

Chapter 12: Mass Storage System



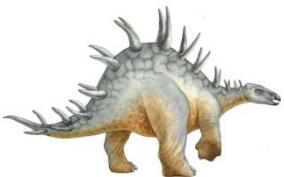
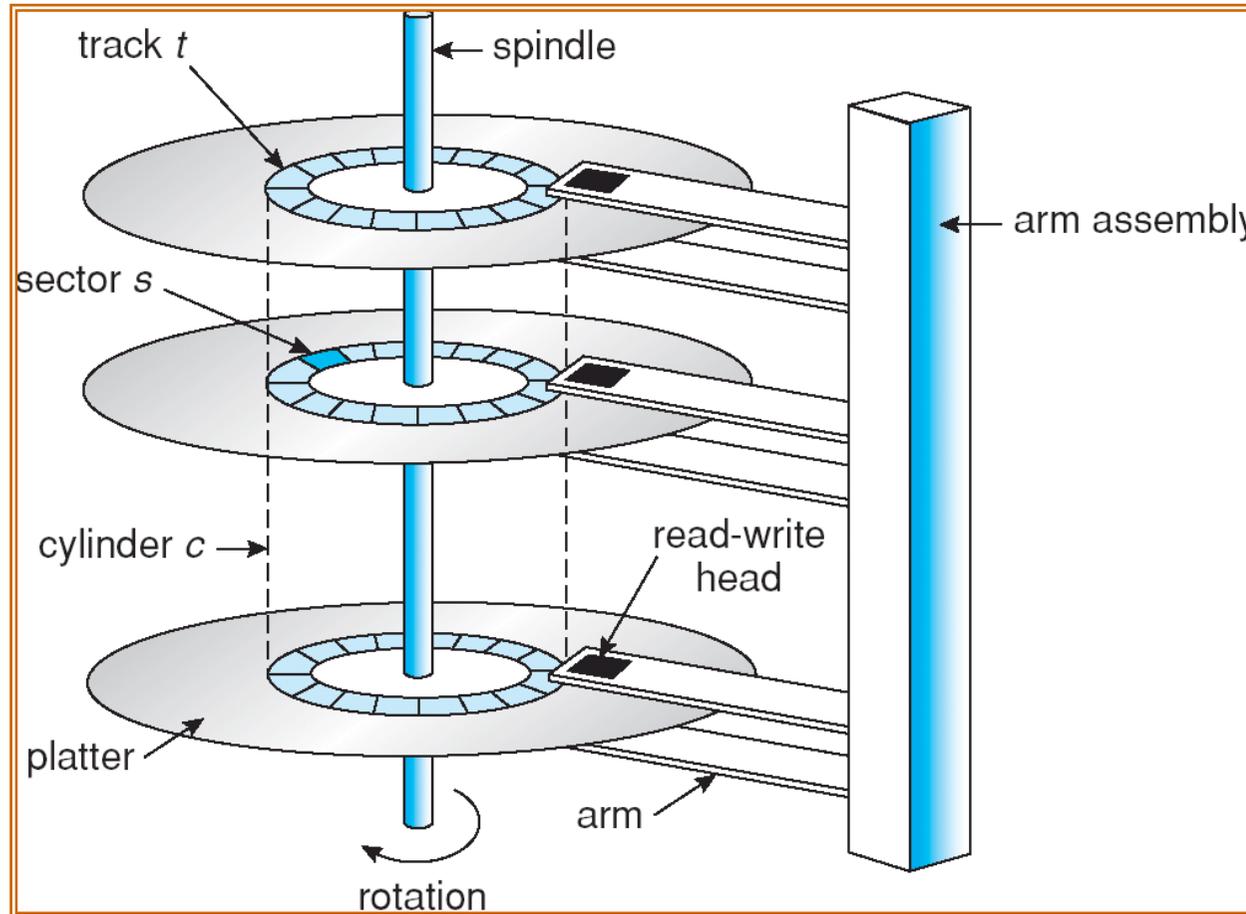
HDD Disk Structure

- Disk drives are addressed as large 1-dimensional arrays of *logical blocks*, where the logical block is the smallest unit of transfer.

- The 1-dimensional array of logical blocks is mapped into the sectors of the disk sequentially.
 - Sector 0 is the 1st sector of the 1st track on the outermost cylinder.
 - Mapping proceeds in order through that track,
 - then the rest of the tracks in that cylinder, (電子的)
 - and then through the rest of the cylinders (기계적) from outermost to innermost.



Moving-head Disk Mechanism



Overview of Mass Storage Structure

- Magnetic disks provide bulk of secondary storage of modern computers
 - Drives rotate at 60 to 200 times per second

- **Transfer rate** is rate at which data flow between drive and computer

- **Access Time**
 - **Seek Time**
 - **Positioning time (random-access time)** is time to move disk arm to desired cylinder
 - **Rotational Latency**
 - time for desired sector to rotate under the disk head (**rotational latency**)

- Drive attached to computer via **I/O bus**
 - Busses vary, including **EIDE, ATA, SATA, USB, Fibre Channel, SCSI**
 - **Host controller** in computer uses bus to talk to **disk controller** built into drive or storage array



HDD Practices

□ Seagate 1TB



사양		
• 모델번호	ST3500413AS	ST31000524AS
• 인터페이스	SATA 6Gb/s	SATA 6Gb/s
• 캐시	16MB	32MB
• 용량	500GB	1TB
• 영역 밀도(평균)	329Gb/in ²	329Gb/in ²
• 보증 섹터	976,773,168	1,953,525,168
물리적		
• 높이	19.98mm (0.787 in)	26.1mm (1.028 in)
• 너비	101.6mm (4.000 in)	101.6mm (4.000 in)
• 길이	146.99mm (5.787 in)	146.99mm (5.787 in)
• 무게(일반)	415g (0.925 lb)	622g (1.371 lb)
성능		
• 회전 속도(RPM)	7,200 RPM	7,200 RPM
• 평균 대기 시간	4.16ms	4.16ms
• 무작위 읽기 탐색 시간	< 8.5ms	< 8.5ms
• 무작위 쓰기 탐색 시간	< 9.5ms	< 9.5ms
신뢰성		
• 연간 고장율	0.32%	0.32%
전원		
• 최대 시작 전류, DC	2	2

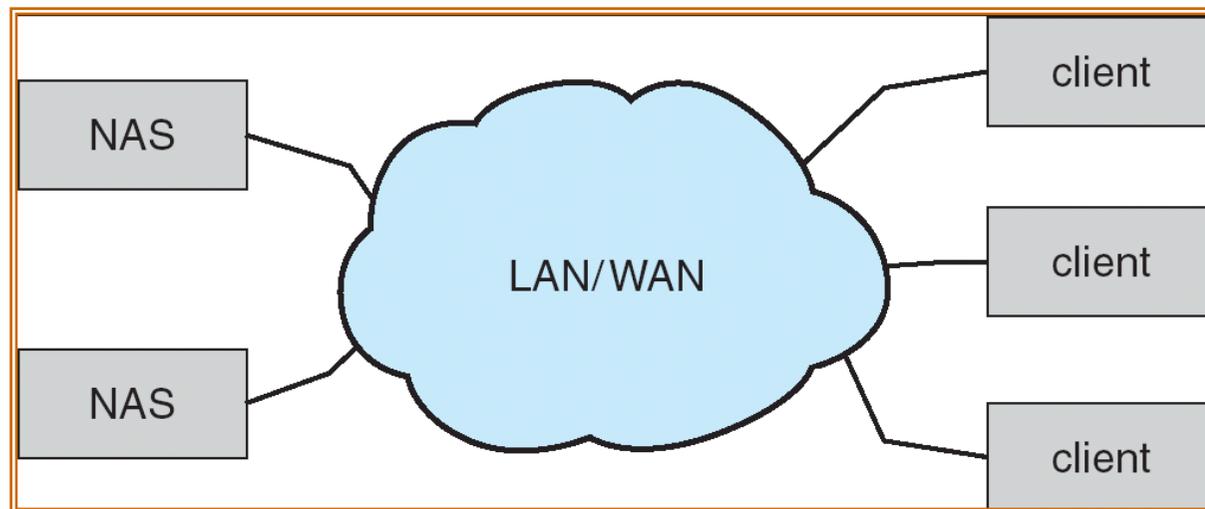
HDD Practices

용량	3TB	2TB
모델번호	ST3000DM001	ST2000DM001
인터페이스	SATA 6Gb/s NCQ	
회전속도	7200RPM	
캐시메모리	64MB	
평균 데이터 전송속도	156MB/s	
최대 데이터 전송속도	210MB/s	
Seagate OptiCache™	Y	
Seagate AcuTrac™	Y	
Seagate SmartAlign™	Y	
크기	101.6 x 146.99 x 26.1 (mm)	
KCC 인증번호	KCC-REM-STX-Barracuda	
정격전압/소비전력	+5V 0.75A, +12V 0.75A / 3.75w, 9w	
동일모델의 출시년월	2011년 6월	
제조사	Seagate Technology	
수입자	Ocean Technology, Daewon CTS, PC Direct, Peach Valley	
제조국	중국, 태국	



Network-Attached Storage(NAS)

- ❑ Network-attached storage (**NAS**) is storage made available over a network rather than over a local connection (such as a bus)
- ❑ NFS and CIFS are common protocols
- ❑ Implemented via remote procedure calls (RPCs) between host and storage
- ❑ New iSCSI protocol uses IP network to carry the SCSI protocol



Network-Attached Storage(NAS)



H/W 제품사양

제품명	ipTIME NAS-II
주요사항	NAS / 3.5인치 2베이 / HDD 미탑재 / 기가비트 유선 /Torrent 지원
포트	DC In, 100/1000Mbps LAN x1, USB 2.0 x1
버튼	Power, RST/USB Copy
LED	Power, Ready, Gigabit, LAN, HDD1, HDD2, USB
CPU	ARM 32bit RISC Processor (256KB L2 Cache)
메모리	256MB (DDR2 DRAM)
지원 하드디스크	8.9Cm (3.5) SATA HDD x2
지원 USB 디스크	NTFS / FAT32 / EXT2 / EXT3
크기	198.5 x 153.5 x 67 mm
무게	752g (HDD 미포함)
최대소비전력	54W
전원	12V 외장 어댑터
OS	Windows 2000, XP, Vista, 7, Linux, Mac OS



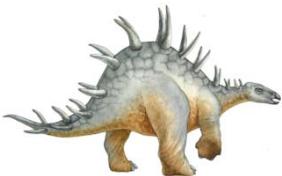
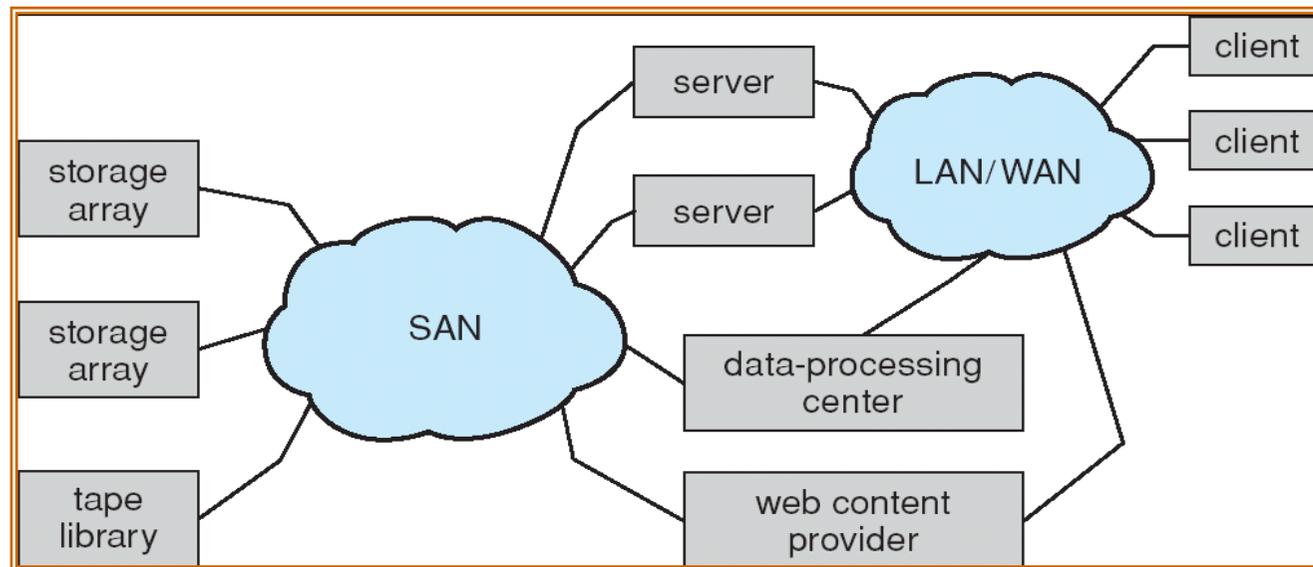
<http://prod.danawa.com/list/?defSite=PC&cate1=862&cate2=32621>

http://www.11st.co.kr/product/SellerProductDetail.tmall?method=getSellerProductDetail&prdNo=737916686&service_id=pcdn

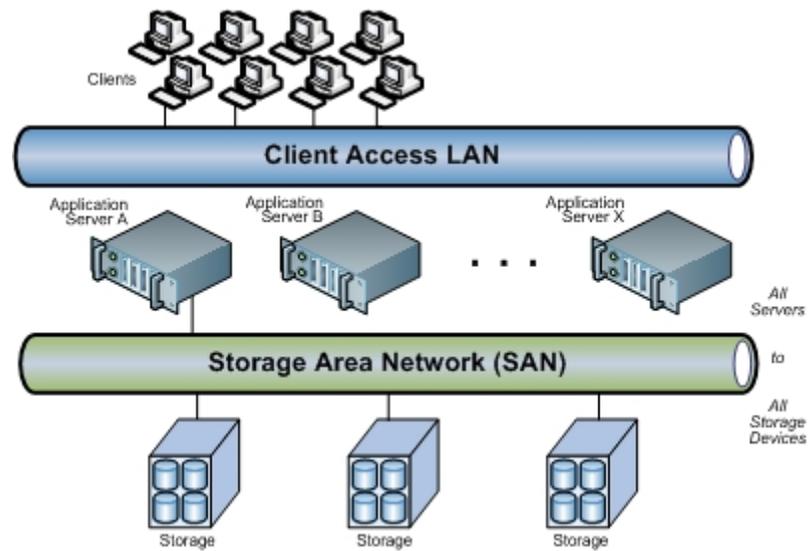


SAN(Storage Area Network)

- ❑ Common in large storage environments (and becoming more common)
- ❑ Multiple hosts attached to multiple storage arrays - flexible



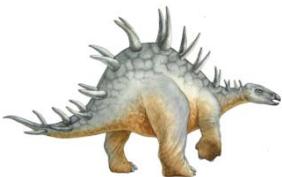
SAN(Storage Area Network)





Disk Scheduling

- ❑ The operating system is responsible for using hardware efficiently — for the disk drives, this means having a fast access time and disk bandwidth.
- ❑ Access time has two major components
 - *Seek time* is the time for the disk are to move the heads to the cylinder containing the desired sector.
 - *Rotational latency* is the additional time waiting for the disk to rotate the desired sector to the disk head.
- ❑ Minimize seek time
- ❑ Seek time \approx seek distance
- ❑ Disk bandwidth is the total number of bytes transferred, divided by the total time between the first request for service and the completion of the last transfer.



Disk Scheduling (Cont.)

- ❑ Several algorithms exist to schedule the servicing of disk I/O requests.
- ❑ We illustrate them with a request queue (0-199).

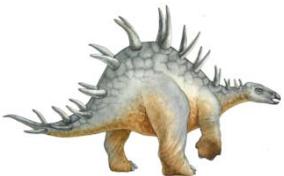
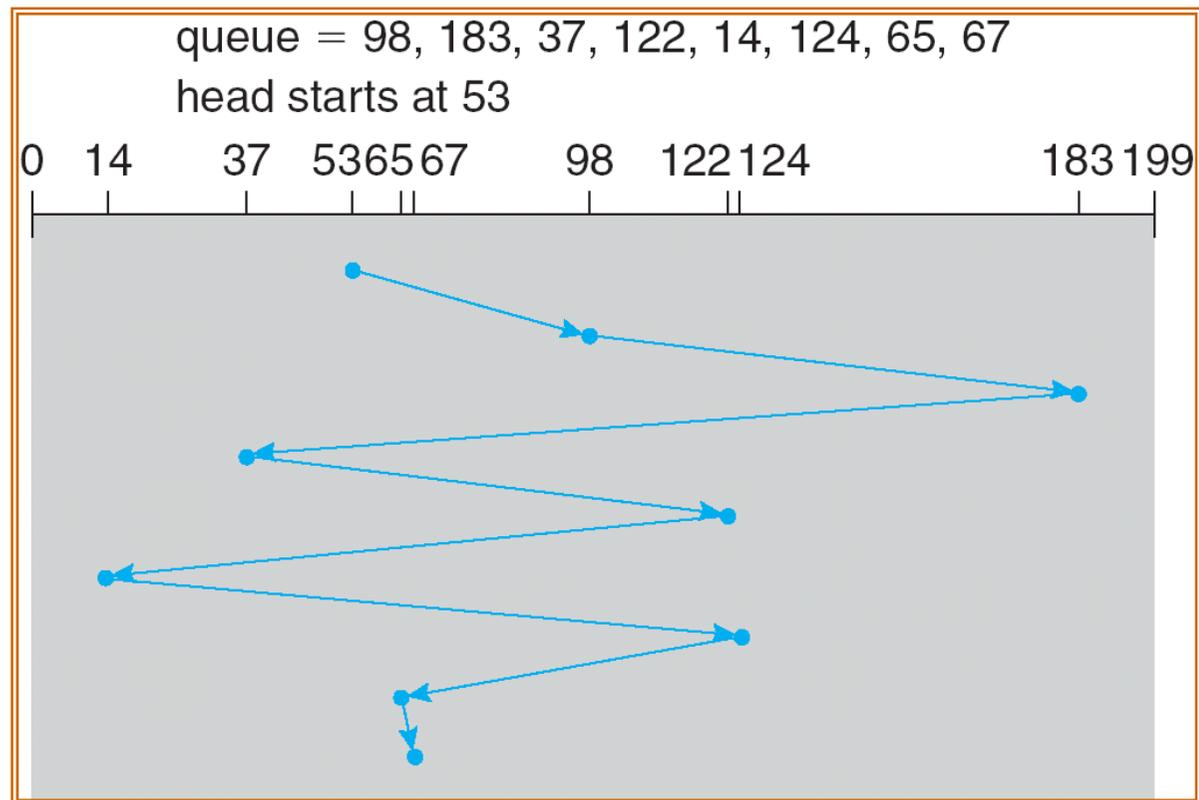
98, 183, 37, 122, 14, 124, 65, 67

Head pointer 53



FCFS

Illustration shows total head movement of 640 cylinders.

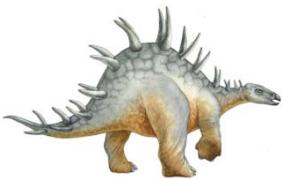
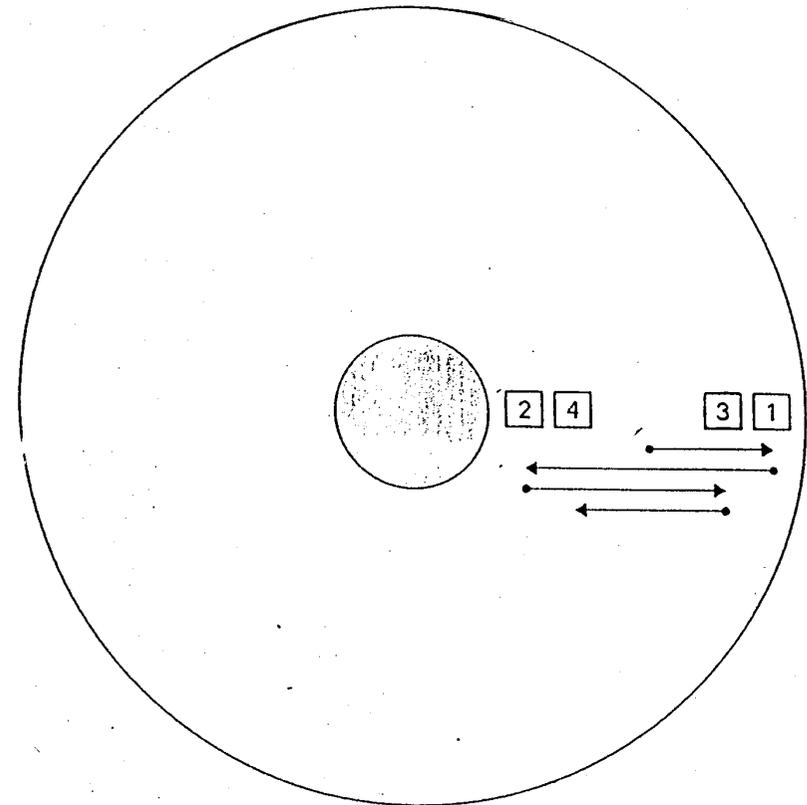


FCFS

■ FCFS (First-Come-First-Served)

스케줄링

- 요청이 도착하는 순서대로 처리
- Disk 부하(load)가 적을 때 유효



SSTF

- Shortest Seek Time First
- Selects the request with the minimum seek time from the current head position.
- SSTF scheduling is a form of SJF scheduling;
 - may cause starvation of some requests.
- 문제점 : 외진 곳이 불리 – unpredictable performance

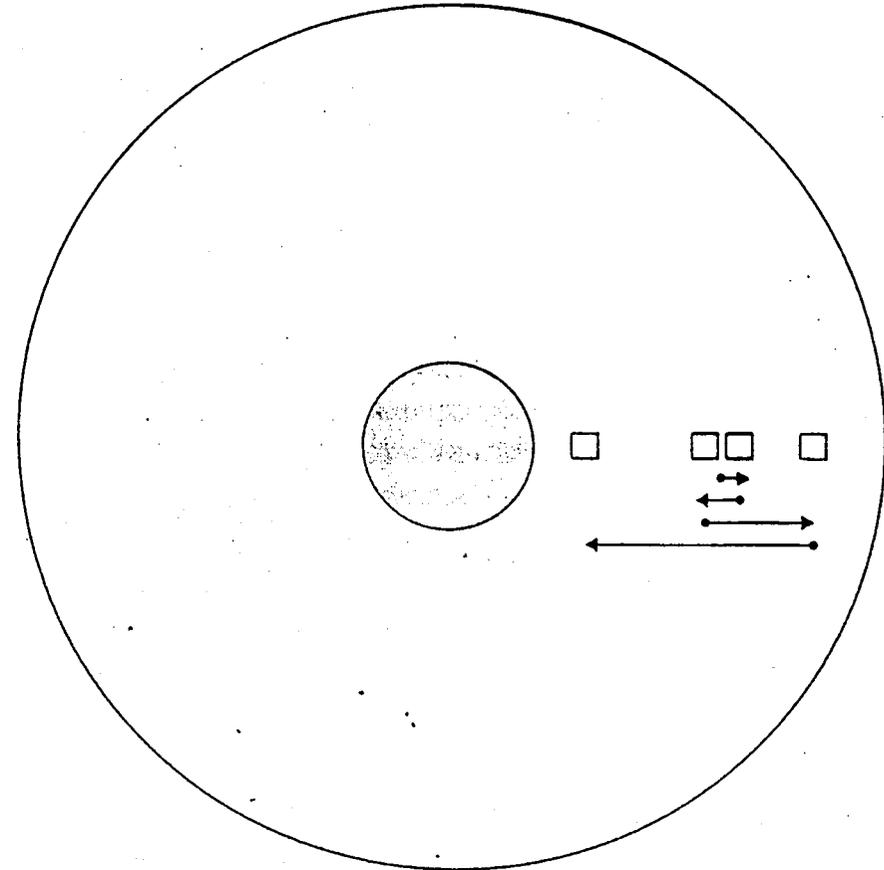


SSTF

■ SSTF(Shortest Seek Time First)

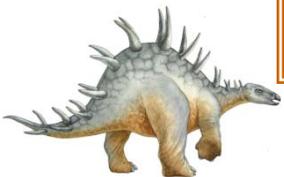
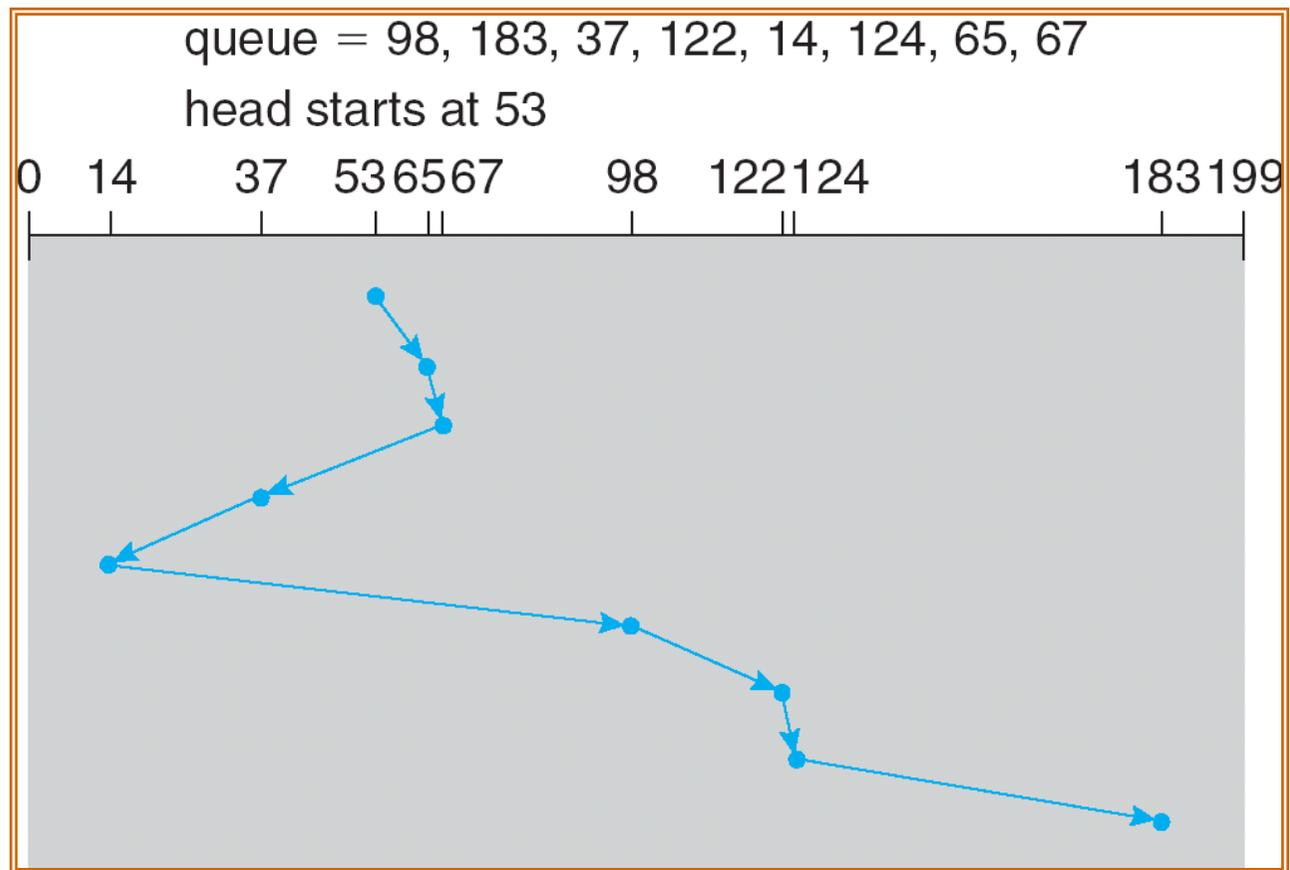
스케줄링

- Seek Distance가 짧은 요청을 먼저 처리
- 바깥쪽 track의 처리 밀도가 낮아 Starvation가능
- 응답 시간의 편차가 큼



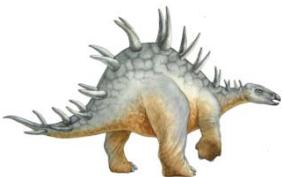
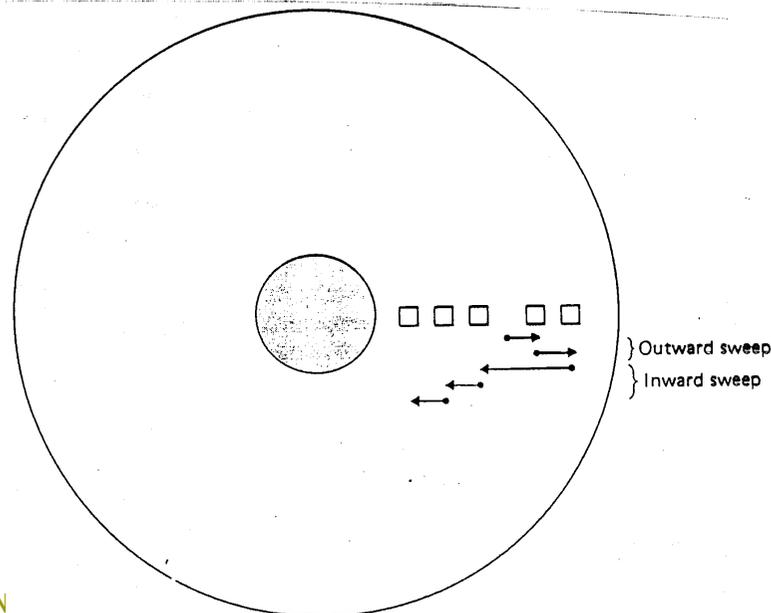
SSTF (Cont.)

shows total head movement of 236 cylinders.



SCAN

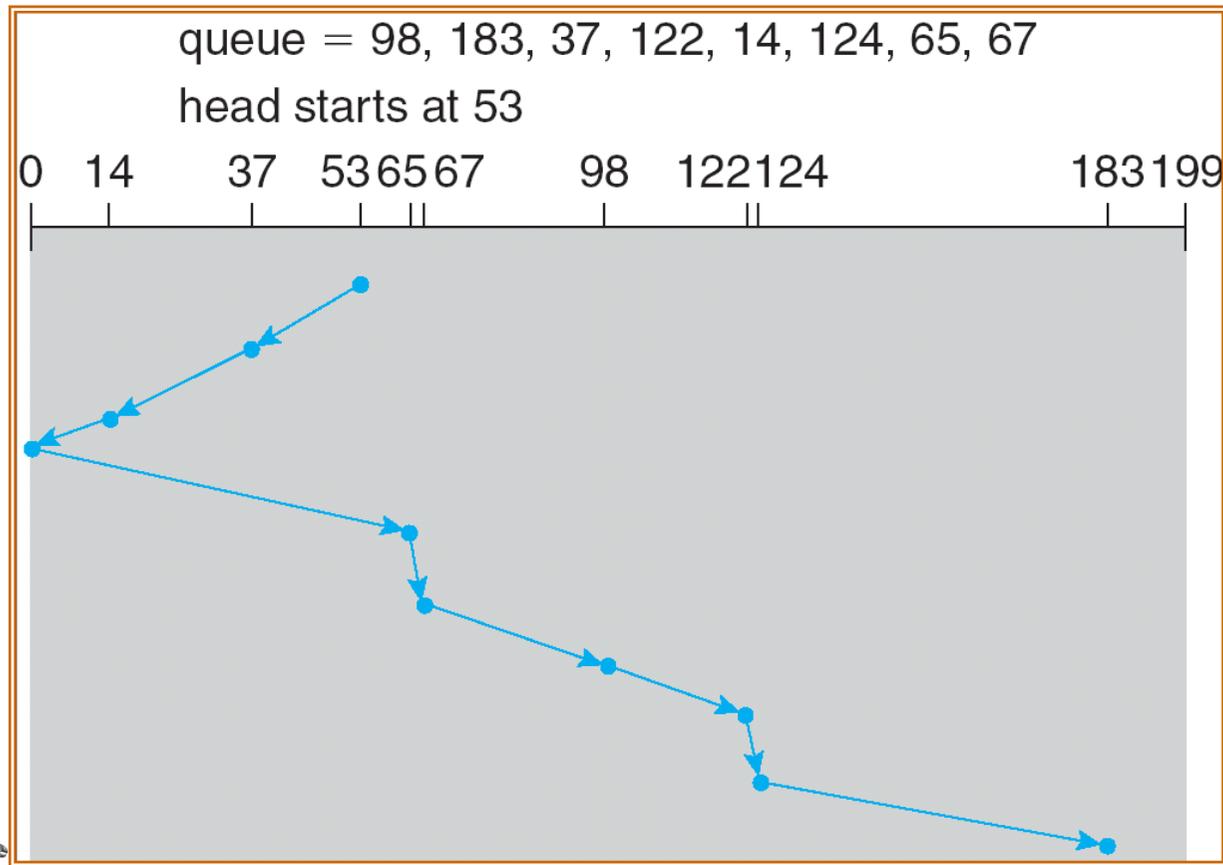
- SCAN스케줄링 (Elevator Algorithm)
 - Head의 진행 방향에서 가장 짧은 거리의 요청을 처리
 - 가운데에 비해 바깥쪽이 차별대우
- Sometimes called the *elevator algorithm*.
 - First, service all requests while going up
 - Then service all requests while going down



SCAN (Cont.)

Illustration shows total head movement of 208 cylinders.

문제점: 가장자리 1회 때마다 - 한가운데는 2번의 기회
--- unfair wait time



C-SCAN

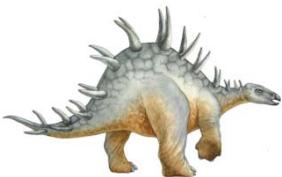
- Provides a more uniform wait time than SCAN.

The head moves from one end of the disk to the other servicing requests as it goes.

When it reaches the other end, however, it immediately returns to the beginning of the disk, without servicing any requests on the return trip.

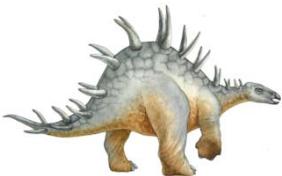
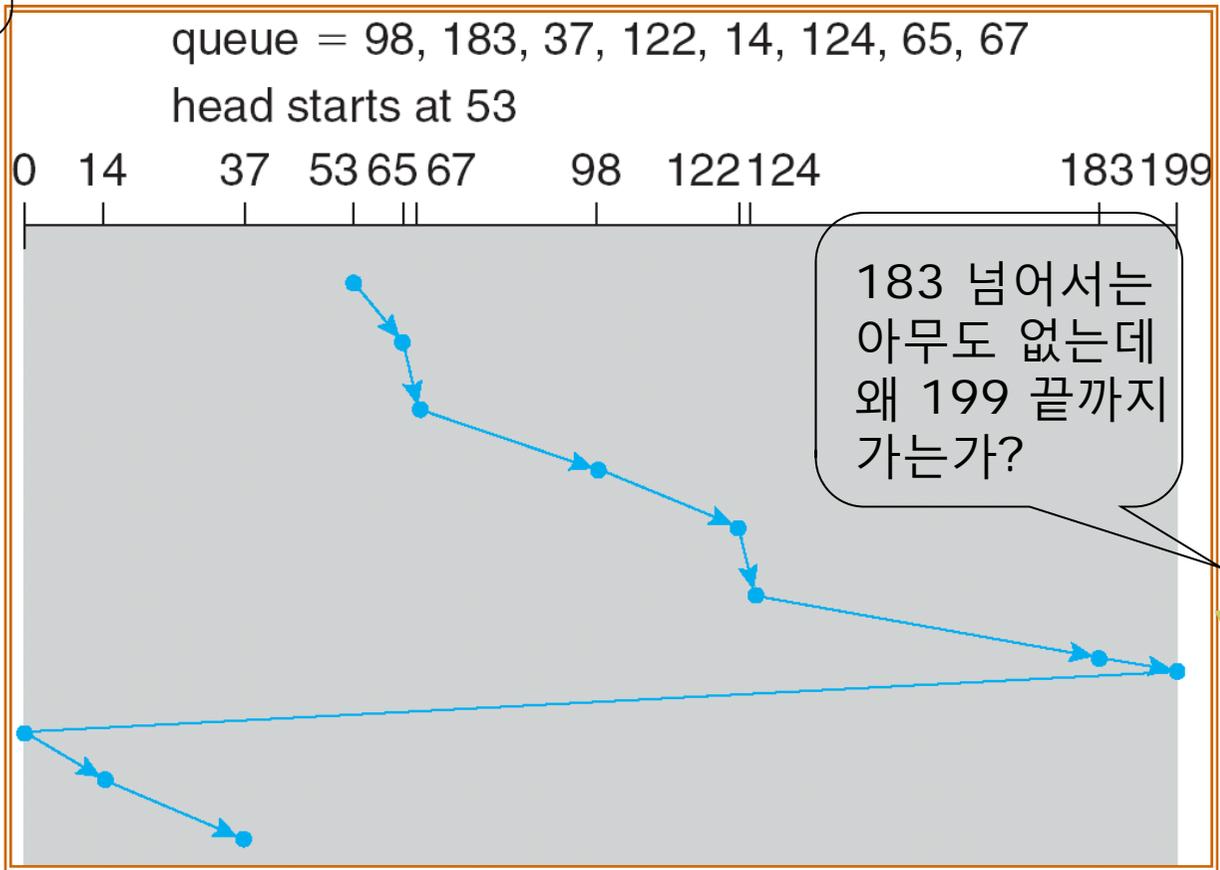
- Treats the cylinders as a circular list that wraps around from the last cylinder to the first one.

- 문제: 가고있는 방향 전방에 아무런 pending request 들이 없는데도 계속 전진?



C-SCAN (Cont.)

14 넘어서는 아무도 없는데 왜 끝까지 가는가?

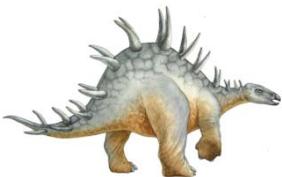


C-LOOK

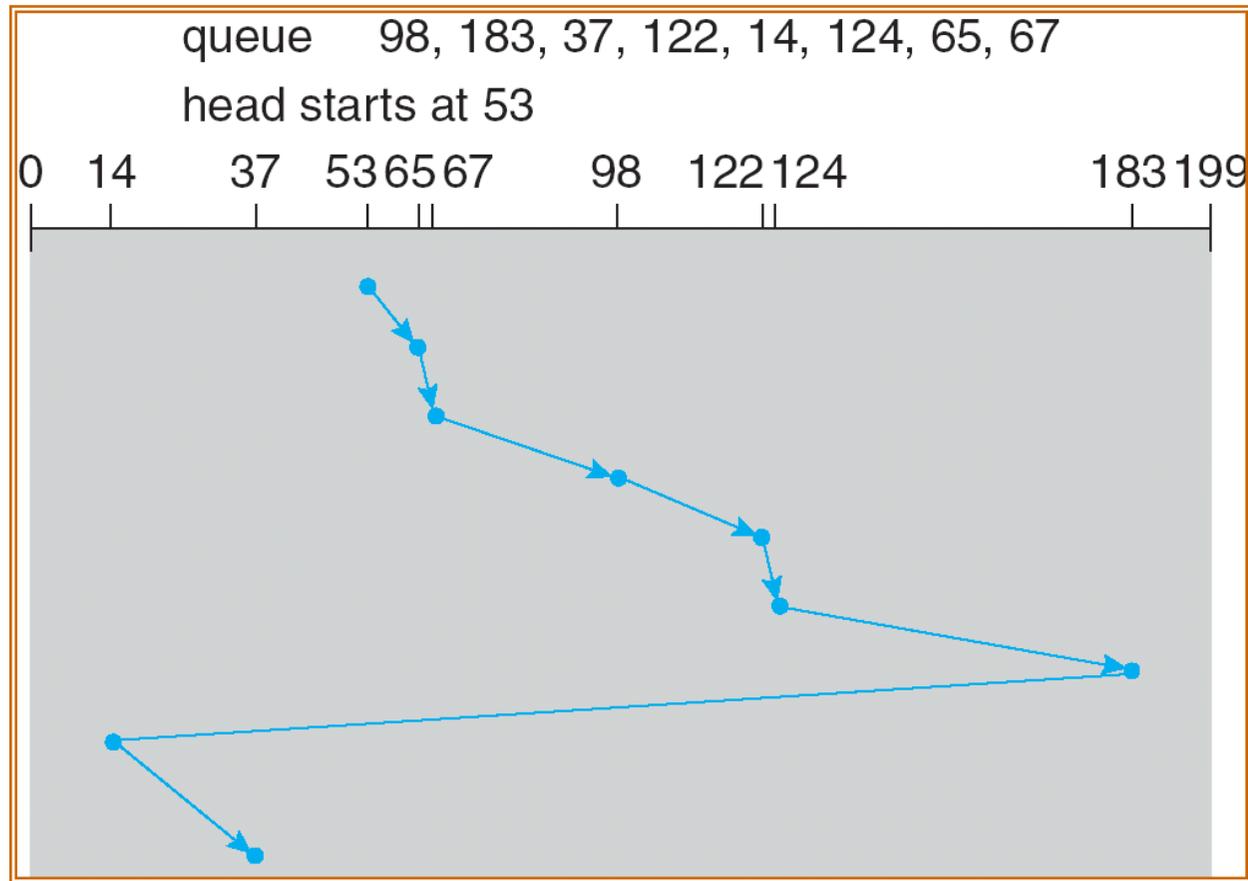
- Version of C-SCAN

Look-for-Request before continuing

Arm only goes as far as the last request in each direction,
then reverses direction immediately

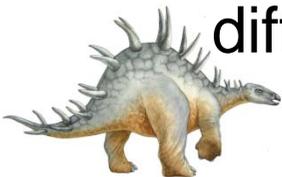


C-LOOK (Cont.)



Selecting a Disk-Scheduling Algorithm

- ❑ SSTF is common and has a natural appeal
- ❑ Either SSTF or LOOK is a reasonable choice for the default algorithm.
- ❑ SCAN and C-SCAN perform better for systems that place a heavy load on the disk.
- ❑ Performance depends on
 - the number and types of requests.
 - Requests influenced by the file-allocation method.
- ❑ The disk-scheduling algorithm -- written as a separate module, allowing it to be replaced with a different algorithm if necessary



Disk Management

physical formatting or Low-level formatting

Dividing a disk into sectors that the disk controller can R/W.

Each sector :

[head + data(보통 512 B) + trailer]

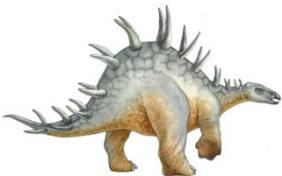
** 미줄부분 --- used by controller

header/trailer contain

sector number,

ECC (controller 용)

- spare sectors/cylinder are reserved for bad blocks



Disk Management

- To use a disk, OS needs its own data structures on the disk.

(1) Partition the disk into one or more groups of cylinders

OS는 이것을 독립적 disk로 취급

각 partition 내에서 logical formatting

(2) Logical formatting or “making a file system”.

data structure for file system, used by OS

space management

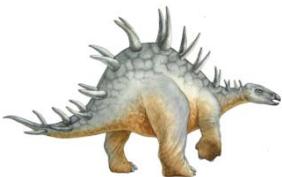
free space

busy space

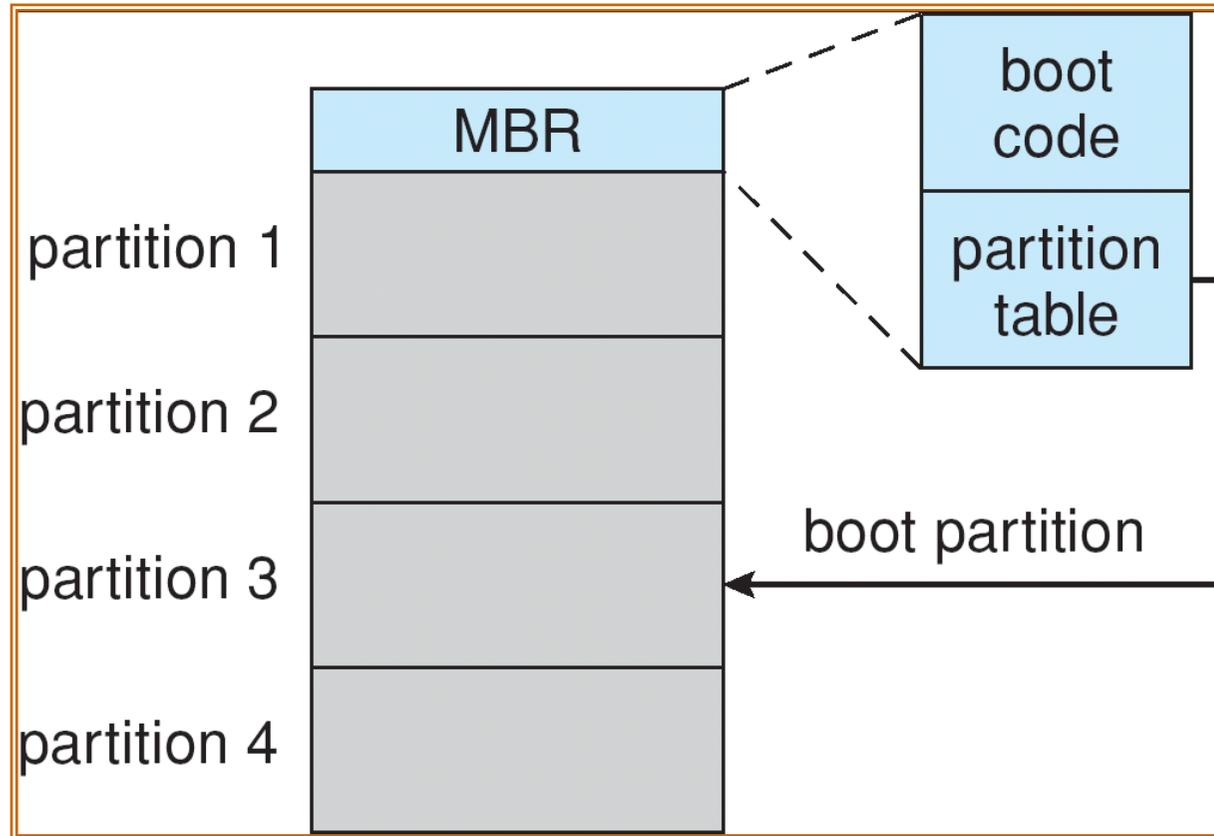
–

일반 파일

directory 파일



Booting from a Disk in Windows 2000

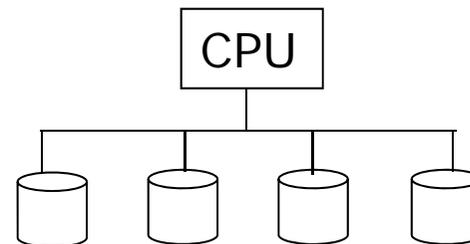
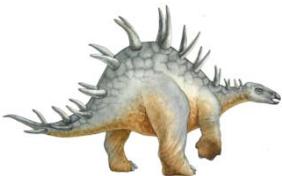


MBR : Master Boot Record



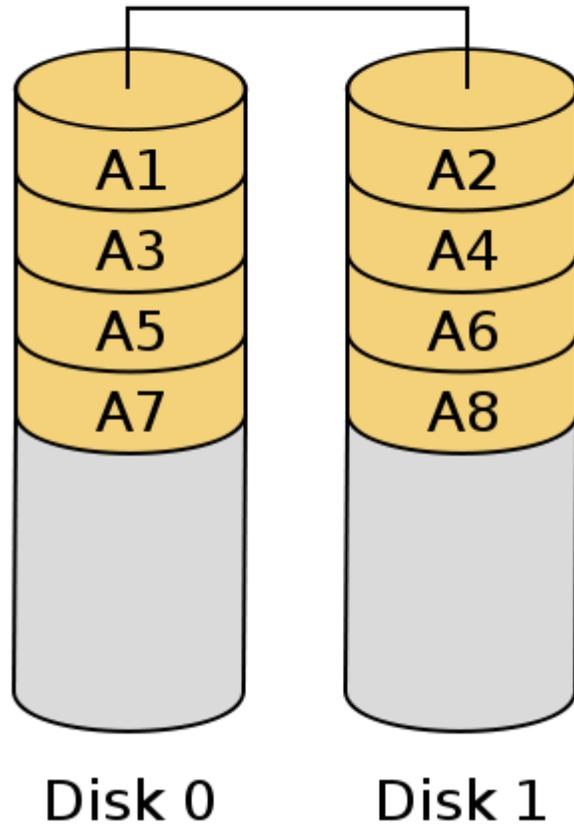
RAID (cont)

- Multiple disks working cooperatively.
- RAID (Redundant Array of Independent Disks)
 - improve performance and/or reliability of the storage.
 - (여러 디스크 병렬처리) (백업용 추가 정보)
 - (복수 개의 disk들을 어떻게 사용하느냐에 따라)
 - Mirroring (or shadowing) keeps duplicate of each disk.
 - Block interleaved parity uses much less redundancy (예: 1/4).
 - Disk striping (or interleaving) transfers N disks in parallel
 - 한번 seek/rotation 후에는 4 배의 transfer
 - Some RAIDS synchronize rotations



RAID (cont)

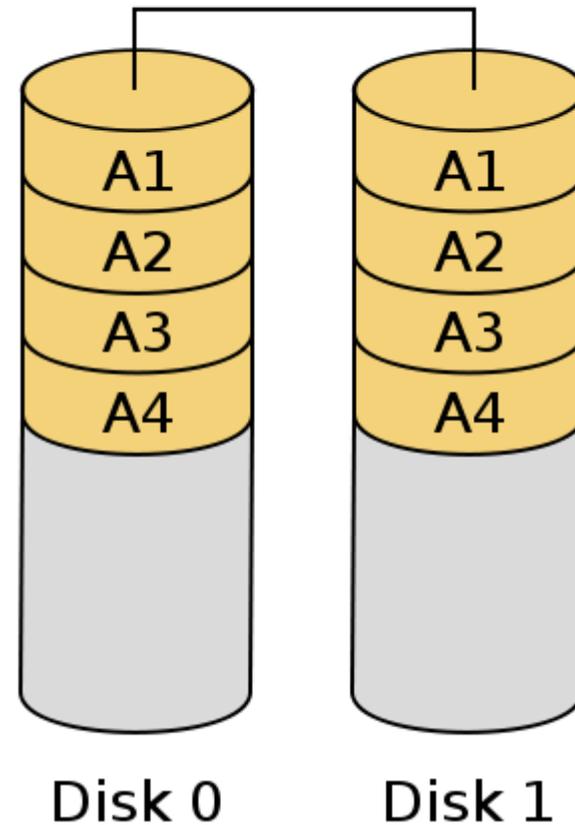
RAID 0



striping



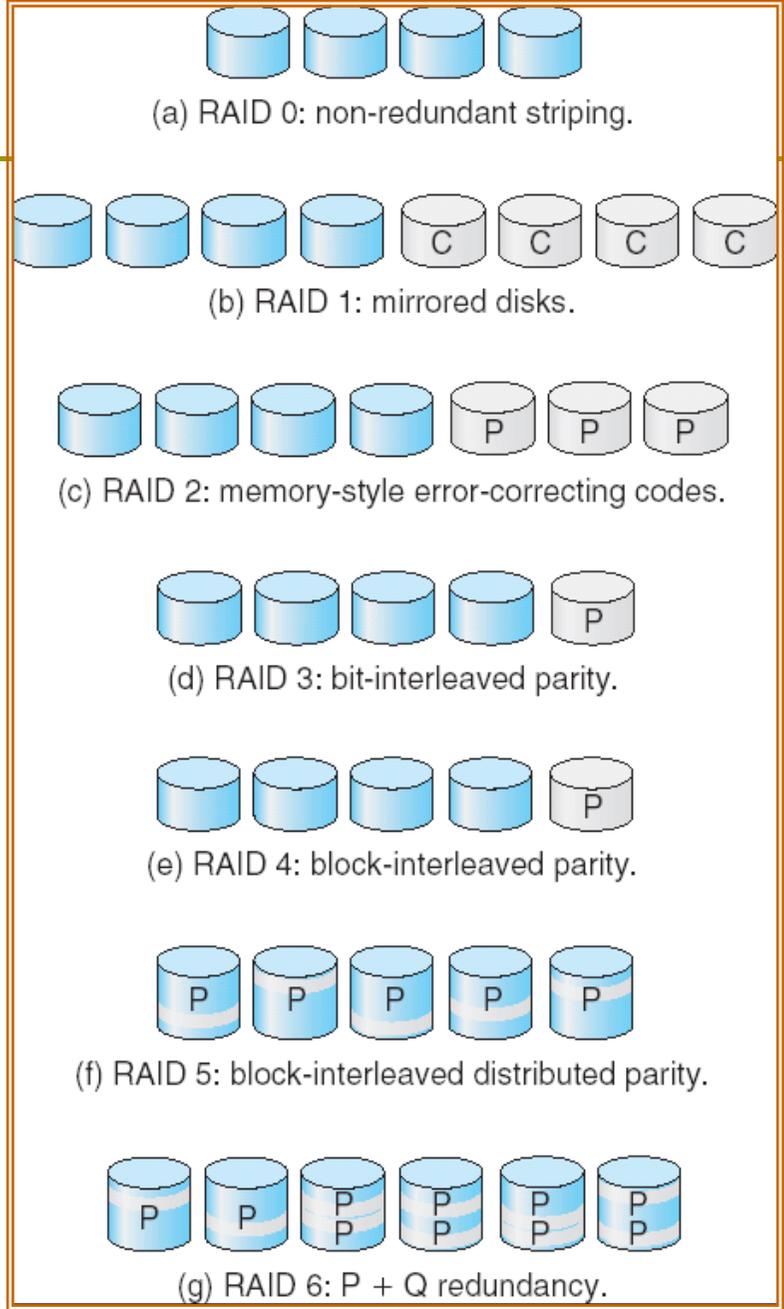
RAID 1



mirroring

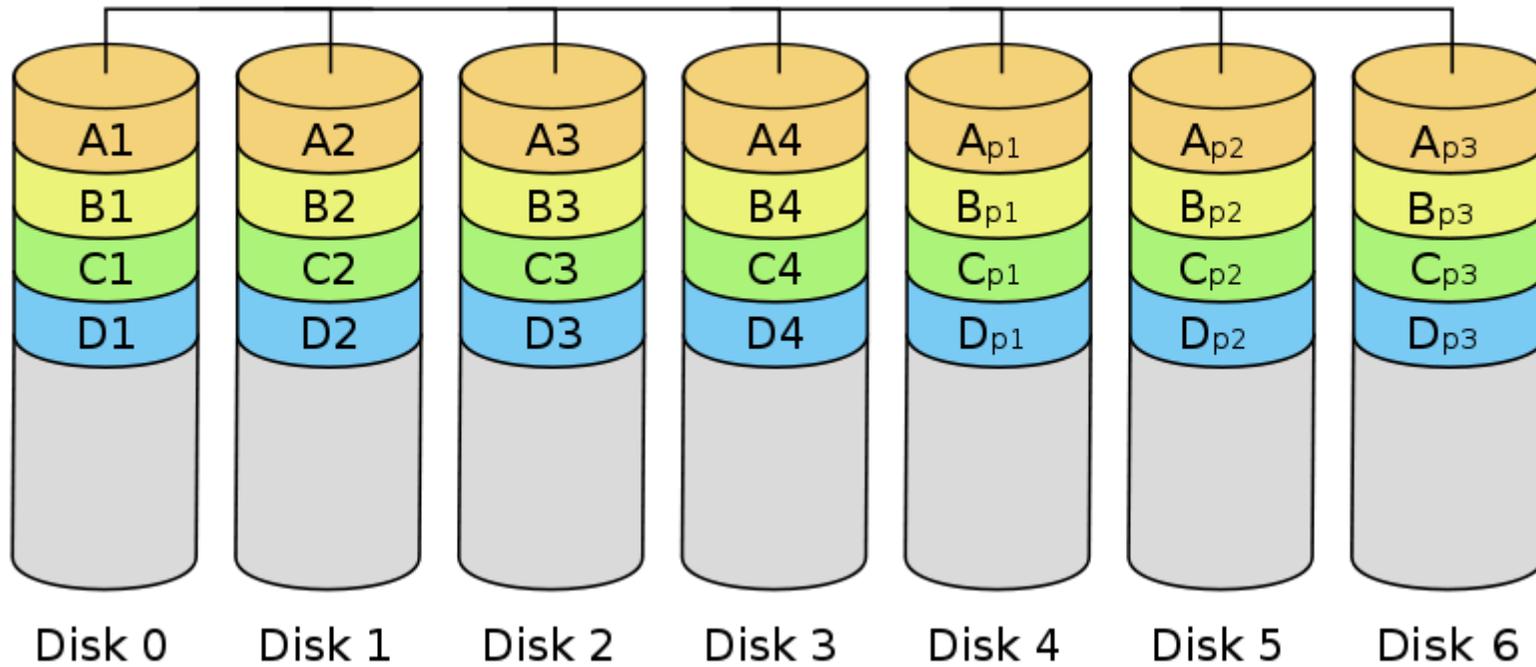


RAID Levels



RAID (cont)

RAID 2

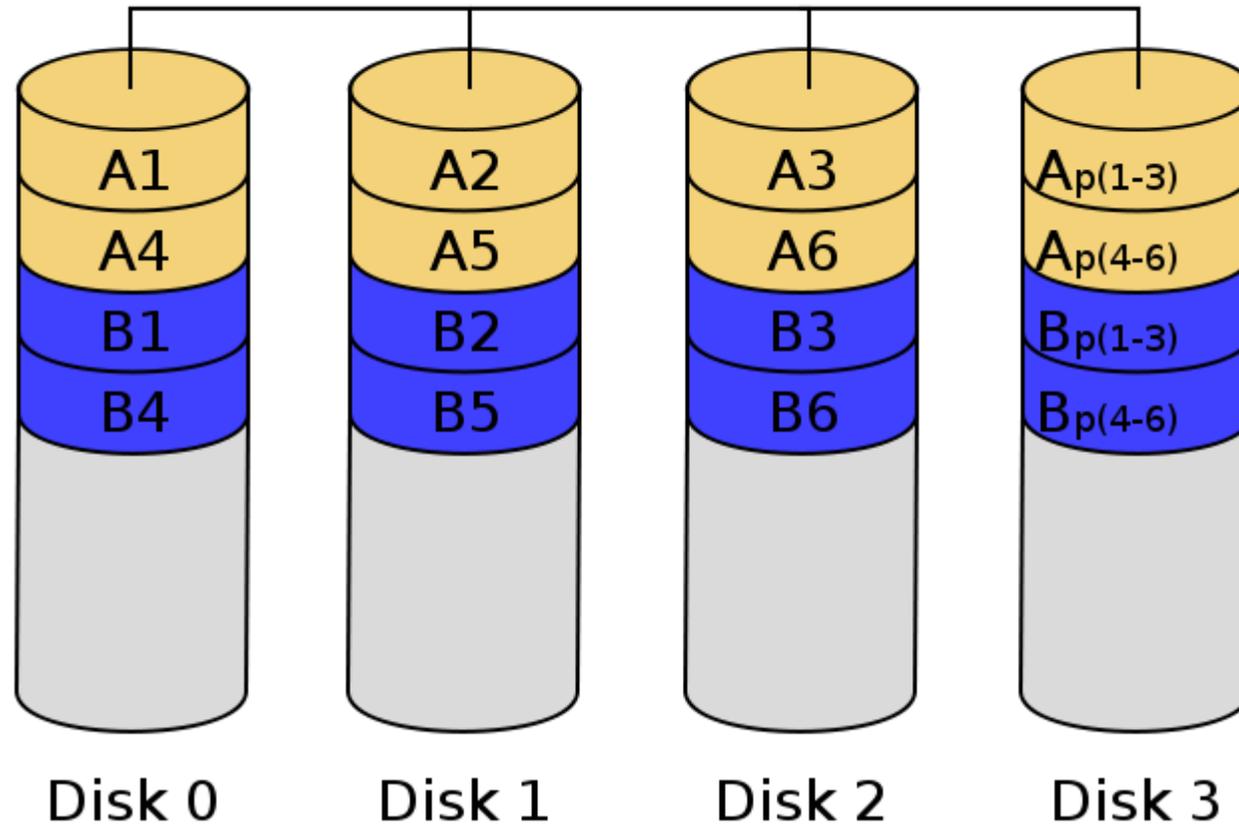


parity disk



RAID (cont)

RAID 3

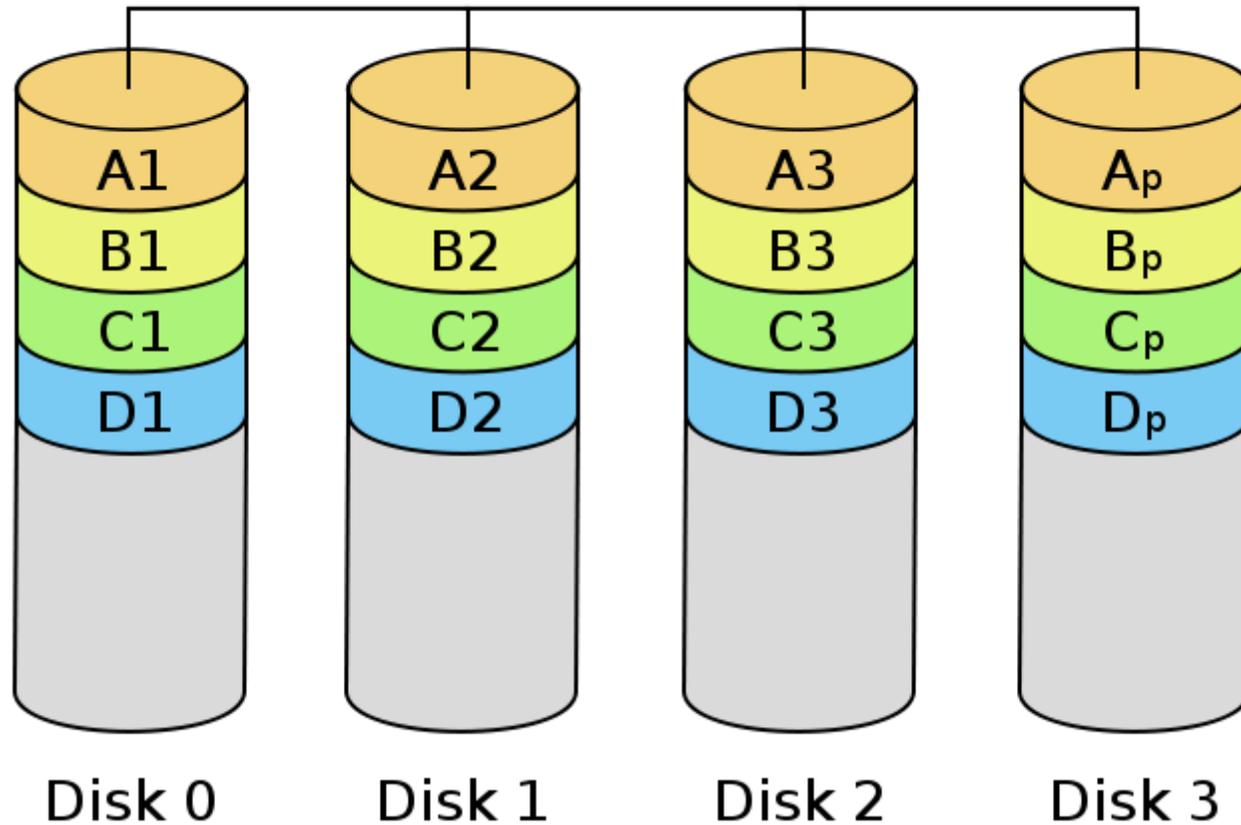


byte-level striping



RAID (cont)

RAID 4

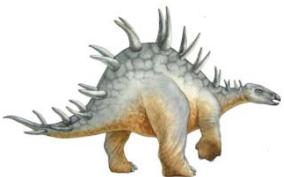
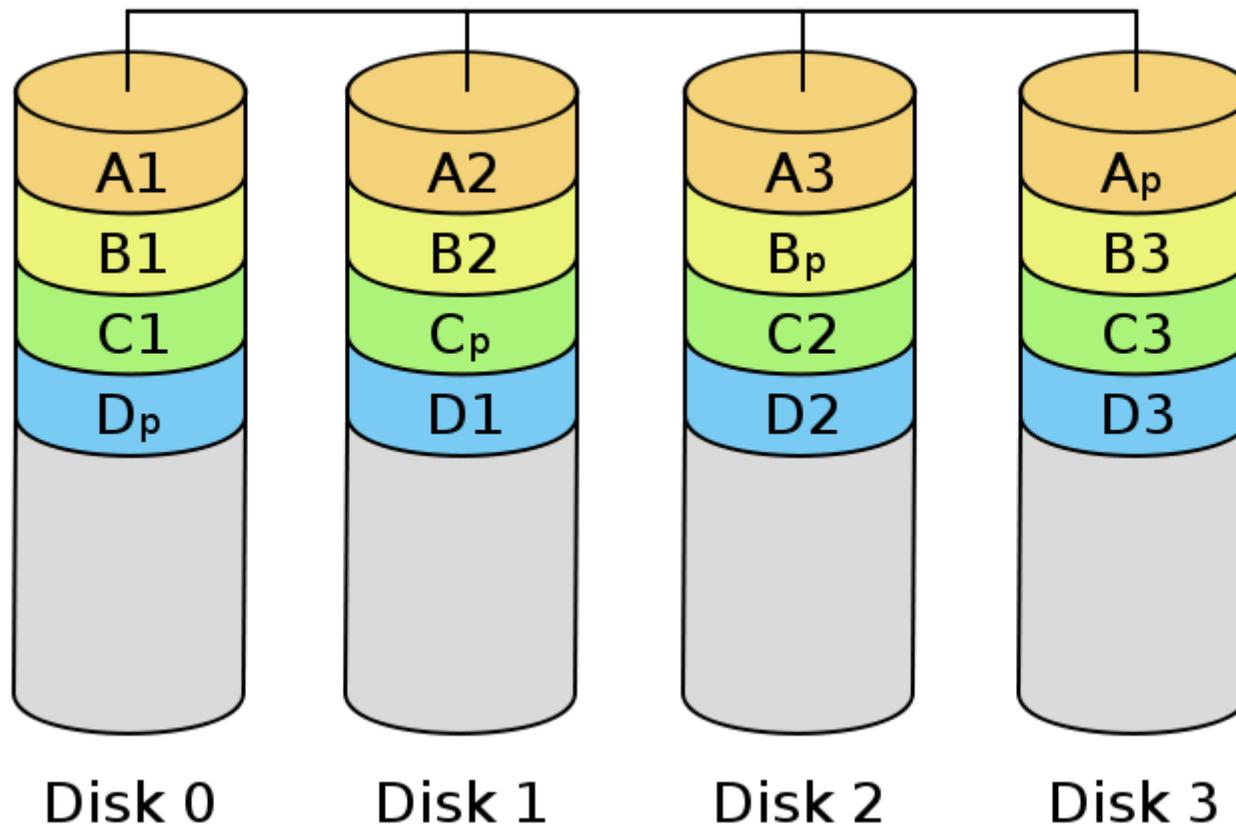


block-level striping



RAID (cont)

RAID 5

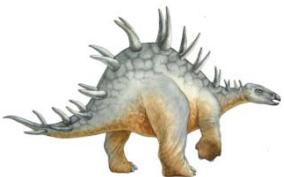
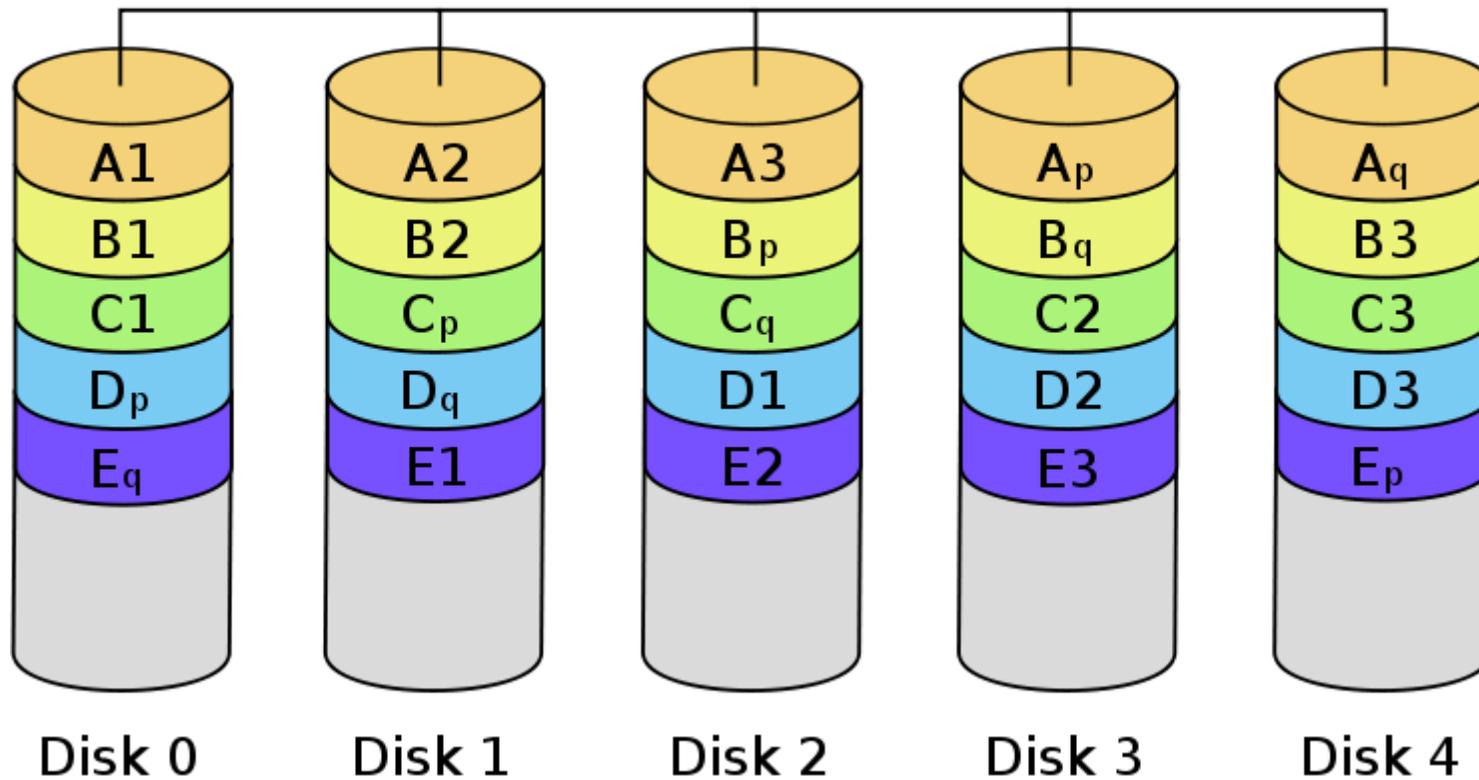


block-level striping, distributed parity



RAID (cont)

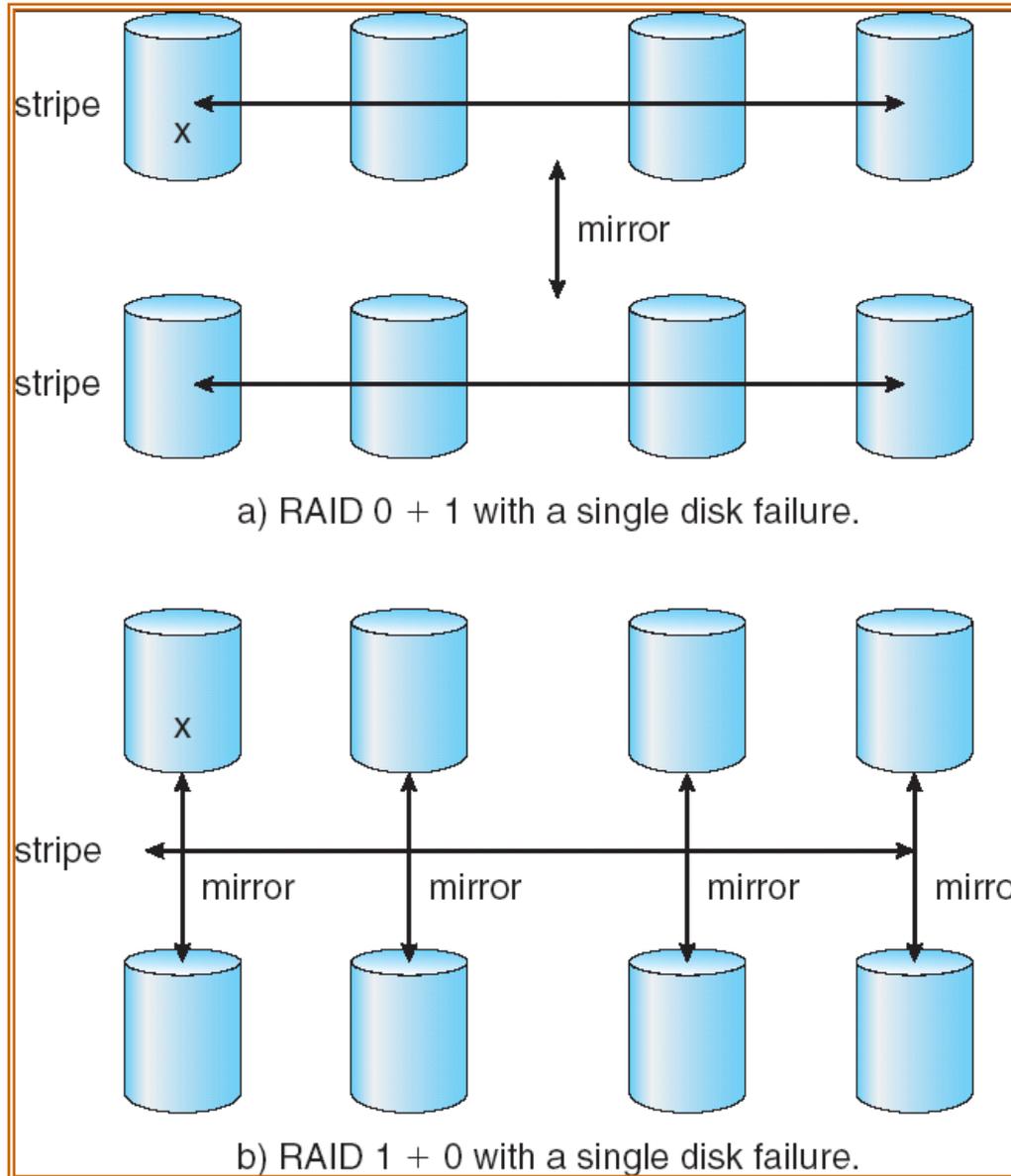
RAID 6



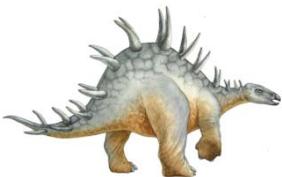
2개 이상의 disk failure



RAID (0 + 1) and (1 + 0)

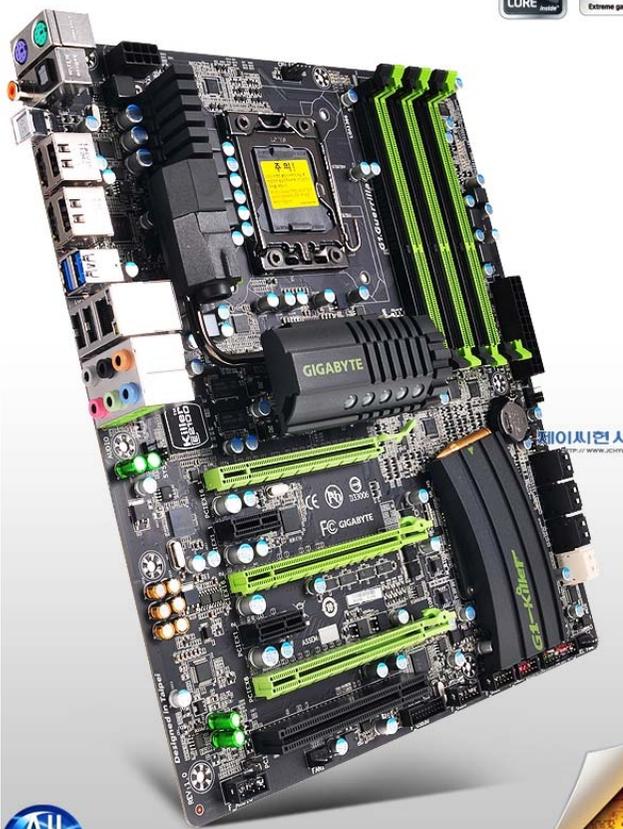


RAID 10



Practice

GIGABYTE] G1.GUERRILLA 메인보드(X58)



구분	상세 내용
플랫폼	ATX
소켓 규격	LGA 1366
CPU	Intel Core™ i7 - L3 cache varies with CPU
칩셋	Intel X58 / ICH 10R
메모리	* DDR3 2200/1333/1066/800 지원, 최대 24GB 지원 듀얼 / 트리플 채널 지원 DIMM 슬롯 6개
확장슬롯	NVIDIA SLI™ / ATI CrossFireX™ - PCI-Express x16(x16) 슬롯 2개 - PCI-Express x16(x8) 슬롯 1개 - PCI-Express x1 슬롯 2개 - PCI 슬롯 1개
저장장치	* S-ATA II 3.0Gb/s 커넥터 6개 - Intel Matrix Storage Technology RAID 0, 1, 5 and 10 * S-ATA III 6.0Gb/s 커넥터 2개 - Intel Matrix Storage Technology RAID 0, 1 * eSATA 3.0Gb/s 커넥터 2개(백패널) - Intel Matrix Storage Technology RAID 0, 1 and JBOD 지원
오디오	Realtek ALC889 8-channel High Definition 오디오 - X-Fi Xtreme Fidelity® and EAX® Advanced HD™ 5.0 기술지원 - S/PDIF in/out
LAN	기가비트 랜 Bigfoot Killer E2100 chip 10/100/1000 LAN
USB	USB2.0 총 10개(백패널 4개/온보드헤더 6개) 지원 * USB3.0/2.0 4개 (백패널 2개/온보드헤더 2개)
IEEE1394	-

제이치현시스템(주)
HTTP://WWW.JCHYUN.COM