



(12)发明专利申请

(10)申请公布号 CN 107544999 A

(43)申请公布日 2018.01.05

(21)申请号 201610487175.9

(22)申请日 2016.06.28

(71)申请人 百度在线网络技术(北京)有限公司

地址 100085 北京市海淀区上地十街10号

百度大厦三层

(72)发明人 马振 冯咀志 吴鹏

(74)专利代理机构 北京品源专利代理有限公司

11332

代理人 孟金喆 高磊

(51)Int.Cl.

G06F 17/30(2006.01)

G06F 21/62(2013.01)

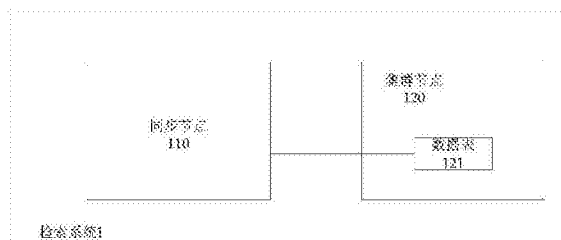
权利要求书4页 说明书9页 附图5页

(54)发明名称

用于检索系统的同步装置及同步方法、检索系统与方法

(57)摘要

本发明公开了一种用于检索系统的同步装置及同步方法、检索系统与方法。检索系统包括同步节点和集群节点,同步节点独立于集群节点且包括一个或多个同步装置,该同步装置包括:至少一个同步触发器,用于在监测到满足同步触发条件时,产生包含待同步数据的数据标识信息的同步触发指示;数据抓取器,用于根据同步触发指示从对应的数据源抓取与所述数据标识信息对应的数据并传输给转储器;转储器,用于将所抓取的数据转储至集群节点中的对应数据表中,集群节点用于存储数据表并提供检索服务。本发明在升级或数据转移时,不需要重启集群节点,满足实际检索需求,提升了检索效率,具有高可用和高并发的特点。



1. 一种应用于检索系统的同步装置,其特征在于,包括:

至少一个同步触发器,用于在监测到满足同步触发条件时,产生同步触发指示,所述同步触发指示包含待同步数据的数据标识信息;

数据抓取器,用于根据所述同步触发指示,从对应的数据源抓取与所述数据标识信息对应的数据,并将所抓取的数据传输给转储器;

转储器,用于将所抓取的数据转储至集群节点中的对应数据表中,所述集群节点,用于存储数据表并提供检索服务。

2. 根据权利要求1所述的同步装置,其特征在于,还包括:

配置模块,用于供用户提供所述同步触发器的配置信息、所述数据抓取器的配置信息和所述转储器的配置信息,

其中,所述待同步数据的数据标识信息是基于所述同步触发器的配置信息以及所述同步触发器监听到的信息获取的,所述数据抓取器和数据源的对应关系以及所述转储器和集群节点中的数据表的对应关系分别基于所述数据抓取器的配置信息和所述转储器的配置信息获取。

3. 根据权利要求1或2所述的同步装置,其特征在于,所述同步触发器包括定时触发器、消息触发器和日志触发器中的至少一种;

以及,在监测到满足同步触发条件时,产生同步触发指示,包括下述至少一种情况:

在通过所述定时触发器监测到当前时间满足设定时刻,或到达设定周期时,产生所述同步触发指示;

在通过所述消息触发器监测到对数据源的操作消息中产生设定消息时,产生所述同步触发指示;

在通过所述日志触发器获取并分析数据源的操作日志,监测到该数据源有数据更新时,产生所述同步触发指示。

4. 根据权利要求1-3中任一所述的同步装置,其特征在于,还包括:

调度器,用于在接收到所述同步触发器发送的至少一个同步触发指示和/或所述数据抓取器抓取的数据后,根据所述调度器中的调度策略,分别调度所述数据抓取器来获取对应的数据和/或调度所述转储器将所抓取的数据存储到对应的数据表中。

5. 根据权利要求4所述的同步装置,其特征在于,所述调度器具体用于:

在接收到所述同步触发器发送的至少一个同步触发指示后,将所接收的至少一个同步触发指示作为任务放入任务池中;

根据调度器中的调度策略,获取待分配给所述数据抓取器的同步触发指示;

将所获取的待分配同步触发指示分配给所述数据抓取器来抓取对应的数据。

6. 根据权利要求5所述的同步装置,其特征在于,所述调度策略包括:最大分配同步触发指示数量、最大转储数据数量以及同步触发指示分配机制。

7. 根据权利要求6所述的同步装置,其特征在于,所述同步触发指示分配机制包括:

优先级分配机制;和/或,

资源节约分配机制。

8. 根据权利要求4所述的同步装置,其特征在于,所述配置模块还用于:

配置所述调度器中的调度策略。

9. 一种检索系统,其特征在于,包括集群节点和同步节点,所述同步节点独立于所述集群节点,包括一个或多个根据权利要求1到8中任一所述的同步装置,所述集群节点用于存储数据表并提供检索服务。

10. 根据权利要求8所述的检索系统,其特征在于,还包括:

权限控制模块,配置在所述集群节点中,用于根据所接收的用户检索请求中的账户标识、操作内容和操作对象以及权限配置表,判断对应于所述检索请求的账户是否具备操作权限,其中所述权限配置表存储账户标识、操作内容和操作对象的映射关系;

若具备操作权限,则将所述检索请求传输给集群节点的检索应用程序接口,在集群节点的数据表中搜索匹配的结果;若不具备操作权限,则屏蔽所述检索请求。

11. 根据权利要求10所述的检索系统,其特征在于,所述账户标识包括应用名称和账户密码,所述操作对象包括数据库索引信息和数据表信息,所述操作内容包括操作权限,以及所述权限配置表包括应用名称、账户密码、数据库索引信息、数据表信息和操作权限的映射关系。

12. 根据权利要求10或11所述的检索系统,其特征在于,还包括:

配置平台,用于提供基于WEB方式的人机交互配置界面,供用户输入配置信息来配置所述权限配置表。

13. 根据权利要求9到12中任一所述的检索系统,其特征在于,所述集群节点是ElasticSearch集群节点。

14. 一种检索系统,其特征在于,包括集群节点和同步节点,所述集群节点用于存储数据表并提供检索服务,所述同步节点独立于所述集群节点,用于同步所述集群节点中的数据表,所述集群节点还包括:权限控制模块,用于在接收到用户检索请求后,根据所接收的用户检索请求中的账户标识、操作内容和操作对象以及权限配置表,判断对应于所述检索请求的账户是否具备操作权限,其中所述权限配置表存储账户标识、操作内容、操作对象和操作权限的映射关系;

若具备操作权限,则将所述检索请求传输给集群节点的检索应用程序接口,在集群节点的数据表中搜索匹配的结果;若不具备操作权限,则屏蔽所述检索请求。

15. 一种应用于检索系统的同步方法,其特征在于,包括:

在通过至少一个同步触发器监测到满足同步触发条件时,产生同步触发指示,所述同步触发指示包含待同步数据的数据标识信息;

根据所述同步触发指示,通过数据抓取器从对应的数据源抓取与所述数据标识信息对应的数据,并将所抓取的数据传输给转储器;

通过所述转储器将所抓取的数据转储至集群节点中的对应数据表中,所述集群节点,用于存储数据表并提供检索服务。

16. 根据权利要求15所述的同步方法,其特征在于,所述方法还包括:

接收用户提供的所述同步触发器的配置信息、所述数据抓取器的配置信息和所述转储器的配置信息,

其中,所述待同步数据的数据标识信息是基于所述同步触发器的配置信息以及所述同步触发器监听到的信息获取的,所述数据抓取器和数据源的对应关系以及所述转储器和集群节点中的数据表的对应关系分别基于所述同步触发器的配置信息、所述数据抓取器的配

置信息和所述转储器的配置信息获取。

17. 根据权利要求15或16所述的同步方法, 其特征在于, 所述同步触发器包括定时触发器、消息触发器和日志触发器中的至少一种;

以及, 在通过至少一个同步触发器监测到满足同步触发条件时, 产生同步触发指示, 包括下述至少一种情况:

在通过所述定时触发器监测到当前时间满足设定时刻, 或到达设定周期时, 产生所述同步触发指示;

在通过所述消息触发器监测到对数据源的操作消息中产生设定消息时, 产生所述同步触发指示;

在通过所述日志触发器获取并分析数据源的操作日志, 监测到该数据源有数据更新时, 产生所述同步触发指示。

18. 根据权利要求15所述的同步方法, 其特征在于, 还包括:

在接收到所述同步触发器发送的至少一个同步触发指示和/或所述数据抓取器抓取的数据后, 根据调度器中的调度策略, 通过所述调度器分别调度所述数据抓取器来获取对应的数据和/或调度所述转储器将所抓取的数据存储到对应的数据表中。

19. 根据权利要求18所述的同步方法, 其特征在于, 在接收到所述同步触发器发送的至少一个同步触发指示后, 根据调度器中的调度策略, 通过所述调度器调度所述数据抓取器来获取对应的数据包括:

在接收到所述同步触发器发送的至少一个同步触发指示后, 将所接收的至少一个同步触发指示作为任务放入任务池中;

根据调度器中的调度策略, 获取待分配给所述数据抓取器的同步触发指示;

将所获取的待分配同步触发指示分配给所述数据抓取器来抓取对应的数据。

20. 根据权利要求18所述的同步方法, 其特征在于, 所述调度策略包括: 最大分配同步触发指示数量、最大转储数据数量以及同步触发指示分配机制。

21. 根据权利要求20所述的同步方法, 其特征在于, 所述同步触发指示分配机制包括:

优先级分配机制; 和/或,

资源节约分配机制。

22. 根据权利要求18所述的同步方法, 其特征在于, 所述调度策略是基于用户提供的配置信息来配置的。

23. 一种检索方法, 其特征在于, 所述方法包括:

在接收到用户检索请求后, 在集群节点的数据表中搜索匹配的结果,

其中, 所述数据表利用同步装置根据权利要求15到22中任一所述的同步方法进行同步, 所述同步装置独立于所述集群节点。

24. 根据权利要求23所述的检索方法, 其特征在于, 所述方法还包括:

在接收到用户检索请求后, 根据所接收的用户检索请求中的账户标识、操作内容和操作对象以及权限配置表, 判断对应于所述检索请求的账户是否具备操作权限, 其中所述权限配置表存储账户标识、操作内容、操作对象和操作权限的映射关系;

若具备操作权限, 则将所述检索请求传输给集群节点的检索应用程序接口, 在集群节点的数据表中搜索匹配的结果; 若不具备操作权限, 则屏蔽所述检索请求。

25.一种检索方法,其特征在于,所述方法包括:

在接收到用户检索请求后,根据所接收的用户检索请求中的账户标识、操作内容和操作对象以及权限配置表,判断对应于所述检索请求的账户是否具备操作权限,其中所述权限配置表存储账户标识、操作内容、操作对象和操作权限的映射关系;

若具备操作权限,则将所述检索请求传输给集群节点的检索应用程序接口,在集群节点的数据表中搜索匹配的结果;

若不具备操作权限,则屏蔽所述检索请求。

26.根据权利要求25所述的检索方法,其特征在于,所述数据表利用同步装置根据权利要求15到22中任一所述的同步方法进行同步,所述同步装置独立于所述集群节点。

用于检索系统的同步装置及同步方法、检索系统及方法

技术领域

[0001] 本发明实施例涉及检索技术,尤其涉及一种用于检索系统的同步装置及同步方法、检索系统及方法。

背景技术

[0002] 现有很多应用场景都需要为用户提供检索服务,例如,糯米运营行为,是基于门店展开的,很多应用程序都需要对门店的信息进行检索,包括多字段查询和排序,对门店名称和品牌名称的中文检索,以及对地理坐标的距离检索等。

[0003] 目前,提供检索服务的检索系统有多种实现方式,例如,采用Elasticsearch(简称ES)、River和Elasticsearch-jetty等工具或组件。

[0004] 以ES为例,ES的检索架构是以多个集群节点来提供检索服务的。每个集群节点用于提供存储数据的数据库,数据库由数据表(Type)组成。数据库设置有索引(Index),集群节点能够基于Index,响应用户的数据检索请求,在数据库和数据表中查询到相应的数据,并反馈给用户。在数据表中,可以由不同类型的数据源来提供数据,数据源的种类包括mysql,mongoDB和rabitMQ,数据表通过调用这些数据源来获取数据。

[0005] 在检索服务中涉及到两个主要的环节,一是数据同步,二是用户的数据操作(可包括读、写、查询等)。现有检索系统针对这两个环节提供的操作都有一些弊端。

[0006] 对于数据同步环节,基于ES系统实现的数据同步,需要针对不同类型的数据源提供不同的同步方式,且实现同步操作的程序以插件的方式安装在集群节点上,这就导致存在两个缺陷:1、对不同的数据源需要单独编制不同的同步插件;2、想运行同步插件时,必须停止集群节点的工作,安装同步插件之后再启动运行。综上所述,若以插件形式加入新的功能,如升级,则需要重启集群,导致服务不够稳定。

[0007] 同时,对于用户的数据操作环节,ES系统的集群节点是无法进行权限控制的,即,任何发送给集群节点的用户数据操作指示,都会被执行。且在集群节点内并没有对不同用户的数据进行物理隔离存储,容易出现误操作或恶意操作的问题。

发明内容

[0008] 本发明提供一种用于检索系统的同步装置及同步方法、检索系统及方法,以实现用户可根据不同设计需求灵活配置同步策略,得到与设计需求匹配的检索系统。

[0009] 根据本发明的一个方面,提供了一种应用于检索系统的同步装置,包括:至少一个同步触发器,用于在监测到满足同步触发条件时,产生同步触发指示,所述同步触发指示包含待同步数据的数据标识信息;数据抓取器,用于根据所述同步触发指示,从对应的数据源抓取与所述数据标识信息对应的数据,并将所抓取的数据传输给转储器;转储器,用于将所抓取的数据转储至集群节点中的对应数据表中,所述集群节点,用于存储数据表并提供检索服务。

[0010] 根据本发明的另一方面,提供了一种检索系统,包括集群节点和同步节点,所述同

步节点独立于所述集群节点,包括一个或多个如上所述的同步装置,所述集群节点用于存储数据表并提供检索服务。

[0011] 根据本发明的另一方面,提供了一种检索系统,包括集群节点和同步节点,所述集群节点用于存储数据表并提供检索服务,所述同步节点独立于所述集群节点,用于同步所述集群节点中的数据表,所述集群节点还包括:权限控制模块,用于在接收到用户检索请求后,根据所接收的用户检索请求中的账户标识、操作内容和操作对象以及权限配置表,判断对应于所述检索请求的账户是否具备操作权限,其中所述权限配置表存储账户标识、操作内容、操作对象和操作权限的映射关系;若具备操作权限,则将所述检索请求传输给集群节点的检索应用程序接口,在集群节点的数据表中搜索匹配的结果;若不具备操作权限,则屏蔽所述检索请求。

[0012] 根据本发明的另一方面,提供了一种应用于检索系统的同步方法,包括:在通过至少一个同步触发器监测到满足同步触发条件时,产生同步触发指示,所述同步触发指示包含待同步数据的数据标识信息;根据所述同步触发指示,通过数据抓取器从数据源抓取与所述数据标识信息对应的数据,并将所抓取的数据传输给转储器;通过所述转储器将所抓取的数据转储至集群节点中的对应数据表中,所述集群节点,用于存储数据表并提供检索服务。

[0013] 根据本发明的另一方面,提供了一种检索方法,包括:在接收到用户检索请求后,在集群节点的数据表中搜索匹配的结果,其中,所述数据表利用同步装置根据如上所述的同步方法进行同步,所述同步装置独立于所述集群节点。

[0014] 根据本发明的另一方面,提供了一种检索方法,包括:在接收到用户检索请求后,根据所接收的用户检索请求中的账户标识、操作内容和操作对象以及权限配置表,判断对应于所述检索请求的账户是否具备操作权限,其中所述权限配置表存储账户标识、操作内容、操作对象和操作权限的映射关系;若具备操作权限,则将所述检索请求传输给集群节点的检索应用程序接口,在集群节点的数据表中搜索匹配的结果;若不具备操作权限,则屏蔽所述检索请求。

[0015] 本发明通过将同步节点独立于集群节点设置,同步节点的安裝或更新均无需重启集群节点,提高了服务的稳定性。根据不同的设计需求配置不同的同步触发条件,产生不同的同步触发指示,以控制数据抓取器和数据转储器执行不同的数据同步策略,满足实际检索需求。经试验在相同的条件下,本发明中单个同步触发器实时从数据源同步100条数据的同步时间小于1秒,单个同步触发器批量同步1000万条数据的同步时间小于1小时,检索系统的响应时间小于500毫秒,提升了检索效率,具有高可用、高并发以及多同步方式的特点。

附图说明

[0016] 图1示出了根据本发明实施例一的一种检索系统的结构示意图;

[0017] 图2示出了根据本发明实施例一的同步装置的一个示例的结构示意图;

[0018] 图3示出了本发明实施例一的用于检索系统的同步方法的一个示例的流程图;

[0019] 图4示出了根据本发明实施例一的同步装置的另一示例的结构示意图;

[0020] 图5示出了根据本发明实施例一的用于检索系统的同步方法的另一示例的流程图;

- [0021] 图6示出了根据本发明实施例一的检索方法的流程图；
- [0022] 图7A-图7D示出了根据本发明的实施例的检索方法的一个实现示例的示意图；
- [0023] 图8示出了根据本发明实施例二的检索系统的结构示意图；
- [0024] 图9示出了根据本发明实施例二的检索方法的一个示例的流程图。

具体实施方式

[0025] 下面结合附图和实施例对本发明作进一步的详细说明。可以理解的是,此处所描述的具体实施例仅仅用于解释本发明,而非对本发明的限定。另外还需要说明的是,为了便于描述,附图中仅示出了与本发明相关的部分而非全部结构。

[0026] 实施例一

[0027] 图1为本发明实施例一提供的检索系统1的结构示意图,本实施例可适用于采用用户配置的同步触发器、生成同步触发指示,根据不同的同步策略将数据同步至设定集群的情况。

[0028] 如图1所示,检索系统1包括同步节点110和集群节点120,所述同步节点110独立于集群节点120。所述集群节点120存储有数据表121,用于在接收到用户发出的检索请求后,在数据表121中进行检索并提供检索结果。所述同步节点110包括一个或多个同步装置,所述同步装置用于对集群节点120中的数据表121进行同步操作。

[0029] 集群节点120为存储数据表121的服务器集群。由至少两个集群节点120构成集群,以分布式的方式构建集群,集群具有分片存储、多副本、负载均衡和无中心的特点,便于集群扩展。例如,可以采用Elasticsearch集群,由于Elasticsearch集群本身具有热备、横向扩展和无中心等特点,因此,可以不用增加负载均衡和热备策略。

[0030] 同步节点110为独立于集群节点120设置的控制终端,该控制终端可以与集群进行数据传输,以将数据由数据源同步至集群。同步节点110中可以创建一个或多个同步装置执行数据同步操作。在其中一个同步装置升级时,其它同步装置111可以继续提供服务,集群节点120不需要停机维护。

[0031] 图2示出了根据本发明实施例一的同步装置的一个示例的结构示意图。如图2所示,同步装置包括至少一个同步触发器111-1,111-2,……,111-n,数据抓取器112和转储器113。

[0032] 同步触发器111用于在监测到满足同步触发条件时,产生同步触发指示,所述同步触发指示包含待同步数据的数据标识信息,并且将所产生的同步触发指示发送给数据抓取器112。在本发明的一个示例中,在同步触发器111监听到信息后,同步触发器111可以基于所述同步触发器111的配置信息以及所述同步触发器111监听到的信息获取所述待同步数据的数据标识信息。

[0033] 在本发明的示例中,所述同步触发器111可以包括定时触发器、消息触发器和日志触发器中的至少一种。定时触发器为在当前时间满足设定时刻,或到达设定周期时,产生将设定数据源的待同步数据同步至目标数据表121的同步触发指示的功能模块。消息触发器为在对数据源的操作消息中产生设定消息时,产生将设定数据源的待同步数据同步至目标数据表121的同步触发指示的功能模块。日志触发器为在数据源的操作日志中包括该数据源的数据更新时,产生将设定数据源的数据同步至目标数据表121中的同步触发指示的功

能模块。

[0034] 所述定时触发器的配置消息可以包括触发周期和要监听的数据标识字段信息,所述触发周期比如可以是设定时刻或设定周期,所述数据标识字段信息可以是要监听的数据的唯一标识信息,比如数据ID等。所述消息触发器的配置消息可以包括消息集群,消息队列名,消息token,所述配置消息用于配置用来接收消息的管道信息。所述日志触发器的配置消息可以包括databus源,要监听的日志数据信息。

[0035] 在监测到满足同步触发条件时,产生同步触发指示,包括下述至少一种情况:

[0036] (1)在通过所述定时触发器监测到当前时间满足设定时刻,或到达设定周期时,基于哪些数据需要进行同步的监听信息,产生包含所述需要进行同步的数据的唯一标识信息的同步触发指示;

[0037] (2)在通过所述消息触发器监测到对数据源的操作消息中产生设定消息时,从所述设定消息中提取出需要进行同步的数据的唯一标识信息,由此产生包含所述需要进行同步的数据的唯一标识信息的同步触发指示;

[0038] (3)在通过所述日志触发器获取并分析数据源的操作日志,监测到该数据源有数据更新时,产生包含所述需要进行同步的数据的唯一标识信息的同步触发指示。

[0039] 在接收到同步触发指示后,数据抓取器112根据所述同步触发指示,从对应的数据源抓取与所述数据标识信息对应的数据,并将所抓取的数据传输给转储器113,比如转储器113的转储队列中。抓取方式可以采用分布式服务框架Dubbo方式或SQL方式。抓取的策略可以是批量抓取或限时抓取。例如,预先设定一次抓取操作从数据源中抓取设定条数的记录。批量抓取为根据预先设定的一次抓取的记录条数,从数据源中抓取设定条数的记录,将所抓取的记录传输至转储器。同时,还可以预先设定一次抓取操作的执行时间。限时抓取为一次抓取的记录条数不满足设定条数,但本次抓取操作的执行时间达到预先设定的执行时间,则将所抓取的记录传输至转储器113。可以在需同步的数据量大于设定阈值时,采用批量抓取策略,减少网络和磁盘IO接口的长时间占用。在需要同步的数量小于设定阈值时,采用限时抓取策略,满足时效性要求。

[0040] 在本发明的一个示例中,所述数据抓取器和数据源的对应关系可以基于所述数据抓取器的配置信息获取。所述数据抓取器的配置信息可以包括对应的数据源地址信息。在本发明的一个示例中,所述数据抓取器可以包括SQL抓取器和Dubbo抓取器。在所述数据抓取器是SQL抓取器时,所述配置信息可以包括数据库地址。此外,优选地,所述配置信息还可以包括数据库用户名,数据库密码,监听语句。在所述数据抓取器是Dubbo抓取器时,所述配置信息可以包括远程服务地址。此外,优选地,所述配置信息还可以包括服务名,方法名,参数。

[0041] 转储器113用于将所抓取的数据转储至集群节点120中的对应数据表121中,以供集群节点120用来提供检索服务。在本发明的一个示例中,所述转储器113和集群节点120中的数据表121的对应关系可以基于所述转储器113的配置信息获取。所述转储器113的配置信息可以包括索引数据的数据结构。

[0042] 此外,优选地,同步装置还可以包括配置模块(未示出),用于供用户提供所述同步触发器的配置信息、所述数据抓取器的配置信息和所述转储器的配置信息。

[0043] 图3示出了本发明实施例一的用于检索系统的同步方法的一个示例的流程图。如

图3所示,在步骤S310,在通过至少一个同步触发器监测到满足同步触发条件时,同步触发器产生同步触发指示,所述同步触发指示包含待同步数据的数据标识信息,并将同步触发指示发送给数据抓取器。在接收到同步触发器发送的同步触发指示后,在步骤S320,数据抓取器根据所述同步触发指示,从对应的数据源抓取与所述数据标识信息对应的数据,并将所抓取的数据传输给转储器。接着,在步骤S330,转储器将所抓取的数据转储至集群节点中的对应数据表中,以供集群节点用来提供检索服务。

[0044] 优选地,在同步触发器产生同步触发指示之前,所述方法还可以包括:接收用户提供的所述同步触发器的配置信息、所述数据抓取器的配置信息和所述转储器的配置信息。所述同步触发器、数据抓取器和转储器的配置信息如上所述,在此不再赘述。

[0045] 在接收到上述配置信息后,可以基于所述同步触发器的配置信息以及所述同步触发器监听到的信息获取所述待同步数据的数据标识信息,可以基于所述数据抓取器的配置信息获取所述数据抓取器和数据源的对应关系,以及可以基于所述转储器的配置信息获取所述转储器和集群节点中的数据表的对应关系。

[0046] 图4示出了根据本发明实施例一的同步装置的另一示例的结构示意图。如图4所示,同步装置包括至少一个同步触发器111'-1,111'-2,……,111'-n,数据抓取器112'、转储器113'和调度器114。

[0047] 同步触发器111'用于在监测到满足同步触发条件时,产生同步触发指示,所述同步触发指示包含待同步数据的数据标识信息,并且将所产生的同步触发指示发送给调度器114。在接收到所述同步触发器发送的至少一个同步触发指示后,调度器114根据所述调度器中的调度策略,将同步触发指示发送给数据抓取器112'来获取对应的数据。

[0048] 数据抓取器112'根据所接收的同步触发指示,从对应的数据源抓取与所述数据标识信息对应的数据,并将所抓取的数据传输给调度器114。调度器114根据所述调度器中的调度策略,将所接收的抓取数据发送给转储器113',由转储器113'将所述抓取数据转储到对应的数据表中。

[0049] 在本发明的一个示例中,在接收到所述同步触发器发送的至少一个同步触发指示后,根据调度器中的调度策略,通过所述调度器调度所述数据抓取器来获取对应的数据可以包括:在接收到所述同步触发器发送的至少一个同步触发指示后,将所接收的至少一个同步触发指示作为任务放入任务池中;根据调度器中的调度策略,获取待分配给所述数据抓取器的同步触发指示;将所获取的待分配同步触发指示分配给所述数据抓取器来抓取对应的数据。

[0050] 在本发明的一个示例中,所述调度策略可以包括:最大分配同步触发指示数量、最大转储数据数量以及同步触发指示分配机制。这里,最大分配同步触发指示数量是指单次分配给数据抓取器的同步触发指示的最大数量,即数据抓取器单次能够处理的最大同步触发指示数量。通常而言,最大分配触发指示数量是指数据抓取器单次能够并发处理的最大线程数。最大转储数据数量是指转储器单次转储的数据的最大数量。通常而言,最大转储数据数量是指转储器单次能够并发处理的最大线程数。此外,优选地,所述同步触发指示分配机制可以包括:优先级分配机制;和/或资源节约分配机制。

[0051] 优先级分配机制是指调度器根据同步触发指示的优先级来确定待分配的同步触发指示,比如根据最大分配同步触发指示数量,按照从高到低的优先级来确定待分配的同

步触发指示。具体地,在接收到同步触发指示后,调度器会根据同步触发指示对应的触发器类型,为同步触发指示分配不同的优先级。比如,定时触发器触发的同步指示,优先级为普通;消息触发器和日志触发器触发的同步触发指示,优先级为优先。针对消息触发器和日志触发器触发的同步触发指示,按照触发的时间排序。

[0052] 资源节约分配机制是指将单个的同步指示,收集为一批同步指示,进行批量抓取和转储,由此降低抓取器和转储器的执行次数,进而降低CPU,网络IO和磁盘IO的消耗。

[0053] 优选地,在所述同步装置110'包括配置模块时,所述配置模块还可以用于配置所述调度器中的调度策略。

[0054] 图5示出了根据本发明实施例一的用于检索系统的同步方法的另一示例的流程图。

[0055] 如图5所示,在步骤S510,在通过至少一个同步触发器监测到满足同步触发条件时,同步触发器产生同步触发指示,所述同步触发指示包含待同步数据的数据标识信息,并且将所产生的同步触发指示发送给调度器。在接收到同步触发器发送的同步触发指示后,在步骤S520,在接收到所述同步触发器发送的至少一个同步触发指示后,调度器根据所述调度器中的调度策略,将同步触发指示发送给数据抓取器来获取对应的数据。接着,在步骤S530,数据抓取器根据所接收的同步触发指示,从对应的数据源抓取与所述数据标识信息对应的数据,并将所抓取的数据传输给调度器。随后,在步骤S540,调度器根据所述调度器中的调度策略,将所接收的抓取数据发送给转储器。然后,在步骤S550,转储器将所述抓取数据转储到对应的数据表中,以供集群节点用来提供检索服务。

[0056] 优选地,在同步触发器产生同步触发指示之前,所述同步方法还可以包括:接收用户提供的所述同步触发器的配置信息、所述数据抓取器的配置信息和所述转储器的配置信息以及所述调度器的调度策略。所述同步触发器、数据抓取器和转储器的配置信息以及所述调度器的调度策略如上所述,在此不再赘述。

[0057] 图6示出了根据本发明实施例一的检索方法的流程图,所述检索方法由图1中示出的检索系统执行。如图6所示,在步骤S610,接收用户输入的检索请求,随后,在步骤S620,基于所接收到检索请求,在集群节点的数据表中搜索匹配的结果,其中,所述数据表利用同步装置根据如上所述的同步方法进行同步,所述同步装置独立于所述集群节点。

[0058] 图7A到图7D示出了根据本发明的实施例的检索方法的一个实现示例的示意图。图7A示出了应用在web页面注册并获取token的示意图;图7B示出了应用在web页面创建本应用下的索引mapping和同步配置的示意图;图7C示出了系统自动完成检索数据的实时同步的示意图;图7D示出了应用通过api检索该应用下的索引数据的示意图。

[0059] 在该示例中,有检索需求的用户首先在RTS系统进行注册,获取app及token,如图7A所示。通过配置模块,指定数据同步的规则及索引的数据结构,如图7B所示。系统自动完成检索数据的同步,如图7C所示。用户可以通过ES原生的api进行数据检索,方便历史项目进行迁移,如图7D所示。

[0060] 综上可知,采用对同步节点110中的同步触发器、抓取器和转储器进行拆分设计的方式,由同步触发器负责指示需同步的数据,由抓取器抓取数据后传输至转储队列,由转储器从转储队列中获取数据,并同步至集群节点120的对应的数据表121中,实现同步节点110的数据无状态特性。因此,对同一批数据(同一个索引对应的数据)进行同步时,可以配置多

个同步策略。由于数据无状态特征,多个同步策略可以并发执行,不会因为同步策略执行的先后顺序和数据更新的先后顺序不同,导致同步数据版本错误。例如,对于同一个索引可以配置至少一个同步触发器,采用不同的同步触发条件,生成同步触发指示,使抓取器和转储器根据所述同步触发指示执行数据同步。由于数据无状态特性,不会因为同步策略执行的先后顺序和数据更新的先后顺序不同,导致同步数据版本错误。

[0061] 本实施例的技术方案,通过将同步节点独立于集群节点设置,同步节点的安装或更新均无需重启集群节点,提高了服务的稳定性。根据不同的设计需求配置不同的同步触发条件,产生不同的同步触发指示,以控制数据抓取器和数据转储器执行不同的数据同步策略,满足实际检索需求,提升了检索效率,具有高可用和高并发的特点。

[0062] 实施例二

[0063] 图8是本发明实施例二中的一种检索系统的结构示意图。本实施例的技术方案在上述实施例的基础上,进一步包括配置平台130和权限控制模块122。

[0064] 权限控制模块122配置在集群节点120中,用于根据所接收的用户检索请求中的账户标识、操作内容和操作对象以及权限配置表,判断对应于所述检索请求的账户是否具备操作权限,其中所述权限配置表存储账户标识、操作内容和操作对象的映射关系。若具备操作权限,则将所述检索请求传输给集群节点的检索应用程序接口,在集群节点的数据表中搜索匹配的结果;若不具备操作权限,则屏蔽所述检索请求。这样设置的好处在于对集群节点120内的数据执行分权控制,满足多用户协同工作时,数据的隔离与共享要求。

[0065] 优选地,所述账户标识可以包括应用名称和账户密码,所述操作对象包括数据库索引信息和数据表信息,所述操作内容包括操作权限,所述操作权限比如是读写权限。所述权限配置表包括应用名称、账户密码、数据库索引信息(即,Index信息)、数据表信息(即,Type信息)和操作权限的映射关系。例如,定义一个专用的index的type用来存储ACL映射规则。其中,ACL映射规则包括定义了应用名称(app)、密钥(token)、索引(indices)、数据表(type)和权限(permission)。在检测到用户输入操作请求时,通过ACL映射规则判断是否可以放行。若该操作请求满足ACL映射规则,则执行所述操作请求,否则,提示用户无操作权限。所述权限配置表可以存储在权限控制模块中,也可以存储在集群节点的其它位置中。

[0066] 按照这种方式,对于用户自行创建的索引或数据表,可以赋予该用户的账户对所创建的索引或数据表的读写权限。对于非用户自行创建的索引或数据表,不具备对索引或数据表的操作权限。若用户要访问不具有操作权限的索引或数据表,则该用户需要向管理员申请,由管理员为该用户的账户分配读权限。这样设置的好处在于实现index级别和type级别的分权控制,可以控制用户对每个index和type的读写权限。

[0067] 所述配置平台130用于提供基于WEB方式的人机交互配置界面,供用户输入配置信息来配置所述权限配置表。其中,所述配置信息包括权限配置参数(包括用户账号和密钥、可操作的索引index)、数据表结构参数(包括数据表名称和数据表类型)、数据抓取配置参数(包括数据源、网络地址url、账号、密钥和抓取方式)和数据同步配置参数(包括数据源、网络地址url、账号、密钥、触发方式和触发频率)。通过配置平台130实现分权控制和数据同步配置,无需用户再自行开发代码,提升研发效率。

[0068] 本实施例的技术方案,通过权限控制模块122实现对集群节点120内的数据执行分权控制,满足多用户协同工作时,数据的隔离与共享要求。通过配置平台130实现分权控制

和数据同步配置,无需用户再自行开发代码,提升研发效率。

[0069] 此外,这里要说明的是,优选地,同步装置中的配置模块也可以设置在配置平台130中。

[0070] 图9示出了根据本发明实施例二的检索方法的一个示例的流程图。

[0071] 如图9所示,在步骤910,集群节点中的权限控制模块通过集群节点中运行的权限控制插件,接收用户的检索请求,所述检索请求包含账户标识、操作内容和操作对象。在一个示例中,所述账户标识可以包括应用名称和账户密码,所述操作对象包括数据库索引信息和数据表信息,所述操作内容包括操作权限。

[0072] 在接收到用户的检索请求后,在步骤S920,从检索请求中提取出账户标识、操作内容和操作对象信息,然后在步骤S930,基于所提取出账户标识、操作内容和操作对象信息,查询权限控制模块中的权项配置表,判断是否具有操作权限,所述权限配置表存储账户标识、操作内容、操作对象和操作权限的映射关系。比如,所述权限配置表包括应用名称、账户密码、数据库索引信息、数据表信息和操作权限的映射关系。

[0073] 例如,定义一个专用的index的type用来存储ACL映射规则,即,权限配置表。其中,ACL映射规则包括定义了应用名称(app)、密钥(token)、索引(indices)、数据表(type)和权限(permission)。

[0074] 所述权限控制插件在拦截用户的操作指示后,提取所述操作指示中的账户标识,根据所述账户标识查询权限配置表,确定该用户可以操作(写和/或读)的索引或数据表。提取所述操作指示中的操作对象,与所确定的用户可以操作的索引或数据表进行匹配。提取对应于匹配成功的索引或数据表的操作内容,将操作内容与用户的操作权限进行比对,由此判断用户是否具有操作权限。

[0075] 若具备操作权限,则在步骤S940,将所述检索请求传输给集群节点的检索应用程序接口,在集群节点的数据表中搜索匹配的结果;

[0076] 若不具备操作权限,则在步骤S950,屏蔽所述检索请求。

[0077] 用户输入的操作指示无效,包括用户对操作指示中包括的数据表无操作权限,或者,用户对操作指示中包括的数据表的操作权限与操作指示中操作内容不符等。例如,用户对操作指示中包括的数据表的操作权限仅具有读操作,但是输入的操作指示中的操作内容是对该数据表执行写操作。

[0078] 在用户输入的操作指示无效时,屏蔽所述操作指示,提示无操作权限。

[0079] 本实施例的技术方案,通过在集群节点中运行权限控制插件,实现账户在索引级别或数据表级别的分权控制,对不同用户的数据进行物理隔离存储,且用户无法对无操作权限的数据执行操作,达到了避免出现误操作或恶意操作的效果。

[0080] 将意识到的是,本发明也扩展到适合于将本发明付诸实践的计算机程序,特别是载体上或者载体中的计算机程序。程序可以以源代码、目标代码、代码中间源和诸如部分编译的形式的目标代码的形式,或者以任何其它适合在按照本发明的方法的实现中使用的形式。也将注意的是,这样的程序可能具有许多不同的构架设计。例如,实现按照本发明的方法或者系统的功能性的程序代码可能被再分为一个或者多个子例程。

[0081] 注意,上述仅为本发明的较佳实施例及所运用技术原理。本领域技术人员会理解,本发明不限于这里所述的特定实施例,对本领域技术人员来说能够进行各种明显的变化、

重新调整和替代而不会脱离本发明的保护范围。因此,虽然通过以上实施例对本发明进行了较为详细的说明,但是本发明不仅仅限于以上实施例,在不脱离本发明构思的情况下,还可以包括更多其他等效实施例,而本发明的范围由所附的权利要求范围决定。

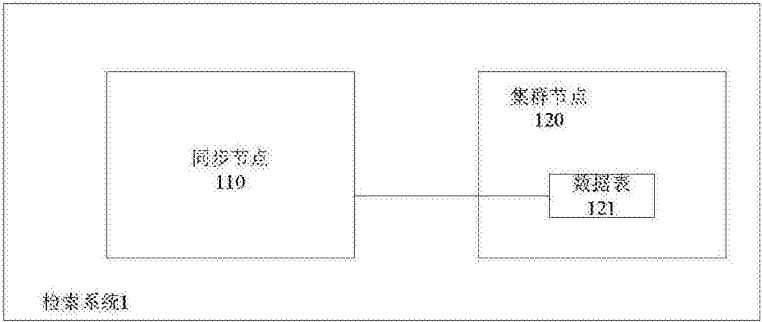


图1

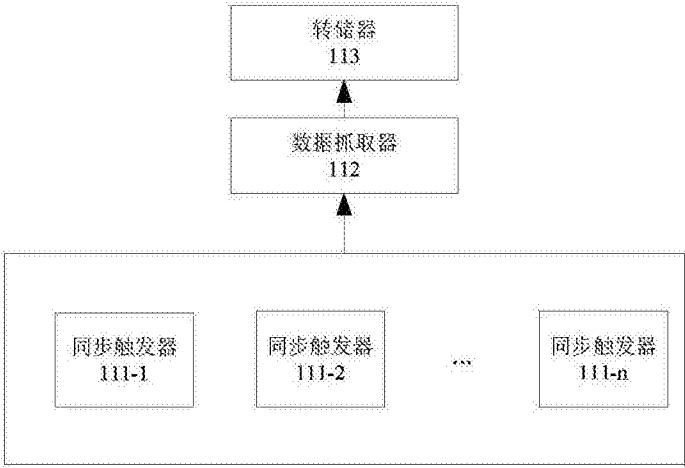


图2

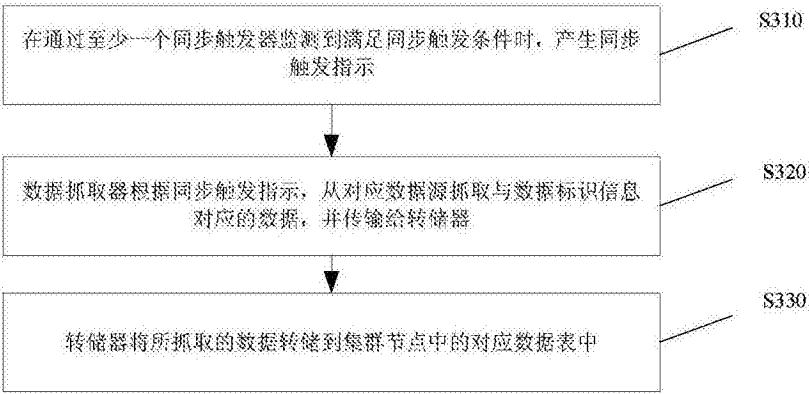


图3

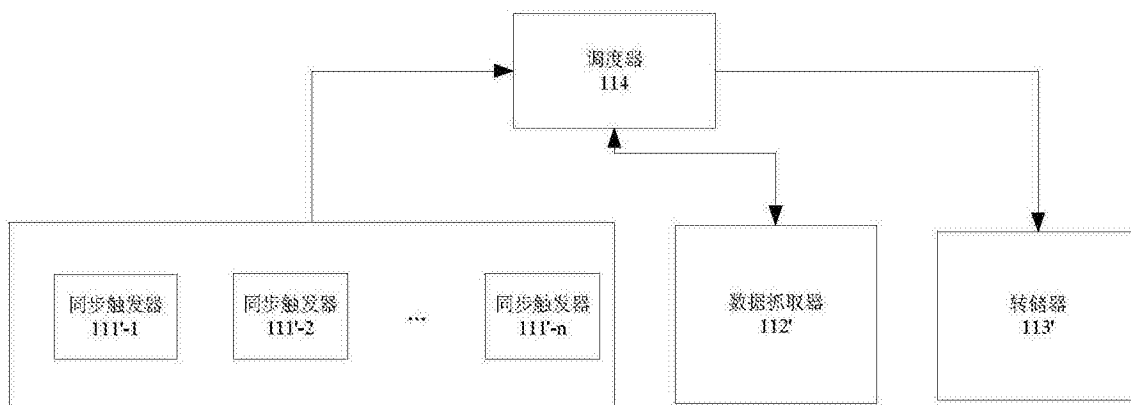


图4

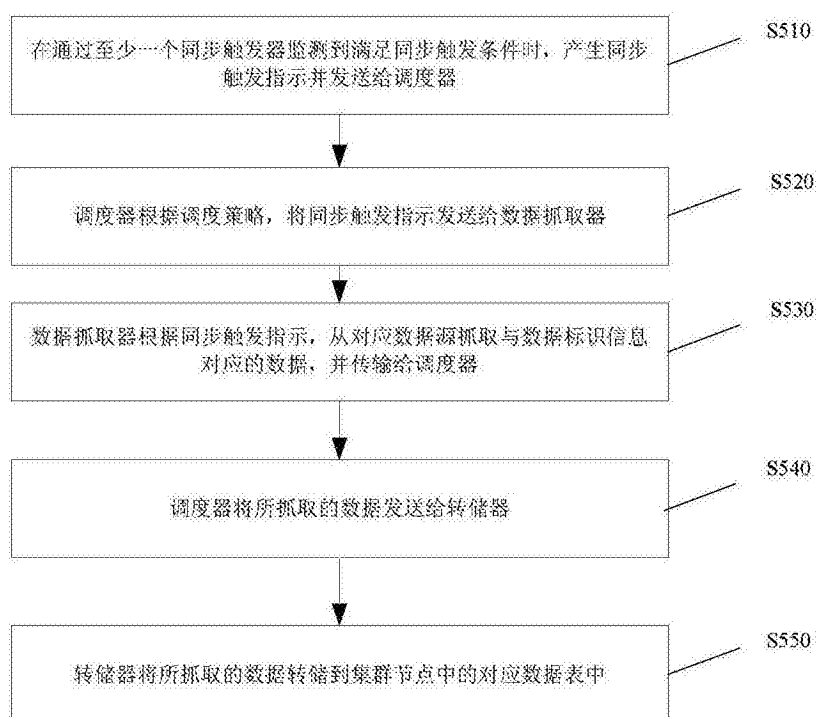


图5

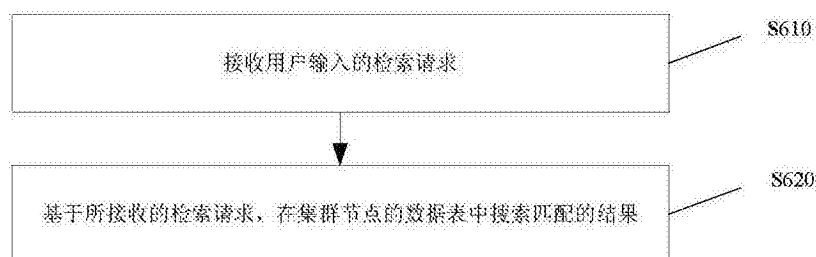


图6



RTS 检查系统

系统介绍

配置管理

应用注册

应用配置

监控管理

应用注册

应用名称:

应用描述:

保存 重置

图7A



RTS 检查系统

系统介绍

配置管理

应用注册

应用配置

监控管理

应用配置

应用名称:

应用描述:

Protocol:

Address:

InterfaceClass:

Method:

ExtraOptions:

参数格式: / key:数据类型,value:数据类型 eg: name(java.lang.String)age(integer)

DriverName: e-Cree...Engine...Database

Username:

Password:

SQL:

图7B



图7C

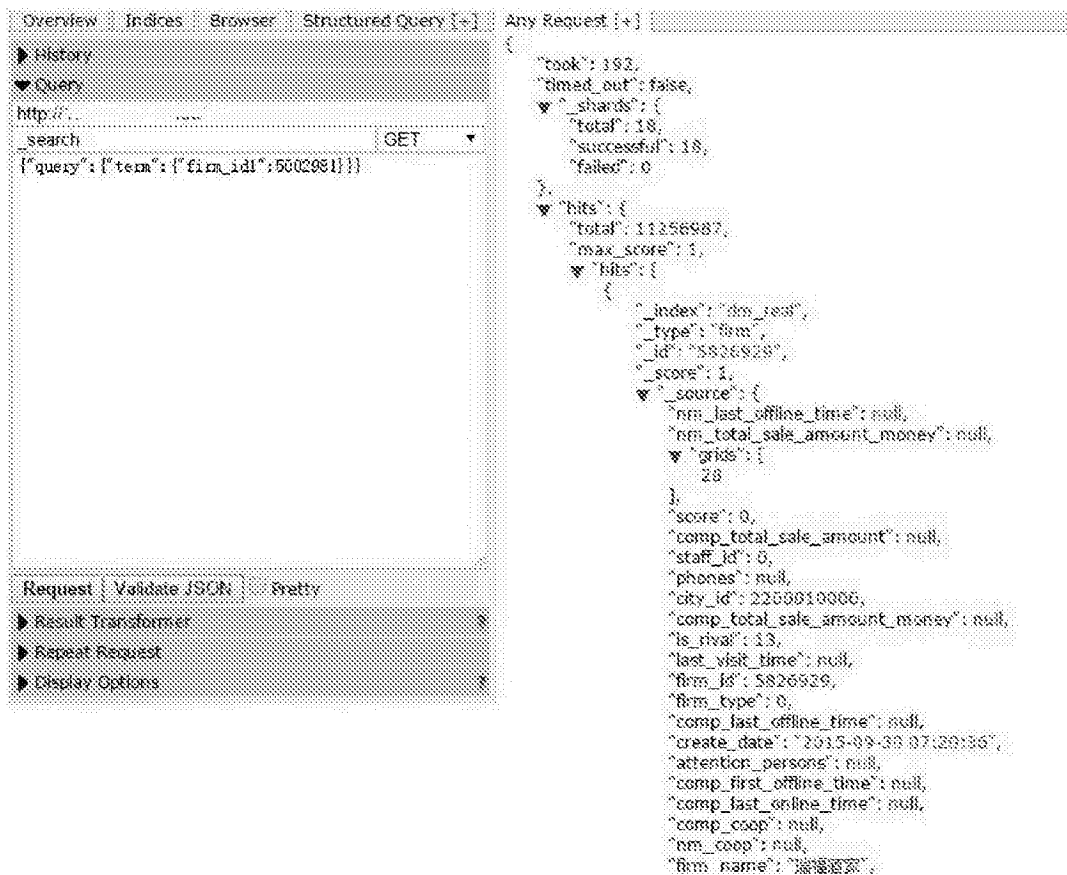


图7D

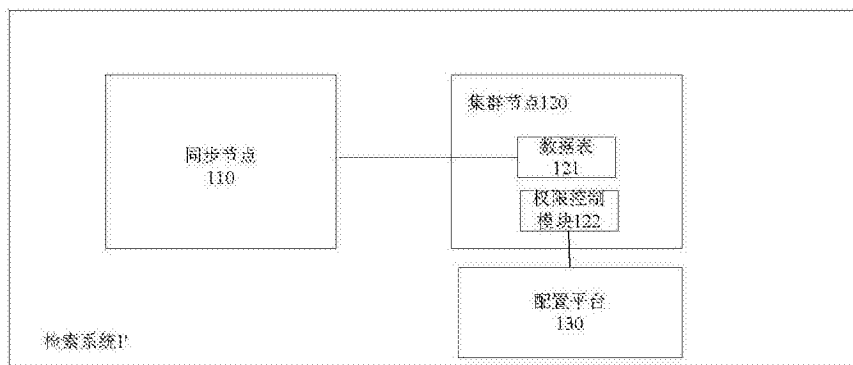


图8

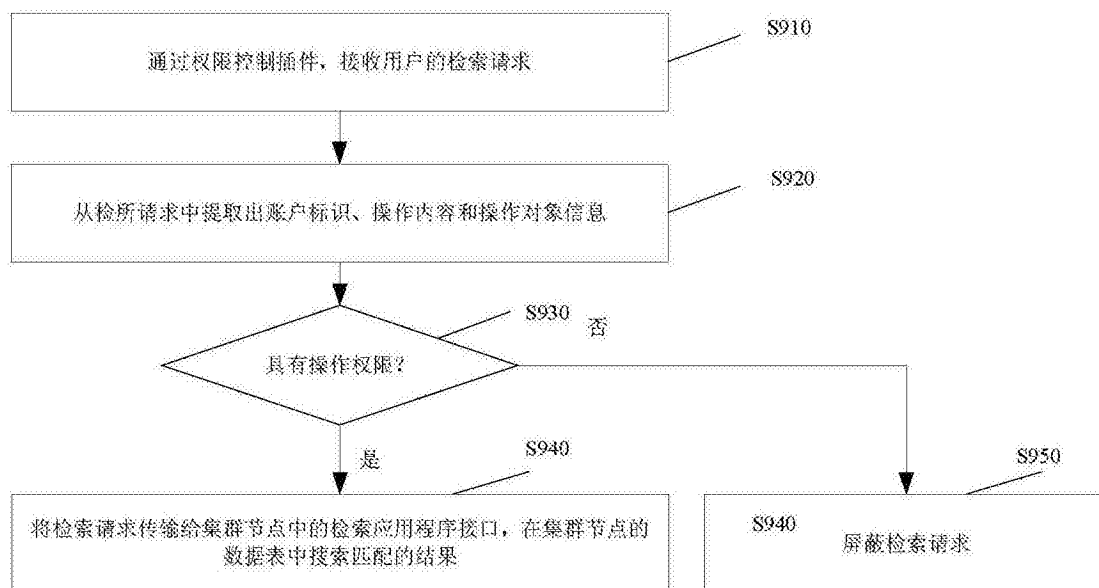


图9