(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2013-0088726
(43) 공개일자 2013년08월08일

(51) 국제특허분류(Int. Cl.)
G11C 7/10 (2006.01) *G06F 12/08* (2006.01)
(21) 출원번호 10-2012-0134041
(22) 출원일자 2012년11월23일
심사청구일자 없음
(30) 우선권주장
13/362,099 2012년01월31일 미국(US)

(71) 출원인
엘에스아이 코포레이션
미국 캘리포니아 95131, 새너제이, 라이더 파크 드라이브 1320
(72) 발명자
이시, 마크
미합중국, 30328 죠지아, 아틀란타, 스톤 밀 트레일 노스이스트 405
바데르딘니, 아난트
미합중국, 30043 죠지아, 로렌스빌, 햄프턴 우즈 웨이 1817
스머던, 게리 제이.
미합중국, 95070 캘리포니아, 사라토가, 피어스 로드 13850
(74) 대리인
특허법인오리진

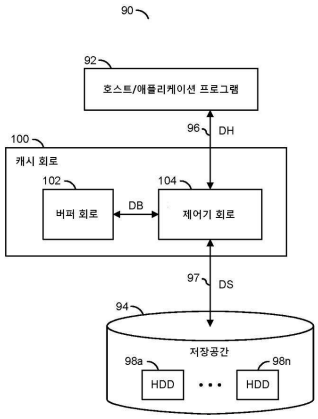
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 **중복 캐시 데이터의 탄력적 캐시**

(57) 요 약

중복 캐시 데이터의 탄력적 캐시처리 장치. 이 장치는 다수의 버퍼 및 회로를 포함할 수 있다. 이 회로는, (i) 기록 데이터를 저장공간에 저장하기 위한 기록 요청을 호스트로부터 수신하고, (ii) 상기 기록 요청에 연관된 중복화 조직을 기반으로 하여 상기 버퍼에 복수의 익스텐트를 할당하고, (iii) 상기 복수의 익스텐트에 상기 기록 데이터를 저장하도록 구성된다. 여기서, (a) 상기 각 익스텐트는 서로 다른 버퍼에 위치하고, (b) 상기 복수의 익스텐트는 상기 기록 요청에 응답하여 서로 동적으로 링크된다.

대 표 도 - 도1



특허청구의 범위

청구항 1

복수의 버퍼, 및

(i) 기록 데이터를 저장공간에 저장하기 위한 기록 요청을 호스트로부터 수신하고, (ii) 상기 기록 요청에 연관된 중복화 조직을 기반으로 하여 상기 버퍼에 복수의 익스텐트를 할당하고, (iii) 상기 복수의 익스텐트에 상기 기록 데이터를 저장하도록 구성된 회로를 포함하며,

(a) 상기 각 익스텐트는 서로 다른 버퍼에 위치하고, (b) 상기 복수의 익스텐트는 상기 기록 요청에 응답하여 서로 동적으로 링크되는 것을 특징으로 하는 중복 캐시 데이터의 탄력적 캐시 장치.

청구항 2

제1항에 있어서,

상기 회로는, 상기 버퍼로부터 상기 저장공간으로 상기 기록 데이터를 복사하는 것에 응답하여 상기 복수의 익스텐트 중 하나 이상을 링크해제하도록 되어 있는 것을 특징으로 하는 중복 캐시 데이터의 탄력적 캐시 장치.

청구항 3

제1항에 있어서,

상기 복수의 익스텐트는 상기 각 익스텐트에 해당되는 각 파라미터를 변경함으로써 동적으로 링크되는 것을 특징으로 하는 중복 캐시 데이터의 탄력적 캐시 장치.

청구항 4

제3항에 있어서,

상기 각 파라미터는 상기 복수의 익스텐트 중 다른 익스텐트를 지시하도록 변경되는 것을 특징으로 하는 중복 캐시 데이터의 탄력적 캐시 장치.

청구항 5

제1항에 있어서,

상기 회로는 상기 호스트로부터 수신한 다른 요청에 관련된 비중복화 조직에 기반하여 하나 이상의 익스텐트를 할당하도록 구성되는 것을 특징으로 하는 중복 캐시 데이터의 탄력적 캐시 장치.

청구항 6

제1항에 있어서,

(i) 상기 회로는 상기 호스트로부터 수신한 다른 기록 요청에 관련된 다른 중복화 조직에 기반하여 다른 복수의 익스텐트를 할당하도록 구성되고, (ii) 상기 다른 중복화 조직은 상기 중복화 조직과 상이한 것을 특징으로 하는 중복 캐시 데이터의 탄력적 캐시 장치.

청구항 7

제1항에 있어서,

상기 중복화 조직은 복수의 독립 디스크가 배열된 조직을 포함하는 것을 특징으로 하는 중복 캐시 데이터의 탄력적 캐시 장치.

청구항 8

제1항에 있어서,

상기 중복 캐시 데이터의 탄력적 캐시 장치는 하나 이상의 집적 회로로 구현되는 것을 특징으로 하는 중복 캐시

데이터의 탄력적 캐시 장치.

청구항 9

중복 캐시 데이터의 탄력적 캐시처리 방법으로서,

(A) 기록 데이터를 저장공간에 저장하기 위해 호스트로부터 장치로 기록 요청을 수신하는 단계,

(B) 상기 기록 요청에 연관된 중복화 조직을 기반으로 하여 복수의 버퍼에 복수의 익스텐트를 할당하는 단계,

(C) 상기 복수의 익스텐트에 상기 기록 데이터를 저장하고, (i) 상기 각 익스텐트는 서로 다른 버퍼에 위치하고, (ii) 상기 복수의 익스텐트는 상기 기록 요청에 응답하여 서로 동적으로 링크되는 단계를 포함하는 것을 특징으로 하는 중복 캐시 데이터의 탄력적 캐시처리 방법.

청구항 10

제9항에 있어서,

상기 기록 데이터를 상기 버퍼로부터 상기 저장공간으로 복사하는 것에 응답하여 상기 복수의 익스텐트 중 하나 이상을 링크해제하는 단계를 더 포함하는 것을 특징으로 하는 중복 캐시 데이터의 탄력적 캐시처리 방법.

청구항 11

제9항에 있어서,

상기 복수의 익스텐트는 상기 각 익스텐트에 해당되는 각 파라미터를 변경함으로써 동적으로 링크되는 것을 특징으로 하는 중복 캐시 데이터의 탄력적 캐시처리 방법.

청구항 12

제11항에 있어서,

상기 각 파라미터는 상기 복수의 익스텐트 중 다른 익스텐트를 지시하도록 변경되는 것을 특징으로 하는 중복 캐시 데이터의 탄력적 캐시처리 방법.

청구항 13

제9항에 있어서,

상기 호스트로부터 수신한 다른 요청에 관련된 비중복화 조직에 기반하여 하나 이상의 익스텐트를 할당하는 단계를 더 포함하는 것을 특징으로 하는 중복 캐시 데이터의 탄력적 캐시처리 방법.

청구항 14

제9항에 있어서,

상기 호스트로부터 수신한 다른 기록 요청에 관련된 다른 중복화 조직에 기반하여 다른 복수의 익스텐트를 할당하는 단계를 더 포함하고, 상기 다른 중복화 조직은 상기 중복화 조직과 상이한 것을 특징으로 하는 중복 캐시 데이터의 탄력적 캐시처리 방법.

청구항 15

제9항에 있어서,

상기 중복화 조직은 복수의 독립 디스크가 배열된 조직을 포함하는 것을 특징으로 하는 중복 캐시 데이터의 탄력적 캐시 방법.

명 세 서

기 술 분 야

[0001] 본 발명은 개괄적으로, 저장장치의 캐시 처리(caching)에 관한 것으로서, 보다 구체적으로는, 중복 캐시(redundant cache) 데이터의 탄력적 캐시(elastic cache)를 구현하는 방법 및/또는 장치에 관한 것이다.

배 경 기 술

[0002] 종래의 저장 시스템에서, 빈번하게 액세스되는 데이터를 캐시 처리함으로써 전체 시스템 및 애플리케이션 프로그램의 성능이 향상된다. 많은 애플리케이션에서, 캐시를 라이트백 모드(write-back mode)로 사용할 경우에는 작업부하상의 이득이 있고 성능이 향상된다. 라이트백 모드로 저장되는 캐시 데이터는 종종, 중복화(redundancy)에 의해 데이터 상실로부터 보호된다. 일반적으로 중복화는 하나의 캐시 제어기에 두 개의 플래시 모듈을 두는 것이다. 두 플래시 모듈을 미러화(mirroring) 하여 영구적으로 짝을 지음으로써, 하나의 플래시 모듈에 캐시 데이터를 보관하기 위한 사용가능 공간이 사실상 줄어들게 된다. 일부 호스트 데이터만 중복화로부터 이득을 받기 때문에 플래시 모듈 내의 사용가능한 공간의 사용이 비효율적으로 된다.

발명의 내용

해결하려는 과제

[0003] 중복 캐시 데이터의 탄력적 캐시를 구현하는 것이 바람직할 것이다.

과제의 해결 수단

[0004] 본 발명은 중복 캐시 데이터의 탄력적 캐시처리 장치에 관한 것이다. 본 발명은 다수의 버퍼(buffer) 및 회로를 포함할 수 있다. 이 회로는, (i) 기록(write) 데이터를 저장공간(storage volume)에 저장하기 위한 기록 요청(write request)을 호스트로부터 수신하고, (ii) 상기 기록 요청에 연관된 중복화 조직(redundancy organization)을 기반으로 하여 상기 버퍼에 복수의 익스텐트(extent)를 할당하고, (iii) 상기 복수의 익스텐트에 상기 기록 데이터를 저장하도록 구성된다. 여기서, (a) 상기 각 익스텐트는 서로 다른 버퍼에 위치하며, (b) 상기 복수의 익스텐트는 상기 기록 요청에 응답하여 서로 동적으로 링크된다.

발명의 효과

[0005] 본 발명의 목적, 특징, 장점은, (i) 효율적인 캐시 공간을 추가적으로 제공할 수 있고, (ii) 통상의 기법에 비해, 주어진 캐시 매체의 크기에 대해서 더 넓은 캐시 적용범위를 제공할 수 있고, (iii) 기존의 익스텐트 구조를 사용할 수 있고, (iv) 기존의 관리 인프라를 활용할 수 있으며, (v) 캐시 중복화에 있어서 다양한 RAID 수준을 구현할 수 있고, (vi) 통상의 기법에 비해, 주어진 호스트 입/출력 동작의 양에 대해서, 버퍼로의 기록 회수를 줄일 수 있고, (vii) 주어진 호스트 입/출력 동작에 대해서, 플래시 메모리 버퍼의 수명을 늘릴 수 있고, 그리고/또는 (viii) 캐시 데이터용의 대형 버퍼를 반도체 드라이버로 구현할 수 있는, 중복 캐시 데이터의 탄력적 캐시를 구현하는 방법 및/또는 장치를 제공하는 것이다.

도면의 간단한 설명

[0006] 본 발명의 상기 및 그 밖의 목적, 특징, 장점은 발명의 상세한 설명과 특허청구범위 그리고 다음과 같은 도면으로부터 명확해질 것이다.

도 1은 시스템의 예시적 구현을 나타내는 블록도이다.

도 2는 본 발명의 바람직한 실시예에 따른 시스템에서의 캐시 회로의 예시적 구현을 나타내는 블록도이다.

도 3은 익스텐트의 예시적 구조도이다.

도 4는 캐시 데이터의 예시도이다.

도 5는 몇 가지 익스텐트 그룹의 예시도이다.

도 6은 미러화된 탄력적인 캐시를 갖는 캐시 구조의 예시도이다.

도 7은 기록 요청을 처리하는 예시적 방법의 흐름도이다.

도 8은 독출 요청을 처리하는 예시적 방법의 흐름도이다.

도 9는 익스텐트를 할당하는 예시적 방법의 흐름도이다.

도 10은 익스텐트를 할당해제하는 예시적 방법의 흐름도이다.

도 11은 캐시 회로의 더티 기록 데이터를 탄력적으로 미러화함에 따른 공간 상의 이득을 나타내는 그래프이다.

발명을 실시하기 위한 구체적인 내용

[0007] 본 발명의 일부 실시예들은 캐시 공간의 기록(write) 부분에 대한 중복화(redundancy)를 제공하기 위하여, 사용가능한 캐시 공간을 탄력적으로 이용할 수 있다. 이 탄력 기법은 개괄적으로, 독출(read)한 데이터 및/또는 보호되지 않은(예컨대, 중복화하지 않은) 기록 데이터를 완충저장(buffering; 버퍼에 저장)하기 위하여 캐시 공간의 전체 사용가능한 부분을 확대하는 지능적 관리 기법으로써 구현된다. 또한, 탄력 기법은 다른 이득들, 예컨대, 데이터 집성(aggregation)(예컨대, 일부 데이터를 익스텐트(extent)를 기준으로 별도의 물리적 장치에서 미러화(mirror)함으로써, 저장소와 캐시가 다수의 노드들(서버) 간에 공유되는 경우에 이 미러화된 익스텐트의 각 부분이 개별 노드에 상주할 수 있게 됨) 및 정책 선택(policy selection)(예컨대, 일부 저장디스크는 캐시 데이터의 보호(또는 중복화)를 사용하고, 일부는 선택된 정책에 의거하여 캐시 데이터를 보호하지 않음)을 나타내도록 모델링할 수도 있다.

[0008] 도 1은 시스템(90)의 예시적 구현의 블록도이다. 이 시스템(또는 장치)(90)은 개괄적으로, 블록(또는 회로)(92), 블록(또는 회로)(94), 네트워크(또는 버스)(96), 네트워크(또는 버스)(97), 및 블록(또는 회로)(100)으로 구성된다. 회로(94)는 전반적으로 다수의 블록(또는 회로)(98a~98n)을 포함한다. 회로(100)는 개괄적으로, 블록(또는 회로)(102)과 블록(또는 회로)(104)을 포함한다. 회로(92~104)는 하드웨어, 소프트웨어, 하드웨어와 소프트웨어의 조합, 또는 그 밖의 구현을 통해 실시가능한 모듈 및/또는 블록으로 구성할 수 있다.

[0009] 양방향 신호(예컨대, DH)가 회로(92)와 회로(100) 사이에서 네트워크(96)를 통해 교환될 수 있다. 이 신호 DH는 회로(92)와 회로(100) 사이에서 데이터(예컨대, 독출 데이터 및/또는 기록 데이터), 요청, 및 상태 정보의 통신을 위해 사용될 수 있다. 양방향 신호(예컨대, DB)가 회로(102)와 회로(104) 사이에서 교환될 수 있다. 이 신호 DB는 회로(102)와 회로(104) 사이에서 캐시 데이터와 메타데이터(metadata)를 전달할 수 있다. 양방향 신호(예컨대, DS)가 회로(94)와 회로(100) 사이에서 네트워크(97)를 통해 교환될 수 있다. 이 신호 DS는 회로(100)와 회로(94) 사이에서 데이터 및 관련 명령어(예컨대, 독출 명령어, 기록 명령어, 구성 명령어)를 전달할 수 있다.

[0010] 회로(92)는 하나 이상의 호스트 회로(또는 서버) 및/또는 하나 이상의 소프트웨어 애플리케이션 프로그램으로 구현할 수 있다. 회로(92)는 개괄적으로, 신호 DH와 DS에서, 회로(100) 및 네트워크(96)를 통해 회로(94)로부터 데이터를 독출하고, 그리고/또는 회로(94)에 데이터를 기록하는 기능을 한다. 이 회로(92)는 또한, 요청(예컨대, 독출 요청, 기록 요청)과 함께 해당 주소 정보(예컨대, 논리 블록 주소(LBA))를 신호 DH로써 회로(100)로 보내는 기능을 할 수 있다. 또한, 이 회로(92)는 신호 DH를 이용해 명령어(예컨대, 사전 호출(pre-fetch), 플러시(flush), 구성(configure) 등)를 회로(100)로 전달할 수 있도록 기능한다. 신호 DH는 또한 회로(100)로부터 회로(92)로 상태 정보(예컨대, 캐시 히트(cache hit), 캐시 미스(cache miss) 등)를 전달하기 위하여 회로(100)에 의해 제어될 수 있다.

[0011] 회로(94)는 하나 이상의 저장공간(storage volume)으로 구현할 수 있다. 회로(94)는 개괄적으로, 기록 명령어에 대한 응답으로서, 신호 DS를 통해 회로(100)로부터 수신한 데이터를 저장하는 기능을 한다. 회로(94)는 또한, 독출 명령어에 응답하여, 신호 DS를 통해 회로(100)로 데이터를 전송하는 기능을 할 수 있다. 저장공간은 논리 공간, 가상 공간, 및/또는 물리 공간으로 구현할 수 있다. 회로(94)는 저장 영역 네트워크(SAN: storage area network), 네트워크 부착 저장소(NAS: network attached storage), 및/또는 디스크 어레이 서브시스템(DAS: disk array subsystem) 구성(아키텍처)의 일부일 수 있다.

[0012] 네트워크(96)는 하나 이상의 디지털 통신망 및/또는 버스로 구현할 수 있다. 네트워크(96)는 개괄적으로, 회로(92)와 회로(100) 사이의 통신을 수행한다. 네트워크(96)의 구현 수단으로는 하나 이상의 인터넷, 이더넷, 광섬유 네트워크, 무선 네트워크, 유선 네트워크, 무선주파수(RF) 통신망, 및/또는 백플레인(backplane) 버스 등을 포함할 수 있지만, 이들에 국한되는 것은 아니다.

[0013] 네트워크(97)는 하나 이상의 디지털 통신망 및/또는 버스로 구현할 수 있다. 네트워크(97)는 개괄적으로, 회로(94)와 회로(100) 사이의 통신을 수행한다. 네트워크(97)의 구현 수단으로는 하나 이상의 인터넷, 이더넷, 광섬유 네트워크, 무선 네트워크, 유선 네트워크, 무선주파수(RF) 통신망, 및/또는 백플레인(backplane) 버스 등을 포함할 수 있지만, 이들에 국한되는 것은 아니다.

[0014] 각 회로(98a~98n)는 대용량 저장 드라이브로 구현할 수 있다. 일부 실시예에서, 각 회로(98a~98n)는 하드디스

크 드라이브(즉, HDD)를 포함할 수 있다. 회로(98a~98n)는 개괄적으로, 신호 DS를 통해 회로(100)에 데이터를 저장하고 받는 기능을 한다. 회로(98a~98n)는 하나 이상의 저장공간으로 배열하여 동작시킬 수 있다. 일부 실시예에서는, 회로(98a~98n)를 하나 이상의 중복화된 복수의 독립 디스크들(예를 들어, RAID)의 형태로서 조직하여 동작시킬 수 있다. RAID 조직에는 RAID 0, RAID 1, RAID 2, RAID 3, RAID 4, RAID 5, RAID 6, RAID 7, RAID 10, RAID 53, 및/또는 RAID 0+1 조직을 포함할 수 있으나, 이들에만 국한되는 것은 아니다. 특정 애플리케이션의 기준을 만족하기 위하여 다른 수준의 RAID 조직을 구현할 수도 있다. 또한, 회로(98a~98n)에 의해서 서로 다른 다수의 RAID 조직들을 동시에 실현할 수도 있다.

[0015] 회로(100)는 캐시 회로로 구현할 수 있다. 회로(100)는 개괄적으로, 회로(92)와 회로(94) 사이에서 교환될 데이터를 캐시처리(caching)하는 기능을 한다. 캐시처리는 회로(92)로부터 회로(94)로 이동될 데이터를 기록하는 라이트백(write-back) 또는 라이트비하인드(write-behind) 정책으로 구현할 수 있다. 더티 기록 데이터(dirty write data)는, 다수의 매체들 간에 분산된 중복화 조직을 이용하여 회로(100)에서 완충저장할 수 있다. 중복화 조직 및 그 분포는, 단일 매체의 실패로 인해 실패한 매체의 기록 데이터의 일부 또는 전부가 액세스 불가 및/또는 상실된 이후에도, 더티 기록 데이터에 대한 액세스가 유지되도록 구성할 수 있다. 회로(100)는, 회로(92)와 회로(94) 내에 또는 중간 지점에 위치한 캐시로 구현할 수 있다. 일부 실시예에서, 회로(100)는 하나 이상의 집적회로로 구현할 수 있다.

[0016] 회로(102)는 다수의 버퍼 회로로 구현할 수 있다. 회로(102)는 회로(92)와 회로(94) 사이에서 전달될 캐시 데이터(또는 완충저장할 데이터)를 저장하도록 기능할 수 있다. 회로(102)는 또한, 캐시 데이터와 관련된 메타데이터를 저장하는 기능을 한다. 캐시 데이터는 익스텐트(extent) 내에 배치될 수 있다. 각 익스텐트는 특정 데이터 파일 또는 소프트웨어 프로그램을 위해 예약되어 있는 저장 공간의 연속적 블록을 대표할 수 있다. 메타데이터는 일반적으로, 익스텐트를 관리하기 위해 필요한 정보를 포함한다. 일부 실시예에서 회로(102)는 전적으로 회로(100) 내부에서 구현할 수도 있고(도시한 바와 같음), 전적으로 회로(100) 외부에서 구현할 수도 있고, 회로(100)의 내부 및 외부에서 부분적으로 구현할 수도 있고, 다수의 노드(또는 서버) 사이에 어그리게이트할 수도 있다.

[0017] 회로(104)는 제어기 장치로 구현할 수 있다. 회로(104)는 회로(92)로부터 기록 요청을 수신하여 회로(94)에 기록 데이터를 저장하고, 기록 요청과 연관된 중복화 조직에 기반하여 회로(102) 내의 다수의 버퍼에 복수의 익스텐트를 할당하고, 기록 데이터를 복수의 익스텐트에 저장하는 기능을 할 수 있다. 각 익스텐트는 통상, 서로 다른 버퍼 내에 위치한다. 복수의 익스텐트들은 기록 요청에 응답하여 서로 동적으로 링크될 수 있다. 회로(104)는 또한, 기록 데이터의 버퍼로부터 회로(94)로의 복사에 응답하여, 다수의 익스텐트 중 적어도 하나를 링크해제하는 기능을 할 수 있다. 일부 실시예에서 회로(104)는 하나 이상의 집적회로로 구현할 수 있다.

[0018] 도 2를 참조하면, 본 발명의 바람직한 실시예에 따른 회로(100)의 예시적 구현예의 블록도가 도시되어 있다. 회로(102)는 개괄적으로, 다수의 블록(또는 회로)(106a~106n)을 포함한다. 회로(104)는 개괄적으로, 블록(또는 모듈)(108), 블록(또는 모듈)(110), 블록(또는 모듈)(112)을 포함한다. 회로(106a~106n)는 하드웨어, 소프트웨어, 하드웨어와 소프트웨어의 조합, 또는 그 밖의 구현을 통해 실시가능한 모듈 및/또는 블록을 대표할 수 있다. 모듈(108~112)은 회로(104)를 제어하기 위해 실행시킬 수 있는 펌웨어(또는 소프트웨어, 또는 프로그램 명령, 또는 코드)로 구현될 수 있다.

[0019] 신호 DH가 인터페이스(114)를 통해 회로(104)에 의해 전송 및 수신될 수 있다. 신호 DS는 인터페이스(116)를 통해 회로(104)에 의해 전송 및 수신될 수 있다. 신호 DB는 다수의 신호(예컨대, DBa~DBn)로써 구현될 수 있다. 각 신호 DBa~DBn은 개괄적으로, 회로(104)의 인터페이스(118a~118n) 및 회로(106a~106n)의 대응하는 인터페이스(120a~120n) 사이의 통신을 설정한다.

[0020] 각 회로(106a~106n)는 버퍼 회로로 구현할 수 있다. 각 회로(106a~106n)는 캐시 데이터와 이에 관련된 메타데이터를 저장하는 기능을 할 수 있다. 일부 실시예에서, 각 회로(106a~106n)는 반도체 드라이브(예컨대, SSD)로 구현할 수 있다. 반도체 드라이브의 통상적인 크기는 1~2 테라바이트의 범위 내일 수 있다. 다른 실시예에서, 각 회로(106a~106n)는 플래시 드라이브로 구현할 수 있다. 플래시 드라이브의 통상적인 크기는 1~32 기가바이트의 범위 내일 수 있다. 또다른 실시예에서, 각 회로(106a~106n)는 이중 데이터속도(double data rate)(예컨대, DDR) 메모리 회로로 구현할 수 있다. DDR 메모리의 통상적인 크기는 1~64 기가바이트의 범위 내일 수 있다. 이는 특정 애플리케이션의 기준에 부합하기 위하여 다른 크기로 구현할 수도 있다. 회로(102) 내의 회로(106a~106n)의 개수는 회로(94) 내의 회로(98a~98n)의 개수보다 많거나, 같거나, 적을 수 있다.

[0021] 모듈(108)은 익스텐트 제어기로 구현할 수 있다. 모듈(108)은 회로(92)로부터 수신한 요청(예컨대, 기록 요청

및/또는 독출 요청)을 처리하기 위하여 회로(102) 내의 하나 이상의 익스텐트를 할당하고 할당해제하는 기능을 할 수 있다. 모듈(108)은 또한, 그룹화된 공간(grouped space)(예컨대, 독출 공간 및/또는 기록 공간)을 생성하기 위하여 회로(102) 내의 각 익스텐트에 관련된 메타데이터의 파라미터를 변경하는 기능을 할 수 있다. 이 그룹화는 중복화 보호 조직에 캐시 데이터를 수용할 수 있는 더 큰 메모리 공간(또는 슈퍼 익스텐트)을 형성하기 위하여 익스텐트들을 함께 링크함으로써 수행할 수 있다. 익스텐트들을 그룹화된 공간으로 할당(또는 링크)하는 것은 통상, 그룹 내의 다음 익스텐트를 지시(point)하기 위하여 관련 익스텐트의 각 파라미터를 변경하는 것을 포함한다. 이러한 지시는 하나 이상의 시작 익스텐트로부터 하나 이상의 최종 익스텐트로의 연쇄로 계속될 수 있다. 최종 익스텐트의 파라미터가 시작 익스텐트를 다시 지시하게 함으로써 링크가 종료된다. 모듈(108)은 또한, 메타데이터 내의 파라미터를 변경하여 그룹화된 공간을 개별 익스텐트 및/또는 더 작은 그룹화된 공간으로 다시 해체할 수 있다. 비중복화 조직에서 캐시 데이터가 완충저장되는 경우에는, 모듈(108)은 하나 이상의 익스텐트의 파라미터를 할당된(또는 무효) 값으로 변경하여, 익스텐트가 현재 캐시 데이터를 완충저장하는 데 사용되고 있음을 나타낼 수 있다. 익스텐트가 캐시 데이터를 완충저장하는 데 더 이상 사용되고 있지 않을 때에는, 모듈(108)은 관련된 파라미터를 할당되지 않은(또는 사용되지 않는) 값으로 변경할 수 있다.

[0022] 모듈(110)은 캐시 제어기로 구현할 수 있다. 모듈(110)은 표준 캐시 동작 및 표준 캐시 정책을 수행할 수 있다. 캐시 동작에는 일반적으로, 회로(102)로 데이터를 쓰고, 회로(102)에서 캐시(또는 완충저장된) 데이터를 읽고, 회로(102)로 메타데이터를 쓰고, 회로(102)로부터 메타데이터를 읽고, 캐시 데이터의 태그를 탐색하고, 일치하는 태그가 발견시에 캐시 히트를 알리고, 일치하는 태그가 발견되지 않을 시에 캐시 미스를 알리고, 회로(102)에서 더티 기록 데이터를 추적하고, 이 더티 기록 데이터를 회로(94)로 플러시하고(flush)(예컨대, 복사), 유효한 및 유효하지 않은 캐시 라인을 추적하고, 유효한 및 유효하지 않은 캐시 워드를 추적하는 등의 동작이 포함되지만, 이들에 국한되는 것은 아니다. 모듈(110)은 또한, 캐시 데이터의 유지 및 제거에 관한 하나 이상의 캐시 정책을 구현하는 기능을 할 수 있다. 캐시 정책에는, 더티 기록 데이터에 대한 라이트백(write-back) 모드, 기록 데이터의 일부 유형에 대한 라이트스루(write-through) 모드, 기록 데이터의 일부 유형의 중복화(또는 보호) 완충저장, 기록 데이터의 다른 유형의 비중복화(또는 비보호) 완충저장, 더티 기록 데이터에 대한 하나 이상의 플러시 정책, 회로(102) 내의 공간을 비우기 위한 하나 이상의 소거 정책 등의 동작이 포함되지만, 이들에 국한되는 것은 아니다.

[0023] 모듈(112)은 RAID 제어기로 구현할 수 있다. 모듈(112)은 개괄적으로, 회로(102)에서 완충저장되어 회로(94)에 저장되어 있는 데이터를, 이 데이터에 지정된 RAID 정책에 기반한 다양한 RAID 구성으로 조직화(또는 배열 또는 구성)한다.

[0024] 보호해야 할 기록 데이터를 회로(92)로부터 수신할 때에, 모듈(112)은 이 기록 데이터를 RAID 조직들 중에서 주어진 조직으로 재배치할 수 있다. 그 다음에, 재배치된 기록 데이터는, 기록 데이터의 크기(예컨대, 기록 데이터에 의해 얼마나 많은 익스텐트 경계(바운더리)가 교차되는지) 및 기록 요청에 해당되는 RAID 조직(예컨대, RAID 0, RAID 3, RAID 5 등)에 기반하여, 둘 이상의 회로(106a~106n)에 저장될 수 있다. 특정 애플리케이션의 기준에 부합하기 위하여 다른 형식의 데이터 보호책을 실시할 수 있다.

[0025] 보호된 캐시 데이터가 회로(102)로부터 회로(92)로 제공되어야 할 때, 모듈(112)은 데이터의 에러를 검사하여 적절하게 이 에러를 수정하는 기능을 할 수 있다. 데이터가 스트라이핑(stripe)되었을 때에는, 모듈(112)은 또한, 요청된 데이터가 회로(92)로 제공되기 전에 스트라이프를 해당 블록 속으로 모아들인다.

[0026] RAID로 보호된 기록 데이터를 회로(102)로부터 회로(94)로 전송하는 것은, 데이터를 RAID 조직 내에 유지시키는 동안에 수행될 수 있다. 예를 들어, 데이터가 미러 구조의 RAID 1 조직 내에 있는 경우에, 이 데이터는 두 회로(106a~106n)(예컨대, 106a 및 106b)로부터 신호 DB와 DS를 통해, 대응되는 두 회로(98a~98n)(예컨대, 98f 및 98g)로 전송될 수 있다. 다른 예를 들자면, 데이터가, 두 스트라이프 익스텐트를 갖는 패리티 익스텐트를 사용하는 RAID 3 조직 내에 있는 경우에, 이 데이터는 세 회로(106a~106n)(예컨대, 106a, 106b, 106c)로부터 네트워크(97)를 통해, 대응되는 세 회로(98a~98n)(예컨대, 98d, 98g, 98k)로 전송될 수 있다.

[0027] RAID로 보호된 저장 데이터를 회로(94)로부터 회로(102)로 전송하는 것은, 회로(94)로부터 모듈(112)로의 전송은 RAID 조직에서 수행될 수 있고, 모듈(112)로부터 회로(102)로의 전송은 비중복화 조직에서 수행될 수 있다. 예를 들어, RAID 3 조직 내에 저장된 데이터는, 세 가지 병행 전송(예컨대, 두 개의 익스텐트 및 패리티 익스텐트)처럼 회로(94)로부터 모듈(112)로 복사될 수 있다. 모듈(112)은 통상의 경우와 동일하게 데이터의 에러 검사 및 수정을 수행할 수 있으며, 저장된 데이터의 하나의 비중복화 사본을 회로(102)로 전송할 수 있다.

[0028] 도 3은 익스텐트(140)의 예시적 구조도이다. 회로(102) 내의 캐시 데이터 및 회로(94) 내에 저장된 데이터는

익스텐트(또는 위도우(widow))의 형태로서 구성할 수 있다. 각 익스텐트는 일반적으로, 데이터를 할당·할당해제·추적·관리하는 하나의 단위로서 취급되는 연속적인 데이터 블록의 집합을 의미한다. 예시한 익스텐트(140)는 통상, 다수의 데이터 라인(142a~142n)을 포함한다. 각 데이터 라인(142a~142n)은 다수의 데이터 블록(144a~144n)을 포함할 수 있다. 각 데이터 블록(144a~144n)은 통상, 다수의 바이트, 워드, 또는 기타 데이터 단위로 구성된다. 또한, 각 데이터 블록(144a~144n)은 그에 대응되는 논리적 블록 주소(예컨대, LBA A ~ LBA N)를 가질 수 있다. 특정 애플리케이션의 기준에 부합하기 위하여 그 밖의 익스텐트 조직을 구현할 수 있다.

[0029] 도 4는 캐시 데이터를 예시하고 있다. 이 예에서는 중복화 구조로서 RAID 1 조직을 도시하고 있다. 단순화를 위해, 두 개의 회로(106a~106n)(예컨대, 106a 및 106b)만 도시한다. 각 회로(106a 및 106b)는 하나 이상의 익스텐트 및 해당 메타데이터로 구성할 수 있다. 일부 익스텐트(예컨대, 기록 익스텐트)는 회로(94)가 맡지 않았던 더티 기록 데이터를 취급(carry)할 수 있다. 다른 익스텐트(예컨대, 독출 익스텐트)는 회로(94)로부터의 캐시 데이터를 몇 가지 기본적인 캐시처리 기준에 의거하여 취급할 수 있다. 더티 기록 데이터를 취급하는 기록 익스텐트는, 회로(106a 또는 106b)가 고장 시에 중복화 보호 조직을 이용해 완충저장될 수 있다. 캐시 독출을 취급하는 다른 모든 익스텐트는 중복화 보호 조직 없이 완충저장될 수 있는데, 왜냐하면, 회로(106a 또는 106b)의 고장시에 상실된 캐시 데이터가, 이 고장 후에 회로(94)로부터 동작 회로(106a 및 106b)로 재집결될 수 있기 때문이다.

[0030] 도 5는 회로(102) 내의 몇 가지 익스텐트 그룹의 도식이다. 도시한 예에서, 회로(102)에 회로(106a 및 106b)가 포함될 수 있다. 특정 애플리케이션의 기준에 부합하기 위해서 더 적거나 더 많은 회로(106a~106n)를 구현할 수도 있다. 도시한 것과 같이, 다수의 익스텐트 그룹, 다수의 그룹화된 유형(예컨대, RAID 1, RAID 3, RAID 4, RAID 5, RAID 10, RAID 0+1 등), 및/또는 다수의 그룹화되지 않은 익스텐트(예컨대, 독출 및 기록 익스텐트)가 회로(102) 내에 동시에 존재할 수 있다.

[0031] 각 익스텐트와 이에 대응되는 메타데이터는 기존의 방식(또는 기법)으로 관리할 수 있다. 구체적으로, 각 익스텐트는, 익스텐트를 추적하고 식별하기 위한 고유의 익스텐트 식별 번호를 갖고 있다. 모든 호스트 입/출력 요청에 대해서, LBA를 익스텐트로 매핑하여 정확한 익스텐트의 위치를 정함으로써 호스트 요청을 처리할 수 있다. 각 익스텐트의 메타데이터에 파라미터(또는 필드)를 부가하여 중복화 보호 조직에 사용되는 다른 익스텐트의 식별 번호를 추적할 수 있다.

[0032] 회로(100)가 단일 익스텐트 내의 기록 데이터를 회로(92)로부터 수신할 때에는, 둘 이상의 익스텐트를 그룹에 할당할 수 있다. 루프 내에서 서로 지시하도록(point)(또는 link) 하기 위하여 해당 파라미터를 업데이트할 수 있다. 익스텐트를 그룹에 할당하는 것은, 단일 그룹 내의 각 익스텐트가 서로 다른 회로(106a~106n)에 위치하도록 실행될 수 있다.

[0033] 기록 데이터가 다수의 익스텐트의 범위에 있고 및/또는 익스텐트 경계에 걸쳐 있는 일부 실시예에서, 두 개 이상의 익스텐트가 모인 익스텐트 조는 대형 익스텐트로서 취급할 수 있다. 두 개 이상의 대형 익스텐트를 그룹에 할당할 수 있고, 이에 해당되는 파라미터를 업데이트하여 루프 내에서 서로를 지시하도록 할 수 있다. 각 대형 익스텐트군의 각 조를 서로 다른 회로(106a~106n)에 위치시킬 수 있다. 예를 들어, RAID 1 그룹(150)은 일반적으로, 회로(106b)에 2개 1조의 익스텐트를 포함하고, 회로(106c)에서는 다른 2개 1조의 익스텐트를 포함한다. 회로(106b) 내의 두 익스텐트의 각 파라미터는 회로(106c) 내의 대응되는 익스텐트를 가리킬 수 있다. 마찬가지로, 회로(106c) 내의 두 익스텐트의 각 파라미터는 회로(106b) 내의 대응되는 익스텐트를 가리킬 수 있다.

[0034] 기록 데이터가 다수의 익스텐트의 범위에 있고/또는 익스텐트 경계에 걸쳐 있는 다른 실시예에서, 인접한 익스텐트를 서로 다른 회로(106a~106n)에 위치시키고 중복화 조직에 대해서 링크시킬 수 있다. 예를 들어, RAID 3 그룹(152) 내의 데이터는 회로(106b)에 완충저장되는 시작 익스텐트, 회로(106c)에 완충저장되는 그 다음 익스텐트, 그리고 회로(106d)에 완충저장되는 패리티 익스텐트를 포함할 수 있다.

[0035] 캐시 정책에 의거하여, 일부 기록 데이터는 비중복화 조직(154)의 회로(103)에 완충저장될 수 있다. 이 보호되지 않는(비보호) 기록 데이터는, 회로(102)의 고장에 의해 데이터 상실이 일어나는 경우에 회로(92)에 의해 교체될 수 있다. 이 비보호 기록 데이터에 있어서는 또한, 잠재적인 데이터 상실을 수용할 수 있는 중요성이 낮을 수 있다.

[0036] 독출 미스(read miss) 시에 수행되는 각 캐시의 채움 동작(cache fill)에 있어서, 두 개 이상이 모인 1조의 익스텐트를, 회로(102)에 채우고자 하는 데이터의 위치 및 양에 따라 하나 이상의 회로(106a~106n)에 할당할 수

있다. 할당된 조의 각 파라미터는, 해당 익스텐트가 비중복화 조직에 할당되었음을 알리는 값으로(예컨대, 공백 값(null value) 또는 할당된 값으로) 변경(또는 프로그래밍) 할 수 있다.

[0037] 익스텐트 그룹 내의 더티 기록 데이터가 회로(94)로 플러시된 후에, 이 그룹의 익스텐트를 할당해제(unallocate)하여 차후의 할당에 사용가능한 공간을 확보할 수 있다. 인접한 익스텐트의 파라미터를 업데이트함으로써 익스텐트를 할당해제하여, 이 할당해제된 익스텐트를 그룹으로부터 링크해제(unlink)할 수 있다. 할당해제되는 익스텐트는 기록 데이터를 보호하기 위하여 사용된 중복화 조직에 의해 결정될 수 있다. 예를 들어, RAID 1 조직의 익스텐트 그룹은 미러화된 익스텐트(mirrored extent) 중 하나를 할당해제할 수 있는데, 두 개의 미러화된 익스텐트가 모든 기록 데이터를 포함하고 있기 때문이다. RAID 3 조직을 갖는 익스텐트 그룹은, 패리티 데이터를 완충저장하기 위하여 지시된 익스텐트를 링크해제할 수 있다. 실제 데이터를 완충저장하는 나머지 익스텐트는 그룹 내에 남아 있을 수 있게 된다.

[0038] 회로(106a~106n)가 고장나면, 표준 복구 기법을 사용하여, 살아 있는 익스텐트로부터의 호스트 요청을 처리할 수 있다. 예를 들어, 고장난 회로(106a~106n)의 블록이 상실된 RAID 5 조직의 익스텐트 그룹(156)은, 살아 있는 회로(106a~106n)의 살아 있는 블록 및 패리티 블록을 이용하여 상실된 블록을 재구축할 수 있다. 재구축된 블록은 추후에 독출/기록 요청을 수행하는 데 사용할 수 있다.

[0039] 도 6은 탄력적 미러 캐시를 갖는 캐시 조직(160)의 예시적 개략도이다. 이 예는 개괄적으로, 두 개의 회로(106a~106n)(예컨대, 106a 및 106b)를 이용한 미러 조직을 나타내고 있다. 공통의 중복화 기록 공간(162)에 익스텐트 1과 익스텐트 10이 둘 다 할당될 수 있다. 이 할당은, (i) 익스텐트 10의 식별 번호를 지시하기 위한, 익스텐트 1에 관련된 파라미터 및 (ii) 익스텐트 1의 식별 번호를 지시하기 위한, 익스텐트 10에 관련된 파라미터를 변경함으로써 수행할 수 있다. 회로(106a)의 버퍼에 저장된 익스텐트 4는, 연관된 파라미터를 공백값(또는 할당된 값)으로 변경함으로써 독출 공간(164)에 할당할 수 있다. 회로(106b)의 버퍼에 저장된 다른 익스텐트 8은, 연관된 파라미터를 공백값(또는 할당된 값)으로 업데이트함으로써 다른 독출 공간(166)에 할당할 수 있다. 다른 익스텐트(예컨대, 참조번호 168의 익스텐트 12)에는 데이터를 넣지 않을 수 있으며, 따라서, 해당 파라미터가 비할당 값(또는 불사용 값)으로 설정될 수 있다. 회로(106b)에 위치한 익스텐트(예컨대, 익스텐트 7~12)에 대한 메타데이터는 메타데이터 공간(172)에서 완충저장될 수 있다.

[0040] 도 7은 기록 요청을 처리하는 예시적 방법(180)의 흐름도이다. 이 방법(또는 절차)(180)은 회로(100)에서 구현될 수 있다. 이 방법(180)은 개괄적으로, 단계(또는 상태)(182), 단계(또는 상태)(184), 단계(또는 상태)(186), 단계(또는 상태)(188), 단계(또는 상태)(190), 및 단계(또는 상태)(192)를 포함한다. 이들 단계 182~192는 하드웨어, 소프트웨어, 하드웨어와 소프트웨어의 조합, 또는 그 밖의 실현 방법에 의해 구현할 수 있는 모듈 및/또는 블록에 해당될 수 있다.

[0041] 단계(182)에서, 회로(104)는 신호 DH를 통해 회로(92)로부터 기록 요청을 수신할 수 있다. 단계(184)에서는 회로(102)에 완충저장된 캐시 데이터가 있는지 회로(104)(예컨대, 모듈(110))가 탐색하여, 기록 데이터를 위한 적절한 익스텐트가 회로(102)에 이미 있는지를 판정한다. 적절한 익스텐트를 회로(102)에서 찾을 수 없다면(예컨대, 캐시 미스), 회로(104)(예컨대, 모듈(108))는 회로(102) 내의 불사용 익스텐트를 회로(106a~106n)로 할당하여 기록 요청을 처리하도록 할 수 있다(단계(186)).

[0042] 익스텐트 그룹이 기록 요청을 처리할 수 있게 되면 모듈(110)은 회로(94)로부터 기록 익스텐트에 해당되는 저장된 데이터를 복사해 올 수 있다(단계(188)). 단계(188)의 일부로서, 만약 회로(94)에서 수신한 저장된 데이터가 보호 조직(또는 형식)에 아직 없다면, 회로(104)(예컨대, 모듈(112))는 이 데이터를 보호 조직 내로 배치시킬 수 있다. 그 다음에 단계(188)에서 이 데이터를 모듈(110)에 의해서 적절한 회로(106a~106n) 내로 완충저장할 수 있다. 단계(190)에서, 모듈(110)과 모듈(112)은 회로(92)로부터 수신한 기록 데이터를 보호 조직에 따라 조직화할 수 있고 이 기록 데이터를 회로(102)에 저장할 수 있다. 기록 데이터가 아직 회로(94)로 다시 복사되지 않았기 때문에(예컨대, 라이트백 정책), 모듈(110)은 새로 기록된 데이터를 더티 데이터로 표시하기 위하여 해당 메타데이터를 업데이트할 수 있다.

[0043] 만일 기록 데이터를 회로(102)에서 찾았다면(예컨대, 캐시 히트), 모듈(108)은 추가 익스텐트를, 캐시 히트가 있었던 하나 이상의 기존 익스텐트에 할당하고 링크시킬 수 있다. 그 다음에, 캐시 미스가 일어난 경우와 동일한 방식으로 단계(188)에서와 같이 회로(94)로부터의 데이터를 이 익스텐트 그룹에 채울(populate) 수 있다. 이 새로운 기록 데이터는 단계(190)에서 모듈(110)에 의해 회로(106a~106n)에 추가될 수 있다.

[0044] 도 8은 독출 요청을 처리하는 예시적 방법(200)의 흐름도이다. 이 방법(또는 절차)(200)은 회로(100)에서 구현

될 수 있다. 이 방법(200)은 개괄적으로, 상태(또는 단계)(202), 상태(또는 단계)(204), 상태(또는 단계)(206), 상태(또는 단계)(208), 및 상태(또는 단계)(210)를 포함한다. 이들 단계 202~210은 하드웨어, 소프트웨어, 하드웨어와 소프트웨어의 조합, 또는 그 밖의 실현 방법에 의해 구현할 수 있는 모듈 및/또는 블록에 해당될 수 있다.

[0045] 단계(202)에서, 회로(104)는 신호 DH를 통해 회로(92)로부터 독출 요청을 수신할 수 있다. 단계(204)에서는 회로(102)에 완충저장된 캐시 데이터가 있는지 회로(104)(예컨대, 모듈(110))가 탐색한다. 이 탐색에 의해서 독출 데이터를 위한 적절한 익스텐트가 회로(102)에 이미 있는지를 판정할 수 있다. 적절한 익스텐트를 회로(102)에서 찾을 수 없다면(예컨대, 캐시 미스), 회로(104)(예컨대, 모듈(108))는 하나 이상의 불사용 익스텐트를 회로(106a~106n)로 할당하여 독출 요청을 처리하도록 할 수 있다(단계(206)).

[0046] 할당된 익스텐트가 독출 요청을 처리할 수 있게 되면, 모듈(110)은 회로(94)로부터 독출 익스텐트에 해당되는 저장된 데이터를 복사해 올 수 있다(단계(208)). 그 다음에 수신한 데이터를 단계(208)에서 모듈(110)에 의해서 적절한 회로(106a~106n) 내로 완충저장할 수 있다. 단계(210)에서, 모듈(110)은 요청된 독출 데이터의 사본을 익스텐트로부터 회로(92)로 전송할 수 있다. 단계(204)에서, 만일 독출 데이터를 회로(102)에서 찾았다면(예컨대, 캐시 히트), 본 도면의 방법(200)을 단계(210)에서 계속 실행할 수 있다. 즉, 회로(104)가 요청된 독출 데이터를 회로(92)로 전송한다.

[0047] 도 9는 익스텐트를 할당하는 예시적 방법(220)의 흐름도이다. 이 방법(또는 절차)(220)은 회로(100)에서 구현될 수 있다. 이 방법(220)은 개괄적으로, 단계(또는 상태)(222), 단계(또는 상태)(224), 단계(또는 상태)(226), 단계(또는 상태)(228), 단계(또는 상태)(230), 단계(또는 상태)(232), 단계(또는 상태)(234), 및 단계(또는 상태)(236)를 포함한다. 이들 단계 222~236은 하드웨어, 소프트웨어, 하드웨어와 소프트웨어의 조합, 또는 그 밖의 실현 방법에 의해 구현할 수 있는 모듈 및/또는 블록에 해당될 수 있다.

[0048] 단계(222)에서, 회로(예컨대, 모듈(108))는 요청을 처리하기 위하여 할당되어야 할 복수의 익스텐트를 결정할 수 있다. 복수의 익스텐트는 버퍼에 저장될 데이터의 양 및 데이터의 조직화 형식(예컨대, 중복화 보호 또는 비중복화)에 근거하여 결정될 수 있다. 단계(224)에서 조직화 형식이 비중복화인 것으로 판단된 경우에는 본 방법(220)의 단계(226)를 실행할 수 있다. 회로(104)(예컨대, 모듈(108))는 회로(92)로부터 수신한 독출 요청에 관련된 비중복화 조직에 기초하여 하나 이상의 익스텐트를 할당할 수 있다. 이 할당은 개괄적으로, 익스텐트의 해당 파라미터를 할당되지 않은 값으로부터 할당된 값(또는 공백값)으로 변경하는 것을 포함한다(단계(226)). 개괄적으로, 결과로서의 하나 이상의 익스텐트(예컨대, 도 6의 익스텐트(164))는 회로(94)로부터의 데이터를 채울 준비가 된다.

[0049] 단계(224)에서, 조직의 형식이 중복화 조직인 것으로 판단되면, 본 방법(220)은 단계(228)에서부터 실행될 수 있다. 단계(228) 내지 단계(236) 동안에, 각 익스텐트에 해당되는 각 파라미터를 변경함으로써 복수의 익스텐트를 서로 동적으로 링크시켜서 복수의 익스텐트 내의 다른 익스텐트를 지시하도록 할 수 있다. 단계(228)에서, 모듈(108)은 카운터를 시작할 수 있다(예컨대, N=1). 단계(230)에서 최초 할당된 익스텐트의 해당 파라미터를 업데이트(또는 변경)하여 다음 할당된 익스텐트를 지시하도록 할 수 있다. 단계(232)에서 카운터가 모듈(108)에 의해 업데이트될 수 있다. 단계(234)에서 다른 익스텐트가 더 있는지 판단할 수 있다. 그룹 내로 링크해야 할 익스텐트가 더 있다면, 단계(230)로 돌아가, 모듈(108)이 다음 익스텐트의 파라미터를 갱신하여 후속 익스텐트를 지시하도록 할 수 있다. 단계(230)에서 단계(234)까지의 루프는 모든 익스텐트가 그룹에 추가될 때까지 계속 진행될 수 있다. 단계(236)에서, 모듈(108)은 마지막 익스텐트의 파라미터를 업데이트하여 최초 익스텐트를 다시 지시하도록 하면서 링크 수행을 종료할 수 있다. 결과로서의 익스텐트 그룹(예컨대, 도 6의 그룹(162))에 데이터를 채울 준비가 될 수 있다.

[0050] 도 10은 익스텐트를 할당해제하는 예시적 방법(240)의 흐름도이다. 이 방법(또는 절차)(240)은 회로(100)에서 구현될 수 있다. 이 방법(240)은 개괄적으로, 단계(또는 상태)(242), 단계(또는 상태)(244), 단계(또는 상태)(246), 및 단계(또는 상태)(248)를 포함한다. 이들 단계 242~248은 하드웨어, 소프트웨어, 하드웨어와 소프트웨어의 조합, 또는 그 밖의 실현 방법에 의해 구현할 수 있는 모듈 및/또는 블록에 해당될 수 있다.

[0051] 단계(242)에서, 회로(104)(예컨대, 모듈(110))는 더티 기록 데이터를 회로(102)로부터 회로(94)로 플러시할 수 있다. 일부 상황에서는, 기록 데이터를 회로(102)로부터 회로(104)로 읽어들여서, 에러를 수정하고, 원래의 형식으로 복원하여, RAID 구성 내로 배치하고, 회로(104)로부터 회로(94)로 전송할 수 있다. 다른 상황에서는, 더티 기록 데이터를 각 회로(106a~106n)로부터 해당 회로(98a~98n)로 있는 그대로 복사할 수 있다. 예를 들어, 익스텐트가 두 개 이상의 데이터 블록과 패리티 블록으로 구성되는 경우에, 각 데이터 블록과 패리티 블록을 신

호 DB와 DS를 통해서 적절한 회로(106a~106n)로부터 해당 회로(98a~98n)로 변경하지 않은 상태로 복사할 수 있다.

[0052] 단계(244)에서, 회로(104)(예컨대, 모듈(108))는, 방금 플러시된 익스텐트 그룹으로부터 어느 익스텐트를 링크해제 및 할당해제할 수 있을지를 결정할 수 있다. 예를 들어, 기록 데이터가 미러화(예컨대, RAID 1 조직)에 의해 보호되는 경우에는 두 익스텐트 중 어느 것이라도 링크해제 및 할당해제할 수 있는데, 그 이유는 두 익스텐트가 모두 기록 데이터 전부를 포함하고 있기 때문이다. 다른 예로서, 기록 데이터가 RAID 4 형태로 보호되는 경우에는, 패리티 데이터를 완충저장하고 있는 익스텐트를 링크해제 및 할당해제할 수 있다. 단계(246)에서, 모듈(108)은 인접 익스텐트 내의 해당 파라미터를 변경하여 선택된 익스텐트를 링크해제함으로써 이 선택된 익스텐트를 바이패스시킬 수 있다. 단계(248)에서, 모듈(108)은 선택된 익스텐트의 파라미터를 할당되지 않은 값으로 변경할 수 있다.

[0053] 도 11은 회로(102)의 더티 기록 데이터를 탄력적으로 미러화함에 따른 공간상의 이득을 나타내는 그래프(260) 도식이다. 세로축으로의 공간상 이득은 일반적으로, 미러화되지 않은 데이터가 완충저장 가능해지게 되는 전체 익스텐트의 백분율을 나타낸다. 그래프(260)에 나타낸 것처럼, 공간상 이득은 데이터의 기록 부분을 나타내는 익스텐트가 적을수록 증가할 수 있다. 더티 기록 데이터를 보호하기 위해 모든 익스텐트를 미러화하면(예컨대, 100%), 공간 절감이 0%가 될 수 있다. 미러 데이터를 완충저장하는 데 사용된 익스텐트의 비율이 적으면, 공간상 이득은 동적으로(또는 탄력적으로) 증가하게 된다. 미러 데이터를 완충저장하는 데 사용되는 익스텐트가 전혀 없으면(예컨대, 0%), 정상적으로는 미러 데이터를 보관하는 데 사용되었을 회로(106a~106n)의 50%가, 그 대신에, 더티하지 않은(non-dirty) 기록 데이터 및/또는 독출 데이터를 보관하는 것이 가능해질 수 있게 된다.

[0054] 회로(100)에 완충저장되는 데이터의 독출/기록 혼합에 따라서, 익스텐트의 탄력적인 그룹화(grouping) 및 탈그룹(ungrouping)에 의해서 추가적인 캐시 공간을 자유롭게 사용할 수 있게 된다. 자유 캐시 공간의 동적 변경에 의해서 개괄적으로, 캐시처리 매체(예컨대, 회로(106a~106n))에 대한 투자 대비, 캐시의 적용범위가 향상된다. 따라서 빈번하게 액세스 되는 데이터가 늘어나는 서버(예컨대, 회로(92))에서 사용가능한 캐시 공간이 추가될 수 있게 된다.

[0055] 본 발명의 일부 실시예는 대체로, 캐시 장치에서 통상 볼 수 있는 대부분의 익스텐트와 메타데이터 관리 인프라를 활용하고 있다. 익스텐트를 탄력적으로 할당하고 할당해제하기 위한 추가 파라미터는 기존의 인프라 위에 구축할 수 있다. 익스텐트 그룹을 생성, 변형, 파괴하는 기법을 미러화 보호(예컨대, RAID 1 수준)로부터 다른 RAID 수준으로 확장시킬 수 있다. 또한, 이들 기법은 개괄적으로, 더티 데이터가 중복화에 의해서 제한될 수 있기 때문에, 주어진 호스트 입/출력 데이터의 양에 대해서, 회로(102)로 기록되는 데이터의 수를 줄이게 된다. 회로(106a~106n)를 플래시 메모리로 구현하는 경우에, 기록 데이터 수가 감소하면 주어진 호스트 입/출력 활동에 대해서 플래시 메모리의 수명이 연장될 수 있다.

[0056] 도 1~11의 도식에 의해 수행되는 기능들은, 본 명세서의 사상에 따라 프로그래밍된 하나 이상의 기존의 범용 프로세서, 디지털 컴퓨터, 마이크로프로세서, 마이크로컨트롤러, RISC(reduced instruction set computer) 프로세서, CISC(complex instruction set computer) 프로세서, SIMD(single instruction multiple data) 프로세서, 신호처리 장치, 중앙처리 장치(CPU), 산술논리 장치(ALU, arithmetic logic unit), 디지털 영상신호처리장치(VDSP), 및/또는 이와 유사한 전자계산 장치를 사용하여 구현될 수 있는바, 이는 당업자에게 자명할 것이다. 적절한 소프트웨어, 펌웨어, 코딩, 루틴, 명령어, op코드, 마이크로코드, 및/또는 프로그램 모듈을, 본 발명의 사상에 기반하여 숙련된 프로그래머가 용이하게 작성할 수 있을 것인바, 이 역시 당업자에게 자명할 것이다. 소프트웨어는 일반적으로, 기계 장치의 하나 이상의 프로세서에 의해 단일 매체 또는 다수 매체로부터 실행된다.

[0057] 본 발명은 또한, 본 명세서에서 설명한 것과 같이, ASIC(주문형 집적회로), 플랫폼 ASIC, FPGA(field programmable gate array), PLD(programmable logic device), CPLD(complex programmable logic device), sea-of-gate, RFIC(무선주파수 IC), ASSP(주문형 표준품; application specific standard product), 하나 이상의 모놀리틱 IC, 플립칩 모듈 및/또는 다중 칩 모듈로서의 하나 이상의 칩(또는 다이) 어레이 제작에 의해, 또는 기존의 구성품 회로의 적절한 네트워크 간의 상호 연결에 의해 구현할 수 있다. 이에 대한 변형은 당업자에게 즉각적으로 자명할 것이다.

[0058] 따라서 본 발명은, 본 발명에 따른 하나 이상의 절차 또는 방법을 수행하도록 기계를 프로그램하는 데 사용할 수 있는 명령어를 포함하는, 저장 매체(들) 및/또는 전달 매체(들) 등의 컴퓨터 제품을 포함할 수 있다. 컴퓨터 제품에 포함된 명령어 및 회로에 관련된 기능들을 기계가 실행하면, 입력 데이터가, 저장 매체상의 하나 이

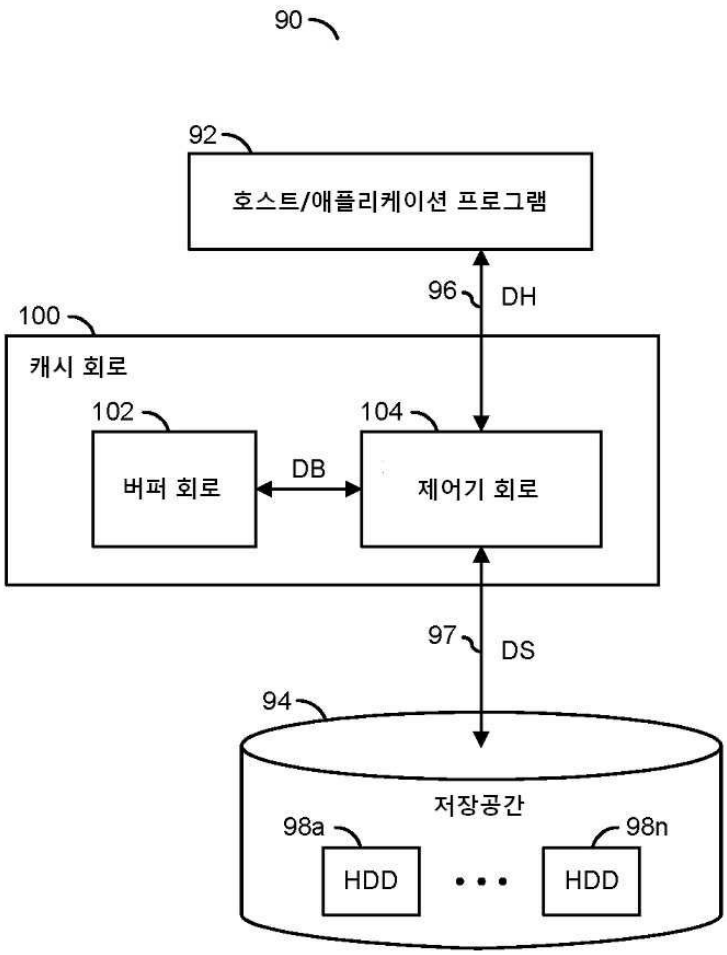
상의 파일, 및/또는 물리적 객체 또는 물질을 대표하는 하나 이상의 출력 신호(청각 및/또는 시각으로 표현되는 신호)로 변환될 수 있다. 저장 매체는 플로피 디스크, 하드 드라이브, 자기 디스크, 광 디스크, CD-ROM, DVD, 및 광자기 디스크 등의 모든 유형의 디스크, 및 ROM, RAM, EPROM, EEPROM, UVPROM(UV erasable programmable ROM), 플래시 메모리 등의 회로, 자기 카드, 광 카드, 및/또는 전자적 명령어를 저장하기에 적합한 그 밖의 모든 매체 유형을 포함할 수 있다.

[0059] 본 발명의 구성요소는 하나 이상의 장치, 유닛, 구성품, 시스템, 기계, 및/또는 장치의 일부 또는 전부를 구성할 수 있다. 장치에는 서버, 워크스테이션, 저장소 배열 제어기, 저장 시스템, 개인용 컴퓨터, 랩톱 컴퓨터, 노트북 컴퓨터, 팜 컴퓨터, PDA, 휴대형 전자 장치, 배터리구동 장치, 셋톱박스, 인코더, 디코더, 트랜스코더, 압축기(컴프레서), 압축해제기(디컴프레서), 전치 프로세서, 후치 프로세서, 셀룰러 전화, 디지털 카메라, 위치 찾기 및/또는 네비게이션 시스템, 의료장비, 헤드업(heads-up) 표시장치, 무선 장치, 오디오 녹음 오디오 저장 및/또는 오디오 재생장치, 영상 녹화 영상 저장 및/또는 영상 재생장치, 게임 플랫폼, 주변장치, 및/또는 모듈 등이 포함될 수 있다. 그러나 이들에 국한되지는 않는다. 당업자는 본 발명의 구성요소를, 특정 애플리케이션의 기준에 부합하도록 다른 유형의 장치로 구현할 수 있음을 이해할 것이다. 본 명세서에서 사용한 용어 "동시에"의 의미는 동일한 시점에서 시작하거나, 동일한 시점에서 종료하거나, 동일한 시간 주기를 갖는 이벤트에 한정되는 의미가 아니라, 일부 공통적인 시간 주기를 공유하는 이벤트를 설명하기 위한 의미로 사용되었다.

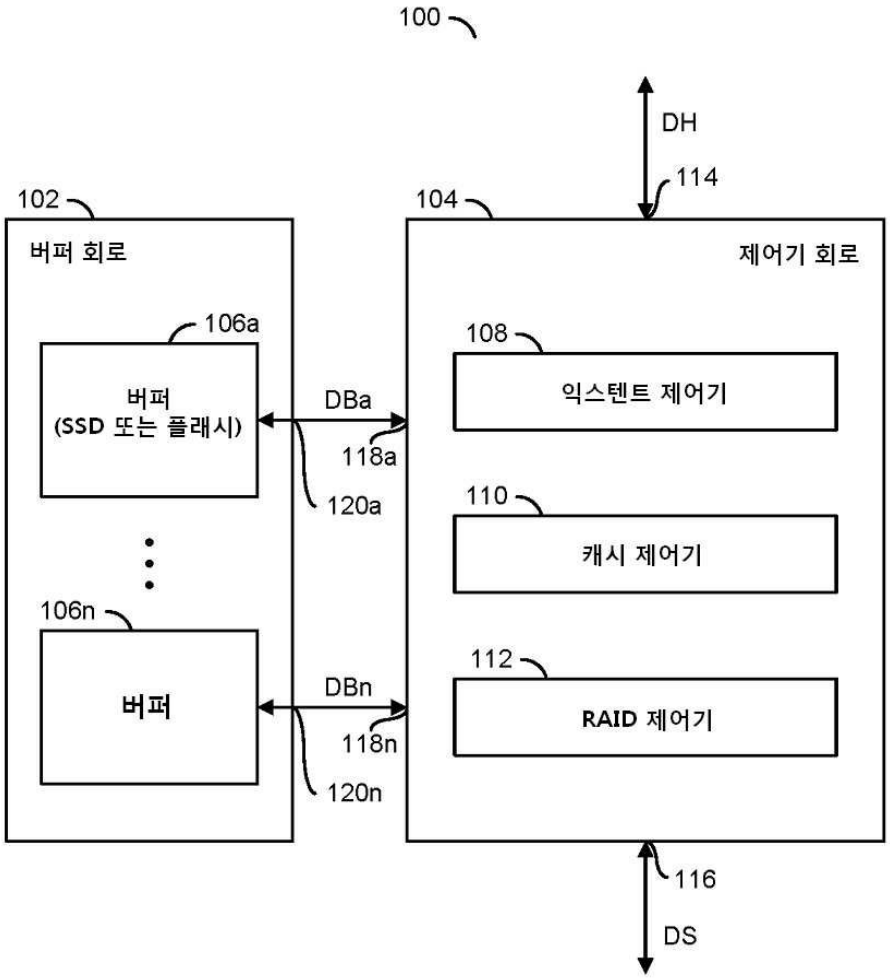
[0060] 본 발명을 그 바람직한 실시예에 관하여 구체적으로 도시하고 설명하였지만, 당업자는 본 발명의 범위를 벗어나지 않도록 형태 및 세부사항을 다양하게 변형할 수 있음을 이해할 것이다.

도면

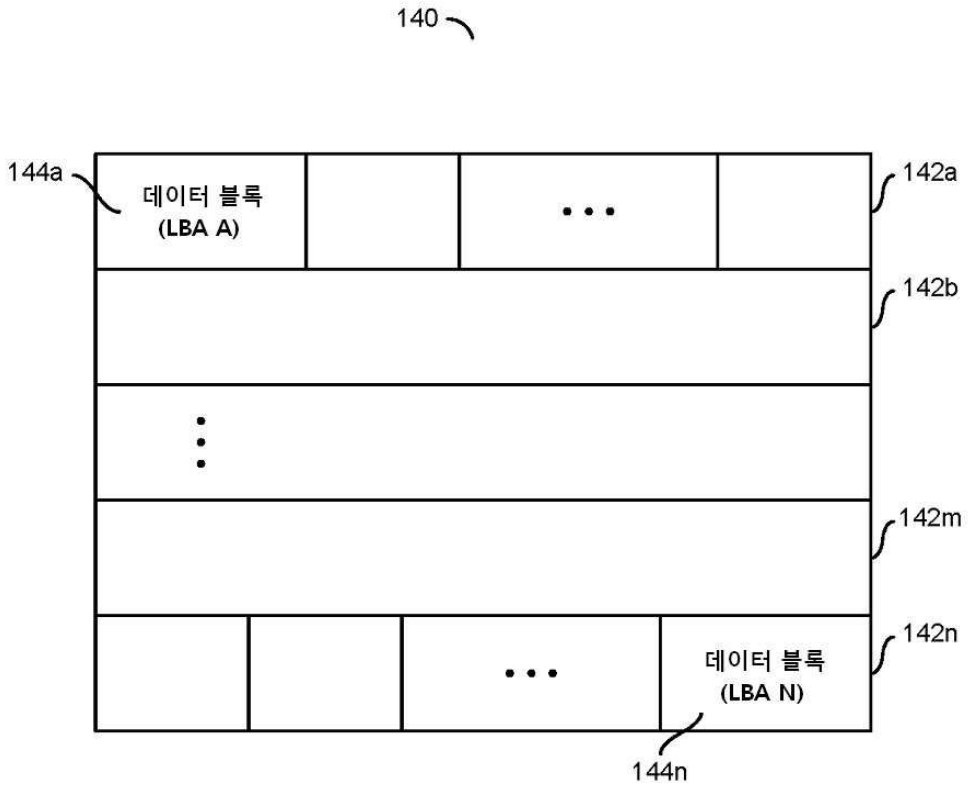
도면1



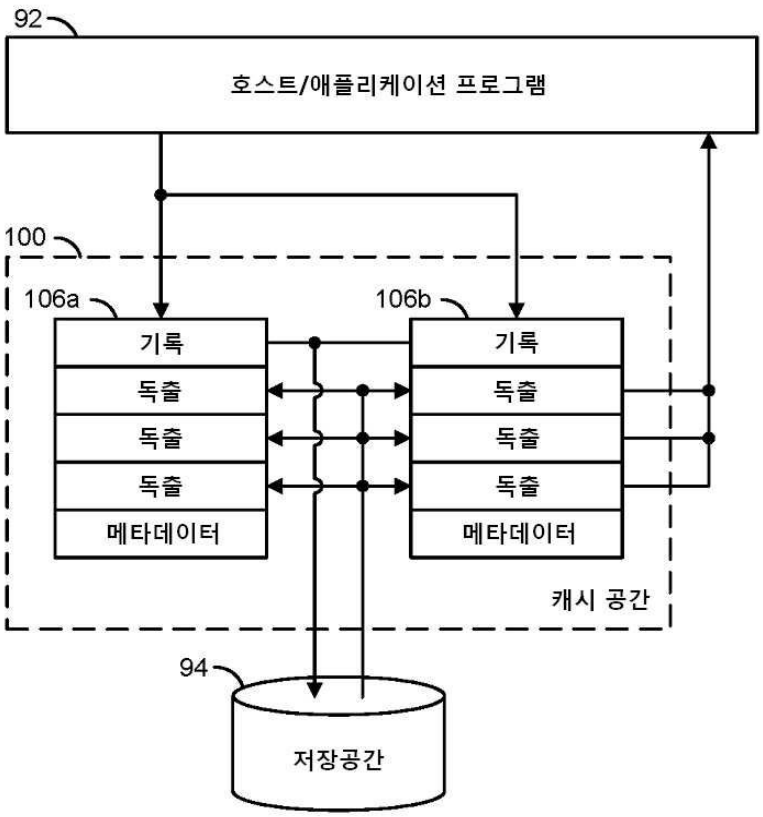
도면2


100
DH
114
102
104
버퍼 회로
제어기 회로
106a
버퍼
(SSD 또는 플래시)
DBa
118a
120a
108
익스텐트 제어기
110
캐시 제어기
106n
버퍼
DBn
118n
120n
112
RAID 제어기
116
DS

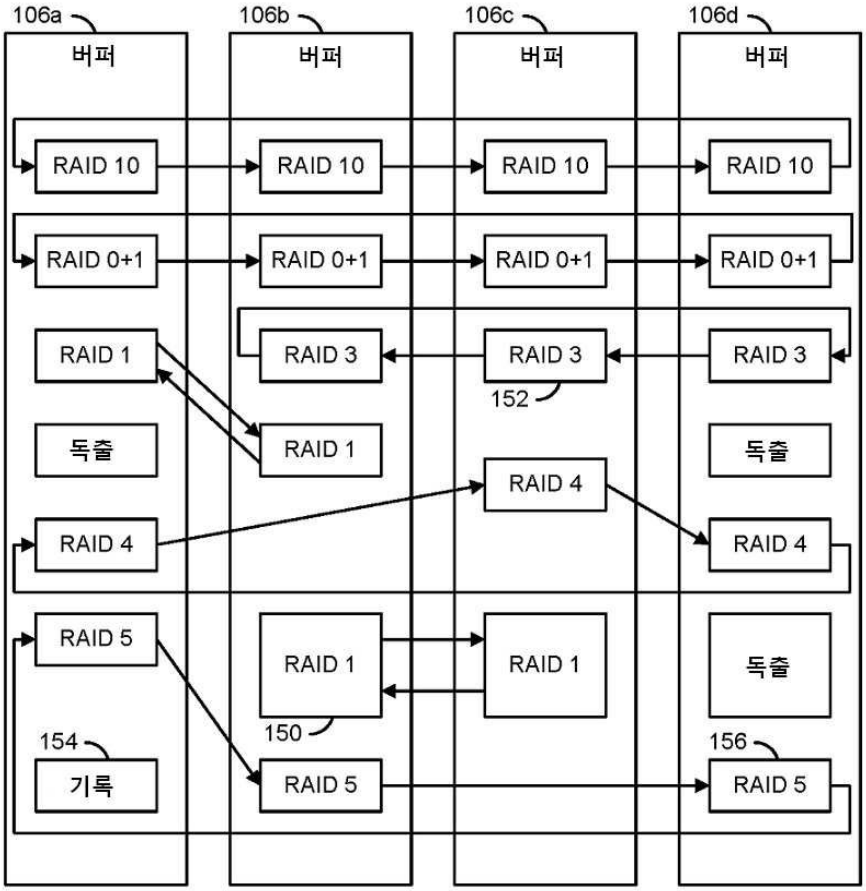
도면3


140
144a
데이터 블록
(LBA A)
142a
142b
142m
142n
데이터 블록
(LBA N)
144n

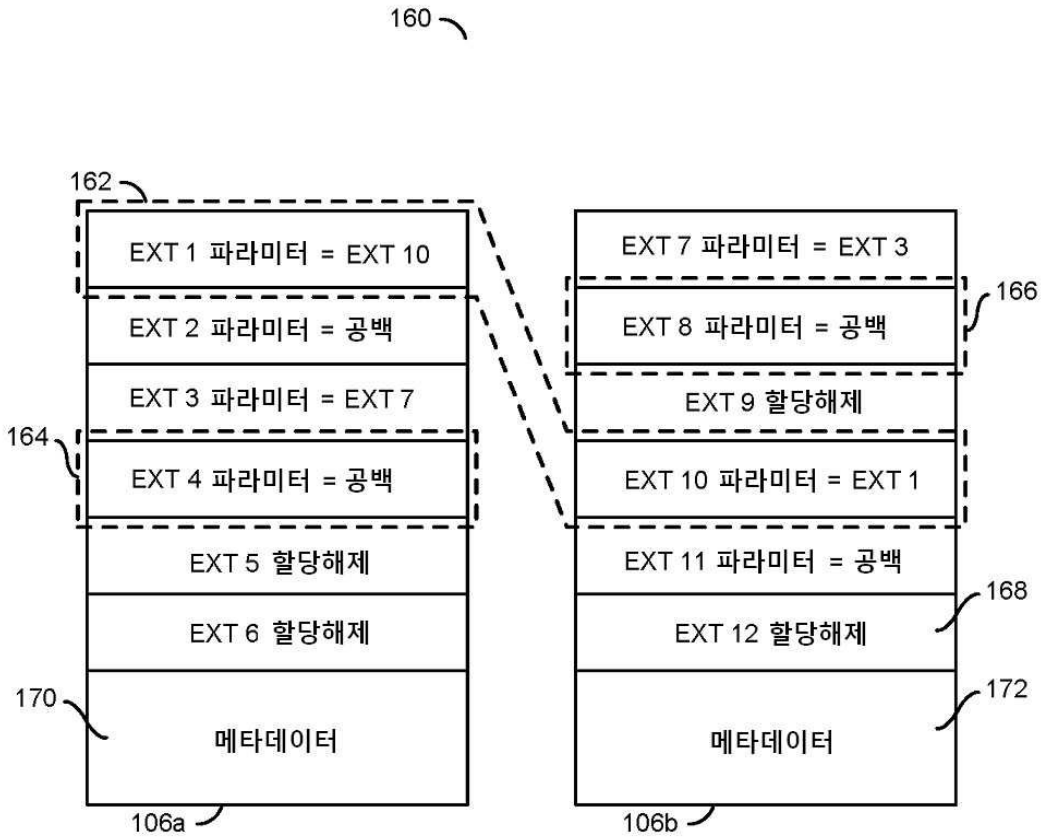
도면4


92
호스트/애플리케이션 프로그램
100
106a
106b
기록
독출
독출
독출
메타데이터
기록
독출
독출
독출
메타데이터
캐시 공간
94
저장공간

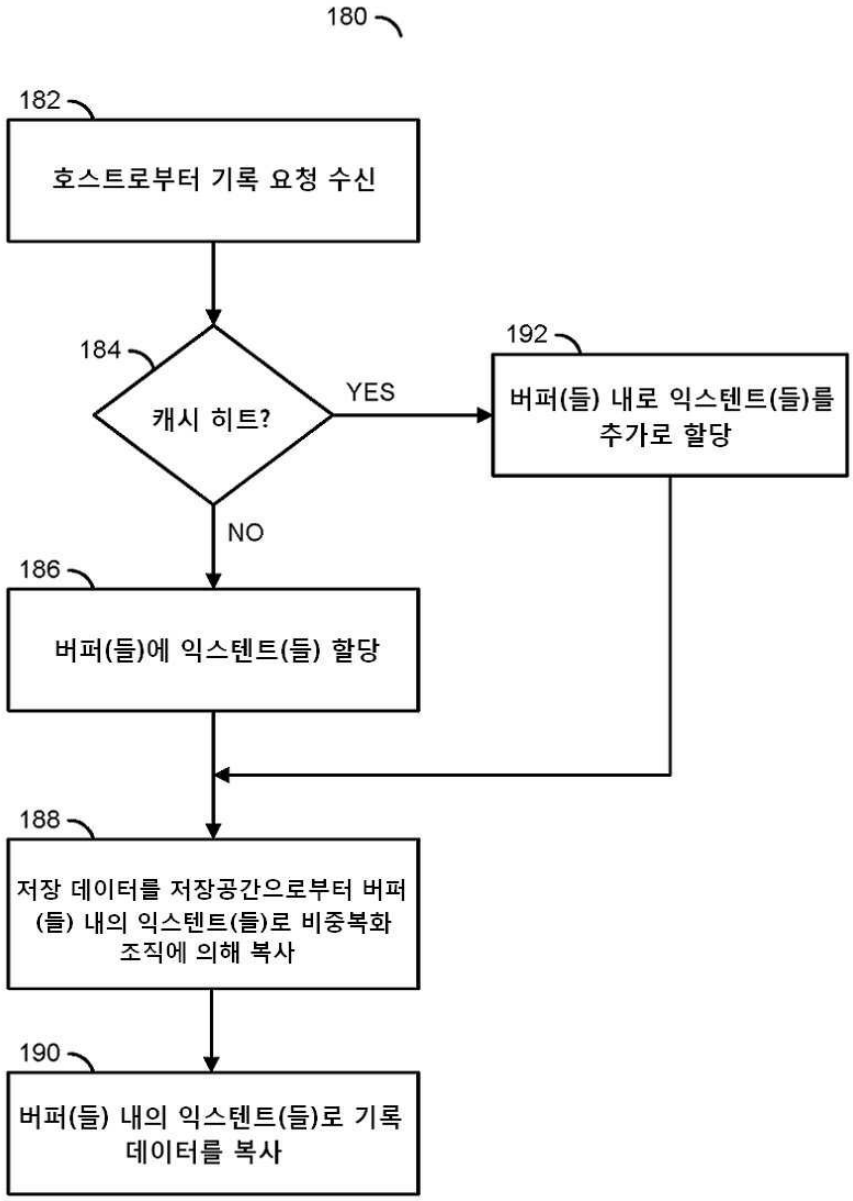
도면5


106a
버퍼
106b
버퍼
106c
버퍼
106d
버퍼
RAID 10
RAID 10
RAID 10
RAID 10
RAID 0+1
RAID 0+1
RAID 0+1
RAID 0+1
RAID 1
RAID 3
RAID 3
RAID 3
152
독출
RAID 1
RAID 4
독출
RAID 4
RAID 4
RAID 5
RAID 1
RAID 1
독출
150
154
156
기록
RAID 5
RAID 5

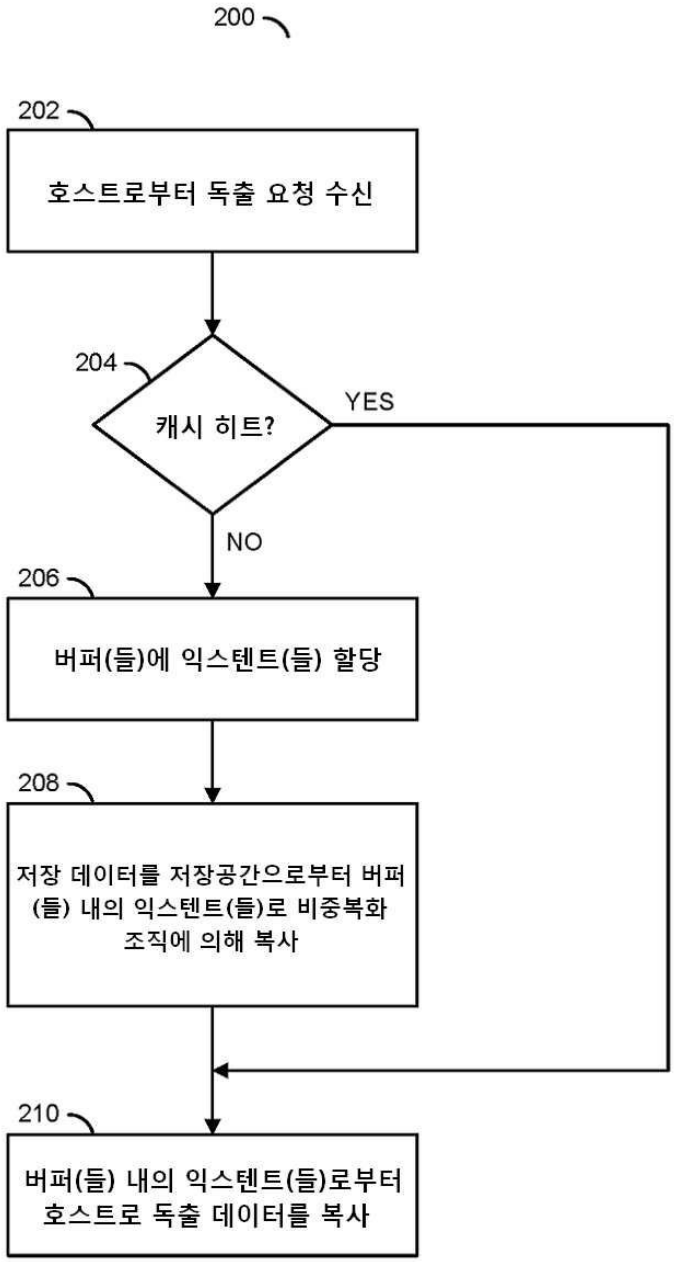
도면6


160
162
EXT 1 파라미터 = EXT 10
EXT 2 파라미터 = 공백
EXT 3 파라미터 = EXT 7
164
EXT 4 파라미터 = 공백
EXT 5 할당해제
EXT 6 할당해제
170
메타데이터
106a
EXT 7 파라미터 = EXT 3
166
EXT 8 파라미터 = 공백
EXT 9 할당해제
EXT 10 파라미터 = EXT 1
EXT 11 파라미터 = 공백
168
EXT 12 할당해제
172
메타데이터
106b

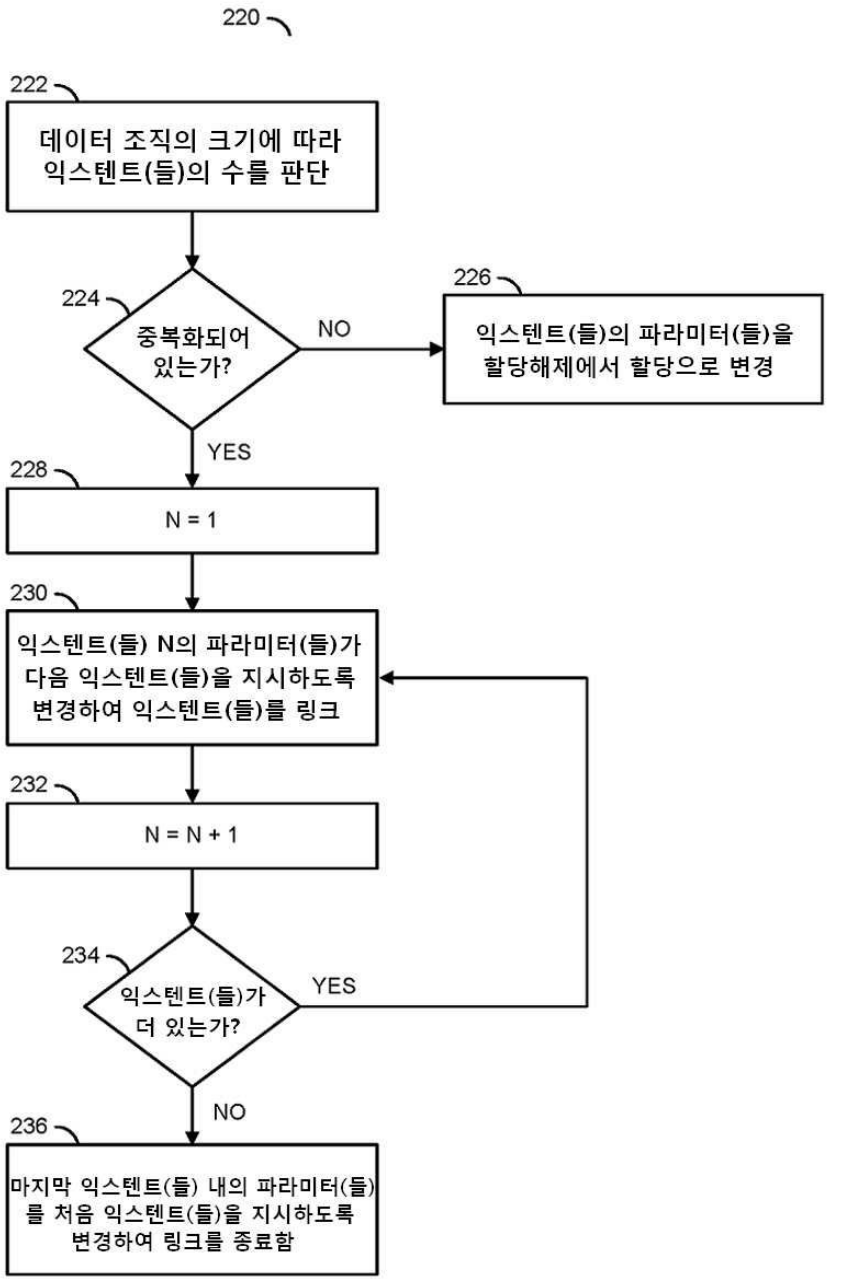
도면7


180
182
호스트로부터 기록 요청 수신
184
캐시 히트?
YES
192
버퍼(들) 내로 익스텐트(들)를 추가로 할당
NO
186
버퍼(들)에 익스텐트(들) 할당
188
저장 데이터를 저장공간으로부터 버퍼(들) 내의 익스텐트(들)로 비중복화 조직에 의해 복사
190
버퍼(들) 내의 익스텐트(들)로 기록 데이터를 복사

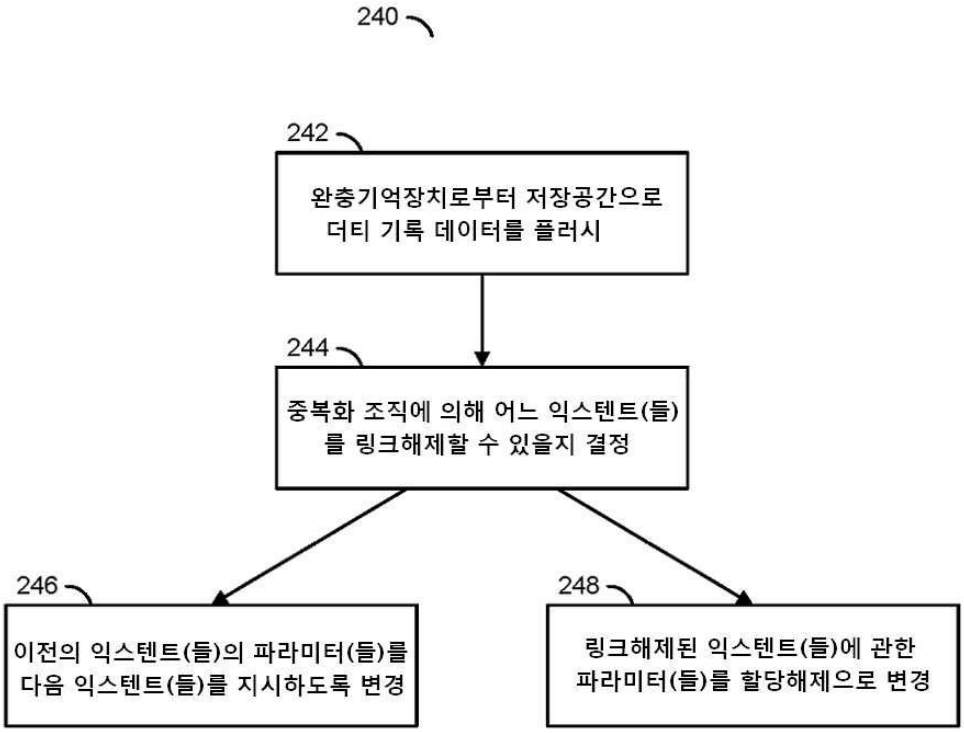
도면8


200
202
호스트로부터 독출 요청 수신
204
캐시 히트?
YES
NO
206
버퍼(들)에 익스텐트(들) 할당
208
저장 데이터를 저장공간으로부터 버퍼(들) 내의 익스텐트(들)로 비중복화 조직에 의해 복사
210
버퍼(들) 내의 익스텐트(들)로부터 호스트로 독출 데이터를 복사

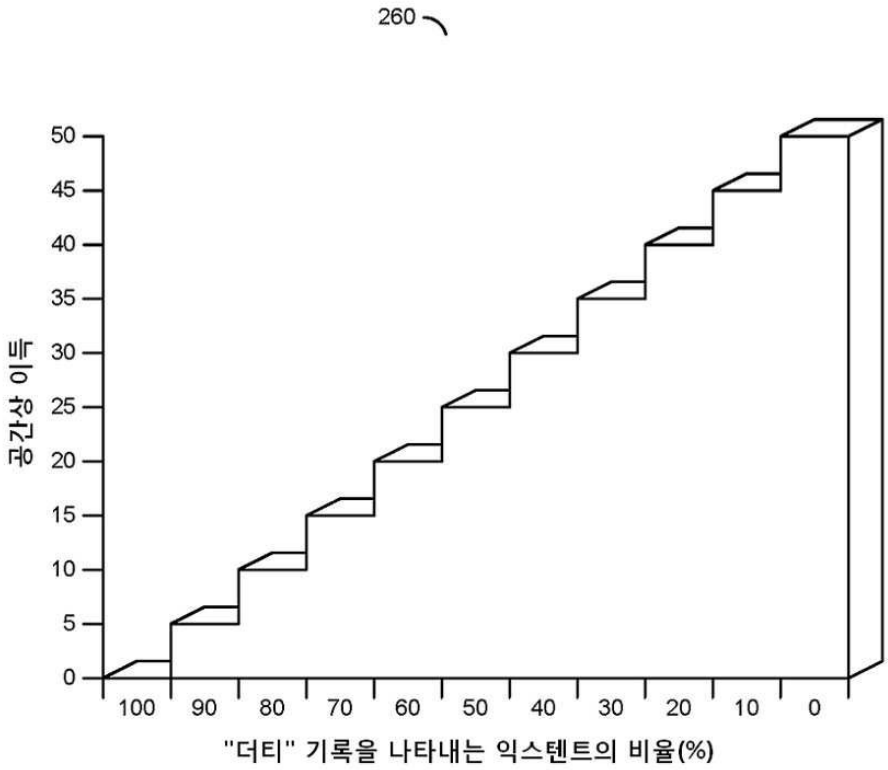
도면9


220
222
데이터 조직의 크기에 따라 익스텐트(들)의 수를 판단
224
중복화되어 있는가?
NO
226
익스텐트(들)의 파라미터(들)을 할당해제에서 할당으로 변경
YES
228
N = 1
230
익스텐트(들) N의 파라미터(들)가 다음 익스텐트(들)을 지시하도록 변경하여 익스텐트(들)를 링크
232
N = N + 1
234
익스텐트(들)가 더 있는가?
YES
NO
236
마지막 익스텐트(들) 내의 파라미터(들)를 처음 익스텐트(들)을 지시하도록 변경하여 링크를 종료함

도면10


240
242
완충기억장치로부터 저장공간으로
더티 기록 데이터를 플러시
244
중복화 조직에 의해 어느 익스텐트(들)
를 링크해제할 수 있을지 결정
246
이전의 익스텐트(들)의 파라미터(들)를
다음 익스텐트(들)를 지시하도록 변경
248
링크해제된 익스텐트(들)에 관한
파라미터(들)를 할당해제으로 변경

도면11


260
공간상 이득
50
45
40
35
30
25
20
15
10
5
0
100
90
80
70
60
50
40
30
20
10
0
"더티" 기록을 나타내는 익스텐트의 비율(%)