



## (12)发明专利

(10)授权公告号 CN 105706084 B

(45)授权公告日 2019.08.06

(21)申请号 201480060691.2

(22)申请日 2014.09.01

(65)同一申请的已公布的文献号

申请公布号 CN 105706084 A

(43)申请公布日 2016.06.22

(30)优先权数据

14/020,517 2013.09.06 US

(85)PCT国际申请进入国家阶段日

2016.05.05

(86)PCT国际申请的申请数据

PCT/US2014/053628 2014.09.01

(87)PCT国际申请的公布数据

W02015/034795 EN 2015.03.12

(73)专利权人 超级医疗系统公司

地址 美国加利福尼亚

(72)发明人 B·伯恩斯 C·马赞蒂

R·迈库姆贝尔 J·米勒

(74)专利代理机构 永新专利商标代理有限公司  
72002

代理人 刘瑜 王英

(51)Int.Cl.

G06F 16/21(2019.01)

G06Q 10/06(2012.01)

G06F 16/25(2019.01)

(56)对比文件

US 2007050394 A1,2007.03.01,

CN 102262707 A,2011.11.30,

US 8738569 B1,2014.05.27,

CN 1790322 A,2006.06.21,

CN 1508728 A,2004.06.30,

CN 103246733 A,2013.08.14,

US 2008071730 A1,2008.03.20,

CN 104516912 A,2015.04.15,

US 2012005241 A1,2012.01.05,

US 2006195492 A1,2006.08.31,

审查员 张雪锋

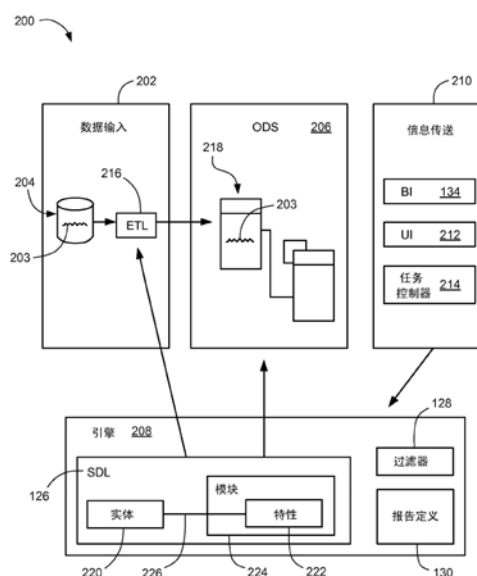
权利要求书3页 说明书25页 附图15页

(54)发明名称

元数据自动化系统

(57)摘要

一种方法可以包括:提供用于定义链接到实体的特性观察的模式定义语言、以及与元数据在模块中分组在一起的特性观察;基于特性观察中的至少一个来生成针对其间具有链接的模块和实体的物理表;以及根据元数据利用数据来填充物理表。



1. 一种用于数据管理的方法,包括:

从计算机存储器向处理器提供用于定义链接到实体的特性观察的模式定义语言、以及与元数据在模块中分组在一起的所述特性观察,所述特性观察是关于所述实体的能够观察的事实;

捕获模式定义模型表,所述模式定义模型表包括用于每个特性观察的行,所述模式定义模型表包括用于实体类型和模块类型的列,所述实体类型定义所述特性观察所链接到的实体的类型,并且所述模块类型定义所述特性观察所链接到的模块的类型,并且所述模式定义模型表包括包含所述元数据的单元;

在能够操作的数据存储中,基于所述特性观察中的至少一个来生成针对所述模式定义模型表内的模块中的每一个的单独物理表,所述单独物理表包括针对所述模式定义模型表中的所述特性观察中的每一个的列,在所述模块和所述实体之间具有链接;以及

根据所述元数据,利用数据来填充所述单独物理表。

2. 根据权利要求1所述的方法,还包括引用所述模式定义语言的所述元数据来在所述单独物理表内定位所选择的特性,并且确定针对所述模块或所述实体的所述单独物理表是否包括所述所选择的特性。

3. 根据权利要求2所述的方法,还包括引用所述模式定义语言的所述元数据来定位所述单独物理表,并且仅在针对所述模块或所述实体的所述单独物理表包括所述所选择的特性的情况下,确定针对所述模块或所述实体的所述单独物理表是否包括后续选择的特性。

4. 根据权利要求1所述的方法,还包括,如果所述特性观察与所选择的特性被分组在同一个模块内,则确定所述特性观察是单跳链接。

5. 根据权利要求1所述的方法,还包括:

当所选择的特性的临床模式与所述特性观察相对应时,将所述实体分类在相匹配的队列中;或者

当所述所选择的特性的所述临床模式不与所述特性观察相对应时,则将所述实体分类为不相匹配的队列。

6. 根据权利要求1所述的方法,还包括:

确定所述模块是否与报告数据定义相关联;以及

如果所述模块不与所述报告数据定义相关联,则引用所述模式定义语言。

7. 根据权利要求1所述的方法,其中,填充所述单独物理表包括引用所述模式定义语言的所述元数据,并且确定所述数据是否是针对所述特性观察的合适的数据类型。

8. 一种能够与处理器关联使用的非暂时性计算机可读介质,包括被配置以执行以下操作的指令:

提供用于定义链接到实体的特性观察的模式定义语言、以及与元数据在模块中分组在一起的所述特性观察,所述特性观察是关于所述实体的能够观察的事实;

捕获模式定义模型表,所述模式定义模型表包括用于每个特性观察的行,所述模式定义模型表包括用于实体类型和模块类型的列,所述实体类型定义所述特性观察所链接到的实体的类型,并且所述模块类型定义所述特性观察所链接到的模块的类型,并且所述模式定义模型表包括包含所述元数据的单元;基于所述特性观察中的至少一个来生成针对所述模式定义模型表内的模块中的每一个的单独物理表,所述单独物理表包括针对所述模式定

义模型表中的所述特性观察中的每一个的列,在所述模块和所述实体之间具有链接;以及根据所述元数据,利用数据来填充所述单独物理表。

9. 根据权利要求8所述的非暂时性计算机可读介质,还包括被配置为执行以下操作的指令:引用所述模式定义语言的所述元数据来在所述单独物理表内定位所选择的特性,并且确定针对所述模块或所述实体的所述单独物理表是否包括所述所选择的特性。

10. 根据权利要求9所述的非暂时性计算机可读介质,还包括被配置为执行以下操作的指令:引用所述模式定义语言的所述元数据来定位所述单独物理表,并且仅在针对所述模块或所述实体的所述单独物理表包括所述所选择的特性的情况下,确定针对所述模块或所述实体的所述单独物理表是否包括后续选择的特性。

11. 根据权利要求8所述的非暂时性计算机可读介质,还包括被配置为执行以下操作的指令:如果所述特性观察与所选择的特性被分组在同一个模块内,则确定所述特性观察是单跳链接。

12. 根据权利要求8所述的非暂时性计算机可读介质,还包括被配置以执行以下操作的指令:

当所选择的特性的临床模式与所述特性观察相对应时,将所述实体分类在相匹配的队列中;或者

当所述所选择的特性的所述临床模式不与所述特性观察相对应时,则将所述实体分类为不相匹配的队列。

13. 根据权利要求8所述的非暂时性计算机可读介质,还包括被配置为执行以下操作的指令:

确定所述模块是否与报告数据定义相关联;以及

如果所述模块不与所述报告数据定义相关联,则引用所述模式定义语言。

14. 根据权利要求8所述的计算机可读介质,其中,被配置为填充所述单独物理表的所述指令进一步包括被配置为引用所述模式定义语言的所述元数据并且确定所述数据是否是针对所述特性观察的合适的数据类型的指令。

15. 一种用于数据管理的系统,包括:

管理引擎,其被配置为:

从计算机存储器向处理器提供用于定义链接到实体的特性观察的模式定义语言、以及与元数据在模块中分组在一起的所述特性观察,所述特性观察是关于所述实体的能够观察的事实;

捕获模式定义模型表,所述模式定义模型表包括用于每个特性观察的行,所述模式定义模型表包括用于实体类型和模块类型的列,并且所述模式定义模型表包括包含所述元数据的单元;

在能够操作的数据存储中,基于所述特性观察中的至少一个来生成针对所述模式定义模型表内的模块中的每一个的单独物理表,所述单独物理表包括针对所述模式定义模型表中的所述特性观察中的每一个的列,在所述模块和所述实体之间具有链接;以及

数据输入块,耦合到所述管理引擎,被配置为根据所述元数据来利用数据填充所述单独物理表。

16. 根据权利要求15所述的系统,还包括信息传递块,其被配置为引用所述模式定义语

言的所述元数据来在所述单独物理表内定位所选择的特性,并且确定针对所述模块或所述实体的所述单独物理表是否包括所述所选择的特性。

17. 根据权利要求16所述的系统,其中,所述信息传递块被配置为引用所述模式定义语言的所述元数据来定位所述单独物理表,并且仅在针对所述模块或所述实体的所述单独物理表包括所述所选择的特性的情况下,确定针对所述模块或所述实体的所述单独物理表是否包括后续选择的特性。

18. 根据权利要求15所述的系统,其中,所述管理引擎被配置为如果所述特性观察与所选择的特性被分组在同一个模块内,则确定所述特性观察是单跳链接。

19. 根据权利要求15所述的系统,其中,所述管理引擎被配置为执行以下操作:

当所选择的特性的临床模式与所述特性观察相对应时,将所述实体分类在相匹配的队列中;或者

当所述所选择的特性的所述临床模式不与所述特性观察相对应时,则将所述实体分类为不匹配的队列。

20. 根据权利要求15所述的系统,其中,所述管理引擎被配置为执行以下操作:

确定所述模块是否与报告数据定义相关联;以及

如果所述模块不与所述报告数据定义相关联,则引用所述模式定义语言。

## 元数据自动化系统

### 发明领域

[0001] 发明的实施例总体上涉及计算机数据库,并且更加特别地,有利于对数据结构的管理,包括其定义和创建、修改、转换和填充(population)。

### 背景技术

[0002] 信息的自动处理对商业而言已经是极大的好处,因为在决策路径的每个点上这都大大提高了决策制定者的有效性和效率。每个企业,无论是政府、商用业务还是非盈利组织,都有管理信息的运营的必要性。

[0003] 例如,在商用业务的情况下,该信息用于治疗病人、获得客户、输入订单、运送产品、向客户开账单、收集发票、向员工和供应商付款、订购产品、审计存货并且保存雇员、用户和供应商之间的交易记录。

[0004] 在正常的事件的过程中,根据每个组织的内部的运营模型,利用软件、计算机硬件、以及数字网络来获取、处理、和整合信息。不幸的是,信息的自动处理充满许多性能不佳的问题,这些问题阻止了可用的、及时的、并且成本有效的数据集成、标准化、以及报告。

[0005] 之前的一个方法专注于构建企业数据仓库以从整个组织中收集整合的和标准化的数据。一般的企业数据仓库要求来自许多源的运营数据被提取、转换、并且加载到第三范式的运营数据存储数据库中,其被再一次提取、转换、并且加载到星形和雪花数据仓储(vault)数据库中。数据仓储数据库然后可以被加载到数据集中,每个专用于特定的部门或功能。

[0006] 企业数据仓库构造和运作处理中的每个数据库都必须被设计、维护、并且填充有定制提取转换加载(ETL)功能。此外,在组织能够生成报告并且开始从企业数据仓库中实现收益之前,开发和使用中的所有阶段都必须以某种形式被完成。

[0007] 尽管企业数据仓库实现了针对整个组织集中管理的标准化数据,但是这产生非常高的成本。实现全面的企业数据仓库所要求的资源可能对于除了极少数之外的所有的都是难以负担的,因为货币成本可能是天文数字。甚至当货币、资源不是限制因素时,用于建立和实现企业数据仓库的时间通常也是以年来度量的。

[0008] 企业数据仓库的另一个缺点起源于企业数据仓库对决策支持应用的专注,其强调摘要信息。这些系统的固有缺点是与客户的标识有关的交易细节丢失。当被应用到诸如客户数据分析等的应用时,企业数据仓库显示出缺点。客户数据分析是将数据与客户的动作、事件、交易、状态等相关联的决策支持分析。摘要信息通常丢失与客户标识有关的细节层面的信息,限制了在这些应用中的企业数据仓库方法的有效性。

[0009] 其它方法专注于直接从组织运营数据来创建部门集中的数据集市。部门集中的数据集市仅需要包含与单个部门相关的数据。由于这点,部门集中的数据集市可能小得多。

[0010] 由于较小的规模,部门集中的数据集市在时间和金钱方面通常花费较少的资源来建立;然而,这些益处也产生过高的成本。部门集中的数据集市不是集中管理的,并且在质量或数据格式方面不具有一致的标准。

[0011] 当部门集中的数据集市被创建时,不一致的标准阻止跨组织的集成。此外,由于每个部门集中的数据集市都是在没有贯穿性的计划的情况下被创建的,因此,要具有数据集市针对每个部门所投资的资源的总量会极大地高于创建单个的计划周详的企业数据仓库。用于维护不一致的部门集中的数据集市的资源也会大大高于维护单个企业数据仓库。

[0012] 当前,没有用于解决提供可用的、及时的、并且成本有效的数据集成、标准化、以及报告的问题的全面的解决方案。在工业中一直以来都感受到这样的需求。

[0013] 之前的发展没有教导或提出任何解决方案来克服上文中所描述的所有的限制,并且因此,本领域技术人员长期没有得到用于克服这些限制的解决方案。

## 发明内容

[0014] 所要求保护的发明针对利用模式定义语言的方法、制品、以及系统,所述模式定义语言具有用于定义链接到实体的特性的元数据并且使所述特性在模块中分组在一起。可以预期,利用模式定义语言来生成针对其间具有链接的模块和实体的并且基于所述特性的物理表。该物理表可以填入符合元数据的数据。

[0015] 还可以预期,可以引用模式定义语言以定位物理表并且确定针对模块或实体的物理表是否包括所选择的特性。同样地,仅在物理表包括所选择的特性的情况下,可以引用模式定义语言来定位物理表并且确定物理表是否包括后续选择的特性。

[0016] 还可以预期,如果特性与所选择的特性被分组在同一个模块内,则可以根据模式定义语言来确定单跳(one-hop)链接。还可以预期,可以基于临床模式是否和与实体相关联的特性相匹配,将实体分类到相匹配的队列或不相匹配的队列中。

[0017] 还可以预期,如果与模块相关联的特性被包括在报告数据定义中,则报告数据定义可以引用模式定义语言以包括该模块。还可以预期,物理表可以分别包括与模块和实体相关联的长度数据和非长度数据。

## 附图说明

[0018] 为了便于可以获得并详细理解本发明的上述特征、优点和目标的方式,可以参照附图中示出的本发明的实施例来得到上文简要概述的本发明的更特别的描述。

[0019] 然而,应当注意的是,附图仅仅示出了该发明的典型的实施例,并且因此不被认为限制本发明的范围,因为本发明可以承认其它同样有效的实施例。在附图内,相同的附图标记想要指代相同或相对应的部件,并且其中:

[0020] 图1呈现了根据本发明的实施例的示例性分布式计算机系统。

[0021] 图2呈现了根据本发明的实施例的数据处理系统的示例性框图。

[0022] 图3呈现了根据本发明的实施例的示例性元数据表。

[0023] 图4呈现了根据本发明的实施例的示例性模式定义模型。

[0024] 图5呈现了根据本发明的实施例的示例性模式定义模型表。

[0025] 图6呈现了根据本发明的实施例的示例性物理模式表。

[0026] 图7呈现了根据本发明的实施例的用于实现和填充运营数据存储的示例性控制流程。

[0027] 图8呈现了根据本发明的实施例的用于生成提取转换加载功能的示例性控制流

程。

[0028] 图9呈现了根据本发明的实施例的用于过滤队列的示例性控制流程。

[0029] 图10呈现了如在商业智能工具中所实现的过滤器的屏幕截图。

[0030] 图11呈现了根据本发明的实施例的用于利用模式来过滤队列的示例性控制流程。

[0031] 图12呈现了如在商业智能工具中所实现的过滤器的屏幕截图。

[0032] 图13呈现了根据本发明的实施例的用于生成报告的示例性控制流程。

[0033] 图14呈现了如在商业智能工具中实现的报告数据定义的屏幕截图。

[0034] 图15呈现了根据本发明的实施例的用于分析报告的控制流程。

## 具体实施方式

[0035] 在本发明的实施例的以下的描述中,参照了作为本发明的一部分的附图,并且在附图中,以示例的方式示出了其中可以实现本发明的示例性实施例。应当理解的是,可以利用其它实施例并且可以做出结构上的改变,而不偏离本发明的范围。

[0036] 以足够的细节描述了以下的实施例,以使本领域技术人员能够做出和使用本发明。应当理解的是,基于本公开,其它实施例将会显而易见,并且可以做出系统、过程、或者机械改变,而不偏离本发明的范围。

[0037] 在以下的说明书中,给出了大量具体的细节以提供对发明的充分理解;然而,显而易见的是,可以没有这些具体细节的情况下实践本发明。为了避免使本发明难以理解,没有详细地公开一些公知的电路、系统配置、和过程步骤。

[0038] 此外,为了对本发明的说明、描述、和理解的清楚和方便,公开并且描述多个实施例的地方有一些公共的特征,相似和相同的特征彼此通常用相同的附图标记来描述。实施例已经被编号为第一实施例、第二实施例等,这是为了描述方便而不是要具有任何其它意义或者提供对本发明的限制。

[0039] 出于说明的目的,本文中所使用的术语“元数据”被定义为关于数据的数据。本文中所使用的术语“系统”表示或指代根据在其中使用该术语的上下文的本发明的方法和装置。

[0040] 本发明的实施例提供了这样一种技术,其应用唯一的模式定义语言并且在构建、生成和填入数据库或数据存储中利用模式定义语言;自动地生成提取-转换-加载功能;启用参考模式定义语言从数据或数据存储中拉取数据的商业智能工具。如在本文中所使用的,商业智能工具一般是指被配置为报告、分析、和呈现数据的软件应用。数据可以存储在数据仓库、数据库、数据存储、数据集市、或其组合中。

[0041] 根据一个方面,模式定义语言可以以呈现由模式定义语言定义的特性、实体和模块的关系的模式定义模型来实现或由模式定义模型来表示。模式定义语言定义了运营数据存储的物理表的结构和框架。模式定义语言可以包括用于将模式定义模型映射到物理数据的一个或多个物理实体的关系和位置。因此,模式定义语言定义并且可以被用于访问包含物理数据的特定集合的物理数据的字段。

[0042] 有利地,本发明的实施例提供了这样的技术,其用于提供较高抽象层次的模式定义语言,并且提供用于与模式定义语言相接以使得商业智能工具能够以较低的成本、以及更短的时间帧来利用运行数据存储的平台。模式定义语言使得商业智能工具能够在较高的

抽象层次操作,因此当底层数据库演进时,工具不再需要被改变。当利用本发明的模式定义语言时,数据库演进是容易的并且是不被中断的

[0043] 有利地,通过提供涉及在之前的步骤中所选择的特性类型的选项,利用模式定义语言来访问和查询物理表的商业智能工具向用户提供直观的体验。模式定义语言还可以被用于向物理表提供简单的结构,实现对运营数据存储进行建模和设计的简单并且高效的解决方案。此外,利用模式定义语言允许提取、转换、和加载功能的自动化,以基于模式定义语言结构和被包含在模式定义模型内的元数据来快速地填充运营数据存储。

[0044] 在下文中,参照了本发明的实施例和具体的示例;然而,应当理解的是,本发明不限于所描述的具体的实施例或示例。相反,以下的特征和元素的任何组合,无论是否与不同的实施例有关,都被预期为实现和实施本发明。此外,尽管本发明的实施例与其它可能的解决方案和/或现有技术相比可能取得优点,但无论特定的优点是否由给定的实施例实现都不会限制本发明。因此,以下的方面、特征、实施例和优点仅仅是说明性的,并且不被认为是所附权利要求的元素或限制,除非在权利要求中明确记载。同样地,对“本发明”的引用将不被解释为对在本文中所公开的任何创造性的主题的概括,并且不应该被解释为所附权利要求的元素或限制,除非在权利要求中明确记载。

[0045] 当所附权利要求中的任何一项被解读为覆盖纯软件和/或固件实现时,至少一个示例中的元素的至少一个由此明确地被定义为包括存储了该软件和/或固件的有形的计算机可读介质。

[0046] 本发明的一个实施例被实现为利用计算机系统来使用的程序产品。程序产品的程序定义了实施例(包括在本文中所描述的方法)的功能,并且可以被包括在各种计算机可读存储介质上。计算机可读存储介质在本文中被定义为制品。示例性的计算机可读存储介质包括但不限于:(i)其上永久地存储有信息的不可写的存储介质(例如,计算机内的只读存储器设备,例如可以由CD-ROM驱动器读取的CD-ROM盘);(ii)其上存储有可改变的信息的可写存储介质(例如,盘式驱动器内的软盘或硬盘驱动器)。当承载着指导本发明的功能的计算机可读指令时,这样的计算机可读存储介质是本发明的实施例。其它介质包括可以通过其将来将信息传递到计算机(例如,通过计算机或电话网络,包括无线通信网络)的通信介质。后来的实施例具体地包括将信息传输到互联网和其它网络或者从互联网和其它网络传输信息。当承载着指导本发明的功能的计算机可读指令时,这样的通信介质是本发明的实施例。宽泛地,计算机可读存储介质和通信介质在本文中都被称为计算机可读介质。

[0047] 一般而言,被执行以实现本发明的实施例的例程可以是操作系统的一部分或者是具体的应用、部件、程序、模块、对象、或者指令的序列。本发明的计算机程序一般包括将由本机计算机转换成机器可读的格式并且因此转化成可执行指令的大量指令。同样,程序包括或者本地地驻留在程序中、或者在存储器中或存储设备上找到的变量和数据结构。此外,在下文中所描述的各种程序可以基于在本发明的具体的实施例中实现它们所针对的应用而被标识;然而,应当认识到,以下任何特定的程序术语都仅仅出于方便而使用,并且因此本发明不应当限于唯一地在由这样的术语标识和/或暗示的任何特定的应用中使用。

[0048] 现在参考图1,其中示出了根据本发明的实施例的示例性分布式计算机系统100。一般地,分布式计算机系统100被示出为分布式环境并且包括计算机系统102和多个网络设备104。计算机系统102可以表示任何类型的计算机、计算机系统或其它可编程的电子设备,



包括客户端计算机、服务器计算机、便携式计算机、嵌入式计算机、基于PC的服务器、微型计算机、中型计算机、大型计算机、适用于支持本发明的方法、装置、和制品的其它计算机。

[0049] 示例性地,计算机系统102包括网络系统。然而,计算机系统102还可以包括独立设备。在任何情况下,应当理解的是,图1仅仅是计算机系统100的一个配置。本发明的实施例可以应用于任何相当的配置,无论计算机系统102是复杂的多用户装置、单用户工作站、还是本身不具有非易失性存储的网络设备。

[0050] 还可以在其中任务由通过通信网络链接的远程处理设备执行的分布式计算环境中实现本发明的实施例。在分布式计算环境中,程序模块可以位于本地和远程的存储器存储设备两者中。就这一点而言,计算机系统102和/或网络设备104中的一个或多个可以是几乎不执行处理的瘦客户端。

[0051] 计算机系统102可以包括多个操作者和外部系统,例如由可操作地连接到直接存取存储设备108的大容量存储接口106、由可操作地连接到显示器112的视频接口110、以及由可操作地连接到多个网络设备104的网络接口114所示出。显示器112可以是用于输出可查看的信息的任何视频输出设备。

[0052] 计算机系统102被示出为包括经由总线118从主存储器120中获取指令和数据的至少一个处理器116。处理器116可以是适用于支持本发明的方法的任何处理器。

[0053] 主存储器120是足够大来保存必要的程序和数据结构的任何存储器。主存储器120可以是包括随机存取存储器、非易失性或备份存储器(例如,可编程或闪速存储器、只读存储器等)在内的存储器设备中的一个或组合。此外,存储器120可以被认为包括物理地位于计算机系统102的其它位置的存储器(例如,被用作虚拟存储器或被存储在大容量存储设备(例如,直接访问存储设备108)上的任何存储容量)或者经由总线118耦合到计算机系统102的另一个计算机上的存储器。

[0054] 存储器120被示出为配置有操作系统122。操作系统122是用于管理计算机系统102的操作的软件。

[0055] 存储器120还包括访问层122、模式定义语言126、过滤器128、报告数据定义130、一个或多个应用132、以及多个商业智能工具134。应用132、商业智能工具134以及访问层124是包括在各种时间驻留在计算机系统102的各种存储器和存储设备中的多个指令的软件产品。当被读取并且由计算机系统102中的一个或多个处理器116执行时,应用132、访问层124、以及模式定义语言126单独地或相结合地使得计算机系统102执行用于执行本发明的各种方面必要的步骤。

[0056] 访问层124(并且更加一般地,任何请求实体,包括操作系统122)被配置为发出针对数据库136的查询。示例性地,数据库136被示出为存储108中的数据库管理系统(DBMS)138的一部分。尽管出于简化的目的仅示出了一个数据库,但是DBMS 138可以包括多个数据库。

[0057] 此外,数据库可以彼此相关地分布。而且,一个或多个数据库可以被分布到网络设备104中的一个或多个上。示例性地,网络系统140被示出为具有包括数据库144的DBMS 142。尽管出于简化的目的仅示出了单个的数据库144与DBMS 142,但是DBMS 142可以包括多个数据库。此外,DBMS 142的数据库可以彼此相关地分布。所有这样的不同实现方式都被广泛地预期。存储108或网络设备104还可以包括由应用132使用以构建过滤器、提供客户反

馈、或者构建针对商业智能工具134的报告数据定义的元数据表、物理表、模式定义语言126、或者模式定义模型。

[0058] 数据库136和144代表任何的数据的集合,无论数据的特定的物理表示。数据的物理表示定义了数据的组织模式。

[0059] 在一个实施例中,数据库136包括运营数据库并且数据库144包括运营数据存储。运营数据库包括被包含在数据存储中的物理数据的至少一部分。根据一个方面,数据存储包含从运营数据库中的物理数据得到的可查询数据。因此,数据存储中的可查询的数据包括运营数据库中的物理数据的子集。除了来自运营数据库的数据的子集之外,数据存储可以包括其它数据。

[0060] 在一个实施例中,可以响应于输入(例如,用户输入)来生成查询。可以使用由模式定义语言126所定义的逻辑字段来组成查询。可以使用过滤器128来执行对数据库144的查询,该过滤器128可以引用模式定义语言126来隔离或定义由数据库144内的物理数据表示的图4的实体404的类别。在下文中参照图9-12更加详细地描述过滤器128的操作。

[0061] 还可以使用报告数据定义130来执行对数据库144的查询,该报告数据定义130可以引用模式定义语言和图4的实体404、模块406、以及特性观察408的结构报告,并且从数据库144内的物理表中拉取数据。在下文中参考图13-15更加详细地描述报告数据定义130的操作。

[0062] 现在参照图2,其中示出了根据本发明的实施例的数据处理系统200示例性框图。数据处理系统200包括数据输入块202,用于从一个或多个外部源204中提取数据203,并且将所提取的数据203加载到与其耦合的运营数据存储206中。例如,外部源204可以是图1的存储设备108的数据库136。运营数据存储206可以被包括在图1的网络设备104的数据库144内。

[0063] 管理引擎208被配置为对运营数据存储206中的数据203的至少一部分进行映射和定位,基于针对任何后续的分配处理、映射、过滤、或报告的模式定义语言126来执行对数据203的位置和关系的运算。

[0064] 数据处理系统200还包括耦合到运营数据存储206和管理引擎208的信息传送块210,以用于将数据203传送给用户。传送功能可以由商业智能工具134实现。商业智能工具134可以是过滤工具,例如分别在下文中的图9和图11中所描述的TransMed系统队列管理器或者TransMed系统临床模式匹配器。商业智能工具134还可以是报告工具,例如,分别在下文中的图13和图15中所描述的TransMed系统队列报告器或者TransMed系统队列分析器。

[0065] 此外,数据处理系统200的信息传送块210包括用于从用户处接收运行时间参数的用户接口212以及用于响应于来自用户接口212的输入而启动管理引擎208的任务控制器214。用户接口212可以包括用于与用户进行双向通信的硬件和软件。任务控制器214也与管理引擎208进行通信,这允许用户经由用户接口212提交用于对数据203进行分组、过滤、以及返回的参数和指令。

[0066] 数据输入块202使用提取-转换-加载(ETL)工具216以从外部源204中的一个或多个中提取数据203,并且然后将所提取的数据203转换成至少一种优选的数据格式。转换后的数据203然后可以被加载到运营数据存储206中的一个或多个物理表218中以用于数据存储。

[0067] 重要的是,管理引擎208可以提供模式定义语言126作为商业智能工具134和物理表218的框架。管理引擎208可以自动地生成ETL工具216以利用模式定义语言126将来自外部源204的数据203填入物理表218。

[0068] 在下文中关于图7和图8描述了使ETL工具216自动化和填充物理表218的操作。管理引擎208可以进一步提供模式定义语言126作为商业智能工具134可以基于其来操作以构建报告和过滤搜索的框架。

[0069] 模式定义语言126包括实体类型220、特性222、以及模块类型224。特性222利用链接226被链接到实体类型220。可以基于特性222是否在一起被观察到而在模块类型224中将特性222分组在一起。

[0070] 实体类型220、特性222、以及模块类型224可以是提供用于形成在下文中关于图4所描述的模式定义模型400的结构的模式定义语言126的元数据构造,模式定义模型400将提供运营数据存储206中的物理表218的结构和关系。模式定义语言126还提供针对数据203将如何填充物理表218的框架。

[0071] 出于示例目的和便于理解,下文将参照健康保健的领域来描述实体类型220、特性222、以及模块类型224。本领域技术人员将认识到这仅仅是出于说明性的目的并且不是想要限制本发明。

[0072] 可以预期,实体类型220可以定义物理对象的类型。实体类型220是图4的实际实体404的抽象和元数据302(图3的),其是数据构造。实际的实体404是具有图4的特性观察408以及生命周期的实体类型220的具体实例。本文中所使用的生命周期被定义为其中物理对象以这样的一种状态存在的时间跨度:合理的观察者能够将该物理对象认为是实体类型220中的一个。特性观察408在本文中被定义为与实体404有关的可观察的事实。

[0073] 相关于健康保健领域,实体类型220可以被示例化为病人、医师、或者设施。实体404可以被实例化为具体的医院、病人、医师、或者付款人。就是说,实体404可以是Jones医生或者Suburban医院。

[0074] 可以预期,模块类型224可以定义产生对特性观察408的具体的记录或时刻。模块类型224是图4的实际模块406的抽象和元数据302,其是数据构造。实际的模块406是模块类型224的具体实例,模块406可以列出通常被一起观察的针对实体404而观察到的特性观察408。本文中所使用的模块406被定义为经常被一起观察的针对实体404的相关的特性观察408的集合。模块类型224可以包括针对特性222的观察的开始和结束日期。模块类型224列出了与模块类型224相关联的特性222,并且标识了模块类型224的开始和结束日期。还可以预期,模块类型224可以仅包括针对没有持续时间的事件的单个日期或时间戳、相同的开始和结束时间戳。

[0075] 模块类型224可以被示例化为交会(入院日期、出院日期)、诊断(诊断日期)、治疗(治疗开始和结束日期)、以及实验室预定(实验室预定日期)。模块406本身可以被示例化为具体的交会、诊断、治疗、实验室预定。也就是说,模块406可以是Mike Johnson在2/5/2012住院到Suburban医院紧急病房,Mike Johnson在2/5/2012被诊断有肺炎,Jones医生在2/5/2012给Mike Johnson用了青霉素来治疗Mike Johnson的肺炎,或者Jones医生在2/5/2012为Mike Johnson预订了没有差分测试的完整的血细胞计数。

[0076] 可以预期,特性222可以定义实体404的或者针对实体404的单个的可观察属性或

特性观察408。特性222具有名称并且指定了图4的特定数据类型410,例如,整数、实数、日期时间、文本、选择、模块指针、或者实体指针。特性222是观察到的特性观察408的抽象和元数据302,其是数据构造。特性观察408是与实体404有关的事实的观察的具体实例。特性观察408匹配特性222数据类型410。

[0077] 特性222可以针对诊断模块类型而被示例化为具有时间的数据类型410的诊断日期,具有选择的数据类型410的ICD9选择、具有文本的数据类型410的医师笔记、或者具有选择的数据类型410的严重性选择。特性222还可以针对交会模块类型而被示例化为具有日期时间的数据类型410的入院日期、具有日期时间的数据类型410的出院日期、具有选择的日期类型410的主要支付者、或者具有选择的数据类型410的出院安置。

[0078] 特性观察408可以针对诊断模块被示例化为诊断日期:“2/5/2012的下午1:30”、ICD9:“480.1”、或者医师的笔记:“病人呼吸困难”。特性观察408可以针对交会模块而被示例化为入院日期:“2/5/2012的上午11:30”、出院日期:“2/7/2012的下午1:30”、主要支付者:“Aetna”、或者出院安置:“家”。

[0079] 特性222可以包括长度和非长度特性观察408。本文中所使用的长度特性观察408被定义以意指与被视为在实体404的生命跨度内发生的事件相对应的特性观察408。通常而言,长度特性观察408具有被观察拥有它们的短于实体404生命周期的持续时间。如在本文中所使用的,非长度特性观察408被定义以意指被视为持续了实体404的生命跨度的特性观察408,例如姓名、血型、或者病人ID。非长度特性可以改变但是不一定与具体的事件有关。实体类型220可以被分配模块类型224中具体的一个以捕获实体404的非长度特性。另一方面,长度特性将被分配实体404的发生的日期,并且将与标准的模块类型224分组。

[0080] 已经发现利用被实现为元数据302并且由图3的元数据表300表示的模式定义语言126语义上表示链接到由模块406分组的特性观察的实体404并且由此提供了物理表218的简单和有效的框架。与元数据表300的模式定义语言126相耦合的该框架使得能够通过引用元数据表300来快速、简单、并且有效地使用物理数据表218。物理数据表218可以被过滤并且用于通过引用元数据表300中的模式定义语言126来生成报告。

[0081] 被捕获为元数据表300中的元数据302的模式定义语言126允许物理数据表218被过滤并且用于以直观的方式生成报告,而同时降低实现成本和时间线。实现时间线可以通过利用模式定义语言126而显著地被降低,因为在典型的数据仓库中,数据结构需要被归一化并且被置于维度表中;然而,通过利用模式定义语言126,数据203可以简单地与元数据构造和数据构造相关联。一旦关联被提供,ETL工具216就可以被自动地生成并且物理表218被填充。运营数据存储206的物理表218一被填充,商业智能工具134就可以利用模式定义语言126来过滤、报告、和查询运营数据存储206中的数据,而不是要求归一化的多个额外的步骤、以及用每个步骤之间的提取、转换、加载来填入维度表。

[0082] 还已经发现,模式定义语言126向特性222提供具有预先定义的数据类型410允许ETL工具216的自动化生成,因此显著地降低了实现成本和时间线。

[0083] 为了使商业智能工具134能够有效地操作并且向用户提供直观的体验,信息传送块210被耦合到管理引擎208以引用模式定义语言126。管理引擎208还可以包括过滤器128和报告数据定义130。过滤器128可以是用于基于实体404的特性观察408来生成实体404的类的一维过滤器。在下文中关于图9-图12更加详细地描述了过滤器。报告数据定义130可以

是实体类型220、模块类型224、以及特性222的集合,它们对于报告是要求或期望的,并且将用于从运营数据存储206中取回数据203。在下文中关于图13-图14更加详细地描述了报告数据定义130。

[0084] 现在参照图3,其中示出了根据本发明的实施例的示例性元数据表300。元数据表300表示图1的模式定义语言126。模式定义语言126可以被用于描述图2的物理数据表218的数据203的元数据302。元数据表300可以包括其间的多个关系链接304。元数据表300可以包括链接到特性表308的实体类型表306。实体类型表306和特性表308可以被链接到模块类型表310。

[0085] 实体类型表306、特性表308、以及模块类型表310可以分别与图2的实体类型220、特性222、以及模块类型224相关。元数据表300可以分别包括作为实体类型表306、模块类型表310、以及特性表308内的行的图4的实体404、模块406以及特性观察408。

[0086] 元数据表300包含通过语义表示来对物理数据表218进行映射的元数据302。作为示例,特性表308可以包含与特性观察408相对应的行。作为示例,ICD9选择是可以被包含为特定表308中的行的特性观察408。特性表308的ICD9选择可以包括到模块406和或实体404的指针或引用,模块406和实体404可以与特性观察408相关联。例如,ICD9特性观察408可以指向诊断模块406并且指向病人实体404。

[0087] 由于元数据表300包括对物理数据表218的结构语义表示或映射,因此元数据表300可以用于定位被包含在物理数据表218内的数据203。元数据表300还可以用于确定物理数据表218的数据203之间的关系链接304。

[0088] 模块类型表306、实体类型表306、以及特性表308的关系示出了在上文中关于图2更加详细地描述的实体类型220、特性222、模块类型224的关系。

[0089] 元数据表300之间的关系链接304是针对数据203将如何最终被构建在图2的运营数据存储206的物理表218内的映射。关系链接304是指向或引用模块406、实体404、或其他数据或元源数据构造的特性观察408的图示。

[0090] 模式定义语言126的关系链接304提供将最终保持数据203的物理表218的关系结构。模式定义语言126可以提供针对物理表218的位置的映射,以及物理表218之间的关系。

[0091] 作为示例,如果用户选择了特性观察408来过滤实体404或者模块406,图1的商业智能工具134可以引用模式定义语言126来定位特性观察408并且确定特性观察408如何链接到实体404和模块406。一旦链接到特性观察408的模块406和实体404由元数据表300确定,则属于相同的模块406或者实体404的其它特性观察408作为客户选择的特性观察408。

[0092] 通过将特性观察408的关系提供给实体404和模块406,图2的管理引擎208可以向用户提供选项以用于选择与所选择的第一特性观察408有关的其它特性观察408,或者可以向用户提供选项以用于选择其它模块406。管理引擎208可以利用模式定义语言126将来自相同模块406的其它特性观察408返回给用户,作为用户的选择选项。以这种方式,可以利用模式定义语言126来仅基于物理表218内所要求和所包含在模式定义语言126内的数据203的结构而向用户提供相关的过滤器或报告选项。

[0093] 元数据表300之间的关系链接304示出了多对一的链接312。实体类型表306和模块类型表310还包括允许实体404之间的分层关系的递归链接314。

[0094] 其它表可以组成元数据表300,包括数据类型表316、选择类型表318、选择表320、

以及授权表322。可以预期,元数据表300可以包括更多表而不偏离本发明。作为示例,元数据表300可以包括模块类型成员表或者度量单位表。还应当理解的是,表可以被组合或删除,而不偏离本发明的范围。

[0095] 现在参考图4,其中示出了根据本发明的实施例的示例性模式定义模型400。模式定义模型400可以是图1的模式定义语言126在数据上下文中的应用。仅仅出于示例性的目的,继续健康保健的示例,实体类型220、模块类型224、以及特性222已经被映射到实体404、模块406、以及特性观察408的元数据302。

[0096] 元数据302可以与实体404、模块406、特性观察408、以及实体类型220、模块类型224、和特性222相关联。如将通过以下的示例性示例所示出的,被包含在模式定义模型400内的元数据302可以定义特性观察408如何链接到实体404以及特性观察408如何在模块406中被分组在一起。

[0097] 作为示例性示例,实体404可以是病人、样本、实验、或者遗传性变型。实体404被示出为与非长度特性观察408分组在一起。作为示例性的示例,与被标识为病人的实体404分组在一起的特性观察408可以包括ID、出生日期、死亡日期、性别、种族、主要支付者、以及当前的生命状态。

[0098] 与模块406分组在一起的特性观察408被示出为长度特性观察408。作为说明性示例,具有针对辐射的元数据302的与模块406分组在一起的特性可以包括加强剂量、加强治疗方式、治疗的数量、辐射部位、辐射体积、以及区域剂量。模块406还可以包括将是其它模块406的外键(foreign key)的指针,例如,在与辐射相对应的模块406中针对治疗模块406的指针。

[0099] 在模式定义模型400中,模块406和实体404没有被描绘为图2的实际数据203,而是相反地,被示出为具有元数据302。特性观察408还被示出为具有数据类型410的形式的元数据302。数据类型410可以指定将最终填充图2的物理表218的数据203的类型。

[0100] 针对特性观察408的数据类型410可以指定整数数据、实数数据、文本数据、长文本数据、日期、日期和时间、选择数据、模块指针数据、以及实体指针数据。利用数据类型410,元数据302允许图2的ETL工具216由图2的管理引擎208自动地生成。将在下文中关于图8更加详细地描述生成ETL工具216。

[0101] 由于图3的实体类型表306和模块类型表310的递归关系,实体404和模块406可以分别包括子实体和子模块。子实体可以是源自特定实体404的实体404。子模块可以是源自特定模块406的模块406。

[0102] 现在参照图5,其中示出了根据本发明的实施例的示例性模式定义模型表500。为了便于描述,以简化的格式示出了模式定义模型表500。

[0103] 模式定义模型表500被示出为具有多个列和行。每一行与特性观察408相对应。表的单元填充有元数据302而不是将要填充图2的运营数据存储206的物理表218的实际数据203。

[0104] 以与图4相似的方式,特性观察408可以链接到实体类型220和模块类型224。实体类型220可以由第一列表示,而模块类型224可以由第二列表示。以类似的方式,特性观察408的行将经过指示特性观察408的其它特征的列。特性观察408可以与显示特性502、序列504、特性名称506、数据类型410、选择类型508、垂直指示符510、目标实体类型512、目标模

块类型514、度量的单元516、多个选择指示符518、以及保护指示符520相关联。

[0105] 第一列的实体类型220可以是与特性观察408相关联的实体404。例如，如果利用之前的健康保健领域的示例，则实体类型220可以包括病人、样本、实验、遗传变型、医院、或者医师。第二列的模块类型224可以是与特性观察408相关联的模块406。模块类型224可以包括病人信息（其是用于包含实体404的非长度特性观察408的特殊模块）、癌症诊断、化学疗法、药物、辐射、手术、治疗、实验室和生命、同意、样本信息、实验信息、遗传变型观察以及其他。

[0106] 第三列的显示特性502可以指示特性观察408是否应该被显示为模块406的默认ID特性。显示特性502可以将特性观察408标识为或者经标识的特性或者去标识的特性。与针对病人实体的病人信息模块相关联的经标识的特性的示例可以是病人ID。与针对病人实体的诊断模块相关联的经标识的特性的示例可以是ICD9值。与针对病人实体的病人信息模块相关联的去标识的特性的示例可以是生日，因为当利用实体404工作中，该特性作为默认值不太有用。

[0107] 第四列的序列504可以指示特性观察408可以在模块406内被显示的顺序。第五列的特性名称506可以通过名称来标识特性。作为药物模块的特性名称506的示例，特性观察408可以被标识为药物剂量、药物频率、药物名称、给药途径、药物治疗、或者药物诊断。模块类型224中的每个都可以包括将与被映射到物理表218中的特性观察408的数据203相对应的唯一的特性名称506。

[0108] 第六列的数据类型410可以指示特性观察408将被存储的数据的类型。特性观察408可以包括整数数据、实数数据、文本数据、长文本数据、日期数据、日期时间数据、选择数据、模块指针、或者实体指针。

[0109] 如果数据类型410与选择数据相对应，则将使用第七列的选择类型508。选择类型508可以是针对与选择数据相关联的特性观察408的所有有效的选择的列表。作为示例，选择类型508可以包括是/否，或者男/女。选择类型508还可以包括较长的列表，例如针对治疗或化学治疗模块的化学治疗的潜在的类型的列表。这种类型的选择类型508可以是包括二氯甲基二乙胺、环磷酰胺、苯丁酸氮芥、美法仑、异环磷酰胺、三胺硫磷、环六亚甲基四胺、六甲嘧胺、甲基苄肼、氮烯唑胺、替莫唑胺、卡莫司汀、环己亚硝脲、链脲菌素、卡铂、顺铂、或者奥沙利铂在内的列表。

[0110] 列8的垂直指示符510可以指示特性观察408将如何被存储在运营数据存储206的物理表218中。如果垂直指示符510是指示特性观察408不是被垂直存储的“N”，则特性观察408将被存储在表示与该特性相关联的模块406的表的列中。如果垂直指示符510是指示特性观察408被垂直存储的“Y”，则特性观察408将被存储在单独的表中，特性观察408中的每个都具有单独的行。

[0111] 如果数据类型410是实体指针，则将使用第九列的目标实体类型512。目标实体类型512可以用于指示由与特性观察408相关联的模块406所引用的实体类型220。作为示例，如果癌症诊断是利用活组织切片检查得出的，则与实体指针相对应并且与癌症诊断模块相关联的特性观察408中的一个可以引用“样本”的实体类型220。

[0112] 如果数据类型410是模块指针，则将使用第十列的目标模块类型514。目标模块类型514可以用于指示由与特性观察408相关联的模块406所引用的模块类型224。作为示例，

与模块指针相对应并且与辐射模块相关联的特性观察408中的一个可以引用“治疗”的模块类型224。

[0113] 第十一列的度量单位516可以用于指示特性观察408(当它们是数值时)的度量的单位。作为示例,度量516的单元可以是磅或者千克。

[0114] 当数据类型410是选择时,可以利用第十二列的多个选择指示符518来指示用户是否可以多个选择还是仅单个选择。作为示例,如果多个选择指示符518是“N”,则仅仅一个选择是有效的,当多个选择指示符518是“Y”时,则多个选择是有效的。

[0115] 第十三列的保护指示符520可以指示特性是否包含敏感信息。例如,如果保护指示符520是“Y”,则特性可以包含病人标识信息并且应该被不同地治疗。

[0116] 模式定义模型表500将被用于生成运营数据存储206的物理表218。模块406中的每个都将是运营数据存储206内的单独的表。特性观察408中的每个都将是与它们相关联的模块406的表内的列。数据203将占据物理表218内的单元。

[0117] 现在参照图6,其中示出了根据本发明的实施例的示例性物理表218。物理表218可以是图2的运营数据存储206的物理表218。

[0118] 如在下文中所描述的,物理表218可以通过引用被包括在图1的模式定义语言126内的图3的元数据302而被生成,单独地或与在图3的元数据表300、图4的模式定义模型400、图5的模式定义模型表500或其组合中表示的模式定义语言126相组合。基于图4的特性观察408中的至少一个,物理表218可以包括针对其间具有链接的图4的模块406和实体404的表。

[0119] 继续之前所讨论的健康保健的示例并且利用模式定义语言126,物理表218被示出为与在模块类型224列或者模式定义模型表500的第二列中所定义的模块406相关。在模式定义模型表500中所定义的模块406的每个都具有在物理表218的完整列表中创建的唯一的物理表602。

[0120] 物理表218可以包括表题604以及接着的列名称606的列表。列名称606是被包括在模式定义模型表500的行中的特性观察408。此外,当模式定义模型表500的特性观察408是模块指针或者实体指针时,这些特性观察408在物理表218中被描绘为链接608。一般地,链接608可以包括一对一关系或者多对一关系。

[0121] 例如,实体表610可以与ODS\_PatientInfo表612具有一对一关系。这意味着对于每个实体,实体表610将包括指向ODS\_PatientInfo表612的一个外键。同样,ODS\_PatientInfo表612将包括指向实体表610的一个外键。

[0122] 物理表218还可以包括多对一链接。例如,ODS\_PatientInfo表612可以包括来自多个ODS\_Medication表614的多个外键,但是每个ODS\_Medication表614将仅包括针对ODS\_PatientInfo表612的单个外键。

[0123] 通过模式定义语言126的使用而呈现的另一个解决方案由垂直指示符510标记提供。通常当垂直指示符510标记是否定的或“N”时,模式定义模型表500中的每个特性观察408将是物理数据表218中的列。这是由ODS\_LabsVital表616所示出的。然而,当特性观察408变得非常多时,垂直指示符510可以被设置成肯定的或者“Y”。当垂直指示符510是肯定的时,其指示特性观察408不应该被构建为物理数据表218内的列,而应该占据被示出为ODS\_LabsVital\_Vertical表618的单独的表内的行。作为示例,诸如血白细胞计数和血红细胞计数之类的特性观察408将与其它多个特性一起占据ODS\_LabsVital\_Vertical表618。



[0124] 关于图1到图6所描述的数据要求以及针对图7、图8、图9、图11、图13、以及图15的流程图的数据要求可以被固定或保存在非暂时性计算机可读介质内。这些数据要求在不使用或实现硬件的情况下不能够访问和修改,并且数据值中的改变可以表示以位的形式硬件记录的物理的、非暂时性的转换。

[0125] 已经描述了针对实现和使用图2的运营数据存储206的数据要求,我们现在描述针对实现和使用运营数据存储206的过程。在关于图7、图8、图9、图11、图13以及图15所描述的流程图中,实现和使用运营数据存储206的过程步骤可以以硬件、软件、固件、或者其任何组合来实现。此外,步骤边界通常变化,而功能是一起实现的,以及在不同的实施例中单独实现。

[0126] 现在参照图7,其中示出了根据本发明的实施例的用于实现和填充运营数据存储的示例性控制流程700。作为示例,图2的运营数据存储206可以以相似的方式被实现和填充。

[0127] 控制流程700包括提供模式定义语言步骤702。可以调用提供模式定义语言步骤702以提供模式定义语言126。通过提供模式定义语言126,设计者可以利用数据和元数据构造。

[0128] 耦合到提供模式定义语言步骤702的是模型模式定义步骤704。在模型模式定义步骤704期间,设计者将利用模式定义语言126来定义使用被包含在现有的文件、现有的运营数据存储、或者现有的数据库中的数据元素的模式定义模型400。模式定义语言126实际上是由设计者通过利用在上文中关于图2和图3更加详细地描述的元数据和数据构造来实现的。

[0129] 模型模式定义步骤704可以被耦合到捕获模式定义模型步骤706。在捕获模式定义模型步骤706期间,模式定义模型400可以在模式定义模型表500内被捕获,所述模式定义模型表500提供了针对特性观察408中的每个的行,并且将其分配给或者相关到图4的实体404和模块406。可以预期,可以略过捕获模式定义模型步骤706,并且可以使用模式定义语言126替换模式定义模型表500,以用于引用和查询元数据302。模式定义语言126可以用作图3的元数据表300、模式定义模型400、或其组合。

[0130] 捕获模式定义模型步骤706可以耦合到生成运营数据存储(ODS)步骤708。生成ODS步骤708包括从被包含在模式定义模型表500内的模式定义语言126中生成运营数据存储206的物理表218,并且在其间创建图6的链接608。

[0131] 具体而言,在生成ODS步骤708期间,首先生成针对模块406中的每个的物理表218中的一个。如果模式定义模型表500的垂直指示符510是“N”,则针对与特性观察408相关联的模块406而在图6的物理表602中创建列。当针对特性观察408创建列时,图5的数据类型410指示器确定该列将包括的数据的类型。

[0132] 在生成ODS步骤708期间,将针对数据类型410选择的特性观察408来生成外键,以用于引用选择表。类似地,在生成ODS步骤708期间,将针对包括模式定义模型表500的目标实体类型512或目标模块类型514中的指示符的特性观察408来生成外键。如果目标实体类型512包括指示符,则将生成引用图6的实体表610的外键。另一方面,如果目标模块类型514包括指示符,则将创建引用模块表中的一个的外键。

[0133] 如果模式定义模型表500的垂直指示符510是“Y”,则将不会针对特性观察408来生

成列。相反,该类型的特性观察408将被载入与模块406相关联的垂直表中,每行一个特性观察408。最终,与长度时间戳相对应的特性观察408将与模块406的物理表218相关联。

[0134] 生成ODS步骤708被耦合到生成ETL步骤710。一旦生成ODS步骤708的物理表218与在捕获模式定义模型步骤706中完成的模式定义模型表500一起被完成,则ETL工具216就可以基于模式定义模型表500和模式定义语言126而被自动地生成。将在下文中关于图8更详细地描述ETL工具216的生成。

[0135] 生成ETL步骤710被耦合到填充ODS步骤712。一旦根据模式定义模型表500和模式定义语言126自动地生成了ETL工具216,并且一旦物理表218被创建,则在填充ODS步骤712中利用ETL工具216以从图2的外部源204中提取数据203,将图2的数据203转换为与数据类型410、选择类型508、或者模式定义模型表500的其它列兼容的格式。一旦数据203与模式定义模型表500和模式定义语言126的要求相兼容,则数据203就被加载到运营数据存储206的物理表218中。

[0136] 应当理解的是,当在图8、图9、图11、图13、和图15的流程图中引用或使用包括模式定义语言126、模式定义模型400、或者模式定义模型表500在内的元素时,可以预期,其中的一个或者某种组合可以代替或者与命名的元素一起被引用或使用。为了便于描述,模式定义模型表500一般被用作流程图的参照示例。

[0137] 现在参照图8,其中示出了根据本发明的实施例的用于生成提取转换加载功能的示例性控制流程800。控制流程800示出了通过引用图1的模式定义语言126来自动地生成图2的ETL工具216的示例性方法。模式定义语言126可以关于包含在图3的元数据表300内的元数据302、图4的模式定义模型400、图5的模式定义模型表500、或者其组合而被引用。如将在下文中被描述的,图2的物理表218可以通过根据元数据302自动地生成的ETL工具216而用图2的数据203来填充。

[0138] 一般地,来自图2的外部源204的数据203在最终到达图2的运营数据存储206之前流经非统一的阶段表802以及统一的阶段表804。图2的管理引擎208可以根据模式定义语言126自动地生成非统一的阶段表802和统一的阶段表804。

[0139] 作为不洁的源数据的数据203,在移动到统一的阶段表804中之前必须首先在非统一的阶段表802中被解析和净化。可以提供适配器和API以解析数据203并且踢出未通过语法验证的记录。洁净的数据203可以直接流到统一的阶段表804中。可以在统一的阶段表804中执行表和数据库级的验证。

[0140] 将数据从非统一的阶段表802和统一的阶段表804移动至运营数据存储206中的ETL工具216例程可以是自动地生成的。下文中关于步骤806-824更加详细地描述了该过程。

[0141] 首先,取回步骤806可以取回作为非统一源数据的数据203。这可以是以平面文件的形式;然而可以预期,数据可以是任何可用的形式。作为平面文件,数据203可以是文本字符串类型的数据203。源数据可以包括与图4的实体404和模块406相对应的数据203。作为示例,非统一阶段表802中的源数据可以包括病人信息平面文件,其可以与保存实体404和图2的实体类型220的非长度数据的特殊的模块类型相对应。

[0142] 非统一阶段表802还可以包括与模块406和图2的模块类型224相对应的模块表。作为示例性的示例,非统一的阶段表802可以包括与诊断模块类型224和癌症诊断模块406相对应的癌症诊断平面文件。

[0143] 与模块406相对应的非统一阶段表802中的数据203可以包括指向与其关联的实体404的外键。另一方面,与实体404相对应的非统一阶段表802中的数据203可以包括主键(例如,病人ID),作为针对实体404以及特别地创建的模块406的主键,所述模块406仅仅是为了保持针对实体404的非长度数据。

[0144] 耦合到取回步骤806的是处理步骤808。处理步骤808可以收集模式定义语言126的元数据302和或将其组织成选择类型统一的阶段表810和选择统一的阶段表812。根据模式定义语言126可用的每个观察到的选择都可以被置于选择类型统一的阶段表810中,并且可以被分配有主键、以及选择类型ID,例如,性别、种族、ICD9等。

[0145] 选择一致的阶段表812可以包括选择类型一致的阶段表810的外键以及针对每个可用的选择的主键。选择可以包括男性、女性、高加索人、西班牙裔、亚裔、173.0恶性的…、174.4佩吉特氏…、或者174.6肿瘤。修改的最后日期也可以被包括在选择类型一致的阶段表810和选择一致的阶段表812中。当执行增量加载时,可用的选择可以来自运营数据存储206。

[0146] 耦合到过程步骤808的是一致的加载步骤814。一致的加载步骤814可以包括一致的阶段表804。包含在针对图4的模块406和特性观察408模式定义语言126的元数据302(其被存储在元数据表300、模式定义模型表500、或者模式定义模型400内)内的定义可以用于确定数据应该如何被加载到统一的阶段表804中。例如,元数据302可以要求图4的数据类型410是日期。在这种情况下,当被加载到统一的阶段表804中时,从非统一的阶段表802中所提取的数据203应该是日期格式的。

[0147] 包含在非统一阶段表802中的实体404可以被加载到统一的阶段表804中。统一的阶段表804的每一行都可以包括实体404的单个实例。实体404中的每个都被分配了统一的阶段表主键。作为实体的子实体的实体404,将被加载并且被给予到父实体的指针。例如,已经从病人身上取样。样本将被认为是子实体并且具有回到病人的指针。指针可以是病人的统一的阶段表主键。实体404本身而不是非长度数据203,被首先加载到统一的阶段表804中。

[0148] 一旦实体404被加载到统一的阶段表804中,与模块406相关联的数据203可以被加载到统一阶段表804中。在统一的加载步骤814中,非统一的阶段表802被打开并被解析。

[0149] 当非统一阶段表802的每一行被处理时,统一的加载步骤814将数据203的值转换成目标数据类型410,确保了非空的特性观察408具有特性值,并且验证了对特性观察408的有效约束。当针对特性408的约束被验证时,针对日期和数值的特性观察408的取值范围的约束被验证,针对文本特性观察408的有效的特征和有效的长度被验证,并且对实体404的有效引用被验证。验证对其它模块406的引用的有效性是在第二遍(a second pass)期间执行的。

[0150] 在统一的加载步骤814期间的任何错误都使得与模块406中的一个相关联的记录被置于错误表816中。成功被处理的每行将被分配唯一的统一的阶段表主键。统一的阶段表804可以包括指示主键和外键的第一行、指示列的数据类型410的第二行、以及可以指示统一的阶段表804的列名称的第三行。统一的阶段表804对运营数据存储206模式进行镜像,除了引用其它模块406的特性观察408之外。

[0151] 继续上文中所描述的平面文件示例,平面文件文本值可以被转换成合适的数类

型410。并且利用统一阶段表主键来代替对实体404的引用。

[0152] 当记录未能通过统一的加载步骤814的有效性检查时,记录可以被置于错误表816中。例如,如果包含针对实体404的非长度数据203的记录包括不合适的出生日期,则记录可以被置于错误表816中。此外,当数据与引用错误表816中的实体404的模块相对应时,模块406记录也被置于错误表816内。

[0153] 一旦模块406在统一的加载步骤814中被加载到统一的阶段表804中,则可以在更新步骤818中更新模块406之间的引用。耦合到统一的加载步骤814的更新步骤818,通过只包括在所有的模块406被载入加载到统一的阶段表804之后的引用,来确保对模块406的合适的依赖。

[0154] 更新步骤818可以在阶段表804内创建针对引用另一个模块406的特性观察408中的每个的两个额外的列。两个额外的列可以包括所引用的模块名称以及所引用的模块外键。作为示例,治疗是指诊断,治疗特性观察408将包括处理诊断和处理诊断外键的额外的列。

[0155] 包括所引用的模块外键的列可以包括所引用的模块406的统一的阶段表主键。在该点,引用无效的、丢失的、或者错误表816中的模块406的特性观察408中的任何一个都被删除并且被置于错误表816内。更新步骤818可以包括适配器,以使得安装可以在过程中在该点插入其本身的、定制的验证。

[0156] 耦合到更新步骤818的是运营数据存储加载步骤820。在更新步骤818之后,统一的阶段表804包括经验证的数据。运营数据存储加载步骤820针对统一的阶段表804中的每个来生成SQL服务合并声明。运营数据存储加载步骤820可以针对选择类型统一的阶段表810、接着是选择统一的阶段表812、接着是包括实体404的统一的阶段表804、接着是包括实体404非长度数据203的统一的阶段表804、并且最后是针对其它模块406的统一的阶段表804而生成SQL服务合并声明。

[0157] 由于统一的阶段表804内的数据203全都被验证,因此只有环境问题可以导致合并失败。环境问题可以包括诸如没有足够的磁盘空间的问题。如果合并失败,则运营数据存储加载步骤820记录错误并且停止。

[0158] 耦合到运营数据存储加载步骤820的是删除步骤822。删除步骤822在到运营数据存储206中的增量加载期间应用。在增量加载期间,删除标记被加载到统一的阶段表804和ODS中。删除步骤822验证要删除的记录在运营数据存储206中是否存在。如果要删除的记录在运营数据存储206中不存在,则将记录路由到错误表816。一旦加载完成,则删除步骤822就利用级联来删除具有删除标记的记录以确保引用完整性。

[0159] 耦合到删除步骤822的是截断步骤824。如果加载成功,则截断步骤824可以截断所有统一的阶段表804以为后续的增量加载做准备。一旦数据203已经被加载到运营数据存储206中,用户可以使用在下文中关于图9-图15更加详细地描述的图1的商业智能工具134来充分利用和访问数据203。

[0160] 现在参考图9,其中示出了根据本发明的实施例的用于过滤队列的示例性控制流900。示例性控制流900可以是关于图1的过滤器128如何对图1的模式定义语言126进行操作的示例性说明。模式定义语言126的元数据302可以从图3的元数据表300、图5的模式定义模型表500、图4的模式定义模型400、或其组合中引用。

[0161] 一旦图2的运营数据存储206被填充,在填充ODS步骤712中,图1的商业智能工具134是完全功能的,因为其对模式定义语言126进行操作。当运营数据存储206被填充时,总队列902将是可用的。总队列902可以包括在运营数据存储206内表示的图4的实体404的全部。

[0162] 总队列902可以是用于过滤运行数据存储206的开始点。总队列902的实体404可以使用模式定义语言126内的图4的特性观察408中的任何一个而被过滤。模式定义语言126内的特性观察408可以由与特性观察408相关联的图3的元数据302来标识。

[0163] 在耦合到填充ODS步骤712的选择特性步骤904中,用户可以选择所选择的特性905。所选择的特性905可以是特性观察408中的一个。作为在上文中所使用的健康保健示例的扩展,用户可能想要利用针对乳腺癌的ICD9诊断来确定实体404的病人类别。

[0164] 商业智能工具134从模式定义语言126中返回所有特性观察408,例如,有效的ICD9选择。用户可以选择特性观察408中的一个作为所选择的特性905,并且进一步过滤所选择的特性905,例如,选择ICD9特性并且仅针对对应于乳腺癌的ICD9码来对其进行过滤。

[0165] 在用户选择了利用其来过滤总队列902的所选择的特性905之后,在查询步骤906中,商业智能工具134可以针对与所选择的实体905相关联的图4的模块406和实体404来查询模式定义语言126。模式定义语言126查询可以返回与模块406、实体404、以及涉及所选择的特性905的其它特性观察408相关的元数据302。此外,与所选择的特性905相关联的模块406和实体404可以在图2的物理数据表218内被标识。商业智能工具134还可以返回与所选择的特性905相关联的图2的实体类型220和模块类型224。

[0166] 在模式定义语言126被查询之后,耦合到查询元数据步骤906的关联特性步骤908可以根据模式定义语言126返回与所选择的特性905相关联的模块406和实体404。如果所选择的特性905仅与实体404相关联,则没有模块406与所选择的特性905相关联并且只有实体404被返回。

[0167] 利用与所选择的特性905相关联的实体404和模块406的返回,耦合到关联特性步骤908的运算步骤910可以运算当前的队列912以及所有的单跳链接914。当运算步骤910运算当前的队列912时,模式定义语言126可以用于定位与具有所选择的特性905的病人相对应的运营数据存储206的物理表218。

[0168] 可以基于模式定义语言126的元数据302来生成查询,以提取与所选择的特性905相关联的实体404的图2的数据203。运算步骤910可以自动地生成SQL指令915以通过对特性观察408、模块406、以及实体404的图6的链接608和物理表602进行定位来查询物理表218并且返回与所选择的特性905相关联的当前队列912。模式定义语言126可以作为可以用于找到针对物理数据表218中的每个的列的元数据表300而被引用,其中,数据203是物理地定位的。

[0169] 商业智能工具134可以将当前的队列912返回为与所选择的特性905相关联的多个实体404。运算步骤910还可以运算单跳链接914并且将其作为选择项呈现给用户以用于进一步细化过滤器128。

[0170] 单跳链接914可以与所选择的特性905所对应的实体404和模块406相对应的其它未经选择的特性观察408。即,单跳链接914是来自与所选择的特性观察408相关联的相同的模块406和实体404的未经选择的特性。与运算未选择的特性一起,运算步骤910还可以将模

块406和实体404返回作为单跳链接914。当与所选择的特性905相关联的模块406和实体404由其它不相关联的模块406或实体404所指向时,这些不相关联的模块406或者实体404可以被返回作为单跳链接914。也就是说,当针对任何特性观察408的图5的目标实体类型512或者图5的目标模块类型514包括指向与所选择的特性905相关联的模块406或者实体404的数据指针时,与具有数据指针的特性观察408相关联的模块406和实体404将被返回作为单跳链接914。

[0171] 为了利用基于乳腺癌的过滤的示例来说明单跳链接914的运算,运算步骤910可以运算针对与所选择的特性905 (例如,年龄、性别、体重、以及对应于实体404的其他非长度特性观察408) 相关联的实体404的单跳链接914。运算步骤910还可以针对与所选择的特性905 (例如,诊断的日期、诊断的年龄、基本位点、或者对应于模块406的其它长度特性观察408) 相关联的模块406来运算单跳链接。

[0172] 运算步骤910还可以返回与所选择的特性905不相关联的其它模块406或实体404的单跳链接914。例如,如果由用户选择的特性观察408包括针对乳腺癌的ICD9,则ICD9特性对应于诊断模块,并且运算步骤910可以随后将选项返回给用户以利用治疗过滤,因为治疗模块包括指向诊断模块的目标模块类型514中的指针。

[0173] 耦合到运算步骤910的是选择另一个特性选项916。如果用户决定不选择更多的特性观察408,则选择另一个特性选项916将在结束步骤918中终止过滤器128的操作。当结束步骤918被调用时,当前的队列912可以仍然被利用或被保存以进一步与商业智能工具134一起使用。

[0174] 如果用户在选择另一个特性选项916中决定选择另一个特性,则后续选择的特性919可以在耦合到选择另一个特性选项916的选择另一个特性步骤920中被选择。用户可以以类似于用户在选择特性步骤904中选择所选择的特性905的方式来选择后续选择的特性919。

[0175] 在选择另一个特性步骤920中与选择特性步骤904的用户体验之间一个区别是在选择另一个特性步骤920中,用户被给与单跳链接914作为选择项,并且显示当前队列912而不是总队列902。

[0176] 一旦用户选择后续选择的特性919,则在耦合到选择另一个特性步骤920的链接特性步骤922中,后续选择的特性919可以被链接到所选择的特性905。链接特性步骤922可以将所选择的特性905链接到后续选择的特性919以便于细化过滤器128并且限制当前的队列912。假设后续选择的特性919将限制过滤器128的当前实体404或者模块406;然而,可以预期,用户可以覆盖该逻辑“与(AND)”操作并且针对后续的特性观察408来利用逻辑“或(OR)”功能。

[0177] 作为说明性示例,用户可以选择对应于与所选择的特性905 (例如,年龄35-40或者男性) 相关联的实体404的单跳链接914中的一个。继续ICD9乳腺癌的示例,链接特性步骤922将第一所选择的ICD9特性 (所选择的特性905) 与后续选择的年龄或者性别特性观察408 (后续所选的特性919) 相链接。链接特性步骤922将对所选择的特性905与后续选择的特性919二者进行组合以进一步细化仅针对年龄35-40之间的或者是男性的病人的乳腺癌诊断的过滤器128。

[0178] 作为第二说明性示例,用户可以选择对应于与所选择的特性905 (例如,诊断日期

2007) 相关联的模块406的单跳链接914中的一个。继续ICD9乳腺癌示例, 链接特性步骤922将第一所选择的ICD9特性(所选择的特性905)与后续选择的日期特性(后续选择的特性919)相链接。链接特性步骤922将对所选择的特性905与后续选择的特性919两者进行组合以进一步细化仅针对2007年的乳腺癌诊断的过滤器128。

[0179] 作为第三说明性示例, 用户可以选择对应于到与所选择的特性905(例如, 治疗类型: 手术) 相关联的模块406或者特性观察408的链接608的单跳链接914中的一个。继续ICD9乳腺癌的示例, 链接特性步骤922将所选择的ICD9特性(所选择的特性905)与后续选择的指针特性(后续选择的特性919)相链接。链接特性步骤922将对所选择的特性905和后续选择的特性919两者进行组合以进一步细化过滤器128仅针对通过手术来治疗的乳腺癌诊断。

[0180] 链接特性步骤922耦合到查询元数据步骤906。一旦客户选择后续选择的特性919, 并且所选择的特性905和后续选择的特性919在链接特性步骤922中被链接, 则查询元数据步骤906将以类似于在上文中所讨论的方式来查询模式定义语言126。

[0181] 一旦查询元数据步骤906已经查询了模式定义语言126, 相关联的特性步骤908可以以类似于在上文中所描述的方式来操作, 除了所有所选择的特性905都被使用并且如逻辑“与”功能一样操作。类似地, 运算步骤910可以以类似于上文中所讨论的方式针对后续选择的特性919来运算单跳链接914, 并且以类似于上文中讨论的方式通过利用模式定义语言126查询运行数据存储206的物理表218来返回当前的队列912。

[0182] 现在参照图10, 其中示出了在商业智能工具126中实现的过滤器128的屏幕截图1000。屏幕截图1000描绘了链接到后续选择的特性919的所选择的特性905的示例。后续选择的特性919可以被链接到所选择的特性905, 作为利用其来过滤总队列902的第二特性观察408。链接可以表示逻辑“与”要求; 然而, 可以预期用户可以切换链接图标以调用总队列902的逻辑“或”过滤。

[0183] 总队列902被示出为过滤器128的开始队列。所选择的特性905被示出为针对乳腺癌所过滤的ICD9值。所选择的特性905被上下文地链接到后续选择的特性919。

[0184] 后续选择的特性919被示出为诊断日期, 其被过滤以返回2010年1月1日与2013年1月1日之间的日期。当前队列912被描绘为与当前队列912内的实体404的数量一起的实体404的分组。

[0185] 现在参照图11, 其中示出了用于利用根据本发明的实施例的模式来过滤队列的示例性控制流1100。示例性控制流1100可以是关于图1的过滤器128如何对图1的模式定义语言126进行操作的另一个或进一步的示例性说明。模式定义语言126的元数据302可以通过利用图3的元数据表300、图5的模式定义模型表500、图4的模式定义模型400、或其组合来引用。

[0186] 这呈现了用于标识图4的实体404的类别并且通过扫描图2的运营数据存储206, 查找与对应于临床事件的图4的特性观察408的特定的系列相匹配的实体404的另一个方法。用户可以选择特性观察408作为所选择的特性905以定义临床模式, 并且使用图1的商业智能工具134来标识与临床模式匹配的以及与临床模式不匹配的实体404。

[0187] 当运营数据存储206被填充时, 图9的总队列902将是可用的。总队列902可以包括被表示在运营数据存储206内的实体404的全部。控制流1100可以根据由运算步骤910运算的总队列902或者根据当前队列912来操作。在该说明性示例中, 提供当前队列912的运算步

骤910将用于通过使用基于模式过滤当前的队列912的控制流1100来进一步细化过滤器128。

[0188] 耦合到运算步骤910的是选择特性步骤1102,其中特性观察408可以被选择为所选择的特性905。用户可以选择的特性观察408包括单跳链接914但是也可以包括不相关的特性观察408。一旦用户选择了所选择的特性905,商业智能工具134将反馈特性和耦合至选择特性步骤1102的展示特性模块1104中的当前的队列912。用户然后具有在耦合到展示特性模块1104的选择选项1106中选择另一个特性的选项。如果用户不想选择更多的特性观察408,则选择选项1106将把用户带到选择特性步骤1102并且用户可以选择后续的特性。以该方式,用户可以选择对应于临床事件的特性观察408并且定义利用其来限制过滤器128的模式。

[0189] 继续在上文中所利用的健康保健领域的示例,假设用户希望标识在出院30天内重新入院的阑尾切除手术的病人。用户可以选择针对阑尾切除手术的特性观察408并且商业智能工具134工具可以利用图9的控制流900找到已经进行了阑尾切除手术的所有实体404。一旦已经标识了与阑尾切除手术特性观察408相关的所有实体404,则用户可以创建在选择特性步骤1102中开始的新的临床模式。用户可以在针对医院入院的特性观察408之后选择针对医院出院的特性观察408。针对医院入院的特性观察408可以由指示在出院30天之内发生入院的特性观察408来限制。该示例临床模式将涉及选择与选择对单跳链接914的最大相距30天的限制相关联的两个特性观察408(出院和入院)。如在上文中所提及的,单跳链接914是属于与所选择的特性905相关联的相同的图4的模块406或者实体404的未经选择的特性观察408。

[0190] 用户利用名称——“阑尾切除手术在30天内重新入院”来保存该模式。

[0191] 如果用户已经通过选择对应于临床模式的特性观察408而定义了临床模式,则用户可以选择运算匹配步骤1108并且保存临床模式。运算匹配步骤1108耦合到引用了模式定义语言126的查询元数据步骤1110,以确定图6的链接608和包含所选择的特性905的图2的物理表218的位置。

[0192] 查询元数据步骤1110耦合到生成SQL步骤1112。生成SQL步骤1112可以利用根据查询元数据步骤1110中的模式定义语言126所返回的物理表218的关系和位置来自动地生成SQL指令114以查询运营数据存储206的物理表218。

[0193] 耦合到生成SQL步骤1112的运行SQL步骤1116可以运行从生成SQL步骤1112中自动地生成的SQL指令1114以查询运营数据存储206的物理表218。耦合到运行SQL步骤1116的是返回步骤1118。返回步骤1118可以返回匹配的队列1120和不匹配的队列1122。

[0194] 匹配队列1120可以是与作为临床模式的所选择的特性905相对应的实体404。非匹配队列1122可以是不与作为临床模式的所选择的特性905相对应的实体404。

[0195] 匹配队列1120和非匹配队列1122可以被保存以由商业智能工具134之后进行处理或进一步限制。当返回步骤1118返回匹配队列1120时,用户可以查看匹配队列1120内的实体404的细节。例如,用户可以查看ID并且可以查看与由所选择的特性905所定义临床模式的匹配何时发生。

[0196] 现在参考图12,其中示出了在图1的商业智能工具126中所实现的图1的过滤器128的屏幕截图1200。屏幕截图1200描绘了特性观察408的示例。特性观察408可以被显示以供



用户在选择框1202中进行选择。

[0197] 用户选择的特性观察408可以在模式框1204中被示出。所选择的特性905与后续选择的特性919可以在模式框1204中被示出为链接在一起。例如,所选择的特性905可以是针对阑尾切除手术所过滤的治疗或者过程。

[0198] 第一后续选择特性1206可以例如是出院。第二后续选择特性1208可以例如是重新入院。可以包括模式过滤器1210以将第一后续选择特性1206以及第二后续选择特性1208的返回限制为小于特性观察408的每个实例的30天。换言之,模式过滤器1210可以将返回的图4的实体404限制为进行了阑尾切除手术的、出院的、并且在出院的30天内重新入院的实体404。

[0199] 现在参考图13,其中示出了根据本发明的实施例的用于生成报告的示例性控制流1300。示例性控制流1300可以是关于图2的报告数据定义130如何对图1的模式定义语言126进行操作的示例性说明。可以使用图3的元数据表300、图5的模式定义模型表500、图4的模式定义模型400、或者其组合来引用模式定义语言126的元数据302。

[0200] 一旦图2的运营数据存储206被填充,用户就能够对运营数据存储206内的图4的实体404中的任何一个进行报告。用户可以定义返回的报告数据。图1的商业智能工具134利用模式定义语言126的元数据302来构建根据图4的实体404、模块406、以及特性观察408所定义的报告数据定义130。商业智能工具134执行报告数据定义130以从运营数据存储206中提取图2的数据203,并且通过一个或多个电子表格来呈现结果。用户还通过下文关于图15更加详细地描述的各种报告查看器来查看并且总结电子表格数据。

[0201] 在开始报告步骤1302中,发起报告数据定义130。开始报告步骤1302包括当前队列912,报告数据定义130可以对其进行操作。可以预期图9的总队列902、图11的匹配队列1120、或者甚至图11的非匹配队列1122都可以由报告数据定义130使用以对其进行操作;然而,出于说明的目的,示出了当前队列912。

[0202] 耦合到开始报告步骤1302的是定义数据步骤1304。定义数据步骤1304包括报告数据定义130。报告数据定义130指定哪些实体404和特性观察408被包括在报告1308中。报告数据定义130包括默认实体404和特性观察408。如在上文中关于图9和图11更加详细地描述的,默认实体404是当前队列912(或者任何当前所选择的队列),而默认特性观察408是实体404ID并且特性观察408用于选择实体404

[0203] 一旦报告数据定义130被创建并且被利用默认特性观察408和实体404定义,用户就具有选项以在耦合至定义数据步骤1304的修改选项1309中修改报告数据定义130。如果用户想要改变报告数据定义130,则可以选择耦合到修改选项1309的修改实体选项1310或者修改特性选项1312。

[0204] 如果修改实体选项1310或者修改特性选项1312是由用户选择的,则模式定义语言126将由商业智能工具134来引用。当用户选择修改报告数据定义130内的实体404时,用户将选择修改实体选项1310。

[0205] 如果用户决定通过添加实体来修改实体404,则引用步骤1314将引用模式定义语言126并且确定到其它实体404的图6的所有连接608,当添加作为当前队列912的子实体的实体404时,这尤其重要。在耦合到引用步骤1314的链接步骤1316中,确定链接608。

[0206] 一旦确定了额外的实体404与之前的实体404之间的链接608,就在运算步骤1318

中运算图9的单跳链接914。单跳链接914可以是与新选择的实体404相对应的特性观察408。即,单跳链接914是与新添加的实体404相关联的特性观察408。模块406还可以被返回为单跳链接914。当新添加的实体404由其它模块406或者实体404指向时,或者可替代地,当新添加的实体404指向其它模块406或者实体404时,这些指向或被指示的模块406或者实体404可以被返回为单跳链接914。

[0207] 运算步骤1318耦合到可以显示选项1322的显示步骤1320。当实体404被附加、添加、或者链接到报告数据定义130时,用户将被提供有选项1322。选项1322包括长度选项,例如,第一、最后、或者全部观察。选项1322还可以包括过滤选项,例如,基于被包括在报告中的特性的类型的过滤。过滤选项将仅保持满足过滤标准的结果。选项1322还包括约束选项,例如,限制的、非限制的。约束选项语义等价于SQL内连接对外连接。

[0208] 在选项1322已经被提供至用户之后,更新步骤1324将用新要求的实体404及其到当前队列912的链接来更新报告数据定义130。当将特性观察408添加到报告数据定义130时,商业智能工具134利用相似的流。

[0209] 当用户想要修改报告1308将显示的特性观察408时,可以选择修改特性选项1312。商业智能工具134将确定用户想要添加的特性观察408是否属于当前在确定选项1326内的报告数据定义130中的模块406。如果模块406当前不是报告数据定义130的一部分,则使用引用步骤1314来引用模式定义语言126。在链接步骤1316中运算链接608,并且最终在运算步骤1318中运算单跳链接914。

[0210] 单跳链接914可以是对应于与由用户添加的特性观察408相对应的实体404和模块406的其它未选择的特性观察408。即,单跳链接914是来自与所添加的特性观察408相关联的相同的模块406和实体404的非添加的特性。与运算非添加的特性观察408一起,运算步骤1318也可以返回模块406和实体404作为单跳链接914。当与所选择的特性905相关联的模块406和实体404由其它不相关联的模块406或实体404所指向时,这些不相关联的模块406或者实体404可以被返回为单跳链接914。也就是说,当具有任何特性观察408的图5的目标实体类型512或者目标模块类型514包括指向与由用户添加的特性观察408相关联的模块406或实体404的数据指针时,具有数据指针的与特性观察408相关联的模块406和实体404将被返回为单跳链接914。

[0211] 以与在上文中刚描述的添加实体404的类似的方式,显示步骤1320将选项1322显示给用户。当添加实体404和模块406时显示的选项1322可以是相同的;然而,可以预期,可以不同地或者取决于所添加的来显示额外的选项或对实体404或模块406唯一的选项。

[0212] 在已经向用户提供了选项1322之后,更新步骤1324将利用新要求的模块406及其到当前的队列912的链接608以及其中的模块406来更新报告数据定义130。当确定选项1326被调用并且确定特性观察408与已经在报告数据定义130内的模块406相关联时,引用步骤1314简单地引用模式定义语言126并且链接步骤1316确定用于将特性观察408并入报告数据定义130所要求的链接608。

[0213] 如果特性观察408与已经在报告数据定义130内的模块406相关联,则运算步骤1318和显示步骤1320不需要被调用来运算单跳链接914或者显示选项1322。更新步骤1324可以被调用以更新报告数据定义130并且将用户返回到修改选项1309。

[0214] 从在上文中所利用的ICD9乳腺癌的示例中继续,用户可能想要对质量特性观察

408进行报告,该质量特性观察408对应于针对由当前队列912所返回的癌症病人而在肿瘤样品实体404上运行但由指示乳腺癌的类型的特性观察408分层的分子实验模块406。这里,用户简单地将特性观察408“Ce1文件质量”和“三阴性”添加到报告数据定义130。当增加Ce1文件质量特性观察408时,引用了模式定义语言126以提供针对乳腺癌诊断特性观察408的到它们的诊断肿瘤实体404的链接608以及还有到在这些肿瘤实体404上运行的分子实验模块406的链接。因为三阴性特性观察408是已经被包括在报告数据定义130内的癌症诊断模块406的成员。

[0215] 当用户不再想要修改报告数据定义130时,用户可以在耦合到修改选项1309的执行步骤1328内执行报告。可以通过耦合到执行步骤1328的引用模式步骤1330来引用模式定义语言126,以确定包含在报告数据定义130内的特性观察408、模块406、以及实体404的运营数据存储206内的物理表218的位置。

[0216] SQL指令1332可以由耦合到引用模式步骤1330的生成SQL步骤1334自动地生成。SQL指令1332可以基于从引用模式步骤1330中所收集的信息来生成,因为针对所请求的数据203的物理表218的位置和其间的链接608可以根据模式定义语言126来确定。

[0217] 一旦SQL指令1332在运营数据存储206上运行,报告1308可以在耦合到生成SQL步骤1334的呈现步骤1336中作为电子表格被呈现给用户。报告1308可以在耦合到呈现步骤1336的报告步骤1338中被过滤、显示和分析。下文中关于图15更详细地讨论了报告步骤1338。

[0218] 作为示例,报告步骤1338可以提供数据透视表报告工具,并且提示用户利用用报告数据定义130所提取的数据203来填充数据透视表的值和维度。

[0219] 现在参照图14,其中示出了在图1的商业智能工具126中所实现的图1的报告数据定义130的屏幕截图1400。屏幕截图1400描绘了特性222连同与其相关的特性观察408的示例。

[0220] 默认被包括的特性观察408可以包括病人ID以及用于过滤图4的实体404的特性观察408。作为说明性示例,在屏幕截图1400中默认的特性观察408是病人ID和ICD9。

[0221] 为了生成针对临时队列1402的报告,用户可以选择应该被包括以用于分析的其它特性观察408。作为示例,用户可以选择Ce1文件质量和三阴性。如所描绘的,商业智能工具126可以包括与新添加的特性观察408相对应的实体404。

[0222] 针对特性观察408,Ce1文件质量、实体404实验和样本可以连同其对应的特性观察408实验ID和样本ID而被添加。如进一步描绘的,针对三阴性的特性观察408已经具有默认包括的模块406。具体而言,癌症诊断已经被包括为与ICD9特性观察408相对应的模块406。

[0223] 现在参照图15,其中示出了根据本发明的实施例的用于分析报告的控制流1500。控制流1500描绘了来自之前的图13的控制流1300的步骤的一部分,其具有额外的步骤以用于分析图13的报告1308。

[0224] 如所描绘的,细化步骤1502可以耦合到拉取数据步骤1504。细化步骤1502和拉取数据步骤1504可以包括控制流1300的步骤。例如,细化步骤1502可以包括图13的定义数据步骤1304、修改选项1309、修改实体选项1310、修改特性选项1312、引用步骤1314、链接步骤1316、运算步骤1318、显示步骤1320、以及更新步骤1324,以便于细化或修改图1的报告数据定义130。

[0225] 拉取数据步骤1504可以包括图13的步骤执行步骤1328、引用模式步骤1330、以及生成SQL步骤1334以拉取报告1308内的图2的数据203并且向用户显示报告1308。一旦报告1308在拉取数据步骤1504中被拉取,则耦合到拉取数据步骤1504的工具选择步骤1506可以提示用户选择分析工具1508。

[0226] 工具选择步骤1506可以包括分析工具选择,例如ANOVA(单向)工具、ANOVA(双向)工具、应急工具、Cox回归工具、Cox回归(多控制变量)工具、Kaplan Meier(两个日期)工具、K均值聚类工具、LiMMA工具、逻辑回归工具、Mann Whitney工具、单样本t测试工具、以及其它分析工具。一旦在工具选择步骤1506中选择了分析工具1508,则图1的商业智能工具134就可以提示用户选择分析工具1508运行所要求的输入。可以在耦合到工具选择步骤1506的提示步骤1510内请求输入。

[0227] 一旦用户定义了输入,在映射步骤1512中,用户就将来自报告1308的数据203映射到输入。一旦数据203被映射到分析工具1508的输入,商业智能工具134将执行分析并且向用户显示该分析。继续上文的ICD9乳腺癌示例,假设用户想要针对图4的乳腺癌病人实体404来比较和分析图4的生存时间特性观察408,其是由对病人实体404进行管理的图4的各种治疗协议特性观察408所分组的。用户对报告数据定义130进行细化以包括完成分析所需要的特性观察408;即——诊断的年龄特性观察408,生存时间特性观察408、治疗协议特性观察408、以及指示病人是否因为乳腺癌去世的标记特性观察408。

[0228] 在拉取数据步骤1504中拉取报告1308。用户可以接着在工具选择步骤1506中选择分析工具1508 Cox回归。用户在映射步骤1512中将来自报告1308的数据203映射到输入,并且执行分析。商业智能工具134执行分析并且以各种格式向用户呈现结果,例如,由治疗协议的每一对的统计比较支持的可视Kaplan Meier曲线。

[0229] 在上文中所描述的附图中的控制流和框图示出了根据本发明的各种实施例的系统、方法、以及计算机程序产品的可能的实现的架构、功能、以及操作。就此,控制流或框图中的每个框都可以表示代码的模块、分段、或者部分,其包括用于实现指定的逻辑功能的一个或多个可执行指令。应当注意,在一些可替代的实现中,在框中记录的功能可以以图中所记录的以外的顺序进行。例如,所示出的连续的框可以实际上大体同时执行,或者框有时可以以相反的顺序执行,取决于所涉及的功能。还应当注意,框图和/或流程图示的每个框、以及框图和/或流程图示中的框的组合可以由执行指定的功能或动作的基于专用硬件的系统、或者专用硬件和计算机指令的组合来实现。

[0230] 有利地,本发明的实施例提供了用于提供较高抽象等级的模式定义语言的技术,并且提供了用于与模式定义语言接合的平台,以使得商业智能工具能够以较低的成本和较短的时间帧来利用运营数据存储。模式定义语言使得商业智能工具能够以较高的抽象等级来操作,因此当底层数据库演进时,工具不再需要改变。当利用本发明的模式定义语言时,数据库演进是简单的并且是无干扰的。

[0231] 结果处理和配置是直接的、成本有效的、不复杂的、高度通用的、精确的、敏感的、以及有效的,并且可以通过调整已知的部件以适应准备好的、高效的、以及经济的应用和利用来实现。

[0232] 尽管前述内容是针对本发明的实施例的,但是可以设计出本发明的其它和进一步的实施例而不偏离本发明的基本范围,并且本发明的范围由以下的权利要求所确定。

[0233] 因此,本发明旨在包含落在所包括的权利要求的范围内的所有这样的替代、修改和变型。在本文中所阐述的或在附图中所示出的所有内容都将被解释为说明性的和非限定性的意义。

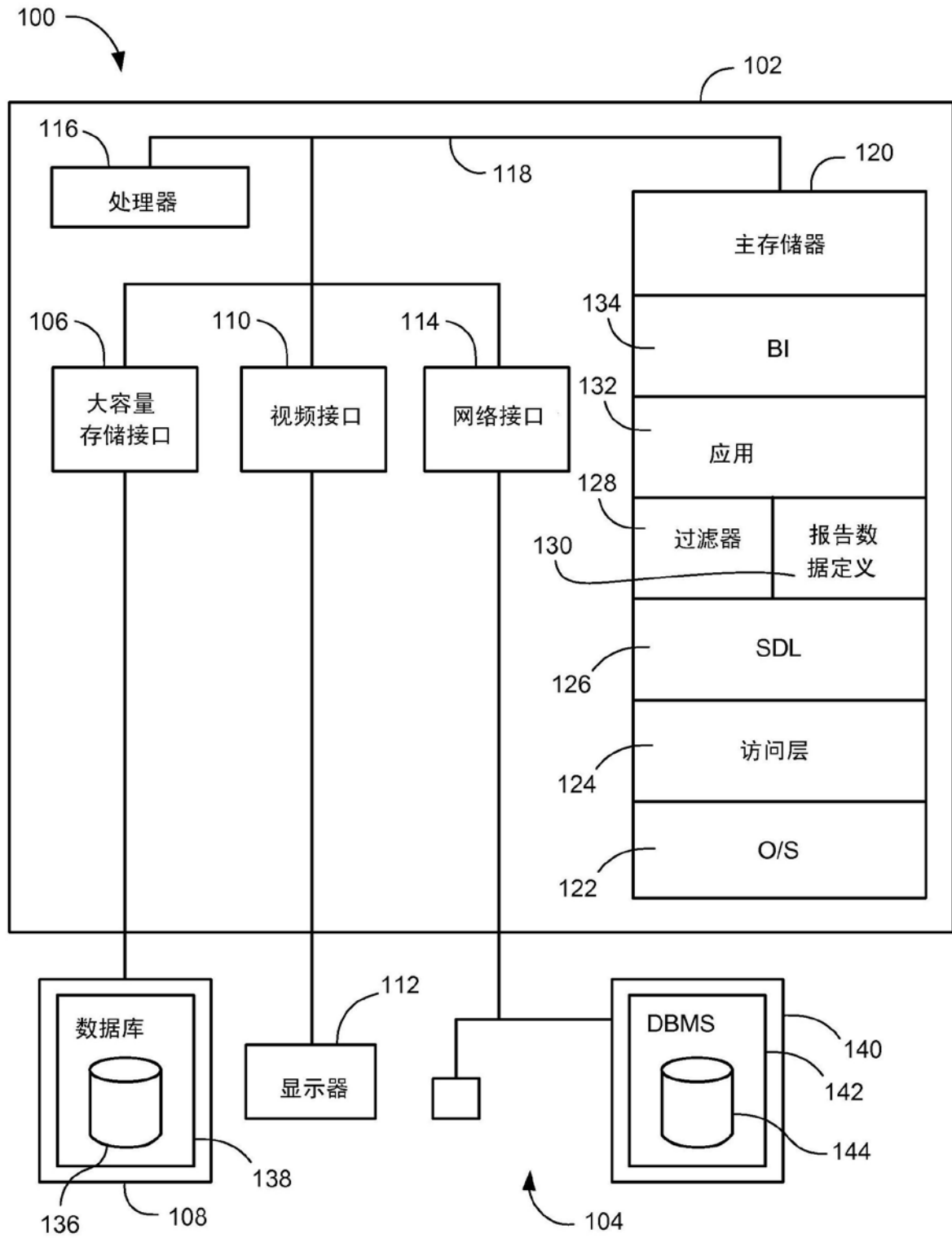


图1

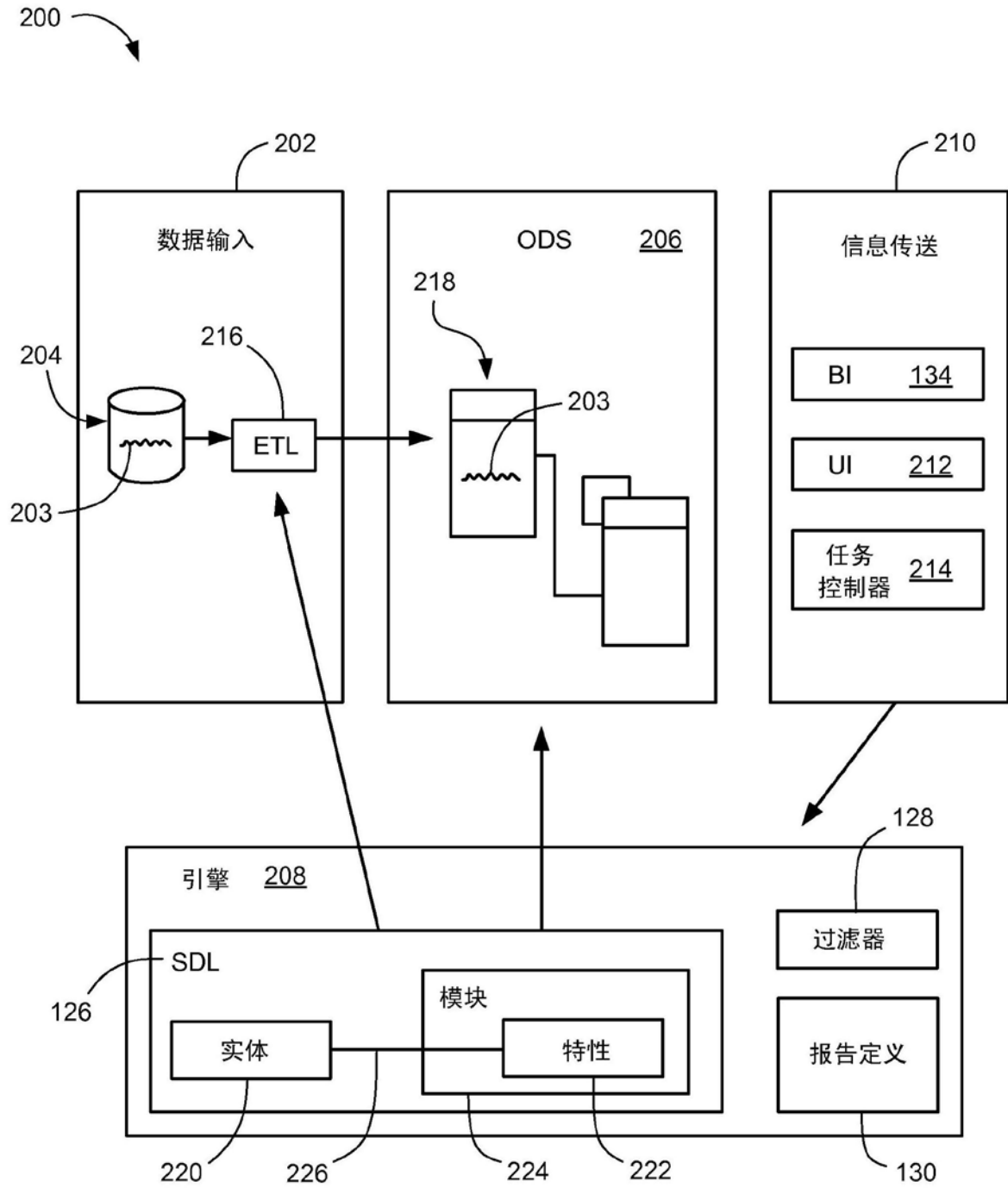


图2

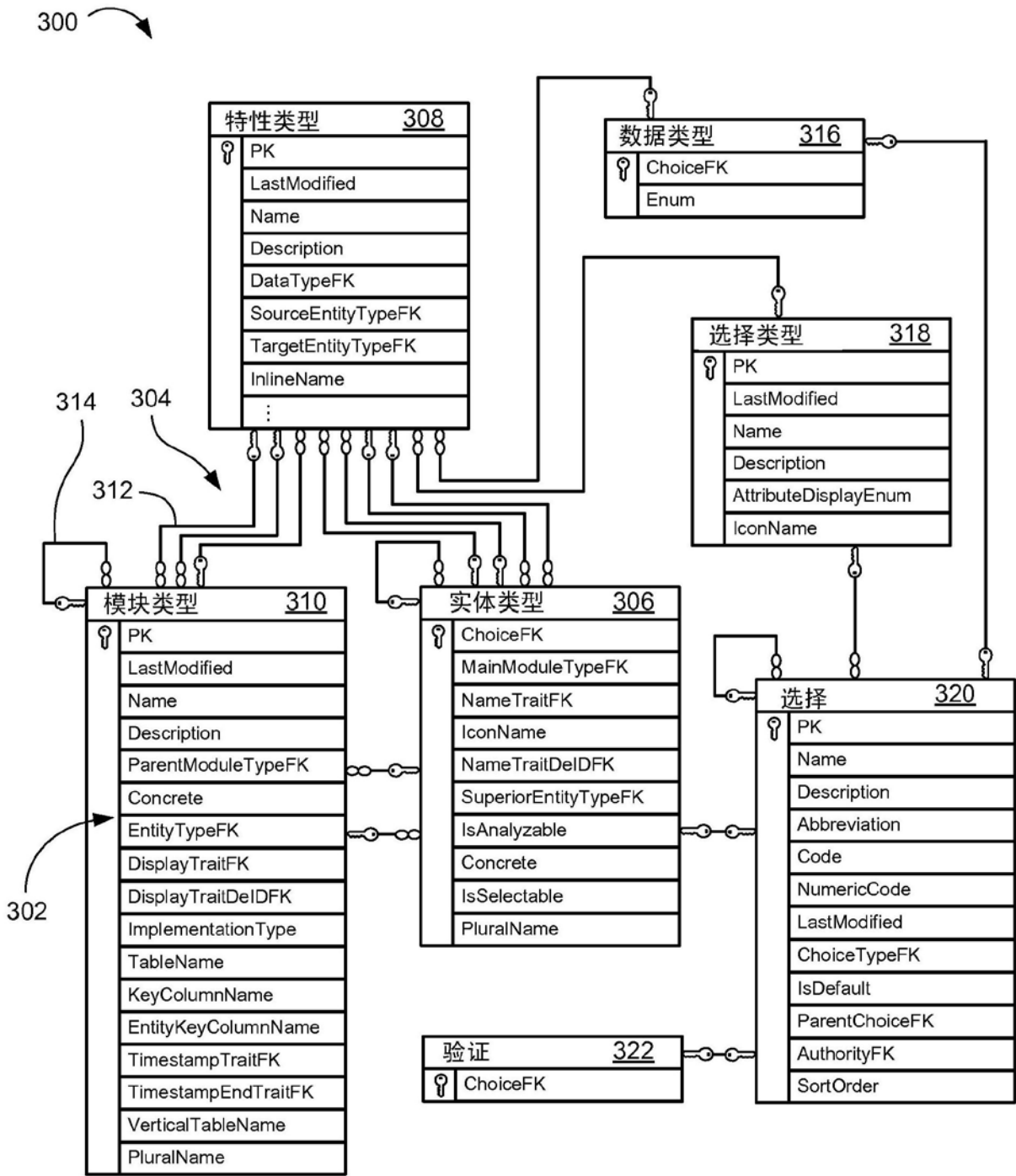


图3



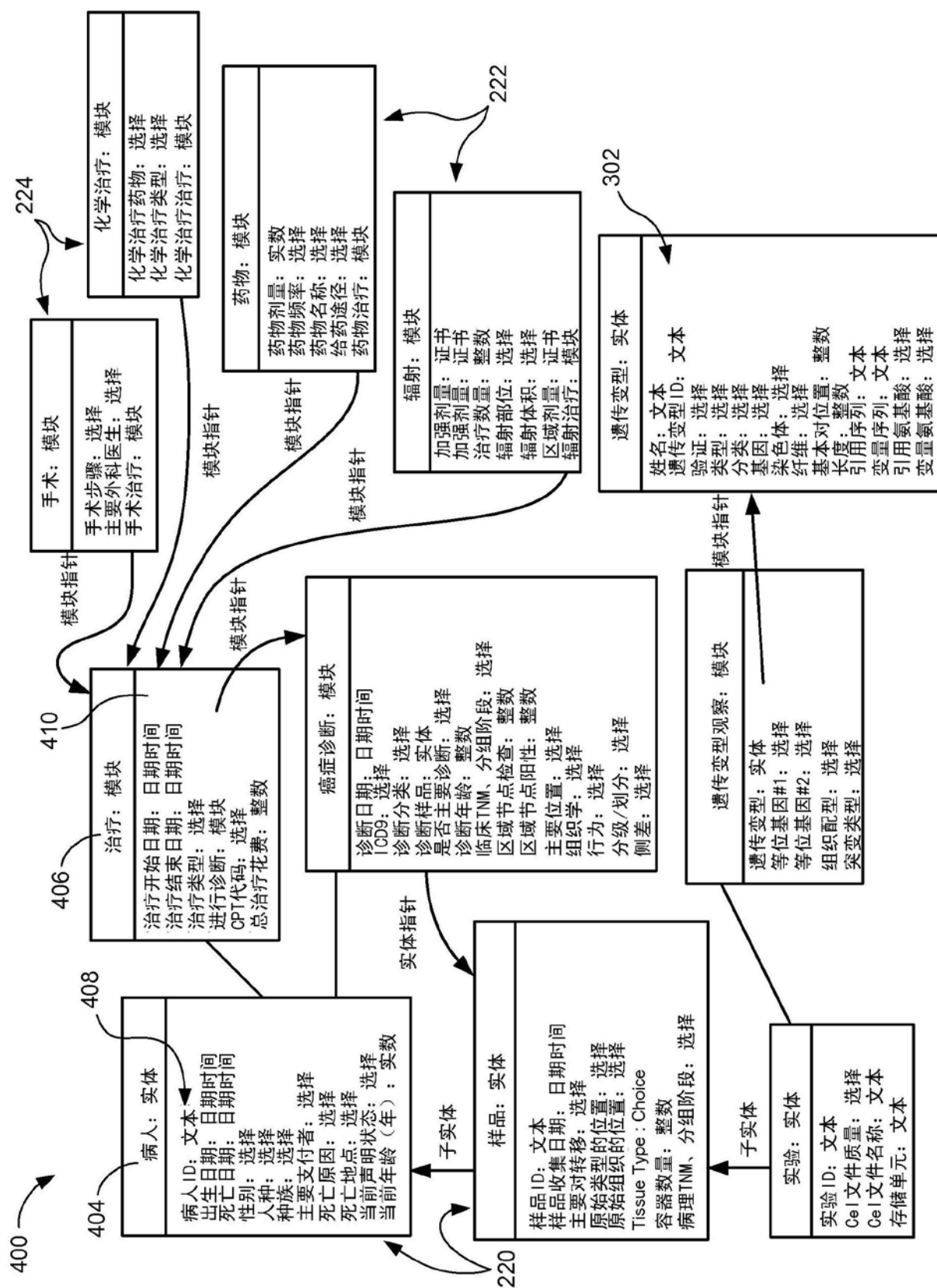


图4

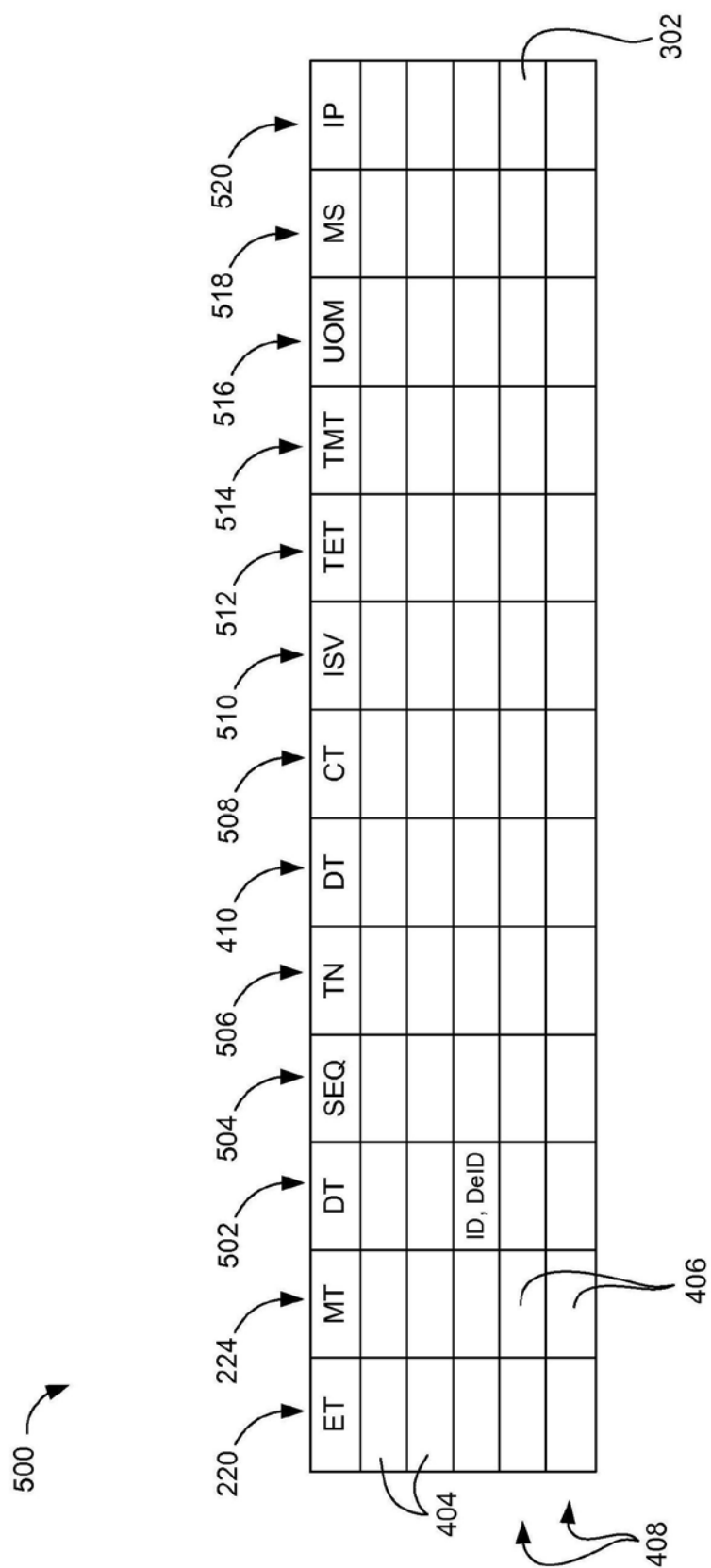


图5

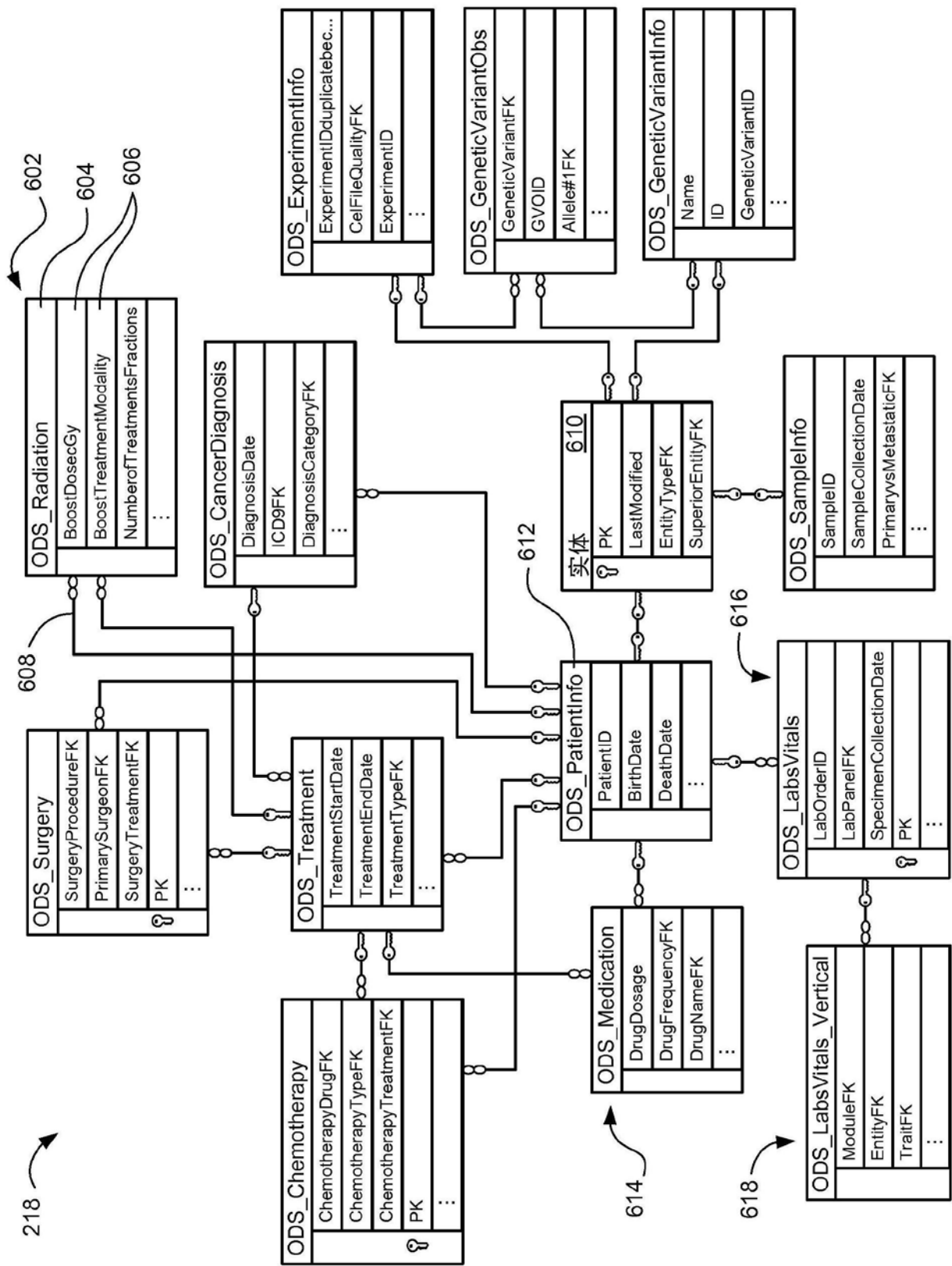


图6

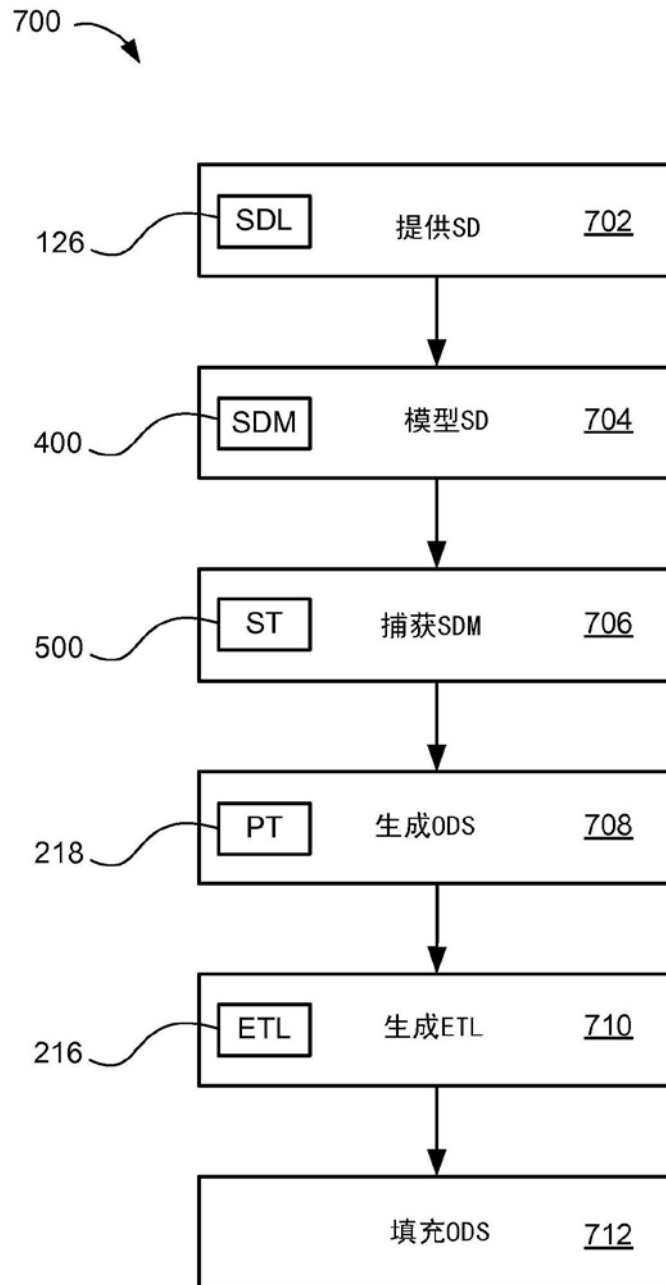


图7

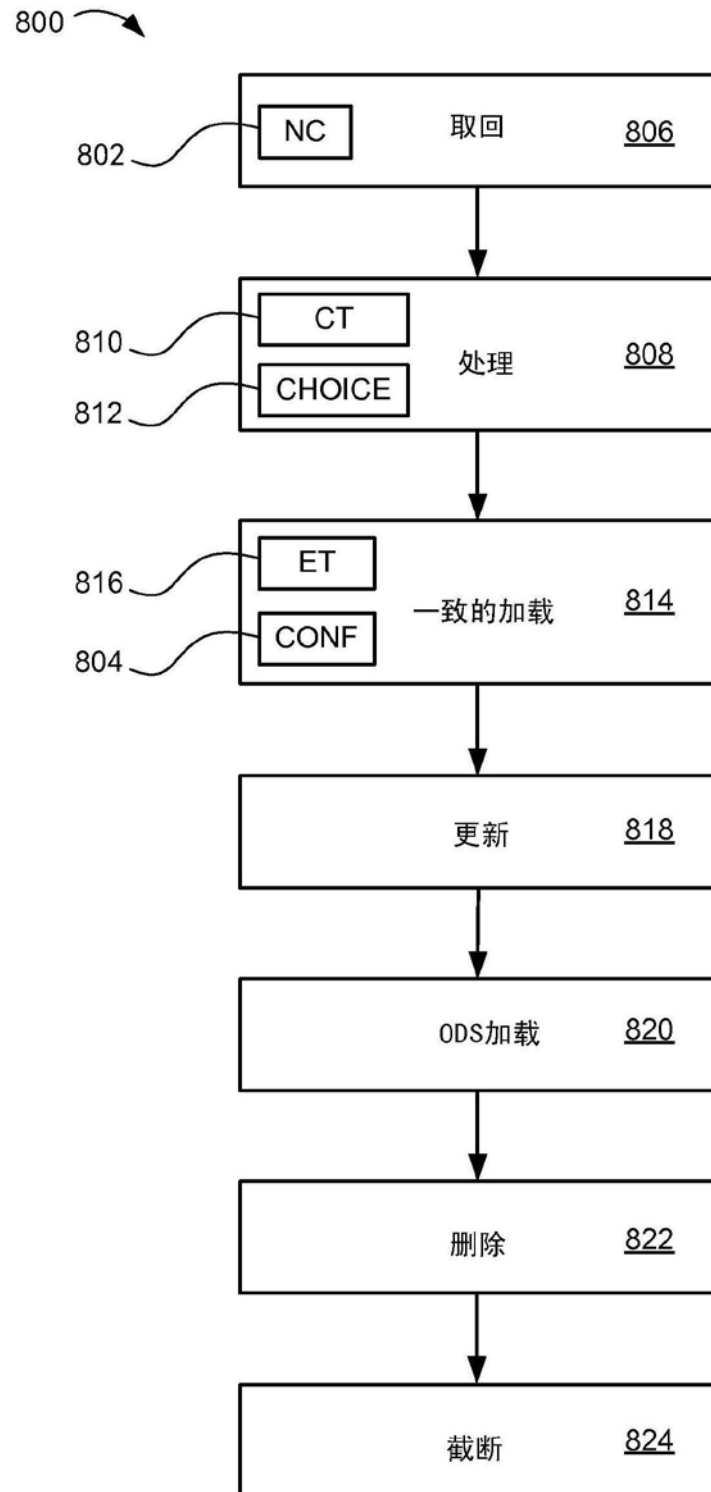


图8

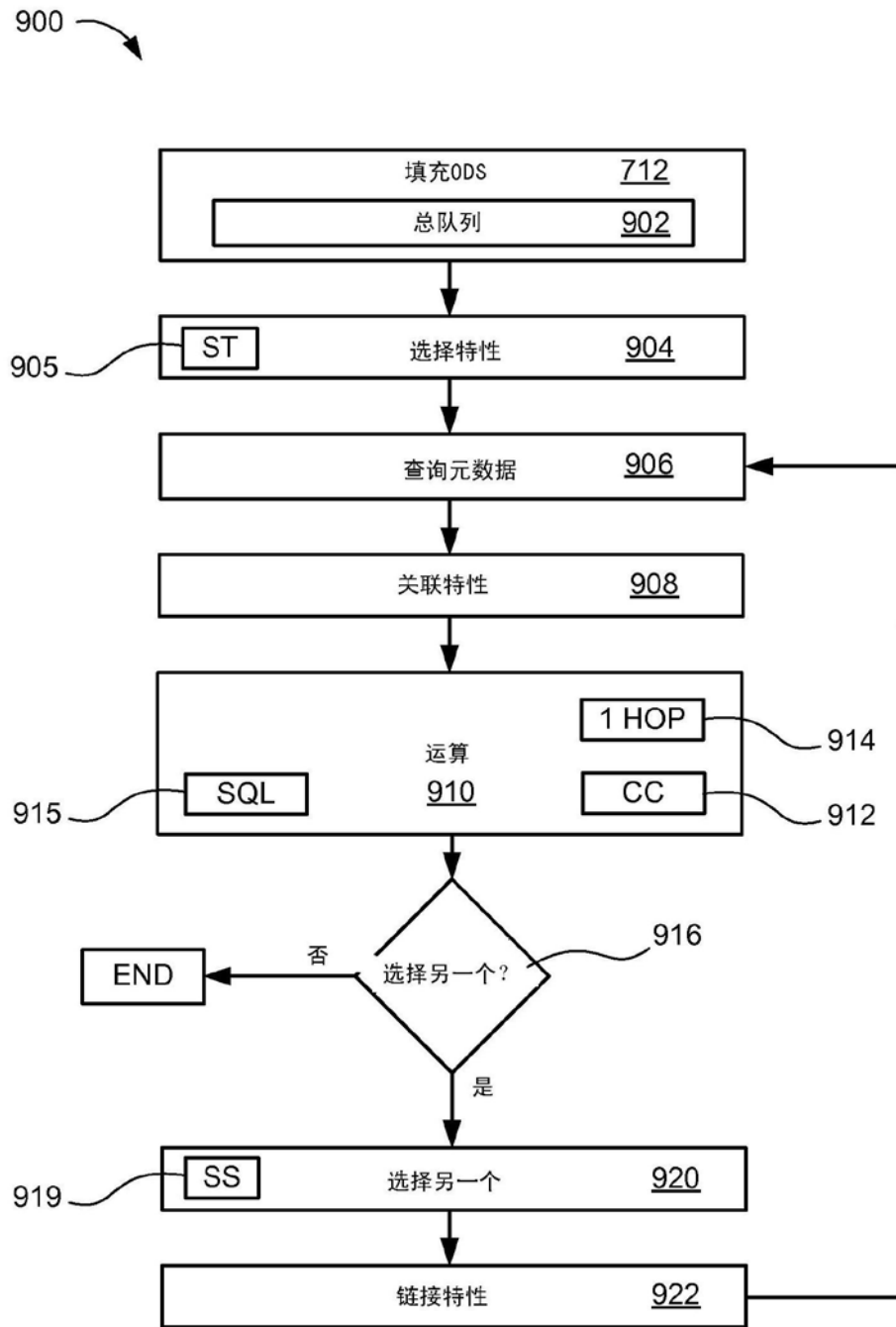


图9

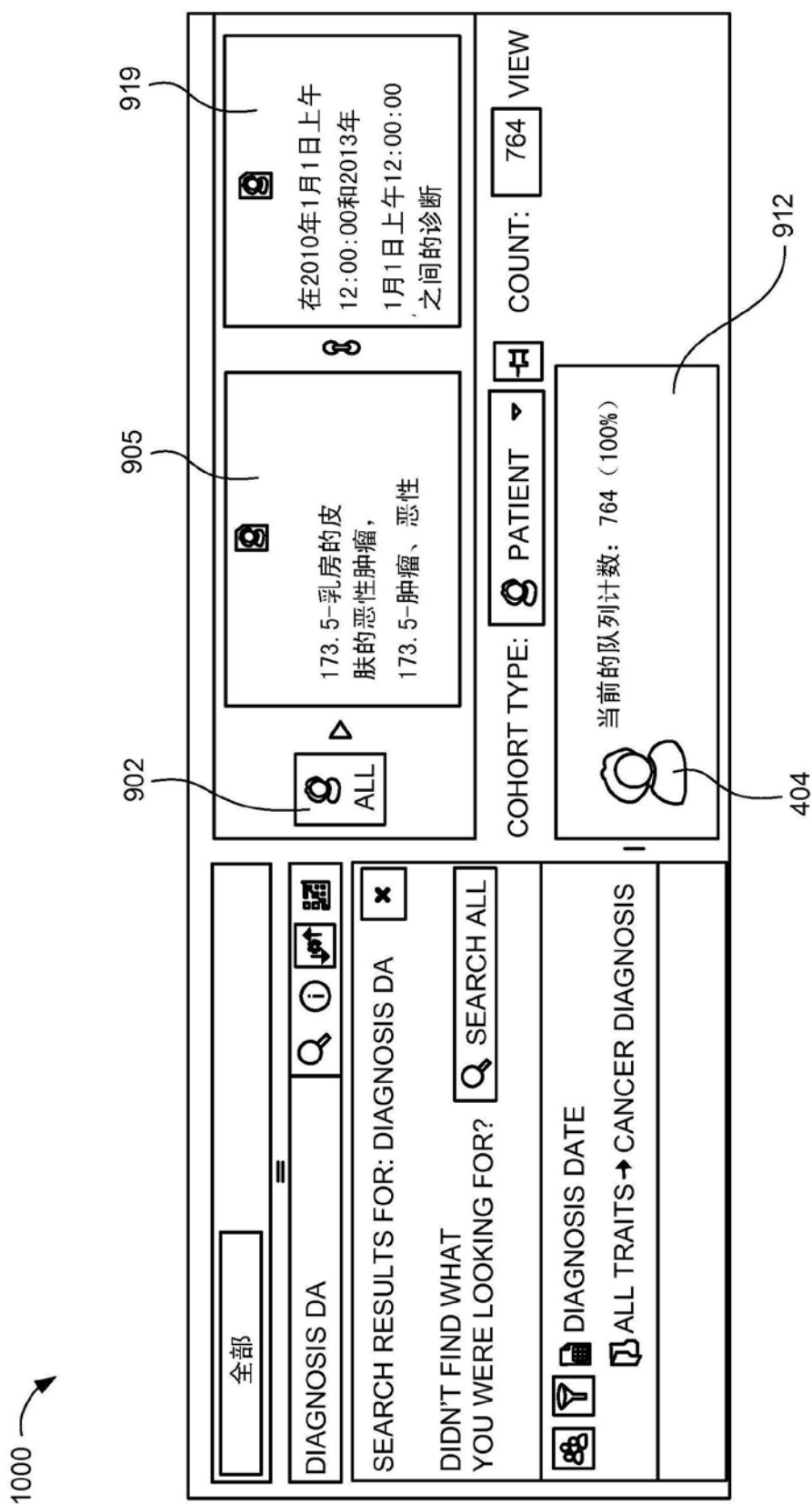


图10

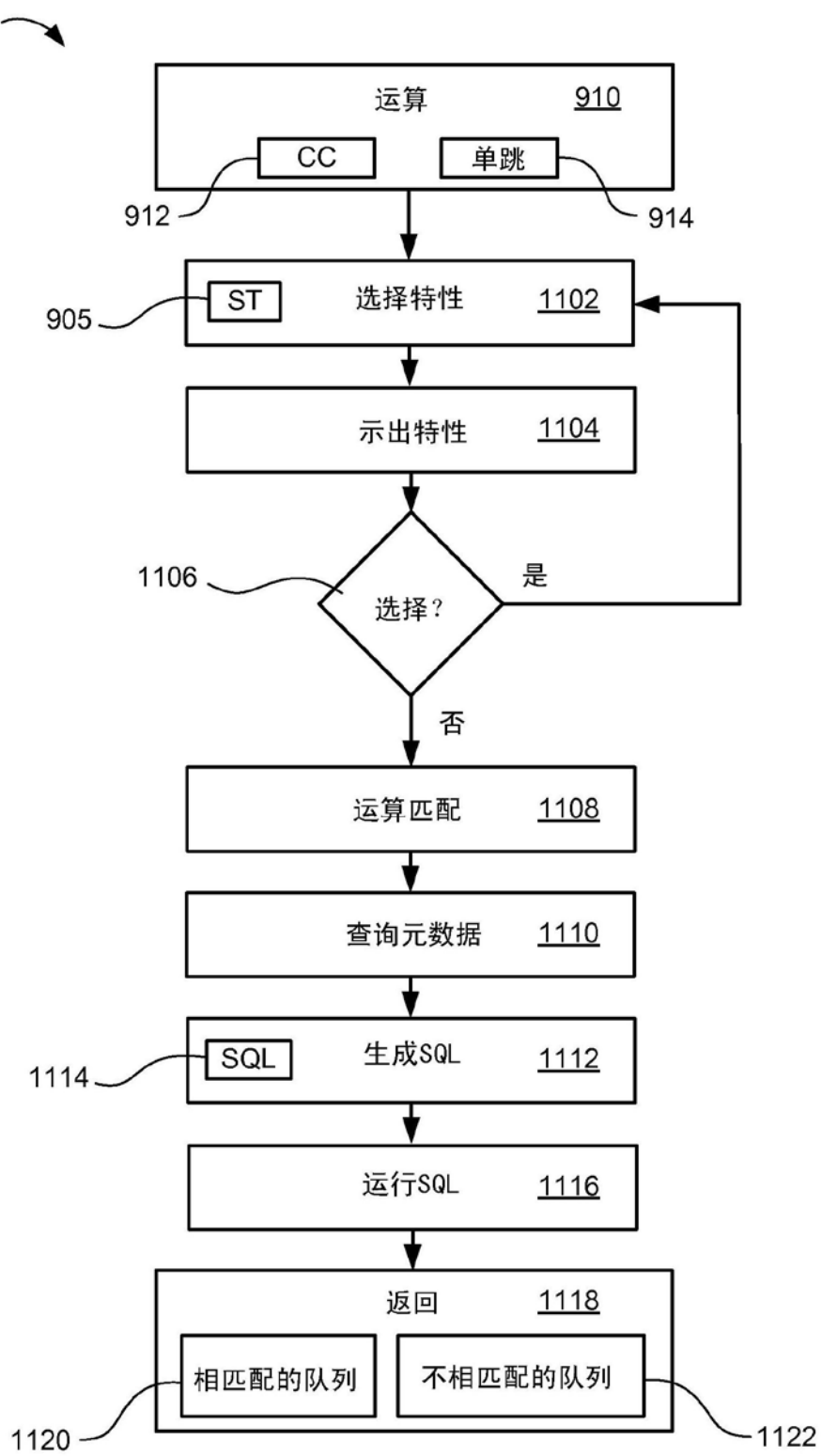


图11



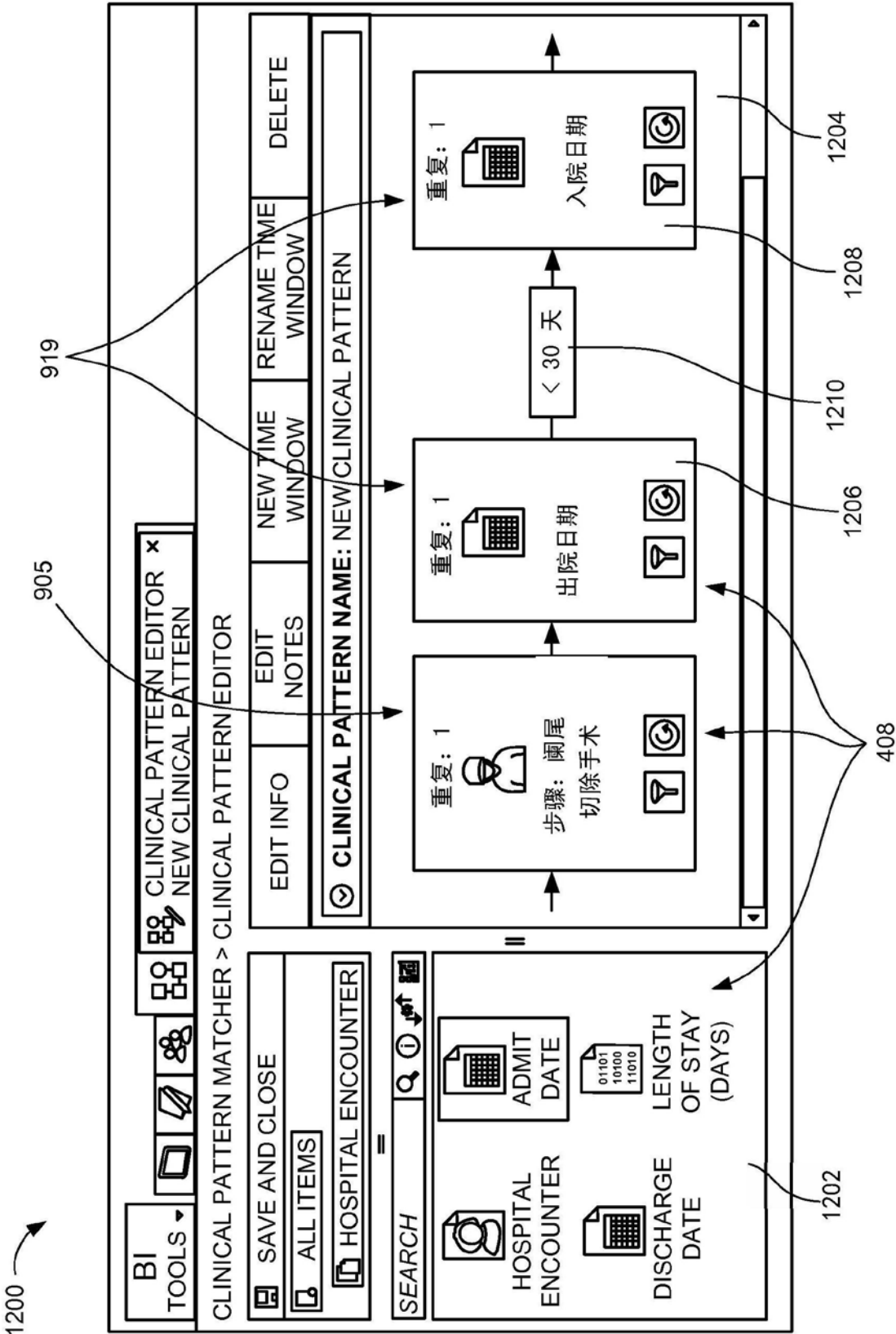


图12

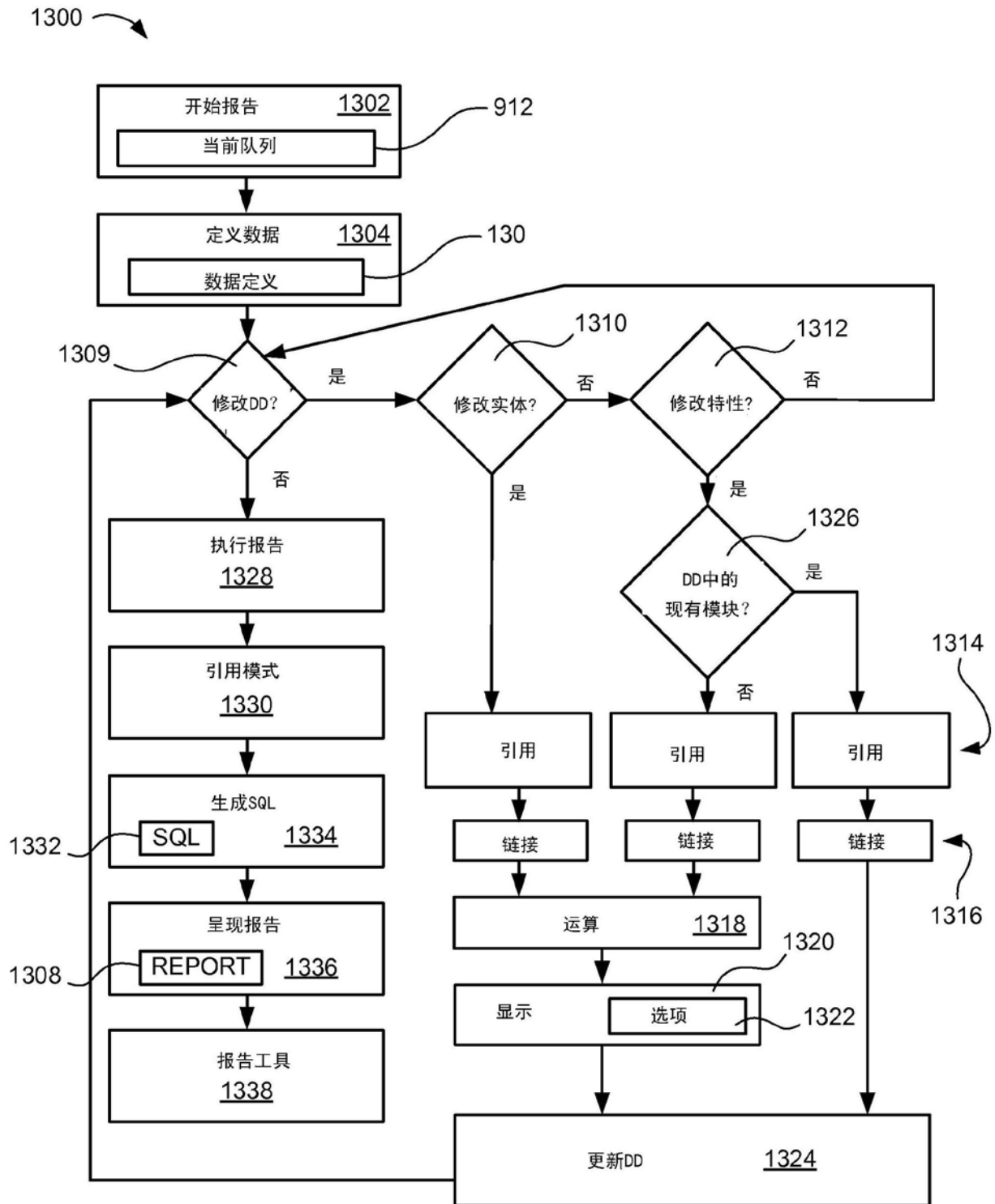


图13

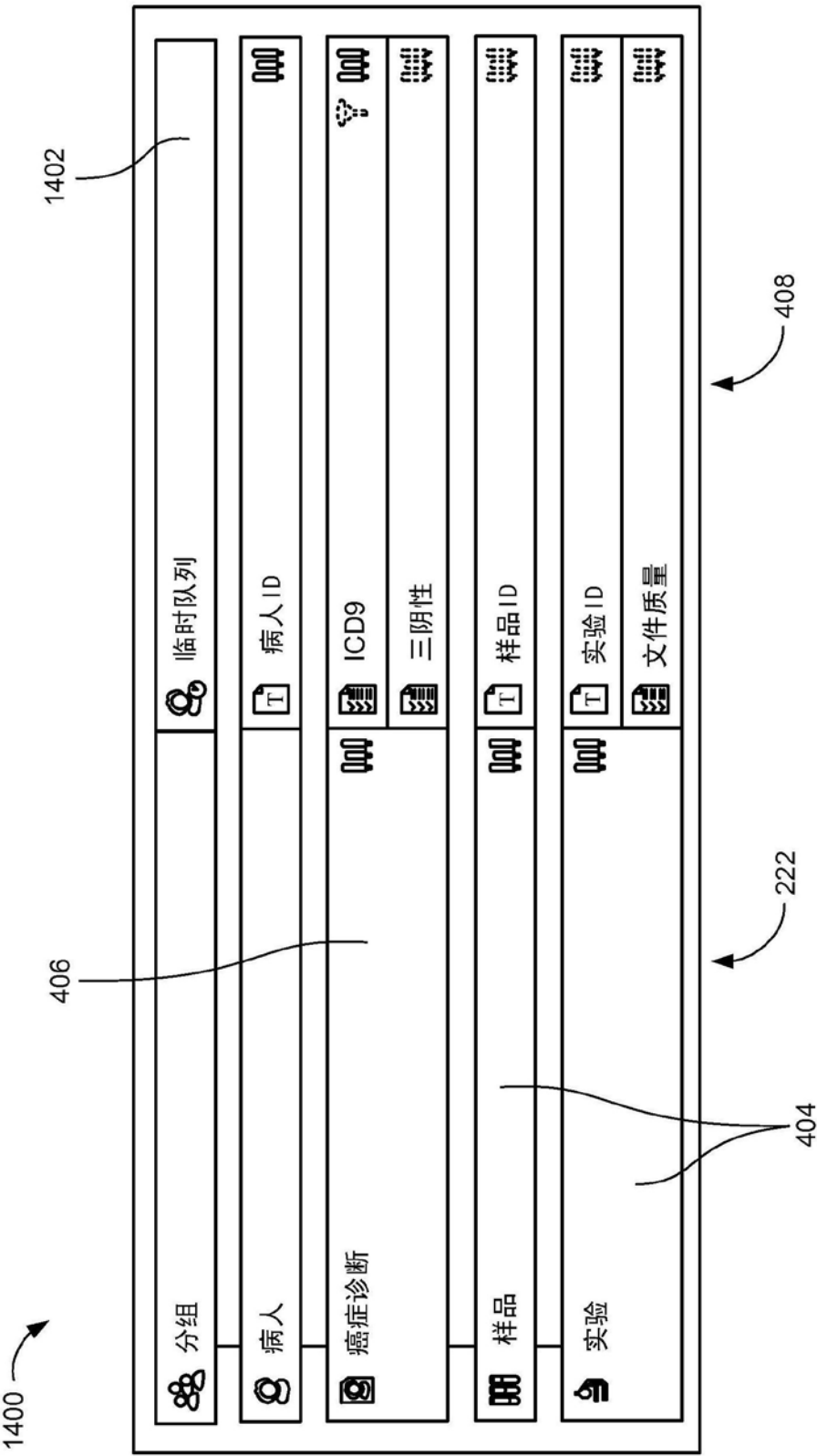


图14

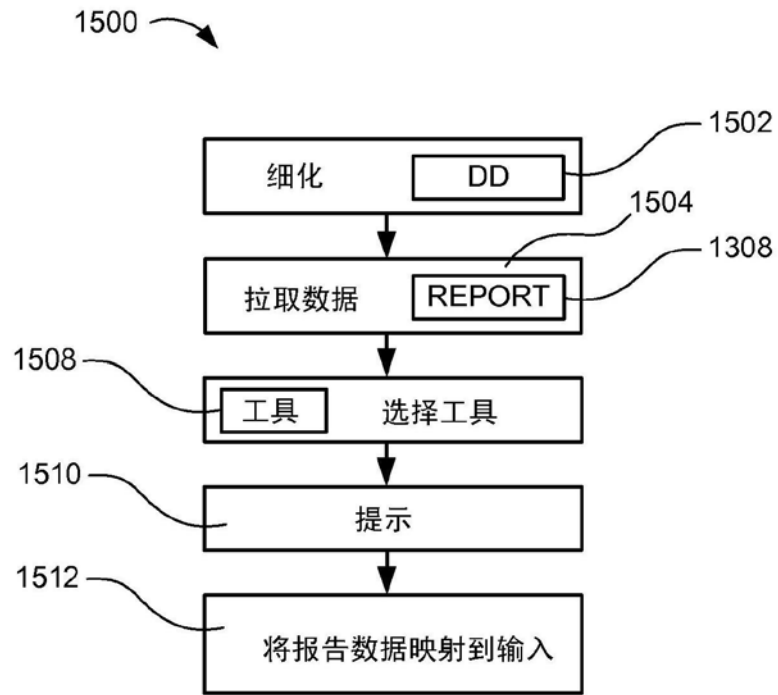


图15