



(12) 发明专利申请

(10) 申请公布号 CN 102208991 A

(43) 申请公布日 2011. 10. 05

(21) 申请号 201010138020. 7

(22) 申请日 2010. 03. 29

(71) 申请人 腾讯科技(深圳)有限公司
地址 518057 广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室

(72) 发明人 刘占亮 赵琳霖 李锐

(74) 专利代理机构 广州三环专利代理有限公司
44202
代理人 郝传鑫 潘中毅

(51) Int. Cl.
H04L 12/24 (2006. 01)

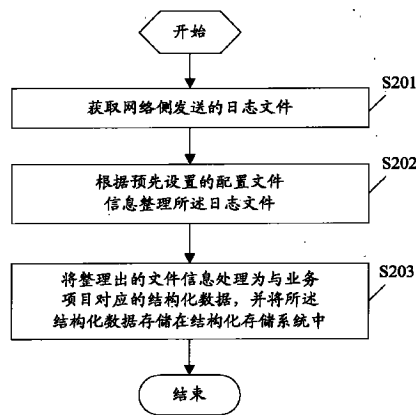
权利要求书 2 页 说明书 6 页 附图 3 页

(54) 发明名称

一种日志处理方法、设备和系统

(57) 摘要

本发明实施例公开了日志处理方法,包括:获取网络侧发送的日志文件;根据预先设置的配置文件信息整理所述日志文件;将整理出的文件信息处理为与业务项目对应的结构化数据,并将所述结构化数据存储于结构化存储系统中。相应地,本发明实施例还公开了一种日志处理设备以及日志处理系统,实施本发明实施例,解决了现有技术中开发成本高的问题,缩短了开发周期,提高了对海量日志的处理能力,通过实时地修改配置文件信息,能够快速敏捷灵活地满足各种统计需求,实时地得到统计结果,并且可以灵活地通过添加日志处理设备和结构化存储系统进行系统的扩容,而且支持业务热点的实时自动发现和自动预警,能够对系统业务模块的健康状况做到实时自动报警。



1. 一种日志处理方法,其特征在于,包括:
获取网络侧发送的日志文件;
根据预先设置的配置文件信息整理所述日志文件;
将整理出的文件信息处理为与业务项目对应的结构化数据,并将所述结构化数据存储
在结构化存储系统中。
2. 如权利要求 1 所述的方法,其特征在于,还包括:
根据业务项目的统计需求实时修改配置文件信息;
所述根据预先设置的配置文件信息整理所述日志文件的步骤包括:
根据所述配置文件信息将所述日志文件整理出业务项目统计需求的文件信息。
3. 如权利要求 1 或 2 所述的方法,其特征在于,所述将整理出的文件处理为与业务项目
对应的结构化数据的步骤包括:
周期性地调用业务项目对应的实时处理脚本分析整理出的文件信息,生成与业务项目
对应的结构化数据。
4. 如权利要求 3 所述的方法,其特征在于,所述生成与业务项目对应的结构化数据的
步骤之后,还包括:
根据所述结构化数据对业务的健康状况进行监控,并进行报警处理;或/和
根据所述结构化数据对业务热点进行自动发现和自动预警。
5. 一种日志处理设备,其特征在于,包括:
获取模块,用于获取网络侧发送的日志文件;
整理模块,用于根据预先设置的配置文件信息整理所述获取模块获取的日志文件;
结构化处理模块,用于将所述整理模块整理出的文件信息处理为与业务项目对应的结
构化数据,并将所述结构化数据存储于结构化存储系统中。
6. 如权利要求 5 所述的日志处理设备,其特征在于,还包括:
配置文件信息设置模块,用于根据业务项目的统计需求实时修改配置文件信息;
所述整理模块根据所述配置文件信息设置模块设置的配置文件信息将所述获取模块
获取的日志文件整理出业务项目统计需求的文件信息。
7. 如权利要求 6 所述的日志处理设备,其特征在于,所述结构化处理模块包括:
分析生成单元,用于周期性地调用业务项目对应的实时处理脚本分析所述整理模块整
理出的文件信息,生成与业务项目对应的结构化数据;
发送单元,用于将所述分析生成单元生成的所述结构化数据发送给结构化存储系统进
行存储。
8. 一种日志处理系统,包括网络侧服务器和日志分发器,其特征在于,所述日志处理系
统还包括至少一个日志处理设备,所述日志处理设备为如权利要求 5-7 任一项所述的日志
处理设备;
所述网络侧服务器将生成的日志文件发送给所述日志分发器,所述日志分发器按照负
载均衡策略将所述日志文件转发向所述日志处理设备,由所述日志处理设备对所述日志
文件进行处理。
9. 如权利要求 8 所述的日志处理系统,其特征在于,所述日志处理系统还包括至少一
个结构化存储系统,用于存储结构化数据。

10. 如权利要求 9 所述的日志处理系统,其特征在于,所述日志处理系统还包括:
监控报警设备,用于根据所述结构化数据对业务的健康状况进行监控,并进行报警处理;或/和
发现预警设备,用于根据所述结构化数据对业务热点进行自动发现和自动预警。

一种日志处理方法、设备和系统

技术领域

[0001] 本发明涉及通信领域,尤其涉及一种日志处理方法、设备和系统。

背景技术

[0002] 在现有的传统互联网业务中,日志是一个必不可少的部分,通过对业务日志的分析处理,我们可以了解到业务的经营状况,同时在服务器出现健康问题的时候,也可以通过日志定位问题所在,方便问题的解决。当前,日志文件都是以原始的格式存放于文件系统或者数据库系统,在对业务运营数据进行分析 and 统计的时候,需要对原始格式的日志数据文件进行处理分析,得到想要的统计结果。

[0003] 现有技术中互联网业务运营数据分析和统计以及 web 服务器、业务服务器运行状况监控的流程如图 1 所示:

[0004] 步骤 S101:提交产品运营数据统计分析、监控需求或产品业务模块性能、健康状况监控需求;

[0005] 步骤 S102:根据需求进行分析,设计针对特定需求的实施方案;

[0006] 步骤 S103:进行系统设计和系统编码,以对原始日志文件进行特定的分析和处理;

[0007] 步骤 S104:交付使用。

[0008] 对每一个业务运营数据统计的需求中,都需要经过图 1 所示的流程,针对特定的需求设计编码,从原始日志中获取想要的统计结果,从而导致开发成本高,开发周期很长。

发明内容

[0009] 本发明实施例在于提供一种日志处理方法、一种日志处理设备和一种日志处理系统,通过设置的配置文件信息对日志文件进行整理,生成业务项目对应的结构化数据,解决了现有技术中开发成本高的问题,缩短了开发周期。

[0010] 为了达到上述技术效果,本发明实施例提出了一种日志处理方法,包括:

[0011] 获取网络侧发送的日志文件;

[0012] 根据预先设置的配置文件信息整理所述日志文件;

[0013] 将整理出的文件信息处理为与业务项目对应的结构化数据,并将所述结构化数据存储于结构化存储系统中。

[0014] 优选地,所述日志处理方法还包括:

[0015] 根据业务项目的统计需求实时修改配置文件信息。

[0016] 优选地,所述根据预先设置的配置文件信息整理所述日志文件的步骤包括:

[0017] 根据所述配置文件信息将所述日志文件整理出业务项目统计需求的文件信息。

[0018] 优选地,所述将整理出的文件处理为与业务项目对应的结构化数据的步骤包括:

[0019] 周期性地调用业务项目对应的实时处理脚本分析整理出的文件信息,生成与业务项目对应的结构化数据。

- [0020] 优选地,所述生成与业务项目对应的结构化数据的步骤之后,还包括:
- [0021] 根据所述结构化数据对业务的健康状况进行监控,并进行报警处理;或/和
- [0022] 根据所述结构化数据对业务热点进行自动发现和自动预警。
- [0023] 相应地,本发明实施例公开了一种日志处理设备,包括:
- [0024] 获取模块,用于获取网络侧发送的日志文件;
- [0025] 整理模块,用于根据预先设置的配置文件信息整理所述获取模块获取的日志文件;
- [0026] 结构化处理模块,用于将所述整理模块整理出的文件信息处理为与业务项目对应的结构化数据,并将所述结构化数据存储于结构化存储系统中。
- [0027] 优选地,所述日志处理设备还包括:
- [0028] 配置文件信息设置模块,用于根据业务项目的统计需求实时修改配置文件信息。
- [0029] 优选地,所述整理模块根据所述配置文件信息设置模块设置的配置文件信息将所述获取模块获取的日志文件整理出业务项目统计需求的文件信息。
- [0030] 优选地,所述结构化处理模块包括:
- [0031] 分析生成单元,用于周期性地调用业务项目对应的实时处理脚本分析所述整理模块整理出的文件信息,生成与业务项目对应的结构化数据;
- [0032] 发送单元,用于将所述分析生成单元生成的所述结构化数据发送给结构化存储系统进行存储。
- [0033] 相应地,本发明实施例公开了一种日志处理系统,包括网络侧服务器和日志分发器,所述日志处理系统还包括至少一个日志处理设备,所述日志处理设备为上述的日志处理设备;
- [0034] 所述网络侧服务器将生成的日志文件发送给所述日志分发器,所述日志分发器按照负载均衡策略将所述日志文件转发向所述日志处理设备,由所述日志处理设备对所述日志文件进行处理。
- [0035] 优选地,所述日志处理系统还包括至少一个结构化存储系统,用于存储结构化数据。
- [0036] 优选地,所述日志处理系统还包括:
- [0037] 监控报警设备,用于根据所述结构化数据对业务的健康状况进行监控,并进行报警处理;或/和
- [0038] 发现预警设备,用于根据所述结构化数据对业务热点进行自动发现和自动预警。
- [0039] 实施本发明实施例,通过设置的配置文件信息对日志文件进行整理,生成业务项目对应的结构化数据,解决了现有技术中开发成本高的问题,缩短了开发周期,提高了对海量日志的处理能力,通过实时地修改配置文件信息,能够快速敏捷灵活地满足各种统计需求,实时地得到统计结果,并且可以灵活地通过添加日志处理设备和结构化存储系统进行系统的扩容,而且支持业务热点的实时自动发现和自动预警,能够对系统业务模块的健康状况做到实时自动报警。

附图说明

- [0040] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现

有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

- [0041] 图 1 是现有技术中日志处理方法的流程示意图；
- [0042] 图 2 是本发明实施例日志处理方法的流程示意图；
- [0043] 图 3 是本发明实施例的日志处理系统的结构示意图；
- [0044] 图 4 是本发明实施例的日志处理设备的结构示意图；
- [0045] 图 5 是本发明日志处理系统的另一实施例的结构示意图。

具体实施方式

[0046] 为了使本发明所要解决的技术问题、技术方案及有益效果更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。

[0047] 请参阅图 2 示出的本发明实施例日志处理方法的流程示意图,包括：

[0048] 步骤 S201 :获取网络侧发送的日志文件；

[0049] 具体地,日志文件是由网络侧服务器按照特定的格式生成,如,以“||”分割的“key = value”格式等,该网络侧服务器可以为 web 服务器或业务服务器等,相应的日志系统有 :log4j、log4cxx、log4cplus 等等,网络侧服务器可以通过用户数据包协议 (UDP, User Datagram Protocol) 方式将生成的日志文件发送给日志分发器,日志分发器根据负载均衡策略将收到的日志文件以 UDP 方式转发向本端的日志处理设备,日志处理设备获取到网络侧发送的日志文件。需要说明的是,该日志处理设备即为处理海量互联网服务日志信息的平台,具有通用的,可实时收集和海量日志信息功能。

[0050] 步骤 S202 :根据预先设置的配置文件信息整理所述日志文件；

[0051] 具体地,日志处理设备根据业务项目的统计需求预先设置配置文件信息,该配置文件信息包括需要统计的业务项目信息,即处理海量互联网服务日志信息的平台预先设有该配置文件信息,如 txt 格式的配置文件,技术人员只需按照格式填写对应的配置文件信息即可完成设置,如技术人员需要统计某业务的某项指标 (如数目),那么技术人员只需打开 txt 格式的配置文件按照格式填写数目项对应的配置文件信息即可,日志处理设备将根据设置好的数目项对应的配置文件信息整理所述日志文件,具体地,例如获取的日志文件是以“||”分割的“key = value”格式而生产的,那么日志处理设备将日志文件中的 key 为数目的相关信息进行整理,如,整理出以数目为文件名的一组文件。

[0052] 步骤 S203 :将整理出的文件信息处理为与业务项目对应的结构化数据,并将所述结构化数据存储于结构化存储系统中。

[0053] 具体地,日志处理设备对步骤 S202 中整理出的文件信息进行处理,将需要统计的项目处理为结构化数据,并存储于结构化存储系统中,结构化存储系统可以为关系型数据库,或者 BigTable 类型的结构化存储系统,如 hadoop/hbase 等。后续地,相关的业务数据分析设备可以根据存储于结构化存储系统中的结构化数据直接快速地进行数据的进一步分析,如进行业务数据报表的生成、业务数据对比、业务数据查询等相关服务,或者直接提供线上服务使用。

[0054] 需要说明的是,技术人员可以根据业务项目的统计需求实时地修改配置文件信

息,即,日志处理设备根据技术人员实时地修改操作可以实时地对配置文件信息进行删除、更改或添加统计项目或日志处理功能,实时修改的配置文件信息可以实时生效、无需重启日志处理设备,大大降低了现有技术中完成同样功能的开发所需的成本,简化了流程,节省了时间,从而提高了日志处理效率。

[0055] 还需要说明的是,日志处理设备可以通过启动多个实例来进行扩展,即可以灵活地通过添加日志处理设备和结构化存储系统进行扩展,日志分发器保证负载均衡。

[0056] 进一步地,步骤 S203 可以具体为,日志处理设备周期性地(如每 5 分钟)调用业务项目对应的实时处理脚本分析步骤 S202 整理出的文件信息,生成与业务项目对应的结构化数据,如,日志处理设备周期性地调用数目统计项目对应的实时处理脚本,分析以数目为文件名的一组文件,生成与数目统计项目对应的结构化数据,然后将所述结构化数据存储到结构化存储系统中。

[0057] 再进一步地,步骤 S203 后,监控报警设备可以根据某一周期内的结构化数据对业务的健康状况进行监控,并进行报警处理,或发现预警设备可以根据所述结构化数据对业务热点进行自动发现和自动预警,如热门榜单、热点关注、用户访客、用户轨迹等数据的统计,并通过邮件或者短信通知技术人员或其他相关用户,当前可以应用到搜吧、图片搜索、音乐搜索、视频搜索、博客搜索、QQ 书签、搜索华尔兹等等中去。

[0058] 上面详细说明了本发明日志处理方法的流程示意图,下面对应地,详细说明本发明实施例的日志处理系统的结构。

[0059] 如图 3 示出的本发明实施例的日志处理系统的结构示意图,日志处理系统包括网络侧服务器 31、日志分发器 32、至少一个日志处理设备 33 和至少一个结构化存储系统 34,具体地,日志处理系统也可以包括多个网络侧服务器 31 和日志分发器 32,本实施例以包括一个网络侧服务器 31、一个日志分发器 32、一个日志处理设备 33 以及一个结构化存储系统 34 为例进行说明,网络侧服务器 31 将生成的日志文件发送给日志分发器 32,日志分发器 32 按照负载均衡策略将所述日志文件转发向日志处理设备 33,由日志处理设备 33 对所述日志文件进行处理,日志处理设备 33 即为处理海量互联网服务日志信息的平台,具有通用的,可实时收集和处理海量日志信息功能,日志处理设备 33 将处理后的日志数据存储到结构化存储系统 34 中。

[0060] 日志处理设备 33 包括获取模块 331、整理模块 332 和结构化处理模块 333,其中

[0061] 获取模块 331 用于获取网络侧发送的日志文件;

[0062] 具体地,日志文件是由网络侧服务器 31 按照特定的格式生成,如,以“||”分割的“key = value”格式等,网络侧服务器 31 可以为 web 服务器或业务服务器等,相应的日志系统有 :log4j、log4cxx、log4cplus 等等,网络侧服务器 31 可以通过 UDP 方式将生成的日志文件发送给日志分发器 32,日志分发器 32 根据负载均衡策略将收到的日志文件以 UDP 方式转发向日志处理设备 33,日志处理设备 33 的获取模块 331 获取到日志文件。

[0063] 整理模块 332 用于根据预先设置的配置文件信息整理获取模块 331 获取的日志文件;

[0064] 具体地,日志处理设备 33 根据业务项目的统计需求预先设置配置文件信息,该配置文件信息包括需要统计的业务项目信息,即处理海量互联网服务日志信息的平台预先设有该配置文件信息,如 txt 格式的配置文件,技术人员只需按照格式填写对应的配置文件

信息即可完成设置,如技术人员需要统计某业务的某项指标(如数目),那么技术人员只需打开 txt 格式的配置文件按照格式填写数目项对应的配置文件信息即可,整理模块 332 将根据设置好的数目项对应的配置文件信息整理所述日志文件,具体地,例如获取模块 331 获取的日志文件是以“||”分割的“key = value”格式而生产的,那么整理模块 332 将日志文件中的 key 为数目的相关信息进行整理,如,整理出以数目为文件名的一组文件。

[0065] 结构化处理模块 333 用于将整理模块 332 整理出的文件信息处理为与业务项目对应的结构化数据,并将所述结构化数据存储于结构化存储系统 34 中。

[0066] 具体地,结构化存储系统 34 可以为关系型数据库,或者 BigTable 类型的结构化存储系统,如 hadoop/hbase 等。后续地,相关的业务数据分析设备可以根据存储在结构化存储系统 34 中的结构化数据可以直接快速地进行数据的进一步分析,如进行业务数据报表的生成、业务数据对比、业务数据查询等相关服务,或者直接提供线上服务使用。

[0067] 下面结合图 4,进一步说明本发明实施例中日志处理设备 33 的结构,日志处理设备 33 包括获取模块 331、整理模块 332 和结构化处理模块 333 外,还包括配置文件信息设置模块 334,用于根据业务项目的统计需求实时修改配置文件信息,具体地,技术人员可以根据业务项目的统计需求通过配置文件信息设置模块 334 实时地修改配置文件信息,即,配置文件信息设置模块 334 可以实时地对配置文件信息进行删除、更改或添加统计项目或日志处理功能,实时修改的配置文件信息可以实时生效、无需重启日志处理设备,大大降低了现有技术中完成同样功能的开发所需的成本,简化了流程,节省了时间,从而提高了日志处理效率;整理模块 332 根据配置文件信息设置模块 334 设置的配置文件信息将获取模块 331 获取的日志文件整理出业务项目统计需求的文件信息。

[0068] 进一步地,结构化处理模块 333 包括:分析生成单元 3331 和发送单元 3332,其中

[0069] 分析生成单元 3331 用于周期性地调用业务项目对应的实时处理脚本分析整理模块 332 整理出的文件信息,生成与业务项目对应的结构化数据;

[0070] 发送单元 3332 用于将分析生成单元 3331 生成的所述结构化数据发送给结构化存储系统 34 进行存储。

[0071] 具体地,分析生成单元 3331 周期性地(如每 5 分钟)调用业务项目对应的实时处理脚本分析整理模块 332 整理出的文件信息,生成与业务项目对应的结构化数据,如,分析生成单元 3331 周期性地调用数目统计项目对应的实时处理脚本,分析以数目为文件名的一组文件,生成与数目统计项目对应的结构化数据,然后通过发送单元 3332 将所述结构化数据存储于结构化存储系统 34 中。

[0072] 如图 5 示出的本发明日志处理系统的另一实施例的结构示意图,日志处理系统包括网络侧服务器 31、日志分发器 32、日志处理设备 33 和结构化存储系统 34 外,还包括监控报警设备 35 和发现预警设备 36,其中

[0073] 监控报警设备 35 用于根据所述结构化数据对业务的健康状况进行监控,并进行报警处理;

[0074] 发现预警设备 36 用于根据所述结构化数据对业务热点进行自动发现和自动预警。

[0075] 具体地,监控报警设备 35 可以根据某一周期内的结构化数据对业务的健康状况进行监控,并进行报警处理,以及发现预警设备 36 可以根据所述结构化数据对业务热点进

行自动发现和自动预警,如热门榜单、热点关注、用户访客、用户轨迹等数据的统计,并通过邮件或者短信通知技术人员或其他相关用户,当前可以应用到搜吧、图片搜索、音乐搜索、视频搜索、博客搜索、QQ 书签、搜索华尔兹等等中去。

[0076] 需要说明的是,日志处理设备 33 可以通过启动多个实例来进行扩展,日志处理系统即可以灵活地通过添加日志处理设备 33 和结构化存储系统 34 进行扩展,日志分发器 32 保证负载均衡。

[0077] 综上所述,通过上述实施例,根据设置的配置文件信息对日志文件进行整理,生成业务项目对应的结构化数据,解决了现有技术中开发成本高的问题,缩短了开发周期,提高了对海量日志的处理能力,通过实时地修改配置文件信息,能够快速敏捷灵活地满足各种统计需求,实时地得到统计结果,并且可以灵活地通过添加日志处理设备和结构化存储系统进行系统的扩容,而且支持业务热点的实时自动发现和自动预警,能够对系统业务模块的健康状况做到实时自动报警。

[0078] 需要说明的是,通过以上的实施方式描述,本领域的技术人员可以清楚地了解到本发明可借助软件加必需的硬件平台的方式来实现,当然也可以全部通过硬件来实施。基于这样的理解,本发明的技术方案对背景技术做出贡献的全部或者部分可以以软件产品的形式体现出来,该计算机软件产品可以存储在存储介质中,如 ROM/RAM、磁碟、光盘等,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本发明各个实施例或者实施例的某些部分所述的方法。

[0079] 以上所揭露的仅为本发明实施例中的较佳实施例而已,当然不能以此来限定本发明之权利范围,因此依本发明权利要求所作的等同变化,仍属本发明所涵盖的范围。

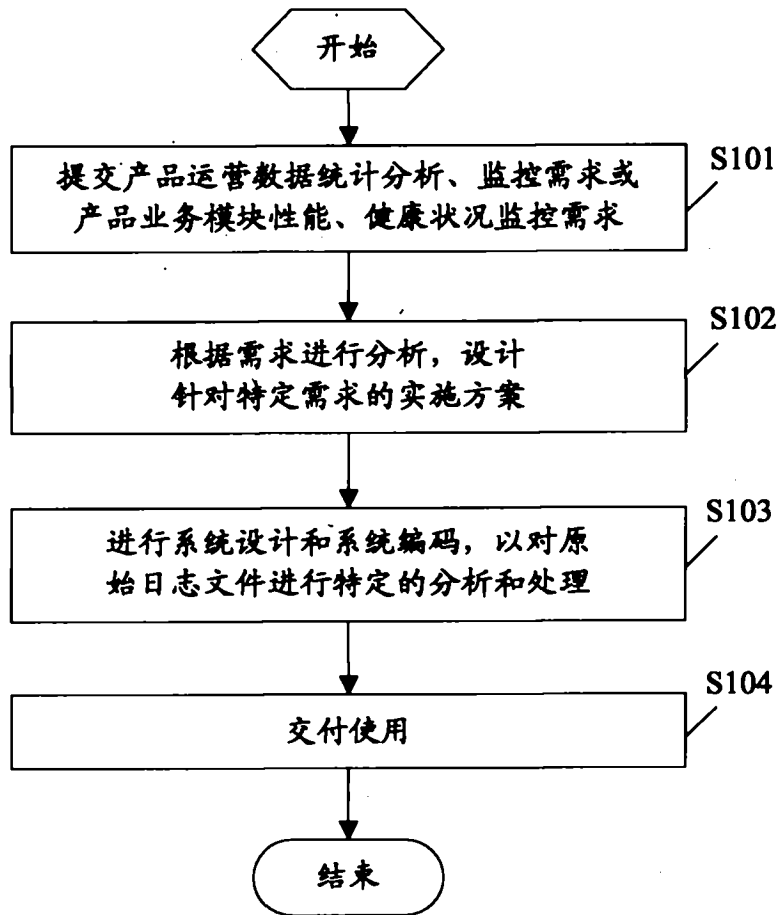


图 1

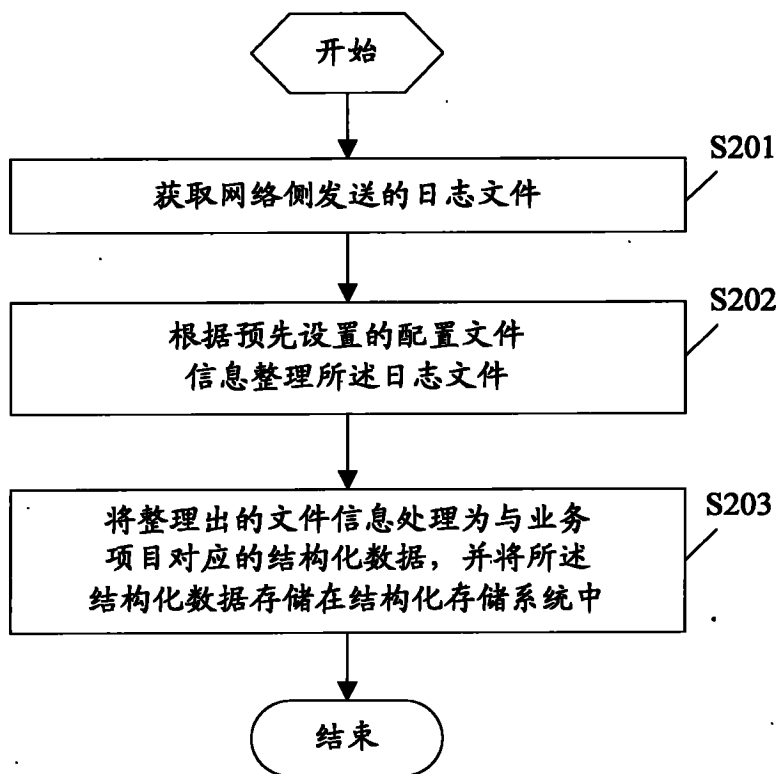


图 2

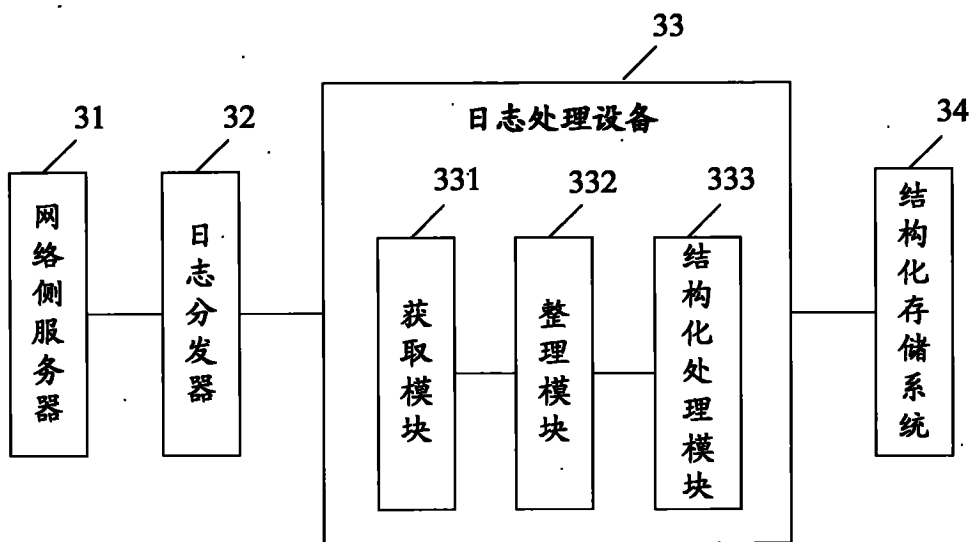


图 3

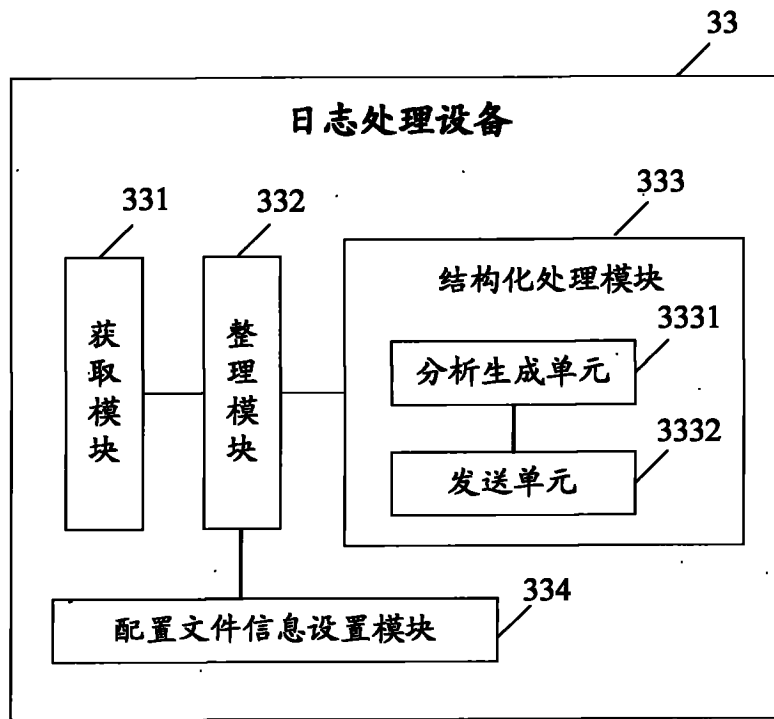


图 4

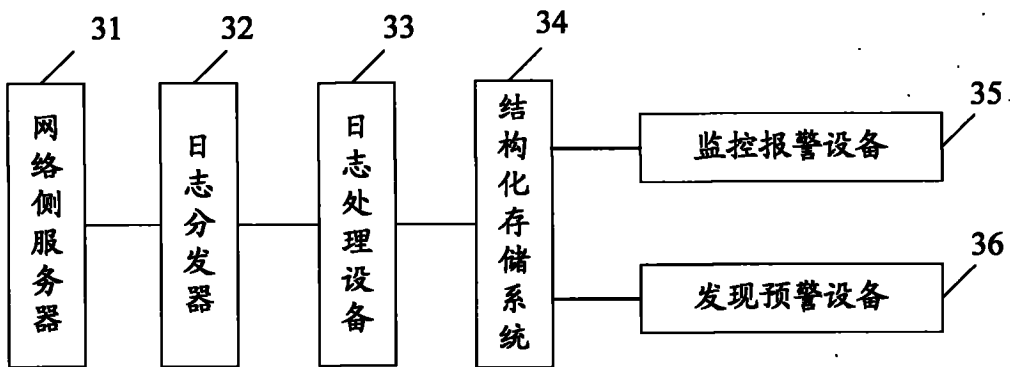


图 5