



(12) 发明专利

(10) 授权公告号 CN 112463832 B

(45) 授权公告日 2022. 10. 25

(21) 申请号 202011365011.1

G06F 16/248 (2019.01)

(22) 申请日 2020.11.27

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 106326429 A, 2017.01.11

申请公布号 CN 112463832 A

US 2010/0287183 A1, 2010.11.11

(43) 申请公布日 2021.03.09

赵璞 等.《高性能分布式搜索引擎Solr的研究与实现》.《电子科技》.2015,

(73) 专利权人 苏州浪潮智能科技有限公司

Ali Y. Aldailamy 等.《Performance

地址 215000 江苏省苏州市吴中经济开发

Evaluation of Distributed Indexing Using

区郭巷街道官浦路1号9幢

Solr and Terrier Information Retrievals》.

(72) 发明人 申胜利 侯满

《2018 Fourth International Conference on

(74) 专利代理机构 北京众达德权知识产权代理

Information Retrieval and Knowledge

有限公司 11570

Management (CAMP)》.2018,

专利代理师 梁凯

审查员 张杨悦

(51) Int. Cl.

G06F 16/2455 (2019.01)

G06F 16/22 (2019.01)

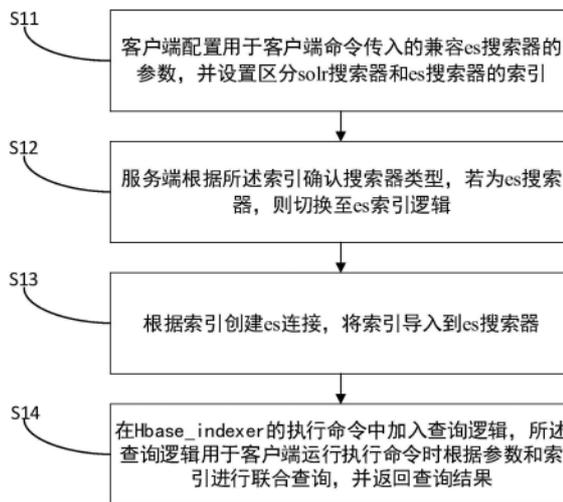
权利要求书2页 说明书7页 附图3页

(54) 发明名称

一种基于hbase-indexer的查询方法、装置及电子设备

(57) 摘要

本发明提供了一种基于hbase-indexer的查询方法、装置、电子设备及存储介质,属于服务器的技术领域,解决了现有技术中hbase-indexer无法将数据索引到es搜索器中的技术性问题。所述方法包括:客户端配置用于客户端命令传入的兼容es搜索器的参数,并设置区分solr搜索器和es搜索器的索引;服务端根据所述索引确认搜索器类型,若为es搜索器,则切换至es索引逻辑,根据索引创建es连接,将索引导入到es搜索器;在hbase-indexer的执行命令中加入查询逻辑,所述查询逻辑用于客户端运行联合查询命令时根据参数和索引进行联合查询,并返回查询结果。



1. 一种基于hbase-indexer的查询方法,其特征在于,应用于服务系统中,包括以下步骤:

客户端配置用于客户端命令传入的兼容es搜索器的参数,并设置区分solr搜索器和es搜索器的索引;

服务端根据所述索引确认搜索器类型,若为es搜索器,则切换至es索引逻辑;

根据索引创建es连接,将索引导入到es搜索器;

在hbase-indexer的执行命令中加入查询逻辑,所述查询逻辑用于客户端运行联合查询命令时根据参数和索引进行联合查询,并返回查询结果。

2. 根据权利要求1所述的基于hbase-indexer的查询方法,其特征在于,在客户端配置用于客户端命令传入的兼容es搜索器的参数,并设置区分solr搜索器和es搜索器的索引的步骤之后,还包括:

在客户端中加入索引校验过程和表权限校验过程。

3. 根据权利要求2所述的基于hbase-indexer的查询方法,其特征在于,所述根据索引创建es连接,将索引导入到es搜索器的步骤之前,还包括:

使用索引参数中的安全参数验证是否有权限;

若是,则通过验证;

若否,则生成异常信息。

4. 根据权利要求1所述的基于hbase-indexer的查询方法,其特征在于,根据索引创建es连接,将索引导入到es搜索器的步骤之后,还包括:

在批量导入历史记录索引的执行命令中添加兼容es的参数。

5. 根据权利要求1所述的基于hbase-indexer的查询方法,其特征在于,还包括:

基于ambari,添加服务控制脚本和自定义组件所需配置,使客户端可以在页面上安装hbase-indexer,并进行开启、停止、重启、删除的操作,以及自动修改参数。

6. 根据权利要求1所述的基于hbase-indexer的查询方法,其特征在于,根据参数和索引进行联合查询,并返回查询结果的步骤,具体包括:

接收查询命令;

获取hbase-indexer索引配置;

根据所述索引配置以及hbase表权限进行安全认证;

若通过认证,根据索引类型进行对应查询;

获取查询结果;

根据查询结果获取hbase的行键;

判断需要返回结果的类型,若需要返回结果,则使用hbase连接,并根据行键搜索结果并返回打印,若只需要返回行键,则返回行键并打印。

7. 根据权利要求6所述的基于hbase-indexer的查询方法,其特征在于,所述根据索引类型进行对应查询的步骤之前,还包括:

判断是否查询命令的参数为索引列且查询的格式正确;

若是,则将查询命令参数转换为对应搜索器中的查询语句。

8. 一种基于hbase-indexer的查询装置,其特征在于,包括:

客户端管理模块,客户端配置用于客户端命令传入的兼容es搜索器的参数,并设置区

分solr搜索器和es搜索器的索引；

服务端处理模块,服务端根据所述索引确认搜索器类型,若为es搜索器,则切换至es索引逻辑,根据索引创建es连接,将索引导入到es搜索器；

查询模块,在hbase-indexer的执行命令中加入查询逻辑,所述查询逻辑用于客户端运行执行命令时根据参数和索引进行联合查询,并返回查询结果。

9. 一种电子设备,包括存储器、处理器,所述存储器中存储有可在所述处理器上运行的计算机程序,其特征在于,所述处理器执行所述计算机程序时实现上述权利要求1至7任一项所述的方法的步骤。

10. 一种计算机可读存储介质,其特征在于,所述计算机可读存储介质存储有机器可运行指令,所述计算机可运行指令在被处理器调用和运行时,所述计算机可运行指令促使所述处理器运行所述权利要求1至7任一项所述的方法。

一种基于hbase-indexer的查询方法、装置及电子设备

技术领域

[0001] 本发明涉及服务器的技术领域,尤其是涉及一种基于hbase-indexer的查询方法、装置、电子设备及存储介质。

背景技术

[0002] Hbase-indexer是NGDATA公司为了将lily子系统里面相关HBase数据存储到Solr(solr是一个搜索服务器,一个开源的、基于Java的信息检索库)而开发的一个软件。HBase数据没有全文检索功能,只能在直接查询行键的场景下保持高效,不能兼容全文检索的功能。Hbase-indexer就是为了这种情况诞生的,但它只能将数据同步到solr中,只有同步模块没有查询模块,多租户安全模式下存在欠缺。目前很多场景更倾向于的使用ES(全称为elasticsearch,是一个搜索服务器,提供了一个分布式多用户能力的全文搜索引擎),因为ES相比slor更有优势。所以,目前Hbase-indexer功能有所欠缺。

发明内容

[0003] 本发明的目的在于提供一种基于hbase-indexer的查询方法、装置、电子设备及存储介质,缓解了现有技术中存在的hbase-indexer无法对es以及solr进行查询的缺陷的技术问题。

[0004] 第一方面,本发明提供的基于hbase-indexer的查询方法,包括以下步骤:

[0005] 客户端配置用于客户端命令传入的兼容es搜索器的参数,并设置区分solr搜索器和es搜索器的索引;

[0006] 服务端根据所述索引确认搜索器类型,若为es搜索器,则切换至es索引逻辑;

[0007] 根据索引创建es连接,将索引导入到es搜索器;

[0008] 在hbase-indexer的执行命令中加入查询逻辑,所述查询逻辑用于客户端运行联合查询命令时根据参数和索引进行联合查询,并返回查询结果。

[0009] 进一步的,所述在客户端中将用于客户端命令传入的参数兼容es搜索器,并设置索引代码区分solr搜索器和es搜索器的步骤之后,还包括:

[0010] 在客户端中加入索引校验过程和表权限校验过程。

[0011] 进一步的,所述根据索引创建es连接,将索引导入到es搜索器的步骤之前,还包括:

[0012] 使用索引参数中的安全参数验证是否有权限;

[0013] 若是,则通过验证;

[0014] 若否,则生成异常信息。

[0015] 进一步的,所述服务端使用es票据进行安全认证的步骤之后,还包括:

[0016] 在批量导入历史记录索引的执行命令中添加兼容es的参数。

[0017] 进一步的,所述基于hbase-indexer的查询方法,还包括:

[0018] 基于ambari(一种基于Web的工具,支持Apache Hadoop集群的供应、管理和监控),

添加服务控制脚本和自定义组件所需配置,使客户端可以在页面上安装hbase-indexer,并进行开启、停止、重启、删除的操作,以及自动修改参数。

[0019] 进一步的,所述根据查询参数和索引进行联合查询,并返回查询结果的步骤,具体包括:

[0020] 接收查询命令;

[0021] 获取hbase-indexer索引配置;

[0022] 根据所述索引配置以及hbase表权限进行安全认证;

[0023] 若通过认证,根据索引类型进行对应查询;

[0024] 获取查询结果;

[0025] 根据查询结果获取hbase的行键;

[0026] 判断需要返回结果的类型,若需要返回结果,则使用hbase连接,并根据行键搜索结果并返回打印,若只需要返回行键,则返回行键并打印。

[0027] 进一步的,所述根据索引类型进行对应查询的步骤之前,还包括:

[0028] 判断是否查询命令的参数为索引列且查询的格式正确;

[0029] 若是,则将查询命令参数转换为对应搜索器中的查询语句。

[0030] 第二方面,本发明还提供一种基于hbase-indexer的查询装置,包括:

[0031] 客户端管理模块,客户端配置用于客户端命令传入的兼容es搜索器的参数,并设置区分solr搜索器和es搜索器的索引;

[0032] 服务端处理模块,服务端根据所述索引确认搜索器类型,若为es搜索器,则切换至es索引逻辑,根据索引创建es连接,将索引导入到es搜索器;

[0033] 查询模块,在hbase-indexer的执行命令中加入查询逻辑,所述查询逻辑用于客户端运行执行命令时根据参数和索引进行联合查询,并返回查询结果。

[0034] 第三方面,本发明还提供一种电子设备,包括存储器、处理器,所述存储器中存储有可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现上述方法的步骤。

[0035] 第四方面,本发明还提供一种计算机可读存储介质,所述计算机可读存储介质存储有机可运行指令,所述计算机可运行指令在被处理器调用和运行时,所述计算机可运行指令促使所述处理器运行所述的方法。

[0036] 本发明提供的基于hbase-indexer的查询方法,通过客户端配置用于客户端命令传入的兼容es搜索器的参数,设置区分solr搜索器和es搜索器的索引,服务端根据所述索引确认搜索器类型并切换至对应的索引逻辑,根据索引创建连接,将索引导入到搜索器;在hbase-indexer的执行命令中加入查询逻辑,所述查询逻辑用于客户端运行执行命令时根据参数和索引进行联合查询,并返回查询结果;完成上述步骤,弥补了现有方案中hbase-indexer无法将数据索引到es搜索器中的缺陷,添加了对应搜索器和hbase的联合查询逻辑。

[0037] 相应地,本发明实施例提供的一种基于hbase-indexer的查询装置、电子设备及计算机可读存储介质,也同样具有上述技术效果。

附图说明

[0038] 为了更清楚地说明本发明具体实施方式或现有技术中的技术方案,下面将对具体实施方式或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图是本发明的一些实施方式,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0039] 图1为本发明实施例提供的基于hbase-indexer的查询方法流程图;

[0040] 图2为本发明实施例提供的基于hbase-indexer的查询流程图;

[0041] 图3为本发明实施例提供的基于hbase-indexer的查询装置示意图。

具体实施方式

[0042] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合附图对本发明的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0043] 本发明实施例中所提到的术语“包括”和“具有”以及它们的任何变形,意图在于覆盖不排他的包含。例如包含了一系列步骤或单元的过程、方法、系统、产品或设备没有限于已列出的步骤或单元,而是可选地还包括其他没有列出的步骤或单元,或可选地还包括对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0044] 如图1所示,本发明实施例提供的基于hbase-indexer的查询方法,包括以下步骤:

[0045] S11:客户端配置用于客户端命令传入的兼容es搜索器的参数,并设置区分solr搜索器和es搜索器的索引;

[0046] S12:服务端根据所述索引确认搜索器类型,若为es搜索器,则切换至es索引逻辑;

[0047] S13:根据索引创建es连接,将索引导入到es搜索器;

[0048] S14:在hbase-indexer的执行命令中加入查询逻辑,所述查询逻辑用于客户端运行联合查询命令时根据参数和索引进行联合查询,并返回查询结果。

[0049] 本发明提供的基于hbase-indexer的查询方法,通过客户端配置用于客户端命令传入的兼容es搜索器的参数,设置区分solr搜索器和es搜索器的索引,服务端根据所述索引确认搜索器类型并切换至对应的索引逻辑,根据索引创建连接,将索引导入到搜索器;在hbase-indexer的执行命令中加入查询逻辑,所述查询逻辑用于客户端运行联合查询命令时根据参数和索引进行联合查询,并返回查询结果;完成上述步骤,弥补了现有方案中hbase-indexer无法将数据索引到es搜索器中的缺陷,添加了对应搜索器和hbase的联合查询逻辑。

[0050] 在一种可能的实施方式中,所述在客户端中将用于客户端命令传入的参数兼容es搜索器,并设置索引代码区分solr搜索器和es搜索器的步骤之后,还包括:

[0051] 在客户端中加入索引校验过程和表权限校验过程。

[0052] 在一种可能的实施方式中,所述根据索引创建es连接,将索引导入到es搜索器的步骤之前,还包括:

[0053] 使用索引参数中的安全参数验证是否有权限,若是,根据索引创建es连接,将索引导入到es搜索器,若否,则生成异常信息。

[0054] 在进行查询之前,会进行身份以及权限等的校验,当校验失败则无法进行使用,保证系统的安全性。

[0055] 在一种可能的实施方式中,所述服务端使用es票据进行安全认证的步骤之后,还包括:

[0056] 在批量导入历史记录的索引的执行命令中添加兼容es的参数。

[0057] 创建索引之后,需要先将历史数据导入到es中,现有方案中是无法将历史记录导入到es中的,通过添加对应参数,使得历史记录可以顺利导入到es中。

[0058] 在一种可能的实施方式中,所述基于hbase-indexer的查询方法,还包括:

[0059] 基于ambari,添加服务控制脚本和自定义组件所需配置,使客户端可以在页面上安装hbase-indexer,并进行开启、停止、重启、删除的操作,以及自动修改参数。

[0060] 如图2所示,在一种可能的实施方式中,所述根据查询参数和索引进行联合查询,并返回查询结果的步骤,具体包括:

[0061] S21:接收查询命令;

[0062] S22:获取hbase-indexer索引配置;

[0063] S23:根据所述索引配置以及hbase表权限进行安全认证,若通过认证,则执行步骤S24,若未通过验证,则执行步骤S210;

[0064] S24:根据索引类型进行对应查询;

[0065] S25:获取查询结果;

[0066] S26:根据查询结果获取hbase的行键;

[0067] S27:判断需要返回结果的类型,若需要返回结果,则执行步骤S28,若只需要返回行键,则执行步骤S29;

[0068] S28:使用hbase连接,并根据行键搜索结果并返回打印;

[0069] S29:返回行键并打印;

[0070] S210:生成异常信息。

[0071] 如图3所示,在一种可能的实施方式中,所述根据索引类型进行对应查询的步骤之前,还包括:

[0072] S240:判断是否查询命令的参数为索引列且查询的格式正确,若是,则执行步骤S241,若否,则执行步骤S242;

[0073] S241:将查询命令参数转换为对应搜索器中的查询语句。

[0074] S242:向客户端发送报错信息。

[0075] 在查询之前,会对搜索的格式进行确定,当格式不正确时无法进行查询,保证查询的有效性。

[0076] 本发明实施例提供的基于hbase-indexer的查询方法,具体实施方式如下:

[0077] 由于hbase-indexer是为了solr设计的,客户端命令里能够传递的参数为solr需要的,如果切换为ES,需要将客户端传入的参数兼容ES的需要,并用“--connection-type-t”来区分solr和ES的索引器。

[0078] 同时,为了适应多租户安全场景的需要,使用客户端缓存的票据,索引的操作执行前验证缓存的票据是否拥有hbase表权限,有则进入下一步,如果没有直接失败。

[0079] 在创建索引时,需要添加ES安全方面所需的参数,据此验证连接ES,并能通过参数

控制索引的副本数和分片数。判断ES中是否存在该索引,存在则失败,不存在则创建;删除索引时删除ES中的索引。

[0080] 根据参数“--connection-type-t”的不同,如果为ES,则进入一种与solr平行逻辑,根据索引的参数创建ES连接,将导入到solr的流程全部换为导入到ES的流程,并将执行出错的行键加入到对应数据结构中。

[0081] 在安全场景下hbase方面的连接使用hbase服务端所用权限对应的票据,和ES的连接使用索引参数中的安全的参数,如果ES非安全模式和安全模式转换,需要客户端更新参数。

[0082] 在批量执行的模块,添加ES所需的参数和索引器类型区分,并改造处理逻辑,导入方式增加ES的导入方式,根据“--connection-type-t”参数选择为ES的导入方式。

[0083] 根据传入的参数或者从zookeeper(一个分布式应用程序协调服务,为分布式应用提供一致性服务,提供配置维护、域名服务、分布式同步、组服务等功能,hbase-indexer的索引参数维护在zookeeper中)中获取的索引配置进行ES连接的安全认证。

[0084] 在hbase-indexer的执行操作里添加query(询问)的选项,并指向查询操作的执行者。

[0085] 根据传入的参数,进入查询逻辑,如果indexer(索引)存在,则由此获取到zookeeper上的配置信息,判断校验查询参数是否为索引列及格式是否正确,转换为对应搜索器中的查询语句,如果多个查询条件,行式索引则为与关系,列式索引为或的关系,并根据传参判断查询方式是否为默认的全文检索方式,如果是精确查询,则条件的查询关系为精确查询。

[0086] 在查询对应搜索器的返回结果中获取hbase的行键,如果为列式索引则需要转换,如果配置有rowkey(hbase-indexer创建索引使用的xml配置文件里的一个属性)在索引中则直接取到该字段。根据参数判断如果需要只返回hbase的行键,则到此结束返回行键。如果需要返回结果,则还需要使用hbase连接,根据行键批量搜索并返回打印。

[0087] 安全模式下使用客户端缓存的票据,校验hbase权限,使用获取到的索引配置里的ES安全配置去认证创建连接并查询。

[0088] 基于ambari,添加服务控制脚本和各种自定义组件所需的配置,使hbase和ES或solr服务都安装好的情况下,才可以安装hbase-indexer,并且在安装后,自动修改hbase需要改动的配置如“hbase.replication=true”等。

[0089] 在开启安全的情况下,自动修改相关的env配置如“-Djava.security.auth.login.config=***”,能够通过页面,进行开启停止重启删除等管理操作。

[0090] 如图3所示,本发明实施例还提供一种基于hbase-indexer的查询装置,包括:

[0091] 客户端管理模块1,客户端配置用于客户端命令传入的兼容es搜索器的参数,并设置区分solr搜索器和es搜索器的索引;

[0092] 服务端处理模块2,服务端根据所述索引确认搜索器类型,若为es搜索器,则切换至es索引逻辑,根据索引创建es连接,将索引导入到es搜索器;

[0093] 查询模块3,在hbase-indexer的执行命令中加入查询逻辑,所述查询逻辑用于客户端运行执行命令时根据参数和索引进行联合查询,并返回查询结果。

[0094] 对应于上述方法,本发明实施例还提供了一种电子设备,包括存储器、处理器,所

述存储器中存储有可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现上述方法的步骤。

[0095] 对应于上述方法,本发明实施例还提供了一种计算机可读存储介质,所述计算机可读存储介质存储有机器可运行指令,所述计算机可运行指令在被处理器调用和运行时,所述计算机可运行指令促使所述处理器运行上述方法的步骤。

[0096] 本发明实施例所提供的装置可以为设备上的特定硬件或者安装于设备上的软件或固件等。本发明实施例所提供的装置,其实现原理及产生的技术效果和前述方法实施例相同,为简要描述,装置实施例部分未提及之处,可参考前述方法实施例中相应内容。所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,前述描述的系统、装置和单元的具体工作过程,均可以参考上述方法实施例中的对应过程,在此不再赘述。

[0097] 在本发明所提供的几个实施例中,应该理解到,所揭露的装置和方法,也可以通过其它的方式实现。以上所描述的装置实施例仅仅是示意性的,例如,附图中的流程图和框图显示了根据本发明的多个实施例的装置、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段或代码的一部分,所述模块、程序段或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现方式中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个连续的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意的,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或动作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0098] 又例如,所述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,再例如,多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些通信接口,装置或单元的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0099] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0100] 另外,在本发明提供的实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。

[0101] 所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(Read-Only Memory,简称ROM)、随机存取存储器(Random Access Memory,简称RAM)、磁碟或者光盘等各种可以存储程序代码的介质。

[0102] 最后应说明的是:以上所述实施例,仅为本发明的具体实施方式,用以说明本发明

的技术方案,而非对其限制,本发明的保护范围并不局限于此,尽管参照前述实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,其依然可以对前述实施例所记载的技术方案进行修改或可轻易想到变化,或者对其中部分技术特征进行等同替换;而这些修改、变化或者替换,并不使相应技术方案的本质脱离本发明实施例技术方案的范围。都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应以权利要求的保护范围为准。

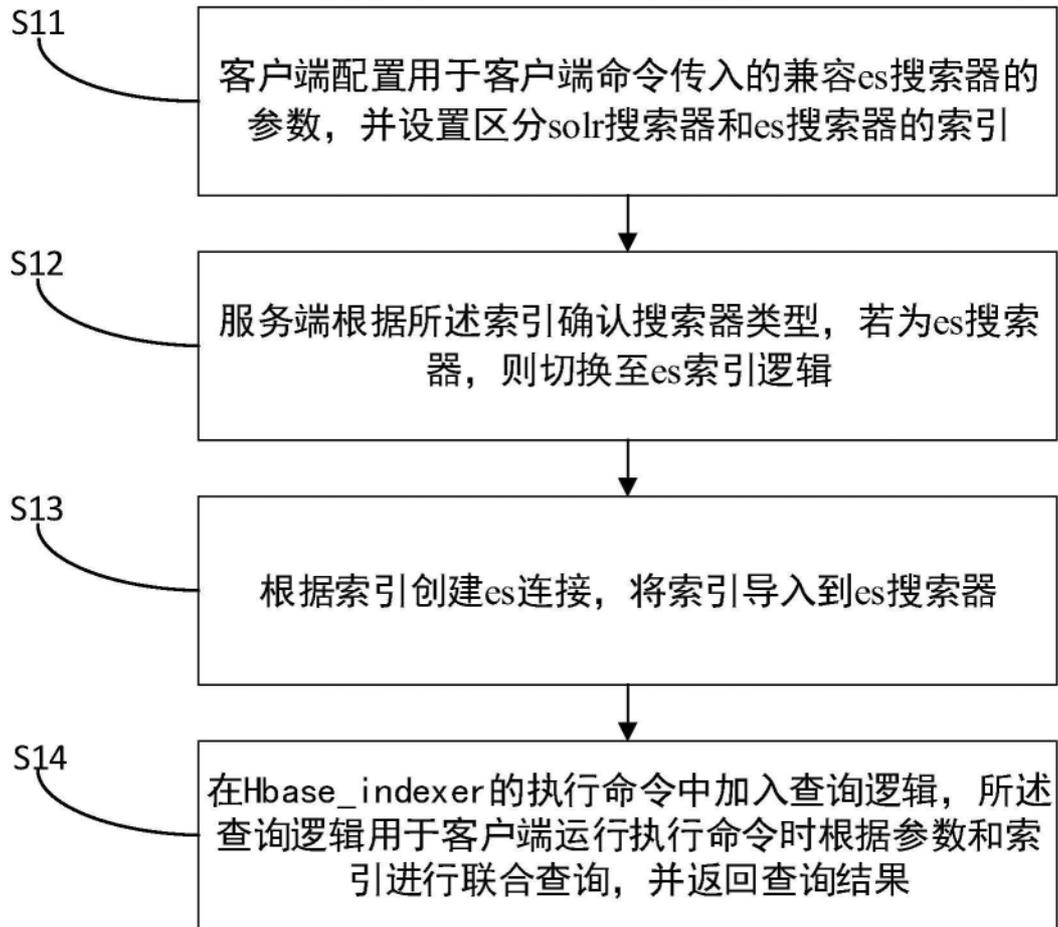


图1

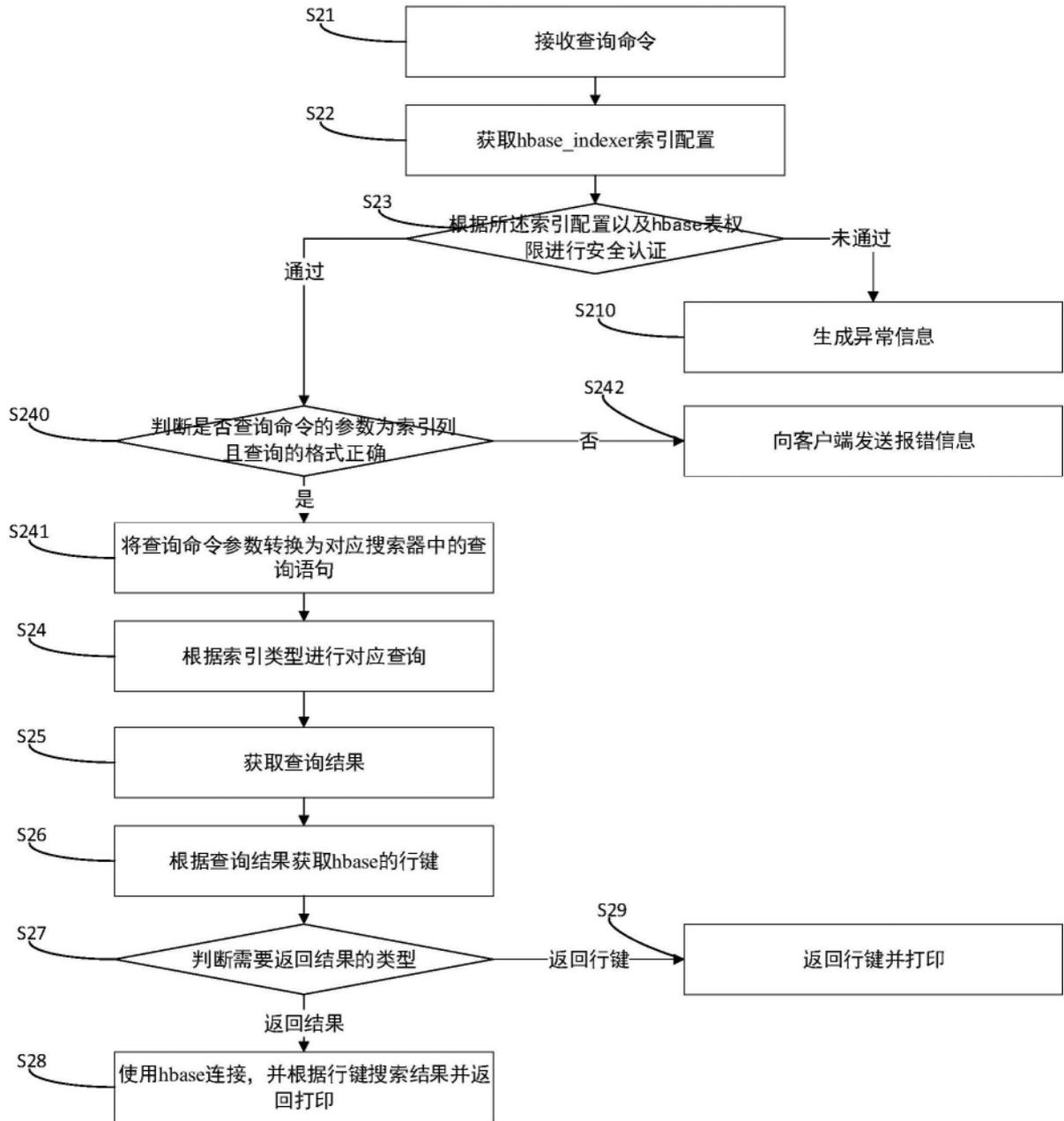


图2

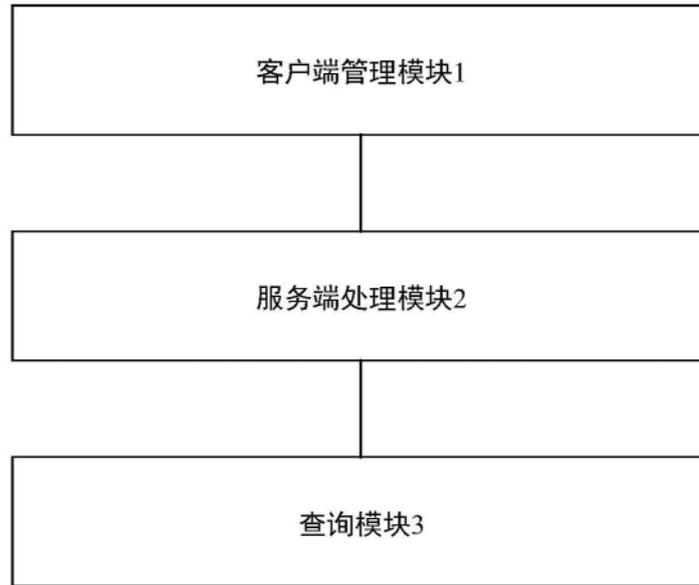


图3