

# CDA. Almacenamiento en centros de datos

Requisitos típicos  
Niveles RAID  
Redes de almacenamiento  
Copias de seguridad

**Centros de datos**  
3º Grado en Ingeniería Informática  
ESEI

Septiembre-2018

# Contenido

- 1 **Requisitos y necesidades típicas**
- 2 **Niveles RAID**
  - Niveles RAID estándar
  - Niveles RAID anidados
- 3 **Redes de almacenamiento**
  - DAS vs SAN vs NAS
  - Elementos SAN
  - Protocolos SAN
  - Sistemas de ficheros sobre SAN
- 4 **Copias de seguridad**
  - Dispositivos
  - Estrategias de copia

# Requisitos y necesidades típicas

## Capacidad

- Muy variable en función del contexto → tendencia creciente
- **Importante:** aumento de tamaño transparente

## Rendimiento

- Variable en función de los dispositivos empleados (tipos de disco, esquema de agrupación, protocolos de red, ...)
- Combinación/estructuración en niveles

## Disponibilidad

- Acceso por red (protocolos)
- Tolerancia a fallos: replicación + recuperación ante fallos

## Facilidad de gestión

- Crecimiento + tolerancia a fallos transparente
- **Tendencia:** virtualización del almacenamiento
  - Capas software (opc. asistidas por hardware) que abstraen el almacenamiento
  - Ejemplos: RAID, LVM (*Logical Volume Management*), protocolos SAN, (SDS: *Software Defined Storage*), ...

## 1 Requisitos y necesidades típicas

## 2 Niveles RAID

- Niveles RAID estándar
- Niveles RAID anidados

## 3 Redes de almacenamiento

- DAS vs SAN vs NAS
- Elementos SAN
- Protocolos SAN
- Sistemas de ficheros sobre SAN

## 4 Copias de seguridad

- Dispositivos
- Estrategias de copia

# RAID

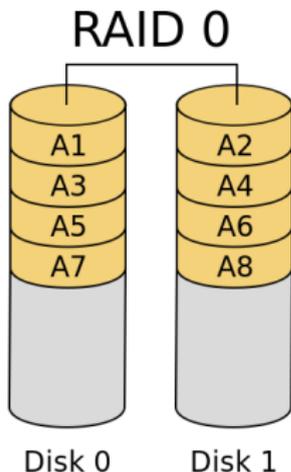
## *Redundant Array of Independent Disks* (originalmente, *Redundant Array of Inexpensive Disks*)

- Combina múltiples dispositivos de almacenamiento (discos) en una única unidad lógica
- **Objetivos:** (cada esquema RAID prioriza uno o más)
  - dotar de redundancia a los datos almacenados (tolerancia a fallos)
  - aumento de capacidad
  - mejora del rendimiento (en lectura fundamentalmente)
  - facilidad de gestión
- **Mecanismos** utilizados
  - **data stripping:** esquemas de "reparto" de los bloques de datos sobre los dispositivos
  - **mirroring:** réplica de los bloques de datos en distintos dispositivos
  - **paridad:** uso de bloques/segmentos con datos de paridad para la recuperación ante errores
  - **spare disks** (discos de reserva): almacenamiento extra que reemplaza [automáticamente (*hot spare*) o de modo manual] a unidades fallidas (elemento adicional/complementario a los niveles estándar)
- Posibilidad de combinar niveles RAID (RAID anidado)
- Implementaciones hardware (controladoras, cabinas de discos) y software (RAID software en GNU/Linux)
- Influencia otras tecnologías de almacenamiento: modos RAID en LVM2

## RAID 0: data striping

**Distribuye** los **bloques** de datos **secuencialmente** entre los  $N$  dispositivos que forman el array.

- No realiza replicación de bloques ni control de paridad.



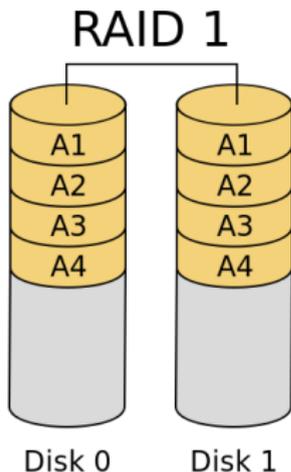
- $N^{\circ}$  mínimo de discos: 2
- Capacidad efectiva:  $N \times$  tamaño disco más pequeño
- No tolera fallos de discos  
(prob. fallo = probabilidad de fallo del "peor disco")
- Mejora rendimiento en escrituras y lecturas
  - permite acceso concurrente a bloques adyacentes ("caen" en discos diferentes)

<https://en.wikipedia.org/wiki/>

Standard\_RAID\_levels

# RAID 1: data mirroring

**Replica cada bloque** en cada uno de los  $N$  discos que forman el array



- Nº mínimo de discos: 2
- Capacidad efectiva: tamaño disco más pequeño
- Tolera fallos de  $(N - 1)$  discos  
(prob. fallo = producto de la prob. de fallo de los discos del array)
- Mejora rendimiento en lecturas
  - permite lecturas concurrentes de bloques (una en cada disco)
  - en escritura no aporta ganancia

[https://en.wikipedia.org/wiki/](https://en.wikipedia.org/wiki/Standard_RAID_levels)

Standard\_RAID\_levels



## Otros niveles RAID estándar

### Requisitos

### RAID

#### RAID estándar

#### RAID anidado

### Redes Almac.

#### SAN/NAS

#### Elementos

#### Protocolos

#### Sist. Ficheros

### Backup

#### Dispositivos

#### Estrategias

- **RAID 2:** Distribuye datos entre los discos (+ calcula paridad) a nivel de bits
  - corrección de errores con códigos Hamming
  - no usado en la práctica
- **RAID 3:** Distribuye datos entre los discos (+ calcula paridad) a nivel de bytes
  - no permite lecturas en paralelo
  - no usado en la práctica
- **RAID 4:** Similar a RAID 5
  - Uno de los dispositivos dedicado en exclusiva a los bloques de paridad en lugar de distribuirlos por el array como en RAID 5
  - Escrituras en disco de paridad → cuello de botella
  - Menor tolerancia a fallos que RAID 5 (disco de paridad → punto único de fallo)
- **RAID 6:** Versión ampliada de RAID 5 con 2 bloques de paridad
  - Usa dos esquemas de cálculo de paridad distintos (cada uno en su propio bloque de paridad)
  - Permite recuperar el fallo de dos discos
  - N° mínimo de discos: 4 (2 datos + 2 paridad)
  - Capacidad efectiva:  $(N - 2) \times$  tamaño disco más pequeño

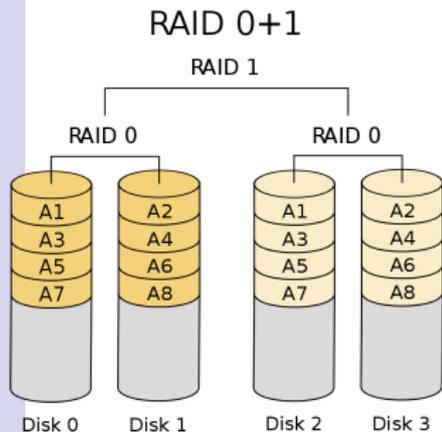
## Niveles RAID no estándar

- **RAID L:** "RAID lineal" (no es parte del estándar RAID)
  - Distribuye los datos usando los  $N$  discos del array uno detrás de otro
  - No "reparte" los bloques secuencialmente sobre los discos como RAID 0 ("concatena" discos uno tras otro)
  - Capacidad efectiva =  $\sum$  tamaño de cada disco
  - No tolera fallos de discos
  - Menor rendimiento en lectura que RAID 0 (no lectura concurrente de bloques contiguos)
  - Otros nombres: **JBOD** (*Just a Bunch Of Disk*)
- **RAID 5E:** RAID 5 con un *spare disk*
  - Dispone de un disco adicional (disco de reserva) que reemplaza el fallo de uno de los discos del array
  - Suele preferirse a RAID 6
    - también consume 2 discos extra y soporta el fallo de hasta 2 discos
    - el cálculo de paridad es menos costoso (sólo una vez)
- **RAID 6E:** RAID 6 con un *spare disk*
  - Dispone de un disco adicional (disco de reserva) que reemplaza el fallo de uno de los discos del array
  - Soporta el fallo de hasta 3 discos

## RAID 01 (RAID 0+1)

Se replican (*mirror*) grupos de  $N/2$  arrays RAID 0 (distribución de bloques) (*mirror of arrays*)

- Normalmente usa 4 discos (*mirror* de 2 arrays RAID 0), aunque son posibles más
- RAID 0 "interno" distribuye bloques
- RAID 1 "externo" replica los 2 RAID 0 "internos"



- Nº mínimo de discos: 4 (2 RAID 0 + 2 RAID 0)
- Capacidad efectiva:  $(N/2) \times$  tamaño disco más pequeño
- Misma tolerancia a fallos que el RAID 1 "externo" (soportaría el fallo de hasta  $N/2$  discos siempre que pertenecieran al mismo RAID 0)
- Normalmente la pérdida de un disco "degrada" el array a un RAID 0 ("inhabilita" el RAID 0 al que pertenece)

<https://en.wikipedia.org/wiki/>

Nested\_RAID\_level



## Otros RAID anidados

- **RAID 50** (RAID 5+0): RAID 0 que combina al menos 2 RAID 5
  - Alternativa para organizar RAID 5 sobre grandes cantidades de discos
  - N<sup>o</sup> mínimo de discos: 6 (3 RAID 5 + 3 RAID 5)
- **RAID 100** (RAID 10+0): RAID 0 que combina al menos 2 RAID 10.
  - Añade un nivel adicional RAID 0 que agrupa varios RAID 10
  - Alternativa para organizar RAID 10 sobre grandes cantidades de discos

## 1 Requisitos y necesidades típicas

## 2 Niveles RAID

- Niveles RAID estándar
- Niveles RAID anidados

## 3 Redes de almacenamiento

- DAS vs SAN vs NAS
- Elementos SAN
- Protocolos SAN
- Sistemas de ficheros sobre SAN

## 4 Copias de seguridad

- Dispositivos
- Estrategias de copia

# Redes de almacenamiento

## Tres modelos de almacenamiento secundario

### DAS (*Direct Attached Storage*) [Almacenamiento de conexión directa]

Modelo empleado internamente en los equipos.

- Dispositivos de almacenamiento conectados internamente
- Acceso a datos a **nivel de bloques**
- No requiere uso de redes de comunicación

### SAN (*Storage Area Network*) [Red de área de almacenamiento]

Uso de una red específica (con protocolos propios) dedicada al acceso a los dispositivos de almacenamiento

- Dispositivos de almacenamiento "compartidos" conectados a través de una **red de almacenamiento dedicada** (red SAN)
- Acceso a datos a **nivel de bloques**
- Se implementan/emulan protocolos de acceso a dispositivos de almacenamiento (ATA, SCSI) que funcionan sobre una red de almacenamiento dedicada (Fibre Channel, TCP/IP, Ethernet) usando protocolos específicos (FCP, AoE, iSCSI, FCoE)

## Redes de almacenamiento (cont.)

### NAS (*Network Attached Storage*) [Almacenamiento conectado en red]

Uso de una red de área local (LAN) sobre la que se comparten ficheros/directorios

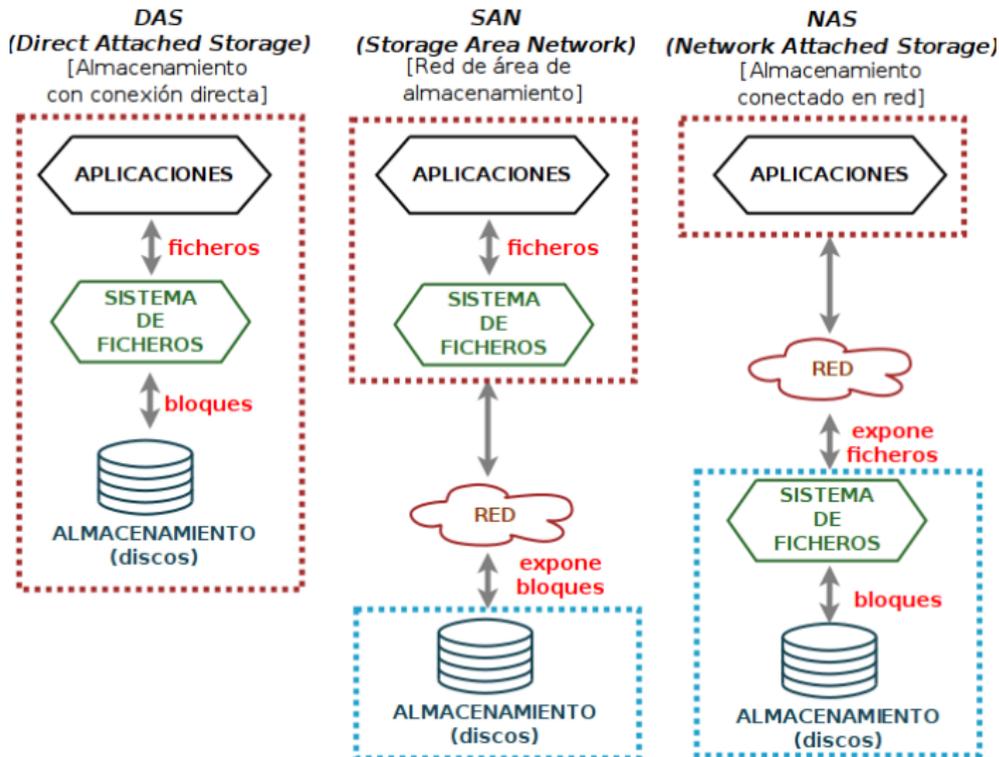
- Dispositivos de almacenamiento conectados a un "servidor NAS" (a través de una **red LAN convencional**)
- Acceso a datos a **nivel de ficheros**
- Uso (en el "servidor NAS") de **sistemas de ficheros de red** ( NFS [*Network File System*], SMB/CIFS [*Server Message Block/Common Internet File System*,...]) y/o protocolos de compartición de archivos (FTP, SFTP, etc)

- Pueden darse combinaciones de NAS sobre SAN (híbridos SAN-NAS)

### Otra alternativa: sistemas de almacenamiento distribuido sobre el cluster

- Usa el almacenamiento (normalmente local, tipo DAS) disponible en los nodos del cluster para conformar un sistema de almacenamiento "compartido" global
- Compatible con todos los esquemas anteriores
- Suelen incluir replicación de bloques en diferentes nodos
- Ejemplos:
  - DRDB (*Distributed Replicated Block Device*) [≈ RAID1 "distribuido"], <https://www.drbd.org/>
  - HDFS (*Hadoop Distributed File System*) [usado en clusters Hadoop]
  - GFS (*Google File System*, también GooFS) [antecedente de HDFS]

# DAS vs SAN vs NAS



Fuente: [https://es.wikipedia.org/wiki/Red\\_de\\_Área\\_de\\_almacenamiento](https://es.wikipedia.org/wiki/Red_de_Área_de_almacenamiento)

# Beneficios/limitaciones SAN

## Beneficios uso SAN

- Permite uso de múltiples discos de forma transparente desde varios equipos (servidores) [compartición + mejor aprovechamiento del espacio]
- Mejora rendimiento en acceso a datos [protocolos especializados + separación de redes [SAN vs LAN]]
  - uso de red dedicada a SAN, libera de tráfico de datos a la LAN "normal"
- Mejora tolerancia a fallos y recuperación ante desastres [arrays/cabinas de discos suelen implementar redundancia RAID, uso de discos de reserva (*spare disk*), alimentación redundantes, etc]
- Simplifica tareas de administración → centralizan gestión del almacenamiento
  - facilita redimensionamiento del espacio (incremento de capacidad [nuevos discos] de forma transparente)
  - facilita tareas recuperación y realización de backup (reducción tiempos de backup)
  - reduce costes de almacenamiento y gestión del almacenamiento
  - centraliza acceso a datos críticos
- Facilita construcción de clusters de servidores que requieran almacenamiento compartido (clusters de balanceo de carga, cluster de alta disponibilidad, clusters de alto rendimiento)
- Simplifica despliegue de soluciones de virtualización

## Limitaciones uso SAN

- Coste de puesta en marcha alto (depende del tipo de protocolos e infraestructura de red empleada)
- No recomendable/aplicable en todos los escenarios: equipos de escritorio finales, servidores de infraestructura (DNS, DHCP, Serv. de dominio, ...), etc

# Elementos y estructura típica SAN

Típicamente elementos de SAN estructurados en tres capas

## - Capa de almacenamiento (*storage layer*)

- Formada por los dispositivos de almacenamiento (discos, cabinas de discos, cintas, bibliotecas de cintas, ...) donde residen los datos [*targets*]
- Incluye la infraestructura de esos dispositivos (racks, fuentes de alimentación, caches internas, controladores RAID, *spare disks*, etc)

## -Capa de infraestructura (*fabric layer*)

- Formada por los dispositivos que dan soporte a la comunicación entre host finales y dispositivos de almacenamiento
- Incluye dispositivos (hubs, switches, routers, etc) y cableado específicos de la infraestructura de red utilizada
- Diversas alternativas dependiendo del protocolo SAN usado
- Posibilidad de definir múltiples rutas (*multipath*) entre equipos finales y dispositivos de almacenamiento

## - Capa de hosts (*host layer*)

- Formada por los equipos finales (servidores) que hacen uso del almacenamiento de bloques proporcionado por la SAN [*initiators*]
- Incluye los HBA (*Host Bus Adapters*): dispositivos físicos (tarjetas + drivers) que conectan a los host finales con la red de almacenamiento (dependen de la tecnología SAN usada)
- También software adicional y drivers: sistemas de ficheros en "cluster", software *multipath* (gestión/balanceo de múltiples rutas "físicas" de acceso a los datos), etc

# Protocolos SAN

## Diversas alternativas en *capa de infraestructura*

- Típicamente: **encapsulación** de protocolos usados en conexión con dispositivos de almacenamiento físicos (discos) sobre protocolos de red
  - SCSI [*Small Computer System Interface*]
  - ATA [*Advanced Technology Attachment*]
- "Exponen" bloques sobre la red
- Protocolos de red específicos vs. adaptación de protocolos generalistas
- Convergencia hacia comunicación sobre redes Ethernet y/o TCP/IP

## Protocolos SAN (cont.)

### [FCP] (*Fibre Channel Protocol*)

- Usa como protocolo de red *Fibre Channel* (FC) [canal de fibra]
  - Velocidades de 1,2,4,8,16 Gbit (recientemente 32, 128 Gbit)
  - Habitualmente sobre fibra óptica, posible sobre par trenzado
  - Uso principal en SAN
  - Organizado en 5 capas: FC0 (física), FC1(≈enlace), FC2(≈red), FC3(servicios), FC4(mapeo de protocolo)
- FPC **encapsula** el protocolo **SCSI sobre FC** (en capa FC4)
- Diversas topologías
 

}	FC punto a punto (2 dispositivos) [FC-P2P]
	FC en anillo arbitrado [FC-AL]
	FC conmutada [FC-SW]
- **FCoE** (*Fibre Channel over Ethernet*): encapsula paquetes FC sobre tramas Ethernet
  - Permite encapsular FC sobre redes Ethernet
  - Ethernet reemplaza a las capas "físicas" de FC (FC0 y FC1)

## Protocolos SAN (cont.)

### [iSCSI] (*internet Small Computer System Interface*)

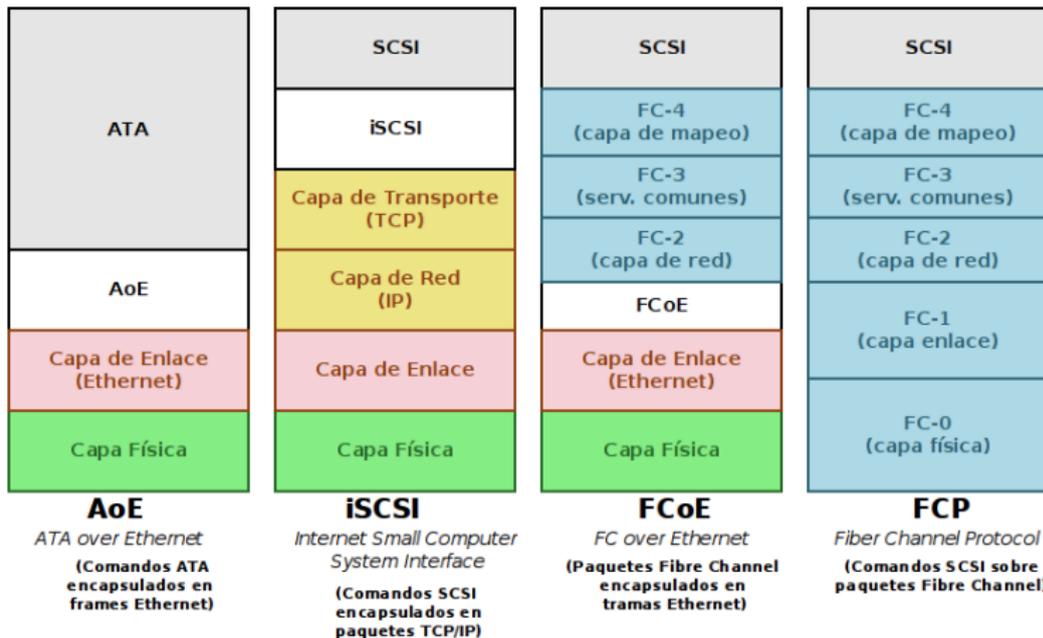
- Funciona sobre redes TCP/IP (en capa de aplicación) → permite enrutamiento
  - **Encapsula** protocolo **SCSI sobre** paquetes **TCP/IP**
  - Emula un dispositivo SCSI que funciona "sobre" la red TCP/IP
- No impone restricciones sobre la infraestructura de red
- Diversas variantes y/o implementaciones alternativas
  - HyperSCSI (encapsula SCSI sobre Ethernet)
  - iSER (*iSCSI Extensions for RDMA [Remote Access Memory]*)

### [AoE] (*ATA over Ethernet*)

- Funciona sobre redes Ethernet (en capa de enlace de datos [nivel 2 OSI]) → no permite enrutamiento
  - **Encapsula** protocolo **ATA sobre** tramas **Ethernet**
  - Emula un dispositivo ATA que funciona "sobre" la red Ethernet
- Coste más reducido que soluciones sobre FCP o iSCSI (HBA ≈ tarjeta Ethernet)
  - Desarrollado por Coraid
  - Cuenta con implementaciones hardware y software (p.e. en GNU/Linux)

# Resumen protocolos SAN

## SISTEMAS DE FICHEROS



## Sistemas de ficheros sobre SAN

SAN expone sobre la red los datos almacenados a **nivel de bloques**

- Equipos finales (servidores) "ven" **dispositivos de bloques**
- Sobre esos dispositivos, equipos finales crean sistemas de ficheros

Sistemas de ficheros "convencionales" (`ext3`, `ntfs`, etc) no funcionan correctamente cuando el **dispositivo de bloques SAN es compartido** por más de un host

- **Asumen** dispositivo de bloques accedido por un **único equipo**
- Acceso concurrente no coordinado  $\Rightarrow$   $\left\{ \begin{array}{l} \text{sobreescritura de datos} \\ \text{corrupción sistema de ficheros} \end{array} \right.$

Uso de **sistemas de ficheros de "cluster"**

- También sistemas de ficheros de "disco compartido"
- Equipos finales del cluster **coordinan** el **acceso** a los bloques de datos y a las estructuras del sistema de ficheros
- Uso de bloqueos distribuidos (*distributed locks*)
- OCFS2 (*Oracle Cluster File System*) [Desarrollado y mantenido por Oracle]
- GFS/GFS2 (*Global File System 2*) [Desarrollado y mantenido por Red Hat]
- Otros: [https://en.wikipedia.org/wiki/Clustered\\_file\\_system#SHARED-DISK](https://en.wikipedia.org/wiki/Clustered_file_system#SHARED-DISK)

## 1 Requisitos y necesidades típicas

## 2 Niveles RAID

- Niveles RAID estándar
- Niveles RAID anidados

## 3 Redes de almacenamiento

- DAS vs SAN vs NAS
- Elementos SAN
- Protocolos SAN
- Sistemas de ficheros sobre SAN

## 4 Copias de seguridad

- Dispositivos
- Estrategias de copia

# Copias de seguridad

Un **backup** o **copia de seguridad** es una réplica exacta de la información resguardada

- instantánea de los datos en un momento determinado
- almacenada en un formato estándar
- es posible su seguimiento (trazabilidad) durante su ciclo de vida

## Necesidad

- Elemento esencial en la recuperación ante catástrofes (DRP, *Disaster Recovery Plan*) y en los planes de continuidad del *negocio* (BCP, *Business Continuity Plan*)
  - Datos son un activo valioso y difícilmente recuperable (recuperación más sencilla con hardware o software)
  - En ocasiones, sujetos a restricciones legales (LOPD, etc)
- Engloba  $\left\{ \begin{array}{l} \text{la copia en sí y su } \mathbf{soporte} \\ \text{planes y procedimientos de } \mathbf{copia} \\ \text{planes y procedimientos de } \mathbf{recuperación} \end{array} \right.$

**Ventana de backup:** Intervalo durante el cual es posible realizar copias de seguridad sin interrumpir/afectar a la operativa normal del sistema

- Tendencia a reducirse  $\Rightarrow$  limita tiempo de copia y condiciona mecanismos y estrategias a usar

## Copias de seguridad (cont.)

### Aspectos a considerar

#### - Tiempo de creación (de copia)

- Tiempo necesario para realizar la copia de seguridad
- Acotado por la ventana de backup y la velocidad de los dispositivos de copia

#### - Tiempo de almacenamiento (de vida, de retención)

- Tiempo máximo que una copia permanece almacenada en un dispositivo
- Al finalizar, copia puede moverse a otro dispositivos (si se usan distintos niveles de backup) o eliminarse liberando espacio de almacenamiento

#### - Tiempo de restauración

- Tiempo necesario para regenerar el sistema a partir de las copias de seguridad una vez producido un fallo

#### - Ubicación de la copia (accesibilidad)

- Varias alternativas:
 

{	<i>on-line</i> [con acceso en línea] <i>near-line</i> [fuera de línea con acceso automático] <i>off-line</i> [fuera de línea con acceso "manual"] <i>off-site</i> [en ubicación separada]
---	--
- Uso de almacenamiento alternativo
  - Copias de seguridad adicionales en ubicaciones externas al sistema (centros de respaldo *off-site*)

#### - Protección ante fallos de dispositivos

- Posibilidad de replicar las copias e seguridad en distintos medios para reducir pérdidas de datos por fallos en dispositivos/soportes de almacenamiento de copias

#### - Coste (criterio clave)

- Diversos aspectos: coste de dispositivos/soportes de almacenamiento, costes de gestión de las copias, coste de la ventana de backup, costes de almacenamiento externo, etc

# Dispositivos para copias de seguridad

## Dispositivos/soportes magnéticos

- Unidades de cinta magnética:
  - Coste reducido, con gran capacidad de almacenamiento
  - Automatización: robots de cintas, bibliotecas de cintas (*tape library*)
  - **Aspecto clave:** acceso secuencial (condiciona modo de realizar copias)
  - Múltiples tecnologías: DDS/DAT (1-160 GB, 3-12 MB/s), LTO (100GB-6TB, 20-300 MB/s)
- Unidades de disco: discos duros y/o discos flexible (en desuso)
  - Acceso directo y mayor velocidad (discos duros > 0,5 GB/s)
  - Capacidad menor y mayor coste

## Dispositivos/soportes ópticos

- CD-ROM, CD-R/RW, DVD-ROM, DVD+/-R/RW, BR, etc.

## Otros

- Unidades magneto-ópticas: iomega ZIP (100 MB - 750 MB)
- Unidades de estado sólido: memorias USB, tarjetas de memoria, discos SSD
- Backups remotos (*cloud storage*)

# Estrategias de copias de seguridad

## Aspectos a definir en las estrategias de backup

- Datos a resguardar
  - depende de la importancia de la información para la organización
- Frecuencia del backup
  - depende de tiempo de creación, del coste de realizar la copia y de las consecuencias de la pérdida de datos no resguardados
- Tiempo de almacenamiento de las copias
  - depende fundamentalmente del dispositivo/soporte de almacenamiento y en parte de los requisitos de la información resguardada

### [1] Copia de seguridad completa

- Es una copia exacta de todos los datos a resguardar en un momento dado
  - Copia y restauración simple
  - Requisitos de almacenamiento pueden ser costosos
  - Moderadamente vulnerable a fallos en los dispositivos/soportes de copia
- Sirve de base a las otras dos estrategias de backup

# Estrategias de copias de seguridad (cont.)

## [2] Copia de seguridad diferencial

- Parte de una copia de seguridad completa (**copia base**)
- Copias sucesivas **sólo** incluyen **cambios respecto a copia completa base**
- Restauración requiere:  $\left\{ \begin{array}{l} \text{copia base} \\ \text{última copia diferencial} \end{array} \right.$ 
  - Usa sólo 2 copias  $\Rightarrow$  falla recuperación sólo si fallan ambas
  - Posibilidad de crecimiento excesivo en el tamaño de las copias diferenciales (degenera en copias totales)

## [3] Copia de seguridad incremental

- Parte de una copia de seguridad completa (**copia base**)
- Copias sucesivas **sólo** incluyen **cambios respecto a copia incremental previa**
- Restauración requiere:  $\left\{ \begin{array}{l} \text{copia base} \\ \text{secuencia completa de copias incrementales} \end{array} \right.$ 
  - Necesita todas las copias  $\Rightarrow$  falla recuperación si falla una de ellas
  - Tamaño contenido de las copias incrementales (menos datos y tiempo de copia)

## Ejemplo

Datos	Copia completa	Copia diferencial	Copia incremental
-----	-----	-----	-----
a b c d e f g h	a b c d e f g h	a b c d e f g h	a b c d e f g h
a M c N e f g h	a M c N e f g h	. M . N . . . .	. M . N . . . .
a M c N e O g h	a M c N e O g h	. M . N . O . .	. . . . O . .
Q M c N e O g P	Q M c N e O g P	Q M . N . O . P	Q . . . . . P
Q M R S T O g P	Q M R S T O g P	Q M R S T O . P	. . R S T . . .
Q M R S T O U P	Q M R S T O U P	Q M R S T O U P	. . . . . U . .
V M R S T O U P	V M R S T O U P	V M R S T O U P	V . . . . . .
V M R X T O U P	V M R X T O U P	V M R X T O U P	. . . X . . . .
V M R X T O Y Z	V M R X T O Y Z	V M R X T O Y Z	. . . . . Y Z