



[12] 发明专利说明书

专利号 ZL 200710013409.7

[45] 授权公告日 2009 年 5 月 6 日

[11] 授权公告号 CN 100485689C

[22] 申请日 2007.1.30

[21] 申请号 200710013409.7

[73] 专利权人 浪潮通信信息系统有限公司
地址 250014 山东省济南市历下区山大路
224 号

[72] 发明人 张 明

[56] 参考文献

US5640551A 1997.6.17

CN1447940A 2003.10.8

US2004/0260769A1 2004.12.23

US5155484A 1992.10.13

CN1684072A 2005.10.19

审查员 孙 娟

[74] 专利代理机构 济南信达专利事务所有限公司
代理人 姜 明

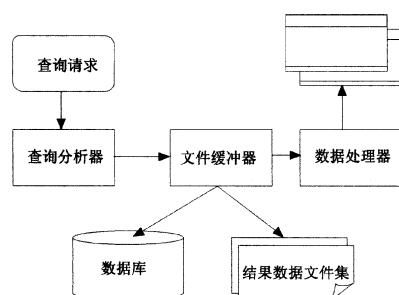
权利要求书 2 页 说明书 6 页 附图 3 页

[54] 发明名称

基于文件系统缓存的数据加速查询方法

[57] 摘要

本发明涉及一种查询系统的数据加速查询方法。本发明适用于针对历史性数据，存储后不变化，相同数据多次查询的特点，提供一种能够基于文件系统缓存的数据加速查询方法，通过将历史查询数据保存，利用历史查询数据作为查询缓存数据，进行的加速查询方法。在系统接收用户查询请求之后，先转到查询分析器，查询分析器的主要作用是分析查询条件，规划查询条件等，然后经过文件缓冲器，分析现有的结果数据文件中是否存在满足当前查询需求，没有则从数据库中进行查询，并将最终的结果文件返回给数据处理器，数据处理器按照最终的查询要求，对数据进行过滤、排序等处理，将结果文件返回上层应用系统。本查询方法在不增加硬件投资下，可以大大的降低数据库和系统资源的占用，大大提高查询速度。



1、基于文件系统缓存的数据加速查询方法，其特征在于，在系统接收用户查询请求之后，先转到查询分析器，查询分析器的主要作用是分析查询条件，规一化查询条件，然后经过文件缓冲器，分析现有的结果数据文件中是否存在满足当前查询需求，没有则从数据库中进行查询，并将最终的结果文件返回给数据处理器，数据处理器按照最终的查询要求，对数据进行过滤、排序等处理，将结果文件返回上层应用系统，具体的执行步骤如下：

步骤（1）、分析查询提交的查询条件，将查询条件分为数据库运算处理和数据处理器处理的两大部分，数据库运算处理主要是对数据的提取和逻辑计算部分的查询条件，数据处理器处理主要包括对结果数据的过滤和排序操作；

步骤（2）、规一化查询条件，将查询条件统一成标准的查询条件，避免某些查询描述上不同而实际上是相同情况，规一化查询包括：

1) 集合判断合并，将小集合的条件转换成大集合，将预先定义子集和全集的关系存在各个系统中表示对应关系的字典表中，利用这个字典表描绘出子集和全集的关系；

2) 查询条件排序，将查询条件按照固定顺序进行排列，采用字符的先后顺序，使其成一个相同的查询条件；

步骤（3）形成统一的查询条件之后，进入到文件缓冲器中，文件缓冲器先找到对应的查询实例，再根据查询条件再登记列表中检索满足当前条件的历史数据文件，该索引列表记录了该查询实例历史数据的查询条件，查询时间，用户，使用次数信息；

步骤（4）根据检索返回的结果，判断是否全部存在所要的数据文件，如果存在转到步骤（8），否则转到步骤（5）；

步骤（5）将分析后的查询条件在数据库中查询，生成所需要的数据文件，并存放到对应的数据目录中；

步骤（6）将生成的数据文件，登记到文件缓冲器的文件检索列表中；

步骤（7）返回需要的结果数据文件编号；

步骤（8）根据返回的文件编号，获取数据文件；

步骤（9）数据处理器对数据返回的数据文件进行处理，从结果文件中还原出需要的查询数据，并且将一些查询条件，本来需要在数据库的运算转移到数据处理器处理，提高对历史数据文件再使用率，从而增加缓冲器的命中率，增加命中率的方法还包括：

1) 选取查询所需要的数据集合，从全集合中过滤出需要的子集数据，合并成所需要的数据集；

2) 选取查询所需要的显示列；

3) 按照查询的要求对结果数据排序；

步骤（10）按照固定的存贮结构，将最终的结果数据返回给上层应用，处理结束。

基于文件系统缓存的数据加速查询方法

1. 技术领域

本发明涉及一种计算机应用技术领域通讯技术领域，具体地说是一种电信查询系统中的数据加速查询方法。适用于针对历史性数据，存储后不变化，存在大量重复查询的情况，进行的基于文件系统缓存的数据加速查询。

2. 技术背景

社会的不断发展，企业进行越来越多的支撑系统建设，信息共享变得尤为重要，信息的共享使得软件系统的访问用户增多，但是随着用户的不断增多，数据的查询量必然会大大增加，而系统就会越来越慢，从而需要投入更多的资金进行设备和软件的升级。造成这样的原因是，用户的每次查询请求，系统都是通过再数据库中进行相应的查找，数据计算等操作，然后将结果返回。用户的查询量增加，系统的数据处理量也随之线性增加。

尤其是在电信网领域中，有很多支撑系统，例如计费运营，网络管理系统，数据基本上都是历史性的数据，数据值不会发生变化。因此针对这类数据只要查询条件不变，查询的结果总是一样的。而且数据量非常大，大多数的查询可能需要进行大数据量查询和计算，才能够得出，处理这样的查询本身就需要很长的时间。这样使得重复的查询请求在数据库中查询和处理了多遍，而结果却是一样的。同样随着用户的不断增多，数据的查询量增加，这种重复性的查询也会随之增多。为此需要投入更多的资金来进行软件和硬件的升级，来提高系统的处理能力，减少用户的等待的时长，而对大数据量查询提速也是非常有限的。

3. 发明内容

本发明针对现有系统的应用情况，历史数据不变化和相同数据多次查询的特点，提供一种能够基于文件系统缓存的数据加速查询方法，通过将历史查询数据保存，利用历史查询数据作为查询缓存数据，进行的加速查询方法。从而能够避免重复查询请求，系统处理多次，使系统不会随着用户的查询量增加，数据处理量也随之线性增加。而且还能够利用历史结果数据，加速数据的查询，提高了系统运行的效率，降低用户的平均响应时间，减少因用户的使用增加等原因，而进行的软件和硬件升级，节约大量资金。

本发明为解决其技术问题，所采用的技术方案如图 1 所示。在系统接收用户查询请求之后，先转到查询分析器，查询分析器的主要作用是分析查询条件，规一

化查询条件等，然后经过文件缓冲器，分析现有的结果数据文件中是否存在满足当前查询需求，没有则从数据库中进行查询，并将最终的结果文件返回给数据处理器，数据处理器按照最终的查询要求，对数据进行过滤、排序等处理，将结果文件返回上层应用系统。

按照上述描述的技术方案，具体的执行步骤如图 2 所示，说明如下：

步骤 1、分析查询提交的查询条件，将查询条件分为数据库运算处理和数据处理器处理的两大部分。数据库运算处理主要是对数据的提取和逻辑计算部分的查询条件。数据处理器处理主要包括对结果数据的过滤和排序等操作。

步骤 2、规一化查询条件，将查询条件统一成标准的查询条件，避免某些查询描述上不同而实际上是相同情况。包括：

1) 集合判断合并，将小集合的条件转换成大集合。例如查询条件是查询山东的济南、青岛、潍坊等 17 个地市所有的数据，而这 17 个地市的属于山东省这个集合，与查询山东结果是一样的，系统会先将这个查询条件转换为查询山东这个集合的数据。子集和全集的关系需要预先定义，一般各个系统中都会存在表示对应关系的字典表，利用这个字典表就可以描绘出子集和全集的关系。

2) 查询条件排序，将查询条件按照固定顺序进行排列，采用字符的先后顺序。例如查询济南、青岛、潍坊这样一个地区条件顺序的数据，与查询青岛、济南、潍坊这样一个地区条件顺序的数据，结果是一样的，系统能够进行统一的顺序排列，使其成一个相同的查询条件。

步骤 3、形成统一的查询条件之后，送入到文件缓冲器中。文件缓冲器先找到对应的查询实例（例如要查询的报表）的历史数据登记列表，再根据查询条件再登记列表中检索满足当前条件的历史数据文件。该索引列表记录了该查询实例历史数据的查询条件，查询时间，用户，使用次数等信息。

检索列表命中率的高低直接影响到加速的效果，因此为了提高查询文件缓冲器检索数据文件的命中率，系统通过采用如下检索方法：

1) 从全集数据中获取子集的数据。如果要查询若干个子集，而这若干个子集恰好属于另外一个全集的一部分，系统将返回该数据文件。例如：查询济南、青岛两个地区的数据，系统先查找是否有对应这两个地区的数据，如果没有再从全省的数据集文件中查找，系统能够利用全省的数据集进行处理。

2) 部分子集查找。如果要查询若干个子集，而这若干个子集只有部分被检索到，那么系统会到步骤 5，只生成没有的部分子集。例如：查询济南、青岛两个地区的数据，系统只找到济南一个地区的数据，同时也没有全省的数据，系统将查询条件转化为只查询青岛，到数据库中进行查询。

为了提高文件缓冲器检索的高效性，降低检索文件所消耗的时间，系统通过采用如下方法，减少检索量。

3) 采用按照查询实例分目录存贮，降低了检索量。系统会自动的针对每一个查询实例（例如：针对每一个报表），形成一个文件缓冲目录，每一个目录都有一个索引列表。每次查询只去检索对应的索引列表。

4) 超期历史数据删除。保证历史的数据文件不至于过多，占用大量的存贮空间，也会造成检索文件造成过大，增大检索的时间。同时过早的历史数据对于用户的查询来说，使用的频率极低，没有缓冲再次使用的意义。系统会将超过保存周期和使用频率低的文件删除，从而减少文件缓冲列表的大小。

5) 查询结果文件聚合。系统为了保证检索的高效性，避免文件的过多，系统会将多个子集文件合并成一个全集文件。例如：将山东省 17 地区的各个子集文件合并成为一个全省的数据文件。

步骤 4、根据检索返回的结果，判断是否全部存在所要的数据文件，如果存在转到步骤 8，否则转到步骤 5。

步骤 5、将分析后的查询条件在数据库中查询，生成所需要的数据文件，并存放到对应的数据目录中。

步骤 6、将生成的数据文件，登记到文件缓冲器的文件检索列表中。

步骤 7、返回需要的结果数据文件编号。

步骤 8、根据返回的文件编号，获取数据文件。

步骤 9、数据处理器对数据返回的数据文件进行处理，从结果文件中还原出需要的查询数据。并且将一些查询条件，本来需要在数据库的运算转移到数据处理器处理，目的是为了提高对历史数据文件再使用率，从而增加缓冲器的命中率。包括：

1) 选取查询所需要的数据集合。从全集合中过滤出需要的子集数据，例如从全省数据集中，获取济南、青岛的数据。或者将多个子集数据合并，例如将济南、青岛两个子集的数据，合并成所需要的数据集。

2) 选取查询所需要的显示列。例如：一个结果数据文件中包含 10 列数据，从中过滤出需要的 6 列数据。

3) 结果排序。按照查询的要求对结果数据排序。

步骤 10、按照固定的存贮结构，将最终的结果数据返回给上层应用，处理结束。

从上面具体处理过程可以看出，查询的历史结果数据被充分的利用在 2 次查询中，提高了文件系统缓存数据的命中率，查询速度和系统处理能力上也会有很大的提高，并且没有改变用户的操作行为。

4. 有益效果

本查询方法在不提高服务器处理能力的下，采用分析历史查询结果，避免每次都查询数据库。遇到重复的查询，通过是检索文件登记列表，直接将历史结果进行返回。同时提供了多种手段，提高检索文件的命中率。而检索数据文件消耗的时间相对于查询数据库和数据处理的时间，可以忽略不计，而系统的运算量非常小。一旦查询被命中，用户几乎感觉不到等待的时间，大大降低数据库和系统资源的使用负荷，避免了用户的查询量增加，系统的数据处理量也随之线性增加，节约了因软件和硬件的升级所需要的大量资金，提高了系统的处理能力。

基于文件系统缓存的数据加速查询方法，具有的优势是：

- (1) 采用文件的方式进行保存，存储方式简单，稳定，不容易损坏。
- (2) 一个文件损坏或者丢失，不影响查询系统的使用，系统健壮性比较高。
- (3) 不采用内存的方式存贮，在服务器出现异常或者突然关机时，不会丢失数据。而检索硬盘和检索内存的速度差别，用户几乎感觉不到。
- (4) 利用文件系统，能够提供比较的缓冲空间，可以保存比较多的缓冲数据。

5. 附图说明

附图 1 是本发明的工作原理框图；

附图 2 是具体处理步骤流程图；

附图 3 是一般性能查询系统的处理过程示意图；

附图 4 是本发明的查询系统处理过程示意图。

6. 具体实施方式

以电信领域综合网络管理的性能系统为例说明本发明的具体应用。性能系统的数据特点是都是历史性的数据，存储以后的数据值不发生变化，而且数据量非常的大，一个普通的省级网管系统来说，一天的数据量能够达到 10G 左右。性能系统属于电信领域的一个重要的支撑平台，要支持各个地市级子公司和省公司各个方面人员的使用，因此访问量也是非常大，必然造成大量的数据被重复处理。一般报表的查询系统处理步骤如图 3 所示，利用本发明在报表系统上，使用该基于文件系统缓存的数据加速查询方法，修改为查询步骤如图 4 所示。先将查询请求转发到文件缓冲器中，由文件缓冲器通过对查询条件分析，数据获取，数据处理等过程，再返回用户要查询的结果集。从而提高系统的处理能力和效率，减少用户查询的平均响应时间。

实施例

例如针对性能系统的报表查询，具体的执行步骤如下：

步骤 1、用户查询报表 A，从页面上设置好要查询的条件并提交查询请求，查

询条件例如日期是 2006-11-21 10:00:00、2006-11-21 20:00:00、2006-11-21 17:00:00，范围是：所属济南所有的区县、淄博、泰安，按照 X 列的降序进行排列。

步骤 2、将查询请求转发到查询分析器，开始分析提交的查询条件，将查询分为数据库运算处理和数据处理器处理的两大部分。

- 数据库运算条件是：日期是 2006-11-21 10:00:00 到 2006-11-21 20:00:00，范围是：济南、泰安、淄博。

- 数据处理器处理条件是按照 X 列的降序进行排列。

步骤 3、规一化查询条件，将查询条件统一成标准的查询条件，避免某些查询描述上不同而实际上是相同情况。通过集合判断方法，得出查询所属济南所有的区县，与查询地区等于济南的查询条件是一样的，因此将查询条件修改为查询济南的数据。并对查询条件按照字母先后顺序进行排序，最终数据库运算条件转换为：日期是 2006-11-21 10:00:00、2006-11-21 17:00:00、2006-11-21 20:00:00，范围是：济南、淄博、泰安。

步骤 4、形成统一的查询条件之后，送入到文件缓冲器中，文件缓冲器根据要查询报表 A 的 ID，找到对应的目录下找到历史数据登记列表，根据查询条件检索登记列表中满足当前条件的历史数据文件。例如查找结果为存在全省集合的 2006-11-21 10:00:00 和 2006-11-21 20:00:00 的数据。

步骤 5、根据检索返回的结果，得出所需要的结果文件只有部分存在，缺少 2006-11-21 17:00:00 的数据，将查询请求转发。

步骤 6、将分析后的查询条件，在数据库中查询日期是 2006-11-21 17:00:00，范围是：济南、淄博、泰安的数据，生成标准格式的数据文件，并存放到对应报表 A 的数据缓冲目录中。

步骤 7、将生成的数据文件，登记到文件缓冲器的文件检索列表中。登记的条件为：日期是 2006-11-21 17:00:00，范围是：济南、淄博、泰安。

步骤 8、返回登记的结果数据文件编号。

步骤 9、根据返回的文件编号，获取数据文件。当前得到的数据文件为：全省集合的 2006-11-21 10:00:00、2006-11-21 20:00:00 数据，和济南、淄博、泰安的 2006-11-21 17:00:00 数据。

步骤 10、数据处理器对数据返回的数据文件进行处理。从结果文件中过滤出需要的数据集合，得到济南、淄博、泰安的 2006-11-21 10:00:00、2006-11-21 17:00:00、2006-11-21 20:00:00 数据，然后按照 X 列的降序进行排序。一般来说查询所选择的源数据量是比较大的，处理出来的结果数据量是非常小的，因此对结果数据进行过滤和排列等操作，花费的处理时间可以忽略。

步骤 11、按照固定的存贮结构，将最终的结果数据返回给上层应用，进行报表展现，反馈给用户，处理结束。

根据上面描述的实例，可以看出采用本发明改造后的报表系统，在没有改变用户的操作行为下，系统的数据处理量变为原来的三分之一，响应速度基本可以到达原来的 3 倍，自然的也提高了系统的处理能力。随着用户查询量的不断增加，重复查询率必然增大，在原有硬件的投资下，加速的效果会更加明显。

综上所述，基于文件系统缓存的网管性能数据加速查询方法，适用于针对历史性数据，存储后不变化，大量重复查询请求的情况，尤其是大型系统会获得更好的效果。可以广泛应用于电信、银行、医疗、教育、财务等领域，例如：医疗系统的病人病历档案分析，财务系统的历史帐务，公司运营等方面的分析。

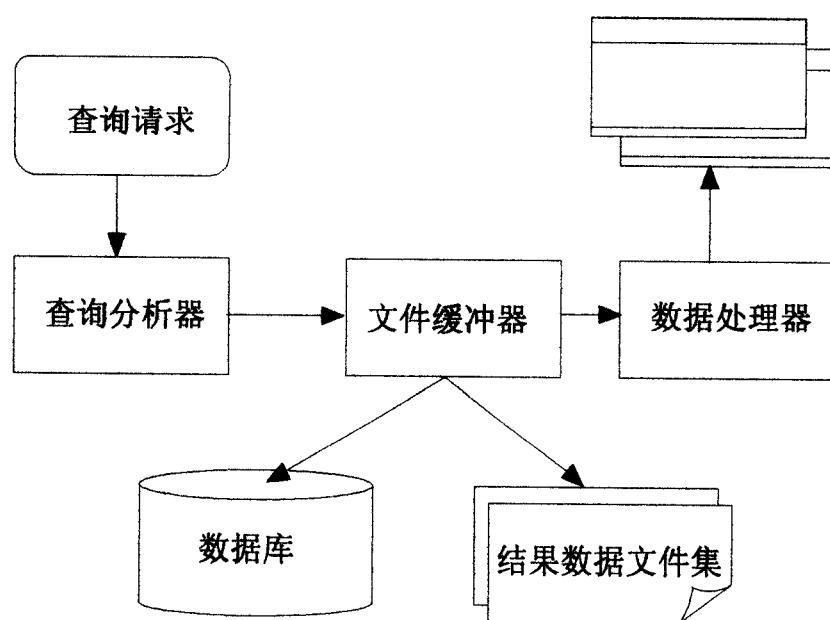


图 1

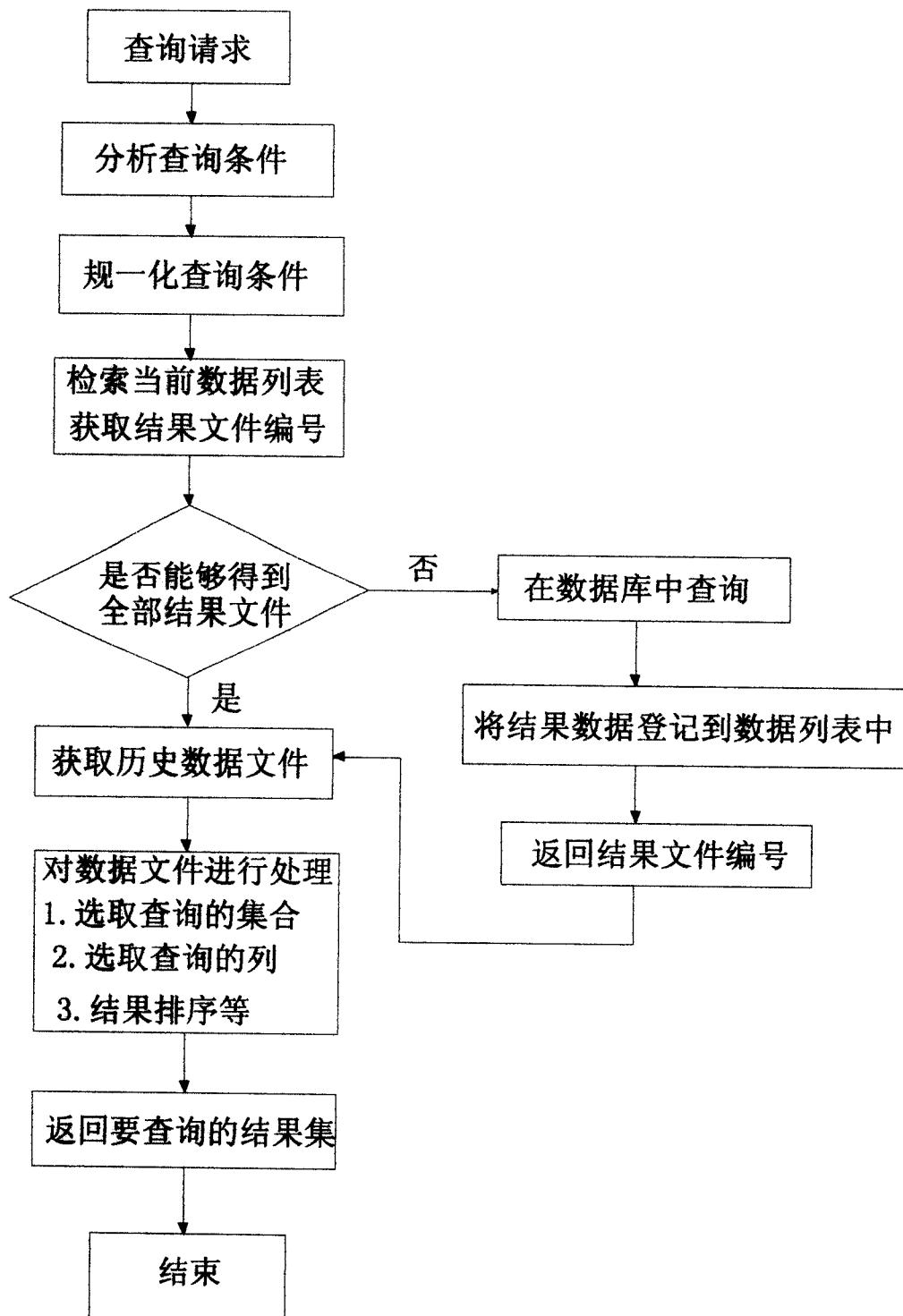


图 2

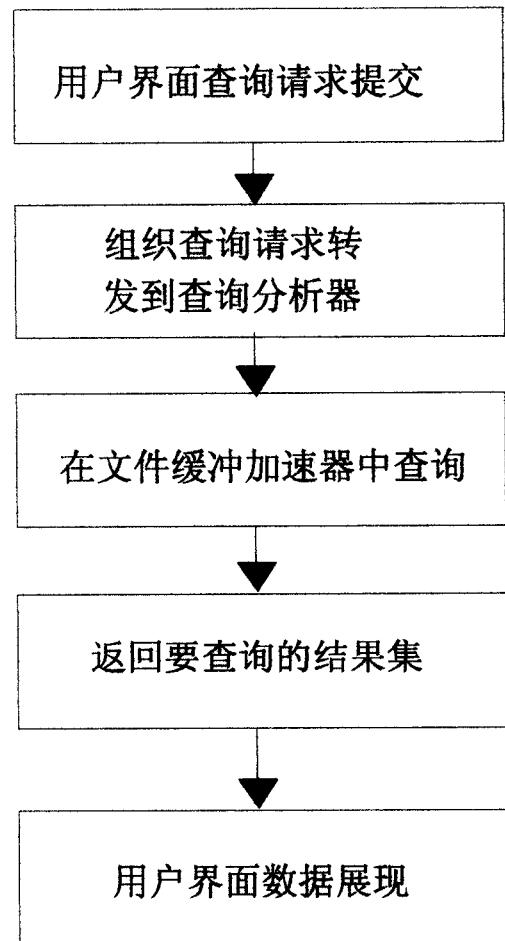
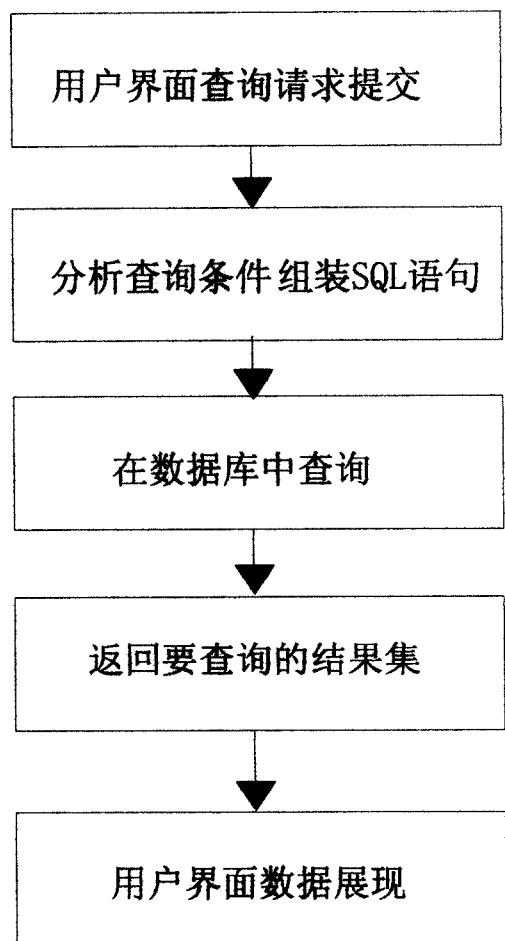


图3

图4