



# 플래시 메모리 기반 시스템 소프트웨어

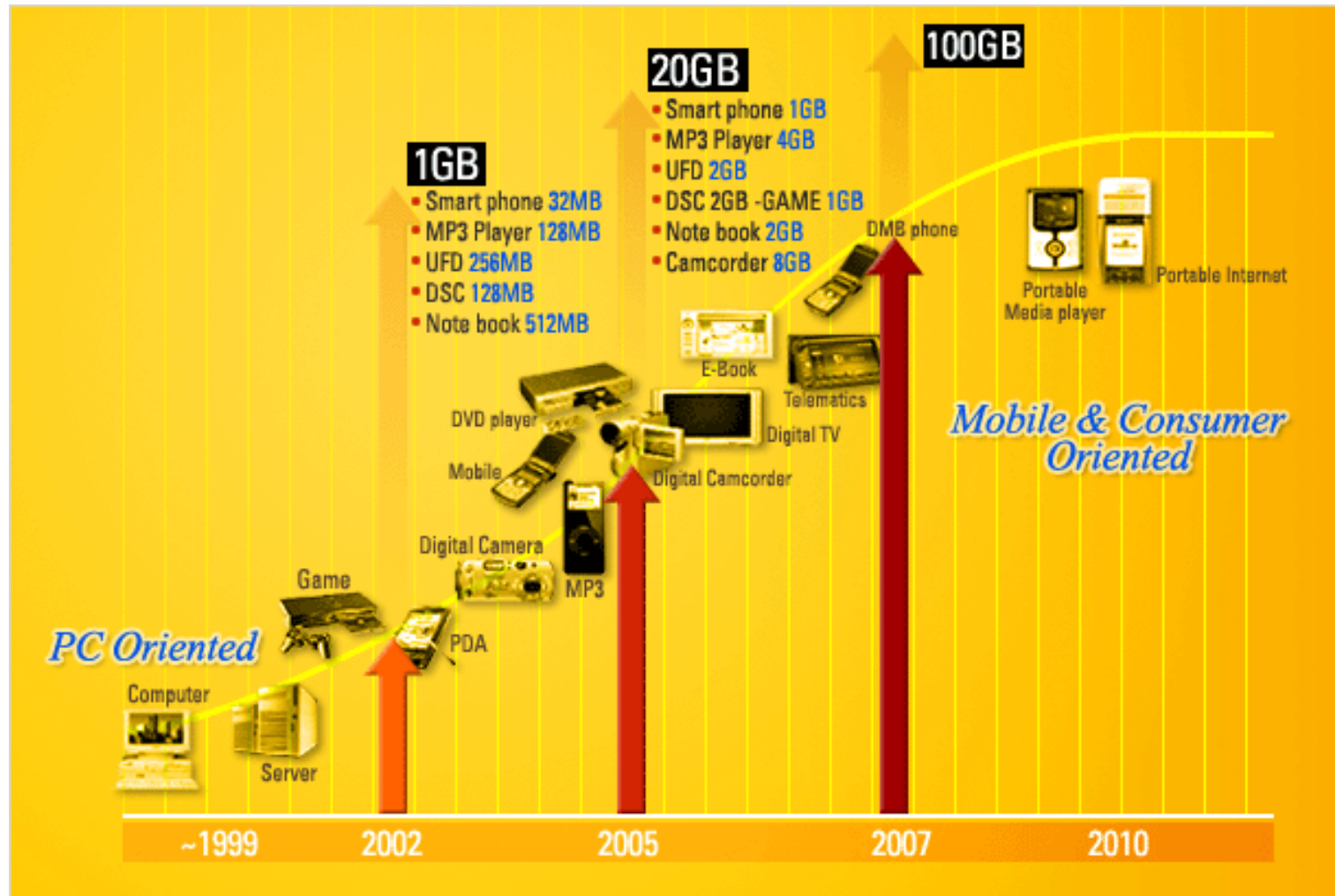
---



# NAND 플래시 메모리

---

# Portable Storage Applications



<http://www.samsung.com/Products/Semiconductor/NANDFlash/index.htm>

# HDD vs. Flash Memory

- HDD 기반 저장장치

- 자기 디스크에 데이터 기록
- 가격 및 성능에서 유리
- 내구성/전력소비/부팅시간에서 불리
- 내장형: 1-inch HDD
- 외장형: Micro Drive

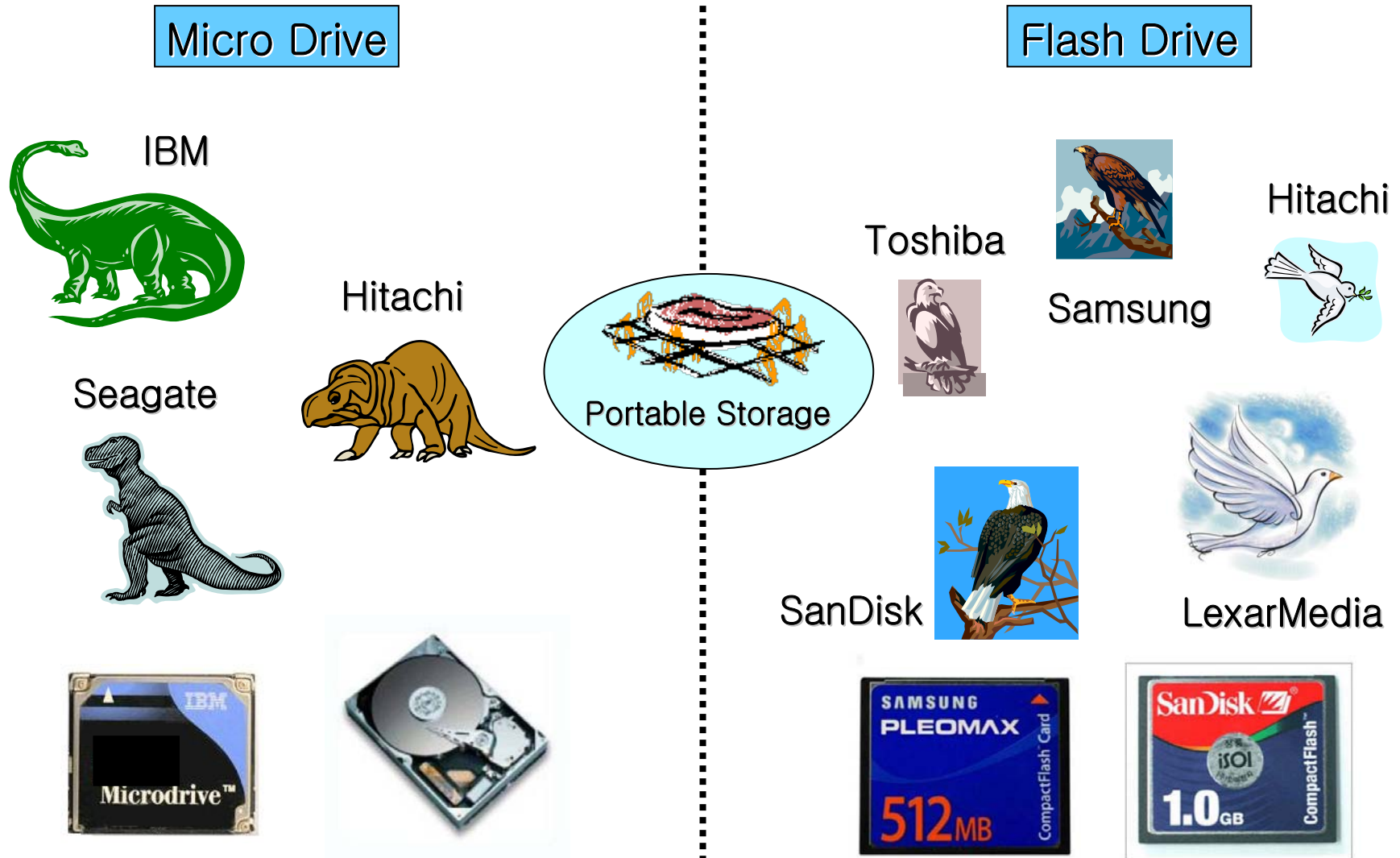


- (NAND) Flash Memory 기반 저장장치

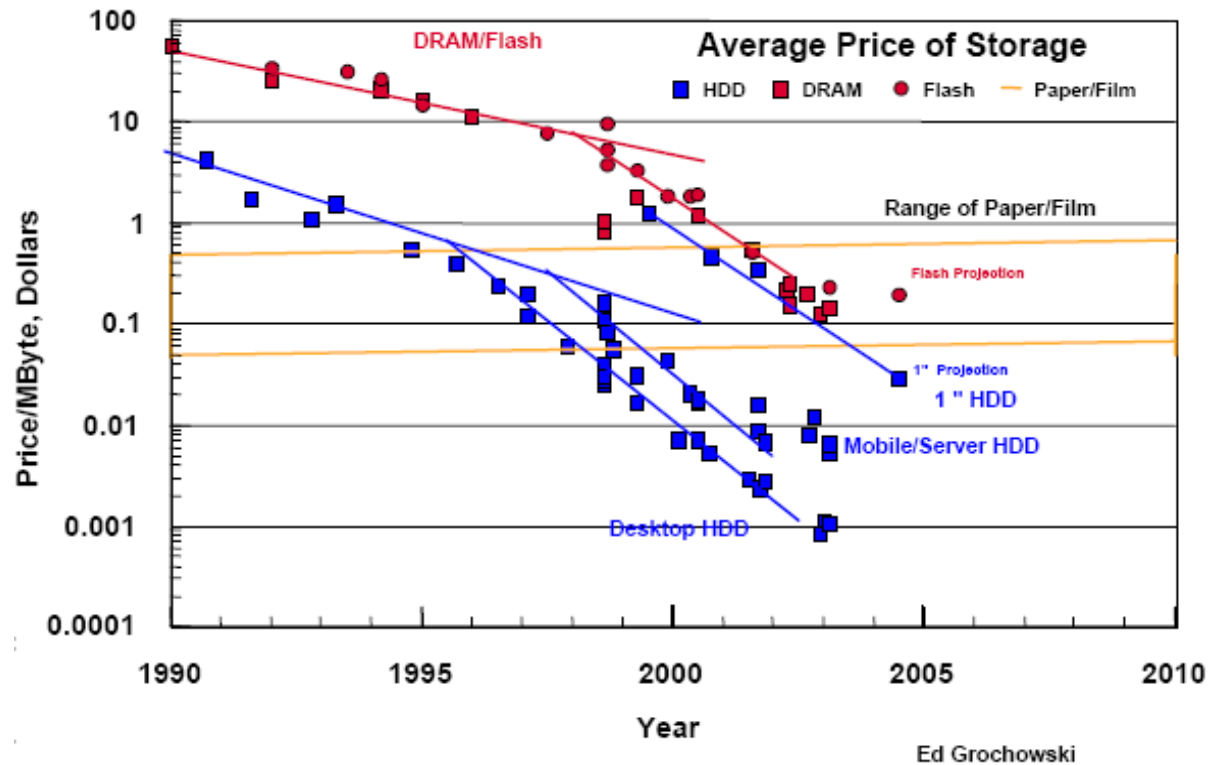
- 플래시 메모리 반도체에 데이터 기록
- 내구성/전력소비/부팅시간에서 유리
- 가격 및 성능에서 불리  
(최근 간격이 좁혀지고 있음)
- 내장형: NAND flash memory chip
- 외장형: CF/SD/MMC card, Memory Stick, USB Drive, Flash SSD (Solid State Disk)



# HDD vs. Flash Memory



# Cost Comparison (2005. 4)



\$153.60(2005.4)



\$78.99(2005.4)



source: <http://www.hitachigst.com/hdd/technolo/overview/chart03.html>

# Why flash memory?

- Faster access
- Lower power
- Shock / Temperature resistance
- Smaller size
- Lighter weight
- Noiseless



# 플래시 메모리 개요

- 플래시 메모리
  - 전원 공급이 중단된 뒤에도 저장 데이터가 보존되는 메모리 반도체
  - 한번 기록한 영역은 **erase** 한 뒤에 **write** 해야 하는 특성이 있음
- 플래시 메모리의 종류
  - NOR 타입 플래시 메모리
    - 바이트 단위 어드레싱 가능 (ROM 처럼 동작)
    - 주로 코드 실행용으로 사용됨
    - Read 성능 측면에서 다소 유리
  - NAND 타입 플래시 메모리
    - 페이지 단위 (**read / write**), 블록 단위 (**erase**) 어드레싱 (디스크처럼 동작)
    - 주로 데이터 저장용으로 사용됨
    - Write 성능, 집적도, 가격, 전력 소비, 수명 (**erase cycles**) 등에서 유리



# Flash Memory Types

Type	NOR	NAND	DINOR	AND
Product	Intel 28F128J3A-150 28F320D18B110 AMD Am29lv641DU Am29DL322D	Samsung K9F5608UOM Toshiba TH58512FTI AMD Am30LV0064D	Mitsubishi M5M29GB/ T160BVP-80	Hitachi HN29W25611
Density	Low	High	Low	High
Erase Block size	Large (64 – 128KB)	Large (16 – 128KB)	Large (32 – 64KB)	Small (2KB)
Page size	1, 2, 8, 32B	512B, 2KB	1, 2, 256B	2KB
Capacity	Low	High	Low	High

# NAND 플래시 메모리 세계 시장 전망 (1)

- NAND 플래시 메모리 시장 전망(From Gartner Group)
  - 2006년 매출액 기준 2005년 대비 37.3% 성장한 147억 4,650만 달러를 기록할 것으로 예상
  - 반면, NOR 플래시 메모리 시장은 매출액 기준 전년대비 4.8% 감소한 67억 4,060만 달러 수준으로 전망
  - NAND 플래시 메모리의 Mb 기준 출하량은 2008년까지 연평균 성장률 146.5%를 기록할 것으로 예상되고 있으며, NAND 플래시의 용량 증가가 매우 빠르게 진행되고 있음을 시사

<표 1> 플래시 메모리 세계 시장 현황 및 전망

구분		2003	2004	2005	2006	2007	2008	2004~2008 CAGR
매출액 (백만 달러)	NAND	4,131.0	7,007.0	10,738.2	14,746.5	14,768.1	17,792.2	26.2%
	NOR	6,583.0	8,429.0	7,083.2	6,740.6	5,877.1	5,663.1	-9.5%
출하량 (백만 개)	NAND	449.3	713.9	1,257.6	1,691.1	1,878.7	2,013.3	29.6%
	NOR	2,435.5	2,872.7	2,834.4	2,636.4	2,180.7	1,757.8	-11.6%
출하량 (백만 Mb)	NAND	20,363.5	65,606.4	231,117.1	627,329.2	<b>1,941,241.7</b>	<b>2,421,866.8</b>	146.5%
	NOR	10,130.9	16,523.8	24,534.2	34,401.5	48,332.4	62,604.8	39.5%
평균단가 (달러)	NAND	9.19	9.81	8.54	8.72	7.86	8.84	-2.6%
	NOR	2.70	2.93	2.50	2.56	2.70	3.22	2.4%

C4EX Gartner Database, 2006.11.

# NAND 플래시 메모리 세계 시장 전망 (2)

- 플래시 메모리 시장은 대용량 데이터 저장을 위한 **NAND** 형과 프로그램 코드용인 **NOR** 형이 주류임
- **NAND** 형 플래시 메모리 관련 기술은 미국 혹은 유럽 지역 보다는 한국, 일본, 대만 등 동북아에서 기술 주도
- 플래시 메모리 분야는, 지금까지 컴퓨터 분야에서 **CPU** 성능과 **DRAM** 집적도 분야에서 통용되어 온 “무어의 법칙”과 비견될, 플래시 메모리 분야의 “황의 법칙”에 따라, 매년 두 배씩 집적도를 높여가고 있음
- “황의 법칙”에 따르면, 2007년경에 64G급의 플래시 메모리가 출현



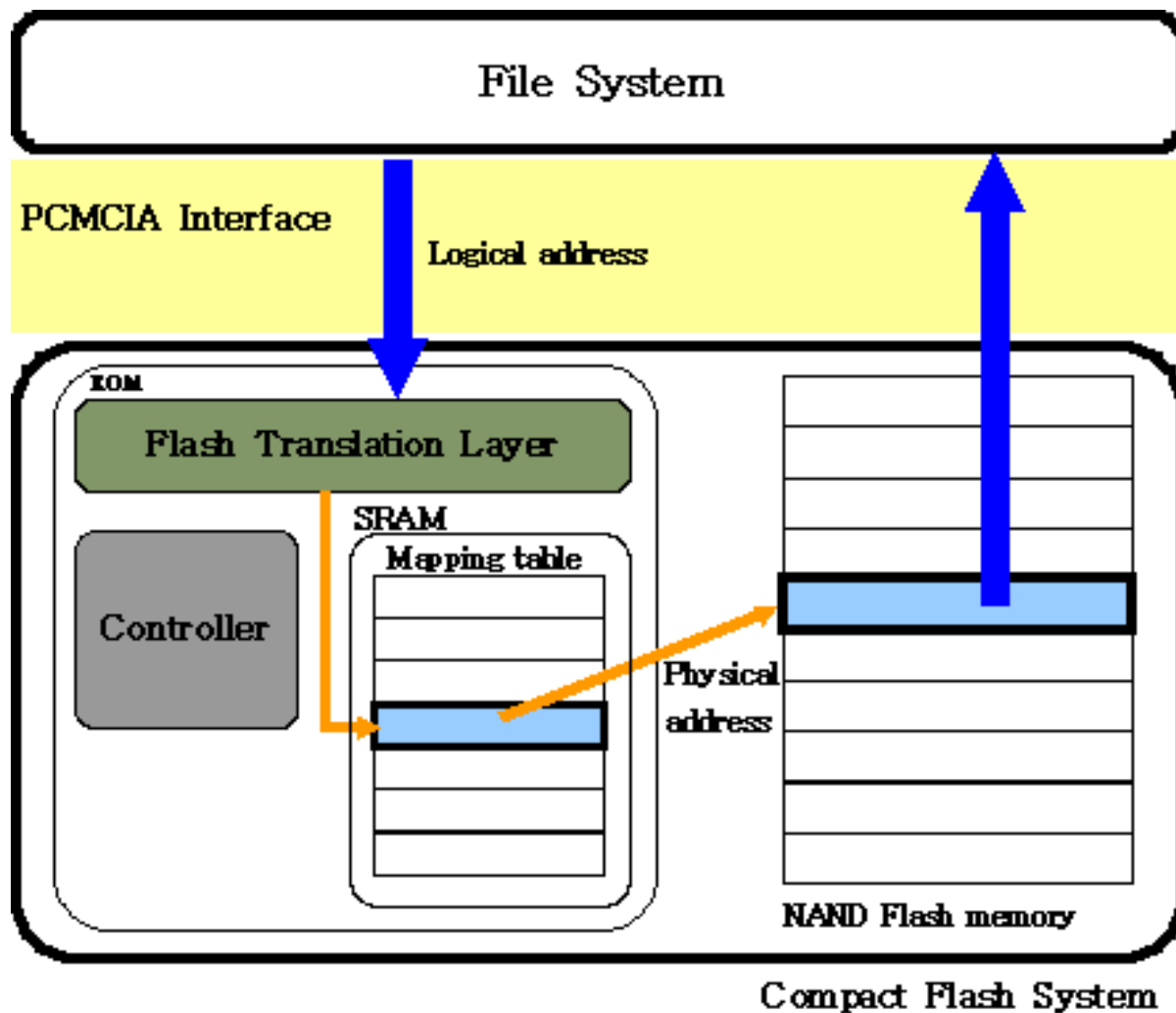
플래시 메모리 용량증가  
("황의 법칙")



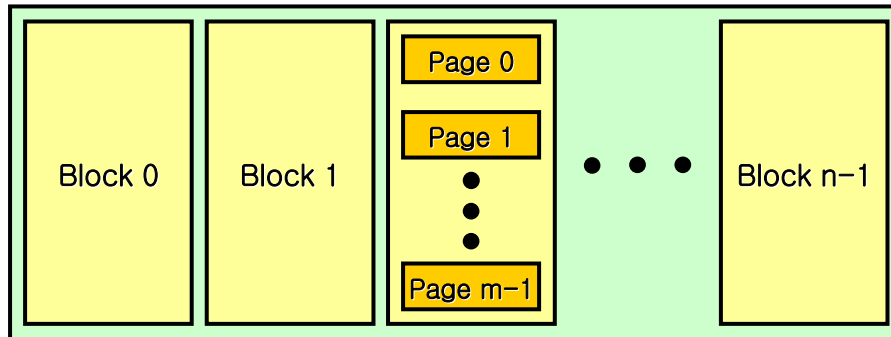
# NAND 플래시 메모리 구조 및 기본 Operation

---

# NAND 플래시 메모리의 구조 (1)



# NAND 플래시 메모리의 구조 (2)



• 소블록 페이지

512 16

• 대블록 페이지

2048

64

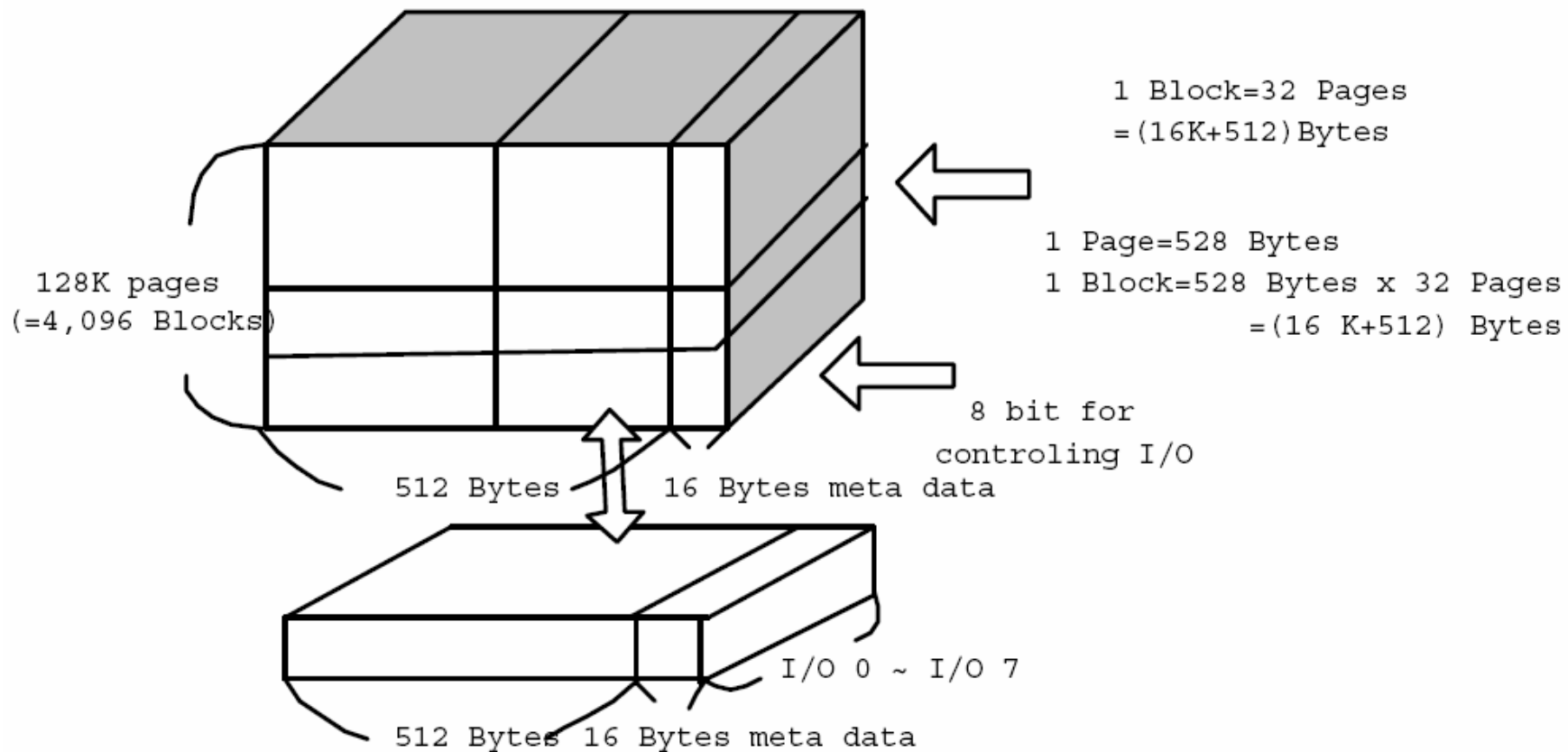
Main Area

Spare Area

## • 기본구조

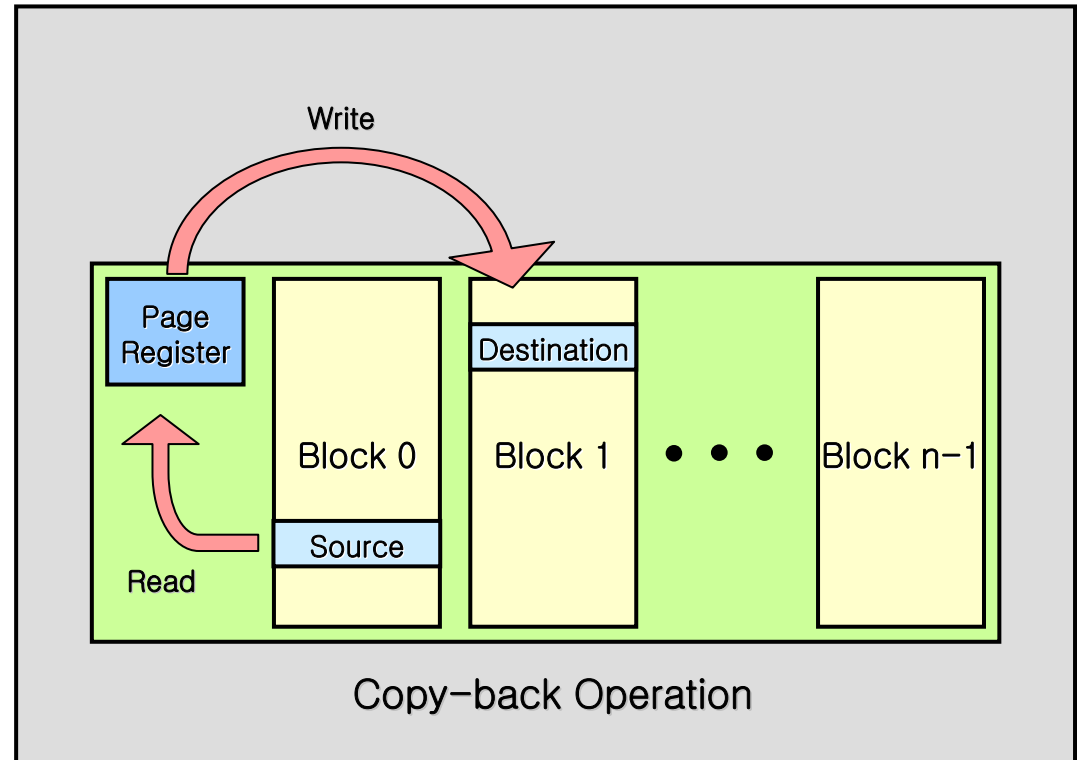
- 소블록 플래시 메모리
  - 16KB 크기의 블록들로 구성
  - 각 블록은 32개의 페이지로 구성
  - 각 페이지는 (512 + 16)byte로 구성
- 대블록 플래시 메모리
  - 128KB 크기의 블록들로 구성
  - 각 블록은 64개의 페이지로 구성
  - 각 페이지는 (2048 + 64) byte로 구성

# NAND 플래시 메모리의 구조 (3)



# NAND 플래시 메모리의 기본 동작

- Read page
  - (chip #, block #, page #)
  - ~20 us
- Write (program) page
  - (chip #, block #, page #)
  - ~200 us
- Erase block
  - (chip #, block #)
  - ~2 ms
- Copy-back page
  - (chip #, target block #, target page #, source block #, source page #)
  - ~200 us





# NAND 플래시 메모리 사용 시 주의 사항

- Bit-Flip Error Handling
  - 일반적으로 1-bit 또는 2-bit ECC (Error Correction Code)를 사용하여 처리함
  - MLC (Multi-Level Cell) NAND의 경우 4-bit 이상의 ECC 사용을 권고
- Bad Block Handling
  - Bad block은 더 이상 사용할 수 없는 블록임
  - Initial bad block 및 run-time bad block 모두 처리해 주어야 함
  - Initial bad block 정보는 spare 영역에 기록되어 있음
  - Write (Program) / erase 동작 결과 에러 리턴시 run-time bad block으로 간주
- Wear-Leveling
  - SLC NAND의 경우 각 블록 당 100,000번 이상의 erase가 가능함
  - MLC NAND의 경우 각 블록 당 10,000번 이상의 erase가 가능함
  - 일부 블록이 집중적으로 wear-out 될 경우 플래시 메모리의 수명이 단축됨
  - 각 블록이 골고루 erase 될 수 있도록 해야 함 (wear-leveling)

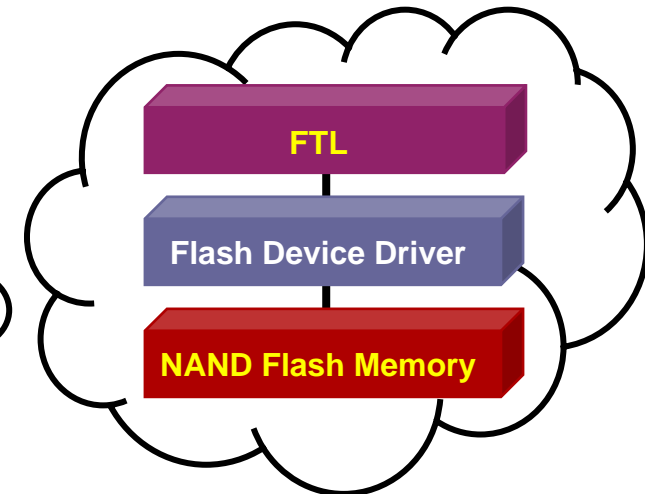


# NAND 플래시 메모리 관리 S/W

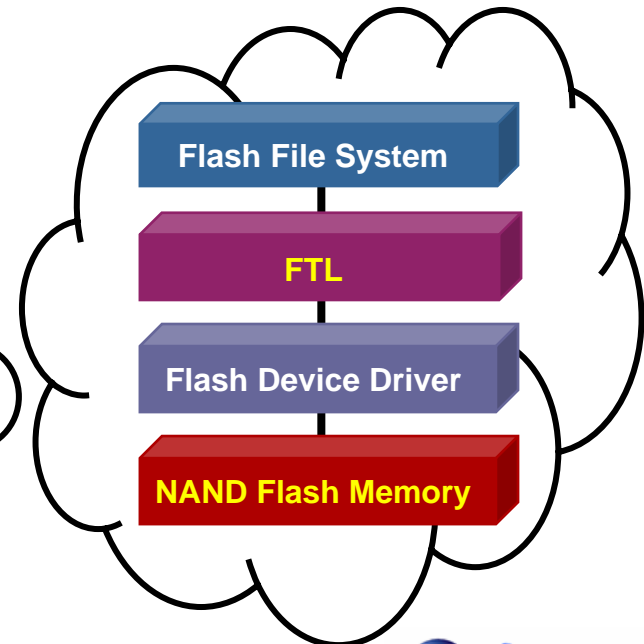
---

# 플래시 메모리 관리 S/W

- 외장형 메모리 카드 사용시
  - 플래시 메모리를 HDD처럼 변환해주는 FTL이 필요

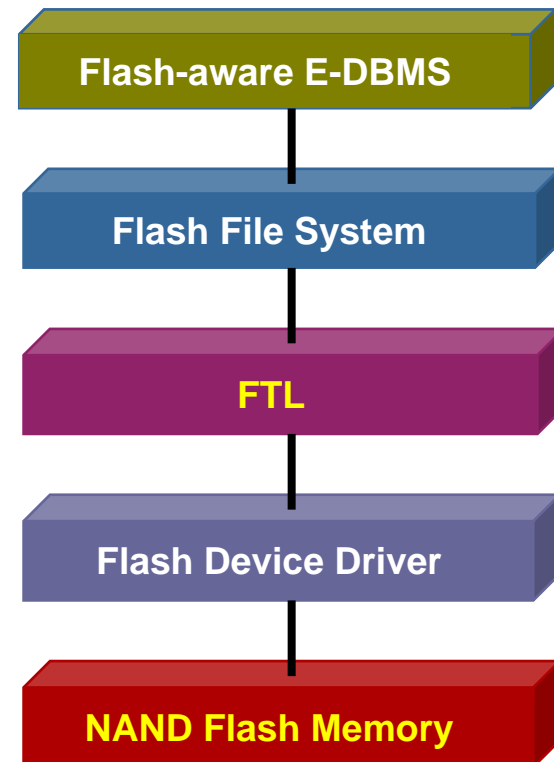


- 내장형 플래시 메모리 칩 사용시
  - 플래시 메모리를 HDD처럼 변환해주는 FTL이 필요
  - 파일 서비스를 제공해 줄 수 있는 파일시스템이 필요

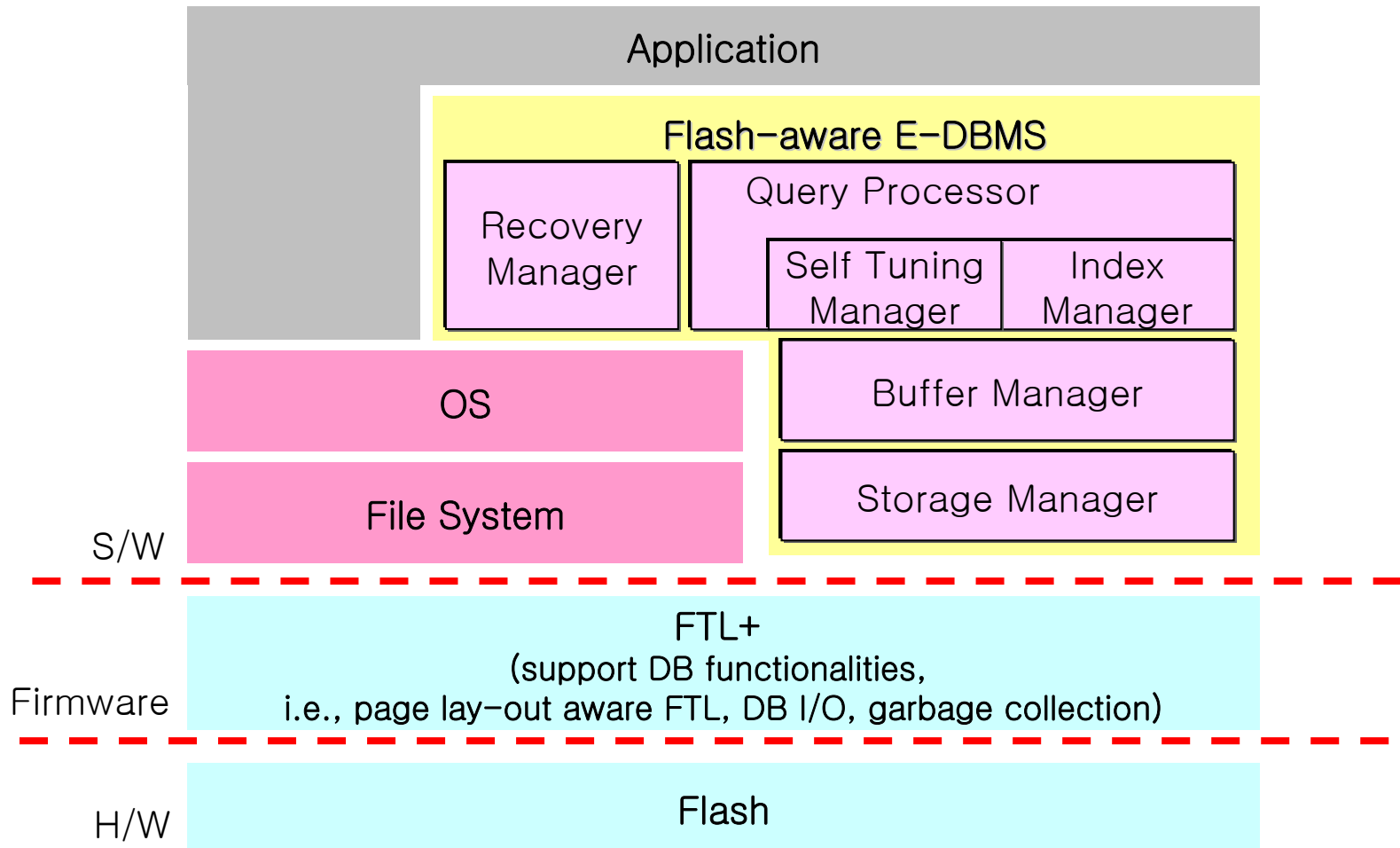


# 플래시 메모리 관리 S/W의 구성 (1)

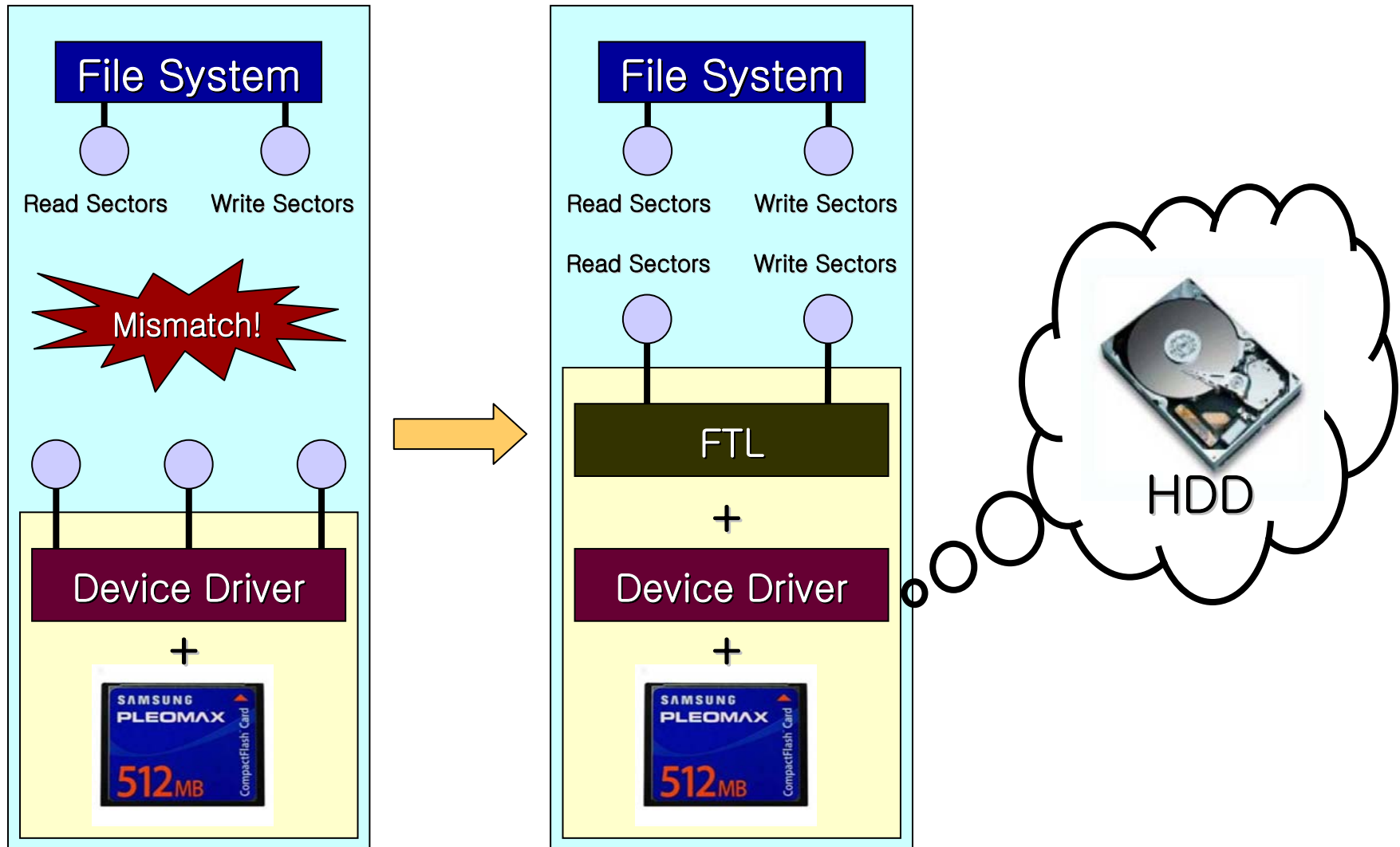
- Flash-aware Embedded DBMS
  - Flash 메모리의 특성을 고려한 임베디드 데이터베이스 시스템
- Flash File System
  - 일반적인 파일 시스템 사용 가능
  - 안정성 및 성능 향상을 위한 플래시 메모리의 특징을 고려하는 것이 바람직함
- FTL (Flash Translation Layer)
  - NAND 플래시 메모리가 하드디스크와 같은 일반적인 블록 장치처럼 보이도록 변환해 줌
  - Wear-leveling 기능을 수행함
- Flash Device Driver
  - 플래시 메모리에 대한 물리적인 접근
  - Bad block 처리
  - ECC 처리



# 플래시 메모리 관리 S/W의 구성 (2)



# FTL (1)



# FTL (2)

- FTL의 기본 기능

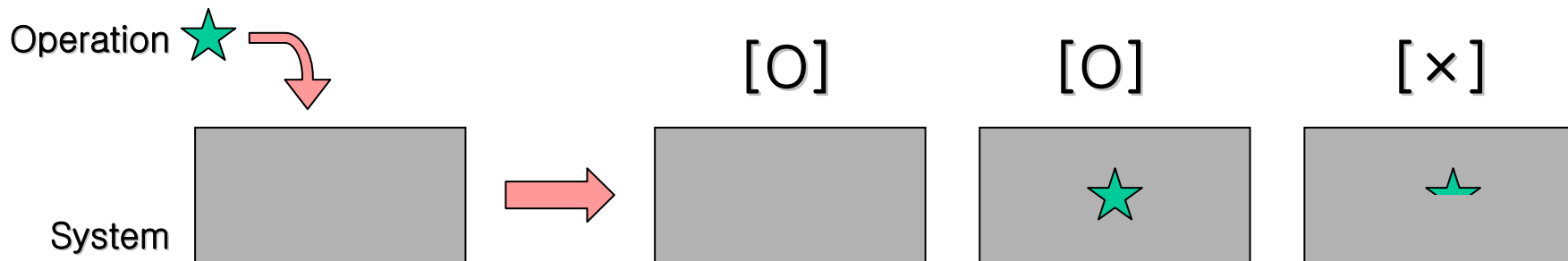
- 주소 변환 : 논리 주소를 플래시 메모리에 대한 물리 주소로 변환

(섹터번호) = (논리블록번호, 논리페이지번호) ➡ (물리블록번호, 물리페이지번호)

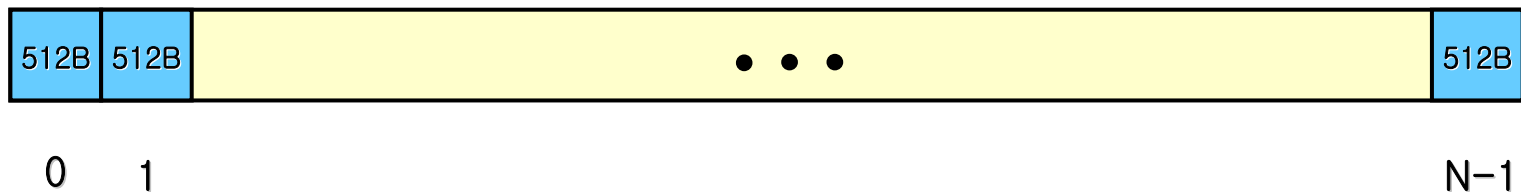
- sector\_read, sector\_write 에 대해 atomicity 보장

- Atomicity 란?

- 어떤 operation 이 시스템에 가해졌을 때 해당 시스템의 상태가 다음과 같이 두 가지로만 정의될 수 있으면, **atomic operation** 이라 일컬어짐
  - Operation 이 가해지기 이전의 시스템 상태
  - Operation 이 완전히 종료된 이후의 시스템 상태



# Logical interface for a disk drive



- Operations
  1. Identify drive(): returns N
  2. Read sectors(start sector #, # of sectors)
  3. Write sectors(start sector #, # of sectors)

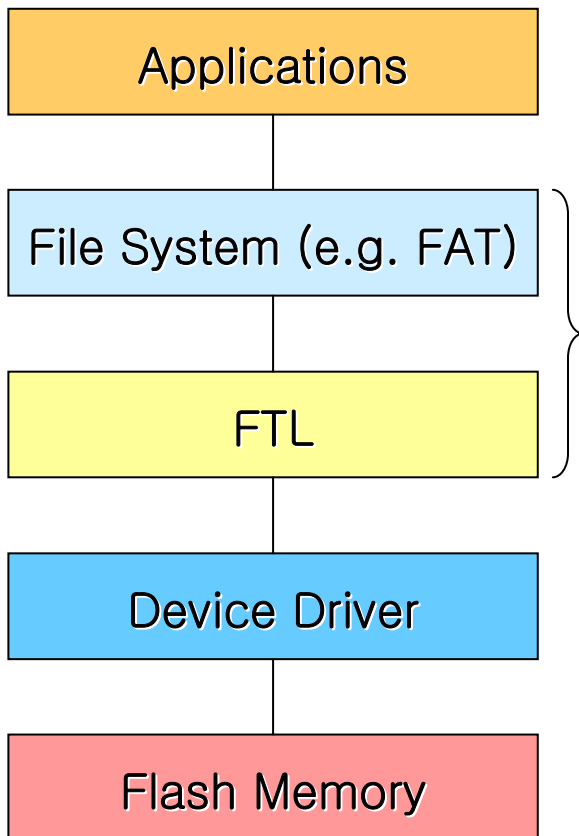


# Flash File System (1)

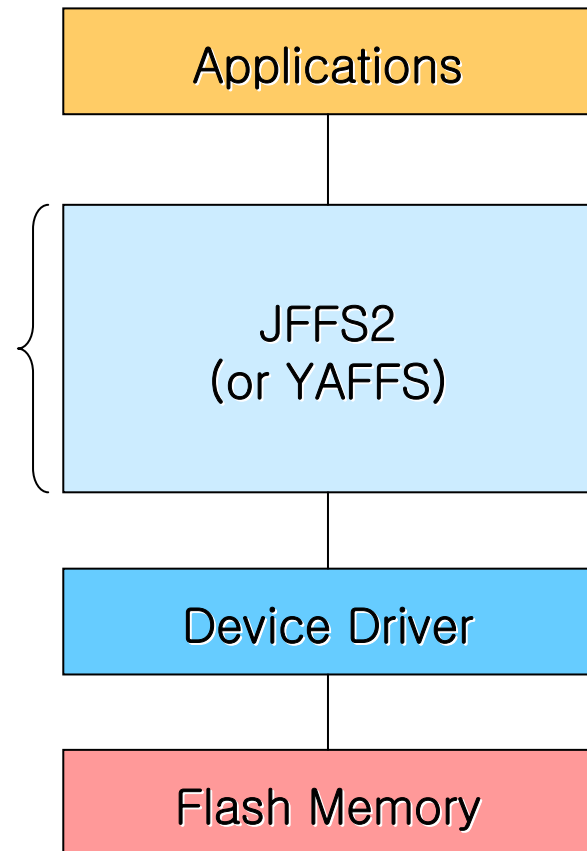
- 일반적인 파일 시스템을 사용하는 경우
  - FTL (Flash Translation Layer)이 필요
  - FAT 파일 시스템이 많이 사용됨
  - 모바일 환경을 고려하여 신뢰성을 보완할 필요가 있음 (Journaling 또는 transaction 기반 수행 메커니즘)
- 플래시 메모리 전용 파일 시스템을 사용하는 경우
  - LFS (Log-Structured File System) 형태의 파일 시스템 (JFFS2, YAFFS)
  - 신뢰성이 높으나 아래와 같은 단점이 있음
    - 파일 시스템 부팅에 적지 않은 시간이 소요됨 (수 초 ~ 수십 초)
    - 파일 시스템 트리 구조를 RAM 에 두어야 하기 때문에 RAM 소요량이 큼
    - Garbage collection 오버헤드가 큼 (write 성능 저하)

# Flash File System (2)

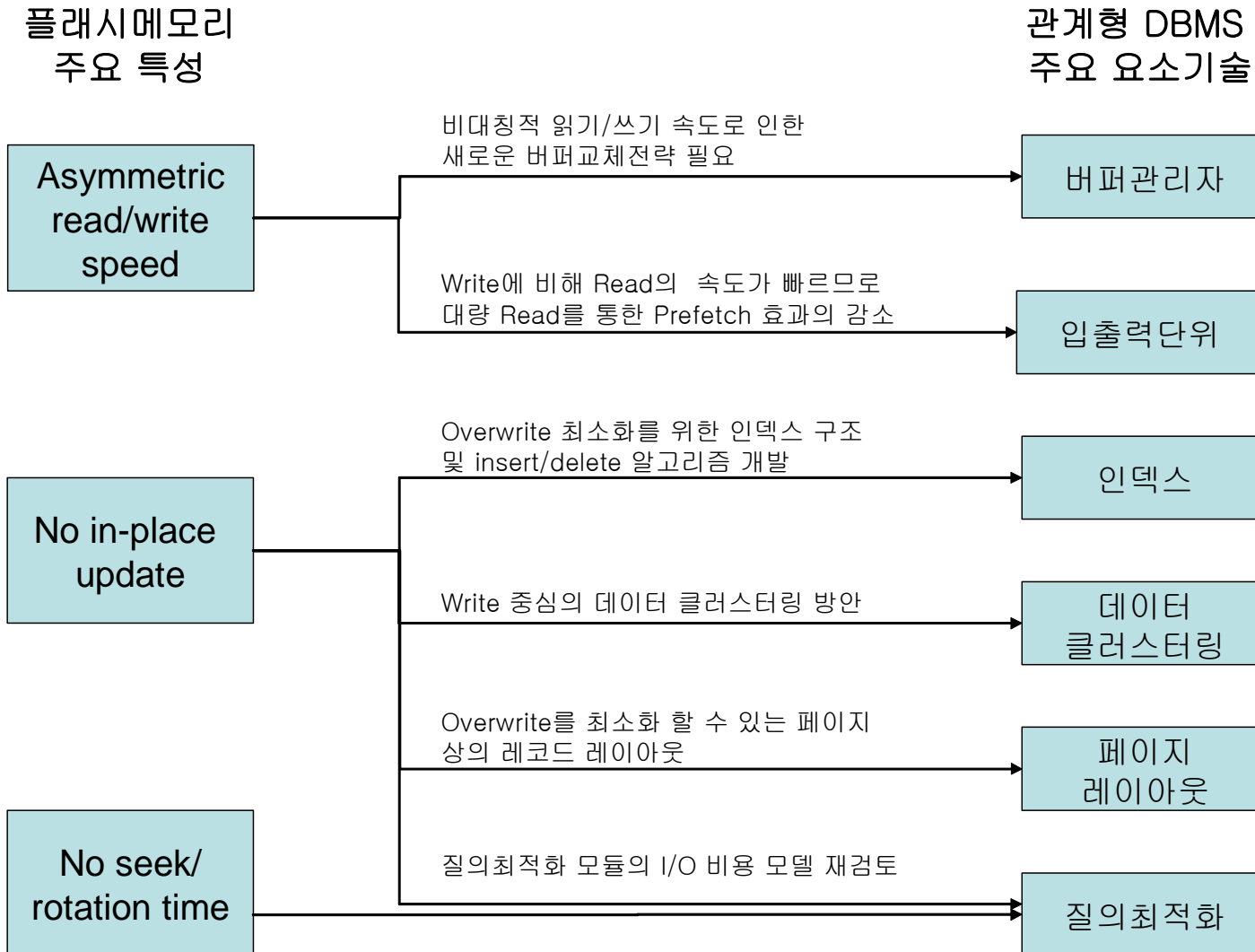
일반 파일 시스템을 사용하는 경우



전용 파일 시스템을 사용하는 경우



# Flash-aware E-DBMS (1)



# Flash-aware E-DBMS (2)

- 버퍼교체전략(buffer replacement algorithm)
  - 플래시 메모리는 메모리 계층구조상(memory hierarchy)에서 역사상 처음으로 읽기와 쓰기 연산의 속도가 비 대칭적(asymmetric)인 기억 장치임
  - 따라서, 기존에 하드디스크와 같이 읽기와 쓰기 연산의 속도가 동일한 기억장치 상에서 유효한 LRU 등과 같은 전통적인 버퍼교체 전략이 부적합할 수 있음
  - 즉, "Hit Ratio"는 읽기와 쓰기의 속도가 동일할 때 의미가 있는 기준이며, 플래시 메모리의 경우 쓰기의 횟수나 그에 따른 erase 횟수의 감소를 위한 알고리즘이 필요
  - 응답시간(response time)이 아니라, 작업량("Throughput") 기준으로 일정정도의 읽기 연산의 회수를 감수하면서도 전체 연산의 시간을 단축하는 방향으로 버퍼 교체 전략 고안이 중요함

# Flash-aware E-DBMS (3)

- 입출력 단위

- 기존 하드디스크 기반의 소프트웨어(파일시스템, 데이터베이스)에서는 블록이 입출력의 기본단위임
- 컴퓨팅 파워와 자원의 증가에 따라, 입출력 단위인 블록의 크기도 대형화되었음(OS: 512B, DB: 16K가 보편적임)
- 하지만, 이는 덮어쓰기가 가능한 하드디스크에는 문제가 없지만, 플래시 메모리의 경우 부분적인 쓰기 연산이 과도한 소거(erase) 연산을 유발하기 때문에 비효율적임
- 플래시 메모리의 경우, 블록의 크기와 FTL(Flash Translation Layer) 알고리즘의 특성에 따라 성능 편차가 아주 크게 됨

# Flash-aware E-DBMS (4)

- 인덱스 구조(index structure)
  - 인덱스는 특히 삽입/삭제 연산에 의해 random write와 노드 분할 및 병합이 빈번하게 발생하는 자료구조임
  - 이러한 이유로, 기존의 B+ 트리 인덱스는 플래시 메모리 상에서 하드 디스크에 비해 현저한 성능 저하가 예상됨
  - 플래시 메모리를 고려한 새로운 색인 구조 및 색인 연산이 필요함
- 데이터 클러스터링(data clustering)
  - 전통적인 데이터 클러스터링 기법은 자주 같이 읽혀지는 데이터를 같은 블록에 위치시킴으로 read시, prefetch 효과로 인해 성능 향상시킴
  - 플래시 메모리의 경우, 성능의 bottleneck이 쓰기와 소거 연산이므로 데이터 클러스터링의 경우도 쓰기 연산과 소거 연산을 최소화하면서 read의 연산도 최적으로 달성할 수 있는 방안이 필요
  - 따라서, 읽기 중심의 클러스터링에서 쓰기 중심의 클러스터링으로 전환 필요함

# Flash-aware E-DBMS (4)

- 페이지 레이아웃(page layout)
  - 전통적인 데이터베이스에서는 한 페이지에 대해 튜플 단위의 레이아웃(row-oriented layout)을 지원하고 있는데, 이는 주로 기존의 OLTP에서 소수의 튜플에 대한 update시의 쓰기 속도의 최적화를 위한 방안임
  - 최근에 데이터웨어하우스와 같은 응용분야를 위해서는 칼럼단위의 읽기를 위한 칼럼 중심의 레이아웃(column-oriented layout)을 지원하는 방안 제시[18]
  - 플래시 메모리의 경우, 하나의 논리적인 데이터베이스 블록이 여러 개의 물리적인 플래시 메모리 섹터로 매핑됨
  - 따라서, 이 경우에 읽기와 쓰기에 대해 모두 좋은 성능을 보일 수 있는 페이지 레이아웃 방안에 대한 재검토 필요

# Flash-aware E-DBMS (5)

- 질의 최적화(query optimization)
  - 플래시 메모리의 두 가지 특징, 즉 **asymmetric read/write time**, **no seek/latency time**은 기존 질의 최적화 모듈의 비용 모델과 최적화에 모두에 많은 영향을 미침
  - 비용 모델의 예를 들면, 하드디스크의 **write**의 경우 덮어쓰기가 가능하기 때문에 **write**에 대한 비용 모델이 단순하지만, 플래시 메모리의 경우 **write**의 비용 모델을 단순 쓰기시간과 소거 시간을 고려한 수식 모델이 필요
  - 또 다른 비용 모델의 예로는, 기존 관계형 **DBMS** 최적화 모듈에서 비용 모델의 가장 중요한 이슈중의 하나는 **full table scan**과 **index scan**의 비용 산정에 관한 것으로, 인덱스의 **clustering factor** 개념에 기반한 복잡한 비용 모델을 채용. 플래시 메모리의 경우, **seek/latency time**이 없기 때문에 비용 모델의 단순화가 필요