관 인 생 략

출 원 번 호 통 지 서

출 원 일 자	2016.02.23
특 기 사 항	심사청구(무) 공개신청(무)
출 원 번 호	10-2016-0021180 (접수번호 1-1-2016-0177348-23)
출원인 명칭	에스케이텔레콤 주식회사(1-1998-004296-6) 외 1명
대 리 인 성 명	이철희(9-1998-000480-5)
발 명 자 성 명	박찬익 성백재
발 명 의 명 칭	빠른 데이터 복원을 위한 데이터 읽기 또는 쓰기 방법

특 허 청 장

<< 안내 >>

1. 귀하의 출원은 위와 같이 정상적으로 접수되었으며, 이후의 심사 진행상황은 출원번호를 통해 확인하실 수 있습니다.

2. 출원에 따른 수수료는 접수일로부터 다음날까지 동봉된 납입영수증에 성명, 납부자번호 등을 기재하여 가까운 우체국 또는 은행에 납부하여야 합니다.
※ 납부자번호 : 0131(기관코드) + 접수번호

3. 귀하의 주소, 연락처 등의 변경사항이 있을 경우, 즉시 [출원인코드 정보변경(경정), 정정신고서]를 제출하여야 출원 이후의 각종 통지서를 정상적으로 받을 수 있습니다.
※ 특허로(patent.go.kr) 접속 > 민원서식다운로드 > 특허법 시행규칙 별지 제5호 서식

4. 특허(실용신안등록)출원은 명세서 또는 도면의 보정이 필요한 경우, 등록결정 이전 또는 의견서 제출기간 이내에 출원서에 최초로 첨부된 명세서 또는 도면에 기재된 사항의 범위 안에서 보정할 수 있습니다.

5. 외국으로 출원하고자 하는 경우 PCT 제도(특허·실용신안)나 마드리드 제도(상표)를 이용할 수 있습니다. 국내출원일을 외국에서 인정받고자 하는 경우에는 국내출원일로부터 일정한 기간 내에 외국에 출원하여야 우선권을 인정받을 수 있습니다.
※ 제도 안내 : http://www.kipo.go.kr-특허마당-PCT/마드리드
※ 우선권 인정기간 : 특허·실용신안은 12개월, 상표·디자인은 6개월 이내
※ 미국특허상표청의 선출원을 기초로 우리나라에 우선권주장출원 시, 선출원이 미공개상태이면, 우선일로부터 16개월 이내에 미국특허상표청에 [전자적교환허가서(PTO/SB/39)를 제출하거나 우리나라에 우선권 증명서류를 제출하여야 합니다.

6. 본 출원사실을 외부에 표시하고자 하는 경우에는 아래와 같이 하여야 하며, 이를 위반할 경우 관련법령에 따라 처벌을 받을 수 있습니다.
※ 특허출원 10-2010-0000000, 상표등록출원 40-2010-0000000

7. 종업원이 직무수행과정에서 개발한 발명을 사용자(기업)가 명확하게 승계하지 않은 경우, 특허법 제62조에 따라 심사단계에서 특허거절결정되거나 특허법 제133조에 따라 등록이후에 특허무효사유가 될 수 있습니다.

8.기타 심사 절차에 관한 사항은 동봉된 안내서를 참조하시기 바랍니다.

2016-02-23

【서지사항】

【서류명】 특허출원서

【출원구분】 특허출원

【출원인】

【명칭】 에스케이 텔레콤주식회사

【출원인코드】 1-1998-004296-6

【출원인】

【명칭】 포항공과대학교 산학협력단

【출원인코드】 2-2004-043336-1

【대리인】

【성명】 이철희

【대리인코드】 9-1998-000480-5

【포괄위임등록번호】 2000-010209-0

【발명의 국문명칭】 빠른 데이터 복원을 위한 데이터 읽기 또는 쓰기 방법

【발명의 영문명칭】 Methods for Providing Data Read or Write for Fast Data Reconstruction

【발명자】

【성명】 박찬익

【성명의 영문표기】 PARK, Chan Ik

【주민등록번호】 610301-1XXXXXX

【우편번호】 37673

【주소】 경상북도 포항시 남구 청암로 77 포항공과대학교 정보통신연구소

【국적】 KR

【발명자】

【성명】 성백재

【성명의 영문표기】 SUNG, Baeg Jae

【주민등록번호】 800819-1XXXXXX

【우편번호】 37673

【주소】 경상북도 포항시 남구 청암로 77 포항공과대학교 정보통신 연구소

【국적】 KR

【출원언어】 국어

【취지】 위와 같이 특허청장에게 제출합니다.

대리인 이철희 (서명 또는 인)

【수수료】

【출원료】	0 면	46,000 원
【가산출원료】	37 면	0 원
【우선권주장료】	0 건	0 원
【심사청구료】	0 항	0 원
【합계】	46,000원	

【발명의 설명】

【발명의 명칭】

빠른 데이터 복원을 위한 데이터 읽기 또는 쓰기 방법{Methods for Providing Data Read or Write for Fast Data Reconstruction}

【기술분야】

【0001】 본 실시예는 빠른 데이터 복원을 위한 데이터 읽기 또는 쓰기 방법에 관한 것이다.

【발명의 배경이 되는 기술】

【0002】 이하에 기술되는 내용은 단순히 본 실시예와 관련되는 배경 정보만을 제공할 뿐 종래기술을 구성하는 것이 아니다.

【0003】 일반적으로 널리 활용되고 있는 RAID(Redundant Array of Independent Disks 또는 Redundant Array of Inexpensive Disks) 6의 기술은 디스크에 오류 발생 시 데이터 복원을 위하여 다른 모든 디스크로 데이터 읽기(Read)를 수행한다. 하지만, RAID 6의 기술은 데이터 복원을 위한 동작으로 인해 데이터 읽기 성능이 절반 수준으로 저하되는 문제가 있다.

【0004】 RAID 6에서 데이터 복원 시 데이터 읽기 성능 저하를 개선하기 위한 기술로는 LRC(Local Reconstruction Code) 기술이 있다. LRC 방식은 분산 스토리지를 위한 기술로서, 로컬 패리티(Local Parity)(P 패리티), 글로벌 패리티(Global Parity)(Q 패리티)를 이용하여 데이터 복원 시 로컬 그룹으로 구분된 디스크의 데

이터 읽기를 수행하여 빠른 데이터 복원을 지원하는 기술이다.

【0005】RAID 6와 LRC 방식의 패리티 배치 정보(데이터 레이아웃)는 도 1a에 도시된 바와 같다. 도 1a에 도시된 RAID 6의 6개 'D 데이터'들은 'P 패리티(로컬 패리티)'와 'Q 패리티(글로벌 패리티)'로 보호되고 있다. 다시 말해, RAID에서의 '패리티'는 데이터 복원용 임시 정보를 의미한다. '패리티'는 데이터를 각 개별 디스크로부터 읽을 수 있으며, 동시에 복수의 데이터 읽기 또는 쓰기가 가능하다. RAID 6의 경우 RAID 5의 확장 컨셉으로 2개의 패리티(P 패리티, Q 패리티)가 쓰인다.

【0006】반면, LRC 방식은 도 1a에 도시된 바와 같이, 'P 패리티(로컬 패리티)'를 2개의 그룹으로 나누었다. 즉, 3개의 'D1 데이터'들은 'P1 패리티(로컬 패리티)'로, 나머지 3개의 'D2 데이터'들은 'P2 패리티(로컬 패리티)'로 보호되고 있다. 6개의 'D1, D2 데이터'들은 'Q 패리티(글로벌 패리티)'로 보호되고 있다. 'P 패리티(로컬 패리티)'와 'Q 패리티(글로벌 패리티)' 개수는 데이터 복원 성능, 데이터 신뢰도에 따라 1개 이상으로 설정될 수 있다.

【0007】도 1b에 도시된 바와 같이, 첫 번째 디스크에 오류가 발생하면, 해당 디스크에 데이터 읽기 요청 시 해당 데이터를 복원하기 위해 발생되는 데이터 읽기 수행 수를 나타낸다. RAID 6는 총 6개의 데이터를 읽어들인다. 반면, LRC 방식은 1그룹의 3개 데이터만 읽어들여서 복원을 수행한다.

【0008】하지만, 모든 디스크의 I/O 성능이 동일한 RAID 어레이는, 빠른 I/O 성능을 위하여 데이터를 모든 데이터 디스크에 균등하게 분배한다. 다시 말해, 모든 데이터 디스크는 평균적으로 동일한 수의 I/O 요청을 받게 되며, 이 경우 위의 특징은 빠른 데이터 복원에 도움을 주지 못한다.

【0009】예컨대, 도 1c에서 RAID 6는 데이터 디스크에 발생한 오류로 인하여, 모든 데이터 디스크는 정상 대비 2배의 데이터 읽기 수행이 발생한다. 데이터 디스크에 발생한 오류로 인해, 총괄 RAID의 데이터 읽기 성능은 절반 수준으로 저하된다. 이해를 돕기 위해 랜덤 읽기(Random Read)를 가정한다. 램덤 읽기에서 데이터 읽기 요청의 병합(Merge)이 발생하지 않는다고 가정한다.

【0010】LRC 방식은 데이터 디스크에 발생한 오류로 인하여, 그룹 1의 2개 데이터 디스크만 정상 대비 2배의 데이터 읽기 수행이 발생한다. LRC 방식에서 그룹 2의 데이터 디스크는 정상인 상황과 동일하다. 하지만, 그룹 1의 2개 데이터 디스크에 병목현상(Bottleneck)이 발생되어, 그룹 2의 3개 데이터 디스크의 활용성(Utilization)은 절반 수준으로 낮아진다. LRC 방식의 총괄 RAID의 데이터 읽기 성능은 절반 수준으로 저하된다. 결과적으로 LRC 방식 역시 RAID 6와 같이 병목현상이 발생하는 문제가 있다.

【0011】종래의 스토리지 장치들은 데이터 손실을 막기 위하여 데이터 보호 기법을 사용한다. 일반적인 스토리지 장치들이 사용하는 추가 패리티 디스크를 활용하는 RAID 기술은 널리 활용되는 데이터 보호 기법이다. 패리티 개수, 패리티 배치 정보(데이터 레이아웃) 등에 따라 복수 개의 RAID 기술로 구분되며 각 다른 특

징을 가진다.

【0012】일반적으로 활용되고 있는 'RAID 6'는 디스크에 오류가 발생하면 절반 수준으로 데이터 읽기 성능이 저하된다. 'RAID 6'에서 데이터 읽기 성능 개선을 위한 LRC 방식 역시 모든 디스크의 I/O 성능이 동일한 RAID 어레이에서 데이터 복원 성능이 개선되지 않는다. 다시 말해, LRC 방식은 고정 로컬 그룹 결정 방법으로 인하여, 디스크에 오류 발생 시 데이터 복원을 위하여 일부 디스크만 활용함으로써 병목현상이 발생하여 데이터 복원 성능이 개선되지 않는 문제가 있다.

【발명의 내용】

【해결하고자 하는 과제】

【0013】본 실시예는 RAID 어레이(Array) 상에서 데이터 읽기 또는 쓰기를 수행할 때, 디스크에 오류 발생 시 나타나는 급격한 데이터 입출력 성능 저하(Bottleneck)를 방지하기 위하여, 동일한 로컬 그룹의 데이터들이 서로 다른 복수 개의 디스크로 분산되어 있는 분산 구조 상에서 빠른 데이터 복원으로 데이터 읽기 또는 쓰기가 수행되도록 하는 빠른 데이터 복원을 위한 데이터 읽기 또는 쓰기 방법을 제공하는 데 목적이 있다.

【과제의 해결 수단】

【0014】본 실시예의 일 측면에 의하면, RAID 어레이(Array) 상에서 데이터 읽기(Read) 방법에 있어서, 애플리케이션(Application)으로부터 적어도 하나의 데이터 위치(Position)를 기준으로 한 적어도 하나의 데이터 읽기 요청을 수신하는

수신 과정; 디스크 분산 구조 상에서 상기 적어도 하나의 데이터 읽기 요청을 처리하기 위한 디스크 번호 및 상기 디스크 번호(Disk Number)에 해당하는 디스크 내의 스트라이프 번호(Stripe Number)를 확인하는 확인 과정; 상기 디스크 번호 중 상기 적어도 하나의 데이터 읽기 요청에서 하나의 데이터 읽기를 수행하고자 하는 특정 디스크에 오류 발생 여부를 확인하는 오류 확인 과정; 상기 특정 디스크에 오류가 발생한 경우, 상기 특정 디스크 내의 상기 하나의 데이터 읽기와 관련된 스트라이프에 대한 데이터가 저장되면서도 상기 특정 디스크와는 다른 복수의 디스크로부터 패리티 복원 관련 데이터를 읽어들이는 읽기 과정; 및 상기 페리티 복원 관련 데이터를 이용하여 상기 스트라이프 내의 읽기를 수행하고자 하는 데이터를 복원하여 읽어들이는 복원 과정을 포함하는 것을 특징으로 하는 빠른 데이터 복원을 위한 데이터 읽기 방법을 제공한다.

【0015】본 실시예의 다른 측면에 의하면, RAID 어레이 상에서 데이터 쓰기(Write) 방법에 있어서, 애플리케이션으로부터 적어도 하나의 데이터 위치를 기준으로 한 적어도 하나의 데이터 쓰기 요청을 수신하는 수신 과정; 디스크 분산 구조 상에서 상기 적어도 하나의 데이터 쓰기 요청을 처리하기 위한 디스크 번호 및 상기 디스크 번호에 해당하는 디스크 내의 스트라이프 번호를 확인하는 확인 과정; 상기 디스크 번호 중 상기 적어도 하나의 데이터 쓰기 요청에서 하나의 데이터 쓰기를 수행하고자 하는 특정 디스크에 오류 발생 여부를 확인하는 오류 확인 과정; 및 상기 특정 디스크에 오류가 발생한 경우, 상기 특정 디스크 내의 상기 하나의 데이터 쓰기와 관련된 스트라이프에 데이터 쓰기가 미수행되도록 하는 쓰기 과정을

포함하는 것을 특징으로 하는 빠른 데이터 복원을 위한 데이터 쓰기 방법을 제공한다.

【발명의 효과】

【0016】 이상에서 설명한 바와 같이 본 실시예에 의하면, RAID 어레이 상에서 데이터 읽기 또는 쓰기를 수행할 때, 디스크에 오류 발생 시 나타나는 급격한 데이터 입출력 성능 저하(Bottleneck)를 방지하기 위하여, 동일한 로컬 그룹의 데이터들이 서로 다른 복수 개의 디스크로 분산되어 있는 분산 구조 상에서 빠른 데이터 복원으로 데이터 읽기 또는 쓰기가 수행되도록 하는 효과가 있다.

【0017】 본 실시예에 의하면, 디스크에 오류 발생 시 데이터 복원을 위하여 디스크 분산 구조를 이용함으로써 병목현상을 없애고, 데이터 복원 성능을 높일 수 있는 효과가 있다.

【0018】 본 실시예에 의하면, RAID 어레이에서는 저장장치(예컨대, 하드 디스크(HDD: Hard Disk Drive), 솔리드 스테이트 디스크(SSD: Solid State Drive) 등) 고장에 대해 데이터 복원을 지원함으로써 어레이 신뢰도를 높이는 목적으로 패리티 정보를 활용하는데 있어서, 패리티 정보를 어레이 신뢰도 향상뿐 아니라 데이터 복원 성능을 향상시킬 수 있는 효과가 있다.

【0019】 본 실시예에 의하면, RAID 어레이에서 신뢰도와 데이터 복원 성능 간의 트레이드 오프(Trade-Off)를 제어할 수 있게 됨에 따라 RAID 어레이 관리에 높은 유연성을 확보할 수 있다. 특히, 올 플래시 어레이(AFA)는 솔리드 스테이트

디스크(SSD)의 성능이 빨라짐에 따라 데이터 복원 성능 개선이 가능한 효과가 있다.

【0020】 본 실시예에 의하면, 복수 개의 디스크가 복수 개의 로컬 그룹으로 구분되며, 동일한 로컬 그룹의 데이터들이 서로 다른 복수 개의 디스크로 분산되어 있는 분산 구조를 이용하여 빠른 데이터 복원이 가능하다. 또한, 본 실시예에 의하면, 어떠한 종류의 올 플래시 어레이(AFA) 제품에서도 적용 가능하며, RAID에 독립적으로 구현 가능한 효과가 있다.

【도면의 간단한 설명】

【0021】 도 1a 내지 1c는 종래의 RAID 기술을 설명하기 위한 도면이다.

도 2는 본 실시예에 따른 빠른 데이터 복원을 위한 스토리지 장치를 개략적으로 나타낸 블럭 구성도이다.

도 3은 본 실시예에 따른 디스크 분산 구조를 나타내기 위한 도면이다.

도 4는 본 실시예에 따른 디스크 분산 구조에서 오류 발생 시 빠른 데이터 복원을 설명하기 위한 도면이다.

도 5는 본 실시예에 따른 디스크 분산 구조의 또 다른 실시예를 나타낸 도면이다.

도 6은 본 실시예에 따른 빠른 데이터 복원을 위한 데이터 읽기 방법을 설명하기 위한 순서도이다.

도 7은 본 실시예에 따른 빠른 데이터 복원을 위한 데이터 쓰기 방법을 설명

하기 위한 순서도이다.

【발명을 실시하기 위한 구체적인 내용】

【0022】 이하, 본 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.

【0023】 도 2는 본 실시예에 따른 빠른 데이터 복원을 위한 스토리지 장치를 개략적으로 나타낸 블럭 구성도이다.

【0024】 본 실시예에 따른 스토리지 장치(200)는 애플리케이션(Application)(210), RAID 제어기(RAID Controller)(220) 및 스토리지(230)를 포함한다. 스토리지 장치(200)에 포함된 구성요소는 반드시 이에 한정되는 것은 아니다.

【0025】 스토리지 장치(200)에 포함된 각 구성요소는 장치 내부의 소프트웨어적인 모듈 또는 하드웨어적인 모듈을 연결하는 통신 경로에 연결되어 상호 간에 유기적으로 동작할 수 있다. 이러한 구성요소는 하나 이상의 통신 버스 또는 신호선을 이용하여 통신한다.

【0026】 도 2에 도시된 스토리지 장치(200)의 각 구성요소는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 소프트웨어적인 모듈, 하드웨어적인 모듈 또는 소프트웨어와 하드웨어의 결합으로 구현될 수 있다.

【0027】 본 실시예에 따른 스토리지 장치(200)는 RAID를 기반으로 동작하는 저장장치를 의미한다. RAID는 복수 배열 독립 디스크로서, 복수 개의 디스크(하드 디스크(HDD) 또는 솔리드 스테이트 디스크(SSD)에 일부 중복된 데이터를 나눠서 저

장하는 기술이다. RAID에서 데이터를 나누는 다양한 방법이 존재하며, 데이터를 나누는 방법들을 ‘레벨’이라 칭하는데, 레벨에 따라 저장장치의 신뢰성을 높이거나 전체적인 성능을 향상시킨다. RAID는 복수 개의 디스크를 하나로 묶어 하나의 논리적 디스크로 작동하게 하는데, 하드웨어적인 방법과 소프트웨어적인 방법이 있다. 하드웨어적인 방법은 운영 체제에 복수 개의 디스크가 하나의 디스크처럼 보이게 한다. 소프트웨어적인 방법은 주로 운영체제 안에서 구현되며, 사용자에게 디스크를 하나의 디스크처럼 보이게 한다.

【0028】본 실시예에 따른 스토리지 장치(200)는 기본적으로 ‘RAID 6’를 동작하나 반드시 이에 한정되는 것은 아니며, RAID 기술의 범위 안에서 다른 RAID 레벨도 적용 가능하다. ‘RAID 6’는 패리티(오류 검출 기능)가 배분(Distributed)되는 스트리핑된 세트(적어도 4 개의 디스크)를 포함하는 방식이다.

【0029】본 실시예에 따른 스토리지 장치(200)는 올 플래시 어레이(AFA)와 같은 RAID 어레이 상황에서, 디스크에 오류가 발생하더라도 디스크 분산 구조를 이용하여 빠른 데이터 복원이 가능하다. 본 실시예에 따른 스토리지 장치(200)는 디스크에 오류 발생 시 데이터 복원을 위하여 디스크 분산 구조를 이용함으로써 병목 현상을 없애고, 데이터 복원 성능을 높일 수 있다.

【0030】본 실시예에 따른 스토리지 장치()는 올 플래시 어레이(AFA: All Flash Array)에도 적용 가능하다.

【0031】올 플래시 어레이(AFA)는 복수 개의 플래시 기반 디스크(SSD)를 어레이(Array) 형태로 결합하여 빠른 I/O(Input/Output) 성능을 지원한다. 올 플래시 어레이(AFA)는 빠른 I/O 성능을 바탕으로 프라이머리 스토리지(Primary Storage)로 활용되는 기술이다. 올 플래시 어레이(AFA)는 주요 스토리지 벤더(Storage Vendor)들을 필두로 활발히 개발되고 있으며, 독자적인 RAID(Redundant Array of Inexpensive Disks) 기술을 활용하여 데이터 보호를 지원하고 있다.

【0032】일반적인 RAID 기술들은 공간 오버헤드, 데이터 신뢰도 간 트레이드오프에 집중하였다. 반면, 올 플래시 어레이(AFA)는 빠른 I/O 성능에 초점을 맞춘 스토리지이다. 올 플래시 어레이(AFA)는 디스크 오류가 발생한 상황에서 나타나는 급격한 데이터 입출력 성능 저하를 막기 위해 빠른 데이터 복원을 위한 RAID 기술에 집중하고 있다.

【0033】애플리케이션(210)은 RAID 제어기(220)로 데이터 입출력을 요청한다. 애플리케이션(210)은 외부 단말기의 운영 체제에 설치될 수 있다. 외부 단말기는 사용자의 키 조작에 따라 네트워크를 경유하여 데이터 통신을 수행하는 전자 기기를 의미한다. 다시 말해, 애플리케이션(210)은 외부 단말기에 설치되어 RAID 제어기(220)로 데이터 읽기 또는 쓰기를 요청한다. 애플리케이션(210)에서 요청한 데이터 입출력은 RAID 제어기(220)에 의해 중재된다. 애플리케이션(210)은 적어도 하나의 데이터 위치(Position)를 기준으로 한 적어도 하나의 데이터 읽기 요청을 RAID 제어기(220)로 전송한다. 애플리케이션(210)은 적어도 하나의 데이터 위치를 기준으로 한 적어도 하나의 데이터 쓰기 요청을 RAID 제어기(220)로

전송한다.

【0034】외부 단말기는 네트워크를 경유하여 RAID 제어기(220)와 통신하기 위한 프로그램 또는 프로토콜을 저장하기 위한 메모리, 해당 프로그램을 실행하여 연산 및 제어하기 위한 마이크로프로세서 등을 구비한다. 외부 단말기는 개인용 컴퓨터(PC: Personal Computer), 랩톱(Laptop)이 바람직하나 반드시 이에 한정되는 것은 아니며, 스마트폰(Smart Phone), 태블릿(Tablet), 개인 휴대 단말기(PDA: Personal Digital Assistant), 게임 콘솔, 휴대형 멀티미디어 플레이어(PMP: Portable Multimedia Player), 무선 통신 단말기(Wireless Communication Terminal), TV, 미디어 플레이어 등과 같은 전자기기일 수 있다.

【0035】외부 단말기는 (i) 각종 기기 또는 유무선 네트워크와 통신을 수행하기 위한 통신 모뎀 등의 통신 장치, (ii) 각종 프로그램과 데이터를 저장하기 위한 메모리, (iii) 프로그램을 실행하여 연산 및 제어하기 위한 마이크로프로세서 등을 구비하는 다양한 장치이다. 적어도 일 실시예에 따르면, 메모리는 램(RAM: Random Access Memory), 롬(ROM: Read Only Memory), 플래시 메모리, 광 디스크, 자기 디스크, 솔리드 스테이트 디스크(SSD: Solid State Disk) 등의 컴퓨터로 판독 가능한 기록/저장매체일 수 있다. 적어도 일 실시예에 따르면, 마이크로프로세서는 명세서상에 기재된 동작과 기능을 하나 이상 선택적으로 수행하도록 프로그램될 수 있다. 적어도 일 실시예에 따르면, 마이크로프로세서는 전체 또는 부분적으로 특정한 구성의 주문형반도체(ASIC: Application Specific Integrated Circuit) 등의 하드웨어로써 구현될 수 있다.

【0036】메모리에 관련 데이터 및 프로그램이 저장되어 있고, 프로세서가 메모리로부터 관련 데이터를 읽어들여 처리한다. 프로세서는 하나의 프로세서가 위각 기능들을 수행할 수 있지만, 복수 개의 프로세서가 분담하여 처리하도록 구현할 수도 있다. 프로세서는 범용 프로세서에서 구현될 수도 있지만, 그 기능을 수행하도록 별도로 제작된 칩으로 구현할 수도 있다.

【0037】RAID 제어기(220)는 기 정의된 패리티 배치 정보(데이터 레이아웃)에 따라 해당 데이터 입출력을 복수 개의 디스크를 활용하여 처리한다. RAID 제어기(220)는 패리티 배치 정보(데이터 레이아웃)에 따라 패리티를 계산하여 저장하고, 디스크에 오류 발생 시 패리티를 활용하여 데이터를 복원하는 기능을 수행한다.

【0038】이하, 본 실시예에 따른 RAID 제어기(220)가 RAID 어레이(Array) 상에서 데이터 읽기를 수행하는 절차에 대해 설명한다.

【0039】RAID 제어기(220)는 애플리케이션(210)으로부터 적어도 하나의 데이터 위치(Position)를 기준으로 한 적어도 하나의 데이터 읽기 요청을 수신한다. RAID 제어기(220)는 데이터 읽기 요청에 포함된 논리적 블럭 주소(Logical Block Address) 또는 오프셋(Offset)을 이용하여 디스크 분산 구조 상에서 데이터 위치에 따른 패리티 배치 정보(데이터 레이아웃)를 확인한다.

【0040】RAID 제어기(220)는 디스크 분산 구조 상에서 적어도 하나의 데이터 읽기 요청을 처리하기 위한 디스크 번호(예컨대, #1, #2, #3... #9)를 확인한다.

RAID 제어기(220)는 디스크 번호(예컨대, #1, #2, #3... #9)(Disk Number)에 해당하는 디스크 내의 스트라이프 번호(예컨대, #0, #1, #2, #3)(Stripe Number)를 확인한다. RAID 제어기(220)는 디스크 번호(예컨대, #1, #2, #3... #9) 중 적어도 하나의 데이터 읽기 요청에서 하나의 데이터 읽기를 수행하고자 하는 특정 디스크(예컨대, #1)에 오류 발생 여부를 확인한다.

【0041】RAID 제어기(220)는 특정 디스크(예컨대, #1)에 오류가 발생한 경우, 특정 디스크(예컨대, #1) 내의 하나의 데이터 읽기와 관련된 스트라이프에 대한 데이터가 저장되면서도 특정 디스크와는 다른 복수의 디스크로부터 패리티 복원 관련 데이터(서로 다른 복수의 디스크 상에서의 동일한 로컬 그룹들의 데이터 및 로컬 패리티(P 패리티))를 읽어들인다.

【0042】다시 말해, RAID 제어기(220)는 특정 디스크(예컨대, #1)에 오류가 발생한 경우, 디스크 분산 구조 상에서 데이터 위치에 따른 패리티 배치 정보(Data Layout)를 확인한다. RAID 제어기(220)는 패리티 배치 정보(데이터 레이아웃)를 참조하여 특정 디스크(예컨대, #1) 내의 하나의 데이터 읽기와 관련된 스트라이프 별로 특정 디스크와 동일한 로컬 그룹들을 확인한다. RAID 제어기(220)는 서로 다른 복수의 디스크 상에서의 동일한 로컬 그룹들의 데이터 및 로컬 패리티를 패리티 복원 관련 데이터로 읽어들인다.

【0043】RAID 제어기(220)는 페리티 복원 관련 데이터를 이용하여 하나의 데이터 읽기와 관련된 스트라이프 내의 읽기를 수행하고자 하는 데이터를 복원한다. 다시 말해, RAID 제어기(220)의 복원 과정에 대해 설명한다. RAID 제어기(220)는

동일한 로컬 그룹의 데이터 및 로컬 패리티를 이용하여 하나의 데이터 읽기와 관련된 스트라이프 내의 읽기를 수행하고자 하는 데이터를 복원한다. RAID 제어기(220)는 스트라이프 내의 읽기를 수행하고자 하는 데이터를 복원하여 데이터 읽기 요청에 대한 응답으로 리턴(Return)한다.

【0044】 이하, RAID 제어기(220)가 RAID 어레이 상에서 데이터 쓰기(Write)를 수행하는 절차에 대해 설명한다.

【0045】 RAID 제어기(220)는 애플리케이션(210)으로부터 적어도 하나의 데이터 위치를 기준으로 한 적어도 하나의 데이터 쓰기 요청을 수신한다. RAID 제어기(220)는 디스크 분산 구조 상에서 적어도 하나의 데이터 쓰기 요청을 처리하기 위한 디스크 번호(예컨대, #1, #2, #3... #9)를 확인한다. RAID 제어기(220)는 확인된 디스크 번호(예컨대, #1, #2, #3... #9)에 해당하는 디스크 내의 스트라이프 번호(예컨대, #0, #1, #2, #3)를 확인한다.

【0046】 RAID 제어기(220)는 디스크 번호에 해당하는 디스크 내의 동일한 로컬 그룹 각각에 대한 로컬 패리티를 각각 확인하고, 디스크 번호에 해당하는 디스크 내의 각 스트라이프에 대한 글로벌 패리티를 각각 확인한다. RAID 제어기(220)는 디스크 번호(예컨대, #1, #2, #3... #9)에 해당하는 디스크 내의 하나의 데이터 쓰기와 관련된 스트라이프에 대한 데이터와 하나의 데이터 쓰기를 수행하고자 하는 로컬 패리티와 글로벌 패리티를 확인한다.

【0047】RAID 제어기(220)는 디스크 번호(예컨대, #1, #2, #3... #9) 중 적어도 하나의 데이터 쓰기 요청에서 하나의 데이터 쓰기를 수행하고자 하는 특정 디스크(예컨대, #1)에 오류 발생 여부를 확인한다. 다시 말해, RAID 제어기(220)는 하나의 데이터 쓰기와 관련된 스트라이프에 대한 데이터, 하나의 데이터 쓰기를 수행하고자 하는 로컬 패리티 또는 글로벌 패리티를 포함한 특정 디스크(예컨대, #1)에 오류 발생 여부를 확인한다.

【0048】RAID 제어기(220)는 특정 디스크(예컨대, #1)에 오류가 발생한 경우, 특정 디스크(예컨대, #1) 내의 하나의 데이터 쓰기와 관련된 스트라이프에 데이터 쓰기가 미수행되도록 한다.

【0049】본 실시예에 따른 디스크 분산 구조에 대해 설명한다. 디스크 분산 구조는 DLRC(Distributed Local Reconstruction Code) 방식을 의미한다. 디스크 분산 구조는 복수 개의 디스크(하드 디스크(HDD) 또는 솔리드 스테이트 디스크(SSD))가 복수 개의 로컬 그룹으로 구분된다. 디스크 분산 구조는 로컬 그룹마다 로컬 패리티가 각각 존재한다. 디스크 분산 구조는 복수 개의 디스크 내의 각 스트라이프 별로 글로벌 패리티가 각각 존재한다. 디스크 분산 구조는 동일한 로컬 그룹의 데이터들이 서로 다른 복수 개의 디스크로 분산되어 있는 분산 구조를 갖는다.

【0050】디스크 분산 구조 내의 로컬 패리티(P 패리티)와 글로벌 패리티(Q 패리티)는 서로 다른 복수 개의 디스크로 분산 저장되는 분산 구조를 갖는다. 디스크 분산 구조 내의 로컬 패리티(P 패리티)와 글로벌 패리티(Q 패리티)는 각각의 디스크에 별도로 저장될 수 있다. 디스크 분산 구조 내의 로컬 패리티(P 패리티)와

글로벌 패리티(Q 패리티) 중 어느 하나의 패리티만이 서로 다른 복수 개의 디스크로 분산되어 있는 분산 구조를 가지며, 나머지 패리티는 하나의 디스크에 별도로 저장될 수 있다.

【0051】 디스크 분산 구조는 동일한 로컬 그룹의 데이터들이 서로 다른 복수 개의 디스크로 분산 저장되어, 하나의 데이터 읽기와 관련된 스트라이프 내의 읽기를 수행하고자 하는 데이터를 복원할 때, 개별 디스크마다 분산 저장된 로컬 그룹에 따라 데이터 읽기 수행 수를 감소(분산)되어 병목현상이 완화되도록 한다.

【0052】 스토리지(230)는 복수 개의 물리적인 디스크를 의미한다. 스토리지(230)는 복수 개의 하드 디스크(HDD) 또는 복수 개의 솔리드 스테이트 디스크(SSD) 또는 하드 디스크(HDD)와 솔리드 스테이트 디스크(SSD)의 조합으로 구현될 수 있다. 스토리지(230)는 RAID 제어기(220)의 제어에 따라 데이터를 저장한다.

【0053】 도 3은 본 실시예에 따른 디스크 분산 구조를 나타내기 위한 도면이다.

【0054】 일반적인 LRC 방식은 모든 디스크의 I/O 성능이 동일한 RAID 어레이에서 데이터 복원 성능이 개선되지 않는다. 다시 말해, 고정 로컬 그룹 결정 방법으로 인하여, LRC 방식은 디스크에 오류 발생 시 데이터 복원을 위하여 일부 디스크만 활용함으로써 병목현상이 발생되어 데이터 복원 성능이 개선되지 않는다.

【0055】 LRC 방식에 비해 본 실시예의 디스크 분산 구조는 분산 로컬 그룹 결정 방법을 사용하여, 디스크에 오류 발생 시 데이터 복원을 위하여 전체 디스크

를 활용함으로써 병목현상을 완화/상쇄하고, 데이터 복원 성능을 높인다.

【0056】 도 3은 본 실시예에 따른 DLRC 방식과 LRC 방식의 패리티 배치 정보(데이터 레이아웃)의 차이를 나타낸다. DLRC 방식과 LRC 방식의 차이점으로는 LRC 방식의 경우 한 개의 디스크 내의 모든 스트라이프(Stripe)(한 데이터 유닛)가 한 개의 로컬 그룹에 고정되어 있다. 반면, DLRC 방식의 경우 한 개의 디스크 내의 스트라이프들은 각 로컬 그룹으로 분산되어 있는 점이다. 전술한 차이점으로 인하여, DLRC 방식이 디스크에 오류 발생 시 데이터 복원 측면에서 이득을 가진다.

【0057】 본 실시예에 따른 디스크 분산 구조는 도 3에 도시된 바와 같이 DLRC 방식으로서, 복수 개의 디스크(하드 디스크(HDD) 또는 솔리드 스테이트 디스크(SSD))가 복수 개의 로컬 그룹으로 구분된다. 디스크 분산 구조는 로컬 그룹마다 로컬 패리티가 각각 존재한다. 디스크 분산 구조는 복수 개의 디스크 내의 각 스트라이프 별로 글로벌 패리티가 각각 존재한다. 디스크 분산 구조는 동일한 로컬 그룹의 데이터들이 서로 다른 복수 개의 디스크로 분산되어 있는 분산 구조를 갖는다. 디스크 분산 구조 내의 로컬 패리티(P 패리티)와 글로벌 패리티(Q 패리티)는, 도 3에 도시된 바와 같이 각각의 디스크에 별도로 저장될 수 있다.

【0058】 도 4는 본 실시예에 따른 디스크 분산 구조에서 오류 발생 시 빠른 데이터 복원을 설명하기 위한 도면이다.

【0059】 도 4에 도시된 바와 같이, LRC 방식과 DLRC 방식의 패리티 배치 정보(데이터 레이아웃)를 기준으로, 첫 번째 디스크에 오류 발생 시 데이터 복원을 위하여 다른 디스크들의 데이터 읽기 수행 수를 보인다. 이해를 돕기 위해 랜덤 읽

기(Random Read)로 가정한다. 랜덤 읽기의 경우, 읽기 요청 병합(Merge)은 발생하지 않는다고 가정한다. 평균적으로 디스크에 오류 발생으로 인해 요청되는 데이터 읽기 요청은 '스트라이프 번호 모드 4(Stripe Number Mod 4)'의 경우 '#0, #1, #2, #3'으로 균등하게 발생하며, 평균적으로 '4 개'의 데이터 복원을 위해 서로 다른 디스크들의 데이터 읽기 수행 수는 도 4에 도시된 바와 같다.

【0060】LRC 방식의 경우, 동일한 로컬 그룹의 3개 디스크로 디스크 읽기 수행이 집중되며, 이로 인하여 '2 개'의 데이터 디스크가 병목현상이 발생되어 데이터 복원 성능이 저하된다. 반면, DLRC 방식의 경우 동일한 로컬 그룹이 다른 '6 개' 디스크로 분산되어 있어, 상대적으로 LRC 방식과 대비하여 개별 디스크에 절반 수준의 데이터 읽기 수행이 발생한다. 이로 인하여 DLRC 방식의 경우 병목현상이 완화/상쇄되어 데이터 복원을 빠르게 수행한다.

【0061】예컨대, 데이터에 오류 발생 시 모든 디스크로 데이터 읽기 요청이 각각 '4 번'씩 발생하였다면, LRC 방식의 경우 두 번째, 세 번째 디스크가 각 '8 번'의 데이터 읽기 수행이 완료되어야 전체 데이터 읽기 요청이 완료되지만, DLRC 방식은 두 번째, 네 번째, 다섯 번째, 여섯 번째 디스크가 각 '6 번'의 데이터 읽기 수행만으로 전체 데이터 읽기 요청을 완료할 수 있다.

【0062】본 실시예의 DLRC 방식은 한 개의 디스크 내의 스트라이프들이 각 로컬 그룹으로 분산되어 있다는 주요한 특징을 가진다. 도 3은 이러한 특징을 보이기 위한 하나의 실시예이며, 실제 발명의 구현에 있어서, 디스크 개수, 로컬 패리티(P 패리티), 글로벌 패리티(Q 패리티) 위치 변동 등에 따른 다양한 형태의 패리

티 배치 정보(데이터 레이아웃)로 구현될 수 있다.

【0063】도 5는 본 실시예에 따른 디스크 분산 구조의 또 다른 실시예를 나타낸 도면이다.

【0064】도 5는 DLRC 방식으로 구성 가능한 다양한 형태의 패리티 배치 정보(데이터 레이아웃) 중 두 가지 실시예(성능, 신뢰성)를 나타낸다. DLRC-P 방식은 성능에 초점을 맞춘 패리티 배치 정보(데이터 레이아웃)으로 'P1 패리티(로컬 패리티)', 'P2 패리티(로컬 패리티)', 'Q 패리티(글로벌 패리티)'가 인접하여 있으며, 스트라이프 데이터 증가에 따라 순환하는 형태를 나타낸다. 이러한 경우 모든 패리티가 순환되므로 데이터 쓰기 성능에 저하가 발생하지 않으며, 디스크에 오류 발생 시 다른 디스크에 발생되는 데이터 읽기 요청이 절반 수준으로 균등하게 발생함으로 가장 빠른 데이터 복원이 가능한 형태이다.

【0065】DLRC-R 방식은 데이터 신뢰성에 초점을 맞춘 형태로, 'Q 패리티(글로벌 패리티)'는 한 개의 디스크 내에 고정되고, 'P1 패리티(로컬 패리티)'는 앞 3개 디스크에서, 'P2 패리티(로컬 패리티)'는 뒤 3개의 디스크에서 순환되는 형태이다. 'Q 패리티(글로벌 패리티)'가 한 개의 디스크 내에 고정됨으로 데이터 쓰기 시 병목현상이 발생되어 성능이 저하되는 단점을 가지나, DLRC-P 방식은 두 개의 디스크에 발생한 오류가 복원 가능한 반면, DLRC-R은 디스크에 발생한 오류 위치에 따라 세 개의 디스크까지 복원 가능하다.

【0066】예컨대, 본 실시예에 따른 스토리지 장치(200)는 첫 번째, 네 번째, 다섯 번째 디스크에 오류가 발생하더라도 복원이 가능하다. 또한, 본 실시예에 따른 스토리지 장치(200)는 디스크 개수 증가, 로컬 패리티(P1, P2 패리티) 개수 증가, 글로벌 패리티(Q 패리티) 개수 증가하는 다양한 실시예를 포함한다.

【0067】본 실시예에 따른 디스크 분산 구조는 도 5에 도시된 바와 같이 DLRC 방식으로서, 복수 개의 디스크(하드 디스크(HDD) 또는 솔리드 스테이트 디스크(SSD))가 복수 개의 로컬 그룹으로 구분된다. 디스크 분산 구조는 로컬 그룹마다 로컬 패리티가 각각 존재한다. 디스크 분산 구조는 복수 개의 디스크 내의 각 스트라이프 별로 글로벌 패리티가 각각 존재한다. 디스크 분산 구조는 동일한 로컬 그룹의 데이터들이 서로 다른 복수 개의 디스크로 분산되어 있는 분산 구조를 갖는다.

【0068】디스크 분산 구조가 성능에 초점을 맞춘 경우, 도 5에 도시된 바와 같이, 디스크 분산 구조 내의 로컬 패리티(P 패리티)와 글로벌 패리티(Q 패리티)는 서로 다른 복수 개의 디스크로 분산 저장된다.

【0069】디스크 분산 구조가 데이터 신뢰성에 초점을 맞춘 경우, 도 5에 도시된 바와 같이, 디스크 분산 구조 내의 로컬 패리티(P 패리티)와 글로벌 패리티(Q 패리티) 중 어느 하나의 패리티만이 서로 다른 복수 개의 디스크로 분산되어 있는 분산 구조를 가지며, 나머지 패리티는 하나의 디스크에 별도로 저장될 수 있다.

【0070】도 6은 본 실시예에 따른 빠른 데이터 복원을 위한 데이터 읽기 방법을 설명하기 위한 순서도이다.

【0071】애플리케이션(210)은 위치(또는 논리적 블럭 주소), 오프셋을 기준으로 데이터 읽기 요청을 발생한다(S610).

【0072】RAID 제어기(220)는 애플리케이션(210)으로부터 적어도 하나의 데이터 위치를 기준으로 한 적어도 하나의 데이터 읽기 요청을 수신한다. RAID 제어기(220)는 데이터 읽기 요청에 포함된 논리적 블럭 주소 또는 오프셋(Offset)을 이용하여 디스크 분산 구조상에서 데이터 위치에 따른 패리티 배치 정보(데이터 레이아웃)를 확인한다.

【0073】RAID 제어기(220)는 디스크 분산 구조 상에서 적어도 하나의 데이터 읽기 요청을 처리하기 위한 디스크 번호(예컨대, #1, #2, #3... #9)를 확인한다. RAID 제어기(220)는 디스크 번호(예컨대, #1, #2, #3... #9)에 해당하는 디스크 내의 스트라이프 번호(예컨대, #0, #1, #2, #3)를 확인한다(S620).

【0074】단계 S620에서 RAID 제어기(220)는 디스크 분산 구조 상에서 데이터 읽기 요청에 포함된 데이터 위치에 따른 패리티 배치 정보(데이터 레이아웃)를 확인한다. 디스크 분산 구조는 동일한 로컬 그룹의 데이터들이 서로 다른 복수 개의 디스크로 분산되어 있는 분산 구조를 갖는다.

【0075】RAID 제어기(220)는 디스크 번호(예컨대, #1, #2, #3... #9) 중 적어도 하나의 데이터 읽기 요청에서 하나의 데이터 읽기를 수행하고자 하는 특정 디

스크(예컨대, #1)에 오류 발생 여부를 확인한다(S630).

【0076】 단계 S630에서 특정 디스크(예컨대, #1)에 오류가 발생한 경우, RAID 제어기(220)는 특정 디스크(예컨대, #1) 내의 하나의 데이터 읽기와 관련된 스트라이프에 대한 데이터가 저장되면서도 특정 디스크와는 다른 복수의 디스크로부터 패리티 복원 관련 데이터(서로 다른 복수의 디스크 상에서의 동일한 로컬 그룹들의 데이터 및 로컬 패리티(P 패리티))를 읽어들인다(S640).

【0077】 단계 S640에서, RAID 제어기(220)는 특정 디스크(예컨대, #1)에 오류가 발생한 경우, 디스크 분산 구조 상에서 데이터 위치에 따른 패리티 배치 정보(데이터 레이아웃)를 확인한다. RAID 제어기(220)는 패리티 배치 정보(데이터 레이아웃)를 참조하여 특정 디스크(예컨대, #1) 내의 하나의 데이터 읽기와 관련된 스트라이프 별로 특정 디스크와 동일한 로컬 그룹들을 확인한다. RAID 제어기(220)는 서로 다른 복수의 디스크 상에서의 동일한 로컬 그룹들의 데이터 및 로컬 패리티를 패리티 복원 관련 데이터로 읽어들인다.

【0078】 RAID 제어기(220)는 페리티 복원 관련 데이터를 이용하여 하나의 데이터 읽기와 관련된 스트라이프 내의 읽기를 수행하고자 하는 데이터를 복원한다(S650). 단계 S650에서, RAID 제어기(220)는 동일한 로컬 그룹의 데이터 및 로컬 패리티를 이용하여 하나의 데이터 읽기와 관련된 스트라이프 내의 읽기를 수행하고자 하는 데이터를 복원한다.

【0079】디스크 분산 구조는 DLRC 방식으로서, 동일한 로컬 그룹 디스크들 개수가 적으며, 오류가 발생한 디스크의 동일한 로컬 그룹이 전체 디스크로 분산되어 있으므로, 병목현상 없이 빠르게 하나의 데이터 읽기와 관련된 스트라이프 내의 읽기를 수행하고자 하는 데이터를 복원할 수 있다.

【0080】단계 S630에서 특정 디스크(예컨대, #1)에 오류가 발생한 경우, RAID 제어기(220)는 하나의 데이터 읽기와 관련된 스트라이프 내의 읽기를 수행하고자 하는 데이터를 읽어들인다(S660). RAID 제어기(220)는 스트라이프 내의 읽기를 수행하고자 하는 데이터를 복원하여 데이터 읽기 요청에 대한 응답으로 리턴한다(S670).

【0081】도 6에서는 단계 S610 내지 단계 S670을 순차적으로 실행하는 것으로 기재하고 있으나, 반드시 이에 한정되는 것은 아니다. 다시 말해, 도 6에 기재된 단계를 변경하여 실행하거나 하나 이상의 단계를 병렬적으로 실행하는 것으로 적용 가능할 것이므로, 도 6은 시계열적인 순서로 한정되는 것은 아니다.

【0082】전술한 바와 같이 도 6에 기재된 본 실시예에 따른 빠른 데이터 복원을 위한 데이터 읽기 방법은 프로그램으로 구현되고 컴퓨터로 읽을 수 있는 기록매체에 기록될 수 있다. 본 실시예에 따른 빠른 데이터 복원을 위한 데이터 읽기 방법을 구현하기 위한 프로그램이 기록되고 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다.

【0083】도 7은 본 실시예에 따른 빠른 데이터 복원을 위한 데이터 쓰기 방

법을 설명하기 위한 순서도이다.

【0084】 애플리케이션(210)은 위치(또는 논리적 블럭 주소), 오프셋을 기준으로 데이터 쓰기 요청을 발생한다(S710).

【0085】 RAID 제어기(220)는 애플리케이션(210)으로부터 적어도 하나의 데이터 위치를 기준으로 한 적어도 하나의 데이터 쓰기 요청을 수신한다. RAID 제어기(220)는 데이터 쓰기 요청에 포함된 논리적 블럭 주소 또는 오프셋을 이용하여 디스크 분산 구조 상에서 데이터 위치에 따른 패리티 배치 정보(데이터 레이아웃)를 확인한다.

【0086】 RAID 제어기(220)는 디스크 분산 구조 상에서 적어도 하나의 데이터 쓰기 요청을 처리하기 위한 디스크 번호(예컨대, #1, #2, #3... #9)를 확인한다. RAID 제어기(220)는 디스크 번호(예컨대, #1, #2, #3... #9)에 해당하는 디스크 내의 스트라이프 번호(예컨대, #0, #1, #2, #3)를 확인한다(S720).

【0087】 RAID 제어기(220)는 디스크 번호(예컨대, #1, #2, #3... #9)에 해당하는 디스크 내의 스트라이프 번호(예컨대, #0, #1, #2, #3)에 해당하는 데이터를 읽어온다(S730). RAID 제어기(220)가 단계 S730을 수행하는 이유는 글로벌 패리티를 계산하기 위함이다.

【0088】 RAID 제어기(220)는 디스크 번호에 해당하는 디스크 내의 동일한 로컬 그룹 각각에 대한 로컬 패리티를 각각 확인하고, 디스크 번호에 해당하는 디스크 내의 각 스트라이프에 대한 글로벌 패리티를 각각 확인한다(S740).

【0089】RAID 제어기(220)는 디스크 번호(예컨대, #1, #2, #3... #9)에 해당하는 디스크 내의 하나의 데이터 쓰기와 관련된 스트라이프에 대한 데이터와 하나의 데이터 쓰기를 수행하고자 하는 로컬 패리티와 글로벌 패리티를 확인한다(S750).

【0090】RAID 제어기(220)는 디스크 번호(예컨대, #1, #2, #3... #9) 중 적어도 하나의 데이터 쓰기 요청에서 하나의 데이터 쓰기를 수행하고자 하는 특정 디스크(예컨대, #1)에 오류 발생 여부를 확인한다(S760). 단계 S760에서 RAID 제어기(220)는 하나의 데이터 쓰기와 관련된 스트라이프에 대한 데이터, 하나의 데이터 쓰기를 수행하고자 하는 로컬 패리티 또는 글로벌 패리티를 포함한 특정 디스크(예컨대, #1)에 오류 발생 여부를 확인한다.

【0091】단계 S760에서 특정 디스크(예컨대, #1)에 오류가 발생하지 않은 경우, RAID 제어기(220)는 하나의 데이터 쓰기와 관련된 스트라이프에 대한 데이터에 쓰기를 수행하고, 하나의 데이터 쓰기를 수행하고자 하는 로컬 패리티 또는 글로벌 패리티를 갱신한다(S770).

【0092】단계 S760에서 특정 디스크(예컨대, #1)에 오류가 발생한 경우, RAID 제어기(220)는 특정 디스크(예컨대, #1) 내의 하나의 데이터 쓰기와 관련된 스트라이프에 데이터 쓰기가 미수행되도록 한다(S780). 단계 S780에서 RAID 제어기(220)는 특정 디스크(예컨대, #1)에 오류가 발생한 경우, 데이터 쓰기를 수행하지 않고(무시하고) 데이터 쓰기 요청에 대한 응답 데이터로서 리턴한다.

【0093】도 7에서는 단계 S710 내지 단계 S780을 순차적으로 실행하는 것으

로 기재하고 있으나, 반드시 이에 한정되는 것은 아니다. 다시 말해, 도 7에 기재된 단계를 변경하여 실행하거나 하나 이상의 단계를 병렬적으로 실행하는 것으로 적용 가능할 것이므로, 도 7은 시계열적인 순서로 한정되는 것은 아니다.

【0094】 전술한 바와 같이 도 7에 기재된 본 실시예에 따른 빠른 데이터 복원을 위한 데이터 쓰기 방법은 프로그램으로 구현되고 컴퓨터로 읽을 수 있는 기록매체에 기록될 수 있다. 본 실시예에 따른 빠른 데이터 복원을 위한 데이터 쓰기 방법을 구현하기 위한 프로그램이 기록되고 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다.

【0095】 이상의 설명은 본 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서, 본 실시예들은 본 실시예의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 실시예의 권리 범위에 포함되는 것으로 해석되어야 할 것이다.

【산업상 이용가능성】

【0096】이상에서 설명한 바와 같이 본 실시예는 RAID 분야에 적용되어, 디스크에 오류 발생 시 나타나는 급격한 데이터 입출력 성능 저하(Bottleneck)를 방지하는 효과를 발생하는 유용한 발명이다.

【부호의 설명】

【0097】200: 스토리지 장치　　210: 애플리케이션

220: RAID 제어기　　230: 스토리지

【특허청구범위】

【청구항 1】

RAID 어레이(Array) 상에서 데이터 읽기(Read) 방법에 있어서,

애플리케이션(Application)으로부터 적어도 하나의 데이터 위치(Position)를 기준으로 한 적어도 하나의 데이터 읽기 요청을 수신하는 수신 과정;

디스크 분산 구조 상에서 상기 적어도 하나의 데이터 읽기 요청을 처리하기 위한 디스크 번호 및 상기 디스크 번호(Disk Number)에 해당하는 디스크 내의 스트라이프 번호(Stripe Number)를 확인하는 확인 과정;

상기 디스크 번호 중 상기 적어도 하나의 데이터 읽기 요청에서 하나의 데이터 읽기를 수행하고자 하는 특정 디스크에 오류 발생 여부를 확인하는 오류 확인 과정;

상기 특정 디스크에 오류가 발생한 경우, 상기 특정 디스크 내의 상기 하나의 데이터 읽기와 관련된 스트라이프에 대한 데이터가 저장되면서도 상기 특정 디스크와는 다른 복수의 디스크로부터 패리티 복원 관련 데이터를 읽어들이는 읽기 과정; 및

상기 패리티 복원 관련 데이터를 이용하여 상기 스트라이프 내의 읽기를 수행하고자 하는 데이터를 복원하여 읽어들이는 복원 과정

을 포함하는 것을 특징으로 하는 빠른 데이터 복원을 위한 데이터 읽기 방법.

【청구항 2】

제 1 항에 있어서,

상기 읽기 과정은 상기 특정 디스크에 오류가 발생한 경우, 상기 디스크 분산 구조 상에서 상기 데이터 위치에 따른 패리티 배치 정보(Data Layout)를 확인하고, 상기 패리티 배치 정보를 참조하여 상기 특정 디스크 내의 상기 하나의 데이터 읽기와 관련된 스트라이프 별로 상기 특정 디스크와 동일한 로컬 그룹들을 확인하고, 서로 다른 복수의 디스크 상에서의 상기 동일한 로컬 그룹들의 데이터 및 로컬 패리티를 상기 패리티 복원 관련 데이터로 읽어들이는 것을 특징으로 하며,

상기 복원 과정은 상기 동일한 로컬 그룹의 데이터 및 상기 로컬 패리티를 이용하여 상기 스트라이프 내의 읽기를 수행하고자 하는 데이터를 복원하는 것을 특징으로 하는 빠른 데이터 복원을 위한 데이터 읽기 방법.

【청구항 3】

제 1 항에 있어서,

상기 디스크 분산 구조는,

DLRC(Distributed Local Reconstruction Code) 방식으로서, 복수 개의 디스크가 복수 개의 로컬 그룹으로 구분되며, 상기 로컬 그룹마다 로컬 패리티가 각각 존재하며, 상기 복수 개의 디스크 내의 각 스트라이프 별로 글로벌 패리티가 각각 존재하며, 동일한 로컬 그룹의 데이터들이 서로 다른 복수 개의 디스크로 분산되어 있는 분산 구조를 갖는 것을 특징으로 하는 빠른 데이터 복원을 위한 데이터 읽기

방법.

【청구항 4】

제 3 항에 있어서,

상기 로컬 패리티와 상기 글로벌 패리티는,

서로 다른 복수 개의 디스크로 분산 저장되는 분산 구조를 가지는 것을 특징으로 하는 빠른 데이터 복원을 위한 데이터 읽기 방법.

【청구항 5】

제 3 항에 있어서,

상기 로컬 패리티와 상기 글로벌 패리티는,

각각의 디스크에 별도로 저장되는 것을 특징으로 하는 빠른 데이터 복원을 위한 데이터 읽기 방법.

【청구항 6】

제 3 항에 있어서,

상기 로컬 패리티와 상기 글로벌 패리티 중 어느 하나의 패리티만이 서로 다른 복수 개의 디스크로 분산되어 있는 분산 구조를 가지며, 나머지 패리티는 하나의 디스크에 별도로 저장되는 것을 특징으로 하는 빠른 데이터 복원을 위한 데이터 읽기 방법.

【청구항 7】

제 1 항에 있어서,

상기 수신 과정은,

상기 데이터 읽기 요청에 포함된 논리적 블럭 주소(Logical Block Address) 또는 오프셋(Offset)을 이용하여 상기 디스크 분산 구조 상에서 상기 데이터 위치에 따른 패리티 배치 정보를 확인하는 것을 특징으로 하는 데이터 복원을 위한 데이터 읽기 방법.

【청구항 8】

RAID 어레이 상에서 데이터 쓰기(Write) 방법에 있어서,

애플리케이션으로부터 적어도 하나의 데이터 위치를 기준으로 한 적어도 하나의 데이터 쓰기 요청을 수신하는 수신 과정;

디스크 분산 구조 상에서 상기 적어도 하나의 데이터 쓰기 요청을 처리하기 위한 디스크 번호 및 상기 디스크 번호에 해당하는 디스크 내의 스트라이프 번호를 확인하는 확인 과정;

상기 디스크 번호 중 상기 적어도 하나의 데이터 쓰기 요청에서 하나의 데이터 쓰기를 수행하고자 하는 특정 디스크에 오류 발생 여부를 확인하는 오류 확인 과정; 및

상기 특정 디스크에 오류가 발생한 경우, 상기 특정 디스크 내의 상기 하나의 데이터 쓰기와 관련된 스트라이프에 데이터 쓰기가 미수행되도록 하는 쓰기 과정

을 포함하는 것을 특징으로 하는 빠른 데이터 복원을 위한 데이터 쓰기

방법.

【청구항 9】

제 8 항에 있어서,

상기 오류 확인 과정은,

상기 디스크 번호에 해당하는 디스크 내의 동일한 로컬 그룹 각각에 대한 로컬 패리티를 각각 확인하고, 상기 디스크 번호에 해당하는 디스크 내의 각 스트라이프에 대한 글로벌 패리티를 각각 확인하는 과정;

상기 디스크 번호에 해당하는 디스크 내의 상기 하나의 데이터 쓰기와 관련된 스트라이프에 대한 데이터와 상기 하나의 데이터 쓰기를 수행하고자 하는 로컬 패리티와 글로벌 패리티를 확인하는 과정; 및

상기 하나의 데이터 쓰기와 관련된 스트라이프에 대한 데이터, 상기 하나의 데이터 쓰기를 수행하고자 하는 로컬 패리티 또는 글로벌 패리티를 포함한 상기 특정 디스크에 오류 발생 여부를 확인하는 오류 확인 과정

을 포함하는 것을 특징으로 하는 빠른 데이터 복원을 위한 데이터 쓰기 방법.

【요약서】

【요약】

빠른 데이터 복원을 위한 데이터 읽기 또는 쓰기 방법을 개시한다.

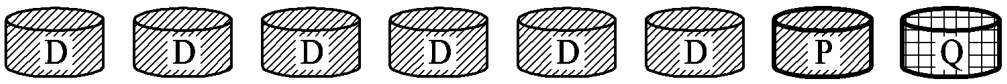
본 실시예는 RAID 어레이(Array) 상에서 데이터 읽기 또는 쓰기를 수행할 때, 디스크에 오류 발생 시 나타나는 급격한 데이터 입출력 성능 저하(Bottleneck)를 방지하기 위하여, 동일한 로컬 그룹의 데이터들이 서로 다른 복수 개의 디스크로 분산되어 있는 분산 구조 상에서 빠른 데이터 복원으로 데이터 읽기 또는 쓰기가 수행되도록 하는 빠른 데이터 복원을 위한 데이터 읽기 또는 쓰기 방법을 제공한다.

【대표도】

도 4

【도면】

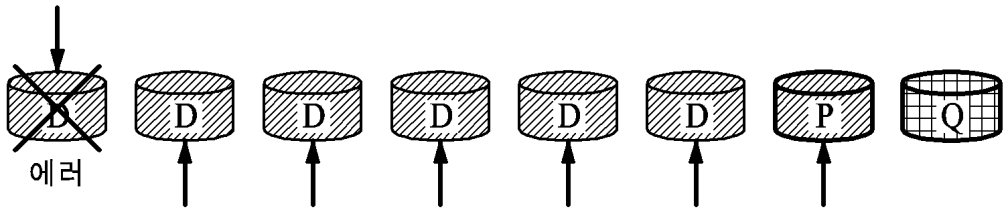
【도 1a】



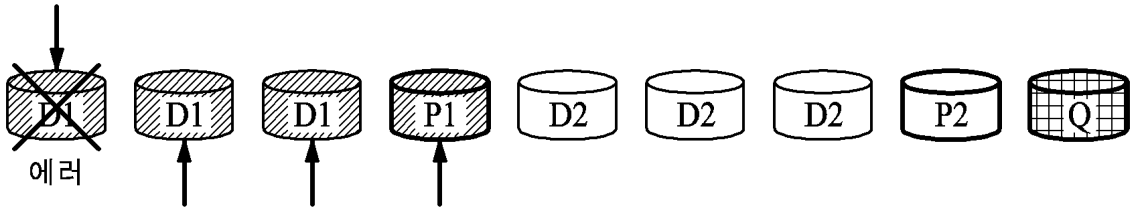
<RAID6>

<LRC>

【도 1b】



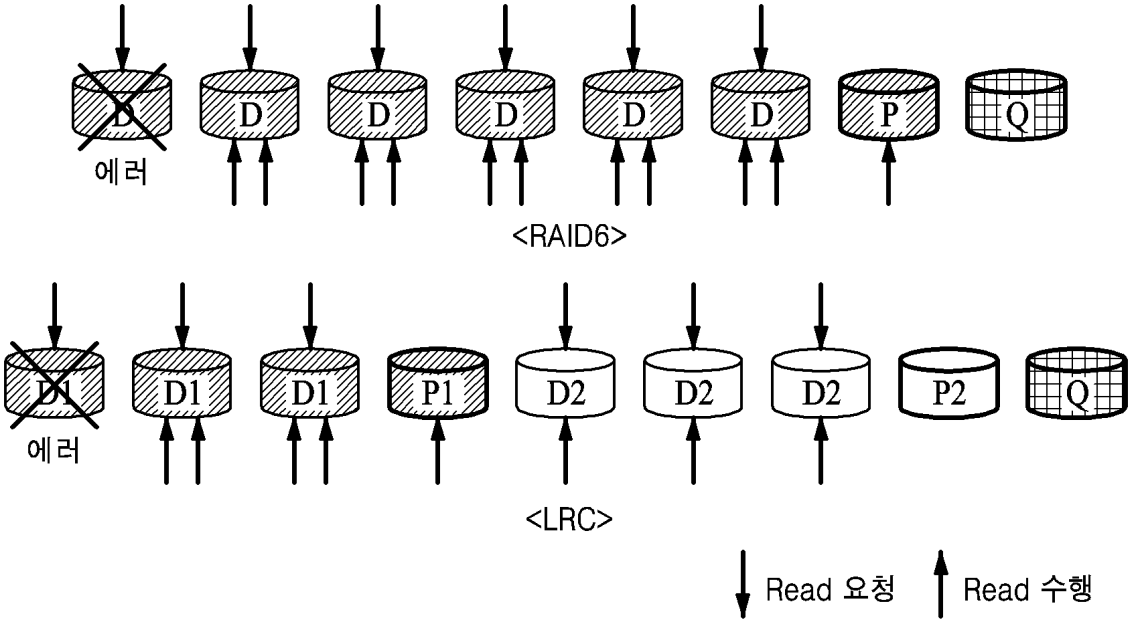
<RAID6>



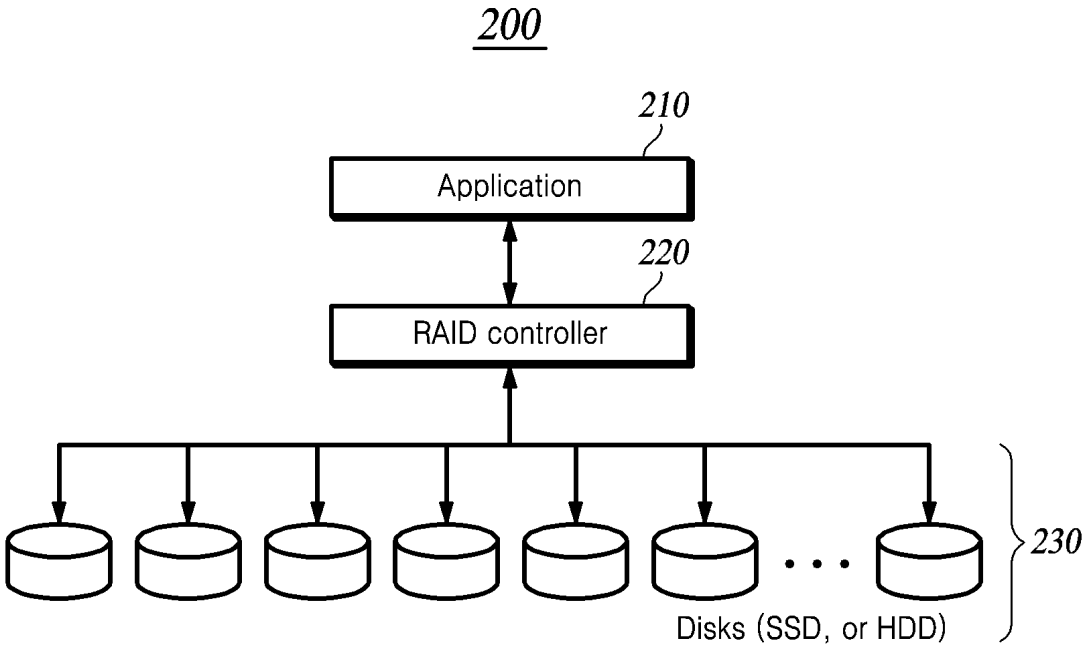
<LRC>

Read 요청 Read 수행

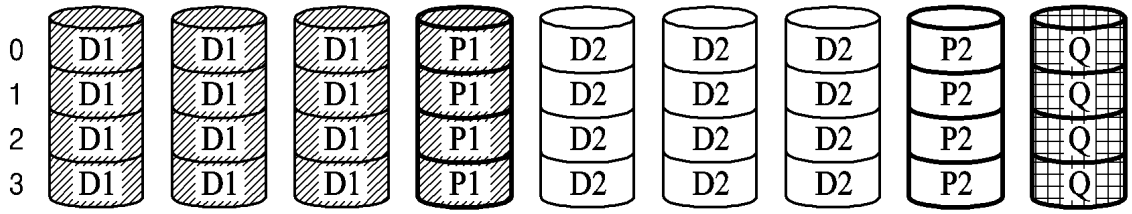
【도 1c】


D
D
D
D
D
D
P
Q
에러
<RAID6>
D1
D1
D1
P1
D2
D2
D2
P2
Q
에러
<LRC>
Read 요청
Read 수행

【도 2】

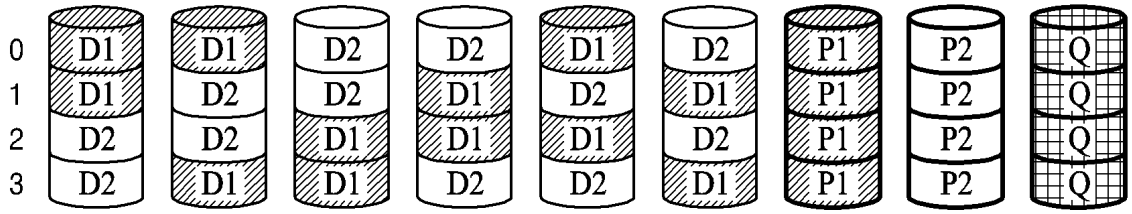

200
210
Application
220
RAID controller
230
Disks (SSD, or HDD)

【도 3】



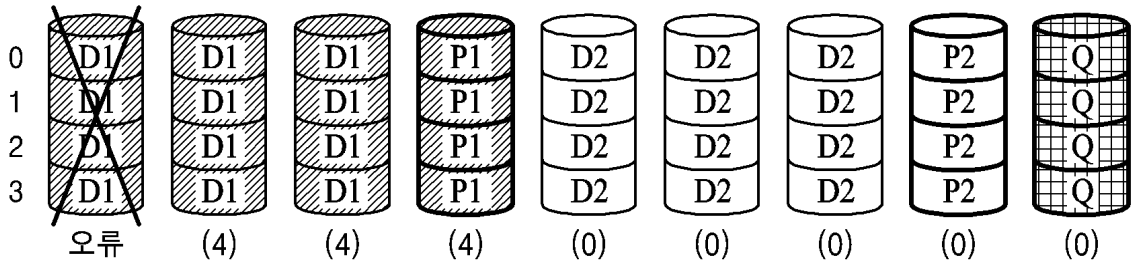
<LRC>

Stripe# mod 4



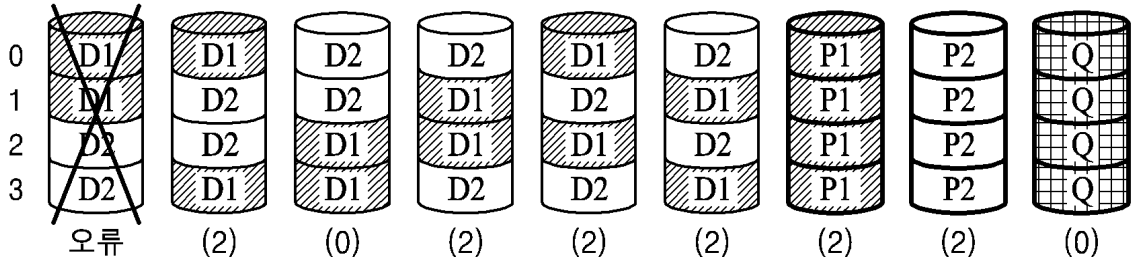
<DLRC>

【도 4】



Stripe# mod 4

<LRC>

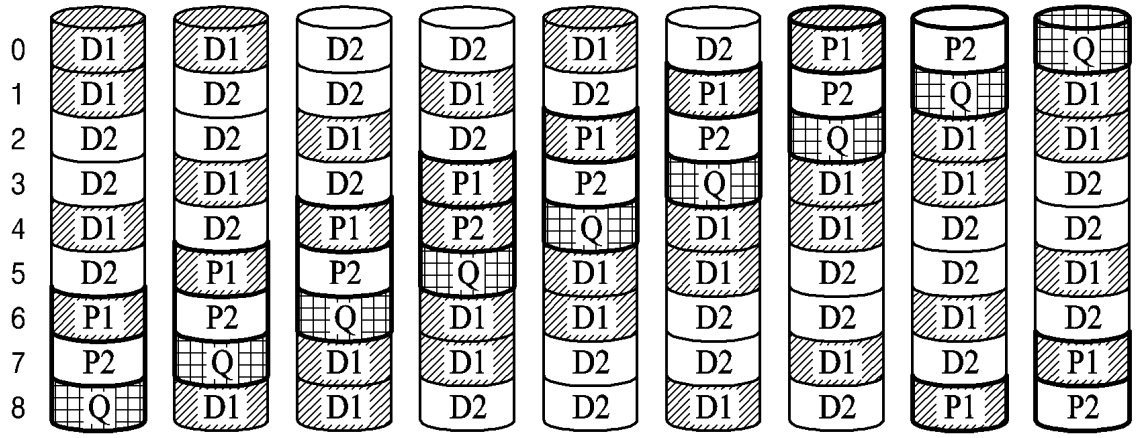


<DLRC>

(x) : 4개 데이터 복원을
위한 read 수행 수

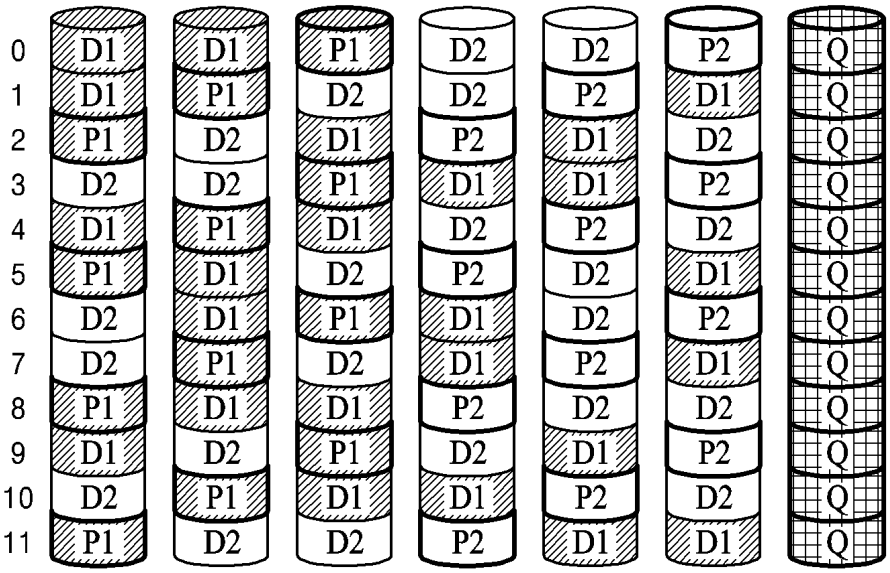
【도 5】

Stripe# mod 9



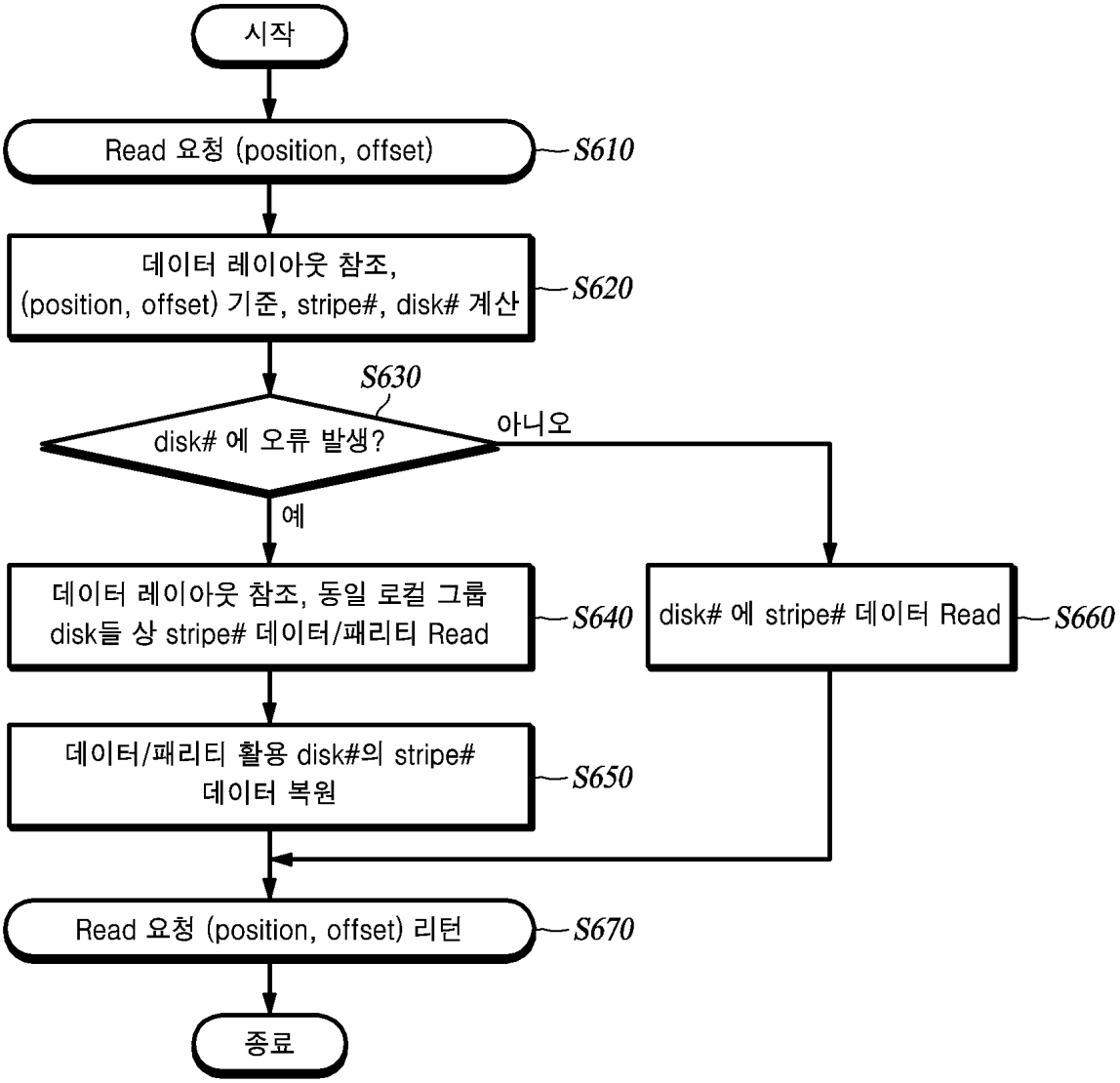
<DLRC-P 성능 초점>

Stripe# mod 12

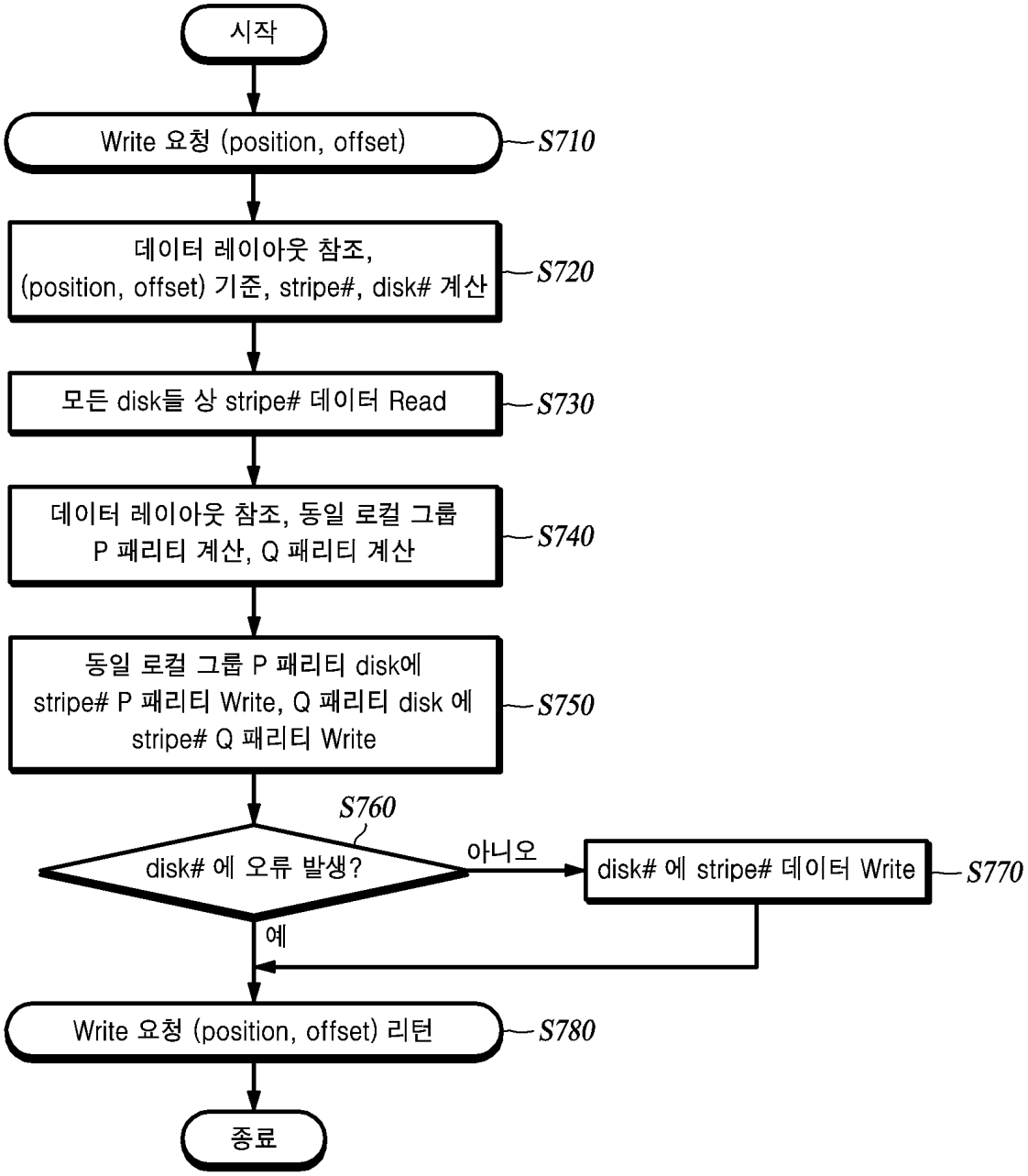


<DLRC-R 데이터 신뢰성 초점>

【도 6】


시작
Read 요청 (position, offset)
S610
데이터 레이아웃 참조,
(position, offset) 기준, stripe#, disk# 계산
S620
S630
disk# 에 오류 발생?
아니오
예
데이터 레이아웃 참조, 동일 로컬 그룹
disk들 상 stripe# 데이터/패리티 Read
S640
disk# 에 stripe# 데이터 Read
S660
데이터/패리티 활용 disk#의 stripe#
데이터 복원
S650
Read 요청 (position, offset) 리턴
S670
종료

【도 7】


시작
Write 요청 (position, offset)
S710
데이터 레이아웃 참조,
(position, offset) 기준, stripe#, disk# 계산
S720
모든 disk들 상 stripe# 데이터 Read
S730
데이터 레이아웃 참조, 동일 로컬 그룹
P 패리티 계산, Q 패리티 계산
S740
동일 로컬 그룹 P 패리티 disk에
stripe# P 패리티 Write, Q 패리티 disk 에
stripe# Q 패리티 Write
S750
S760
disk# 에 오류 발생?
아니오
disk# 에 stripe# 데이터 Write
S770
예
Write 요청 (position, offset) 리턴
S780
종료