reinicia de los sistemas en caso de que se produzca una pérdida de datos.

Una réplica ha de proporcionar:

- **La capacidad de recuperación:** permite la restauración de los datos de los volúmenes de producción en caso de que se produzca una pérdida de los datos. Se ha de proporcionar un mínimo de y RTO y un RPO concreto que nos garanticen la reanudación de las operaciones comerciales en los volúmenes de producción.

- **La capacidad de reinicio:** garantiza la coherencia de los datos de la réplica, posibilitando la reanudación de las operaciones de negocio utilizando para ello la información contenida en las réplicas.

La replicación se pueden clasificar en dos grandes categorías: ***locales y remotos***

4613.1 Replicación Local

La replicación local hace referencia al proceso creación de réplicas dentro del mismo array de discos o el mismo centro de datos.

4613.1.1 Tecnologías de replicación local

Las replicaciones Host-based (basadas en replicación en host local) y Storage-based (basadas en almacenamiento) son las dos principales tecnologías adoptadas para la replicación local. La replicación de archivos del sistema y la replicación basada en LVM son ejemplos de la tecnología Host-based de replicación local. La replicación de almacenamiento basada en matrices de disco puede llevarse a cabo con soluciones distintas, la duplicación de todo el volumen, la replicación pointer-based de todo el volumen, y la replicación basadas en punteros y virtual.

4613.1.1.1 Basada en replicación en host local

En este tipo de replicación, los administradores del sistema llevan a cabo el proceso de copia y restauración en la propia máquina, pudiendo basarse la recuperación en una replicación integral del volumen mediante LVM (Logical Volume Manager), o bien mediante instantáneas del sistema de ficheros.

- Replicación del volumen mediante LVM: El LVM se encarga de crear y controlar el volumen de host a nivel lógico y está formado por tres componentes: los discos físicos, los volúmenes lógicos y los grupos de volúmenes. En la replicación de volúmenes basado en LVM, cada partición lógica en un volumen se asigna a dos particiones físicas en dos discos diferentes. De esa forma se consigue un espejo que permite redundancia y recuperación directa en caso de necesitar replicar.

- Instantánea de archivos del sistema: Consiste en crear una réplica a base de instantáneas del sistema de ficheros mediante la utilización de metadatos almacenados en un mapa de bits. Estos metadatos van

reflejando el cambio que se va produciendo en el sistema de ficheros y van almacenando un registro de las direcciones accedidas mediante operaciones de lectura/escritura. Este sistema requiere de una fracción del espacio utilizado por el sistema de ficheros original.

4613.1.1.2 Basada en arrays de discos

En este tipo de replicación se hace uso de matrices de discos que pueden estar distribuidas dentro del CPD. El entorno operativo es el que lleva a cabo el proceso de replicación de un determinado sistema de ficheros, sin necesidad de que los recursos de acogida (CPU y memoria) del anfitrión intervengan en el proceso de replicación.

4613.2 La replicación remota

La replicación remota consiste en el proceso de creación de replicas del conjunto de datos en lugares con otra ubicación física. Las réplicas remotas ayudan a las organizaciones a mitigar los riesgos asociados a las interrupciones regionales del servicio, que pueden estar provocadas por diferentes causas, por ejemplo, desastres naturales. La infraestructura en la que los datos se almacenan inicialmente se llama fuente. La réplica, o infraestructura remota en la que se almacena la copia se le llama blanco.

4613.2.1 Tecnologías de replicación remota

La más habitual es la tecnología de replicación basada en host remoto, que utiliza uno o más componentes de la máquina para realizar y gestionar la operación de replicación. Existen dos enfoques fundamentales para la replicación basada en host remoto: Replicación remota basada en LVM y replicación de bases de datos a través de trasvase de registros.

4613.2.1.1 Replicación remota basada en LVM

En este modelo, la replicación se efectúa y gestiona a nivel de grupo de volúmenes. El LVM de la máquina origen es el encargado de gestionar y transmitir la información del volumen al LVM de la máquina remota. El LVM de la máquina remota se encarga de recibir los datos y realiza la operación de réplica del volumen.

Antes del inicio de la replicación, se deben configurar los sistemas fuente y remoto para que los sistemas de archivos, los volúmenes y la agrupación de volúmenes sea idéntica en ambos. El punto de partida, o sincronización inicial, se puede realizar de diferentes formas, siendo la más frecuente la restauración en el punto remoto de una copia de seguridad de los datos de origen.

En la replicación remota basada en LVM se soportan dos modos de transferencia de datos, que son el sincrónico y el asincrónico. En el modo asíncrono, las operaciones de escritura se van almacenando en una cola de registros gestionada por el LVM y se van enviando al host remoto en el orden en el que son recibidas. En caso de fallo de la red, las operaciones siguen acumulándose en la cola de registros.

En la replicación síncrona, las operaciones de escritura deben estar comprometidas tanto en origen como en destino. Las operaciones de escritura consecutivas no pueden ocurrir en fuente ni destino hasta que las operaciones previas hayan finalizado. Esto garantiza que los datos de la fuente y destino son exactamente los mismos en todo momento. Esto hace posible que el RPO en caso de fallo sea cero o cercano a cero. Sin embargo, como contraprestación al nivel de seguridad, el tiempo de respuesta es mucho mayor. El grado de impacto en el tiempo de respuesta depende de la distancia entre ambos sitios (fuente y destino), del ancho de banda disponible y de la infraestructura de conectividad de red.

4613.2.1.2 Basada en trasvase de registros

La replicación de bases de datos a través de trasvase de registros consiste en la captura de las transacciones realizadas en la base de datos fuente, que son almacenadas en registros que se transmiten periódicamente de un host fuente a un host destino. El host destino recibe el conjunto de registros y realiza las operaciones oportunas en la base de datos replicada. El proceso inicial de producción y reproducción requiere que todos los componentes importantes de la base de datos se repliquen en el sitio remoto.

Los sistemas gestores de bases de datos permiten definir un intervalo de tiempo para el envío de los ficheros de registro, o bien configurar un tamaño predeterminado de los mismos. Cuando un registro supera el intervalo de tiempo establecido o alcanza su tamaño máximo, se cierra, y se abre un nuevo fichero para registrar las transacciones. Los registros cerrados van siendo enviados desde la fuente al destino garantizando que la base de datos replicada en destino sea consecuente con la fuente hasta el último registro cerrado. El RPO en el sitio remoto dependerá del tamaño del fichero de registro y de la frecuencia de cambio de registro en la fuente.

4614 REPLICACIÓN LOCAL Y REMOTA, ESTRATEGIAS DE RECUPERACIÓN

La replicación es el proceso de creación de una copia exacta de los datos. La creación de una o varias réplicas de los datos de producción es una de las maneras de proporcionar continuidad al del negocio (BC).

Estos modelos pueden ser utilizados para operaciones de recuperación y reinicia de los sistemas en caso de que se produzca una pérdida de datos.

Una réplica ha de proporcionar:

- **La capacidad de recuperación:** permite la restauración de los datos de los volúmenes de producción en caso de que se produzca una pérdida de los datos. Se ha de proporcionar un mínimo de y RTO y un RPO concreto que nos garanticen la reanudación de las operaciones comerciales en los volúmenes de producción.

- **La capacidad de reinicio:** garantiza la coherencia de los datos de la réplica, posibilitando la reanudación de las operaciones de negocio utilizando para ello la información contenida en las réplicas.

La replicación se pueden clasificar en dos grandes categorías: ***locales y remotos***

4614.1 Replicación Local

La replicación local hace referencia al proceso creación de réplicas dentro del mismo array de discos o el mismo centro de datos.

4614.1.1 Tecnologías de replicación local

Las replicaciones Host-based (basadas en replicación en host local) y Storage-based (basadas en almacenamiento) son las dos principales tecnologías adoptadas para la replicación local. La replicación de archivos del sistema y la replicación basada en LVM son ejemplos de la tecnología Host-based de replicación local. La replicación de almacenamiento basada en matrices de disco puede llevarse a cabo con soluciones distintas, la duplicación de todo el volumen, la replicación pointer-based de todo el volumen, y la replicación basadas en punteros y virtual.

4614.1.1.1 Basada en replicación en host local

En este tipo de replicación, los administradores del sistema llevan a cabo el proceso de copia y restauración en la propia máquina, pudiendo basarse la recuperación en una replicación integral del volumen mediante LVM (Logical Volume Manager), o bien mediante instantáneas del sistema de ficheros.

- Replicación del volumen mediante LVM: El LVM se encarga de crear y controlar el volumen de host a nivel lógico y está formado por tres componentes: los discos físicos, los volúmenes lógicos y los grupos de volúmenes. En la replicación de volúmenes basado en LVM, cada partición lógica en un volumen se asigna a dos particiones físicas en dos discos diferentes. De esa forma se consigue un espejo que permite redundancia y recuperación directa en caso de necesitar replicar.

- Instantánea de archivos del sistema: Consiste en crear una réplica a base de instantáneas del sistema de ficheros mediante la utilización de metadatos almacenados en un mapa de bits. Estos metadatos van

reflejando el cambio que se va produciendo en el sistema de ficheros y van almacenando un registro de las direcciones accedidas mediante operaciones de lectura/escritura. Este sistema requiere de una fracción del espacio utilizado por el sistema de ficheros original.

4614.1.1.2 Basada en arrays de discos

En este tipo de replicación se hace uso de matrices de discos que pueden estar distribuidas dentro del CPD. El entorno operativo es el que lleva a cabo el proceso de replicación de un determinado sistema de ficheros, sin necesidad de que los recursos de acogida (CPU y memoria) del anfitrión intervengan en el proceso de replicación.

4614.1.2 La replicación remota

La replicación remota consiste en el proceso de creación de replicas del conjunto de datos en lugares con otra ubicación física. Las réplicas remotas ayudan a las organizaciones a mitigar los riesgos asociados a las interrupciones regionales del servicio, que pueden estar provocadas por diferentes causas, por ejemplo, desastres naturales. La infraestructura en la que los datos se almacenan inicialmente se llama fuente. La réplica, o infraestructura remota en la que se almacena la copia se le llama blanco.

4614.1.3 Tecnologías de replicación remota

La más habitual es la tecnología de replicación basada en host remoto, que utiliza uno o más componentes de la máquina para realizar y gestionar la operación de replicación. Existen dos enfoques fundamentales para la replicación basada en host remoto: Replicación remota basada en LVM y replicación de bases de datos a través de trasvase de registros.

4614.1.3.1 Replicación remota basada en LVM

En este modelo, la replicación se efectúa y gestiona a nivel de grupo de

volúmenes. El LVM de la máquina origen es el encargado de gestionar y transmitir la información del volumen al LVM de la máquina remota. El LVM de la máquina remota se encarga de recibir los datos y realiza la operación de réplica del volumen.

Antes del inicio de la replicación, se deben configurar los sistemas fuente y remoto para que los sistemas de archivos, los volúmenes y la agrupación de volúmenes sea idéntica en ambos. El punto de partida, o sincronización inicial, se puede realizar de diferentes formas, siendo la más frecuente la restauración en el punto remoto de una copia de seguridad de los datos de origen.

En la replicación remota basada en LVM se soportan dos modos de transferencia de datos, que son el sincrónico y el asincrónico. En el modo asíncrono, las operaciones de escritura se van almacenando en una cola de registros gestionada por el LVM y se van enviando al host remoto en el orden en el que son recibidas. En caso de fallo de la red, las operaciones siguen acumulándose en la cola de registros.

En la replicación síncrona, las operaciones de escritura deben estar comprometidas tanto en origen como en destino. Las operaciones de escritura consecutivas no pueden ocurrir en fuente ni destino hasta que las operaciones previas hayan finalizado. Esto garantiza que los datos de la fuente y destino son exactamente los mismos en todo momento. Esto hace posible que el RPO en caso de fallo sea cero o cercano a cero. Sin embargo, como contraprestación al nivel de seguridad, el tiempo de respuesta es mucho mayor. El grado de impacto en el tiempo de respuesta depende de la distancia entre ambos sitios (fuente y destino), del ancho de banda disponible y de la infraestructura de conectividad de red.

4614.1.3.2 Basada en trasvase de registros

La replicación de bases de datos a través de trasvase de registros consiste

en la captura de las transacciones realizadas en la base de datos fuente, que son almacenadas en registros que se transmiten periódicamente de un host fuente a un host destino. El host destino recibe el conjunto de registros y realiza las operaciones oportunas en la base de datos replicada. El proceso inicial de producción y reproducción requiere que todos los componentes importantes de la base de datos se repliquen en el sitio remoto.

Los sistemas gestores de bases de datos permiten definir un intervalo de tiempo para el envío de los ficheros de registro, o bien configurar un tamaño predeterminado de los mismos. Cuando un registro supera el intervalo de tiempo establecido o alcanza su tamaño máximo, se cierra, y se abre un nuevo fichero para registrar las transacciones. Los registros cerrados van siendo enviados desde la fuente al destino garantizando que la base de datos replicada en destino sea consecuente con la fuente hasta el último registro cerrado. El RPO en el sitio remoto dependerá del tamaño del fichero de registro y de la frecuencia de cambio de registro en la fuente.

4615 BIBLIOGRAFÍA

- Windows Server 2008 Hyper-V : kit de recursos. Larson, Robert. Anaya, D.L. 2009
- Grid computing : experiment management, tool integration, and scientific workflows. Prodan, Radu Berlin: Springer, cop. 2007
- Virtualización na Wikipedia: http://es.wikipedia.org/wiki/Virtualizaci%C3%B3n
- Green IT: Tecnologías para la Eficiencia Energética en Sistemas TI.

Marisa López-Vallejo, Eduardo Huedo Cuesta y Juan Garbajosa Sopeña.

- Dot-cloud : the 21st century business platform built on cloud computing. Fingar, Peter Tampa (FL) : Meghan-Kiffer Press, cop. 2009
- System & Disaster Recovery Planning. Richard Dolewski
- Information Storage and Management: Storing, Managing, and Protecting Digital Information. G. Somasundaram y Alok Shrivastava.
- Backup & Recovery. W. Curtis Preston y O'Reilly Media.
- Redes de área de Almacenamento na Wikipedia. http://es.wikipedia.org/wiki/Red_de_área_de_almacenamiento
- Cristopher Poelker y Alex Nikitin. Storage Area Networks for Dummies.
- G. Somasundaram, Alok Shirvastava "Information, Storage and Management: Storing, Managin and Protecting Digital Information". John Wiley & Sons. April 06, 2009.
- · Jason Buffington "Data protection for Virtual Data Centers". Sybex. August 02, 2010.
- · Doug Lowe "Networking for Dummies". John Willey & sons. May 29, 2007.

Autor: Francisco Javier Rodriguez Martínez

Subdirector de Sistemas Escola Superior Enxeñaría Informática Ourense

Colegiado del CPEIG