

Introducción al “Capacity planning” para servicios

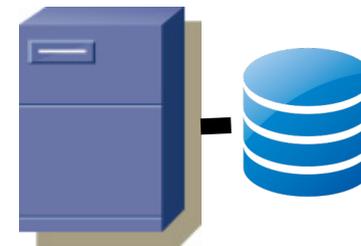
Area de Ingeniería Telemática
<http://www.tlm.unavarra.es>

Grado en Ingeniería en Tecnologías de
Telecomunicación, 4º

Tiempos de acceso a disco

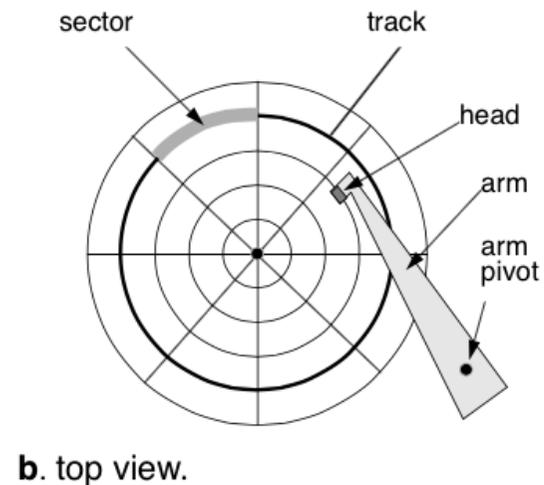
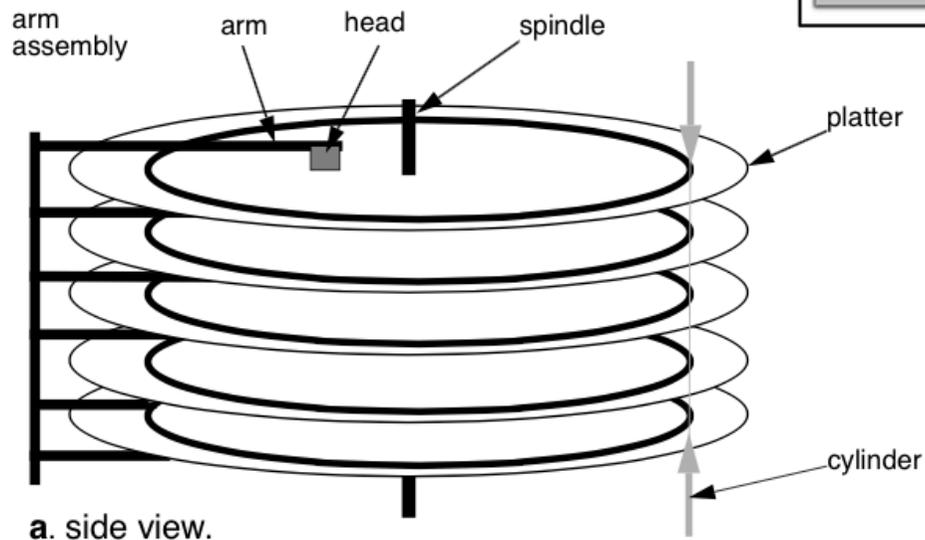
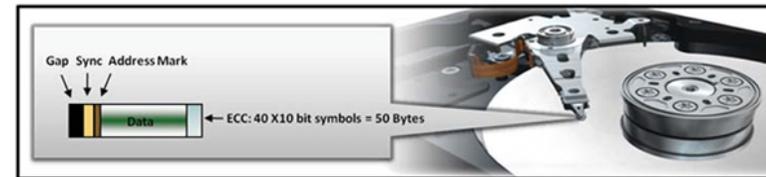
DAS

- DAS = Direct-Attached Storage
- Internos o externos
- Diferentes tipos de buses
 - SCSI: Small Computer System Interface (discos, impresoras, escáners, etc)
 - IDE: Integrated Drive Electronics (ATA = Advanced Technology Attachment)
 - SATA: Serial ATA (eSATA = external SATA)
 - SAS: Serial Attached SCSI
 - USB: Universal Serial Bus
 - IEEE 1394 (FireWire)
 - Fiber Channel



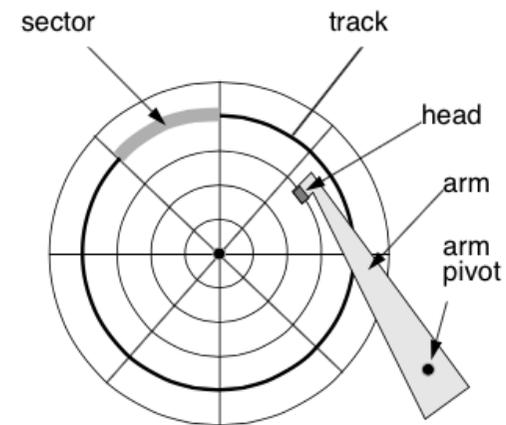
Arquitectura del disco

- Platos (*platters*): material magnético
- Pistas (*tracks*): trayectoria circular
- Sectores: unidad mínima direccionable, todos del mismo tamaño, tradicionalmente 512 Bytes (hay ahora unidades con 4 KiB)
- Cilindros: pila vertical de pistas
- El tamaño del disco (en pulgadas) condiciona su capacidad y consumo
- Brazo y cabeza de lectura/escritura (por cada plato)
- Eje de rotación (*spindle*)



Rotación del disco

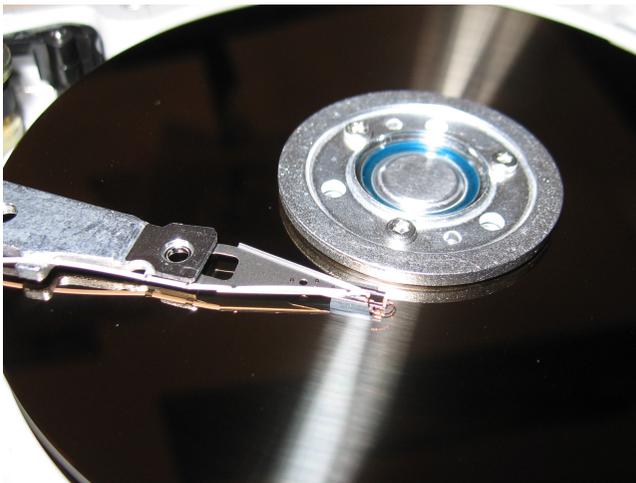
- Todos los platos rotan al unísono
- Típicas velocidades de rotación: 7.200 rpm, 10.000 rpm, 15.000 rpm
- La cabeza debe avanzar hasta la pista donde están los datos
- Debe esperar a que el plato rote hasta que el sector que busca se encuentre debajo
- Entonces podrá leer o escribir (no a la vez)
- Lee del cilindro, así que cuando termina la pista pasa a la del mismo cilindro en otro disco
- Esas 4 operaciones llevan tiempo (posicionarse, esperar a que gire, leer/escribir y opcionalmente cambiar de plato o de pista)



b. top view.

Tiempos básicos

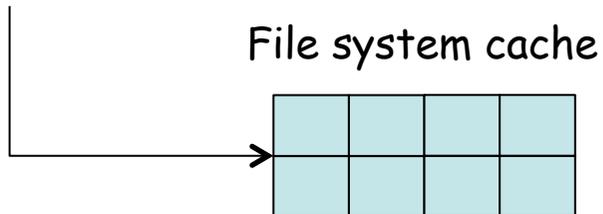
- “*Seek time*”: Tiempo necesario para colocar la cabeza lectora en la pista deseada
- “*Rotational latency*”: Tiempo de espera a que el sector deseado alcance la cabeza
- “*Transfer time*”: Tiempo para transferir los datos del/al sector
- “*Bus transfer time*”: el protocolo del bus (SCSI, SATA) añade mensajes (handshakes)



File System Cache

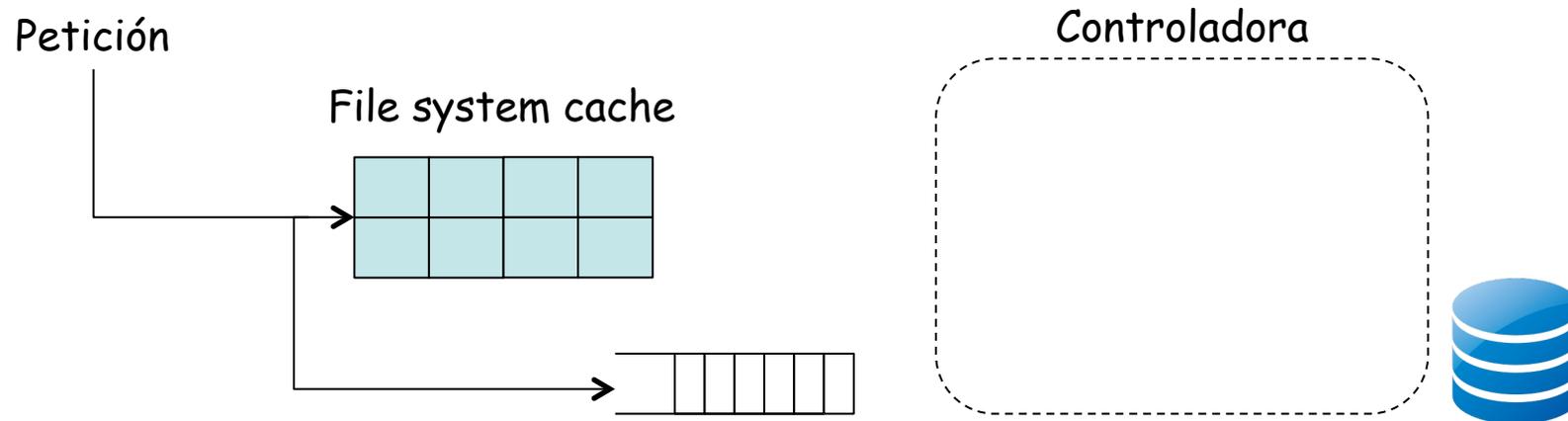
- Para evitar los accesos al disco se implementan caches
- El sistema de ficheros suele tener una (en el kernel)
- Puede encontrar ahí el bloque (*cache hit*) o no (*miss*)

Petición



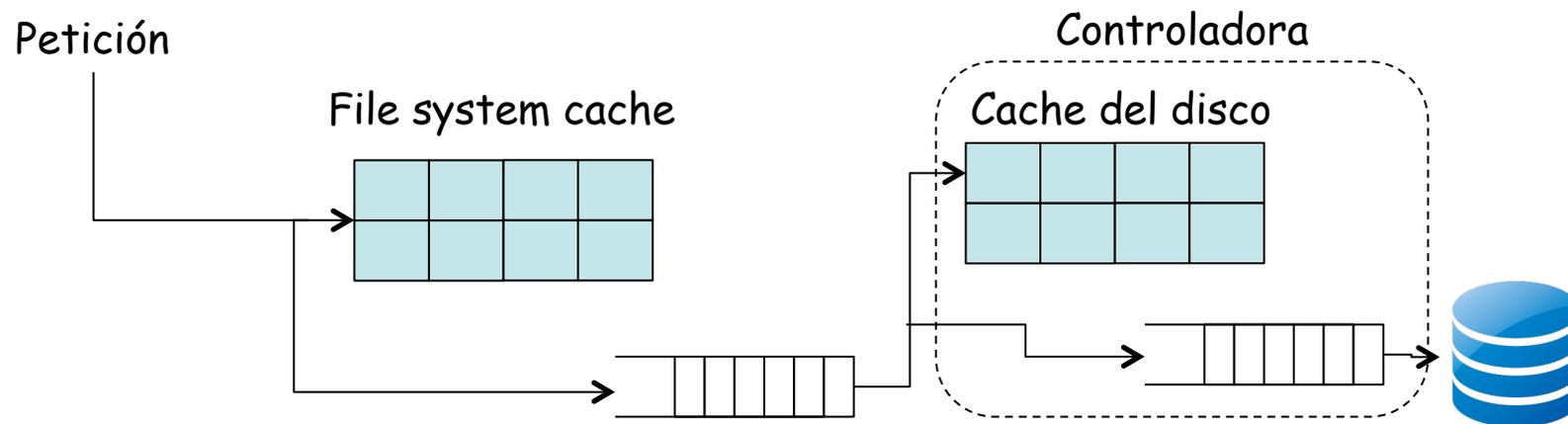
Cola del driver

- La petición al disco se encola (peticiones de más procesos)
- Tras la cola va a la controladora del disco (...)



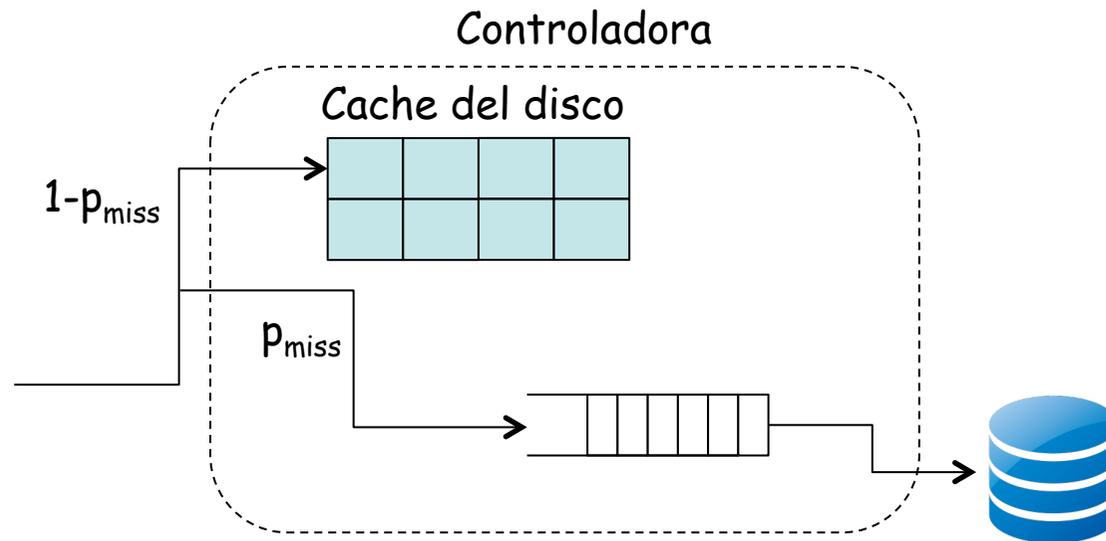
Cache de la controladora

- La controladora implementa otra cache
- Si falla irá a una cola de peticiones para el disco
- Puede reordenar las peticiones de esa cola para reducir los tiempos de acceso al disco (*scheduling*)
- Puede leer más de lo pedido (*read-ahead, pre-fetch*) y guardarlo en la cache para futuras peticiones



Tiempo de servicio del disco

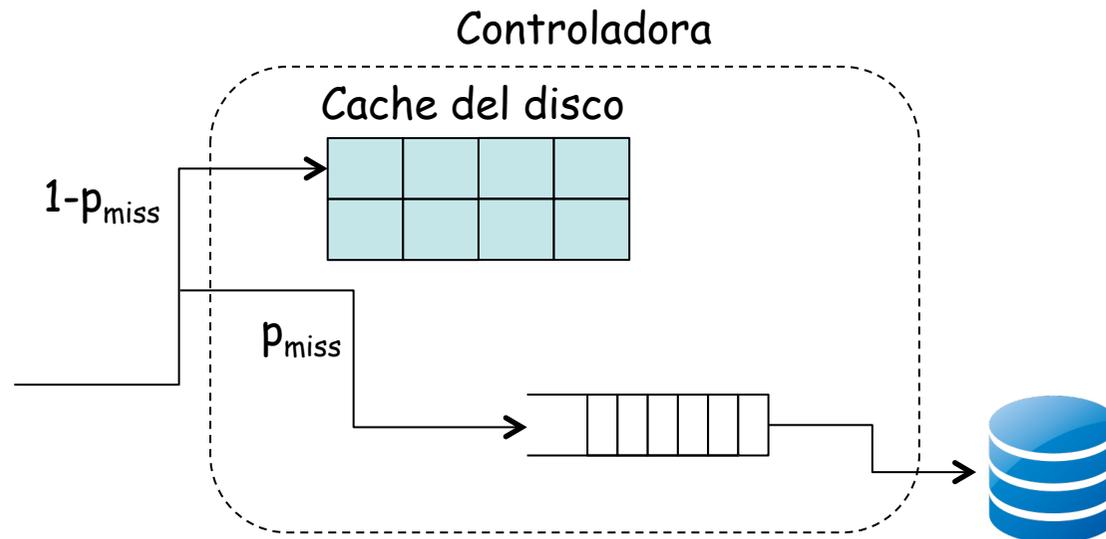
- Veamos solo la parte de la controladora y su tiempo medio



Tiempo de servicio del disco

- Según lo que pida la aplicación habrá una probabilidad de NO encontrarlo en la cache p_{miss}
- Esta comprobación llevará un tiempo
- Si se lee del disco se llevan esos bloques a la cache
- La petición lo que hace es leer de la cache así que siempre habrá un tiempo de leer de ella
- *ControllerTime* cuenta todo eso

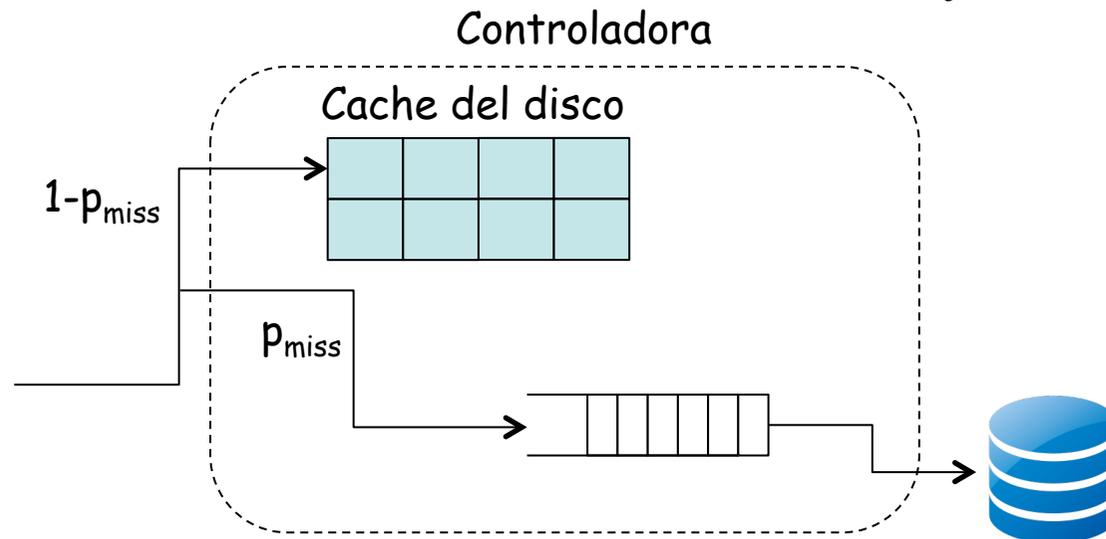
$$\overline{S_d} = ControllerTime + p_{miss} DiskAccessTime$$



Tiempo de servicio del disco

- Estimaremos el tiempo medio de acceso a disco
- Estará compuesto por:
 - El *seek time* medio entre todas las posiciones posibles de la cabeza
 - La *rotational latency* media
 - El tiempo de transferencia: en función del tamaño de bloque y la velocidad de transferencia

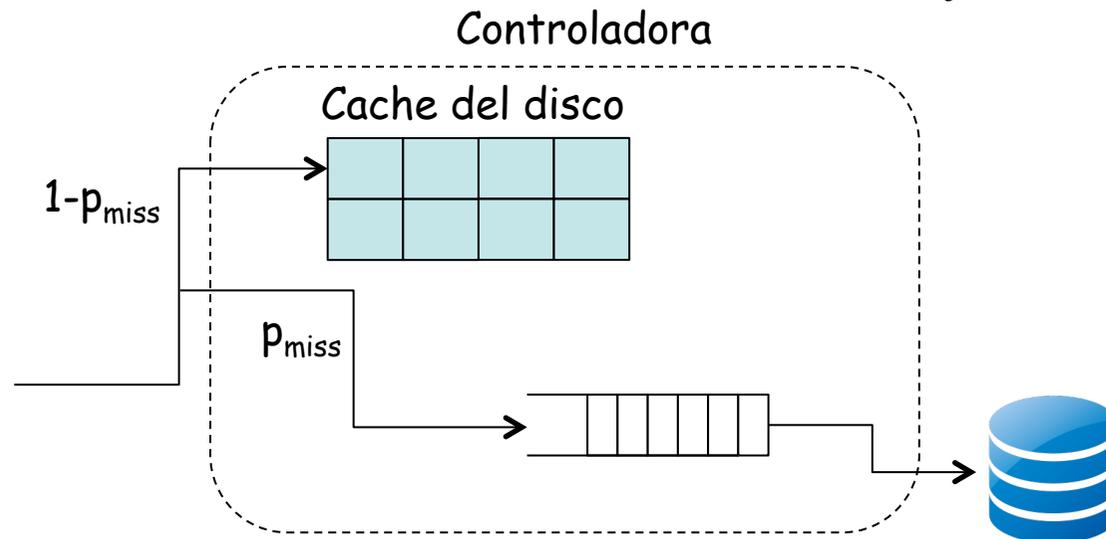
$$\overline{S_d} = \text{ControllerTime} + p_{\text{miss}} \left(\text{SeekTime} + \text{RotationalLatency} + \frac{\text{BlockSize}}{\text{TransferRate}} \right)$$



Tiempo de servicio del disco

- Muchos de estos valores dependen de la carga, es decir, la secuencia de peticiones que se va haciendo:
 - Los aciertos/fallos en la cache (p_{miss}) dependerán de qué bloques se hayan pedido con anterioridad, del tamaño de la misma y la política de reemplazo
 - SeekTime dependerá de la pista en que se encuentren los datos
 - RotationalLatency dependerá de la velocidad de rotación y del sector en que se encuentren los datos (y podría depender de la anterior lectura)

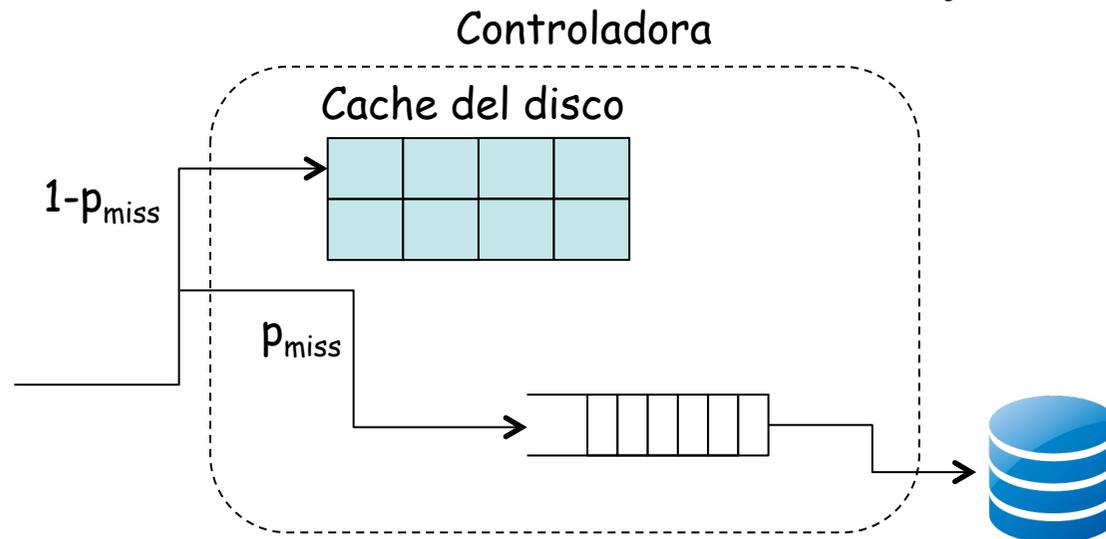
$$\overline{S}_d = ControllerTime + p_{miss} (SeekTime + RotationalLatency + \frac{BlockSize}{TransferRate})$$



Tiempo de servicio del disco

- Se suelen hacer estimaciones de este valor promedio de tiempo para diferentes cargas
- Al hablar de cargas *aleatorias* suelen ser cargas de peticiones de bloques distribuidos de forma uniforme por el disco

$$\overline{S_d} = ControllerTime + p_{miss} \left(SeekTime + RotationalLatency + \frac{BlockSize}{TransferRate} \right)$$

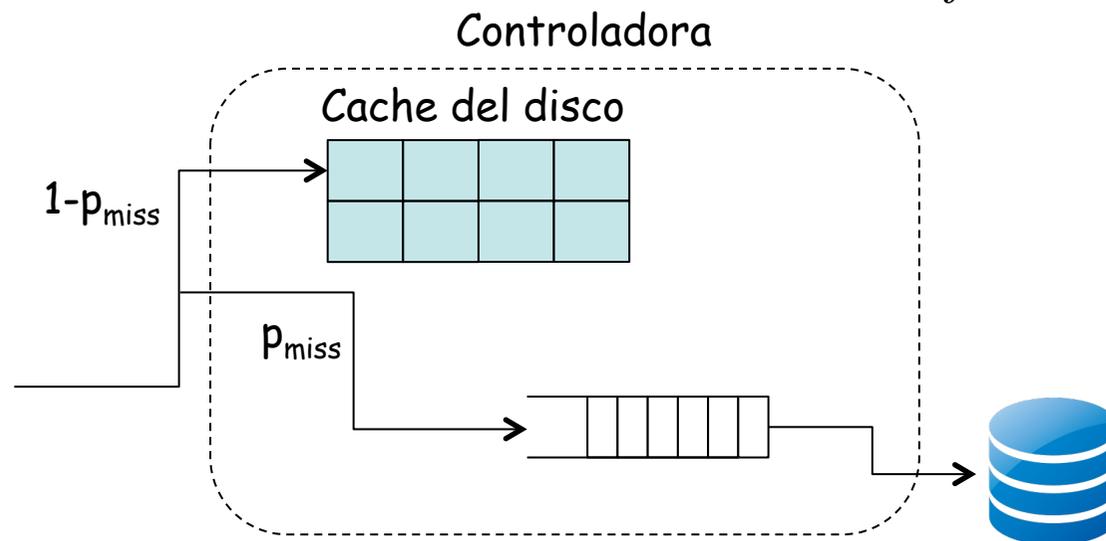


Carga aleatoria

- Podemos suponer que $p_{miss} = 1$
- RotationalLatency, en media es media rotación, lo cual se obtiene de la velocidad de rotación del disco
- El SeekTime medio lo suele dar el fabricante
- Con eso podemos calcular el tiempo medio total
- Y el inverso será el número medio de operaciones por segundo (IOPS = *Input/output operations per second*)

(rpm)	Avg. rot (ms)
5400	5.5
7200	4.2
10000	3
15000	2

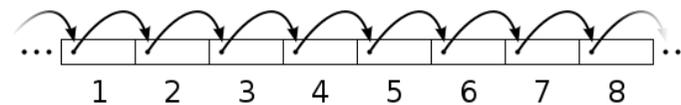
$$\overline{S_d} = ControllerTime + p_{miss} (SeekTime + RotationalLatency + \frac{BlockSize}{TransferRate})$$



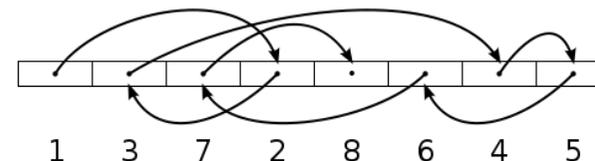
Carga secuencial

- Normalmente una lectura no buscará solo un bloque sino varios
- Se habla de carga *secuencial* cuando los bloques se encuentran consecutivos en el disco o a grupos (*runs*)
- Ejemplo: bloques 4, 350, 351, 352, 353, 80, 104, 105, 106, 107, 108, 243
- En estos casos, en cada *run* se puede sacar provecho a un *read-ahead* que daría un hit en la cache para todos los bloques menos el primero en el *run*

Sequential access



Random access



Tipos de discos

Discos para escritorio



Especificaciones	4 TB ¹	3 TB ¹	2 TB ¹	1 TB ¹	500 GB ¹
Números de modelo estándar	ST4000DM000	ST3000DM001	ST2000DM001	ST1000DM003	ST500DM002 ²
Números de modelo de Seagate Secure™	—	ST3000DM002	ST2000DM002	ST1000DM004	—
Nombre de modelo	Desktop HDD	Anteriormente Barracuda®	Anteriormente Barracuda	Anteriormente Barracuda	Anteriormente Barracuda
Opciones de la interfaz	SATA a 6 Gb/s con NCQ	SATA a 6 Gb/s con NCQ	SATA a 6 Gb/s con NCQ	SATA a 6 Gb/s con NCQ	SATA a 6 Gb/s con NCQ
Rendimiento					
Caché, multisegmentada (MB)	64	64	64	64	16
Velocidades de transferencia SATA compatibles (Gb/s)	6,0 / 3,0 / 1,5	6,0 / 3,0 / 1,5	6,0 / 3,0 / 1,5	6,0 / 3,0 / 1,5	6,0 / 3,0 / 1,5
Tiempo medio de búsqueda, lectura (ms)	12	<8,5	<8,5	<8,5	<8,5
Tiempo medio de búsqueda, escritura (ms)	12	<9,5	<9,5	<9,5	<9,5
Velocidad de transferencia de datos media, lectura / escritura (MB/s)	146	156	156	156	125
Velocidad de transferencia de datos sostenida máxima, lectura OD (MB/s)	180	210	210	210	144
Configuración / organización					
Cabezales / discos	8 / 4	6 / 3	4 / 2	2 / 1	2 / 1
Bytes por sector	4.096	4.096	4.096	4.096	4.096 o 512 ²
Voltaje					
Tolerancia del voltaje, incluido el ruido (5 V)	±5%	±5%	±5%	±5%	±5%
Tolerancia del voltaje, incluido el ruido (12 V)	±10%	+10% / -7,5%	+10% / -7,5%	+10% / -7,5%	+10% / -7,5%
Fiabilidad / Integridad de los datos					
Ciclos de inicio / parada de contacto	—	—	—	—	50.000
Ciclos de carga / descarga	300.000	300.000	300.000	300.000	—
Errores de lectura no recuperables por bits leídos, máx.	1 por 10 ¹⁴	1 por 10 ¹⁴	1 por 10 ¹⁴	1 por 10 ¹⁴	1 por 10 ¹⁴
Índice de carga de trabajo (TB/año)	55	55	55	55	55
Horas de funcionamiento	2.400	2.400	2.400	2.400	2.400
Garantía limitada (años) ³	2	2	2	2	2

Discos para escritorio



Especificaciones	4 TB ¹	3 TB ¹	2 TB ¹	1 TB ¹	500 GB ¹
Números de modelo estándar	ST4000DM000	ST3000DM001	ST2000DM001	ST1000DM003	ST500DM002 ²
Números de modelo de Seagate Secure™	—	ST3000DM002	ST2000DM002	ST1000DM004	—
Nombre de modelo	Desktop HDD	Anteriormente Barracuda®	Anteriormente Barracuda	Anteriormente Barracuda	Anteriormente Barracuda
Opciones de la interfaz	SATA a 6 Gb/s con NCQ	SATA a 6 Gb/s con NCQ	SATA a 6 Gb/s con NCQ	SATA a 6 Gb/s con NCQ	SATA a 6 Gb/s con NCQ
Rendimiento					
Caché, multisegmentada (MB)	64	64	64	64	16
Velocidades de transferencia SATA compatibles (Gb/s)	6,0 / 3,0 / 1,5	6,0 / 3,0 / 1,5	6,0 / 3,0 / 1,5	6,0 / 3,0 / 1,5	6,0 / 3,0 / 1,5
Tiempo medio de búsqueda, lectura (ms)	12	<8,5	<8,5	<8,5	<8,5
Tiempo medio de búsqueda, escritura (ms)	12	<9,5	<9,5	<9,5	<9,5
Velocidad de transferencia de datos media (lectura / escritura) (MB/s)	146	156	156	156	125
		210	210	210	144
Configuración / organización					
Cabezales / discos	8 / 4	6 / 3	4 / 2	2 / 1	2 / 1
Bytes por sector	4.096	4.096	4.096	4.096	4.096 o 512 ²
Voltaje					
Tolerancia del voltaje, incluido el ruido (5 V)	±5%	±5%	±5%	±5%	±5%
Tolerancia del voltaje, incluido el ruido (12 V)	±10%	+10% / -7,5%	+10% / -7,5%	+10% / -7,5%	+10% / -7,5%
Fiabilidad / Integridad de los datos					
Ciclos de inicio / parada de contacto	—	—	—	—	50.000
Ciclos de carga / descarga	300.000	300.000	300.000	300.000	—
Errores de lectura no recuperables por bits leídos, máx.	1 por 10 ¹⁴	1 por 10 ¹⁴	1 por 10 ¹⁴	1 por 10 ¹⁴	1 por 10 ¹⁴
Índice de carga de trabajo (TB/año)	55	55	55	55	55
Horas de funcionamiento	2.400	2.400	2.400	2.400	2.400
Garantía limitada (años) ³	2	2	2	2	2

NCQ = Native Command Queuing

Discos para escritorio



Especificaciones	4 TB ¹	3 TB ¹	2 TB ¹	1 TB ¹	500 GB ¹
Números de modelo estándar	ST4000DM000	ST3000DM001	ST2000DM001	ST1000DM003	ST500DM002 ²
Números de modelo de Seagate Secure™	—	ST3000DM002	ST2000DM002	ST1000DM004	—
Nombre de modelo	Desktop HDD	Anteriormente Barracuda®	Anteriormente Barracuda	Anteriormente Barracuda	Anteriormente Barracuda
Opciones de la interfaz	SATA a 6 Gb/s con NCQ	SATA a 6 Gb/s con NCQ	SATA a 6 Gb/s con NCQ	SATA a 6 Gb/s con NCQ	SATA a 6 Gb/s con NCQ
Rendimiento					
Caché, multisegmentada (MB)	64	64	64	64	16
Velocidades de transferencia SATA compatibles (Gb/s)	6,0 / 3,0 / 1,5	6,0 / 3,0 / 1,5	6,0 / 3,0 / 1,5	6,0 / 3,0 / 1,5	6,0 / 3,0 / 1,5
Tiempo medio de búsqueda, lectura (ms)	12	<8,5	<8,5	<8,5	<8,5
Tiempo medio de búsqueda, escritura (ms)	12	<9,5	<9,5	<9,5	<9,5
Velocidad de transferencia de datos media, lectura / escritura (MB/s)	146	156	156	156	125
Velocidad de transferencia de datos sostenida máxima, lectura OD (MB/s)	180	210	210	210	144
Configuración / organización					
Cabezales / discos	8 / 4	6 / 3	4 / 2	2 / 1	2 / 1
Bytes por sector	4.096	4.096	4.096	4.096	4.096 o 512 ²
Voltaje					
Tolerancia del voltaje, incluido el ruido (5 V)	±5%	±5%	±5%	±5%	±5%
Tolerancia del voltaje, incluido el ruido (12 V)	±10%	+10% / -7,5%	+10% / -7,5%	+10% / -7,5%	+10% / -7,5%
Fiabilidad / Integridad de los datos					
Ciclos de inicio / parada de contacto	—	—	—	—	50.000
Ciclos de carga / descarga	300.000	300.000	300.000	300.000	—
Errores de lectura no recuperables por bits leídos, máx.	1 por 10 ¹⁴	1 por 10 ¹⁴	1 por 10 ¹⁴	1 por 10 ¹⁴	1 por 10 ¹⁴
Índice de carga de trabajo (TB/año)	55	55	55	55	55
Horas de funcionamiento	2.400	2.400	2.400	2.400	2.400
Garantía limitada (años) ³	2	2	2	2	2

Discos para escritorio



2.6 Seek time

Seek measurements are taken with nominal power at 25°C ambient temperature. All times are measured using drive diagnostics. The specifications in the table below are defined as follows:

- Track-to-track seek time is an average of all possible single-track seeks in both directions.
- Average seek time is a true statistical random average of at least 5000 measurements of seeks between random tracks, less overhead.

Typical seek times (ms)	Read	Write
Track-to-track	1.0	1.2
Average	12.0	12.0
Average latency	5.1	

2.7 Start/stop times

The start/stop times listed below are for the 3TB and 4TB models.

Power-on to ready (in seconds)	15 (typical) 18 (max)
Standby to ready (in seconds)	15 (typical) 18 (max)
Ready to spindle stop (in seconds)	10 (typical) 11 (max)

Time-to-ready may be longer than normal if the drive power is removed without going through normal OS powerdown procedures.

Disco “enterprise”



- 15K (15.000 rpm)

Especificaciones	512 nativo ⁴		Emulación 512		4K nativo	
	600 GB ¹	300 GB ¹	600 GB ¹	300 GB ¹	600 GB ¹	
Modelo estándar	ST600MP0005	ST300MP0005	ST600MP0035	ST300MP0035	ST600MP0065	S
Modelo Seagate Secure™	ST600MP0015 ²	ST300MP0015 ²	ST600MP0045 ²	ST300MP0045 ²	ST600MP0075 ²	S
Modelo Seagate Secure FIPS 140-2	ST600MP0025 ^{2,3}	—	ST600MP0055 ^{2,3}	—	ST600MP0085 ^{2,3}	
Interfaz	SAS a 12 Gb/s	SAS a 12 Gb/s	SAS a 12 Gb/s	SAS a 12 Gb/s	SAS a 12 Gb/s	S
Rendimiento						
Tiempo medio de latencia (ms)	2,0	2,0	2,0	2,0	2,0	
Velocidad de transferencia sostenida (Diámetro externo a interno) MiB/s	De 233 a 160	De 233 a 160	De 246 a 180	De 246 a 180	De 246 a 180	
Máx. velocidad de transferencia instantánea (puerto dual SAS) MB/s	2.400	2.400	2.400	2.400	2.400	
Caché, multisegmentada (MB)	128	128	128	128	128	
Configuración / fiabilidad						
Discos / Cabezales	3 / 6	2 / 4	3 / 6	2 / 4	3 / 6	
Errores de lectura no recuperables por bits leídos, máx.	1 por 10 ¹⁶	1 por 10 ¹⁶	1 por 10 ¹⁶	1 por 10 ¹⁶	1 por 10 ¹⁶	
Porcentaje de errores anual (AFR)	0,44%	0,44%	0,44%	0,44%	0,44%	
Garantía limitada basada en uso (años) ⁵	5	5	5	5	5	
Administración de energía						

Disco “enterprise”

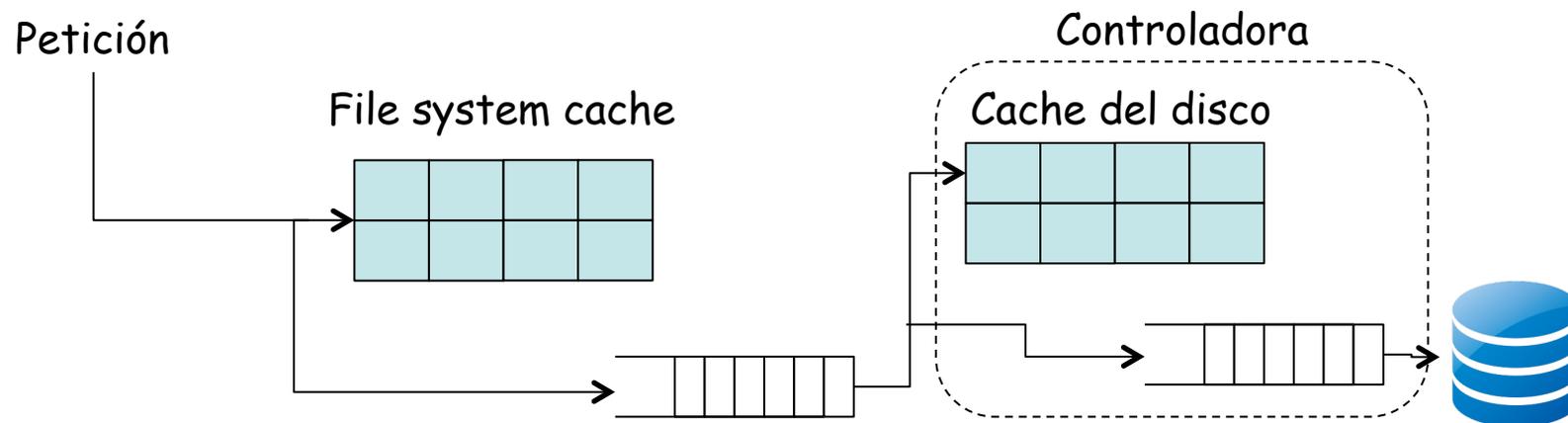


Enterprise Performance 15K HDD and Enterprise Turbo SSHD drives have the following standard features:

- Perpendicular recording technology
- 3.0 / 6.0 / 12.0 Gb Serial Attached SCSI (SAS) interface
- Integrated dual port SAS controller supporting the SCSI protocol
- Support for SAS expanders and fanout adapters
- Firmware downloadable using the SAS interface
- 128-deep task set (queue)
- Supports up to 32 initiators
- Jumperless configuration
- User-selectable logical block size for 512N/512e (512, 520, 524, or 528 bytes per logical block)
- User-selectable logical block size for 4KN (4096, 4160, 4192, or 4224 bytes per logical block)
- Industry standard SFF 2.5-in dimensions
- Programmable logical block reallocation scheme
- Flawed logical block reallocation at format time
- Programmable auto write and read reallocation
- Reallocation of defects on command (Post Format)
- ECC maximum burst correction length of 520 bits for 512 byte blocks and 1400 bits for 4k byte blocks
- No preventive maintenance or adjustments required
- Dedicated head landing zone and automatic shipping lock
- Embedded servo design
- Automatic shipping lock
- Self diagnostics performed when power is applied to the drive
- Zone bit recording (ZBR)
- Vertical, horizontal, or top down mounting
- Dynamic spindle brake
- 128MB data buffer (see Section 4.5)
- 32GB hybrid read cache
- 8MB Hybrid NVC-backed write cache
 Approximately 512K on standard models (512e, 4KN)
- Drive Self Test (DST)

Modelo de rendimiento de disco

- Hay modelos con decenas de parámetros
- Con un buen modelo se puede predecir el comportamiento de un disco por simulación
- Hay aplicaciones que hacen pruebas de los discos con diferentes patrones de carga



Discos SSD

- Solid State Disk
- En realidad no es un “disco” sino memoria, pero la utilizamos como almacenamiento permanente
- NAND Flash
- No tiene sentido hablar de tiempo de posicionar la cabeza lectora o de esperar a que rote el disco



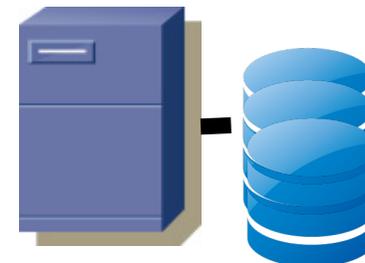
Disco SSD

- La velocidad máxima (1.200 MBps) es para una lectura/escritura secuencial de 128KB
- La tasa sostenida (750/500 MBps) es para lecturas/escrituras de bloques de 128KB
- Las celdas de almacenamiento se desgastan; el disco intenta equilibrar el uso de todas las celdas
- Estas tasas caen con el uso (5 años pueden ser del 60%)

Especificaciones	800 GB ¹	400 GB ¹	200 GB ¹
Modelos SED	ST800FM0053 ²	ST400FM0073 ²	ST200FM0073 ²
Modelos FIPS 140-2	ST800FM0063 ^{2,3}		
Modelos que no son SED	ST800FM0043	ST400FM0053	ST200FM0053
Opciones de la interfaz	SAS a 12 Gb/s	SAS a 12 Gb/s	SAS a 12 Gb/s
Tipo de flash NAND	MLC	MLC	MLC
Rendimiento			
Velocidad máx. de transferencia de datos sostenida (MB/s)	750	750	750
Velocidad de transferencia de datos de E/S (MB/s)	1.200	1.200	1.200
Entradas / salidas por segundo por vatio (IOPS/W)	27.160	29.650	27.990
Velocidad de comandos de lectura / escritura secuencial (MB/s) máxima, 128 KB	750 / 500	750 / 500	750 / 400
Velocidad de comandos de lectura / escritura aleatoria (IOPS) máxima, 4 KB	110.000 / 40.000	110.000 / 40.000	110.000 / 25.000
Configuración / fiabilidad			
Errores de lectura no recuperables por bits leídos, máx.	1 por 10 ¹⁶	1 por 10 ¹⁶	1 por 10 ¹⁶
Porcentaje de errores anual (AFR)	0,44%	0,44%	0,44%
Total de terabytes escritos (TBW) durante el período de garantía ⁴	14.600	7.300	3.650
Garantía limitada con uso de soportes (años) ⁵	5	5	5
Administración de energía			

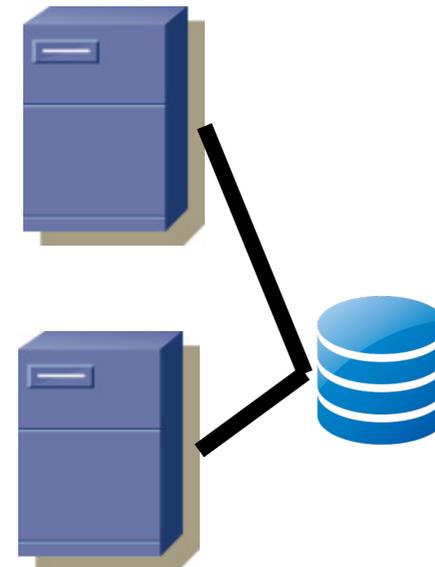
Otros discos

- Esto eran soluciones DAS con un solo disco
- Podemos agrupar varios discos y repartir los datos entre ellos
- Hablaríamos entonces de soluciones RAID
- Pueden ofrecer mayor seguridad ante fallos, mayor rendimiento o una combinación de ambas
- No vamos a entrar en más detalle sobre estas opciones y su rendimiento



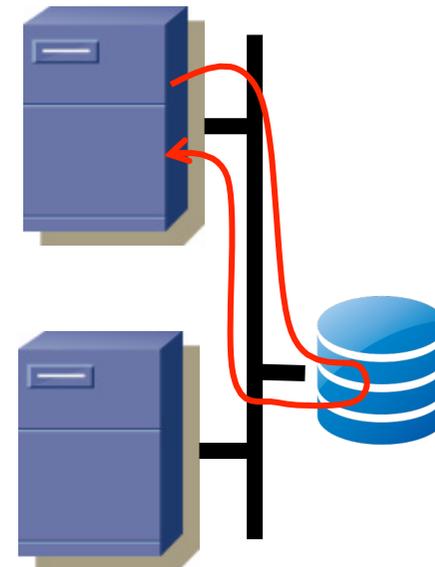
Otros discos

- Podemos compartir el disco entre varios hosts
- Hablamos entonces de soluciones NAS y SAN y tenemos una red para comunicar el servidor con el disco (...)



Otros discos

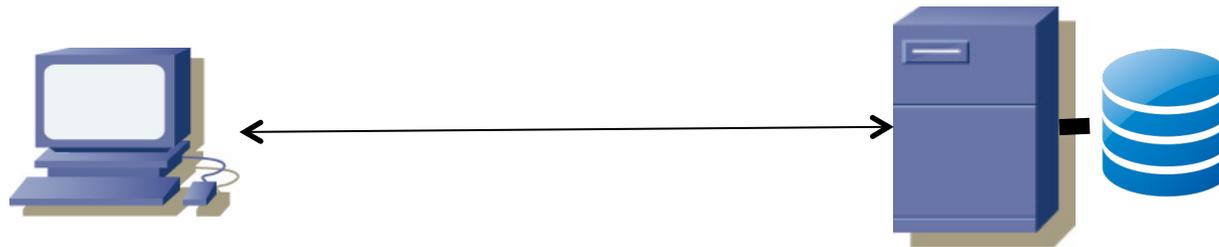
- Podemos compartir el disco entre varios hosts
- Hablamos entonces de soluciones NAS y SAN y tenemos una red para comunicar el servidor con el disco
- Básicamente el disco es en sí un servidor en red
- El análisis presentado sería entonces para el extremo final donde se encuentra el disco
- Y para analizar el tiempo de acceso del servidor al disco debemos tener en cuenta la red y los protocolos empleados en ella



Evolución del modelo de servidor

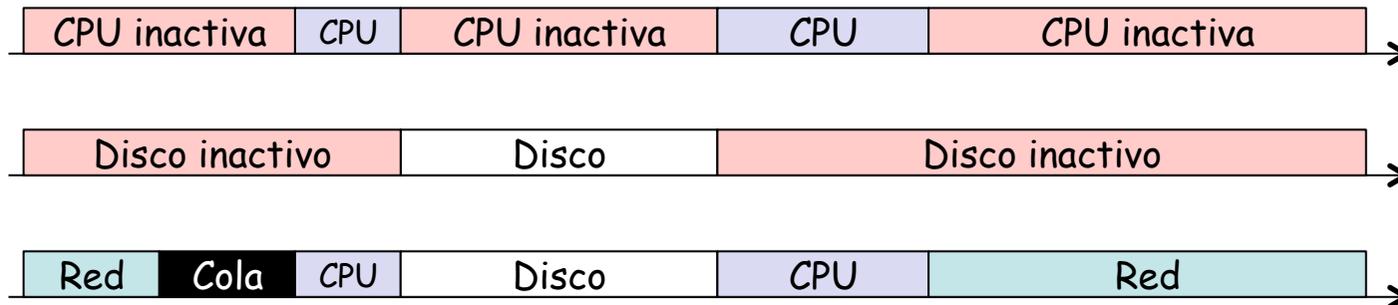
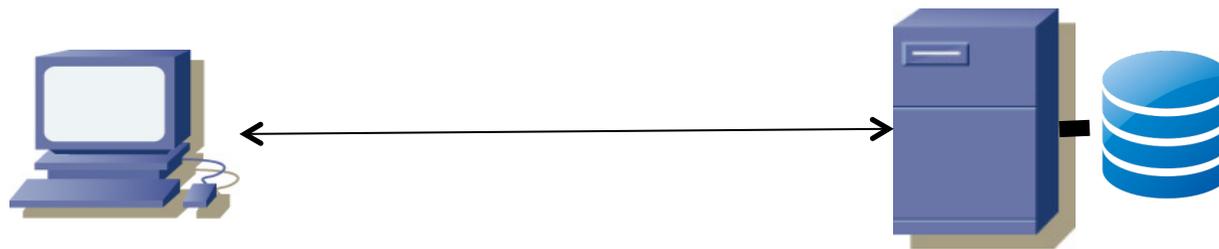
Modelo de un servidor

- Hemos supuesto que se atienden las peticiones en serie
- Esto es poco eficiente porque (...)



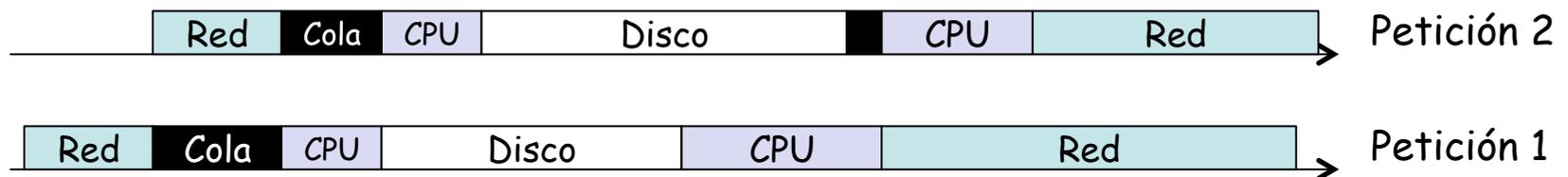
Modelo de un servidor

- Hemos supuesto que se atienden las peticiones en serie
- Esto es poco eficiente porque no estamos usando todo el tiempo los recursos (CPU, disco, red)
- Podríamos tener varios procesos/hilos activos, atender a varias peticiones simultáneamente



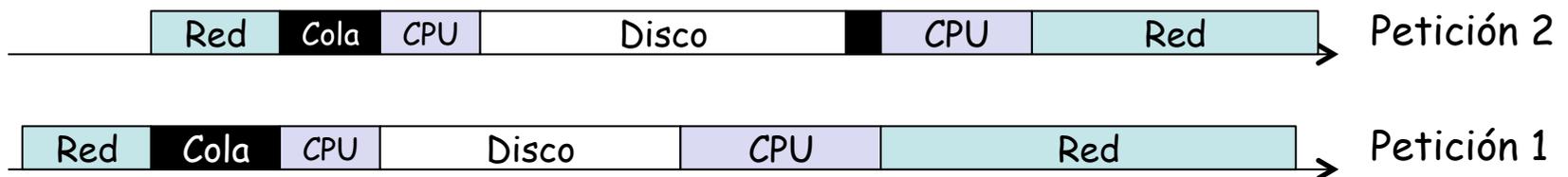
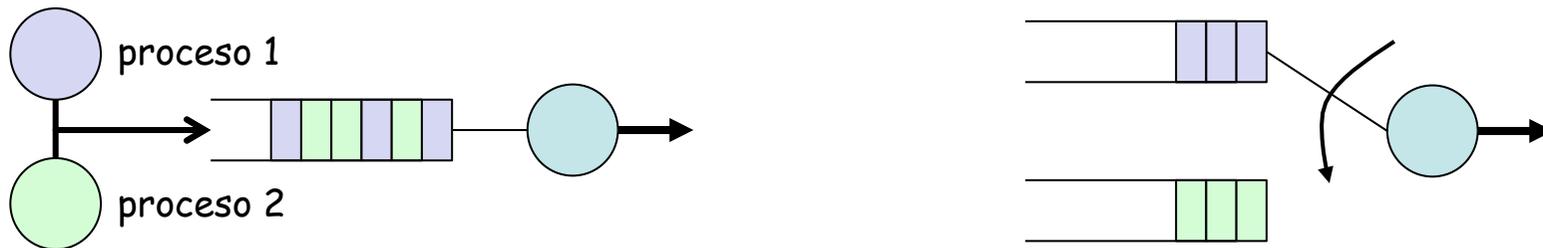
Modelo multi-proceso/thread

- Hemos supuesto que se atienden las peticiones en serie
- Esto es poco eficiente porque no estamos usando todo el tiempo los recursos (CPU, disco, red)
- Podríamos tener varios procesos/hilos activos, atender a varias peticiones simultáneamente
- Si la CPU es solo 1 (1 core) el S.O. tendrá que repartir el uso de la misma (*scheduling*) por lo que puede tener que esperar un proceso
- Los accesos a disco también pueden colisionar
 - Entrarán en funcionamiento las colas en el kernel o la controladora
 - Se reparte el uso del disco así que si los dos quieren leer, cada uno verá que tarda más en obtener los datos pues espera al otro
 - Aunque con cachés podría uno sacar provecho a lecturas del otro



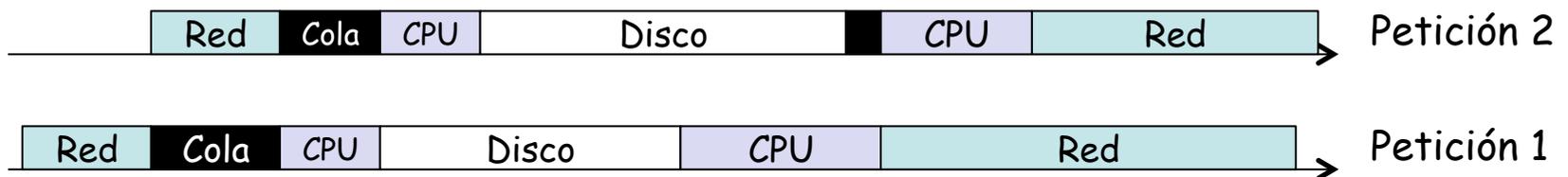
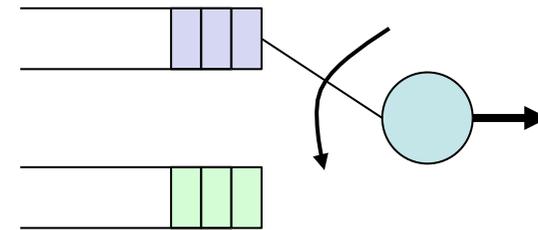
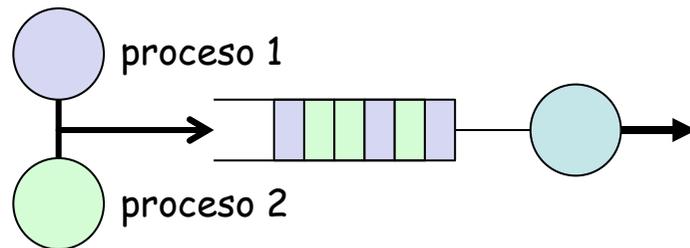
Modelo multi-proceso/thread

- Si tenemos múltiples procesos/hilos pero solo una CPU, en el fondo tenemos:
 - Una cola con las peticiones de uso de la CPU de los procesos
 - Cuántas peticiones depende de cuánto quieran usar la CPU los procesos
 - Un servidor que es la CPU que solo puede ejecutar instrucciones de uno
 - Cuánto tiempo de uso de CPU se asigne a cada proceso cada vez depende del planificador
 - Esto es muy similar a los planificadores en un enlace de comunicaciones y un RR o WRR es lo más frecuente
 - Aunque se puede dar más prioridad a un proceso que a otros y son posibles las soluciones con *pre-emption*



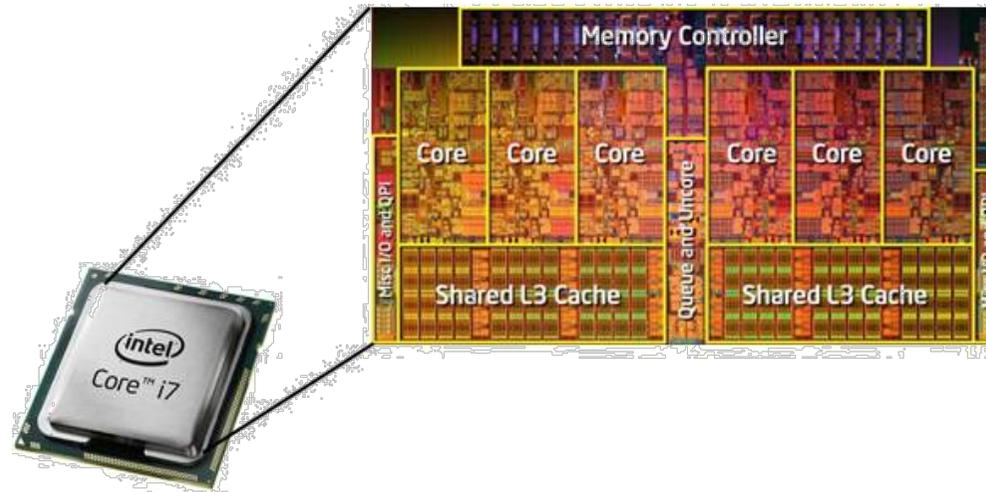
Modelo multi-proceso/thread

- En general permitir varios threads con baja carga es bueno pues reduce el tiempo de espera de la petición a que el thread esté libre
- Es decir, para una carga baja se puede esperar una reducción del tiempo de espera medio al aumentar el número de threads
- Llegado cierto número de threads no mejora pues lo que limitará es la contienda en el acceso a la CPU
- Con alta carga (muchas peticiones) y muchos threads corremos el riesgo de dedicar mucho tiempo al cambio de contexto entre ellos



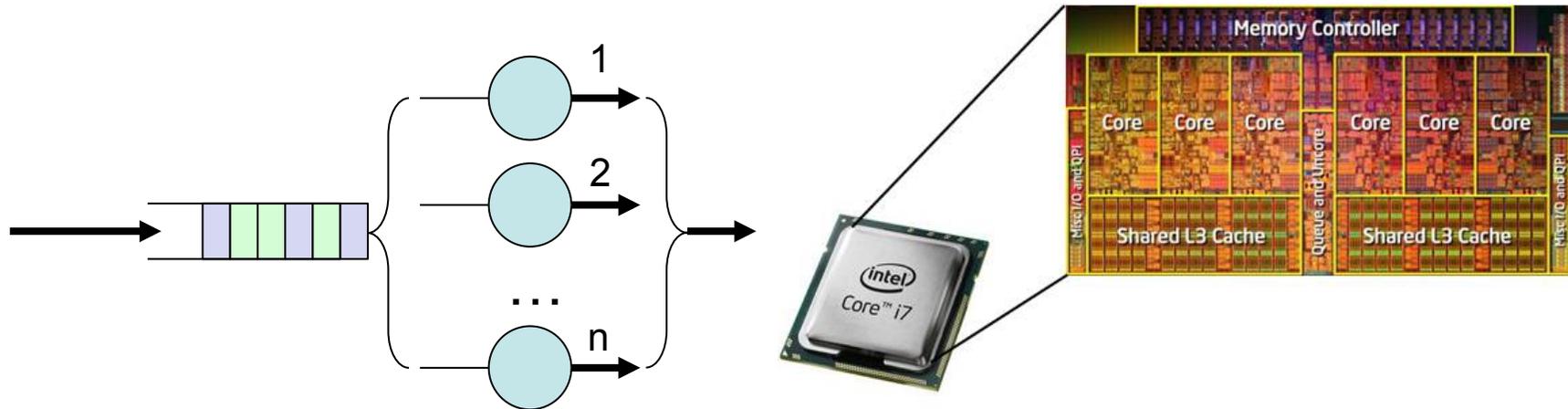
Modelo multi-proceso/thread

- Si tenemos:
 - múltiples CPUs o
 - una CPU con múltiples “cores” o
 - múltiples CPUs con múltiples “cores” o
 - múltiples cores con múltiples “hilos” (CPU superescalar)
 - o múltiples CPUs con múltiples “cores” con múltiples “hilos” de ejecución
- tenemos varias unidades de servicio
- Ejemplo: CPU core i7 con 4 cores y *hyperthreading* sería 1 CPU x 4 cores x 2 hilos por core



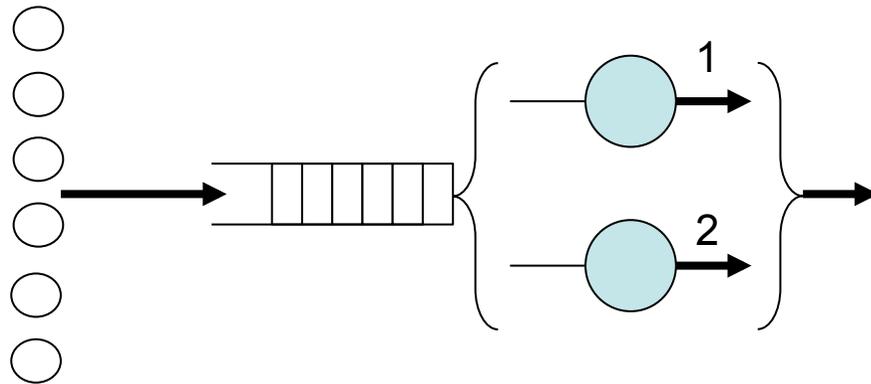
Modelo multi-proceso/thread

- Si hay más cores/hilos que procesos no tendremos contienda
- Si hay menos cores que procesos de nuevo necesitamos repartir el uso
- El mismo proceso no tiene por qué ejecutarse todo el tiempo en el mismo core aunque se puede forzar



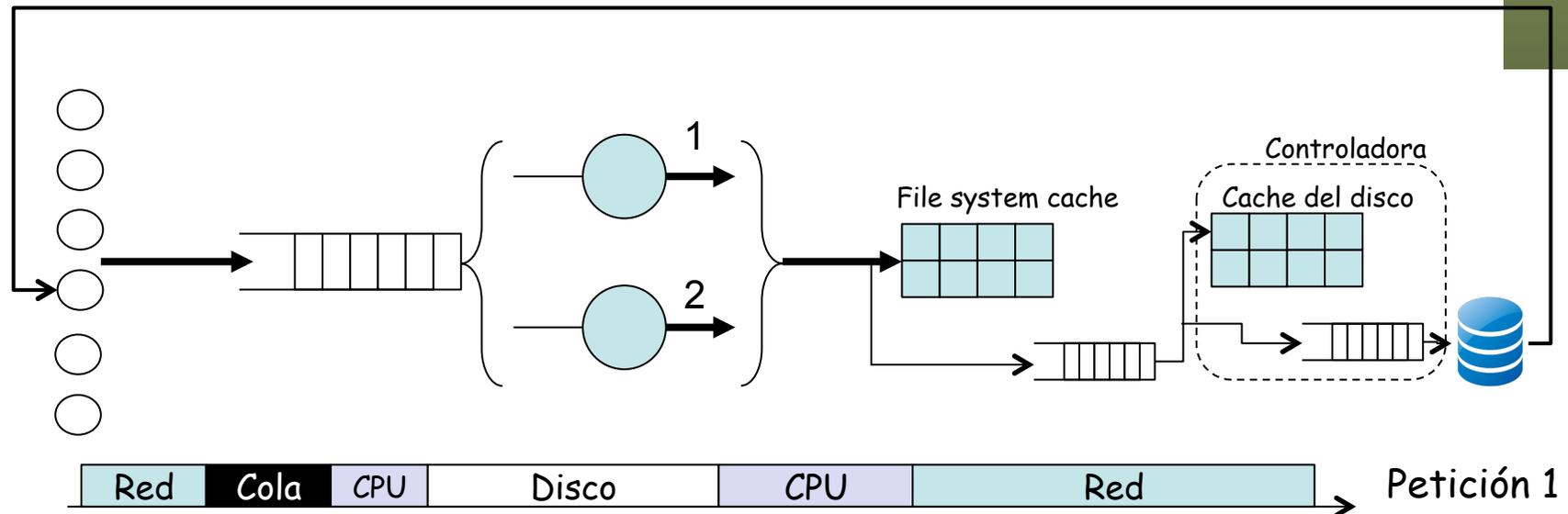
Modelo multi-proceso/thread

- Es normal que haya más procesos que cores porque los procesos no siempre necesitan la CPU
- Aunque entonces puede haber periodos de tiempo en que tengan que esperar a que se libere una CPU/core
- (...)



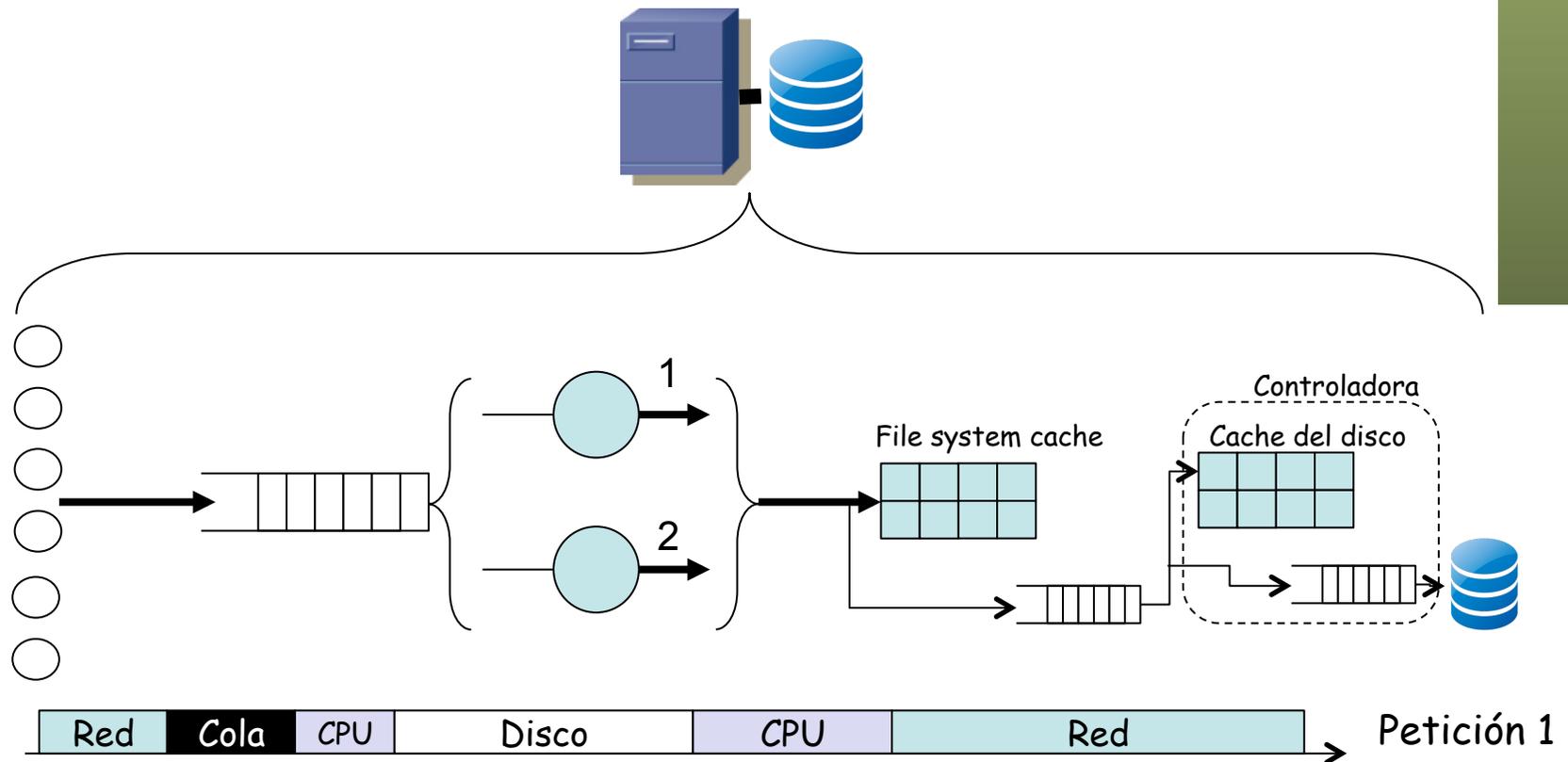
Modelo multi-proceso/thread

- Es normal que haya más procesos que cores porque los procesos no siempre necesitan la CPU
- Aunque entonces puede haber periodos de tiempo en que tengan que esperar a que se libere una CPU/core
- Y no olvidemos que detrás puede estar el disco porque tras obtener tiempo de CPU requiera leer de disco
- Y después puede requerir de nuevo tiempo de CPU



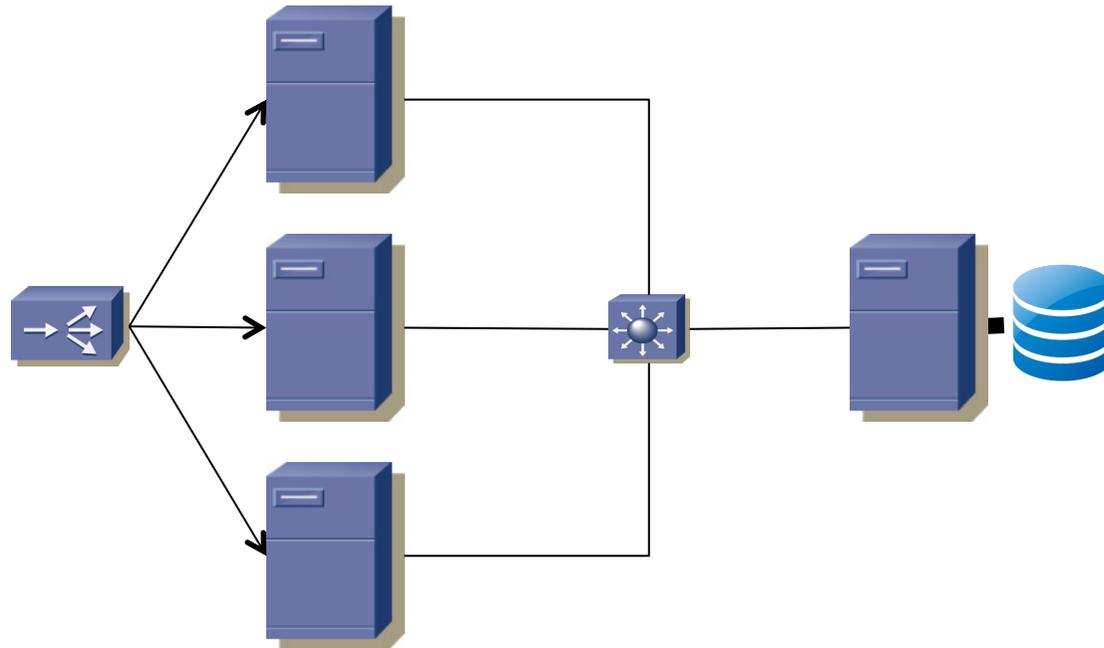
Modelo multi-proceso/thread

- ¿Y la arquitectura *n-Tier*?



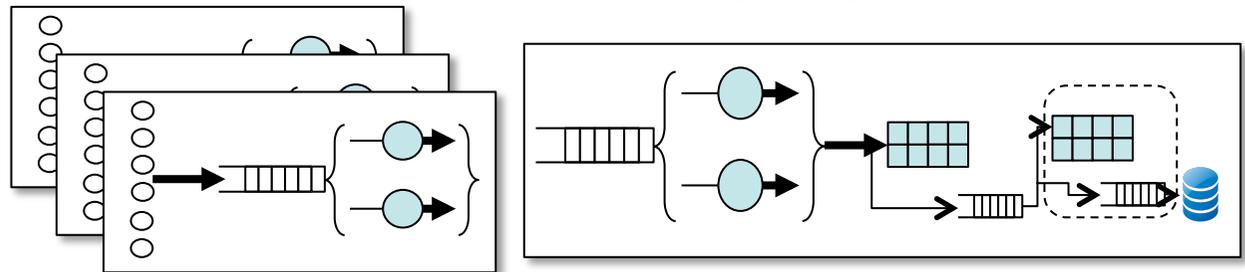
Multi-capa

- Aquí no se está contabilizando el acceso al disco local de los servidores web



Servidores Web

Servidor de aplicaciones
+
Base de datos



Arquitectura y su análisis

- Como podemos ver el sistema “es complejo”
- Demasiados factores que interactúan
- Además nos falta conocer bien bastantes elementos que son muy comunes en la escala media-grande:
 - NAS: almacenamiento en red en modo acceso a fichero, protocolos típicos y sus limitaciones (SMB, NFS, etc)
 - SAN: almacenamiento en red en modo acceso a bloque, protocolos (FiberChannel, iSCSI), características de la red y el servicio, etc
 - Discos en RAID vs JBOD
 - Bases de datos y sus servidores
 - Servidores de aplicaciones, cachés distribuidas
 - Virtualización
 - Balanceo según carga
 - etc etc etc

