

# NVDIMM 기반 EXT4 파일시스템 외부 저널링 성능평가

박종혁<sup>o</sup>, 이상원

성균관대학교

{akindo19, swlee}@skku.edu

## Performance Analysis of External Journaling on NVDIMM-based EXT4 File System

Jong-Hyeok Park<sup>o</sup>, Sang-Won Lee

Sungkyunkwan University

### 요약

최근 비휘발성 메모리 기술에 대한 연구가 활발히 진행되었지만 대부분의 연구가 상용화되지 않는 비현실적인 성능치를 가정한 시뮬레이션 중심으로 진행되었다. 본 논문에서는 상용화된 NVDIMM 및 SSD를 EXT4파일 시스템의 외부 저널링 디바이스로 사용할 때 Filebench의 Varmail 벤치마킹을 통해 대역폭 성능을 평가하였다. NVDIMM을 외부 저널링 디바이스로 사용할 때 가장 높은 대역폭과 가장 낮은 지연을 보임을 알 수 있었다. 이를 통해 NVDIMM의 특성을 고려한 EXT4 파일 시스템의 저널링 기법 연구의 방향에 대한 통찰력을 제공한다.

### 1. 서론

Intel 3D X-point 기술의 발표와 함께 데이터베이스 커뮤니티를 중심으로 PCM, STT-MRAM과 플래시 메모리 이외의 NVM(Non-Volatile Memory)을 활용하는 로깅 및 쓰기 최적화에 대한 연구들이 활발히 진행되고 있다. [1]

그러나 대부분의 연구는 상용화된 제품이 아닌 시뮬레이터 기반의 실험이며, 만약 NVM 기술이 상용화 되더라도 DRAM과 동일한 접근 속도를 내지 못하는 것과 경제성을 고려한다면 현재 플래시 기반의 컴퓨팅 환경에서는 NVM 기반의 최신 관련기술은 현실적인 기술 대안이 될 수 없다.

반면, DRAM 모듈에 플래시 저장장치와 배터리 백업(BBU) 기능을 갖춘 NVDIMM(Non-Volatile DIMM) 제품이 등장했고, NVDIMM은 DRAM과 동일한 속도로 바이트 단위 읽기쓰기 연산이 가능하며 비휘발성 속성을 가진다. 그러나 경제적인 측면을 고려한다면, NVDIMM은 DRAM을 대체할 수 없고 플래시메모리 기반의 저장장치 아키텍처에 NVDIMM이 추가되는 이중 시스템으로 동작하는 것이 가장 효율적인 방안이 될 것이다.

EXT4 파일 시스템은 저널링으로 인한 Bandwidth 저하 문제를 해결하기 위해 외부의 디바이스에 저널링을 할 수 있는 외부 저널링(External Journaling) 모드를 제공한다. NVDIMM의 빠른 읽기 쓰기 연산과 외부 저널링을 활용하면 파일시스템의 저널 대역폭(Bandwidth) 저하 문제를 해결할 수 있다.

따라서 본 논문에서는 NVDIMM 기반의 이중 시스템 환경에서 Filebench 벤치마킹을 통해 EXT4 파일시스템

의 저널 디바이스를 NVDIMM와 SSD를 사용하였을 경우 각각의 성능을 평가했다. 성능평가 결과를 바탕으로 NVDIMM의 특성을 고려한 저널 기법이 가질 수 있는 성능향상에 대해 논한다.

본 논문의 구성은 다음과 같다. 2장에서는 EXT4 파일 시스템의 저널 모드와 외부 저널링에 대해 알아보고, NVDIMM의 특성을 고려한 외부 저널링 기법에 대해 알아본다. 3장에서는 본 논문에서 수행한 성능평가 환경과 결과를 보인다. 마지막으로 4장에서는 결론 및 향후 연구를 제시하고 논문을 마무리한다.

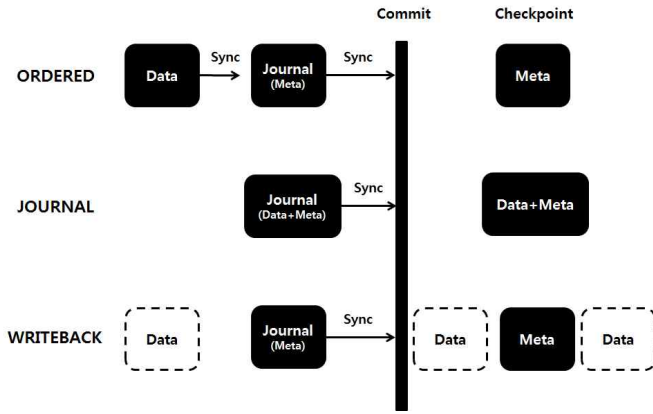
### 2. External Journaling 과 NVDIMM

파일 시스템은 전력공급의 문제나 갑작스런 시스템 충돌이 발생할 경우, 효율적으로 일관성을 보존하기 위해 저널링 기법을 사용한다.

저널링 기법은 파일시스템 내에 업데이트가 발생할 때마다, 실제 스토리지의 위치에 기록하기 전에 복사본을 저널 영역에 먼저 기록하고, 충돌이 발생하는 경우 저널을 읽어 변경사항을 스토리지에 반영하여 복구한다.

EXT4 파일시스템에서 제공하는 저널 모드는 Ordered, Journal, Writeback으로 총 3가지이며, 동작과정은 [그림 1]과 같다. [2] Ordered 모드(EXT4 기본설정)는 데이터를 저널링 없이 기록하고, 이후 메타데이터를 저널링하여 기록하는 방식이다. Journal 모드는 데이터를 저널링한 뒤, 메타데이터를 저널링하고 실제 데이터를 쓰는 방식이다. 따라서 Ordered모드의 경우, 시스템 충돌시 메타데이터는 저널링되어 보존 되지만 데이터는 손실될 수 있

다. 반면, Journal 모드는 데이터도 저널링하기 때문에 시스템 충돌이 발생해도 데이터복구가 가능하다. 마지막으로 Writeback 모드는 데이터 저널링을 전혀 수행하지 않고 저널 모드중에 성능이 가장 좋지만 시스템 충돌이 발생할 경우 데이터가 유실될 가능성이 높다.

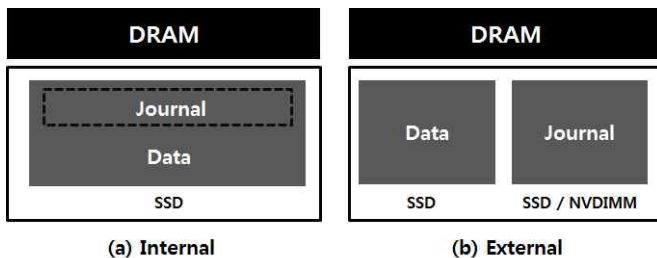


[그림 1] EXT4 파일시스템의 저널모드

저널은 파일시스템과 동일한 디바이스에 저장되므로 저널쓰기와 데이터쓰기의 경합으로 인해 대역폭을 감소시킨다. 특히 스토리지가 포화상태에 가까워 질 때, 저널링 오버헤드로 인해 성능이 크게 저하될 수 있다. [3]

외부 저널링은 위에서 언급한 대역폭 감소문제를 해결하기 위해 외부 디바이스에 저널링하는 방식이며 저널쓰기와 데이터 쓰기를 병렬 처리한다.

[그림 1]에서 볼 수 있듯이, 저널쓰기와 데이터쓰기를 하나의 디바이스에 수행하는 Internal 모드와 각각 다른 디바이스에 수행하는 External 모드가 있다. 외부 저널링이 가지는 병렬성을 활용하면 저널 디바이스가 SSD여도 대역폭이 증가된다. 그러나 저널 디바이스를 NVDIMM으로 설정할 경우, NVDIMM의 빠른 I/O 속도와 저지연 (low-latency) 특성을 활용하면 대역폭을 더욱 크게 증가시킬 수 있다.



[그림 2] EXT4 저널링 위치에 따른 기법

### 3. 성능 평가 및 분석

본 논문에서 성능측정에 사용한 저장장치는 Samsung SSD 850 PRO이고, NVDIMM은 Netlist NVvault DDR4 16GB이다. 자세한 실험 환경은 [표 1]과 같다.

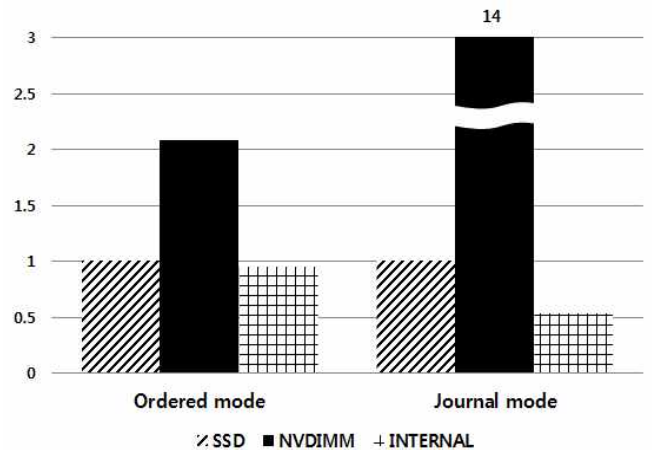
Filebench 벤치마킹에서 사용한 워크로드는 Varmail이다. Varmail 워크로드는 평균적인 작업 파일크기가 16KB이고 쓰기위주의 워크로드이며, fsync() 함수를 통해 다른 워크로드들 보다 더 많은 디스크 I/O를 발생시킨다. [4]

운영체제	Ubuntu 16.04.2 LTS
프로세서	Intel® Xeon E5-2640 2.60GHz (32 Core)
메모리(RAM)	32GB
저장장치	Samsung SSD 850 Pro 256GB
NVDIMM	Netlist NVvault DDR4 16GB
벤치마크	Filebench

[표 1] 실험환경

본 연구는 저널링에 대한 대역폭 감소 문제를 해결하는 것에 초점을 맞췄기 때문에 Writeback 모드를 제외한 두가지 저널 모드에 대해 실험을 진행하였다.

Internal은 외부 저널링을 사용하지 않고 저널과 데이터 쓰기를 하나의 SSD에 수행한 것이고, SSD와 NVDIMM은 각각을 외부 저널링 디바이스로 설정하여 수행한 것이다. [그림 3]은 각각의 경우에 대해 SSD 결과 값으로 정규화한 값이다.



[그림 3] 외부 저널링 디바이스에 따른 대역폭

두가지 저널 모드 모두 NVDIMM을 외부 저널링 디바이스로 사용하였을 때 대역폭이 가장 높았다. 또한 Journal 모드의 경우, 뚜렷한 대역폭 차이를 보인다. NVDIMM을 외부 저널링 디바이스로 사용하였을 때

512MB/s의 대역폭을 가지며 이는 SSD를 External 외부 저널링 디바이스로 사용하였을 때보다 14배 높다. 반면 Ordered 모드의 경우 메타 데이터만을 저널링하기 때문에 Journal 모드는 물론 Internal과 비교했을 때 차이가 작다.

#### 4. 결론 및 향후 연구

본 논문에서는 EXT4 파일 시스템의 저널모드와 외부 저널링을 분석하고, NVDIMM을 외부 저널링 디바이스로 설정할 경우 대역폭 성능을 분석하였다. NVDIMM의 빠른 IO와 속도와 저지연 특성을 활용하면 저널쓰기에 대한 오버헤드가 줄어들기 때문에 높은 대역폭을 가질 수 있다.

성능 평가 결과, NVDIMM을 외부 저널링 디바이스로 설정할 경우, Ordered 모드에 대해서는 2배, Journal 모드에 대해서는 14배 높은 대역폭을 가진다. Journal 모드의 경우, 데이터와 메타데이터를 함께 저널링하기 때문에 큰 차이를 보이지만, Ordered 모드는 메타데이터만 저널링하기 때문에 차이가 크지 않다.

하지만, 지속성을 보장하기 위해 fsync()를 자주 호출하는 DBMS 로깅 시스템을 고려한다면 NVDIMM을 외부 저널링 디바이스로 사용하는 이점을 최대한 활용할 수 있다.

향후 연구에는 NVDIMM의 비휘발성과 저지연 특성을 고려한 파일 시스템의 저널링 알고리즘 개선 연구를 해 볼 계획이다.

#### 5. 참고문헌

[1] Joy Arulraj, Matthew Perron, and Andrew Pavlo. 2016. Write-behind logging. Proc. VLDB Endow. 10, 4 (November 2016), 337-348.

[2] Vijayan Prabhakaran, Andrea C. Arpaci-Dusseau, and Remzi H. Arpaci-Dusseau. 2005. Analysis and evolution of journaling file systems. In Proceedings of the annual conference on USENIX Annual Technical Conference (ATEC '05). USENIX Association, Berkeley, CA, USA, 8-8.

[3] Pedro Eugenio Rocha and Luis C. E. Bona. 2012. Analyzing the Performance of an Externally Journalled Filesystem. In Proceedings of the 2012 Brazilian Symposium on Computing System Engineering (SBESC '12). IEEE Computer Society, Washington, DC, USA, 93-98.

[4] 송용주, 김정훈, 강동현, 이민호, 엄영익. (2016). 시스템 환경이 Filebench 벤치마크에 미치는 영향 분석. 정보과학회논문지, 43(4), 411-418.

#### 사사

본 연구는 미래창조과학부 및 정보통신기술진흥센터의 SW컴퓨팅산업원천기술개발사업(SW스타랩)의 연구결과로 수행되었음 (IITP-2015-0-00314)

이 논문은 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단-차세대정보·컴퓨팅기술개발사업의 지원을 받아 수행된 연구임(No. NRF-2015M3C4A7065696).

이 논문은 2017년도 삼성전자(클러스터과제)의 지원을 받아 수행된 연구임.