



# (12)发明专利申请

(10)申请公布号 CN 110232106 A

(43)申请公布日 2019. 09. 13

(21)申请号 201910341618.7

(51)Int.Cl.

(22)申请日 2019.04.26

G06F 16/31(2019.01)

(71)申请人 安徽四创电子股份有限公司

G06F 16/33(2019.01)

地址 230088 安徽省合肥市高新技术产业  
开发区习友路3366号

G06F 16/332(2019.01)

(72)发明人 王利梅 唐飞 何丹娜 张海荣  
吴涛 李凯 闻志 吴爱清 沙健  
熊飞 刘江明 尚兵兵 王微  
杨良闯 周小将 陈迪宇 朱志超  
唐杰 王文杰 汪志发 舒贵阳  
李慧 胡昊

(74)专利代理机构 合肥和瑞知识产权代理事务  
所(普通合伙) 34118

代理人 王挺

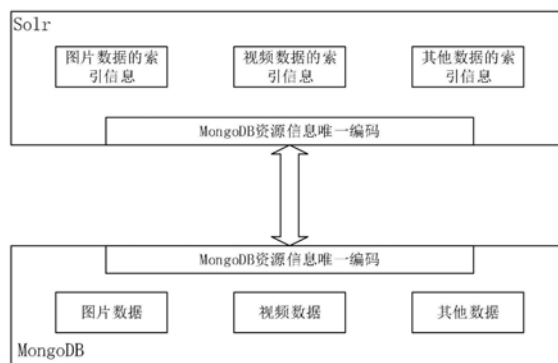
权利要求书1页 说明书4页 附图2页

## (54)发明名称

一种基于MongoDB和Solr的海量数据存储及  
快速检索方法

## (57)摘要

本发明公开了一种基于MongoDB和Solr的海量数据存储及快速检索方法,对于非结构化数据的存储,调用MongoDB数据库存储非结构化数据,并创建其对应的MongoDB资源信息唯一编码,同时,还调用Solr创建该非结构化数据的索引信息,并将该非结构化数据的索引信息和MongoDB资源信息唯一编码存储至Solr的索引库中;对于非结构化数据的快速检索,先根据查询语句在Solr的索引库中进行查找,查找出与该查询语句相对应的非结构化数据的索引信息和MongoDB资源信息唯一编码;再在MongoDB数据库中查找出该MongoDB资源信息唯一编码所对应的非结构化数据。本发明满足了日常业务中各种非结构化数据存储及快速检索的需求,实现了海量的非结构化数据的快速检索以快速获取所需要的信息。



1. 一种基于MongoDB和Solr的海量数据存储及快速检索方法,其特征在于,对于非结构化数据的存储,包括以下具体步骤:

S101,获取到非结构化数据后,调用MongoDB数据库存储该非结构化数据的信息,即将该非结构化数据存储至MongoDB数据库中,并为该非结构化数据创建与其对应的MongoDB资源信息唯一编码;

且每一个非结构化数据均对应一个MongoDB资源信息唯一编码;

S102,调用Solr创建该非结构化数据的索引信息,并将该非结构化数据的索引信息和MongoDB资源信息唯一编码存储至Solr的存储空间即索引库中;

对于非结构化数据的检索,包括以下具体步骤:

S201,输入检索条件进行非结构化数据的检索,Solr根据该检索条件创建查询语句;

S202,Solr先根据该查询语句在Solr的索引库中进行查找,查找出与该查询语句相对应的非结构化数据的索引信息,以及该索引信息所对应的MongoDB资源信息唯一编码;

S203,再根据该对应的MongoDB资源信息唯一编码在MongoDB数据库中进行查找,查找出该MongoDB资源信息唯一编码所对应的非结构化数据的信息。

2. 根据权利要求1所述的一种基于MongoDB和Solr的海量数据存储及快速检索方法,其特征在于,步骤S101中,调用MongoDB的API接口存储非结构化数据的信息;步骤S203中,调用MongoDB的API接口获取非结构化数据的信息。

3. 根据权利要求1所述的一种基于MongoDB和Solr的海量数据存储及快速检索方法,其特征在于,步骤S102中,调用SolrJ接口将该非结构化数据的索引信息和MongoDB资源信息唯一编码存储至Solr的索引库中。

4. 根据权利要求1所述的一种基于MongoDB和Solr的海量数据存储及快速检索方法,其特征在于,步骤S102中,调用Solr中的basic\_config配置创建该非结构化数据的索引信息。

5. 根据权利要求1所述的一种基于MongoDB和Solr的海量数据存储及快速检索方法,其特征在于,步骤S201中,调用SolrJ接口创建查询语句。

6. 根据权利要求1所述的一种基于MongoDB和Solr的海量数据存储及快速检索方法,其特征在于,非结构化数据包括视频数据、图片数据、其他数据;所述其他数据是指文档类的数据,包括Word、Excel、Pdf数据;

MongoDB数据库存储有不同类型的非结构化数据;Solr索引库存储有不同类型的非结构化数据的索引信息和MongoDB资源信息唯一编码。

7. 根据权利要求6所述的一种基于MongoDB和Solr的海量数据存储及快速检索方法,其特征在于,步骤S202中,SolrCloud平台接收所创建的查询语句,并在Solr的索引库中针对不同类型的非结构化数据进行分布式查找,查找出与该查询语句相对应的非结构化数据的索引信息,以及该索引信息所对应的MongoDB资源信息唯一编码。

8. 根据权利要求1所述的一种基于MongoDB和Solr的海量数据存储及快速检索方法,其特征在于,步骤S201中,引入Solr的中文分词机制,以实现全文检索。

## 一种基于MongoDB和Solr的海量数据存储及快速检索方法

### 技术领域

[0001] 本发明涉及大数据快速检索的技术领域,尤其是一种基于MongoDB和Solr的海量数据存储及快速检索方法。

### 背景技术

[0002] 随着计算机信息技术的发展及大数据的应用,在我们的日常生活中充斥着越来越多的数据,例如,对于公安机构而言,利用城市视频监控资源侦破案件并形成各种案件的资料信息,积累了海量的如图片、视频等非结构化数据。因此,如何从这些海量的非结构化数据中快速获取所需要的信息成为当前海量数据检索的一个热点及难点问题。

[0003] 经检索,中国专利公开号为CN106682148A公开了一种基于Solr数据搜索的方法及装置,该专利中提出:选用Hbase或MongoDB的底层存储数据库,并根据记录的字段生成rowkey;在查询数据时,根据Solr创建后的索引从底层存储中获取对应的数据。

[0004] 采用该专利所提出的方法是可以对数据进行存储和检索,该方法是直接用Solr创建的索引在底层存储数据库中进行检索,并对JVM内存及Solr的内存配置、磁盘占用和事务日志等方面进行优化,实现大数据的检索。但是,由于JVM内存及Solr的内存配置、磁盘占用和事务日志等相关内容本身就相对繁琐复杂,针对海量数据时,对这些方面的优化需要经过反复试验,不易操作实现。

### 发明内容

[0005] 为了克服上述现有技术中的缺陷,本发明提供一种基于MongoDB和Solr的海量数据存储及快速检索方法,满足了日常业务中各种非结构化数据存储及快速检索的需求,实现了海量的非结构化数据的快速检索以快速获取所需要的信息。

[0006] 为实现上述目的,本发明采用以下技术方案,包括:

[0007] 一种基于MongoDB和Solr的海量数据存储及快速检索方法,对于非结构化数据的存储,包括以下具体步骤:

[0008] S101,获取到非结构化数据后,调用MongoDB数据库存储该非结构化数据的信息,即将该非结构化数据存储至MongoDB数据库中,并为该非结构化数据创建与其对应的MongoDB资源信息唯一编码;

[0009] 且每一个非结构化数据均对应一个MongoDB资源信息唯一编码;

[0010] S102,调用Solr创建该非结构化数据的索引信息,并将该非结构化数据的索引信息和MongoDB资源信息唯一编码存储至Solr的存储空间即索引库中;

[0011] 对于非结构化数据的检索,包括以下具体步骤:

[0012] S201,输入检索条件进行非结构化数据的检索,Solr根据该检索条件创建查询语句;

[0013] S202,Solr先根据该查询语句在Solr的索引库中进行查找,查找出与该查询语句相对应的非结构化数据的索引信息,以及该索引信息所对应的MongoDB资源信息唯一编码;

[0014] S203,再根据该对应的MongoDB资源信息唯一编码在MongoDB数据库中进行查找,查找出该MongoDB资源信息唯一编码所对应的非结构化数据的信息。

[0015] 步骤S101中,调用MongoDB的API接口存储非结构化数据的信息;步骤S203中,调用MongoDB的API接口获取非结构化数据的信息。

[0016] 步骤S102中,调用SolrJ接口将该非结构化数据的索引信息和MongoDB资源信息唯一编码存储至Solr的索引库中。

[0017] 步骤S102中,调用Solr中的basic\_config配置创建该非结构化数据的索引信息。

[0018] 步骤S201中,调用SolrJ接口创建查询语句。

[0019] 非结构化数据包括视频数据、图片数据、其他数据;所述其他数据是指文档类的数据,包括Word、Excel、Pdf数据;

[0020] MongoDB数据库存储有不同类型的非结构化数据;Solr索引库存储有不同类型的非结构化数据的索引信息和MongoDB资源信息唯一编码。

[0021] 步骤S202中,SolrCloud平台接收所创建的查询语句,并在Solr的索引库中针对不同类型的非结构化数据进行分布式查找,查找出与该查询语句相对应的非结构化数据的索引信息,以及该索引信息所对应的MongoDB资源信息唯一编码。

[0022] 步骤S201中,引入Solr的中文分词机制,以实现全文检索。

[0023] 本发明的优点在于:

[0024] (1)本发明在存储非结构化数据时,不仅创建了该非结构化数据对应的MongoDB资源信息唯一编码,还调用Solr创建了该非结构化数据的索引信息,并且,将该非结构化数据的索引信息和MongoDB资源信息唯一编码存储至Solr的存储空间即索引库中,以方便在后续检索过程时,先在solr的索引库中进行一级检索,再在MongoDB数据库中进行二级检索,满足了日常业务中各种非结构化数据存储及快速检索的需求。

[0025] 本发明对比现有技术中的直接用Solr创建的索引在底层存储即MongoDB中检索数据的方式,当MongoDB中存储了海量的非结构化数据时,本发明能有效的提高海量数据的检索速度。

[0026] (2)本发明对比现有技术,无需对JVM内存及Solr的内存配置、磁盘占用和事务日志等方面进行全面优化,易操作实现。

[0027] (3)本发明在检索非结构化数据时,先利用查询语句在solr的索引库查找出对应的索引信息和MongoDB资源信息唯一编码,即先在solr的索引库中进行一级检索,再利用该MongoDB资源信息唯一编码在MongoDB数据库查找出对应的非结构化数据的信息,即再在MongoDB数据库中进行二级检索,以快速查找出所需的非结构化数据,从而实现海量非结构化数据的秒级检索。

[0028] (4)MongoDB数据库主要用于存储非结构化数据,是一个基于分布式文件存储的数据库,介于关系型数据库和非关系型数据之间,是非关系型数据库中功能最丰富,也最接近于关系型数据库的数据库;MongoDB资源信息唯一编码是指MongoDB数据库中的数据主键,类似关系型数据库的数据ID,利用数据主键查询数据使查询速度更快。

[0029] (5)本发明创建非结构化数据对应的MongoDB资源信息唯一编码,方便在MongoDB数据库可以根据MongoDB资源信息唯一编码进行检索,提高检索响应速度,缩减检索时间;

[0030] (6)本发明调用Solr中的basic\_config配置创建该非结构化数据的索引信息,方

便在solr的索引库中可以根据索引信息进行检索,提高检索响应速度,缩减检索时间。

[0031] (7) 本发明利用SolrCloud平台接收所创建的查询语句,并在Solr的索引库中针对不同类型的非结构化数据进行分布式查找,以实现集中式的配置信息、自动容错、近实时的搜索,以及查询时的自动负载均衡的功能。

[0032] (8) 本发明还引入了Solr的中文分词机制,方便快速实现全文检索,提高用户检索效率和工作效率。

## 附图说明

[0033] 图1为本发明一种基于MongoDB和Solr的海量数据存储及快速检索方法的整体架构图。

[0034] 图2为本发明的一种基于MongoDB和Solr的海量数据存储方法的流程图。

[0035] 图3为本发明的一种基于MongoDB和Solr的海量数据快速检索方法的流程图。

## 具体实施方式

[0036] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅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。

[0037] Solr是一个独立的企业级搜索应用服务器,基于Lucene的全文搜索,同时又对Lucene进行了扩展,提供了比Lucene更为丰富的查询语言,实现了可配置性和可扩展性,以及对查询性能进行了优化。其中,可配置是指Solr的部分属性通过配置即可实现使用;可扩展是指:当数据量较大时,过多的Solr索引也无法满足海量数据的快速检索,可以通过同时使用复制和分片,那么每个分片将对应一个master和若干slave,将Solr扩展实现快速检索。

[0038] Solr对外提供类似于Web\_service的API接口,即SolrJ接口,用户可以通过http请求,向搜索引擎服务器即Solr提交的XML文件创建索引信息;也可通过http GET操作提出查找请求即检索条件,并得到XML格式的返回结果,即在Solr索引库中的请求查找的结果

[0039] MongoDB数据库主要用于存储非结构化数据,是一个基于分布式文件存储的数据库,介于关系型数据库和非关系型数据之间,是非关系型数据库中功能最丰富,也最接近于关系型数据库的数据库。

[0040] 由图1和图3所示,一种基于MongoDB和Solr的海量数据存储方法,包括以下具体步骤:

[0041] S101,获取到非结构化数据后,利用MongoDB数据库存储该非结构化数据的信息,调用MongoDB的API接口将该非结构化数据存储至对应类型的MongoDB数据库后,并为该非结构化数据创建其对应的MongoDB资源信息唯一编码;且每一个非结构化数据均对应一个MongoDB资源信息唯一编码;

[0042] 所述MongoDB资源信息唯一编码是指MongoDB数据库中的数据主键,类似关系型数据库的数据ID,利用数据主键查询数据使查询速度更快;

[0043] S102,调用Solr中的basic\_config配置创建该非结构化数据的索引信息,并调用

SolrJ接口将该非结构化数据的索引信息和MongoDB资源信息唯一编码存储至Solr的存储空间即索引库中。

[0044] 由于,所述非结构化数据的类型包括:视频数据、图片数据、其他数据;所述其他数据是指Word、Excel、Pdf等非结构化数据;因此,所述MongoDB数据库存储有不同类型的非结构化数据;所述Solr的索引库中存储有不同类型的非结构化数据的索引信息和MongoDB资源信息唯一编码。

[0045] 由图2和图3所示,一种基于MongoDB和Solr的海量数据快速检索方法,包括以下具体步骤:

[0046] S201,输入检索条件进行非结构化数据的检索,Solr根据该检索条件创建查询语句,具体地,调用SolrJ接口创建查询语句;

[0047] 本实施例中,引入了Solr的中文分词机制,方便快速实现全文检索,提高用户检索效率和工作效率;

[0048] S202,SolrCloud平台接收所创建的查询语句,先根据该查询语句在Solr的索引库中针对不同类型的非结构化数据进行分布式查找,查找出与该查询语句相对应的非结构化数据的索引信息,以及该索引信息所对应的MongoDB资源信息唯一编码;

[0049] 所述SolrCloud是基于Solr和Zookeeper的分布式搜索方案,其主要思想是使用Zookeeper作为集群的配置信息中心,以实现集中式的配置信息、自动容错、近实时的搜索,以及查询时的自动负载均衡的功能。

[0050] S203,再根据该对应的MongoDB资源信息唯一编码在MongoDB数据库中进行查找,查找出该MongoDB资源信息唯一编码所对应的非结构化数据,并调用MongoDB的API接口获取非结构化数据的信息。

[0051] 以上仅为本发明创造的较佳实施例而已,并不用以限制本发明创造,凡在本发明创造的精神和原则之内所作的任何修改、等同替换和改进等,均应包含在本发明创造的保护范围之内。

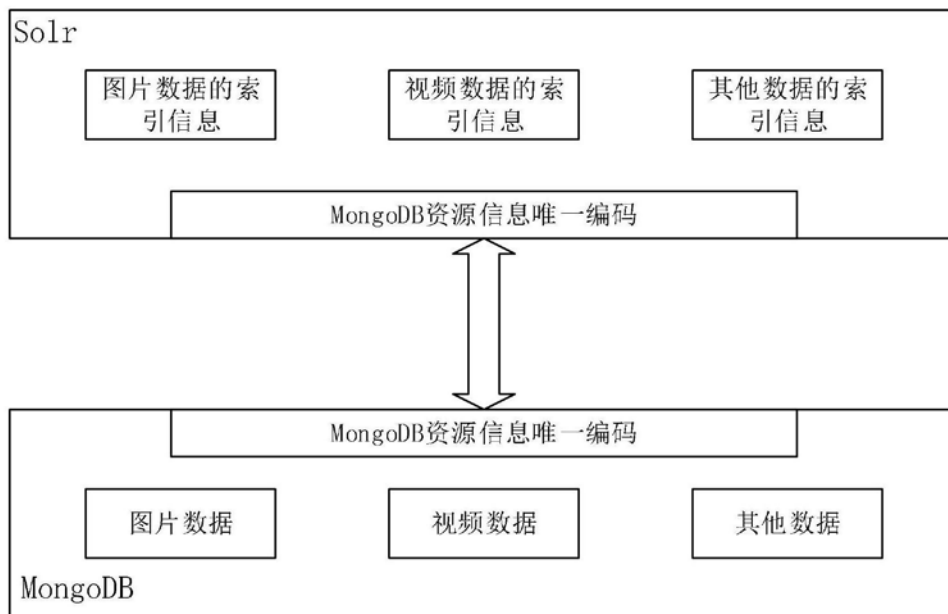


图1

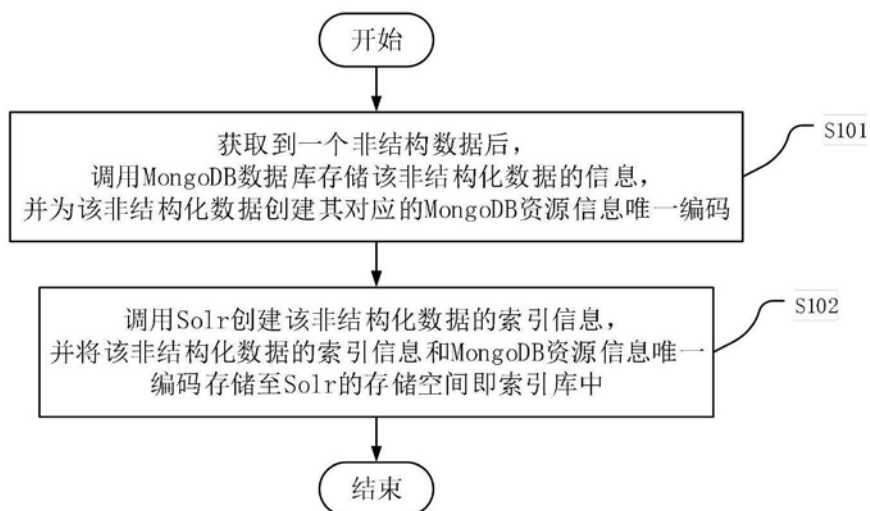


图2

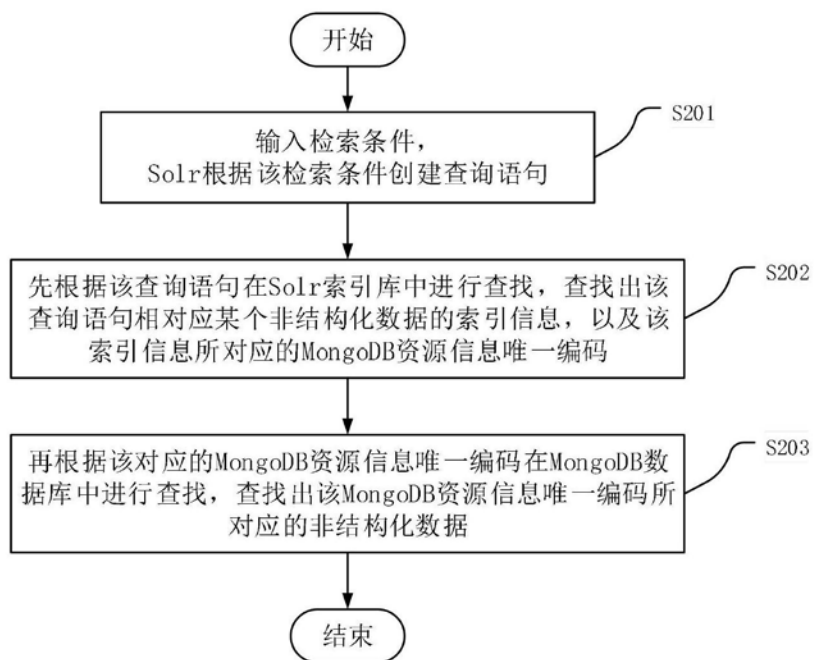


图3