(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2014년01월06일
(11) 등록번호 10-1345572
(24) 등록일자 2013년12월20일

(51) 국제특허분류(Int. Cl.)
G06F 21/78 (2013.01) *G06F 11/08* (2006.01)
(21) 출원번호 10-2010-0053772
(22) 출원일자 2010년06월08일
심사청구일자 2012년03월07일
(65) 공개번호 10-2010-0131949
(43) 공개일자 2010년12월16일
(30) 우선권주장
61/268,055 2009년06월08일 미국(US)
(56) 선행기술조사문헌
US20060004957 A1
JP2009076075 A
WO2008070173 A1
KR1020070029358 A
전체 청구항 수 : 총 16 항

(73) 특허권자
엘에스아이 코포레이션
미국 캘리포니아 95131, 새너제이, 라이더 파크 드라이브 1320
(72) 발명자
베르트, 루카
미국 조지아 30041 커밍 마운트클레어 드라이브 820
(74) 대리인
장훈

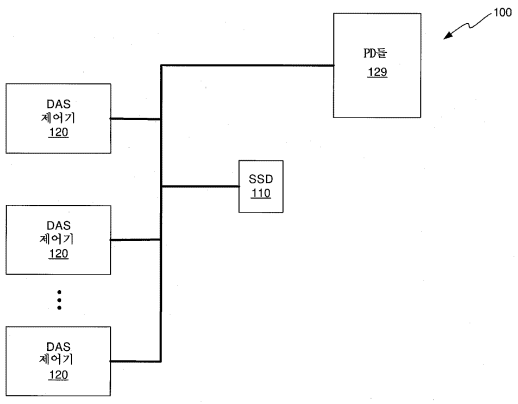
심사관 : 이석형

(54) 발명의 명칭 **직접 연결 저장 장치 (DAS) 시스템에서 캐시 데이터의 무결성을 보호하기 위한 방법 및 장치**

(57) 요 약

DAS 시스템의 DAS 제어기들에 외부에 고체 상태 디스크들(SSDs)의 어레이가 WB 캐싱 동작들을 실행하기 위해 WB 캐시 메모리로서 DAS 제어기들에 의해 이용되는, RAID 기술을 구현하는 DAS 시스템이 제공된다. 외부 SSD 어레이를 WB 캐시 메모리로서 이용함으로써 DAS 시스템은 DAS 시스템의 복잡성을 현저하게 증가시키지 않고 캐시 동작들을 실행하기 위해 이용되는 대역폭량을 증가시키지 않고 완전히 캐시 코히런트하게 된다. 또한, WB 캐시 메모리로서 외부 SSD 어레이를 이용하는 것은 DAS 제어기들을 미러할 필요성을 제거한다.

대 표 도



특허청구의 범위

청구항 1

직접 연결 저장 장치(direct-attached storage; DAS) 시스템에 있어서,

물리적 디스크 드라이브들(PDs)의 저가(또는 독립) 디스크들의 중복 어레이(Redundant Array of Inexpensive(or Independent) Disks; RAID) 어레이로서 구성된 복수의 자기 하드 디스크 드라이브들(HDDs);

캐시 메모리로서 구성된 고체 상태 디스크들(SSDs)의 어레이; 및

상기 PD들의 RAID 어레이 및 상기 SSD 어레이에 연결된 적어도 제 1 및 제 2 DAS 제어기들로서, 각각의 DAS 제어기는 중앙 처리 장치(CPU), 국부적 메모리 디바이스, 및 입력/출력(I/O) 인터페이스 디바이스를 갖고, 상기 CPU들 각각 및 상기 국부적 메모리 디바이스들 각각은 상기 PD들의 RAID 어레이의 RAID 구성에 일관된 기술의 RAID 레벨을 실행하도록 구성되고, 상기 CPU들 각각은, 상기 각각의 DAS 제어기에 수신된 데이터가 상기 SSD 어레이의 캐시 메모리에 일시적으로 저장되고 후속하여 상기 PD들의 RAID 어레이의 상기 PD들 중 하나 이상에 저장되게 하는 캐싱 알고리즘(caching algorithm)을 실행하도록 구성되고, 상기 데이터는 상기 데이터와 연관된 메타데이터(metadata)를 갖고, 상기 각각의 CPU들에 의해 실행되는 상기 캐싱 알고리즘은 상기 데이터가 상기 SSD 어레이 내의 블록들에 저장되게 하고, 각각의 블록은 데이터 무결성 필드(data integrity field; DIF)를 포함하고, 상기 각각의 CPU들은 상기 각각의 DAS 제어기에 수신된 상기 데이터에 연관된 상기 메타데이터가 상기 SSD 어레이 내의, 상기 연관된 데이터가 저장되는 블록들 중 적어도 하나의 블록에 연관된 상기 DIF들 중 적어도 하나의 DIF에 저장되게 하는, 상기 적어도 제 1 및 제 2 DAS 제어기들을 포함하는, 직접 연결 저장 장치(DAS) 시스템.

청구항 2

삭제

청구항 3

삭제

청구항 4

삭제

청구항 5

제 1 항에 있어서,

상기 각각의 CPU들은, 상기 SSD 어레이에 일시적으로 저장되는 상기 데이터에 캐시 일관성(cache coherency)이 제공되게 하기 위해, 상기 SSD 어레이 내의 상기 데이터의 일시적 저장에 관련하여 RAID 기술의 한 레벨을 실행하도록 구성되는, 직접 연결 저장 장치 (DAS) 시스템.

청구항 6

삭제

청구항 7

제 1 항에 있어서,

각각의 CPU에 의해 실행되는 상기 캐싱 알고리즘은 상기 메타데이터의 해시(hash)를 계산하는 해싱 알고리즘을 포함하고, 각각의 해시는 상기 SSD 어레이 내의, 상기 연관된 데이터가 저장되는 블록들 중 적어도 하나의 블록에 연관된 상기 DIF들 중 상기 적어도 하나의 DIF에 저장되는, 직접 연결 저장 장치 (DAS) 시스템.

청구항 8

제 1 항에 있어서,

상기 데이터에 연관된 상기 메타데이터는 상기 데이터의 소스의 표시, 상기 데이터 스트림의 길이의 표시, 및 상기 데이터가 수정되었는지의 여부의 표시를 포함하는, 직접 연결 저장 장치 (DAS) 시스템.

청구항 9

직접 연결 저장 장치 (DAS) 시스템에서 캐시된 데이터의 무결성을 보호하기 위한 방법에 있어서,

중앙 처리 장치(CPU), 국부적 메모리 디바이스, 및 입력/출력(I/O) 인터페이스 디바이스를 포함하는 제 1 DAS 제어기에서, 데이터를 수신하는 단계;

상기 제 1 DAS 제어기의 CPU에서, 상기 제 1 DAS 제어기에 수신된 상기 데이터가 고체 상태 디스크(SSD) 어레이의 하나 이상의 SSD들 내의 블록들에 일시적으로 저장되고 후속하여 물리적 디스크 드라이브들(PDs)의 저가(또는 독립) 디스크들의 중복 어레이(Redundant Array of Inexpensive(or Independent) Disks; RAID) 어레이로서 구성된 하나 이상의 자기 하드 디스크 드라이브들(HDDs)에 저장되게 하는 캐싱 알고리즘을 실행하는 단계로서, 상기 데이터는 상기 데이터와 연관된 메타데이터를 갖고, 상기 SSD 어레이의 하나 이상의 SSD들 내의, 상기 CPU들에 의해 일시적으로 데이터가 저장된 각각의 블록은 데이터 무결성 필드(DIF)를 포함하고, 상기 CPU는, 상기 DAS 제어기에 수신된 상기 데이터에 연관된 상기 메타데이터가 상기 SSD 어레이 내의, 상기 연관된 데이터가 저장되는 블록들 중 적어도 하나의 블록에 연관된 상기 DIF들 중 적어도 하나의 DIF에 저장되게 하는, 상기 제 1 DAS 제어기의 CPU에서 캐싱 알고리즘을 실행하는 단계;

CPU, 국부적 메모리 디바이스, 및 I/O 인터페이스 디바이스를 포함하는 제 2 DAS 제어기에서, 데이터를 수신하는 단계; 및

상기 제 2 DAS 제어기의 CPU에서, 상기 제 2 DAS 제어기에 수신된 상기 데이터가 상기 SSD 어레이의 하나 이상의 SSD들 내의 블록들에 일시적으로 저장되고 후속하여 상기 PD들의 RAID 어레이로서 구성된 하나 이상의 자기 HDD들에 저장되게 하는 캐싱 알고리즘을 실행하는 단계로서, 상기 제 2 DAS 제어기에 수신된 상기 데이터는 상기 데이터에 연관된 메타데이터를 갖고, 상기 SSD 어레이의 하나 이상의 SSD들 내의, 상기 제 2 DAS 제어기의 상기 CPU들에 의해 일시적으로 데이터가 저장된 각각의 블록은 무결성 필드(DIF)를 포함하고, 상기 제 2 DAS 제어기의 상기 CPU는, 상기 제 2 DAS 제어기에 수신된 상기 데이터에 연관된 상기 메타데이터가 상기 SSD 어레이 내의, 상기 연관된 데이터가 저장되는 블록들 중 적어도 하나의 블록에 연관된 상기 DIF들 중 적어도 하나의 DIF에 저장되게 하는, 상기 제 2 DAS 제어기의 CPU에서 캐싱 알고리즘을 실행하는 단계를 포함하는, 직접 연결 저장 장치 (DAS) 시스템에서 캐시된 데이터의 무결성을 보호하기 위한 방법.

청구항 10

삭제

청구항 11

삭제

청구항 12

삭제

청구항 13

제 9 항에 있어서,

상기 각각의 CPU들은, 상기 SSD 어레이에 일시적으로 저장되는 데이터에 캐시 일관성이 제공되게 하기 위해, 상기 SSD 어레이 내의 데이터의 일시적 저장에 관련하여 RAID 기술의 한 레벨을 실행하도록 구성되는, 직접 연결 저장 장치 (DAS) 시스템에서 캐시된 데이터의 무결성을 보호하기 위한 방법.

청구항 14

삭제

청구항 15

제 9 항에 있어서,

각각의 CPU에 의해 실행되는 상기 캐싱 알고리즘은 상기 메타데이터의 해시를 계산하는 해싱 알고리즘을 포함하고, 각각의 해시는 상기 SSD 어레이 내의, 상기 연관된 데이터가 저장되는 블록들 중 적어도 하나의 블록에 연관된 상기 DIF들 중 상기 적어도 하나의 DIF에 저장되는, 직접 연결 저장 장치 (DAS) 시스템에서 캐시된 데이터의 무결성을 보호하기 위한 방법.

청구항 16

제 9 항에 있어서,

상기 데이터에 연관된 상기 메타데이터는 상기 데이터의 소스의 표시, 상기 데이터 스트림의 길이의 표시, 및 상기 데이터가 수정되었는지의 여부의 표시를 포함하는, 직접 연결 저장 장치 (DAS) 시스템에서 캐시된 데이터의 무결성을 보호하기 위한 방법.

청구항 17

직접 연결 저장 장치 (DAS) 시스템에서 캐시된 데이터의 무결성을 보호하기 위한 컴퓨터 프로그램이 저장된 컴퓨터-판독가능한 매체로서, 상기 컴퓨터 프로그램은 상기 컴퓨터-판독가능한 매체 상에 저장된 컴퓨터 명령들을 포함하는, 상기 컴퓨터-판독가능한 매체에 있어서, 상기 명령들은,

제 1 DAS 제어기에서 데이터를 수신하기 위한 제 1 세트의 명령들; 및

상기 제 1 DAS 제어기에 수신된 상기 데이터가 고체 상태 디스크(SSD) 어레이의 하나 이상의 SSD들 내의 블록들에 일시적으로 저장되고 후속하여 물리적 디스크 드라이브들(PDs)의 저가(또는 독립) 디스크들의 중복 어레이(Redundant Array of Inexpensive(or Independent) Disks; RAID) 어레이로서 구성된 하나 이상의 자기 하드 디스크 드라이브들(HDDs)에 저장되게 하는 캐싱 알고리즘을 상기 제 1 DAS 제어기에서 실행하기 위한 제 2 세트의 명령들을 포함하고,

각각의 블록은 데이터 무결성 필드(DIF)를 포함하고, 상기 데이터는 상기 데이터에 연관된 메타데이터를 갖고, 상기 캐싱 알고리즘은 상기 DAS 제어기에 수신된 상기 데이터에 연관된 상기 메타데이터가, 상기 SSD 어레이 내의, 상기 연관된 데이터가 저장되는 상기 블록들 중 적어도 하나의 블록에 연관된 상기 DIF들 중 적어도 하나의 DIF에 저장되게 하는, 컴퓨터-판독가능한 매체.

청구항 18

삭제

청구항 19

삭제

청구항 20

삭제

청구항 21

제 17 항에 있어서,

상기 DAS 제어기는 상기 SSD 어레이에 일시적으로 저장되는 상기 데이터에 캐시 일관성이 제공되게 하기 위해, 상기 SSD 어레이 내의 상기 데이터의 일시적 저장에 관련하여 RAID 기술의 한 레벨을 실행하도록 구성되는, 컴퓨터-판독가능한 매체.

청구항 22

삭제

청구항 23

제 17 항에 있어서,

상기 DAS 제어기에 의해 실행되는 상기 캐싱 알고리즘은 상기 메타데이터의 해시를 계산하는 해싱 알고리즘을 포함하고, 각각의 해시는 상기 SSD 어레이 내의, 상기 연관된 데이터가 저장되는 블록들 중 적어도 하나의 블록에 연관된 상기 DIF들 중 상기 적어도 하나의 DIF에 저장되는, 컴퓨터-판독가능한 매체.

청구항 24

제 17 항에 있어서,

상기 데이터에 연관된 상기 메타데이터는 상기 데이터의 소스의 표시, 상기 데이터 스트림의 길이의 표시, 및 상기 데이터가 수정되었는지의 여부의 표시를 포함하는, 컴퓨터-판독가능한 매체.

청구항 25

직접 연결 저장 장치 (DAS) 시스템에 있어서,

물리적 디스크 드라이브들(PDs)의 저가(또는 독립) 디스크들의 중복 어레이(Redundant Array of Inexpensive(or Independent) Disks; RAID) 어레이로서 구성된 복수의 자기 하드 디스크 드라이브들(HDDs);

캐시 메모리로서 구성된 고체 상태 디스크들(SSDs)의 어레이; 및

상기 PD들의 RAID 어레이 및 상기 SSD 어레이에 연결된 적어도 제 1 및 제 2 DAS 제어기들로서, 각각의 DAS 제어기는 중앙 처리 장치(CPU), 국부적 메모리 디바이스, 및 입력/출력(I/O) 인터페이스 디바이스를 갖고, 상기 CPU들 각각 및 상기 국부적 메모리 디바이스들 각각은 상기 PD들의 RAID 어레이의 RAID 구성에 일관된 기술의 RAID 레벨을 실행하도록 구성되고, 상기 CPU들 각각은, 상기 각각의 DAS 제어기에 수신된 데이터가 상기 SSD 어레이의 캐시 메모리에 일시적으로 저장되고 후속하여 상기 PD들의 RAID 어레이의 상기 PD들 중 하나 이상에 저장되게 하는 캐싱 알고리즘을 실행하도록 구성되고, 상기 각각의 CPU들은 상기 SSD 어레이에 일시적으로 저장되는 상기 데이터에 캐시 일관성이 제공되게 하기 위해, 상기 SSD 어레이 내의 상기 데이터의 일시적 저장에 관련하여 RAID 기술의 한 레벨을 실행하도록 구성되는, 상기 적어도 제 1 및 제 2 DAS 제어기들을 포함하는, 직접 연결 저장 장치 (DAS) 시스템.

청구항 26

제 25 항에 있어서,

상기 데이터는 상기 데이터에 연관된 메타데이터를 갖고, 상기 DAS 제어기들의 상기 국부적 메모리 디바이스들 각각은 캐시 메모리로서 사용되는 부분을 갖고, 상기 각각의 CPU들에 의해 실행되는 WB 캐싱 알고리즘은, 상기 각각의 DAS 제어기에 수신된 상기 데이터에 연관된 상기 메타데이터가 상기 각각의 국부적 메모리 디바이스의 각각의 캐시 메모리 부분에 일시적으로 저장되고 후속하여 PD들의 상기 RAID 어레이의 상기 PD들 중 하나 이상에 저장되게 하는, 직접 연결 저장 장치 (DAS) 시스템.

청구항 27

직접 연결 저장 장치 (DAS) 시스템에서 캐시된 데이터의 무결성을 보호하기 위한 방법에 있어서,

중앙 처리 장치(CPU), 국부적 메모리 디바이스, 및 입력/출력(I/O) 인터페이스 디바이스를 포함하는 제 1 DAS 제어기에서, 데이터를 수신하는 단계; 및

상기 제 1 DAS 제어기의 CPU에서, 상기 제 1 DAS 제어기에 수신된 상기 데이터가 고체 상태 디스크(SSD) 어레이의 하나 이상의 SSD들 내에 일시적으로 저장되고 후속하여 물리적 디스크 드라이브들(PDs)의 저가(또는 독립) 디스크들의 중복 어레이(Redundant Array of Inexpensive(or Independent) Disks; RAID) 어레이로서 구성된 하나 이상의 자기 하드 디스크 드라이브들(HDDs)에 저장되게 하는 캐싱 알고리즘을 실행하는 단계로서, 상기 CPU는, 상기 SSD 어레이에 일시적으로 저장되는 상기 데이터에 캐시 일관성이 제공되게 하기 위해, 상기 SSD 어레이 내의 데이터의 일시적 저장에 관련하여 RAID 기술의 한 레벨을 실행하도록 구성되는, 상기 제 1 DAS 제어기의 CPU에서 캐싱 알고리즘을 실행하는 단계를 포함하는, 직접 연결 저장 장치 (DAS) 시스템에서 캐시된 데이터의 무결성을 보호하기 위한 방법.

청구항 28

제 27 항에 있어서,

상기 데이터는 상기 데이터에 연관된 메타데이터를 갖고, 상기 DAS 제어기의 상기 국부적 메모리 디바이스는 캐시 메모리로서 사용되는 부분을 갖고, 상기 CPU에 의해 실행되는 WB 캐싱 알고리즘은, 상기 DAS 제어기에 수신된 상기 데이터에 연관된 상기 메타데이터가 상기 국부적 메모리 디바이스의 캐시 메모리 부분에 일시적으로 저장되고 후속하여 PD들의 상기 RAID 어레이의 상기 PD들 중 하나 이상에 저장되게 하는, 직접 연결 저장 장치(DAS) 시스템에서 캐시된 데이터의 무결성을 보호하기 위한 방법.

명 세 서

기 술 분 야

[0001] 관련 출원들에 대한 상호-참조

[0002] 본 출원은 전체를 참조로 여기에 포함시키는 "METHOD TO EFFICIENTLY USE SSD AS WB CACHE ELEMENT IN BOTH PRIVATE AND SHARED DAS CONFIGURATIONS" 명칭의 출원 번호가 61/268,055인 2009년 6월 8월에 출원된 미국 예비 특허 출원의 출원일에 대한 우선권 및 이의 이익을 청구한다.

[0003] 본 발명은 일반적으로 데이터 저장 시스템들에 관한 것으로, 특히 직접 연결 저장 장치 (DAS) 시스템에 캐시된 데이터의 무결성을 보호하기 위한 방법 및 장치에 관한 것이다.

배 경 기 술

[0004] 저장 어레이 또는 디스크 어레이는 복수의 자기 하드 디스크 드라이브들(HDD) 또는 유사한 영속 저장 유닛들을 포함하는 데이터 저장 디바이스이다. 저장 어레이는 많은 량의 데이터가 효율적으로 저장되게 할 수 있다. 서버 또는 워크스테이션은 저장 어레이가 서버 또는 워크스테이션에 국부적이 되게 저장 어레이에 직접 연결될 수도 있다. 서버 또는 워크스테이션이 저장 어레이에 직접 연결되는 경우들에 있어서, 저장 어레이는 전형적으로 직접 연결 저장 장치 (DAS) 시스템이라고 한다. 대안적으로, 서버 또는 워크스테이션은 저장 어레이 네트워크(SAN)를 통해 저장 어레이에 원격적으로 부착될 수도 있다. SAN 시스템들에서, 저장 어레이가 서버 또는 워크스테이션에 국부적이 아닐지라도, 어레이의 디스크 드라이브들은 서버 또는 워크스테이션의 운영 시스템(OS)에겐 국부적 부착된 것으로 여겨진다.

[0005] DAS 시스템들 및 SAN 시스템들은 흔히 저가(또는 독립) 디스크들의 중복 어레이(RAID) 시스템들로서 흔히 구성된다. RAID 시스템들은 저장 신뢰성을 개선하기 위해서 및/또는 입력/출력(I/O) 성능을 개선하기 위해서 저장 중복성(storage redundancy)을 이용한다. 일반적으로, RAID 시스템들은 더 큰 레벨들의 성능, 신뢰성 및/또는 더 큰 데이터 볼륨 크기들을 달성하기 위해서, 전형적으로 물리적 디스크 드라이브들(PDs)이라고 하는 2개 이상의 자기 HDD들을 동시에 이용한다. "RAID"라는 구(phrase)는 일반적으로 데이터를 분할하여 복수의 PD들 간에 데이터를 복제하는 컴퓨터 데이터 저장 기법들을 기술하기 위해 이용된다. RAID 시스템들에서, 하나 이상의 PD들은 RAID 가상 디스크 드라이브(VD)로서 셋업된다. RAID VD에서, 데이터는 복수의 PD들에 걸쳐 분산될 수 있을 것이지만, VD는 이용자에 의해서 그리고 서버 또는 워크스테이션의 OS에 의해서는 단일 디스크로서 보여진다.

[0006] RAID 시스템으로서 구성되는 DAS 시스템에서, DAS 제어기는 RAID 제어기로서 기능한다. 이러한 시스템에서, RAID 제어기는 이의 국부적인 메모리의 일부를 캐시 메모리로서 이용한다. 캐시 메모리는 PD들에 기입될 데이터를 임시로 저장하는데 이용된다. 이 목적을 위해 이용되는 하나의 유형의 캐시 메모리 구성은 기입 백(write back; WB) 캐시 메모리 구성으로서 알려져 있다. WB 캐시 메모리 구성들에서, 캐시 명령들은 전형적으로 데이터가 캐시 메모리로 이동되는 즉시 완료된다. 이러한 구성들에서, 캐시된 데이터의 무결성을 유지하는 것은 데이터가 일단 캐시되면 데이터가 PD들에 기입되기로 되는 사실에 기인하여 페일오버(failover) 또는 페일백(failback) 이벤트가 발생하는 경우에 문제가 될 수 있다. 결국, 페일오버 또는 페일백 이벤트의 발생으로 캐시된 데이터가 변질되는 결과가 확실히 초래되지 않게 하기 위해 조치들이 강구되어야 한다. 환원하여, DAS 시스템은 캐시 일관성(cache coherency)을 제공해야 한다. 캐시 일관성을 제공하기 위해서, 캐시된 데이터는 도 1 내지 도 3을 참조하여 이제 기술하는 바와 같이, 전형적으로 다른 메모리 디바이스에 복제된다.

[0007] 도 1은 RAID 기술을 구현하는 전형적인 DAS 시스템(2)의 블록도이다. 시스템(2)은 서버(3), RAID 제어기(4), 및 주변 상호연결(PCI) 버스(5)를 포함한다. RAID 제어기(4)는 중앙 처리 장치(CPU)(6), 메모리 디바이스(7), 및 I/O 인터페이스 디바이스(8)를 포함한다. 메모리 디바이스(7)의 저장 공간의 일부는 캐시 메모리로서 이용된다. 대안적으로, RAID 제어기(4)는 캐시 메모리로서 이용하기 위해 별도의 메모리 디바이스(도시되지 않음)를 포함할 수도 있다. I/O 인터페이스 디바이스(8)는 시리얼 어태치드 SCSI(SAS) 및/또는 시리얼 어드밴스드 테크놀로

지 어태치먼트(SATA) 표준들과 같은, 공지의 데이터 송신 프로토콜 표준들에 따라 데이터 송신을 실행하도록 구성된다. I/O 인터페이스 디바이스(8)는 복수의 PD들(9)에 및 이들로부터 데이터의 송신을 제어한다. RAID 제어기(4)는 PCI 버스(5)를 통해 서버 CPU(11) 및 서버 메모리 디바이스(12)와 통신한다. 서버 메모리 디바이스(12)는 서버 CPU(11)에 의한 실행을 위한 소프트웨어 프로그램들 및 데이터를 저장한다.

[0008] 전형적인 기입 동작 동안에, 서버 CPU(11)는 기입 요청 명령들을 PCI 버스(5)를 통해 RAID 제어기(4)에 보낸다. RAID 제어기(4)의 CPU(6)는 데이터가 RAID 제어기(4)의 메모리 디바이스(7)에서의 캐시 메모리에 일시적으로 저장되게 한다. 이어서 데이터는 메모리 디바이스(7)로부터 I/O 인터페이스 디바이스(8)를 통해 하나 이상의 PD들(9)에 송신된다. 메모리 디바이스(7)는 RAID VD의 가상 주소들과 PD들(9)의 물리적 주소들 간에 매핑을 실행하기 위한 코어 로직을 내장한다. RAID 제어기(4)의 CPU(6)는 시스템(2)의 RAID 레벨에 따라 패리티 계산들과 같은 계산들을 실행한다. 시스템(2)의 현재 RAID 레벨이 패리티를 이용하는 경우에, I/O 인터페이스 디바이스(8)는 패리티 비트들이 하나 이상의 PD들(9)에 저장되게 한다.

[0009] 전형적인 판독 동작 동안에, 서버 CPU(11)는 대응하는 판독 요청을 PCI 버스(5)를 통해 RAID 제어기(4)에 보낸다. 메모리 디바이스(7)에서의 보유된 로직을 이용하여, RAID 제어기 CPU(6)는 요청을 처리하고, 요청된 데이터가 메모리 디바이스(7)에서의 캐시 메모리에 보유되어 있다면, 요청된 데이터를 메모리 디바이스(7)의 캐시 메모리로부터 검색한다. 요청된 데이터가 메모리 디바이스(7)에서의 캐시 메모리에 보유되어 있지 않다면, RAID 제어기 CPU(6)는 요청된 데이터를 PD들(9)로부터 검색되게 한다. 검색된 데이터는 PCI 버스(5)를 통해 서버 CPU(11)에 송신되어 판독 요청을 충족한다.

[0010] 도 2는 도 1에 도시된 복수의 RAID 제어기들(4) 및 RAID 제어기들(4)에 의해 공유되는 도 1에 도시된 PD들(9)의 어레이를 포함하는 공지된 공유 DAS 시스템(23)의 블록도이다. 공유 DAS 시스템(23)에서 캐시 일관성을 제공하기 위해서, RAID 제어기들(4) 중 하나의 제어기의 메모리 디바이스(7)에 캐시된 데이터는 다른 RAID 제어기들(4) 중 하나의 제어기의 메모리 디바이스(7)에 복제 또는 미러되어(mirrored) RAID 제어기들(4)은 캐시 미러링(cache mirroring) 면에서 쌍이 된다. 캐시된 데이터의 복제는 도 2에 화살표들(24)로 나타내었다. 이러한 유형의 캐시 일관성 기술이 일반적으로 효과적이긴 하나, 주어진 쌍의 두 RAID 제어기들(4)에서 페일오버 또는 페일백 이벤트가 발생한다면, 이 미러된 쌍에 대한 캐시된 데이터의 무결성은 위태롭게 된다.

[0011] 도 3은 RAID 제어기들(4) 각각의 메모리 디바이스(7)에 캐시된 데이터를 다른 RAID 제어기들(4) 각각의 메모리 디바이스들(7)에 복제함으로써 캐시 일관성이 제공되는 도 2에 도시된 공유 DAS 시스템(23)의 블록도이다. 캐시된 데이터의 복제는 도 3에 화살표들(24, 25)로 나타내었다. 이러한 유형의 캐시 일관성 기술이 일반적으로 효과적이긴 하지만, 이러한 기술의 물리적 구현은 극히 복잡하고 많은 량의 대역폭을 이용한다. 또한, 시스템(23)이 확장되고(scaled out) 더 많은 수의 RAID 제어기들(4)이 시스템(23)에 추가될 때, 시스템(23)의 복잡성 및 캐시 미러링을 위해 이용되는 대역폭량은 지수함수적으로 증가한다. 이들 이유로, 이러한 캐시 일관성 해결책은 대부분의 경우에 비현실적이다.

[0012] DAS 시스템에서 캐시 일관성 문제에 대한 또 다른 해결책은 WB 캐시 구성 대신에 WT 캐시 구성을 이용하는 것이다. 그러나, WB 캐시 구성 대신에 WT 캐시 구성을 이용하는 것은 일반적으로 DAS 시스템의 I/O 성능을 저하시키며, 따라서 경쟁 시장에서 많은 저장 애플리케이션들엔 적합하지 않다. 캐시 일관성 문제가 SAN 제어기들을 이용하여 쉽게 처리될 수 있지만, 이러한 해결책은 비교적 고가이며 많은 경우들에 있어서 구현하기엔 너무 비용이 든다.

발명의 내용

해결하려는 과제

[0013] 따라서, 캐시된 데이터의 무결성을 적합하게 보호하며 DAS 시스템들에서 이용되는 공지된 캐시 일관성 해결책들의 전술한 한계들을 극복하는 필요성이 DAS 시스템에 대해서 존재한다.

과제의 해결 수단

[0014] 본 발명은 캐시된 데이터의 무결성을 보호하기 위한 DAS 시스템, 방법 및 컴퓨터-판독가능한 매체를 제공한다. DAS 시스템은 PD들의 RAID 어레이로서 구성된 복수의 자기 HDD들을 포함하고, 캐시 메모리로서 구성된 고체 상태 디스크들(SSDs)의 어레이, 및 PD들의 RAID 어레이와 SSD 어레이에 연결된 적어도 제 1 및 제 2 DAS 제어기들을 포함한다. 각각의 DAS 제어기는 CPU, 국부적 메모리 디바이스, 및 I/O 인터페이스 디바이스를 갖는다. CPU들 각각 및 국부적 메모리 디바이스들의 각각은 PD들의 RAID 어레이의 RAID 구성에 일관된 기술의 RAID 레벨을 실

행하도록 구성된다. CPU들 각각은 각각의 DAS 제어기에 수신된 데이터가 SSD 어레이의 캐시 메모리에 일시적으로 저장되고 후속적으로 PD들의 RAID 어레이의 PD들 중 하나 이상에 저장되게 하는 캐싱 알고리즘을 실행하도록 구성된다.

[0015] DAS 시스템에서 캐시된 데이터의 무결성을 보호하기 위한 방법은: 제 1 DAS 제어기에서, 데이터를 수신하는 단계; 제 1 DAS 제어기의 CPU에서, 제 1 DAS 제어기에 수신된 데이터가 SSD 어레이의 하나 이상의 SSD들에 일시적으로 저장되고 후속적으로 PD들의 RAID 어레이로서 구성된 하나 이상의 자기 HDD들에 저장되게 하는 캐싱 알고리즘을 실행하는 단계; 제 2 DAS 제어기에서, 데이터를 수신하는 단계; 및 제 2 DAS 제어기의 CPU에서, 제 2 DAS 제어기에 수신된 데이터가 SSD 어레이의 하나 이상의 SSD들에 일시적으로 저장되고 후속적으로 PD들의 RAID 어레이로서 구성된 하나 이상의 자기 HDD들에 저장되게 하는 캐싱 알고리즘을 실행하는 단계를 포함한다.

[0016] 컴퓨터-판독가능한 매체는 DAS 제어기에 의한 실행을 위한 제 1 세트의 명령들 및 제 2 세트의 명령들을 포함한다. 제 1 세트의 명령들은 DAS 제어기에서 데이터를 수신한다. 제 2 세트의 명령들은 제 1 DAS 제어기에 수신된 데이터가 SSD 어레이의 하나 이상의 SSD들에 일시적으로 저장되고 후속적으로 PD들의 RAID 어레이로서 구성된 하나 이상의 자기 HDD들에 저장되게 하는 캐싱 알고리즘을 제 1 DAS 제어기에서 실행한다.

[0017] 본 발명의 이들 및 다른 특징들 및 잇점들은 다음의 설명, 도면들 및 청구항들로부터 명백하게 될 것이다.

도면의 간단한 설명

[0018] 도 1은 RAID 기술을 구현하는 공지의 DAS 시스템의 블록도.

도 2는 도 1에 도시된 복수의 RAID 제어기들이 도 1에 도시된 PD들의 어레이를 공유하며, 쌍들의 RAID 제어기들의 메모리 디바이스들에 캐시된 데이터를 미러링함으로써 캐시 일관성이 제공되는 공유 DAS 시스템의 블록도.

도 3은 모든 다른 RAID 제어기들의 메모리 디바이스들에서의 각각의 RAID 제어기의 캐시된 데이터를 미러링함으로써 캐시 일관성이 제공되는 도 2에 도시된 공유 DAS 시스템의 블록도.

도 4는 DAS 시스템의 DAS 제어기들에 외부에 적어도 하나의 공유된 고체 상태 디스크(SSD)가 DAS 시스템의 PD들에 기입될 데이터를 캐시하기 위해 WB 캐시 메모리로서 이용되는 일 실시예에 따른 공유 DAS 시스템의 블록도.

도 5는 도 4에 도시된 DAS 제어기들 중 하나의 제어기의 블록도.

도 6은 일 실시예에 따라 도 4에 도시된 DAS 제어기들 중 하나의 제어기의 CPU에 의해 실행되는 WB 캐싱 알고리즘을 나타내는 흐름도.

도 7은 연관된 메타데이터가 DAS 제어기들 내부의 캐시 메모리에 캐시되고 데이터가 SSD 어레이에 캐시되는 일 실시예에 따른 도 4에 도시된 DAS 시스템의 블록도.

도 8은 데이터 및 연관된 메타데이터가 SSD 어레이에 캐시되는 일 실시예에 따른 도 4에 도시된 DAS 시스템의 블록도.

도 9는 도 8에 도시된 DAS 제어기들에 의해 실행되는 WB 캐싱 알고리즘을 나타내는 흐름도.

발명을 실시하기 위한 구체적인 내용

[0019] 본 발명에 따라서, DAS 시스템의 DAS 제어기들 외부에 하나의 어레이의 고체 상태 디스크들(SSD)이 DAS 제어기들에 의해 WB 캐시 동작들을 실행하기 위해 WB 캐시 메모리로서 이용되는, RAID 기술을 구현하는 DAS 시스템이 제공된다. 외부 SSD 어레이를 WB 캐시 메모리로서 이용하는 것은 DAS 시스템의 복잡성을 현저히 증가시키지도 않고 캐시 동작들을 실행하기 위해 이용되는 대역폭량을 증가시키지 않으면서 DAS 시스템이 완전히 캐시 코히런트되게 한다. 또한, 외부 SSD를 WB 캐시 메모리로서 이용하는 것은 도 1 내지 도 3을 참조하여 위에 기술된 방식으로 DAS 제어기들을 미러링할 필요성을 제거한다.

[0020] 도 4는 DAS 시스템(100)의 SSD 어레이(110)가 WB 캐시 메모리로서 DAS 시스템(100)의 복수의 DAS 제어기들(120)에 의해 공유되는 일 실시예에 따른 본 발명의 DAS 시스템(100)의 블록도이다. 도 5는 도 4에 도시된 DAS 제어기들(120) 중 하나의 제어기의 블록도이다. DAS 시스템(100)의 DAS 제어기들(120)은 도 1에 도시된 RAID 제어기(4)의 구성과 동일 또는 유사한 구성들을 갖는다. DAS 제어기들(120) 각각은 캐시 동작들에 관해서는 제외하고, RAID 제어기(4)가 동작하는 바와 동일한 방식으로 동작한다. 이에 따라, 도 1에 도시된 RAID 제어기(4)에 대한 경우와 같이, 도 4에 도시된 DAS 제어기들(120) 각각은 RAID 제어기로서 구성된다. DAS 시스템(100)은

RAID 기술을 채용한다. 본 발명은 DAS 시스템(100)에 채용되는 RAID의 레벨에 관하여 제한되지 않는다.

[0021] RAID는 상이한 시스템 설계들에 대응하는 7개의 기본 레벨들을 가지며, 이들 레벨들 중 임의의 하나 이상은 DAS 시스템(100)에 구현될 수 있다. 상이한 RAID 레벨들이 공지되어 있을지라도, 이들 RAID 레벨들의 논의가 이제 제공될 것이다. 전형적으로 RAID 레벨들(0 내지 6)이라고 하는 7개의 기본 RAID 레벨들은 다음과 같다. RAID 레벨 0은 개선된 데이터 신뢰성 및 증가된 I/O 성능을 달성하기 위해 스트라이핑을 이용한다. "스트라이핑(striping)"이라는 용어는 단일 데이터 파일과 같이 논리적으로 순차적인 데이터가 단편화되고(fragmented) 라운드-로빈(round-robin) 형태로 복수의 PD들에 할당됨을 의미한다. 이에 따라, 데이터가 기입될 때 복수의 PD들에 걸쳐 데이터가 "스트라이프된다(striped)"라고 한다. 스트라이핑은 성능을 개선하며 추가의 저장 용량을 제공한다. RAID 레벨 1은 패리티없이 미러링을 이용한다. "미러링"이라는 용어는 데이터가 연속적으로 이용될 수 있게 하기 위해서 데이터가 실시간으로 개별적 PD들에 복제됨을 의미한다. 이러한 유형의 복제는 데이터 중복성을 제공한다. RAID 레벨 2는 중복성 및 스트라이핑을 이용한다. RAID 레벨 2에서, 중복성은 PD들 상의 비트들에 대해 계산되어 복수의 PD들 상에 저장되는 해밍 코드들을 이용하여 달성된다. PD가 고장난 경우, 데이터를 재구성하기 위해서 패리티 비트들이 이용될 수 있다.

[0022] RAID 레벨 3 시스템들은 인터리빙된 패리티 비트들(interleaved parity bits) 및 전용 패리티 PD를 결합한 바이트-레벨 스트라이핑(byte-level striping)을 이용한다. 바이트-레벨 스트라이핑 및 중복성의 이용은 성능을 개선하고 내고장성(fault tolerance)을 갖춘 시스템을 제공한다. RAID 레벨 3 시스템은 패리티 없이 계속하여 동작할 수 있고 패리티 PD가 고장난 경우에 어떠한 성능 패널티도 겪지 않는다. RAID 레벨 4는 RAID 레벨 4 시스템들이 바이트-레벨 또는 워드-레벨 스트라이핑 대신 블록-레벨 스트라이핑을 채용하는 것을 제외하곤 근본적으로 RAID 레벨 3과 동등하다. 각각의 스트라이프는 비교적 크기 때문에, 단일 파일은 블록으로 저장될 수 있다. 각각의 PD는 독립적으로 동작하며 많은 상이한 I/O 요청들은 병렬로 취급될 수 있다. 오류 검출은 블록-레벨 패리티 비트 인터리빙을 이용함으로써 달성된다. 인터리빙된 패리티 비트들은 별도의 단일 패리티 PD에 저장된다.

[0023] RAID 레벨 5는 분산된 패리티와 함께 스트라이핑을 이용한다. 분산 패리티를 구현하기 위해서, 시스템이 동작하기 위해 PD들 중 하나를 제외한 모든 PD들이 있어야 한다. PD들 중 어느 하나의 PD의 고장은 PD의 대체를 필요하게 한다. 그러나, PD들 중 단일 PD의 고장은 시스템이 고장나게 하지 않는다. RAID 레벨 6은 이중 분산된 패리티와 함께 스트라이핑을 이용한다. RAID 레벨 6 시스템들은 적어도 4개의 PD들의 이용을 필요로 하며, PD들 중 2개는 분산 패리티 비트들을 저장하기 위해 이용된다. 시스템은 2개의 PD들이 고장나더라도 계속하여 동작할 수 있다. 이중 패리티는 각각의 VD가 상당수의 PD들로 구성되는 시스템들에서 점점 더 중요해지고 있다. 단일 패리티를 이용하는 RAID 레벨 시스템들은 고장난 드라이브가 복구될 때까지 데이터 손실에 취약하다. RAID 레벨 6 시스템들에서, 이중 패리티의 이용은 다른 VD들 중 하나의 VD의 PD가 첫번째 고장난 PD의 복구 완료 전에 고장난 경우에 고장난 PD를 가지는 VD가 데이터 손실의 위험 없이 복구될 수 있게 한다.

[0024] 다시 도 5를 참조하면, DAS 제어기(120)는 CPU(130), 메모리 디바이스(140), 및 I/O 인터페이스 디바이스(150)를 포함한다. I/O 인터페이스 디바이스(150)는 PD들(129)에 및 이들로부터 데이터의 송신을 제어한다. I/O 인터페이스 디바이스(150)는 전형적으로 예를 들면, 전매 데이터 송신 프로토콜들 뿐만 아니라, 그외 공지된 데이터 송신 프로토콜들이 이 목적을 위해 이용될 지라도, SAS 및/또는 SATA 표준들 및 이들의 변형들과 같은, 공지된 데이터 송신 프로토콜 표준들에 따라 데이터 송신을 실행하도록 구성된다.

[0025] DAS 시스템(100)의 동작들이 이제 도 4 및 도 5를 참조하여 기술될 것이다. 전형적인 기입 동작 동안, DAS 제어기(120)의 CPU(130)는 PD들(129) 중 하나 이상에 기입될 데이터를 외부 서버 또는 워크스테이션(도시되지 않음)으로부터 수신한다. DAS 제어기 CPU(130)는 PD들(129)에 기입될 데이터를 일시적으로 저장하기 위해서 SSD 어레이(110)를 WB 캐시 메모리로서 이용한다. SSD(110)에 데이터를 WB 캐시하는 프로세스가 도 6을 참조하여 더 상세히 이하 기술된다. 데이터가 SSD(110)에 WB 캐시된 이후에 어떤 시점에서, DAS 제어기 CPU(130)는 캐시된 데이터가 SSD(110)로부터 송신되고 하나 이상의 PD들(129)에 저장되게 한다. 메모리 디바이스(140)는 RAID VD의 가상 주소들과 PD들(129)의 물리적 주소들 간에 매핑을 실행하기 위한 코어 로직을 포함한다. DAS 제어기(120)는 캐시된 데이터가 PD들(129)에서의 대응하는 물리적 주소들에 저장되게 한다. DAS 제어기 CPU(130)는 DAS 시스템(100)의 RAID 레벨에 따라 패리티 계산들과 같은 계산들을 실행한다. 예를 들면, DAS 시스템(100)의 RAID 레벨이 패리티를 이용한다면, DAS 제어기 CPU(130)는 패리티 비트들을 계산하고 I/O 인터페이스 디바이스(150)는 하나 이상의 PD들(129)에 패리티 비트들이 저장되게 한다.

[0026] 전형적인 판독 동작 동안에, DAS 제어기 CPU(130)는 외부 서버 또는 워크스테이션(도시되지 않음)으로부터 판독 요청을 수신하며, 메모리 디바이스(140)에 보유된 로직을 이용하여, 데이터가 판독되어야 하는 하나 이상의 PD

들(129)의 물리적 주소들을 결정하기 위해 판독 요청을 처리한다. 이어서 DAS 제어기 CPU(130)는 데이터가 PD들(129)에 있는 주소들로부터 요청된 데이터를 검색하게 하고 외부 서버 또는 워크스테이션(도시되지 않음)에 보내지게 한다. 메모리 디바이스(140)의 일부 또는 DAS 제어기(120) 내에 이외 다른 메모리 디바이스(도시되지 않음)이 판독 캐시 메모리로서 이용될 수 있고, 이 경우 CPU(130)는 요청된 데이터가 판독 캐시 메모리에 보유된 것으로 CPU(130)가 결정한 경우 PD들(129)로부터가 아니라 판독 캐시 메모리로부터 데이터를 판독할 것이다.

[0027] 도 6은 일 실시예에 따라 DAS 제어기들(120) 중 하나의 제어기의 CPU(130)에 의해 실행되는 WB 캐싱 알고리즘을 나타내는 흐름도이다. DAS 제어기들(120) 각각의 CPU들(130)은 WB 캐싱 알고리즘을 실행한다. 그러나, 간략성을 위해서, DAS 제어기들(120) 중 하나의 제어기만을 참조하여 알고리즘이 기술될 것이다. 서버 또는 워크스테이션(도시되지 않음)이 PD들(129)에 기입될 DAS 제어기(120)에 데이터를 보낼 때, DS 제어기(120)는 블록(201)으로 나타낸 바와 같이, 데이터를 수신한다. 이어서 CPU(130)는 블록(203)으로 나타낸 바와 같이, 캐시 "히트(hit)" 또는 캐시 "미스(miss)"가 발생하였는지의 여부를 결정하기 위해 수신된 데이터를 처리한다. 캐시 "히트"는 SSD(110)에서의 캐시 메모리에 데이터가 현재 보유되어 있는 것으로 CPU(130)가 결정하였음을 의미한다. 캐시 "미스"는 캐시 메모리에 데이터가 현재 보유되어 있지 않은 것으로 CPU(130)가 결정하였음을 의미한다.

[0028] CPU(130)가 캐시 미스가 발생한 것으로 결정하였다면, CPU(130)는 블록(205)으로 나타낸 바와 같이, 데이터가 SSD 어레이(110)의 캐시 메모리에 기입되게 한다. 데이터가 SSD 어레이(110)의 캐시 메모리에 기입된 이후에 어떤 시점에서, CPU(130)는 블록(206)으로 나타낸 바와 같이, PD들(129)에서의 물리적 주소들에 데이터가 저장되게 한다. 블록(203)에서, CPU(130)가 캐시 히트가 발생한 것으로 결정하면, CPU(130)는 블록(206)으로 나타낸 단계에서 SSD 어레이(110)의 캐시 메모리에 보유된 대응하는 데이터가 SSD 어레이(110)에서의 대응하는 물리적 주소들에 저장되게 한다.

[0029] 도 1 내지 도 3을 참조하여 위에 기술된 공지된 캐시 일관성 방법들과는 반대로, 본 발명에 따라서, SSD 어레이(110)에 대해 어떤 레벨의 RAID 기술을 이용함으로써 캐시 일관성이 제공된다. 구체적으로, DAS 제어기들(120)이 SSD 어레이(110)에서의 캐시 메모리에 데이터를 저장할 때, RAID 기술은 데이터가 캐시되는 SSD 어레이(110)의 SSD가 고장난 경우에 데이터가 확실히 복구될 수 있게 하는데 이용된다. 예를 들면, RAID 레벨 0은 DAS 제어기들(120) 각각이 SSD 어레이(110)에서의 캐시 메모리에 데이터를 저장할 때 데이터가 SSD 어레이(110)의 복수의 SSD들에 걸쳐 스트라이프되게 하기 위해 이용될 수 있다. 예를 들면, RAID 레벨 1이 이용된다면, DAS 제어기들(120) 각각이 SSD 어레이(110)에서의 캐시 메모리에 데이터를 저장할 때, 데이터가 SSD 어레이(110)의 복수의 SSD들에 복제 또는 미러된다. SSD 어레이(110)의 SSD들 중 하나의 SSD가 고장난다면, SSD 어레이(110)에 구현되는 기술의 RAID 레벨은 데이터가 복구되게 할 것이다. 이에 따라, DAS 시스템(100)은 완전히 캐시 코히런트하게 된다. 본 발명은 SSD 어레이(110)의 캐시 메모리에 캐시되는 데이터에 대해 캐시 일관성을 보장하기 위해 이용되는 RAID 레벨에 관하여 제한되지 않는다.

[0030] 또한, SSD 어레이(110)는 전형적으로, 각각의 DAS 제어기들(120)에 의해 이용되는 각각의 부분들로 파티션되나 필수적이진 않다. 예를 들면, N은 1 이상인 양의 정수인, 총 N개의 DAS 제어기들(120)이 있다고 할 때, SSD 어레이(110)의 저장 용량은 N개의 같은 부분들로 분할될 것이며, 각 부분은 각각의 DAS 제어기(120)에 의해 이용된다. 이러한 식으로 SSD 어레이(110)를 파티션하는 것은 DAS 제어기들(120)이 SSD 어레이(110)에 액세스할 때 액세스 충돌을 피하게 한다. 그러나, DAS 제어기들(120) 중 하나가 고장난 경우, 다른 DAS 제어기들(120) 중 하나의 제어기는 SSD 어레이(110)에 저장되어 있고 고장난 DAS 제어기(120)에 연관된 데이터에 액세스할 수 있다.

[0031] 공지된 바와 같이, 데이터는 이와 연관되고 예를 들면 데이터의 소스의 아이덴티티(identity)(즉, 해시 캐시 태그들), 데이터 스트림의 길이, 및 데이터가 수정되었는지의 여부(즉, 상태 표시)와 같은, 데이터의 속성들을 규정하는 메타데이터를 갖는다. DAS 시스템들에서, 메타데이터는 전형적으로 DAS 제어기 내부의 동적 랜덤 액세스 메모리(DRAM)에 저장된다. 본 발명의 하나의 실시예에 따라, 메타데이터는 DAS 제어기(120)의 메모리 디바이스(140)에 저장되는 반면 대응하는 데이터는 도 4 내지 도 6을 참조하여 위에 기술된 방식으로 SSD 어레이(110)의 캐시 메모리에 저장된다. 메타데이터를 취급하기 위한 이 실시예는 이제 도 7을 참조하여 기술된다.

[0032] 도 7은 데이터가 SSD 어레이(110)에 캐시되고 연관된 메타데이터가 DAS 제어기들(120) 내부의 캐시 메모리에 캐시되는 일 실시예에 따른 도 4에 도시된 DAS 시스템(100)의 블록도이다. 이 실시예에 따라, DAS 제어기들(120) 중 하나의 제어기의 CPU(130)가 데이터를 SSD 어레이(110)의 캐시 메모리에 저장할 때, 대응하는 메타데이터는 DAS 제어기(120)의 메모리 디바이스(140)의 캐시 메모리 부분(도시되지 않음)에 또는 DAS 제어기(120)의 어떤 다른 메모리 디바이스(도시되지 않음)에 저장된다. 또한, 메타데이터에 대한 캐시 일관성을 제공하기 위해서, DAS 제어기(120)에서의 캐시 메모리에 저장되는 메타데이터는 이웃한 DAS 제어기들(120) 간을 지나는 화살표들

로 도 7에 나타낸 바와 같이, 하나 이상의 다른 DAS 제어기들(120)에서의 캐시 메모리에 미러된다. 이렇게 하여, DAS 제어기들(120) 중 하나가 고장나더라도, 메타데이터가 미러된 다른 DAS 제어기(120)의 캐시 메모리로부터 대응하는 메타데이터가 복구될 수 있다.

[0033] 도 8은 데이터 및 연관된 메타데이터가 SSD 어레이(110)에 캐시된 일 실시예에 따른 도 4에 도시된 DAS 시스템(100)의 블록도이다. SSD들에서, 데이터는 전형적으로 미리 결정된 수의 바이트들(B)의 블록들, 예를 들면 블록당 520 B로 기입된다. 블록들의 형식은 INCITS(InterNational Committee on Information Technology Standards)의 표준 T10에 의해 통제된다. 표준 T10은 순환 중복 체크(cyclic redundancy check; CRC) 비트들, 애플리케이션 태그 비트들, 및 참조 태그 비트들과 같은 보호 정보를 포함하는 데이터 무결성 필드(data integrity field; DIF)를 각각의 데이터 블록이 포함함을 제공한다. 표준 T10에 의해 규정되는 DIF는 각 블록의 끝에 8 B로 구성된다. 그러므로, 각각의 블록은 512 B의 데이터와 8 B의 DIF로 구성된다. 이 실시예에 따라, 데이터 블록에 연관된 메타데이터를 나타내기 위해서 각 데이터 블록의 8 B DIF가 이용된다. 이에 따라, 데이터 및 이의 연관된 메타데이터가 SSD 어레이(110)의 캐시 메모리에 함께 저장된다.

[0034] 데이터 및 이의 연관된 메타데이터를 함께 캐시하는 것은 동작들 동안에 파워 고장의 경우에, 데이터가 업데이트되고 이의 연관된 메타데이터는 업데이트되지 않거나, 또는 데이터는 업데이트되지 않고 이의 연관된 메타데이터가 업데이트되는 것을 보장한다. 예를 들면, 데이터 및 메타데이터가 상이한 메모리 디바이스들에 서로 무관하게 캐시된다면, 하나의 메모리 디바이스의 파워 고장은 고장난 메모리 디바이스에 저장된 데이터 또는 메타데이터는 업데이트되지 않게 하고 파워가 고장나지 않은 다른 메모리 디바이스에 저장된 데이터 또는 메타데이터는 업데이트되게 한다. SSD 어레이(110)의 캐시 메모리에 데이터 및 메타데이터를 함께 캐시하는 것은 이 문제를 제거한다.

[0035] 이 예시적인 실시예에 따라, SSD 어레이(110)의 캐시 메모리에서의 캐시 라인은 64 KB로 구성되고, 여기에서 k = 1024이다. 주어진 캐시 라인의 각각의 블록은 512 B의 데이터로 구성되기 때문에, 각 캐시 라인은 128 블록들로 구성된다(즉, (64 B x 1024)/512 B = 128). 주어진 캐시 라인의 128 블록들의 각각의 DIF는 8 B의 메타데이터로 구성된다. 그러나, 각각의 캐시 라인에 대해서, 단지 약 64 B의 메타데이터만이 필요하다. 64 B의 메타데이터는 8 B의 부분들로 분할된다. 이에 따라, 주어진 라인에 대해 이용되는 8 블록들의 DIF들만이 메타데이터용으로 필요하게 된다. 실시예에 따라, 각각의 128-블록 캐시 라인의 첫 번째 8개의 블록들은 캐시 라인에 포함된 데이터에 연관된 8 B의 메타데이터용으로 할당된다. 중복성을 제공하기 위해서, 동일한 8 B의 메타데이터가 캐시 라인의 다음 8 블록들의 8개의 DIF들의 각각에 복제된다. 64 B의 메타데이터의 해시는 캐시 라인의 마지막 8 블록들의 8개의 DIF들에 포함된다. 이 방식은 캐시 라인에 기입되는 첫 번째 및 마지막 블록들이 해시 의해 서로 연관되는 점에서 원자성(atomicity)을 보장한다.

[0036] 도 9는 도 8에 도시된 DAS 제어기들(120)에 의해 실행되는 WB 캐싱 알고리즘을 나타내는 흐름도이다. 이 알고리즘은 도 5, 도 8 및 도 9를 참조하여 기술될 것이다. DAS 제어기들(120) 각각의 CPU들(130)은 WB 캐싱 알고리즘을 실행한다. 그러나, 간략성을 위해서, 알고리즘은 DAS 제어기들(120) 중 하나의 제어기만을 참조하여 기술될 것이다. 서버 또는 워크스테이션(도시되지 않음)이 데이터를 PD들(129)에 기입될 DAS 제어기(120)에 보낼 때, DAS 제어기(120)는 블록(301)으로 나타낸 바와 같이, 데이터 및 이의 연관된 메타데이터를 수신한다. 이어서 CPU(130)는 블록(303)으로 나타낸 바와 같이, 캐시 "히트" 또는 캐시 "미스"가 발생하였는지의 여부를 결정하기 위해 수신된 데이터를 처리한다.

[0037] CPU(130)가 캐시 미스가 발생한 것으로 결정하였다면, CPU(130)는 블록(305)으로 나타낸 바와 같이, 메타데이터 해시를 계산하고 데이터, 메타데이터 및 메타데이터 해시가 SSD 어레이(110)의 캐시 메모리에 기입되게 한다. 메타데이터 해시를 계산하기 위한 다양한 공지된 해싱 알고리즘들이 존재한다. 이 목적을 위해 임의의 적합한 공지된 해싱 알고리즘이 이용될 수 있다. 그러므로, 간략성을 위해서, 메타데이터 해시를 계산하기 위해 이용되는 알고리즘을 여기에선 기술되지 않을 것이다.

[0038] SSD 어레이(110)의 캐시 메모리에 데이터가 기입된 이후에 어떤 시점에서, CPU(130)는 블록(306)으로 나타낸 바와 같이, 데이터 및 이의 연관된 메타데이터가 PD들(129)에서의 물리적 주소들에 저장되게 한다. 블록(303)에서, CPU(130)가 캐시 히트가 발생한 것으로 결정하면, CPU(130)는 블록(306)으로 나타낸 단계에서 SSD 어레이(110)의 캐시 메모리에 보유된 대응하는 데이터 및 메타데이터가 SSD 어레이(110)에서의 대응하는 물리적 주소들에 저장되게 한다.

[0039] 도 6 및 도 9를 참조로 위에 기술된 WB 캐싱 알고리즘은 다양한 방식들로 구현될 수 있음에 유의한다. WB 알고리즘들은 전형적으로 단지 하드웨어로만 또는 하드웨어 및 소프트웨어 또는 펌웨어의 조합으로 CPU들(130)에서

실행된다. 이 목적을 위해 이용되는 소프트웨어 또는 펌웨어 명령들은 예를 들면, DAS 제어기들(120)의 메모리 디바이스들(140)에서와 같은, 컴퓨터-판독가능한 매체에 저장된다.

[0040] 위에 기술된 실시예들에 대해 많은 변형들이 행해질 수 있고 모든 이러한 변형들은 본 발명의 범위 내에 있음을 당업자는 알 것이다. 예를 들면, DIF를 이용하지 않는 SSD들은 시중에서 구입할 수 있다. 이러한 SSD들은 데이터, 메타데이터 및 메타데이터 해시들을 캐시하기 위해 본 발명과 함께 이용에 적합하다. 본 발명은 SSD 어레이(110)에서의 SSD들의 어떤 특정한 유형 또는 구성을 이용하는 것으로 국한되지 않는다. 또한 본 발명은 DAS 제어기(120)의 구성에 관하여 제한되지 않는다. 도 5에 도시된 DAS 제어기(120)의 구성은 본 발명과 함께 이용에 적합한 DAS 제어기 구성의 단지 일례이다.

[0041] 본 발명은 본 발명의 원리들 및 개념들을 보일 목적으로 실시예들을 참조하여 위에 기술되었음에 유의해야 한다. 당업자들은 많은 수정들이 여기 기술된 실시예들에 대해 행해질 수 있고 모든 이러한 수정들이 본 발명의 범위 내에 있음을 알 것이다.

부호의 설명

[0042] 100: DAS 시스템

110: SSD 어레이

120: DAS 제어기

129: 물리적 디스크 드라이브들

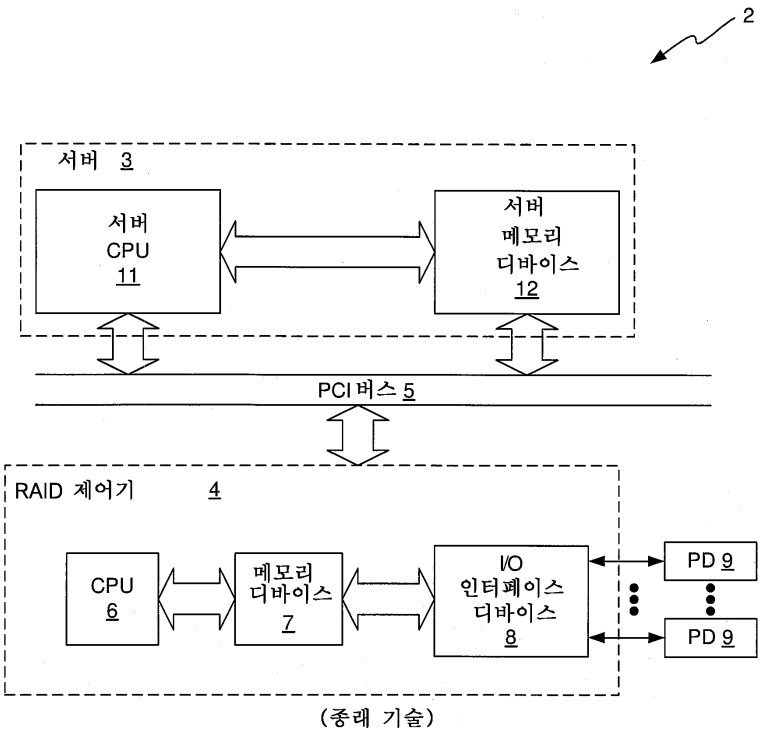
130: 중앙 처리 장치

140: 메모리 디바이스

150: 입력/출력 인터페이스 디바이스

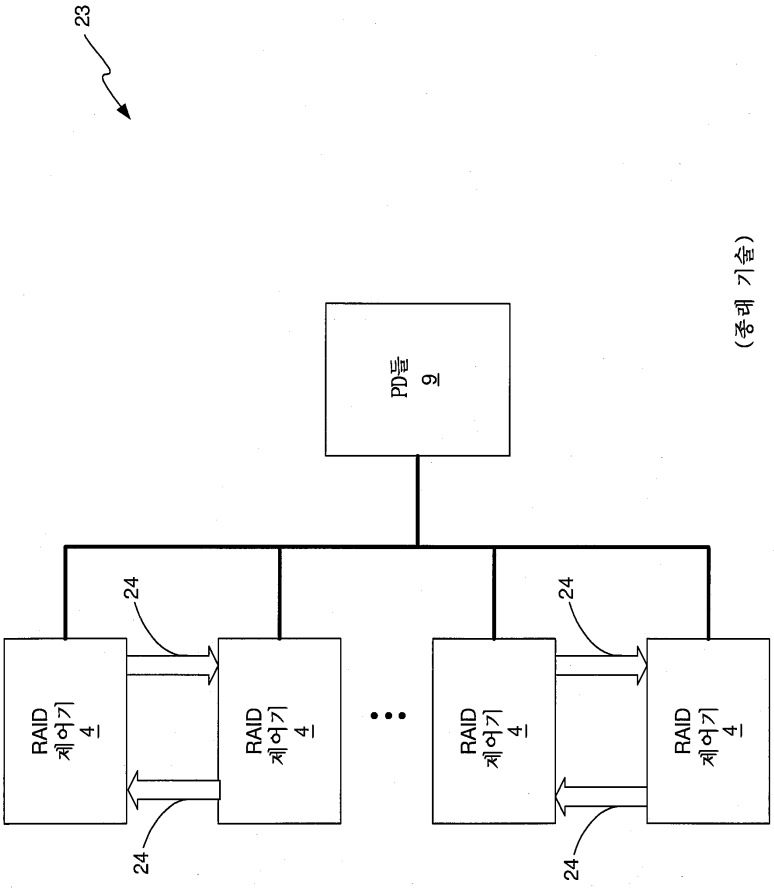
도면

도면1

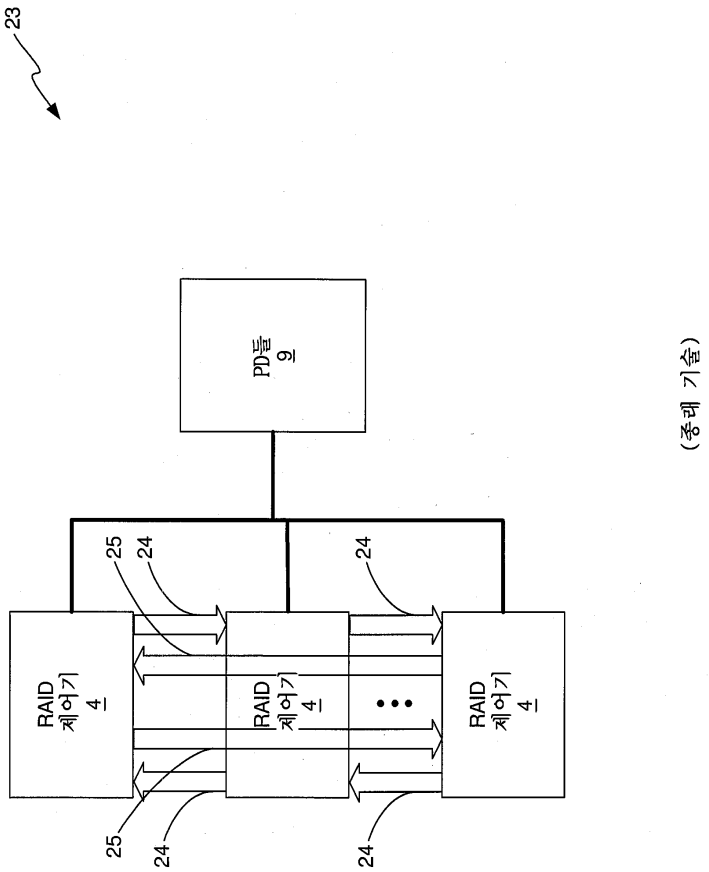


(종래 기술)

도면2

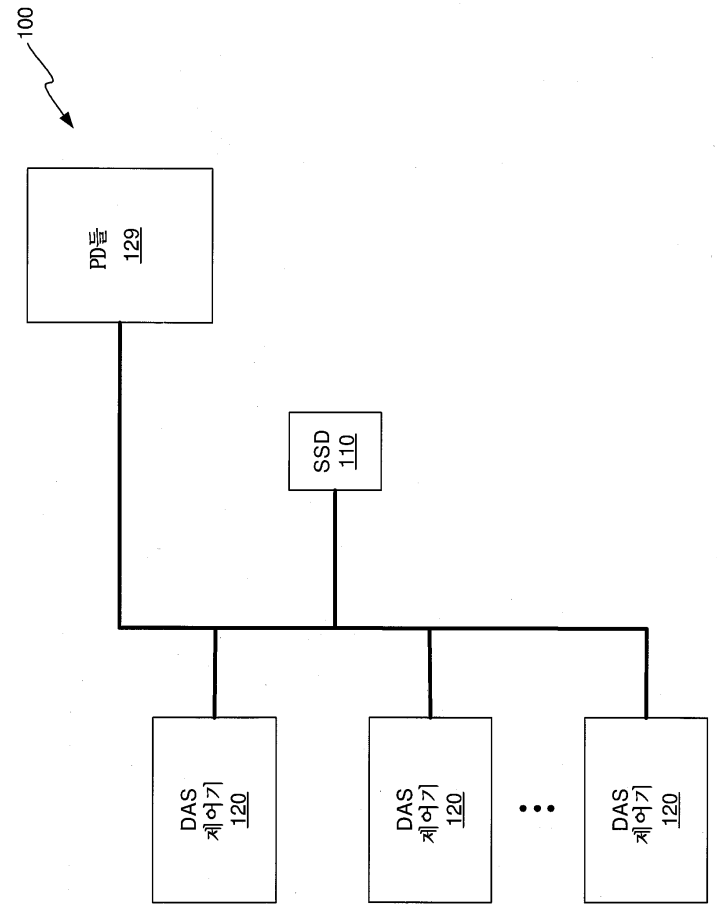


도면3

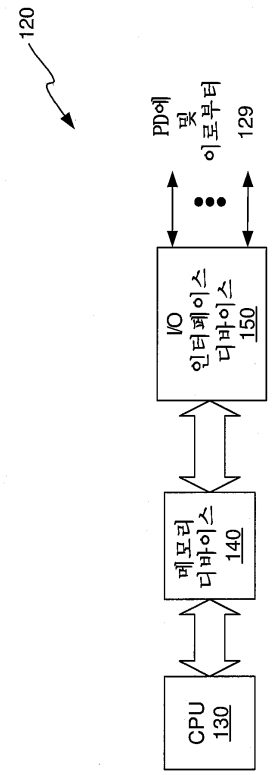


(종래 기술)

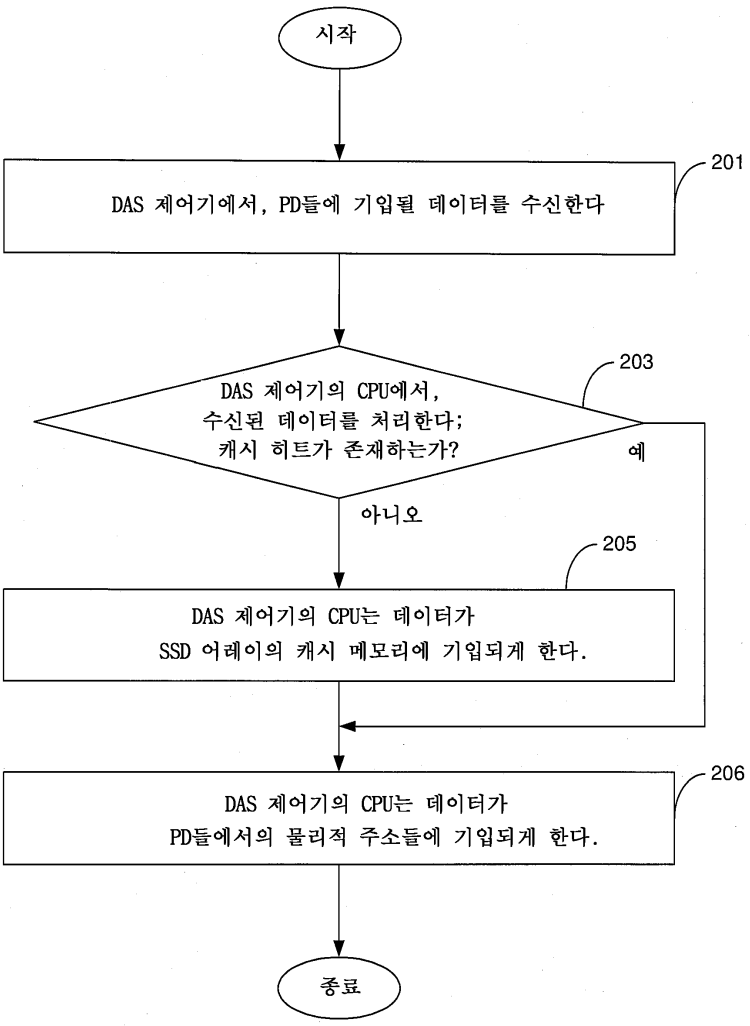
도면4



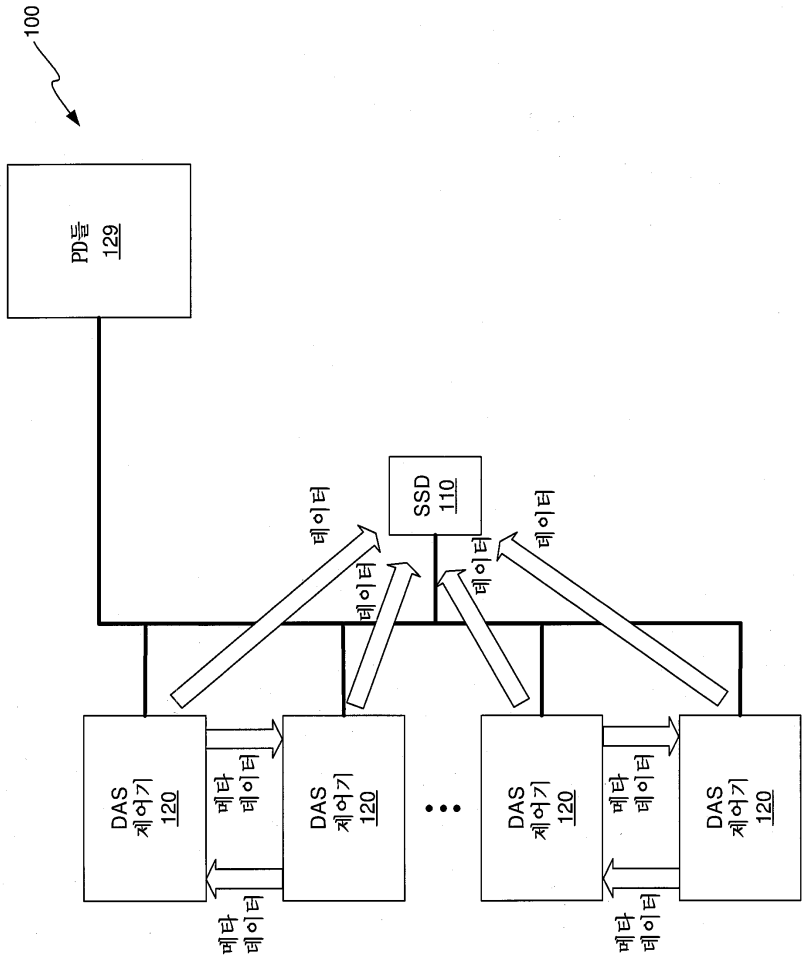
도면5



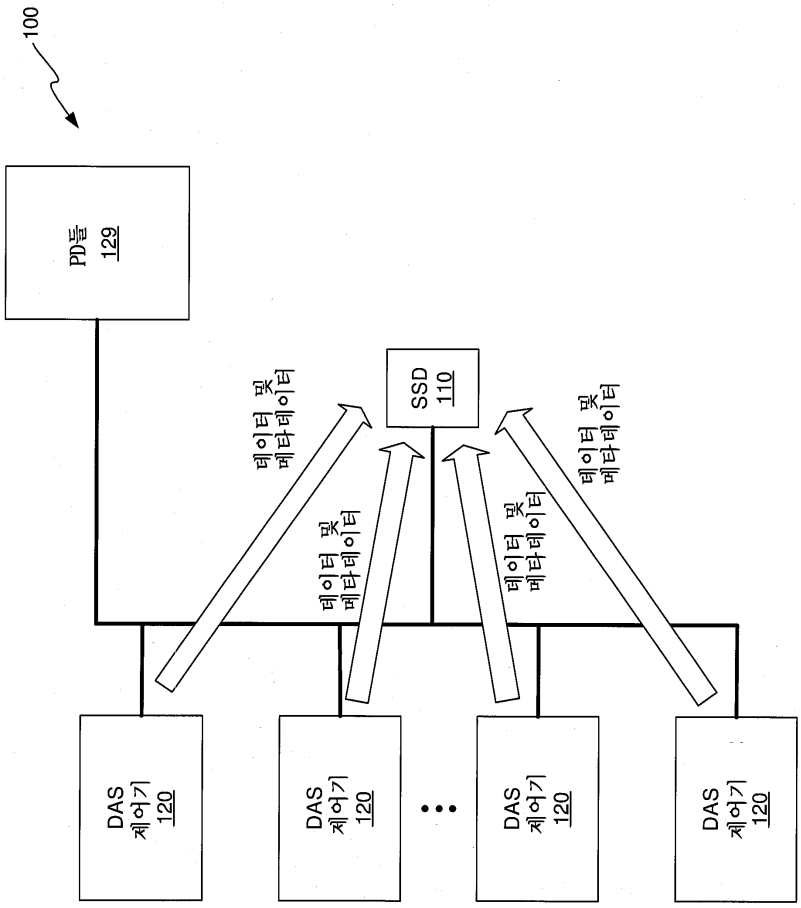
도면6


시작
DAS 제어기에서, PD들에 기입될 데이터를 수신한다
201
DAS 제어기의 CPU에서,
수신된 데이터를 처리한다;
캐시 히트가 존재하는가?
203
예
아니오
205
DAS 제어기의 CPU는 데이터가
SSD 어레이의 캐시 메모리에 기입되게 한다.
DAS 제어기의 CPU는 데이터가
PD들에서의 물리적 주소들에 기입되게 한다.
206
종료

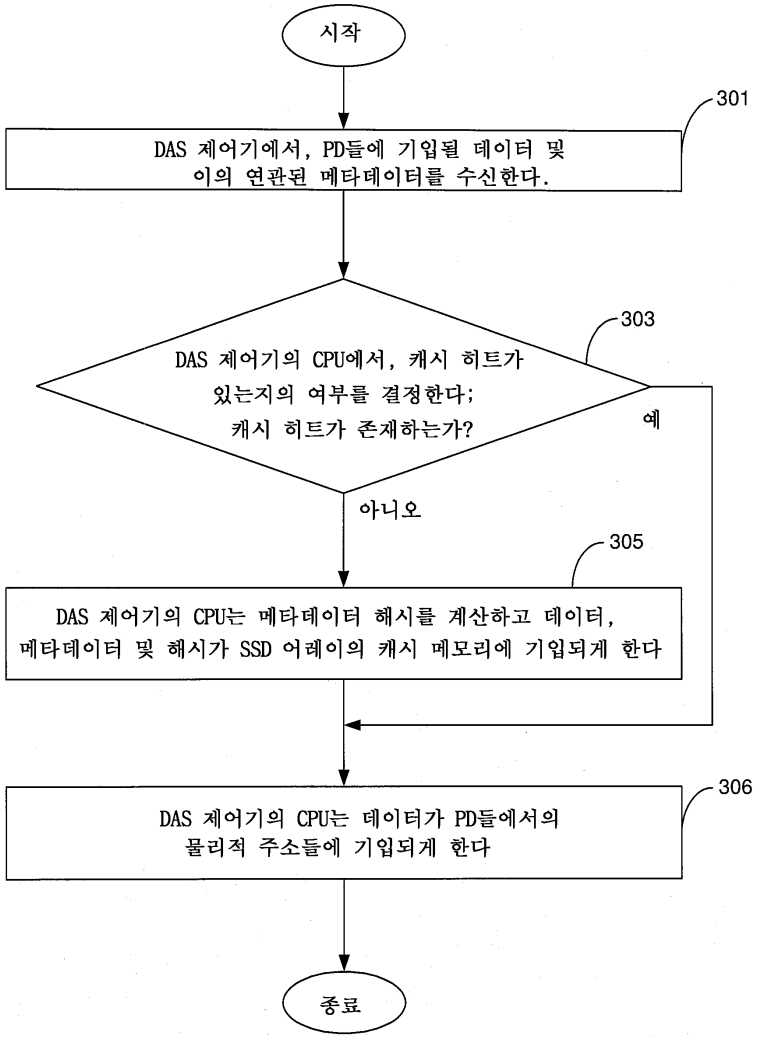
도면7


100
PD들 129
SSD 110
데이터
데이터
데이터
데이터
DAS 제어기 120
DAS 제어기 120
DAS 제어기 120
DAS 제어기 120
메타 데이터
메타 데이터
메타 데이터
메타 데이터

도면8


100
PD들
129
SSD
110
데이터 및
메타데이터
데이터 및
메타데이터
데이터 및
메타데이터
데이터 및
메타데이터
DAS
제어기
120
DAS
제어기
120
DAS
제어기
120
DAS
제어기
120

도면9


시작
301
DAS 제어기에서, PD들에 기입될 데이터 및
이의 연관된 메타데이터를 수신한다.
303
DAS 제어기의 CPU에서, 캐시 히트가
있는지의 여부를 결정한다;
캐시 히트가 존재하는가?
예
아니오
305
DAS 제어기의 CPU는 메타데이터 해시를 계산하고 데이터,
메타데이터 및 해시가 SSD 어레이의 캐시 메모리에 기입되게 한다
306
DAS 제어기의 CPU는 데이터가 PD들에서의
물리적 주소들에 기입되게 한다
종료