



(12) 发明专利

(10) 授权公告号 CN 109981731 B

(45) 授权公告日 2021.06.15

(21) 申请号 201910117769.4

(56) 对比文件

(22) 申请日 2019.02.15

CN 101873005 A, 2010.10.27

(65) 同一申请的已公布的文献号

US 2013031562 A1, 2013.01.31

申请公布号 CN 109981731 A

US 2011194563 A1, 2011.08.11

(43) 申请公布日 2019.07.05

杨媛琦. 基于负载均衡的Multi-UAV任务分配算法的研究.《中国优秀硕士学位论文全文数据库》.2016,

(73) 专利权人 联想(北京)有限公司

Liyun Zuo. A Multiqueue Interlacing Peak Scheduling Method Based on Tasks' Classification in Cloud Computing.《IEEE》.2016,

地址 100085 北京市海淀区上地西路6号

审查员 李旭佳

(72) 发明人 余健伟 王耀晖 杨帆 张成松

(74) 专利代理机构 北京派特恩知识产权代理有限公司 11270

代理人 张振伟 张颖玲

(51) Int.Cl.

权利要求书2页 说明书13页 附图6页

H04L 29/08 (2006.01)

H04L 12/26 (2006.01)

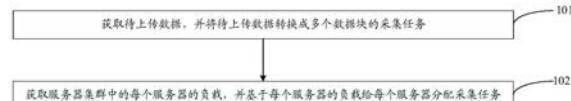
G06F 16/28 (2019.01)

(54) 发明名称

一种数据处理方法和设备

(57) 摘要

本发明实施例公开了一种数据处理方法，所述方法包括：获取待上传数据，并将所述待上传数据转换成多个数据块的采集任务；获取服务器集群中的每个服务器的负载，并基于所述每个服务器的负载给每个服务器分配所述采集任务；其中，所述服务器集群中包括至少两个用于采集所述待上传数据并上传至数据处理平台的服务器。本发明的实施例同时还公开了一种数据处理设备。



1.一种数据处理方法,所述方法包括:

获取待上传数据,并将所述待上传数据转换成多个数据块的采集任务;

设置每个所述数据块的采集任务的任务名称和每个服务器对应的服务地址;

定义每个所述数据块的采集任务的任务名称、每个所述数据块的采集任务的采集状态、每个所述数据块的采集任务对应的服务地址和指示关于采集任务的情况的含义之间的第一关系;

定义每个服务器的服务地址、服务器的额定任务数、每个服务器的服务状态和指示关于服务器运行情况的含义之间的第二关系;

获取所述服务器集群中的所述每个服务器的负载,并基于所述每个服务器的负载、所述第一关系和所述第二关系给所述每个服务器分配所述采集任务;其中,所述服务器集群中包括至少两个用于采集所述待上传数据并上传至数据处理平台的服务器。

2.根据权利要求1所述的方法,其特征在于,所述方法还包括:

通过每个服务器的心跳信息,获取所述服务器集群中的处于异常状态且分配有第一采集任务的第一服务器;

若所述第一采集任务与所述数据块的采集任务匹配,设置所述第一服务器的服务状态为第一状态,并设置所述第一采集任务的采集状态的标识为第一标识;其中,第一状态指示服务器异常;所述第一标识指示所述第一采集任务未被采集。

3.根据权利要求1或2所述的方法,其特征在于,所述方法还包括:

获取所述数据块的采集任务中需要重新分配服务器且未被采集的第二采集任务;

从所述服务器集群中获取服务器上运行的任务数小于额定任务数且服务状态正常的第二服务器;其中,所述额定任务数指的是服务器上最大可运行的任务数;

按照所述第二服务器的额定任务数,将所述第二采集任务分配给所述第二服务器;

若所述第二服务器上运行的任务数等于额定任务数,设置所述第二服务器的服务状态为第二状态;其中,所述第二状态指示服务器的服务状态正常且运行的任务数等于额定任务数。

4.根据权利要求1或2所述的方法,其特征在于,所述方法还包括:

获取所述数据块的采集任务中采集异常且未被采集的第三采集任务;

从所述服务器集群中获取服务器上运行的任务数小于额定任务数且服务状态正常的第二服务器,并计算第二服务器上剩余可运行的任务数;

基于所述第二服务器上剩余可运行的任务数,将所述第三采集任务分配至所述第二服务器。

5.根据权利要求1或2所述的方法,其特征在于,所述方法还包括:

从所述服务器集群中,获取服务器上运行的任务数大于额定任务数且服务状态正常的第四服务器,并获取服务器上运行的任务数小于额定任务数且服务状态正常的第二服务器;

基于所述第二服务器的额定任务数和所述第二服务器上运行的任务数,计算所述第二服务器上剩余可运行的任务数;

停止所述第四服务器中超出额定任务的采集任务的线程;

将所述第四服务器中的第一数量个超出额定任务的采集任务分配至所述第二服务器,

并设置所述第一数量个超出额定任务的采集任务的采集状态的标识为第二标识；其中，所述第二标识指示采集任务重新分配了服务器且当前已被采集；所述第一数量为所述第二服务器上剩余可运行的任务数量。

6. 根据权利要求1或2所述的方法，其特征在于，所述方法还包括：

获取所述数据块的采集任务中已分配服务器且未被采集的第四采集任务；

创建并执行采集任务线程，并设置所述第四采集任务的采集状态标识为第三标识；其中，所述第三标识指示采集任务当前正在被采集；

获取所述数据块的采集任务中需要重新分配服务器且已被采集的第五采集任务；

中止用于采集所述第五采集任务的线程，并设置所述第五采集任务采集状态的标识为第四标识；其中，所述第四标识指示采集任务需要重新分配服务器且当前未被采集。

7. 一种数据处理设备，其特征在于，所述设备包括：处理器、存储器和通信总线；

所述通信总线用于实现处理器和存储器之间的通信连接；

所述处理器用于执行存储器中存储的数据处理程序，以实现以下步骤：

获取待上传数据，并将所述待上传数据转换成多个数据块的采集任务；

设置每个所述数据块的采集任务的任务名称和每个服务器对应的服务地址；

定义每个所述数据块的采集任务的任务名称、每个所述数据块的采集任务的采集状态、每个所述数据块的采集任务对应的服务地址和指示关于采集任务的情况的含义之间的第一关系；

定义每个服务器的服务地址、服务器的额定任务数、每个服务器的服务状态和指示关于服务器运行情况的含义之间的第二关系；

获取所述服务器集群中的所述每个服务器的负载，并基于所述每个服务器的负载、所述第一关系和所述第二关系给所述每个服务器分配所述采集任务；其中，所述服务器集群中包括至少两个用于采集所述待上传数据并上传至数据处理平台的服务器。

一种数据处理方法和设备

技术领域

[0001] 本发明涉及通信领域中的数据处理技术,尤其涉及一种数据处理方法和设备。

背景技术

[0002] 在现阶段中不同性质数据,分散存储在各自相符合的数据仓库中,如数据可以包括设备控制系统(Device Control System,DCS)实时数据,实验室信息管理系统(Laboratory Information Management System,LIMS)数据,质量管理系统(Quality Management System,QMS)数据等。随着大数据应用的逐渐广泛,如何实时高效的将各个分散在不同仓库的海量数据拉取到大数据平台,同时对待不同权重数据采取不同的采集策略,逐渐成为大数据平台数据采集环节一大问题。

[0003] 为解决上述问题,相对技术中出现了如下技术方案:方案A、针对不同数据开发不同采集服务,对海量数据的一个仓库,拆分数据块部署多个采集服务进行采集。各采集服务独立,针对同一个数据块采集服务部署主备,实现高可用。方案B、分布式部署多个服务,各服务按设置采集任务进行采集。方案C、由数据主动拉取改为不同数据源主动推送数据到大数据平台。但是,上述方案存在各服务之间没有关联,所需资源过多且无法均衡采集,以及采集不能做到高可用的问题。

发明内容

[0004] 为解决上述技术问题,本发明实施例期望提供一种数据处理方法和设备,解决了相对技术中存在的服务之间没有关联的问题,降低了所需资源,且可以实现均衡采集;同时,采集过程中实现了服务的高可用性。

[0005] 本发明的技术方案是这样实现的:

[0006] 一种数据处理方法,所述方法包括:

[0007] 获取待上传数据,并将所述待上传数据转换成多个数据块的采集任务;

[0008] 获取服务器集群中的每个服务器的负载,并基于所述每个服务器的负载给每个服务器分配所述采集任务;其中,所述服务器集群中包括至少两个用于采集所述待上传数据并上传至数据处理平台的服务器。

[0009] 可选的,所述获取服务器集群中的每个服务器的负载,并基于所述每个服务器的负载给每个服务器分配所述采集任务,包括:

[0010] 通过每个服务器的心跳信息,获取所述服务器集群中的处于异常状态且分配有第一采集任务的第一服务器;

[0011] 若所述第一采集任务与所述数据块的采集任务匹配,设置所述第一服务器的服务状态为第一状态,并设置所述第一采集任务的采集状态的标识为第一标识;其中,第一状态指示服务器异常;所述第一标识指示所述第一采集任务未被采集。

[0012] 可选的,所述获取服务器集群中的每个服务器的负载,并基于所述每个服务器的负载给每个服务器分配所述采集任务,包括:

- [0013] 获取所述数据块的采集任务中需要重新分配服务器且未被采集的第二采集任务；
[0014] 从所述服务器集群中获取服务器上运行的任务数小于额定任务数且服务状态正常的第二服务器；其中，所述额定任务数指的是服务器上最大可运行的任务数；
[0015] 按照所述第二服务器的额定任务数，将所述第二采集任务分配给所述第二服务器；
[0016] 若所述第二服务器上运行的任务数等于额定任务数，设置所述第二服务器的服务状态为第二状态；其中，所述第二状态指示服务器的服务状态正常且运行的任务数等于额定任务数。
[0017] 可选的，所述获取服务器集群中的每个服务器的负载，并基于所述每个服务器的负载给每个服务器分配所述采集任务，包括：
[0018] 获取所述数据块的采集任务中采集异常且未被采集的第三采集任务；
[0019] 从所述服务器集群中获取服务器上运行的任务数小于额定任务数且服务状态正常的第二服务器，并计算第二服务器上剩余可运行的任务数；
[0020] 基于所述第二服务器上剩余可运行的任务数，将所述第三采集任务分配至所述第二服务器。
[0021] 可选的，所述方法还包括：
[0022] 若所述第三采集任务中存在未分配至服务器的第一子采集任务，且所述第二服务器上运行的任务数等于额定任务数，设置所述第二服务器的服务状态为第二状态，并从所述服务器集群中获取服务器上运行的任务数等于额定任务数且服务状态正常的第三服务器；其中，所述第三服务器中包括第二服务器；
[0023] 计算所述第三服务器上剩余可运行的任务数，并基于所述第三服务器上剩余可运行的任务数，将所述第一子采集任务依次分配至所述第三服务器。
[0024] 可选的，所述方法还包括：
[0025] 若所述第一子采集任务中存在未分配至服务器的第二子采集任务，且所述第三服务器上运行的任务数大于额定任务数，设置所述第三服务器的服务状态为第三状态，并从所述服务器集群中获取服务器上运行的任务数大于额定任务数且服务状态正常的第四服务器；其中，所述第四服务器中包括第三服务器，所述第三状态指示服务器的服务状态正常且运行的任务数大于额定任务数；
[0026] 将所述第二子采集任务依次分配至所述第四服务器。
[0027] 可选的，所述获取服务器集群中的每个服务器的负载，并基于所述每个服务器的负载给每个服务器分配所述采集任务，包括：
[0028] 从所述服务器集群中，获取服务器上运行的任务数大于额定任务数且服务状态正常的第四服务器，并获取服务器上运行的任务数小于额定任务数且服务状态正常的第二服务器；
[0029] 基于所述第二服务器的额定任务数和所述第二服务器上运行的任务数，计算所述第二服务器上剩余可运行的任务数；
[0030] 停止所述第四服务器中超出额定任务的采集任务的线程；
[0031] 将所述第四服务器中的第一数量个超出额定任务的采集任务分配至所述第二服务器，并设置所述第一数量个超出额定任务的采集任务的采集状态的标识为第二标识；其

中,所述第二标识指示采集任务重新分配了服务器且当前已被采集;所述第一数量为所述第二服务器上剩余可运行的任务数量。

[0032] 可选的,所述获取服务器集群中的每个服务器的负载,并基于所述每个服务器的负载给每个服务器分配所述采集任务,包括:

[0033] 获取所述数据块的采集任务中已分配服务器且未被采集的第四采集任务;

[0034] 创建并执行采集任务线程,并设置所述第四采集任务的采集状态标识为第三标识;其中,所述第三标识指示采集任务当前正在被采集;

[0035] 获取所述数据块的采集任务中需要重新分配服务器且已被采集的第五采集任务;

[0036] 中止用于采集所述第五采集任务的线程,并设置所述第五采集任务采集状态的标识为第四标识;其中,所述第四标识指示采集任务需要重新分配服务器且当前未被采集。

[0037] 可选的,所述方法还包括:

[0038] 设置每个所述数据块采集任务的任务名称和每个服务器对应的服务地址;

[0039] 定义每个所述数据块采集任务的任务名称、每个所述数据块采集任务的采集状态、每个所述数据块采集任务对应的服务地址和指示关于采集任务的情况的含义之间的第一关系;

[0040] 定义每个服务器的服务地址、服务器的额定任务数、每个服务器的服务状态和指示关于服务器运行情况的含义之间的第二关系;其中,所述第一关系和所述第二关系用于给所述服务器分配所述采集任务时使用。

[0041] 一种数据处理设备,所述设备包括:处理器、存储器和通信总线;

[0042] 所述通信总线用于实现处理器和存储器之间的通信连接;

[0043] 所述处理器用于执行存储器中存储的数据处理程序,以实现以下步骤:

[0044] 获取待上传数据,并将所述待上传数据转换成多个数据块的采集任务;

[0045] 获取服务器集群中的每个服务器的负载,并基于所述每个服务器的负载给每个服务器分配所述采集任务;其中,所述服务器集群中包括至少两个用于采集所述待上传数据并上传至数据处理平台的服务器。

[0046] 本发明的实施例所提供的数据处理方法和设备,获取待上传数据并将待上传数据转换成多个数据块的采集任务,获取服务器集群中的每个服务器的负载,并基于每个服务器的负载给每个服务器分配采集任务;其中,服务器集群中包括至少两个用于采集待上传数据并上传至数据处理平台的服务器,如此,在给服务器分配采集任务时是基于各个服务器的负载来进行的,且无需针对每个数据块部署一个服务器,解决了相对技术中存在的服务之间没有关联的问题,降低了所需资源,且可以实现均衡采集;同时,采集过程中实现了服务的高可用性。

附图说明

[0047] 图1为本发明的实施例提供的一种数据处理方法的流程示意图;

[0048] 图2为本发明的实施例提供的另一种数据处理方法的流程示意图;

[0049] 图3为本发明的实施例提供的又一种数据处理方法的流程示意图;

[0050] 图4为本发明的另一实施例提供的一种数据处理方法的流程示意图;

[0051] 图5为本发明的另一实施例提供的另一种数据处理方法的流程示意图;

- [0052] 图6为本发明的另一实施例提供的又一种数据处理方法的流程示意图；
- [0053] 图7为本发明的又一实施例提供的一种数据处理方法的流程示意图；
- [0054] 图8为本发明的实施例提供的一种数据处理设备的结构示意图。

具体实施方式

[0055] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述。

- [0056] 本发明的实施例提供一种数据处理方法，参照图1所示，该方法包括以下步骤101、获取待上传数据，并将待上传数据转换成多个数据块的采集任务。其中，步骤101获取待上传数据并将待上传数据转换成多个数据块的采集任务可以由数据处理设备来实现；待上传数据可以包括设备控制系统(Device Control System, DCS)实时数据，实验室信息管理系统(Laboratory Information Management System, LIMS)数据，质量管理系统(Quality Management System, QMS)数据等数据。本申请实施例中是将待上传数据抽象成多个数据块的采集任务。
- [0057] 步骤102、获取服务器集群中的每个服务器的负载，并基于每个服务器的负载给每个服务器分配采集任务。其中，服务器集群中包括至少两个用于采集待上传数据并上传至数据处理平台的服务器。
- [0058] 步骤103、获取服务器集群中的每个服务器的负载，并基于每个服务器的负载给每个服务器分配采集任务可以由数据处理设备来实现；数据处理设备可以根据服务器的负载确定服务器的服务状态，并根据服务器的服务状态给该服务器分配采集任务。
- [0059] 本发明的实施例所提供的数据处理方法，获取待上传数据并将待上传数据转换成多个数据块的采集任务，获取服务器集群中的每个服务器的负载，并基于每个服务器的负载给每个服务器分配采集任务；其中，服务器集群中包括至少两个用于采集待上传数据并上传至数据处理平台的服务器，如此，在给服务器分配采集任务时是基于各个服务器的负载来进行的，且无需针对每个数据块部署一个服务器，解决了相对技术中存在的服务之间没有关联的问题，降低了所需资源，且可以实现均衡采集；同时，采集过程中实现了服务的高可用性。
- [0060] 基于前述实施例，本发明的实施例提供一种数据处理方法，参照图2所示，该方法包括以下步骤：
- [0061] 步骤201、数据处理设备获取待上传数据，并将待上传数据转换成多个数据块的采集任务。
- [0062] 步骤202、数据处理设备通过每个服务器的心跳信息，获取服务器集群中的处于异常状态且分配有第一采集任务的第一服务器。其中，数据处理设备可以通过实时监测每个服务器的心跳信息，进而根据服务器的心跳信息来确定服务器的服务状态；在确定每个服务器的服务状态后从服务器集群中获取处于异常状态且分配有采集任务的服务器；且，第一采集任务可以是任一采集待上传数据的任务。
- [0063] 步骤203、若第一采集任务与数据块的采集任务匹配，数据处理设备设置第一服务

器的服务状态为第一状态，并设置第一采集任务的采集状态的标识为第一标识。

[0068] 其中，第一状态指示服务器异常；第一标识指示第一采集任务未被采集。

[0069] 在本发明的其他实施例中，如果第一采集任务属于数据块的采集任务，那么说明第一采集任务与数据块的采集任务匹配；此时，数据处理设备可以设置该第一服务器的服务状态为异常状态，且设置第一采集任务的采集状态为未被采集。

[0070] 在本发明的其它实施例中，参照图3所示，该方法还可以包括以下步骤：

[0071] 步骤204、数据处理设备获取数据块的采集任务中需要重新分配服务器且未被采集的第二采集任务。

[0072] 步骤205、数据处理设备从服务器集群中获取服务器上运行的任务数小于额定任务数且服务状态正常的第二服务器。

[0073] 其中，额定任务数指的是服务器上最大可运行的任务数。

[0074] 步骤206、数据处理设备按照第二服务器的额定任务数，将第二采集任务分配给第二服务器。

[0075] 其中，再给第二服务器分配第二采集任务时可以是按照第二服务器上最大可运行的任务数来分配的。

[0076] 步骤207、若第二服务器上运行的任务数等于额定任务数，数据处理设备设置第二服务器的服务状态为第二状态。

[0077] 其中，第二状态指示服务器的服务状态正常且运行的任务数等于额定任务数。

[0078] 在本发明的其他实施例中，如果第二服务器上运行的任务数达到第二服务器上可运行的最大任务数，此时数据处理设备可以设置该第二服务器的服务状态为正常状态，且第二服务器的上运行的任务数为额定任务数。

[0079] 其中，步骤204-207也可以是在步骤201之后执行的。

[0080] 需要说明的是，本实施例中与其它实施例中相同步骤和相同内容的说明，可以参照其它实施例中的描述，此处不再赘述。

[0081] 本发明的实施例所提供的数据处理方法，在给服务器分配采集任务时是基于各个服务器的负载来进行的，且无需针对每个数据块部署一个服务器，解决了相对技术中存在的服务之间没有关联的问题，降低了所需资源，且可以实现均衡采集；同时，采集过程中实现了服务的高可用性。

[0082] 基于前述实施例，本发明的实施例提供一种数据处理方法，参照图4所示，该方法包括以下步骤：

[0083] 步骤301、数据处理设备获取待上传数据，并将待上传数据转换成多个数据块的采集任务。

[0084] 步骤302、数据处理设备通过每个服务器的心跳信息，获取服务器集群中的处于异常状态且分配有第一采集任务的第一服务器。

[0085] 步骤303、若第一采集任务与数据块的采集任务匹配，数据处理设备设置第一服务器的服务状态为第一状态，并设置第一采集任务的采集状态的标识为第一标识。

[0086] 其中，第一状态指示服务器异常；第一标识指示第一采集任务未被采集。

[0087] 步骤304、数据处理设备获取数据块的采集任务中采集异常且未被采集的第三采集任务。

[0088] 其中,第三采集任务可以指的是还没有被任何服务器采集,且采集异常的采集任务。

[0089] 步骤305、数据处理设备从服务器集群中获取服务器上运行的任务数小于额定任务数且服务状态正常的第二服务器,并计算第二服务器上剩余可运行的任务数。

[0090] 其中,第二服务器是服务器集群中服务器上还能够运行其他采集任务的服务器;第二服务器上剩余可运行的任务数是用第二服务器的额定任务数减去第二服务器上当前运行的任务数后得到的。

[0091] 步骤306、数据处理设备基于第二服务器上剩余可运行的任务数,将第三采集任务分配至第二服务器。

[0092] 其中,第二服务器上可分配的采集任务的数量小于或等于第二服务器上剩余可运行的任务数。

[0093] 步骤307、若第三采集任务中存在未分配至服务器的第一子采集任务,且第二服务器上运行的任务数等于额定任务数,数据处理设备设置第二服务器的服务状态为第二状态,并从服务器集群中获取服务器上运行的任务数等于额定任务数且服务状态正常的第三服务器。

[0094] 其中,第三服务器中包括第二服务器。

[0095] 在本发明的其他实施例中,如果第三采集任务中存在还没分配服务器的采集任务,且第二服务器上运行的任务数达到第二服务器上可运行的最大任务数,此时数据处理设备可以设置该第二服务器的服务状态为正常状态,且第二服务器的上运行的任务数为额定任务数。

[0096] 步骤308、数据处理设备计算第三服务器上剩余可运行的任务数,并基于第三服务器上剩余可运行的任务数,将第一子采集任务依次分配至第三服务器。

[0097] 其中,第三服务器上剩余可运行的任务数是用第三服务器的额定任务数减去第二服务器上当前运行的任务数后得到的。在给第三服务器分配采集任务时可以依次将第一子采集任务分配给第三服务器直至第一子采集任务分配完成且第三服务器上运行的任务数小于或等于额定任务数;或,依次将第一子采集任务分配给第三服务器直至第三服务器上运行的任务数等于额定任务数。

[0098] 在本发明的其它实施例中,参照图5所示,该方法还包括以下步骤:

[0099] 步骤309、若第一子采集任务中存在未分配至服务器的第二子采集任务,且第三服务器上运行的任务数大于额定任务数,数据处理设备设置第三服务器的服务状态为第三状态,并从服务器集群中获取服务器上运行的任务数大于额定任务数且服务状态正常的第四服务器。

[0100] 其中,第四服务器中包括第三服务器,第三状态指示服务器的服务状态正常且运行的任务数大于额定任务数。

[0101] 在本发明的其他实施例中,如果第一子采集任务中存在还没分配服务器的采集任务,且第三服务器上运行的任务数超过第三服务器上可运行的最大任务数,此时数据处理设备可以设置该第三服务器的服务状态为正常状态,且第三服务器的上运行的任务数大于额定任务数。

[0102] 步骤310、数据处理设备将第二子采集任务依次分配至第四服务器。

[0103] 在给第四服务器分配采集任务时可以依次将第二子采集任务分配给第四服务器直至第二子采集任务分配完成且第四服务器上运行的任务数小于或等于额定任务数；或，依次将第二子采集任务分配给第四服务器直至第四服务器上运行的任务数等于额定任务数。

[0104] 需要说明的是，本实施例中与其它实施例中相同步骤和相同内容的说明，可以参照其它实施例中的描述，此处不再赘述。

[0105] 其中，步骤304-310也可以是在步骤301之后执行的。

[0106] 本发明的实施例所提供的数据处理方法，在给服务器分配采集任务时是基于各个服务器的负载来进行的，且无需针对每个数据块部署一个服务器，解决了相对技术中存在的服务之间没有关联的问题，降低了所需资源，且可以实现均衡采集；同时，采集过程中实现了服务的高可用性。

[0107] 基于前述实施例，本发明的实施例提供一种数据处理方法，参照图6所示，该方法包括以下步骤：

[0108] 步骤401、数据处理设备获取待上传数据，并将待上传数据转换成多个数据块的采集任务。

[0109] 步骤402、数据处理设备通过每个服务器的心跳信息，获取服务器集群中的处于异常状态且分配有第一采集任务的第一服务器。

[0110] 步骤403、若第一采集任务与数据块的采集任务匹配，数据处理设备设置第一服务器的服务状态为第一状态，并设置第一采集任务的采集状态的标识为第一标识。

[0111] 其中，第一状态指示服务器异常；第一标识指示第一采集任务未被采集。

[0112] 步骤404、数据处理设备从服务器集群中，获取服务器上运行的任务数大于额定任务数且服务状态正常的第四服务器，并获取服务器上运行的任务数小于额定任务数且服务状态正常的第二服务器。

[0113] 步骤405、数据处理设备基于第二服务器的额定任务数和第二服务器上运行的任务数，计算第二服务器上剩余可运行的任务数。

[0114] 其中，第二服务器上剩余可运行的任务数是用第二服务器的额定任务数减去第二服务器上当前运行的任务数得到的。

[0115] 步骤406、数据处理设备停止第四服务器中超出额定任务的采集任务的线程。

[0116] 其中，停止第四服务器中超出额定任务的采集任务的线程后，第四服务器不会再给第四服务器中超出额定任务的采集任务提供任何服务。

[0117] 步骤407、数据处理设备将第四服务器中的第一数量个超出额定任务的采集任务分配至第二服务器，并设置第一数量个超出额定任务的采集任务的采集状态的标识为第二标识。

[0118] 其中，第二标识指示采集任务重新分配了服务器且当前已被采集；第一数量为第二服务器上剩余可运行的任务数量。

[0119] 在本发明的其他实施中，可以将第四服务器中的第一数量个超出额定任务的采集任务分配至可为其提供服务的其他服务器；第一数量是其他服务器上剩余可运行的任务数；或者，第一数量是第四服务器中超出额定任务的采集任务的数量。

[0120] 在本发明的其他实施例中，如图7所示，步骤403之后也可以执行以下步骤：

[0121] 步骤408、数据处理设备获取数据块的采集任务中已分配服务器且未被采集的第四采集任务。

[0122] 步骤409、数据处理设备创建并执行采集任务线程，并设置第四采集任务的采集状态标识为第三标识。

[0123] 其中，第三标识指示采集任务当前正在被采集。

[0124] 步骤410、数据处理设备获取数据块的采集任务中需要重新分配服务器且已被采集的第五采集任务。

[0125] 其中，第五采集任务是因为某些原因需要重新为其分配服务器且已被分配了服务器的采集任务。

[0126] 步骤411、数据处理设备中止用于采集第五采集任务的线程，并设置第五采集任务采集状态的标识为第四标识。

[0127] 其中，第四标识指示采集任务需要重新分配服务器且当前未被采集。

[0128] 在本发明的其他实施例中，数据处理设备可以终止第五采集任务的线程，并设置第五采集任务的采集状态为需要重新分配服务器且当前未被采集的状态。

[0129] 需要说明的是，本实施例中与其它实施例中相同步骤和相同内容的说明，可以参照其它实施例中的描述，此处不再赘述。

[0130] 其中，步骤404-411也可以是在步骤401之后执行的。

[0131] 本发明的实施例所提供的数据处理方法，在给服务器分配采集任务时是基于各个服务器的负载来进行的，且无需针对每个数据块部署一个服务器，解决了相对技术中存在的服务之间没有关联的问题，降低了所需资源，且可以实现均衡采集；同时，采集过程中实现了服务的高可用性。

[0132] 基于前述实施例，在本发明的其它实施例中，该方法还包括以下步骤：

[0133] A、设置每个数据块采集任务的任务名称和每个服务器对应的服务地址。

[0134] B、定义每个数据块采集任务的任务名称、每个数据块采集任务的采集状态、每个数据块采集任务对应的服务地址和指示关于采集任务的情况的含义之间的第一关系。

[0135] 其中，每个数据块采集任务的任务名称和每个服务器对应的服务地址可以按照如下所示的表1中所示的示例来设置，且每个数据块采集任务的任务名称、每个数据块采集任务的采集状态、每个数据块采集任务对应的服务地址和指示关于采集任务的情况的含义之间的第一关系也可以按照表1中所示的示例来设置，其指示关于采集任务的情况的含义可以如表1中所示的表示含义所示。当然，表1中只是进行示例说明，并未限定只能是如表1中所示。

[0136]

任务	所属采集 ip	采集状态	表示含义
Task1	0.0.0.0	0	表示任务 Task1 尚未被任何采集服务确认加载采集，且当前未被采集
Task1	0.0.0.1	0	表示任务 Task1 加载采集服务异常退出，且当前未被采集
Task1	10.100.124.209	0	表示任务 Task1 被采集服务 10.100.124.209 确认加载采集 但当前未被采集
Task1	10.100.124.209	1	表示任务 Task1 被采集服务 10.100.124.209 确认加载采集，且当前已正在采集
Task1	10.100.124.209	2	表示任务 Task1 需要重新分配采集服务，但当前已被 10.100.124.209 服务采集
Task1	10.100.124.209	3	表示任务 Task1 需要重新分配采集采集服务，且当前未被 10.100.124.209 服务采集

[0137] 表1

[0138] C、定义每个服务器的服务地址、服务器的额定任务数、每个服务器的服务状态和指示关于服务器运行情况的含义之间的第二关系。

[0139] 其中，第一关系和第二关系用于给服务器分配采集任务时使用。

[0140] 在本发明的其他实施例中，每个服务器的服务地址、服务器的额定任务数、每个服务器的服务状态和指示关于服务器运行情况的含义之间的第二关系可以按照如下表2中所示的示例来设置，其指示关于服务器运行情况的含义可以如表2中所示的表示含义所示。当然，表2中只是进行示例说明，并未限定只能是如表2中所示。

[0141]

采集服务 ip	最大并发任务数	服务状态	表示含义
10.100.124.209	5	0	表示采集服务 10.100.124.209 最大并发采集任务数为 5，服务状态未启用或异常中止

[0142]	10.100.124.209	5	1	表示采集服务 10.100. 124.209 最大并发采集任务数为 5, 服务状态正常执行但实际并发执行任务数小于设定最大值
	10.100.124.209	5	2	表示采集服务 10.100. 124.209 最大并发采集任务数为 5, 服务状态正常执行且实际并发执行任务数等于设定最大值
	10.100.124.209	5	3	表示采集服务 10.100. 124.209 最大并发采集任务数为 5, 服务状态正常执行且实际并发执行任务数大于设定最大值

[0143] 表2

[0144] 基于前述实施例,本发明的实施例提供一种数据处理设备,该设备可以应用于图1~7对应的实施例提供的数据处理方法中,参照图8所示,该设备可以包括:处理器51、存储器52和通信总线53,其中:

- [0145] 通信总线53用于实现处理器51和存储器52之间的通信连接;
- [0146] 处理器51用于执行存储器52中存储的数据处理程序,以实现以下步骤:
 - [0147] 获取待上传数据,并将待上传数据转换成多个数据块的采集任务;
 - [0148] 获取服务器集群中的每个服务器的负载,并基于每个服务器的负载给每个服务器分配采集任务;
 - [0149] 其中,服务器集群中包括至少两个用于采集待上传数据并上传至数据处理平台的服务器。
 - [0150] 在本发明的其他实施例中,处理器51用于执行存储器52中存储的数据处理程序中的获取服务器集群中的每个服务器的负载,并基于每个服务器的负载给每个服务器分配采集任务,以实现以下步骤:
 - [0151] 通过每个服务器的心跳信息,获取服务器集群中的处于异常状态且分配有第一采集任务的第一服务器;
 - [0152] 若第一采集任务与数据块的采集任务匹配,设置第一服务器的服务状态为第一状态,并设置第一采集任务的采集状态的标识为第一标识;
 - [0153] 其中,第一状态指示服务器异常;第一标识指示第一采集任务未被采集。
 - [0154] 在本发明的其他实施例中,处理器51用于执行存储器52中存储的数据处理程序中的获取服务器集群中的每个服务器的负载,并基于每个服务器的负载给每个服务器分配采集任务,以实现以下步骤:
 - [0155] 获取数据块的采集任务中需要重新分配服务器且未被采集的第二采集任务;
 - [0156] 从服务器集群中获取服务器上运行的任务数小于额定任务数且服务状态正常的第二服务器;

- [0157] 其中，额定任务数指的是服务器上最大可运行的任务数；
[0158] 按照第二服务器的额定任务数，将第二采集任务分配给第二服务器；
[0159] 若第二服务器上运行的任务数等于额定任务数，设置第二服务器的服务状态为第二状态；
[0160] 其中，第二状态指示服务器的服务状态正常且运行的任务数等于额定任务数。
[0161] 在本发明的其他实施例中，处理器51用于执行存储器52中存储的数据处理程序中的获取服务器集群中的每个服务器的负载，并基于每个服务器的负载给每个服务器分配采集任务，以实现以下步骤：
[0162] 获取数据块的采集任务中采集异常且未被采集的第三采集任务；
[0163] 从服务器集群中获取服务器上运行的任务数小于额定任务数且服务状态正常的第二服务器，并计算第二服务器上剩余可运行的任务数；
[0164] 基于第二服务器上剩余可运行的任务数，将第三采集任务分配至第二服务器。
[0165] 在本发明的其他实施例中，处理器51用于执行存储器52中存储的数据处理程序，还可以实现以下步骤：
[0166] 若第三采集任务中存在未分配至服务器的第一子采集任务，且第二服务器上运行的任务数等于额定任务数，设置第二服务器的服务状态为第二状态，并从服务器集群中获取服务器上运行的任务数等于额定任务数且服务状态正常的第三服务器；
[0167] 其中，第三服务器中包括第二服务器；
[0168] 计算第三服务器上剩余可运行的任务数，并基于第三服务器上剩余可运行的任务数，将第一子采集任务依次分配至第三服务器。
[0169] 在本发明的其他实施例中，处理器51用于执行存储器52中存储的数据处理程序，还可以实现以下步骤：
[0170] 若第一子采集任务中存在未分配至服务器的第二子采集任务，且第三服务器上运行的任务数大于额定任务数，设置第三服务器的服务状态为第三状态，并从服务器集群中获取服务器上运行的任务数大于额定任务数且服务状态正常的第四服务器；
[0171] 其中，第四服务器中包括第三服务器，第三状态指示服务器的服务状态正常且运行的任务数大于额定任务数；
[0172] 将第二子采集任务依次分配至第四服务器。
[0173] 在本发明的其他实施例中，处理器51用于执行存储器52中存储的数据处理程序中的获取服务器集群中的每个服务器的负载，并基于每个服务器的负载给每个服务器分配采集任务，以实现以下步骤：
[0174] 从服务器集群中，获取服务器上运行的任务数大于额定任务数且服务状态正常的第四服务器，并获取服务器上运行的任务数小于额定任务数且服务状态正常的第二服务器；
[0175] 基于第二服务器的额定任务数和第二服务器上运行的任务数，计算第二服务器上剩余可运行的任务数；
[0176] 停止第四服务器中超出额定任务的采集任务的线程；
[0177] 将第四服务器中的第一数量个超出额定任务的采集任务分配至第二服务器，并设置第一数量个超出额定任务的采集任务的采集状态的标识为第二标识；

[0178] 其中,第二标识指示采集任务重新分配了服务器且当前已被采集;第一数量为第二服务器上剩余可运行的任务数量。

[0179] 在本发明的其他实施例中,处理器51用于执行存储器52中存储的数据处理程序中的获取服务器集群中的每个服务器的负载,并基于每个服务器的负载给每个服务器分配采集任务,以实现以下步骤:

[0180] 获取数据块的采集任务中已分配服务器且未被采集的第四采集任务;

[0181] 创建并执行采集任务线程,并设置第四采集任务的采集状态标识为第三标识;

[0182] 其中,第三标识指示采集任务当前正在被采集;

[0183] 获取数据块的采集任务中需要重新分配服务器且已被采集的第五采集任务;

[0184] 中止用于采集第五采集任务的线程,并设置第五采集任务采集状态的标识为第四标识;

[0185] 其中,第四标识指示采集任务需要重新分配服务器且当前未被采集。

[0186] 在本发明的其他实施例中,处理器51用于执行存储器52中存储的数据处理程序,还可以实现以下步骤:

[0187] 设置每个数据块采集任务的任务名称和每个服务器对应的服务地址;

[0188] 定义每个数据块采集任务的任务名称、每个数据块采集任务的采集状态、每个数据块采集任务对应的服务地址和指示关于采集任务的情况的含义之间的第一关系;

[0189] 定义每个服务器的服务地址、服务器的额定任务数、每个服务器的服务状态和指示关于服务器运行情况的含义之间的第二关系;

[0190] 其中,第一关系和第二关系用于给服务器分配集任务时使用。

[0191] 需要说明的是,本实施例中处理器所执行的步骤的具体实现过程,可以参照图1~7对应的实施例提供的数据处理方法中的实现过程,此处不再赘述。

[0192] 本发明的实施例所提供的数据处理设备,在给服务器分配采集任务时是基于各个服务器的负载来进行的,且无需针对每个数据块部署一个服务器,解决了相对技术中存在的服务之间没有关联的问题,降低了所需资源,且可以实现均衡采集;同时,采集过程中实现了服务的高可用性。

[0193] 基于前述实施例,本发明的实施例提供一种计算机可读存储介质,该计算机存储介质存储有一个或多个程序,该一个或多个程序可以被处理器执行,以实现如图1~7对应的实施例提供的数据处理方法中的步骤。

[0194] 本领域内的技术人员应明白,本发明的实施例可提供为方法、系统、或计算机程序产品。因此,本发明可采用硬件实施例、软件实施例、或结合软件和硬件方面的实施例的形式。而且,本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器和光学存储器等)上实施的计算机程序产品的形式。

[0195] 本发明是参照根据本发明实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0196] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0197] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0198] 以上所述,仅为本发明的较佳实施例而已,并非用于限定本发明的保护范围。

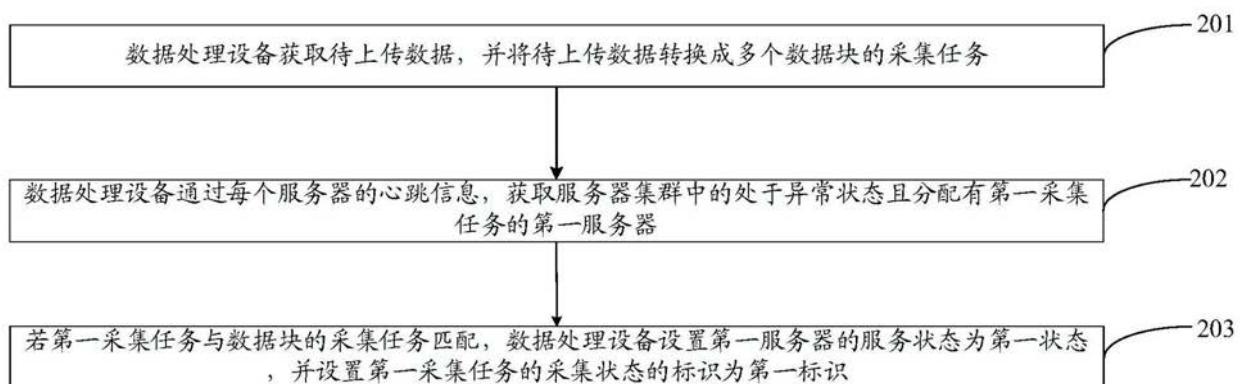
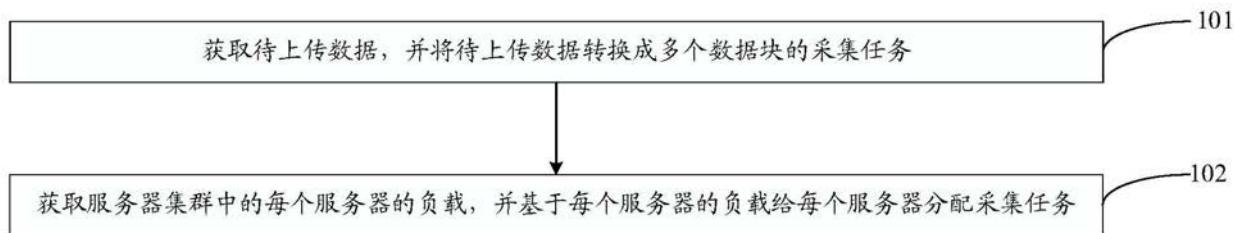




图3



图4

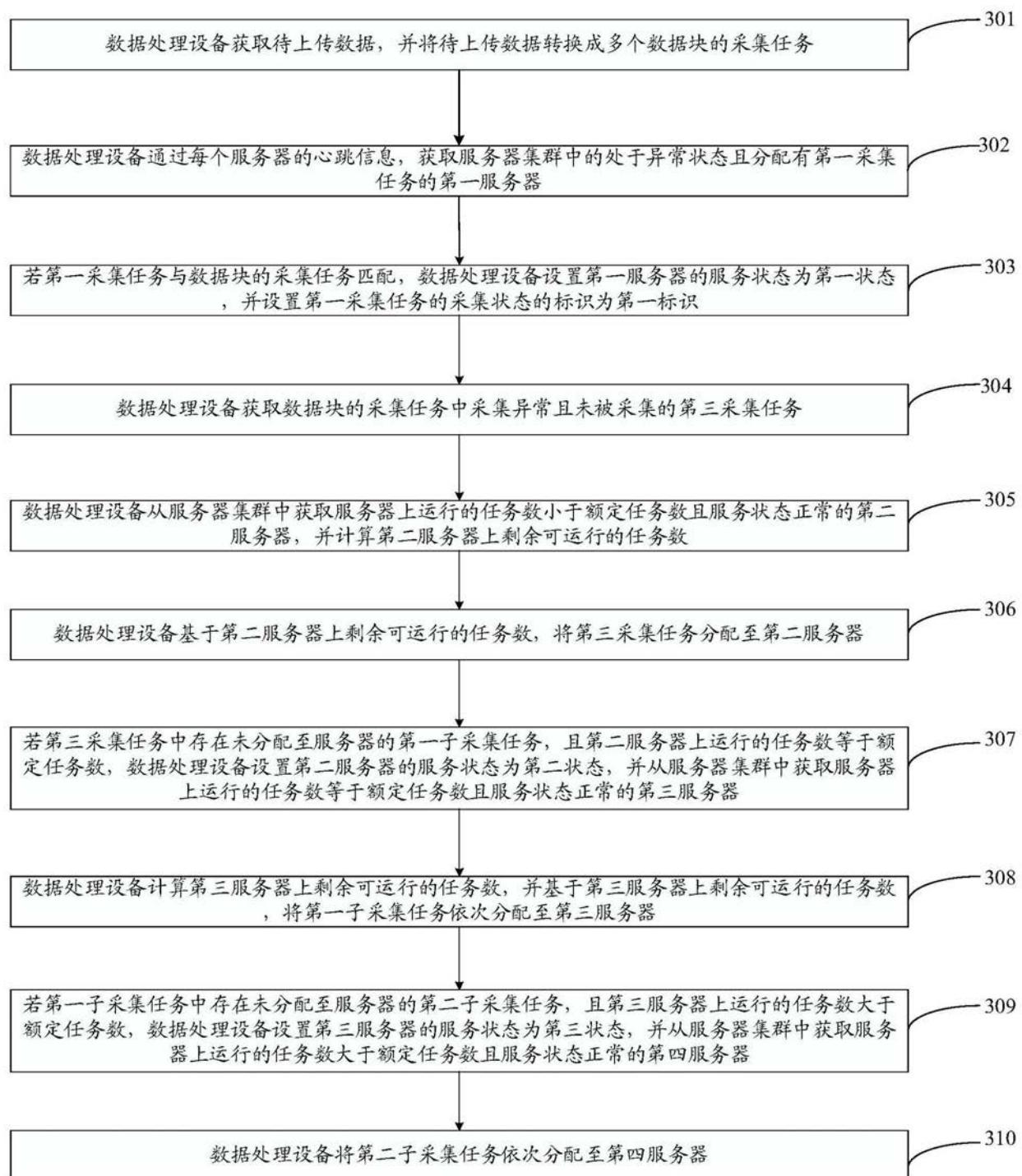


图5



图6



图7

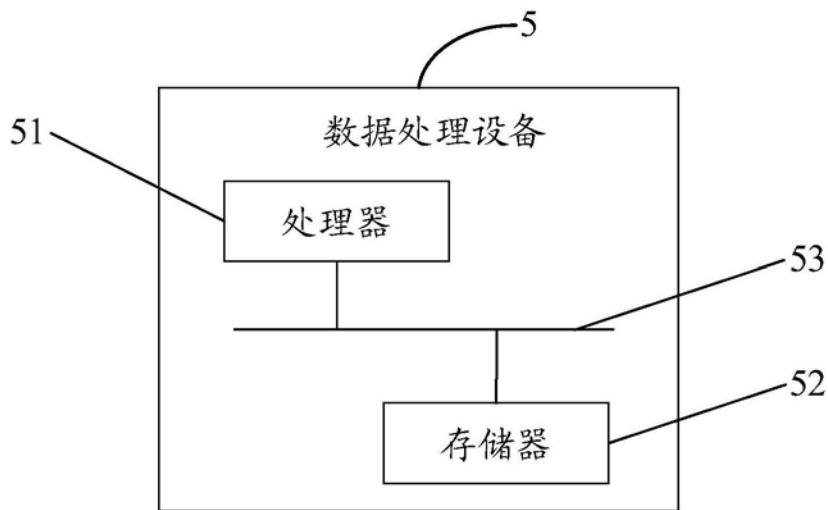


图8