(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2010년11월05일
(11) 등록번호 10-0992024
(24) 등록일자 2010년10월29일

(51) Int. Cl.
G06F 12/16 (2006.01) *G06F 3/06* (2006.01)
(21) 출원번호 10-2006-7022769
(22) 출원일자(국제출원일자) 2005년04월26일
심사청구일자 2008년02월28일
(85) 번역문제출일자 2006년10월30일
(65) 공개번호 10-2007-0009660
(43) 공개일자 2007년01월18일
(86) 국제출원번호 PCT/EP2005/051862
(87) 국제공개번호 WO 2005/109167
국제공개일자 2005년11월17일
(30) 우선권주장
10/842,047 2004년05월06일 미국(US)
(56) 선행기술조사문헌
JP13035080 A
US20030225794 A1
전체 청구항 수 : 총 8 항

(73) 특허권자
인터내셔널 비지네스 머신즈 코포레이션
미국 10504 뉴욕주 아몬크 뉴오차드 로드
(72) 발명자
하지 아민
미국 캘리포니아주 95120 산 호세 퀘일 크릭 서클 1211
(74) 대리인
송승필, 김태홍

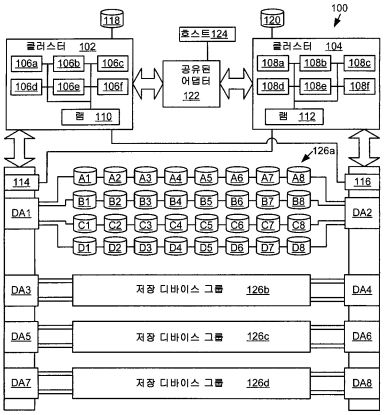
심사관 : 이상헌

(54) 부가적인 및 자율적인 보호 방법을 이용하여 저장 장치들의어레이에 데이터를 저장하는 방법 및 시스템

(57) 요 약

본 발명의 일 양태는 저장 장치들의 어레이에 데이터를 저장하는 방법에 관한 것이다. 이 방법의 일례는 제 1 저장 장치 및 제 2 저장 장치에 제 1 스트립을 기록하는 단계를 포함한다. 또한, 이 일례는 제 2 저장 장치 및 제 3 저장 장치에 제 2 스트립을 기록하는 단계를 포함한다. 이 일례는 제 3 저장 장치 및 제 4 저장 장치에 제 3 스트립을 기록하는 단계를 포함한다.

대 표 도 - 도1



특허청구의 범위

청구항 1

저장 장치들의 어레이에 데이터를 저장하는 방법에 있어서,

제 1 저장 장치 및 제 2 저장 장치에 제 1 스트립(strip)을 기록하는 단계;

상기 제 2 저장 장치 및 제 3 저장 장치에 제 2 스트립을 기록하는 단계;

상기 제 3 저장 장치 및 제 4 저장 장치에 제 3 스트립을 기록하는 단계;

상기 저장 장치들의 어레이에 대하여 최대 스트립 LBA 를 결정하는 단계;

제1 데이터(primary data)에 대하여 이용가능한 스트립 LBA 들의 절반을 저장하는 단계; 및

데이터의 회전된 카피본들에 대하여 이용가능한 스트립 LBA 들의 절반을 저장하는 단계

를 포함하는 데이터 저장 방법.

청구항 2

삭제

청구항 3

삭제

청구항 4

저장 장치들의 어레이에 데이터를 저장하는 방법으로서,

제 1 저장 장치, 제 2 저장 장치 및 제 3 저장 장치에 제 1 스트립을 기록하는 단계;

상기 제 2 저장 장치, 상기 제 3 저장 장치 및 제 4 저장 장치에 제 2 스트립을 기록하는 단계;

상기 제 3 저장 장치, 상기 제 4 저장 장치 및 제 5 저장 장치에 제 3 스트립을 기록하는 단계;

상기 저장 장치들의 어레이에 대하여 최대 스트립 LBA 를 결정하는 단계;

제 1 데이터에 대하여 이용가능한 스트립 LBA 들 중 적어도 33 퍼센트를 저장하는 단계; 및

데이터의 회전된 카피본들에 대하여 이용가능한 스트립 LBA 들 중 적어도 66 퍼센트를 저장하는 단계

를 포함하는 데이터 저장 방법.

청구항 5

삭제

청구항 6

저장 장치들의 어레이에 데이터를 저장하는 방법에 있어서,

제 1 저장 장치 및 제 2 저장 장치에 제 1 스트립(strip)을 기록하는 단계;

상기 제 2 저장 장치 및 제 3 저장 장치에 제 2 스트립을 기록하는 단계;

상기 제 3 저장 장치 및 제 4 저장 장치에 제 3 스트립을 기록하는 단계;

파라미터 N 에 대한 값을 설정하는 단계로서, 저장 장치들의 어레이 내의 각각의 저장 장치가 적어도 N 개의 스트립 LBA를 가지는, 상기 설정 단계;

저장될 스트라이드의 개수 j를 식별하는 단계;

2j가 N-1 이하인지를 결정하는 단계;

2j가 N-1 이하인 경우에는, 상기 어레이의 제 1 저장 장치의 LBA 에 그리고 상기 어레이의 제 2 저장 장치의

LBA 에, 스트립 s1j를 기록하는 단계;

상기 어레이의 제 2 저장 장치의 LBA 에 그리고 상기 어레이의 제 3 저장 장치의 LBA 에, 스트립 s2j를 기록하는 단계; 및

상기 어레이의 제 3 저장 장치의 LBA 에 그리고 상기 어레이의 제 4 저장 장치의 LBA 에, 스트립 s3j를 기록하는 단계

를 포함하는 데이터 저장 방법.

청구항 7

삭제

청구항 8

제 6 항에 있어서,

저장 장치들의 어레이의 각 저장 장치에 대하여, 상기 저장 장치의 스트립 LBA 들의 전체 개수를 결정하는 단계; 및

스트립 LBA 들의 가장 작은 전체 개수를 식별하는 단계

를 더 포함하며, 상기 파라미터 N에 대한 값을 설정하는 단계는 스트립 LBA 들의 가장 작은 전체 개수와 동일하게 N 을 설정하는 단계를 포함하는 것인, 데이터 저장 방법.

청구항 9

삭제

청구항 10

삭제

청구항 11

제 6 항에 있어서,

상기 어레이의 제 1 저장 장치의 스트립 LBAj 에, 그리고 상기 어레이의 제 2 저장 장치의 스트립 LBAj+1에 스트립 s1j를 기록하는 단계;

상기 어레이의 제 2 저장 장치의 스트립 LBAj에, 그리고 상기 어레이의 제 3 저장 장치의 스트립 LBAj+1에 스트립 s2j를 기록하는 단계; 및

상기 어레이의 제 3 저장 장치의 스트립 LBAj 에, 그리고 상기 어레이의 제 4 저장 장치의 스트립 LBAj+1 에 스트립 s3j를 기록하는 단계

를 더 포함하는 데이터 저장 방법.

청구항 12

삭제

청구항 13

삭제

청구항 14

삭제

청구항 15

제 6 항에 있어서,

매핑 테이블을 확립하는 단계;

기록 명령을 수신하는 단계;

카피 플래그(copy flag)가 "예"값을 가지는지 여부를 결정하는 단계;

"예"값을 가지는 경우에는, 상기 매핑 테이블에 따라 저장 장치들의 어레이 내의 대응하는 저장 장치에 스트라이드의 각각의 스트립을 기록하는 단계;

상기 매핑 테이블에 따라 상기 저장 장치들의 어레이 내의 하나 이상의 대응하는 저장 장치에 상기 스트라이드의 각각의 스트립의 적어도 하나의 카피본을 기록하는 단계;

"예"값을 가지지 않는 경우에는, 상기 매핑 테이블에 따라 저장 장치들의 어레이 내의 대응하는 저장 장치에 스트라이드의 각각의 스트립을 기록하는 단계

를 포함하는 데이터 저장 방법.

청구항 16

삭제

청구항 17

삭제

청구항 18

삭제

청구항 19

삭제

청구항 20

저장 장치들의 어레이에 데이터를 저장하는 저장 시스템에 있어서,

제 1 저장 장치 및 제 2 저장 장치에 제 1 스트립을 기록하는 수단;

상기 제 2 저장 장치 및 제 3 저장 장치에 제 2 스트립을 기록하는 수단;

상기 제 3 저장 장치 및 제 4 저장 장치에 제 3 스트립을 기록하는 수단;

상기 저장 장치들의 어레이에 대하여 최대 스트립 LBA 를 결정하는 수단;

제 1 데이터(primary data)에 대하여 이용가능한 스트립 LBA 들의 절반을 저장하는 수단; 및

데이터의 회전된 카피본들에 대하여 이용가능한 스트립 LBA 들의 절반을 저장하는 수단

을 포함하는 데이터 저장 시스템.

청구항 21

삭제

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

청구항 36

삭제

청구항 37

삭제

청구항 38

삭제

청구항 39

제 1 항, 제 4 항, 제 6 항, 제 8 항, 제 11 항 또는 제 15 항 중 어느 한 항에 기재된 단계들 전체를 수행하도록 구성되는 프로그램 코드 수단을 포함하는 컴퓨터 프로그램을 기록한, 컴퓨터 판독 가능한 기록 매체로서, 상기 프로그램은 컴퓨터 상에서 동작하는 것인, 컴퓨터 판독 가능한 기록 매체.

명 세 서

기 술 분 야

[0001] 본 발명은 컴퓨팅 시스템에 데이터를 저장하는 것에 관한 것이다. 보다 상세하게로는, 본 발명의 몇몇 일례들은 데이터 손실로부터의 개선된 보호방법(protection)을 제공하는 방식으로 저장 장치들의 어레이에 데이터를 저장하는 것에 관한 것이다.

배 경 기 술

[0002] 중요한 데이터는 종종 컴퓨팅 시스템들의 저장 장치들에 저장된다. 저장 장치들이 고장날 수 있고, 그 고장난 저장 장치들 내의 데이터가 소실될 수 있기 때문에, 하나 이상의 저장 장치들이 고장나는 경우에 데이터 손실을 방지하고 데이터를 복원하기 위한 기술들이 개발되고 있다.

[0003] 데이터 손실을 방지하기 위한 하나의 기술은, 저장 어레이의 멤버인, 저장 장치(디스크 드라이브 등)에 패리티 정보를 저장하는 단계, 및 그 어레이 내의 하나 이상의 다른 나머지 저장 장치들에 고객(customer) 데이터를 저장하는 단계를 포함한다. (여기서, 디스크 드라이브는 공용시에 간소화되는, "디스크"로서 지칭될 수도 있다.) 이러한 기술에 있어서, 저장 장치가 고장나면, 패리티 정보는 그 고장난 저장 장치 상에 있는 데이터를 복원하는데 사용될 수 있다. 또한, 충분한 패리티 정보가 또 다른 저장 장치에 부가되면, 부가적인 패리티 정보는 하나 보다 더 많은 고장난 저장 장치로부터 데이터를 복원하는데 사용될 수 있다. 데이터 미러링이라고 불리는, 데이터 손실을 방지하는 또 다른 기술은, 별개의 저장 장치 상에 데이터의 사본 카피를 생성하는 단계를 포함한다. 저장 장치가 고장나는 경우에, 데이터는 데이터의 카피본으로부터 복원될 수 있다.

[0004] RAID(Redundant Array of Inexpensive(or Independent) Disks)는 증가된 성능 및 용량을 가지는 데이터 저장 시스템을 제공하는데 사용될 수도 있다. 데이터 미러링 및 패리티 정보 저장 또는 양자의 결합물은, 데이터를 보호하기 위하여 RAID 어레이 상에 구현될 수도 있다. 또한, 스트리핑(striping)이라 불리는 기술을 이용할 수도 있으며, 여기서 데이터 기록 및 패리티 정보가 스트립들로 분할되어 스트립들의 개수가 상기 어레이 내의 디스크 개수와 동일하게 된다. 각각의 스트립은 디스크 양단의 부하를 밸런싱하고 성능을 개선시키기 위하여 RAID 어레이의 각각의 다른 디스크들에 기록 또는 "스트립"된다. RAID 내의 드라이브들 전체에 걸쳐서 하나의 통로를 구비하는 스트립들의 그룹은, 스트라이드로 불린다. 몇몇 RAID 프로토콜들은 고안되어 있으며, 여기서 서로 다른 미러링, 패리티 및 스트리핑 장치들이 사용된다. 일례로서, RAID 에서, 6 개의 디스크, 5 개의 데이터 스트립 및 하나의 패리티 스트립으로 이루어지는 5 개의 어레이는, 이 디스크들 전반에 걸쳐서 패리티 정보를 회전시킨 상태로, 6 개의 디스크 전반에 걸쳐서 스트립된다. 이 디스크들 전반에 걸친 패리티의 회전은, 그 어레이에 대한 패리티 업데이트가 디스크들 전반에 걸쳐서 공유됨을 보증한다. RAID(5)는, 1 의 리던던시를 제공하며, 이는 어레이 내의 디스크들 중 임의의 하나 및 이들 중 하나만이 고장나는 경우에 전체 데이터가 복구될 수 있음을 의미한다.

[0005] 하나 보다 더 많은 저장 장치가 고장난 이후에 저장 장치 리던던시를 더 크게 제공하여 데이터 복구를 허용하는 기술들이 공지되어 있지만, 이러한 기술들은 일반적으로 부가적인 저장 장치들에 부가적인 패리티 정보를 저장할 것을 요청하거나(예를 들어, 더 높은 해밍 코드들을 이용함으로써) 또는 부가적인 저장 장치들에 부가적인 미러링을 요청한다. RAID 6 은 RAID 5 와 유사한 배열을 가지지만, 각각의 스트라이드 내에 2 개의 패리티 스트립을 필요로 하므로 2 의 리던던시를 제공한다. 동일한 데이터 저장 용량의 RAID 6 어레이에 대한 저장 효율은, RAID 6 이 부가적인 디스크를 요청하기 때문에 RAID 5 어레이에 저장 효율보다 낮다. 또한, 패리티 정보로부터 손실된 데이터를 복원하는 것은 시간 소모적인 것이 될 수 있다. 따라서, 알려진 기술들은, 증가된 고장(fault) 방지능력(tolerance) 및 고속 데이터 복구에 대한 필요성에 대하여 가중되는 바람직하지 않은 용량과 성능 트레이드오프를 가진다.

발명의 상세한 설명

[0006] 제 1 양태에 따르면, 저장 장치들의 어레이에 데이터를 저장하는 방법이 제공되며, 이 방법은, 제 1 저장 장치 및 제 2 저장 장치에 제 1 스트립을 기록하는 단계; 제 2 저장 장치 및 제 3 저장 장치에 제 2 스트립을 기록하는 단계; 및 제 3 저장 장치 및 제 4 저장 장치에 제 3 스트립을 기록하는 단계를 포함한다.

[0007] 바람직하기로는, 스트립들 중 하나 이상은 패리티 스트립이다. 바람직산 실시형태에서, 이 방법은 파라미터 N 에 대한 값을 설정하는 단계로서, 저장 장치들의 어레이 내의 각각의 저장 장치가 적어도 N 개의 스트립 LBA를 가지는, 상기 설정 단계; 저장될 스트라이드의 개수 j를 식별하는 단계; 3j가 N-1 미만인지를 결정하는 단계; 만일 미만인 경우에는, 그 어레이 내의 제 1 저장 장치의 LBA 에, 그 어레이 내의 제 2 저장 장치의 LBA 에, 그

리고 그 어레이 내의 제 3 저장 장치의 LBA 에 스트립 s1j를 기록하는 단계; 어레이의 제 2 저장 장치의 LBA 에, 어레이의 제 3 저장 장치의 LBA 에 그리고 어레이의 제 4 저장 장치의 LBA 에 스트립 s2j를 기록하는 단계; 어레이의 제 3 저장 장치의 LBA 에, 어레이의 제 4 저장 장치의 LBA 에 그리고 어레이의 제 5 저장 장치의 LBA 에 스트립 s3j를 기록하는 단계를 더 포함한다. 바람직하기로는, 3j 가 N-1 미만이 아니라고 결정되면, 상기 오퍼레이션들은, 제 1 저장 장치의 LBA에 스트립 s1j를 기록하는 단계; 제 2 저장 장치의 LBA에 스트립 s2j를 기록하는 단계; 제 3 저장 장치의 LBA에 스트립 s3j를 기록하는 단계를 더 포함한다. 보다 상세하게는, 상기 오퍼레이션들은, 저장 장치들의 어레이의 각 저장 장치에 대하여, 그 저장 장치의 스트립 LBA 들의 전체 개수를 결정하는 단계; 및 스트립 LBA 들의 가장 작은 전체 개수를 식별하는 단계를 더 포함하며, 여기서 파라미터 N에 대한 값을 설정하는 단계는 스트립 LBA 들의 가장 작은 전체 개수와 동일하게 N 을 설정하는 단계를 포함한다.

[0008] 바람직하기로는, 3j 가 N-1 미만이라고 결정되면, 상기 오퍼레이션들은 어레이의 제 4 저장 장치의 LBA 에, 어레이의 제 5 저장 장치의 LBA 에, 그리고 어레이의 제 6 저장 장치의 LBA 에 스트립 s4j를 기록하는 단계; 제 5 저장 장치의 LBA에, 제 6 저장 장치의 LBA 에 그리고 제 1 저장 장치의 LBA 에 스트립 s5j를 기록하는 단계; 제 6 저장 장치의 LBA 에, 제 1 저장 장치의 LBA 에 그리고 제 2 저장 장치의 LBA 에 스트립 s6j를 기록하는 단계를 더 포함한다. 보다 바람직하기로는, 3j 가 N-1 미만이 아니라고 결정된 경우에, 상기 오퍼레이션들은 제 4 저장 장치의 LBA 에 스트립 s4j를 기록하는 단계; 제 5 저장 장치의 LBA 에 스트립 s5j를 기록하는 단계; 제 6 저장 장치의 LBA 에 스트립 s6j를 기록하는 단계를 더 포함한다.

[0009] 바람직한 실시형태에서, 이 방법은, 파라미터 N 의 값을 설정하는 단계로서, 저장 장치들의 어레이의 각각의 저장 장치는 적어도 N 개의 스트립 LBA를 가지는 것인, 상기 설정 단계; 저장될 스트라이드의 개수 j를 식별하는 단계; 3j 가 N-1 미만인지를 결정하는 단계; 만일 N-1 미만이면, 어레이의 제 1 저장 장치의 LBAj 에, 어레이의 제 2 저장 장치의 LBAj+2에 그리고 어레이의 제 3 저장 장치의 LBAj+1에 스트립 s1j를 기록하는 단계; 어레이의 제 2 저장 장치의 LBAj에 제 3 저장 장치의 LBAj+2에, 어레이의 제 3 저장 장치의 LBAj+2 에 그리고 어레이의 제 4 저장 장치의 LBAj+1에 스트립 s2j를 기록하는 단계; 어레이의 제 3 저장 장치의 LBAj에, 어레이의 제 4 저장 장치의 LBAj+2에 그리고 어레이의 제 5 저장 장치의 LBAj+1에 스트립 s3j를 기록하는 단계를 더 포함한다.

[0010] 바람직하기로는, 3j 가 N-1 미만이 아니라고 결정되면, 상기 오퍼레이션들은 제 1 저장 장치의 LBA(3j-N+2)에 스트립 s1j를 기록하는 단계; 제 2 저장 장치의 LBA(3j-N+2)에 스트립 s2j를 기록하는 단계; 제 3 저장 장치의 LBA(3j-N+2)에 스트립 s3j를 기록하는 단계를 더 포함한다. 보다 바람직하기로는, 상기 오퍼레이션들은 저장 장치들의 어레이의 각각의 저장 장치에 대하여, 그 저장 장치의 스트립 LBA 들의 전체 개수를 결정하는 단계; 및 스트립 LBA 들의 가장 작은 전체 개수를 식별하는 단계를 더 포함하며, 파라미터 N에 대한 값을 설정하는 단계는 스트립 LBA 들의 가장 작은 전체 개수와 동일하게 N 을 설정하는 단계를 포함한다.

[0011] 바람직하기로는, 3j가 N-1 미만이라고 결정되면, 상기 오퍼레이션들은 어레이의 제 4 저장 장치의 LBAj에, 제 5 저장 장치의 LBAj+2에 그리고 제 6 저장 장치의 LBAj+1에 스트립 s4j를 기록하는 단계; 어레이의 제 5 저장 장치의 LBAj에, 제 6 저장 장치의 LBAj+2에, 그리고 제 5 저장 장치의 LBAj+1에 스트립 s5j를 기록하는 단계; 및 제 6 저장 장치의 LBAj에, 제 1 저장 장치의 LBAj+2에, 그리고 제 2 저장 장치의 LBAj+1에 스트립 s6j를 기록하는 단계를 더 포함한다. 보다 바람직하기로는, 3j 가 N-1 미만이 아니라고 결정되면, 상기 오퍼레이션들은, 제 4 기억 장치의 LBA(3j-N+2)에 스트립 s5j를 기록하는 단계; 및 제 6 저장 장치의 LBA(3j-N+2)에 스트립 s6j를 기록하는 단계를 더 포함한다.

[0012] 제 2 양태에 따르면, 저장 장치들의 어레이에 데이터를 저장하는 저장 시스템이 제공되며, 이 시스템은 제 1 저장 장치 및 제 2 저장 장치에 제 1 스트립을 기록하는 수단; 제 2 저장 장치 및 제 3 저장 장치에 제 2 스트립을 기록하는 수단; 및 제 3 저장 장치 및 제 4 저장 장치에 제 3 스트립을 기록하는 수단을 포함한다.

[0013] 제 3 양태에 따르면, 컴퓨터 프로그램이 컴퓨터 상에서 동작하는 경우에, 상기 방법의 전 단계들을 수행하도록 구성되는 프로그램 코드 수단을 구비하는 컴퓨터 프로그램이 제공된다.

[0014] 본 발명의 하나의 양태는, 저장 장치들의 어레이에 데이터를 저장하는 방법이다. 이 방법의 일례는, 제 1 저장 장치와 제 2 저장 장치에 제 1 스트립을 기록하는 단계를 포함한다. 또한, 이 예는 제 2 저장 장치 및 제 3 저장 장치에 제 2 스트립을 기록하는 단계를 포함한다. 이 예는 제 3 저장 장치 및 제 4 저장 장치에 제 3 스트립을 기록하는 단계를 더 포함한다.

[0015] 본 발명의 방법 양태의 일부 다른 예들은, 부가적인 스트립들 및 디스크들에 대하여 디스크 어레이 전반에 걸쳐서 데이터 스트라이드들을 스트립하는 단계, 그 어레이의 제 1 디스크 상에 그 스트라이드의 제 1 스트립을 기

록 또는 업데이트하는 단계, 제 2 디스크 상의 제 2 스트립을 기록 또는 업데이트하는 단계 등을 포함한다. 이 방법은 부가적으로 제 1 디스크가 이 어레이의 마지막 디스크의 스트립의 카피본을 가지며 제 2 디스크가 제 1 디스크 상에 그 스트립의 카피본을 가지도록 하나의 디스크에 의해 회전되는, 스트립들 각각의 카피본을 제조하는 단계를 포함한다.

[0016] 본 발명의 다른 양태들은 아래의 섹션들에 기재되어 있으며, 예를 들어 저장 시스템, 디지털 프로세싱 장치에 의해 실행될 수 있는 머신 판독가능한 명령들의 프로그램을 명확하게 구현하는 신호 운반 매체를 포함하여 저장 장치들의 어레이에 데이터를 저장하는 동작들을 수행한다.

[0017] 본 발명의 몇몇 일례들은, 기본 RAID 구성의 저장 장치들 이외의 저장 장치들을 이용하지 않고, 기본 RAID 구성에 의해 제공되는 방지능력 보다 더 높은 저장 장치 고장 허용 오차를 유리하게 제공한다. 따라서, 본 발명의 일부 일례들은, RAID 내의 이용가능한 디스크 공간만을 이용하여, 주어진 디스크 개수에 대하여 기본 RAID 코드의 상부에 부가적인 리던던시를 더한다. 부가적으로, 본 발명의 일부 일례들은 유익하게도 저장 장치의 초기 사용 동안에 높은 고장 방지능력을 제공하며, 이는 높은 고장 레이트들에 의해 특성화되는 사용 주기이다. 또한, 본 발명의 일부 일례들은, 데이터를 신속하게 복원할 수 있다. 또한, 본 발명은 복수의 다른 이점 및 이익을 제공하며, 이는 이하의 설명부터 명백하게 된다.

실 시 예

[0035] 본 발명의 본질, 목적 및 이점은 첨부된 도면들과 함께 이하의 상세한 설명을 참조한다면 당업자에게 더욱 명확하게 된다.

[0036] Ⅰ. 하드웨어 구성요소 및 상호접속

[0037] 본 발명의 일 양태는 저장 장치들의 어레이에 데이터를 저장하는 저장 시스템에 관한 것이다. 예를 들어, 이 저장 시스템은, 도 1 에 나타낸 저장 시스템(100)의 전체 또는 일부분에 의해 구현될 수도 있다. 예를 들어, 저장 시스템(100)은 IBM(International Business Machines) 주식회사에 의해 제조된 모델 800 ESS(Enterprise Storage Server)를 사용하여 주로 구현될 수도 있다.

[0038] 저장 시스템(100)은 제 1 클러스터(102) 및 제 2 클러스터(104)를 포함한다. 다른 실시형태에서, 저장 시스템(100)은 단일 클러스터 또는 2 개 이상의 클러스터를 포함할 수도 있다. 각각의 클러스터는 하나 이상의 프로세서를 가진다. 예를 들어, 각 클러스터는 4 또는 6 개의 프로세서를 가질 수도 있다. 도 1 에 나타낸 예에서, 제 1 클러스터(102)는 6 개의 프로세서(106a, 106b, 106c, 106d, 106e 및 106f)를 가지며, 2 개의 클러스터(104)는 또한 6 개의 프로세서(108a, 108b, 108c, 108d, 108e 및 108f)를 가진다. 충분한 컴퓨팅 능력을 가지는 임의의 프로세서들을 사용할 수도 있다. 예를 들어, 각각의 프로세서(106a-f, 108a-f)는, IBM 주식회사에 의해 제조된 PowerPC RISC 프로세서일 수도 있다. 또한, 제 1 클러스터(102)는 제 1 메모리(110)를 포함하며, 이와 유사하게 제 2 클러스터(104)는 제 2 메모리(112)를 포함한다. 예를 들어, 메모리들(110, 112)은 RAM 일 수도 있다. 이 메모리들(110, 112)은 예를 들어 데이터, 애플리케이션 프로그램들 및 프로세서(106a-f, 108a-f)에 의해 실행된 다른 프로그래밍 명령들을 저장하는데 사용될 수도 있다. 2 개의 클러스터(102, 104)는 단일 인클로져(enclosure)에 또는 별개의 인클로져에 위치될 수도 있다. 다른 실시형태에서, 각각의 클러스터(102, 104)는 슈퍼컴퓨터, 메인프레임 컴퓨터, 컴퓨터 워크스테이션 및/또는 퍼스널 컴퓨터와 대체될 수 있다.

[0039] 제 1 클러스터(102)는 장치 어댑터(DA1, DA3, DA5, DA7;후술함)의 제 1 그룹에 포함되는, NVRAM(114; non-volatile random access memory)에 결합된다. 이와 유사하게, 제 2 클러스터(104)는 장치 어댑터(DA2, DA4, DA6, DA8; 후술함)의 제 2 그룹에 포함되는, NVRAM(116)에 결합된다. 또한, 제 1 클러스터(102)는 NVRAM(116)에 결합되고, 제 2 클러스터(104)는 NVRAM(114)에 결합된다. 예를 들어, 클러스터(102)에 의해 동작되는 데이터는 메모리(110)에 저장되고, 또한 NVRAM(116)에 저장되어, 만일 클러스터(102)가 동작하지 않게 되면, 데이터는 소실되지 않게 되고 클러스터(104)에 의해 동작될 수 있다. 이와 유사하게, 예를 들어, 클러스터(104)에 의해 동작되는 데이터는 메모리(112)에 저장되고, 또한 NVRAM(114)에 저장되어, 만일 클러스터(104)가 동작하지 않게 되면, 데이터는 소실되지 않게 되고 클러스터(102)에 의해 동작될 수 있다. NVRAM(114, 116)은 예를 들어 전력 없이 약 48 시간까지 데이터를 유지할 수 있다.

[0040] 제 1 클러스터(102) 내에서, 2 개 이상의 프로세서(106a-f)는 동일한 작업들에 함께 종사하기 위하여 그룹화될 수도 있다. 그러나, 작업들은 프로세서들(106a-f) 사이에서 분할될 수 있다. 이와 유사하게, 제 2 클러스터(104)내에서, 2 개 이상의 프로세서(108a-f)는 동일한 작업들에 함께 종사하기 위하여 그룹화될 수도 있다. 다른 방법으로, 작업들은 프로세서들(108a-f) 사이에서 분할될 수도 있다. 2 개의 클러스터(102, 104) 사이의 상

호 작용에 대하여, 클러스터(102, 104)는 독립적으로 작업에 대한 결정을 내린다. 그러나, 작업들은 다른 클러스터(102, 104)들 내의 프로세서들(106a-f, 108a-f)에 의해 공유될 수 있다.

[0041] 제 1 클러스터(102)는 예를 들어 제 1 하드 드라이브(118)와 같은 제 1 부트(boot) 장치에 결합된다. 이와 유사하게, 제 2 클러스터(104)는 예를 들어 제 2 하드 드라이브(120)와 같은 제 2 부트 장치에 결합된다.

[0042] 클러스터들(102, 104) 각각은, 클러스터들(102, 104)에 의해 공유된, 어댑터(12)에 결합된다. 또한, 이 공유된 어댑터(122)들은 호스트 어댑터로 불릴 수 있다. 이 공유된 어댑터(122)들은 예를 들어 PCI 슬롯, PCI 슬롯에 후킹되는 베이(bay)들일 수도 있으며, 이는 어느 하나의 클러스터(102, 104)에 의해 동작될 수 있다. 예를 들어, 이 공유된 어댑터(122)들은 SCSI, ESCON, FICON 또는 파이버 채널 어댑터일 수도 있고, 호스트(124)와 같이 하나 이상의 PC 및/또는 다른 호스트들과의 통신을 용이하게 할 수도 있다. 예를 들어, 호스트(124)는 IBM 주식회사로부터 입수가능한 z시리즈 서버, 또는 Netfinity 서버일 수도 있다.

[0043] 또한, 제 1 클러스터(102)는 장치 어댑터들(DA1, DA3, DA5, DA7)(이는 전용 어댑터들로 불리 수도 있음)의 제 1 그룹에 결합되며, 제 2 클러스터(104)는 장치 어댑터들(DA2, DA4, DA6, DA8)의 제 2 그룹에 결합된다. 장치 어댑터들(DA1, DA3, DA5, DA7) 각각은, 제 1 클러스터(102)와 저장 장치 그룹들(126a, 126b, 126c, 126d)중 하나의 그룹 사이의 인터페이스이며, 이와 유사하게 장치 어댑터들(DA2, DA4, DA6, DA8) 각각은, 제 2 클러스터(104)와 저장 장치 그룹들(126a, 126b, 126c, 126d)중 하나의 그룹 사이의 인터페이스이다. 보다 자세하게는, 장치 어댑터들(DA1 및 DA2)은 저장 장치 그룹(126a)에 결합되며, 장치 어댑터들(DA3 및 DA4)는 저장 장치 그룹(126b)에 결합되며, 장치 어댑터들(DA5 및 DA6)은 저장 장치 그룹(126c)에 결합되며, 장치 어댑터들(DA7 및 DA8)은 저장 장치 그룹(126d)에 결합된다. 다른 실시형태에서, 더 크거나 또는 더 작은 개수의 장치 어댑터들(DA1-8) 및 저장 장치 그룹(126a-d)들이 사용될 수 있다. 저장 장치 그룹들(126a-d)은 클러스터(102, 104)에 의해 공유된다. 다른 실시형태에서, 하나 이상의 저장 장치 그룹들이 제 1 클러스터(102)와 제 2 클러스터(104)와 다른 사이트에 위치될 수 있다.

[0044] 예를 들어, 각각의 (저장) 장치 어댑터(DA1-8)는 SSA(Serial Storage Architecture) 어댑터일 수도 있다. 다른 방법으로, 하나 이상의 장치 어댑터(DA1-8)들은, 예를 들어 SCSI 또는 파이버 채널 어댑터들과 같은 다른 타입의 어댑터로 구현될 수 있다. 각각의 어댑터(DA1-8)는, 본 발명의 하나 이상의 일례 또는 본 발명의 일부분들을 수행하기 위한, 소프트웨어, 펌웨어 및/또는 마이크로코드를 포함할 수도 있다. 예를 들어, CPI(Commmon Parts Interconnect)는 각각의 클러스터(102, 104)에 각각의 장치 어댑터(DA1-8)를 결합하는데 사용될 수도 있다.

[0045] 장치 어댑터들의 각 쌍(DA1 및 DA2, DA3 및 DA4, DA5 및 DA6, DA7 그리고 DA8)은 저장 장치들의 2 개의 루프에 결합된다. 예를 들어, 장치 어댑터들(DA1 및 DA2)은, 저장 장치들(A1, A2, A3, A4, A5, A6, A7, A8)의 제 1 스트링 및 저장 장치들(B1, B2, B3, B4, B5, B6, B7, B8)의 제 2 스트링을 포함하는 저장 장치들의 제 1 루프에 결합된다. 루프 상태의 저장 장치들의 제 1 및 제 2 스트링은 통상적으로 저장 장치들의 개수와 동일한 개수를 가지므로 균형잡힌 루프를 유지한다. 이와 유사하게, 장치 어댑터들(DA1 및 DA2)은 또한 저장 장치들(C1, C2, C3, C4, C5, C6, C7, C8)의 제 1 스트링 및 저장 장치들(D1, D2, D3, D4, D5, D6, D7, D8)의 제 2 스트링을 포함하는 저장 장치들의 제 2 루프에 결합된다. 저장 장치들(A1, A2, A3, A4, A5, A6, A7 및 A8)과 같은 8 개의 저장 장치의 집합은 8 팩으로 지칭될 수도 있다. 비록 요청되지는 않지만, 루프는 일반적으로 최소 6 개의 저장 장치를 가진다. 다른 실시형태들에서, 더 큰 또는 더 작은 개수의 저장 장치들이 각각의 루프에 포함될 수 있다. 예를 들어, 32, 48 또는 다른 개수의 저장 장치들이 각각의 루프에 포함될 수 있다. 통상적으로, 루프 내의 저장 장치들의 스트링들은 저장 장치들의 개수와 동일한 개수를 가진다. 저장 장치들의 각각 루프는, 저장 장치들의 루프가 결합되는 각각의 장치 어댑터을 사용하여 직렬 루프를 형성한다. 예를 들어, 저장 장치들(A1, A2, A3, A4, A5, A6, A7 및 A8; B1, B2, B3, B4, B5, B6, B7, B8)을 포함하는 저장 장치들의 루프는, 장치 어댑터(DA1)를 사용하여 직렬 루프를 형성하고, 또한 장치 어댑터(DA2)를 사용하여 직렬 루프를 형성한다. 이러한 배열은, 각각의 직렬 루프가 루프 내의 각각의 저장 장치와 루프에 결합되는 각각의 장치 어댑터 사이에 중복적인 통신 경로들을 제공하기 때문에 신뢰성을 증가시킨다.

[0046] 저장 장치들(126a, 126b, 126c, 126d)의 각각의 그룹 내의 저장 장치들은, 하나 이상의 저장 장치 어레이로 그룹화될 수 있으며, 이들 각각은 예를 들어 RAID(Redundant Array of Inexpensive(or Independent) Disks)일 수도 있다. 또한, RAID 어레이들은 RAID 랭크로 불릴 수도 있다. 제 1 및 제 2 클러스터(102, 104)(또는 호스트(124))로부터 수신된 판독 및 기록 요청에 응답하여, (저장) 장치 어댑터들(DA1-8)은 이들이 결합되는 RAID 어레이들의 각 저장 장치를 개별적으로 어드레스화할 수 있다. 특정 RAID 어레이 내의 저장 장치들은, 장치 어댑

터들의 쌍 사이에서, 동일한 루프에 또는 다른 루프들에 존재할 수도 있다. 예를 들어, RAID 어레이들이 단일 루프 내에 있는 저장 장치들로부터 형성되고, 제 1 RAID 어레이는 저장 장치들(A1, A2, A3, A4, B1, B2 및 B3)을 포함할 수도 있고, 제 2 RAID 어레이는 저장 장치들(A6, A7, A8, B5, B6, B7 및 B8)을 포함할 수도 있으며, 저장 장치들(B4 및 A5)은 어느 하나의 RAID 어레이에 의해 사용될 수 있는 예비품으로서 지정된다. 이 예에서, 각 RAID 어레이는 A1, A2, A3, A4, A5, A6, A7, A8의 8 팩 및 B1, B2, B3, B4, B5, B6, B7, B8의 8 팩으로부터 저장 장치들을 포함하므로, 각각의 RAID 어레이는 장치 어댑터들(DA1, DA2)중 하나에 근접한다. 예를 들어, RAID 어레이들이 다른 루프들에 있는 저장 장치들로부터 형성되며, 제 1 RAID 어레이는 저장 장치들(A1, A2, B1, B2, C1, C2 및 D1)을 포함할 수도 있고, 제 2 RAID 어레이는 저장 장치들(A3, A4, B3, B4, C3, D3 및 D4)을 포함할 수도 있고, 제 3 RAID 어레이는 저장 장치들(A5, A6, B6, C5, C6, D5 및 D6)을 포함할 수도 있고, 제 4 RAID 어레이는 저장 장치들(A8, B7, B8, C7, C8, D7 및 D8)을 포함할 수도 있으며, 저장 장치들(D2, C4, B5 및 A7) 은 임의의 4 개의 RAID 어레이에 의해 사용될 수 있는 예비품으로서 지정된다. 이 예에서, RAID 어레이들 및 이 RAID 어레이들에 대하여 이용가능한 예비 저장 장치들은 동일한 쌍의 장치 어댑터들에 결합된다. 그러나, RAID 어레이, 이 RAID 어레이에 대하여 이용가능한 예비 저장 장치들은 다른 쌍의 장치 어댑터들에 결합될 수 있다. 또한, RAID 어레이 및 이 RAID 어레이에 대하여 이용가능한 예비 저장 장치들은 단일 루프 또는 다른 루프들에 존재할 수도 있다.

[0047] 데이터 및 만일 요청되는 경우의, 패리티 정보는 임의의 원하는 배열로 RAID 어레이의 저장 장치들에 저장될 수 있으며, 이는 RAID 어레이 내의 저장 장치들의 전체 또는 일부에 걸쳐서 스트리핑 및/또는 미러링하는 것을 포함할 수도 있다. 예를 들어, RAID 어레이 내의 6 개의 저장 장치는 데이터를 저장하는데 사용될 수도 있으며, RAID 어레이 내의 7 번째 저장 장치는 패리티 정보를 저장하는데 사용될 수도 있다. 또 다른 일례에서, RAID 어레이 내의 7 개의 저장 장치들은 데이터를 저장하는데 사용될 수 있고, RAID 어레이 내의 8 번째 저장 장치는 패리티 정보를 저장하는데 사용될 수 있다. 또 다른 예로서, 데이터 및 패리티 정보가 RAID 어레이 내의 저장 장치들 전체에 저장될 수도 있다. 다른 실시형태에서, RAID 어레이들은 7 개 미만의 저장 장치 또는 8 개 보다 많은 저장 장치를 가져야 한다. 예를 들어, RAID 어레이는 데이터 및 패리티 정보를 저장하는데 각각 사용되는 5 개 도는 6 개의 저장 장치로 이루어져 있다. 또한, 더블(double) 패리티 정보는, 제 1 저장 장치 고장 이후에 복원을 완료하기 이전에 발생하는 제 2 저장 장치 고장으로부터 복원을 허용하도록 저장될 수도 있다. 예를 들어, RAID 어레이는 데이터를 저장하는데 사용되는 6 개의 저장 장치 및 패리티 정보를 저장하는데 사용되는 2 개의 저장 장치로 이루어져 있다. 또 다른 예에서, 7 개의 저장 장치는 데이터에 대하여 사용될 수 있으며, 또 다른 7 개의 저장 장치들은 첫번째 7 개의 저장 장치에 데이터를 미러링하는데 사용될 수 있으며, 2 개 보다 많은 저장 장치들이 패리티 정보를 저장하는데 사용될 수 있으며, 이들 모두는 함께 9 개의 저장 장치의 고장(9 개의 고장 허용 오차)으로부터 복원시킨다.

[0048] 일반적으로, 저장 장치 그룹(126a-d) 내의 저장 장치들은 데이터를 저장하기 위한 임의의 적절한 장치들일 수도 있으며, 자기, 광학, 자기-광학, 전기 또는 데이터를 저장하기 위한 임의의 다른 적절한 기술을 이용할 수도 있다. 예를 들어, 저장 장치들은 하드 디스크 드라이브, 광 디스크 또는 디스크들(예를 들어, CD-R, CD-RW, WORM, DVD-R, DVD+R, DVD-RW 또는 DVD+RW), 플로피 디스크, 자기 데이터 저장 디스크 또는 디스켓, 자기 테이프, 디지털 광 테이프, EPROM, EEPROM 또는 플래시 메모리일 수 있다. 이 저장 장치들은 각각 동일한 타입의 장치를 가지거나 또는 동일한 타입의 기술을 이용할 필요는 없다. 예를 들어, 각 저장 장치는 예를 들어 146 기가 바이트의 용량을 가지는 하드 드라이브일 수 도 있다. 일 례에서, 각 저장 장치 그룹(126a-d)은, IBM 주식회사에 의해 제조되는 모델 2105 엔터프라이즈 저장 서버에 저장 인클로져일 수도 있다.

[0049] 하나 이상의 장치 어댑터(DA1-8)와 하나 이상의 저장 장치 그룹(126a-d)의 적어도 일부분을 함께 가지는 제 1 클러스터(102) 및/또는 제 2 클러스터(104)는, 저장 시스템 또는 저장 장치로 지칭될 수 있다. 하나 이상의 저장 장치 그룹(126a-d)의 일부분을 가지거나 또는 이것이 없는, 하나 이상의 장치 어댑터(DA1-8)는 저장 시스템 또는 저장 장치로서 지칭될 수도 있다.

[0050] 예시적인 컴퓨터 장치(200)를 도 2 에 도시한다. 예를 들어, 호스트(124)(및 다른 실시형태에서) 클러스터(102) 및/또는 클러스터(104)는 컴퓨팅 장치(200)의 실시형태로 구현될 수 있다. 이 컴퓨팅 장치(200)는 프로세서(202)(프로세싱 장치로 불릴 수도 있음)를 포함하며, 몇몇 일례에서, 하나 보다 더 많은 프로세서(202)를 포함할 수 있다. 예를 들어, 프로세서는 IBM 주식회사로부터 입수가능한 PowerPC RISC 프로세서 또는 인텔 주식회사에 의해 제조된 프로세서일 수도 있다. 이 프로세서(202)는, 예를 들어 윈도우 2000, AIX, SolarisTM, 리눅스, 유닛스 또는 HP-UX™ 과 같은 임의의 적절한 동작 시스템을 동작시킬 수 있다. 컴퓨팅 장치(200)는 예를 들어 퍼스널 컴퓨터, 워크스테이션, 메인프레임 컴퓨터 또는 슈퍼컴퓨터인 임의의 적절한 컴퓨터 상에 구현

될 수 있다. 또한, 컴퓨팅 장치(200)는 프로세서(202)에 모두 결합되는, 저장 장치(204), 네트워크 인터페이스(206) 및 입/출력 장치(208)를 포함한다. 저장 장치(204)는 예를 들어 RAM 및 비휘발성 메모리(21)일 수 있는 일차(primary) 메모리(210)를 포함할 수도 있다. 이 비휘발성 메모리(210)는 예를 들어 하드 디스크 드라이브, 광학 매체 또는 자기-광학 매체로부터의 판독 및 기록을 위한 드라이브, 테이프 드라이브, 비휘발성 RAM(NVRAM) 또는 임의의 다른 적절한 타입의 저장 장치일 수 있다. 저장 장치(204)는 데이터 및 프로세서에 의해 실행되는 애플리케이션 프로그램 및/또는 다른 프로그래밍 명령들을 저장하는데 사용될 수도 있다. 네트워크 인터페이스(206)는 임의의 적절한 유선 또는 무선 네트워크 또는 통신 링크에의 액세스를 제공할 수도 있다.

[0051] Ⅱ. 동작

[0052] 상술된 하드웨어 실시형태 이외에, 본 발명의 다른 양태들은 저장 장치들의 어레이에 데이터를 저장하는 동작에 관한 것이다.

[0053] A. 신호 운반 매체

[0054] 도 1 및 도 2 와 관련하여, 본 발명의 방법 양태들은, 장치 어댑터(DA1-8), 클러스터(102) 및/또는 클러스터(104)(및/또는 호스트(124)) 중 하나 이상을 가짐으로써 구현될 수 있고, 코드로서 지칭되기도 하는 머신 판독가능한 명령들의 시퀀스를 실행한다. 이 명령들은 다양한 타입의 신호 운반 매체에 존재한다. 이와 관련하여, 본 발명의 일부 양태들은, 신호 운반 매체 또는 디지털 프로세싱 장치에 의해 실행가능한 머신 판독가능한 명령들의 프로그램을 명백하게 구현하는 신호 운반 매체를 구비하여 저장 장치들의 어레이에 데이터를 저장하는 동작들을 수행하는 프로그램된 제품에 관한 것이다.

[0055] 이 신호 운반 매체는, 예를 들어 RAM(110), RAM(112), NVRAM(114), NVRAM(116), 1 차(primary) 메모리(210), 비휘발성 메모리(212) 및/또는 장치 얻댑터(DA1-8)내의 펌웨어를 구비할 수도 있다. 다른 방법으로, 이 명령들은 도 3 에 나타낸 광 데이터 저장 디스크(300)와 같은 신호 운반 매체에 구현될 수도 있다. 광 디스크는 예를 들어 CD-ROM, CD-R, CD-RW, WORM, DVD-R, DVD+R, DVD-RW 또는 DVD+RW과 같은 임의의 타입의 신호 운반 디스크일 수 있다. 또한, 저장 시스템(100) 또는 그 밖의 장치에 포함되어 있는지 여부에 대해서, 이 명령들은, 예를 들어 "하드 드라이브", RAID 어레이, 자기 데이터 저장 디스켓(플로피 디스크 등), 자기 테이프, 디지털 광 테이프, RAM, ROM, EPROM, EEPROM, 플래시 메모리, 프로그램가능한 로직, 임의의 다른 타입의 펌웨어, 자기-광학 저장 장치, 페이퍼 펀치 카드, 또는 전기, 광학 및/또는 무선 링크일 수도 있는 디지털 및/또는 아날로그 통신 링크와 같은 송신 매체를 포함하는 임의의 다른 적절한 신호 운반 매체를 포함할 수도 있는 임의의 여러가지 머신 판독가능한 데이터 저장 매체들 또는 미디어에 저장될 수도 있다. 예를 들어, 일부 실시형태에서, 명령들 또는 코드는 네트워크를 통하여 파일 서버로부터 또는 다른 송신 매체로부터 액세스될 수 있으며, 이 명령들 또는 코드를 포함하는 신호 운반 매체는, 네트워크 송신 라인, 무선 송신 미디어, 공간을 통하여 전파하는 신호, 무선파 및/또는 적외선 신호들과 같은 송신 매체를 포함할 수도 있다. 다른 방법으로, 신호 운반 매체는 예를 들어 집적 회로 칩, PGA(Programmable Gate Array) 또는 ASIC과 같은 하드웨어 로직에 구현될 수 있다. 예를 들어, 머신 판독가능한 명령들은 마이크로코드를 포함하거나 또는 "C++"과 같은 언어로부터 컴파일된 소프트웨어 오브젝트 코드를 포함할 수도 있다.

[0056] B. 동작의 전체 시퀀스

[0057] 1. 동작 시퀀스의 제 1 실시예

[0058] 설명의 편의를 위하여, 임의의 의도된 제한없이, 본 발명의 예시적인 방법 양태들을 도 1 에 나타내고 상술한 저장 시스템(100)을 참조하여 설명한다. 본 발명의 방법 양태의 일례를 도 4 에 나타내며, 이는 저장 장치들의 어레이에 데이터를 저장하는 방법에 대한 시퀀스(400)를 나타낸다.

[0059] 시퀀스(400)의 동작들은 장치 어댑터(DA1-8), 클러스터(102) 및/또는 클러스터(104)(및/또는 호스트(104))중 하나 이상에 의해 수행된다. 도 4 를 참조하면, 시퀀스(400)는 오퍼레이션(402)을 포함하며, 이 오퍼레이션(402)으로 개시한다. 오퍼레이션 (402)은 최대 스트립 개수이며, 이들과 연관된 LBA(Logical Block Address)에 의해 식별되는, 어레이에 대한 값 "N"을 결정하는 단계를 포함하며, 이 LBA는 어레이의 자장 장치 각각에 저장될 수 있다. 예를 들어, 저장 장치들의 어레이는 저장 장치 그룹들(126a-d) 중 하나 이상에 저장 장치들의 일부 또는 전부를 포함할 수도 있다. 상술한 바와 같이, 몇몇 예에 있어서, 저장 장치들은 하드 디스크 드라이브일 수도 있다.

[0060] 저장 장치들의 어레이의 저장 장치들에 기록되는 스트립들의 최대 개수 N을 결정하기 위하여, 저장 어댑터는 어레이의 각각의 장치에 질의(query)한 후, 이 어레이의 가장 작은 용량의 저장 장치가 지원할 수 있는 최대값과

동일하게 스트립들의 개수 N을 설정한다. 그러나, 다른 일례들에서, 저장 어댑터는 최대값을 더 작은 값으로 제한할 수도 있다. 대부분의 경우에, RAID 어레이의 저장 장치들 전체는 동일한 저장 용량을 가지며, 동일한 개수의 이용가능한 스트립 LBA 들을 가질 수 있다.

[0061] 통상적으로, 각각의 스트립은 복수의 데이터 블록을 포함하며, 여기서 각 데이터 블록은 대응하는 LBA에 저장된다. 스트립의 제 1 블록의 LBA 는 스트립 LBA로 불린다. 예를 들어, 각 스트립은 64 개 블록을 포함할 수도 있으며, 여기서 각각의 블록은, 예를 들어 데이터의 512 바이트를 포함한다. 스트립의 각 데이터 블록은, 대응하는 스트립 LBA 과 블록 오프셋의 합으로 어드레싱될 수 있으며, 스트립 LBA 는 스트립의 제 1 데이터 블록의 어드레스이며, 오프셋은 스트립 LBA로부터 타켓 데이터 블록 LBA 까지의 블록 개수이다. 통상적으로 스트립들이 공통적인 길이를 가지기 때문에, 스트라이드의 각 스트립의 개시 LBA 는 통상적으로 어레이의 각 저장 장치에 대하여 동일한 값을 가진다. 따라서, 스트라이드의 전체 데이터 블록들은, 타겟 저장 장치(예를 들어 디스크), 스트립 LBA 및 오프셋을 식별함으로써 어드레스화될 수 있다. "스트립 LBA에 기록" 이라는 용어는 주어진 스트립 LBA 에서 개시하는 스트립과 연관되는 임의의 블록 또는 블록들의 전체에의 기록을 설명하기 위한 속기법(shorthand)로서 사용된다.

[0062] 또한, 시퀀스(400)는, 저장 장치들의 어레이의 새로운 LBA에의 기록 횟수의 카운트를 유지하기 위하여, 1 과 같은 초기값으로 카운터를 설정하는 단계를 포함하는 오퍼레이션(404)을 포함할 수도 있다.

[0063] 또한, 시퀀스(400)는 랜덤한 입력 기록 LBA들과 이 어레이의 저장 장치들에 기록되는 순서화된 LBA 들 사이에 1 대 1 매핑을 확립하는 단계를 포함하는, 오퍼레이션(406)을 포함할 수도 있다. 오퍼레이션(406)은 매핑 알고리즘에 기초하는, 매핑 테이블을 설정하는 단계를 포함할 수도 있다. 또한, 매핑 테이블을 설정하는 것은 매핑 테이블을 지정하는 것으로 불리며, 캐시 내에 공간을 유지하는 것을 포함한다. 일례로서, 매핑 테이블은 어댑터 메모리에 저장될 수도 있다.

[0064] 어댑터 메모리는 비휘발성 메모리일 수도 있으므로, 매핑 테이블은 저장 장치(예를 들어, 디스크)가 리셋되는 경우에 소실되지 않는다.

[0065] 랜덤한 입력 기록 LBA들과 이 어레이의 저장 장치들에 기록되는 순서화된 LBA 들 사이의 1 대 1 매핑을 확립하면, 회전된 카피본에 대하여 인접한 LBA 들을 유지하는 알고리즘을 포함할 수 있다. 도 5 에 나타낸 알고리즘은 데이터 및 5 개의 디스크 어레이에 그 데이터의 단일 회전된 카피본을 기록하기 위한 알고리즘의 일례이다. 이러한 알고리즘을 이용하여, 인접한 LBA 들은 회전된 카피본들에 대하여 저장되며, 개선된 판독 및 기록 효율을 제공한다. 그러나, 일반적으로는 임의의 1 대 1 매핑 알고리즘이 사용될 수도 있다. 도 5 와 관련하여, s1j, s2j, s3j, s4j 및 s5j는, Sj=s1j+s2j+s3j+s4j+s5j 가 되도록 스트립 Sj를 구성하는 스트립들이다. 또한, LBAm 은 매핑 알고리즘 및 테이블(도 6 에 나타냄)에 의해 결정되는 바와 같이 스트라이드 Sj에 대하여 매핑된 LBA 이다. 도 5 에 관하여, 스트라이드 Sj를 기록하는 것은, 각각의 디스크의 2 개의 스트립 LBA에 기록하는 것을 포함하며, 여기서 제 2 LBA 에의 기록은 또 다른 디스크에 기록되는 데이터의 회전된 카피본이다. 예를 들어, 디스크(1)상에, 스트라이드 Sj를 기록하는 경우에, LBAm 에서 개시하는 스트립 s1j가 기록되며, LBAm+1 에서 개시하는 스트립 s5j의 카피본이 기록된다. 디스크(2)상에서, LBAm에서 개시하는 스트립 s2j 가 기록되며, LBAm+1에서 개시하는 스트립 s1j의 카피본이 기록된다. 디스크(3)상에서, LBAm에서 개시하는 스트립 s3j 가 기록되며, LBAm+1에서 개시하는 스트립 s2j의 카피본이 기록된다. 디스크(4)상에서, LBAm에서 개시하는 스트립 s4j 가 기록되며, 스트립 s3j의 카피본이 LBAm+1에 기록된다. 디스크(5)상에서, 스트립 s5j 가 LBAm에서 개시하여 기록되고, 스트립 s4j 의 카피본이 LBAm+1에 기록된다. 개시하는 LBA들은 각 스트립의 블록들의 개수의 함수이다. 예를 들어, 스트라이드 1 은 LBA 0 에서 개시할 수 있고, 스트라이드 2 는 LBA 128 에서 개시할 수 있다. 도 6 은 이용가능한 스트립 LBA 들 전체에 대한 FIFO(first-in-first-out) 접근방법을 이용하는, 각 스트라이드의 각 스트립의 단일 회전된 카피본을 저장하기 위한, 도 5 에 나타낸 알고리짐에 기초한 LBA 매핑 테이블을 나타낸다.

[0066] 또 다른 예로서, 도 7 은 매핑 알고리즘을 나타내고, 도 8 은 대응하는 매핑 테이블을 나타내며, 여기서 FIFO 접근법은 5 개의 디스크 어레이에 데이터의 2 개의 회전된 카피본의 저장을 구현하는데 사용된다. (다른 실시형태들에서, 2 개의 회전된 카피본 보다 많은 카피본이 저장될 수 있다.) 도 7 을 참조하면, s1j, s2j, s3j, s4j 및 s5j 는, Sj=s1j+s2j+s3j+s4j+s5j가 되도록 스트라이드 Sj를 구성하는 스트립들이다. 또한, LBAm는 매핑 알고리즘 및 테이블에 의해 결정되는 바와 같이 스트라이드 Sj에 대하여 매핑된 LBA이다. 도 8 을 참조하면, 스트라이드 Sj를 기록하는 것은, 각각의 디스크의 3 개의 LBA에 기록하는 것을 포함하며, 여기서 제 2 및 제 3 LBA 에의 기록은 다른 디스크들에 기록되는 스트립들의 회전된 카피본이다. 예를 들어, 디스크 1 상

에서, 스트라이드 Sj를 기록하는 경우에, 스트립 s1j 가 LBAm 에서 개시하여 기록되고, 스트립 s5j의 카피본이 LBAm+1에서 개시하여 기록되며, 스트립 s4j의 카피본이 LBAm+2 에서 개시하여 기록된다. 디스크 2 상에서, 스트립 s2j 는 LBAm에서 개시하여 기록되며, 스트립 s1j의 카피본이 LBAm+1에서 개시하여 기록되며, 스트립 s5j의 카피본이 LBAm+2 에서 개시하여 기록된다. 디스크 3 상에서, 스트립 s3j는 LBAm에서 개시하여 기록되고, 스트립 s2j의 카피본은 LBAm+1에서 개시하여 기록되며, 스트립 s1j의 카피본은 LBAm+2에서 개시하여 기록된다. 디스크 4 상에서, 스트립 s4j가 LBAm 에서 개시하여 기록되며, 스트립 s3j의 카피본이 LBAm+1에 기록되며, 스트립 s2j의 카피본이 LBAm+2에서 개기하여 기록된다. 디스크 5 상에서, 스트립 s5j 가 LBAm 에서 개시하여 기록되며, 스트립 s4j의 카피본이 LBAm+1 에 기록되며, 스트립 s3j 의 카피본이 LBAm+2 에서 개시하여 기록된다.

[0067] 또 다른 일례에서, 매핑 알고리즘은 밴드에 대하여 매핑된 LBA들의 세트 또는 입력(incoming) 기록 LBA 들의 밴드들의 세트를 유지할 수도 있다. 예를 들어, LBA들은 입력 기록 LBA 들이 논리적으로 서로 근접하게 유지되도록 하는 방식으로 유지될 수도 있다. 일부 일례들에서, 알고리즘은 특정 애플리케이션 및/또는 동작 시스템에 대한 동작에 대하여 변경될 수 있다. 이 예에서, LBA들의 밴드가 저장되며, 이 저장된 밴드에 존재하지 않는 LBA들은, 예를 들어 FIFO 접근법을 이용할 수도 있다. 도 9 는 첫번째 10 개의 LBA들에 대하여 저장된 LBA 밴드 매핑의 일례를 나타낸다. 도 10 은 하나의 회전된 카피본에 대한, 매핑 테이블을 나타내며, 여기서 도 10 에 나타낸 10 개의 LBA 의 밴드의 저장된 매핑 및 FIFO 매핑이 결합된다. 이 예에서, 매핑 테이블은, 입력 기록 LBA 가 이 매핑 테이블에 이미 존재하지 않는 경우에만 업데이트된다. FIFO 알고리즘이 저장된 밴드의 외부에 있는 LBA 들에 대하여 사용된다. 저장된 밴드들을 이용한다는 개념은 일반화될 수 있으며, 하나의 밴드 보다 더 많은 밴드를 포함하도록 확장될 수도 있다.

[0068] 각 스트로브의 오리지널 및 하나의 카피본이 저장되는 실시형태들에 대하여, 오퍼레이션들은 또한 1 차 데이터에 대하여 이용가능한 LBA들의 절반을 저장하는 단계, 및 데이터의 회전된 카피본들에 대하여 이용가능한 LBA들의 절반을 저장하는 단계를 포함한다. 각 스트로브의 오리지널 카피본 및 2 개의 카피본이 저장되는 실시형태들에 대하여, 오퍼레이션들은 또한 1 차 데이터에 대하여 이용가능한 LBA들의 1/3을 저장하는 단계, 및 데이터의 회전된 카피본들에 대하여 이용가능한 LBA 들의 2/3를 저장하는 단계를 포함할 수도 있다. 저장 공간의 보유는 도 5 내지 도 10 에 나타낸 알고리즘 및 테이블과 같은 1 대 1 매핑 알고리즘 및 테이블을 이용하여 저장 장치 어댑터(DA1-8)에 의해 은연중에 수행될 수도 있다. 클러스터(102, 104)로부터 수신된 데이터를 기록하기 위한 요청에 응답하여, 저장 장치 어댑터(DA1-8)는 데이터의 제 1 카피본 및 임의의 제 2 카피본을 기록할 수도 있고, 기록되는 것의 트랙을 유지할 수도 있으며, 여기서 카피본은 매핑 테이블을 이용하여 기록된다.

[0069] 도 4 를 참조하면, 시퀀스(400)는 기록 명령이 수신되었는지를 결정하는 단계를 포함하는 오퍼레이션(408)을 포함할 수도 있다. 기록 명령이 수신되지 않으면, 오퍼레이션(408)은 기록 명령이 수신될 때 까지 반복될 수도 있다. 기록 명령이 수신되면, 시퀀스(400)는, 이전에 기록되지 않은 LBA(새로운 LBA)에 기록이 행해졌는지를 판정하는 단계를 포함하는 오퍼레이션(410)을 포함할 수도 있다. 이전에 기록되었던 LBA 에 기록이 행해졌다고 판정되면, 시퀀스(400)는 매핑 테이블을 체킹하는 단계를 포함하는 오퍼레이션(412), 및 매핑 테이블에 따라 스트립들을 기록하기 위하여 기록을 실행하는 단계를 포함하는 오퍼레이션(413)을 포함할 수도 있다. 기록을 실행하는 단계는, 스트라이드의 각 스트립에 대하여, 매핑 테이블에 표시되는 LBA 에 스트립을 기록하는 단계, 및 대응하는 카피 플래그의 값이 "예"인 경우에, 매핑 테이블에 표시되는 바와 같이 각각의 스트립의 하나 이상의 회전된 카피본을 기록하는 단계를 포함한다.

[0070] 만약 오퍼레이션(410)에서 이전에 기록되지 않았던 LBA에 기록하는 것이 결정된다면, 시퀀스(400) 또한 오퍼레이션(414)을 포함할 수 있는데, 그 오퍼레이션(414)은 카운터를 증가시키는 것을 포함한다. 시퀀스(400)는 또한 스트립 LBA 및 맵핑된 스트립 LBA 사이에서 그 맵핑을 지시하는 맵핑 테이블을 업데이트하는 것을 포함한다. 오퍼레이션(416)은 또한 맵핑 테이블 내의 대응하는 엔트리를 위한 카피 플래그에 대한 "예" 또는 "아니오" 값을 설정하는 것을 포함한다. "예" 또는 "아니오" 값을 설정하는 것은 설정하기 위한 값을 결정하는 것을 포함한다. 예컨대, 카피 플래그가 "아니오" 값으로 설정되어야 하는지 여부를 결정하는 것은 카운터가, 비복사(no-copy) 임계 값보다 크거나 동일한(또한, 보다 작지 않은 경우로 설명될 수도 있음) 값을 구비하는지 여부를 결정하는 것을 포함한다. 예컨대, 비-복사 임계값이 N의 퍼센티지일 수도 있는데, 여기서 퍼센티지는 맵핑 알고리즘의 함수이다. 예컨대, 도 6의 맵핑 테이블에 대하여, 카피 플래그는 카운터가 N/2+1의 값에 도달하는 경우 "아니오"로 설정될 것이다. 시퀀스(400)는 또한 오퍼레이션(418)을 포함하는데, 그 오퍼레이션은 카운터에 대응하는 값에 대하여 카피 플래그가 "예" 또는 "아니오" 인지 여부를 결정하는 것을 포함한다. 만약 카피 플래그의 값이 "예" 인 경우, 시퀀스(400)는 오퍼레이션(420)을 포함하고, 그 오퍼레이션(420)은 스트라이드 내의

각 스트립에 대하여 스트립 및 그 스트립의 회전된 카피본을 맵핑 테이블 내의 지시된 LBA들로 기록하는 것을 포함한다. 시퀀스(400)는 또한 오퍼레이션(422)을 포함하는데, 이 오퍼레이션은 카운터가 N과 동일한 값을 구비하는지 여부를 결정하는 것을 포함하고, 만약 그렇다면, 시퀀스가 종료하고, 그렇지 않다면 시퀀스가 오퍼레이션(408)에서 계속될 수 있다.

[0071] 만약 오퍼레이션(418)에서 카피 플래그가 그 카운터의 대응하는 값에 대하여 "아니오" 값을 구비한다고 결정된다면, 시퀀스(400)는 또한 스트라이드 내의 각 스트립에 대하여, 그 스트립들의 어떠한 복사도 기록함이 없이, 맵핑 테이블 내에 지시된 LBA로 스트립을 기록하는 것을 포함하는 오퍼레이션(424)을 포함한다. 시퀀스(400)는 또한 카운터가 N과 동일한 값을 구비하는지 여부를 결정하는 것을 포함하고, 만약 그렇다면 시퀀스는 종료할 수 있고, 그렇지 않다면 그 시퀀스는 오퍼레이션(408)에서 계속될 수 있다.

[0072] 2. 동작 시퀀스의 제2 실시예

[0073] 도 11은 저장 장치의 어레이 내에 데이터를 저장하기 위한 방법에 대한 시퀀스(1100)에 대한 흐름도이다. 시퀀스(1100)의 동작들은 하나 이상의 장치 어댑터(DA1-8), 클러스터(102), 및/또는 클러스터(104)[및/또는 호스트(104)]에 의해 실행될 수 있다. 도 11a를 참조하면, 시퀀스(1100)는 오퍼레이션(1102)를 포함할 수 있고, 그 오퍼레이션과 함께 시작될 수도 있는데, 그 오퍼레이션(1102)은 저장 장치들의 어레이 내의 각 저장 장치에 대하여, 그 관련 논리 블럭 어드레스들(LBAs)에 의해 식별되는 스트립들의 전체 수를 결정하는 것을 포함하고, 그 관련 논리 블록 어드레스들은 저장 장치에 저장될 수 있다. 이것은 또한 어레이 내의 각 저장 장치 상의 스트립 LBA들의 전체 수를 결정하는 것으로 기술될 수도 있다. 예컨대, 저장 장치들의 어레이는 하나 이상의 저장 장치 그룹들(126a-d) 내의 저장 장치의 일부 또는 전부를 포함할 수 있다.

[0074] 시퀀스(1100)는 또한 오퍼레이션(1104)를 포함할 수도 있는데, 이 오퍼레이션은 가장 작은 용량을 구비한 어레이의 저장 장치 내에 저장될 수 있는 스트립들의 최대수를 식별하는 것을 포함한다. 이것은 또한 가장 작은 용량을 구비한 어레이 내의 저장 장치들 상의 스트립 LBA들의 수를 식별하는 것으로 기술될 수 있다. 시퀀스(1100)는 또한 오퍼레이션(1106)을 포함하는데, 이 오퍼레이션은 어레이 내의 가장 작은 용량의 저장 장치(들) 내에 저장될 수 있는 파라미터 N을 스트립들의 최대 수와 동일하게 설정할 수 있으며, 이것은 스트립 LBA들의 수와 동일하게 N을 설정하는 것으로 기술될 수 있다. 각 스트라이드의 원본과 1 개의 카피본을 저장하는 실시예에 있어서, 그 오퍼레이션들은 또한 주요 데이터에 대하여 사용 가능한 스트립 LAB들의 절반을 저장하고, 데이터의 회전된 카피들에 대하여 사용 가능한 스트립 LAB들의 절반을 저장하는 것을 포함한다. 각 스트라이드의 원본과 2 개의 카피본을 저장하는 실시예에 있어서, 오퍼레이션들은 또한 주요 데이터에 대하여 사용 가능한 스트립 LAB들의 3분의 1을 저장하고, 데이터의 회전된 카피본들에 대하여 사용 가능한 스트립 LAB들의 3분의 2를 저장하는 것을 포함한다. 저장 공간의 유지는 예컨대, 일 대 일 맵핑 알고리즘과 도 5 내지 10에 도시된 것과 같은 테이블과 같은 저장 장치 어댑터(DA1-8)에 의해 잠재적으로 실행될 수 있다. 일반적으로, 클러스터(102, 104)(또는 호스트 124)로부터 수신된 데이터를 기록하기 위한 요청에 응답하여, 저장 장치 어댑터(DA1-8)는 데이터의 주요 카피 및 임의의 부 카피들의 기록을 실행하고, 또한 예컨대, 맵핑 테이블을 사용함으로써 무엇을 기록하고, 어디에 기록할지를 추적할 수 있다.

[0075] 시퀀스(1100)는 또한 오퍼레이션(1108)을 포함하는데, 이 오퍼레이션은 저장될 스트라이드 S_j 의 수 j를 식별하는 것을 포함한다. 시퀀스(1100)는 또한 오퍼레이션(1110)을 포함하는데, 이 오퍼레이션은 각 스트립의 원본 및 단일 카피가 저장되는 예에 있어서, 2j 이 N-1 보다 작거나 동일한지 여부를 결정한다. 만약 오퍼레이션(1110)에서 2j 가 N-1 보다 작거나 동일하다고 결정되면, 시퀀스(1100)는 하나 이상의 오퍼레이션들(1112, 1114, 1116, 및 1118)을 포함할 수 있다. 오퍼레이션(1112)은 어레이 내의 제1 저장 장치 내의 LBA 예컨대, LBAj와, 어레이 내의 제2 저장 장치 내의 LAB, 예컨대 LABj+1로 스트립 s1j를 기록하는 것을 포함한다. 예컨대, 제1 및 제2 저장 장치는 저장 장치 그룹들(126a-d) 내에 포함될 수 있다. 오퍼레이션(1114)은 스트립 s2j를 제2 저장 장치 내의 LBA 예컨대, LABj+1 와 어레이 내의 제3 저장 장치 내의 LAB 예컨대, LABj+1 로 기록하는 것을 포함한다. 오퍼레이션(1116)은 스트립 s3j를 제3 저장 장치 내의 LAB 예컨대, LBAj+1와, 어레이 내의 제4 저장 장치 내의 LAB 예컨대, LBAj+1에 기록하는 것을 포함한다. 스트립 s1j, s2j, s3j 는 오퍼레이션(1108) 내에서 식별되는 스트라이드 j의 구성요소일 수 있다. 하나 이상의 스트립(s1j, s2j, s3j)들은 패리티 스트립일 수 있다. 또한, 스트라이드(j)가 추가적인 스트립들을 가지는 경우, 스트라이드(j)의 추가적인 스트립들은 저장될 수 있다. 예를 들어, 스트립(s4j)은 예컨대 LBAj와 같은 어레이의 4번째 저장 장치의 LBA에, 및, 예컨대, LBAj+1과 같은 어레이의 5번째 저장 장치의 LBA에 기록될 수 있으며, 스트립(s5j)은 예컨대 LBAj와 같은 5번째 저장 장치의 LBA에, 및, 예컨대 LBAj+1과 같은 어레이의 6번째 저장 장치의 LBA에 기록될 수 있고,

스트립(s6j)은 예컨대 LBAj와 같은 6번째 저장 장치의 LBA에, 및 예컨대 LBAj+1과 같은 1번째 저장 장치의 LBA에 기록될 수 있다. 하나 이상의 스트립(s1j, s2j, s3j, s4j, s5j, s6j)은 패리티 스트립일 수 있다. 다른 실시예들에서, 스트라이드(j)의 3보다 크거나 작은 수의 스트립, 또는 스트라이드(j)의 6보다 크거나 작은 수의 스트립이 유사한 방법으로 저장 장치에 기록될 수 있으며, 이때 각 스트립은 2개 이상의 저장 장치에 기록된다.

[0076] 오퍼레이션(1118)은 어레이에 저장할 추가의 스트라이드가 있는지를 결정하는 단계를 포함하고, 만일 있다면, 하나 이상의 오퍼레이션(1108 - 1118)이 다시 수행될 수 있다. 오퍼레이션(1118) 내에, 저장할 추가의 스트라이드가 없는 것으로 결정되면, 시퀀스(1100)가 종결한다.

[0077] 대안적인 실시예에서는, 오퍼레이션(1110)은 3j가 N-1보다 작은지 결정하는 단계를 포함한다. 이러한 대안적인 실시예에서는, 오퍼레이션(1110)에서, 3j가 N-1보다 작은 것으로 결정되면, 그 다음 시퀀스(1100)는 오퍼레이션(1112, 1114, 1116, 1118)의 대안적인 실시예들을 포함할 수 있다. 예를 들어, 오퍼레이션(1112)은 예컨대 LBAj와 같은 어레이의 1번째 저장 장치의 LBA에, 및 예컨대 LBAj+2와 같은 어레이의 2번째 저장 장치의 LBA에, 및 예컨대 LBAj+1과 같은 어레이의 3번째 저장 장치에 스트립(s1j)을 기록하는 단계를 포함할 수 있다. 이러한 대안적 실시예에서는 오퍼레이션(1114)은 예컨대 LBAj와 같은 2번째 저장 장치의 LBA에, 예컨대 LBAj+2과 같은 3번째 저장 장치의 LBA에, 예컨대 LBAj+1과 같은 어레이의 4번째 저장 장치의 LBA에 스트립(s2j)을 기록하는 단계를 포함할 수 있다. 이 대안적인 실시예에서, 오퍼레이션(1116)은 예컨대 LBAj와 같은 3번째 저장 장치의 LBA에, 예컨대 LBAj+2과 같은 4번째 저장 장치의 LBA에, 예컨대 LBAj+1과 같은 어레이의 5번째 저장 장치의 LBA에 스트립(s3j)를 기록하는 단계를 포함할 수 있다. 이 대안적인 실시예에서 스트라이드(j)의 추가의 스트립이 유사한 방법으로 저장될 수 있다. 예를 들어, 스트립(s4j)은 예컨대 LBAj와 같은 어레이의 4번째 저장 장치의 LBA에, 및 예컨대 LBAj+2과 같은 어레이의 5번째 저장 장치의 LBA에, 및 예컨대 LBAj+1과 같은 어레이의 6번째 저장 장치의 LBA에 기록될 수 있고; 스트립(s5j)은 예컨대 LBAj와 같은 5번째 저장 장치의 LBA에, 및 예컨대 LBAj+2과 같은 어레이의 6번째 저장 장치의 LBA에, 및 예컨대 LBAj+1과 같은 어레이의 1번째 저장 장치의 LBA에 기록될 수 있고; 스트립(s6j)은 예컨대 LBAj와 같은 6번째 저장 장치의 LBA에, 및 예컨대 LBAj+2과 같은 1번째 저장 장치의 LBA에, 예컨대 LBAj+1과 같은 2번째 저장 장치의 LBA에 기록될 수 있다. 다른 실시예들에서는, 스트라이드(j)는 3보다 크거나 작은(또는 6보다 크거나 작은) 스트립의 수를 가질 수 있고, 이러한 실시예에서는 스트라이드(j)의 스트립들은 오퍼레이션(1112, 1114, 1116)에서 기술된 방법으로 저장 장치에 기록될 수 있으며, 이때 각 스트립은 3개의 저장 장치에 기록된다. 다른 대안적인 실시예에서는, 각 스트라이드의 추가의 카피가 유사한 방법으로 저장될 수 있다. 오퍼레이션(1118)은 어레이에 저장되는 추가의 스트라이드가 있는지 결정하는 단계를 포함하고, 만일 있다면, 하나 이상의 오퍼레이션(1108 - 1118)이 전술한 이 대안적인 실시예에서와 같이 다시 수행된다. 오퍼레이션(1118)에서, 저장되는 추가의 스트라이드가 없는 것으로 결정되면, 시퀀스(1100)는 종결한다.

[0078] 도 11a-b에 도시된 첫번째 실시예를 다시 참조하면, 각 스트라이드의 오리지널 카피와 하나의 추가의 카피본이 기록되면, 만일 오퍼레이션(1110)에서 2j가 N-1보다 작지 않거나 동일한 것으로 결정되면, 그 다음 시퀀스(1100)가 하나 이상의 오퍼레이션(1120, 1122, 1124, 1126)을 포함할 수 있다. 도 11b를 참조하면, 오퍼레이션(1120)은 예컨대 LBA(2j-N+1)와 같은 1번째 저장 장치의 LBA에 스트립(s1j)을 기록하는 단계를 포함한다. 오퍼레이션(1122)은 예컨대 LBA(2j-N+1)과 같은 2번째 저장 장치의 LBA에 스트립(s2j)를 기록하는 단계를 포함한다. 오퍼레이션(1124)은 예컨대 LBA(2j-N+1)과 같은 3번째 저장 장치의 LBA에 스트립(s3j)를 기록하는 단계를 포함한다. 만일 스트라이드(j)에 추가의 스트립이 있다면, 유사한 방법으로 저장될 것이다. 예를 들어, 스트립(s4j)이 예컨대 LBA(2j-N+1)과 같은 4번째 저장 장치의 LBA에 기록될 수 있고, 스트립(s5j)이 예컨대 LBA(2j-N+1)과 같은 5번째 저장 장치의 LBA에 기록될 수 있고, 스트립(s6j)이 예컨대 LBA(2j-N+1)과 같은 6번째 저장 장치의 LBA에 기록될 수 있다. 다른 실시예들에서는, 스트라이드(j)는 3보다 크거나 작은(또는 6보다 크거나 작은) 스트립의 수를 가질 수 있고, 이 실시예에서 스트라이드(j)의 스트립들은 오퍼레이션(1120, 1122, 1124)에서 기술된 방법에서와 같이 저장 장치에 기록될 수 있다. 오퍼레이션(1126)은 어레이에 저장되는 추가의 스트라이드가 있는지 결정하는 단계를 포함하고, 만일 있다면, 하나 이상의 오퍼레이션(1108 - 1126)이 다시 수행될 수 있다. 만일 저장할 추가의 스트라이드가 없다면, 시퀀스(1100)가 종결한다.

[0079] 대안적인 실시예에서는 오퍼레이션(1110)은 3j가 N-1보다 작은지 결정하는 단계를 포함하고, 만일 3j가 N-1보다 작지 않다면, 그 다음 시퀀스(1100)는 대안적인 실시예의 오퍼레이션(1120, 1122, 1124, 1126)을 포함할 수 있다. 예를 들어, 도 11b를 참조하면, 오퍼레이션(1120)은 예컨대 LBA(3j-N+2)와 같은 1번째 저장 장치의 LBA에 스트립(s1j)을 기록하는 단계를 포함할 수 있다. 이 대안적인 실시예에서는 오퍼레이션(1122)은 예컨대 LBA(3j-N+2)와 같은 2번째 저장 장치의 LBA에 스트립(s2j)을 기록하는 단계를 포함할 수 있다. 또한, 이 대안

적인 실시예에서는, 오퍼레이션(1124)은 예컨대 LBA(3j-N+2)와 같은 3번째 저장 장치의 LBA에 스트립(s3j)을 기록하는 단계를 포함할 수 있다. 만일 스트라이드(j)에 추가의 스트립이 있다면, 그들은 유사한 방법으로 저장될 것이다. 예를 들면, 스트립 s4j는 네 번째 저장 장치내의 LBA에 기록될 수 있을 것이다(예를 들어, LBA(3j-N+2)). 그리고 스트립 s5j는 다섯번째 저장장치내의 LBA에 기록될 수 있을 것이다(예를 들어, LBA(3j-N+2). 그리고 스트립 s6j는 여섯번째 저장장치내의 LBA에 기록될 수 있을 것이다(예를 들어, LBA(3j-N+2). 다른 실시예에 있어서, 스트라이드j 는 3보다 크거나 작은(또는 6보다 크거나 작은) 스트립의 숫자를 가질 수도 있을 것이다. 그리고, 이러한 실시예에 있어서 스트라이드 j의 스트립은 동작 1120,1122, 및 1123에 설명된 방법에 의해 저장장치에 기록될 수 있을 것이다. 동작 1126은 어레이에 저장할 수 있는 추가의 스트라이드가 있는지 여부를 결정하는 단계를 포함한다. 그리고 그러하다면, 1108 내지 1126 중 하나 또는 그 이상의 동작은 이러한 대안의 실시예에 설명된 바오 같이 다시 수행될 수 있을 것이다. 만약, 저장할 추가의 스트라이드가 없다면, 시퀀스 1100은 끝이 날 것이다.

[0080] 상기에서 언급된 시퀀스의 하나의 예는 다음과 같이 요약될 수 있을 것이다.;처리는 N개의 가용한 LBA들과 함께, m개의 디스크 드라이브의 어레이 상에서 수행될 수 있을 것이다. 여기서, 각 스트라이드 S는 패리티 스트립 Sj=(s1j+s2j+...+smj)을 포함하는 m개의 스트립(s1,s2,...sm)으로 구성된다. . 새로운 스트라이드 Sj는 LBAj에서 시작되며 기록된다. 여기서, j=0,1,2,3,...N-1, N=기록을 위한 가용한 LBA의 숫자(메타 데이터를 포함하여)가용한 n은 2j와 같도록 설정될 수 있다. 데이터를 바람직한 패턴으로 저장하기 위해서, 만약 n이 N-1보다 작거나 같다면, LBAn에서 시작하여, s1j를 디스크1에 s2j를 디스크2에 그리고 smj를 디스크m에 기록한다. 선행된 절차는 단지 하나의 예에 불과하다. 그리고 데이터의 기록 패턴과 데이터의 카피는 또한 일대일 대응을 갖는 다른 저장 패턴에서도 일반화될 수 있을 것이다.

[0081] C. 추가의 논의

[0082] 본원 발명의 다른 예들에서 실용화되는 두 번째 카피들은 다양한 기술을 사용하여 기록될 수 있을 것이다. 예를 들면, RAID 어레이에 첨부된 하나 또는 그 이상의 장치 어댑터 DA1-8은 실시간 모드에서 어레이 카피본를 만드는데 사용될 수 있을 것이다. 실시간으로 장치 어댑터 버퍼는 이전의 데이터 스트립과, 어레이 멤버를 타켓으로 하는 오리지널 데이터 스트립과 쌍을 이루는 데스티지(DESTAGE)를 홀드하는 데 사용될 수 있을 것이다. 듀얼 카피를 할 수 있는 공간이 없는 경우라면 새로운 오리지널 데이터 스트라이드는 카피 스트립의 가장 오래된 스트립 위에 기록될 수 있을 것이다. 카피본이 중복기록된 오래된 데이터의 오리지널 스트라이드는 손상되지 않은 상태로 남을 것이며, 베이스 RAID 코드에 의해 제공되는 RAID 보호가 여전히 보장될 것이다. 중복기록되지 않은 오리지널 스트라이드는 높은 리던던시 보호를 계속 갖게 될 것이다. 결과적으로 모든 카피 스트라이드는 중복기록될 것이며, 최소한의 베이스 RAID 보호만이 남을 것이다.

[0083] 실시간으로 데이터의 카피본을 기록하기 보다는 RAID 어레이에 부착된 하나 또는 그 이상의 장치 어댑터 DA1-8가 백그라운드 모드에서는 어레이 카피본을 만드는데 사용될 것이다. 백그라운드 모드에서는 장치 어댑터 DA1-8은 각 어레이 멤버로부터 스트립을 읽을 수 있고 그들을 오리지날 스트라이드에 대하여 쉬프트된 시퀀스에 기록할 것이다.

[0084] 본 발명의 일부 예는 RAID 스트라이드 카피의 듀얼 세트 혹은 그보다 높은 수준의 세트를 소정 개수의 디스크 전반에 걸쳐서 스트립하는 것을 포함한다. 각각의 제1 스트라이드는 m 순차 스트립으로 이루어지고, 각 스트립은 어레이의 m 드라이브 중 하나에 기록된다. 스트립 중 적어도 하나는, 예컨대 잔여 스트립의 배타적 논리합(XORing)에 의해 구성된 패리티 스트립일 수 있다. 제1 스트라이드에 있어서 스트립의 제2 카피는 어레이 내의 디스크에 대하여 회전되어 어레이 내의 디스크의 제2 준-피지컬 미러를 제공한다.

[0085] 도 12는 RAID 5가 베이스 어레이인 6개의 디스크 어레이용으로 각 스트라이드의 싱글 카피가 제조되는 본 발명의 실시예를 보여준다. 또한, 다른 패리티 RAID 방식(RAID 51, 더블 패리티 등)도 이러한 리던던시 증가에 의해 (혹은 본 발명의 다른 실시예에서는 이중 혹은 삼중 미러링 등과 같은 추가적인 리던던시에 의해) 향상될 수 있다. 제1 저장 스트라이드는 A, B, C ...로 지정되고, 1 드라이브(즉, 스트레치가 1임)로 회전된 스트라이드의 카피, 즉 제2 세트가 A', B', C'...로 지정된다. 따라서, A', B' 및 C'는 이들의 언프라임 카운터파트 A, B, C의 미러 이미지인 제2 데이터 스트라이드이다. 전술한 바와 같이, 역시 1 드라이브로 회전된(혹은 소정의 다른 수의 드라이브로 회전된) 제2(혹은 제3) 카피 등과 같은 추가적인 카피를 사용하여 보다 높은 수준의 리던던시를 제공할 수 있다. 이 예에서의 각 스트라이드는, 데이터 스트립 A1, A2, A3, A4 및 A5와 관련된 패리티 스트립을 나타내는 소정의 패리티 스트립(예컨대 Ap)을 갖는다. 따라서, i=1, 2, 3, 4, 5인 경우의 A1, B1, C1, ...이 제1 데이터 스트립이고, Ap, Bp, Cp ...는 관련 패리티 스트립이다. 이 예에서, 제1 스트립과 제2

스트립은 모두 각각 1의 스트레치를 갖는다(각각의 후속 스트라이드는 1 드라이브만큼 회전된다). 그러나, 다른 스트레치(예컨대, 2, 3, 4, 또는 5)를 사용할 수 있다.

[0086] 도 13은 손실 데이터의 재구성(패리티 복원)을 사용하지 않고도 하나의 디스크 고장 이후에 재구성을 행하는 예를 보여준다. 각 손실 스트립은 인접 드라이브로부터의 카피에 의해 스페어 드라이브에서 재구성되는데, 이러한 재구성은 제2 스트립 A'2, B'1, C'p,...로부터 제1 스트립 A2, B1, Cp,...을 재구성하는 것으로 시작된다.

[0087] 도 14는 패리티 재구성을 사용하지 않고도 2개의 비인접 디스크 고장 이후에 재구성을 행하는 예를 보여준다. 이 도면은 임의의 2개의 비인접 디스크 고장으로부터 회복하는 능력을 예시하는데, 이 능력은 베이스 RAID 5보다 높은 수준의 내성이다. 각 스트립은 인접 드라이브로부터의 카피에 의해 스페어 드라브에서 재구성된다. 예컨대, 우선 제1 스트립 A2, B1, Cp,...이 제2 스트립 A'2, B'1, C'p,...로부터 재구성되고, 이후에 제2 스페어 A4, B3, C2,...에서 제1 스트립의 재구성이 행해진다. 이 예에서, 패리티 스트립 Ap, Bp, Cp,...를 이용한 데이터 복원은 불필요한 데, 이는 손상된 드라이브가 인접해 있지 않기 때문이다.

[0088] 도 15는 2개의 인접 디스크 고장 이후에 재구성을 행하는 예를 보여준다. 이 도면은 인접한 곳에서 고장이 일어난 경우에도 임의의 2개의 고장으로부터 회복하는 능력을 예시하는데, 이 능력은 베이스 RAID 5보다 높은 수준의 내성이다. 이러한 재구성은 하나의 스페어 디스크 드라이브의 제1 스트립으로 최소화된 패리티 복원을 사용한다. 도 15의 (a)는 어레이에 있어서 2개의 고장난 드라이브를 확인시켜준다. 도 15의 (b)는 제1 스트립이 인접 드라이브로부터의 카피에 의해 복원되는 것을 보여준다. 도 15의 (c)는 제1 스트립 A2, B1, Cp,...이 패리티 복원을 이용하여 복원되는 것을 보여준다. 도 15의 (d)는 제2 스트립 A'1, B'p, C'5이 인접 하드 디스크 드라이브 상의 인접 스페어 제2 스트립 A1, Bp, C5로부터의 카피에 의해 복원되고, 제2 스트립 A'2, B'1, C'p이 인접 하드 디스크 드라이브 상의 인접 스페어 제2 스트립 A2, B1, Cp로부터의 카피에 의해 복원되는 것을 보여준다.

[0089] 본 발명의 일부 예와 관련하여 전술한 바와 같이, RAID 어레이에 있는 소정 수의 디스크에 있어서, 제1 RAID 스트라이드의 회전 카피를 사용하면, 드라이브 고장에 대한 내성(리던던시)가 베이스 RAID 보다 높아진다. 또한, 본 발명의 일부 예는, 제1 RAID 기억이 제2 카피(혹은 다른 실시예에 있어서는 제3 혹은 다른 수의 카피)와 겹칠 때, 드라이브 고장에 대한 내성을 베이스 RAID 어레이보다 나쁘지 않은 수준까지 점진적으로 감소시키는, 최적 리던던시를 위한 자체 튜닝 프로세스를 제공한다. 본 발명의 일부 예는, 소정 수의 디스크가 베이스 RAID 시스템에서보다 고객의 데이터에 대한 자체 보호를 향상시키고, 사용되는 디스크 공간의 양이 커짐에 따라 상기 자체 보호가 튜닝되며, 하나 이상의 드라이브가 고장날 때 효율적으로 자체 치유가 이루어지는, 자율적 RAID 시스템을 제공한다.

[0090] RAID 어레이의 주어진 개수의 디스크 드라이브들에 대하여, 본 발명의 일부 일례들은, 자유 디스크 공간을 이용하여 베이스 RAID 코드에 의해 제공되는 상술한 리던던시 기록을 통하여 RAID 어레이의 유효한 고장 방지능력을 증가시킨다. 어레이 드라이브들의 세트 넘버에 대하여, 각각의 RAID 카피본은, 베이스 RAID 보다 더 많은 드라이브 고장 방지능력을 제공한다. 예를 들어, RAID 5 베이스 코드를 가진 6 개의 멤버 어레이에 대하여, 본 발명이 사용되지 않는 경우에, 데이터는, 하나 보다 더 많은 드라이브 멤버가 고장날 때만 복원될 수도 있다. 이와 반대로, 본 발명의 몇몇 일례들에 대하여, 회전된 RAID 스트립들의 단일 카피본이 존재하며, 데이터는 2 개의 드라이브 멤버들이 고장나더라도 복원될 수도 있다. 본 발명의 일례에 대하여, 회전된 RAID 스트립들의 2 개의 카피본이 저장되며, 데이터는 3 개의 디스크 고장이 동시에 발생하는 경우에도 복원될 수도 있다.

[0091] 본 발명의 일부 실시예들은 RAID 어레이의 초기 사용 동안에 더 높은 RAID 보호를 제공하며, 이는 보호가 가장 필요한 경우이고, 대부분의 자유 공간이 이용가능하다. 새로운 디스크 어레이의 초기 사용은, 새로운 하드 디스크 드라이브(HDD)의 초기 사망율이, 많은 POH(Power On Hours)에 대하여 동작한 이후에 HDD 고장율 보다 더 높게 된다.

[0092] 본 발명의 일부 실시예에 따르면, 어레이 중의 소정 수의 디스크를 위한 베이스 RAID 어레이의 유효 데이터 용량의 100%를 사용할 수 있도록 한다. 이는 이전의 (고객) 데이터를 베이스 RAID 코드의 고장 방지능력(fault tolerance)에 대하여 점진적으로 노출시키는 희생을 통하여 달성된다. 어레이 디스크의 고장 방지능력은 추가의 디스크 공간 사용에 따라 자발적으로 감소하지만, 주요 RAID의 고장 방지능력 아래로는 떨어지지 않는다. 이에 따라 데이터의 보호는 항상 적어도 베이스 RAID 코드의 것에 대응한다.

[0093] 본 발명의 일부 예에 따르면, 가장 이전의 데이터로 시작하는 데이터의 회전된 카피는 궁극적으로 새로운 (고객) 데이터 위에 기록되므로, 단지 주요 데이터만이 남게 된다. 단지 주요 데이터만이 남게 되는 데이터의

서브세트의 경우에, 예컨대 RAID 5는 데이터 복구를 위하여 단지 하나의 디스크 고장만을 허용할 것이다. 회전된 카피본이 새로운 데이터에 의하여 위에 기록되지 않는 어레이 중의 데이터는 여전히 보다 높은 디스크 고장 방지능력을 가질 것이다. 단지 주요 데이터만이 남게 되는 데이터의 서브세트는, 어레이의 모든 데이터 용량이 사용될 때까지 추가의 고객 데이터가 어레이에 저장되는 때에 성장한다. 일부 예에 따르면, 조작 시스템이 기존의 보조 카피본의 위치에 대해 기록하기를 원하고, 디스크가 채워지지 않은 경우에, 저장 장치 어댑터 AD1-8는 기존의 보조 카피본을 다른 위치로 이동시킬 수도 있고, 이전에 기록된 보조 카피본을 판독 및 재저장하지 않고 저장 위치를 재할당할 수도 있다.

[0094] RAID 5와 관련하여 "싱글 미러(single mirror; 각 스트립의 1개의 회전된 카피)"를 위한 보호가 도 16에 도시되어 있다. 보다 구체적으로, 도 16은 싱글 미러를 활용하고 있는 본 발명의 예에 따른 "자율적 RAID 5"를 위하여 임의의 2개의 하드 디스크 드라이브 고장에 대하여 보호된 데이터의 비율을 도시하고 있다. 일반적으로, 본 발명의 일부 예는 "자율적 고객 데이터 보호를 위한 RAID 저장 장치"로 지칭될 수 있다. RAID 5와 함께 사용될 때에, 본 발명의 일부 예는 "자율적 RAID 5"로 지칭될 수 있다. 도 16 에 도시된 바와 같이, 가용 디스크 공간의 대략 50%를 사용할 때까지 모든 데이터에 대하여 2개의 디스크 고장 방지능력이 존재한다. 이와 달리, 베이스 RAID 5의 고장 보호(fault protection)는 도 17의 하부에서 제로 퍼센트 수평선으로 표시되어 있다. 따라서 이러한 싱글 미러의 예는 최대 2개의 디스크 고장 방지능력(fault tolerance)를 제공하며, 이는 베이스 RAID 5의 싱글 디스크 고장 허용요차에 비하여 상당히 개선된 것이다.

[0095] "더블 미러(double mirror; 각 스트립의 2개의 회전된 카피)"를 위한 보호가 도 17에 도시되어 있다. 보다 구체적으로, 도 17은 더블 미러를 활용하고 있는 본 발명의 예에 따른 "자발적 RAID 5"를 위하여 임의의 3개의 하드 디스크 드라이브 고장에 대하여 보호된 데이터의 비율을 도시하고 있다. 도 17에 도시된 바와 같이, 가용 디스크 공간의 대략 33.33%를 사용할 때까지 모든 데이터에 대하여 3개의 디스크 고장 방지능력이 존재한다. 이와 달리, 베이스 RAID 5의 고장 보호(fault protection)는 도 17의 하부에서 제로 퍼센트 수평선으로 표시되어 있다. 따라서 이러한 더블 미러의 예는 최대 3개의 디스크 고장 방지능력(fault tolerance)를 제공하며, 이는 베이스 RAID 5의 싱글 디스크 고장 방지능력(fault tolerance)에 비하여 상당히 개선된 것이다.

[0096] 본 발명의 일부 실시예는 하나 이상의 드라이브가 고장난 경우의 복원 시간을 현저하게 감소시킴으로써 복원 중의 (고객) 데이터 손실에 대한 추가의 RAID 견고성(robustness)을 제공한다. 예로서, 어레이 손실 또는 〔킬스트립(killstrip)으로 지칭될 수 있는〕 하나 이상의 스트립의 손실에 기인하여 데이터 손실이 발생할 수 있다. 본 발명의 예에 의해 제공되는 보조 카피본으로 인하여, 패리티 복구를 통하여 손실된 주요 데이터를 복구하는 단계를 제거할 수 있거나, 고장의 수에 따라 그리고 인접 드라이브에서 고장이 발생하는 가의 여부에 따라 주요 데이터를 복원하기 위하여 패리티 복구(parity recover)를 사용할 필요가 있는 횟수를 현저하게 감소시킬 수 있다. 본 발명의 일부 실시예에 있어서는 복원 시간을 단축할 수 있는데, 그 이유는 스트립을 서바이빙 디스크(surviving disk)로부터 핫 스페어(hot spare)에 카피할 필요가 있는 횟수가 스트라이드 중의 각각의 서바이빙 디스크 상의 스트립을 판독하고 소실 스트립을 복구하도록 XOR 연산하는 패리티 재구성을 통하여 각각의 로스트 스트립을 재구성할 필요가 있는 횟수보다 훨씬 작기 때문이다.

[0097] 또한 본 발명의 일부 실시예는, 드라이브 중 하나가 판독 요청에 응답하여 느려질 경우에 선점형 재구성(Preemptive reconstruct)보다 빠르게 데이터를 판독한다. 스트라이드 중의 남아 있는 데이터 스트립 모두를 판독하고, 응답에 느리게 반응하는 스트립(slow-to-respond-strip)의 데이터를 재구성하도록 패리티 스트립에 의해 상기 데이터 스트립들을 XOR 연산하는 대신에, 손실된 스트립 중의 데이터의 카피본을 주요 스트립과 함께 인접 드라이브로부터 판독할 수 있기 때문에, 데이터를 보다 빠르게 판독할 수 있다.

[0098] 이상에서 본 발명의 예시적인 여러 실시예를 개시하였지만, 당업자는 첨부의 청구범위에 의해 한정되는 본 발명의 범위를 벗어나지 않으면서 상기 실시예에 대한 다양한 변형 및 수정이 있을 수 있다는 것을 알 것이다. 또한, 본 발명의 구성 요소들에 대해서는 단수로서 상세한 설명에 설명하거나 청구범위에 한정하고 있지만, 단수와 관련하여 명백하게 제한하고 있지 않는 한은 상기 구성 요소들은 복수로 존재할 수도 있다.

도면의 간단한 설명

[0018] 도 1 은 본 발명의 일례에 따른 저장 시스템의 하드웨어 구성요소들 및 상호접속에 대한 블록도이다.

[0019] 도 2 는 본 발명의 일례에 따른 컴퓨팅 장치의 하드웨어 구성요소들 및 상호접속에 대한 블록도이다.

[0020] 도 3 은 본 발명의 일례에 따른 신호 운반 매체의 일례이다.

[0021] 도 4 는 본 발명의 일례에 따른 데이터 백업을 위한 동작 시퀀스의 흐름도이다.

[0022] 도 5 는 본 발명의 일례에 따른 스트라이드의 스트립들의 회전된 카피본을 제공하는 매핑 알고리즘이다.

[0023] 도 6 은 본 발명의 일례에 따른 스트라이드의 스트립들의 회전된 카피본을 제공하는 매핑 테이블이다.

[0024] 도 7 은 본 발명의 일례에 따른 스트라이드의 스트립들의 2 개의 회전된 카피본을 제공하는 매핑 알고리즘이다.

[0025] 도 8 은 본 발명의 일례에 따른 스트라이드의 스트립들의 2 개의 회전된 카피본을 제공하는 매핑 테이블이다.

[0026] 도 9 는 본 발명의 일례에 따른 저장된 LBA 밴드 매핑의 일례이다.

[0027] 도 10 은 본 발명의 일례에 따른, 스트라이드의 스트립들의 회전된 카피본을 제공하기 위하여, 보유된 밴드 및 FIFO 알고리즘을 이용한 매핑 테이블이다.

[0028] 도 11a 내지 도 11b 는 본 발명의 일례에 따라 데이터를 백업하는 동작 시퀀스의 흐름도이다.

[0029] 도 12 는 본 발명의 일례에 따른 디스크 어레이에 데이터 및 그 데이터의 카피본을 저장방식을 나타낸다.

[0030] 도 13 은 본 발명의 일례에 따른 디스크 어레이 내의 데이터를 복원하는 방식을 나타낸다.

[0031] 도 14 는 본 발명의 일례에 따른 디스크 어레이 내의 데이터를 복원하는 방식을 나타낸다.

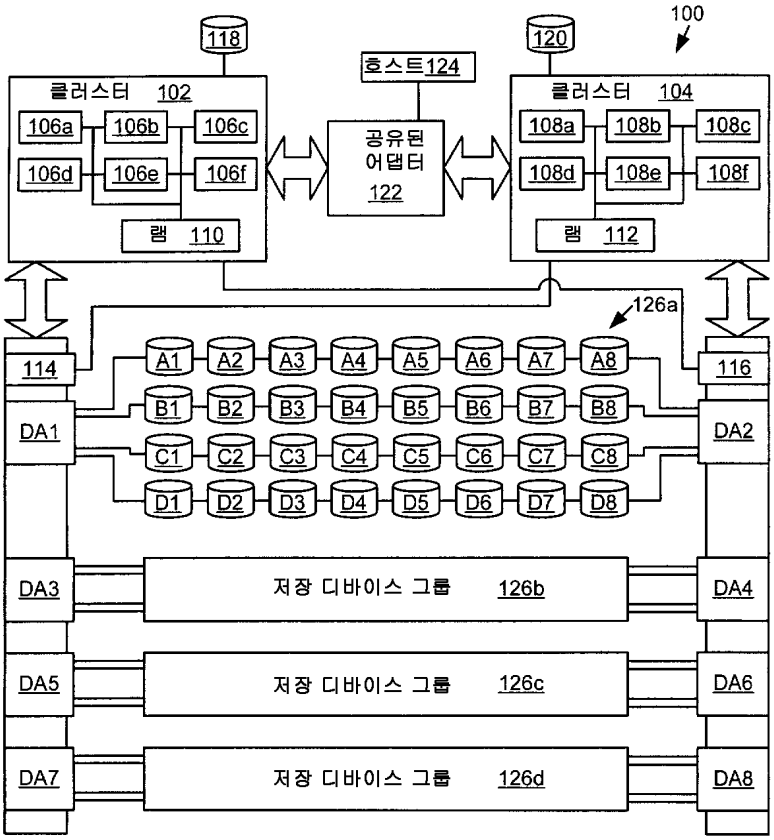
[0032] 도 15 는 본 발명의 일례에 따른 디스크 어레이 내의 데이터를 복원하는 방식을 나타낸다.

[0033] 도 16 은 하나의 회전된 카피본을 이용하여 본 발명의 일례에 따라 임의의 2 개의 하드 디스크 드라이브에 대하여 보호된 데이터의 퍼센티지를 나타내는 그래프이다.

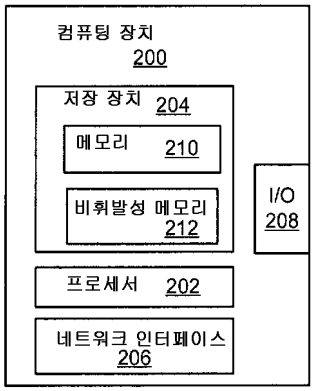
[0034] 도 17 은 2 개의 회전된 카피본을 이용하여 본 발명의 일례에 따라 임의의 3 개의 하드 디스크 드라이브에 대하여 보호된 데이터의 퍼센티지를 나타내는 그래프이다.

도면

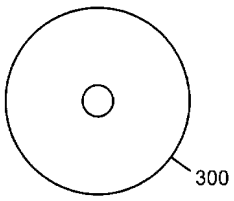
도면1



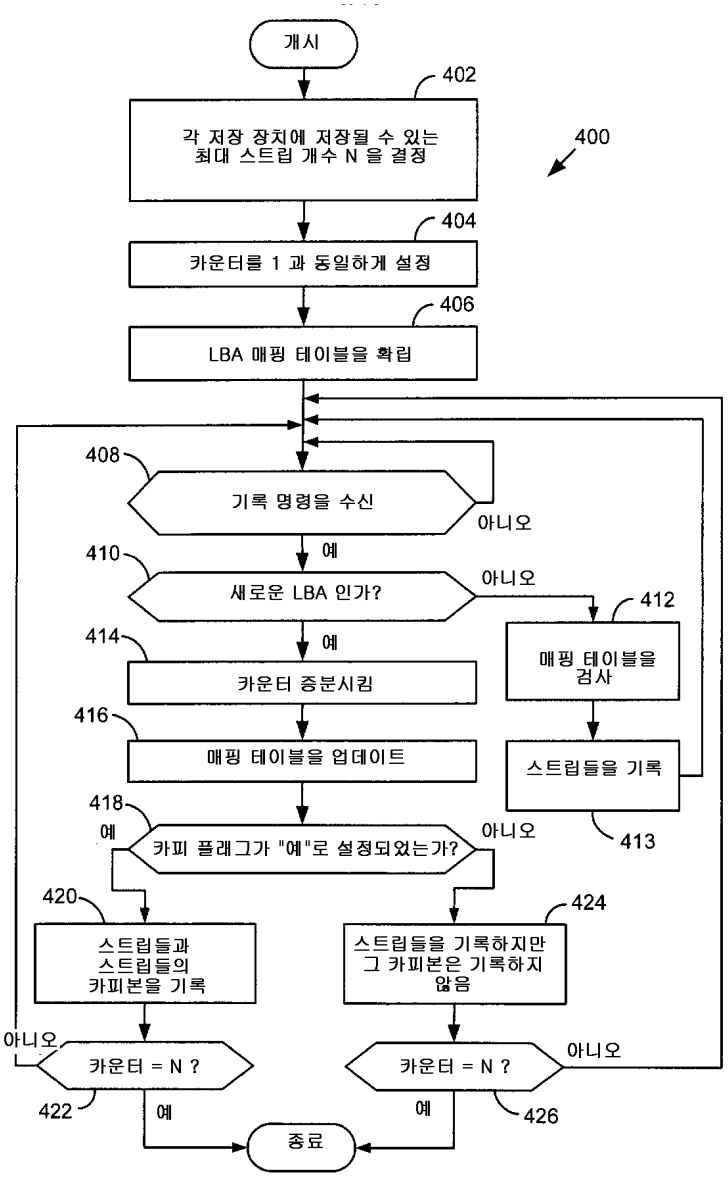
도면2


컴퓨팅 장치
200
저장 장치 204
메모리 210
비휘발성 메모리 212
I/O 208
프로세서 202
네트워크 인터페이스 206

도면3


300

도면4



도면5

	LBA	디스크 1	디스크 2	디스크 3	디스크 4	디스크 5
스트립	LBAm	s1j	s2j	s3j	s4j	s5j
회전된 카피본	LBAm+1	s5j	s1j	s2j	s3j	s4j

도면6

1 개의 회전된 카피본에 대한 LBA 매핑 테이블
FIFO 알고리즘

카운터	호스트로부터의 랜덤한 새로운 입력 스트림 LBA	FIFO 알고리즘을 이용하여 매핑된 스트림 LBA	회전된 카피본의 매핑된 LBA	카피 플래그 (Y/N)
1	LBA 21	LBA 1	LBA 2	YES
2	LBA N-7	LBA 3	LBA 4	YES
3	LBA N-1	LBA 5	LBA 6	YES
4	LBA 4	LBA 7	LBA 8	YES
5	LBA N-88	LBA 9	LBA 10	YES
6	LBA N	LBA 11	LBA 12	YES
	LBA 1	LBA 13	LBA 14	YES
.	.			
.	.			
N/2	LBA 15	LBA (N-1)	LBA N	YES
N/2+1	LBA 42	LBA 2	NA(중복기록된 카피본)	NO
	LBA 50	LBA 4	NA	NO
		LBA 6	NA	NO
	LBA N-25	LBA 8	NA	NO
	LBA 22	LBA 10	NA	NO
.				
.				
N-3	LBA N-73	LBA N-6	NA	NO
N-2	LBA 30	LBA N-4	NA	NO
N-1	LBA 6	LBA N-2	NA	NO
N	LBA 100	LBA N	NA	NO

도면7

5 개의 디스크 어레이에 대한 스트라이드 Sj 의 + 2 개의 회전된 카피본을 기록

	LBA	디스크 1	디스크 2	디스크 3	디스크 4	디스크 5
스트립	LBAm	s1j	s2j	s3j	s4j	s5j
회전된 카피본 #1	LBAm+1	s5j	s1j	s2j	s3j	s4j
회전된 카피본 #2	LBAm+2	s4j	s5j	s1j	s2j	s3j

도면8

카운터	호스트로부터의 랜덤한 새로운 스트림 LBA	FIFO 알고리즘을 이용하여 매핑된 스트림 LBA	회전된 카피본 #1 의 매핑된 LBA	카피본 #1 플래그	회전된 카피본 #2d 의 매핑된 LBA	카피본 #2 플래그
1	LBA 41	LBA 1	LBA 2	YES	LBA 3	YES
2	LBA N-76	LBA 4	LBA 5	YES	LBA 6	YES
3	LBA N-1	LBA 7	LBA 8	YES	LBA 9	YES
4	LBA 4	LBA 10	LBA 11	YES	LBA 12	YES
5	LBA N-4	LBA 13	LBA 14	YES	LBA 15	YES
...	...	...	...	...	...	...
...	...	...	...	...	...	...
N/3 (정수)	LBA 101	LBA 3	NA	NO	NA	NO
N/3+1	LBA 7	LBA 6	NA	NO	NA	NO
N/3+2	LBA 311	LBA 9	NA	NO	NA	NO
N/3+3	LBA 867	LBA 12	NA	NO	NA	NO
N/3+4	LBA 2	LBA 15	NA	NO	NA	NO
...	...	...	...	...	...	...
...	...	...	...	...	...	...
2N/3 (정수)	LBA 12	LBA 2	NA	NO	NA	NO
2N/3+1	LBA N	LBA 5	NA	NO	NA	NO
2N/3+2	LBA 1	LBA 8	NA	NO	NA	NO
2N3+3	LBA N-43	LBA 11	NA	NO	NA	NO
2N3+4	LBA 366	LBA 14	NA	NO	NA	NO
...	...	...	...	...	...	...
...	...	...	...	...	...	...
N-2	LBA 479	LBA 2N/3-5	NA	NO	NA	NO
N-1	LBA 174	LBA 2N/3-2	NA	NO	NA	NO
N	LBA 9	LBA 2N/3+1	NA	NO	NA	NO

도면9

예비 LBA 밴드 매핑
(예비 LBA 1-m, 테이블은 m=10 인 경우를 나타냄)

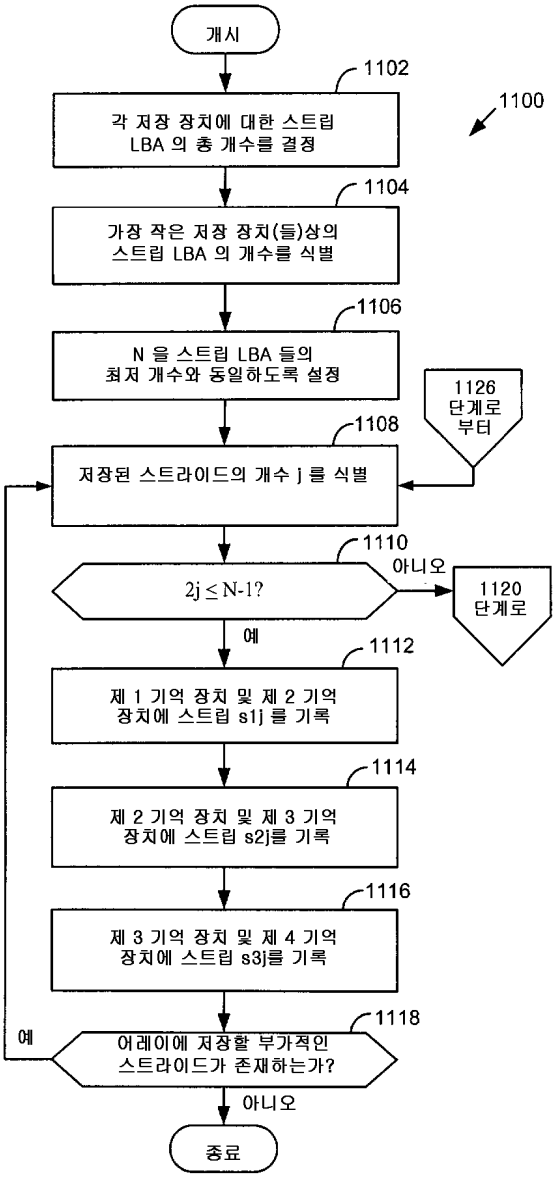
입력 스트림 LBA 의 예비 밴드	매핑된 스트림 LBA	카운터가 N/2 이하인 경우에 회전된 카피본 LBA
LBA 1	LBA 1	LBA 2
LBA 2	LBA 3	LBA 4
LBA 3	LBA 5	LBA 6
LBA 4	LBA 7	LBA 8
LBA 5	LBA 9	LBA 10
LBA 6	LBA 11	LBA 12
LBA 7	LBA 13	LBA 14
LBA 8	LBA 15	LBA 16
LBA 9	LBA 17	LBA 18
LBA 10	LBA 19	LBA 20
...	...	...
...	...	...
...	...	...
LBA k	LBA 2k-1	LBA 2k

도면10

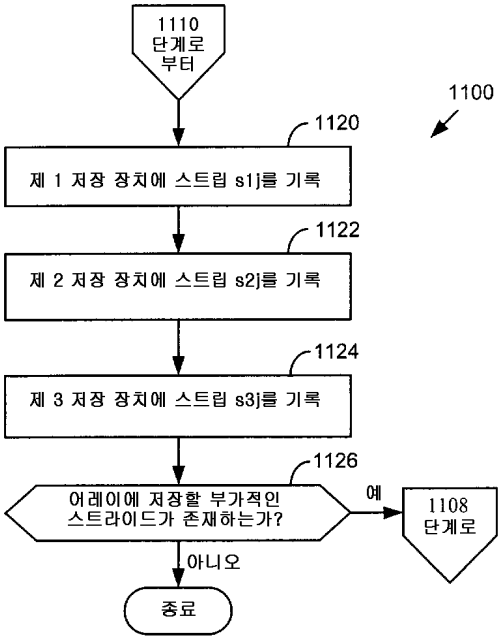
1 개의 회전된 카피본에 대한 LBA 매핑 테이블
예비 밴드 + FIFO 알고리즘
(예비 LBA 1, 2,..., m, 테이블은 m=10 인 경우를 나타냄)

카운터	호스트로 부터의 랜덤한 새로운 입력 스트림 LBA	예비 LBA 밴드 + 예비되지 않은 LBA 에 대한 FIFO 알고리즘 내에 매핑된 스트림 LBA	회전된 카피본의 매핑된 LBA	카피 플래그 (Y/N)
1	LBA 33	LBA 21	LBA 22	YES
2	LBA N-7	LBA 23	LBA 24	YES
3	LBA N-1	LBA 25	LBA 26	YES
4	**LBA 3**	**LBA 5**	**LBA 6**	**YES**
5	LBA N-88	LBA 27	LBA 28	YES
6	**LBA 9**	**LBA 17**	**LBA 18**	**YES**
	LBA 1	**LBA 1**	**LBA 2**	**YES**
.	.			
.	.			
N/2	LBA YY	LBA XX	LBA XX+1	YES
N/2+1	LBA 42	LBA 22	NA (중복기록된 카피본)	NO
	LBA 50	LBA 24	NA	NO
	LBA 2	**LBA 3**	**NA**	**NO**
	LBA N-25	LBA 28	NA	NO
	LBA 22	LBA 30	NA	NO
.				
.				
N-3	LBA N-73	LBA N-6	NA	NO
N-2	LBA 30	LBA N-4	NA	NO
N-1	**LBA 7**	**LBA 13**	**NA**	**NO**
N	LBA 100	LBA N	NA	NO

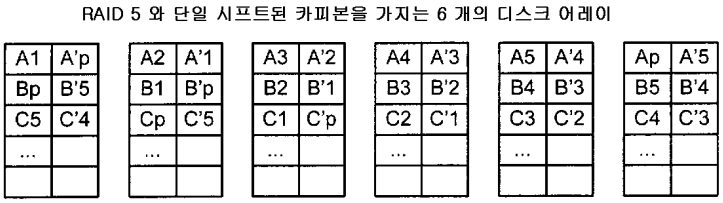
도면11a


개시
1102
각 저장 장치에 대한 스트립 LBA 의 총 개수를 결정
1100
1104
가장 작은 저장 장치(들)상의 스트립 LBA 의 개수를 식별
1106
N 을 스트립 LBA 들의 최저 개수와 동일하도록 설정
1126 단계로 부터
1108
저장된 스트라이드의 개수 j 를 식별
1110
2j ≤ N-1?
아니오
1120 단계로
예
1112
제 1 기억 장치 및 제 2 기억 장치에 스트립 s1j 를 기록
1114
제 2 기억 장치 및 제 3 기억 장치에 스트립 s2j를 기록
1116
제 3 기억 장치 및 제 4 기억 장치에 스트립 s3j를 기록
1118
어레이에 저장할 부가적인 스트라이드가 존재하는가?
예
아니오
종료

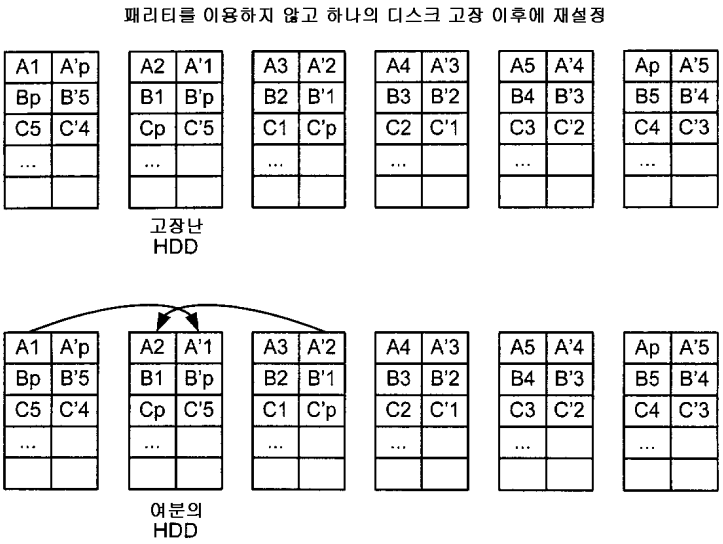
도면11b



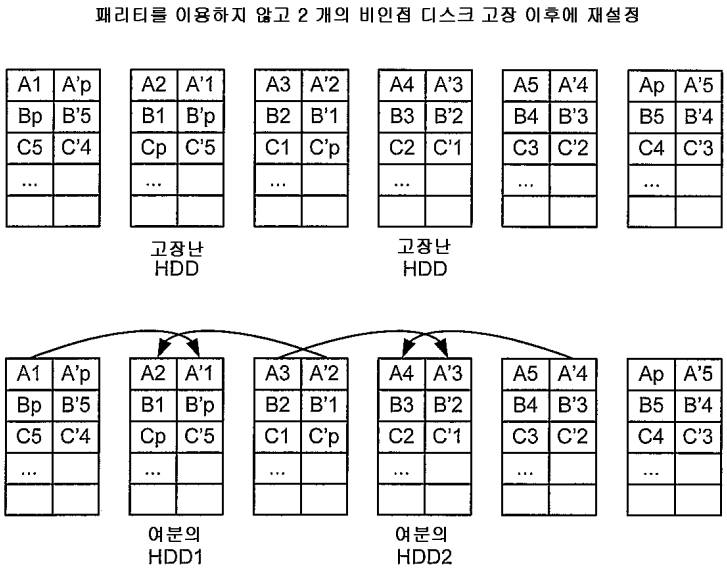
도면12



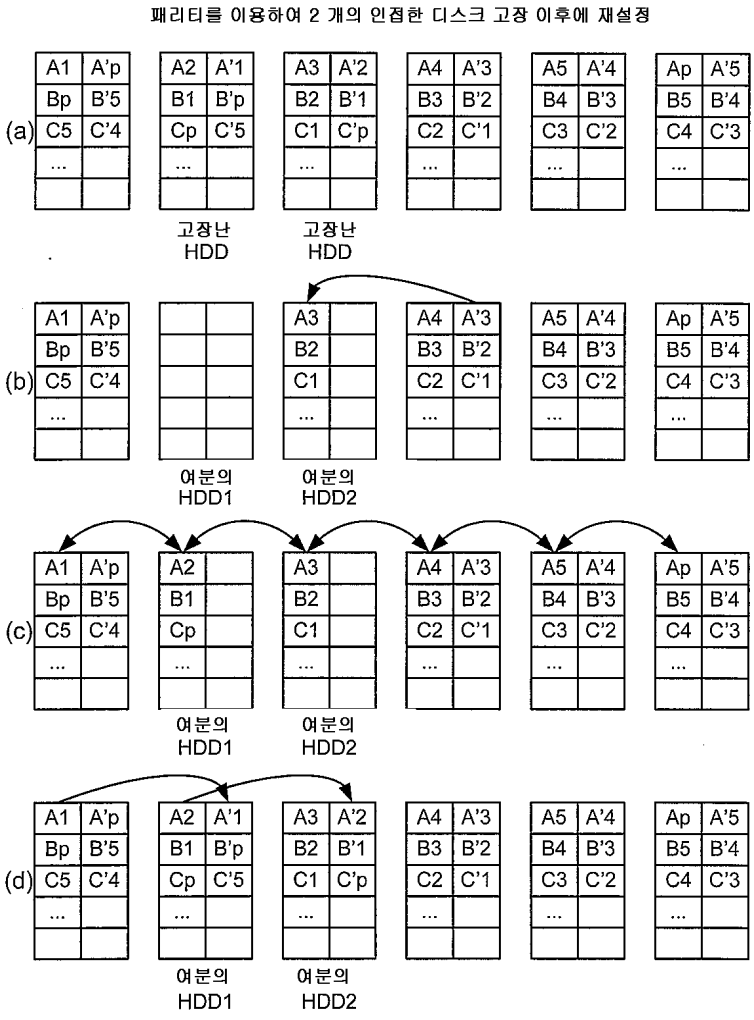
도면13



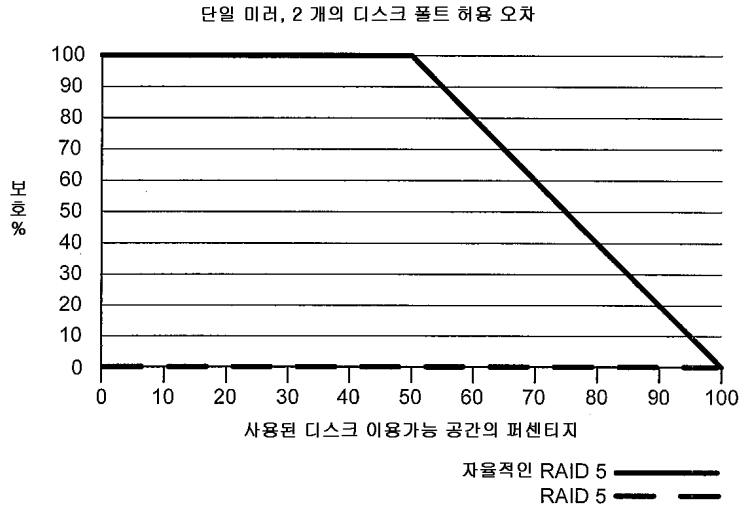
도면14


패리티를 이용하지 않고 2 개의 비인접 디스크 고장 이후에 재설정
고장난 HDD
고장난 HDD
여분의 HDD1
여분의 HDD2

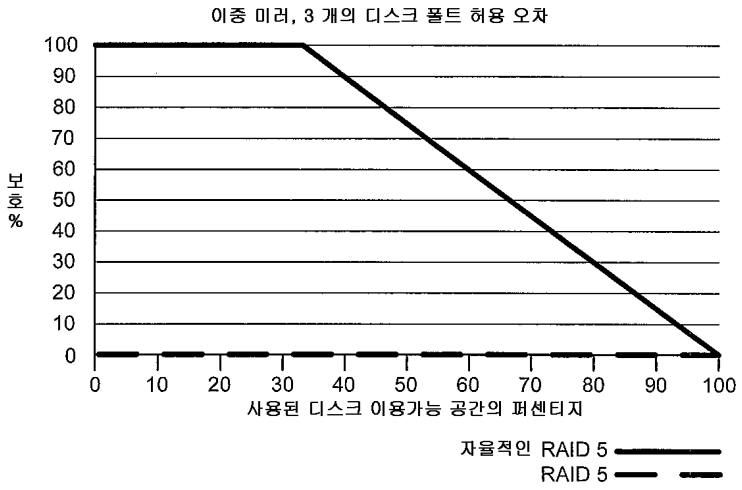
도면15


패리티를 이용하여 2 개의 인접한 디스크 고장 이후에 재설정
(a)
(b)
(c)
(d)
고장난 HDD
고장난 HDD
여분의 HDD1
여분의 HDD2

도면16


단일 미러, 2 개의 디스크 폴트 허용 오차
보호 %
100
90
80
70
60
50
40
30
20
10
0
0 10 20 30 40 50 60 70 80 90 100
사용된 디스크 이용가능 공간의 퍼센티지
자율적인 RAID 5
RAID 5

도면17


이중 미러, 3 개의 디스크 폴트 허용 오차
보호 %
100
90
80
70
60
50
40
30
20
10
0
0 10 20 30 40 50 60 70 80 90 100
사용된 디스크 이용가능 공간의 퍼센티지
자율적인 RAID 5
RAID 5