

비휘발성 메모리 환경에서 인메모리 키밸류 스토어의 성능 비용 간 트레이드-오프 분석

An Empirical Study of In-memory Key-value Store Performance in Intel Optane Persistent Memory

요 약

최근 intel Optane DC Persistent Memory (DCPMM)의 출시와 함께 고용량 비휘발성 메모리를 상용 시스템에 활용하고자 하는 시도가 활발히 이루어지고 있다. 본 논문은 DCPMM의 주요 응용인 인메모리 키밸류 스토어(Redis)를 타겟으로 비휘발성 메모리 환경에서 구동될 때 발생할 수 있는 성능저하를 관찰하고 이를 통해 비휘발성 메모리를 활용한 성능과 비용 간의 적절한 트레이드오프에 대해 논의한다.

1. 서 론

지난 수십년간 프로그램의 복잡도 증가 및 데이터 기반 응용의 확산은 대용량 메인 메모리에 대한 요구를 지속적으로 증가시켜왔다. 이에 단일 면적 당 많은 비트를 저장하기 위한 메모리 스케일링 기술은 학계와 업계 모두에서 활발하게 연구되어 왔다. 무어의 법칙에서 보여주듯이 메모리 반도체의 집적도는 18개월에 2배로 증가해오며 눈부신 향상을 이루어 왔다. 그러나 최근 DRAM의 경우 셀 간의 간섭 현상이 심해져 10nm 이하의 셀 스케일링은 불가능할 것으로 예측되며 무어의 법칙이 종료될 것이라는 전망이 우세하다[1].

비휘발성 메모리는 이러한 메모리 기술의 한계점을 극복하고 고용량·고성능·고신뢰성 메모리를 제공하기 위해 최근 활발하게 연구되고 있는 기술이다. 고집적이 가능한 새로운 물질을 메모리로 활용해 대용량 메모리에 대한 요구를 효율적으로 대응하고자 하는 것이다. 대표적인 비휘발성 메모리로 PCM (Phase Change Memory)를 들 수 있으며 GST라는 특수 물질의 상변환을 통해 비트를 표현한다. 이 물질은 DRAM 대비 수 배 이내의 성능저하만을 일으키면서도 전원이 소실되어도 상태를 유지하기 때문에 DRAM과 달리 데이터의 영속성도 보장해주어 메모리와 스토리지 영역에서 모두 사용 가능한 이상적인 메모리 기술로 각광을 받아왔다.

최근 intel에서 공개한 Intel Optane DC Persistent Memory (DCPMM)는 PCM과 유사한 기술을 상용화 한 차세대 메모리이다. 이전까지의 비휘발성 메모리 분야의 연구가 가상 장치를 가정하고 이루어진 반면, 실제 비휘발성 메모리 기술이 현실화됨에 따라 실용적 관점에서 이를 활용하고자 하는 연구가 전세계적으로 주목을 받고 있다. 본 논문은 이러한 연구들과 유사한 동기를 가지고 비휘발성 메모리의

도입이 인메모리 키밸류 스토어(In-memory Key-value Store)에서 어떠한 성능과 가격 간의 트레이드오프를 제공하는가에 대한 연구를 수행하였다.

DCPMM은 바이트 단위의 접근을 허용하면서도 슬롯당 수백GB의 용량을 제공하여 고용량 메인 메모리의 구현을 가능하게 한다. 또한 데이터의 영속성을 제공하기 때문에 메모리 내부에서 데이터를 관리하는 인메모리 키밸류 스토어에서 활용도가 높을 것으로 예측되고 있다. 본 연구진은 대표적인 인메모리 키밸류 스토어인 Redis[2]를 타겟으로 비휘발성 메모리 도입에 따른 성능 변화에 대해 관찰하였다.

2. 비휘발성 메모리 도입에 따른 키밸류 스토어 성능 변화

현재 Intel 은 상용 어플리케이션이 DCPMM을 효율적으로 활용할 수 있도록 PMDK(Persistent Memory Development Kit) 라이브러리를 제공하고 있다. PMDK는 비휘발성 메모리에서 데이터를 관리할 수 있도록 해주는 라이브러리이다. Intel 은 주요 타겟 응용인 Redis가 DCPMM 메모리 위에서 데이터를 저장하고 관리할 수 있도록 PMDK 기반의 Redis 확장 버전인 Pmem-redis를 개발하였다[3, 4]. 본 논문은 DRAM에서 데이터를 관리하는 원본 Redis와 비휘발성 메모리를 활용해 데이터를 관리하는 Pmem-redis 의 성능을 비교 분석한다. 이를 통해 비휘발성 메모리 도입에 따른 성능변화를 관찰하고 향후 이질적 메모리를 효율적으로 활용하기 위한 기법 연구의 초석이 되는 데이터를 확보하고자 한다.

그림 1은 기존 Redis 와 Pmem-redis의 전체적인 구조를 비교하여 보여준다. Redis는 기본적으로 데이터를 키(Key)와 값(Value)의 쌍으로 저장하며 해쉬 테이블을 이용해 데이터를 인덱싱한다. 기존 Redis의 경우 키와 값을 모두 DRAM에

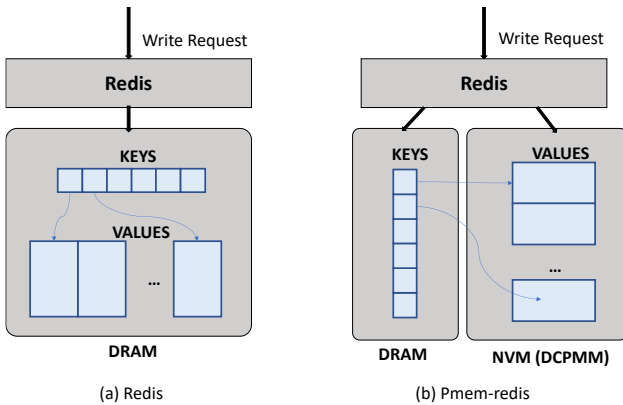


그림 1. Redis와 Pmem-redis 의 저장구조 비교

저장하는 반면 Pmem-redis는 키는 DRAM에 저장하고 값은 NVM(DCPMM)에 저장한다. Pmem-redis는 크기가 작고 접근이 빈번한 키는 상대적으로 성능이 빠른 DRAM에 저장하고 크기가 큰 값은 고용량 DCPMM에 저장하여 저장시스템의 성능을 최적화하는 구조이다.

Pmem-redis 에서 한가지 유의할 점은 비휘발성 메모리를 활용하기는 하지만 부분적으로 활용하기 때문에 데이터의 손실을 막을 수 없다는 것이다. 따라서 실제 데이터 손실을 방지하기 위해서는 기존과 동일하게 영속성을 제공하기 위한 기법을 사용해야 한다. 기존 Redis는 휘발성 메모리에 데이터를 저장하기 때문에 영속성을 보장해야 하는 경우에는 변경사항을 AOF(Append Only File)에 로깅하거나 주기적으로 데이터셋을 스냅샷 형태로 RDB(Redis Database) 파일에 저장하는 방식을 사용한다. Pmem-redis 의 경우 AOF를 이용하여 영속성을 보장할 수 있는데 값 자체는 비휘발성 메모리에 저장되기 때문에 별도의 로깅을 수행하지 않는다. AOF에는 키와 해당 키에 대응되는 값이 저장된 비휘발성 메모리의 주소를 기록하여 전원 소실 시에 데이터를 복구한다.

3. 성능평가

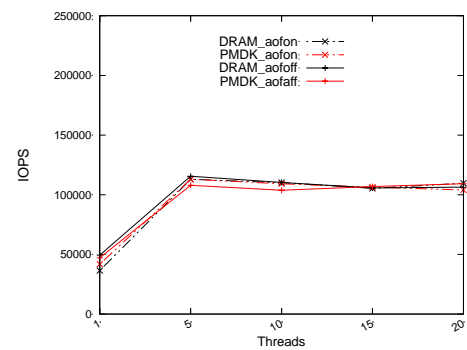
본 논문은 상이한 구조를 가지는 Redis와 Pmem-redis의 성능을 비교해 봄으로써 비휘발성 메모리 사용에 따른 성능 저하를 관찰하고 이를 개선하기 위한 방안에 대해 살펴보고자 한다. Redis는 분산 인메모리 키밸류 스토어이기 때문에 1개의 Master와 2개의 Slave로 구성된 Redis 환경에서 성능평가를 수행하였다. 성능 분석 도구는 Redis의 벤치마크인 Memtier Benchmark[5]를 사용했으며 워크로드는 일련의 랜덤 키와 밸류를 생성한 후 해당 워크로드를 반복적으로 시행하는 순차 Read/Write 워크로드를 사용하였다. 키와 값은 각각 16바이트와 1KB로 설정하였다. 실험은 Intel Optane DC Persistent Memory가 탑재된 Supermicro 서버에서 이루어졌으며 구체적인 시스템 환경은 표1과 같다. DCPMM 메모리는 Ext4-DAX를 이용하여 App. Direct Mode로 접근되었다. Redis가 데이터의 영속성을 보장할 때와 보장하지 않을 때 각각 비휘발성 메모리로 인한 성능 변화를 관찰하기 위해 AOF 기능을 활성화 시킨 경우(aofon)와

| Benchmark Tool | Memtier Benchmark 1.2.17 | |
|----------------|---|------------------|
| Database | Redis 4.0.0 | PMDK Redis 4.0.0 |
| OS | Ubuntu Desktop 18.04 LTS 64Bit | |
| CPU | Inter Xeon Gold 5215 CPU @ 2.50GHz (10 cores 2 sockets) | |
| RAM | DDR4 256GB | |
| Optane Memory | 128GB * 4 | |

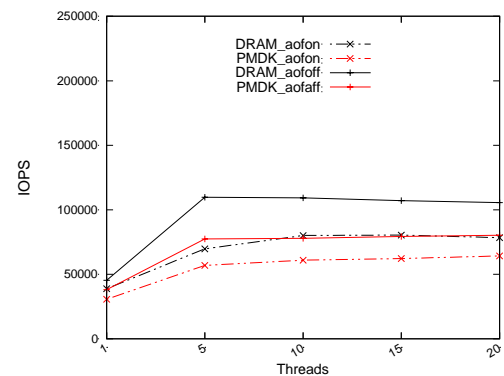
표 1. 성능 평가 시스템 환경

비활성화 시킨 경우(aoff)에 대해 각각 성능을 평가하였다. 실험결과는 총 10번을 수행하고 중간 값을 사용하였다.

그림 2는 Memtier Benchmark에서 GET과 SET 연산에 대한 Redis와 Pmem-redis 의 성능을 비교하여 보여준다. x축은 Redis에 요청을 보내는 클라이언트의 스레드 숫자로 증가할수록 요청이 동시다발적으로 유입된다. GET에서는 Redis와 Pmem-redis의 성능격차가 거의 관찰되지 않았다. 반면 SET의 경우에는 설정 별로 상당한 성능 차이가 관찰되었다. 우선 AOF를 사용하지 않을 때 DRAM과 NVM간의 성능 격차가 상대적으로 크게 나타났다. AOF 비활성화 모드에서 스레드가 50이상인 경우 기존 Redis와 비교했을 때 Pmem-redis는 평균 27%의 성능저하를



(a) GET의 성능



(b) SET의 성능

그림 2. Redis와 Pmem-redis 성능 비교

나타났다. 반면 AOF를 실행하는 경우에는 성능격차가 21%로 데이터 로깅으로 인한 스토리지 접근이 발생하여 메인 메모리의 속도 차이가 전체 성능에 끼치는 영향이 감소하기 때문으로 판단된다. 한편, 쓰레드 수 증가에 따른 가시적인 성능변화는 관찰되지 않았는데 이는 Redis 자체가 단일 쓰레드로 구현되어 내부적으로 동시다발적 실행을 수행하지 않기 때문으로 판단된다. 이러한 결과는 읽기 위주의 워크로드에서는 비휘발성 메모리의 사용이 성능과 비용 간의 적절한 트레이드오프를 허용함을 알 수 있다. 현재 비용당 바이트로 메모리의 집적도를 추산했을 때 DRAM과 DCPMM의 집적도 차이는 거의 없거나 오히려 DCPMM이 더 낮을 수 있다. 그러나 DCPMM의 경우 단일 슬롯당 제공하는 용량이 DRAM의 수십배에 이른다는 점을 고려할 때 TCO(Total Cost of Ownership) 관점에서는 DCPMM을 활용하는 것이 상당한 비용을 절감할 수 있도록 해준다.

5. 결론

본 논문에서는 최근 새롭게 출시된 Intel Optane DC Persistent Memory 환경에서 인메모리 키밸류 스토어가 워크로드에 따라 어떠한 성능 변화를 나타내는지 관찰하였다. 읽기에서는 성능격차가 미미한 반면 쓰기에서는 상당한 성능차이를 나타냈다. 또한 Redis가 데이터의 영속성을 보장하는 경우와 그렇지 않은 경우의 성능 격차가 상이한 것을 관찰하였다. 이러한 실험결과는 DCPMM 메모리와 Pmem-Redis를 사용하고자 하는 다양한 응용들이 성능과 비용 간의 합리적인 트레이드오프를 할 수 있도록 가이드라인을 제시해 줄 수 있을 것으로 기대된다. 향후에는 소프트웨어 최적화를 통해 비용 증가 없이 DRAM과 DCPMM 성능격차를 최소화할 수 있는 방안에 대해 연구하고자 한다.

참고문헌

- [1] <https://semiengineering.com/dram-scaling-challenges-grow/>
- [2] <https://redis.io/documentation>
- [3] <https://github.com/pmem/pmem-redis>
- [4] https://ci.spdk.io/download/2019-summit-prc/02_Presentation_13_Accelerating_Redis_with_Intel_Optane_DC_Persistent_Memory_Dennis.pdf
- [5] https://github.com/RedisLabs/memtier_benchmark