



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0160470
(43) 공개일자 2024년11월11일

(51) 국제특허분류(Int. Cl.)
G06F 3/06 (2006.01)
(52) CPC특허분류
G06F 3/0659 (2013.01)
G06F 3/0604 (2013.01)
(21) 출원번호 10-2023-0057336
(22) 출원일자 2023년05월02일
심사청구일자 없음

(71) 출원인
에스케이하이닉스 주식회사
경기도 이천시 부발읍 경충대로 2091
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
이승진
서울특별시 마포구 독막로 227-8 (신수동)
이창규
서울특별시 마포구 마포대로 73 (도화동) 813
(뒷면에 계속)
(74) 대리인
김선중

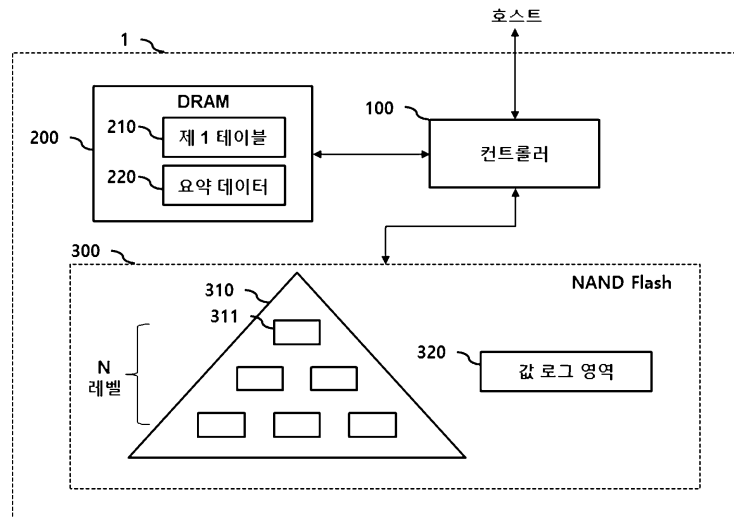
전체 청구항 수 : 총 16 항

(54) 발명의 명칭 키-값 기반 데이터 저장 장치 및 그 동작 방법

(57) 요약

본 기술에 의한 키-값 기반 데이터 저장 장치는 제 1 테이블 및 요약 데이터를 저장하는 휘발성 메모리; LSM 트리 구조를 형성하는 다수의 제 2 테이블을 저장하는 LSM 트리 영역과 키에 대응하는 값을 저장하는 값 로그 영역을 포함하는 비휘발성 메모리; 및 하나 또는 둘 이상의 영역 질의 명령을 처리하는 컨트롤러를 포함하되, 요약 데이터는 제 1 테이블과 다수의 제 2 테이블에 대한 현재 상태를 나타내는 글로벌 버전을 포함하는 버전 데이터를 포함하고, 컨트롤러는 어느 하나의 영역 질의 명령을 처리하기 시작할 때 글로벌 버전을 복제된 복제 버전을 버전 데이터에 추가하고 어느 하나의 영역 질의 명령의 처리 과정에서 복제 버전을 참조하는 제어 회로를 포함한다.

대표도 - 도1



(52) CPC특허분류

G06F 3/0658 (2013.01)

G06F 2212/1016 (2013.01)

(72) 발명자

김영재

서울특별시 마포구 백범로 35 AS911

박인혁

경기도 이천시 부발읍 경충대로 2091

정우석

경기도 이천시 부발읍 경충대로 2091

명세서

청구범위

청구항 1

제 1 테이블 및 요약 데이터를 저장하는 휘발성 메모리;

LSM 트리 구조를 형성하는 다수의 제 2 테이블을 저장하는 LSM 트리 영역과 키에 대응하는 값을 저장하는 값 로그 영역을 포함하는 비휘발성 메모리; 및

하나 또는 둘 이상의 영역 질의 명령을 처리하는 컨트롤러

를 포함하되,

상기 요약 데이터는 상기 제 1 테이블과 상기 다수의 제 2 테이블에 대한 현재 상태를 나타내는 글로벌 버전을 포함하는 버전 데이터를 포함하고,

상기 컨트롤러는 어느 하나의 영역 질의 명령을 처리하기 시작할 때 상기 글로벌 버전을 복제한 복제 버전을 상기 버전 데이터에 추가하고 상기 어느 하나의 영역 질의 명령의 처리 과정에서 상기 복제 버전을 참조하는 제어 회로

를 포함하는 키-값 기반 데이터 저장 장치.

청구항 2

청구항 1에 있어서, 상기 제어 회로는 상기 어느 하나의 영역 질의 명령의 처리가 종료되는 경우 상기 복제 버전을 상기 버전 데이터에서 삭제하는 키-값 기반 데이터 저장 장치.

청구항 3

청구항 1에 있어서, 상기 글로벌 버전은 상기 제 1 테이블에 대응하는 제 1 테이블 정보와 상기 다수의 제 2 테이블에 대응하는 다수의 제 2 테이블 정보를 포함하되, 제 2 테이블 정보는 대응하는 제 2 테이블의 주소 및 키 범위 정보를 포함하는 키-값 기반 데이터 저장 장치.

청구항 4

청구항 1에 있어서, 상기 요약 데이터는 상기 다수의 제 2 테이블에 대응하는 다수의 제 2 테이블 정보를 포함하는 메타데이터 풀을 더 포함하고,

상기 글로벌 버전은 상기 다수의 제 2 테이블에 대응하여 상기 메타데이터 풀에 포함된 제 2 테이블 정보에 대한 포인터 정보를 포함하는 키-값 기반 데이터 저장 장치.

청구항 5

청구항 4에 있어서, 상기 다수의 제 2 테이블 정보는 각각 레퍼런스 카운터를 포함하되, 상기 제어 회로는 상기 어느 하나의 영역 질의 명령의 처리가 시작되는 경우 대응하는 복제 버전에 연관된 다수의 제 2 테이블 정보에서 레퍼런스 카운터를 증가시키고 상기 어느 하나의 질의 명령의 처리가 종료되는 경우 상기 대응하는 복제 버전에 연관된 다수의 제 2 테이블 정보에서 레퍼런스 카운터를 감소시키는 키-값 기반 데이터 저장 장치.

청구항 6

청구항 5에 있어서, 상기 LSM 트리 영역에서 어느 하나의 제 2 테이블이 삭제되는 경우 상기 제어 회로는 상기 메타데이터 풀에서 대응하는 제 2 테이블 정보의 레퍼런스 카운터를 참조하여 상기 대응하는 제 2 테이블 정보의 삭제 여부를 결정하는 키-값 기반 데이터 저장 장치.

청구항 7

청구항 6에 있어서, 상기 제어 회로는 상기 글로벌 버전을 생성하기 위하여 생성 시점에 상기 LSM 트리 영역에

포함된 다수의 제 2 테이블에 대응하는 다수의 제 2 테이블 정보를 상기 메타데이터 풀에 추가하고, 상기 다수의 제 2 테이블 정보에 포함된 레퍼런스 카운터의 값을 1로 설정하는 키-값 기반 데이터 저장 장치.

청구항 8

청구항 7에 있어서, 상기 제어 회로는 상기 LSM 트리 영역에 제 2 테이블을 추가하는 경우 레퍼런스 카운터 값을 1로 설정하여 대응하는 제 2 테이블 정보를 상기 메타데이터 풀에 추가하고, 상기 LSM 트리 영역에서 제 2 테이블을 삭제하는 경우 상기 메타데이터 풀에서 대응하는 제 2 테이블 정보의 레퍼런스 카운터의 값을 1 감소시키는 키-값 기반 데이터 저장 장치.

청구항 9

청구항 8에 있어서, 상기 제어 회로는 상기 메타데이터 풀에서 레퍼런스 카운터의 값이 0인 제 2 테이블 정보를 삭제하는 키-값 기반 데이터 저장 장치.

청구항 10

제 1 테이블 및 요약 데이터를 포함하는 휘발성 메모리 및 LSM 트리 구조를 이루는 다수의 제 2 테이블을 포함하는 LSM 트리 영역과 값 로그 영역을 포함하는 비휘발성 메모리를 포함하는 키-값 기반 데이터 저장 장치의 동작 방법으로서,

상기 제 1 테이블과 상기 다수의 제 2 테이블에 대한 현재 상태를 나타내는 글로벌 버전을 상기 요약 데이터에 추가하는 단계;

어느 하나의 영역 질의 명령을 처리하기 시작할 때 상기 글로벌 버전을 복제한 복제 버전을 상기 버전 데이터에 추가하는 단계;

상기 어느 하나의 영역 질의 명령의 처리 과정에서 상기 복제 버전을 참조하는 단계; 및

상기 어느 하나의 영역 질의 명령의 처리가 종료되는 경우 상기 복제 버전을 상기 버전 데이터에서 삭제하는 단계

를 포함하는 키-값 기반 데이터 저장 장치의 동작 방법.

청구항 11

청구항 10에 있어서, 상기 요약 데이터는 상기 다수의 제 2 테이블에 대응하는 다수의 제 2 테이블 정보를 포함하는 메타데이터 풀을 더 포함하고,

상기 글로벌 버전은 상기 다수의 제 2 테이블에 대응하여 상기 메타데이터 풀에 포함된 제 2 테이블 정보에 대한 포인터 정보를 포함하는 키-값 기반 데이터 저장 장치의 동작 방법.

청구항 12

청구항 11에 있어서,

상기 다수의 제 2 테이블 정보는 각각 레퍼런스 카운터를 포함하되,

상기 복제 버전을 상기 버전 데이터에 추가하는 단계는 복제 버전에 연관된 다수의 제 2 테이블 정보에서 레퍼런스 카운터를 증가시키는 단계를 포함하고,

상기 복제 버전을 상기 버전 데이터에서 삭제하는 단계는 복제 버전에 연관된 다수의 제 2 테이블 정보에서 레퍼런스 카운터를 감소시키는 단계를 포함하는 키-값 기반 데이터 저장 장치의 동작 방법.

청구항 13

청구항 12에 있어서, 상기 LSM 트리 영역에서 어느 하나의 제 2 테이블이 삭제되는 경우 상기 메타데이터 풀에서 대응하는 제 2 테이블 정보의 레퍼런스 카운터를 참조하여 상기 대응하는 제 2 테이블 정보의 삭제 여부를 결정하는 단계를 더 포함하는 키-값 기반 데이터 저장 장치의 동작 방법.

청구항 14

청구항 13에 있어서, 상기 글로벌 버전을 상기 요약 데이터에 추가하는 단계는 상기 글로벌 버전의 생성 시점에서 상기 LSM 트리 영역에 포함된 다수의 제 2 테이블에 대응하는 다수의 제 2 테이블 정보를 상기 메타데이터 풀에 추가하고, 상기 다수의 제 2 테이블 정보에 포함된 레퍼런스 카운터의 값을 1로 설정하는 단계를 포함하는 키-값 기반 데이터 저장 장치의 동작 방법.

청구항 15

청구항 14에 있어서, 상기 LSM 트리 영역에 제 2 테이블을 추가하는 경우 레퍼런스 카운터 값을 1로 설정하여 대응하는 제 2 테이블 정보를 상기 메타데이터 풀에 추가하는 단계 및, 상기 LSM 트리 영역에서 제 2 테이블을 삭제하는 경우 상기 메타데이터 풀에서 대응하는 제 2 테이블 정보의 레퍼런스 카운터의 값을 1 감소시키는 단계를 더 포함하는 키-값 기반 데이터 저장 장치의 동작 방법.

청구항 16

청구항 15에 있어서, 상기 메타데이터 풀에서 레퍼런스 카운터의 값이 0인 제 2 테이블 정보를 삭제하는 단계를 더 포함하는 키-값 기반 데이터 저장 장치의 동작 방법.

발명의 설명

기술 분야

[0001] 본 기술은 키-값 기반 데이터 저장 장치와 그 동작 방법에 관한 것으로서 보다 구체적으로는 요약 데이터를 이용하여 영역 질의 명령을 처리하는 키-값 기반 데이터 저장 장치와 그 동작 방법에 관한 것이다.

배경 기술

- [0002] 키-값 기반 데이터 저장 장치는 키를 이용하여 값을 처리하는 새로운 형태의 데이터 저장 장치로서 쓰기(Put), 읽기(Get), 삭제>Delete), 스캔(Scan) 동작 등을 수행한다.
- [0003] 영역 질의(Range Query)는 데이터베이스 동작에서 자주 사용되는 동작으로서 일정한 키 범위에 대응하는 다수의 값을 읽는 동작을 수행한다.
- [0004] LSM 트리 구조를 사용하는 키-값 기반 데이터 저장 장치는 요약 테이블을 이용하여 SSTable을 관리한다.
- [0005] 낸드 플래시 메모리 기반의 종래의 키-값 기반 데이터 저장 장치는 영역 질의 명령을 처리하는 도중에 PUT, DELETE 명령을 수행할 수 있는데 이 경우 컴팩션 동작이 수행되어 LSM 트리 구조에 변경이 발생할 수 있다.
- [0006] 종래의 키-값 기반 데이터 저장 장치는 이러한 상황에 대비하여 요약 테이블을 관리하지 못하기 때문에 컴팩션 동작이 수행되기 전을 기준으로 영역 질의 명령을 완료하는 것이 불가능한 문제가 있다.

선행기술문헌

특허문헌

- [0007] (특허문헌 0001) KR 10-2018-0030319 A
- (특허문헌 0002) KR 10-2020-0053512 A
- (특허문헌 0003) US 11,550,504 B2

비특허문헌

- [0008] (비특허문헌 0001) Lee, Chang-Gyu, et al. "iLSM-SSD: An intelligent LSM-tree based key-value SSD for data analytics." 2019 IEEE 27th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS). IEEE, 2019

발명의 내용

해결하려는 과제

- [0009] 본 기술은 영역 질의 명령을 처리하는 도중 LSM 트리 구조에 변경이 생기더라도 영역 질의 명령이 제공된 시점을 기준으로 영역 질의 명령을 완료할 수 있는 키-값 기반의 데이터 저장 장치와 그 동작 방법을 제공한다.

과제의 해결 수단

- [0010] 본 발명의 일 실시예에 의한 키-값 기반 데이터 저장 장치는 제 1 테이블 및 요약 데이터를 저장하는 휘발성 메모리; LSM 트리 구조를 형성하는 다수의 제 2 테이블을 저장하는 LSM 트리 영역과 키에 대응하는 값을 저장하는 값 로그 영역을 포함하는 비휘발성 메모리; 및 하나 또는 둘 이상의 영역 질의 명령을 처리하는 컨트롤러를 포함하되, 요약 데이터는 제 1 테이블과 다수의 제 2 테이블에 대한 현재 상태를 나타내는 글로벌 버전을 포함하는 버전 데이터를 포함하고, 컨트롤러는 어느 하나의 영역 질의 명령을 처리하기 시작할 때 글로벌 버전을 복제한 복제 버전을 버전 데이터에 추가하고 어느 하나의 영역 질의 명령의 처리 과정에서 복제 버전을 참조하는 제어 회로를 포함한다.
- [0011] 본 발명의 일 실시예에 의한 키-값 기반 데이터 저장 장치의 동작 방법은 제 1 테이블 및 요약 데이터를 포함하는 휘발성 메모리 및 LSM 트리 구조를 이루는 다수의 제 2 테이블을 포함하는 LSM 트리 영역과 값 로그 영역을 포함하는 비휘발성 메모리를 포함하는 키-값 기반 데이터 저장 장치의 동작 방법으로서, 제 1 테이블과 다수의 제 2 테이블에 대한 현재 상태를 나타내는 글로벌 버전을 요약 데이터에 추가하는 단계; 어느 하나의 영역 질의 명령을 처리하기 시작할 때 글로벌 버전을 복제한 복제 버전을 상기 버전 데이터에 추가하는 단계; 어느 하나의 영역 질의 명령의 처리 과정에서 복제 버전을 참조하는 단계; 및 어느 하나의 영역 질의 명령의 처리가 종료되는 경우 복제 버전을 버전 데이터에서 삭제하는 단계를 포함한다.

발명의 효과

- [0012] 본 기술을 적용함으로써 영역 질의 명령을 처리하는 도중 LSM 트리 구조에 변경이 생기더라도 영역 질의 명령이 제공된 시점을 기준으로 영역 질의 명령을 완료할 수 있다.

도면의 간단한 설명

- [0013] 도 1은 본 발명의 일 실시예에 의한 키-값 기반 데이터 저장 장치를 나타내는 블록도.
 도 2는 본 발명의 일 실시예에 의한 컨트롤러를 나타내는 블록도.
 도 3은 본 발명의 일 실시예에 의한 요약 데이터의 구조를 나타내는 설명도.
 도 4는 본 발명의 일 실시예에 의한 복제 버전 생성 방법을 나타내는 설명도.

발명을 실시하기 위한 구체적인 내용

- [0014] 이하에서는 첨부한 도면을 참조하여 본 발명의 실시예를 개시한다.
- [0015] 이하에서는 LSM 트리 구조를 이용하여 키를 관리하고 LSM 트리 영역과는 별도의 값 로그 영역에 키에 대응하는 값을 저장하는 키-값(Key-Value, KV) 방식의 SSD를 예로 들어 본 실시예를 개시한다.
- [0016] LSM 트리 구조를 이용하여 키를 관리하고 값을 별도의 로그 영역에 저장하는 KV SSD 자체는 비특허문헌 1에 개시된 바와 같이 잘 알려진 것이므로 종래 기술에 대한 반복적인 설명은 생략한다.
- [0017] 도 1은 본 발명의 일 실시예에 의한 KV SSD(1)를 나타내는 블록도이다.
- [0018] KV SSD(1)는 컨트롤러(100), 디램(200), 낸드 플래시 메모리(300)를 포함한다.
- [0019] 컨트롤러(100)는 호스트에서 제공된 명령을 수신하여 읽기, 쓰기, 또는 소거 동작을 수행할 수 있다. 이는 비특허문헌 1에 개시되어 있으므로 반복적인 설명은 생략한다.
- [0020] 또한 본 기술에 의한 컨트롤러(100)는 영역 질의 명령을 처리하며 이 과정에서 요약 데이터에 대한 관리 동작을 수행한다.

- [0021] 컨트롤러(100)는 영역 질의 명령을 처리하는 과정에서 디램(200)과 낸드 플래시 메모리(300)를 제어할 수 있다.
- [0022] 디램(200)은 제 1 테이블(210)과 요약 데이터(220)를 저장한다. 제 1 테이블(210)은 MemTable로 지칭될 수 있다. 본 기술에 의한 요약 데이터(220)의 구조와 그 관리 방법에 대해서는 아래에서 구체적으로 개시한다.
- [0023] 낸드 플래시 메모리(300)는 키를 저장하는 LSM 트리 영역(310)과 키에 대응하는 값을 저장하는 값 로그 영역(320)을 포함한다.
- [0024] LSM 트리 영역(310)은 다수의 제 2 테이블(311)을 포함하며 다수의 제 2 테이블은 N 개의 레벨을 가지는 트리 구조로 배치된다(N은 자연수). 제 2 테이블은 SSTable로 지칭될 수 있다.
- [0025] 하나의 제 2 테이블은 다수의 키와 이에 대응하는 옵셋을 저장한다.
- [0026] 값 로그 영역(320)은 키에 대응하는 값을 저장하는 영역으로서 하나의 키에 대응하는 값의 주소는 키에 대응하는 옵셋과 베이스 주소를 이용하여 계산될 수 있다.
- [0027] 도 2는 본 발명의 일 실시예에 의한 컨트롤러(100)를 나타내는 블록도이다.
- [0028] 컨트롤러(100)는 제어 회로(110), 디램 컨트롤러(120), 및 플래시 채널 컨트롤러(130)를 포함한다.
- [0029] 제어 회로(110)는 호스트에서 제공된 영역 질의 명령을 해석하고 이에 따라 지정된 범위에 포함된 다수의 키에 대응하는 다수의 값을 출력하는 동작을 수행한다.
- [0030] 제어 회로(110)는 하드웨어, 소프트웨어, 또는 하드웨어와 소프트웨어의 조합으로 구성될 수 있다.
- [0031] 디램 컨트롤러(120)는 제어 회로(110)의 제어에 따라 디램(200)에 대한 읽기 또는 쓰기 동작을 제어한다.
- [0032] 플래시 채널 컨트롤러(130)는 제어 회로(110)의 제어에 따라 낸드 플래시 메모리(300)에 대한 읽기, 쓰기, 또는 소거 동작을 제어한다.
- [0033] 본 실시예에서 플래시 채널 컨트롤러(130)는 하나 또는 둘 이상의 채널을 개별적으로 제어하며 각각의 채널에는 다수개의 플래시 메모리가 연결된다.
- [0034] 디램 컨트롤러(120)나 플래시 채널 컨트롤러(130) 자체는 종래의 기술이므로 이에 대한 반복적인 설명은 생략한다.
- [0035] 제어 회로(110)는 호스트 인터페이스(111), 키-값 명령 처리부(112), 프리페치 제어부(113), 컨트롤러 인터페이스(114)를 포함한다. 이하에서, 키-값 명령 처리부(112)를 KV 명령 처리부(112)로 지칭할 수 있다.
- [0036] 호스트 인터페이스(111)는 호스트에서 전달된 요청에서 키-값 명령을 추출한다.
- [0037] 예를 들어 호스트는 NVMe 프로토콜에 따라 키-값 명령을 전송할 수 있으며 호스트 인터페이스(111)는 NVMe 프로토콜에 따라 전달된 요청에서 키-값 명령을 추출한다.
- [0038] 키-값 명령 처리부(112)는 키-값 명령인 PUT, GET, DELETE 등의 명령을 처리할 수 있으며 이는 종래의 기술을 통해 잘 알려진 것이므로 이에 대한 반복적인 설명은 생략한다.
- [0039] 또한 본 실시예에서 키-값 명령 처리부(112)는 영역 질의 명령에 따라 컨트롤러 인터페이스(114)를 제어하여 키-값 명령을 처리한다.
- [0040] 영역 질의 명령은 주어진 범위에 속한 키에 대응하는 데이터를 읽기 위한 동작으로서 본 발명에 있어서 영역 질의 명령을 처리하는 방식은 특정한 방식으로 제한되는 것은 아니다.
- [0041] 예를 들어 영역 질의 명령은 영역의 시작 키를 찾는 제 1 명령과 시작 키 이후 순차적으로 다음 키에 대응하는 데이터를 읽는 다수의 제 2 명령을 포함할 수 있다. 제 1 명령을 Seek 명령 또는 Seek로 지칭할 수 있고, 제 2 명령을 Next 명령 또는 Next로 지칭할 수 있다.
- [0042] 본 실시예에서 영역 질의 명령은 영역 질의 명령이 처리되기 시작한 시점을 기준으로 주어진 영역에 대응하는 데이터를 출력하는 방식으로 동작할 것을 전제한다.
- [0043] 즉, 영역 질의 명령에 대한 처리가 완료되기 전에 다른 키-값 명령에 의해 데이터가 갱신되더라도 갱신되기 이전의 데이터를 출력해야 하는 것이다.
- [0044] 디램(200)은 요약 데이터(220)를 저장하는데, 요약 데이터(220)는 LSM 트리 영역(310)에 저장된 모든 제 2 테이블

블에 대한 정보를 저장한다.

- [0045] 키-값 명령 처리부(112)는 영역 질의 명령을 처리하는 도중 요약 데이터(220)를 참조하여 주어진 영역에 포함된 키를 찾을 수 있다.
- [0046] 잘 알려진 바와 같이 LSM 트리 구조를 사용하는 키-값 기반 데이터 저장 장치는 PUT, DELETE 등의 명령을 처리하는 동작 및 컴팩션 동작을 수행함에 따라 LSM 트리 구조에 변경이 생긴다.
- [0047] 요약 데이터 관리부(113)는 영역 질의 명령을 처리하는 과정에서 요약 데이터(220)를 관리한다.
- [0048] 요약 데이터 관리부(113)는 영역 질의 명령의 처리가 시작된 후 처리가 완료되기 이전에 LSM 트리 구조에 변경이 생기더라도 영역 질의 명령이 정상적으로 처리될 수 있도록 요약 데이터를 관리한다.
- [0049] 도 3은 본 발명의 일 실시예에 의한 요약 데이터(220)의 구조를 나타내는 설명도이다.
- [0050] 요약 데이터(220)는 메타데이터 풀(230)과 버전 데이터(240)를 포함한다.
- [0051] 본 실시예에서 메타데이터 풀(230)은 LSM 트리 영역(310)에 포함된 모든 제 2 테이블에 대한 정보를 테이블 형태로 저장하는데 데이터 구조가 이에 한정된 것은 아니다.
- [0052] 본 실시예에서 메타데이터 풀(230)은 ID 필드, 물리주소 필드, 메타정보 필드, 레퍼런스 카운터 필드를 포함한다.
- [0053] ID는 테이블의 행 번호에 대응한다.
- [0054] 물리주소 필드는 LSM 트리 영역(310) 내에서 대응하는 제 2 테이블이 위치한 물리주소를 나타낸다.
- [0055] 메타정보는 제 2 테이블에 관한 다양한 정보를 포함할 수 있다. 예를 들어 LSM 트리 내에서 대응하는 제 2 테이블이 위치한 레벨, 상위 레벨에 위치한 제 2 테이블의 ID, 하위 레벨에 위치한 제 2 테이블의 ID, 해당 제 2 테이블에 저장된 다수의 키 정보 등을 포함한다.
- [0056] 레퍼런스 카운터는 대응하는 제 2 테이블을 참조하는 키-값 명령의 개수를 표시한다.
- [0057] 요약 데이터 관리부(113)는 키-값 명령 수행 동작이나 컴팩션 동작에 의해 LSM 트리 영역(310)에 새로운 제 2 테이블이 생성되거나 기존의 제 2 테이블이 삭제되는 경우 대응하는 행을 메타데이터 풀(230)에 추가하거나 삭제한다.
- [0058] LSM 트리 영역(310)에 제 2 테이블이 추가, 삭제되는 경우 변경된 LSM 트리 구조에 부합하도록 메타데이터 풀(230)에서 대응하는 테이블 행의 정보를 수정하고, 이에 대응하는 이웃하는 다른 테이블 행의 정보를 수정하는 것은 당업자가 용이하게 알 수 있는 것이다.
- [0059] LSM 트리 영역(310)에서 제 2 테이블을 삭제하였을 때 이에 대응하는 행을 메타데이터 풀(230)에서 삭제할지는 레퍼런스 카운터의 값에 따라 달라진다. 이에 대해서는 아래에서 구체적으로 개시한다.
- [0060] 버전 데이터(240)는 하나 또는 둘 이상의 LSM 트리 버전(241)을 포함한다.
- [0061] LSM 트리 버전(241)은 버전 생성 시점에서 제 1 테이블에 대한 정보와 LSM 트리 구조에 포함된 다수의 제 2 테이블에 대한 정보를 포함한다.
- [0062] 이에 따라 LSM 트리 버전(241)은 제 1 테이블 정보(211)와 다수의 제 2 테이블 정보(2411)를 포함한다.
- [0063] 제 1 테이블 정보(211)는 버전 생성 시점에서의 제 1 테이블의 정보를 복제한 것이다.
- [0064] 각각의 제 2 테이블 정보(2411)는 대응하는 제 2 테이블에 대한 정보를 저장한다.
- [0065] 일 실시예에서 제 2 테이블 정보(2411)는 대응하는 제 2 테이블에 대한 물리 주소와 메타정보를 그대로 저장할 수 있다. 이 경우 메타데이터 풀(230)은 중복된 정보이므로 생략될 수 있다.
- [0066] 다만 본 실시예에서는 LSM 트리 버전(241)을 다수 개 생성하는데 이 경우 중복된 데이터로 인하여 요약 데이터(220)의 크기가 증가할 수 있다.
- [0067] 본 실시예에서는 제 2 테이블 정보(2411)로서 대응하는 제 2 테이블에 대한 포인터를 저장할 수 있다. 예를 들어, 메타데이터 풀(230)에서 제 2 테이블에 대응하는 행의 ID를 포인터로서 저장할 수 있다.
- [0068] 이러한 방식은 LSM 트리 버전(241)이 다수 개 존재하는 경우에 요약 데이터(220)의 크기를 줄일 수 있는 점에서

유리하다.

- [0069] 본 실시예에서 LSM 트리 버전(241)은 LSM 트리 영역(310)의 최신 정보를 반영한 글로벌 버전과 특정 시점의 글로벌 버전을 복제한 복제 버전을 포함한다.
- [0070] 글로벌 버전은 하나만 존재하며 LSM 트리 구조에 변경이 발생하는 경우마다 갱신되며, 복제 버전은 하나 또는 둘 이상 존재할 수 있다.
- [0071] 본 실시예에서 요약 데이터 관리부(113)는 영역 질의 명령이 실행될 때마다 해당 시점에서의 글로벌 버전을 복제한 복제 버전을 생성한다.
- [0072] 도 4는 LSM 트리 버전(241)의 생성 방법을 설명하는 설명도이다.
- [0073] 전술한 바와 같이 요약 데이터 관리부(113)는 LSM 트리 영역(310)의 최신 정보를 반영하여 글로벌 버전을 갱신한다.
- [0074] 이하에서, t1에서의 글로벌 버전을 버전 1(V1)로 지칭하고, t2에서의 글로벌 버전을 버전 2(V2)로 지칭하고, t3에서의 글로벌 버전을 버전 3(V3)으로 지칭하고, t4에서의 글로벌 버전을 버전 4(V4)로 지칭한다.
- [0075] 예를 들어, t1에서 t2 사이에 컴팩션 동작이 수행되면 LSM 트리 구조에 변동이 발생하므로 요약 데이터 관리부(113)는 t2에서 글로벌 버전을 버전2로 변경한다.
- [0076] 마찬가지로, t2에서 t3 사이에 컴팩션 동작이 수행되면 LSM 트리 구조에 변동이 발생하므로 요약 데이터 관리부(113)는 t3에서 글로벌 버전을 버전3으로 변경하고, t3에서 t4 사이에 컴팩션 동작이 수행되면 LSM 트리 구조에 변동이 발생하므로 요약 데이터 관리부(113)는 t4에서 글로벌 버전을 버전4로 변경한다.
- [0077] 본 실시예에서 복제 버전은 영역 질의 명령이 실행되는 경우 생성된다.
- [0078] 예를 들어, t1과 t2 사이에서 컴팩션 동작이 수행되기 전에 제 1 영역 질의 명령(Q1)이 실행되면 요약 데이터 관리부(113)는 해당 시점에서의 글로벌 버전인 버전 1(V1)을 복제한 제 1 복제 버전(CV1)을 버전 데이터(240)에 추가한다.
- [0079] 키-값 명령 처리부(112)는 제 1 영역 질의 명령(Q1)의 처리 도중에 글로벌 버전이 변경되더라도 제 1 복제 버전(CV1)을 기준으로 제 1 영역 질의 명령(Q1)을 처리한다.
- [0080] 요약 데이터 관리부(113)는 제 1 영역 질의 명령(Q1)의 처리가 완료되면 제 1 복제 버전(CV1)을 삭제한다.
- [0081] 마찬가지로, t3과 t4 사이에서 컴팩션 동작이 수행되기 전에 제 2 영역 질의 명령(Q2)이 실행되면 요약 데이터 관리부(113)는 해당 시점에서의 글로벌 버전인 버전 3(V3)을 복제한 제 3 복제 버전(CV3)을 버전 데이터(240)에 추가한다.
- [0082] 키-값 명령 처리부(112)는 제 2 영역 질의 명령(Q2)의 처리 도중에 글로벌 버전이 변경되더라도 제 3 복제 버전(CV3)을 기준으로 제 2 영역 질의 명령(Q2)을 처리한다.
- [0083] 요약 데이터 관리부(113)는 제 2 영역 질의 명령(Q2)의 처리가 완료되면 제 3 복제 버전(CV3)을 삭제한다.
- [0084] 요약 데이터 관리부(113)는 글로벌 버전을 최초로 생성하였을 때 메타데이터 풀(230)에서 모든 행의 레퍼런스 카운터를 1로 설정한다.
- [0085] 요약 데이터 관리부(113)는 글로벌 버전을 갱신하는 경우 메타데이터 풀(230)에서 모든 행의 레퍼런스 카운터를 조절한다.
- [0086] 예를 들어, 컴팩션 동작에 의해 병합되어 삭제된 제 2 테이블에 대응하는 레퍼런스 카운터는 1 감소시키고, 기존에 존재하는 제 2 테이블의 레퍼런스 카운터는 현재값을 유지하며, 새로 생성된 제 2 테이블의 레퍼런스 카운터는 1로 설정한다.
- [0087] 요약 데이터 관리부(113)는 복제 버전이 생성되면 메타데이터 풀(230)에서 복제 버전에 존재하는 제 2 테이블에 대응하는 행의 레퍼런스 카운터를 1만큼 증가시키고, 복제 버전이 삭제되면 메타데이터 풀(230)에서 복제 버전에 존재하였던 제 2 테이블에 대응하는 행의 레퍼런스 카운터를 1만큼 감소시킨다.
- [0088] 전술한 바와 같이 본 실시예에서 요약 데이터 관리부(113)는 LSM 트리 영역(310)에서 삭제된 제 2 테이블에 대응하는 행을 메타데이터 풀(230)에서 삭제하기 전에 레퍼런스 카운터를 참조해야 한다.

- [0089] 예를 들어 컴팩션 동작에 의해 병합되어 제 2 테이블이 삭제되는 경우를 가정한다. 만일 버전 1에 존재하던 제 2 테이블이 LSM 트리 영역(310)에서 삭제되었고 이에 대응하는 행이 메타데이터 풀(230)에서 함께 삭제되었다고 가정한다.
- [0090] 이 경우 버전 1에 존재하던 제 2 테이블에 대응하는 행이 메타데이터 풀(320)에서 삭제되면 제 1 복제 버전(CV1)을 기준으로 처리되고 있는 제 1 영역 질의 명령(Q1)은 포인터에 의해 참조되는 값을 얻지 못하여 정상적으로 처리되지 못할 수 있다.
- [0091] 따라서, 요약 데이터 관리부(113)는 메타데이터 풀(230)에서 제 2 테이블을 삭제하기 전에 해당 테이블을 사용하고 있는 복제 버전이 존재하는지 판단해야 한다.
- [0092] 전술한 바와 같이, 어느 하나의 제 2 테이블에 대응하는 레퍼런스 카운터는 해당 제 2 테이블 정보를 사용하는 복제 버전이 생성되는 경우 1 증가하고 삭제되는 경우 1 감소한다.
- [0093] 즉, 제 2 테이블이 LSM 트리 영역(310)에서 삭제되었다고 해도 해당 제 2 테이블의 정보를 참조하는 영역 질의 명령이 수행되는 도중이라면 대응하는 행의 레퍼런스 카운터 값은 1이상의 값으로 유지된다.
- [0094] 따라서 요약 데이터 관리부(113)는 삭제된 제 2 테이블에 대응하는 레퍼런스 카운터의 값이 0보다 크면 대응하는 행을 메타데이터 풀(230)에 유지해야 한다.
- [0095] 요약 데이터 관리부(113)는 메타데이터 풀(230)의 레퍼런스 카운터를 갱신하는 경우 레퍼런스 카운터의 값이 0이 되는 경우 해당 행을 삭제할 수 있다.
- [0096] 요약 데이터 관리부(113)는 일정한 주기로 메타데이터 풀(230)의 레퍼런스 카운터 값을 조회하여 0의 값을 가지는 행을 메타데이터 풀(230)에서 삭제할 수 있다.
- [0097] 본 발명의 권리범위는 이상의 개시로 한정되는 것은 아니다. 본 발명의 권리범위는 청구범위에 문언적으로 기재된 범위와 그 균등범위를 기준으로 해석되어야 한다.

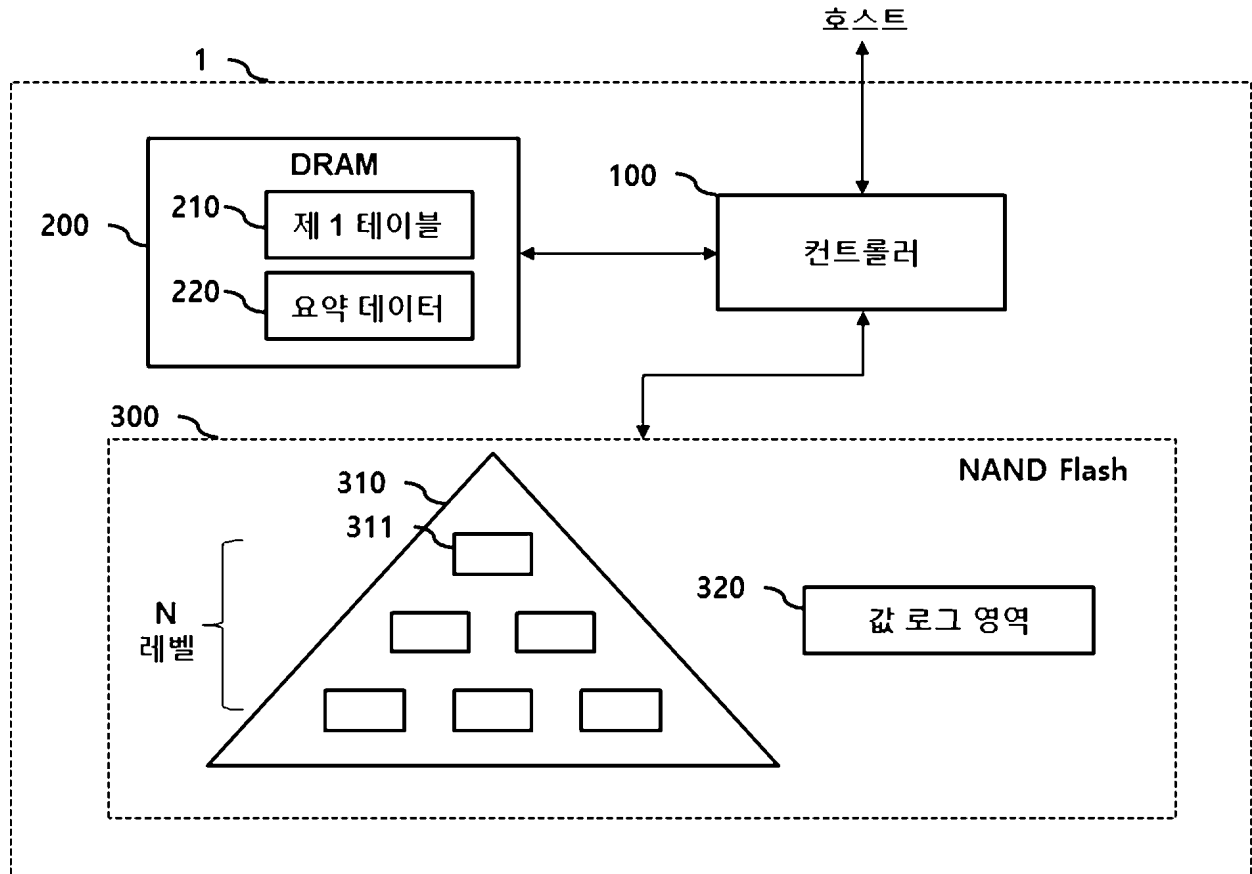
부호의 설명

- [0098] 1: 키-값 기반 데이터 저장 장치, KV SSD
- 100: 컨트롤러
- 110: 제어 회로
- 111: 호스트 인터페이스
- 112: 키-값 명령 처리부
- 113: 요약 테이블 관리부
- 114: 컨트롤러 인터페이스
- 120: 디램 컨트롤러
- 130: 플래시 채널 컨트롤러
- 200: 디램
- 210: 제 1 테이블, MemTable
- 220: 요약 데이터
- 230: 메타데이터 풀
- 240: 버전 데이터
- 241: LSM 트리 버전
- 300: 낸드 플래시 메모리
- 310: LSM 트리 영역
- 311: 제 2 테이블, SSTable

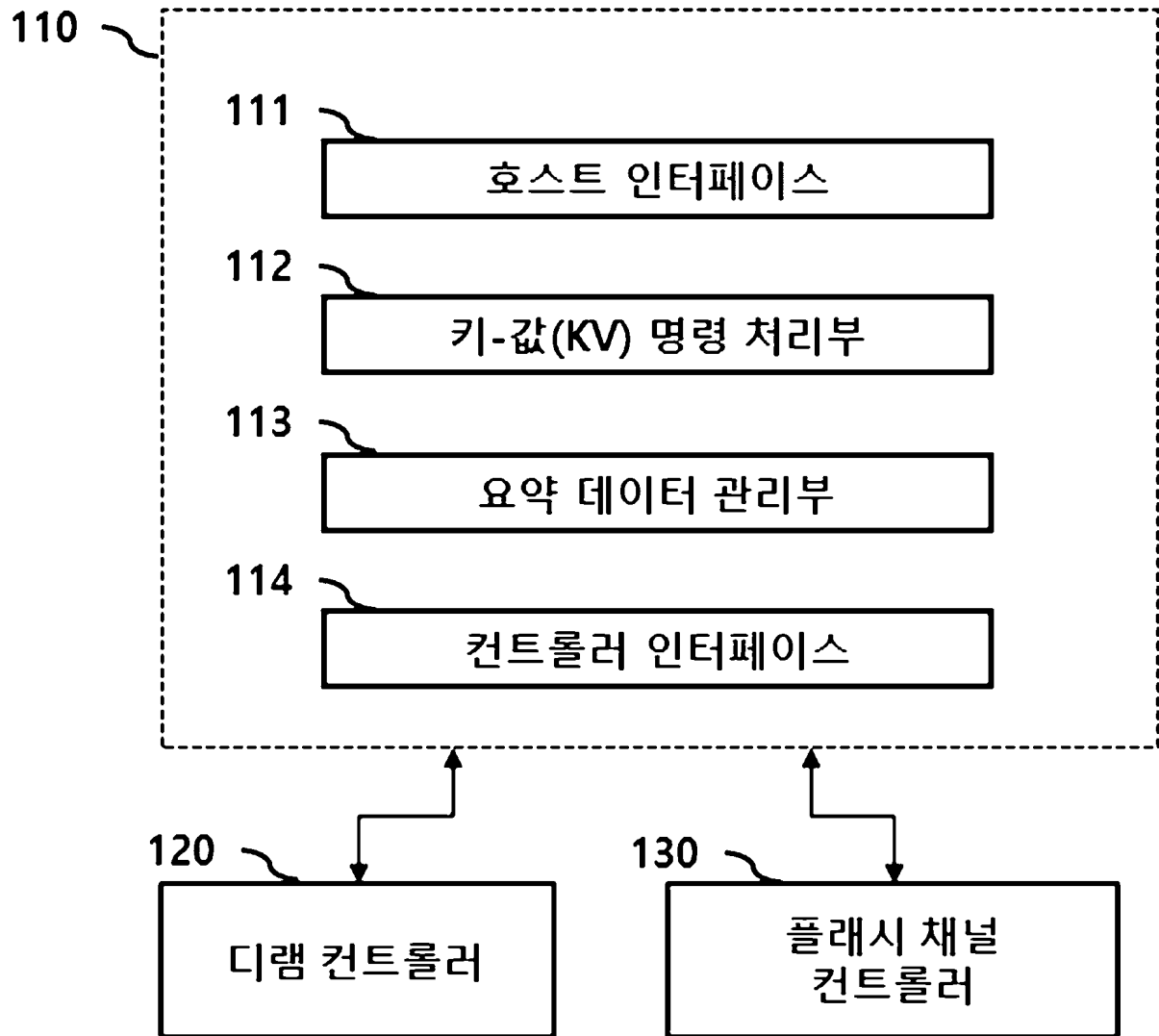
320: 값 로그 영역

도면

도면1

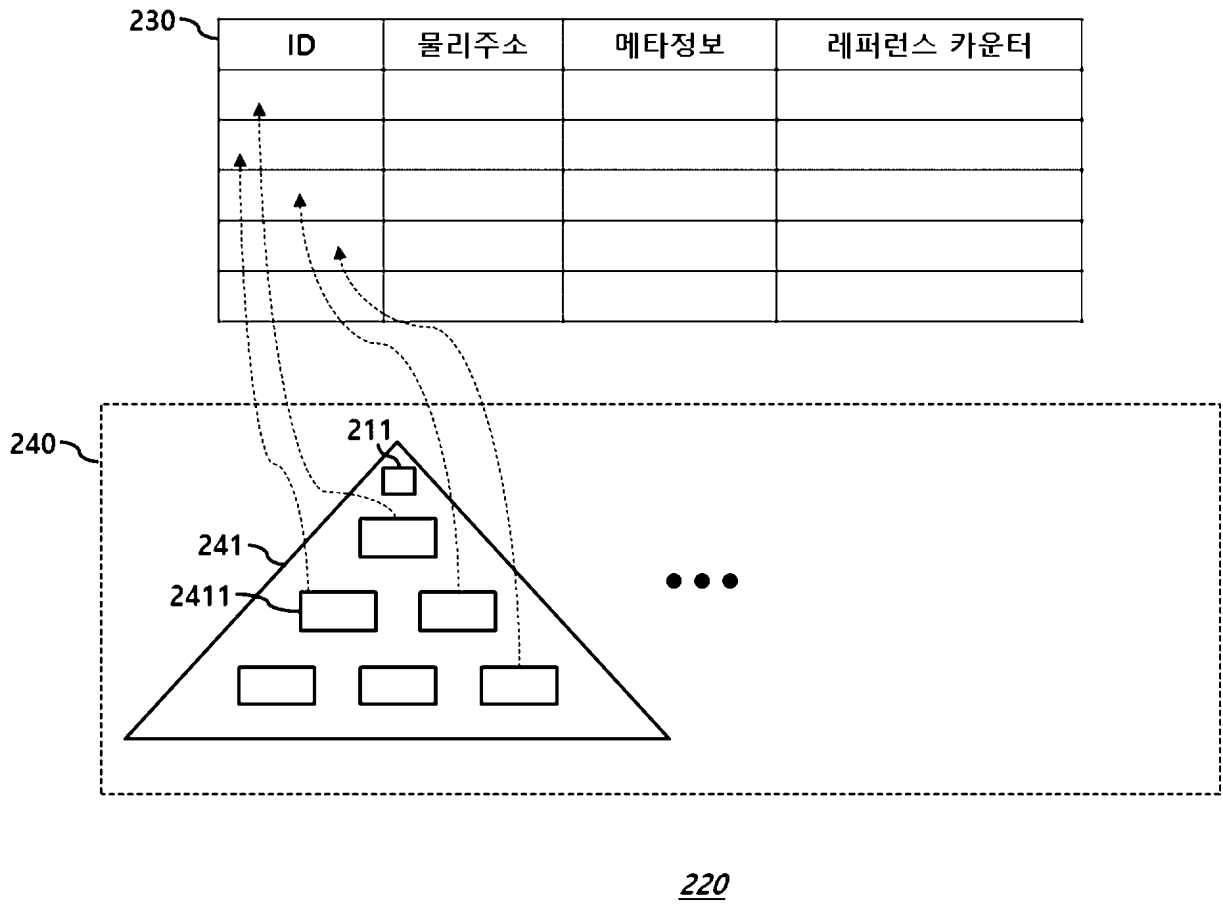


도면2



100

도면3



도면4

