



(12) 发明专利

(10) 授权公告号 CN 103034536 B

(45) 授权公告日 2015. 11. 25

(21) 申请号 201210533228. 8

(22) 申请日 2012. 12. 11

(30) 优先权数据

13/323, 213 2011. 12. 12 US

(73) 专利权人 微软技术许可有限责任公司

地址 美国华盛顿州

(72) 发明人 P·杜尔诺夫 L·伊伦-布里斯

M·库托年科 C·桑德斯

G·古普塔 A·哈萨恩

I·S·M·菲洛 A·沙哈

T·普夫雷格 S·赛德

S·里瓦斯卡 U·阿扎德

(74) 专利代理机构 上海专利商标事务所有限公

司 31100

代理人 陈斌

(51) Int. Cl.

G06F 9/48(2006. 01)

H04L 29/08(2006. 01)

(56) 对比文件

US 2010/0042720 A1, 2010. 02. 18,

US 2010/0088150 A1, 2010. 04. 08,

US 2010/0235903 A1, 2010. 09. 16,

审查员 武守秋

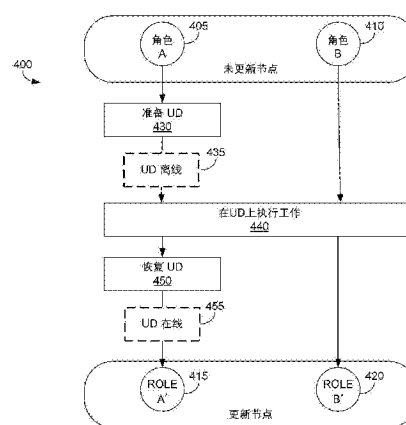
权利要求书3页 说明书20页 附图8页

(54) 发明名称

提高有状态应用的可用性

(57) 摘要

本发明公开了用于提高有状态应用的可用性的方法和系统。提供了用于促进云计算网络的结构控制器与运行在云计算网络上的服务应用之间的协调的系统和计算机可读介质。最初,选择包括服务应用的角色实例的更新域(UD),其中该服务应用代表被选定作为接收在其上执行的租户任务的目标的有状态应用。协调过程包括:准备UD以执行租户任务;将UD的角色实例停用至离线状态;允许租户任务执行;以及在完成租户任务执行时将角色实例恢复至在线状态。准备UD包括:通知在服务应用内建立的管理角色关于结构控制器执行租户任务的意图;以及接收传达受租户任务影响的角色实例的内部状态的复制存在的管理角色响应。



1. 一种用于促进云计算网络的结构控制器与运行在云计算网络中的服务应用之间的交互的方法,所述方法包括:

选择 (710) 第一更新域,所述第一更新域包括所述服务应用的一个或多个角色实例,其中所述一个或多个角色实例在在线状态下操作;

准备 (712) 所述第一更新域以用于执行租户任务,其中准备所述第一更新域包括:

(a) 通知 (714) 所述服务应用内的管理角色关于所述结构控制器执行租户任务的意图;以及

(b) 基于确定受租户任务影响的一个或多个角色实例的内部状态是否被复制在所述服务应用的不受租户任务影响的部分上,从所述管理角色接收 (716) 响应;

发起 (718) 将所述第一更新域的一个或多个角色实例停用至离线状态;以及

允许 (720) 所述租户任务在所述第一更新域上执行。

2. 如权利要求 1 所述的方法,其特征在于,所述方法进一步包括,完成租户任务的执行后,将所述第一更新域的一个或多个角色实例恢复 (722) 至在线状态。

3. 如权利要求 2 所述的方法,其特征在于,将所述第一更新域的一个或多个角色实例恢复至在线状态包括:

验证 (724) 所述受租户任务影响的一个或多个角色实例是起作用的;以及

通知 (726) 所述管理角色关于租户任务在所述一个或多个角色实例上的执行完成,由此提示所述服务应用继续采用所述一个或多个角色实例。

4. 如权利要求 3 所述的方法,其特征在于,将所述一个或多个角色实例停用至离线状态包括使主存所述第一更新域内的一个或多个角色实例的一组节点不工作,而且其中将所述一个或多个角色实例恢复至在线状态包括使主存所述第一更新域内的一个或多个角色实例的所述一组节点工作。

5. 如权利要求 4 所述的方法,其特征在于,所述一组节点中的每一个代表能够运行所述服务应用的一个或多个角色实例的物理机或虚拟机,而且其中所述角色实例代表支持所述服务应用的功能的相应组件程序。

6. 如权利要求 1 所述的方法,其特征在于,所述方法还包括选择角色实例的第二更新域以在其上运行租户任务,其中第一更新域和第二更新域在组员关系上互斥,且各自代表在云计算网络上传播租户任务的分立阶段。

7. 一种用于在执行租户任务时调度高优先级任务的计算机化方法,所述方法包括:

接收 (810) 在服务应用上执行租户任务的指示;

标识 (812) 更新域,所述更新域包括所述服务应用的一个或多个角色实例,其中所述一个或多个角色实例在在线状态下操作;

准备 (814) 所述更新域以执行所述租户任务;

发起 (816) 将所述更新域的一个或多个角色实例停用至离线状态;

接收 (818) 执行高优先级任务的指示,其中所述高优先级任务在优先级方案中预定为抢先于所述租户任务;

使所述租户任务的部署让位 (820) 于所述高优先级任务,其中所述让位的过程包括:

(a) 指示 (822) 所述服务应用执行将所述一个或多个角色实例恢复至在线状态的操作的截断集合;

(b) 将所述租户任务挂起 (824) ;以及

(c) 发起 (826) 所述高优先级任务在所述一个或多个角色实例上的执行 ;

完成所述高优先级任务的执行后,继续 (828) 所述租户任务的部署,其中继续的过程包括 :

(a) 重新发起 (830) 将所述更新域的一个或多个角色实例停用至离线状态 ;以及

(b) 允许 (832) 所述租户任务在所述一个或多个角色实例上的执行 ;以及

完成所述租户任务的执行后,将所述更新域的一个或多个角色实例恢复 (824) 至在线状态。

8. 如权利要求 7 所述的计算机化方法,其特征在于,准备所述更新域以执行所述租户任务包括 :

通知所述服务应用内的管理角色关于执行所述租户任务的意图 ;以及

基于确定受租户任务影响的一个或多个角色实例的内部状态被复制在所述服务应用的不受租户任务影响的部分上,从所述管理角色接收肯定响应 ;或

基于确定受租户任务影响的一个或多个角色实例的内部状态受限于所述一个或多个角色实例,从所述管理角色接收否定响应。

9. 如权利要求 8 所述的计算机化方法,其特征在于,还包括 :

从所述管理角色接收到肯定响应时,允许所述租户任务在所述一个或多个角色实例上执行 ;以及

从所述管理角色接收到否定响应时 :

(a) 当执行租户任务的指示是顾客发起时,推迟所述租户任务在所述一个或多个角色实例上的执行,以允许内部状态的复制 ;以及

(b) 当执行租户任务的指示是平台发起时,继续所述租户任务在所述一个或多个角色实例上的执行。

10. 一种用于执行使租户任务的逐步部署与服务应用的部分协调的方法的计算机系统,包括 :

所述服务应用 (350) 的一个或多个角色实例 (321-323),其中所述角色实例代表支持所述服务应用的功能的组件程序 ;

租户改变管理器 (580),用于指导所述租户任务的部署,其中指导部署的过程包括 :传达部署所述租户任务的指示 ;发起将所述一个或多个角色实例停用至离线状态 ;允许所述租户任务在所述一个或多个角色实例上执行 ;以及发起将所述一个或多个角色实例恢复至在线状态 ;以及

所述服务应用的管理角色 (597-599),用于监测所述一个或多个角色实例的内部状态,其中,在接收到部署所述租户任务的指示时,所述管理角色负责确定所监测到的受所述租户任务影响的一个或多个角色实例的内部状态是否被复制在所述服务应用的未受所述租户任务影响的部分上,并负责将对所述内部状态的可用性的指示传达至所述租户改变管理器。

11. 一种用于促进云计算网络的结构控制器与运行在云计算网络中的服务应用之间的交互的系统,所述系统包括 :

用于选择第一更新域的装置,所述第一更新域包括所述服务应用的一个或多个角色实

例,其中所述一个或多个角色实例在在线状态下操作;

用于准备所述第一更新域以用于执行租户任务的装置,其中用于准备所述第一更新域的装置包括:

(a) 用于通知所述服务应用内的管理角色关于所述结构控制器执行租户任务的意图的装置;以及

(b) 用于基于确定受租户任务影响的一个或多个角色实例的内部状态是否被复制在所述服务应用的不受租户任务影响的部分上,从所述管理角色接收响应的装置;

用于发起将所述第一更新域的一个或多个角色实例停用至离线状态的装置;以及

用于允许所述租户任务在所述第一更新域上执行的装置。

12. 如权利要求 11 所述的系统,其特征在于,进一步包括,用于完成租户任务的执行后,将所述第一更新域的一个或多个角色实例恢复至在线状态的装置。

13. 如权利要求 12 所述的系统,其特征在于,用于将所述第一更新域的一个或多个角色实例恢复至在线状态的装置包括:

用于验证所述受租户任务影响的一个或多个角色实例是起作用的装置;以及

用于通知所述管理角色关于租户任务在所述一个或多个角色实例上的执行完成,由此提示所述服务应用继续采用所述一个或多个角色实例的装置。

14. 如权利要求 13 所述的系统,其特征在于,用于将所述一个或多个角色实例停用至离线状态的装置包括用于使主存所述第一更新域内的一个或多个角色实例的一组节点不工作的装置,而且其中用于将所述一个或多个角色实例恢复至在线状态的装置包括用于使主存所述第一更新域内的一个或多个角色实例的所述一组节点工作的装置。

15. 如权利要求 14 所述的系统,其特征在于,所述一组节点中的每一个代表能够运行所述服务应用的一个或多个角色实例的物理机或虚拟机,而且其中所述角色实例代表支持所述服务应用的功能的相应组件程序。

16. 如权利要求 11 所述的系统,其特征在于,还包括用于选择角色实例的第二更新域以在其上运行租户任务的装置,其中第一更新域和第二更新域在组员关系上互斥,且各自代表在云计算网络上传播租户任务的分立阶段。

提高有状态应用的可用性

技术领域

[0001] 本发明涉及用于提高有状态应用的可用性的方法和系统。

背景技术

[0002] 一般而言,分布式服务应用被主存(host)在云计算网络中(跨越多个节点),且主要是为了通过服务应用组件的冗余来促进高可用性。这些服务应用通常被划分成多个部分,这多个部分包括一组服务应用组件。使用这些部分有助于确保整个服务应用不会在对主存该服务应用的节点进行更新或维修期间变得不可用。在操作中,当这些部分单独地离线时,允许该服务应用的余下部分继续在线操作。因此,在理想情况下,该服务应用的用户不会遇到该服务应用的可用性失效。

[0003] 当前,服务应用的划分是基于位置相关标准(例如,驻留在数据中心内的公共节点上的服务应用组件)或功能相关标准(例如,执行某种角色的所有服务应用组件不能同时离线)。用于将服务应用组件分组成多个部分的这些标准对于有状态应用(stateful application)(即依赖于存储在其组件外部的数据的易失性软件)而言在某种程度上是有效的。然而,这些标准对于保持持久内部状态的有状态应用而言是无效的,该状态在服务应用组件处持续。即,利用上述标准无法考虑服务应用组件的当前内部状态,由此在使先前形成的服务应用的部分离线时,导致该服务应用的潜在不可用。因此,本申请中介绍了一种解决上述问题的技术,实现了保持内部状态的分布式服务应用的高可用性。

发明内容

[0004] 提供本发明内容是为了以简化的形式介绍将在以下具体实施方式中进一步描述的概念。本发明内容并非旨在标识所要求保护的主题的关键特征或必要特征,也不旨在用于帮助确定所要求保护的主题的范围。

[0005] 本发明的实施例涉及用于促进保持内部状态的服务应用及其部分的高可用性的系统、方法以及计算机存储介质。一般而言,这些“有状态”服务应用包括云计算网络(下文称为“云”)内的节点(例如物理机或虚拟机)分布的多个部分,其中这些节点被容纳在一个或多个数据存储中。典型地,云配备有结构控制器,该结构控制器一般负责围绕对支持云的基础功能的计算机资源的健康进行维护和管理多种任务。

[0006] 例如,结构控制器可通过选定有状态服务应用的部分所驻留的节点作为目标并重新配置这些节点,以负责组织协调对运行在云上的有状态服务应用的更新。在这些更新期间,结构控制器在选定这些节点作为目标时考虑服务应用部分的可用性,以确保有状态服务应用均可操作。此外,结构控制器通过考虑服务应用部分的内部状态来确定要选定作为目标的节点,由此试图确保在更新期间不使内部状态完全离线。

[0007] 在示例性实施例中,结构控制器依赖于用于组织协调更新来以有状态服务应用保持在线和完全功能的方式调度节点的重新配置的软件组件(在本文中称为“租户改变管理器(tenant-change manager)”)。在操作中,租户改变管理器采用在结构控制器与有状态服

务应用的专用组件之间交互的协议。这些专用组件(在本文中称为“管理角色”)在有状态服务应用内实现,以监测有状态服务应用的功能部分(诸如有状态服务应用的角色实例)的内部状态。

[0008] 通过租户改变管理器促进的该交互通过考虑角色实例的内部状态的可用性来确保有状态服务应用内的改变操作的正确排序,以保持整个有状态服务应用的恒定高可用性。作为示例,对内部状态可用性的该考虑涉及:管理角色推断改变操作期间的内部状态,并将该信息报告给租户改变管理器。租户改变管理器采用该信息来建立用于在有状态服务应用的特定角色实例上实现改变操作或“租户任务”的开始时间。因此,租户改变管理器与管理角色之间的交互通过考虑分布式角色实例的个体内部状态而访问有状态服务应用的真实可用性。

附图说明

[0009] 以下参考附图详细描述本发明的实施例,附图中:

[0010] 图 1 是适用于实现本发明的实施例的示例性计算环境的框图;

[0011] 图 2 是适用于实现本发明的实施例的数据中心的框图,该数据中心配置成容许和支持分布在数据中心的多个节点上的服务应用的操作;

[0012] 图 3 是示出根据本发明的实施例的示例性主存环境的框图,其中多个角色实例被组合以形成更新域,以用于逐步实现数据中心内的租户任务;

[0013] 图 4 是根据本发明的实施例的用于有状态节点和无状态节点的处理流程的图形表示;

[0014] 图 5 是根据本发明的实施例的包括用于执行实现租户任务的分立操作的多个组件的云计算网络的框图;

[0015] 图 6 是根据本发明的实施例的描绘结构控制器和服务应用之间的交互的数据中心的示例性拓扑的图形表示;

[0016] 图 7 是示出根据本发明的实施例的用于促进云计算网络的结构控制器与服务应用之间的交互的整体方法的流程图;以及

[0017] 图 8 是根据本发明的实施例的示出用于在执行租户任务时调度高优先级任务的整体方法的流程图。

[0018] 详细描述

[0019] 本文用细节来描述本发明的实施例的主题以满足法定要求。然而,该描述本身并非旨在限制本专利的范围。相反,发明人设想所要求保护的主体还可结合其他当前或未来技术按照其他方式来具体化,以包括不同的步骤或类似于本文档中所描述的步骤的步骤组合。

[0020] 一般而言,本发明的实施例在云计算网络中引入了通过允许有状态服务应用保持对它们的顾客的高可用性来支持对有状态服务应用的更新和修复的机制。具体而言,这些机制能够在有状态服务应用的多个部分(在本文中称为“角色实例”)的升级、配置改变、向上或向下扩展以及重新定位过程期间管理这些有状态服务应用的持续可用性。在示例性实施例中,这些机制依赖于有状态服务应用与结构控制器之间的交互,结构控制器管理有状态服务应用的角色实例的生命周期。如下文详细讨论,在一个实例中,该机制包括驻留在结

构控制器处的租户改变管理器和驻留在有状态服务应用内的管理角色,管理角色交互以智能地调度更新或修复。

[0021] 如本文中所使用,短语“有状态服务应用”一般表示其中状态嵌入服务应用相应的角色实例或程序组件中的那些服务应用。例如,在负责显露特定的在线接口以用于进行银行业务的金融机构服务应用内,状态可包含顾客数据(例如账户信息),顾客数据被保持在服务应用的角色实例内部。即,与上传至远程位置相反,顾客数据被本地主存在服务应用角色实例处。因此,当服务应用分布在云计算网络内的多个节点上时,顾客提供的账户信息或“状态”也分布在云计算网络的节点上。此外,因为存在与服务应用进行银行业务的多个顾客,所以每个角色实例可具有稍不同的内部状态。

[0022] 因此,伴随着知晓作为云计算网络基础的平台的可用性和健康状况,理解其中运行的分布式服务应用的内部状态对于调度维护和恢复操作的定时是有帮助的。例如,当执行维护和恢复操作时,可能使服务应用的多个部分离线。如果没有特定内部状态的副本仍保持在线,则使这些部分离线会不利地影响甚至停用整个服务应用。因此,本发明的实施例涉及用于理解多个内部状态的位置并协调维护和恢复操作以使得每个内部状态的副本保持在线的机制。

[0023] 因此,在理解内部状态位置的情况下协调维护和恢复操作有助于实现角色实例的内部状态的高可靠性。在一个实施例中,协调可包括在相同角色的多个实例上复制特定状态,以允许使角色实例中本地主存该特定状态的角色之一实例离线。例如,如果角色当前具有在数据中心上分布的五个实例,且如果这些角色实例中的特定一个角色实例存储顾客想要维持的状态,则在使该特定角色实例离线之前,将所存储的该状态复制、传播以及维持在五个实例中的其它实例处。以此方式,在多个节点上的多个角色实例上复制所存储状态确保该特定状态的可靠性。因此,即使该特定角色实例在维护和恢复操作范围之外失效,另一角色实例保持失效的角色实例的所存储状态。

[0024] 虽然已经描述了鉴于维护和恢复操作来复制状态的多个不同实施例,但应理解和领会的是,其它类型的适合的服务应用也可经由本文中描述的过程来更新,而且本发明的实施例不限于本文中描述的那些有状态服务应用。例如,可按照与在云计算网络的基础平台内执行租户任务时相似的方式来解决无状态服务应用。如本文中所使用,短语“无状态服务应用”一般表示由云计算网络主存的服务,该服务将其存储数据卸载至外部实体,诸如能够管理所存储数据的另一数据中心。该数据中心可在不同的物理机上保持不同的状态,但能够通过创建许多冗余(例如多个机器上的多个副本)来管理这些状态。

[0025] 因此,在一个方面中,本发明的实施例涉及一个或多个机器可读介质,一个或多个机器可读介质上具体化有计算机可执行指令,计算机可执行指令在执行时执行用于促进云计算网络的结构控制器与运行在云计算网络中的服务应用之间的交互的方法。该方法最初可包括选择包括该服务应用的一个或多个角色实例的第一更新域(UD)。典型地,角色实例运行在在线条件下,且代表支持该服务应用的功能的组件程序(例如该服务应用的角色的单个复制)。

[0026] 该方法还可包括准备第一 UD 以用于执行租户任务(例如平台发起的更新、顾客发起的更新、平台发起的维修、或顾客发起的维修)。在示例性实施例中,准备第一 UD 包括至少以下步骤:通知服务应用内的管理角色关于执行租户任务的结构控制器意图;以及基于

确定受租户任务影响的角色实例的内部状态是否在该服务应用的不受租户任务影响的一部分上复制,从该管理角色接收响应。在实施例中,该响应是由管理角色经过检查角色实例内部状态和 / 或经过采取其它动作之后产生的。例如,这些其它动作可包括利用在云计算网络的基础结构内建立的信道来检索附加信息(例如租户角色的状态)。经过准备第一 UD 的角色实例之后,发起将这些角色实例停用至离线状态。一般而言,停用角色实例包括使主存第一 UD 内的一个或多个角色实例的一组节点不工作。一旦角色实例已采取离线状态,就允许租户任务在第一 UD 上执行。

[0027] 再进一步,该方法可包括,在完成租户任务的执行后,就将第一 UD 的角色实例恢复至在线状态。在示例性实施例中,将第一 UD 的实例恢复至在线状态包括至少以下步骤:验证受租户任务影响的角色实例在起作用;以及通知管理角色关于角色实例上的租户任务的执行完成,由此提示该服务应用继续利用这些角色实例来执行任务。一般而言,将第一 UD 内的角色实例恢复至在线状态还包括使主存这些角色实例的该组节点工作。在将第一 UD 内的角色实例恢复至在线状态后,则该方法可通过选择角色实例的第二 UD 以在其上运行租户任务而继续。典型地,第一 UD 和第二 UD 在组员关系上是互斥的,可分别代表在云计算网络上传播租户任务时的分立阶段。

[0028] 在另一方面中,本发明的实施例涉及用于在执行租户任务时调度高优先级任务的计算机化方法。该计算机化方法包括执行以下逻辑步骤:接收在服务应用上执行租户任务的指示;以及标识包括该服务应用的一个或多个角色实例的更新域(UD)。在此时,角色实例在在线状态下操作。所准备的 UD 是为了执行租户任务,而且随后将 UD 的实例停用至离线状态。在示例性实施例中,准备 UD 以用于执行租户任务的过程包括多个逻辑步骤,包括以下:通知服务应用内的管理角色关于执行该租户任务的意图;以及接收来自管理角色的肯定响应,或接收来自管理角色的否定响应。

[0029] 典型地,在管理角色确定受租户任务影响的角色实例的内部状态被复制在该服务应用的不受租户任务影响的部分上之后,接收肯定响应。在租户改变管理器(在结构控制器内操作)接收来自管理角色的肯定响应后,租户任务可能被允许在角色实例上执行。替代地,在管理角色确定受租户任务影响的角色实例的内部状态受限于那些角色实例之后,接收否定响应。在租户改变管理器接收到来自管理角色的否定响应之后,多个操作可接着发生。在一个实施例中,当执行租户任务的指示是顾客发起的,则租户任务在角色实例上的执行被推迟,以允许内部状态的复制。在另一实施例中,当执行租户任务的指示是平台发起的,则租户任务在角色实例上的执行被提示继续。

[0030] 在某个后续时间点,接收执行高优先级任务的指示。如本文中所使用,短语“高优先级”不意味着限制性,而是代表被预先确定为在优先级方案内抢先于租户任务的任何任务。在接收到执行高优先级任务的指示后,租户任务的部署让位于高优先级任务。在示例性实施例中,让位的过程包括多个逻辑步骤,包括以下:指令服务应用执行将角色实例恢复至在线状态的操作的截断集合;将租户任务挂起;以及发起高优先级任务在角色实例上的执行。

[0031] 在完成高优先级任务的执行后,继续租户任务的部署。在示例性实施例中,继续的过程包括多个逻辑步骤,包括以下:重新发起将 UD 的角色实例的停用至离线状态;以及允许租户任务在角色实例上执行。一旦完成租户任务的执行,就可将 UD 的角色实例恢复至在

线状态。在示例性实施例中,将角色实例恢复至在线状态的过程包括多个逻辑步骤,包括以下:验证受租户任务影响的角色实例在起作用;以及通知管理角色关于角色实例上的租户任务的执行完成,由此提示该服务应用继续利用这些角色实例来执行多个任务。

[0032] 在又一方面中,本发明的实施例涉及用于执行使租户任务的部署与服务应用的组件程序协调的方法的计算机系统。一般地,该计算机系统包括耦合到计算机存储介质的处理单元,其中该计算机存储介质存储可由该处理单元执行的多个计算机软件组件。最初,计算机软件组件包括该服务应用的一个或多个角色实例、租户改变管理器、管理角色、主机代理、线路协议、调度组件以及节点状态驱动器。在实施例中,角色实例代表支持该服务应用的功能的组件程序。租户改变管理器典型地配置用于指导租户任务的部署。指导部署的过程一般包括:传达租户任务部署的指示;发起将角色实例停用至离线状态;允许租户任务在角色实例上执行;以及发起将角色实例恢复至在线状态。

[0033] 作为服务应用的组件的管理角色典型地配置用于监测角色实例的内部状态。在实施例中,在接收到租户任务的部署的指示后,管理角色就负责确定所监测到的受租户任务影响的角色实例的内部状态是否在该服务应用的不受租户任务影响的部分上复制。此外,管理角色可配置用于将内部状态可用性的指示传达至租户改变管理器。

[0034] 主机代理典型地驻留在主存角色实例的节点上。在操作中,主机代理通过线路协议向服务应用展现角色实例的预期状态。如本文中所使用,短语“预期状态”一般代表角色实例采取离线状态或在线状态的目标。调度组件一般配置用于根据优先级方案使租户任务被高优先级任务中断。而且,节点状态驱动器一般配置用于将角色实例停用至离线状态、执行租户任务、以及在租户改变管理器的指导下将角色实例恢复至在线状态。

[0035] 如上所述,与使服务应用的所有角色实例同时离线相比,租户任务通常被划分成多个阶段并在服务应用的角色实例分组上逐步进行。因此,本发明的实施例涉及标识主存环境或数据中心内的更新域。如本文中所利用,短语“更新域”不意味着限制性,而是一般表示主存该服务应用的一个或多个角色实例并且可同时离线的一组节点(例如物理机或虚拟机)的描述。在特定实施例中,通过更新域描述的该组节点代表可能同时不可用(例如在有意地使那些节点离线以向其安装软件)、同时确保依赖于那些节点的服务应用保持可用的节点。即,更新域的该组节点典型地包括在执行租户任务时被允许同时不起作用的有限数量的节点,以确保在运行在该组节点上的角色实例在更新域以外具有复制。因此,更新域被设置在适当的位置,以向服务应用的拥有者确保仅该服务应用的最小数量的角色实例会同时失效。

[0036] 建立更新域一般包括遵守由主存环境施加的多个规则。例如,存在对可加入公共更新域的角色 A 的实例数量的限制。或者,存在对可与角色 A 的实例一起加入公共更新域的角色 B 的实例数量的限制。一般而言,如本文中使用的“角色”提供该服务应用的起作用部分的模板描述。例如,在在线购物服务应用的情况下,存在在三个角色之间划分的三个核心任务,分别是:角色 A (被分配呈现第一 GUI 以接收购买订单的任务)、角色 B (被分配执行购买订单的任务)、以及角色 C (被分配呈现第二 GUI 以传达购买订单的状态的任务)。在示例性实施例中,各个角色代表该服务应用的特定类的程序组件。典型地,服务模型描述在数据中心内放置每个角色的多少个实例,其中每个实例是特定类的程序组件或角色的复制。换言之,每个角色代表每类程序组件的实例的集合,其中该服务应用可具有用于执行其

功能的任何数量的程序组件类。

[0037] 角色通常具有对其应用的配置设定。在一个实例中,该角色的配置设定可包括由该角色的所有实例共享的集体设定。在另一实例中,配置设定可包括每个角色实例特定的个体设定。这些个体设定可包括在角色实例处存储的状态,其中该状态是该角色实例所专有的。参考上述在线购物服务应用,存储在角色 B 的实例处的状态可以是由某个客户作出的特定购买订单的细节。

[0038] 在简要描述了本发明的实施例的概览后,以下描述适于实现本发明的实施例的示例性操作环境。

[0039] 操作环境

[0040] 首先具体参考图 1,示出了用于实现本发明的各实施方式的示例性操作环境,并将其概括地指定为计算设备 100。计算设备 100 只是合适的计算环境的一个示例,并且不旨在对本发明的使用范围或功能提出任何限制。也不应该将计算设备 100 解释为对所示出的任一组件或其组合有任何依赖性 or 要求。

[0041] 本发明可以在由计算机或诸如个人数据助理或其他手持式设备之类的其他机器执行的计算机代码或机器可使用指令(包括诸如程序模块之类的计算机可执行指令)的一般上下文中描述。一般而言,包括例程、程序、对象、组件、数据结构等等的程序模块是指执行特定任务或实现特定抽象数据类型的代码。本发明可以在各种系统配置中实施,这些系统配置包括手持式设备、消费电子产品、通用计算机、专用计算设备等等。本发明也可以在其中任务由通过通信网络链接的远程处理设备执行的分布式计算环境中实施。

[0042] 参考图 1,计算设备 100 包括直接或间接耦合以下设备的总线 110:存储器 112、一个或多个处理器 114、一个或多个呈现组件 116、输入 / 输出(I/O) 端口 118、输入 / 输出组件 120、和说明性电源 122。总线 110 可以是一条或多条总线(诸如地址总线、数据总线、或其组合)。虽然为了清楚起见利用线条示出了图 1 的各框,但是实际上,各组件的轮廓并不是那样清楚,并且比喻性地来说,线条更精确地将是灰色的和模糊的。例如,可以将诸如显示设备等呈现组件认为是 I/O 组件。而且,处理器具有存储器。发明人认识到这是本领域的特性,并重申,图 1 的图示只是例示可以结合本发明的一个或多个实施方式来使用的示例性计算设备。诸如“工作站”、“服务器”、“膝上型计算机”、“手持式设备”等分类之间没有区别,它们全部都被认为是在图 1 的范围之内的并且被称为“计算设备”。

[0043] 计算设备 100 通常包括各种计算机可读介质。计算机可读介质可以是可由计算设备 100 访问的任何可用介质,而且包含易失性和非易失性介质、可移动和不可移动介质。作为示例而非限制,计算机可读介质可包括计算机存储介质和通信介质。计算机存储介质包括以用于存储诸如计算机可读指令、数据结构、程序模块或其它数据等信息的任何方法或技术实现的易失性和非易失性、可移动和不可移动介质。计算机存储介质包括,但不限于, RAM、ROM、EEPROM、闪存或其它存储器技术、CD-ROM、数字多功能盘(DVD)或其它光盘存储、磁带盒、磁带、磁盘存储或其它磁性存储设备、或能用于存储所需信息且可以由计算设备 100 访问的任何其它介质。通信介质一般将计算机可读指令、数据结构、程序模块或其它数据包含在经调制的数据信号中,诸如载波或其它传输介质并且包括任何信息传递介质。术语“经调制数据信号”是指具有以在信号中编码信息的方式被设定或改变其一个或多个特征的信号。作为示例而非限制,通信介质包括有线介质,诸如有线网络或直接线路连接,以及无线

介质,诸如声学、RF、红外线和其他无线介质。上述的任意组合也应包含在计算机可读介质的范围内。

[0044] 存储器 112 包括易失性和 / 或非易失性存储器形式的计算机存储介质。存储器可以是可移动的、不可移动的、或其组合。示例性硬件设备包括固态存储器、硬盘驱动器、光盘驱动器等。计算设备 100 包括从诸如存储器 112 或 I/O 组件 120 等各种实体读取数据的一个或多个处理器。呈现组件 116 向用户或其他设备呈现数据指示。示例性呈现组件包括显示设备、扬声器、打印组件、振动组件等等。

[0045] I/O 端口 118 允许计算设备 100 逻辑上耦合至包括 I/O 组件 120 的其他设备,其中某些设备可以是内置的。说明性组件包括话筒、操纵杆、游戏手柄、圆盘式卫星天线、扫描仪、打印机、无线设备等等。

[0046] 数据中心的体系结构

[0047] 现在转到图 2,示出了根据本发明的实施例的框图,其中示出了数据中心 200,配置成容许和支持根据服务模型的服务应用的组件程序或角色实例的操作。本领域普通技术人员将理解和领会,图 2 中示出的数据中心 200 仅仅是主存环境的一个合适的部分的示例,并且不旨在暗示关于本发明的实施例的用途或功能的范围的任何限制。数据中心 200 也不应解释为具有与其中示出的任何单个资源或资源的组合有关的任何依赖性 or 要求。此外,尽管为了清楚起见用线条示出了图 2 的各个框,但是在实际上,各组件的轮廓并不是那样清楚,并且比喻性地来说,线条更精确地将是灰色的和模糊的。

[0048] 数据中心 200 包括经由网络云 240 互连的多个资源。如本文中描述的这些资源可包括软件组件(例如结构控制器 295)以及有形硬件元件,诸如机架 A205、机架 B210、机架 C215、顶机架交换机 230、电源板 235、服务器 270、计算设备 280、更新域 251 以及更新域 252。网络云 240 互连这些资源,使得在多个物理资源上分布放置的服务应用的实例可识别其它实例的位置,以在这些实例之间建立通信。此外,网络云 240 通过连接服务应用的实例的信道以及该服务模型需要的任何其它元件来促进该通信。该网络可包括但不限于一个或多个局域网(LAN)和 / 或广域网(WAN)。这样的联网环境常见于办公室、企业范围计算机网络、内联网和因特网中。因此,网络不在此进一步描述。

[0049] 一般而言,数据中心 200 容许一种分布式系统,其中服务应用的多个组件程序或角色实例在执行这些组件程序的处理设备上独立运行,但经由网络云 240 互连。有时候,这些计算资源或节点会经历多个更新。更新域 251 和 252 典型地提供如下的一组保证:服务应用的角色实例将分布在数据中心 200 中的资源上,而且不会在底层平台的更新期间全部同时离线。相应地,保留了服务应用的操作的一致性。

[0050] 返回数据中心 200 的配置,服务器 270、计算设备 280、故障域 251 和 252 以及具有与其耦合的刀片(例如刀片 A260、刀片 B261、刀片 C262)的机架 A205、B210 以及 C215 中的每一个可以是任何类型的计算设备,诸如以上参考图 1 所描述的计算设备 100。仅作为示例而非限制,服务器 270、计算设备 280 以及刀片 A260、刀片 B261 以及刀片 C262 中的每一个可以是个人计算机、台式计算机、膝上型计算机、手持设备、移动手持设备、消费电子设备以及类似物。在实施例中,数据中心 200 可包括以上列出的任何数量的物理资源,且可包括计算设备的任何集合或能够执行该服务应用的其它机器。

[0051] 此外,上述物理资源中的一个或多个可进一步主存驻留在其上的结构控制器 295,

以经由通过数据中心 200 的有线或无线网络路径来部署、访问以及管理该服务应用。然而，应理解，本发明的实施例不限于在图 2 中示出的这样的物理资源上的实现，而可以在本发明的实施例范围内的多种不同类型的计算设备和设施中的任一种上实现。换言之，所示出的数据中心 200 的资源描绘了仅旨在用于讨论目的的示例性配置；因此，计算工业中已知的任何合适的资源布局可被使用并由本发明构想。

[0052] 具体而言，数据中心 200 的示例性资源用于引入更新域的概念。如上所讨论，更新域代表在数据中心 200 内执行租户任务时变得不可用的节点的组或集合。如上所讨论，节点可指代数据中心 200 中的刀片（例如刀片 A260、刀片 B261 以及刀片 C262）、计算机（例如计算设备 280）、机器（例如服务器 270）或能够运行服务应用的组件程序或实例的任何其它设备。因此，抽象出更新域的优点是为了理解数据中心内的哪些资源组在执行租户任务的一个阶段时一起离线，以及如何隔离其它资源以免离线。

[0053] 虽然已经描述了数据中心 200 和更新域 251 和 252 的多种不同配置，但本领域普通技术人员应当理解和领会，可基于在节点上运行的角色实例的身份来抽象出将一组节点分组的其它类型的合适更新域，而且本发明的实施例不限于本文中示出的那些更新域 251 和 252。

[0054] 服务应用的更新域

[0055] 现在转到图 3，示出了根据本发明的实施例的示例性主存环境 300 的图形表示。具体而言，主存环境 300 包括结构控制器 295、服务应用 350 以及用于支持结构控制器 295 和服务应用 350 的操作的数据中心 200。如上所讨论，数据中心 200 包括至少部分地基于服务应用 350 的角色 A、B 和 C 的冗余被组织成更新域 305 和 310 的资源，其中角色位于实例 311、312、313、321、322 以及 323 处。一般而言，更新域 305 和 310 分别描述了根据更新方案或“任务列表”而被有意地致使同时不可用的一个或多个节点。在实施例中，任务列表（参见图 6 的附图标记 610）根据优先级来组织被调度和 / 或待决的租户任务。如下文更完整讨论，使任务列表的至少一部分显露给服务应用 350 的管理角色，以使服务应用 350 准备好接收即将来临的租户任务。

[0056] 如本文中所使用，短语“租户任务”指的是由结构控制器 295 对服务应用 350 执行的不可变的一组改变操作。这些改变操作会影响服务应用 350 的可用性。在实施例中，当前存在由主存环境 300 定义的固定数量的租户任务类型。然而，这些任务类型可在未来扩展。着重于服务应用 350 的示例性租户任务包括以下：服务应用更新任务（例如维护动作）；以及服务应用代码修改（例如升级动作或推出新版本）；服务应用拓扑改变；服务应用配置改变；在采用重启（例如停止和开始操作）的节点 320 的虚拟机（VM）内的服务应用访客 OS 改变；采用重启和重铺的服务应用维修（例如改变访客 OS 的注册表）。一般而言，维修在 VM 无法通过其它进程恢复时发起。

[0057] 其它类型的租户任务可着重于主存环境 300 的维护。这些租户任务类型的示例包括以下：基础结构更新（当出故障时使实例 321、322 和 323 离线）；将应用 350 从一个地理位置迁移至另一地理位置（例如由于硬件和其它平台问题导致的实例 321、322 或 323 的重新定位）；以及资源更新（例如对监测运行在节点 320 的 VM 上的实例 321、322 和 323 的代理的更新）。虽然着重于改变主存环境 300，但这些类型的租户任务也可产生实例 321、322 或 323 的可用性的中断。

[0058] 租户任务的执行可由主存环境 300 的需求(例如更新操作系统软件的版本)驱动,或可由主存环境 300 的管理员来引起。当选择租户任务以在数据中心 200 内实现时,租户任务的实际执行可按阶段逐步地进行。即,租户任务被划分成多个阶段,其中每个阶段的执行潜在地影响服务应用 350 的部分(例如一个或多个角色实例)。例如,如果服务应用 350 具有角色 A 的十个实例,则可将租户任务分成十个阶段,阶段可进一步分成子集。在实施例中,取决于租户任务类型、该租户任务专有的限制、或服务应用 350 的部署布局,整个更新域或更新域的子集可受阶段的影响。此外,阶段执行的排序是租户任务类型所专属的,和/或受服务应用 350 的服务模型管控。

[0059] 在操作中,执行每个阶段可包括在不同的更新域执行租户任务,直到解决每个更新域为止。例如,租户任务的第一阶段可包括停用更新域 305 内的节点 320 以执行维护操作,而租户任务的第二阶段可包括停用更新域 310 内的节点。停用节点 320 隐含地而且并发地使角色 321、322 以及 323 离线。然而,由于角色 A 位于实例 311 和 321 处,角色 B 位于实例 322 和 312 处,以及角色 C 位于实例 323 和 313 处,使角色 321、322 以及 323 离线将不一定使服务应用 350 不可用。

[0060] 有时候,当角色仅位于最小数量的实例处时,在租户阶段期间实际上仅使更新域的子集离线。即,当角色实例 321、322 和 323 被分组到相同更新域 305 中时,可使角色实例 321、322 和 323 的任何组合同时离线。不过,在第一阶段期间,租户任务的效果仅限于更新域 305 的那些角色实例 321、322 以及 323。

[0061] 如果服务应用 350 是有状态应用,而且如果角色实例 321、322 和 323(包含在更新域 305 中)的状态未在角色实例 311、312 和 313(包含在更新域 310 中)中复制,则在实现租户任务的第一阶段时可能发生信息的不可用。因此,以下讨论的本发明的实施例涉及协调租户任务的技术,使得在迭代地执行租户任务的多个阶段时,不使存储在角色实例处的状态不可用。

[0062] 本发明的实施例进一步包括对分布式服务应用 350 的拓扑的更新。有利地,该拓扑更新允许服务应用 350 的角色实例为了互相通信的目的而检查该服务应用 350 的其它角色实例。在实例中,该拓扑更新包括用于在服务应用的角色实例上传播有关该服务应用拓扑的信息的机制。作为示例,所传播的信息可包括重新定位的角色实例的改变地址的描述和/或可散布至整个服务应用 350 的其它数据。

[0063] 现在将参考图 3 描述该拓扑更新。最初,如果角色 A 的实例 321 想要与角色 B 的实例 322 通信,则实例 321 应当知晓实例 322 的位置(例如 IP 地址)以便发起该通信。典型地,分别使用 IP 地址到角色实例 321 和 322 的名称的映射,以便促进角色实例 321 与 322 之间的通信。该映射被保持在实例 321 和 322 中的每一个处。

[0064] 经过实例 321 和 322 的最初部署,映射是当前的,因为实例 321 和 322 的位置自从节点 320 上的实例化之后已经保持静态。随时间流逝,经过对服务应用 350 进行的动态更新(例如重新定位、添加或删除),实例 321 和 322 中的一个或二者可能经历原始 IP 地址的改变。实例 321 和 322 中的一个或多个的这些改变的 IP 地址可能使服务应用的其它实例具有不准确的映射,从而导致服务应用拓扑的不一致理解和相互通信的中断。例如,如果由于实例 322 已经移动,实例 321 处保持的映射包括与服务应用拓扑有关的不准确信息,则实例 321 可能无法定位实例 322 或与实例 322 交谈。

[0065] 为了解决以上讨论的问题,配置了拓扑更新以施加冗余手段,该冗余手段记忆 IP 地址相对于更新域的改变。例如,当两个角色实例 321 和 311 分别被添加至两个分开的更新域 305 和 310 时,在第一更新域 305 内操作的服务应用 350 被通知第一角色实例 321 被放置在其中,同时在第二更新域 310 内操作的服务应用 350 被通知第二角色实例 311 被放置在其中。然后,经过分发通知之后,第一和第二角色实例 321 和 311 在它们各自的更新域 305 和 310 中被实例化和启动。因此,驻留在其它角色实例 322、323、312 和 313 的映射处的拓扑信息被更新两次,给予服务应用 350 关于新角色实例 321 和 311 的定位位置的一致理解。

[0066] 对更新域(UD)应用租户任务

[0067] 现在转到图 4,图形表示出根据本发明的实施例的用于在有状态服务应用(例如角色 A)和无状态服务应用(例如角色 B)二者上执行租户任务的阶段的流程。当在无状态服务应用上执行租户任务时,通常不考虑角色实例中的一个或多个上具体存储的内部状态。因此,可按照调度使无状态应用的角色实例离线。例如,当租户任务的阶段包括更新包含无状态应用的角色实例 B410 的 UD 时,租户任务在角色实例 B410 上执行(参见图 4 的框 440),从而导致经更新的角色实例 B' 420。在该示例中,没有理由与无状态应用的其它角色实例协调,因为这些状态被安全地上传到远程数据存储(即未本地存储在角色实例 B410 处)。

[0068] 然而,当在有状态服务应用上执行租户任务时,存在对角色实例上存储的内部状态的考虑。因此,可按照协调方式使有状态服务应用的角色实例离线。在示例性实施例中,该协调包括在 UD 处执行租户任务时采用至少以下阶段:准备 UD(参见框 430);在 UD 上执行工作(参见框 440);以及恢复 UD(参见框 450)。当准备 UD 的角色实例 A405 以接收租户任务时,有状态服务应用的管理组件(参见图 5 的附图标记 597、598 和 599)负责确定是否会使角色实例 A405 离线,并标识存储在角色实例 A405 上的内部状态。

[0069] 在实施例中,确定是否会使角色实例 A405 离线包括至少以下步骤:识别结构控制器执行特定租户任务的意图(例如接收来自结构控制器的角色实例 A405 的目标状态,该目标状态警告管理角色关于将使角色实例 A405 进入离线状态);接收来自结构控制器的描述租户任务类型的信息(例如要在角色实例 A405 上执行的操作和那些操作的效果);使结构控制器暂停执行租户任务(例如直到管理角色以批准继续执行作为响应或直到预定时间段期满为止);采取动作以查明在使角色实例 A405 进入离线状态时服务应用是否将保持可用;以及利用消息对结构控制器作出响应。因此,管理角色提示服务应用重新配置自身以不再依赖于或指望 UD 内进入离线状态的角色实例 A405。以此方式,响应于结构控制器而从管理角色发送的消息可指示角色实例 A405 准备好离线。

[0070] 如果经过评估经调度而在角色实例 A405 上执行的租户任务类型并查明该服务应用将在使角色实例 A405 进入离线状态时不保持可用,则管理角色可能不同意使 UD 离线。以此方式,管理角色可以利用拒绝允许租户任务在角色实例 A405 上继续的消息来对结构控制器作出响应。考虑来自管理角色的传递拒绝租户任务的消息,结构控制器智能地决定如何部分地基于租户任务类型来继续。例如,如果租户任务类型包括对主存环境的基础平台的维修,则结构控制器可不管管理角色所试图的拒绝而继续租户任务在角色实例 A405 上的执行。如果租户任务类型包括仅对该服务应用的更新,则结构控制器可留意管理角色的拒绝,并取消租户任务在角色实例 A405 上的执行。如果租户任务类型包括对主存角色实例

A405 的节点的 OS 的维护动作,则结构控制器可试图与管理角色协调使角色实例 A405 进入离线状态。在一个实例中,协调包括管理角色请求暂停,暂时中断租户任务的执行以允许角色实例 A405 将其内部状态复制到针对租户任务当前被选定为目标的 UD 之外的另一角色实例。在另一实例中,协调包括调停故障转移,该调停故障转移通过可控地延迟使角色实例 A405 进入离线状态,且通过复制其内部状态以在 UD 上产生冗余,同时使该 UD 的其它角色实例进入离线状态。

[0071] 当管理角色和 / 或结构控制器确定可使角色实例 A405 进入离线状态(参见框 435),则可在包含该角色实例 A405 的 UD 上开始工作(参见框 440)。在该工作阶段,结构控制器根据租户任务的定义来执行各种任务(例如执行维护动作)。典型地,在该工作阶段期间,结构控制器进行该租户任务的执行,且不与管理角色交流。一旦完成租户任务的任务,在继续下一阶段之前,角色实例 A405 被允许稳定化,使得其内部状态不再为瞬态。

[0072] 经过角色实例 A405 稳定化之后,可恢复 UD (参见框 450)。在恢复阶段期间,由结构控制器通知服务应用的管理角色关于租户任务完成。进一步,在恢复阶段期间,在使 UD 进入在线状态(参见框 455)以及指定角色实例 A405 为经更新的角色实例 A' 415 之前,管理角色负责确保所有受影响的角色实例起作用并且稳定。如所能看到,在 UD 准备阶段 430 与 UD 恢复阶段 450 期间,存在服务应用的管理角色与结构控制器之间的协调。具体而言,在这两个阶段 430 和 450 期间,管理角色计算服务应用的可用性以确定租户任务是否继续至下一阶段。取决于任务类型和服务应用账户设定,该应用可以或不阻止该任务执行,但在任何情况下,它可推迟该任务执行。

[0073] 有时候,可在租户任务待决时(例如角色 A405 正在阶段 430、440 或 450 中的一个处进行处理)发起高优先级任务。当高优先级任务被检测为发起时,待决的租户任务被抢先。在示例性实施例中,抢先可包括将租户任务的执行让位,允许高优先级任务在角色 A405 上执行,并在高优先级任务执行完成时继续租户任务。

[0074] 在抢先期间,例如,调度组件(参考图 5 的标记 530)可向节点状态驱动器(参考图 5 的附图标记 460)发送让位请求,节点状态驱动器负责推进角色 A405 处的阶段 430、440 和 450。该让位请求暂时地挂起待决的租户任务,导致租户任务停止进展,同时调用高优先级任务以在角色 A405 上开始。换言之,调度组件指示节点状态驱动器执行适当的操作,这些操作使租户任务从活动状态转换为不活动状态,且使高优先级任务从不活动状态转换为活动状态,因为典型地,每次仅一个任务在角色 A405 上执行。这些适当的操作可包括强行使角色 A405 进入在线状态,使得角色 A405 基本恢复,并且使角色 A405 进入离线状态,使得高优先级任务可在角色 A405 上进行工作。

[0075] 完成高优先级任务之后,当租户任务正在等待继续时,可由调度程序确定是否重新召集被中断的租户任务。作为示例,可测量角色 A405 的状态(例如在角色 A 上运行的软件的当前版本),并基于该状态,租户任务可在角色 A405 上重新召集或被取消。因此,抢先于待决租户任务的进程可包括多次确定和可改变(例如添加、去除、修改)租户任务的典型过程内的阶段 430、440 和 450 的动作。

[0076] 用于实现租户任务的组件

[0077] 参考图 5,框图示出了根据本发明的实施例的云计算网络 500,包括用于执行实现租户任务的分立操作的多个组件。一般而言,网络 500 描绘了租户改变管理器 580 与服务

应用 350 的管理角色 597-599 之间的交互的一个示例。该交互可影响网络 500 上和 / 或服务应用 350 上的租户任务的实现的协调。

[0078] 最初,租户改变管理器 580 代表图 3 的结构控制器 295 的组件,负责与服务应用 350 的管理角色 597-599 交互,并部分地基于那些交互来指导租户任务的部署。指导部署的过程一般包括:将租户任务部署的指示传达至向管理角色 597-599;发起将角色实例 521-523 停用至离线状态;允许租户任务在角色实例 521-523 上执行;以及发起将角色实例 521-523 恢复至在线状态。作为服务应用 350 的组件的管理角色 597-599 典型地分别配置用于监测角色实例 521-523 的内部状态,角色实例 521-523 被分组在更新域 305 内。在实施例中,在接收到租户任务部署的指示后,管理角色 597-599 负责确定所监测到的角色实例 521-523 的内部状态是否分别在服务应用 350 的不受租户任务影响的部分上(诸如更新域 310 的节点 A、B 和 C 上)被复制。如本文中所描绘,租户任务最初选定更新域 305 作为目标,这影响分别在节点 X、Y 和 Z 上被主存的组件 501、502 和 503 的可用性。此外,管理角色 597-599 可被配置用于将内部状态的可用性指示传达给租户改变管理器 580。

[0079] 现在将参考图 5 来详细描述租户改变管理器 580。如上所述,租户改变管理器 580 一般负责控制对云计算网络 500 的基础平台的更新和 / 或维修(例如调度和组织协调租户任务)。控制的过程可包括租户改变管理器 580 管理管理角色 597-599 与目标状态机构 510 之间的交互。典型地,每个租户任务与特定的目标状态机构相关联,该目标状态机构作为驱动执行该租户任务所固有的工作的状态机来操作。因此,由租户改变管理器 580 管理的交互可包括以下步骤:调度针对更新域 305 的节点 X、Y 和 Z 的租户任务;确定节点 X、Y 和 Z 上主存的组件 501、502 和 503 的身份;识别组件 501、502 和 503 分别包括作为服务应用 350 的部分的角色实例 521、522 和 523 后,将执行所调度租户任务的意图传达至管理角色 597-599;接收来自管理角色 597-599 的消息,该消息传达与角色实例 521、522 以及 523 的内部状态有关的信息;以及通过考虑角色实例 521、522 和 523 的内部状态来处理租户任务的执行。在操作中,角色实例 521、522 和 523 的内部状态指导目标状态机构 510 关于它在使节点 X、Y 和 Z 从在线状态转换成离线状态时应当如何继续。

[0080] 通过给管理角色 597-599 分别分配监测角色实例 521-523 的内部状态的责任,允许租户改变管理器 580 保持对服务应用 350 的实现的具体细节不知晓,以便高效地操作。因此,管理角色 597-599 与租户改变管理器 580 的合作在结构控制器与应用 350 之间有效地引入了通用协议,该协议能实现所有以下操作:将与服务应用 350 的状态有关的充分信息提供给基础平台以满足任何应用管理目标;向服务应用 350 提供对由基础平台发起的任何正在进行的维护活动的可见性;以及向服务应用 350 提供正确地定时平台发起的活动的机会。

[0081] 现在将相对于图 5 的管理角色 597-599 来详细描述示例性管理角色的性能。一般而言,示例性管理角色被编程为理解本地存储在服务应用 350 的一个或多个角色实例处的内部状态,并在网络 500 的基础平台上部署租户任务(例如维护、更新以及恢复操作)期间传达该理解。因此,如本文中所使用,短语“管理角色”泛指服务应用中的负责监测和评估服务应用的内部状态的组件。此外,示例性的管理角色可负责与结构控制器协调操作,该结构控制器在执行租户任务时被涉及。

[0082] 将这些责任委派给管理角色后,结构控制器就被提供对内部状态的可访问查看,

内部状态对于服务应用的角色实例而言是本地的。因此,角色实例内部状态的复制可以是已知的并由结构控制器考虑。实际上,通过收集和分发描述角色实例内部状态的信息,管理角色试图帮助确保在租户任务期间使节点进入离线状态时,结构控制器将不会使特定内部状态的所有复制都离线。因此,可改进在结构控制器直接监测数据中心内的所有有状态应用的所有角色实例的内部状态时将会发生的低效率,不过经由管理角色使内部状态对于结构控制器仍然可见。

[0083] 在一个示例中,如果服务应用 350 具有角色 A (例如接收来自顾客的一组数字图像)、角色 B (例如接收后格式化这些数字图像)、以及角色 C (例如将经过格式化的数字图像在 UI 显示器上呈现),则可创建跟踪角色 A、B 和 C 中实例的内部状态(例如特定顾客数据)的第四组角色 D。因此,管理角色 D (例如管理角色 597-599)具有对服务应用 350 的内部细节的密切理解。例如,管理角色 D 可以知晓数字图像的每个复制品存储的位置以及存储该数字图像的角色实例(例如角色实例 521-523)当前是否可用。此外,管理角色 D 可能知晓个体地本地存储该数字图像的角色实例的数量。

[0084] 在实施例中,可通过以下步骤来建立管理角色 597-599:将服务应用 350 标识为结构控制器已将其选择用于与其组成高可见性的有状态应用;以及指定该应用 350 的一个或多个角色实例(例如管理角色 597-599)来承担理解构成该服务应用 350 的其它角色实例(例如角色实例 521-523)的内部状态的责任。一般而言,从服务模型的立场,管理角色 597-599 代表被服务模型定义中的标志指定的常规服务应用角色。此外,管理角色 597-599 建立后,服务应用 350 可配置管理角色 597-599 中的一个或多个以监测和报告服务应用 350 的多个独立部分。

[0085] 典型地,服务应用 350 被编程为决定将管理角色 597-599 置于由结构控制器推行的一组条件内的何处。例如,结构控制器可施加两个管理角色不能共享同一节点的条件,从而保持多个角色实例高度可用。在该实例中,参考图 5,未在与被监测的其它角色实例 521-523 (即服务角色实例相对于管理角色实例)相同的节点 X、Y 和 Z (例如物理机或虚拟机)上建立管理角色 597-599。

[0086] 建立并放置管理角色 597-599 后,通常使管理角色 597-599 能够通过向租户改变管理器 580 显露管理角色 597-599 对角色实例 521-523 的内部状态的理解,而能够影响被监测的角色实例 521-523 上的租户任务的协调,其中租户改变管理器 580 如以上讨论地那样管理租户任务的实现。对角色实例内部状态的理解可通过爬行(crawl)角色实例 521-523 和任意地实现复杂逻辑来实现,该复杂逻辑使对服务应用 350 自身可用性的理解基于对内部状态的复制的数量和位置。一旦内部状态的理解已知,则该理解被传达至租户改变管理器 530。在由管理角色 597-599 采用的用于传达角色实例 521-523 的内部状态的理解的方案示例中,管理角色 597-599 可评估角色实例 521-523,并传达封装租户改变管理器 580 所需要的评估的消息。或者,评估和传达的过程可由管理角色 597-599 自动地进行(例如在预定的时间段或在检测到触发事件时)。

[0087] 在实施例中,租户改变管理器 580 可包括用于执行与协调租户任务相关联的特定任务的多个组件。如图 5 所示,示例性组件包括执行引擎 520、调度组件 530、租户任务管理组件 540 以及任务相容性组件 550。任务相容性组件 550 配置成处理租户任务的抢先过程,其中一个任务(后部署)比另一待决租户任务(先部署)优先级高。租户任务管理组件 540 协

调租户任务的部署、将指令发送至调度组件 530、并跟踪租户任务的进展。调度组件 530 一般负责调度租户任务的执行,同时将当前在不同服务应用上执行的待决租户任务的类型考虑在内。此外,调度组件可进一步配置用于根据优先级方案使一个或多个租户任务被高优先级任务中断(以下参考图 6 讨论)。

[0088] 在实施例中,执行引擎 520 代表通用有限状态(FSM)机,配置成调用租户任务的发起和执行。典型地,执行引擎 520 包括描述用于执行特定租户任务的工作而进行的一组步骤的框架。该框架也可提供该组步骤之间的排序以及与从一个步骤进行到另一步骤的相关联的链接。这些链接涉及在移动至另一步骤之前在执行第一步骤时应当满足的条件。例如,执行引擎 520 可读取来自调度组件 530 和 / 或目标状态机构 510 的输入和输出,以驱动更新域 305 的状态转换,其中在更新域 305 处的状态转换的完成允许在该更新域上正确地执行租户任务,并进展至更新域 310。

[0089] 在特定实施例中,当充当 FSM 时,执行引擎 520 由与租户任务的转换(例如不活动状态和活动状态)以及角色实例的转换(例如在线状态和离线状态)相关联的有限数量状态组成,其中每个转换包括在一个状态开始和在另一状态结束的一组动作。转换典型地由触发器发起,该触发器诸如检测到的事件(例如写 / 存储操作)。在一个实例中,触发器可包括接收到来自调度组件 530 的提示执行引擎 520 在更新域 305 和 310 上执行租户任务的指令。一般而言,FSM 允许一次仅驻留在一个状态中,该状态在本文中称为待决租户任务的当前状态。

[0090] 网络 500 的基础平台包括辅助租户改变管理器 580 智能地部署和执行租户任务的附加组件。这些附加组件可包含在结构控制器的范围内,或这些组件可脱离结构控制器单独存在(不过仍支持结构控制器)。在实施例中,基础平台的组件包括但不限于以下组件:根升级引擎 590、目标状态机构 510 以及节点状态驱动器 460。

[0091] 根升级引擎 590 一般设计成选择一组节点(例如更新域 310 的节点 A、B 和 C 或更新域 305 的节点 X、Y 和 Z),该组节点可在不违犯服务应用 350 的基础平台或服务模型的预定规则的情况下被更新。例如,这些规则可确保被选择用于接收租户任务的一组节点主存属于不多于一个更新域的角色实例。一旦选择了用于接收租户任务的该组节点,就向租户改变管理器 580 提出所选的该组节点以指导租户任务的部署。

[0092] 目标状态机构 510 一般与管理角色 597-599 通信,且配置成将角色实例 521-523 转换至如执行引擎 520 所指示的预期状态并保持该预期状态。作为示例,在更新域 305 的准备阶段(参见图 4 的附图标记 430)期间,角色实例 521-523 的预期状态可以是离线状态 435。另一方面,在更新域 305 的恢复阶段(参见图 4 的附图标记 450)期间,角色实例 521-523 的预期状态可以是在线状态 455。

[0093] 在目标状态机构 510 的操作的一个示例中,执行引擎 520 和 / 或调度组件 530 可将指令 560 发送至目标状态机构 510,请求目标状态机构 510 将更新域 305 转换至预期状态。如本文中所使用,短语“预期状态”一般代表角色实例 521-523 采取离线状态或在线状态的目标。一旦采取了适当的行动以将角色实例 521-523 转换至预期状态,则可通过线路协议(示为无箭头的虚线)将角色实例 521-523 的实际状态报告给相应的主机代理(作为角色实例 521-523 驻留在公共节点上),相应的主机代理将实际状态转发至租户改变管理器 580。租户改变管理器 580 进而可将租户任务的实际状态传递给目标状态机构 510,以辅助

目标状态机构 510 确定如何继续执行正在进行的租户任务。

[0094] 在另一实施例中,更新域 305 的节点 X、Y 和 Z 到达预期状态后,目标状态机构 510 可将角色实例 521-523 的状态传达至管理角色 597-599。在离线状态是预期状态的实施例中,通信可包括发送第一消息,该第一消息通知管理角色 597-599 关于应当考虑停用角色实例 521-523,以使服务应用 350 准备好不指望使用角色实例 521-523。在在线状态是预期状态的实施例中,通信可包括发送第二消息,该第二消息通知管理角色 597-599 关于应当考虑启用角色实例 521-523,以使服务应用 350 能够依赖于角色实例 521-523。

[0095] 在一些实例中,传递预期状态的消息被直接转发至运行在受租户任务影响的节点 X、Y 和 Z 的相应根分区上的主机代理。例如,在将这些消息分发至适当的管理角色 597-599 之前,这些消息可被发送至图 6 的主机代理 615。同样,主机代理用于向服务应用 350 揭示角色实例 521-523 的状态。进一步地,通过在节点 X、Y 和 Z 上本地存储这些消息,这些消息能够在结构控制器故障转移下和 / 或主存管理角色 597-599 的节点分别离线(例如在重启或重铺操作期间)时存活。因此,网络 500 的基础平台被设计成,不管当前在进展中的租户任务如何,都保持对消息的一致可见性。

[0096] 在实施例中,经由消息保持的预期状态可被转移至主存服务应用 350 的管理角色的实例的每个节点。然后,管理角色 597-599 可实现用于通过规则地检查这些消息来周期性地轮询更新域 305 和 310 的预期状态的机制。这些轮询用于检索预期状态和其它信息,并使其对服务应用 350 可用。

[0097] 一旦目标状态机构 510 设定并传达预期状态,节点状态驱动器 460 就一般配置用于根据预期状态在租户改变管理器 580 的指导下将角色实例 521-523 停用至离线状态、执行租户任务、以及将角色实例 521-523 恢复至在线状态。

[0098] 遵照本发明的实施例,节点可代表任何形式的计算设备,诸如例如个人计算机、台式计算机、膝上型计算机、移动设备、消费电子设备、服务器、图 1 的计算设备 100 以及类似物。典型地,该节点包括或连接至一些形式计算单元(例如中央处理单元、微处理器等等)以支持角色实例和运行在该节点上的其它组件的操作。如本文中所使用的那样,短语“计算单元”通常是指具有处理能力和存储存储器的专用计算设备,该专用计算设备支持一个或多个操作系统或其他基础软件。在一个实例中,计算单元被配置为有形硬件元件或机器,有形硬件元件或机器作为节点的一部分或操作耦合至节点,以使每个端点能执行多种处理和操作。在另一实例中,计算单元可包含耦合至节点所容纳的计算机可读介质的处理器(未示出)。一般而言,计算机可读介质至少临时地存储可由处理器执行的多个计算机软件组件(例如角色实例 521-523)。如本文所使用的,术语“处理器”不旨在是限制性的,并且可包含具有计算能力的计算单元的任何要素。在这种能力中,处理器可被配置成处理指令的有形物品。在示例性实施例中,处理可以包括取指令、解码 / 解释指令、执行和写回指令。

[0099] 如图 5 所示,更新域 305 的节点 X、Y 和 Z 和更新域 310 的节点 A、B 和 C 各自代表虚拟机或物理机。当代表虚拟机时,这些节点可包括图 1 的存储器 112 的部分和 / 或图 1 的处理器 114 的部分。如在此所使用的那样,短语“虚拟机”无意为限制性的,并且可以指由处理单元执行以支持服务应用的功能的任何软件、应用、操作系统或程序。在实施例中,虚拟机代表节点的相应的访客分区,其中该访客分区被启用以主存服务应用 350 或服务应用 350 的至少一部分。

[0100] 一般而言,服务应用 350 由以分布方式主存服务应用 350 的云计算网络 500 的租户(例如顾客)所拥有。该节点可进一步支持运行在该节点上的主机代理(参见图 6 的附图标记 615)。在实施例中,主机代理 615 驻留在该节点的根分区上,其中根分区一般负责管理来自虚拟机的请求,诸如执行盘 I/O 操作的请求。

[0101] 应理解和领会,图 5 中示出的云计算网络 500 仅仅是一个合适的计算系统环境的实例,且并不旨在暗示关于本发明的实施例的使用范围或功能的任何限制。例如,云计算网络 500 可以是公共云、私有云或专用云。云计算网络 500 不应被解释为具有与本文中示出的任何单个组件或组件的组合有关的任何依赖性 or 要求。此外,尽管为了清楚起见用线条示出了图 5 的各个框,但是在实际上,各组件的轮廓并不是那样清楚,并且比喻性地来说,线条更精确地将是灰色的和模糊的。另外,可以采用任何数目的物理机、虚拟机、数据中心、端点或其组合来在本发明实施例的范围内实现所期望的功能。

[0102] 云计算网络 500 典型地包括数据中心,配置成主存和支持由云计算网络 350 的租客/顾客所拥有的分布式服务应用的角色实例的操作。如本文所使用的短语“服务应用”泛指在云计算网络 500 之上运行或访问云计算网络 500 内的存储位置的任何软件或软件的部分。在一个实施例中,角色实例(例如角色实例 521-523)可代表参与支持服务应用(例如服务应用 350)的功能的软件一部分或组件程序。应理解和领会,图 5 中示出的每个角色实例仅仅是支持服务应用的合适部分的示例,且并不旨在暗示关于本发明实施例的使用范围或功能的任何限制。

[0103] 在云计算网络 500 的情境下,所示出的多个组件可在节点内部通信、通过横跨数据中心动态建立的连接在物理节点上通信、以及通过物理网络拓扑与远程网络资源(例如企业专用网络)外部地通信。这些连接可包括经由网络云(未示出)互连在数据中心的物理资源上分布的组件。网络云互连这些资源,使得一个组件可识别另一组件的位置以在它们之间建立通信。例如,网络云可在连接公共服务应用的角色实例的信道上建立该通信。举例来说,这些信道可包括但不限于一个或多个局域网(LAN)和/或广域网(WAN)。这样的联网环境常见于办公室、企业范围计算机网络、内联网和因特网中。因此,不在此进一步描述该网络。

[0104] 此外,可采用任何数量的组件在本发明实施例的范围内实现所期望的功能。尽管为了清楚起见用线条示出了图 5 的各组件,但是在实际上,各组件的轮廓并不是那样清楚,并且按比喻的说法,线条更精确地将是灰色的或模糊的。进一步,尽管将图 5 的一些组件描绘为单个框,但这些描绘在本质上和数量上是示例性的,且不应理解为限制性(例如,尽管仅示出了一个执行引擎 520,但可建立更多执行引擎以管理与各个租户任务相关联的转换)。

[0105] 优先和抢先于租户任务

[0106] 现在转到图 6,根据本发明的实施例示出了描绘结构控制器 295 和服务应用 250 之间的交互的数据中心的示例性拓扑的图形表示。最初,结构控制器 295 可保持当前待决的租户任务(例如租户任务 611-613)的任务列表 610。调度组件 530 配置成通知服务应用 350 的至少一个管理角色 620 关于影响服务应用 350 的那些当前活动的和待决的租户任务以及在任务列表 610 内出现的任何相关修改(例如租户任务 611-613 的重新排序或待决租户任务的抢先)。

[0107] 有时候,如上所述,可在租户任务正在待决并被部署在活动状态时发起高优先级任务。管控如何处理高优先级任务的策略 605 对于调度组件 530 可用,其中不同的策略 605 可导致用于执行租户任务的不同方式。在操作中,策略 605 管控调度组件 530 对围绕租户任务 611-613 的冲突的反应(例如超时和取消消息)。在实施例中,该反应由策略 605 所揭示的租户任务因素驱动。

[0108] 这些因素之一涉及发起租户任务的实体的类型。租户任务可被分成两类:应用发起或顾客发起;以及基础结构发起。当顾客向他们的服务应用 350 发起租户任务(例如版本更新)时,该租户任务被放置在任务列表 610 中,并向管理角色 620 显露。一旦顾客发起的租户任务被发起(例如在到达任务列表 610 的次序顶部时),结构控制器 295 就可查询管理角色 620 以确定当前在由顾客发起的租户任务选定为目标的角色实例上是否存在另一租户任务待决。如果基础结构发起的租户任务(例如由基础平台内部发起的维护操作或包括重启受影响节点的新的根操作系统版本的安装)当前被部署在被选定为目标的角色实例上,则顾客发起的租户任务被挂起(转换至挂起状态)或完全终止(转换至被取消状态)。

[0109] 当基础结构向服务应用 350 发起租户任务时,该任务可被置于任务列表 610 中,或被分配高优先级并且立即被部署。在检测到基础结构发起的租户任务的部署后,结构控制器 295 可查询管理角色 620 以确定当前在由基础结构发起的租户任务所选定为目标的角色实例上是否存在另一租户任务待决。如果顾客发起的租户任务当前被部署在被选定为目标的角色实例上,则可进行若干步骤。一个步骤包括确定顾客发起的租户任务是否被给予任何特权,这些特权将使其优先级排名比基础结构发起的租户任务的优先级高。如果未被给予特权,则另一步骤包括抢先于顾客发起的租户任务。在实施例中,抢先包括使顾客发起的租户任务暂时停止工作(转换至不活动状态),同时允许基础结构发起的租户任务开始(转换至活动状态)。因此,待决的顾客发起的租户任务将不会阻止基础结构发起的租户任务。

[0110] 在其它实施例中,调度组件 530 配置用于建立任务列表 610,并在特定租户任务到达任务列表 610 顶部时调用该特定租户任务的执行。当构建任务列表 610 时,调度组件 530 可依赖于优先级方案来产生任务列表 610 内的租户任务排序。此外,调度组件 530 被配置用于,当在选定共同节点作为目标的两个或多个租户任务之间存在冲突时,选择适当的租户任务以立即执行。该选择可基于咨询优先级方案并查明哪个租户任务应当根据该优先级方案首先发起。典型地,优先级方案包含由基础平台建立的策略 605。

[0111] 现在将讨论用于采用优先级方案的示例性过程。最初,当待决租户任务正在对象角色实例处执行时,新租户任务到达调度组件 530。此时,新租户任务处于不活动状态,待决租户任务处于活动状态,而且对象角色实例处于离线状态。新租户任务的优先级可根据优先级方案来确定。如果所确定的优先级低于待决租户任务的优先级,则调度组件 530 可继续执行待决租户任务(保持于活动状态),同时根据所确定的优先级将新租户任务(保持于不活动状态)插入任务列表 610 的租户任务 611-613 队列中。

[0112] 然而,如果所确定的优先级高于待决租户任务的优先级,则调度组件 530 可通过对待决租户任务执行任务抢先序列的仲裁动作而抢先于待决租户任务的执行。这些仲裁操作可包括将待决租户任务转换至不活动状态,并将新租户任务转换至活动状态。当将待决租户任务转换至不活动状态时,调度组件 530 的目标是使对象角色实例尽快地从当前离线状态进入在线状态。为此,待决租户任务的执行在对象角色实例处被挂起,或被允许完成执

行(例如执行图 4 的工作阶段 440)并通过明确定义的状态转换路径退出。当待决租户任务被挂起时,新的租户任务在对象角色实例上执行,而且在完成新租户任务的执行后,在中断点处在对象角色实例上继续先前待决的租户任务。

[0113] 在调度组件 530 允许待决租户任务在对象角色实例上完成执行的实例中,可实施超时逻辑以确保待决租户任务在预定的时间量中完成。在操作时,在对象角色实例在预定时间量期满之前未能采取在线状态,则超时逻辑可强制待决租户任务过早停止执行,以停止阻止新租户任务的部署。然而,超时逻辑的优势是允许待决租户任务的当前阶段(例如图 4 的阶段 430、440 以及 450)运行以在对象角色实例上不中断地完成,从而使服务应用 350 在抢先期间保持一致状态。

[0114] 在示例性实施例中,由超时逻辑分配的用于了结待决租户任务的预定时间量可部分地基于分配给较高优先级的新租户任务所选定为目标的服务应用的特权级别而改变。换言之,特权级别一般代表允许服务应用行为影响结构控制器 295 在挂起待决租户任务之前所等待的时间的程度。在实例中,与分配给被分配低特权级别的服务应用的短时间量相比,被分配高特权级别的服务应用可被分配较长的时间量来完成“准备”和“恢复”阶段。例如,关键内部应用可被给予比第三方主存应用高的特权级别。基础平台的管理员建立用于执行租户任务的策略 605 后,这些特权级别可被修改。在另一实施例中,被分配给服务应用的特权级别可用于控制其它操作,诸如在租户任务故障的情况下何时产生操作警报。

[0115] 虽然以上讨论了在抢先事件下执行的动作的多个实施例,但应当领会和理解,租户任务的不同优先级可不同地处理,且可在调度组件 530 处调用替代的动作。

[0116] 当在服务应用 350 上执行租户任务时,结构控制器 295 可向管理角色 620 显露服务应用 350 的角色实例的运行时健康信息。该健康信息可经由消息(例如 XML 文档)对管理角色 620 可用,该消息具有结构控制器 295 发布至主机代理 615 的实例的时间戳,该消息然后经由线路协议 616 被发送至服务应用 350。一般而言,线路协议 616 在主机代理 615 和服务应用 350 的角色之间传递通信。如图 6 所示,典型地,存在线路协议 616 与主机代理 615 之间的一对一对应关系。

[0117] 在实施例中,使健康信息可用的过程包括查询服务应用 350 的角色实例(这些角色实例进而将它们的健康信息报告给结构控制器 350),并在允许管理角色 620 查看消息之前聚集健康信息。在查看该消息后,管理角色 620 可检测已遇到问题(例如无法返回在线状态)的一个或多个角色实例,并进行补救动作(例如请求结构控制器 295 重启失效的角色实例)。如此,对角色实例的健康信息的理解允许管理角色 620 评估租户任务的结果,和 / 或计算是否继续至租户任务的下一阶段。

[0118] 在实施例中,健康信息还可包含内部状态信息,该内部状态信息描述服务应用 350 的角色实例处的特定内部状态的存在。一般而言,内部状态信息可用于影响结构控制器 295 发起租户任务在服务应用 350 一部分上的执行的定时。如图 6 所示,仅出于说明目的,如果服务应用 350 包括角色 A 的三个实例(A₁、A₂、和 A₃)和角色 B 的三个实例(B₁、B₂、和 B₃),则内部状态信息可描述角色 A 的哪些实例本地存储第一内部状态,以及角色 B 的哪些实例本地存储第二内部状态。当租户改变管理器 580 向管理角色 620 传达对被选定作为目标以采取离线状态的那些角色实例的指示时,在部署待决租户任务的特定阶段后,管理角色 620 就可咨询健康信息以确定使被选定作为目标的角色实例进入离线状态是否安全。

[0119] 例如,在要求关于使角色实例 A_1 (保持第一内部状态) 和 B_1 (保持第二内部状态) 进入离线状态的反馈时,管理角色 620 可咨询健康信息来确定第一内部状态是否在角色 A 的两个或更多个实例处被复制,以及第二内部状态是否在角色 B 的两个或更多个实例处被复制。当内部状态信息揭示第一和第二内部状态在被选定作为目标的角色实例(A_1 和 B_1) 之外被复制时,管理角色 620 可发出待决租户任务可如所调度地被部署的信号。然而,当内部状态信息揭示第一和第二内部状态未在被选定作为目标的角色实例(A_1 和 B_1) 之外被复制时,管理角色 620 可请求待决租户任务被延迟,以在角色实例 A_2 、 A_3 、 B_2 以及 B_3 中的一个或多个上分别复制第一和第二内部状态。该延迟允许管理角色 620 有时间保留第一和第二内部状态的可用性。如上所述,所请求的延迟由结构控制器 295 考虑,但不一定被允许(例如基于待决租户任务的优先级)。

[0120] 过程流程

[0121] 现在参考图 7,示出根据本发明实施例的用于促进云计算网络的结构控制器与服务应用之间的交互的整体方法 700 的流程图。尽管术语“步骤”和 / 或“框”可在此处用于指示所采用方法的不同元素,但除非而且仅当明确描述了各个步骤的顺序时,该术语不应被解释为意味着此处公开的多个步骤之中或之间的任何特定顺序。方法 700 最初可包括选择包括该服务应用的一个或多个角色实例的第一更新域(UD),如框 710 所示。典型地,角色实例运行在在线条件下,且代表支持该服务应用的功能的相应组件程序(例如该服务应用的角色实例的单个复制)。

[0122] 该方法 700 还可包括准备第一 UD 以用于执行租户任务(例如平台发起的更新、顾客发起的更新、平台发起的维修、或顾客发起的维修),如框 712 所示。在示例性实施例中,准备第一 UD 包括至少以下步骤:通知服务应用内的管理角色关于结构控制器执行租户任务的意图(参见框 714);以及基于确定受租户任务影响的角色实例的内部状态是否在该服务应用的不受租户任务影响的部分上被复制,从管理角色接收响应(参见框 716)。准备第一 UD 的角色实例之后,发起将这些角色实例停用至离线状态,如框 718 所示。一般而言,停用角色实例包括使主存第一 UD 内的一个或多个角色实例的一组节点不工作。一旦角色实例已采取离线状态,允许租户任务在第一 UD 上执行,如框 720 所示。

[0123] 再进一步,该方法 700 可包括,完成租户任务的执行后,将第一 UD 的角色实例恢复至在线状态,如框 722 所示。在示例性实施例中,将第一 UD 的实例恢复至在线状态包括至少以下步骤:验证受租户任务影响的角色实例在起作用(参见框 724);以及通知管理角色关于角色实例上的租户任务的执行完成(参见框 726),由此提示该服务应用继续利用这些角色实例来执行任务。一般而言,将第一 UD 内的角色实例恢复至在线状态还包括使主存这些角色实例的该组节点工作。将第一 UD 内的角色实例恢复至在线状态后,则该方法 700 可通过选择角色实例的第二 UD 以在其上运行租户任务而继续。典型地,第一 UD 和第二 UD 在组员关系上是互斥的,可分别代表在云计算网络上传播租户任务时的分立阶段。

[0124] 现在转到图 8,根据本发明实施例示出用于在执行租户任务时调度高优先级任务的整体方法 800 的流程图。该方法 800 包括接收在服务应用上执行租户任务的指示(参见框 810),并标识包括该服务应用的一个或多个角色实例的更新域(UD)(参见框 812)。在此时,角色实例在在线状态下操作。准备 UD 以用于执行租户任务(参见框 814),而且随后将 UD 的实例停用至离线状态(参见框 816)。在示例性实施例中,准备 UD 以用于执行租户任务

的过程包括多个逻辑步骤,包括以下:通知服务应用内的管理角色关于执行该租户任务的意图;以及接收来自管理角色的肯定响应,或接收来自管理角色的否定响应。

[0125] 在某个后续时间点,接收执行高优先级任务的指示,如框 818 处所示。如本文中所使用,短语“高优先级”不意味着限制性,而是代表被预先确定为在优先级方案内抢先于租户任务的任何任务。在接收到执行高优先级任务的指示后,则如框 820 处所示,租户任务的部署让位于高优先级任务。在示例性实施例中,让位的过程包括多个逻辑步骤,包括以下:指令服务应用执行将角色实例恢复至在线状态的操作的截断集合(参见框 822);将租户任务挂起(参见框 824);以及发起高优先级任务在角色实例上的执行(参见框 826)。

[0126] 完成高优先级任务的执行后,继续租户任务的部署,如框 828 处所示。在示例性实施例中,继续的过程包括多个逻辑步骤,包括以下:重新发起将 UD 的角色实例停用至离线状态(参见框 830);以及允许租户任务在角色实例上执行(参见框 832)。一旦完成租户任务的执行,可将 UD 的角色实例恢复至在线状态,如框 834 处所示。在示例性实施例中,将角色实例恢复至在线状态的过程包括多个逻辑步骤,包括以下:验证受租户任务影响的角色实例在起作用;以及通知管理角色关于角色实例上的租户任务的执行完成,由此提示该服务应用继续利用这些角色实例来执行多个任务。

[0127] 参考具体实施例描述了本发明的实施例,具体实施例在所有方面都旨在是说明性的而非限制性的。在不背离本发明范围的情况下各替换实施例对本发明的各实施例的所属领域的普通技术人员将变得显而易见。

[0128] 从前面的描述可以看出,本发明很好地适用于实现上文所阐述的所有目的和目标,并且具有对于该系统和方法是显而易见且固有的其他优点。可理解的是,某些特征和子组合是有用的,并且可以加以利用而无需参考其它特征和子组合。这由权利要求所构想的,并在权利要求的范围内。

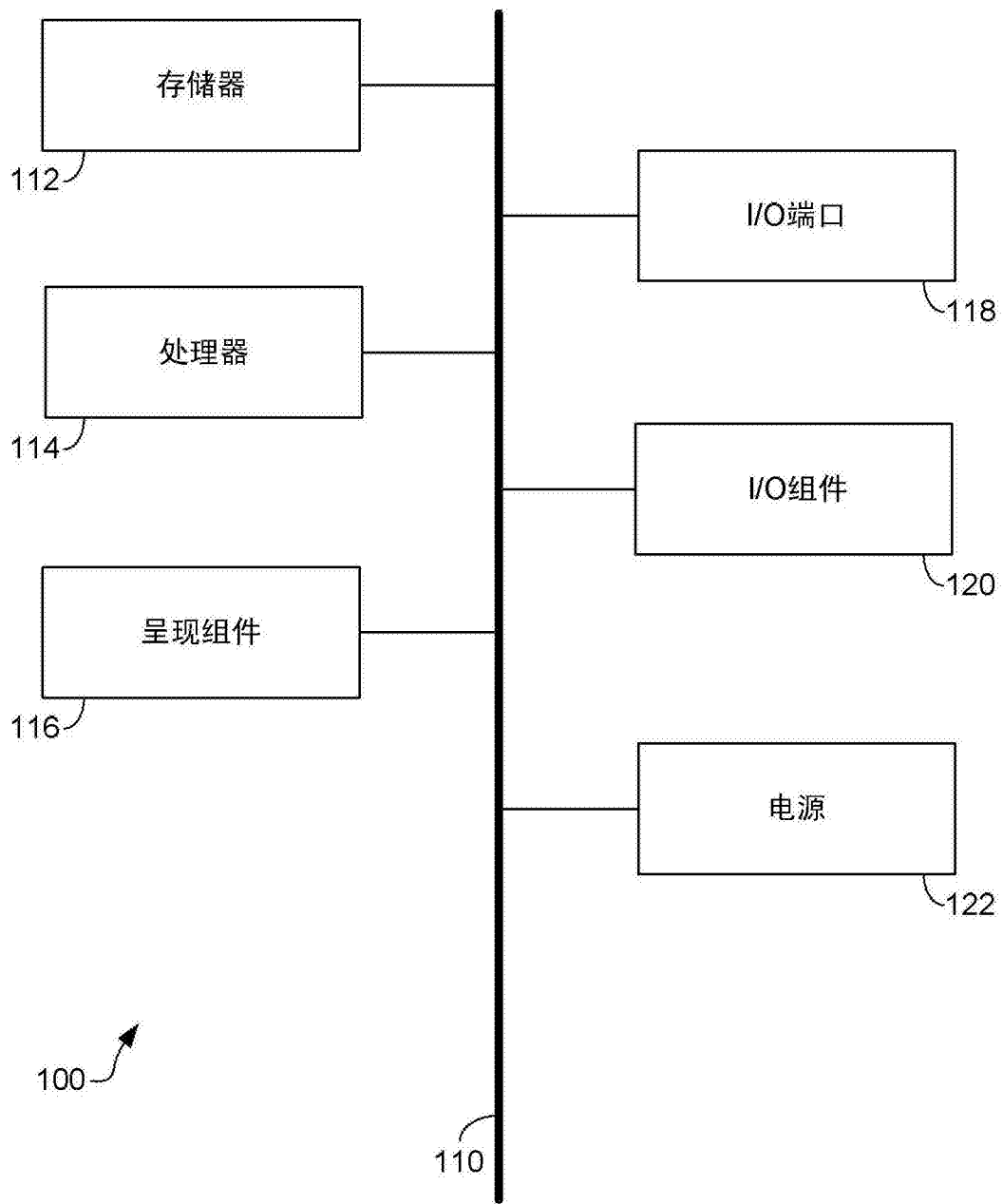


图 1

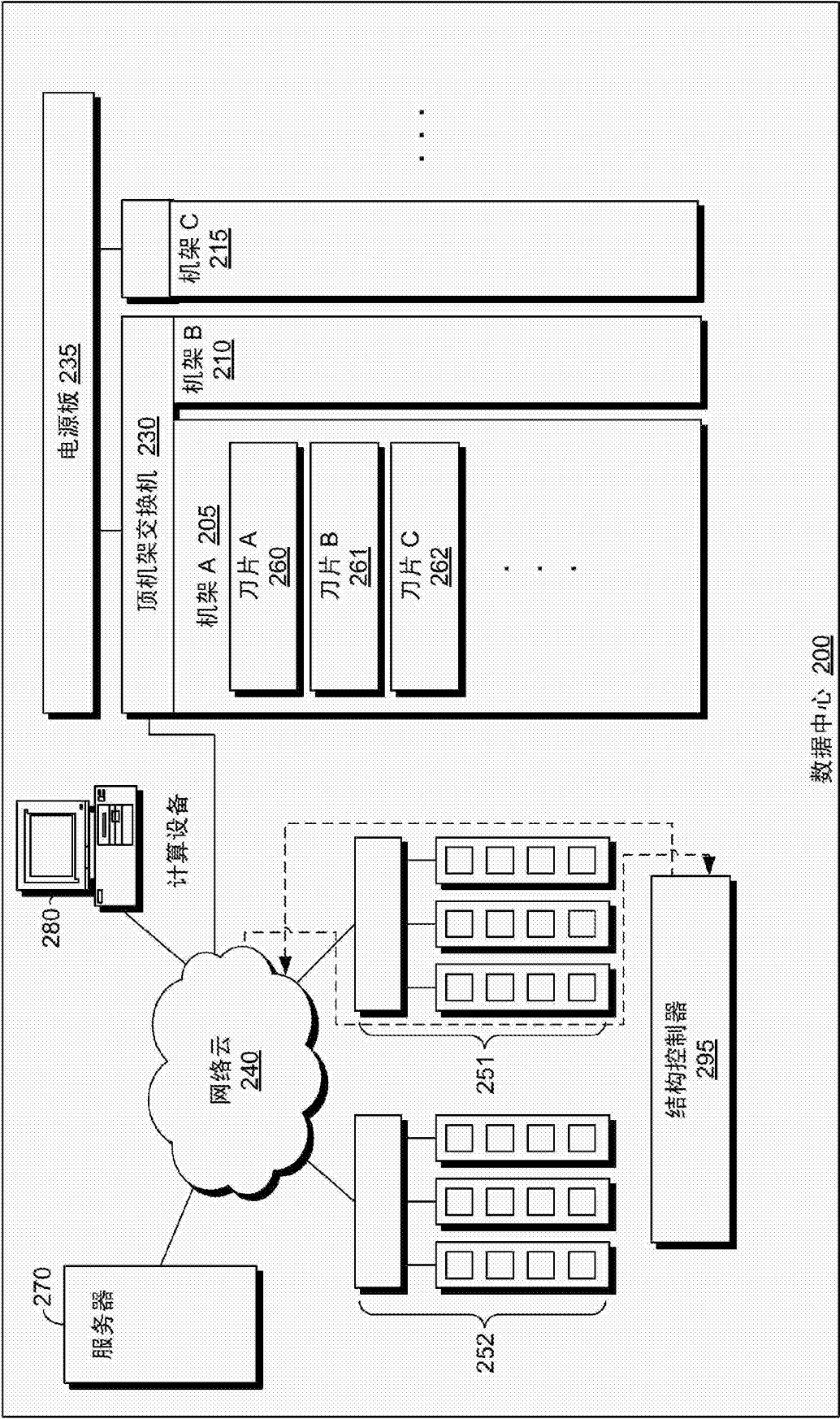


图 2

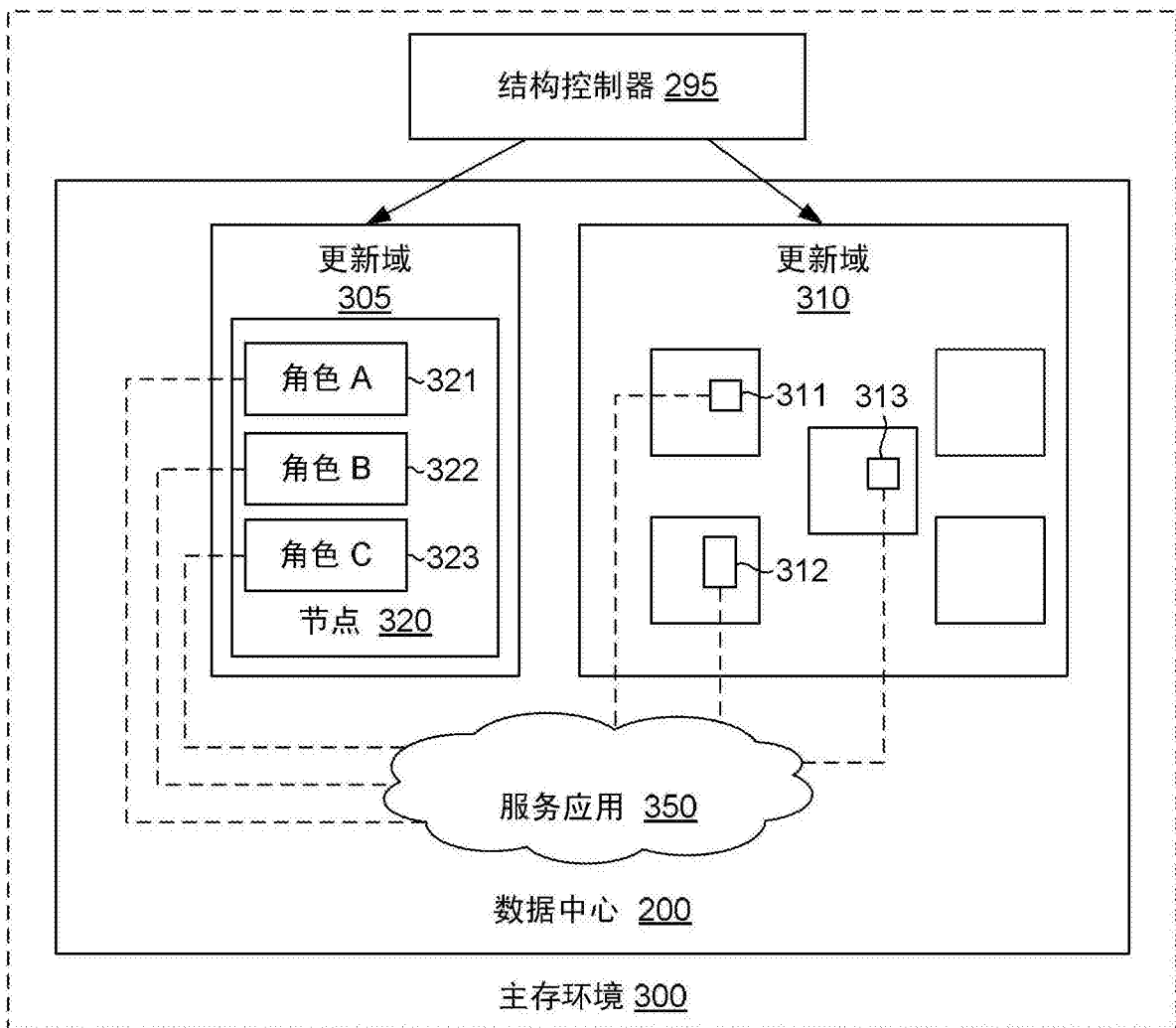


图 3

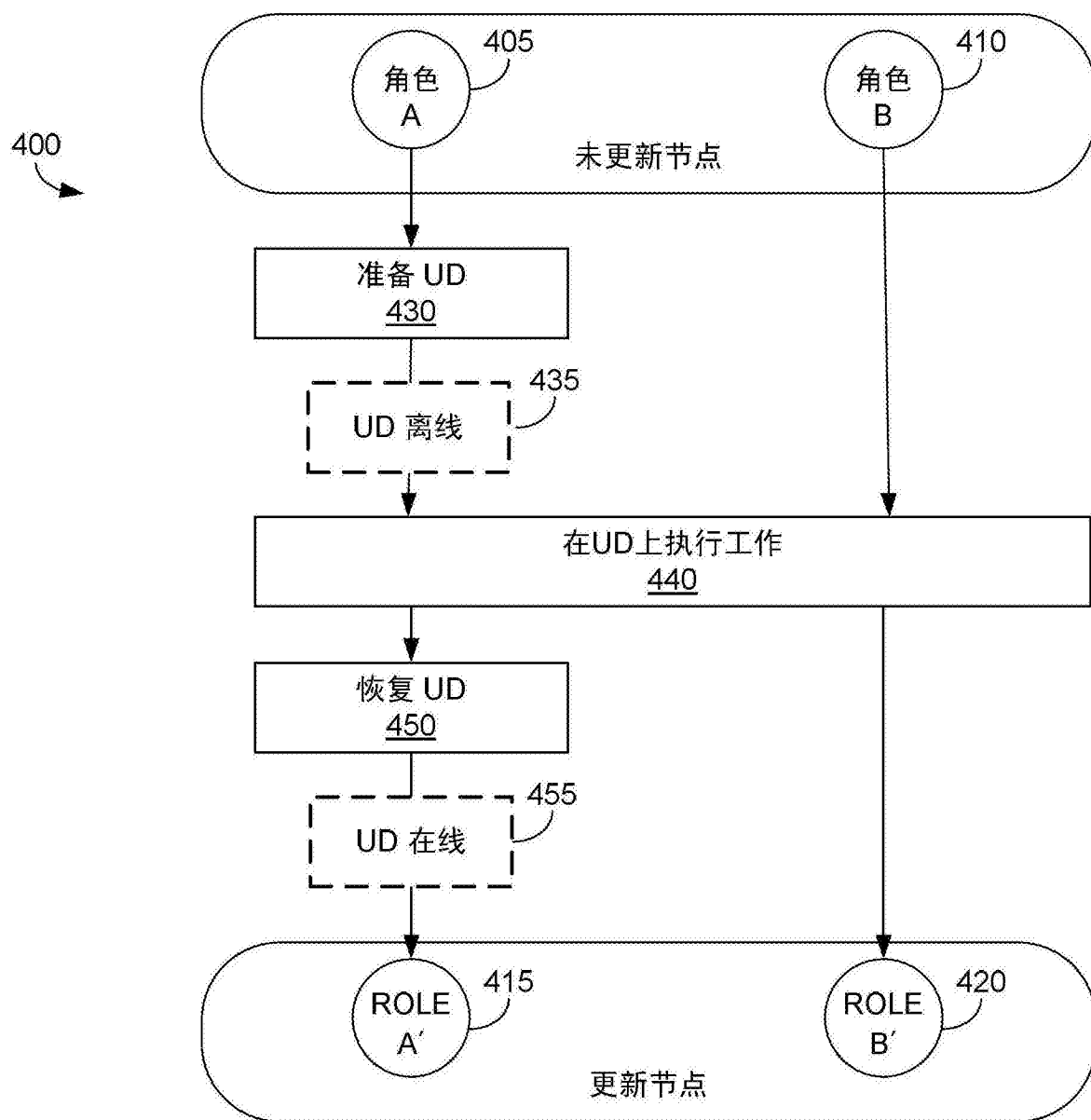


图 4

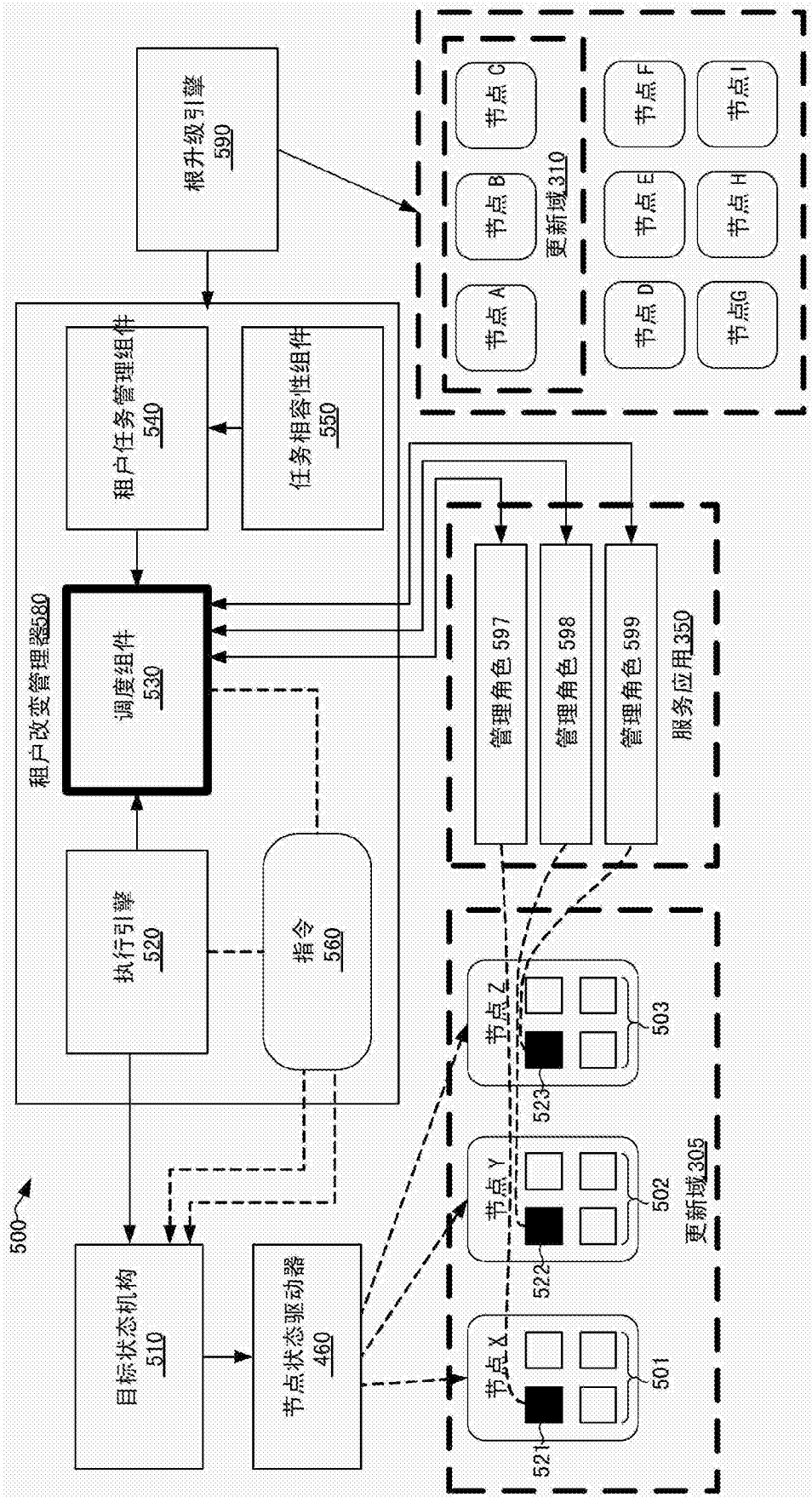


图 5

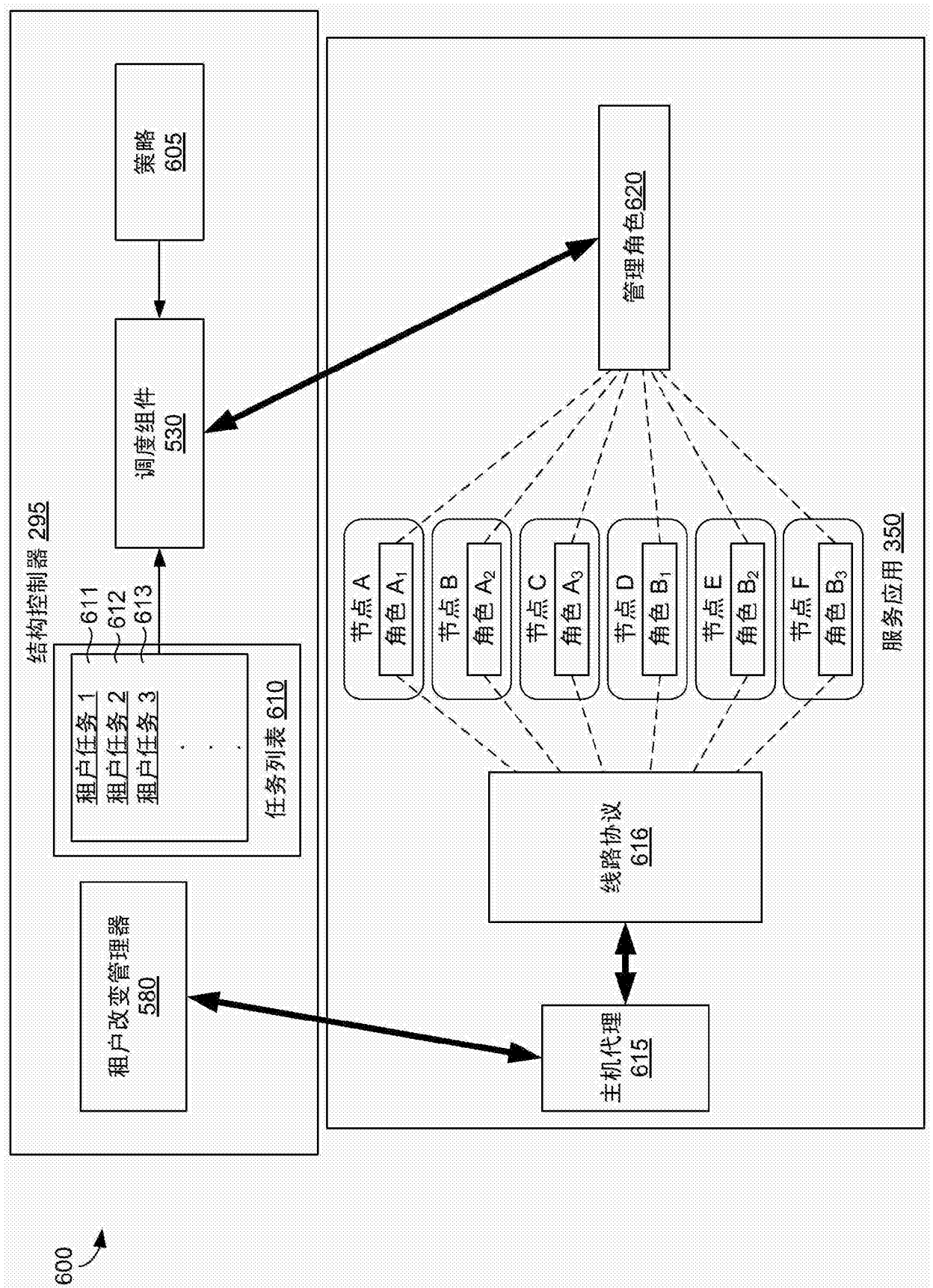


图 6

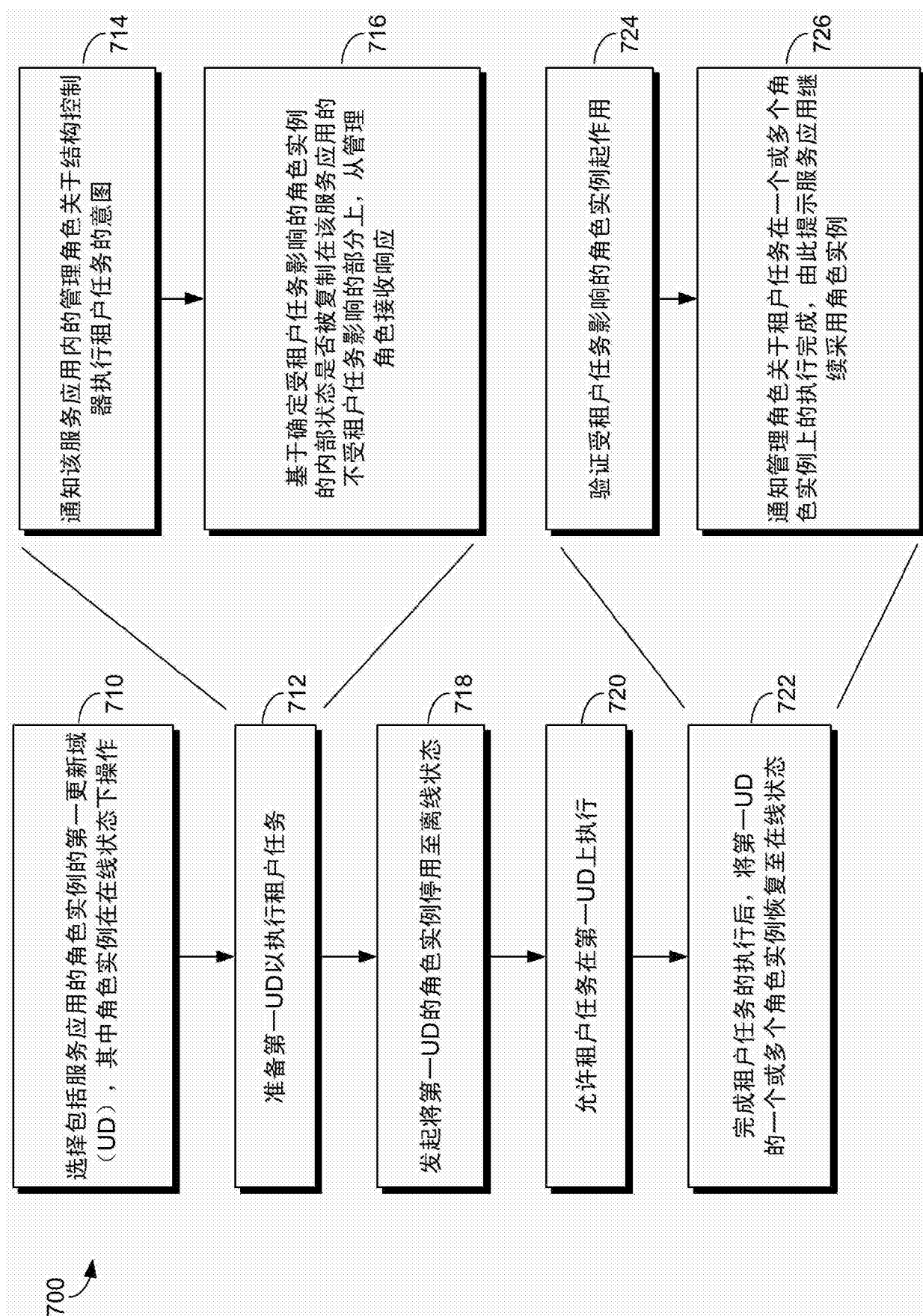


图 7

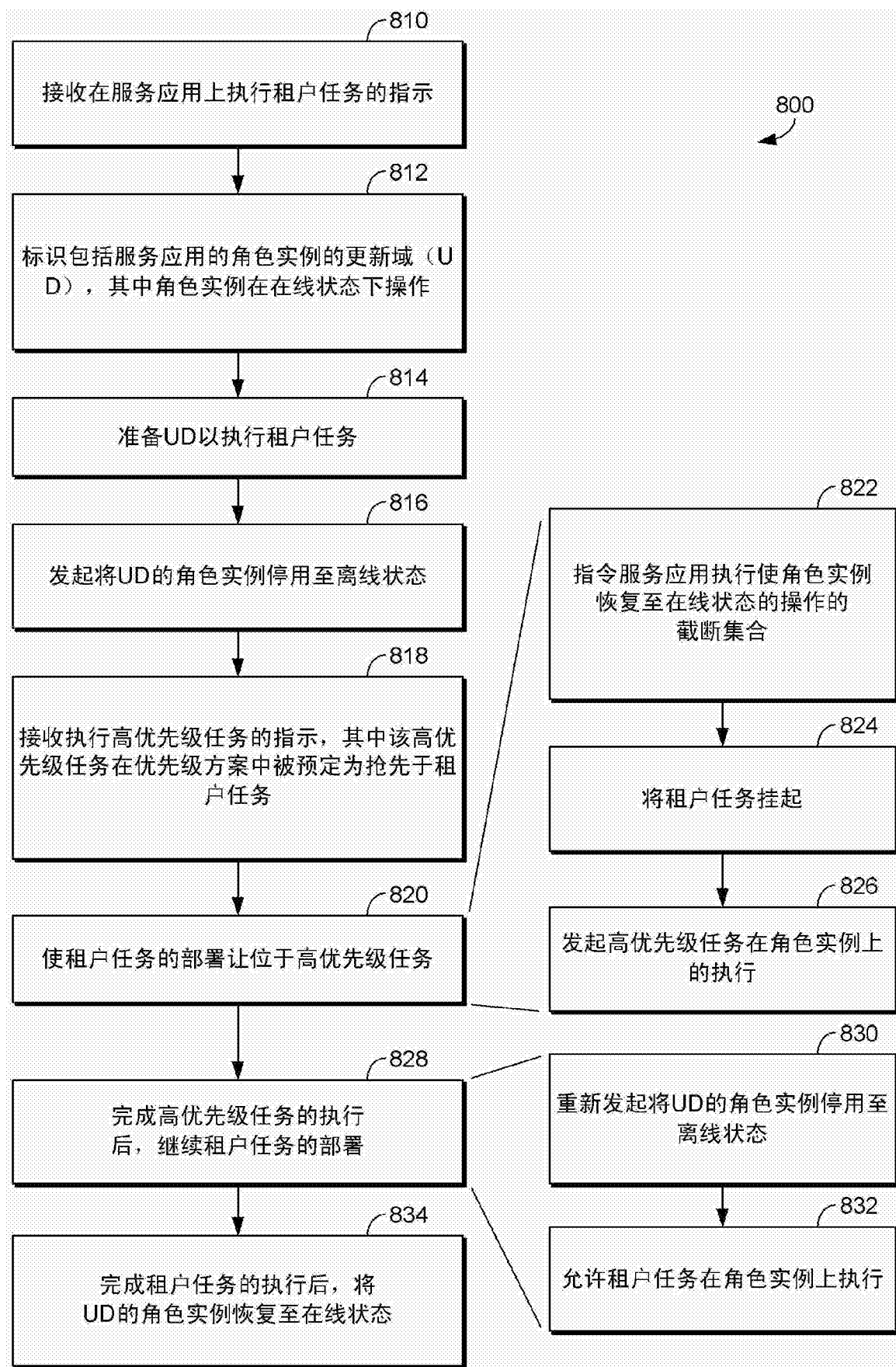


图 8