(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 113347249 B
(45)授权公告日 2022.11.29

(21)申请号 202110605581.1
(22)申请日 2021.05.31
(65)同一申请的已公布的文献号
申请公布号 CN 113347249 A
(43)申请公布日 2021.09.03
(73)专利权人 中国工商银行股份有限公司
地址 100140 北京市西城区复兴门内大街55号
(72)发明人 孙伟 庄琴 王东青 李治中
(74)专利代理机构 北京三友知识产权代理有限公司 11127
专利代理师 任默闻 王涛
(51)Int.Cl.
H04L 67/10(2022.01)
H04L 67/1008(2022.01)

(56)对比文件
CN 112162865 A,2021.01.01
US 2017078373 A1,2017.03.16
CN 111694663 A,2020.09.22
CN 111221887 A,2020.06.02
US 2019370263 A1,2019.12.05
CN 111367984 A,2020.07.03
CN 112596806 A,2021.04.02
CN 112307122 A,2021.02.02
US 2019057122 A1,2019.02.21
史宝山.大数据湖存储模式建设探讨.《广播电视信息》.2018,(第06期),

审查员 杜少凤

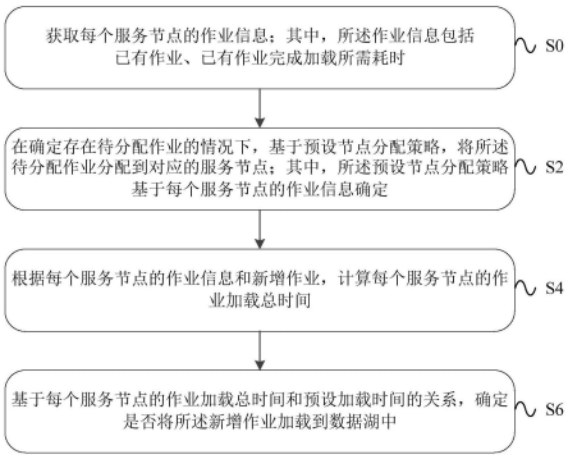
权利要求书2页 说明书10页 附图2页

(54)发明名称

一种作业加载方法、装置及设备

(57)摘要

本说明书实施例提供了一种作业加载方法、装置及设备,所述作业加载方法、装置及设备可用于大数据技术领域。所述方法包括获取每个服务节点的作业信息;其中,作业信息包括已有作业、已有作业完成加载所需耗时;在确定存在待分配作业的情况下,基于预设节点分配策略,将待分配作业分配到对应的服务节点;其中,预设节点分配策略基于每个服务节点的作业信息确定;根据每个服务节点的作业信息和新增作业,计算每个服务节点的作业加载总时间;基于每个服务节点的作业加载总时间和预设加载时间的关系,确定是否将新增作业加载到数据湖中。利用本说明书实施例可以在满足下游对作业加载时长要求的同时,充分利用生产上每台服务器的资源。



CN 113347249 B

1.一种作业加载方法,其特征在于,应用于数据湖开发管理系统,所述数据湖开发管理系统中包括多个服务节点,所述服务节点用于将作业加载到数据湖中,所述方法包括:

获取每个服务节点的作业信息;其中,所述作业信息包括已有作业、已有作业完成加载所需耗时;

在确定存在待分配作业的情况下,基于预设节点分配策略,将所述待分配作业分配到对应的服务节点;其中,所述预设节点分配策略基于每个服务节点的作业信息确定;

根据每个服务节点的作业信息和新增作业,计算每个服务节点的作业加载总时间;

基于每个服务节点的作业加载总时间和预设加载时间的关系,确定是否将所述新增作业加载到数据湖中;

其中,所述基于预设节点分配策略,将所述待分配作业分配到对应的服务节点,包括:

基于预设加载时间和每个服务节点的已有作业完成加载所需耗时,确定每个服务节点的固定权重;

基于每个服务节点的固定权重,确定第一次分配处理时每个服务节点的非固定权重;

根据第一次分配处理时每个服务节点的非固定权重,确定第一次分配处理时第一待分配作业的服务节点。

2.根据权利要求1所述的方法,其特征在于,所述根据第一次分配处理时每个服务节点的非固定权重,确定第一次分配处理时第一待分配作业的服务节点,包括:

将所述非固定权重中数值最大的服务节点作为所述第一次分配处理时第一待分配作业的服务节点。

3.根据权利要求1所述的方法,其特征在于,还包括:

基于每个服务节点的固定权重和第一次分配处理时每个服务节点的非固定权重,确定第二次分配处理时每个服务节点的非固定权重;

相应的,根据第二次分配处理时每个服务节点的非固定权重,确定第二次分配处理时第二待分配作业的服务节点。

4.根据权利要求3所述的方法,其特征在于,所述基于每个服务节点的固定权重和第一次分配处理时每个服务节点的非固定权重,确定第二次分配处理时每个服务节点的非固定权重,包括:

获取第一数值;其中,所述第一数值为第一次分配处理时所有服务节点对应的非固定权重中最大值;

计算第一次分配处理时所有服务节点的非固定权重之和;

基于所述第一数值和所述第一次分配处理时所有服务节点的非固定权重之和,更新第一数值,获得第一次分配处理更新后的非固定权重;

根据所述第一次分配处理更新后的非固定权重和固定权重比,获得第二次分配处理时每个服务节点的非固定权重。

5.根据权利要求3所述的方法,其特征在于,所述根据每个服务节点的作业信息和新增作业,计算每个服务节点的作业加载总时间,包括:

在所述第二次分配处理时每个服务节点的非固定权重与所述固定权重相同时,统计每个服务节点上新增作业的数量;

根据每个服务节点的作业信息和新增作业的数量,计算每个服务节点的作业加载总时

间。

6.根据权利要求5所述的方法,其特征在于,所述基于每个服务节点的作业加载总时间和预设加载时间的关系,确定是否将所述新增作业加载到数据湖中,包括:

在目标服务节点的作业加载总时间小于预设值时,将目标服务节点上新增作业的节点信息保存到数据库的配置信息中;其中,所述预设值根据所述预设加载时间确定;

基于所述配置信息生成执行脚本;

将所述执行脚本提交到版本管理系统,以使所述版本管理系统发布所述执行脚本;

将新增作业加载到数据湖中。

7.根据权利要求6所述的方法,其特征在于,所述基于每个服务节点的作业加载总时间和预设加载时间的关系,确定是否将所述新增作业加载到数据湖中,还包括:

在目标服务节点的作业加载总时间大于等于预设值时,发送新增服务节点的提示信息;

基于所述提示信息,获取新增服务节点的作业信息;

基于新增服务节点的作业信息和除所述目标服务节点外其余服务节点的作业信息,确定预设节点分配策略;

相应的,在确定存在待分配作业的情况下,基于预设节点分配策略,将所述待分配作业分配到对应的服务节点;根据每个服务节点的作业信息和新增作业,计算每个服务节点的作业加载总时间;基于每个服务节点的作业加载总时间和预设加载时间的关系,确定是否将所述新增作业加载到数据湖中。

8.一种作业加载装置,其特征在于,包括:

获取模块,用于获取每个服务节点的作业信息;其中,所述作业信息包括已有作业、已有作业完成加载所需耗时;

分配模块,用于在确定存在待分配作业的情况下,基于预设节点分配策略,将所述待分配作业分配到对应的服务节点;其中,所述预设节点分配策略基于每个服务节点的作业信息确定;

计算模块,用于根据每个服务节点的作业信息和新增作业,计算每个服务节点的作业加载总时间;

确定模块,用于基于每个服务节点的作业加载总时间和预设加载时间的关系,确定是否将所述新增作业加载到数据湖中;

其中,所述分配模块还用于基于预设加载时间和每个服务节点的已有作业完成加载所需耗时,确定每个服务节点的固定权重;基于每个服务节点的固定权重,确定第一次分配处理时每个服务节点的非固定权重;根据第一次分配处理时每个服务节点的非固定权重,确定第一次分配处理时第一待分配作业的服务节点。

9.一种作业加载设备,其特征在于,包括至少一个处理器以及存储计算机可执行指令的存储器,所述处理器执行所述指令时实现权利要求1-7中任意一项所述方法的步骤。

10.一种计算机可读存储介质,其特征在于,其上存储有计算机指令,所述指令被执行时实现权利要求1-7中任一项所述方法的步骤。

一种作业加载方法、装置及设备

技术领域

[0001] 本申请涉及大数据技术领域，特别涉及一种作业加载方法、装置及设备。

背景技术

[0002] 目前数据湖主要以HADOOP集群为平台构建，这样贴源集中可以存储含业务价值的数据，也可以实现覆盖全集团、境内外、行内外来源的上百个应用，为全行提供数据共享服务。

[0003] 随着源业务数据文件全入湖的推行，每个月度版本新增上千个文件入湖，其中，文件入湖可以通过作业进行加载。然而，随着作业数的大量增加容易导致各服务器上作业数分配不合理，从而无法在期望的时间内完成文件入湖，对下游业务造成影响。

[0004] 因此，业内亟需一种可以解决上述技术问题的技术方案。

发明内容

[0005] 本说明书实施例提供了一种作业加载方法、装置及设备，可以满足下游对作业加载时长的要求，还可以充分利用生产上每台服务器的资源。

[0006] 本说明书提供的一种作业加载方法、装置及设备是包括以下方式实现的。

[0007] 一种作业加载方法，应用于数据湖开发管理系统，所述数据湖开发管理系统中包括多个服务节点，所述服务节点用于将作业加载到数据湖中，所述方法包括：获取每个服务节点的作业信息；其中，所述作业信息包括已有作业、已有作业完成加载所需耗时；在确定存在待分配作业的情况下，基于预设节点分配策略，将所述待分配作业分配到对应的服务节点；其中，所述预设节点分配策略基于每个服务节点的作业信息确定；根据每个服务节点的作业信息和新增作业，计算每个服务节点的作业加载总时间；基于每个服务节点的作业加载总时间和预设加载时间的关系，确定是否将所述新增作业加载到数据湖中。

[0008] 一种作业加载装置，包括：获取模块，用于获取每个服务节点的作业信息；其中，所述作业信息包括已有作业、已有作业完成加载所需耗时；分配模块，用于在确定存在待分配作业的情况下，基于预设节点分配策略，将所述待分配作业分配到对应的服务节点；其中，所述预设节点分配策略基于每个服务节点的作业信息确定；计算模块，用于根据每个服务节点的作业信息和新增作业，计算每个服务节点的作业加载总时间；确定模块，用于基于每个服务节点的作业加载总时间和预设加载时间的关系，确定是否将所述新增作业加载到数据湖中。

[0009] 一种作业加载设备，包括至少一个处理器以及存储计算机可执行指令的存储器，所述处理器执行所述指令时实现本说明书实施例中任意一个方法实施例的步骤。

[0010] 一种计算机可读存储介质，其上存储有计算机指令，所述指令被执行时实现本说明书实施例中任意一个方法实施例的步骤。

[0011] 本说明书提供的一种作业加载方法、装置及设备。一些实施例中可以获取每个服务节点的作业信息，其中，作业信息包括已有作业、已有作业完成加载所需耗时，进而在确

定存在待分配作业的情况下，基于预设节点分配策略，将待分配作业分配到对应的服务节点，其中，预设节点分配策略基于每个服务节点的作业信息确定。还可以根据每个服务节点的作业信息和新增作业，计算每个服务节点的作业加载总时间，基于每个服务节点的作业加载总时间和预设加载时间的关系，确定是否将新增作业加载到数据湖中。本说明书实施例中基于预设节点分配策略将每个服务节点上的作业数控制在有效范围内，不仅可以满足下游对作业加载时长的要求，还可以充分利用生产上每台服务器的资源。

附图说明

[0012] 此处所说明的附图用来提供对本说明书的进一步理解，构成本说明书的一部分，并不构成对本说明书的限定。在附图中：

[0013] 图1是本说明书提供的一种作业加载方法的一个实施例的流程示意图；

[0014] 图2是本说明书提供的一种作业加载装置的一个实施例的模块结构示意图；

[0015] 图3是本说明书提供的一种作业加载服务器的一个实施例的硬件结构框图。

具体实施方式

[0016] 为了使本技术领域的人员更好地理解本说明书中的技术方案，下面将结合本说明书实施例中的附图，对本说明书实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本说明书中的一部分实施例，而不是全部的实施例。基于本说明书中的一个或多个实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都应当属于本说明书实施例保护的范围。

[0017] 下面以一个具体的应用场景为例对本说明书实施方案进行说明。具体的，图1是本说明书提供的一种作业加载方法的一个实施例的流程示意图。虽然本说明书提供了如下述实施例或附图所示的方法操作步骤或装置结构，但基于常规或者无需创造性的劳动在所述方法或装置中可以包括更多或者部分合并后更少的操作步骤或模块单元。

[0018] 本说明书提供的一种实施方案可以应用于数据湖开发管理系统，数据湖开发管理系统中可以包括多个服务节点，所述服务节点可以用于将作业加载到数据湖中。其中，每个服务节点可以是服务器。所述服务器可以包括单台计算机设备，也可以包括多个服务器组成的服务器集群，或者分布式系统的服务器结构等。

[0019] 需要说明的是，下述实施例描述并不对基于本说明书的其他可扩展到的应用场景中的技术方案构成限制。具体的一种实施例如图1所示，本说明书提供的一种作业加载方法的一种实施例中，所述方法可以包括以下步骤。

[0020] S0：获取每个服务节点的作业信息；其中，所述作业信息包括已有作业、已有作业完成加载所需耗时。

[0021] 其中，每个服务节点可以用于将作业加载到数据湖中。每个服务节点可以是服务器。所述服务器可以包括单台计算机设备，也可以包括多个服务器组成的服务器集群，或者分布式系统的服务器结构等。

[0022] 一些实施场景中，作业信息至少可以包括已有作业、已有作业完成加载所需耗时。一些实施场景中，作业信息还可以包括入湖文件名、文件字段信息、表名、作业名、对应服务节点名、期望完成时间、字段分隔符、换行符、文件编码格式和下传频度等。其中，作业信息

也可以理解为是元信息。例如,生产上有三台服务器A、B和C负责作业加载入湖,当前状态是A服务器上已有作业有Y1个,Y1个已有作业完成加载所需耗时为T1,B服务器上已有作业有Y2个,Y2个已有作业完成加载所需耗时为T2,C服务器上已有作业有Y3个,Y3个已有作业完成加载所需耗时为T3,下游业务处理系统对每台服务器完成作业加载时效为T,即期望完成时间为T。

[0023] 一些实施场景中,可以由开发人员在开发平台页面录入作业信息。

[0024] 一些实施场景中,开发人员在开发平台页面录入作业信息后,数据湖开发管理系统可以根据录入信息的时间,将相关信息按照时间线性顺序存储到数据库中,相应的,还可以同步展现在开发平台页面上。

[0025] 本说明书实施例中,通过按照时间维度存储作业信息,记录作业从第一次入湖,到后续不断修改,再到删除的全流程快照,这样可以对每个入湖作业做到整个生命周期的管理。

[0026] 一些实施场景中,在将作业信息录入后,可以通过开发平台页面显示每个作业的全流程快照,从而清晰的观察每个作业的变化,提高对作业信息的管理效率。

[0027] 本说明书实施例中,通过预先在数据库中录入或配置作业信息,这样,在需要时,可以直接从中获取,从而提高后续处理效率。

[0028] S2:在确定存在待分配作业的情况下,基于预设节点分配策略,将所述待分配作业分配到对应的服务节点;其中,所述预设节点分配策略基于每个服务节点的作业信息确定。

[0029] 本说明书实施例中,在获取每个服务节点的作业信息后,可以基于预设节点分配策略,将待分配作业分配到对应的服务节点。其中,预设节点分配策略可以基于每个服务节点的作业信息确定。其中,待分配作业的数量可以为一个或多个。待分配作业可以理解为需要分配服务节点的作业。

[0030] 一些实施例中,所述基于预设节点分配策略,将所述待分配作业分配到对应的服务节点,可以包括:基于预设加载时间和每个服务节点的已有作业完成加载所需耗时,确定每个服务节点的固定权重;基于每个服务节点的固定权重,确定第一次分配处理时每个服务节点的非固定权重;根据第一次分配处理时每个服务节点的非固定权重,确定第一次分配处理时第一待分配作业的服务节点。其中,预设加载时间可以理解为期望完成时间。

[0031] 一些实施场景中,可以将预设加载时间和每个服务节点的已有作业完成加载所需耗时做差,然后将差值作为每个服务节点的固定权重。其中,固定权重可以表示每个服务节点上预留加载时间。预设节点分配策略的原理是基于服务节点上预留加载时间为权重进行动态生成。

[0032] 一些实施场景中,在确定每个服务节点的固定权重后,可以计算每个服务节点的固定权重比,然后将固定权重比对应的数值分别作为第一次分配处理时每个服务节点的非固定权重。

[0033] 例如,生产上有三台服务器A、B和C负责作业加载入湖,当前状态是A服务器上已有作业有Y1个,Y1个已有作业完成加载所需耗时为T1,B服务器上已有作业有Y2个,Y2个已有作业完成加载所需耗时为T2,C服务器上已有作业有Y3个,Y3个已有作业完成加载所需耗时为T3,下游业务处理系统对每台服务器完成作业的预设加载时间为T,则基于预设加载时间和每个服务节点的已有作业完成加载所需耗时确定的每个服务节点的固定权重可以为Zx

＝T-Tx(x＝1、2、3)，相应的，固定权重比为Z1:Z2:Z3，则第一次分配处理时A、B、C三个服务节点的非固定权重分别为Z1、Z2、Z3。

[0034] 一些实施场景中，在确定每个服务节点的非固定权重后，可以基于每个服务节点的非固定权重，确定第一次分配处理时第一待分配作业的服务节点。

[0035] 一些实施场景中，所述根据第一次分配处理时每个服务节点的非固定权重，确定第一次分配处理时第一待分配作业的服务节点，可以包括：将所述非固定权重中数值最大的服务节点作为所述第一次分配处理时第一待分配作业的服务节点。其中，第一待分配作业可以是待分配作业中任意一个作业。例如，A、B、C三个服务节点的固定权重比为4:2:1，则第一次分配处理时A、B、C三个服务节点的非固定权重分别为4，2，1，此时由于最大数值为4，则可以将待分配作业分配到4对应的A服务节点上。

[0036] 一些实施场景中，在确定第一次分配处理时第一待分配作业的服务节点后，还可以基于每个服务节点的固定权重和第一次分配处理时每个服务节点的非固定权重，确定第二次分配处理时每个服务节点的非固定权重；相应的，根据第二次分配处理时每个服务节点的非固定权重，确定第二次分配处理时第二待分配作业的服务节点。

[0037] 一些实施场景中，由于待处理作业可以包括多个，所以为了使数据湖开发管理系统中每个服务节点得到高效利用，需要进行多次分配处理才能将所有待分配作业分配到对应的服务节点上。也就是，在第一次分配处理后，可以根据第一次分配处理时每个服务节点的非固定权重和固定权重，确定第二次分配处理时每个服务节点的非固定权重，进而可以根据第二次分配处理时每个服务节点的非固定权重确定第二次分配处理时第二待分配作业的服务节点。其中，第二待分配作业可以为第一次分配处理后剩余待分配作业中任意一个作业。

[0038] 一些实施场景中，所述基于每个服务节点的固定权重和第一次分配处理时每个服务节点的非固定权重，确定第二次分配处理时每个服务节点的非固定权重，可以包括：获取第一数值；其中，所述第一数值为第一次分配处理时所有服务节点对应的非固定权重中最大值；计算第一次分配处理时所有服务节点的非固定权重之和；基于所述第一数值和所述第一次分配处理时所有服务节点的非固定权重之和，更新第一数值，获得第一次分配处理更新后的非固定权重；根据所述第一次分配处理更新后的非固定权重和所述固定权重比，获得第二次分配处理时每个服务节点的非固定权重。

[0039] 一些实施场景中，在确定下一次分配处理(记为P_{i+1})时每个服务节点的非固定权重时，可以先获取本次分配处理(记为P_i)中所有服务节点对应的非固定权重中最大数值(记为a)，然后计算P_i中所有非固定权重的和(记为b)，用(a-b)替换P_i中a，获得P_i中更新后的非固定权重，最后将P_i中更新后的非固定权重与固定权重相加，获得P_{i+1}时的非固定权重。

[0040] 一些实施场景中，在确定第二次分配处理时每个服务节点的非固定权重后，可以比较第二次分配处理时每个服务节点的非固定权重与每个服务节点的固定权重是否相同，若不相同，则可以继续基于上述相同方式确定每次分配处理时每个服务节点的非固定权重，并基于确定的非固定权重确定待分配作业的服务节点。具体实现方式相似，对此不做赘述。

[0041] 一些实施场景中，在确定第二次分配处理时每个服务节点的非固定权重与每个服

务节点的固定权重相同时，可以根据每个服务节点的作业信息和新增作业，计算每个服务节点的作业加载总时间。

[0042] S4：根据每个服务节点的作业信息和新增作业，计算每个服务节点的作业加载总时间。

[0043] 本说明书实施例中，在将待分配作业分配到对应的服务节点后，可以根据每个服务节点的作业信息和新增作业，计算每个服务节点的作业加载总时间。

[0044] 一些实施例中，所述根据每个服务节点的作业信息和新增作业，计算每个服务节点的作业加载总时间，可以包括：在所述第二次分配处理时每个服务节点的非固定权重与所述固定权重相同时，统计每个服务节点上新增作业的数量；根据每个服务节点的作业信息和新增作业的数量，计算每个服务节点的作业加载总时间。其中，当分配处理时每个服务节点的非固定权重与固定权重第一次相同时可以理解为第一轮分配结束，当分配处理时每个服务节点的非固定权重与固定权重第二次相同时可以理解为第二轮分配结束。

[0045] 一些实施场景中，根据每个服务节点的作业信息和新增作业的数量，计算每个服务节点的作业加载总时间时，首先可以根据作业信息中已有作业、已有作业完成加载所需耗时确定每个作业的加载耗时，然后确定每个服务节点上已有作业和新增作业的总数量，最后基于每个作业的加载耗时和服务节点上作业的总数量，确定该服务节点的作业加载总时间。例如，上述A服务器上原有Y1个作业，完成加载需要耗时为T1，平均一个作业耗时T＝T1/Y1，一轮分配处理后，A服务器新增了A1个作业，则服务器A的作业加载总时间为Ta＝(Y1+A1)×T。

[0046] S6：基于每个服务节点的作业加载总时间和预设加载时间的关系，确定是否将所述新增作业加载到数据湖中。

[0047] 本说明书实施例中，在确定每个服务节点的作业加载总时间后，可以基于每个服务节点的作业加载总时间和预设加载时间的关系，确定是否将新增作业加载到数据湖中。其中，预设加载时间可以理解为期望完成时间。预设加载时间可以根据实际场景中下游业务处理系统的需求进行设定，本说明书对此不做限定。

[0048] 一些实施例中，所述基于每个服务节点的作业加载总时间和预设加载时间的关系，确定是否将所述新增作业加载到数据湖中，可以包括：在目标服务节点的作业加载总时间小于预设值时，将目标服务节点上新增作业的节点信息保存到数据库的配置信息中；其中，所述预设值根据所述预设加载时间确定；基于所述配置信息生成执行脚本；将所述执行脚本提交到版本管理系统，以使所述版本管理系统发布所述执行脚本；将新增作业加载到数据湖中。其中，目标服务节点可以是参与分配的所有服务节点中的任意一个服务节点。预设值可以是预设加载时间的90％、80％等，本说明书对此不做限定。

[0049] 一些实施场景中，在目标服务节点的作业加载总时间小于预设值时，可以说明对待分配作业进行分配的结果正确，此时可以将目标服务节点上新增作业的节点信息保存到数据库的配置信息中，进一步可以基于配置信息生成相应的执行脚本。其中，执行脚本可以理解为文件入湖信息代码，其可以包括配置信息初始化、建表脚本、作业配置节点脚本等。

[0050] 一些实施场景中，在生成执行脚本后，可以将执行脚本提交到版本管理系统GIT，以便GIT发布执行脚本，这样作业在已分配的服务节点上就可以进行文件入湖加载。其中，执行脚本通过版本管理系统发布后可以在生产上运行。

[0051] 一些实施场景中，所述基于每个服务节点的作业加载总时间和预设加载时间的关系，确定是否将所述新增作业加载到数据湖中，还可以包括：在目标服务节点的作业加载总时间大于等于预设值时，发送新增服务节点的提示信息；基于所述提示信息，获取新增服务节点的作业信息；基于新增服务节点的作业信息和除所述目标服务节点外其余服务节点的作业信息，确定预设节点分配策略；相应的，在确定存在待分配作业的情况下，基于预设节点分配策略，将所述待分配作业分配到对应的服务节点；根据每个服务节点的作业信息和新增作业，计算每个服务节点的作业加载总时间；基于每个服务节点的作业加载总时间和预设加载时间的关系，确定是否将所述新增作业加载到数据湖中。

[0052] 一些实施场景中，目标服务节点的作业加载总时间大于等于预设值时，则可以在开发平台页面显示提示信息，这样，后续不再向该服务节点分配待处理作业，即在下一轮分配过程中，该目标服务节点不再参与新增作业的分配，目标服务节点只处理已经分配的作业，而新增服务节点加入到预设节点分配策略中。其中，提示信息可以为“请新增服务节点”、“该服务节点已超额，请增加新服务节点”等。

[0053] 本说明书实施例中，在目标服务节点的作业加载总时间大于等于预设值时，通过重新新增服务节点，可以实现对服务节点的高效利用。

[0054] 本说明书实施例中，通过多轮服务节点的分配，可以将多个作业按照对应权重数较为合理的分配到每台服务器上。

[0055] 下面以一个具体实施例对本说明书实施例提高的作业加载方式进行说明。本实施例中，有三台服务器A、B、C负责文件加载入湖为例，当前状态是A服务器上有Y1个作业，完成加载需要耗时为T1，B服务器上有Y2个作业，完成加载需要耗时为T2，C服务器上有Y3个作业，完成加载需要耗时为T3。下游业务处理系统期望在T时间内完成加载。每台服务器上预留加载时间为：Zx＝T-Tx(x＝1、2、3)，即A、B、C服务器的固定权重比为Z1:Z2:Z3。现将M个待处理作业分配到服务器A、B和C。非固定权重在每次分配处理时会根据一定的规则变化，第一次分配处理时的非固定权重为Z1、Z2、Z3。本实施例中，以固定权重比为4:2:1为例。

[0056] 具体的，如表1所示，第一次分配处理时，A、B、C的“非固定权重”(即获取服务器前的非固定权重)分别是4、2、1，因为4是其中最大的，4对应的是A服务器，所以这次选到的服务器是A(即此次将待处理作业中一个作业分配到服务器A上)。在确定此次选到的服务器后，可以用当前被选中的服务器A的非固定权重-各个服务器的非固定权重之和(即4-7＝-3)更新被选中服务器A的非固定权重，没被选中的服务器的“非固定权重”不做变化，从而获得此次更新后A、B、C的“非固定权重”(获取服务器后的非固定权重)为-3、2、1。

[0057] 第二次分配处理时，计算第一次分配处理得到的更新后A、B、C的“非固定权重”和“固定权重”的和，获得此次A、B、C的“非固定权重”(即获取服务器前的非固定权重)分别是1、4、2，因为4是其中最大的，4对应的是B服务器，所以这次选到的服务器是B(即此次将待处理作业中剩余作业中一个作业分配到服务器B上)。在确定此次选到的服务器后，可以用当前被选中的服务器B的非固定权重-各个服务器的非固定权重之和(即4-7＝-3)更新被选中服务器B的非固定权重，没被选中的服务器的“非固定权重”不做变化，从而获得此次更新后A、B、C的“非固定权重”(获取服务器后的非固定权重)为1、-3、2。

[0058] 第三次分配处理，计算第二次分配处理得到的更新后A、B、C的“非固定权重”和“固定权重”的和，获得此次A、B、C的“非固定权重”(即获取服务器前的非固定权重)分别是5、-

1、3，因为5是其中最大的，5对应的是A服务器，所以这次选到的服务器是A(即此次将待处理作业中剩余作业中一个作业分配到服务器A上)。在确定此次选到的服务器后，可以用当前被选中的服务器A的非固定权重-各个服务器的非固定权重之和(即5-7＝-2)更新被选中服务器A的非固定权重，没被选中的服务器的“非固定权重”不做变化，从而获得此次更新后A、B、C的“非固定权重”(获取服务器后的非固定权重)为-2、-1、3。

[0059] 以此类推，当第八次分配处理时，计算第七次分配处理得到的更新后A、B、C的“非固定权重”和“固定权重”的和，获得此次A、B、C的“非固定权重”(即获取服务器前的非固定权重)分别是4、2、1，其与固定权重相同说明此时完成了一轮动态服务器节点分配。

[0060] 本实施例中，在完成一轮动态服务器节点分配后，可以统计每台服务器上作业数(已有作业和新增作业之和)，计算每台服务器完成加载的总时间。例如，上述A服务器上原有Y1个作业，完成加载需要耗时为T1，平均一个作业耗时T＝T1/Y1，新增了A1个作业，则总时间为Ta＝(Y1+A1)×T。进一步，在获得每台服务器完成加载的总时间后，可以判断其是否在完成期望时间(T的90％)内，如在，则说明分配正确。如不在，则说明该服务器不能再新增作业，此时，可以在开发平台页面提示新增服务器，这样，后续不再向该服务器分配待处理作业，即在下一轮分配过程中，该服务器不再参与新增作业的分配，其只处理已经分配的作业，而新增服务器加入到下一轮动态服务器节点分配中。这样，经过多轮服务器节点分配，可以将M个作业按照对应权重合理的分配到每台服务器上。

[0061] 表1

[0062]

序号	获取服务器前的非固定权重	选中的服务器	获取服务器后的非固定权重
1	4,2,1	A	-3,2,1
2	1,4,2	B	1,-3,2
3	5,-1,3	A	-2,-1,3
4	2,1,4	C	2,1,-3
5	6,3,-2	A	-1,3,-2
6	3,5,-1	B	3,-2,-1
7	7,0,0	A	0,0,0
8	4,2,1	A	-3,2,1

[0063] 需要说明的是，一些实施场景中，当获取服务器前的非固定权重中出现相同的非固定权重时，可以从中相同的非固定权重中随机选择一个服务器作为选中的服务器，还可以将本轮已进行的分配处理中选中次数最少的服务器作为选中的服务器，然后按照相似步骤进行处理，具体处理过程可以相互参照，对此不作赘述。当然，上述只是进行示例性说明，所属领域技术人员在本申请技术精髓的启示下，还可能做出其它变更，但只要其实现的功能和效果与本申请相同或相似，均应涵盖于本申请保护范围内。

[0064] 本说明书实施例中，基于预设节点分配策略控制每个服务节点上的新增作业数，不仅可以保证作业在期望的时间内完成加载，而且可以自动判断是否需要新增服务节点。

[0065] 本说明书实施例，在自动化数据湖开发管理系统中通过预设节点分配策略将每台服务器上的作业数控制在有效的范围内，既可以满足下游对作业加载时长的要求，又可以充分利用生产上每台服务器的资源。

[0066] 当然，上述只是进行示例性说明，本说明书实施例不限于上述举例，所属领域技术

人员在本申请技术精髓的启示下，还可能做出其它变更，但只要其实现的功能和效果与本申请相同或相似，均应涵盖于本申请保护范围内。

[0067] 从以上的描述中，可以看出，本申请实施例可以获取每个服务节点的作业信息，其中，作业信息包括已有作业、已有作业完成加载所需耗时，进而在确定存在待分配作业的情况下，基于预设节点分配策略，将待分配作业分配到对应的服务节点，其中，预设节点分配策略基于每个服务节点的作业信息确定。还可以根据每个服务节点的作业信息和新增作业，计算每个服务节点的作业加载总时间，基于每个服务节点的作业加载总时间和预设加载时间的关系，确定是否将新增作业加载到数据湖中。本说明书实施例中基于预设节点分配策略将每个服务节点上的作业数控制在有效范围内，不仅可以满足下游对作业加载时长的要求，还可以充分利用生产上每台服务器的资源。

[0068] 本说明书中上述方法的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参照即可，每个实施例重点说明的都是与其他实施例的不同之处。相关之处参见方法实施例的部分说明即可。

[0069] 基于上述所述一种作业加载方法，本说明书一个或多个实施例还提供一种作业加载装置。所述的装置可以包括使用了本说明书实施例所述方法的系统(包括分布式系统)、软件(应用)、模块、组件、服务器、客户端等并结合必要的实施硬件的装置。基于同一创新构思，本说明书实施例提供的一个或多个实施例中的装置如下面的实施例所述。由于装置解决问题的实现方案与方法相似，因此本说明书实施例具体的装置的实施可以参见前述方法的实施，重复之处不再赘述。以下所使用的，术语“单元”或者“模块”可以实现预定功能的软件和/或硬件的组合。尽管以下实施例所描述的装置较佳地以软件来实现，但是硬件，或者软件和硬件的组合的实现也是可能并被构想的。

[0070] 具体地，图2是本说明书提供的一种作业加载装置的一个实施例的模块结构示意图，如图2所示，本说明书提供的一种作业加载装置可以包括：获取模块120，分配模块122，计算模块124，确定模块126。

[0071] 获取模块120，可以用于获取每个服务节点的作业信息；其中，所述作业信息包括已有作业、已有作业完成加载所需耗时；

[0072] 分配模块122，可以用于在确定存在待分配作业的情况下，基于预设节点分配策略，将所述待分配作业分配到对应的服务节点；其中，所述预设节点分配策略基于每个服务节点的作业信息确定；

[0073] 计算模块124，可以用于根据每个服务节点的作业信息和新增作业，计算每个服务节点的作业加载总时间；

[0074] 确定模块126，科研用于基于每个服务节点的作业加载总时间和预设加载时间的关系，确定是否将所述新增作业加载到数据湖中。

[0075] 需要说明的，上述所述的装置根据方法实施例的描述还可以包括其他的实施方式，具体的实现方式可以参照相关方法实施例的描述，在此不作一一赘述。

[0076] 本说明书还提供一种作业加载设备的实施例，包括处理器及用于存储处理器可执行指令的存储器，所述指令被所述处理器执行时实现包括以下步骤：获取每个服务节点的作业信息；其中，所述作业信息包括已有作业、已有作业完成加载所需耗时；在确定存在待分配作业的情况下，基于预设节点分配策略，将所述待分配作业分配到对应的服务节点；其

中,所述预设节点分配策略基于每个服务节点的作业信息确定;根据每个服务节点的作业信息和新增作业,计算每个服务节点的作业加载总时间;基于每个服务节点的作业加载总时间和预设加载时间的关系,确定是否将所述新增作业加载到数据湖中。

[0077] 需要说明的,上述所述的设备根据方法或装置实施例的描述还可以包括其他的实施方式。具体的实现方式可以参照相关方法实施例的描述,在此不作一一赘述。

[0078] 本说明书所提供的方法实施例可以在移动终端、计算机终端、服务器或者类似的运算装置中执行。以运行在服务器上为例,图3是本说明书提供的一种作业加载服务器的一个实施例的硬件结构框图,该服务器可以是上述实施例中的作业加载装置或作业加载设备。如图3所示,服务器10可以包括一个或多个(图中仅示出一个)处理器100(处理器100可以包括但不限于微处理器MCU或可编程逻辑器件FPGA等的处理装置)、用于存储数据的存储器200、以及用于通信功能的传输模块300。本领域普通技术人员可以理解,图3所示的结构仅为示意,其并不对上述电子装置的结构造成限定。例如,服务器10还可包括比图3中所示更多或者更少的组件,例如还可以包括其他的处理硬件,如数据库或多级缓存、GPU,或者具有与图3所示不同的配置。

[0079] 存储器200可用于存储应用软件的软件程序以及模块,如本说明书实施例中的作业加载方法对应的程序指令/模块,处理器100通过运行存储在存储器200内的软件程序以及模块,从而执行各种功能应用以及数据处理。存储器200可包括高速随机存储器,还可包括非易失性存储器,如一个或者多个磁性存储装置、闪存、或者其他非易失性固态存储器。在一些实例中,存储器200可进一步包括相对于处理器100远程设置的存储器,这些远程存储器可以通过网络连接至计算机终端。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0080] 传输模块300用于经由一个网络接收或者发送数据。上述的网络具体实例可包括计算机终端的通信供应商提供的无线网络。在一个实例中,传输模块300包括一个网络适配器(Network Interface Controller,NIC),其可通过基站与其他网络设备相连从而可与互联网进行通讯。在一个实例中,传输模块300可以为射频(Radio Frequency,RF)模块,其用于通过无线方式与互联网进行通讯。

[0081] 上述对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下,在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外,在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中,多任务处理和并行处理也是可以的或者可能是有利的。

[0082] 本说明书提供的上述实施例所述的方法或装置可以通过计算机程序实现业务逻辑并记录在存储介质上,所述的存储介质可以计算机读取并执行,实现本说明书实施例所描述方案的效果。所述存储介质可以包括用于存储信息的物理装置,通常是将信息数字化后再以利用电、磁或者光学等方式的媒体加以存储。所述存储介质可以包括:利用电能方式存储信息的装置如,各式存储器,如RAM、ROM等;利用磁能方式存储信息的装置如,硬盘、软盘、磁带、磁芯存储器、磁泡存储器、U盘;利用光学方式存储信息的装置如,CD或DVD。当然,还有其他方式的可读存储介质,例如量子存储器、石墨烯存储器等等。

[0083] 本说明书提供的上述作业加载方法或装置实施例可以在计算机中由处理器执行

相应的程序指令来实现，如使用windows操作系统的c++语言在PC端实现、linux系统实现，或其他例如使用android、iOS系统程序设计语言在智能终端实现，以及基于量子计算机的处理逻辑实现等。

[0084] 需要说明的是说明书上述所述的装置、设备、系统根据相关方法实施例的描述还可以包括其他的实施方式，具体的实现方式可以参照对应方法实施例的描述，在此不作一一赘述。

[0085] 本申请中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参照即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于硬件+程序类实施例而言，由于其基本相似于方法实施例，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

[0086] 为了描述的方便，描述以上装置时以功能分为各种模块分别描述。当然，在实施本说明书一个或多个时可以把部分模块的功能在同一个或多个软件和/或硬件中实现，也可以将实现同一功能的模块由多个子模块或子单元的组合实现等。

[0087] 本发明是参照根据本发明实施例的方法、装置、设备、系统的流程图和/或方框图来描述的。应理解可由计算机程序指令实现，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指定的功能的装置。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。

[0088] 本领域技术人员应明白，本说明书一个或多个实施例可提供为方法、系统或计算机程序产品。因此，本说明书一个或多个实施例可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。

[0089] 以上所述仅为本说明书一个或多个实施例的实施例而已，并不用于限制本本说明书一个或多个实施例。对于本领域技术人员来说，本说明书一个或多个实施例可以有各种更改和变化。凡在本申请的精神和原理之内所作的任何修改、等同替换、改进等，均应包含在权利要求范围之内。

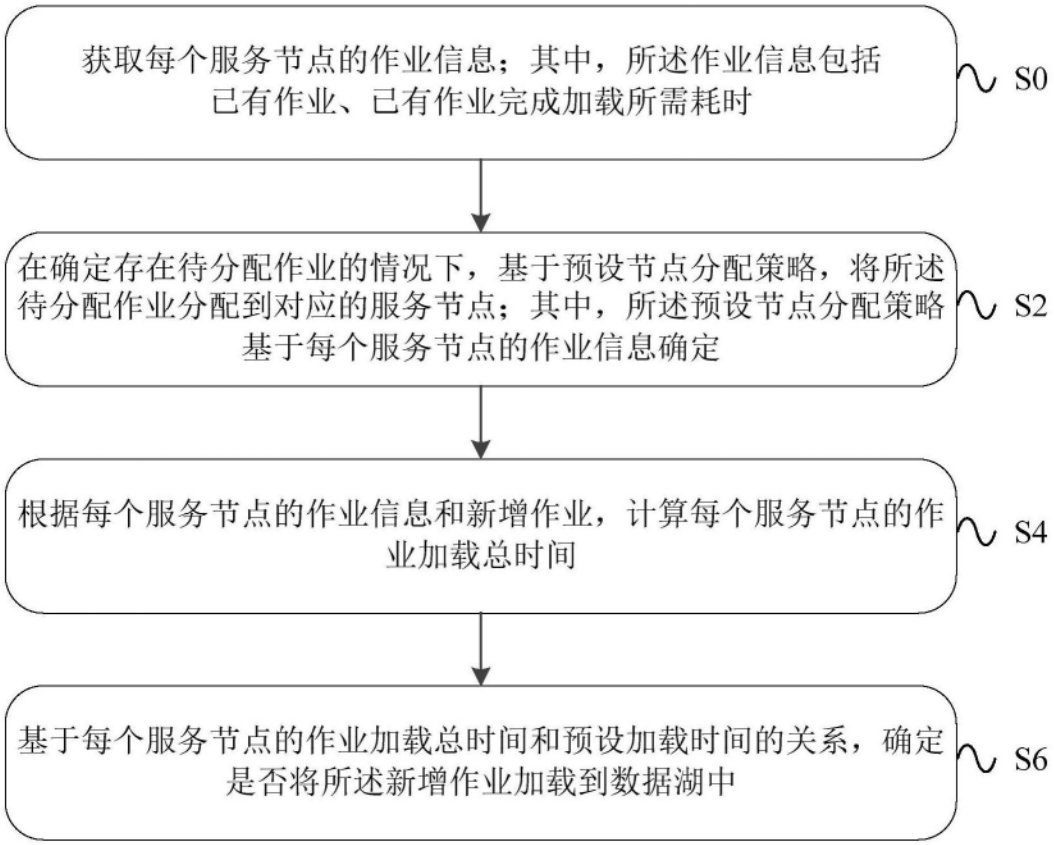

获取每个服务节点的作业信息；其中，所述作业信息包括已有作业、已有作业完成加载所需耗时
S0
在确定存在待分配作业的情况下，基于预设节点分配策略，将所述待分配作业分配到对应的服务节点；其中，所述预设节点分配策略基于每个服务节点的作业信息确定
S2
根据每个服务节点的作业信息和新增作业，计算每个服务节点的作业加载总时间
S4
基于每个服务节点的作业加载总时间和预设加载时间的关系，确定是否将所述新增作业加载到数据湖中
S6

图1

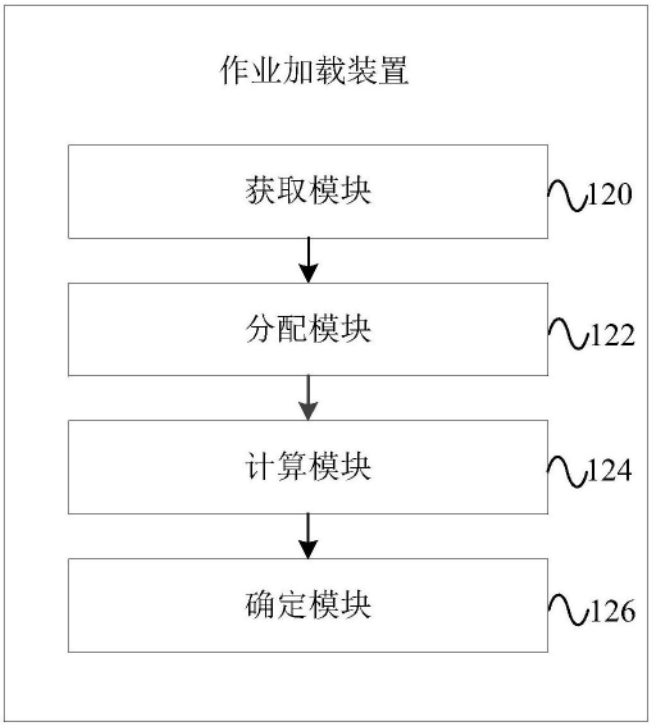

作业加载装置
获取模块
120
分配模块
122
计算模块
124
确定模块
126

图2

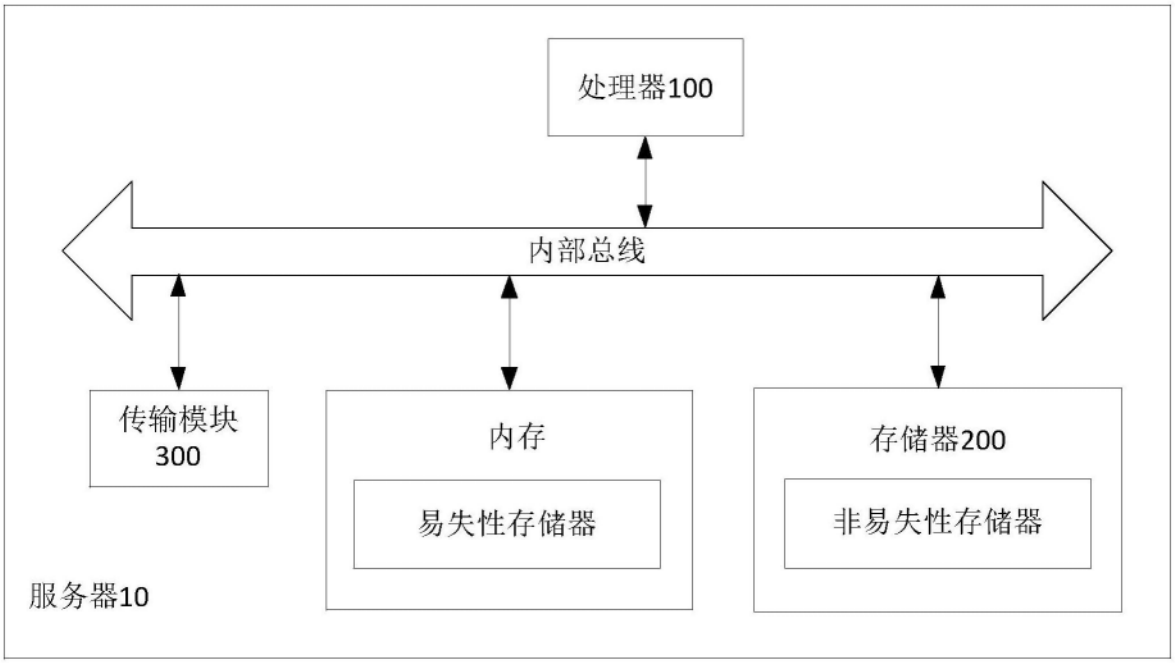


图3