(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 103678092 B
(45)授权公告日 2017.01.25

(21)申请号 201310743923.1

(22)申请日 2013.12.30

(65)同一申请的已公布的文献号
申请公布号 CN 103678092 A

(43)申请公布日 2014.03.26

(73)专利权人 北京网康科技有限公司
地址 100190 北京市海淀区中关村东路66号长城大厦A座3层

(72)发明人 金丰 斯俊伟 梁志勇 孙凌阁

(74)专利代理机构 北京亿腾知识产权代理事务所 11309
代理人 陈霁

(51)Int.Cl.
G06F 11/34(2006.01)
G06F 17/30(2006.01)
H04L 12/24(2006.01)

(56)对比文件
CN 102915269 A,2013.02.06,
CN 102053983 A,2011.05.11,
KR 10-1252471 B1,2013.04.03,
US 2013/0282788 A1,2013.10.24,

审查员 鲍旭恒

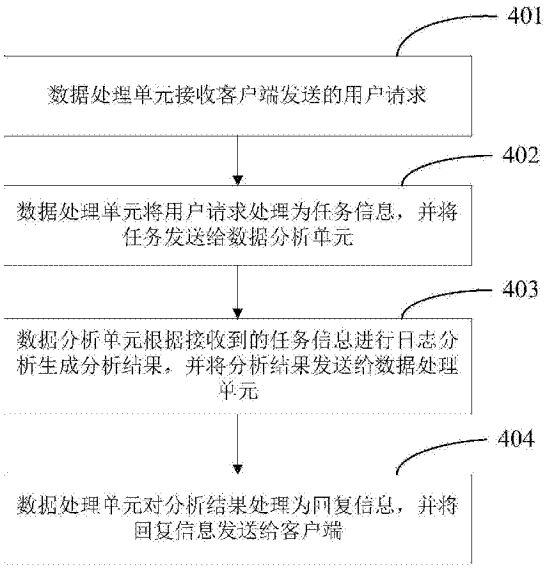
权利要求书2页 说明书6页 附图2页

(54)发明名称

日志分析方法及系统

(57)摘要

本发明涉及一种日志分析方法及系统，其日志分析方法包括以下步骤：数据处理单元接收客户端发送的用户请求；数据处理单元将用户请求处理为任务信息，并将所述任务信息发送给数据分析单元；数据分析单元根据接收到的任务信息进行日志分析生成分析结果，并将分析结果发送给数据处理单元；数据处理单元对分析结果处理为回复信息，并将回复信息发送给所述客户端。本发明的日志分析方法及系统具有实时性，可以对原始数据在任何维度进行日志信息分析，能够满足用户多维度数据查询的需求。



CN 103678092 B

1.一种日志分析方法，其特征在于，包括以下步骤：

数据处理单元接收客户端发送的用户请求；

所述数据处理单元用于判断所述用户请求是否转换为数据调用参数；

当所述用户请求没有转换为数据调用参数时，所述数据处理单元将所述用户请求转换为数据调用参数，将所述数据调用参数生成任务集合；

所述数据处理单元用于判断所述任务集合是否存储在本地缓存中；

当所述任务集合没有存储在本地缓存当中时，所述数据处理单元将所述任务集合对应的任务信息发送给数据分析单元；

所述数据处理单元将所述用户请求处理为任务信息，并将所述任务信息发送给所述数据分析单元；

所述数据分析单元根据接收到的所述任务信息进行日志分析生成分析结果，并将所述分析结果发送给数据处理单元；

所述数据处理单元对所述分析结果处理为回复信息，并将所述回复信息发送给所述客户端。

2.如权利要求1所述的日志分析方法，其特征在于，所述数据处理单元用于判断所述任务集合是否存储在本地缓存中具体包括：

如果所述任务集合存储在本地缓存当中，则将所述任务集合在本地的缓存关联至该任务；如果所述任务集合没有存储在本地缓存当中，则所述任务集合对应的任务信息发送给所述数据分析单元。

3.如权利要求1所述的日志分析方法，其特征在于，所述数据处理单元用于判断所述用户请求是否转换为数据调用参数具体还包括：

如果所述用户请求转换为数据调用参数，则直接返回；如果所述用户请求没有转换为数据调用参数，则将用户请求转换为数据调用参数，并在监控时间内实时监控本次日志分析的执行状态。

4.如权利要求3所述的日志分析方法，其特征在于，所述在监控时间内实时监控本次日志分析的执行状态具体包括：

如果在所述监控时间内，所述数据分析单元将分析结果发送给所述数据处理单元，则所述数据处理单元将当前状态更新为完成；

如果在所述监控时间内，所述数据分析单元将错误信息发送给所述数据处理单元，则所述数据处理单元将当前状态更新为失败，所述数据处理单元记录所述错误信息；

如果在所述监控时间内，所述数据分析单元没有发送信息给所述数据处理单元，则所述数据处理单元将当前状态更新为超时，并且所述数据处理单元记录所述超时。

5.如权利要求1所述的日志分析方法，其特征在于，所述数据处理单对将所述分析结果处理为所述回复信息具体包括：

所述数据处理单元接收所述数据分析单元发送的所述分析结果，所述数据处理单元对所述分析结果进行二次汇总分析，并将所述二次汇总分析的结果处理为回复信息。

6.一种日志分析系统，其特征在于，所述系统包括：

数据处理单元，用于用户请求的收集、分析，在判断所述用户请求没有转换为数据调用参数时，将所述用户请求转换为数据调用参数后，将所述数据调用参数生成任务集合，以及

在判断所述任务集合没有存储在本地缓存中时，将所述任务集合对应的任务信息发送给数据分析单元，并且接收所述数据分析单元发送的分析结果，将所述分析结果发送给客户端；

数据分析单元，用于接收所述数据处理单元发送的任务信息，根据所述任务信息对日志数据进行分析并生成所述分析结果，并将所述分析结果发送给所述数据处理单元。

7.如权利要求6所述的日志分析系统，其特征在于，所述数据处理单元包括：数据访问接口，数据收集调度模块，数据分析模块，第一存储模块，第二存储模块，数据收集模块；

所述数据访问接口，用于接收所述用户请求，并判断所述用户请求是否存储在本地缓存中；如果所述用户请求缓存在本地缓存中，则将所述用户请求返回；如果所述用户请求没有缓存在本地缓存中，则将所述用户请求转换为所述收集调度模块和收据分析模块的调用参数，并接收所述数据分析模块的分析结果；

所述数据收集调度模块，接收所述数据访问接口的调用参数，将来自所述数据访问接口的调用参数转换为任务集合，并下发任务集中的所有任务，用于所述用户请求的转换、监控、环境的初始化与清理；

所述数据分析模块，对所述第二存储模块中的汇总数据进行汇总分析，并将汇总数据分析的结果发送给所述数据访问接口，用于对所述数据处理单元中的所有数据进行二次汇总分析；

所述数据收集模块用于判断所述任务信息的每一任务信息是否已经缓存在第二存储模块中；如果，所述每一任务信息缓存在所述第二存储模块当中，则所述每一任务请求关联到对应的任务；如果所述每一任务信息没有缓存在所述第二存储模块当中，则将所述每一任务请求对应的任务信息下发给数据分析模块，并接收所述数据分析模块的分析结果，用于任务的下发与数据的收集；

所述第一存储模块，用于接收所述数据收集调度模块的所述任务集合并存储；

所述第二存储模块，用于对所述数据分析模块发送的分析结果进行存储。

日志分析方法及系统

技术领域

[0001] 本发明涉及计算机领域，尤其涉及一种日志分析方法及系统。

背景技术

[0002] 随着计算机技术的飞速发展，数据量迅速增加，数据的积累也越来越大，这也就意味着海量数据时代已经到来。在进行数据的传输、交换和处理过程中，安全性是一个非常重要的因素，因此，许多与信息处理相关的设备（如防火墙、入侵监测系统、路由器和服务器等）都会产生日志。日志记录了设备上和网络上每天发生的各种各样的事情，我们可以通过对日志的查询、统计等方法来了解各个设备和整个网络的状况。

[0003] 然而，面对海量的日志数据，如何才能更快更准确的统计、分析出想要获取的信息已经成为我们要面对的一个难题。为此，Princeton University的Ariel Rabkin等人提出了一种“分布式数据分析方法”，即将原始日志数据进行前期的聚集、汇总、入库、分析，从而降低了数据规模，从而达到了“实时性”的要求。但是，现有的日志分析技术中，大部分均以“日志数据集中存储”为前提，且日志分析并不具有实时性；尤其在广域网的环境下，由于“日志信息传输”与分析时“各设备间信息交互”更会导致实时性显著下降。而且现有技术中，不但在数据前期汇总时增加了数据的传输规模，而且付出了“损失其它维度数据信息”的代价，无法满足多维度数据查询的需求；并且现有技术还存在成本高、效率低及耗时等缺点。

发明内容

[0004] 本发明的目的是针对上述问题，提供了一种具有效率高、实时性强的日志分析方法及系统。

[0005] 为实现上述目的，本发明提供了一种日志分析方法，包括以下步骤：

[0006] 数据处理单元接收客户端发送的用户请求；

[0007] 所述数据处理单元将所述用户请求处理为任务信息，并将所述任务信息发送给所述数据分析单元；

[0008] 所述数据分析单元根据接收到的所述任务信息进行日志分析生成分析结果，并将所述分析结果发送给数据处理单元；

[0009] 所述数据处理单元对所述分析结果处理为回复信息，并将所述回复信息发送给所述客户端。

[0010] 优选地，所述数据处理单元将所述用户请求处理为任务信息具体包括：

[0011] 数据处理单元将所述用户请求转换为数据调用参数，将所述调用参数生成任务集合，并将所述任务集合对应的任务信息发送给数据分析单元。

[0012] 优选地，所述数据处理单元将所述用户请求处理为任务信息具体还包括：

[0013] 判断所述任务集合是否存储在本地缓存中；

[0014] 如果所述任务集合存储在本地缓存当中，则将所述任务集合在本地的缓存关联至

该任务；如果所述任务集合没有存储在本地缓存当中，则所述任务集合对应的任务信息发送给所述数据分析单元。

[0015] 优选地，所述数据处理单元将所述用户请求处理为任务信息具体还包括：

[0016] 判断所述用户请求是否转换为数据调用参数；

[0017] 如果所述用户请求转换为数据调用参数，则直接返回；如果所述用户请求没有转换为数据调用参数，则将用户请求没有转换为数据调用参数，并在监控时间内实时监控本次日志分析的执行状态。

[0018] 优选地，所述在监控时间内实时监控本次日志分析的执行状态具体包括：

[0019] 如果在所述监控时间内，所述数据分析单元将分析结果发送给所述数据处理单元，则所述数据处理单元将当前状态更新为完成；

[0020] 如果在所述监控时间内，所述数据分析单元将错误信息发送给数艘数据处理单元，则所述数据处理单元将当前状态更新为失败，所述数据处理单元记录所述错误信息；

[0021] 如果在所述监控时间内，所述数据分析单元没有发送信息给所述数据处理单元，则所述数据处理单元将当前状态更新为超时，并且所述数据处理单元记录所述超时。

[0022] 优选地，所述数据处理单对将所述分析结果处理为所述回复信息具体包括：

[0023] 所述数据处理单元接收所述数据分析单元发送的所述分析结果，所述数据处理单元对所述分析结果进行二次汇总分析，并将所述二次汇总分析的结果处理为回复信息。

[0024] 一种日志分析系统，所述系统包括：

[0025] 数据处理单元，用于用户请求的收集、分析，将所述用户请求转换为任务信息，并且接收数据分析单元发送的分析结果，将所述分析结果发送给客户端；

[0026] 数据分析单元，用于接收所述数据处理单元发送的任务信息，根据所述任务信息对日志数据进行分析并生成所述分析结果，并将所述分析结果发送给所述数据处理单元。

[0027] 优选地，所述数据处理单元包括：数据访问接口，数据收集调度模块，

[0028] 数据分析模块，第一存储模块，第二存储模块，数据收集模块；

[0029] 所述数据访问接口，用于接收所述用户请求，将用户请求转换为所述收集调度模块和数据分析模块的调用参数，并接收所数据分析模块的分析结果；

[0030] 所述数据收集调度模块，接收所述数据访问接口的调用参数，将来自所述数据访问接口的调用参数转换为任务集合，并下发任务集中的所有任务，用于所述用户请求的转换、监控、环境的初始化与清理；

[0031] 所述数据分析模块，对所述第二存储模块中的汇总数据进行汇总分析，并将汇总数据分析的结果发送给所述数据访问接口，用于对所述数据处理单元中的所有数据进行二次汇总分析；

[0032] 所述数据收集模块，接收所述数据收集调度模块下发的任务，将所述任务信息下发给所述数据分析模块，并接收所述数据分析模块的分析结果，用于任务的下发与数据的收集；

[0033] 所述第一存储模块，用于接收所述数据收集调用模块的所述任务集合并存储；

[0034] 所述第二存储模块，用于对所述数据分析模块发送的分析结果进行存储。

[0035] 本发明带来的有益效果是：本发明的日志分析方法及系统具有实时性，可以对原始数据在任何维度进行日志信息分析，能够满足用户多维度数据查询的需求。

附图说明

[0036] 图1为本发明一实施例的原理框图；

[0037] 图2为本发明实施例中日志分析方法的流程图；

[0038] 图3为本发明一实施例中日志分析系统部署的示意图；

[0039] 图4为本发明图1中的系统原理示意图。

具体实施方式

[0040] 下面通过附图和实施例，对本发明的技术方案做进一步的详细描述。

[0041] 图1是本发明一实施例的原理框图。

[0042] 如图1所示，采用本法对日志分析的一实施例中，用户通过客户端100发送用户请求11，客户端100将用户请求11发送给数据处理单元200。在本实施例中，用户请求11可以是日志的精确查询，也可以是数据的统计（如服务器IP地址、流量等），还可以是用户的自定义查询，根据用户的需要进行全维度的查询。以上用户请求11包括但不限于此。

[0043] 在图1中，数据处理短语200接收客户端100用户请求11，数据处理单元200会将用户请求11进行转换、存储及分析等操作，使用户请求11处理为任务信息12发送给数据分析单元300。数据分析单元300根据接收到的任务信息12对本地的日志数据进行分析，并将日志分析处理为分析结果13发送给数据处理单元200。数据处理单元200根据接收到来自数据分析单元300发送的分析结果13进行处理，并将分析结果300处理为回复信息14发送给客户端100。用户通过客户端100返回回复信息14得到所要查询的数据信息。

[0044] 图2是本发明一实施例的流程图。

[0045] 如图2所示，在步骤401中，数据处理单元200接收客户端100发送的用户请求11。用户将所需的信息通过客户端100以用户请求11的形式发送给数据处理单元。用户的请求信息可以是多样的，如上面所述，此处不再赘述。

[0046] 在图2中，步骤402为数据处理单元200将用户请求11处理为任务信息12，并将任务信息12发送给所述数据分析单元300。

[0047] 当数据处理单元200接收到来自客户端100发送的用户请求11之后，数据处理单元200首先判断并分析用户请求结果是否存在于本地缓存当中，如果数据处理单元200查找到用户请求结果已经存在于本地缓存当中，则直接返回。如果数据处理单元200经过判断分析本次用户请求结果未存在本地缓存当中，则数据处理单元200将用户请求11转换为调用参数，并且设置分析时间t，在时间t内实时监控本次分析的执行状态。

[0048] 如果在监控时间t内，数据处理单元200接收到来自数据分析单元300发送的分析结果13，则数据处理200单元将当前状态更新为完成。

[0049] 如果在监控时间t内，数据处理单元200接收到来自数据分析单元300发送的错误信息，则数据处理单元200将当前状态更新为失败，并且数据处理单元200记录错误信息。

[0050] 如果在监控时间t内，数据处理单元200没有接收到来自数据分析单元300发送的分析结果13，则数据处理单元将当前状态更新为超时，并且数据处理单元记录超时状态。

[0051] 数据处理单元200将调用参数转换为任务集合，并且存储任务集合。当一次用户请求11发送到数据处理单元200时，数据处理单元200会对本次用户请求11最后转换为的任务

集合进行判断分析，当本次任务集合存储在本地缓存当中时，数据处理单元200则将任务集合关联到本次任务当中；如果数据处理单元200经过判断分析没有发现本次任务集合在本地缓存中存储，则数据处理单元200将本次任务集合以任务信息12的形式发送给数据分析单元300。

[0052] 在步骤403中，数据分析单元300根据接收到的所述任务信息12进行日志分析生成分析结果13，并将分析结果13发送给数据处理单元200。

[0053] 在步骤404中，数据处理单元200对分析结果13处理为回复信息14，并将回复信息14发送给客户端100。数据处理单元200接收来自数据分析单元300发送的分析结果13，并对分析结果13进行二次汇总分析。数据处理单元将二次汇总分析的结果存储并处理为回复信息14发送给客户端100。

[0054] 如图3所示是本发明一实施例中日志分析系统部署的示意图，本发明实施例中日志分析的系统优选分布式存储，数据处理单元200部署在数据中心服务器310，数据分析单元300部署在广域网内的日志服务器320，各个设备上的日志信息存储在网络内的日志服务器320。数据分析单元300可以对存储在广域网内日志服务器上的原始日志数据进行前期的聚集、汇总、入库及分析等处理，可以降低日志数据的规模，能够达到实时性的效果。

[0055] 上述步骤为本发明日志分析方法的优选步骤，具体保护范围以权利要求书为依据。其中，步骤404与步骤403不一定是“先后”的顺序关系，即当数据分析单元300返回部分分析结果13时，步骤404也可以执行，而且可以在一次“分析过程”中执行多次，以满足用户数据查询的实时性需求。

[0056] 图4为本发明图1中一实施例的系统原理图。

[0057] 如图4所示，数据处理单元200包括：

[0058] 数据访问接口201，连接客户端100、数据收集调度模块202和数据分析模块203。数据访问接口201接受来自客户端的用户请求11，并将用户请求11解析为收集请求，用于对数据收集调度模块202和数据分析模块调用，并用于将回复信息14返回给客户端100。

[0059] 数据收集调度模块202，连接数据访问接口、第一存储模块及数据收集模块204，其用于接收数据访问接口201发送的收集请求，并将请求任务进行转换、监控，以及用于对环境的初始化与清理。数据收集调度模块202接收数据访问接口201发送的收集请求并将收集请求转化为可以在不同设备上执行的一系列任务信息，并将任务信息转换为任务集合存储在第一存储模块205。

[0060] 数据分析模块203，连接数据访问接口201和第二存储模块，其用于对数据中心服务器310的所有数据进行二次汇总分析。

[0061] 数据收集模块204，连接数据访问接口201、第一存储模块205、第二存储模块506及数据分析单元300，其用于任务集合的下发与数据收集。数据收集模块204管理各个设备的连接，并记录与每个设备连接的上下文信息，并且数据收集模块204将任务集合中的每个对应的任务信息下发到相应的设备当中。数据收集模块204也用于将数据分析单元300发送的分析结果13发送到第二存储模块。

[0062] 第一存储模块205，连接数据收集调度模块202和数据收集模块204，用于存储数据收集调度模块202生成的任务集合。

[0063] 第二存储模块206，连接数据收集调度模块202、数据分析模块203和数据收集模块

204，其用于存储数据分析单元300发送的数据，并且第二存储模块可用于数据收集调度模块202创建日志分析使用的临时数据结构。数据存储接收数据收集模块204发送的分析结果13并进行存储。

[0064] 如图4及本发明说明书其它附图所示，本发明的工作原理如下：

[0065] 用户将需要查询的信息(如日志的精确查询、统计信息和自定义查询等信息)通过客户端100输入，客户端100将接受到来自用户的查询信息并将转换为用户请求11发送给数据访问接口201，数据访问接口201将接收到的用户请求11解析为收集请求，用于对数据收集调度模块202和数据分析模块调用。数据访问接口201首先分析判断用户请求11是否存储在本地缓存当中，如果用户请求11存在于本地缓存当中，那么数据访问接口201会直接返回；如果用户请求11不存在与本地缓存当中，数据访问接口201，将用户请求11转换为收集请求，既数据收集调度模块202和数据分析模块203的调用参数。数据访问接口201首先调用数据收集调度模块202，并且设置监控时间t。如果在监控时间t内，数据处理单元200接收到来自数据分析单元300发送的分析结果13，则数据处理200单元将当前状态更新为完成。如果在监控时间t内，数据处理单元200接收到来自数据分析单元300发送的错误信息，则数据处理单元200将当前状态更新为失败，并且数据处理单元200记录错误信息。如果在监控时间t内，数据处理单元200没有接收到来自数据分析单元300发送的分析结果13，则数据处理单元将当前状态更新为超时，并且数据处理单元记录超时状态。

[0066] 数据访问接口201在监控时间内监控本次日志分析的执行状态。数据收集调度模块202接收数据访问接口201发送的收集请求并将收集请求转化为可以在不同设备上执行的一系列任务信息，并将任务信息转换为任务集合存储在第一存储模块205。数据收集模块204判断任务集合中的每一任务信息是否已经被缓存在第二存储模块当中，如果任务集合中的每一任务信息已经被缓存在第二存储模块当中，那么该任务请求关联到该任务；否则，如果任务集合中的任务信息没有缓存在第二存储模块当中，那么数据收集模块则将任务集合中对应的任务信息发送到数据分析单元300。

[0067] 数据分析单元300部署在日志服务器320，数据分析单元300将日志服务器320聚集到的设备上的日志信息进行数据分析，并将任务信息进行日志分析生成分析结果13，数据分析单元300将分析结果发送给数据收集模块204。在进行初次客户端发送用户请求11时，数据处理单元200会将该用户请求11对应的脚本模块发送到相关的日志服务器320。其中，用户请求11的解析是将用户请求11字符串解析为内部表述结构，并根据日志服务器320的类别，为相关的日志服务器320各生成一条任务信息，组成任务集合。

[0068] 数据收集模块204接收数据分析单元300发送的分析结果，数据收集模块204将当前的任务状态更新为完成，并且将分析结果13存储到第二存储模块206当中。数据分析模块203对存储在第二存储模块206当中的汇总数据进行分析，并将分析的结果发送到数据访问接口201，数据访问接口201将分析的结果处理为回复信息发送给客户端100。

[0069] 综上所述，采用本发明的日志分析方法及系统可以用于在海量的日志信息中实现实时的对所需信息进行查询，具有实时、快速和高效等优点。

[0070] 专业人员应该还可以进一步意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。

这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本发明的范围。

[0071] 结合本文中所公开的实施例描述的方法或算法的步骤可以用硬件、处理器执行的软件模块，或者二者的结合来实施。软件模块可以置于随机存储器(RAM)、内存、只读存储器(ROM)、电可编程ROM、电可擦除可编程ROM、寄存器、硬盘、可移动磁盘、CD-ROM、或技术领域内所公知的任意其它形式的存储介质中。

[0072] 以上所述的具体实施方式，对本发明的目的、技术方案和有益效果进行了进一步详细说明，所应理解的是，以上所述仅为本发明的具体实施方式而已，并不用于限定本发明的保护范围，凡在本发明的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

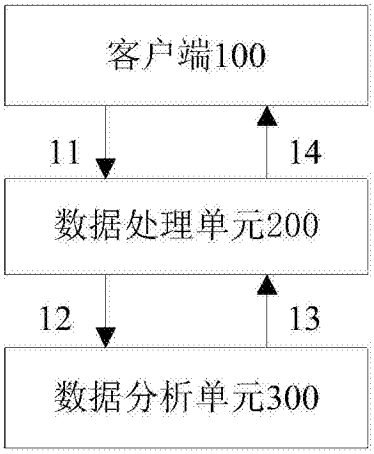

客户端100
11
14
数据处理单元200
12
13
数据分析单元300

图1

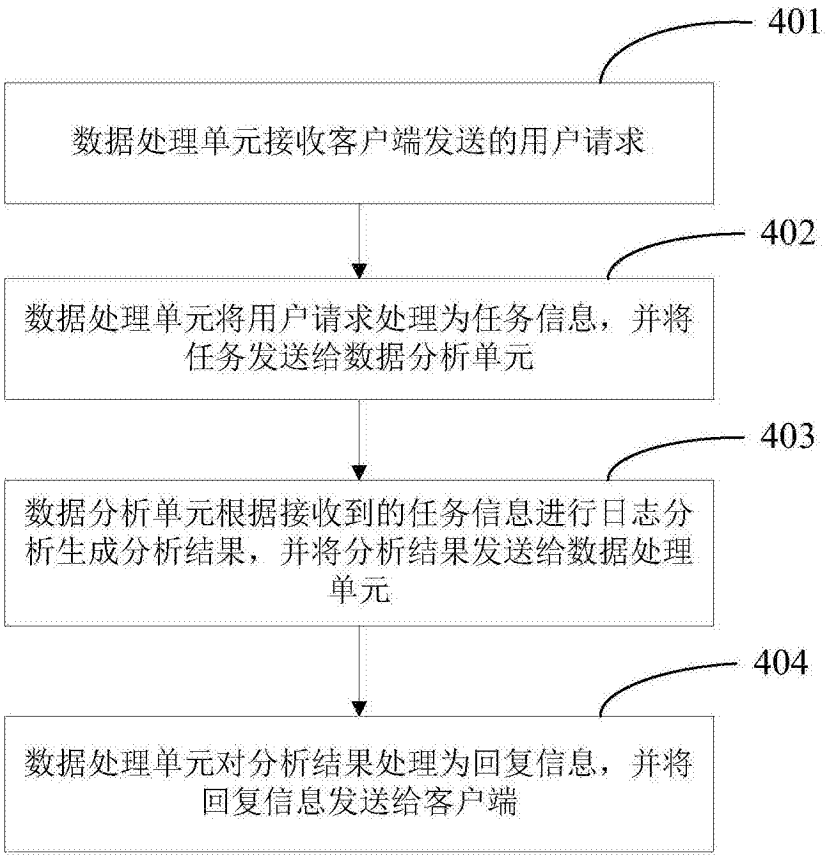

401
数据处理单元接收客户端发送的用户请求
402
数据处理单元将用户请求处理为任务信息，并将任务发送给数据分析单元
403
数据分析单元根据接收到的任务信息进行日志分析生成分析结果，并将分析结果发送给数据处理单元
404
数据处理单元对分析结果处理为回复信息，并将回复信息发送给客户端

图2

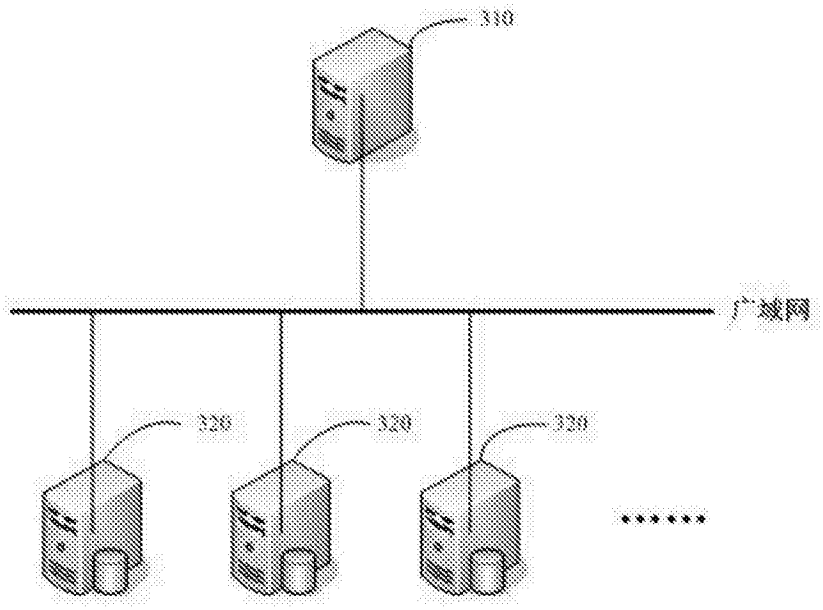

310
广域网
320
320
320

图3

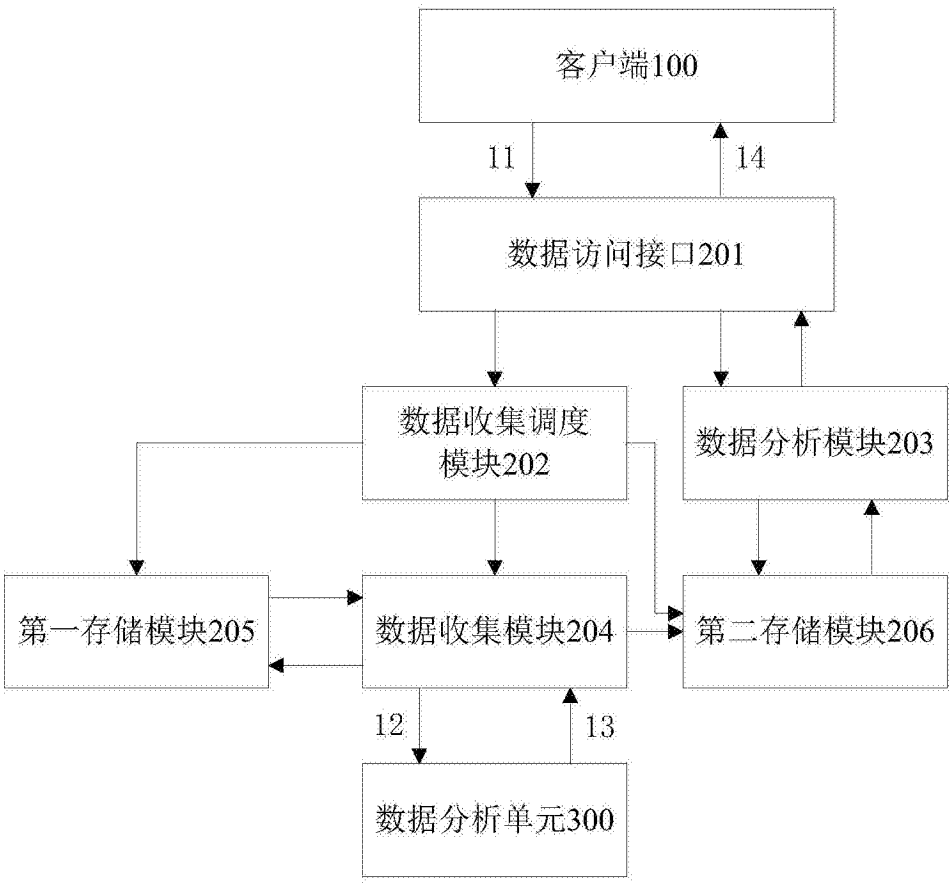

客户端100
11
14
数据访问接口201
数据收集调度模块202
数据分析模块203
第一存储模块205
数据收集模块204
第二存储模块206
12
13
数据分析单元300

图4