



(12)发明专利申请

(10)申请公布号 CN 108763527 A

(43)申请公布日 2018. 11. 06

(21)申请号 201810548802.4

(22)申请日 2018.05.31

(71)申请人 北京明朝万达科技股份有限公司  
地址 100097 北京市海淀区蓝靛厂南路25号嘉友国际大厦北区2层

(72)发明人 张鹏 孙加光 喻波 王志海  
秦凯 安鹏 吴宏毅

(74)专利代理机构 北京润泽恒知识产权代理有限公司 11319

代理人 莎日娜

(51) Int. Cl.  
G06F 17/30(2006.01)

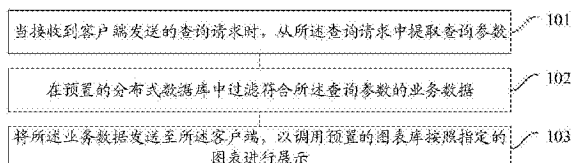
权利要求书2页 说明书11页 附图2页

(54)发明名称

一种业务数据的搜索方法和装置

(57)摘要

本发明实施例提供了一种业务数据的搜索方法和装置,该方法包括:当接收到客户端发送的查询请求时,从所述查询请求中提取查询参数;在预置的分布式数据库中过滤符合所述查询参数的业务数据;将所述业务数据发送至所述客户端,以调用预置的图表库按照指定的图表进行展示。本发明实施例基于分布式数据库本身性能,通过过滤业务数据,避免计算匹配度分值,使得查询效率得到了极大的提升,上亿条数据进行查询或聚合统计,都可以秒级查询出结果,在查询效率提高的同时,统计维度也得到了提升。



1. 一种业务数据的搜索方法,其特征在于,包括:

当接收到客户端发送的查询请求时,从所述查询请求中提取查询参数;

在预置的分布式数据库中过滤符合所述查询参数的业务数据;

将所述业务数据发送至所述客户端,以调用预置的图表库按照指定的图表进行展示。

2. 根据权利要求1所述的方法,其特征在于,所述在预置的分布式数据库中过滤符合所述查询参数的业务数据,包括:

将所述查询参数封装至过滤器filter中;

将所述过滤器filter封装至查询器query中;

调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数的业务数据。

3. 根据权利要求2所述的方法,其特征在于,所述调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数的业务数据,包括:

在预置的分布式数据库中查找所述查询参数对应的一个或多个存储区域;

调用所述查询器query在所述一个或多个存储区域进行搜索,以过滤符合所述查询参数的业务数据。

4. 根据权利要求2所述的方法,其特征在于,所述调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数的业务数据,包括:

在预置的分布式数据库中查找所述查询参数对应的索引文件;

调用所述查询器query在所述索引文件对应的业务数据中进行搜索,以过滤符合所述查询参数的业务数据;

其中,所述索引文件通过如下的至少一种方式生成:

将业务数据中指定的第一字段的类型设置为目标类型;

采用除所述目标类型之外的其他类型的字段生成索引文件;

和/或,

在业务数据中确定指定的第二字段;

对所述第二字段整体生成索引文件;

和/或,

判断业务数据是否为嵌套文档;

若是,则拆分所述业务数据;

若否,则采用所述业务数据生成索引文件。

5. 根据权利要求2所述的方法,其特征在于,所述分布式数据库具有多个数据库节点,所述调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数的业务数据,包括:

将每个数据库节点的内存设置为指定的值,以将指针限制在有效的长度内;

基于所述内存调用所述查询器query在所述数据库节点中进行搜索,以过滤符合所述查询参数的业务数据。

6. 根据权利要求2所述的方法,其特征在于,所述调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数的业务数据,还包括:

从所述查询请求中提取查询方式;

若所述查询方式为数据搜索,则确定数据条数;

调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数、且数量为所述数据条数的业务数据;

若所述查询方式为数据聚合,则调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数的、原始的业务数据;

将所述原始的业务数据进行聚合,作为新的业务数据进行输出;

禁止输出所述原始的业务数据。

7. 根据权利要求1-6任一项所述的方法,其特征在于,所述将所述业务数据发送至所述客户端,以调用预置的图表库按照指定的图表进行展示,包括:

将所述业务数据按照图表库的格式封装为查询结果;

将所述查询结果发送至所述客户端,以调用所述图表库读取所述查询结果、生成指定的图表进行展示。

8. 一种业务数据的搜索装置,其特征在于,包括:

查询参数提取模块,用于当接收到客户端发送的查询请求时,从所述查询请求中提取查询参数;

业务数据过滤模块,用于在预置的分布式数据库中过滤符合所述查询参数的业务数据;

业务数据发送模块,用于将所述业务数据发送至所述客户端,以调用预置的图表库按照指定的图表进行展示。

9. 根据权利要求8所述的装置,其特征在于,所述业务数据过滤模块包括:

过滤器封装子模块,用于将所述查询参数封装至过滤器filter中;

查询器封装子模块,用于将所述过滤器filter封装至查询器query中;

搜索子模块,用于调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数的业务数据。

10. 根据权利要求9所述的装置,其特征在于,所述搜索子模块包括:

存储区域查询单元,用于在预置的分布式数据库中查找所述查询参数对应的一个或多个存储区域;

存储区域搜索单元,用于调用所述查询器query在所述一个或多个存储区域进行搜索,以过滤符合所述查询参数的业务数据。

## 一种业务数据的搜索方法和装置

### 技术领域

[0001] 本发明涉及计算机处理的技术领域,特别是涉及一种业务数据的搜索方法和一种业务数据的搜索装置。

### 背景技术

[0002] 各个企业部署的Web服务器,为用户提供各项服务,存储了海量的业务数据。

[0003] 目前,Web服务器都是基于MySQL、Oracle等关系型数据库存储这些业务数据,在业务数据的数量在百万级以内时,能够比较快速的进行数据的分类统计。

[0004] 但是,当业务数据的数据量超过百万级、千万级,甚至过亿级时,关系型数据库的查询效率明显降低,甚至查询失败。

### 发明内容

[0005] 本发明实施例提出了一种业务数据的搜索方法和装置,以解决业务数据的数据量较大时,关系型数据库查询效率低的问题。

[0006] 依据本发明的一个方面,提供了一种业务数据的搜索方法,包括:

[0007] 当接收到客户端发送的查询请求时,从所述查询请求中提取查询参数;

[0008] 在预置的分布式数据库中过滤符合所述查询参数的业务数据;

[0009] 将所述业务数据发送至所述客户端,以调用预置的图表库按照指定的图表进行展示。

[0010] 可选地,所述在预置的分布式数据库中过滤符合所述查询参数的业务数据,包括:

[0011] 将所述查询参数封装至过滤器filter中;

[0012] 将所述过滤器filter封装至查询器query中;

[0013] 调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数的业务数据。

[0014] 可选地,所述调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数的业务数据,包括:

[0015] 在预置的分布式数据库中查找所述查询参数对应的一个或多个存储区域;

[0016] 调用所述查询器query在所述一个或多个存储区域进行搜索,以过滤符合所述查询参数的业务数据。

[0017] 可选地,所述调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数的业务数据,包括:

[0018] 在预置的分布式数据库中查找所述查询参数对应的索引文件;

[0019] 调用所述查询器query在所述索引文件对应的业务数据中进行搜索,以过滤符合所述查询参数的业务数据;

[0020] 其中,所述索引文件通过如下的至少一种方式生成:

[0021] 将业务数据中指定的第一字段的类型设置为目标类型;

- [0022] 采用除所述目标类型之外的其他类型的字段生成索引文件；
- [0023] 和/或，
- [0024] 在业务数据中确定指定的第二字段；
- [0025] 对所述第二字段整体生成索引文件；
- [0026] 和/或，
- [0027] 判断业务数据是否为嵌套文档；
- [0028] 若是，则拆分所述业务数据；
- [0029] 若否，则采用所述业务数据生成索引文件。
- [0030] 可选地，所述分布式数据库具有多个数据库节点，所述调用所述查询器query在预置的分布式数据库中进行搜索，以过滤符合所述查询参数的业务数据，包括：
- [0031] 将每个数据库节点的内存设置为指定的值，以将指针限制在有效的长度内；
- [0032] 基于所述内存调用所述查询器query在所述数据库节点中进行搜索，以过滤符合所述查询参数的业务数据。
- [0033] 可选地，所述调用所述查询器query在预置的分布式数据库中进行搜索，以过滤符合所述查询参数的业务数据，还包括：
- [0034] 从所述查询请求中提取查询方式；
- [0035] 若所述查询方式为数据搜索，则确定数据条数；
- [0036] 调用所述查询器query在预置的分布式数据库中进行搜索，以过滤符合所述查询参数、且数量为所述数据条数的业务数据；
- [0037] 若所述查询方式为数据聚合，则调用所述查询器query在预置的分布式数据库中进行搜索，以过滤符合所述查询参数的、原始的业务数据；
- [0038] 将所述原始的业务数据进行聚合，作为新的业务数据进行输出；
- [0039] 禁止输出所述原始的业务数据。
- [0040] 可选地，所述将所述业务数据发送至所述客户端，以调用预置的图表库按照指定的图表进行展示，包括：
- [0041] 将所述业务数据按照图表库的格式封装为查询结果；
- [0042] 将所述查询结果发送至所述客户端，以调用所述图表库读取所述查询结果、生成指定的图表进行展示。
- [0043] 根据本发明的另一方面，提供了一种业务数据的搜索装置，包括：
- [0044] 查询参数提取模块，用于当接收到客户端发送的查询请求时，从所述查询请求中提取查询参数；
- [0045] 业务数据过滤模块，用于在预置的分布式数据库中过滤符合所述查询参数的业务数据；
- [0046] 业务数据发送模块，用于将所述业务数据发送至所述客户端，以调用预置的图表库按照指定的图表进行展示。
- [0047] 可选地，所述业务数据过滤模块包括：
- [0048] 过滤器封装子模块，用于将所述查询参数封装至过滤器filter中；
- [0049] 查询器封装子模块，用于将所述过滤器filter封装至查询器query中；
- [0050] 搜索子模块，用于调用所述查询器query在预置的分布式数据库中进行搜索，以过

滤符合所述查询参数的业务数据。

[0051] 可选地,所述搜索子模块包括:

[0052] 存储区域查询单元,用于在预置的分布式数据库中查找所述查询参数对应的一个或多个存储区域;

[0053] 存储区域搜索单元,用于调用所述查询器query在所述一个或多个存储区域进行搜索,以过滤符合所述查询参数的业务数据。

[0054] 可选地,所述搜索子模块包括:

[0055] 索引文件查找单元,用于在预置的分布式数据库中查找所述查询参数对应的索引文件;

[0056] 索引文件搜索单元,用于调用所述查询器query在所述索引文件对应的业务数据中进行搜索,以过滤符合所述查询参数的业务数据;

[0057] 其中,所述索引文件通过如下的至少一种方式生成:

[0058] 目标类型设置单元,用于将业务数据中指定的第一字段的类型设置为目标类型;

[0059] 第一文件生成单元,用于采用除所述目标类型之外的其他类型的字段生成索引文件;

[0060] 和/或,

[0061] 字段指定单元,用于在业务数据中确定指定的第二字段;

[0062] 第二文件生成单元,用于对所述第二字段整体生成索引文件;

[0063] 和/或,

[0064] 嵌套文档判断单元,用于判断业务数据是否为嵌套文档;若是,则调用业务数据拆分单元,若否,则调用第三文件生成单元;

[0065] 业务数据拆分单元,用于拆分所述业务数据;

[0066] 第三文件生成单元,用于采用所述业务数据生成索引文件。

[0067] 可选地,所述分布式数据库具有多个数据库节点,所述搜索子模块包括:

[0068] 内存设置单元,用于将每个数据库节点的内存设置为指定的值,以将指针限制在有效的长度内;

[0069] 内存搜索单元,用于基于所述内存调用所述查询器query在所述数据库节点中进行搜索,以过滤符合所述查询参数的业务数据。

[0070] 可选地,所述搜索子模块包括:

[0071] 查询方式提取单元,用于从所述查询请求中提取查询方式;

[0072] 数据条数确定单元,用于若所述查询方式为数据搜索,则确定数据条数;

[0073] 条数过滤单元,用于调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数、且数量为所述数据条数的业务数据;

[0074] 原始数据过滤单元,用于若所述查询方式为数据聚合,则调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数的、原始的业务数据;

[0075] 原始数据聚合单元,用于将所述原始的业务数据进行聚合,作为新的业务数据进行输出;

[0076] 原始数据禁止输出单元,用于禁止输出所述原始的业务数据。

[0077] 可选地,所述业务数据发送模块包括:

[0078] 查询结果封装子模块,用于将所述业务数据按照图表库的格式封装为查询结果;  
[0079] 查询结果发送子模块,用于将所述查询结果发送至所述客户端,以调用所述图表库读取所述查询结果、生成指定的图表进行展示。

[0080] 本发明实施例包括以下优点:

[0081] 在本发明实施例中,当接收到客户端发送的查询请求时,从查询请求中提取查询参数,在预置的分布式数据库中过滤符合查询参数的业务数据,将业务数据发送至客户端,以调用预置的图表库按照指定的图表进行展示,一方面,基于分布式数据库本身性能,通过过滤业务数据,避免计算匹配度分值,使得查询效率得到了极大的提升,上亿条数据进行查询或聚合统计,都可以秒级查询出结果,在查询效率提高的同时,统计维度也得到了提升,另一方面,将过滤出的业务数据使用图表进行展示,提高了业务数据展示的直观性,从而提高了用户体验。

### 附图说明

[0082] 图1是本发明一个实施例的一种业务数据的搜索方法的步骤流程图;

[0083] 图2是本发明一个实施例的一种系统结构示意图;

[0084] 图3是本发明一个实施例的一种业务数据的搜索装置的结构框图。

### 具体实施方式

[0085] 为使本发明的上述目的、特征和优点能够更加明显易懂,下面结合附图和具体实施方式对本发明作进一步详细的说明。

[0086] 参照图1,示出了本发明一个实施例的一种业务数据的搜索方法的步骤流程图,具体可以包括如下步骤:

[0087] 步骤101,当接收到客户端发送的查询请求时,从所述查询请求中提取查询参数。

[0088] 如图2所示,本发明实施例可以应用在Web服务器中,该Web服务器可以连接一个或多个客户端,如Web客户端。

[0089] 用户在客户端中输入用户帐号、密码等信息,登录Web服务器,在Web页面输入相关的查询参数,生成查询请求,发送至Web服务器。

[0090] Web服务器若接收到客户端的查询请求时,则可以对查询请求进行解析,从中提取查询参数,如日期范围、金额范围、用户范围,等等。

[0091] 步骤102,在预置的分布式数据库中过滤符合所述查询参数的业务数据。

[0092] 如图2所示,在后台预先部署分布式数据库,在该分布式数据库中存储大量的业务数据。

[0093] 依据客户端的查询请求,将相关的查询参数封装为分布式数据库可读的查询参数,并发送至该分布式数据库,在该分布式数据库中过滤复合该查询参数的业务数据。

[0094] 基于分布式设计,随着业务数据的增大,可以给分布式数据库添加新的数据库节点,则可以保证业务数据的查询效率,仍能实时快速的查询出所需的业务数据。

[0095] 为使本领域技术人员更好地理解本发明实施例,在本说明书中,将ElasticSearch数据库作为分布式数据库的一种示例进行说明。

[0096] ElasticSearch数据库是一个基于Lucene的搜索服务器,它提供了一个分布式多

用户能力的全文搜索引擎。

[0097] 因Elasticsearch数据库是分布式的,通过管理节点来提供高扩展和高可用,一个集群(cluster)是一组具有相同cluster.name的节点集合,他们协同工作,共享数据并提供故障转移和扩展功能,因此,可以动态的进行集群环境扩充,且其能够达到实时快速稳定搜索,为业务数据的查询、展现提供了良好的基础。

[0098] 在本示例中,Web服务器可以使用Java语言,接收Web客户端的查询参数,将Web客户端的查询参数封装为ElasticSearch数据库的查询参数,并发送给ElasticSearch数据库,等待ElasticSearch数据库返回业务数据。

[0099] 在本发明的一个实施例中,步骤102可以包括如下子步骤:

[0100] 子步骤S11,将所述查询参数封装至过滤器filter中。

[0101] 子步骤S12,将所述过滤器filter封装至查询器query中。

[0102] 子步骤S13,调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数的业务数据。

[0103] 在本发明实施例中,采用ElasticSearch数据库等分布式数据库进行业务数据的存储和查询。

[0104] 为了使查询效率更快,当Web服务器封装查询参数时,将所有查询参数都放入过滤器filter的关键字中,然后将封装好的过滤器filter放入到查询器query的关键字中,而不是将所有查询参数直接放入查询器query的关键字中。

[0105] 因为查询器query的关键字中的查询参数在ElasticSearch数据库等分布式数据库进行查询时,会先计算业务数据的匹配度分值(0-1),然后与查询参数中的匹配度分值进行比较(默认为1),再返回合适的业务数据。

[0106] 而过滤器filter的关键字中的查询参数在ElasticSearch数据库等分布式数据库进行查询时,是直接判断是否满足查询参数,不需要计算匹配度分值,从而提高查询性能,而且,ElasticSearch数据库等分布式数据库默认会缓存过滤器filter的查询结果(即业务数据),如果后续的查询参数一致,将直接从缓存中取业务数据,从而进一步提高查询性能。

[0107] 在本发明的一个实施例中,子步骤S13进一步可以包括如下子步骤:

[0108] 子步骤S21,在预置的分布式数据库中查找所述查询参数对应的一个或多个存储区域。

[0109] 子步骤S22,调用所述查询器query在所述一个或多个存储区域进行搜索,以过滤符合所述查询参数的业务数据。

[0110] 在本发明实施例中,业务数据按照一定的方式进行拆分存储在相应的存储区域中,例如,日志数据按照天存储。

[0111] 在查询时,可以查询对应的一个或多个存储区域,从该存储区域中进行搜索,大大地减小数据搜索范围,进而提高查询效率。

[0112] 在本发明的一个实施例中,子步骤S13进一步可以包括如下子步骤:

[0113] 子步骤S31,在预置的分布式数据库中查找所述查询参数对应的索引文件。

[0114] 子步骤S32,调用所述查询器query在所述索引文件对应的业务数据中进行搜索,以过滤符合所述查询参数的业务数据。

[0115] 在本发明实施例中,可以预先对业务数据生成索引文件,通过索引文件索引查询



参数对应的业务数据,在该业务数据中进一步过滤符合该查询参数的业务数据。

[0116] 其中,索引文件通过如下的至少一种方式生成:

[0117] 子步骤S41,将业务数据中指定的第一字段的类型设置为目标类型。

[0118] 子步骤S42,采用除所述目标类型之外的其他类型的字段生成索引文件。

[0119] 在此方式中,对于业务数据中不需要进行搜索的第一字段,如附件名称、备注等,将其类型定义为目标类型object,该类型的数据不会在分布式数据库中建立索引文件,而其他类型的数据则在分布式数据库中建立索引文件,从而节约索引空间,减少对分布式数据库的整体性能影响。

[0120] 和/或,

[0121] 子步骤S43,在业务数据中确定指定的第二字段。

[0122] 子步骤S44,对所述第二字段整体生成索引文件。

[0123] 在此方式中,对于业务数据中不需要分词的第二字段,不使用分词器,而是整个第二字段的数据生成索引文件,因为分词器会将该第二字段的数据按照进行拆分,拆分之后分别生成索引文件,这样不但会造成索引空间的浪费,更是对数据的查询效率产生影响。

[0124] 和/或,

[0125] 子步骤S45,判断业务数据是否为嵌套文档;若是,则执行子步骤S46,若否,则执行子步骤S47。

[0126] 子步骤S46,拆分所述业务数据。

[0127] 子步骤S47,采用所述业务数据生成索引文件。

[0128] 在此方式中,所谓嵌套文档,可以指在一条业务数据中嵌套两个或两个以上独立的文档。

[0129] 例如,在一条业务数据中,具有某个用户的用户信息,同时也具有该用户属下用户的用户信息,则该业务数据属于嵌套文档。

[0130] 虽然分布式数据库针对嵌套文档的查询和聚合都做了相应的支持,但是,也应该尽量避免这样情况,因为对嵌套文档做增加、修改或者删除时,整个文档都要重新被索引,嵌套文档越多,这带来的成本就越大,而且嵌套文档的会使查询效率降低。

[0131] 如果确定某条业务数据为嵌套文档,则对其进行拆分,使之成为独立的文档,否则,可以对其生成索引文件。

[0132] 在本发明的一个实施例中,分布式数据库具有多个数据库节点,子步骤S13进一步可以包括如下子步骤:

[0133] 子步骤S51,将每个数据库节点的内存设置为指定的值,以将指针限制在有效的长度内。

[0134] 子步骤S52,基于所述内存调用所述查询器query在所述数据库节点中进行搜索,以过滤符合所述查询参数的业务数据。

[0135] 在本发明实施例中,由于ElasticSearch数据库等分布式数据库对内存要求比较高,所以随着业务数据的数据量的增加,需要给ElasticSearch数据库等分布式数据库中的每个数据库节点(如ElasticSearch服务器)增加内存,也可采用集群部署方案,以保证查询效率。

[0136] 但是,单台的数据库节点(如ElasticSearch服务器)内存设置不要超过指定的值

(如32G),原因如下:

[0137] 由于ElasticSearch数据库等分布式数据库是基于Java语言开发,在Java中,所有的对象都分配在堆上,然后有一个指针引用它。

[0138] 指向这些对象的指针大小通常是CPU (Central Processing Unit/Processor,中央处理器) 字长的大小,指针指向了值的精确位置。

[0139] 对于32位系统,内存最大可使用4G,对于64位系统可以使用更大的内存。但是,64位系统的指针意味着更大的浪费,因为指针本身大了,不仅浪费内存,而且更大的指针在主内存和缓存器之间移动数据的时候,会占用更多的带宽。

[0140] Java使用一个叫内存指针压缩的技术来解决这个问题,它的指针不再表示对象在内存中的精确位置,而是表示偏移量。

[0141] 这意味着32位系统的指针可以引用40亿个对象,而不是40亿个字节。最终,也就是说堆内存长到32G的物理内存,也可以用32bit的指针表示。

[0142] 一旦越过30G-32G的边界,指针就会切回普通对象的指针,每个对象的指针都变长了,就会使用更多的CPU、内存、带宽等资源,也就是说,实际上失去了更多的内存。

[0143] 事实上,当内存到达40G-50G的时候,有效内存才相当于使用内存对象指针压缩技术时候的32G。

[0144] 所以,即便足够的内存,也尽量不要超过32G,因为它浪费了内存,降低了CPU的性能,还要让GC (Java垃圾回收机制) 应对大内存。

[0145] 在本发明的一个实施例中,子步骤S13进一步可以包括如下子步骤:

[0146] 子步骤S61,从所述查询请求中提取查询方式。

[0147] 子步骤S62,若所述查询方式为数据搜索,则确定数据条数。

[0148] 子步骤S63,调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数、且数量为所述数据条数的业务数据。

[0149] 子步骤S64,若所述查询方式为数据聚合,则调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数的、原始的业务数据。

[0150] 子步骤S65,将所述原始的业务数据进行聚合,作为新的业务数据进行输出。

[0151] 子步骤S66,禁止输出所述原始的业务数据。

[0152] 在本发明实施例中,客户端在生成查询请求时,指定了查询方式。

[0153] 若查询方式为数据搜索,即分页查询,确定每页查询的数据条数,从而过滤该数据条数内的业务数据。

[0154] 若查询方式为数据聚合,则过滤出合适的、原始的业务数据,则进行聚合操作,如合并某个字段的总量、统计业务数据的总数量等,作为新的业务数据,从分布式数据库中作为查询结果输出。

[0155] 同时,禁止输出原始的业务数据 (ElasticSearch数据库等分布式数据库的聚合查询默认输出原始的业务数据),从而提高查询效率。

[0156] 步骤103,将所述业务数据发送至所述客户端,以调用预置的图表库按照指定的图表进行展示。

[0157] 如图2所示,为了更友好地将业务数据展示给用户,Web服务器将查询到的业务数据,发送至客户端,客户端则可以调用预置的的图表库按照指定的图表进行展示,如条状

图、饼状图、扇形图、曲线图,等等。

[0158] 在本发明的一个实施例中,步骤103可以包括如下子步骤:

[0159] 子步骤S71,将所述业务数据按照图表库的格式封装为查询结果。

[0160] 子步骤S72,将所述查询结果发送至所述客户端,以调用所述图表库读取所述查询结果、生成指定的图表进行展示。

[0161] 在本发明实施例中,图表库已经通过Web页面的JSON脚本发送至客户端,若Web服务器从分布式数据库中过滤出所需的业务数据,则按照用户指定的图表的类型,将该业务数据封装为图表库可读的查询结果,并返回客户端。

[0162] 客户端调用该图表库,解析该查询结果,按照该类型的图表进行渲染、显示。

[0163] 为使本领域技术人员更好地理解本发明实施例,在本说明书中,将ECharts作为图表库的一种示例进行说明。

[0164] ECharts是一个Javascript的图表库,可以流畅的运行在个人电脑和移动设备上,兼容当前绝大部分浏览器,可以提高开发效率,降低维护成本。

[0165] ECharts提供直观,生动,可交互,可高度个性化定制的数据可视化图表,拖拽重计算、数据视图、值域漫游等特性大大增强了用户体验,赋予了用户对数据进行挖掘、整合的能力。

[0166] ECharts支持折线图(区域图)、柱状图(条状图)、散点图(气泡图)、K线图、饼图(环形图)、雷达图(填充雷达图)、和弦图、力导向布局图、地图、仪表盘、漏斗图、事件河流图等图表。

[0167] Web服务器将业务数据根据业务的需求,封装为Echarts便于解析的JSON字符串(查询结果),并返回给客户端,客户端调用Echarts解析该JSON字符串,渲染指定的图表进行显示。

[0168] 在本发明实施例中,当接收到客户端发送的查询请求时,从查询请求中提取查询参数,在预置的分布式数据库中过滤符合查询参数的业务数据,将业务数据发送至客户端,以调用预置的图表库按照指定的图表进行展示,一方面,基于分布式数据库本身性能,通过过滤业务数据,避免计算匹配度分值,使得查询效率得到了极大的提升,上亿条数据进行查询或聚合统计,都可以秒级查询出结果,在查询效率提高的同时,统计维度也得到了提升,另一方面,将过滤出的业务数据使用图表进行展示,提高了业务数据展示的直观性,从而提高了用户体验。

[0169] 进一步地,因为在MySQL或Oracle等关系型数据库中,查询上亿的数据,需要最少几十分钟乃至数小时,而复杂的聚合查询根本不可能完成,故本发明实施例采用ElasticSearch数据库等分布式数据库进行大数据存储及查询,这样不但使得业务数据在查询效率上提高,也可根据各种具体需求进行报表统计展示,使得业务数据得到更充分的利用。

[0170] 需要说明的是,对于方法实施例,为了简单描述,故将其都表述为一系列的动作组合,但是本领域技术人员应该知悉,本发明实施例并不受所描述的动作顺序的限制,因为依据本发明实施例,某些步骤可以采用其他顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于优选实施例,所涉及的动作并不一定是本发明实施例所必须的。

[0171] 参照图3,示出了本发明一个实施例的一种业务数据的搜索装置的结构框图,具体可以包括如下模块:

[0172] 查询参数提取模块301,用于当接收到客户端发送的查询请求时,从所述查询请求中提取查询参数;

[0173] 业务数据过滤模块302,用于在预置的分布式数据库中过滤符合所述查询参数的业务数据;

[0174] 业务数据发送模块303,用于将所述业务数据发送至所述客户端,以调用预置的图表库按照指定的图表进行展示。

[0175] 在本发明的一个实施例中,所述业务数据过滤模块302包括:

[0176] 过滤器封装子模块,用于将所述查询参数封装至过滤器filter中;

[0177] 查询器封装子模块,用于将所述过滤器filter封装至查询器query中;

[0178] 搜索子模块,用于调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数的业务数据。

[0179] 在本发明的一个实施例中,所述搜索子模块包括:

[0180] 存储区域查询单元,用于在预置的分布式数据库中查找所述查询参数对应的一个或多个存储区域;

[0181] 存储区域搜索单元,用于调用所述查询器query在所述一个或多个存储区域进行搜索,以过滤符合所述查询参数的业务数据。

[0182] 在本发明的一个实施例中,所述搜索子模块包括:

[0183] 索引文件查找单元,用于在预置的分布式数据库中查找所述查询参数对应的索引文件;

[0184] 索引文件搜索单元,用于调用所述查询器query在所述索引文件对应的业务数据中进行搜索,以过滤符合所述查询参数的业务数据;

[0185] 其中,所述索引文件通过如下的至少一种方式生成:

[0186] 目标类型设置单元,用于将业务数据中指定的第一字段的类型设置为目标类型;

[0187] 第一文件生成单元,用于采用除所述目标类型之外的其他类型的字段生成索引文件;

[0188] 和/或,

[0189] 字段指定单元,用于在业务数据中确定指定的第二字段;

[0190] 第二文件生成单元,用于对所述第二字段整体生成索引文件;

[0191] 和/或,

[0192] 嵌套文档判断单元,用于判断业务数据是否为嵌套文档;若是,则调用业务数据拆分单元,若否,则调用第三文件生成单元;

[0193] 业务数据拆分单元,用于拆分所述业务数据;

[0194] 第三文件生成单元,用于采用所述业务数据生成索引文件。

[0195] 在本发明的一个实施例中,所述分布式数据库具有多个数据库节点,所述搜索子模块包括:

[0196] 内存设置单元,用于将每个数据库节点的内存设置为指定的值,以将指针限制在有效的长度内;

[0197] 内存搜索单元,用于基于所述内存调用所述查询器query在所述数据库节点中进行搜索,以过滤符合所述查询参数的业务数据。

[0198] 在本发明的一个实施例中,所述搜索子模块包括:

[0199] 查询方式提取单元,用于从所述查询请求中提取查询方式;

[0200] 数据条数确定单元,用于若所述查询方式为数据搜索,则确定数据条数;

[0201] 条数过滤单元,用于调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数、且数量为所述数据条数的业务数据;

[0202] 原始数据过滤单元,用于若所述查询方式为数据聚合,则调用所述查询器query在预置的分布式数据库中进行搜索,以过滤符合所述查询参数的、原始的业务数据;

[0203] 原始数据聚合单元,用于将所述原始的业务数据进行聚合,作为新的业务数据进行输出;

[0204] 原始数据禁止输出单元,用于禁止输出所述原始的业务数据。

[0205] 在本发明的一个实施例中,所述业务数据发送模块303包括:

[0206] 查询结果封装子模块,用于将所述业务数据按照图表库的格式封装为查询结果;

[0207] 查询结果发送子模块,用于将所述查询结果发送至所述客户端,以调用所述图表库读取所述查询结果、生成指定的图表进行展示。

[0208] 对于装置实施例而言,由于其与方法实施例基本相似,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

[0209] 在本发明实施例中,当接收到客户端发送的查询请求时,从查询请求中提取查询参数,在预置的分布式数据库中过滤符合查询参数的业务数据,将业务数据发送至客户端,以调用预置的图表库按照指定的图表进行展示,一方面,基于分布式数据库本身性能,通过过滤业务数据,避免计算匹配度分值,使得查询效率得到了极大的提升,上亿条数据进行查询或聚合统计,都可以秒级查询出结果,在查询效率提高的同时,统计维度也得到了提升,另一方面,将过滤出的业务数据使用图表进行展示,提高了业务数据展示的直观性,从而提高了用户体验。

[0210] 进一步地,因为在MySQL或Oracle等关系型数据库中,查询上亿的数据,需要最少几十分钟乃至数小时,而复杂的聚合查询根本不可能完成,故本发明实施例采用ElasticSearch数据库等分布式数据库进行大数据存储及查询,这样不但使得业务数据在查询效率上提高,也可根据各种具体需求进行报表统计展示,使得业务数据得到更充分的利用。

[0211] 本说明书中的各个实施例均采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似的部分互相参见即可。

[0212] 本领域内的技术人员应明白,本发明实施例的实施例可提供为方法、装置、或计算机程序产品。因此,本发明实施例可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本发明实施例可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0213] 本发明实施例是参照根据本发明实施例的方法、终端设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图

中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理终端设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理终端设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0214] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理终端设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0215] 这些计算机程序指令也可装载到计算机或其他可编程数据处理终端设备上,使得在计算机或其他可编程终端设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程终端设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0216] 尽管已描述了本发明实施例的优选实施例,但本领域内的技术人员一旦得知了基本创造性概念,则可对这些实施例做出另外的变更和修改。所以,所附权利要求意欲解释为包括优选实施例以及落入本发明实施例范围的所有变更和修改。

[0217] 最后,还需要说明的是,在本文中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者终端设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者终端设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者终端设备中还存在另外的相同要素。

[0218] 以上对本发明所提供的一种业务数据的搜索方法和一种业务数据的搜索装置,进行了详细介绍,本文中应用了具体个例对本发明的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本发明的方法及其核心思想;同时,对于本领域的一般技术人员,依据本发明的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本发明的限制。

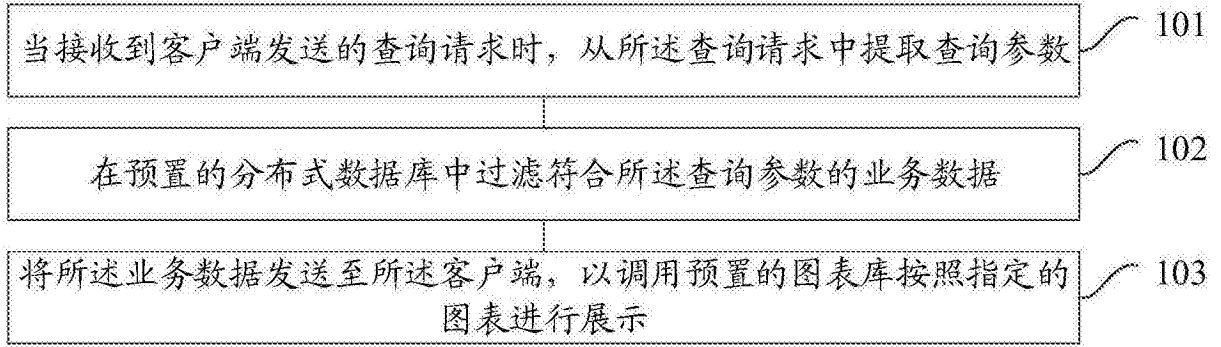


图1

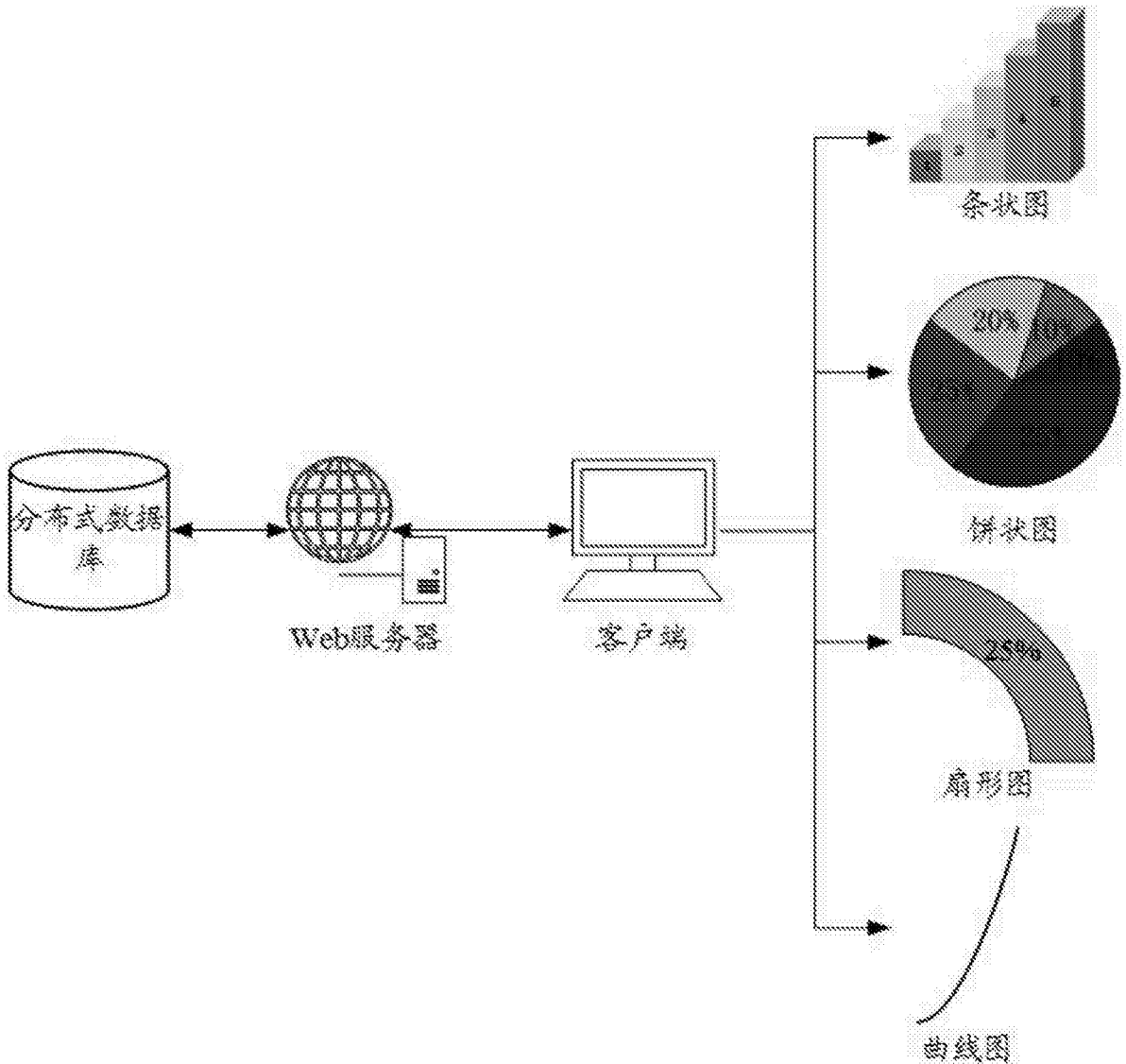


图2

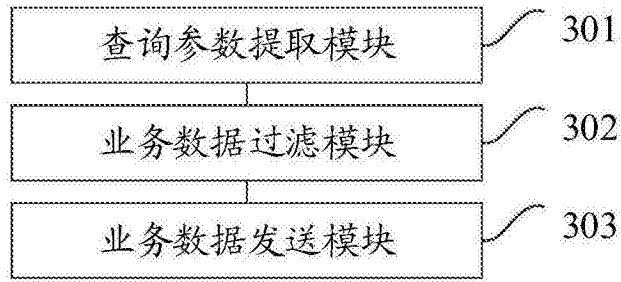


图3