(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 106201848 A
(43)申请公布日 2016.12.07

(21)申请号 201610514809.5

(22)申请日 2016.06.30

(71)申请人 北京奇虎科技有限公司
地址 100088 北京市西城区新街口外大街28号D座112室(德胜园区)
申请人 奇智软件(北京)有限公司

(72)发明人 王义辉 徐胜国 王素梅 沈迪 李铮

(74)专利代理机构 北京市隆安律师事务所 11323
代理人 权鲜枝 何立春

(51)Int.Cl.
G06F 11/34(2006.01)

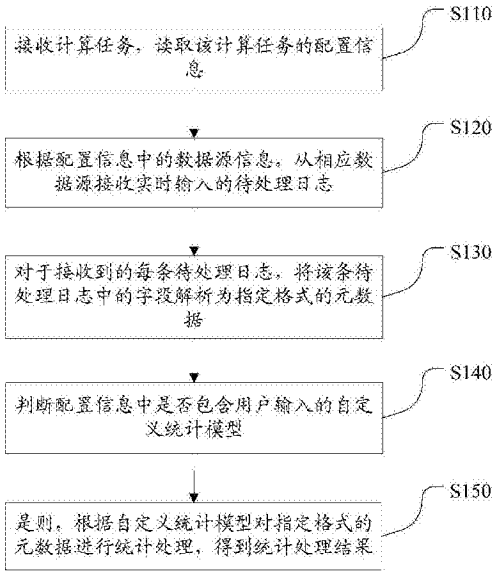
权利要求书2页 说明书15页 附图2页

(54)发明名称

一种实时计算平台的日志处理方法和装置

(57)摘要

本发明公开了一种实时计算平台的日志处理方法和装置。该方法包括:接收计算任务,读取该计算任务的配置信息;根据配置信息中的数据源信息,从相应数据源接收实时输入的待处理日志;将每条待处理日志中的字段解析为指定格式的元数据;判断配置信息中是否包含用户输入的自定义统计模型;是则,根据自定义统计模型对指定格式的元数据进行统计处理,得到统计处理结果。依据本方案,实时计算平台接收各种计算任务并执行各计算任务对应的日志处理过程,将待处理日志解析为统一的格式有利于后续统计处理的开展,且统计处理的过程支持基于用户输入的自定义统计模型而进行,实现了计算任务的可定制化,也实现了实时计算平台对可定制化的计算任务的动态适配。



CN 106201848 A

1.一种实时计算平台的日志处理方法,其中,该方法包括:

接收计算任务,读取该计算任务的配置信息;

根据所述配置信息中的数据源信息,从相应数据源接收实时输入的待处理日志;

对于接收到的每条待处理日志,将该条待处理日志中的字段解析为指定格式的元数据;

判断所述配置信息中是否包含用户输入的自定义统计模型;

是则,根据所述自定义统计模型对指定格式的元数据进行统计处理,得到统计处理结果。

2.如权利要求1所述的方法,其中,所述根据所述自定义统计模型对指定格式的元数据进行统计处理包括:

对用户输入的自定义统计模型进行解析,动态地将所述自定义统计模型解析为以实时计算平台可运行的语言表达的统计模型;

根据该解析出的统计模型,对指定格式的元数据进行统计。

3.如权利要求2所述的方法,其中,用户输入的对应于该条待处理日志的数据源的自定义统计模型是以DSL语言表达的自定义统计模型。

4.如权利要求1所述的方法,其中,该方法进一步包括:

预设多个基本统计模板;

当所述配置信息中不包含用户输入的自定义统计模型且所述配置信息中包含用户从预设的多个基本统计模板中选择的一个基本统计模板时,

根据用户选择的基本统计模板,对指定格式的元数据进行统计。

5.如权利要求4所述的方法,其中,所述基本统计模板包括如下一种或多种:

页面浏览量的统计模板,独立访客数的统计模板,访客的访问次数的统计模板,独立IP数的统计模板。

6.一种实时计算平台的日志处理装置,其中,该装置包括:

任务接收单元,适于接收计算任务,读取该计算任务的配置信息;

日志接收单元,适于根据所述配置信息中的数据源信息,从相应数据源接收实时输入的待处理日志;

解析单元,适于对于接收到的每条待处理日志,将该条待处理日志中的字段解析为指定格式的元数据;

统计单元,适于判断所述配置信息中是否包含用户输入的自定义统计模型;是则,根据所述自定义统计模型对指定格式的元数据进行统计处理,得到统计处理结果。

7.如权利要求6所述的装置,其中,

所述统计单元,适于对用户输入的自定义统计模型进行解析,动态地将所述自定义统计模型解析为以实时计算平台可运行的语言表达的统计模型;根据该解析出的统计模型,对指定格式的元数据进行统计。

8.如权利要求7所述的装置,其中,用户输入的对应于该条待处理日志的数据源的自定义统计模型是以DSL语言表达的自定义统计模型。

9.如权利要求6所述的装置,其中,

所述统计单元,进一步适于预设多个基本统计模板;适于当所述配置信息中不包含用

户输入的自定义统计模型且所述配置信息中包含用户从预设的多个基本统计模板中选择的一个基本统计模板时，根据用户选择的基本统计模板，对指定格式的元数据进行统计。

10.如权利要求9所述的装置，其中，所述基本统计模板包括如下一种或多种：

页面浏览量的统计模板，独立访客数的统计模板，访客的访问次数的统计模板，独立IP数的统计模板。

一种实时计算平台的日志处理方法和装置

技术领域

[0001] 本发明涉及互联网技术领域，具体涉及一种实时计算平台的日志处理方法和装置。

背景技术

[0002] 随着互联网技术的不断发展，互联网大数据的趋势日益显著，每一条互联网的业务线都在不断地产生新的打点日志，对产生的日志进行进一步地处理以对互联网业务的运行进行反馈是相当重要的工作之一。现有技术中，当工作人员希望对某一业务线的数据源输出的待处理日志进行处理时，需要根据相应的处理需求手动编写完整的数据处理程序，不同的日志处理需求需要重新编写不同的程序，对于不同业务线的工作人员来说，维护成本高、学习周期长，十分费时费力，使得数据处理效率低下，不符合大数据发展趋势。

发明内容

[0003] 鉴于上述问题，提出了本发明以便提供一种克服上述问题或者至少部分地解决上述问题的一种实时计算平台的日志处理方法和装置。

[0004] 依据本发明的一个方面，提供了一种实时计算平台的日志处理方法，其中，该方法包括：

[0005] 接收计算任务，读取该计算任务的配置信息；

[0006] 根据所述配置信息中的数据源信息，从相应数据源接收实时输入的待处理日志；

[0007] 对于接收到的每条待处理日志，将该条待处理日志中的字段解析为指定格式的元数据；

[0008] 判断所述配置信息中是否包含用户输入的自定义统计模型；

[0009] 是则，根据所述自定义统计模型对指定格式的元数据进行统计处理，得到统计处理结果。

[0010] 可选地，所述根据所述自定义统计模型对指定格式的元数据进行统计处理包括：

[0011] 对用户输入的自定义统计模型进行解析，动态地将所述自定义统计模型解析为以实时计算平台可运行的语言表达的统计模型；

[0012] 根据该解析出的统计模型，对指定格式的元数据进行统计。

[0013] 可选地，用户输入的对应于该条待处理日志的数据源的自定义统计模型是以DSL语言表达的自定义统计模型。

[0014] 可选地，该方法进一步包括：

[0015] 预设多个基本统计模板；

[0016] 当所述配置信息中不包含用户输入的自定义统计模型且所述配置信息中包含用户从预设的多个基本统计模板中选择的一个基本统计模板时，

[0017] 根据用户选择的基本统计模板，对指定格式的元数据进行统计。

[0018] 可选地，所述基本统计模板包括如下一种或多种：

[0019] 页面浏览量的统计模板，独立访客数的统计模板，访客的访问次数的统计模板，独立IP数的统计模板。

[0020] 可选地，所述指定格式的元数据为由字段和字段取值构成的键值对形式。

[0021] 可选地，所述将该条待处理日志中的字段解析为指定格式的元数据包括：

[0022] 根据所述配置信息中的解析条件，通过调用该条待处理日志对应的解析器将该条待处理日志中符合所述解析条件的字段解析为指定格式的元数据。

[0023] 可选地，在得到统计处理结果之后，该方法进一步包括：

[0024] 根据所述配置信息中的存储规则，将所述统计处理结果保存到相应的存储介质中。

[0025] 可选地，该方法进一步包括：预存多个基本解析器，每个基本解析器适配于一种基本数据格式；

[0026] 所述通过调用该条待处理日志对应的解析器将该条待处理日志中符合所述解析条件的字段解析为指定格式的元数据包括：

[0027] 当该条待处理日志的格式为单一基本数据格式时，从预存的多个基本解析器中查找适配于该基本数据格式的基本解析器，通过调用查找到的基本解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

[0028] 可选地，所述通过调用该条待处理日志对应的解析器将该条待处理日志中符合所述解析条件的字段解析为指定格式的元数据还包括：

[0029] 当该条待处理日志的格式为多种基本数据格式的组合时，对于每种基本数据格式，从预存的多个基本解析器中查找适配于该基本数据格式的基本解析器，通过调用查找到的多个基本解析器的组合将该条待处理日志中富恶化解析条件的字段解析为指定格式的元数据。

[0030] 可选地，所述通过调用该条待处理日志对应的解析器将该条待处理日志中符合所述解析条件的字段解析为指定格式的元数据包括：

[0031] 根据该条待处理日志的格式，确定适配于该条待处理日志的一个或多个解析函数；

[0032] 创建该条待处理日志对应的解析器，在该解析器中动态注册所述一个或多个解析函数；

[0033] 通过调用所创建的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

[0034] 可选地，在所述通过调用该条待处理日志对应的解析器将该条待处理日志中符合所述解析条件的字段解析为指定格式的元数据之后，该方法进一步包括：

[0035] 将所调用的解析器放入指定全局变量数据库中。

[0036] 可选地，所述所述通过调用该条待处理日志对应的解析器将该条待处理日志中符合所述解析条件的字段解析为指定格式的元数据包括：

[0037] 根据该条待处理日志的格式，从所述指定全局变量数据库中查找该条待处理日志对应的解析器；

[0038] 如果查找到，直接通过调用查找到的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据；

[0039] 如果未查找到，创建该条待处理日志对应的解析器，通过调用所创建的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

[0040] 依据本发明的另一个方面，提供了一种实时计算平台的日志处理装置，其中，该装置包括：

[0041] 任务接收单元，适于接收计算任务，读取该计算任务的配置信息；

[0042] 日志接收单元，适于根据所述配置信息中的数据源信息，从相应数据源接收实时输入的待处理日志；

[0043] 解析单元，适于对于接收到的每条待处理日志，将该条待处理日志中的字段解析为指定格式的元数据；

[0044] 统计单元，适于判断所述配置信息中是否包含用户输入的自定义统计模型；是则，根据所述自定义统计模型对指定格式的元数据进行统计处理，得到统计处理结果。

[0045] 可选地，所述统计单元，适于对用户输入的自定义统计模型进行解析，动态地将所述自定义统计模型解析为以实时计算平台可运行的语言表达的统计模型；根据该解析出的统计模型，对指定格式的元数据进行统计。

[0046] 可选地，用户输入的对应于该条待处理日志的数据源的自定义统计模型是以DSL语言表达的自定义统计模型。

[0047] 可选地，所述统计单元，进一步适于预设多个基本统计模板；适于当所述配置信息中不包含用户输入的自定义统计模型且所述配置信息中包含用户从预设的多个基本统计模板中选择的一个基本统计模板时，根据用户选择的基本统计模板，对指定格式的元数据进行统计。

[0048] 可选地，所述基本统计模板包括如下一种或多种：

[0049] 页面浏览量的统计模板，独立访客数的统计模板，访客的访问次数的统计模板，独立IP数的统计模板。

[0050] 可选地，所述指定格式的元数据为由字段和字段取值构成的键值对形式。

[0051] 可选地，所述解析单元，适于根据所述配置信息中的解析条件，通过调用该条待处理日志对应的解析器将该条待处理日志中符合所述解析条件的字段解析为指定格式的元数据。

[0052] 可选地，该装置进一步包括：

[0053] 存储处理单元，适于根据所述配置信息中的存储规则，将所述统计处理结果保存到相应的存储介质中。

[0054] 可选地，所述解析单元，进一步适于预存多个基本解析器，每个基本解析器适配于一种基本数据格式；以及适于当该条待处理日志的格式为单一基本数据格式时，从预存的多个基本解析器中查找适配于该基本数据格式的基本解析器，通过调用查找到的基本解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

[0055] 可选地，所述解析单元，还适于当该条待处理日志的格式为多种基本数据格式的组合时，对于每种基本数据格式，从预存的多个基本解析器中查找适配于该基本数据格式的基本解析器，通过调用查找到的多个基本解析器的组合将该条待处理日志中富恶化解析条件的字段解析为指定格式的元数据。

[0056] 可选地，所述解析单元，适于根据该条待处理日志的格式，确定适配于该条待处理

日志的一个或多个解析函数；创建该条待处理日志对应的解析器，在该解析器中动态注册所述一个或多个解析函数；通过调用所创建的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

[0057] 可选地，所述解析单元，进一步适于在所述通过调用该条待处理日志对应的解析器将该条待处理日志中符合所述解析条件的字段解析为指定格式的元数据之后，将所调用的解析器放入指定全局变量数据库中。

[0058] 可选地，所述解析单元，适于根据该条待处理日志的格式，从所述指定全局变量数据库中查找该条待处理日志对应的解析器；如果查找到，直接通过调用查找到的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据；如果未查找到，创建该条待处理日志对应的解析器，通过调用所创建的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

[0059] 由上述可知，本发明提供的技术方案执行了实时计算平台上的日志处理流程，该日志处理流程是依据实时计算平台接收到的计算任务的配置信息而展开的，根据配置信息中的数据源信息从相应的数据源获取待处理日志，由于不同数据源输出的待处理日志的格式不相同，先将接收到的待处理日志解析为统一格式的元数据，再对接收到的待处理日志进行统计处理，具体地，当配置信息中包含用户输入的自定义统计模型时，根据该自定义统计模型对待处理日志进行统计处理。依据本方案，实时计算平台为不同的日志处理需求开设了统一的接口，接收各种计算任务并执行各计算任务对应的日志处理过程，在日志处理过程中，将待处理日志解析为统一的格式有利于后续统计处理的开展，且统计处理的过程支持基于用户输入的自定义统计模型而进行，实现了计算任务的可定制化，也实现了实时计算平台对可定制化的计算任务的动态适配，能够尽可能地满足不同的日志处理需求，快速有效地得到所需要的统计处理结果。

[0060] 上述说明仅是本发明技术方案的概述，为了能够更清楚了解本发明的技术手段，而可依照说明书的内容予以实施，并且为了让本发明的上述和其它目的、特征和优点能够更明显易懂，以下特举本发明的具体实施方式。

附图说明

[0061] 通过阅读下文优选实施方式的详细描述，各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的，而并不认为是对本发明的限制。而且在整个附图中，用相同的参考符号表示相同的部件。在附图中：

[0062] 图1示出了根据本发明一个实施例的一种实时计算平台的日志处理方法的流程图；

[0063] 图2示出了根据本发明一个实施例的一种实时计算平台的日志处理装置的示意图；

[0064] 图3示出了根据本发明另一个实施例的一种实时计算平台的日志处理装置的示意图。

具体实施方式

[0065] 下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开

的示例性实施例，然而应当理解，可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反，提供这些实施例是为了能够更透彻地理解本公开，并且能够将本公开的范围完整的传达给本领域的技术人员。

[0066] 图1示出了根据本发明一个实施例的一种实时计算平台的日志处理方法的流程图。如图1所示，该方法包括：

[0067] 步骤S110，接收计算任务，读取该计算任务的配置信息。

[0068] 步骤S120，根据配置信息中的数据源信息，从相应数据源接收实时输入的待处理日志。

[0069] 步骤S130，对于接收到的每条待处理日志，将该条待处理日志中的字段解析为指定格式的元数据。

[0070] 步骤S140，判断配置信息中是否包含用户输入的自定义统计模型。

[0071] 本步骤中，用户输入的自定义统计模型是指：由用户自己编写的指示日志处理过程应遵循的统计规则的统计模型。

[0072] 步骤S150，是则，根据自定义统计模型对指定格式的元数据进行统计处理，得到统计处理结果。

[0073] 可见，图1所示的方法描述了实时计算平台上的日志处理流程，该日志处理流程是依据实时计算平台接收到的计算任务的配置信息而展开的，根据配置信息中的数据源信息从相应的数据源获取待处理日志，由于不同数据源输出的待处理日志的格式不相同，先将接收到的待处理日志解析为统一格式的元数据，再对接收到的待处理日志进行统计处理，具体地，当配置信息中包含用户输入的自定义统计模型时，根据该自定义统计模型对待处理日志进行统计处理。依据本方案，实时计算平台为不同的日志处理需求开设了统一的接口，接收各种计算任务并执行各计算任务对应的日志处理过程，在日志处理过程中，将待处理日志解析为统一的格式有利于后续统计处理的开展，且统计处理的过程支持基于用户输入的自定义统计模型而进行，实现了计算任务的可定制化，也实现了实时计算平台对可定制化的计算任务的动态适配，能够尽可能地满足不同的日志处理需求，快速有效地得到所需要的统计处理结果。

[0074] 在本发明的一个实施例中，所接收的计算任务的配置信息是由用户配置输入的，具体地，实时计算平台的前端与用户进行交互，根据用户配置输入的配置信息创建计算任务，如以网页页面的形式向用户展示多个输入框，用户通过在输入框中进行输入设置来完成配置信息的配置过程；前端将创建的计算任务提交到实时计算平台，实时计算平台接收计算任务，读取该计算任务的配置信息，根据配置信息中的数据源信息，从相应的数据源接收实时输入的待处理日志，对待处理日志进行解析，将待处理日志中的字段解析为指定格式的元数据，根据计算任务的配置信息中的用户输入的自定义统计模型，对指定格式的元数据进行统计处理，得到统计处理结果。该实时计算平台为不同的日志处理需求开设了统一的接口，用户无需为日志处理过程编写完整的程序代码，只需要将对应于日志处理需求的配置信息输入到前端创建计算任务即可，便于实施，十分省时省力，实时计算平台的整合性好、实时性高、效率高、与用户交互友好，且可以同时运行多个计算任务，符合当前的大数据发展趋势。

[0075] 在本发明的一个实施例中，步骤S150中根据自定义统计模型对指定格式的元数据

进行统计处理包括：对用户输入的自定义统计模型进行解析，动态地将自定义统计模型解析为以实时计算平台可运行的语言表达的统计模型；根据该解析出的统计模型，对指定格式的元数据进行统计。

[0076] 具体地，用户输入的对应于该条待处理日志的数据源的自定义统计模型是以DSL语言表达的自定义统计模型；DSL是具有受限表达性的一种计算机程序设计语言，DSL语言属于声明式变成，其特点是语法简单、可维护性高、学习成本低，对于绝大多数具有日志处理需求的用户来说，都可以掌握以DSL语言来表达自定义统计模型，因此，在本例中以DSL语言作为用户配置自定义统计模型的语言，当然，其他具有上述特点的语言均可以作为用户配置自定义统计模型的语言，在此不做限制，由于DSL语言表达的的自定义统计模型是没办法生成实时计算平台可识别的编程语如Java语言，因此需要相应的解析器对该自定义统计模型进行解释，动态地生成实时计算平台可运行的语言，相当于把自定义统计模型翻译成一个更复杂的统计模型，但是这个更复杂的统计模型是无法固化下来的。

[0077] 例如，接收一个计算任务，读取该计算任务的配置信息，根据该配置信息中的数据源信息，从相应的数据源接收实时输入的待处理日志，将当前接收到的待处理日志解析为指定格式的元数据，解析结果包括：(k1,v1)、(k2,v2)和(k3,v3)，且该配置信息中包含了用户输入的以DSL语言表达的自定义统计模型，自定义统计模型如下所示：

[0078] DSpark::input(…)

[0079] →filter(“a”＝“b”)

[0080] →Map(array(k1,k2))

[0081] →groupBy(k1)

[0082] →count(…)

[0083] →output(…)

[0084] 调用DSL解析器对用户输入的该自定义统计模型进行解析，动态地将自定义统计模型解析为以实时计算平台可运行的语言表达的统计模型，获知上述自定义统计模型的含义是：DSpark，有一个输入源，在这里面传一些参数(input(…)，参数具体没有写出)，然后做一个过滤，过滤条件是a＝b(filter(“a”＝“b”))，判断是否满足该过滤条件，接着从前面解析出的结果包括(k1,v1)、(k2,v2)和(k3,v3)中拿出k1和k2的字段(Map(array(k1,k2)))，进行分步统计，对k1进行汇总(groupBy(k1))，后面还可以计算总数(count(…))，后面直接输出(output(…))。则在对用户输入的自定义统计模型进行解析后，根据该解析出的统计模型，按照该统计模型指示的统计规则，对指定格式的元数据进行统计，得到统计处理结果。

[0085] 前文提到实时计算平台接收到的计算任务的配置信息是由用户输入设置的，如用户通过交互页面上的输入框进行输入设置，对于每个输入框，用户可以在该输入框中手动输入自定义字符，也可以通过在与该输入框关联的模板库中进行选择设置，则在本方案中，用户可以在配置信息中设置自定义统计模型，也可以不设置自定义统计模型而是直接从预设的模板库中进行选择。因此，在本发明的一个实施例中，图1所示的方法进一步包括：预设多个基本统计模板，所述多个基本统计模板保存在模板库中；当配置信息中不包含用户输入的自定义统计模型且该配置信息中包含用户从预设的多个基本统计模板中选择的一个基本统计模板时，根据用户选择的基本统计模板，对指定格式的元数据进行统计。

[0086] 具体地，基本统计模板包括如下一种或多种：页面浏览量的统计模板，独立访客数的统计模板，访客的访问次数的统计模板，独立IP数的统计模板。众所周知地，页面浏览量(PV)、独立访客数(UV)、访客的访问次数(VV)、独立IP数(IP)是网站分析工作中非常重要的几项分析指标；页面浏览量是衡量一个网站或网页用户访问量的数据，具体地，PV值就是所有访问者在预设统计时间内如24小时(0点到24点)内看了某个网站多少个页面或某个网页多少次，也就是页面刷新的次数，每一次页面刷新，就算做一次PV流量，其度量方法就是从浏览器发出一个对网络服务器的请求(Request)，网络服务器接到这个请求后，会将该请求对应的一个网页(Page)发送给浏览器，从而产生了一个PV，那么在这里只要是这个请求发送给了浏览器，无论这个页面是否完全打开(下载完成)，那么都是应当计为1个PV；独立访客数指访问某个站点或点击某个网页的不同ID的人数。在同一天内，UV只记录第一次进入网站的具有独立ID的访问者，在同一天内再次访问该网站则不计数，UV提供了一定时间内不同观众数量的统计指标，而没有反应出网站的全面活动；独立IP数可以理解为独立IP的访问用户，指1天内使用不同IP地址的用户访问网站的数量，同一IP无论访问了几个页面，独立IP数均为1，但是假如说两台机器访问而使用的是同一个IP，那么只能算是一个IP的访问，IP和UV之间的数据不会有太大的差异，通常UV量和比IP量高出一点，每个UV相对于每个IP更准确地对应一个实际的浏览者。

[0087] 例如，实时计算平台接收计算任务，读取计算任务的配置信息，读取该计算任务的配置信息，根据该配置信息中的数据源信息，从相应的数据源接收实时输入的待处理日志，将接收到的待处理日志解析为指定格式的元数据，且该配置信息中不包含用户输入的自定义统计模型且该配置信息中包含用户选择的独立访客数的统计模板，则根据该独立访客数的统计模板从上述指定格式的元数据中取出所有表示访问指定网页的用户ID的字段，对这些字段进行去重，将去重后剩下的字段的数量作为统计得到的独立访客数。

[0088] 此外，用户在进行配置信息的设置时，也可以选择多个基本统计模型的组合的形式，如一个计算任务的配置信息中包含用户选择的独立访客数的统计模板和页面浏览量的统计模板，则在对指定格式的元数据进行统计处理时，可以分别根据独立访客数的统计模板和页面浏览量的统计模板统计相应的独立访客数和页面浏览量。进一步地，基本统计模板还可以包括TopN统计模板，当一个计算任务的配置信息中包含用户选择的TopN统计模板时，则在对解析出的指定格式的元数据进行统计处理的过程中，对指定格式的元数据进行排序统计。

[0089] 通常情况下，上述几种基本统计模板基本覆盖了70％-80％的日志处理需求，对于不满足的日志处理需求，则可以采用前文所述的自定义统计模型进行配置，可见，本方案提供的实时计算平台基本可以满足所有的日志处理需求，并且可以同时运行多个计算任务，从各数据源不断接收待处理日志经过统计处理后不断输出统计处理结果，具有极高的日志处理效率。

[0090] 在本发明的一个实施例中，指定格式的元数据为由字段和字段取值构成的键值对形式，即key-value的形式，该形式的元数据能够反映待处理日志中的所有数据参数。

[0091] 在本发明的一个实施例中，步骤S130中将该条待处理日志中的字段解析为指定格式的元数据包括：根据配置信息中的解析条件，通过调用该条待处理日志对应的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

[0092] 其中，通过调用该条待处理日志对应的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据包括以下方式：

[0093] 方式一，根据该条待处理日志的格式，确定适配于该条待处理日志的一个或多个解析函数；创建该条待处理日志对应的解析器，在该解析器中动态注册一个或多个解析函数；通过调用所创建的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

[0094] 其中，解析函数包括如下一种或多种：Base64decode函数、base64encode函数、urldecode函数、urlencode函数、isNum函数、isVer函数、getDay函数、getHour函数、getMin函数，其中，Base64decode函数用于对Base64编码的数据进行解码，base64encode函数用于对数据进行Base64编码，urldecode函数用于还原url编码字符串，urlencode函数用于对字符串进行url编码，isNum函数用于判断是否是数字，isVer函数用于判断是否是版本，getDay函数用于获取时间的日期信息，getHour函数用于获取时间的小时信息，getMin函数用于获取时间的分钟信息。本实施例通过在创建的解析器中动态注册解析待处理日志所需的解析函数，实现了对解析器的动态定制，可以动态地适配待处理日志的形式的多样化。

[0095] 方式二，预存多个基本解析器，每个基本解析器适配于一种基本数据格式。具体地，基本解析器包括如下一种或多种：Apache日志解析器、Nginx日志解析器、数组日志解析器、Json日志解析器、分割符解析器，Apache日志解析器适配于Apache日志的数据格式，Nginx日志解析器适配于Nginx日志的数据格式，数组日志解析器适配于数组日志的数据格式，Json日志解析器适配于Json日志的数据格式，分割符解析器适配于以指定分割符进行字段分割的数据格式。

[0096] 当该条待处理日志的格式为单一基本数据格式时，从预存的多个基本解析器中查找适配于该基本数据格式的基本解析器，通过调用查找到的基本解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

[0097] 当该条待处理日志的格式为多种基本数据格式的组合时，对于每种基本数据格式，从预存的多个基本解析器中查找适配于该基本数据格式的基本解析器，通过调用查找到的多个基本解析器的组合将该条待处理日志中富恶化解析条件的字段解析为指定格式的元数据。

[0098] 例如，接收到的待处理日志是Apache日志，对应于Apache日志的数据格式，是单一基本数据格式，则对接收到的待处理日志进行解析的过程是：从预存的多个基本解析器中查找到Apache日志解析器，通过调用该Apache日志解析器将待处理日志中的字段解析为指定格式的元数据。或者，接收到的待处理日志是由分割符进行字段分割的，如“字段1&字段2”，其中“&”是分割符，字段1是数组格式，字段2是Json格式，则在对接收到的待处理日志进行解析时，需要调用分割符解析器、数组日志解析器和Json日志解析器的组合将待处理日志中的字段解析为指定格式的元数据，数组日志解析器和Json日志解析器是并列式组合，分割符解析器与该并列式组合形成层级式组合，具体地，先通过调用分隔符解析器分别解析出字段1和字段2，再通过调用数组日志解析器对字段1进行解析，以及通过调用Json日志解析器对字段2进行解析。

[0099] 进一步地，在通过调用该条待处理日志对应的解析器将该条待处理日志中符合所述解析条件的字段解析为指定格式的元数据之后，上述方法进一步包括：将所调用的解析

器放入指定全局变量数据库中。则通过调用该条待处理日志对应的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据包括：根据该条待处理日志的格式，从所述指定全局变量数据库中查找该条待处理日志对应的解析器；如果查找到，直接通过调用查找到的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据；如果未查找到，创建该条待处理日志对应的解析器，通过调用所创建的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

[0100] 例如，从同一数据源接收到待处理日志1和待处理日志2，待处理日志1和待处理日志2具有相同的数据格式，先对待处理日志1进行解析，创建待处理日志1对应的解析器1，将待处理日志1中的字段解析为指定格式的元数据，在解析之后，将解析器1放入指定全局变量数据库中，使得该解析器1作为全局变量存在可以被方便地调用，则在对待处理日志2进行解析时，先从指定全局变量数据库中查找是否有待处理日志2对应的解析器，由于待处理数据2与待处理数据1的数据格式相同，解析器1同样适配于待处理日志2，因此，直接通过调用指定全局变量数据库中的解析器1对待处理日志2进行解析，避免了适配于相同数据格式的解析器的重复创建，避免不必要的系统资源的使用，且直接全局地找解析器的过程比重新创建解析器的过程快得多，加快了解析过程，保证日志处理过程的实时性。

[0101] 在本发明的一个实施例中，在得到统计处理结果之后，图1所示的方法进一步包括：根据配置信息中的存储规则，将统计处理结果保存到相应的存储介质中。其中，所述存储介质包括如下一种或多种：Redis数据库，大存储Redis数据库，Mysql数据库，HBase数据库，HDFS数据库，GreenPlum数据库。不同的存储介质具有不同的特性，可以根据存储需求选择合适的存储介质，如Redis数据库是基于key-value形式在内存中进行存储的，但当数据量达到一定程度时，可以采用基于磁盘进行存储的大存储Redis数据库来分担存储压力，或者也可以采用分布式存储的GreenPlum数据库来分担存储压力，使得往存储介质中写数据以及从存储介质中读取数据的过程较为快捷，保证实时计算平台的实时性、有效性和稳定性。后续可以从存储介质中读取统计处理结果，将统计处理结果展示给用户，供用户查看以及在线搜索查询。

[0102] 在一个具体的例子中，在将统计处理结果保存到存储介质中之前，还可以对统计处理结果进行聚合处理，以减轻存储介质的压力，或者，在实时程度要求范围之内，设定触发存储的条件，在得到统计处理结果后，不直接进行存储，而是在满足触发存储的条件后进行存储，同样可以减轻存储压力。

[0103] 图2示出了根据本发明一个实施例的一种实时计算平台的日志处理装置的示意图。如图2所示，该实时计算平台的日志处理装置200包括：

[0104] 任务接收单元210，适于接收计算任务，读取该计算任务的配置信息。

[0105] 日志接收单元220，适于根据配置信息中的数据源信息，从相应数据源接收实时输入的待处理日志。

[0106] 解析单元230，适于对于接收到的每条待处理日志，将该条待处理日志中的字段解析为指定格式的元数据。

[0107] 统计单元240，适于判断配置信息中是否包含用户输入的自定义统计模型；是则，根据自定义统计模型对指定格式的元数据进行统计处理，得到统计处理结果。

[0108] 可见，图2所示的装置执行了实时计算平台上的日志处理流程，该日志处理流程是

依据实时计算平台接收到的计算任务的配置信息而展开的，根据配置信息中的数据源信息从相应的数据源获取待处理日志，由于不同数据源输出的待处理日志的格式不相同，先将接收到的待处理日志解析为统一格式的元数据，再对接收到的待处理日志进行统计处理，具体地，当配置信息中包含用户输入的自定义统计模型时，根据该自定义统计模型对待处理日志进行统计处理。依据本方案，实时计算平台为不同的日志处理需求开设了统一的接口，接收各种计算任务并执行各计算任务对应的日志处理过程，在日志处理过程中，将待处理日志解析为统一的格式有利于后续统计处理的开展，且统计处理的过程支持基于用户输入的自定义统计模型而进行，实现了计算任务的可定制化，也实现了实时计算平台对可定制化的计算任务的动态适配，能够尽可能地满足不同的日志处理需求，快速有效地得到所需要的统计处理结果。

[0109] 在本发明的一个实施例中，统计单元240，适于对用户输入的自定义统计模型进行解析，动态地将自定义统计模型解析为以实时计算平台可运行的语言表达的统计模型；根据该解析出的统计模型，对指定格式的元数据进行统计。其中，用户输入的对应于该条待处理日志的数据源的自定义统计模型是以DSL语言表达的自定义统计模型。

[0110] 在本发明的一个实施例中，统计单元240，进一步适于预设多个基本统计模板；适于当所述配置信息中不包含用户输入的自定义统计模型且所述配置信息中包含用户从预设的多个基本统计模板中选择的一个基本统计模板时，根据用户选择的基本统计模板，对指定格式的元数据进行统计。其中，基本统计模板包括如下一种或多种：页面浏览量的统计模板，独立访客数的统计模板，访客的访问次数的统计模板，独立IP数的统计模板。

[0111] 在本发明的一个实施例中，所述指定格式的元数据为由字段和字段取值构成的键值对形式。

[0112] 在本发明的一个实施例中，解析单元230，适于根据配置信息中的解析条件，通过调用该条待处理日志对应的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

[0113] 进一步地，解析单元230，进一步适于预存多个基本解析器，每个基本解析器适配于一种基本数据格式；以及适于当该条待处理日志的格式为单一基本数据格式时，从预存的多个基本解析器中查找适配于该基本数据格式的基本解析器，通过调用查找到的基本解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据；以及，解析单元230，还适于当该条待处理日志的格式为多种基本数据格式的组合时，对于每种基本数据格式，从预存的多个基本解析器中查找适配于该基本数据格式的基本解析器，通过调用查找到的多个基本解析器的组合将该条待处理日志中富恶化解析条件的字段解析为指定格式的元数据。

[0114] 在本发明的一个实施例中，解析单元230，适于根据该条待处理日志的格式，确定适配于该条待处理日志的一个或多个解析函数；创建该条待处理日志对应的解析器，在该解析器中动态注册所述一个或多个解析函数；通过调用所创建的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

[0115] 进一步地，解析单元230，进一步适于在通过调用该条待处理日志对应的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据之后，将所调用的解析器放入指定全局变量数据库中。则解析单元230，适于根据该条待处理日志的格式，从指定全

局变量数据库中查找该条待处理日志对应的解析器；如果查找到，直接通过调用查找到的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据；如果未查找到，创建该条待处理日志对应的解析器，通过调用所创建的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

[0116] 图3示出了根据本发明另一个实施例的一种实时计算平台的日志处理装置的示意图。如图3所示，该实时计算平台的日志处理装置300包括：任务接收单元310、日志接收单元320、解析单元330、统计单元340和存储处理单元350。

[0117] 其中，任务接收单元310、日志接收单元320、解析单元330、统计单元340和图2所示的任务接收单元210、日志接收单元220、解析单元230、统计单元240具有对应相同的功能，相同的部分在此不再赘述。

[0118] 存储处理单元350，适于根据配置信息中的存储规则，将统计处理结果保存到相应的存储介质中。

[0119] 需要说明的是，图2-图3所示装置的各实施例与图1所示方法的各实施例对应相同，上文中已有详细说明，在此不再赘述。

[0120] 综上所述，本发明提供的技术方案执行了实时计算平台上的日志处理流程，该日志处理流程是依据实时计算平台接收到的计算任务的配置信息而展开的，根据配置信息中的数据源信息从相应的数据源获取待处理日志，由于不同数据源输出的待处理日志的格式不相同，先将接收到的待处理日志解析为统一格式的元数据，再对接收到的待处理日志进行统计处理，具体地，当配置信息中包含用户输入的自定义统计模型时，根据该自定义统计模型对待处理日志进行统计处理。依据本方案，实时计算平台为不同的日志处理需求开设了统一的接口，接收各种计算任务并执行各计算任务对应的日志处理过程，在日志处理过程中，将待处理日志解析为统一的格式有利于后续统计处理的开展，且统计处理的过程支持基于用户输入的自定义统计模型而进行，实现了计算任务的可定制化，也实现了实时计算平台对可定制化的计算任务的动态适配，能够尽可能地满足不同的日志处理需求，快速有效地得到所需要的统计处理结果。

[0121] 需要说明的是：

[0122] 在此提供的算法和显示不与任何特定计算机、虚拟装置或者其它设备固有相关。各种通用装置也可以与基于在此的示教一起使用。根据上面的描述，构造这类装置所要求的结构是显而易见的。此外，本发明也不针对任何特定编程语言。应当明白，可以利用各种编程语言实现在此描述的本发明的内容，并且上面对特定语言所做的描述是为了披露本发明的最佳实施方式。

[0123] 在此处所提供的说明书中，说明了大量具体细节。然而，能够理解，本发明的实施例可以在没有这些具体细节的情况下实践。在一些实例中，并未详细示出公知的方法、结构和技术，以便不模糊对本说明书的理解。

[0124] 类似地，应当理解，为了精简本公开并帮助理解各个发明方面中的一个或多个，在上面对本发明的示例性实施例的描述中，本发明的各个特征有时被一起分组到单个实施例、图、或者对其的描述中。然而，并不应将该公开的方法解释成反映如下意图：即所要求保护的本发明要求比在每个权利要求中所明确记载的特征更多的特征。更确切地说，如下面的权利要求书所反映的那样，发明方面在于少于前面公开的单个实施例的所有特征。因此，

遵循具体实施方式的权利要求书由此明确地并入该具体实施方式，其中每个权利要求本身都作为本发明的单独实施例。

[0125] 本领域那些技术人员可以理解，可以对实施例中的设备中的模块进行自适应性地改变并且把它们设置在与该实施例不同的一个或多个设备中。可以把实施例中的模块或单元或组件组合成一个模块或单元或组件，以及此外可以把它们分成多个子模块或子单元或子组件。除了这样的特征和/或过程或者单元中的至少一些是相互排斥之外，可以采用任何组合对本说明书(包括伴随的权利要求、摘要和附图)中公开的所有特征以及如此公开的任何方法或者设备的所有过程或单元进行组合。除非另外明确陈述，本说明书(包括伴随的权利要求、摘要和附图)中公开的每个特征可以由提供相同、等同或相似目的的替代特征来代替。

[0126] 此外，本领域的技术人员能够理解，尽管在此所述的一些实施例包括其它实施例中所包括的某些特征而不是其它特征，但是不同实施例的特征的组合意味着处于本发明的范围之内并且形成不同的实施例。例如，在下面的权利要求书中，所要求保护的实施例的任意之一都可以以任意的组合方式来使用。

[0127] 本发明的各个部件实施例可以以硬件实现，或者以在一个或者多个处理器上运行的软件模块实现，或者以它们的组合实现。本领域的技术人员应当理解，可以在实践中使用微处理器或者数字信号处理器(DSP)来实现根据本发明实施例的一种实时计算平台的日志处理装置中的一些或者全部部件的一些或者全部功能。本发明还可以实现为用于执行这里所描述的方法的一部分或者全部的设备或者装置程序(例如，计算机程序和计算机程序产品)。这样的实现本发明的程序可以存储在计算机可读介质上，或者可以具有一个或者多个信号的形式。这样的信号可以从因特网网站上下载得到，或者在载体信号上提供，或者以任何其他形式提供。

[0128] 应该注意的是上述实施例对本发明进行说明而不是对本发明进行限制，并且本领域技术人员在不脱离所附权利要求的范围的情况下可设计出替换实施例。在权利要求中，不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本发明可以借助于包括有若干不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干装置的单元权利要求中，这些装置中的若干个可以是通过同一个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

[0129] 本发明公开了A1、一种实时计算平台的日志处理方法，其中，该方法包括：

[0130] 接收计算任务，读取该计算任务的配置信息；

[0131] 根据所述配置信息中的数据源信息，从相应数据源接收实时输入的待处理日志；

[0132] 对于接收到的每条待处理日志，将该条待处理日志中的字段解析为指定格式的元数据；

[0133] 判断所述配置信息中是否包含用户输入的自定义统计模型；

[0134] 是则，根据所述自定义统计模型对指定格式的元数据进行统计处理，得到统计处理结果。

[0135] A2、如A1所述的方法，其中，所述根据所述自定义统计模型对指定格式的元数据进

行统计处理包括：

[0136] 对用户输入的自定义统计模型进行解析，动态地将所述自定义统计模型解析为以实时计算平台可运行的语言表达的统计模型；

[0137] 根据该解析出的统计模型，对指定格式的元数据进行统计。

[0138] A3、如A2所述的方法，其中，用户输入的对应于该条待处理日志的数据源的自定义统计模型是以DSL语言表达的自定义统计模型。

[0139] A4、如A1所述的方法，其中，该方法进一步包括：

[0140] 预设多个基本统计模板；

[0141] 当所述配置信息中不包含用户输入的自定义统计模型且所述配置信息中包含用户从预设的多个基本统计模板中选择的一个基本统计模板时，

[0142] 根据用户选择的基本统计模板，对指定格式的元数据进行统计。

[0143] A5、如A4所述的方法，其中，所述基本统计模板包括如下一种或多种：

[0144] 页面浏览量的统计模板，独立访客数的统计模板，访客的访问次数的统计模板，独立IP数的统计模板。

[0145] A6、如A1所述的方法，其中，所述指定格式的元数据为由字段和字段取值构成的键值对形式。

[0146] A7、如A1所述的方法，其中，所述将该条待处理日志中的字段解析为指定格式的元数据包括：

[0147] 根据所述配置信息中的解析条件，通过调用该条待处理日志对应的解析器将该条待处理日志中符合所述解析条件的字段解析为指定格式的元数据。

[0148] A8、如A7所述的方法，其中，在得到统计处理结果之后，该方法进一步包括：

[0149] 根据所述配置信息中的存储规则，将所述统计处理结果保存到相应的存储介质中。

[0150] A9、如A7所述的方法，其中，该方法进一步包括：预存多个基本解析器，每个基本解析器适配于一种基本数据格式；

[0151] 所述通过调用该条待处理日志对应的解析器将该条待处理日志中符合所述解析条件的字段解析为指定格式的元数据包括：

[0152] 当该条待处理日志的格式为单一基本数据格式时，从预存的多个基本解析器中查找适配于该基本数据格式的基本解析器，通过调用查找到的基本解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

[0153] A10、如A9所述的方法，其中，所述通过调用该条待处理日志对应的解析器将该条待处理日志中符合所述解析条件的字段解析为指定格式的元数据还包括：

[0154] 当该条待处理日志的格式为多种基本数据格式的组合时，对于每种基本数据格式，从预存的多个基本解析器中查找适配于该基本数据格式的基本解析器，通过调用查找到的多个基本解析器的组合将该条待处理日志中富恶化解析条件的字段解析为指定格式的元数据。

[0155] A11、如A7所述的方法，其中，所述通过调用该条待处理日志对应的解析器将该条待处理日志中符合所述解析条件的字段解析为指定格式的元数据包括：

[0156] 根据该条待处理日志的格式，确定适配于该条待处理日志的一个或多个解析函

数；

[0157] 创建该条待处理日志对应的解析器，在该解析器中动态注册所述一个或多个解析函数；

[0158] 通过调用所创建的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

[0159] A12、如A7所述的方法，其中，在所述通过调用该条待处理日志对应的解析器将该条待处理日志中符合所述解析条件的字段解析为指定格式的元数据之后，该方法进一步包括：

[0160] 将所调用的解析器放入指定全局变量数据库中。

[0161] A13、如A12所述的方法，其中，所述所述通过调用该条待处理日志对应的解析器将该条待处理日志中符合所述解析条件的字段解析为指定格式的元数据包括：

[0162] 根据该条待处理日志的格式，从所述指定全局变量数据库中查找该条待处理日志对应的解析器；

[0163] 如果查找到，直接通过调用查找到的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据；

[0164] 如果未查找到，创建该条待处理日志对应的解析器，通过调用所创建的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

[0165] 本发明还公开了B14、一种实时计算平台的日志处理装置，其中，该装置包括：

[0166] 任务接收单元，适于接收计算任务，读取该计算任务的配置信息；

[0167] 日志接收单元，适于根据所述配置信息中的数据源信息，从相应数据源接收实时输入的待处理日志；

[0168] 解析单元，适于对于接收到的每条待处理日志，将该条待处理日志中的字段解析为指定格式的元数据；

[0169] 统计单元，适于判断所述配置信息中是否包含用户输入的自定义统计模型；是则，根据所述自定义统计模型对指定格式的元数据进行统计处理，得到统计处理结果。

[0170] B15、如B14所述的装置，其中，

[0171] 所述统计单元，适于对用户输入的自定义统计模型进行解析，动态地将所述自定义统计模型解析为以实时计算平台可运行的语言表达的统计模型；根据该解析出的统计模型，对指定格式的元数据进行统计。

[0172] B16、如B15所述的装置，其中，用户输入的对应于该条待处理日志的数据源的自定义统计模型是以DSL语言表达的自定义统计模型。

[0173] B17、如B14所述的装置，其中，

[0174] 所述统计单元，进一步适于预设多个基本统计模板；适于当所述配置信息中不包含用户输入的自定义统计模型且所述配置信息中包含用户从预设的多个基本统计模板中选择的一个基本统计模板时，根据用户选择的基本统计模板，对指定格式的元数据进行统计。

[0175] B18、如B17所述的装置，其中，所述基本统计模板包括如下一种或多种：

[0176] 页面浏览量的统计模板，独立访客数的统计模板，访客的访问次数的统计模板，独立IP数的统计模板。

[0177] B19、如B14所述的装置，其中，所述指定格式的元数据为由字段和字段取值构成的键值对形式。

[0178] B20、如B14所述的装置，其中，

[0179] 所述解析单元，适于根据所述配置信息中的解析条件，通过调用该条待处理日志对应的解析器将该条待处理日志中符合所述解析条件的字段解析为指定格式的元数据。

[0180] B21、如B20所述的装置，其中，该装置进一步包括：

[0181] 存储处理单元，适于根据所述配置信息中的存储规则，将所述统计处理结果保存到相应的存储介质中。

[0182] B22、如B20所述的装置，其中，

[0183] 所述解析单元，进一步适于预存多个基本解析器，每个基本解析器适配于一种基本数据格式；以及适于当该条待处理日志的格式为单一基本数据格式时，从预存的多个基本解析器中查找适配于该基本数据格式的基本解析器，通过调用查找到的基本解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

[0184] B23、如B22所述的装置，其中，

[0185] 所述解析单元，还适于当该条待处理日志的格式为多种基本数据格式的组合时，对于每种基本数据格式，从预存的多个基本解析器中查找适配于该基本数据格式的基本解析器，通过调用查找到的多个基本解析器的组合将该条待处理日志中富恶化解析条件的字段解析为指定格式的元数据。

[0186] B24、如B20所述的装置，其中，

[0187] 所述解析单元，适于根据该条待处理日志的格式，确定适配于该条待处理日志的一个或多个解析函数；创建该条待处理日志对应的解析器，在该解析器中动态注册所述一个或多个解析函数；通过调用所创建的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

[0188] B25、如B20所述的装置，其中，

[0189] 所述解析单元，进一步适于在所述通过调用该条待处理日志对应的解析器将该条待处理日志中符合所述解析条件的字段解析为指定格式的元数据之后，将所调用的解析器放入指定全局变量数据库中。

[0190] B26、如B25所述的装置，其中，

[0191] 所述解析单元，适于根据该条待处理日志的格式，从所述指定全局变量数据库中查找该条待处理日志对应的解析器；如果查找到，直接通过调用查找到的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据；如果未查找到，创建该条待处理日志对应的解析器，通过调用所创建的解析器将该条待处理日志中符合解析条件的字段解析为指定格式的元数据。

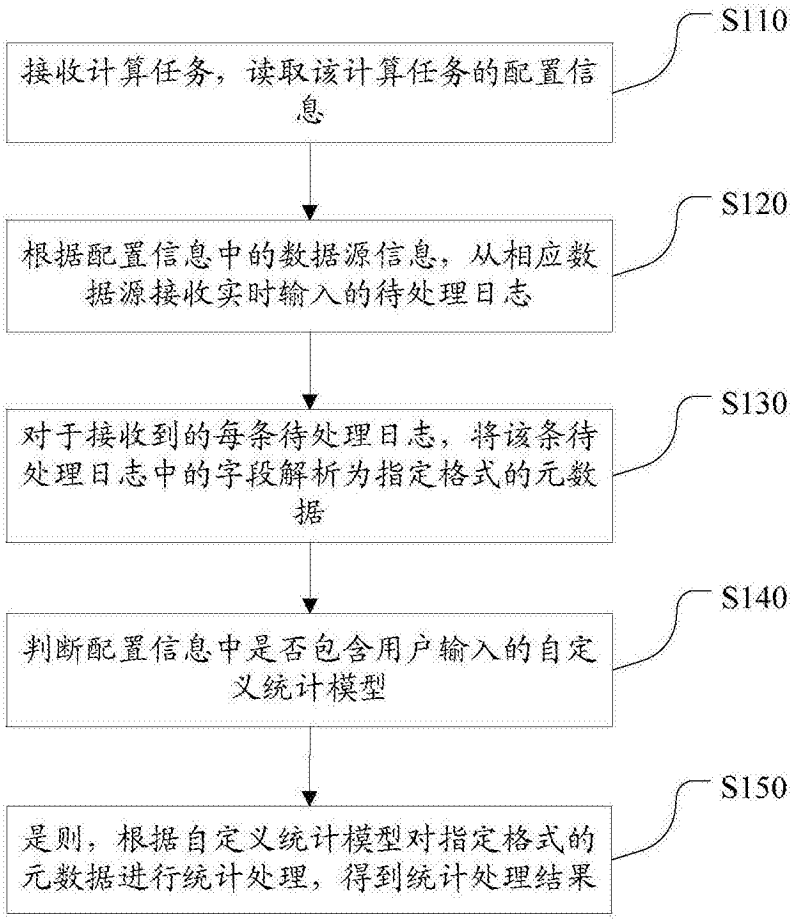

S110
接收计算任务，读取该计算任务的配置信息
S120
根据配置信息中的数据源信息，从相应数据源接收实时输入的待处理日志
S130
对于接收到的每条待处理日志，将该条待处理日志中的字段解析为指定格式的元数据
S140
判断配置信息中是否包含用户输入的自定义统计模型
S150
是则，根据自定义统计模型对指定格式的元数据进行统计处理，得到统计处理结果

图1

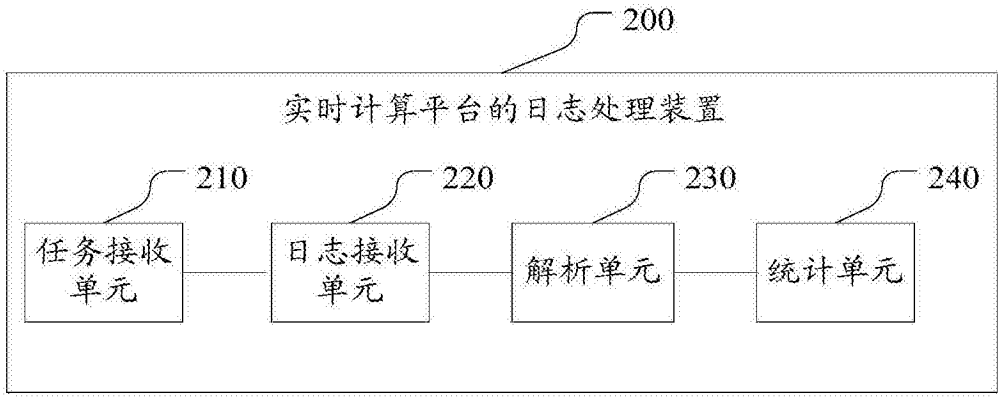

200
实时计算平台的日志处理装置
210
任务接收单元
220
日志接收单元
230
解析单元
240
统计单元

图2

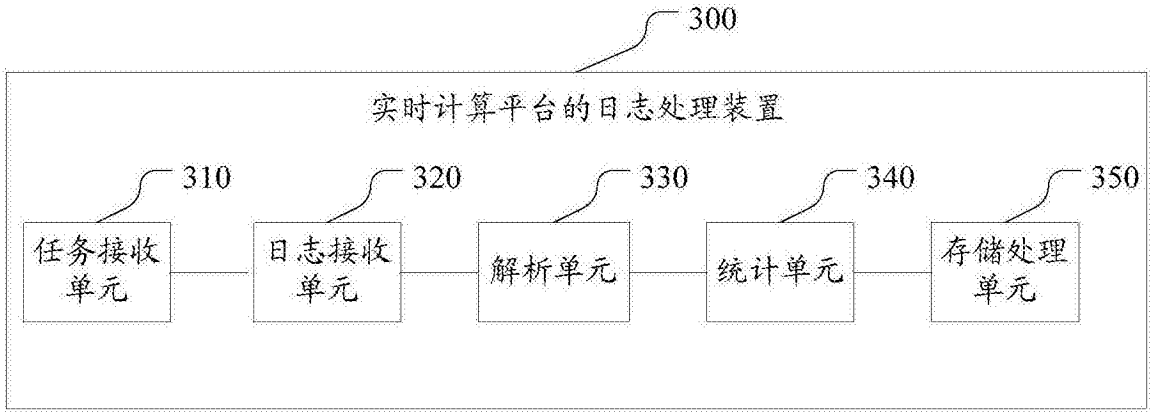

300
实时计算平台的日志处理装置
310
任务接收单元
320
日志接收单元
330
解析单元
340
统计单元
350
存储处理单元

图3