



(12) 发明专利申请

(10) 申请公布号 CN 112506486 A

(43) 申请公布日 2021.03.16

(21) 申请号 202011483122.2

G06F 16/242 (2019.01)

(22) 申请日 2020.12.15

G06F 16/28 (2019.01)

(71) 申请人 康键信息技术(深圳)有限公司

地址 518000 广东省深圳市前海深港合作区前湾一路1号A栋201室(入驻深圳市前海商务秘书有限公司)

(72) 发明人 张华

(74) 专利代理机构 深圳市沃德知识产权代理事务所(普通合伙) 44347

代理人 高杰 于志光

(51) Int. Cl.

G06F 8/20 (2018.01)

G06F 8/30 (2018.01)

G06F 16/22 (2019.01)

G06F 16/23 (2019.01)

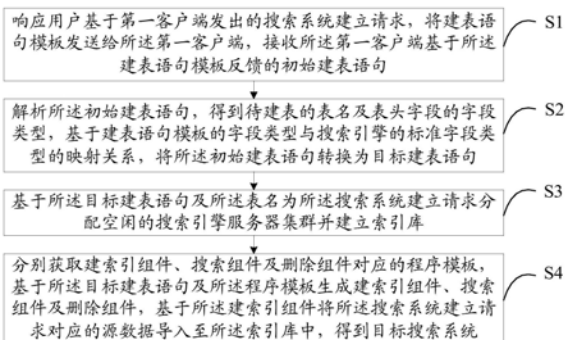
权利要求书2页 说明书11页 附图2页

(54) 发明名称

搜索系统建立方法、装置、电子设备及可读存储介质

(57) 摘要

本发明涉及数据处理,揭露一种搜索系统建立方法,包括:接收第一客户端用户基于建表语句模板反馈的初始建表语句;解析初始建表语句,得到待建表的表名及表头字段的字段类型,基于建表语句模板的字段类型与搜索引擎的标准字段类型的映射关系,将初始建表语句转换为目标建表语句;基于目标建表语句建立索引库;分别获取建索引组件、搜索组件及删除组件对应的程序模板,基于目标建表语句及程序模板生成建索引组件、搜索组件及删除组件,基于建索引组件将源数据导入至索引库中,得到目标搜索系统。本发明还提供一种搜索系统建立装置、电子设备及可读存储介质。本发明降低了搜索系统开发难度、提高了开发效率。



1. 一种搜索系统建立方法,其特征在于,所述方法包括:

响应用户基于第一客户端发出的搜索系统建立请求,将建表语句模板发送给所述第一客户端,接收所述第一客户端基于所述建表语句模板反馈的初始建表语句;

解析所述初始建表语句,得到待建表的表名及表头字段的字段类型,基于建表语句模板的字段类型与搜索引擎的标准字段类型的映射关系,将所述初始建表语句转换为目标建表语句;

基于所述目标建表语句及所述表名为所述搜索系统建立请求分配空闲的搜索引擎服务器集群并建立索引库;

分别获取建索引组件、搜索组件及删除组件对应的程序模板,基于所述目标建表语句及所述程序模板生成建索引组件、搜索组件及删除组件,基于所述建索引组件将所述搜索系统建立请求对应的源数据导入至所述索引库中,得到目标搜索系统。

2. 如权利要求1所述的搜索系统建立方法,其特征在于,所述基于所述目标建表语句及所述程序模板生成建索引组件、搜索组件及删除组件,包括:

解析所述目标建表语句,得到待建表的表头字段名称及其格式,基于所述表头字段名称及其格式生成实体类;

基于所述实体类与建索引组件对应的程序模板生成建索引组件;

从所述目标建表语句中提取属性为可搜索字段的字段名称列表,基于所述字段名称列表、所述实体类及搜索组件对应的程序模板生成搜索组件;

基于所述实体类与删除组件对应的程序模板生成删除组件。

3. 如权利要求1所述的搜索系统建立方法,其特征在于,所述方法还包括:

当接收到用户基于第二客户端发出的索引库扩容请求时,将扩容配置界面展示到所述第二客户端的显示界面,接收用户基于所述扩容配置界面输入的目标索引库及待分片字段;

将所述待分片字段的key值输入预设计算公式进行计算,基于计算得到的数值将所述目标索引库中的数据分配到所述搜索引擎服务器集群的不同节点中。

4. 如权利要求3所述的搜索系统建立方法,其特征在于,所述预设计算公式为:

$$K\%N==0 \quad (1)$$

$$K\%N==1 \quad (2)$$

…….

$$K\%N==(N-1) \quad (N)$$

其中,K为待分片字段的key值,N为扩容后的总节点数;

满足公式(1)的数据将被存储至第一个节点,满足公式(2)的数据将被存储至第二个节点,……,满足公式(N)的数据将被存储至第N个节点。

5. 如权利要求1所述的搜索系统建立方法,其特征在于,所述方法还包括:

实时或定时获取所述搜索系统所属的服务器的资源利用率,当所述资源利用率超过第一阈值时,对所述搜索系统执行限流及降级处理。

6. 如权利要求1所述的搜索系统建立方法,其特征在于,所述方法还包括:

当接收到用户基于第三客户端发出的数据写入请求时,判断待写入数据的数据量是否大于第二阈值,若判断大于第二阈值,则对所述待写入数据执行异步写入操作。

7. 如权利要求1所述的搜索系统建立方法,其特征在于,所述建索引组件对应的程序模板中包括写入索引库的API接口名称、源数据名称、字段版本信息及动作发生时间信息;

所述搜索组件对应的程序模板中包括排序参数、分页参数、搜索字段及返回字段。

8. 一种搜索系统建立装置,其特征在于,所述装置包括:

响应模块,用于响应用户基于第一客户端发出的搜索系统建立请求,将建表语句模板发送给所述第一客户端,接收所述第一客户端基于所述建表语句模板反馈的初始建表语句;

解析模块,用于解析所述初始建表语句,得到待建表的表名及表头字段的字段类型,基于建表语句模板的字段类型与搜索引擎的标准字段类型的映射关系,将所述初始建表语句转换为目标建表语句;

分配模块,用于基于所述目标建表语句及所述表名为所述搜索系统建立请求分配空闲的搜索引擎服务器集群并建立索引库;

生成模块,用于分别获取建索引组件、搜索组件及删除组件对应的程序模板,基于所述目标建表语句及所述程序模板生成建索引组件、搜索组件及删除组件,基于所述建索引组件将所述搜索系统建立请求对应的源数据导入至所述索引库中,得到目标搜索系统。

9. 一种电子设备,其特征在于,所述电子设备包括:

至少一个处理器;以及,

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的搜索系统建立程序,所述搜索系统建立程序被所述至少一个处理器执行,以使所述至少一个处理器能够执行如权利要求1至7中任一项所述的搜索系统建立方法。

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上存储有搜索系统建立程序,所述搜索系统建立程序可被一个或者多个处理器执行,以实现如权利要求1至7任一项所述的搜索系统建立方法。

搜索系统建立方法、装置、电子设备及可读存储介质

技术领域

[0001] 本发明涉及数据处理领域,尤其涉及一种搜索系统建立方法、装置、电子设备及可读存储介质。

背景技术

[0002] 一个完整的搜索系统包括数据源、索引服务、搜索服务,不同的业务数据搜索系统对应不同的数据源、不同的索引流程、不同的搜索逻辑及日志监控服务。当需要新开发一个业务数据的搜索系统时,需要开发一整套的代码,使得开发周期长、开发难度大、开发效率低下。因此,亟需一种搜索系统建立方法,以降低搜索系统开发难度、提高开发效率。

发明内容

[0003] 鉴于以上内容,有必要提供一种搜索系统建立方法,旨在降低搜索系统开发难度、提高开发效率。

[0004] 本发明提供的搜索系统建立方法,包括:

[0005] 响应用户基于第一客户端发出的搜索系统建立请求,将建表语句模板发送给所述第一客户端,接收所述第一客户端基于所述建表语句模板反馈的初始建表语句;

[0006] 解析所述初始建表语句,得到待建表的表名及表头字段的字段类型,基于建表语句模板的字段类型与搜索引擎的标准字段类型的映射关系,将所述初始建表语句转换为目标建表语句;

[0007] 基于所述目标建表语句及所述表名为所述搜索系统建立请求分配空闲的搜索引擎服务器集群并建立索引库;

[0008] 分别获取建索引组件、搜索组件及删除组件对应的程序模板,基于所述目标建表语句及所述程序模板生成建索引组件、搜索组件及删除组件,基于所述建索引组件将所述搜索系统建立请求对应的源数据导入至所述索引库中,得到目标搜索系统。

[0009] 可选的,所述基于所述目标建表语句及所述程序模板生成建索引组件、搜索组件及删除组件,包括:

[0010] 解析所述目标建表语句,得到待建表的表头字段名称及其格式,基于所述表头字段名称及其格式生成实体类;

[0011] 基于所述实体类与建索引组件对应的程序模板生成建索引组件;

[0012] 从所述目标建表语句中提取属性为可搜索字段的字段名称列表,基于所述字段名称列表、所述实体类及搜索组件对应的程序模板生成搜索组件;

[0013] 基于所述实体类与删除组件对应的程序模板生成删除组件。

[0014] 可选的,所述方法还包括:

[0015] 当接收到用户基于第二客户端发出的索引库扩容请求时,将扩容配置界面展示到所述第二客户端的显示界面,接收用户基于所述扩容配置界面输入的目标索引库及待分片字段;

[0016] 将所述待分片字段的key值输入预设计算公式进行计算,基于计算得到的数值将所述目标索引库中的数据分配到所述搜索引擎服务器集群的不同节点中。

[0017] 可选的,所述预设计算公式为:

[0018] $K\%N==0$ (1)

[0019] $K\%N==1$ (2)

[0020] …….

[0021] $K\%N==(N-1)$ (N)

[0022] 其中,K为待分片字段的key值,N为扩容后的总节点数;

[0023] 满足公式(1)的数据将被存储至第一个节点,满足公式(2)的数据将被存储至第二个节点,……,满足公式(N)的数据将被存储至第N个节点。

[0024] 可选的,所述方法还包括:

[0025] 实时或定时获取所述搜索系统所属的服务器的资源利用率,当所述资源利用率超过第一阈值时,对所述搜索系统执行限流及降级处理。

[0026] 可选的,所述方法还包括:

[0027] 当接收到用户基于第三客户端发出的数据写入请求时,判断待写入数据的数据量是否大于第二阈值,若判断大于第二阈值,则对所述待写入数据执行异步写入操作。

[0028] 可选的,所述建索引组件对应的程序模板中包括写入索引库的API接口名称、源数据名称、字段版本信息及动作发生时间信息;

[0029] 所述搜索组件对应的程序模板中包括排序参数、分页参数、搜索字段及返回字段。

[0030] 为了解决上述问题,本发明还提供一种搜索系统建立装置,所述装置包括:

[0031] 响应模块,用于响应用户基于第一客户端发出的搜索系统建立请求,将建表语句模板发送给所述第一客户端,接收所述第一客户端基于所述建表语句模板反馈的初始建表语句;

[0032] 解析模块,用于解析所述初始建表语句,得到待建表的表名及表头字段的字段类型,基于建表语句模板的字段类型与搜索引擎的标准字段类型的映射关系,将所述初始建表语句转换为目标建表语句;

[0033] 分配模块,用于基于所述目标建表语句及所述表名为所述搜索系统建立请求分配空闲的搜索引擎服务器集群并建立索引库;

[0034] 生成模块,用于分别获取建索引组件、搜索组件及删除组件对应的程序模板,基于所述目标建表语句及所述程序模板生成建索引组件、搜索组件及删除组件,基于所述建索引组件将所述搜索系统建立请求对应的源数据导入至所述索引库中,得到目标搜索系统。

[0035] 为了解决上述问题,本发明还提供一种电子设备,所述电子设备包括:

[0036] 至少一个处理器;以及,

[0037] 与所述至少一个处理器通信连接的存储器;其中,

[0038] 所述存储器存储有可被所述至少一个处理器执行的搜索系统建立程序,所述搜索系统建立程序被所述至少一个处理器执行,以使所述至少一个处理器能够执行上述搜索系统建立方法。

[0039] 为了解决上述问题,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上存储有搜索系统建立程序,所述搜索系统建立程序可被一个或者多个处理器执

行,以实现上述搜索系统建立方法。

[0040] 相较于现有技术,本发明首先将建表语句模板发送给第一客户端,接收第一客户端用户基于建表语句模板反馈的初始建表语句,基于建表语句模板的字段类型与搜索引擎的标准字段类型的映射关系,将初始建表语句转换为目标建表语句;接着,分配空闲的搜索引擎服务器集群,并基于目标建表语句建立索引库,本发明提供了建表语句模板,使得用户不需要额外学习索引库建库方法,仅在建表语句模板中填入对应的参数即可实现建立索引库,大大提高了索引库建立效率;最后,基于目标建表语句及建索引组件、搜索组件及删除组件对应的程序模板,生成建索引组件、搜索组件及删除组件,基于建索引组件将源数据导入至所述索引库中,得到目标搜索系统。因此,本发明降低了搜索系统开发难度、提高了开发效率。

附图说明

[0041] 图1为本发明一实施例提供的搜索系统建立方法的流程示意图;

[0042] 图2为本发明一实施例提供的搜索系统建立装置的模块示意图;

[0043] 图3为本发明一实施例提供的实现搜索系统建立方法的电子设备的结构示意图;

[0044] 本发明目的的实现、功能特点及优点将结合实施例,参照附图做进一步说明。

具体实施方式

[0045] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅用以解释本发明,并不用于限定本发明。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。

[0046] 需要说明的是,在本发明中涉及“第一”、“第二”等的描述仅用于描述目的,而不能理解为指示或暗示其相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一个该特征。另外,各个实施例之间的技术方案可以相互结合,但是必须是以本领域普通技术人员能够实现为基础,当技术方案的结合出现相互矛盾或无法实现时应当认为这种技术方案的结合不存在,也不在本发明要求的保护范围之内。

[0047] 本发明提供一种搜索系统建立方法。参照图1所示,为本发明一实施例提供的搜索系统建立方法的流程示意图。该方法可以由一个电子设备执行,该电子设备可以由软件和/或硬件实现。

[0048] 本实施例中,搜索系统建立方法包括:

[0049] S1、响应用户基于第一客户端发出的搜索系统建立请求,将建表语句模板发送给所述第一客户端,接收所述第一客户端基于所述建表语句模板反馈的初始建表语句。

[0050] 当前,通常使用ES (ElasticSearch,搜索引擎框架) 建库语句建立索引库(索引库类似于数据表),然而,ES建库语句较为复杂,且语句数量庞大。本实施例提供了一种建表语句模板,该建表语句模板的语法为MySQL语法,MySQL语法为一种更为常见的语法,从而开发人员不需另外学习ES建库方法,且MySQL语法可用更加简洁的语言实现建表功能(通过验证,针对某一建表需求,MySQL建表语句只有十行,而ES建表语句需要七十多行),采用MySQL

语法的建表语句模板可大大提高建表效率。

[0051] S2、解析所述初始建表语句,得到待建表的表名及表头字段的字段类型,基于建表语句模板的字段类型与搜索引擎的标准字段类型的映射关系,将所述初始建表语句转换为目标建表语句。

[0052] 解析所述初始建表语句,可得到待建表的表名、表头字段的字段名、字段类型和字段格式,这些内容为第一客户端用户填充至建表语句模板中的,所述字段类型包括字符型、数值型等,所述字段格式为字段对应的值的格式。同时,所述初始建表语句中还包括各字段是否需要分词、是否需要存储、是否为可搜索字段。

[0053] 所述搜索引擎的标准字段类型即为ES建表语句的字段类型,生成的目标建表语句为ES建表语句,通过建表语句模板的字段类型与搜索引擎的标准字段类型的映射关系,可将MySQL语法的初始建表语句转化为ES格式的目标建表语句。

[0054] S3、基于所述目标建表语句及所述表名为所述搜索系统建立请求分配空闲的搜索引擎服务器集群并建立索引库。

[0055] 生成的索引库类似于数据表,字段名为数据表的表头部分的字段名,即生成的索引库为只含表头的空白数据表。

[0056] 本实施例中,可以随机为所述搜索系统建立请求分配空闲的搜索引擎服务器集群节点用以建立索引库,也可以将索引库建立在用户指定的集群节点上。

[0057] S4、分别获取建索引组件、搜索组件及删除组件对应的程序模板,基于所述目标建表语句及所述程序模板生成建索引组件、搜索组件及删除组件,基于所述建索引组件将所述搜索系统建立请求对应的源数据导入至所述索引库中,得到目标搜索系统。

[0058] 所述基于所述目标建表语句及所述程序模板生成建索引组件、搜索组件及删除组件,包括:

[0059] A11、解析所述目标建表语句,得到待建表的表头字段名称及其格式,基于所述表头字段名称及其格式生成实体类;

[0060] 所述表头字段的格式为对字段的具体配置,例如,date类型的字段,有一个format属性,表示该字段的值以什么格式输入输出。

[0061] 实体类定义了索引库的表头中各字段的值的获取方法,通过实体类可从源数据中抓取索引库中各字段对应的值。所述实体类(model类)的生成是将ES中的索引库映射到java中的一个POJO类(可直接根据目标建表语句生成)。

[0062] A12、基于所述实体类与建索引组件对应的程序模板生成建索引组件;

[0063] 所述建索引组件包括建索引API组件和建索引响应组件,所述建索引API组件用于导入数据并将数据存储至索引库中,所述建索引响应组件用于输出建索引响应结果信息(响应成功或失败、响应耗时、响应失败原因)。

[0064] 所述建索引组件对应的程序模板中定义了写入索引库的API接口名称(系统根据索引库名称、操作类型(增删改查)自动生成的)、源数据名称、字段版本信息(每次更新后字段版本会加一,用于解决并发访问导致的数据更新结果不一致的问题)、timestamp动作发生时间信息。

[0065] A13、从所述目标建表语句中提取属性为可搜索字段的字段名称列表,基于所述字段名称列表、所述实体类及搜索组件对应的程序模板生成搜索组件;

[0066] 例如,目标建表语句中有一个index属性,表示该字段是否需要索引(是否可搜索),当index属性为true时,表示可搜索;当index属性为false时,表示不可搜索。

[0067] 所述搜索组件包括搜索API组件和搜索响应组件,所述搜索API组件用于执行搜索操作,所述搜索响应组件用于存储检索到的数据,如果执行错误会返回错误代码和报错信息。

[0068] 所述搜索组件对应的程序模板中定义了排序参数、分页参数、搜索字段及返回字段。

[0069] A14、基于所述实体类与删除组件对应的程序模板生成删除组件。

[0070] 所述删除组件包括删除API组件和删除响应组件,所述删除API组件用于执行删除操作,所述删除响应组件用于记录删除结果(删除成功或失败,若删除失败,会返回错误代码和报错信息)。

[0071] 所述建索引组件、搜索组件及删除组件为搜索系统的功能组件,根据建索引组件可将源数据导入至建好的索引库中,此时搜索系统搭建完成,后续用户可通过建索引组件更新索引库中的数据,通过搜索组件搜索索引库中与检索字段对应的数据,通过删除组件删除待删除数据。

[0072] 本实施例中,所述方法还包括:

[0073] B11、当接收到用户基于第二客户端发出的索引库扩容请求时,将扩容配置界面展示到所述第二客户端的显示界面,接收用户基于所述扩容配置界面输入的目标索引库及待分片字段;

[0074] B12、将所述待分片字段的key值输入预设计算公式进行计算,基于计算得到的数值将所述目标索引库中的数据分配到所述搜索引擎服务器集群的不同节点中。

[0075] 本实施例中,所述预设计算公式为:

[0076] $K \% N = 0$ (1)

[0077] $K \% N = 1$ (2)

[0078] …….

[0079] $K \% N = (N - 1)$ (N)

[0080] 其中,K为待分片字段的key值,N为扩容后的总节点数;

[0081] 满足公式(1)的数据可被存储至第一个节点,满足公式(2)的数据可被存储至第二个节点,……,满足公式(N)的数据可被存储至第N个节点。

[0082] 例如,若“hospital”索引库原来存储在ES集群的A节点上,现需将“hospital”索引库扩容至B节点上(即存储在A、B两个节点上),则N为2,可将K除以2无余数对应的数据存储到A节点,将K除以2余数为1对应的数据存储到B节点。

[0083] 该索引库的新数据直接按照上述规则分别存储到A、B节点上,原有的老数据也会根据该规则重新拆分,分别存储到两个节点上。

[0084] 在本发明的另一个实施例中,所述方法还包括:

[0085] 实时或定时获取所述搜索系统所属的服务器的资源利用率,当所述资源利用率超过第一阈值时,对所述搜索系统执行限流及降级处理。

[0086] 所述限流,可以是限制最大并发访问量;所述降级,可以是将复杂的操作降级为普通的操作(例如,商品页面上本来是实时显示库存的,当服务器运行负载过高时,可降级为

定时更新显示库存)。

[0087] 本实施例中,所述方法还包括:

[0088] 当接收到用户基于第三客户端发出的数据写入请求时,判断待写入数据的数据量是否大于第二阈值,若判断大于第二阈值,则对所述待写入数据执行异步写入操作。

[0089] 由上述实施例可知,本发明提出的搜索系统建立方法,首先,将建表语句模板发送给第一客户端,接收第一客户端用户基于建表语句模板反馈的初始建表语句,基于建表语句模板的字段类型与搜索引擎的标准字段类型的映射关系,将初始建表语句转换为目标建表语句;接着,分配空闲的搜索引擎服务器集群,并基于目标建表语句建立索引库,本发明提供了建表语句模板,使得用户不需要额外学习索引库建库方法,仅在建表语句模板中填入对应的参数即可实现建立索引库,大大提高了索引库建立效率;最后,基于目标建表语句及建索引组件、搜索组件及删除组件对应的程序模板,生成建索引组件、搜索组件及删除组件,基于建索引组件将源数据导入至所述索引库中,得到目标搜索系统。因此,本发明降低了搜索系统开发难度、提高了开发效率。

[0090] 如图2所示,为本发明一实施例提供的搜索系统建立装置的模块示意图。

[0091] 本发明所述搜索系统建立装置100可以安装于电子设备中。根据实现的功能,所述搜索系统建立装置100可以包括响应模块110、解析模块120、分配模块130及生成模块140。本发明所述模块也可以称之为单元,是指一种能够被电子设备处理器所执行,并且能够完成固定功能的一系列计算机程序段,其存储在电子设备的存储器中。

[0092] 在本实施例中,关于各模块/单元的功能如下:

[0093] 响应模块110,用于响应用户基于第一客户端发出的搜索系统建立请求,将建表语句模板发送给所述第一客户端,接收所述第一客户端基于所述建表语句模板反馈的初始建表语句。

[0094] 当前,通常使用ES (ElasticSearch,搜索引擎框架) 建库语句建立索引库(索引库类似于数据表),然而,ES建库语句较为复杂,且语句数量庞大。本实施例提供了一种建表语句模板,该建表语句模板的语法为MySQL语法,MySQL语法为一种更为常见的语法,从而开发人员不需另外学习ES建库方法,且MySQL语法可用更加简洁的语言实现建表功能(通过验证,针对某一建表需求,MySQL建表语句只有十行,而ES建表语句需要七十多行),采用MySQL语法的建表语句模板可大大提高建表效率。

[0095] 解析模块120,用于解析所述初始建表语句,得到待建表的表名及表头字段的字段类型,基于建表语句模板的字段类型与搜索引擎的标准字段类型的映射关系,将所述初始建表语句转换为目标建表语句。

[0096] 解析所述初始建表语句,可得到待建表的表名、表头字段的字段名、字段类型和字段格式,这些内容为第一客户端用户填充至建表语句模板中的,所述字段类型包括字符型、数值型等,所述字段格式为字段对应的值的格式。同时,所述初始建表语句中还包括各字段是否需要分词、是否需要存储、是否为可搜索字段。

[0097] 所述搜索引擎的标准字段类型即为ES建表语句的字段类型,生成的目标建表语句为ES建表语句,通过建表语句模板的字段类型与搜索引擎的标准字段类型的映射关系,可将MySQL语法的初始建表语句转化为ES格式的目标建表语句。

[0098] 分配模块130,用于基于所述目标建表语句及所述表名为所述搜索系统建立请求

分配空闲的搜索引擎服务器集群并建立索引库。

[0099] 生成的索引库类似于数据表,字段名为数据表的表头部分的字段名,即生成的索引库为只含表头的空白数据表。

[0100] 本实施例中,可以随机为所述搜索系统建立请求分配空闲的搜索引擎服务器集群节点用以建立索引库,也可以将索引库建立在用户指定的集群节点上。

[0101] 生成模块140,用于分别获取建索引组件、搜索组件及删除组件对应的程序模板,基于所述目标建表语句及所述程序模板生成建索引组件、搜索组件及删除组件,基于所述建索引组件将所述搜索系统建立请求对应的源数据导入至所述索引库中,得到目标搜索系统。

[0102] 所述基于所述目标建表语句及所述程序模板生成建索引组件、搜索组件及删除组件,包括:

[0103] A21、解析所述目标建表语句,得到待建表的表头字段名称及其格式,基于所述表头字段名称及其格式生成实体类;

[0104] 所述表头字段的格式为对字段的具体配置,例如,date类型的字段,有一个formate属性,表示该字段的值以什么格式输入输出。

[0105] 实体类定义了索引库的表头中各字段的值的获取方法,通过实体类可从源数据中抓取索引库中各字段对应的值。所述实体类(model类)的生成是将ES中的索引库映射到java中的一个POJO类(可直接根据目标建表语句生成)。

[0106] A22、基于所述实体类与建索引组件对应的程序模板生成建索引组件;

[0107] 所述建索引组件包括建索引API组件和建索引响应组件,所述建索引API组件用于导入数据并将数据存储至索引库中,所述建索引响应组件用于输出建索引响应结果信息(响应成功或失败、响应耗时、响应失败原因)。

[0108] 所述建索引组件对应的程序模板中定义了写入索引库的API接口名称(系统根据索引库名称、操作类型(增删改查)自动生成的)、源数据名称、字段版本信息(每次更新后字段版本会加一,用于解决并发访问导致的数据更新结果不一致的问题)、timestamp动作发生时间信息。

[0109] A23、从所述目标建表语句中提取属性为可搜索字段的字段名称列表,基于所述字段名称列表、所述实体类及搜索组件对应的程序模板生成搜索组件;

[0110] 例如,目标建表语句中有一个index属性,表示该字段是否需要索引(是否可搜索),当index属性为true时,表示可搜索;当index属性为false时,表示不可搜索。

[0111] 所述搜索组件包括搜索API组件和搜索响应组件,所述搜索API组件用于执行搜索操作,所述搜索响应组件用于存储检索到的数据,如果执行错误会返回错误代码和报错信息。

[0112] 所述搜索组件对应的程序模板中定义了排序参数、分页参数、搜索字段及返回字段。

[0113] A24、基于所述实体类与删除组件对应的程序模板生成删除组件。

[0114] 所述删除组件包括删除API组件和删除响应组件,所述删除API组件用于执行删除操作,所述删除响应组件用于记录删除结果(删除成功或失败,若删除失败,会返回错误代码和报错信息)。

[0115] 所述建索引组件、搜索组件及删除组件为搜索系统的功能组件,根据建索引组件可将源数据导入至建好的索引库中,此时搜索系统搭建完成,后续用户可通过建索引组件更新索引库中的数据,通过搜索组件搜索索引库中与检索字段对应的数据,通过删除组件删除待删除数据。

[0116] 本实施例中,所述生成模块140还用于:

[0117] B21、当接收到用户基于第二客户端发出的索引库扩容请求时,将扩容配置界面展示到所述第二客户端的显示界面,接收用户基于所述扩容配置界面输入的目标索引库及待分片字段;

[0118] B22、将所述待分片字段的key值输入预设计算公式进行计算,基于计算得到的数值将所述目标索引库中的数据分配到所述搜索引擎服务器集群的不同节点中。

[0119] 本实施例中,所述预设计算公式为:

[0120] $K \% N = 0$ (1)

[0121] $K \% N = 1$ (2)

[0122] …….

[0123] $K \% N = (N - 1)$ (N)

[0124] 其中,K为待分片字段的key值,N为扩容后的总节点数;

[0125] 满足公式(1)的数据可被存储至第一个节点,满足公式(2)的数据可被存储至第二个节点,……,满足公式(N)的数据可被存储至第N个节点。

[0126] 例如,若“hospital”索引库原来存储在ES集群的A节点上,现需将“hospital”索引库扩容至B节点上(即存储在A、B两个节点上),则N为2,可将K除以2无余数对应的数据存储在A节点,将K除以2余数为1对应的数据存储在B节点。

[0127] 该索引库的新数据直接按照上述规则分别存储到A、B节点上,原有的老数据也会根据该规则重新拆分,分别存储到两个节点上。

[0128] 在本发明的另一个实施例中,所述生成模块140还用于:

[0129] 实时或定时获取所述搜索系统所属的服务器的资源利用率,当所述资源利用率超过第一阈值时,对所述搜索系统执行限流及降级处理。

[0130] 所述限流,可以是限制最大并发访问量;所述降级,可以是将复杂的操作降级为普通的操作(例如,商品页面上本来是实时显示库存的,当服务器运行负载过高时,可降级为定时更新显示库存)。

[0131] 本实施例中,所述生成模块140还用于:

[0132] 当接收到用户基于第三客户端发出的数据写入请求时,判断待写入数据的数据量是否大于第二阈值,若判断大于第二阈值,则对所述待写入数据执行异步写入操作。

[0133] 由上述实施例可知,本发明提出的搜索系统建立装置100,首先,将建表语句模板发送给第一客户端,接收第一客户端用户基于建表语句模板反馈的初始建表语句,基于建表语句模板的字段类型与搜索引擎的标准字段类型的映射关系,将初始建表语句转换为目标建表语句;接着,分配空闲的搜索引擎服务器集群,并基于目标建表语句建立索引库,本发明提供了建表语句模板,使得用户不需要额外学习索引库建库方法,仅在建表语句模板中填入对应的参数即可实现建立索引库,大大提高了索引库建立效率;最后,基于目标建表语句及建索引组件、搜索组件及删除组件对应的程序模板,生成建索引组件、搜索组件及删

除组件,基于建索引组件将源数据导入至所述索引库中,得到目标搜索系统。因此,本发明降低了搜索系统开发难度、提高了开发效率。

[0134] 如图3所示,为本发明一实施例提供的实现搜索系统建立方法的电子设备的结构示意图。

[0135] 所述电子设备1是一种能够按照事先设定或者存储的指令,自动进行数值计算和/或信息处理的设备。所述电子设备1可以是计算机、也可以是单个网络服务器、多个网络服务器组成的服务器组或者基于云计算的由大量主机或者网络服务器构成的云,其中云计算是分布式计算的一种,由一群松散耦合的计算机集组成的一个超级虚拟计算机。

[0136] 在本实施例中,电子设备1包括,但不仅限于,可通过系统总线相互通信连接的存储器11、处理器12、网络接口13,该存储器11中存储有搜索系统建立程序10,所述搜索系统建立程序10可被所述处理器12执行。图3仅示出了具有组件11-13以及搜索系统建立程序10的电子设备1,本领域技术人员可以理解的是,图3示出的结构并不构成对电子设备1的限定,可以包括比图示更少或者更多的部件,或者组合某些部件,或者不同的部件布置。

[0137] 其中,存储器11包括内存及至少一种类型的可读存储介质。内存为电子设备1的运行提供缓存;可读存储介质可为如闪存、硬盘、多媒体卡、卡型存储器(例如,SD或DX存储器等)、随机访问存储器(RAM)、静态随机访问存储器(SRAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、可编程只读存储器(PROM)、磁性存储器、磁盘、光盘等的非易失性存储介质。在一些实施例中,可读存储介质可以是电子设备1的内部存储单元,例如该电子设备1的硬盘;在另一些实施例中,该非易失性存储介质也可以是电子设备1的外部存储设备,例如电子设备1上配备的插接式硬盘,智能存储卡(Smart Media Card, SMC),安全数字(Secure Digital, SD)卡,闪存卡(Flash Card)等。本实施例中,存储器11的可读存储介质通常用于存储安装于电子设备1的操作系统和各类应用软件,例如存储本发明一实施例中的搜索系统建立程序10的代码等。此外,存储器11还可以用于暂时地存储已经输出或者将要输出的各类数据。

[0138] 处理器12在一些实施例中可以是中央处理器(Central Processing Unit, CPU)、控制器、微控制器、微处理器、或其他数据处理芯片。该处理器12通常用于控制所述电子设备1的总体操作,例如执行与其他设备进行数据交互或者通信相关的控制和处理等。本实施例中,所述处理器12用于运行所述存储器11中存储的程序代码或者处理数据,例如运行搜索系统建立程序10等。

[0139] 网络接口13可包括无线网络接口或有线网络接口,该网络接口13用于在所述电子设备1与客户端(图中未画出)之间建立通信连接。

[0140] 可选的,所述电子设备1还可以包括用户接口,用户接口可以包括显示器(Display)、输入单元比如键盘(Keyboard),可选的用户接口还可以包括标准的有线接口、无线接口。可选的,在一些实施例中,显示器可以是LED显示器、液晶显示器、触控式液晶显示器以及OLED(Organic Light-Emitting Diode,有机发光二极管)触摸器等。其中,显示器也可以适当的称为显示屏或显示单元,用于显示在电子设备1中处理的信息以及用于显示可视化的用户界面。

[0141] 应该了解,所述实施例仅为说明之用,在专利申请范围上并不受此结构的限制。

[0142] 所述电子设备1中的所述存储器11存储的搜索系统建立程序10是多个指令的组

合,在所述处理器12中运行时,可以实现:

[0143] 响应用户基于第一客户端发出的搜索系统建立请求,将建表语句模板发送给所述第一客户端,接收所述第一客户端基于所述建表语句模板反馈的初始建表语句;

[0144] 解析所述初始建表语句,得到待建表的表名及表头字段的字段类型,基于建表语句模板的字段类型与搜索引擎的标准字段类型的映射关系,将所述初始建表语句转换为目标建表语句;

[0145] 基于所述目标建表语句及所述表名为所述搜索系统建立请求分配空闲的搜索引擎服务器集群并建立索引库;

[0146] 分别获取建索引组件、搜索组件及删除组件对应的程序模板,基于所述目标建表语句及所述程序模板生成建索引组件、搜索组件及删除组件,基于所述建索引组件将所述搜索系统建立请求对应的源数据导入至所述索引库中,得到目标搜索系统。

[0147] 具体地,所述处理器12对上述搜索系统建立程序10的具体实现方法可参考图1对应实施例中相关步骤的描述,在此不赘述。

[0148] 进一步地,所述电子设备1集成的模块/单元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。所述计算机可读介质可以是非易失性的,也可以是非易失性的。所述计算机可读介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM,Read-Only Memory)。

[0149] 所述计算机可读存储介质上存储有搜索系统建立程序10,所述搜索系统建立程序10可被一个或者多个处理器执行,本发明计算机可读存储介质具体实施方式与上述搜索系统建立方法各实施例基本相同,在此不作赘述。

[0150] 在本发明所提供的几个实施例中,应该理解到,所揭露的设备,装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述模块的划分,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。

[0151] 所述作为分离部件说明的模块可以是或者也可以不是物理上分开的,作为模块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模块来实现本实施例方案的目的。

[0152] 另外,在本发明各个实施例中的各功能模块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。上述集成的单元既可以采用硬件的形式实现,也可以采用硬件加软件功能模块的形式实现。

[0153] 对于本领域技术人员而言,显然本发明不限于上述示范性实施例的细节,而且在不背离本发明的精神或基本特征的情况下,能够以其他的具体形式实现本发明。

[0154] 因此,无论从哪一点来看,均应将实施例看作是示范性的,而且是非限制性的,本发明的范围由所附权利要求而不是上述说明限定,因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本发明内。不应将权利要求中的任何附关联图标记视为限制所涉及的权利要求。

[0155] 本发明所指区块链是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式。区块链(Blockchain),本质上是一个去中心化的数据库,是一串使用

密码学方法相关联产生的数据块,每一个数据块中包含了一批次网络交易的信息,用于验证其信息的有效性(防伪)和生成下一个区块。区块链可以包括区块链底层平台、平台产品服务层以及应用服务层等。

[0156] 此外,显然“包括”一词不排除其他单元或步骤,单数不排除复数。系统权利要求中陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第二等词语用来表示名称,而并不表示任何特定的顺序。

[0157] 最后应说明的是,以上实施例仅用以说明本发明的技术方案而非限制,尽管参照较佳实施例对本发明进行了详细说明,本领域的普通技术人员应当理解,可以对本发明的技术方案进行修改或等同替换,而不脱离本发明技术方案的精神和范围。

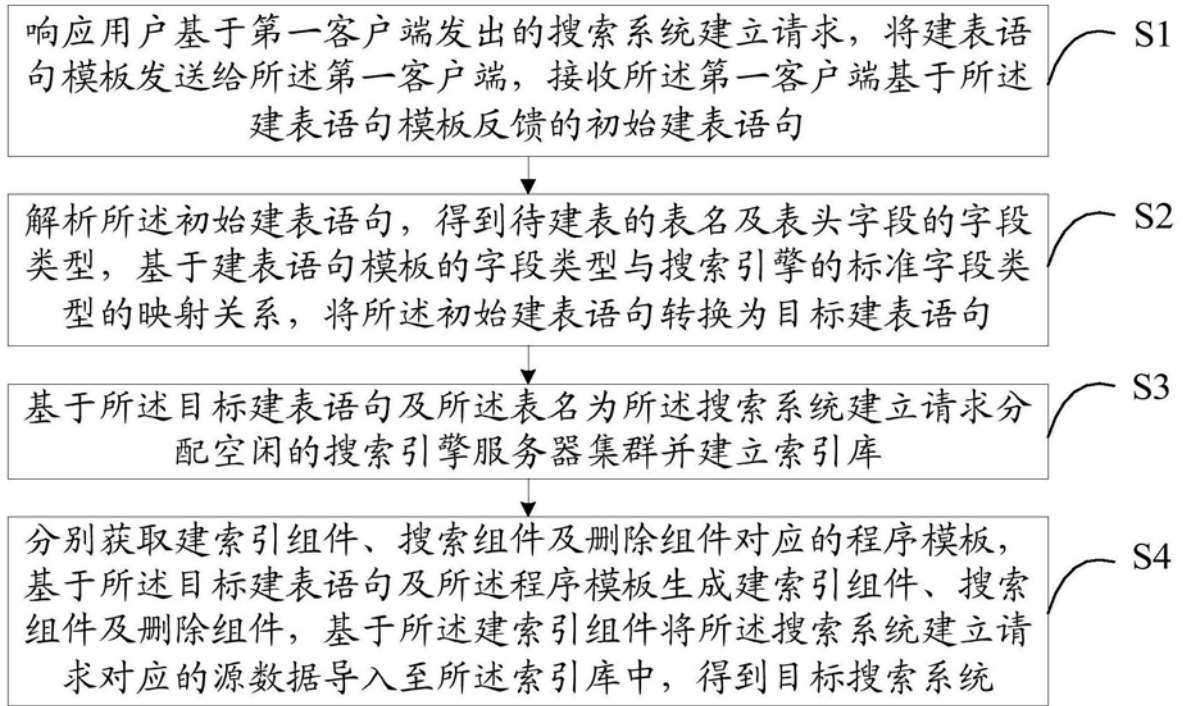


图1



图2

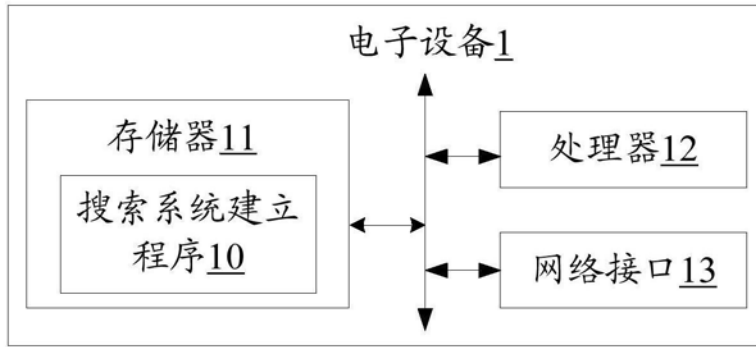


图3