

# Tema 8

## Requisitos hardware y software en servidores web (1ª parte)



Pedro A. Castillo Valdivieso  
Depto Arquitectura y Tecnología de Computadores  
Universidad de Granada  
[pacv@ugr.es](mailto:pacv@ugr.es)

# Índice

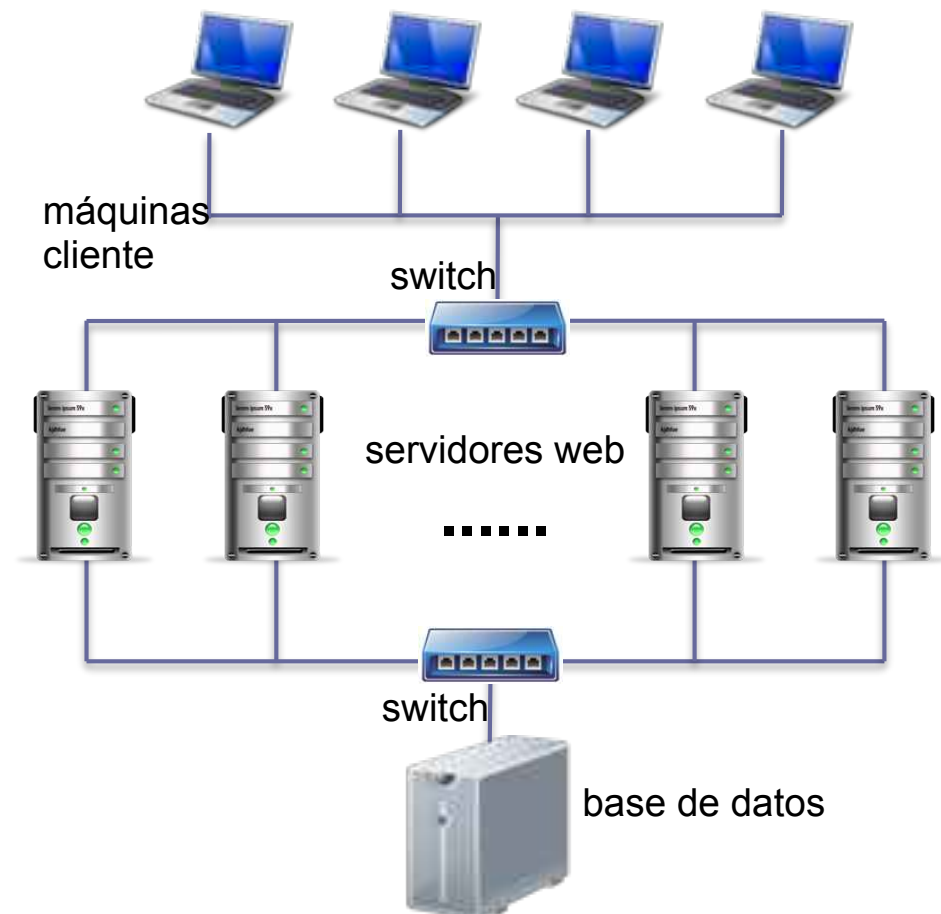
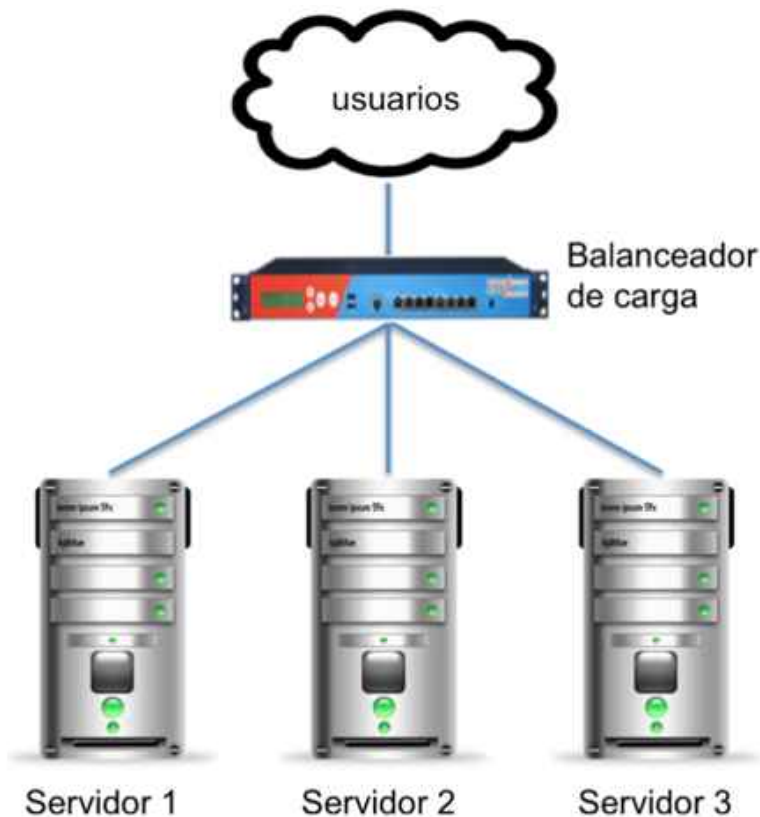
## [ 1. Introducción ]



2. Elementos de la granja web
3. Necesidades del servidor web
4. Hardware para servidores
5. Software para servidores
6. Conclusiones

# Introducción

Usar **clústers web** incrementa la capacidad de ofrecer servicios usando tecnologías hardware y software estándar.



# Introducción

## Clasificación:

- High Performance Computing Clusters (HPCC) o clústers de alto rendimiento.
- High Availability Computing Clusters (HACC) o clústers de alta disponibilidad.
- High Throughput Computing Clusters (HTCC) o clústers de alta eficiencia.

# Índice



1. Introducción

[ 2. Elementos de la granja web ]

3. Necesidades del servidor web

4. Hardware para servidores

5. Software para servidores

6. Conclusiones

# Elementos de la granja web

La granja web necesita varios componentes de software y hardware para poder funcionar.

En cuanto al hardware, cabe destacar:

- servidores
- almacenamiento
- conexiones de red

El software de la granja web es muy importante (última sección del tema).

# Elementos de la granja web

## Servidores

Los servidores web no necesitan una capacidad de procesamiento especialmente alta, ya que los servicios **HTTP no consumen demasiada CPU**.

La cantidad de **memoria RAM** es más **importante** en este tipo de máquinas.

Algunas páginas en un sitio serán servidas mucho más que el resto.

Conviene ***cachear*** las imágenes más comunes. Así las páginas que las muestren se servirán más rápido, haciendo que el tiempo medio de servicio sea más rápido.

# Elementos de la granja web

## Servidores

Los sitios web siguen la regla 80-20: el 80% del tráfico corresponde al 20% de las páginas almacenadas.

Si *cacheamos* ese 20% de páginas en RAM, haremos que nuestros clientes tengan tiempos de servicio más rápidos.

Debemos asegurarnos de tener máquinas con gran cantidad de RAM, expandible a un máximo que vendrá dado en función del tamaño estimado de ese conjunto de páginas más populares.



# Elementos de la granja web

## Servidores

Al adquirir el **hardware**, debemos elegir componentes para minimizar la latencia de disco, así como CPU y RAM que **excedan las necesidades de ese momento**.

Dependiendo del peso de las aplicaciones, para sitios con menos de un millón de páginas servidas al mes, una máquina tipo i7 con 16GB de RAM y un SO bien configurado puede ser suficiente.

Para volúmenes de tráfico mayores, esa configuración puede ser insuficiente.

# Elementos de la granja web

## Almacenamiento

Se pueden evitar latencias usando **discos de estado sólido** de alta velocidad (para hacer *caching*).

El almacenamiento puede consistir en una NAS (Almacenamiento conectado a la red), una SAN (Área de almacenamiento en red), o almacenamiento interno en el servidor.

El protocolo más comúnmente utilizado es NFS.

# Elementos de la granja web

## Almacenamiento

- **Arrays de almacenamiento compartido: SSA**  
Dispositivo específico que incluye varios discos en rack
- **Área de almacenamiento en red: SAN**  
Red de almacenamiento especializada que conecta dispositivos de almacenamiento a los servidores (discos, cintas, servidores, etc.).
- **Almacenamiento conectado a la red: NAS**  
Dispositivo que actúa como un servidor de ficheros, pero ahorrando los recursos de tener una máquina más.

# Elementos de la granja web

## Almacenamiento

Tecnologías para la conexión de discos internos:

- IDE o ATA: velocidades de 33, 66, 100, 133 y 166 MB/s
- SATA: velocidades de 150, 300 y 600 MB/s
- SCSI: velocidades de 160, 320, 640 MB/s. Proporciona altos rendimientos.
- SAS: aúna SATA2 y SCSI. Velocidades de 300 y 600 MB/s
- Las unidades de cinta (DLT) son utilizadas para copias de seguridad por su bajo coste.

# Elementos de la granja web

## Conexiones de red

Los nodos del clúster pueden conectarse mediante una simple red Ethernet con tarjetas de red comunes, o utilizarse tecnologías especiales de alta velocidad:

- Fast Ethernet,
- Gigabit Ethernet,
- Myrinet,
- InfiniBand,
- SCI

# Elementos de la granja web

## Conexiones de red. Ethernet

Son las redes **más utilizadas** en la actualidad, debido a su relativo bajo coste. Su tecnología limita el tamaño de los paquetes. Realizan excesivas comprobaciones de error y sus **protocolos no son eficientes**, y sus velocidades de transmisión pueden limitar el rendimiento de los clústeres.

La opción más utilizada en la actualidad es **Gigabit Ethernet** (1 Gbit/s), siendo emergente la solución 10 Gigabit Ethernet (10 Gbit/s). La latencia de estas tecnologías está en torno a los 30 a 100  $\mu$ s, dependiendo del protocolo de comunicación empleado.

Aunque no sea la mejor solución de red para un sistema de altas prestaciones, se suele utilizar para las tareas administrativas.

# Elementos de la granja web

## Conexiones de red. Myrinet (Myrinet 2000 y Myri-10G)

Su latencia es de 99 a 10  $\mu$ s, y su ancho de banda es de 2 a 10 Gbit/s.

Es la red de baja latencia más utilizada en la actualidad, tanto en clústeres como en MPP; está presente en más de la mitad de los sistemas del top500.

Tiene dos bibliotecas de comunicación a bajo nivel (GM y MX). Sobre estas bibliotecas están implementadas MPICH-GM, MPICH-MX, Sockets-GM y Sockets MX, para aprovechar las excelentes características de Myrinet. Existen también emulaciones IP sobre TCP/IP, IPoGM e IPoMX

# Elementos de la granja web

## Conexiones de red. InfiniBand

Es una red surgida de un estándar desarrollado específicamente **para realizar la comunicación en clústers**. Una de sus mayores ventajas es que mediante la **agregación de canales** (x1, x4 y x12) permite obtener anchos de banda muy elevados. La conexión básica es de 2 Gbit/s efectivos y con 'quad connection' x12 alcanza los 96 Gbit/s.

Define una conexión entre un nodo de computación y un nodo de I/O. La conexión va desde un Host Channel Adapter (HCA) hasta un Target Channel Adapter (TCA). Se está usando principalmente para acceder a arrays de discos SAS.



# Elementos de la granja web

## Conexiones de red. SCI (*scalable coherent interface*)

IEEE standard 1596-1992

Latencia teórica es de 1,43  $\mu$ s y ancho de banda de **5333 Mbit/s bidireccional**.

Topologías de anillo (1D), toro (2D) e hipercubo (3D) sin necesidad de switch, se tiene una red adecuada para clústers de pequeño y mediano tamaño.

El software sobre SCI está menos desarrollado que sobre Myrinet, pero los rendimientos obtenidos son superiores.

# Tipos de conexión de red

## Ejercicio T8.1:

*Buscar información sobre tipos de conexiones de red utilizados especialmente en arquitecturas de servidores web de altas prestaciones.*

# Índice



1. Introducción
2. Elementos de la granja web
- [ 3. Necesidades del servidor web ]**
4. Hardware para servidores
5. Software para servidores
6. Conclusiones

# Necesidades del servidor web

Determinar las necesidades de un servidor es muy complejo.

El diseño debe permitir el **escalado del sistema**

y

debemos realizar continuamente **monitorización** para determinar cuándo y dónde escalar el sistema.

# Necesidades del servidor web

Podemos pensar que tener una sola máquina servidora no es la mejor opción (falta de redundancia). Sin embargo, antes de nada debemos determinar los requisitos reales del sitio.

El sistema de base de datos debería residir en una máquina separada, protegida con un cortafuegos.

En cuanto a la seguridad, debemos determinar qué software/servicios dejar disponibles:

p.ej. puede ser una mala idea colocar un software para blogs como wordpress en la misma máquina que el software y servicios de comercio electrónico.

# Necesidades del servidor web

Para **planear la capacidad** del servidor (necesidades en cuanto a hardware y software) **no existe una fórmula.**

Hay que hacer un **primer estudio**, instalar y **configurar**, y una vez esté en funcionamiento, **monitorizar** constantemente el sistema para determinar si funciona correctamente.

No se pueden hacer suposiciones del tipo

*“como la máquina va a servir páginas estáticas, necesitaremos una potencia de CPU de X”, o bien*

*“como el tamaño medio de las páginas que vamos a servir será de 25KB, el ancho de banda que vamos a necesitar será Y”.*

# Necesidades del servidor web

Hay muchas variables a tener en cuenta => tarea compleja.

Hay tres tareas que pueden ayudar a determinar los requisitos de hardware del sistema:

1. Estimar el hardware en función de las necesidades de la empresa.
2. Utilizar un software de monitorización para buscar cuellos de botella en la configuración inicial.
3. Monitorizar el sistema durante todo el tiempo que esté en uso.

# HW: Necesidades de la empresa

Las transacciones web consisten en una petición de información y su correspondiente servicio (en un tiempo razonable).

Estas comunicaciones pueden servir contenido estático o dinámico.



# HW: Necesidades de la empresa

Servir contenido estático consume pocos recursos, ya que servir ficheros es una tarea ligera, además pueden estar en caché, y probablemente no haya que controlar información de estado de navegación.

Servir contenido estático presenta **pocos riesgos** en cuanto a la seguridad y el mantenimiento, la monitorización o depuración a realizar son casi nulas.

```
<html>
<body>
  Contenido estático en <BR/>
  un documento HTML.
</body>
</html>
```

# HW: Necesidades de la empresa

Servir contenido dinámico implica **ejecutar código** de los programas: incremento del uso de la CPU, memoria y ancho de banda.

**No se puede almacenar contenido en caché.**

Podemos ampliar la memoria, instalar controladoras más rápidas para los discos e incluso discos SSD.

Hay más **riesgos** en cuanto a la **seguridad** y suponen un mayor trabajo de depuración y mantenimiento.

```
<?php>
if ($numResultados > 0) {
    while($fila=mysql_fetch_assoc($resultado)){
        echo " Id " . $fila['id'] . " <br>";
        echo " Nombre " . $fila['nombre'] . "<br>";
    }
}
```

# Estimación de necesidad de CPU

**La máquina servidora llevará a cabo las siguientes tareas:**

- Ejecutar un sistema operativo
- Ejecutar servicios que no tienen que ver con el servidor web.
- Ejecutar los servicios del servidor web.
- Ejecutar programas externos necesarios para generar el contenido dinámico.

El sistema operativo debe consumir cerca del 10% de la capacidad de procesamiento del sistema.

Los procesos del servidor web necesitan poca capacidad de procesamiento.

El cuello de botella se presenta cuando el servidor recibe un alto número de conexiones.

# Estimación de necesidad de CPU

Las aplicaciones adicionales que necesita el servidor web para generar el contenido estático necesitan más CPU.

El desarrollo de aplicaciones PHP o CGI necesitan un intérprete externo al servidor para procesar el programa y generar el contenido que se enviará al cliente

No hay dos aplicaciones web con los mismos requisitos.  
Una misma aplicación puede necesitar diferente CPU en diferentes momentos.

# Estimación de necesidad de MEMORIA

El uso de memoria sigue patrones **similares** a los comentados **respecto a la CPU**. Así, el software en ejecución que demandará memoria es:

- El sistema operativo.
- Servicios adicionales al servidor web.
- Los servicios del servidor web.
- Programas externos necesarios para generar el contenido dinámico.

Un servidor con una gran capacidad de memoria puede albergar una caché de suficiente tamaño para almacenar y servir los contenidos estáticos más comunes.

# Estimación de necesidad de ANCHO DE BANDA

El ancho de banda debe soportar el máximo de peticiones HTTP en momentos de picos de carga. Para estimar el ancho de banda, primero hay que determinar:

- El número de clientes que se conectarán al servidor por segundo.
- El número de bytes que el cliente enviará al servidor en cada petición.
- El número de bytes que el servidor enviará al cliente en cada petición.
- El porcentaje del ancho de banda que se reserva para el servidor web.

Multiplicar el número de clientes/seg por el total de bytes transferidos en cada petición, y dividirlo entre el porcentaje de ancho de banda reservado para el servidor.

La conexión debe tener más del doble del ancho de banda del que hayamos estimado en este punto.

# Buscar cuellos de botella

Una vez que hemos montado el sistema, importante **monitorizar cómo está funcionando**, si cubre nuestras necesidades, y **detectar cuellos de botella**.

Realizar **tests de carga** con benchmarks (apache benchmark, httpperf, u openwebload) o poner el sistema en producción de forma controlada (accesible a **betatesters**) para monitorizar y analizar cómo se comporta:

<http://www.thegeekstuff.com/2011/07/iostat-vmstat-mpstat-examples/>

<http://www.cyberciti.biz/tips/top-linux-monitoring-tools.html>

<http://www.tecmint.com/command-line-tools-to-monitor-linux-performance/>

# Buscar cuellos de botella

**Analizar los logs del servidor** examinando el uso de la memoria, CPU y conexión:

- **Memoria:** poner una caché grande para servir contenido estático. Si falta memoria, incrementar RAM y optimizar los objetos (gráficos).
- **CPU:** vigilar la carga para mantenerla sobre el 50% (nunca cerca del 100%).
- **Conexión:** el uso de red no debe ser cercano al 100%. Debe aceptar picos de demanda en los que el uso de la red puede verse duplicado.



# Monitorizar el resto del tiempo

Una vez en producción, debemos seguir monitorizando.

Nos permitirá obtener la disponibilidad del sitio y de las aplicaciones que sirve.

Debemos tener en cuenta los siguientes aspectos:

- El número de usuarios tenderá a crecer, y esperarán tener un servicio cada vez mejor.
- El ancho de banda que las aplicaciones usarán crecerá (los usuarios enviarán y demandarán más información).

# Monitorizar el resto del tiempo

Almacenar en los logs sólo lo que realmente se necesita  
Usar una localización (partición) segura y amplia.

El **análisis de los logs** nos permitirá determinar qué contenido es el más demandado (hacerlo eficiente).  
También nos permitirá hacer análisis del tipo de usuario y de los patrones de navegación (para marketing).

Podremos identificar fallos (errores tipo 404 ó 500).

A partir de los logs podremos descubrir posibles ataques.

# Ejemplo de HW para servidor medio

Podemos encontrar **ejemplos** y “**consejos**” acerca de qué hardware es adecuado para un servidor web:

Rackform nServ A161

Opteron 6128 2.0GHz, 8-Core

16GB (4 x 4GB) DDR3-1333MHz

2 x Intel 82574L Gigabit Ethernet Controllers

Integrated IPMI 2.0 with Dedicated LAN

LSI 9260-4i 6Gb/s SAS/SATA RAID

4 X 1TB Seagate Constellation ES

Optical Drive: Low-Profile DVD+/-RW Drive

350W Power Supply

Gigabit LAN

También existen **herramientas propietarias** que pueden ayudar en la configuración de ciertos entornos, como en el caso de HP Sizer para Microsoft SharePoint.

# Herramientas automáticas para determinar las necesidades del servidor

## HP Sizer para Microsoft SharePoint



HP Sizer for Microsoft SharePoint 2010



<http://bit.ly/bG5vAh>  
<http://bit.ly/178nGXX>  
<http://bit.ly/cz3ZqH>  
<http://bit.ly/16lCa6P>

Profile	Intel	Price
<b>Server solutions with SAN</b>		
		<b>\$77,316</b>
<b>Recommended Configuration</b>		
Web Front End and Query Search Server	1 x ProLiant BL480c Gen8 Server 2P Intel® Xeon® E5-2630 v2 2.93GHz / 15MB Cache 16,384 MB RAM (Per Server) Disk (DAS) - 2 Disks Spare (0) (Per Server) System Drive (SS) - 2 Disks (72GB 6G SAS 15K2.5m SEC ENT HDD) RAID10 Spare (0)	
Index Search Server	1 x ProLiant BL480c Gen8 Server 2P Intel® Xeon® E5-2630 v2 2.93GHz / 15MB Cache 16,384 MB RAM (Per Server) Disk (DAS) - 2 Disks Spare (0) (Per Server) System Drive (SS) - 2 Disks (72GB 6G SAS 15K2.5m SEC ENT HDD) RAID10 Spare (0)	
SQL Server	1 x ProLiant BL480c Gen8 Server 2P Intel® Xeon® E5-2630 v2 2.93GHz / 15MB Cache 24,576 MB RAM (Per Server)	

# Índice



1. Introducción
2. Elementos de la granja web
3. Necesidades del servidor web
- [ 4. Hardware para servidores ]**
5. Software para servidores
6. Conclusiones