



(12)发明专利

(10)授权公告号 CN 105786587 B

(45)授权公告日 2019.11.26

(21)申请号 201410814195.3

(22)申请日 2014.12.23

(65)同一申请的已公布的文献号
申请公布号 CN 105786587 A

(43)申请公布日 2016.07.20

(73)专利权人 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 邹韬

(74)专利代理机构 北京同达信恒知识产权代理有限公司 11291

代理人 冯艳莲

(51)Int.Cl.
G06F 9/455(2006.01)

(56)对比文件

US 2011029970 A1,2011.02.03,说明书第[0004]-[0028]段.

CN 102932413 A,2013.02.13,全文.

US 2013179574 A1,2013.07.11,说明书第[0042],[0056]-[0062]段,附图2-7.

审查员 杨昕瑞

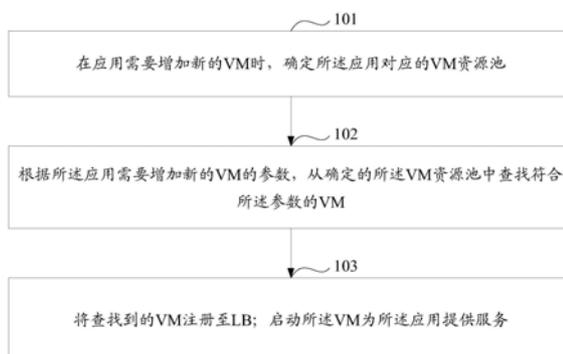
权利要求书3页 说明书11页 附图4页

(54)发明名称

一种虚拟机VM的伸缩方法和设备

(57)摘要

本发明公开了一种虚拟机VM的伸缩方法和设备,内容包括:在应用需要增加新的VM时,确定所述应用对应的VM资源池;根据所述应用需要增加新的VM的参数,从确定的所述VM资源池中查找符合所述参数的VM;将查找到的所述VM注册至负载均衡器LB;启动所述VM为所述应用提供服务,也就是说,在应用需要增加新的VM时,在预先创建的VM资源池中选择符合应用需求的VM,注册并为该应用提供服务,省略了创建虚拟机的流程,缩短了虚拟机创建的时间,有效地提升了虚拟机的伸缩效率。



1. 一种虚拟机VM的伸缩方法,其特征在于,包括:

在应用需要增加新的虚拟机VM时,确定所述应用对应的VM资源池;其中,在应用需要增加新的虚拟机VM时,确定所述应用对应的VM资源池,具体包括:在应用需要增加新的VM时,根据预先建立的实例化的应用标识与VM资源池的之间的对应关系和需要增加新的VM的应用的标识,确定所述应用的标识对应的VM资源池;

根据所述应用需要增加新的VM的参数,从确定的所述VM资源池中查找符合所述参数的VM;

将查找到的所述VM注册至负载均衡器LB;

启动所述VM为所述应用提供服务;

所述方法还包括:

判断所述应用对应的VM资源池中剩余VM的个数是否与预设的VM的个数相同;

若所述应用对应的VM资源池中剩余VM的个数小于预设的VM的个数时,确定需要增加的VM的个数;根据所述应用的参数或者预设的VM配置规则,确定需要增加的每一个VM的参数;利用确定的所述需要增加的每一个VM的参数,创建需要增加的每一个VM;

若所述应用对应的VM资源池中剩余VM的个数大于预设的VM的个数时,确定需要删除的VM的个数,并根据确定的删除VM的个数,从VM资源池中选择VM,删除选择的VM。

2. 如权利要求1所述的伸缩方法,其特征在于,根据所述应用需要增加新的VM的参数,从确定的所述VM资源池中查找符合所述参数的VM,包括:

在从确定的所述VM资源池中查找符合所述参数的VM时,若确定的所述VM资源池中不存在符合所述参数的VM,进一步查找确定的所述VM资源池中是否存在空闲的VM;

若查找结果是存在空闲的VM,那么根据所述应用需要增加新的VM的参数,对查找到的所述空闲的VM的参数进行修改;

将修改后的所述空闲的VM作为查找到的且符合所述参数的VM。

3. 如权利要求1至2任一所述的伸缩方法,其特征在于,通过以下方式创建应用对应的VM资源池,包括:

计算应用对应的VM资源池中需要配置VM的个数;

根据所述应用的参数或者预设的VM配置规则,确定所述VM资源池中每一个VM的参数;

利用确定的所述每一个VM的参数,创建所述VM资源池中每一个VM。

4. 如权利要求3所述的伸缩方法,其特征在于,计算应用对应的VM资源池中需要配置VM的个数,包括:

根据应用的资源需求与设定的VM个数的比例之间的映射关系,计算满足所述应用的资源需求所需要配置的VM的个数。

5. 如权利要求3所述的伸缩方法,其特征在于,根据所述应用的参数或者预设的VM配置规则,确定所述VM资源池中每一个VM的参数,包括:

根据所述应用的参数,确定所述应用对应的所述VM资源池中每一个VM的参数,其中,所述VM资源池中任意两个VM的参数相同;

或者,

根据预设的VM配置规则,确定多个应用对应的VM资源池;并

根据所述多个应用的参数,确定所述VM资源池中VM的参数,其中,所述VM资源池中至少

两个VM的参数不相同。

6. 如权利要求1、2、4或5所述的伸缩方法,其特征在于,所述方法还包括:

在所述应用确定负载减小时,从LB中去注册VM;

在确定所述应用对应的VM资源池支持VM回收时,将去注册的VM重新移入所述应用对应的VM资源池。

7. 如权利要求6所述的伸缩方法,其特征在于,将去注册的VM重新移入所述应用对应的VM资源池,包括:

在确定去注册的VM的参数与所述应用对应的VM资源池所要求的VM的参数不一致时,按照所述应用对应的VM资源池所要求的VM的参数,对去注册的VM的参数进行修改;

将修改后的该VM重新移入所述应用对应的VM资源池。

8. 一种虚拟机VM的伸缩设备,其特征在于,包括:

确定模块,用于在应用需要增加新的虚拟机VM时,确定所述应用对应的VM资源池;所述确定模块,具体用于:在应用需要增加新的VM时,根据预先建立的实例化的应用标识与VM资源池的之间的对应关系和需要增加新的VM的应用的标识,确定所述应用的标识对应的VM资源池;

查找模块,用于根据所述应用需要增加新的VM的参数,从确定的所述VM资源池中查找符合所述参数的VM;

启动模块,用于将查找到的所述VM注册至负载均衡器LB;启动所述VM为所述应用提供服务;

所述伸缩设备还包括:计算模块,其中:

计算模块,用于判断所述应用对应的VM资源池中剩余VM的个数是否与预设的VM的个数相同;

若所述应用对应的VM资源池中剩余VM的个数小于预设的VM的个数时,确定需要增加的VM的个数;根据所述应用的参数或者预设的VM配置规则,确定需要增加的每一个VM的参数;利用确定的所述需要增加的每一个VM的参数,创建需要增加的每一个VM;

若所述应用对应的VM资源池中剩余VM的个数大于预设的VM的个数时,确定需要删除的VM的个数,并根据确定的删除VM的个数,从VM资源池中选择VM,删除选择的VM。

9. 如权利要求8所述的伸缩设备,其特征在于,

所述查找模块,具体用于在从确定的所述VM资源池中查找符合所述参数的VM时,若确定的所述VM资源池中不存在符合所述参数的VM,进一步查找确定的所述VM资源池中是否存在空闲的VM;

若查找结果是存在空闲的VM,那么根据所述应用需要增加新的VM的参数,对查找到的所述空闲的VM的参数进行修改;

将修改后的所述空闲的VM作为查找到的且符合所述参数的VM。

10. 如权利要求8至9任一所述的伸缩设备,其特征在于,所述伸缩设备还包括:创建模块,其中:

所述创建模块,用于通过以下方式创建应用对应的VM资源池,包括:

计算应用对应的VM资源池中需要配置VM的个数;

根据所述应用的参数或者预设的VM配置规则,确定所述VM资源池中每一个VM的参数;

利用确定的所述每一个VM的参数,创建所述VM资源池中每一个VM。

11. 如权利要求10所述的伸缩设备,其特征在于,

所述创建模块,具体用于根据应用的资源需求与设定的VM个数的比例之间的映射关系,计算满足所述应用的资源需求所需要配置的VM的个数。

12. 如权利要求10所述的伸缩设备,其特征在于,

所述创建模块,具体用于根据所述应用的参数,确定所述应用对应的所述VM资源池中每一个VM的参数,其中,所述VM资源池中任意两个VM的参数相同;

或者,

根据预设的VM配置规则,确定多个应用对应的VM资源池;并

根据所述多个应用的参数,确定所述VM资源池中VM的参数,其中,所述VM资源池中至少两个VM的参数不相同。

13. 如权利要求8、9、11或12所述的伸缩设备,其特征在于,所述伸缩设备还包括:收缩模块,其中:

所述收缩模块,用于在所述应用确定负载减小时,从LB中去注册VM;在确定所述应用对应的VM资源池支持VM回收时,将去注册的VM重新移入所述应用对应的VM资源池。

14. 如权利要求13所述的伸缩设备,其特征在于,

所述收缩模块,具体用于在确定去注册的VM的参数与所述应用对应的VM资源池所要求的VM的参数不一致时,按照所述应用对应的VM资源池所要求的VM的参数,对去注册的VM的参数进行修改;

将修改后的该VM重新移入所述应用对应的VM资源池。

一种虚拟机VM的伸缩方法和设备

技术领域

[0001] 本发明涉及计算机技术领域,尤其涉及一种虚拟机VM的伸缩方法和设备。

背景技术

[0002] 作为自互联网革命以来IT(英文:Information Technology;中文:信息科技和产业)产业最深刻的变革,云计算实现了对业务、对底层资源的屏蔽。VM(英文:Virtual Manufacturing;中文:虚拟机)伸缩是指根据虚拟机的负载,对虚拟机的内部参数或者虚拟机的个数进行修改,以满足应用对资源的需求。

[0003] 例如:应用A由虚拟机B和虚拟机C通过负载均衡方式对其进行服务。通常情况下,虚拟机B和虚拟机C提供的资源能够满足应用A的要求,但是,在业务量增加时,虚拟机B和虚拟机C提供的资源将无法满足不同应用A的要求,此时需要执行VM伸缩,为应用A新增加一个虚拟机D,这样,由虚拟机B、虚拟机C和虚拟机D共同提供资源为应用A服务。这里增加一个虚拟机D就是VM伸缩的方式。

[0004] 但是,目前增加虚拟机D的方式主要包括以下步骤:

[0005] 第一步:虚拟化管理系统从虚拟机B和虚拟机C中获取应用A的参数,并根据所述应用A的参数,确定伸缩规则。

[0006] 例如:应用A的参数为CPU占用达到80%,根据CPU占用达到80%与设定的伸缩规则之间的对应关系,得到CPU占用达到80%对应的伸缩规则,例如:增加一个虚拟机。

[0007] 第二步:虚拟化管理系统利用选择的伸缩规则,创建虚拟机D。

[0008] 其中,虚拟机D以虚拟机B和虚拟机C作为源虚拟机进行创建,得到的虚拟机D与虚拟机B和虚拟机C的参数一致。

[0009] 第三步:虚拟化管理系统启动虚拟机D,并注册创建的虚拟机D,实现虚拟机B、虚拟机C和虚拟机D共同提供资源为应用A服务。

[0010] 但是,目前实现虚拟机伸缩的方式耗时比较长,降低了虚拟机伸缩效率,不能满足应用业务的需求。

发明内容

[0011] 有鉴于此,本发明实施例提供了一种虚拟机VM的伸缩方法和设备,用以解决目前虚拟机伸缩耗时较长,伸缩效率较低的问题。

[0012] 第一方面,提供了一种虚拟机VM的伸缩方法,包括:

[0013] 在应用需要增加新的VM时,确定所述应用对应的VM资源池;

[0014] 根据所述应用需要增加新的VM的参数,从确定的所述VM资源池中查找符合所述参数的VM;

[0015] 将查找到的所述VM注册至LB;

[0016] 启动所述VM为所述应用提供服务。

[0017] 结合第一方面,在第一种可能的实施方式中,根据所述应用需要增加新的VM的参

数,从确定的所述VM资源池中查找符合所述参数的VM,包括:

[0018] 在从确定的所述VM资源池中查找符合所述参数的VM时,若确定的所述VM资源池中
没有符合所述参数的VM,进一步查找确定的所述VM资源池中是否存在空闲的VM;

[0019] 若查找结果是存在空闲的VM,那么根据所述应用需要增加新的VM的参数,对查找
到的所述空闲的VM的参数进行修改;

[0020] 将修改后的所述空闲的VM作为查找到的且符合所述参数的VM。

[0021] 结合第一方面,或者结合第一方面的第一种可能的实施方式,在第二种可能的实
施方式中,通过以下方式创建应用对应的VM资源池,包括:

[0022] 计算应用对应的VM资源池中需要配置VM的个数;

[0023] 根据所述应用的参数或者预设的VM配置规则,确定所述VM资源池中每一个VM的参
数;

[0024] 利用确定的所述每一个VM的参数,创建所述VM资源池中每一个VM。

[0025] 结合第一方面的第二种可能的实施方式,在第三种可能的实施方式中,计算应用
对应的VM资源池中需要配置VM的个数,包括:

[0026] 根据应用的资源需求与设定的VM个数的比例之间的映射关系,计算满足所述应用
的资源需求所需要配置的VM的个数。

[0027] 结合第一方面的第二种可能的实施方式,或者结合第一方面的第三种可能的实施
方式,在第四种可能的实施方式中,根据所述应用的参数或者预设的VM配置规则,确定所述
VM资源池中每一个VM的参数,包括:

[0028] 根据所述应用的参数,确定所述应用对应的所述VM资源池中每一个VM的参数,其
中,所述VM资源池中任意两个VM的参数相同;

[0029] 或者,

[0030] 根据预设的VM配置规则,确定多个应用对应的VM资源池;并

[0031] 根据所述多个应用的参数,确定所述VM资源池中VM的参数,其中,所述VM资源池
中至少两个VM的参数不相同。

[0032] 结合第一方面,或者结合第一方面的第一种可能的实施方式,或者结合第一方面
的第二种可能的实施方式,或者结合第一方面的第三种可能的实施方式,或者结合第一方面
的第四种可能的实施方式,在第五种可能的实施方式中,所述方法还包括:

[0033] 在所述应用确定负载减小时,从LB中去注册VM;

[0034] 在确定所述应用对应的VM资源池支持VM回收时,将去注册的VM重新移入所述应用
对应的VM资源池。

[0035] 结合第一方面的第五种可能的实施方式,在第六种可能的实施方式中,将去注册
的VM重新移入所述应用对应的VM资源池,包括:

[0036] 在确定去注册的VM的参数与所述应用对应的VM资源池所要求的VM的参数不一致
时,按照所述应用对应的VM资源池所要求的VM的参数,对去注册的VM的参数进行修改;

[0037] 将修改后的该VM重新移入所述应用对应的VM资源池。

[0038] 结合第一方面,或者结合第一方面的第一种可能的实施方式,或者结合第一方面
的第二种可能的实施方式,或者结合第一方面的第三种可能的实施方式,或者结合第一方面
的第四种可能的实施方式,或者结合第一方面的第五种可能的实施方式,或者结合第一

方面的第六种可能的实施方式,在第七种可能的实施方式中,所述方法还包括:

[0039] 判断所述应用对应的VM资源池中剩余VM的个数是否与预设的VM的个数相同;

[0040] 若所述应用对应的VM资源池中剩余VM的个数小于预设的VM的个数时,确定需要增加的VM的个数;

[0041] 根据所述应用的参数或者预设的VM配置规则,确定需要增加的每一个VM的参数;

[0042] 利用确定的所述需要增加的每一个VM的参数,创建需要增加的每一个VM。

[0043] 第二方面,提供了一种虚拟机VM的伸缩设备,包括:

[0044] 确定模块,用于在应用需要增加新的VM时,确定所述应用对应的VM资源池;

[0045] 查找模块,用于根据所述应用需要增加新的VM的参数,从确定的所述VM资源池中查找符合所述参数的VM;

[0046] 启动模块,用于将查找到的所述VM注册至LB;启动所述VM为所述应用提供服务。

[0047] 结合第二方面,在第一种可能的实施方式中,所述查找模块,具体用于在从确定的所述VM资源池中查找符合所述参数的VM时,若确定的所述VM资源池中不存在符合所述参数的VM,进一步查找确定的所述VM资源池中是否存在空闲的VM;

[0048] 若查找结果是存在空闲的VM,那么根据所述应用需要增加新的VM的参数,对查找到的所述空闲的VM的参数进行修改;

[0049] 将修改后的所述空闲的VM作为查找到的且符合所述参数的VM。

[0050] 结合第二方面,或者结合第二方面的第一种可能的实施方式,在第二种可能的实施方式中,所述伸缩设备还包括:创建模块,其中:

[0051] 所述创建模块,用于通过以下方式创建应用对应的VM资源池,包括:

[0052] 计算应用对应的VM资源池中需要配置VM的个数;

[0053] 根据所述应用的参数或者预设的VM配置规则,确定所述VM资源池中每一个VM的参数;

[0054] 利用确定的所述每一个VM的参数,创建所述VM资源池中每一个VM。

[0055] 结合第二方面的第二种可能的实施方式,在第三种可能的实施方式中,所述创建模块,具体用于根据应用的资源需求与设定的VM个数的比例之间的映射关系,计算满足所述应用的资源需求所需要配置的VM的个数。

[0056] 结合第二方面的第二种可能的实施方式,或者结合第二方面的第三种可能的实施方式,在第四种可能的实施方式中,所述创建模块,具体用于根据所述应用的参数,确定所述应用对应的所述VM资源池中每一个VM的参数,其中,所述VM资源池中任意两个VM的参数相同;

[0057] 或者,

[0058] 根据预设的VM配置规则,确定多个应用对应的VM资源池;并

[0059] 根据所述多个应用的参数,确定所述VM资源池中VM的参数,其中,所述VM资源池中至少两个VM的参数不相同。

[0060] 结合第二方面,或者结合第二方面的第一种可能的实施方式,或者结合第二方面的第二种可能的实施方式,或者结合第二方面的第三种可能的实施方式,或者结合第二方面的第四种可能的实施方式,在第五种可能的实施方式中,所述伸缩设备还包括:收缩模块,其中:

[0061] 所述收缩模块,用于在所述应用确定负载减小时,从LB中去注册VM;在确定所述应用对应的VM资源池支持VM回收时,将去注册的VM重新移入所述应用对应的VM资源池。

[0062] 结合第二方面的第五种可能的实施方式,在第六种可能的实施方式中,所述收缩模块,具体用于在确定去注册的VM的参数与所述应用对应的VM资源池所要求的VM的参数不一致时,按照所述应用对应的VM资源池所要求的VM的参数,对去注册的VM的参数进行修改;

[0063] 将修改后的该VM重新移入所述应用对应的VM资源池。

[0064] 结合第二方面,或者结合第二方面的第一种可能的实施方式,或者结合第二方面的第二种可能的实施方式,或者结合第二方面的第三种可能的实施方式,或者结合第二方面的第四种可能的实施方式,或者结合第二方面的第五种可能的实施方式,或者结合第二方面的第六种可能的实施方式,在第七种可能的实施方式中,所述伸缩设备还包括:计算模块,其中:

[0065] 计算模块,用于判断所述应用对应的VM资源池中剩余VM的个数是否与预设的VM的个数相同;

[0066] 若所述应用对应的VM资源池中剩余VM的个数小于预设的VM的个数时,确定需要增加的VM的个数;

[0067] 根据所述应用的参数或者预设的VM配置规则,确定需要增加的每一个VM的参数;

[0068] 利用确定的所述需要增加的每一个VM的参数,创建需要增加的每一个VM。

[0069] 本发明有益效果如下:

[0070] 本发明实施例在应用需要增加新的VM时,确定所述应用对应的VM资源池;根据所述应用需要增加新的VM的参数,从确定的所述VM资源池中查找符合所述参数的VM;将查找到的所述VM注册至LB;启动所述VM为所述应用提供服务,也就是说,在应用需要增加新的VM时,在预先创建的VM资源池中选择符合应用需求的VM,注册并为该应用提供服务,省略了创建虚拟机的流程,缩短了虚拟机创建的时间,有效地提升了虚拟机的伸缩效率。

附图说明

[0071] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简要介绍,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,对于本领域的普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0072] 图1为本发明实施例提供的一种VM的伸展方法的流程示意图;

[0073] 图2为本发明实施例提供的一种VM的收缩方法的流程示意图;

[0074] 图3为本发明实施例提供的一种VM资源池的创建方法的流程示意图;

[0075] 图4为本发明实施例提供的一种VM的伸缩设备的结构示意图;

[0076] 图5为本发明实施例提供的一种VM的伸缩设备的结构示意图。

具体实施方式

[0077] 为了实现本发明的目的,本发明实施例提供了一种虚拟机VM的伸缩方法和设备,在应用需要增加新的VM时,确定该应用对应的VM资源池;根据该应用需要增加新的VM的参数,从确定的VM资源池中查找符合该参数的VM;将查找到的所述VM注册至负载均衡器(英

文:load balancer,缩写:LB);启动该VM为所述应用提供服务,也就是说,在应用需要增加新的VM时,在预先创建的VM资源池中选择符合应用需求的VM,注册并为该应用提供服务,省略了创建虚拟机的流程,缩短了虚拟机创建的时间,有效地提升了虚拟机的伸缩效率。

[0078] 下面结合说明书附图对本发明各个实施例作进一步地详细描述。显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,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它实施例,都属于本发明保护的范围。

[0079] 图1为本发明实施例提供的一种VM的伸展方法的流程示意图。所述方法可以如下所述。

[0080] 本发明实施例提出在虚拟资源规划阶段确定应用的VM预留规则;在应用实例化阶段,根据VM预留规则完成对应用的VM资源池的初始化;在应用运行阶段需要增加VM时,从该应用对应的VM资源池中选择符合该应用需要的VM,注册并启动选择的VM,缩短了VN创建的流程,有效地提升了虚拟机的伸缩效率。

[0081] 步骤101:在应用需要增加新的VM时,确定所述应用对应的VM资源池。

[0082] 在步骤101中,由于在应用的实例化阶段,根据VM预留规则完成对应用的VM资源池的初始化,可以为每一个实例化的应用初始化对应的VM资源池,即实例化的应用与VM资源池之间满足一对一的关系,那么在初始化VM资源池时,建立实例化的应用标识与VM资源池之间的对应关系;还可以为多个实例化的应用初始化对应的VM资源池,即实例化的应用与VM资源池之间满足多对一的关系,也就是说一个VM资源池为多个应用所共用,那么在初始化VM资源池时,建立多个实例化的应用标识与VM资源池之间的对应关系。

[0083] 在应用需要增加新的VM时,根据预先建立的实例化的应用标识与VM资源池的之间的对应关系和需要增加新的VM的应用的标识,确定所述应用的标识对应的VM资源池。

[0084] 此外,在确定所述应用对应的VM资源池时,还可以根据预先建立的实例化的应用标识与VM资源池的之间的对应关系,判断所述应用对应的VM资源池是否为多个应用共用的VM资源池。

[0085] 步骤102:根据所述应用需要增加新的VM的参数,从确定的所述VM资源池中查找符合所述参数的VM。

[0086] 在步骤102中,对于步骤101确定的所述应用对应的VM资源池,获取该VN资源池中每一个VM的参数。

[0087] 将该VM资源池中每一个VM的参数与所述应用需要增加新的VM的参数进行比较。

[0088] 若比较结果是该VM资源池中存在至少一个VM的参数与所述应用需要增加新的VM的参数相同,那么确定该VM资源池中存在符合所述应用需要增加新的VM的参数的VM,触发执行步骤103。

[0089] 若比较结果是该VM资源池中不存在与所述应用需要增加新的VM的参数相同的VM,那么确定该VM资源池中不存在符合所述应用需要增加新的VM的参数的VM,进一步查找确定的所述VM资源池中是否存在空闲的VM。

[0090] 若进一步查找结果是存在空闲的VM,那么根据所述应用需要增加新的VM的参数,对查找到的所述空闲的VM的参数进行修改。

[0091] 将修改后的所述空闲的VM作为查找到的且符合所述参数的VM。

[0092] 需要说明的是,本发明实施例中描述的VM的参数可以是指VM的虚拟资源,例如:中央处理器(英文:central processing unit;缩写:CPU)资源参数、计算资源参数、内存资源参数、存储资源参数等等。

[0093] 不同的VM在被创建时,可以采用统一的参数进行创建;也可以采用不同的参数进行创建。

[0094] VM在被创建时,可以根据应用的不同,选择不同的参数进行创建;也可以根据不同的VM配置规则,确定创建的VM的参数,这里不做具体限定。

[0095] 步骤103:将查找到的VM注册至LB;启动所述VM为所述应用提供服务。

[0096] 在步骤103中,在查找到满足条件的VM时,在VM资源池中对该VM进行标注,表明该VM已被使用。

[0097] 此时,应用对应的新增VM的伸展操作结束。

[0098] 通过本发明实施例的方案,在应用需要增加新的VM时,在预先创建的VM资源池中选择符合应用需求的VM,注册并为该应用提供服务,省略了创建虚拟机的流程,缩短了虚拟机创建的时间,有效地提升了虚拟机的伸缩效率。

[0099] 图2为本发明实施例提供的一种VM的收缩方法的流程示意图。

[0100] 本发明实施例可以在上述图1所描述的VM伸展的基础之上,在应用负载减小时,触发执行VM的收缩方法。所述方法可以如下所述。

[0101] 步骤201:在所述应用确定负载减小时,触发执行VM的收缩操作。

[0102] 在步骤201中,所述应用确定负载减小的方法包括但不限于:

[0103] 当所述应用的CPU的使用率小于设定阈值时,确定负载减小,不需要启动过多的VM,此时需要释放一部分VM,即触发执行VM的收缩操作。

[0104] 步骤202:从LB中去注册VM。

[0105] 在步骤202中,在从LB中去注册VM时,停止应用。

[0106] 步骤203:确定所述应用对应的VM资源池,并判断确定的所述VM资源池是否支持VM回收;若支持,则触发执行步骤204;否则,直接删除去注册的VM。

[0107] 步骤204:将去注册的VM重新移入所述应用对应的VM资源池。

[0108] 在步骤204中,在将去注册的VM重新移入所述应用对应的VM资源池,具体包括:

[0109] 首先,停止VM的工作。

[0110] 其次,判断确定去注册的VM的参数与所述应用对应的VM资源池所要求的VM的参数是否一致。

[0111] 若确定去注册的VM的参数与所述应用对应的VM资源池所要求的VM的参数一致时,则将去注册的VM重新移入所述应用对应的VM资源池。

[0112] 若确定去注册的VM的参数与所述应用对应的VM资源池所要求的VM的参数不一致时,则按照所述应用对应的VM资源池所要求的VM的参数,对去注册的VM的参数进行修改;将修改后的该VM重新移入所述应用对应的VM资源池。

[0113] 这里需要说明的是,将去注册的VM重新移入所述应用对应的VM资源池可以是将标注该VM已被使用的标识去掉即可。

[0114] 由于在VM收缩的过程中,有些VM资源池不支持回收被使用的VM,那么在VM资源池中的VM被使用之后,将被直接删除,这样将导致VM资源池中的VM的数量减少;有些VM资源池

支持回收被使用的VM,那么在VM资源池中的VM被使用之后,被使用的VM将重新移入VM资源池,这样将导致VM资源池中的VM的数量增加,因此,为了保证VM资源池中VM数量保持稳定,可选地,本发明实施例在完成VM的伸缩操作之后,所述方法还包括:

[0115] 判断所述应用对应的VM资源池中剩余VM的个数是否与预设的VM的个数相同。

[0116] 若所述应用对应的VM资源池中剩余VM的个数小于预设的VM的个数时,确定需要增加的VM的个数;根据所述应用的参数或者预设的VM配置规则,确定需要增加的每一个VM的参数;利用确定的所述每一个VM的参数,创建需要增加的每一个VM。

[0117] 若所述应用对应的VM资源池中剩余VM的个数大于预设的VM的个数时,确定需要删除的VM的个数,并根据确定的删除VM的个数,从VM资源池中选择VM,删除选择的VM。

[0118] 通过本发明实施例的方案,在应用的负载发生变化需要减少VM时,触发VM的伸缩操作,将需要减少的VM回收至VM资源池中,以保证后续继续使用。

[0119] 图3为本发明实施例提供的一种VM资源池的创建方法的流程示意图。本发明实施例中VM资源池的创建发生在应用实例化阶段,根据VM预留规则完成VM资源池的初始化。所述方法可以如下所述。

[0120] 步骤301:根据VM预留规则,确定应用对应的VM资源池的类型以及大小。

[0121] 在步骤301中,根据应用执行业务的类型,确定该应用对应的VM资源池是专用VM资源池还是共用VM资源池。

[0122] 所谓专用VM资源池是指由该应用唯一使用,即建立该应用与该专用VM资源池之间一对一关系;

[0123] 所谓共用VM资源池是指由多个应用共同使用,即建立应用与该共用VM资源池之间的多对一关系。

[0124] 根据VM预留规则,计算应用对应的VM资源池中需要配置VM的个数。

[0125] 具体地,根据应用的资源需求与设定的VM个数的比例之间的映射关系,计算满足所述应用的资源需求所需要配置的VM的个数。

[0126] 例如:根据预设的应用的资源需求与设定的配置VM的百分比关系和该应用的资源需求,计算满足该应用的资源需求所需要配置的VM的个数。

[0127] 或者,根据VM预留规则中设定的VM的个数,为该应用对应的VM资源池中配置设定个数的VM。

[0128] 步骤302:根据所述应用的参数或者预设的VM配置规则,确定所述VM资源池中每一个VM的参数。

[0129] 在步骤302中,根据所述应用的参数,确定所述VM资源池中每一个VM的参数。

[0130] 若确定该应用对应的VM资源池由该应用专用,那么在确定所述VM资源池中每一个VM的参数时,可以确定该VM资源池中包括的VM的参数相同。

[0131] 也就是说,该应用对应的VM资源池中包括的VM是统一规格的,即VM的参数都相同。

[0132] 或者,根据预设的VM配置规则,确定多个应用对应的VM资源池,并根据所述多个应用的参数,确定所述VM资源池中VM的参数。

[0133] 若根据预设的VM配置规则,确定一个VM资源池由多个应用使用,那么对于这样的VM资源池中VM的参数,可以根据该多个应用的参数进行确定。

[0134] 例如:根据一个应用的参数,确定一组VM的参数,这样,在该VM资源池中,这一组VM

的参数可以相同；此外，VM资源池中的其他VM的参数可以通过其他应用的参数确定，这样，通过不同应用的参数确定的VM的参数可以不同。

[0135] 步骤303：利用确定的所述每一个VM的参数，创建所述VM资源池中每一个VM。

[0136] 通过本发明实施例的方案，在应用需要增加新的VM时，在预先创建的VM资源池中选择符合应用需求的VM，注册并为该应用提供服务，省略了创建虚拟机的流程，缩短了虚拟机创建的时间，有效地提升了虚拟机的伸缩效率。

[0137] 图4为本发明实施例提供的一种VM的伸缩设备的结构示意图。所述伸缩设备包括：确定模块41、查找模块42和启动模块43，其中：

[0138] 确定模块41，用于在应用需要增加新的VM时，确定所述应用对应的VM资源池；

[0139] 查找模块42，用于根据所述应用需要增加新的VM的参数，从确定的所述VM资源池中查找符合所述参数的VM；

[0140] 启动模块43，用于将查找到的所述VM注册至LB；启动所述VM为所述应用提供服务。

[0141] 具体地，所述查找模块42，具体用于在从确定的所述VM资源池中查找符合所述参数的VM时，若确定的所述VM资源池中不存在符合所述参数的VM，进一步查找确定的所述VM资源池中是否存在空闲的VM；

[0142] 若查找结果是存在空闲的VM，那么根据所述应用需要增加新的VM的参数，对查找到的所述空闲的VM的参数进行修改；

[0143] 将修改后的所述空闲的VM作为查找到的且符合所述参数的VM。

[0144] 可选地，所述伸缩设备还包括：创建模块44，其中：

[0145] 所述创建模块44，用于通过以下方式创建应用对应的VM资源池，包括：

[0146] 计算应用对应的VM资源池中需要配置VM的个数；

[0147] 根据所述应用的参数或者预设的VM配置规则，确定所述VM资源池中每一个VM的参数；

[0148] 利用确定的所述每一个VM的参数，创建所述VM资源池中每一个VM。

[0149] 具体地，所述创建模块44，具体用于根据应用的资源需求与设定的VM个数的比例之间的映射关系，计算满足所述应用的资源需求所需要配置的VM的个数。

[0150] 具体地，所述创建模块44，具体用于根据所述应用的参数，确定所述应用对应的所述VM资源池中每一个VM的参数，其中，所述VM资源池中任意两个VM的参数相同；

[0151] 或者，

[0152] 根据预设的VM配置规则，确定多个应用对应的VM资源池；并

[0153] 根据所述多个应用的参数，确定所述VM资源池中VM的参数，其中，所述VM资源池中至少两个VM的参数不相同。

[0154] 可选地，所述伸缩设备还包括：收缩模块45，其中：

[0155] 所述收缩模块45，用于在所述应用确定负载减小时，从LB中去注册VM；在确定所述应用对应的VM资源池支持VM回收时，将去注册的VM重新移入所述应用对应的VM资源池。

[0156] 具体地，所述收缩模块45，具体用于在确定去注册的VM的参数与所述应用对应的VM资源池所要求的VM的参数不一致时，按照所述应用对应的VM资源池所要求的VM的参数，对去注册的VM的参数进行修改；

[0157] 将修改后的该VM重新移入所述应用对应的VM资源池。

[0158] 可选地,所述伸缩设备还包括:计算模块46,其中:

[0159] 计算模块46,用于判断所述应用对应的VM资源池中剩余VM的个数是否与预设的VM的个数相同;

[0160] 若所述应用对应的VM资源池中剩余VM的个数小于预设的VM的个数时,确定需要增加的VM的个数;

[0161] 根据所述应用的参数或者预设的VM配置规则,确定需要增加的每一个VM的参数;

[0162] 利用确定的所述需要增加的每一个VM的参数,创建需要增加的每一个VM。

[0163] 需要说明的是,本发明实施例所述伸缩设备可以通过硬件方式实现,也可以通过软件方式实现,所述伸缩设备在应用需要增加新的VM时,在预先创建的VM资源池中选择符合应用需求的VM,注册并为该应用提供服务,省略了创建虚拟机的流程,缩短了虚拟机创建的时间,有效地提升了虚拟机的伸缩效率。

[0164] 图5为本发明实施提供的一种VM的伸缩设备的结构示意图。本发明实施例所述的伸缩设备可以具备上述的VM的伸缩功能。所述伸缩设备可以采用通用计算机系统结构,计算机系统可具体是基于处理器的计算机。所述路由器实体包括处理器51、接口52和存储器53。处理器51和接口52连接,并且和存储器53连接。例如可以通过总线连接处理器51、接口52和存储器53。其中:

[0165] 处理器51可以是中央处理器(英文:central processing unit,缩写:CPU),或者是CPU和硬件芯片的组合。

[0166] 接口52可以为以下一种或多种:提供有线接口的网络接口控制器(英文:network interface controller,缩写:NIC),例如以太网NIC,该以太网NIC可以提供铜线和/或光纤接口;提供无线接口的NIC,例如无线局域网(英文:wireless local area network,缩写:WLAN)NIC。

[0167] 存储器53用于存储程序代码,所述处理器51从存储器中获得存储的程序代码,按照获得的程序代理执行相应地处理。

[0168] 存储器53可以是易失性存储器(英文:volatile memory),例如随机存取存储器(英文:random-access memory,缩写:RAM);或者非易失性存储器(英文:non-volatile memory),例如只读存储器(英文:read-only memory,缩写:ROM),快闪存储器(英文:flash memory),硬盘(英文:hard disk drive,缩写:HDD)或固态硬盘(英文:solid-state drive,缩写:SSD);或者上述种类的存储器的组合。存储器53还可以包括内容可寻址存储器(英文:content-addressable memory,缩写:CAM)。

[0169] 具体地,所述处理器51执行所述存储器53中存放的程序,执行以下操作:

[0170] 所述处理器51在应用需要增加新的虚拟机VM时,确定所述应用对应的VM资源池;

[0171] 根据所述应用需要增加新的VM的参数,从确定的所述VM资源池中查找符合所述参数的VM;

[0172] 将查找到的所述VM注册至负载均衡器LB;

[0173] 启动所述VM为所述应用提供服务。

[0174] 所述处理器51根据所述应用需要增加新的VM的参数,从确定的所述VM资源池中查找符合所述参数的VM,包括:

[0175] 在从确定的所述VM资源池中查找符合所述参数的VM时,若确定的所述VM资源池中

没有符合所述参数的VM,进一步查找确定的所述VM资源池中是否存在空闲的VM;

[0176] 若查找结果是存在空闲的VM,那么根据所述应用需要增加新的VM的参数,对查找到的所述空闲的VM的参数进行修改;

[0177] 将修改后的所述空闲的VM作为查找到的且符合所述参数的VM。

[0178] 所述处理器51通过以下方式创建应用对应的VM资源池,包括:

[0179] 计算应用对应的VM资源池中需要配置VM的个数;

[0180] 根据所述应用的参数或者预设的VM配置规则,确定所述VM资源池中每一个VM的参数;

[0181] 利用确定的所述每一个VM的参数,创建所述VM资源池中每一个VM。

[0182] 所述处理器51计算应用对应的VM资源池中需要配置VM的个数,包括:

[0183] 根据应用的资源需求与设定的VM个数的比例之间的映射关系,计算满足所述应用的资源需求所需要配置的VM的个数。

[0184] 所述处理器51根据所述应用的参数或者预设的VM配置规则,确定所述VM资源池中每一个VM的参数,包括:

[0185] 根据所述应用的参数,确定所述应用对应的所述VM资源池中每一个VM的参数,其中,所述VM资源池中任意两个VM的参数相同;

[0186] 或者,

[0187] 根据预设的VM配置规则,确定多个应用对应的VM资源池;并

[0188] 根据所述多个应用的参数,确定所述VM资源池中VM的参数,其中,所述VM资源池中至少两个VM的参数不相同。

[0189] 所述处理器51,还用于在所述应用确定负载减小时,从LB中去注册VM;

[0190] 在确定所述应用对应的VM资源池支持VM回收时,将去注册的VM重新移入所述应用对应的VM资源池。

[0191] 所述处理器51将去注册的VM重新移入所述应用对应的VM资源池,包括:

[0192] 在确定去注册的VM的参数与所述应用对应的VM资源池所要求的VM的参数不一致时,按照所述应用对应的VM资源池所要求的VM的参数,对去注册的VM的参数进行修改;

[0193] 将修改后的该VM重新移入所述应用对应的VM资源池。

[0194] 所述处理器51,还用于判断所述应用对应的VM资源池中剩余VM的个数是否与预设的VM的个数相同;

[0195] 若所述应用对应的VM资源池中剩余VM的个数小于预设的VM的个数时,确定需要增加的VM的个数;

[0196] 根据所述应用的参数或者预设的VM配置规则,确定需要增加的每一个VM的参数;

[0197] 利用确定的所述需要增加的每一个VM的参数,创建需要增加的每一个VM。

[0198] 所述伸缩设备在应用需要增加新的VM时,在预先创建的VM资源池中选择符合应用需求的VM,注册并为该应用提供服务,省略了创建虚拟机的流程,缩短了虚拟机创建的时间,有效地提升了虚拟机的伸缩效率。

[0199] 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步骤是可以通程序来指令处理器完成,所述的程序可以存储于计算机可读存储介质中,所述存储介质是非短暂性(英文:non-transitory)介质,例如随机存取存储器,只读存储器,快闪存储

器,硬盘,固态硬盘,磁带(英文:magnetic tape),软盘(英文:floppy disk),光盘(英文:optical disc)及其任意组合。

[0200] 本发明是参照根据本发明实施例的方法、设备和计算机程序产品的流程图、方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。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的装置。

[0201] 以上所述,仅为本发明较佳的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到的变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应该以权利要求的保护范围为准。

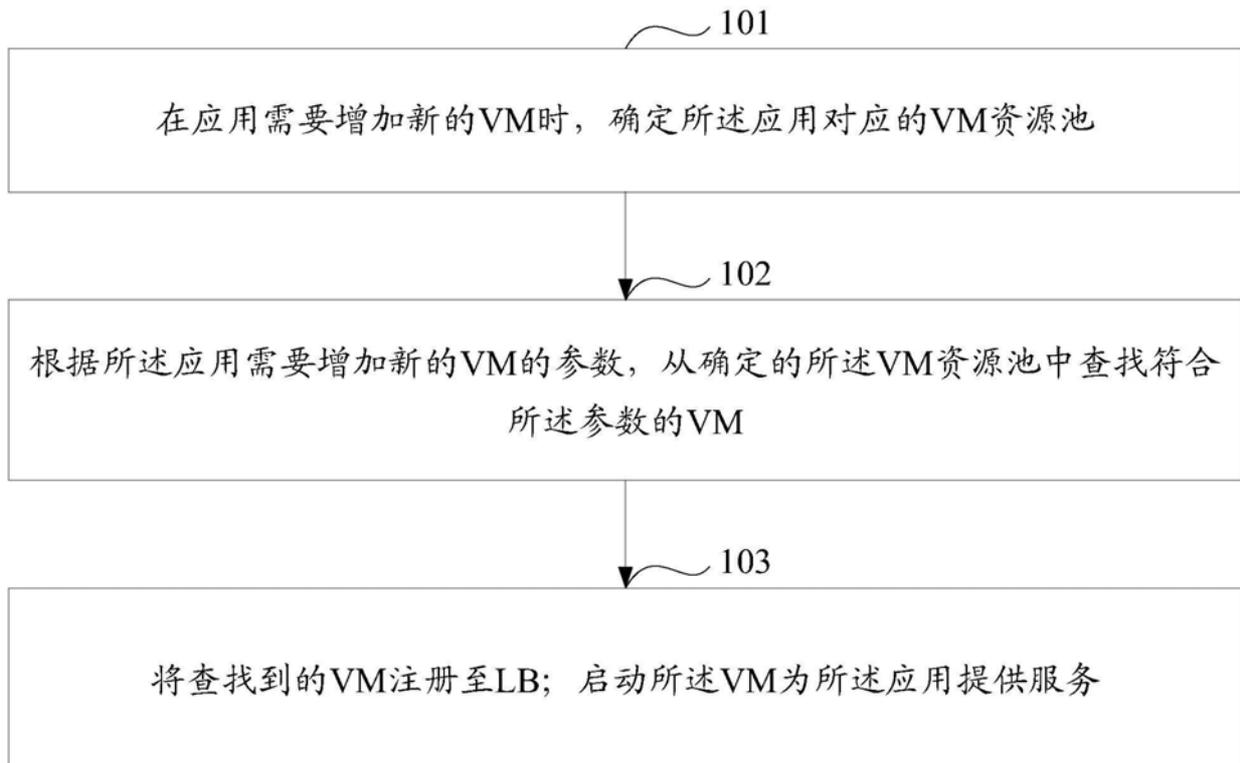


图1

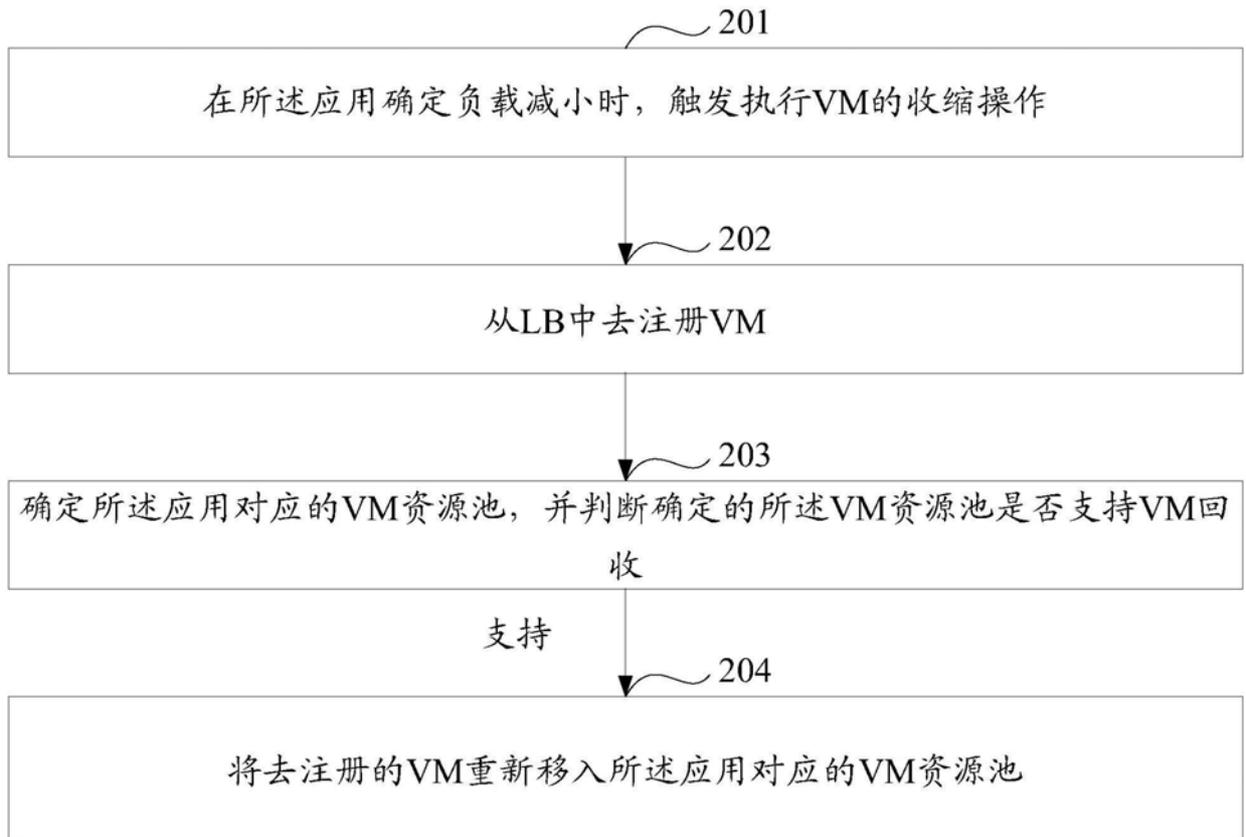


图2

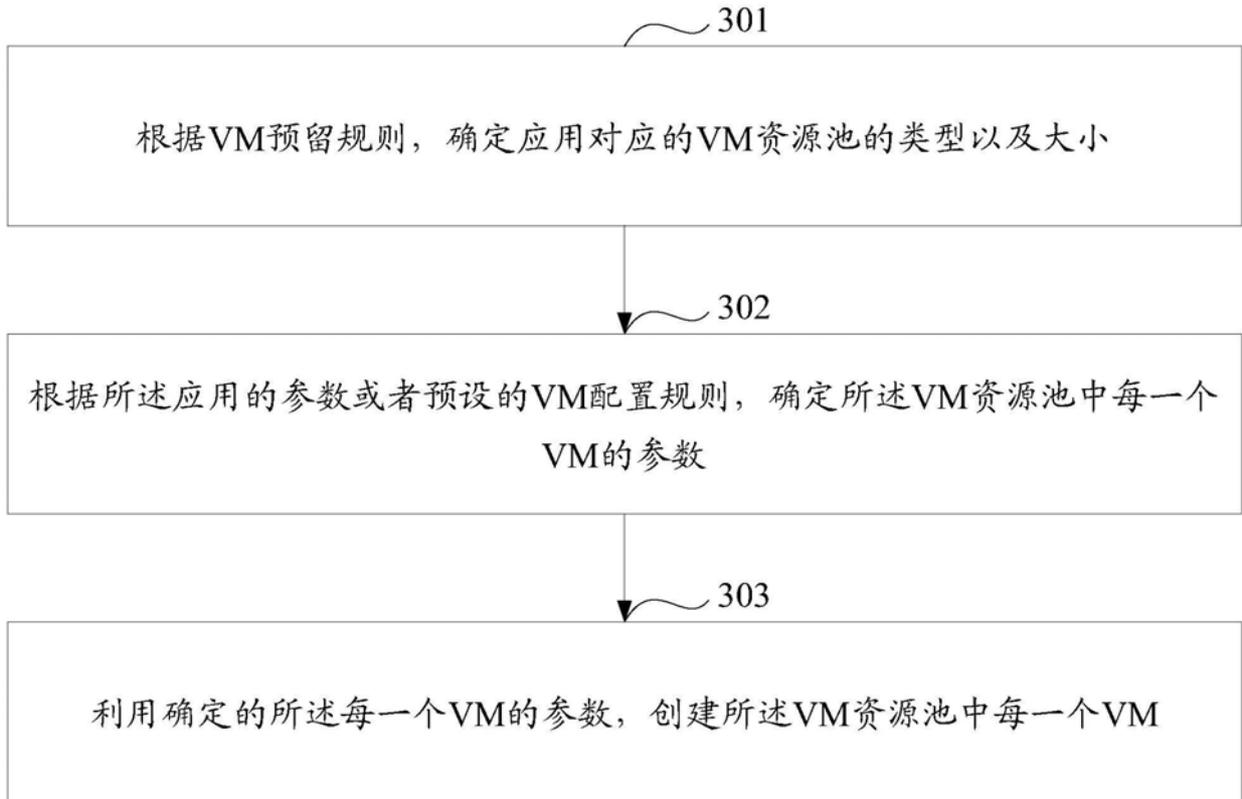


图3



图4

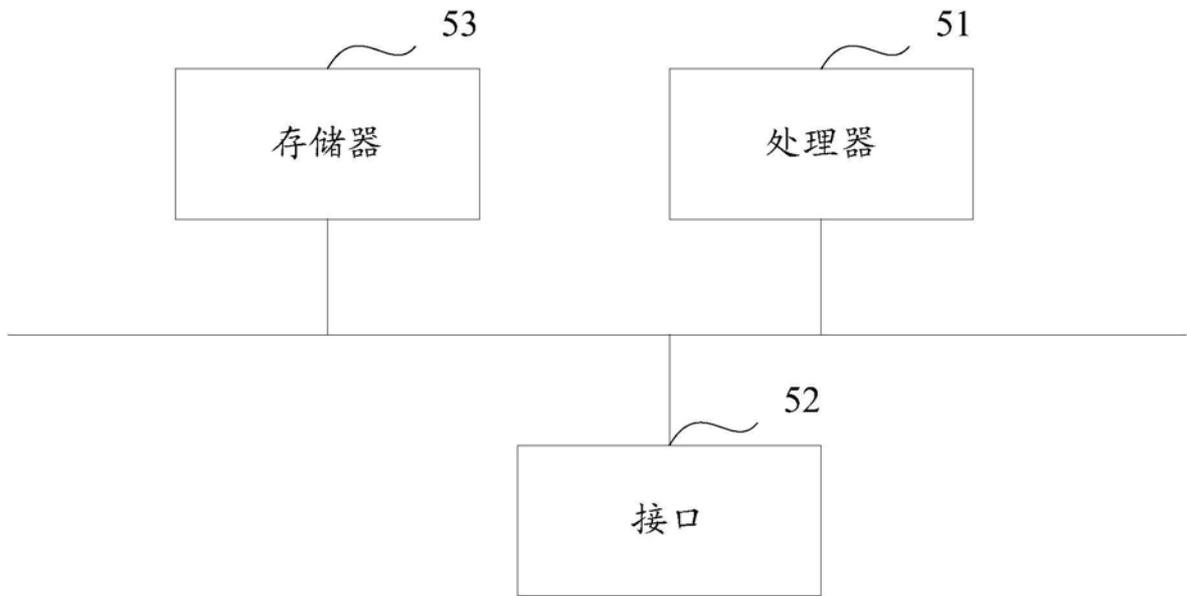


图5