UF1473: Salvaguarda y seguridad de los datos

Elaborado por: Alberto Gómez García

Edición: 5.0

EDITORIAL ELEARNING S.L.

ISBN: 978-84-16275-51-9 • Depósito legal: MA 1795-2014

Impreso en España - Printed in Spain

Presentación

– Identificación de la Unidad Formativa

Bienvenido a la Unidad Formativa **UF1473: Salvaguarda y seguridad de los datos**. Esta Unidad Formativa pertenece al Módulo Formativo **MF0225_3: Gestión de bases de datos**, que forma parte del Certificado de Profesionalidad **IFCT0310: Administración de bases de datos**, de la familia profesional Informática y comunicaciones.

– Presentación de los contenidos

La finalidad de esta unidad formativa es enseñar al alumno a implantar la política de control de acceso en los gestores de bases de datos, a planificar y realizar copias de seguridad, así como a realizar la recuperación de datos en caso necesario y habilitar el acceso a las Bases de Datos de acuerdo a criterios de confidencialidad, integridad y disponibilidad. Para ello, se analizarán los procesos de salvaguarda y recuperación de datos y se estudiarán las bases de datos distribuidas, además de profundizar en la seguridad de los datos y en su transferencia.

– Objetivos

Al finalizar esta unidad formativa aprenderás a:

- Mantener la seguridad de los accesos a las bases de datos garantizando la confidencialidad.
- Garantizar la salvaguarda y recuperación de la información almacenada en las bases de datos de acuerdo a las necesidades de cada una de ellas.
- Exportar e importar datos de la Base de Datos garantizando su integridad.

Índice

UD1. Salvaguarda y recuperación de datos

UD2. Bases de datos distribuidas desde un punto de vista orientado a la distribución de los datos y la ejecución de las consultas

UD3. Seguridad de los datos

UD4. Transferencia de datos

UD1
Salvaguarda y recuperación de datos

1.1. Descripción de los diferentes fallos posibles (tanto físicos como lógicos) que se pueden plantear alrededor de una base de datos
1.2. Enumeración y descripción de los elementos de recuperación ante falloslógicos que aportan los principales SGBD estudiados
1.3. Distinción de los diferentes tipos de soporte utilizados para la salvaguarda de datos y sus ventajas e inconvenientes en un entorno de backup
1.4. Concepto de RAID y niveles más comúnmente utilizados en las empresas:
    1.4.1. RAID5, RAID6
    1.4.2. Clasificación de los niveles RAID por sus tiempos de reconstrucción
1.5. Servidores remotos de salvaguarda de datos
1.6. Diseño y justificación de un plan de salvaguarda y un protocolo de recuperación de datos para un supuesto de entorno empresarial
1.7. Tipos de salvaguardas de datos:
    1.7.1. Completa
    1.7.2. Incremental
    1.7.3. Diferencial
1.8. Definición del concepto de RTO (Recovery Time Objective) y RPO (Recovery Point Objective)
1.9. Empleo de los mecanismos de verificación de la integridad de las copias de seguridad

1.1. Descripción de los diferentes fallos posibles (tanto físicos como lógicos) que se pueden plantear alrededor de una base de datos

Una base de datos y el sistema en donde esté alojada, se pueden ver expuestos a una gran multitud de posibles fallos.

Hoy en día, es de vital importancia para las organizaciones el proteger la información que estas manejan, ya que si no pueden disponer de ella en el momento que lo necesiten, la continuidad del negocio puede verse comprometida.

Por este motivo, invierten muchos recursos en conseguir sistemas cada vez más fiables y eficientes. Estos sistemas han de ser capaces de sobreponerse por sí solos a un posible fallo.

Ya que estos fallos pueden ser de muy diversa índole, los sistemas tendrán que contemplar diferentes herramientas y metodologías para blindarse ante los desastres.

Durante los últimos años, los desarrolladores de bases de datos han invertido una gran cantidad de recursos en conseguir que sus SGBD tengan cada vez mayor tolerancia a los fallos.

En la actualidad, todos los SGBD profesionales suelen integrarse dentro de una arquitectura distribuida que de soporte para tolerancia a fallos.

Estos sistemas se consideran de alta disponibilidad, ofreciendo continuidad en el servicio aún en presencia de fallo.

Con el uso masivo de Internet, algunos sistemas se han visto desbordados por el crecimiento inesperado de la demanda de sus servicios, quedando estos obsoletos.

Para paliar esto, muchos fabricantes de servidores de bases de datos, ya incorporan en sus productos potentes mecanismos de replicación. Con ello, se consigue que quiénes utilicen sus servidores, tengan la seguridad de que la información que almacenan va a estar siempre disponible, aun cuando alguno de ellos presente algún fallo.

Un ejemplo de ello podría ser el acceso a las bases de datos de Google, cuando un usuario carga su página web para buscar algo, realmente está accediendo a uno de los servidores de los muchos que Google tiene repartidos por el planeta, sin ser consciente de a cuál de ellos está accediendo, ni cuántos de ellos se encuentran inactivos.

Lo importante para el usuario es que el acceso a la información sea transparente y está se le muestre en el menor tiempo posible.

Cuando hablamos de fallos, debemos hacer una distinción entre lo que serían fallos lógicos y lo que serían fallos físicos.

Los fallos lógicos son aquellos que se producen en un programa en cuestión, como puede ser en el sistema operativo o en la base de datos. Cuando esto ocurre, se verá interrumpida la operación que en ese mismo instante se estaba procesando; en ocasiones podrán continuar normalmente el resto de operaciones pendientes y en otras ocasiones, se interrumpirá completamente la ejecución de dicho programa.

Por el contrario, los fallos físicos son aquellos que afectan a algún componente hardware del sistema. Dependiendo de la gravedad, el sistema podrá detenerse total o parcialmente, provocando la discontinuidad de su labor.

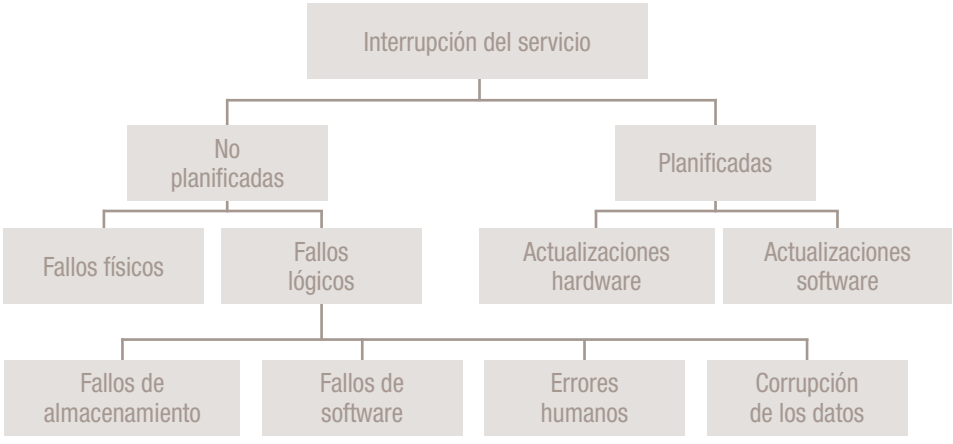
Cuando se trata de SGBD en entornos profesionales, el sistema debe estar preparado para recuperarse no sólo de fallos lógicos,comouna mala terminación en un proceso de transacción, sino también de fallos físicos, como podría ser la interrupción del suministro eléctrico al servidor.

Las fallas lógicopuede afectar solo a una operación concreta, pero por el contrario, un fallo físico puede afectar al sistema completo, ya que el propio servidor dejará de funcionar y de atender las peticiones de todos los usuarios.

Las interrupciones en el sistema no siempre serán debidas a fallos en el mismo, sino que a veces también estarán planificadas, como por ejemplo al tener que actualizar el hardware del propio servidor.

Estas interrupciones planificadas en el servicio, también tendrán que ser soportadas y solventadasde forma que el usuario no note estos parones.

Por tanto podríamos realizar el siguiente esquema con respecto a las interrupciones del servicio:



Con respecto a las tareas planificadas, si se prevé que haya que apagar el sistema para llevar a cabo las tareas de mantenimiento, tanto de hardware como de software, se tendrá que hacer en un horario que afecte lo menos posible a los accesos de los usuarios, es decir, cuando menor número de accesos se registre de media y siempre avisando con antelación del corte del suministro.

Normalmente suele ser en horario nocturno, siempre y cuando al servidor no se acceda desde otros países con distintos usos horarios. Por ejemplo, si nuestro servidor es manejado habitualmente por usuarios españoles y usuarios de Nueva Zelanda, entonces no habrá un horario concreto de menor acceso, ya que Nueva Zelanda es la antípoda de España, con lo que cuando aquí es noche cerrada, allí estarán a medio día.

El término **antípoda** es utilizado en Geografía para señalar el punto más distante de otro sobre la superficie terrestre. Este lugar coincide normalmente con el situado al otro lado de la diagonal terrestre.

En estos casos, se hará imprescindible el contar con un servidor replicado, para que al desactivar uno de ellos (para el posible mantenimiento), el otro entre en funcionamiento y así el sistema nunca se vea interrumpido.

Los fallos más graves normalmente vendrán de los no planificados, ya que serán los que no esperemos que ocurran hasta que realmente han ocurrido.

Vamos a comentarlos por separado.

– **Fallos físicos**

  Cuando hablamos de fallos físicos nos referimos a fallos del hardware o de las instalaciones. En ocasiones estos fallos pueden considerarse verdaderos desastres, como por ejemplo un incendio.

  Por suerte, en la mayoría de las ocasiones solo supondrán algún fallo concreto en alguna de las piezas, que una vez sustituida, el sistema seguirá operando con normalidad.

  De tratarse de un fallo hardware en el servidor, dependiendo de la pieza en la que se produzca el fallo, este se podrá ver totalmente interrumpido,

como por ejemplo si fallase la fuente de alimentación, o solamente se verá afectado una parte de él, como por ejemplo si se estropease la unidad lectora de DVD, esta no se podría utilizar, pero el servidor seguiría funcionando con normalidad y en consiguiente el SGBD también.

Para evitar en mayor medida estos posibles parones, los servidores modernos cuentan con sistemas duplicados de aquellas piezas con una mayor posibilidad de rotura, como son las fuentes de alimentación o los discos duros.

En uno de estos servidores, de fallar una de las fuentes, la otra entraría en funcionamiento de forma automática, sin que se aprecie una caída del sistema. Algo parecido también ocurre con los discos duros.

Se caiga el sistema o siga esté funcionando, será necesario que un técnico cualificado repare en la menor brevedad posible los fallos físicos que presente el servidor.

Otro tipo de fallos físicos son los producidos en las instalaciones que tengan efecto directo con la interrupción del servicio.

Dentro de estos fallos podemos mencionar los desastres naturales como incendios, terremotos, inundaciones, etc., los cuales causarán un daño extremo en los sistemas, siendo la única forma de asegurarnos de que no caiga la continuidad del negocio, el disponer de servidores replicados en otros lugares geográficos.

Otro posible fallo del sistema puede ser la interrupción del suministro eléctrico en la organización. Para protegernos bastará con contar con un SAI conectado al servidor.

SAI son las siglas de Sistema de Alimentación Ininterrumpida. Se trata de aparatos especialmente preparados para regular la corriente eléctrica, evitando que se produzcan subidas o bajadas de tensión, eliminar el ruido eléctrico y seguir proporcionando electricidad durante un tiempo, cuando el suministro eléctrico principal se interrumpe.

– **Fallos lógicos**

Los fallos lógicos son los que en mayor medida, como administradores podremos intentar prevenir.

Según el esquema anterior los podemos clasificar en:

- Fallos de almacenamiento
- Fallos de software
- Errores humanos
- Corrupción de los datos

Los fallos de almacenamiento se producen cuando se el sistema de archivos detecta alguna incoherencia entre la información lógica de lo que supuestamente hay almacenado y la información física que realmente esta almacenada.

Cuando ocurre uno de estos fallos, normalmente quedará destruida físicamente parte de nuestra base de datos.

La solución para poder recuperar el sistema a un estado coherente después de uno de estos fallos, nos obligará a restaurar una copia de seguridad previa y posteriormente volver a reproducir todas aquellas sentencias que hayan alterado la base de datos desde el momento posterior a la copia de seguridad, hasta cuando se produjo el fallo.

Si justo cuando se produjo el fallo de almacenamiento, se estaba realizando algún tipo de transacción, y esta se quedó a medias, esta última transacción habrá que llevarla a cabo completa y no solo aquellas sentencias que se realizaron antes del fallo.

Todos los sistemas de almacenamiento guardan un índice de lo que tienen almacenado.Como norma general cuando un usuario está "navegando" en una unidad entre sus carpetas y archivos, lo está haciendo realmente en el índice de esa unidad; cuando al final intenta abrir o ejecutar algo, es cuando la unidad se va a buscar realmente la información necesaria para leerla, si cuando localiza la zona en donde según su índice, debiera estar el archivo buscado y no lo encuentra, se produce un fallo de coherencia entre la información física y la lógica, con el consiguiente error.

La parte del sistema responsable de la administración de los archivos del almacenamiento secundario se denomina **Sistema de Archivos**.

El sistema de archivos es la parte del sistema operativo responsable de permitir compartir de forma segura la información de los archivos y de proporcionar un acceso controlado a los mismos.

Entre los sistemas de archivos más usuales presentes en el mercado se encuentran FAT (DOS/Windows), NTFS (Windows NT/2000/XP/2003/7/8), HPFS (OS/2) y ext2fs/ext3fs (Linux).

Con respecto a los fallos de software, estos pueden provenir del sistema operativo o del propio SGBD.

Algún programa puede realizar operaciones que causen un overflow de un entero o la división por cero, así mismo puede ocurrir que se pasen valores erróneos a algún parámetro o que se detecte un error en la lógica de un programa, o que sencillamente no se encuentren los datos del programa.

Además, si el usuariose conecta por consola, en algunos sistemas puede explícitamente interrumpir una transacción durante su ejecución, por ejemplo pulsando Control+C.

Para evitar los posibles fallos del sistema operativo, es aconsejable tenerlo debidamente actualizado y configurado, siendo necesario un chequeo cada cierto tiempo para comprobar su consistencia.

Con respecto al SGBD, si se detectan errores concretos, estos deberán ser remitidos al fabricante y desarrollador del producto, para que de no existir, cree una actualización que los corrija.

Todos los SGBD profesionales cuentan con contratos de soporte, en los que personal cualificado estará a nuestra disposición para ayudarnos a configurar nuestro sistema y solventar aquellos fallos que se produzcan.

Los errores humanos pueden producir por un descuido o intencionadamente, y en general modificarán la información almacenada en nuestra base de datos, ya sea porque la han eliminado o porque han modificado algún valor.

Es importante de que una vez tengamos conocimiento del fallo, los datos afectados sean repuestos a su estado correcto por medio de la restauración de una copia de respaldo.

Para evitar en lo más posible estos errores, tendremos que asignar correctamente los privilegios a los usuarios del sistema, para que solo aquellos que tengan permiso, puedan realmente realizar alguna modificación en la información.

El resto de usuarios se limitará a acceder y listar la información, sin posibilidad de modificarla.

Por último, la corrupción de los datos se produce cuando no podemos acceder de una forma correcta a los mismos.

Las causas para que se produzca una corrupción en la información, pueden ser diversas, pero el resultado siempre será el mismo, que no podremos volver a utilizar la información corrupta.

La forma de devolver el sistema a un estado coherente, será una vez más por medio de la restauración de una copia previa de seguridad.

1.2. Enumeración y descripción de los elementos de recuperación ante falloslógicos que aportan los principales SGBD estudiados

Todos los SGBD profesionales cuentan con una serie de aplicaciones y herramientas destinadas a la recuperación de los datos ante un posible fallo a nivel lógico.

De no disponer de estas opciones, cuando se produjese un error, sería muy probable la pérdida de información.

Vemos a ver por tanto las opciones que contemplan al respecto los sistemas:

- Oracle DB
- Microsoft SQL Server
- MySQL de Oracle

– **Oracle DB**

Cuando pretendemos que nuestro sistema tenga siempre sus datos a punto y sea capaz de recuperarse de un posible fallo, estamos ante un sistema de alta disponibilidad.

La alta disponibilidad la podemos aplicar al propio servidor, a solo los datos o a ambos.

Algunas de las características de alta disponibilidad ofrecidas por Oracle DB son las siguientes:

- **Disponibilidad del servidor**

  La disponibilidad del servidor pasa por garantizar el acceso ininterrumpido a los servicios de las bases de datos a pesar de los posibles fallos inesperados de una o más máquinas que alojan el servidor de la base de datos, lo cual puede producirse debido a fallos de hardware o de software.

  Oracle posee una arquitectura privada de cloudcomputing, en la que Oracle Real ApplicationClusters, ofrece la protección más efectiva contra este tipo de fallos.

Cloud computing, son propuestas tecnológicas que permiten ofrecer servicios de computación a través de Internet.

Oracle Real ApplicationClusters (RAC) es una tecnología de agrupación en clústeres de bases de datos que permite que dos o más equipos (nodos) de un grupo de servidores accedan al mismo tiempo a una sola base de datos compartida.

Aunque este sistema de base de datos pueda incluir varios nodos, aparecerá ante la aplicación como una sola base de datos unificada.

Esta arquitectura permite beneficiarse de las ventajas que ofrece la disponibilidad y la escalabilidad, como por ejemplo:

- Tolerancia a fallos dentro del grupo de servidores, especialmente a fallos de los equipos.
- Flexibilidad y rentabilidad en la planificación de las capacidades del sistema, ya que el sistema puede escalar a cualquier capacidad según la demanda y a medida que cambian las necesidades de la empresa.

La ventaja principalde Oracle Real ApplicationClusterses la tolerancia a fallos,gracias a disponerde varios nodos.

Como los nodos físicos se ejecutan en forma independiente, el fallo de uno o más nodos no afecta a los demás. Este tipo de arquitecturas también permite que un grupo de nodos se conecte o desconecte de forma transparente, mientras que el resto de nodos conectadossiguen dando servicio a la base de datos.

Si un futuro aumenta la demanda de capacidad,Oracle Real ApplicationClusterspermite agregar tantos nodos como sean necesarios para dotar al sistema de la potencia necesaria. Este método de escalabilidad es cómodo y económico ya que no obliga a sustituir ninguna máquina..