



(12) 发明专利

(10) 授权公告号 CN 115086334 B

(45) 授权公告日 2024.05.28

(21) 申请号 202210685021.6

(22) 申请日 2022.06.15

(65) 同一申请的已公布的文献号
申请公布号 CN 115086334 A

(43) 申请公布日 2022.09.20

(73) 专利权人 北京奇艺世纪科技有限公司
地址 100080 北京市海淀区北一街2号爱奇艺
艺创新大厦10、11层

(72) 发明人 刘建康 张若麟 张晓婉

(74) 专利代理机构 北京集佳知识产权代理有限公司 11227
专利代理师 王宝筠

(51) Int. Cl.

H04L 67/1029 (2022.01)

H04L 67/1008 (2022.01)

H04L 67/1023 (2022.01)

(56) 对比文件

CN 109800125 A, 2019.05.24

CN 103516537 A, 2014.01.15

CN 105007312 A, 2015.10.28

CN 107172193 A, 2017.09.15

CN 109032324 A, 2018.12.18

CN 109471722 A, 2019.03.15

CN 111221651 A, 2020.06.02

CN 111481921 A, 2020.08.04

CN 111506423 A, 2020.08.07

CN 112306686 A, 2021.02.02

CN 113206876 A, 2021.08.03

CN 113301126 A, 2021.08.24

CN 113783922 A, 2021.12.10

CN 114237808 A, 2022.03.25

WO 2017190798 A1, 2017.11.09

WO 2022042724 A1, 2022.03.03

WO 2022048674 A1, 2022.03.10

审查员 赵潜

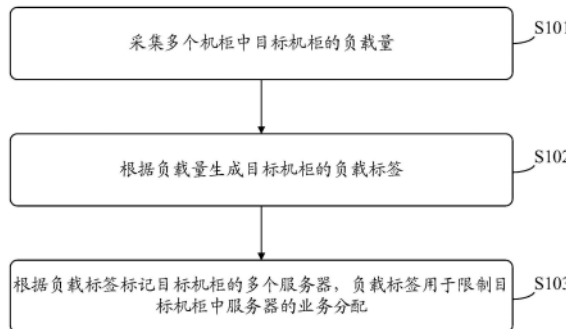
权利要求书2页 说明书7页 附图5页

(54) 发明名称

一种服务器标记方法及相关装置

(57) 摘要

本申请公开了一种服务器标记方法及相关装置,该方法应用于机房,机房包括多个机柜,每个机柜包括多个服务器,方法包括:采集多个机柜中目标机柜的负载量;根据负载量生成目标机柜的负载标签;根据负载标签标记目标机柜的多个服务器,负载标签用于限制目标机柜中服务器的业务分配。通过服务器所在的机柜的负载量生成负载标签。本申请实施例提供的标记方法可以通过服务器上的负载标签,可以在一定程度上避免机柜的成本过高,或机柜的整体功率过高导致或机柜的整体掉电,从而使得服务器达到较好的经济效益和较好的稳定性。



1. 一种服务器标记方法,其特征在于,所述方法应用于机房,所述机房包括多个机柜,每个机柜包括多个服务器,所述方法包括:

采集所述多个机柜中目标机柜的负载量;

根据所述负载量生成所述目标机柜的负载标签;所述负载标签指示所述目标机柜的负载等级;

根据所述负载标签标记所述目标机柜的多个服务器,所述负载标签用于限制所述目标机柜中服务器的业务分配;

当所述目标机柜的负载量低于第一参数时,所述负载标签为低负载标签;所述低负载标签指示所述目标机柜中的多个服务器被优先分配业务;

当所述目标机柜的负载量高于所述第一参数但低于第二参数时,所述负载标签为中负载标签;所述中负载标签指示当所述机房中不存在所述低负载标签的机柜时,所述目标机柜中的多个服务器将被分配业务;

当所述目标机柜的负载量高于所述第二参数但低于第三参数时,所述负载标签为高负载标签;所述高负载标签指示当所述机房中不存在所述低负载标签和所述中负载标签的机柜时,所述目标机柜中的多个服务器将被分配业务;

当所述目标机柜的负载量高于所述第三参数时,所述负载标签为危险负载标签;所述危险负载标签指示所述目标机柜具有掉电的风险,所述目标机柜中的服务器不再被分配业务。

2. 根据权利要求1所述的方法,其特征在于,当所述负载标签为高负载标签时,所述方法还包括:

向用户发送提示信息,所述提示信息用于提示用户分析所述目标机柜的业务的趋势。

3. 根据权利要求2所述的方法,其特征在于,当所述负载标签为危险负载标签时,所述方法还包括:

向用户发送报警信息,所述报警信息用于指示所述目标机柜具有掉电的风险。

4. 根据权利要求1-3任一项所述的方法,其特征在于,所述负载量包括机柜中多个服务器的总电流,和/或,机柜中多个服务器的总功率。

5. 根据权利要求1所述的方法,其特征在于,所述目标机柜的负载标签被周期性地更新,所述根据所述负载标签标记所述目标机柜的多个服务器,包括:

根据被更新后的负载标签标记所述目标机柜的多个服务器。

6. 根据权利要求1所述的方法,其特征在于,所述服务器由业务调度中心分配业务,所述方法还包括:

获得用户设置的目标业务对应的业务调度配置;

根据所述目标机柜的负载标签,给所述业务调度配置增加负载标签限制,并将包括所述负载标签的业务调度配置发送至所述业务调度中心。

7. 根据权利要求1所述的方法,其特征在于,所述机房包括中转机,所述根据所述负载量生成目标机柜的负载标签,包括:

所述中转机将所述机房中的目标机柜的负载量发送至数据分析中心;

所述数据分析中心根据所述目标机柜的负载量生成所述负载标签。

8. 一种服务器标记装置,其特征在於,所述装置应用于机房,所述机房包括多个机柜,每个机柜包括多个服务器,所述装置包括:

采集模块,用于采集所述多个机柜中目标机柜的负载量;所述负载量包括机柜中多个服务器的总电流或机柜中多个服务器的总功率;

生成模块,用于根据所述负载量生成所述目标机柜的负载标签;

标记模块,用于根据所述负载标签标记所述目标机柜的多个服务器,所述负载标签用于限制所述目标机柜中服务器的业务分配;

当所述目标机柜的负载量低于第一参数时,所述负载标签为低负载标签;所述低负载标签指示所述目标机柜中的多个服务器被优先分配业务;

当所述目标机柜的负载量高于所述第一参数但低于第二参数时,所述负载标签为中负载标签;所述中负载标签指示当所述机房中不存在所述低负载标签的机柜时,所述目标机柜中的多个服务器将被分配业务;

当所述目标机柜的负载量高于所述第二参数但低于第三参数时,所述负载标签为高负载标签;所述高负载标签指示当所述机房中不存在所述低负载标签和所述中负载标签的机柜时,所述目标机柜中的多个服务器将被分配业务;

当所述目标机柜的负载量高于所述第三参数时,所述负载标签为危险负载标签;所述危险负载标签指示所述目标机柜具有掉电的风险,所述目标机柜中的服务器不再被分配业务。

9. 一种计算机可读存储介质,其特征在於,所述计算机可读存储介质用于存储计算机程序,所述计算机程序用于执行权利要求1-7中任一项所述的方法。

一种服务器标记方法及相关装置

技术领域

[0001] 本申请涉及计算机领域,尤其涉及一种服务器标记方法及相关装置。

背景技术

[0002] Kubernetes (K8s) 是一个用于管理云平台中多个主机上的容器化的应用。在调度系统 (scheduler) 的作用下,可以将业务在各个服务器 (node) 间转移切换。K8s通过将业务分配到不同的服务器,使得服务器的负载较为均衡,从而可以提高设备利用率,降低系统故障。但目前的业务分配方式仅考虑了服务器本身的负载,使得服务器的耗能较高且稳定性较差。

发明内容

[0003] 为了解决上述技术问题,本申请提供了一种服务器标记方法及相关装置,通过服务器所在的机柜的负载量,对服务器进行标记,从而可以限制服务器的业务分配,使得服务器达到较好的经济效益和较好的稳定性。

[0004] 为了实现上述目的,本申请实施例提供的技术方案如下:

[0005] 本申请实施例提供一种服务器标记方法,所述方法应用于机房,所述机房包括多个机柜,每个机柜包括多个服务器,所述方法包括:

[0006] 采集所述多个机柜中目标机柜的负载量;

[0007] 根据所述负载量生成目标机柜的负载标签;所述负载标签指示所述目标机柜的负载等级;

[0008] 根据所述负载标签标记所述目标机柜的多个服务器,所述负载标签用于限制所述目标机柜中服务器的业务分配。

[0009] 作为一种可能的实施方式,当所述目标机柜的负载量低于第一参数时,所述负载标签为低负载标签;

[0010] 所述低负载标签指示所述目标机柜中的多个服务器被优先分配业务。

[0011] 作为一种可能的实施方式,当所述目标机柜的负载量高于第一参数但低于第二参数时,所述负载标签为中负载标签;

[0012] 所述中负载标签指示当所述机房中不存在低负载标签的机柜时,所述目标机柜中的多个服务器将被分配业务。

[0013] 作为一种可能的实施方式,当所述目标机柜的负载量高于第二参数但低于第三参数时,所述负载标签为高负载标签;

[0014] 所述高负载标签指示当所述机房中不存在低负载标签和中负载标签的机柜时,所述目标机柜中的多个服务器将被分配业务;

[0015] 当所述负载标签为高负载标签时,所述方法还包括:

[0016] 向用户发送提示信息,所述提示信息用于提示用户分析所述目标机柜的业务的的增长趋势。

[0017] 作为一种可能的实施方式,当所述目标机柜的负载量高于第三参数时,所述负载标签为危险负载标签;

[0018] 所述危险负载标签指示所述目标机柜具有掉电的风险,所述目标机柜中的服务器不再被分配业务;

[0019] 当所述负载标签为危险负载标签时,所述方法还包括:

[0020] 向用户发送报警信息,所述报警信息用于指示所述目标机柜具有掉电的风险。

[0021] 作为一种可能的实施方式,所述负载量包括机柜中多个服务器的总电流,和/或,机柜中多个服务器的总功率。

[0022] 作为一种可能的实施方式,所述目标机柜的负载标签被周期性地更新,所述根据所述负载标签标记所述目标机柜的多个服务器,包括:

[0023] 根据被更新后的负载标签标记所述目标机柜的多个服务器

[0024] 作为一种可能的实施方式,所述服务器由业务调度中心分配业务,所述方法还包括:

[0025] 获得用户设置的目标业务对应的业务调度配置;

[0026] 根据所述目标机柜的负载标签,给所述业务调度配置增加负载标签限制,并将包括所述负载标签的业务调度配置发送至所述业务调度中心。

[0027] 作为一种可能的实施方式,所述机房包括中转机,所述根据所述负载量生成目标机柜的负载标签,包括:

[0028] 所述中转机将所述机房中的目标机柜的负载量发送至数据分析中心;

[0029] 所述数据分析中心根据所述目标机柜的负载量生成所述负载标签。

[0030] 根据上述的服务器标记方法,本申请还提供了一种服务器标记装置,所述装置应用于机房,所述机房包括多个机柜,每个机柜包括多个服务器,所述装置包括:

[0031] 采集模块,用于采集所述多个机柜中目标机柜的负载量;所述负载量包括机柜中多个服务器的总电流或机柜中多个服务器的总功率;

[0032] 生成模块,用于根据所述负载量生成目标机柜的负载标签;

[0033] 标记模块,用于根据所述负载标签标记所述目标机柜的多个服务器,所述负载标签用于限制所述目标机柜中服务器的业务分配。

[0034] 本申请还提供了一种计算机可读存储介质,所述计算机可读存储介质用于存储计算机程序,所述计算机程序用于执行上述的方法。

[0035] 通过上述技术方案可知,本申请具有以下有益效果:

[0036] 本申请实施例提供了一种服务器标记方法,该方法应用于机房,机房包括多个机柜,每个机柜包括多个服务器,方法包括:采集多个机柜中目标机柜的负载量;根据负载量生成目标机柜的负载标签;根据负载标签标记目标机柜的多个服务器,负载标签用于限制目标机柜中服务器的业务分配。

[0037] 由此可知,本申请实施例提供的服务器标记方法,通过服务器所在的机柜的负载量生成负载标签,对服务器进行标记,从而使得在对该服务器进行业务分配时可以参考该负载标签。如此,本申请实施例提供的标记方法可以通过服务器上的负载标签,将业务整理分配到一个或者少量几个的机柜上,从而使得正在运行的机柜数量减少,可以在一定程度上避免机柜的成本过高,或机柜的整体功率过高导致或机柜的整体掉电,从而使得服务器

达到较好的经济效益和较好的稳定性。

附图说明

[0038] 为了更清楚地说明本申请实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图是本申请的一些实施例,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。

[0039] 图1为本申请实施例提供的一种服务器标记方法的流程图;

[0040] 图2为本申请实施例提供的一种服务器的标记方法的执行主体示意图;

[0041] 图3为本申请实施例提供的一种目标业务的业务调度配置示意图;

[0042] 图4为本申请实施例提供的另一种目标业务的业务调度配置示意图;

[0043] 图5为本申请实施例提供的一种机柜负载的示意图;

[0044] 图6为本申请实施例提供的一种服务器标记装置的示意图。

具体实施方式

[0045] 为了帮助更好地理解本申请实施例提供的方案,在介绍本申请实施例提供的方法之前,先介绍本申请实施例方案的应用的场景。

[0046] Kubernetes (K8s) 是一个用于管理云平台中多个主机上的容器化的应用。在调度系统(scheduler)的作用下,可以将业务在各个服务器(nodes)间转移切换。K8s通过将业务分配到不同的服务器,使得服务器的负载较为均衡,从而可以提高设备利用率,降低系统故障。但目前的业务分配方式仅考虑了服务器本身的负载,而没有考虑机柜的整体负载。本申请人发现,目前机柜运行的成本较高,而且当机柜整体用电量超过一定的功率,因机柜电气线路的标准问题,超过一定的功率机柜会意外掉电,导致该机柜上的多个服务器不可用,如果该机柜也给交换机路由器等网络设备供电,则影响的范围更广,造成重大灾难。因此,目前的业务调度方式并不能使服务器达到较好的经济效益和较好的稳定性。

[0047] 为了解决上述的技术问题,本申请实施例提供了一种服务器标记方法,该方法应用于机房,机房包括多个机柜,每个机柜包括多个服务器,方法包括:采集多个机柜中目标机柜的负载量;负载量包括机柜中多个服务器的总电流或机柜中多个服务器的总功率;根据负载量生成目标机柜的负载标签;根据负载标签标记目标机柜的多个服务器,负载标签用于限制目标机柜中服务器的业务分配。

[0048] 由此可知,本申请实施例提供的服务器标记方法,通过服务器所在的机柜的负载量生成负载标签,对服务器进行标记,从而使得在对该服务器进行业务分配时可以参考该负载标签。如此,本申请实施例提供的标记方法可以通过服务器上的负载标签,避免机柜的成本增加,或机柜的整体功率过高导致或机柜的整体掉电,从而使得服务器达到较好的经济效益和较好的稳定性。

[0049] 为使本申请的上述目的、特征和优点能够更加明显易懂,下面结合附图和具体实施方式对本申请实施例作进一步详细的说明。

[0050] 参见图1,该图为本申请实施例提供的一种服务器标记方法的流程图。

[0051] 如图1所示,本申请实施例提供的服务器标记方法应用于机房,机房包括多个机

柜,每个机柜包括多个服务器,该方法包括:

[0052] S101:采集多个机柜中目标机柜的负载量;

[0053] S102:根据负载量生成目标机柜的负载标签。

[0054] S103:根据负载标签标记目标机柜的多个服务器,负载标签用于限制目标机柜中服务器的业务分配。

[0055] 需要说明的是,本申请实施例中的负载标签用于指示所述目标机柜的负载等级,不同负载等级的机柜对应不同的负载标签。负载量包括机柜中多个服务器的总电流,和/或,机柜中多个服务器的总功率。本申请实施例中的目标机柜可以包括一个机柜,也可以包括多个机柜,本申请实施例在此不做限定。当目标机柜包括多个机柜时,可以根据多个机柜的负载量生成多个机柜的负载标签,并根据机柜对应的负载标签标记该机柜。

[0056] 在实际的应用中,如果当前的业务量较少,此时如果将业务分配到多个不同机柜上,使得多个机柜都在以较低的功率运行,这导致整个机房的整体经济效益较低,即机房的运行成本相对于业务量较低。而本申请实施例提供的标记方法可以通过服务器上的负载标签,将业务整理分配到一个或者少量几个的机柜上,从而使得正在运行的机柜数量减少,进而减少机房的运行成本,从而使得服务器达到较好的经济效益。

[0057] 在实际的应用中,如果当前的业务量较多,此时如果将业务分配到负载较高的机柜上,会导致该机柜的整体功率过高。此时因机柜电气线路的标准问题,超过一定的功率该机柜会发生意外掉电,导致该机柜上的多个服务器不可用。而本申请实施例提供的标记方法可以通过服务器上的负载标签,避免机柜的功率过高的情况发生,从而可以避免机柜因为功率过高而产生掉电,提高了机柜中服务器运行的稳定性。

[0058] 本申请实施例中的负载标签可以用于指示目标机柜的负载。在实际的应用中,根据机柜的负载大小可以将机柜分为多个负载等级,每个负载等级对应一个负载标签。作为一个示例,可以通过机柜中多个服务器的总电流或总功率划分负载等级。为了更好地理解本申请提供的方案,下面将通过机柜中多个服务器的总电流或总功率划分等级的方案作为示例来介绍本申请的方案。

[0059] 作为一个示例,当目标机柜的负载量低于第一参数阈值时,目标机柜的负载标签为低负载标签(power-low);低负载标签指示目标机柜中的多个服务器被优先分配业务。当目标机柜的负载量高于第一参数阈值但低于第二参数阈值时,目标机柜的负载标签为中负载标签(power-mid);中负载标签指示当不存在低负载标签的服务器时,目标机柜中的多个服务器将被分配业务。

[0060] 当目标机柜的负载量高于第二参数阈值但低于第三参数阈值时,目标机柜的负载标签为高负载标签(power-high);高负载标签指示当不存在低负载标签和中负载标签的服务器时,目标机柜中的多个服务器将被分配业务。当出现高负载标签的服务器时,本申请实施例提供的方法还可以向用户发送提示信息,从而让用户分析业务的增长趋势,做长周期的预测,制定设备的采购计划,合理分配采购的预算,进行采购等相关工作。

[0061] 当目标机柜的负载量高于第三参数阈值时,负载标签为危险负载标签(power-danger);危险负载标签指示目标机柜具有掉电的风险,目标机柜中的服务器不再被分配业务。当出现危险负载标签的服务器时,本申请实施例提供的方法还可以向用户发送报警信息,通知用户对目标机柜进行紧急处理。

[0062] 需要说明的是,本申请实施例可以将机柜分为四个等级,也可以将机柜分为三个等级或者五个等级,甚至更多个等级,本申请实施例在此不做限定。

[0063] 由于机柜的负载量会随时时间而进行变化,为了更加准确地标记机柜中的服务器,本申请实施例提供的服务器标记方法可以周期性地采集目标机柜的负载量,并根据采集得到的负载量周期性地更新目标机柜的负载标签,然后根据被更新后的负载标签标记目标机柜的多个服务器。作为一个示例,本申请实施例中的周期可以为1分钟。需要说明的是,本申请实施例中的周期可以为固定周期,也可以根据服务器的实际运行情况进行调整。例如在服务器的业务较为繁忙时,缩短负载标签更新的周期;在服务器的业务较为稀少时,拉长负载标签更新的周期。

[0064] 参见图2,该图为本申请实施例提供的一种服务器的标记方法的执行主体示意图。

[0065] 如图2所示,本申请实施例中机房中的中转机将机房中机柜(例如机柜1、机柜2和机柜3)的负载量上传至采集中心。需要说明的是,中转机可以为机房中的一台服务器,中转机在将机房中机柜的负载量上传至采集中心时,也可以将机柜的负载量保存到本机。采集中心可以连接多个机房,采集中心在接收到机房上传的负载量后,对负载量做数据分析,以及数据规范化,标准化,统一单位,去除噪点等处理。

[0066] 然后采集中心将处理后的负载量发送至数据分析中心,数据分析中心根据目标机柜的负载量生成负载标签,并根据负载标签对机柜中的服务器进行标记。然后数据分析中心可以获得用户设置的目标业务对应的业务调度配置;根据目标机柜的负载标签,给业务调度配置增加负载标签限制,并将增加后的业务调度配置发送至业务调度中心。作为一个示例,可以给目标业务的业务调度配置修改或增加“affinity”配置项内容,此内容和用户设置的固定标签取交集,联合生效。具体地,给业务调度配置增加负载标签限制时,可以根据遍历调度系统中,用户预设的配置信息,优先使用power-low标签,当符合power-low的node全部被分配完之后,则配置次优先标签power-mid,依次类推,直到能给目标业务分配到服务器为止。

[0067] 作为一个示例,目标业务的业务调度配置可以包含指示该业务分配至“power-low”标签的机柜中的服务器的代码。业务调度中心在分配该业务时,就会根据该业务的业务配置将该目标业务就会被分配至具有“power-low”标签的服务器。如果机房中存在多个具有“power-low”标签的服务器,业务调度中心可以将该业务随机分配到其中的一个服务器上,但如果该业务的业务配置中还存在其他服务器限制信息,例如服务器的型号或服务器本身的负载,则业务调度中心将根据其他的服务器限制信息,在多个具有“power-low”标签的服务器的服务器中选择一个目标服务器分配该业务。

[0068] 下面通过两个附图进行示例性的介绍。

[0069] 参见图3,该图为本申请实施例提供的一种目标业务的业务调度配置示意图。

[0070] 参见图4,该图为本申请实施例提供的另一种目标业务的业务调度配置示意图。

[0071] 如图3所示,目标业务配置没有增加负载标签限制时的示意图。如图4所示,目标业务配置以及增加了负载标签限制,即增加了“affinity”配置项内容。在“affinity”配置项内容包括“values:-power-low”限定该业务将被分配到具有“power-low”的负载标签的机柜中的服务器。在增加了负载标签限制后,目标业务会更倾向于被分配到负载标签等级低的机柜中的服务器,从而避免机柜的整体功率过高导致的机柜的整体掉电。

[0072] 为了更好地理解,本申请实施例所提供的方案所带来的效果,下面将通过一个示例来进行介绍。

[0073] 参见图5,该图为本身实施例提供的一种机柜负载的示意图。

[0074] 如图5所示,假设node-A-01和node-B-01服务器是固态硬盘(Solid State Disk或Solid State Drive,简称SSD)速度非常快,但是价格非常贵,容量比较小,适合做小数据量的高速关系型数据库主机;机柜上其他的服务器是普通硬盘,速度比较慢,但是容量非常大,价格非常便宜,适合做全局存储,存储海量的图片,视频等。这里的固态硬盘只是为了举例说明,不同的设备有不同的特性,适用于不同的业务场景。

[0075] 假设某业务S需要使用SSD硬盘的设备才能更好工作,则该业务的部分配置文件如图2所示:disktype:ssd键值,表示业务S只能选择SSD硬盘的node。当其业务扩大,需要再增加一个Pod以扩大对用户的服务能力时,目前调度系统有两个node可以选择。机柜A上共10台设备,机柜整体电量功率已经使用了90%。node-A-01无负荷。

[0076] 机柜B上10台node是10%负载,机柜整体电量功率已经使用了10%。此时业务调度中心需要在node-A-1,与node-B-1之间选择时,大概率会选择无负荷的node-A-1,而这是一个很坏的调度选择。而通过本身实施例提供的服务器标记方法,采集机柜电量并分析后,通过给node打上标签,例如power-low,power-mid,power-high,power-danger,标记node所处的机柜整体功率为低,中,高,危险的标签,例如,将机柜A的负载标签设置为power-high,机柜B的负载标签设置为power-low,并修改业务的配置文件使得业务被指示分配给在负载等级更低的机柜B中的服务器。业务调度中心就可以将业务S分配至node-B-1,从而避免机柜A整体功率过高导致的机柜A的意外掉电。

[0077] 综上所述,本申请实施例提供的服务器标记方法,通过服务器所在的机柜的负载量生成负载标签,对服务器进行标记,从而使得在对该服务器进行业务分配时可以参考该负载标签。如此,本申请实施例提供的标记方法可以通过服务器上的负载标签,可以避免机柜的成本增加,或机柜的整体功率过高导致或机柜的整体掉电,从而使得服务器达到较好的经济效益和较好的稳定性。

[0078] 根据上述实施例提供的服务器标记方法,本申请实施例还提供了一种服务器标记装置。

[0079] 参见图6,该图为本申请实施例提供的一种服务器标记装置的示意图。

[0080] 如图6所示,本申请实施例提供的服务器标记装置应用于机房,机房包括多个机柜,每个机柜包括多个服务器,该装置包括:

[0081] 采集模块100,用于采集多个机柜中目标机柜的负载量;负载量包括机柜中多个服务器的总电流或机柜中多个服务器的总功率。

[0082] 生成模块200,用于根据负载量生成目标机柜的负载标签。

[0083] 标记模块300,用于根据负载标签标记目标机柜的多个服务器,负载标签用于限制目标机柜中服务器的业务分配。

[0084] 本申请实施例提供的服务器标记装置,通过服务器所在的机柜的负载量生成负载标签,对服务器进行标记,从而使得在对该服务器进行业务分配时可以参考该负载标签。如此,本申请实施例提供的标记装置可以通过服务器上的负载标签,可以避免机柜的成本过高,或机柜的整体功率过高导致或机柜的整体掉电,从而使得服务器达到较好的经济效益

和较好的稳定性。

[0085] 根据上述的服务器标记方法和服务器标记装置,本申请实施例还提供了一种计算机可读存储介质,计算机可读存储介质用于存储计算机程序,计算机程序用于执行上述的服务器标记方法。

[0086] 通过以上的实施方式的描述可知,本领域的技术人员可以清楚地了解到上述实施例方法中的全部或部分步骤可借助软件加必需的通用硬件平台的方式来实现。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品可以存储在存储介质中,如ROM/RAM、磁碟、光盘等,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者诸如媒体网关等网络通信设备,等等)执行本申请各个实施例或者实施例的某些部分所述的方法。

[0087] 需要说明的是,本说明书中各个实施例采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似部分互相参见即可。对于实施例公开的方法而言,由于其与实施例公开的系统相对应,所以描述的比较简单,相关之处参见系统部分说明即可。

[0088] 还需要说明的是,在本文中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0089] 对所公开的实施例的上述说明,使本领域专业技术人员能够实现或使用本申请对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的,本文中所定义的一般原理可以在不脱离本申请的精神或范围的情况下,在其它实施例中实现。因此,本申请将不会被限制于本文所示的这些实施例,而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

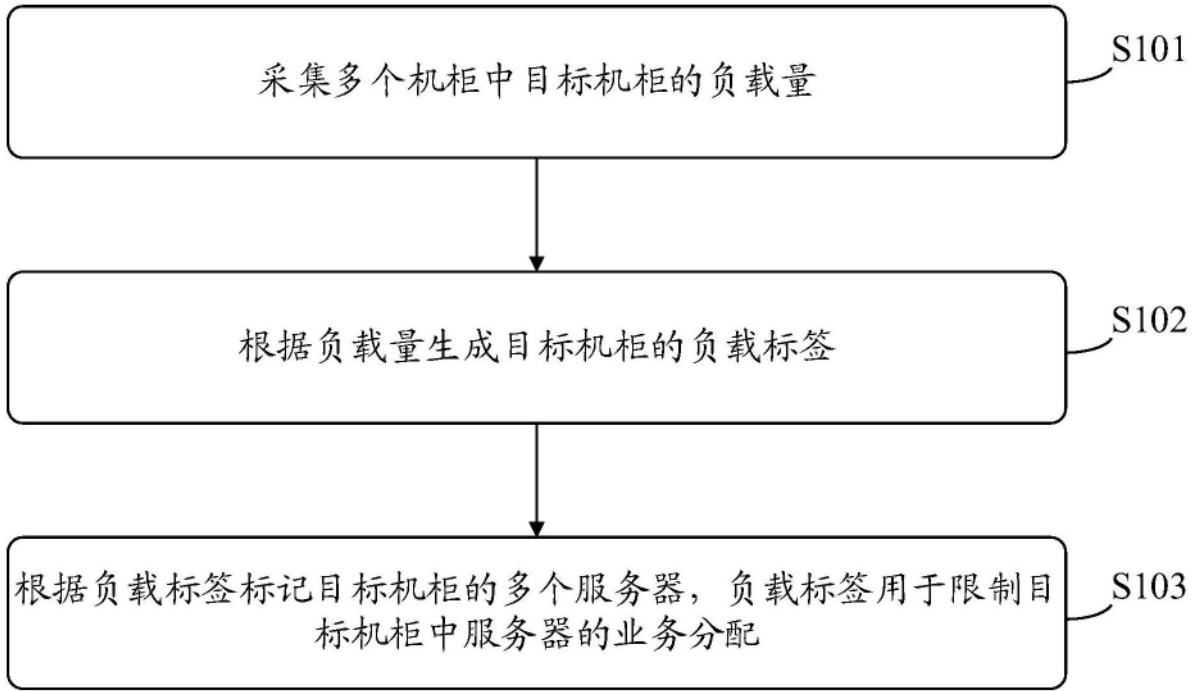


图1

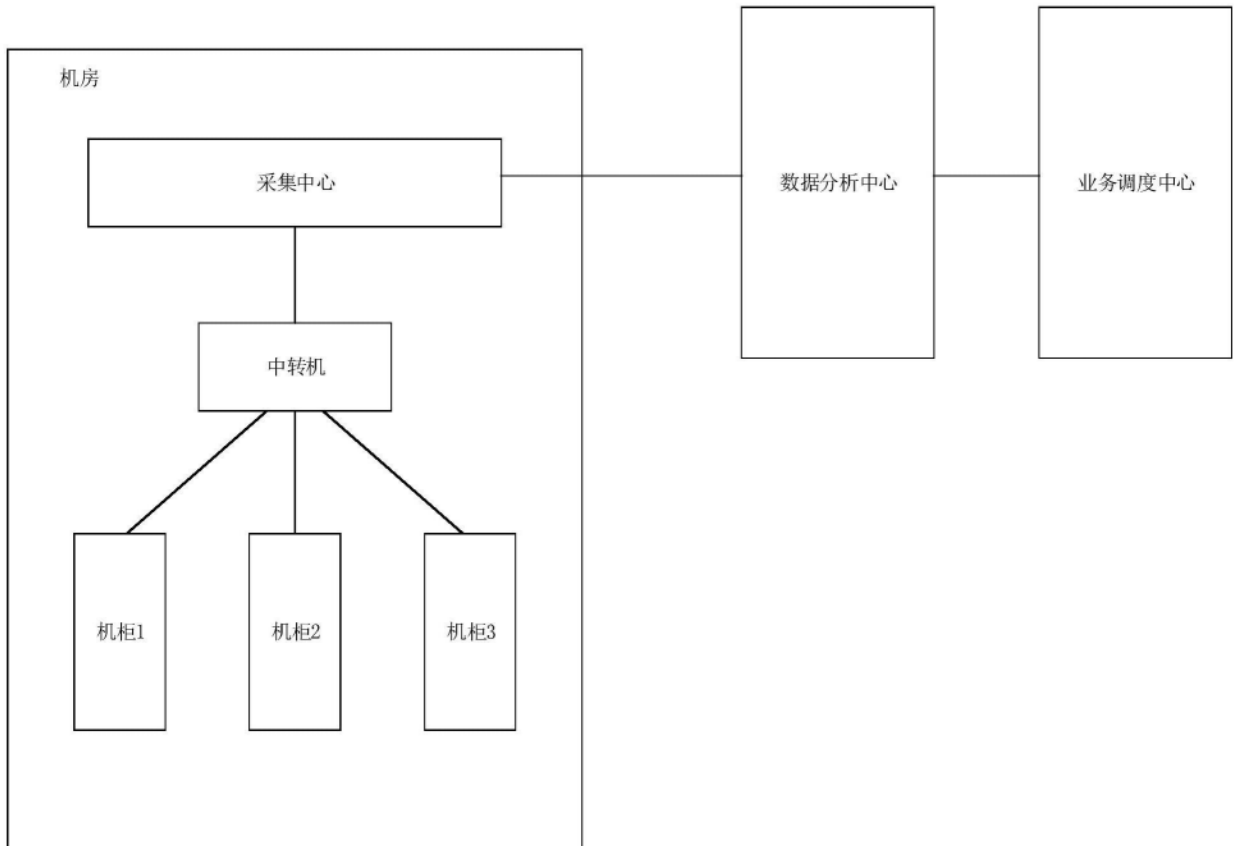


图2

```
apiVersion: v1
kind: Pod
metadata:
  name: mysql-ssd
labels:
  env: test
spec:
  containers:
  - name: mysql
    image: mysql
  nodeSelector:
    disktype: ssd
```

图3

```
apiVersion: v1
kind: Pod
metadata:
  name: mysql-ssd
  labels:
    env: test
spec:
  containers:
  - name: mysql
    image: mysql
  nodeSelector:
    disktype: ssd
  affinity:
    nodeAffinity:
      requiredDuringSchedulingIgnoredDuringExecution:
        nodeSelectorTerms:
        - matchExpressions:
          - key: iqiyi.com/rack-power
            operator: In
            values:
            - power-low
```

图4

node-A-01 空负载	node-B-01 10%负载
node-A-02 满负载	node-B-02 10%负载
node-A-03 满负载	node-B-03 10%负载
node-A-04 满负载	node-B-04 10%负载
node-A-05 满负载	node-B-05 10%负载
node-A-06 满负载	node-B-06 10%负载
node-A-07 满负载	node-B-07 10%负载
node-A-08 满负载	node-B-08 10%负载
node-A-09 满负载	node-B-09 10%负载
node-A-10 满负载	node-B-10 10%负载

机柜A-负载90%

机柜B-负载10%

图5



图6