



등록특허 10-2220246



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년02월24일
(11) 등록번호 10-2220246
(24) 등록일자 2021년02월19일

- (51) 국제특허분류(Int. Cl.)
G06F 11/36 (2006.01)
(52) CPC특허분류
G06F 11/36 (2013.01)
G06F 11/3688 (2013.01)
(21) 출원번호 10-2016-7019431
(22) 출원일자(국제) 2014년12월17일
심사청구일자 2018년12월03일
(85) 번역문제출일자 2016년07월18일
(65) 공개번호 10-2016-0118235
(43) 공개일자 2016년10월11일
(86) 국제출원번호 PCT/US2014/070747
(87) 국제공개번호 WO 2015/095275
국제공개일자 2015년06월25일
(30) 우선권주장
61/917,727 2013년12월18일 미국(US)

(56) 선행기술조사문헌
KR100898527 B1*

(뒷면에 계속)

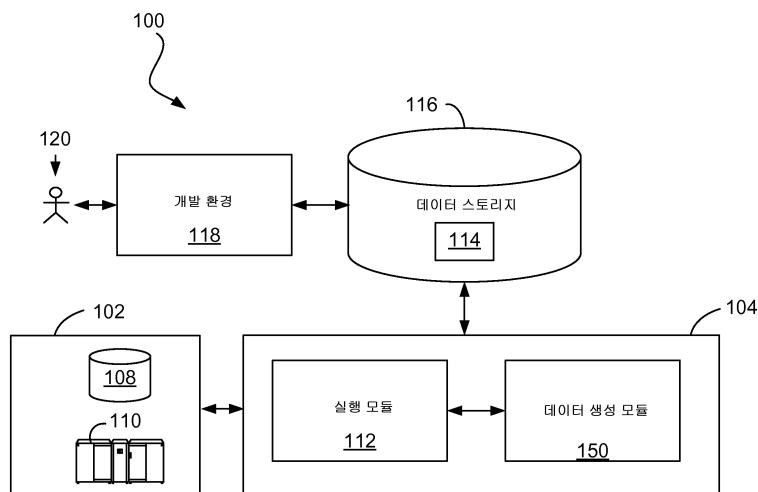
전체 청구항 수 : 총 49 항

심사관 : 홍경아

(54) 발명의 명칭 데이터 생성

(57) 요약

방법은 하나 이상의 레코드를 처리하는 동안 하나 이상의 규칙 각각이 데이터 처리 애플리케이션에 의해 실행되는 횟수를 기반으로 하나 이상의 필드 각각에 대해 컨텐츠 기준을 결정하는 단계; 상기 컨텐츠 기준을 기반으로 상기 각 필드에 대한 컨텐츠를 생성하는 단계; 및 하나 이상의 레코드에서 상기 각 필드를 상기 생성된 컨텐츠로 파풀레이팅하는 단계를 포함한다.

대 표 도 - 도1

(56) 선행기술조사문현

KR1020110027821 A*

KR1020120135218 A*

US20120197887 A1*

US08478706 B2

*는 심사관에 의하여 인용된 문현

명세서

청구범위

청구항 1

데이터 레코드의 처리 동안 처리 규칙이 실행되도록 데이터 레코드에 대한 콘텐츠를 생성하기 위한 방법으로서, 데이터 처리 애플리케이션에 의해 데이터 레코드의 제1세트를 처리하는 동안 데이터 처리 애플리케이션의 처리 규칙이 실행된 제1횟수를 나타내는 실행 정보를 수신하는 단계 - 상기 처리 규칙이 특정 데이터 레코드의 처리 동안 데이터 처리 애플리케이션에 의해 실행되는지 여부는 상기 특정 데이터 레코드의 하나 이상의 데이터 필드 각각의 값에 직접적으로 또는 간접적으로 종속됨 -;

상기 처리 규칙이 실행된 제1횟수가 임계 횟수 미만인 경우에, 하나 이상의 필드 각각에 대해 생성되도록 콘텐츠의 특성을 나타내는 컨텐츠 기준을 결정하는 단계;

상기 컨텐츠 기준에 따라 컨텐츠를 생성하는 단계; 및

상기 컨텐츠 기준에 따라 생성된 컨텐츠로 데이터 레코드의 제2세트의 하나 이상의 데이터 필드를 패플레이팅 (populating)하는 단계 - 상기 생성된 컨텐츠로 패플레이팅된 데이터 레코드의 제2세트가 데이터 처리 애플리케이션에 의해 처리되는 경우에, 상기 생성된 컨텐츠로 패플레이팅된 데이터 레코드의 제2세트는 임계 횟수 이상인 제2횟수로 처리 규칙을 실행하도록 하기에 충분함 -

을 포함하고,

상기 데이터 레코드의 제2세트는 상기 데이터 처리 애플리케이션을 테스트하기 위하여 데이터 처리 애플리케이션에 의해 처리되도록 구성되는, 데이터 레코드에 대한 콘텐츠를 생성하기 위한 방법.

청구항 2

제1항에 있어서,

상기 컨텐츠를 생성하는 단계는 각각의 상기 필드의 포맷 사양을 기반으로 컨텐츠를 생성하는 단계를 포함하는, 데이터 레코드에 대한 콘텐츠를 생성하기 위한 방법.

청구항 3

제2항에 있어서,

상기 포맷 사양은 컨텐츠 타입을 나타내는, 데이터 레코드에 대한 콘텐츠를 생성하기 위한 방법.

청구항 4

제3항에 있어서, 상기 컨텐츠 타입은 스트링, 소수, 정수, 날짜, 또는 일시 중 하나 이상을 포함하는, 데이터 레코드에 대한 콘텐츠를 생성하기 위한 방법.

청구항 5

제2항 내지 제4항 중 어느 한 항에 있어서,

상기 포맷 사양을 사용자 인터페이스를 통해 수신하는 단계를 포함하는, 데이터 레코드에 대한 콘텐츠를 생성하기 위한 방법.

청구항 6

제1항 내지 제4항 중 어느 한 항에 있어서,

상기 데이터 레코드의 제2세트의 하나 이상의 데이터 필드를 패플레이팅하는 단계는

하나 이상의 데이터 필드를 포함하는 새로운 데이터 레코드를 생성하는 단계; 및

하나 이상의 필드를 상기 새로운 데이터 레코드로 패플레이팅하는 단계를 포함하는, 데이터 레코드에 대한 콘텐츠를 생성하기 위한 방법.

청구항 7

제1항 내지 제4항 중 어느 한 항에 있어서,

상기 생성된 컨텐츠로 상기 하나 이상의 데이터 필드를 패플레이팅하는 단계는 데이터 레코드의 제2세트의 하나 이상의 데이터 필드에 대한 기존 컨텐츠를 변경하는 단계를 포함하는, 데이터 레코드에 대한 콘텐츠를 생성하기 위한 방법.

청구항 8

제1항 내지 제4항 중 어느 한 항에 있어서,

상기 하나 이상의 데이터 필드를 패플레이팅하는 단계는 하나 이상의 기존 데이터 레코드에 대한 데이터 필드를 생성하는 단계를 포함하는, 데이터 레코드에 대한 콘텐츠를 생성하기 위한 방법.

청구항 9

제1항 내지 제4항 중 어느 한 항에 있어서,

상기 컨텐츠 기준은 타겟 값, 타겟 값 범위, 타겟 값 분포, 최대치, 또는 최소치 중 하나 이상을 포함하는, 데이터 레코드에 대한 콘텐츠를 생성하기 위한 방법.

청구항 10

제1항 내지 제4항 중 어느 한 항에 있어서,

상기 컨텐츠 기준은 상기 생성된 컨텐츠에 대한 특정 값을 포함하는, 데이터 레코드에 대한 콘텐츠를 생성하기 위한 방법.

청구항 11

제1항 내지 제4항 중 어느 한 항에 있어서,

상기 컨텐츠 기준은 상기 하나 이상의 데이터 필드 중 둘 이상에 포함될 타겟 컨텐츠 조합을 나타내는, 데이터 레코드에 대한 콘텐츠를 생성하기 위한 방법.

청구항 12

제1항 내지 제4항 중 어느 한 항에 있어서,

상기 컨텐츠 기준을 사용자 인터페이스를 통해 수신하는 단계를 포함하는, 데이터 레코드에 대한 콘텐츠를 생성하기 위한 방법.

청구항 13

제1항 내지 제4항 중 어느 한 항에 있어서,

컨텐츠를 생성하는 단계는 생성모드를 기반으로 컨텐츠를 생성하는 단계를 포함하는, 데이터 레코드에 대한 콘텐츠를 생성하기 위한 방법.

청구항 14

제13항에 있어서,

상기 생성 모드는 (i) 다수의 레코드 각각에 대해 하나 이상의 필드에 대한 랜덤 컨텐츠 생성, (ii) 다수의 레코드 각각에 대해 하나 이상의 필드에 대한 고유 컨텐츠 생성, 및 (iii) 사용가능한 컨텐츠 세트로부터 다수의 레코드 각각에 대해 하나 이상의 필드에 대한 컨텐츠 선택 중 하나 이상을 포함하는, 데이터 레코드에 대한 콘텐츠를 생성하기 위한 방법.

청구항 15

제1항 내지 제4항 중 어느 한 항에 있어서,

상기 생성된 컨텐츠를 포함하는 데이터 레코드의 제2세트를 상기 데이터 처리 애플리케이션에 제공하는 단계를 포함하는, 데이터 레코드에 대한 컨텐츠를 생성하기 위한 방법.

청구항 16

제1항 내지 제4항 중 어느 한 항에 있어서,

상기 컨텐츠 기준을 결정하는 단계는 상기 처리 규칙이 상기 임계 횟수 이상의 횟수로 상기 데이터 처리 애플리케이션에 의해 실행되도록 상기 컨텐츠 기준을 결정하는 단계를 포함하는, 데이터 레코드에 대한 컨텐츠를 생성하기 위한 방법.

청구항 17

삭제

청구항 18

데이터 레코드의 처리 동안 처리 규칙이 실행되도록 데이터 레코드에 대한 컨텐츠를 생성하기 위해 명령어들을 저장하는 비일시적인 컴퓨터 판독가능한 매체로서,

상기 명령어들은 컴퓨터 시스템으로 하여금:

데이터 처리 애플리케이션에 의해 데이터 레코드의 제1세트를 처리하는 동안 데이터 처리 애플리케이션의 처리 규칙이 실행된 제1횟수를 나타내는 실행 정보를 수신하고 - 상기 처리 규칙이 특정 데이터 레코드의 처리 동안 데이터 처리 애플리케이션에 의해 실행되는지 여부는 상기 특정 데이터 레코드의 하나 이상의 데이터 필드 각각의 값에 직접적으로 또는 간접적으로 종속됨 -,

상기 처리 규칙이 실행된 제1횟수가 임계 횟수 미만인 경우에, 하나 이상의 필드 각각에 대해 생성되도록 컨텐츠의 특성을 나타내는 컨텐츠 기준을 결정하고,

상기 컨텐츠 기준에 따라 컨텐츠를 생성하고,

상기 컨텐츠 기준에 따라 생성된 컨텐츠로 데이터 레코드의 제2세트의 하나 이상의 데이터 필드를 파풀레이팅 (populating)하도록 하고, 상기 생성된 컨텐츠로 파풀레이팅된 데이터 레코드의 제2세트가 데이터 처리 애플리케이션에 의해 처리되는 경우에, 상기 생성된 컨텐츠로 파풀레이팅된 데이터 레코드의 제2세트는 임계 횟수 이상인 제2횟수로 처리 규칙을 실행하도록 하기에 충분하며,

상기 데이터 레코드의 제2세트는 상기 데이터 처리 애플리케이션을 테스트하기 위하여 데이터 처리 애플리케이션에 의해 처리되도록 구성되는, 컴퓨터 판독가능한 매체.

청구항 19

데이터 레코드의 처리 동안 처리 규칙이 실행되도록 데이터 레코드에 대한 컨텐츠를 생성하기 위한 컴퓨팅 시스템으로서,

메모리에 연결된 적어도 하나의 프로세서를 포함하고, 상기 프로세서와 메모리는:

데이터 처리 애플리케이션에 의해 데이터 레코드의 제1세트를 처리하는 동안 데이터 처리 애플리케이션의 처리 규칙이 실행된 제1횟수를 나타내는 실행 정보를 수신하고 - 상기 처리 규칙이 특정 데이터 레코드의 처리 동안 데이터 처리 애플리케이션에 의해 실행되는지 여부는 상기 특정 데이터 레코드의 하나 이상의 데이터 필드 각각의 값에 직접적으로 또는 간접적으로 종속됨 -,

상기 처리 규칙이 실행된 제1횟수가 임계 횟수 미만인 경우에, 하나 이상의 필드 각각에 대해 생성되도록 컨텐츠의 특성을 나타내는 컨텐츠 기준을 결정하고,

상기 컨텐츠 기준에 따라 컨텐츠를 생성하고,

상기 컨텐츠 기준에 따라 생성된 컨텐츠로 데이터 레코드의 제2세트의 하나 이상의 데이터 필드를 파풀레이팅

(populating)하도록 구성되고, 상기 생성된 컨텐츠로 파풀레이팅된 데이터 레코드의 제2세트가 데이터 처리 애플리케이션에 의해 처리되는 경우에, 상기 생성된 컨텐츠로 파풀레이팅된 데이터 레코드의 제2세트는 임계 횟수 이상인 제2횟수로 처리 규칙을 실행하도록 하기에 충분하며,

상기 데이터 레코드의 제2세트는 상기 데이터 처리 애플리케이션을 테스트하기 위하여 데이터 처리 애플리케이션에 의해 처리되도록 구성되는, 컴퓨팅 시스템.

청구항 20

데이터 레코드의 처리 동안 처리 규칙이 실행되도록 데이터 레코드에 대한 콘텐츠를 생성하기 위한 컴퓨팅 시스템에 있어서,

데이터 처리 애플리케이션에 의해 데이터 레코드의 제1세트를 처리하는 동안 데이터 처리 애플리케이션의 처리 규칙이 실행된 제1횟수를 나타내는 실행 정보를 수신하는 수단 - 상기 처리 규칙이 특정 데이터 레코드의 처리 동안 데이터 처리 애플리케이션에 의해 실행되는지 여부는 상기 특정 데이터 레코드의 하나 이상의 데이터 필드 각각의 값에 직접적으로 또는 간접적으로 종속됨 -;

상기 처리 규칙이 실행된 제1횟수가 임계 횟수 미만인 경우에, 하나 이상의 필드 각각에 대해 생성되도록 콘텐츠의 특성을 나타내는 컨텐츠 기준을 결정하는 수단;

상기 컨텐츠 기준에 따라 컨텐츠를 생성하는 수단; 및

상기 컨텐츠 기준에 따라 생성된 컨텐츠로 데이터 레코드의 제2세트의 하나 이상의 데이터 필드를 파풀레이팅 (populating)하는 수단 - 상기 생성된 컨텐츠로 파풀레이팅된 데이터 레코드의 제2세트가 데이터 처리 애플리케이션에 의해 처리되는 경우에, 상기 생성된 컨텐츠로 파풀레이팅된 데이터 레코드의 제2세트는 임계 횟수 이상인 제2횟수로 처리 규칙을 실행하도록 하기에 충분함 -

을 포함하고,

상기 데이터 레코드의 제2세트는 상기 데이터 처리 애플리케이션을 테스트하기 위하여 데이터 처리 애플리케이션에 의해 처리되도록 구성되는, 컴퓨팅 시스템.

청구항 21

삭제

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

제19항에 있어서,

상기 컨텐츠를 생성하는 것은 각각의 상기 필드의 포맷 사양을 기반으로 컨텐츠를 생성하는 것을 포함하는 컴퓨팅 시스템.

청구항 34

제33항에 있어서,

상기 포맷 사양은 컨텐츠 타입을 나타내는 컴퓨팅 시스템.

청구항 35

제34항에 있어서,

상기 컨텐츠 타입은 스트링, 소수, 정수, 날짜, 또는 일시 중 하나 이상을 포함하는 컴퓨팅 시스템.

청구항 36

제33항 내지 제35항 중 어느 한 항에 있어서,

상기 프로세서는 상기 포맷 사양을 사용자 인터페이스를 통해 수신하도록 구성되는 컴퓨팅 시스템.

청구항 37

제19항에 있어서,

상기 데이터 레코드의 제2세트의 하나 이상의 데이터 필드를 파풀레이팅하는 것은

하나 이상의 데이터 필드를 포함하는 새로운 데이터 레코드를 생성하는 것, 및

하나 이상의 필드를 상기 새로운 데이터 레코드로 파풀레이팅하는 것을 포함하는, 컴퓨팅 시스템.

청구항 38

제19항에 있어서,

상기 생성된 컨텐츠로 상기 하나 이상의 데이터 필드를 파풀레이팅하는 것은 데이터 레코드의 제2세트의 하나 이상의 데이터 필드에 대한 기존 컨텐츠를 변경하는 것을 포함하는, 컴퓨팅 시스템.

청구항 39

제19항에 있어서,

상기 하나 이상의 데이터 필드를 파풀레이팅하는 것은 하나 이상의 기존 데이터 레코드에 대한 데이터 필드를

생성하는 것을 포함하는, 컴퓨팅 시스템.

청구항 40

제19항에 있어서,

상기 컨텐츠 기준은 타겟 값, 타겟 값 범위, 타겟 값 분포, 최대치, 또는 최소치 중 하나 이상을 포함하는 컴퓨팅 시스템.

청구항 41

제19항에 있어서,

상기 컨텐츠 기준은 상기 생성된 컨텐츠에 대한 특정 값들을 포함하는 컴퓨팅 시스템.

청구항 42

제19항에 있어서,

상기 컨텐츠 기준은 상기 하나 이상의 데이터 필드 중 둘 이상에 포함될 타겟 컨텐츠 조합을 나타내는 컴퓨팅 시스템.

청구항 43

제19항에 있어서,

상기 프로세서는 상기 컨텐츠 기준을 사용자 인터페이스를 통해 수신하도록 구성되는 컴퓨팅 시스템.

청구항 44

제19항에 있어서,

컨텐츠를 생성하는 것은 생성모드를 기반으로 컨텐츠를 생성하는 것을 포함하는 컴퓨팅 시스템.

청구항 45

제44항에 있어서,

상기 생성 모드는 (i) 다수의 레코드 각각에 대해 하나 이상의 필드에 대한 랜덤 컨텐츠 생성, (ii) 다수의 레코드 각각에 대해 하나 이상의 필드에 대한 고유 컨텐츠 생성, 및 (iii) 사용가능한 컨텐츠 세트로부터 다수의 레코드 각각에 대해 하나 이상의 필드에 대한 컨텐츠 선택 중 하나 이상을 포함하는 컴퓨팅 시스템.

청구항 46

제19항에 있어서,

상기 프로세서는 상기 생성된 컨텐츠를 포함하는 데이터 레코드의 제2세트를 상기 데이터 처리 애플리케이션에 제공하도록 구성되는 컴퓨팅 시스템.

청구항 47

제19항에 있어서,

상기 컨텐츠 기준을 결정하는 것은 상기 처리 규칙이 상기 임계 횟수 이상의 횟수로 상기 데이터 처리 애플리케이션에 의해 실행되도록 상기 컨텐츠 기준을 결정하는 것을 포함하는, 컴퓨팅 시스템.

청구항 48

삭제

청구항 49

제18항에 있어서,

상기 컨텐츠를 생성하는 것은 각각의 상기 필드의 포맷 사양을 기반으로 컨텐츠를 생성하는 것을 포함하는 컴퓨팅 시스템.

터 판독가능한 매체.

청구항 50

제49항에 있어서,

상기 포맷 사양은 컨텐츠 타입을 나타내는 컴퓨터 판독가능한 매체.

청구항 51

제50항에 있어서,

상기 컨텐츠 타입은 스트링, 소수, 정수, 날짜, 또는 일시 중 하나 이상을 포함하는 컴퓨터 판독가능한 매체.

청구항 52

제49항 내지 제51항 중 어느 한 항에 있어서,

상기 명령어들은 상기 컴퓨터 시스템으로 하여금 상기 포맷 사양을 사용자 인터페이스를 통해 수신하도록 하는 컴퓨터 판독가능한 매체.

청구항 53

제18항에 있어서,

상기 데이터 레코드의 제2세트의 하나 이상의 데이터 필드를 패플레이팅하는 것은,

하나 이상의 데이터 필드를 포함하는 새로운 데이터 레코드를 생성하는 것, 및

하나 이상의 필드를 상기 새로운 데이터 레코드로 패플레이팅하는 것을 포함하는, 컴퓨터 판독가능한 매체.

청구항 54

제18항에 있어서,

상기 생성된 컨텐츠로 상기 하나 이상의 데이터 필드를 패플레이팅하는 것은 데이터 레코드의 제2세트의 하나 이상의 데이터 필드에 대한 기존 컨텐츠를 변경하는 것을 포함하는, 컴퓨터 판독가능한 매체.

청구항 55

제18항에 있어서,

상기 하나 이상의 데이터 필드를 패플레이팅하는 것은 하나 이상의 기존 데이터 레코드에 대한 데이터 필드를 생성하는 것을 포함하는, 컴퓨터 판독가능한 매체.

청구항 56

제18항에 있어서,

상기 컨텐츠 기준은 타겟 값, 타겟 값 범위, 타겟 값 분포, 최대치, 또는 최소치 중 하나 이상을 포함하는 컴퓨터 판독가능한 매체.

청구항 57

제18항에 있어서,

상기 컨텐츠 기준은 상기 생성된 컨텐츠에 대한 특정 값들을 포함하는 컴퓨터 판독가능한 매체.

청구항 58

제18항에 있어서,

상기 컨텐츠 기준은 상기 하나 이상의 데이터 필드 중 둘 이상에 포함될 타겟 컨텐츠 조합을 나타내는 컴퓨터 판독가능한 매체.

청구항 59

제18항에 있어서,

상기 명령어들은 상기 컴퓨터 시스템으로 하여금 상기 컨텐츠 기준을 사용자 인터페이스를 통해 수신하도록 하는 컴퓨터 판독가능한 매체.

청구항 60

제18항에 있어서,

컨텐츠를 생성하는 것은 생성모드를 기반으로 컨텐츠를 생성하는 것을 포함하는 컴퓨터 판독가능한 매체.

청구항 61

제60항에 있어서,

상기 생성 모드는 (i) 다수의 레코드 각각에 대해 하나 이상의 필드에 대한 랜덤 컨텐츠 생성, (ii) 다수의 레코드 각각에 대해 하나 이상의 필드에 대한 고유 컨텐츠 생성, 및 (iii) 사용가능한 컨텐츠 세트로부터 다수의 레코드 각각에 대해 하나 이상의 필드에 대한 컨텐츠 선택 중 하나 이상을 포함하는 컴퓨터 판독가능한 매체.

청구항 62

제18항에 있어서,

상기 명령어들은 상기 컴퓨터 시스템으로 하여금 상기 생성된 컨텐츠를 포함하는 데이터 레코드의 제2세트를 상기 데이터 처리 애플리케이션에 제공하도록 하는 컴퓨터 판독가능한 매체.

청구항 63

제18항에 있어서,

상기 컨텐츠 기준을 결정하는 것은 상기 처리 규칙이 상기 임계 횟수 이상의 횟수로 상기 데이터 처리 애플리케이션에 의해 실행되도록 상기 콘텐츠 기준을 결정하는 것을 포함하는, 컴퓨터 판독가능한 매체.

청구항 64

삭제

청구항 65

삭제

발명의 설명

기술 분야

[0001] 본 출원은 2013년 12월 18일 출원된 미국 출원 일련번호 61/917,727호에 대해 우선권 주장을 하며, 상기 특허출원은 원용에 의해 본 발명에 포함된다.

[0002] 본 발명은 데이터 생성에 관한 것이다.

배경 기술

[0003] 데이터 처리 애플리케이션을 개발하는 동안, 개발자들은 프로덕션 환경 외부에서 근무할 수 있고 프로덕션 데이터에 접근하지 못할 수도 있다. 데이터 처리 애플리케이션이 실제의 데이터를 이용하여 올바르게 실행되도록 하기 위해, 데이터 처리 애플리케이션을 개발 및 테스트하는 동안 현실적인 데이터가 사용될 수 있다.

발명의 내용

[0004] 일반적인 양태에서, 방법은 하나 이상의 레코드를 처리하는 동안 하나 이상의 규칙이 데이터 처리 애플리케이션에 의해 실행되는 횟수를 나타내는 데이터를 수신하는 단계; 각 규칙이 상기 데이터 처리 애플리케이션에 의해

실행된 횟수를 기반으로, 하나 이상의 특정 필드 각각에 대한 컨텐츠 기준을 결정하는 단계; 상기 컨텐츠 기준을 기반으로 상기 특정 필드 각각에 대한 컨텐츠를 생성하는 단계; 및 상기 특정 필드를 상기 생성된 컨텐츠로 패플레이팅(populating)하는 단계를 포함한다.

- [0005] 실시예들은 다음과 같은 하나 이상의 특징들을 포함할 수 있다.
- [0006] 컨텐츠를 생성하는 단계는 각 특정 필드에 대한 포맷 사양을 기반으로 하여 컨텐츠를 생성하는 단계를 포함한다. 상기 포맷 사양은 스트링, 소수, 정수, 날짜, 또는 일시(datetime) 중 하나 이상을 포함하는 컨텐츠 타입을 포함한다. 상기 방법은 사용자 인터페이스를 통해 포맷 사양을 수신하는 단계를 포함한다.
- [0007] 상기 방법은 상기 특정 필드 중 하나 이상을 포함하는 새로운 레코드를 생성하는 단계; 및 상기 새로운 레코드에 대한 특정 필드를 패플레이팅하는 단계를 포함한다.
- [0008] 상기 특정 필드 각각을 상기 생성된 컨텐츠로 패플레이팅하는 단계는 상기 하나 이상의 특정 필드에 대한 기준 컨텐츠를 변경하는 단계를 포함한다.
- [0009] 상기 각 필드를 패플레이팅하는 단계는 하나 이상의 기존 레코드에 대한 특정 필드를 생성하는 단계를 포함한다.
- [0010] 상기 컨텐츠 기준은 타겟 값, 타겟 값들의 범위, 타겟 값들의 분포, 최대값, 또는 최소값 중 하나 이상을 포함한다.
- [0011] 상기 컨텐츠 기준은 상기 생성된 컨텐츠에 대한 특정 값들을 포함한다.
- [0012] 상기 컨텐츠 기준은 상기 하나 이상의 필드 중 둘 이상에 포함될 타겟 컨텐츠들의 조합을 나타낸다.
- [0013] 상기 방법은 사용자 인터페이스를 통해 상기 컨텐츠 기준을 수신하는 단계를 포함한다.
- [0014] 상기 컨텐츠를 생성하는 단계는 생성 모드를 기반으로 컨텐츠를 생성하는 단계를 포함한다.
- [0015] 상기 생성 모드는 (i) 다수의 레코드 각각에 대해 하나 이상의 필드에 대한 컨텐츠 생성, (ii) 다수의 레코드 각각에 대해 하나 이상의 필드에 대한 고유 컨텐츠 생성, 및 (iii) 사용가능한 컨텐츠 세트 중 다수의 레코드 각각에 대해 하나 이상의 필드에 대한 컨텐츠 선택 중 하나 이상을 포함한다.
- [0016] 상기 방법은 상기 생성된 컨텐츠를 포함한 하나 이상의 레코드를 상기 데이터 처리 애플리케이션에 제공하는 단계를 포함한다.
- [0017] 상기 컨텐츠 기준은 특정 규칙이 상기 데이터 처리 애플리케이션에 의해 실행되도록 결정된다.
- [0018] 상기 컨텐츠 기준을 결정하는 단계는 최소 임계 횟수보다 적게 실행된 특정 규칙을 식별하는 단계; 및 상기 특정 규칙이 적어도 상기 최소 임계 횟수만큼 실행되게 하는 상기 컨텐츠 기준을 결정하는 단계를 포함한다.
- [0019] 상기 포맷 사양은 컨텐츠 타입을 포함한다.
- [0020] 일반적인 양태에서, 방법은 사용자 인터페이스를 통해 포맷 사양을 수신하는 단계; 사용자 인터페이스를 통해 컨텐츠 기준을 수신하는 단계; 사용자 인터페이스를 통해 컨텐츠가 생성될 모드 표시를 수신하는 단계; 상기 표시된 모드에 따라 컨텐츠를 생성하고, 상기 생성된 컨텐츠는 상기 포맷 사양과 컨텐츠 기준을 만족하는 단계; 및 레코드에 필드를 생성하는 단계로서, 상기 필드는 상기 생성된 컨텐츠를 포함하는, 필드 생성 단계를 포함한다.
- [0021] 실시예들은 다음과 같은 하나 이상의 특징들을 포함할 수 있다.
- [0022] 상기 모드는 (i) 다수의 레코드 각각에 대한 랜덤 컨텐츠 생성, (ii) 다수의 레코드 각각에 대한 고유 컨텐츠 생성, 및 (iii) 사용가능한 컨텐츠 세트로부터 다수의 레코드 각각에 대한 컨텐츠 선택, 중 하나 이상을 포함한다.
- [0023] 상기 컨텐츠 기준을 수신하는 단계는 상기 사용자 인터페이스에 디스플레이된 하나 이상의 컨텐츠 기준 중 상기 컨텐츠 기준의 선택을 수신하는 단계를 포함한다.
- [0024] 상기 포맷 사양을 수신하는 단계는 상기 사용자 인터페이스에 디스플레이된 하나 이상의 포맷 사양 중 상기 포맷 사양에 대한 선택을 수신하는 단계를 포함한다.
- [0025] 상기 포맷 사양은 스트링, 소수, 정수, 날짜, 또는 일시와 같은 컨텐츠 타입을 나타낸다.

- [0026] 상기 컨텐츠 기준은 상기 컨텐츠에 대한 최대값, 최소값, 및 값들의 범위 중 하나 이상을 나타낸다.
- [0027] 상기 컨텐츠 기준은 상기 컨텐츠에 대한 특정 값들을 포함한다.
- [0028] 상기 컨텐츠 기준은 상기 컨텐츠에 대한 타겟 분포를 포함한다.
- [0029] 상기 컨텐츠 기준은 상기 둘 이상의 필드에 포함될 타겟 컨텐츠 조합을 나타낸다.
- [0030] 상기 컨텐츠 생성 모드에 대한 표시를 수신하는 단계는 상기 사용자 인터페이스에 디스플레이된 하나 이상의 컨텐츠 생성 모드들 중 상기 컨텐츠 생성 모드의 선택을 수신하는 단계를 포함한다.
- [0031] 상기 방법은 새로운 레코드를 생성하는 단계를 포함한다. 레코드에 필드를 생성하는 단계는 상기 새로운 레코드에 필드를 생성하는 단계를 포함한다.
- [0032] 상기 방법은 제2포맷 사양, 제2컨텐츠 기준, 및 제2모드에 대한 표시를 수신하는 단계; 상기 표시된 제2모드에 따라 제2컨텐츠를 생성하고, 상기 생성된 컨텐츠는 상기 제2포맷 사양과 상기 제2컨텐츠 기준을 만족하는 단계; 및 상기 레코드 내 기존 필드에 대한 컨텐츠를 상기 생성된 제2컨텐츠로 대체하는 단계를 포함한다.
- [0033] 일반적인 양태에서, 비일시적인 컴퓨터 기록가능한 매체는 컴퓨터 시스템이, 하나 이상의 레코드들이 처리되는 동안 하나 이상의 규칙들 각각이 데이터 처리 애플리케이션에 의해 실행되는 횟수를 기반으로 하나 이상의 필드들 각각에 대한 컨텐츠 기준을 결정하고; 상기 컨텐츠 기준을 기반으로 상기 필드들 각각에 대한 컨텐츠를 생성하며; 하나 이상의 레코드 내 필드들 각각을 상기 생성된 컨텐츠로 파풀레이팅하도록 하는 명령어들을 저장한다.
- [0034] 일반적인 양태에서, 컴퓨팅 시스템은 하나 이상의 레코드들이 처리되는 동안 하나 이상의 규칙들 각각이 데이터 처리 애플리케이션에 의해 실행되는 횟수를 기반으로 하나 이상의 필드들 각각에 대한 컨텐츠 기준을 결정하고; 상기 컨텐츠 기준을 기반으로 상기 필드들 각각에 대한 컨텐츠를 생성하며; 하나 이상의 레코드의 필드들 각각을 상기 생성된 컨텐츠로 파풀레이팅하도록 구성된 적어도 하나의 프로세서를 포함한다.
- [0035] 일반적인 양태에서, 컴퓨팅 시스템은 하나 이상의 레코드들이 처리되는 동안 하나 이상의 규칙들 각각이 데이터 처리 애플리케이션에 의해 실행되는 횟수를 기반으로 하나 이상의 필드들 각각에 대한 컨텐츠 기준을 결정하는 수단; 상기 컨텐츠 기준을 기반으로 상기 필드들 각각에 대한 컨텐츠를 생성하는 수단; 및 하나 이상의 레코드의 필드들 각각을 상기 생성된 컨텐츠로 파풀레이팅하는 수단을 포함한다.
- [0036] 일반적인 양태에서, 상기 데이터 처리 애플리케이션을 개발하는 동안 데이터 처리 애플리케이션을 테스트하는 방법은 상기 데이터 처리 애플리케이션을 사용하는 제1데이터 레코드들을 처리하는 단계를 포함하고, 상기 데이터 처리 애플리케이션은 하나 이상의 규칙들을 구현한다. 제1데이터 레코드들 각각은 다수의 필드들을 갖는다. 상기 방법은 상기 제1데이터 레코드를 처리하는 동안 상기 데이터 처리 애플리케이션의 각 규칙이 실행되는 횟수를 그 규칙에 대한 타겟 값과 비교하는 단계를 포함한다. 상기 방법은 비교 결과를 기반으로, 상기 데이터 처리 애플리케이션을 사용하여 처리될 제2데이터 레코드의 데이터에 대한 타겟 특성을 나타내는 컨텐츠 기준을 결정하는 단계를 포함한다. 상기 방법은 상기 제1데이터 레코드 포맷 및 상기 컨텐츠 기준에 따라 상기 제2데이터 레코드의 하나 이상의 필드에 대한 컨텐츠를 생성하는 단계; 상기 데이터 처리 애플리케이션을 사용하여 상기 제2데이터 레코드를 처리하는 단계; 상기 제2데이터 레코드를 처리하는 동안 상기 데이터 처리 애플리케이션의 각 규칙이 실행되는 횟수를 그 규칙에 대한 타겟 값과 비교하고 각 규칙에 대한 타겟 값이 만족되는지를 결정하는 단계; 및 상기 제2데이터 레코드를 사용하여 상기 데이터 처리 애플리케이션을 테스트하는 단계를 포함한다.
- [0037] 실시예들은 다음과 같은 하나 이상의 특징들을 포함할 수 있다.
- [0038] 상기 방법은 클라이언트 장치에서 상기 제1데이터 레코드를 수신하는 단계를 포함하고, 상기 데이터 처리 애플리케이션은 상기 클라이언트 장치에서 실행된다. 상기 클라이언트 장치는 컴퓨팅 장치를 포함한다.
- [0039] 상기 방법은 제거 서버(remove server)로부터 상기 제1데이터 레코드를 수신하는 단계를 포함한다. 상기 방법은 LAN 또는 WAN 연결을 통해 서버에 통신가능하게 연결되는 클라이언트 장치에서 제1데이터 레코드를 수신하는 단계를 포함한다.
- [0040] 상기 방법은 상기 데이터 처리 애플리케이션이 상기 제2데이터 레코드를 처리할 때 상기 제2데이터 레코드의 데이터가 각 규칙에 대한 상기 타겟 값이 만족되도록 상기 컨텐츠 기준을 결정하는 단계를 포함한다.
- [0041] 상기 제2데이터 레코드의 하나 이상의 필드들에 대한 컨텐츠를 생성하는 단계는 상기 제1데이터 레코드의 하나

이상의 필드들 내 컨텐츠를 변경하는 단계를 포함하고, 상기 변경된 제1데이터 레코드는 상기 제2데이터 레코드이다.

- [0042] 상기 제2데이터 레코드의 하나 이상의 필드들에 대한 컨텐츠를 생성하는 단계를 새로운 데이터 레코드를 생성하는 단계를 포함한다.
- [0043] 상기 방법은 상기 제2데이터 레코드를 스토리지 장치에 저장하는 단계를 포함한다.
- [0044] 상기 방법은 상기 테스트 결과를 기반으로 상기 데이터 처리 애플리케이션을 변경하는 단계를 포함한다.
- [0045] 상기 데이터 처리 애플리케이션을 테스트하는 단계는 상기 데이터 처리 애플리케이션을 사용하여 상기 제2데이터 레코드를 처리한 결과를 기반으로 상기 데이터 처리 애플리케이션을 변경하는 단계를 포함한다.
- [0046] 상기 방법은 상기 제2데이터 레코드 내 데이터가 프로덕션 환경에서 상기 데이터 처리 애플리케이션에 의해 처리될 것으로 기대되는 데이터와 실질적으로 유사하도록 상기 컨텐츠 기준을 결정하는 단계를 포함한다.
- [0047] 양태들은 하나 이상의 다음과 같은 장점을 포함할 수 있다. 데이터는 임의의 포맷 사양과 컨텐츠 기준에 따라, 예를 들어 간단한 사용자 인터페이스를 통해 생성될 수 있다. 데이터 생성은, 예를 들어 기존 데이터 세트가 충분한 레코드를 갖지 않거나 원하는 특성의 레코드를 갖지 않는다면, 기존 데이터의 보충에 사용될 수 있다. 데이터 생성은 또한, 예를 들어 애플리케이션이 사용 가능하지 않은 데이터를 요청한 경우, 스크래치(scratch)로부터의 데이터 생성에 사용될 수 있다.
- [0048] 일부 예에서, 현실적인 데이터 처리 애플리케이션의 테스트를 위해 생성될 수 있다. 예를 들어, 상기 데이터 처리 애플리케이션의 모든 규칙들이 적어도 한번 실행되게 하는, 예를 들어, 상기 데이터 처리 애플리케이션에서 모든 가능한 동작을 트리거하는 테스트 데이터 세트가 생성될 수 있다. 그러한 데이터가 존재하지 않는다면, 데이터가 생성될 수 있다. 예를 들어, 완전한 데이터 세트가 생성된다면, 생성된 데이터 세트는 상기 데이터 처리 애플리케이션에 의해 처리되어 상기 데이터 처리 애플리케이션 실행을 테스트할 수 있다. 상기 생성된 데이터 세트는 상기 데이터 처리 애플리케이션의 모든 가능한 동작이 실행되게 하므로, 상기 데이터 처리 애플리케이션은 완전히 테스트될 수 있다. 상기 데이터 처리 애플리케이션이 원하는 것처럼 수행되지 않는다면 (예를 들어, 상기 데이터 처리 애플리케이션의 동작이 원하는 것처럼 수행되지 않는다면), 상기 데이터 처리 애플리케이션은 적절하게 변경되어 만족할만한 결과를 얻을 때까지 테스트될 수 있다.
- [0049] 본 발명의 다른 특징과 장점들은 다음의 상세한 설명과 청구범위로부터 명백해질 것이다.

도면의 간단한 설명

- [0050] 도 1은 데이터 생성 시스템에 대한 블록도이다.
- 도 2는 데이터 생성 모듈에 대한 블록도이다.
- 도 3A는 예시적인 소스 파일의 일부이다.
- 도 3B 내지 3D는 예시적인 타겟 파일의 일부이다.
- 도 4는 소스 윈도우이다.
- 도 5A는 소스 윈도우이다.
- 도 5B는 데이터 생성 윈도우이다.
- 도 6은 타겟 윈도우이다.
- 도 7은 보조 소스 윈도우이다.
- 도 8은 레코드 포맷 윈도우이다.
- 도 9는 스트링 특성 윈도이다.
- 도 10은 소수 특성 윈도우이다.
- 도 11은 정수 특성 윈도우이다.
- 도 12는 날짜 특성 윈도우이다.

도 13은 일시(datetime) 특성 원도우이다.

도 14A 내지 14C는 규칙 생성을 위한 원도우이다.

도 15A 내지 15C는 규칙 생성을 위한 원도우이다.

도 16은 규칙 생성을 위한 원도우이다.

도 17A 및 17B는 주제 영역 데이터세트에 대한 원도우이다.

도 18은 흐름도이다.

도 19는 데이터 처리 애플리케이션의 블록도이다.

도 20은 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0051]

본 명세서에서는 다양한 구성가능한 옵션에 따른 데이터 생성에 대한 일반적인 접근을 설명한다. 예를 들어, 생성된 데이터의 데이터 타입은 사용자에 의해 특정될 수 있다. 데이터 타입은, 예를 들어, 스트링, 소수, 정수, 날짜, 및 일시(datetime)를 포함할 수 있다. 생성된 데이터는, 예를 들어, 생성된 소수 또는 정수 데이터에 대한 허용치 범위, 생성된 스트링 데이터에 대한 평균 스트링 길이, 생성된 데이터에 사용될 수 있는 값 또는 문자 세트, 및 다른 특성과 같이 생성된 데이터의 특성에 제한을 가하는 하나 이상의 구성가능한 컨텐츠 기준을 만족 할 수 있다. 데이터 생성은 기존 소스 레코드의 하나 이상의 필드 내 값을 변경하는 단계, 레코드에 새로운 필드를 생성하고 패플레이팅하여 소스 레코드를 추가하는 단계, 또는 전체적으로 새로운 레코드를 생성하는 단계를 포함할 수 있다. 일부 예에서, 구성가능한 옵션들은 사용자 인터페이스를 통해 사용자에 의해 특정될 수 있다.

[0052]

일부 예들에서, 레코드 세트는 데이터 처리 애플리케이션에 의해 처리될 수 있다. 데이터 처리 애플리케이션은 입력 레코드에 종속하는 것과 같이 실행이 하나 이상의 변수 값에 좌우되는 규칙을 구현할 수 있다. 일부 경우에서, 데이터 처리 애플리케이션이 레코드 세트를 처리할 때 그 레코드 세트에 대한 데이터가 생성되어 원하는 실행 결과를 얻게 한다. 예를 들어, 데이터 처리 애플리케이션의 특정 규칙이 수회 실행되게 하거나 다른 원하는 실행 결과를 내도록 하는 레코드 세트에 대한 데이터가 생성될 수 있다. 애플리케이션 실행 결과를 기반으로, 어떤 데이터가 생성되게 할 것인지, 예를 들어 원하는 실행 결과가 더 잘 얻어지도록 결정이 이뤄질 수 있다. 예를 들어, 데이터 처리 애플리케이션에 의해 처리되기를 원하는 프로덕션 데이터와 실질적으로 유사한 현실적인 데이터 세트가 생성될 수 있다. 생성된 현실적인 데이터 세트는 데이터 처리 애플리케이션의 테스트, 예를 들어, 데이터 처리 애플리케이션의 모든 양태가 정확하게 동작하도록 보장하는데 사용될 수 있다. 테스트 결과를 기반으로, 데이터 처리 애플리케이션은, 적절하다면 프로덕션 환경에서 사용되기 전에 변경될 수 있다.

[0053]

도 1은 데이터 생성 기술이 사용될 수 있는 예시적인 데이터 처리 시스템 (100)을 보여준다. 시스템(100)은 스토리지 장치와 같은 하나 이상의 데이터 소스들 또는 온라인 데이터 스트림으로의 연결을 포함할 수 있는 소스(102)를 포함하고, 각각은 다양한 포맷(예를 들어, 데이터베이스 테이블, 스프레드시트 파일, 일반 텍스트(flat text) 파일 또는 메인프레임에 의해 사용되는 고유 포맷(native format))의 데이터를 저장 또는 공급할 수 있다. 실행 환경(104)은 실행 모듈(112)을 포함한다. 실행 환경(104)에서는, 예를 들어, UNIX 운영체제 버전과 같은 적절한 운영체제의 제어하에 있는 하나 이상의 범용 컴퓨터가 호스팅할 수 있다. 예를 들어, 실행 환경(104)은 다수의 CPU 또는 프로세서 코어들을 사용하는 컴퓨터 시스템 구성을 포함하는 다중 노드 병렬 컴퓨팅 환경을 포함할 수 있고, 다수의 CPU 또는 프로세서 코어들은 근거리 (예를 들어, 대칭형 다중처리(SMP) 컴퓨터와 같은 멀티프로세서 시스템) 또는 근거리 분산 (예를 들어, 클러스터 또는 대규모 병렬 처리(MPP) 시스템으로 연결된 다수의 프로세서들), 또는 원격, 또는 원격 분산 (예를 들어, LAN 및/또는 WAN을 통해 연결된 다수의 프로세서들), 또는 이들의 결합된 형태이다.

[0054]

소스(102)는 실행 환경(104)의 근거리에 있는, 예를 들어, 실행 환경(104)을 호스팅하는 컴퓨터에 연결된 스토리지 매체(예를 들어, 하드 드라이브(108))일 수 있거나, 실행 환경(104)의 원격에 있는, 예를 들어, 실행 환경(104)을 호스팅하는 컴퓨터와 (예를 들어, 클라우드 컴퓨팅 인프라에 의해 제공되는) 원격연결을 통해 통신하는 원격 시스템(예를 들어, 메인 프레임(110))에서 호스팅될 수 있다.

[0055]

실행 모듈(112)은 소스(102)로부터 데이터를 판독하거나 판독된 데이터를 처리한다. 실행 모듈(112)은 데이터에

대해 계산과 같은 프로세스를 실행하는 하나 이상의 데이터 처리 애플리케이션을 포함한다. 출력 데이터(114)는 소스(102)에 다시 저장되거나 실행 환경(104)에 접근가능한 데이터 스토리지 시스템(116)에 저장되거나 달리 사용될 수 있다.

[0056]

데이터 스토리지 시스템(116)은 또한 개발자(120)가 데이터 처리 애플리케이션, 소스(102), 또는 둘 다를 구성할 수 있는 개발 환경(118)에 접근할 수 있다. 개발환경(118)은, 일부 구현에서, 애플리케이션들을 (데이터 처리 컴포넌트들 또는 데이터 세트를 나타내는) 정점들 사이에서 (작업 요소, 즉, 데이터 플로우를 나타내는) 방향성 있는 링크들로 연결된 정점들을 포함하는 데이터 플로우 그래프로 개발하는 시스템이다. 예를 들어, 그러한 환경은 본 명세서에서 참조로 포함된 “그래프 기반 애플리케이션들에 대한 파라미터 관리(Managing Parameters for Graph-Based Applications)”라는 제목의 미국 공개번호 2007/0011668호에 상세하게 기술되어 있다. 그러한 그래프 기반 계산을 실행하는 시스템은 “그래프로 표현된 계산 실행”이라는 제목의 미국 특허번호 5,966,072호에 기술되어 있다. 시스템(100)과 연계되어 만들어진 데이터 플로우 그래프는 프로세스간 정보 이동을 위해 및 프로세스들에 대한 실행 순서를 규정하기 위해 그래프 컴포넌트로 표현된 개별 프로세스로/로부터 정보를 입력하거나/얻는 방법을 제공한다. 시스템(100)은 임의의 사용가능한 방법들 (예를 들어, 그래프 링크에 따른 통신 경로는 TCP/IP 또는 UNIX 도메인 소켓(domain socket)을 사용하거나 공유 메모리를 사용하여 프로세스간 데이터를 통과시킨다)로부터 프로세스간 통신 방법을 선택하는 알고리즘을 포함한다.

[0057]

데이터 생성 모듈(150)은 실행 모듈에 의한 처리를 위해 데이터를 생성할 수 있다. 데이터 생성 모듈(150)은 소스(102)로부터의 데이터와 같은 기존 데이터를 변경 또는 추가할 수 있다. 예를 들어, 데이터 생성 모듈(150)은 소스 데이터 레코드에 하나 이상의 필드에 대한 값을 변경하거나 소스 데이터 레코드에 하나 이상의 새로운 필드들을 파플레이팅할 수 있다. 데이터 생성 모듈(150)은 또한 소스 데이터 레코드의 포맷을 기반으로 또는 특정된 포맷을 기반으로 전체적으로 새로운 데이터 레코드를 생성할 수 있다.

[0058]

도 2를 참조하면, 데이터 생성 모듈(150)은 타겟(24)에 저장될 데이터를 생성하는 데이터 생성 엔진(20)을 포함한다. 일부 예에서, 데이터 생성 엔진(20)은 데이터 소스(22) 내 데이터를 기반으로 데이터를 생성할 수 있다. 일부 예에서, 데이터 생성 엔진(20)은 스크래치로부터, 즉, 소스 데이터가 아닌 데이터를 생성할 수 있다. 데이터 소스(22)는 파일, 데이터베이스, 또는 파라미터 세트이거나 다른 데이터 소스일 수 있다. 데이터 소스(22)는 하나 이상의 레코드를 포함하며, 각각은 하나 이상의 데이터 필드를 가질 수 있다. 예를 들어, 예시적인 데이터 소스는 소매점 고객들에 대한 고객 레코드들을 저장하는 (우리가 여기서 ‘고객 데이터베이스’로 부르는) 데이터 베이스일 수 있다. 데이터베이스의 각 레코드는 개별 고객을 나타낸다. 각 레코드는 다수의 필드를 가질 수 있다. 데이터 소스(22)는 필드 수, 각 필드 내 데이터 타입 (예를 들어, 스트링, 정수, 소수, 날짜, 또는 일시), 및 각 필드 내 데이터 특성 (예를 들어, 값의 허용 범위, 최대 허용치, 허용 문자 리스트, 또는 다른 제한들)과 같은 레코드 포맷을 특정하는 레코드 포맷을 가질 수 있다. 일부 예에서, 데이터는 스크래치로부터 생성되어 데이터 소스가 제공되지 않는다.

[0059]

예를 들어, 도 3을 참조하면, 예시적인 소스 파일(200)의 일부는 6개의 레코드들(202)을 포함하고, 각 레코드는 다음의 필드들: 고객 ID (cust_ID) 필드 (204), 이름 (name) 필드 (206), 고객 나이 (cust_age) 필드(208), 및 상태 (state) 필드(210)를 갖는다.

[0060]

도 2를 참조하면, 일부 예에서 데이터 소스(22)는 사용자(26)에 의해 식별될 수 있다. 예를 들어, 데이터 생성 모듈(150)의 사용자 인터페이스 엔진(28)은 사용자 인터페이스(30)가, 예를 들어, 개인용 컴퓨터 또는 휴대 컴퓨팅 장치와 같은 컴퓨팅 장치(32)의 디스플레이 인터페이스를 통해 사용자(26)에게 디스플레이되게 할 수 있다. 사용자(26)는 사용자 인터페이스(30)를 통해 데이터 소스(22)를 식별한다. 생성된 데이터는 파일, 데이터베이스, 파라미터 세트 또는 다른 데이터 구조와 같은 타겟(24)에 저장될 수 있다. 일부 예에서, 타겟 파일(24)은 사용자(26)에 의해, 예를 들어, 사용자 인터페이스(30)에 의해 식별될 수 있다. 일부 예에서, 타겟의 레코드 포맷은 소스의 레코드 포맷과 일치할 수 있다. 즉, 타겟에 대해 특정된 레코드 포맷은 소스에서 기존 레코드의 레코드 포맷과 동일할 수 있다. 일부 예에서, 타겟의 레코드 포맷은 소스의 레코드 포맷과 다르고, 소스의 일부 또는 모든 레코드 포맷은 타겟으로 매핑될 수 있다. 예를 들어, 소스의 하나 이상의 포맷은 타겟의 레코드 포맷에 포함될 수 있다.

[0061]

데이터 생성 모듈(150)은, 예를 들어, 데이터베이스(36), 파일, 또는 다른 데이터 구조에 저장된 구성 데이터(34)를 기반으로 타겟(24)에 대한 데이터를 생성할 수 있다. 구성 데이터(34)는 타겟(24)에 사용될 데이터 생성 접근, 타겟(24)에 대한 컨텐츠 생성 모드, 타겟(24)에 대해 생성될 데이터의 데이터 타입, 타겟에 대해 생성될 데이터에 대한 컨텐츠 기준, 및 타겟(24)에 대해 생성될 데이터에 대한 다른 구성 정보를 특정할 수 있다. 이

구성 데이터(34)는 이하에서 더 상세하게 논의된다. 일부 예에서, 타겟(240에 대한 데이터 생성에 사용되는 일부 또는 모든 구성 데이터(34)는 사용자(26)에 의해, 예를 들어, 사용자 인터페이스(30)를 통해 특정될 수 있다. 일부 실시예에서, 일부 또는 모든 구성 데이터(34)는 구성 엔진(39)에 의해, 예를 들어, 데이터 소스(22)의 분석을 기반으로, 또는 타겟의 원하는 특성에 대한 정보를 기반으로 결정될 수 있다.

[0062] 일부 예에서, 데이터 생성 엔진(20)은 구성 데이터(34)에 따라 데이터 소스(22)에서 기준 소스 레코드의 하나 이상의 필드들에 대한 값들을 변경하여 타겟(24)에 대한 데이터를 생성한다. 변경된 레코드는 타겟(24)에 저장될 수 있다. 일부 예에서, 주어진 필드에 대한 모든 값들은 변경될 수 있다. 예를 들어, 모든 레코드들에 대해 주어진 필드 값들의 분포가 구성 데이터(34)로 나타내진 것처럼 타겟 분포와 일치하도록 각 레코드의 주어진 필드에 값이 할당될 수 있다. 예를 들어, 분포는 사용자(26)에 의해 특정되거나 구성 엔진(39)에 의해 결정될 수 있다. 일부 예에서, 주어진 필드에 대한 모든 값보다 더 적은 값들, 예를 들어, 구성 데이터(34)로 나타내진 것처럼 특정된 기준을 만족하지 못하는 값들만 변경될 수 있다. 예를 들어, 주어진 필드에 대한 허용 값들의 특정 범위 밖에 있는 값들이 변경될 수 있다.

[0063] 예를 들어, 도 3B를 참조하면, 예시적인 타겟 파일(220)의 일부는 소스 파일(200)을 기반으로 생성된 레코드(222)를 포함한다. 이 예에서 구성 데이터(34, 도 2)는 고객의 나이(cust_age) 필드 (208)는 최대값 40을 가질 수 있다. 구성 데이터(34)로 특정된 기준을 만족시키기 위해, 소스 파일(200)에서 cust_age 필드(208)에 대한 값은 cust_age 필드에 대한 최대값이 40이 되도록 데이터 생성 엔진(20)에 의해 변경되었다. 즉, (소스 파일(200) 내) 특정 레코드(222a)의 cust_age 필드(208) 값이 45에서 (타겟 파일(220) 내) 25로 변경되었다. 타겟 파일에서 생성된 레코드(222)의 다른 필드 값은 소스 레코드(202)의 값들에 대응한다.

[0064] 도 2를 참조하면, 일부 예에서, 데이터 생성 엔진(20)은 구성 데이터(34)에 따라 데이터 소스(22)의 기준 소스 레코드에 하나 이상의 새로운 필드를 추가하여 데이터를 생성할 수 있다. 늘어난 레코드 (즉, 하나 이상의 새로운 필드를 갖는 레코드)는 타겟(24)에 저장될 수 있다. 구성 데이터(34)는 새로운 필드의 수, 새로운 필드에 대한 데이터 타입과 값들, 및 새로운 필드의 특성을 결정하는 명령어를 제공할 수 있다. 일부 예에서, 구성 데이터(34)는 새로운 필드에 대한 값들이 데이터 소스(22)의 소스 데이터에서 기준 필드에 대한 데이터를 기반으로 생성되도록 특정할 수 있다. 일부 예에서, 구성 데이터(34)는 새로운 필드에 대한 값들이 기준 소스 데이터를 기반으로 하지 않고, 예를 들어, 특정한 특성 (예를 들어, 구성 데이터(34)에 의해 특정된 특성)에 따라 스크래치로부터 생성되도록 특정할 수 있다.

[0065] 예를 들어, 도 3C를 참조하면, 예시적인 타겟 파일(230)의 일부는 소스 파일(200)을 기반으로 생성된 레코드(232)를 포함한다. 이 예에서 구성 데이터(34, 도 2)는 새로운 고객의 나이(cust_age) 필드(208)에 대한 값을 5 개의 사용 가능한 나이 범위(age_range) 버킷(bucket)들 중 하나로 구분하고 생성되는 값들을 갖는 새로운 나이 범위(age_range) 필드(234) 생성을 특정한다. 구성 데이터(34)는 특정 최대 및 최소값에 종속되지만 소스 파일(200) 내 데이터와는 아무 관계없는 값들로 새로운 고객 수입(cust_income) 필드(236) 생성을 더 특정한다.

[0066] 도 2를 참조하면, 일부 예에서, 데이터 생성 엔진은 구성 데이터(34)에 따라 데이터 소스(22)의 기준 소스 레코드에 하나 이상의 새로운 필드를 추가하여 데이터를 생성할 수 있다. 늘어난 레코드 (즉, 기준 소스 레코드 및 새로운 레코드 모두)는 타겟(24)에 저장될 수 있다. 새로운 레코드는 소스 레코드와 동일한 레코드 포맷을 가질 수 있다. 구성 데이터(34)는 새로운 레코드의 수, 새로운 레코드 필드에 대한 값들, 및 새로운 필드의 특성을 결정하는 명령어를 제공할 수 있다. 일부 예에서, 구성 데이터(34)는 새로운 레코드에서 하나 이상의 필드들에 대한 값들이, 예를 들어 특정된 특성에 따르지만, 데이터 소스(22)의 기준 소스 데이터를 기반으로 하지 않는, 스크래치로부터 생성되는 것을 특정할 수 있다. 일부 예에서, 구성 데이터(34)는 새로운 레코드에서 하나 이상의 필드들에 대한 값들이 구성 데이터(34)에 의해 특정된 프로파일을 만족하도록 생성되는 것을 특정할 수 있다. 예를 들어, 프로파일은 모든 레코드에서 특정한 레코드에 대한 값들이 평균 또는 분포와 같은 특성을 만족하도록 특정할 수 있다. 예를 들어, 고객 데이터베이스 소스에서, 새로운 레코드는 모든 레코드에 걸쳐 고객 나이(cust_age) 필드에 대한 값들이 구성 데이터(34)에 의해 특정된 특정 분포를 만족하도록 생성될 수 있다.

[0067] 예를 들어, 도 3D를 참조하면, 예시적인 타겟 파일(240)의 일부는 소스 파일(200)을 기반으로 생성된 레코드(202, 242)를 포함한다. 이 예에서, 구성 데이터(34, 도 2)는 모든 레코드에 걸쳐 고객의 나이(cust_age) 필드(208)에 대한 값이 40과 45 사이에서 평균값을 갖고 필드별 최대값 60에 종속되도록 4개의 새로운 레코드(242) 생성을 특정한다. 이 예에서, 구성 데이터(34)는 각 필드에 대한 값들이, 예를 들어 필드에 대한 특정된 특성에 따라 랜덤하게 생성되도록 특정한다. 원본 소스 레코드(202)는 또한 타겟 파일(240)에 포함된다.

[0068] 일부 예에서, 데이터 생성 엔진(20)은 구성 데이터(34)에 따라 데이터 생성에 대한 한 번 이상의 접근을 적용할

수 있다. 예를 들어, 데이터 생성 엔진(20)은 다음 접근들의 조합을 적용할 수 있다: 하나 이상의 필드값들이 변경될 수 있고, 소스 레코드들에 하나 이상의 새로운 필드들이 추가될 수 있고, 소스 레코드에 하나 이상의 새로운 레코드가 추가될 수 있다.

[0069] 일부 예들에서, 생성된 레코드만, 예를 들어, 파일, 데이터베이스, 또는 파라미터 세트와 같은 타겟에 저장되고, (만약에 있다면) 소스 레코드는 타겟에 저장되지 않는다. 일부 예에서, 소스는, 예를 들어 사용자에 의해 특정될 수 있고, 생성된 레코드는 소스의 레코드 포맷, 소스의 하나 이상의 필드의 프로파일, 또는 소스의 다른 특성을 기반으로 데이터 생성 엔진(20)에 의해 생성될 수 있다. 일부 예에서, 소스는 특정되지 않고, 레코드는 구성 데이터(34)에 따라 데이터 생성 엔진(20)에 의해 스크래치로부터 자동으로 생성된다. 예를 들어, 레코드 포맷 및 하나 이상의 필드의 특성은 구성 데이터(34)에 의해 특정될 수 있다.

[0070] 일부 예에서, 소스의 레코드 포맷은 타겟에 매핑될 수 있다. 예를 들어, 구성 데이터(34)는 소스의 레코드 포맷이 타겟에 의해 채택되는 것을 지시할 수 있다. 예를 들어, 소스의 레코드 포맷은 구성 데이터(34)에 따라 타겟에 적용될 수 있고, 새로운 레코드는 소스의 레코드 포맷에 따라 데이터 구성 엔진(20)에 의해 스크래치로부터 생성될 수 있다. 일부 예에서, 데이터가 다수의 소스로부터 생성될 수 있고, 각 소스의 레코드 포맷은 부분적으로 또는 완전히 타겟에 매핑될 수 있다. 예를 들어, 관심 필드의 포맷은 각 소스로부터 타겟으로 매핑될 수 있다.

[0071] 일부 예에서, 소스의 레코드 포맷은 타겟으로 매핑되거나 변경될 수 있다. 예를 들어, 구성 데이터(34)는 필드 이름이 소스에서 타겟으로 변경되는 것 또는 필드가 소스에서 제거되는 것을 특정할 수 있다.

[0072] 도 4를 참조하면, 데이터 생성 모듈(150)은 소스 윈도우(300)를 제공하는 사용자 인터페이스(예를 들어, 도 2의 사용자 인터페이스(30))를 제공하여 사용자가 데이터 소스를 식별할 수 있게 한다. 소스 윈도우(300)는 사용자가 파일 또는 데이터 베이스와 같은 소스 타입을 특정하게 하는 소스 타입 메뉴(302)를 포함한다. 도 4의 예에서, 소스는 파일이다. 소스 윈도우(300)는 사용자가 소스로의 경로(예를 들어, 파일 소스로의 경로 또는 데이터 베이스 소스에 대한 구성 파일로의 경로) 및 소스 코드의 레코드 포맷 식별자(예를 들어, 레코드 포맷 파일로의 경로)와 같은 소스의 식별자(304)를 특정할 수 있게 한다. 일부 예에서, 소스가 데이터베이스인 경우, 사용자는 데이터베이스로부터 소스 데이터를 얻는 데 사용되는 질의(query, 예를 들어 SQL 질의)를 특정할 수 있다. 소스 윈도우(300)는 옵션을 제공하여 사용자가 새로운 레코드(306) 생성 여부, 또 그런 경우, 생성될 새로운 레코드의 수를 지시하게 한다. 소스 윈도우(300)는 사용자가 소스에 대한 다른 정보(308)를 볼 수 있게 한다. 예를 들어, 사용자는 소스의 레코드 포맷, 소스 데이터, 또는 소스 데이터의 프로파일을 볼 수 있다. 일부 예에서, 데이터 플로우 그래프와 같이 실행가능한 애플리케이션은 소스 데이터 생성에 사용될 수 있다.

[0073] 도 5A 및 5B를 참조하면, 일부 예에서, 사용자 인터페이스(30)의 소스 윈도우(300)은 사용자가 소스를 특정하지 않고도 데이터가 스크래치로부터 생성되도록 특정하게 할 수 있다. 특히 소스 윈도우(300)는 사용자가 소스 타입 메뉴(302)에서 생성된 데이터를 소스 타입으로 선택할 수 있게 한다. 사용자가 생성된 데이터를 소스 타입으로 선택한다면, 데이터 생성 윈도우(400)는 사용자 인터페이스(30)에 디스플레이된다. 데이터 생성 윈도우(400)는 사용자가 (아래에서 논의될) 데이터 생성에 사용될 방법(402)을 표시하고 생성될 새로운 레코드(404)의 수를 표시할 수 있게 한다. 일부 예에서, 디폴트 데이터 생성 모드 또는 타겟의 레코드 포맷에서 특정된 디폴트 값들과 같은 디폴트 데이터 생성 특성은, 사용자가 디폴트를 무시하지 않는다면, 스크래치로부터의 데이터 생성에 사용될 수 있다. 예를 들어, 필드 생성 무시 옵션(406)은 사용자가 타겟에서 하나 이상의 필드에 대한 데이터 생성 특성을 특정하도록 한다. 일부 예에서 규칙 옵션(408)은 사용자가 개별 필드들에 대한 좀 더 복잡한 데이터 생성 규칙들을 만들게 한다.

[0074] 도 6을 참조하면, 사용자 인터페이스(30)는 사용자가 타겟을 식별할 수 있게하는 타겟 윈도우(500)를 제공한다. 타겟 윈도우(500)에서 타겟 타입 메뉴(502)는 사용자가 파일 또는 데이터베이스와 같은 타겟 타입을 특정할 수 있게 한다. 타겟 윈도우(500)는 또한 사용자가 타겟(예를 들어, 타겟 파일로의 경로 또는 타겟 데이터베이스에 대한 구성 파일로의 경로)의 식별자(504)를 특정할 수 있게 한다. 일부 예에서, 소스의 레코드 포맷은, 예를 들어 타겟 윈도우(500)를 통해 다른 레코드 포맷이 특정되지 않는다면, 타겟에 대한 레코드 포맷으로 사용된다. 예를 들어, 데이터가 소스 없이 스크래치로부터 생성된다면, 레코드 포맷 파일은 타겟에 대한 레코드 포맷을 특정하는 타겟 윈도우를 통해 식별될 수 있다. 일부 예에서, 데이터 플로우 그래프와 같은 실행가능한 애플리케이션은 생성된 데이터를 타겟에 기록하는데 사용될 수 있다. 소스와 타겟이 식별되면, 타겟 윈도우(500)는 사용자에게 데이터 생성을 위한 다양한 구성가능한 옵션에 대한 접근을 제공하는 실행 버튼(506)을 제공한다. 타겟 윈도우(500)는 또한 사용자가 소스에 대한 다른 정보(508)를 볼 수 있게 한다. 예를 들어, 사용자는 타겟 데이터

세트를 볼 수 있고, 타겟 데이터 세트의 프로파일을 볼 수 있고, 타겟의 레코드 포맷을 볼 수 있다. 일부 예에서, 타겟 윈도우는, 소스로부터 판독 및/또는 타겟에 기입된 레코드의 수, 생성된 레코드 수, 또는 다른 통계와 같은 타겟을 위한 데이터 생성과 연계된 통계에 대한 접근을 제공한다.

[0075] 데이터 생성 엔진(20)은 데이터 생성을 위한 몇 가지 접근을 제공한다. 사용자 인터페이스(30)의 데이터 생성 윈도우는 사용자가 원하는 데이터 생성 접근을 특정할 수 있게 한다. 예를 들어, 필드 변경, 필드 생성, 레코드 생성, 기준 소스, 및 부모 데이터세트 접근이 데이터 생성에 사용될 수 있다.

[0076] 필드 변경: 필드 변경 접근에서, 타겟 데이터 레코드가 소스 데이터 레코드와 동일한 포맷을 갖지만 다른 컨텐츠를 갖도록 소스 레코드 내 하나 이상 필드에 대한 값들이 변경될 수 있다. 일부 예에서, 주어진 필드 내 모든 값들이 변경될 수 있다. 예를 들어, 모든 레코드에 걸쳐 주어진 필드에서 값들의 분포가 타겟 분포와 일치하도록 값이 할당될 수 있다. 일부 예에서, 주어진 필드에 대한 모든 값들보다 적은 값들, 예를 들어, 특정된 기준을 만족하지 못하는 값들만 변경될 수 있다. 예를 들어, 특정한 필드에 대한 허용 값들의 특정 범위 밖에 있는 값들이 변경될 수 있다.

[0077] 필드 생성: 필드 생성 접근에서, 기준 레코드에 대해 하나 이상의 새로운 필드가 생성될 수 있다. 일부 예에서, 소스 데이터에서 기준 필드에 대한 데이터를 기반으로 새로운 필드 값들이 생성될 수 있다. 일부 예에서, 예를 들어, 기준 소스 데이터를 기반으로 하지 않지만, 특정된 특성에 따라 새로운 필드 값들이 스크래치로부터 생성될 수 있다.

[0078] 레코드 생성: 레코드 생성 접근에서 새로운 레코드가 생성될 수 있다. 새로운 레코드가 생성되어야 할 때, 생성될 레코드의 수가 (예를 들어, 도 5B에서와 같이) 특정될 수 있다. 일부 예에서, 새로운 레코드의 레코드 포맷이 특정될 수 있다. 예를 들어, 타겟이 기준 레코드와 새로 생성된 레코드 둘 다로 파풀레이팅되는 경우, 새로운 레코드의 레코드 포맷은 소스 레코드의 레코드 포맷과 동일할 수 있다. 타겟이 새로 생성된 레코드로만 파풀레이팅되면, 사용자는 생성된 레코드에 적용될 레코드 포맷을, 필드의 수, 각 필드에 대한 데이터 타입, 각 필드에 대한 데이터 특성 (예를 들어, 최대값, 최소값, 허용된 문자 세트, 및 다른 특성) 및 레코드 포맷의 다른 특징들과 같이, 특정할 수 있다.

[0079] 기준 데이터세트: 기준 데이터세트 접근에서, 데이터는 새로 생성된 자식 데이터 세트에 대해 생성된다. 예를 들어, 기준 고객 데이터세트를 기반으로 거래 레코드를 포함한 새로운 데이터 세트가 생성될 수 있다. 기준 데이터 세트 접근에서, 새로운 레코드의 특정된 개수는 기준 소스 레코드별로 생성된다. 예를 들어, 도 7을 참조하면, 기준 데이터 세트 윈도우(600)는 사용자가 경로(602), 소스 데이터세트의 데이터 조작 언어(DML, 604), 기준 소스 데이터 세트에서 데이터 레코드당 생성될 타겟 레코드의 수(606), 및 타겟 내 레코드 생성에 사용되지 않을 소스 레코드의 백분율을 특정하게 한다. 일부 예에서, 보조 소스는 타겟 레코드의 어떤 필드를 파풀레이팅하는데 사용될 데이터를 포함할 수 있지만, 소스 또는 타겟의 레코드 포맷과 일치하지 않는 레코드 포맷도 가질 수 있다. 그러한 예에서, 하나 이상의 관심 필드는 보조 소스로부터 타겟 레코드로 매핑될 수 있다.

[0080] 주제영역 접근: 주제영역 접근에서, 기준 소스 데이터 세트는 주제영역 계층에서 부모 데이터 세트일 수 있고, 데이터를 생성하는 단계는 부모 데이터 세트와 관련된 자식 데이터 세트를 생성하는 단계를 포함할 수 있다. 예를 들어, 부모 데이터 세트(즉, 소스)는 고객 레코드 세트일 수 있고, 자식 데이터 세트(즉, 타겟)는 고객별 하나 이상의 거래 레코드 세트일 수 있다. 자식 데이터 세트 내 레코드를 부모 세트 내 해당 레코드로 연결하는 키(key) 필드가 특정될 수 있고, 부모 데이터 레코드 내 하나 이상의 필드는 생성된 자식 데이터 레코드 내 해당 필드로 매핑될 수 있다. 예를 들어, 고객 ID (cust_ID) 필드는 고객 레코드와 거래 레코드를 연결하는 키 필드일 수 있다. 키 필드당 생성될 자식 레코드의 수가 특정될 수 있다. 일부 예에서, 자식 레코드 생성에 사용되지 않는 부모 레코드의 백분율이 특정될 수 있다. 일부 예에서, 자식 레코드의 레코드 포맷이 특정될 수 있다.

[0081] 일부 예에서, 데이터는 포맷 사양, 컨텐츠 기준, 혹은 둘 다에 따라 생성될 수 있다. 포맷 사양은 생성될 데이터의 포맷 사양이다. 예를 들어, 포맷 사양은 생성될 데이터의 데이터 타입(예를 들어, 스트링, 소수, 정수, 날짜, 또는 일시)을 나타낼 수 있다. 컨텐츠 기준은 생성될 데이터의 특성을 제한하는 기준이다. 예시적인 컨텐츠는, 예를 들어, 허용된 값들의 범위, 최대 허용치, 허용된 문자 리스트, 또는 다른 컨텐츠 기준을 포함할 수 있다. 일부 예에서, 포맷 사양과 컨텐츠 기준은 타겟 레코드의 레코드 포맷으로 특정된다. 일부 예에서, 사용자 인터페이스(30)는 사용자가 필드별 포맷 사양 또는 컨텐츠 기준과 같은 필드 특성을 특정할 수 있게 하는 필드 윈도우를 제공할 수 있다.

- [0082] 도 8를 참조하면, 사용자 인터페이스의 레코드 포맷 윈도우(700)는 사용자가 타겟의 하나 이상의 필드에 대한 데이터 타입 및 데이터 특성을 포함한 타겟 레코드 포맷을 편집할 수 있게 한다. 레코드 포맷 윈도우(700)는 타겟의 레코드 포맷에 있는 필드 리스트(702)를 디스플레이한다. 리스트(702)는 또한 각 필드의 데이터 타입을 나타낸다. 일부 예에서, 타겟이 레코드 포맷에 있는 필드들은 또한 데이터 소스의 레코드 포맷으로 나타난다. 타겟 레코드 포맷 및 소스 레코드 포맷 둘 다에 나타난 필드들은 리스트(702)에서, 예를 들어 애스터리스크 (즉, **)로 표시될 수 있다. 예를 들어, last_transaction 필드와 customer_since 필드는 타겟 레코드 포맷과 소스 레코드 포맷 둘 다에 나타난다. 표시되지 않은 필드는 타겟 레코드 포맷에만 나타난다. 일부 예에서, 타겟 레코드 포맷에는 나타나지 않고 소스 레코드 포맷에 나타난 필드는 리스트(702)에 디스플레이되지 않는다.
- [0083] 레코드 포맷 윈도우(700)는 데이터 생성 특성을 특정하기 위해 사용자가, 예를 들어, 원하는 필드 이름을 클릭하거나 텁핑(tapping)하거나, 또는 드래그하여 타겟 레코드 포맷의 하나 이상의 필드를 선택하게 할 수 있다. 타겟 레코드 포맷의 선택된 필드 리스트(704)가 디스플레이된다. 즉, 리스트(704) 내 필드는 사용자가 데이터 생성 특성을 특정하려고 하는 타겟 레코드 포맷의 필드들이다. 일부 예에서, 사용자가 타겟 레코드 포맷 필드의 일부에 대해서만 데이터 생성 특성을 특정하려고 한다면, 리스트(704)는 타겟 레코드 포맷 내 모든 필드들의 리스트(702)의 서브세트일 수 있다. 예를 들어, 도 8의 예에서 custid, fname, lname, 및 building_num 필드들이 선택되었다.
- [0084] 사용자 인터페이스(30)는 사용자가 리스트(704)에 디스플레이된 선택된 각 필드에 대한 레코드 포맷을 편집할 수 있게 한다. 예를 들어, 각 선택된 필드들에 대해, 사용자는 필드에 대한 데이터 타입을 지정하기, 컨텐츠 생성 모드를 필드에 할당하기, 및 필드에 대한 데이터 특성을 특정하기를 조합할 수 있다. 사용자 인터페이스는 데이터 타입 윈도우, 컨텐츠 생성 윈도우 및 각 선택된 필드에 대한 데이터 특성 윈도우 중 하나 이상을 디스플레이하고, 따라서 사용자는 각 선택된 필드에 대한 다양한 특징들을 특정하게 할 수 있다.
- [0085] 사용자 인터페이스(30)는 사용자가 필드에 대한 데이터 타입을 지정할 수 있게 한다. 데이터 타입 윈도우는 리스트(704) 내 특정한 필드에 대한 사용자 선택에 응답하는 사용자 인터페이스(30)에 디스플레이되어 사용자가 해당 필드에 대한 데이터 타입을 지정할 수 있게 한다. 사용자는, 스트링, 소수, 정수, 날짜, 및 일시(예를 들어, 날짜와 시간)과 같이 데이터 생성 모듈(150)에 의해 지원되는 임의의 데이터 타입을 지정할 수 있다. 하나의 데이터 타입이 각 필드에 대해 지정될 수 있다. 일부 예에서, 데이터 타입이 주어진 필드에 대해 지정되지 않는다면, 해당 필드에 대한 레코드 포맷에 의해 지시된 데이터 타입이 유지된다. 예를 들어, 도 8의 예에서, 해당 필드에 대한 레코드 포맷으로 표시된 것처럼 필드별 데이터 타입은 필드 이름 다음의 괄호로 보여진다. 타겟 레코드 포맷의 각 필드에 지정된 데이터 타입은 구성 데이터(34)로 저장된다.
- [0086] 사용자 인터페이스(30)는 사용자가 컨텐츠 생성 모드를 하나 이상의 선택된 필드들에 할당할 수 있게 한다. 컨텐츠 생성 윈도우는 리스트(704) 내 특정한 필드에 대한 사용자 선택에 대한 응답으로 사용자 인터페이스(30)에 디스플레이되어 사용자가 해당 필드에 컨텐츠 생성 모드를 할당할 수 있게 한다. 필드별 컨텐츠 생성 모드는 데이터 생성 엔진(20)에 의해 데이터가 필드별로 생성되는 방법을 나타낸다. 예를 들어, 예시적인 컨텐츠 생성 모드는 고유 모드, 랜덤 모드, 디폴트 모드 및 리스트 모드를 포함할 수 있다. 고유 모드에서는, 각 레코드의 선택된 필드에 대해 고유값이 생성될 수 있다. 예를 들어, 내림차순의 고유 소수, 스트링, 또는 정수 값 세트가 생성될 수 있다. 데이터 값들은 디폴트로 일별 내림차순으로 또는 19000101과 같은 특정된 날짜 베이시스로부터, 또는 다른 날짜 베이시스로부터 생성될 수 있다. 일시 값들은 디폴트로 분별 내림차순으로 또는 현재의 날짜와 시간과 같은 특정된 일시 베이시스로부터 생성될 수 있다. 랜덤 모드에서는, 각 레코드의 선택된 필드에 대해 랜덤값이 생성될 수 있다. 디폴트 모드에서는, 타겟의 레코드 포맷으로 특정된 디폴트 특성 또는 디폴트 값들이 값 생성에 사용될 수 있다. 예를 들어, 일부 예에서, 디폴트 데이터 생성 모드는 랜덤 모드 일 수 있고, 디폴트 값의 범위 및 특성은 각 데이터 타입에 적용될 수 있다. 리스트 모드에서, 값들은 허용 값 리스트에서 선택된다. 일부 예에서, 컨텐츠 생성 모드가 주어진 필드에 할당되지 않는다면, 디폴트 모드는 필드별 데이터에 대한 데이터 생성에 사용되거나, (예를 들어 타겟 레코드의 레코드 포맷으로 특정된 것과 같은) 디폴트 값이 해당 필드에 대한 값으로 사용될 수 있다. 일부 예에서, 하나 이상의 필드에 대한 컨텐츠 생성 모드는 타겟 레코드의 레코드 포맷으로 특정될 수 있다. 타겟 레코드 포맷의 각 필드에 대한 컨텐츠 생성 모드는 구성 데이터(34)로 저장될 수 있다.
- [0087] 고유 모드, 랜덤 모드, 또는 디폴트 모드의 데이터 생성이 필드에 할당된다면, 사용자 인터페이스(30)는 사용자가 하나 이상의 선택된 필드에 대한 데이터 특성을 특정할 수 있게 한다. 데이터 특성은 해당 필드에 대한 데이터 타입에 종속될 수 있고, 해당 필드에 대해 생성될 수 있는 허용치를 제한할 수 있다. 예를 들어, 소수 필드 또는 정수 필드와 같은 수치 필드에 대한 데이터 특성은 최대 허용치, 최소 허용치, 허용값 범위, 또는 모든 레

코드에 결친 필드별 평균치를 나타낼 수 있다. 다른 데이터 특성도 하기와 같이 적용될 수 있다. 일부 예에서, 필드 에러율 (즉, 에러를 포함하는 레코드의 개수 또는 백분율) 또한 하나 이상의 필드에 대해 특정될 수 있다. 데이터 특성이 특정 필드에 대해 특정될 수 있게 하기 위해, 해당 필드의 데이터 타입에 특정된 특성 윈도우가리스트(704) 내 해당 필드에 대한 사용자 선택에 반응하는 사용자 인터페이스(30)에 디스플레이된다. 예를 들어, 스트링 특성 윈도우, 소수 특성 윈도우, 정수 특성 윈도우, 날짜 특성 윈도우, 또는 일시 특성 윈도우, 또는 다른 데이터 타입용 윈도우가 사용자 인터페이스(30)에 디스플레이될 수 있다. 타겟 레코드 포맷의 각 필드에 대한 데이터 특성은 구성 데이터(34)로 저장될 수 있다.

[0088] 도 9를 참조하면, 스트링 특성 윈도우(800)는 사용자가 스트링 필드에 대한 데이터 특성을 특정할 수 있게 한다. 도 9의 예에서, 데이터 특성은 평균 스트링 길이(802), 최대 스트링 길이(804), 및 특정한 스트링 필드에 대한 허용가능한 문자 세트(806)를 포함한다. 사용자는 또한 특정한 스트링 필드에 대한 NULL값들(808)의 백분율을 특정할 수 있다. 일부 예에서, 다른 데이터 특성도 특정될 수 있다. 일부 예에서, 사용자가 스트링 필드에 대한 하나 이상의 데이터 특성을 특정하지 않는다면, 데이터 특성에 대한 디폴트 값을 적용할 수 있다. 예를 들어, 평균 스트링 길이(802)의 디폴트 값은 8일 수 있다; 최대 길이(804)의 디폴트 값은 16일 수 있다; 허용가능한 문자 세트(806)의 디폴트 세트는 문자들 {a-z, A-Z, 0-9, 및 스페이스(space)}을 포함할 수 있다; 그리고 NULL 백분율(808)의 디폴트 값은 0일 수 있다.

[0089] 특정한 예에서, 특정한 필드에 대해 사용자가 특정한 평균 스트링 길이(802)는 4개 문자, 최대 스트링 길이(804)는 8개 문자, 허용가능한 문자 세트(806)는 {a-z, A-Z}, 및 NULL 백분율 값(808)은 0이다. 이 예에서, 사용자가 특정한 필드에 랜덤 모드 데이터 생성을 할당한다면, 랜덤 스트링은 특정된 평균 길이와 최대 길이 내에서 허용가능한 문자 세트로부터 각 레코드 내 해당 필드에 대해 생성될 것이다. 사용자가 특정한 필드에 고유 모드 데이터 생성을 할당한다면, 고유 스트링은 각 레코드의 그 필드에 대해 (즉, 두 레코드가 특정한 필드에 대해 동일하게 생성된 스트링을 갖지 않도록) 생성될 것이다.

[0090] 도 10을 참조하면, 소수 특성 윈도우(900)는 사용자가 소수 필드에 대한 데이터 특성을 특정할 수 있게 한다. 도 10의 예에서, 데이터 특성은 특정한 소수 필드에 대해 허용치 범위(902) 및 NULL 값의 백분율(904)을 포함한다. 일부 예에서, 허용된 소수 위치의 최대 수, 모든 생성된 레코드의 소수 필드에 대한 평균치, 또는 다른 특성과 같은 다른 데이터 특성도 특정될 수 있다. 일부 예에서, 사용자가 하나 이상의 데이터 특성을 특정하지 않는다면, 디폴트 값을 적용할 수 있다. 예를 들어, 디폴트 범위(902)는 0 내지 1000일 수 있고, NULL 값의 디폴트 백분율(904)은 0일 수 있다.

[0091] 특정한 예에서, 사용자는 특정한 필드에 대해 허용치 범위(902)를 0 내지 10 및 NULL 값의 백분율(904)을 10%로 특정하였다. 이 예에서, 사용자가 특정 필드에 랜덤 모드 데이터 생성을 할당한다면, 허용치 범위(902) 이내의 랜덤 소수는 각 레코드의 해당 필드에 대해 생성될 것이다. 레코드의 10%가 NULL 값으로 주어질 것이다. 사용자가 특정 필드에 고유 모드 데이터 생성을 할당한다면, 허용치 범위(902) 이내의 고유 소수가 각 레코드의 해당 필드에 대해 (즉, 두 레코드가 특정 필드에 대해 동일한 값을 갖지 않도록) 생성될 것이다.

[0092] 도 11을 참조하면, 정수 특성 윈도우(180)는 사용자가 정수 필드에 대한 데이터 특성을 특정할 수 있게 한다. 도 11의 예에서, 데이터 특성은 허용치 범위(182)를 포함한다. 일부 예에서, 특정 정수 필드에 대한 NULL 값의 백분율(184) 또는 모든 생성된 레코드에 대한 정수 필드의 평균값과 같은 다른 특성도 특정될 수 있다. 일부 예에서, 사용자가 하나 이상의 데이터 특성을 특정하지 않는다면, 디폴트 값을 적용할 수 있다. 예를 들어, 디폴트 범위(182)는 0 내지 127일 수 있다.

[0093] 도 12를 참조하면, 날짜 특성 윈도우(190)는 사용자가 날짜 필드에 대한 데이터 특성을 특정할 수 있게 한다. 도 12의 예에서, 데이터 특성은 년도, 달, 또는 일과 같은 날짜 범위 메트릭(192)을 포함한다. 날짜 특성은 또한 YYYY가 년도, MM이 달, DD가 날을 나타낼 때 날짜 범위 계산(예를 들어, YYYYMMDD로 쓰인 것)과 관련된 베이시스로 사용되는 날짜인 날짜 베이시스(194)를 포함한다. 날짜 특성은 또한 특정한 날짜 필드에 대해 날짜 베이시스(194)와 관련된 허용 날짜 범위 (196) 및 NULL 또는 비어있는(blank) 값의 백분율(198)을 포함한다. 일부 예에서, 다른 데이터 특성도 특정될 수 있다. 일부 예에서, 사용자가 하나 이상의 데이터 특성을 특정하지 않는다면, 디폴트 값을 적용할 수 있다. 예를 들어, 디폴트 데이터 범위 메트릭(192)은 년도일 수 있고, 디폴트 날짜 베이시스(194)는 오늘일 수 있으며, 허용 날짜 범위(196)는 (디폴트 날짜 베이시스와 관련해) -7 내지 0 년 일 수 있고, 디폴트 NULL 값 백분율(198)은 1%일 수 있다.

[0094] 특정한 예에서, 사용자는 특정한 필드에 대해 날짜 범위 메트릭(192)을 일, 데이터 베이시스(194)를 2013. 1. 1, 허용 날짜 범위(196)를 365 (즉, 2013. 1. 1의 날짜 베이시스를 초과해 365일) 및 NULL 값 백분율(198)을 0

으로 특정했다. 이 예에서, 사용자가 특정한 필드에 랜덤 모드 데이터 생성을 할당한다면, 2013. 1. 1과 2014. 1. 1 사이에서 랜덤하게 선택된 날짜가 각 레코드의 해당 필드에 할당될 것이다. 사용자가 특정한 필드에 고유 모드 데이터 생성을 할당한다면, 2013. 1. 1과 2014. 1. 1 사이의 고유 날짜가 각 레코드의 해당 필드에 (즉, 두 레코드가 특정한 필드에 대해 동일한 날짜를 갖지 않도록) 할당될 것이다.

[0095] 도 13을 참조하면, 일시 특성 윈도우(250)는 사용자가 일시 필드에 대한 데이터 특성을 특정하도록 한다. 도 13의 예에서, 데이터 특성은 년도, 달, 또는 일과 같은 일시 범위 메트릭(252)을 포함한다. 날짜 특성을 또한 일시 베이시스(254)를 포함하고, 여기서 YYYY가 년도, MM이 달, DD가 날, HH24가 24시간 시계상의 시간, MI가 분, SS가 초, 그리고 NNN이 1000분의 1초를 나타낼 때, 날짜 범위 계산(예를 들어, YYYYMMDD HH24MISS.NNN으로 쓰인 것)과 관련된 베이시스로 사용되는 날짜와 시간인 일시 베이시스(254)를 포함한다. 날짜 특성은 또한 특정한 일시 필드에 대해 일시 베이시스(254)와 관련된 허용 일시 범위 (256) 및 NULL 값의 백분율(258)을 포함한다. 일부 예에서, 다른 데이터 특성이 특정될 수 있다. 일부 예에서 사용자가 하나 이상의 데이터 특성을 특정하지 않는다면, 디폴트 값을 적용할 수 있다. 일부 예에서, 디폴트 일시 범위 메트릭(252)은 일(day)일 수 있고, 디폴트 일시 베이시스(254)는 지금일 수 있으며, 디폴트 허용 일시 범위(256)는 -7 내지 0일일 수 있고, 디폴트 NULL 값 백분율(258)은 0일 수 있다.

[0096] 일부 예에서, 하나 이상의 특성 윈도우 (예를 들어, 스트링 특성 윈도우(800), 소스 특성 윈도우(900), 정수 특성 윈도우(180), 날짜 특성 윈도우(190), 또는 일시 특성 윈도우(250))는 사용자가, 예를 들어, 필드에 대한 필드에러율 값을 선택 또는 특정하여 해당 필드의 디폴트 에러율을 무시하게 할 수 있다. 예를 들어, 사용자는 다음의 에러율에서 선택할 수 있다: 0, 에러가 있는 2개 필드 중 1, 10개 필드 중 1, 100개 필드 중 1, 1000개 필드 중 1, 10000개 필드 중 1, 및 100000개 필드 중 1. 일부 예에서, 사용자는 에러 백분율을 선택 또는 특정할 수 있다.

[0097] 일부 예에서, 사용자 인터페이스(30)는 사용자에게 하나 이상의 데이터 타입에 대한 디폴트 데이터 특성을 무시하는 메커니즘을 제공한다. 예를 들어, 사용자는 모든 스트링 데이터에 대한 평균 스트링 길이를 무시할 수 있다. 사용자가 특정 데이터 타입에 대해 디폴트 특성을 무시하는 경우, 그 무시 특성은 사용자가 데이터 특성을 특정하지 않은 해당 데이터 타입의 모든 데이터 생성에 적용될 수 있다.

[0098] 리스트 모드 데이터 생성이 필드에 할당된다면, 필드에 대한 허용치 리스트가 제공될 수 있게 하는 사용자 인터페이스(30)에 리스트 윈도우가 디스플레이될 수 있다. 예를 들어, 리스트 모드 데이터 생성이 빌딩 수(building_num) 필드에 할당된다면, 사용자는 빌딩 수(building_num) 필드에 대한 수치로 제공될 수 있는 건물 수 리스트 (예를 들어 {10, 20, 30, 30, 50})를 제공할 수 있다. 데이터가 데이터 생성 엔진(150)에 의해 빌딩 수(building_num) 필드에 대해 생성된 경우, 각 레코드의 필드값은 빌딩 수 리스트로부터 랜덤하게 선택될 수 있다. 일부 예에서, 다른 제한조건이 적용할 수 있다. 예를 들어, 필드값은 리스트상의 각 항목이 동일한 횟수로 선택되는 제한조건에 따라 리스트에서 선택될 수 있다.

[0099] 일부 예에서, 데이터 생성 또는 소스에서 타겟으로의 포맷 매핑에 대한 고급규칙이 생성될 수 있다. 예를 들어, 고급 규칙은 한 필드 내 값들을 다른 필드로 매핑하는 규칙, 한 필드 내 값들을 사용하여 다른 필드 내 값들을 계산하는 규칙, 또는 소스에서 하나 이상의 필드를 사용하여 타겟에서 단일 필드를 생성하는 규칙을 생성할 수 있다. 일부 예에서, 고급 규칙은 데이터 소스 필드 내 값들을 타겟 필드에 대해 생성될 값들과 연관시킬 수 있다. 일부 예에서, 고급 규칙은 타겟의 제1필드 내 값들을 타겟의 제2필드에 대해 생성될 값들과 연관시킬 수 있다. 사용자 인터페이스(30)는 사용자가 하나 이상의 타겟 필드에 대해 데이터 생성을 위한 고급 규칙을 생성할 수 있게 하는, 도 14A 내지 14C, 15A 내지 15C, 및 16에 도시된 구성 화면 및 규칙 작성기 등과 같은 규칙 작성기 기능을 제공한다. 예를 들어, 사용자 인터페이스(30)는 사용자가 필드와 연계된 규칙을 특정할 수 있게 하여, 규칙이 필드를 파풀레이팅하는 데이터의 생성에 사용될 수 있는 규칙 윈도우를 제공할 수 있다. 사용자 인터페이스(30)는 사용자가 규칙을 정의할 수 있게 하는 규칙 작성기를 제공할 수 있다.

[0100] 예를 들어, 도 14A 내지 14C를 참조하면, 사용자 인터페이스(30)의 규칙 작성기 기능은 사용자가 한 필드 내 값들을 다른 필드로 매핑하는 고급 규칙을 생성할 수 있게 한다. 이 예에서, 사용자는 값이 해당 전체 주 이름인 주 이름(state_name) 필드를 기반으로, 값이 뉴잉글랜드 주의 두 글자 약자 리스트에서 선택된 주 코드(state_cd) 필드를 생성할 수 있게 하는 규칙을 만들고 싶어한다. 이 규칙을 만들기 위해, 사용자는 주 코드(state_cd) 필드(도 14A)에 대한 구성 윈도우(350)를 통해 주 코드(state_cd) 필드로 리스트 생성 모드를 할당하고, 주 코드(state_cd) 필드값이 선택될 수 있는 리스트(352)(이 경우, 뉴 잉글랜드 주의 두 글자 약자 리스트: MA, CT, RI, VT, ME, NH)를 특정할 수 있다. 규칙 윈도우(354, 도 14B)에서, 사용자는 주 이름(state_name) 필

드(356)가 주 이름(state_name)필드(356)에 데이터를 파풀레이팅하는데 사용될 수 있는 규칙인 "주 이름 계산" (358)과 연계되는 것을 표시할 수 있다. 사용자는 규칙 작성기(360)에서 주 이름(state_name) 필드에 대한 "주 이름 계산" 규칙을 정의할 수 있다 (도 14C). 이 예에서, "주 이름 계산" 규칙은 각 주 코드(state_cd)필드 값에 대해 주 이름(state_name) 필드로 할당되는 값을 특정할 수 있다.

[0101] 예를 들어, 도 15A 내지 15C를 참조하면, 사용자 인터페이스(30)의 규칙 작성기 기능은 사용자가 타겟 내 제1필드 값을 사용하여 타겟 내 제2필드 값을 계산하는 고급 규칙을 만들 수 있게 한다. 이 예에서, 사용자는 수입(income)필드 값을 사용하여 세금율(tax_rate) 필드 값을 계산하는 규칙을 만들기를 원한다. 이 규칙을 만들기 위해, 사용자는 수입(income) 필드에 대한 컨텐츠 생성 윈도우(450)를 통해 랜덤 생성 모드를 수입(income) 필드에 할당하고(도 15A), 수입(income) 필드에 대한 허용치 범위, 빈 값들의 백분율과 같은 데이터 특성(452)을 특정할 수 있다. 규칙 윈도우(454, 도 15B)에서, 사용자는 세율(tax_rate) 필드(456)가 세율(tax_rate) 필드(456)에 데이터를 파풀레이팅하는데 사용될 수 있는 "세율 계산" 규칙(458)과 연계됨을 표시할 수 있다. 사용자는 규칙 작성기(460)에서 세율(tax_rate) 필드에 대한 규칙을 정의할 수 있다 (도 15C). 이 예에서, "세율 계산" 규칙은 각 수입(income) 필드 값에 대해 세율(tax_rate) 필드에 할당되는 값을 특정할 수 있다. "세율 계산" 규칙은 또한, 예를 들어, 수입(income) 필드가 비었거나, NULL, 또는 에러인 경우, 세율(tax_rate) 필드에 할당되는 디폴트 값(462)을 특정할 수 있다.

[0102] 예를 들어, 도 16을 참조하면, 사용자 인터페이스(30)의 규칙 작성기 기능은 사용자가 다수의 소스 필드를 단일 타겟 필드로 매핑하는 고급 규칙을 생성할 수 있게 한다. 일부 예에서, 하나 이상의 소스 필드의 레코드 포맷은 타겟 필드의 레코드 포맷과 다를 수 있다. 이 예에서, 사용자는 성명(fullname) 필드에 대한 데이터 생성에 이름(firstname) 필드와 성(lastname) 필드값을 사용하는 표현문(expression)을 특정하는 규칙을 생성하기를 원한다. 이 규칙을 생성하기 위해, 규칙 윈도우(550)에서 사용자는 성명(fullname) 필드(554)에 대한 규칙(552)을 소스의 이름(firstname) 필드(556), 스페이스, 소스의 성(lastname) 필드(558)의 조합("이름" + " " + "성")으로 정의할 수 있다.

[0103] 데이터는 하나 이상의 사양(들), 컨텐츠 기준 또는 기준들, 데이터 생성 모드, 및 고급 규칙을 기반으로 생성될 수 있다. 예를 들어, 소스 레코드 데이터가 변경될 수 있고, 하나 이상의 새로운 필드가 생성되어 데이터로 파풀레이팅될 수 있으며, 및/또는 새로운 레코드가 생성되어 데이터로 파풀레이팅될 수 있다.

[0104] 일부 예에서, 주제 영역이 생성될 수 있다. 주제 영역은 적어도 하나의 필드에서 중요 관계를 통해 계층에서 서로 관련된 데이터 세트를 수집한 것이다. 예를 들어, 기존 부모 데이터세트는 고객 레코드 세트일 수 있고, 자식 데이터세트는 각 기존 고객에 대한 하나 이상 생성된 거래 레코드 세트일 수 있다.

[0105] 도 17A를 참조하면, 일례에서, 연결(join) 윈도우(70)는 사용자가 하나 이상의 기존 부모 데이터세트 필드를 특정하여 생성될 자식 데이터세트와 함께 키 필드로 사용되도록 특정하게 한다. 도 17A의 예에서, 기존 부모 데이터세트는 고객 데이터세트이고, 거래 데이터세트는 연결 키 필드인 고객 ID(cust_id)와 함께 자식으로 생성될 것이다. 도 17B를 참조하면, 연결 윈도우(75)는 사용자가 자식 데이터세트에 대한 하나 이상의 연결 키 필드를 특정하게 한다. 이 예에서, 고객 ID(cust_id) 필드는 이미 박스(76)에서 부모 데이터세트로부터 연결 키 필드로 특정되었다. 박스(78)에서, 필드는 자식 데이터세트로부터 연결 키 필드로 특정될 수 있다. 일부 예에서, 대응하는 자식 레코드가 없는 부모 레코드 또는 대응하는 부모 레코드가 없는 자식 레코드와 같은 비연결(non-joining) 레코드는 주제 영역 데이터세트에서 제외될 수 있다. 일부 예에서, 비연결 레코드도 주제 영역 데이터세트에 포함될 수 있다.

[0106] 일부 예에서, 주제 영역은 참조 무결성(referential integrity)을 위해 생성될 수 있고, 이는 주제 영역이 주제 영역의 계층적 데이터세트의 참조 무결성을 보장하는 데이터 레코드만을 제공함을 의미한다. 참조 무결성에 의해, 우리는 주제 영역의 각 레코드가 주제영역 내 적어도 다른 레코드와 연결됨을 의미한다. 예를 들어, 예시적인 주제영역은 계정 식별자를 각 거래 레코드와 연관시키는 제1데이터세트를 포함할 수 있다. 주제 영역의 다른 데이터세트에서, 각 고객 식별자는 많은 계정 식별자들 및 많은 제품 식별자들과 연계될 수 있다. 주제영역의 참조 무결성 보장은 주제영역 데이터세트에 포함된 모든 레코드들이 주제영역 데이터세트 내 하나 이상의 다른 레코드들과 유효하게 연결됨을 보장함을 의미한다. 일부 예에서, 주제영역은 참조 무결성을 보장하는 적용 가능한 서브세팅 규칙과 데이터 레코드로 인해 출력되는 모든 데이터 레코드를 포함하도록 생성될 수 있다. 즉, 예를 들어, 참조 무결성 기준을 만족하지 않는 레코드도 주제영역에 포함되어 서브세팅 규칙을 만족시킬 수 있다. 서브세팅 규칙에 관한 추가 설명은 "데이터 레코드 선택"이라는 제목의 미국 출원번호 13/827,558호에 기재되어 있으며, 그 내용은 본 명세서에 참조로 포함된다.

- [0107] 일부 예에서, 하나 이상의 기존 데이터세트를 기반으로 가상 데이터세트가 생성될 수 있다. 예를 들어, 고객 데이터세트와 거래 데이터세트를 기반으로, 고객별 거래 수 또는 고객별 달러 소비량과 같은 값들을 갖는 하나 이상의 필드를 포함하는 새로운 롤업(rollup) 데이터세트가 생성될 수 있다. 그러한 롤업 데이터세트는, 예를 들어, 대량 구매 고객 또는 소량 구매 고객에 대해 동작하는 데이터 처리 애플리케이션에서 사용될 수 있다.
- [0108] 도 2를 참조하면, 일부 예에서 분석 엔진(38)은 자동화된 분석을 수행하여 하나 이상의 포맷 사양(들), 컨텐츠 기준 또는 기준들, 데이터 생성 모드, 및 생성된 데이터에 대한 고급 규칙을 구현할 수 있다. 예를 들어, 기본 예로서, 사용자가 날짜 베이시스를 2000. 5. 1., 허용 날짜 범위를 10년으로 특정한다면, 분석 엔진(38)은 사용자가 정확한 날짜를 특정하지 않아도 허용 날짜 범위로서 2000. 5. 1.을 2010. 5. 1로 식별할 수 있다.
- [0109] 다른 예에서, 분석 엔진(38)은 사용자에 의해 특정된 평균 스트링 길이 및 최대 스트링 길이를 달성할 다수의 레코드에 걸쳐 스트링 길이 분포를 결정할 수 있다. 예를 들어, 분석 엔진은 디폴트 최대 스트링 길이 및 디폴트 평균 스트링 길이 사이에서 디폴트 타겟 비율을 기반으로 스트링 길이 분포를 결정할 수 있다. 디폴트 타겟 비율은 최대 스트링 길이에 적용되어 가우시안(Gaussian) 분포와 같은 스트링 길이 분포에서 피크(peak) 스트링 길이를 식별할 수 있다. 예를 들어, 일례에서, 디폴트 타겟 비율이 0.5가 되도록 디폴트 최대 스트링 길이는 16이고, 디폴트 평균 스트링 길이는 8이다. 이 디폴트 타겟 비율은, 예를 들어, 랜덤 스트링 생성과 같은 스트링 생성에 적용될 수 있다. 예를 들어, 최대 스트링 길이가 6으로 제한된 필드에서, 디폴트 타겟 비율이 적용되어 피크 스트링 길이가 3으로 결정될 수 있다. 따라서 스트링은 피크 스트링 길이 3 주변에서 가우시안 분포로 생성될 것이다. 다른 예에서, 최대 스트링 길이가 25로 제한된 필드에서, 디폴트 타겟 비율이 적용되어 피크 스트링 길이가 12 내지 13으로 결정될 수 있다. 따라서 스트링은 피크 스트링 길이 12 내지 13 주변에서 가우시안 분포로 생성될 것이다.
- [0110] 일부 예에서, 분석 엔진(38)은 하나 이상의 포맷 사양(들), 컨텐츠 기준 또는 기준들, 데이터 생성 모드, 및 자동으로, 예를 들어, 소스 레코드에 대한 자동 분석에 의해 고급 규칙을 결정할 수 있다. 예를 들어, 분석 엔진(38)은 기존 데이터를 어떻게 변경할 것인지 및/또는 특정 필드에 대해 새로 데이터를 어떻게 생성할 것인지를 결정하여 해당 필드에 대한 값들의 타겟 분포를 달성할 수 있다.
- [0111] 도 18을 참조하면, 데이터 생성(650)에 대한 일반적인 접근에서, 필드(652)에 대한 포맷 사양이 수신된다. 포맷 사양은, 예를 들어, 스트링, 소수, 정수, 날짜, 또는 일시와 같은 필드에 대해 생성될 데이터에 대한 데이터 타입일 수 있다. 예를 들어, 시스템(100)은 데이터 생성(650)에 대한 일반적인 접근에 따라 데이터를 생성할 수 있다. 일부 예에서, 포맷 사양은 사용자 인터페이스(30) (예를 들어, 사용자 인터페이스(30)를 통해 제공되는 데이터 타입 원도우)와 같은 사용자 인터페이스를 통해 사용자로부터 수신될 수 있다.
- [0112] 컨텐츠 기준은 필드(654)에 대해 수신된다. 컨텐츠 기준은, 예를 들어, 필드에 대해 생성될 데이터에 대한 데이터 특성을 수 있다. 예를 들어, 소수 필드 또는 정수 필드와 같은 수치 필드에 대한 데이터 특성은 최대 허용치, 최소 허용치, 허용값 범위, 또는 모든 레코드에 걸친 필드에 대한 평균치를 나타낼 수 있다. 스트링 필드에 대한 데이터 특성은 최대 허용 문자 수, 최소 허용 문자 수, 평균 문자 수, 또는 허용 문자 세트를 나타낼 수 있다. 데이터 필드 또는 일시 필드와 같은 시간 필드에 대한 데이터 특성은 날짜 또는 일시 메트릭, 날짜 또는 일시 베이시스, 또는 날짜 또는 일시 범위를 나타낼 수 있다. 일부 예에서, 필드 에러율 (즉, 에러를 포함하는 레코드의 개수 또는 백분율) 또한 하나 이상의 필드에 대해 특정될 수 있다. 일부 예에서, 컨텐츠 기준은 사용자 인터페이스(30)(예를 들어 스트링 특성 원도우(800), 소수 특성 원도우(900), 정수 특성 원도우(180), 날짜 특성 원도우(190), 또는 일시 특성 원도우(250))와 같은 사용자 인터페이스를 통해 사용자로부터 수신될 수 있다.
- [0113] 필드에 대한 컨텐츠 생성 모드의 지시가 수신된다(656). 컨텐츠 생성 모드는 컨텐츠가 필드에 대해 생성되어야 하는 모드이다. 예를 들어, 컨텐츠 생성 모드는 다음 모드들: 고유 모드, 랜덤 모드, 디폴트 모드, 및 리스트 모드로부터 선택될 수 있다. 고유 모드에서는, 선택된 필드에 대해 고유값이 생성될 수 있다. 랜덤 모드에서는, 선택된 필드에 대해 랜덤값이 생성될 수 있다. 디폴트 모드에서는, 예를 들어, 타겟의 레코드 포맷으로 특정된 디폴트 특성 또는 디폴트 값들이 값 생성에 사용될 수 있다. 리스트 모드에서는, 허용치 리스트에서 값들이 선택된다. 일부 예에서, 컨텐츠 생성 모드가 주어진 필드에 대해 할당되지 않는다면, 디폴트 모드가 해당 필드에 대한 데이터 생성에 사용될 수 있다. 일부 예에서, 컨텐츠 생성 모드의 표시가 사용자 인터페이스(30) (예를 들어, 사용자 인터페이스(30)를 통해 제공되는 컨텐츠 생성 원도우)와 같은 사용자 인터페이스를 통해 사용자로부터 수신될 수 있다.
- [0114] 다수의 레코드 각각에 대해 필드에 대한 컨텐츠는 표시된 컨텐츠 생성 모드에 따라 생성된다(658). 필드에 대해

생성된 컨텐츠는 필드에 대한 포맷 사양과 컨텐츠 기준을 만족한다. 일부 예에서, 새로운 필드는 기존 다수의 레코드들 각각에 대해 생성될 수 있고, 컨텐츠는 표시된 컨텐츠 생성 모드에 따라 새로운 필드에 대해 생성될 수 있다. 일부 예에서, 새로운 다수의 레코드들이 생성되고, 각 레코드는 컨텐츠가 표시된 컨텐츠 모드에 따라 생성될 수 있는 하나 이상의 필드를 갖는다. 예를 들어, 새로운 다수의 레코드는 각각의 기준 레코드를 기반으로 생성될 수 있다.

[0115] 일부 예에서, 데이터 생성은 하나 이상의 원하는 특징을 갖는 레코드 세트 생성에 사용될 수 있다. 예를 들어, 데이터 생성은 데이터 처리 애플리케이션에 의한 처리를 위해 필드 내 원하는 레코드 수 및/또는 원하는 값의 분포를 갖는 레코드 세트 생성에 사용될 수 있다. 예시적인 데이터 처리 애플리케이션은, 예를 들어, 고객 청구서(예를 들어, 전화고객에 대한 청구서)를 위한 데이터 처리 애플리케이션, 고객 레코드를 처리 (예를 들어, 고객 구매 패턴에서 인구통계학적 경향을 식별)하기 위한 데이터 처리 애플리케이션, 및 다른 데이터 처리 애플리케이션 탑재를 포함한다.

[0116] 데이터 처리 애플리케이션은 종종 실행시 하나 이상의 변수 값에 종속되는 (예를 들어, 트리거되는) 규칙을 구현한다. 예를 들어, 변수는 입력 데이터에 대응하는 입력 변수들 또는 하나 이상의 입력 변수들에 의존하는 파생 변수들일 수 있다. 데이터 처리 애플리케이션의 효과적인 테스트를 위해, 모든 규칙들이 적어도 최소 대응 횟수로 실행되도록, 및/또는 모든 규칙들이 최대 대응 횟수만큼 실행되도록 (예를 들어, 애플리케이션 내 완전한 코드 커버리지가 달성되도록), 애플리케이션 내 모든 규칙들이 실행되도록 입력 데이터가 충분히 제공될 수 있다.

[0117] 도 19를 참조하면, 일부 예에서, 입력 데이터 레코드(50)은 데이터 처리 애플리케이션(52)에 입력 데이터로서 제공될 수 있다. 데이터 처리 애플리케이션(52)은 하나 이상의 규칙(54)을 구현한다. 규칙은, 예를 들어, 데이터를 한 포맷에서 다른 포맷으로 변환하거나, 데이터 대한 결정을 하거나, 또는 입력 데이터 세트를 기반으로 새로운 데이터를 생성하는데 사용될 수 있는 기준 세트이다. 규칙에 대한 추가 설명은 2007년 4월 10일 출원된 미국 출원 일련번호 11/733,434호에 기재되어 있으며, 그 내용은 본 명세서에 참조로 포함된다.

[0118] 데이터 처리 애플리케이션(52)에 의해 구현된 각 규칙(54)은, 규칙에 대한 해당 조건문이 만족되면 실행될 수 있고, 해당 조건문이 만족되지 않으면 데이터 처리 애플리케이션(52)에 의해 실행되지 않을 수 있다. 규칙(54)은 적어도 조건문 및 실행문을 포함하는 사양에 의해 특정될 수 있다. 조건문이 만족될 때 (예를 들어, 조건문이 ‘참’으로 평가될 때), 실행문이 평가될 수 있다. 조건문은, 예를 들어 입력 데이터(50)에 해당하는 입력 변수 또는 하나 이상의 입력 변수들에 종속하는 파생 변수들일 수 있는, 하나 이상의 변수 값에 종속(예를 들어, 트리거)될 수 있다. 일부 예에서, 애플리케이션은 트리거되는 모든 규칙(54)을 실행한다. 일부 예에서, 애플리케이션은 규칙(54)들의 일부, 또는 단 하나 (예를 들어, 트리거된 제1규칙(54))와 같이 트리거된 모든 규칙(54)보다 적은 규칙을 실행한다. 그러한 데이터 처리 애플리케이션의 실행에 대한 추가 설명은 “데이터 레코드 선택”이라는 제목의 미국 출원번호 13/827,558호에 기재되어 있고, 그 내용은 본 명세서에 참조로 포함된다.

[0119] 그 규칙은 해당 조건문이 만족될 때 실행될 수 있는 임의의 규칙일 수 있다. 일례에서, 소매점 고객에 대한 마케팅 제안을 식별하는 데이터 처리 애플리케이션 내 규칙은 20,000달러 미만의 수입을 가진 고객에 대한 특정 마케팅 제안을 식별하는 규칙일 수 있다. 일례에서, 전화회사 고객에 대한 청구서를 생성하는 데이터 처리 애플리케이션 규칙은 지난 달 휴대폰 사용 시간이 고객 서비스 계획에서 허용된 휴대폰 시간을 초과한 고객의 청구서에 추가 요금을 적용하는 규칙일 수 있다.

[0120] 데이터 처리 애플리케이션(52)에 의해 생성된 실행 데이터(56)는 실행하거나 실행하지 않은 애플리케이션(52) 내 규칙(54), 애플리케이션(52) 내 각 규칙(54)이 실행된 횟수, 또는 다른 정보와 같은 데이터 처리 애플리케이션(52)의 실행에 대한 정보를 제공할 수 있다. 실행 데이터(56)는 데이터 생성 모듈(150)로 제공될 수 있다. 실행 데이터(56)를 기반으로, 데이터 생성 모듈(150, 도 2) 내 결과 엔진(40)은, 비실행(unexecuted) 규칙(54)이 실행되게 한 입력 데이터, 특정 규칙(54)을 특정된 횟수동안 실행되게 한 입력 데이터, 또는 다른 원하는 실행 결과를 야기한 입력 데이터와 같은 데이터 처리 애플리케이션(52)에 대해 원하는 입력 데이터 또는 입력 데이터 기준을 식별할 수 있다. 예를 들어, 원하는 실행 결과는 사용자에 의해, 예를 들어 사용자 인터페이스(30)를 통해 특정될 수 있다.

[0121] 데이터 생성 모듈(150)의 데이터 생성 엔진(20)은 데이터(58)를 생성하여 원하는 입력 데이터 기준을 만족시킨다. 생성된 데이터(58)는 데이터 처리 애플리케이션(52)에 입력 데이터로서 제공될 수 있다. 즉, 예를 들어, 데이터 생성 엔진(20)은 데이터를 생성하여 데이터 처리 애플리케이션 내에서 이전에 비실행된 규칙이 실행되게

하고, 특정한 규칙이 수회 실행되게 하거나 다른 원하는 실행 결과를 야기할 수 있다. 일례에서, 수입(income) 필드에 대해 1,000,000 달러보다 큰 값을 갖는 레코드가 생성되어 수입(income) > 1,000,000 달러를 갖는 입력 데이터에 대해서만 실행하는 규칙의 실행을 야기할 수 있다. 일례에서, 제1규칙은 성별(gender) 필드가 “F” 값을 가지면 실행되고, 제2규칙은 성별(gender) 필드가 “M” 값을 가지면 실행된다. 데이터 처리 애플리케이션의 초기 실행시, 실행 데이터(56)가 제2규칙이 50회 실행되는 동안 제1규칙이 단 한 번 실행되는 것을 나타낸다면, 데이터는 데이터 생성 엔진(20)에 의해 생성되어, 예를 들어, 사용자에 의해 특정된 바와 같이 원하는 실행 결과를 만족할 수 있어 제1 및 제2규칙은 거의 동일한 횟수로 실행된다. 예를 들어, 데이터 생성 엔진(20)은 성별(gender)=‘F’를 갖는 추가 레코드를 생성하거나 일부 기준 레코드의 성별(gender) 필드에 대한 값을 변경할 수 있다.

[0122] 도 20을 참조하면, 데이터 생성 모듈(150)에 의해 구현된 일반 접근(750 단계)은 데이터 처리 애플리케이션에 적합한 테스트 데이터 세트를 생성한다. 예를 들어, 테스트 데이터 세트는 데이터 처리 애플리케이션에 의해 처리되어, 원하는 횟수만큼 데이터 처리 애플리케이션의 각 규칙 실행과 같은 원하는 실행 결과를 만들어낼 수 있다. 접근에서(750단계) 하나 이상의 레코드들이 하나 이상의 규칙들을 구현하는 데이터 처리 애플리케이션에 위한 처리를 위해 제공된다(752단계). 각 레코드는 하나 이상의 필드를 가질 수 있다. 데이터 처리 애플리케이션이 실행될 수 있다 (754단계).

[0123] 각 규칙이 데이터 처리 애플리케이션에 의해 실행되는 횟수를 기반으로, 컨텐츠 기준이 하나 이상의 레코드의 특정 필드에 대해 데이터 생성 모듈(150)의 결과 엔진(40)에 의해 결정될 수 있다 (756단계). 일부 예에서, 데이터 처리 애플리케이션에 의해 실행될 때, 컨텐츠 기준에 따라 생성된 데이터가 원하는 실행 결과를 낼 수 있도록, 특정 필드에 대해, 이전에 실행되지 않은 규칙이 실행되게 하거나 특정 규칙이 수회 실행되게 하는 것과 같은, 컨텐츠 기준이 결정될 수 있다. 컨텐츠 기준은, 예를 들어, 필드에 대해 생성될 데이터에 대한 데이터 특성일 수 있다. 예를 들어, 소수 필드 또는 정수 필드와 같은 수치 필드에 대한 데이터 특성은 최대 허용치, 최소 허용치, 허용값 범위, 또는 모든 레코드에 걸친 필드에 대한 평균치를 나타낼 수 있다. 스트링 필드에 대한 데이터 특성은 최대 허용 문자 수, 최소 허용 문자 수, 평균 문자 수, 또는 허용 문자 세트를 나타낼 수 있다. 데이터 필드 또는 일시 필드와 같은 시간 필드에 대한 데이터 특성은 날짜 또는 일시 메트릭, 날짜 또는 일시 베이스, 또는 날짜 또는 일시 범위를 나타낼 수 있다.

[0124] 컨텐츠는 컨텐츠 기준을 기반으로 및 특정 필드에 대한 포맷 사양을 기반으로 특정 필드에 대해 생성될 수 있다 (758단계). 일부 예에서, 기존 레코드의 특정 필드에 대한 컨텐츠가 변경될 수 있다. 일부 예에서, 새로운 레코드가 생성되거나 새로운 레코드에서 특정 필드에 대한 컨텐츠가 생성될 수 있다.

[0125] 데이터 처리 애플리케이션은 새로 생성된 컨텐츠를 입력 데이터로 사용하여 실행될 수 있다(760단계). 일부 예에서, 특정 필드에 대해 변경된 값들을 포함하는 기존 레코드는 데이터 처리 애플리케이션으로 제공될 수 있다. 일부 예에서, 기존 레코드는 새로 생성된 레코드와 함께 데이터 처리 애플리케이션으로 제공될 수 있다. 일부 예에서, 새로 생성된 레코드만 데이터 처리 애플리케이션으로 제공될 수 있다.

[0126] 적어도 새로 생성된 컨텐츠를 입력 데이터로 사용한 데이터 처리 애플리케이션 실행 결과가 분석될 수 있다(762 단계). 데이터 처리 애플리케이션 내 규칙이 수회 실행되는 것과 같이 원하는 실행 결과가 나오면(764단계), 데이터 생성 과정이 완료된다(766단계). 원하는 실행 결과가 나오지 않으면(764단계), 추가 컨텐츠 기준이 결정될 수 있고(757단계), 새로운 컨텐츠가 다시 생성되어 (758단계) 데이터 처리 애플리케이션의 입력 데이터로 사용될 수 있다.

[0127] 일부 예에서, 접근(750 단계)이 구현되어 데이터 처리 애플리케이션 내 모든 규칙이 적어도 한번 실행되게 하는 데이터 처리 애플리케이션용 테스트 데이터 세트를 생성할 수 있다. 즉, 데이터 처리 애플리케이션을 철저하게 테스트하기 위해, 데이터 처리 애플리케이션에서 모든 가능한 동작을 트리거하는 데이터가 제공될 수 있다. 그러한 데이터가 존재하지 않으면, 해당 데이터가 생성될 수 있다. 예를 들어, 데이터 생성 과정이 완료되면(766 단계), 생성된 데이터 세트가 데이터 처리 애플리케이션에 의해 처리되어 데이터 처리 애플리케이션의 실행을 테스트할 수 있다(768단계). 생성된 데이터 세트가 데이터 처리 애플리케이션의 모든 가능한 동작이 실행되게 하므로, 데이터 처리 애플리케이션이 철저히 테스트될 수 있다. 데이터 처리 애플리케이션이 원하는 것처럼 수행되지 않는다면 (예를 들어, 데이터 처리 애플리케이션의 동작이 원하는 것처럼 수행되지 않는다면), 데이터 처리 애플리케이션은 적절하게 변경되어(770단계) 만족할만한 결과를 얻을 때까지 테스트될 수 있다.

[0128] 일부 예에서, 접근 (750단계)이 구현되어 데이터 처리 애플리케이션에 의한 처리를 위해 현실적인 데이터 세트가 생성될 수 있다. 예를 들어, 소매 회사용 데이터 처리 애플리케이션은 패턴을 구매한 고객의 나타내는 트렌

드 보고서를 생성할 수 있다. 고객의 프라이버시를 보호하기 위해, 현실에 맞지만 허구적인 데이터 레코드 세트가 데이터 처리 애플리케이션에 의해 처리를 위해 생성될 수 있다. 일부 예에서, 생성된 데이터는 민감한 개인 정보를 드러내지 않고도 데이터 세트의 레코드에 걸친 값의 분포 (예를 들어, 고객의 수입 분포)와 같은 특성을 실제의 고객 데이터와 공유할 수 있다. 일부 예에서, 현실적인 데이터 세트는 물리적인 과정을 나타내는 데이터, 금융 거래를 나타내는 데이터, 인간 또는 동물 행위를 나타내는 데이터, 또는 다른 데이터일 수 있다.

- [0129] 실시예 1은, 하나 이상의 레코드들이 처리되는 동안 하나 이상의 규칙들 각각이 데이터 처리 애플리케이션에 의해 실행되는 횟수를 기반으로 하나 이상의 필드들 각각에 대한 컨텐츠 기준을 결정하는 단계; 컨텐츠 기준을 기반으로 필드들 각각에 대한 컨텐츠를 생성하는 단계; 및 하나 이상의 레코드의 필드들 각각을 생성된 컨텐츠로 파풀레이팅하는 단계를 포함하는 방법에 관한 것이다.
- [0130] 실시예 2는 실시예 1에 관한 것으로, 컨텐츠를 생성하는 단계는 각 필드의 포맷 사양을 기반으로 컨텐츠를 생성하는 단계이다.
- [0131] 실시예 3은 실시예 1 또는 2에 관한 것으로, 포맷 사양은 컨텐츠의 타입을 나타낸다.
- [0132] 실시예 4는 실시예 3에 관한 것으로, 컨텐츠 타입은 스트링, 소수, 정수, 날짜, 또는 일시 중 하나 이상을 포함한다.
- [0133] 실시예 5는 실시예 2 내지 4중 어느 하나에 관한 것으로, 사용자 인터페이스를 통해 포맷 사양을 수신하는 것을 포함한다.
- [0134] 실시예 6은 선행 실시예들 중 어느 하나에 관한 것으로, 하나 이상의 필드를 포함하는 새로운 레코드를 생성하는 단계; 및 새로운 레코드의 필드를 파풀레이팅하는 단계를 포함한다.
- [0135] 실시예 7은 선행 실시예중 어느 하나에 관한 것으로, 각 필드를 생성된 컨텐츠로 파풀레이팅하는 단계는 하나 이상의 필드에 대한 기준 컨텐츠를 변경하는 단계를 포함한다.
- [0136] 실시예 8는 선행 실시예중 어느 하나에 관한 것으로, 각 필드를 파풀레이팅하는 단계는 하나 이상의 기준 레코드에 대한 특별한 필드를 생성하는 단계를 포함한다.
- [0137] 실시예 9는 실시예 8는 선행 실시예중 어느 하나에 관한 것으로, 컨텐츠 기준은 타겟 값, 타겟 값 범위, 타겟 값의 분포, 최대치, 또는 최소치 중 하나 이상을 포함한다.
- [0138] 실시예 10은 선행 실시예중 어느 하나에 관한 것으로, 컨텐츠 기준은 생성된 컨텐츠에 대한 특별한 값들을 포함한다.
- [0139] 실시예 11는 선행 실시예중 어느 하나에 관한 것으로, 컨텐츠 기준은 하나 이상의 필드 중 둘 이상에 포함될 타겟 컨텐츠들의 조합을 나타낸다.
- [0140] 실시예 12는 선행 실시예중 어느 하나에 관한 것으로, 컨텐츠 기준은 사용자 인터페이스를 통해 수신하는 단계를 포함한다.
- [0141] 실시예 13은 선행 실시예중 어느 하나에 관한 것으로, 컨텐츠를 생성하는 단계는 생성모드를 기반으로 컨텐츠를 생성하는 단계를 포함한다.
- [0142] 실시예 14는 실시예 13에 관한 것으로, 생성 모드는 (i) 다수의 레코드 각각에 대해 하나 이상의 필드에 대한 랜덤 컨텐츠 생성, (ii) 다수의 레코드 각각에 대해 하나 이상의 필드에 대한 고유 컨텐츠 생성, 및 (iii) 사용 가능한 컨텐츠 세트로부터 다수의 레코드 각각에 대해 하나 이상의 필드에 대한 컨텐츠 선택 중 하나 이상을 포함한다.
- [0143] 실시예 15는 선행 실시예중 어느 하나에 관한 것으로, 생성된 컨텐츠를 포함하는 하나 이상의 레코드를 데이터 처리 애플리케이션에 제공하는 단계를 포함한다.
- [0144] 실시예 16은 선행 실시예중 어느 하나에 관한 것으로, 컨텐츠 기준은 특정 규칙이 데이터 처리 애플리케이션에 의해 실행되도록 결정된다.
- [0145] 실시예 17은 선행 실시예중 어느 하나에 관한 것으로, 컨텐츠 기준을 결정하는 단계는 최소 임계 횟수 미만으로 실행된 특정 규칙을 식별하는 단계; 및 특정 규칙이 적어도 최소 임계 횟수만큼 실행되게 하는 컨텐츠 기준을

결정하는 단계를 포함한다.

- [0146] 실시예 18은 명령어들을 저장하는 비일시적인 컴퓨터 판독가능한 매체에 관한 것으로, 명령어들은 컴퓨터 시스템이 하나 이상의 레코드를 처리하는 동안 하나 이상의 규칙 각각이 데이터 처리 애플리케이션에 의해 실행되는 횟수를 기반으로 하나 이상의 필드 각각에 대해 컨텐츠 기준을 결정하고, 상기 컨텐츠 기준을 기반으로 상기 각 필드에 대한 컨텐츠를 생성하고; 및 하나 이상의 레코드에서 상기 각 필드를 상기 생성된 컨텐츠로 파풀레이팅하게 한다.
- [0147] 실시예 19는 적어도 하나의 프로세서를 포함하는 컴퓨팅 시스템에 관한 것으로, 적어도 하나의 프로세서는 하나 이상의 레코드를 처리하는 동안 하나 이상의 규칙 각각이 데이터 처리 애플리케이션에 의해 실행되는 횟수를 기반으로 하나 이상의 필드 각각에 대해 컨텐츠 기준을 결정하고, 상기 컨텐츠 기준을 기반으로 상기 각 필드에 대한 컨텐츠를 생성하고; 및 하나 이상의 레코드에서 상기 각 필드를 상기 생성된 컨텐츠로 파풀레이팅하도록 구성된다.
- [0148] 실시예 20은 컴퓨팅 시스템에 관한 것으로, 하나 이상의 레코드를 처리하는 동안 하나 이상의 규칙 각각이 데이터 처리 애플리케이션에 의해 실행되는 횟수를 기반으로 하나 이상의 필드 각각에 대해 컨텐츠 기준을 결정하는 수단; 상기 컨텐츠 기준을 기반으로 상기 각 필드에 대한 컨텐츠를 생성하는 수단; 및 하나 이상의 레코드에서 상기 각 필드를 상기 생성된 컨텐츠로 파풀레이팅하는 수단을 포함한다.
- [0149] 실시예 21은 데이터 처리 애플리케이션을 개발하는 동안 상기 데이터 처리 애플리케이션을 테스트하는 방법에 관한 것으로, 그 방법은 데이터 처리 애플리케이션을 사용하여 제1데이터 레코드를 처리하는 단계로서, 데이터 처리 애플리케이션은 하나 이상의 규칙을 구현하며, 제1데이터 레코드는 다수의 필드를 포함하는, 제1데이터 레코드 처리 단계; 제1데이터 레코드를 처리하는 동안 데이터 처리 애플리케이션의 각 규칙이 실행되는 횟수를 규칙에 대한 타겟 값과 비교하는 단계; 비교 결과를 기반으로, 데이터 처리 애플리케이션을 사용하여 처리될 제2데이터 레코드의 데이터에 대한 타겟 특성을 나타내는 컨텐츠 기준을 결정하는 단계; 제1데이터 레코드의 포맷에 따라 및 컨텐츠 기준에 따라 제2데이터 레코드의 하나 이상의 필드에 대한 컨텐츠를 생성하는 단계; 데이터 처리 애플리케이션을 사용하여 제2데이터 레코드를 처리하는 단계; 제2데이터 레코드를 처리하는 동안 데이터 처리 애플리케이션의 각 규칙이 실행되는 횟수를 규칙의 타겟 값과 비교하고 각 규칙에 대한 타겟 값이 만족하는지를 결정하는 단계; 및 제2데이터 레코드를 사용하여 데이터 처리 애플리케이션을 테스트하는 단계를 포함한다.
- [0150] 실시예 22는 실시예 21에 관한 것으로, 클라이언트 장치에서 제1데이터 레코드를 수신하는 단계를 포함하고, 데이터 처리 애플리케이션은 클라이언트 장치에서 실행된다.
- [0151] 실시예 23은 실시예 22에 관한 것으로, 클라이언트 장치는 컴퓨팅 장치를 포함한다.
- [0152] 실시예 24는 실시예 21 또는 22에 관한 것으로, 제거 서버로부터 제1데이터 레코드를 수신하는 단계를 포함한다.
- [0153] 실시예 25는 실시예 24에 관한 것으로, LAN 또는 WAN 연결을 통해 서버에 연결되는 클라이언트 장치에서 제1데이터 레코드를 수신하는 단계를 포함한다.
- [0154] 실시예 26은 실시예 21 내지 15에 관한 것으로, 데이터 처리 애플리케이션이 제2데이터 레코드를 처리할 때 제2데이터 레코드의 데이터가 각 규칙에 대한 타겟 값을 만족하도록 컨텐츠 기준을 결정하는 단계를 포함한다.
- [0155] 실시예 27은 실시예 21 내지 26에 관한 것으로, 제2데이터 레코드의 하나 이상의 필드들에 대한 컨텐츠를 생성하는 단계는 제1데이터 레코드의 하나 이상의 필드들의 컨텐츠를 변경하는 단계를 포함하고, 변경된 제1데이터 레코드는 제2데이터 레코드이다.
- [0156] 실시예 28은 실시예 21 내지 27에 관한 것으로, 제2데이터 레코드의 하나 이상의 필드에 대한 컨텐츠를 생성하는 단계는 새로운 데이터 레코드를 생성하는 단계를 포함한다.
- [0157] 실시예 29는 실시예 21 내지 28에 관한 것으로, 스토리지 장치에 제2데이터 레코드를 저장하는 단계를 포함한다.
- [0158] 실시예 30은 실시예 21내지 29에 관한 것으로, 테스트 결과를 기반으로 데이터 처리 애플리케이션을 변경하는 단계를 포함한다.
- [0159] 실시예 31은 실시예 21 내지 30에 관한 것으로, 데이터 처리 애플리케이션을 테스트하는 단계는 데이터 처리 애플리케이션을 테스트하는 단계를 포함한다.

플리케이션을 사용한 제2데이터 레코드 처리 결과를 기반으로 데이터 처리 애플리케이션을 변경하는 단계를 포함한다.

[0160] 실시예 32는 실시예 21 내지 31에 관한 것으로, 제2데이터 레코드의 데이터가 프로덕션 환경에서 데이터 처리 애플리케이션에 의해 처리될 것으로 예상되는 데이터와 실질적으로 유사하도록 컨텐츠 기준을 결정하는 단계를 포함한다.

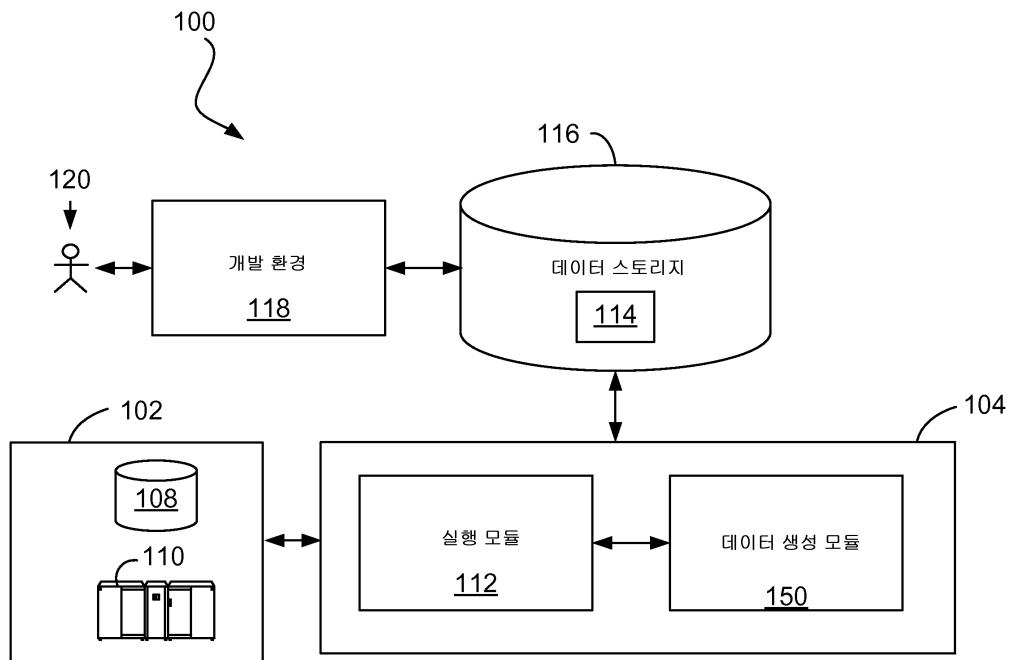
[0161] 상술한 데이터 생성 기술은 적절한 소프트웨어를 실행하는 컴퓨팅 시스템을 사용하여 구현될 수 있다. 예를 들어, 컴퓨팅 시스템은 시스템(100), 개발 환경(118), 데이터 생성 모듈(150) 및 다른 컴퓨팅 시스템을 포함할 수 있다. 예를 들어, 소프트웨어는 (분산, 클라이언트/서버, 또는 그리드(grid)와 같은 다양한 구조를 가질 수 있는) 하나 이상의 프로그램된 또는 프로그래머블 컴퓨팅 시스템에서 실행되는 하나 이상의 컴퓨터 프로그램 프로시저들을 포함하고, 각 컴퓨팅 시스템은 적어도 하나의 프로세서, 적어도 하나의 데이터 스토리지 시스템(휘발성 및/또는 비휘발성 메모리 및/또는 스토리지 구성요소를 포함), (적어도 하나의 입력 장치 또는 포트를 사용하여 입력을 수신하고, 적어도 하나의 출력 장치 또는 포트를 사용하여 출력을 제공하는) 적어도 하나의 사용자 인터페이스를 포함한다. 소프트웨어는, 예를 들어, 설계, 구성, 및 데이터 플로우 그래프의 실행과 관련된 서비스를 제공하는 하나 이상의 대형 프로그램 모듈을 포함할 수 있다. 프로그램 모듈들 (예를 들어, 데이터 플로우 그래프의 구성요소들)은 데이터 구조 또는 데이터 저장소에 저장된 데이터 모델에 따른 조직화된 데이터로 구현될 수 있다.

[0162] 소프트웨어는 CD-ROM 또는 다른 컴퓨터로 판독가능한 (예를 들어 범용 또는 특정 컴퓨팅 시스템 또는 장치에 의해 판독가능한) 매체와 같은 실재적이고 비일시적인 매체로 제공되거나, (예를 들어 전파 신호로 부호화되어) 네트워크의 통신 매체를 통해 실행될 컴퓨팅 시스템의 실재적이고 비일시적인 매체로 전달될 수 있다. 일부 또는 모든 처리는 특정 컴퓨터에서 또는 코프로세서(coprocessor) 또는 필드 프로그래밍 지원 게이트 어레이(FPGA) 또는 전용, 주문형 집적회로(ASIC)와 같은 특정 하드웨어를 사용하여 수행될 수 있다. 그 처리는 소프트웨어에 의해 특정된 다른 계산 부분이 다른 컴퓨팅 요소에 의해 수행되는 분산 방식으로 구현될 수 있다. 스토리지 장치의 매체가 컴퓨터에 의해 판독되어 본 명세서에서 기술된 처리를 수행할 때, 그러한 컴퓨터 프로그램 각각은 범용 또는 특정 목적의 프로그래머블 컴퓨터에 의해 접근가능한 스토리지 장치의 컴퓨터로 판독가능한 스토리지 매체(예를 들어, 고체 상태 메모리 또는 매체, 자기 또는 광 매체)에 저장되거나 다운로드되어 컴퓨터를 구성 및 운영한다. 본 발명의 시스템은 또한 컴퓨터 프로그램으로 구성된 실재적이고 비일시적인 매체로 구현된 것으로 간주되고, 이렇게 구성된 매체는 컴퓨터가 특정되고 미리 정의된 방식으로 동작하게 하여 본 명세서에서 기술된 하나 이상의 처리 단계들을 수행한다.

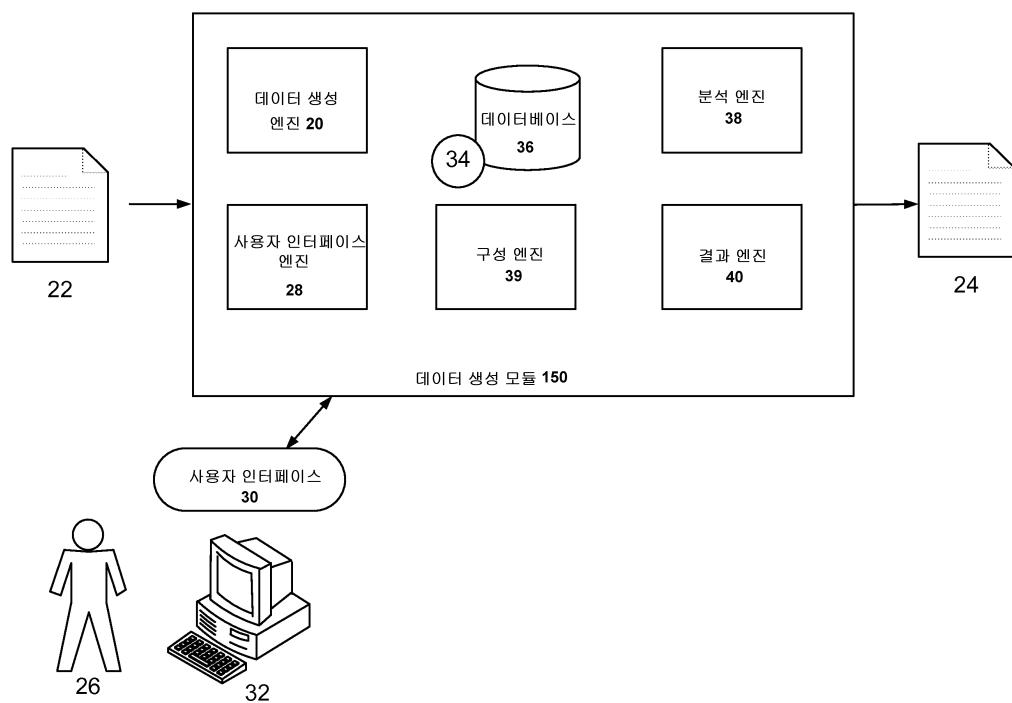
[0163] 본 발명의 다수의 실시예들이 기술되었다. 그럼에도, 상술한 설명은 다음의 청구범위에 의해 정의되는 본 발명의 범위를 설명하려는 것으로 본 발명의 범위를 제한하려는 것은 아니다. 따라서 다른 실시예들 또한 다음의 청구범위 내에 있다. 예를 들어, 다양한 실시예들이 본 발명의 범위를 벗어나지 않고 구현될 수 있다. 또한 상술한 일부 단계들은 독립된 순서를 가질 수 있고 따라서 상술한 바와 다른 순서로 수행될 수 있다.

도면

도면1



도면2



도면3a

200 ↗ \sum^{204} \sum^{206} \sum^{208} \sum^{210}

cust_ID	name	cust_age	state
21324	Jane Smith	28	MA
45621	Harry White	45	NY
21354	Sarah Chen	31	MA
56324	Sally Harris	34	RI
54326	Mary Jones	38	CT
43215	Mike Johnson	29	MA

도면3b

220 ↗ \sum^{204} \sum^{206} \sum^{208} \sum^{210}

cust_ID	name	cust_age	state
21324	Jane Smith	28	MA
45621	Harry White	25	NY
21354	Sarah Chen	31	MA
56324	Sally Harris	34	RI
54326	Mary Jones	38	CT
43215	Mike Johnson	29	MA

222 ↗ \sum^{222a}

도면3c

230 ↗

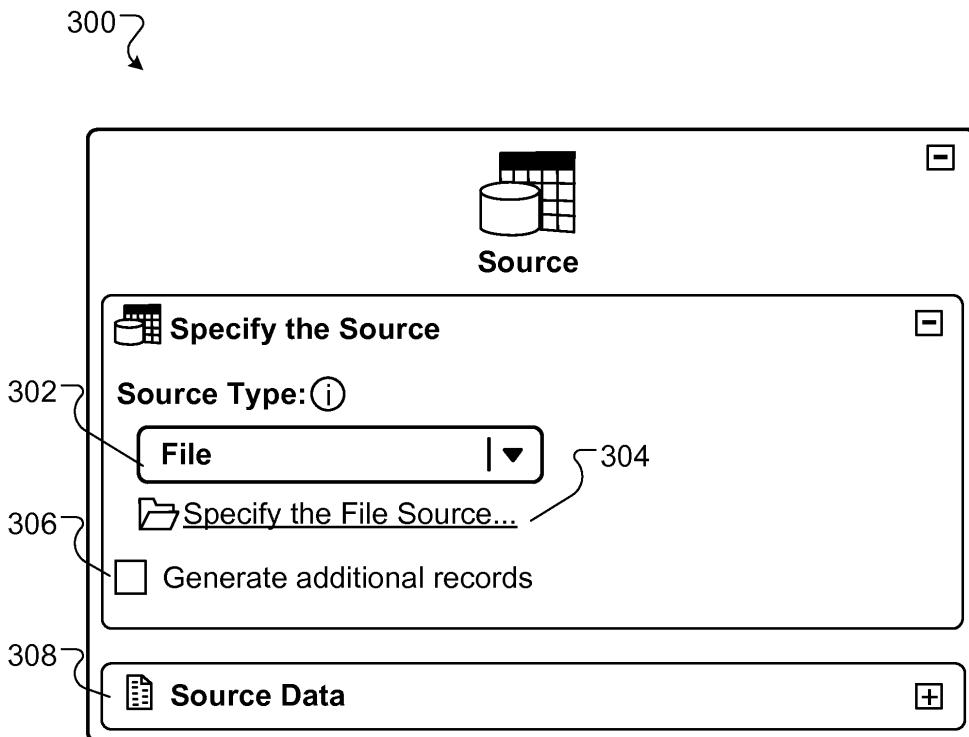
		\sum^{204}	\sum^{206}	\sum^{208}	\sum^{210}	\sum^{234}	\sum^{236}
232 ↘	cust_ID	name	cust_age	state	Age_range	Cust_income	
	21324	Jane Smith	19	MA	18-25	\$20,467	
	45621	Harry White	45	NY	40-55	\$45,326	
	21354	Sarah Chen	31	MA	25-40	\$60,245	
	56324	Sally Harris	34	RI	25-40	\$98,760	
	54326	Mary Jones	38	CT	25-40	\$85,198	
	43215	Mike Johnson	29	MA	25-40	\$70,870	

도면3d

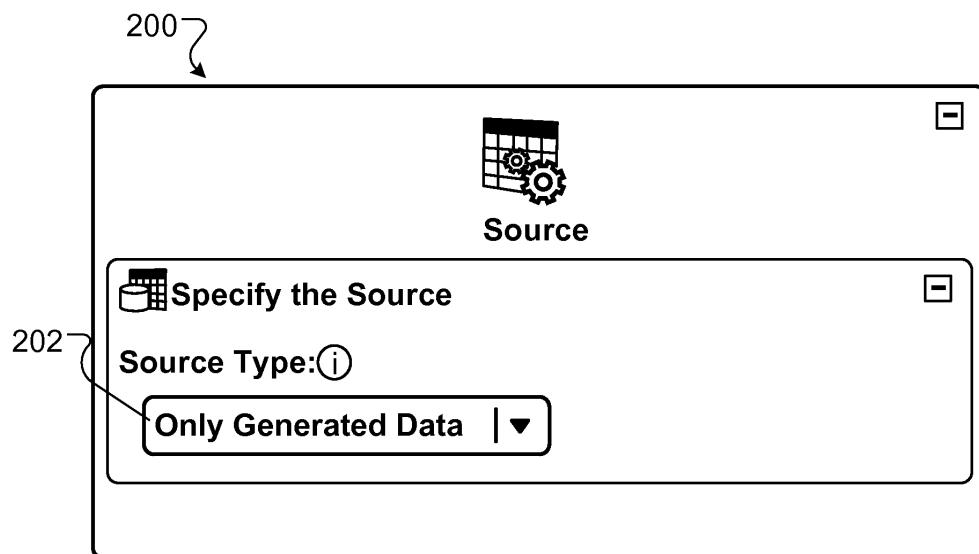
240 ↗

	\sum^{204}	\sum^{206}	\sum^{208}	\sum^{210}
202 ↘	cust_ID	name	cust_age	state
	21324	Jane Smith	28	MA
	45621	Harry White	45	NY
	21354	Sarah Chen	31	MA
	56324	Sally Harris	34	RI
	54326	Mary Jones	38	CT
	43215	Mike Johnson	29	MA
	54321	Erin Davis	48	MA
	56532	Gina Simons	60	NY
	32423	Beth Philips	55	CA
242 ↘	43218	David O'Brien	51	NH

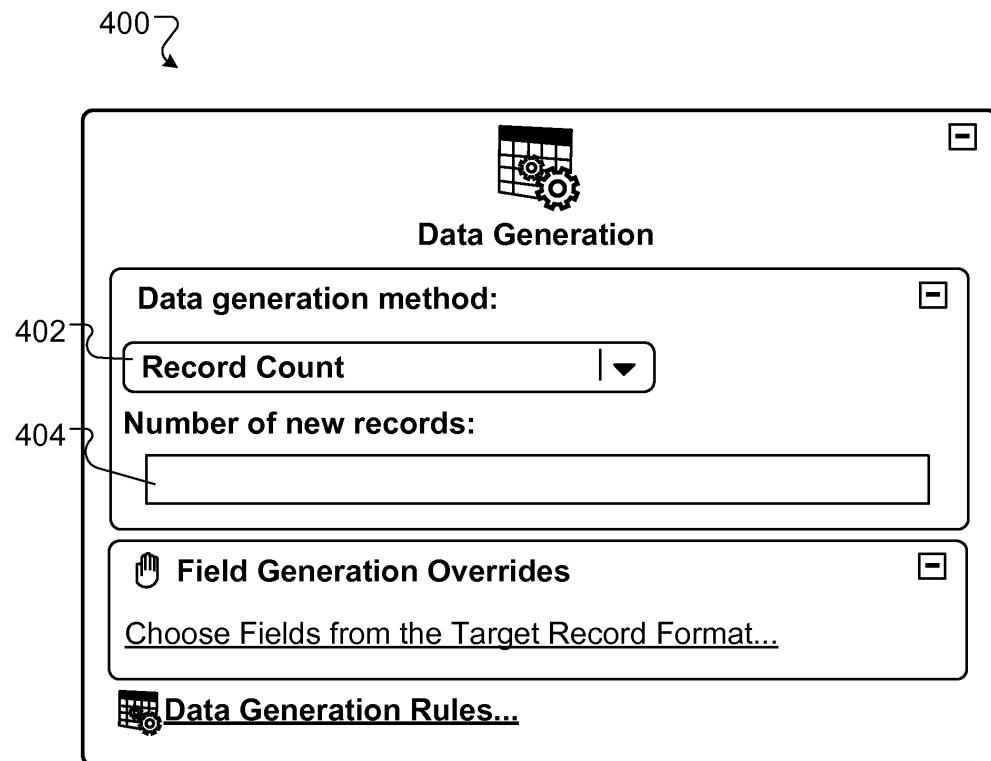
도면4



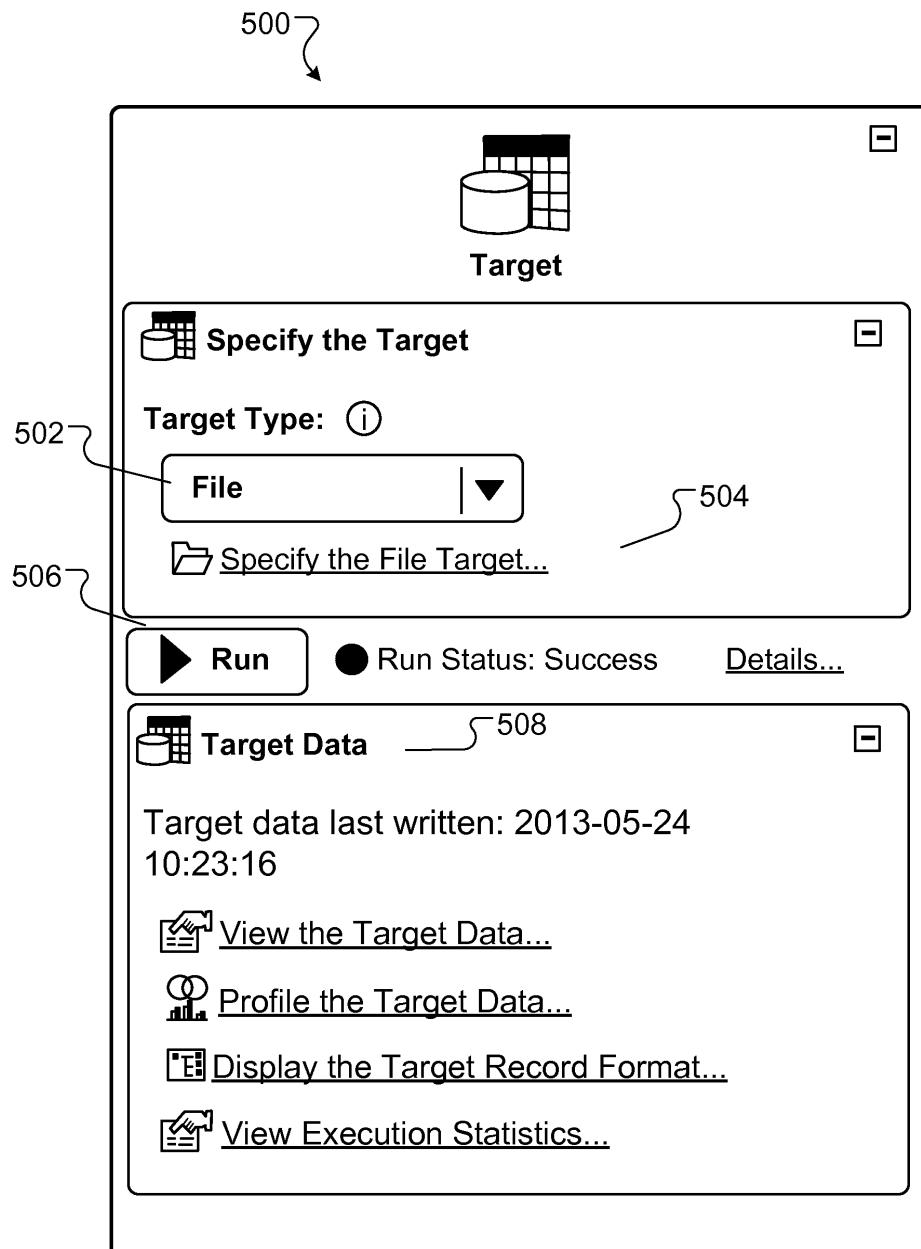
도면5a



도면5b



도면6



도면7

600 ↗

Existing Data File Details

602 **Existing Data File Path:** [Text Box] **Browse...**

604 **Existing Data File DML:** [Text Box] **Browse...**

Deduplicate the existing data file by key

606 **Number of records to generate per input record in the existing dataset
(specify as <min> to <max>, or <exact? to <exact>)** [Text Box] ⓘ

608 **Nonproducing percentage:** ⓘ [Text Box] 0

OK **Cancel**

도면8

700 ↗

Target Record Format

Notes:
Fields in the target record format but not in the source record format are marked with double asterisks (**).
Fields in the source record format but not in the target record format are not shown.

Ordered List:

Input Items:	selected items:
<input type="text" value="Filter..."/> <ul style="list-style-type: none"> <input type="checkbox"/> custid (decimal) 702 <input type="checkbox"/> fname (string) <input type="checkbox"/> lname (string) <input type="checkbox"/> place (string) <input type="checkbox"/> building_num (decimal) <input type="checkbox"/> preferred (decimal) <input type="checkbox"/> newline (string) <input type="checkbox"/> last_transaction (datetime) ** <input type="checkbox"/> customer_since (date) ** 	<ul style="list-style-type: none"> <input type="checkbox"/> custid (decimal) 704 <input type="checkbox"/> fname (string) <input type="checkbox"/> lname (string) <input type="checkbox"/> building_num (decimal)
<input type="button" value="Add"/> <input type="button" value="Remove"/> <input type="button" value="Remove All"/> <input type="button" value="Move Up"/> <input type="button" value="Move Down"/>	
<input type="button" value="Apply"/>	
OK Cancel	

도면9

800

Optional generation characteristics

802 Average string length (up to 10 characters) ⓘ
8

804 Maximum string length (up to 10 characters) ⓘ
10

806 Allowable characters (default is a-z, A-Z and 0-9) ⓘ
[empty]

808 Percentage of NULL values (default is 0) ⓘ
0

도면10

900

Optional generation characteristics

The allowed value range is: -999999999999 to 999999999999 ⓘ 902
0 to 1000

Percentage of NULL values (default is 0%) ⓘ 904
0

도면11

180 ↗

Optional generation characteristics

The allowed value range is: -2147483648 to 2147483647: ⓘ

0 to 127

182 ↗

도면12

190 ↗

Optional generation characteristics

Date Range Metric: 192

TODAY

194 ↗ Date basis (Form is YYYYMMDD) ⓘ

-7 to -1

196 ↗ Range is -7 to -1 Months) ⓘ

Percentage of blank values (default is 1%) ⓘ

도면13

250 ↗

Default datetime characteristics	
252 ↗	Datetime Range Metrid: Days ▼
254 ↗	Datetime Basis (Use NOW for current datetime) ⓘ NOW
256 ↗	Datetime Range (Default is '-7 to 0' Days) ⓘ -7 to 0
258 ↗	Percentage of blank or NULL values (default is 0%) ⓘ 0

도면14a

350 ↗

state_cd	
Specify the generation method for state_cd:	List of Values ▼
A comma-separated list of values Spaces are allowed:	
MA,CT,RI,VT,ME,NH	

도면14b

354 ↗

		Filter Output Name... (X)		
		Output/Internal Name		Expression / Rule
1	employee_id (0)	employee_id	in.	in.employee_id
2	firstname (g5ufrBr)	firstname	in.	in.firstname
3	lastname (7yYMkDV3)	lastname	in.	in.lastname
4	state_cd (CT)	state_cd	in.	in.state_cd
5	state_name (QutzptJg)	state_name	Compute state_name	Compute state_name

356 ↗ 358 ↗

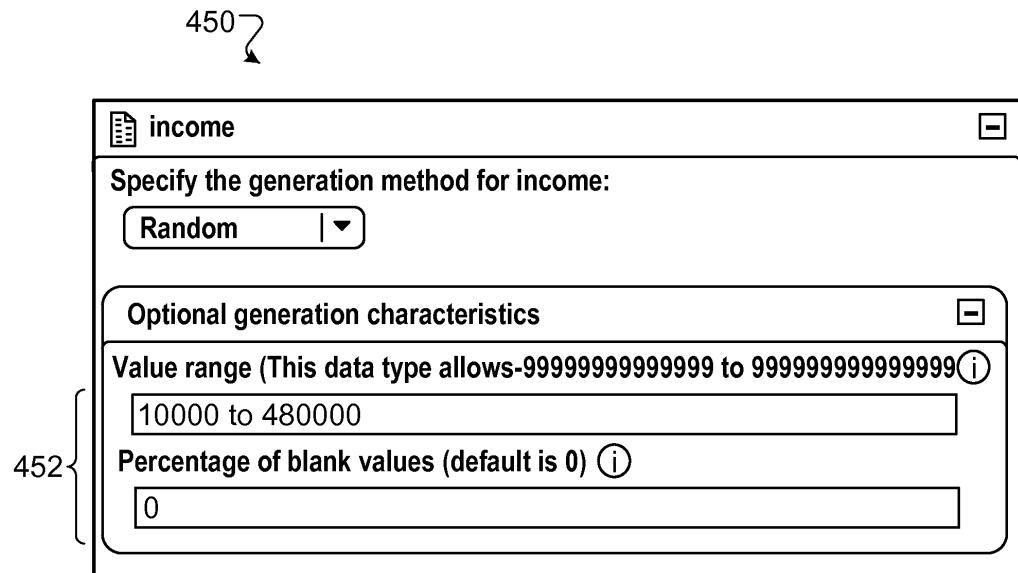
도면 14c

360 ↘

Rule: Compute state_name

	Trigger (Only the first true case will fire)	Output
	▷ state_cd (CT)	▷ state_name (Connecticut)
1	“CT”	“Connecticut”
2	“MA”	“Massachusetts”
3	“RI”	“Rhode Island”
4	“ME”	“Main”
5	“VT”	“Vermont”
6	“NH”	“New Hampshire”
*	any	Auto Map: state_name

도면15a



도면15b

454 ↗

Filter Output Name... (X) |

	Output/Internal Name		Expression / Rule
1	employee_id		in.employee_id
2	firstname		in.firstname
3	lastname		in.lastname
4	state_cd		in.state_cd
5	state_name		Compute state_name
6	income		in.income
7	tax_rate	→	Compute tax_rate

456 ↘ 458 ↘

도면 15c

460 ↘

Rule: Compute state_rate

	Trigger (Only the first true case will fire)	Output
	▷ income	▷ tax_rate
1	< 8925	10
2	< 36250	15
3	< 87850	25
4	< 183250	28
5	< 398350	33
6	< 400000	35
*	any	Default Variable Expression: 39.6

462 ↗

도면16

550

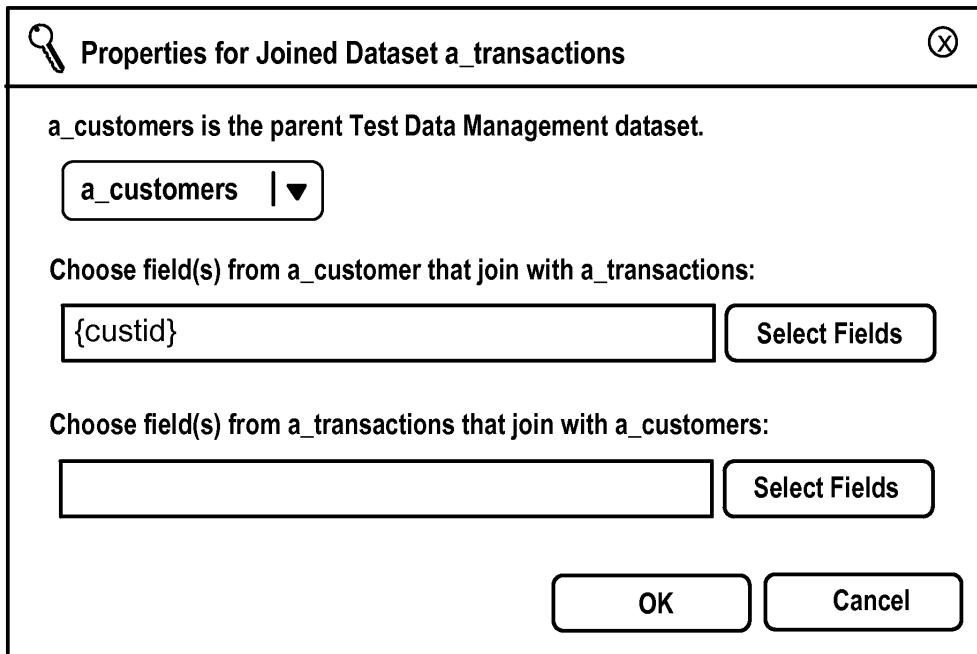
Inputs		Filter Output Name...	(X)	552
	Output/Internal Name	Expression / Rule	Computed Value	
12	employee_id (1)	in.employee_id	1	
A	firstname (agFMdsxa)	firstname + " " lastname	Jackie Norris	
A	state_cd (RI)	in.state_cd	RI	
A	state_name (Rhode Island)	in.state_name	Rhode Island	
12	income (460983)	in.income	460983	
12	tax_rate (39.6)	in.tax_rate	39.6	
A	ssn (573-05-3583)	in.ssn	573-05-3583	
▼ E _xx_source_dml		* + Add new Internal		
556	employee_id (1)			
A	firstname (Jackie)			
A	lastname (Norris)			
558	state_cd (RI)			
A	state_name (Rhode Island)			
12	income (460983)			
12	tax_rate (39.6)			
A	ssn (573-05-3583)			

도면 17a

Choose field(s) from a_customers that join with a_transactions

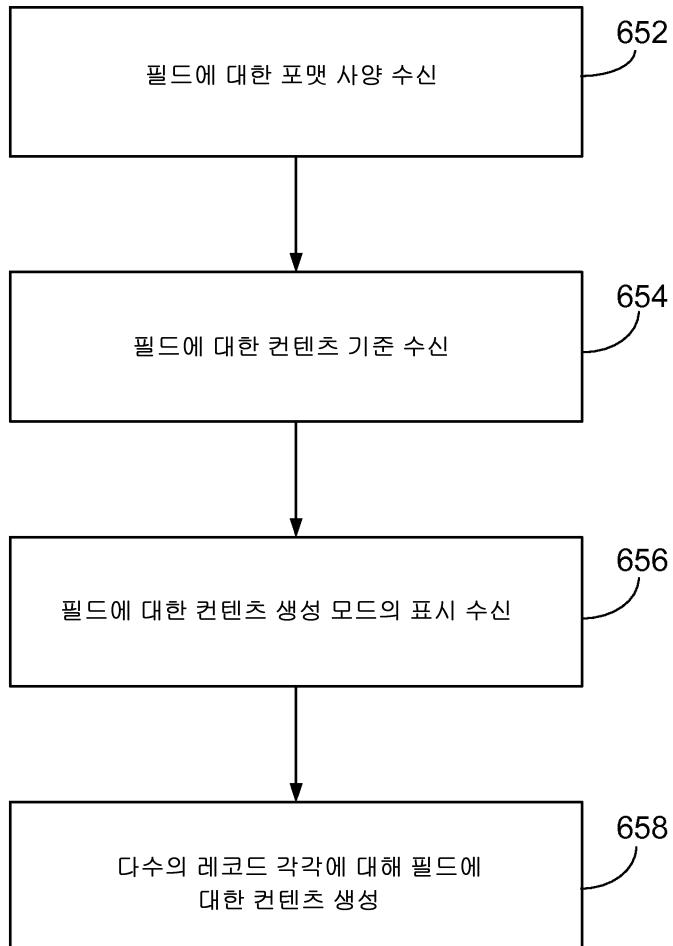
Record description:	Selected fields:
<input type="text" value="Filter..."/> X <ul style="list-style-type: none"> <input checked="" type="checkbox"/> custid <input type="checkbox"/> fname <input type="checkbox"/> lname <input type="checkbox"/> place <input checked="" type="checkbox"/> building_num <input checked="" type="checkbox"/> preferred <input type="checkbox"/> newline 	Selected fields: <ul style="list-style-type: none"> <input checked="" type="checkbox"/> custid
	<div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <input type="button" value="Add>"/> <input type="button" value="<Remove"/> <input type="button" value="<<Remove All"/> <input type="button" value="Move Up"/> <input type="button" value="Move Down"/> </div>

도면17b

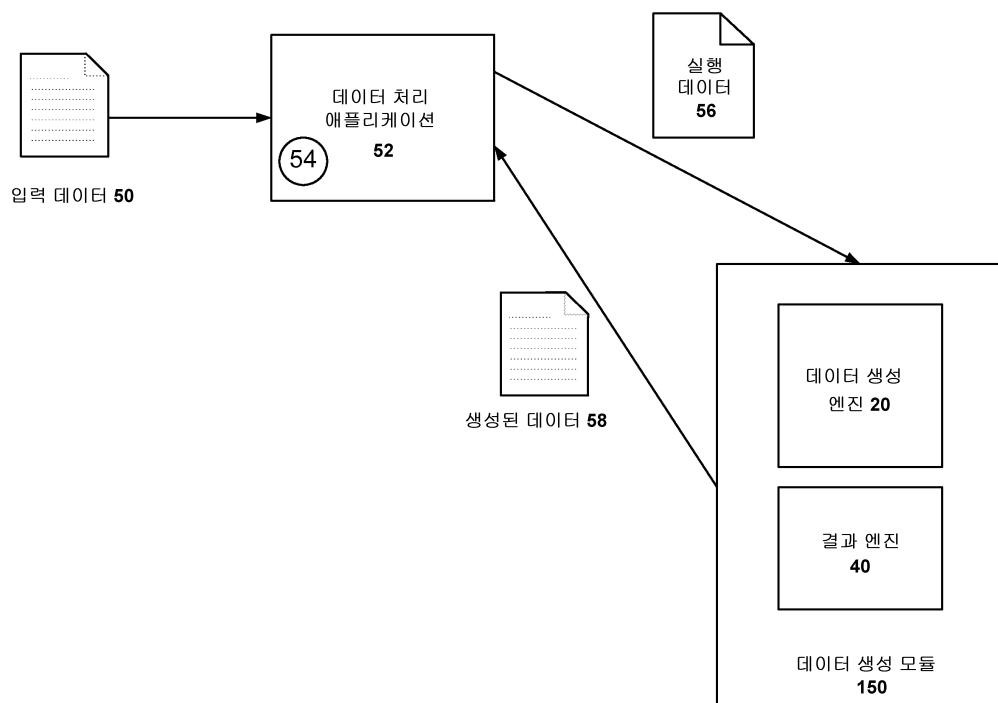


도면18

650



도면19



도면20

