



(12) 发明专利申请

(10) 申请公布号 CN 104520815 A

(43) 申请公布日 2015. 04. 15

(21) 申请号 201480001794. 1

(51) Int. Cl.

(22) 申请日 2014. 03. 17

G06F 9/50(2006. 01)

(85) PCT国际申请进入国家阶段日
2015. 01. 13

(86) PCT国际申请的申请数据
PCT/CN2014/073495 2014. 03. 17

(71) 申请人 华为技术有限公司
地址 518129 广东省深圳市龙岗区坂田华为
总部办公楼

(72) 发明人 王光明 周红兵

(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 熊永强

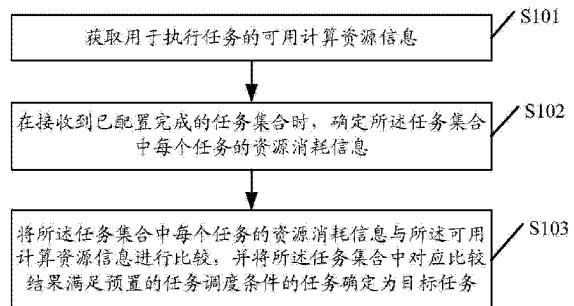
权利要求书5页 说明书16页 附图7页

(54) 发明名称

一种任务调度的方法、装置及设备

(57) 摘要

本发明实施例公开了一种任务调度的方法、装置及设备,其中,所述方法包括:获取用于执行任务的可用计算资源信息;在接收到已配置完成的任务集合时,确定所述任务集合中每个任务的资源消耗信息;将所述任务集合中每个任务的资源消耗信息与所述可用计算资源信息进行比较,并将所述任务集合中对应比较结果满足预置的任务调度条件的任务确定为目标任务。另外,生成的任务调度图包括所述任务集合中各任务、各任务之间的依赖关系、资源消耗信息、目标任务的执行进度等信息。采用本发明,可优化任务调度,使得任务调度优化更加直观,从而提高数据仓库的处理性能。



1. 一种任务调度的方法,其特征在于,包括:
获取用于执行任务的可用计算资源信息;
在接收到已配置完成的任务集合时,确定所述任务集合中每个任务的资源消耗信息;
将所述任务集合中每个任务的资源消耗信息与所述可用计算资源信息进行比较,并将所述任务集合中对应比较结果满足预置的任务调度条件的任务确定为目标任务。

2. 如权利要求 1 所述的方法,其特征在于,还包括:
根据所述任务集合中每个任务的资源依赖关系和 / 或时间依赖关系,生成任务调度图;

其中,所述任务调度图包括标识所述任务集合中每个任务的资源标记和标识任务间任务依赖关系和 / 或时间依赖关系的关系标记。

3. 如权利要求 2 所述的方法,其特征在于,所述任务调度图还包括所述任务集合中每个任务的资源消耗信息标记和从所述任务集合中确定出的目标任务的执行进度;

其中,所述资源消耗信息标记用于标识所述任务集合中每个任务的资源消耗信息;所述目标任务的执行进度是根据目标任务的资源消耗信息和当前检测到的所述目标任务在执行过程中的已消耗资源信息获得的。

4. 如权利要求 1 至 3 任一项所述的方法,其特征在于,在接收到已配置完成的任务集合时,所述确定出所述任务集合中每个任务的资源消耗信息,包括:

从任务资源消耗映射表中查找所述任务集合中每个任务在上一调度周期的实际资源消耗信息;

根据上一调度周期的实际资源消耗信息估算所述任务集合中每个任务在当前调度周期的资源消耗信息;

其中,所述任务资源消耗映射表为预先设置的记录有所述任务集合中各任务在上一调度周期执行后的实际资源消耗信息。

5. 如权利要求 1 至 4 任一项所述的方法,其特征在于,所述将所述每个任务的资源消耗信息与所述可用计算资源信息进行比较,并将所述任务集合中对应比较结果满足预置的任务调度条件的任务确定为目标任务,包括:

将所述任务集合中每个任务的所述资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较;

当比较结果为任务的所述资源消耗量小于所述可用资源量时,则记录该任务对应的比较结果满足预置的任务调度条件;

将所述任务集合中记录比较结果满足所述任务调度条件的至少一个任务确定为当前调度周期的目标任务。

6. 如权利要求 1 至 4 任一项所述的方法,其特征在于,所述将所述每个任务的资源消耗信息与所述可用计算资源信息进行比较,并将所述任务集合中对应比较结果满足预置的任务调度条件的任务确定为目标任务,包括:

根据所述任务集合中每个任务的资源消耗信息指示的任务集群类型和所述可用计算资源信息指示的可用集群类型,从所述任务集合中确定出所述任务集群类型与所述可用集群类型匹配的任务子集合;

将所述任务子集合中任务的资源消耗信息指示的资源消耗量与所述可用计算资源信

息指示的可用资源量进行比较；

当比较结果为任务的所述资源消耗量小于所述可用资源量时，则记录该任务对应的比较结果满足预置的任务调度条件；

将所述任务子集中记录比较结果满足所述任务调度条件的至少一个任务作为当前调度周期的目标任务。

7. 如权利要求 1 至 6 任一项所述的方法，其特征在于，所述将所述每个任务的资源消耗信息与所述可用计算资源信息进行比较，并将所述任务集中对应比较结果满足预置的任务调度条件的任务确定为目标任务之后，还包括：

控制所述任务集中确定的目标任务执行；

当所述目标任务执行完成后，获取所述目标任务在当前调度周期的实际资源消耗信息；

根据所述当前调度周期的实际资源消耗信息，更新预置的任务资源消耗映射表中记录的所述目标任务的实际资源消耗信息。

8. 如权利要求 3 所述的方法，其特征在于，还包括：

当所述任务集中每个任务的所述资源消耗信息指示的资源消耗量大于所述可用计算资源信息所指示的可用资源量时，发出提示消息以提示优化该任务；

或者，所述任务调度图中生成的任务执行进度小于预设阈值时，发出提示消息以提示优化该任务。

9. 一种任务调度的装置，其特征在于，包括：

可用计算资源信息获取模块，用于获取用于执行任务的可用计算资源信息；

任务资源消耗信息确定模块，用于在接收到已配置完成的任务集合时，确定所述任务集合中每个任务的资源消耗信息；

比较确定模块，用于将所述任务资源消耗信息确定模块确定的每个任务的资源消耗信息与所述可用计算资源信息获取模块获得的可用计算资源信息进行比较，并根据所述比较结果，将所述任务集中对应比较结果满足预置的任务调度条件的任务确定为目标任务。

10. 如权利要求 9 所述的装置，其特征在于，还包括：

任务调度图生成模块，用于根据所述任务集合中每个任务的任务依赖关系和 / 或时间依赖关系，生成任务调度图；

其中，所述任务调度图包括标识所述任务集合中每个任务的任务标记和标识任务间任务依赖关系和 / 或时间依赖关系的关系标记。

11. 如权利要求 10 所述的装置，其特征在于，所述任务调度图生成模块生成的任务调度图还包括所述任务集合中每个任务的资源消耗信息标记和从所述任务集合中确定出的目标任务的执行进度；

其中，所述资源消耗信息标记用于标识所述任务集合中每个任务的资源消耗信息；所述目标任务的执行进度是根据目标任务的资源消耗信息和当前检测到的所述目标任务在执行过程中的已消耗资源信息获得的。

12. 如权利要求 9 至 11 任一项所述的装置，其特征在于，所述任务资源消耗信息确定模块，包括：

查找单元，用于在接收到已配置完成的任务集合时，从任务资源消耗映射表中查找所

述任务集合中每个任务在上一调度周期的实际资源消耗信息；

估算单元,用于根据所述查找单元查找的实际资源消耗信息估算所述任务集合中每个任务在当前调度周期的资源消耗信息；

其中,所述任务资源消耗映射表为预先设置的记录有所述任务集合中各任务在上一调度周期执行后的实际资源消耗信息。

13. 如权利要求 9 至 12 任一项所述的装置,其特征在于,所述比较确定模块,包括:

第一比较单元,用于将所述任务集合每个任务的所述资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较；

第一记录单元,用于当所述第一比较单元的比较结果为任务的所述资源消耗量小于所述可用资源量时,则记录该任务对应的比较结果匹配预置的任务调度条件；

第一确定单元,用于将所述任务集合中所述第一记录单元记录比较结果满足所述任务调度条件的至少一个任务确定为当前调度周期的目标任务。

14. 如权利要求 9 至 12 任一项所述的装置,其特征在于,所述比较确定模块,包括:

类型匹配单元,用于将所述任务集合中每个任务的资源消耗信息指示的任务集群类型和所述可用计算资源信息指示的可用集群类型进行比较,从所述任务集合中确定出所述任务集群类型与所述可用集群类型匹配的任务子集合；

第二比较单元,用于将所述类型匹配单元确定的任务子集合中任务的资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较；

第二记录单元,用于当所述第二比较单元的比较结果为任务的所述资源消耗量小于所述可用资源量时,则记录该任务对应的比较结果满足预置的任务调度条件；

第二确定单元,用于将所述任务子集合中所述第二记录单元记录的比较结果满足所述任务调度条件的至少一个任务作为当前调度周期的目标任务。

15. 如权利要求 9 至 14 任一项所述的装置,其特征在于,还包括:

控制模块,用于控制所述任务集合中确定的目标任务执行；

实际资源消耗信息获取模块,用于当所述控制模块控制所述目标任务执行完成后,获取所述目标任务在当前调度周期的实际资源消耗信息；

更新模块,用于根据所述实际资源信息获取模块获取的当前调度周期所述目标任务的实际资源消耗信息,更新预置的任务资源消耗映射表中记录的所述目标任务的实际资源消耗信息。

16. 如权利要求 11 所述的装置,其特征在于,还包括:

提示消息发出模块,用于当所述任务集合中每个任务的所述资源消耗信息指示的资源消耗量大于所述可用计算资源信息指示的可用资源量时,或者所述任务调度图中生成的任务的执行进度小于预设阈值时,发出提示消息以修改该任务的依赖关系。

17. 一种计算机存储介质,其特征在于,

所述计算机存储介质可存储有程序,该程序执行时包括如权利要求 1 至 8 任一项所述的步骤。

18. 一种任务调度设备,其特征在于,包括:处理器、通信接口,

所述通信接口,用于接收数据仓库中已配置的任务集合和可用计算资源信息；

所述处理器,用于获取用于执行任务的可用计算资源信息;在接收到已配置完成的任

务集合时,确定所述任务集合中每个任务的资源消耗信息;将所述任务集合中每个任务的资源消耗信息与所述可用计算资源信息进行比较,并将所述任务集合中对应比较结果满足预置的任务调度条件的任务确定为目标任务。

19. 如权利要求 18 所述的设备,其特征在于,所述处理器还用于:

根据所述任务集合中每个任务的任务依赖关系和/或时间依赖关系,生成任务调度图;

其中,所述任务调度图包括标识所述任务集合中每个任务的任务标记和标识任务间任务依赖关系和/或时间依赖关系的关系标记。

20. 如权利要求 19 所述的设备,其特征在于,所述处理器生成的任务调度图还包括所述任务集合中每个任务的资源消耗信息标记和从所述任务集合中确定出的目标任务的执行进度;

其中,所述资源消耗信息标记用于标识所述任务集合中每个任务的资源消耗信息;所述目标任务的执行进度是根据目标任务的资源消耗信息和当前检测到的所述目标任务在执行过程中的已消耗资源信息获得的。

21. 如权利要求 18 至 20 任一项所述的设备,其特征在于,所述处理器具体用于:

从任务资源消耗映射表中查找所述任务集合中每个任务在上一调度周期的实际资源消耗信息;

根据所述实际资源消耗信息估算所述任务集合中每个任务在当前调度周期的资源消耗信息;

其中,所述任务资源消耗映射表为预先设置的记录有所述任务集合中各任务在上一调度周期执行后的实际资源消耗信息。

22. 如权利要求 18 至 21 任一项所述的设备,其特征在于,所述处理器具体用于:

将所述任务集合每个任务的所述资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较;

当比较结果为任务的所述资源消耗量小于所述可用资源量时,则记录该任务对应的比较结果满足预置的任务调度条件;

将所述任务集合中记录比较结果满足所述任务调度条件的至少一个任务确定为当前调度周期的目标任务。

23. 如权利要求 18 至 21 任一项所述的设备,其特征在于,所述处理器具体用于:

根据所述任务集合中每个任务的资源消耗信息指示的任务集群类型和所述可用计算资源信息指示的可用集群类型,将所述任务集合中确定出所述任务集群类型与所述可用集群类型匹配的任务子集合;

将所述任务子集合中任务的资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较;

当比较结果为任务的所述资源消耗量小于所述可用资源量时,则记录该任务对应的比较结果满足预置的任务调度条件;

将所述任务子集合中记录比较结果满足所述任务调度条件的至少一个任务作为当前调度周期的目标任务。

24. 如权利要求 18 至 23 任一项所述的设备,其特征在于,所述处理器还用于:

控制所述任务集合中确定的目标任务执行；

当所述目标任务执行完成后，获取所述目标任务在当前调度周期的实际资源消耗信息；

根据所述当前调度周期的实际资源消耗信息，更新预置的任务资源消耗映射表中记录的所述目标任务的实际资源消耗信息。

25. 如权利要求 20 所述的设备，其特征在于，所述处理器还用于：

当所述任务集合中每个任务的所述资源消耗信息指示的资源消耗量大于所述可用计算资源信息指示的可用资源量时，发出提示消息以提示优化该任务；

或者所述任务调度图中显示的任务的执行进度小于预设阈值时，发出提示消息以提示优化该任务。

26. 如权利要求 18 至 25 任一项所述的设备，其特征在于，还包括：显示器，

所述显示器，用于显示所述处理器执行的结果。

一种任务调度的方法、装置及设备

技术领域

[0001] 本发明涉及计算机技术领域,尤其涉及一种任务调度的方法、装置及设备。

背景技术

[0002] 数据仓库用于从数据库中获取信息,其任务对应一个或多个数据处理步骤,各任务之间可能存在单向依赖、独立或时间约束等特点。任务调度系统是数据仓库的重要组成部分,用于调度数据仓库的计算资源完成各个任务处理。

[0003] 现有的数据仓库任务调度的过程为:任务调度系统根据数据开发人员在任务部署系统中配置的任务集合,以及任务集合中各任务依赖关系等,驱动任务执行;每个任务执行完成后,再依次或者根据任务间的依赖关系执行下一个任务,并不会考虑下一个任务所需资源量的大小,导致任务执行过程资源利用率低,处理性能低。

发明内容

[0004] 本发明实施例提供一种任务调度的方法、装置及设备,可提高数据处理性能和资源利用率。

[0005] 本发明第一方面提供了一种任务调度的方法,包括:

[0006] 获取用于执行任务的可用计算资源信息;

[0007] 在接收到已配置完成的任务集合时,确定所述任务集合中每个任务的资源消耗信息;

[0008] 将所述任务集合中每个任务的资源消耗信息与所述可用计算资源信息进行比较,并将所述任务集合中对应比较结果满足预置的任务调度条件的任务确定为目标任务。

[0009] 在第一种可能的实现方式中,所述方法还包括:

[0010] 根据所述任务集合中每个任务的任务依赖关系和/或时间依赖关系,生成任务调度图;

[0011] 其中,所述任务调度图包括标识所述任务集合中每个任务的任务标记和标识任务间任务依赖关系和/或时间依赖关系的关系标记。

[0012] 结合第一方面的第一种可能的实现方式,在第二种可能的实现方式中,所述任务调度图还包括所述任务集合中每个任务的资源消耗信息标记和从所述任务集合中确定出的目标任务的执行进度;

[0013] 其中,所述资源消耗信息标记用于标识所述任务集合中每个任务的资源消耗信息;所述目标任务的执行进度是根据目标任务的资源消耗信息和当前检测到的所述目标任务在执行过程中的已消耗资源信息获得的。

[0014] 结合第一方面,或第一方面的第一种可能的实现方式,或第一方面的第二种可能的实现方式,在第三种可能的实现方式中,在接收到已配置完成的任务集合时,所述确定出所述任务集合中每个任务的资源消耗信息,包括:

[0015] 从任务资源消耗映射表中查找所述任务集合中每个任务在上一调度周期的实际

资源消耗信息；

[0016] 根据上一调度周期的实际资源消耗信息估算所述任务集合中每个任务在当前调度周期的资源消耗信息；

[0017] 其中,所述任务资源消耗映射表为预先设置的记录有所述任务集合中各任务在上一调度周期执行后的实际资源消耗信息。

[0018] 结合第一方面,或第一方面的第一种可能的实现方式,或第一方面的第二种可能的实现方式,或第一方面的第三种可能的实现方式,在第四种可能的实现方式中,所述将所述每个任务的资源消耗信息与所述可用计算资源信息进行比较,并将所述任务集合中对应比较结果满足预置的任务调度条件的任务确定为目标任务,包括:

[0019] 将所述任务集合每个任务的所述资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较;

[0020] 当比较结果为任务的所述资源消耗量小于所述可用资源量时,则记录该任务对应的比较结果满足预置的任务调度条件;

[0021] 将所述任务集合中记录比较结果满足所述任务调度条件的至少一个任务确定为当前调度周期的目标任务。

[0022] 结合第一方面,或第一方面的第一种可能的实现方式,或第一方面的第二种可能的实现方式,或第一方面的第三种可能的实现方式,在第五种可能的实现方式中,所述将所述每个任务的资源消耗信息与所述可用计算资源信息进行比较,并将所述任务集合中对应比较结果满足预置的任务调度条件的任务确定为目标任务,包括:

[0023] 根据所述任务集合中每个任务的资源消耗信息指示的任务集群类型和所述可用计算资源信息指示的可用集群类型,从所述任务集合中确定出所述任务集群类型与所述可用集群类型匹配的任务子集合;

[0024] 将所述任务子集合中任务的资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较;

[0025] 当比较结果为任务的所述资源消耗量小于所述可用资源量时,则记录该任务对应的比较结果满足预置的任务调度条件;

[0026] 将所述任务子集合中记录比较结果满足所述任务调度条件的至少一个任务作为当前调度周期的目标任务。

[0027] 结合第一方面,或第一方面的第一种可能的实现方式,或第一方面的第二种可能的实现方式,或第一方面的第三种可能的实现方式,或第一方面的第四种可能的实现方式,或第一方面的第五种可能的实现方式,在第六种可能的实现方式中,所述将所述每个任务的资源消耗信息与所述可用计算资源信息进行比较,并将所述任务集合中对应比较结果满足预置的任务调度条件的任务确定为目标任务之后,还包括:

[0028] 控制所述任务集合中确定的目标任务执行;

[0029] 当所述目标任务执行完成后,获取所述目标任务在当前调度周期的实际资源消耗信息;

[0030] 根据所述当前调度周期的实际资源消耗信息,更新预置的任务资源消耗映射表中记录的所述目标任务的实际资源消耗信息。

[0031] 结合第一方面的第二种可能的实现方式,在第七种可能的实现方式中,所述方法

还包括：

[0032] 当所述任务集合中每个任务的所述资源消耗信息指示的资源消耗量大于所述可用计算资源信息所指示的可用资源量时，发出提示消息以提示优化该任务；

[0033] 或者，所述任务调度图中生成的任务执行进度小于预设阈值时，发出提示消息以提示优化该任务。

[0034] 本发明第二方面提供一种任务调度的装置，包括：

[0035] 可用计算资源信息获取模块，用于获取用于执行任务的可用计算资源信息；

[0036] 任务资源消耗信息确定模块，用于在接收到已配置完成的任务集合时，确定所述任务集合中每个任务的资源消耗信息；

[0037] 比较确定模块，用于将所述任务资源消耗信息确定模块确定的每个任务的资源消耗信息与所述可用计算资源信息获取模块获得的可用计算资源信息进行比较，并根据所述比较结果，将所述任务集合中对应比较结果满足预置的任务调度条件的任务确定为目标任务。

[0038] 结合第二方面，在第一种可能的实现方式中，所述装置还包括：

[0039] 任务调度图生成模块，用于根据所述任务集合中每个任务的任务依赖关系和 / 或时间依赖关系，生成任务调度图；

[0040] 其中，所述任务调度图包括标识所述任务集合中每个任务的任务标记和标识任务间任务依赖关系和 / 或时间依赖关系的关系标记。

[0041] 结合第二方面的第一种可能的实现方式，在第二种可能的实现方式中，所述任务调度图生成模块生成的任务调度图还包括所述任务集合中每个任务的资源消耗信息标记和从所述任务集合中确定出的目标任务的执行进度；

[0042] 其中，所述资源消耗信息标记用于标识所述任务集合中每个任务的资源消耗信息；所述目标任务的执行进度是根据目标任务的资源消耗信息和当前检测到的所述目标任务在执行过程中的已消耗资源信息获得的。

[0043] 结合第二方面，或第二方面的第一种可能的实现方式，或第二方面的第二种可能的实现方式，在第三种可能的实现方式中，所述任务资源消耗信息确定模块，包括：

[0044] 查找单元，用于在接收到已配置完成的任务集合时，从任务资源消耗映射表中查找所述任务集合中每个任务在上一调度周期的实际资源消耗信息；

[0045] 估算单元，用于根据所述查找单元查找的实际资源消耗信息估算所述任务集合中每个任务在当前调度周期的资源消耗信息；

[0046] 其中，所述任务资源消耗映射表为预先设置的记录有所述任务集合中各任务在上一调度周期执行后的实际资源消耗信息。

[0047] 结合第二方面，或第二方面的第一种可能的实现方式，或第二方面的第二种可能的实现方式，或第二方面的第三种可能的实现方式，在第四种可能的实现方式中，所述比较确定模块，包括：

[0048] 第一比较单元，用于将所述任务集合每个任务的所述资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较；

[0049] 第一记录单元，用于当所述第一比较单元的比较结果为任务的所述资源消耗量小于所述可用资源量时，则记录该任务对应的比较结果匹配预置的任务调度条件；

[0050] 第一确定单元,用于将所述任务集合中所述第一记录单元记录比较结果满足所述任务调度条件的至少一个任务确定为当前调度周期的目标任务。

[0051] 结合第二方面,或第二方面的第一种可能的实现方式,或第二方面的第二种可能的实现方式,或第二方面的第三种可能的实现方式,在第五种可能的实现方式中,所述比较确定模块,包括:

[0052] 类型匹配单元,用于将所述任务集合中每个任务的资源消耗信息指示的任务集群类型和所述可用计算资源信息指示的可用集群类型进行比较,从所述任务集合中确定出所述任务集群类型与所述可用集群类型匹配的任务子集合;

[0053] 第二比较单元,用于将所述类型匹配单元确定的任务子集合中任务的资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较;

[0054] 第二记录单元,用于当所述第二比较单元的比较结果为任务的所述资源消耗量小于所述可用资源量时,则记录该任务对应的比较结果满足预置的任务调度条件;

[0055] 第二确定单元,用于将所述任务子集合中所述第二记录单元记录的比较结果满足所述任务调度条件的至少一个任务作为当前调度周期的目标任务。

[0056] 结合第二方面,或第二方面的第一种可能的实现方式,或第二方面的第二种可能的实现方式,或第二方面的第三种可能的实现方式,或第二方面的第四种可能的实现方式,或第二方面的第五种可能的实现方式,在第六种可能的实现方式中,所述装置还包括:

[0057] 控制模块,用于控制所述任务集合中确定的目标任务执行;

[0058] 实际资源消耗信息获取模块,用于当所述控制模块控制所述目标任务执行完成后,获取所述目标任务在当前调度周期的实际资源消耗信息;

[0059] 更新模块,用于根据所述实际资源信息获取模块获取的当前调度周期所述目标任务的实际资源消耗信息,更新预置的任务资源消耗映射表中记录的所述目标任务的实际资源消耗信息。

[0060] 结合第二方面的第二种可能的实现方式,在第七种可能的实现方式中,所述装置还包括:

[0061] 提示消息发出模块,用于当所述任务集合中每个任务的所述资源消耗信息指示的资源消耗量大于所述可用计算资源信息指示的可用资源量时,或者所述任务调度图中生成的任务的执行进度小于预设阈值时,发出提示消息以修改该任务的依赖关系。

[0062] 本发明第三方面提供一种计算机存储介质,

[0063] 所述计算机存储介质可存储有程序,该程序执行时包括第一方面提供了一种任务调度的方法的部分或全部步骤。

[0064] 本发明第四方面提供一种任务调度设备,包括:处理器、通信接口,

[0065] 所述通信接口用于接收数据仓库中已配置的任务集合和可用计算资源信息;

[0066] 所述处理器用于获取用于执行任务的可用计算资源信息;在接收到已配置完成的任务集合时,确定所述任务集合中每个任务的资源消耗信息;将所述任务集合中每个任务的资源消耗信息与所述可用计算资源信息进行比较,并将所述任务集合中对应比较结果满足预置的任务调度条件的任务确定为目标任务。

[0067] 结合第四方面,在第一种可能的实现方式中,所述处理器还用于:

[0068] 根据所述任务集合中每个任务的依赖关系和/或时间依赖关系,生成任务调

度图；

[0069] 其中,所述任务调度图包括标识所述任务集合中每个任务的任务标记和标识任务间任务依赖关系和 / 或时间依赖关系的关系标记。

[0070] 结合第四方面的第一种可能的实现方式,在第二种可能的实现方式中,所述处理器生成的任务调度图还包括所述任务集合中每个任务的资源消耗信息标记和从所述任务集合中确定出的目标任务的执行进度；

[0071] 其中,所述资源消耗信息标记用于标识所述任务集合中每个任务的资源消耗信息 ;所述目标任务的执行进度是根据目标任务的资源消耗信息和当前检测到的所述目标任务在执行过程中的已消耗资源信息获得的。

[0072] 结合第四方面,或第四方面的第一种可能的实现方式,或第四方面的第二种可能的实现方式,在第三种可能的实现方式中,所述处理器具体用于：

[0073] 从任务资源消耗映射表中查找所述任务集合中每个任务在上一调度周期的实际资源消耗信息；

[0074] 根据所述实际资源消耗信息估算所述任务集合中每个任务在当前调度周期的资源消耗信息；

[0075] 其中,所述任务资源消耗映射表为预先设置的记录有所述任务集合中各任务在上一调度周期执行后的实际资源消耗信息。

[0076] 结合第四方面,或第四方面的第一种可能的实现方式,或第四方面的第二种可能的实现方式,或第四方面的第三种可能的实现方式,在第四种可能的实现方式中,所述处理器具体用于：

[0077] 将所述任务集合每个任务的所述资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较；

[0078] 当比较结果为任务的所述资源消耗量小于所述可用资源量时,则记录该任务对应的比较结果满足预置的任务调度条件；

[0079] 将所述任务集合中记录比较结果满足所述任务调度条件的至少一个任务确定为当前调度周期的目标任务。

[0080] 结合第四方面,或第四方面的第一种可能的实现方式,或第四方面的第二种可能的实现方式,或第四方面的第三种可能的实现方式,在第五种可能的实现方式中,所述处理器还用于：

[0081] 根据所述任务集合中每个任务的资源消耗信息指示的任务集群类型和所述可用计算资源信息指示的可用集群类型,将所述任务集合中确定出所述任务集群类型与所述可用集群类型匹配的任务子集合；

[0082] 将所述任务子集合中任务的资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较；

[0083] 当比较结果为任务的所述资源消耗量小于所述可用资源量时,则记录该任务对应的比较结果满足预置的任务调度条件；

[0084] 将所述任务子集合中记录比较结果满足所述任务调度条件的至少一个任务作为当前调度周期的目标任务。

[0085] 结合第四方面,或第四方面的第一种可能的实现方式,或第四方面的第二种可能

的实现方式,或第四方面的第三种可能的实现方式,或第四方面的第四种可能的实现方式,或第四方面的第五种可能的实现方式,在第六种可能的实现方式中,所述处理器还用于:

[0086] 控制所述任务集合中确定的目标任务执行;

[0087] 当所述目标任务执行完成后,获取所述目标任务在当前调度周期的实际资源消耗信息;

[0088] 根据所述当前调度周期的实际资源消耗信息,更新预置的任务资源消耗映射表中记录的所述目标任务的实际资源消耗信息。

[0089] 结合第四方面的第二种可能的实现方式,在第七种可能的实现方式中,所述处理器还用于:

[0090] 当所述任务集合中每个任务的所述资源消耗信息指示的资源消耗量大于所述可用计算资源信息指示的可用资源量时,发出提示消息以提示优化该任务;

[0091] 或者所述任务调度图中显示的任务的执行进度小于预设阈值时,发出提示消息以提示优化该任务。

[0092] 结合第四方面,或第四方面的第一种可能的实现方式,或第四方面的第二种可能的实现方式,或第四方面的第三种可能的实现方式,或第四方面的第四种可能的实现方式,或第四方面的第五种可能的实现方式,或第四方面的第六种可能的实现方式,或第四方面的第七种可能的实现方式,在第八种可能的实现方式中,所述设备还包括:显示器,

[0093] 所述显示器,用于显示所述处理器执行的结果。

[0094] 由上可见,本发明的一些可行的实施方式通过获取用于执行任务的可用计算资源信息,并确定任务集合中每个任务的资源消耗信息;将所述资源消耗信息与所述可用计算资源信息进行比较,将比较结果满足预置的任务调度条件的任务确定为目标任务,优化了任务调度过程的资源利用率,提高了数据仓库的处理性能。

附图说明

[0095] 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的前提下,还可以根据这些附图获得其他的附图。

[0096] 图1为本发明实施例提供的一种任务调度的方法的流程示意图;

[0097] 图2为本发明实施例提供的另一种任务调度的方法的流程示意图;

[0098] 图3为本发明实施例提供的又一种任务调度的方法的流程示意图;

[0099] 图4为本发明实施例提供的一种任务调度的装置的结构示意图;

[0100] 图5为本发明实施例提供的另一种任务调度的装置的结构示意图;

[0101] 图6为图5中任务资源消耗信息确定模块的具体结构示意图;

[0102] 图7为图5中比较确定模块的具体结构示意图;

[0103] 图8为本发明实施例提供的一种数据仓库任务调度系统的结构示意图;

[0104] 图9为本发明实施例提供的一种具体的任务调度图;

[0105] 图10为本发明实施例提供的一种具体的数据仓库任务调度流程图;

[0106] 图11为本发明实施例提供的一种任务调度设备的结构示意图。

具体实施方式

[0107] 请参见图 1, 为本发明实施例提供的一种任务调度的方法的流程示意图, 所述方法可应用在数据仓库和云计算系统中, 具体可由执行任务调度的设备或服务器实现, 本发明实施例的所述方法具体包括:

[0108] S101, 获取用于执行任务的可用计算资源信息;

[0109] 在本发明的实施例中, 该步骤具体可以从数据仓库集成的资源监控系统中获取数据仓库当前的可用计算资源信息(所述资源监控系统可在所述数据仓库中部署整体资源监控代理, 用于监控数据仓库集群中全部的可用计算资源), 所述可用计算资源信息指示的可用资源量可根据 CPU 使用情况和可用内存空间等确定; 另外, 若数据仓库中的可用资源量包括不同的功能集群类型, 如源数据接收、批量数据处理、实时数据处理、数据传输等, 即每个集群上执行不同类型的任务, 则所述可用计算资源信息不仅指示可用的资源量, 还指示所述可用资源量对应的功能集群类型。

[0110] S102, 在接收到已配置完成的任务集合时, 确定所述任务集合中每个任务的资源消耗信息;

[0111] 具体的, 用户可通过任务部署系统来完成任务的部署和配置, 然后所述任务部署系统在本端部署任务及任务的依赖关系等信息, 即本端接收的任务部署系统配置的任务集合, 所述任务集合包括各任务及其之间的依赖关系。

[0112] 可选的, 所述确定所述任务集合中每个任务的资源消耗信息具体为: 首先, 从数据仓库中资源监控系统获取上一调度周期中所述任务集合中各任务的资源消耗量(所述资源监控系统在所述数据仓库集群中部署任务资源监控代理, 其可根据各种任务类型具体开发, 用于监控每个任务消耗的计算资源); 然后, 估算当前调度周期中各任务的资源消耗信息。具体的, 由于同一任务相邻调度周期消耗的资源量大致相等, 因此可将上一调度周期中各任务的资源消耗量作为各任务在当前调度周期中的资源消耗信息。

[0113] 可选的, 所述任务集合中每个任务上一调度周期的资源消耗信息也可以从本端存储的任务资源消耗映射表中查找获得, 其中, 所述任务资源消耗映射表为预先设置的记录有所述任务集合中各任务在上一调度周期执行后的实际资源消耗信息。

[0114] S103, 将所述任务集合中每个任务的资源消耗信息与所述可用计算资源信息进行比较, 并将所述任务集合中对应比较结果满足预置的任务调度条件的任务确定为目标任务。

[0115] 可选地, 所述将所述任务集合中每个任务的资源消耗信息与所述可用计算资源信息进行比较具体为将所述资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较; 将所述比较结果为任务的资源消耗量小于所述可用资源量的任务记录为该任务的比较结果满足预置的任务调度条件; 将所述满足预置的任务调度条件的至少一个任务作为目标任务。

[0116] 其中, 任务调度条件为只调度资源消耗量小于可用资源量的任务。

[0117] 另外, 当所述可用资源量包括不同的功能集群类型时, 所述 S103 还包括根据所述任务集合中每个任务的资源消耗信息指示的任务集群类型和所述可用计算资源信息指示的可用集群类型(即前述的可用资源量对应的可用集群类型)进行匹配, 从所述任务集合

中匹配出与可用集群类型匹配的至少一个任务,这些任务构成任务子集合。

[0118] 将匹配出的任务子集合中任务的资源消耗信息指示的资源消耗量小于所述可用计算资源指示的可用资源量的至少一个任务作为目标任务。

[0119] 本发明实施例通过获取用于执行任务的可用计算资源信息,并确定任务集合中每个任务的资源消耗信息;将所述资源消耗信息与所述可用计算资源信息进行比较,将比较结果满足预置的任务调度条件的任务确定为目标任务,所述方法根据可用计算资源信息调度任务优化了资源利用率,提高了数据仓库的处理性能。

[0120] 图2为本发明实施例的另一种任务调度的方法的流程示意图,所述方法可应用在数据仓库或云计算系统中,具体可由执行任务调度的设备或服务器实现,所述方法具体包括:

[0121] S201,获取用于执行任务的可用计算资源信息;

[0122] 本步骤的执行动作与上述实施例的S101相同,这里不再详述。

[0123] S202,从任务资源消耗映射表中查找所述任务集合中每个任务在上一调度周期的实际资源消耗信息;

[0124] 其中,所述任务资源消耗映射表为预先设置的记录有所述任务集合中各任务在上一调度周期执行后的实际资源消耗信息,所述任务资源消耗映射表中的实际资源消耗信息可以从资源监控系统获取,或直接从资源监控系统在数据仓库集群中部署的任务资源监控代理处获取。

[0125] S203,根据上一调度周期的实际资源消耗信息估算所述任务集合中每个任务在当前调度周期的资源消耗信息;

[0126] S204,将所述任务集合中每个任务的所述资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较;

[0127] S205,当比较结果为任务的所述资源消耗量小于所述可用资源量时,则记录该任务对应的比较结果满足预置的任务调度条件;

[0128] S206,将所述任务集合中记录比较结果满足所述任务调度条件的至少一个任务确定为当前调度周期的目标任务。

[0129] S207,控制所述任务集合中确定的目标任务执行;

[0130] S208,根据所述任务集合中各任务以及各任务间的依赖关系、当前调度周期的资源消耗信息以及所述目标任务正在执行过程中已消耗的资源信息,生成任务调度图;

[0131] 其中,所述任务调度图包括标识所述任务集合中每个任务的标记和标识任务间任务依赖关系和/或时间依赖关系的关系标记,所述任务集合中每个任务的资源消耗信息标记和从所述任务集合中确定出的目标任务的执行进度。

[0132] 所述资源消耗信息标记用于标识所述任务集合中每个任务的资源消耗信息;所述目标任务的执行进度是根据目标任务的资源消耗信息和当前检测到的所述目标任务在执行过程中的已消耗资源信息获得的。

[0133] 可选的,可在接收到任务集合时生成所述任务调度图(此时所述任务调度图包括各任务标记及依赖关系标记);在任务调度过程中,生成所述任务集合中各任务的资源消耗信息标记;在确定出待调度的目标任务时,继续生成所述目标任务的执行进度。本端可将包括这些信息的任务调度图发送到终端进行显示,以使开发或维护人员对任务调度过程进

行实时监控或管理。

[0134] S209,当所述目标任务执行完成后,获取所述目标任务在当前调度周期的实际资源消耗信息;

[0135] 可选的,所述获取所述目标任务在当前调度周期的实际资源消耗信息可从资源监控系统获取,或直接从所述资源监控系统在数据仓库集群中部署的任务资源监控代理获取。

[0136] S210,根据所述当前调度周期的实际资源消耗信息,更新预置的任务资源消耗映射表中记录的所述目标任务的实际资源消耗信息;

[0137] S211,当所述任务集合中每个任务的所述资源消耗信息指示的资源消耗量大于所述可用计算资源信息所指示的可用资源量时,发出提示消息以提示优化该任务;

[0138] S212,所述任务调度图中生成的任务执行进度小于预设阈值时,发出提示消息以提示优化该任务。

[0139] 步骤 S211 和 S212 中,所述发出提示消息以提示优化该任务,具体的,可通过变化所述任务调度图中该任务的颜色标记等进行提示,或通过向弹出告警提示框的形式提示该任务需开发人员或维修人员进行维护。

[0140] 本发明实施例通过获取用于执行任务的可用计算资源信息,并确定任务集合中各任务的资源消耗信息,通过可用计算资源信息与资源消耗信息的比较,将比较结果符合预置的任务调度条件的至少一个任务确定为目标任务;同时在所述目标任务执行过程中可以在生成的任务调度图中生成并显示所述目标任务的执行进度,为优化任务调度提供了直观依据,提高数据仓库的处理性能。

[0141] 图 3 为本发明实施例提供的又一种任务调度的方法的流程示意图,所述方法可应用在数据仓库或云计算系统中,具体可由执行任务调度的设备或服务器实现,本发明实施例所述方法的步骤 S301 至 S303 与上述实施例的 S201 至 S203 相同,这里不再详述。

[0142] 当所述可用计算资源信息包括不同的功能集群类型时,本发明实施例所述方法执行如下步骤:

[0143] S304,根据所述任务集合中每个任务的资源消耗信息指示的任务集群类型和所述可用计算资源信息指示的可用集群类型,从所述任务集合中确定出所述任务集群类型与所述可用集群类型匹配的任务子集合;

[0144] 其中,所述任务子集合中的任务数可以小于或等于所述任务集合。

[0145] S305,将所述任务子集合中任务的资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较;

[0146] S306,当比较结果为任务的所述资源消耗量小于所述可用资源量时,则记录该任务对应的比较结果满足预置的任务调度条件;

[0147] S307,将所述任务子集合中记录的比较结果满足所述任务调度条件的至少一个任务作为当前调度周期的目标任务。

[0148] S308,控制所述任务集合中确定的目标任务执行;

[0149] S309,根据所述任务集合中每个任务的任务依赖关系和/或时间依赖关系,生成任务调度图;

[0150] 其中,所述任务调度图包括标识所述任务集合中每个任务的任务标记和标识任务

间任务依赖关系和 / 或时间依赖关系的关系标记。

[0151] 所述任务调度图还包括所述任务集合中每个任务的资源消耗信息标记和从所述任务集合中确定出的目标任务的执行进度 ; 其中, 所述资源消耗信息标记用于标识所述任务集合中每个任务的资源消耗信息 ; 所述目标任务的执行进度是根据目标任务的资源消耗信息和当前检测到的所述目标任务在执行过程中的已消耗资源信息获得的。

[0152] 本发明实施例的 S310 至 S313 与上述发明实施例的 S209 至 S212 的步骤相同, 这里不再详述。

[0153] 本发明实施例在可用计算资源信息包含不同的功能集群类型时, 从所述任务集合中确定出任务的资源消耗信息指示的资源消耗功能集群类型与所述可用功能集群类型匹配的任务子集合, 再从所述任务子集合中确定出至少一个目标任务, 从而优化了数据仓库的资源利用率, 并提高了数据仓库的处理性能。

[0154] 可选的, 可将上述两个实施例增加判断可用资源信息是否为单一的功能集群类型的步骤, 则上述发明实施例可构成一个技术方案以优化任务调度过程, 提高处理性能。

[0155] 图 4 为本发明实施例提供的一种任务调度的装置的结构示意图, 所述装置具体包括 :

[0156] 可用计算资源信息获取模块 1, 用于获取用于执行任务的可用计算资源信息 ;

[0157] 在本发明实施例中, 可用计算资源信息获取模块 1 可以从数据仓库集成的资源监控系统中获取数据仓库当前的可用计算资源信息, 所述可用计算资源信息指示的可用资源量可根据 CPU 使用情况和可用内存空间等确定 ; 另外, 若数据仓库包括不同的功能集群类型时, 如源数据接收、批量数据处理、实时数据处理、数据传输等, 即每个集群上执行不同类型的任务, 则所述可用计算资源信息不仅指示可用的资源量, 还指示可用资源量对应的功能集群类型。

[0158] 任务资源消耗信息确定模块 2, 用于在接收到已配置完成的任务集合时, 确定所述任务集合中每个任务的资源消耗信息 ;

[0159] 可选的, 所述任务资源消耗信息确定模块 2 确定所述任务集合中每个任务的资源消耗信息具体为 : 首先, 所述任务资源消耗信息确定模块 2 从数据仓库中部署的任务资源监控代理获取上一调度周期中所述任务集合中各任务的资源消耗量 ; 然后, 所述任务资源消耗信息确定模块 2 估算当前调度周期中各任务的资源消耗信息。由于同一任务相邻调度周期消耗的资源量大致相等, 因此可将上一调度周期中各任务的资源消耗量作为各任务在当前调度周期中的资源消耗信息。

[0160] 可选的, 所述任务集合中每个任务上一调度周期的资源消耗信息可以由所述任务资源消耗信息确定模块 2 从任务资源消耗映射表中查找获得, 其中, 所述任务资源消耗映射表为预先设置的记录有所述任务集合中各任务在上一调度周期执行后的实际资源消耗信息。

[0161] 比较确定模块 3, 用于将所述任务资源消耗信息确定模块 2 确定的每个任务的资源消耗信息与所述可用计算资源信息获取模块 1 获得的可用计算资源信息进行比较, 并根据所述比较结果, 将所述任务集合中对应比较结果满足预置的任务调度条件的任务确定为目标任务。

[0162] 可选地, 所述比较确定模块 3 将所述任务集合中每个任务的资源消耗信息与所述

可用计算资源信息进行比较具体为将所述资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较；将所述比较结果为任务的资源消耗量小于所述可用资源量的任务记录为该任务的比较结果满足预置的任务调度条件；将所述满足预置的任务调度条件的至少一个任务作为目标任务。

[0163] 其中，任务调度条件为只调度资源消耗量小于可用资源量的任务。

[0164] 另外，当所述可用资源量包括不同的功能集群时，所述比较确定模块 3 还根据所述任务集合中每个任务的资源消耗信息指示的任务集群类型和所述可用计算资源信息指示的可用集群类型（即前述的可用资源量对应的可用集群类型）进行匹配，从所述任务集合中确定集群类型匹配的至少一个任务，这些任务构成任务子集合。

[0165] 所述比较确定模块 3 将匹配出的任务子集合中任务的资源消耗信息指示的资源消耗量小于所述可用计算资源指示的可用资源量的至少一个任务作为目标任务。

[0166] 本发明实施例中，可用计算资源信息获取模块 1 获取用于执行任务的可用计算资源信息，任务资源消耗信息确定模块 2 确定任务集合中每个任务的资源消耗信息；比较确定模块 3 将所述资源消耗信息与所述可用计算资源信息进行比较，将比较结果满足预置的任务调度条件的任务确定为目标任务，根据可用计算资源信息调度任务优化了资源利用率，提高了数据仓库的处理性能。

[0167] 图 5 为本发明实施例提供的另一种任务调度的装置的结构示意图，所述装置除包括上述实施例的可用计算资源获取模块 1、任务资源消耗信息确定模块 2 和比较确定模块 3 之外，还包括：

[0168] 任务调度图生成模块 4，用于根据所述任务集合中每个任务的任务依赖关系和/或时间依赖关系，生成任务调度图；

[0169] 其中，所述任务调度图包括标识所述任务集合中每个任务的任务标记和标识任务间任务依赖关系和/或时间依赖关系的关系标记。

[0170] 所述任务调度图生成模块 4 生成的任务调度图还包括所述任务集合中每个任务的资源消耗信息标记和从所述任务集合中确定出的目标任务的执行进度；

[0171] 可选地，所述任务调度图生成模块 4 可为任务调度装置的可视化模块，根据所述任务集合中每个任务的任务依赖关系和/或时间依赖关系，所述任务集合中每个任务的资源消耗信息和从所述任务集合中确定出的目标任务的执行进度生成任务调度图，使得这些信息在任务调度过程中可直观查看。

[0172] 如图 6 所示，所述任务资源消耗信息确定模块 2，包括：

[0173] 查找单元 21，用于在接收到已配置完成的任务集合时，从任务资源消耗映射表中查找所述任务集合中每个任务在上一调度周期的实际资源消耗信息；

[0174] 估算单元 22，用于根据所述查找单元查找的实际资源消耗信息估算所述任务集合中每个任务在当前调度周期的资源消耗信息；

[0175] 其中，所述任务资源消耗映射表为预先设置的记录有所述任务集合中各任务在上一调度周期执行后的实际资源消耗信息。

[0176] 如图 7 所示，所述比较确定模块 3，包括：

[0177] 第一比较单元 31，用于将所述任务集合每个任务的所述资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较；

[0178] 第一记录单元 32,用于当所述第一比较单元的比较结果为任务的所述资源消耗量小于所述可用资源量时,则记录该任务对应的比较结果匹配预置的任务调度条件;

[0179] 第一确定单元 33,用于将所述任务集合中所述第一记录单元记录比较结果满足所述任务调度条件的至少一个任务确定为当前调度周期的目标任务。

[0180] 当所述可用计算资源信息指示的可用资源量包括不同的功能集群类型时,如图 7 所示,所述比较确定模块 3,还包括:

[0181] 类型匹配单元 34,用于将所述任务集合中每个任务的资源消耗信息指示的任务集群类型和所述可用计算资源信息指示的可用集群类型进行比较,从所述任务集合中确定出所述任务集群类型与所述可用集群类型匹配的任务子集合;

[0182] 第二比较单元 35,用于将所述类型匹配单元确定的任务子集合中任务的资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较;

[0183] 第二记录单元 36,用于当所述第二比较单元的比较结果为任务的所述资源消耗量小于所述可用资源量时,则记录该任务对应的比较结果满足预置的任务调度条件;

[0184] 第二确定单元 37,用于将所述任务子集合中所述第二记录单元记录的比较结果满足所述任务调度条件的至少一个任务作为当前调度周期的目标任务。

[0185] 可选的,所述装置还包括:

[0186] 控制模块 5,用于控制所述任务集合中确定的目标任务执行;

[0187] 实际资源消耗信息获取模块 6,用于当所述控制模块控制所述目标任务执行完成后,获取所述目标任务在当前调度周期的实际资源消耗信息;

[0188] 更新模块 7,用于根据所述实际资源信息获取模块获取的当前调度周期所述目标任务的实际资源消耗信息,更新预置的任务资源消耗映射表中记录的所述目标任务的实际资源消耗信息。

[0189] 可选的,所述装置还包括:

[0190] 提示消息发出模块 8,用于当所述任务集合中每个任务的所述资源消耗信息指示的资源消耗量大于所述可用计算资源信息指示的可用资源量时,或者所述任务调度图中生成的任务的执行进度小于预设阈值时,发出提示消息以提示修改该任务或其依赖关系。

[0191] 本发明实施例通过比较确定模块 3 根据可用计算资源信息获取模块 1 和任务资源消耗信息模块 2 的结果从所述任务集合中确定出至少一个目标任务,从而优化了数据仓库的资源利用率,并提高了数据仓库的处理性能。

[0192] 结合图 8、图 9 和图 10 介绍本发明实施例提供的一种数据仓库中任务调度的方法,数据仓库是一个为决策者提供联机分析处理所需要信息的仓储,其任务之间可能存在单向依赖或不存在依赖关系或拥有不同的优先级等关系,面对数据仓库中众多的处理任务,如何提高数据仓库性能和资源利用率是关键,本发明实施例提供了如图 8 所示的数据仓库系统,所述数据仓库系统主要包括任务部署系统、数据仓库集群和任务调度系统。

[0193] 任务部署系统在数据仓库集群中完成各任务部署和任务配置,并在任务调度系统中配置各任务的依赖关系、时序关系,数据开发人员或维护人员可通过任务部署系统提供的接口对部署的任务进行监控和管理;所述任务部署系统还在数据仓库集群中部署整体资源监控代理和任务资源监控代理实时获取数据仓库中的可用计算资源信息和任务的资源消耗信息,并依据上述实施例所述的任一种任务调度的方法对数据仓库集群执行任务调

度,使各任务在数据仓库集群中有序的进行,并提高资源利用率。

[0194] 所述任务调度系统可以包括上述实施例所述的任务调度的装置,在本发明实施例中,所述任务调度装置中所述可用计算资源信息获取模块具体用于完成数据仓库整体可用计算资源信息的状态管理,并保持与数据仓库资源真实情况实时同步;所述比较确定模块用于管理各种调度规则(除上述发明实施例所述的任务调度方法,还对数据仓库的任务集合进行更新、归档,并生成各种报表传递给终端显示),即比较确定模块是任务调度系统的核心;当所述比较确定模块确定了待调度的目标任务,所述控制模块还可以控制所述目标任务执行;任务调度系统中的任务调度图生成模块还实现了调度的可视化即生成任务调度图,通过所述任务调度图中显示的任务、任务间的依赖关系、任务的资源消耗信息和目标任务的执行进度向管理人员直观的展示任务调度情况;当所述目标任务执行完成后,任务调度系统中的实际资源消耗信息获取模块还可以实时获取所述目标任务在当前调度周期的实际资源消耗信息;更新模块根据所述实际资源消耗信息获取模块的结果实时更新任务资源消耗映射表中记录的信息;另外,任务调度系统还具有提示消息发出模块用于当所述任务集合中每个任务的所述资源消耗信息指示的资源消耗量大于所述可用计算资源信息指示的可用资源量时,或者所述任务调度图中生成的任务的执行进度小于预设阈值时,向终端发出提示消息以提示数据开发人员或维护人员重点优化该任务。

[0195] 具体的,如图9所示,所述任务调度系统生成的任务调度图可用圆圈表示每个任务,可用任务标记命名,如任务1、任务2、任务3、任务4、任务5、任务6、任务7;各任务之间的连线表示任务间的依赖关系(排在右边的任务依赖左边的任务),如任务4的执行依赖任务2和任务3的执行结果;圆圈的大小表示该任务预估的资源消耗信息指示的资源消耗量,如任务1预估的资源消耗量大于任务2预估的资源消耗量;圆圈的空白或阴影表示任务的执行状态,同时具有空白和阴影的圆圈表示目标任务正在执行,空白与阴影的比例表示任务的执行进度,如任务1、任务2、任务3、任务4和任务7表示未执行,任务5表示已经执行,任务6表示正在执行,并已执行了任务的四分之一。所述任务调度图也可以通过其他方式表示任务的属性、状态、依赖关系、时序关系等信息。

[0196] 基于图8的数据仓库任务调度系统,本发明实施例还给出了如图10所示的一种具体的数据仓库任务调度流程,包括:

[0197] S401,资源监控系统获取数据仓库当前的可用计算资源信息;

[0198] S402,任务调度系统从所述资源监控系统获取所述可用计算资源信息;

[0199] S403,数据开发人员通过任务部署系统在任务调度系统中部署任务集合,所述任务集合包括部署的各任务以及任务间的依赖关系;

[0200] S404,所述任务调度系统从任务资源消耗信息映射表中查找所述任务集合中各任务在上一调度周期中的实际资源消耗信息,根据所述实际资源消耗信息估算所述任务集合中各任务在当前调度周期的资源消耗信息;

[0201] 可选的,所述任务调度系统也可以从所述资源监控系统实时获取所述任务集合中各任务在上一调度周期的实际资源消耗信息。

[0202] S405,所述任务调度系统根据所述可用计算资源信息和各任务在上一调度周期的实际资源消耗信息确定满足预置的待调度条件的至少一个目标任务;

[0203] S406,将确定的所述目标任务指令发送给数据仓库;

[0204] S407,所述数据仓库执行所述目标任务;

[0205] S408,所述任务调度系统从所述资源监控系统获取所述目标任务在当前调度周期中的实际资源消耗信息;

[0206] S409,根据所述目标任务以及所述目标任务在当前调度周期中的实际资源消耗信息,任务调度系统更新所述任务集合和任务资源消耗信息映射表。

[0207] 可选的,所述任务调度系统在控制执行完确定的目标任务后,更新所述任务集合,进一步地,可通过预置的任务计数器或任务扫描器检测所述任务集合中是否还有未执行的任务;若所述检测的结果为所述任务集合中没有未执行的任务则停止任务调度;若所述检测的结果为所述任务集合中还有未执行的任务,可针对更新后的任务集合中未执行的任务根据本发明实施例提供的任务调度方法启动任务调度。

[0208] 另外,当数据仓库每次增加或删减某个任务时,均需执行步骤S403。当所述可用计算资源信息包括不同类型的功能集群时,所述步骤S405所确定的目标任务的功能集群类型还必须与可用计算资源信息指示的可用功能集群类型匹配,例如大数据处理任务必须在大数据处理集群上执行等。

[0209] 本发明实施例根据数据仓库的可用计算资源信息和任务预估的资源消耗信息的匹配来进行任务调度,从而可以最大限度的利用计算资源,提高数据仓库的处理性能。

[0210] 本发明实施例还提供了一种计算机存储介质,其中,所述计算机存储介质可存储有程序,该程序执行时包括上述图1至图3所述实施例中记载的任务调度的方法的部分或全部步骤。

[0211] 请参见图11,为本发明实施例提供的一种任务调度设备的结构示意图,所述任务调度设备可以包括处理器01、通信接口02(所述处理器01的数量可以为一个或多个,图11中以一个处理器01为例)。本发明的一些实施例中,处理器01和通信接口02可通过通信总线或其他方式连接,其中,图11以通信总线连接为例。

[0212] 所述通信接口02,用于接收数据仓库中已配置的任务集合和可用计算资源信息;

[0213] 所述处理器01用于获取用于执行任务的可用计算资源信息;在接收到已配置完成的任务集合时,确定所述任务集合中每个任务的资源消耗信息;将所述任务集合中每个任务的资源消耗信息与所述可用计算资源信息进行比较,并将所述任务集合中对应比较结果满足预置的任务调度条件的任务确定为目标任务。

[0214] 所述处理器01还用于:

[0215] 根据所述任务集合中每个任务的任务依赖关系和/或时间依赖关系,生成任务调度图;

[0216] 其中,所述任务调度图包括标识所述任务集合中每个任务的任务标记和标识任务间任务依赖关系和/或时间依赖关系的关系标记。

[0217] 所述处理器01生成的任务调度图还包括所述任务集合中每个任务的资源消耗信息标记和从所述任务集合中确定出的目标任务的执行进度;

[0218] 其中,所述资源消耗信息标记用于标识所述任务集合中每个任务的资源消耗信息;所述目标任务的执行进度是根据目标任务的资源消耗信息和当前检测到的所述目标任务在执行过程中的已消耗资源信息获得的。

[0219] 所述处理器01具体用于:

[0220] 从任务资源消耗映射表中查找所述任务集合中每个任务在上一调度周期的实际资源消耗信息；

[0221] 根据所述实际资源消耗信息估算所述任务集合中每个任务在当前调度周期的资源消耗信息；

[0222] 其中,所述任务资源消耗映射表为预先设置的记录有所述任务集合中各任务在上一调度周期执行后的实际资源消耗信息。

[0223] 所述处理器 01 具体用于：

[0224] 将所述任务集合每个任务的所述资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较；

[0225] 当比较结果为任务的所述资源消耗量小于所述可用资源量时,则记录该任务对应的比较结果满足预置的任务调度条件；

[0226] 将所述任务集合中记录比较结果满足所述任务调度条件的至少一个任务确定为当前调度周期的目标任务。

[0227] 所述处理器 01 具体用于：

[0228] 根据所述任务集合中每个任务的资源消耗信息指示的任务集群类型和所述可用计算资源信息指示的可用集群类型,将所述任务集合中确定出所述任务集群类型与所述可用集群类型匹配的任务子集合；

[0229] 将所述任务子集合中任务的资源消耗信息指示的资源消耗量与所述可用计算资源信息指示的可用资源量进行比较；

[0230] 当比较结果为任务的所述资源消耗量小于所述可用资源量时,则记录该任务对应的比较结果满足预置的任务调度条件；

[0231] 将所述任务子集合中记录比较结果满足所述任务调度条件的至少一个任务作为当前调度周期的目标任务。

[0232] 所述处理器 01 还用于：

[0233] 控制所述任务集合中确定的目标任务执行；

[0234] 当所述目标任务执行完成后,获取所述目标任务在当前调度周期的实际资源消耗信息；

[0235] 根据所述当前调度周期的实际资源消耗信息,更新预置的任务资源消耗映射表中记录的所述目标任务的实际资源消耗信息。

[0236] 所述处理器 01 还用于：

[0237] 当所述任务集合中每个任务的所述资源消耗信息指示的资源消耗量大于所述可用计算资源信息指示的可用资源量时,发出提示消息以提示优化该任务；

[0238] 或者所述任务调度图中显示的任务的执行进度小于预设阈值时,发出提示消息以提示优化该任务。

[0239] 可选的,本发明实施例所述的任务调度设备还包括显示器 03,

[0240] 所述显示器 03,用于显示所述处理器 01 执行的结果,如显示所述处理器 01 生成的任务调度图。

[0241] 本发明实施例通过获取用于执行任务的可用计算资源信息,根据接收到的各任务的资源消耗信息查找与所述可用计算资源信息匹配的目标任务,优化了任务调度过程,并

且根据生成的调度图实时查看任务的资源消耗信息和执行进度为任务调度优化提供直观依据,从而提高数据仓库的性能。

[0242] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的程序可存储于一计算机可读取存储介质中,该程序在执行时,可包括如上述各方法的实施例的流程。其中,所述的存储介质可为磁碟、光盘、只读存储记忆体 (Read-Only Memory, ROM) 或随机存储记忆体 (Random Access Memory, RAM) 等。

[0243] 以上所揭露的仅为本发明较佳实施例而已,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。

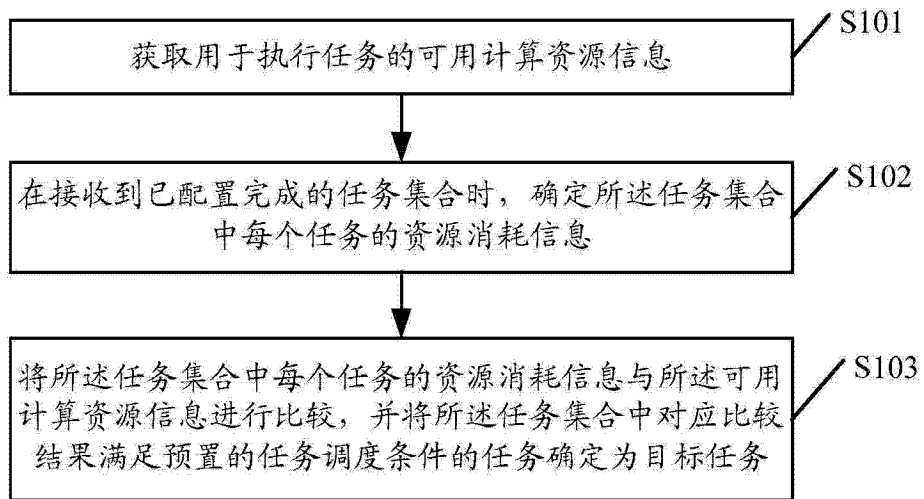


图 1

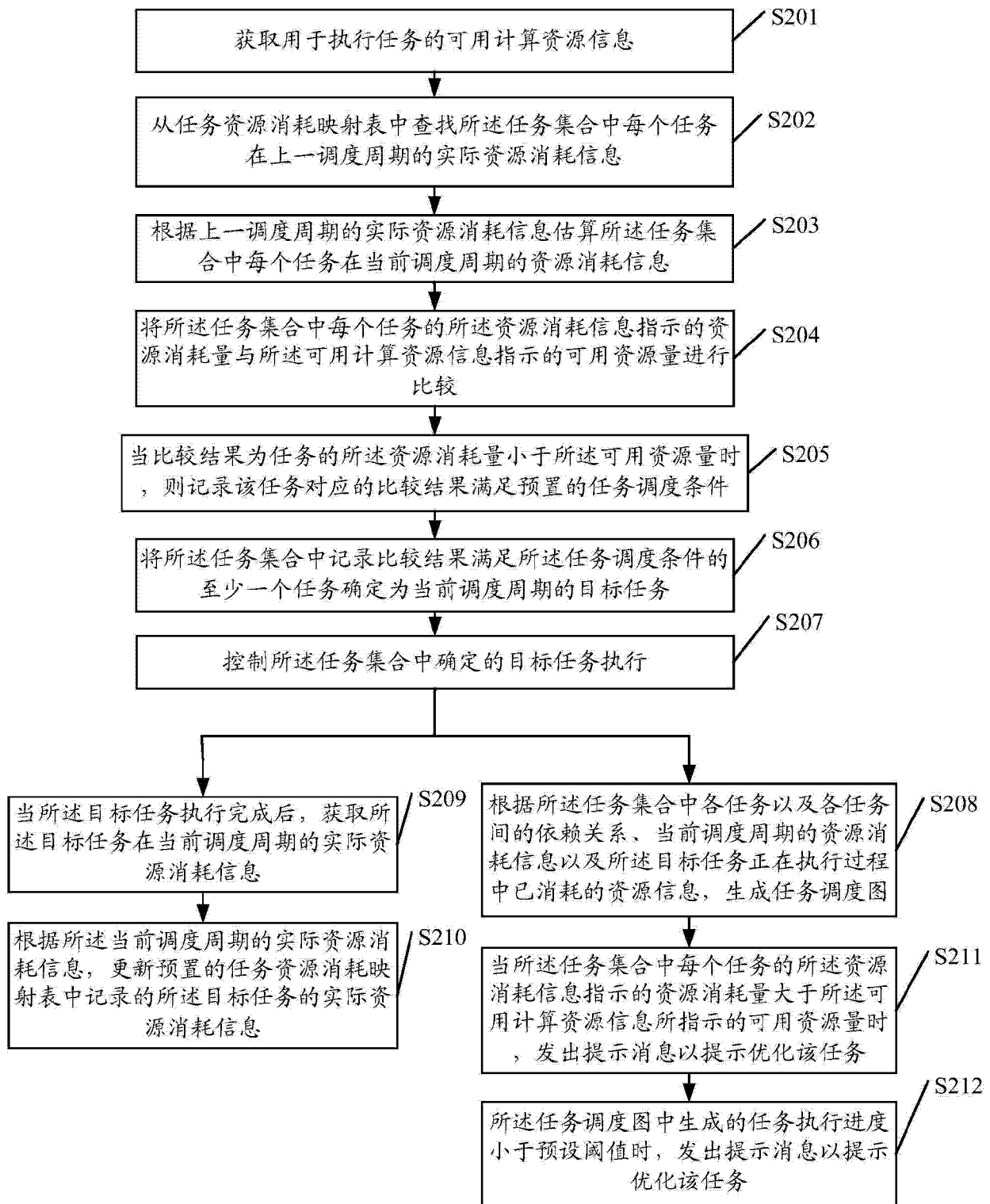


图 2

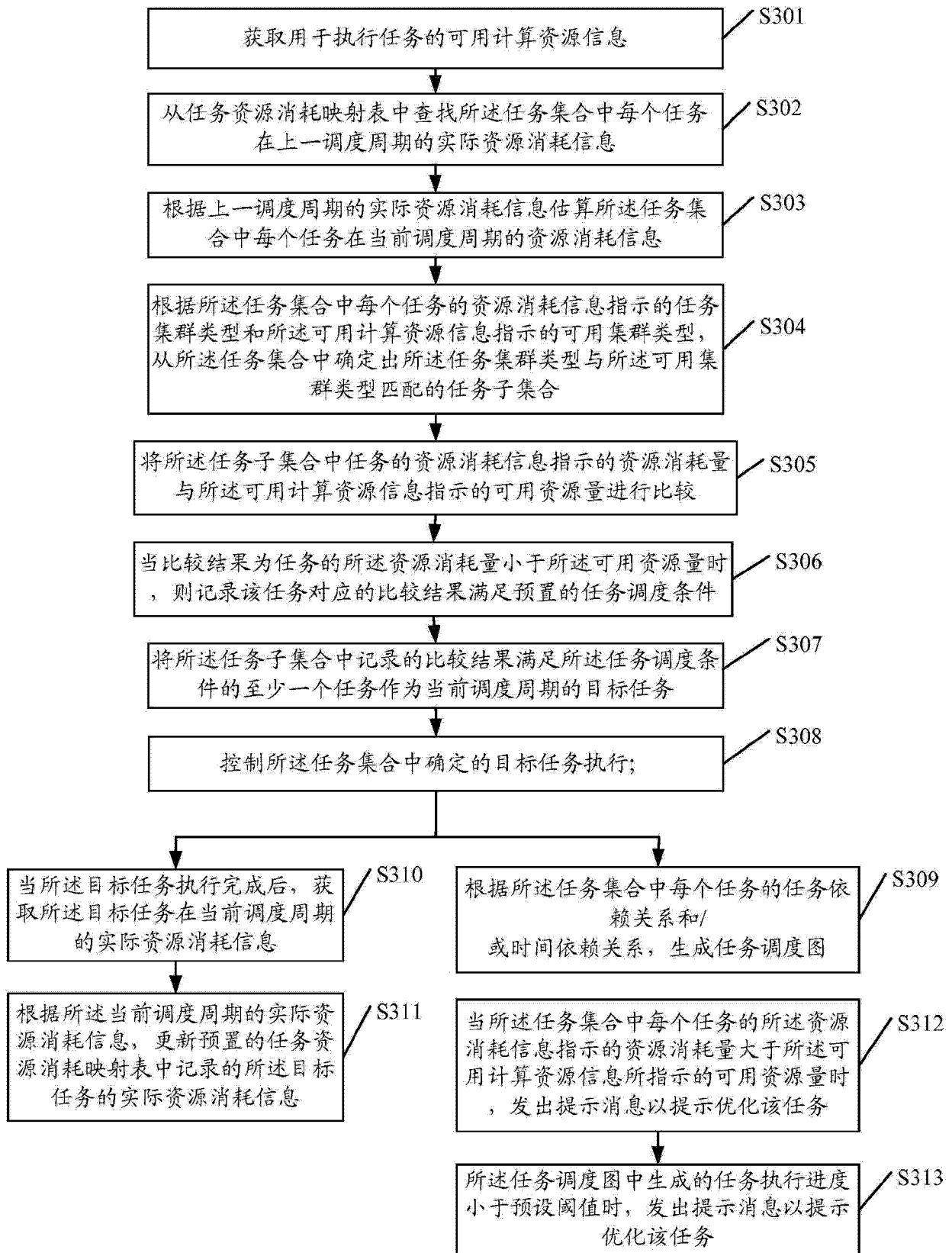


图 3

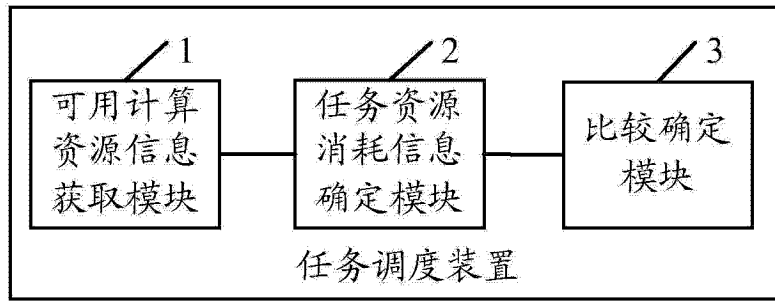


图 4

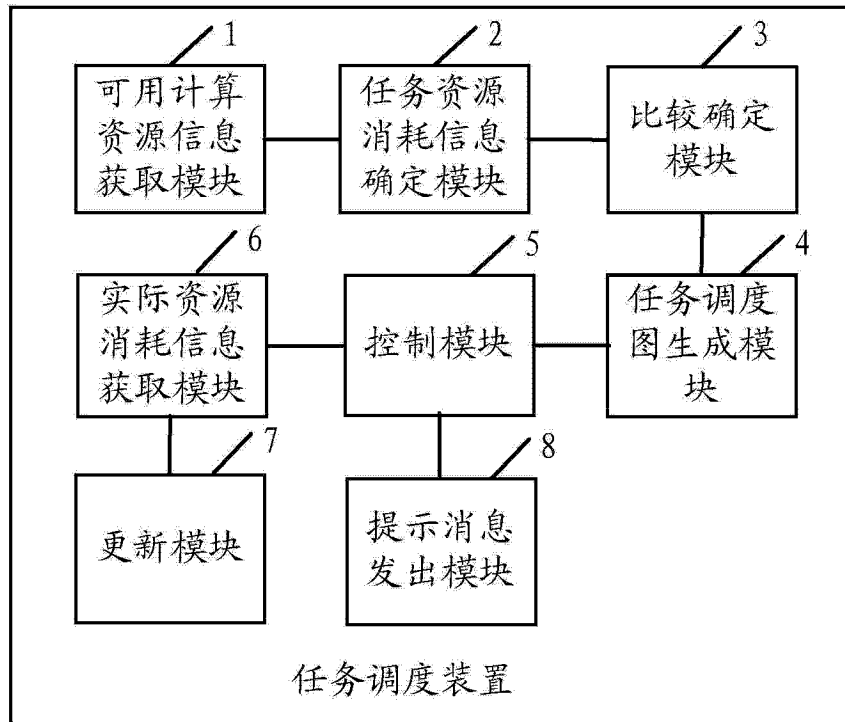


图 5

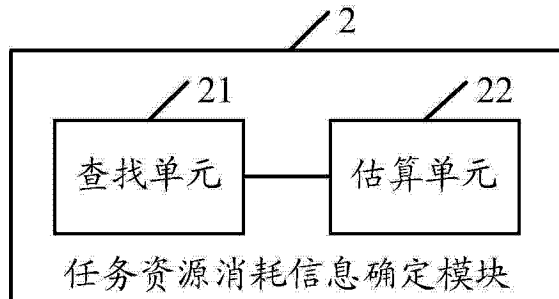


图 6

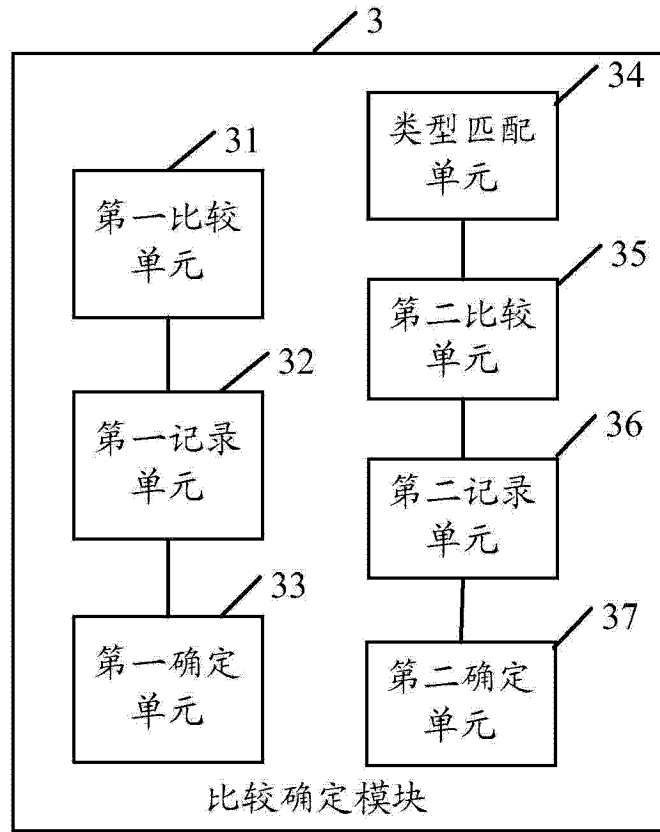


图7

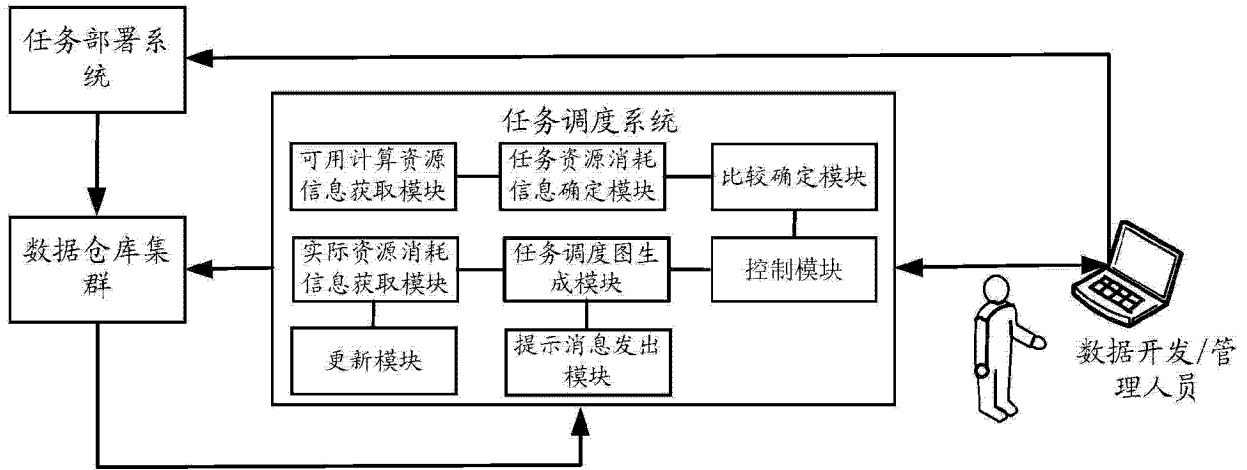


图8

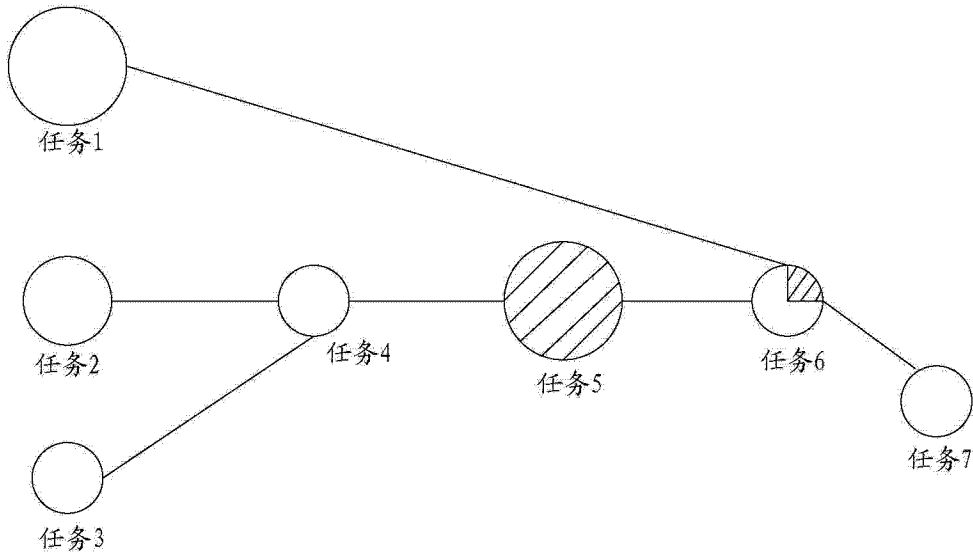


图 9

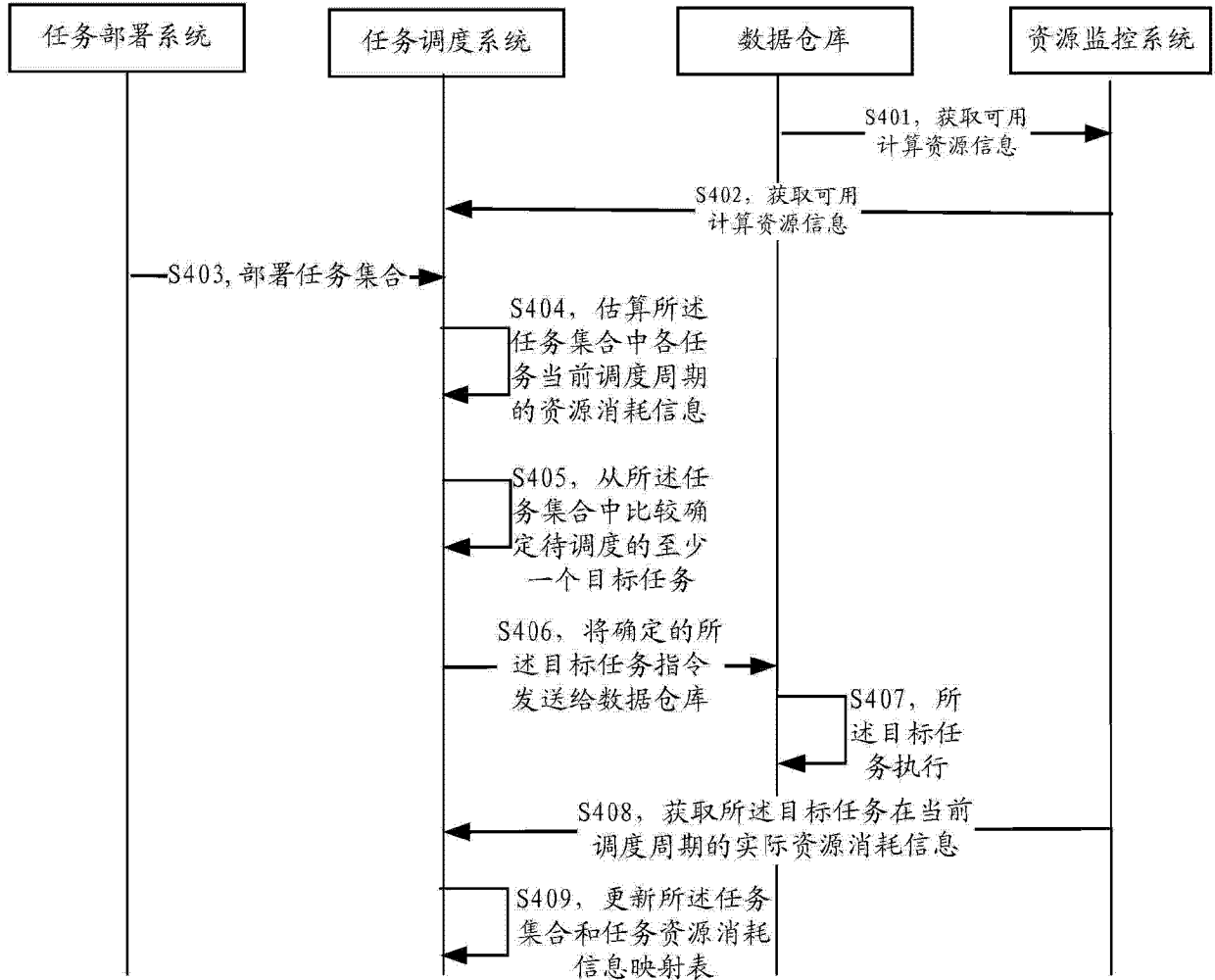


图 10

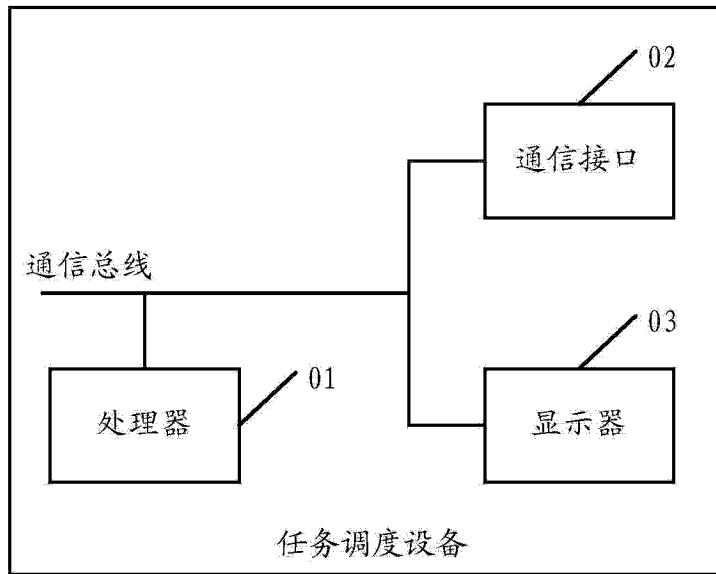


图 11