

## 명세서

### 발명의 명칭: 데이터 적재 및 처리 시스템 및 그 방법

#### 기술분야

- [1] 본 발명은 데이터 적재 및 처리 시스템 및 그 방법에 관한 것으로서, 특히 수집 서버에서 데이터를 수집하기 전에 대상 리소스의 상태를 확인하고, 상기 확인된 리소스의 상태를 근거로 해당 리소스에 대한 데이터 수집 및 적재 방식을 결정하고, 데이터 서버에 데이터 수집 요청을 전달하고, 상기 결정된 데이터 수집 및 적재 방식으로 데이터 서버로부터 데이터를 수집하고, 상기 수집된 데이터를 빅데이터 플랫폼에 적재하는 데이터 적재 및 처리 시스템 및 그 방법에 관한 것이다.

#### 배경기술

- [2] 빅 데이터(big data)는 기존의 데이터베이스로는 수집 · 저장 · 분석 등을 수행하기가 어려울 만큼 방대한 양의 데이터를 나타낸다.
- [3] 이러한 빅 데이터는 복수의 사용자가 생성한 데이터를 수집, 저장 및 분석하기 위해서 추가 자원이 필요하며, 이는 데이터를 수집하는 것뿐만 아니라 데이터를 저장하고 분석하기 위한 플랫폼을 필요로 한다.

#### 발명의 상세한 설명

##### 기술적 과제

- [4] 본 발명의 목적은 수집 서버에서 데이터를 수집하기 전에 대상 리소스의 상태를 확인하고, 상기 확인된 리소스의 상태를 근거로 해당 리소스에 대한 데이터 수집 및 적재 방식을 결정하고, 데이터 서버에 데이터 수집 요청을 전달하고, 상기 결정된 데이터 수집 및 적재 방식으로 데이터 서버로부터 데이터를 수집하고, 상기 수집된 데이터를 빅데이터 플랫폼에 적재하는 데이터 적재 및 처리 시스템 및 그 방법을 제공하는 데 있다.

##### 과제 해결 수단

- [5] 본 발명의 실시예에 따른 데이터 적재 및 처리 시스템은 리소스에 포함되며, 데이터의 수집 대상인 데이터 서버; 상기 리소스에 포함되며, 데이터의 적재 대상인 빅데이터 플랫폼; 및 리소스의 사용량 및 데이터 량을 확인하고, 상기 확인된 리소스의 상태를 근거로 데이터 서버 및 빅데이터 플랫폼에 대한 데이터 수집 및 적재 방식을 결정하고, 상기 데이터 서버에 데이터 수집 요청을 전달하고, 상기 결정된 데이터 수집 및 적재 방식으로 상기 수집 서버의 미리 설정된 수집 영역에 상기 데이터 서버로부터 데이터를 수집하고, 상기 결정된 데이터 수집 및 적재 방식을 근거로 상기 수집된 데이터를 상기 빅데이터 플랫폼에 적재하는 수집 서버를 포함할 수 있다.
- [6] 본 발명과 관련된 일 예로서 상기 수집 서버는, 상기 데이터 서버의 사용량, 상기 데이터 서버의 데이터 량 및 상기 빅데이터 플랫폼의 사용량을 확인할 수

있다.

- [7] 본 발명과 관련된 일 예로서 상기 수집 서버는, 상기 결정된 데이터 수집 및 적재 방식을 근거로 상기 수집된 데이터를 상기 빅데이터 플랫폼 내의 디스크 및 메모리 중 적어도 하나에 적재할 수 있다.
- [8] 본 발명과 관련된 일 예로서 상기 수집 서버는, 상기 수집 서버에서 상기 빅데이터 플랫폼으로의 상기 수집된 데이터 적재시 업데이트 함수 또는 쿼리가 지원되지 않을 때, 상기 빅데이터 플랫폼의 최종 적재 영역의 데이터와 상기 수집된 데이터를 조인(join)하여, 새로운 파티션의 데이터를 생성하고, 상기 생성된 새로운 파티션의 데이터를 상기 빅데이터 플랫폼에 대치하여 적재할 수 있다.
- [9] 본 발명과 관련된 일 예로서 상기 수집 서버는, 상기 수집 서버에서 상기 빅데이터 플랫폼으로의 상기 수집된 데이터 적재시 삭제 함수 또는 쿼리가 지원되지 않을 때, 상기 수집 서버의 새로 수집된 데이터를 상기 빅데이터 플랫폼의 해당 파티션에 대치하여 적재할 수 있다.
- [10] 본 발명의 실시예에 따른 데이터 적재 및 처리 방법은 수집 서버에 의해, 리소스의 사용량 및 데이터량을 확인하는 단계; 상기 수집 서버에 의해, 상기 확인된 리소스의 상태를 근거로 데이터 서버 및 빅데이터 플랫폼에 대한 데이터 수집 및 적재 방식을 결정하는 단계; 상기 수집 서버에 의해, 상기 데이터 서버에 데이터 수집 요청을 전달하는 단계; 상기 데이터 서버에 의해, 상기 수집 서버로부터 전달되는 상기 데이터 수집 요청을 수신하고, 상기 수신된 데이터 수집 요청을 저장하는 단계; 상기 수집 서버에 의해, 상기 결정된 데이터 수집 및 적재 방식으로 상기 수집 서버의 미리 설정된 수집 영역에 상기 데이터 서버로부터 데이터를 수집하는 단계; 및 상기 수집 서버에 의해, 상기 결정된 데이터 수집 및 적재 방식을 근거로 상기 수집된 데이터를 상기 빅데이터 플랫폼에 적재하는 단계를 포함할 수 있다.
- [11] 본 발명과 관련된 일 예로서 상기 데이터 서버 및 빅데이터 플랫폼에 대한 데이터 수집 및 적재 방식을 결정하는 단계는, 확인되는 데이터량 및 파티션 단위 구분 가능 여부에 따라 미리 설정된 복수의 데이터 수집 및 적재 방식 중 어느 하나의 데이터 수집 및 적재 방식을 결정할 수 있다.
- [12] 본 발명과 관련된 일 예로서 상기 데이터 수집 및 적재 방식은, 미리 설정된 1 유형, 2A 유형, 2B 유형, 3A 유형, 3B 유형, 4 유형 및 5 유형 중 어느 하나이며, 상기 1 유형은, 상기 데이터 서버의 전체 데이터를 상기 수집 서버의 미리 설정된 수집 영역으로 수집하고, 상기 빅데이터 플랫폼의 최종 적재 영역의 기존 데이터를 삭제한 후, 상기 수집된 데이터를 상기 빅데이터 플랫폼에 적재하는 방식이고, 상기 2A 유형은, 상기 데이터 서버의 새로 추가된 데이터만 상기 수집 서버의 수집 영역으로 수집하고, 상기 빅데이터 플랫폼의 최종 적재 영역에 상기 수집된 데이터를 추가하여 적재하는 방식이고, 상기 2B 유형은, 상기 데이터 서버의 새로 추가 및 수정된 데이터만 상기 수집 서버의 수집 영역으로

수집하고, 리소스(resource) 내 완료테이블의 기본 데이터와 중복되지 않도록 미리 설정된 기본 키(Primary Key)를 이용하여 상기 빅데이터 플랫폼의 최종 적재 영역에 상기 수집된 데이터를 적재하는 방식이고, 상기 3A 유형은, 타깃 영역이 파티션으로 구성되며, 상기 데이터 서버의 파티션 중 추가된 데이터가 존재하는 파티션에 해당하는 데이터만 상기 수집 서버의 수집 영역으로 수집하고, 상기 빅데이터 플랫폼의 최종 적재 영역에 상기 수집된 데이터를 추가하여 적재하는 방식이고, 상기 3B 유형은, 타깃 영역이 파티션으로 구성되며, 상기 데이터 서버의 파티션 중 추가 및 수정된 데이터만 상기 수집 서버의 수집 영역으로 수집하고, 상기 빅데이터 플랫폼의 최종 적재 영역의 기존 데이터와 중복되지 않도록 상기 기본 키를 이용하여 상기 빅데이터 플랫폼의 최종 적재 영역 내의 파티션에 상기 수집된 데이터를 적재하는 방식이고, 상기 4 유형은, 타깃 영역이 파티션으로 구성되며, 상기 데이터 서버의 파티션 중 추가 및 수정이 발생한 파티션 또는, 상기 데이터 서버의 파티션 중 삭제가 발생한 파티션을 상기 수집 서버의 수집 영역으로 수집하고, 상기 빅데이터 플랫폼의 최종 적재 영역의 데이터 양을 확인하고, 상기 확인된 빅데이터 플랫폼의 최종 적재 영역의 데이터 양이 미리 설정된 기준값보다 작은 경우, 상기 빅데이터 플랫폼의 최종 적재 영역의 기존 파티션에 상기 수집된 데이터를 덮어 적재하는 방식이고, 상기 5 유형은, 타깃 영역이 파티션으로 구성되며, 상기 데이터 서버의 파티션 중 추가 및 수정이 발생한 파티션 또는, 상기 데이터 서버의 파티션 중 삭제가 발생한 파티션을 상기 수집 서버의 수집 영역으로 수집하고, 상기 빅데이터 플랫폼의 최종 적재 영역의 데이터 양을 확인하고, 상기 확인된 빅데이터 플랫폼의 최종 적재 영역의 데이터 양이 미리 설정된 기준값보다 크거나 같은 경우, 상기 수집된 데이터 중 일부를 상기 3B 유형으로 상기 빅데이터 플랫폼의 최종 적재 영역에 적재하고, 상기 수집된 데이터 중 나머지를 상기 4 유형으로 상기 빅데이터 플랫폼의 최종 적재 영역에 적재하는 방식일 수 있다.

- [13] 본 발명과 관련된 일 예로서 상기 결정된 데이터 수집 및 적재 방식으로 상기 수집 서버의 미리 설정된 수집 영역에 상기 데이터 서버로부터 데이터를 수집하는 단계는, 상기 결정된 데이터 수집 및 적재 방식으로 상기 데이터 서버를 구성하는 DBMS에 저장된 복수의 데이터를 미리 설정된 JDBC 또는 ODBC를 이용해서 수집하는 과정; 상기 결정된 데이터 수집 및 적재 방식으로 상기 데이터 서버에 저장된 로그나 파일들을 미리 설정된 FTP, 오픈 소스 및 에이전트 중 적어도 하나를 이용해서 수집하는 과정; 및 상기 결정된 데이터 수집 및 적재 방식으로 웹 데이터들을 웹 크롤링을 이용해서 수집하는 과정 중 적어도 하나의 과정을 포함할 수 있다.
- [14] 본 발명과 관련된 일 예로서 상기 결정된 데이터 수집 및 적재 방식을 근거로 상기 수집된 데이터를 상기 빅데이터 플랫폼에 적재하는 단계는, 상기 수집 서버에서 상기 빅데이터 플랫폼으로의 상기 수집된 데이터 적재시 업데이트

함수 또는 쿼리가 지원되지 않을 때, 상기 수집 서버에 의해, 상기 빅데이터 플랫폼의 최종 적재 영역의 데이터와 상기 수집된 데이터를 조인(join)하여, 새로운 파티션의 데이터를 생성하는 과정; 및 상기 생성된 새로운 파티션의 데이터를 상기 빅데이터 플랫폼에 대치하여 적재하는 과정을 포함할 수 있다.

### 발명의 효과

- [15] 본 발명은 수집 서버에서 데이터를 수집하기 전에 대상 리소스의 상태를 확인하고, 상기 확인된 리소스의 상태를 근거로 해당 리소스에 대한 데이터 수집 및 적재 방식을 결정하고, 데이터 서버에 데이터 수집 요청을 전달하고, 상기 결정된 데이터 수집 및 적재 방식으로 데이터 서버로부터 데이터를 수집하고, 상기 수집된 데이터를 빅데이터 플랫폼에 적재함으로써, 수집 대상 및 적재 대상의 현재 상태에 따라 최적의 데이터 수집 및 적재 방식에 의해 데이터를 수집 및 적재하여 전체 시스템 운영 효율을 높일 수 있는 효과가 있다.

### 도면의 간단한 설명

- [16] 도 1은 본 발명의 실시예에 따른 데이터 적재 및 처리 시스템의 구성을 나타낸 블록도이다.
- [17] 도 2 내지 도 8은 본 발명의 실시예에 따른 데이터 수집 및 적재 방식의 예를 나타낸 도이다.
- [18] 도 9는 본 발명의 실시예에 따른 업데이트 함수 또는 쿼리가 지원되지 않을 경우의 데이터 적재 예를 나타낸 도이다.
- [19] 도 10은 본 발명의 실시예에 따른 삭제 함수 또는 쿼리가 지원되지 않을 경우의 데이터 적재 예를 나타낸 도이다.
- [20] 도 11은 본 발명의 실시예에 따른 데이터 적재 및 처리 방법을 나타낸 흐름도이다.

### 발명의 실시를 위한 최선의 형태

- [21] 본 발명에서 사용되는 기술적 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아님을 유의해야 한다. 또한, 본 발명에서 사용되는 기술적 용어는 본 발명에서 특별히 다른 의미로 정의되지 않는 한, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 의미로 해석되어야 하며, 과도하게 포괄적인 의미로 해석되거나, 과도하게 축소된 의미로 해석되지 않아야 한다. 또한, 본 발명에서 사용되는 기술적인 용어가 본 발명의 사상을 정확하게 표현하지 못하는 잘못된 기술적 용어일 때에는 당업자가 올바르게 이해할 수 있는 기술적 용어로 대체되어 이해되어야 할 것이다. 또한, 본 발명에서 사용되는 일반적인 용어는 사전에 정의되어 있는 바에 따라, 또는 전후 문맥상에 따라 해석되어야 하며, 과도하게 축소된 의미로 해석되지 않아야 한다.
- [22] 또한, 본 발명에서 사용되는 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한 복수의 표현을 포함한다. 본 발명에서 "구성된다" 또는 "포함한다" 등의

용어는 발명에 기재된 여러 구성 요소들 또는 여러 단계를 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지 않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다.

- [23] 또한, 본 발명에서 사용되는 제 1, 제 2 등과 같이 서수를 포함하는 용어는 구성 요소들을 설명하는데 사용될 수 있지만, 구성 요소들은 용어들에 의해 한정되어서는 안 된다. 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제 1 구성 요소는 제 2 구성 요소로 명명될 수 있고, 유사하게 제 2 구성 요소도 제 1 구성 요소로 명명될 수 있다.
- [24] 이하, 첨부된 도면을 참조하여 본 발명에 따른 바람직한 실시예를 상세히 설명하되, 도면 부호에 관계없이 동일하거나 유사한 구성 요소는 동일한 참조 번호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다.
- [25] 또한, 본 발명을 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다. 또한, 첨부된 도면은 본 발명의 사상을 쉽게 이해할 수 있도록 하기 위한 것일 뿐, 첨부된 도면에 의해 본 발명의 사상이 제한되는 것으로 해석되어서는 아니 됨을 유의해야 한다.
- [26] 도 1은 본 발명의 실시예에 따른 데이터 적재 및 처리 시스템(10)의 구성을 나타낸 블록도이다.
- [27] 도 1에 도시한 바와 같이, 데이터 적재 및 처리 시스템(10)은 데이터 서버(100), 빅데이터 플랫폼(200) 및 수집 서버(300)로 구성된다. 도 1에 도시된 데이터 적재 및 처리 시스템(10)의 구성 요소 모두가 필수 구성 요소인 것은 아니며, 도 1에 도시된 구성 요소보다 많은 구성 요소에 의해 데이터 적재 및 처리 시스템(10)이 구현될 수도 있고, 그보다 적은 구성 요소에 의해서도 데이터 적재 및 처리 시스템(10)이 구현될 수도 있다.
- [28] 상기 데이터 서버(100), 상기 빅데이터 플랫폼(200) 및 상기 수집 서버(300) 각각은 다른 단말들과의 통신 기능을 수행하기 위한 통신부(미도시), 다양한 정보 및 프로그램(또는 애플리케이션)을 저장하기 위한 저장부(미도시), 다양한 정보 및 프로그램 실행 결과를 표시하기 위한 표시부(미도시), 상기 다양한 정보 및 프로그램 실행 결과에 대응하는 음성 정보를 출력하기 위한 음성 출력부(미도시), 각 단말의 다양한 구성 요소 및 기능을 제어하기 위한 제어부(미도시) 등을 포함할 수 있다.
- [29] 상기 데이터 서버(100), 상기 빅데이터 플랫폼(200), 상기 수집 서버(300) 등과 통신한다.
- [30] 또한, 상기 데이터 서버(100)는 웹 서버, 데이터베이스 서버, 프록시 서버 등의 형태로 구현될 수 있다. 또한, 상기 데이터 서버(100)에는 네트워크 부하 분산 메커니즘, 내지 상기 데이터 서버(100)가 인터넷 또는 다른 네트워크 상에서

동작할 수 있도록 하는 다양한 소프트웨어 중 하나 이상이 설치될 수 있으며, 이를 통해 컴퓨터화된 시스템으로 구현될 수 있다. 또한, 네트워크는 http 네트워크일 수 있으며, 전용 회선(private line), 인트라넷 또는 임의의 다른 네트워크일 수 있다. 나아가, 상기 데이터 서버(100)와 다른 서버/단말의 연결은 데이터가 임의의 해커 또는 다른 제3자에 의한 공격을 받지 않도록 보안 네트워크로 연결될 수 있다. 또한, 상기 데이터 서버(100)는 복수의 데이터베이스 서버를 포함할 수 있으며, 이러한 데이터베이스 서버가 분산 데이터베이스 서버 아키텍처를 비롯한 임의의 유형의 네트워크 연결을 통해 상기 데이터 서버(100)와 별도로 연결되는 방식으로 구현될 수 있다.

- [31] 또한, 상기 데이터 서버(100)는 CPU(미도시), 디스크(DISK)(미도시), 메모리(Memory)(미도시), DBMS(Database Management System: 데이터베이스 관리 시스템)(미도시) 등을 포함(또는 구성)한다. 여기서, 상기 DBMS는 다양한 서버 및/또는 장치들로부터 수집된 복수의 데이터(또는 테이블)를 저장(또는 관리)한다.
- [32] 또한, 상기 수집 서버(300)에서 상기 데이터 서버(100)에 저장된 하나 이상의 데이터(또는 테이블)에 대한 수집 과정을 시도하기 전에, 상기 데이터 서버(100)는 상기 수집 서버(300)로부터 전달되는 상기 데이터 수집 요청을 수신한다.
- [33] 또한, 상기 데이터 서버(100)는 상기 수신된 데이터 수집 요청을 저장한다.
- [34] 또한, 상기 데이터 서버(100)는 상기 수신된 데이터 수집 요청을 근거로 상기 수집 서버(300)에서 상기 데이터 서버(100)에 저장된 데이터의 수집을 시도하는 경우, 해당 수집 서버(300)의 액세스를 허용한다.
- [35] 상기 빅데이터 플랫폼(200)은 상기 데이터 서버(100), 상기 수집 서버(300) 등과 통신한다.
- [36] 또한, 상기 빅데이터 플랫폼(200)은 빅데이터 서버로, CPU(미도시), 디스크(미도시), 메모리(미도시), DBMS 등을 포함(또는 구성)한다.
- [37] 또한, 상기 빅데이터 플랫폼(200)은 상기 수집 서버(300)에서 결정된 데이터 수집 및 적재 방식으로 상기 데이터 서버(100)로부터 수집된 데이터를 상기 빅데이터 플랫폼(200)에 적재한다.
- [38] 즉, 상기 빅데이터 플랫폼(200)은 상기 수집 서버(300)와 연동하여, 상기 수집 서버(300)에서 결정된 데이터 수집 및 적재 방식을 근거로 상기 데이터 서버(100)로부터 수집된 데이터를 상기 빅데이터 플랫폼(200) 내의 디스크 및/또는 메모리에 적재한다.
- [39] 상기 수집 서버(300)는 상기 데이터 서버(100), 상기 빅데이터 플랫폼(200) 등과 통신한다.
- [40] 또한, 상기 수집 서버(300)는 CPU(미도시), 디스크(미도시), 메모리(미도시), DBMS 등을 포함(또는 구성)한다.
- [41] 또한, 상기 수집 서버(300)는 대상 리소스인 데이터 서버(100) 및 빅데이터

플랫폼(200)의 사용량 및 상기 데이터 서버(100)의 데이터 량을 확인(또는 체크)한다. 여기서, 상기 리소스는 수집 대상인 상기 데이터 서버(100) 및 적재 대상인 상기 빅데이터 플랫폼(200)의 사용량 및 수집 대상의 데이터 량을 나타낸다.

- [42] 즉, 상기 수집 서버(300)는 상기 데이터 서버(100) 및 상기 빅데이터 플랫폼(200)에 구성된 CPU, 디스크(DISK) 및 메모리(Memory) 중 적어도 하나의 사용량을 확인한다.
- [43] 또한, 상기 수집 서버(300)는 상기 데이터 서버(100)의 데이터 량을 확인한다.
- [44] 또한, 상기 수집 서버(300)는 상기 확인된 리소스의 상태(예를 들어 상기 데이터 서버(100)의 사용량, 상기 빅데이터 플랫폼(200)의 사용량, 상기 데이터 서버(100)의 데이터 량 등 포함) 등을 근거로 상기 데이터 서버(100) 및 상기 빅데이터 플랫폼(200)에 대한 데이터 수집 및 적재 방식을 결정한다.
- [45] 즉, 상기 수집 서버(300)는 상기 확인되는 데이터 량 및 파티션 단위 구분 가능 여부에 따라 미리 설정된 복수의 데이터 수집 및 적재 방식 중 어느 하나의 데이터 수집 및 적재 방식을 결정(또는 선택/선정)한다. 이때, 상기 복수의 데이터 수집 및 적재 방식(또는 수집 및 적재 유형)은 1 유형, 2A 유형, 2B 유형, 3A 유형, 3B 유형, 4 유형, 5 유형 등을 포함한다. 또한, 상기 파티션은 데이터가 적재되는(또는 저장되는) 기간(예를 들어 년 단위, 월 단위, 일 단위 등 포함), 지역(예를 들어 국가, 행정 구역(일 예로 시, 도, 군 등 포함) 등 포함) 등에 따라 설정(또는 정의/구별)할 수 있다.
- [46] 여기서, 도 2에 도시된 바와 같이, 상기 1 유형(또는 1 유형의 데이터 수집 방법/Full Dump 유형)은 상기 데이터 서버(100)의 전체 데이터를 상기 수집 서버(300)의 미리 설정된 수집 영역으로 수집하고, 상기 빅데이터 플랫폼(200)의 최종 적재 영역의 기존 데이터를 삭제한 후, 상기 수집된 데이터를 상기 빅데이터 플랫폼(200)에 적재하는 방식이다.
- [47] 또한, 도 3에 도시된 바와 같이, 상기 2A 유형(또는 2A 유형의 데이터 수집 방법/Append 유형)은 상기 데이터 서버(100)의 새로 추가된 데이터만 상기 수집 서버(300)의 수집 영역으로 수집하고, 상기 빅데이터 플랫폼(200)의 최종 적재 영역에 상기 수집된 데이터를 추가하여 적재하는 방식이다.
- [48] 또한, 도 4에 도시된 바와 같이, 상기 2B 유형(또는 2B 유형의 데이터 수집 방법/Update 유형)은 상기 데이터 서버(100)의 추가 및 수정된 데이터만 상기 수집 서버(300)의 수집 영역으로 수집하고, 리소스 내 완료테이블의 기본 데이터와 중복되지 않도록 미리 설정된 기본 키(Primary Key)를 이용하여 상기 빅데이터 플랫폼(200)의 최종 적재 영역에 상기 수집된 데이터를 적재하는 방식이다.
- [49] 또한, 도 5에 도시된 바와 같이, 상기 3A 유형(또는 3A 유형의 데이터 수집 방법/Partial Append 유형)은 타깃 영역(또는 상기 빅데이터 플랫폼(200)의 타깃 영역)이 파티션으로 구성되며, 상기 데이터 서버(100)의 파티션 중 추가된

데이터가 존재하는 파티션에 해당하는 데이터만 상기 수집 서버(300)의 수집 영역으로 수집하고, 상기 빅데이터 플랫폼(200)의 최종 적재 영역에 상기 수집된 데이터를 추가하여 적재하는 방식이다.

- [50] 또한, 도 6에 도시된 바와 같이, 상기 3B 유형(또는 3B 유형의 데이터 수집 방법/Partial Update 유형)은 타깃 영역(또는 상기 빅데이터 플랫폼(200)의 타깃 영역)이 파티션으로 구성되며, 상기 데이터 서버(100)의 파티션 중 추가 및 수정된 데이터만 상기 수집 서버(300)의 수집 영역으로 수집하고, 상기 빅데이터 플랫폼(200)의 최종 적재 영역의 기존 데이터와 중복되지 않도록 상기 기본 키를 이용하여 상기 빅데이터 플랫폼(200)의 최종 적재 영역 내의 파티션에 상기 수집된 데이터를 적재하는 방식이다.
- [51] 또한, 도 7에 도시된 바와 같이, 상기 4 유형(또는 4 유형의 데이터 수집 방법/Partition And Remove 유형)은 타깃 영역(또는 상기 빅데이터 플랫폼(200)의 타깃 영역)이 파티션으로 구성되며, 상기 데이터 서버(100)의 파티션 중 추가 및 수정이 발생한(또는 일어난) 파티션 또는, 상기 데이터 서버(100)의 파티션 중 삭제가 발생한 파티션을 상기 수집 서버(300)의 수집 영역으로 수집하고, 상기 빅데이터 플랫폼(200)의 최종 적재 영역의 데이터 양을 확인하고, 상기 확인된 빅데이터 플랫폼(200)의 최종 적재 영역의 데이터 양이 미리 설정된 기준값보다 작은 경우(또는 데이터의 파티션 범위가 미리 설정된 기준 단위(예를 들어 3개월)보다 작은 경우), 상기 빅데이터 플랫폼(200)의 최종 적재 영역의 기존 파티션에 상기 수집된 데이터를 덮어 적재하는 방식이다.
- [52] 또한, 도 8에 도시된 바와 같이, 상기 5 유형(또는 5 유형의 데이터 수집 방법/Partition and Complex 유형)은 타깃 영역(또는 상기 빅데이터 플랫폼(200)의 타깃 영역)이 파티션으로 구성되며, 상기 데이터 서버(100)의 파티션 중 추가 및 수정이 발생한(또는 일어난) 파티션 또는, 상기 데이터 서버(100)의 파티션 중 삭제가 발생한 파티션을 상기 수집 서버(300)의 수집 영역으로 수집하고, 상기 빅데이터 플랫폼(200)의 최종 적재 영역의 데이터 양을 확인하고, 상기 확인된 빅데이터 플랫폼(200)의 최종 적재 영역의 데이터 양이 미리 설정된 기준값보다 크거나 같은 경우(또는 데이터의 파티션 범위가 미리 설정된 기준 단위(예를 들어 3개월)보다 크거나 같은 경우), 상기 수집된 데이터 중 일부를 상기 3B 유형으로 상기 빅데이터 플랫폼(200)의 최종 적재 영역에 적재(또는 상기 빅데이터 플랫폼(200)의 최종 적재 영역의 기존 데이터와 중복되지 않도록 상기 기본 키를 이용하여 상기 빅데이터 플랫폼(200)의 최종 적재 영역 내의 파티션에 상기 수집된 데이터를 적재)하고, 상기 수집된 데이터 중 나머지를 상기 4 유형으로 상기 빅데이터 플랫폼(200)의 최종 적재 영역에 적재(또는 상기 빅데이터 플랫폼(200)의 최종 적재 영역의 기존 파티션에 상기 수집된 데이터를 덮어 적재)하는 방식이다.
- [53] 이와 같이, 상기 데이터 수집 및 적재 방식 중 1 유형, 2A 유형 및 2B 유형과, 3A 유형, 3B 유형, 4 유형 및 5 유형의 구분 기준은 데이터를 파티션 단위로 구분

가능한지 여부일 수 있다.

- [54] 또한, 상기 수집 서버(300)는 상기 데이터 서버(100)에 데이터 수집 요청을 전달(또는 전송)한다.
- [55] 즉, 상기 수집 서버(300)는 상기 결정된 데이터 수집 및 적재 방식으로 데이터를 수집하기 위해서 상기 데이터 서버(100)에 데이터 수집 요청을 전달(또는 전송)한다.
- [56] 또한, 상기 수집 서버(300)는 상기 결정된 데이터 수집 및 적재 방식으로 해당 수집 서버(300)의 미리 설정된 수집 영역에 상기 데이터 서버(100)로부터 데이터를 수집한다.
- [57] 즉, 상기 수집 서버(300)는 상기 결정된 데이터 수집 및 적재 방식으로 상기 데이터 서버(100)를 구성하는 DBMS(데이터베이스 관리 시스템)에 저장된 복수의 데이터(또는 테이블)를 미리 설정된 JDBC(Java Database Connectivity: 자바 데이터베이스 접속성) 또는 ODBC(Open Database Connectivity: 개방형 데이터베이스 접속성)를 이용해서 수집하고, 상기 수집된 데이터를 상기 수집 영역에 저장(또는 임시 저장/수집)한다.
- [58] 또한, 상기 수집 서버(300)는 상기 결정된 데이터 수집 및 적재 방식으로 상기 데이터 서버(100)에 저장된 로그(log)나 파일들을 미리 설정된 FTP(File Transfer Protocol), 오픈 소스(Open Source), 자체 개발 에이전트(미도시) 등을 이용해서 수집하거나 또는, 웹 데이터들을 웹 크롤링(Web Crawling) 등을 이용해서 수집하고, 상기 수집된 데이터를 상기 수집 영역에 저장(또는 임시 저장/수집)할 수도 있다.
- [59] 또한, 상기 수집 서버(300)는 상기 결정된 데이터 수집 및 적재 방식을 근거로 상기 수집된 데이터를 상기 빅데이터 플랫폼(200)에 적재한다.
- [60] 즉, 상기 수집 서버(300)는 상기 결정된 데이터 수집 및 적재 방식을 근거로 상기 수집된 데이터(또는 상기 수집 영역에 저장된 상기 수집된 데이터)를 상기 빅데이터 플랫폼(200) 내의 디스크 및/또는 메모리에 적재한다.
- [61] 이때, 상기 수집 서버(300)에서 상기 빅데이터 플랫폼(200)으로의 상기 수집된 데이터 적재시 업데이트 함수 또는 쿼리가 지원되지 않은 경우, 도 9에 도시된 바와 같이, 상기 수집 서버(300)는 상기 빅데이터 플랫폼(200)의 최종 적재 영역의 데이터와 상기 수집된 데이터를 조인(join)하여, 새로운 파티션의 데이터를 생성하고, 상기 생성된 새로운 파티션의 데이터를 상기 빅데이터 플랫폼(200)에 대치하여 적재할 수 있다.
- [62] 또한, 상기 수집 서버(300)에서 상기 빅데이터 플랫폼(200)으로의 상기 수집된 데이터 적재시 삭제 함수 또는 쿼리가 지원되지 않은 경우, 도 10에 도시된 바와 같이, 상기 수집 서버(300)는 상기 수집 서버(300)의 새로 수집된 데이터를 상기 빅데이터 플랫폼(200)의 해당 파티션에 대치하여 적재할 수 있다.
- [63] 본 발명의 실시예에서는, 상기 수집 서버(300)가 독립적으로 운영되는 것을 설명하고 있으나, 이에 한정되는 것은 아니며, 상기 수집 서버(300)는 상기

빅데이터 플랫폼(200) 내에(또는 상기 빅데이터 플랫폼(200) 내의 수집 모듈로) 구성되어 운영될 수도 있다.

- [64] 이와 같이, 수집 서버에서 데이터를 수집하기 전에 대상 리소스의 상태를 확인하고, 상기 확인된 리소스의 상태를 근거로 해당 리소스에 대한 데이터 수집 및 적재 방식을 결정하고, 데이터 서버에 데이터 수집 요청을 전달하고, 상기 결정된 데이터 수집 및 적재 방식으로 데이터 서버로부터 데이터를 수집하고, 상기 수집된 데이터를 빅데이터 플랫폼에 적재할 수 있다.
- [65] 이하에서는, 본 발명에 따른 데이터 적재 및 처리 방법을 도 1 내지 도 11을 참조하여 상세히 설명한다.
- [66] 도 11은 본 발명의 실시예에 따른 데이터 적재 및 처리 방법을 나타낸 흐름도이다.
- [67] 먼저, 수집 서버(300)는 대상 리소스인 데이터 서버(100) 및 빅데이터 플랫폼(200)의 사용량 및 상기 데이터 서버(100)의 데이터량을 확인(또는 체크)한다. 여기서, 상기 리소스는 수집 대상인 상기 데이터 서버(100) 및 적재 대상인 상기 빅데이터 플랫폼(200)의 사용량 및 수집 대상의 데이터량을 나타낸다.
- [68] 즉, 상기 수집 서버(300)는 상기 데이터 서버(100) 및 상기 빅데이터 플랫폼(200)에 구성된 CPU, 디스크(DISK) 및 메모리(Memory) 중 적어도 하나의 사용량을 확인한다.
- [69] 또한, 상기 수집 서버(300)는 상기 데이터 서버(100)의 데이터량을 확인한다.
- [70] 일 예로, 제 1 수집 서버(300)는 제 1 데이터 서버(100)에 포함된 제 1 CPU, 제 1 디스크 및 제 1 메모리의 사용량과, 상기 제 1 데이터 서버의 데이터량을 확인한다.
- [71] 또한, 상기 제 1 수집 서버는 제 1 빅데이터 플랫폼(200)에 포함된 제 2 CPU, 제 2 디스크 및 제 2 메모리의 사용량을 확인한다.
- [72] 다른 일 예로, 제 2 수집 서버(300)는 제 1 데이터 서버(100)에 포함된 제 1 CPU, 제 1 디스크 및 제 1 메모리의 사용량과, 상기 제 1 데이터 서버의 데이터량을 확인한다.
- [73] 또한, 상기 제 2 수집 서버는 제 1 빅데이터 플랫폼(200)에 포함된 제 2 CPU, 제 2 디스크 및 제 2 메모리의 사용량을 확인한다(S1110).
- [74] 이후, 상기 수집 서버(300)는 상기 확인된 리소스의 상태(예를 들어 상기 데이터 서버(100)의 사용량, 상기 빅데이터 플랫폼(200)의 사용량, 상기 데이터 서버(100)의 데이터량 등 포함) 등을 근거로 상기 데이터 서버(100) 및 상기 빅데이터 플랫폼(200)에 대한 데이터 수집 및 적재 방식을 결정한다.
- [75] 즉, 상기 수집 서버(300)는 상기 확인되는 데이터량 및 파티션 단위 구분 가능 여부에 따라 미리 설정된 복수의 데이터 수집 및 적재 방식 중 어느 하나의 데이터 수집 및 적재 방식을 결정(또는 선택/선정)한다. 이때, 상기 복수의 데이터 수집 및 적재 방식(또는 수집 및 적재 유형)은 1 유형, 2A 유형, 2B 유형, 3A 유형,

3B 유형, 4 유형, 5 유형 등을 포함한다. 여기서, 상기 파티션은 데이터가 적재되는(또는 저장되는) 기간(예를 들어 년 단위, 월 단위, 일 단위 등 포함), 지역(예를 들어 국가, 행정 구역(일 예로 시, 도, 군 등 포함) 등 포함) 등에 따라 설정(또는 정의/구별)할 수 있다.

- [76] 일 예로, 상기 제 1 수집 서버는 상기 확인된 제 1 데이터 서버에 포함된 제 1 CPU, 제 1 디스크 및 제 1 메모리의 사용량과, 상기 제 1 데이터 서버의 데이터량과, 상기 확인된 제 1 빅데이터 플랫폼에 포함된 제 2 CPU, 제 2 디스크 및 제 2 메모리의 사용량을 근거로 미리 설정된 복수의 데이터 수집 및 적재 방식 중에서 1 유형의 데이터 수집 및 적재 방식을 결정한다.
- [77] 다른 일 예로, 상기 제 2 수집 서버는 상기 확인된 제 1 데이터 서버에 포함된 제 1 CPU, 제 1 디스크 및 제 1 메모리의 사용량과, 상기 제 1 데이터 서버의 데이터량과, 상기 확인된 제 1 빅데이터 플랫폼에 포함된 제 2 CPU, 제 2 디스크 및 제 2 메모리의 사용량을 근거로 미리 설정된 복수의 데이터 수집 및 적재 방식 중에서 3B 유형의 데이터 수집 및 적재 방식을 결정한다(S1120).
- [78] 이후, 상기 수집 서버(300)는 상기 데이터 서버(100)에 데이터 수집 요청을 전달(또는 전송)한다.
- [79] 즉, 상기 수집 서버(300)는 상기 결정된 데이터 수집 및 적재 방식으로 데이터를 수집하기 위해서 상기 데이터 서버(100)에 데이터 수집 요청을 전달(또는 전송)한다.
- [80] 또한, 상기 데이터 서버(100)는 상기 수집 서버(300)로부터 전달되는 상기 데이터 수집 요청을 수신하고, 상기 수신된 데이터 수집 요청을 저장한다.
- [81] 일 예로, 상기 제 1 수집 서버는 상기 결정된 1 유형의 데이터 수집 및 적재 방식으로 상기 제 1 데이터 서버로부터 데이터를 수집하기 위해서 상기 제 1 데이터 서버에 제 1 데이터 수집 요청을 전송한다.
- [82] 또한, 상기 제 1 데이터 서버는 상기 제 1 수집 서버로부터 전송되는 상기 제 1 데이터 수집 요청을 수신한다.
- [83] 다른 일 예로, 상기 제 2 수집 서버는 상기 결정된 3B 유형의 데이터 수집 및 적재 방식으로 상기 제 1 데이터 서버로부터 데이터를 수집하기 위해서 상기 제 1 데이터 서버에 제 2 데이터 수집 요청을 전송한다.
- [84] 또한, 상기 제 1 데이터 서버는 상기 제 2 수집 서버로부터 전송되는 상기 제 2 데이터 수집 요청을 수신한다(S1130).
- [85] 이후, 상기 수집 서버(300)는 상기 결정된 데이터 수집 및 적재 방식으로 해당 수집 서버(300)의 미리 설정된 수집 영역에 상기 데이터 서버(100)로부터 데이터를 수집한다.
- [86] 즉, 상기 수집 서버(300)는 상기 결정된 데이터 수집 및 적재 방식으로 상기 데이터 서버(100)를 구성하는 DBMS(데이터베이스 관리 시스템)에 저장된 복수의 데이터(또는 테이블)를 미리 설정된 JDBC(자바 데이터베이스 접속성) 또는 ODBC(개방형 데이터베이스 접속성)를 이용해서 수집하고, 상기 수집된

데이터를 상기 수집 영역에 저장(또는 임시 저장/수집)한다.

- [87] 또한, 상기 수집 서버(300)는 상기 결정된 데이터 수집 및 적재 방식으로 상기 데이터 서버(100)에 저장된 로그(log)나 파일들을 미리 설정된 FTP, 오픈 소스, 자체 개발 에이전트(미도시) 등을 이용해서 수집하거나 또는, 웹 데이터들을 웹 크롤링(Web Crawling) 등을 이용해서 수집하고, 상기 수집된 데이터를 상기 수집 영역에 저장(또는 임시 저장/수집)할 수도 있다.
- [88] 일 예로, 상기 제 1 수집 서버는 상기 결정된 1 유형의 데이터 수집 및 적재 방식으로 상기 제 1 데이터 서버를 구성하는 DBMS에 저장된 전체 데이터를 상기 JDBC를 이용해서 수집하고, 상기 수집된 전체 데이터를 상기 제 1 수집 서버의 제 1 수집 영역에 저장한다.
- [89] 다른 일 예로, 상기 제 2 수집 서버는 상기 결정된 3B 유형의 데이터 수집 및 적재 방식으로 상기 제 1 데이터 서버를 구성하는 DBMS에 저장된 전체 데이터 중에서 파티션 중 추가 및 수정된 데이터만을 상기 JDBC를 이용해서 수집하고, 상기 수집된 추가 및 수정된 데이터를 상기 제 2 수집 서버의 제 2 수집 영역에 저장한다(S1140).
- [90] 이후, 상기 수집 서버(300)는 상기 결정된 데이터 수집 및 적재 방식을 근거로 상기 수집된 데이터를 상기 빅데이터 플랫폼(200)에 적재한다.
- [91] 즉, 상기 수집 서버(300)는 상기 결정된 데이터 수집 및 적재 방식을 근거로 상기 수집된 데이터(또는 상기 수집 영역에 저장된 상기 수집된 데이터)를 상기 빅데이터 플랫폼(200) 내의 디스크 및/또는 메모리에 적재한다.
- [92] 이때, 상기 수집 서버(300)에서 상기 빅데이터 플랫폼(200)으로의 상기 수집된 데이터 적재시 업데이트 함수 또는 쿼리가 지원되지 않은 경우, 도 9에 도시된 바와 같이, 상기 수집 서버(300)는 상기 빅데이터 플랫폼(200)의 최종 적재 영역의 데이터와 상기 수집된 데이터를 조인(join)하여, 새로운 파티션의 데이터를 생성하고, 상기 생성된 새로운 파티션의 데이터를 상기 빅데이터 플랫폼(200)에 대치하여 적재할 수 있다.
- [93] 또한, 상기 수집 서버(300)에서 상기 빅데이터 플랫폼(200)으로의 상기 수집된 데이터 적재시 삭제 함수 또는 쿼리가 지원되지 않은 경우, 도 10에 도시된 바와 같이, 상기 수집 서버(300)는 상기 수집 서버(300)의 새로 수집된 데이터를 상기 빅데이터 플랫폼(200)의 해당 파티션에 대치하여 적재할 수 있다.
- [94] 일 예로, 상기 제 1 수집 서버는 상기 결정된 1 유형의 데이터 수집 및 적재 방식을 근거로 상기 수집된 상기 DBMS에 저장된 전체 데이터를 상기 제 1 빅데이터 플랫폼 내의 제 2 디스크에 적재한다.
- [95] 다른 일 예로, 상기 제 2 수집 서버는 상기 결정된 3B 유형의 데이터 수집 및 적재 방식을 근거로 상기 수집된 추가 및 수정된 데이터를 상기 제 1 빅데이터 플랫폼 내의 제 2 디스크에 적재한다(S1150).
- [96] 본 발명의 실시예는 앞서 설명된 바와 같이, 수집 서버에서 데이터를 수집하기 전에 대상 리소스의 상태를 확인하고, 상기 확인된 리소스의 상태를 근거로 해당

리소스에 대한 데이터 수집 및 적재 방식을 결정하고, 데이터 서버에 데이터 수집 요청을 전달하고, 상기 결정된 데이터 수집 및 적재 방식으로 데이터 서버로부터 데이터를 수집하고, 상기 수집된 데이터를 빅데이터 플랫폼에 적재하여, 수집 대상 및 적재 대상의 현재 상태에 따라 최적의 데이터 수집 및 적재 방식에 의해 데이터를 수집 및 적재하여 전체 시스템 운영 효율을 높일 수 있다.

- [97] 전술된 내용은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 수정 및 변형이 가능할 것이다. 따라서, 본 발명에 개시된 실시예들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아니다. 본 발명의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

#### **발명의 실시를 위한 형태**

- [98] 발명의 실시를 위한 형태는 위의 발명의 실시를 위한 최선의 형태에서 함께 기술되었다.

#### **산업상 이용가능성**

- [99] 본 발명은 수집 서버에서 데이터를 수집하기 전에 대상 리소스의 상태를 확인하고, 상기 확인된 리소스의 상태를 근거로 해당 리소스에 대한 데이터 수집 및 적재 방식을 결정하고, 데이터 서버에 데이터 수집 요청을 전달하고, 상기 결정된 데이터 수집 및 적재 방식으로 데이터 서버로부터 데이터를 수집하고, 상기 수집된 데이터를 빅데이터 플랫폼에 적재함으로써, 수집 대상 및 적재 대상의 현재 상태에 따라 최적의 데이터 수집 및 적재 방식에 의해 데이터를 수집 및 적재하여 전체 시스템 운영 효율을 높일 수 있어 산업상 이용가능성이 있다.