

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6306599号
(P6306599)

(45) 発行日 平成30年4月4日(2018.4.4)

(24) 登録日 平成30年3月16日(2018.3.16)

(51) Int.Cl.

F 1

G06F 17/30 (2006.01)
G06F 12/00 (2006.01)G06F 17/30 220Z
G06F 17/30 230Z
G06F 12/00 513Z

請求項の数 13 (全 43 頁)

(21) 出願番号	特願2015-537697 (P2015-537697)
(86) (22) 出願日	平成25年8月2日(2013.8.2)
(65) 公表番号	特表2015-537294 (P2015-537294A)
(43) 公表日	平成27年12月24日(2015.12.24)
(86) 国際出願番号	PCT/US2013/053373
(87) 国際公開番号	W02014/062278
(87) 国際公開日	平成26年4月24日(2014.4.24)
審査請求日	平成28年8月1日(2016.8.1)
(31) 優先権主張番号	61/715,428
(32) 優先日	平成24年10月18日(2012.10.18)
(33) 優先権主張国	米国(US)
(31) 優先権主張番号	13/804,049
(32) 優先日	平成25年3月14日(2013.3.14)
(33) 優先権主張国	米国(US)

(73) 特許権者	502303739 オラクル・インターナショナル・コーポレーション
(74) 代理人	110001195 特許業務法人深見特許事務所
(72) 発明者	コジナ, ルドミラ アメリカ合衆国、O 2 1 3 8 マサチューセツ州、ケンブリッジ、ホルワースー・ストリート、37

最終頁に続く

(54) 【発明の名称】関連情報伝播システム

(57) 【特許請求の範囲】

【請求項 1】

データ要素の関連情報を表示するためにコンピュータで実行される方法であつて、
1つ以上のデータテーブルを備えるデータウェアハウスのデータ系統をたどることを備え、前記データ系統は、複数のデータ要素を含み、
前記複数のデータ要素のうちの一次データ要素と、前記一次データ要素に関連付けられている関連オブジェクトとの間の関連付けを表わす関連オブジェクトマッピングシステムテーブル内に第1の関連オブジェクトマッピングシステムレコードを格納することを備え、前記一次データ要素と前記関連オブジェクトとの間の前記関連付けは一次リンクとして同定され、

前記データ系統を用いて、前記複数のデータ要素のうちの二次データ要素と前記関連オブジェクトとの間の関連付けを同定することを備え、前記二次データ要素は前記データ系統内の前記一次データ要素に関連付けられており、

前記データ系統の前記二次データ要素と前記関連オブジェクトとの間の前記関連付けを表わし、かつ前記一次データ要素と前記二次データ要素との間の関連付けを表わす前記関連オブジェクトマッピングシステムテーブル内に第2の関連オブジェクトマッピングシステムレコードを格納することを備え、前記二次データ要素と前記関連オブジェクトとの間の前記関連付けは二次リンクとして同定される、方法。

【請求項 2】

各々のデータテーブルは1つ以上のデータレコードを含み、

前記一次データ要素は一次データレコードの列を含み、
前記関連オブジェクトは、データ要素以外の情報として当該データ要素に関連付けられる関連情報を含み。

前記第1の関連オブジェクトマッピングシステムレコードは、前記一次データ要素のアイデンティティ、前記関連オブジェクトのアイデンティティ、および一次リンクインジケータを含み。

前記二次データ要素は二次データレコードの列を含み。

前記第2の関連オブジェクトマッピングシステムレコードは、前記二次データ要素のアイデンティティ、前記関連オブジェクトのアイデンティティ、および二次リンクインジケータを含む、請求項1に記載のコンピュータで実行される方法。 10

【請求項3】

ユーザインターフェイス内に前記一次データ要素または前記二次データ要素を表示することと、

前記一次データ要素または前記二次データ要素のいずれかを選択することと、

前記第1の関連オブジェクトマッピングシステムレコードまたは前記第2の関連オブジェクトマッピングシステムレコードのいずれかを用いて前記ユーザインターフェイス内に前記関連オブジェクトを表示することとをさらに備える、請求項1または2に記載のコンピュータで実行される方法。

【請求項4】

前記関連オブジェクトの前記表示は関連オブジェクト種類によってフィルタされる、請求項3に記載のコンピュータで実行される方法。 20

【請求項5】

前記第1の関連オブジェクトマッピングシステムレコードは、前記一次データ要素のデータモデルを同定するデータ要素モデルアイデンティティ、前記一次データ要素のデータテーブルを同定するデータ要素テーブルアイデンティティ、および前記一次データ要素の列を同定するデータ要素列アイデンティティを含み。

前記第2の関連オブジェクトマッピングシステムレコードは、前記二次データ要素のデータモデルを同定するデータ要素モデルアイデンティティ、前記二次データ要素のデータテーブルを同定するデータ要素テーブルアイデンティティ、および前記二次データ要素の列を同定するデータ要素列アイデンティティを含む、請求項1～4のいずれかに記載のコンピュータで実行される方法。 30

【請求項6】

前記関連オブジェクトは、不一致、コメント、ステータスフラグ、状態インジケータ、タグ、画像、録音録画、ファイル、リンク、または文書のうち1つである、請求項1～5のいずれかに記載のコンピュータで実行される方法。

【請求項7】

前記二次データ要素は前記一次データ要素のソースデータ要素である、請求項1～6のいずれかに記載のコンピュータで実行される方法。

【請求項8】

前記第2の関連オブジェクトマッピングシステムレコードは前記二次データ要素が好ましいソースデータ要素であるという指標を含む、請求項7に記載のコンピュータで実行される方法。 40

【請求項9】

前記一次データ要素の前記アイデンティティおよび前記二次データ要素の前記アイデンティティは各々サロゲートキーである、請求項1～8のいずれかに記載のコンピュータで実行される方法。

【請求項10】

前記一次データ要素は前記一次データレコードを含み、

前記二次データ要素は前記二次データレコードを含む、請求項2に記載のコンピュータで実行される方法。 50

【請求項 1 1】

前記一次データ要素を削除することと、
前記第 1 の関連オブジェクトマッピングシステムレコードを削除することと、
前記第 2 の関連オブジェクトマッピングシステムレコードを削除することと、
前記関連オブジェクトの状態を閉じた状態に修正することとをさらに備え、前記閉じた
状態は、前記関連オブジェクトへのリンクを有さない、請求項 1 ~ 10 のいずれかに記載
のコンピュータで実行される方法。

【請求項 1 2】

請求項 1 ~ 11 のいずれかに記載の方法をコンピュータに実行させる、プログラム。

【請求項 1 3】

請求項 1 2 に記載のプログラムを格納したメモリと、

前記メモリに結合され、前記プログラムを実行するためのプロセッサとを備える、シス
テム。

【発明の詳細な説明】**【技術分野】****【0 0 0 1】**

関連出願への相互参照

この出願は、2012年10月18日に出願された米国仮特許出願連続番号第 61 / 7
15,428 号の優先権を主張し、その主題がここに引用により援用される。

【0 0 0 2】

分野

1 つの実施形態はコンピュータシステムに向けられ、より特定的にはデータを管理する
コンピュータシステムに向けられる。

【背景技術】**【0 0 0 3】**

背景

報告および分析に用いられるデータを格納するのにデータウェアハウスを用いることができる。コンピュータシステムは、複数の分散されたソースからデータを集め、データモデル中の集積データをデータウェアハウス内に格納することができる。ユーザは、格納されたデータに基づくデータ検討、データ分析、およびデータマイニングの準備のために変換をデータに適用することができる。さらに、データを、コメント、エラー報告、タグ、およびフラグ / 状態情報などの関連情報としばしば関連付けることができる。

【発明の概要】**【課題を解決するための手段】****【0 0 0 4】**

要約

1 つの実施形態は、関連情報の可視性を伝播させるシステムである。システムは、1 つ
以上のデータテーブルを含むデータウェアハウスのデータ系統を追跡する。システムは、
データ系統の一次データ要素と関連オブジェクトとの間の関連付けを表わす関連オブジェ
クトマッピングシステムテーブル内に第 1 の関連オブジェクトマッピングシステム記録を
さらに格納し、一次データ要素と関連オブジェクトとの間の関連付けは一次リンクとして
同定される。システムは、データ系統を用いて、二次データ要素と関連オブジェクトとの
間の関連付けをさらに同定し、二次データ要素はデータ系統内の一次データ要素に関する
。システムは、データ系統の二次データ要素と関連オブジェクトとの間の関連付けを表わ
し、かつ一次データ要素と二次データ要素との間の関連付けを表わす関連オブジェクトマ
ッピングシステムテーブル内に第 2 の関連オブジェクトマッピングシステム記録をさらに
格納し、二次データ要素と関連オブジェクトとの間の関連付けは二次リンクとして同定さ
れる。

【0 0 0 5】

添付の図面と関連して捉えられるべき好ましい実施形態の以下の詳細な説明から、一層

10

20

30

40

50

の実施形態、詳細、利点、および修正例が明らかになるであろう。

【図面の簡単な説明】

【0006】

【図1】発明の実施形態を実現することができるシステムのブロック図である。

【図2】発明の実施形態に従う、システムのための論理データモデルのブロック図である。

【図3】発明の実施形態に従う、例示的なデータ系統のブロック図である。

【図4】発明の実施形態に従う、データ系統を表示する例示的なユーザインターフェイスの図である。

【図5】発明の実施形態に従う、システムについての物理的モデルのブロック図である。 10

【図6】発明の実施形態に従う、データ系統マッピングテーブルのブロック図である。

【図7】発明の実施形態に従う、シャドウテーブルのブロック図である。

【図8】発明の実施形態に従う、データ系統サブシステムの物理的モデルのブロック図である。

【図9】発明の実施形態に従う、関連オブジェクトマッピングテーブルのブロック図である。

【図10】発明の実施形態に従う、関連情報サブシステムの物理的モデルのブロック図である。

【図11】発明の実施形態に従う、統合(union)変換の例示的なデータ系統追跡の図である。 20

【図12】発明の実施形態に従う、ピボット変換の例示的なソースデータ系統追跡の図である。

【図13】発明の実施形態に従う、ピボット変換の例示的な目標データ系統追跡の図である。

【図14】発明の実施形態に従う、アンピボット変換の例示的なソースデータ系統追跡の図である。

【図15】発明の実施形態に従う、アンピボット変換の例示的な目標データ系統追跡の図である。

【図16】発明の実施形態に従う、合体(join)変換の例示的なデータ系統追跡の図である。 30

【図17】発明の実施形態に従う、ケース変換の例示的なデータ系統追跡の図である。

【図18】発明の実施形態に従う、データ系統モジュールの機能性のフロー図である。

【図19】発明の実施形態に従う、関連情報伝播モジュールの機能性のフロー図である。

【発明を実施するための形態】

【0007】

詳細な説明

1つの実施形態では、データウェアハウス内に格納することができる1組のデータに対して施される1つ以上の変換の道筋を通したデータ系統追跡を提供することができるシステムが提供され、1組のデータは1つ以上のテーブル内に格納され、各々のテーブルは1つ以上のデータ記録を含む。「変換」または「データ変換」は、第1の形式の1つ以上のデータ記録を第2の形式の1つ以上のデータ記録に転換することであり、データ記録は1つ以上のデータ値を含むことができる。所与の目標データ要素について、システムは、目標データ要素を発生した1組のソースデータ要素および1組のソースデータ要素から目標データ要素への経路を同定することができ、「データ要素」は、データ記録の列の中に格納されるデータ値を含む。同様に、所与のソースデータ要素について、システムは、ソースデータ要素から導出される正確な組の目標データ要素およびソースデータ要素から目標データ要素の各々への経路を同定することができる。このように、システムは、データウェアハウス内の格納されたデータに加えて、データ系統追跡情報を与えて、格納されたデータの徹底的な分析をさらに容易にすることができます。ある実施形態では、データウェアハウス内の格納されたデータを1つ以上の「データ記録」内に格納することができ、データ 40

系統追跡情報を1つ以上の「システム記録」内に格納することができる。ここで定義されるように、「データ記録」は、データウェアハウスのデータを格納する任意の種類の記録であることができ、「システム記録」は、データ系統追跡情報を格納する任意の種類の記録であることができる。

【0008】

当該実施形態に従うと、ソースデータ要素から目標データ要素へおよびその逆にデータ系統を追跡するために、システムは、各々の変換毎に以下のデータ系統追跡情報を格納することができる：すなわち、(a) 1つ以上のソーステーブル定義中の1つ以上の列から、変換によって行なわれるデータ再編成に関するメタデータ情報である「列マッピングデータ」として同定される変換の目標テーブル中の各々の列への1つ以上のメタデータマッピング；(b) 1つ以上のソーステーブル中の1つ以上の実際のデータ記録から、「記録マッピングデータ」と同定される変換の実行によってポピュレートされる目標テーブル中の各々の実際のデータ記録への1つ以上のマッピング；および(c) 変換の種類などの、変換に特定的な付加的なデータ系統追跡情報、である。変換が定義されるか、または既存の変換プログラムの挙動を記述するように変換を格納することができる場合に、列マッピングデータを格納することができる。そのような変換列マッピングデータは、(a) ソーステーブルおよびソース列、(b) 目標テーブルおよび目標列、(c) 変換の種類、ならびに(d) 特定的な変換に必要な付加的な情報を含むことができる。変換によって処理されるデータ記録と関連付けられる記録マッピングデータは、変換が実行されるときに格納することができる。記録マッピングデータは、a) ソースデータ記録のアイデンティティ、b) 目標データ記録のアイデンティティ、およびc) 特定的な変換のための付加的な情報を含むことができる。データ記録のアイデンティティは、単一の列であることができるか、または、主キーもしくは固有キーとしても公知である、データテーブル内のデータ記録をその値が一意に同定する複数の列を含むことができる。当該実施形態に従うと、変換の目標テーブル中の各々のデータ記録は、その目標データ記録のための変換で1つ以上の列についてデータを与えたソースデータ記録のアイデンティティを格納することができ、このアイデンティティは「ソースキー」として同定される。ある実施形態では、これらの同定値を、データ記録を同定するのに用いることができる单一の値または「サロゲートソースキー」に組合せることができる。ある実施形態では、サロゲートキーは、ソースデータ記録を目標データ記録に変換するのに用いられる変換に関する付加的な情報を含むことができる。目標データ記録が複数のソースデータ記録からのデータを含有する状況では、目標テーブルを適切な数の列で拡張することができるため、目標データ記録は適切な数のソースサロゲートキーを格納することができる。次にシステムは、列マッピングデータとして格納される対応のソーステーブルと関連付けられる記録マッピングデータおよび目標列を保持する、目標テーブル中の1つ以上のソースサロゲートキー列同士の間の関係を格納することができる。この組合せられたデータをシステムが用いて、データ要素についてのソースまたは目標データ系統追跡を提供することができる。

【0009】

別の実施形態では、データ要素にリンクされる「関連情報」または「関連オブジェクト」の可視性をサポートし、かつ表示およびアクセスのために関連情報（すなわち関連オブジェクト）を関係するソースデータ要素および目標データ要素に伝播するようにデータ系統追跡情報を強化することができるシステムが提供される。システムは、データについての外部ソースへのデータ系統内の変換の経路を通した外部システムへの関連情報（すなわち関連オブジェクト）のエクスポートもサポートすることができる。

【0010】

当該実施形態に従うと、「関連オブジェクト」は、データウェアハウス内に格納されるデータ要素と関連付けられる任意の関連情報（すなわち、関連オブジェクトが関連付けられているデータ要素、または既にデータ系統によって関連付けられているデータウェアハウス中の任意の他のデータ要素以外の任意の情報）であることができる。関連オブジェクトの一種が演算オブジェクトである。1つ以上の関連オブジェクトを「関連情報」として

10

20

30

40

50

も同定することができる。関連オブジェクトの例が「不一致」である。「不一致」は、データ要素の1つ以上の特性が予測される特性から乖離していることの指標である。のような乖離は、1つ以上のデータ駆動規則に基づいて人が検出するかまたは機械が検出することができる。さらに、システムのユーザが手動で不一致を作成することができるか、または、データの確認の結果、たとえば、システムが自動的に作成することができる。不一致の例はエラー報告である。関連オブジェクトの他の例は、「コメント」、「ステータスフラグ」、「状態インジケータ」、「タグ」、「画像」、「録音録画」、「ファイル」、「リンク」、および「文書」を含むことができるが、これらに限定されない。「コメント」は、データ要素に属するユーザによって入力されるテキストの集まりである。「ステータスフラグ」は、データ要素のステータスの指標である。「状態インジケータ」は、データ要素の状態の指標である。「タグ」は、データ要素に属する語または句である。「画像」は、データ要素に属する画像である。「録音録画」は、データ要素に属する任意の音声または映像の録音録画である。「ファイル」は、データ要素に属するコンピュータファイルである。「リンク」は、データ要素に属するハイパーテキストマークアップ言語（「H T M L」）である。「文書」は、データ要素に属する外部文書である。関連オブジェクトは、データ要素またはデータ記録全体のいずれかにリンクされてもよい。10

【0011】

当該実施形態に従うと、システムは、関連オブジェクトと、関連オブジェクトに関連付けられるデータ要素またはデータ記録との間のリンクを格納することができる。関連オブジェクトが最初に作成されたまたは元から関連付けられていた（というのも、関連オブジェクトは先存する関連オブジェクトであることができるので）データ要素またはデータ記録を一次リンクとしてマークすることができる。システムによって生成されて、データ要素またはデータ記録の上流およびデータ要素またはデータ記録の下流の両者を含むデータ系統内の関連オブジェクトの可視性を提供するリンクを二次リンクとしてマークすることができる。このように、当該実施形態に従うと、不一致などの関連オブジェクトは、データ系統内のデータ要素またはデータ記録の上流およびデータ要素またはデータ記録の下流の両方で可視であることができる。この可視性を提供するため、システムは、（a）関連オブジェクトとデータ要素（またはデータ記録）との間の一次リンクを格納し、（b）データ要素（またはデータ記録）のソースおよび目標系統を追跡し、（c）一次データ要素からの、ソースおよび目標追跡中の関連オブジェクトと各データ要素との間の1つ以上の二次リンクを格納し、かつ（d）関連オブジェクト可視性の表示のためのユーザインターフェイスに一次および二次リンク情報の両方を与えることができる。さらに、関連オブジェクトの存在を検出するためにデータ系統を横断することに対して、上流または下流のデータ要素から関連オブジェクトへの直接アクセスを得るために、システムが生成するリンクを用いることができる。20

【0012】

図1は、発明の1つの実施形態を実現することができるシステム10のブロック図を示す。システム10は、バス12またはシステム10の構成要素同士の間で情報を通信するための他の通信機構を含む。システム10は、情報を処理しつつ命令または動作を実行するためにバス12に動作するように結合されるプロセッサ22も含む。プロセッサ22は、任意の種類の汎用または特定用途プロセッサであってもよい。システム10はさらに、プロセッサ22によって実行すべき命令および情報を格納するためのメモリ14を含む。メモリ14は、ランダムアクセスメモリ（「RAM」）、読み専用メモリ（「ROM」）、磁気もしくは光学ディスクなどの静的記憶装置、または任意の他の種類の機械もしくはコンピュータ読み出可能媒体の任意の組合せを備えることができる。システム10はさらに、ネットワークへのアクセスを与えるように、ネットワークインターフェイスカードまたは他の通信インターフェイスなどの通信装置20を含む。その結果、ユーザは、直接にまたはネットワークもしくは任意の他の方法によって遠隔に、システム10とインターフェイスし得る。30

【0013】

コンピュータ読出可能媒体は、プロセッサ22によってアクセス可能な任意の利用可能な媒体であってもよい。コンピュータ読出可能媒体は、揮発性および不揮発性媒体の両方、取外し可能なおよび取外し不可能な媒体、通信媒体、ならびに記憶媒体を含んでもよい。通信媒体は、コンピュータ読出可能命令、データ構造、プログラムモジュール、または搬送波もしくは他の伝送機構などの変調データ信号中の他のデータを含んでもよく、当該技術分野で公知の任意の他の形態の情報送達媒体を含んでもよい。記憶媒体は、RAM、フラッシュメモリ、ROM、消去可能プログラマブル読出専用メモリ(「EPROM」)、電気的消去可能プログラマブル読出専用メモリ(「EEPROM」)、レジスタ、ハードディスク、取外し可能ディスク、コンパクトディスク読出専用メモリ(「CD-ROM」)、または当該技術分野で公知の任意の他の形態の記憶媒体を含んでもよい。

10

【0014】

プロセッサ22は、バス12を介して液晶ディスプレイ(「LCD」)などのディスプレイ24に動作するように結合されることもできる。ディスプレイ24はユーザに対して情報を表示することができる。キーボード26およびコンピュータマウスなどのカーソル制御装置28もバス12に動作するように結合されて、ユーザがシステム10とインターフェイスできるようにすることができる。

【0015】

1つの実施形態に従うと、メモリ14は、プロセッサ22によって実行されると機能性を提供し得るソフトウェアモジュールを格納することができる。モジュールは、オペレーティングシステム15、関連情報伝播モジュール16、および他の機能モジュール18を含むことができる。オペレーティングシステム15は、システム10のためのオペレーティングシステム機能性を提供することができる。関連情報伝播モジュール16は、以下により詳細に記載するように、関連情報の可視性を伝播させるための機能性を提供することができる。ある実施形態では、関連情報伝播モジュール16は複数のモジュールを備えることができ、各々のモジュールは関連情報の可視性を伝播させるための特定的な個々の機能性を提供する。システム10はより大きなシステムの一部であることもできる。このように、システム10は、付加的な機能性を含むように1つ以上の付加的な機能モジュール18を含むことができる。たとえば、機能モジュール18は、オラクルコーポレーションからの「Oracle Life Sciences Data Hub」製品のモジュールなどの付加的な機能性を提供するモジュールを含んでもよい。1つの実施形態では、機能モジュール18は、以下により詳細に記載するように、データ系統を追跡するための機能性を提供することができるデータ系統モジュールを含んでもよい。

20

【0016】

プロセッサ22は、バス12を介してデータベース34に動作するように結合されることもできる。データベース34は論理的に関連したデータ記録、システム記録、またはデータファイルの集積された集合の中のデータを格納することができる。データベース34は、演算データベース、分析データベース、データウェアハウス、分散データベース、エンタープライズデータベース、外部データベース、ナビゲーションデータベース、メモリ内データベース、ドキュメント指向データベース、リアルタイムデータベース、リレーショナルデータベース、オブジェクト指向データベース、データ管理ワークベンチ、または当該技術分野で公知の任意の他のデータベースであることができる。

30

【0017】

図2は、発明の実施形態に従う、システムのための論理データモデルのブロック図を示す。論理データモデルは、データ系統伝播モジュール200を含むことができる。データ系統伝播モジュール200は、以下により詳細に記載するように、データ系統を追跡するための機能性および関連情報の可視性を伝播するための機能性を提供することができるモジュールである。ある実施形態では、データ系統伝播モジュール200は、データ系統を追跡するための機能性を提供する別個のデータ系統モジュールと、関連情報の可視性を伝播するための機能性を提供する別個の関連情報伝播モジュールとを含むことができる。

40

【0018】

50

当該実施形態に従うと、データ系統伝播モジュール200は、記録マッピングデータ205、列マッピングデータ210、および関連オブジェクトマッピングデータ215を発生することができる。先に記載したように、記録マッピングデータ205は、変換の目標テーブル中のデータ記録中に格納されるソースサロゲートキー（すなわち、ソースデータ記録のアイデンティティ）を含む。これも先に記載したように、列マッピングデータ210は、変換のソーステーブルおよび列、変換の目標テーブルおよび列、変換の種類、ならびに変換に必要な任意の付加的な情報を含む。さらに、列マッピングデータ210を変換定義から導出することができる。最後に、これも先に記載したように、関連オブジェクトマッピングデータ215は、関連オブジェクトと（一次リンクとして同定された）関連オブジェクトと関連付けられるデータ要素またはデータ記録との間のリンクを含む。関連オブジェクトマッピングデータ215は、データ系統伝播モジュール200によって生成されて（二次リンクとして同定される）上流および下流データモデル中の関連オブジェクトの可視性を提供するリンクも含む。そのような可視性は、データ表示において関連オブジェクトと関連付けられるデータ要素またはデータ記録を特に視覚的に同定するための関連オブジェクトの効率的なアクセスを含むことができる。データ系統伝播モジュール200はさらに、公開アプリケーションプログラミングインターフェイス（「API」）220を含むことができる。公開API220は、システムのユーザが列マッピングデータをデータ系統伝播モジュール200に入力することができるよう設けることができるAPIである。これによりユーザは、カスタム手続型言語／構造化問合せ言語（「PL/SQL」）プログラムまたはカスタム統計分析システム（「SAS」）プログラムなどのカスタムコンピュータプログラムを用いてデータについてのマッピングを格納できるようになる。
。

【0019】

当該実施形態に従うと、論理データモデルはさらに、変換モジュール225、確認モジュール230、およびクエリービルダ235を含むことができる。変換モジュール225は、マッピングされたデータの一意なソースを同定し、ソースサロゲートキーを1つ以上の目標データテーブルの定義に追加することができる。さらに、1つ以上の変換定義に基づいて、変換モジュール225は、1つ以上の変換についての列マッピングデータ210を格納することができる。変換モジュール225はさらに、同定されるカスタム変換についてのカスタム列マッピングデータをユーザが変換モジュール225に与えられるようすることができる。変換モジュール225は確認モジュール230およびクエリービルダ235を実現することができ、確認モジュール230は、1つ以上のデータ記録を読み出し、1つ以上のデータ記録のデータ値をテストし、かつデータ値についての1つ以上の不一致を作成することができ、クエリービルダ235は、変換と同じ目標列マッピングへのメタデータソース列を維持しながら、1つ以上のデータテーブルから入力するデータのクエリーまたはデータベースビューを構築することができる。ある実施形態では、変換マッピングデータは「好ましいソース経路」も含むことができる。変換の目標列が複数の直近のソース列を有する場合、1つの列を好ましいソース列として同定することができる。この情報を列マッピングデータ210の一部として追跡することができる。このように、「好ましいソース経路」は、好ましいソース列の変換を通る経路である。換言すると、好ましいソース経路は、複数の可能なソース経路から、他の可能なソース経路よりも好ましいとして同定される単一のソース経路である。これは、クエリーを添付するための単一のソースデータ要素またはソースデータを与えたソースシステムに送られ得る他の関連オブジェクトに一意な経路を与えるように行なうことができる。

【0020】

（変換実行処理240として図2に図示される）変換の実行の間、記録マッピングデータ205を1つ以上のソースサロゲートキーの中に格納することができ、1つ以上のサロゲートキーを、（図2にデータテーブル245として図示される）変換の1つ以上の目標テーブルの1つ以上の補助列の中に格納することができる。当該実施形態に従うと、記録マッピングデータ205および列マッピングデータ215と他のデータ系統追跡情報との
。

組合せを用いて、以下にさらに詳細に記載するように、（図2に系統追跡250として図示される）ソースまたは目標データ要素のデータ系統の追跡を表示することができる。

【0021】

当該実施形態に従うと、論理データモデルはさらに、関連オブジェクトモジュール255およびアダプタ260を含むことができる。関連オブジェクトモジュール255は、各関連オブジェクト毎に関連オブジェクトマッピングデータ215内にデータを入力することができる。関連オブジェクトマッピングデータ215からのデータを用いて、（図2にはハイライト関連オブジェクト250として図示される）データ要素に関連付けられる関連オブジェクトを有する1つ以上のデータ要素の表示を視覚的にハイライトすることができる。ある実施形態に従うと、関連オブジェクトモジュール255は、関連オブジェクトに関連付けられる各データ要素毎に好ましいソース経路を要求することができる。さらに、アダプタ260を用いて、1つ以上の関連オブジェクトをシステム内に作成することができる。アダプタ260は、関連オブジェクトモジュール255を用いてシステム内の1つ以上の関連オブジェクトを作成することができ、次に関連オブジェクトモジュール255は、データ系統伝播モジュール200を用いて、ソースまたは目標データ要素のデータ系統の追跡内で関連オブジェクトの可視性を伝播することができる。ある実施形態では、論理データモジュールはさらに、1つ以上のデータテーブルからの1つ以上のデータ要素を表示しつつ1つ以上のデータ要素と関連付けられる1つ以上の関連オブジェクトの存在をハイライトする（図2には図示しない）付加的なモジュールを含むことができる。

【0022】

図3は、発明の実施形態に従う、例示的なデータ系統のブロック図を示す。ある実施形態に従うと、システムは、データウェアハウス内に格納されるデータのデータ系統の追跡を生成することができる。このデータ系統追跡内で、システムのユーザは、目標テーブル中のデータ要素を選択し、選択されたデータ要素の導出元である1つ以上のソースデータ要素への1つ以上の変換ステップを通した上流への追跡を要求することができる。さらに、ユーザは、ソーステーブル中のデータ要素を選択して、選択されたデータ要素から導出される1つ以上の目標データ要素への1つ以上の変換ステップを通した下流への追跡を要求することができる。変換に関して、当該実施形態に従うと、異なる種類の変換が存在することができる。そのような変換の種類は、（a）（おそらくはフィルタおよび／または導出を含む）直接マッピング変換、（b）統合変換、（c）（おそらく単一のおよび／または繰返しデータを含む）合体変換、（d）横縦アンピボット変換、および（e）縦横ピボット変換を含むことができる。異なる変換の種類を以下により詳細に記載する。ある実施形態では、複数段階変換のためにデータを格納するように作成される1つ以上の段階実施テーブルを通してデータ系統を含むようにデータ系統追跡を生成することができる。当該実施形態に従うと、段階実施テーブルは、必要に応じて1つ以上の変換のために利用することができるデータウェアハウスのテーブルであり、複数段階変換シナリオのための中間格納構造として働く目的を果たすことができる。

【0023】

直接マッピング変換は、ソーステーブル中のソースデータ記録の1つ以上の列を目標テーブルの目標データ記録中で名称変更するおよび／または並べ替える変換である。ある実施形態では、ソーステーブル中のソースデータ記録の1つ以上の列を直接マッピング変換内で変更せずに伝播させることができる。統合変換は、目標テーブルが2つ以上のソーステーブルからの1つ以上のソースデータ記録の一部またはすべてを含むように、2つ以上のソーステーブルからの1つ以上のソースデータ記録を单一の目標テーブルの2つ以上の目標データ記録に組合せる変換である。合体変換は、目標テーブルの目標データ記録が2つ以上のソーステーブルからの1つ以上のソースデータ記録のすべての列を含むように、2つ以上のソーステーブルからの1つ以上のソースデータ記録を单一の目標テーブルの单一の目標データ記録に組合せる変換である。

【0024】

さらに、横縦アンピボット変換は、ソーステーブルからの1つ以上のソースデータ記録

10

20

30

40

50

を目標テーブルの 1 つ以上の目標データ記録に転換し、各々のソースデータ記録の 1 つ以上の列の各組を目標テーブルの別個の目標データ記録に変換する変換である。たとえば、ソーステーブルが列 A、B、C を含む単一のソースデータ記録を有する場合、アンピボット変換は、ソースデータ記録を、3 つの目標データ記録、すなわち、列 A を含む第 1 の目標データ記録、列 B を含む第 2 の目標データ記録、および列 C を含む第 3 の目標データ記録、に変換する。縦横ピボット変換は、ソーステーブルからの 1 つ以上のソースデータ記録を目標テーブルの目標データ記録に転換し、1 つ以上のソースデータ記録を単一の目標データ記録に組合せる変換である。たとえば、ソーステーブルが、3 つのソースデータ記録、すなわちデータ記録 1、データ記録 2、およびデータ記録 3 を含む場合、ピボット変換は、3 つのソースデータ記録を、データ記録 1、2、および 3 の列のうち 1 つ以上を含む単一の目標データ記録に変換する。ケース変換は、目標データ記録のデータ値が、条件に基づいて異なるソーステーブルからあり得る 1 つ以上のソース列からポピュレートされる変換である。10

【0025】

ある実施形態に従うと、システムのユーザインターフェイス内での表示のためにデータ系統追跡情報を利用可能にすることができる。例示的な実施形態に従ってユーザインターフェイスに与えることができる情報をここでより詳細に記載する。選択されたデータ要素のソース系统については、選択されたデータ要素を発生したデータ要素の階層を与えることができる。これは、すべての関連変換を通してデータウェアハウス中の生のソースデータへ遡って横断するデータ要素の経路である。経路はデータを通るいくつかの経路分岐を含むことができる。同様に、選択されたデータ要素の目標系統については、選択されたデータ要素から導出されるデータ要素の階層を与えることができる。20

【0026】

当該実施形態に従うと、以下の属性のうち 1 つ以上を経路中の各データ要素毎に与えることができる：すなわち、(a) 階層内のデータ要素の配置（たとえば、データ要素のアイデンティティおよび階層内のその親データ要素のアイデンティティ）；(b) データモデル名；(c) テーブル名；(d) 列名；(e) データ要素のデータ種類；(f) データ要素の値；(g) 好ましいソースルート（はいまたはいいえ）；(h) 主キー；(i) データがマスクされたデータである（あり得る）か否かのインジケータ；(j) データが内部段階実施テーブルからのものであるか否かのインジケータ；(k) 1 つ以上の関連オブジェクト（不一致など）がデータ要素に主にリンクされていることのインジケータ；または(l) 基礎となるデータ（たとえば、テーブルオブジェクトアイデンティティ、列オブジェクトアイデンティティ、サロゲートキー値、および任意の他の関係のデータ）、である。「データモデル」は、データウェアハウスの 1 つ以上のデータテーブルと関連付けられるデータスキーマである。30

【0027】

図 3 は、発明の実施形態に従う、一連の医学的研究に関する患者データのデータ系統を図示する。しかしながら、他の実施形態では、データは、任意の種類の生命科学または医療データであることができる。さらに、他の実施形態では、データは任意の種類のデータであることができる。図 3 に示される実施形態に従うと、データウェアハウスは、ソースデータモデル 310、全現場データモデル 320、検討データモデル 330、および分析データモデル 340 の 4 つのデータモデルを含むことができる。ソースデータモデル 310 は、人口統計テーブル 311、現場 A 訪問テーブル 312、および現場 B 訪問テーブル 313 の 3 つのテーブルを含む。人口統計テーブル 311 は、研究におけるすべての被験者の人口統計データを含み、人口統計データは各々の被験者の身長および身長の単位を含む。現場 A 訪問テーブル 312 は、研究現場 A における各被験者毎の訪問当たりの 1 つのデータ記録（すなわち訪問データ）を含み、訪問データは被験者の体重および体重の単位を含む。現場 B 訪問テーブル 313 は、研究現場 B における各被験者毎の訪問当たりの 1 つのデータ記録（すなわち訪問データ）を含み、訪問データは被験者の体重および体重の単位を含む。全現場データモデル 320 は、訪問テーブル 321 という 1 つのテーブルを4050

含む。訪問テーブル321は、現場A訪問テーブル312内に含まれる訪問データと現場B訪問テーブル313内に含まれる訪問データとの統合変換によって作成される現場A訪問テーブル312と現場B訪問テーブル313との統合である。検討データモデル330はマイルストーン訪問テーブル331という1つのテーブルを含む。マイルストーン訪問テーブル331は、ソースフィルタ変換を用いて、訪問テーブル321内に含む訪問データから導出される各患者毎の重要な訪問のサブセットである（たとえば、各被験者毎の訪問1、4、7、および10）。分析データモデル340はB M I テーブル341という1つのテーブルを含む。B M I テーブル341は、人口統計テーブル311内に含まれる人口統計データとマイルストーン訪問テーブル331内に含まれる訪問データとの合体変換によって作成される人口統計テーブル311とマイルストーン訪問テーブル331との合体であり、体格指数（「B M I」）値は、被験者の体重、体重の単位、身長、および身長の単位を用いた導出である。

【0028】

図4は、発明の実施形態に従う、データ系統400を表示する例示的なユーザインターフェイスを示す。例示的な実施形態に従うと、図3のB M I テーブル341は以下のデータを含むことができる。

【0029】

【表1】

研究	被験者	訪問	現場	性別	年齢	年齢単位	BMI	BMIU
BMI	1001	4	MGH	女性	34	才	22.7	KG/M**2
BMI	1002	4	MGH	女性	24	才	24.5	KG/M**2
BMI	1003	4	MGH	男性	42	才	21.9	KG/M**2
BMI	1004	4	MGH	男性	28	才	26.3	KG/M**2
BMI	1005	4	MGH	女性	34	才	24.6	KG/M**2
BMI	1006	4	MGH	女性	29	才	28.8	KG/M**2
BMI	1007	4	McLean	女性	42	才	33.3	KG/M**2
BMI	1008	4	McLean	女性	35	才	22.5	KG/M**2
BMI	1009	4	McLean	男性	34	才	33.6	KG/M**2
BMI	1010	4	McLean	女性	54	才	30.6	KG/M**2

10

20

30

40

【0030】

当該実施形態に従うと、システムは、システムのユーザインターフェイス内に上記データを表示することができる。システムのユーザは、女性、年齢42歳、B M I 値33.3である被験者1007についてソースデータを調査したいかもしれない。ユーザは、ユーザインターフェイス内で被験者1007についてのデータ値33.3を選択し、ユーザがデータ値のソースデータ系を検討することを望んでいると示すことができる。応答して、システムは、ユーザインターフェイス内に図4のデータ系統400を表示することができ、これをさらにより詳細に説明する。

【0031】

図示される実施形態に従うと、データ系統400は、複数の行（すなわち、行401、

50

402、403、404、405、406、407、408、および409)を含み、各行はソースデータ要素のうち1つについての情報を含む。行は、図3に示される変換の組を通る経路を横断する。このように、直近の上位データ要素は、データシステム400内で選択されたデータ要素よりも下の階層レベルに存在し、それらの上位が同様の態様に従う。このように、階層経路は、選択されたデータ要素に最も近いデータ要素から上流へ、生のソースデータ要素へと横断される。したがって、図4に示される実施形態では、行401が選択されたデータ要素(すなわち、被験者1007についてのデータ値33.3)を表わす。行402、403、404、および407は、行401によって表わされるデータ要素についての直近の上位データ要素を表わす。このように、行401によって表わされる選択されたデータ要素の直近の上位テーブルは、図3の人口統計テーブル311およびマイルストーン訪問テーブル331である。行405は、行404によって表わされるデータ要素についての直近の上位データ要素を表わす。行406は、行405によって表わされるデータ要素についての直近の上位データ要素を表わす。行408は、行407によって表わされるデータ要素についての直近の上位データ要素を表わす。行409は、行408によって表わされるデータ要素についての直近の上位データ要素を表わす。このように、行404によって表わされるデータ要素および行407によって表わされるデータ要素の直近の上位テーブルは訪問テーブル321である。さらに、405によって表わされるデータ要素および行408によって表わされるデータ要素の直近の上位テーブルは現場B訪問テーブル313である。データシステム400はソースデータシステムの例であるが、代替的な実施形態では目標データシステムをユーザインターフェイス内に表示することができ、選択されたデータ要素についての1つ以上の下位データ要素を表示することができ、目標データシステムは複数の行を含み、各行は目標データ要素のうち1つについての情報を含む。
10

【0032】

データシステム400はさらに複数の列(すなわち列410、420、430、440、および450)を含む。列410は、データ要素のデータモデル名、テーブル名、および列名を表わす。列420はデータ要素のデータ値を表わす。列430はデータ要素のデータ種類およびサイズを表わす。列440はデータ要素が好ましいソース経路の一部であるか否か(すなわち「はい」または「いいえ」)の指標を表わす。列450はデータ要素についての主キー値を表わす。

【0033】

このように、データシステム400はデータ要素の階層を表示するツリーテーブルであることができ、列の見出しが静的であり、ユーザはツリーテーブル構成要素を用いて階層の部分を開くまたは閉じることができる。さらに、1つの実施形態では、(図4にデータ要素460として図示される)関連オブジェクトにリンクされるデータ要素を背景色でハイライトすることができます。ハイライトされた背景色はデータ要素と関連オブジェクトとの間の一次リンクを表わすことができる。一次リンクのインスタンスでは、ユーザはハイライトされたデータ要素を選択し、データ要素にリンクされた関連オブジェクトの詳細を表示するよう要求することができる。

【0034】

1つ以上の実施形態では、データシステム400が内部段階実施テーブルを通る経路を含めれば、対応するデータ値を(図4には図示しない)星印などの記号でマークすることができる。さらに、データシステム400の下に脚注を表示して記号の意味を説明することができる。加えて、1つの実施形態では、データ要素が、セキュリティがテーブルまたは列に適用されているテーブルまたは列からのものであり、かつマスクされたデータが実際のデータの代わりに表示されているのであれば、データ要素を(図4には記号470として図示される)記号「§」などの記号でマークすることができる。さらに、データシステム400の下に脚注を表示して記号の意味を説明することができる。さらに、1つの実施形態では、データに適用されているセキュリティのために(上流または下流のいずれかの)追跡の経路中のデータのうちの一部へのアクセスをユーザが有していないために、またはデータ記録がテーブルから削除されて削除が下流のテーブルにまだ伝播されていないために、完全な
20
30
40
50

データ系統追跡を完了することができないことがある。いずれの場合も、当該実施形態に従うと、ユーザインターフェイスは、（図4には図示しない）データ追跡が終了されたことを示すメッセージを表示することができる。

【0035】

このように、1つの実施形態では、システムのユーザは、データブラウザを通してデータ一覧ページへ移動してユーザインターフェイス内に表示されるデータグリッド上のデータ要素を選択することができる。「データグリッド」は、データテーブルの行および列またはデータのビューを提示する表示である。データ要素を選択する際、ユーザは、選択されたデータ要素の導出元であるソースデータをユーザが見たいと示すことができる。応答して、システムは、ソースデータ要素の集合を入手して、各ソースデータ要素毎にデータモデル、テーブル、および列についての対応の情報とともにソースデータ要素値の集合を表示することができる。選択されたデータ要素についてのソースデータを構成する経路のツリーとして集合をユーザインターフェイス中に表示することができるよう集合が編成されかつ集合が情報を含むことができる。好みしいソース経路を表示できるように情報を含むこともできる。10

【0036】

別の実施形態では、ユーザは、データブラウザを通してデータ一覧ページへ移動してユーザインターフェイス内に表示されるデータグリッド上のデータ要素を選択することができる。データ要素を選択する際、ユーザは、選択されたデータ要素から導出される目標データをユーザが見たいと示すことができる。応答して、システムは、目標データ要素の集合を入手して、各目標データ要素毎のデータモデル、テーブル、および列についての対応の情報とともに目標データ要素値の集合を表示することができる。選択されたデータ要素についての目標データを構成する経路のツリーとして集合をユーザインターフェイス中に表示できるように、集合が編成されかつ集合が情報を含むことができる。20

【0037】

さらに、関連オブジェクトに関して、1つの実施形態では、データウェアハウス内に作成されるかまたはそうでない場合はデータウェアハウスのデータ要素と関連付けられる関連オブジェクトを、これらのオブジェクトが異なる関連モデルに伝播されたかのように上流または下流データモデルで閲覧することができる。システムは、1つのデータモデル中に作成された関連オブジェクトを他のデータモデル中のその対応のソースおよび目標データ要素と自動的に関連付けることができる。ソースおよび目標データモデル中の関連オブジェクトとデータ要素との間の関連付けをシステムのユーザインターフェイス内に表示することができ、関連オブジェクトの存在を関連データ要素および関連オブジェクトカテゴリとともに表示することができる。さらに、目標データ要素を導出するのに複数のソースデータ要素を用いる場合、変換定義の間にソースデータ要素のうちの1つを好みしいソースデータ要素として選択することができる。好みしいソースデータ要素は、ソースデータを通して上流にデータ系統経路を形成することができる。最も上流のソースデータ要素を用いて関連オブジェクトを外部システムにエクスポートすることができる。さらに、関連オブジェクトと関連付けられたデータが更新されると、データが更新されたことを示すように関連オブジェクトの1つ以上のフラグを設定可能である。30

【0038】

先に説明したように、データ系統400を表示するユーザインターフェイス内の表示のために、データ要素と関連オブジェクトとの間のリンクを利用可能にすることができる。データ要素またはデータ記録と関連付けられる関連オブジェクトを有するデータ要素またはデータ記録を同定するためのシステムがユーザインターフェイスに与える情報をさらに詳細にここで説明する。システムは情報をユーザインターフェイスに与えることができ、これによりデータ要素と関連付けられた関連オブジェクトを有する各々のデータ要素を表示においてハイライトする。これは、関連オブジェクトが元々作成されるかまたはそれ以外のやり方で関連付けられる（すなわち一次リンク）テーブルの中のデータ要素を含むことができる。40

【 0 0 3 9 】

1つの実施形態では、システムは、(a) (関連オブジェクトおよびそのプロパティにアクセスするのに用いることができる) 関連オブジェクトアイデンティティ、(b) (テーブルオブジェクトアイデンティティ、列オブジェクトアイデンティティ、およびサロゲートキーなどの) データ要素参照、(c) データ要素と関連付けられる関連オブジェクトが一次リンクであるかまたは二次リンクであるかの指標、(d) リンクが (二次リンクのための) 一次データ要素のソース経路または目標経路の中にあるか否かの指標、および(e) 関連オブジェクトの状態を与えることができる。当該実施形態に従うと、データ要素 10 460 によって図4に示されるように、システムは、背景色でデータ要素をハイライトして、データ要素が関連オブジェクトにリンクされていることを示すことができる。システムは、第1の背景色を用いてデータ要素をハイライトして一次リンクを示すことができる。これに代えて、システムは、第2の背景色を用いてデータ要素をハイライトして二次リンクを示すことができる。先に説明したように、ユーザは、関連オブジェクトがリンクされているデータ要素のソースまたは目標系統を要求することができる。ユーザが関連オブジェクトへの一次リンクを含むデータ要素に到達すると、ユーザは、関連オブジェクトの1つ以上の特性を含む関連オブジェクトを閲覧することができる。ユーザは、関連オブジェクトを外部ソースにエクスポートすることもできる。代替的な実施形態では、データ要素を色でハイライトしてデータ要素が関連オブジェクトに関連付けられていることを示すよりもむしろ、システムは代替的な指標を用いることができる。代替的な指標のいくつかの例は、アイコン、色付けされた境界、または音声指標を含むが、これらに限定されない 20 。

【 0 0 4 0 】

30 このように、1つの実施形態では、システムのユーザは、データ一覧へ移動し、ユーザインターフェイス内に表示されるデータグリッド上のデータ要素を選択し、関連オブジェクトを作成し、選択されたデータ要素と関連オブジェクトとを関連付けることができる。次にシステムは、関連オブジェクトを、選択されたデータ要素に関するデータ要素と関連付けることができる。システムは新たに作成された関連オブジェクトを選択されたデータ要素にリンクさせ、リンクを一次リンクとして指定することができる。システムはさらに、データウェアハウス内の既存の変換と関連付けられる列マッピングデータおよび記録マッピングデータを用いることにより、選択されたデータ要素についてのすべてのソースおよび目標データ要素を帰納的に見つけることができる。システムはその後、同じ関連オブジェクトを関係のソースおよび目標データ要素にリンクさせてこれらのリンクを二次リンクとして指定することができる。このように、各々の関係のソースまたは目標データモデルにおいて、関連オブジェクトへの二次リンクを有するデータ要素をシステムのユーザインターフェイス内でハイライトすることができる。

【 0 0 4 1 】

別の実施形態では、システムのユーザはデータ一覧へ移動することができ、システムは関連オブジェクトと一次ソースデータとの間の一次リンクおよび二次リンクを与えることができる。システムはさらに、各々の関連オブジェクトの状態などの関連オブジェクトからのプロパティを入手して表示をフィルタすることができます。また別の実施形態では、ユーザは、1つ以上の関連オブジェクトへのリンクを含むデータ要素を削除することができる。システムは、削除されたデータに関連付けられている関連オブジェクトへの一次および/または二次リンクを外すことができる。一次リンクについてのデータが削除されると、システムは、データ要素と関連付けられている関連オブジェクトへの一次および二次リンクを外すことができ、システムは関連オブジェクトを削除することができる。二次リンクについてのデータが削除されると、システムは二次リンクを外し、それ以上のリンクを外す（および必要に応じて任意のそれ以上のリンクを外す）べきか否かを判断することができる。より一層の実施形態では、システムは、ユーザが、列マッピングデータおよび関連オブジェクトの可視性をサポートするリンク済関連オブジェクトにアクセスできるクエリーを作成できるようにする。 40 50

【0042】

(図2のデータ系統伝播モジュール200などの)データ系統伝播モジュールについての制御の流れにおいて、1つの実施形態に従うと、1つ以上のデータモデルをデータウェアハウス内に定義することができる。各データモデル毎のメタデータが作成されるかまたはロードされる。(図2の変換モジュール225などの)変換モジュールを用いて1つ以上の目標テーブルについて1つ以上の変換が定義される。必要に応じて、特定の変換のために、各目標テーブルに1つ以上の補助列が追加される。これらの補助列を以下により詳細にさらに説明する。その後、データ系統マッピングテーブルは、データウェアハウスの1つ以上のデータテーブル内に格納されるデータとは区別される列マッピングに関するデータ系統追跡情報(すなわちメタデータ情報)でポピュレートされる。なお、データ系統マッピングテーブルを以下により詳細にさらに説明する。補助列アイデンティティは列マッピングメタデータの一部であり、これにより、マッピングされた目標列を、マッピングされたソース列についてのソーステーブル記録のアイデンティティを保持する補助列とリンクさせることができる。1つ以上のデータモデルに1つ以上のテーブルがインストールされる。1つ以上のソーステーブルがデータでロードされる。変換が実行され、1つ以上の目標テーブルをポピュレートする。1つ以上の目標テーブルがデータでロードされた後に、データウェアハウスの1つ以上のデータテーブル内に格納されるデータからこれも区別されるデータ系統追跡情報(すなわち記録マッピング情報)でシャドウテーブルがポピュレートされるかまたは更新される。なお、シャドウテーブルを以下により詳細にさらに説明する。最終的に、関連オブジェクトについての1つ以上の二次リンクがデータ内で伝播される。関連オブジェクトの存在を検出するためにデータ系統マッピングテーブルおよびシャドウテーブルのデータ系統追跡情報を横断することなく、二次リンクを用いて関連オブジェクトとの上流または下流データ要素の関連付け入手することができる。

10

20

【0043】

このように、ある実施形態に従うと、(図2のデータ系統伝播モジュール200などの)データ系統伝播モジュールは、1つ以上の関連付けを維持することができ、2種類のオブジェクト、すなわち、(a)テーブル中のデータ記録のデータ要素、および(b)テーブル中の完全なデータ記録、同士の間を追跡することができる。(a)に関して、データ要素を一意に同定するため、およびデータ要素入手するのに必要な場合にインストールされたオブジェクトの名前を追跡するためにも、(1)データウェアハウスのデータモデルを同定するモデルオブジェクトアイデンティティ、(2)データウェアハウスのテーブルを同定するテーブルインスタンスオブジェクトアイデンティティ、(3)サロゲートキーを同定したサロゲートキー値、および(4)テーブルの列を同定する列オブジェクトアイデンティティを用いることができる。(b)に関して、データ記録内の特定的なデータ要素よりもむしろ、いくつかの関連オブジェクトをデータ記録にリンクさせることができる。このシナリオでは、データ記録内に特定的な列が存在しないので、テーブルの列を同定する列オブジェクトアイデンティティよりもむしろ、サロゲートキーについての列オブジェクトアイデンティティを用いる。このように、(1)データウェアハウスのデータモデルを同定するモデルオブジェクトアイデンティティ、(2)データウェアハウスのテーブルを同定するテーブルインスタンスオブジェクトアイデンティティ、(3)サロゲートキーを同定したサロゲートキー値、および(4)サロゲートキーについての列オブジェクトアイデンティティを用いてデータ記録を一意に同定することができる。代替的な実施形態では、(データウェアハウスのすべてのデータモデルにわたって)データウェアハウスのテーブルを一意に同定する一意なテーブル識別子を用いてデータ記録を一意に同定することができる。

30

40

【0044】

さらに、当該実施形態に従うと、データ系統伝播モジュールは、関連オブジェクトとの関連付けを維持しつこれを伝播する。データ系統伝播モジュールはさらに、(a)1つ以上のソースまたは目標データ要素への特定的なデータ要素のマッピング(すなわち、データ系統追跡可能性)、および(b)関連オブジェクトとデータ要素またはデータ記録と

50

の間のマッピング（すなわち、関連オブジェクト可視性）という2種類の関係マッピングをさらに与える。

【0045】

当該実施形態に従うと、データ系統追跡可能性を与えるため、データ系統伝播モジュールは、1つ以上のテーブルがデータウェアハウスを用いて定義された後に、データ系統追跡情報でデータ系統マッピングテーブルをポピュレートする。なお、データ系統マッピングテーブルをさらに以下により詳細に説明する。1つ以上の変換がデータウェアハウス内で実行されると、1つ以上の補助列がポピュレートされ、シャドウテーブルが1つ以上のデータ要素のデータ系統を追跡するためのデータ系統追跡情報でロードされる。なお、補助列およびシャドウテーブルをさらに以下により詳細に説明する。関連オブジェクトが作成されるおよび／またはデータ要素と関連付けられると、データ系統伝播モジュールは、他のプロセスによる効率的な視覚化のためにソースおよび目標データ要素と関連付けられる関連オブジェクトをインデックス付けすることができる。このように、データ系統マッピングテーブルおよびシャドウテーブルのデータ系統追跡情報を横断せずに、関連オブジェクトとの上流または下流データ要素の関連付けを直接に入手することができる。10

【0046】

データモデルおよび変換定義フェーズの間、変換の目標テーブルを同定することができ、1つ以上のメタデータ定義として同定されたテーブルに1つ以上の補助列を追加することができる。補助列の数は変換の種類に依存することができる。1つ以上の補助列の主な用途は、変換のソーステーブルからのソースサロゲートキー値を格納することである。先に説明したようなサロゲートキーはデータ記録のアイデンティティであり、ソースサロゲートキーはソースデータ記録のアイデンティティである。1つ以上の補助列を、1つ以上のソースサロゲートキー値または変換に適切な他のデータ値でポピュレートすることができる。データ系統伝播モジュールは、補助列内に格納されたサロゲートキー値（または他の適切なデータ値）を用いて、目標データに寄与するソースデータ要素の経路を通じて遡って追跡することができる。20

【0047】

先に説明したように、必要に応じて、特定的な変換のために1つ以上の目標テーブルに1つ以上の補助列を追加することができる。異なる種類の変換のために追加することができる補助列の例は以下のとおりである。30

【0048】

【表2】

変形	補助列
統合	1つのソースサロゲートキー列（ソースサロゲートキーはその第1の構成要素としてソーステーブルアイデンティティを含有することができる）
合体	ソースサロゲートキー列：合体に寄与する各テーブル毎に1つ
横縦アンピボット	1つのソースサロゲートキー列
縦横ピボット	1つのソースサロゲートキー列（値は、サロゲートキーの根を保持するように修正可能、すなわちピボット列値を保持する位置のプレースホルダーを用いて）
直接マップ	1つのソースサロゲートキー列（これは直接マップ型導出を含むことができる）
ケース選択	ソースサロゲートキー列：ソーステーブルおよび列にリンクするフィルタ値（ケース識別子）について1列およびソースサロゲートキーについて1列

【0049】

当該実施形態に従うと、データ要素のデータ系統を追跡することは1つ以上のサロゲー50

トキーに依存する。各々のテーブルは、テーブルについてのインデックス付けされた一意なキーとして働くことができるサロゲートキー列を含むことができる。変換の目標テーブルからソーステーブルへ遡って追跡するために、ソーステーブルからのサロゲートキーを目標テーブルの中に格納する。単純な変換の例では、目標テーブル中の選択されたデータ要素についてのソースデータ要素を追跡するため、(以下により詳細に記載する)データ系統マッピングテーブルの中で目標テーブルをロックアップして、ソーステーブル、ソース列、およびソースサロゲートキーが格納される補助列入手する。補助列を用いてソースサロゲートキーの値が入手され、選択されたデータ要素を含有する目標テーブル記録の中に格納される。ソーステーブル、ソース列、およびソースサロゲートキー値を用いて、選択された目標データ要素についてのソースデータ値を見出す。より複雑な変換のためのデータ系統追跡を以下により詳細に説明する。

10

【0050】

選択されたソースデータ要素についての目標データ要素を追跡するため、(以下により詳細に記載する)シャドウテーブルとして同定されるデータ系統追跡情報の別個のテーブルを維持する。シャドウテーブルは、目標データ記録へのソースデータ記録の1つ以上のマッピングを含む。より具体的には、シャドウテーブルは、列とソースキーとの間のメタデータ関係を同定する文脈マップとともに特定のソースキーを用いて、非正規化ソース記録 - 目標記録マッピングを含むことができる。このように、当該実施形態に従うと、変換の実行の終わりに、シャドウテーブルは、データ系統マッピングテーブルを用いて、目標データ要素をソースデータ要素から追跡できるようにするマッピングデータのインデックス付け版を与える1つ以上のシステム記録を含むことができる。

20

【0051】

図5は、発明の実施形態に従う、システムについての物理的モデルのブロック図を示す。図5はシステムについての代表的な物理的モデルを示す。ある実施形態では、システムは、図5に示される物理的モデルの構成要素をすべて含むわけではないことがある。他の実施形態では、システムは、図5に示されていない付加的な構成要素を含むことがある。(図2のデータ系統伝播モジュールなどの)データ系統伝播モジュールは、1つ以上の構成要素を含む物理的モデルを含むことができ、各々の構成要素は1つ以上のテーブルを含む。図5に示される実施形態では、物理的モデルは、目標データ要素からソースデータ要素を追跡するためのデータ系統追跡情報(すなわちメタデータ情報)を含むデータ系統マッピング構成要素510を含む。データ系統マッピング構成要素510は、図6と関連して以下により詳細にさらに記載されるデータ系統マッピングテーブルを含む。さらに、物理的モデルは、目標データ要素をソースデータ要素から追跡するためのインデックス付けデータ系統追跡情報を含むシャドウテーブル構成要素520を含む。シャドウテーブル構成要素520は、図7と関連して以下により詳細にさらに記載されるシャドウテーブルを含む。加えて、物理的モデルは、1つ以上の変換についてのデータを含む変換構成要素530を含む。さらに、物理的モデルは、関連オブジェクトとデータ要素またはデータ記録との間のリンクに関するデータを含む関連オブジェクトマッピング構成要素540を含む。関連オブジェクトマッピング構成要素540は、図9と関連して以下により詳細にさらに記載される関連オブジェクトマッピングテーブルを含む。加えて、物理的モデルは、1つ以上のソーステーブルおよび1つ以上の目標テーブルについてのアプリケーションテーブル定義を含むテーブル定義構成要素550を含む。1つの実施形態では、テーブル定義構成要素550は、データモデルレベルでの1つ以上のアプリケーションテーブル定義(たとえば図5のテーブル「CDR_MODEL」)、テーブルレベルでの1つ以上のアプリケーションテーブル定義(たとえば、図5のテーブル「CDR_TABLE」、「CDR_TABLE_CONS」、および「CDR_TABLES_CONS_COLS」)、および列レベルでの1つ以上のアプリケーションテーブル定義(図5のテーブル「CDR_COLUMN」および「CDR_TABLES_CONS_COLS」)を含むことができる。さらに、物理的モデルは、1つ以上の関連オブジェクトについてのテーブル定義を含む関連オブジェクト構成要素560を含む。1つの実施形態では、関連オブジェク

30

40

50

ト構成要素 560 は、1つ以上の不一致についての1つ以上のテーブル定義（たとえば図5のテーブル「D M E _ D I S C R E P A N C Y」）および1つ以上のフラグについての1つ以上のテーブル定義（たとえば図5のテーブル「D M E _ F L A G」）を含むことができる。加えて、物理的モデルは、物理的モデル中の他のオブジェクトのすべてをグループ分けする親オブジェクトを含むである研究構成要素 570 を含む。

【0052】

ともに、テーブル定義構成要素 550 とともに、データ系統マッピング構成要素 510 、シャドウテーブル構成要素 520 、および変換構成要素 530 は、図8と関連して以下により詳細にさらに記載されるデータ系統サブシステムを提供する。さらに、テーブル定義構成要素 550 および関連オブジェクト構成要素 560 とともに関連オブジェクトマッピング構成要素 540 は、図10と関連して以下により詳細にさらに記載される関連情報サブシステムを提供する。10

【0053】

図6は、発明の実施形態に従うデータ系統マッピングテーブル 600 のブロック図を示す。データ系統マッピングテーブル 600 は、目標データ要素からソースデータ要素を追跡するための情報を含む。目標データ要素からソースデータ要素を追跡するための情報は、データウェアハウスの1つ以上のデータテーブル内に格納されるデータとは区別されるデータ系統追跡情報（すなわちメタデータ情報）である。データモデル公開フェーズの間にデータ系統マッピングテーブル 600 をポピュレートすることができ、データ系統マッピングテーブル 600 は、各データ変換毎にソースデータ要素と目標データ要素との間を追跡するための情報を含む。データ系統マッピングテーブル 600 は、オブジェクトアイデンティティにより、列レベルマップオブジェクト、目標テーブルおよび列、ソーステーブルおよび列、ソースサロゲートキーを保持する補助列、ならびに特定的な変換に必要な任意の付加的な情報に対する参照を格納する。20

【0054】

ある実施形態に従うと、データ系統マッピングテーブル 600 内に格納される情報は、データ系統を追跡するのに用いられるデータ系統追跡情報の第1の主要な部分を構成する。データ系統を追跡するのに用いられるデータ系統追跡情報の第2の主要な部分は、ソースデータ記録へ遡って追跡する変換の目標テーブルのテーブル定義に加えられる1つ以上の補助列である。補助列は、目標テーブルの中のデータ記録に寄与する各々のソーステーブルからのサロゲートキーを含有することができる。必要な場合、補助列は、ソーステーブルのテーブルアイデンティティを保持する、および／または特定的な変換についての「フィルタ」値を同定することができる。データ系統マッピングテーブル 600 および1つ以上の補助列は、目標データ要素からソースデータ要素へ遡って追跡する（図2のデータ系統伝播モジュール 200 などの）データ系統伝播モジュールが用いるデータを含む。データ系統を追跡するのに用いられるデータ系統追跡情報の第3の主要な部分は、図7と関連してより詳細にさらに記載されるシャドウテーブルである。30

【0055】

図6に示される実施形態では、（列マッピングデータを含む）データ系統マッピングテーブル 600 は、変換テーブルレベルマップアイデンティティに対応する（すなわち変換を同定する）「T M A P _ I D」列を含む。データ系統マッピングテーブル 600 は、ソース列がどのように目標列に関するかを規定する（すなわち変換種類を同定する）「T M A P _ T Y P E _ R C」列をさらに含む。データ系統マッピングテーブル 600 はさらに、ソースデータモデルアイデンティティに対応する「S R C _ D M _ O B J _ I D」列、ソーステーブルアイデンティティに対応する「S R C _ T A B _ O B J _ I D」列、およびソース列アイデンティティに対応する「S R C _ C O L _ O B J _ I D」列を含む。データ系統マッピングテーブル 600 はさらに、目標データモデルアイデンティティに対応する「T R G _ D M _ O B J _ I D」列、目標テーブルアイデンティティに対応する「T R G _ T A B _ O B J _ I D」列、および目標列アイデンティティに対応する「T R G _ C O L _ O B J _ I D」列を含む。データ系統マッピングテーブル 600 はさらに、デー4050

タ記録についての「S R C _ T A B _ O B J _ I D」列内に格納されるソーステーブルアイデンティティによって同定されるソーステーブルについてのソーステーブルデータ記録サロゲートキーを格納する目標テーブル中の補助列についてのオブジェクトアイデンティティに対応する「S R C _ S K E Y _ C O L _ O B J _ I D」列を含む。図示される実施形態に従うと、データ系マッピングテーブル600はさらに、ある変換種類についてのソーステーブルまたは目標テーブルのいずれかの中のフィルタ列アイデンティティについての情報をオプションで与える「F I L T E R _ C O L _ O B J _ I D」列およびある変換種類についてのテーブルフィルタ列中のフィルタ値についての情報をオプションで与える「F I L T E R _ V A L U E」列を含む。

【0056】

10

図7は、発明の実施形態に従う、シャドウテーブル700のブロック図を示す。シャドウテーブル700は、ソースデータ要素から目標データ要素へ追跡する際に、データ系追跡情報の効率的なルックアップを提供することができる最適化された中央インデックスである。シャドウテーブル700は、データウェアハウス内の各データテーブル毎の別個のクエリーを作成する必要なく、データ系処理が単一のテーブル内で直接にソース記録-目標記録関係を解決できるようにすることができる。そのようなデータ系処理は、数桁より効率的であることができる。さらに、シャドウテーブル700は、データウェアハウスの1つ以上のデータテーブルからソース記録-目標記録関係を抽出することができ、データテーブル自身にクエリーすることなく、1つ以上のデータテーブルの系統を横断するためのインデックスとして働くことができる。

【0057】

20

図7に示される実施形態では、シャドウテーブル700は、目標テーブルインスタンスアイデンティティに対応する「T R G _ T A B _ O B J _ I D」列を含む。シャドウテーブル700はさらに、ソーステーブルインスタンスアイデンティティに対応する「S R C _ T A B _ O B J _ I D」列を含む。シャドウテーブル700は、ソーステーブルデータ記録サロゲートキーを格納する目標テーブル中の補助列についてのオブジェクトアイデンティティに対応する「S R C _ S K E Y _ C O L _ O B J _ I D」列をさらに含む。シャドウテーブル700はさらに、目標テーブルデータ記録のサロゲートキー値に対応する「T R G _ S K E Y _ V A L U E」および前述の目標テーブルデータ記録の補助列についての値でもあるソーステーブルデータ記録のサロゲートキー値に対応する「S R C _ S K E Y _ V A L U E」を含む。図示される実施形態に従うと、シャドウテーブル700はさらに、ある変換種類についてのテーブルフィルタ列中のフィルタ値についての情報をオプションで与える「F I L T E R _ V A L U E」列を含む。

【0058】

30

図8は、発明の実施形態に従う、データ系サブシステムの物理的モデルのブロック図を示す。図8に示される実施形態では、物理的モデルは、図6のデータ系マッピングテーブル600および図7のシャドウテーブル700を含む。物理的モデルはさらに、1つ以上の変換種類を表わすデータを含む変換種類テーブル800を含む。加えて、物理的モデルは、1つ以上のソーステーブルおよび1つ以上の目標テーブルについてのテーブル定義を含むテーブル定義構成要素810を含み、ソーステーブルおよび目標テーブルはデータテーブルの例である。図示される実施形態では、テーブル定義構成要素810は、データモデルレベルでの1つ以上のテーブル定義(たとえば、図8中の「C D R _ M O D E L」テーブル)、テーブルレベルでの1つ以上のテーブル定義(たとえば、図8中の「C D R _ T A B L E」、「C D R _ T A B L E _ C O N S」、および「C D R _ T A B L E S _ C O N S _ C O L S」)、および列レベルでの1つ以上のテーブル定義(たとえば、図8中の「C D R _ C O L U M N」および「C D R _ T A B L E S _ C O N S _ C O L S」)を含むことができる。

【0059】

40

当該実施形態に従うと、データがデータウェアハウス内で生成されるかまたはその中にロードされ、かつ1つ以上の変換がデータウェアハウスについて定義されると、(図2の

50

データ系統伝播モジュール 200などの)データ系統伝播モジュールは、データ系統サブシステムの物理的モデル内にデータ系統追跡情報を生成することができる。より具体的には、1つ以上の変換のうちの各変換毎に、データ系統伝播モジュールは、データ系統マッピングテーブル 600内に各々の目標 - ソースマッピングを格納することができる。データ系統伝播モジュールは、1つ以上の目標テーブルの1つ以上の目標データ値を1つ以上のソーステーブルの1つ以上のソースデータ値にさらにマッピングすることができる。データ系統伝播モジュールは、テーブル定義構成要素 810の1つ以上のテーブル定義を修正することによって、1つ以上の補助列を1つ以上の目標テーブルにさらに追加することができ、各々の補助列はソーステーブルを同定するソースサロゲートキーを格納することができる。さらに、1つ以上の変換のうちの各変換毎に、データ系統伝播モジュールはシャドウテーブル 700内に各々のソース - 目標マッピングを格納することができる。データ系統伝播モジュールは、1つ以上のソーステーブルの1つ以上のソースデータ値を1つ以上の目標テーブルの1つ以上の目標データ値にさらにマッピングすることができる。10

【0060】

データ系統マッピングテーブル 600内に格納される1つ以上の目標 - ソースマッピングについて、データ系統伝播モジュールは、データ系統マッピングテーブル 600中の各目標列データ値毎に以下の情報を格納することができる: すなわち、(a)ソーステーブルインスタンスオブジェクトアイデンティティおよびソース列オブジェクトアイデンティティ; (b)目標テーブルインスタンスオブジェクトアイデンティティおよび目標列オブジェクトアイデンティティ; (c)目標データ値を発生するのに用いられるソースデータ記録についてのソースサロゲートキー値を各々の補助列が格納する目標テーブルに加えられた各々の補助列のオブジェクトアイデンティティ; ならびに (d)マッピングが好ましいソースルートであることのインジケータ(たとえばフラグ)などの、マッピングに必要な任意の付加的なデータ、である。1つよりも多くの目標列の中に値をポピュレートするのに同じソースデータ記録を用いる場合、データ系統マッピングテーブル 600内の複数のマッピングシステム記録において、ソースサロゲートキー列アイデンティティを格納する同じ補助列を繰返すことができる。20

【0061】

データ系統追跡情報はデータウェアハウスの変換についてのサポートを与えることができ、例示的な変換種類は、直接マップ変換、転換変換、導出変換、統合変換、縦横ピボット変換、横縦ピボット変換、ケース変換、およびカスタム種類の変換を含む。ある実施形態では、データ系統マッピング技術を用いてデータ系統追跡情報を発生させることができる。データ系統マッピング技術は、1つ以上のソース列を目標列にマッピングすることおよび1つ以上の補助列を追加して1つ以上のソースサロゲートキーを保持することに係り、各々のソースサロゲートキーは目標列の値に寄与するソーステーブルに対応する。30

【0062】

直接マップ変換、転換変換、および導出変換は、データ系統マッピング技術を用いてデータ系統マッピングテーブル 600中に単純明快なエントリを発生させることができ、1つ以上のソース列が各々の目標列にマッピングする。これらの変換はしばしば、单一のソーステーブルから単一の目標テーブルにマッピングするので、ソースサロゲートキー値を格納する補助列が1つしかないことがしばしばである。40

【0063】

データ系統マッピング技術を用いて統合変換およびケース変換をサポートするのにもデータ系統追跡情報を用いることができる。しかしながら、データ系統マッピング技術は、変換当たり1つよりも多くの補助列がソースデータ要素への追跡を正すことを確実にするのを要件とすることができます。たとえば、1つの目標テーブルへの5つのソーステーブルの統合において、目標テーブル中の任意のデータ記録は、5つのソーステーブルのうち任意のものからであり得る。特定のデータ記録の出処を同定する際の曖昧さを回避するため、(ソーステーブル当たり1つの)5つの補助列を含むように目標テーブルを拡張することができる。統合変換を実行する場合、正しいソーステーブルにマッピングされる補助列50

の中にソースデータ記録サロゲートキーが挿入される。このように、目標テーブル中の各データ記録毎に、5つの補助列のうちの1つのみがソースサロゲートキー値を含有するであろう、また他の4つの補助列は空のままである。

【0064】

ケース変換に同じシナリオを適用可能である。ケース変換では、異なるソーステーブルからのものであり得る異なるソース列から、目標テーブルの1列中の異なるデータ記録からのデータがポピュレートされ得る。たとえば、第1のデータ記録では、目標ケース列はソースTab1.C01Xからポピュレートされ得る。第2のデータ記録では、目標ケース列はソースTab1.C01Yからポピュレートされ得る。第3のデータ記録では、目標ケース列はソースTab2.C01Xからポピュレートされ得る。各々の割当てはケース変換の条件によって制御される。このように、ソースデータ要素と目標データ要素との間を追跡するためには、目標列をポピュレートするのに用い得るソーステーブルと列との各組合せを同定しなければならない。可能なソーステーブルと列との各組合せ毎に1つの補助列が追加されるであろう。統合変換と同様に、どのケース条件がそのデータ記録について真であるかに依存して、補助列のうち1つのみが各データ記録中の値でポピュレートされる。さらに、目標列への列データマッピングが存在し、各々のマッピングソース記録は異なるソースキー列アイデンティティを含む。10

【0065】

縦横ピボット変換および横縦ピボット変換をサポートするようにデータ系統マッピング技術を修正することができる。修正は、たとえば、ピボット列についての情報および変換のために用いられるフィルタ値の一覧などの、データ系統マップテーブル600内の付加的な情報を格納することに係ることができる。たとえば、縦横ピボット変換は、各々が「ピボット」列中の異なる「フィルタ」値を有する複数のデータ記録を入力として取ることができ、目標テーブル中に「フィルタ」値の各々毎の1つ以上の列である複数の列を作成することができる。結果は、複数のソースデータ記録が単一のデータ記録に寄与した目標テーブルであり得る。したがって、目標データ記録に寄与するソースデータ記録の各々毎の1つ以上の列である複数の補助列を目標テーブルに追加することができる。20

【0066】

ある実施形態では、データ系統マッピングデータを格納するのに高度な技術を用いることができる。1つの高度な技術では、統合変換のため、データ系統マッピングデータを格納するための高度な技術は、補助列内に格納されるソースサロゲートキー内にソーステーブルインスタンスオブジェクトアイデンティティを含むことに係ることができる。目標データ記録からソースデータ記録へ遡って追跡するアルゴリズムは、目標データ記録の補助列内に格納されるソースサロゲートキー値を用いて、特定されたソーステーブル中のソースデータ記録をルックアップすることができ、ソーステーブルをソースサロゲートキー値から抽出することができる。統合変換とケース変換との両方について、データ系統追跡情報を見格納するための別の高度な技術は、「フィルタ値」を格納する「フィルタ」補助列およびソースサロゲートキーを格納する別の補助列を含むことに係ることができる。統合変換については、「フィルタ値」を格納する「フィルタ」補助列がソーステーブルインスタンスオブジェクトアイデンティティを格納する。このように、目標データ記録からソースデータ記録へ遡って追跡するアルゴリズムは、目標データ記録の補助列内に格納されるソースサロゲートキー値を用いてソーステーブル中のソースデータ記録をルックアップすることができ、ソーステーブルが「フィルタ」補助列中に同定される。ケース変換については、ユーザは、テキスト文字列である「フィルタ」値を与えて、ケース目標列をポピュレートし得るソーステーブルと列とのあらゆる組合せを同定することができる。これらの「フィルタ」値は、適切なソーステーブル/列の組合せおよびケース変換目標テーブル/列の組合せとともにデータ系統マッピングテーブル600中に格納される。目標テーブルの各データ記録において、対応の「フィルタ値」を「フィルタ」補助列の中に入力することができる。このように、目標データ記録からソースデータ記録へ遡って追跡するアルゴリズムは、「フィルタ値」およびデータ系統マッピングテーブル600内に格納される情報304050

とともに、目標データ記録からのソースサロゲートキー値を用いることができる。

【0067】

複数の補助列という要件を回避するため、データ系統追跡情報を格納するのに他の高度な技術を用いることができる。1つの高度な技術は「スマートサロゲートキー」に係り、「スマートサロゲートキー」は、その前にソーステーブルアイデンティティがついた1つ以上のソースサロゲートキー値の継続である。その結果、スマートサロゲートキーを用いてソーステーブルアイデンティティを入手することができ、したがって、ソーステーブルアイデンティティを格納する別個の補助列は必要ない。さらに、スマートサロゲートキーを格納するには単一の補助列しか必要なく、これにより、統合変換、縦横ピボット変換、および横縦ピボット変換の場合の複数の補助列の要件を除外くであろう。別の高度な技術は、縦横ピボット変換および横縦ピボット変換についてのスマートサロゲートキー内にピボット列値を「フィルタ値」として含むことに係る。10

【0068】

ある実施形態に従うと、データ要素またはデータ記録のいずれかにリンクされる関連オブジェクトの可視性をサポートし、かつ関連オブジェクトを関係のソースデータ要素および目標データ要素に伝播させるように、前述のデータ系統追跡情報または他のデータ系統追跡情報を強化することができる。さらに、データの外部ソースへの、データ系統内の変換の経路を通した関連オブジェクトのエクスポートをサポートするように、前述のデータ系統追跡情報または他のデータ系統追跡情報を強化することもできる。ある実施形態に従うと、関連オブジェクトマッピングテーブルは、関連オブジェクトとデータ要素またはデータ記録との間の1つ以上のリンクについてのデータを格納することができる。（図2のデータ系統伝播モジュールなどの）データ系統伝播モジュールは、関連オブジェクトおよびデータ要素（またはデータ記録）との関連付けを同定することができ、関連オブジェクトを一次リンクとして作成し、関連オブジェクトマッピングテーブル内に一次リンクをさらに格納することができる。データ系統伝播モジュールは、一次データ要素（またはデータ記録）のソースおよび目標データ要素／データ記録への二次リンクをさらに作成しあつ格納することができる。これは、システムのユーザがソースまたは目標データを閲覧している場合およびユーザが一次データ要素（またはデータ記録）を閲覧している場合に関連オブジェクトの効率的な視覚化を提供することができる。このように、当該実施形態に従うと、各々の関連オブジェクトが1つの一次リンクを有することができ、1つ以上の二次リンクを有することができる。2030

【0069】

図9は、発明の実施形態に従う、関連オブジェクトマッピングテーブル900のプロック図を示す。関連オブジェクトマッピングテーブル900は、関連オブジェクトとデータ要素またはデータ記録との間の1つ以上のリンクに関するデータを格納することができる。関連オブジェクトマッピングテーブル900は、関連オブジェクトとデータ要素（またはデータ記録）との間の一次リンクに関するデータを格納することができ、一次リンクは、関連オブジェクトがデータ要素（またはデータ記録）のために作成された（すなわち元々関連付けられた）ことを同定する。さらに、関連オブジェクトマッピングテーブル900は、関連オブジェクトとデータ要素（またはデータ記録）との間の二次リンクに関するデータを格納することができ、二次リンクは、データ要素（またはデータ記録）が関連オブジェクトに関連付けられる一次データ要素（またはデータ記録）のソースまたは目標データ要素（またはデータ記録）のいずれかであることを同定する。関連オブジェクトとの上流または下流データ要素の関連付けを入手するのに、関連オブジェクトとデータ要素との間の一次リンクに関するデータおよび関連オブジェクトと1つ以上のデータ要素との間の1つ以上の二次リンクに関するデータを用いることができる。さらに、データ系統を横断して関連オブジェクトの存在を検出することなく、関連付け入手することができる。40

【0070】

図9に示される実施形態では、関連オブジェクトマッピングテーブル900は、関連オブジェクトアイデンティティに対応する（すなわち、関連オブジェクトを同定する）「O50

P O O B J _ I D」列を含む。関連オブジェクトマッピングテーブル900はさらに、関連オブジェクト種類アイデンティティに対応する（すなわち、関連オブジェクトの種類を同定する）「O P O O B J _ T Y P E _ I D」列を含む。代替的な実施形態では、関連オブジェクトのアイデンティティは、アイデンティティ値および少なくとも1つの値を含む複合アイデンティティであることができる。

【0071】

図示される実施形態に従うと、関連オブジェクトマッピングテーブル900はさらに、データ要素モデルアイデンティティに対応する（すなわち、関連オブジェクトが関連付けられているデータ要素のデータテーブルのデータモデルを同定する）「D M _ O B J _ I D」列を含む。関連オブジェクトマッピングテーブル900は、データ要素テーブルアイデンティティに対応する（すなわち、関連オブジェクトが関連付けられているデータ要素のデータテーブルを同定する）「T A B _ O B J _ I D」列も含む。関連オブジェクトマッピングテーブル900はさらに、データ要素列アイデンティティに対応する（すなわち、関連オブジェクトが関連付けられているデータ要素の列を同定する）「C O L _ O B J _ I D」列を含む。関連オブジェクトマッピングテーブル900は、データ要素データ記録アイデンティティに対応する（すなわち、関連オブジェクトが関連付けられているデータ要素のデータ記録を同定する）「S K E Y _ V A L U E」列も含む。データ要素データ記録アイデンティティはサロゲートキーでもある。

【0072】

図示される実施形態に従うと、関連オブジェクトマッピングテーブル900は、関連オブジェクトとデータ要素との間の関連付けが一次リンクであるかまたは二次リンクであるかを同定する「I S _ P R I M A R Y _ R C」列も含む。関連オブジェクトマッピングテーブル900は、関連オブジェクトがデータ要素またはデータ記録と関連付けられているか否かを同定する「I S _ R E C O R D _ L E V E L _ R C」列をさらに含む。関連オブジェクトマッピングテーブル900は、データ要素が関連オブジェクトについての好みいソースデータ要素であるか否か、およびどの値がデータ系統マッピングテーブル600から受け継がれているかを同定する「P R E F E R R E D _ R O U T E _ R C」列も含む。関連オブジェクトマッピングテーブル900はさらに、関連オブジェクトとデータ要素との間のリンクが一次リンクであるか、上流二次リンクであるか、または下流一次リンクであるかを同定した「T R A C E _ T O _ R C」列を含む。

【0073】

図10は、発明の実施形態に従う関連情報サブシステムの物理的モデルのブロック図を示す。図10に示される実施形態では、物理的モデルは図9の関連オブジェクトマッピングテーブル900を含む。物理的モデルはさらに、1つ以上の関連オブジェクトを表わすデータを含む関連オブジェクト構成要素1000を含む。1つの実施形態では、関連オブジェクト構成要素1000は、1つ以上の関連オブジェクト種類を表わすデータを含むテーブル（たとえば図10中の「D M E _ O P E R A T I O N A L _ O B J E C T _ T Y P E」テーブル）を含むことができる。当該実施形態に従うと、関連オブジェクト構成要素1000はさらに、特定された関連オブジェクト種類の1つ以上の関連オブジェクトを表わすデータを含む1つ以上のテーブルを含むことができる。図示される実施形態では、「C D R _ D I S C R E P A N C Y」テーブルは、不一致である1つ以上の関連オブジェクトを表わすデータを含み、「C D R _ F L A G」テーブルは、ステータスフラグである1つ以上の関連オブジェクトを表わすデータを含み、「C D R _ C O M M E N T」テーブルは、コメントである1つ以上の関連オブジェクトを表わすデータを含む。加えて、物理的モデルは、1つ以上のソーステーブルおよび1つ以上の目標テーブルについてのテーブル定義を含むテーブル定義構成要素1010を含む。図示される実施形態では、テーブル定義構成要素1010は、データモデルレベルでの1つ以上のテーブル定義（たとえば、図10中の「C D R _ M O D E L」テーブル）、テーブルレベルでの1つ以上のテーブル定義（たとえば、図10中の「C D R _ T A B L E」テーブル）、および列レベルでの1つ以上のテーブル定義（たとえば図10中の「C D R _ C O L U M N」テーブル）を含むこ

10

20

30

40

50

とができる。

【0074】

ある実施形態に従うと、システムのユーザまたはシステム内で実行される自動プログラムは、データ要素を選択し、選択されたデータ要素に関する不一致などの関連オブジェクトを作成することができる。システムは、(図2のデータ系統伝播モジュールなどの)データ系統伝播モジュールを呼出すことができ、データ系統伝播モジュールは、新たな関連オブジェクトと(一次リンクと同定される)選択されたデータ要素との間の一次関連付けを作成することができる。データ系統伝播モジュールはさらに、関連オブジェクトマッピングテーブル900内に一次関連付け(すなわち、一次リンク)を格納することができる。データ系統伝播モジュールは、選択されたデータ要素から1つ以上のソースデータ要素および/または目標データ要素への完全なデータ系統経路を追跡することにより、新たな関連オブジェクトと1つ以上のソースデータ要素および/または目標データ要素との間の1つ以上の二次関連付けをさらに作成することができる。データ系統伝播モジュールは、関連オブジェクトマッピングテーブル900内に1つ以上の二次関連付け(すなわち、1つ以上の二次リンク)をさらに格納することができる。あるシナリオでは、ソースデータ要素へと上流に追跡する場合、複数のソースデータ要素が单一の目標データ要素にマッピングされてもよい。関連オブジェクトマッピングテーブル900中で、関連オブジェクトについての好ましいソースデータ要素であるソースデータ要素をそのようなものとして同定することができる。好ましいソースデータ要素を好ましい経路として同定することができる、好ましい経路データ系統追跡情報を(図6および図8のデータ系統マッピングテーブル600などの)データ系統マッピングテーブルの中に格納することができる。10

【0075】

ある実施形態に従うと、システムは、関連オブジェクトマッピングテーブル900内に格納される一次リンクを用いて関連オブジェクトが作成された(またはそれ以外のやり方で関連付けられた)データテーブルのデータモデルの表示において関連オブジェクトの存在を表示することができる。システムは、関連オブジェクトマッピングテーブル900内に格納される1つ以上の二次リンクを用いて上流および/または下流データモデル中の関連オブジェクトをさらに表示することができる。特定の目標データ要素について複数のソースデータ要素が存在する場合、システムは、1つ以上の好ましいソースデータ要素を好ましい経路としてマークすることができる。システムがデータモデルを表示する場合、システムは、関連オブジェクトが一次リンクまたは二次リンクによってデータ要素に関連付けられていてもいなくても、データモデルのデータテーブルの各々のデータ要素と関連付けられる1つ以上の関連オブジェクトも表示することができる。20

【0076】

1つの実施形態では、関連オブジェクトへの一次リンクを有するデータ要素を含有するデータ記録が削除されると、関連オブジェクトの状態を閉じた状態に修正することができる。関連オブジェクトについてのデータを関連オブジェクトマッピングテーブル900から取除くことができる。これに代えて、関連オブジェクトについてのデータを関連オブジェクトマッピングテーブル900内に保持することができるが、「削除済み」フラグまたはデータでのタイムスタンプを関連付けるなどして)データが「事実上削除」されたと示すようにこれを修正することができる。関連オブジェクトを削除するよりもむしろ保持して、監査目的のために保存することができる。一次リンクを介して関連オブジェクトと関連付けられたデータ要素が削除されると、関連オブジェクトへの一次およびすべての二次リンクを削除できるようになり、関連オブジェクトの状態を閉じた状態に修正することができる。二次リンクを介して関連オブジェクトと関連付けられているデータ要素が削除されると、二次リンクを削除することができる。30

【0077】

データ系統を追跡し、関連情報(すなわち関連オブジェクト)の可視性を伝播する付加的な局面がある実施形態に従ってここでより詳細に説明する。ある実施形態では、データ系統が外部システム(すなわちソースシステム)中で発生したソースデータ記録/フィー40

ルドに到達するまで、システムは、任意のデータ記録上の任意のデータフィールドからソース側データ系統中のソースデータフィールド / 記録を通って遡るユーザまたはシステムナビゲーションをサポートすることができる。システムは次に、それに対して行為をなすべきソースデータ記録（またはデータフィールド）をユーザが選べるようにすることができる。行為の例は、データ記録（またはデータフィールド）値についてソースシステムへクエリーまたはメッセージを送ること、文書を送ること、ソースシステムに対して直接または間接に行行為を行なうことができる A P I を呼出すこと、およびソースデータ記録 / フィールドを有するソースシステムのアプリケーションウインドウまたは文脈中でのソースシステム中での明示（manifestation）を開くことである。

【0078】

10

上記機能性への拡張は、各々のテーブル（または列）から直前のソーステーブル（または列）への「好ましいソース経路」を定義する変換定義メタデータに係る。この好ましい経路は、ユーザが明示的にソースデータ系統を移動する必要なく、外部システムで発生したデータ記録（またはデータフィールド）中のソース系統の単一の分岐をシステムが選択できるようにする。次にシステムは、外部システム中のそのデータ記録（またはデータフィールド）に対して上述の行為を行なうことができる。

【0079】

他のある実施形態では、表表示または表示されるフィールドを含むフォーマット済形態などのユーザインターフェイス中にデータ記録およびそれらのデータフィールドを表示する際に、関連オブジェクトマッピングテーブルなどの関連情報（すなわち、1つ以上の関連オブジェクト）の最適化されたインデックスを用いて、関連情報と関連付けられるデータ記録および / またはデータフィールドについて関連情報が存在するという、ハイライトまたはアイコンなどの視覚的指標を与えることができる。

20

【0080】

システムはまた、ユーザが、表示される情報をその情報の何らかの属性によって関連情報のサブセット（すなわち1つ以上の関連オブジェクト）に制限できるようにすることもできる。例は、（不一致、エラー報告、文書、コメントなどの）関連オブジェクトの種類と、（文書の種類、エラー報告のステータス、またはタグもしくはステータスフラグのカテゴリなどの）特定の種類の関連オブジェクトの亜種とを含む。この機能性は、システムがテーブル自体の先を見て制限を行なう必要がないように、データ記録（またはデータフィールド）をすべての他の関係のデータ記録（またはフィールド）への関連オブジェクトと関連付ける関連オブジェクトテーブル内に関連オブジェクトの種類または亜種についての情報を含むことによって最適化することができる。

30

【0081】

システムは、ユーザが、関連オブジェクトと関連付けられるデータ記録（またはデータフィールド）の関係に基づいて表示される関連情報（すなわち関連オブジェクト）を現在表示されている関連オブジェクトに制限できるようにもすることができる。例は、データ系統中の表示されるデータ要素の上流のデータ記録（またはデータフィールド）と関連付けられる関連オブジェクトのみを表示すること、表示されたデータ要素の下流のデータ記録（またはデータフィールド）と関連付けられる関連オブジェクトのみを表示すること、（一次リンクを介して）表示されたデータ要素と直接に関連付けられる関連オブジェクトのみを表示すること、または特定的なソースもしくは目標テーブルもしくは特定的な組のソーステーブルもしくは目標テーブルからの関連オブジェクトのみを表示することを含む。これは、システムがテーブル自体の先を見て制限を行なわなくてもよいように、すべての他の関係のデータ記録（またはデータフィールド）に関連オブジェクトと関連付けられるデータ記録（またはデータフィールド）を関連付ける関連オブジェクトテーブルの中に各々の関連オブジェクトの相対的な位置についての情報を含めることによって、最適化されることがある。

40

【0082】

ある実施形態に従うと、関連情報（すなわち関連オブジェクト）の視覚的指標を表示す

50

るために、特定的な表示されるデータフィールド（またはデータ記録）に各々の関連オブジェクトを効率的に関連付ける必要がある可能性がある。1つの実施形態では、システムは、各々の関連オブジェクトが表形式になるよう関連データ記録についてのアイデンティティに対してまずピボットを行なう構造化問合せ言語（SQL）の能力を用いることによって関連付けを行なうことができ、各列の名前またはアイデンティティは、表示されている情報と同じ表構造を有し、ピボットされたテーブル中のデータ要素は、関連オブジェクトが存在する場合にのみデータ値を有する。次にシステムは、表示されている実際のデータとピボットされたテーブルとの間の所与のデータ記録アイデンティティについての名前または列アイデンティティの一致を用いて、関連オブジェクトの存在の指標を効率的にオンすることができる。別の実施形態では、システムは、特定のデータ記録アイデンティティまたは特定の列アイデンティティについての情報を取出す効率的なアクセス方法を有するローカルデータ構造中に関連オブジェクトを別個に取出すことによって、関連付けを行なうことができる。次にシステムは、データ記録アイデンティティまたは列アイデンティティを用いてこのローカルデータ構造にアクセスすることにより、表示されたデータ系統中のデータ記録およびデータフィールドを効率的に同定することができる。1つの実施形態では、データ要素のデータ系統について、システムは、データ系統中の各データ記録（またはデータフィールド）毎に、1つ以上の関連オブジェクトの存在も表示することができる。10

【0083】

ある実施形態では、システムのユーザインターフェイスが1つ以上の関連オブジェクトの存在の指標を表示する場合、システムは、たとえば、マウスクリック、マウス右クリックおよびメニュー選択、または目標データ記録（もしくはデータフィールド）へのカーソルナビゲーションおよびキーボード行為を用いてユーザが特定のデータ記録（またはデータフィールド）に対する行為を惹起できるようにする。惹起された行為は、関連オブジェクトの一覧を表示することと、関連オブジェクトについての詳細へのナビゲーションとを含むことができる。たとえば、システムは、コメントが付与されている各々のデータテーブル（またはデータフィールド）毎の指標を用いて関連コメントのポップアップテーブルを表示することができる。別の例は、不一致についての詳細な情報を選択しつつ表示するのに用いることができる関連不一致の一覧を表示することであり得る。さらに、代わりに、システムは、コメントに応答するまたは不一致のステータスを変更するなど、ユーザが関連オブジェクトに対して行為を行なうことができてもできなくても、ユーザをインターフェイス中に置くことができる。20

【0084】

ある実施形態では、システムは、異なる種類の関連オブジェクトにアクセスするのに必要なユーザの権利および特権もしくはユーザインターフェイス中の関連オブジェクトの可視性を制限する特定の上流もしくは下流データテーブルに付けられる関連オブジェクトにアクセスするのに必要な特権、ユーザが利用可能な関連オブジェクトについての詳細のレベル、または関連オブジェクトに対してユーザが行為を行なうことができることについてのセキュリティ情報を用いることができる。30

【0085】

ある実施形態に従うと、システムは、上流または下流データ記録およびデータ記録と関連付けられる（不一致またはコメントなどの）関連オブジェクトと関連付けられる追跡タイムスタンプ情報を用いて、ユーザが現在閲覧しているデータの前または後のデータが上流または下流データと整合するか否かをユーザが理解できるようにすることができる。たとえば、ユーザがデータ要素を閲覧していて現在のデータ要素のデータ系統中のデータに付けられた関連コメントを示す機能を呼出す場合、システムは、ソースデータ系統中のソースデータおよび関連コメントがデータ要素よりも最近のものであるか否かを示すことができ、したがって、現在のデータテーブルにまだ伝播していない変更を反映し得る。同様に、目標データ系統において、ユーザは現在のデータよりもデータおよび関連コメントがより古いか否かを把握することができ、したがって、まだ更新されていないデータを反映4050

させて、現在のデータ要素に既になされた変更を反映し得る。同じ能力の別の明示は、ユーザが現在のデータと関連付けられる好ましいソースデータ要素に対するクエリーを提起する行為を開始して、ソースデータ要素が現在のデータ要素よりも最近修正されていることをシステムが検出してユーザに警告することである。この機能性は、ユーザが、既に修正されたかもしれないデータに対する系統設備を通して行為を行なうのを防止することができる。

【0086】

先に説明したように、システムは、サロゲートキーの使用によってデータ格納を最小限にすることことができ、最適化されたルックアップのための単一のテーブルの中にソースサロゲートキーを抽出するシャドウテーブルの使用によりルックアップ性能を最適化することができる。1つの実施形態に従うと、この設計に対する代替策はサロゲートキーの生成および使用を要件としない。この実施形態では、ソーステーブルの完全な主キーが1つ以上の目標テーブル内にソースキーとして格納され、目標テーブル上の付加的なデータフィールドが追加されて、ソース主キーフィールドと関連付けられるピボット値またはテーブルセレクタなどの、サロゲートキー中で捕捉される任意の補助情報を保持する。この実施形態では、システムメタデータは、ソースキーを構成するデータフィールドの完全な組を用いて動作し、かつサロゲートキーを用いて動作する各々のそのようなソースキーおよび任意の機能を記述する。さらに、各々の個別のテーブルをクエリーする必要なく、必要に応じて各々のソースキーの構成要素についての値を保持するのに用いられる、各ソースまたは目標キー毎に取っておかかる多数の列を有する同様のシャドウテーブルを用いることによって、最適化されたクエリー性能のためにシャドウテーブルへサロゲートキーを抽出する最適化を達成することができる。各テーブル毎のキー構造についてのシステムメタデータは次に、システムが、完全なソースキー情報を保持するように特定のテーブルが用いる列に対する適切な合体によってこの単一のシャドウテーブルに対してクエリーを構築できるようにするであろう。ソースキーの組に対する適切な継続されたインデックスと、目標キーの組に対する別の継続インデックスとは、単一のシャドウテーブルを用いて性能を最適化するであろう。

【0087】

ある実施形態では、システムは、1つ以上のアプリケーションのメタデータ記述を用いてデータ系統のランタイム要素を生成するアプリケーション生成器として働くことができる。たとえば、ユーザが1つ以上のソーステーブルと目標テーブルとの間のマッピングを定義するユーザインターフェイスが存在することができる。マッピングは、変換の種類（たとえば、合体、ピボット、またはアンピボット）に依存して異なる可能性がある。定義のプロセスは、1つ以上のソースデータ要素と1つ以上の目標データ要素との間の列レベルでのマッピングを記録することができる。次に、変換を配備する一部として、システムは、(a)目標テーブルがデータベースにインストールされる前に1つ以上のソースキーを格納する1つ以上の補助列を目標テーブルに追加する、(b)1つ以上のソースデータ要素と目標データ要素との間の関係についてのメタデータマッピング情報を記録する、かつ(c)変換を行なうように実行可能なコンピュータコードを生成する、ことができる。コンピュータコードは、1つ以上のソースデータ要素と目標データ要素との間の関係を記録するために、目標テーブル中の1つ以上のソースキーをポピュレートする論理を含むことができる。ある実施形態に従うと、生成されたコンピュータコードは、ユーザの介入なしに、同一のまたは同様の構造の新しいアプリケーションの配備において再利用することができる。

【0088】

図11は、発明の実施形態に従う、統合変換の例示的なデータ系統追跡を図示する。例示的な実施形態では、ソーステーブル「LAB1」、「LAB2」、および「LAB3」（すなわち、ソーステーブル1110、1120、および1130）が統合変換を介して目標テーブル「LAB_SUM」（すなわち目標テーブル1140）に組合される。さらに、例示的な実施形態では、（図2のデータ系統伝播モジュール200などの）データ系

10

20

30

40

50

統伝播モジュールは、目標テーブル 1140 中の以下の目標データ要素についてのソースデータ要素を見出すことができる：列「RBC」およびデータ記録アイデンティティ（すなわちサロゲートキー）「[LAB_SUM] ~ 103 ~ 1」。データ系統伝播モジュールはデータ系統マッピングテーブル 1100 をルックアップして、システム記録 9、25、および 35 という 3 つのシステム記録の中に、目標テーブルとして目標テーブル「LAB_SUM」（すなわち目標テーブル 1130）を、および目標列として列「RBC」を見出すことができる。これらの 3 つのシステム記録は、「UNION」の変換マップ種類（すなわち列「TMAP_TYPE」）を有し、それらは、データ系統マッピングテーブル 1100 の「SRC_SK EY_COLNAME」列中に示される列「LAB_SUM.UNION_SRC_KEY」中にそれらのソースサロゲートキーを格納する。目標テーブル 1140 において、データ系統伝播モジュールは、目標データ記録について列「UNION_SRC_KEY」の中を見ることができる。この列は、「[LAB1] ~ 103 ~ 1 ~ LAB1」を含有し、これはスマートサロゲートキーであるので、この値の中の第 1 の構成要素（すなわち、「[LAB1]」）がソーステーブルアイデンティティである。このように、データ系統伝播モジュールは、ソーステーブルがソーステーブル「LAB1」（すなわちソーステーブル 1110）であると判断することができる。データ系統伝播モジュールは次に、データ系統マッピングテーブル 1100 中に見出される 3 つのシステム記録に戻ることができ、ソーステーブルがソーステーブル 1110 であるという知識を備えて、データ系統伝播モジュールは、データ系統マッピングテーブル 1100 内に、正しいシステム記録、すなわちシステム記録 9 を同定することができる。このように、データ系統伝播モジュールは、サロゲートキー値「[LAB1] ~ 103 ~ 1 ~ LAB1」を用いて、ソースデータ要素が、ソーステーブル 1110 中、列「RBC」中にあると同定することができる。データ系統伝播モジュールは、ソースデータ値が「5.3」であるとさらに同定することができる。
10
20

【0089】

さらに、例示的な実施形態では、データ系統伝播モジュールは、ソーステーブル 1110 中の以下のソースデータ要素についての目標データ要素を見出すことができる：列「RBC」およびデータ記録アイデンティティ（すなわちサロゲートキー）「[LAB1] ~ 103 ~ 1 ~ LAB1」。データ系統伝播モジュールはデータ系統マッピングテーブル 1100 の中を見て、システム記録 9 の中に、ソーステーブルとしてソーステーブル「LAB1」（すなわちソーステーブル 1110）を見出し、ソース列として列「RBC」を見出すことができる。この情報から、データ系統伝播モジュールは、これが目標テーブル 1140 の「RBC」列中に目標データ要素を探す必要があると判断することができる。データ系統伝播モジュールは次に、目標テーブル「LAB_SUM」、ソーステーブル「LAB1」、「SRC_SK EY_COLNAME」「UNION_SRC_SK EY」、およびソーステーブルサロゲートキー「[LAB1] ~ 103 ~ 1 ~ LAB1」の組合せを用いて、シャドウテーブル 1150 を探索して、目標テーブル 1140 中の所望のデータ記録のサロゲートキーが「[LAB_SUM] ~ 103 ~ 1」であると判断することができる。このように、データ系統伝播モジュールは、サロゲートキー値「[LAB_SUM] ~ 103 ~ 1」を用いて、目標データ要素が、目標テーブル 1110 の中、データ記録の列「RBC」中にあると同定することができる。データ系統伝播モジュールはさらに、目標データ値が「5.3」であると同定することができる。
30
40

【0090】

図 12 は、発明の実施形態に従う、ピボット変換の例示的なソースデータ系統追跡を示す。これは、縦方向ソーステーブル「VS」（すなわちソーステーブル 1210）から横方向目標テーブル「VITALS」（すなわち目標テーブル 1220）への変換である。ソーステーブル 1210 は、生命兆候を同定するピボット列、列「VSTSTCD」を有し、生命兆候値は列「VSORES」に格納される。「VSTSTCD」列の中には、「SYSBP」、「DIASBP」、「PULSE」、「HEIGHT」、および「WEIGHT」の 5 つの異なる生命兆候値が存在する。それらはデータ系統マッピングテーブル
50

1200の中にフィルタ値（すなわち、列「F I L T E R _ V A L U E」中の値）として格納され、各々は目標テーブル1220中にその対応の列を有する。さらに、ソーステーブル1210で、各々のスマートサロゲートキー値（すなわち、列「S K」中の値）は、サロゲートキー値の最後の構成要素として、その「V S T S T C D」列からのフィルタ値を含有する。目標テーブル1220は、列「S Y S B P」、「D I A S B P」、「P U L S E」、「H T」、および「W T」である集められた5つの生命兆候の各々毎に1つの列を含む。生命兆候の値は各々の被験者毎に適切な列の中に格納される。目標テーブル1220中の各々のデータ記録は、ソーステーブル1210中の5つのデータ記録を消費する。すなわち、ソーステーブル1210中のソース列「V S T S T C D」からのフィルタ値（すなわち「S Y S B P」、「D I A S B P」、「P U L S E」、「H E I G H T」、および「W E I G H T」）は、目標テーブル1220中の列（すなわち、列「S Y S B P」、「D I A S B P」、「P U L S E」、「H T」、および「W T」）に変換される。
10

【0091】

例示的な実施形態では、（図2のデータ系統伝播モジュール200などの）データ系統伝播モジュールは、目標テーブル1220中の以下の目標データ要素についてのソースデータ要素を見出すことができる：すなわち、列「P U L S E」およびデータ記録アイデンティティ（すなわちサロゲートキー）「[V I T A L S] ~ 101 ~ 2」である。データ系統伝播モジュールは、データ系統マッピングテーブル1200をロックアップして、目標テーブル「V I T A L S」（すなわち目標テーブル1220）および目標列「P U L S E」について、システム記録をただ1つ、すなわち、（図12ではフィルタ済みデータ系統マッピングテーブル1230として同定される）システム記録3、のみを見出すことができる。このシステム記録は、ピボット列「V S T S T C D」を有するT A L L _ S K I N N Y _ S O U R C E テーブル「V S」についての「P I V O T」の変換マップ種類と、ソースデータ項目列「V S O R E S」とを有する。同じシステム記録である、（図12ではフィルタ済みデータ系統マッピングテーブル1230として同定される）システム記録3は、目標記録補助列を定義して、「V S _ S K 1」としてソース記録サロゲートキー値を得る。アルゴリズムに従うと、この列からの値（[V S] ~ 101 ~ 2）を、（図2のデータ系統伝播モジュール200などの）データ系統伝播モジュールによってフィルタ値「P U L S E」と組合せて（すなわち「[V S] ~ 101 ~ 2 ~ P U L S E」）、ソーステーブル中の正しい記録を見出すことができる。このように、データ系統伝播モジュールはソースデータ値が「68」であるとさらに同定することができる。
20
30

【0092】

図13は、発明の実施形態に従う、ピボット変換の例示的な目標データ系統追跡を示す。先に説明したように、これは、縦方向のソーステーブル「V S」（すなわち図12のソーステーブル1210）から横方向の目標テーブル「V I T A L S」（すなわち図12の目標テーブル1220）への変換である。例示的な実施形態では、（図2のデータ系統伝播モジュール200などの）データ系統伝播モジュールは、ソーステーブル1210中に以下のソースデータ要素についての目標データ要素を見出すことができる：すなわち、列「V S O R E S」およびデータ記録アイデンティティ（すなわちサロゲートキー）「[V S] ~ 101 ~ 2 ~ P U L S E」である。データ系統伝播モジュールは、図12のデータ系統マッピングテーブル1200の中を見て、5つのシステム記録であるシステム記録1、2、3、4、および5中にソース列として列「V S O R E S」を見出すことができる。データ系統伝播モジュールは次に、列「V S T S T C D」（すなわち「P U L S E」）中のソースデータ記録データ値を用いてデータ系統マッピングテーブル1200中のシステム記録をフィルタして、データ系統マッピングテーブル1200中の正しいシステム記録を判断することができる。このように、データ系統伝播モジュールは、システム記録3を（図13でフィルタ済みデータ系統マッピングテーブル1300として同定される）データ系統マッピングテーブル1200中の正しいシステム記録として同定することができる。この情報から、データ系統伝播モジュールは、これが目標テーブル1220の「P U L S E」列中に目標データ要素を探す必要があると判断することができる。ピボット変換に
40
50

についてのアルゴリズムに従うと、データ系統伝播モジュールはソースサロゲートキー値からフィルタ値を削除くことができ、次に、サロゲートキー「[V S ~ 1 0 1 ~ 2]」を用いてシャドウテーブル1310を探索して、目標テーブル1220中の所望のデータ記録のサロゲートキーが「[VITALS] ~ 1 0 1 ~ 2」であると判断することができる。このように、データ系統伝播モジュールは、サロゲートキー値「[VITALS] ~ 1 0 1 ~ 2」を用いて、目標データ要素が、目標テーブル1220の中、列「PULSE」中にあると同定することができる。データ系統伝播モジュールは目標データ値が「68」であるとさらに同定することができる。

【0093】

図14は、発明の実施形態に従う、アンピボット変換の例示的なソースデータ系統追跡を示す。これは、横方向ソーステーブル「VITALS」（すなわちソーステーブル1410）から縦方向の目標テーブル「VS」（すなわち目標テーブル1420）への変換である。ソーステーブル1410は、集められた5つの生命信号、すなわち、列「SYSBP」、「DIASBP」、「PULSE」、「HT」、および「WT」、の各々毎に1つの列を有する。生命兆候の値は各被験者毎に適切な列の中に格納される。目標テーブル1420は、生命兆候を同定するピボット列である列「VSTSTCD」を有し、生命兆候の値は列「VSORES」に格納される。「VSTSTCD」列には、5つの異なる生命兆候の値、すなわち「SYSBP」、「DIASBP」、「PULSE」、「HEIGHT」、および「WEIGHT」が存在する。これらはデータ系統マッピングテーブル1400中のフィルタ値（すなわち列「FILTER_VALUE」中の値）として格納され、各々はソーステーブル1410中にその対応の列を有する。加えて、データ系統マッピングテーブル1400において、目標テーブル「VS」は、「tall skinny」テーブルとしてマークされる。目標テーブル1420において、各々のスマートサロゲートキー値（すなわち列「SK」の値）は、サロゲートキー値の最後の構成要素としてその「VSTSTCD」列からのフィルタ値を含有する。ソーステーブル1410中の各々のデータ記録は目標テーブル1420中の5つのデータ記録に変換される。ソーステーブル1410中の5つの生命兆候列（すなわち、列「SYSBP」、「DIASBP」、「PULSE」、「HT」、および「WT」）の各々は、目標テーブル1420の列「VSTSTCD」中のフィルタ値になる（すなわち、「SYSBP」、「DIASBP」、「PULSE」、「HEIGHT」、および「WEIGHT」）。 10 20 30

【0094】

例示的な実施形態では、（図2のデータ系統伝播モジュール200などの）データ系統伝播モジュールは、目標テーブル1420中の以下の目標データ要素についてのソースデータ要素を見出すことができる：すなわち、列「VSORES」およびデータ記録アイデンティティ（すなわちサロゲートキー）「[VS] ~ 1 0 1 ~ 2 ~ PULSE」である。データ系統伝播モジュールはデータ系統マッピングテーブル1400の中を見て、5つのシステム記録であるシステム記録1、2、3、4、および5の中に、目標テーブルとして目標テーブル「VS」（すなわち目標テーブル1420）を見出し、目標列として列「VSORES」を見出すことができる。目標列「VSTSTCD」は目標テーブル「VS」ピボット列としてマークされる。データ系統伝播モジュールは次に、列「VSTSTCD」（すなわち「PULSE」）中の目標データ記録データ値を用いてデータ系統マッピングテーブル1400中のシステム記録をフィルタして、データ系統マッピングテーブル1400中の正しいシステム記録を判断することができる。このように、データ系統伝播モジュールは、（図14ではフィルタ済みデータ系統マッピングテーブル1430として同定される）データ系統マッピングテーブル1400中の正しいシステム記録としてシステム記録3を同定することができる。このシステム記録は、「UNPIVOT」の変換マップ種類（すなわち、列「TMAP_TYPE」）、「VSTSTCD」のピボット列（すなわち列「PIVOT_COL」、「VITALS」のソーステーブル（すなわち、列「SRC_TAB」）、および「PULSE」のソース列（すなわち列「SRC_COL」））を有する。目標テーブル1420のソースサロゲートキー値（すなわち列「VITAL」 40 50

S_SK1」)から、データ系統伝播モジュールは、ソースデータ記録のサロゲートキーが「[VITALS]~101~2」であると判断することができる。このように、データ系統伝播モジュールは、サロゲートキー値「[VITALS]~101~2」を用いて、ソースデータ要素が、ソーステーブル1410中、列「PULSE」中にあると同定することができる。データ系統伝播モジュールはソースデータ値が「68」であるとさらに同定することができる。

【0095】

図15は、発明の実施形態に従う、アンピボット変換の例示的な目標データ系統追跡を示す。先に説明したように、これは、横方向のソーステーブル「VITALS」(すなわち、図14のソーステーブル1410)から縦方向の目標テーブル「VS」(すなわち図14の目標テーブル1420)への変換である。例示的な実施形態では、(図2のデータ系統伝播モジュール200などの)データ系統伝播モジュールは、ソーステーブル1410中の以下のソースデータ要素についての目標データ要素を見出すことができる：すなわち、列「PULSE」およびデータ記録アイデンティティ(すなわちサロゲートキー「[VITALS]~101~2」)である。データ系統伝播モジュールは、図14のデータ系統マッピングテーブル1400をルックアップして、(図15ではフィルタ済みデータ系統マッピングテーブル1500と同定される)システム記録3中に、ソーステーブルとしてソーステーブル「VITALS」(すなわちソーステーブル1410)を見出し、ソース列として列「PULSE」を見出すことができる。このシステム記録は、「UNPIVOT」の変換マップ種類(すなわち列「TMAP_TYPE」)、「VSTSTCD」のピボット列(すなわち列「PIVOT_COL」)、「VS」の目標テーブル(すなわち列「TRG_TAB」)、および「VSORES」の目標列(すなわち列「TRG_COL」)を有する。この情報から、データ系統伝播モジュールは、それが目標テーブル1420の「VSORES」列中に目標データ要素を探す必要があると判断することができる。データ系統伝播モジュールは次に、ソーステーブル1410のサロゲートキー「[VITALS]~101~2」を用いてシャドウテーブル1510を探索して、目標テーブル1420中の所望のデータ記録のサロゲートキーが「[VS]~101~2~PULSE」であると判断することができる。シャドウテーブル中の記録の数を減らすために、(図2のデータ系統伝播モジュール200などの)データ系統伝播モジュールは、アンピボット変換のために、列「TRG_SKKEY_VALUE」中に、省略された値を、すなわちフィルタ値なしに、格納する。次にデータ系統伝播モジュールは、フィルタ値を、これを目標記録識別子として用いる前にシャドウテーブルから取出した列「TRG_SKKEY_VALUE」中の値に付ける。このように、データ系統伝播モジュールは、サロゲートキー値「[VS]~101~2~PULSE」を用いて、目標データ要素が、目標テーブル1420中、列「VSORES」中にあると同定することができる。データ系統伝播モジュールは目標データ値が「68」であるとさらに同定することができる。

【0096】

図16は、発明の実施形態に従う、合体変換の例示的なデータ系統追跡を図示する。例示的な実施形態では、ソーステーブル「DM」と「LAB1」と(すなわちソーステーブル1610と1620と)が合体されて目標テーブル「LAB1_SUM」(すなわち目標テーブル1630)を形成する。データ系統マッピングテーブル1600は、目標テーブル1630の各列毎にシステム記録を含む。各々のシステム記録は、ソーステーブルアイデンティティ(すなわち列「SRC_TAB」)、ソース列アイデンティティ(すなわち列「SRC_COL」)、およびソースサロゲートキー列アイデンティティ(すなわち列「SRC_SKKEY_COLNAME」)を含む。ソースサロゲートキー列は、目標テーブル1620中のデータ記録に寄与したソーステーブル1610(すなわち列「DM_SKKEY」)またはソーステーブル1620(すなわち列「LAB1_SKKEY」)のいずれかの中のデータ記録についてのソースサロゲートキーを格納するよう目標テーブル1630に追加される補助列である。合体変換については、この例では2つのテーブル(すなわちソーステーブル1610および1620)である合体に寄与する各テーブル毎

10

20

30

40

50

にソースサロゲートキー補助列が存在する。ソーステーブル 1610 は、列「AGE」、「AGE_U」、および「GENDER」の中の各研究被験者毎の人口統計データを表わすデータ記録を含む。ソーステーブル 1620 は、列「RBC」、「HEMOGLOBIN」、および「HEMATOCRIT」中の異なる被験者についての実験室テストの結果を表わすデータ記録を含む。これらの 2 つのテーブルは合体されて目標テーブル 1630 を作成する。目標テーブル 1630 の各々のデータ記録は、ソーステーブル 1610 および 1620 の各々のソーステーブルからの値を含む。したがって、目標テーブル 1630 の各々のデータ記録は、2 つのソースサロゲートキー値（すなわち、各々の寄与するソーステーブルからの 1 つのソースサロゲートキー値）を格納する。これらのソースサロゲートキー値は、列「DM_SKKEY」および列「LAB1_SKKEY」に格納される。

10

【0097】

例示的な実施形態では、（図 2 のデータ系統伝播モジュール 200 などの）データ系統伝播モジュールは、目標テーブル 1630 中の以下の目標データ要素についてのソースデータ要素を見出すことができる：すなわち、列「HEMOGLOBIN」およびデータ記録アイデンティティ（すなわちサロゲートキー）「[LAB1_SUM] ~ 105 ~ 1 ~ LAB1」である。データ系統伝播モジュールはデータ系統マッピングテーブル 1600 をルックアップして、システム記録 10 中に、目標テーブルとして目標テーブル「LAB1_SUM」（すなわち目標テーブル 1630）を見出し、目標列として列「HEMOGLOBIN」を見出すことができる。これから、データ系統伝播モジュールは、ソースデータ要素がソーステーブル「LAB1」（すなわちソーステーブル 1620）およびソース列「HEMOGLOBIN」の中にあると知る。データ系統伝播モジュールは、ソースサロゲートキー値が目標テーブル 1630 の列「LAB1_SKKEY」中に見出されるということも知る。目標テーブル 1630 において、データ系統伝播モジュールは、目標データ記録について列「LAB1_SKKEY」の中を見ることができる。この列はサロゲートキー値「[LAB1] ~ 105 ~ 1 ~ LAB1」を含有し、データ系統伝播モジュールは、ソーステーブル、ソース列、およびサロゲートキー値を知る。このように、データ系統伝播モジュールは、サロゲートキー値「[LAB1] ~ 105 ~ 1 ~ LAB1」を用いて、ソースデータ要素が、ソーステーブル 1620 中、列「HEMOGLOBIN」中にあると同定することができる。データ系統伝播モジュールはソースデータ値が「111」であるとさらに同定することができる。

20

【0098】

さらに、例示的な実施形態では、データ系統伝播モジュールはソーステーブル 1620 の中に以下のソースデータ要素についての目標データ要素を見出すことができる：すなわち、列「HEMOGLOBIN」およびデータ記録アイデンティティ（すなわちサロゲートキー）「[LAB1] ~ 105 ~ 1 ~ LAB1」である。データ系統伝播モジュールはデータ系統マッピングテーブル 1600 をルックアップして、システム記録 10 中に、ソーステーブルとしてソーステーブル「LAB1」（すなわちソーステーブル 1620）を見出し、ソース列として列「HEMOGLOBIN」を見出すことができる。この情報から、データ系統伝播モジュールは、それが目標テーブル「LAB1_SUM」（すなわち目標テーブル 1630）の「HEMOGLOBIN」列中に目標データ要素を探す必要があると判断することができる。データ系統伝播モジュールは次に、ソーステーブル 1620 のサロゲートキー「[LAB1] ~ 105 ~ 1 ~ LAB1」を用いてシャドウテーブル 1640 を探索して、目標テーブル 1630 中の所望のデータ記録のサロゲートキーが「[LAB1_SUM] ~ 105 ~ 1 ~ LAB1」であると判断することができる。このように、データ系統伝播モジュールは、サロゲートキー値「[LAB1_SUM] ~ 105 ~ 1 ~ LAB1」を用いて、目標データ要素が、目標テーブル 1630 中、列「HEMOGLOBIN」中にあると同定することができる。データ系統伝播モジュールは目標データ値が「111」であるとさらに同定することができる。

40

【0099】

図 17 は発明の実施形態に従うケース変換の例示的なデータ系統追跡を示す。例示的な

50

実施形態では、ソーステーブル「N R A N G E」（すなわちソーステーブル1710）は、各人口統計群毎に異なる血液検査についての正常範囲の値を含む。ソーステーブル「L A B」（すなわちソーステーブル1720）は一群の被験者についての実験室テスト結果を含む。目標テーブル「L A B _ S U M」（すなわち目標テーブル1730）は、ソーステーブル1720からの列プラスソーステーブル1710からの対応の正常範囲および正常範囲単位を含有する列を含む。目標テーブル1730は、補助列、すなわち1つのフィルタ列（すなわち列「C A S E _ F V A L」と2つのソースサロゲートキー列（すなわち、列「L A B _ S K E Y」および「N R A N G E _ S K E Y」と）をさらに含む。

【0100】

例示的な実施形態では、（図2のデータ系統伝播モジュール200などの）データ系統伝播モジュールは、目標テーブル1730中に以下の目標データ要素についてのソースデータ要素を見出すことができる：すなわち、列「R B C _ N R」およびデータ記録アイデンティティ（すなわちサロゲートキー「[L A B _ S U M] ~ 1 0 3 ~ 1」である。データ系統伝播モジュールはデータ系統マッピングテーブル1700をルックアップして、4つのシステム記録であるシステム記録11、12、13、および14の中に目標テーブルとして目標テーブル「L A B _ S U M」（すなわち目標テーブル1730）を見出し、目標列として列「R B C _ N R」を見出すことができる。これらの4つのシステム記録は「C A S E」の変換マップ種類（すなわち、列「T M A P _ T Y P E」）を有する。データ系統伝播モジュールは、目標テーブル1730中の2つの補助列を用いて、データ系統マッピングテーブル1700中の正しいシステム記録を判断する：すなわち、列「N R A N G E _ S K E Y」および列「C A S E _ F V A L」である）。列「N R A N G E _ S K E Y」はソーステーブルサロゲートキーを格納し、データ系統マッピングテーブル1700中に対応のアイデンティティ（すなわち列「S R C _ S K E Y _ C O L N A M E」）を有する。列「C A S E _ F V A L」はケース条件節タグを格納し、データ系統マッピングテーブル1700中に対応のアイデンティティ（すなわち列「F I L T E R _ C O L」）を有する。データ系統伝播モジュールは、目標テーブル1730中のデータ記録がフィルタ列（すなわちCOLUMN「C A S E _ F V A L」）中に値「M A L E」を含むと判断することができる。データ系統伝播モジュールは、データ系統マッピングテーブル1700中の正しいシステム記録が「F I L T E R _ V A L」列中の一致するアイデンティティを有するシステム記録であるとさらに判断することができる。このように、4つのシステム記録から、データ系統伝播モジュールは、「F I L T E R _ V A L」列中に「M A L E」を含むシステム記録がシステム記録11であると判断することができる。このように、ソースデータ要素は、ソーステーブル「N R A N G E」（すなわち、ソーステーブル1710）からのものであり、列「M A L E」を有する。データ系統伝播モジュールは、ソースサロゲートキー値が目標テーブル1730の「N R A N G E _ S K E Y」列中に位置し、ソースサロゲートキー値が「[N R A N G E] ~ L A B 1 ~ R B C」であるとさらに判断することができる。データ系統伝播モジュールはソースデータ値が「4 . 7 - 6 . 1」であるとさらに判断することができる。

【0101】

さらに、例示的な実施形態では、データ系統伝播モジュールは、ソーステーブル1710中に以下のソースデータ要素についての目標データ要素を見出すことができる：すなわち、列「M A L E」およびデータ記録アイデンティティ（すなわちサロゲートキー）「[N R A N G E] ~ L A B 1 ~ R B C」である。データ系統伝播モジュールはデータ系統マッピングテーブル1700をルックアップして、システム記録11中に、ソーステーブルとしてソーステーブル「N R A N G E」（すなわちソーステーブル1710）を見出し、ソース列として列「M A L E」を見出すことができる。この情報から、データ系統伝播モジュールは、それが目標テーブル「L A B _ S U M」（すなわち目標テーブル1730）の「R B C _ N R」列中に目標データ要素を探す必要があると判断することができる。データ系統伝播モジュールは次に、ソーステーブル1710のサロゲートキー「[N R A N G E] ~ L A B 1 ~ R B C」を用いてシャドウテーブル1740を探索して、目標テー

10

20

30

40

50

ブル 1 7 3 0 中の所望のデータ記録のサロゲートキーが「[LAB_SUM] ~ 103 ~ 1」および「[LAB_SUM] ~ 104 ~ 1」であると判断することができる。このように、データ系統伝播モジュールは、サロゲートキー値「[LAB_SUM] ~ 103 ~ 1」および「[LAB_SUM] ~ 104 ~ 1」を用いて、2つの目標データ要素が、目標テーブル 1 7 3 0 中、列「RBC_NR」中にあると同定することができる。データ系統伝播モジュールはこれらの目標データ要素の両者についての目標データ値が「4 . 7 - 6 . 1」であるとさらに同定することができる。

【0102】

図 18 は、発明の実施形態に従う、(図 1 の機能モジュール 18 の一部であるデータ系統モジュールなどの)データ系統モジュールの機能性のフロー図を示す。1つの実施形態では、(以下に記載される)図 18 のフロー図の機能性および(これも以下に記載される)図 19 のフロー図の機能性は、各々がメモリまたは何らかの他のコンピュータ読出可能なもしくは有形の媒体の中に格納されかつプロセッサによって実行されるソフトウェアによって実現される。他の実施形態では、各々の機能性は、(たとえば特定用途向け集積回路(A S I C)、プログラマブルゲートアレイ(P G A)、フィールドプログラマブルゲートアレイ(F P G A)などの使用を通して)ハードウェアによってまたはハードウェアとソフトウェアとの任意の組合せによって行なわれ得る。

【0103】

フローが開始し、1805 に進む。1805 で、目標データ要素が1つ以上のデータテーブルを含むデータハウスのデータ系統の一部として1つ以上のソースデータ要素にマッピングされ、各々のデータテーブルは1つ以上のデータ記録を含む。目標データ要素は目標データ記録の列を含む。各々のソースデータ要素はソースデータ記録の列を含む。データ変換を用いて、1つ以上のソースデータ要素から、目標データ要素を導出する。1つの実施形態では、データハウスは1つ以上の順序付けされたデータモデルを含む。各々の順序付けされたデータモデルは1つ以上のデータテーブルを含む。さらに、データ変換を用いて、先のデータモデルの少なくとも1つのテーブルから、後続のデータモデルの1つ以上のデータテーブルを形成する。1つの実施形態では、データ変換は、直接マッピング変換、統合変換、合体変換、横縦アンピボット変換、縦横ピボット変換、またはケース変換のうちの1つである。フローは1810 に進む。

【0104】

1810 で、目標データ記録の1つ以上の補助列内に1つ以上のソースサロゲートキーが格納される。各々のソースサロゲートキーは、対応のソースデータ要素を含むソースデータ記録のアイデンティティを含む。1つの実施形態では、1つ以上のサロゲートキーのうちの各々のサロゲートキーは、ソースデータ記録を備えるソースデータテーブルのアイデンティティを含む。1つの実施形態では、1つ以上のソースデータ要素への目標データ要素のマッピングに基づいて、1つ以上の補助列が目標データ記録に自動的に追加される。フローは1815 に進む。

【0105】

1815 で、各ソースデータ要素毎に、データ系統マッピングシステム記録が目標データ要素および対応のソースデータ要素のマッピングを表わすデータ系統マッピングシステムテーブル内に格納される。データ系統マッピングシステム記録は、対応のソースデータ要素、目標データ要素、および1つ以上の補助列を含む。当該実施形態に従うと、データ系統マッピングシステム記録はデータ変換の種類も含む。フローは1820 に進む。

【0106】

1820 で、ソースデータ要素は、データ系統の一部として1つ以上の目標データ要素にマッピングされる。1つ以上の目標データ要素は、データ変換を用いてソースデータ要素から導出される。フローは1825 に進む。

【0107】

1825 で、各目標データ要素毎に、ソースデータ要素および対応の目標データ要素のマッピングを表わすシャドウシステムテーブル内にシャドウシステム記録が格納される。

10

20

30

40

50

シャドウシステム記録は、ソースデータ要素を同定するソースサロゲートキー、対応の目標データ要素を同定する目標サロゲートキー、および1つ以上の補助列を含む。このように、当該実施形態に従うと、データ系統は、1つ以上のデータ系統マッピングシステム記録、1つ以上のシャドウシステム記録、および1つ以上のソースサロゲートキーを含む。フローは1830に進む。

【0108】

1830で、データ系統を用いてマッピングされる1つ以上のデータ要素からデータ要素が選択される。フローは1835に進む。

【0109】

1835で、選択されたデータ要素にマッピングされる1つ以上のデータ要素がデータ系統を用いてユーザインターフェイス内に表示される。1つの実施形態では、選択されたデータ要素にマッピングされる1つ以上のデータ要素の表示はさらに、各データ要素毎に、データ系統内のデータ要素の位置を表示することを含む。フローは1840に進む。

【0110】

1840で、各目標データ要素毎に、1つ以上のソースデータ要素のうちの1つのソースデータ要素が好ましいソースデータ要素として選択される。フローは1845に進む。

【0111】

1845で、選択された1つ以上の好ましいソースデータ要素に基づいて、データ系統の好ましいソース経路が作成される。フローは1850に進む。

【0112】

1850で、目標データ記録の補助列内にピボット列アイデンティティが格納される。ピボット列アイデンティティはソースデータ記録のピボット列を同定する。フローは1855に進む。

【0113】

1855で、目標データ記録の第1の補助列内にフィルタ列アイデンティティが格納される。フィルタ列アイデンティティはソースデータ記録のフィルタ列を同定する。フローは1860に進む。

【0114】

1860で、目標データ記録の第2の補助列内にフィルタ値が格納され、フィルタ値はソースデータ記録のフィルタ値である。そしてフローが終了する。

【0115】

図19は、発明の実施形態に従う、(図1の関連情報伝播モジュール16などの)関連情報伝播モジュールの機能性のフロー図を示す。フローは1905で始まる。1905で、データウェアハウスのデータ系統追跡が追跡される。データウェアハウスは1つ以上のデータテーブルを含む。各々のデータテーブルは1つ以上のデータ記録を含む。1つの実施形態では、データウェアハウスは1つ以上の順序付けされたデータモデルを含む。各々の順序付けされたデータモデルは1つ以上のデータテーブルを含む。さらに、データ変換を用いて、先のデータモデルの少なくともテーブルから、後続のデータモデルの1つ以上のデータテーブルが形成される。1つの実施形態では、データ変換は、直接マッピング変換、統合変換、合体変換、横縦アンピボット変換、縦横ピボット変換、またはケース変換のうちの1つである。フローは1910に進む。

【0116】

1910で、一次データ要素と関連オブジェクトとの間の関連付けが同定される。一次データ要素は一次データ記録の列を含む。関連オブジェクトは関連情報を備える。一次データ要素と関連オブジェクトとの間の関連付けは一次リンクとして同定される。1つの実施形態では、関連オブジェクトは、不一致、コメント、ステータスフラグ、状態インジケータ、タグ、画像、録音録画、ファイル、リンク、または文書の1つである。また、1つの実施形態では、一次データ要素は、一次データ記録の列を含む一次データ要素の代わりに一次データ記録を含む。

【0117】

10

20

30

40

50

1915で、一次リンクを表わす関連オブジェクトマッピングシステムテーブル内に第1の関連オブジェクトマッピングシステム記録が格納される。第1の関連オブジェクトマッピングシステム記録は、一次データ要素のアイデンティティ、関連オブジェクトのアイデンティティ、および一次リンクインジケータを含む。1つの実施形態では、第1の関連オブジェクトマッピングシステム記録は、一次データ要素のデータモデルを同定するデータ要素モデルアイデンティティ、一次データ要素のデータテーブルを同定するデータ要素テーブルアイデンティティ、および一次データ要素の列を同定するデータ要素列アイデンティティを含む。フローは1920に進む。

【0118】

1920で、データ系統を用いて、二次データ要素と関連オブジェクトとの間の関連付けが同定される。二次データ要素は二次データ記録の列を含む。二次データ要素はデータ系統内の一次データ要素に関する。二次データ要素と関連オブジェクトとの間の関連付けは二次リンクとして同定される。1つの実施形態では、二次データ要素は一次データ要素のソースデータ要素である。また、1つの実施形態では、一次データ要素のアイデンティティおよび二次データ要素のアイデンティティは各々サロゲートキーである。さらに、1つの実施形態では、二次データ要素は、二次データ記録の列を含む二次データ要素の代わりに二次データ記録を含む。1つの実施形態では、二次データ要素と関連オブジェクトとの間の関連付けは、データ系統を用いて一次データ要素に基づいて二次データ要素を同定することによって同定される。ある実施形態では、付加的な二次データ要素と関連オブジェクトとの間の付加的な関連付けもデータ系統を用いて同定される。フローは1925に進む。

【0119】

1925で、二次リンクを表わし、かつ一次データ要素と二次データ要素との間の関連付けを表わす関連オブジェクトマッピングシステムテーブル内に第2の関連オブジェクトマッピングシステム記録が格納される。第2の関連オブジェクトマッピングシステム記録は、二次データ要素のアイデンティティ、関連オブジェクトのアイデンティティ、および二次リンクインジケータを含む。1つの実施形態では、第2の関連オブジェクトマッピングシステム記録は、二次データ要素のデータモデルを同定するデータ要素モデルアイデンティティ、二次データ要素のデータテーブルを同定するデータ要素テーブルアイデンティティ、および二次データ要素の列を同定するデータ要素列アイデンティティを含む。さらに、1つの実施形態では、第2の関連オブジェクトマッピングシステム記録は、二次データ要素が好ましいソースデータ要素であるという指標を含む。データ系統を用いて付加的な二次データ要素と関連オブジェクトとの間の付加的な関連付けも同定される実施形態では、関連オブジェクトマッピングシステムテーブル内に付加的な関連オブジェクトマッピングシステム記録が格納される。フローは1930に進む。

【0120】

1930で、一次データ要素または二次データ要素がユーザインターフェイス内に表示される。1つの実施形態では、一次データ要素と二次データ要素との両方がユーザインターフェイス内に表示される。次にフローは1935に進む。

【0121】

1935で、一次データ要素または二次データ要素のいずれかが選択される。次にフローは1940に進む。

【0122】

1940で、第1の関連オブジェクトマッピングシステム記録または第2の関連オブジェクトマッピングシステム記録のいずれかを用いてユーザインターフェイス内に関連オブジェクトが表示される。より具体的には、一次データ要素が選択されると、第1の関連オブジェクトマッピングシステム記録を用いてユーザインターフェイス内に関連オブジェクトが表示される。二次データ要素が選択されると、第2の関連オブジェクトマッピングシステム記録を用いてユーザインターフェイス内に関連オブジェクトが表示される。1つの実施形態では、関連オブジェクトの表示は関連オブジェクトの種類によってフィルタされ

10

20

30

40

50

る。フローは 1945 に進む。

【0123】

1945 で、一次データ要素が削除される。フローは 1950 に進む。

1950 で、第 1 の関連オブジェクトマッピングシステム記録が削除される。フローは 1955 に進む。

【0124】

1955 で、第 2 の関連オブジェクトマッピングシステム記録が削除される。フローは 1960 に進む。

【0125】

1960 で、関連オブジェクトの状態が閉じた状態に修正される。フローは終了する。 10

このように、1つの実施形態では、データウェアハウス内に格納することができる1組のデータのデータ系統追跡を与えることができるシステムを提供することができ、1組のデータは1つ以上のテーブル内に格納され、各々のテーブルは1つ以上のデータ記録を含む。所与の目標データ要素については、システムは、目標データ要素を発生した1組のソースデータ要素を同定することができ、「データ要素」はデータ記録の列を含む。同様に、所与のソースデータ要素については、システムは、ソースデータ要素から導出される1組の目標データ要素およびソースデータ要素から1組の目標データ要素への経路を同定することができる。システムによって与えられるデータ経路追跡は格納されたデータの徹底的な分析をさらに容易にすることができる。

【0126】

さらに、別の実施形態では、データ要素にリンクされる「関連情報」または「関連オブジェクト」の可視性をサポートし、関係のソースデータ要素および目標データ要素に関連情報（すなわち関連オブジェクト）を伝播するようにデータ系統追跡を強化することができるシステムを提供することができる。実施形態に従うと、システムは、関連オブジェクトと、関連オブジェクトに関連付けられるデータ要素またはデータ記録との間の一次リンクおよび二次リンクの両者を格納することができる。このように、当該実施形態に従うと、データ系統のデータ要素の上流およびデータ要素の下流の両方で、不一致などの関連オブジェクトが可視となることができる。この関連オブジェクトの可視性は、格納されたデータの徹底的な分析をさらに容易にすることができる。 20

【0127】

この明細書を通じて記載した発明の特徴、構造、または特性を、1つ以上の実施形態において任意の好適な態様で組合せててもよい。たとえば、この明細書を通じての「1つの実施形態」、「いくつかの実施形態」、「ある実施形態」、「ある複数の実施形態」、または他の同様の文言の使用は、実施形態と関連して記載した特定の特徴、構造、または特性が本発明の少なくとも1つの実施形態に含まれてもよいことを指すものである。このように、この明細書を通じての句「1つの実施形態」、「いくつかの実施形態」、「ある実施形態」、「ある複数の実施形態」、または他の同様の文言の出現は同じ群の実施形態を必ずしもすべて指すものではなく、記載の特徴、構造、または特性を1つ以上の実施形態で任意の好適な態様で組合せててもよい。 30

【0128】

当業者は、以上で論じたような発明が、開示されるのとは異なる順序のステップで、および/または異なる構成中の要素を用いて実践されてもよいことを容易に理解するであろう。したがって、発明はこれらの好ましい実施形態に基づいて記載されたが、当業者には、発明の意味および精神内に留まりながら、ある修正例、変形例、および代替的な構築が明らかになるであろうことが明らかであろう。したがって、発明の境界を判断するには、添付の請求項を参照すべきである。 40

【図1】

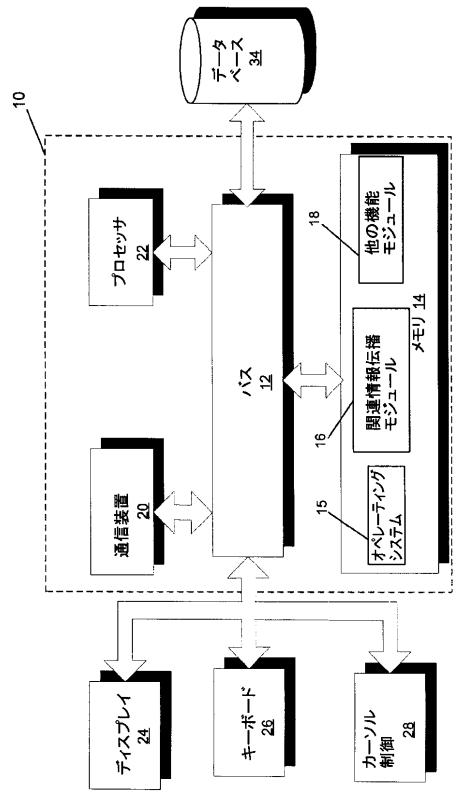


Fig. 1

【図2】

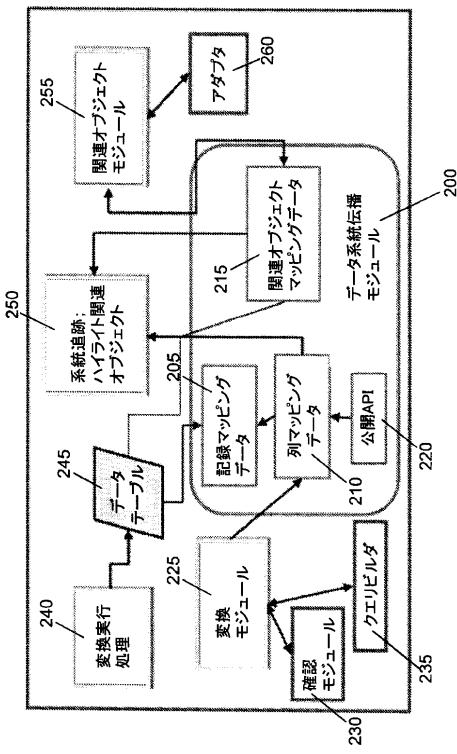


Fig. 2

【図3】

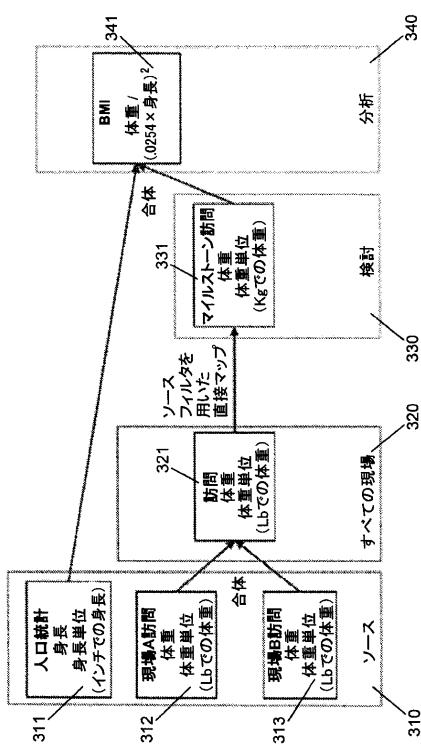


Fig. 3

【図4】

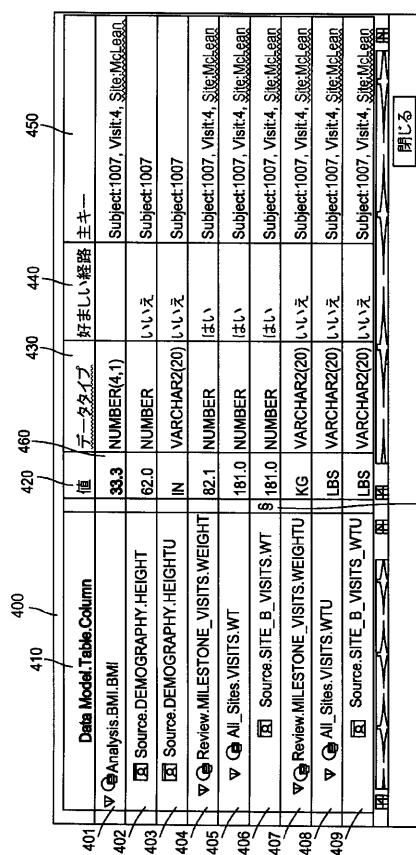


Fig. 4

【図5】

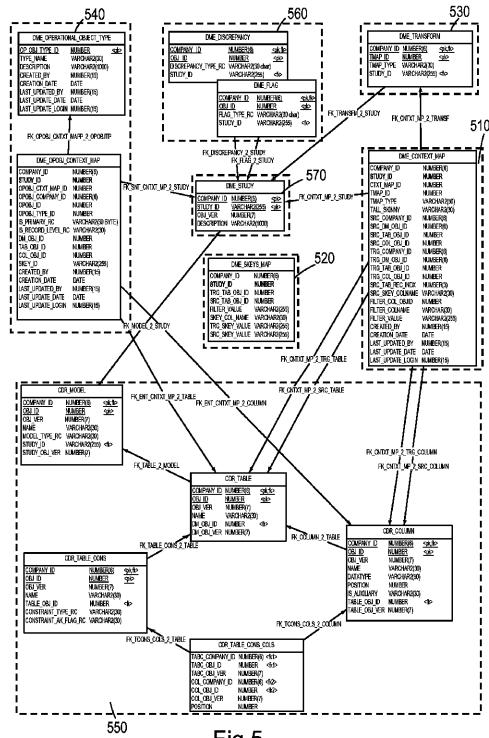


Fig.5

【図6】

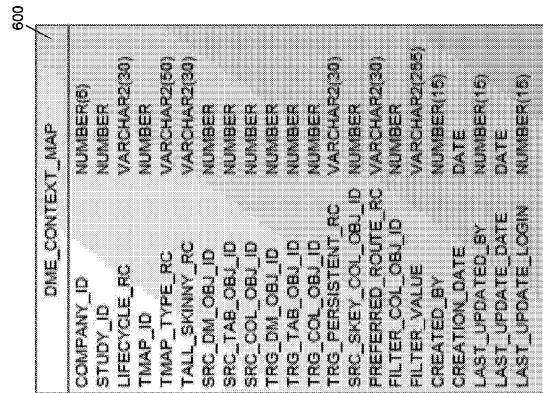


Fig. 6

【図7】

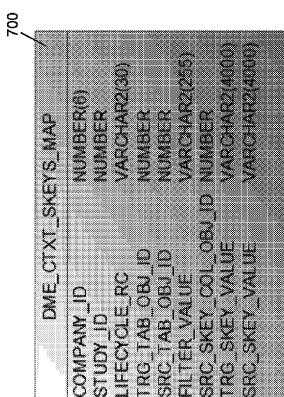


Fig. 7

【図8】

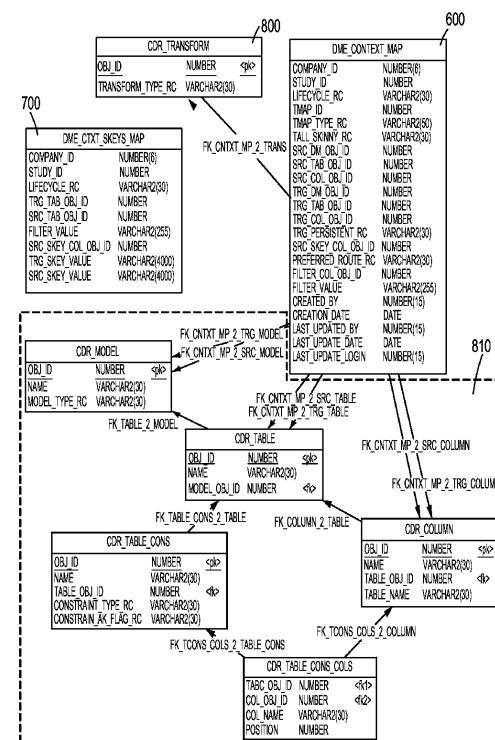


Fig.8

【図9】

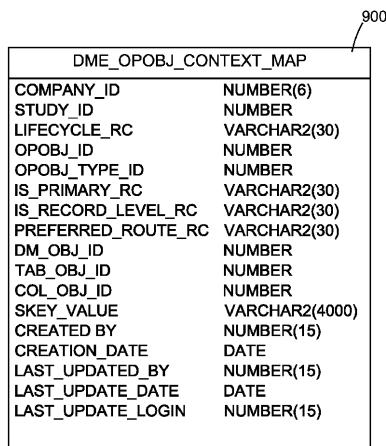


Fig.9

【図10】

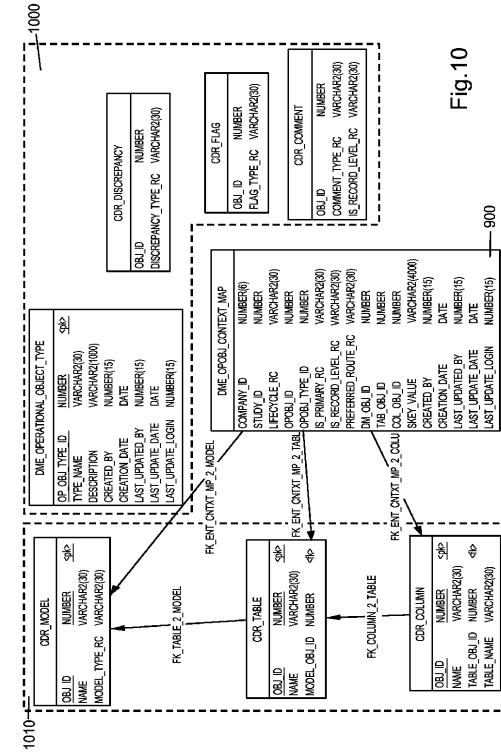


Fig. 10

【 図 1 1 】

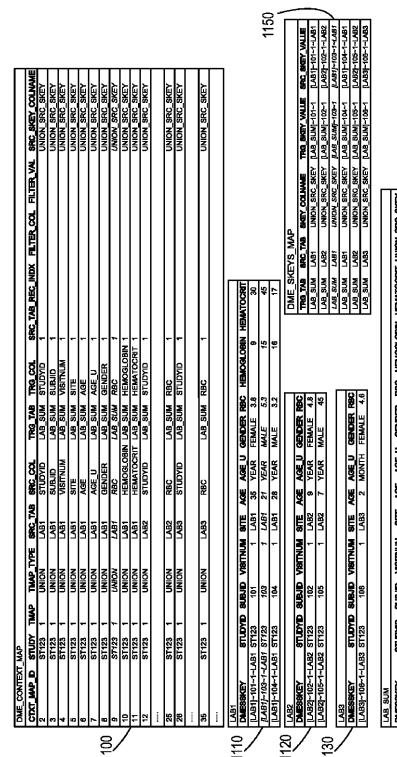
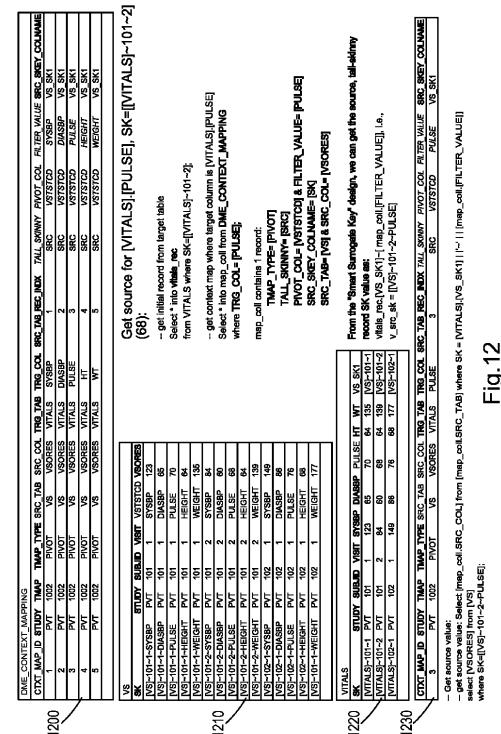


Fig. 11

【図12】



【図17】

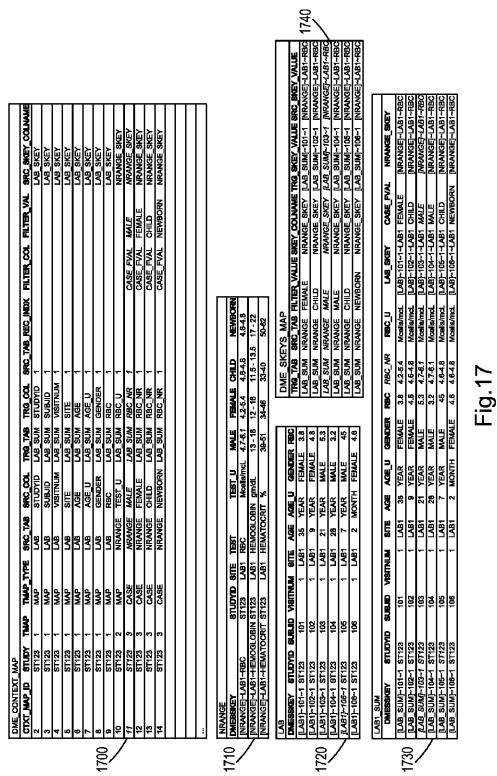


Fig. 17

【 図 1 8 】

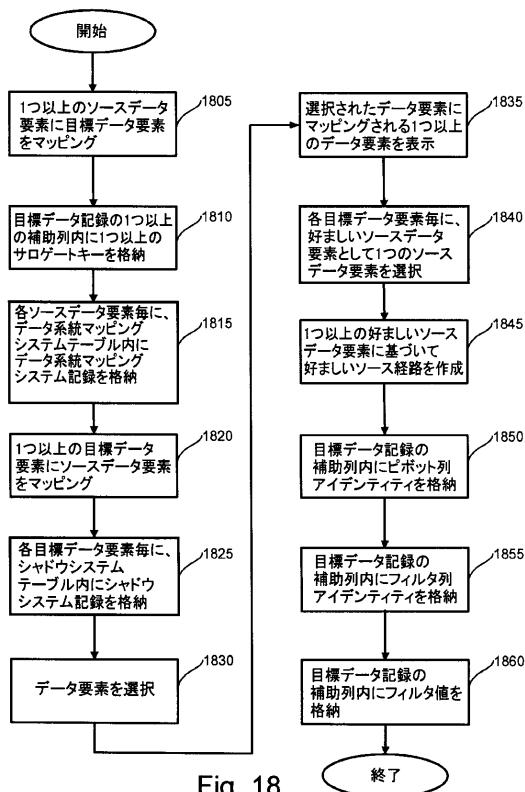
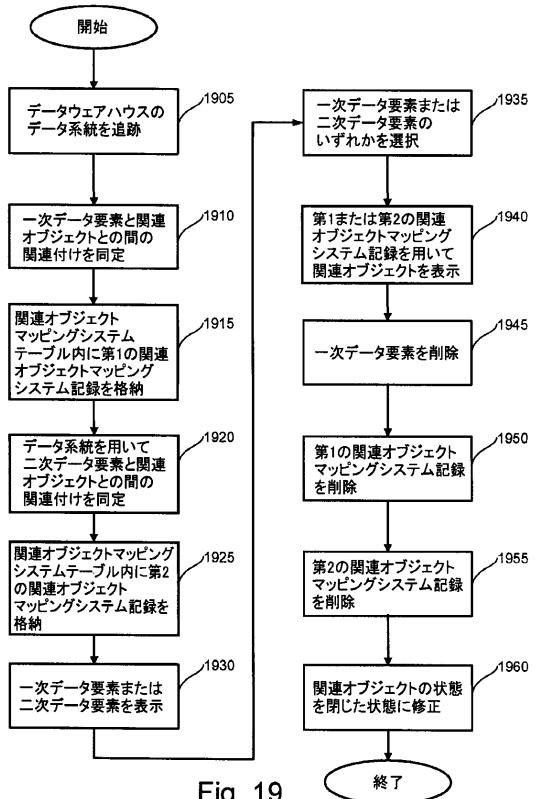


Fig. 18

【 図 19 】



フロントページの続き

(72)発明者 リース , ジョン・ケイ

アメリカ合衆国、02446 マサチューセッツ州、ブルックライン、ワインチェスター・ストリート、126

審査官 田中 秀樹

(56)参考文献 米国特許第06343295(US, B1)

米国特許第07117219(US, B1)

米国特許出願公開第2011/0313969(US, A1)

(58)調査した分野(Int.Cl., DB名)

G06F 17/30

G06F 12/00