

# 고속 SSD와 대용량 HDD를 위한 하이브리드 저장 시스템에 대한 연구

변시우\* 허문행\*

\*안양대학교 디지털미디어학과

e-mail:swbyun@anyang.ac.kr

## A Study of Hybrid Storage System for High-Speed SSD and Large-Volume HDD

Siwoo Byun\*, Moonhang Hur\*

\*Dept. of Digital Media, Anyang University

### 요 약

최근 산업체나 연구소에서 사용되는 중대형 데이터 서버에서 발생하는 가장 큰 오버헤드(bottleneck)중의 하나가 바로 저장 시스템이라 할 수 있다. 대표적인 저장 시스템으로서 전통적인 하드디스크 드라이브(HDD)와 더불어 최근 상용화된 Solid State Drive(SSD)가 주목 받기 시작하였다. 본 연구의 목표는 이러한 대용량 HDD 및 고속 SSD에 대하여, 저장 장치로서의 약점을 상호 보완하면서, 장점을 강화하여, 최종적으로 입출력 성능을 극대화할 수 있는 새로운 융합형 데이터 저장 시스템(Hybrid-RAID) 개발이다. Hybrid 저장 시스템 개발을 위한 최근의 기술 연구 및 기반 기술의 설계 및 구현에 대하여 논하며, hybrid-clustering 기술을 제안한다.

### 1. 서론

하드 디스크(HDD: Hard Disk Drive)는 지난 30년 전부터 대부분의 데이터 서버에서 사용되는 절대적인 저장 미디어이다. 그러나 최근 중대형 데이터 서버에서의 저장 시스템으로서 최근 상용화된 Solid State Drive(SSD)가 주목 받기 시작하였다.

본 연구는 이러한 대용량 HDD 및 고속 SSD에 대하여, 저장 장치로서의 약점을 상호 보완하면서, 장점을 강화하여, 최종적으로 입출력 성능을 극대화할 수 있는 새로운 융합형 데이터 저장 시스템(Hybrid-RAID) 개발이다. Hybrid 저장 시스템 개발을 위한 최근의 기술 연구 및 기반 기술의 설계 및 구현에 대하여 논하며, hybrid-clustering 기술을 제안한다.

### 2. 관련 연구

하드 디스크[6]에 대한 대안으로 최근에 주목받고 있는 저장장치가 바로 플래시 메모리[4,7]를 장착한 SSD이다. 결론적으로 용량과 가격측면에서는 하드 디스크가 매우 우세하고, 접근 성능 및 저전력 측면에서는 SSD가 매우 우세하다.[2]

이미 1990년대에 등장한 SSD는 하드 디스크에 비

하여 많은 장점에도 불구하고, 그동안 가격 경쟁력이 매우 부족하여, 특수한 분야(군사, 항공, 우주, 선박 등)에서만 주로 사용되었다. 그러나, 최근에는 플래시 메모리의 급속한 가격 하락에 힘입어, 서버, 휴대용 노트북 등 고급 제품을 중심으로 기업 및 일반 소비자 시장으로 확대되고 있다.[5,8]



\*제조사 : SAMSUNG, 2008, HP Server에 내장됨  
[그림 1] SATA I/F로 연결되는 64GB SLC-type의 SSD

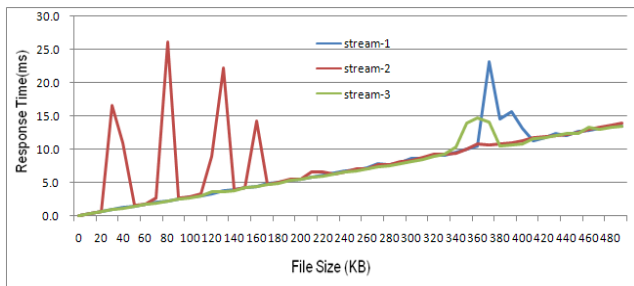
### 3. SSD의 불안정한 I/O 특성의 문제

SSD의 문제점은 실제 사용자가 작업하다가, 저장 연산이 집중되는 경우, SSD가 평소속도보다 상당히 느린 속도(20-30배)로 반응한다는 문제이다. 물론, 가끔씩 발생하지만, 하드 디스크에 비해서는 상당히 빈번히 발생하고, 홀딩 시간도 하드 디스크의 보다 훨씬 더 길다.

이 현상은 물론 읽기 연산에서는 거의 발생하지

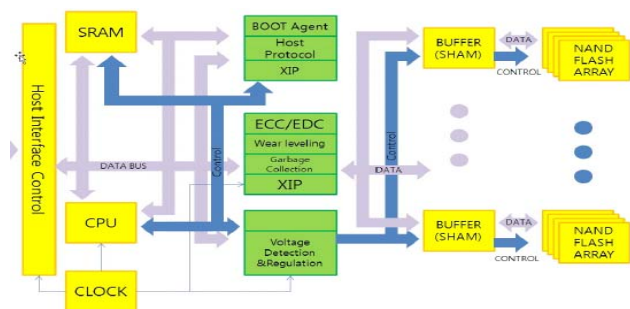
않는다. 하지만, 소프트웨어를 설치하거나, 특히, 큰 콘텐츠를 복사 받는 경우 윈도우 운영체제가 심한 경우 10초 이상씩 정지(흔히 프리징[freezing]이라고 함)가 되는데, 이러한 체험 사례는 인터넷에서도 심각하게 보고되어 있다.

본 연구에서도, 프리징 현상을 확인하기 위하여, 비교 실험하였다. 이러한 불안정한 응답성은, 아래의 그래프에서도 확인 가능하다. 실험환경은 10KB~500KB 까지 10KB 단위로 50개의 저장 스트림을 발생시켜서 응답시간(Response Time)을 측정한 결과이다. 참고로 일반 시스템 환경에서는 95% 이상이 500KB 미만의 데이터 파일이다. 실험결과, 아래 그래프에서 다수의 문제 피크(Peak)가 발생하는데, 스트림-2의 경우처럼 한번 홀딩이 되면 연이어서 계속(4차례) 발생한다는 것이 문제점(프리징발생)이다.



[그림 2] 자체 테스트 결과 [SSD Write Stream. 2009.1월]

그 이유는, 쓰기 연산의 집중으로 인하여, 고가의 SRAM 버퍼가 포화되어, 더 이상 지연시키지 못하고, Buffer에서 밀려 있던 플래시 메모리의 소거 작업과 수반되는 마모 평준화(웨어레벨링) 연산과 느린 CPU, 카피 오버헤드 등의 복합적인 원인 때문이다. (아래 그림에서 관련 SRAM, Buffer, CPU, 부품 참조). 이 경우, 사용자는 좋은 성능을 보인 SSD에 대한 기대가 한 순간에 무너지는 체험을 하게 된다. 마찬가지로, SSD에 데이터가 별로 없는 제품구입 초기에는 없다가, 데이터가 점점 더 쌓여갈수록 프리징현상이 빈번해 진다.



[그림 3] SSD의 주요 부품과 구성도

또한, 사용자는 잘 체감하지 못하는 짧은 수명의 문제점도 발생한다(수명열화 현상). 사용자가 특정 물품을 구매할 때, 제품 설명서를 꼼꼼히 읽어 보는 경우는 많지 않다. 특히, 저장 장치를 구매하면, 대부분의 사용자는 중요한 항목인 저장수명을 의심하지 않는다. 하드 디스크의 경우 특별한 외부 충격이 없는 한 거의 반영구적이며, 기계적인 부분이 훼손되더라도, 데이터가 저장된 디스크 원판은 다시 복구되는 경우가 많았다. 반면에 SSD는 큰 충격을 주더라도 쉽게 데이터가 파괴되지 않는 반면, 백만번 정도의 저장 횟수를 초과하면, 읽기만 가능하고, 더 이상 저장은 불가능하다. (원인:NAND Flash Cell의 악화, “자연열화현상”, 밴더들은 최대 10년 이내로 보증함). 이 수명 제한 때문에, 대용량의 데이터를 빈번히 사용하는 데이터센터라면 그 불안감이 상당할 것이다. 즉, 사용자는 이러한 중요한 보증수명의 문제를 평소에는 전혀 체감하지 못하다가, 나중에 수명초과가 되면, 작업 중이라도 갑작스럽게 데이터 저장 장애가 시작된다는 점이다.

#### 4. 이질형 Hybrid-RAID에 의한 개선 방안

과거부터 하드 디스크 저장 시스템 업계에서는 저장 성능과 응답 성능의 문제해결을 위하여 여러 개의 디스크를 연동한 RAID[1,3] 기술을 사용하여 왔었다. 마찬가지로 SSD도 성능과 안정성을 모두 개선하여야 하므로, 이 RAID 기술의 활용이 필요하다. 단, 단일 매체로 구성된 RAID는 아무리 그 구성 방법을 바꾸어도, 그 단일 매체의 자체의 본질적인 약점을 제대로 극복하기는 어렵다. 예를 들어, HDD만으로 아무리 혁신적인 RAID기법을 개발하더라도, HDD의 구성만 달라질 뿐, 본질적인 기계적인 약점은 그 속에 그대로 내재되어 있기 때문이다. SSD의 경우도 마찬가지이다.

따라서, 반드시 서로의 약점을 보완해줄 성질이 다른 이질형(Heterogeneous)-Hybrid-RAID 기술이 필요하다. 즉, 한 매체의 약점이 다른 매체에서는 강점인 반대 특성을 활용하여, 그 약점을 흡수한다.

[표 1] HDD와 SSD의 장단점 및 대비 특성표

특성들	용량당 비용	고용량성	읽기 성능	저장 안정성	전력 효율	수명 안정성
HDD	낮음	높음	낮음	높음	낮음	높음
SSD	높음	낮음	높음	낮음	높음	낮음

이러한 관점에서 기존 방법론의 한계를 극복하여, HDD의 대용량성과 SSD의 고속성 등의 장점을 살리며, 단점을 상호 보완하여, 전체적인 시너지 효과를 낼 수 있는 다음과 같은 고성능 Hybrid-Raid 기

술의 활용을 제안한다.

### 1) 이질형 매체 특성 고려 클러스터링

일반적으로, 하드 디스크의 경우는 읽기-위주의 데이터 클러스터링이 적합하고, 반면에 SSD의 경우는 쓰기-위주의 데이터 클러스터링이 적합하다. 기존의 하드 디스크와 같은 저장장치들은 기계적인 헤드/디스크 이동 오버헤드를 줄이기 위하여 읽기 중심으로 데이터/인덱스 블록을 순차적으로 모아서 판독시 성능을 개선하도록 발전하였다.(read시의 prefetch 효과) 그러나, 플래시 메모리 SSD는 하드 디스크와는 달리 기계적인 움직임이 없으므로, 데이터 저장 위치에 따라서 입출력 오버헤드가 발생하지 않는 임의 접근(Random Access) 장치이다. 따라서, 읽기 중심으로 데이터를 모을 필요가 전혀 없고, 대신에 쓰기 연산의 오버헤드가 매우 심하므로, 변경이 발생한 데이터는 같은 데이터 클러스터에 모아서 한번에 소거 연산과 쓰기 연산을 진행하여야만 훨씬 더 효율적이다. 오히려 역지로 순차접근을 위하여 주변에 있는 기존 데이터의 이동이 부수적으로 발생한다면, 이는 CPU 낭비뿐 아니라, 플래시 메모리 상에 불필요한 쓰기 및 소거 연산의 발생으로 전체 시스템의 성능이 심각히 저하된다. 따라서, 이질적 특성을 가진 개별 저장 장치 특성에 맞추도록 독자적인 클러스터 관리 기법의 설계가 필요하다. 본 연구에서는 HDD와 SSD의 입출력 특성이 서로 상반되므로, HDD는 읽기 위주로 데이터를 클러스터링하고, SSD는 쓰기 위주로 데이터를 모아쓰는 이질형 매체-적응형 클러스터링 기법을 구사한다.

### 2) 빈도 특성 고려 클러스터링

사용자 데이터 중에는 빈번히 쓰기가 이루어지는 핫-속성의 데이터(파일의 속성 데이터, 작업 중인 데이터)와 잘 접근이 되지 않는 콜드-속성의 데이터(오래된 백업 데이터, 한번 보고 나면 잘 안보는 영상 콘텐츠)가 있다. 플래시 메모리는 핫-속성의 다수의 데이터들을 같은 집합 영역으로 클러스터링을 유도하면, 개별적으로 수행 해야 할 쓰기와 소거 연산을 집합 개념으로 한 번에 몰아서 쓰기/소거 연산을 수행하여, 고질적인 쓰기-연산의 오버헤드를 크게 줄일 수 있다. 특히, 쓰기가 빈번하여 변경이 자주 발생하는 데이터들 끼리 동일 클러스터에 모으면, 상대적으로 마구 섞여 있는 경우보다 전체적으로 소거(클리닝)오버헤드가 줄어서 결국은 수명도 연장된다. (마치, 개구쟁이 아이들만 모아서 한 방만 쓰게 하면, 그 방만 청소하면 되므로, 여러 방에 분산하는 것보다 청소(소거) 비용이 줄어드는 이치와

같다.) 또한, 플래시 메모리는 용량대비 고가이므로, 잘 접근이 되지 않는 콜드 속성의 데이터는 압축해서 저장하면 매우 공간효율적이다. 아래의 그림은 SSD의 효율적인 저장과 쓰기 연산의 부담 감소를 위하여 설계한 기본적인 핫-콜드 클러스터링 기법의 초안인데, 추후 더 개선한다.

### 3) 크기 특성 고려 클러스터링

소용량 데이터 클러스터링: 일반 사용자 컴퓨터의 파일시스템을 분석해보면, 대다수의 파일은 그 크기가 작으며, 전체 저장 장치에서 차지하는 용량도 적다.[13] 따라서, 일반적인 소용량 데이터의 경우는 하드 디스크 보다 페이지 크기가 작은 플래시 메모리(SSD)를 사용하도록 유도하면, 파일 단편화 낭비를 최소화 할 수 있으므로, 전체 저장 공간 활용 측면에서 훨씬 더 효율적이다. 또한, 대다수의 I/O 접근은 작은 파일에 집중되는 경향이 크다. 따라서, 일반적인 작은 파일의 경우는 고속의 플래시 메모리에 저장하도록 설계하면 파일의 접근 성능도 역시 향상시킬 수 있다.

대용량 데이터 클러스터링: 전체 저장 용량의 대부분은 큰 사이즈의 파일이 차지한다. 최근 멀티미디어 및 UCC 데이터의 증가로 대용량화가 지속되고 있으며, 이 큰 파일들은 인터넷을 통하여 다운로드 받아 저장되는 경우도 매우 많다. 이 파일들의 특성은 대용량으로 저장되며, 대부분 읽기 연산 위주로 접근되며, 비디오 재생, 스트리밍 방식으로 관독된다. 따라서 저장 비용상 플래시 메모리의 전체 저장이 절대 불가능하며, 하드 디스크에 저장되어야 한다. 또한, 최근 하드 디스크의 기동후의 접근 성능이 충분하므로, 재생시 스트리밍에 필요한 데이터 전송은 거의 문제가 없다. 다만, 대용량 데이터의 재생초기의 응답 성능 향상을 위하여 그 대용량 파일의 메타 데이터나 첫 데이터 블록만큼은 플래시 메모리에 저장해도 그 용량이 작으므로 부담 없이 개선 효과를 볼 수 있다.

요약하면, 작은 사이즈의 파일은 대부분 자주 접근되며 읽기 쓰기가 빈번한 동적인 특성을 가진 반면, 대용량의 파일은 접근 빈도가 낮고 대부분 읽기 연산 위주의 정적인 접근 특성을 가지고 있다. 기존의 LRU, LFU, MRU, 교체 기법과 Prefix, Sliding window, Segment, Layered, Version 캐싱 등의 기법들도 모두 하나의 매체만을 대상으로 개발된 방법이며, 특히 플래시 메모리의 특성에 대한 고려가 부족하다. 무조건 사용이 적게 되었다고 교체를 하게 되면, 읽기 연산에 비하여 쓰기 연산의 비용이 훨씬 더 큰 플래시 메모리 상에서는 성능 저하가 크게 발

생하게 된다. 더욱이 쓰기 연산은 매우 느린 소거 연산을 수반하게 되므로, 이를 방지하지 못한다면, 그 성능 저하는 더 심각해질 수 있다. 따라서 이를 충분히 보완 설계하여야 하며, LRU, LFU, MRU, two-level LRU, LRU-K, FBR, FLRU, Sliding Window, Segment 등등의 기존 기법들은 특히 매체 고유 특징에 대한 고려가 부족하므로, 복합 매체의 위와 같은 특성을 반영한 최적화된 새로운 클러스터링 기법이 필요하다.

#### 4) 저장비용성능옵티마이저

연산 관련 비용과 성능 요인을 모델링하고 분석하는 최적화 기술이 필요하다. 특히, SSD는 기존 최적화 모듈과는 다른 새로운 수식이 요구된다, 예를 들면, 하드디스크는 write의 경우 덮어쓰기(overwrite)가 가능하기 때문에 write에 대한 비용 모델이 단순하지만, 플래시 메모리의 경우는 덮어쓰기가 불가능하므로, 기존의 write의 비용 모델을 수정하여, 쓰기 시간과 소거 시간을 고려한 전혀 다른 모델이 필요하다.

더 나아가서 데이터베이스 관점에서 full table scan과 index scan의 비용 문제에 관한 것으로, 기존 비용 모델의 재검토가 필요하다고 한다. 또한, 정렬(sort) 알고리즘에서는 다른 정렬방식 보다, 읽기는 많으나, 쓰기는 상대적으로 적은 합병(merge)방식의 정렬효과가 크게 기대된다. 조인(Join)에서는 write를 필요로 하지 않는 nested loop가 더 낫은 최적의 조인 방법이 될 수도 있다. 즉, 검색 최적화와 관련한 예로는, 일반적으로 sort-merge나 hash-join이 더 좋은 조인 알고리즘으로 알려져 있으나, 이 두 가지 모두 데이터 용량이 클 경우, 임시 디스크에 write를 필요로 한다. 따라서, 임시 디스크에 write를 필요로 하지 않는 nested loop가 SSD에는 더 좋은 최적의 조인 방법이 될 수 있다. 또한, SSD의 데이터와 인덱스를 random write에서 SSD의 특성을 살린 append only의 sequential write 연산으로 변경하고, 나중에 재 병합과정에서 random write를 최소화하는 기법을 활용하고자 한다. 또한 SSD 및 HDD에 맞게 더 최적화하여 보완하고자 한다.

#### 4. 결론 및 향후 과제

본 논문에서는 전통적인 하드디스크 드라이브(HDD)와 더불어 최근 주목 받고 있는 차세대 저장장치인 Solid State Drive에 대한 시장 동향과 기술 이슈에 대하여 분석하였다.

또한, 대용량 HDD 및 고속 SSD에 대하여, 저장

장치로서의 약점을 상호 보완하면서, 장점을 강화하여, 최종적으로 입출력 성능을 극대화할 수 있는 새로운 융합형 RAID 시스템을 제안하였다. 향후 시스템 설계 및 컴퓨터 시뮬레이션을 완료한 후, 구체적인 저장 시스템의 구현과 성능 실험을 수행하고, 그 결과를 토대로 최종 시스템 구현하고자 한다.

#### 참고문헌

- [1] 김호진, 황인철, 맹승렬, 윤현수, "SDIO에서 RAID 레벨 5의 효율적인 구현," 정보과학회 학술발표대회, 제31권 제1호, pp. 64-66, 2004.
- [2] 배영현, "고성능 플래시 메모리 SSD 설계 기술", 정보과학회지 vol 25, No. 6, pp. 18-28, 2007.6.
- [3] 최귀열, 박계원, "지능형 I/O 구조를 갖는 RAID 시스템의 성능향상을 위한 연구," 정보통신학회지 제10권 11호, pp. 1989-1995, 2007.
- [4] Chanik Park, Jaeyu Seo, Dongyoung Seo, Shinhan Kim, Bumsu Kim, "Cost-Efficient Memory Architecture Design of NAND Flash Memory Embedded Systems", 21st International Conference on Computer Design, San Jose, California, pp. 474-479, October 13-15, 2003.
- [5] Li-Pin Chang, Tei-Wei Kuo, "An Efficient Management Scheme for Large-Scale Flash Memory Storage Systems," Proc. of ACM SAC'04, pp. 862-868, Nicosia, Cyprus, March 14-17, 2004.
- [6] SpinPoint T Series, <http://www.samsung.com/Products/HardDisk/SpinPointTSeries/index.asp>, 2007.
- [7] What is Flash?, <http://www.samsung.com/Products/Semiconductor/Flash/WhatisFlash/Flash.htm>, 2007.
- [8] Yim K., "A Novel Memory Hierarchy for Flash Memory Based Storage Systems", Journal of Semiconductor Technology and Science, 5:(4), pp. 262-269. Dec. 2005.