



# [12] 发明专利说明书

专利号 ZL 200510114011.3

[45] 授权公告日 2008 年 12 月 3 日

[11] 授权公告号 CN 100440212C

[22] 申请日 2005.10.13

[21] 申请号 200510114011.3

[30] 优先权

[32] 2004.10.14 [33] US [31] 10/965,184

[73] 专利权人 国际商业机器公司

地址 美国纽约

[72] 发明人 埃里克·L·巴斯尼斯

约翰·M·桑托斯沃索

[56] 参考文献

CN1477560A 2004.2.25

US20030093407A1 2003.5.15

US5926807A 1999.7.20

CN1341901A 2002.3.27

审查员 陈颖

[74] 专利代理机构 北京市柳沈律师事务所

代理人 邸万奎 黄小临

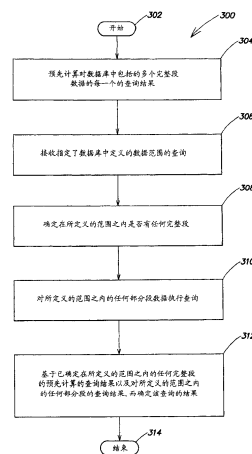
权利要求书 4 页 说明书 7 页 附图 4 页

## [54] 发明名称

用于处理数据库查询的方法和设备

## [57] 摘要

在第一方面，提供了一种方法，包括以下步骤：(1) 预先计算对数据库中包含的多个完整段数据的每一个的查询结果；(2) 接收指定了数据库中定义的数据范围的查询；(3) 确定在所定义的范围之内是否有任何完整段；(4) 对所定义的范围之内的任何部分段数据执行查询；以及(5) 基于已确定在所定义的范围之内的任何完整段的预先计算的查询结果以及对所定义的范围之内的任何部分段的查询结果，而确定该查询的结果。还提供了许多其他方面。



1. 一种方法，包括：

预先计算对数据库中包括的多个完整段数据的每一个的查询结果；

接收指定了数据库中定义的数据范围的查询；

确定在所定义的范围之内是否有任何完整段；

对所定义的范围之内的任何部分段数据执行查询；以及

基于已确定在所定义的范围之内的任何完整段的预先计算的查询结果以及对所定义的范围之内的任何部分段的查询结果，而确定该查询的结果。

2. 根据权利要求1的方法，其中基于已确定在所定义的范围之内的任何完整段的预先计算的查询结果以及对所定义的范围之内的任何部分段的查询结果而确定该查询的结果的步骤包括：通过检索该预先计算的查询结果，并确定对所定义的范围内的任何部分段数据的查询结果，而确定该查询的结果。

3. 根据权利要求1的方法，还包括：

在数据库中存储预先计算的查询结果；以及

从数据库中去掉对大于指定年限的日期范围的数据执行的预先计算的查询结果。

4. 根据权利要求1的方法，其中数据库中的数据基于与该数据关联的日期而被分组为一个或多个完整段和部分段。

5. 根据权利要求1的方法，其中预先计算对数据库中包括的多个完整段数据的每一个的查询结果的步骤包括：响应于接收到指定了数据库中定义的数据范围的查询，而预先计算对数据库中包括的多个完整段数据的每一个的查询结果。

6. 一种用于处理对信息的数据库查询的方法，包括：

接收对数据库中的第一部分数据执行查询的请求；

确定对数据库的第一部分的查询结果基于至少一个预先计算结果，其中该预先计算结果是对数据库中包括在第一部分中的第二部分数据的查询结果；

对数据库中的第三部分数据执行查询，其中该第三部分是第一部分中除了第二部分之外的部分；以及

基于该预先计算结果和对第三部分的查询结果而计算对第一部分的查询

结果。

7. 根据权利要求6的方法，其中基于该预先计算结果和对第三部分的查询结果而计算对第一部分的查询结果的步骤包括：合并该预先计算结果和对第三部分数据的查询结果。

8. 根据权利要求6的方法，其中该第二部分数据多于该第三部分数据。

9. 根据权利要求6的方法，其中基于该预先计算结果和对第三部分的查询结果而计算对第一部分的查询结果的步骤包括：通过检索该预先计算结果，并计算对第三部分的查询结果，而计算对第一部分的查询结果。

10. 根据权利要求6的方法，还包括：

在数据库中存储预先计算的查询结果；以及

从数据库中去掉对大于指定年限的日期范围的数据执行的预先计算的查询结果。

11. 根据权利要求6的方法，其中数据库中的数据基于与该数据关联的日期而被分组为第一、第二和第三部分数据。

12. 根据权利要求11的方法，其中日、月和年中的至少一个与该第二部分数据关联。

13. 根据权利要求6的方法，其中响应于接收到对数据库中的第一部分数据执行查询的请求，而计算该预先计算结果。

14. 根据权利要求6的方法，还包括：

监控该预先计算结果；以及

基于监控而确定是否继续存储该预先计算结果。

15. 一种设备，包括：

用于预先计算对数据库中包括的多个完整段数据的每一个的查询结果的单元；

用于接收指定了数据库中定义的数据范围的查询的单元；

用于确定在该定义的范围之内是否有任何完整段的单元；

用于对所定义的范围之内的任何部分段数据执行查询的单元；以及

用于基于已确定在所定义的范围之内的任何完整段的预先计算的查询结果以及对所定义的范围之内的任何部分段的查询结果，而确定该查询的结果的单元。

16. 根据权利要求15的设备，其中用于基于已确定在所定义的范围之内

的任何完整段的预先计算的查询结果以及对所定义的范围之内的任何部分段的查询结果而确定该查询的结果的单元包括：用于通过检索该预先计算的查询结果，并确定对所定义的范围之内的任何部分段数据的查询结果，而确定该查询的结果的单元。

17. 根据权利要求 15 的设备，其中还包括：

用于在数据库中存储该预先计算的查询结果的单元；以及

用于从数据库中去除对大于指定年限的日期范围的数据执行的预先计算的查询结果的单元。

18. 根据权利要求 15 的设备，其中数据库中的数据基于与该数据关联的日期而被分组为一个或多个完整段和部分段。

19. 根据权利要求 15 的设备，其中用于预先计算对数据库中包括的多个完整段数据的每一个的查询结果的单元包括：用于响应于接收到指定了数据库中定义的数据范围的查询，而预先计算对数据库中包括的多个完整段数据的每一个的查询结果的单元。

20. 一种用于处理数据库查询的设备，包括：

用于接收对数据库中的第一部分数据执行查询的请求的单元；

用于确定对数据库的第一部分的查询结果基于预先计算结果的单元，其中该预先计算结果是对数据库中包括在第一部分中的第二部分数据的查询结果；

用于对数据库中的第三部分数据执行查询的单元，其中该第三部分是第一部分中除了第二部分之外的部分；以及

用于基于该预先计算结果和对第三部分的查询结果而计算对第一部分的查询结果的单元。

21. 根据权利要求 20 的设备，其中用于基于该预先计算结果和对第三部分的查询结果而计算对第一部分的查询结果的单元包括：用于合并该预先计算结果和对第三部分数据的查询结果的单元。

22. 根据权利要求 20 的设备，其中该第二部分数据多于该第三部分数据。

23. 根据权利要求 20 的设备，其中用于基于该预先计算结果和对第三部分的查询结果而计算对第一部分的查询结果的单元包括：用于通过检索该预先计算结果，并计算对第三部分的查询结果，而计算对第一部分的查询结果的单元。

24. 根据权利要求 20 的设备, 其中还包括:  
用于在数据库中存储该预先计算的查询结果的单元; 以及  
用于从数据库中去掉对大于指定年限的日期范围的数据执行的预先计算的查询结果的单元。

25. 根据权利要求 20 的设备, 其中数据库中的数据基于与该数据关联的日期而被分组为第一、第二和第三部分数据。

26. 根据权利要求 25 的设备, 其中日、月和年中的至少一个与该第二部分数据关联。

27. 根据权利要求 20 的设备, 其中还包括: 用于响应于对数据库中的第一部分数据执行查询的请求, 而计算该预先计算结果的单元。

28. 根据权利要求 20 的设备, 其中还包括:  
用于监控该预先计算结果的单元; 以及  
用于基于监控而确定是否继续存储该预先计算结果的单元。

## 用于处理数据库查询的方法和设备

### 技术领域

本发明一般涉及计算机系统，并更具体地涉及用于处理数据库查询的方法和设备。

### 背景技术

数据库频繁地接收到对数据库中某数据范围的诸如总和、平均值等信息的查询。传统上，响应于这类查询，数据库访问该数据范围内的每条记录以处理该查询，其中每一记录包括一个或多个数据项。然而，这种处理数据库查询的方法导致大处理开销。

因此，期望用于数据库查询的改善处理的方法和设备。

### 发明内容

根据本发明的第一方面，提供了第一方法。该第一方法包括以下步骤：

(1)预先计算对数据库中包括的多个完整段数据的每一个的查询结果；(2)接收指定了数据库中定义的数据范围的查询；(3)确定在所定义的范围之内是否有任何完整段；(4)对所定义的范围之内的任何部分段数据执行查询；以及(5)基于已确定在所定义的范围之内的任何完整段的预先计算的查询结果以及对所定义的范围之内的任何部分段的查询结果，而确定该查询的结果。

根据本发明的第二方面，提供了一种用于处理对信息的数据库查询的方法。该方法包括以下步骤：(1)接收对数据库中的第一部分数据执行查询的请求；(2)确定对数据库的第一部分的查询结果基于至少一个预先计算结果，其中该预先计算结果是对数据库中包括在第一部分中的第二部分数据的查询结果；(3)对数据库中的第三部分数据执行查询，其中该第三部分是第一部分中除了第二部分之外的部分；以及(4)基于该预先计算结果和对第三部分的查询结果而计算对第一部分的查询结果。还提供了许多其他方面，如根据本发明这些其他方面的方法、系统、设备和计算机程序产品。这里描述的每一计算机程序产品可由计算机可读介质(例如载波信号、软盘、压缩盘、DVD、硬盘

驱动器、随机存取存储器等)承载。

根据以下详细描述、所附权利要求以及附图，本发明的其他特征和方面将变得更加清楚。

#### 附图说明

图 1 是根据本发明的实施例的用于处理数据库查询的设备的方框图。

图 2 是可在根据本发明的实施例的用于处理数据库查询的设备中使用的数据库的例子的方框图。

图 3 是根据本发明的实施例的用于处理数据库查询的第一示例方法。

图 4 是根据本发明的实施例的用于处理数据库查询的第二示例方法。

#### 具体实施方式

本发明提供了用于有效处理数据库查询的方法和设备。更具体地，本方法和设备避免了不得不访问由数据库查询指定的范围的大部分(例如整个范围)。结果，与传统数据库查询处理相比，本方法和设备提供了用于处理数据库查询的性能增益。

图 1 是根据本发明的实施例的用于处理数据库查询的设备的方框图。参考图 1，该用于处理数据库查询的设备 100 包括与处理器 104 耦接并用于存储数据的存储器 102。该处理器 104 与包括有数据库 108 的例如盘驱动器等的存储装置 106 耦接。下面参考图 2 描述数据库 108 的细节。处理器 104 适配为运行适配为有效查询数据库 108 的代码。更具体地，处理器 104 适配为运行适配为执行根据本发明的实施例的第一和/或第二示例方法的代码。下面将分别参考图 3 和 4 来描述第一和第二示例方法的细节。另外，设备 100 可包括用于与其他设备进行通信的通信(例如 com)端口 110。

图 2 是可在根据本发明的实施例的用于处理数据库查询的设备中使用的数据库的例子的方框图。参考图 2，数据库 108 可存储与例如特定消费者或应用相关的数据。这样的数据可以以用户(如数据库管理员)指定的方式进行组织。例如，数据库 108 可包括代表日期的第一列 202 数据。更具体地，第一列 202 包括从 2000 年第一季度到 2004 年第三季度中每一季度的数据项(例如在与列对应的行上)。数据库 108 包括代表收入的第二列 204 数据。例如，第二列 204 的数据包括与第一列 202 中的每一季度对应的代表收入的数据项。

而且,数据库 108 包括基于数据库 108 中的其他数据的(例如通过从中计算出的)数据。更具体地,数据库 108 可包括第三列 206 数据,其每一项代表一年的总收入(例如这一年每一季度的收入数据之和)。类似地,数据库 108 可包括第四列 208 数据,其每一项代表一年的平均季度收入(例如这一年各季度的收入数据的平均值)。例如,第三列 206 和/或第四列 208 的每一数据项基于第二列 204 中与第一列 202 的特定一年的各季度对应的数据项。如下面所述,数据库 108 中的数据可被分组在一起以形成完整段 210 - 216 中的一个或更多(例如多个)。尽管图 2 的数据库 108 中的数据包括四个完整段 210 - 216,但是数据库 108 中的数据可包括更大或更小数目的完整段。而且,完整段可包括比图 2 所示的完整段 210 - 216 更大或更小数量的数据。

上述数据库 108 的结构(例如行和列中的数据安排)是示例性的。该数据库 108 可以以不同的方式构成。例如,数据库 108 可为关系数据库、平面文件(flat-file)数据库、对象数据库、分布式数据库等。而且,对填充数据库 108 的数据的描述也是示例性的。该数据库 108 可包括不同的数据。

现在参考图 2 和图 3 来描述用于处理数据库查询的设备 100 的操作,图 3 中图示了根据本发明的实施例的用于处理数据库查询的第一示例方法 300。参考图 3,在步骤 302,方法 300 开始。在步骤 304,预先计算数据库中包括的多个完整段数据的每一个的查询结果。如上所述,数据库 108 可被分组为一个或更多(例如多个)完整段。例如,数据库 108 的第二列 204 中与第一列 202 的特定一年(例如 2000、2001、2002 或 2003)的各季度对应的数据项可被分组为各完整段 210 - 216(虚框所示)。所以,示例数据库 108 包括四个完整段。频繁接收到对数据库(例如对数据库中某数据范围)的诸如总和、平均值等的信息的查询。传统上,响应于这类查询,数据库访问该数据范围内的每条记录以处理该查询,其中每一记录包括一个或更多项。然而,在步骤 304,(例如在计算对该数据范围的信息查询的结果之前)对数据库 108 中的每一完整段 210 - 216 执行了对例如总和、平均值等信息的查询。以这种方式,在接收和/或计算从数据库中某数据范围查询信息的结果之前,计算对多个完整段 210 - 216 中的每一个的查询结果(例如预先计算)。这样的结果可作为一项存储在数据库 108 中(例如在数据库 108 的第三列 206 或第四列 208)。尽管该例子示出了明确添加到数据库表格中用于存储预先计算的值的其他列,但是可替代使用混合索引(hybrid index)。该混合索引可存储该预先计算的值作为元数据,



并且利用该索引的查询可从该索引而不是直接从该表格中检索完整段的元数据。

例如，计算分别代表 2000、2001、2002 和 2003 年的总收入的每一完整段 210 - 216 的数据总和。附加地或可替换地，计算分别代表 2000、2001、2002 和 2003 年的平均季度收入的每一完整段 210 - 216 的数据平均值。尽管数据库 108 的第三列 206 存储年数据，第四列 208 存储平均季度数据，但是混合索引可配置为存储代表其他时间段的数据(例如每日、每月等)。

在步骤 306，接收指定数据库中所定义数据范围的查询。更具体地，如上所述，可由数据库接收对数据库中某数据范围的诸如总和、平均值等信息的查询。例如，可向该数据库查询从 2000 年第三季度到 2003 年第二季度的总收入。

在步骤 308，确定在所定义的范围中是否有任何完整段。例如，设备 100 确定在预定义范围中有两个完整段 212 - 214。因为通过确定在定义的范围中是否有任何完整段 210 - 216 来预先计算对数据库 108 中的每一完整段 210 - 216 的这类查询结果，所以设备 100 可避免不得不访问所定义的范围中的每条记录以处理该查询。更具体地，该设备 100 可不需要(例如响应于该查询)访问与这样的完整段 212 - 214 对应的每条记录而处理该查询。

在步骤 310，对所定义范围内的任何部分段的数据执行查询。更具体地，设备 100 可确定该定义的范围包括一个或多个部分段(例如完整段的部分)的数据。在以上例子中，设备 100 确定该定义的范围包括代表 2000 年的收入的完整段 210 的部分(例如 2000 年最后两季度的收入)，并包括代表 2003 年的收入的完整段 216 的部分(例如 2003 年前两季度的收入)。设备 100 对每个这样的部分段定义的范围执行查询。例如，该设备 100 可通过访问该范围中的各项来确定 2000 年第三和第四季度的总收入。类似地，该设备 100 可通过访问该范围中的各项来确定 2003 年第一和第二季度的总收入。以这种方式，确定对所定义范围内的任何部分段数据的查询结果。

在步骤 312，基于已确定在所定义范围内的任何完整段的预先计算的查询结果和对所定义范围内的任何部分段的查询结果，确定指定该定义范围的查询的结果。更具体地，访问已确定在所定义范围内的每一完整段 212 - 214 的预先计算的查询结果。例如，访问第三列 206 中与 2001 年的总收入(例如“8000”)和 2002 年的总收入(例如“12000”)对应的项。将这样的项加到对

部分段的查询结果，以产生指定所定义范围的查询的结果。在采用混合索引的实施例中，访问该混合索引中的预先计算的查询结果。其后，执行步骤 314。在步骤 314，方法 300 结束。

现在参考图 2 和图 4 来描述用于处理数据库查询的设备 100 的操作，图 4 中图示了根据本发明的实施例的用于处理数据库查询的第二示例方法 400。参考图 4，在步骤 402，方法 400 开始。在步骤 404，接收对数据库中的第一部分数据进行查询的请求。更具体地，设备 100 可对数据库 108 的定義的范围查询信息。该定义的范围充当该第一部分。例如，该设备 100 可向该数据库查询从 2000 年第三季度到 2003 年第二季度的总收入，在此情况下，第一部分中包括从 2000 年第三季度到 2003 年第二季度的收入项。

在步骤 406，确定是否已为该数据库中的至少一个第二部分的数据预先计算了查询结果，其中该第二部分包括在第一部分中。第二部分可为例如全部包括在第一部分数据中的 2001 年季度收入数据和/或 2002 年季度收入数据。第二部分可能具有与其关联的预先计算结果。所以，在步骤 406，设备 100 可确定是否已预先计算了 2001 年的总收入(例如 2001 年的季度收入之和)和/或 2002 年的季度收入数据。如果在步骤 406 确定还没有为数据库中的至少一个第二部分的数据预先计算查询结果，则执行步骤 408。或者，如果确定已为数据库中的至少一个第二部分的数据预先计算了查询结果，则执行步骤 410。

在步骤 408，对该第二部分执行查询，并将这样的查询结果作为预先计算结果而存储。例如，与方法 300 的步骤 304 类似，该设备可在获得该第一部分数据的查询结果之前，针对完全包括在第一部分数据中的第二部分数据查询信息。例如，设备 100 可通过向数据库查询从 2001 年第一季度到 2001 年第四季度的总收入而预先计算 2001 年的总收入(在此情况下，第二部分中包括从 2001 年第一季度到 2001 年第四季度的收入项)，并将该查询结果作为预先计算结果而存储在数据库 108 的第三列 206 中。类似地，该设备 100 可通过向数据库查询从 2002 年第一季度到 2002 年第四季度的总收入而预先计算 2002 年的总收入(在此情况下，第二部分中包括从 2002 年第一季度到 2002 年第四季度的收入项)，并将该查询结果作为预先计算结果而存储在数据库 108 的第三列 206 中。以这种方式，对第二部分数据执行查询，并按照需要(例如仅响应于对第一部分数据执行查询的请求)存储这种查询的结果，从而避免

不必要的处理。其后，执行步骤 410。

在步骤 410，确定对数据库的第一部分的查询结果基于至少一个预先计算结果。一旦接收到对数据库中的第一部分数据执行查询的请求，设备 100 便可确定可以采用该预先计算结果来确定该查询结果。

在步骤 412，检索该预先计算结果。例如，该设备可访问数据库 108 的第三列 206 并检索 2001 年和/或 2002 年的总收入值。

在步骤 414，对数据库中的第三部分数据执行查询，其中该第三部分是第一数据中除了第二部分之外的部分。例如，设备 100 可向数据库查询 2000 年第三和第四季度以及 2003 年第一和第二季度的总收入，在此情况下，在第三部分中包括来自 2000 年第三和第四季度以及 2003 年第一和第二季度的收入项。

在步骤 416，基于至少一个预先计算结果和对第三部分数据的查询结果而计算对第一部分的查询结果。例如，将至少一个预先计算结果加到对第三部分数据的查询结果，以得到对数据库中的第一部分数据的查询结果。其后，执行步骤 418。在步骤 418，方法 400 结束。

通过使用方法 300 或方法 400，可确定指定了数据库中所定义的数据范围的查询的结果，而无需响应于该查询而访问与所定义的范围对应的每一记录。以这种方式，方法 300 和方法 400 可有效地查询数据库，从而提供超出传统数据库查询的性能增益。

前面的描述仅公开了本发明的示范实施例。落入本发明的范围之内的以上公开的设备和方法的变形对于本领域普通技术人员来说是显而易见的。例如，尽管在以上一个或多个实施例中，数据库中的数据按照日期而分组为多个段或多个部分，但是也可基于与该数据关联的其它参数而将数据库中的数据分组为多个段或多个部分。而且，在一个或多个实施例中，可监控完整段或第二部分中的数据，并因此监控存储在混合索引中的基于这样的数据的结果，以确定访问该数据的频率。可丢弃很少使用的数据。在一些实施例中，可使用算法来确定该数据被确定的频率。基于监控，该设备可确定是否继续存储基于该数据的结果。以这种方式，不使用的数据不会不必要地占用数据库空间。而且，可采用阈值年限，使得仅存储对最近日期范围的数据(例如通过阈值年限定义，如上一、月、星期等中存储的数据)执行的预先计算结果。例如，如果数据库包括长时间的数据，但仅访问最近日期范围内的数据，则

根据本方法和设备，仅存储对这样的日期范围执行的预先计算结果，从而节约数据库空间。可去除对过时的日期范围执行的预先计算结果。

因此，尽管已结合其示范实施例而公开了本发明，但应理解，其他实施例可落入所附权利要求限定的本发明的精神和范围中。

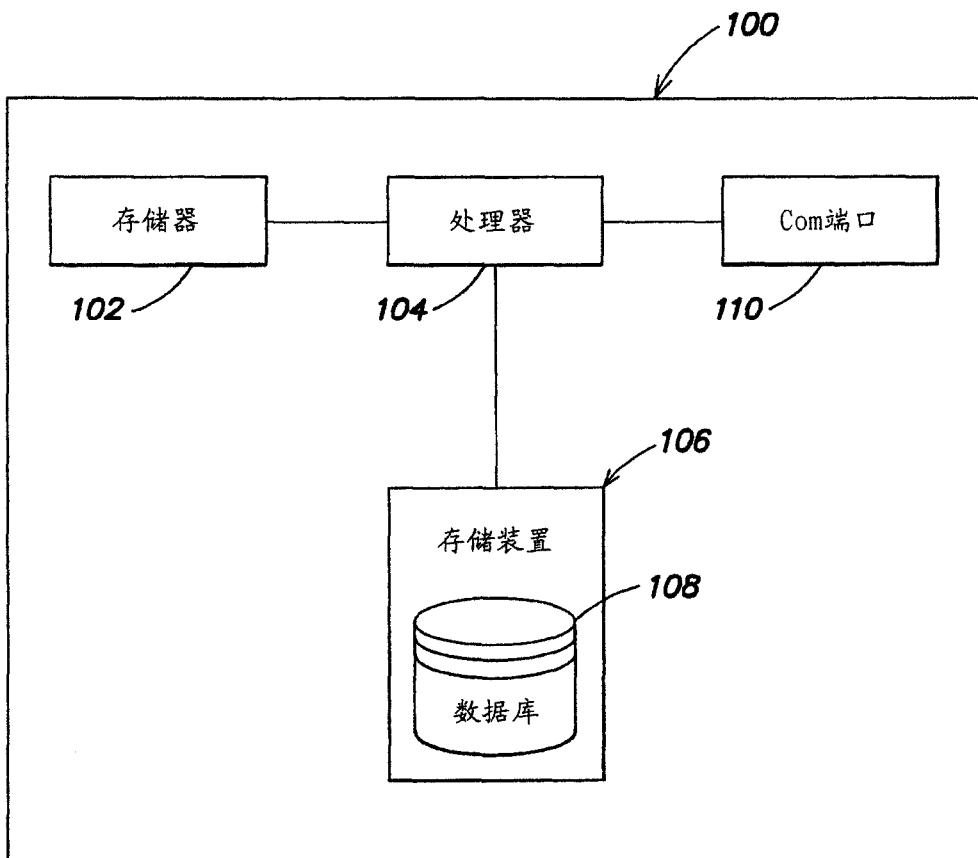


图 1

	A <sup>202</sup>	B <sup>204</sup>	C <sup>206</sup>	D <sup>208</sup>
1	日期	收入	年总和	年平均值
2	Q1 2000	1000	210	
3	Q2 2000	1000		
4	Q3 2000	1000		
5	Q4 2000	1000		4000
6	Q1 2001	2000	212	
7	Q2 2001	2000		
8	Q3 2001	2000		
9	Q4 2001	2000		8000
10	Q1 2002	3000	214	
11	Q2 2002	3000		
12	Q3 2002	3000		
13	Q4 2002	3000		12000
14	Q1 2003	4000	216	
15	Q2 2003	4000		
16	Q3 2003	4000		
17	Q4 2003	4000		16000
18	Q1 2004	5000		
19	Q2 2004	5000		
20	Q3 2004	5000		

• • •

•  
•  
•

108

图 2

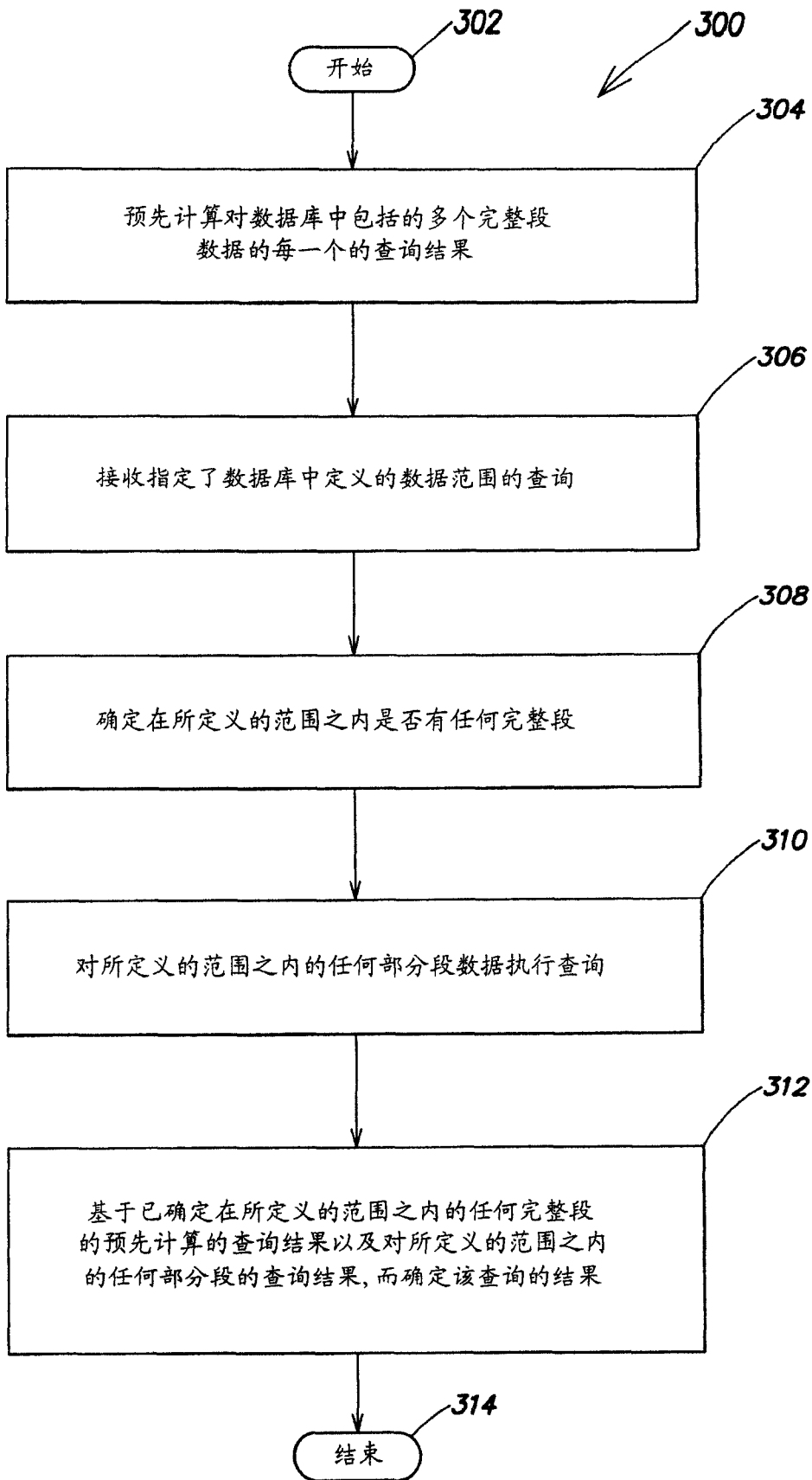


图 3

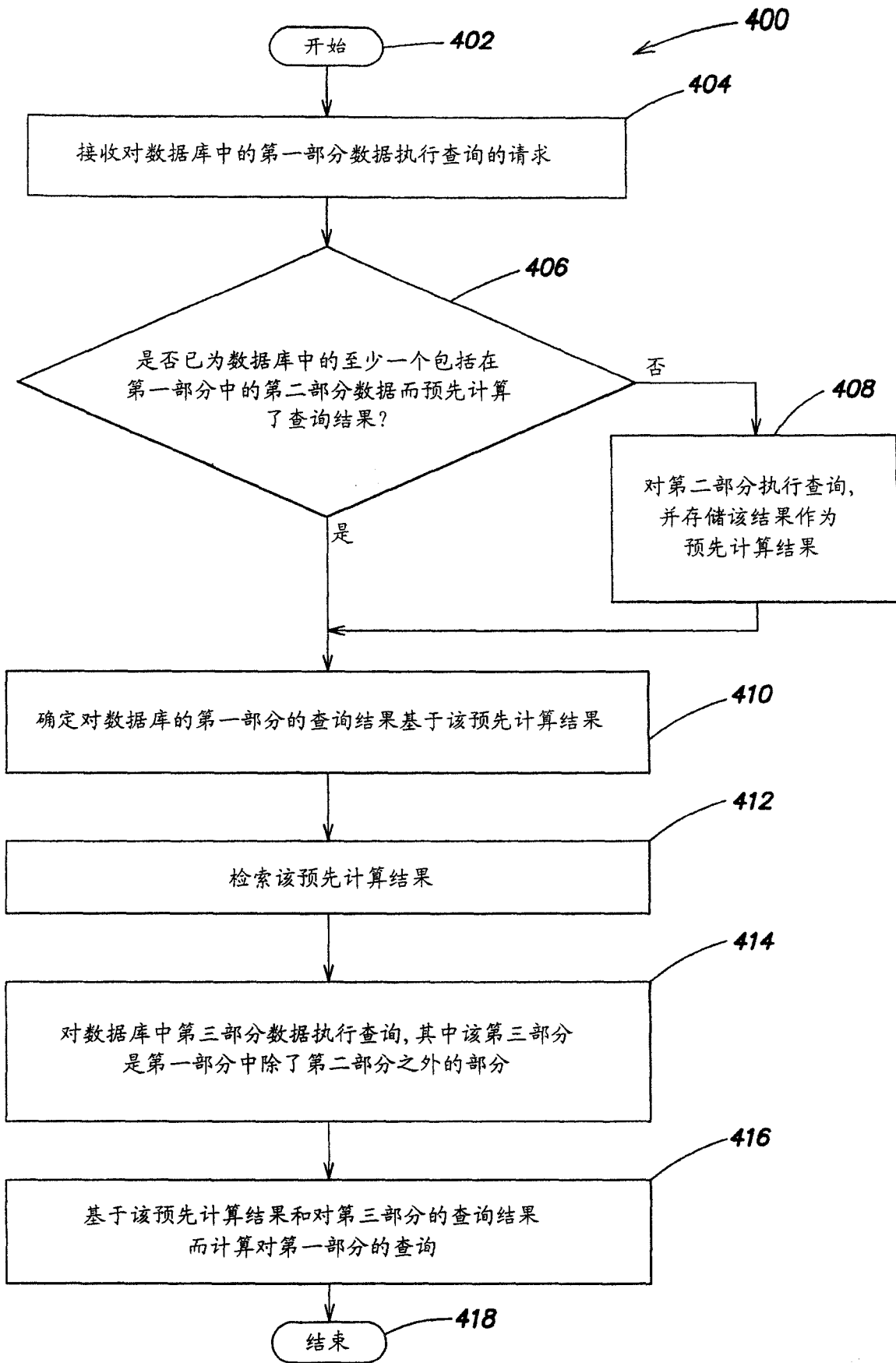


图 4