

# 쓰기 지연을 활용한 플래시 SSD 기반 DBMS의 저장 장치 관리 기법

김강년<sup>o</sup> 강운학 이상원  
성균관대학교 전자전기컴퓨터공학과

[kangnuni@skku.edu](mailto:kangnuni@skku.edu), [woonagi319@skku.edu](mailto:woonagi319@skku.edu), [swlee@skku.edu](mailto:swlee@skku.edu)

## Storage Management Scheme of Flash SSD based DBMS using Delayed Write

Kang-Nyeon Kim<sup>o</sup> Woon-Hak Kang Sang-Won Lee  
Department of electrical and Computer Engineering, SungKyunKwan University

### 1. 서론

최근 낸드 플래시 메모리는 집적도 향상에 따라 용량이 커지고 가격은 지속적으로 하락 하고 있다. 특히 MLC칩의 등장으로 SLC칩 대비 두 배의 집적도 향상이 가능해졌고 향후 1셀에 3비트 이상이 저장되는 낸드 플래시 메모리 칩이 등장할 것으로 보여 이러한 추세는 지속 될 것으로 보인다.

플래시 SSD는 하드 디스크와 동일한 인터페이스로 사용할 수 있는 플래시 메모리 기반 저장 장치로 최근 개인용 컴퓨터와 서버 시장에서 하드 디스크를 대체하기 시작 했다. 플래시 SSD의 특성은 순차 접근 처리가 빠르고, 임의 읽기 성능 또한 하드 디스크 보다 좋은 것으로 알려져 있다. 하지만 임의 쓰기의 경우 쓰기 증폭(Write Amplification)의 심화로 인해 플래시 SSD의 성능과 수명을 크게 저하시킨다. 또한 플래시 메모리의 발전 추세가 집적도가 향상될 때마다 내구성이 크게 감소하고 프로그래밍과 삭제 연산의 단위는 커지는 것으로 나타나고 있어 소량의 임의 쓰기가 플래시 SSD에 주는 부담이 커질 것으로 예상 된다.

임의 쓰기가 유발하는 쓰기 증폭 문제를 완화하기 위해 SSD 제조 업체들은 내구성이 좋은 SLC칩을 사용하거나 복잡한 컨트롤러 기술을 적용한 서버용 제품들을 출시 하고 있다. 이러한 제품들은 임의 쓰기 요청이 빈번한 OLTP(On-Line Transaction Processing) 환경에서 비교적 향상된 성능과 제품 수명을 보이지만 가격이 높아지는 것이 일반적이다.

이처럼 플래시 SSD 기술 발전이 OLTP 환경의 DBMS 성능 향상에 기여 하고 있지만 플래시 SSD를 보다 효율적으로 활용하기 위해서는 디스크 기반으로 설계된 기존 DBMS의 저장 장치 관리 구조를 재설계할 필요가 있다. 따라서 본 논문은 쓰기 지연 기법을 통해 수명 관리와 성능 측면에서 플래시 SSD에 효율적인 저장 장치 관리 기법을 제안 하고 이를 성능 분석을 통해 검증 한다.

### 2. 본론

OLTP 환경에서 기존의 DBMS가 플래시 SSD에 어떤 영향을 주는지 알아 보기 위해 PostgreSQL 상에서 TPC-C 벤치마크를 수행하고 I/O 트레이스를 수집했다. 트레이스 분석 결과 OLTP 환경에서는 광범위한 임의 쓰기가 발생하지만 20%의 데이터 페이지에 대한 쓰기 회수가 전체 쓰기 회수의 72%를

---

본 연구는 서울시의 지원으로 수행한 '서울시 산학연 협력사업(PA090903)'과 지식경제부 및 정보통신 산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음 (NIPA-2010-(C1090-1021-0008))

차지하는 지역성이 존재하는 것으로 확인 되었다. 이러한 결과를 통해 버퍼 교체 시점 마다 데이터 페이지를 디스크에 쓰는 기존의 방식이 덮어쓰기가 많은 임의 쓰기를 유발하여 플래시 SSD 내부의 쓰기 증폭을 심화시키는 것을 확인 했다. 이를 해결하기 위해 본 논문은 버퍼 교체 시점의 데이터 페이지 쓰기를 체크포인트 시점까지 지연시킨다.

이러한 쓰기 지연을 통해 얻는 장점은 쓰기 절약(Write Saving)과 플래시 SSD의 병렬성 이용으로 요약 된다. 버퍼 교체에 의한 페이지 쓰기를 체크포인트 시점까지 미루면 같은 페이지에 대한 덮어쓰기가 여러 번 발생해도 체크포인트 시점에 한번만 쓴다. 덮어쓰기 회수를 줄이는 것은 잠재적으로 삭제 회수 또한 줄이기 때문에 성능과 내구성 측면에서 모두 좋은 결과를 가져 온다. 또한 버퍼 교체 때마다 단일 데이터 페이지 단위로 쓰기 요청 하던 기존 방식과 달리 체크포인트 시점에 일괄적으로 쓰기 때문에 최근의 SSD들이 갖고 있는 병렬성을 활용할 수 있다.

하지만 쓰기 지연의 대가로 버퍼에서 방출된 데이터 페이지를 다시 읽을 때 로그 저장 장치에서 로그를 읽고 반영하여 최신 데이터 페이지를 생성해야 하는 읽기 페널티(Read Penalty)가 발생 한다. OLTP 트레이스 분석 결과에 의하면 읽기 횟수가 쓰기 횟수에 비해 4.5배 이상으로 많기 때문에 읽기 페널티 문제는 더욱 심각해진다. 이 문제를 해결 하기 위해 본 논문에서는 로그 버퍼의 동작을 변경 한다. 커밋이나 버퍼 교체 시에 기존 방식과 마찬가지로 로그 레코드를 로그 저장 장치에 기록 하지만 로그 버퍼를 비우지 않고 체크포인트 시점까지 모든 로그 레코드들을 보존한다. 이는 최신 데이터 페이지를 생성하기 위해 로그를 읽어 올 때, 로그 저장 장치가 아닌 로그 버퍼에 접근 하여 읽기 페널티를 줄이기 위함이다. 따라서 로그 버퍼의 크기는 체크 포인트 주기내의 모든 로그레코드들을 저장 할 수 있어야 한다. 시스템 오류(System Failure)로 로그 버퍼가 손실되어도 커밋 시점에 로그를 저장 장치에 저장 했기 때문에 복구에 영향을 미치지 않는다.

로그 버퍼가 꽉 차면 체크포인트 연산이 수행 된다. 체크포인트는 버퍼 내의 더티 페이지들을 방출 할 뿐만 아니라 버퍼 교체 시점에 쓰기를 지연 해둔 모든 데이터 페이지들을 저장 장치에서 다시 읽고 로그 버퍼의 모든 로그를 반영하는 병합 연산을 수행한다. 병합 연산으로 생성된 최신 데이터들을 저장 장치에 쓰고 로그 버퍼를 비우는 것으로 체크포인트가 완료 된다. 덮어쓰기가 전혀 발생 하지 않은 데이터 페이지의 경우 읽기와 로그 레코드를 적용하는 페널티가 부과 된다. 하지만 덮어쓰기가 여러 번 발생한 데이터 페이지는 쓰기 절약으로 인해 페널티를 상쇄시키는 이익이 발생한다.

본 연구에서 제안하는 기법을 검증하기 위해 비용 모델에 기반한 성능 분석을 수행 했다. 분석 결과 기존 방식 대비 플래시 SSD에 대한 쓰기 회수를 43% 감소 했고 총 I/O 시간은 28% 감소한 것으로 확인 되었다.

### 3. 결론

쓰기 증폭 문제는 덮어쓰기가 불가능한 플래시 메모리의 특성에 기인하기 때문에 근본적으로 피할 수 없다. 하지만 플래시 SSD 기술의 발전과 더불어 상위 계층의 덮어쓰기 최소화를 통해 보다 쓰기 증폭을 완화하고 플래시 SSD를 보다 효율적으로 사용할 수 있다. 이 점에 착안한 본 연구는 쓰기 지연을 통해 덮어쓰기를 최소화하여 내구성과 성능 측면에서 플래시 SSD에 적합한 DBMS의 저장 장치 관리 기법을 제시 했다.