



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2014-0104882
(43) 공개일자 2014년08월29일

(51) 국제특허분류(Int. Cl.)
G06F 17/30 (2006.01) G06F 17/40 (2006.01)
(21) 출원번호 10-2013-0066978
(22) 출원일자 2013년06월12일
심사청구일자 2014년05월16일
(30) 우선권주장
1020130016699 2013년02월18일 대한민국(KR)

(71) 출원인
단국대학교 산학협력단
경기 용인시 수지구 죽전로 152, 내 (죽전동, 단국대학교)
(72) 발명자
민영근
서울 노원구 한글비석로48길 6, 6동 316호 (상계동, 한신2차아파트)
김윤호
경기도 파주시 다율동 월드메르디앙2차아파트 204동 303호
(74) 대리인
특허법인이상

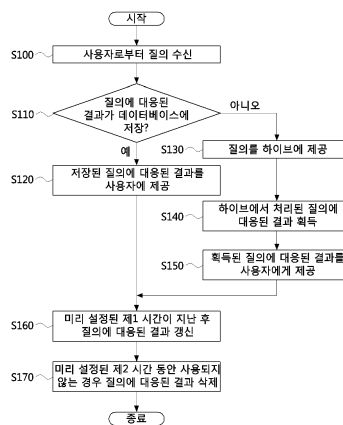
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 **하둑 기반의 질의 처리 방법 및 장치**

(57) 요약

하둑 기반의 질의 처리 방법 및 장치가 개시된다. 질의 처리 방법은, 사용자 단말로부터 질의를 수신하는 단계, 질의에 대응된 결과가 질의 처리 장치 내의 데이터베이스에 저장되어 있는지 판단하는 단계 및 질의에 대응된 결과가 데이터베이스에 저장되어 있는 경우, 저장된 질의에 대응된 결과를 사용자 단말에 제공하는 단계를 포함한다. 따라서, 사용자의 질의를 신속하게 처리할 수 있다.

대표도 - 도2



이 발명을 지원한 국가연구개발사업

과제고유번호 10039143

부처명 지식경제부

연구사업명 산업융합원천기술개발사업

연구과제명 데이터 센트릭 컴퓨팅을 위한 매니코어 시스템용 GBps급 병렬 I/O 가상화 및 DB 성능 최적
화 기술개발

기여율 1/1

주관기관 단국대학교 산학협력단

연구기간 2011.05.01 ~ 2014.02.28

특허청구의 범위

청구항 1

하둡(hadoop) 기반의 질의 처리 장치에서 수행되는 질의 처리 방법에 있어서,

사용자 단말로부터 질의(query)를 수신하는 단계;

상기 질의에 대응된 결과가 상기 질의 처리 장치 내의 데이터베이스(database)에 저장되어 있는지 판단하는 단계; 및

상기 질의에 대응된 결과가 상기 데이터베이스에 저장되어 있는 경우, 저장된 상기 질의에 대응된 결과를 상기 사용자 단말에 제공하는 단계를 포함하는 것을 특징으로 하는 질의 처리 방법.

청구항 2

청구항 1에 있어서,

상기 질의 처리 방법은,

상기 질의에 대응된 결과가 상기 데이터베이스에 저장되어 있지 않은 경우, 상기 질의를 하이브(hive)에 제공하는 단계;

상기 하이브를 통해 처리된 상기 질의에 대응된 결과를 획득하는 단계; 및

상기 질의에 대응된 결과를 상기 사용자 단말에 제공하는 단계를 더 포함하는 것을 특징으로 하는 질의 처리 방법.

청구항 3

청구항 2에 있어서,

상기 질의 처리 방법은,

상기 하이브를 통해 처리된 상기 질의에 대응된 결과를 상기 데이터베이스에 저장하는 단계를 더 포함하는 것을 특징으로 하는 질의 처리 방법.

청구항 4

청구항 1에 있어서,

상기 질의 처리 방법은,

상기 데이터베이스에 저장된 임의의 질의에 대응된 결과가 미리 설정된 시간 동안 사용되지 않는 경우, 상기 임의의 질의에 대응된 결과를 삭제하는 것을 특징으로 하는 질의 처리 방법.

청구항 5

청구항 1에 있어서,

상기 질의 처리 방법은,

임의의 질의에 대응된 결과가 상기 데이터베이스에 저장된 시점부터 미리 설정된 시간이 지난 경우, 상기 임의의 질의에 대응된 결과를 갱신하는 것을 특징으로 하는 질의 처리 방법.

청구항 6

청구항 1에 있어서,

상기 데이터베이스는,

상기 질의와 상기 질의에 대응된 결과를 기반으로 생성된 관계형 데이터베이스(relational database)인 것을 특징으로 하는 질의 처리 방법.

청구항 7

하둡(hadoop) 기반의 질의 처리 장치에 있어서,

사용자 단말로부터 수신한 질의를 분석하여 상기 질의에 대응된 결과가 데이터베이스(database)에 저장되어 있는지 판단하고, 상기 질의에 대응된 결과가 상기 데이터베이스에 저장되어 있는 경우 상기 질의에 대응된 결과를 상기 사용자 단말에 제공하는 프로세서(processor); 및

질의 및 질의에 대응된 결과를 저장하는 상기 데이터베이스를 포함하는 것을 특징으로 하는 질의 처리 장치.

청구항 8

청구항 7에 있어서,

상기 프로세서는,

상기 질의에 대응된 결과가 상기 데이터베이스에 저장되어 있지 않은 경우 상기 질의를 하이브(hive)에 제공하고, 상기 하이브를 통해 처리된 상기 질의에 대응된 결과를 획득하고, 상기 질의에 대응된 결과를 상기 사용자 단말에 제공하는 것을 특징으로 하는 질의 처리 장치.

청구항 9

청구항 8에 있어서,

상기 데이터베이스는,

상기 하이브를 통해 처리된 상기 질의에 대응된 결과를 저장하는 것을 특징으로 하는 질의 처리 장치.

청구항 10

청구항 7에 있어서,

상기 프로세서는,

상기 데이터베이스에 저장된 임의의 질의에 대응된 결과가 미리 설정된 시간 동안 사용되지 않는 경우, 상기 임의의 질의에 대응된 결과를 삭제하는 것을 특징으로 하는 질의 처리 장치.

청구항 11

청구항 7에 있어서,

상기 프로세서는,

임의의 질의에 대응된 결과가 상기 데이터베이스에 저장된 시점부터 미리 설정된 시간이 지난 경우, 상기 임의의 질의에 대응된 결과를 갱신하는 것을 특징으로 하는 질의 처리 장치.

청구항 12

청구항 7에 있어서,

상기 데이터베이스는,

상기 질의와 상기 질의에 대응된 결과를 기반으로 생성된 관계형 데이터베이스(relational database)인 것을 특징으로 하는 질의 처리 장치.

명세서

기술분야

[0001] 본 발명은 질의 처리 방법 및 장치에 관한 것으로, 더욱 상세하게는 빅 데이터를 처리하기 위해 사용되는 하둡을 기초로 하여 사용자의 질의를 신속하게 처리하기 위한 질의 처리 방법 및 장치에 관한 것이다.

배경기술

- [0002] 하둡(hadoop)은 빅 데이터(big data)를 처리하기 위해 사용되는 기술 중 하나로, 크게 하둡 분산 파일 시스템(hadoop distributed file system, HDFS)과 맵 리듀스(map reduce)로 구성된다.
- [0003] 맵 리듀스는 실제 빅 데이터를 처리하는 프로그램으로서, 맵 단계와 리듀스 단계로 구분할 수 있다. 맵 단계에서는 입력 데이터를 키-값(key-value)으로 매핑(mapping)하며, 특정 키에 대한 값들을 병합하여 중간 데이터로 출력한다. 리듀스 단계에서는 맵 단계를 통해 생성된 중간 데이터를 입력으로 받아 사용자가 원하는 방향으로 최종 분석, 처리하여 그 결과를 출력한다.
- [0004] 맵 리듀스 프로그램은 일반적으로 자바(java) 프로그래밍 언어를 통해 개발된다. 즉, 빅 데이터를 분석 및 처리하기 위해 사용자는 자바 프로그래밍 언어를 다룰 줄 알아야 하며, 이는 기존의 SQL(structured query language)을 통해 데이터를 분석하던 사용자의 입장에서 큰 진입 장벽이 된다.
- [0005] 이러한 문제점을 해결하기 위해 하이브(hive)가 개발되었으며, 하이브는 HiveQL(hive query language)라고 불리는 SQL과 상당히 유사한 인터페이스(interface)를 제공한다. 빅 데이터를 분석 및 처리하기 위해 HiveQL로 작성된 질의는 내부적으로 낮은-레벨(low-level)에서 맵 리듀스 잡(job)으로 변환되어 동작하게 된다. 이로 인해, 실제 빅 데이터의 사용자는 자바 프로그래밍 언어를 사용하여 맵 리듀스 잡을 작성할 필요 없이 HiveQL과 같은 질의 자체에만 집중할 수 있다.
- [0006] 이러한 하이브를 사용함으로써 맵 리듀스 프로그램에 대한 접근성이 향상되었으나, 하이브를 사용하는 경우에도 질의를 처리하기 위해 맵 리듀스 프로그램을 수행해야 하므로, 실질적으로 빅 데이터의 분석 및 처리 성능이 향상되지 못하는 문제점이 있다.

발명의 내용

해결하려는 과제

- [0007] 상기와 같은 문제점을 해결하기 위한 본 발명의 목적은, 관계형 데이터베이스를 통해 사용자의 질의를 신속히 처리하기 위한 질의 처리 방법을 제공하는 데 있다.
- [0008] 상기와 같은 문제점을 해결하기 위한 본 발명의 다른 목적은, 관계형 데이터베이스를 통해 사용자의 질의를 신속히 처리하기 위한 질의 처리 장치를 제공하는 데 있다.

과제의 해결 수단

- [0009] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 질의 처리 방법은, 사용자 단말로부터 질의를 수신하는 단계, 상기 질의에 대응된 결과가 상기 질의 처리 장치 내의 데이터베이스에 저장되어 있는지 판단하는 단계 및 상기 질의에 대응된 결과가 상기 데이터베이스에 저장되어 있는 경우, 저장된 상기 질의에 대응된 결과를 상기 사용자 단말에 제공하는 단계를 포함한다.
- [0010] 여기서, 상기 질의 처리 방법은, 상기 질의에 대응된 결과가 상기 데이터베이스에 저장되어 있지 않은 경우, 상기 질의를 하이브에 제공하는 단계, 상기 하이브를 통해 처리된 상기 질의에 대응된 결과를 획득하는 단계 및 상기 질의에 대응된 결과를 상기 사용자 단말에 제공하는 단계를 더 포함할 수 있다.
- [0011] 여기서, 상기 질의 처리 방법은, 상기 하이브를 통해 처리된 상기 질의에 대응된 결과를 상기 데이터베이스에 저장하는 단계를 더 포함할 수 있다.
- [0012] 여기서, 상기 질의 처리 방법은, 상기 데이터베이스에 저장된 임의의 질의에 대응된 결과가 미리 설정된 시간 동안 사용되지 않는 경우, 상기 임의의 질의에 대응된 결과를 삭제할 수 있다.
- [0013] 여기서, 상기 질의 처리 방법은, 임의의 질의에 대응된 결과가 상기 데이터베이스에 저장된 시점부터 미리 설정된 시간이 지난 경우, 상기 임의의 질의에 대응된 결과를 갱신할 수 있다.
- [0014] 여기서, 상기 데이터베이스는, 상기 질의와 상기 질의에 대응된 결과를 기반으로 생성된 관계형 데이터베이스일 수 있다.
- [0015] 상기 다른 목적을 달성하기 위한 본 발명의 일 실시예에 따른 질의 처리 장치는, 사용자 단말로부터 수신한 질의를 분석하여 상기 질의에 대응된 결과가 데이터베이스에 저장되어 있는지 판단하고, 상기 질의에 대응된 결과가 상기 데이터베이스에 저장되어 있는 경우 상기 질의에 대응된 결과를 상기 사용자 단말에 제공하는 프로세서, 및 질의 및 질의에 대응된 결과를 저장하는 상기 데이터베이스를 포함한다.

- [0016] 여기서, 상기 프로세서는, 상기 질의에 대응된 결과가 상기 데이터베이스에 저장되어 있지 않은 경우 상기 질의를 하이브에 제공하고, 상기 하이브를 통해 처리된 상기 질의에 대응된 결과를 획득하고, 상기 질의에 대응된 결과를 상기 사용자 단말에 제공할 수 있다.
- [0017] 여기서, 상기 데이터베이스는, 상기 하이브를 통해 처리된 상기 질의에 대응된 결과를 저장할 수 있다.
- [0018] 여기서, 상기 프로세서는, 상기 데이터베이스에 저장된 임의의 질의에 대응된 결과가 미리 설정된 시간 동안 사용되지 않는 경우, 상기 임의의 질의에 대응된 결과를 삭제할 수 있다.
- [0019] 여기서, 상기 프로세서는, 임의의 질의에 대응된 결과가 상기 데이터베이스에 저장된 시점부터 미리 설정된 시간이 지난 경우, 상기 임의의 질의에 대응된 결과를 갱신할 수 있다.
- [0020] 여기서, 상기 데이터베이스는, 상기 질의와 상기 질의에 대응된 결과를 기반으로 생성된 관계형 데이터베이스일 수 있다.

발명의 효과

- [0021] 본 발명에 의하면, 맵 리듀스 과정을 통해 처리된 질의에 대한 정보를 포함한 관계형 데이터베이스를 구축할 수 있고, 추후 동일하거나 유사한 질의에 대해서는 맵 리듀스 과정을 거치지 않고 관계형 데이터베이스를 사용하여 처리할 수 있으므로, 사용자의 질의를 신속하게 처리할 수 있다.

도면의 간단한 설명

- [0022] 도 1은 하둡 기반의 질의 처리 시스템을 도시한 블록도이다.
- 도 2는 본 발명의 일 실시예에 따른 질의 처리 방법을 도시한 흐름도이다.
- 도 3은 완전 부분 질의에 대한 일 실시예를 도시한 개념도이다.
- 도 4는 불완전 부분 질의에 대한 일 실시예를 도시한 개념도이다.
- 도 5는 관계형 데이터베이스와 하이브의 관계를 도시한 개념도이다.
- 도 6은 본 발명의 일 실시예에 따른 질의 처리 장치를 도시한 블록도이다.
- 도 7은 본 발명의 일 실시예에 따른 프로세서의 구성을 도시한 블록도이다.
- 도 8은 본 발명의 일 실시예에 따른 질의 처리 장치가 적용된 하둡 기반의 질의 처리 시스템을 도시한 블록도이다.

발명을 실시하기 위한 구체적인 내용

- [0023] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세하게 설명하고자 한다.
- [0024] 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.
- [0025] 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.
- [0026] 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.
- [0027] 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 조

합한 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.

- [0028] 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.
- [0029] 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.
- [0030] 도 1은 하둡(hadoop) 기반의 질의 처리 시스템을 도시한 블록도이다.
- [0031] 도 1을 참조하면, 사용자 단말(10)은 질의(query)를 질의 처리 장치(20)에 제공할 수 있다. 사용자 단말(10)은 데스크탑 컴퓨터(Desktop Computer), 랩탑 컴퓨터(Laptop Computer), 태블릿(Tablet) PC 등을 의미할 수 있다.
- [0032] 질의 처리 장치(20)는 사용자 단말(10)로부터 수신한 질의에 대응된 결과가 질의 처리 장치(20) 내에 저장되어 있는 경우, 저장된 질의에 대응된 결과를 사용자 단말(10)에 제공할 수 있다. 한편, 사용자 단말(10)로부터 수신한 질의에 대응된 결과가 질의 처리 장치(20) 내에 저장되어 있지 않은 경우, 질의 처리 장치(20)는 질의를 하이브(hive)(30)에 제공할 수 있다. 질의 처리 장치(20)의 구체적인 구성 및 각 구성의 기능에 대해서는 도 6을 참조하여 후술하도록 한다.
- [0033] 하이브(30)는 질의 처리 장치(20)로부터 수신한 질의에 대응된 결과가 하이브(30) 내에 저장되어 있는 경우, 저장된 질의에 대응된 결과를 질의 처리 장치(20)에 제공할 수 있다. 한편, 질의 처리 장치(20)로부터 수신한 질의에 대응된 결과가 하이브(30) 내에 저장되어 있지 않은 경우, 하이브(30)는 하둡(40)을 기초로 하여 질의를 처리할 수 있다.
- [0034] 하이브(30)는 SQL(structured query language)와 유사한 HiveQL(hive query language) 기반의 인터페이스(interface)를 제공하며, 빅 데이터를 분석 및 처리하기 위해 HiveQL로 작성된 질의는 내부적으로 낮은-레벨(low-level)에서 맵 리듀스(map reduce) 작업으로 변환되어 동작하게 된다.
- [0035] 하둡(40)은 크게 하둡 분산 파일 시스템(hadoop distributed file system, HDFS)과 맵 리듀스로 구성된다. 하둡(40)은 하이브(30)로부터 수신된 질의를 맵 리듀스를 기반으로 처리할 수 있고, 처리된 결과를 하이브(30)에 제공할 수 있다.
- [0036] 도 2는 본 발명의 일 실시예에 따른 질의 처리 방법을 도시한 흐름도이다.
- [0037] 도 2를 참조하면, 하둡 기반의 질의 처리 방법은, 사용자 단말로부터 질의를 수신하는 단계(S100), 질의에 대응된 결과가 데이터베이스(database)에 저장되어 있는지 판단하는 단계(S110) 및 질의에 대응된 결과가 데이터베이스에 저장되어 있는 경우, 저장된 질의에 대응된 결과를 사용자 단말에 제공하는 단계(S120)를 포함한다.
- [0038] 또한, 하둡 기반의 질의 처리 방법은, 질의에 대응된 결과가 데이터베이스에 저장되어 있지 않은 경우 질의를 하이브에 제공하는 단계(S130), 하이브를 통해 처리된 질의에 대응된 결과를 획득하는 단계(S140) 및 질의에 대응된 결과를 사용자 단말에 제공하는 단계(S150)를 포함할 수 있다.
- [0039] 하둡 기반의 질의 처리 방법은 도 6에 도시된 질의 처리 장치(20)에서 수행될 수 있으며, 질의 처리 장치(20)는 프로세서(processor)(21) 및 데이터베이스(22)를 포함한다.
- [0040] 단계 S100에서, 질의 처리 장치는 사용자 단말로부터 질의를 수신할 수 있다. 사용자 단말은 질의에 대응된 결과를 요청하기 위해 사용자가 사용하는 것으로, 데스크탑 컴퓨터, 랩탑 컴퓨터, 태블릿 PC 등을 의미할 수 있다.
- [0041] 단계 S110에서, 질의 처리 장치는 사용자 단말로부터 수신된 질의에 대응된 결과가 질의 처리 장치 내의 데이터베이스에 저장되어 있는지 판단할 수 있다. 예를 들어, 사용자 단말로부터 수신된 질의가 'SELECT year, month

FROM airline_delay WHERE year=2013' 인 경우, 질의 처리 장치는 SELECT 구문, FROM 구문, WHERE 구문별로 질의에 대응된 결과가 질의 처리 장치 내의 데이터베이스에 저장되어 있는지 판단할 수 있다.

- [0042] 사용자 단말로부터 수신한 질의가 질의 처리 장치 내의 데이터베이스에 저장된 질의에 포함되는 경우(즉, 완전 부분 질의), 질의 처리 장치는 사용자 단말로부터 수신된 질의에 대응된 결과가 자신의 데이터베이스에 모두 저장되어 있는 것으로 판단할 수 있다.
- [0043] 도 3은 완전 부분 질의에 대한 일 실시예를 도시한 개념도이다.
- [0044] 도 3을 참조하면, 사용자 단말로부터 수신한 질의(300)가 'SELECT Col1, Col2', 'FROM tb11', 'WHERE Col1>100'이고 질의 처리 장치 내의 데이터베이스에 저장된 질의(200)가 'SELECT Col1, Col2, Col3', 'FROM tb11'인 경우, 사용자 단말로부터 수신한 질의(300)는 질의 처리 장치 내의 데이터베이스에 저장된 질의(200)에 포함되므로(즉, 완전 부분 질의), 질의 처리 장치는 사용자 단말로부터 수신된 질의에 대응된 결과가 자신의 데이터베이스에 모두 저장되어 있는 것으로 판단할 수 있다. 따라서, 이와 같은 경우 질의 처리 장치는 다음 단계로 S120을 수행할 수 있다.
- [0045] 한편, 사용자 단말로부터 수신한 질의가 질의 처리 장치 내의 데이터베이스에 저장된 질의의 일부인 경우(즉, 불완전 부분 질의), 질의 처리 장치는 사용자 단말로부터 수신된 질의에 대응된 결과가 자신의 데이터베이스에 일부 저장되어 있는 것으로 판단할 수 있다.
- [0046] 도 4는 불완전 부분 질의에 대한 일 실시예를 도시한 개념도이다.
- [0047] 도 4를 참조하면, 사용자 단말로부터 수신한 질의(300)가 'SELECT Col1, Col2, Col3', 'FROM tb11', 'WHERE Col1>50'이고 질의 처리 장치 내의 데이터베이스에 저장된 질의(200)가 'SELECT Col1, Col2', 'FROM tb11', 'WHERE Col1<100'인 경우, 사용자 단말로부터 수신한 질의(300)는 질의 처리 장치 내의 데이터베이스에 저장된 질의(200)에 일부 포함되므로(즉, 불완전 부분 질의), 질의 처리 장치는 사용자 단말로부터 수신된 질의에 대응된 결과가 자신의 데이터베이스에 일부 저장되어 있는 것으로 판단할 수 있다.
- [0048] 이와 같은 경우, 질의 처리 장치는 자신의 데이터베이스에 저장된 질의에 대해서는 단계 S120을 통해 처리(즉, 자신의 데이터베이스를 사용하여 처리)할 수 있고, 자신의 데이터베이스에 저장되지 않은 질의에 대해서는 단계 S130, 단계 S140, 단계 S150을 통해 처리(즉, 하이브를 통해 처리)할 수 있다.
- [0049] 다시 도 2를 참조하면, 단계 S120에서 질의 처리 장치는 자신의 데이터베이스에 저장된 질의에 대응된 결과를 사용자 단말에 제공할 수 있다. 즉, 질의 처리 장치는 사용자 단말로부터 수신한 질의에 대응한 질의를 자신의 데이터베이스 내에서 검색하고, 검색된 질의에 대응된 결과를 사용자 단말에 제공할 수 있다.
- [0050] 단계 S130에서, 질의 처리 장치는 자신의 데이터베이스에 저장되지 않은 질의를 하이브에 제공할 수 있다. 하이브는 자신의 스토리지(storage) 내에 질의 처리 장치로부터 수신한 질의에 대응된 결과가 저장되어 있는 경우, 저장된 질의에 대응된 결과를 질의 처리 장치에 제공할 수 있다. 한편, 자신의 스토리지 내에 질의 처리 장치로부터 수신한 질의에 대응된 결과가 저장되어 있지 않은 경우, 하이브는 맵 리듀스를 사용하여 질의에 대응된 결과를 생성할 수 있고, 생성된 질의에 대응된 결과를 질의 처리 장치에 제공할 수 있다.
- [0051] 단계 S140에서, 질의 처리 장치는 하이브에서 처리된 질의에 대응된 결과를 획득할 수 있다. 예를 들어, 질의 처리 장치는 하이브에서 처리된 질의에 대응된 결과를 오픈 소스(open source)인 스콥(sqoop)을 사용하여 데이터베이스로 캐싱(caching)할 수 있다. 즉, 질의 처리 장치는 하이브에서 처리된 질의에 대응된 결과를 데이터베이스에 저장할 수 있다.
- [0052] 단계 S150에서, 질의 처리 장치는 하이브로부터 획득된 질의에 대응된 결과를 사용자 단말에 제공할 수 있다.
- [0053] 여기서, 데이터베이스는 질의 처리 장치에 포함된 하나의 구성으로, 관계형 데이터베이스(relational database)를 의미할 수 있다. 데이터베이스는 하이브에서 처리된 결과를 기반으로 미리 구축될 수 있다. 즉, 질의 처리 장치는 하이브를 통해 처리된 질의, 질의에 대응된 결과를 저장하여 데이터베이스를 구축할 수 있다.
- [0054] 도 5는 관계형 데이터베이스와 하이브의 관계를 도시한 개념도이다.

- [0055] 도 5를 참조하면, 관계형 데이터베이스(22)와 하이브(30)는 논리적으로 계층적 구조를 가질 수 있다. 즉, 관계형 데이터베이스(22)는 논리적으로 하이브(30)의 상부에 위치할 수 있다. 관계형 데이터베이스(22)는 하이브(30)보다 빠른 데이터 입출력 속도를 가지는 스토리지를 의미할 수 있고, 하이브(30)는 관계형 데이터베이스(22)보다 느린 데이터 입출력 속도를 가지는 스토리지를 의미할 수 있다. 하이브(30)에 저장된 질의에 대응된 결과들 중에서 자주 사용되는 질의에 대응된 결과는 관계형 데이터베이스(22)에 저장(즉, 캐시(cache))될 수 있으며, 이러한 캐싱 과정은 도 6에 도시된 질의 처리 장치에 의해 수행될 수 있다.
- [0056] 이와 같이 관계형 데이터베이스를 구축함으로써, 질의 처리 장치는 사용자 단말로부터 수신한 질의를 먼저 관계형 데이터베이스(22)를 통해 처리할 수 있고, 관계형 데이터베이스(22)를 통해 처리할 수 없는 경우(즉, 질의에 대응된 결과가 관계형 데이터베이스(22)에 저장되어 있지 않은 경우) 하이브(30)를 통해 질의를 처리할 수 있다.
- [0057] 다시 도 2를 참조하면, 하둡 기반의 질의 처리 방법은 미리 설정된 제1 시간이 지난 경우 질의 처리 장치 내의 데이터베이스에 저장된 질의에 대응된 결과를 갱신하는 단계(S160)를 더 포함할 수 있다. 예를 들어, 제1 질의에 대응된 결과가 질의 처리 장치 내의 데이터베이스에 저장된 시점부터 미리 설정된 제1 시간이 지난 경우, 질의 처리 장치는 제1 질의와 대응된 질의를 하이브에서 검색할 수 있고, 검색된 질의에 대응된 결과를 하이브로부터 캐싱할 수 있다. 이때, 질의 처리 장치는 하이브에서 검색된 질의에 대응된 결과와 자신의 데이터베이스에 저장된 제1 질의에 대응된 결과가 다른 경우 검색된 질의에 대응된 결과를 하이브로부터 캐싱할 수 있으며, 하이브에서 검색된 질의에 대응된 결과와 자신의 데이터베이스에 저장된 제1 질의에 대응된 결과가 동일한 경우 캐싱 과정을 수행하지 않을 수 있다. 미리 설정된 제1 시간은 사용자의 요구에 따라 다양하게 설정될 수 있다.
- [0058] 이와 같은 갱신 과정을 수행함으로써, 질의 처리 장치는 자신의 데이터베이스에 저장된 정보를 최신의 정보로 유지할 수 있다.
- [0059] 하둡 기반의 질의 처리 방법은 질의 처리 장치 내의 데이터베이스에 저장된 질의에 대응된 결과가 미리 설정된 제2 시간 동안 사용되지 않는 경우 질의에 대응된 결과를 삭제하는 단계(S170)를 더 포함할 수 있다. 예를 들어, 제1 질의에 대응된 결과가 질의 처리 장치 내의 데이터베이스에 저장된 시점(또는, 제1 질의에 대응된 결과가 이전에 사용된 시점)부터 미리 설정된 제2 시간 동안 사용되지 않은 경우, 질의 처리 장치는 자신의 데이터베이스에 저장된 제1 질의에 대응된 결과를 삭제할 수 있다. 미리 설정된 제2 시간은 사용자의 요구에 따라 다양하게 설정될 수 있으며, 미리 설정된 제2 시간은 미리 설정된 제1 시간보다 길게 설정될 수 있다.
- [0060] 이와 같은 삭제 과정을 수행함으로써, 질의 처리 장치는 불필요한 정보가 데이터베이스에 저장되는 것을 방지할 수 있다.
- [0061] 본 발명에 따른 질의 처리 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.
- [0062] 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.
- [0063] 도 6은 본 발명의 일 실시예에 따른 질의 처리 장치를 도시한 블록도이다.
- [0064] 도 6을 참조하면, 질의 처리 장치(20)는 프로세서(21) 및 데이터베이스(22)를 포함한다. 프로세서(21)는 사용자 단말로부터 수신한 질의를 분석하여 질의에 대응된 결과가 데이터베이스(22)에 저장되어 있는지 판단할 수 있고, 질의에 대응된 결과가 데이터베이스(22)에 저장되어 있는 경우 질의에 대응된 결과를 상기 사용자 단말에 제공할 수 있다.

- [0065] 프로세서(21)는 사용자 단말로부터 질의를 수신할 수 있다. 사용자 단말은 질의에 대응된 결과를 요청하기 위해 사용자가 사용하는 것으로, 데스크탑 컴퓨터, 랩탑 컴퓨터, 태블릿 PC 등을 의미할 수 있다.
- [0066] 프로세서(21)는 사용자 단말로부터 수신된 질의에 대응된 결과가 데이터베이스(22)에 저장되어 있는지 판단할 수 있다. 예를 들어, 사용자 단말로부터 수신된 질의가 'SELECT year, month FROM airline_delay WHERE year=2013' 인 경우, 프로세서(21)는 SELECT 구문, FROM 구문, WHERE 구문별로 질의에 대응된 결과가 데이터베이스(22)에 저장되어 있는지 판단할 수 있다.
- [0067] 사용자 단말로부터 수신한 질의가 데이터베이스(22)에 저장된 질의에 포함되는 경우(즉, 완전 부분 질의), 프로세서(21)는 사용자 단말로부터 수신된 질의에 대응된 결과가 데이터베이스(22)에 모두 저장되어 있는 것으로 판단할 수 있다. 예를 들어, 앞서 설명한 도 3에 도시된 경우에서, 프로세서(21)는 사용자 단말로부터 수신된 질의에 대응된 결과가 데이터베이스(22)에 모두 저장되어 있는 것으로 판단할 수 있다.
- [0068] 한편, 사용자 단말로부터 수신한 질의가 데이터베이스(22)에 저장된 질의의 일부인 경우(즉, 불완전 부분 질의), 프로세서(21)는 사용자 단말로부터 수신된 질의에 대응된 결과가 데이터베이스(22)에 일부 저장되어 있는 것으로 판단할 수 있다. 예를 들어, 앞서 설명한 도 4에 도시된 경우에서, 프로세서(21)는 사용자 단말로부터 수신된 질의에 대응된 결과가 데이터베이스(22)에 일부 저장되어 있는 것으로 판단할 수 있다.
- [0069] 이와 같은 경우, 프로세서(21)는 데이터베이스(22)에 저장된 질의에 대해서는 데이터베이스(22)를 사용하여 처리할 수 있고, 데이터베이스(22)에 저장되지 않은 질의에 대해서는 하이브를 통해 처리할 수 있다.
- [0070] 질의에 대응된 결과가 데이터베이스(22)에 저장되어 있는 경우, 프로세서(21)는 데이터베이스(22)에 저장된 질의에 대응된 결과를 사용자 단말에 제공할 수 있다. 즉, 프로세서(21)는 사용자 단말로부터 수신한 질의에 대응한 질의를 데이터베이스(22) 내에서 검색하고, 검색된 질의에 대응된 결과를 사용자 단말에 제공할 수 있다.
- [0071] 한편 질의에 대응된 결과가 데이터베이스(22)에 저장되어 있지 않은 경우, 프로세서(21)는 질의를 하이브에 제공할 수 있다. 하이브는 자신의 스토리지 내에 프로세서(21)로부터 수신한 질의에 대응된 결과가 저장되어 있는 경우, 저장된 질의에 대응된 결과를 프로세서(21)에 제공할 수 있다. 한편, 자신의 스토리지 내에 프로세서(21)로부터 수신한 질의에 대응된 결과가 저장되어 있지 않은 경우, 하이브는 맵 리듀스를 사용하여 질의에 대응된 결과를 생성할 수 있고, 생성된 질의에 대응된 결과를 프로세서(21)에 제공할 수 있다.
- [0072] 프로세서(21)는 하이브에서 처리된 질의에 대응된 결과를 획득할 수 있다. 예를 들어, 프로세서(21)는 하이브에서 처리된 질의에 대응된 결과를 오픈 소스인 스쿼를 사용하여 데이터베이스(22)로 캐싱할 수 있다. 즉, 프로세서(21)는 하이브에서 처리된 질의에 대응된 결과를 데이터베이스(22)에 저장할 수 있다. 프로세서(21)는 하이브로부터 획득된 질의에 대응된 결과를 사용자 단말에 제공할 수 있다.
- [0073] 프로세서(21)는 미리 설정된 제1 시간이 지난 경우 데이터베이스(22)에 저장된 질의에 대응된 결과를 갱신할 수 있다. 예를 들어, 제1 질의에 대응된 결과가 데이터베이스(22)에 저장된 시점부터 미리 설정된 제1 시간이 지난 경우, 프로세서(21)는 제1 질의와 대응된 질의를 하이브에서 검색할 수 있고, 검색된 질의에 대응된 결과를 하이브로부터 캐싱할 수 있다. 이때, 프로세서(21)는 하이브에서 검색된 질의에 대응된 결과와 데이터베이스(22)에 저장된 제1 질의에 대응된 결과가 다른 경우 검색된 질의에 대응된 결과를 하이브로부터 캐싱할 수 있으며, 하이브에서 검색된 질의에 대응된 결과와 데이터베이스(22)에 저장된 제1 질의에 대응된 결과가 동일한 경우 캐싱 과정을 수행하지 않을 수 있다. 미리 설정된 제1 시간은 사용자의 요구에 따라 다양하게 설정될 수 있다.
- [0074] 프로세서(21)는 데이터베이스(22)에 저장된 질의에 대응된 결과가 미리 설정된 제2 시간 동안 사용되지 않는 경우 질의에 대응된 결과를 삭제할 수 있다. 예를 들어, 제1 질의에 대응된 결과가 데이터베이스(22)에 저장된 시점(또는, 제1 질의에 대응된 결과가 이전에 사용된 시점)부터 미리 설정된 제2 시간 동안 사용되지 않은 경우, 프로세서(21)는 데이터베이스(22)에 저장된 제1 질의에 대응된 결과를 삭제할 수 있다. 미리 설정된 제2 시간은 사용자의 요구에 따라 다양하게 설정될 수 있으며, 미리 설정된 제2 시간은 미리 설정된 제1 시간보다 길게 설정될 수 있다.
- [0075] 도 7은 본 발명의 일 실시예에 따른 프로세서의 구성을 도시한 블록도이다.
- [0076] 도 7을 참조하면, 프로세서(21)는 캐시 모듈(211), 하이브 모듈(212), 스쿼 모듈(213) 및 저장부(214)를 포함한다.

다. 여기서, 프로세서(21)는 도 6에 도시된 질의 처리 장치의 프로세서를 의미한다.

[0077] 캐시 모듈(211)은 질의에 대응된 결과가 질의 처리 장치 내의 데이터베이스에 저장되어 있는지 판단하는 과정을 수행할 수 있다. 하이브 모듈(212)은 질의가 하이브에서 처리되도록 제어할 수 있다. 스쿱 모듈(213)은 하이브에서 처리된 결과를 질의 처리 장치 내의 데이터베이스에 캐싱하는 과정을 수행할 수 있다. 저장부(214)는 캐시 모듈(211), 하이브 모듈(212) 및 스쿱 모듈(213)을 통해 처리되는 결과와 처리된 결과를 저장할 수 있다.

[0078] 다시 도 6을 참조하면, 데이터베이스(21)는 질의 및 질의에 대응된 결과를 저장할 수 있으며, 하이브를 통해 처리된 질의에 대응된 결과도 저장할 수 있다. 데이터베이스(21)는 질의와 질의에 대응된 결과를 기반으로 생성된 관계형 데이터베이스를 의미할 수 있다.

[0079] 도 8은 본 발명의 일 실시예에 따른 질의 처리 장치가 적용된 하둡 기반의 질의 처리 시스템을 도시한 블록도이다.

[0080] 도 8을 참조하면, 프로세서(21)는 도 6에 도시된 질의 처리 장치의 프로세서를 의미하고, RDBMS(relational database management system)(22)는 도 6에 도시된 질의 처리 장치의 데이터베이스를 의미한다.

[0081] 프로세서(21)는 사용자 단말(10)로부터 질의를 수신할 수 있고, 수신된 질의에 대응된 결과가 RDBMS(22)에 저장되어 있는 경우 저장된 질의에 대응된 결과를 사용자 단말(10)에 제공할 수 있다. 한편, 사용자 단말(10)로부터 수신된 질의에 대응된 결과가 RDBMS(22)에 저장되어 있지 않은 경우, 프로세서(21)는 하이브(30)에 질의를 제공할 수 있다. 하이브(30)는 하둡(40)을 기초로 하여 질의에 대응된 결과를 생성할 수 있고, 생성된 질의에 대응된 결과를 프로세서(21)에 제공할 수 있다. 즉, 프로세서(21)는 하이브(30)에 의해 생성된 질의에 대응된 결과를 오픈 소스인 스쿱을 사용하여 캐싱할 수 있다. 프로세서(21)는 하이브(30)로부터 캐싱된 질의에 대응된 결과를 RDBMS(22)에 저장할 수 있고, 질의에 대응된 결과를 사용자 단말(10)에 제공할 수 있다.

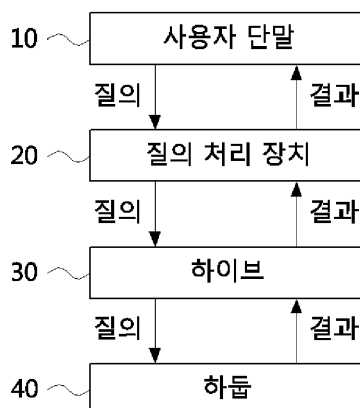
[0082] 이상 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

부호의 설명

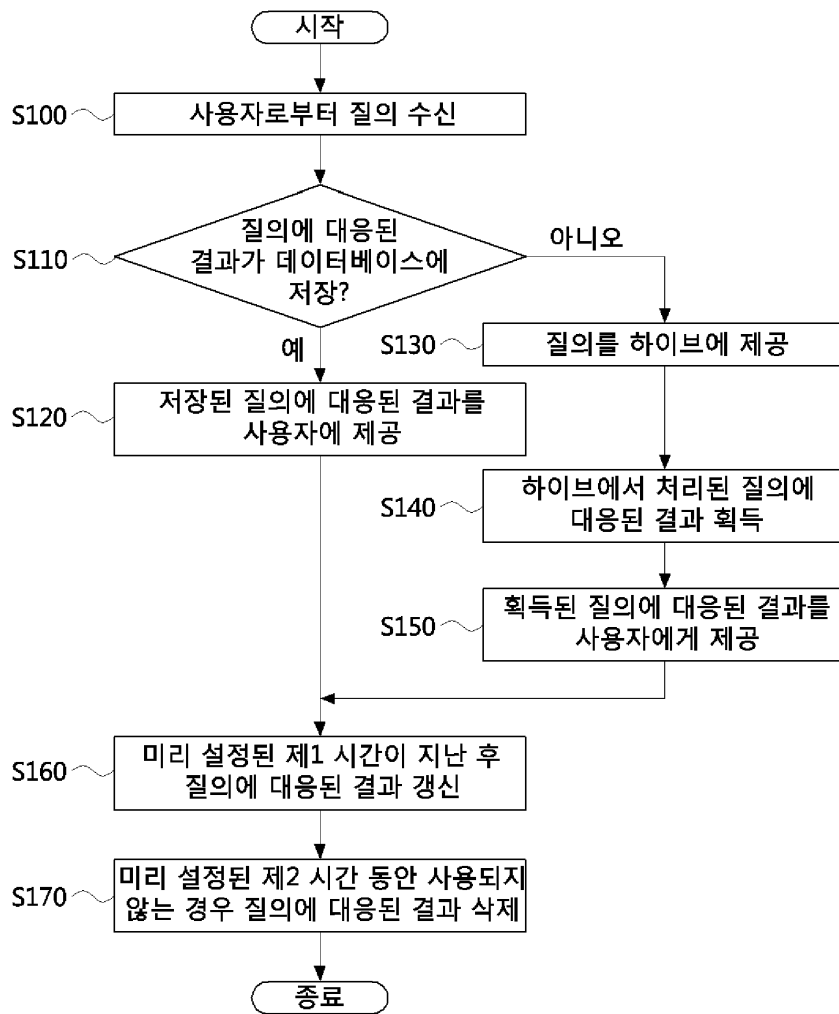
- [0083] 20: 질의 처리 장치
- 21: 프로세서
- 22: 데이터베이스

도면

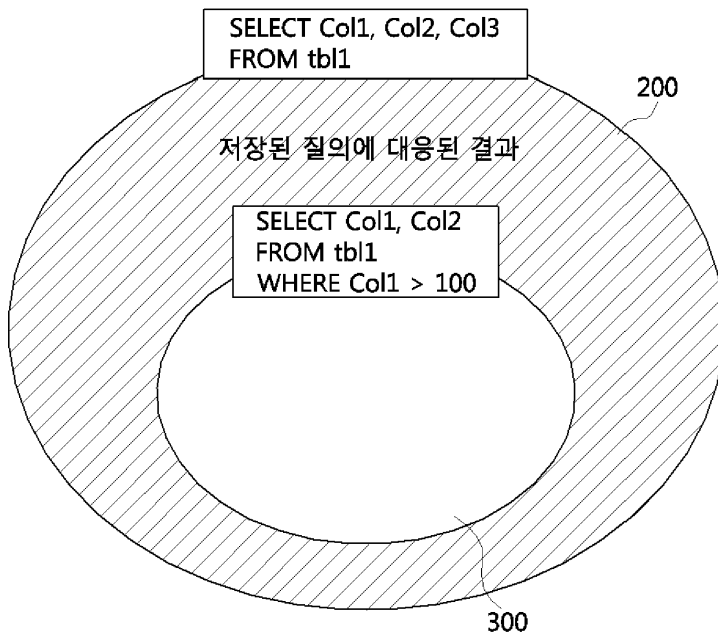
도면1



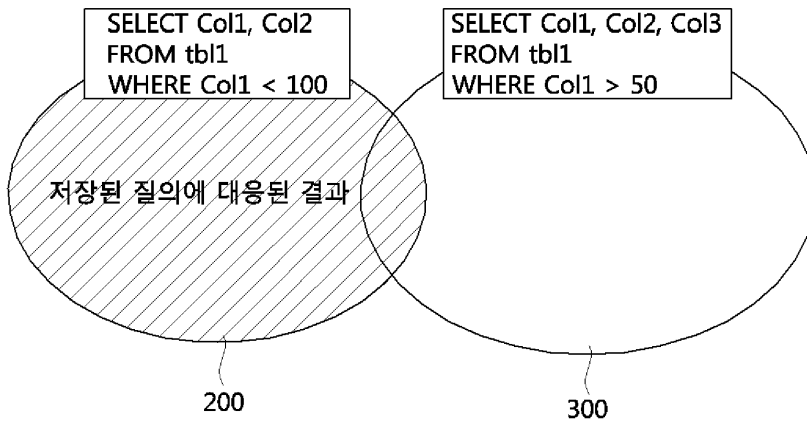
도면2



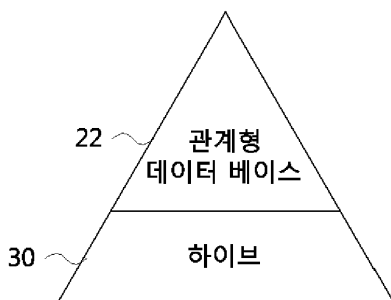
도면3



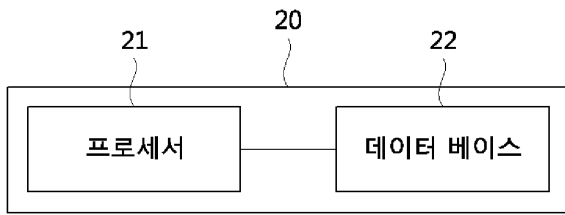
도면4



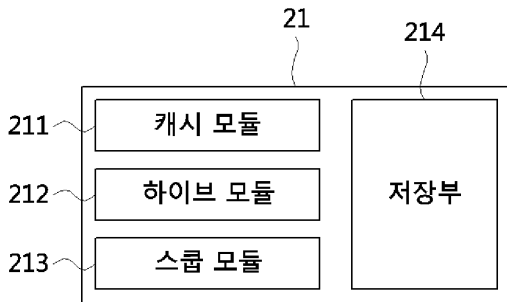
도면5



도면6



도면7



도면8

