Universidad de Granada
Escuela Técnica Superior de Ingeniería Informática y Telecomunicaciones
Grado en Ingeniería Informática

Centros de Procesamiento de Datos

Actividad : Copias de seguridad en CPD

Javier Martín Gómez
Juan Jesús Torres Prieto

2018/2019

Índice

1. Directrices fundamentales a tener en cuenta a la hora de hacer copias de seguridad

- Llevar a cabo todas las copias de seguridad desde un sitio central. La centralización facilita la administración de las copias, pero tiene la desventaja de que tienen que hacerse en remoto, lo que supone un coste mayor de tiempo comparado con una copia local.
- Etiquetar completamente cada medio de copia. Se debe conocer los sistemas de ficheros guardados, la fecha de la copia, el formato, la sintaxis de las órdenes que se han usado para generarlas, las órdenes necesarias para restaurar la copia y cualquier otra información que se considere relevante.
- Escoge un intervalo de copia prudente. Cuantas más copias se hacen, menos datos se pierden en un fallo, pero hacer copias consumen recursos de sistema y tiempo del administrador. Esto depende de cada uso de caso y del coste que se pueda asumir. En sistemas de mucho uso, pueden hacerse copias una vez al día, mientras que si tienen poco uso, pueden hacerse copias una o varias veces a la semana.
- Elige bien qué sistemas de ficheros copiar. Si un directorio no se modifica mucho no es necesario copiarlo tanto como los directorios de usuario. Por ejemplo, el directorio /etc no se suele cambiar todos los días, así que se pueden copiar sus datos a otra partición que tenga copias de seguridad frecuentes. Directorios como /tmp no necesitan ser copiados porque todos sus datos son temporales.
- Las copias de seguridad diarias deben caber en un solo medio. Lo ideal sería que todas las copias importantes cupieran en una cinta, ya que así al restaurarla no habría que estar presente montando una cinta tras otra. El problema es que los discos son cada vez más grandes y baratos, por lo que los sistemas se hacen cada vez más duros de copiar.
- Mantén las copias off-site. Es decir, que no esté guardada en el disco duro de la máquina de origen. Las imágenes de disco y el RAID no son copias de seguridad. El motivo principal es estar protegido frente a desastres que puedan destruir los datos del centro.
- Protege las copias de seguridad. Los datos copiados son muy importantes, así que no solo debes tenerlas off-site, sino también cifradas y protegidas bajo llave.
- Limita la actividad durante las copias. Los cambios mientras se hace una copia puede que den como resultado un copia errónea, así que lo ideal sería hacerlas cuando nadie o casi nadie está usando la máquina, al final de la jornada de trabajo o los fines de semana. Las bases de datos son especialmente sensibles y requieren procedimientos especiales, se deben detener o poner en modo limitado durante las copias. En la práctica es imposible asumir la pérdida de uptime necesaria para hacer una copia de

seguridad tradicional, así que la única forma fiable de hacerla es con una imagen (snapshot) de disco.
Todos los sistemas operativos actuales tienen medios para hacer imágenes de disco. Esto te permite copiar un sistema de ficheros en activo. En linux se pueden hacer con Logical Volume Manager (LVM).
Las imágenes se pueden crear rápidamente gracias a un sistema copy-on-write. Una vez se crea la imagen no se copian datos, sino que los cambios a partir de entonces se guardan en un sitio distinto del disco. Esto recuerda a una copia incremental, solo que las imágenes funcionan a nivel de sistema de ficheros.
Por todo esto, las imágenes son una herramienta útil para crear copias de seguridad de verdad, pero no sustituyen a las copias off-site ya que se hacen en el mismo disco, y son especialmente necesarias para las bases de datos, ya que solo se detienen un instante a hacer la imagen y luego se puede copiar la base de datos a otro disco tranquilamente mientras sigue operativa.

- Verifica que las copias se pueden restaurar: Hay que leer las copias justo después de hacerlas. Una buena prueba inicial es comprobar que el número de archivos es correcto. Hay que restaurar las copias periódicamente para comprobar que se pueden usar, especialmente si son cintas de hace unos años, porque se pueden haber estropeado.
- Crea un ciclo de vida para los medios. Hay que ir rotando los discos y cintas que se usen para copiar. Comprobar el tiempo de vida media que da el fabricante.
- Organiza los datos en torno a las copias de seguridad. Hay que recopilar la siguiente información: Los tipos de datos que manejas, la volatilidad de cada tipo de dato, la frecuencia de copia necesaria… a partir de ahí hay que desarrollar una arquitectura de almacenamiento, teniendo siempre en cuenta las copias de seguridad y el crecimiento. Por ejemplo, si pones los directorios de proyectos y las home de usuario en un sistema de ficheros dedicado es más fácil administrar y proteger los datos que si estuvieran cada uno en un ordenador. Cuando una máquina se estropea, suele ser más fácil volcar una imagen nueva en ella que tratar de averiguar qué ha fallado y arreglarlo.
- Siempre prepararse para lo peor. Una vez esté establecido el procedimiento de copias de seguridad, hay que asumir siempre que en cualquier momento puede haber un fallo total del sistema. Hay que tener en cuenta siempre cuantos datos se podrían perder y cuánto tiempo se tardaría en restaurar el sistema, incluso cuánto se tardaría en comprar e instalar nuevo hardware. Más formalmente: RTO (Recovery Time Objective) representa el máximo tiempo que un negocio puede esperar a recuperarse y RPO (Recovery Point Objective) indica qué tan reciente debe ser una copia de seguridad que se restaure. Para definir estos tiempos, sería necesario reunir a los dueños de los datos con el equipo técnico y equilibrar el coste asumible con las necesidades del negocio.

2. Tipos de medios de almacenamiento

- CD/DVD/Bluray: Buenos para copiar sistemas pequeños y aislados. Si se almacenan correctamente pueden durar 100 años.
- Discos duros externos: Fáciles de conectar por USB o eSATA. Son buenos para copias off-site porque se extraen fácilmente.
- Cintas magnéticas: Son muy sensibles a campos electromagnéticos y tienen solo unos años de retención de datos, pero tienen una gran capacidad de almacenamiento. El último estándar de cintas LTO-8, soporta hasta 12 TB de almacenamiento y 300 MBps de transferencia de datos, mientras que las cintas LTO-4 de hace 10 años tenían 800 GB de capacidad. Como son cintas, el acceso a los datos es secuencial, por eso las copias de seguridad suelen ser su único caso de uso. Si el tamaño de una cinta no es suficiente, existen cambiadores automáticos de cintas que cambian una por otra cuando se ha llenado, no requiriendo así intervención humana.
- Discos duros: Gracias a la reducción de sus precios, cada vez se usan más las copias disco a disco. Si se hacen a través de la red son una buena opción de copia off-site. A diferencia de las cintas, como son de acceso aleatorio son buenos para recuperar solo archivos puntuales: se puede poner un disco en red con una copia del trabajo del día anterior y si un trabajador borra por error algún archivo lo puede recuperar el mismo desde ahí sin llamar a un administrador. Hay que recordar que si un disco está on-line no está lo suficientemente protegido contra atacantes o fallos del CPD, siempre hay que tener una copia off-site.
- La nube: Datos almacenados en la red por un proveedor externo. Permite almacenar datos fácilmente en múltiples lugares geográficos. Suelen ser de pago por uso, es decir cobran según la cantidad que almacenes y el tiempo que esté almacenada. Solo son útiles si la conexión a internet es lo suficientemente rápida como para no afectar el tráfico de red necesario para el trabajo. Si la información es confidencial, debe transmitirse cifrada.

3. Copias completas e incrementales

Generalmente existen dos tipos de copias: Copias completas y copias incrementales. Una copia completa contiene todo el sistema de ficheros y una incremental contiene sólo los cambios desde la copia anterior. Las copias incrementales son buenas porque reducen el ancho de banda que se consume y el almacenamiento usado en las copias diarias. Como la mayoría de los archivos no cambia el ahorro es considerable.
Algunos programas también identifican ficheros o bloques de datos idénticos y copian las partes repetidas solo una vez. A esto se le llama deduplicación.

4. Copias de seguridad con Dump

Dump es una herramienta en línea de órdenes para copiar sistemas de ficheros enteros. Permite hacer copias que quepan en varios discos, preservando los permisos y fechas de modificación, maneja bien los archivos huecos (que tengan una cantidad grande de ceros consecutivos), y permite hacer copias incrementales. Las copias incrementales usan un sistema de niveles. El nivel 0 copia todo el sistema de ficheros, el nivel N copia todos los cambios que hubo desde la última copia menor que N.
El primer argumento que se le pasa es el nivel de copia incremental.
La opción -u actualiza automáticamente el archivo /etc/dumpdates con la información de las copias incrementales hechas para poder restaurarlas.
La opción -f indica a qué dispositivo se va a guardar la copia. Por defecto dump apunta a la primera unidad de cinta. También se puede copiar a un dispositivo remoto con la sintaxis hostname:dispositivo.

Ejemplo de uso:
$ sudo dump -0u -f maquinaremota:/dev/sdb /

Las copias de dump se restauran con restore. Permite usar un modo interactivo con -i para restaurar archivos puntuales y la opción -r que restaura todo el sistema de ficheros en el directorio de trabajo actual.

5. Copias de seguridad con Bacula

Bacula es una herramienta empresarial para copias de seguridad muy sofisticada, que administra copias, restauración y verificación de archivos por la red. Se trata de un sistema modular compatible tanto con varios sistemas tipo UNIX como con windows, puede usar bases de datos como MySQL, PostgreSQL o SQLite, crea sumas de verificación sha1 o md5 para cada archivo, permite cifrado tanto del tráfico de red como del almacenamiento, permite ejecutar scripts tras una copia y puede centralizar la administración de copias de toda una red de ordenadores.

Como bacula es modular, consta de varios componentes:
-Director: Es el demonio que coordina las copias, restauraciones y verificaciones.
-Consola: A través de ella nos comunicamos con el director. Puede ser una línea de órdenes o una interfaz gráfica.
-Almacenamiento: Es el demonio que lee y escribe discos de copia. Se debe ejecutar en la máquina conectada al disco que hace las copias.
-Cliente de ficheros: Se ejecuta en cada una de las máquinas que hay que copiar. Cada una envía los atributos y datos de los ficheros al demonio de almacenamiento cuando se solicita una copia.

-Catálogo: Es una base de datos relacional en la que se guarda toda la información sobre cada archivo y volumen que está copiado. Esto hace bacula muy eficiente ya que sabe en cualquier momento que discos son los necesarios para restaurar unos archivos concretos sin tener que ir a buscar el disco conectarlo y leerlo para ver si están ahí los archivos.

Como componente opcional, existe un CD de rescate basado en linux con el demonio de ficheros y varios scripts que facilitan el rescate de un sistema en caso de fallo.

6. Centros de respaldo

Un centro de respaldo absorbe las operaciones del CPD principal en caso de emergencia como terremotos, incendios, atentados que aunque son infrecuentes no son improbables.

Están situados habitualmente a unos 20-40km del CPD principal, dependiendo de las necesidades de telecomunicaciones entre ambos centros.

El equipamiento hardware no tiene porque ser idéntico al del CPD principal ya que no todos los procesos son críticos, en cambio el software si debe ser idéntico para que el centro de respaldo sea compatible con el CPD principal.

7. Bibliografía

[1] https://es.wikipedia.org/wiki/Centro_de_respaldo

UNIX and Linux System Administration Handbook, 4th Edition

https://public.dhe.ibm.com/common/ssi/ecm/ts/en/tsd03243usen/tsd03243-usen-00_TSD03243USEN.pdf