

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06F 15/16 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200780006818.2

[43] 公开日 2009年3月18日

[11] 公开号 CN 101390080A

[22] 申请日 2007.2.23

[21] 申请号 200780006818.2

[30] 优先权

[32] 2006.2.28 [33] US [31] 11/363,895

[86] 国际申请 PCT/US2007/004799 2007.2.23

[87] 国际公布 WO2007/100700 英 2007.9.7

[85] 进入国家阶段日期 2008.8.27

[71] 申请人 微软公司

地址 美国华盛顿州

[72] 发明人 G·萨瑞恩 M·A·伊萨德

Y·于 G·库马

[74] 专利代理机构 上海专利商标事务所有限公司
代理人 陈斌

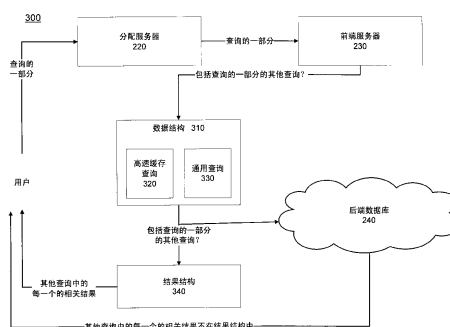
权利要求书 3 页 说明书 11 页 附图 8 页

[54] 发明名称

基于查询部分来供应高速缓存的查询结果

[57] 摘要

各实施例构想了一种用于获得关于查询的一部分的相关结果并为服务器的高速缓存生成一组经更新的查询的系统和方法。可以从包括服务器高速缓存和一组通用查询的数据结构中识别并获得以查询的一部分开头的其他查询。一旦其他查询被标识，从服务器高速缓存或后端数据库获得关于其他查询的结果。可以包括被删除的和额外的查询的一组通用查询可用于为服务器生成一组经更新的查询。将可能基于分配函数而属于服务器的遗漏查询和被删除的查询插入到一队列中，该队列随后被传送给服务器的高速缓存。该传送可以发生在低活动率或空闲状态期间。



1. 一种提供查询结果的方法，所述方法包括：
获得一查询的一部分；
将所述查询的所述部分分配给服务器（230）；
获得以所述查询的所述部分开头的其他查询；以及
基于所述查询的所述部分获得部分或所有所述其他查询的结果。
2. 如权利要求 1 所述的方法，其特征在于，还包括：
计算所述查询的所述部分的表示；
其中将所述查询的所述部分分配给服务器（230）包括基于所述表示将所述查询的所述部分分配给服务器（230）。
3. 如权利要求 1 所述的方法，其特征在于，还包括：
确定所述查询的所述部分是否与一预定条目匹配；以及
将后续字符分配为所述查询的所述部分。
4. 如权利要求 1 所述的方法，其特征在于，还包括：
将高速缓存查询（320）和一组通用查询（330）存储在一数据结构（310）中；
其中获得以所述查询的所述部分开头的其他查询包括从所述数据结构（310）获得以所述查询的所述部分开头的其他查询。
5. 如权利要求 4 所述的方法，其特征在于，还包括：
修改所述数据结构（310）以反映对所述高速缓存查询（320）所作的改变以及对所述一组通用查询（330）所作的改变。
6. 如权利要求 4 所述的方法，其特征在于，还包括：
根据使用频率存储所述高速缓存查询（320）和所述一组通用查询（330）；以及
按所述使用频率的降序显示以所述查询的所述部分开头的所述其他查询中的每一个的结果。
7. 如权利要求 1 所述的方法，其特征在于，获得以所述查询的所述部分开头的所述其他查询中的每一个的结果包括从所述服务器（230）的高速

缓存（410）获得所述其他查询中的每一个的结果。

8. 如权利要求 1 所述的方法，其特征在于，获得以所述查询的所述部分开头的所述其他查询中的每一个的结果包括从一后端服务器（240）获得所述其他查询中的每一个的结果。

9. 如权利要求 1 所述的方法，其特征在于，还包括：

如果不能获得以所述查询的所述部分开头的所述其他查询，则提供一空结果。

10. 一种查询系统（300），包括：

接收一查询的一部分的分配服务器（220）；

存储高速缓存查询（320）和一组通用查询（330）的数据结构（310）；

提供关于所述高速缓存查询（320）和所述一组通用查询组（330）的结果的结果结构（340）；以及

基于所述查询的所述部分咨询所述数据结构（310）和所述结果结构（340）以提供相关结果的前端服务器（230）。

11. 如权利要求 10 所述的系统，其特征在于，还包括后端数据库（240）。

12. 如权利要求 10 所述的系统，其特征在于，所述分配服务器（220）将所述查询的所述部分分配给所述前端服务器（230）。

13. 如权利要求 10 所述的系统，其特征在于，所述查询的所述部分是所述查询的前缀。

14. 一种构造一组通用查询（330）的方法，所述方法包括：

编译从前端服务器（230）的高速缓存（410）删除的条目；

形成由所述删除的条目组成的一组通用查询（330），其中所述删除的条目来自一预定时间范围；以及

将所述一组通用查询（330）分配给另一前端服务器（230）。

15. 如权利要求 14 所述的方法，其特征在于，将所述一组通用查询（330）分配给另一前端服务器（230）包括将所述一组通用查询（330）分配给多个前端服务器（230）。

16. 如权利要求 14 所述的方法，还包括：

基于一预定列表生成额外的查询；以及

将所述额外的查询添加到所述一组通用查询（330）。

17. 如权利要求 14 所述的方法，还包括：

为前端服务器（230）计算一组经更新的查询，其中所述一组经更新的查询由被分配给所述前端服务器（230）的所述一组通用查询中的所分配的查询组成；以及

为所述前端服务器（230）构造一数据结构（310），其中所述数据结构的内容由所述一组经更新的查询和高速缓存查询（320）组成。

18. 如权利要求 17 所述的方法，其特征在于，还包括：

计算遗漏查询，其中所述遗漏查询属于所述前端服务器（230）的高速缓存（410）；以及

将所述遗漏查询插入到一队列中。

19. 如权利要求 18 所述的方法，其特征在于，还包括：

将所述遗漏查询从所述队列移除；

将所述遗漏查询从所述队列发送到一后端数据库（240）；以及

将所述遗漏查询插入到所述前端服务器（230）的所述高速缓存（410）中。

20. 如权利要求 17 所述的方法，其特征在于，还包括：

确定从所述服务器（230）的所述高速缓存（410）过期的一过期查询是否属于所述一组经更新的查询；

将所述过期查询插入到一队列中；

将所述过期查询从所述队列中移除；

将所述过期查询从所述队列发送到一后端数据库（240）；以及

将所述过期查询插入到所述前端服务器（230）的所述高速缓存（410）中。

基于查询部分来供应高速缓存的查询结果

背景

web 搜索引擎将用户的搜索或者查询定向到分配服务器，该服务器又将该查询转发给前端服务器。前端服务器可以包含查询和相应的结果，或可以咨询后端数据库以获得查询结果。web 搜索引擎通常从用户输入相同搜索项而接收重复的查询。新近查询及其结果可以在前端服务器的前端高速缓存中保存一段预定的时间，因为咨询前端高速缓存以获得所需结果比咨询后端数据库更有效且耗时更少，其中该后端数据库包含全部 web 索引且一般比前端高速缓存大。全部 web 索引可以包括诸如网页等所有可获得的 web 结果；因此，由于索引比存储在前端高速缓存上的结果要大得多，会导致耗时的搜索。然而，如果在前端高速缓存中没有找到所需结果，就向后端数据库咨询。

除了存储新近查询和结果之外，前端高速缓存还可以用作有用搜索的记录，它可以包括用户更频繁地咨询的结果或系统管理员相信将被咨询的结果。

用户通常必须输入完整的搜索项或查询以便呈现相关结果。如果输入诸如前几个字符等查询的一部分，通常会将错误的消息或不正确的结果呈现给用户。另外，例如，如果用户不确定为了接收结果而输入的查询，则用户必须单独输入若干查询并筛选结果以找到所需网页。如果用户输入查询的一部分并接收以所输入的部分开头的一组查询的相关结果将是理想的。另外，当用户继续输入查询时由此扩展该部分时，该结果随着查询改变而自动调整将是理想的。

另外，如果通过进一步利用前端高速缓存来保存资源且更快速地完成操作，那么这种部分查询机制则更理想。

概述

通过标识以查询的一部分开头的其他查询并获得其他查询的结果来获得查询的一部分的相关结果。该查询的一部分可被分配给接收以该查询的相同部分开头的其他查询的前端服务器。其他查询可存储在包括服务器高速缓存和一

组通用查询的数据结构中。其他查询的结果可存储在服务器高速缓存或后端数据库中。

通过编译从服务器高速缓存删除的已删除查询来形成一组通用查询。该组通用查询还可以包括来自预定列表的额外查询。

可以使用该组通用查询来计算对服务器高速缓存的一组经更新的查询。遗漏查询可以是属于服务器高速缓存但也许不存储在服务器高速缓存中的查询。遗漏查询可以是该组通用查询的一部分。遗漏查询可以在空闲或低活动率期间插入到队列中并随后被置于服务器高速缓存中。同样地，从服务器高速缓存删除的过期查询也可被插入队列中并随后被置于服务器高速缓存中。

提供本概述以便以简化形式介绍将在以下详细描述中进一步描述的一些概念。本概述并不旨在确定所要求保护的主题的关键特征或必要特征，也不旨在用于帮助确定所要求保护的主题的范围。

附图简述

结合各附图阅读以下详细描述将更好地理解前述摘要和以下详细描述。示例性实施例在各图中示出，然而应当理解，各实施例不局限于此处描述的具体方法和手段中。在各图中：

图 1 是表示一示例性计算设备的框图；

图 2 是表示一示例性服务器组织的框图；

图 3 是表示一查询系统的框图；

图 4 是表示一查询构造系统的框图；

图 5a 和图 5b 是示出提供查询结果的方法的一实施例的流程图；

图 6 是示出构造一组通用查询的方法的一实施例的流程图；以及

图 7 是示出更新一组高速缓存查询的方法的一实施例的流程图。

详细描述

参见图 1，用于实现本发明的示例性系统包括计算机 110 形式的通用计算设备。计算机 110 的组件包括但不限于：处理单元 120、系统存储器 130 和将包括系统存储器在内的各个系统组件耦合到处理单元 120 的系统总线 121。系

统总线 121 可以是几种类型的总线结构中的任何一种，包括存储器总线或存储控制器、外围总线、以及使用各种总线体系结构中的任何一种的局部总线。作为例子而非限制，这样的体系结构包括工业标准体系结构（ISA）总线、微通道体系结构（MCA）总线、增强型 ISA（EISA）总线、视频电子技术标准协会（VESA）局部总线、以及外围部件互连（PCI）总线（也被称为夹层（Mezzanine）总线）。

计算机 110 通常包括各种计算机可读介质。计算机可读介质可以是能由计算机 110 访问的任何可用介质，且包括易失性和非易失性介质、以及可移动和不可移动介质。作为例子而非限制，计算机可读介质可以包括计算机存储介质和通信介质。计算机存储介质可以包括以用于存储如计算机可读指令、数据结构、程序模块或其他数据这样的信息的任何方法或技术实现的易失性和非易失性、可移动和不可移动介质。计算机存储介质包括但不限于：RAM、ROM、EEPROM、闪存或者其他存储器技术、CD-ROM、数字多功能盘（DVD）或者其他光盘存储、磁带盒、磁带、磁盘存储或者其他磁存储设备、或者能够用来存储所需信息并且能够由计算机 110 访问的任何其它介质。通信介质通常以诸如载波或者其他传输机制的已调制数据信号来体现计算机可读指令、数据结构、程序模块或者其他数据，并且包括任何信息传递介质。术语“已调制数据信号”是指以在该信号中编码信息的方式来设置或改变其一个或多个特性的信号。作为例子而非限制，通信介质包括有线介质，诸如有线网络或者直接线连接，以及无线介质，诸如声学、射频（RF）、红外以及其它无线介质。任何上述组合都应包括在计算机可读介质的范围之内。

系统存储器 130 包括易失性和/或非易失性存储器形式的计算机可读存储介质，诸如只读存储器（ROM）131 和随机存取存储器（RAM）132。基本输入/输出系统 133（BIOS）一般被存储在 ROM 131 中，BIOS 包含有助于例如在启动过程中在计算机 110 内部的元件之间传输信息的基本例程。RAM 132 一般包含可由处理单元 120 立即访问和/或当前正由处理单元 120 进行操作的数据和/或程序模块。作为示例而非限制，图 1 示出了操作系统 134、应用程序 135、其他程序模块 136 和程序数据 137。

计算机 110 还可以包括其它可移动/不可移动、易失性/非易失性计算机存

储介质。仅仅作为示例，图 1 示出了从不可移动、非易失性磁介质读取或向其写入的硬盘驱动器 141，从可移动、非易失性磁盘 152 读取或向其写入的磁盘驱动器 151，以及从诸如 CD-ROM 或其它光学介质等可移动、非易失性光盘 156 读取或向其写入的光盘驱动器 155。能在示例性操作环境中使用的其它可移动/不可移动、易失性/非易失性计算机存储介质包括但不限于，盒式磁带、闪存卡、数字多功能盘、数字录像带、固态 RAM、固态 ROM 等等。硬盘驱动器 141 通常通过诸如接口 140 等不可移动存储器接口连接到系统总线 121，而磁盘驱动器 151 和光盘驱动器 155 通常通过诸如接口 150 等可移动存储器接口连接到系统总线 121。

以上讨论且示于图 1 中的驱动器及其相关联计算机存储介质为计算机 110 提供了对计算机可读指令、数据结构、程序模块及其它数据的存储。在图 1 中，例如，硬盘驱动器 141 被示为存储操作系统 144、应用程序 145、其它程序模块 146 和程序数据 147。应当注意，这些组件可以和操作系统 134、应用程序 135、其他程序模块 136 和程序数据 137 相同或者相异。操作系统 144、应用程序 145、其他程序模块 146 和程序数据 147 在此处被给予了不同的编号以示出至少它们是不同的副本。用户可以经由输入设备，诸如键盘 162 和通常称之为鼠标、跟踪球或者触摸垫的指示设备 161 将命令和信息输入到计算机 110 中。其它输入设备（未示出）可包括话筒、操纵杆、游戏手柄、圆盘式卫星天线、扫描仪等。这些及其它输入设备经常经由耦合到系统总线的用户输入接口 160 耦合到处理单元 120，但是可以通过其它接口和总线结构，诸如并行端口、游戏端口或通用串行总线（USB）进行连接。监视器 191 或其他类型的显示设备也通过诸如视频接口 190 等接口耦合到系统总线 121。除了监视器之外，计算机还可以包括诸如扬声器 197 和打印机 196 等其他外围输出设备，它们可以通过输出外围接口 190 进行连接。

计算机 110 可以利用与诸如远程计算机 180 等一个或多个远程计算机的逻辑连接在网络化环境中操作。远程计算机 180 可以是个人计算机、服务器、路由器、网络 PC、对等设备或者其它常见的网络节点，而且通常包括许多或者所有以上相对于计算机 110 所描述的元件，尽管在图 1 中仅示出了存储器存储设备 181。所描述的逻辑连接包括局域网（LAN）171 和广域网（WAN）173，

但是也可以包括其它网络。这种网络环境常见于办公室、企业范围计算机网络、内联网和因特网。

当在 LAN 网络环境中使用时，计算机 110 经由网络接口或适配器 170 连接到 LAN 171。当在 WAN 网络环境中使用时，计算机 110 通常包括调制解调器 172 或者用于经由如因特网等 WAN 173 建立通信的其他装置。可以是内置或者外置的调制解调器 172 可经由用户输入接口 160 或者其他适当的机制连接到系统总线 121。在网络化环境中，相对于计算机 110 描述的程序模块或其部分可存储在远程存储器设备中。作为示例而非限制，图 1 将远程应用程序 185 示为驻留在存储器设备 181 上。应当明白，所示网络连接是示例性的，也可以使用在计算机之间建立通信链路的其它手段。

以上所描述的本发明的各方法的全部或部分可以用硬件、软件或两者的组合来实施。当以软件来实施时，本发明的各方法或其某些方面或部分可用当被计算系统执行时使该计算系统执行本发明的各方法的程序代码的形式来实施。如上对该术语的定义，该程序代码可存储在任意计算机可读介质上。

图 2 表示可用来为诸如通过例如计算机 110 输入到因特网搜索引擎中的查询或查询的部分等查询提供结果的服务器组织 200。查询源自因特网 210 上，其中用户可以插入他们想要获得的结果（诸如针对查询的网页）的查询。可在搜索引擎上输入查询，搜索引擎可以允许用户插入感兴趣的查询且可以显示相关结果供用户访问。因特网 210 可以连接到从因特网 210 接收查询的分配服务器 220。多个分配服务器 220 可以连接到因特网 210。图 2 示出了 3 个分配服务器 220（220a、220b 和 220c），但分配服务器 220 的数量不受限制。查询可以通过例如查询的起源网际协议（IP）地址来分配给分配服务器 220 中的特定一个。可以利用其他方案将查询分配给分配服务器 220 之一。或者，可以随机地将查询发送给分配服务器 220 中的任何一个。

接收查询的分配服务器 220 可以将查询发送给前端服务器 230，诸如图 2 的服务器组织 220 中所示的前端服务器 230a、230b、230c 和 230n 中的一个。服务器组织 200 不受前端服务器 230 的特定个数的限制。从分配服务器 220 接收查询的前端服务器 230 可基于将查询链接到前端服务器 230 的映射来做到这一点。例如，查询可被散列化以产生一个进而被用于映射函数的值，该函数将

某个查询置于特定前端服务器 230 中且可以确保相同查询总是或几乎总是被定向到相同的前端服务器 230。映射函数可以确定前端服务器 230 中的哪一个将接收查询。不需要基于映射函数来分配前端服务器 230，而是可以例如且不限于随机地接收查询。

高速缓存或本地存储可以包含查询及其相应的结果。高速缓存可以是前端服务器 230 的一部分，且高速缓存的查询和结果可以是诸如来自一预定时间段（如 3 小时）的一组新近查询和结果。当前端服务器 230 从分配服务器 220 接收查询时，前端服务器 230 可以搜索其高速缓存以确定该查询及其相应的结果是否已存储在高速缓存中。如果查询在高速缓存中，则前端服务器 230 则可以返回来自高速缓存的结果。该结果可以被提供给输入查询的用户。

前端服务器 230 可与包含例如网页等现有结果的后端数据库 240 通信。查询结果可被存储在后端数据库 240 中。如果输入的查询不在前端服务器 230 的高速缓存中，则可以咨询后端数据库 240 以便向用户返回相关的查询结果。因为后端数据库 240 一般比前端服务器 230 的高速缓存大得多，所以咨询后端数据库比咨询前端服务器 230 的高速缓存要花更长的时间。因此，咨询后端数据库 240 仅发生在例如当前端服务器 230 不包含所请求的结果或当前端服务器 230 遇到诸如服务器故障等问题的时候。

图 3 是表示可用于获得对查询的一部分的结果的查询系统 300 的框图。查询系统 300 可以包括用于接收查询的一部分的分配服务器 220。查询的一部分可以是例如查询的前 3 个字符。想得到查询结果的用户可以将查询的一部分键入或输入到因特网搜索引擎中。可基于查询的一部分所起源的计算机 110 的 IP 地址将该查询的该部分定向到分配服务器 220。查询的该部分可被随机地定向到分配服务器 220。

分配服务器 220 可以负责将查询的一部分分配给前端服务器 230。为了这项分配任务，分配服务器 220 可以使用散列函数。例如，分配服务器 220 可以将查询的该部分分配给前端服务器 230，从而具有相同部分的其他查询将被分配给相同的前端服务器 230。在查询的该部分被分配给前端服务器 230 之后，分配服务器 220 可以将查询的该部分发送给所分配的前端服务器 230。

一旦前端服务器 230 接收了查询的该部分，前端服务器 230 可以负责执行

搜索以便获得具有与查询的该部分相同的部分的其他查询。为了获得这些查询，前端服务器 230 可以咨询数据结构 310。虽然在图 3 中被示为单独的组件，但是数据结构 310 可以是前端服务器 230 的一部分。数据结构 310 可以包括诸如高速缓存查询 320 等服务器高速缓存。高速缓存查询 320 可以包括已被保存在数据结构 310 中一段预定时间的先前的查询。因此，例如，高速缓存查询 320 可以包括最新近的查询。高速缓存查询 320 可以被保存在数据结构 310 中以便简单且快速地访问新近查询。数据结构 310 还可以包括一组通用查询 330。该组通用查询的形成将在以下更详细地描述。

数据结构可以是扩充的特里 (trie) 结构，它是促进与查询或查询的一部分相关的信息的快速检索的结构。例如，查询的可能部分以及与查询的每个部分相关的额外信息可以被存储在特里结构中。与查询的每个部分相关的信息可以是例如以查询的该部分开头的预定个数的最频繁请求的查询的列表。该列表可按频率降序排序，从而首先列出以查询的该部分开头的最常请求的查询。

当从数据结构 310 获得可以包括查询的该部分的其他查询列表时，就咨询一结果结构 340，该结构可以是高速缓存结构。该结果结构可以是前端服务器 230 的一部分。结果结构 340 包含查询的结果。在咨询结果结构 340 之后，获得以查询的该部分开头的查询结果并将其提供给用户。这些结果可以是例如多个网页选集。

后端服务器 240 还可作为查询系统 300 的一部分来包括。如果包括查询的该部分的查询中的一个或多个的结果不位于结果结构 340 中，则可以咨询后端数据库 240。如图 3 所示，还将从后端数据库 240 获得的结果提供给用户。

通用查询构造系统 400 的框图表示在图 4 中示出。系统 400 可被分配创建通用查询 330 的任务，通用查询可用于获得具有与查询的该部分相同的部分的查询。通用查询构造系统 400 可以包括从分配服务器 220 接收查询的该部分的前端服务器 230。服务器高速缓存 410 可以是前端服务器 230 的一部分。服务器高速缓存 410 可以是存储诸如已被保存了一段预定时间的查询等先前查询的前端服务器 230 的高速缓存结构。前端服务器 230 的服务器高速缓存 410 可以是与数据结构 310 的高速缓存查询 320 相同的组件，它可以是前端服务器 230 的一部分。

查询可以经常从前端服务器 230 的服务器高速缓存 410 删除。删除可由于服务器高速缓存 410 的大小限制而发生；可能需要删除查询以便为较新的查询（或查询的一部分）创建存储空间。或者，可以周期性地发生删除。当从服务器高速缓存 410 删除查询时，日志条目 420 可以接收并跟踪被删除的查询。一段预定时间过去之后，日志条目 420 可以向日志条目编译器 430 提供被删除的查询，后者编译被删除的条目。日志条目编译器 430 可以基于例如在一周内出现不止一次的条目等先前确定的准则来编译被删除的条目。一旦日志条目编译器 430 完成了对被删除的条目的编译，它就可以将此编译提供给数据结构 310 的通用查询 330。数据结构 310 然后存储通用查询以便用于将具有与搜索的查询的该部分相同的部分的查询提供给结果结构 340。

通用查询 330 可以用作高速缓存查询 320 的补充。高速缓存查询 320 可以是更新近的查询，而不一定是新近查询的通用查询 330 可以从相对较长的预定时间（如一周）频繁地请求的查询。

图 5a 和 5b 示出了提供查询结果的方法。在 505 处，获得查询的一部分。查询的该部分可由想获得与查询相关的结果的用户启动，该用户将查询的该部分输入到例如因特网搜索引擎中。可在分配服务器 220 上获得查询的该部分。在 510 处，检查查询的该部分以便确定查询的该部分是否与预定条目匹配。如果查询的该部分与预定条目匹配，则在 515 处，将后续字符作为查询的一部分来分配。因为可能存在用户并无意使其成为查询的一部分的多个频繁出现的条目，因此该比较可以排除提供对不需要的查询的查询结果。例如，如果用户输入以“http”和“www”开头的此类查询，则用户一般不想搜索这些查询，而是对这些频繁出现的条目之后的字符感兴趣。因此，通过将诸如“www”之后的 3 个字符等后续字符作为查询的一部分来分配，可以搜索查询的合适部分。可将查询的该部分与其他预定条目比较。

在后续字符作为查询的一部分分配之后，如果查询的该部分与预定条目不匹配，则提供查询结果的方法分别从 515 和 510 前进到 520。在 520 处，可以计算查询的该部分的表示。例如，该表示可以是查询的该部分的散列函数。其他表示也是可能的。

接着 520，在 525 处，可将查询的该部分分配给前端服务器 230。当在 520

处计算查询的该部分的表示时，然后可使用该表示以便将查询的该部分分配给前端服务器 230。可使用其他分配准则来将查询的该部分分配给前端服务器。或者，分配可以是随机选择。在 530 处，可将查询的该部分发送给指定的前端服务器 230。

535 确定诸如以上根据图 3 和图 4 描述的数据结构 310 等数据结构是否包含以查询的该部分开头的查询。如果数据结构 310（它可以包括高速缓存查询 320，而后者可以包括已被保存在数据结构 310 中一段预定时间的先前查询和通用查询 330）不包含以查询的该部分开头的查询，则在 540 处向用户提供空结果列表。然而，如果数据结构 310 包含以查询的该部分开头的查询，则该方法前进到 545（图 5b）。在 545 处，可从数据结构 310 获得以查询的该部分开头的查询。

在 550 处，对每个获得的以查询的该部分开头的查询作出判定以查明结果是否被存储在诸如结果结构 340 等作为前端服务器 230 的一部分的服务器高速缓存中。在 560 处，如果关于所获得的以查询的该部分开头的查询的结果不在结果结构 340 中，则从后端数据库 240 获得结果。在 555 处，如果关于所获得的以查询的该部分开头的查询的结果在结果结构 340 中，则从结果结构 340 中获得结果。对每个所获得的包括查询的该部分的查询执行 555 处的判定。这可以确保为包括查询的该部分的每个查询提供结果，而不管结果是来自后端数据库 240 还是来自结果结构 340。提供查询的方法在提供结果的 565 处终止。

图 6 是示出构造一组通用查询的方法的流程图。在 610 处，基于预定列表生成查询。该预定列表可以包括系统管理员相信应为一组通用查询 330 的一部分的查询。例如，预定查询可以是还没有被用户在查询搜索中使用过但有望在不久的将来变得流行且被广泛搜索的一部新电影。通过将这种查询添加到作为数据结构 310 一部分的通用查询 330 中，针对查询或查询的一部分的搜索就不需向后端数据库 240 的咨询。

在 615 处，编译从诸如前端服务器 230 的服务器高速缓存 410 或高速缓存查询 320 等服务器的高速缓存删除的查询。可在预定时间内编译被删除的查询。如以上根据图 4 的通用查询构造系统 400 所描述的，被删除的查询可由日志条目 420 跟踪并由日志编译器 430 编译。

在 620 处，形成一组通用查询 330。通用查询可以包括在 610 处生成的预定查询的组合或编译以及在 615 处生成的被删除的查询。在形成通用查询 330 之后，在 625 处，将通用查询 330 分配给前端服务器 230。如果被删除的查询是从诸如图 2 的前端服务器 230b 等特定前端服务器删除的查询，则通用查询则可被分配给系统 200 的其他前端服务器 230a、230c 和 230n。

图 7 是示出更新可以作为前端服务器 230 的一部分的前端服务器的数据结构 310 的高速缓存查询 320 的方法的流程图。高速缓存查询 320 可以包括新近查询。在 710 处，计算一组经更新的查询。该组经更新的查询可以包括来自通用查询 330 的查询，该查询将基于例如操作来将相同查询和查询的相同部分定向到相同的前端服务器 230 的散列和映射函数而被分配给前端服务器 230。可以由前端服务器 230 来计算该组经更新的查询。在 720 处，该组经更新的查询可与高速缓存查询 320 一起存储在数据结构 310 中。这有助于更新高速缓存查询 320。

在 730 处，确定在该组经更新的查询中但不在高速缓存查询 320 中的查询（遗漏查询）。如果没有查询与此准则匹配，则该方法继而搜索与此准则匹配的查询。方法继续以使遗漏查询被添加到高速缓存查询 320。在 740 处，将遗漏查询添加到队列。

该方法可以前进到 750，其中分析系统以确定该系统是否为轻负荷。每次前端服务器 230 向后端系统 240 发送请求时，来自后端系统 240 的响应可以包括可用于确定后端 240 的当前状态的反馈，该状态可以包括当前后端负荷。例如，反馈可以包括在队列中等待被后端 240 处理的请求的平均数量或后端 240 在查询处理时间限制内能够获得的答复的质量。前端服务器 230 然后可以使用一算法来基于反馈确定当前后端 240 是应被视为重负荷还是轻负荷。如果系统为轻负荷，则该方法则前进到 760。

从 740 或 750 中任一个到 760 处，从队列中移除遗漏查询。在 770 处，可将遗漏查询传送给后端数据库 240。在 780 处，后端数据库 240 用遗漏查询更新高速缓存查询 320。遗漏查询不再从高速缓存查询 320 “遗漏”。

应当理解，所公开的各实施例可以被整体或部分地在一个或多个计算系统或设备中实现。图 1 示出了其中可实现或实施各方面的计算系统 100 的一个示

例的功能性组件。如此处所用的，术语“计算系统”、“计算机系统”以及“计算机”是指包括能够执行或以别的方式处理程序代码和/或数据的处理器的任何机器、系统或设备。不旨在任何限制，计算系统的示例包括个人计算机（PC）、微型计算机、大型计算机、瘦客户机、网络 PC、服务器、工作站、膝上型电脑、手持式计算机、可编程消费者电子产品、多媒体控制台、游戏控制台、卫星接收器、机顶盒、自动取款机、街机、移动电话、个人数字助理（PDA）以及任何其他基于处理器的系统或机器。术语“程序代码”和“代码”是指由处理器执行或以其他方式处理的任何指令集。程序代码和/或数据可用执行特定功能的例程、程序、对象、模块、数据结构等的形式来实现。

应当注意，前述示例仅是出于解释的目的提供的，且不应被理解为限制性的。虽然已参考各实施例描述了本发明，但是应当理解，此处所使用的词是描述和说明性的词，而非限制的词。此外，虽然此处已经参考具体方法、材料和示例描述了各实施例，但是各实施例不旨在限于此处所公开的细节；相反，各实施例可扩展到所有功能上等效的结构、方法和用途，这些都在所附权利要求书的范围内。

计算环境 100

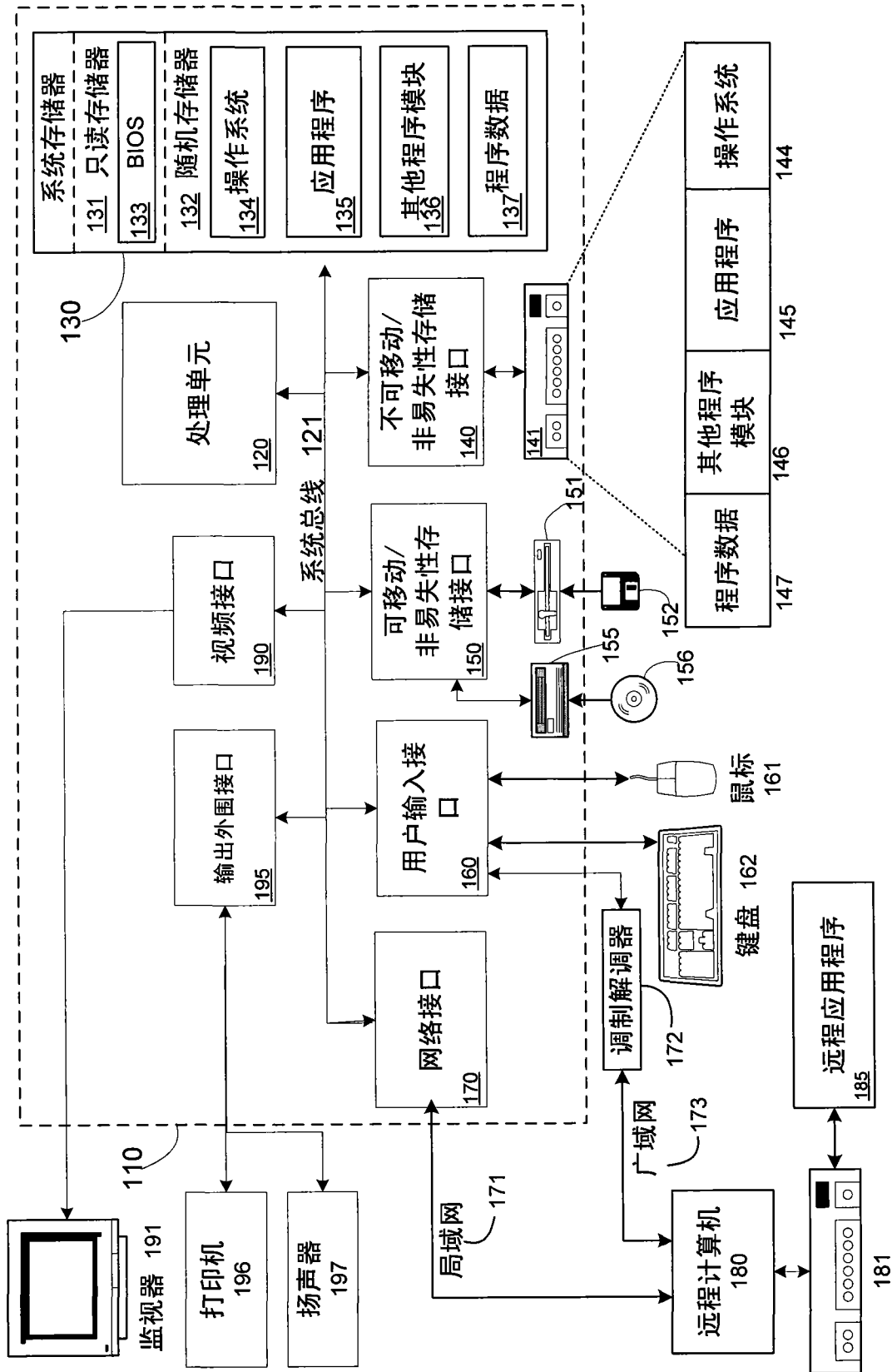
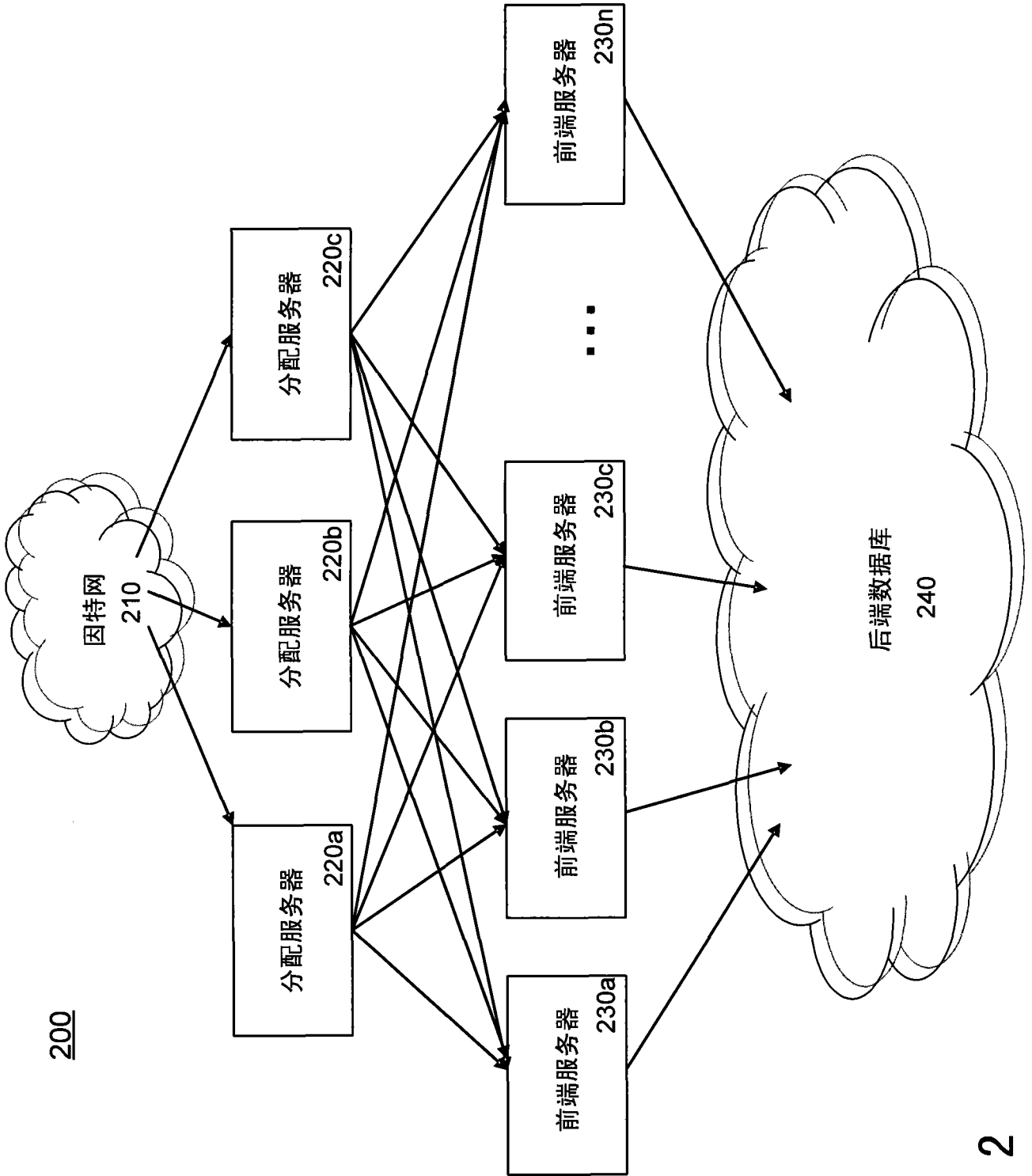


图 1



200

图 2

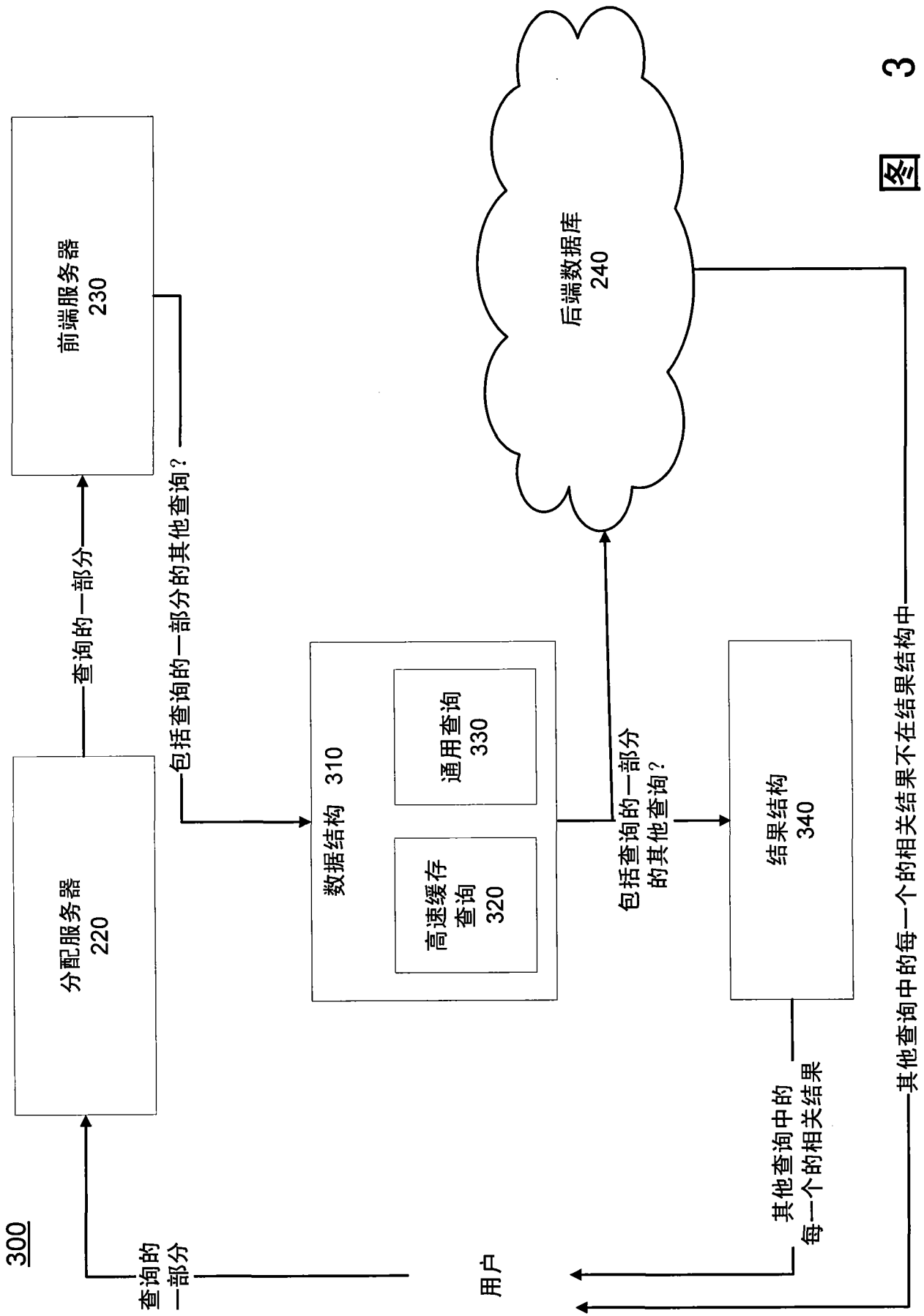


图 3

400

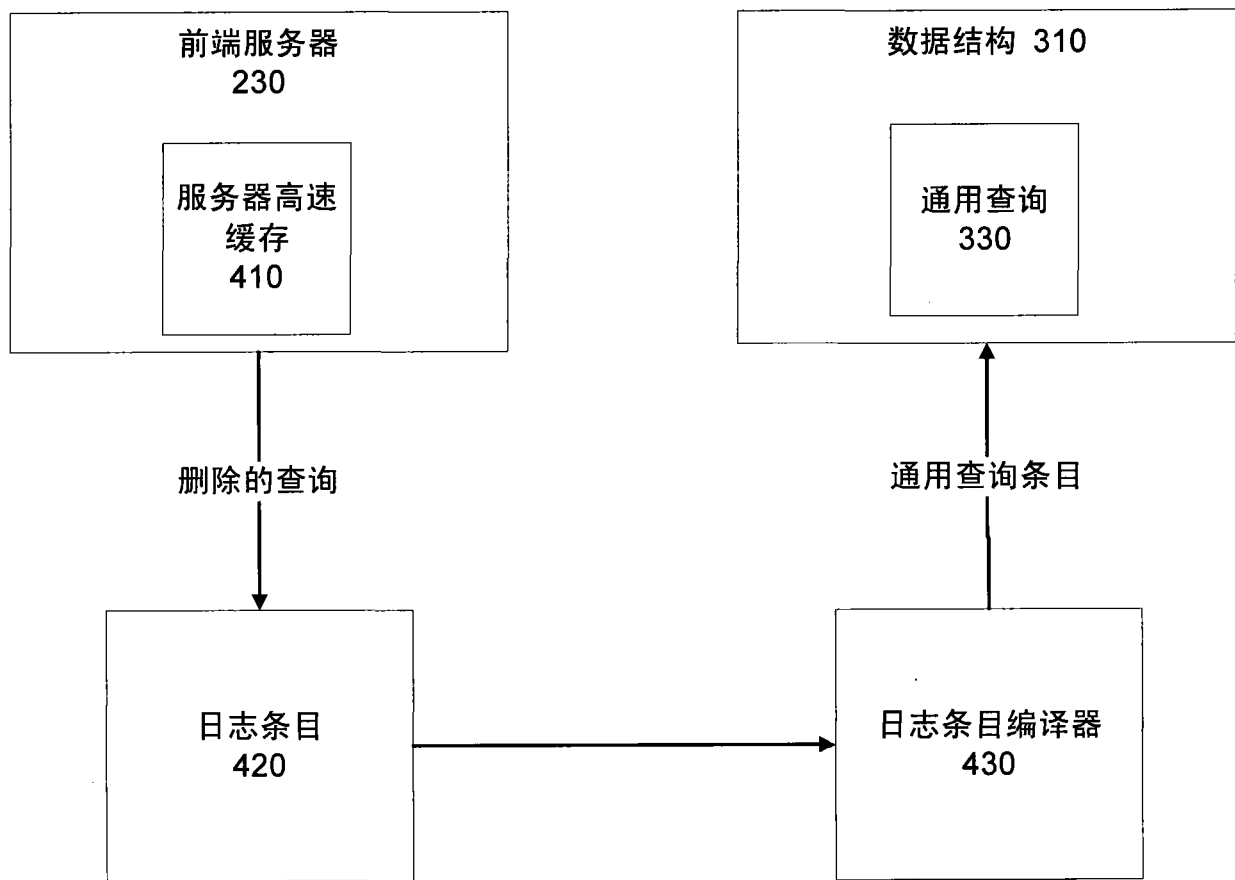


图 4

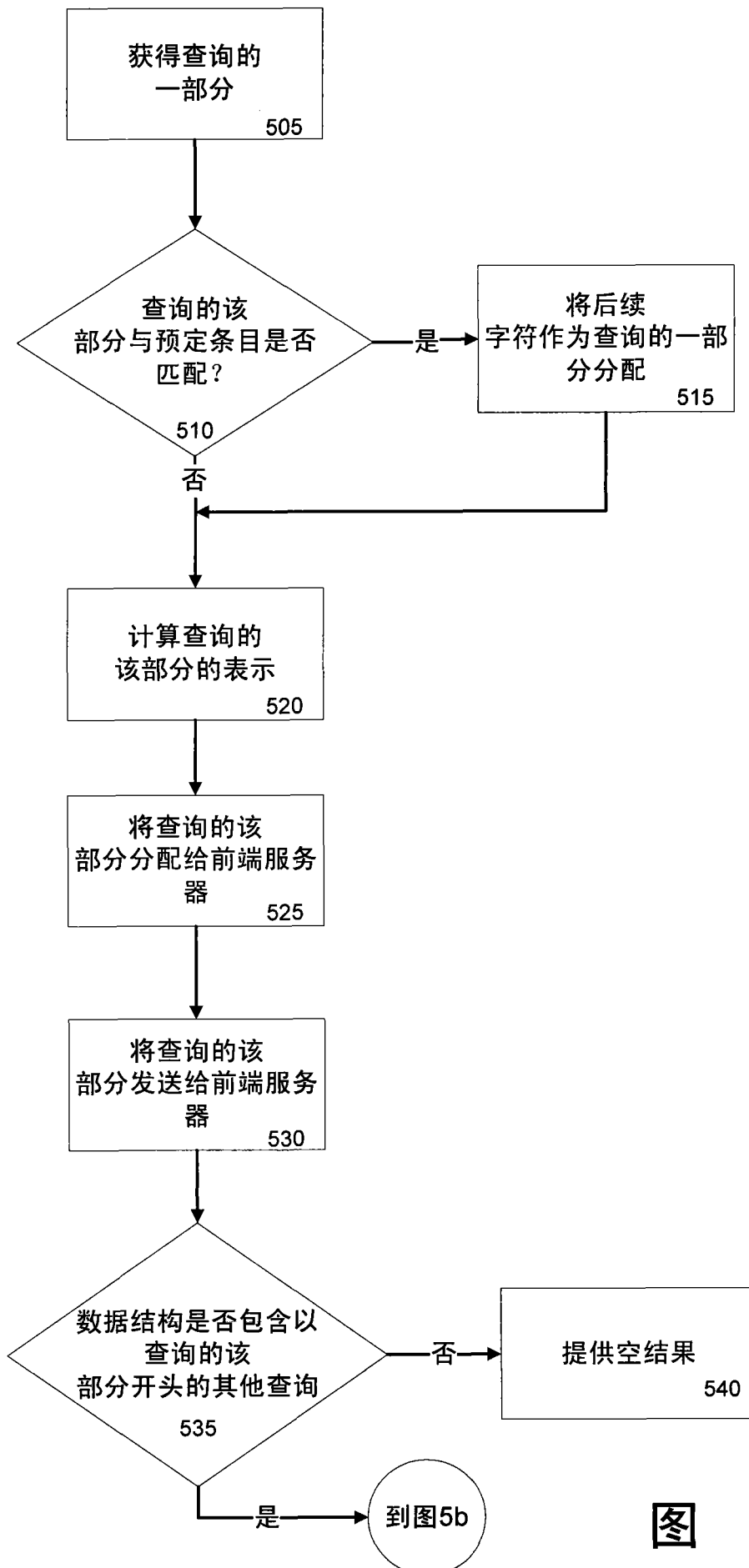


图 5a

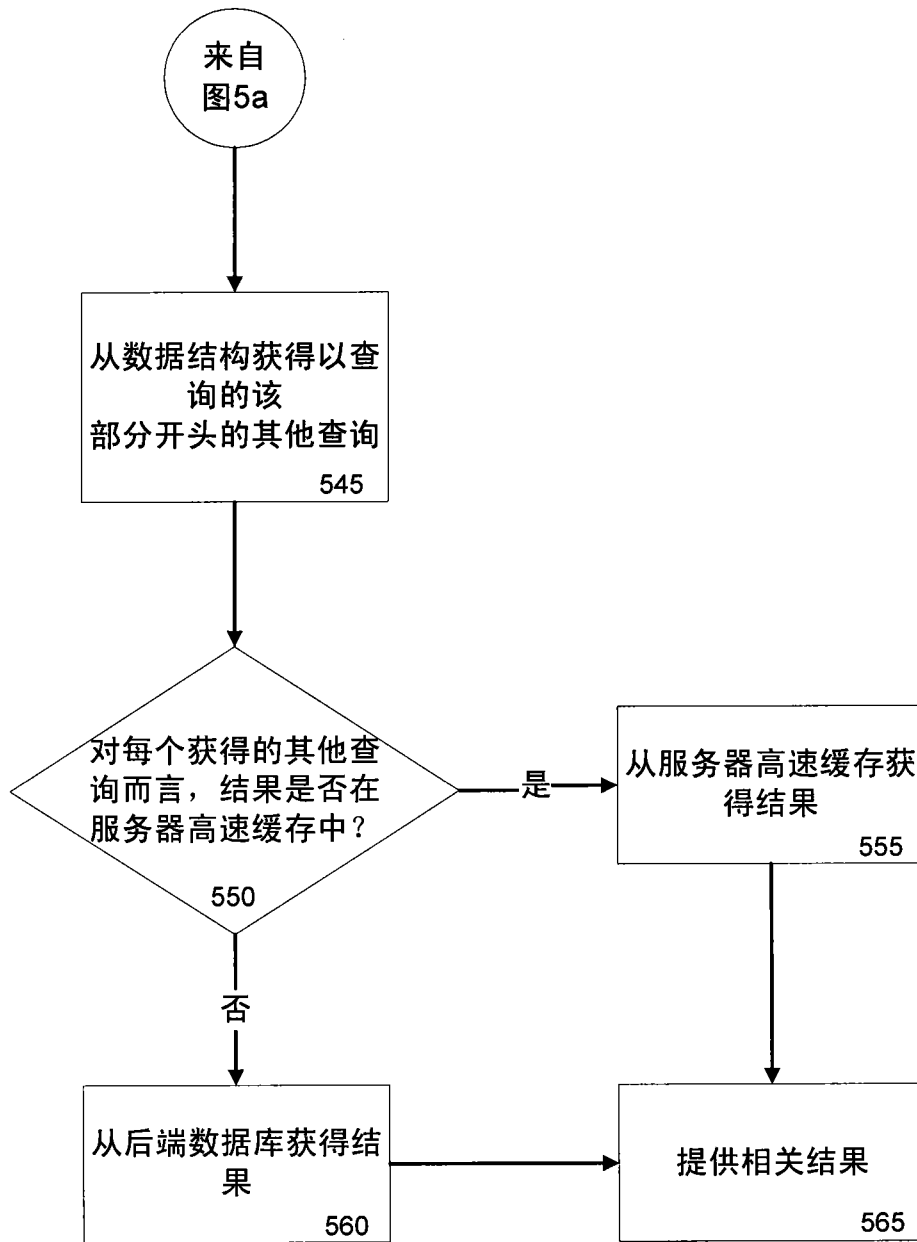


图 5b

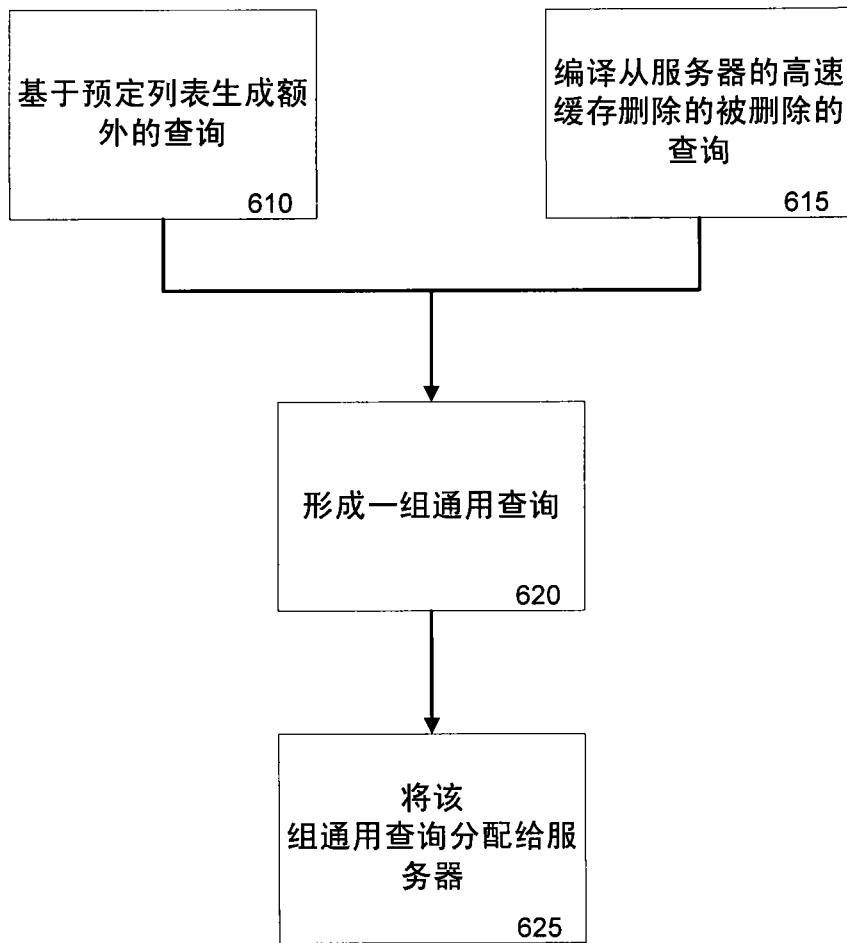


图 6

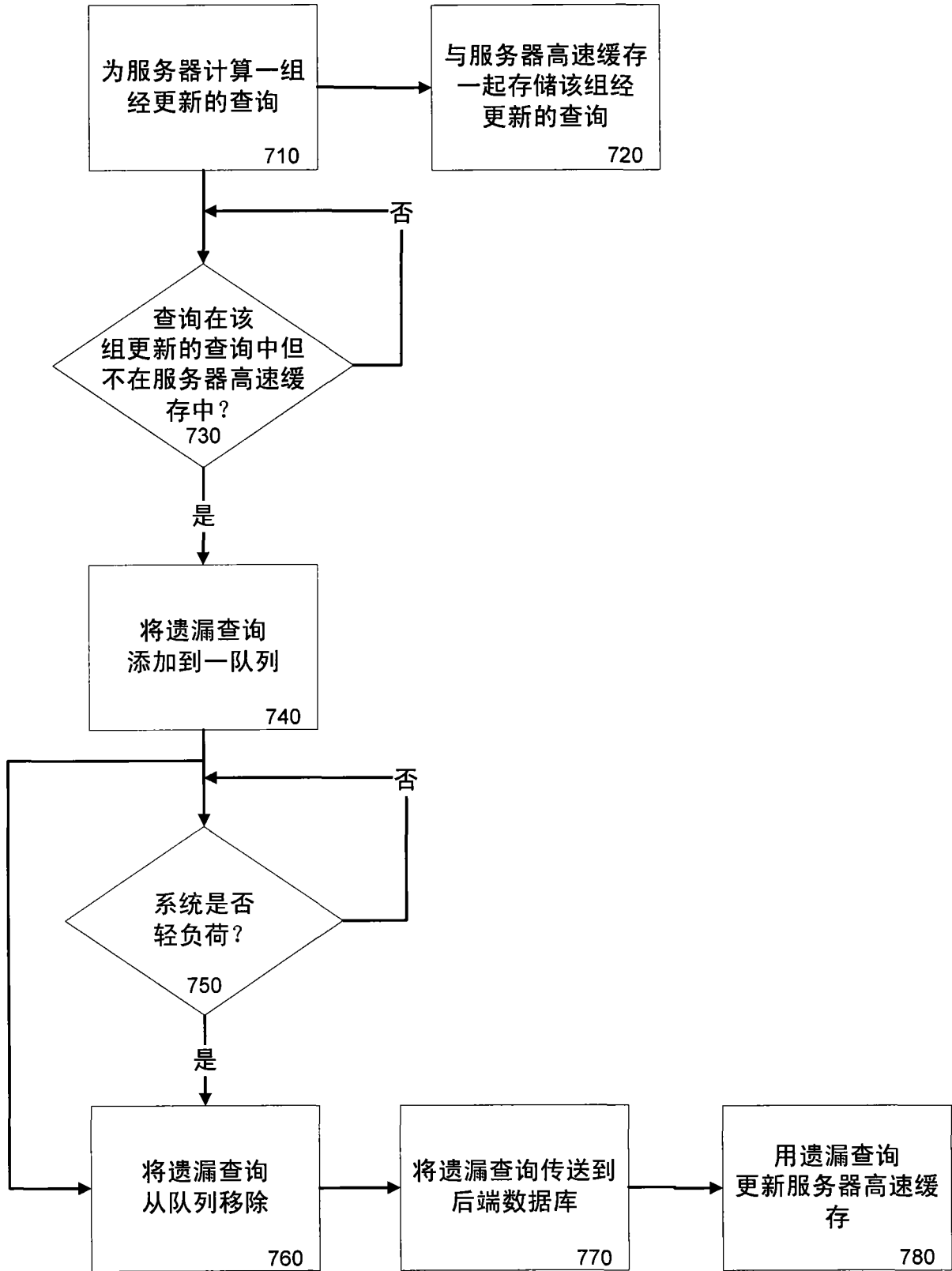


图 7