



(12) 发明专利申请

(10) 申请公布号 CN 114450685 A

(43) 申请公布日 2022. 05. 06

(21) 申请号 202080066809.8

(72) 发明人 A·戈亚尔

(22) 申请日 2020.08.07

(74) 专利代理机构 中国贸促会专利商标事务所
有限公司 11038

(30) 优先权数据

62/884,931 2019.08.09 US

62/884,933 2019.08.09 US

16/986,158 2020.08.05 US

16/986,160 2020.08.05 US

专利代理师 周磊

(85) PCT国际申请进入国家阶段日

2022.03.23

(51) Int.Cl.

G06F 21/62 (2013.01)

G06F 9/455 (2006.01)

G06F 9/50 (2006.01)

G06F 21/31 (2013.01)

H04L 47/78 (2022.01)

H04L 47/80 (2022.01)

(86) PCT国际申请的申请数据

PCT/US2020/045514 2020.08.07

(87) PCT国际申请的公布数据

W02021/030219 EN 2021.02.18

(71) 申请人 甲骨文国际公司

地址 美国加利福尼亚

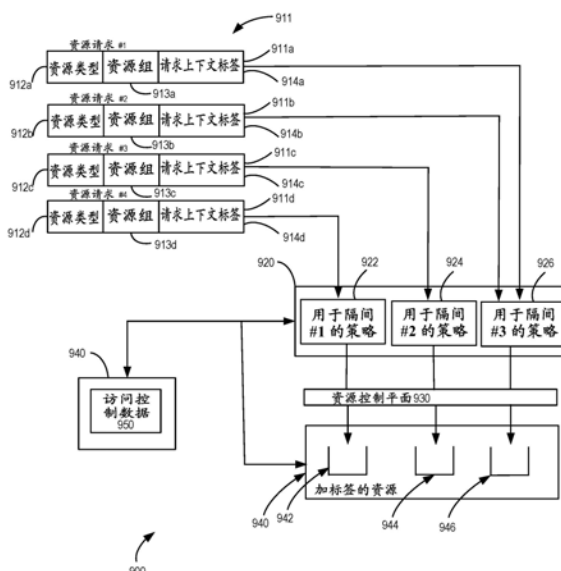
权利要求书6页 说明书28页 附图12页

(54) 发明名称

云基础设施环境中基于标签的资源限制或
配额的系统和方法

(57) 摘要

本文描述的系统和方法支持云基础设施环境中基于标签的资源限制或配额。云管理员一般没有能力限制现有云中的资源使用情况。授予用户创建资源的许可允许他们创建任何数量的资源，最多达到预定义的账户限制。标签与资源相关联，从而通过允许微调的成本控制允许管理员将用户的资源使用情况限制到适当的水平。确定与和供应资源的请求的请求特点对应的资源标签相关联的租赁中资源的使用情况，并将其与多个基于标签的配额进行比较，其中供应资源的请求基于所确定的使用情况超过所述多个基于标签的配额之一而被丢弃。



1. 一种用于相关联的云基础设施环境中资源使用情况的基于标签的控制的系统,该系统包括:

计算机,包括一个或多个微处理器;

租赁,在相关联的云基础设施环境中定义;以及

存储器设备,与计算机可操作地耦合,该存储器设备存储由计算机可执行的逻辑,用于提供对租赁中的资源使用情况的基于标签的控制,该存储器存储基于标签的配额数据,所述基于标签的配额数据代表租赁中资源供应的多个基于标签的配额,

其中接收在租赁中供应资源的请求,该请求包括代表请求的请求特点的请求特点数据;

其中确定与和请求的请求特点对应的资源标签相关联的租赁中资源的使用情况,并将其与所述多个基于标签的配额进行比较,

其中基于所确定的使用情况超过所述多个基于标签的配额之一而丢弃供应资源的请求。

2. 根据权利要求1所述的系统,其中:

租赁包括存储资源的多个隔间,其中所述多个隔间中的每个隔间提供该隔间内的资源集相对于其它隔间中的资源的一个或多个其它集合的隔离;

跨所述多个隔间共同确定与和请求的请求特点对应的资源标签相关联的租赁中资源的使用情况,并且将跨所述多个隔间共同确定的使用情况与租赁的所述多个基于标签的配额进行比较;以及

基于跨所述多个隔间共同确定的使用情况超过所述多个基于标签的配额之一,而丢弃供应资源的请求。

3. 根据权利要求1或2所述的系统,其中:

存储在存储器设备中的基于标签的配额数据代表对应的多个资源类型的租赁中的租赁中资源供应的多个基于标签的配额;

在租赁中供应资源的请求的请求特点数据包括代表所请求资源的第一资源类型的资源类型数据;

确定与和第一资源类型对应的资源标签相关联的租赁中资源的使用情况,并将其与第一资源类型的租赁中资源供应的第一基于标签的配额进行比较;以及

基于所确定的与和第一资源类型对应的资源标签相关联的租赁中资源的使用情况超过第一资源类型的租赁中资源供应的第一基于标签的配额,而丢弃供应第一资源类型的资源的请求。

4. 根据权利要求1、2或3所述的系统,其中:

存储在存储器设备中的基于标签的配额数据代表分配给租赁的对应多个用户组的租赁中资源供应的多个基于标签的配额;

在租赁中供应资源的请求的请求特点数据包括代表指派给请求资源的系统的用户的用户组类别的用户组数据;

确定与和用户组类别对应的资源标签相关联的租赁中资源的使用情况,并将其与分配给用户组类别的租赁中资源供应的第一基于标签的配额进行比较;以及

基于所确定的与和用户组类别对应的资源标签相关联的租赁中资源的使用情况超过

分配给用户组类别的租赁中资源供应的第一基于标签的配额,而丢弃供应资源的请求。

5. 根据权利要求1至4中的任一项所述的系统,其中:

存储在存储器设备中的基于标签的配额数据代表分配给租赁的多个用户组的多个资源类型的租赁中的租赁中资源供应的基于标签的配额;

在租赁中供应资源的请求的请求特点数据包括:i) 代表所请求资源的第一资源类型的资源类型数据;以及ii) 代表指派给请求资源的系统的用户的用户组类别的用户组数据;

确定与和供应给第一用户组类别的第一资源类型对应的资源标签相关联的租赁中资源的使用情况,并将其与分配给第一用户组类别的第一资源类型的租赁中资源供应的第一基于标签的配额进行比较;以及

基于所确定的与和供应给第一用户组类别的第一资源类型对应的资源标签相关联的租赁中资源的使用情况超过分配给第一用户组类别的第一资源类型的租赁中资源供应的第一基于标签的配额,而丢弃供应资源的请求。

6. 根据权利要求1至5中的任一项所述的系统,其中:

存储在存储器设备中的基于标签的配额数据代表分配给租赁的多个用户组的多个资源类型的租赁中的租赁中资源供应的基于标签的配额;

接收跟踪租赁的资源使用情况的第一请求,该第一请求包括代表第一用户组类别和第一资源类型的请求特点数据;

确定与和由第一用户组类别供应的第一资源类型对应的资源标签相关联的租赁中资源的第一使用情况;以及

资源使用情况跟踪数据是基于所确定的由第一用户组供应的第一资源类型的租赁中资源的第一使用情况生成的。

7. 根据权利要求1至6中的任一项所述的系统,其中:

基于所确定的使用情况超过所述多个基于标签的配额之一而丢弃供应资源的请求;

接收供应租赁的资源的覆盖请求;

基于系统接收到覆盖请求而选择性地供应资源;以及

资源使用情况超额数据是基于响应于系统接收到覆盖请求而选择性地供应资源而生成的。

8. 一种用于相关联的云基础设施环境中资源使用情况的基于标签的控制的方法,该方法包括:

由计算机在相关联的云基础设施环境中提供租赁,该计算机包括与该计算机可操作地耦合的一个或多个处理器和存储器设备,该存储器设备存储基于标签的控制逻辑;

将基于标签的配额数据存储在存储器设备中,基于标签的配额数据代表租赁中资源供应的多个基于标签的配额;

接收在租赁中供应资源的请求,该请求包括代表请求的请求特点的请求特点数据;

由执行基于标签的控制逻辑的所述一个或多个处理器确定与和请求的请求特点对应的资源标签相关联的租赁中资源的使用情况;

由执行基于标签的控制逻辑的所述一个或多个处理器将所确定的使用情况与所述多个基于标签的配额进行比较;以及

由执行基于标签的控制逻辑的所述一个或多个处理器基于所确定的使用情况超过所

述多个基于标签的配额之一而丢弃供应资源的请求。

9. 根据权利要求8所述的方法, 其中:

提供租赁包括提供存储资源的多个隔间, 其中所述多个隔间中的每个隔间提供该隔间内的资源集相对于其它隔间中的资源的一个或多个其它集合的隔离;

确定使用情况包括跨所述多个隔间共同确定与和请求的请求特点对应的资源标签相关联的租赁中资源的使用情况;

比较包括将跨所述多个隔间共同确定的使用情况与租赁的所述多个基于标签的配额进行比较; 以及

丢弃包括基于跨所述多个隔间共同确定的使用情况超过所述多个基于标签的配额之一而丢弃供应资源的请求。

10. 根据权利要求8或9所述的方法, 其中:

存储基于标签的配额数据包括将如下基于标签的配额数据存储在存储器设备中: 代表对应的多个资源类型的租赁中的租赁中资源供应的多个基于标签的配额的基于标签的配额数据;

接收请求包括接收包括代表所请求资源的第一资源类型的资源类型数据的请求;

确定包括确定与和第一资源类型对应的资源标签相关联的租赁中资源的使用情况被确定;

比较包括将所确定的使用情况与第一资源类型的租赁中资源供应的第一基于标签的配额进行比较; 以及

丢弃请求包括基于所确定的与和第一资源类型对应的资源标签相关联的租赁中资源的使用情况超过第一资源类型的租赁中资源供应的第一基于标签的配额而丢弃请求。

11. 根据权利要求8、9或10所述的方法, 其中:

存储基于标签的配额数据包括将如下基于标签的配额数据存储在存储器设备中: 代表分配给租赁的对应多个用户组的租赁中资源供应的多个基于标签的配额的基于标签的配额数据;

接收请求包括接收包括代表指派给请求资源的系统的用户的用户组类别的资源类型数据的请求;

确定包括确定与和用户组类别对应的资源标签相关联的租赁中资源的使用情况;

比较包括将所确定的使用情况与分配给用户组类别的租赁中资源供应的第一基于标签的配额进行比较; 以及

丢弃请求包括基于所确定的与和用户组类别对应的资源标签相关联的租赁中资源的使用情况超过分配给用户组类别的租赁中资源供应的第一基于标签的配额来丢弃请求。

12. 根据权利要求8至11中的任一项所述的方法, 其中:

存储基于标签的配额数据包括将如下基于标签的配额数据存储在存储器设备中: 代表分配给租赁的多个用户组的多个资源类型的租赁中的租赁中资源供应的基于标签的配额的基于标签的配额数据;

接收请求包括接收包括资源类型数据的请求, 该资源类型数据包括: i) 代表所请求资源的第一资源类型的资源类型数据; 以及 ii) 代表指派给请求资源的系统的用户的用户组类别的用户组数据;

确定包括确定与和供应给第一用户组类别的第一资源类型对应的资源标签相关联的租赁中资源的使用情况；

比较包括将所确定的使用情况与分配给第一用户组类别的第一资源类型的租赁中资源供应的第一基于标签的配额进行比较；以及

丢弃请求包括基于所确定的与和供应给第一用户组类别的第一资源类型对应的资源标签相关联的租赁中资源的使用情况超过分配给第一用户组类别的第一资源类型的租赁中资源供应的第一基于标签的配额而丢弃请求。

13. 根据权利要求8至12中的任一项所述的方法，还包括：

将如下基于标签的配额数据存储在存储器设备中：代表分配给租赁的多个用户组的多个资源类型的租赁中的租赁中资源供应的基于标签的配额的基于标签的配额数据；

接收跟踪租赁的资源使用情况的第一请求被接收到，该第一请求包括代表第一用户组类别和第一资源类型的请求特点数据；

确定与和由第一用户组类别供应的第一资源类型对应的资源标签相关联的租赁中资源的第一使用情况；以及

基于所确定的由第一用户组供应的第一资源类型的租赁中资源的第一使用情况，生成资源使用情况跟踪数据。

14. 根据权利要求8至13中的任一项所述的方法，还包括：

接收供应租赁的资源的覆盖请求；

基于系统接收到覆盖请求而选择性地供应资源；以及

基于响应于系统接收到覆盖请求而选择性地供应资源而生成资源使用情况超额数据。

15. 一种计算机可读介质，其上具有用于相关联的云基础设施环境中资源使用情况的基于标签的控制的指令，所述指令在由计算机读取和执行时使所述计算机执行包括以下的步骤：

由计算机在相关联的云基础设施环境中提供租赁，该计算机包括与该计算机可操作地耦合的一个或多个处理器和存储器设备，该存储器设备存储基于标签的控制逻辑；

将基于标签的配额数据存储在存储器设备中，基于标签的配额数据代表租赁中资源供应的多个基于标签的配额；

接收在租赁中供应资源的请求，该请求包括代表请求的请求特点的请求特点数据；

由执行基于标签的控制逻辑的所述一个或多个处理器确定与和请求的请求特点对应的资源标签相关联的租赁中资源的使用情况；

由执行基于标签的控制逻辑的所述一个或多个处理器将所确定的使用情况与所述多个基于标签的配额进行比较；以及

由执行基于标签的控制逻辑的所述一个或多个处理器基于所确定的使用情况超过所述多个基于标签的配额之一而丢弃供应资源的请求。

16. 根据权利要求15所述的介质，其中：

提供租赁包括提供存储资源的多个隔间，其中所述多个隔间中的每个隔间提供该隔间内的资源集相对于其它隔间中的资源的一个或多个其它集合的隔离；

确定使用情况包括跨所述多个隔间共同确定与和请求的请求特点对应的资源标签相关联的租赁中资源的使用情况；

比较包括将跨所述多个隔间共同确定的使用情况与租赁的所述多个基于标签的配额进行比较;以及

丢弃包括基于跨所述多个隔间共同确定的使用情况超过所述多个基于标签的配额之一而丢弃供应资源的请求。

17. 根据权利要求15或16所述的介质, 其中:

存储基于标签的配额数据包括将如下基于标签的配额数据存储在存储器设备中: 代表对应的多个资源类型的租赁中的租赁中资源供应的多个基于标签的配额的基于标签的配额数据;

接收请求包括接收包括代表所请求资源的第一资源类型的资源类型数据的请求;

确定包括确定与和第一资源类型对应的资源标签相关联的租赁中资源的使用情况被确定;

比较包括将所确定的使用情况与第一资源类型的租赁中资源供应的第一基于标签的配额进行比较; 以及

丢弃请求包括基于所确定的与和第一资源类型对应的资源标签相关联的租赁中资源的使用情况超过第一资源类型的租赁中资源供应的第一基于标签的配额而丢弃请求。

18. 根据权利要求15、16或17所述的介质, 其中:

存储基于标签的配额数据包括将如下基于标签的配额数据存储在存储器设备中: 代表分配给租赁的对应多个用户组的租赁中资源供应的多个基于标签的配额的基于标签的配额数据;

接收请求包括接收包括代表指派给请求资源的系统的用户的用户组类别的资源类型数据的请求;

确定包括确定与和用户组类别对应的资源标签相关联的租赁中资源的使用情况;

比较包括将所确定的使用情况与分配给用户组类别的租赁中资源供应的第一基于标签的配额进行比较; 以及

丢弃请求包括基于所确定的与和用户组类别对应的资源标签相关联的租赁中资源的使用情况超过分配给用户组类别的租赁中资源供应的第一基于标签的配额来丢弃请求。

19. 根据权利要求15至18中的任一项所述的介质, 其中:

存储基于标签的配额数据包括将如下基于标签的配额数据存储在存储器设备中: 代表分配给租赁的多个用户组的多个资源类型的租赁中的租赁中资源供应的基于标签的配额的基于标签的配额数据;

接收请求包括接收包括资源类型数据的请求, 该资源类型数据包括: i) 代表所请求资源的第一资源类型的资源类型数据; 以及 ii) 代表指派给请求资源的系统的用户的用户组类别的用户组数据;

确定包括确定与和供应给第一用户组类别的第一资源类型对应的资源标签相关联的租赁中资源的使用情况;

比较包括将所确定的使用情况与分配给第一用户组类别的第一资源类型的租赁中资源供应的第一基于标签的配额进行比较; 以及

丢弃请求包括基于所确定的与和供应给第一用户组类别的第一资源类型对应的资源标签相关联的租赁中资源的使用情况超过分配给第一用户组类别的第一资源类型的租赁

中资源供应的第一基于标签的配额而丢弃请求。

20. 根据权利要求15至19中的任一项所述的介质,还包括:

将如下基于标签的配额数据存储在存储器设备中:代表分配给租赁的多个用户组的多个资源类型的租赁中的租赁中资源供应的基于标签的配额的基于标签的配额数据;

接收跟踪租赁的资源使用情况的第一请求被接收到,该第一请求包括代表第一用户组类别和第一资源类型的请求特点数据;

确定与和由第一用户组类别供应的第一资源类型对应的资源标签相关联的租赁中资源的第一使用情况;以及

基于所确定的由第一用户组供应的第一资源类型的租赁中资源的第一使用情况,生成资源使用情况跟踪数据。

云基础设施环境中基于标签的资源限制或配额的系统和方法

[0001] 版权声明

[0002] 本专利文档的部分公开内容包含受版权保护的材料。版权所有者不反对任何人对专利文档或专利公开内容进行传真复制,因为它出现在专利商标局的专利文档或记录中,但除此之外保留所有版权权利。

[0003] 优先权声明和相关申请的交叉引用:

[0004] 本申请要求于2019年8月9日提交的标题为“SYSTEM AND METHOD FOR TAG BASED RESOURCE LIMITS OR QUOTAS IN A CLOUD INFRASTRUCTURE ENVIRONMENT”的美国临时专利申请No.62/884,931;以及2019年8月9日提交的标题为“SYSTEM AND METHOD FOR TAG BASED REQUEST CONTEXT IN A CLOUD INFRASTRUCTURE ENVIRONMENT”的美国临时专利申请No.62/884,933的优先权权益;上述申请中的每一个都通过引用整体并入本文。

[0005] 本申请还要求以下申请的优先权权益:于2020年8月5日提交的标题为“SYSTEM AND METHOD FOR TAG BASED RESOURCE LIMITS OR QUOTAS IN A CLOUD INFRASTRUCTURE ENVIRONMENT”的美国专利申请No.16/986,158;于2020年8月5日提交的标题为“SYSTEM AND METHOD FOR TAG BASED REQUEST CONTEXT IN A CLOUD INFRASTRUCTURE ENVIRONMENT”的美国专利申请No.16/986,160;上述申请中的每一个都通过引用整体并入本文。

技术领域

[0006] 本文描述的实施例一般而言与云基础设施环境有关,诸如基础设施即服务(IaaS),并且特别涉及用于提供用于在此类云基础设施环境内提供资源约束的系统和方法的系统和方法。

背景技术

[0007] 云基础设施环境可以包括使用户和客户端(在整个说明书中,术语“客户端”和“客户”可以互换使用)能够在高度可用的被托管环境中构建和运行广泛的应用和服务的互补云服务的集合。

[0008] 年复一年,越来越多的企业和组织正在将任务关键应用和系统迁移到云基础设施环境。这种转变有多种原因。例如,许多企业正在移动到云,以便降低运营、维护和构建本地基础设施的成本和复杂性。同样,云基础设施还允许更快速的信息技术(IT)递送机制。一些企业和组织还将云基础设施环境视为通过适应更灵活的系统来获得竞争优势的一种手段。

[0009] 在IaaS(基础设施即服务)模型内,云提供商可以提供、托管和管理基础设施组件,这些组件在传统设置中将在每个客户/客户端的位置进行本地部署(on-premise)。传统上在本地提供的此类组件可以包括硬件(例如数据仓库和数据中心、服务器、存储装置、联网硬件)以及软件(诸如虚拟化软件)。

[0010] IaaS提供商除了提供传统上本地部署的硬件和软件之外,还可以向他们的客户端和客户提供服务。作为示例,可以允许客户端和客户定制他们的IaaS订阅以满足他们的需求,这进而允许详细和细分的计费 and 发票。IaaS还可以支持诸如负载平衡、冗余性、复制和

恢复之类的特征。因为此类服务是由IaaS提供商(而不是客户)提供和支持的,因此这让客户端和客户更加专注于通过更多地推动其服务的自动化和编排来改善他们的服务。

[0011] 云基础设施使用户和客户端能够都在同一平台上无缝地运行传统企业应用以及云原生应用,从而减少操作开销并启用两种类型的工作负载之间的直接连接性。

发明内容

[0012] 本文描述的是用于在云基础设施环境中提供基于标签的资源限制或配额的系统和方法。本文描述的系统和方法支持云基础设施环境中基于标签的资源限制/配额。细粒度的方法可以跨越多个容器基于标签来提供资源限制。标签是主要用在资源治理中的机制,并且云提供商也将其用于成本治理。系统和方法可以创建一种机制来通过标签在组级别控制成本。系统和方法为云基础设施环境提供隔间配额。

[0013] 云管理员具有限制现有云中的资源使用的能力,但仅在高级别并且没有特定用户、特定资源和资源类型的视角或容器分层结构级别的资源的视角。根据实施例,资源与标签相关联,使得它们的使用可以在必要或期望的资源级别被控制、跟踪和/或以其它方式处置。提供资源配额和/或限制允许管理员和其他人将用户的资源使用限制在适当的级别,从而允许经微调的成本控制。

[0014] 根据实施例,可以在账户创建时向客户指派由云基础设施环境定义的服务级别限制。这些服务级别限制限制了客户跨整个租赁(例如,跨具有多个隔间的多个区域,以及跨越多个容器)可以创建的资源总数。租赁和隔间管理员可以利用基于标签的资源配额来设置特定于资源的限制。如果没有基于与资源相关联的标签的此类限制,那么被授权启动实例的用户可以消耗整个租赁中所有可用容量。基于标签的资源限制解决了这个问题,并且与服务限制不同,它由客户端和客户经由例如控制台、SDK或API来设置和定制。基于标签的资源限制可以应用在服务限制之上,并通过嵌套的隔间层次结构继承。这允许隔间管理员限制资源消耗并围绕可接受的资源使用来设置边界。

附图说明

[0015] 图1图示了根据实施例的用于提供云基础设施环境的系统。

[0016] 图2图示了根据实施例的用于在云基础设施环境内提供云基础设施区域的系统。

[0017] 图3示出了根据实施例的云基础设施环境系统,其图示了用于跨越云基础设施区域的策略管理和控制的隔间、隔间策略、子隔间与子隔间策略之间的关系。

[0018] 图4示出了云基础设施环境400,其图示了当移动隔间时,隔间、隔间策略、子隔间与子隔间策略之间的关系。

[0019] 图5示出了云基础设施环境500,其图示了在移动隔间时策略的含义。

[0020] 图6示出了根据示例实施例的一对定义的标签。

[0021] 图7示出了根据示例实施例的基于标签对云基础设施环境中的资源强制执行配额或限制的系统的体系架构,该标签包括例如资源标签和资源请求上下文标签。

[0022] 图8示出了使用请求上下文标签和/或资源标签来在云基础设施环境中限制诸如资源的供应之类的使用的系统。

[0023] 图9是根据实施例的在云基础设施环境中提供基于资源请求上下文标签的限制/

配额的系统的功能示意图。

[0024] 图10是示出根据示例实施例的用于基于请求上下文在云基础设施环境中对供应资源进行限制或强加配额的方法的流程图。

[0025] 图11是根据实施例的在云基础设施环境中提供基于标签的资源限制/配额的系统的功能示意图。

[0026] 图12是示出根据示例实施例的用于基于资源标签在云基础设施环境中对供应资源进行限制或强加配额的方法的流程图。

具体实施方式

[0027] 根据本文的实施例,术语“裸机主机”、“裸机实例”或“裸机计算实例”可以指提供以供云基础设施环境使用的物理主机(“裸机”)机器。裸机计算实例直接在裸机服务器上运行,无需管理程序。当供应裸机实例时,用户/客户端可以维护物理CPU、存储器和网络接口卡(NIC)的单独控制。用户/客户端可以配置和利用每个物理机器的全部能力,就好像它是在他们自己的数据中心中运行的硬件一样。用户/客户端不与任何其它租户共享物理机器。

[0028] 根据本文的实施例,云基础设施环境可以物理地托管在区域和可用性域中。术语“区域”可以指局部地理地区。术语“可用性域”可以表示位于区域内的一个或多个数据中心。区域由一个或多个可用性域组成。云基础设施可以是特定于区域的,诸如虚拟云网络,或者是特定于可用性域的,诸如计算实例。可用性域彼此隔离,具有容错性,并且不太可能同时发生故障或受到另一个可用性域的故障的影响。当用户/客户端配置他们的云服务时,用户/客户端可以使用多个可用性域来确保高可用性并防止资源故障。一些资源必须在同一个可用性域内创建,诸如实例和附连到它的存储卷。

[0029] 根据本文的实施例,术语“领域”可以指区域的逻辑集合。领域彼此隔离并且不共享任何数据。客户端的租赁可以存在于单个领域中,并且可以访问属于那个领域的每个区域。例如,可以为商业区域和政府云区域提供领域。

[0030] 根据本文的实施例,术语“控制台”可以指用户/客户端可以用来访问和管理云基础设施环境的用户接口(例如,基于web的)。

[0031] 根据本文的实施例,术语“租赁”可以指云基础设施环境内的安全且隔离的分区。当客户端/用户向云基础设施环境注册时,云基础设施环境可以为用户/客户端(或他们的公司)创建租赁。租赁可以是用户/客户端创建、组织和管理其云资源的地方。

[0032] 根据本文的实施例,术语“隔间”可以指相关资源(诸如实例、虚拟云网络、块卷(block volume))的集合,这些资源只能由管理员已给予许可的某些组访问。隔间允许客户端/用户组织和控制对其云资源的访问。当客户端/用户开始使用控制台中的资源工作时,隔间可以为客户端/用户正在查看的内容提供过滤器。

[0033] 根据本文的实施例,术语“租赁”可以指包括所有租户的云资源的根隔间。可以在租赁(根隔间)内创建附加隔间,并使用对应的策略来控制对每个隔间中的资源的访问。当用户/客户端创建云资源(诸如实例、块卷或云网络)时,用户/客户端可以指定资源应当属于哪个隔间。

[0034] 根据本文的实施例,术语“虚拟云网络”可以指在其上运行实例的传统网络(包括子网、路由表和网关)的虚拟版本。云网络驻留在单个区域内,但包括该区域的所有可用性

域。云网络中定义的每个子网可以或者位于单个可用性域中,或者可以跨越区域中的所有可用性域。在可以启动实例之前,可以设置一个至少一个云网络。

[0035] 根据本文的实施例,术语“实例”可以指在云中运行的计算主机。与传统的基于软件的虚拟机相反,云基础设施环境实例可以允许用户/客户端使用被托管的物理硬件,从而提供更高的安全性和性能。

[0036] 根据本文的实施例,术语“映像”可以指虚拟硬盘驱动器的模板,该模板为实例定义操作系统和其它软件,例如,LINUX。当启动实例时,可以通过选择其映像来定义其特点。可以提供映像的集合,并且可以使用来自已经配置(例如,作为模板)的现有实例的映像以便使用具有相同软件和定制的实例。

[0037] 根据本文的实施例,术语“形状”可以指分配给实例的CPU的数量和存储器的量的规范。在负载平衡中,形状可以确定负载平衡器为入口加出口流量预先供应的总最大容量(带宽)。可用的形状可以包括例如100Mbps、400Mbps和8000Mbps。

[0038] 根据本文的实施例,术语“密钥对”可以指使用云基础设施环境的认证机制。密钥对由私钥文件和公钥文件组成。可以上传公钥,而私钥保存在用户/客户端的设备上。私钥对用户/客户端来说是私有的,就像密码一样。可以根据不同的规范生成密钥对,诸如实例SSH密钥对,以及API签署密钥对。实例SSH密钥对被用于建立与实例的安全外壳(SSH)连接。当供应实例时,可以提供公钥,该公钥保存在实例的经授权的密钥文件中。为了登录到实例,提供私钥,该私钥用公钥进行验证。API签署密钥对是PEM(隐私增强邮件)格式并且被用于在提交API请求时对用户/客户端进行认证。

[0039] 根据本文的实施例,术语“块卷”可以指为云基础设施环境实例提供持久块存储的虚拟盘。可以以与例如计算机上的物理硬盘驱动器相似的方式使用块卷来例如存储数据和应用。卷可以从一个实例分离并附连到另一个实例,而不会丢失数据。

[0040] 根据本文的实施例,术语“对象存储”可以指可以被用于将数据作为对象进行存储和管理的存储体系架构。数据文件可以是任何类型。一旦数据被上传到对象存储,就可以从任何地方访问它。当需要存储大量数据并且数据不是很频繁地改变时,可以使用对象存储。对象存储的典型用途包括但不限于数据备份、文件共享、存储非结构化数据(例如,日志和传感器生成的数据)。

[0041] 根据本文的实施例,术语“桶”(bucket)可以指用于存储数据和文件的逻辑容器。桶可以包含无限数量的对象。

[0042] 如上所述,云基础设施环境可以包括使用户和客户端能够在高度可用的被托管的环境中构建和运行广泛的应用和服务的互补云服务的集合。

[0043] 图1示出了根据实施例的用于提供云基础设施环境的系统。

[0044] 根据实施例,可以在多个硬件和软件资源112上运行的云基础设施环境100可以包括控制台接口102和API 104。此外,云基础设施环境100可以支持多个治理服务110、身份和访问管理(IAM)服务120以及供应服务130。云基础设施环境100还可以支持多个资源140,例如在层中,诸如计算机资源层150、网络资源层160和存储资源层170。云基础设施环境100还可以支持与每个资源相关联的多个标签,包括例如一般与资源140相关联的资源标签140a、与计算机资源150相关联的计算机资源标签150a、与网络资源160相关联的网络资源标签160a、以及与存储资源170相关联的存储资源标签170a。

[0045] 根据实施例,例如,诸如具有设备硬件(处理器、存储器……等)12的计算设备10之类的客户端设备可以经由诸如广域网(WAN)、局域网(LAN)或互联网之类的网络与云基础设施环境通信。客户端设备可以包括管理员应用14,其可以包括用户界面16。

[0046] 根据实施例,在云基础设施环境内,可以支持租赁。在注册和部署时,可以为每个客户端/客户创建租赁,该租赁可以包括云基础设施内的安全且隔离的分区,客户端可以在其中创建、组织和管理他们的云资源。

[0047] 根据实施例,控制台接口102和API 104可以为客户端提供对云基础设施环境的相应部分的访问和控制。根据实施例,控制台接口可以包括直观的图形界面,其允许客户端创建和管理资源、实例、云网络和存储卷,以及管理与客户端相关联的用户,并在客户端范围内设置许可。同样,API 104可以包括例如利用HTTPS(安全超文本传输协议)的REST API。

[0048] 根据实施例,控制台接口或API的一个示例可以是配置管理工具(例如,Ansible)。配置管理工具可以被用于云基础设施供应、编排和配置管理。配置管理工具可以允许客户端将配置和供应云基础设施、部署和更新软件资产以及编排复杂的操作过程自动化。

[0049] 根据实施例,云基础设施环境的治理服务110向客户端提供工具以帮助客户端启用简单的资源治理、管理成本以及控制对云基础设施的访问。作为示例,治理服务提供加标签(tagging),这允许客户端出于信息或操作原因而将标签应用于其资源。可以控制定义的标签,以避免将不正确的标签应用于资源。标签还可以为管理脚本提供灵活的选定目标机制。同样,治理服务可以允许受管理的预算,并从一个地方跟踪所有实际和预测的支出。这允许客户端通过成本分析仪表板掌握使用情况,并按隔间和标签进行过滤,以分析部门、团队和项目的支出。这种数据也可以被导出,用于详细的资源利用报告以及与现有云管理和商业智能工具的集成。治理服务还可以记录事件,事件以后可以被检索、存储和分析以用于跨云基础设施权利和隔间的安全性、合规性和资源优化。

[0050] 根据示例实施例,治理服务提供加标签,从而允许客户端、管理员等在资源正被实例化时出于信息或操作原因将标签应用于他们的资源。根据进一步的示例实施例,治理服务还提供加标签,从而允许客户端和其他人在资源已被实例化之后出于信息或操作原因将标签应用于他们的资源,从而允许使用标签来追溯系统中的资源配额或限制的强制执行。

[0051] 根据实施例,身份和访问管理(IAM)服务120可以为IAM服务中的每个客户端/客户/用户创建与用户凭证(例如,用户名和密码)相关联的用户简档。客户端也可以经由IAM服务在云基础设施中被准予管理员特权。

[0052] 根据实施例,身份和访问管理服务可以与云基础设施环境集成。在客户端注册后。IAM服务可以在身份服务中创建单独的用户凭证,然后可以允许单点登录到云基础设施服务以及访问附加的云服务。

[0053] 根据实施例,供应服务130可以供应例如云基础设施服务内的租赁,诸如在资源140内。可以通过例如控制台接口或经由一个或多个API(诸如API 104)来访问和控制该供应服务。供应服务可以允许客户端供应和管理计算主机,这些主机可以被称为实例。客户端可以根据需要启动实例以满足计算和应用要求。在客户端启动实例之后,可以从例如客户端设备访问所供应的实例。供应服务还可以提供重新启动实例、从实例附连和分离卷,以及终止实例。

[0054] 根据实施例,由云基础设施环境提供的资源140可以被分解成多个层,诸如计算资

源层150、网络资源层160和存储资源层170。

[0055] 根据实施例,计算资源层150可以包括多个资源,诸如例如裸机实例152、虚拟机154、边缘服务156和容器158。例如,计算资源层可以被用于供应和管理裸机计算实例,根据需要供应实例以部署和运行应用,就像在本地数据中心中一样。根据示例实施例的云基础设施环境100支持与每个资源相关联的多个标签140a,包括例如与包括例如裸机实例152、虚拟机154、边缘服务156和容器158的计算机资源150相关联的计算机资源标签150a。

[0056] 根据实施例,云基础设施环境可以提供对计算资源层内的一个或多个物理主机(“裸机”)机器的控制。裸机计算实例直接在裸机服务器上运行,无需管理程序。当供应裸机计算实例时,客户端可以保持对物理CPU、存储器和网络接口卡(NIC)的单独控制。裸机计算实例可以被配置并利用每个物理机器的全部能力,就好像它是在本地自己的数据中心中运行的硬件一样。因此,裸机计算实例一般不在租户之间共享。

[0057] 根据实施例,与基于软件的虚拟环境相反,裸机计算实例可以经由相关联的物理硬件提供高级别的安全性和性能。

[0058] 根据实施例,云基础设施环境可以提供对计算资源层内的多个虚拟机的控制。例如,可以从可以确定虚拟机操作系统以及其它软件的映像来启动虚拟机计算主机。例如,可以基于从其启动虚拟机的映像来确定可用于那个虚拟机实例的资源类型和数量。

[0059] 根据实施例,虚拟机(VM)计算实例可以包括在物理裸机硬件之上运行的独立计算环境。虚拟化使得运行彼此隔离的多个虚拟机成为可能。例如,VM可以用于运行不要求整个物理机器的性能和资源(CPU、存储器、网络带宽、存储)的应用。

[0060] 在一些实施例中,虚拟机实例可以在与裸机实例相同的硬件上运行,这可以提供对使用相同的云优化的硬件、固件、软件堆栈和联网基础设施的充分利用。

[0061] 根据实施例,云基础设施环境可以在计算资源层内提供多个图形处理单元(GPU)计算实例。加速的计算要求跨每项服务的一致快速的基础设施。利用GPU实例,客户端可以更高效地处理和分析海量数据集,使其对于复杂的机器学习(ML)、人工智能(AI)算法和许多工业HPC应用有用。GPU计算实例可以被供应为虚拟化的计算实例(其中多个GPU计算实例共享相同的裸机硬件),或者可以被供应为为每个GPU计算实例提供专用硬件的裸机实例。

[0062] 根据实施例,云基础设施环境可以在计算资源层内提供多个容器化的计算实例。独立的容器引擎服务可以被用于构建容器化的应用并将其启动到云。例如,容器服务可以被用于构建、部署和管理云原生应用。容器服务可以指定容器化的应用所需的计算资源,然后容器引擎可以经由供应服务来供应在云基础设施环境内(例如,在租赁的上下文中)使用的所需计算资源。

[0063] 根据实施例,可以被使用的一个这样的容器服务引擎是Kubernetes,用于将跨主机的集群部署、缩放和管理容器化应用自动化的开源系统。此类容器服务可以将构成应用的容器分组为逻辑单元,以便容易地管理和发现。

[0064] 根据实施例,网络资源层160可以包括多个资源,诸如例如虚拟云网络(VCN)162、负载均衡器164、边缘服务166和连接服务168。根据示例实施例的云基础设施环境100支持与每个资源相关联的多个标签,包括例如一般与资源140相关联的资源标签140a,以及与虚拟云网络(VCN)162、负载均衡器164、边缘服务166和连接服务168相关联的网络资源标签160a。

[0065] 根据实施例,云基础设施环境可以在联网资源层处提供多个虚拟云网络162。虚拟云网络可以包括客户端实例可以在其上运行的传统网络(包括子网、路由表和网关)的虚拟版本。云网络驻留在单个区域内,但包括该区域的所有可用性域。云网络中定义的每个子网可以或者位于单个可用性域中,或者跨越区域中的所有可用性域(推荐)。在启动实例之前,可以配置至少一个云网络。在某些实施例中,可以经由互联网网关配置VCN以处置公共流量、VPN连接或快速连接服务以安全地扩展本地网络。

[0066] 根据实施例,云基础设施环境可以在联网资源层处提供多个负载平衡器164。负载平衡服务可以提供从一个入口点到可从虚拟云网络(VCN)到达的多个服务器的自动化流量分布。各种负载平衡可以提供公共或私有IP地址以及供应的带宽。

[0067] 根据实施例,负载平衡器可以改进资源利用、缩放并帮助确保高可用性。可以配置多个负载平衡策略,并且可以提供特定于应用的健康检查以确保负载平衡器仅将流量指向健康的实例。负载平衡器可以通过在将不健康的应用服务器从服务中移除以进行维护之前排出流量来缩短维护窗口。

[0068] 根据实施例,负载平衡服务使得能够结合VCN创建公共或私有负载平衡器。公共负载平衡器具有可从互联网访问的公共IP地址。私有负载平衡器具有来自托管子网的IP地址,该IP地址仅在VCN内可见。可以为IP地址配置多个侦听器,以对传输不同层的流量(例如,层4和层7(TCP和HTTP)流量)进行负载平衡。公共和私有负载平衡器都可以将数据流量路由到可从VCN到达的任何后端服务器。

[0069] 根据实施例,公共负载平衡器可以接受来自互联网的流量,可以创建被指派有公共地址的公共负载平衡,其用作传入流量的入口点。

[0070] 根据实施例,公共负载平衡器在范围上是区域性的。如果区域包括多个可用性域,那么公共负载平衡器可以具有例如区域性子网或两个特定于可用性域(特定于AD)的子网,每个子网位于分开的可用性域中。利用区域性子网,负载平衡器可以创建主负载平衡器和备用负载平衡器,每个负载平衡器位于不同的可用性域中,以确保即使在可用性域中断期间的可访问性。如果在多个特定于AD的子网中创建负载平衡,那么一个子网可以托管主负载平衡器,而另一个子网托管备用负载平衡器。如果主负载平衡器发生故障,那么公共IP地址可以切换到辅助负载平衡器。该服务将两个负载平衡器视为等效。

[0071] 根据实施例,如果区域仅包括一个可用性域,那么服务仅要求一个子网(或者区域性的或者特定于AD的)来托管主和备用负载平衡器两者。除了指派的浮动公共IP地址之外,主和备用负载平衡器可以各自具有来自主机子网的私有IP地址。如果存在可用性域中断,那么负载平衡器没有故障转移。

[0072] 根据实施例,还可以提供私有负载平衡,以便将负载平衡器与互联网隔离并简化安全姿态。负载平衡器服务可以为用作传入流量的入口点的负载平衡器指派私有地址。

[0073] 根据实施例,私有负载平衡器可以由服务创建以服务于仅一个子网来托管主和备用负载平衡器两者。负载平衡器可以是区域性的或特定于AD的,具体取决于主机子网的范围。负载平衡器只能从包含该主机子网的VCN内访问,或者受安全规则的进一步限制。

[0074] 根据实施例,指派的浮动私有IP地址对于主机子网是本地的。主和备用负载平衡器各自要求来自主机子网的额外私有IP地址。

[0075] 根据实施例,如果存在可用性域中断,那么多AD区域内的区域性子网中创建的

私有负载平衡器提供故障转移能力。在特定于AD的子网或单个可用性域区域内的区域性子网中创建的私有负载平衡器没有响应可用性域中断的故障转移能力。

[0076] 根据实施例,云基础设施环境可以在联网资源层处提供多个边缘服务166。一般而言,边缘服务包括允许客户端管理、保护和维护域和端点的多个服务。例如,这些包括DNS(域名系统)、DDoS(服务的分布式拒绝)保护和电子邮件递送。这些服务使客户端能够优化性能、阻止网络攻击和扩展通信。

[0077] 根据实施例,云基础设施环境可以在联网资源层处提供多个连接服务168。此类连接服务可以提供在客户端数据中心或现有网络与云基础设施环境之间创建专用的、私有连接的简单方式。连接服务可以提供高带宽,以及可靠和一致的网络。

[0078] 根据实施例,存储资源层170可以包括多个资源,诸如例如块卷172、文件存储174、对象存储176和本地存储178。根据示例实施例的云基础设施环境100支持与每个资源相关联的多个标签,包括例如一般与资源140相关联的资源标签140a,以及与块卷172、文件存储174、对象存储176和本地存储178相关联的存储资源标签170a。

[0079] 根据实施例,块卷172提供支持广泛的I/O密集型工作负载的高性能网络存储容量。客户端可以使用块卷来扩展计算实例的存储容量,提供可以跨计算实例被迁移的持久的数据存储,并托管大型数据库。

[0080] 根据实施例,文件存储174允许客户端创建可扩展的分布式企业级网络文件系统。文件存储支持语义、快照能力和静态数据加密。

[0081] 根据实施例,对象存储为非结构化数据提供高吞吐量存储。对象存储服务为大量分析数据或丰富内容(像图像和视频)提供近乎无限的存储容量。块卷可以备份到对象存储以增加持久性。

[0082] 根据实施例,本地存储178可以为I/O密集型应用提供例如固态驱动器形式的高速且可靠的存储。例如,这些可以在裸机实例中提供。本地存储为VM的和裸机计算实例提供高存储性能。一些示例包括关系数据库、数据仓库、大数据、分析、AI和HPC应用。

[0083] 图2图示了根据实施例的用于在云基础设施环境内提供云基础设施区域的系统。

[0084] 根据实施例,以上在图1中描述的云基础设施环境的实例可以托管在不同的区域中,称为云基础设施区域220。如上所述,这些可以经由控制台、SDK或API由客户端网络200通过诸如互联网210之类的网络来访问。每个云基础设施区域可以包括管理服务222、计算服务224、存储服务226、边缘服务228、网络服务230和物理基础设施232。

[0085] 根据实施例,云基础设施可以托管在区域和可用性域中。区域可以是本地化的地理地区,并且可用性域可以是位于区域内的一个或多个数据中心。区域由一个或多个可用性域组成。大多数云基础设施资源可以是特定于区域的,诸如虚拟云网络,或者是特定于可用性域的,诸如计算实例。可用性域之间和区域之间的流量被加密。

[0086] 根据实施例,可用性域彼此隔离、容错并且不太可能同时发生故障。因为可用性域不共享基础设施(诸如电源或冷却系统)或内部可用性域网络,所以区域内一个可用性域的故障不太可能影响同一区域内其它可用性域的可用性。

[0087] 根据实施例,同一区域内的可用性域可以通过低时延、高带宽网络彼此连接,该网络可以提供到互联网和本地部署的高可用性连接性,并在多个可用性域中构建复制的系统以用于高可用性和灾难恢复。

[0088] 根据实施例,区域独立于其它区域并且可以在地理上分离(例如,跨越国家或大陆)。然后,这导致将应用部署在最有可能最常使用该应用的区域内。

[0089] 但是,根据实施例,应用也可以出于各种原因部署在不同的区域中。例如,这可以包括当诸如天气系统之类的事件使某个区域离线时风险减轻。此外,出于战略原因,诸如税收域或其它商业或社会准则,应用可以部署在其它区域中。

[0090] 根据实施例,存在跨区域可用的若干服务。这些包括例如管理服务222、计算服务224、存储服务226、边缘服务228和网络服务230。

[0091] 根据实施例,隔间允许客户端组织和控制对云资源的访问。隔间是只能由被管理员给予许可的某些组访问的相关资源(诸如实例、虚拟云网络、块卷)的集合。例如,一个隔间可以包含构成公司人力资源(HR)系统的生产的所有服务器和存储卷,而作为进一步的示例,其它隔间可以各自分开地专用于公司的法律、营销、会计、运营和信息技术(IT)系统。在示例中,只有对那个隔间具有许可的用户才能管理和/或访问那些服务器和卷。在进一步的示例中,隔间可以包含构成公司人力资源(HR)系统以及法律、营销、会计、运营和信息技术(IT)系统的集合的生产的所有服务器和存储卷。在示例中,只有对那些资源的部分具有许可的用户才能管理和/或访问服务器和卷上的那些部分。

[0092] 示例实施例的隔间包括一个或多个逻辑组并且不一定是物理存储器容器,但是如果期望或必要,那么它们可以是物理存储器容器。当在控制台内工作时,隔间可以充当允许查看的内容的过滤器。隔间是可以用于云资源的主要构建块,并且它们可以被用于组织和隔离资源,以便更轻松地管理和保护对那些资源的访问。

[0093] 根据实施例,隔间可以具有若干层。例如,可以将租赁视为保持客户端的所有云资源的根隔间。可以在那个根隔间(租赁)内创建附加隔间,并且可以创建对应的策略来控制对每个隔间中资源的访问。当客户端创建云资源(诸如计算、存储、VCN、IP地址和/或DNS实例、块卷或云网络)时,可以将此类资源指向特定的一个或多个隔间。隔间可以跨越区域。

[0094] 故障域

[0095] 根据实施例,故障域可以包括可用性域内的硬件和基础设施的分组。每个可用性域可以包括三个故障域。故障域允许实例被分布,以便它们不在单个可用性域内的同一物理硬件上。影响一个故障域的硬件故障或计算硬件维护不会影响其它故障域中的实例。

[0096] 根据实施例,诸如计算、裸机DB系统或虚拟机DB系统实例之类的资源的放置可以可选地在启动时指定故障域或新实例。通过在当前故障域处终止资源并在另一个故障域处启动资源的新实例,资源可以在放置之后附加地改变故障域。

[0097] 根据实施例,可以出于多种原因使用故障域,诸如防止意外硬件故障和防止由于维护引起的计划的中断。

[0098] 可用性

[0099] 根据实施例,可以提供服务可用性。云基础设施环境内的区域可以提供核心基础设施服务和资源,包括:

[0100] • 计算:计算(裸机&虚拟机、DenseIO&标准)、用于Kubernetes的容器引擎、注册表

[0101] • 存储:块卷、文件存储、对象存储、归档存储

[0102] • 联网:虚拟云网络、负载均衡、快速连接

[0103] • 数据库:数据库、Exadata云服务、自主数据仓库、自主事务

[0104] • 处理

[0105] • 边缘:DNS

[0106] • 平台:身份和访问管理、加标签、审计

[0107] 根据实施例,以上服务和资源一般是可用的,而其它服务和资源也可以附加地可用(例如,基于区域需求或客户请求)。作为示例,新的云服务可以基于各种考虑因素尽快在地区中可用,考虑因素包括区域性客户需求、在适用的情况下实现监管合规性的能力、资源可用性和其它因素。由于低时延互连骨干网,客户可以在其它地理区域使用云服务并在其所在区域不可用时获得有效结果,前提是数据驻留要求不会阻止他们这样做。

[0108] 根据实施例,可以在全局可用性、区域性可用性、单区域可用性和域可用性的上下文中考虑资源可用性。一般而言,IAM资源是全局可用的,DC系统、实例和卷是特定于可行性域的。大多数其它资源都是区域性的。

[0109] 根据实施例,全局可用资源的示例可以包括API签署密钥、隔间、动态组、联合资源、组、策略、标签命名空间、标签密钥和用户。

[0110] 根据实施例,区域性可用资源的示例可以包括警报、应用、桶(虽然桶是区域性资源,但是当使用用于API调用的正确的特定于区域的对象存储URL时,它们可以从任何位置被访问)、集群、云事件规则、客户位置处的装备(CPE)、DHCP选项集、动态路由网关(DRG)、加密密钥、功能、映像、互联网网关、作业、密钥库、负载均衡器、本地对等网关(LPG)、度量、NAT网关、网络安全组、节点池、ons-订阅、ons-话题、储存库、保留的公共Ips、路由表、安全列表、服务网关、堆栈、子网(当创建子网时,它可以被声明为区域性的或特定于可用性域)、虚拟云网络(VCN)和卷备份(卷备份可以作为新卷被恢复到存储它们的同一区域内的任何可用性域)。

[0111] 根据实施例,特定于可用性域的资源示例可以包括DB系统、临时公共Ips、实例(实例可以附连到同一可用性域中的卷)、子网(当创建子网时,它可以被声明为区域性的或特定于可用性域)和卷(卷可以附连到同一可用性域中的实例)。

[0112] 隔间

[0113] 根据实施例,管理员可以管理云基础设施环境内的隔间。

[0114] 根据实施例,标签可以应用于隔间内的资源。例如,标签可以被用于根据模式(诸如业务需求模式)组织资源。可以在创建资源时将标签应用于资源,或者可以在现有资源上更新标签。与每个资源相关联的标签可以包括例如一般与资源140相关联的资源标签140a(图1)、与计算机资源150相关联的计算机资源标签150a、与网络资源160相关联的网络资源标签160a、以及与存储资源170相关联的存储资源标签170a。

[0115] 根据实施例,隔间对于云基础设施的构建和组织是重要的。可以在隔间之间移动资源,并且可以显示(例如,经由用户界面)在当前区域内按隔间组织的资源。当用资源工作和管理资源时,可以首先选择隔间。

[0116] 根据实施例,隔间是租赁范围的,并且可以跨越区域。当创建隔间时,可以使该隔间在租赁订阅的每个区域内可用。

[0117] 根据实施例,可以删除隔间。为了删除隔间,该隔间可以在删除之前移除其中的所有资源。

[0118] 根据实施例,删除隔间的动作可以是异步的并且发起工作请求。在执行工作请求

时,隔间的状态变为“正在删除”。如果工作请求失败,那么隔间不会被删除并且它返回到活动状态。

[0119] 根据实施例,在云基础设施环境内创建的每个隔间可以具有某些特性。例如,可以为每个隔间指派唯一标识符(ID),并且可以附加地和可选地提供有可修改的描述以及名称。根据实施例,子隔间(或子隔间)可以在基本隔间下以分层方式定义。

[0120] 根据实施例,对隔间和子隔间的访问和控制可以限于管理员或具有足够凭证的其他用户。凭证可以与不同级别的隔间访问权相关联。例如,管理员可以具有查看和访问所有隔间并使用租赁的任何隔间内的资源的许可,但具有更受限访问权的用户将没有这种级别的访问和控制。

[0121] 图3示出了根据实施例的云基础设施环境300,其图示了用于跨越云基础设施区域的策略管理和控制的隔间360、隔间策略365、子隔间370与子隔间策略375之间的关系。

[0122] 根据实施例,如上所述,上图1中所述的云基础设施环境的实例可以被托管在不同区域(诸如云基础设施区域320、330、340、350)中。如上所述,这些可以经由控制台、SDK或API由客户网络301经由诸如互联网之类的网络302来访问。

[0123] 根据实施例,客户网络301可以包括例如单个计算机、客户计算机网络或其它这样的网络。

[0124] 根据实施例,虽然图中未示出,但每个云基础设施区域可以包括多个服务,每个服务包括多个资源,诸如管理服务、计算服务、存储服务、边缘服务、网络服务和物理基础设施。

[0125] 根据实施例,云基础设施可以托管在区域和可用性域中。区域可以是本地化的地理地区,并且可用性域可以是位于区域内的一个或多个数据中心。区域由一个或多个可用性域组成。大多数云基础设施资源可以是特定于区域的,诸如虚拟云网络,或者是特定于可用性域的,诸如计算实例。可用性域之间和区域之间的流量被加密。

[0126] 根据实施例,可用性域彼此隔离、容错并且不太可能同时发生故障。因为可用性域不共享基础设施(诸如电源或冷却系统)或内部可用性域网络,所以区域内一个可用性域的故障不太可能影响同一区域内其它可用性域的可用性。

[0127] 根据实施例,同一区域内的可用性域可以通过低时延、高带宽网络彼此连接,该网络可以提供到互联网和本地部署的高可用性连接性,并在多个可用性域中构建复制的系统以用于高可用性和灾难恢复。

[0128] 根据实施例,区域独立于其它区域并且可以在地理上分离(例如,跨越国家或大陆)。然后,这导致将应用部署在最有可能最常使用该应用的区域内。

[0129] 但是,根据实施例,应用也可以出于各种原因部署在不同的区域中。例如,这可以包括当诸如天气系统之类的事件使某个区域离线时风险减轻。此外,出于战略原因,诸如税收域或其它商业或社会准则,应用可以部署在其它区域中。

[0130] 根据实施例,存在跨区域可用的若干服务。这些包括例如管理服务、计算服务、存储服务、边缘服务和网络服务。

[0131] 根据实施例,隔间允许客户端组织和控制对云资源的访问。隔间是只能由被管理员给予许可的某些组访问的相关资源(诸如实例、虚拟云网络、块卷)的集合。隔间可以被认为逻辑组,而不是物理容器。当在控制台内工作时,隔间可以充当允许查看的内容的过滤

器。

[0132] 根据实施例,隔间可以具有若干层。例如,可以将租赁305视为保持客户端的所有云资源的根隔间。隔间可以以分层方式组织,诸如隔间360是租赁隔间下的一级,而子隔间370是隔间360下的附加层。根据实施例,每个隔间可以与一个或多个隔间策略(诸如隔间策略365和子隔间策略375)相关联。图中未示出租户隔间策略。

[0133] 根据实施例,在创建隔间或子隔间(诸如隔间360和子隔间370)期间、当时或之后,可以为每个隔间和子隔间编写/创建诸如隔间策略365和子隔间策略375之类的策略。如果没有适当的策略,对隔间和/或子隔间的访问可以被限制到具有租赁305级别的许可的用户。

[0134] 根据实施例,当在隔间内创建隔间(即,子隔间)时,子隔间从其层级更高的隔间继承访问许可。

[0135] 根据实施例,在创建隔间或子隔间策略后,该策略可以包括指示该策略附连到哪个隔间的规范。这种规范可以包含限制对策略的子序列控制、修改或删除的访问的控制。在一些实施例中,可以将策略附连到租赁、父隔间或策略所针对的特定隔间。

[0136] 根据实施例,可以将新资源放置到隔间中。这可以通过在创建新资源后指定作为目标的隔间来完成(隔间是创建资源所需的多条信息之一)。这可以经由控制台接口来完成。

[0137] 根据实施例,现有资源也可以移动到不同的隔间。大多数资源在创建之后可以被移动。有一些资源不能从一个隔间移动到另一个隔间。

[0138] 根据实施例,一些资源具有附连的资源依赖性,而一些则没有。当父资源移动时,并非所有附连的依赖项(dependency)的行为方式都相同。

[0139] 根据实施例,对于一些资源,附连的依赖项与父资源一起移动到新隔间。父资源立即移动,但在一些情况下,附连的依赖项异步移动,并且在移动完成之前在新隔间中不可见。

[0140] 根据实施例,对于其它资源,附连的资源依赖项不移动到新的隔间。此类附连的资源可以被独立移动。

[0141] 根据实施例,在资源被移动到新隔间之后,治理新隔间的策略立即应用并影响对资源的访问。取决于隔间组织的结构,计量、计费和警报也会受到影响。

[0142] 根据实施例,在创建之后,可以将隔间移动到例如同一个租赁内的不同父隔间。在移动隔间后,该隔间的所有内容(包括子隔间和资源)都会随隔间一起移动。

[0143] 图4图示了根据实施例的用于云基础设施环境内的隔间迁移的系统。

[0144] 根据实施例,如上所述,上图1中描述的云基础设施环境400的实例可以托管在不同区域中。可以在云基础设施环境内定义诸如租赁405、隔间A 460和隔间B 470之类的隔间,并且这些隔间可以跨越区域。如上所述,此类隔间可以经由控制台、SDK或API由客户端网络401经由诸如互联网之类的网络402来访问。

[0145] 根据实施例,客户端网络401可以包括例如单个计算机、客户计算机网络或其它此类网络。

[0146] 根据实施例,隔间允许客户端组织和控制对云资源的访问。隔间是只能由管理员给予许可的某些组访问的相关资源(诸如实例、虚拟云网络、块卷)的集合。隔间可以被认为

是逻辑组,而不是物理容器。当在控制台内工作时,隔间可以充当允许查看的内容的过滤器。

[0147] 根据实施例,隔间可以具有若干层。例如,可以将租赁405视为保持客户端的所有云资源的根隔间。隔间可以按分层方式组织,诸如隔间A 460和隔间B 470是租赁隔间下的一级,而子隔间A 465定义在隔间A下,子隔间B 475定义在隔间B下。根据实施例,每个隔间可以与一个或多个隔间策略(未示出)相关联。

[0148] 根据实施例,例如,在租赁内定义的隔间可以通过例如重新定义隔间或子隔间而被移动。

[0149] 根据实施例,为了移动隔间,可以接收具有足够许可的请求。即,来自属于例如对最低共享父隔间的“管理所有资源”许可的组的用户对当前隔间和移动隔间的目的地隔间的请求。

[0150] 即,例如,必须从具有足够许可的用户接收将子隔间A 465(从465移动到476)从隔间A 460移动到隔间B 470的请求。因为租赁405是源隔间(隔间A 460)和目的地隔间(隔间B 470)的最低共享父隔间,所以如图所示移动子隔间A的请求必须从在租赁405隔间内具有“管理所有资源”许可的用户接收。

[0151] 根据实施例,在另一个示例中,如果从仅在隔间A内具有“管理所有资源”许可的用户接收到将子隔间A 465从隔间A移动到隔间B的请求,那么该请求可能会失败,因为来自用户的请求无法管理目的地隔间(即,隔间B)内的资源。

[0152] 根据实施例,在将隔间移动到新的父隔间后,新的父隔间的访问策略生效并且先前的父隔间的策略不再适用。在一些情况下,当用指定层次结构的策略来移动嵌套的隔间时,可以自动更新策略以确保一致性。

[0153] 因此,根据实施例,先前应用于子隔间A的隔间A 460的隔间策略将不再适用于子隔间A向隔间B的迁移。然后,隔间B的隔间策略将改为适用于子隔间A。这将在下图的描述中得到更多解释。

[0154] 图5示出了在云基础设施环境内的隔间迁移期间用于策略管理和强制执行的系统。

[0155] 根据实施例,并且更具体而言,图5示出了在其中移动隔间的隔间层次结构,以及用于不同策略的结果。

[0156] 根据实施例,如上所述,上图1中描述的云基础设施环境500的实例可以托管在不同区域中。可以在云基础设施环境内定义诸如租赁505、隔间A 560和隔间B 565以及隔间D 566之类的隔间,并且这些隔间可以跨越区域。如上所述,此类隔间可以经由控制台、SDK或API由客户网络501经由诸如互联网之类的网络502来访问。

[0157] 根据实施例,客户网络501可以包括例如单个计算机、客户计算机网络或其它此类网络。

[0158] 根据实施例,隔间可以具有若干层。例如,可以将租赁505视为根隔间,其保持客户端的所有云资源。隔间可以以分层方式组织,诸如隔间A 560是租赁下的一级。隔间B 565和隔间D 566然后被组织为隔间A 560下的又另一级,而子隔间C 570被示为最初位于隔间B下的一级。根据实施例,每个隔间可以与一个或多个隔间策略(诸如隔间B策略582、隔间A策略580和隔间D策略583)相关联。例如,此类策略可以治理对隔间的访问的用户/客户端许可,

以及对任何给定隔间内的资源的访问和控制的许可。如上所述,隔间策略可以添加到彼此(即,“堆栈”),使得访问隔间B 565的用户将使其与隔间B 565的交互除了隔间A策略580之外还受隔间B策略582治理/限制。

[0159] 根据实施例,例如,假定隔间B策略582允许组(组1)管理隔间A-B中的实例系列(隔间层次结构包括隔间B是隔间A的子隔间)。

[0160] 根据实施例,还假定隔间D策略583允许另一个组(组2)管理隔间A-D中的实例系列(隔间层次结构包括隔间D是隔间A的子隔间)。

[0161] 根据实施例,在隔间C从隔间B移动到隔间D后,组1的成员不再能够管理隔间C中的实例系列,而组2的成员现在可以管理隔间C中的实例系列。

[0162] 根据实施例,在某些情况下,在移动隔间后,可以自动更新某些策略。例如,当策略附连到当前和目标父级的共享祖先时,可以自动更新将隔间层次结构向下指定到要移动的隔间的策略。

[0163] 返回去参考图5,例如,根据实施例,假定隔间A策略允许组(组X)的成员管理隔间B:C中的桶。在将隔间C移动到隔间D时,由于隔间B和D之间的共享祖先(隔间A),因此可以自动更新隔间A策略以允许组X的成员管理隔间D:C中的桶。

[0164] 根据实施例,附连到租赁的策略同样可以在隔间在租赁内移动后自动更新。

[0165] 但是,根据实施例,并非所有策略都在隔间移动后自动更新。例如,参考图5,在隔间C从隔间B移动到隔间D的情况下。假定隔间B策略允许管理隔间C中的桶(在移动之前)。当移动隔间C时,隔间B策略不会自动更新。代替地,该策略不再有效并且可以被移除(例如,手动或自动)。

[0166] 基于标签的资源限制/配额

[0167] 根据实施例,云管理员一般不具有限制现有云中的资源使用的能力。授予用户创建资源的许可允许他们创建任何数量的资源,最多达到预定义的账户限制。基于标签的资源限制或配额允许将在隔间内创建或使用资源的能力限制到适当的级别,从而允许微调成本控制。

[0168] 根据实施例,可以在账户创建时向客户指派由云基础设施环境定义的服务级别限制。这些服务级别限制限制了客户可以跨整个租赁(例如,跨具有多个隔间的多个区域)创建的资源的总数。租赁和隔间管理员可以利用基于标签的资源限制或配额来设置特定于资源的硬限制。如果对单个资源没有此类隔间限制,那么被授权启动实例的用户可以消耗整个租赁中所有可用容量。基于标签的资源限制或配额解决了这个问题,并且与服务限制不同,它由客户端和客户经由例如控制台、SDK或API来设置和定制。基于标签的资源限制或配额可以应用在服务限制之上,并通过嵌套的隔间层次结构继承。这允许云管理员限制资源消耗并围绕可接受的资源使用来设置边界。

[0169] 根据实施例,基于标签的资源限制或配额使租赁和隔间管理员能够更好地控制在云基础设施环境中如何消耗资源,从而使管理员能够借助于例如控制台、SDK或API轻松地将资源分配给用户或用户组。隔间配额是管理租赁内客户端支出的强大工具集。

[0170] 根据实施例,当客户端在其租赁中具有跨多个隔间的资源(例如,实例、VCN、负载均衡器和块卷)时,可能难以跟踪用于特定目的的资源,或汇总它们、报告它们或对它们执行批量动作。加标签允许客户端定义键和值并将它们与资源相关联。标签可以被用于帮助

组织和列出资源。一般而言,存在两种类型的标签:自由形式的标签和定义的标签,但是根据示例实施例,其它类型的标签也是可能的。

[0171] 根据实施例,自由形式的标签可以由键和价值组成。例如,“环境:生产”是自由形式的标签的示例,其中“环境”是键,“生产”是值。多个自由形式的标签可以应用于单个资源。

[0172] 根据实施例,定义的标签比自由形式的标签提供更多的特征和控制。在客户端创建定义的标签键之前,可以为定义的标签设置标签命名空间。命名空间可以被认为是标签键的集合的容器。定义的标签支持策略,以允许控制谁可以应用你定义的标签。标签命名空间是客户端可以对其应用策略的实体。

[0173] 根据实施例,为了将定义的标签应用于资源,用户可以首先选择标签命名空间,然后选择命名空间内的标签键,然后用户可以指派值。管理员可以控制允许哪些用户组使用每个命名空间。根据示例实施例,标签可以在它们被创建时应用于资源。在另一个实施例中,标签可以在它们被供应之后应用于资源,从而允许在使用标签来追溯系统中的资源配额或限制的强制执行。根据实施例,标签可以被用于启用成本治理和/或资源治理。以那种方式,标签可以被用于跟踪成本。

[0174] 图6示出了根据实施例的两个定义的标签。设置两个标签命名空间,命名空间1 600和命名空间2 620(例如,命名空间1可以是“Operations”(“操作”),命名空间2可以是“HumanResources”)。标签键605、610、625和630在命名空间中定义。在每个命名空间内,标签键可以是唯一的,但标签键名称可以跨命名空间被重复。例如,两个命名空间都可以包含名为“Environment”(“环境”)的键。

[0175] 根据实施例,许多云提供商提供服务限制以防止容量短缺。但是,很多时候,客户端也希望设置限制,以免他们自己陷入资源超额(overage)的境地。一种这样的方法是在组级别指定限制。组可以包含用户、资源或用户想要指定的任何内容的集合。组级别限制由特定租赁/账户的管理员/用户设置。

[0176] 根据实施例,作为示例,租赁管理员或组管理员可以在组级别配置限制以控制他们的云成本。这向客户/公司保证,特定组不会超过由租赁或账户管理员设置的使用限制,因为云资源会发出使用数据,这些数据会转换成与在那个组内创建的(一个或多个)资源相关联的美元成本。在这个意义上,组可以被认为是成本中心组,因为由那个组的成员创建的(一个或多个)资源的所有云成本都被分配给客户/公司作为单个汇总的一次性(lump)费用或针对单个组账户被汇总等。

[0177] 但是,根据实施例,存在问题,因为存在用于让管理员控制成本的非常简单的机制。作为示例,管理员可以在资源超额中包括或排除特定组,或者他们可以通过拒绝访问来排除特定用户创建资源。但是,这些是粗粒度机制,并且客户一般没有指定为他们提供灵活性和更好的控制来管理他们的云资源的细粒度机制的能力。

[0178] 根据实施例,细粒度方法可以提供基于标签的资源限制,其中可以由特定组的一个或多个成员供应的所有资源与代表该组的成本中心标签相关联或以其它方式为其提供该成本中心标签。这种性质的成本中心标签提供了用于资源治理的机制,并且云提供商也将它们用于成本治理。系统和方法可以创建一种机制来通过标签在组级别控制成本。作为示例,用户可以指定细粒度的规则/策略,诸如:

[0179] Set limits<resource type>to 10in group A where target.resource.tag=

“finance”

[0180] 根据实施例,标签可以与资源相关联,并且通过以上方法,标签可以在组级别保护客户免受资源超额的影响。在以上示例中,当标签值为组A中的“finance”时,用户成员将无法创建超过10个<resource type>实例。这种在组级别限制成本的方法为客户端提供了细粒度的机制,以更好地管理成本和资源使用。

[0181] 根据实施例,一种系统在云基础设施环境中提供基于资源请求上下文标签的限制/配额。该系统执行一种用于基于请求上下文在云基础设施环境中对供应资源进行限制或强加配额的方法。

[0182] 根据实施例,一种系统在云基础设施环境中提供基于标签的资源限制/配额。该系统执行一种用于基于资源标签在云基础设施环境中对供应资源进行限制或强加配额的方法。

[0183] 根据实施例,在高度安全的环境中,客户希望保护对资源及其数据的访问。作为示例,客户通过将资源分类为高度机密、机密、受限、公共等来保护他们的资源,并基于按每个约束他们的准许(clearance)授予用户访问权。有权访问分类的资源/数据的用户在访问数据时需要非常小心,以确保没有信息泄漏,并避免在共享上下文中跨数据分类对资源进行无意修改。本文的实施例提供了改进的系统和方法,用于,即使用户可能具有更高的访问权,也使用户能够在给定会话/上下文的数据分类边界内限制访问或修改资源。

[0184] 根据实施例,云体系架构可以使用基于角色的访问控制技术通过在角色内安排(slot)用户来限制/允许对特定安全边界的访问。但是,这些技术是粗粒度的。责任在于用户对它们正在访问、修改或执行的内容要勤奋和谨慎。

[0185] 根据实施例,当前公开的系统和方法保护用户在他们不打算时不访问资源以执行某些操作。一个示例是当用户可以访问高度分类的资源时,访问数据的其它分类。例如,假定用户具有只需要通过过滤掉高度分类的资源来访问未分类的资源的脚本。在此,用户/脚本必须知道哪些数据/资源是分类的,哪些不是。

[0186] 根据实施例,另一个示例是当用户仅对业务关闭之后关掉非业务关键的基础设施感兴趣,但想要继续运行业务关键的其它资源时。在此,用户必须确切知道要关闭哪些实例。

[0187] 根据实施例,本文的系统和方法经由附连到对资源的一个或多个请求的标签来提供对此类问题的解决方案。来自脚本的请求代表用户执行操作,例如关掉实例、读取资源的细节、更新资源、访问分类的资源等。用户发送这些请求,请求具有附连到请求上下文的一个或多个标签。当这些请求对资源执行动作时,可以检查来自请求上下文的这些标签,然后可以或者拒绝或者授予访问权以执行用户想要执行的操作,即使最初用户能够访问以执行那个操作。这通过执行用户不打算针对特定(一个或多个)资源的操作,来避免无意访问高度分类的数据或资源。这可以属于基于标签的自动化。它可以满足各种各样的用例,而不仅仅是与安全相关的领域。

[0188] 根据实施例,当通过发送请求来访问资源时,代表资源请求上下文的标签可以被用于经由标签创建安全边界。

[0189] 根据实施例,当用户想要访问资源以执行某个操作时,用户一般可以通过软件开发工具包进行调用或者将直接进行API调用。用户可以传递一个或多个标签,这些标签被添

加到发送到资源的请求。在授权用户时,资源实施方式可以检查与请求相关联的标签,如果标签与用户标签匹配,那么请求被允许,否则请求被拒绝/否认。

[0190] 图7示出了根据示例实施例的用于让用户在系统中启动实例的体系架构流程,该实例在基于标签(包括例如资源标签和/或资源请求上下文标签)在云基础设施环境中对资源强制执行配额或限制。

[0191] 根据实施例,在步骤1处,用户701请求启动区域700中的实例,其中该请求与在区域700内定义的一个或多个标签(例如,自由形式的标签或定义的标签)相关联。在示例实施例中,标签代表来自用户的对资源的请求的上下文。

[0192] 根据实施例,在步骤2处,负载均衡器720将请求转发到API,诸如公共代理API 725。

[0193] 根据实施例,在步骤3处,API在身份数据平面735处执行认证。

[0194] 根据实施例,在步骤4处,启动实例的请求可以被转发到计算控制平面(CCP) 730。然后,CCP可以在步骤5处在身份数据平面处附加地认证请求。

[0195] 根据实施例,在步骤6处,计算控制平面可以从限制服务数据平面740提取(可替代地,可以在资源控制平面处执行该提取操作)与作为来自用户701的请求的一部分的标签相关联的所有基于标签的限制/配额750。例如,此类基于标签的配额可以跨整个租赁应用,并且不必特定于隔间。

[0196] 根据实施例,在步骤7处,计算控制平面可以进行检查以查看所请求的使用是否会违反任何750中的任何基于标签的限制/配额。如果所请求的使用没有违反任何隔间配额,那么可以将请求处理到例如数据库745。如果请求的使用将违反沿着隔间树的任何基于标签的限制/配额,那么可以丢弃请求并且可以发送消息以通知用户。

[0197] 根据实施例,基于标签的限制/配额跨例如整个租赁被应用,包括在其中启用租赁的所有区域。租赁内的隔间。当确定所请求的实例是否会违反任何基于标签的限制/配额时,CCP可以确保新请求(例如,经由服务开发人员工具包(SDK))不违反任何配额。如果请求超出配额,那么CCP可以使请求失败。

[0198] 根据实施例,CCP的SDK可以存在于云基础设施环境的每个控制平面中。

[0199] 根据实施例,CCP可以与限制服务数据平面一起工作以针对基于标签的限制/配额检查新请求。

[0200] 根据实施例,当客户端触发创建云资源的操作时,请求转到资源控制平面。资源控制平面向限制服务询问是否可以考虑在组级别配置的限制的情况下安全地启动资源。如果限制服务器向资源控制平面返回否定/失败响应,那么不创建资源。

[0201] 在内部,当限制服务接收到请求时,它检查与资源创建相关联的标签,并在存储库或数据库或存储器中查找以查看具有那个特定标签的资源的当前使用情况。如果标签值超过了在组级别配置的限制,那么它向资源控制平面发送失败/否定响应,从而导致资源创建失败。

[0202] 为资源限制/配额使用标签

[0203] 图8示出了使用请求上下文标签和/或资源标签来在云基础设施环境中限制诸如资源的供应之类的使用的系统。

[0204] 根据实施例,客户端设备810可以向云基础设施环境提交请求811,其中该请求与

标签812 (诸如例如代表请求的上下文的标签) 相关联。客户端设备可以经由诸如控制台 802、API 804或SDK 806之类的接口之一向云基础设施环境800 (在设备硬件801上运行) 提交请求。

[0205] 根据实施例,可以将请求转发到资源840,资源840可以包括多个层 (如上所述), 诸如包括加标签的资源851-854的计算资源层850、包括加标签的资源861-864的网络资源层860,以及包括加标签的资源871-874的存储资源层870。

[0206] 根据实施例,在客户端设备享有高安全性/优先级特权 (例如,欺诈保护或智能) 的情况下,客户端设备可以针对请求利用宽范围的标签。作为示例,与请求相关联的标签可以被用于帮助用户避免误触受保护的资源。例如,假定每天下午6点之后可以关掉某些计算资源以节省成本。此类计算资源具有低安全标签。但是,同一系统还有其它无法关掉的计算资源 (必须24/7运行)。此类计算资源具有高度安全的标签。为了关掉低特权的计算资源,用户在没有加标签的请求的情况下将必须去每个资源并只选择要关掉的那些资源,或者编写脚本来这样做,以避免关掉需要运行24/7的高特权资源。这是难以维护的,因为每天都可以添加和移除计算资源。

[0207] 根据实施例,然后,用户可以代替地提供加标签的请求以关掉低特权资源。即,与请求相关联的标签将与可以在下午6点关掉的计算资源共享低安全标签。然后,加标签的可以去往网络中的所有资源,但它会绕过没有请求所具有的标签的高度保护的资源。以这种方式,用于关掉计算资源的加标签的请求仅关掉低安全标签匹配的那些计算资源,并且请求绕过了高度安全的加标签的计算资源。

[0208] 根据实施例,作为另一个示例,假定客户端810想要将高度安全的计算实例附连到块卷。在这种情况下,附连高度安全的计算实例的请求可以用高度安全的标签来加标签。然后,通过这样做,这可以确保高度安全的计算实例不会附连到具有较低安全标签的块卷,而是代替地可以请求只能将高度安全的计算实例附连到类似加标签的 (高度安全的) 块卷。换句话说,请求不会将高度安全的计算附连到低安全块卷 (标签检查将失败)。

[0209] 基于请求上下文标签的限制/配额

[0210] 图9是根据实施例的在云基础设施环境中提供基于资源请求标签的限制/配额的系统900的功能示意图。一般而言,客户端设备810 (图8) 可以向云基础设施环境提交请求911,以供应多个加标签的资源940当中的新资源。多个加标签的资源940可以是例如计算资源层850的加标签的计算资源851-854、网络资源层860的加标签的网络资源861-864和/或存储资源层870的加标签的存储资源871-874中的任何一个。

[0211] 根据示例实施例,资源是在云基础设施环境中供应的项,可以是各种类型,其可以包括例如计算资源类型、存储服务资源类型、VCN服务类型、IP地址服务类型、DNS服务类型、DDoS服务保护类型和电子邮件递送服务类型,等等。本文的示例实施例不限于这些特定的资源类型,并且也可以包括其它类型,其中进一步的资源类型示例在下面通过示例的方式来原因。

[0212] 示例资源类型

[0213] 计算

[0214] 存储

[0215] VCN

[0216] IP地址

[0217] DNS

[0218] 根据示例实施例,供应的资源和服务是可加标签的。在示例实施例中,对资源的请求和资源与一个或多个标签相关联。此外,可以基于资源的标签值对资源进行分组。本质上,在示例实施例中,与每个供应的资源相关联的资源标签是键值对,其可以附连到云基础设施环境中供应的资源或以其它方式与供应的资源相关联。标签键值对可以被用于多种目的,包括例如各种核算目的。在示例中,与供应的资源相关联的资源键值对标签被用于对资源进行分组,其中由资源标签提供的分组启用云基础设施环境中所有有价值的东西的成本治理和/或资源治理。

[0219] 在云基础设施环境中供应的资源可以基于指派给每个资源的标签值或基于可能必要或期望的其它准则被分组在容器942、944、946中。作为示例,第一容器942可以存储已经在云中供应并与客户端用户701(图7)的核算组相关联的计算和存储资源,第二容器944可以保持已经在云中供应并与客户端用户701的操作组相关联的计算资源,并且第三容器946可以保持已经在云中供应并与客户端用户701的人力资源组相关联的计算资源。本文的示例实施例不限于这些组并且也可以使用其它组,其中进一步的示例组在下面通过示例的方式来原因。

[0220] 示例资源组

[0221] 操作

[0222] 人力资源

[0223] 财务

[0224] 核算

[0225] 法律

[0226] 市场营销

[0227] 根据示例,资源实例被逻辑上布置到不同的容器942、944、946中,从而提供分开的成本中心,并且每个成本中心被指派标签值,其中可以向最终用户收费以访问各种成本中心,其中成本中心的标签被用于成本治理和核算。一些客户可以在其最终组价格内对多个成本中心进行建模。对于每个成本中心可以有分开的标签,并且云可以针对每个标签指派有限的资源集。

[0228] 在示例实施例中,一个或多个策略920为最终用户提供一种机制,以保护他们自己免于访问比标准访问合约所允许的资源更多的资源,其中最终用户在访问标准合约的标准合约条款下允许的更多资源会发生严重超额。例如,客户可以发送对资源的请求,该请求经由资源控制平面930激活多个计算资源(诸如例如100个计算资源)的供应,其中合约条款仅允许以标准费率供应90个计算机资源,并且其中任何超额(在示例中为10)都会给客户最终用户带来附加的费用。

[0229] 如上所述并且根据实施例,细粒度的方法可以基于标签提供资源限制。标签是主要用于资源治理的机制,并且云提供商也将其用于成本治理。系统和方法可以创建通过标签在组级别控制成本的机制。作为示例,用户可以指定细粒度的规则/策略,诸如:

[0230] Set limits<resource type>to 10in group A where target.resource.tag="finance"

[0231] 根据实施例,标签可以与资源和对资源的请求或两者相关联,并且利用以上方法,标签可以在一个或多个所选择级别(诸如例如在组级别)保护客户免于资源超额。在以上示例中,当组A中的标签值为“finance”时,用户将无法创建超过10个<resource type>的实例。这种在组级别限制成本的方法向客户端提供细粒度的机制以更好地管理成本和资源使用。

[0232] 根据示例实施例,第一策略922在组级别上保护客户免于分组在第一容器942中的资源的资源超额。类似地,第二策略924在组级别上保护客户免于分组在第二容器944中的资源的资源超额,并且第三策略926在组级别上保护客户免于分组在第三容器946中的资源的资源超额。

[0233] 根据示例,资源请求911a-911d中的每一个包括具有代表被请求的资源实例的类型(Resource Type)的数据、代表被请求的资源实例的组(Resource Group)的数据以及代表资源请求的上下文(Context Tag)的数据的字段。在这点上,资源请求911a-911d中的每一个包括资源类型字段912a-912d,其具有代表被请求的资源实例的类型(Resource Type)的数据。此外,资源请求911a-911d中的每一个包括资源组字段913a-913d,其代表被请求的资源实例的组(Resource Group)。此外,资源请求911a-911d中的每一个包括请求上下文字段914a-914d,其包括代表资源请求的上下文(Context Tag)的请求上下文数据。

[0234] 根据示例,计算控制平面可以从限制服务数据平面740(图7)提取(可替代地,可以在资源控制平面930处执行提取操作)与作为来自用户701的请求的一部分的标签相关联的所有基于标签的限制/配额950。此类基于标签的配额可以跨例如整个租赁而应用,并且不必特定于隔间。

[0235] 根据实施例,计算控制平面可以进行检查以查看所请求的使用是否会违反存储器950中的任何配额或限制规则的任何基于标签的限制/配额。如果所请求的使用没有违反任何隔间配额,那么可以将请求处理到例如数据库。如果所请求的使用将违反沿着隔间树的任何基于标签的限制/配额,那么可以丢弃请求并且可以发送消息以通知用户。

[0236] 根据示例实施例,基于标签的限制/配额被应用于租赁内的隔间。根据下面将要描述的示例实施例,基于标签的限制/配额跨例如整个租赁(包括在其中启用租赁的所有区域,基本上跨越多个隔间)被应用。当确定所请求的实例是否会违反任何基于标签的限制/配额时,计算控制平面可以确保新请求(例如,经由服务开发人员工具包(SDK))不违反任何配额。如果请求超过配额,那么计算控制平面可以使请求失败。

[0237] 根据实施例,计算控制平面的SDK可以存在于云基础设施环境的每个控制平面中。根据实施例,计算控制平面可以与限制服务数据平面结合工作以对照基于标签的限制/配额来检查新请求。

[0238] 根据实施例,标签可以与资源和对资源的请求或两者相关联,并且利用这种方法,标签可以保护资源免于无意访问并且还可以保护客户用户免于无意的资源超额。

[0239] 在这方面,使用请求上下文标签914a-914d来提供系统900,用于控制相关联的云基础设施环境中的资源940的处置。该系统可以包括一个或多个计算机,该计算机包括一个或多个微处理器,该微处理器定义相关联的云基础设施环境中的租赁。在示例实施例中,与计算机可操作地耦合的存储器设备940存储可由计算机执行的逻辑,用于提供对租赁中资源的处置的控制。存储器设备还可以存储代表用于允许处置租赁中的资源的多个所需的凭

证门级别的访问控制数据950。可以接收处置租赁中的第一资源的请求911a-911d,其中请求911a-911d包括代表请求的请求上下文信息的请求上下文标签数据914a-914d。确定与所请求的第一资源相关联的第一特权级别分类,并且将该请求的请求上下文信息与多个所需凭证门级别中的用于允许处置具有第一特权级别分类的租赁中的资源的第一所需凭证门级别进行比较。根据示例实施例,处置第一资源的请求基于请求上下文信息与第一所需凭证门级别匹配而被选择性地授权。

[0240] 处置第一资源的请求可以包括在租赁中供应第一资源的请求。根据示例实施例,基于请求上下文信息匹配第一所需凭证门级别,在租赁中选择性地供应第一资源。

[0241] 根据示例实施例,在租赁中供应第一资源的请求可以包括来自系统用户的在租赁中供应裸机计算实例请求,以向相关联的云基础设施环境中的计算资源层内的一个或多个物理主机的用户提供控制。在这个示例中,该请求的请求上下文标签数据包括代表用户的用户标识信息的用户上下文标签数据913a-913d。存储在存储器设备中的访问控制数据包括代表用于允许处置租赁中的资源的多个用户所需凭证门级别的用户访问控制数据,其中用户的用户标识信息与多个用户所需凭证门级别中的用于允许处置具有第一特权分类的租赁中的资源的第一用户所需凭证门级别进行比较。在这个示例中,基于用户的用户标识信息与第一用户所需凭证门级别匹配,将裸机计算实例选择性地供应给租赁中的用户。

[0242] 根据示例实施例,在租赁中供应第一资源的请求可以包括在租赁中供应具有第一特权级别分类的多个资源的请求。在这种情况下,基于请求上下文信息与第一所需凭证门级别匹配而选择性地供应租赁中的多个资源。

[0243] 根据示例实施例,请求的请求上下文标签数据可以包括代表用户的用户标识信息的用户上下文标签数据913a-913d,并且处置第一资源的请求可以包括处置租赁中供应并与第二特权级别分类相关联的所有资源的请求。系统900进行操作以将该请求的用户标识信息与多个所需凭证门级别中的用于允许处置具有第二特权级别分类的租赁中的资源的第二所需凭证门级别进行比较。系统900基于用户标识信息与第二所需凭证门级别匹配而选择性地授权处置与第二特权级别分类相关联的在租赁中供应的所有资源的请求。

[0244] 根据示例实施例,租赁可以包括存储与第二特权级别分类相关联的资源的多个隔间942、944、946,其中多个隔间中的每个隔间提供该隔间内的与第二特权级别分类相关联的资源集相对于其它隔间中的与第二特权级别分类相关联的资源的一个或多个其它集合的隔离。系统900基于用户标识信息与第二所需凭证门级别匹配并跨越多个隔间而选择性地授权处置与第二特权级别分类相关联的在租赁中供应的所有资源的请求。

[0245] 由系统接收的用于处置租赁中的第一资源的请求的请求上下文数据可以包括以下中的一个或多个:代表正被请求的资源实例的类型的资源类型数据912a-912d,和/或代表正被请求的资源实例的组的资源组数据913a-913d。

[0246] 根据实施例,当客户端触发创建云资源的操作时,请求转到资源控制平面。资源控制平面向限制服务询问是否可以在考虑在组级别配置的限制的情况下安全地启动资源。如果限制服务器向资源控制平面返回否定/失败响应,那么不创建资源。

[0247] 在内部,当限制服务接收到请求时,它检查与资源创建相关联的标签,并在存储库或数据库或存储器中查找以查看具有那个特定标签的资源的当前使用情况。如果标签值超过了在组级别配置的限制,那么它向资源控制平面发送失败/否定响应,使得资源创建失

败。

[0248] 仅出于描述示例实施例的功能的目的而不是出于限制其的目的,对于分组在第一容器942中的资源,第一策略922可以根据诸如以下的细粒度规则/策略来在财务 (Finance) 组级别保护客户免于资源超额:

[0249] `Set limits<Compute>to 10in Container#1 (642)where target.resource.tag = "Finance"`

[0250] 在以上特定示例中,用户将不能在财务容器942中创建超过10个计算资源的实例。以这种方式,用于在容器942中供应资源的请求911d将由策略922处理以确定该请求是否可以被执行或以其它方式执行而不会引起任何超额。

[0251] 类似地,对于分组在第二容器644中的资源,第二策略924可以根据诸如以下的细粒度规则/策略在操作组级别保护客户免于资源超额,诸如:

[0252] `Set limits<db storage>to 100in Container#2 (644)where target.resource.tag = "Operations"`

[0253] 在以上示例中,用户将不能在操作容器644中创建超过100个存储资源的实例。以这种方式,用于在容器944中供应资源的请求911c将由策略924处理以确定请求是否可以被执行或以其它方式被执行而不会引起任何超额。

[0254] 而且类似地,对于分组在第三容器946中的资源,第三策略926可以根据诸如以下的细粒度规则/策略在人力资源组级别保护客户免于资源超额,诸如:

[0255] `Set limits<Compute>to 10and<db storage>to 100in Container#3 (646)where target.resource.tag = "Human Resources"`

[0256] 在以上示例中,用户将不能在人力资源容器946中创建超过10个计算资源实例或超过100个存储资源实例。以这种方式,用于在容器946中供应资源的请求911a、911b将由策略926处理以确定请求是否可以被执行或以其它方式执行而不会引起任何超额。

[0257] 图10是示出根据示例实施例的用于基于请求上下文在云基础设施环境中对处置资源 (诸如例如对供应资源) 进行限制或强加配额的方法1000的流程图。现在参考该图,在步骤1010中接收在云基础设施环境中供应资源的请求。该请求例如可以是上述的第一请求911a。该请求可以包括例如代表请求的请求上下文信息的请求上下文标签数据914a。

[0258] 在步骤1020处检查请求911a并确定对所请求资源的请求的上下文。所确定的资源类型可以是例如供应计算资源类型的请求或供应存储资源类型的请求。

[0259] 在步骤1030处确定与所请求资源相关联的特权级别分类。根据示例实施例,资源可以与反映访问该资源所需的特权级别的标记相关联地存储。特权级别可以存储在一个或多个特权级别标签中或存储为一个或多个特权级别标签,诸如例如图8中在851-854、861-864、871-874和图1中在140a、16a、170a处所示。例如,所确定的资源组可以是例如对供应被分组在人力资源组中的资源的请求。

[0260] 根据所确定的请求的组,可以在步骤1040处进行资源请求911a到特定容器的指派。在该示例中,根据该示例,资源请求911a可以例如指派给第三容器946。与所指派的隔间对应的策略可以选择性地应用于资源请求。在所示示例中,第三策略926可以选择性地应用于资源请求911a。如果在步骤1040中确定根据资源请求911a的内容供应附加资源不会超过被指派给隔间946的策略926中指定的限制和/或配额,那么在请求911a中请求的实例可以

在隔间946中创建。

[0261] 在步骤1050中,将请求911a的请求上下文信息914a与多个所需凭证门级别中的用于允许处置具有第一特权级别分类的租赁中的资源的第一所需凭证门级别进行比较。

[0262] 如果在步骤1060中确定请求上下文信息与第一所需凭证门级别匹配,那么在步骤1070中授权处置第一资源的请求,并且可以处置该资源,诸如例如可以创建所请求资源的实例,可以供应所请求的资源,可以终止或以其它方式取消供应所请求的资源等等。

[0263] 如果在步骤1060中确定请求上下文信息与第一所需凭证门级别不匹配,那么在步骤1080中丢弃处置第一资源的请求。

[0264] 基于资源标签的限制/配额

[0265] 图11是根据实施例的在云基础设施环境中提供基于标签的资源限制/配额的系统1100的功能示意图。

[0266] 一般而言,客户端设备810(图8)可以向云基础设施环境提交请求1111,以在多个加标签的资源1040当中供应新资源。多个加标签的资源1140可以是例如计算资源层850的加标签的计算资源851-854、网络资源层860的加标签的网络资源861-864和/或存储资源层870的加标签的存储资源871-874中的任何一个。

[0267] 根据示例实施例,供应的资源和服务是可加标签的。在示例实施例中,资源与一个或多个标签相关联。此外,可以基于其标签值对资源进行分组。本质上,在示例实施例中,与每个供应的资源相关联的资源标签是键值对,其可以附连到云基础设施环境中供应的资源或以其它方式与供应的资源相关联。标签键值对可以用于多种目的,包括例如各种核算目的。在示例中,与供应的资源相关联的资源键值对标签被用于对资源进行分组,其中由资源标签提供的分组启用云基础设施环境中所有有价值的东西的成本治理和/或资源治理。

[0268] 在云基础设施环境中供应的资源可以基于指派给每个资源的标签值被分组在容器1142、1144、1146中。

[0269] 根据示例,资源实例被逻辑上布置到不同的容器1142、1144、1146中,从而提供分开的成本中心,并且每个成本中心被指派标签值,其中可以向最终用户收费以访问各种成本中心,其中成本中心的标签被用于成本治理和核算。一些客户可以在其最终组价格内对多个成本中心进行建模。对于每个成本中心可以有分开的标签,并且云可以针对每个标签指派有限的资源集。

[0270] 在示例实施例中,一个或多个策略1120为最终用户提供一种机制,以保护他们自己免于访问比标准访问合约所允许的更多的资源,其中最终用户在访问标准合约的标准合约条款下允许的更多资源会发生严重超额。例如,客户可以发送对资源的请求,该请求经由资源控制平面1130激活多个计算资源(诸如例如100个计算资源)的供应,其中合约条款仅允许以标准费率供应90个计算机资源,并且其中任何超额(在示例中为10)都会给客户最终用户带来附加的费用。

[0271] 如上所述并且根据实施例,细粒度方法可以基于标签并且特别是基于与资源相关联的标签来提供资源限制。与资源相关联的标签提供可以在资源治理110(图1)中使用的机制,并且云提供商也可以将它们用于成本治理。系统和方法可以创建一种机制,以通过标签在包括组级别在内的级别控制成本。作为示例,用户可以指定细粒度的规则/策略,诸如:

[0272] Set limits<resource type>to 10in group A where target.resource.tag=

“finance”

[0273] 根据实施例,标签可以与资源相关联,并且利用以上方法,标签可以在一个或多个所选择级别(诸如例如在组级别)保护客户免受资源超额的影响。在以上示例中,当组A中的标签值为“finance”时,用户将无法创建超过10个<resource type>的实例。这种在组级别限制成本的方法向客户端提供细粒度的机制以更好地管理成本和资源使用。

[0274] 根据示例实施例,第一策略1120a保护客户免受与第一标签相关联并且可以分组在容器1142、1144、1146中的任何一个中的资源的超额的影响。类似地,第二策略1120b保护客户免受与第二标签相关联并且可以分组在容器1142、1144、1146中的任何一个中的资源的超额的影响。同样类似地,第三和第四策略1120c、1120d保护客户免受与第三和第四标签相关联并且也可以分组在容器1142、1144、1146中的任何一个中的资源的超额的影响。

[0275] 根据示例,资源请求1111a-1111d中的每一个包括具有代表被请求的资源实例的类型(Resource Type)的数据、代表被请求的资源实例的组(Resource Group)的数据以及代表资源请求的上下文(Context Tag)的数据的字段。在这点上,资源请求1111a-1111d中的每一个包括资源类型字段1112a-1112d,其具有代表被请求的资源实例的类型(Resource Type)的数据。此外,资源请求1111a-1111d中的每一个包括资源组字段1113a-1113d,其代表被请求的资源实例的组(Resource Group)。此外,资源请求1111a-1111d中的每一个包括请求上下文字段1114a-1114d,其包括代表资源请求的上下文(Context Tag)的请求上下文数据。此外,资源请求1111a-1111d中的每一个包括请求特点字段1115a-1115d,其包括代表资源请求的特点(Characteristic Tag)的请求特点数据。为了便于讨论,请求特点字段1115a-1115d被单独示出和描述,但是应该理解的是,请求特点可以从资源类型信息或从资源组信息或从形成资源请求的任何其它信息中导出。

[0276] 根据示例,计算控制平面可以从限制服务数据平面740(图7)提取(可替代地,可以在资源控制平面1130处执行提取操作)与作为来自用户701的请求的一部分的标签相关联的所有基于标签的限制/配额1150。此类基于标签的配额可以跨例如整个租赁应用,并且不必特定于隔间。

[0277] 根据实施例,计算控制平面可以进行检查以查看所请求的使用是否会违反存储器1150中的任何配额或限制规则的任何基于标签的限制/配额。如果所请求的使用没有违反任何基于资源标签的配额,那么可以将请求处理到例如数据库。如果所请求的使用将违反沿着隔间树的任何基于资源标签的限制/配额,那么可以丢弃请求并且可以发送消息以通知用户。

[0278] 根据示例实施例,基于资源标签的限制/配额被应用于租赁内的隔间。根据下面将要描述的示例实施例,基于资源标签的限制/配额跨例如整个租赁(包括在其中启用租赁的所有区域,基本上跨越多个隔间)被应用。当确定所请求的实例是否会违反任何基于资源标签的限制/配额时,计算控制平面可以确保新请求(例如,经由服务开发人员工具包(SDK))不违反任何配额。如果请求超过配额,那么计算控制平面可以使请求失败。

[0279] 根据实施例,计算控制平面的SDK可以存在于云基础设施环境的每个控制平面中。根据实施例,计算控制平面可以与限制服务数据平面结合工作以对照基于资源标签的限制/配额来检查新请求。

[0280] 根据实施例,当客户端触发创建云资源的操作时,请求转到资源控制平面。考虑到

在所选择级别(诸如例如在组级别)配置的限制,资源控制平面向限制服务询问是否可以安全地启动资源。如果限制服务器向资源控制平面返回否定/失败响应,那么不创建资源,并且丢弃请求。

[0281] 在内部,当限制服务接收到请求时,它检查与资源创建相关联的标签,并在存储库或数据库或存储器中查找以查看具有那个特定标签的资源的当前使用情况。资源可以与多于一个标签相关联,并且资源可以与多个不同的标签相关联。如果标签值超过在所选择级别(诸如例如在组级别)配置的限制,那么它向资源控制平面发送失败/否定响应,使得资源创建失败。

[0282] 根据实施例,存储器设备1140存储基于标签的配额数据1150,该数据代表租赁305(图3)中资源供应的多个基于标签的配额。接收在租赁中供应资源的请求1111a-1111d。请求1111a-1111d可以包括代表请求的请求特点的请求特点数据1115a-1115d。根据实施例,确定与和请求的请求特点对应的资源标签150a、160a、170a(图1)相关联的租赁中资源的使用情况并且将其与多个基于标签的配额和/或限制1150进行比较。基于所确定的使用情况超过多个基于标签的配额之一,来丢弃供应资源的请求。

[0283] 根据实施例的示例,租赁305包括存储资源的多个隔间1142、1144、1146,其中多个隔间中的每个隔间提供该隔间内的资源集相对于其它隔间内的资源的一个或多个其它集合的隔离。跨多个隔间共同确定与和请求的请求特点对应的资源标签相关联的租赁中的资源的使用情况,并且将跨多个隔间共同确定的使用情况与租赁的多个基于标签的配额进行比较。供应资源的请求1111a-1111d基于跨多个隔间共同确定的使用情况超过多个基于标签的配额之一而被丢弃。

[0284] 根据实施例的示例,存储在存储器设备1140中的基于标签的配额数据1150代表对应的多个资源类型的租赁中的租赁中资源供应的多个基于标签的配额。此外,在租赁中供应资源的请求1111a-1111d的请求特点数据1115a-1115d包括代表所请求资源的第一资源类型1112a-1112d的资源类型数据。在示例实施例中,与和第一资源类型1112a对应的资源标签相关联的租赁中的资源的使用情况被确定并且被与第一资源类型的租赁中资源供应的第一基于标签的配额1120a进行比较。供应第一资源类型1112a的资源的请求1111a,基于所确定的与和第一资源类型对应的资源标签相关联的资源的使用情况超过例如作为第一资源类型的租赁中资源供应的策略1120a而存储的第一基于标签的配额,而被丢弃。

[0285] 根据实施例的示例,存储在存储器设备1140中的基于标签的配额数据1150代表分配给租赁的对应多个用户组的租赁中资源供应的多个基于标签的配额。在租赁中供应资源的请求1111a-1111d的请求特点数据1115a-1115d包括用户组数据1113a-1113d,其代表指派给请求资源的系统的用户的用户组类别。确定与和用户组类别对应的资源标签相关联的租赁中的资源的使用情况,并将其与分配给用户组类别的租赁中的资源供应的第一基于标签的配额1120b进行比较。供应资源的请求,基于所确定的与和用户组类别对应的资源标签相关联的租赁中的资源的使用情况超过分配给用户组类别的租赁中的资源供应的第一基于标签的配额,而被丢弃。

[0286] 根据实施例的示例,存储在存储器设备中的基于标签的配额数据代表分配给租赁的多个用户组的多个资源类型的租赁中的租赁中资源供应的基于标签的配额。在租赁中供应资源的请求的请求特点数据可以包括例如i)代表所请求资源的第一资源类型的资源类

型数据;和/或ii)代表指派给请求资源的系统的用户的用户组类别的用户组数据。在该示例中,确定与和供应给第一用户组类别的第一资源类型对应的资源标签相关联的租赁中的资源的使用情况,并将其与分配给第一用户组类别的第一资源类型的租赁中的资源供应的第一基于标签的配额进行比较。此外,供应资源的请求基于所确定的与和供应给第一用户组类别的第一资源类型对应的资源标签相关联的租赁中资源的使用情况超过分配给第一用户组类别的第一资源类型的租赁中资源供应的第一基于标签的配额而被丢弃。

[0287] 根据实施例的示例,存储在存储器设备1140中的基于标签的配额数据1150代表分配给租赁的多个用户组的多个资源类型的租赁中的租赁中资源供应的基于标签的配额。可以从用户701接收跟踪租赁的资源使用情况的第一请求,其中第一请求包括代表第一用户组类别1113a-1113d和第一资源类型1112a-1112d的请求特点数据1115a-1115d。

[0288] 确定与和由第一用户组类别供应的第一资源类型对应的资源标签相关联的租赁中资源的第一使用情况,并且基于所确定的由第一用户组供应的第一资源类型的租赁中资源的第一使用情况,生成资源使用情况跟踪数据。

[0289] 根据实施例的示例,供应资源的请求基于所确定的使用情况超过多个基于标签的配额之一而被丢弃。但是,可以从用户701接收供应租赁的资源的覆盖请求。根据实施例的示例,基于系统接收到覆盖请求而选择性地供应资源。此外,资源使用超额数据是基于响应于系统接收到覆盖请求而选择性地供应的资源而生成的。

[0290] 图12是示出根据示例实施例的用于基于资源标签在云基础设施环境中对供应资源进行限制或施加配额的方法1200的流程图。

[0291] 根据方法1200,在步骤1210中,提供租赁并且将基于标签的限制/配额数据存储在存储器中。在其优选形式中,租赁由包括与计算机可操作地耦合的一个或多个处理器和存储器设备的计算机在相关联的云基础设施环境中提供。存储器设备存储可由计算机执行的基于标签的控制逻辑,以提供对相关联的云基础设施环境中的资源使用情况的基于标签的控制。在示例实施例中,存储在存储器设备中的基于标签的配额数据代表租赁中资源供应的多个基于标签的配额。

[0292] 在步骤1220处接收到在租赁中供应资源的请求。该请求包括代表请求的请求特点的请求特点数据。

[0293] 在步骤1230中,确定所请求资源的资源类型。

[0294] 在步骤1240中,与和请求的请求特点对应的资源标签相关联的租赁中的资源的使用情况由执行基于标签的控制逻辑的一个或多个处理器确定。

[0295] 在步骤1250中,请求基于所请求的资源的组而被指派给隔间中的一个或多个。

[0296] 执行基于标签的控制逻辑的一个或多个处理器确定与标签相关联的资源的使用情况并且在步骤1260处将确定的使用情况与多个基于标签的配额进行比较。

[0297] 资源请求在步骤1290中由执行基于标签的控制逻辑的一个或多个处理器基于在步骤1270处确定使用情况超过多个基于标签的配额之一而供应资源的请求而被丢弃。

[0298] 所请求资源的实例在步骤1280中由执行基于标签的控制逻辑的一个或多个处理器基于在步骤1270处确定使用情况不超过多个基于标签的配额之一而供应资源的请求而被创建。

[0299] 根据实施例的示例,通过提供存储资源的多个隔间来提供租赁,其中多个隔间中

的每个隔间提供隔间内的资源集相对于其它隔间中的资源的一个或多个其它集合的隔离。此外,确定使用情况包括跨多个隔间共同确定与和请求的请求特点对应的资源标签相关联的租赁中资源的使用情况,并且比较包括将跨多个隔间共同确定的使用情况针对租赁的多个基于标签的配额进行比较。此外,丢弃请求包括基于跨多个隔间共同确定的使用情况超过多个基于标签的配额之一来丢弃供应资源的请求。

[0300] 根据实施例的示例,存储基于标签的配额数据包括将如下基于标签的配额数据存储在存储器设备中:该数据代表租赁的对应多个资源类型的租赁中资源供应的多个基于标签的配额。此外,接收请求包括接收包括代表所请求资源的第一资源类型的资源类型数据的请求,并且确定包括确定与和第一资源类型对应的资源标签相关联的租赁中资源的使用情况。此外,比较包括将所确定的使用情况与第一资源类型的租赁中的资源供应的第一基于标签的配额进行比较,并且丢弃请求包括基于所确定的与第一资源类型对应的资源标签相关联的租赁中资源的使用情况超过第一资源类型的租赁中资源供应的第一基于标签的配额来丢弃请求。

[0301] 根据实施例的示例,存储基于标签的配额数据包括将如下基于标签的配额数据存储在存储器设备中:该数据代表分配给租赁的对应多个用户组的租赁中资源供应的多个基于标签的配额,并且接收请求包括接收包括代表指派给请求资源的系统的用户的用户组类别的资源类型数据的请求。此外,确定包括确定与和用户组类别对应的资源标签相关联的租赁中资源的使用情况,并且比较包括将所确定的使用情况与分配给用户组类别的租赁中资源供应的第一基于标签的配额进行比较。此外,丢弃请求包括基于所确定的与和用户组类别对应的资源标签相关联的租赁中资源的使用情况超过分配给用户组类别的租赁中资源供应的第一基于标签的配额而丢弃请求。

[0302] 根据实施例的示例,存储基于标签的配额数据包括将如下基于标签的配额数据存储在存储器设备中:该数据代表分配给租赁的多个用户组的多个资源类型的租赁中的租赁中资源供应的基于标签的配额,并且接收请求包括接收包括资源类型数据的请求,该资源类型数据包括:i)代表所请求资源的第一资源类型的资源类型数据;以及ii)代表指派给请求资源的系统的用户的用户组类别的用户组数据。此外,确定包括确定与资源标签相关联的租赁中资源的使用情况,该资源标签与供应给第一用户组类别的第一资源类型对应,并且比较包括将所确定的使用情况与分配给第一用户组类别的第一资源类型的租赁中的资源供应的第一基于标签的配额进行比较。此外,丢弃请求包括基于所确定的与和提供给第一用户组类别的第一资源类型对应的资源标签相关联的租赁中资源的使用情况超过分配给第一用户组类别的第一资源类型的租赁中资源供应的第一基于标签的配额来丢弃请求。

[0303] 根据实施例的示例,在存储器设备中存储代表分配给租赁的多个用户组的多个资源类型的租赁中的租赁中资源供应的基于标签的配额的基于标签的配额数据。该方法还包括接收跟踪租赁的资源使用情况的第一请求,其中第一请求包括代表第一用户组类别和第一资源类型的请求特点数据。此外,确定与和由第一用户组类别供应的第一资源类型对应的资源标签相关联的租赁中资源的第一使用情况,并基于所确定的由第一用户组供应的第一资源类型的租赁中资源的第一使用情况生成资源使用情况跟踪数据。

[0304] 该方法还包括接收供应租赁的资源的覆盖请求,基于系统接收到覆盖请求而选择性地供应资源,以及基于响应于系统接收到覆盖请求而选择性地供应资源来生成资源使用

情况超额数据。

[0305] 根据各种实施例,本文的教导可以使用一个或多个常规的通用或专用计算机、计算设备、机器或微处理器方便地实现,包括根据本公开的教导的一个或多个处理器、存储器和/或编程的计算机可读存储介质。熟练的程序员可以基于本公开的教导容易地准备适当的软件编码,这对于软件领域的技术人员来说是显而易见的。

[0306] 在一些实施例中,本文的教导可以包括计算机程序产品,该计算机程序产品可以使用计算机可读介质(或多种介质)来传送(例如,存储或携带)指令,以供一个或多个可编程处理器或计算机执行本教导的任何过程。计算机可读介质的一个示例是具有存储在其上/其中的指令的非暂态计算机可读存储介质(多种介质),指令可以被用于对计算机进行编程以执行本教导的任何过程。此类存储介质的示例可以包括但不限于硬盘驱动器、硬盘、硬盘驱动器、固定盘或其它机电数据存储设备、软盘、光盘、DVD、CD-ROM、微驱动器和磁电机-光盘、ROM、RAM、EPROM、EEPROM、DRAM、VRAM、闪存设备、磁卡或光卡、纳米系统或适于指令和/或数据的非暂态存储的其它类型的存储介质或设备。计算机可读介质的另一个示例是在计算机系统之间或在给定计算机系统的组件之间传送此类指令的瞬态介质。瞬态介质的示例可以包括但不限于载波和传输信号。

[0307] 因此,从一个角度来看,已经描述了在云基础设施环境中支持基于标签的资源限制或配额的系统和方法。云管理员一般没有能力限制现有云中的资源使用情况。授予用户创建资源的许可允许他们创建任何数量的资源,最多达到预定义的账户限制。标签与资源相关联,从而通过允许微调成本控制来允许管理员将用户的资源使用情况限制到适当的水平。确定与和供应资源的请求的请求特点对应的资源标签相关联的租赁中资源的使用情况,并将其与多个基于标签的配额进行比较,其中供应资源的请求基于所确定的使用情况超过多个基于标签的配额之一而被丢弃。

[0308] 前述描述是为了说明和描述的目的而提供的。其并非旨在是详尽的或将保护范围限制到所公开的精确形式。对于本领域技术人员来说,许多修改和变化将是显而易见的。例如,虽然本文提供的示例中有几个说明了与企业软件应用组件(诸如Oracle Fusion Applications);云环境(诸如Oracle云基础设施);以及云服务(诸如Oracle Fusion Analytics)一起使用;但是根据各种实施例,本文描述的系统和方法可以与其它类型的企业软件应用、云环境、云服务、云计算或其它计算环境一起使用。

[0309] 选择和描述实施例是为了最好地解释本教导的原理及其实际应用,从而使本领域的其他技术人员能够理解各种实施例以及适于预期的特定用途的各种修改。旨在由以下权利要求及其等同物定义范围。

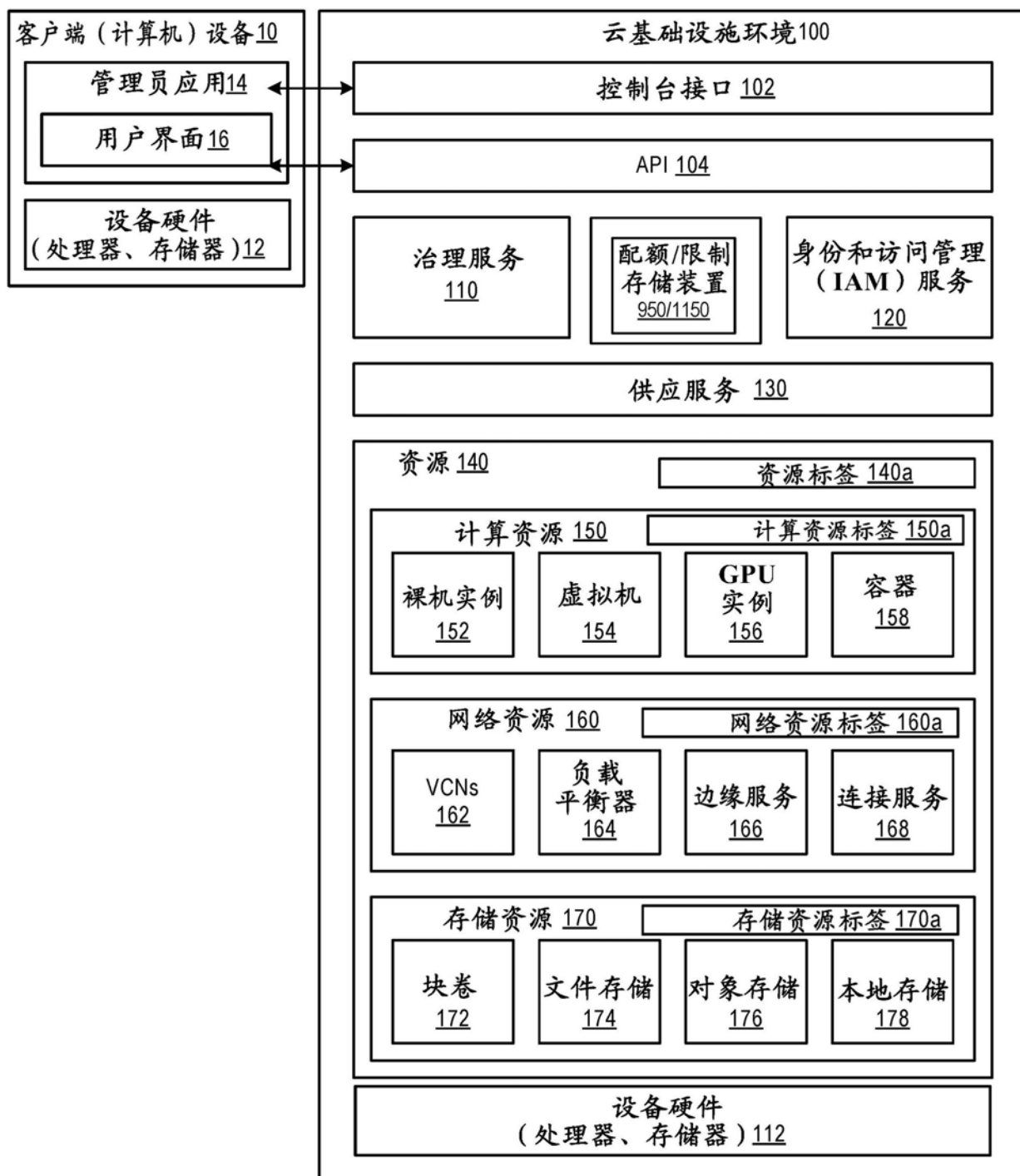


图1

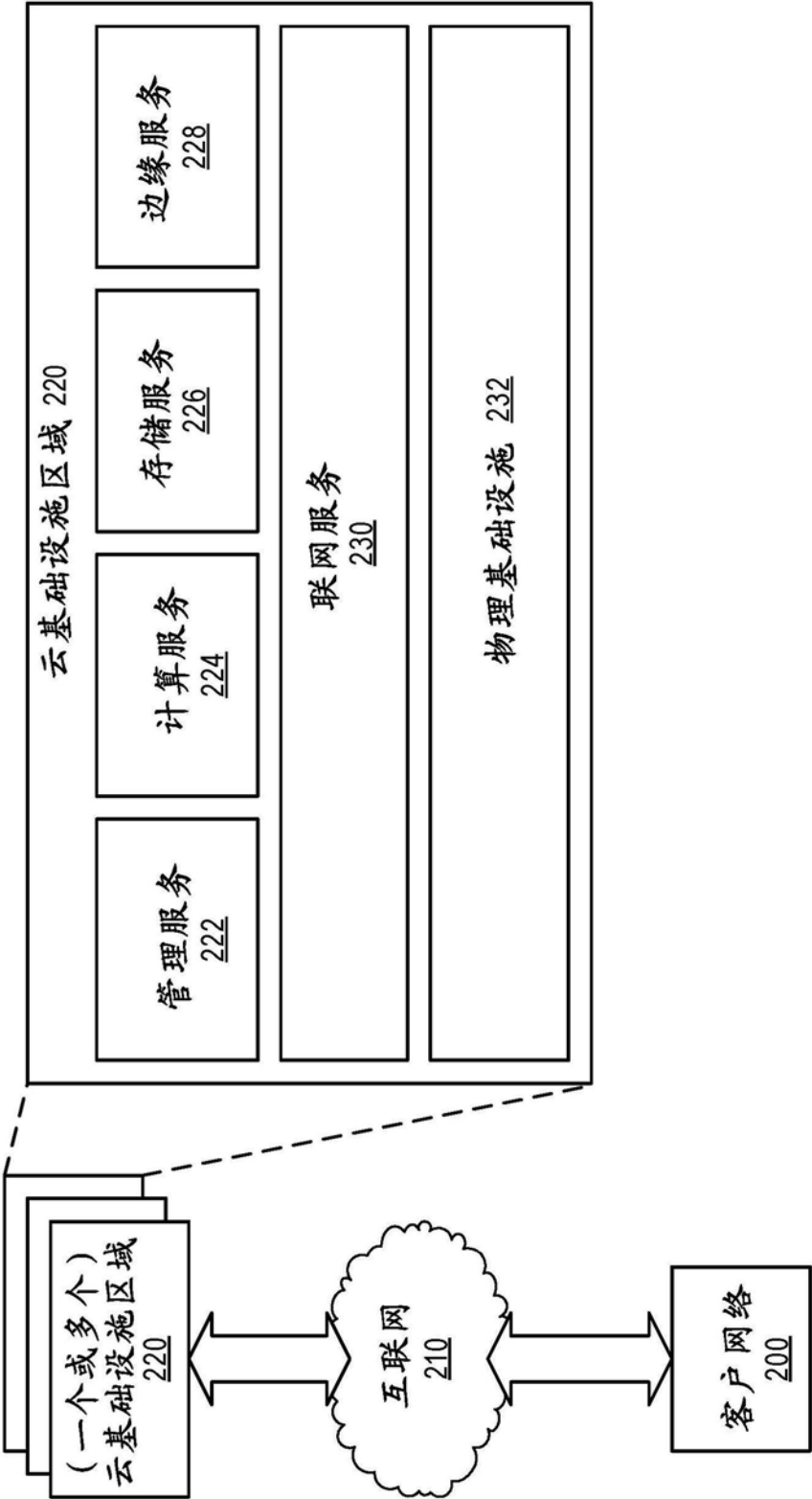


图2

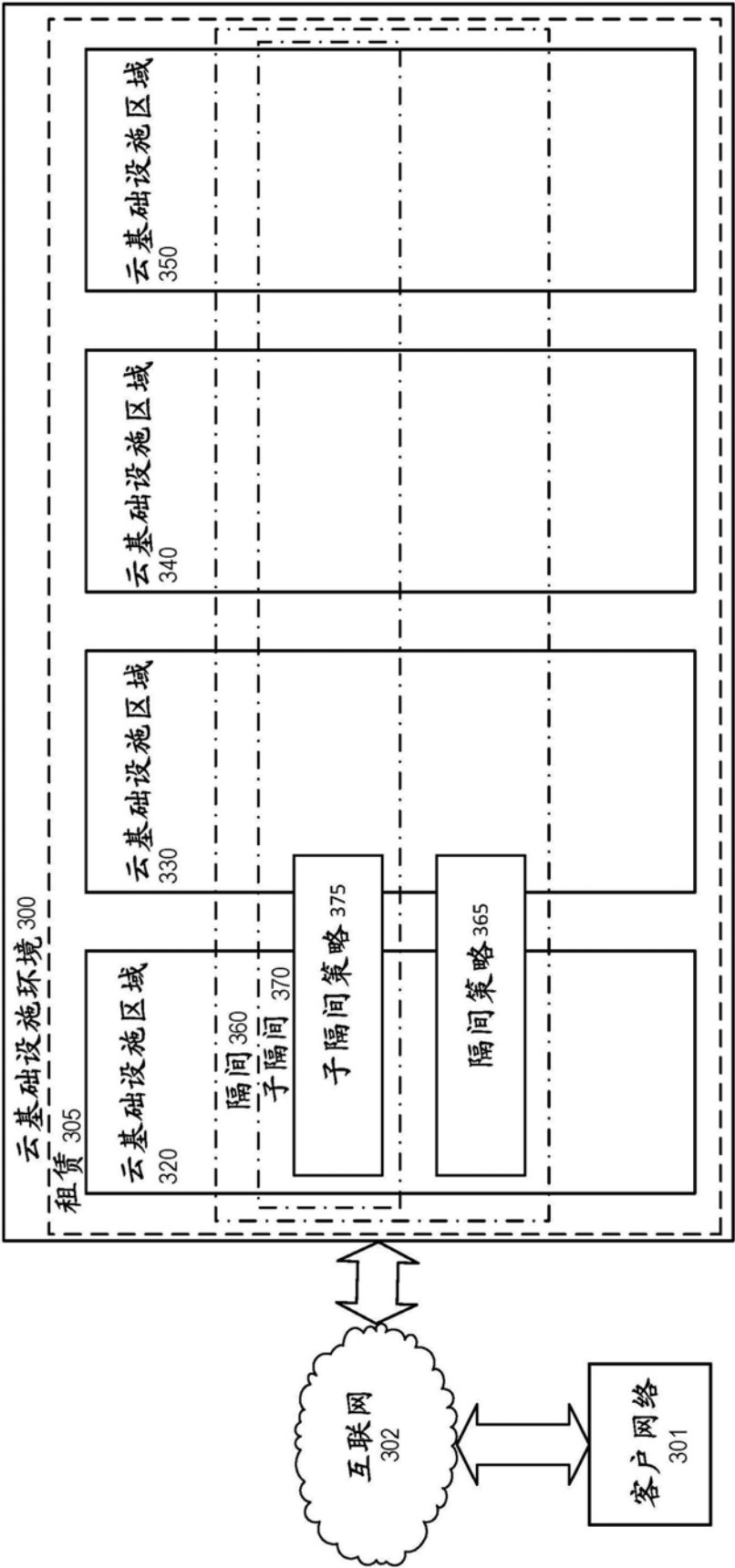


图3

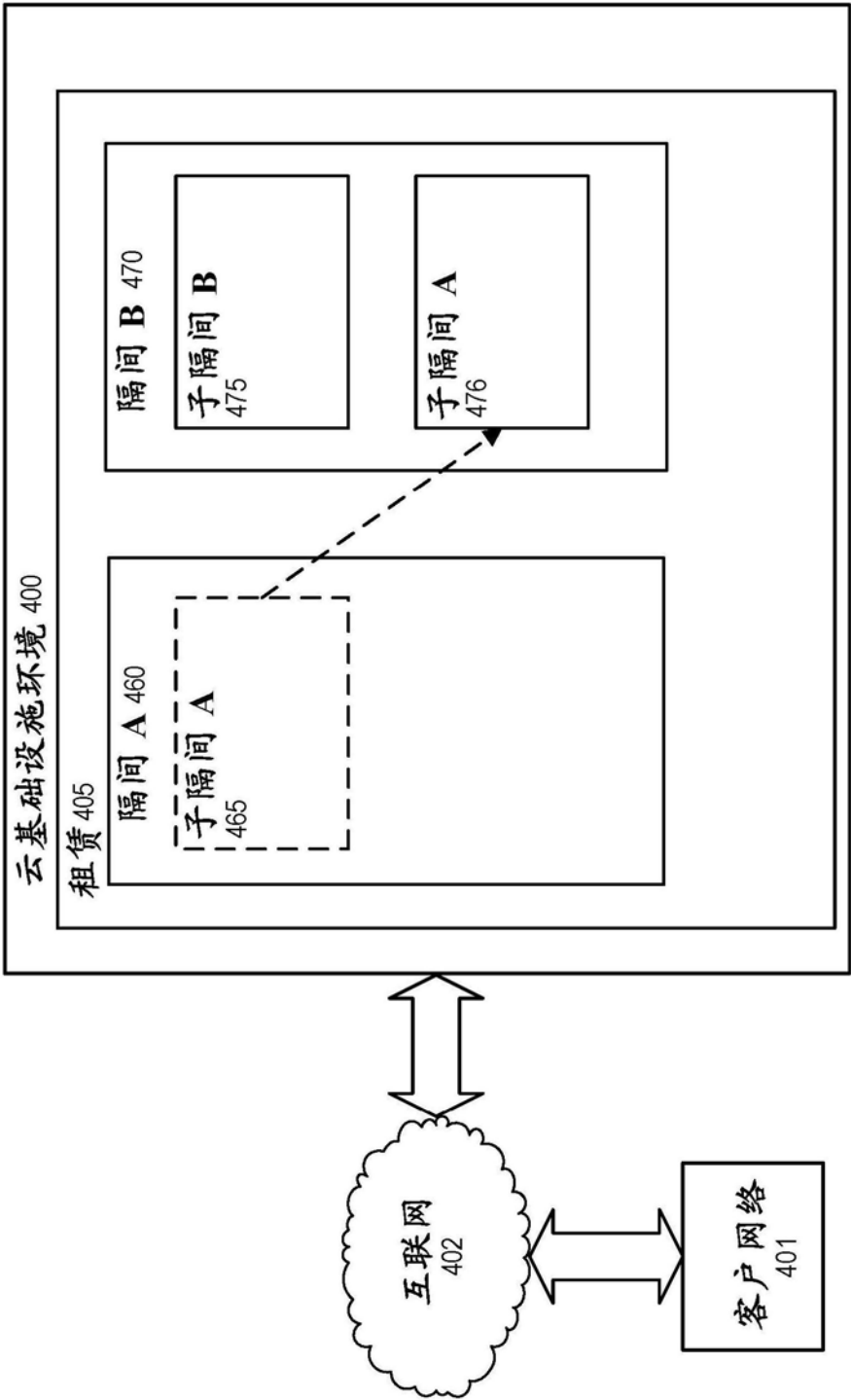


图4

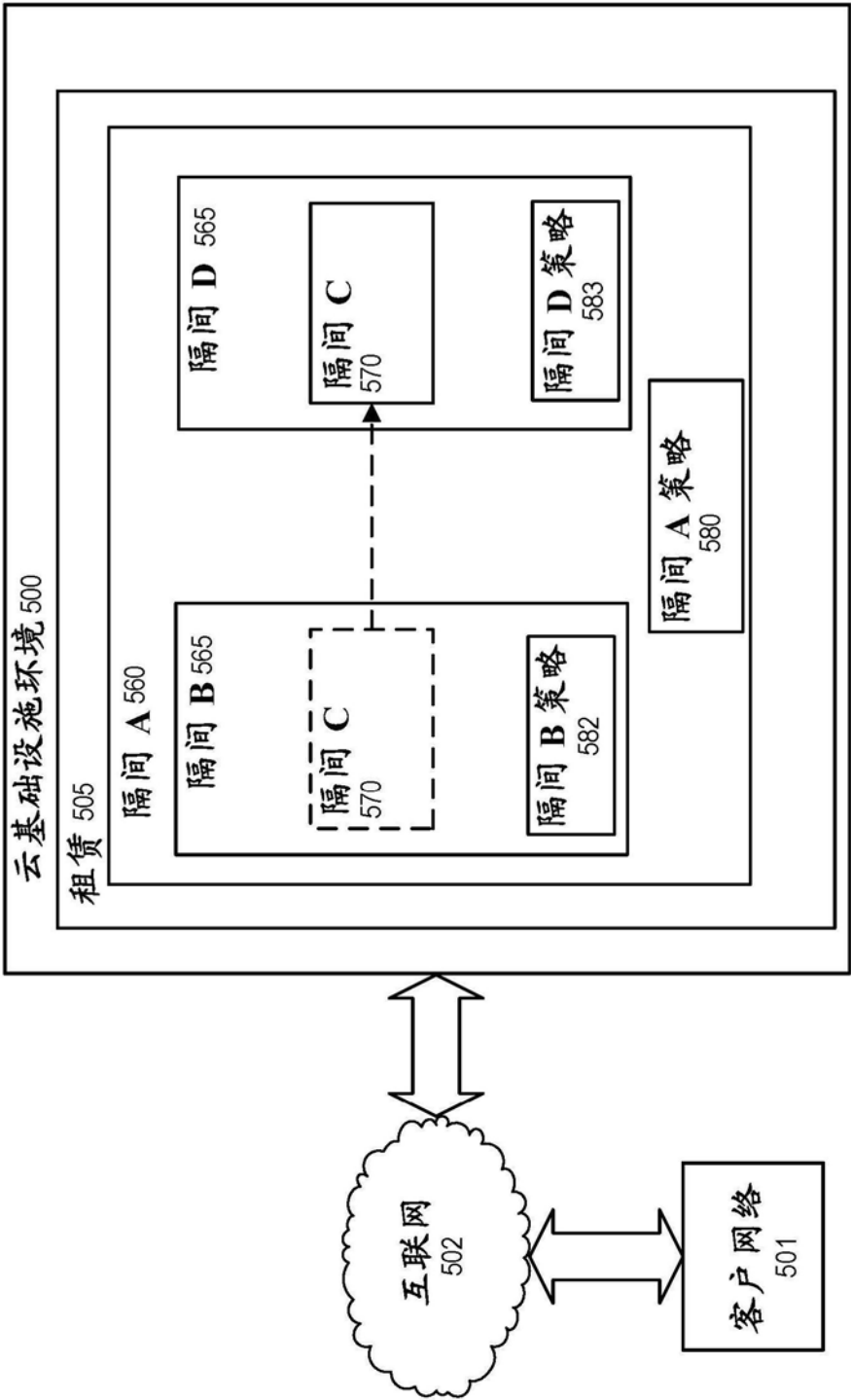


图5

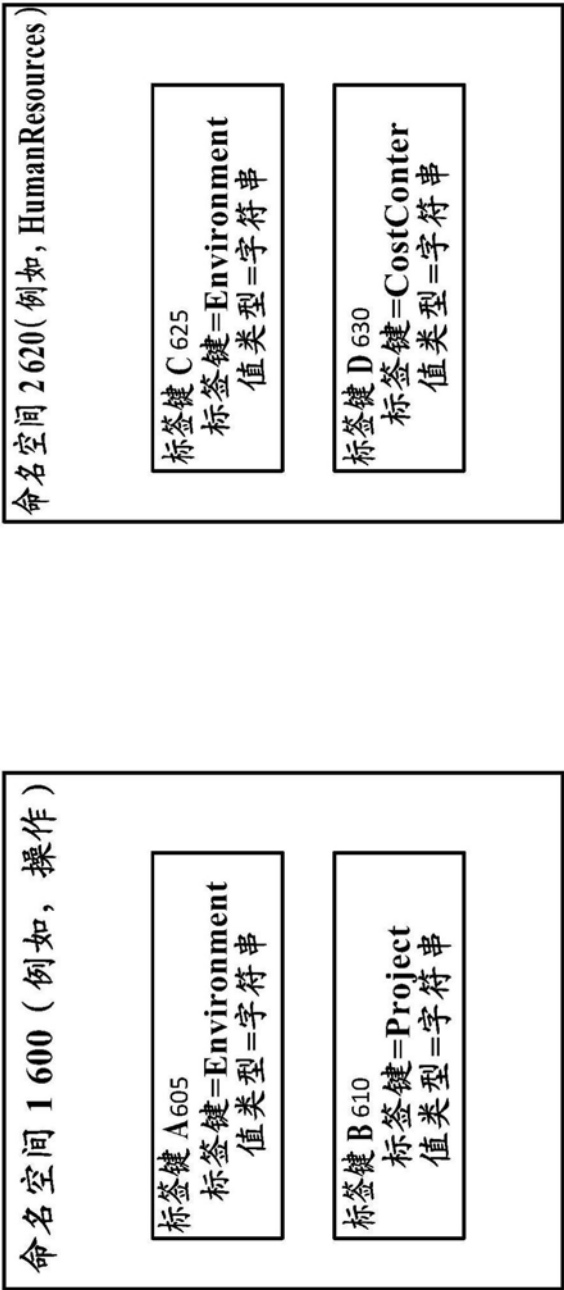


图6

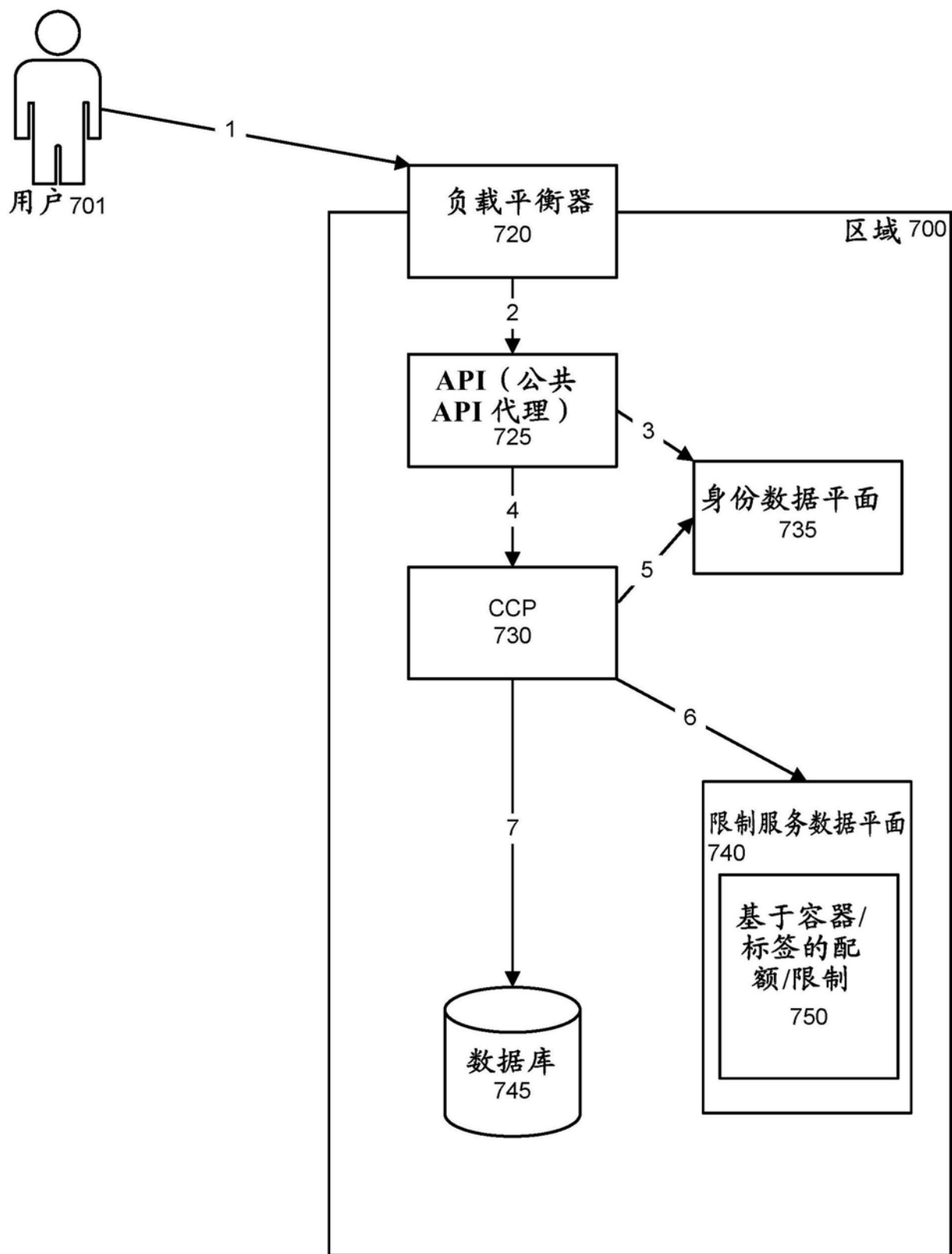


图7

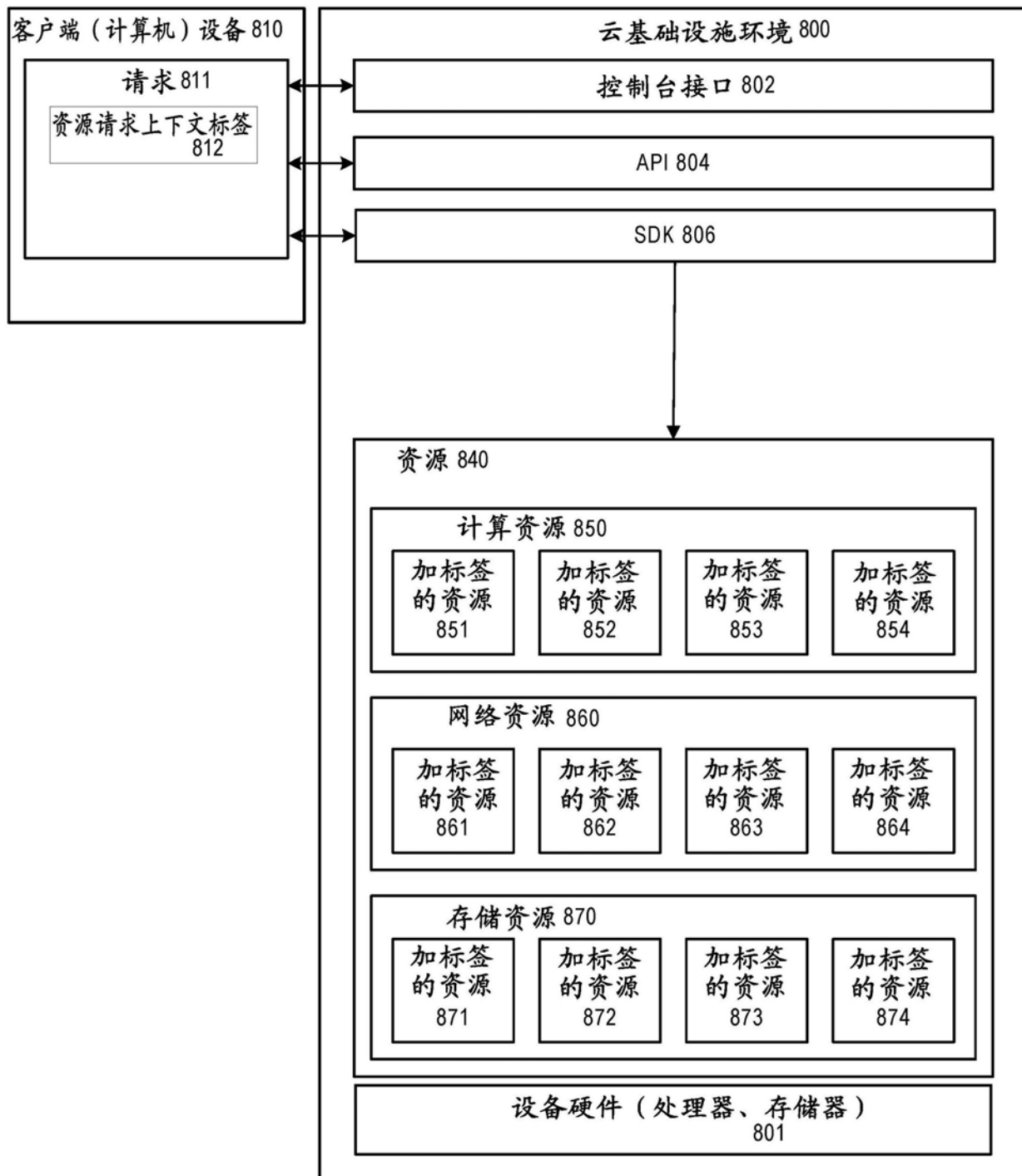


图8

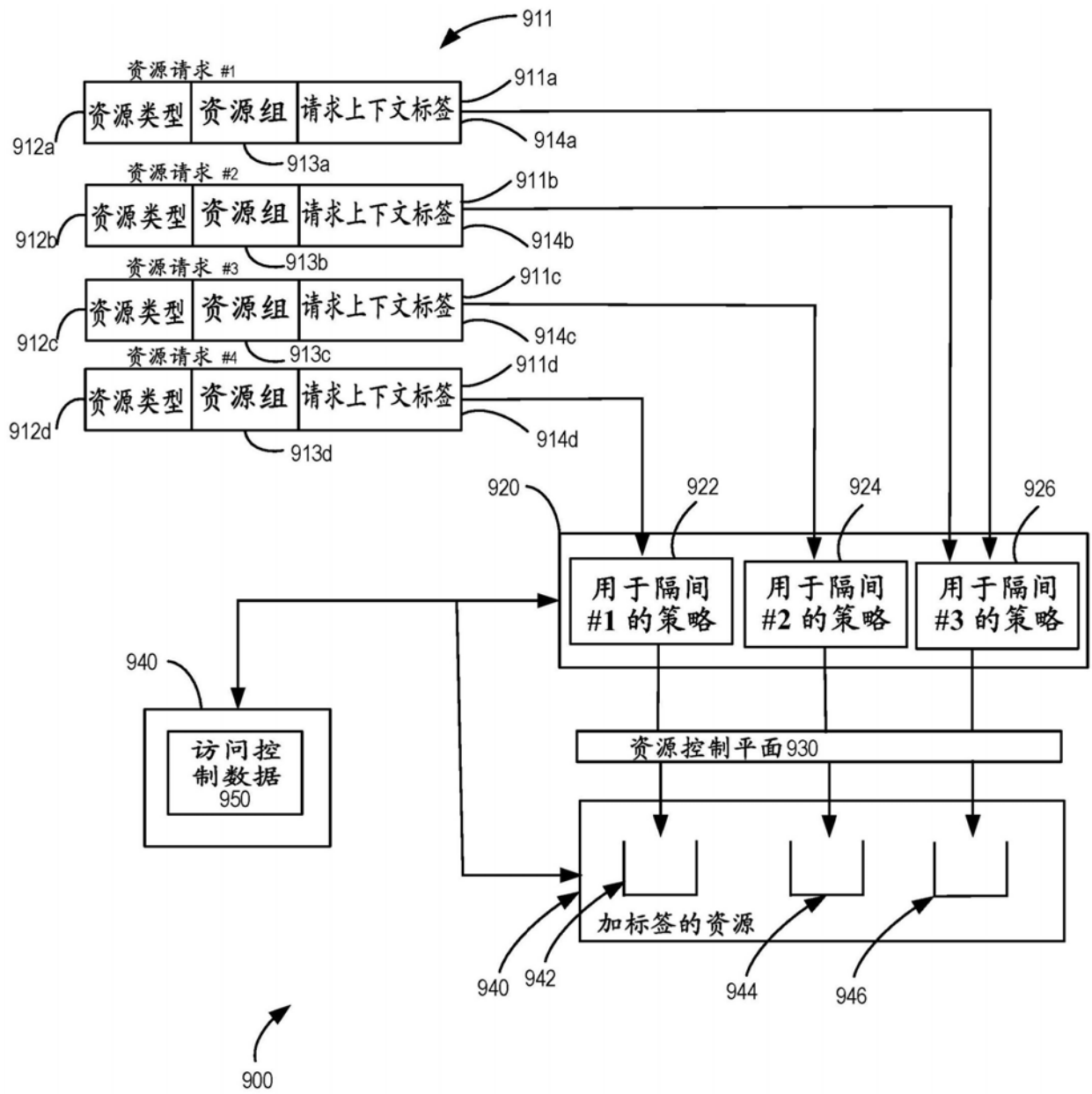


图9

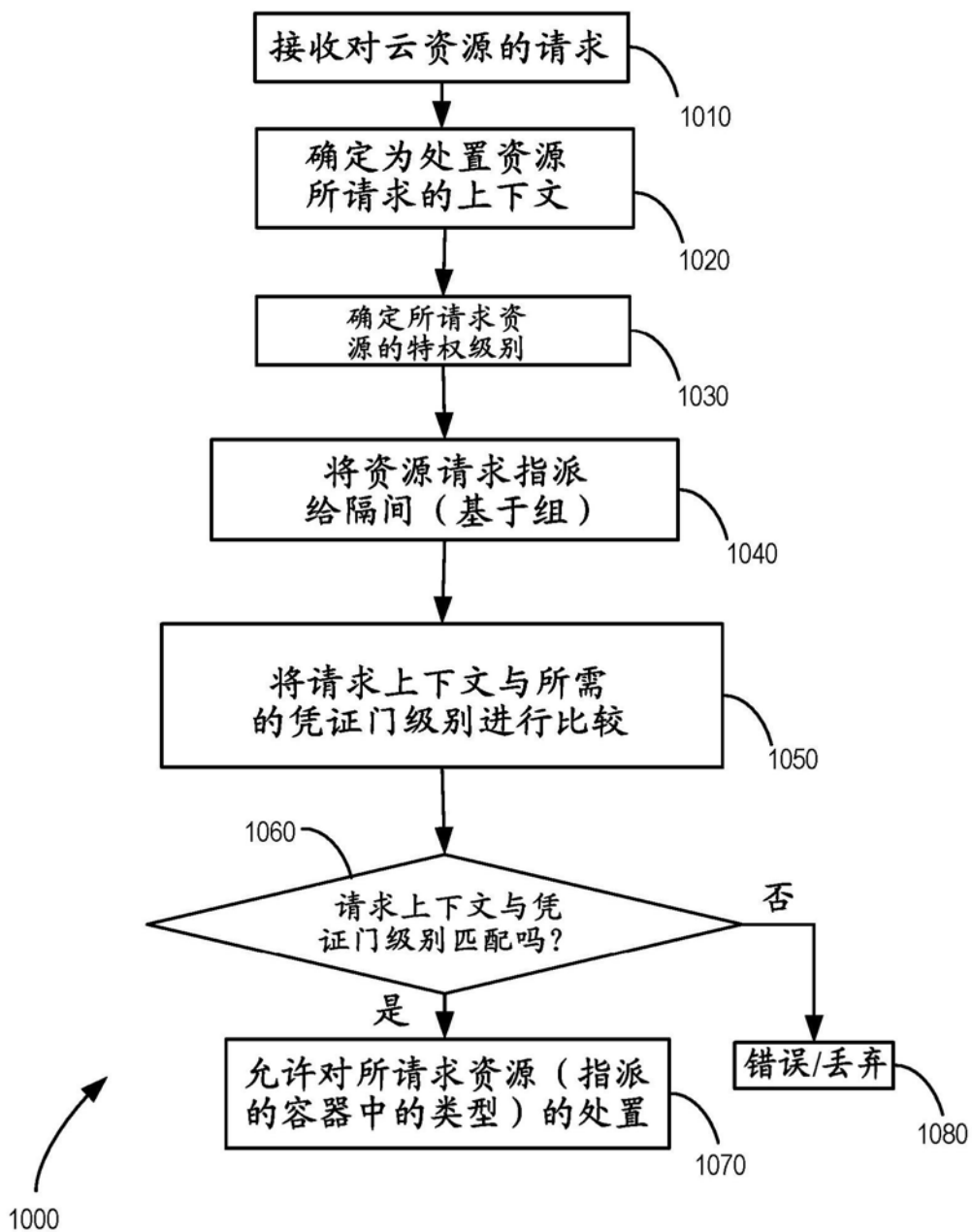


图10

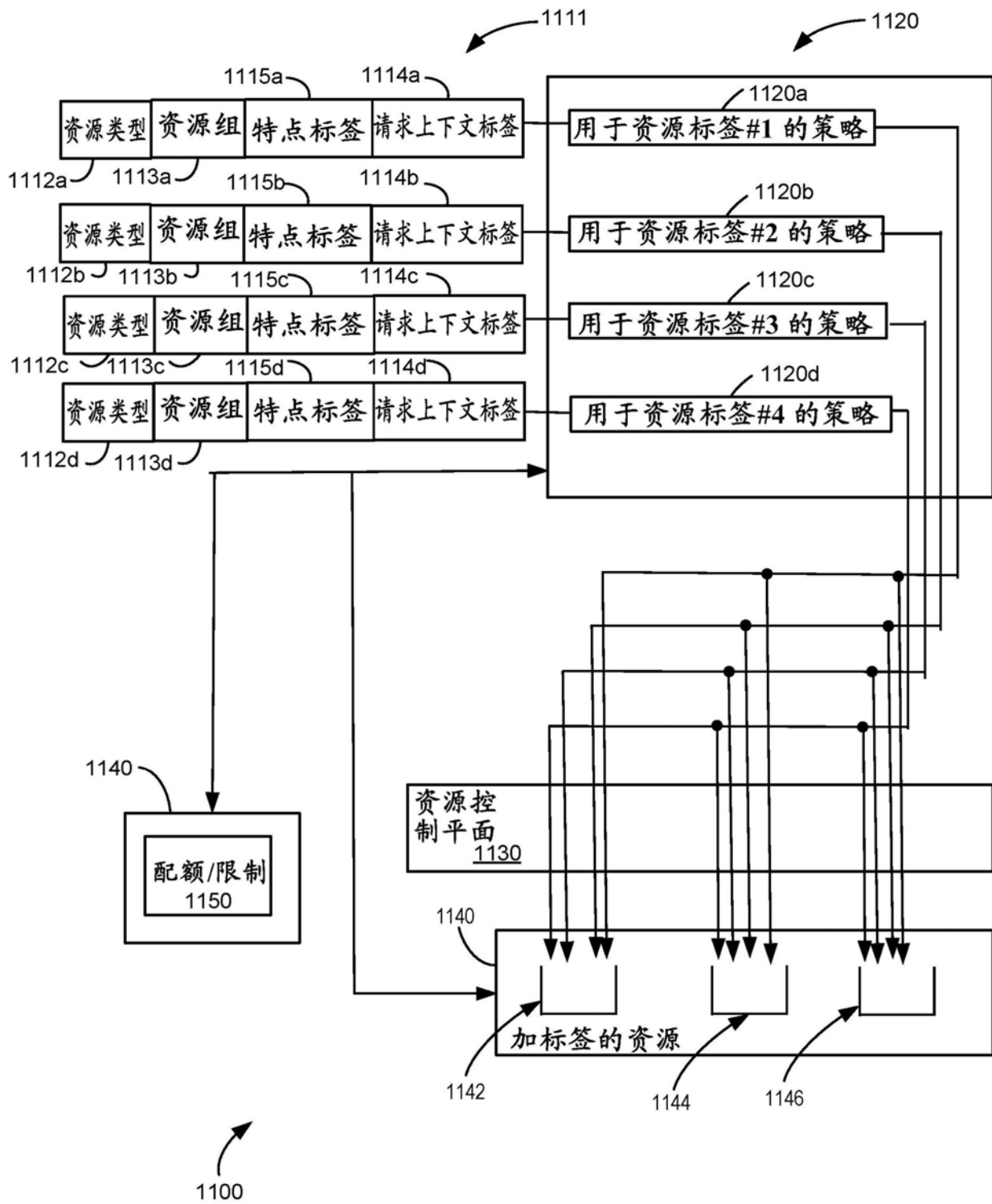


图11

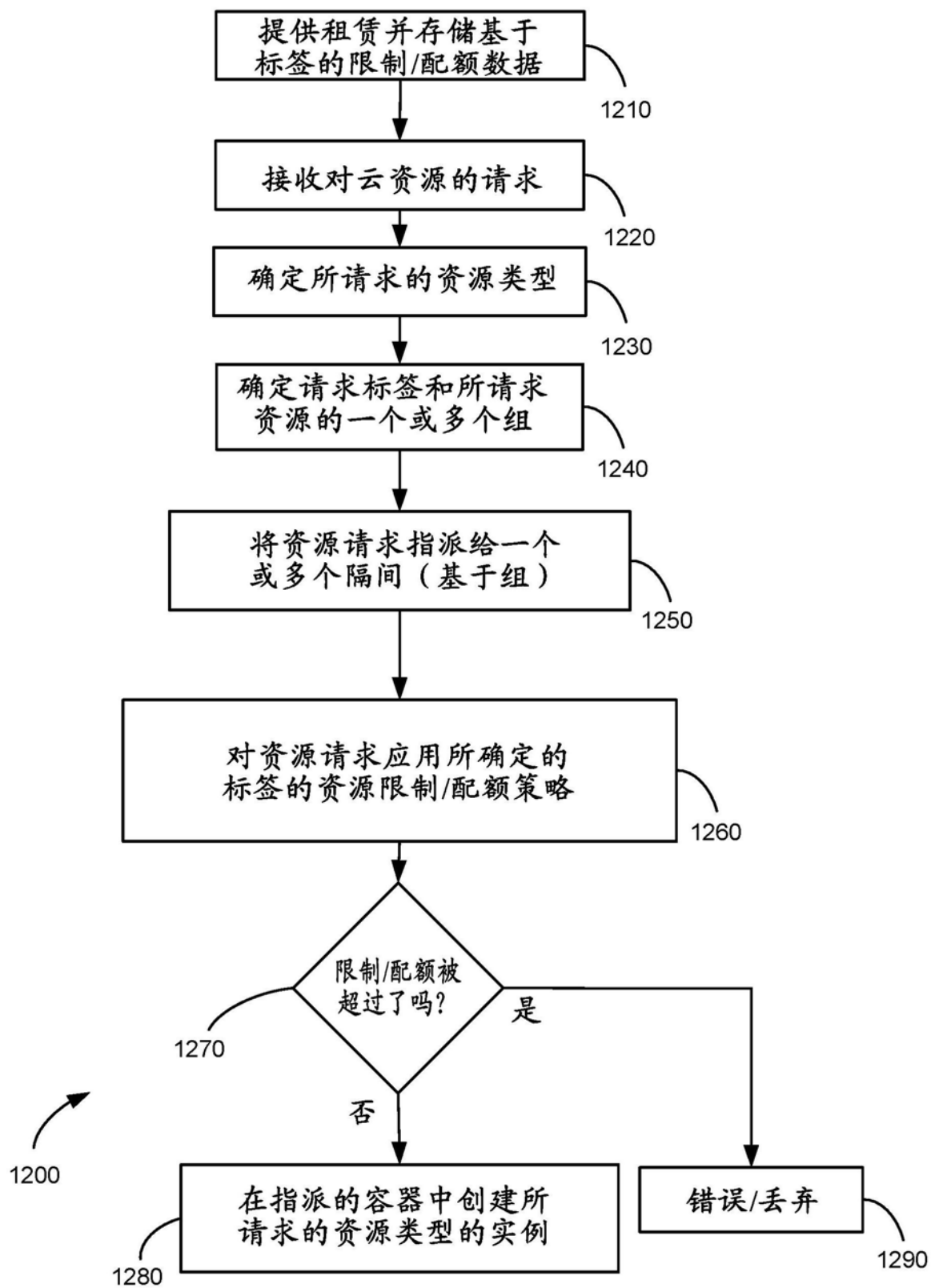


图12