



## (12)发明专利申请

(10)申请公布号 CN 110109870 A

(43)申请公布日 2019.08.09

(21)申请号 201810086508.6

(22)申请日 2018.01.24

(71)申请人 江苏友上科技实业有限公司

地址 210000 江苏省南京市雨花区西春路1号

(72)发明人 陈刚

(51)Int.Cl.

G06F 16/13(2019.01)

G06F 16/182(2019.01)

G06F 16/9535(2019.01)

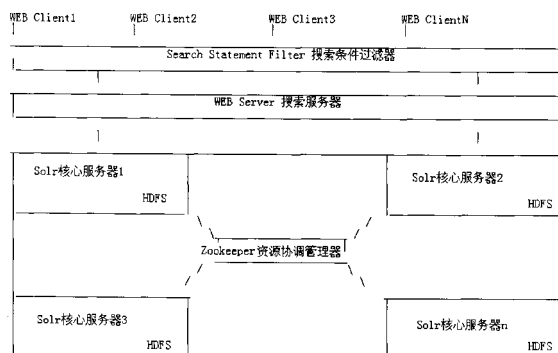
权利要求书1页 说明书3页 附图1页

### (54)发明名称

一种基于Solr的海量数据快速检索系统

### (57)摘要

本发明提供了一种基于Solr的海量数据快速检索系统,包括:Web客户端,其用于提供Web网页供用户输入关键词进行检索;搜索条件过滤器,其用于将从关键词形成搜索条件;Web搜索服务器,其用于将搜索条件发送到Solr核心服务器请求检索;多个Solr核心服务器,其用于接收搜索条件,对其进行分析,搜索到符合条件的数据,并将结果返回给Web搜索服务器;Zookeeper资源协调管理器,其用于将多个Solr核心服务器协同起来,进行海量数据的存储与检索、备份与容错;多个Solr核心服务器中还包括Hadoop分布式文件系统(HDFS),其用于将海量数据分割为块,存储于多个Solr核心服务器中,并将块的键值对映射到内存中。本发明实现了海量数据的快速索引功能。



1. 一种基于Solr的海量数据快速检索系统,其特征在于,包括Web客户端、搜索条件过滤器、Web搜索服务器、多个Solr核心服务器和Zookeeper资源协调管理器,其中:

所述Web客户端,其用于提供Web网页供用户输入关键词进行检索;

所述搜索条件过滤器,其用于将从所述关键词形成搜索条件;

所述Web搜索服务器,其用于将所述搜索条件发送到所述Solr核心服务器请求检索;

所述多个Solr核心服务器,其用于接收所述搜索条件,对其进行分析,搜索到符合条件的数据,并将结果返回给所述Web搜索服务器;

所述Zookeeper资源协调管理器,其用于将所述多个Solr核心服务器协同起来,进行所述海量数据的存储与检索、备份与容错;其中,

所述多个Solr核心服务器中还包括Hadoop分布式文件系统(HDFS),其用于将海量数据分割为块,存储于所述多个Solr核心服务器中,并将块的键值对映射到内存中。

2. 如权利要求1所述的基于Solr的海量数据快速检索系统,其中所述Web客户端为多个。

3. 如权利要求1所述的基于Solr的海量数据快速检索系统,其中所述海量数据为海量交通数据。

4. 如权利要求3所述的基于Solr的海量数据快速检索系统,其中所述SOLR核心服务器用于将车辆车牌、颜色、品牌等车辆信息,以及路口名称等路口信息形成的索引文件,利用SOLR的分词组件进行解析。

5. 如权利要求3或4所述的基于Solr的海量数据快速检索系统,其中所述SOLR核心服务器用于接收所述web客户端的查询内容,对其进行词法、语法、语言分析,然后搜索索引文件,搜索符合条件的数据,并将结果通过特定格式返回给用户。

6. 如权利要求5所述的基于Solr的海量数据快速检索系统,其中所述特定格式为xml或json格式。

## 一种基于Solr的海量数据快速检索系统

### 技术领域

[0001] 本发明涉及海量数据检索领域,尤其是一种基于Solr的海量数据快速检索系统。

### 背景技术

[0002] 随着互联网、物联网以及各种移动终端设备的快速发展,数据正以前所未有的速度迅猛的增长,如何存储并快速检索这庞大的数据是每个行业正面临的问题。目前的技术还是采用传统的关系型数据库,利用分布式关系型数据库的方式可以提升查询速度,但是这种方式的核心技术并没有改变,依然不能满足庞大数据的快速检索需要。

[0003] Solr作为高性能的搜索服务器,能够提供快速、强大的全文检索功能。Solr提供了针对大规模文档数据进行检索的查询语言,查询功能丰富,包括匹配单一字符、匹配0或多个字符、基于编辑距离的模糊查询、邻近查询(查找相隔一定距离的单词)、范围查询等等。Solr查询语法还支持多个查询条件的组合,以及查询的字段筛选、分页等特征。

[0004] 针对现有技术无法实现的海量数据实时快速检索问题,本发明提出一种基于Solr的海量数据快速检索系统。

### 发明内容

[0005] 本发明的目的在于提供一种基于Solr的海量数据快速检索系统。

[0006] 本发明的技术方案如下:

[0007] 一种基于Solr的海量数据快速检索系统,其特征在于,包括Web客户端、搜索条件过滤器、Web搜索服务器、多个Solr核心服务器和Zookeeper资源协调管理器,其中:

所述Web客户端,其用于提供Web网页供用户输入关键词进行检索;

所述搜索条件过滤器,其用于将从所述关键词形成搜索条件;

所述Web搜索服务器,其用于将所述搜索条件发送到所述Solr核心服务器请求检索;

所述多个Solr核心服务器,其用于接收所述搜索条件,对其进行分析,搜索到符合条件的数据,并将结果返回给所述Web搜索服务器;

所述Zookeeper资源协调管理器,其用于将所述多个Solr核心服务器协同起来,进行所述海量数据的存储与检索、备份与容错;其中,

所述多个Solr核心服务器中还包括Hadoop分布式文件系统(HDFS),其用于将海量数据分割为块,存储于所述多个Solr核心服务器中,并将块的键值对映射到内存中。

[0008] Web客户端可以为多个。

[0009] 海量数据可以为海量交通数据。

[0010] SOLR核心服务器用于将车辆车牌、颜色、品牌等车辆信息,以及路口名称等路口信息形成的索引文件,利用SOLR的分词组件进行解析。

[0011] SOLR核心服务器用于接收所述web客户端的查询内容,对其进行词法、语法、语言分析,然后搜索索引文件,搜索符合条件的数据,并将结果通过特定格式返回给用户。

[0012] 特定格式可以为xml或json格式。

[0013] 本发明的技术效果在于利用zookeeper和hdfs的分布式存储功能,利用多个Solr核心服务器实现了海量数据的存储,利用zookeeper和solr建立的全文分布式索引功能,实现了快速索引的功能。

## 附图说明

[0014] 图1为本发明的基于Solr的海量数据快速检索系统的示意图。

## 具体实施方式

[0015] 本发明提供了一种基于Solr的海量数据快速检索系统,如图1所示,包括Web客户端、搜索条件过滤器、Web搜索服务器、多个Solr核心服务器和Zookeeper资源协调管理器,其中:

Web客户端,其用于提供Web网页供用户输入关键词进行检索;

搜索条件过滤器,其用于将从关键词形成搜索条件;

Web搜索服务器,其用于将搜索条件发送到Solr核心服务器请求检索;

多个Solr核心服务器,其用于接收搜索条件,对其进行分析,搜索到符合条件的数据,并将结果返回给Web搜索服务器;

Zookeeper资源协调管理器,其用于将多个Solr核心服务器协同起来,进行所述海量数据的存储与检索、备份与容错;其中,

多个Solr核心服务器中还包括Hadoop分布式文件系统(HDFS),其用于将海量数据分割为块,存储于多个Solr核心服务器中,并将块的键值对映射到内存中。

[0016] 通过WEB客户端,用户可以通过Web网页根据不同的条件组合进行数据的快速检索。本系统主要是提供用户所需查询数据的过滤关键字。同时将用户输入的关键字形成的搜索条件向SOLR服务请求检索,最终将solr检索出的结果直观的展示给用户。

[0017] SOLR核心服务器提供了本系统最核心的搜索功能,能够实时的进行索引、全文检索、动态聚类、分面搜索等;在本系统中主要是:(1)将车辆车牌、颜色、品牌等车辆信息以及路口名称等路口信息形成的索引文件利用SOLR的分词组件进行解析,SOLR分词组件支持对中文文本类型的分词;(2)接收web客户端的查询内容,对其进行词法、语法、语言分析,然后搜索索引文件,搜索符合条件的数据,并将结果通过特定格式(xml或json格式)返回给请求者(用户)。

[0018] 通过该Zookeeper资源协调管理器,所有的SOLR服务器能够一起协同起来,实现海量数据的存储与检索、备份与容错,让所有的SOLR服务器能够实现任务的并行处理;在本系统中主要是存储索引配置文件、监控集群状态的改变。同时管理SOLR以及HDFS文件系统。

[0019] 本发明是以海量的交通数据,以及solr搜索引擎为基础,需要实现存储数据、搜索的功能。存储数据主要是将交通数据根据配置好的solr索引解析为索引文档数据存储于索引库中,通常作为hdfs文件存储(hdfs由zookeeper管理,解决了海量数据存储空间的问题)。搜索功能是提供web网页供用户输入关键字的查询,并将关键字的搜索条件与solr服务器进行交互,并且能够接收solr检索出的数据。

[0020] 为了实现该发明的具体功能,首先需要搭建solr服务平台,包括solr核心服务、zookeeper管理器、hdfs文件存储系统的环境搭建。其中在solr服务搭建中需要以交通数据

为基础根据具体需求,确定哪些字段需要被索引,配置solr索引文件。同时需要对索引数据进行分片,每个分片由多台服务器来共同完成,solr会确立每个分片的leader,形成分布式的索引,来大大降低索引的时间。其次要实现搜索功能,需要提供网页查询功能。网页查询功能使用springMVC的设计模式,通过tomcat来进行发布。其中需要进行html网页的开发,以及运用java接收用户填写的关键字搜索条件,将其解析为solr服务所需要的检索语言。

[0021] 最后应当说明的是,以上实施例仅用以说明本发明的技术方案而非对其限制;尽管参照较佳实施例对本发明进行了详细的说明,所属领域的普通技术人员应当理解,依然可对本发明的具体实施方式进行修改或者对部分技术特征进行等同替换而不脱离本发明技术方案的精神,其均应当涵盖本发明请求保护的技术方案范围当中。

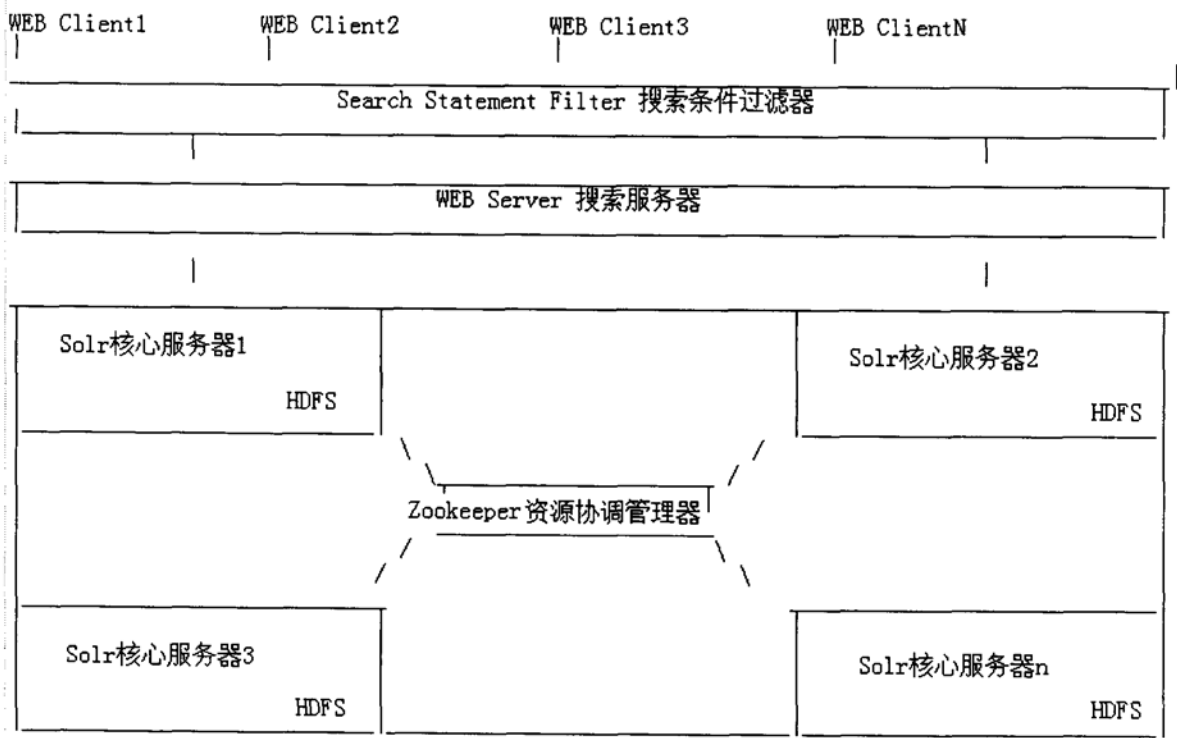


图1