(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 103870480 A
(43)申请公布日 2014.06.18

(21)申请号 201210536320.X

(22)申请日 2012.12.12

(71)申请人 财团法人资讯工业策进会
地址 中国台湾台北市和平东路二段106号11F

(72)发明人 蔡林峻 锺松刚 吴建兴

(74)专利代理机构 北京律诚同业知识产权代理有限公司 11006
代理人 徐金国

(51)Int.Cl.
G06F 17/30(2006.01)

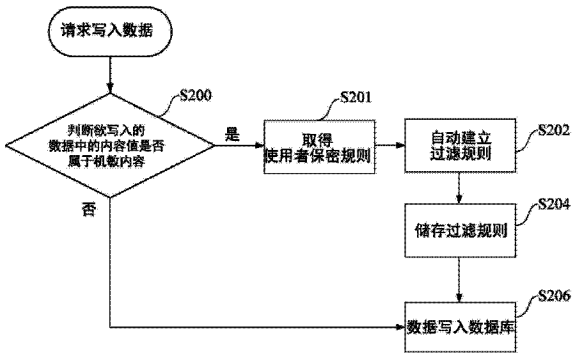
权利要求书2页 说明书6页 附图5页

(54)发明名称

动态数据遮罩方法以及数据库系统

(57)摘要

本发明揭露一种动态数据遮罩方法以及数据库系统。动态数据遮罩方法适用于包含多笔数据的数据库，每一笔数据包含多个内容值以及对应所述多个内容值的多个栏位标签。动态数据遮罩方法包含：当一笔数据欲写入至该数据库时，判断欲写入的该笔数据中的多个内容值及多个栏位标签是否属于机敏内容；若该笔数据的其中一个内容值或栏位标签本身属于机敏内容，将该其中一个内容值所对应的一栏位标签或该其中一个栏位标签本身设为机敏栏位，并动态建立一过滤规则对应该栏位标签；以及，储存该过滤规则，并将该笔数据写入该数据库。



CN 103870480 A

1. 一种动态数据遮罩方法，其特征在于，适用于一数据库其用以保存多笔数据，其中每一笔数据包含多个内容值以及对应所述多个内容值的多个栏位标签，该动态数据遮罩方法包含：

当一笔数据欲写入至该数据库时，判断欲写入的该笔数据中的多个内容值及多个栏位标签是否属于机敏内容；

若该笔数据的其中一个内容值属于机敏内容或者该笔数据的其中一个栏位标签本身属于机敏内容，将该其中一个内容值所对应的一栏位标签或者该其中一个栏位标签本身设为机敏栏位，并动态建立一过滤规则对应该栏位标签；以及

储存该过滤规则，并将该笔数据写入该数据库。

2. 根据权利要求 1 所述的动态数据遮罩方法，其特征在于，于该笔数据写入至该数据库的过程中，该动态数据遮罩方法还包含：

取得一使用者保密规则，其包含多个相异的使用者等级，其中对应该栏位标签动态建立该过滤规则的步骤中，进一步根据该使用者保密规则对应同一栏位标签分别建立多个相异的过滤规则，以对应所述多个相异的使用者等级。

3. 根据权利要求 1 所述的动态数据遮罩方法，其特征在于，还包含：

当请求读取该数据库时，判断请求的一栏位标签是否为机敏栏位；

若请求的该栏位标签属于机敏栏位，载入请求的该栏位标签所对应的该过滤规则；

根据对应的该过滤规则，对请求的该栏位标签对应的该内容值进行遮罩处理；以及

回复经遮罩处理后的该内容值。

4. 根据权利要求 3 所述的动态数据遮罩方法，其特征在于，于请求读取该数据库的过程中，该动态数据遮罩方法还包含：

取得目前请求的一使用者等级，其中载入该过滤规则的步骤中，是同时根据请求的该栏位标签以及该使用者等级进而载入相对应的该过滤规则。

5. 根据权利要求 1 所述的动态数据遮罩方法，其特征在于，该动态数据遮罩方法是根据一演算法或一查表方式判断欲写入的所述多个内容值是否属于机敏内容，其中该演算法是选自正规表示法、机器学习法及签章演算法至少其一。

6. 一种数据库系统，其特征在于，包含：

一数据库，用以保存多笔数据，每一笔数据包含多个内容值以及对应所述多个内容值的多个栏位标签；

一数据处理单元，与该数据库通讯连接，用以处理写入或读取该数据库的请求，

其中，当一笔数据请求写入至该数据库时，该数据处理单元判断欲写入的该笔数据中的多个内容值及多个栏位标签是否属于机敏内容，若该笔数据的其中一个内容值或其中一个栏位标签本身属于机敏内容，该数据处理单元将该其中一个内容值所对应的一栏位标签或该其中一个栏位标签本身设为机敏栏位，并动态建立一过滤规则对应该栏位标签。

7. 根据权利要求 6 所述的数据库系统，其特征在于，当请求读取该数据库时，该数据处理单元判断请求的一栏位标签是否为机敏栏位，若请求的该栏位标签属于机敏栏位，该数据处理单元载入请求的该栏位标签所对应的该过滤规则，根据对应的该过滤规则，该数据处理单元对请求的该栏位标签对应的该内容值进行遮罩处理，并回复经遮罩处理后的该内容值。

8. 根据权利要求 6 所述的数据库系统，其特征在于，该数据处理单元为一网络网关、整合于一网络网关上的一控制电路、或整合于该数据库上的一控制电路。

9. 根据权利要求 6 所述的数据库系统，其特征在于，该数据库为一非关系型数据库、或一关系型数据库。

10. 根据权利要求 6 所述的数据库系统，其特征在于，该数据处理单元储存有一使用者保密规则，其包含多个相异的使用者等级，当该数据处理单元对应该栏位标签动态建立该过滤规则时，该数据处理单元进一步根据该使用者保密规则对应同一栏位标签分别建立多个相异的过滤规则，以对应所述多个相异的使用者等级，当该数据处理单元读取该数据库时，该数据处理单元判断目前请求的该使用者等级，该数据处理单元是同时根据请求的该栏位标签以及该使用者等级进而载入相对应的该过滤规则。

动态数据遮罩方法以及数据库系统

技术领域

[0001] 本发明是有关于一种数据处理方法，且特别是有关于一种可用以保护机敏内容的数据处理方法及其数据库系统。

背景技术

[0002] 近年来云端网络快速发展，许多的重要信息（如个人的身份资料、帐单、信件、公司的商业文件、政府公文等）皆存放在各种云端数据库中，使用者可透过网络方便且快速地存取数据库中的各种信息。

[0003] 传统的数据库架构，如关联性数据库管理系统(Relational Database Management System,RDBMS)及基于结构化查询语言(Structured Query Language,SQL)的关联性数据库，已无法负荷云端时代来临所带来的大量数据储存需求。因此，近年来非关联性数据库（如NoSQL）架构兴起，实际例子如Google BigTable、Facebook Cassandra、Yahoo Hbase、Amazon DynamoDB等数据库。

[0004] 传统的关联性数据库中具有事先设定好的栏位以及各栏位的内容值，因应不同的需求或使用者数据需重新设计适当的栏位标签，以及栏位标签与内容值的对应关系。

[0005] 非关联性数据库则相对具有较高的即时性，每一笔数据可各自具有多个内容值及相对应的多个栏位标签。因此，非关联性的数据库架构（如NoSQL）比传统的关联性数据库管理系统适合处理当前非结构化的大量云端数据存取。

[0006] 现今云端数据库处理到重要的敏感性数据（如个人的身份证号码、电话号码、通讯地址等）时，需要有不同程度的遮罩(masking)处理，例如将电话号码由0921345678处理为09xxxxx678，借此保障使用者的机敏内容。

[0007] 目前常见的数据遮罩技术包含静态数据遮罩技术(Static Data Masking)以及动态数据遮罩技术(Dynamic Data Masking)。

[0008] 其中，静态数据遮罩技术针对关联式数据库进行机敏数据遮罩，将遮罩后的数据内容存入去识别化的数据库供所有使用者使用。然而，静态数据遮罩技术所产生的去识别化数据库，数据库中的已遮罩数据无法即时更新，且无法依使用者身份分别提供不同的遮罩方式，应用范围有限。

[0009] 其中，动态数据遮罩技术(Dynamic Data Masking)可依据使用者身份分别提供即时机敏数据去识别化。目前一般的动态数据遮罩技术是透过拦截结构化查询语言(SQL)指令，并修改答复封包（将答复封包加上遮罩），来达到保护机敏数据。

[0010] 既有动态数据遮罩技术必须知道目标数据库的栏位是否属于机敏栏位（需要管理者预先设置），然而，非关联性数据库（如NoSQL）所具有的栏位将随着新建数据写入时而动态变化。随着非关联性数据库中的数据增加，栏位的总体数目将相对应地增加。由于非关联性数据库的特性，管理人员无法有效定义相关栏位属性及过滤规则。因此，传统做法中预先设置机敏栏位并且拦截修改SQL指令保护来机敏数据的方式，无法应用至新的非关联性数据库上。

[0011] 此外，传统的动态数据遮罩技术仅在使用者读取数据时拦截查询指令并修改答复封包，并不会在数据写入时进行分析判断。由于数据储存与读取之间并没有自动建立关联性，因此造成管理人员必须自行定义相关栏位属性及过滤规则，容易造成机敏数据外泄。

发明内容

[0012] 为解决上述问题，本发明提出一种动态数据遮罩方法及其数据库系统，其中本发明的方法在数据写入数据库的写入阶段中，对储存至数据库的内容值 (value) 进行扫描，并根据内容值动态建立过滤规则。在读取阶段中，便基于先前动态建立的过滤规则即时进行数据遮罩。于本发明中的过滤规则是在写入阶段中依据内容值是否符合机敏特性而自动判断产生，管理人员不需自行定义机敏栏位或过滤规则，因此同时适用于新式的非关联性数据库与传统的关联性数据库。此外，本发明可进一步依据使用者身份等级不同，提供不同的机敏数据查询结果。

[0013] 本发明的一方面是在提供一种动态数据遮罩方法，适用于包含多笔数据的数据库，每一笔数据包含多个内容值以及对应所述多个内容值的多个栏位标签。动态数据遮罩方法包含：当一笔数据欲写入至该数据库时，判断欲写入的该笔数据中的多个内容值及多个栏位标签是否属于机敏内容；若该笔数据的其中一个内容值属于机敏内容，将该其中一个内容值所对应的一栏位标签设为机敏栏位，并动态建立一过滤规则对应该栏位标签；或者，若该笔数据的其中一个栏位标签本身属于机敏内容，将该其中一个栏位标签设为机敏栏位，并动态建立一过滤规则对应该其中一个栏位标签；以及，储存该过滤规则，并将该笔数据写入该数据库。

[0014] 本发明的另一方面是在提供一种数据库系统，其包含数据库以及数据处理单元。数据库包含多笔数据，每一笔数据包含多个内容值以及对应所述多个内容值的多个栏位标签。数据处理单元与数据库通讯连接，用以处理写入或读取该数据库的请求。其中当一笔数据请求写入至该数据库时，该数据处理单元判断欲写入的该笔数据中的多个内容值是否属于机敏内容，若该笔数据的其中一个内容值属于机敏内容，该数据处理单元将该其中一个内容值所对应的一栏位标签设为机敏栏位，并动态建立一过滤规则对应该栏位标签。

附图说明

[0015] 为让本发明的上述和其他目的、特征、优点与实施例能更明显易懂，所附附图的说明如下：

[0016] 图 1 绘示根据本发明的一实施例中一种数据库系统的示意图；

[0017] 图 2 绘示根据本发明的一实施例中一种动态数据遮罩方法在写入阶段的流程图；

[0018] 图 3 绘示根据本发明的一实施例中动态数据遮罩方法在读取阶段的流程图；

[0019] 图 4 绘示根据本发明的另一实施例中一种动态数据遮罩方法在写入阶段的流程图；以及

[0020] 图 5 绘示根据本发明的另一实施例中动态数据遮罩方法在读取阶段的流程图。

具体实施方式

[0021] 请参阅图 1，其绘示根据本发明的一实施例中一种数据库系统 100 的示意图。如图

1 所示，数据库系统 100 包含数据库 120 以及数据处理单元 140。数据库 120 可用以保存多笔数据，每一笔数据包含多个内容值以及对应所述多个内容值的多个栏位标签。数据处理单元 140 与数据库 120 通讯连接，数据处理单元 140 用以处理写入或读取数据库 120 的请求。于此实施例中，数据库系统 100 可进一步包含与数据处理单元 140 通讯连接的过滤规则数据库 160，但本发明并不以此为限。

[0022] 于此实施例中，数据处理单元 140 可为网络网关，使用者终端 180 可经由网络网关（数据处理单元 140）进而写入或读取数据库 120 中的内容。须补充的是，使用者终端 180 并不局限于特定使用者，它可能是任何的数据来源，例如，有可能数据库系统 100 的拥有者同时也是所谓的“使用者”，因此“使用者终端”这个词的定义不仅限于数据库系统 100 的数据来源，举例来说，亦可为欲读取数据库系统 100 的请求者，欲修改 / 控制数据库系统 100 的管理者等。

[0023] 本发明并不以限于数据处理单元 140 为网络网关，于其他实施例中，数据处理单元 140 亦可为整合于网络网关上的控制电路，或是整合于数据库 120 上的控制电路。此外，本发明中的数据库 120 可为非关系型数据库（如 NoSQL）或是关系型数据库。

[0024] 于此实施例中，数据库系统 100 在数据写入与读取的过程中可执行动态数据遮罩方法，借此来保护机敏内容的安全性。此动态数据遮罩方法的详细做法请一并参考图 2 及图 3，图 2 绘示根据本发明的一实施例中一种动态数据遮罩方法在写入阶段的流程图，图 3 绘示根据本发明的一实施例中动态数据遮罩方法在读取阶段的流程图。

[0025] 如图 1 与图 2 所示，假设使用者终端 180 请求将一笔数据写入至数据库 120。此时，数据处理单元 140 执行步骤 S200 判断欲写入的该笔数据中的多个内容值及多个栏位标签是否属于机敏内容。实际应用中，数据处理单元 140 可根据演算法判断欲写入的内容值 / 栏位标签当中是否属于机敏内容。实际应用中，判断机敏内容的演算法可采用正规表示法 (Regular Expression, regex)、机器学习法 (Machine Learning) 及签章演算法 (Signature) 等演算法中至少其一。

[0026] 或是，于另一实施例中，数据处理单元 140 亦可利用查表方式判断欲写入的内容值当中是否存在有属于机敏内容的内容值，于此例中，数据处理单元 140 须建有常见机敏内容的表格，如姓氏、地址格式或特定关键字。

[0027] 若步骤 S200 中判断该笔欲写入的数据中的多个内容值及多个栏位标签是否属于机敏内容，数据处理单元 140 便执行步骤 S202，自动建立过滤规则。其中若该笔数据的其中一个内容值属于机敏内容，步骤 S202 是将此一内容值所对应的栏位标签设为机敏栏位，并动态建立对应此栏位标签的过滤规则；另一方面，若该笔数据的其中一个栏位标签本身属于机敏内容，步骤 S202 是将该其中一个栏位标签设为机敏栏位，并动态建立一过滤规则对应该其中一个栏位标签。

[0028] 假设，欲写入的数据如下表一：

[0029]

栏位标签		内容值
user001	email	abc123@gmail.com
user001	passport_num	3456789012
user001	text	大家好

[0030] 表一

[0031] 如上表一所举的例子，当欲写入的数据其中一个内容值为 abc123@gmail.com。此时步骤 S200 判断此一内容值涉及机敏内容，步骤 S202 便可将对应的栏位标签 user001.email 设定为机敏栏位，并动态建立对应此一栏位标签 user001.email 的过滤规则。例如，过滤规则可为将内容值的字串第一字符至第三字符以其他字符（例如 * 字符）替代。储存的过滤规则以程序语言表示的例子可为 :MaskRule(substr(user001.email,1,3)||`***’)。

[0032] 又或者，如上表一所举的例子，当欲写入的数据其中一个栏位标签本身为护照号码 (passport_num)。此时步骤 S200 判断此一栏位标签本身涉及机敏内容，步骤 S202 便可将对应的栏位标签 user001.passport_num 设定为机敏栏位，并动态建立对应此一栏位标签 user001.passport_num 的过滤规则。

[0033] 另一方面，若步骤 S200 判断不存在属于机敏内容的内容值，便执行步骤 S206 将数据写入数据库 120。步骤 S200 可判断内容值“大家好”未涉及机敏内容，便不需对栏位标签 user001.text 产生过滤规则。

[0034] 此时，数据处理单元 140 便执行步骤 S204 将对应此一栏位标签 (user001.email) 的过滤规则存入过滤规则数据库 160。当过滤规则自动产生之后，数据处理单元 140 便执行步骤 S206，将使用者终端 180 欲建立的数据写入数据库 120。须补充的是，数据库 120 中所储存的是未经遮罩处理的完整数据。

[0035] 此外，过滤规则数据库 160 可为独立于数据库 120 外的另一单独数据库，但本发明并不以此为限。于另一实施例中，过滤规则数据库 160 亦可整合于数据库 120 中，数据处理单元 140 可将写入数据与过滤规则分别存在数据库 120 中的不同记忆空间。

[0036] 另一方面，须补充的是，本实施例中将数据写入数据库的步骤 (S206) 与产生并储存过滤规则的步骤 (S202 及 S204) 并不限于特定的先后次序关系，实际应用中，将数据写入数据库的步骤 S206 与产生并储存过滤规则的步骤 S202 及 S204 先后次序可互换，或亦可平行处理。

[0037] 本实施例的动态数据遮罩方法及其数据库系统在上述数据写入的阶段，便动态地根据写入数据中的内容值选择性产生过滤规则，并可将原始写入数据存入数据库中。相较既有静态数据遮罩技术，本实施例可保留完整的写入数据内容在数据库中。相较既有动态数据遮罩技术，本实施例在数据写入的阶段便分析数据内容并自动产生过滤规则。

[0038] 接着，如图 1 与图 3 所示，假设使用者终端 180 请求读取数据库 120 中的一笔数据（包含指定的至少一个栏位标签）或某一栏位标签的多笔数据。此时，数据处理单元 140 执行步骤 S300 判断请求的栏位标签是否为机敏栏位。

[0039] 若步骤 S300 判断请求的该栏位标签属于机敏栏位，数据处理单元 140 便执行步骤

S302 载入请求的栏位标签所对应的该过滤规则。

[0040] 随后，执行步骤 S304，数据处理单元 140 从数据库 120 中读出使用者终端 180 请求读取的数据内容（数据库中保存完整数据内容），并且数据处理单元 140 根据对应的过滤规则对请求的栏位标签对应的内容值进行遮罩处理。举例来说，若使用者终端 180 请求的栏位标签为 user001.email，此时可载入过滤规则，例如将第一字符至第三字符以 * 字符替代。

[0041] 随后，数据处理单元 140 执行步骤 S306 将对应请求的栏位标签且经遮罩处理后（如步骤 S304）的内容值回复至使用者终端 180。此例子中，回复给使用者终端 180 的内容值便为经遮罩处理的样式，如“***123@gmail.com”。借此，达到机敏数据的保护效果。

[0042] 另一方面，若步骤 S300 判断请求的栏位标签为非机敏栏位，则可直接进行步骤 S306 将对应请求的栏位标签的内容值回复至使用者终端 180。

[0043] 此外，本发明的动态数据遮罩方法及数据库系统 100 可进一步根据不同的使用者等级，产生不同的机敏数据过滤结果。请一并参阅图 4 及图 5，图 4 绘示根据本发明的另一实施例中一种动态数据遮罩方法在写入阶段的流程图，图 5 绘示根据本发明的另一实施例中动态数据遮罩方法在读取阶段的流程图。

[0044] 在图 4 及图 5 的实施例中，动态数据遮罩方法可进一步根据不同的使用者等级产生不同的机敏数据过滤结果。

[0045] 在数据写入阶段，请一并参阅图 1 及图 4，并可对照图 2，于图 4 的实施例中相较图 2 进一步包含步骤 S201，取得一使用者保密规则。于此实施例中，使用者保密规则可储存在数据处理单元 140。使用者保密规则中包含多个相异的使用者等级，例如 :访客、内部员工、系统管理员等使用者等级。

[0046] 于图 4 的实施例中，当数据处理单元 140 执行步骤 S202 对应栏位标签动态建立过滤规则时，数据处理单元 140 进一步根据使用者保密规则对应同一栏位标签分别建立多个相异的过滤规则，以对应上述相异的使用者等级。

[0047] 例如，对应同一栏位标签 user001.email 的过滤规则，例如，访客等级的过滤规则可为将内容值的字串全部字符以 * 字符替代，内部员工等级的过滤规则可为将内容值的字串第一至第三字符以 * 字符替代，系统管理员等级的过滤规则可为不替代任何字串。

[0048] 也就是说，对应同一栏位标签 user001.email 根据多个使用者等级建立三个独立的过滤规则，其过滤规则彼此间可为相同规则或相异规则。

[0049] 另一方面，在数据读取阶段，请一并参阅图 1 及图 5，并可对照图 3，于图 5 的实施例中相较图 3 进一步包含步骤 S301，取得目前使用者终端 180 上的使用者等级。

[0050] 随后，在载入过滤规则的步骤 S302 中，数据处理单元 140 是同时根据请求的栏位标签以及使用者等级进而载入相对应的过滤规则。

[0051] 也就是说，针对栏位标签 user001.email 的读取请求，访客等级所看到经遮罩处理的回复内容值可为“**********”，内部员工等级所看到经遮罩处理的回复内容值可为“***123@gmail.com”，系统管理员等级所看到经遮罩处理的回复内容值可为“abc123@gmail.com”。借此，达到对应不同使用者有高弹性的数据库存取操作。

[0052] 综上所述，本发明提出一种动态数据遮罩方法及其数据库系统，其中本发明的方法在数据写入数据库的写入阶段中，对储存至数据库的内容值及栏位标签进行扫描，并根

据内容值及栏位标签动态建立过滤规则。在读取阶段中，便基于先前动态建立的过滤规则即时进行数据遮罩。于本发明中的过滤规则是在写入阶段中依据内容值及栏位标签是否符合机敏特性而自动判断产生，管理人员不需自行定义机敏栏位或过滤规则，因此同时适用于新式的非关联性数据库与传统的关联性数据库。此外，本发明可进一步依据使用者身份等级不同，提供不同的机敏数据查询结果。

[0053] 虽然本发明已以实施方式揭露如上，然其并非用以限定本发明，任何熟悉此技艺者，在不脱离本发明的精神和范围内，当可作各种的更动与润饰，因此本发明的保护范围当视所附的权利要求书所界定的范围为准。

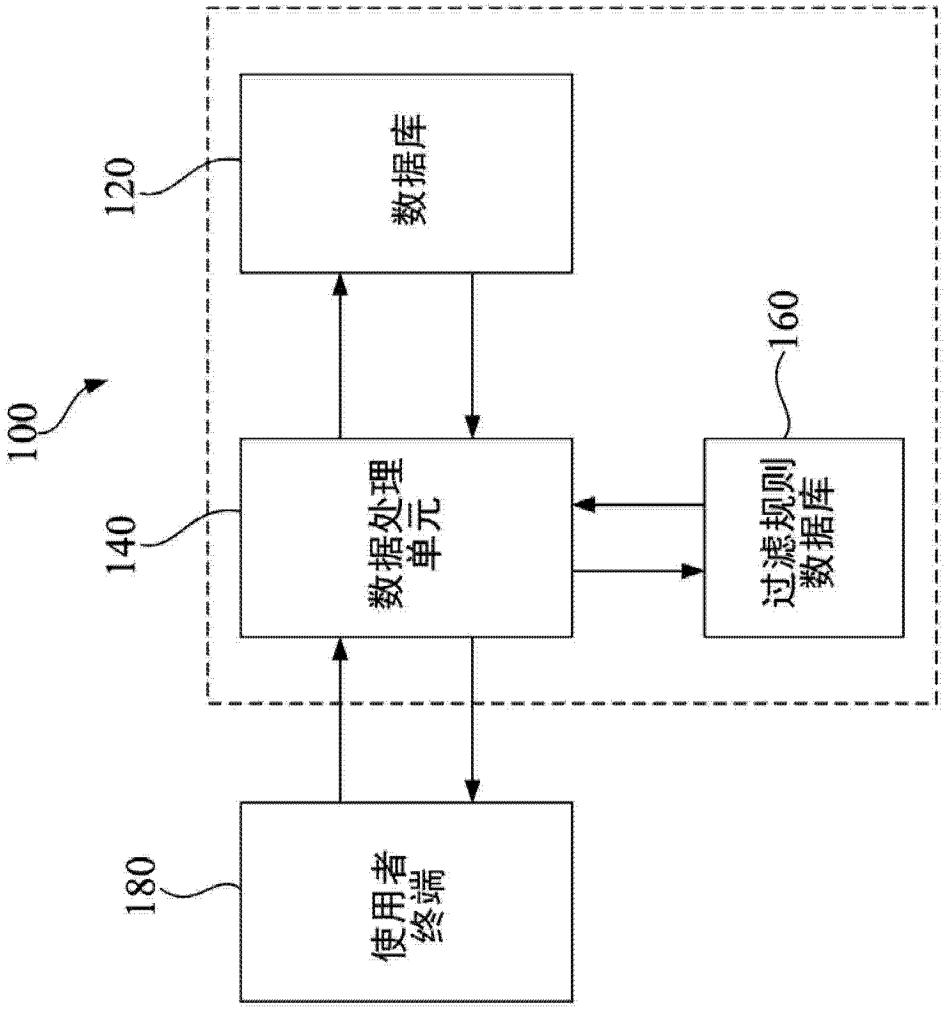


图 1

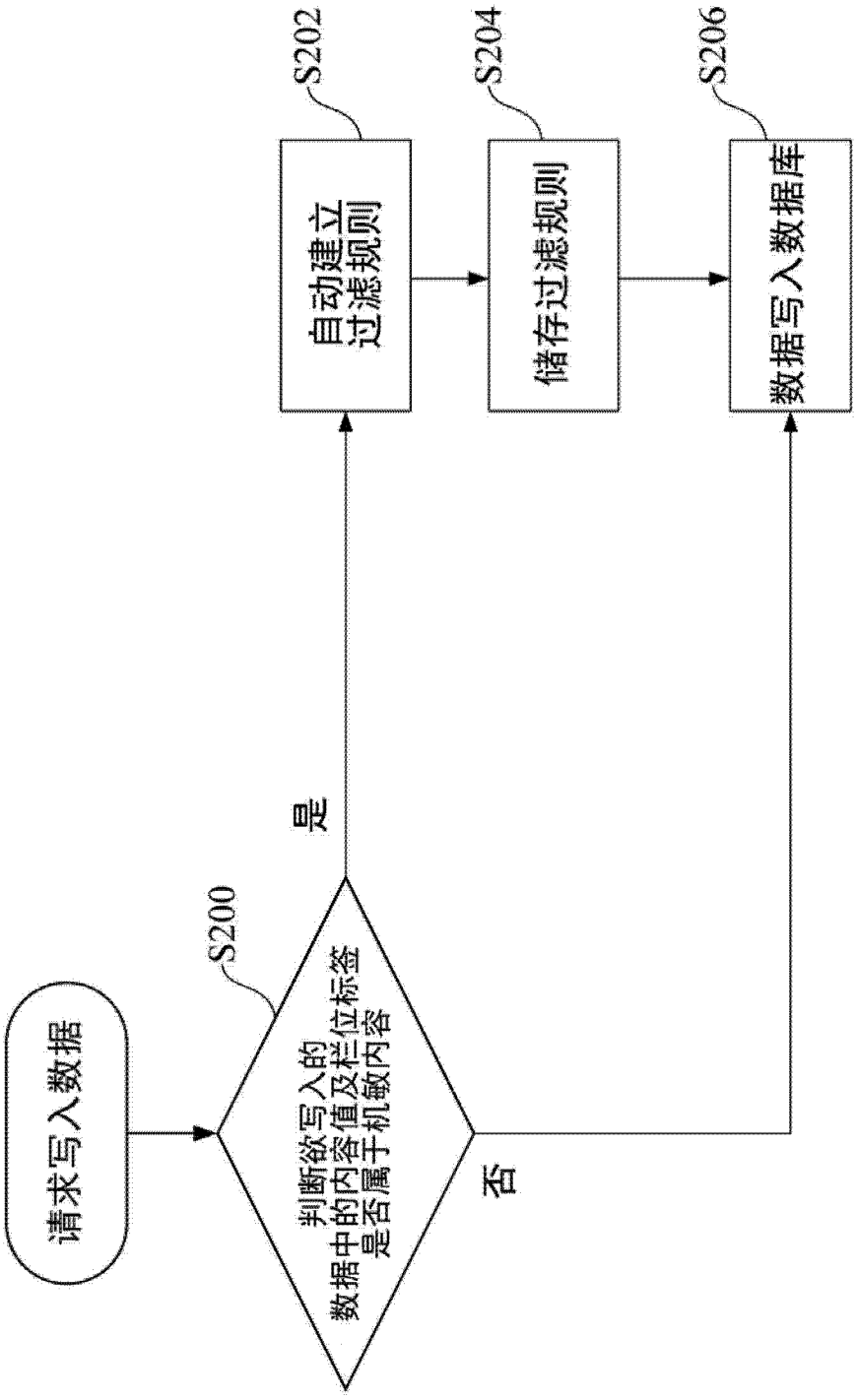

请求写入数据
S200
判断欲写入的
数据中的内容值及栏位标签
是否属于机敏内容
是
否
S202
自动建立
过滤规则
S204
储存过滤规则
S206
数据写入数据库

图 2

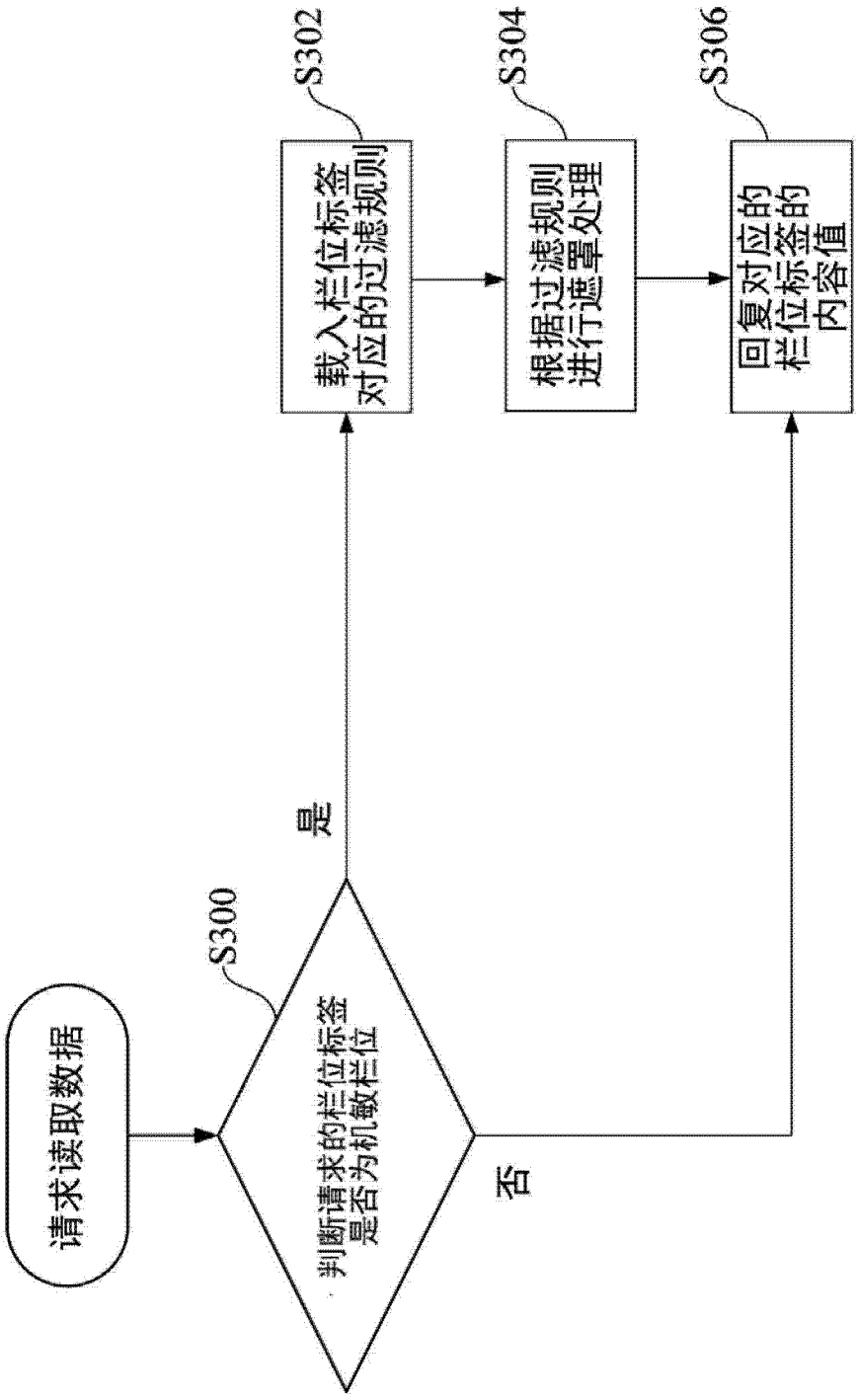

请求读取数据
S300
判断请求的栏位标签是否为机敏栏位
是
否
S302
载入栏位标签对应的过滤规则
S304
根据过滤规则进行遮罩处理
S306
回复对应的栏位标签的内容值

图 3

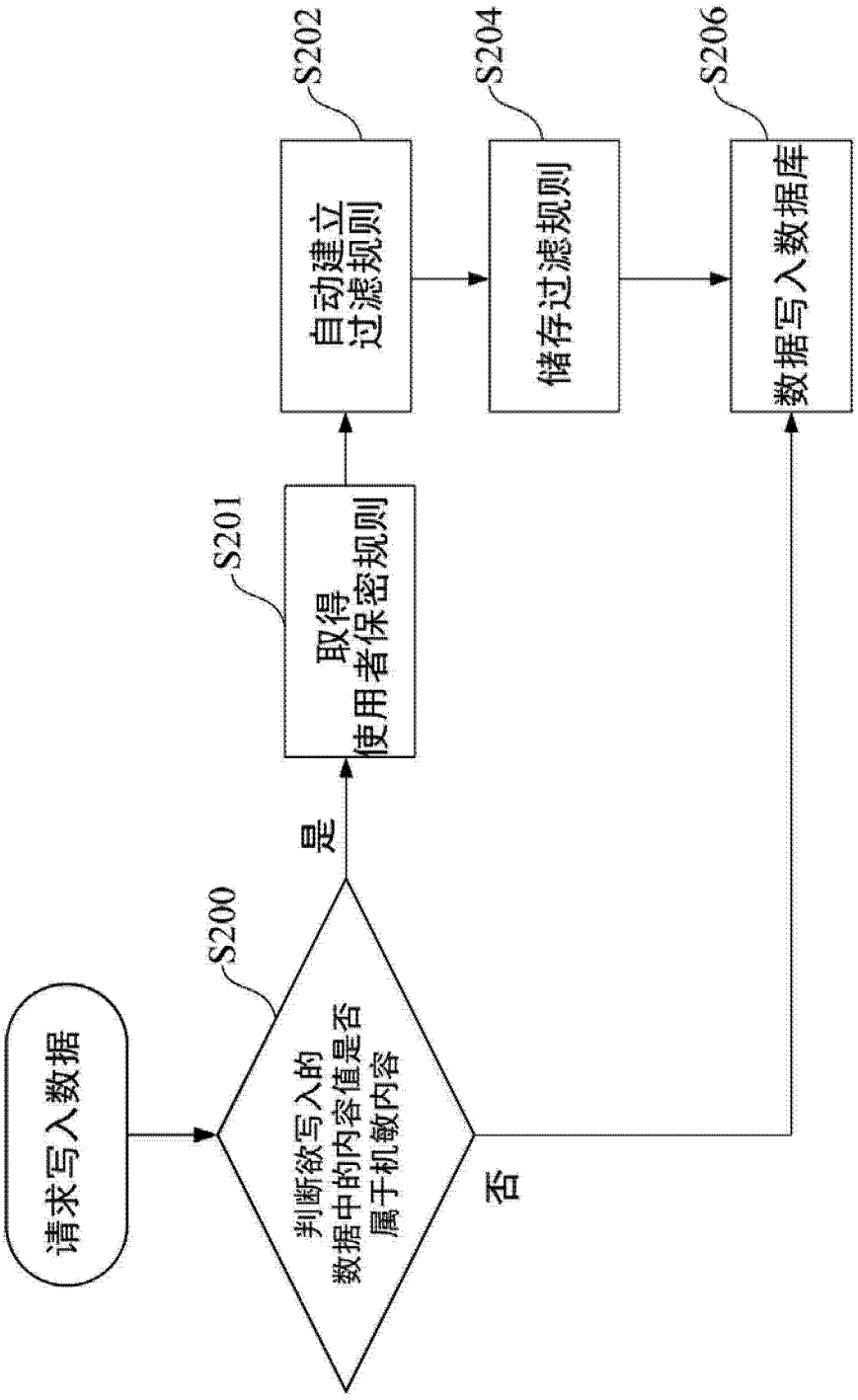

请求写入数据
S200
判断欲写入的
数据中的内容值是否
属于机敏内容
是
否
S201
取得
使用者保密规则
S202
自动建立
过滤规则
S204
储存过滤规则
S206
数据写入数据库

图 4

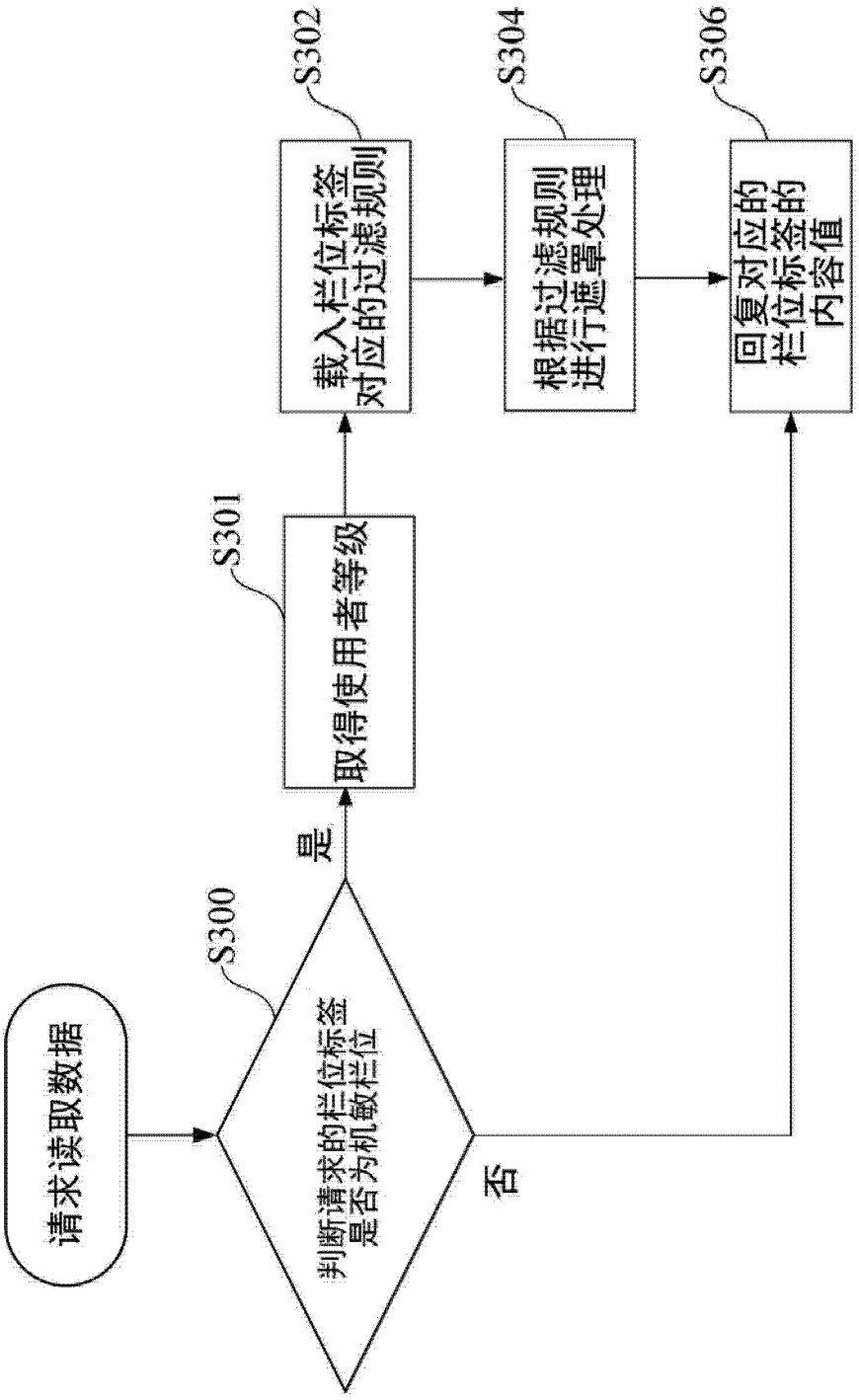

请求读取数据
S300
判断请求的栏位标签是否为机敏栏位
是
否
S301
取得使用者等级
S302
载入栏位标签对应的过滤规则
S304
根据过滤规则进行遮罩处理
S306
回复对应的栏位标签的内容值

图 5