



(12) 发明专利

(10) 授权公告号 CN 109902070 B

(45) 授权公告日 2023.12.12

(21) 申请号 201910060016.4

(22) 申请日 2019.01.22

(65) 同一申请的已公布的文献号
申请公布号 CN 109902070 A

(43) 申请公布日 2019.06.18

(73) 专利权人 华中师范大学
地址 430079 湖北省武汉市武昌区珞瑜路
152号

(72) 发明人 刘智 刘三女牙 杨宗凯 周东波
李敏士 孙建文 赵亮

(74) 专利代理机构 苏州达权专利代理事务所
(普通合伙) 32737
专利代理师 陆坚

(51) Int. Cl.
G06F 16/18 (2019.01)
G06F 16/182 (2019.01)
G06F 16/13 (2019.01)
G06F 16/14 (2019.01)
H04W 24/04 (2009.01)

(56) 对比文件

CN 106211226 A, 2016.12.07
CN 107391746 A, 2017.11.24
CN 108228755 A, 2018.06.29
CN 108920659 A, 2018.11.30
KR 20170067118 A, 2017.06.15
US 2006184529 A1, 2006.08.17
US 2011314148 A1, 2011.12.22
US 2018307576 A1, 2018.10.25
US 9305010 B1, 2016.04.05
CN 106503079 A, 2017.03.15
CN 107291928 A, 2017.10.24
CN 107660283 A, 2018.02.02
CN 103944973 A, 2014.07.23
US 2011191394 A1, 2011.08.04
US 2015310044 A1, 2015.10.29
a-du.logstash同步mongodb数据到
elasticsearch.《博客园URL:https://
www.cnblogs.com/a-du/p/8298418.html》
.2018, 网页第1-2页.

审查员 郑扬银

权利要求书1页 说明书3页 附图3页

(54) 发明名称

一种面向WiFi日志数据的解析存储搜索方法

(57) 摘要

本发明属于计算机信息技术领域,提供一种面向WiFi日志数据的解析存储搜索方法,包括:
(1) 提取解析WiFi日志的日期、时间、宿主IP、事件类型、消息类型、消息码、事件码、路由器地址和路由器IP; (2) 将解析的数据以键值对的形式保存至MongoDB分布式文件存储数据库; (3) 开启Logstash管道将MongoDB数据库复制到Elasticsearch搜索引擎中搜索; (4) 按照事务完整性导出MongoDB数据库的数据为二进制JSON文件,并加载到Hadoop平台。本发明方法能够对WiFi日志数据进行有效的解析和存储,同时提高了搜索的效率以及离线计算效率。

CN 109902070 B



1. 一种面向WiFi日志数据的解析存储搜索方法,其特征在于该方法包括以下步骤:

(1) 提取解析WiFi日志的日期、时间、宿主IP、事件类型、消息类型、消息码、事件码、路由器地址和路由器IP内容;具体包括:

S11、创建一个使用默认大小输入缓冲区的缓冲字符输入流,使用字符输入流读取日志文件直到读取为空,读取到一行日志后,判断日志文件的类型,如果是warning警告类型或者err错误类型,则解析出路由的空间实体位置和IP地址,随后与info信息类型和notice注意类型一起解析出日期、时间、宿主IP、事件类型、消息类型、消息码、事件码;

(2) 将解析的数据以键值对的形式保存至MongoDB分布式文件存储数据库;具体包括:

S21、使用列表List存储MongoDB数据库中的文档Document,当列表List中数量达到一万条进行一次插入数据库操作,直到所有数据插入完毕,不足一万条的数据最后插入;

(3) 开启Logstash管道将MongoDB数据库复制到Elasticsearch搜索引擎中搜索;具体包括:

S31、在MongoDB数据库中开启复制集,设定好副本集的名称后将MongoDB 数据库和副本集成员连接,初始化副本集;

S32、配置MongoDB数据库和Elasticsearch搜索引擎的地址与端口,然后同步Elasticsearch搜索引擎和MongoDB数据库中的数据;

(4) 按照事务完整性,使用shell编写文件定时导出 MongoDB 数据库的数据为二进制JSON文件,并加载到 Hadoop平台;具体包括:

S41、将 MongoDB 数据库的数据导出为二进制JSON文件;具体的,使用命令mongodump指定宿主host的IP地址、端口地址、数据库的名称、集合的名称以及导出的路径;

S42、将二进制JSON文件加载进分布式系统Hadoop的分布式文件系统HDFS当中;具体的,使用Hadoop的命令“hdfs dfs-put”导入二进制JSON文件;

S43、创建好映射表后,在Hadoop分布式集群中使用数据进行离线计算。

一种面向WiFi日志数据的解析存储搜索方法

技术领域

[0001] 本发明属于计算机信息技术领域,具体涉及一种面向WiFi日志数据的解析存储搜索方法。

技术背景

[0002] 在计算机软硬件迅速发展的今天,人们的移动设备都需要连接WiFi进行使用,对于WiFi产生的日志文件存在非常大的价值。对这些WiFi日志文件进行用户行为分析,可以对用户进行精准的提供个性服务,并且可以对设备管理人员提供必要的决策。同时,对大数据的处理分析正成为新一代信息技术融合应用的结点。通过对不同来源数据的管理、处理、分析与优化,将创造出巨大的经济和社会价值。

[0003] 然而,对于WiFi的日志数据,每天都会产生海量的数据,目前对这些海量的数据进行筛选和提取其中有实际价值的内容成为了一个巨大的难题,对大数据进行采集、导入、预处理、统计和挖掘的挑战在于数据量过大,现有技术并不能很好的及时解析和处理这些数据,从而导致无法进行后续的分析。

发明内容

[0004] 本发明的目的就是为了解决现有技术中的不足,提出一种面向WiFi日志数据的解析存储搜索方法,能够对WiFi日志数据进行有效的解析和存储,同时提高了搜索的效率以及离线计算效率。

[0005] 本发明的目的是通过以下技术方案实现的。

[0006] 一种面向WiFi日志数据的解析存储搜索方法,包括以下步骤:

[0007] (1) 提取解析WiFi日志的日期、时间、宿主IP、事件类型、消息类型、消息码、事件码、路由器地址和路由器IP等内容;

[0008] (2) 将解析的数据以键值对的形式保存至MongoDB分布式文件存储数据库;

[0009] (3) 开启Logstash管道将MongoDB数据库复制到Elasticsearch搜索引擎中搜索;

[0010] (4) 按照事务完整性导出MongoDB数据库的数据为二进制JSON文件,并加载到Hadoop平台。

[0011] 在上述技术方案中,步骤(1)具体包括:

[0012] S11、创建一个使用默认大小输入缓冲区的缓冲字符输入流,使用字符输入流读取日志文件直到读取为空,读取到一行日志后,判断日志文件的类型,如果是warning警告类型或者err错误类型,则解析出路由的空间实体位置和IP地址,随后与info信息类型和notice注意类型一起解析出日期、时间、宿主IP、事件类型、消息类型、消息码、事件码。

[0013] 在上述技术方案中,步骤(2)具体包括:

[0014] S21、使用列表List存储MongoDB数据库中的文档Document,当列表List中数量达到一万条进行一次插入数据库操作,直到所有数据插入完毕,不足一万条的数据最后插入。

[0015] 在上述技术方案中,所述步骤(3)具体包括:

[0016] S31、在MongoDB数据库中开启复制集,设定好副本集的名称后将MongoDB数据库和副本集成员连接,初始化副本集;

[0017] S32、配置MongoDB数据库和Elasticsearch搜索引擎的地址与端口,然后同步Elasticsearch搜索引擎和MongoDB数据库中的数据。

[0018] 在上述技术方案中,步骤(4)具体包括:

[0019] S41、将MongoDB数据库的数据导出为二进制JSON文件;

[0020] S42、将二进制JSON文件加载进分布式系统Hadoop的分布式文件系统HDFS当中。

[0021] 本发明与现有技术相比,具有如下优势:

[0022] 1、针对错误或警告等不同类型的WiFi日志数据有不同的解析操作,能够得到错误或警告等不同日志数据中的信息。

[0023] 2、同步到Elasticsearch搜索引擎中提供了更为强大的搜索功能。

[0024] 3、将MongoDB数据库的数据加载到Hadoop分布式文件系统HDFS,提高映射归约效率。

附图说明

[0025] 图1是本发明实施例的总体架构图。

[0026] 图2是本发明实施例的方法流程图。

[0027] 图3是本发明实施例的日志解析原理图。

[0028] 图4是本发明实施例的复制模块原理图。

[0029] 图5是本发明实施例的分布式模块原理图。

具体实施方式

[0030] 为使本发明的目的、技术方案和优点更加清楚,下面将结合附图对本发明实施方式作进一步地详细描述,但不以任何方式限制本发明的范围。

[0031] 如图1、2所示,本发明实施例提供了一种面向WiFi日志数据的解析存储搜索方法,包括以下步骤:

[0032] (1) 利用数据解析模块提取解析WiFi日志的日期、时间、宿主IP、事件类型、消息类型、消息码、事件码、路由器地址和路由器IP内容;

[0033] (2) 利用入库模块将解析的数据以键值对的形式保存至MongoDB分布式文件存储数据库;

[0034] (3) 利用复制模块开启Logstash管道将MongoDB数据库复制到Elasticsearch搜索引擎中搜索;

[0035] (4) 利用分布式模块按照事务完整性导出MongoDB数据库的数据为二进制JSON文件,并加载到Hadoop平台。

[0036] 如图3所示,在得到日志文件后,首先连接到MongoDB数据库,使用Java创建一个使用默认大小输入缓冲区的缓冲字符输入流,随后创建一个存储文档的列表。

[0037] 使用字符输入流读取日志文件直到读取为空。读取到一行日志后,判断日志文件的类型,如果是warning警告类型或者err错误类型,则解析出路由的空间实体位置和IP地址。随后与info信息类型和notice注意类型一起解析出日期、时间、宿主IP、事件类型、消息

类型、消息码、事件码。

[0038] 例如,提取出类型数组:["_id","date","host_IP","info_type","event","event_code","AP_address","AP_IP","AP_MAC"]。

[0039] 然后将解析的文档放入列表中,当列表的元素数量达到一万条后,直接插入进MongoDB数据库的集合中。最后当文档数目少于一万条时,再直接插入进集合。最后关闭数据库的连接。

[0040] 在MongoDB数据库中开启复制集,设定好副本集的名称后将MongoDB数据库和副本集成员连接,初始化副本集。

[0041] 如图4所示,在Logstash管道中,首先要编辑配置文件,在输入input中写入MongoDB数据库的路径uri为mongodb://username:password@ip:port,配置数据库的路径、名称与集合。在输出output中写入Elasticsearch搜索引擎的宿主hosts为ip:port,可选文档id和索引。在本实例中,在Logstash管道中对数据进行再次过滤,具体为配置过滤条件filter中定义时间date为时间戳的格式。随后开启服务,即可在Elasticsearch搜索引擎中进行文档的索引与搜索。

[0042] 例如,存储的键值对为:

```
[0043] {"_id":ObjectId("aaa"),"date":ISODate("2018-01-01T11:11:41.000Z"),"host_IP":"10.100.10.1","info_type":"debug","event":"warning","event_code":"501003","AP_address":"AP-5#-F4-0401","AP_IP":"10.100.20.100","AP_MAC":"10:10:10:10:10:10",}
```

[0044] 即可使用文档中的任意字段在搜索引擎中搜索出所有对应的数据。

[0045] 如图5所示,使用shell编写文件定时导出MongoDB数据库中的二进制JSON文件,具体为使用命令mongodump指定宿主host的IP地址、端口地址、数据库的名称、集合的名称以及导出的路径,然后使用Hadoop的命令“hdfs dfs-put”导入二进制JSON文件,创建好映射表后即可在Hadoop分布式集群中使用数据进行离线计算。

[0046] 本说明书中未作详细描述的内容,属于本专业技术人员公知的现有技术。

[0047] 应当说明的是,以上包含的本发明实施例的说明,是为了详细说明本发明的技术特征。在不脱离本发明原理的前提下,所作出的若干改进与修饰也受本发明的保护,因此本发明的保护范围应当以本申请的权利要求所界定的内容为准。

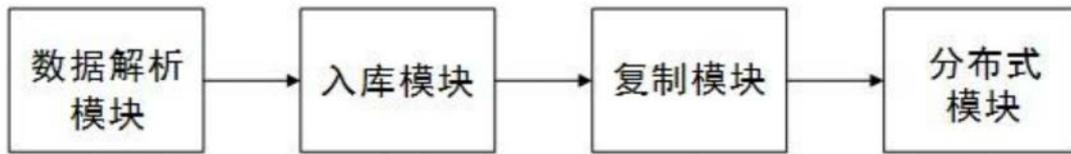


图1



图2

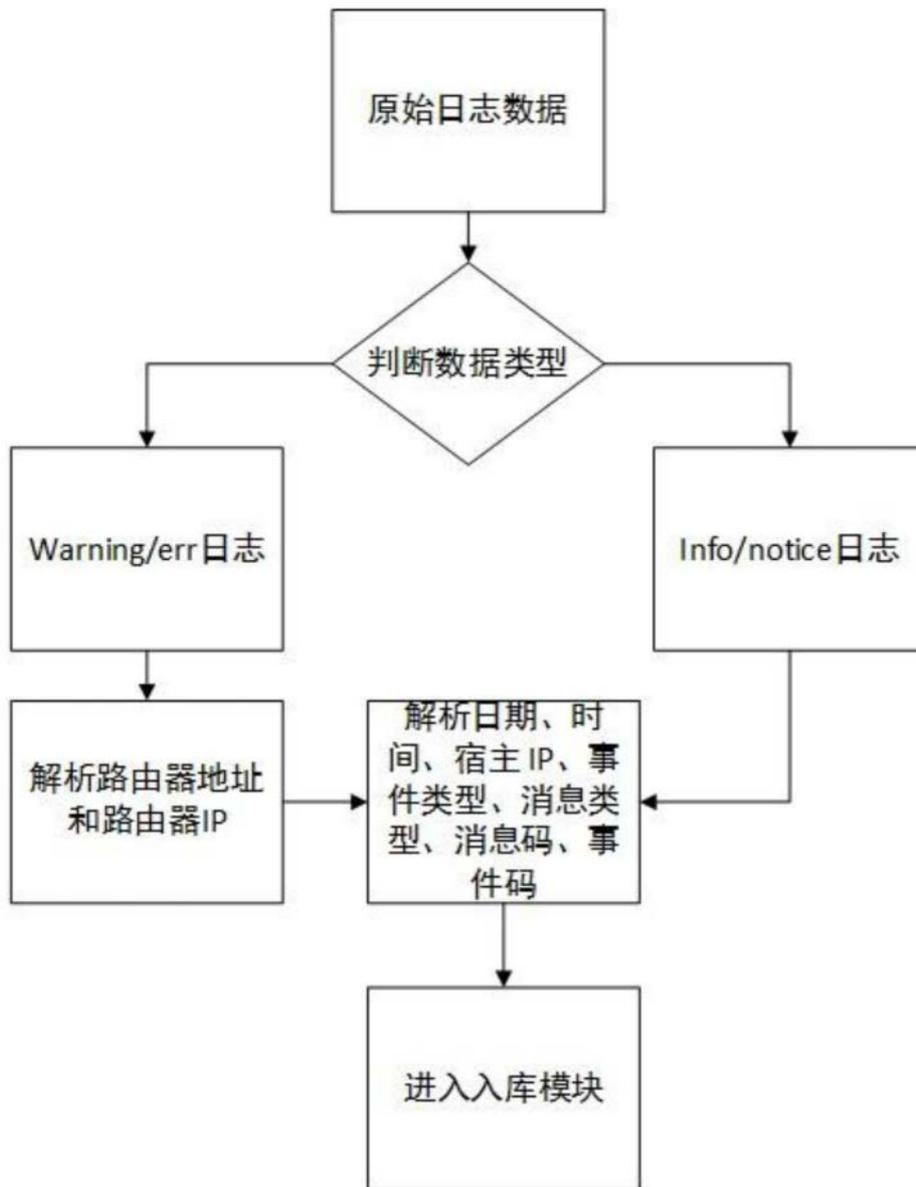


图3



图4



图5