



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0160468  
(43) 공개일자 2024년11월11일

(51) 국제특허분류(Int. Cl.)  
G06F 12/0862 (2016.01) G06F 3/06 (2006.01)  
(52) CPC특허분류  
G06F 12/0862 (2013.01)  
G06F 3/0604 (2013.01)  
(21) 출원번호 10-2023-0057331  
(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

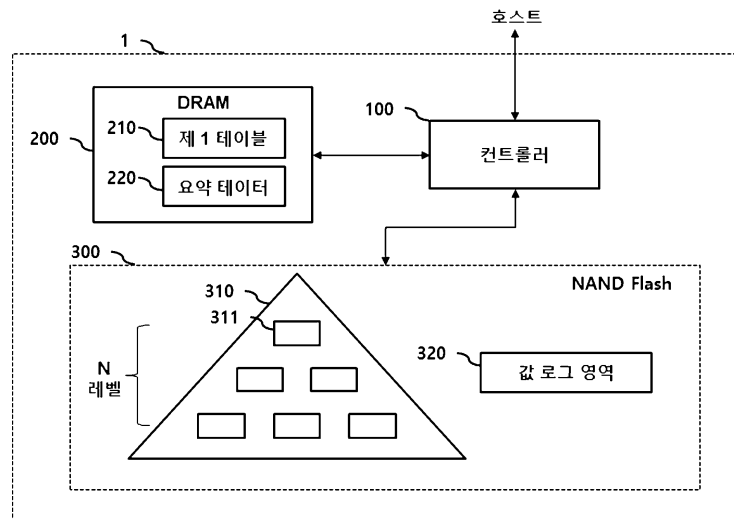
전체 청구항 수 : 총 13 항

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

(57) 요약

키-값 기반 데이터 저장 장치는 제 1 테이블을 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 영역 질의 명령을 처리하는 컨트롤러를 포함하되, 컨트롤러는 키에 대응하는 하나 또는 둘 이상의 버퍼링된 제 2 테이블을 저장하는 프리페치 키 버퍼; 및 영역 질의 명령을 처리하는 도중 버퍼링된 제 2 테이블에서 조회하지 않은 키의 개수가 미리 지정한 개수 이하이면 버퍼링된 제 2 테이블에 인접한 제 2 테이블을 비휘발성 메모리에서 읽어 프리페치 키 버퍼에 저장하는 키 프리페치 동작을 제어하는 제어 회로를 포함한다.

대표도 - 도1



(52) CPC특허분류

**G06F 3/0614** (2013.01)

**G06F 3/0658** (2013.01)

**G06F 3/0659** (2013.01)

**G06F 2212/1024** (2013.01)

**G06F 2212/1032** (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 테이블을 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

영역 질의 명령을 처리하는 컨트롤러

를 포함하되, 상기 컨트롤러는

키에 대응하는 하나 또는 둘 이상의 버퍼링된 제 2 테이블을 저장하는 프리페치 키 버퍼; 및

상기 영역 질의 명령을 처리하는 도중 상기 버퍼링된 제 2 테이블에서 조회하지 않은 키의 개수가 미리 지정된 개수 이하이면 상기 버퍼링된 제 2 테이블에 인접한 제 2 테이블을 상기 비휘발성 메모리에서 읽어 상기 프리페치 키 버퍼에 저장하는 키 프리페치 동작을 제어하는 제어 회로

를 포함하는 키-값 기반 데이터 저장 장치.

#### 청구항 2

청구항 1에 있어서, 상기 제어 회로는 프리페치 데이터 버퍼를 더 포함하고, 상기 제어 회로는 영역 질의 명령을 처리하는 과정에서 현재 키에 대응하는 데이터를 상기 프리페치 데이터 버퍼에서 읽고, 상기 현재 키에서 일정한 거리만큼 이격된 키에 대응하는 값을 상기 값 로그 영역에서 읽어 상기 프리페치 데이터 버퍼에 저장하는 데이터 프리페치 동작을 추가로 제어하는 키-값 기반 데이터 저장 장치.

#### 청구항 3

청구항 2에 있어서, 상기 컨트롤러는

상기 제어 회로의 제어에 따라 디램인 상기 휘발성 메모리를 제어하는 디램 컨트롤러; 및

상기 제어 회로의 제어에 따라 낸드 플래시 메모리인 상기 비휘발성 메모리를 제어하는 플래시 채널 컨트롤러

를 더 포함하는 키-값 기반 데이터 저장 장치.

#### 청구항 4

청구항 3에 있어서, 상기 제어 회로는

상기 영역 질의 명령을 처리하면서 키에 대응하는 값을 읽기 위해 상기 디램 컨트롤러 및 상기 플래시 채널 컨트롤러를 제어하는 키-값 명령 처리부; 및

상기 키 프리페치 동작과 상기 데이터 프리페치 동작을 수행하기 위해 상기 키-값 명령 처리부를 제어하는 프리페치 제어부

를 더 포함하는 키-값 기반 데이터 저장 장치.

#### 청구항 5

청구항 4에 있어서, 상기 제어 회로는 호스트에서 제공된 영역 질의 요청을 디코딩하여 영역 질의 명령을 추출하는 호스트 인터페이스 및 상기 디램 컨트롤러 및 상기 플래시 채널 컨트롤러에 신호를 제공하는 컨트롤러 인터페이스를 더 포함하는 키-값 기반 데이터 저장 장치.

#### 청구항 6

청구항 1에 있어서, 상기 영역 질의 명령은 시작 키에 대응하는 제 1 명령과 상기 제 1 명령에 후속하는 다수의

제 2 명령을 포함하고,

상기 제어 회로는 상기 제 1 명령을 수행하는 경우 상기 시작 키를 포함하는 하나 또는 둘 이상의 제 2 테이블을 상기 비휘발성 메모리에서 읽어 상기 프리페치 키 버퍼에 저장하고, 상기 제 1 테이블 및 상기 하나 또는 둘 이상의 제 2 테이블을 조회하여 가장 최근에 갱신된 상기 시작 키에 대응하는 데이터 읍셋을 찾고 상기 데이터 읍셋을 이용하여 상기 시작 키에 대응하는 값을 읽도록 상기 휘발성 메모리 및 상기 비휘발성 메모리를 제어하는 키-값 기반 데이터 저장 장치.

#### 청구항 7

청구항 6에 있어서, 상기 제어 회로는 상기 제 2 명령을 수행하는 경우 상기 프리페치 키 버퍼에 저장된 상기 하나 또는 둘 이상의 제 2 테이블 중에서 상기 제 2 명령에 대응하는 현재 키 이후에 추가로 조회할 수 있는 키의 개수가 미리 정해진 값보다 작은 대상 테이블을 선택하고, 상기 대상 테이블에 인접한 제 2 테이블을 읽어 상기 프리페치 키 버퍼에 저장하도록 상기 휘발성 메모리 및 상기 비휘발성 메모리를 제어하는 키-값 기반 데이터 저장 장치.

#### 청구항 8

청구항 7에 있어서, 상기 제어 회로는 상기 제 2 명령을 수행하는 경우 상기 현재 키에 대응하는 데이터를 읽는 동작을 병렬적으로 수행하는 키-값 기반 데이터 저장 장치.

#### 청구항 9

제 1 테이블을 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,

영역 질의 명령의 시작 키에 대응하는 값을 읽는 제 1 명령을 처리하는 단계; 및

상기 시작 키에 후속하는 다수의 키에 대응하는 값을 읽도록 상기 제 1 명령에 후속하는 다수의 제 2 명령을 순차적으로 처리하는 단계

를 포함하되,

상기 제 1 명령을 처리하는 단계는

상기 시작 키를 포함하는 하나 또는 둘 이상의 제 2 테이블을 상기 LSM 트리 영역에서 읽어 프리페치 키 버퍼에 저장하는 단계;

상기 제 1 테이블과 상기 프리페치 키 버퍼에 저장된 하나 또는 둘 이상의 제 2 테이블 중 어느 하나에 저장된 시작 키에 대응하는 데이터 읍셋을 선택하는 단계; 및

상기 데이터 읍셋에 대응하는 값을 상기 값 로그 영역에서 읽는 단계

를 포함하는 키-값 기반 데이터 저장 장치의 동작 방법.

#### 청구항 10

청구항 9에 있어서, 상기 다수의 제 2 명령 중 어느 하나의 제 2 명령을 처리하는 단계는

상기 프리페치 키 버퍼에서 상기 어느 하나의 제 2 명령에 대응하는 현재 키를 포함하는 하나 또는 둘 이상의 제 2 테이블 중에서 현재 키 이후에 추가로 조회할 수 있는 키의 개수가 미리 정해진 값보다 작은 대상 테이블을 선택하는 단계; 및

상기 대상 테이블에 인접한 제 2 테이블을 상기 LSM 트리 영역에서 읽어 상기 프리페치 키 버퍼에 저장하는 추가 프리페치 단계

를 포함하는 키-값 기반 데이터 저장 장치의 동작 방법.

#### 청구항 11

청구항 10에 있어서, 상기 어느 하나의 제 2 명령을 처리하는 단계는 상기 현재 키에 대응하는 값을 읽는 단계를 더 포함하되,

상기 값을 읽는 단계는 상기 추가 프리페치 단계와 병렬적으로 수행되는 키-값 기반 데이터 저장 장치의 동작 방법.

#### 청구항 12

청구항 11에 있어서, 상기 어느 하나의 제 2 명령을 처리하는 단계는 상기 현재 키로부터 일정한 간격만큼 이격된 키에 대응하는 값을 상기 값 로그 영역에서 읽어 프리페치 데이터 버퍼에 저장하는 데이터 프리페치 단계를 더 포함하는 키-값 기반 데이터 저장 장치의 동작 방법.

#### 청구항 13

청구항 12에 있어서, 상기 값을 읽는 단계는 상기 프리페치 데이터 버퍼에서 상기 현재 키에 대응하는 값을 읽는 단계를 포함하는 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 프리페치 동작을 수행하여 영역 질의 명령을 효율적으로 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 특정 키 범위에 대응하는 키-값 쌍을 읽는 동작을 수행한다.

[0004] 낸드 플래시 메모리 기반의 종래의 키-값 데이터 저장 장치는 영역 질의 명령을 처리하는 과정에서 키와 데이터를 읽기 위해 낸드 플래시 메모리에 빈번하게 접근하는데 낸드 플래시 메모리 접근 시간이 상대적으로 매우 길기 때문에 영역 질의 명령을 처리하는 시간이 증가하고 이로 인하여 전체적인 성능이 저하되는 문제가 발생한다.

[0005] 특히 값을 읽기 전에 키를 읽는 동작을 추가로 수행하는 경우 대기 시간이 더욱 증가하는 문제가 있다.

### 선행기술문헌

#### 특허문헌

[0006] (특허문헌 0001) KR 10-2021-0034479 A

(특허문헌 0002) KR 10-2021-0123236 A

(특허문헌 0003) US 11,550,504 B2

#### 비특허문헌

[0007] (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

[0008] 본 기술은 영역 질의 명령을 효율적으로 처리할 수 있는 키-값 데이터 저장 장치와 그 동작 방법을 제공한다.

### 과제의 해결 수단

[0009] 본 기술에 의한 키-값 기반 데이터 저장 장치는 제 1 테이블을 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 영역 질의 명령을 처리하는 컨트롤러를 포함하되, 컨트롤러는 키에 대응하는 하나 또는 둘 이상의 버퍼링된 제 2 테이블을 저장하는 프리페치 키 버퍼; 및 영역 질의 명령을 처리하는 도중 버퍼링된 제 2 테이블에서 조회하지 않은 키의 개수가 미리 지정한 개수 이하이면 버퍼링된 제 2 테이블에 인접한 제 2 테이블을 비휘발성 메모리에서 읽어 프리페치 키 버퍼에 저장하는 키 프리페치 동작을 제어하는 제어 회로를 포함한다.

[0010] 본 기술에 의한 키-값 기반 데이터 저장 장치의 동작 방법은 제 1 테이블을 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, 영역 질의 명령의 시작 키에 대응하는 값을 읽는 제 1 명령을 처리하는 단계; 및 시작 키에 후속하는 다수의 키에 대응하는 값을 읽도록 제 1 명령에 후속하는 다수의 제 2 명령을 순차적으로 처리하는 단계를 포함하되, 제 1 명령을 처리하는 단계는 시작 키를 포함하는 하나 또는 둘 이상의 제 2 테이블을 LSM 트리 영역에서 읽어 프리페치 키 버퍼에 저장하는 단계; 제 1 테이블과 프리페치 키 버퍼에 저장된 하나 또는 둘 이상의 제 2 테이블 중 어느 하나에 저장된 시작 키에 대응하는 데이터 옵션을 선택하는 단계; 및 데이터 옵션에 대응하는 값을 값 로그 영역에서 읽는 단계를 포함한다.

### 발명의 효과

[0011] 본 기술은 키-값 데이터 저장 장치에서 키 또는 키에 대응하는 값을 미리 읽어 옴으로써 영역 질의 명령을 처리하는 경우 대기 시간을 줄여 데이터 저장 장치의 동작 성능을 향상시킬 수 있다.

### 도면의 간단한 설명

[0012] 도 1은 본 발명의 일 실시예에 의한 키-값 기반 데이터 저장 장치를 나타내는 블록도.

도 2는 본 발명의 일 실시예에 의한 제어 회로를 나타내는 블록도.

도 3은 본 발명의 일 실시예에 의한 제어 회로의 동작을 나타내는 순서도.

도 4는 본 발명의 일 실시예에 의한 제어 회로의 키 프리페치 동작을 나타내는 설명도.

도 5는 본 발명의 일 실시예에 의한 제어 회로의 값 프리페치 동작을 나타내는 설명도.

### 발명을 실시하기 위한 구체적인 내용

[0013] 이하에서는 첨부한 도면을 참조하여 본 발명의 실시예를 개시한다.

[0014] 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.

[0015] 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.

[0016] 도 1은 본 발명의 일 실시예에 의한 KV SSD(1)를 나타내는 블록도이다.

[0017] KV SSD(1)는 컨트롤러(100), 디램(200), 낸드 플래시 메모리(300)를 포함한다.

[0018] 컨트롤러(100)는 호스트에서 제공된 명령을 수신하여 읽기, 쓰기, 또는 소거 동작을 수행할 수 있다. 이는 비특허문헌 1에 개시되어 있으므로 반복적인 설명은 생략한다.

[0019] 본 기술에 의한 컨트롤러(100)는 영역 질의 명령을 추가로 처리하며 이 과정에서 키 또는 값에 대한 프리페치 동작을 제어한다.

[0020] 이하에서 키에 대한 프리페치 동작을 키 프리페치 동작으로 지칭하고 값에 대한 프리페치 동작을 값 프리페치 동작 또는 데이터 프리페치 동작으로 지칭한다. 이에 대해서는 아래에서 다시 구체적으로 개시한다.

[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 있으며 요약 데이터(220)는 Summary로 지칭될 수 있다.
- [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. 하나의 제 2 테이블은 다수의 키와 이에 대응하는 값을 저장한다.
- [0025] 값 로그 영역(320)은 키에 대응하는 값을 저장하는 영역으로서 하나의 키에 대응하는 값의 주소는 키에 대응하는 값과 베이스 주소를 이용하여 계산될 수 있다.
- [0026] 도 2는 본 발명의 일 실시예에 의한 컨트롤러(100)를 나타내는 블록도이다.
- [0027] 컨트롤러(100)는 제어 회로(110), 디램 컨트롤러(120), 플래시 채널 컨트롤러(130), 프리페치 키 버퍼(140), 및 프리페치 데이터 버퍼(150)를 포함한다.
- [0028] 제어 회로(110)는 호스트에서 제공된 영역 질의 명령을 해석하고 인덱스 프리페치 동작과 값 프리페치 동작을 제어한다.
- [0029] 제어 회로(110)는 하드웨어, 소프트웨어, 또는 하드웨어와 소프트웨어의 조합으로 구성될 수 있다.
- [0030] 디램 컨트롤러(120)는 제어 회로(110)의 제어에 따라 디램(200)에 대한 읽기 또는 쓰기 동작을 제어한다.
- [0031] 플래시 채널 컨트롤러(130)는 제어 회로(110)의 제어에 따라 낸드 플래시 메모리(300)에 대한 읽기, 쓰기, 또는 소거 동작을 제어한다.
- [0032] 본 실시예에서 플래시 채널 컨트롤러(130)는 하나 또는 둘 이상의 채널을 개별적으로 제어하며 각각의 채널에는 다수개의 플래시 메모리가 연결된다.
- [0033] 디램 컨트롤러(120)나 플래시 채널 컨트롤러(130) 자체는 종래의 기술이므로 이에 대한 반복적인 설명은 생략한다.
- [0034] 프리페치 키 버퍼(140)는 LSM 트리 영역(310)에서 미리 읽어온 다수의 제 2 테이블을 저장한다.
- [0035] 프리페치 데이터 버퍼(150)는 값 로그 영역(320)에서 미리 읽어온 다수의 데이터를 저장한다.
- [0036] 데이터를 프리페치 하는 경우 현재 키로부터 D번째에 위치한 키에 대응하는 데이터를 미리 읽는데 이때 D를 프리페치 거리(Degree)로 지칭할 수 있으며 이때 D는 자연수이다. 이에 대해서는 아래에서 구체적으로 개시한다.
- [0037] 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.
- [0038] 호스트 인터페이스(111)는 호스트에서 전달된 요청에서 키-값 명령을 추출한다.
- [0039] 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.
- [0040] 키-값 명령 처리부(112)는 일반적인 키-값 명령인 PUT, GET 명령을 처리할 수 있으나 이는 종래의 기술을 통해 잘 알려진 것이므로 이에 대한 반복적인 설명은 생략한다.
- [0041] 본 실시예에서 키-값 명령 처리부(112)는 영역 질의 명령에 따라 프리페치 제어부(113), 컨트롤러 인터페이스(114)를 제어하여 키-값 명령을 처리한다.
- [0042] 영역 질의 명령은 일정한 키 범위의 데이터를 읽기 위한 명령으로서 시작 키를 찾는 제 1 명령과 시작 키 이후 순차적으로 다음 키에 대응하는 데이터를 읽는 다수의 제 2 명령을 포함할 수 있다.
- [0043] 제 1 명령을 Seek 명령 또는 Seek로 지칭할 수 있고, 제 2 명령을 Next 명령 또는 Next로 지칭할 수 있다.
- [0044] 전술한 바와 같이 디램(200)은 요약 데이터(220)를 저장하는데, 요약 데이터(220)는 LSM 트리 영역(310)에 저장된 모든 제 2 테이블 각각에 대한 메타 정보를 저장한다. 배터리, 수퍼 캐پ 등의 비상 전원 공급 장치를 사용하여 전원이 차단되는 상황에서 디램(200)에 전원을 공급하여 요약 데이터(220)를 보호할 수 있다.

- [0045] 메타 정보는 예를 들어 제 2 테이블의 레벨, 해당 제 2 테이블의 주소, 상위 레벨에 위치한 제 2 테이블의 주소, 하위 레벨에 위치한 제 2 테이블의 주소, 해당 제 2 테이블에 저장된 다수의 키 정보 등을 포함한다. 이때 주소는 낸드 플래시 메모리(300) 내에서의 주소를 나타낸다.
- [0046] 본 실시예에서 영역 질의 명령에 의한 세부 동작은 다음과 같이 구분될 수 있다.
- [0047] 이하에서는 영역 질의 명령을 수행하는 첫 번째 키를 시작 키로 지칭하고, 영역 질의 명령을 수행하는 도중 현재 읽을 데이터에 대응하는 키를 현재 키로 지칭하고, 현재 키 다음의 키를 다음 키로 지칭한다.
- [0048] (a) 메모리 참조 및 계산
- [0049] 제 1 명령을 수행하는 경우에는 디램(200)의 요약 데이터(210)를 참조하여 시작 키를 포함하는 제 2 테이블을 찾아 프리페치 키 버퍼(140)에 저장하는 동작에 대응한다.
- [0050] 이때 레벨별로 시작 키를 포함하는 다수의 제 2 테이블이 존재하는 경우 이들 모두를 프리페치 키 버퍼(140)에 저장한다. 키-값 명령 제어부(112)는 제 1 테이블 및 버퍼링된 다수의 제 2 테이블 중에서 가장 최근에 갱신된 시작 키를 선택한다.
- [0051] 제 2 명령을 수행하는 경우에는 제 2 테이블에서 현재 이후 조회할 수 있는 키의 개수를 판단하는 동작에 대응한다.
- [0052] 본 실시예에서는 제 2 테이블에서 조회할 수 있는 키의 개수가 일정한 개수 이하가 되면 다음에 위치한 제 2 테이블을 미리 읽기 위한 키 프리페치 동작을 수행한다.
- [0053] 제 2 테이블은 크기 순으로 정렬된 일정한 개수의 키를 포함하므로, 현재 키에 대응하는 배열 인덱스를 커서로 표시하는 경우 커서를 이용하여 추가로 조회할 수 있는 키의 개수를 판단할 수 있다. 커서가 배열 인덱스의 최대값을 초과하는 경우 더 이상 읽을 수 있는 키가 없음을 의미한다.
- [0054] (b) 제 2 테이블 추가 읽기
- [0055] 버퍼링된 제 2 테이블에 인접한 다음 제 2 테이블을 찾아 프리페치 키 버퍼(140)에 저장하는 동작에 대응한다.
- [0056] (c) 증가 및 비교
- [0057] 제 1 테이블 및 버퍼링된 하나 또는 둘 이상의 제 2 테이블에서 현재 키 다음의 키를 검색하고 이들 중 가장 작은 키를 찾는 동작에 대응한다. 가장 작은 키가 다수 개 존재하는 경우에는 가장 최근에 갱신된 키를 다음 키로 선택한다. 이때 선택된 제 2 테이블에서 커서의 위치를 갱신한다.
- [0058] (d) 값 읽기
- [0059] 현재 키에 대응하는 값을 값 저장 영역(320)에서 읽는 동작에 대응한다.
- [0060] (e) 호스트 전달
- [0061] 값 저장 영역(320)에서 읽은 데이터를 호스트로 전달하는 동작에 대응한다.
- [0062] 본 실시예에서는 제 2 테이블을 미리 읽는 키 프리페치 동작을 수행하여 대기 시간을 줄일 수 있다. 또한 본 실시예에서는 값을 미리 읽는 데이터 프리페치 동작을 수행하여 대기 시간을 줄일 수 있다.
- [0063] 이하에서는 키 프리페치 동작과 데이터 프리페치 동작에 대해서 보다 구체적으로 개시한다.
- [0064] 도 3은 본 발명의 일 실시예에 의한 컨트롤러(100)의 동작을 나타내는 순서도이다.
- [0065] 본 실시예에서 영역 질의 명령은 제 1 명령(Seek)을 수행한 후 다수의 제 2 명령(Next)을 순차적으로 수행한다. 도 3은 제 1 명령 수행 후 제 2 명령을 수행하는 경우에 대응한다.
- [0066] 이하에서는 먼저 제 1 명령을 수행하는 경우에 대해서 개시한다.
- [0067] 제 1 명령에 따라 키-값 명령 제어부(112)는 요약 테이블(220)을 참조하여 시작 키를 포함하는 제 2 테이블을 읽어 프리페치 키 버퍼(140)에 저장한다.
- [0068] LSM 트리를 이용하는 KV SSD에서는 동일한 키를 포함하는 제 2 테이블이 다수의 레벨에 분산되어 존재할 수 있다. 이 경우 레벨 별로 시작 키를 포함하는 제 2 테이블을 모두 프리페치 키 버퍼(140)에 저장할 수 있다.
- [0069] 이후 키-값 명령 제어부(112)는 제 1 테이블(210)과 프리페치 키 버퍼(140)에 존재하는 다수의 시작 키를 비교

하여 가장 최근에 갱신된 시작 키를 찾는다.

- [0070] 키-값 명령 제어부(112)는 시작 키에 대응하는 읽기 동작을 실행한다. 시작 키에 대응하는 데이터는 호스트로 출력된다.
- [0071] 본 실시예에서 데이터 프리페치 동작은 선택적으로 수행될 수 있다. 제 1 명령 실행 시 수행되는 데이터 프리페치 동작과 제 2 명령 실행 시 수행되는 데이터 프리페치 동작은 프리페치 데이터 범위에 있어서 차이가 있을 수 있다.
- [0072] 제 1 명령 실행 시 수행되는 실행되는 데이터 프리페치 동작을 개시하면 다음과 같다.
- [0073] 데이터 프리페치 동작시 프리페치 거리를 D로 표시하면 프리페치 제어부(113)는 시작 키의 다음 키부터 시작 키로부터 D번째 키까지의 데이터를 읽도록 키-값 명령 제어부(112)를 제어한다.
- [0074] 예를 들어 D=2인 경우, 프리페치 제어부(113)는 시작 키의 첫 번째 다음 키와, 두 번째 다음 키에 대한 읽기 동작을 수행하도록 키-값 명령 제어부(112)를 제어한다.
- [0075] 현재 키의 다음 키를 찾기 위해서 전술한 (c) 동작을 수행한다. 즉, 버퍼링된 다수의 제 2 테이블 및 제 1 테이블 중에서 현재 키를 포함하는 테이블을 찾고 이들 중 현재 키의 다음 키를 모두 비교한다. 비교 결과 가장 작은 키를 다음 키로 선택하되, 가장 작은 키가 다수 개 있는 경우 가장 최근에 갱신된 키를 다음 키로 결정한다.
- [0076] 이후 다수의 제 2 명령을 수행한다.
- [0077] 제 2 명령이 실행되면 현재 키에 대응하는 값은 프리페치 데이터 버퍼(150)에 존재하므로 이를 읽어 호스트로 전송한다. 데이터 프리페치 동작을 수행하는 경우 현재 키로부터 D 번째 다음 키 하나에 대한 읽기 동작을 추가로 수행한다.
- [0078] 도 3은 제 2 명령이 실행될 때 컨트롤러의 동작을 나타낸다.
- [0079] 먼저 새로운 제 2 명령이 입력되는지 판단한다(S10).
- [0080] 제 2 명령이 입력되지 않으면 대기하고 그렇지 않으면 다음 단계로 이동한다.
- [0081] 먼저 현재 키에 대응하는 데이터를 출력하고, 버퍼링된 하나 또는 둘 이상의 제 2 테이블에서 커서의 위치를 확인한다(S100).
- [0082] 현재 키에 대응하는 데이터는 프리페치 데이터 버퍼(150)에 저장되어 있으므로 바로 출력이 가능하다.
- [0083] 커서는 현재 키의 위치를 나타낸다. 제 2 테이블은 순차적으로 정렬된 일정한 개수의 키를 포함하고 각각의 키의 위치는 인덱스로 표시할 수 있다. 즉 커서는 현재 키에 대응하는 인덱스로 볼 수 있다. 전술한 바와 같이 어느 하나의 키에 대응하는 제 2 테이블은 여러 레벨에 걸쳐 다수 개가 존재할 수 있다. 커서는 다수의 제 2 테이블 각각에 대해서 다른 값을 가질 수 있다.
- [0084] 이후 커서가 임계점에 도달했는지 판단한다(S110).
- [0085] 임계점은 미리 정해진 인덱스로서 제 2 테이블마다 공통된 값을 가진다.
- [0086] 제 2 테이블의 마지막 인덱스에서 커서를 뺀 값은 앞으로 조회할 수 있는 키의 개수에 대응하며 이를 잔여 키 개수로 표시할 수 있다.
- [0087] 제 2 테이블의 마지막 인덱스에서 임계점을 뺀 값을 한계 잔여 키 개수로 지칭할 수 있다.
- [0088] 만일 현재 키에 대응하는 제 2 테이블이 다수 개 존재하는 경우 적어도 하나의 제 2 테이블에서 커서가 임계점에 도달하면 Yes로 판단하고 모든 제 2 테이블에서 커서가 임계점에 도달하지 않으면 No로 판단한다.
- [0089] 단계(S110)에서 No로 판단하면 데이터 프리페치 동작을 수행한다(S120). 데이터 프리페치 동작을 수행하기 위하여 현재 키에서 D 번째 키를 찾고 이에 대응하는 데이터를 읽어 프리페치 데이터 버퍼(150)에 저장한다.
- [0090] 이때 현재 키에서 D번째 키는 버퍼링된 테이블 내에 저장되어 있는 것이 바람직하다. 이에 따라 한계 잔여 키 개수가 D 이상이 되도록 임계점을 설정하는 것이 바람직하다.
- [0091] 단계(S110)에서 Yes로 판단하면 키 프리페치 동작이 수행된다(S130).
- [0092] 제 2 테이블이 다수 개 존재하는 경우 커서가 임계점에 도달한 제 2 테이블을 대상 테이블로 지칭한다.

- [0093] LSM 트리 구조에서 동일한 레벨에 포함된 다수의 제 2 테이블은 키 순서대로 정렬된다. 이에 따라 프리페치 제어부(113)는 요약 데이터(220)를 참조하여 대상 테이블 다음의 키를 포함하는 제 2 테이블을 선택할 수 있다. 프리페치 제어부(113)는 대상 테이블 다음의 제 2 테이블을 프리페치 하도록 키-값 명령 제어부(112)를 제어하고 읽은 제 2 테이블은 프리페치 키 버퍼(140)에 저장한다.
- [0094] 만일 대상 테이블이 해당 레벨의 마지막 제 2 테이블인 경우에는 해당 레벨에 대해서는 키 프리페치 동작을 수행하지 않는다.
- [0095] 대상 테이블이 다수 개 존재하는 경우에는 다수의 대상 테이블 각각에 대해서 위에서 설명한 키 프리페치 동작을 수행한다.
- [0096] 이후 전술한 데이터 프리페치 동작을 수행하고(S120) 대기 단계(S10)로 이동한다.
- [0097] 도 4는 본 발명의 키 프리페치 동작의 효과를 나타내는 설명도이다.
- [0098] 도 4의 설명도는 영역 질의 명령 중 제 2 명령(Next)이 지속되는 경우를 예시한 것이다.
- [0099] 도 4는 낸드 플래시 메모리가 두 개의 채널(NFC1, NFC2)에 분산되어 배치된 경우를 가정한다. 즉 플래시 채널 컨트롤러(130)는 두 개의 채널에 대해서 독립적으로 낸드 플래시 메모리를 제어한다. 또한 도 4에서 a, b, c, d, 및 e 동작은 앞서 개시한 동작과 동일하다.
- [0100] (A)는 키 프리페치 동작을 수행하지 않는 경우의 동작을 나타낸다.
- [0101] 첫 번째 제 2 명령(Next1)은 현재 키가 버퍼링된 제 2 테이블에 존재하는 것을 가정한다(a).
- [0102] 이에 따라 버퍼링된 제 2 테이블에서 현재 키에 대응하는 읍셋을 읽고(c), 읍셋을 이용하여 낸드 플래시 메모리(300)에서 데이터를 읽고(d), 읽은 데이터를 호스트로 전송한다(e).
- [0103] 두 번째 제 2 명령(Next2)은 현재 키가 버퍼링된 제 2 테이블에 존재하지 않는 경우를 가정한다(a).
- [0104] 이에 따라 다음 제 2 테이블을 낸드 플래시 메모리(300)에서 읽어 버퍼링 테이블로 저장하고(b), 버퍼링된 제 2 테이블에서 현재 키에 대응하는 읍셋을 읽어와(c), 읍셋을 이용하여 낸드 플래시 메모리(300)에서 데이터를 읽고(d), 읽은 데이터를 호스트로 전송한다(e).
- [0105] 세 번째 제 2 명령(Next3)은 첫 번째 제 2 명령(Next1)과 동일하다.
- [0106] 이와 같이 두 번째 제 2 명령을 실행하는 경우 데이터를 읽기 전에 제 2 테이블을 읽는 동작을 수행해야 하므로 호스트로 데이터를 전달하기까지 대기 시간이 증가한다.
- [0107] (B)는 키 프리페치 동작을 수행하는 경우를 나타낸다.
- [0108] 이에 따라 a 동작 시에는 커서가 임계점에 도달했는지 판단하게 된다.
- [0109] 첫 번째 제 2 명령(Next1)은 커서가 임계점에 도달한 것을 가정한다(a).
- [0110] 이에 따라 먼저 버퍼링된 제 2 테이블에서 현재 키에 대응하는 읍셋을 읽어와(c), 읍셋을 이용하여 낸드 플래시 메모리(300)에서 데이터를 읽고(d), 이를 호스트로 전송한다(e).
- [0111] 이와 함께, 대상 테이블 다음의 제 2 테이블을 낸드 플래시 메모리(300)에서 읽어 프리페치 키 버퍼(140)에 저장하는(b) 키 프리페치 동작을 수행한다.
- [0112] 두 번째 제 2 명령(Next2)과 세 번째 제 2 명령(Next3)은 커서가 임계점에 도달하지 않은 것을 가정한다.
- [0113] 이 경우는 (A)의 첫 번째 제 2 명령(Next1) 실행 방법과 동일하므로 설명을 반복하지 않는다.
- [0114] 키 프리페치 동작을 수행하면 값을 읽기 전에 제 2 테이블을 읽는 동작을 수행하지 않으므로 대기 시간이 감소한다.
- [0115] (B)에서 데이터를 읽는 동작은 제 1 채널(NFC1)에 연결된 낸드 플래시 메모리에서 수행되고, 제 2 테이블을 읽는 동작은 제 2 채널(NFC2)에 연결된 낸드 플래시 메모리에서 수행되는 것으로 표시하였다.
- [0116] 만일 제 2 테이블이 저장된 낸드 플래시 메모리와 데이터가 저장된 낸드 플래시 메모리가 동일한 채널에 연결된다면 두 동작(b, d)이 순차적으로 진행되어야 하며 이 과정에서 시간 지연이 발생할 수는 있다.
- [0117] 그러나, 이러한 채널 경합으로 인한 시간 지연은 데이터 할당 방식을 효율적으로 수행하여 극복할 수 있는 문제

이며 본 발명과는 별개의 것으로 다루어야 한다.

- [0118] 도 5는 본 발명의 값 프리페치 동작의 효과를 나타내는 설명도이다.
- [0119] 도 5의 설명도는 영역 질의 명령 중 제 2 명령(Next)이 지속되는 과정을 나타낸다.
- [0120] 도 5에서 A는 도 4의 a, b, c 단계에 대응하고, d와 e는 도 4의 d, e와 동일하다.
- [0121] 도 5에서 D = 2인 것으로 가정한다.
- [0122] (A)와 같이 값 프리페치 동작을 수행하지 않는 경우 제어 회로(110)는 A 동작 이후 e 동작을 수행하기 전에 낸드 플래시 메모리에서 데이터를 읽는 동안 대기해야 한다.
- [0123] 첫 번째 제 2 명령(Next1)을 예로 들면 제어 회로(110)는 A 동작을 수행한 후 낸드 플래시 메모리에서 d 동작을 수행하는 동안 대기한 후 e 동작을 수행할 수 있다.
- [0124] 이에 비하여 (B)와 같이 값 프리페치 동작을 수행하는 경우에는 제어 회로(110)에서 A 동작을 수행한 후 낸드 플래시 메모리의 동작을 대기할 필요 없이 프리페치된 데이터를 이용하여 e 동작을 수행할 수 있어 대기 시간이 줄어들 수 있다.
- [0125] (B)에서 세 번째 제 2 명령(Next3)을 예로 들면 A 동작을 수행한 후 이에 대응하는 프리페치된 데이터를 이용하여 e 동작을 수행한다.
- [0126] 이때 제어 회로(110)는 프리페치 데이터 버퍼(150)에 저장된 데이터를 호스트에 전송하므로 적어도 낸드 플래시 메모리(300)에 접근하는데 필요한 시간만큼 대기 시간이 감소한다.
- [0127] e 동작을 수행하는 동안 프리페치 제어부(113)는 현재 키에서 D번째 다음 키에 대한 값 프리페치 동작을 수행한다.
- [0128] 현재 키를 3번 키라고 하면 5번 키에 대응하는 데이터(V5)를 읽고 이를 프리페치 데이터 버퍼(150)에 저장한다. 이때 프리페치된 데이터(V5)는 다섯 번째 제 2 명령(Next5) 실행 시 출력될 수 있다.
- [0129] 본 발명의 권리범위는 이상의 개시로 한정되는 것은 아니다. 본 발명의 권리범위는 청구범위에 문언적으로 기재된 범위와 그 균등범위를 기준으로 해석되어야 한다.

### 부호의 설명

- [0130] 1: 키-값 방식의 데이터 저장 장치, KV SSD
- 100: 컨트롤러
- 110: 제어 회로
- 111: 호스트 인터페이스
- 112: 키-값 명령 처리부
- 113: 프리페치 제어부
- 114: 컨트롤러 인터페이스
- 120: 디램 컨트롤러
- 130: 플래시 채널 컨트롤러
- 140: 프리페치 키 버퍼
- 150: 프리페치 데이터 버퍼
- 200: 디램
- 210: 제 1 테이블, MemTable
- 220: 요약 데이터
- 300: 낸드 플래시 메모리

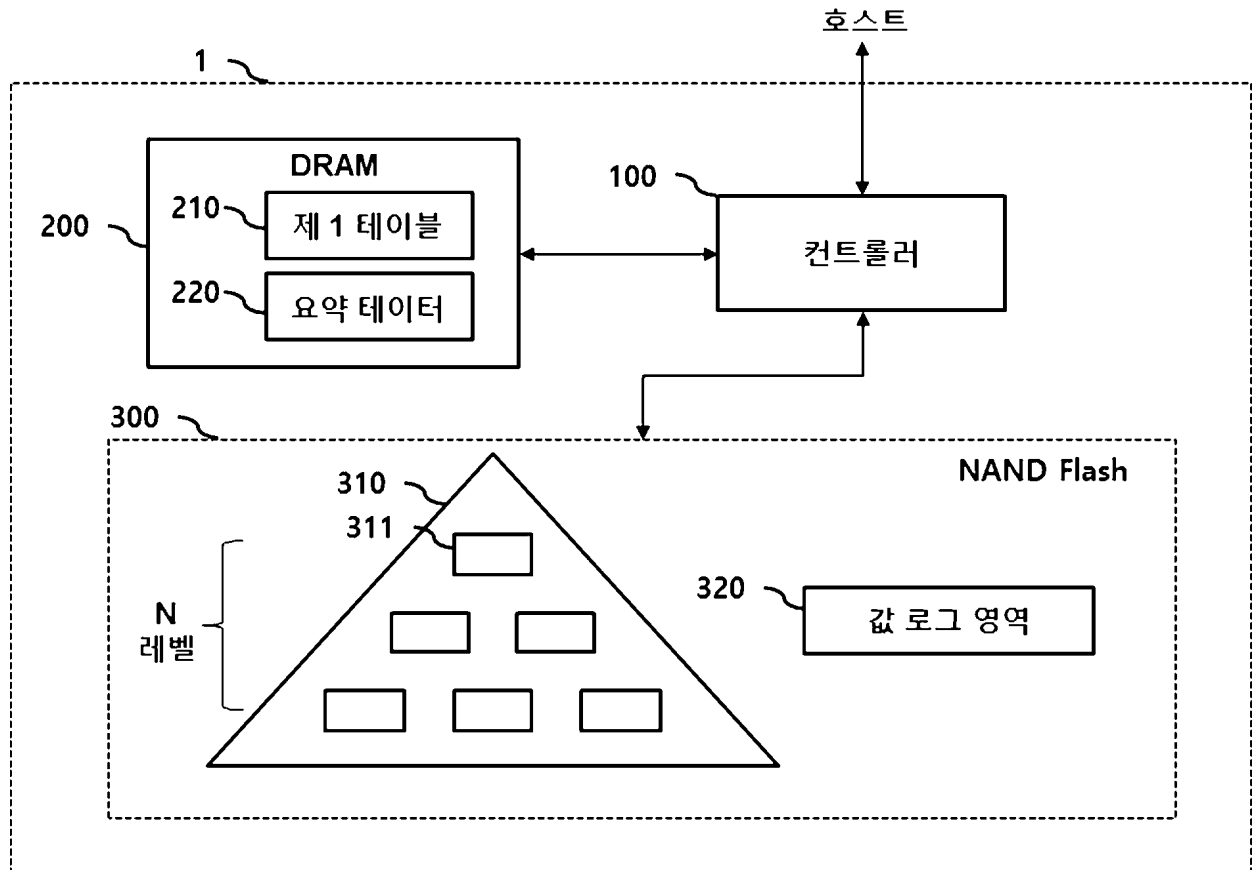
310: LSM 트리 영역

311: 제 2 테이블, SSTable

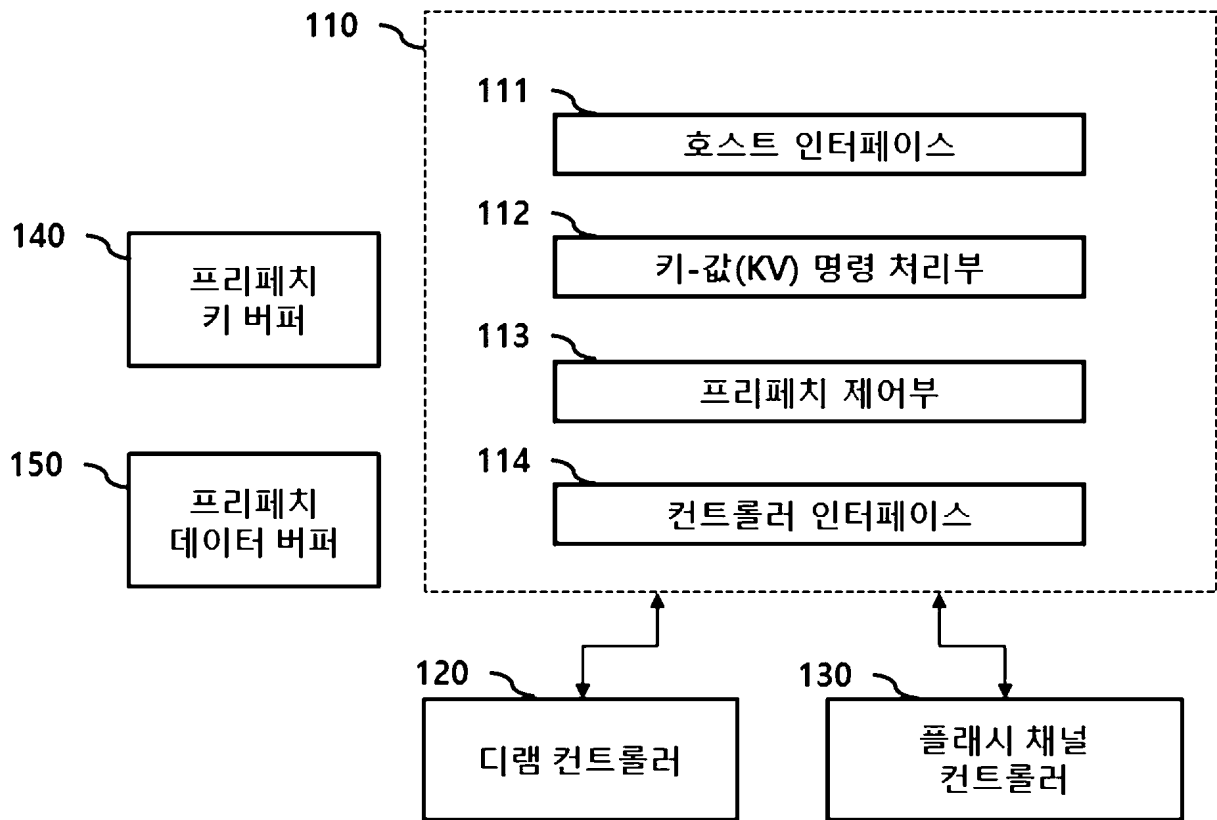
320: 값 로그 영역

도면

도면1

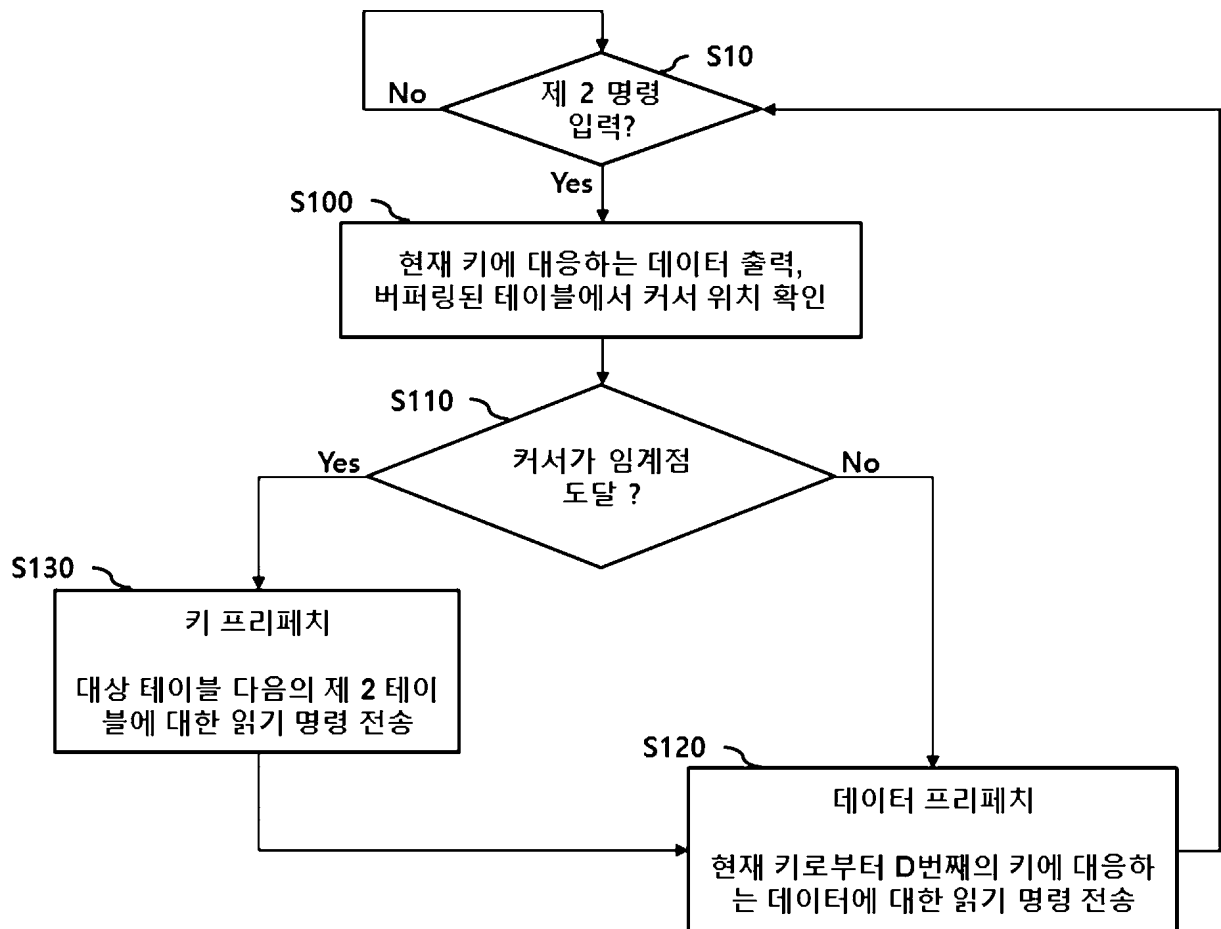


도면2

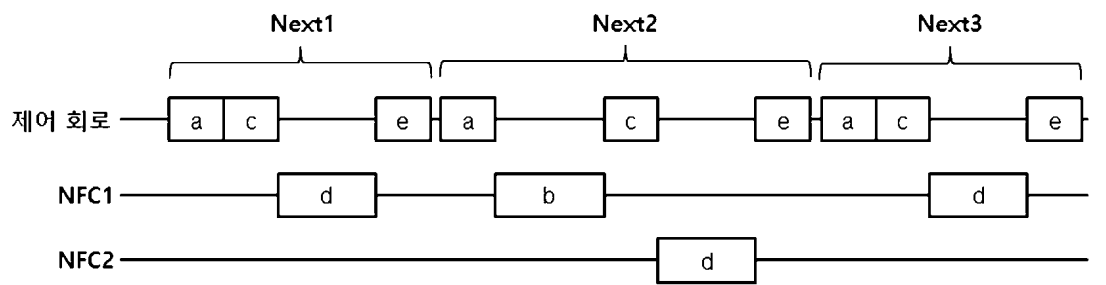


100

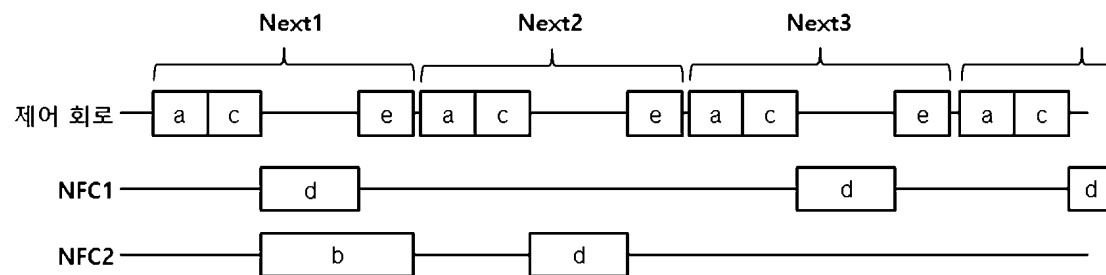
도면3



도면4

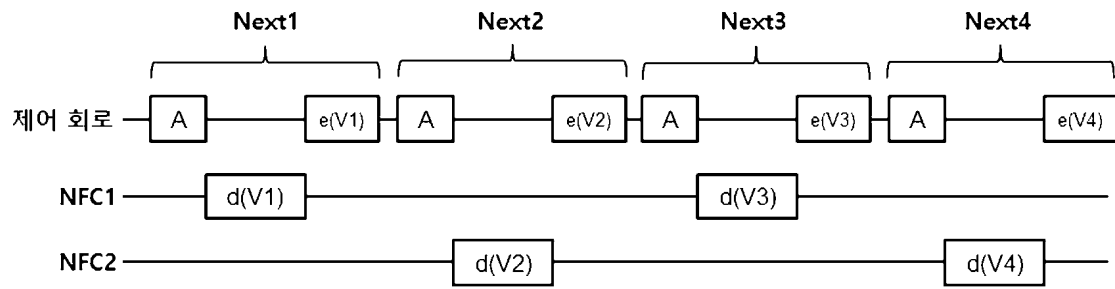


(A) 키 프리페치 동작을 수행하지 않는 경우

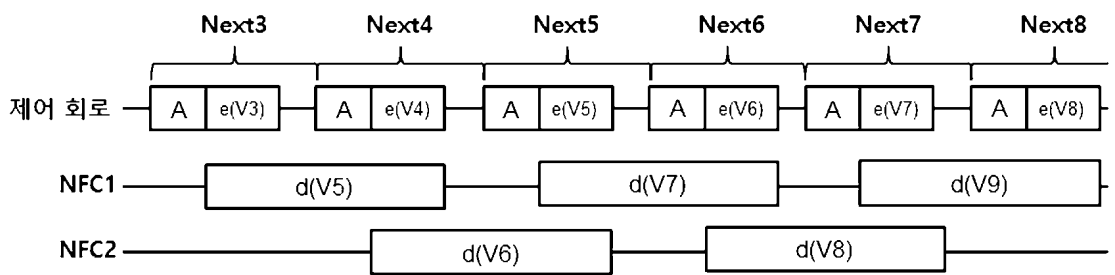


(B) 키 프리페치 동작을 수행하는 경우

도면5



(A) 값 프리페치 동작을 수행하지 않는 경우



(B) 값 프리페치 동작을 수행하는 경우