



**(19) 대한민국특허청(KR)**  
**(12) 등록특허공보(B1)**

(45) 공고일자 2013년02월05일  
(11) 등록번호 10-1229851  
(24) 등록일자 2013년01월30일

(51) 국제특허분류(Int. Cl.)  
G06F 17/40 (2006.01) G06F 12/00 (2006.01)  
(21) 출원번호 10-2011-0143176  
(22) 출원일자 2011년12월27일  
심사청구일자 2011년12월27일  
(56) 선행기술조사문헌  
KR1020090129355 A\*  
KR1020110032290 A\*  
KR1020110055263 A\*  
\*는 심사관에 의하여 인용된 문헌

(73) 특허권자  
한국과학기술원  
대전광역시 유성구 대학로 291(구성동)  
(72) 발명자  
박규호  
대전 유성구 구성동 한국과학기술원 6-3208  
김철민  
대전 유성구 구성동 한국과학기술원 전자동 3243호  
(뒷면에 계속)  
(74) 대리인  
김성호

전체 청구항 수 : 총 7 항

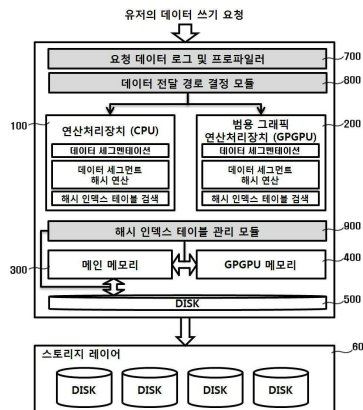
심사관 : 안지현

(54) 발명의 명칭 **데이터의 병렬적 중복제거 시스템**

**(57) 요약**

본 발명의 병렬적 중복제거 시스템은 쓰기 요청된 데이터에 대해서 병렬적으로 중복 제거 연산을 수행할 수 있는 제1처리장치 및 제2처리장치; 및 상기 제1처리장치 또는 상기 제2처리장치를 통해 중복 제거 연산을 수행한 후 중복되지 않은 데이터를 저장할 수 있는 스토리지를 포함한다.

**대표도 - 도2**



(72) 발명자

**김동진**

대전 유성구 구성동 한국과학기술원 전자동 3217호

**박기웅**

서울특별시 노원구 광운로15길 48 (월계동)

---

## 특허청구의 범위

### 청구항 1

쓰기 요청된 데이터에 대해서 병렬적으로 중복 제거 연산을 수행할 수 있는 제1처리장치 및 제2처리장치;

상기 제1처리장치 또는 상기 제2처리장치를 통해 중복 제거 연산을 수행한 후 중복되지 않은 데이터를 저장할 수 있는 스토리지;

해시 인덱스 테이블을 각각 저장할 수 있는 상기 제1처리장치의 메인 메모리, 상기 제2처리장치의 메모리, 및 디스크 스토리지;

상기 쓰기 요청된 데이터의 중복 제거 여부에 관한 정보를 포함하는 데이터 로그; 및

상기 해시 인덱스 테이블을 관리하는 해시 인덱스 테이블 관리모듈을 포함하며,

상기 해시 인덱스 테이블 관리모듈은 상기 데이터 로그 및 상기 쓰기 요청된 데이터의 길이에 근거하여 해시 인덱스를 상기 제1처리장치의 메인 메모리, 상기 제2처리장치의 메모리, 및 상기 디스크 스토리지에 저장하는 데이터의 병렬적 중복제거 시스템.

### 청구항 2

제1항에 있어서,

상기 제1처리장치는 중앙 연산 처리장치이고 상기 제2처리장치는 그래픽 연산 처리장치인 것을 특징으로 하는 데이터의 병렬적 중복제거 시스템.

### 청구항 3

제1항에 있어서,

상기 쓰기 요청된 데이터 중에서 상기 스토리지에 동시에 쓰기될 수 있는 데이터들을 그룹화하고, 상기 그룹화된 데이터들에 대해서 상기 제1처리장치 및 상기 제2처리장치에서 중복 제거 연산을 병렬적으로 수행하는 것을 특징으로 하는 데이터의 병렬적 중복제거 시스템.

### 청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서,

상기 쓰기 요청된 데이터에 대한 중복 제거 연산을 상기 제1처리장치와 상기 제2처리장치 중 어떤 처리장치에서 수행할지를 판단하는 프로파일러를 더 포함하는 것을 특징으로 하는 데이터의 병렬적 중복제거 시스템.

### 청구항 5

제4항에 있어서,

상기 프로파일러는 데이터 스트림의 크기가 소정의 임계값 이상인 경우 상기 제2처리장치에서 중복 제거 연산을 수행하도록 하고, 상기 임계값 미만인 경우 상기 제1처리장치에서 중복 제거 연산을 수행하도록 하는 것을 특징으로 하는 데이터의 병렬적 중복제거 시스템.

### 청구항 6

제5항에 있어서,

상기 임계값은 상기 제1처리장치에서 상기 제2처리장치로의 데이터 전송 지연값, 및 상기 제1처리장치 및 상기 제2처리장치 각각의 사용량에 근거하여 결정되는 것을 특징으로 하는 데이터의 병렬적 중복제거 시스템.

### 청구항 7

삭제

**청구항 8**

삭제

**청구항 9**

삭제

**청구항 10**

제1항에 있어서,

상기 제2처리장치의 메모리에는 상기 제1처리장치의 메인 메모리에 비해 상대적으로 긴 데이터에 대한 해시 인덱스가 저장되는 것을 특징으로 하는 데이터의 병렬적 중복제거 시스템.

**명세서**

**기술분야**

[0001] 본 발명은 데이터의 병렬적 중복제거 시스템에 관한 것입니다.

**배경기술**

[0002] 현재 클라우드 서비스와 같은 대규모 데이터 서비스가 제공되고 있다. 이러한 형태의 서비스는 향후 빠른 속도로 증가될 추세에 있다. 이러한 대규모 데이터 서비스를 위해서는 대용량의 스토리지(storage)를 구비하는 것이 필수적이다. 하지만, 스토리지를 늘이는 것은 비용의 증가를 야기하므로 저비용으로 서비스를 제공하는 것이 어려워진다. 따라서, 용량 대비 스토리지의 이용 효율을 높이는 것이 바람직하다.

[0003] 이와 같은 일환으로 중복된 데이터를 제거함으로써 스토리지의 가용영역을 확대하고자 하는 노력이 있어왔다. 중복된 데이터를 제거함으로써 가용 데이터의 대역폭이 넓어질 수 있다. 하지만, 현재까지는 이러한 중복성 데이터 제거는 일반적으로 백업(backup) 시스템에서 사용되어 왔다. 예컨대, 은행 또는 증권사등에서 데이터 백업용 시스템에서 사용된다. 이때, 중복된 데이터를 제거하기 위해서는 이를 위한 추가의 연산이 요구된다. 하지만, 백업 시스템에 대해 중복 제거 연산을 수행하는 경우 그 속도는 문제되지 않았다.

[0004] 하지만, 데이터의 양이 점점 많아지는 추세에 따라, 중복성 제거가 백업 스토리지뿐만 아니라 일반 스토리지에도 사용될 필요성이 야기되고 있다. 또한, 중복성 제거시에 추가의 컴퓨팅으로 인한 병목 현상을 제거할 수 있는 기술에 대한 필요성이 대두 되고 있다.

**선행기술문헌**

**특허문헌**

[0005] (특허문헌 0001) 한국공개공보 제10-2010-0054093호 (2010.05.24)

**발명의 내용**

**해결하려는 과제**

[0006] 본 발명은 종래의 필요성을 충족시키기 위해 안출된 것으로서, 일반 스토리지에도 적용될 수 있도록 데이터의 중복 제거시의 속도를 향상할 수 있는 데이터의 병렬적 중복제거 시스템을 제공하기 위한 것이다.

[0007] 본 발명이 이루고자 하는 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 본 발명의 기재로부터 당해 분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

**과제의 해결 수단**

[0008] 본 발명에 따른 데이터의 병렬적 중복제거 시스템은 쓰기 요청된 데이터에 대해서 병렬적으로 중복 제거 연산을

수행할 수 있는 제1처리장치 및 제2처리장치; 상기 제1처리장치 또는 상기 제2처리장치를 통해 중복 제거 연산을 수행한 후 중복되지 않은 데이터를 저장할 수 있는 스토리지; 해시 인덱스 테이블을 각각 저장할 수 있는 상기 제1처리장치의 메인 메모리, 상기 제2처리장치의 메모리, 및 디스크 스토리지; 상기 쓰기 요청된 데이터의 중복 제거 여부에 관한 정보를 포함하는 데이터 로그; 및 상기 해시 인덱스 테이블을 관리하는 해시 인덱스 테이블 관리모듈을 포함하며, 상기 해시 인덱스 테이블 관리모듈은 상기 데이터 로그 및 상기 쓰기 요청된 데이터의 길이에 근거하여 해시 인덱스를 상기 제1처리장치의 메인 메모리, 상기 제2처리장치의 메모리, 및 상기 디스크 스토리지에 저장한다.

[0009] 제1처리장치는 중앙 연산 처리장치이고 제2처리장치는 그래픽 연산 처리장치일 수 있다.

**발명의 효과**

[0010] 본 발명에 따르면 데이터의 중복제거 속도를 향상시켜 일반 스토리지 시스템에도 적용할 수 있는 데이터의 병렬적 중복제거 시스템을 제공할 수 있다. 본 발명에 따르면 그래픽 연산 처리 장치를 이용하여 전송한 데이터의 병렬적 중복제거 시스템을 제공할 수 있다. 또한, 본 발명에 따르면 시스템의 상황에 따라 중복제거 연산을 수행하는 처리장치가 변경되도록 함으로써 시스템 효율을 높일 수 있다. 또한, 본 발명에 따르면 중복 데이터를 제거함으로써 가용 데이터 대역폭을 늘릴 수 있다.

**도면의 간단한 설명**

- [0011] 도1은 본 발명의 실시예에 따른 데이터의 병렬적 중복제거 시스템의 하드웨어 구조를 예시한다.
- 도2는 본 발명의 실시예에 따른 데이터의 병렬적 중복제거 시스템의 소프트웨어 구조를 예시한다.
- 도3은 본 발명의 실시예에 따른 프로파일러의 데이터 전달경로 결정 과정을 예시한다.
- 도4는 데이터의 크기에 따른 중앙 연산 처리장치와 그래픽 연산 처리장치의 데이터 처리 성능 그래프이다.
- 도5는 본 발명의 실시예에 따라 그룹화를 사용하는 경우와 사용하지 않는 경우의 연산 과정을 비교한다.

**발명을 실시하기 위한 구체적인 내용**

[0012] 이하, 본 발명의 바람직한 실시예의 상세한 설명이 첨부된 도면들을 참조하여 설명된다. 그러나, 본 발명의 실시형태는 여러 가지의 다른 형태로 변형될 수 있으며, 본 발명의 범위가 이하 설명하는 실시형태로만 한정되는 것은 아니다. 도면에서의 요소들의 형상 및 크기 등은 보다 명확한 설명을 위해 과장될 수 있으며, 도면들 중 인용부호들 및 동일한 구성요소들에 대해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 인용부호들로 표시됨을 유의해야 한다. 참고로 본 발명을 설명함에 있어서 관련된 공지 기능 혹은 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.

[0013] 도1은 본 발명의 실시예에 따른 데이터의 병렬적 중복제거 시스템의 하드웨어 구조를 예시한다. 본 발명의 실시예에 따른 데이터의 병렬적 중복제거 시스템은 사용자의 데이터 쓰기 요청에 따라 병렬적으로 중복 제거 연산을 수행할 수 있는 제1처리장치(100) 및 제2처리장치(200)를 포함한다. 도1 및 이하의 설명에서는 제1처리장치(100)로서 중앙 연산 처리장치(CPU: Central Processing Unit)이 예시되고 제2처리장치(200)로는 그래픽 연산 처리장치(GPGPU: Genaral Purpose Graphic Processing unit)이 예로서 사용된다.

[0014] 이때, 제1처리장치(100)의 메모리로서 메인 메모리(300) 및 제2처리장치(200)의 메모리로서 GPU 메모리(400)를 더 포함할 수 있다. 이때, 제1처리장치(100)는 메인 메모리(300)로 자유롭게 접근하여 데이터를 얻을 수 있으나 GPU 메모리(400)로는 직접 접근하여 데이터를 얻을 수 없다. 제1처리장치(100)는 다만 메인 메모리(300)를 통해 GPU 메모리(400)에 접근할 수 있다. 마찬가지로 제2처리장치(200)는 GPU 메모리(400)에는 자유롭게 접근하여 데이터를 얻을 수 있으나 메인 메모리(300)로는 직접 접근하여 데이터를 얻을 수 없다. 따라서, 제1처리장치(100)가 GPU 메모리(400)에 접근하여 데이터를 사용하기 위해서는 동일 데이터를 메인 메모리(300)로 전송하는 과정을 거쳐야 한다.

[0015] 본 발명의 실시예에 따른 데이터의 병렬적 중복제거 시스템은 제1처리장치(100) 또는 제2처리장치(200)를 통해 중복 제거 연산을 수행한 후 중복되지 않은 데이터를 저장할 수 있는 스토리지(600)를 포함할 수 있다. 스토리지(600)와는 별개로 제1처리장치(100) 및 제2처리장치(200)의 연산을 도울 수 있는 디스크 스토리지(500)를 더 포함할 수 있다.

- [0016] 도2는 본 발명의 실시예에 따른 데이터의 병렬적 중복제거 시스템의 소프트웨어 구조를 예시한다. 제1처리장치(100) 및 제2처리장치(200) 각각은 데이터의 중복 제거 연산을 수행할 수 있다. 도2에서는 중복 제거 연산으로서 해시 연산을 사용하는 것이 예시된다. 하지만 이는 단지 예시일 뿐이면 다른 종류의 중복 제거 연산을 사용하는 것도 가능하다.
- [0017] 예컨대, 해시 연산을 이용한 중복 제거 연산은 아래와 같은 순서로 수행될 수 있다: 1) 쓰기 요청된 데이터를 일정 단위로 자르는 데이터 세그멘테이션(data segmentation)을 수행한다. 2) 잘린 데이터 세그먼트 별로 해시 연산(hash operation)을 수행한다. 3) 해시 인덱스 테이블(hash index table)을 검색하여 상기 2) 단계에 따라 획득한 해시값이 이미 해시 인덱스 테이블에 존재하는지를 판단한다. 해시값이 기존 해시 인덱스 테이블에 존재한다면 이는 중복된 데이터임을 의미한다. 따라서, 해시값 비교 결과, 중복된 데이터가 아닌 경우에만 스토리지(600)에 해당 데이터를 저장할 수 있다.
- [0018] 이와 같이 해시값을 비교함으로써 실제 데이터를 직접 비교할 시에 발생할 수 있는 오버헤드(overhead)를 줄일 수 있다. 해시값 비교를 위해 이전의 요청 데이터에 대한 해시값들로 구성된 해시 인덱스 테이블이 존재한다. 이러한 해시 인덱스 테이블은 메인 메모리(300), GPU 메모리(400) 및 디스크 스토리지(500)에도 일부 저장될 수 있다. 이때, 메인 메모리(300)와 디스크 스토리지(500) 사이에는 마이그레이션 정책(migration policy)이 존재할 수 있다.
- [0019] 본 발명의 실시예에 따른 데이터의 병렬적 중복제거 시스템에서 소프트웨어 모듈로서 데이터 로그 및 프로파일러(700)를 포함할 수 있다. 여기서, 프로파일러는 쓰기 요청된 데이터에 대한 중복 제거 연산을 제1처리장치(100)와 제2처리장치(200) 중 어떤 처리장치에서 수행할지를 판단한다.
- [0020] 예컨대, 프로파일러(700)는 데이터 스트림의 크기가 소정의 임계값 이상인 경우 상기 제2처리장치에서 중복 제거 연산을 수행하도록 하고, 임계값 미만인 경우 제1처리장치에서 중복 제거 연산을 수행하도록 판단할 수 있다. 연산 속도는 제2처리장치(200)인 그래픽 연산 처리장치가 더 빠르기 때문에 데이터 스트림의 크기가 상대적으로 긴 데이터에 대해서는 제2처리장치(200)에서 중복 제거 연산을 수행하도록 GPU 루틴을 택한다. 또한, 데이터 스트림의 크기가 임계값보다 작은 경우에는 제1처리장치(100)인 중앙 연산 처리장치에서 중복 제거 연산을 수행하도록 CPU 루틴을 택할 수 있다. 이는 제2처리장치(200)에서 중복 제거 연산시 제1처리장치(100)에서 제2처리장치(200)로 해당 데이터를 전송해야 하는데, 데이터 크기가 작은 경우 그 전송으로 인한 손해가 제2처리장치(200)의 고속 연산으로 인한 이득보다 커지기 때문이다.
- [0021] 상기 임계값(threshold)은 제1처리장치(100)에서 제2처리장치(200)로의 데이터 전송 지연값을 근거로 하여 결정될 수 있다. 이때, 제1처리장치(100) 및 제2처리장치(200) 각각의 사용량을 고려할 수 있다. 예컨대, 제1처리장치(100)의 사용량이 많을 경우 임계값을 감소시키고 제2처리장치(200)의 사용량이 많을 경우 임계값을 증가시킬 수 있다.
- [0022] 예컨대, 아래의 수학적식1에 의해서 임계값(TS)를 결정할 수 있다.
- [0023] 
$$TS = TS_0 - \alpha \cdot Utilization_{CPU} + \beta \cdot Utilization_{GPU}$$
 수학적식1
- [0024] 여기서,  $TS_0$ : 기본 사용량에서의 임계값이고,  $\alpha$ : CPU 사용량에 따른 감소계수, 그리고  $\beta$ : GPU 사용량에 따른 증가계수를 나타낸다.
- [0025] 도3은 본 발명의 실시예에 따른 프로파일러의 데이터 전달경로 결정 과정을 예시한다. 즉, 데이터 스트림에 대한 쓰기 요청이 들어오고(810), 데이터 전송 지연값, 및 해당 시간에 제1처리장치(100)와 제2처리장치(200)의 사용량을 근거로 임계값이 결정된다(820). 이후 데이터 스트림의 크기가 임계값보다 큰지 작은지(830)에 따라 해당 데이터 스트림에 대한 중복 제거 연산은 GPU 루틴(840)을 따르거나 CPU 루틴(830)을 따르는 것으로 판단될 수 있다.
- [0026] 본 발명의 실시예에 따른 시스템은, 이와 같은 판단에 따라 실제로 데이터를 제1처리장치(100) 또는 제2처리장치(200)로 보내는 데이터 전달 경로 결정 모듈(800)을 더 포함할 수 있다. 도2에서는 데이터 전달 경로 결정 모듈(800)이 프로파일러(700)와는 별개의 모듈로 도시되고 있으나 상기 데이터 전달 경로 결정 모듈(800)은 프로파일러(700)에 통합되어 구현되는 것도 가능하다.
- [0027] 전술한 바와 같은 프로파일러(700)의 판단의 근거가 될 수 있는 데이터의 크기에 따른 중앙 연산 처리장치와 그래픽 연산 처리장치의 데이터 처리 성능 그래프가 도4에 도시된다. 데이터 크기가  $TS_0$ 보다 작을 때에는 메모리



전송 지연이 포함된 그래픽 연산 처리장치의 데이터 처리 성능보다 중앙 연산 처리장치의 데이터 처리 성능이 높다. 데이터 크기가 TS<sub>0</sub>보다 클 때는 메모리 전송 지연이 포함된 그래픽 연산 처리장치의 데이터 처리 성능이 중앙 연산 처리장치의 데이터 처리 성능보다 높은 것을 확인할 수 있다. 이때, 그래픽 연산 처리장치와 중앙 연산 처리장치의 사용량에 따라서 임계값은 TS<sub>0</sub>보다 크거나 작은 값을 가질 수 있다.

- [0028] 쓰기 요청된 데이터는 그룹화를 통해 제1처리장치(100)와 제2처리장치(200)에서 병렬적으로 중복 제거 연산이 수행될 수 있다. 예컨대, 쓰기 요청된 데이터 중에서 스토리지(600)에 동시에 쓰기될 수 있는 데이터들을 그룹화하고, 상기 그룹화된 데이터들에 대해서 제1처리장치(100) 및 제2처리장치(200)에서 중복 제거 연산을 병렬적으로 수행할 수 있다. 이때, 이와 같이 그룹화된 데이터 이외의 데이터에 대해서는 이전에 결정된 중복 제거 연산 위치는 변하지 않는다. 이러한 데이터 그룹화는 중복 제거 연산이 병렬적으로 수행된 경우라도 스토리지(600)에 저장하는 과정에서 발생하는 지연을 줄일 수 있다.
- [0029] 도5는 본 발명의 실시예에 따라 그룹화를 사용하는 경우와 사용하지 않는 경우의 연산 과정을 예시한다.
- [0030] 데이터 스트림 A, B 및 C에 대해서 쓰기 요청이 순서대로 들어오는 경우를 예로 들 수 있다. 데이터 A는 우측 4개의 저장공간을 사용하고, 데이터 B는 오른쪽 6개의 저장공간을 사용하고, 그리고 데이터 C는 좌측 2개의 저장공간을 사용한다. 이때, 데이터 A와 데이터 C는 동시에 병렬적으로 스토리지(600)에 쓰기가 이루어질 수 있다. 따라서, 본 발명의 실시예에서는 데이터A와 데이터C를 그룹화할 수 있다.
- [0031] 이렇게 그룹화된 데이터A와 C에 대해서는 각각 제2처리장치(200)와 제1처리장치(100)에서 동시에 병렬적으로 중복 제거 연산이 수행되도록 할 수 있다. 이후, 본 발명의 실시예에 따르면 중복 제거 연산이 병렬적으로 수행된 데이터 A 및 C에 대해서 동시에 스토리지(600)에 쓰기가 이루어질 수 있다. 이에 따라 그룹화를 사용하지 않는 경우에 비해서 쓰기 시간이 단축되는 것을 도5로부터 알 수 있다. 즉, 본 발명의 실시예에 따른 그룹화를 이용하는 경우 같은 양의 데이터를 스토리지(600)에 저장할 때 더 적은 시간이 소모될 수 있다.
- [0032] 본 발명의 실시예에 따른 데이터의 병렬적 중복 제거 시스템은 해시 인덱스 테이블 관리 모듈(900)을 더 포함할 수 있다. 해시 인덱스 테이블 관리 모듈(900)은 중복 제거 연산시에 효율적으로 중복 여부를 판단할 수 있도록 해시 인덱스를 관리할 수 있다.
- [0033] 해시 인덱스 테이블은 예컨대 메인 메모리(300), GPU 메모리(400) 및 디스크 스토리지(500)에 각각 저장되어 관리될 수 있다. 이때, 해시 인덱스 테이블 관리 모듈(900)은 중복 제거 연산의 효율을 높일 수 있도록 해시 인덱스 테이블을 적절히 메인 메모리(300), GPU 메모리(400) 및 디스크 스토리지(500)에 저장할 수 있다.
- [0034] 예컨대, 디스크 스토리지(500)는 메인 메모리(300) 및 GPU 메모리(400)에 비해서 속도가 떨어지는 반면 용량이 크기 때문에 큰 인덱스 테이블을 저장할 수 있다. 이때, 메모리 속도의 차이 때문에, 메인 메모리(300) 또는 GPU 메모리(400)에 보관하는 해시 인덱스 테이블과 디스크 스토리지(500)에 보관하는 해시 인덱스 테이블 사이에는 중요도의 차이가 있을 수 있다.
- [0035] 본 발명의 실시예에 따른 데이터 로그 및 프로파일러(700) 모듈에서 데이터 로그는 쓰기 요청된 데이터의 중복 제거 여부에 관한 정보를 포함한다. 이때, 자주 나타나는 중복 데이터들의 인덱스는 메인 메모리(300) 또는 GPU 메모리(400)에 저장될 수 있다. 그렇지 않은 데이터들의 인덱스는 디스크 스토리지(500)에 저장될 수 있다. 즉, 제1처리장치(100)의 메인 메모리 및 제2처리장치(200)의 메모리에는 디스크 스토리지(500)에 비해 상대적으로 자주 출현하는 중복 데이터에 대한 해시 인덱스가 저장될 수 있다.
- [0036] 이때, 제2처리장치(200)의 메모리에는 제1처리장치(100)의 메인 메모리에 비해 상대적으로 긴 데이터에 대한 해시 인덱스가 저장될 수 있고, 그 이외의 인덱스들은 메인 메모리(300)에 저장될 수 있다. 이는 본 발명의 실시예에 따른 프로파일러에서 상대적으로 긴 데이터에 대해서 제2처리장치(200)에서 중복 제거 연산을 수행하도록 하는 것보다 상통한다. GPU 메모리(400)와 메인 메모리(300) 사이의 인덱스 테이블은 서로 중복되거나 서로 포함 관계에 있을 수 있다.
- [0037] 이상에서 살펴본 바와 같이, 본 발명의 실시예에 따르면 중앙 연산 처리장치 및 그래픽 연산 처리장치 모두에서 병렬적으로 데이터의 중복 제거 연산을 처리하도록 함으로써 중복 제거 연산에 따른 오버 헤드를 줄일 수 있다. 즉, 중복 제거 연산의 속도가 크게 향상될 수 있다. 또한, 효율적인 인덱스 테이블의 관리 및/또는 데이터 그룹화를 통해서 중복 제거 연산의 속도 및/또는 쓰기 속도를 향상시킬 수 있다. 이에 따라 오버 헤드 없이 일반 스토리지 시스템에서도 중복 제거 연산을 수행할 수 있다.
- [0038] 이상 첨부된 도면을 참조하여 본 발명의 실시예를 설명하였지만, 본 발명이 속하는 기술분야의 당업자는 본 발

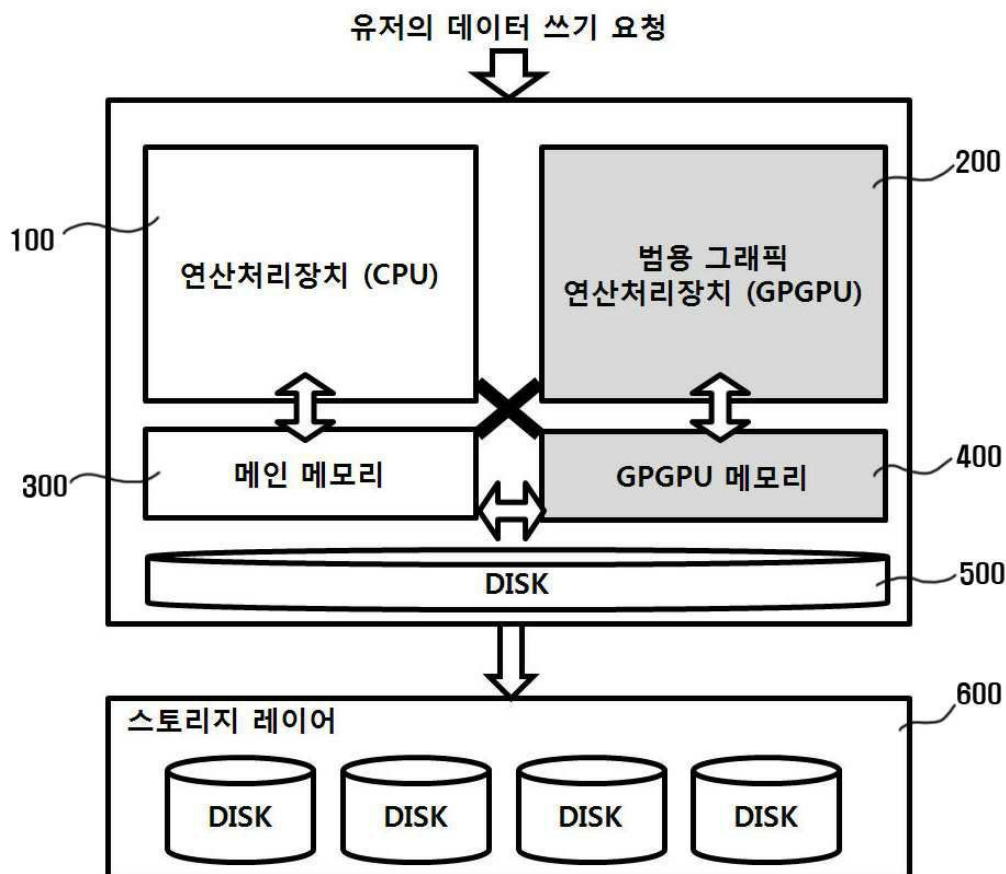
명이 그 기술적 사상이나 필수적 특징을 변경하지 않고 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적인 것이 아닌 것으로서 이해되어야 하고, 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 등가개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

**부호의 설명**

- 100: 제1처리장치(중앙 연산 처리장치)
- 200: 제2처리장치(그래픽 연산 처리장치)
- 300: 메인 메모리
- 400: 그래픽 연산 처리장치의 메모리
- 500: 디스크 스토리지
- 600: 스토리지
- 700: 데이터 로그 및 프로파일러
- 800: 데이터 전달 경로 결정 모듈
- 900: 해시 인덱스 테이블 관리 모듈

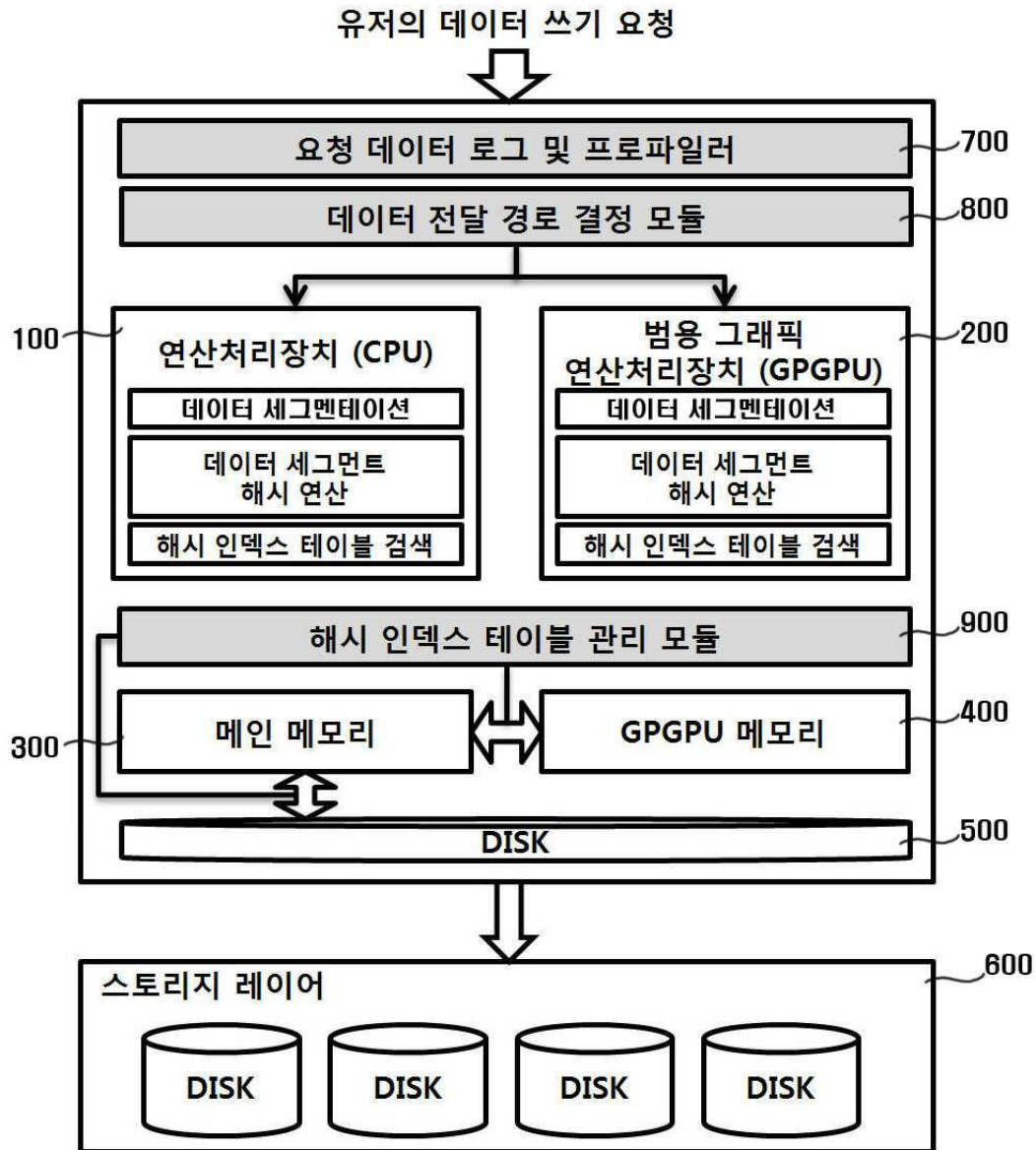
**도면**

**도면1**

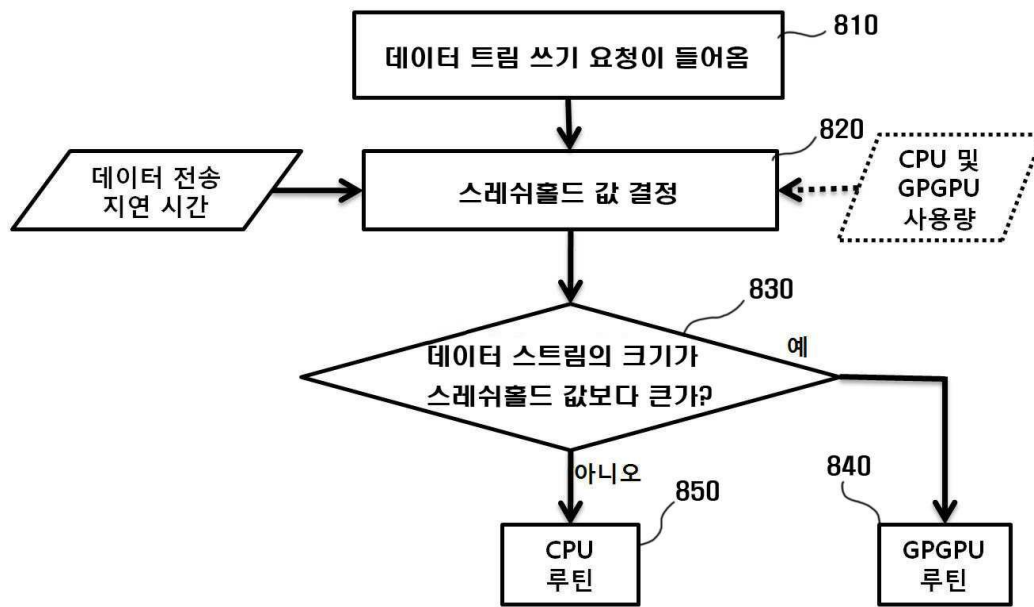




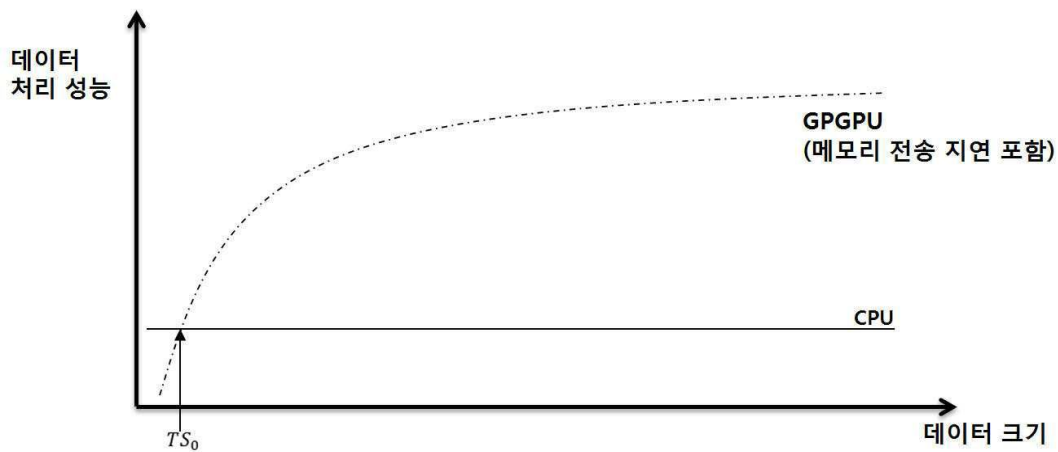
도면2



도면3



도면4



도면5

