



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2017년05월30일
(11) 등록번호 10-1740962
(24) 등록일자 2017년05월23일

(51) 국제특허분류(Int. Cl.)
G06F 17/30 (2006.01) G06F 3/06 (2006.01)
G06F 9/455 (2006.01)
(52) CPC특허분류
G06F 17/302 (2013.01)
G06F 17/3012 (2013.01)
(21) 출원번호 10-2016-0035398
(22) 출원일자 2016년03월24일
심사청구일자 2016년03월24일
(56) 선행기술조사문헌
JP2015035034 A
KR1020140055451 A*
KR1020120065072 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
유혁
서울특별시 강남구 광평로39길 3, 801동 101호(수서동, 동익아파트)
박준영
서울특별시 강남구 역삼로81길 18-2 B동 502호 (대치동, 한스타운)
(74) 대리인
김등용, 김홍석

전체 청구항 수 : 총 6 항

심사관 : 홍경아

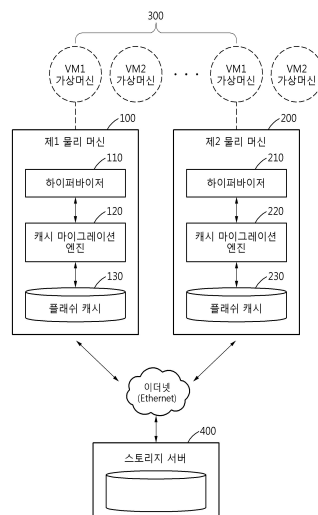
(54) 발명의 명칭 가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 시스템 및 그 관리 방법

(57) 요약

본 발명의 가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 시스템은, 네트워크로 연결된 클라우드 환경의 스토리지 서버를 포함하는 클라우드 스토리지 시스템에 있어서, 플래시 캐시를 모니터링 하여 가상 머신의 마이그레이션 실행시 캐시 메타 데이터를 전송하는 제1 물리 머신; 및 상기 제 1 물리 머신으로부터 캐시 메타 데이터를 수신하여 캐시 리스트를 재구성하는 제 2 물리 머신을 포함한다.

본 발명은 각 호스트 서버의 플래시 캐시를 모니터링 하고, 가상 머신 마이그레이션이 발생하면 유휴된 플래시 캐시의 메타 데이터 정보를 새로운 호스트 서버로 이동시켜 쿨드한 플래시 캐시를 가장 핫한 캐시 블록부터 재구성함으로써 핫 스타트를 달성할 수 있다.

대표도 - 도2



(52) CPC특허분류

G06F 3/067 (2013.01)

G06F 9/45504 (2013.01)

(72) 발명자

박현찬

서울특별시 성북구 안암동2가 401호

고현석

경기도 성남시 분당구 미금로 114 (구미동, 하얀마을그랜드빌라) 303동 102호

이 발명을 지원한 국가연구개발사업

과제고유번호 R01261510660001002

부처명 미래창조과학부

연구관리전문기관 정보통신기술진흥센터

연구사업명 정보통신방송연구개발사업

연구과제명 (SW 스타랩) 성능 및 보안 SLA 보장이 가능한 차세대 클라우드 인프라SW 개발

기 여 율 1/2

주관기관 고려대학교 산학협력단

연구기간 2015.03.01 ~ 2016.02.29

이 발명을 지원한 국가연구개발사업

과제고유번호 B01261510460001002

부처명 미래창조과학부

연구관리전문기관 정보통신기술진흥센터

연구사업명 정보통신방송연구개발사업

연구과제명 SDN 2.0 실현을 위한 네트워크 가상화 플랫폼 핵심 기술 및 서비스 연구

기 여 율 1/2

주관기관 고려대학교 산학협력단

연구기간 2015.03.01 ~ 2016.02.29

공지예외적용 : 있음

명세서

청구범위

청구항 1

네트워크로 연결된 클라우드 환경의 스토리지 서버를 포함하는 클라우드 스토리지 시스템에 있어서,

플래시 캐시를 모니터링 하여 가상 머신의 마이그레이션 실행시 캐시 메타 데이터를 전송하는 제 1 물리 머신; 및

상기 제 1 물리 머신으로부터 캐시 메타 데이터를 수신하여 캐시 리스트를 재구성하는 제 2 물리 머신을 포함하고,

상기 제 1 물리 머신은

데이터 캐시 블록을 저장하는 플래시 캐시;

상기 가상 머신의 마이그레이션을 실행하는 하이퍼바이저; 및

상기 플래시 캐시를 모니터링 하여 상기 가상 머신에 대해 마이그레이션이 실행되면 상기 플래시 캐시의 메타 데이터를 전송하는 캐시 마이그레이션 엔진을 포함하고,

상기 캐시 마이그레이션 엔진은

핫 캐시 리스트를 요청하는 핫 캐시 메신저;

상기 플래시 캐시의 캐시 블록을 핫/콜드 캐시로 구별하여 핫한 캐시 블록 부터 각 캐시 블록이 저장하고 있는 LBA(Logical Block Address)를 정렬하여 생성된 핫 캐시 리스트 및 모니터링 된 적중률을 포함한 메타 데이터를 제공하는 핫 캐시 모니터를 포함하는,

가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 시스템.

청구항 2

삭제

청구항 3

삭제

청구항 4

제1항에 있어서,

상기 제 2 물리 머신은

데이터 캐시 블록을 저장하는 제2 플래시 캐시;

상기 가상 머신의 마이그레이션을 실행하는 제2 하이퍼바이저; 및

상기 제 1 물리 머신으로부터 전송된 메타 데이터를 수신하여 핫 캐시 리스트를 재구성하는 제2 캐시 마이그레이션 엔진을 포함하는 가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 시스템.

청구항 5

네트워크로 연결된 클라우드 환경의 스토리지 서버를 포함하는 클라우드 스토리지 시스템에서,

전송 측 물리 머신의 플래시 캐시 상태를 모니터링하는 단계;

가상 머신의 마이그레이션이 발생하면, 상기 모니터링 된 플래시 캐시의 메타 데이터를 전송하는 단계; 및
 상기 전송된 메타 데이터를 수신하여 수신 측 물리 머신의 플래시 캐시를 재구성하는 단계를 포함하고,
 상기 메타 데이터는 상기 플래시 캐시의 캐시 블록을 핫/콜드 캐시로 구별하여 핫한 캐시 블록부터 각 캐시 블록이 저장하고 있는 LBA(Logical Block Address)를 정렬하여 생성된 핫 캐시 리스트 및 모니터링 된 적중률을 포함하는,
 가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 관리 방법.

청구항 6

삭제

청구항 7

제5항에 있어서,

상기 메타 데이터를 전송하는 단계 이후, 상기 전송 측 물리 머신은 상기 수신 측 물리 머신으로부터 전송 완료 응답을 받으면 상기 전송 측 물리 머신의 플래시 캐시의 메타 데이터를 삭제하는 단계를 더 포함하는 스토리지 성능 향상을 위한 클라우드 스토리지 관리 방법.

청구항 8

네트워크로 연결된 클라우드 환경의 스토리지 서버를 포함하는 클라우드 스토리지 시스템에서,

전송 측 물리 머신의 플래시 캐시 상태를 모니터링하는 단계;

가상 머신의 마이그레이션이 발생하면, 상기 모니터링 된 플래시 캐시의 메타 데이터를 전송하는 단계; 및

상기 전송된 메타 데이터를 수신하여 수신 측 물리 머신의 플래시 캐시를 재구성하는 단계를 포함하고,

상기 수신 측 물리 머신에서 플래시 캐시를 재구성하는 단계에서 재구성하려는 캐시 블록이 이미 수신 측 물리 서버에서 발생한 캐시적중실패로 인해 재구성 되었을 경우, 해당 블록을 제외하여 재구축을 실행하는,

가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 관리 방법.

청구항 9

네트워크로 연결된 클라우드 환경의 스토리지 서버를 포함하는 클라우드 스토리지 시스템에서,

전송 측 물리 머신의 플래시 캐시 상태를 모니터링하는 단계;

가상 머신의 마이그레이션이 발생하면, 상기 모니터링 된 플래시 캐시의 메타 데이터를 전송하는 단계; 및

상기 전송된 메타 데이터를 수신하여 수신 측 물리 머신의 플래시 캐시를 재구성하는 단계를 포함하고,

상기 수신 측 물리 머신에서 플래시 캐시를 재구성하는 단계에서 상기 마이그레이션 전 마지막으로 모니터링 된 캐시 적중률을 달성하거나 비용-이득 분석을 통해 캐시 재구성 비용이 캐시 적중실패 비용보다 커지면 재구성 실행을 중단하는,

스토리지 성능 향상을 위한 클라우드 스토리지 관리 방법.

발명의 설명

기술 분야

본 발명은 가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 시스템 및 그 관리 방법에 관한 것으로, 특히 각 호스트 서버의 캐시 마이그레이션 엔진을 구성하여 호스트 플래시 캐시를 모니터링하고, 가상 머신의

[0001]

마이그레이션 발생시 워업된 플래시 캐시 엔트리 정보를 새로운 호스트 서버로 이동시켜 이를 재구축하여 핫 스타트를 할 수 있는 가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 시스템 및 그 관리 방법에 관한 것이다.

배경 기술

- [0002] 최근 스마트폰 등의 다양한 IT 콘텐츠 생산 기기들이 널리 사용되고 있으며 이를 이용해 생산된 데이터들이 주로 클라우드에 저장되면서, 클라우드 데이터 센터의 데이터의 양이 급증하고 있다. 이에 따라 클라우드에서 스토리지의 역할이 더욱 더 중요해지고 있으며, 특히 복잡하고 다양해지는 다수 사용자의 요청을 고품질의 일관된 성능으로 처리하는 것이 중요하다.
- [0003] 클라우드 스토리지 시스템에서는 주로 네트워크로 연결된 공유 스토리지 시스템(shared storage system)을 구축하여 클라우드의 관리성, 확장성, 가용성을 높인다. 사용자의 가상 머신 이미지는 공유 스토리지 서버에 저장되어 있으며, 실행 시 할당받은 호스트 서버에서 네트워크를 통해 공유 스토리지에 접근하여 작업을 수행한다. 이때, 공유 스토리지로의 접근은 네트워크 자원, 스토리지 자원에 대한 다수 사용자의 경쟁으로 인해 성능이 크게 낮거나 유동적일 수 있다. 이러한 문제를 해결하기 위해 각 물리 머신은 호스트 캐시 스토리지를 사용할 수 있다. 그리고 이때 SSD(solid-state drives)와 같은 빠른 성능의 대용량 플래시 기반 스토리지를 널리 사용한다.
- [0004] 도 1은 일반적인 클라우드 시스템의 구성을 개략적으로 도시한 도면이다. 도 1에 도시된 바와 같이, 클라우드 시스템은, 플래시 캐시를 포함한 호스트 서버(10, 20), 각 호스트 서버(10, 20)의 가상 서버(30) 및 각 호스트 서버(10, 20)와 네트워크로 연결된 공유된 스토리지 서버(40)를 포함하여 구성한다.
- [0005] 이러한, 클라우드 시스템은 공유 스토리지를 단독으로 사용할 때 보다 호스트 플래시 캐시를 사용할 때 비교적 높고 일관된 성능을 제공할 수 있다. 그러나 클라우드 관리자는 성능 최적화, 저전력 운용, 로드밸런싱과 같은 자원 관리를 위해 실행되고 있는 가상 머신을 다른 호스트 서버로 옮기는 가상 머신 마이그레이션을 수시로 수행한다. 이때, 기존 클라우드 관리 기법에서는 가상 머신 마이그레이션을 수행할 때 플래시 캐시에 대한 고려를 전혀 하지 않아 심각한 스토리지 성능 저하가 발생한다. 가상 머신 마이그레이션은 클라우드 관리자의 필요에 의해 실행되기 때문에 클라우드 사용자는 마이그레이션을 예상할 수 없을뿐더러 마이그레이션으로 인한 가상 머신의 성능 저하를 원하지 않는다.
- [0006] 클라우드 사용자의 가상 머신이 마이그레이션 되면 기존 호스트 서버에 워업(warm-up)된 캐시를 잃고 새로운 서버에서 콜드 스타트(cold-start)를 하게 된다. 결국, 사용자의 가상 머신은 다시 한 번 플래시 캐시를 워업하는 시간이 소요됨에 따라 스토리지 성능 저하가 발생한다. 이는 클라우드 사용자가 직접적으로 체감하는 클라우드 서비스의 품질을 저하시키는 원인이 된다.

발명의 내용

해결하려는 과제

- [0007] 본 발명은 각 호스트 서버의 캐시 마이그레이션 엔진을 구성하여 호스트 플래시 캐시를 모니터링하고, 가상 머신의 마이그레이션 발생시 워업된 플래시 캐시 엔트리 정보를 새로운 호스트 서버로 이동시켜 이를 재구축하여 핫 스타트를 할 수 있는 가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 시스템 및 그 관리 방법을 제공하는 것을 목적으로 한다.
- [0008] 또한, 본 발명은 클라우드 상태에서 따라 플래시 캐시를 재구축하는 비용-이득을 분석하여 유동적인 크기의 캐시 워업을 수행하는 가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 시스템 및 그 관리 방법 제공하는 것을 목적으로 한다.

과제의 해결 수단

- [0009] 상기의 과제를 달성하기 위한 본 발명의 일실시예에 따른 가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 시스템은, 네트워크로 연결된 클라우드 환경의 스토리지 서버를 포함하는 클라우드 스토리지 시스템에 있어서, 플래시 캐시를 모니터링 하여 가상 머신의 마이그레이션 실행시 캐시 메타 데이터를 전송하는 제 1 물리 머신; 및 상기 제 1 물리 머신으로부터 캐시 메타 데이터를 수신하여 캐시 리스트를 재구축하는 제 2 물리 머신을 포함하는 점에 그 특징이 있다.
- [0010] 여기서, 특히 상기 제 1 물리 머신은, 데이터 캐시 블록을 저장하는 플래시 캐시; 가상 머신의 마이그레이션을

실행하는 하이퍼바이저; 및 상기 플래시 캐시를 모니터링 하여 상기 가상 머신에 대해 마이그레이션이 실행되면 상기 플래시 캐시의 메타 데이터를 전송하는 캐시 마이그레이션 엔진을 포함하는 점에 그 특징이 있다.

[0011] 여기서, 특히 상기 캐시 마이그레이션 엔진은 핫 캐시 리스트를 요청하는 핫 캐시 메신저; 상기 플래시 캐시의 캐시 블록을 핫/콜드 캐시로 구별하여 핫한 캐시 블록부터 각 캐시 블록이 저장하고 있는 LBA(Logical Block Address)를 정렬하여 생성된 핫 캐시 리스트 및 모니터링 된 적중률을 포함한 메타 데이터를 제공하는 핫 캐시 모니터를 포함하는 점에 그 특징이 있다.

[0012] 여기서, 특히 상기 제 2 물리 머신은 데이터 캐시 블록을 저장하는 플래시 캐시; 가상 머신의 마이그레이션을 실행하는 하이퍼바이저; 및 상기 제 1 물리 머신으로부터 전송된 메타 데이터를 수신하여 핫 캐시 리스트를 재구성하는 캐시 마이그레이션 엔진을 포함하는 점에 그 특징이 있다.

[0013] 또한, 상기의 과제를 달성하기 위한 본 발명의 또 다른 실시예에 따른 가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 관리 방법은, 네트워크로 연결된 클라우드 환경의 스토리지 서버를 포함하는 클라우드 스토리지 시스템에서, 전송 측 물리 머신의 플래시 캐시 상태를 모니터링 하는 단계; 가상 머신의 마이그레이션이 발생하면, 상기 모니터링 된 플래시 캐시의 메타 데이터를 전송하는 단계; 및 상기 전송된 메타 데이터를 수신하여 이를 이용하여 수신 측 물리 머신의 플래시 캐시를 재구성하는 단계를 포함하는 점에 그 특징이 있다.

[0014] 여기서, 특히 상기 메타 데이터는 상기 플래시 캐시의 캐시 블록을 핫/콜드 캐시로 구별하여 핫한 캐시 블록부터 각 캐시 블록이 저장하고 있는 LBA(Logical Block Address)를 정렬하여 생성된 핫 캐시 리스트 및 모니터링 된 적중률을 포함하는 점에 그 특징이 있다.

[0015] 여기서, 특히 상기 메타 데이터를 전송하는 단계 이후, 상기 전송 측 물리 머신은 상기 수신 측 물리 머신으로부터 전송 완료 응답을 받으면 상기 전송 측 물리 머신의 플래시 캐시의 메타 데이터를 삭제하는 단계를 더 포함하는 점에 그 특징이 있다.

[0016] 여기서, 특히 상기 수신 측 물리 머신에서 플래시 캐시를 재구성하는 단계에서 재구성하려는 캐시 블록이 이미 수신 측 물리 서버에서 발생한 캐시적중실패로 인해 재구축 되었을 경우, 해당 블록을 제외하여 재구축을 실행하는 점에 그 특징이 있다.

[0017] 여기서, 특히 상기 수신 측 물리 머신에서 플래시 캐시를 재구성하는 단계에서 상기 마이그레이션 전 마지막으로 모니터링 된 캐시 적중률을 달성하거나 비용-이득 분석을 통해 캐시 재구축 비용이 캐시 적중실패 비용보다 커지면 재구축 실행을 중단하는 점에 그 특징이 있다.

발명의 효과

[0018] 본 발명의 효과는 다음과 같다.

[0019] 첫째, 각 호스트 서버의 플래시 캐시를 모니터링 하고, 가상 머신 마이그레이션이 발생하면 워업된 플래시 캐시의 메타 데이터 정보를 새로운 호스트 서버로 이동시켜 콜드한 플래시 캐시를 가장 핫한 캐시 블록부터 재구축함으로써 핫 스타트를 달성할 수 있다.

[0020] 둘째, 플래시 캐시를 재구축하는 비용이 콜드 스타트로 인한 손해보다 커지는 것에 대한 비용-이득을 통해 자동으로 이윤배반(trade-off) 구간을 분석하고 그에 따라 유동적인 크기의 캐시 워업을 수행할 수 있다.

도면의 간단한 설명

[0021] 도 1은 일반적인 클라우드 시스템의 구성을 개략적으로 도시한 도면.

도 2는 본 발명의 실시예에 따른 가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 시스템의 구성을 개략적으로 도시한 도면.

도 3은 상기 도 2의 전송 측 제 1 물리 머신의 캐시 마이그레이션 엔진의 구성을 개략적으로 도시한 도면.

도 4는 상기 도 2의 수신 측 제 2 물리 머신의 캐시 마이그레이션 엔진의 구성을 개략적으로 도시한 도면.

도 5는 본 발명의 또 다른 실시예에 따른 가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 관리 방법에 관한 순서도.

발명을 실시하기 위한 구체적인 내용

- [0022] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시 예를 가질 수 있는 바, 특정 실시 예들을 도면에 예시하고 이를 상세한 설명을 통해 상세히 설명하고자 한다. 그러나 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.
- [0023] 본 발명을 설명함에 있어서, 관련된 공지 기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다. 또한, 본 명세서의 설명 과정에서 이용되는 숫자(예를 들어, 제1, 제2 등)는 하나의 구성요소를 다른 구성요소와 구분하기 위한 식별기호에 불과하다.
- [0024] 또한, 본 명세서에서, 일 구성요소가 다른 구성요소와 "연결된다" 거나 "접속된다" 등으로 언급된 때에는, 상기 일 구성요소가 상기 다른 구성요소와 직접 연결되거나 또는 직접 접속될 수도 있지만, 특별히 반대되는 기재가 존재하지 않는 이상, 중간에 또 다른 구성요소를 매개하여 연결되거나 또는 접속될 수도 있다고 이해되어야 할 것이다.
- [0025] 이하, 첨부된 도면들을 참조하여 본 발명의 가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 시스템 및 그 관리 방법에 관하여 상세히 설명한다.
- [0026] 먼저, 도 2는 본 발명의 일 실시 예에 따른 가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 시스템의 구성을 개략적으로 도시한 도면이다. 도 2에 도시된 바와 같이, 본 발명의 일 실시 예에 따른 가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 시스템은, 플래시 캐시를 모니터링 하여 가상 머신의 마이그레이션 실행시 캐시 메타 데이터를 전송하는 제 1 물리 머신(100), 상기 제 1 물리 머신(100)으로부터 캐시 메타 데이터를 수신하여 캐시 리스트를 재구성하는 제 2 물리 머신(200), 상기 제 1, 제 2 물리 머신(100, 200)에서 각각 생성된 가상 머신들(300) 및 상기 제 1 물리 머신(100) 및 상기 제 2 물리 머신(200)을 스토리지 네트워크, 예컨대 이더넷(Ethernet)으로 연결하여 데이터를 저장하는 공유 스토리지 서버(400)를 포함하여 구성된다.
- [0027] 상기 제 1 물리 머신(100) 및 상기 제 2 물리 머신(200)은 각 호스트 서버에 해당하는 것으로, 상기 제 1 물리 머신(100)은 데이터 캐시 블록을 저장하는 플래시 캐시(130), 가상 머신의 마이그레이션을 실행하는 하이퍼바이저(110), 및 상기 플래시 캐시(130)를 모니터링 하여 상기 가상 머신에 대한 마이그레이션이 실행되면 상기 플래시 캐시(130)의 메타 데이터를 전송하는 캐시 마이그레이션 엔진(120)을 포함하여 구성되고, 상기 제 2 물리 머신(200)은 데이터 캐시 블록을 저장하는 플래시 캐시(230), 가상 머신을 마이그레이션 실행하는 하이퍼바이저(210) 및 상기 제 1 물리 머신(100)으로부터 전송된 메타 데이터를 수신하여 핫 캐시 리스트를 재구성하는 캐시 마이그레이션 엔진(230)을 포함하여 구성된다. 이때, 상기 플래시 캐시(130)의 메타 데이터는 마이그레이션과 무관하게 생성 및/또는 관리되므로 추가적인 생성 동작은 수행되지 않을 수 있다.
- [0028] 상기 하이퍼바이저(Hypervisor)(110, 210)는 가상화 시스템을 구축하기 위한 소프트웨어 계층으로 운영체제와 하드웨어 사이에 존재하며, 각 가상 머신에게 논리적으로 분리된 하드웨어를 제공하게 된다. 즉, 상기 하이퍼바이저(110, 210)는 가상 머신 모니터(VMM:virtual machine monitor)로 명명될 수도 있으며 가상 머신을 생성하고 실행하는 펌웨어 또는 소프트웨어를 지칭할 수 있다.
- [0029] 보다 구체적으로, 상기 하이퍼바이저(110, 210) 상에는 적어도 하나의 가상 머신(300)이 실행될 수 있다. 가상 머신(VM1, VM2)은 운영 체제 계층 및 어플리케이션 계층을 포함할 수 있다. 여기에서, 각 가상 머신에 대한 운영 체제 계층은 서로에 대하여 독립적일 수 있다. 즉, 하이퍼바이저(110, 210)는 하드웨어 계층의 자원을 가상 머신(VM1, VM2)에게 분배하고, 서로에 대하여 독립적으로 동작할 수 있도록, 자원 분배를 수행할 수 있다. 또한, 하이퍼바이저(110, 210)는 디스크 이미지를 이용하여 각 가상 머신으로부터의 데이터 입출력을 처리할 수 있다.
- [0030] 도 3은 상기 도 2의 전송 측 제 1 물리 머신의 캐시 마이그레이션 엔진의 구성을 개략적으로 도시한 도면이다. 도 3에 도시된 바와 같이, 상기 캐시 마이그레이션 엔진(120)은 핫 캐시 리스트를 요청하는 핫 캐시 메신저(123); 상기 플래시 캐시(130)의 캐시 블록을 핫/콜드 캐시로 구별하여 핫한 캐시 블록부터 각 캐시 블록이 저장하고 있는 LBA(Logical Block Address)를 정렬하여 생성된 핫 캐시 리스트 및 모니터링 된 적중률을 포함한 메타 데이터를 제공하는 핫 캐시 모니터(121)를 포함하여 구성된다.
- [0031] 상기 핫 캐시 모니터(121)는 호스트 서버의 운영체제 내부에서 캐시 마이그레이션에 필요한 플래시 캐시의 상태를 모니터링하게 된다. 보다 구체적으로, 모니터링을 통해 모든 플래시 캐시 블록을 가장 핫한 블록부터 콜드한 블록까지 구분하게 된다. 즉, 상기 핫 캐시 모니터(121)는 핫 캐시 전달자(123)로부터 핫 캐시 리스트(122)를 요청받으면 가장 핫한 캐시블록부터 각 캐시 블록이 저장하고 있는 LBA(Logical Block Address)를 나열하여 전

송한다.

- [0032] 여기서, 핫/콜드 캐시 구별 방법에는 가장 최근에 접근한 캐시 블록을 가장 핫한 캐시로 판별하는 LRU(Least Recently Used) 방식 및 가장 빈번하게 접근한 캐시 블록을 가장 핫한 캐시로 판별하는 LFU(Least Frequently Used) 방식이 있으며, 시스템에서 운영하는 관리자가 시스템에서 동작하는 작업에 특성에 맞는 방식을 적절하게 선택할 수 있다. 이때 플래시 캐시를 관리하는 소프트웨어는 콜드 캐시를 구별하여 효율적인 캐시 교체(cache replacement)를 하기 위해 핫/콜드 캐시를 구별하는 기능을 포함하고 있기 때문에 상기 핫/콜드 캐시 구별을 위해 추가적인 오버헤드가 크게 필요치 않게 된다. 따라서, 핫 캐시 모니터(121)는 핫/콜드 구분을 위해 추가해야 할 기능은 없으며 단순히 핫 캐시 리스트를 요청받으면 이전 자료구조를 이용하여 핫-콜드 순으로 데이터의 LBA를 정렬하여 반환하면 된다.
- [0033] 핫 캐시 메신저(123)는 호스트 서버의 역할에 따라 전송자(sender)와 수신자(receiver)로 구분된다. 즉, 가상 머신이 이전에 수행되던 호스트 서버인 제 1 물리 머신(100)에서는 전송자가 동작하고, 마이그레이션으로 옮겨 갈 호스트 서버인 제 2 물리 머신(200)의 즉 후술될 핫 캐시 메신저는 수신자로 동작한다.
- [0034] 상기 핫 캐시 메신저(123)(전송자)가 상기 핫 캐시 메신저(223)(수신자)에게 전달하는 데이터는 핫 캐시 모니터(121)가 관리하고 있던 플래시 캐시(130)의 메타데이터, 즉 핫 캐시 리스트(122)와 마지막으로 모니터링 된 적 중물이다.
- [0035] 여기서, 상기 핫 캐시 리스트(122)는 실제 캐시 블록이 아닌 각 캐시 블록이 저장하고 있는 데이터의 LBA(Logical Block Address)로서 그 크기 크지 않아 전송 오버헤드가 작다. 예를 들어 20GB의 플래시 캐시에서 4KB 단위로 캐시 블록이 구성된 경우 핫 캐시 리스트의 최대 크기는 20MB(4B x 20GB/4KB)이다. 상기 핫 캐시 메신저(123)(전송자)는 핫 캐시 리스트(122)를 전송한 후, 핫 캐시 메신저(223)(수신자)로부터 응답을 받으면 플래시 캐시와 관련 메타 데이터를 서버에서 삭제한다.
- [0036] 도 4는 상기 도 2의 수신 측 제 2 물리 머신의 캐시 마이그레이션 엔진의 구성을 개략적으로 도시한 도면이다. 도 4에 도시된 바와 같이, 상기 캐시 마이그레이션 엔진(220)은 상기 제1 물리 머신으로부터 핫 캐시 리스트를 포함하는 메타 데이터를 수신하는 핫 캐시 메신저(223); 상기 수신된 메타 데이터의 핫 캐시 리스트(222)를 재구축하는 핫 캐시 모니터(221)를 포함하여 구성된다.
- [0037] 상기 핫 캐시 메신저(223)(수신자)는 가상 머신 마이그레이션으로 인해 캐시를 재구축해야 할 때는 상기 전달받은 핫 캐시 리스트(122)를 이용하여 가장 핫한 캐시 블록부터 캐시가 재구축 될 수 있도록 한다. 여기서, 상기 가상 머신 마이그레이션(virtual machine migration)은 하나의 물리적 서버에서 다른 물리적 서버로 현재 동작하고 있는 가상 머신을 이동시키는 기술이다. 가상 머신 마이그레이션을 통해 다수의 낮은 활용률의 물리적 서버에 공간을 많이 차지하는 것을 막기 위한 접근 방식, 효율적인 자원 운용이 가능하여 비용을 절감할 수 있다.
- [0038] 그리고 만약 재구축하려는 캐시 블록이 이전 제 1 물리 머신(110)인 호스트 서버에서 발생한 캐시 적중실패로 인해 재구축 되었을 경우, 해당 블록을 제외하여 재구축한다. 핫 캐시 리스트를 전송받은 상기 핫 캐시 메신저(223)(수신자)가 상기 핫 캐시 메신저(123)(전송자)에게 응답을 보내기 전에는 핫 캐시 리스트의 캐시 블록들은 이전 호스트 서버와 스토리지 서버(400)에 동시에 존재하게 된다. 이때, 각 서버와의 네트워크 연결 상태가 다를 수 있어 네트워크 상태를 고려하여 원하는 서버에서 캐시 블록을 읽어 오도록 할 수 있다.
- [0039] 상기 핫 캐시 메신저(223)(수신자)는 마이그레이션 전 마지막으로 모니터링 된 캐시 적중률을 달성하거나 비용-이득 분석을 통해 캐시 재구축 비용이 오히려 캐시 적중실패 비용보다 커질 때 캐시 재구축을 멈춘다. 예를 들어, 워크로드의 WSS(Working Set Size)가 플래시 캐시의 크기보다 월등히 작고 지역성이 높은 워크로드인 경우 호스트 플래시 캐시로 얻을 수 있는 이득이 적으면 캐시 읽음이 빠르게 진행되기 때문에 추가적인 비용을 들여 모든 캐시 블록을 미리 재구축 할 필요가 없다. 따라서 이와 같은 경우에 비용-이득 분석을 통해 최적의 재구축 크기를 찾아 선택적으로 재구축을 실행할 것인지 캐시 읽음을 그대로 진행할 것인지 여부를 판단하여 처리하게 된다.
- [0040] 한편, 상기 캐시 마이그레이션 엔진(220)은 가상화 환경에서 작동하기 때문에 특정 가상 머신의 스토리지 성능을 높이기 위해 서버의 모든 리소스를 사용할 수 없다. 따라서 상기 핫 캐시 메신저(223)(수신자)는 캐시를 재구축할 때 사용하는 리소스를 조절하여 가상 머신의 다른 작업 또는 다른 가상 머신의 성능에 부정적인 영향을 미치지 않도록 수행한다.
- [0041] 상기 플래시 캐시(Host-side flash cache)(130,230)는 SSD(Solid-state drives)와 같이 빠른 성능의 대용량 플래시 기반 스토리지를 플래시 캐시로 사용하고, 또한 플래시 캐시는 가상 머신의 원격 공유 스토리지 서버의 캐

시로서 사용될 수 있다.

- [0042] 상기 가상 머신(virtual machine)(300)은 컴퓨팅 환경을 소프트웨어로 구현한 것으로, 물리적 서버를 다중화 하여 완전한 시스템 플랫폼을 제공하여 완전한 운영 체제의 실행이 가능하다.
- [0043] 상기 공유 스토리지 서버(shared storage server)(400)는 호스트 서버(100, 200)들과 네트워크로 연결하여 다수의 호스트 서버가 스토리지를 공유하고, 가상화된 서버에서 동작하는 가상 머신이 원격의 스토리지 서버에 연결되어 전송되는 데이터를 저장하게 된다.
- [0044] 또한, 도 5는 본 발명의 또 다른 실시 예에 따른 가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 관리 방법에 관한 순서도이다.
- [0045] 도 5에 도시된 바와 같이, 본 발명의 또 다른 실시 예에 따른 가상환경의 스토리지 성능 향상을 위한 클라우드 스토리지 관리 방법은, 먼저 네트워크로 연결된 클라우드 환경의 스토리지 서버를 포함하는 클라우드 스토리지 시스템에서 전송 측 제 1 물리 머신(100)의 플래시 캐시 상태를 모니터링하게 된다(S501).
- [0046] 보다 구체적으로, 상기 핫 캐시 모니터(121)는 호스트 서버의 운영체제 내부에서 캐시 마이그레이션에 필요한 플래시 캐시의 상태를 모니터링 하게 된다. 보다 구체적으로, 모니터링을 통해 모든 플래시 캐시 블록을 가장 핫한 블록부터 콜드한 블록까지 구분하게 된다.
- [0047] 여기서, 핫/콜드 캐시 구별 방법에는 가장 최근에 접근한 캐시 블록을 가장 핫한 캐시로 판별하는 LRU(Least Recently Used) 방식 및 가장 빈번하게 접근한 캐시 블록을 가장 핫한 캐시로 판별하는 LFU(Least Frequently Used) 방식이 있으며, 시스템에서 운영하는 관리자가 시스템에서 동작하는 작업에 특성에 맞는 방식을 적절하게 선택할 수 있다. 이때 플래시 캐시를 관리하는 소프트웨어는 콜드 캐시를 구별하여 효율적인 캐시 교체(cache replacement)를 하기 위해 핫/콜드 캐시를 구별하는 기능을 포함하고 있기 때문에 상기 핫/콜드 캐시 구별을 위해 추가적인 오버헤드가 크게 필요치 않게 된다. 따라서 핫 캐시 모니터(121)는 핫/콜드 구분을 위해 추가해야 할 기능은 없으며 단순히 핫 캐시 리스트를 요청받으면 이전 자료구조를 이용하여 핫-콜드 순으로 데이터의 LBA를 정렬하여 반환하면 된다.
- [0048] 그리고 가상 머신의 마이그레이션이 발생하면 클라우드 관리자는 수신 측 제 2 물리 머신(200)이 위치한 네트워크 주소, 예컨대 IP 주소를 알리게 된다(S502). 상기 가상 머신 마이그레이션(virtual machine migration)은 하나의 물리적 서버에서 다른 물리적 서버로 현재 동작하고 있는 가상 머신을 이동시키는 기술이다. 가상 머신 마이그레이션을 통해 다수의 낮은 활용률의 물리적 서버에 공간을 많이 차지하는 것을 막기 위한 접근 방식, 효율적인 자원 운용이 가능하여 비용을 절감할 수 있다.
- [0049] 그 다음, 전송 측 제 1 물리 머신(100)의 핫 캐시 메신저로부터 상기 핫 캐시 모니터에 모니터링 된 플래시 캐시 상태의 메타 데이터를 요청하면 이를 핫 캐시 모니터에서 핫 캐시 리스트를 포함하는 메타 데이터를 생성하게 된다(S503). 즉, 상기 핫 캐시 모니터(121)는 핫 캐시 전달자(123)로부터 핫 캐시 리스트(122)를 요청받으면 가장 핫한 캐시블록부터 각 캐시 블록이 저장하고 있는 LBA(Logical Block Address)를 나열하여 핫 캐시 리스트를 전송한다.
- [0050] 여기서, 상기 메타 데이터는 상기 플래시 캐시의 캐시 블록을 핫/콜드 캐시로 구별하여 핫한 캐시 블록부터 각 캐시 블록이 저장하고 있는 LBA(Logical Block Address)를 정렬하여 생성된 핫 캐시 리스트 및 모니터링 된 적중률을 포함한다.
- [0051] 이어서, 상기 생성된 메타 데이터를 전송하는 단계 이후, 상기 전송 측 물리 머신은 상기 수신 측 물리 머신으로부터 전송 완료 응답을 받으면 상기 전송 측 물리 머신의 플래시 캐시의 메타 데이터를 삭제하게 된다(S504).
- [0052] 그리고 상기 전송된 메타 데이터를 수신하여 이를 이용하여 수신 측 물리 머신의 플래시 캐시를 재구축을 실행하게 하게 된다(S505).
- [0053] 보다 구체적으로, 상기 핫 캐시 메신저(223)(수신자)는 가상 머신 마이그레이션으로 인해 캐시를 재구축해야 할 때는 상기 전달받은 핫 캐시 리스트(122)를 이용하여 가장 핫한 캐시 블록부터 캐시가 재구축 될 수 있도록 한다. 상기 재구축하려는 캐시 블록이 이전 제 1 물리 머신(110)인 호스트 서버에서 발생한 캐시 적중실패로 인해 재구축 되었을 경우, 해당 블록을 제외하여 재구축한다. 핫 캐시 리스트를 전송받은 상기 핫 캐시 메신저(223)(수신자)가 상기 핫 캐시 메신저(123)(전송자)에게 응답을 보내기 전에는 핫 캐시 리스트의 캐시 블록들은 이전 호스트 서버와 스토리지 서버(400)에 동시에 존재하게 된다. 이때, 각 서버와의 네트워크 연결 상태가 다를 수

있어 네트워크 상태를 고려하여 원하는 서버에서 캐시 블록을 읽어 오도록 할 수 있다.

[0054] 상기 핫 캐시 메신저(223)(수신자)는 마이그레이션 전 마지막으로 모니터링 된 캐시 적중률을 달성하거나 비용-이득 분석을 통해 캐시 재구축 비용이 오히려 캐시 적중실패 비용보다 커질 때 캐시 재구축을 멈춘다. 예를 들어, 워크로드의 WSS(Working Set Size)가 플래시 캐시의 크기보다 월등히 작고 지역성이 높은 워크로드인 경우 호스트 플래시 캐시로 얻을 수 있는 이득이 적으면 캐시 워업을 빠르게 진행되기 때문에 추가적인 비용을 들여 모든 캐시 블록을 미리 재구축 할 필요가 없다. 따라서 이와 같은 경우에 비용-이득 분석을 통해 최적의 재구축 크기를 찾아 선택적으로 재구축을 실행할 것인 캐시 워업을 그대로 진행할 것인지 여부를 판단하여 처리하게 된다.

[0055] 따라서 상기와 같은 실시 예에 의하면 각 물리 머신(호스트 서버)의 플래시 캐시를 모니터링하고, 가상 머신 마이그레이션이 발생하면 워업된 플래시 캐시의 메타 데이터 정보를 새로운 호스트 서버로 이동시켜 콜드한 플래시 캐시를 가장 핫한 캐시 블록부터 재구축함으로써 핫 스타트를 달성할 수 있다.

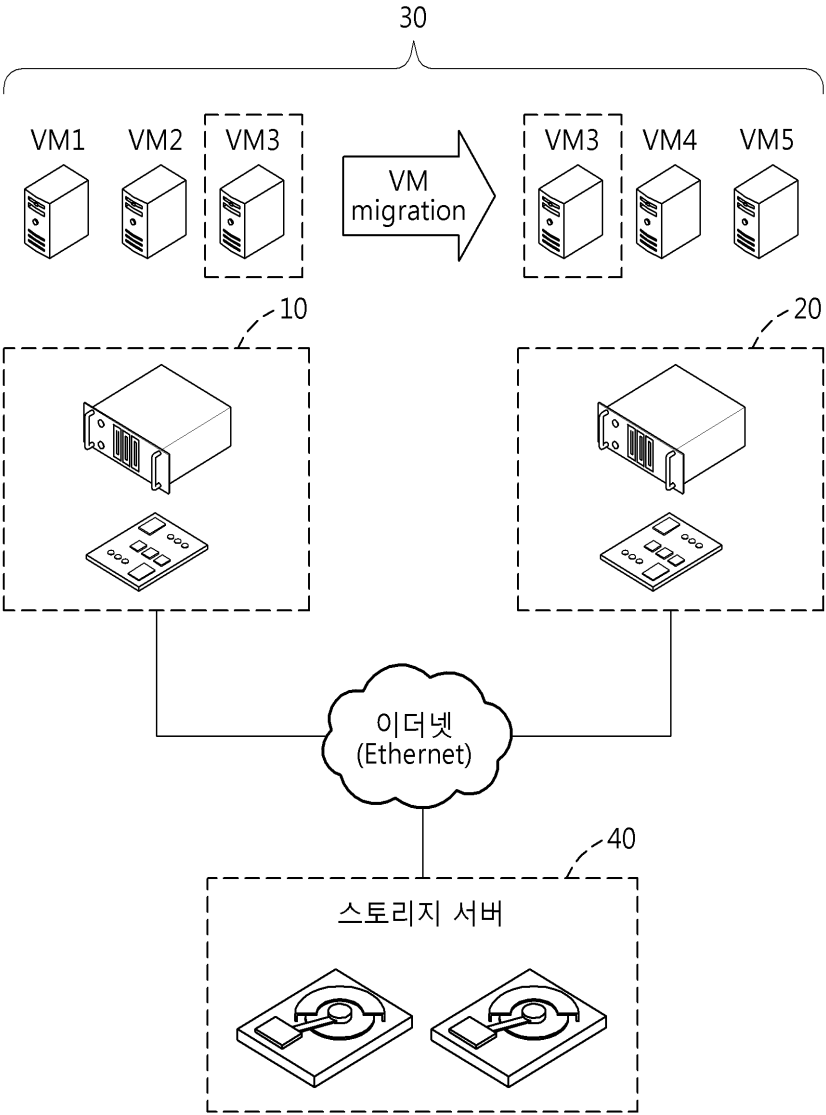
[0056] 본 발명의 권리범위는 상술한 실시 예에 한정되는 것이 아니라 첨부된 특허청구범위 내에서 다양한 형태의 실시 예로 구현될 수 있다. 특허청구범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 누구든지 변형 가능한 다양한 범위까지 본 발명의 청구범위 기재의 범위 내에 있는 것으로 본다.

부호의 설명

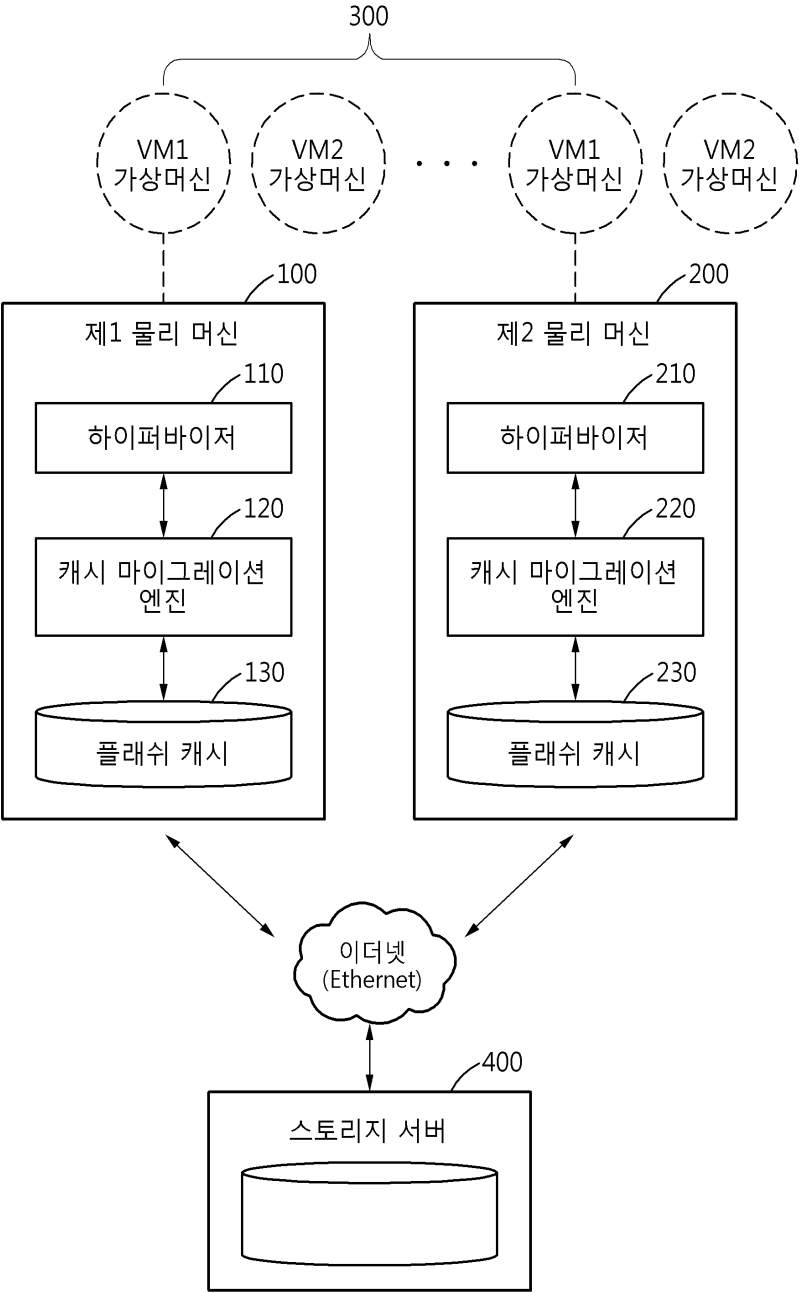
[0057] 100 --- 제 1 물리 머신
110, 210 --- 하이퍼바이저
120, 220 --- 캐시 마이그레이션 엔진
121, 221 --- 핫 캐시 모니터
122, 222 --- 핫 캐시 리스트
123, 223 --- 핫 캐시 메신저
130, 230 --- 플래시 캐시
200 --- 제 2 물리 머신
300 --- 가상 머신
400 --- 스토리지 서버

도면

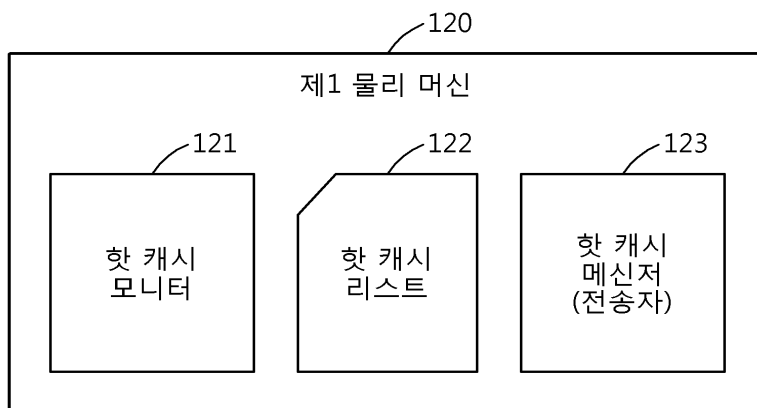
도면1



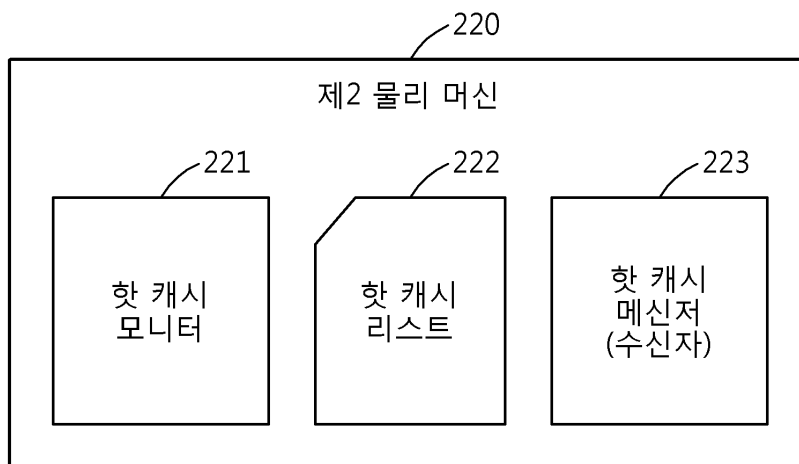
도면2



도면3



도면4



도면5

