



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2015년03월24일

(11) 등록번호 10-1505263

(24) 등록일자 2015년03월17일

(51) 국제특허분류(Int. Cl.)

G06F 11/14 (2006.01)

(21) 출원번호 10-2013-0024356

(22) 출원일자 2013년03월07일

심사청구일자 2013년03월07일

(65) 공개번호 10-2014-0110288

(43) 공개일자 2014년09월17일

(56) 선행기술조사문헌

JP1996055049 A

JP2012238038 A

KR1020120074818 A

"플래쉬 메모리 SSD 기반 해쉬 조인 알고리즘의 성능평가", 정보과학회논문지, 컴퓨팅의 실제 및 레터 제16권 제11호, 2010.11, pp.1031-1040

전체 청구항 수 : 총 18 항

(73) 특허권자

포항공과대학교 산학협력단

경상북도 포항시 남구 청암로 77 (지곡동)

(72) 발명자

박찬익

경북 포항시 남구 지곡로 155, 9동 1503호 (지곡동, 교수아파트)

박세진

대구 북구 칠곡중앙대로51길 6, 103동 1506호 (태전동, 협화맨션)

(74) 대리인

특허법인이상

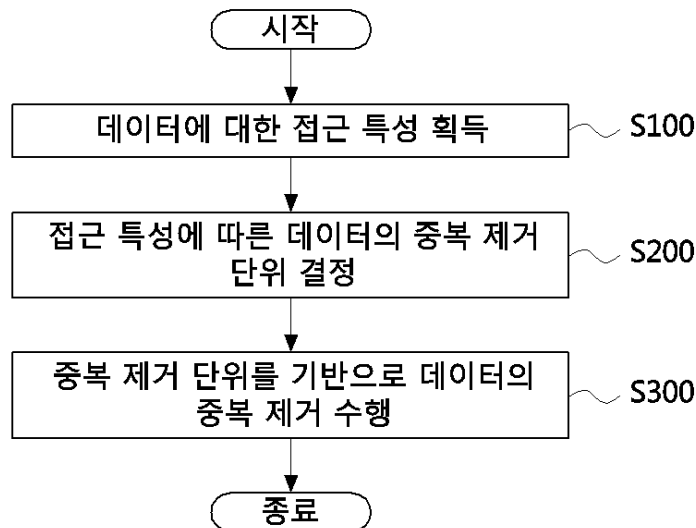
심사관 : 김정진

(54) 발명의 명칭 데이터 중복 제거 방법 및 장치

(57) 요약

데이터 중복 제거 방법 및 장치가 개시된다. 데이터의 중복 제거 방법은, 데이터의 입력 요청 또는 출력 요청을 기반으로 데이터에 대한 접근 특성을 획득하는 단계, 접근 특성을 기반으로 데이터의 중복 제거 단위를 결정하는 단계, 및 중복 제거 단위를 기반으로 데이터에 대한 중복 제거를 수행하는 단계를 포함한다. 따라서, 낮은 입출력 레이턴시를 제공할 수 있다.

대표도 - 도1



이 발명을 지원한 국가연구개발사업

과제고유번호 20120005286

부처명 교육과학기술부

연구관리전문기관 재단법인 한국연구재단

연구사업명 중견연구자지원사업

연구과제명 가상화 환경에서의 네트워크 스토리지 시스템 입출력 성능 개선을 위한 다중 수준 캐시 알고리즘 연구

기 여 율 1/1

주관기관 포항공과대학교 산학협력단

연구기간 2012.05.01 ~ 2013.04.30

특허청구의 범위

청구항 1

데이터 중복 제거 장치에서 수행되는 데이터 중복 제거 방법에 있어서,
 데이터의 입력 요청 또는 출력 요청을 기반으로 상기 데이터에 대한 접근 특성을 획득하는 단계;
 상기 접근 특성을 기반으로 상기 데이터의 중복 제거 단위를 결정하는 단계; 및
 상기 중복 제거 단위를 기반으로 상기 데이터에 대한 데이터 블록(block)을 생성하고, 상기 데이터 블록에 대한 고유의 식별자를 생성하고, 상기 고유의 식별자가 인덱스 테이블(index table)에 존재하는지 여부에 기초하여 상기 데이터에 대한 중복 제거를 수행하는 단계를 포함하는 데이터 중복 제거 방법.

청구항 2

청구항 1에 있어서,
 상기 접근 특성은,
 상기 데이터에 대한 접근 시간, 상기 데이터에 대한 변경 시간, 상기 데이터에 대한 순차적 접근 횟수 및 상기 데이터에 대한 임의적 접근 횟수 중 적어도 하나를 포함하는 것을 특징으로 하는 데이터 중복 제거 방법.

청구항 3

청구항 2에 있어서,
 상기 접근 특성을 획득하는 단계는,
 상기 데이터의 입력 요청을 수신한 경우, 상기 데이터의 입력 요청에 대한 시간 정보를 기반으로 상기 접근 시간 및 상기 변경 시간을 획득하는 단계; 및
 상기 데이터의 입력 요청의 연속성을 기반으로 상기 순차적 접근 횟수 또는 상기 임의적 접근 횟수를 획득하는 단계를 포함하는 것을 특징으로 하는 데이터 중복 제거 방법.

청구항 4

청구항 2에 있어서,
 상기 접근 특성을 획득하는 단계는,
 상기 데이터의 출력 요청을 수신한 경우, 상기 데이터의 출력 요청에 대한 시간 정보를 기반으로 상기 접근 시간을 획득하는 단계; 및
 상기 데이터의 출력 요청의 연속성을 기반으로 상기 순차적 접근 횟수 또는 상기 임의적 접근 횟수를 획득하는 단계를 포함하는 것을 특징으로 하는 데이터 중복 제거 방법.

청구항 5

청구항 2에 있어서,
 상기 중복 제거 단위를 결정하는 단계는,
 상기 데이터에 대한 현재 접근 시간과 상기 데이터에 대한 이전(以前) 변경 시간에 대한 제1 차이를 산출하는 단계;
 상기 제1 차이가 미리 정의된 제1 시간 이하인 경우, 상기 중복 제거 단위를 데이터에 대한 중복 제거 가능성이 가장 낮은 제4 중복 제거 단위로 결정하는 단계;
 상기 제1 차이가 미리 정의된 제1 시간을 초과하는 경우, 상기 데이터에 대한 현재 접근 시간과 상기 데이터에 대한 이전(以前) 접근 시간에 대한 제2 차이를 산출하는 단계;

상기 제2 차이가 미리 정의된 제2 시간을 초과하는 경우, 상기 중복 제거 단위를 데이터에 대한 중복 제거 가능성이 가장 높은 제1 중복 제거 단위로 결정하는 단계;

상기 제2 차이가 미리 정의된 제2 시간 이하인 경우, 상기 데이터에 대한 임의적 접근 횟수가 순차적 접근 횟수 이상이면 상기 중복 제거 단위를 제1 중복 제거 단위보다 중복 제거 가능성이 낮은 제2 중복 제거 단위로 결정하는 단계; 및

상기 제2 차이가 미리 정의된 제2 시간 이하인 경우, 상기 데이터에 대한 임의적 접근 횟수가 순차적 접근 횟수 미만이면 상기 중복 제거 단위를 제2 중복 제거 단위보다 중복 제거 가능성이 낮은 제3 중복 제거 단위로 결정하는 단계를 포함하는 것을 특징으로 하는 데이터 중복 제거 방법.

청구항 6

청구항 5에 있어서,

상기 제4 중복 제거 단위는,

상기 데이터에 대한 중복 제거를 수행하지 않는 것을 의미하는 것을 특징으로 하는 데이터 중복 제거 방법.

청구항 7

청구항 1에 있어서,

상기 데이터에 대한 중복 제거를 수행하는 단계는,

상기 고유의 식별자가 인덱스 테이블 내에 존재하는지 판단하는 단계;

상기 고유의 식별자가 상기 인덱스 테이블 내에 존재하는 경우, 상기 고유의 식별자에 대응된 데이터 블록을 제거하는 단계; 및

상기 고유의 식별자가 상기 인덱스 테이블 내에 존재하지 않는 경우, 상기 고유의 식별자와 상기 고유의 식별자에 대응된 데이터 블록을 저장하는 단계를 포함하는 것을 특징으로 하는 데이터 중복 제거 방법.

청구항 8

청구항 7에 있어서,

상기 고유의 식별자를 생성하는 단계는,

해시 알고리즘(hash algorithm)을 사용하여 상기 데이터 블록에 대한 고유의 식별자를 생성하는 것을 특징으로 하는 데이터 중복 제거 방법.

청구항 9

청구항 1에 있어서,

상기 중복 제거 단위는,

데이터에 대한 중복 제거 가능성을 기반으로 적어도 하나의 중복 제거 단위로 분류되는 것을 특징으로 하는 데이터 중복 제거 방법.

청구항 10

데이터의 요청 입력 또는 출력 요청을 기반으로 상기 데이터에 대한 접근 특성을 획득하고, 상기 접근 특성을 기반으로 상기 데이터의 중복 제거 단위를 결정하고, 상기 중복 제거 단위를 기반으로 상기 데이터에 대한 데이터 블록(block)을 생성하고, 상기 데이터 블록에 대한 고유의 식별자를 생성하고, 상기 고유의 식별자가 인덱스 테이블(index table)에 존재하는지 여부에 기초하여 상기 데이터에 대한 중복 제거를 수행하는 처리부; 및

상기 처리부에서 처리되는 정보 및 처리된 정보를 저장하는 저장부를 포함하는 데이터 중복 제거 장치.

청구항 11

청구항 10에 있어서,

상기 접근 특성은,

상기 데이터에 대한 접근 시간, 상기 데이터에 대한 변경 시간, 상기 데이터에 대한 순차적 접근 횟수 및 상기 데이터에 대한 임의적 접근 횟수 중 적어도 하나를 포함하는 것을 특징으로 하는 데이터 중복 제거 장치.

청구항 12

청구항 11에 있어서,

상기 처리부는,

상기 데이터의 입력 요청을 수신한 경우, 상기 데이터의 입력 요청에 대한 시간 정보를 기반으로 상기 접근 시간 및 상기 변경 시간을 획득하고, 상기 데이터의 입력 요청의 연속성을 기반으로 상기 순차적 접근 횟수 또는 상기 임의적 접근 횟수를 획득하는 것을 특징으로 하는 데이터 중복 제거 장치.

청구항 13

청구항 11에 있어서,

상기 처리부는,

상기 데이터의 출력 요청을 수신한 경우, 상기 데이터의 출력 요청에 대한 시간 정보를 기반으로 상기 접근 시간을 획득하고, 상기 데이터의 출력 요청의 연속성을 기반으로 상기 순차적 접근 횟수 또는 상기 임의적 접근 횟수를 획득하는 것을 특징으로 하는 데이터 중복 제거 장치.

청구항 14

청구항 11에 있어서,

상기 처리부는,

상기 중복 제거 단위를 결정하는 경우, 상기 데이터에 대한 현재 접근 시간과 상기 데이터에 대한 이전(以前) 변경 시간에 대한 제1 차이를 산출하고, 상기 제1 차이가 미리 정의된 제1 시간 이하인 경우 상기 중복 제거 단위를 데이터에 대한 중복 제거 가능성이 가장 낮은 제4 중복 제거 단위로 결정하고, 상기 제1 차이가 미리 정의된 제1 시간을 초과하는 경우 상기 데이터에 대한 현재 접근 시간과 상기 데이터에 대한 이전(以前) 접근 시간에 대한 제2 차이를 산출하고, 상기 제2 차이가 미리 정의된 제2 시간을 초과하는 경우 상기 중복 제거 단위를 데이터에 대한 중복 제거 가능성이 가장 높은 제1 중복 제거 단위로 결정하고,

상기 제2 차이가 미리 정의된 제2 시간 이하인 경우, 상기 데이터에 대한 임의적 접근 횟수가 순차적 접근 횟수 이상이면 상기 중복 제거 단위를 제1 중복 제거 단위보다 중복 제거 가능성이 낮은 제2 중복 제거 단위로 결정하고,

상기 제2 차이가 미리 정의된 제2 시간 이하인 경우, 상기 데이터에 대한 임의적 접근 횟수가 순차적 접근 횟수 미만이면 상기 중복 제거 단위를 제2 중복 제거 단위보다 중복 제거 가능성이 낮은 제3 중복 제거 단위로 결정하는 것을 특징으로 하는 데이터 중복 제거 장치.

청구항 15

청구항 14에 있어서,

상기 제4 중복 제거 단위는,

상기 데이터에 대한 중복 제거를 수행하지 않는 것을 의미하는 것을 특징으로 하는 데이터 중복 제거 장치.

청구항 16

청구항 10에 있어서,

상기 처리부는,

상기 데이터에 대한 중복 제거를 수행하는 경우, 상기 고유의 식별자가 인덱스 테이블 내에 존재하는지 판단하고, 상기 고유의 식별자가 상기 인덱스 테이블 내에 존재하는 경우 상기 고유의 식별자에 대응된 데이터 블록을 제거하고, 상기 고유의 식별자가 상기 인덱스 테이블 내에 존재하지 않는 경우 상기 고유의 식별자와 상기 고유

의 식별자에 대응된 데이터 블록을 저장하는 것을 특징으로 하는 데이터 중복 제거 장치.

청구항 17

청구항 16에 있어서,

상기 처리부는,

상기 고유의 식별자를 생성하는 경우, 해시 알고리즘(hash algorithm)을 사용하여 상기 데이터 블록에 대한 고유의 식별자를 생성하는 것을 특징으로 하는 데이터 중복 제거 장치.

청구항 18

청구항 10에 있어서,

상기 중복 제거 단위는,

데이터에 대한 중복 제거 가능성을 기반으로 적어도 하나의 중복 제거 단위로 분류되는 것을 특징으로 하는 데이터 중복 제거 장치.

명세서

기술분야

[0001] 본 발명은 데이터 중복 제거 기술에 관한 것으로, 더욱 상세하게는 낮은 입출력 레이턴시(latency)를 제공하기 위한 데이터 중복 제거 방법 및 장치에 관한 것이다.

배경 기술

[0002] 데이터 중복 제거 기술은 데이터 저장 장치 내에 중복된 데이터를 제거하여 더 많은 저장 공간을 확보하기 위한 기술을 의미한다. 현재 많은 기업, 공공기관 등에서 데이터의 안전한 보관을 위해 데이터의 백업(backup)을 주기적으로 수행하고 있다. 백업 데이터는 그 특성상 많은 중복적인 요소를 가지며, 이에 따라 백업 데이터의 저장 공간의 효율을 향상시키기 위해 데이터 중복 제거 기술이 사용되고 있다. 이와 같은 데이터 중복 제거 기술은 백업 데이터 저장 장치의 특성상 낮은 입출력 레이턴시(latency)를 필요로 하지 않기 때문에 중복 제거율을 높이는 기술을 중심으로 발전해 오고 있다.

[0003] 그러나, 이러한 데이터 중복 제거 기술은 중복 제거를 위한 복잡한 알고리즘(algorithm)을 기초로 수행되기 때문에, 노트북(notebook), 스마트폰(smart phone), 태블릿(tablet) PC 등과 같은 휴대용 단말에 적용하기 어려운 문제점이 있다. 즉, 이와 같은 휴대용 단말에 데이터 중복 제거 기술을 적용하는 경우, 순차적으로 저장된 데이터의 물리적 순서가 바뀌게 되므로 데이터의 입출력 속도가 심각하게 느려지는 문제점이 있다.

발명의 내용

해결하려는 과제

[0004] 상기와 같은 문제점을 해결하기 위한 본 발명의 목적은, 데이터의 입출력 특성에 따라 적응적으로 결정된 중복 제거율을 기반으로 데이터의 중복을 제거하기 위한 데이터 중복 제거 방법을 제공하는 데 있다.

[0005] 상기와 같은 문제점을 해결하기 위한 본 발명의 다른 목적은, 데이터의 입출력 특성에 따라 적응적으로 결정된 중복 제거율을 기반으로 데이터의 중복을 제거하기 위한 데이터 중복 제거 장치를 제공하는 데 있다.

과제의 해결 수단

[0006] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 데이터의 중복 제거 방법은, 데이터의 입력 요청 또는 출력 요청을 기반으로 상기 데이터에 대한 접근 특성을 획득하는 단계, 상기 접근 특성을 기반으로 상기 데이터의 중복 제거 단위를 결정하는 단계, 및 상기 중복 제거 단위를 기반으로 상기 데이터에 대한 중복 제거를 수행하는 단계를 포함한다.

[0007] 여기서, 상기 접근 특성은, 상기 데이터에 대한 접근 시간, 상기 데이터에 대한 변경 시간, 상기 데이터에 대한 순차적 접근 횟수 및 상기 데이터에 대한 임의적 접근 횟수 중 적어도 하나를 포함할 수 있다.

- [0008] 여기서, 상기 접근 특성을 획득하는 단계는, 상기 데이터의 입력 요청을 수신한 경우, 상기 데이터의 입력 요청에 대한 시간 정보를 기반으로 상기 접근 시간 및 상기 변경 시간을 획득하는 단계, 및 상기 데이터의 입력 요청의 연속성을 기반으로 상기 순차적 접근 횟수 또는 상기 임의적 접근 횟수를 획득하는 단계를 포함할 수 있다.
- [0009] 여기서, 상기 접근 특성을 획득하는 단계는, 상기 데이터의 출력 요청을 수신한 경우, 상기 데이터의 출력 요청에 대한 시간 정보를 기반으로 상기 접근 시간을 획득하는 단계, 및 상기 데이터의 출력 요청의 연속성을 기반으로 상기 순차적 접근 횟수 또는 상기 임의적 접근 횟수를 획득하는 단계를 포함할 수 있다.
- [0010] 여기서, 상기 중복 제거 단위를 결정하는 단계는, 상기 데이터에 대한 현재 접근 시간과 상기 데이터에 대한 이전 변경 시간에 대한 제1 차이를 산출하는 단계, 상기 제1 차이가 미리 정의된 제1 시간 이하인 경우, 상기 중복 제거 단위를 데이터에 대한 중복 제거 가능성이 가장 낮은 제4 중복 제거 단위로 결정하는 단계, 상기 제1 차이가 미리 정의된 제1 시간을 초과하는 경우, 상기 데이터에 대한 현재 접근 시간과 상기 데이터에 대한 이전 접근 시간에 대한 제2 차이를 산출하는 단계, 상기 제2 차이가 미리 정의된 제2 시간을 초과하는 경우, 상기 중복 제거 단위를 데이터에 대한 중복 제거 가능성이 가장 높은 제1 중복 제거 단위로 결정하는 단계, 상기 제2 차이가 미리 정의된 제2 시간 이하인 경우, 상기 데이터에 대한 임의적 접근 횟수가 순차적 접근 횟수 이상이면 상기 중복 제거 단위를 제1 중복 제거 단위보다 중복 제거 가능성이 낮은 제2 중복 제거 단위로 결정하는 단계, 및 상기 제2 차이가 미리 정의된 제2 시간 이하인 경우, 상기 데이터에 대한 임의적 접근 횟수가 순차적 접근 횟수 미만이면 상기 중복 제거 단위를 제2 중복 제거 단위보다 중복 제거 가능성이 낮은 제3 중복 제거 단위로 결정하는 단계를 포함할 수 있다.
- [0011] 여기서, 상기 제4 중복 제거 단위는, 상기 데이터에 대한 중복 제거를 수행하지 않는 것을 의미할 수 있다.
- [0012] 여기서, 상기 데이터에 대한 중복 제거를 수행하는 단계는, 상기 중복 제거 단위를 기반으로 상기 데이터에 대한 적어도 하나의 데이터 블록을 생성하는 단계, 상기 데이터 블록에 대한 고유의 식별자를 생성하는 단계, 상기 고유의 식별자가 인덱스 테이블 내에 존재하는지 판단하는 단계, 상기 고유의 식별자가 상기 인덱스 테이블 내에 존재하는 경우, 상기 고유의 식별자에 대응된 데이터 블록을 제거하는 단계, 및 상기 고유의 식별자가 상기 인덱스 테이블 내에 존재하지 않는 경우, 상기 고유의 식별자와 상기 고유의 식별자에 대응된 데이터 블록을 저장하는 단계를 포함할 수 있다.
- [0013] 여기서, 상기 고유의 식별자를 생성하는 단계는, 해시 알고리즘을 사용하여 상기 데이터 블록에 대한 고유의 식별자를 생성할 수 있다.
- [0014] 여기서, 상기 중복 제거 단위는, 데이터에 대한 중복 제거 가능성을 기반으로 적어도 하나의 중복 제거 단위로 분류될 수 있다.
- [0015] 상기 다른 목적을 달성하기 위한 본 발명의 일 실시예에 따른 데이터 중복 제거 장치는, 데이터의 요청 입력 또는 출력 요청을 기반으로 상기 데이터에 대한 접근 특성을 획득하고, 상기 접근 특성을 기반으로 상기 데이터의 중복 제거 단위를 결정하고, 상기 중복 제거 단위를 기반으로 상기 데이터에 대한 중복 제거를 수행하는 처리부, 및 상기 처리부에서 처리되는 정보 및 처리된 정보를 저장하는 저장부를 포함한다.
- [0016] 여기서, 상기 접근 특성은, 상기 데이터에 대한 접근 시간, 상기 데이터에 대한 변경 시간, 상기 데이터에 대한 순차적 접근 횟수 및 상기 데이터에 대한 임의적 접근 횟수 중 적어도 하나를 포함할 수 있다.
- [0017] 여기서, 상기 처리부는, 상기 데이터의 입력 요청을 수신한 경우, 상기 데이터의 입력 요청에 대한 시간 정보를 기반으로 상기 접근 시간 및 상기 변경 시간을 획득하고, 상기 데이터의 입력 요청의 연속성을 기반으로 상기 순차적 접근 횟수 또는 상기 임의적 접근 횟수를 획득할 수 있다.
- [0018] 여기서, 상기 처리부는, 상기 데이터의 출력 요청을 수신한 경우, 상기 데이터의 출력 요청에 대한 시간 정보를 기반으로 상기 접근 시간을 획득하고, 상기 데이터의 출력 요청의 연속성을 기반으로 상기 순차적 접근 횟수 또는 상기 임의적 접근 횟수를 획득할 수 있다.
- [0019] 여기서, 상기 처리부는, 상기 중복 제거 단위를 결정하는 경우, 상기 데이터에 대한 현재 접근 시간과 상기 데이터에 대한 이전 변경 시간에 대한 제1 차이를 산출하고, 상기 제1 차이가 미리 정의된 제1 시간 이하인 경우 상기 중복 제거 단위를 데이터에 대한 중복 제거 가능성이 가장 낮은 제4 중복 제거 단위로 결정하고, 상기 제1 차이가 미리 정의된 제1 시간을 초과하는 경우 상기 데이터에 대한 현재 접근 시간과 상기 데이터에 대한 이전 접근 시간에 대한 제2 차이를 산출하고, 상기 제2 차이가 미리 정의된 제2 시간을 초과하는 경우 상기 중복 제

거 단위를 데이터에 대한 중복 제거 가능성이 가장 높은 제1 중복 제거 단위로 결정하고, 상기 제2 차이가 미리 정의된 제2 시간 이하인 경우, 상기 데이터에 대한 임의적 접근 횟수가 순차적 접근 횟수 이상이면 상기 중복 제거 단위를 제1 중복 제거 단위보다 중복 제거 가능성이 낮은 제2 중복 제거 단위로 결정하고, 상기 제2 차이가 미리 정의된 제2 시간 이하인 경우, 상기 데이터에 대한 임의적 접근 횟수가 순차적 접근 횟수 미만이면 상기 중복 제거 단위를 제2 중복 제거 단위보다 중복 제거 가능성이 낮은 제3 중복 제거 단위로 결정할 수 있다.

- [0020] 여기서, 상기 제4 중복 제거 단위는, 상기 데이터에 대한 중복 제거를 수행하지 않는 것을 의미할 수 있다.
- [0021] 여기서, 상기 처리부는, 상기 데이터에 대한 중복 제거를 수행하는 경우, 상기 중복 제거 단위를 기반으로 상기 데이터에 대한 적어도 하나의 데이터 블록을 생성하고, 상기 데이터 블록에 대한 고유의 식별자를 생성하고, 상기 고유의 식별자가 인덱스 테이블 내에 존재하는지 판단하고, 상기 고유의 식별자가 상기 인덱스 테이블 내에 존재하는 경우 상기 고유의 식별자에 대응된 데이터 블록을 제거하고, 상기 고유의 식별자가 상기 인덱스 테이블 내에 존재하지 않는 경우 상기 고유의 식별자와 상기 고유의 식별자에 대응된 데이터 블록을 저장할 수 있다.
- [0022] 여기서, 상기 처리부는, 상기 고유의 식별자를 생성하는 경우, 해시 알고리즘을 사용하여 상기 데이터 블록에 대한 고유의 식별자를 생성할 수 있다.
- [0023] 여기서, 상기 중복 제거 단위는, 데이터에 대한 중복 제거 가능성을 기반으로 적어도 하나의 중복 제거 단위로 분류될 수 있다.

발명의 효과

- [0024] 본 발명에 의하면, 데이터의 입출력 특성을 기반으로 중복 제거율(즉, 청크(chunk)의 크기)을 적응적으로 결정할 수 있고, 적응적으로 결정된 중복 제거율을 기반으로 데이터의 중복을 제거할 수 있으므로, 낮은 입출력 레이턴시(latency)를 제공할 수 있다.

도면의 간단한 설명

- [0025] 도 1은 본 발명의 일 실시예에 따른 데이터 중복 제거 방법을 도시한 흐름도이다.
- 도 2는 데이터에 대한 접근 특성을 나타낸 표이다.
- 도 3은 본 발명의 일 실시예에 따른 데이터 중복 제거 방법에 있어서 접근 특성 획득 단계를 도시한 흐름도이다.
- 도 4는 본 발명의 일 실시예에 따른 데이터 중복 제거 방법에 있어서 중복 제거 단위 결정 단계를 도시한 흐름도이다.
- 도 5는 중복 제거 단위에 대한 특성을 나타낸 표이다.
- 도 6은 본 발명의 일 실시예에 따른 데이터 중복 제거 방법에 있어서 중복 제거 수행 단계를 도시한 흐름도이다.
- 도 7은 데이터 블록에 대한 고유의 식별자를 생성하는 과정을 도시한 개념도이다.
- 도 8은 본 발명의 일 실시예에 따른 데이터 중복 제거 장치를 도시한 블록도이다.

발명을 실시하기 위한 구체적인 내용

- [0026] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세하게 설명하고자 한다.
- [0027] 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.
- [0028] 제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의

조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.

- [0029] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.
- [0030] 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0031] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가진 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0032] 이하, 첨부한 도면들을 참조하여, 본 발명의 바람직한 실시예를 보다 상세하게 설명하고자 한다. 본 발명을 설명함에 있어 전체적인 이해를 용이하게 하기 위하여 도면상의 동일한 구성요소에 대해서는 동일한 참조부호를 사용하고 동일한 구성요소에 대해서 중복된 설명은 생략한다.
- [0033] 도 1은 본 발명의 일 실시예에 따른 데이터 중복 제거 방법을 도시한 흐름도이다.
- [0034] 도 1을 참조하면, 본 발명의 일 실시예에 따른 데이터 중복 제거 방법은 데이터의 입력 요청 또는 출력 요청을 기반으로 데이터에 대한 접근 특성을 획득하는 단계(S100), 접근 특성을 기반으로 데이터의 중복 제거 단위를 결정하는 단계(S200), 및 중복 제거 단위를 기반으로 데이터에 대한 중복 제거를 수행하는 단계(S300)를 포함한다.
- [0035] 여기서, 도 1에 도시된 각 단계는 도 8에 도시된 데이터 중복 제거 장치에서 수행될 수 있으며, 데이터 중복 제거 장치의 구체적인 구성과 그 기능에 대해서는 후술하도록 한다.
- [0036] 도 2는 데이터에 대한 접근 특성을 나타낸 표이다.
- [0037] 도 2를 참조하면, 데이터에 대한 접근 특성은 데이터에 대한 접근 시간(a_time), 데이터에 대한 변경 시간(m_time), 데이터에 대한 순차적 접근 횟수(seqCount), 데이터에 대한 임의적 접근 횟수(randCount) 등을 포함할 수 있다. 여기서, 데이터에 대한 접근 특성은 상기 정보 등에 한정되지 않고 데이터의 입력 또는 출력에 대한 특성을 나타낼 수 있는 정보를 포함할 수 있다.
- [0038] 데이터에 대한 접근 시간(a_time)은 데이터의 입력 요청(즉, 데이터의 쓰기 요청) 또는 데이터의 출력 요청(즉, 데이터의 읽기 요청)을 수신한 시간을 의미한다. 데이터의 입력 요청을 받은 경우, 데이터 중복 제거 장치는 입력 요청 시간을 해당 데이터에 대한 접근 시간으로 획득할 수 있고, 획득한 접근 시간을 저장(즉, 현재 데이터에 대한 a_time 필드를 현재 시간으로 기록)할 수 있다. 한편, 이미 저장된 접근 시간이 있는 경우, 데이터 중복 제거 장치는 이미 저장된 접근 시간을 최근에 획득한 접근 시간으로 갱신할 수 있다.
- [0039] 한편, 데이터의 출력 요청을 받은 경우, 데이터 중복 제거 장치는 출력 요청 시간을 해당 데이터에 대한 접근 시간으로 획득할 수 있고, 획득한 접근 시간을 저장(즉, 현재 데이터에 대한 a_time 필드를 현재 시간으로 기록)할 수 있다. 한편, 이미 저장된 접근 시간이 있는 경우, 데이터 중복 제거 장치는 이미 저장된 접근 시간을 최근에 획득한 접근 시간으로 갱신할 수 있다.
- [0040] 데이터에 대한 변경 시간(m_time)은 데이터의 입력 요청을 수신한 시간을 의미한다. 데이터의 입력 요청을 받은 경우, 데이터 중복 제거 장치는 입력 요청 시간을 해당 데이터에 대한 변경 시간으로 획득할 수 있고, 획득한 변경 시간을 저장(즉, 현재 데이터에 대한 m_time 필드를 현재 시간으로 기록)할 수 있다. 한편, 이미 저장된 변경 시간이 있는 경우, 데이터 중복 제거 장치는 이미 저장된 변경 시간을 최근에 획득한 변경 시간으로 갱신

할 수 있다.

- [0041] 데이터에 대한 순차적 접근 횟수(seqCount)는 현재의 데이터 요청과 이전의 데이터 요청이 연속된(즉, 현재의 데이터 요청과 이전의 데이터 요청의 물리적 또는 논리적 블록번호가 연속적인 경우, 또는 그 요청이 연속적인 추세를 가지는 경우) 횟수를 의미할 수 있고, 데이터에 대한 임의적 접근 횟수(randCount)는 현재의 데이터 요청과 이전 데이터 요청이 연속되지 않은 횟수를 의미할 수 있다. 여기서, 순차적 접근 횟수와 임의적 접근 횟수는 현재까지 누적된 횟수를 의미할 수 있다.
- [0042] 현재의 데이터 요청과 이전의 데이터 요청이 연속되는 경우, 데이터 중복 제거 장치는 현재 데이터에 대한 seqCount 필드의 값을 1 증가시킬 수 있다. 반면, 현재의 데이터 요청과 이전의 데이터 요청이 연속되지 않는 경우, 데이터 중복 제거 장치는 현재 데이터에 대한 randCount 필드의 값을 1 증가시킬 수 있다.
- [0043] 도 3은 본 발명의 일 실시예에 따른 데이터 중복 제거 방법에 있어서 접근 특성 획득 단계를 도시한 흐름도이다.
- [0044] 도 3을 참조하면, 데이터에 대한 접근 특성을 획득하는 단계(S100)는, 수신된 요청이 데이터의 입력 요청 또는 데이터의 출력 요청에 해당하는지 판단하는 단계(S110), 데이터의 입력 요청을 수신한 경우 데이터의 입력 요청에 대한 시간 정보를 기반으로 접근 시간 및 변경 시간을 획득하는 단계(S120), 및 데이터의 입력 요청의 연속성을 기반으로 순차적 접근 횟수 또는 임의적 접근 횟수를 획득하는 단계(S130)를 포함할 수 있다.
- [0045] 더불어, 데이터에 대한 접근 특성을 획득하는 단계(S100)는 데이터의 출력 요청을 수신한 경우 데이터의 출력 요청에 대한 시간 정보를 기반으로 접근 시간을 획득하는 단계(S140), 및 데이터의 출력 요청의 연속성을 기반으로 순차적 접근 횟수 또는 임의적 접근 횟수를 획득하는 단계(S150)를 포함할 수 있다.
- [0046] 단계 S110에서, 데이터 중복 제거 장치는 수신된 요청이 데이터의 입력 요청에 해당하는지 데이터의 출력 요청에 해당하는지 판단할 수 있다. 수신된 요청이 데이터의 입력 요청에 해당하는 경우, 데이터 중복 제거 장치는 다음 단계로 단계 S120, 단계 S130을 수행할 수 있다. 반면, 수신된 요청이 데이터의 출력 요청에 해당하는 경우, 데이터 중복 제거 장치는 다음 단계로 단계 S140, 단계 S150을 수행할 수 있다.
- [0047] 단계 S120에서, 데이터 중복 제거 장치는 데이터의 입력 요청에 대한 시간 정보를 기반으로 접근 시간 및 변경 시간을 획득할 수 있다. 즉, 데이터 중복 제거 장치는 데이터의 입력 요청을 수신한 시간을 접근 시간 및 변경 시간으로 획득할 수 있고, 획득한 접근 시간 및 획득한 변경 시간을 데이터베이스(database)에 저장할 수 있다. 이때, 이미 저장된 접근 시간이 존재하는 경우 데이터 중복 제거 장치는 이미 저장된 접근 시간을 상기 획득한 접근 시간으로 갱신할 수 있고, 이미 저장된 변경 시간이 존재하는 경우 데이터 중복 제거 장치는 이미 저장된 변경 시간을 상기 획득한 변경 시간으로 갱신할 수 있다.
- [0048] 단계 S130에서, 데이터 중복 제거 장치는 데이터의 입력 요청과 해당 데이터에 대한 이전(以前) 요청의 연속성을 기반으로 순차적 접근 횟수 또는 임의적 접근 횟수를 획득할 수 있다. 즉, 데이터 중복 제거 장치는 데이터의 입력 요청과 해당 데이터에 대한 이전(以前) 요청이 연속되는 경우 순차적 접근 횟수를 1 증가시킬 수 있고, 데이터의 입력 요청과 해당 데이터에 대한 이전(以前) 요청이 연속되지 않는 경우 임의적 접근 횟수를 1 증가시킬 수 있다.
- [0049] 예를 들어, 데이터베이스에 저장된 접근 특성 중 순차적 접근 횟수가 7 이고 임의적 접근 횟수가 5 인 경우, 데이터 중복 제거 장치는 데이터의 입력 요청과 해당 데이터에 대한 이전(以前) 요청이 연속되는 경우 순차적 접근 횟수를 8 로 갱신할 수 있고, 데이터의 입력 요청과 해당 데이터에 대한 이전(以前) 요청이 연속되지 않는 경우 임의적 접근 횟수를 6 으로 갱신할 수 있다.
- [0050] 여기서, 단계 S120을 먼저 수행한 후 단계 S130을 수행하는 것으로 설명하였으나, 단계 S120과 단계 S130의 수행 순서는 이에 한정되지 않는다. 즉, 단계 S130은 단계 S120과 동시에 수행될 수 있고, 또는 단계 S120보다 먼저 수행될 수 있다.
- [0051] 단계 S140에서, 데이터 중복 제거 장치는 데이터의 출력 요청에 대한 시간 정보를 기반으로 접근 시간을 획득할 수 있다. 즉, 데이터 중복 제거 장치는 데이터의 출력 요청을 수신한 시간을 접근 시간으로 획득할 수 있고, 획득한 접근 시간을 데이터베이스에 저장할 수 있다. 이때, 이미 저장된 접근 시간이 존재하는 경우 데이터 중복 제거 장치는 이미 저장된 접근 시간을 상기 획득한 접근 시간으로 갱신할 수 있다.

- [0052] 단계 S150에서, 데이터 중복 제거 장치는 데이터의 출력 요청과 해당 데이터에 대한 이전(以前) 요청의 연속성을 기반으로 순차적 접근 횟수 또는 임의적 접근 횟수를 획득할 수 있다. 즉, 데이터 중복 제거 장치는 데이터의 출력 요청과 해당 데이터에 대한 이전(以前) 요청이 연속되는 경우 순차적 접근 횟수를 1 증가시킬 수 있고, 데이터의 입력 요청과 해당 데이터에 대한 이전(以前) 요청이 연속되지 않는 경우 임의적 접근 횟수를 1 증가시킬 수 있다.
- [0053] 예를 들어, 데이터베이스에 저장된 접근 특성 중 순차적 접근 횟수가 7 이고 임의적 접근 횟수가 5 인 경우, 데이터 중복 제거 장치는 데이터의 출력 요청과 해당 데이터에 대한 이전(以前) 요청이 연속되는 경우 순차적 접근 횟수를 8 로 갱신할 수 있고, 데이터의 출력 요청과 해당 데이터에 대한 이전(以前) 요청이 연속되지 않는 경우 임의적 접근 횟수를 6 으로 갱신할 수 있다.
- [0054] 여기서, 단계 S140을 먼저 수행한 후 단계 S150을 수행하는 것으로 설명하였으나, 단계 S140과 단계 S150의 수행 순서는 이에 한정되지 않는다. 즉, 단계 S150은 단계 S140과 동시에 수행될 수 있고, 또는 단계 S140보다 먼저 수행될 수 있다.
- [0055] 도 4는 본 발명의 일 실시예에 따른 데이터 중복 제거 방법에 있어서 중복 제거 단위 결정 단계를 도시한 흐름도이다.
- [0056] 도 4를 참조하면, 중복 제거 단위를 결정하는 단계(S200)는, 데이터에 대한 현재 접근 시간과 데이터에 대한 이전(以前) 변경 시간에 대한 제1 차이를 산출하는 단계(S210), 제1 차이가 미리 정의된 제1 시간을 초과하는지 판단하는 단계(S220), 제1 차이가 미리 정의된 제1 시간 이하인 경우, 중복 제거 단위를 데이터에 대한 중복 제거 가능성이 가장 낮은 제4 중복 제거 단위로 결정하는 단계(S230), 제1 차이가 미리 정의된 제1 시간을 초과하는 경우, 데이터에 대한 현재 접근 시간과 데이터에 대한 이전(以前) 접근 시간에 대한 제2 차이를 산출하는 단계(S240), 제2 차이가 미리 정의된 제2 시간 이하인지 판단하는 단계(S250), 제2 차이가 미리 정의된 제2 시간을 초과하는 경우, 중복 제거 단위를 데이터에 대한 중복 제거 가능성이 가장 높은 제1 중복 제거 단위로 결정하는 단계(S260), 제2 차이가 미리 정의된 제2 시간 이하인 경우, 데이터에 대한 임의적 접근 횟수가 데이터에 대한 순차적 접근 횟수 이하인지 판단하는 단계(S270), 제2 차이가 미리 정의된 제2 시간 이하인 경우, 데이터에 대한 임의적 접근 횟수가 순차적 접근 횟수 이상이면 중복 제거 단위를 제1 중복 제거 단위보다 중복 제거 가능성이 낮은 제2 중복 제거 단위로 결정하는 단계(S280), 및 제2 차이가 미리 정의된 제2 시간 이하인 경우, 데이터에 대한 임의적 접근 횟수가 순차적 접근 횟수 미만이면 중복 제거 단위를 제2 중복 제거 단위보다 중복 제거 가능성이 낮은 제3 중복 제거 단위로 결정하는 단계(S290)을 포함할 수 있다.
- [0057] 이하 단계 S200에서 결정되는 중복 제거 단위에 대해 도 5를 참조하여 상세하게 설명한다.
- [0058] 도 5는 중복 제거 단위에 대한 특성을 나타낸 표이다.
- [0059] 도 5를 참조하면, 중복 제거 단위는 제1 중복 제거 단위, 제2 중복 제거 단위, 제3 중복 제거 단위, 제4 중복 제거 단위로 분류할 수 있다. 제1 중복 제거 단위는 자주 사용되지 않는 데이터에 적용될 수 있으며, 자주 사용되지 않는 데이터는 데이터에 대한 현재 접근 시간과 이전(以前) 접근 시간의 차이가 미리 정의된 임계값보다 큰 데이터를 의미할 수 있다. 제1 중복 제거 단위는 모든 중복 제거 단위 중 가장 작은 청크(chunk)를 사용하며, 이에 따라 모든 중복 제거 단위 중 가장 높은 데이터 중복 제거율을 제공할 수 있다. 즉, 제1 중복 제거 단위는 낮은 레이턴시(latency)보다 높은 데이터 중복 제거율을 제공하기 위해 사용될 수 있다.
- [0060] 제2 중복 제거 단위는 순차적 접근보다 임의적 접근이 자주 발생하는 데이터에 적용될 수 있다. 제2 중복 제거 단위는 제1 중복 제거 단위보다 크고 제3 중복 제거 단위보다 작은 크기의 청크를 사용하며, 이에 따라 모든 중복 제거 단위 중 상대적으로 높은 데이터 중복 제거율(즉, 제1 중복 제거 단위보다 낮고 제3 중복 제거 단위보다 높은 중복 제거율)을 제공할 수 있다. 즉, 임의적 접근이 자주 발생하는 데이터의 경우 물리적인 뒤틀림이 발생하여도 임의적 접근 성능에 문제가 발생하지 않기 때문에, 제2 중복 제거 단위는 높은 중복 제거율을 제공하기 위해 상대적으로 작은 크기의 청크를 사용할 수 있다.
- [0061] 제3 중복 제거 단위는 임의적 접근보다 순차적 접근이 자주 발생하는 데이터에 적용될 수 있다. 제3 중복 제거 단위는 제2 중복 제거 단위보다 큰 크기의 청크를 사용하며, 이에 따라 모든 중복 제거 단위 중 상대적으로 낮은 중복 제거율(즉, 제2 중복 제거 단위보다 낮은 중복 제거율)을 제공할 수 있다. 즉, 순차적인 접근이 자주

발생하는 데이터에 대해 낮은 입출력 레이턴시를 제공하기 위해, 제3 중복 제거 단위는 상대적으로 큰 크기의 청크를 사용할 수 있다.

[0062] 제4 중복 제거 단위는 입력이 자주 발생하는 데이터에 적용될 수 있다. 제4 중복 제거 단위는 제3 중복 제거 단위보다 큰 크기의 청크를 사용할 수 있으며, 이에 따라 모든 중복 제거 단위 중 가장 작은 중복 제거율(즉, 제3 중복 제거 단위보다 낮은 중복 제거율)을 제공할 수 있다. 한편, 제4 중복 제거 단위는 데이터에 대한 중복 제거를 수행하지 않는 것을 의미할 수도 있다. 즉, 데이터의 중복 제거는 출력 위주의 데이터에 유리하므로, 입력 위주의 데이터의 경우 중복 제거를 수행하지 않을 수 있다.

[0063] 여기서, 중복 제거 단위의 분류는 상기 설명에 한정되지 아니하고 다양하게 구성될 수 있다. 예를 들어, 중복 제거 단위를 3개 분류 또는 5개 분류로 구성할 수 있다. 중복 제거 단위가 3개 분류로 구성되는 경우, 제1 중복 제거 단위는 가장 높은 중복 제거율을 제공할 수 있고, 제2 중복 제거 단위는 제1 중복 제거 단위보다 낮은 중복 제거율을 제공할 수 있고, 제3 중복 제거 단위는 제2 중복 제거 단위보다 낮은 중복 제거율(즉, 가장 낮은 중복 제거율)을 제공할 수 있다.

[0064] 다시 도 4를 참조하면, 단계 S210에서, 데이터 중복 제거 장치는 데이터에 대한 현재 접근 시간과 데이터에 대한 이전(以前) 변경 시간에 대한 제1 차이를 산출할 수 있다. 즉, 데이터 중복 제거 장치는 데이터의 입력 요청 또는 출력 요청으로부터 획득한 현재 접근 시간과, 동일한 데이터에 대한 이전(以前) 변경 시간(즉, 데이터의 이전(以前) 입력 요청으로부터 획득한 변경 시간)의 차이인 제1 차이를 산출할 수 있다. 여기서, 제1 차이는 해당 데이터가 얼마나 자주 변경되는지(즉, 데이터의 입력 요청이 얼마나 자주 발생하는지)를 의미할 수 있다.

[0065] 단계 S220에서, 데이터 중복 제거 장치는 제1 차이가 미리 정의된 제1 시간을 초과하는지 판단할 수 있다. 여기서, 미리 정의된 제1 시간은 데이터 중복 제거를 수행하기 위한 기준이 되는 시간을 의미하며, 사용자의 설정에 따라 다른 값을 가질 수 있다. 예를 들어, 미리 정의된 제1 시간은 1 시간, 2 시간, 3 시간 등으로 설정될 수 있다. 제1 차이가 미리 정의된 제1 시간 이하인 경우 데이터 중복 제거 장치는 다음 단계로 단계 S230을 수행할 수 있고, 제1 차이가 미리 정의된 제1 시간을 초과하는 경우 데이터 중복 제거 장치는 다음 단계로 단계 S240을 수행할 수 있다.

[0066] 단계 S230에서, 데이터 중복 제거 장치는 중복 제거 단위를 데이터에 대한 중복 제거 가능성이 가장 낮은 제4 중복 제거 단위로 결정할 수 있다. 즉, 제1 차이가 미리 정의된 제1 시간 이하인 경우 이는 입력이 자주 발생하는 데이터를 의미하므로, 데이터 중복 제거 장치는 중복 제거 단위 중 데이터 중복 제거율이 가장 낮은(또는, 데이터 중복 제거를 수행하지 않는) 제4 중복 제거 단위를 선택할 수 있다.

[0067] 단계 S240에서, 데이터 중복 제거 장치는 데이터에 대한 현재 접근 시간과 데이터에 대한 이전(以前) 접근 시간에 대한 제2 차이를 산출할 수 있다. 즉, 데이터 중복 제거 장치는 데이터의 입력 요청 또는 출력 요청으로부터 획득한 현재 접근 시간과, 동일한 데이터에 대한 이전(以前) 접근 시간(즉, 데이터의 이전(以前) 입력 요청 또는 이전(以前) 출력 요청으로부터 획득한 접근 시간)의 차이인 제2 차이를 산출할 수 있다. 여기서, 제2 차이는 해당 데이터에 대한 접근이 얼마나 자주 발생하는지를 의미할 수 있다.

[0068] 단계 S250에서, 데이터 중복 제거 장치는 제2 차이가 미리 정의된 제2 시간 이하인지 판단할 수 있다. 여기서, 미리 정의된 제2 시간은 데이터에 대한 접근이 발생할 가능성이 낮은 데이터를 구별하기 위해 기준이 되는 시간을 의미하며, 사용자의 설정에 따라 다른 값을 가질 수 있다. 예를 들어, 미리 정의된 제2 시간은 1 시간, 2 시간, 3 시간 등으로 설정될 수 있다. 제2 차이가 미리 정의된 제2 시간을 초과하는 경우 데이터 중복 제거 장치는 다음 단계로 단계 S260을 수행할 수 있고, 제2 차이가 미리 정의된 제2 시간 이하인 경우 데이터 중복 제거 장치는 다음 단계로 단계 S270을 수행할 수 있다.

[0069] 단계 S260에서, 데이터 중복 제거 장치는 중복 제거 단위를 데이터에 대한 중복 제거 가능성이 가장 높은 제1 중복 제거 단위로 결정할 수 있다. 즉, 제2 차이가 미리 정의된 제2 시간을 초과하는 경우 이는 자주 사용되지 않는 데이터(즉, 접근 가능성이 낮은 데이터)를 의미하므로, 데이터 중복 제거 장치는 중복 제거 단위 중 데이터 중복 제거율이 가장 높은 제1 중복 제거 단위를 선택할 수 있다.

[0070] 단계 S270에서, 데이터 중복 제거 장치는 데이터에 대한 임의적 접근 횟수가 데이터에 대한 순차적 접근 횟수 미만인지 판단할 수 있다. 임의적 접근 횟수가 순차적 접근 횟수 이상인 경우 데이터 중복 제거 장치는 다음 단계로 S280을 수행할 수 있고, 임의적 접근 횟수가 순차적 접근 횟수 미만인 경우 데이터 중복 제거 장치는 다음

단계로 단계 S290을 수행할 수 있다.

- [0071] 단계 S280에서, 데이터 중복 제거 장치는 중복 제거 단위를 제1 중복 제거 단위보다 중복 제거 가능성이 낮은 제2 중복 제거 단위로 결정할 수 있다. 즉, 데이터에 대한 임의적 접근 횟수가 순차적 접근 횟수 이상인 경우는 임의적 접근이 자주 발생하는 데이터를 의미하므로, 데이터 중복 제거 장치는 중복 제거 단위 중 데이터 중복 제거율이 상대적으로 높은 제2 중복 제거 단위를 선택할 수 있다.
- [0072] 단계 S290에서, 데이터 중복 제거 장치는 중복 제거 단위를 제2 중복 제거 단위보다 중복 제거 가능성이 낮은 제3 중복 제거 단위로 결정할 수 있다. 즉, 데이터에 대한 임의적 접근 횟수가 순차적 접근 횟수 미만인 경우는 순차적 접근이 자주 발생하는 데이터를 의미하므로, 데이터 중복 제거 장치는 중복 제거 단위 중 데이터 중복 제거율이 상대적으로 낮은 제3 중복 제거 단위를 선택할 수 있다.
- [0073] 도 6은 본 발명의 일 실시예에 따른 데이터 중복 제거 방법에 있어서 중복 제거 수행 단계를 도시한 흐름도이고, 도 7은 데이터 블록에 대한 고유의 식별자를 생성하는 과정을 도시한 개념도이다.
- [0074] 이하 도 6 및 도 7을 참조하여, 중복 제거를 수행하는 단계(S300)에 대해 상세하게 설명한다.
- [0075] 본 발명의 일 실시예에 따른 데이터 중복 제거 방법에 있어서 중복 제거를 수행하는 단계(S300)는, 중복 제거 단위를 기반으로 데이터에 대한 적어도 하나의 데이터 블록을 생성하는 단계(S310), 데이터 블록에 대한 고유의 식별자(identifier)를 생성하는 단계(S320), 고유의 식별자가 인덱스 테이블(index table) 내에 존재하는지 판단하는 단계(S330), 고유의 식별자가 인덱스 테이블 내에 존재하지 않는 경우, 고유의 식별자와 고유의 식별자에 대응된 데이터 블록을 저장하는 단계(S340), 및 고유의 식별자가 인덱스 테이블 내에 존재하는 경우, 고유의 식별자에 대응된 데이터 블록을 제거하는 단계(S350)를 포함할 수 있다.
- [0076] 단계 S310에서, 데이터 중복 제거 장치는 중복 제거 단위를 기반으로 데이터에 대한 적어도 하나의 데이터 블록을 생성할 수 있으며, 중복 제거 단위는 청크의 크기를 의미한다. 즉, 데이터 중복 제거 장치는 제1 중복 제거 단위에 대응된 청크의 크기를 기초로 데이터 블록을 생성할 수 있고, 데이터 중복 제거 장치는 제2 중복 제거 단위에 대응된 청크의 크기를 기초로 데이터 블록을 생성할 수 있고, 데이터 중복 제거 장치는 제3 중복 제거 단위에 대응된 청크의 크기를 기초로 데이터 블록을 생성할 수 있고, 데이터 중복 제거 장치는 제4 중복 제거 단위에 대응된 청크의 크기를 기초로 데이터 블록을 생성할 수 있다. 한편, 제4 중복 제거 단위가 데이터 중복 제거를 수행하지 않는 것을 의미하는 경우, 데이터 중복 제거 장치는 데이터에 대한 데이터 블록을 생성하지 않을 수 있다.
- [0077] 상기에서 설명한 단계 S310을 기초로, 데이터 중복 제거 장치는 데이터(30, 도 7)로부터 복수의 데이터 블록(31, 도 7)을 생성할 수 있다.
- [0078] 단계 S320에서, 데이터 중복 제거 장치는 데이터 블록에 대한 고유의 식별자를 생성할 수 있다. 여기서, 데이터 중복 제거 장치는 해시 알고리즘(예를 들어, SHA-1, SHA-2, SHA-3 등)을 사용하여 데이터 블록에 대한 고유의 식별자를 생성할 수 있다. 데이터 블록에 대한 고유의 식별자를 생성하는 방법은 상기 설명에 한정되지 않고, 공지된 다양한 방법을 사용하여 데이터 블록에 대한 고유의 식별자를 생성할 수 있다.
- [0079] 상기에서 설명한 단계 S320을 기초로, 데이터 중복 제거 장치는 각각의 데이터 블록(31, 도 7)으로부터 고유의 식별자(32, 도 7)를 생성할 수 있다.
- [0080] 단계 S330에서, 데이터 중복 제거 장치는 고유의 식별자가 인덱스 테이블 내에 존재하는지 판단할 수 있다. 인덱스 테이블은 고유의 식별자와 고유의 식별자에 대응된 데이터 블록을 포함할 수 있다. 여기서, 고유의 식별자가 인덱스 테이블 내에 존재하는 경우 이는 고유의 식별자에 대응된 데이터 블록이 이미 저장되어 있음을 나타내고, 고유의 식별자가 인덱스 테이블 내에 존재하지 않는 경우 이는 고유의 식별자에 대응된 데이터 블록이 저장되어 있지 않음을 나타낸다. 고유의 식별자가 인덱스 테이블 내에 존재하지 않는 경우 데이터 중복 제거 장치는 다음 단계로 단계 S340을 수행할 수 있고, 고유의 식별자가 인덱스 테이블 내에 존재하는 경우 데이터 중복 제거 장치는 다음 단계로 단계 S350을 수행할 수 있다.
- [0081] 단계 S340에서, 데이터 중복 제거 장치는 고유의 식별자와 고유의 식별자에 대응된 데이터 블록을 저장할 수 있다. 즉, 고유의 식별자에 대응된 데이터 블록이 저장되어 있지 않은 상태이므로, 데이터 중복 제거 장치는 중복 제거를 수행하지 않고 고유의 식별자와 데이터 블록을 데이터베이스(또는, 인덱스 테이블)에 저장할 수 있다.

- [0082] 단계 S350에서, 데이터 중복 제거 장치는 고유의 식별자에 대응된 데이터 블록을 제거할 수 있다. 즉, 고유의 식별자에 대응된 데이터 블록이 이미 저장되어 있는 상태이므로, 데이터 중복 제거 장치는 중복 제거(즉, 고유의 식별자에 대응된 데이터 블록 삭제)를 수행할 수 있다.
- [0083] 본 발명의 일 실시예에 따른 데이터 중복 제거 방법들은 다양한 컴퓨터 수단을 통해 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능 매체에 기록되는 프로그램 명령은 본 발명을 위해 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다.
- [0084] 컴퓨터 판독 가능 매체의 예에는 롬(rom), 램(ram), 플래시 메모리(flash memory) 등과 같이 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러(compiler)에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터(interpreter) 등을 사용해서 컴퓨터에 의해 실행될 수 있는 고급 언어 코드를 포함한다. 상술한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 적어도 하나의 소프트웨어 모듈로 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [0085] 도 8은 본 발명의 일 실시예에 따른 데이터 중복 제거 장치를 도시한 블록도이다.
- [0086] 도 8을 참조하면, 본 발명의 일 실시예에 따른 데이터 중복 제거 장치는 처리부(10) 및 저장부(20)를 포함한다.
- [0087] 처리부(10)는 데이터의 입력 요청 또는 출력 요청을 기반으로 데이터에 대한 접근 특성을 획득할 수 있고, 접근 특성을 기반으로 데이터의 중복 제거 단위를 결정할 수 있고, 중복 제거 단위를 기반으로 데이터에 대한 중복 제거를 수행할 수 있다.
- [0088] 여기서, 처리부(10)는 논리적 구성인 접근 특성 획득부(11), 중복 제거 단위 결정부(12), 중복 제거 수행부(13) 및 인덱스 테이블 관리부(14)를 포함할 수 있다. 한편, 처리부(60)는 물리적 구성인 프로세서(processor) 및 메모리(memory)를 포함할 수 있다. 프로세서는 범용의 프로세서(예를 들어, CPU(Central Processing Unit) 및/또는 GPU(Graphics Processing Unit) 등) 또는 데이터 중복 제거 방법의 수행을 위한 전용의 프로세서를 의미할 수 있다. 메모리에는 데이터 중복 제거 방법의 수행을 위한 프로그램 코드(program code)가 저장될 수 있다. 즉, 프로세서는 메모리에 저장된 프로그램 코드를 독출할 수 있고, 독출된 프로그램 코드를 기반으로 데이터 중복 제거 방법의 각 단계를 수행할 수 있다.
- [0089] 여기서, 데이터에 대한 접근 특성은 데이터에 대한 접근 시간(a_time, 도 2 참조), 데이터에 대한 변경 시간(m_time, 도 2 참조), 데이터에 대한 순차적 접근 횟수(seqCount, 도 2 참조), 데이터에 대한 임의적 접근 횟수(randCount, 도 2 참조) 등을 포함할 수 있다. 여기서, 데이터에 대한 접근 특성은 상기 정보 등에 한정되지 않고 데이터의 입력 또는 출력에 대한 특성을 나타낼 수 있는 정보를 포함할 수 있다.
- [0090] 데이터에 대한 접근 시간(a_time)은 데이터의 입력 요청(즉, 데이터의 쓰기 요청) 또는 데이터의 출력 요청(즉, 데이터의 읽기 요청)을 수신한 시간을 의미한다. 데이터에 대한 변경 시간(m_time)은 데이터의 입력 요청을 수신한 시간을 의미한다. 데이터에 대한 순차적 접근 횟수(seqCount)는 현재의 데이터 요청과 이전의 데이터 요청이 연속되는(즉, 현재의 데이터 요청과 이전의 데이터 요청이 동일함) 횟수를 의미하고, 데이터에 대한 임의적 접근 횟수(randCount)는 현재의 데이터 요청과 이전의 데이터 요청이 연속되지 않는(즉, 현재의 데이터 요청과 이전의 데이터 요청이 동일하지 않음) 횟수를 의미한다.
- [0091] 접근 특성을 획득하는 경우, 처리부(10)는 데이터의 입력 요청에 대한 시간 정보를 기반으로 접근 시간 및 변경 시간을 획득할 수 있고, 데이터의 입력 요청의 연속성을 기반으로 순차적 접근 횟수 또는 임의적 접근 횟수를 획득할 수 있다. 또한, 처리부(10)는 데이터의 출력 요청에 대한 시간 정보를 기반으로 접근 시간을 획득할 수 있고, 데이터의 출력 요청의 연속성을 기반으로 순차적 접근 횟수 또는 임의적 접근 횟수를 획득할 수 있다. 여기서, 접근 특성을 획득하는 과정은 처리부(10) 내의 접근 특성 획득부(11)에서 수행될 수 있다.
- [0092] 구체적으로, 처리부(10)는 수신된 요청이 데이터의 입력 요청에 해당하는지 데이터의 출력 요청에 해당하는지 판단할 수 있다. 수신된 요청이 데이터의 입력 요청에 해당하는 경우, 처리부(10)는 데이터의 입력 요청을 수신한 시간을 접근 시간 및 변경 시간으로 획득할 수 있고, 획득한 접근 시간 및 획득한 변경 시간을 저장부(20)에 저장할 수 있다. 더불어, 처리부(10)는 데이터의 입력 요청과 해당 데이터에 대한 이전(以前) 요청이 연속되는

경우 순차적 접근 횟수를 1 증가시킬 수 있고, 데이터의 입력 요청과 해당 데이터에 대한 이전(以前) 요청이 연속되지 않는 경우 임의적 접근 횟수를 1 증가시킬 수 있다.

[0093] 한편, 수신된 요청이 데이터의 출력 요청에 해당하는 경우, 처리부(10)는 데이터의 출력 요청을 수신한 시간을 접근 시간으로 획득할 수 있고, 획득한 접근 시간을 데이터베이스에 저장할 수 있다. 더불어, 처리부(10)는 데이터의 출력 요청과 해당 데이터에 대한 이전(以前) 요청이 연속되는 경우 순차적 접근 횟수를 1 증가시킬 수 있고, 데이터의 입력 요청과 해당 데이터에 대한 이전(以前) 요청이 연속되지 않는 경우 임의적 접근 횟수를 1 증가시킬 수 있다.

[0094] 중복 제거 단위를 결정하는 경우, 처리부(10)는 데이터에 대한 현재 접근 시간과 데이터에 대한 이전(以前) 변경 시간에 대한 제1 차이를 산출할 수 있고, 제1 차이가 미리 정의된 제1 시간 이하인 경우 중복 제거 단위를 데이터에 대한 중복 제거 가능성이 가장 낮은 제4 중복 제거 단위로 결정할 수 있다. 한편, 제1 차이가 미리 정의된 제1 시간을 초과하는 경우, 처리부(10)는 데이터에 대한 현재 접근 시간과 데이터에 대한 이전(以前) 접근 시간에 대한 제2 차이를 산출할 수 있다.

[0095] 여기서, 처리부(10)는 제2 차이가 미리 정의된 제2 시간을 초과하는 경우 중복 제거 단위를 데이터에 대한 중복 제거 가능성이 가장 높은 제1 중복 제거 단위로 결정할 수 있고, 제2 차이가 미리 정의된 제2 시간 이하인 경우 데이터에 대한 임의적 접근 횟수가 순차적 접근 횟수 이상인지 판단할 수 있다.

[0096] 여기서, 처리부(10)는 데이터에 대한 임의적 접근 횟수가 순차적 접근 횟수 이상인 경우 중복 제거 단위를 제1 중복 제거 단위보다 중복 제거 가능성이 낮은 제2 중복 제거 단위로 결정할 수 있고, 데이터에 대한 임의적 접근 횟수가 순차적 접근 횟수 미만인 경우 중복 제거 단위를 제2 중복 제거 단위보다 중복 제거 가능성이 낮은 제3 중복 제거 단위로 결정할 수 있다.

[0097] 상기에서 설명한 중복 제거 단위를 결정하는 과정은 처리부(10) 내의 중복 제거 단위 결정부(12)에서 수행될 수 있다.

[0098] 여기서, 중복 제거 단위는 제1 중복 제거 단위, 제2 중복 제거 단위, 제3 중복 제거 단위, 제4 중복 제거 단위로 분류할 수 있다. 제1 중복 제거 단위는 자주 사용되지 않는 데이터에 적용될 수 있고, 제2 중복 제거 단위는 순차적 접근보다 임의적 접근이 자주 발생하는 데이터에 적용될 수 있고, 제3 중복 제거 단위는 임의적 접근보다 순차적 접근이 자주 발생하는 데이터에 적용될 수 있고, 제4 중복 제거 단위는 입력이 자주 발생하는 데이터에 적용될 수 있다.

[0099] 구체적으로, 처리부(10)는 데이터의 입력 요청 또는 출력 요청으로부터 획득한 현재 접근 시간과, 동일한 데이터에 대한 이전(以前) 변경 시간(즉, 데이터의 이전(以前) 입력 요청으로부터 획득한 변경 시간)의 차이인 제1 차이를 산출할 수 있다. 여기서, 제1 차이는 해당 데이터가 얼마나 자주 변경되는지(즉, 데이터의 입력 요청이 얼마나 자주 발생하는지)를 의미할 수 있다.

[0100] 처리부(10)는 제1 차이가 미리 정의된 제1 시간을 초과하는지 판단할 수 있다. 여기서, 미리 정의된 제1 시간은 데이터 중복 제거를 수행하기 위한 기준이 되는 시간을 의미하며, 사용자의 설정에 따라 다른 값을 가질 수 있다.

[0101] 제1 차이가 미리 정의된 제1 시간 이하인 경우 이는 입력이 자주 발생하는 데이터를 의미하므로, 처리부(10)는 중복 제거 단위 중 데이터 중복 제거율이 가장 낮은(또는, 데이터 중복 제거를 수행하지 않는) 제4 중복 제거 단위를 선택할 수 있다.

[0102] 한편, 제1 차이가 미리 정의된 제1 시간을 초과하는 경우, 처리부(10)는 데이터의 입력 요청 또는 출력 요청으로부터 획득한 현재 접근 시간과, 동일한 데이터에 대한 이전(以前) 접근 시간(즉, 데이터의 이전(以前) 입력 요청 또는 이전(以前) 출력 요청으로부터 획득한 접근 시간)의 차이인 제2 차이를 산출할 수 있다. 여기서, 제2 차이는 해당 데이터에 대한 접근이 얼마나 자주 발생하는지를 의미할 수 있다.

[0103] 처리부(10)는 제2 차이가 미리 정의된 제2 시간 이하인지 판단할 수 있다. 여기서, 미리 정의된 제2 시간은 데이터에 대한 접근이 발생할 가능성이 낮은 데이터를 구별하기 위해 기준이 되는 시간을 의미하며, 사용자의 설정에 따라 다른 값을 가질 수 있다.

[0104] 제2 차이가 미리 정의된 제2 시간을 초과하는 경우 이는 자주 사용되지 않는 데이터(즉, 접근 가능성이 낮은 데이터)를 의미하므로, 처리부(10)는 중복 제거 단위 중 데이터 중복 제거율이 가장 높은 제1 중복 제거 단위를 선택할 수 있다. 한편, 제2 차이가 미리 정의된 제2 시간 이하인 경우, 처리부(10)는 데이터에 대한 임의적 접근

근 횡수가 데이터에 대한 순차적 접근 횡수 미만인지 판단할 수 있다.

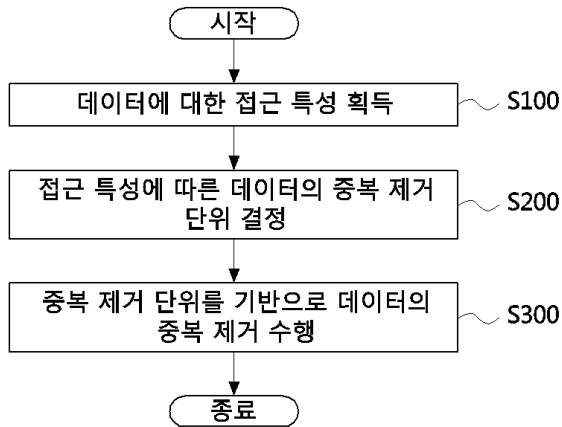
- [0105] 데이터에 대한 임의적 접근 횡수가 순차적 접근 횡수 이상인 경우 이는 임의적 접근이 자주 발생하는 데이터를 의미하므로, 처리부(10)는 중복 제거 단위 중 데이터 중복 제거율이 상대적으로 높은 제2 중복 제거 단위를 선택할 수 있다. 한편, 데이터에 대한 임의적 접근 횡수가 순차적 접근 횡수 미만인 경우는 순차적 접근이 자주 발생하는 데이터를 의미하므로, 처리부(10)는 중복 제거 단위 중 데이터 중복 제거율이 상대적으로 낮은 제3 중복 제거 단위를 선택할 수 있다.
- [0106] 데이터의 중복 제거를 수행하는 경우, 처리부(10)는 중복 제거 단위를 기반으로 데이터에 대한 적어도 하나의 데이터 블록을 생성할 수 있고, 데이터 블록에 대한 고유의 식별자를 생성할 수 있고, 고유의 식별자가 인덱스 테이블 내에 존재하는지 판단할 수 있고, 고유의 식별자가 인덱스 테이블 내에 존재하는 경우 고유의 식별자에 대응된 데이터 블록을 제거할 수 있고, 고유의 식별자가 인덱스 테이블 내에 존재하지 않는 경우 고유의 식별자와 고유의 식별자에 대응된 데이터 블록을 저장할 수 있다.
- [0107] 여기서, 데이터의 중복을 제거하는 과정은 처리부(10) 내의 중복 제거 수행부(13)에서 수행될 수 있고, 인덱스 테이블 내에 정보를 저장, 삭제, 갱신하는 과정은 처리부(10) 내의 인덱스 테이블 관리부(14)에서 수행될 수 있다.
- [0108] 구체적으로, 처리부(10)는 제1 중복 제거 단위에 대응된 청크의 크기를 기초로 데이터 블록을 생성할 수 있고, 또는 제2 중복 제거 단위에 대응된 청크의 크기를 기초로 데이터 블록을 생성할 수 있고, 또는 제3 중복 제거 단위에 대응된 청크의 크기를 기초로 데이터 블록을 생성할 수 있고, 또는 제4 중복 제거 단위에 대응된 청크의 크기를 기초로 데이터 블록을 생성할 수 있다. 한편, 제4 중복 제거 단위가 데이터 중복 제거를 수행하지 않는 것을 의미하는 경우, 처리부(10)는 데이터에 대한 데이터 블록을 생성하지 않을 수 있다.
- [0109] 처리부(10)는 해시 알고리즘(예를 들어, SHA-1, SHA-2, SHA-3 등)을 사용하여 데이터 블록에 대한 고유의 식별자를 생성할 수 있다. 데이터 블록에 대한 고유의 식별자를 생성하는 방법은 상기 설명에 한정되지 않고, 공지된 다양한 방법을 사용하여 데이터 블록에 대한 고유의 식별자를 생성할 수 있다.
- [0110] 처리부(10)는 고유의 식별자가 인덱스 테이블 내에 존재하는지 판단할 수 있다. 고유의 식별자가 인덱스 테이블 내에 존재하지 않는 경우 이는 고유의 식별자에 대응된 데이터 블록이 저장되어 있지 않은 상태이므로, 처리부(10)는 중복 제거를 수행하지 않고 고유의 식별자와 데이터 블록을 저장부(20)에 저장할 수 있다. 한편, 고유의 식별자가 인덱스 테이블 내에 존재하는 경우 이는 고유의 식별자에 대응된 데이터 블록이 이미 저장되어 있는 상태이므로, 처리부(10)는 중복 제거(즉, 고유의 식별자에 대응된 데이터 블록 삭제)를 수행할 수 있다.
- [0111] 저장부(20)는 처리부(10)에서 처리되는 정보 및 처리된 정보를 저장할 수 있다. 예를 들어, 저장부(20)는 데이터의 입력 요청, 데이터의 출력 요청, 데이터에 대한 접근 특성, 제1 차이, 미리 정의된 제1 시간, 제2 차이, 미리 정의된 제2 시간, 인덱스 테이블, 중복 제거 단위 정보 등을 저장할 수 있다.
- [0112] 이상 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

부호의 설명

- [0113] 10: 처리부
- 11: 접근 특성 획득부
- 12: 중복 제거 단위 결정부
- 13: 중복 제거 수행부
- 14: 인덱스 테이블 관리부
- 20: 저장부

도면

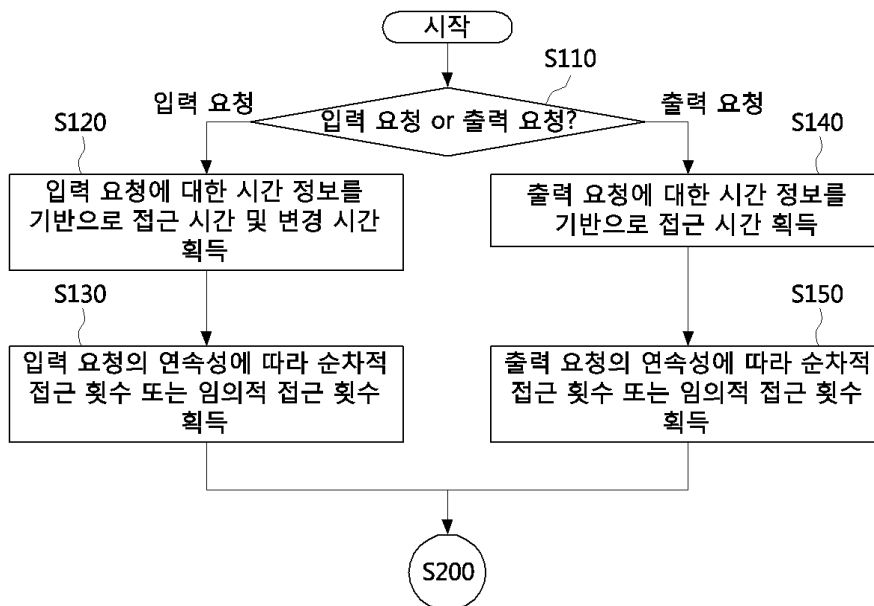
도면1



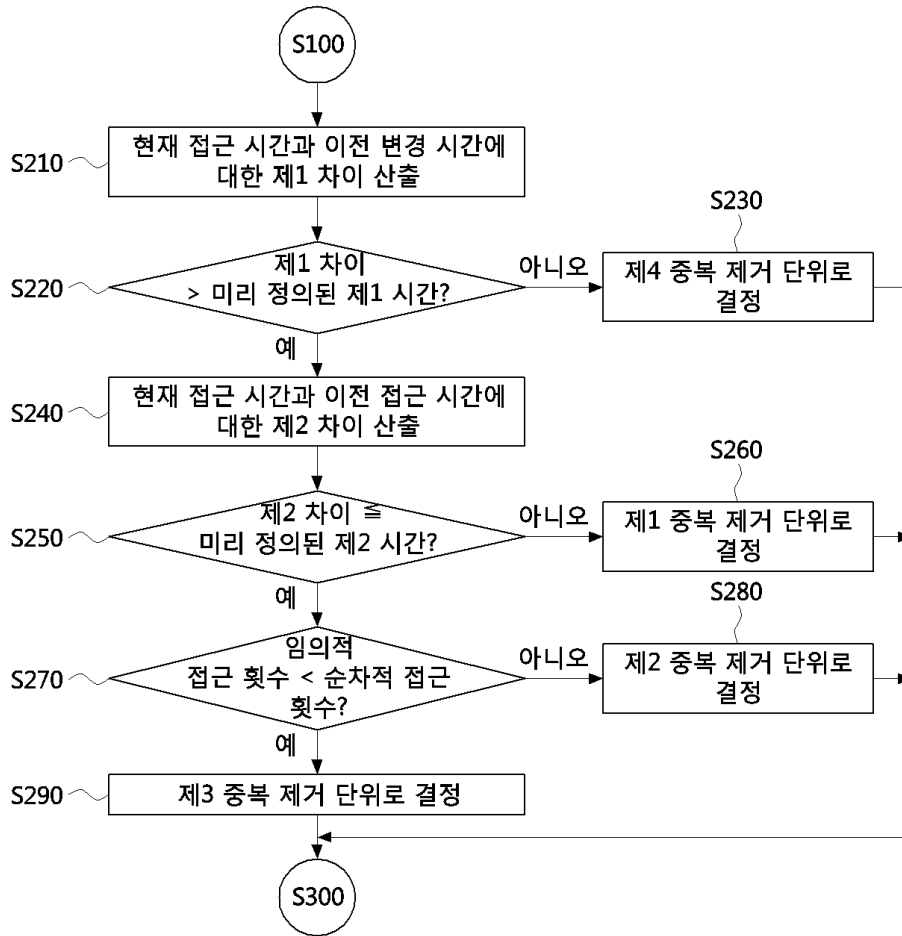
도면2

타입	필드	비고
시간형	m_time	최근 변경 시간
시간형	a_time	최근 접근 시간
정수형	seqCount	순차적 접근 횟수
정수형	randCount	임의적 접근 횟수

도면3



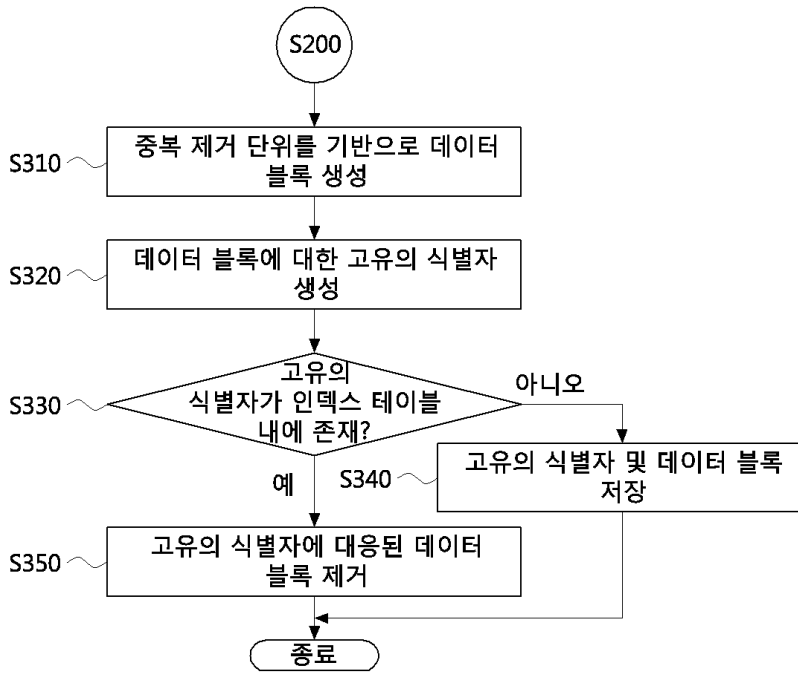
도면4



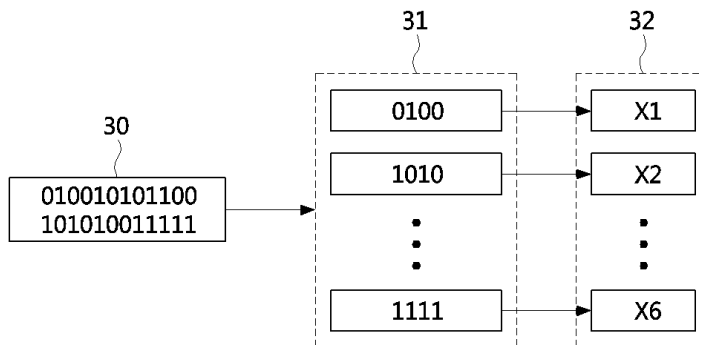
도면5

분류	중복 제거 단위	비고
자주 사용되지 않는 데이터	제1 중복 제거 단위	- 가장 작은 청크 사용 - 가장 큰 중복 제거율
임의적 접근이 자주 발생하는 데이터	제2 중복 제거 단위	- 상대적으로 작은 청크 사용 - 상대적으로 큰 중복 제거율
순차적 접근이 자주 발생하는 데이터	제3 중복 제거 단위	- 상대적으로 큰 청크 사용 - 상대적으로 작은 중복 제거율
입력이 자주 발생하는 데이터	제4 중복 제거 단위	- 중복 제거를 수행하지 않음

도면6



도면7



도면8

