



(12)发明专利

(10)授权公告号 CN 103713951 B

(45)授权公告日 2018.11.13

(21)申请号 201310537815.9

克里斯托夫·布朗

(22)申请日 2007.03.29

昆顿·R·胡利

(65)同一申请的已公布的文献号

申请公布号 CN 103713951 A

(74)专利代理机构 中科专利商标代理有限责任
公司 11021

(43)申请公布日 2014.04.09

代理人 袁飞

(30)优先权数据

11/395,463 2006.03.31 US

(51)Int.Cl.

G06F 9/50(2006.01)

G06F 9/455(2006.01)

(62)分案原申请数据

200780020255.2 2007.03.29

(73)专利权人 亚马逊技术有限公司

地址 美国内华达州

(72)发明人 罗兰·帕特森-琼斯

克里斯托弗·C·平卡姆

本杰明·托布勒

威廉·R·范比林

加百利·斯密特

(56)对比文件

US 2002/0002613 A1,2002.01.03,

US 6185598 B1,2001.02.06,

US 2004/0261076 A1,2004.12.23,

CN 1675621 A,2005.09.28,

US 2005/0004952 A1,2005.01.06,

审查员 唐佩

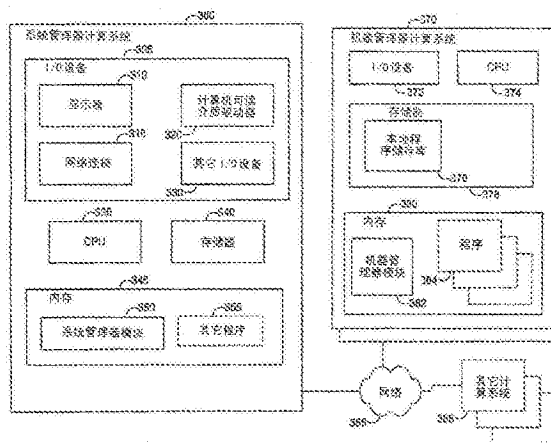
权利要求书3页 说明书17页 附图7页

(54)发明名称

管理由多个计算系统执行程序

(57)摘要

说明了用于管理程序在多个计算系统(诸如被组织成多个组的计算系统)上的执行的技术。程序执行服务代表多个客户或其它用户管理程序的执行,并例如部分基于先前存储的程序的一个以上的副本的、可以从中获取所要执行的程序的副本位置,选择合适的计算系统来执行程序的一个以上的实例。举例而言,在某些情况下,选择合适的用以执行程序实例计算系统是部分基于是否在物理或逻辑上同其它资源(如程序的存储副本、正在执行的程序的副本、和/或可用计算系统)相邻的方式实现的。



1. 一种计算机执行的方法,包括:

程序执行服务的一个或多个计算系统从程序执行服务的客户端接收一个或多个请求,接收到的一个或多个请求至少包括对操作系统的指示以及对要执行的操作系统的实例的数目的指示;

由所述一个或多个计算系统选择程序执行服务的一个或多个计算节点,以用于执行所指示的操作系统;以及

由所述一个或多个计算系统引起一个或多个所选计算节点在一个或多个虚拟机内执行所指示的操作系统的一个或多个实例,执行的引起至少部分基于所接收的对要执行的操作系统的实例的数目的指示。

2. 根据权利要求1所述的方法,还包括:由所述一个或多个计算系统提供接口。

3. 根据权利要求2所述的方法,其中,所提供的接口包括图形用户界面,并且一个或多个请求经由图形用户界面接收。

4. 根据权利要求2所述的方法,其中,所提供的接口包括应用程序接口“API”,并且一个或多个请求的接收基于客户端的远程计算系统对API的调用。

5. 根据权利要求1所述的方法,其中,引起一个或多个所选计算节点执行所指示的操作系统的一个或多个实例包括:引起一个或多个所选计算节点在与所指示的实例数目相等的数目的虚拟机内执行所指示的操作系统的所指示的数目的实例。

6. 根据权利要求1所述的方法,其中,所接收的对要执行的操作系统的实例的数目的指示指定要执行的所指示的操作系统的实例的最小数目或要执行的所指示的操作系统的实例的最大数目中的至少一个,并且引起一个或多个所选计算节点执行所指示的操作系统的一个或多个实例是根据所指定的实例的最小和/或最大数目执行的。

7. 根据权利要求1所述的方法,其中,所接收的对要执行的操作系统的实例的数目的指示指定启动执行所指示的操作系统的一个或多个实例的一个或多个时间,并且引起一个或多个所选计算节点执行所指示的操作系统的一个或多个实例是根据所指定的一个或多个时间执行的。

8. 根据权利要求1所述的方法,其中,所接收的对要执行的操作系统的实例的数目的指示指定终止执行所指示的操作系统的一个或多个实例的一个或多个时间,所述方法还包括:由一个或多个计算机系统根据所指定的一个或多个时间来管理一个或多个虚拟机的执行。

9. 根据权利要求1所述的方法,其中,所接收的对要执行的操作系统的实例的数目的指示指定一个或多个准则,所述一个或多个准则与要用于执行所指示的操作系统的一个或多个计算相关资源有关。

10. 根据权利要求9所述的方法,其中,所指定的计算相关资源准则与以下至少一项有关:所指示的内存容量、所指示的处理器使用量、所指示的网络带宽大小、所指示的硬盘空间大小或所指示的交换空间大小。

11. 根据权利要求9所述的方法,其中,所指定的计算相关资源准则与对将由所述一个或多个虚拟机使用的计算相关资源的一个或多个量的指示有关。

12. 根据权利要求1所述的方法,其中,所述一个或多个所选计算节点是程序执行服务的位于两个或多个不同地理位置的更大量计算节点的一部分,其中,所接收的对要执

行的操作系统的实例的数目的指示包括:对执行所指示的操作系统的一个或更多个地理位置的指示,并且对一个或更多个计算节点的选择至少部分基于所指示的一个或更多个地理位置。

13. 根据权利要求1所述的方法,其中,所述一个或更多个请求标识用户账户,并且所述方法还包括:确定是否授权用户账户执行所指示的操作系统。

14. 根据权利要求1所述的方法,还包括:

由一个或更多个计算机系统产生执行所指示的操作系统的一个或更多个实例的费用信息,并且费用信息至少部分基于一个或更多个所选计算节点执行一个或更多个虚拟机的时间量。

15. 一种计算机执行的方法,包括:

经由程序执行服务的接口接收运行镜像的一个或更多个实例的请求,所述镜像包括程序,并且所述请求还包括对要运行的镜像的实例的数目的指示;

基于所述指示,选择程序执行服务的一个或更多个计算节点以作为一个或更多个实例的主机,其中,使用一个或更多个虚拟机运行所述一个或更多个实例;以及

对被选为运行一个或更多个虚拟机的一个或更多个计算节点进行管理。

16. 根据权利要求15所述的方法,其中,所述接口包括图形用户界面。

17. 根据权利要求15所述的方法,其中,所述接口包括应用程序接口“API”,并且请求的接收基于远程计算系统对API的调用。

18. 根据权利要求15所述的方法,其中,所接收的指示指定要运行的镜像的实例的数目。

19. 根据权利要求15所述的方法,其中,所接收的指示指定:要运行的镜像的实例的最小数目、或者要运行的镜像的实例的最大数目。

20. 根据权利要求15所述的方法,其中,所接收的指示指定:开始运行镜像的一个或更多个实例的一个或更多个时间、或者终止镜像的一个或更多个实例的一个或更多个时间。

21. 根据权利要求15所述的方法,其中,所接收的指示指定与要用于运行镜像的一个或更多个实例的一个或更多个计算相关资源有关的一个或更多个准则,所指定的准则与所指示的内存容量、所指示的处理器使用量、所指示的网络带宽大小、所指示的硬盘空间大小或所指示的交换空间大小有关。

22. 根据权利要求15所述的方法,其中,所述一个或更多个所选计算节点是程序执行服务的位于两个或更多个不同地理位置的多个计算节点的一部分,其中,所接收的指示包括:对要运行镜像的至少一些实例的一个或更多个地理位置的指示,并且对一个或更多个计算节点的选择至少部分基于所指示的一个或更多个地理位置。

23. 根据权利要求15所述的方法,还包括:

确定用于运行镜像的一个或更多个实例的费用,所确定的费用至少部分基于运行一个或更多个实例的时间量。

24. 一种用于程序执行服务的系统,包括:

一个或更多个计算系统,其中,各个计算系统具有一个或更多个处理器;以及

至少一个内存,所述内存包括指令,当所述指令被一个或更多个处理器中的至少一个执行时,使系统:

经由接口从客户端接收请求,所述接口由程序执行服务提供用于配置对程序的执行,所接收的请求包括对要执行的所指示的程序的实例的数目的指示;

选择程序执行服务的一个或更多个计算节点,用于执行所指示的程序;以及

由所选择的一个或更多个计算节点代表客户端对所指示的程序的一个或更多个实例的执行进行管理,对执行的管理至少部分基于所接收的指示。

25.根据权利要求24所述的系统,其中,所述接口包括应用程序接口“API”,并且请求的接收基于客户端的远程计算系统对API的调用。

26.根据权利要求24所述的系统,其中,所接收的指示指定要执行的所指示的程序的实例的数目。

27.根据权利要求24所述的系统,其中,所接收的指示指定以下至少一项:启动执行所指示的程序的一个或更多个实例的一个或更多个时间、或者终止执行所指示的程序的一个或更多个实例的一个或更多个时间。

28.根据权利要求24所述的系统,其中,所接收的指示指定一个或更多个准则,所述一个或更多个准则与要用于执行所指示的程序的一个或更多个计算相关资源有关,所指定的准则包括以下至少一项:所指示的内存容量、所指示的处理器使用量、所指示的网络带宽大小、所指示的硬盘空间大小或所指示的交换空间大小。

29.根据权利要求24所述的系统,其中,对执行的管理包括:在第一时间,启动由所选择的一个或更多个计算节点对所指示的程序的一个或更多个实例的执行;并且所述内存还包括指令,当所述指令被执行时,使系统:

在比第一时间晚的第二时间经由所述接口接收附加信息,所述附加信息指定与执行所指示的程序有关的一个或更多个修改;以及

响应于接收到所述附加信息,动态修改所指示的程序的一个或更多个实例中至少一个实例的执行。

管理由多个计算系统执行程序

[0001] 分案申请说明

[0002] 本申请是申请日为2007年3月29日、申请号为200780020255.2 (国际申请号PCT/US2007/007601) 的、题为“管理由多个计算系统执行程序”的发明专利申请的分案申请。

技术领域

[0003] 以下公开主要涉及管理程序在多个计算机系统上的执行,诸如通过在计算系统组之间以支持高效地获取被执行的程序副本的方式来交换程序副本。

背景技术

[0004] 容纳大量互联计算系统的数据中心目前已十分常见,此类数据中心的示例包括:由单个组织以及代表单个组织运营的私有数据中心,以及由实体(如企业)运营的、在多种商业模式下为客户提供对计算资源的访问权限的公共数据中心。举例而言,某些公共数据中心的运营商为客户所拥有的各种硬件提供网络访问、功能,以及安全安装设施,而其它公共数据中心的运营商则提供“全套服务”设施,后者还包括运营商的客户所使用的实际硬件资源。然而,由于通常的数据中心的规模和范围不断扩大,供应、管理物理计算资源的任务变得越来越复杂。

[0005] 商品硬件虚拟化技术的出现为对管理具有不同需求的大量客户的大规模计算资源的问题提供了部分的解决方案。概况地说,虚拟化技术使多个客户能够有效和安全地共享各种计算资源。举例而言,诸如由VMWare、XEN、或User-Mode Linux提供的虚拟化技术使单个物理计算机能在多个用户之间共享。具体而言,可以为各个用户提供一个以上由单个物理计算机所寄居(hosted)的虚拟机,作为软件模拟的各虚拟机起独立逻辑计算系统的作用。各虚拟机使用户感到自己是给定硬件计算资源的唯一操作者和管理者,同时还在各种虚拟机之间提供应用程序隔离和安全功能。此外,某些虚拟化技术能够提供跨一个以上物理资源的虚拟资源,例如,单个虚拟机可以具有多个实际上跨多个独立物理计算系统的多个虚拟处理器。

[0006] 然而,在数据中心以虚拟或物理的方式寄居了大量针对一组不同客户的应用程序或系统的情况下,将产生一个问题,该问题涉及管理软件应用程序副本的存储、分发、以及获取。举例而言,应用程序可能很大,以至于(即使并非不可能)利用足够的存储资源来存储数据中心中的每个计算系统上的每个寄居应用程序的本地副本的代价是很高的。然而,如果可选地维护一个频繁向数据中心中需要执行那些应用程序的各计算系统发送应用程序副本的集中存储位置,那么就网络带宽资源而言这种方式的代价也是很高的。以这种可选方式,网络带宽将被应用程序复制传输独占,并且将会使正在执行的应用程序无法为其正常工作获得足够的网络带宽。此外,在等待应用程序副本传输等工作完成时,会对应用程序的执行引入很大的启动时延等。此类问题可能因其它因素(如频繁引入所要执行的新应用程序和/或频繁部署应用程序的后续版本)而进一步加剧。

[0007] 因此,鉴于这些问题,以有效的方式向执行应用程序的计算系统提供用于分发应

用程序副本的技术以及提供各种其他优势是十分有益的。

发明内容

[0008] 根据本公开的第一方面,提供了一种计算机执行的方法。所述方法包括:程序执行服务的一个或多个计算系统从程序执行服务的客户端接收一个或多个请求,接收到的一个或多个请求至少包括对操作系统的指示以及对要执行的操作系统的实例的数目的指示;由所述一个或多个计算系统选择程序执行服务的一个或多个计算节点,以用于执行所指示的操作系统;以及由所述一个或多个计算系统引起一个或多个所选计算节点在一个或多个虚拟机内执行所指示的操作系统的实例,执行的引起至少部分基于所接收的对要执行的操作系统的实例的数目的指示。

[0009] 根据本公开的第二方面,提供了一种计算机执行的方法。所述方法包括:经由程序执行服务的接口接收运行镜像的一个或多个实例的请求,所述镜像包括程序,并且所述请求还包括对要运行的镜像的实例的数目的指示;基于所述指示,选择程序执行服务的一个或多个计算节点以作为一个或多个实例的主机,其中,使用一个或多个虚拟机运行所述一个或多个实例;以及对被选为运行一个或多个虚拟机的一个或多个计算节点进行管理。

[0010] 根据本公开的第三方面,提供了一种用于程序执行服务的系统。所述系统包括:一个或多个计算系统,其中,各个计算系统具有一个或多个处理器;以及至少一个内存,所述内存包括指令,当所述指令被一个或多个处理器中的至少一个执行时,使系统:经由接口从客户端接收请求,所述接口由程序执行服务提供用于配置对程序的执行,所接收的请求包括对要执行的所指示的程序的实例的数目的指示;选择程序执行服务的一个或多个计算节点,用于执行所指示的程序;以及由所选择的一个或多个计算节点代表客户端对所指示的程序的一个或多个实例的执行进行管理,对执行的管理至少部分基于所接收的指示。

附图说明

[0011] 图1是一幅网络图,示出了多个计算系统在其中交换和执行程序的示例实施例。

[0012] 图2示出了存储和交换程序副本的多组计算系统的示例。

[0013] 图3是一幅框图,示出了适于管理程序在多个计算系统上的执行的计算系统的示例。

[0014] 图4A-4B示出了系统管理器模块的工作过程的实施例的流程图。

[0015] 图5示出了机器管理器模块的工作过程的实施例的流程图。

[0016] 图6示出了程序执行服务客户的工作过程的实施例的流程图。

具体实施方式

[0017] 下面对于对在多个计算系统上执行程序进行管理的技术予以说明。在某些实施例中,所述技术是代表程序执行服务执行的,所述程序执行服务用于代表该服务的多个用户(例如,客户)执行多个程序。在某些实施例中,程序执行服务可以利用多种因素(如选定计算系统可以从中获取所要执行的程序和/或用以执行程序实例的可用计算资源的副本

的、先前存储的一个以上的程序副本的位置),以选择合适的计算系统来执行程序实例。举例而言,在某些实施例中,可以部分基于对已存储了程序的本地副本的计算系统的判断,来选择合适的计算系统来执行程序实例。在另一实施例中,可以部分基于对与每个存储了这样一个本地副本的一个以上的计算系统(在物理上和/或逻辑上)足够邻近的计算系统(如与确定的计算系统属于同一组的一个以上的其它计算系统)的判断,来选择合适的计算系统。

[0018] 在某些实施例中,可用于执行程序的多个计算系统可以包括通过能够在计算机之间交换数据的一个以上网络或其它数据交换媒介相互连接的多个物理计算机。多个计算系统可以,例如,位于同一物理位置(例如,数据中心),还可以被分成多个组,并且可以由一个以上的系统管理器模块以及由多个机器管理器模块加以管理,所述系统管理器模块总体负责所述多个计算系统,所述机器管理器模块各与一组相关,以管理各组中的计算系统。至少某些计算机可以分别包含足够的资源(例如,足够的可写内存和/或一个以上的足够的内存、CPU周期、或其它CPU使用度量、网络带宽、交换空间等),以同时执行多个程序。举例而言,在某些此类实施例中至少某些计算机可以分别寄居多个虚拟机节点,所述虚拟机节点可以分别为独立用户执行一个以上的程序。如前所述,至少在某些实施例中,可以诸如基于在物理或逻辑上十分接近、或具有公共数据交换媒介等准则,将受程序执行服务管理的多个计算系统组织成多个独立的组(例如,每个计算机属于一个单独的组)。在一示例中,组的公共数据交换媒介可由在组中的计算系统之间提供高带宽通信的单个网络交换机和/或机架背板予以提供(例如,与网络交换机或机架背板相连接的部分或全部计算系统可以是组中的成员)。每组计算系统还可以通过一个以上其它数据交换媒介(例如与组的公共数据交换媒介相比带宽较低的其它数据交换媒介(例如,基于以太网的配线、无线连接,或其它数据连接))与其它计算系统(例如,其它组的计算系统,或不受程序执行服务管理的远程计算系统)相连接。此外,至少在某些实施例中,某些或全部计算系统可以分别具有本地程序储存库(例如,硬盘或其它本地存储机制),所述程序储存库可用于例如在执行程序之前或执行程序时,存储执行所需的程序的本地副本。此外,至少在某些实施例中,每组多个计算系统可以使用该组中的一个以上的计算系统来存储程序的本地副本,以供组中其它计算系统使用。

[0019] 在一说明性实施例中,程序执行服务可以包括执行于一个以上的计算系统上的软件设施,以管理程序的执行。软件设施可以包括一个以上的为各组一个以上的计算系统所用的机器管理器模块,所述机器管理器模块管理该组中计算系统的程序获取、程序存储以及程序执行。举例而言,可以为各独立物理计算机配备独立的机器管理器模块,例如物理计算机的机器管理器模块可以执行于该计算机的至少一个虚拟机上。此外,在某些实施例中,软件设施可以包括一个以上执行于一个以上的计算系统上的系统管理器模块,所述系统管理器模块为正在用于执行程序的全部的多个计算系统管理程序的获取、存储以及执行。如以下将更为详细地予以讨论的那样,系统管理器模块可以适当地同机器管理器模块进行交互。

[0020] 至少在某些实施例中,可以对请求立即执行程序实例的当前执行请求予以响应,开始在一个以上计算系统上执行程序的一个以上的实例。可选地,执行的启动可以基于,先前接收到的、预定或保留在将来(当前时刻)执行这些程序实例的程序执行请求。可以通过

多种方式(如直接从用户(例如,通过由程序执行服务提供的交互式控制台或其它GUI),或从自动启动其它程序或其自身的一个以上实例(例如,通过由程序执行服务提供的API(或应用程序接口),如使用Web服务的API)的执行的程序实例)接收程序执行请求。

[0021] 程序执行请求可以包括多种在启动执行一个以上的程序实例过程中所用的信息(如先前注册的或供未来执行的程序的指示,以及需要同时执行的程序实例的数量,如表示为一个单独的实例的期望数量、期望实例的最小和最大数量,等)。此外,在某些实施例中,程序执行请求可以包括多种其它类型的信息,诸如:用户账户的指示、或先前注册的用户或其它指示(例如,用于识别先前存储的程序和/或判断是否所请求的程序实例是否得到授权的指示);付费源的指示,用于向程序执行服务付款,以执行程序实例;程序实例执行的预先付款或其它授权的指示(例如,事先购买的对一段时间、程序执行实例的数目、资源使用量等的有效预订);和/或需要立即执行和/或存储起来以备稍后执行的程序的可执行或其它副本。此外,在某些实施例中,程序执行请求还可以包括许多其它类型的、对执行一个以上的程序实例的偏好和/或要求。此类偏好和/或要求可以包括:对于需要在特定地理和/或逻辑位置中(如在容纳有多个可用计算机的多个数据中心中的一个数据中心的、在彼此邻近的多个计算系统中、和/或在与执行一个以上其它所指示的程序实例(例如,同一程序或其它程序的实例)的计算系统相邻近的一个以上的计算系统中)执行的某些或全部程序实例的指示。此类偏好和/或要求还可以包括:对在执行过程中各自被分配以所指示的资源的某些或全部程序实例的指示。

[0022] 在一所指示的时刻接收到请求执行程序的一个以上的实例之后,程序执行服务就确定用于执行程序实例的一个以上的计算系统。在某些实施例中,确定所用计算系统是在请求的时刻执行的,即使程序实例在将来执行也是如此。在其它实施例中,可以推迟到稍后的时间再确定将来执行程序的一个以上的程序实例所用的计算系统,例如可以基于那时的可用信息来确定将来的执行时刻。可以通过多种方式(包括(例如,在预先注册的时刻)基于在程序请求中指定或者为程序和/或相关用户指定的任何偏好和/或要求)来确定使用哪个计算系统来执行各个程序实例。举例而言,如果为执行程序实例的偏好和/或要求的资源确定了准则,那么可以至少部分基于计算系统是否具有足够的可用来满足这些资源准则的资源,来确定合适的用于执行程序实例的计算系统。

[0023] 在某些实施例中,程序执行服务可以根据所要执行的一个以上的先前存储的程序副本的位置,来确定使用哪个计算系统来执行要执行的程序。具体而言,如前所述,至少在某些实施例中,可用于执行程序的多个计算系统可以被组织成组(诸如各个计算系统可以属于多组之一)。因此,判断计算系统是否适合用于执行程序的实例可以部分基于是否那个计算系统的组中一个以上的计算系统存储了程序的本地副本。通过选择已在本地存储了程序副本的计算系统或属于具有一个以上的本地储存副本的组的计算系统来执行程序实例,可以获得各种好处,例如基于获取程序副本来减少程序执行启动时延。尽管组内的一个计算系统存储了要执行的程序的本地副本,然而由于多种原因(例如,如果具有本地存储副本的计算系统当前没有足够的资源来执行程序实例,如果具有本地存储副本的计算系统正在执行一个以上的程序实例,等等),程序执行服务可以选择组内的一个以上的其它计算系统在当前执行程序实例。

[0024] 在又一实施例中,程序执行服务可以根据各种其它因素,来选择一个以上的计算

系统来执行程序实例。举例而言,当用户请求同时执行所指示的程序的多个实例时,程序执行服务可能选择将程序实例的执行分布在属于不同组的计算系统间,以例如在特定组网络中断或发生其它问题的情况下提供增强的可靠性。类似地,在某些实施例中,(即使单个计算系统具有足够的资源来执行多个实例),也可以在多个计算系统上而不是单个计算系统上执行程序的多个实例。这种程序实例的分布可以,例如,在执行程序所有实例的单个计算系统失败或同该单个计算系统失去连接的情况下,提供增强的可靠性。此外,如果受程序执行服务管理的计算系统在物理(例如,地理上)是分开的,程序执行服务可由用户指示,或优选在位于单个数据中心内的计算系统上执行程序的多个实例,以例如为执行程序实例间的通信提供相对较高的网络带宽。可选地,(例如如果程序实例几乎不存在相互通信,和/或如果多个程序实例支持多个在地理上分散的独立的终端用户或应用程序),程序执行服务可以接受指示,或优选在多个独立的数据中心中执行多个程序实例。

[0025] 在程序执行服务确定了用于执行程序实例的一个以上计算系统之后,程序执行服务可以通过多种方式启动这些程序实例的执行。举例而言,系统管理器模块可以向选定的计算系统提供指令或各种其它执行信息。例如,此类其它信息可以包括:对存储或可能存储了程序的本地副本的一个以上其它计算机的指示。向选定的计算系统提供的其它类型的信息可以包括:关于程序实例执行时长的指示、关于分配给程序实例的资源的指示、向程序实例提供的访问权限的指示、对如何管理程序实例的执行的任何限制(例如,(如果存在的话)用于使程序实例能够进行发送或接收的通信的类型)的指示等。

[0026] 在选定的计算系统获知应执行所指示的程序的一个以上实例之后,选定的计算系统尝试根据任何接收到的指令或其它相关信息(例如,预先定义的偏好或要求)执行程序实例。至少在某些实施例中,程序执行通知可以由与选定的计算系统相关的机器管理器模块(例如,在选定计算系统上执行的机器管理器模块,或代表选定计算系统所属的组执行的机器管理器模块)接收。在此类实施例中,机器管理器模块可用于管理程序实例的执行。举例而言,在选定计算系统未存储所要执行的所指示的程序的本地副本的情况下,机器管理器模块可以获得或获取用于执行以及用于可选的本地存储的程序副本。举例而言,获取程序副本可以包括:同通知中指示的或已知至少有可能存储了程序的本地副本的一个以上的计算机或其它系统(例如,数据存储系统)相联系,以请求或获取程序副本。在不同实施例中,获取程序副本可以通过不同方式予以实现,如以下详细讨论的那样,获取程序副本的方式可以包括连同接收到的指示执行程序实例的通知一起接收程序副本。如以下详细讨论的那样,至少在某些实施例中,程序执行服务可以采用多种其它行为来管理程序的执行。

[0027] 在另一方案中,可以提供API,所述API允许其它程序可编程地发出执行程序实例的请求,还可以可编程地执行各种其它类型的管理、供应操作。此类操作包括但不限于:创建用户账户、保留执行资源、注册所要执行的新程序、管理组和访问策略、监控和管理程序实例的执行等。例如,由API提供的函数可以由代表用户的客户计算系统和设备(包括由在程序执行服务的计算系统上执行的程序实例)来调用。

[0028] 为了说明性的目的,以下对某些实施例进行了说明,其中,以特定方式管理特定类型的计算系统上的特定类型程序的执行。为了说明性的目的,提供这些示例,并且为简洁起见,对这些示例进行了简化。并且本发明的技术可用于各种其它情况,以下对其中一些情况进行了说明,并且本发明的技术不局限于同虚拟机、数据中心或其它特定类型的计算系统

或计算系统配置一起使用。

[0029] 图1是一幅网络图,示出了多个计算系统在其中(在诸如程序执行服务的控制下)交换和执行程序的示例实施例。具体而言,在本示例中,程序执行服务管理在位于数据中心100内的多个计算系统上的程序执行。在该示例实施例中,数据中心100包括多个机架105,每个机架包括多个计算系统110a-c以及机架支撑计算系统122。在该示例中,计算系统110a-c分别寄居一个以上的虚拟机节点120,以及用于管理虚拟机的独立节点管理器115。在该示例中,各虚拟机120可用于提供独立的计算机环境以执行程序的实例。在该示例中,机架支撑计算系统122向机架的其他本地计算系统,还可能向位于数据中心的其它计算系统提供各种实用工具服务。举例而言,实用工具服务可能包括:为其它计算系统进行数据和/或程序存储、执行一个以上的机器管理模块以支持其它计算系统,等等。各个计算系统110可以可选地具有一个独立的机器管理模块(例如,被配备为计算系统的节点管理器的一部分)和/或具有用于存储程序的本地副本的本地存储器(未示出)。在该示例中,计算系统110a-c以及机架支撑计算系统122的共享公共的数据交换媒介,并且可能均属于一组的一部分。该公共数据交换媒介可同一个以上由例如数据中心100内的其它机架或计算系统所共享的外部数据交换媒介相连接。

[0030] 此外,示例数据中心100还包括:与节点管理器125共享公共数据交换媒介的附加计算系统130a-b以及135,并且节点管理器125管理计算系统130a-b以及135。在所述示例中,计算系统135还寄居了多个虚拟计算机,作为用于为一个以上的用户执行程序实例的执行环境,然而计算系统130a-b没有寄居独立的虚拟机。在该示例中,可选计算系统145位于数据中心100和外部网络170的连接处。可选计算系统145可提供多种服务,以例如充当网络代理、管理传入和/或传出数据的传输等等。此外,可选系统管理器计算系统140还被示为协助管理程序在位于数据中心内的其它计算系统上(或可选地在位于一个以上的其它数据中心160中的计算系统上)的执行。可选系统管理器计算系统140可执行系统管理器模块。如上所述,除了管理程序的执行以外,系统管理器模块还可以提供多种服务,包括:用户账户的管理(例如,创建、删除、计费);对所要执行的程序进行注册、存储,以及分配;收集和处理与程序执行相关的性能以及审核数据;从客户或其它用户获取执行程序所需的费用等。

[0031] 在该示例中,数据中心100通过网络170(如,互联网)与多个其它系统相连接,所述多个其它系统包括:可由数据中心100或第三方的操作员操作的附加计算系统180、同样可由数据中心100或第三方的操作员操作的附加数据中心160、以及可选的系统管理器150。除了提供多种其它服务,系统管理器150可以以类似于系统管理器140的方式,在位于一个以上的数据中心100和/或160中的计算系统上管理程序的执行。虽然示例系统管理器150在该示例中被描述为位于任意特定数据中心的外部,但是在其它实施例中它可以位于数据中心(如数据中心160之一)的内部。

[0032] 图2示出了两组(例如代表程序执行服务)存储和交换程序副本的计算系统的示例。应当理解的是,在实际的实施例中,组、计算系统、以及程序的数量可能比图2所示的组要大的多。举例而言,作为一说明性实施例,每组可以有40个计算系统,每个数据中心可以有100个组,因此每个数据中心可以有4000个计算系统,并且每个计算系统可寄居15个虚拟机以执行客户的程序实例。此外,如果每组包括具有2兆兆字节(terabytes)存储容量的专用计算系统,则每组可存储2000个千兆字节(gigabytes)的虚拟机镜像程序副本,每个数据

中心总共200,000个副本。可选地,如果每组40个计算系统分别具有10万兆的本地存储容量,则每组可存储4000个千兆虚拟机镜像程序副本,每个数据中心总共400,000个副本。如果每个寄存虚拟机执行一个程序,那么这样的数据中心可一次执行多达60,000个程序实例。将意识到的是,在其它实施例中可以使用其它数量的组、计算系统、以及程序,并且可以存储和执行尺寸小得多的和/或具有可变尺寸的程序。

[0033] 在该示例中,图2示出了2个组,组A 200和组B 250。组A包括三个分别名为MA1、MA2、以及MA3的计算机210a至c。类似地,组B250包括三个名为MB1、MB2、以及MB3的计算机260a至c。每组可具有不同数量不同类型的计算机,并且在某些实施例中,计算机可以是属于多组或不属于任何组的成员。正如在别处详细说明的那样,各组计算机共享该组的公共数据交换媒介(未示出)。

[0034] 在一说明性实施例中,图2的各计算机可执行一个以上的程序实例,并且可在本地程序储存库(例如,作为由诸如硬盘或其它存储设备等提供的永久性存储器的一部分)中存储一个以上的本地程序副本。举例而言,计算机MA1在其程序储存库220a中存储程序P1、P2、P3、P5以及P9的本地副本,并且如方框230a所示,当前正在执行程序P1的实例。在该示例中,各计算机上的程序储存库的存储容量被限制为最多存储五个程序副本,并且各计算系统的执行资源被限制为最多同时执行两个程序实例。在该示例中采用的对程序储存库空间以及执行程序数量的限制仅仅是为了用于说明,在其它实施例中各计算系统还可具有独立的资源。此外,虽然在许多实施例中,程序储存库的空间比执行程序实例时的可用内存空间大一个以上的数量级,但这种情况并不一定是必须的。在其它实施例中,同时执行的程序的最大数量可以大于、小于或等于本地存储于程序储存库中的程序副本的数量。因此,至少某些计算机或其它系统可以仅提供本地程序储存库和可用资源之一以执行程序实例。最后,如在别处将予以详细说明的那样,在某些实施例中,可以在某些情况下将至少某些程序的本地存储副本从存储器中移除或清除,以例如在程序储存库达到其容量时,为其它程序副本腾出空间。在某些实施例中,可以在某些情况下终止或停止至少某些程序的正在执行的实例,以例如在程序执行资源达到其容量时,为其它正在执行的程序实例腾出空间。

[0035] 为了说明的目的,此处给出了程序执行服务的一实施例的若干种操作的示例。程序执行服务可采用一个或多个指定的、预定义的和/或习得的(learned)策略,来影响执行程序实例在计算机上的布置,在该示例中采用了如下所示的一组简单的策略。首先,如果可能,在一组以上的计算机上执行多个程序实例。第二,如果可能,在一台以上的计算机上执行多个程序实例。第三,如果可能,在已在其程序储存库中存储了该程序副本的计算机上执行程序实例。第四,如果可能,在具有至少一台已在其程序储存库中存储了该程序的本地副本的计算机的组中的成员计算机上执行程序实例。最后,如果可能,在具有最多可用执行资源的计算机上执行程序实例。

[0036] 下面开始说明为这六个计算系统管理程序执行的说明性示例,假定程序执行服务的客户已请求执行程序P7的两个实例。在这种情况下,鉴于上述策略,程序执行服务的示例实施例可能选择在组A中执行P7的一个实例并在组B中执行P7的一个实例,因为该布置倾向于将副本分布在一个以上的组中。在组A的计算机之间,该组没有任何一台计算机存储程序的本地副本,由于计算机MA3正在执行两个程序(P8和P9),因此程序执行服务可能选择不在于MA3上执行P7的副本。在计算机MA1和MA2之间,由于MA2当前未执行任何程序,因此将选择

MA2来执行P7的副本。在所述实施例中,机器MA2将从位于组A外部的一个以上的计算系统获取程序P7的副本,以用于执行并可选地在储存库220b中进行本地存储。举例而言,机器MA2可从程序执行服务的所有计算机所用的远程程序储存库和/或从程序执行服务的外部位置获取程序P7的副本。对于组B的计算机,由于没有一个计算系统存储程序的本地副本,并且每台计算机正在执行一个程序,因此程序执行服务可选择三台计算机中的任意一台来执行P7程序实例。然而,由于MB3当前只在其程序储存库中存储了一个程序副本,因此程序执行服务可能选择机器MB3。因此,如果需要的话无需从其程序储存库中清除已存储的程序副本,机器MB3就可以存储程序P7的本地副本。

[0037] 接下来,再次从图2所示的初始条件出发,假定程序执行服务的客户已请求执行程序P6的两个实例。在这种情况下,鉴于上述策略,由于该布置将把实例分配在一个以上的组中,因此程序执行服务的示例实施例可能再次选择在组A中执行P6的一个实例,并在组B中执行P6的一个实例。在组A的计算机之间,由于没有一个计算系统存储了程序P6的本地副本,并且计算机MA2是最不忙碌的,因此将可能再次选择计算机MA2。在组B的同等繁忙的计算机之中,尽管只有MB2存储了程序的本地副本,但是由于策略优选在同一组内的多台计算机上分配单个程序的副本,因此可能不会选择计算机MB2。然而,值得注意的是,由于MB2的程序储存库中已存储P6的副本,因此采用表现为比起可靠性而言更重视效率的不同策略的其它实施例,事实上可能刚好选择在计算机MB2上执行P6。由于无需从MB3程序储存库中清除任何程序副本,因此在剩余的候选计算机MB3和MB1之间,程序执行服务可能再次优选MB3。因此,在该实施例中,机器MB3将从MB2获取程序P6的副本,以用于执行,并可能将其存储在本地储存库270c中。

[0038] 接下来,再次从图2中所示的初始条件出发,假定程序执行服务的客户已请求执行程序P4的一个实例。在这种情况下,鉴于上述策略,程序执行服务的示例实施例将可能选择在计算机MB1上执行P4。具体而言,由于不存在正在执行的P4的实例,并且只请求执行一个实例,因此不适用优选在多组间分配程序实例、以及优选避免在单独的计算机上放置程度的多个执行实例的策略。因此,由于MB1已在其程序储存库中存储了程序P4的本地副本,因此将可能选择MB1来执行P4。

[0039] 接下来,再次从图2中所示的初始条件出发,假定程序执行服务的客户已请求执行程序P10的一个实例。在这种情况下,鉴于上述策略,程序执行服务的示例实施例将可能选择在MA2上执行P10。与前一示例相同,不适用优选在多组间分配用以执行的程序实例以及避免在单台计算机上放置程序的多个实例策略。并且,虽然因为计算机MA3已在储存库中存储了P10的副本,故计算机MA3是一个很吸引人的候选计算机,但是由于MA3已达到了两个执行程序(P8和P9)的上限,因而不具备在当前执行P10的能力。这使得由于MA1和MA2与在其储存库中存储了程序P10的本地副本的计算机(MA3)位于同一组中,计算机MA1和MA2优先于组B中的任何一台计算机。在MA1和MA2之间,由于MA2最不繁忙,因此可能选择MA2,并且MA2将从MA3中获取程序P10的副本。

[0040] 接下来,再次从图2中所示的初始条件出发,假定程序执行服务的示例实施例的客户已请求执行程序P3的6个附加实例。在这种情况下,鉴于上述策略,程序执行服务将可能在计算机MA2上执行两个实例并在计算机MA1、MB1、MB2和MB3上各执行一个实例。由于计算机MA3已处于两个执行程序(P8和P9)的上限,因此可能不会在计算机MA3上执行任何实例。

注意在这种情况下,某些实施例可能从那些程序储存库不具有多余容量以存储程序P3的本地副本的计算机中清除所存储的程序的本地副本。举例而言,在选择总是在程序执行前在本地程序储存库中存储所要执行的程序的副本的实施例中,计算机MA1和MB1可从其各自的程序储存库中清除一个本地程序副本。还应注意,在这种情况下,与优选在多台计算机之间分配执行程序多个实例的策略相反,计算机MA2和MB3将可能最终分别执行P3的两个实例。然而,由于在给定示例中不存在额外的、用以执行P3程序实例的计算机,因此如果想要满足请求,程序执行服务将选择在单台计算机上执行P3的多个实例。可选地,在某些实施例中,程序执行服务可为策略分配不同的权重,使程序执行服务可选择执行少于所请求的数目的实例,以例如在计算机MA1、MA2、MA3以及MB3上各执行一个单独的实例。类似地,在某些实施例中,如果请求六个以上的程序P3的附加实例,并且程序和/或请求者的优先级足够高,那么程序执行服务可以(例如通过终止另一个程序实例(例如MA3上的程序P8和/或P9的实例)的执行和/或通过在一个当前执行的程序实例正常结束之后为P3保留下一可用程序实例执行的方式)选择执行P3的附加实例。

[0041] 继续参考当前示例,由于同组A中的计算机MA1和MA2一样,组B中的MB2和MB3均存储了该程序的本地副本,因此计算机MB1具有多个可用于获取执行所用的程序P3的副本的来源。在该实施例中,MB1将请求本组的MB2和MB3提供程序P3的一部分(例如,第一组X比特以及第二组X比特,其中,X是由程序执行服务选定的数目)。机器MB1接着监控从计算机接收到响应的速度,并请求响应速度更快的计算机提供该程序的至少大部分(有可能是全部)的剩余部分。在其它实施例中,可以其它方式为计算机MB1获取程序P3的副本,所述方式可以是诸如:只从计算机MB2和MB3之一请求程序副本、(除了或并不从组B中的MB2和MB3还或而)从组A中的计算机MA1和/或MA2请求至少部分程序副本等。

[0042] 图3是一幅方框图,示出了适于诸如通过执行程序执行服务系统的实施例等方式来管理在被管理的多个计算系统上的程序执行的示例计算系统。在该示例中,计算系统300执行系统管理器模块的实施例,以协调在被管理的多个计算系统上的程序执行。在某些实施例中,计算系统300可对应于图1中的系统管理器140或150。此外,一个以上的机器管理器计算系统370分别执行机器管理器模块382,以便于由一个以上的相关计算系统获取和执行程序。在某些实施例中,一个以上的机器管理器模块可以分别对应于图1的节点管理器115或125中的一个。在该示例中,提供多个机器管理器计算系统,并且其中每一个机器管理器计算系统充当受系统管理器模块管理的程序执行服务的多个计算系统中的一个。在所述示例中,独立的机器管理器模块执行于各个计算系统370上。在其它实施例中,各机器管理器计算系统上的机器管理器模块可管理一个以上的其它计算系统(例如,其它计算系统388)。

[0043] 在该示例实施例中,计算系统300包括:中央处理单元(“CPU”)335、存储器340、内存345、以及各种输入/输出(“I/O”)设备305,并且所述I/O设备包括:显示器310、网络连接315、计算机可读介质驱动器320、以及其它I/O设备330。未示出的其它I/O设备可包括:键盘、鼠标或其它定位设备、麦克风、扬声器等。在所述实施例中,在内存345中执行系统管理器模块350,以管理其它计算系统上的程序执行,并且还可以可选地在内存345中执行一个以上的其它程序355。计算系统300以及计算系统370通过网络386相互连接并同其它计算系统388相连接。

[0044] 类似地,各计算系统370包括:CPU374、各种I/O设备372、存储器376、以及内存380。

在所述实施例中,在内存380中执行机器管理器模块382,以为程序执行服务(如代表程序执行服务的客户)管理一个以上的其他程序384在计算系统上的执行。在某些实施例中,某些或全部计算系统370可寄居多个虚拟机。倘若如此,各个执行程序384可以是执行于单个寄居虚拟机上的完整的虚拟机镜像(例如,具有操作系统和一个以上应用程序)。机器管理器模块可类似地执行于另一寄居虚拟机,如具有特权的、能够监测其它寄居虚拟机的虚拟机。在其它实施例中,执行程序实例384和机器管理器模块382可作为执行于计算系统370上的单个操作系统(未示出)上的独立进程来执行。因此,在该示例实施例中,程序执行服务的能力是由系统管理器350和机器管理器模块382之间的交互予以提供的,所述系统管理器350和机器管理器模块382利用网络386进行通信,以共同管理程序在被管理的计算系统上的分配、获取以及执行。

[0045] 将意识到的是,计算系统(如计算系统300和370)仅仅用于说明,而不是用来限制本发明的范围的。计算系统300和370可同其它未示出的设备相连接,所述其它设备包括网络可访问的数据系统或其它数据存储设备。更一般地,计算机或计算系统或数据存储系统可以包括可相互作用并执行所述各类的功能的硬件或软件的任意组合,包括但是不局限于:台式计算机或其它计算机、数据库服务器、网络存储设备以及其它网络设备、PDA、蜂窝电话、无线电话、寻呼机、电子记事簿、互联网设备、基于电视的(如,采用机顶盒和/或个人/数字录像机的)系统、以及各种其它包含适当的相互通信功能的消费产品。此外,在其它实施例中,可以将由所述系统模块提供的功能合并于更少的模块中或分布在附加模块中。类似地,在某些实施例中,可能不提供某些所述模块的功能和/或可以使用其它附加功能。

[0046] 还应理解的是,虽然将各个项目说明为在使用时存储在内存中或存储器上,但是为了进行内存管理和保持数据完整性,这些项目或者其部分可在内存和其它存储设备之间发生转移。可选地,在其它实施例中,某些或全部软件组件和/或模块可执行于另一设备的内存中,并利用计算机间的通信同所述计算系统进行通信。某些或全部系统模块或数据结构也可以(例如,作为软件指令或结构体数据)被存储在计算机可读介质(如可由适当的驱动器或通过合适的连接读取的硬盘、内存、网络、或便携介质)上。系统模块和数据结构也可作为所产生的数据信号(例如,作为载波或其它模拟或数字传播信号的一部分)在包括基于无线和基于有线/电缆的媒介在内的各种计算机可读传输媒介上传输,并且可以采取各种形式(例如,作为单独或复用的模拟信号,或作为多个离散的数字分组或帧)。在其它实施例中,此类计算机程序产品也可采用其它形式。因此,本发明可以用其它计算机系统配置予以实施。

[0047] 图4A-4B示出了系统管理器模块工作过程400的实施例的流程图。该过程可通过例如执行图1中的系统管理器模块140和/或图3中的系统管理器模块350的方式予以实现,以例如代表程序执行服务管理多个程序在多个计算系统上的执行。

[0048] 该过程起始于步骤405,并接收与一个以上的程序的执行相关的状态消息或请求。接着过程进入步骤410,判断接收到的消息或请求的类型。如果判定已接收到执行一个以上的所指示的程序的一个以上的实例的请求,过程就进入步骤415。在步骤415中,该过程确定一组以上的计算系统,以执行所指示的程序。在步骤420中,该过程在一个以上的所指示的组的每组中选择一个以上的计算系统来执行所指示的程序的实例。可基于多种因素(如一个组是否具有一个以上的、存储了程序的一个以上的本地副本的计算系统、适合的计算资

源的可用性、以及组中计算系统的位置) 选择一个以上的组。在所指定的组中选择一个计算系统可类似地基于各种因素(如所存储的程序的本地副本在组中计算系统间的位置、以及计算资源的可用性)。如前所述,在不同实施例中,可用不同的指定策略以及其它准则(包括由用户或其他请求者指定的准则)作为组和计算系统选择的一部分。此外,在其它实施例中,可能不会独立地选择组以及特定的计算系统,以例如仅仅挑选最适合的一个以上的计算系统而不对组进行考虑(例如,在不使用组的情况下)。

[0049] 接下来,在步骤425,过程通过诸如发送包括执行程序实例的指令在内的消息,向选定的计算系统和/或与这些计算系统相关的一个以上的机器管理器模块提供所要执行的程序的指示。在所述实施例中,独立的机器管理器模块执行于各个计算系统上,并且接收该消息。如前所述,可向机器管理器模块提供多种类型的信息,包括:若干指示,用于说明如何确定需要从中获取所要执行的程序的副本的一个以上的计算系统。可选地,在某些实施例中,系统管理器可直接向计算系统提供所指示的程序的副本和/或在无需机器管理器模块或其它附加模块介入的情况下,启动程序在计算系统上的执行。

[0050] 否则,如果例如在步骤410中判定从用户接收到注册新程序的请求,该过程就进入步骤440,并存储程序的指示以及任何相关的管理信息(如注册程序的用户的标识)。接下来,在步骤445中,过程可选地开始向一个以上的计算系统分配所指示的程序的副本。举例而言,在某些实施例中,系统管理器可选择将所存储的所指示的程序的本地副本分配(seed)到一个以上的数据中心内的一个以上的计算系统和/或程序储存库中,以提高稍后启动程序执行的效率。

[0051] 否则,如果在步骤410中判定接收到反映一个以上的被管理计算系统的操作的状态消息,过程就进入步骤450,并更新一个以上计算系统的状态信息。举例而言,机器管理器模块可能判定相关计算系统修改了正在执行的程序实例和/或所存储的本地程序副本,并且可能相应地向系统管理提供状态消息。在某些实施例中,将由机器管理器模块周期性地发送状态信息,以使系统管理器能够获知在选择合适的用于执行程序的计算系统时所用的被管理的计算系统的运行状态。在其它实施例中,可在其它时刻(如相关改变发生的任何时刻)发送状态信息。在其它实施例中,系统管理器模块可以按照需要从机器管理器模块请求信息。状态消息可包括多种类型的信息,如当前执行于特定计算系统上的程序的数目和标识、当前存储在特定计算机上的本地程序储存库中的程序副本的数目和标识、计算系统的性能相关和资源相关的信息(例如,CPU利用率、网络、硬盘、内存等)、计算系统的配置信息、以及与特定计算系统上的硬件或软件相关的错误或失败条件的报告。

[0052] 如果在步骤410中判定收到任何其它类型的请求,过程就进入步骤455,并视情况执行其它所指示的操作。举例而言,此类操作可包括:响应来自系统中其它组件的状态查询、暂停或终止一个以上的正在执行的程序的执行、将当前执行的程序从一个计算系统迁移至另一个计算系统、关闭或重启系统管理等。

[0053] 在步骤425、445、450以及455之后,过程进入步骤430并可选地执行任何常规(housekeeping)任务,如:计算用户付费信息、更新显示信息、向节点管理器或其它组件发送周期性查询、轮替日志或其它信息等。接着,过程进入步骤495并判断是否继续。倘若继续,过程就返回步骤405,否则就进入步骤499并返回。

[0054] 图5示出了机器管理器模块的工作过程500的流程图。该过程可通过例如执行图3

中的机器管理器模块382和/或图1中的节点管理器115或125的方式予以实现,以例如便于为一个以上的被管理的相关计算系统获取程序副本和执行程序实例。在所述实施例中,各机器管理器模块的工作过程是代表被配置为即执行一个以上的程序实例又存储一个以上的本地程序副本的单个计算系统执行的,其中,机器管理器模块与参考图4A-B所描述的系统管理器模块的工作过程协同工作,为程序执行服务管理被管理的计算系统的程序执行。

[0055] 该过程起始于步骤505,并从诸如系统管理器模块接收与执行一个以上的程序有关的请求。过程进入步骤510以判断是否接收到执行或存储所指示的程序的请求。如果是,过程就进入步骤515,以判断所指示的程序当前是否存储在被管理的计算系统的本地程序储存库中。否则,过程就进入步骤540,以判断本地程序储存库是否具有足够容量来存储所指示的程序。如果没有,过程就进入步骤545,并例如根据在步骤505中接收到的请求中所指示的那样或基于机器管理器模块所采用的清除策略,从本地程序储存库中清除一个以上的程序。在步骤545之后,或如果在步骤540中确定本地程序储存库的确具有足够容量用来存储所指示的程序的本地副本,过程就进入步骤550,并从一个以上的确定的其它计算系统获取所指示的程序的本地副本。过程可以通过多种方式(包括基于作为在步骤505中接收到的请求的一部分的信息)确定存储着程序的本地副本的其它计算系统。此外,可以采用一种以上的其它技术,例如还可以使用向邻近计算系统广播、向中心目录请求、和/或对等数据交换等技术。在其它实施例中,可以在步骤505中将程序副本同请求一起提供。接着,过程进入步骤555,并在本地程序储存库中存储所获得的所指示的程序的副本。在步骤555之后,或者如果在步骤515中确定所指示的程序已存储在储存库中,过程就进入步骤520,以判断是否接收到需要执行程序的指示。如果收到,过程就进入步骤525,并且启动执行所指示的程序。

[0056] 如果在步骤510中判定未接收到存储或执行程序的请求,过程就进入步骤535,并视情况执行其它所指示的操作。举例而言,其它操作可以包括:比如对接收到的请求予以响应和/或基于收集到的关于程序性能的信息(如程序行为混乱或过度使用资源),暂停或终止一个以上的程序的执行。此外,其它操作可以包括:对于与当前正在执行的程序或本地程序储存库的内容有关的状态信息的请求予以响应等。

[0057] 在步骤535、525之后,或者如果在步骤520中判定未接收到执行程序的指示,过程就进入步骤530,并向一个以上的系统管理器模块发送状态信息消息。在所述实施例中,过程在每次操作后向系统管理器模块发送状态信息消息,从而使系统管理器能够得知受节点管理器管理的计算系统的状态。在其它实施例中,状态信息可以在其它时刻并以其它方式发送。在步骤530之后,过程进入步骤595并判断是否继续。倘若继续,过程就返回步骤505,否则,就进入步骤599并返回。虽然此处并未说明,但是过程还可以根据需要在各时刻执行各种内务处理操作。

[0058] 图6示出了程序执行服务客户的工作过程的实施例的流程图。该过程可通过例如驻留在图1所示的计算系统180之一上的应用程序予以实现,以例如提供交互式控制台,令使用者能够与程序执行服务进行交互。该过程可以可选地反映出由程序执行服务以交互方式向用户提供的和/或以编程方式向用户程序提供的的能力。可选地,该过程可以是由程序执行服务在被管理的计算系统之一上执行的程序之一的一部分,以例如使此类程序能够动态地执行附加的程序实例,以实现负载均衡、满足增长或减少的需求等目的。

[0059] 该过程起始于步骤605,并接收与一个以上的程序的执行相关的请求。在步骤610

中,过程确定接收消息的类型。如果请求有关于新程序注册(或先前注册的程序的新版本),过程就进入步骤615,并向程序执行服务(例如,向系统管理器模块)发送新程序需要注册的指示。所述指示可以包括:程序的副本或如何获取程序的指令。如果在步骤610中确定请求与执行程序相关,过程就进入步骤615,以向程序执行服务(例如,向系统管理器模块)发送请求,以执行所要执行的程序的一个以上的实例。举例而言,过程可利用先前从程序执行服务接收到的指示来确定程序和/或将代表哪个用户执行程序实例。如果在步骤610中判定接收到某些其它类型的请求,过程就进入步骤625,并视情况执行其它所指示的操作。举例而言,过程可以向程序执行服务发送请求从而保留计算资源以在将来执行一个以上的所指示的程序实例,向程序执行服务发送与一个以上的程序的当前或先前执行有关的状态查询,(例如,作为向程序执行服务注册用户的一部分)提供或更新与用户相关的信息、撤销或删除之前注册的程序、暂停或终止一个以上的程序实例的执行等。

[0060] 在步骤615、625、或630之后,过程进入步骤620,并可选地执行额外的常规任务,以例如更新显示信息、存储接收到的响应于步骤615、625或630的从程序执行服务(未示出)返回的信息、对程序执行服务进行周期性的状态查询等。在步骤620之后,过程进入步骤695以判断是否继续处理。倘若继续,过程就返回步骤606,倘若不继续,就进入步骤699并返回。

[0061] 本领域的技术人员还将意识到,在某些实施例中,由上述过程提供的功能可通过其它方式(如将所述功能分入更多过程,或合并为更少的过程)予以提供。类似地,在某些实施例中,在诸如其它所述过程分别缺少或包括这样的功能,或当所提供的功能的数量发生改变的情况下,所述过程可以提供比所描述的功能数量更多或更少的功能。此外,虽然可能将各个操作说明为是以特定方式(例如,串行或并行)和/或特定顺序予以执行的,但是本领域技术人员将意识到,在其它实施例中,可按照其它顺序或其它方式来执行操作。本领域技术人员还将意识到,以上讨论的数据结构可采用不同方式(如通过将单个数据结构分解为多个数据结构,或将多个数据结构合并为单个数据结构)加以组织。类似地,在某些实施例中,在诸如其它所述数据结构分别缺少或包括这样的信息,或当所存储的信息的数量或类型发生改变时,所述数据结构可以存储比所描述的数量更多或更少的信息。

[0062] 如上所述,各实施例将程序执行服务的计算系统组织成一个以上的组,以便于实现与程序执行相关的策略。此外,计算系统可以通过其它方式(如采用组层次结构)加以组织。举例而言,最小的组可以各包含单个计算系统,并且各计算系统将被分配到各自的组中。通过单个网络交换机连接的单个机器组还可以包含在交换机级组(switch-level group)中,交换机级组包含所有通过单个网络交换机进行物理连接的计算系统。交换机级组还可以包含在数据中心级组中,该数据中心级组包含位于给定数据中心内的所有计算系统。数据中心级组还可以包含在全体组(universal group)中,全体组包含多个数据中心中的所有计算系统。按照这样的组织结构,位于各级的组对位于组中其它计算系统上的程序副本的访问速度通常逐级变慢,单个机器组提供最快的访问速度,全体组提供最慢的访问速度。由于程序执行服务可以搜索既存储了所要执行的特定程序的副本又具有用于执行程序的必不可少的资源的最小的组,因此这样的组织使得可以高效地实施用于指导对执行程序进行最优布置的各种策略的应用。可选地,其它实施例可能完全不会以组的方式对程序执行服务中的计算系统进行建模。例如,此类实施例可以向与某些或全部网络交换机相连接的或位于某些或全部硬件机架上的专用数据存储计算或其它系统分配某些或全部程序

的副本,然后只随机地向计算系统分配所要执行的程序。

[0063] 如前所述,对于将计算系统和/或组选作执行程序 and/或接收程序副本分配的候选系统和/或组而言,不同的实施例可能实施不同策略。在很多情况下,不同的程序放置策略可能需要在诸如可靠性和效率(例如,启动延迟、网络延迟、或吞吐量等)等因素间进行折中。布置策略可以考虑以下因素,如:请求执行一个以上程序的用户的偏好;当前执行的程序的数目、标识、以及位置;当前请求执行的程序的数目以及标识;安排用于在将来执行的程序的数目以及标识;之前存储的程序副本的位置;网络架构;地理位置等。此外,在某些实施例中,在某些情况下可以基于用户请求或其它因素对默认的应用策略进行覆盖或修改。举例而言,特定实施例提供一组默认策略,这组默认策略可以被用户在其执行一个以上程序的请求中所表达的偏好所覆盖。

[0064] 在计算系统受跨多个数据中心的程序执行服务的管理的实施例中,程序执行服务可以优先选择在同一数据中心内执行单个程序的多个实例,和/或同一数据中心内执行同一用户的多个单独实例。这样的策略往往使这些程序能够利用带宽相对较高的数据中心内数据交互,与实现程序实例之间的通信交换。另一方面,某些实施例可能优先选择将这些程序实例分布在多个数据中心内,以在可能导致整个数据中心不能正常工作的功率、网络、或其它大规模中断(如程序实例几乎不与其它此类程序实例通信)的情况下确保可靠性。此类分布或合并这种程序实例的偏好可以类似地应用于各其它级别的计算系统组织(诸如物理子网、组、以及个人计算系统)中。此外,某些实施例可以采用可用于在多个候选计算系统之间进行选择的策略,如果不进行选择,将无法在程序执行服务的布置策略之下区分这些候选计算系统。举例而言,一实施例可以从一组同等条件的候选计算系统中随机选择计算系统,而另一实施例可以选择具有最低资源利用率的计算系统,而另一实施例可以循环(round-robin)选择此计算系统。

[0065] 此外,对于程序执行,不同实施例可以实施不同的用于在本地程序存储库中存储程序副本的策略。举例而言,某些实施例可以始终在在容纳本地程序存储库的计算系统上执行之前(或期间、或之后),将程序的本地副本存储在本地程序存储库中。可选地,在其它实施例中,可以仅仅将某些程序存储在这样的本地程序存储库中。此外,在程序存储库容量不足无法存储给定程序的本地副本时,不同实施例可以采用不同方法。举例而言,为了腾出空间以存储新程序,某些实施例将选择清除或移除存储在程序存储库中的一个以上的程序副本,以例如清除最近最少使用的副本、最旧的副本、随机副本、以不同方式选定的副本、仍存储在某些其它相关程序存储库(如同一组中的一个以上的其它计算系统的程序存储库)中的程序的副本等。在其它实施例中,在给定的程序存储库空间已满时不执行清除操作(而是例如周期性地(如每天、重启时等)从程序存储库中删除所有程序,或仅当将程序从程序执行服务中撤销时删除程序)。

[0066] 在某些实施例中,程序可以被分解成多个、可能是固定大小的数据块。通过以该方式分解程序,获取程序副本的计算系统可以向多个其它已经在其程序存储库中存储了所需程序的计算系统分配请求。当某些其它的多个计算系统对程序块的请求予以响应时,获取计算系统可以向这些响应计算系统请求更多程序块。因此,同较少地响应或不响应的计算系统相比,更偏重由具有足够的可用资源的计算系统提供程序块。

[0067] 某些实施例可以通过诸如仅仅传递与可能已经存储在本地程序存储库中的其它

程序不同的程序部分的方式进行优化,以提高程序传递的效率。对于同一程序的多个、递增(incremental)版本,或共享大部分代码或数据的不同程序,此方法是十分有利的。举例而言,如果程序被分解为多个、可能为固定大小的块,那么在起初向程序执行服务注册程序时,可以计算每个块的校验和并将其保存。之后,当需要获取程序以用于执行时,计算系统可以将程序块校验和同与存在于一个以上的程序储存库中的程序块相关的校验和进行比较,然后仅仅获取尚未存储的程序块。可选地,某些实施例可以将程序表示为一组一个以上文件,如可执行文件、数据文件、以及库文件。在这种情况下,两个程序可能共同拥有一个以上的文件(如,库文件),并且给定计算系统可以选择仅仅获取与已储存在计算系统程序储存库中的文件不同的、需要获取以用于执行的程序的文件。

[0068] 某些实施例将提供具有全都是固定大小的程序,然而其它实施例可以支持不同大小的程序。在计算诸如内存或程序储存库的系统资源的程序利用率时,固定大小的程序可以简化程序的处理。在提供具有不同大小的程序的实施例中,在存储程序的本地副本和/或执行程序实例时,可应用不同算法(包括诸如最优匹配、首次匹配等各种装箱算法)优化固定大小资源(诸如内存或硬盘空间)的利用,以限制碎片的数量。

[0069] 此外,某些实施例可以提供,在请求执行程序之前向多个被管理的计算系统分配或分发程序副本的功能。虽然某些实施例将提供至少一个通用程序储存库,用以在首次注册程序时对程序进行存储,但是由于无法在任何相对于执行该程序的计算系统而言的本地程序储存库中找到所述程序,因此在程序首次执行时,这些实施例将遭受很大的时延。如果这样的实施例被配置为在本地程序储存库中存储所要执行的程序的副本,那么与初始执行相比,后续执行将招致相对较小的启动时延。初次执行程序启动时延相对较长的问题可以通过在请求执行程序之前分配或分布程序副本的方式得到解决。此类实施例可以向对于提供程序执行服务的一个以上的数据中心而言的本地程序储存库分配程序的一个以上的副本。采用这种方式,当第一次请求执行程序时,通常可以在相对于计算系统或被选中用于执行程序的计算系统而言的本地(例如,至少在同一数据中心中的)程序储存库中找到程序。

[0070] 此外,某些实施例可以在同时或重叠启动执行单个程序的多个实例的情况下进行优化。在这种环境下,可能的情况是,需要由多个不同的计算系统几乎在同时获取所要执行的程序的副本。如果各个计算系统独立地从远程程序储存库获取程序的副本,那么由于各个计算系统同时开始在网络上传输相同数据,因此可能会导致过度利用网络或其它资源。在某些环境下,同步或安排多个计算系统获取程序的一个以上的副本的顺序,以更好地利用系统资源(例如,通过最小化不必要的网络利用),对于多个计算系统而言可能是十分有利的。举例而言,当选定的用于执行程序的多个计算系统是同组的一部分,且要从该组外部的一个以上的计算系统获取程序副本时,对于多个计算系统中的第一计算系统,首先从该组外部的计算系统获取程序副本(并将其存储在本地程序储存库中)可能是十分有利的。在第一计算系统已获取程序副本之后,其余的计算系统可以通过用于该组的公共数据交换媒介从第一计算系统获取副本。

[0071] 此外,当多个计算系统中的每个计算系统要获取程序副本时,可以利用多种其它技术来有效地利用网络和/或其它计算资源。举例而言,可以选择多个计算系统中的第一计算系统来管理向多个计算系统中的其它计算系统分发程序副本。如果多个计算系统中没有

任何一个计算系统具有存储在本地程序储存库中的程序的存储副本,选定的计算系统就可以发起从远程位置传输程序的至少某些部分(例如,块)。当选定的计算系统接收到程序的若干部分时,选定的计算系统可以将接收到的部分向多个计算系统中的其它计算系统进行多播。由于较少的冗余数据分组将被发送到连接多个计算系统的网络中,因此与其它网络通信机制(例如,由多个计算系统中的每个计算系统基于TCP进行传输)相比,多播机制可以产生更好地利用网络。可选地,如果多个计算系统中的一个以上的计算系统已在本地程序储存库中存储了程序副本,那么选定的计算系统可以指示具有程序的存储副本的一个以上的计算系统中的至少某些计算系统向多个计算系统中的其它计算系统多播程序的至少某些部分(例如,块),从而分散块的传输负载,并最小化对于网络的其它计算系统和/或部分的影响。在基于多播机制分发程序之后,多个计算系统中的一个以上的计算系统可以利用其它通信机制(例如,TCP),以获取还未接收到(例如,由于丢弃网络包的缘故)的程序的任意部分。其它分发机制可以包括:以将负载分布在网络中的多个计算系统的其它计算系统和/或部分上的循环或其它方式,分发对部分程序的请求。

[0072] 在某些实施例中,还可以使用更多的技术。举例而言,如果使用基于多播的分发机制从组中的一计算系统向本组中的另一计算系统分发程序的某些部分,那么由于多播的缘故,可以采用多种技术来阻止或限制组外的任何网络业务。举例而言,可以为多播分组指定较短的生存时间和/或使用分组寻址技术,使交换机不向未同该交换机相连的计算系统传输多播分组。此外,某些实施例可以实施多种策略,以最小化网络资源的使用量,最小化与传输或执行用以执行的程序副本无关的计算机上的负载,和/或提供网络和/或计算资源的可预测的性能。举例而言,无论对多播和/或点对点传输,某些实施例可以限制计算系统向其它计算系统传输程序副本的速度。此外,当中间网络设备在子网之间传输承载着程序副本的若干部分的数据分组时,某些实施例可以限制传输速度和/或限制中间网络设备(例如,交换机,路由器等)可以使用的网络带宽的比例。中间网络设备可以基于例如特定类型和/或发往特定地址(例如,属于特定范围内的多播IP地址)和/或端口来识别此类数据分组。在某些实施例中,可以对如上所述的多种机制加以组合以实现各种网络利用策略。

[0073] 在某些实施例中,还可以使用多种技术来将一个以上的执行程序实例从一个以上的计算系统迁移至一个以上的其它计算系统。在一方案中,所述迁移可以反映出与执行程序实例的初始计算系统相关的问题(例如,计算系统和/或对计算系统的访问失败)。在另一方案中,迁移可以使其它程序实例在初始计算系统上执行,以例如执行更高优先级的程序,或将程序实例的执行合并有限数目的计算系统上,以例如使原来执行程序实例的计算系统能够由于诸如维护、紧急保存等需要而得以关闭。作为一个特定示例,如果在计算系统上执行的一个以上的程序实例需要的资源多于从该计算系统所能获得的资源,一个以上的程序实例就可能需要迁移至具有更多资源的一个以上的其它计算系统。可用资源的过度使用可能是由于各种原因(诸如一个以上的计算系统所具有的资源比预期的要少,一个以上计算系统所使用的资源比预期(允许)的要多,或者在某些实施例中,相对于一个以上的保留的或正在执行的程序实例可能需要的资源而言,有意的过量使用一个以上的计算系统的可用资源)所导致的。举例而言,如果程序实例预期的资源需求在可用资源范围内,那么最大资源需求可能会超出可用资源。如果程序实例执行需要的实际资源超过可用资源也可能导致过度使用可用资源。程序迁移可以采用多种方式来执行,以例如将在初始计算系统上本

地保存的程序副本转移到目标目的地计算系统,和/或在目标目的地计算系统上开始执行执行于初始计算系统上的程序的新的实例。如果可能的话,迁移可发生在初始执行程序实例结束以前,以例如使当前的执行状态信息能够被传递到新的正在执行的程序实例中,和/或实现初始和新的程序实例间进行其它协调。

[0074] 某些实施例可以以收取费用的方式向多个客户提供程序执行服务。在这样的环境下,客户可以交付一定费用,以向程序执行服务注册或提供程序,并请求执行此程序。可以使用各种计费模型,使客户能够例如按时间(例如,分钟、小时、天等)购买对程序执行服务资源(例如,网络带宽、内存、存储器、处理器)的各种配置的访问权限,购买对一个以上的预定虚拟或物理硬件配置的访问权限,以附加费用购买额外付费服务(例如,提供执行优先权,以例如在发起执行非优先客户的程序之前发起执行优先客户的程序;提供程序储存库的放置优先权,以例如在清除优先客户的程序之前清除属于非优先客户的程序等);购买基于每实例执行在指定时段内执行程序实例的能力,等等。

[0075] 如上所述,某些实施例可以利用虚拟计算系统,并且倘若如此由程序执行服务执行的程序可以包括完整的虚拟计算机镜像。在此类实施例中,所要执行的程序可以包括:完整的操作系统、文件系统、和/或其它数据、还可能包括一个以上的用户级进程。在其它实施例中,所要执行的程序可以包括一种以上协同工作以提供某些功能的其它类型的可执行程序。在另外的其它实施例中,所要执行的程序可以包括一组物理或逻辑指令和数据,所述指令和数据可本地执行于供给计算系统上,也可以通过虚拟计算系统、解释器,或其它软件实现的硬件抽象间接执行。更一般地,在某些实施例中,所要执行的程序可以包括:一个以上的应用程序、应用框架、库、存档文件、类文件、脚本、配置文件、数据文件等。

[0076] 虽然对在程序执行服务中利用互通系统管理器模块和用于管理程序执行的机器管理器模块的组合来执行程序的实施例进行了说明,但是也可以考虑其它实现,以及在多个程序执行服务模块间分配职责。举例而言,在某些实施例中,单个模块或组件可以负责管理程序在某些或全部所管理的物理计算系统或虚拟机上的执行。举例而言,程序可以通过多种远程执行技术(例如,rexec、rsh等)直接执行于目标计算系统上。

[0077] 本领域的技术人员将认识到,虽然上述示例实施例被用在用数据中心提供程序执行服务的环境中,但是上述示例实施例也可以用于其它实现情形。举例而言,所述设施可用于由企业或其它机构(例如,大学)运营的部门级内联网环境,以为其员工和/或成员提供便利。可选地,所述技术可以为分布式计算系统所采用,从而以分布式方式执行大规模(例如,科学)计算任务,所述分布式计算系统包括分别受多个第三方管理和操作的节点。

[0078] 根据前述内容,将会理解到,虽然此处为了进行说明描述了特定的实施例,但是可以在不背离本发明的精神和范围的前提下,对本发明做大量的改动。因此,本发明只受所附权利要求以及在权利要求中所陈述的内容的限制。此外,虽然下面以特定的权利要求形式对本发明的某些方案进行了说明,但是发明者考虑了具有任意可用权利要求形式的本发明的各种方案。举例而言,虽然目前本发明只有某些方案被描述为包含于计算机可读介质,但是同样还可以将其它方案包含于计算机可读介质。

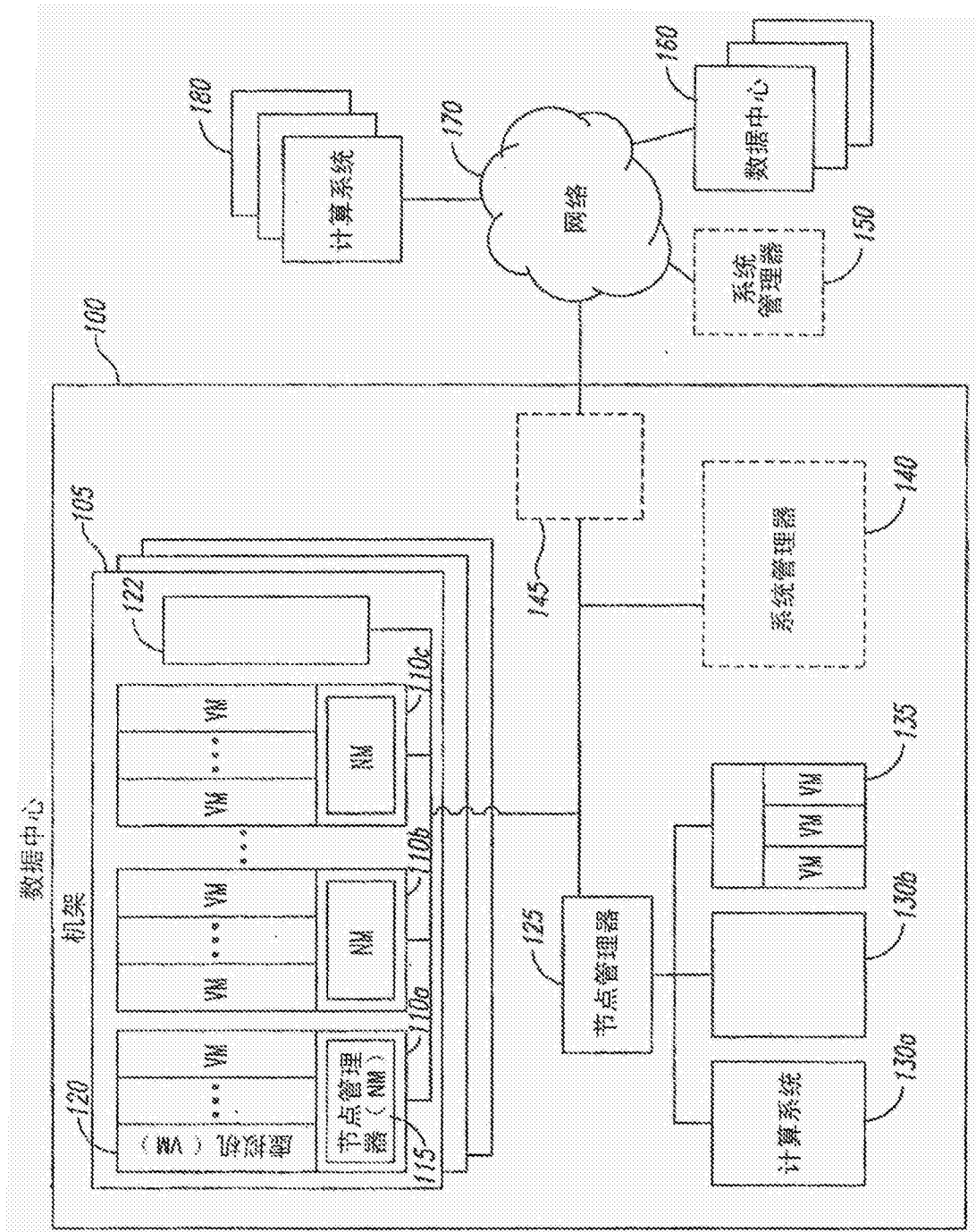


图1

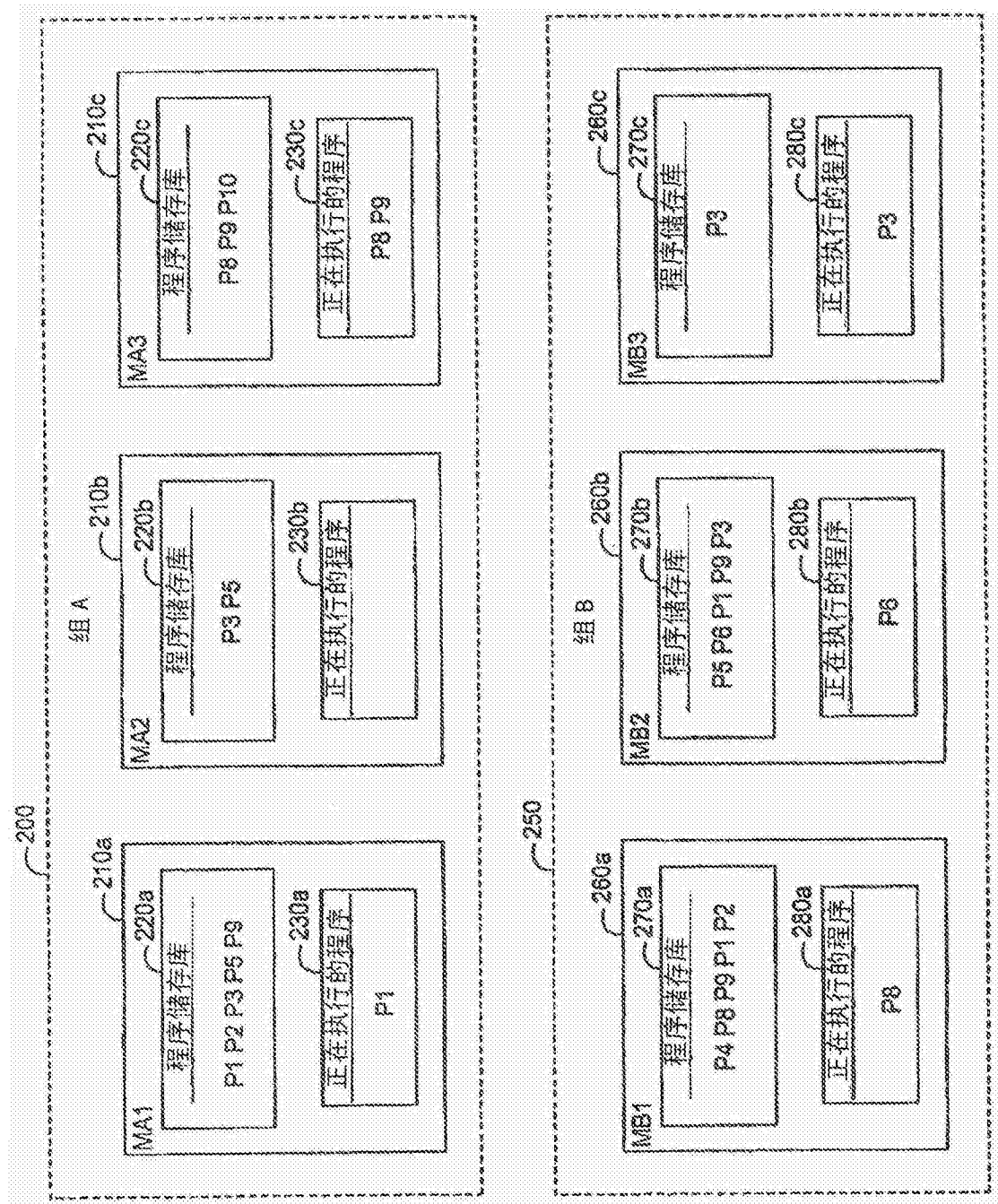


图2

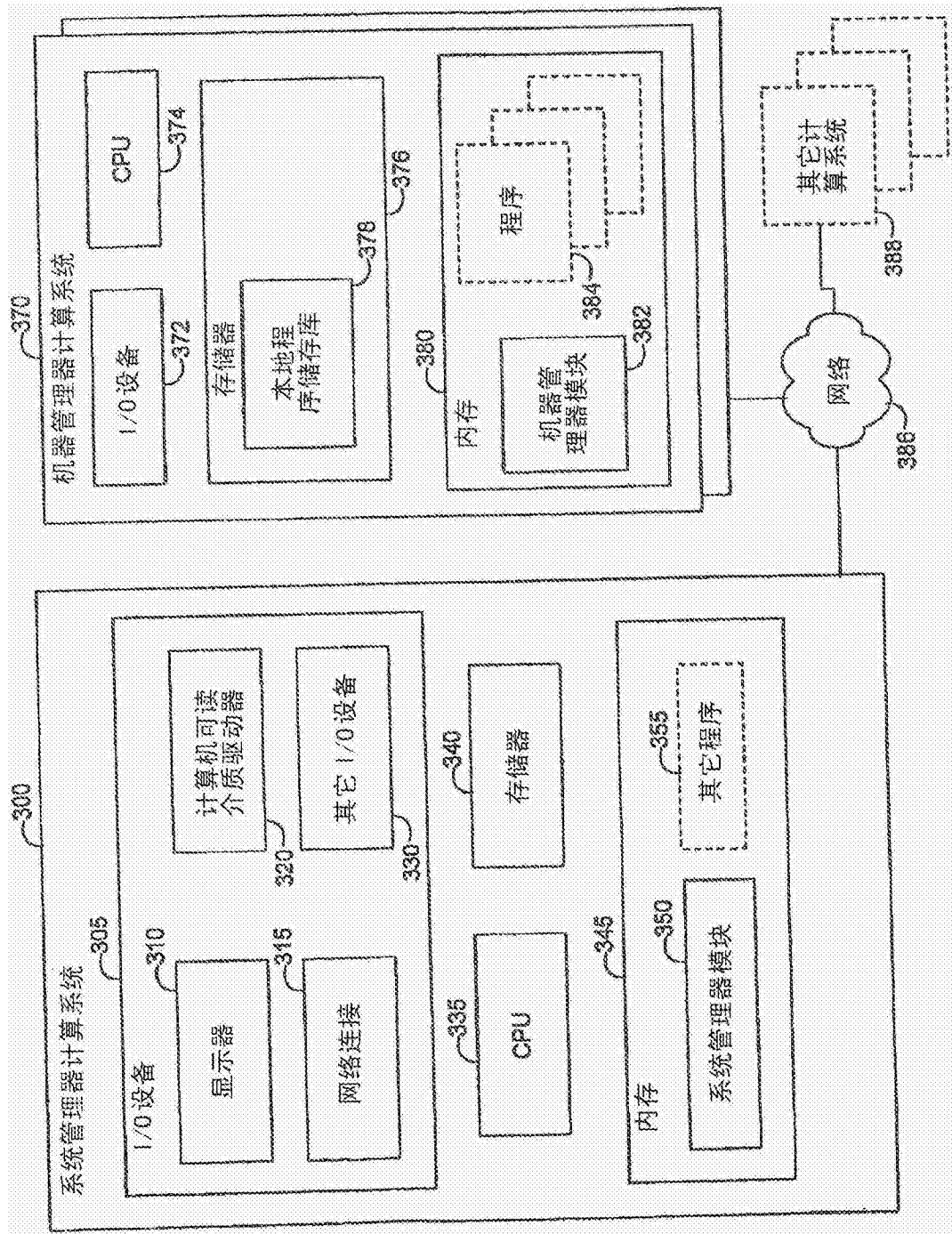


图3

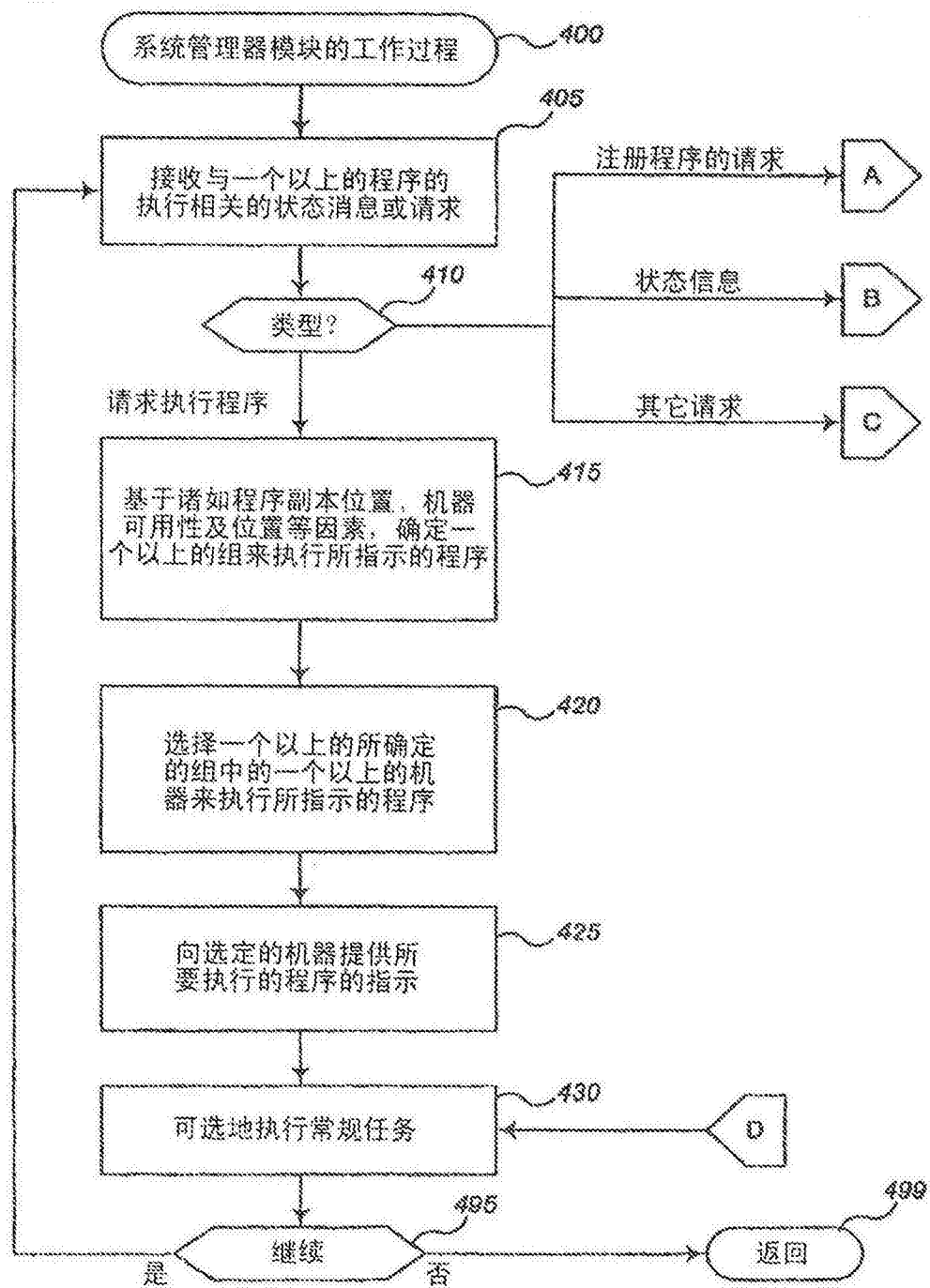


图4A

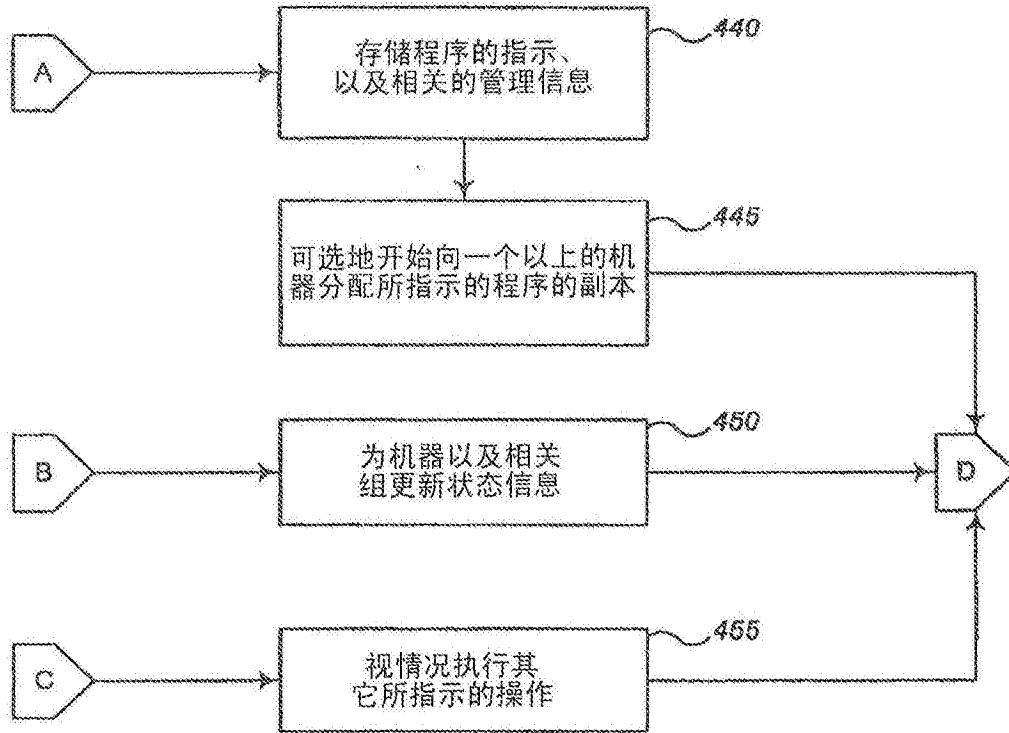


图4B

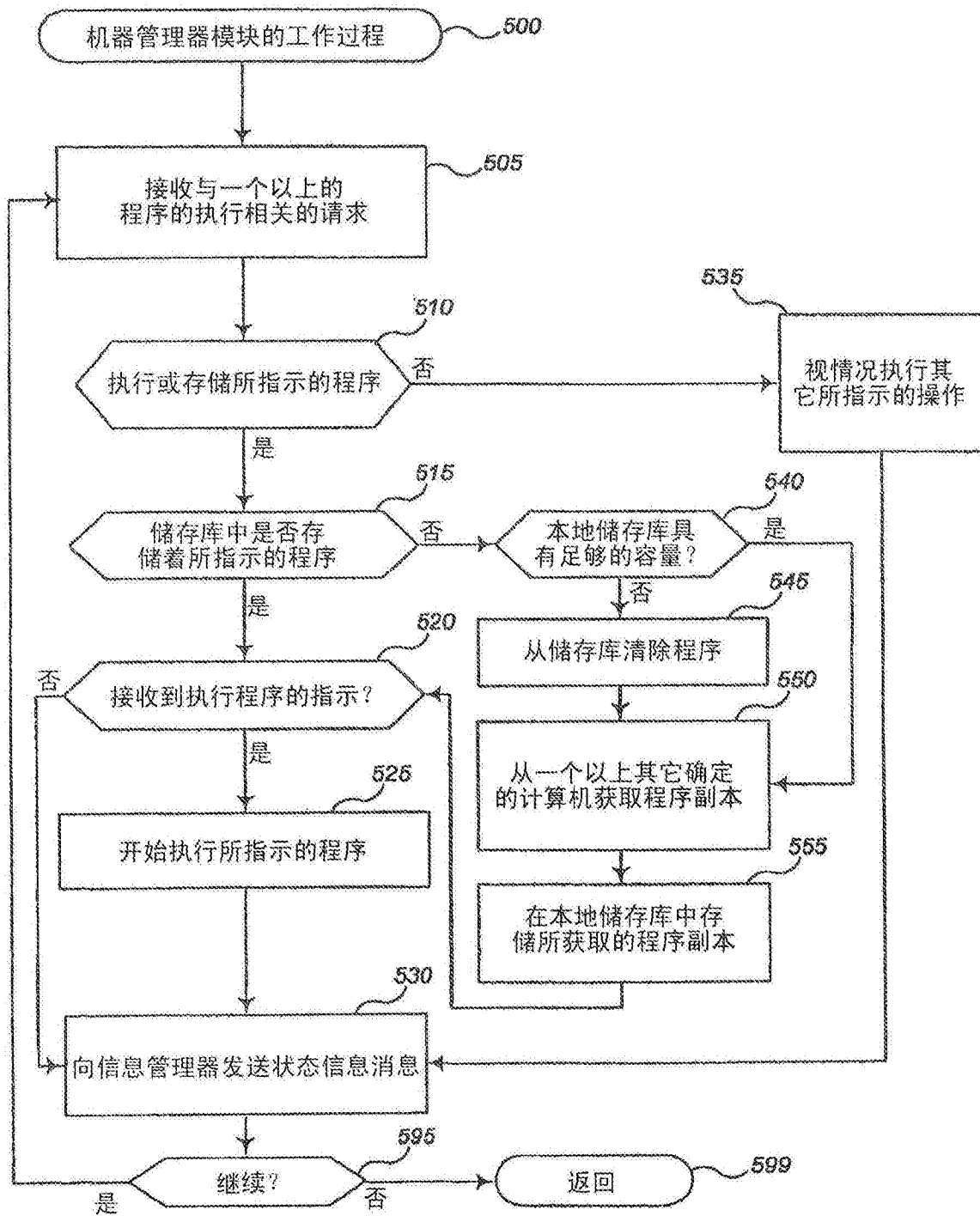


图5

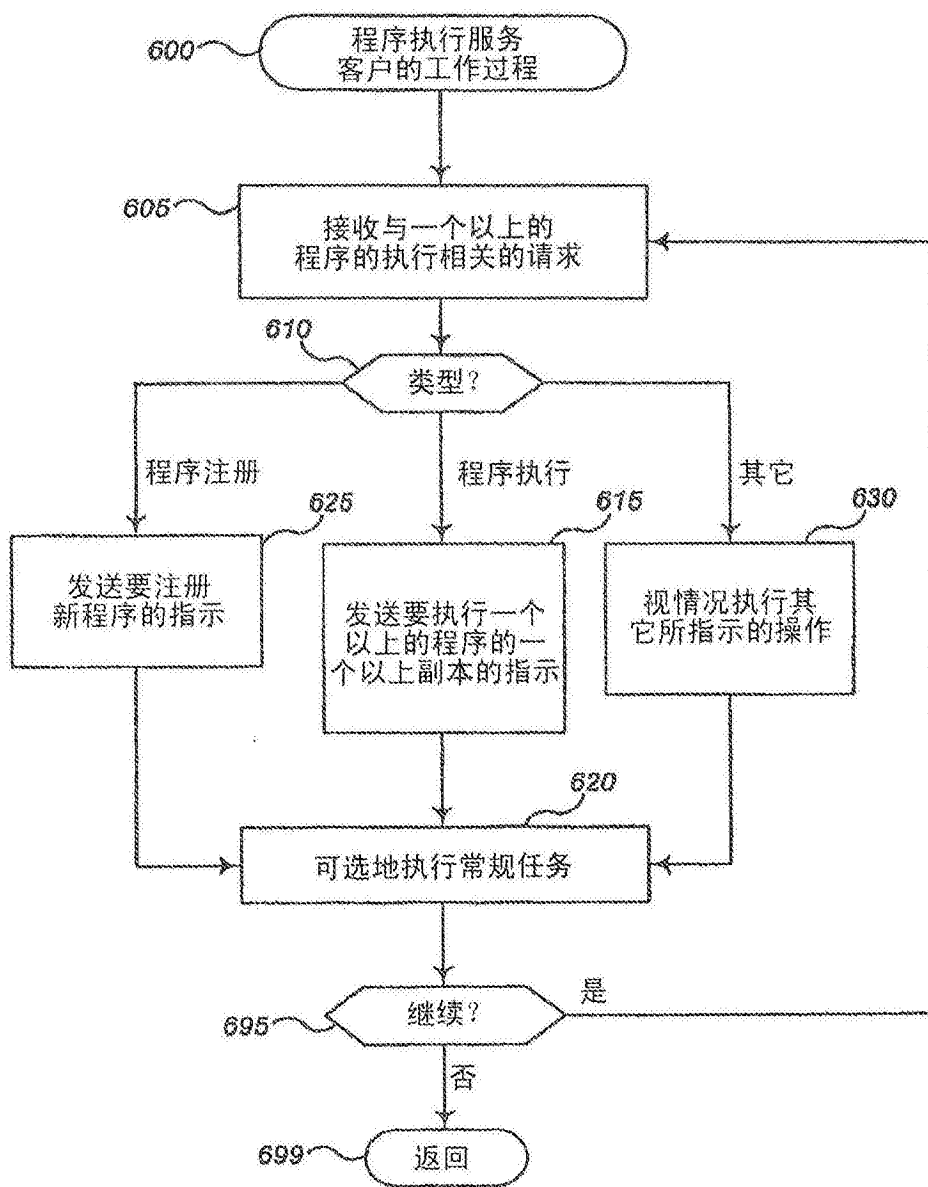


图6