



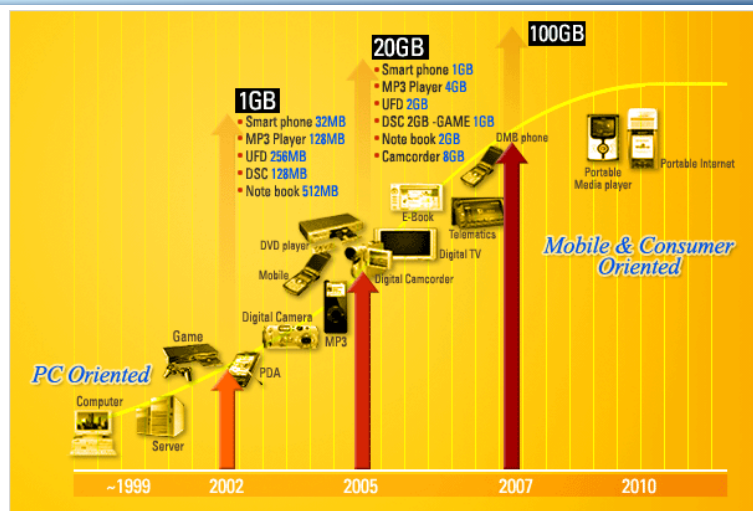
Issues in an Embedded DBMS for NAND Flash Memory

Dong-Ho Lee

Assistant Professor
Knowledge and Data Engineering Lab.
Dept. of Computer Science and Engineering
Hanyang University



Portable Storage Applications



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



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 성장을 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

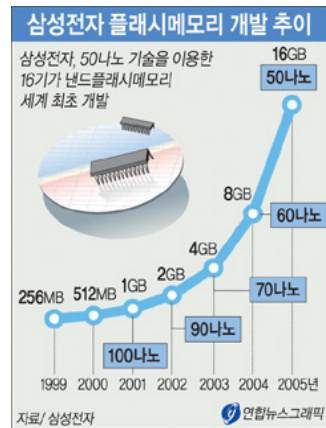
구분		2003	2004	2005	2006	2007	2008	2004~2008 CAGR
매출액 (억 달러)	NAND	41310	73070	117312	147465	147311	117312	26.2%
	NOR	65810	64210	70812	67416	58721	58811	-3.5%
출하량 (억 개)	NAND	4413	7131	12576	18511	18717	20113	29.6%
	NOR	24355	23727	28344	26314	21017	17518	-1.6%
출하량 (억 개 Mb)	NAND	213615	653064	2311171	6223212	1,011,2117	2,411,3558	146.5%
	NOR	111313	111210	213112	114115	113314	112018	31.3%
평균당가 (달러)	NAND	919	101	154	172	708	114	2.6%
	NOR	270	291	250	256	270	322	2.4%

출처: Gartner Group, 2006년 11월



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 세계 시장 전망 (3)



[전자신문, 2006/03/22]

- 32GB급 플래시메모리 개발, 노트북PC 시장 공략
- 2006년 NAND 플래시 메모리가 PC시장에 진입하는 원년
- 2008년 세계 노트북용 HDD 시장의 30%를 플래시가 대체
- 32GB 개발은 앞으로 모바일 PC 이외에도 모든 디지털 제품에 NAND 플래시가 본격 확대되는 신호탄

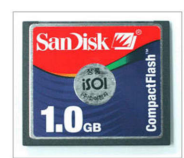


HDD vs. Flash Memory

- HDD 기반 저장장치
 - 자기 디스크에 데이터를 기록
 - 가격 및 성능에서 유리
 - 내구성/전력소비/부팅시간에서 불리



- NAND 플래시 메모리 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: 각종 플래시 메모리 Card, USB Drive



Why flash memory?

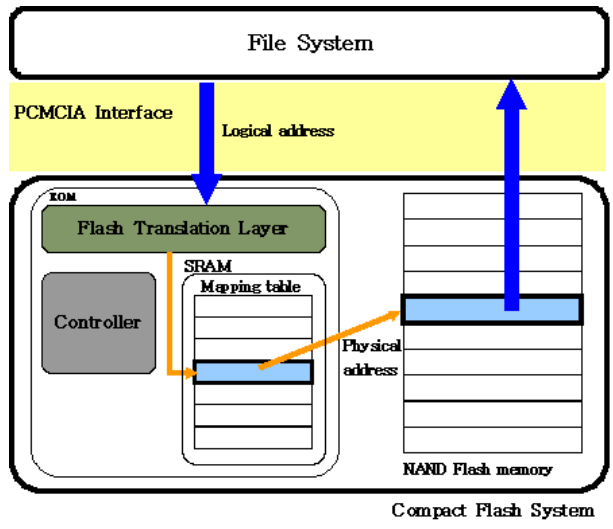
- 빠른 데이터 접근
- 저전력
- 강한 내구성
- 작고 가벼움



플래시 메모리의 개요

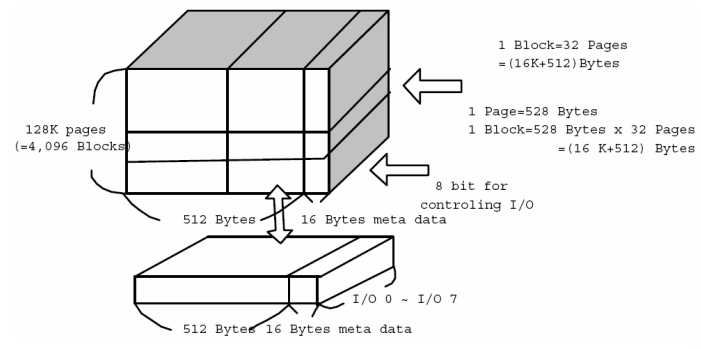
- 플래시 메모리
 - 전원 공급이 중단된 뒤에도 데이터가 보존되는 메모리 반도체
 - 한번 기록된 영역은 **in-place update**가 되지 않기 때문에, **erase** 연산을 한 뒤에 **write** 해야 하는 특성이 있음
- 플래시 메모리의 종류
 - NOR 플래시 메모리
 - 바이트 단위 어드레싱 가능
 - 코드 실행용
 - NAND 플래시 메모리
 - 페이지 단위의 read/write 연산,
 - 블록 단위의 erase 연산
 - 데이터 저장용
 - 가격, 전력 소비, 수명 등에서 NOR보다 유리
 - 이외에도 DINOR, AND 타입 플래시 메모리도 존재

NAND 플래시 메모리의 구조 (1)



NAND 플래시 메모리의 구조 (2)

- 기본구조
 - 소블록 플래시 메모리
 - 16KB 크기의 블록들로 구성, 각 블록은 32개의 페이지로 구성
 - 각 페이지는 512 byte의 main area와 16 byte의 spare area로 구성
 - 대블록 플래시 메모리
 - 128KB 크기의 블록들로 구성, 각 블록은 64개의 페이지로 구성
 - 각 페이지는 2048 byte의 main area와 64 byte의 spare area로 구성

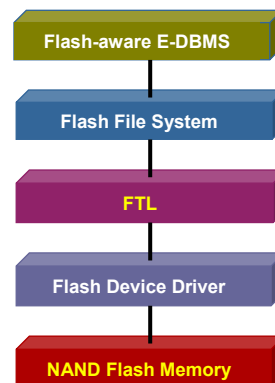


NAND 플래시 메모리의 주요 특징

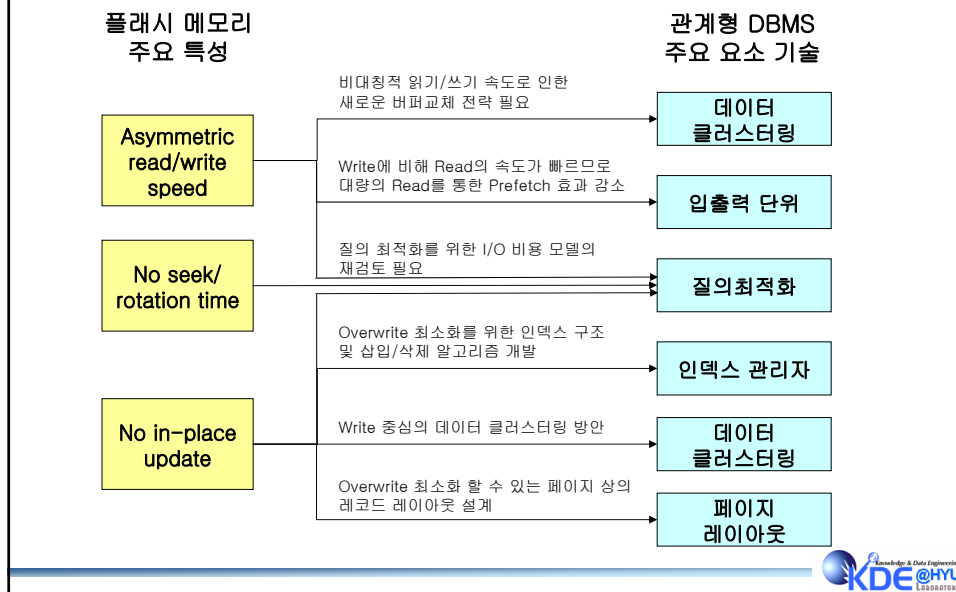
- No in-place Update
 - 한번 Write한 영역에 Overwrite가 불가능(Erase 연산을 한 후에 Write 가능)
- Read/Write 연산과 Erase 연산 단위
 - Read/Write 단위 : sector (or page)
 - Erase 단위 : block
- Read/Write/Erase 속도 차이
 - Read page : ~ 20us
 - Write page : ~ 200us
 - Erase block : ~ 2ms

NAND 플래시 메모리 관리 S/W의 구성

- 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 (논리적 주소를 물리적 주소로 사상)
 - Wear-leveling 기능을 수행함
- Flash Device Driver
 - 플래시 메모리에 대한 물리적인 접근
 - Bad block 처리
 - ECC 처리



Flash-aware E-DBMS (1)



Flash-aware E-DBMS (2)

- 버퍼교체전략(buffer replacement algorithm)
 - 플래시 메모리는 메모리 계층 구조상(memory hierarchy)에서 역사상 처음으로 읽기와 쓰기 연산의 속도가 비 대칭적(asymmetric)인 기억 장치임
 - 기존의 하드디스크와 같이 읽기와 쓰기 연산의 속도가 동일한 저장장치 상에서 효과적인 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
 - 일정 정도의 읽기 연산의 횟수를 감수하면서도 전체 연산의 시간을 단축하는 방향으로 버퍼 교체 전략 고안이 중요함

Flash-aware E-DBMS (3)

- 입출력 단위
 - 기존 하드디스크 기반의 소프트웨어(화일시스템, 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
 - 따라서, 기존의 B+ 트리 인덱스는 플래시 메모리 상에서 하드디스크에 비해 현저한 성능 저하가 예상됨
 - 또한, 인덱스 노드의 크기를 어떻게 설계하느냐에 따라서도 write와 erase 연산의 편차가 심하게 날 수 있음
 - 플래시 메모리를 고려한 새로운 색인 구조 및 write와 erase 연산을 최소화할 수 있는 삽입/삭제 알고리즘이 필요함

Flash-aware E-DBMS (5)

- 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 (6)

- 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
 - 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 (7)

- 질의 최적화(query optimization)
 - 플래시 메모리의 특징, 즉 **read/write** 단위와 **erase** 단위의 불일치, **asymmetric read/write time**, 그리고 **no seek/latency time**은 기존 질의 최적화 모듈의 IO 비용 모델에 대한 재검토를 필요로 함
 - 하드디스크의 **write**의 경우 덮어쓰기가 가능하기 때문에 **write**에 대한 비용 모델이 단순하지만, 플래시 메모리의 경우 **write**의 비용 모델은 단순 쓰기시간과 소거-쓰기 시간을 고려한 수식 모델이 필요
 - 기존 관계형 **DBMS** 질의 최적화 모듈에서 비용 모델의 가장 중요한 이슈중의 하나는 **full table scan**과 **index scan**의 비용 산정에 관한 것으로, 인덱스의 **clustering factor** 개념에 기반 둔 복잡한 비용 모델을 채용
 - 플래시 메모리의 경우, **seek/latency time**이 없기 때문에 이를 고려한 비용 모델이 필요