

# 플래시 메모리를 이용한 DBMS 리두 로그 영역에 대한 성능평가

김상우, 김재명, 이상원  
 성균관대학교 전자전기컴퓨터공학과  
 e-mail : {spun, jam02, swlee}@skku.edu

## Performance Evaluation of DBMS Redo Log Area using a Flash Memory

Sang-Woo Kim, Jae-Myung Kim, Sang-Won Lee  
 Electrical and Computer Engineering, Sungkyunkwan University

### 요 약

최근 휴대기기용 저장매체로 플래시 메모리가 각광받고 있다. 그러나 플래시 메모리의 휴대성 측면 외에 빠른 속도 측면을 주목하여, 플래시 메모리를 DBMS의 리두 로그 영역 저장장치로 활용할 때 나타나는 성능 특성을 실험을 통하여 알아본다. 그리고, 결과를 분석하여, 앞으로의 플래시 메모리를 이용한 데이터베이스 시스템의 연구방향을 제안한다

### 1. 서론

플래시 메모리(Flash Memory)는 지난 몇 년간 휴대기기 - PDA, MP3 플레이어, 디지털 카메라, 핸드폰 등 - 의 저장매체로 급부상하였다. 이러한 배경에는 플래시 메모리 가격 하락과 더불어 플래시 메모리가 하드디스크에 비해 가지는 장점 - 부피, 무게 전력소모, 내구성 - 때문이다.

위의 장점 외에 플래시 메모리는 하드디스크와는 근본적으로 다른 성능 특성을 보인다. 지금까지 컴퓨터 공학 역사를 지배해왔던 하드디스크는 기계적인 움직임을 필요로 하지만, 플래시 메모리는 전자적인 방법으로 구동되므로 언제나 동일한 빠른 속도로 데이터에 접근하여 사용할 수 있다.

DBMS 리두 로그 영역의 IO 패턴은 플래시 메모리에 적합하다. 실제 Solid State Disk(SSD)에 적용하여 실험한 결과, 로그 쓰기를 위해 대기한 평균시간은 하드디스크의 1/10 임을 확인하였다. 그럼에도 총 수행시간은 큰 차이가 없었고, 분석을 통해 플래시 메모리에서 나타나는 이러한 이상현상을 설명한다.

논문의 구성은 다음과 같다. 2 장에서는 플래시 메모리의 속도를 하드디스크와 비교하고, 3 장에서는 DBMS의 리두 로그 영역에 대해 설명한다. 4 장에서는 실제로 SSD를 이용하여 하드디스크와의 성능을 비교 및 분석한다. 마지막으로 5 장에서는 결론 및 향후 연구에 대해 기술한다.

### 2. 플래시 메모리의 속도

플래시 메모리는 하드디스크와 본질적으로 다른 특성을 가지고 있다[1]. 표 1은 플래시 메모리의 동작속도

를 간략하게 요약한 것이다.

장치	동작속도		
	읽기	쓰기	지우기
HDD <sup>1</sup>	12700 us(2KB)	13700 us(2KB)	N/A
NAND <sup>2</sup>	131 us(2KB)	306 us(2KB)	2000us(128KB)

<표 1>. 접근속도(HDD와 NAND 플래시)

즉, 동작속도를 보았을 때 장치의 IO per second (IOps) 측면에서 하드디스크에 비해 우위에 있다. 플래시 메모리를 사용한 디스크에서 얻을 수 있는 대략적인 IOps에 대해서는 [2]의 실험을 참고하기 바란다. IOps 수치가 절대적으로 중요한 지표인 엔터프라이즈 DBMS 환경에서 플래시 디스크를 하드디스크 대신 사용하는 것 만으로도 더 좋은 성능을 낼 수 있을 것이라 짐작한 것이 본 논문 연구의 시작이 되었다.

### 3. DBMS의 리두 로그 영역

DBMS의 리두 로그 영역은 Write-Ahead-Logging 프로토콜[3]을 지원하기 위해, 사용자 혹은 시스템이 행하는 모든 작업을 데이터 기록 이전에 로그를 우선적으로 기록하는 공간이다. 이를 위해 메모리에는 리두 로그 버퍼가 존재하는데, 이 버퍼에 기록된 내용들은 1) 특정한 주기 혹은 2) 사용자가 Commit 명령 수행 시 또는 3) 리두 로그 버퍼가 다 찼을 때, 디스크에 기록하는 방식으로 작동된다. DBMS의 신뢰성

<sup>1</sup> Seagate Barracuda 7200.7 ST380011A, seek time 과 rotation time 을 포함한 평균 접근 시간

<sup>2</sup> 참조문헌 [2]

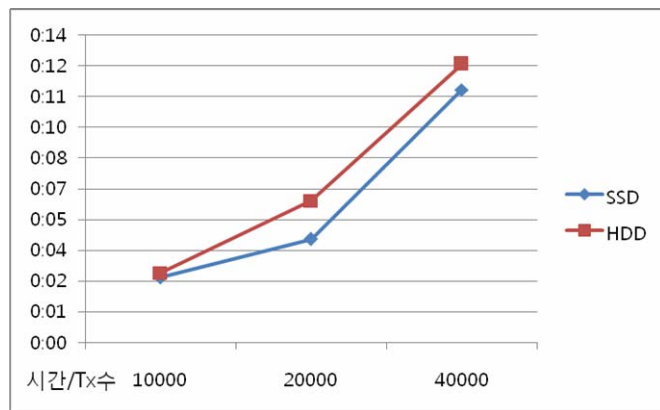
(Durability) 보장을 위하여, 리두 로그는 반드시 디스크에 쓰여야 되며, 모든 변경 내용에 대해서 기록이 일어나야 하므로 많은 IO 가 발생하게 된다. 그리고 순차 쓰기를 잘 하는 디스크의 특성 때문에 지금까지의 DBMS 들은 리두 로그에 많은 양의 쓰기 작업을 해야 하더라도 잘 동작되어 왔다. 그런데 플래시 메모리가 더 높은 IOps 수치를 보여주므로 리두 로그 영역을 플래시 메모리 상에 생성하고, 많은 양의 요청을 주었을 때, 더 좋은 성능을 낼 수 있을지를 4 장의 실험을 통하여 확인하겠다.

4. 성능평가

현재 국내외에서 가장 많이 사용되는 상용 DBMS 의 리두 로그 영역만 7200 rpm 의 IDE 하드디스크와 개발 버전의 SSD 를 Raw device 모드로 설정한 후, 여러 개의 세션 상에서 수 만 건의 트랜잭션을 수행하여, 수행시간을 비교하는 방법으로 실험하였다. 리두 로그에 대한 성능실험이므로 단 건 Insert 명령 후 Commit 명령을 수행하도록 트랜잭션을 작성하였다. 그리고 공정한 실험을 위해 DBMS 의 버퍼 캐시는 수백 MB 로 넉넉하게 설정하였고, 하드디스크와 SSD 모두 쓰기 버퍼 캐시를 사용하지 않도록 설정 후 실험하였다. 펜티엄 4-2GHz, 1GB 메인 메모리, Linux 커널 버전 2.6 계열에서 수행한 실험의 결과는 표 2 와 그림 1 와 같다.

		세션 1	세션 2	세션 3	세션 4
SSD	10k Tx	184	189	182	174
	20k Tx	283	245	249	386
	40k Tx	691	695	757	696
HDD	10k Tx	198	194	196	194
	20k Tx	399	400	394	399
	40k Tx	780	782	785	780

<표 2>. 세션별 실험결과



(그림 1). 수행 평균 시간

표 2 의 세션 1~4 는 동시에 수행한 세션의 수를 의미한다. 10k, 20k, 40k 는 각각 1 만, 2 만, 4 만 건의

1Byte Insert 를 각각의 세션에서 수행했음을 의미한다. 실험결과는 수행 시간(초)을 의미한다.

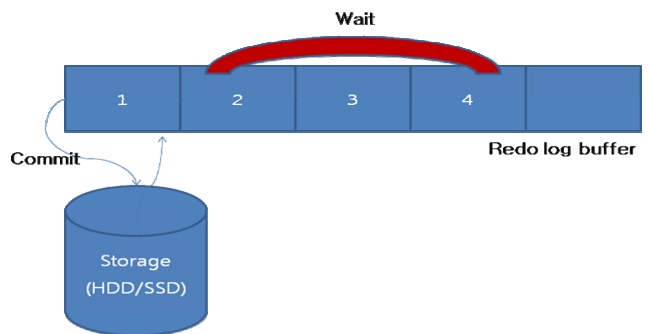
그리고 또 다른 실험으로 512Byte 의 트랜잭션을 2 만건 Insert 하고 시스템 카탈로그[3] 정보를 통하여 해당 세션의 수행시간과 트랜잭션당 평균 대기 시간을 출력하도록 하였다. 이 실험의 결과는 표 3 과 같다. 표 3 은 수행시간(초)과 대기시간(초)을 전체 세션에 대해 취합하였고, 수행시간에서 대기시간이 차지하는 비율을 나타내고 있다.

	수행시간	대기시간	비율
SSD	1501	97	6.5%
HDD	1610	1012	62.8%

<표 3>. 수행 및 대기시간

표 3 에서 볼 수 있듯이, SSD 는 하드디스크에 비해 기다리는 시간이 10 배 정도 적은 수치를 보인다. 다시 말해, 플래시 메모리를 사용할 때 실제 대기 시간과 개별 평균 대기 시간이 1/10 로 줄어든 것을 알 수 있다. 그러나 표 2 와 그림 2 의 결과를 보면, 전체적으로 봤을 때 예상과 다르게 플래시 메모리의 높은 IOps 에도 불구하고, SSD 에서 뚜렷한 성능의 향상은 얻을 수 없었다.

이는 DBMS 엔진 내부 구조와 SSD 내의 Flash Translation Layer(FTL)[4]의 구현 내용을 해당 기업이 공개하지 않으므로 현재로서는 다음과 같은 요인 때문이라고 짐작된다. 표 2 의 SSD 결과를 보면 하드디스크와 다르게 세 개의 세션에 비해 유독 한 세션의 시간이 오래 걸리는데, 이는 한 세션이 다른 세션들에 비해 유독 많이 기다렸기 때문으로 생각된다.



(그림 2). 로그 큐 예제

그림 2 의 예를 보면, 1 번이 Commit 되는 동안에 큐에 쌓이는 양이 SSD 인 경우 쓰기작업이 빨리 수행되어 3 개까지 쌓일 수 있으나, 하드디스크는 10 개까지 쌓인다고 가정할 때, 다음 번 Commit 단위에서는 SSD 는 한번에 3 개를 쓰기작업을 수행할 것이고, 하드디스크는 10 개를 쓰기작업을 수행하게 될 것이다. 즉, 생산자-소비자간의 소비속도를 고려해야 할 것이고, 이를 위해 큐잉 이론적으로 접근이 필요할 것이다. 표 2 의 실험을 통해 결과를 확인할 수 있으나, 이 실험

외에도 더 많은 수의 세션에서 실험해 보았지만, 특별한 패턴이 발견되지 않았던 것이 그 이유이다.

## 5. 결론 및 향후 연구

플래시 메모리는 하드디스크에 비해서 충분히 높은 I/Os 수치를 얻을 수 있다. 특히 하드디스크 기반의 DBMS에서는 랜덤 읽기가 성능 병목의 주 원인인데, 플래시 디스크를 사용하면 랜덤 읽기 측면에서 큰 성능 향상을 얻을 수 있다. 로그 영역만 보았을 때도, 쓰기 성능은 제외하더라도, 읽기 성능이 빠르므로, 플래시 메모리에 로그 영역을 사용하면 고장 회복(crash recovery)을 매우 빠르게 수행할 수 있다.

그러나 현재 하드디스크 기반의 DBMS를 그대로 사용하기에는 4장에서 언급하였던 문제들 때문에, 기대한 것만큼의 성능향상은 없을 것으로 예상되며, 이를 위해 DBMS의 버퍼 관리자, 저장장치 관리자 측면에서 개선이 필요하다. [5]은 이와 관련된 연구로 이러한 문제에 대한 새로운 접근방법을 제안하였다.

또한, 로그 버퍼 영역에서 플래시 메모리의 빠른 속도를 이용하기 위해서 큐잉 이론을 고려하여 연구를 수행해야 하며, 이 문제가 해결된다면 플래시 메모리를 이용한 DBMS는 (특히 로그영역에 대하여) 성능이 많이 개선될 것으로 보인다.

향후 연구로 다양한 실험을 통하여, 로그 영역의 쓰기 작업에 대하여 일정한 법칙을 추론할 것이다. 또한 현재까지 CPU의 연산 비용과 DBMS의 접속프로그램에 따른 네트워크 비용 등에 대한 반영이 부족했으므로, 이 점을 고려하여 연구를 수행할 것이다.

## Acknowledgment

본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT 연구센터 지원사업(IITA-2006-(C1090-0603-0046))과 정통부 및 정보통신연구진흥원의 정보통신선도기반기술개발사업의 연구결과로 수행되었습니다.

## 참고문헌

- [1] TOSHIBA, "NAND Flash Applications Design Guide", 2004
- [2] Jim Gray, "FLASH Disk Opportunity for Server-Applications", <http://research.microsoft.com/~Gray/papers/FlashDiskPublic.doc>
- [3] R. Ramakrishnan, J. Gehrke, "Database Management Systems, 3rd Ed.", McGraw-Hill, 2004
- [4] Intel Corp., "Understanding the Flash Translation Layer(FTL) Specification", 1998
- [5] Sang-Won Lee, Bongki Moon, "Design of a Flash based DBMS: An In-page Logging Approach", ACM SIGMOD International Conference on Management of Data, 2007 (to appear)