

Tema 7

Almacenamiento de datos



José Manuel Soto Hidalgo
Dpto. Arquitectura y Tecnología de Computadores
Universidad de Granada

jmsoto@ugr.es

Índice



- [1. Introducción]
- 2. Tecnologías hardware para BD
- 3. Tecnología RAID
- 4. SSA
- 5. SAN
- 6. NAS
- 7. Conclusiones

Introducción

El sistema de almacenamiento de datos resulta clave en un sistema web de altas prestaciones.

Parte del sistema donde se guarda la información, ya sea en una BD o en archivos.



Introducción

Diseñar teniendo en mente ciertos requisitos en cuanto a escalabilidad.

Todo usuario que llegue al sistema accederá a los datos almacenados, y debemos estar preparados para servir datos a un número creciente de usuarios.



Introducción

Los gestores de BD y el diseño de éstas deben ser robustas para soportar múltiples accesos concurrentes.

Podemos mejorar las prestaciones de los sistemas de almacenamiento:

- **ampliación vertical** (adquirir un mejor hardware más rápido y actualizado)
- **ampliación horizontal** (replicar el almacenamiento entre varios servidores); puede resultar más efectivo en cuanto a la escalabilidad

Introducción

Posibles problemas de realizar la replicación y repartir la carga:

- El coste de nuevos servidores y almacenamiento.
- La configuración de métodos y rutinas de replicación y sincronización.
- La latencia en los procesos de replicación.
- La necesidad de un sistema de balanceo de carga adecuado entre los servidores de BD.

Introducción

Estrategias alternativas a la replicación completa para mejorar el sistema de almacenamiento y BD:

Realizar distribución funcional:

dividir la BD global en varias secciones relativas a aplicaciones diferentes (p.ej. inventario, usuarios, mensajería, etc).

Configurar varios servidores que hospedarán cada sección de la BD.

Es complicado mantener la integridad de los datos entre las diferentes secciones.

Introducción

Estrategias alternativas a la replicación completa para mejorar el sistema de almacenamiento y BD:

Segmentar la BD:

hacer una división lógica de la BD, p.ej. en función del tipo de clientes o según períodos contables.

Cada segmento queda almacenado en un servidor de BD, quedando repartida así la carga.

Es complicado mantener la integridad de los datos entre las diferentes divisiones.

Introducción

Existen productos de **BD propietarios** en los cuales se pueden usar **extensiones** que facilitan la interacción entre varios servidores para gestionar una sola gran BD.

Suelen depender estrechamente de un sistema operativo o de un sistema de distribución muy concretos.

- Oracle11g <http://www.oracle.com/us/products/database/overview/index.html>
 - SQL Server 2008 R2 <http://msdn.microsoft.com/en-us/library/ms191440.aspx>
 - Apache Cassandra <http://cassandra.apache.org/>
-
- Configurar con MySQL un cluster de BD (**práctica 5**).

Índice



1. Introducción
- [2. Tecnologías hardware para BD]**
3. Tecnología RAID
4. SSA
5. SAN
6. NAS
7. Conclusiones

Tecnologías hardware

El sistema de almacenamiento y de BD es un punto fundamental en cualquier sistema web actual.

Una mala configuración afectará a las prestaciones.

Hay que ser cuidadosos con el hardware y software.

El hardware del resto del sistema web puede actualizarse en cualquier momento casi sin que los usuarios lo noten.

El de la BD es crítico, ya que no se podrá actualizar de forma fácil una vez que esté en funcionamiento.

Tecnologías hardware

Factores a tener en cuenta al diseñar la arquitectura de BD:

- El número de sesiones concurrentes en la BD puede afectar al rendimiento de la granja web completa (conexiones costosas).
- El tipo de accesos a la BD también influye.
- Las búsquedas que devuelvan resultados muy grandes afectarán al rendimiento de CPU, almacenamiento y red.

Tecnologías hardware

Factores a tener en cuenta al diseñar la arquitectura de BD:

- El tamaño total de la BD determinará el espacio para almacenamiento, y el tiempo necesario para hacer copias de seguridad y restaurarlas.
- Conviene utilizar hardware redundante para los servidores.
- Una gran cantidad de accesos a la BD por cada petición HTTP puede sobrecargar la conexión de red entre los servidores web y de BD.

Tecnologías hardware

Factores a tener en cuenta al diseñar la arquitectura de BD:

- Arquitectura de la BD basada en un cluster.
- Una BD se podrá escalar en el futuro si desde el principio se instaló hardware con capacidad de ampliación (CPU, memoria, etc) y se configuró el software de forma adecuada.

Índice



1. Introducción
2. Tecnologías hardware para BD
- [3. Tecnología RAID]
4. SSA
5. SAN
6. NAS
7. Conclusiones

Almacenamiento basado en RAID

¿Qué es RAID y qué no es RAID?



Almacenamiento basado en RAID

¿Qué es RAID y qué no es RAID?

Redundant Array of Independent Disks

<http://es.wikipedia.org/wiki/RAID>



Almacenamiento basado en RAID

RAID (*conjunto redundante de discos independientes*) es un sistema de almacenamiento que usa múltiples discos duros entre los que se distribuyen o replican los datos.

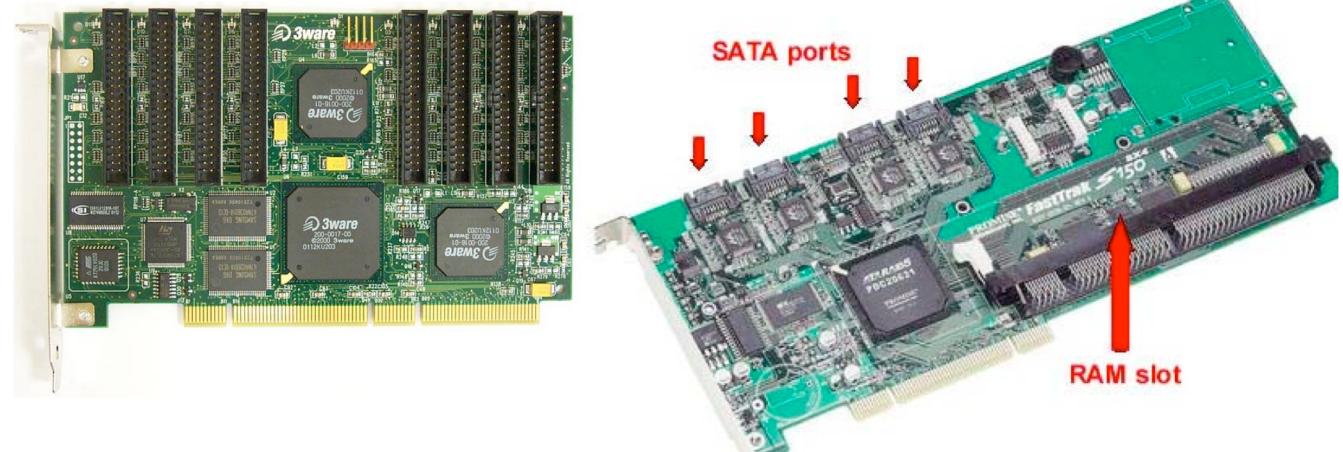
Ofrece mayor integridad, mayor tolerancia a fallos, mayor rendimiento y mayor capacidad.

La idea inicial es combinar varios dispositivos en un conjunto que ofrece mayor capacidad, fiabilidad y velocidad que un solo dispositivo de última generación más caro.

Almacenamiento basado en RAID

Un RAID por hardware es mucho más rápido que uno configurado por software.

From Computer Desktop Encyclopedia
© 2004 The Computer Language Co., Inc.



Por software son mucho más flexibles:

- permiten construir RAID de particiones en lugar de discos completos
- agrupar en un mismo RAID discos conectados en varias controladoras.

Almacenamiento basado en RAID

The image displays two windows illustrating RAID management. On the left, the Windows Computer Management interface shows a list of volumes (E:, F:, H:, M:, N:, System) and a detailed view of Disk 0, Disk 1, and Disk 2, each containing multiple NTFS volumes. On the right, the SoftRAID application shows a graphical representation of three physical disks connected to a single logical volume named 'DataMirror' (size: 999 GB, Mirror, no errors, Safeguard Enabled). A terminal window at the bottom shows the command `sudo mdadm --create /dev/md0 --level=raid1 --raid-device=3 --name=RAID1Array --size=10M --verbose /dev/sdd2 /dev/sde2 /dev/sdf2` being run in an Ubuntu environment.

Volume	Layout	Type	File System	Status	Capacity	Free Sp
New Volume (E:)	RAID-5	Dynamic	NTFS	Healthy	1.95 GB	1.94 GB
New Volume (F:)	Simple	Dynamic	NTFS	Healthy	1.16 GB	1.15 GB
New Volume (H:)	Spanned	Dynamic	NTFS	Healthy	8.97 GB	8.93 GB
New Volume (M:)	Mirror	Dynamic	NTFS	Healthy	1000 MB	992 MB
New Volume (N:)	Striped	Dynamic	NTFS	Healthy	1.95 GB	1.94 GB
System (G:)	Partition	Basic	NTFS	Healthy (System)	7.75 GB	5.32 GB

Disk	Capacity	File Systems
Disk 0	8.00 GB	4x 1000 MB NTFS (Healthy)
Disk 1	8.00 GB	3x 1000 MB NTFS (Healthy)
Disk 2	8.00 GB	1x 1000 MB NTFS (Healthy), 1x Unallocated (7.02 GB)

Legend: █ Unallocated █ Primary partition █ Simple volume █ Spanned volume

SoftRAID

Disks

- SATA bus 0, ID 0 (Upper)
size: 200 GB • available: 134 MB
OWC Mercury Extreme SSD-E v.343A
- SATA bus 0, ID 1 (Bay 2)
size: 1.00 TB • available: 512 bytes
WDC WD1001FALS-41Y6A0 v.05.03D
- SATA bus 0, ID 2 (Bay 3)
size: 1.00 TB • available: 512 bytes
WDC WD1001FALS-40Y6A0 v.05.01D

Volumes

- DataMirror**
size: 999 GB • available: 999 GB • HF: Mirror
no errors • Safeguard Enabled
- disk identifier: disk3
total bytes: 999,851,249,664
Optimized for Workstation
created: Feb 4, 2011 2:06:41 PM
last validated: never
I/O requests: 1,059 • I/O errors: -
- Mule** (10.6.6 startup)
size: 199 GB • available: 160 GB • HF: non-RAID

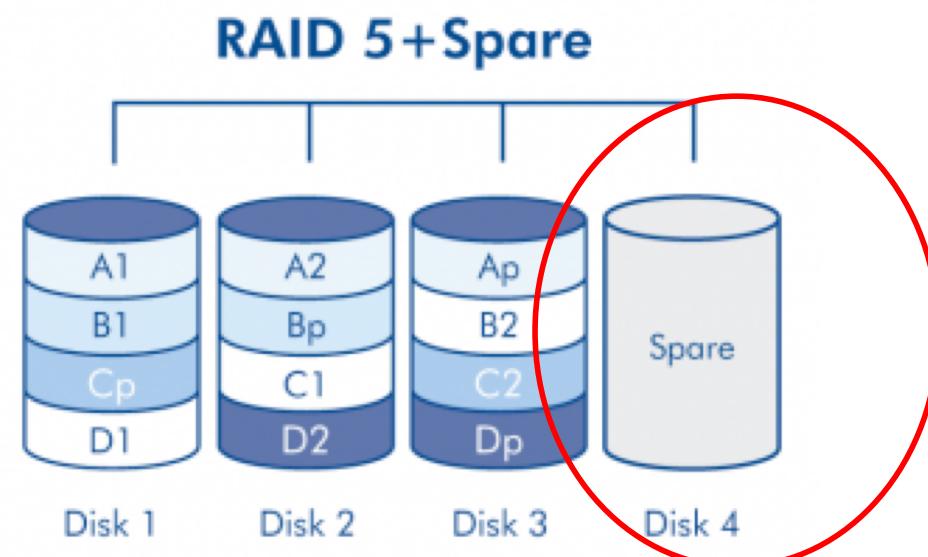
SoftRAID v 4.0.1 • Driver v 4.0.1

```
ubuntu@vb-ubuntu1310:~$ sudo mdadm --create /dev/md0 --level=raid1 --raid-device=3 --name=RAID1Array --size=10M --verbose /dev/sdd2 /dev/sde2 /dev/sdf2
[sudo] password for ubuntu:
mdadm: /dev/sdd2 appears to contain an ext2fs file system
      size=102400K  mtime=Thu Jan  1 09:00:00 1970
```

Almacenamiento basado en RAID

La tecnología RAID soporta el uso de varios discos de reserva (hot spare), para usarse inmediatamente y de forma automática tras el fallo de uno de los discos.

Esto reduce el tiempo del período de reparación al acortar el tiempo de reconstrucción del RAID.



Almacenamiento basado en RAID

Niveles RAID

Hay diversos métodos de almacenamiento, llamados niveles, con diferente complejidad:

- RAID 0: Conjunto dividido
- RAID 1: Conjunto en espejo
- RAID 5: Conjunto dividido con paridad distribuida

http://en.wikipedia.org/wiki/Standard_RAID_levels

Podemos anidar niveles RAID: que un RAID pueda usarse como elemento básico de otro en lugar de discos físicos.

Almacenamiento basado en RAID

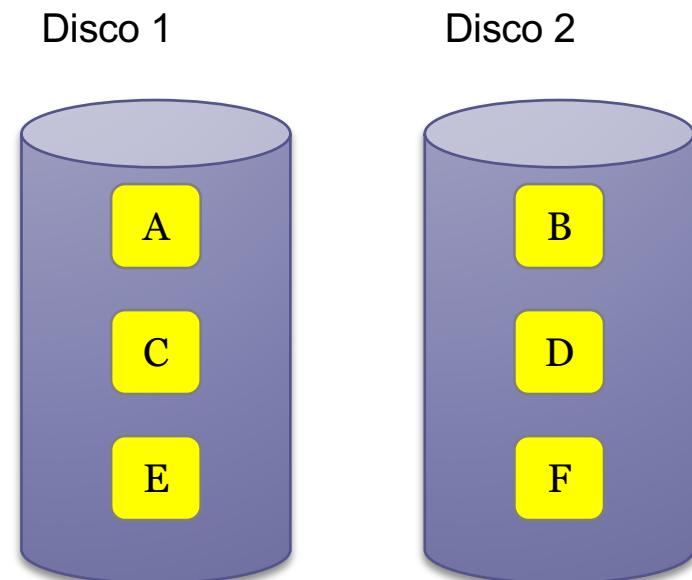
RAID 0

Reparte los datos entre varios discos => incremento de la velocidad de lectura y escritura.

Se puede acceder a varios bloques consecutivos al mismo tiempo.

Esta configuración **no ofrece protección contra fallos** en los discos, ya que no se escribe información duplicada o información de paridad.

striping



Almacenamiento basado en RAID

RAID 0

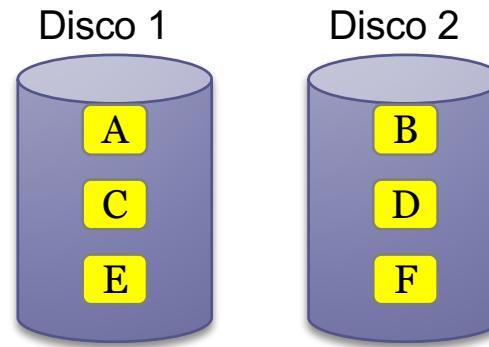
La velocidad de transferencia (ideal) se puede ver como la suma de las velocidades de transferencia de todos los discos.

Se suele usar en configuraciones de servidor NFS.

Almacenamiento basado en RAID

Ejercicio:

¿Qué tamaño de unidad de unidad RAID se obtendrá al configurar un RAID 0 a partir de dos discos de 100 GB y 100 GB?



¿Qué tamaño de unidad de unidad RAID se obtendrá al configurar un RAID 0 a partir de tres discos de 200 GB cada uno?

Almacenamiento basado en RAID

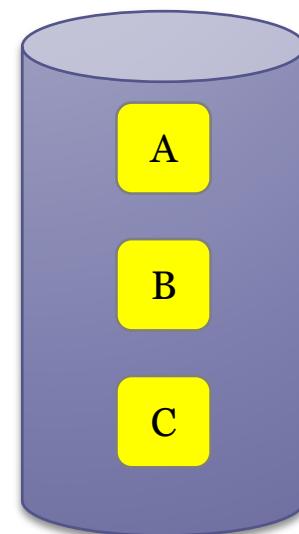
RAID 1

Crea una copia exacta (o espejo) de un conjunto de datos en dos o más discos. Ofrece gran fiabilidad, ya que para que el conjunto falle es necesario que lo hagan todos sus discos.

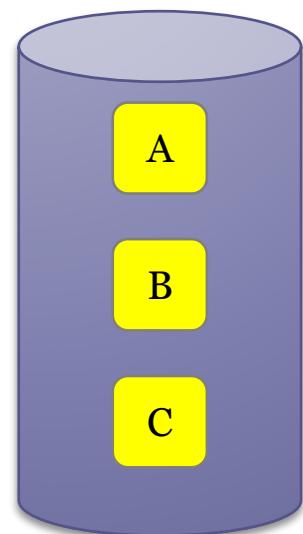
Como los discos que forman el RAID 1 tienen hardware independiente, se puede leer simultáneamente dos datos diferentes en dos discos diferentes, por lo que su rendimiento se duplica.

mirroring

Disco 1



Disco 2



Almacenamiento basado en RAID

RAID 1

Útil si la seguridad de los datos es más importante que la capacidad de almacenamiento total.

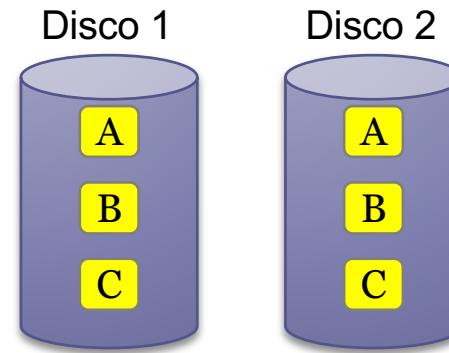
Se recomienda tener controladoras independientes para cada disco.

Ventajas desde el punto de vista administrativo: se puede poner un disco inactivo para hacer backup de los datos, mientras que el otro sigue dando servicio.

Almacenamiento basado en RAID

Ejercicio:

¿Qué tamaño de unidad de unidad RAID se obtendrá al configurar un RAID 1 a partir de dos discos de 100 GB y 100 GB?



¿Qué tamaño de unidad de unidad RAID se obtendrá al configurar un RAID 1 a partir de tres discos de 200 GB cada uno?

Almacenamiento basado en RAID

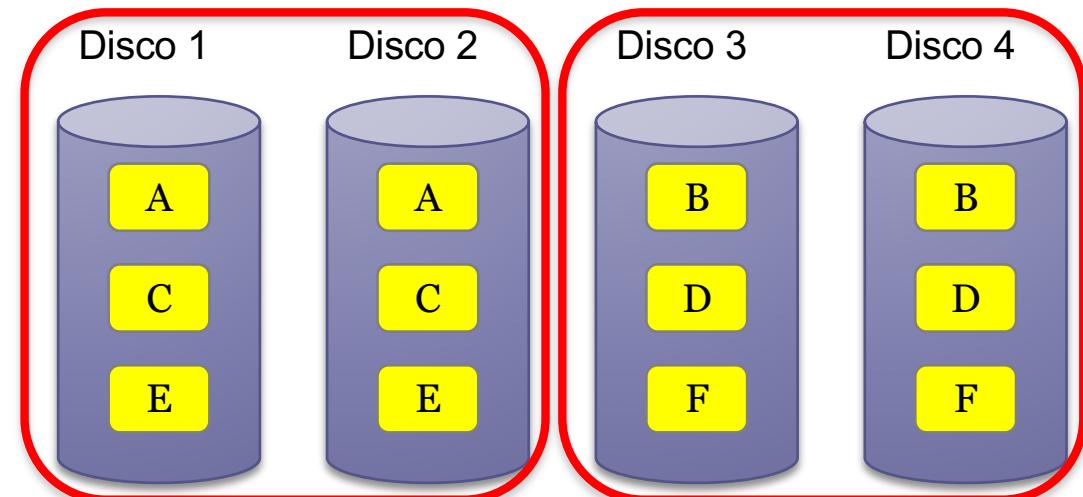
RAID 10 o RAID 1+0 es una **división de espejos**

La **duplicación** (RAID 1) significa grabar los datos en dos o más discos al mismo tiempo.

Si un disco falla por completo, la duplicación conserva la información.

El **reparto de los datos** (RAID 0) divide los datos en fragmentos y los graba en conjuntos distintos de forma sucesiva.

Mejora el rendimiento porque el equipo puede obtener datos de más de un conjunto a la vez.



RAID Level	Description	Performance and Fault Tolerance
10 (1+0)	RAID 0 (stripe) built with RAID 1 (mirror) arrays	<p>RAID 1+0 provides high levels of I/O performance, data redundancy, and disk fault tolerance. Because each member device in the RAID 0 is mirrored individually, multiple disk failures can be tolerated and data remains available as long as the disks that fail are in different mirrors.</p> <p>You can optionally configure a spare for each underlying mirrored array, or configure a spare to serve a spare group that serves all mirrors.</p>
10 (0+1)	RAID 1 (mirror) built with RAID 0 (stripe) arrays	<p>RAID 0+1 provides high levels of I/O performance and data redundancy, but slightly less fault tolerance than a 1+0. If multiple disks fail on one side of the mirror, then the other mirror is available. However, if disks are lost concurrently on both sides of the mirror, all data is lost.</p> <p>This solution offers less disk fault tolerance than a 1+0 solution, but if you need to perform maintenance or maintain the mirror on a different site, you can take an entire side of the mirror offline and still have a fully functional storage device. Also, if you lose the connection between the two sites, either site operates independently of the other. That is not true if you stripe the mirrored segments, because the mirrors are managed at a lower level.</p> <p>If a device fails, the mirror on that side fails because RAID 1 is not fault-tolerant. Create a new RAID 0 to replace the failed side, then resynchronize the mirrors.</p>

Almacenamiento basado en RAID

¿Qué puede y qué no puede hacer RAID?



- Permite acceder a los datos aunque falle un disco.
- Puede mejorar el rendimiento de ciertas aplicaciones (para archivos grandes mantiene tasas de transferencia altas).



- No protege los datos (p.ej. por virus).
- No simplifica la recuperación de un desastre.
- No mejora el rendimiento para todas las aplicaciones.
- No facilita el traslado del almacenamiento a un sistema nuevo.

Tutoriales

Instalación Ubuntu server con RAID1

<http://www.youtube.com/watch?v=DS4uKJ9pfnk>

Instalación ubuntu server 12.04 LTS precise pangoline con RAID 1 software

<http://www.youtube.com/watch?v=y17EfNs0TBc>

Como crear un RAID 1 en Windows

<http://www.youtube.com/watch?v=g5I-1IXgwRo>

Raid 1 - Sincronización espejo en Windows Server 2008

<http://www.youtube.com/watch?v=k92yKphhKYE>

Instalación y configuración RAID de dos discos duros en un Mac PRo

<http://www.youtube.com/watch?v=O5VuJSRjLT8>

Xserve RAID Install

<http://www.youtube.com/watch?v=WxsQ2Y1iW7w>

Índice

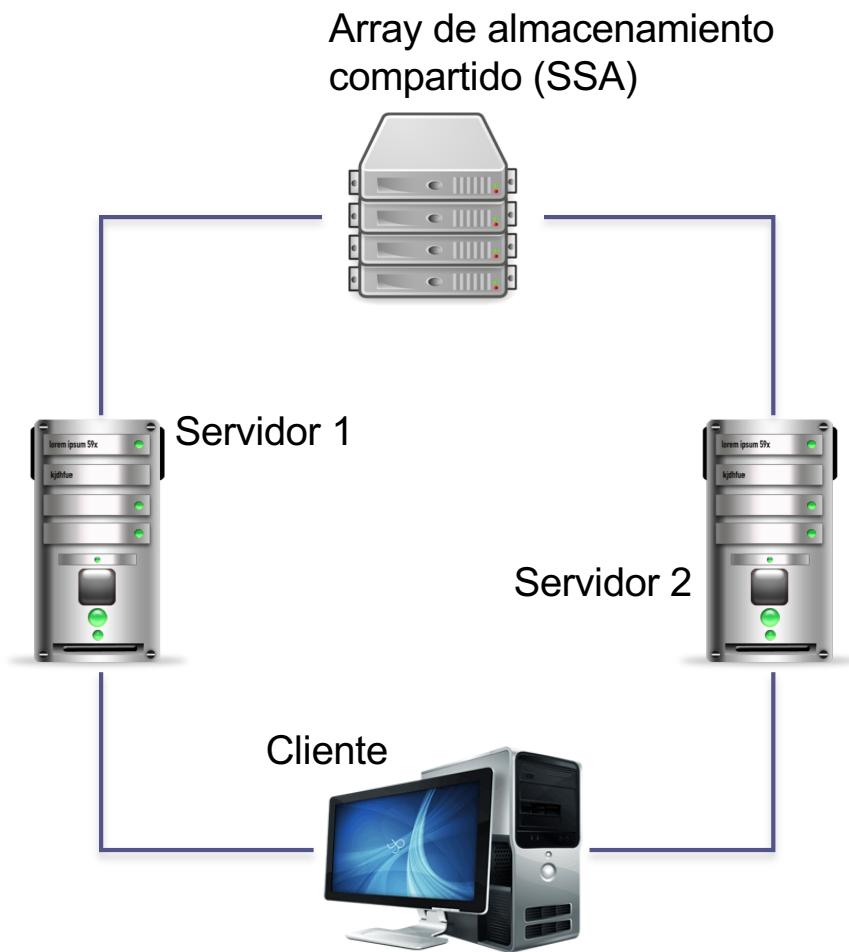


1. Introducción
2. Tecnologías hardware para BD
3. Tecnología RAID
4. SSA
5. SAN
6. NAS
7. Conclusiones

Arrays de almacenamiento compartido: SSA

Forma simple de almacenamiento externo.

Dispositivo específico que incluye varios discos en rack:



Arrays de almacenamiento compartido: SSA

- Posee una interfaz para conectar los discos a las controladoras (normalmente SCSI).
- Número limitado de puertos para hacer la conexión entre servidores y almacenamiento.
- Se suele usar para disponer del almacenamiento necesario para archivos y BD en clusters.
- La posibilidad de manejo y la flexibilidad de un SSA es limitada. Aceptan cambios en caliente de discos y varias configuraciones RAID.
- Dispositivos desarrollados por una empresa con unas especificaciones y herramientas propietarias.



Índice

1. Introducción
2. Tecnologías hardware para BD
3. Tecnología RAID
4. SSA
5. SAN
6. NAS
7. Conclusiones

Área de almacenamiento en red: SAN

Red de almacenamiento especializada que conecta dispositivos de almacenamiento a los servidores.

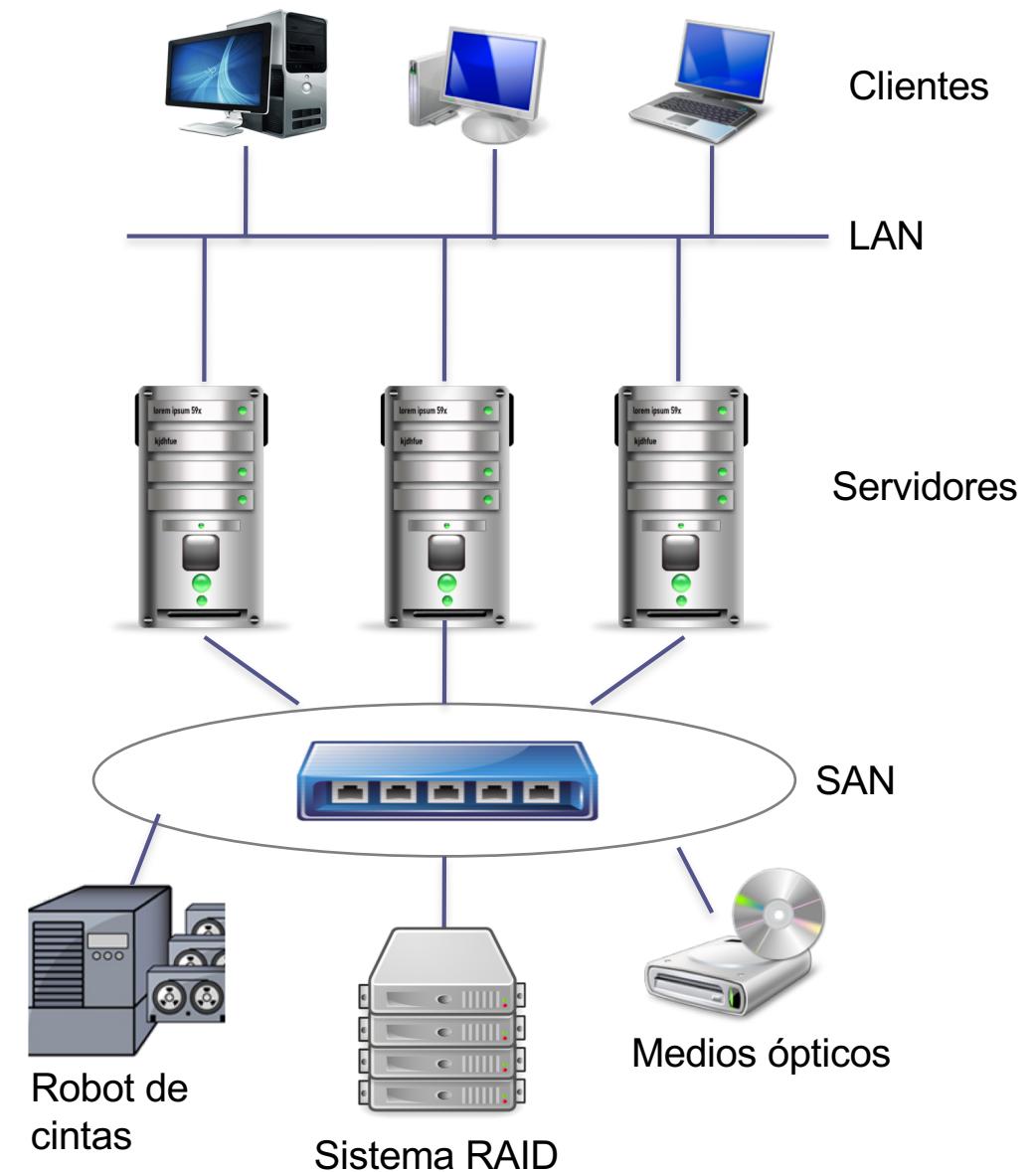
Conjunto de dispositivos interconectados (discos, cintas, etc.) y servidores conectados a un canal de comunicación e intercambio de datos común (concentrador de alta velocidad).

Área de almacenamiento en red: SAN

Esquema:

Gran flexibilidad y
facilidad de manejo del
almacenamiento.

Se puede actualizar
cualquier componente.



Área de almacenamiento en red: SAN

- Red de alta velocidad (mínimo de 1Gbps). Es como un bus de un ordenador, pero compartido entre varias máquinas.
- Utiliza hardware de red muy especializado.
- Una SAN ofrece una capa de abstracción entre los dispositivos de almacenamiento y los servidores, y permite que el espacio físico de almacenamiento crezca.
- Se puede usar para almacenar archivos, compartir datos entre los servidores, mirroring de discos y backups.
- Puede operar con SSA y NAS.
- Permite que se añadan nuevos dispositivos al sistema (servidores o almacenamiento).

Índice



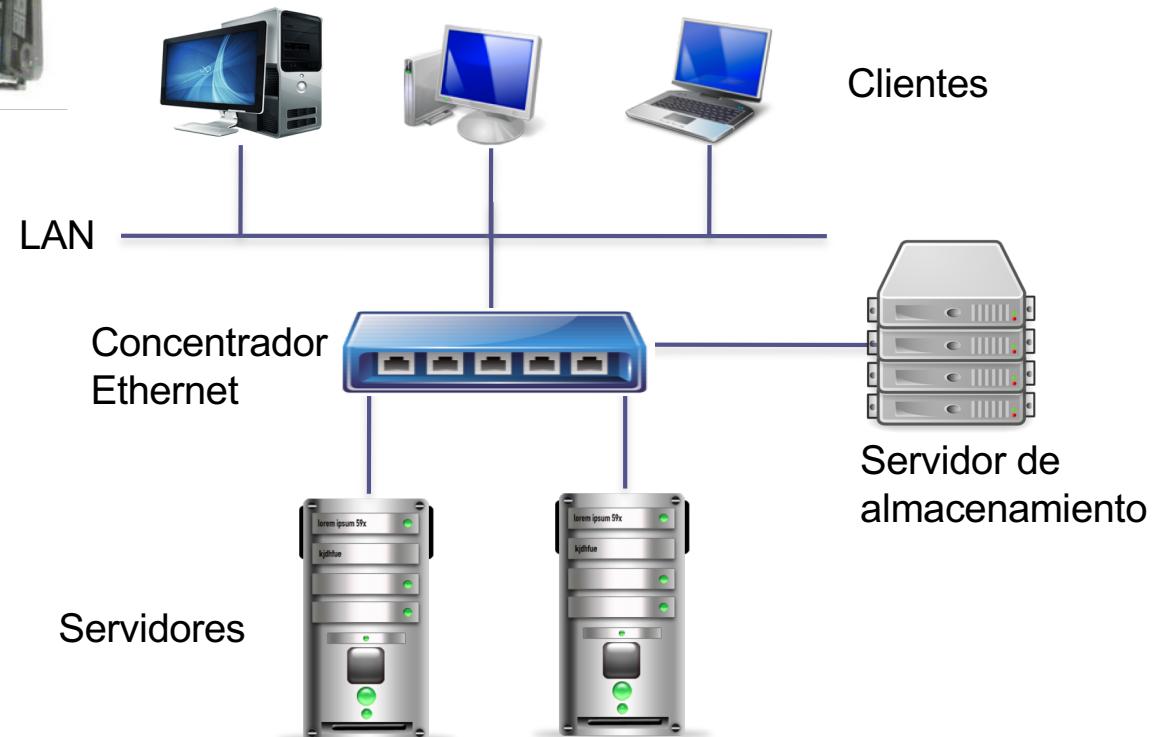
1. Introducción
2. Tecnologías hardware para BD
3. Tecnología RAID
4. SSA
5. SAN
6. NAS
7. Conclusiones

Almacenamiento conectado a la red: NAS

Dispositivo que actúa como un servidor de ficheros, pero ahorrando los recursos de tener una máquina más.



para almacenar copias de seguridad, y para ofrecer espacio de almacenamiento compartido



Almacenamiento conectado a la red: NAS

- Conjunto de discos organizados en un dispositivo de red con IP y que puede conectarse a una red Ethernet.
- Utilizando algún protocolo, como Internetwork Packet Exchange (de Microsoft), NetBEUI (de Microsoft), Network File System (NFS, de Sun) o IPE (de Novell).
- Aparece como otro servidor más en la red.
- Usan software específico para configurarlos y manejarlos (creación de unidades, gestión de permisos, etc).
- Utilizan configuraciones RAID.

Almacenamiento basado en RAID

Ejercicio 7.1:

¿Qué es Lustre File System?

¿Qué ventajas tiene? ¿Qué arquitectura plantea?

Ejemplo de NAS. *openmediavault*

Sistema de almacenamiento en red basado en Debian:

<http://www.openmediavault.org/>

<http://en.wikipedia.org/wiki/OpenMediaVault>

Distribución Linux basada en Debian pensada para configurar un NAS con un PC.

Servicios: ssh, sftp, smb/cifs, rsync

Requisitos hardware: 1GByte de RAM, 2 GByte de disco para el sistema operativo, y los discos duros que usen para el servicio de almacenamiento en red.

<https://ostechnix.wordpress.com/2013/01/17/openmediavault-setup-your-own-nasnetwork-attached-storage-box-in-minutes/>

Ejemplo de NAS. *openmediavault*

The screenshot shows the OpenMediaVault web interface. The top navigation bar includes the logo "OpenMediaVault" and the tagline "The open network attached storage solution". A search bar and language selection ("English") are also present. The main content area is titled "Storage | Physical Disks". On the left, a sidebar menu lists several categories: System (General Settings, Date & Time, Network, Notification, Power Management, Certificates, Cron Jobs, Update Manager, Plugins), Storage (Physical Disks, RAID Management, Filesystems, S.M.A.R.T.), Access Right Management (User, Group, Shared Folders), and Services (TFTP, SNMP, SSH). The "Physical Disks" item under Storage is currently selected. The main panel displays a table of physical disks:

Device	Model	Serial Number	Vendor	Capacity
/dev/sda	TRANSCEND	[REDACTED]	ATA	3.73 GiB
/dev/sdb	WDC WD15EARS-00Z5B1	[REDACTED]	ATA	1.36 TiB
/dev/sdc	WDC WD15EARS-00Z5B1	[REDACTED]	ATA	1.36 TiB
/dev/sdd	WDC WD15EARS-00Z5B1	[REDACTED]	ATA	1.36 TiB
/dev/sde	WDC WD15EARS-00MVWB0	[REDACTED]	ATA	1.36 TiB
/dev/sdf	WDC WD15EARS-00MVWB0	[REDACTED]	ATA	1.36 TiB

At the bottom of the table, a message states "Displaying items 1 - 6 of 6".

Almacenamiento conectado a la red

Ejercicio T7.2:

Buscar información sobre los sistemas de ficheros en red más utilizados en la actualidad y comparar sus características. Hacer una lista de ventajas e inconvenientes de todos ellos, así como grandes sistemas en los que se utilicen.



Índice

1. Introducción
2. Tecnologías hardware para BD
3. Tecnología RAID
4. SSA
5. SAN
6. NAS
- [7. Conclusiones]**

Conclusiones

La arquitectura de almacenamiento de la granja web resulta fundamental para la disponibilidad de las aplicaciones.

El escalado del sistema de BD resultará muy importante a lo largo de la vida del sistema web.

La configuración de un sistema RAID supondrá una mejora en la disponibilidad y en la seguridad de nuestros datos.

Además, se mejorará la capacidad del sistema de entrada/salida a disco.

Conclusiones

Otras soluciones pasan por la instalación y configuración de **sistemas avanzados de almacenamiento** usando tecnologías de red (SSA, SAN y NAS).

Estas tecnologías ofrecen **flexibilidad y la posibilidad de escalar** el sistema de almacenamiento en el futuro.

En resumen, la **arquitectura de BD** del sistema web debe ser lo más robusta posible, con capacidad para crecer (**ampliable y escalable**).

Convendrá realizar un buen análisis, adquirir un buen hardware e instalar un buen software al principio.