



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2018년10월25일  
(11) 등록번호 10-1911793  
(24) 등록일자 2018년10월19일

(51) 국제특허분류(Int. Cl.)  
*G06F 17/00* (2006.01) *G06F 17/30* (2006.01)  
*G06F 9/44* (2018.01)

(21) 출원번호 10-2013-7010459

(22) 출원일자(국제) 2011년10월25일  
심사청구일자 2016년09월30일

(85) 번역문제출일자 2013년04월24일

(65) 공개번호 10-2013-0130706

(43) 공개일자 2013년12월02일

(86) 국제출원번호 PCT/US2011/057623

(87) 국제공개번호 WO 2012/061109  
국제공개일자 2012년05월10일

(30) 우선권주장  
61/406,438 2010년10월25일 미국(US)

(56) 선행기술조사문현  
US20040225632 A1\*

\*는 심사관에 의하여 인용된 문현

전체 청구항 수 : 총 23 항

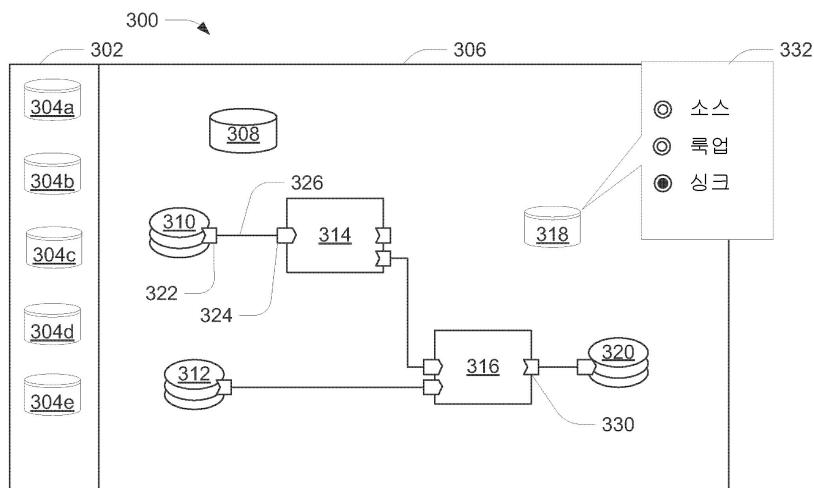
심사관 : 이복현

(54) 발명의 명칭 컴퓨터 프로그램을 나타내는 데이터플로우 그래프 내의 데이터세트 객체의 관리

**(57) 요 약**

그래프 기반의 데이터 처리를 위한 데이터세트 객체를 관리하는 방법은, 대응하는 데이터세트를 각각 나타내는 하나 이상의 데이터세트 객체로 이루어진 그룹(114)을 데이터 저장 시스템(116)에 저장하는 단계와, 그룹 내의 적어도 제1 데이터세트 객체와 데이터 처리 시스템 내에서의 데이터의 처리를 위한 데이터플로우 그래프의 적어도 제1 노드를 연관(210, 236)시키는 단계를 포함하며, 제1 노드는 데이터플로우 그래프 내의 링크에 의해 표현되는 데이터의 흐름에서의 데이터의 소스 또는 싱크를 나타내며, 제1 데이터세트 객체는 상이한 변환 로직이 제1 노드에 의해 처리되는 데이터에 적용되는 복수의 모드(402, 404, 406, 410, 412, 418)를 포함한다.

**대 표 도**



(72) 발명자

스탠필 크레이그 더블유.

미국 01773 매사추세츠주 린콜린 혁클베리 힐 로드  
43

웨이스 애덤

미국 02420 매사추세츠주 렉싱턴 로슨 애비뉴 15

## 명세서

### 청구범위

#### 청구항 1

데이터 처리 시스템에서 데이터를 처리하는 데이터플로우 그래프와 함께 사용되도록 데이터세트 객체를 관리하는 방법으로서,

하나 이상의 데이터세트 객체의 그룹을 데이터 저장 시스템에 저장하는 단계 – 상기 그룹 내의 각각의 데이터세트 객체는 개개의 데이터세트를 표현하고, 상기 그룹은 제1 데이터세트를 표현하는 제1 데이터세트 객체를 포함하며, 상기 제1 데이터세트 객체는 복수의 모드를 가지고, 상기 복수의 모드 각각은 개개의 변환 로직에 대응하며, 상기 복수의 모드는 적어도:

상기 제1 데이터세트로부터 데이터 레코드를 판독하기 위한 제1 모드로서, 상기 제1 모드는 상기 제1 데이터세트로부터 판독된 데이터 레코드에 적용될 제1 변환 로직과 연관되는, 제1 모드; 및

상기 제1 데이터세트에 데이터 레코드를 기입하기 위한 제2 모드로서, 상기 제2 모드는 데이터 레코드가 상기 제1 데이터세트에 기입되기 전에 데이터 레코드에 적용될 제2 변환 로직과 연관되며, 상기 제1 변환 로직은 상기 제2 변환 로직과는 상이한, 제2 모드를 포함함 –;

상기 데이터플로우 그래프를 특정하는 정보에 액세스하는 단계 – 상기 데이터플로우 그래프는 복수의 노드 및 상기 복수의 노드를 연결하는 복수의 링크를 포함하고, 상기 복수의 링크는 상기 복수의 노드에서 노드들 사이의 데이터의 흐름을 표현함 –;

상기 하나 이상의 데이터세트 객체의 그룹으로부터 상기 제1 데이터세트 객체를 식별하는 단계;

상기 제1 데이터세트 객체에 대해 상기 복수의 모드로부터 모드를 식별하는 단계;

상기 데이터플로우 그래프를 특정하는 액세스된 정보에 기초하여, 상기 데이터플로우 그래프의 복수의 노드에 제1 노드를 부가하는 단계; 및

상기 제1 노드에 의해 처리된 데이터가 식별된 모드와 연관되는 변환 로직에 따라 처리되도록 상기 데이터플로우 그래프의 상기 제1 노드와 상기 제1 데이터세트 객체 간에 연관을 생성하는 단계를 포함하는, 데이터세트 객체를 관리하는 방법.

#### 청구항 2

제1항에 있어서,

상기 제1 노드와 상기 제1 데이터세트 객체 간에 연관을 생성하는 단계는:

링크를 이용하여 상기 데이터플로우 그래프에서 상기 제1 노드를 적어도 하나의 다른 노드에 연결하는 단계를 포함하는, 데이터세트 객체를 관리하는 방법.

#### 청구항 3

제1항에 있어서,

상기 제1 데이터세트 객체의 상기 복수의 모드 중 하나는 상기 데이터플로우 그래프와는 상이한 제2 데이터플로우 그래프에 의해 표현되는, 데이터세트 객체를 관리하는 방법.

#### 청구항 4

제1항에 있어서,

상기 제1 데이터세트 객체에 대한 상기 복수의 모드는 데이터세트로의 데이터 레코드를 판독하기 위한 다수의 모드를 포함하고, 상기 다수의 모드는 상기 제1 모드를 포함하며, 상기 다수의 모드 각각은 상기 제1 데이터세트로부터 판독된 데이터 레코드에 적용될 개개의 변환 로직과 연관되는, 데이터세트 객체를 관리하는 방법.

**청구항 5**

제1항에 있어서,

상기 제1 데이터세트 객체에 대한 상기 복수의 모드는 데이터세트에 데이터 레코드를 기입하기 위한 다수의 모드를 포함하고, 상기 다수의 모드는 상기 제2 모드를 포함하며, 상기 다수의 모드 각각은 데이터 레코드가 상기 제1 데이터세트에 기입되기 전에 데이터 레코드에 적용될 개개의 변환 로직과 연관되는, 데이터세트 객체를 관리하는 방법.

**청구항 6**

제1항에 있어서,

상기 제1 데이터세트에서의 데이터 레코드는 XML 포맷으로 저장되며, 상기 제1 변환 로직은 실행되는 경우 상기 데이터 레코드를 상기 XML 포맷으로부터 일련의 순차적 데이터 레코드로 변환하는, 데이터세트 객체를 관리하는 방법.

**청구항 7**

제1항에 있어서,

상기 제2 변환 로직은 실행되는 경우 일련의 순차적 데이터 레코드를 XML 포맷으로 변환하는, 데이터세트 객체를 관리하는 방법.

**청구항 8**

제1항에 있어서,

상기 복수의 모드로부터 모드를 식별하는 단계는 그래픽 사용자 인터페이스를 통해 제공되는 사용자 입력에 기초하여 수행되며, 상기 사용자 입력은 상기 모드의 선택을 나타내는, 데이터세트 객체를 관리하는 방법.

**청구항 9**

제1항에 있어서,

상기 제1 데이터세트 객체를 식별하는 단계는:

캔버스와 상기 그룹 내의 데이터세트 객체의 카탈로그를 표시하는 그래픽 사용자 인터페이스에서, 상기 제1 데이터세트 객체의 선택 및 상기 캔버스 상의 위치에서 그 배치를 나타내는 제1 사용자 입력을 수신하는 단계를 포함하는, 데이터세트 객체를 관리하는 방법.

**청구항 10**

컴퓨터 프로그램을 저장하는 비일시적 컴퓨터 판독가능 저장 매체로서,

적어도 하나의 프로세서에 의해 실행되는 경우, 상기 적어도 하나의 프로세서로 하여금, 데이터 처리 시스템에서 데이터를 처리하는 데이터플로우 그래프와 함께 사용되도록 데이터세트 객체를 관리하는 방법을 수행하도록 하고, 상기 방법은:

하나 이상의 데이터세트 객체의 그룹을 데이터 저장 시스템에 저장하는 단계 – 상기 그룹 내의 각각의 데이터세트 객체는 개개의 데이터세트를 표현하고, 상기 그룹은 제1 데이터세트를 표현하는 제1 데이터세트 객체를 포함하며, 상기 제1 데이터세트 객체는 복수의 모드를 가지고, 상기 복수의 모드 각각은 개개의 변환 로직에 대응하며, 상기 복수의 모드는 적어도:

상기 제1 데이터세트로부터 데이터 레코드를 판독하기 위한 제1 모드로서, 상기 제1 모드는 상기 제1 데이터세트로부터 판독된 데이터 레코드에 적용될 제1 변환 로직과 연관되는, 제1 모드; 및

상기 제1 데이터세트에 데이터 레코드를 기입하기 위한 제2 모드로서, 상기 제2 모드는 데이터 레코드가 상기 제1 데이터세트에 기입되기 전에 데이터 레코드에 적용될 제2 변환 로직과 연관되며, 상기 제1 변환 로직은 상기 제2 변환 로직과는 상이한, 제2 모드를 포함함 –;

상기 데이터플로우 그래프를 특정하는 정보에 액세스하는 단계 – 상기 데이터플로우 그래프는 복수의 노드 및

상기 복수의 노드를 연결하는 복수의 링크를 포함하고, 상기 복수의 링크는 상기 복수의 노드에서 노드들 사이의 데이터의 흐름을 표현함 –;

상기 하나 이상의 데이터세트 객체의 그룹으로부터 상기 제1 데이터세트 객체를 식별하는 단계;

상기 제1 데이터세트 객체에 대해 상기 복수의 모드로부터 모드를 식별하는 단계;

상기 데이터플로우 그래프를 특정하는 액세스된 정보에 기초하여, 상기 데이터플로우 그래프의 복수의 노드에 제1 노드를 부가하는 단계; 및

상기 제1 노드에 의해 처리된 데이터가 식별된 모드와 연관되는 변환 로직에 따라 처리되도록 상기 데이터플로우 그래프의 상기 제1 노드와 상기 제1 데이터세트 객체 간에 연관을 생성하는 단계를 포함하는, 컴퓨터 판독가능 저장 매체.

### 청구항 11

제10항에 있어서,

상기 제1 데이터세트 객체에 대한 상기 복수의 모드는 데이터세트로의 데이터 레코드를 판독하기 위한 다수의 모드를 포함하고, 상기 다수의 모드는 상기 제1 모드를 포함하며, 상기 다수의 모드 각각은 상기 제1 데이터세트로부터 판독된 데이터 레코드에 적용될 개개의 변환 로직과 연관되는, 컴퓨터 판독가능 저장 매체.

### 청구항 12

제10항에 있어서,

상기 제1 데이터세트 객체에 대한 상기 복수의 모드는 데이터세트에 데이터 레코드를 기입하기 위한 다수의 모드를 포함하고, 상기 다수의 모드는 상기 제2 모드를 포함하며, 상기 다수의 모드 각각은 데이터 레코드가 상기 제1 데이터세트에 기입되기 전에 데이터 레코드에 적용될 개개의 변환 로직과 연관되는, 컴퓨터 판독가능 저장 매체.

### 청구항 13

제10항에 있어서,

상기 제1 데이터세트에서의 데이터 레코드는 XML 포맷으로 저장되며, 상기 제1 변환 로직은 실행되는 경우 상기 데이터 레코드를 상기 XML 포맷으로부터 일련의 순차적 데이터 레코드로 변환하는, 컴퓨터 판독가능 저장 매체.

### 청구항 14

제10항에 있어서,

상기 복수의 모드로부터 모드를 식별하는 단계는 그래픽 사용자 인터페이스를 통해 제공되는 사용자 입력에 기초하여 수행되며, 상기 사용자 입력은 상기 모드의 선택을 나타내는, 컴퓨터 판독가능 저장 매체.

### 청구항 15

제10항에 있어서,

상기 제1 데이터세트 객체를 식별하는 단계는:

캔버스와 상기 그룹 내의 데이터세트 객체의 카탈로그를 표시하는 그래픽 사용자 인터페이스에서, 상기 제1 데이터세트 객체의 선택 및 상기 캔버스 상의 위치에서의 그 배치를 나타내는 제1 사용자 입력을 수신하는 단계를 포함하는, 컴퓨터 판독가능 저장 매체.

### 청구항 16

데이터 처리 시스템에서 데이터를 처리하는 데이터플로우 그래프와 함께 사용되도록 데이터세트 객체를 관리하기 위한 컴퓨팅 시스템으로서,

하나 이상의 데이터세트 객체의 그룹을 저장하는 데이터 저장 시스템; 및

상기 데이터 저장 시스템에 연결되는 적어도 하나의 프로세서를 포함하고,

상기 그룹 내의 각각의 데이터세트 객체는 개개의 데이터세트를 표현하고, 상기 그룹은 제1 데이터세트를 표현하는 제1 데이터세트 객체를 포함하며, 상기 제1 데이터세트 객체는 복수의 모드를 가지고, 상기 복수의 모드 각각은 개개의 변환 로직에 대응하며, 상기 복수의 모드는 적어도:

상기 제1 데이터세트로부터 데이터 레코드를 판독하기 위한 제1 모드로서, 상기 제1 모드는 상기 제1 데이터세트로부터 판독된 데이터 레코드에 적용될 제1 변환 로직과 연관되는, 제1 모드;

상기 제1 데이터세트에 데이터 레코드를 기입하기 위한 제2 모드로서, 상기 제2 모드는 데이터 레코드가 상기 제1 데이터세트에 기입되기 전에 데이터 레코드에 적용될 제2 변환 로직과 연관되며, 상기 제1 변환 로직은 상기 제2 변환 로직과는 상이한, 제2 모드를 포함하며,

상기 적어도 하나의 프로세서는:

상기 데이터플로우 그래프를 특정하는 정보에 액세스하고 – 상기 데이터플로우 그래프는 복수의 노드 및 상기 복수의 노드를 연결하는 복수의 링크를 포함하고, 상기 복수의 링크는 상기 복수의 노드에서 노드들 사이의 데이터의 흐름을 표현함 –;

상기 하나 이상의 데이터세트 객체의 그룹으로부터 상기 제1 데이터세트 객체를 식별하며;

상기 제1 데이터세트 객체에 대해 상기 복수의 모드로부터 모드를 식별하고;

상기 데이터플로우 그래프를 특정하는 액세스된 정보에 기초하여, 상기 데이터플로우 그래프의 복수의 노드에 제1 노드를 부가하고;

상기 제1 노드에 의해 처리된 데이터가 식별된 모드와 연관되는 변환 로직에 따라 처리되도록 상기 데이터플로우 그래프의 상기 제1 노드와 상기 제1 데이터세트 객체 간에 연관을 생성하도록 구성되는, 데이터세트 객체를 관리하기 위한 컴퓨팅 시스템.

## 청구항 17

제16항에 있어서,

상기 제1 데이터세트 객체에 대한 상기 복수의 모드는 데이터세트로의 데이터 레코드를 판독하기 위한 다수의 모드를 포함하고, 상기 다수의 모드는 상기 제1 모드를 포함하며, 상기 다수의 모드 각각은 상기 제1 데이터세트로부터 판독된 데이터 레코드에 적용될 개개의 변환 로직과 연관되는, 데이터세트 객체를 관리하기 위한 컴퓨팅 시스템.

## 청구항 18

제16항에 있어서,

상기 제1 데이터세트 객체에 대한 상기 복수의 모드는 데이터세트에 데이터 레코드를 기입하기 위한 다수의 모드를 포함하고, 상기 다수의 모드는 상기 제2 모드를 포함하며, 상기 다수의 모드 각각은 데이터 레코드가 상기 제1 데이터세트에 기입되기 전에 데이터 레코드에 적용될 개개의 변환 로직과 연관되는, 데이터세트 객체를 관리하기 위한 컴퓨팅 시스템.

## 청구항 19

제16항에 있어서,

상기 제1 데이터세트에서의 데이터 레코드는 XML 포맷으로 저장되며, 상기 제1 변환 로직은 실행되는 경우 상기 데이터 레코드를 상기 XML 포맷으로부터 일련의 순차적 데이터 레코드로 변환하는, 데이터세트 객체를 관리하기 위한 컴퓨팅 시스템.

## 청구항 20

제16항에 있어서,

상기 제2 변환 로직은 실행되는 경우 일련의 순차적 데이터 레코드를 XML 포맷으로 변환하는, 데이터세트 객체를 관리하기 위한 컴퓨팅 시스템.

## 청구항 21

제16항에 있어서,

상기 복수의 모드로부터 모드를 식별하는 단계는 그래픽 사용자 인터페이스를 통해 제공되는 사용자 입력에 기초하여 수행되며, 상기 사용자 입력은 상기 모드의 선택을 나타내는, 데이터세트 객체를 관리하기 위한 컴퓨팅 시스템.

## 청구항 22

제16항에 있어서,

상기 제1 데이터세트 객체를 식별하는 단계는:

캔버스와 상기 그룹 내의 데이터세트 객체의 카탈로그를 표시하는 그래픽 사용자 인터페이스에서, 상기 제1 데이터세트 객체의 선택 및 상기 캔버스 상의 위치에서의 그 배치를 나타내는 제1 사용자 입력을 수신하는 단계를 포함하는, 데이터세트 객체를 관리하기 위한 컴퓨팅 시스템.

## 청구항 23

데이터 처리 시스템에서 데이터를 처리하는 데이터플로우 그래프와 함께 사용되도록 데이터세트 객체를 관리하기 위한 컴퓨팅 시스템으로서,

하나 이상의 데이터세트 객체의 그룹을 데이터 저장 시스템에 저장하기 위한 수단 — 상기 그룹 내의 각각의 데이터세트 객체는 개개의 데이터세트를 표현하고, 상기 그룹은 제1 데이터세트를 표현하는 제1 데이터세트 객체를 포함하며, 상기 제1 데이터세트 객체는 복수의 모드를 가지고, 상기 복수의 모드 각각은 개개의 변환 로직에 대응하며, 상기 복수의 모드는 적어도:

상기 제1 데이터세트로부터 데이터 레코드를 판독하기 위한 제1 모드로서, 상기 제1 모드는 상기 제1 데이터세트로부터 판독된 데이터 레코드에 적용될 제1 변환 로직과 연관되는, 제1 모드; 및

상기 제1 데이터세트에 데이터 레코드를 기입하기 위한 제2 모드로서, 상기 제2 모드는 데이터 레코드가 상기 제1 데이터세트에 기입되기 전에 데이터 레코드에 적용될 제2 변환 로직과 연관되며, 상기 제1 변환 로직은 상기 제2 변환 로직과는 상이한, 제2 모드를 포함함 —;

상기 데이터플로우 그래프를 특정하는 정보에 액세스하기 위한 수단 — 상기 데이터플로우 그래프는 복수의 노드 및 상기 복수의 노드를 연결하는 복수의 링크를 포함하고, 상기 복수의 링크는 상기 복수의 노드에서 노드들 사이의 데이터의 흐름을 표현함 —;

상기 하나 이상의 데이터세트 객체의 그룹으로부터 상기 제1 데이터세트 객체를 식별하기 위한 수단;

상기 제1 데이터세트 객체에 대해 상기 복수의 모드로부터 모드를 식별하기 위한 수단;

상기 데이터플로우 그래프를 특정하는 액세스된 정보에 기초하여, 상기 데이터플로우 그래프의 복수의 노드에 제1 노드를 부가하기 위한 수단; 및

상기 제1 노드에 의해 처리된 데이터가 식별된 모드와 연관되는 변환 로직에 따라 처리되도록 상기 데이터플로우 그래프의 상기 제1 노드와 상기 제1 데이터세트 객체 간에 연관을 생성하기 위한 수단을 포함하는, 데이터세트 객체를 관리하기 위한 컴퓨팅 시스템.

## 발명의 설명

### 기술 분야

[0001]

본 개시는 데이터세트 객체의 관리에 관한 것이다.

### 배경 기술

[0002]

데이터플로우 그래프는 데이터를 처리한다. 데이터플로우 그래프는 데이터 소스, 데이터 싱크, 및 성분을 포함한다. 데이터 소스, 데이터 싱크 및 성분은 하나의 성분으로부터 다른 성분으로, 데이터 소스로부터 성분으로, 또는 성분으로부터 데이터 싱크로 데이터가 흐르도록 하는 성분에 의해 연결되어 있다. 데이터플로우 그래프는

연결을 나타내는 링크(또는 "아크"라고도 함)에 의해 연결된 데이터 소스, 데이터 싱크 및 성분을 나타내는 노드("정점"이라고도 함)를 갖는 방향성 그래프에 의해 시작적으로 표현될 수 있다.

### 발명의 내용

[0003] 하나의 관점에서, 일반적으로, 그래프 기반의 데이터 처리(graph-based data processing)를 위한 데이터세트 객체를 관리하는 방법은, 대응하는 데이터세트(data set)를 각각 나타내는 하나 이상의 데이터세트 객체(data set object)로 이루어진 그룹을 데이터 저장 시스템(data storage system)에 저장하는 단계; 및 그룹 내의 적어도 제1 데이터세트 객체와 데이터 처리 시스템 내에서의 데이터의 처리를 위한 데이터플로우 그래프(dataflow graph)의 적어도 제1 노드(node)를 연관(associate)시키는 단계를 포함하며, 제1 노드는 데이터플로우 그래프 내의 링크에 의해 표현되는 데이터의 흐름에서의 데이터의 소스(source) 또는 싱크(sink)를 나타내며, 제1 데이터세트 객체는 상이한 변환 로직(transformational logic)이 제1 노드에 의해 처리되는 데이터에 적용되는 복수의 모드를 포함한다.

[0004] 관점은 이하의 특징 중의 하나 이상을 포함할 수 있다.

[0005] 연관된 데이터세트 객체와 연관된 노드를 연관시키는 단계는, 데이터플로우 그래프에 대응하는 데이터 구조에 액세스하는 단계로서, 데이터플로우 그래프가 제1 노드를 포함하는 복수의 노드와 노드를 연결하며 노드 사이에서의 데이터의 흐름을 나타내는 링크를 포함하는, 데이터 구조에 액세스하는 단계; 사용자 입력에 기초하여, 그룹으로부터, 입력 포트 또는 출력 포트 중의 하나 이상과 변환 로직을 각각 포함하는 복수의 모드를 포함하는 제1 데이터세트 객체의 선택을 수신하는 단계; 사용자 입력에 기초하여, 제1 데이터세트 객체에 대한 모드의 선택을 수신하는 단계; 선택된 모드를 수신한 후에, 데이터플로우 그래프 내에, 선택된 모드의 변환 로직, 임의의 입력 포트 및 임의의 출력 포트를 포함하는 제1 노드를 생성하는 단계; 및 링크를 사용하여 데이터플로우 그래프 내의 하나 이상의 노드에 제1 노드를 연결하는 단계를 포함한다.

[0006] 제1 데이터세트 객체는 하나 이상의 파라미터를 더 포함하며, 제1 노드는 파라미터에 대한 값을 더 포함한다.

[0007] 복수의 모드 중의 제1 모드의 변환 로직은 데이터 저장장소(data store)로부터 데이터 레코드(data record)를 판독하는 변환 로직을 포함한다.

[0008] 복수의 모드 중의 제2 모드의 변환 로직은 데이터 저장장소에 데이터 레코드를 기입하는 변환 로직을 포함한다.

[0009] 제1 노드는 제1 데이터세트 객체를 식별하는 정보를 포함한다.

[0010] 본 방법은, 사용자 입력에 기초하여, 그룹으로부터 제1 데이터세트 객체의 선택을 수신하는 단계; 및 제1 데이터세트 객체를 식별하는 노드를 각각 포함하는 복수의 데이터플로우 그래프를 식별하는 단계를 더 포함할 수 있다.

[0011] 데이터세트 객체의 각각의 모드는 노드 및 노드를 연결하며 노드 사이에서 데이터의 흐름을 나타내는 링크를 포함하는 별개의 데이터플로우 그래프에 의해 표현된다.

[0012] 연관된 데이터세트 객체와 연관된 노드를 연관시키는 단계는, 제1 노드를 포함하는 복수의 데이터플로우 그래프에 대응하는 데이터 구조에 액세스하는 단계로서, 데이터플로우 그래프는 제1 노드를 포함하고, 각각의 데이터플로우 그래프는 복수의 노드와 노드를 연결하는 링크를 포함하며, 노드는 노드 사이에서의 데이터의 흐름을 나타내는 링크와 변환 로직을 포함하는, 데이터 구조에 액세스하는 단계; 동일한 데이터 저장장소에 액세스하는 복수의 데이터플로우 그래프의 복수의 노드를 식별하는 단계로서, 하나 이상의 식별된 노드가 하나 이상의 파라미터를 포함하는, 식별하는 단계; 데이터세트 객체를 생성하는 단계로서, 식별된 노드의 변환 로직에 기초하여, 생성된 데이터세트 객체에 대한 복수의 모드를 식별하는 단계, 및 식별된 노드와 연관된 파라미터에 기초하여, 생성된 데이터 객체에 대한 파라미터를 식별하는 단계를 포함하는, 데이터세트 객체를 생성하는 단계; 및 생성된 데이터세트 객체를 그룹 내에 저장하는 단계를 포함한다.

[0013] 본 방법은, 생성된 데이터세트 객체를 식별하는 정보에 기초하여, 데이터 저장장소에 액세스하기 위해 복수의 노드의 각각을 갱신(update)하는 단계를 더 포함할 수 있다.

[0014] 동일한 데이터 저장장소에 액세스하는 복수의 데이터플로우 그래프의 복수의 노드를 식별하는 단계는, 복수의 데이터플로우 그래프의 각각을 실행(execute)하는 단계; 데이터플로우 그래프를 실행하는 중에, 복수의 데이터플로우 그래프의 노드의 파라미터에 할당된 파라미터 값을 기록(record)하는 단계; 및 기록된 파라미터 값에 기초하여, 복수의 노드를 식별하는 단계를 포함한다.

[0015] 다른 관점으로서, 일반적으로, 컴퓨터로 관리가능한 저장매체는 그래프 기반의 데이터 처리를 위한 데이터세트 객체를 관리하기 위한 컴퓨터 프로그램을 저장한다. 이러한 컴퓨터 프로그램은, 컴퓨팅 시스템으로 하여금, 대응하는 데이터세트를 각각 나타내는 하나 이상의 데이터세트 객체로 이루어진 그룹을 데이터 저장 시스템에 저장하도록 하고, 및 그룹 내의 적어도 제1 데이터세트 객체와 데이터 처리 시스템 내에서의 데이터의 처리를 위한 데이터플로우 그래프의 적어도 제1 노드를 연관시키도록 하는 명령어를 포함하며, 제1 노드는 데이터플로우 그래프 내의 링크에 의해 표현되는 데이터의 흐름에서의 데이터의 소스 또는 싱크를 나타내며, 제1 데이터세트 객체는 상이한 변환 로직이 제1 노드에 의해 처리되는 데이터에 적용되는 복수의 모드를 포함한다.

[0016] 다른 관점으로서, 일반적으로, 그래프 기반의 데이터 처리를 위한 데이터세트 객체를 관리하기 위한 컴퓨팅 시스템은, 대응하는 데이터세트를 각각 나타내는 하나 이상의 데이터세트 객체로 이루어진 그룹을 저장하는 데이터 저장 시스템; 및 데이터 저장 시스템에 결합되어, 그룹 내의 적어도 제1 데이터세트 객체와 데이터 처리 시스템 내에서의 데이터의 처리를 위한 데이터플로우 그래프의 적어도 제1 노드를 연관시키도록 구성된 하나 이상의 프로세서를 포함하며, 제1 노드는 데이터플로우 그래프 내의 링크에 의해 표현되는 데이터의 흐름에서의 데이터의 소스 또는 싱크를 나타내며, 제1 데이터세트 객체는 상이한 변환 로직이 제1 노드에 의해 처리되는 데이터에 적용되는 복수의 모드를 포함한다.

[0017] 다른 관점으로서, 일반적으로, 그래프 기반의 데이터 처리를 위한 데이터세트 객체를 관리하기 위한 컴퓨팅 시스템은, 대응하는 데이터세트를 각각 나타내는 하나 이상의 데이터세트 객체로 이루어진 그룹을 저장하기 위한 수단; 및 그룹 내의 적어도 제1 데이터세트 객체와 데이터 처리 시스템 내에서의 데이터의 처리를 위한 데이터플로우 그래프의 적어도 제1 노드를 연관시키기 위한 수단을 포함하며, 제1 노드는 데이터플로우 그래프 내의 링크에 의해 표현되는 데이터의 흐름에서의 데이터의 소스 또는 싱크를 나타내며, 제1 데이터세트 객체는 상이한 변환 로직이 제1 노드에 의해 처리되는 데이터에 적용되는 복수의 모드를 포함한다.

[0018] 관점은 이하의 장점을 중 하나 이상을 포함할 수 있다. 데이터세트 객체는 다수의 데이터플로우 그래프에서 재사용될 수 있다. 데이터 레코드의 생성은 단순화될 수 있다. 데이터의 존성이 판정될 수 있다. 애플리케이션 개발이 단순화될 수 있다.

[0019] 본 발명의 다른 특징과 장점에 대해서는, 이하의 상세한 설명과 청구범위로부터 명백해질 것이다.

### 도면의 간단한 설명

[0020] 도 1은 데이터세트 객체가 사용될 수 있는 데이터 처리 시스템(100)의 예를 나타낸다.

도 2a 및 2b는 데이터세트 객체를 관리하는 데에 사용될 수 있는 절차(200)의 예를 위한 플로차트이다.

도 3은 개발 환경의 사용자 인터페이스의 예를 나타낸다.

도 4는 데이터세트 객체와 연관된 퍼스널리티의 예를 나타낸다.

도 5는 카탈로그 내의 데이터세트 객체와 데이터플로우 그래프 간의 관계를 나타낸다.

도 6은 카탈로그 내의 데이터세트 객체와 기준 노드의 연관을 나타낸다.

도 7은 파라미터 값에 기초한 데이터세트 객체와 노드의 연관을 나타낸다.

### 발명을 실시하기 위한 구체적인 내용

[0021] 도 1은 데이터세트 객체 기술이 사용될 수 있는 데이터 처리 시스템(100)의 예를 나타낸다. 이 데이터 처리 시스템(100)은 다양한 저장 포맷(예를 들어, 데이터베이스 테이블, 스프레드시트 파일, 플랫 텍스트 파일, 또는 메인프레임에 의해 사용되는 네이티브 포맷) 중의 임의의 형태로 데이터를 저장할 수 있는 온라인 데이터 스토리지에의 접속이나 저장 장치 등과 같은 데이터의 하나 이상의 소스를 포함할 수 있는 데이터 소스(102)를 포함한다. 실행 환경(104)은 실행 실행 모듈(pre-execution module)(106)과 실행 모듈(112)을 포함한다. 실행 환경(104)은 UNIX 운영체제 등의 적절한 운영체제의 제어 하에서 하나 이상의 범용 컴퓨터에 대해 호스트 기능을 할 수 있다. 예를 들어, 실행 환경(104)은 로컬(예를 들어, SMP 컴퓨터 등의 다중 프로세서 시스템) 또는 로컬 분산된(예를 들어, 클러스터 또는 MPP로서 연결된 다중 프로세서), 또는 원격이나 원격 분산된(예를 들어, 근거리 통신망(LAN) 및/또는 광역 통신망(WAN)에 의해 연결된 다중 프로세서), 또는 이들이 임의로 조합된 다수의 중앙 처리장치(CPU)를 이용하는 컴퓨터 시스템의 구성을 포함하는 다중 노드 병렬 컴퓨팅 환경을 포함할 수 있다.

[0022] 실행 실행 모듈(106)은 데이터플로우 그래프를 컴파일하거나, 컴파일한 데이터플로우 그래프를 실행 환경(104)

에 액세스가능한 데이터 저장 시스템(116)과의 사이에서 저장시키거나 로딩하는 등의 데이터플로우 그래프 등의 실행가능한 프로그램을 실행하기 위한 준비로서 다양한 작업을 수행하거나, 데이터플로우 그래프를 관리하는 것과 관련된 작업을 수행하도록 구성된다. 실행 모듈(112)은 데이터 소스(102)로부터 데이터를 판독하고, 판독한 데이터를, 데이터플로우 그래프의 형태로 표현된 프로그램을 포함하는 실행가능한 프로그램을 사용해서 처리한다.

[0023] 데이터 소스(102)를 제공하는 저장 장치는 예를 들어, 실행 환경(104)을 실행하는 컴퓨터에 접속된 저장 매체 [예를 들어, 하드 드라이브(108)]에 저장되는 실행 환경(104)에 대해 로컬에 위치하거나, 원격 접속을 통해 실행 환경(104)을 실행하는 컴퓨터와 통신을 행하여 원격 시스템[예를 들어, 메인 프레임(110)]에 대해 호스트 기능을 하는 실행 환경(104)에 대해 원격에 위치할 수 있다.

[0024] 실행 모듈(112)은 데이터플로우 그래프를 실행하기 위한 실행 환경(104)에 액세스가능한 데이터 저장 시스템(116)에 저장된 데이터세트 객체의 카탈로그(114)를 사용한다. 이 카탈로그(114)는 전처리 모듈(106)에 의해 관리되어, 여러 상이한 모드로 설정될 수 있는 데이터세트 객체를 제공할 수 있는데, 이에 대해서는 후술한다. 카탈로그는 데이터세트 객체의 그룹을 데이터 구조 또는 특정의 위치(예를 들어, 파일 시스템 디렉토리)에 저장하거나, 다양한 위치에 저장된 데이터 객체의 그룹을 식별하는 정보를 포함할 수 있다. 일례로, 카탈로그는 검색 등의 동작을 할 수 있도록 하는 인덱스 또는 메타데이터 등의 정보를 포함한다. 데이터 저장 시스템(116)은 개발 환경(118)에 액세스할 수 있으며, 개발 환경에서는 개발자(120)가 데이터플로우 그래프를 개발하고 카탈로그(114)와 상호작용할 수 있다.

[0025] 일례로, 개발 환경(118)은 애플리케이션을 정점과 방향성 링크를 포함하는 데이터플로우 그래프로서 개발하기 위한 시스템으로서, 정점(성분 또는 데이터세트를 나타냄)은 정점들 사이가 방향성 링크(작업 요소의 흐름을 나타냄)에 의해 연결된다. 예를 들어, 이러한 환경은, 본원에서 인용에 의해 포함하는 "Managing Parameters for Graph-Based Applications"란 명칭의 미국 공개번호 2007/0011668호에 상세하게 개시되어 있다. 이러한 그래프 기반의 계산을 실행하기 위한 시스템은, 본원에서 인용에 의해 포함하는 "EXECUTING COMPUTATIONS EXPRESSED AS GRAPHS"란 명칭의 미국특허 5,566,072호에 개시되어 있다. 이러한 시스템에 의해 이루어진 데이터플로우 그래프는 그래프 성분으로 표현된 개별 프로세스에 정보를 삽입하거나 프로세스로부터 정보를 취득하는 과정, 이러한 프로세스 사이에서 정보를 이동시키는 과정, 프로세스를 위한 실행 순서를 정하는 과정을 포함하는 방법을 제공한다. 이 시스템은 프로세스 간 통신 방법을 선택(예를 들어, 그래프의 링크에 의한 통신 경로는 TCP/IP 또는 UNIX 도메인 소켓을 이용하거나, 프로세스 사이에서 데이터를 전달하기 위한 공유 메모리를 사용할 수 있음)하는 알고리즘을 포함한다.

[0026] 실행 모듈(112)은 여러 형태의 데이터베이스 시스템을 포함하는 다양한 유형의 시스템으로부터 데이터를 수신할 수 있다. 데이터는 널 값(null value)을 포함할 수 있는, 각각의 필드("속성" 또는 "컬럼"이라고도 함)에 대한 값을 갖는 레코드로서 구성될 수 있다. 실행 모듈(112)은, 데이터 소스로부터 데이터를 판독하면, 해당 데이터 소스 내의 레코드에 관한 몇몇 개시 포맷 정보로 시작하는 것이 전형적이다. 어떤 경우에는, 데이터 소스의 레코드 구조를 처음에는 알지 못할 수 있으며, 데이터 소스를 분석한 이후에 정해질 수도 있다. 레코드에 관한 개시 정보로는 개별 값을 나타내는 비트의 수, 레코드 내의 필드의 순서, 및 비트로 나타내는 값의 유형(예를 들어, 스트링, 부호있는 정수/부호없는 정수) 등이 있다.

[0027] 개발 환경(118)은 개발자(120)가 실행 환경(104)에서 실행될 데이터플로우 그래프를 구성 및/또는 편집하도록 하는 데에 사용되는 인터페이스를 제공한다. 이 데이터플로우 그래프는 데이터 저장 시스템(116)에 저장된 실행 가능한 데이터 구조와 연관되며, 노드 및 노드를 연결하는 링크를 포함하는 방향성 그래프로서 개발 환경(118) 내에 표현된다. 개발 환경은 개발 중인 데이터플로우 그래프에 대응하는 데이터 구조에 액세스할 수 있다. 데이터플로우 그래프의 노드는 데이터 소스, 데이터 싱크, 및 데이터 처리 성분을 나타낸다. 링크는 노드 사이에서의 데이터의 흐름을 나타낸다. 개발 환경(118)은 또한 데이터세트 객체의 카탈로그(114)에 액세스할 수 있다. 카탈로그(114)는 데이터 저장 시스템(116) 내의 데이터 저장 장소 또는 다른 데이터 저장 장소에 위치할 수 있다.

[0028] 카탈로그(114)는 사용자 입력에 기초하여 선택될 수 있는 하나 이상의 데이터세트 객체를 포함한다. 데이터세트 객체는 이러한 데이터세트 객체가 제공하는 기능성(functionality)을 판정하는 다수의 모드["퍼스널리티"(personality)라고 함]를 포함하는 데이터플로우 그래프 노드이다. 예를 들어, 데이터세트 객체는 데이터 소스 퍼스널리티와 데이터 싱크 퍼스널리티를 포함할 수 있다. 데이터플로우 그래프 내의 노드로서 사용하기 위한 데이터세트 객체가 선택되면, 이 데이터세트 객체는 예를 들어, 사용자 입력에 기초하여, 퍼스널리티 중의 하나

를 채택한다. 예를 들어, 개발 환경(118)은 카탈로그의 표현을 포함하는 그래픽 사용자 인터페이스(GUI)와 데이터플로우 그래프의 표현(예를 들어, 링크로 연결된 노드)을 포함하는 캔버스(canvas)를 포함할 수 있다. 사용자 [예를 들어, 개발자(120)]는 캔버스 위에서 카탈로그로부터 데이터세트 객체를 드래그 및 드롭할 수 있다. 데이터세트 객체의 표현이 캔버스 위에 위치하게 되면, 사용자는 그 데이터세트 객체에 대한 퍼스넬러티를 선택할 수 있다.

[0029] 선택된 퍼스넬러티는 데이터플로우 그래프 내에 신규 노드를 생성하는 데에 사용된다. 선택된 퍼스넬러티는 몇 가지 방식으로 데이터를 변경하는 변환 로직(transformational logic)과 연관된다. 예를 들어, 데이터 소스 퍼스넬러티는 데이터 저장장소(예를 들어, 데이터베이스 테이블, 파일, 또는 다른 데이터 저장장소)로부터 데이터 레코드를 판독하고, 몇 가지 변환을 수행하는 변환 로직을 포함할 수 있다. 데이터 싱크 퍼스넬러티는 몇 가지 변환을 수행한 후에 데이터 저장장소에 데이터 레코드를 기록하는 변환 로직을 포함할 수 있다. 일례로, 데이터 세트 객체의 퍼스넬러티의 변환 로직은 데이터플로우 그래프를 이용해서 구현될 수 있다. 일례로, 데이터세트 객체의 각각의 퍼스넬러티는 노드 및 노드를 연결하는 링크를 포함하는, 여러 상이한 각각의 데이터플로우 그래프와 연관된다.

[0030] 각각의 퍼스넬러티는 입력 포트(예를 들어, 데이터 싱크 퍼스넬러티), 출력 포트(예를 들어, 데이터 소스 퍼스넬러티), 또는 입력 포트 및 출력 포트 모두(예를 들어, 데이터 변환 퍼스넬러티)를 갖거나, 포트를 갖지 않을 수 있다(예를 들어, 루프 테이블 퍼스넬러티). 데이터세트 객체에 대한 퍼스넬러티의 선택을 수신하는 것에 의해 생성되는 신규 노드는 변환 로직과 선택된 퍼스넬러티의 포트를 반영한다. 일례로, 데이터세트 객체는 데이터 소스의 위치를 식별한다. 데이터세트 객체는 데이터 소스의 위치를 식별하는 파라미터를 포함할 수 있다. 일례로, 파라미터는 데이터플로우 그래프의 실행 중에 판정되는 값을 포함하는 동적 파라미터(dynamic parameters)를 사용해서 데이터 소스의 위치를 식별한다. 데이터 소스에 액세스하는 데이터세트 객체는 데이터의 포맷과 데이터가 압축되었는지 암호화되었는지 여부를 식별할 수 있다.

[0031] 생성된 노드는 데이터플로우 그래프에 포함될 수 있는데, 생성된 노드를 링크를 사용해서 데이터플로우 그래프 내의 하나 이상의 노드에 연결시키는 것에 의해 포함될 수 있다. 일례로, 생성된 노드는 데이터를 데이터 소스로부터 데이터플로우 그래프에 제공하거나, 데이터를 데이터플로우 그래프로부터 데이터 저장장소에 저장한다.

[0032] 도 2a는 데이터세트 객체를 관리하기 위한 시스템에 의해 구현될 수 있는 프로세스(200)의 예를 나타내는 플로차트이다. 설명의 편의를 위해, 프로세스는 이 프로세스를 수행하는 시스템과 관련해서 설명한다.

[0033] 프로세스(200)는 데이터플로우 그래프에 대응하는 데이터 구조에 액세스한다(202). 데이터 구조는 노드 및 노드를 연결하는 링크를 포함하는데, 링크는 노드들 사이에서의 데이터의 흐름을 나타낸다.

[0034] 프로세스(200)는 데이터세트 객체로 이루어진 카탈로그를 제공한다(204). 앞서 설명한 바와 같이, 카탈로그는 하나 이상의 데이터세트 객체를 포함할 수 있다.

[0035] 프로세스(200)는 카탈로그로부터 데이터세트 객체의 선택을 수신한다(206). 예를 들어, 사용자는 캔버스 위에서 카탈로그로부터 데이터세트 객체를 드래그하는 것에 의해 데이터 카탈로그로부터 데이터세트 객체를 선택할 수 있다. 앞서 설명한 바와 같이, 데이터세트 객체는 다수의 퍼스넬러티를 포함하며, 각각의 퍼스넬러티는 변환 로직과, 하나 이상의 입력 포트 및/또는 하나 이상의 출력 포트를 포함한다.

[0036] 프로세스(200)는 데이터세트 객체에 대한 퍼스넬러티의 선택을 수신한다(208). 앞서 설명한 바와 같이, 사용자는 예를 들어, 무선 버튼이나 드롭다운 메뉴 상의 값을 선택하는 것에 의해, 선택된 데이터세트 객체의 퍼스넬러티를 선택할 수 있다.

[0037] 프로세스(200)는 데이터플로우 그래프 내에 노드를 생성한다(210). 생성된 노드는 선택된 퍼스넬러티의 임의의 입력 및 출력 포트와 변환 로직을 포함한다.

[0038] 프로세스(200)는 생성된 노드를 데이터플로우 내의 하나 이상의 노드에 연결한다(212).

[0039] 도 2b는 데이터세트 객체를 관리하기 위한 시스템에 의해 구현될 수 있는 프로세스(230)의 예를 나타내는 플로차트이다. 설명의 편의를 위해, 프로세스는 이 프로세스를 수행하는 시스템과 관련해서 설명한다.

[0040] 프로세스(230)는 다수의 데이터플로우 그래프에 액세스한다(232). 앞서 설명한 바와 같이, 각각의 데이터플로우 그래프는 노드와 이러한 노드를 연결하는 링크를 포함한다. 노드는 변환 로직과 링크를 포함하는데, 링크는 노드들 사이에서의 데이터의 흐름을 나타낸다.

[0041] 프로세스(230)는 동일한 데이터 저장장소에 액세스하는 다수의 노드를 식별한다(234). 각각의 노드는 파라미터와 변환 로직을 포함한다.

[0042] 프로세스(230)는 데이터세트 객체를 생성한다(236). 데이터세트 객체를 생성하기 위해, 프로세스는 다수의 노드의 변환 로직에 기초해서 다수의 퍼스넬러티를 식별한다(238). 프로세스(230)는 또한 노드와 연관된 파라미터에 기초해서 파라미터를 식별한다(240). 프로세스(230)는 생성된 데이터세트 객체를 데이터 카탈로그에 저장한다(242).

[0043] 도 3은 개발 환경, 예를 들어 도 1의 개발 환경(118)의 사용자 인터페이스(300)의 예를 나타낸다. 사용자 인터페이스(300)는 다수의 데이터세트 객체(304a-304e)를 포함하는 데이터세트 객체로 이루어진 카탈로그(302)를 포함한다. 사용자 인터페이스(300)는 또한 캔버스(306)를 포함하며, 이 캔버스에 의해 사용자는 데이터플로우 그래프를 작성 또는 변경할 수 있다. 캔버스는 데이터플로우 그래프 내에 포함된 노드의 표현을 표시한다. 본 예에서, 데이터플로우 그래프는 제1 데이터 소스(310), 제2 데이터 소스(312), 제1 성분 노드(314), 제2 성분 노드(316), 데이터 싱크(320), 및 루업 테이블(308)을 포함한다. 데이터플로우 그래프를 통한 데이터의 흐름은 링크에 의해 정의된다. 예를 들어, 제1 데이터 소스(310)는 링크(326)에 의해 성분 노드(314)의 입력 포트(324)에 연결되는 출력 포트(322)를 포함한다.

[0044] 사용자(도시 안 됨)는 데이터세트 객체[예를 들어, 데이터세트 객체(304b)]를 카탈로그(302)로부터 캔버스(306)로 드래그해서 신규 노드(318)를 생성할 수 있다. 사용자는 신규 노드(318)에 대한 퍼스넬러티를 선택하는데, 예를 들어 특성 인터페이스 요소(properties interface element)(332) 상의 무선 버튼을 이용해서 퍼스넬러티를 선택하는 것에 의해 신규 노드에 대한 퍼스넬러티를 선택한다. 퍼스넬러티가 선택되면, 신규 노드(318)는 해당 퍼스넬러티와 연관된 입력 포트, 출력 포트 및/또는 변환 로직을 제시한다. 앞서 설명한 바와 같이, 신규 객체는 사용자에 의해 설정될 수 있는 파라미터를 포함할 수 있다.

[0045] 도 4는 데이터세트 객체와 연관된 퍼스넬러티의 예를 나타낸다. 데이터세트 객체(400)는 하나 이상의 파라미터(414)와 연관될 수 있다. 예를 들어, 데이터세트 객체(400)는 레이블, 파일명, 레코드 포맷, 데이터베이스 테이블명, 또는 임의의 다른 정보를 정의하는 값을 가진 파라미터와 연관될 수 있다. 일례로, 파라미터 값은 실행 중에 동적으로 정해질 수 있다.

[0046] 데이터세트 객체(400)는 다수의 퍼스넬러티를 포함할 수 있으며, 각각의 퍼스넬러티는 자신의 입력 포트, 출력 포트, 및/또는 변환 로직을 포함하는 자신의 특징을 갖는다. 예를 들어, 데이터세트 객체는 파일, 데이터베이스 테이블 또는 다른 데이터 구조로부터 데이터를 판독하고, 판독한 데이터를 데이터플로우 그래프에 제공하는 데이터 소스 퍼스넬러티(404)를 포함할 수 있다. 데이터세트 객체(400)는 입력 포트에서의 데이터 레코드의 흐름을 허용하고, 이러한 데이터 레코드를 데이터베이스 테이블, 파일 또는 몇몇 다른 데이터 저장장소 등의 데이터 저장장소에 저장시키는 데이터 싱크 퍼스넬러티(410)를 포함할 수 있다. 데이터세트 객체(400)는 다수의 병렬 데이터 구조로부터 데이터를 판독하고 판독한 데이터를 데이터 플로우 그래프에 제공하는 병렬 데이터 소스 퍼스넬러티(402)를 포함할 수 있다. 마찬가지로, 데이터세트 객체(400)는 데이터 레코드의 스트림을 다수의 데이터 저장장소로 분할하는 병렬 데이터 싱크 퍼스넬러티(412)를 포함할 수 있다.

[0047] 데이터세트 객체(400)는 루업 테이블 퍼스넬러티(418)를 포함할 수 있다. 일반적으로, 루업 테이블은 입력 포트 또는 출력 포트를 포함하지 않는다. 루업 테이블 퍼스넬러티는 다른 노드에서의 신속한 데이터 검색을 위해 데이터 소스를 메모리로 판독한다.

[0048] 데이터세트 객체(400)는 하나 이상의 성분 퍼스넬러티, 예를 들어 퍼스넬러티(406)를 포함할 수 있다. 일반적으로, 성분 퍼스넬러티는 하나 이상의 입력 포트와 하나 이상의 출력 포트를 포함할 수 있으며, 데이터플로우에 대해 동작을 수행할 수 있다. 예를 들어, 성분은 데이터플로우를 필터링해서 특정의 값을 갖는 데이터 레코드를 식별할 수 있다.

[0049] 일례로, 파라미터는 데이터세트 객체 퍼스넬러티와 연관될 수 있다. 예를 들어, 병렬 데이터 싱크 퍼스넬러티(412)는 데이터플로우가 분할될 파일의 이름을 정의하는 파라미터(416)와 연관된다.

[0050] 일례로, 데이터세트 객체는 다수의 데이터 싱크 및 소스를 포함할 수 있다. 예를 들어, 데이터세트 객체는 데이터를 데이터베이스 테이블에 기입하는 제1 데이터 싱크 퍼스넬러티, 데이터를 플랫 파일에 기입하는 제2 데이터 싱크 퍼스넬러티, 및 파일 전송 프로토콜(FTP)을 사용하여 데이터를 원격 서버에 기입하는 제3 데이터 싱크 퍼스넬러티를 포함할 수 있다. 마찬가지로, 데이터세트 객체는 데이터베이스 테이블로부터 데이터를 판독하는 제1 데이터 소스 퍼스넬러티, 플랫 파일로부터 데이터를 판독하는 제2 데이터 소스 퍼스넬러티, 및 FTP를 사용하여

원격 서버에 있는 데이터에 액세스하는 제3 데이터 소스 퍼스널러티를 포함할 수 있다. 일례로, 다른 프로토콜이나 데이터 액세스 방법도 사용할 수 있다[예를 들어, 하이퍼텍스트 전송 프로토콜(HTTP) 및 확장 마크업 언어(XML) 문서]. 일례로, 데이터세트 객체에 대한 모든 데이터 소스 퍼스널러티는 동일한 출력을 생성하며 데이터세트 객체에 대한 모든 데이터 싱크 퍼스널러티는 동일한 입력을 허용한다. 따라서, 데이터플로우 그래프 내에 포함되는 어느 데이터 소스 퍼스널러티 또는 데이터 싱크 퍼스널러티라도 데이터플로우 그래프 내의 다른 노드에 대해 투명할 수 있다.

[0051] 앞서 설명한 바와 같이, 일례로, 데이터세트 퍼스널러티는 데이터에 대해 변환을 수행하는 데이터플로우 그래프로서 구현되는 변환 로직과 연관될 수 있다. 예를 들어, 데이터세트 객체의 데이터 소스 퍼스널러티는 XML 파일을 판독하고, 계층적 XML 데이터를 처리를 위한 일련의 순차적 레코드로 변환할 수 있다. 마찬가지로, 데이터세트 객체의 데이터 싱크 퍼스널러티는 일련의 순차적인 레코드를 허용하고, 레코드를 XML 파일에 저장할 계층적 XML 데이터로 변환할 수 있다.

[0052] 데이터세트 객체에 연관될 수 있는 다른 퍼스널러티는 데이터 소스로부터 데이터를 재생하고, 레코드를 캡신하는 데이터 소스에서 캡신하는 퍼스널러티를 포함한다.

[0053] 도 5는 카탈로그 내의 데이터세트 객체와 데이터플로우 그래프 간의 관계를 나타낸다. 카탈로그로부터의 데이터세트 객체가 데이터플로우 그래프 내에 노드를 생성하는 데에 사용되는 경우, 성분과 데이터세트 객체 간의 관계가 유지된다. 본 예에서는, 데이터플로우 그래프(506a) 내에 노드(508a)를 생성하기 위해 데이터세트 객체(504)가 사용되었다. 마찬가지로, 데이터플로우 그래프(505b-505e) 내에 노드(508b-508e)를 생성하기 위해 동일한 데이터세트 객체(504)가 사용되었다.

[0054] 라인(510a)에 의해 나타낸 바와 같이, 카탈로그(502)는 데이터세트 객체(5004)와 노드(508a) 간의 관계를 유지한다. 라인(510b-510e)에 의해 나타낸 바와 같이, 데이터세트 객체(504)와 노드(508b-508e) 간의 관계가 마찬가지로 유지된다. 일례로, 노드(508a-508e)는 노드를 생성하기 위해 사용된 데이터세트 객체(504)의 아이덴티티를 유지한다.

[0055] 일례로, 데이터플로우 그래프(506a-506e)는 데이터세트 객체(504)를 참조하는 것에 의해 노드(508a-508e)를 구현한다. 이러한 구현 예에서, 데이터세트 객체(504)의 퍼스널러티에 대해 이루어진 변경은 실행되는 데이터플로우 그래프(506a-506e)에 자동으로 반영된다.

[0056] 개발 환경, 예를 들어 도 1의 개발 환경(118)의 사용자는 데이터플로우 그래프 간의 의존성을 분석하기 위해 유지된 관계를 이용할 수 있다. 예를 들어, 사용자는 데이터세트 객체(504) 또는 데이터세트 객체에 의해 식별된 데이터 소스에 대한 변경에 의해 영향을 받을 데이터플로우 그래프(506a-506e)의 각각을 식별할 수 있다. 식별된 데이터플로우 그래프는 이러한 변경에 부합하도록 필요에 따라 변경될 수 있다.

[0057] 도 6은 기존 노드를 카탈로그 내의 데이터세트 객체와 연관시키는 것을 나타낸다. 일례로, 사용자는 카탈로그로부터 데이터세트 객체를 이용하지 않고 데이터 저장장소에 액세스하는 생성된 노드(예를 들어, 데이터 소스, 데이터 싱크, 또는 루프 테이블)를 가질 수 있다. 이것은 적절한 데이터세트 객체가 데이터플로우 그래프를 작성할 때에는 존재하지 않거나, 단순히 개발자가 제공된 것을 사용하는 것을 무시했기 때문에 생길 수 있다. 도 1을 참조하면, 데이터 저장 시스템(116)과 협력하는 개발 환경(118), 및 잠재적으로 실행 환경(104)은 이들 노드를 식별하고, 이들을 적절한 데이터세트 객체와 관련시킨다. 이에 대하여 이하에 설명한다.

[0058] 도 6을 참조하면, 데이터 싱크(612)는 데이터 저장장소(618)에 대한 싱크이며, 데이터 저장장소(618)에 액세스하기 위한 변환 로직을 포함한다. 개발 환경은 카탈로그 내의 데이터세트 객체(620)가 데이터 저장장소(618)에 대한 싱크인 데이터 싱크 퍼스널러티(도시 안 됨)를 포함한다고 판정한다. 화살표(624)로 나타내는 바와 같이, 데이터 싱크(612)는 데이터세트 객체(620)와 연관된다. 일례로, 데이터 싱크(612) 내의 변환 로직은 데이터세트 객체(620)의 데이터 싱크 퍼스널러티 내의 변환 로직으로 교체된다. 다른 예에서, 변환 로직은 교체되지 않는다. 일례로, 데이터 싱크(612)는 데이터세트 객체(620)에 별개의 퍼스널러티로서 추가될 수 있다[예를 들어, 데이터세트 객체(620)가 데이터 싱크 퍼스널러티를 포함하지 않는 경우].

[0059] 다른 예로서, 데이터 싱크(614)는 데이터 저장장소(626)에 대한 싱크이며, 데이터 저장장소(626)를 액세스하기 위한 변환 로직을 포함한다. 개발 환경은 데이터 싱크(614)에 대응하는 카탈로그(622) 내에 데이터세트 객체가 존재하지 않는다고 판정한다. 결과적으로, 개발 환경은 데이터 싱크(614)에 기초해서 데이터 싱크 퍼스널러티(도시 안 됨)를 가진 신규 데이터세트 객체(630)를 작성한다. 신규 데이터세트 객체(630)는 카탈로그(622)에 추가된다. 신규 데이터세트 객체(630)는 데이터 싱크(614)와 연관된다.

[0060] 일례로, 개발 환경은 노드(데이터 소스, 데이터 싱크, 루업 테이블, 또는 다른 성분)를 연관시킬 수 있는데, 해당 노드의 파라미터의 값에 기초하여 연관시킨다. 예를 들어, 데이터세트 객체의 파라미터와 파라미터 값이 노드의 파라미터와 파라미터 값에 일치하는 경우, 노드는 데이터세트 객체와 연관될 수 있다. 일례로, 노드 및 데이터세트 객체는, 파라미터와 파라미터 값이 정확히 일치하지 않는 경우에도 연관될 수 있다. 예를 들어, 일부 파라미터는 요구되는 파라미터로 간주될 수 있다(예를 들어, 파라미터와 파라미터 값 모두가 노드 및 데이터세트 객체가 연관될 수 있기 전에는 일치해야 한다). 다른 파라미터는 존재하지 않아도 되거나 데이터세트 객체와 연관될 노드에 대한 데이터세트 객체의 파라미터 값과 일치하지 않는 임의 선택적 파라미터로 간주될 수 있다.

[0061] 앞서 설명한 바와 같이, 일부 파라미터는 그래프가 실행될 때까지 판정되지 않은 값을 가질 수 있다. 일례로, 시스템은 실행 동안 데이터플로우 그래프의 노드에 대한 파라미터의 값을 추적할 수 있다. 실행 동안 판정된 파라미터와 이들의 값은, 앞서 설명한 바와 같이, 데이터세트 객체의 파라미터와 이들의 값과 비교될 수 있다.

[0062] 도 7을 참조하면, 일례로, 실행 중에 판정된 파라미터 및 이들의 값을, 실행 중에 데이터세트 객체와 연관된 다른 노드의 파라미터 및 이들의 값과 비교한다. 예를 들어, 데이터 카탈로그(702)는 "\$inputfile"의 값을 가진 파라미터(706) "FileName"을 갖는 데이터세트 객체(704)를 포함한다. 여기서, "\$inputfile"은 파라미터(706)가 실행 중에 판정되는 값을 갖는 동적 파라미터라는 것을 나타낸다. 연결 라인(708)으로 나타낸 바와 같이, 데이터세트 객체(704)는 데이터세트 객체(704)의 데이터 소스 퍼스널러티로서 미리 생성된 데이터 소스(710)와 연관된다. 실행 중에, 파라미터(712) "FileName"의 값은 "X:\Input\_20101001.dat"인 것으로 판정된다. 데(704)와 연관되어 있지 않은 데이터 싱크(714)는 "X:\Input\_20101001.dat"의 값을 가진 파라미터(716) "FileName"을 포함한다. 일치하는 파라미터(706, 716)와, 데이터세트 객체(704)와 미리 연관된 다른 파라미터(712)의 값과 일치하는 파라미터(716)의 값에 기초하여, 데이터 싱크(714)는 데이터세트 객체의 데이터 싱크 퍼스널러티로서 데이터세트 객체(704)와 연관된 것으로 판정된다. 일반적으로, 노드(데이터 소스, 데이터 싱크, 또는 루업 테이블)을 데이터세트 객체에 일치시키는 것은, 일치하는 복수의 파라미터와 이들의 값을 필요로 할 것이다. 예를 들어, 노드를 데이터세트 객체에 연관시키는 것은 데이터 저장장소의 위치와 레코드 포맷을 식별하기 위한 일치하는 파라미터를 필요로 할 수 있다.

[0063] 상기 설명한 데이터세트 객체 방법은 컴퓨터에서 실행하기 위한 소프트웨어를 사용해서 구현될 수 있다. 예를 들어, 소프트웨어는 하나 이상의 데이터 저장 시스템(휘발성 및 비휘발성 메모리 및/또는 저장 요소를 포함), 하나 이상의 입력 장치 또는 포트, 및 하나 이상의 출력 장치 또는 포트를 각각 포함하는, 하나 이상의 프로그램된 또는 프로그램 가능한 컴퓨터 시스템(분산형, 클라이언트/서버형, 또는 그리드형 등의 다양한 구조가 될 수 있음)에서 실행하는 하나 이상의 컴퓨터 프로그램에서 절차를 구성한다. 소프트웨어는 계산 그래프의 설계 및 구성에 관련된 다른 서비스를 제공하는 더 큰 프로그램의 하나 이상의 모듈을 형성할 수 있다. 그래프의 노드 및 요소는 컴퓨터로 판독 가능한 매체에 저장된 데이터 구조 또는 데이터 레포지토리 내에 저장된 데이터 모델에 부합하는 다른 조직화된 데이터로서 구현될 수 있다.

[0064] 소프트웨어는 범용 또는 전용의 프로그램 가능한 컴퓨터로 판독 가능한 CD-ROM과 같은 저장 매체에 제공되거나, 네트워크와 같은 통신 매체를 통해 실행 가능한 컴퓨터로 전달(전파 신호로 부호화되는)될 수 있다. 모든 기능은, 전용의 컴퓨터상에서, 또는 코프로세서와 같은 전용의 하드웨어를 사용해서 수행될 수 있다. 소프트웨어는, 해당 소프트웨어에 의해 특정된 연산의 상이한 부분이 여러 컴퓨터에서 수행되는 분산 방식으로 구현되어 도 된다. 이러한 각각의 컴퓨터 프로그램은, 본 명세서에서 설명하는 절차를 수행하도록 저장 매체 또는 디바이스가 컴퓨터 시스템에 의해 판독될 때에 컴퓨터를 구성 및 운영하기 위한, 범용 또는 전용의 프로그램 가능한 컴퓨터에 의해 판독 가능한 저장 매체 또는 디바이스(예를 들어, 고체 메모리 또는 매체, 자기 또는 광학 매체)에 저장되거나 다운로드되도록 하는 것이 바람직하다. 본 발명의 시스템은 컴퓨터 프로그램에 맞게 구성된, 컴퓨터로 판독 가능한 저장 매체로서 구현될 수도 있는데, 이러한 저장 매체는 컴퓨터 시스템으로 하여금, 본 명세서에서 설명한 기능의 수행을 위해 특정되고 미리 정해진 방식으로 동작할 수 있도록 구성된다.

[0065] 본 발명에 대하여 많은 실시예를 설명하였다. 그렇지만, 본 발명의 범위를 벗어남이 없이 다양한 변형이 가능하다는 것을 알 수 있을 것이다. 예를 들어, 상기 설명한 단계들 중 몇몇은 반드시 그 순서대로 수행되지 않아도 되며, 설명된 것과 다른 순서대로 수행되어도 된다.

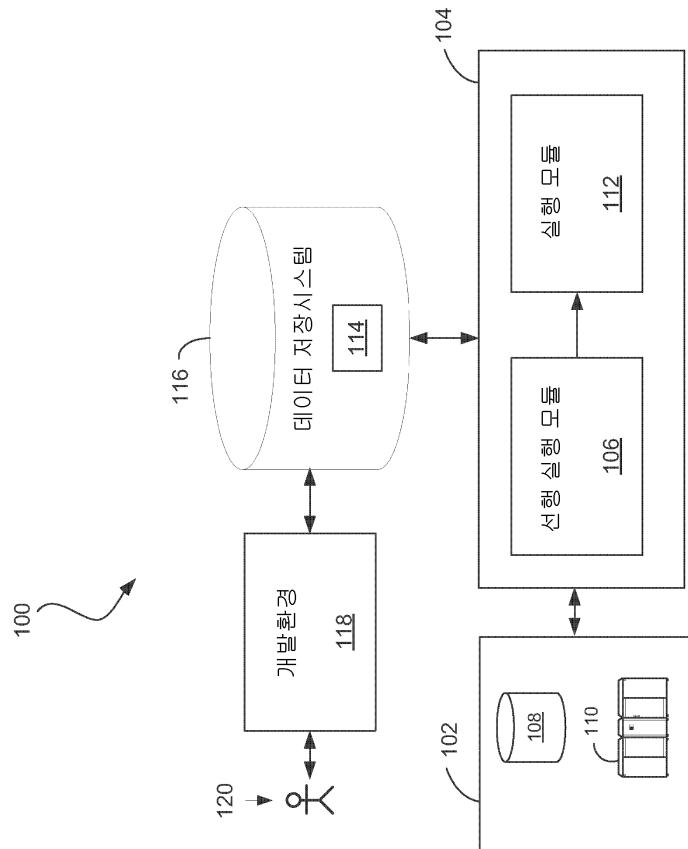
[0066] 이상의 설명은 청구범위에 의해 정해지는 본 발명의 범위를 제한하기 위한 것이 아니라 예시일 뿐이다. 예를 들어, 앞서 설명한 많은 기능적 단계들은 전체적인 과정에 실질적인 영향을 미치지 않으면서, 다른 순서로 수행되어도 된다. 다른 실시예는 이하의 청구범위에 포함된다.

[0067] 본 출원은 2010년 10월 25일에 출원된 미국출원 61/406,438호에 대하여 우선권을 주장하며, 그 내용을 인용해서

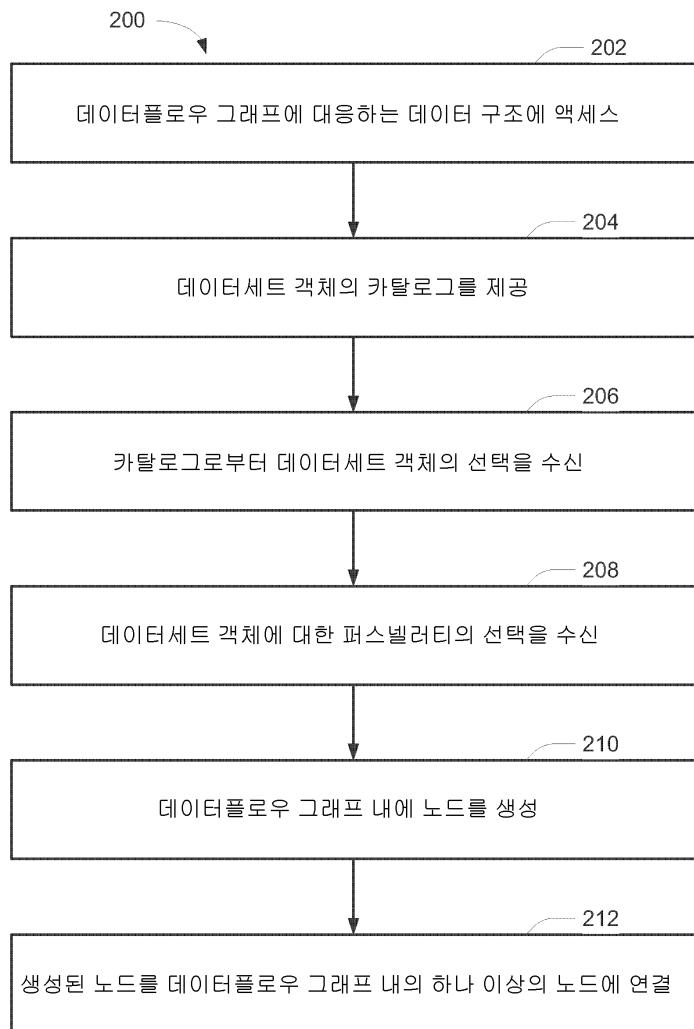
본원에 포함한다.

## 도면

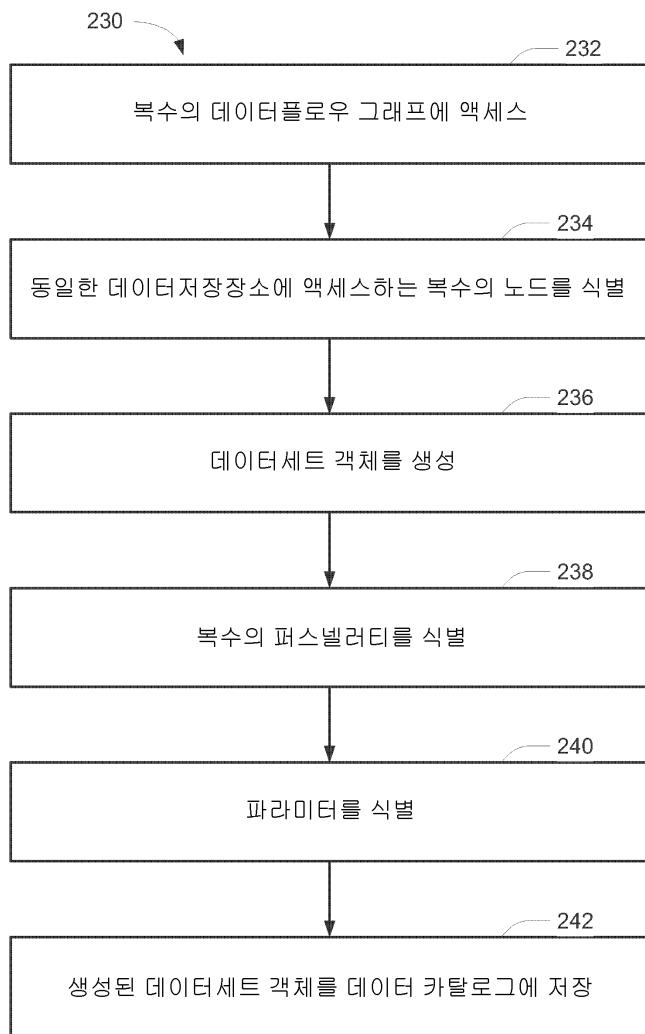
### 도면1



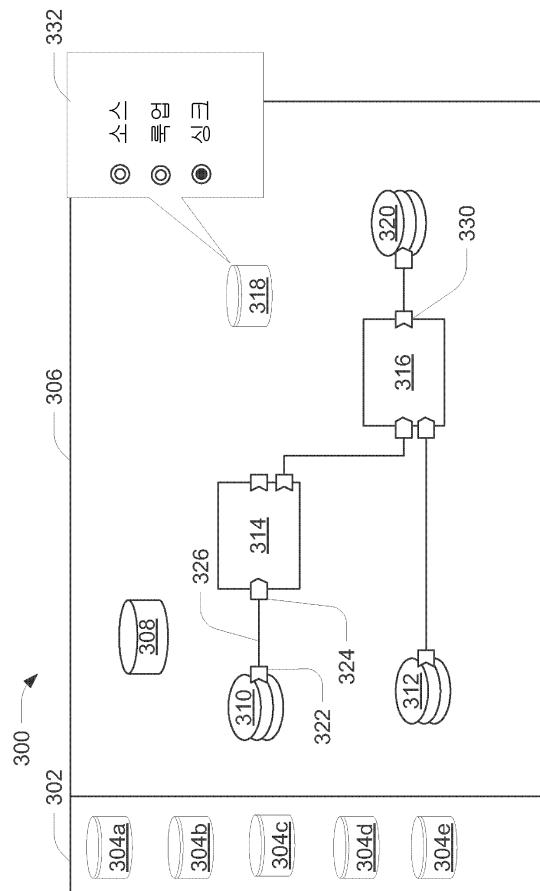
## 도면2a



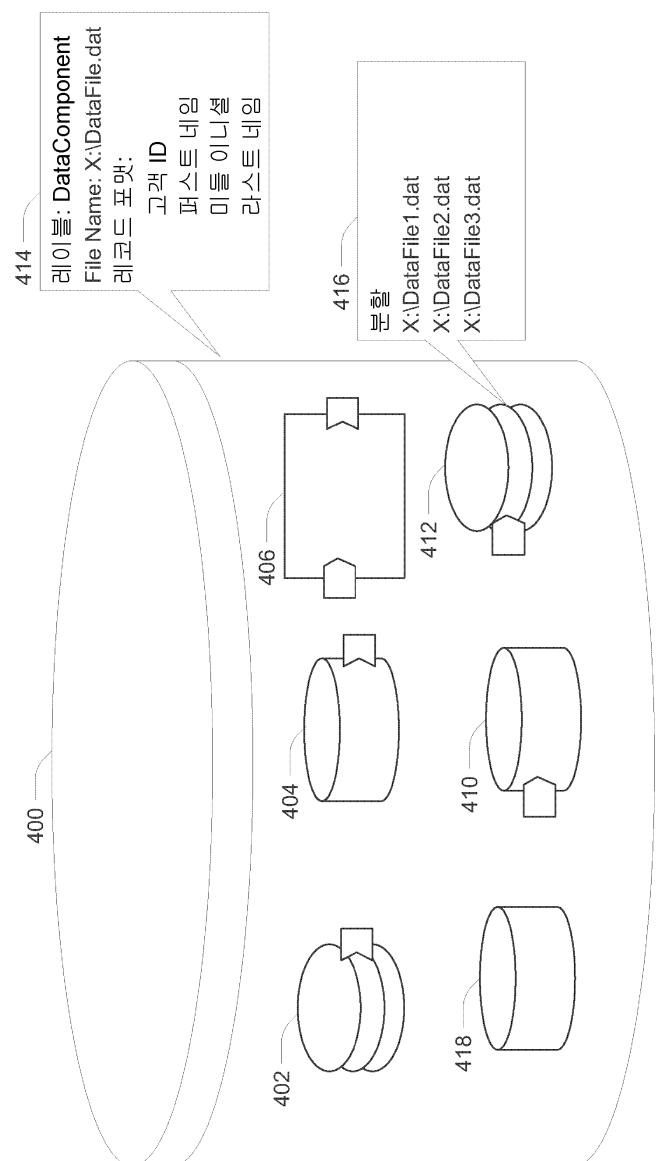
## 도면2b



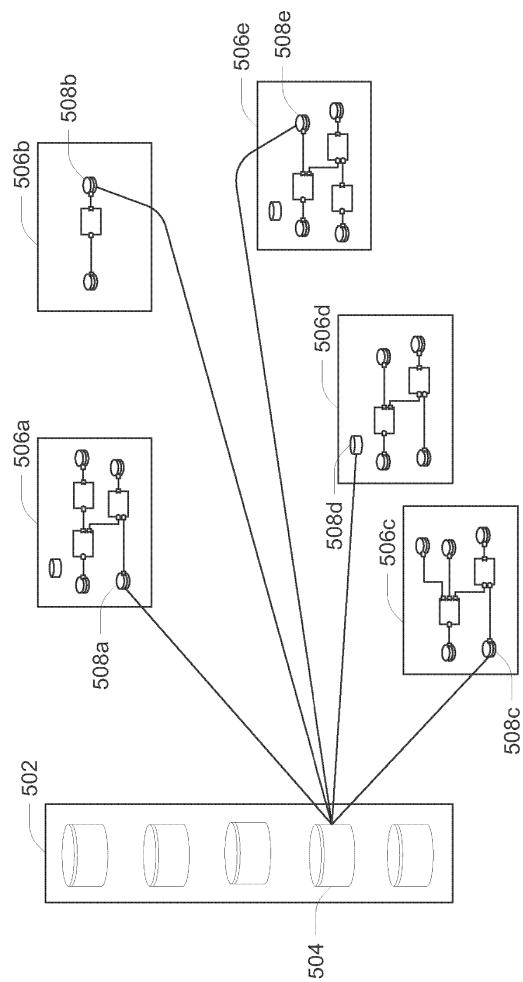
## 도면3



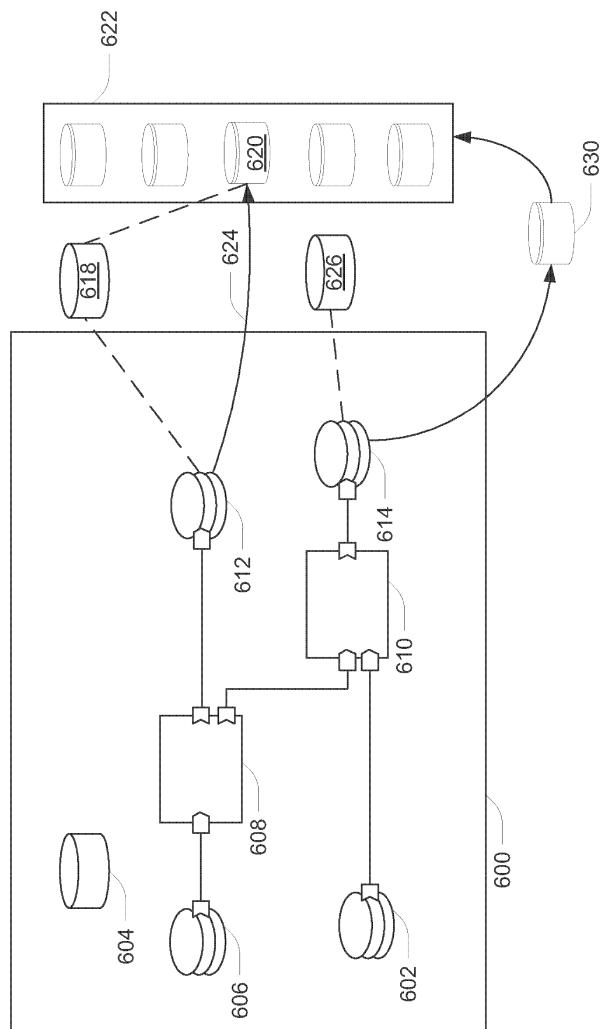
## 도면4



## 도면5



도면6



## 도면7

