



(12)发明专利申请

(10)申请公布号 CN 113448717 A

(43)申请公布日 2021.09.28

(21)申请号 202010219853.X

(22)申请日 2020.03.25

(71)申请人 北京沃东天骏信息技术有限公司  
地址 100176 北京市北京经济技术开发区  
科创十一街18号院2号楼4层A402室  
申请人 北京京东世纪贸易有限公司

(72)发明人 高田林

(74)专利代理机构 中原信达知识产权代理有限  
责任公司 11219  
代理人 王安娜 陈继越

(51)Int.Cl.  
G06F 9/50(2006.01)  
H04L 29/08(2006.01)

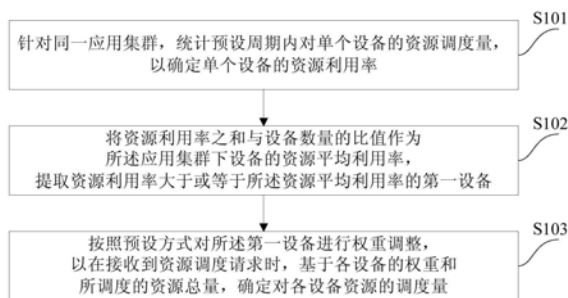
权利要求书2页 说明书9页 附图4页

(54)发明名称

一种资源调度方法和装置

(57)摘要

本发明公开了一种资源调度方法和装置,涉及计算机技术领域。该方法的一具体实施方式包括:针对同一应用集群,统计预设周期内对单个设备的资源调度量,以确定单个设备的资源利用率;将资源利用率之和与设备数量的比值作为所述应用集群下设备的资源平均利用率,提取资源利用率大于或等于所述资源平均利用率的第一设备;按照预设方式对第一设备进行权重调整,以在接收到资源调度请求时,基于各设备的权重和所调度的资源总量,确定对各设备资源的调度量。该实施方式在分布式集群环境下,根据现有集群CPU利用率的平均值,自动化调整利用率超过平均值的设备,如采用降低权重方法,且整体无需手动操作,实现整个集群CPU利用率的动态平衡。



1. 一种资源调度方法,其特征在于,包括:

针对同一应用集群,统计预设周期内对单个设备的资源调度量,以确定单个设备的资源利用率;

将资源利用率之和与设备数量的比值作为所述应用集群下设备的资源平均利用率,提取资源利用率大于或等于所述资源平均利用率的第一设备;

按照预设方式对所述第一设备进行权重调整,以在接收到资源调度请求时,基于各设备的权重和所调度的资源总量,确定对各设备资源的调度量。

2. 根据权利要求1所述的方法,其特征在于,所述按照预设方式对所述第一设备进行权重调整,包括:

获取对所述第一设备的预设权重,判断所述第一设备的资源利用率是否大于或等于所述预设权重;

若是,则将所述第一设备的权重调整为第一权重;

若否,则计算所述第一设备的资源利用率与所述资源平均利用率的差值,结合所述预设权重,得到第二权重,进而将所述第一设备的权重调整为所述第二权重。

3. 根据权利要求1所述的方法,其特征在于,还包括:

获取单个设备的IP地址和预设权重并上传至数据中心,以通过所述数据中心将单个设备的数据传输至资源调度模块;其中,所述数据包括IP地址、资源利用率和预设权重;

在所述确定对各设备资源的调度量之后,还包括:

所述资源调度模块根据调度结果,将设备的IP地址传输至发送所述资源调度请求的第二设备,以使所述第二设备通过传输的IP地址进行资源调度。

4. 根据权利要求3所述的方法,其特征在于,所述通过所述数据中心将单个设备的数据传输至资源调度模块,包括:

所述资源调度模块在监测到所述数据中心内设备的数据发生变化时,从所述数据中心获取设备变化后的数据。

5. 根据权利要求3所述的方法,其特征在于,在所述将设备的IP地址传输至发送所述资源调度请求的第二设备之前,还包括:

获取各第二设备对资源的调度请求量,确定调度请求量之和为调度量的至少一个第二设备,进而建立设备与所述至少一个第二设备之间的映射关系。

6. 根据权利要求3所述的方法,其特征在于,在所述将设备的IP地址传输至发送所述资源调度请求的第二设备之前,还包括:

按照所确定的对各设备资源的调度量进行排序,得到设备队列;

按照各第二设备对资源的调度请求量进行排序,得到第二设备队列;

按照轮训方式,将所述第二设备队列中的第二设备分配至所述设备队列中的设备,进而建立设备与第二设备之间的映射关系。

7. 一种资源调度装置,其特征在于,包括:

资源计算模块,用于针对同一应用集群,统计预设周期内对单个设备的资源调度量,以确定单个设备的资源利用率;

设备提取模块,用于将资源利用率之和与设备数量的比值作为所述应用集群下设备的资源平均利用率,提取资源利用率大于或等于所述资源平均利用率的第一设备;

权重调整模块,用于按照预设方式对所述第一设备进行权重调整,以在接收到资源调度请求时,基于各设备的权重和所调度的资源总量,确定对各设备资源的调度量。

8.一种电子设备,其特征在于,包括:

一个或多个处理器;

存储装置,用于存储一个或多个程序,

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求1-6中任一所述的方法。

9.一种计算机可读介质,其上存储有计算机程序,其特征在于,所述程序被处理器执行时实现如权利要求1-6中任一所述的方法。

## 一种资源调度方法和装置

### 技术领域

[0001] 本发明涉及计算机技术领域,尤其涉及一种资源调度方法和装置。

### 背景技术

[0002] 在互联网环境下,用户访问量通常是千万甚至亿级,因而提供服务的应用集群数量少则上百多则上千台,每台设备CPU(Central Proc essing Unit,中央处理器)资源的高低都会影响到整个集群服务的质量。

[0003] 在现有设备管理中,由于设备硬件老化或资源分配不均衡问题,导致有些设备CPU的利用率过高,有些则过低,导致服务提供不稳定问题。开发人员通常选用资源监控报警方式,将CPU利用率过高的设备手动下线会降低其负载权重值,以通过减少资源访问降低CPU资源消耗。

[0004] 在实现本发明的过程中,发明人发现现有技术至少存在如下问题:

[0005] 分布式集群中的设备数量众多,选用手动处理CPU利用率过高设备的方式,耗时又耗力。

### 发明内容

[0006] 有鉴于此,本发明实施例提供一种资源调度方法和装置,至少能够解决现有技术中针对CPU利用率过高的设备,需手动调整的问题。

[0007] 为实现上述目的,根据本发明实施例的一个方面,提供了一种资源调度方法,包括:

[0008] 针对同一应用集群,统计预设周期内对单个设备的资源调度量,以确定单个设备的资源利用率;

[0009] 将资源利用率之和与设备数量的比值作为所述应用集群下设备的资源平均利用率,提取资源利用率大于或等于所述资源平均利用率的第一设备;

[0010] 按照预设方式对所述第一设备进行权重调整,以在接收到资源调度请求时,基于各设备的权重和所调度的资源总量,确定对各设备资源的调度量。

[0011] 可选的,所述按照预设方式对所述第一设备进行权重调整,包括:

[0012] 获取对所述第一设备的预设权重,判断所述第一设备的资源利用率是否大于或等于所述预设权重;

[0013] 若是,则将所述第一设备的权重调整为第一权重;

[0014] 若否,则计算所述第一设备的资源利用率与所述资源平均利用率的差值,结合所述预设权重,得到第二权重,进而将所述第一设备的权重调整为所述第二权重。

[0015] 可选的,还包括:

[0016] 获取单个设备的IP地址和预设权重并上传至数据中心,以通过所述数据中心将单个设备的数据传输至资源调度模块;其中,所述数据包括IP地址、资源利用率和预设权重;

[0017] 在所述确定对各设备资源的调度量之后,还包括:

[0018] 所述资源调度模块根据调度结果,将设备的IP地址传输至发送所述资源调度请求的第二设备,以使所述第二设备通过传输的IP地址进行资源调度。

[0019] 可选的,所述通过所述数据中心将单个设备的数据传输至资源调度模块,包括:所述资源调度模块在监测到所述数据中心内设备的数据发生变化时,从所述数据中心获取设备变化后的数据。

[0020] 可选的,在所述将设备的IP地址传输至发送所述资源调度请求的第二设备之前,还包括:获取各第二设备对资源的调度请求量,确定调度请求量之和为调度量的至少一个第二设备,进而建立设备与所述至少一个第二设备之间的映射关系。

[0021] 可选的,在所述将设备的IP地址传输至发送所述资源调度请求的第二设备之前,还包括:

[0022] 按照所确定的对各设备资源的调度量进行排序,得到设备队列;

[0023] 按照各第二设备对资源的调度请求量进行排序,得到第二设备队列;

[0024] 按照轮训方式,将所述第二设备队列中的第二设备分配至所述设备队列中的设备,进而建立设备与第二设备之间的映射关系。

[0025] 为实现上述目的,根据本发明实施例的另一方面,提供了一种资源调度装置,包括:

[0026] 资源计算模块,用于针对同一应用集群,统计预设周期内对单个设备的资源调度量,以确定单个设备的资源利用率;

[0027] 设备提取模块,用于将资源利用率之和与设备数量的比值作为所述应用集群下设备的资源平均利用率,提取资源利用率大于或等于所述资源平均利用率的第一设备;

[0028] 权重调整模块,用于按照预设方式对所述第一设备进行权重调整,以在接收到资源调度请求时,基于各设备的权重和所调度的资源总量,确定对各设备资源的调度量。

[0029] 可选的,所述权重调整模块,用于:

[0030] 获取对所述第一设备的预设权重,判断所述第一设备的资源利用率是否大于或等于所述预设权重;

[0031] 若是,则将所述第一设备的权重调整为第一权重;

[0032] 若否,则计算所述第一设备的资源利用率与所述资源平均利用率的差值,结合所述预设权重,得到第二权重,进而将所述第一设备的权重调整为所述第二权重。

[0033] 可选的,还包括服务提供模块,用于:

[0034] 获取单个设备的IP地址和预设权重并上传至数据中心,以通过所述数据中心将单个设备的数据传输至资源调度模块;其中,所述数据包括IP地址、资源利用率和预设权重;

[0035] 资源调度模块,用于:

[0036] 所述资源调度模块根据调度结果,将设备的IP地址传输至发送所述资源调度请求的第二设备,以使所述第二设备通过传输的IP地址进行资源调度。

[0037] 可选的,所述服务提供模块,用于:所述资源调度模块在监测到所述数据中心内设备的数据发生变化时,从所述数据中心获取设备变化后的数据。

[0038] 可选的,所述资源调度模块,还用于:获取各第二设备对资源的调度请求量,确定调度请求量之和为调度量的至少一个第二设备,进而建立设备与所述至少一个第二设备之间的映射关系。

- [0039] 可选的,所述资源调度模块,还用于:
- [0040] 按照所确定的对各设备资源的调度量进行排序,得到设备队列;
- [0041] 按照各第二设备对资源的调度请求量进行排序,得到第二设备队列;
- [0042] 按照轮训方式,将所述第二设备队列中的第二设备分配至所述设备队列中的设备,进而建立设备与第二设备之间的映射关系。
- [0043] 为实现上述目的,根据本发明实施例的再一方面,提供了一种资源调度电子设备。
- [0044] 本发明实施例的电子设备包括:一个或多个处理器;存储装置,用于存储一个或多个程序,当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现上述任一所述的资源调度方法。
- [0045] 为实现上述目的,根据本发明实施例的再一方面,提供了一种计算机可读介质,其上存储有计算机程序,所述程序被处理器执行时实现上述任一所述的资源调度方法。
- [0046] 根据本发明所述提供的方案,上述发明中的一个实施例具有如下优点或有益效果:在分布式集群环境下,根据现有集群CPU利用率的平均值,自动化调整利用率超过平均值的设备,如采用降低权重方法,且整体无需手动操作,实现整个集群CPU利用率的动态平衡。
- [0047] 上述的非惯用的可选方式所具有的进一步效果将在下文中结合具体实施方式加以说明。

## 附图说明

- [0048] 附图用于更好地理解本发明,不构成对本发明的不当限定。其中:
- [0049] 图1是根据本发明实施例的一种资源调度方法的主要流程示意图;
- [0050] 图2是根据本发明实施例的一种可选的资源调度方法的流程示意图;
- [0051] 图3是根据本发明实施例的一种资源调度方法框架示意图;
- [0052] 图4是根据本发明实施例的一种资源调度装置的主要模块示意图;
- [0053] 图5是本发明实施例可以应用于其中的示例性系统架构图;
- [0054] 图6是适于用来实现本发明实施例的移动设备或服务器的计算机系统的结构示意图。

## 具体实施方式

[0055] 以下结合附图对本发明的示范性实施例做出说明,其中包括本发明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,而不会背离本发明的范围和精神。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。

[0056] 参见图1,示出的是本发明实施例提供的一种资源调度方法的主要流程图,包括如下步骤:

[0057] S101:针对同一应用集群,统计预设周期内对单个设备的资源调度量,以确定单个设备的资源利用率;

[0058] S102:将资源利用率之和与设备数量的比值作为所述应用集群下设备的资源平均利用率,提取资源利用率大于或等于所述资源平均利用率的第一设备;

[0059] S103:按照预设方式对所述第一设备进行权重调整,以在接收到资源调度请求时,基于各设备的权重和所调度的资源总量,确定对各设备资源的调度量。

[0060] 上述实施方式中,对于步骤S101和S102,本发明主要应用于分布式集群环境下,用以实现设备CPU资源利用率的动态调度。

[0061] 控制中心设置有资源调度模块,该模块启动后,会首先计算整个应用集群下所有可用设备CPU利用率的平均值,具体地:

[0062] 1) 计算单个设备的CPU利用率,例如计算预设时间段/预设周期内对该设备CPU的调度量求得;

[0063] 2) 统计设备数量,即统一应用集群下的可用设备数量;

[0064] 3) 计算设备CPU利用率之和/可用设备数量,将所得商作为整个应用集群下设备的CPU平均利用率。

[0065] 以上可以设置定时任务,以周期性求取应用集群下设备的CPU平均利用率,例如每5分钟更新一次。

[0066] 根据计算所得CPU平均利用率,将应用集群下的设备归为两类,一类为CPU利用率大于等于CPU平均利用率的第一设备,另一类为剩余设备。考虑到CPU利用率过高可能会影响到整个应用集群的性能,因此本发明主要针对第一设备进行处理。

[0067] 对于步骤S103,资源调度模块周期性调整各设备对CPU调度量的权重,以实现动态调整CPU利用率的目的;其中,权重为资源调度的分配策略,权重越大意味着分配的资源调度请求量越多。

[0068] 判断第一设备的CPU利用率是否超出其预设权重,例如100%:

[0069] 1) 若有,则将其权重直接归位至第一权重,例如0,即表示该第一设备后续将不参与负载,无流量/资源调度分配;

[0070] 2) 若否,则根据预定算法进行权重调整,具体地:预设权重-(CPU利用率-CPU平均利用率) $\times$ 100%。

[0071] 而后,在接收到第二设备传输的资源调度请求后,资源分配模块根据重新调整好的权重对各设备进行资源调度量分配,调度量较小的设备CPU利用率会逐步下降。同时在达到下一个周期时,该应用集群下的所有设备参与下一次的权重调整,最终动态调整整体CPU利用率逐步降低到平均值以下。

[0072] 上述实施例所提供的方法,在分布式集群环境下,根据设备的CPU利用率动态调整整个应用集群下设备的负载权重,以避免部分设备CPU利用率过高,影响整个集群服务质量的情况。

[0073] 参见图2,示出了根据本发明实施例的一种可选的资源调度方法流程示意图,包括如下步骤:

[0074] S201:针对同一应用集群,统计预设周期内对单个设备的资源调度量,以确定单个设备的资源利用率;

[0075] S202:获取单个设备的IP地址和预设权重并上传至数据中心,以通过所述数据中心将设备的数据传输至资源调度模块;其中,所述数据包括IP地址、资源利用率和预设权重;

[0076] S203:将资源利用率之和与设备数量的比值作为所述应用集群下设备的资源平均

利用率,提取资源利用率大于或等于所述资源平均利用率的第一设备;

[0077] S204:获取对所述第一设备的预设权重,判断所述第一设备的资源利用率是否大于或等于所述预设权重;

[0078] S205:若是,则将所述第一设备的权重调整为第一权重;

[0079] S206:若否,则计算所述第一设备的资源利用率与所述资源平均利用率的差值,结合所述预设权重,得到第二权重,进而将所述第一设备的权重调整为所述第二权重

[0080] S207:在接收到资源调度请求时,基于各设备的权重和所调度的资源总量,确定对各设备资源的调度量;

[0081] S208:所述资源调度模块根据调度结果,将设备的IP地址传输至发送所述资源调度请求的第二设备,以使所述第二设备通过所传输的IP地址进行资源调度。

[0082] 上述实施方式中,对于步骤S201、S203~S207可参见图1所示步骤S101~S103的描述,在此不再赘述。

[0083] 上述实施方式中,对于步骤S202,本发明中所提及的设备,可视为是提供资源调度的服务端。服务提供模块(server)和资源计算模块可作为不同部件,部署在同一设备中。

[0084] 服务提供模块启动后,会主动获取设备的IP地址和预设权重(例如100%),并上传至分布式数据中心。在服务提供模块上传数据完毕后,自动触发资源计算模块的启动。资源计算模块启动后,会统计单个设备在预设周期内对CPU的调度量,进而计算得到CPU利用率(秒级和分钟级),并将该CPU利用率同步至分布式数据中心。

[0085] 需要说明的是,设备启动必须优于第二设备(发送资源调度请求),且需要在服务提供模块同步数据之后,才可以启动资源计算模块,否则易导致服务不可用的情况。

[0086] 与分布式数据中心关联的模块,除了设备中的服务提供模块和资源计算模块外,还包括资源调度模块,用以动态调整设备的CPU利用率/权重。

[0087] 资源调度模块启动后,订阅分布式数据中心内设备的数据(包括IP地址、CPU利用率以及预设断种),当监测到其数据发生变更时会主动从分布式数据中心拉取数据,以此实现数据同步的目的。

[0088] 需要说明的是,资源调度模块与分布式数据中心不是一个整体。分布式数据中心可视为是基于zookeeper(此处仅为示例)的控制中心,用于存储数据。

[0089] 另外,分布式数据中心与服务提供模块之间设有心跳监测,当监测到服务提供模块提供的数据状态异常时,会自动从数据中心剔除掉,再将删除的数据同步至资源调度模块。

[0090] 对于步骤S208,第二设备目的在于调度设备中的CPU资源,因此通常在设备启动之后再启动,也可以在设备启动前启动,并等待设备启动后再发送资源调度请求。

[0091] 对于所确定的服务于第二设备的设备,可以通过如下方式计算:

[0092] 1) 比对第二设备对资源的调度请求量与所确定的对设备的资源调度量,最终得出调度请求量之和的第二设备,例如,对设备a的资源调度量为100,而第二设备I对资源的调度请求量为50,第二设备II和第二设备III分别为20和50,最终得到设备a-第二设备I、第二设备II和第二设备III之间的映射关系;

[0093] 2) 顺序轮训方式:

[0094] ①首先根据对设备的资源调度量,建立设备队列,例如设备a、设备b、设备c、设备



d……

[0095] 根据第二设备对资源的调度请求量,建立第二设备队列,例如第二设备I、第二设备II、第二设备III……

[0096] ②按照轮训方式,将第二设备I分配至设备a中、第二设备II分配至设备b中、以此类推。

[0097] 需要说明的是,轮训时需要比对第二设备对资源的调度请求量,是否大于相应设备的资源调度量(或调度量-已分配的调度请求量),若大于,则只分配一部分,并将该第二设备移动至第二设备队列的队尾中,此时一个第二设备可对应于多个设备,两者呈现多对多的关系。

[0098] 3) 随机轮训方式

[0099] 与上述2) 区别之处在于,随机将第二设备分配至设备中,同时比对资源调度请求量和资源调度量。

[0100] 在确定设备与第二设备之间的映射关系后,将设备IP地址传输至相应第二设备中。第二设备只有在获取到设备IP地址后,才可以与设备进行通信,进而访问设备中的CPU资源。

[0101] 上述实施例所提供的方法,基于资源调度请求量之和、顺序轮训方式、随机轮训方式等,建立设备与第二设备之间的映射关系,用以实现设备中资源调度合理化。

[0102] 参见图3,示出了根据本发明实施例的一种资源调度框架示意图,包括分布式数据中心、设备、第二设备和资源调度模块:

[0103] 1) 设备,包含服务提供模块(server)和资源计算模块:

[0104] ①服务提供模块,主动获取设备的IP地址和预设权重并上传至分布式数据中心;

[0105] ②资源计算模块,响应于服务提供模块上传数据完毕,统计单个设备在预设周期内对CPU的调度量以计算得到CPU利用率,之后将该CPU利用率同步至分布式数据中心;

[0106] 2) 分布式数据中心:用以存储设备的数据,包括IP地址、预设权重和CPU利用率;

[0107] 3) 资源调度模块,与分布式数据中心关联,用以获取设备数据,并根据设备数据动态调整各设备的权重,以及确定设备与第二设备之间的映射关系;

[0108] 4) 第二设备,用以发送资源调度请求。

[0109] 本发明实施例所提供的方法,在分布式集群环境下,自动化调整CPU利用率过高的设备,采用降低权重的方式,以将整个集群中设备的CPU利用率达到动态平衡。

[0110] 参见图4,示出了本发明实施例提供的一种资源调度装置400的主要模块示意图,包括:

[0111] 资源计算模块401,用于针对同一应用集群,统计预设周期内对单个设备的资源调度量,以确定单个设备的资源利用率;

[0112] 设备提取模块402,用于将资源利用率之和与设备数量的比值作为所述应用集群下设备的资源平均利用率,提取资源利用率大于或等于所述资源平均利用率的第一设备;

[0113] 权重调整模块403,用于按照预设方式对所述第一设备进行权重调整,以在接收到资源调度请求时,基于各设备的权重和所调度的资源总量,确定对各设备资源的调度量。

[0114] 本发明实施装置中,所述权重调整模块403,用于:

[0115] 获取对所述第一设备的预设权重,判断所述第一设备的资源利用率是否大于或等

于所述预设权重；

[0116] 若是,则将所述第一设备的权重调整为第一权重；

[0117] 若否,则计算所述第一设备的资源利用率与所述资源平均利用率的差值,结合所述预设权重,得到第二权重,进而将所述第一设备的权重调整为所述第二权重。

[0118] 本发明实施装置还包括服务提供模块404(图中未标出),用于:

[0119] 获取单个设备的IP地址和预设权重并上传至数据中心,以通过所述数据中心将单个设备的数据传输至资源调度模块;其中,所述数据包括IP地址、资源利用率和预设权重;

[0120] 资源调度模块405(图中未标出),用于:

[0121] 所述资源调度模块根据调度结果,将设备的IP地址传输至发送所述资源调度请求的第二设备,以使所述第二设备通过传输的IP地址进行资源调度。

[0122] 本发明实施装置中,所述服务提供模块404,用于:

[0123] 所述资源调度模块在监测到所述数据中心内设备的数据发生变化时,从所述数据中心获取设备变化后的数据。

[0124] 本发明实施装置中,所述资源调度模块405,还用于:

[0125] 获取各第二设备对资源的调度请求量,确定调度请求量之和为调度量的至少一个第二设备,进而建立设备与所述至少一个第二设备之间的映射关系。

[0126] 本发明实施装置中,所述资源调度模块405,还用于:

[0127] 按照所确定的对各设备资源的调度量进行排序,得到设备队列;

[0128] 按照各第二设备对资源的调度请求量进行排序,得到第二设备队列;

[0129] 按照轮训方式,将所述第二设备队列中的第二设备分配至所述设备队列中的设备,进而建立设备与第二设备之间的映射关系。

[0130] 另外,在本发明实施例中所述装置的具体实施内容,在上面所述方法中已经详细说明了,故在此重复内容不再说明。

[0131] 图5示出了可以应用本发明实施例的示例性系统架构500。

[0132] 如图5所示,系统架构500可以包括终端设备501、502、503,网络504和服务器505(仅仅是示例)。网络504用以在终端设备501、502、503和服务器505之间提供通信链路的介质。网络504可以包括各种连接类型,例如有线、无线通信链路或者光纤电缆等等。

[0133] 用户可以使用终端设备501、502、503通过网络504与服务器505交互,以接收或发送消息等。终端设备501、502、503上可以安装有各种通讯客户端应用。

[0134] 终端设备501、502、503可以是具有显示屏并且支持网页浏览的各种电子设备,包括但不限于智能手机、平板电脑、膝上型便携计算机和台式计算机等等。

[0135] 服务器505可以是提供各种服务的服务器,例如对用户利用终端设备501、502、503所浏览的购物类网站提供支持的后台管理服务器(仅为示例)。

[0136] 需要说明的是,本发明实施例所提供的方法一般由服务器505执行,相应地,装置一般设置于服务器505中。

[0137] 应该理解,图5中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要,可以具有任意数目的终端设备、网络和服务器。

[0138] 下面参考图6,其示出了适于用来实现本发明实施例的终端设备的计算机系统600的结构示意图。图6示出的终端设备仅仅是一个示例,不应对本发明实施例的功能和使用范

围带来任何限制。

[0139] 如图6所示,计算机系统600包括中央处理单元(CPU)601,其可以根据存储在只读存储器(ROM)602中的程序或者从存储部分608加载到随机访问存储器(RAM)603中的程序而执行各种适当的动作和处理。在RAM 603中,还存储有系统600操作所需的各种程序和数据。CPU 601、ROM 602以及RAM 603通过总线604彼此相连。输入/输出(I/O)接口605也连接至总线604。

[0140] 以下部件连接至I/O接口605:包括键盘、鼠标等的输入部分606;包括诸如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分607;包括硬盘等的存储部分608;以及包括诸如LAN卡、调制解调器等的网络接口卡的通信部分609。通信部分609经由诸如因特网的网络执行通信处理。驱动器610也根据需要连接至I/O接口605。可拆卸介质611,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器610上,以便于从其上读出的计算机程序根据需要被安装入存储部分608。

[0141] 特别地,根据本发明公开的实施例,上文参考流程图描述的过程可以被实现为计算机软件程序。例如,本发明公开的实施例包括一种计算机程序产品,其包括承载在计算机可读介质上的计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中,该计算机程序可以通过通信部分609从网络上被下载和安装,和/或从可拆卸介质611被安装。在该计算机程序被中央处理单元(CPU)601执行时,执行本发明的系统中限定的上述功能。

[0142] 需要说明的是,本发明所示的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是一——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本发明中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本发明中,计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0143] 附图中的流程图和框图,图示了按照本发明各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,上述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要

注意的是,框图或流程图中的每个方框、以及框图或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0144] 描述于本发明实施例中所涉及到的模块可以通过软件的方式实现,也可以通过硬件的方式来实现。所描述的模块也可以设置在处理器中,例如,可以描述为:一种处理器包括资源计算模块、设备提取模块、权重调整模块。其中,这些模块的名称在某种情况下并不构成对该模块本身的限定,例如,权重调整模块还可以被描述为“设备权重调整模块”。

[0145] 作为另一方面,本发明还提供了一种计算机可读介质,该计算机可读介质可以是上述实施例中描述的设备中所包含的;也可以是单独存在,而未装配入该设备中。上述计算机可读介质承载有一个或者多个程序,当上述一个或者多个程序被一个该设备执行时,使得该设备包括:

[0146] 针对同一应用集群,统计预设周期内对单个设备的资源调度量,以确定单个设备的资源利用率;

[0147] 将资源利用率之和与设备数量的比值作为所述应用集群下设备的资源平均利用率,提取资源利用率大于或等于所述资源平均利用率的第一设备;

[0148] 按照预设方式对所述第一设备进行权重调整,以在接收到资源调度请求时,基于各设备的权重和所调度的资源总量,确定对各设备资源的调度量。

[0149] 根据本发明实施例的技术方案,在分布式集群环境下,根据现有集群CPU利用率的平均值,自动化调整利用率超过平均值的设备,如采用降低权重方法,且整体无需手动操作,实现整个集群CPU利用率的动态平衡。

[0150] 上述具体实施方式,并不构成对本发明保护范围的限制。本领域技术人员应该明白的是,取决于设计要求和因素,可以发生各种各样的修改、组合、子组合和替代。任何在本发明的精神和原则之内所作的修改、等同替换和改进等,均应包含在本发明保护范围之内。

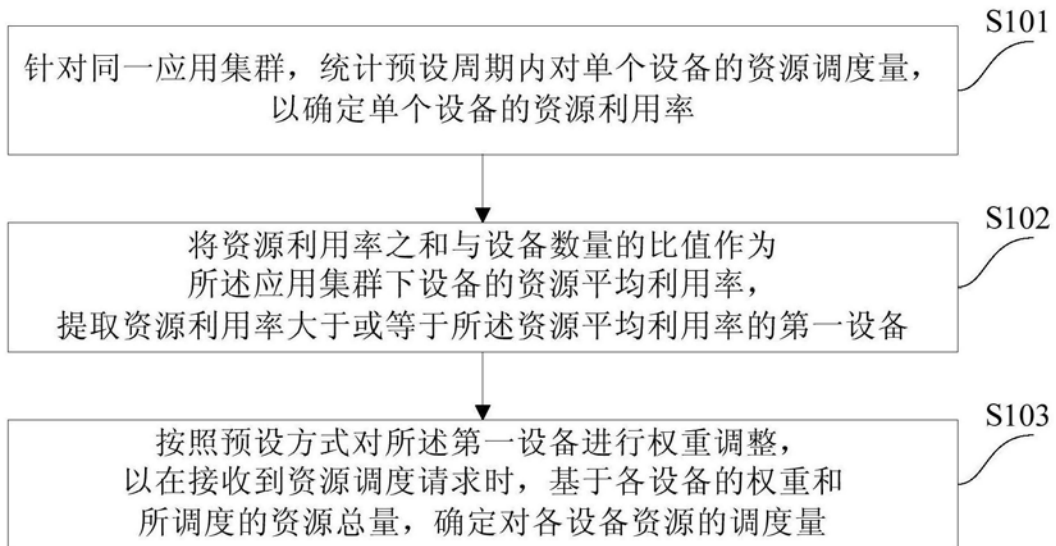


图1

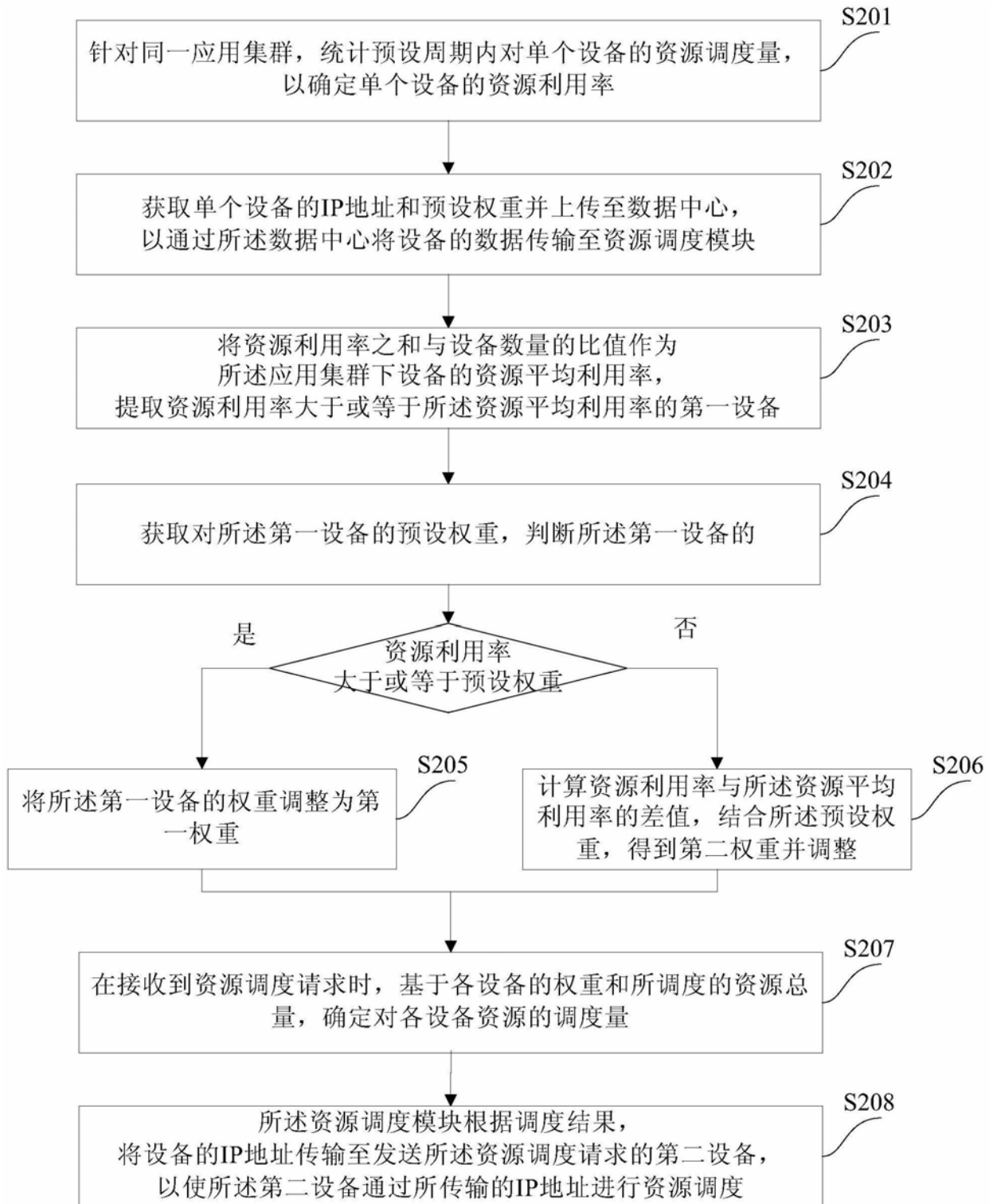


图2

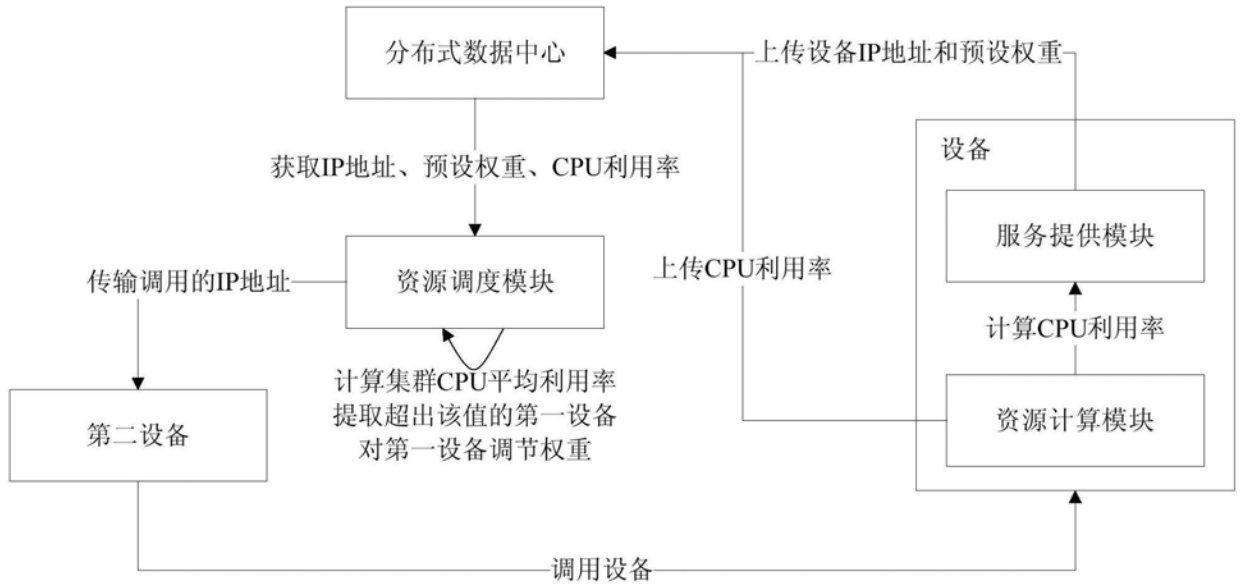


图3

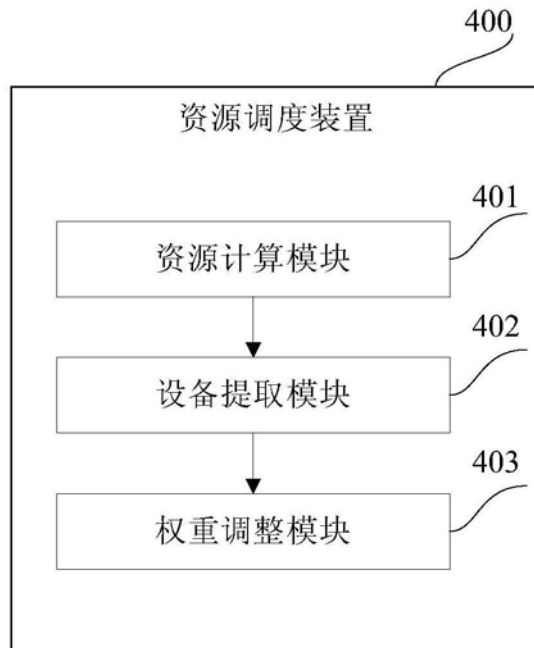


图4

500

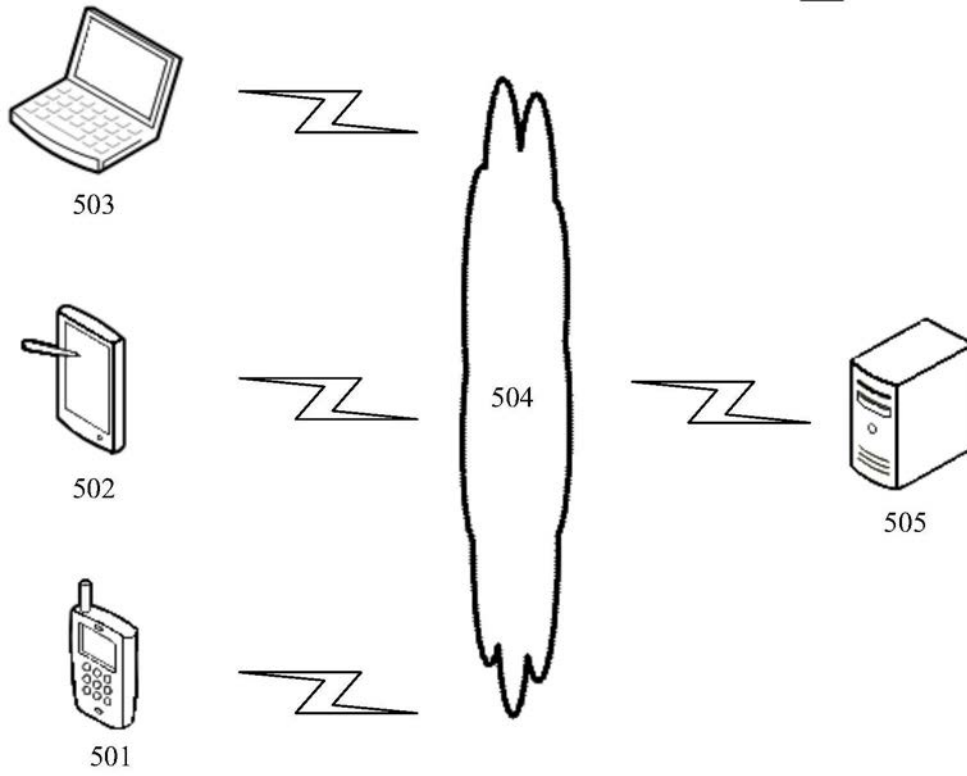


图5

600

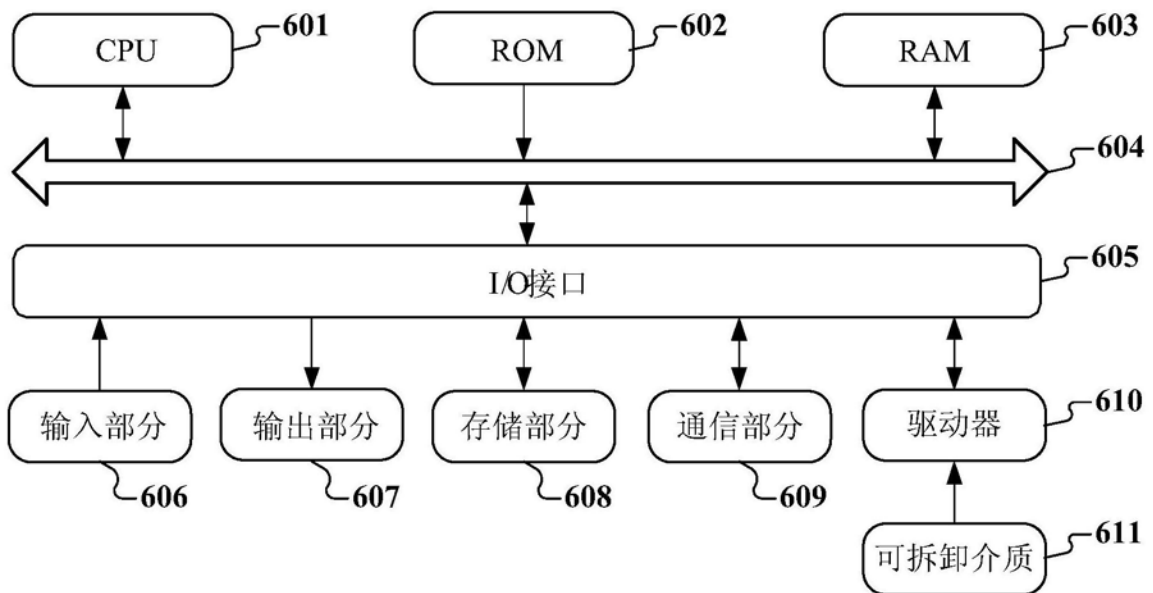


图6