



(12) 发明专利申请

(10) 申请公布号 CN 115599468 A

(43) 申请公布日 2023. 01. 13

(21) 申请号 202211375907.7

(22) 申请日 2022.11.04

(71) 申请人 北京百度网讯科技有限公司
地址 100085 北京市海淀区上地十街10号
百度大厦2层

(72) 发明人 张词校 唐杰 冯智 黄志

(74) 专利代理机构 北京柏杉松知识产权代理事
务所(普通合伙) 11413
专利代理师 丁芸 马敬

(51) Int. Cl.
G06F 9/445 (2018.01)

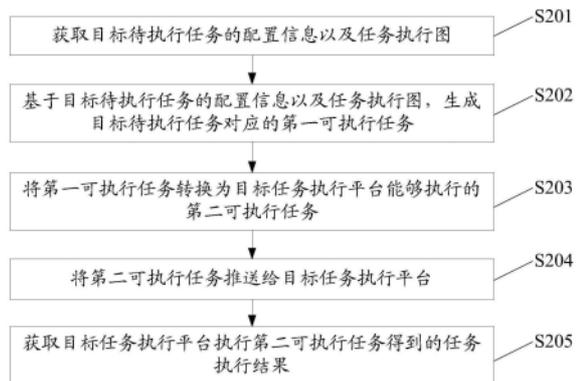
权利要求书4页 说明书18页 附图6页

(54) 发明名称

一种任务处理方法、系统、电子设备及存储介质

(57) 摘要

本公开提供了一种任务处理方法、系统、电子设备及存储介质,涉及数据处理技术领域,尤其涉及任务处理等技术领域。具体实现方案为:获取目标待执行任务的配置信息以及任务执行图;基于目标待执行任务的配置信息以及任务执行图,生成目标待执行任务对应的第一可执行任务;将第一可执行任务转换为目标任务执行平台能够执行的第二可执行任务;将第二可执行任务推送给目标任务执行平台;获取目标任务执行平台执行第二可执行任务得到的任务执行结果,表示目标待执行任务的执行结果,实现了任务的自定义配置执行,扩展了任务处理的业务场景。



1. 一种任务处理方法,包括:

获取目标待执行任务的配置信息以及任务执行图;其中,所述目标待执行任务的配置信息以及任务执行图为预先针对所述目标待执行任务自定义设置的;所述任务执行图中包括:执行所述目标待执行任务的目标任务执行平台;

基于所述目标待执行任务的配置信息以及任务执行图,生成所述目标待执行任务对应的第一可执行任务;

将所述第一可执行任务转换为所述目标任务执行平台能够执行的第二可执行任务;

将所述第二可执行任务推送给所述目标任务执行平台;

获取所述目标任务执行平台执行所述第二可执行任务得到的任务执行结果,所述第二可执行任务的任务执行结果,表示所述目标待执行任务的任务执行结果。

2. 根据权利要求1所述的方法,还包括:

判断当前实例节点是否达到负载均衡上限条件;

在所述当前实例节点没有达到负载均衡上限条件的情况下,从待执行任务列表中取出一个待执行任务作为目标待执行任务;

对所述目标待执行任务进行加锁处理;

在对所述目标待执行任务加锁成功的情况下,执行获取所述目标待执行任务的配置信息以及任务执行图。

3. 根据权利要求2所述的方法,其中,所述判断当前实例节点是否达到负载均衡上限条件,包括:

基于目标实例列表,确定当前有效的实例节点个数;所述目标实例列表为:各实例节点间隔预设时间段将自身信息和对应的时间信息,更新至实例列表中得到的;所述当前有效的实例节点表示:所述目标实例列表中该实例节点的时间信息,与当前系统时间之间的时间差在预设时间范围内;

确定当前的任务总数,所述任务总数包括待执行任务的个数以及正在执行任务的个数;

基于所述当前有效的实例节点个数、以及所述当前的任务总数,判断当前实例节点是否达到负载均衡上限条件。

4. 根据权利要求3所述的方法,其中,所述基于所述当前有效的实例节点个数、以及所述当前的任务总数,判断当前实例节点是否达到负载均衡上限条件,包括:

将所述当前的任务总数和所述当前有效的实例节点个数的商,与1的和,确定为负载上限阈值;

确定当前实例节点正在执行任务的个数;

在所述当前实例节点正在执行任务的个数小于所述负载上限阈值的情况下,确定当前实例节点没有达到负载均衡上限条件;

在所述当前实例节点正在执行任务的个数不小于所述负载上限阈值的情况下,确定当前实例节点达到负载均衡上限条件。

5. 根据权利要求2所述的方法,其中,所述对所述目标待执行任务进行加锁处理,包括:

在所述目标待执行任务的状态处于解锁状态的情况下,将所述目标待执行任务的状态更新为加锁状态。

6. 根据权利要求2所述的方法,还包括:

在对所述目标待执行任务加锁不成功的情况下,返回执行从待执行任务列表中取出一个待执行任务作为目标待执行任务。

7. 根据权利要求2所述的方法,其中,所述目标待执行任务的执行图中还包括:执行所述目标待执行任务的各任务执行节点对应的信息、以及各任务执行节点的执行顺序和连接关系信息;所述目标待执行任务的配置信息包括:各任务执行节点在执行所述目标待执行任务时用到的属性信息;

所述基于所述目标待执行任务的配置信息以及任务执行图,生成所述目标待执行任务对应的第一可执行任务,包括:

基于各所述任务执行节点对应的信息、各所述任务执行节点的执行顺序和连接关系信息、以及各任务执行节点在执行所述目标待执行任务时用到的属性信息,实例化生成所述目标待执行任务对应的第一可执行任务。

8. 根据权利要求2所述的方法,其中,所述目标任务执行平台包括:本地计算引擎,Spark计算引擎以及Flink计算引擎。

9. 根据权利要求2所述的方法,在得到所述第二可执行任务的执行结果之后,还包括:

对所述目标待执行任务进行解锁处理,并从所述待执行任务列表中删除所述目标待执行任务。

10. 根据权利要求2-9任一所述的方法,还包括:

确定是否存在处于加锁状态的加锁任务;

在存在处于加锁状态的加锁任务的情况下,针对每一加锁任务,确定该加锁任务是否被正常执行;

在该加锁任务没有被正常执行的情况下,对该加锁任务进行解锁处理,并将解锁处理得到的解锁任务加入所述待执行任务列表中。

11. 根据权利要求10所述的方法,其中,所述加锁任务中包含执行该加锁任务的实例节点的目标标识;所述确定该加锁任务是否被正常执行,包括:

基于所述目标标识,判断执行该加锁任务的实例节点是否为当前实例节点;

在执行该加锁任务的实例节点为当前实例节点的情况下,判断所述当前实例节点的任务执行列表中是否存在该加锁任务;

在所述当前实例节点的任务执行列表中不存在该加锁任务的情况下,确定该加锁任务被正常执行;

在所述当前实例节点的任务执行列表中不存在该加锁任务的情况下,确定该加锁任务没有被正常执行;

在执行该加锁任务的实例节点不是当前实例节点的情况下,判断执行该加锁任务的实例节点是否在目标实例列表中;

在执行该加锁任务的实例节点在目标实例列表中的情况下,确定该加锁任务被正常执行;

在执行该加锁任务的实例节点不在目标实例列表中的情况下,确定该加锁任务没有被正常执行。

12. 一种任务处理系统,包括:

信息获取模块,用于获取目标待执行任务的配置信息以及任务执行图;其中,所述目标待执行任务的配置信息以及任务执行图为预先针对所述目标待执行任务自定义设置的;所述任务执行图中包括:执行所述目标待执行任务的目标任务执行平台;

任务生成模块,用于基于所述目标待执行任务的配置信息以及任务执行图,生成所述目标待执行任务对应的第一可执行任务;

任务转换模块,用于将所述第一可执行任务转换为所述目标任务执行平台能够执行的第二可执行任务;

任务分配模块,用于将所述第二可执行任务推送给所述目标任务执行平台;

结果获取模块,用于获取所述目标任务执行平台执行所述第二可执行任务得到的任务执行结果,所述第二可执行任务的任务执行结果,表示所述目标待执行任务的任务执行结果。

13. 根据权利要求12所述的系统,还包括:

负载均衡模块,用于判断当前实例节点是否达到负载均衡上限条件;

任务确定模块,用于在所述负载均衡模块判断出所述当前实例节点没有达到负载均衡上限条件的情况下,从待执行任务列表中取出一个待执行任务作为目标待执行任务;

任务加锁模块,用于对所述目标待执行任务进行加锁处理,并在对所述目标待执行任务加锁成功的情况下,触发所述信息获取模块执行获取所述目标待执行任务的配置信息以及任务执行图。

14. 根据权利要求13所述的系统,其中,所述目标待执行任务的任务执行图中还包括:执行所述目标待执行任务的各任务执行节点对应的信息、以及各任务执行节点的执行顺序和连接关系信息;所述目标待执行任务的配置信息包括:各任务执行节点在执行所述目标待执行任务时用到的属性信息;所述目标任务执行平台包括:本地计算引擎,Spark计算引擎以及Flink计算引擎;

所述任务生成模块,具体用于:基于各所述任务执行节点对应的信息、各所述任务执行节点的执行顺序和连接关系信息、以及各任务执行节点在执行所述目标待执行任务时用到的属性信息,实例化生成所述目标待执行任务对应的第一可执行任务。

15. 根据权利要求13-14任一所述的系统,还包括:

任务状态确定模块,用于确定是否存在处于加锁状态的加锁任务;

任务执行情况确定模块,用于在所述任务状态确定模块确定存在处于加锁状态的加锁任务的情况下,针对每一加锁任务,确定该加锁任务是否被正常执行;

任务恢复模块,用于在所述任务执行情况确定模块确定该加锁任务没有被正常执行的情况下,对该加锁任务进行解锁处理,并将解锁处理得到的解锁任务加入所述待执行任务列表中。

16. 根据权利要求15所述的系统,其中,所述加锁任务中包含执行该加锁任务的实例节点的目标标识;所述确定该加锁任务是否被正常执行,包括:

基于所述目标标识,判断执行该加锁任务的实例节点是否为当前实例节点;

在执行该加锁任务的实例节点为当前实例节点的情况下,判断所述当前实例节点的任务执行列表中是否存在该加锁任务;

在所述当前实例节点的任务执行列表中存在该加锁任务的情况下,确定该加锁任务被正常执行;

在所述当前实例节点的任务执行列表中不存在该加锁任务的情况下,确定该加锁任务没有被正常执行;

在执行该加锁任务的实例节点不是当前实例节点的情况下,判断执行该加锁任务的实例节点是否在目标实例列表中;

在执行该加锁任务的实例节点在目标实例列表中的情况下,确定该加锁任务被正常执行;

在执行该加锁任务的实例节点不在目标实例列表中的情况下,确定该加锁任务没有被正常执行。

17. 一种电子设备,包括:

至少一个处理器;以及

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行权利要求1-11中任一项所述的方法。

18. 一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行根据权利要求1-11中任一项所述的方法。

19. 一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现根据权利要求1-11中任一项所述的方法。

一种任务处理方法、系统、电子设备及存储介质

技术领域

[0001] 本公开涉及数据处理技术领域,进一步涉及任务处理、任务调度等技术领域,尤其涉及一种任务处理方法、系统、电子设备及存储介质。

背景技术

[0002] 随着企业业务规模的不断扩大,用户数量的不断增加,企业对用户身份安全的重视程度不断的加深,针对企业自身的安全管理以及对用户的行为分析等对应的任务数量越来越多,扩展任务处理的业务场景,提高任务的执行效率,能够使得企业更好地实现自身的安全管理,以及提高对用户行为分析的时效性。

发明内容

[0003] 本公开提供了一种任务处理方法、系统、电子设备及存储介质。

[0004] 根据本公开的一方面,提供了一种任务处理方法,包括:

[0005] 获取目标待执行任务的配置信息以及任务执行图;其中,所述目标待执行任务的配置信息以及任务执行图为预先针对所述目标待执行任务自定义设置的;所述任务执行图中包括:执行所述目标待执行任务的目标任务执行平台;

[0006] 基于所述目标待执行任务的配置信息以及任务执行图,生成所述目标待执行任务对应的第一可执行任务;

[0007] 将所述第一可执行任务转换为所述目标任务执行平台能够执行的第二可执行任务;

[0008] 将所述第二可执行任务推送给所述目标任务执行平台;

[0009] 获取所述目标任务执行平台执行所述第二可执行任务得到的任务执行结果,所述第二可执行任务的任务执行结果,表示所述目标待执行任务的任务执行结果。

[0010] 根据本公开的另一方面,提供了一种任务处理系统,包括:

[0011] 信息获取模块,获取目标待执行任务的配置信息以及任务执行图;其中,所述目标待执行任务的配置信息以及任务执行图为预先针对所述目标待执行任务自定义设置的;所述任务执行图中包括:执行所述目标待执行任务的目标任务执行平台;

[0012] 任务生成模块,用于基于所述目标待执行任务的配置信息以及任务执行图,生成所述目标待执行任务对应的第一可执行任务;

[0013] 任务转换模块,用于将所述第一可执行任务转换为所述目标任务执行平台能够执行的第二可执行任务;

[0014] 任务分配模块,用于将所述第二可执行任务推送给所述目标任务执行平台;

[0015] 结果获取模块,用于获取所述目标任务执行平台执行所述第二可执行任务得到的任务执行结果,所述第二可执行任务的任务执行结果,表示所述目标待执行任务的任务执行结果。

[0016] 根据本公开的另一方面,提供了一种电子设备,包括:

- [0017] 至少一个处理器;以及
- [0018] 与所述至少一个处理器通信连接的存储器;其中,
- [0019] 所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行本公开中任一项所述的方法。
- [0020] 根据本公开的另一方面,提供了一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行本公开中任一项所述的方法。
- [0021] 根据本公开的另一方面,提供了一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现本公开中任一项所述的方法。
- [0022] 本公开实施例,实现了任务的自定义配置执行,扩展了任务处理的业务场景。
- [0023] 应当理解,本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征,也不用于限制本公开的范围。本公开的其他特征将通过以下的说明书而变得容易理解。

附图说明

- [0024] 附图用于更好地理解本方案,不构成对本公开的限定。其中:
- [0025] 图1是用于实现本公开实施例的任务处理方法的系统架构图;
- [0026] 图2是根据本公开任务处理方法的一种示意图;
- [0027] 图3是根据本公开任务执行图的一种示意图;
- [0028] 图4是根据本公开任务执行驱动器的一种示意图;
- [0029] 图5是根据本公开任务生成及分发执行的一种示意图;
- [0030] 图6是根据本公开任务处理方法的另一种示意图;
- [0031] 图7是根据本公开确定实例节点是否达到负载均衡上限的一种示意图;
- [0032] 图8是根据本公开任务数据展示的一种示意图;
- [0033] 图9是根据本公开加锁任务恢复的一种示意图;
- [0034] 图10是根据本公开确定任务是否被正常执行的一种示意图;
- [0035] 图11是根据本公开任务处理系统的一种示意图;
- [0036] 图12是用来实现本公开实施例的任务处理方法的电子设备的框图。

具体实施方式

[0037] 以下结合附图对本公开的示范性实施例做出说明,其中包括本公开实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本公开的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0038] 企业的IAM(Identity and Access Management,身份识别与访问管理)经过多次产品迭代和演进,在基础的账户/权限管理功能层面,已进入相对比较成熟的阶段。但随着企业业务规模的不断扩大,用户数量的不断增加,企业对用户身份安全的重视程度不断的加深,仍有企业频发各类身份安全相关的问题,使得企业扩展更高级别的安全管理和分析功能的身份识别势在必行,当前最前沿的用户身份安全管理和分析功能是自适应身份认证。

[0039] UEBA(User Entity Behavior Analysis,用户和实体行为分析)技术是自适应身

份认证的基础。UEBA最主要的功能是基于用户(包含自然人和设备实体等,统称为用户)在企业系统中的行为模式,分析用户是否存在异常和是否存在安全风险等,并将分析结果反馈给系统或者提示给系统管理人员,以便于系统管理人员给出相应的处置。

[0040] 然而,随着企业业务规模的不断扩大,用户数量的不断增加,小型企业可能每日会有几百条用户行为记录,而大型企业以及公有云等每日的用户行为数据量可能达到十亿甚至百亿级别,使得待进行用户行为分析的数据量巨大,相应的,待进行用户行为分析的任务数量也较大,且,不同企业其业务场景往往差异较大,使得如何处理不同业务场景下的数据任务成为亟待解决的问题。

[0041] 为解决上述问题,本公开提供了一种任务处理方法,获取目标待执行任务的配置信息以及任务执行图;其中,所述目标待执行任务的配置信息以及任务执行图为预先针对所述目标待执行任务自定义设置的,所述任务执行图中包括:执行所述目标待执行任务的目标任务执行平台;基于所述目标待执行任务的配置信息以及任务执行图,生成所述目标待执行任务对应的第一可执行任务;将所述第一可执行任务转换为所述目标任务执行平台能够执行的第二可执行任务;将所述第二可执行任务推送给所述目标任务执行平台;获取所述目标任务执行平台执行所述第二可执行任务得到的任务执行结果,所述第二可执行任务的任务执行结果,表示所述目标待执行任务的任务执行结果。

[0042] 本公开实施例中,可以预先针对不同的目标待执行任务,自定义设置该目标待执行任务的配置信息以及任务执行图,进而基于该自定义设置的配置信息以及任务执行图,生成目标待执行任务对应的第一可执行任务,实现了任务的自定义配置以及生成,扩展了任务处理的业务场景。进一步的,将第一可执行任务转换为目标任务执行平台能够执行的第二可执行任务,将转换后的第二可执行任务推送给目标任务执行平台进行执行,保证了任务执行过程中任务计算能力的灵活可伸缩性,使得任务的处理能够适应于不同企业的内部执行环境。

[0043] 本公开实施例提供的任务处理方法,可以应用于基于UEBA的用户异常行为分析场景中,相应的,本公开实施例中的任务可以为用户异常行为分析任务。当然,本公开实施例的任务处理方法也可以应用于其他需要执行任务的任何场景中。本公开实施例中以基于UEBA的用户异常行为分析任务为例进行说明。一个例子中,用于实现本公开实施例的任务处理方法的系统架构如图1所示。

[0044] 图1为基于UEBA的用户异常行为分析系统,该系统中任务控制模块用于实现本公开实施例中的任务处理过程,具体的,任务控制模块负责控制任务的整个生命周期,包括任务的开启、分配、组装、转换、重试、终止以及完成,根据任务生命周期的不同,任务控制模块可以分为:任务重试子模块、任务负载均衡子模块、任务组装子模块以及任务转换子模块。

[0045] 为防止任务的重复执行,任务在执行的过程中会被加锁,但系统可能会出现断电、宕机等突发情况,使得任务被中断而无法解锁,任务重试子模块能够将中断或执行失败但仍处于加锁状态的任务进行解锁恢复,并尝试重新执行。任务负载均衡子模块,负责将任务均衡的分发到不同的实例节点进行执行,以确保每个实例节点的负载均衡。示例性的,一个实例节点可以对应于一个机器设备。任务组装子模块,可以将用户自定义设置的任务配置信息以及任务执行图(TaskGraph),进行整合并组装成内存中一个完整的可执行任务,组装完成的可执行任务中包含一系列的任务执行逻辑(或称为算子),组装后的可执行任务能够

被执行、终止以及完成等。任务转换子模块,能够将实例节点本地组装好的可执行任务,通过语法转换,转换为不同执行平台下的可执行任务,从而使得任务可以推送至不同的执行平台上执行,以保证系统计算能力的灵活伸缩。

[0046] 图1系统中的控制执行即为控制并执行组装完成的可执行任务,该可执行任务中包含一系列任务执行逻辑(或算子),由系统中各组件进行执行,组件比如包括:数据接入组件、数据建模组件、风险引擎组件等,系统还包括对外服务组件等。其中,数据接入组件中可以包括:数据采集、数据过滤、数据转换以及事件视图等处理。数据建模组件中可以包括:基线建模、实体关联以及实体行为等的建模。风险引擎组件中可以包括:基线分析、聚类分析、静态规则以及风险评分等处理。对外服务组件中可以包括:配置接口、控制台以及风险数据等。

[0047] 不同数据源在数据获取方式、数据含义、数据格式以及数据完整性等方面通常有较大的差异,数据接入组件为消除不同数据源数据的差异,针对不同数据源获取的数据进行数据采集、数据过滤、数据转换以及事件视图等操作,以实现数据的视图化。

[0048] 数据接入组件可以从外部依赖获取数据,外部依赖表示不同的数据源,比如设备日志数据、审计日志数据、接口日志数据、数据库日志数据以及网络流量数据等。数据接入组件对不同数据源的数据可以以文件的形式,通过API(Application Programming Interface,应用程序接口)或者kafka队列等进行数据采集。数据采集之后,利用预先设置的规则,对采集的数据进行过滤,比如任务为统计用户登录频率,相应的规则可以是登录接口路径的表达式等。进一步的,针对过滤后的数据,将该数据中的关键字(或值)与预设字段进行对应,转换为键值对,并对过滤后的数据中的用户身份进行解析,得到用户对应的实体信息,再基于得到的键值对与实体信息生成事件视图。示例性的,获取到的数据为{"name": "zhangsan", "age": 20, "height": 180},订阅的视图为["name", "age"],提取出的数据(即生成的事件视图)为{"name": "zhangsan", "age": 20},该事件视图可以以数据流的形式传输给数据建模组件。还可以对该事件视图进行存储,比如存储至Mysql数据库、Redis(Remote dictionary server,远程字典服务)数据库或者图数据库中。

[0049] 图1系统中数据建模组件根据数据接入组件提供的数据,生成用户行为的一系列描述数据,然后以数据流的形式传输给风险引擎组件。数据建模组件根据数据接入组件提供的数据,比如,生成包含账户登录频率、访问地点等的实体行为信息,生成包含账户与账户、账户与设备等的实体关联行为信息,以及生成包含确定是否超出系统规则和阈值的异常信息等的基线建模信息,其中,系统规则和阈值可以是根据已有的用户访问环境、访问行为经验等抽象出的场景和模式生成的。

[0050] 图系统1中风险引擎组件基于数据建模组件生成的一系列描述数据,分析用户事件是否异常,主要分析的方式包含:基线分析、聚类分析、静态规则以及风险评分等。其中,风险引擎组件在对数据进行分析的过程中可以调用其他分布式计算引擎,比如Spark计算引擎等。

[0051] 图1系统中风险引擎组件分析的结果以数据流的形式传输给对外服务组件,对外服务组件中提供了配置接口以及控制台,对管理员开放,进而管理员可以通过配置接口以及控制台进行任务的配置和操作。对外服务组件中还提供了风险数据等的查询接口,使得IdaaS(Alibaba Cloud Identity as a Service,阿里云应用身份服务)、IAM以及其他应用

能够通过对外服务组件查询风险数据等。

[0052] 下面对本公开实施例提供的任务处理方法进行详细说明。

[0053] 本公开实施例提供的任务处理方法可以应用于电子设备,如服务器设备、集群设备以及云服务设备等等,优选的,用于执行该任务处理方法的执行主体可以是任一实例节点,一个实例节点对应于一个机器设备。本公开实施例提供的任务处理方法应用场景可以是基于UEBA的用户异常行为分析场景等。UEBA支持的用户异常行为分析场景可以包括:短时间异地登录、非常用地区登录、非常用IP(Internet Protocol,互联网协议)地址登录、非常用时间登录、短时间内频繁登录、非白名单IP登录、密钥访问频率异常、密钥未使用监测等。

[0054] 参见图2,图2为本公开实施例提供的一种任务处理方法的流程示意图,包括以下步骤:

[0055] S201,获取目标待执行任务的配置信息以及任务执行图。

[0056] 其中,目标待执行任务的配置信息以及任务执行图为预先针对目标待执行任务自定义设置的,任务执行图中包括:执行目标待执行任务的目标任务执行平台。

[0057] 本公开实施例中,以上述图1所示的系统是基于UEBA的用户异常行为分析系统为例进行说明,该系统中包含可完全自定义的任务执行模型(图1中未示出),该任务执行模型支持用户自己开发、配置完整的任务执行过程,从而使得系统能够支持各种不同的任务分析场景。具体的,用户可以通过图1所示的对外服务组件中的配置接口和控制台,在任务执行模型中自定义设置任务的配置信息以及任务执行图等,以动态的扩展任务的业务场景。

[0058] 不同量级的企业,其内部产生的用户行为数据的量级是不同的。对于中小型企业,可能每日的用户行为数据量只有几千到几万条,这种量级的数据,通常在本地物理机/虚拟机内存中即可完成数据的分析。对于大型企业,每日的用户行为数据量可能会达到千万或者上亿条的级别,这种量级的数据,企业一般无法提供相应规格的物理机/虚拟机作为计算资源,故而需要专有的计算引擎实现。

[0059] 在对目标待执行任务进行自定义配置时,可以根据该目标待执行任务的任务量、所需消耗的资源等,为该目标待执行任务配置执行其的目标任务执行平台。一个例子中,可以通过该目标待执行任务占用计算资源的大小、计算的复杂度以及计算的频率等,估计本地物理机/虚拟机是否可以完成该目标待执行任务,如果能够完成,将该目标待执行任务的目标任务执行平台配置为本地计算引擎,如果本地计算引擎不足以支撑目标待执行任务的完成,那么将该目标待执行任务的目标任务执行平台配置为专有的计算引擎。示例性的,任务执行图中所包括的执行目标待执行任务的目标任务执行平台,具体可以是目标任务执行平台的名称标识或者地址信息等。

[0060] 在一种可能的实施方式中,上述目标任务执行平台可以包括:本地计算引擎,Spark计算引擎以及Flink计算引擎。

[0061] 本公开实施例中,同时支撑任务在本地计算引擎(本地物理机/虚拟机),Spark计算引擎以及Flink计算引擎中执行,保证了任务执行过程中任务计算能力的灵活可伸缩性,使得任务的处理能够适应于不同企业的内部执行环境。

[0062] 基于UEBA的用户异常行为分析系统的实例节点在检测到需要对任务进行处理的情况下,获取目标待执行任务的配置信息以及任务执行图。

[0063] S202,基于目标待执行任务的配置信息以及任务执行图,生成目标待执行任务对应的第一可执行任务。

[0064] 一个例子中,实例节点利用上述图1所示的系统中的任务组装子模块,对目标待执行任务的配置信息以及任务执行图进行组装,以实例化生成目标待执行任务对应的第一可执行任务。

[0065] 在一种可能的实施方式中,目标待执行任务的执行图中还可以包括:执行目标待执行任务的各任务执行节点对应的信息、以及各任务执行节点的执行顺序和连接关系信息;目标待执行任务的配置信息包括:各任务执行节点在执行目标待执行任务时用到的属性信息。其中,任务执行节点对应的信息可以是任务执行节点的标识、名称等等。

[0066] 一个完整的可以执行的任务包括:一个任务执行图,该任务执行图中包括执行任务的各任务执行节点(TaskExecuteNode),每个任务执行节点都可以是一个指定的执行逻辑或者算子(TaskExecutor)。一个任务执行图是一个有向无环图,图中定义了各任务执行节点之间的执行顺序和连接关系。

[0067] 示例性的,一个任务执行图如图3所示,该任务执行图中包括:执行任务的各任务执行节点A、B、C、D和E,以及任务执行节点A、B、C、D和E之间的执行顺序和连接关系,即任务执行节点A运行完毕之后,运行任务执行节点B和C,任务执行节点B和C运行完毕之后,运行任务执行节点D,任务执行节点D运行完毕之后,运行任务执行节点E。

[0068] 每个任务执行节点均有对应的任务执行配置信息,任务执行配置信息定义了该任务执行节点在执行任务的过程中用到的各种可配置参数,即该任务执行节点在执行目标待执行任务时用到的属性信息。

[0069] 示例性的,任务执行节点对应的算子为统计异地登录事件,算子中定义了一个最小时间间隔,以及在最小时间间隔内发生了异地登录时,将此次登录统计为异地登录异常事件的分析逻辑。相应的,该任务执行节点的任务执行配置信息包括:最小时间间隔,异地登录事件的判定,以及异地登录异常事件的判定方式(即在最小时间间隔内发生了异地登录)。

[0070] 示例性的,任务执行节点对应的算子为统计访问频率事件,算子中定义了一个设定时间段,以及在该设定时间段内发生的访问次数超过设定阈值时,将此次访问统计为访问频率异常事件的分析逻辑。相应的,该任务执行节点的任务执行配置信息包括:设定时间段,设定阈值,以及访问频率异常事件的判定方式(即在设定时间段内发生的访问次数超过设定阈值)。

[0071] 一个例子中,每个任务执行节点对应一个类型(type),不同类型的任务执行节点执行逻辑不同,针对每个任务执行节点的类型均可以自定义设置。示例性的,每个任务执行节点对应的类型可以是上述图1所示系统中预埋的常用分析器(SystemTaskExecutor)类型,也可以是按照UEBA提供的接口标准实现的分析器(CustomTaskExecutor)上传至UEBA服务中的类型,这些数据分析的类型可以由用户命名。即实现了任务执行节点的自定义配置,扩展了任务分析的业务场景。

[0072] 相应的,基于目标待执行任务的配置信息以及任务执行图,生成目标待执行任务对应的第一可执行任务,包括:基于各任务执行节点对应的信息、各任务执行节点的执行顺序和连接关系信息、以及各任务执行节点在执行目标待执行任务时用到的属性信息,实例

化生成目标待执行任务对应的第一可执行任务。

[0073] 对各任务执行节点、各任务执行节点的执行顺序和连接关系、以及各任务执行节点在执行目标待执行任务时用到的属性信息进行组装,实例化生成目标待执行任务对应的第一可执行任务。具体的,将各任务执行节点、各任务执行节点的执行顺序和连接关系、以及各任务执行节点在执行目标待执行任务时用到的属性信息,组装成内存中的可执行任务对象,即生成可以执行的实例化程序对象,对象中指定了各任务执行节点之间的关联关系(比如任务执行节点之间的执行顺序、启动以及终止等等)。

[0074] 本公开实施例中,目标待执行任务的执行图中包括:执行目标待执行任务的各任务执行节点对应的信息、以及各任务执行节点的执行顺序和连接关系信息,目标待执行任务的配置信息包括:各任务执行节点在执行目标待执行任务时用到的属性信息,以便于能够基于各任务执行节点对应的信息、各任务执行节点的执行顺序和连接关系信息、以及各任务执行节点在执行目标待执行任务时用到的属性信息,实例化生成目标待执行任务对应的第一可执行任务,实现了任务可自定义设置,动态扩展了任务分析场景,使得能够以高可扩展的方式满足用户对各种业务场景任务的分析需求。

[0075] 参见图2,S203,将第一可执行任务转换为目标任务执行平台能够执行的第二可执行任务。

[0076] 一个例子中,实例节点利用上述图1所示的系统中的任务转换子模块,将第一可执行任务,通过语法转换,转换为目标任务执行平台能够执行的第二可执行任务。

[0077] 在一种可能的实施方式中,在UEBA系统(即上述基于UEBA的用户异常行为分析系统)中定义了任务执行驱动器(TaskGraphDriver),如图4所示,相应于上述目标任务执行平台,任务执行驱动器可以包括:本地任务执行驱动器(LocalTaskGraphDriver)、Spark任务执行驱动器(SparkTaskGraphDriver)以及Flink任务执行驱动器(FlinkTaskGraphDriver)。

[0078] 任务执行驱动器的作用是:能够将预先自定义配置好的任务执行图,转换为对应任务执行平台上能够执行的可执行任务(TaskWrapper)。示例性的,如图5所示,任务执行驱动器能够将预先自定义配置好的任务执行图,转换为对应任务执行平台上能够执行的可执行任务(TaskWrapper)。

[0079] 示例性的,执行目标待执行任务的目标任务执行平台为Spark计算引擎,使用parseTaskGraph(对任务执行图作句法分析)方法将第一可执行任务转换为Spark计算引擎能够执行的第二可执行任务,比如,Spark的RDD(Resilient Distributed Datasets,弹性分布式数据集)定义的任务等。其中,parseTaskGraph根据计算引擎的不同,实现不同的语法转换。

[0080] S204,将第二可执行任务推送给目标任务执行平台。

[0081] 将转换后的第二可执行任务推送给目标任务执行平台,以使目标任务执行平台执行该第二可执行任务,并在完成第二可执行任务之后返回第二可执行任务的执行结果。

[0082] S205,获取目标任务执行平台执行第二可执行任务得到的任务执行结果。

[0083] 一个例子中,可以接收目标任务执行平台完成第二可执行任务之后返回的任务执行结果,也可以从目标任务执行平台中读取目标任务执行平台执行第二可执行任务得到的

任务执行结果等。

[0084] 其中,第二可执行任务的任务执行结果,表示目标待执行任务的任务执行结果。

[0085] 本公开实施例中,可以预先针对不同的目标待执行任务,自定义设置该目标待执行任务的配置信息以及任务执行图,进而基于该自定义设置的配置信息以及任务执行图,生成目标待执行任务对应的第一可执行任务,实现了任务的自定义配置以及生成,扩展了任务处理的业务场景。进一步的,将第一可执行任务转换为目标任务执行平台能够执行的第二可执行任务,将转换后的第二可执行任务推送给目标任务执行平台进行执行,保证了任务执行过程中任务计算能力的灵活可伸缩性,使得任务的处理能够适应于不同企业的内部执行环境。

[0086] 参见图6,图6为本公开实施例提供的另一种任务处理方法的流程示意图,包括以下步骤:

[0087] S601,判断当前实例节点是否达到负载均衡上限条件。

[0088] 当前实例节点判断自身是否达到负载均衡上限条件,一个例子中,负载均衡上限条件可以是当前实例节点正在执行的任务数量是否达到预设数值,或者当前实例节点的内存利用率是否达到设定利用率等等。

[0089] 本公开实施例中,每一当前实例节点都是无状态的。

[0090] S602,在当前实例节点没有达到负载均衡上限条件的情况下,从待执行任务列表中取出一个待执行任务作为目标待执行任务。

[0091] 在当前实例节点达到负载均衡上限条件的情况下,执行步骤S609当前实例节点结束任务的执行。在当前实例节点没有达到负载均衡上限条件的情况下,查询数据库中所有状态为待执行的待执行任务列表,进一步在该待执行任务列表中有待执行任务的情况下,从待执行任务列表中取出一个待执行任务作为目标待执行任务。一个例子中,可以对待执行任务列表中的待执行任务进行遍历,逐一取出待执行任务列表中的待执行任务。

[0092] S603,对目标待执行任务进行加锁处理。

[0093] 为了保证任意时刻一个待执行任务只能被一个实例节点执行,在当前实例节点没有达到负载均衡上限条件的情况下,从待执行任务列表中取出一个待执行任务作为目标待执行任务之后,对该目标待执行任务进行抢占,即加锁处理。

[0094] 在一种可能的实施方式中,对目标待执行任务进行加锁处理,包括:在目标待执行任务的状态处于解锁状态的情况下,将目标待执行任务的状态更新为加锁状态。

[0095] 示例性的,可以采用语句“update task set status= ‘locked’ where status= ‘unlocked’ where id=1”,对目标待执行任务进行加锁处理,该语句表示:标识为1的目标待执行任务,如果其状态为解锁状态,则将其状态设置为加锁状态。

[0096] 由于同一时间只有一个实例节点能够成功执行该语句,其余实例节点都会由于这个任务已经被更新成了locked(加锁)状态而执行失败,这样就达到了任意时刻一个任务只能被一个实例节点执行的目的,只有加锁成功的这个实例节点可以执行该任务。

[0097] 本公开实施例中,在目标待执行任务的状态处于解锁状态的情况下,将目标待执行任务的状态更新为加锁状态,抢占该目标待执行任务,使得任意时刻一个待执行任务只能被一个实例节点执行,避免了任务被重复执行。

[0098] 一个例子中,步骤S601-步骤S603的实现,可以通过当前实例节点利用上述图1

所示的系统中的任务负载均衡子模块完成的。

[0099] S604,在对目标待执行任务加锁成功的情况下,执行获取目标待执行任务的配置信息以及任务执行图。

[0100] 其中,目标待执行任务的配置信息以及任务执行图为预先针对目标待执行任务自定义设置的,任务执行图中包括:执行目标待执行任务的目标任务执行平台。

[0101] 在对目标待执行任务加锁成功的情况下,执行该目标待执行任务。在对目标待执行任务加锁不成功的情况下,返回执行步骤S602中从待执行任务列表中取出一个待执行任务作为目标待执行任务,直至待执行任务列表中没有待执行任务。

[0102] 本公开实施例中,在当前实例节点没有达到负载均衡上限条件,且对目标待执行任务加锁不成功的情况下,重新从待执行任务列表中取出一个待执行任务作为目标待执行任务,直至待执行任务列表中没有待执行任务,实现了负载均衡基础上任务的执行。

[0103] S605,基于目标待执行任务的配置信息以及任务执行图,生成目标待执行任务对应的第一可执行任务。

[0104] S606,将第一可执行任务转换为目标任务执行平台能够执行的第二可执行任务。

[0105] S607,将第二可执行任务推送给目标任务执行平台。

[0106] S608,获取目标任务执行平台执行第二可执行任务得到的任务执行结果。

[0107] 第二可执行任务的任务执行结果,表示目标待执行任务的任务执行结果。

[0108] 其中,步骤S604获取目标待执行任务的配置信息以及任务执行图,以及步骤S605-步骤S608的实现过程,可参考上述步骤S201-步骤S205的实现过程,本公开实施例在此不再赘述。

[0109] S609,结束。

[0110] 本公开实施例中,采用无状态的实例节点自动进行任务的负载均衡,能够实现实例节点的任意扩缩容,在保证各实例节点能够负载所有任务的情况下,增加了任务处理的可伸缩性。实例节点通过对任务进行加锁抢占,保证了任意时刻一个待执行任务只能被一个实例节点执行,避免了任务被重复执行,进而提高了任务执行的实时性和准确性。并且,可以预先针对不同的目标待执行任务,自定义设置该目标待执行任务的配置信息以及任务执行图,进而基于该自定义设置的配置信息以及任务执行图,生成目标待执行任务对应的第一可执行任务,实现了任务的自定义配置以及生成,扩展了任务处理的业务场景。进一步的,将第一可执行任务转换为目标任务执行平台能够执行的第二可执行任务,将转换后的第二可执行任务推送给目标任务执行平台进行执行,保证了任务执行过程中任务计算能力的灵活可伸缩性,使得任务的处理能够适应于不同企业的内部执行环境。

[0111] 在一种可能的实施方式中,如图7所示,上述步骤S601判断当前实例节点是否达到负载均衡上限条件的实施过程,包括:

[0112] S701,基于目标实例列表,确定当前有效的实例节点个数。

[0113] 目标实例列表为:各实例节点间隔预设时间段将自身信息和对应的时间信息,更新至实例列表中得到的;当前有效的实例节点表示:目标实例列表中该实例节点的时间信息,与当前系统时间之间的时间差在预设时间范围内。

[0114] 每个实例节点可以间隔预设时间段,将自身信息和对应的时间信息,更新到数据库的实例列表集合(uebaInstanceList)中,存储为目标实例列表。该目标实例列表中即存

储有每个实例节点最新的自身信息和对应的时间信息。其中,自身信息可以是实例节点的标识或名称等,对应的时间信息即为将自身信息更新至数据库时的时间戳信息,数据库比如可以是Redis数据库或者Mysql数据库等,预设时间段可以根据需求进行配置,比如可配置为5秒或10秒或20秒等。

[0115] 示例性的,预设时间段为 t ,可以将目标实例列表中存储的实例节点的时间信息,与当前系统时间之间的时间差在 $2t$ 内的实例节点,确定为当前有效的实例节点。

[0116] S702,确定当前的任务总数,任务总数包括待执行任务的个数以及正在执行任务的个数。

[0117] 查询数据库中存储的任务表,从该任务表中确定当前的任务总数,当前任务总数包括待执行任务的个数以及正在执行任务的个数。

[0118] 示例性的,实例节点表示为UEBA实例,如图8所示,各UEBA实例(即实例节点,图8中展示了3个)间隔预设时间段,将自身信息和对应的时间信息,更新到Redis数据库的目标实例列表中,进而通过查询该目标实例列表,就能够确定当前有效的实例节点个数。以及查询Mysql数据库中的任务表,以确定当前的任务总数。

[0119] S703,基于当前有效的实例节点个数、以及当前的任务总数,判断当前实例节点是否达到负载均衡上限条件。

[0120] 一个例子中,根据所确定的当前有效的实例节点个数以及当前的任务总数,可以计算到每个实例节点平均处理任务的个数,以此来确定实例节点是否达到负载均衡上限条件。

[0121] 本公开实施例中,根据所确定的当前有效的实例节点个数以及当前的任务总数,按照任务的数量实现负载均衡,使得任务能够均匀的分布到所有实例节点中执行。

[0122] 在一种可能的实施方式中,上述步骤S703基于当前有效的实例节点个数、以及当前的任务总数,判断当前实例节点是否达到负载均衡上限条件的实施过程,包括:

[0123] 将当前的任务总数和当前有效的实例节点个数的商,与1的和,确定为负载上限阈值;

[0124] 确定当前实例节点正在执行任务的个数;

[0125] 在当前实例节点正在执行任务的个数小于负载上限阈值的情况下,确定当前实例节点没有达到负载均衡上限条件;

[0126] 在当前实例节点正在执行任务的个数不小于负载上限阈值的情况下,确定当前实例节点达到负载均衡上限条件。

[0127] 示例性的,当前的任务总数表示为 m ,当前有效的实例节点个数表示为 n ,则负载上限阈值表示为: $m/n+1$ 。

[0128] 每一实例节点维护有自己当前正在执行任务的列表,通过查询当前实例节点的任务执行列表(current_execute_tasks),可以知晓当前实例节点正在执行任务的个数。

[0129] 在当前实例节点正在执行任务的个数小于负载上限阈值的情况下,确定当前实例节点没有达到负载均衡上限条件,否则,确定当前实例节点达到负载均衡上限条件。

[0130] 本公开实施例中,按照任务的数量,将任务均匀的分布在各实例节点中,使得任务能够均匀的分布到所有实例节点中执行。

[0131] 在一种可能的实施方式中,上述在得到第二可执行任务的任务执行结果之后,还

可以包括:对目标待执行任务进行解锁处理,并从待执行任务列表中删除目标待执行任务。

[0132] 本公开实施例中,任务在执行的过程中会被加锁,在执行完任务之后,对加锁的任务进行解锁,并从待执行任务列表中删除该执行完成的任务,以完整的结束该任务。

[0133] 实例节点在执行任务的过程中,可能会由于系统断电、死机、CPU或内存占用过高等因素导致实例节点直接停止运行,使得任务执行失败,该执行失败的任务因为还在执行中无法实现任务的解锁,进而一直处于加锁状态,无法被再次执行。

[0134] 为了处理这类由于系统断电、死机、CPU或内存占用过高等因素导致实例节点直接停止运行引起的执行失败的任务,本公开实施例中,通过实例节点利用上述图1所示的系统中的任务重试子模块实现执行失败任务的自动恢复和执行。

[0135] 在一种可能的实施方式中,如图9所示,在上述实施例的基础上还可以执行以下步骤,优选的,以下步骤可以在判断当前实例节点是否达到负载均衡上限条件之前执行。步骤包括:

[0136] S901,确定是否存在处于加锁状态的加锁任务。

[0137] 实例节点查询当前是否存在处于加锁状态的加锁任务,如果存在,则执行步骤S902-步骤S903以恢复该加锁任务的执行,如果不存在,则无需进行加锁任务的恢复。

[0138] S902,在存在处于加锁状态的加锁任务的情况下,针对每一加锁任务,确定该加锁任务是否被正常执行。

[0139] 在存在处于加锁状态的加锁任务的情况下,遍历处于加锁状态的每一加锁任务,确定每一加锁任务是否被正常执行,对没有被正常执行的加锁任务进行恢复。

[0140] 一个例子中,针对每一加锁任务,确定该加锁任务是否被实例节点执行,如果是,表示该加锁任务被正常执行,否则,表示该加锁任务没有被正常执行。

[0141] S903,在该加锁任务没有被正常执行的情况下,对该加锁任务进行解锁处理,并将解锁处理得到的解锁任务加入待执行任务列表中。

[0142] 本公开实施例中,确定是否存在处于加锁状态的加锁任务,在存在处于加锁状态的加锁任务的情况下,进一步针对每一加锁任务,确定该加锁任务是否被正常执行,在该加锁任务没有被正常执行的情况下,对该加锁任务进行解锁处理,并将解锁处理得到的解锁任务加入待执行任务列表中,实现了没有被正常执行任务的自动恢复和执行,提高了任务处理的可靠性。

[0143] 在一种可能的实施方式中,上述加锁任务中包含执行该加锁任务的实例节点的目标标识,如图10所示,确定该加锁任务是否被正常执行的实施过程,包括:

[0144] S1001,基于目标标识,判断执行该加锁任务的实例节点是否为当前实例节点。

[0145] 在上述对目标待执行任务进行加锁处理时,还可以将执行该加锁处理的实例节点的标识添加至加锁任务中,对目标待执行任务进行加锁处理的实例节点,即为执行该目标待执行任务加锁后得到的加锁任务的实例节点,加锁成功后的加锁任务中包含执行该加锁任务的实例节点的目标标识。示例性的,目标标识可以是实例节点的名称或IP地址等。

[0146] 示例性的,上述可以采用语句“update task set status='locked',execute_node='当前节点host'where status='unlocked'and id='1'”,对目标待执行任务进行加锁处理,该语句表示:标识为1的目标待执行任务,如果其状态为解锁状态,则将其状态设置为加锁状态,且设置执行该目标待执行任务的实例节点为当前节点host(实例节点的名

称)。

[0147] 将该加锁任务中包含的目标标识,与当前实例节点的标识进行匹配,在目标标识与当前实例节点的标识相同的情况下,表示执行该加锁任务的实例节点为当前实例节点,否则,表示执行该加锁任务的实例节点不是当前实例节点。

[0148] S1002,在执行该加锁任务的实例节点为当前实例节点的情况下,判断当前实例节点的任务执行列表中是否存在该加锁任务。

[0149] 每一实例节点维护有自己当前正在执行任务的列表,在执行该加锁任务的实例节点为当前实例节点的情况下,进一步查询当前实例节点的任务执行列表,判断当前实例节点的任务执行列表中是否存在该加锁任务。如果存在,表示该加锁任务在当前实例节点中被正常执行,如果不存在,表示该加锁任务没有被执行但是被加锁了,此时需要对该加锁任务进行解锁恢复。

[0150] S1003,在当前实例节点的任务执行列表中存在该加锁任务的情况下,确定该加锁任务被正常执行。

[0151] S1004,在当前实例节点的任务执行列表中不存在该加锁任务的情况下,确定该加锁任务没有被正常执行。

[0152] S1005,在执行该加锁任务的实例节点不是当前实例节点的情况下,判断执行该加锁任务的实例节点是否在目标实例列表中。

[0153] 在执行该加锁任务的实例节点不是当前实例节点的情况下,进一步查询数据库的目标实例列表中是否存在执行该加锁任务的实例节点。如果执行该加锁任务的实例节点在目标实例列表中,则表示该加锁任务在实例节点上被正常执行,如果执行该加锁任务的实例节点不在目标实例列表中,表示执行该加锁任务的实例节点已经不存在了,当前没有任何实例节点执行该加锁任务,此时需要对该加锁任务进行解锁恢复。

[0154] S1006,在执行该加锁任务的实例节点在目标实例列表中的情况下,确定该加锁任务被正常执行。

[0155] S1007,在执行该加锁任务的实例节点不在目标实例列表中的情况下,确定该加锁任务没有被正常执行。

[0156] 本公开实施例中,先判断执行加锁任务的实例节点是否为当前实例节点,在是的情况下进一步查询加锁任务是否在当前实例节点的任务执行列表中,以确定加锁任务是否被正常执行,在不是的情况下进一步查询执行加锁任务的实例节点是否在目标实例列表中,以确定加锁任务是否被正常执行,进而准确的判断处于加锁状态的任务是否被正常执行,以在处于加锁状态的任务没有被正常执行的情况下,对加锁任务进行解锁恢复,以提高任务处理的可靠性。

[0157] 示例性的,本公开实施例提供的一种任务处理方法包括以下步骤:

[0158] 步骤1),确定是否存在处于加锁状态的加锁任务;

[0159] 步骤2),在存在处于加锁状态的加锁任务的情况下,针对每一加锁任务,确定该加锁任务是否被正常执行;

[0160] 步骤3),在该加锁任务没有被正常执行的情况下,对该加锁任务进行解锁处理,并将解锁处理得到的解锁任务加入待执行任务列表中;

[0161] 步骤4),在该加锁任务被正常执行的情况下,执行步骤5);

[0162] 步骤5),判断当前实例节点是否达到负载均衡上限条件;

[0163] 步骤6),在当前实例节点没有达到负载均衡上限条件的情况下,从待执行任务列表中取出一个待执行任务作为目标待执行任务;

[0164] 步骤7),在当前实例节点达到负载均衡上限条件的情况下,不做处理。

[0165] 步骤8),对目标待执行任务进行加锁处理;

[0166] 步骤9),在对目标待执行任务加锁成功的情况下,获取目标待执行任务的配置信息以及任务执行图;其中,目标待执行任务的配置信息以及任务执行图为:预先针对目标待执行任务自定义设置的,任务执行图中包括:执行目标待执行任务的目标任务执行平台;

[0167] 步骤10),在对目标待执行任务加锁不成功的情况下,返回步骤6)执行从待执行任务列表中取出一个待执行任务作为目标待执行任务;

[0168] 步骤11),基于目标待执行任务的配置信息以及任务执行图,生成目标待执行任务对应的第一可执行任务;

[0169] 步骤12),将第一可执行任务转换为目标任务执行平台能够执行的第二可执行任务;其中,任务执行平台包括:本地计算引擎,Spark计算引擎以及Flink计算引擎;

[0170] 步骤13),将第二可执行任务推送给目标任务执行平台;

[0171] 步骤14),获取目标任务执行平台执行第二可执行任务得到的任务执行结果。

[0172] 本公开实施例中,针对实例节点在执行任务的过程中,可能会由于系统断电、死机、CPU或内存占用过高等因素导致实例节点直接停止运行,使得任务执行失败,该执行失败的任务因为还在执行中无法实现任务的解锁,进而一直处于加锁状态,无法被再次执行的情况,在确定加锁任务没有被正常执行时,对该加锁任务进行解锁处理,并将解锁处理得到的解锁任务加入待执行任务列表中,实现任务的自动恢复和执行,提高了任务处理的可靠性。

[0173] 采用无状态的实例节点自动进行任务的负载均衡,能够实现实例节点的任意扩缩容,在保证各实例节点能够负载所有任务的情况下,增加了任务处理的可伸缩性。实例节点通过对任务进行加锁抢占,保证了任意时刻一个待执行任务只能被一个实例节点执行,避免了任务被重复执行,进而提高了任务执行的实时性和准确性。

[0174] 预先针对不同的目标待执行任务,自定义设置该目标待执行任务的配置信息以及任务执行图,进而基于该自定义设置的配置信息以及任务执行图,生成目标待执行任务对应的第一可执行任务,实现了任务的自定义配置以及生成,扩展了任务处理的业务场景。进一步的,将第一可执行任务转换为目标任务执行平台能够执行的第二可执行任务,将转换后的第二可执行任务推送给目标任务执行平台进行执行,保证了任务执行过程中任务计算能力的灵活可伸缩性,使得任务的处理能够适应于不同企业的内部执行环境。

[0175] 同时支撑任务在本地计算引擎(本地物理机/虚拟机),Spark计算引擎以及Flink计算引擎中执行,保证了任务执行过程中任务计算能力的灵活可伸缩性,使得任务的处理能够适应于不同企业的内部执行环境。

[0176] 本公开实施例还提供了一种任务处理系统,参见图11,该系统包括:

[0177] 信息获取模块1101,获取目标待执行任务的配置信息以及任务执行图;其中,目标待执行任务的配置信息以及任务执行图为预先针对目标待执行任务自定义设置的;任务执行图中包括:执行目标待执行任务的目标任务执行平台;

[0178] 任务生成模块1102,用于基于目标待执行任务的配置信息以及任务执行图,生成目标待执行任务对应的第一可执行任务;

[0179] 任务转换模块1103,用于将第一可执行任务转换为目标任务执行平台能够执行的第二可执行任务;

[0180] 任务分配模块1104,用于将第二可执行任务推送给目标任务执行平台;

[0181] 结果获取模块1105,用于获取目标任务执行平台执行第二可执行任务得到的任务执行结果,第二可执行任务的任务执行结果,表示目标待执行任务的任务执行结果。

[0182] 本公开实施例中,预先针对不同的目标待执行任务,自定义设置该目标待执行任务的配置信息以及任务执行图,进而基于该自定义设置的配置信息以及任务执行图,生成目标待执行任务对应的第一可执行任务,实现了任务的自定义配置以及生成,扩展了任务处理的业务场景。进一步的,将第一可执行任务转换为目标任务执行平台能够执行的第二可执行任务,将转换后的第二可执行任务推送给目标任务执行平台进行执行,保证了任务执行过程中任务计算能力的灵活可伸缩性,使得任务的处理能够适应于不同企业的内部执行环境。

[0183] 在一种可能的实施方式中,上述系统还包括:

[0184] 负载均衡模块,用于判断当前实例节点是否达到负载均衡上限条件;

[0185] 任务确定模块,用于在负载均衡模块判断出当前实例节点没有达到负载均衡上限条件的情况下,从待执行任务列表中取出一个待执行任务作为目标待执行任务;

[0186] 任务加锁模块,用于对目标待执行任务进行加锁处理,并在对目标待执行任务加锁成功的情况下,触发所述信息获取模块执行获取目标待执行任务的配置信息以及任务执行图。

[0187] 在一种可能的实施方式中,上述负载均衡模块,包括:

[0188] 第一确定子模块,用于基于目标实例列表,确定当前有效的实例节点个数;目标实例列表为:各实例节点间隔预设时间段将自身信息和对应的时间信息,更新至实例列表中得到的;当前有效的实例节点表示:目标实例列表中该实例节点的时间信息,与当前系统时间之间的时间差在预设时间范围内;

[0189] 第二确定子模块,用于确定当前的任务总数,任务总数包括待执行任务的个数以及正在执行任务的个数;

[0190] 负载均衡子模块,用于基于当前有效的实例节点个数、以及当前的任务总数,判断当前实例节点是否达到负载均衡上限条件。

[0191] 在一种可能的实施方式中,上述负载均衡子模块,具体用于:

[0192] 将当前的任务总数和当前有效的实例节点个数的商,与1的和,确定为负载上限阈值;

[0193] 确定当前实例节点正在执行任务的个数;

[0194] 在当前实例节点正在执行任务的个数小于负载上限阈值的情况下,确定当前实例节点没有达到负载均衡上限条件;

[0195] 在当前实例节点正在执行任务的个数不小于负载上限阈值的情况下,确定当前实例节点达到负载均衡上限条件。

[0196] 在一种可能的实施方式中,上述任务加锁模块,具体用于:

[0197] 在目标待执行任务的状态处于解锁状态的情况下,将目标待执行任务的状态更新为加锁状态。

[0198] 在一种可能的实施方式中,上述系统还包括:

[0199] 任务获取模块,用于在任务加锁模块对目标待执行任务加锁不成功的情况下,触发任务确定模块执行从待执行任务列表中取出一个待执行任务作为目标待执行任务。

[0200] 在一种可能的实施方式中,上述目标待执行任务的任务执行图中还包括:执行目标待执行任务的各任务执行节点对应的信息、以及各任务执行节点的执行顺序和连接关系信息;目标待执行任务的配置信息包括:各任务执行节点在执行目标待执行任务时用到的属性信息;目标任务执行平台包括:本地计算引擎,Spark计算引擎以及Flink计算引擎;

[0201] 上述任务生成模块1102,具体用于:基于各任务执行节点对应的信息、各任务执行节点的执行顺序和连接关系信息、以及各任务执行节点在执行目标待执行任务时用到的属性信息,实例化生成目标待执行任务对应的第一可执行任务。

[0202] 在一种可能的实施方式中,上述系统还包括:

[0203] 任务删除模块,用于在得到第二可执行任务的任务执行结果之后,对目标待执行任务进行解锁处理,并从待执行任务列表中删除目标待执行任务。

[0204] 在一种可能的实施方式中,上述系统还包括:

[0205] 任务状态确定模块,用于确定是否存在处于加锁状态的加锁任务;

[0206] 任务执行情况确定模块,用于在任务状态确定模块确定存在处于加锁状态的加锁任务的情况下,针对每一加锁任务,确定该加锁任务是否被正常执行;

[0207] 任务恢复模块,用于在任务执行情况确定模块确定该加锁任务没有被正常执行的情况下,对该加锁任务进行解锁处理,并将解锁处理得到的解锁任务加入待执行任务列表中。

[0208] 在一种可能的实施方式中,上述加锁任务中包含执行该加锁任务的实例节点的目标标识;上述确定该加锁任务是否被正常执行,包括:

[0209] 基于目标标识,判断执行该加锁任务的实例节点是否为当前实例节点;

[0210] 在执行该加锁任务的实例节点为当前实例节点的情况下,判断当前实例节点的任务执行列表中是否存在该加锁任务;

[0211] 在当前实例节点的任务执行列表中存在该加锁任务的情况下,确定该加锁任务被正常执行;

[0212] 在当前实例节点的任务执行列表中不存在该加锁任务的情况下,确定该加锁任务没有被正常执行;

[0213] 在执行该加锁任务的实例节点不是当前实例节点的情况下,判断执行该加锁任务的实例节点是否在目标实例列表中;

[0214] 在执行该加锁任务的实例节点在目标实例列表中的情况下,确定该加锁任务被正常执行;

[0215] 在执行该加锁任务的实例节点不在目标实例列表中的情况下,确定该加锁任务没有被正常执行。

[0216] 本公开的技术方案中,所涉及的用户个人信息的收集、存储、使用、加工、传输、提供和公开等处理,均符合相关法律法规的规定,且不违背公序良俗。需要说明的是,本实施

例中的人头模型并不是针对某一特定用户的人头模型,并不能反映出某一特定用户的个人信息。需要说明的是,本实施例中的二维人脸图像来自于公开数据集。

[0217] 根据本公开的实施例,本公开还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。

[0218] 本公开提供的一种电子设备,包括:

[0219] 至少一个处理器;以及

[0220] 与至少一个处理器通信连接的存储器;其中,

[0221] 存储器存储有可被至少一个处理器执行的指令,指令被至少一个处理器执行,以使至少一个处理器能够执行本公开中任一项的方法。

[0222] 本公开提供的一种存储有计算机指令的非瞬时计算机可读存储介质,其中,计算机指令用于使计算机执行本公开中任一项的方法。

[0223] 本公开提供的一种计算机程序产品,包括计算机程序,计算机程序在被处理器执行时实现本公开中任一项的方法。

[0224] 图12示出了可以用来实施本公开的实施例的示例电子设备1200的示意性框图。电子设备旨在表示各种形式的数字计算机,诸如,膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置,诸如,个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例,并且不意在限制本文中描述的和/或者要求的本公开的实现。

[0225] 如图12所示,设备1200包括计算单元1201,其可以根据存储在只读存储器(ROM)1202中的计算机程序或者从存储单元1208加载到随机访问存储器(RAM)1203中的计算机程序,来执行各种适当的动作和处理。在RAM 1203中,还可存储设备1200操作所需的各种程序和数据。计算单元1201、ROM 1202以及RAM 1203通过总线1204彼此相连。输入/输出(I/O)接口1205也连接至总线1204。

[0226] 设备1200中的多个部件连接至I/O接口1205,包括:输入单元1206,例如键盘、鼠标等;输出单元1207,例如各种类型的显示器、扬声器等;存储单元1208,例如磁盘、光盘等;以及通信单元1209,例如网卡、调制解调器、无线通信收发机等。通信单元1209允许设备1200通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0227] 计算单元1201可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元1201的一些示例包括但不限于中央处理单元(CPU)、图形处理单元(GPU)、各种专用的人工智能(AI)计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器(DSP)、以及任何适当的处理器、控制器、微控制器等。计算单元1201执行上文所描述的各个方法和处理,例如任务处理方法。例如,在一些实施例中,任务处理方法可被实现为计算机软件程序,其被有形地包含于机器可读介质,例如存储单元1208。在一些实施例中,计算机程序的部分或者全部可以经由ROM1202和/或通信单元1209而被载入和/或安装到设备1200上。当计算机程序加载到RAM 1203并由计算单元1201执行时,可以执行上文描述的任务处理方法一个或多个步骤。备选地,在其他实施例中,计算单元1201可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行任务处理方法。

[0228] 本文中以上描述的系统和技术各种实施方式可以在数字电子电路系统、集成电

路系统、场可编程门阵列 (FPGA)、专用集成电路 (ASIC)、专用标准产品 (ASSP)、芯片上系统的系统 (SOC)、复杂可编程逻辑设备 (CPLD)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括：实施在一个或者多个计算机程序中，该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释，该可编程处理器可以是专用或者通用可编程处理器，可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令，并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0229] 用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器，使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行，作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0230] 在本公开的上下文中，机器可读介质可以是有形的介质，其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备，或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器 (RAM)、只读存储器 (ROM)、可擦除可编程只读存储器 (EPROM 或快闪存储器)、光纤、便捷式紧凑盘只读存储器 (CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0231] 为了提供与用户的交互，可以在计算机上实施此处描述的系统和技术，该计算机具有：用于向用户显示信息的显示装置 (例如，CRT (阴极射线管) 或者 LCD (液晶显示器) 监视器)；以及键盘和指向装置 (例如，鼠标或者轨迹球)，用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互；例如，提供给用户的反馈可以是任何形式的传感反馈 (例如，视觉反馈、听觉反馈、或者触觉反馈)；并且可以用任何形式 (包括声输入、语音输入或者、触觉输入) 来接收来自用户的输入。

[0232] 可以将此处描述的系统和技术实施在包括后台部件的计算系统 (例如，作为数据服务器)、或者包括中间件部件的计算系统 (例如，应用服务器)、或者包括前端部件的计算系统 (例如，具有图形用户界面或者网络浏览器的用户计算机，用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信 (例如，通信网络) 来将系统的部件相互连接。通信网络的示例包括：局域网 (LAN)、广域网 (WAN) 和互联网。

[0233] 计算机系统可以包括客户端和服务端。客户端和服务端一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务端的关系。服务器可以是云服务器，也可以为分布式系统的服务器，或者是结合了区块链的服务器。

[0234] 应该理解，可以使用上面所示的各种形式的流程，重新排序、增加或删除步骤。例如，本公开中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行，只

要能够实现本公开公开的技术方案所期望的结果,本文在此不进行限制。

[0235] 上述具体实施方式,并不构成对本公开保护范围的限制。本领域技术人员应该明白的是,根据设计要求和因素,可以进行各种修改、组合、子组合和替代。任何在本公开的精神和原则之内所作的修改、等同替换和改进等,均应包含在本公开保护范围之内。

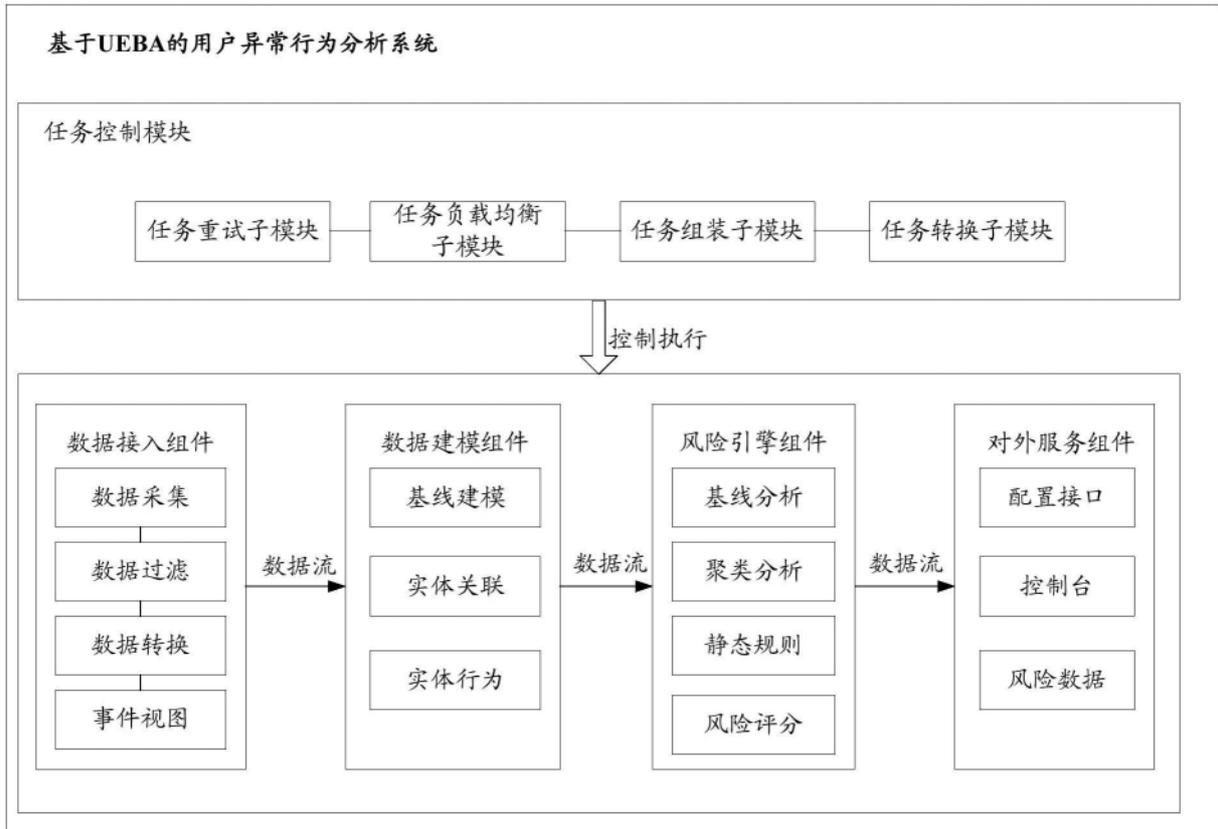


图1

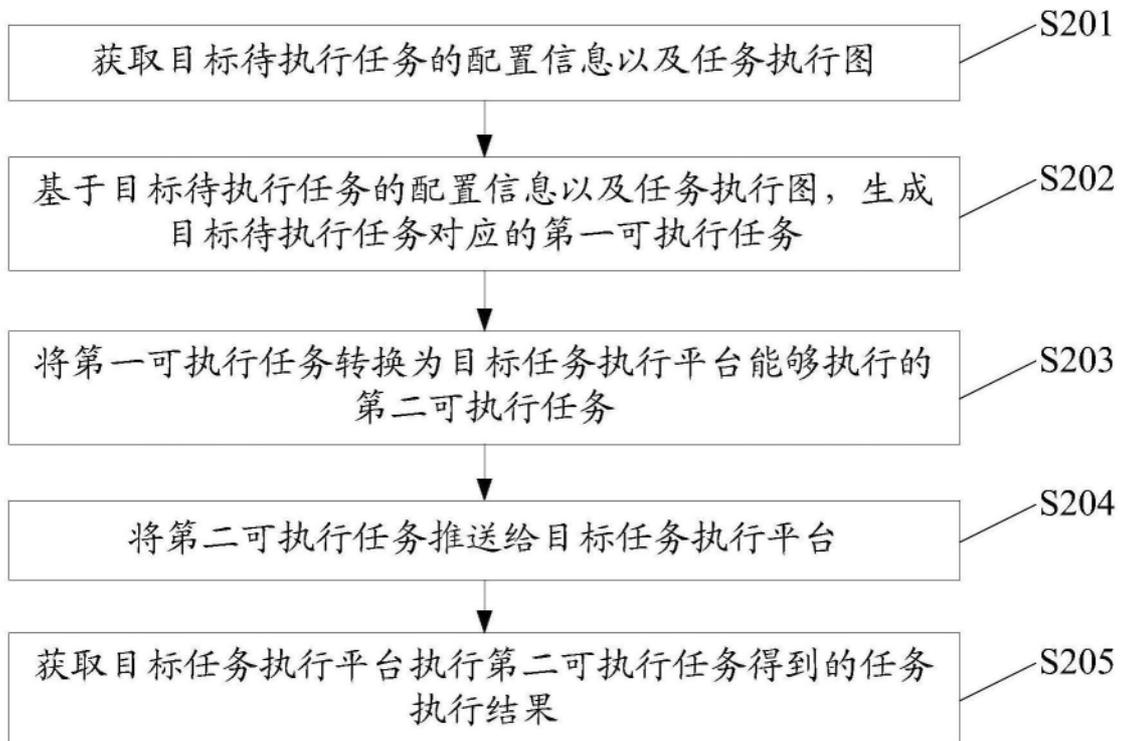


图2

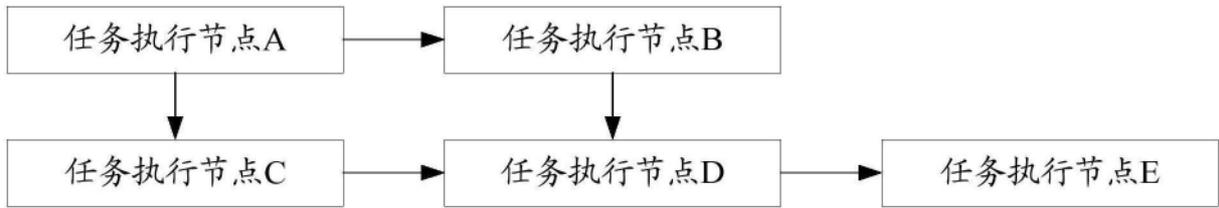


图3

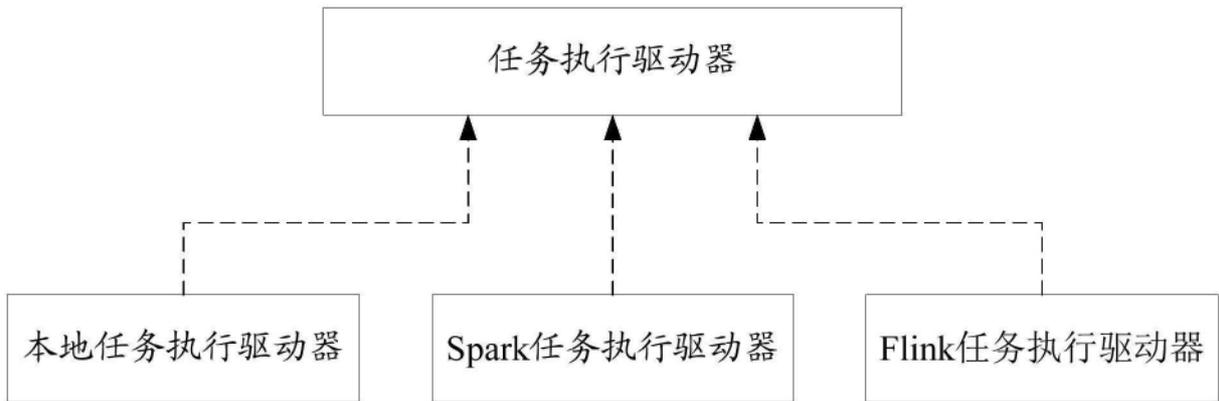


图4

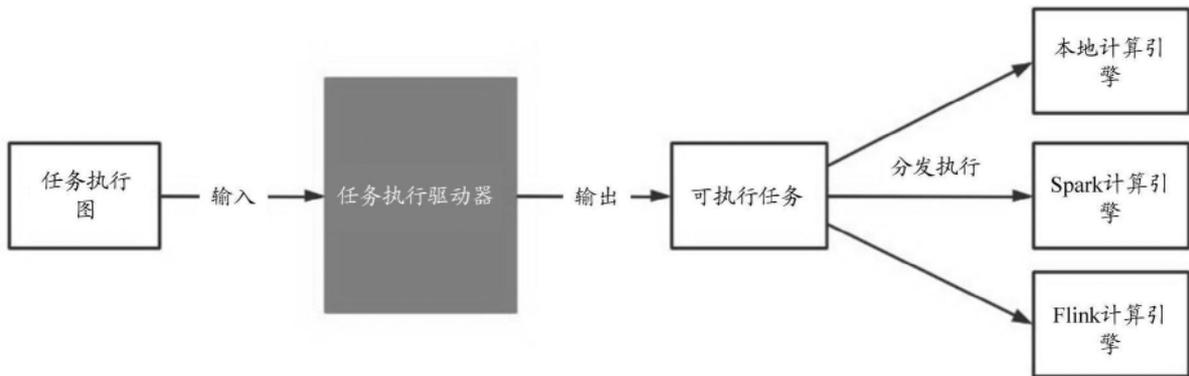


图5

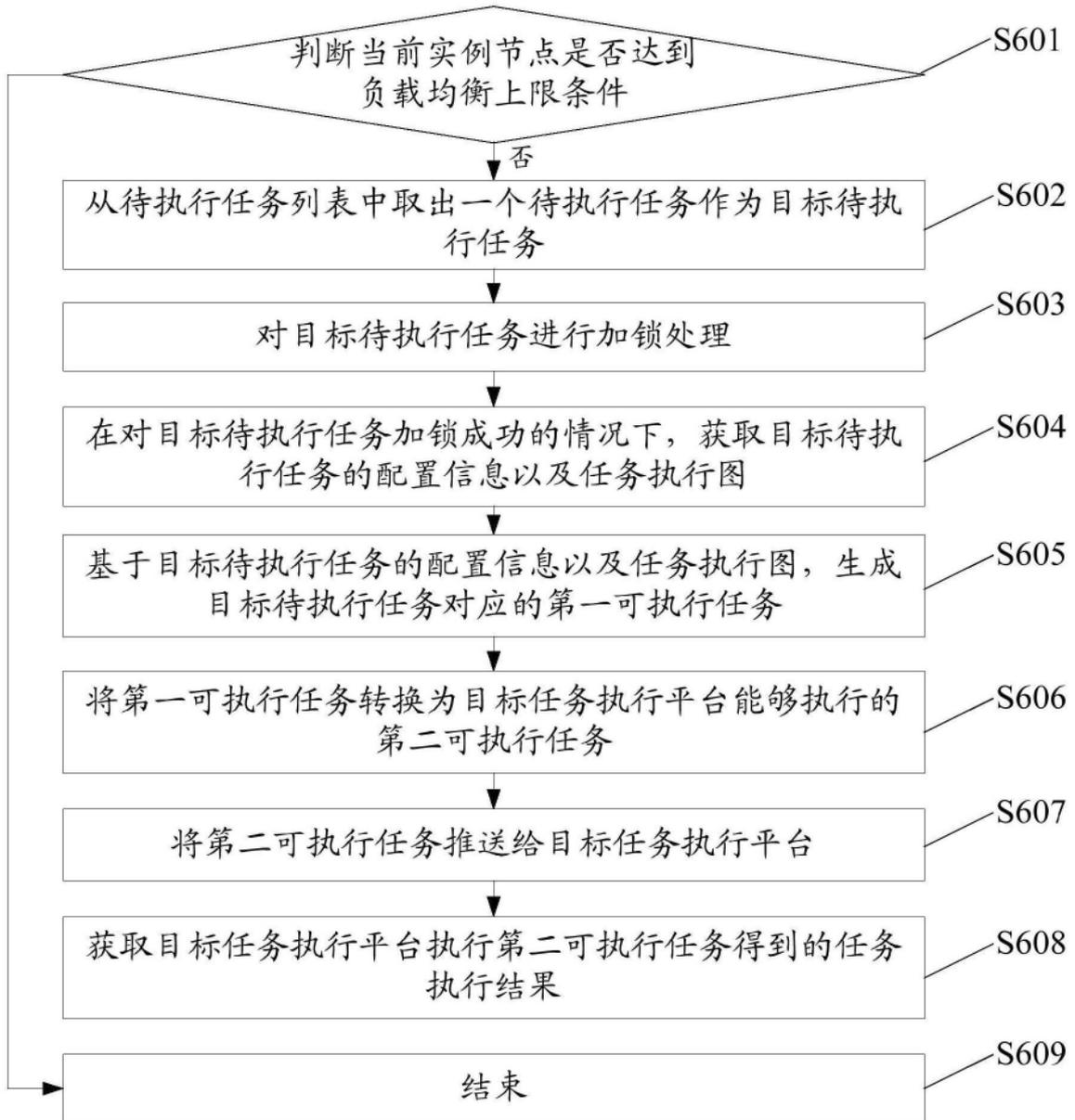


图6

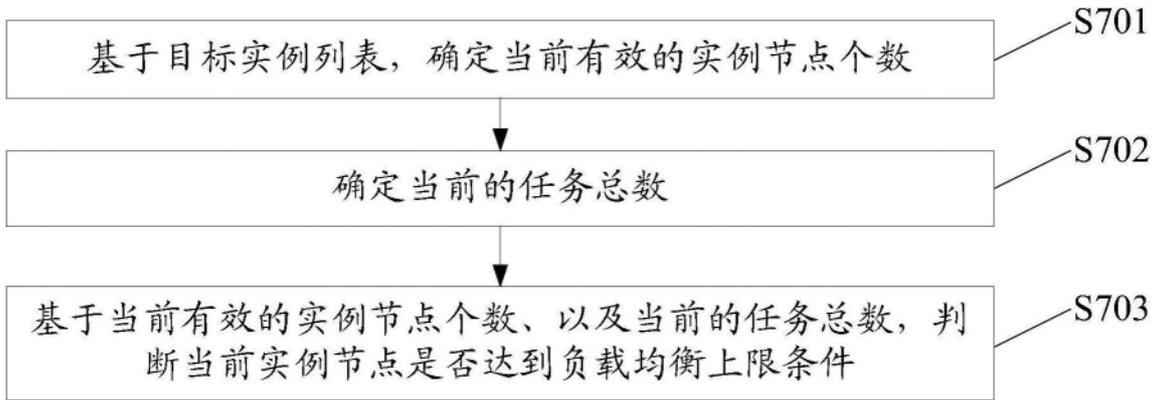


图7

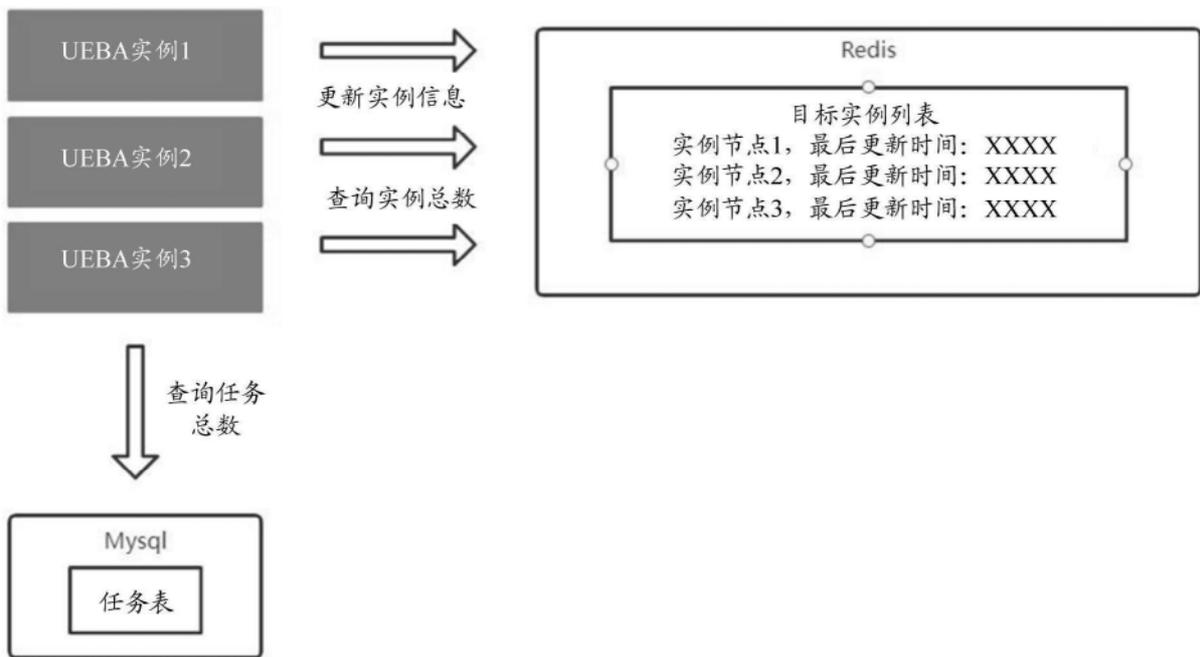


图8

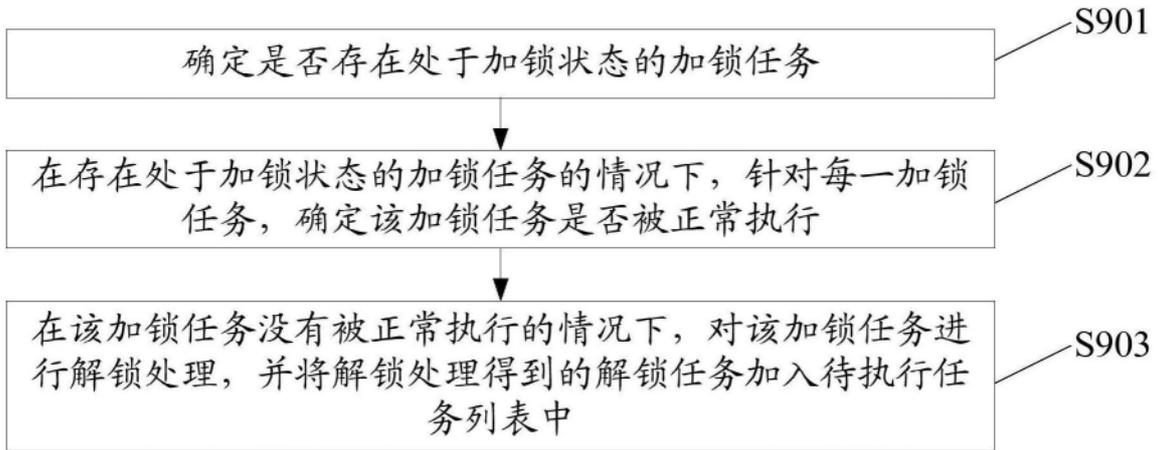


图9

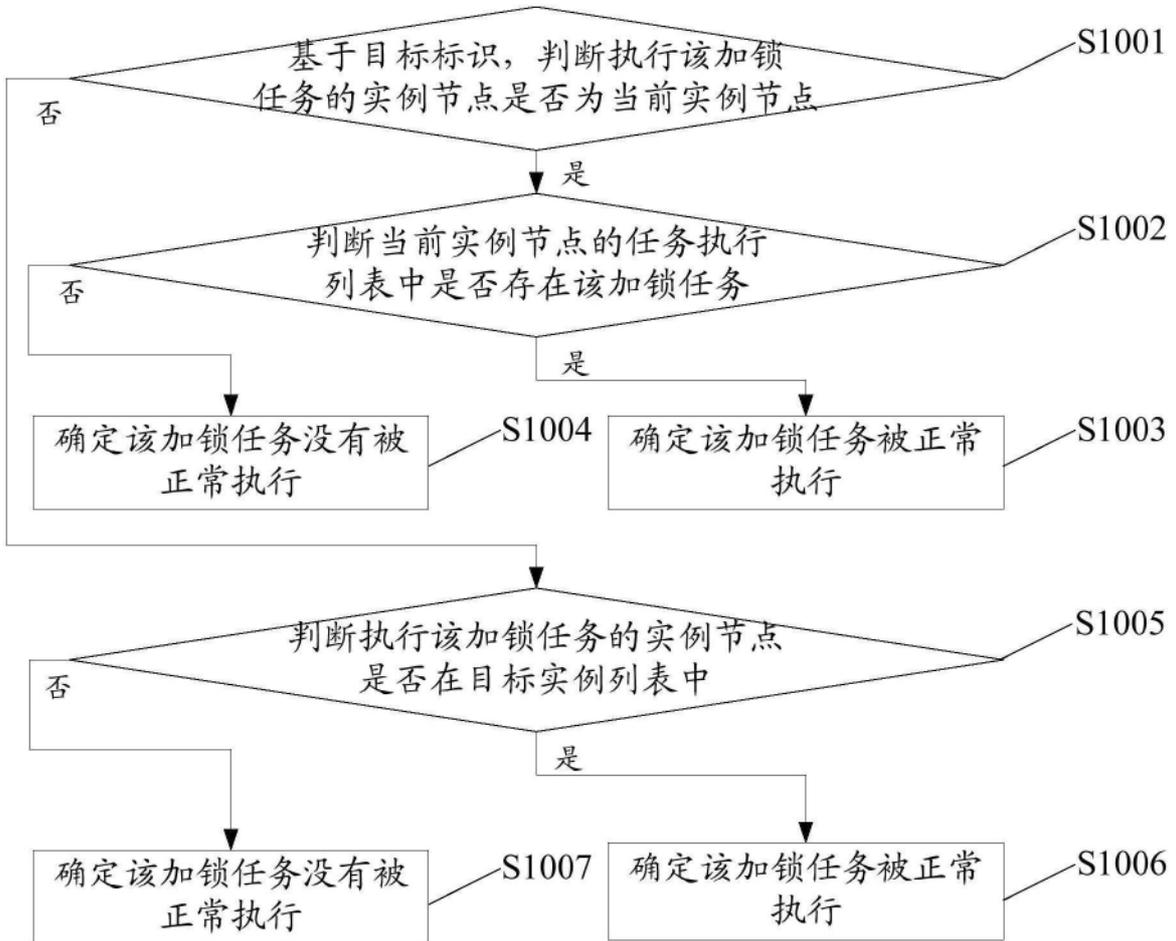


图10

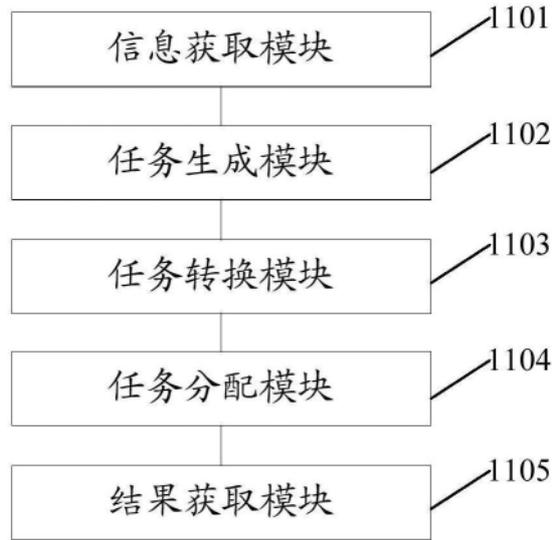


图11

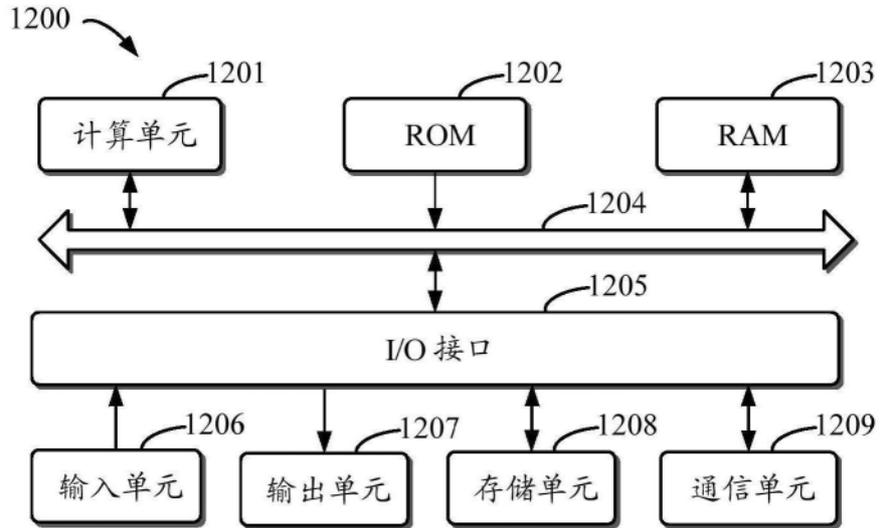


图12