(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 115380275 A
(43)申请公布日 2022.11.22

(21)申请号 202180025714.6
(22)申请日 2021.02.19
(30)优先权数据
16/801,590 2020.02.26 US
(85)PCT国际申请进入国家阶段日
2022.09.29
(86)PCT国际申请的申请数据
PCT/US2021/018750 2021.02.19
(87)PCT国际申请的公布数据
WO2021/173440 EN 2021.09.02
(71)申请人 甲骨文国际公司
地址 美国加利福尼亚
(72)发明人 K·C·格罗斯 S·R·桑都尔
王光超
(74)专利代理机构 中国贸促会专利商标事务所有限公司 11038
专利代理师 边海梅
(51)Int.Cl.
G06F 9/50(2006.01)
G06F 11/30(2006.01)
G06N 20/00(2006.01)
G05B 15/02(2006.01)
G05B 17/02(2006.01)

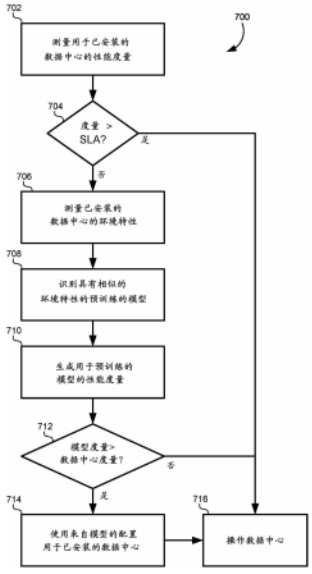
权利要求书2页 说明书20页 附图11页

(54)发明名称

用于数据中心的启用模式识别的自主配置优化

(57)摘要

一种确定用于数据中心的最优配置的基于模型的方法可以使用环境室来表征在温度和海拔的不同组合下各种数据中心配置的性能。当不同的配置在每个温度/海拔组合下执行压力工作负载时，可以从它们记录遥测数据，并且可以使用遥测数据来训练模型的对应库。当正在配置新数据中心时，新数据中心的温度/海拔可以被用于从相似的温度/海拔选择预训练的模型。能够将当前配置的性能与模型的性能进行比较，并且如果模型性能更好，那么可以使用基于模型的新配置作为用于数据中心的最优配置。



CN 115380275 A

1.一种包括指令的非暂态计算机可读介质,指令在由一个或多个处理器执行时使所述一个或多个处理器执行操作,包括:

从数据中心接收第一性能度量,其中:

数据中心包括第一配置;以及

第一性能度量是在数据中心执行工作负载时生成的;

在数据中心执行工作负载时接收用于数据中心的一个或多个环境特性;

识别使用来自一个或多个数据中心配置的数据训练的模型,该一个或多个数据中心配置在具有与数据中心的所述一个或多个环境特性相似的环境特性的环境中操作;

从模型生成第二性能度量;

将第二性能度量与第一性能度量进行比较;以及

从模型确定用于数据中心的第二配置。

2.根据权利要求1所述的非暂态计算机可读介质,其中模型是从多个模型当中识别出来的,其中所述多个模型中的每个模型是使用来自在具有不同环境特性的环境中操作的一个或多个数据中心配置的数据来训练的。

3.根据权利要求1所述的非暂态计算机可读介质,其中所述一个或多个数据中心配置被放置在环境室中以在所述一个或多个数据中心配置执行工作负载时控制环境特性。

4.根据权利要求3所述的非暂态计算机可读介质,其中环境室控制所述一个或多个数据中心配置周围的温度和模拟的海拔。

5.根据权利要求4所述的非暂态计算机可读介质,其中环境室被配置为模拟海拔并在该海拔的多个温度下收集整套遥测信号。

6.根据权利要求5所述的非暂态计算机可读介质,其中所述多个温度包括近似15℃和35℃之间的温度。

7.根据权利要求4所述的非暂态计算机可读介质,其中环境室模拟近似海平面和5000英尺之间的多个海拔。

8.根据权利要求7所述的非暂态计算机可读介质,其中所述多个温度以近似1℃的间隔递增。

9.根据权利要求1所述的非暂态计算机可读介质,其中所述一个或多个环境特性包括数据中心周围的环境温度和安装数据中心的海拔。

10.根据权利要求1所述的非暂态计算机可读介质,其中识别使用来自在具有与所述一个或多个环境特性相似的环境特性的环境中操作的一个或多个数据中心配置的数据来训练的模型包括:

执行最近邻算法以识别与所述一个或多个环境特性最相似的环境特性。

11.根据权利要求1所述的非暂态计算机可读介质,其中最近邻算法最小化用于模型的温度与用于数据中心的温度之间的差异,并且最小化模型的模拟的海拔与数据中心的海拔之间的差异。

12.根据权利要求1所述的非暂态计算机可读介质,其中第一配置包括数据中心中的处理器的数量和类型。

13.根据权利要求1所述的非暂态计算机可读介质,其中第一配置包括数据中心中的硬盘驱动器的数量和类型。

14.根据权利要求1所述的非暂态计算机可读介质,其中第一配置包括数据高速缓存尺寸。

15.根据权利要求1所述的非暂态计算机可读介质,其中数据中心包括云数据中心。

16.根据权利要求1所述的非暂态计算机可读介质,其中操作还包括:

将第一性能度量与服务水平协议(SLA)进行比较;以及

确定第一性能度量不满足SLA,其中模型是响应于确定第一性能度量不满足SLA而被识别的。

17.根据权利要求1所述的非暂态计算机可读介质,其中操作还包括:

确定第二性能度量超过第一性能度量;以及

提供第二配置以由数据中心植入。

18.根据权利要求1所述的非暂态计算机可读介质,其中第一性能度量包括用于数据中心中的每个处理器的处理器性能,以及用于数据中心中的每个硬盘驱动器的I/O性能。

19.一种系统,包括:

一个或多个处理器;以及

一个或多个存储器设备,包括指令,指令在由所述一个或多个处理器执行时使所述一个或多个处理器执行操作,包括:

从数据中心接收第一性能度量,其中:

数据中心包括第一配置;以及

第一性能度量是在数据中心执行工作负载时生成的;

在数据中心执行工作负载时接收用于数据中心的一个或多个环境特性;

识别使用来自一个或多个数据中心配置的数据训练的模型,该一个或多个数据中心配置在具有与数据中心的所述一个或多个环境特性相似的环境特性的环境中操作;

从模型生成第二性能度量;

将第二性能度量与第一性能度量进行比较;以及

从模型确定用于数据中心的第二配置。

20.一种用于自主确定用于数据中心的最优配置的方法,该方法使用在环境特性的各种组合下预训练的模型的库,该方法包括:

从数据中心接收第一性能度量,其中:

数据中心包括第一配置;以及

第一性能度量是在数据中心执行工作负载时生成的;

在数据中心执行工作负载时接收用于数据中心的一个或多个环境特性;

识别使用来自一个或多个数据中心配置的数据训练的模型,该一个或多个数据中心配置在具有与数据中心的所述一个或多个环境特性相似的环境特性的环境中操作;

从模型生成第二性能度量;

将第二性能度量与第一性能度量进行比较;以及

从模型确定用于数据中心的第二配置。

用于数据中心的启用模式识别的自主配置优化

[0001] 相关申请的交叉引用

[0002] 本申请要求于2020年2月26日提交的标题为“PATTERN-RECOGNITION ENABLED AUTONOMOUS CONFIGURATION OPTIMIZATION FOR DATA CENTERS”的美国专利申请No.16/801,590的权益和优先权,该申请通过引用整体并入本文。

背景技术

[0003] 数据中心可以包括计算机系统和相关联组件的任何集合,诸如处理系统、电信系统和数据存储系统。数据中心可以包括用于执行处理操作的多个服务器或CPU核心、用于通信的网络装备、用于存储数据的存储器设备,和/或用于基础设施、电源、数据存储和/或数据通信的冗余或备份组件。现代数据中心可以是安装在客户的内部部署设施上的私有/内部数据中心,而其它数据中心可以是服务提供商提供的面向公众的、基于云的数据中心。基于云的数据中心提供可供云租户用来执行其数据操作的硬件和/或软件。

[0004] 数据中心配置可以包括定义数据中心的构成和操作的任何配置参数。数据中心的配置可以有助于安装和操作各种硬件和软件组件以满足操作要求。一般而言,确定最优数据中心配置是个非常复杂的过程,但选择合适的配置是确保数据中心性能令人满意的关键。次优配置会导致长时延、低吞吐量、多次I/O超时,并最终违反服务水平协议。由于数据中心中各种软件/硬件组件之间交互的复杂性,目前认为确定最优配置更像是一门艺术,而不是由经验结果支持的科学。这常常导致需要数天或数周才能完成的试错过程。因此,用于确定最优数据中心配置的技术改进可以是有益的。

发明内容

[0005] 数据中心中计算机硬件的性能变得越来越依赖于环境因素,诸如环境温度和海拔。在较高温度下,必须增加风扇速度以在CPU密集型和存储器I/O密集型工作负载期间冷却数据中心。在较高的海拔,必须增加风扇速度以补偿稀薄的空气。进而,增加的风扇速度会转化为更多的振动和噪声,这干扰硬盘驱动器的机械操作。增加的温度还造成CPU速度和电压的更多节流(throttling)。如果不适当考虑这些环境因素,那么无法系统地优化数据中心配置的性能。

[0006] 本文描述的实施例通过使用参数化的机器学习模式识别技术优化数据中心配置来解决这些和其它技术问题。可以将各种数据中心配置放置在环境室中,该环境室在不同的环境条件下(诸如不同的海拔和温度组合)循环。可以在环境室中对数据中心执行压力工作负载,并且可以在环境条件的每个集合下提供完整的遥测读数。然后可以使用遥测数据来在每个海拔/温度组合下训练模型,诸如MSET模型或MSET2模型。

[0007] 当安装新的数据中心时,可以使用默认数据中心配置来执行压力工作负载,并且可以实时捕获遥测信号。可以针对默认条件计算不同的性能度量,包括CPU效率、I/O效率和吞吐量,以及其它服务质量(QoS)度量。如果默认配置的性能不令人满意,那么新数据中心的操作条件可以被用于从预训练的模型的库中选择最接近匹配的模型。在安装新数据中心

之前,所选择的预训练的模型可以已经使用在环境室中类似环境条件下操作的配置进行了训练。例如,新数据中心的海拔和/或温度可以被用于选择在相似海拔和/或温度下进行了训练的最接近匹配模型。然后可以使用该模型来确定要用于新数据中心的最优配置。代替对默认配置的试错调整的多次迭代,这个方法使用最多两个不同的配置测试来提供最优配置,从而将设置新配置所需的时间从几天减少到大约几小时。

附图说明

[0008] 通过参考本说明书的其余部分和附图,可以实现对各种实施例的本质和优点的进一步理解,其中,贯穿若干附图使用相似的附图标记指代相似的组件。在一些情况下,子标签与附图标记相关联以表示多个相似组件中的一个。当在没有指定现有子标签的情况下参考附图标记时,意在指代所有这样的多个相似组件。

[0009] 图1图示了根据一些实施例的用于数据中心的冷却系统的简化图。

[0010] 图2图示了根据一些实施例的对象存储系统的简化图。

[0011] 图3图示了根据一些实施例的可以被用于在各种温度和/或海拔训练模型的库的系统。

[0012] 图4图示了根据一些实施例的用于控制环境室的时间和/或高度增量的循环算法的流程图。

[0013] 图5图示了根据一些实施例的用于在每个温度/海拔组合下训练模型的过程。

[0014] 图6图示了根据一些实施例的可以从模型训练过程生成的模型库。

[0015] 图7图示了根据一些实施例的用于使用在环境特性的各种组合下预训练的模型的库来自主地确定用于数据中心的最优配置的方法的流程图。

[0016] 图8图示了根据一些实施例的用于选择具有与数据中心的环境特性相似的环境特性的预训练的模型的方法的流程图。

[0017] 图9图示了用于实现实施例中的一些的分布式系统的简化框图。

[0018] 图10图示了系统环境的组件的简化框图,通过该系统环境由实施例系统的组件提供的服务可以作为云服务来提供。

[0019] 图11图示了其中可以实现各种实施例的示例性计算机系统。

具体实施方式

[0020] 优化用于现代数据中心的配置正在成为向内部部署和基于云的数据中心的客户递送可预测的服务质量(QoS)的关键过程。传统上,数据中心的配置不太考虑环境特性,诸如温度和海拔。数据中心将被配置为优化数据中心的CPU和I/O性能,使得配置将满足客户的服务水平协议(SLA)。然后假设这种配置在一系列环境条件下有效。例如,10年前,20℃和30℃之间的温度波动没有显著影响数据中心的性能。类似地,海平面和5,000英尺之间的高度变化也不会显著影响数据中心硬件/软件的操作。一般而言,在数据中心计算的最初25-30年中使用的计算机硬件的性质是使得CPU和I/O性能仅受环境温度或数据中心海拔的改变的影响最小。因此,用于设计数据中心的配置可以很容易地在具有不同操作温度和/或海拔的多个数据中心中实施。

[0021] 但是,CPU和存储器技术的进步以及CPU密集型和I/O密集型工作负载的日益普及

已经造成数据中心性能现在变得严重依赖于温度和海拔的情况。这种依赖性会对现代数据中心配置影响最大,包括使用动态电压和频率缩放(DVFS)的激进CPU设计和/或使用自旋硬盘驱动器(HDD)而不是固态驱动器(SDD)的存储器存储。无论制造商如何,温度和海拔对性能的影响都适用于所有信息技术(IT)系统。温度和海拔也已被证明会同样影响内部部署数据中心和云数据中心以及内部HDD和外部存储布置,诸如网络附加存储(NAS)。此外,随着每一代新一代CPU和HDD对环境温度/海拔的依赖性呈指数增长,这个问题将在未来继续增长。

[0022] 图1图示了根据一些实施例的用于数据中心的冷却系统的简化图。大多数服务器和存储系统使用风扇100对内部组件进行气冷。此外,许多存储系统可以包括也配备冷却风扇的内部电源。这些风扇100可以配备可变速度以适应数据中心中的可变温度。随着数据中心环境温度的升高,风扇速度也会按比例增加以冷却内部组件。

[0023] 已发现现代HDD 102对即使是最轻微的振动也极为敏感。例如,机械和/或声学振动严重降级HDD 102的I/O速率。随着环境温度增加,并且风扇100的速度增加以进行补偿,来自风扇100的机械振动也增加。这些机械振动可能会干扰HDD 102的读/写操作,从而降级HDD 102的性能。此外,来自风扇的声学噪声也会随着风扇速度的增加而增加。这种声学噪声也影响HDD并进一步降级I/O性能。因此,随着数据中心在更高的温度下操作,风扇速度增加所产生的机械和声学振动可测量地降低HDD 102的I/O性能。这种源自更高温度的影响在未来可能会增加,因为数据中心在世界目前正在将环境温度提高到30℃至35℃以上,以试图节省用于空调的能源。但是,在较高温度下对应的性能降级增加执行客户工作负载所需的时间,这进而抵消增加操作温度带来的潜在能源节约。

[0024] 除了增加温度的影响之外,还已经发现海拔的改变会造成HDD102的性能进一步降级。例如,高海拔数据中心将位于空气比位于海平面的数据中心时更薄的区域中。这些较高海拔处越稀薄的空气具有越少的冷却能力。这要求风扇100进一步增加其速度以增加通过服务器的气流。由于越暖和的温度,这种增加的风扇速度可以被添加到上述增加中,并且会进一步生成破坏HDD 102性能的机械和/或声学振动。例如,与海平面处的类似数据中心配置相比,风扇在高海拔环境中的运行速度通常快13%(例如,Denver Colorado)。

[0025] 图1还图示了可以是数据中心配置的一部分的CPU。除了影响各种存储器组件的I/O性能外,温度的增加还会对CPU 106在使用DVFS时的性能产生不利影响。使用DVFS的系统一般调整关于CPU和其它外围设备的功率和/或速度设置,以优化资源消耗。DVFS允许CPU 106使用工作负载所需的最小功率量来执行CPU操作负载,以最大化功率节省并改善CPU 106的寿命。系统监视工作负载并动态调整操作电压和/或时钟频率以匹配CPU的性能与工作负载的要求。

[0026] 现代CPU和图形处理单元(GPU)使用非常灵敏和激进的DVFS方案来连续地节流频率和/或操作电压。但是,较高的温度一般在CPU 106中造成更多的泄漏电流,因此DVFS频率节流随着温度的每一度增加而呈指数增加。例如,随着数据中心的海拔从海平面增加到10,000英尺,从15℃到35℃的增加示出CPU性能降级了50%以上。由于温度和海拔而引起的CPU和I/O性能降级的这种组合为安装在新操作环境中的任何数据中心带来了巨大的挑战。

[0027] 以前,用于配置数据中心的过程是迭代过程,其涉及多个试错周期以调整配置的各个方面以满足用于数据中心的SLA要求。当安装数据中心时,使用默认配置,然后评估数据中心的性能以确定默认配置是否满足SLA要求。与正在安装的数据中心相比,默认配置常

常是在操作温度和/或海拔不同的环境中创建的。由于上述较高温度和/或海拔的影响，新数据中心的性能常常与最初为其设计默认配置的数据中心的性能不匹配。如果新数据中心的性能不令人满意，那么调整数据中心配置中的各种参数，并为数据中心实现新配置并对照SLA进行测试。这个过程将重复进行，直到识别出令人满意的数据中心配置。这要求多次迭代，这对数据中心的设置时间添加了显著的延迟。

[0028] 除了初始配置新数据中心所需的多次迭代之外，安装之后和操作期间温度和海拔的改变通常会使得最初令人满意的配置在I/O密集型或CPU密集型工作负载期间无法满足SLA要求。即使默认配置最终针对新数据中心进行了优化，但出于上述原因，新数据中心中温度的增加也会降级CPU 106的性能和HDD 102的性能。这种性能的降级对客户造成超时错误并增加了完成客户工作负载所需的时间，因此在未满足SLA要求时导致客户不满意并导致事件升级。诊断这种性能降级的根本原因是个非常困难的过程，特别是在由温度和/或海拔差异的波动造成时。为了在配置新数据中心之后解决这种情况，数据中心管理员常常必须提供全新的配置。然后新配置常常会面临与数据中心的原始配置中观察到的相同的超时错误、性能降级和一致性挑战。

[0029] 本文描述的实施例通过使用自动化配置优化框架以在给定任何客户用例场景可能需要的存储器、CPU和I/O性能的工作负载的情况下确定最优数据中心配置来解决这些和其它技术问题。这些实施例可以利用环境室来控制各种不同数据中心配置的环境特性。例如，热/海拔室可以被用于循环通过大范围的温度和/或海拔，使得可以对照环境特性的不同组合来测量数据中心配置的性能。由于在不同的温度和海拔组合下测试了各种数据中心配置，因此可以训练与每个海拔/温度组合对应的模型的库。在已经训练模型的库之后，这个库可以被用于在新数据中心被安装和配置时确定它们的最优配置。已安装的数据中心的环境特性可以被用于在库中选择最近邻模型。然后可以将类似的操作要求提供给模型以预测各种性能度量，诸如CPU性能、存储器I/O性能等。如果由模型生成的性能度量优于已安装的数据中心的性能，那么可以使用由该模型提供的配置来配置数据中心。这可以要求为数据中心安装和测试最多两种不同的配置，以确保已对其进行了最优配置。

[0030] 本文描述的实施例与任何类型的数据中心兼容。例如，一些数据中心可以包括由客户私有和/或控制的内部部署数据中心，这些数据中心通常安装在客户的位置。一些数据中心还可以包括基于云的数据中心，包括由服务提供商托管并可供客户使用的位于中心的硬件/软件。一些数据中心可以包括内部存储，其中盘驱动器位于服务器内部。其它数据中心可以包括通信地耦合到服务器的外部数据存储设备。基于对象的存储设备也可以被一些数据中心用来将数据作为对象进行管理。本文描述的配置方法与这些类型的数据中心或存储技术中的任何一种兼容。

[0031] 图2图示了根据一些实施例的对象存储系统200的简化图。这个对象存储系统200仅作为可以使用本文描述的自主配置优化技术的一类数据中心的一个示例提供。但是，这个对象存储系统200并不意味着限制，并且这些技术可以与任何类型的数据中心一起使用。本领域普通技术人员将认识到，这些技术不受数据中心类型的限制，而是可以根据本公开应用于任何数据中心。

[0032] 客户可以使用存储云服务作为存储解决方案以将他们的数据/文件存储在公共云上。对象存储系统200可以包括内部部署服务器204，其包括存储在文件目录206中的客户数

据。服务器204可以通信地耦合到作为使用对象存储系统200的数据中心的一部分的附加服务器实例202a、220b。服务器204可以包括安装在客户站点的云存储软件设备(CSSA)。CSSA可以充当云存储网关,其将内部部署应用和工作流连接到公共云。例如,CSSA可以使用Oracle Cloud Infrastructure Storage Software Appliance®来实现。CSSA可以与公共云中的存储云服务212通信,其中可以存储客户的对象数据。CSSA可以管理客户的数据和/或应用与公共云存储设备之间的I/O业务。

[0033] CSSA的有效配置对于实现最大存储I/O吞吐量和/或计算性能可以是至关重要的。CSSA的任何次优配置都会导致CSSA的功能性不令人满意,其特征在于吞吐量或计算性能不佳。这会导致客户可立即识别出的“超时”故障。由于CSSA的配置引起的性能差会导致违反由云服务提供的SLA要求。这对于CPU密集型和I/O密集型工作负载尤其如此。

[0034] CSSA可以包括一个或多个配置参数210。这些参数应当被配置为满足客户的性能规范。这些性能规范常常被表达为基于并发用户的数量、所存储文件的数量和/或尺寸、目录层次结构、I/O操作的数量(读、写、追加等)等的I/O吐吞量需求。满足这些性能规范的数据中心的性能可以由CSSA的可配置参数210支配。这些是可以被调谐以解决客户性能需求的参数。配置参数可以包括定义数据中心的设置和/或操作的任何硬件和/或软件参数。配置参数可以包括数据中心中的CPU的数量;数据中心中的CPU的速度和/或类型;数据中心中使用的HDD的数量;数据中心中使用的硬件的类型,包括网卡、盘控制器和数据库;高速缓存尺寸;和/或任何其它参数。配置参数还可以包括被用于控制硬件操作的软件设置。如本文所使用的,配置或配置参数可以包括本领域普通技术人员将理解的数据中心的任何方面。

[0035] 与任何数据中心一样,对象存储系统200可以提供许多不同的配置,这些配置可以在安装和操作期间被优化。客户工作负载混合与CSSA之间的交互常常非常复杂,并且推荐最优配置解决方案可以是实现令人满意的性能的关键。上述配置的任何次优组合都可能导致长时延、低吞吐量、多个I/O超时和/或SLA违反。如上所述,随着性能对海拔和温度的依赖性越来越大,为对象存储系统200或任何其它数据中心推荐最优配置变得越来越困难。

[0036] 为了快速且高效地为数据中心提供最优配置,本文描述的实施例可以使用两阶段过程。在第一阶段,可以针对温度和海拔的许多不同组合下的各种配置训练模型的库。然后可以将这个预训练的模型的库提供给在安装或重新配置新的或现有数据中心时使用的第二阶段。用于训练模型的库的第一阶段可以被离线执行,而不考虑用于客户的特定数据中心安装。代替地,可以使用环境室来训练模型,该环境室循环通过针对不同数据中心配置的温度和海拔的各种组合。这个过程可以在没有为客户安装任何特定数据中心的压力的情况下执行,因此可以随着时间的推移开发非常全面和稳健的经训练的模型的库。在第二阶段期间,可以访问并使用这个库来实时推荐用于给定数据中心的最优配置,而不会出现显著延迟。

[0037] 图3图示了根据一些实施例的可以被用于在各种温度和/或海拔下训练模型的库的系统。该系统可以包括环境室300,该环境室300被配置为调整和控制用于被评估的数据中心配置的操作环境的一个或多个环境特性。许多不同的环境条件可以由环境室300监视和/或控制,诸如温度、海拔、湿度、压力等。作为示例,本文描述的实施例可以使用被配置为控制腔室300内部的环境温度并模拟腔室300中的各种高度的热-海拔室。出于上面讨论的性能原因,温度和海拔是环境特性,其可以是与数据中心性能最相关的,但它们并不意味着

限制。当训练、监视和使用用于数据中心的模型时，其它实施例可以使用其它环境特性。

[0038] 不同数据中心的各种配置可以安装在环境室300内。例如，图3中描绘的环境室300中的数据中心的配置300可以包括多个服务器、HDD和/或可以用于构建数据中心的其它硬件。配置300还可以包括影响或控制数据中心的操作的各种配置参数。在一些实施例中，数据中心配置的类型可以被划分为少量简化的类别，诸如低、中和/或高性能数据中心。例如，第一(例如，低性能)类别的数据中心可以包括有限数量的CPU和/或HDD。第二(例如，中等性能)类别的数据中心可以包括更大数量的多核CPU和/或GPU，以及更大数量的HDD。第三(例如，高性能)类别的数据中心可以被称为“工程化的系统”，并且可以使用最快的CPU/GPU并提供数PB(千万亿字节)的存储器。在这些类别的数据中心的每一个中，附加的配置参数会更改数据中心中使用的硬件组件和/或软件的数量。

[0039] 当数据中心的配置300安装在环境室300中时，控制器302可以向数据中心的配置310提供多变量工作负载脚本304。数据中心然后可以在环境室300中执行多变量工作负载脚本304。如下所述，工作负载脚本304可以在诸如温度和海拔之类的环境特性的每个不同组合下执行。多变量工作负载脚本304可以包括使得CPU工作负载从空闲变为最大负载的极端压力工作负载，从而随着时间正弦循环以完全评估CPU性能的范围。此外，多变量工作负载脚本304可以被配置为类似地将HDD的操作从空闲循环到最大负载。多变量工作负载脚本304可以被设计为覆盖数据中心在各种客户安装处可以经历的不同工作负载的范围。

[0040] 控制器302可以包括循环算法308，其使环境室300循环通过不同环境特性的各种置换。在这个示例中，循环算法可以在所选择的模拟海拔增量处循环通过温度范围。在一些实施例中，这个过程可以在多个小时的时间间隔内(例如，超过24小时)执行。

[0041] 在温度和海拔的每个不同组合下，多变量工作负载脚本304可以由环境室300中的数据中心执行，并且来自配置的遥测数据可以由控制器302收集。在一些实施例中，当数据中心执行多变量工作负载脚本304时，可以使用诸如Oracle的连续系统遥测线束(CSTH)之类的遥测线束来收集实时遥测数据。遥测数据可以包括数据中心在操作期间的任何电气和/或环境特性，包括分布式温度、环境温度、用电量等。遥测数据还可以产生用于数据中心的性能度量。例如，性能度量可以包括配置310中用于每个CPU的性能度量。对于每个CPU，遥测线束312可以捕获指示用于那个CPU的最大性能百分比的数据。(这也可以指示节流发生的量的倒数。)对于每个HDD，可以生成表示被记录为最大I/O速率的百分比的I/O速率的性能度量。用于HDD的性能度量还可以包括I/O吞吐量的Mb/s数量的指示。这些性能度量可以被用于确定数据中心的配置310是否在温度和海拔的每个组合下满足客户的SLA要求。

[0042] 由遥测线束312捕获的遥测数据可以提供给模型训练过程306，模型训练过程306使用遥测数据来训练特定于当前温度和海拔组合的模型。模型训练过程306可以加载与温度和/或海拔的每个增量对应的模型。当数据中心在那个温度和海拔组合下执行多变量工作负载脚本304时接收的遥测数据可以被提供给控制器302，使得每个温度/海拔组合产生用于训练对应模型的整套遥测信号。在已经执行了通过每个温度和海拔增量的完整循环之后，控制器302可以具有在温度和海拔的所有组合上的完整遥测时间序列信号库。

[0043] 图4图示了根据一些实施例的用于控制环境室的时间和/或高度增量的循环算法308的流程图。该算法可以通过将环境室中的模拟海拔初始化为初始值来开始(402)。贯穿本公开，各种海拔的数学表示—模拟的和测得的—都可以使用变量H来引用。初始海拔(H_L)

可以被设置为任何值。在一些实施例中，初始海拔可以被设置在海平面处(即，海拔0英尺)。该算法然后可以确定当前模拟海拔H是否小于最大模拟海拔(404)。最大海拔(H_H)也可以被设置为任何值，包括5000英尺、7000英尺、10000英尺等。

[0044] 接下来，算法可以将环境室中的温度初始化为初始值(410)。贯穿本公开，可以使用变量T来引用温度的数学表示。初始温度(T_L)可以被设置为任何值。在一些实施例中，初始温度可以被设置为0℃、5℃、15℃、20℃等。该算法然后可以确定当前温度T是否小于最大温度(412)。最高温度(T_H)可以被设置为任何值，诸如15℃、20℃、25℃、30℃、35℃、40℃等。

[0045] 该算法然后可以使数据中心在指定的温度和海拔组合下运行多变量工作负载脚本(414)。在一些实施例中，算法可以等待环境室中的海拔和/或温度随着这些值被初始设置或递增而稳定。由于多变量工作负载脚本在指定的海拔和温度下执行，因此该算法可以在那个海拔和温度下收集整套遥测信号，以用于训练对应的模型(418)，如下文更详细描述的。

[0046] 在脚本已经被执行并且遥测信号已经在当前温度/海拔组合下被捕获之后，算法接下来可以递增温度(416)。温度可以被递增任何值，诸如0.5℃、1.0℃、2.0℃、2.5℃、5.0℃等。该算法然后可以确定递增的温度是否大于应当为其测试这个数据中心配置的最大温度。如果尚未达到最大温度，那么该算法可以再次运行多变量工作负载脚本(414)并在新的温度/高度组合下收集整套遥测信号(418)。这个循环可以继续在当前海拔下的温度范围内收集遥测信号，直到达到最高温度。

[0047] 在达到最高温度(412)之后，算法然后可以增加由环境室模拟的海拔(406)。海拔可以递增任何值，诸如100英尺、200英尺、250英尺、500英尺、1000英尺等。在海拔被递增之后，算法可以确定新海拔是否大于要为其评估这个数据中心配置的最大海拔(404)。如果当前海拔仍然小于最大海拔，那么算法可以如上所述在新海拔下的整个温度范围内递增。但是，当达到最大海拔时，用于特定数据中心配置的温度/海拔特性可以完成。在这个时候，系统可以提供在温度和海拔的所有组合的整套遥测时间序列信号库，用于在那些组合中每一个下训练模型。

[0048] 图5图示了根据一些实施例的用于在每个温度/海拔组合下训练模型的过程。在执行多元脚本期间在每个时间/海拔组合下收集的遥测数据的集合502可以被提供给过程306以训练模型。不同的实施例可以使用不同的模型。例如，一些实施例可以使用模式识别模型，诸如多变量状态估计技术(MSET)建模技术、MSET2建模技术或任何其它非线性、非参数建模方法。可以使用任何遥测数据来训练模型，以学习影响数据中心的性能的所有配置参数之间的相关模式。例如，MSET2(其已用于其它应用中的预后和网络安全性用例)可以学习独立变量(诸如CPU性能和I/O吞吐量)的交互，作为用于特定数据中心配置的环境条件的函数。此外，在每个温度/海拔组合下，可以测试多种配置来训练模型。例如，虽然图3中所示的过程表征单一配置，但实际上可以表征多种配置，其中每种配置通过图4中所示的温度/海拔组合的全范围提供遥测信号。

[0049] 用于训练模型库的过程可以是正在进行的累积计算，该累积计算用从各个数据中心累积的数据离线执行。虽然上述示例依赖于从受控表征环境中的环境室提供的遥测数据，但这并不意味着限制。其它实施例可以使用从其它数据中心环境(包括客户使用的实况数据中心安装)累积的遥测数据。这种离线训练过程的结果是预训练的模型的库，其可以被

用于生成性能度量，诸如在给定温度/海拔组合下用于不同配置的服务质量(QoS)度量。预训练的模型的库也可以被用于生成推荐的配置。

[0050] 自主配置过程的第二阶段可以包括实时推荐。当为客户或云环境安装新的数据中心时，模式识别算法可以从训练模式过渡到实时“推荐”模式。例如，可以选择最接近数据中心的测得的温度/海拔组合的特定MSET模型506，并且模型506可以被用于输出应当被用于基于所有其它相关变量的给定环境条件的新配置。例如，可以为预测方法提供用于新数据中心的数据中心要求的集合508。遥测输入可以被提供给MSET模型506以生成性能度量，诸如在给定温度/海拔组合下的QoS度量。然后可以将这些QoS度量与数据中心的当前配置的测得的QoS度量进行比较，并且如果当前配置不是最优的，那么可以为新配置提供推荐510。

[0051] 这种预后机器学习技术可以使用在各种温度/海拔下用数据中心配置训练的模型。当客户用例作为查询被提供给这种方法时，该方法可以使用预训练的MSET模型的库来基于客户的工作负载要求和在新数据中心安装处测得的环境条件为客户推荐最优数据中心配置选项。在一些实施例中，安装数据中心的技术人员可以使用智能电话、平板计算机或其它计算设备上的应用或“app”来访问预训练的模型的库。app可以从数据中心接收性能数据，包括测得的温度/海拔，并且可以在模型库中识别最匹配的模型。一旦模型被识别出，MSET就可以被用于输出最低成本配置，该配置提供最优性能度量以满足SLA要求。一些实施例还可以提供给定数据中心的海拔和环境温度可以实现的最大性能。相反，一些实施例可以在给定数据中心工作负载和SLA要求的情况下提供推荐的环境条件的集合。

[0052] 图6图示了根据一些实施例的可以从模型训练过程306生成的模型库602。如上所述，模型库602可以包括由图4中的温度/海拔循环算法生成的用于温度和海拔的每个组合的条目。可以在图3中所示的环境室中使用多种不同的数据中心配置在每个温度/海拔组合下训练模型506。填充模型库602的模型506的数量可以取决于由图4中的循环算法生成的温度/海拔的范围。模型506的数量也可以取决于数据中心每次执行工作负载脚本之间使用的温度/海拔增量。温度/海拔增量越小，模型库602可以越细粒度。这增加了新数据中心安装可以与模型库602中预训练的模型紧密匹配的可能性。

[0053] 应当强调的是，用于配置新数据中心安装的过程的这个第二阶段可以在几小时甚至几分钟内执行。因为模型库之前已经填充了预训练的模型，所以可以将来自新数据中心安装的模型输入快速提供给现有模型以生成对应的性能度量，该度量可以与新数据中心的当前配置的性能度量进行比较。如果最接近的模型生成比当前配置更好的QoS结果，那么由模型提供的配置可以用来代替新数据中心安装的当前配置。这可以与以前用于配置数据中心的过程进行比较。一般而言，没有用于确定当前配置是否最优的系统化、参数化方法。代替地，将调整用于数据中心的配置的各种参数，直到可以满足SLA要求。这常常是个反复试错的过程，需要多次迭代，并且可以花费数天或数周。本文描述的这种自主数据中心配置过程通过将需要针对SLA要求进行测试的不同配置的数量限制为最多两个，从而显著减少了这个时间。

[0054] 图7图示了根据一些实施例的用于使用在环境特性的各种组合下预训练的模型库来自主地确定数据中心的最优配置的方法的流程图700。该方法可以包括从数据中心接收或测量第一性能度量(702)。数据中心可以包括新数据中心安装或其配置正在被更新的先前安装的数据中心。例如，现有数据中心可能在更高的工作负载或变化的温度下开始无法

满足其SLA要求，并且这种方法可以被执行以便以更优化的方式重新配置数据中心。数据中心可以具有在本文称为“第一配置”的现有配置，以将这种配置与其它配置区分开来。后续配置可以被称为“第二配置”仅仅是为了将其与第一配置区分开来。术语第一/第二并不意味着指示配置之间的次序、优先级、重要性或任何其它比较质量。类似地，“第一性能度量”可以这样称呼以区分由数据中心生成的性能度量与由预训练的模型中的一个生成的“第二”性能度量。第一性能度量可以包括任何QoS度量，诸如作为最大性能百分比的CPU性能、I/O吞吐量、I/O速率、时延等。可以在数据中心执行诸如上述多变量工作负载脚本之类的工作负载时实时地测量和/或生成第一性能度量。

[0055] 可选地，该方法可以包括确定第一性能度量是否满足SLA要求(704)。例如，可以将超时错误的数量与SLA要求进行比较。测得的带宽或时延可以与SLA要求进行比较。CPU性能或功耗可以与SLA要求进行比较。在一些情况下，如果第一性能度量满足或超过SLA要求，那么该方法可以简单地确定数据中心的当前配置是足够的，并且可以使用当前配置来操作数据中心(716)。注意的是，一般会有多个SLA要求和多个性能度量与这些SLA要求进行比较。这种方法仅列举单个性能指标和SLA要求作为示例，但应当理解的是，其它实施例可以要求每个性能度量满足对应的SLA要求，然后才允许数据中心在当前配置中继续操作而无需任何调整。

[0056] 该方法还可以包括在数据中心执行工作负载时接收或测量用于数据中心的一个或多个环境特性(706)。如上文关于图3描述的相似遥测系统可以被用于数据中心以在多变量工作负载脚本的执行期间收集遥测数据。遥测数据可以包括环境条件，诸如温度和/或海拔。

[0057] 该方法还可以包括在模型库中识别具有相似环境特性的预训练的模型(708)。该模型可以预先使用来自在受控环境(诸如上面关于图3描述的环境室)中操作的一个或多个数据中心配置的数据进行训练。该方法可以接收环境特性并从在相似的温度/海拔组合下训练的模型库中自主地选择对应的模型。在一些情况下，数据中心的当前环境条件可以与用于预训练的模型的环境条件相匹配。例如，对于处于海平面并在25℃下操作的数据中心，模型库可以包括在那些对应的环境条件下训练的对应模型。但是，情况常常可以是数据中心的当前操作条件可以介于模型库中可用的温度/海拔组合之间。当精确匹配不可用时，该方法可以选择具有与数据中心的环境特性最相似的环境特性的模型。例如，一些实施例可以使用最近邻算法来选择预训练的模型，如下面关于图8详细描述的。

[0058] 该方法还可以包括为预训练的模型生成性能度量(710)。这个性能度量在本文中可以被称为“第二性能度量”以将其与上述第一性能度量区分开来。这个性能度量可以是与第一性能度量相同的类型，使得可以比较它们。这个性能度量可以包括本文描述的任何QoS度量。再次，这种方法可以包括生成多个性能度量，每个都可以与来自数据中心配置的性能度量进行比较，并且可以要求每个都通过SLA要求。可以针对所考虑的每个性能度量重复该方法的这个步骤。

[0059] 该方法还可以包括将来自数据中心配置的性能度量与由预训练的模型生成的性能度量进行比较(712)。如果来自数据中心的性能度量优于来自模型的性能度量，那么可以认为数据中心的配置是最优的，并且数据中心可以继续以现有配置操作(716)。但是，如果来自预训练的模型的性能度量优于来自数据中心配置的性能度量，那么该方法可以包括根

据模型确定用于数据中心的新配置并将来自模型的配置用于已安装的数据中心(714)。因为该模型是使用各种不同的配置参数进行训练的,所以该模型可以针对那个海拔/温度组合输出数据中心配置,其将生成最佳QoS结果。

[0060] 如果模型生成了更优化的数据中心配置,那么可以在数据中心中实现新的数据中心配置(716)。然后可以测试新配置以确保新配置满足SLA要求,并且可以完成配置过程。注意的是,可能只需要测试两种不同的配置以确定最优配置。如果第一配置优于由对应的预训练的模型生成的配置,那么可以认为第一配置是最优的。但是,如果由预训练的模型生成的第二配置生成更好的QoS度量,那么可以认为第二配置是最优的,无需任何进一步的配置测试。

[0061] 应当认识到的是,图7中所示的特定步骤提供了根据各种实施例优化数据中心的配置的特定方法。根据替代实施例,也可以执行其它步骤序列。例如,本发明的替代实施例可以以不同的次序执行上述步骤。而且,图7中所示的各个步骤可以包括多个子步骤,这些子步骤可以以适于各个步骤的各种次序执行。此外,可以根据特定应用添加或移除附加的步骤。本领域的普通技术人员将认识到许多变化、修改和替代。

[0062] 图8图示了根据一些实施例的用于选择具有与数据中心的环境特性相似的环境特性的预训练的模型的方法的流程图800。如上所述,数据中心的环境特性与预训练的模型的环境特性之间的确切匹配可能并不总是可用的。取决于生成模型库时使用的温度/海拔增量,一些数据中心可以在与预训练的模型对应的温度/海拔之间的温度/海拔下操作。因此,一些实施例可以通过识别具有特性的最相似环境的模型来选择模型。这可以包括使用最近邻算法来识别模型库中最接近的温度/海拔匹配。流程图800中的方法说明了识别最近邻模型的一个示例方法。

[0063] 该方法可以包括接收所安装的数据中心的环境温度和海拔(802)。温度/海拔可以作为在测试如上所述配置的第一数据中心时使用的遥测数据的一部分被接收。该方法还可以包括访问用于生成模型库的温度和海拔(804)。例如,M个温度值的序列可以与N个海拔值的序列组合以索引模型库中的条目。该方法然后可以包括计算库中M个温度值和N个海拔值的每个组合与数据中心的环境温度和海拔之间的距离,诸如欧几里得距离(806)。在一些实施例中,该方法可以首先计算并最小化海拔之间的距离,然后最小化环境温度之间的距离。然后该方法可以识别库中与数据中心的温度/海拔具有最小距离的温度/海拔对(808)。可以使用以下等式计算这个距离的最小化:

[0064] $$\min \sum_{M} \sum_{N} \sqrt{(T_i - T_0)^2 + \left(H_j - H_0\right)^2}$$

[0065] 在这个等式中,T_0和H_0表示数据中心的环境温度和海拔,而T_i和H_j表示模型库中的温度和海拔值的范围。一旦已经识别出最小化这个等式的T_i和H_j的值,这些值就可以被用于索引模型库并检索要使用的对应最近邻模型。

[0066] 本文公开的实施例可以使用各种计算环境,包括联网的计算环境、云服务环境,以及包括处理器、存储器、指令集等的计算机体系架构。以下附图描述可以被用于实现本文描述的各种实施例的硬件/软件系统。这些硬件/软件系统也可以被用于与本文描述的实施例交互。

[0067] 此外，这些方法的每个步骤可以由计算机系统自动执行，和/或可以提供有涉及用户的输入/输出。例如，用户可以为方法中的每个步骤提供输入，并且这些输入中的每一个可以响应于请求这种输入的具体输出，其中输出是由计算机系统生成的。可以响应于对应的请求输出而接收每个输入。此外，输入可以从用户接收、作为数据流从另一个计算机系统接收、从存储器位置检索、通过网络检索、从web服务请求等。同样，可以将输出提供给用户、作为数据流提供给另一个计算机系统、保存在存储器位置中、通过网络发送、提供给web服务等。简而言之，本文描述的方法的每个步骤都可以由计算机系统执行，并且可以涉及与计算机系统之间可能涉及或可能不涉及用户的任何数量的输入、输出和/或请求。可以说不涉及用户的那些步骤是由计算机系统自动执行的，而无需人工干预。因此，根据本公开将理解的是，本文描述的每种方法的每个步骤可以被更改为包括去往和来自用户的输入和输出，或者可以由计算机系统自动地完成而无需人为干预，其中任何确定都是由处理器做出的。此外，本文描述的每个方法的一些实施例可以被实现为存储在有形的，非暂态的存储介质上以形成有形的软件产品的指令集。

[0068] 图9描绘了用于实现实施例之一的分布式系统900的简化图。在所示的实施例中，分布式系统900包括一个或多个客户端计算设备902、904、906和908，每个客户端计算设备通过一个或多个网络910执行和操作客户端应用，诸如web浏览器、专有客户端(例如，Oracle Forms)等。服务器912可以经由网络910与远程客户端计算设备902、904、906和908通信地耦合。

[0069] 在各种实施例中，服务器912可以适于运行由系统的一个或多个组件提供的一个或多个服务或软件应用。在一些实施例中，这些服务可以作为基于web的服务或云服务或者在软件即服务(SaaS)模型下被供应给客户端计算设备902、904、906和/或908的用户。操作客户端计算设备902、904、906和/或908的用户进而可以利用一个或多个客户端应用来与服务器912交互以利用由这些组件提供的服务。

[0070] 在图中描绘的配置中，系统900的软件组件918、920和922被示出为在服务器912上实现。在其它实施例中，系统900的一个或多个组件和/或由这些组件提供的服务也可以由客户端计算设备902、904、906和/或908中的一个或多个来实现。然后，操作客户端计算设备的用户可以利用一个或多个客户端应用来使用由这些组件提供的服务。这些组件可以用硬件、固件、软件或其组合来实现。应该认识到的是，各种不同的系统配置是可能的，这些配置可能与分布式系统900不同。图中所示的实施例因此是用于实现实施例系统的分布式系统的一个示例，而不旨在是限制性的。

[0071] 客户端计算设备902、904、906和/或908可以是便携式手持设备(例如，**iPhone®**、蜂窝电话、**iPad®**、计算平板电脑、个人数字助理(PDA))或可穿戴设备(例如，Google **Glass®**头戴式显示器)，运行诸如Microsoft Windows **Mobile®**和/或各种移动操作系统(诸如iOS、Windows Phone、Android、BlackBerry 10、Palm OS等)的软件，并且启用互联网、电子邮件、短消息服务(SMS)、**Blackberry®**或其它通信协议。客户端计算设备可以是通用个人计算机，作为示例，包括运行各种版本的Microsoft **Windows®**的个人计算机和/或膝上型计算机。客户端计算设备可以是运行各种可商业获得的**UNIX®**或类UNIX操作系统(包括但不限于各种GNU/Linux操作系统，诸如例如Google Chrome OS)

中的任何操作系统的工作站计算机。替代地或附加地，客户端计算设备902、904、906和908可以是能够通过(一个或多个)网络910进行通信的任何其它电子设备，诸如瘦客户端计算机、启用互联网的游戏系统(例如，具有或不具有Kinect®手势输入设备的Microsoft Xbox游戏控制台)和/或个人消息传送设备。

[0072] 虽然示出了具有四个客户端计算设备的示例性分布式系统900，但是可以支持任何数量的客户端计算设备。其它设备(诸如具有传感器等的设备等)可以与服务器912交互。

[0073] 分布式系统900中的(一个或多个)网络910可以是本领域技术人员熟悉的、可以使用各种可商业获得的协议中的任何协议来支持数据通信的任何类型的网络，其中协议包括但不限于TCP/IP(传输控制协议/互联网协议)、SNA(系统网络体系架构)、IPX(互联网分组交换)、AppleTalk等。仅仅作为示例，(一个或多个)网络910可以是局域网(LAN)，诸如基于以太网、令牌环等的LAN。(一个或多个)网络910可以是广域网和互联网。它可以包括虚拟网络，包括但不限于虚拟专用网络(VPN)、内联网、外联网、公共交换电话网(PSTN)、红外网络、无线网络(例如，依据电气和电子协会(IEEE)802.11协议套件、蓝牙®和/或任何其它无线协议的任何协议而操作的网络)；和/或这些网络和/或其它网络的任何组合。

[0074] 服务器912可以由一个或多个通用计算机、专用服务器计算机(作为示例，包括PC(个人计算机)服务器、UNIX®服务器、中档服务器、大型计算机、机架安装的服务器等)、服务器场、服务器集群或任何其它适当的布置和/或组合组成。在各种实施例中，服务器912可以适于运行在前述公开中描述的一个或多个服务或软件应用。例如，服务器912可以与用于执行以上根据本公开的实施例描述的处理的服务器对应。

[0075] 服务器912可以运行包括以上讨论的任何操作系统中的操作系统，以及任何可商业获得的服务器操作系统。服务器912还可以运行各种附加的服务器应用和/或中间层应用中的任何一种，包括HTTP(超文本传输协议)服务器、FTP(文件传输协议)服务器、CGI(公共网关接口)服务器、JAVA®服务器、数据库服务器等。示例性数据库服务器包括但不限于可从Oracle、Microsoft、Sybase、IBM(国际商业机器)等商业获得的那些数据库服务器。

[0076] 在一些实施方式中，服务器912可以包括一个或多个应用，以分析和整合从客户端计算设备902、904、906和908的用户接收到的数据馈送和/或事件更新。作为示例，数据馈送和/或事件更新可以包括但不限于，Twitter®馈送、Facebook®更新或者从一个或多个第三方信息源和连续数据流接收到的实时更新，实时更新可以包括与传感器数据应用、金融价格收报机(financial ticker)、网络性能测量工具(例如，网络监视和流量管理应用)、点击流分析工具、汽车交通监视等相关的实时事件。服务器912还可以包括一个或多个应用，以经由客户端计算设备902、904、906和908的一个或多个显示设备来显示数据馈送和/或实时事件。

[0077] 分布式系统900还可以包括一个或多个数据库914和916。数据库914和916可以驻留在各种位置。作为示例，数据库914和916中的一个或多个可以驻留在服务器912本地(和/或驻留在服务器912中)的非暂态存储介质上。可替代地，数据库914和916可以远离服务器912并且经由基于网络的连接或专用的连接与服务器912进行通信。在实施例的一个集合中，数据库914和916可以驻留在存储区域网络(SAN)中。类似地，用于执行服务器912所具有的功能的任何必要文件都可以适当地本地存储在服务器912上和/或远程存储。在实施例的

一个集合中，数据库914和916可以包括适于响应SQL格式的命令而存储、更新和检索数据的关系数据库，诸如由Oracle提供的数据库。

[0078] 图10是根据一些实施例的可以将由实施例系统的一个或多个组件提供的服务作为云服务提供的系统环境1000的一个或多个组件的简化框图。在所示的实施例中，系统环境1000包括可以由用户使用以与提供云服务的云基础设施系统1002交互的一个或多个客户端计算设备1004、1006和1008。客户端计算设备可以被配置为操作客户端应用，诸如web浏览器、专有客户端应用(例如，Oracle Forms)或某种其它应用，这些应用可以由客户端计算设备的用户用来与云基础设施系统1002交互以使用由云基础设施系统1002提供的服务。

[0079] 应当认识到的是，图中描绘的云基础设施系统1002可以具有与所描绘的组件不同的其它组件。另外，图中所示的实施例仅仅是可以结合本发明的实施例的云基础设施系统的一个示例。在一些其它实施例中，云基础设施系统1002可以具有比图中所示更多或更少的组件，可以组合两个或更多个组件，或者可以具有组件的不同配置或布置。

[0080] 客户端计算设备1004、1006和1008可以是与上面针对902、904、906和908所描述的设备类似的设备。

[0081] 虽然示例性系统环境1000被示出具有三个客户端计算设备，但是任何数量的客户端计算设备可以被支持。其它设备(诸如具有传感器的设备等)可以与云基础设施系统1002交互。

[0082] (一个或多个)网络1010可以促进客户端1004、1006和1008与云基础设施系统1002之间的数据的通信和交换。每个网络可以是本领域技术人员所熟悉的可以使用各种商业上可获得的协议(包括上面针对(一个或多个)网络910所描述的那些协议)中的任何一种支持数据通信的任何类型的网络。

[0083] 云基础设施系统1002可以包括一个或多个计算机和/或服务器，其可以包括上面针对服务器912所描述的那些计算机和/或服务器。

[0084] 在某些实施例中，由云基础设施系统提供的服务可以包括按需对云基础设施系统的用户可用的许多服务，诸如在线数据存储和备份解决方案、基于Web的电子邮件服务、被托管的办公室(office)套件和文档协作服务、数据库处理、受管理的技术支持服务等。由云基础设施系统提供的服务可以动态扩展以满足云基础设施系统的用户的需要。由云基础设施系统提供的服务的具体实例化在本文中被称为“服务实例”。一般而言，从云服务提供商的系统经由通信网络(诸如互联网)对用户可用的任何服务被称为“云服务”。通常，在公共云环境中，构成云服务提供商的系统的服务器和系统与客户自己的室内服务器和系统不同。例如，云服务提供商的系统可以托管应用，并且用户可以经由诸如互联网的通信网络按需订购和使用应用。

[0085] 在一些示例中，计算机网络云基础设施中的服务可以包括对存储装置、被托管的数据库、被托管的Web服务器、软件应用或由云供应商向用户提供的其它服务的受保护的计算机网络访问，或者如本领域中另外已知的。例如，服务可以包括通过互联网对云上的远程存储装置进行密码保护的访问。作为另一个示例，服务可以包括基于Web服务的被托管的关系数据库和脚本语言中间件引擎，以供联网的开发人员私有使用。作为另一个示例，服务可以包括对在云供应商的网站上托管的电子邮件软件应用的访问。

[0086] 在某些实施例中，云基础设施系统1002可以包括以自助服务、基于订阅、弹性可扩

展、可靠、高度可用和安全的方式递送给客户的应用、中间件和数据库服务供应的套件。这种云基础设施系统的示例是由本受让人提供的Oracle公共云。

[0087] 在各种实施例中，云基础设施系统1002可以适于自动供给、管理和跟踪客户对云基础设施系统1002供应的服务的订阅。云基础设施系统1002可以经由不同的部署模型来提供云服务。例如，可以依据公共云模型提供服务，其中云基础设施系统1002被销售云服务的组织拥有(例如，被Oracle拥有)，并且服务对一般公众或不同行业的企业可用。作为另一个示例，可以依据私有云模型来提供服务，其中云基础设施系统1002仅针对单个组织操作，并且可以为该组织内的一个或多个实体提供服务。还可以依据社区云模型来提供云服务，其中云基础设施系统1002和由云基础设施系统1002提供的服务由相关社区中的若干组织共享。云服务还可以依据混合云模型被提供，该混合云模型是两个或更多个不同模型的组合。

[0088] 在一些实施例中，由云基础设施系统1002提供的服务可以包括在软件即服务(SaaS)类别、平台即服务(PaaS)类别、基础设施即服务(IaaS)类别或包括混合服务的其它服务类别下提供的一个或多个服务。客户经由订阅订单可以订购由云基础设施系统1002提供的一个或多个服务。云基础设施系统1002然后执行处理以提供客户的订阅订单中的服务。

[0089] 在一些实施例中，由云基础设施系统1002提供的服务可以包括但不限于应用服务、平台服务和基础设施服务。在一些示例中，应用服务可以由云基础设施系统经由SaaS平台提供。SaaS平台可以被配置为提供落入SaaS类别的云服务。例如，SaaS平台可以提供在集成开发和部署平台上构建和递送按需应用套件的能力。SaaS平台可以管理和控制用于提供SaaS服务的底层软件和基础设施。通过利用由SaaS平台提供的服务，客户可以利用在云基础设施系统上执行的应用。客户可以获取应用服务，而无需客户购买单独许可证和支持。可以提供各种不同的SaaS服务。示例包括但不限于为大型组织提供销售绩效管理、企业集成和商务灵活性的解决方案的服务。

[0090] 在一些实施例中，平台服务可以由云基础设施系统经由PaaS平台提供。PaaS平台可以被配置为提供落入PaaS类别的云服务。平台服务的示例可以包括但不限于使组织(诸如Oracle)能够在共享的公共体系架构上整合现有应用以及充分利用平台提供的共享服务来构建新应用的能力的服务。PaaS平台可以管理和控制用于提供PaaS服务的底层软件和基础设施。客户可以获取由云基础架构系统提供的PaaS服务，而无需客户购买单独的许可证和支持。平台服务的示例包括但不限于Oracle Java云服务(JCS)、Oracle数据库云服务(DBCS)等。

[0091] 通过利用由PaaS平台提供的服务，客户可以采用由云基础设施系统支持的编程语言和工具，并且还控制所部署的服务。在一些实施例中，由云基础设施系统提供的平台服务可以包括数据库云服务、中间件云服务(例如，Oracle融合中间件服务)和Java云服务。在一个实施例中，数据库云服务可以支持共享服务部署模型，该模型使得组织能够汇集数据库资源并且以数据库云的形式向客户供应数据库即服务。中间件云服务可以为客户提供开发和部署各种商务应用的平台，并且Java云服务可以为客户提供在云基础设施系统中部署Java应用的平台。

[0092] 各种不同的基础设施服务可以由云基础设施系统中的IaaS平台提供。基础设施服务促进底层计算资源(诸如存储装置、网络和其它基础计算资源)的管理和控制，以供客户

利用由SaaS平台和PaaS平台提供的服务。

[0093] 在某些实施例中，云基础设施系统1002还可以包括基础设施资源1030，用于向云基础设施系统的客户提供用于提供各种服务的资源。在一个实施例中，基础设施资源1030可以包括预先集成和优化的硬件(诸如服务器、存储装置和联网资源)的组合，以执行由PaaS平台和SaaS平台提供的服务。

[0094] 在一些实施例中，云基础设施系统1002中的资源可以由多个用户共享并且根据需要动态重新分配。此外，可以将资源分配给在不同时区的用户。例如，云基础设施系统1030可以使在第一时区中的第一组用户能够在指定的小时数内利用云基础设施系统的资源，并且然后使相同资源能够被重新分配给位于不同时区的另一组用户，从而使资源的利用率最大化。

[0095] 在某些实施例中，可以提供由云基础设施系统1002的不同组件或模块以及由云基础设施系统1002提供的服务共享的多个内部共享服务1032。这些内部共享服务可以包括但不限于：安全和身份服务、集成服务、企业储存库服务、企业管理器服务、病毒扫描和白名单服务、高可用性、备份和恢复服务、启用云支持的服务、电子邮件服务、通知服务、文件传输服务等。

[0096] 在某些实施例中，云基础设施系统1002可以提供云基础设施系统中的云服务(例如，SaaS、PaaS和IaaS服务)的综合管理。在一个实施例中，云管理功能可以包括用于供给、管理和跟踪由云基础设施系统1002接收到的客户订阅等的能力。

[0097] 在一个实施例中，如图中所描绘的，云管理功能可以由一个或多个模块提供，诸如订单管理模块1020、订单编排模块1022、订单供给模块1024、订单管理和监视模块1026，以及身份管理模块1028。这些模块可以包括一个或多个计算机和/或服务器或者使用一个或多个计算机和/或服务器来提供，这些计算机和/或服务器可以是通用计算机、专用服务器计算机、服务器场、服务器集群或任何其它适当的布置/或组合。

[0098] 在示例性操作1034中，使用客户端设备(诸如客户端设备1004、1006或1008)的客户可以通过请求由云基础设施系统1002提供的一个或多个服务并且下订阅由云基础设施系统1002供应的一个或多个服务来的订单来与云基础设施系统1002交互。在某些示例中，客户可以访问云用户界面(UI)、云UI 1012、云UI 1014和/或云UI 1016并经由这些UI下订阅订单。云基础设施系统1002响应于客户下订单而接收到的订单信息可以包括识别客户以及客户想要订阅的云基础设施系统1002供应的一个或多个服务的信息。

[0099] 在客户下订单之后，经由云UI，1012、1014和/或1016，接收订单信息。

[0100] 在操作1036处，订单存储在订单数据库1018中。订单数据库1018可以是由云基础设施系统1018操作和与其它系统元件一起操作的若干数据库之一。

[0101] 在操作1038处，订单信息被转发到订单管理模块1020。在一些情况下，订单管理模块1020可以被配置为执行与订单相关的计费和记账功能，诸如验证订单，并且在验证后，预订订单。

[0102] 在操作1040处，将关于订单的信息传送到订单编排模块1022。订单编排模块1022可以利用订单信息为客户下的订单编排服务和资源的供给。在一些情况下，订单编排模块1022可以使用订单供给模块1024的服务来编排资源的供给以支持订阅的服务。

[0103] 在某些实施例中，订单编排模块1022使得能够管理与每个订单相关联的业务流程

并应用业务逻辑来确定订单是否应该进行到供给。在操作1042处，在接收到新订阅的订单时，订单编排模块1022向订单供给模块1024发送请求以分配资源并配置履行订阅订单所需的那些资源。订单供给模块1024使得能够为客户订购的服务分配资源。订单供给模块1024提供由云基础设施系统1000提供的云服务和用于供给用于提供所请求的服务的资源的物理实现层之间的抽象层。因此，订单编排模块1022可以与实现细节隔离，诸如服务和资源是否实际上即时供给或预先供给并仅在请求后才分配/指派。

[0104] 在操作1044处，一旦供给了服务和资源，就可以通过云基础设施系统1002的订单供给模块1024向客户端设备1004、1006和/或1008上的客户发送所提供服务的通知。

[0105] 在操作1046处，订单管理和监视模块1026可以管理和跟踪客户的订阅订单。在一些情况下，订单管理和监视模块1026可以被配置为收集订阅订单中的服务的使用统计，诸如，所使用的存储量、传输的数据量、用户的数量，以及系统运行时间和系统停机时间量。

[0106] 在某些实施例中，云基础设施系统1000可以包括身份管理模块1028。身份管理模块1028可以被配置为提供身份服务，诸如云基础设施系统1000中的访问管理和授权服务。在一些实施例中，身份管理模块1028可以控制关于希望利用由云基础设施系统1002提供的服务的客户的信息。这样的信息可以包括认证这些客户的身份的信息以及描述这些客户被授权相对于各种系统资源(例如，文件、目录、应用、通信端口、存储器段等)执行哪些动作的信息。身份管理模块1028还可以包括对关于每个客户的描述性信息以及关于如何和由谁来访问和修改这些描述性信息的管理。

[0107] 图11图示了其中可以实现各种实施例的示例性计算机系统1100。系统1100可以用于实现上述计算机系统中的任何一个。如图所示，计算机系统1100包括经由总线子系统1102与多个外围子系统通信的处理单元1104。这些外围子系统可以包括处理加速单元1106、I/O子系统1108、存储子系统1118和通信子系统1124。存储子系统1118包括有形计算机可读存储介质1122和系统存储器1110。

[0108] 总线子系统1102提供用于让计算机系统1100的各种组件和子系统按意图彼此进行通信的机制。虽然总线子系统1102被示意性地示出为单条总线，但是总线子系统的替代实施例可以利用多条总线。总线子系统1102可以是若干种类型的总线结构中的任何类型，这些总线类型包括存储器总线或存储器控制器、外围总线、以及使用各种总线体系架构中的任何体系架构的局部总线。例如，这种体系架构可以包括工业标准体系架构(ISA)总线、微通道体系架构(MCA)总线、增强型ISA(EISA)总线、视频电子标准协会(VESA)局部总线和外围组件互连(PCI)总线，这些总线可以被实现为按IEEE P1086.1标准制造的Mezzanine总线。

[0109] 可以被实现为一个或多个集成电路(例如，常规微处理器或微控制器)的处理单元1104控制计算机系统1100的操作。一个或多个处理器可以被包括在处理单元1104中。这些处理器可以包括单核处理器或多核处理器。在某些实施例中，处理单元1104可以被实现为一个或多个独立的处理单元1132和/或1134，其中在每个处理单元中包括单核处理器或多核处理器。在其它实施例中，处理单元1104也可以被实现为通过将两个双核处理器集成到单个芯片中形成的四核处理单元。

[0110] 在各种实施例中，处理单元1104可以响应于程序代码执行各种程序并且可以维护多个并发执行的程序或进程。在任何给定的时间，要被执行的程序代码中的一些或全部代

码可以驻留在(一个或多个)处理器1104中和/或存储子系统1118中。通过适当的编程,(一个或多个)处理器1104可以提供上述各种功能。计算机系统1100可以附加地包括处理加速单元1106,该处理加速单元1106可以包括数字信号处理器(DSP)、专用处理器等。

[0111] I/O子系统1108可以包括用户接口输入设备和用户接口输出设备。用户接口输入设备可以包括键盘、诸如鼠标或轨迹球的指向设备、结合到显示器中的触摸板或触摸屏、滚动轮、点击轮、拨盘、按钮、开关、小键盘、具有语音命令辨识系统的音频输入设备、麦克风以及其它类型的输入设备。用户接口输入设备可以包括,例如,运动感测和/或手势辨识设备,诸如Microsoft **Kinect®**运动传感器,该运动传感器使得用户能够通过使用手势和语音命令的自然用户接口来控制诸如Microsoft **Xbox®**360游戏控制器的输入设备并与输入设备交互。用户接口输入设备也可以包括眼睛姿势辨识设备,诸如从用户检测眼睛活动(例如,当拍摄照片和/或进行菜单选择时的"眨眼")并且将眼睛姿势变换为进入输入设备(例如,Google **Glass®**)中的输入的Google **Glass®**眨眼检测器。此外,用户接口输入设备可以包括使用户能够通过语音命令与语音辨识系统(例如,**Siri®**导航器)交互的语音辨识感测设备。

[0112] 用户接口输入设备也可以包括但不限于:三维(3D)鼠标、操纵杆或指向棒(pointing stick)、游戏面板和绘图板,以及音频/视觉设备,诸如扬声器、数码相机、数码摄录机、便携式媒体播放器、网络摄像头、图像扫描仪、指纹扫描仪、条形码阅读器3D扫描仪、3D打印机、激光测距仪和视线跟踪设备。此外,用户接口输入设备可以包括,例如,医学成像输入设备,诸如计算机断层扫描、磁共振成像、正电子发射断层显像、医疗超声设备。用户接口输入设备也可以包括,例如,诸如MIDI键盘、数字乐器等的音频输入设备。

[0113] 用户接口输出设备可以包括显示子系统、指示灯,或者诸如音频输出设备的非可视显示器等。显示子系统可以是阴极射线管(CRT)、诸如使用液晶显示器(LCD)或等离子显示器的平板设备、投影设备、触摸屏等。一般而言,术语"输出设备"的使用旨在包括用于从计算机系统1100向用户或其它计算机输出信息的所有可能类型的设备和机制。例如,用户接口输出设备可以包括但不限于:可视地传达文本、图形和音频/视频信息的各种显示设备,诸如监视器、打印机、扬声器、耳机、汽车导航系统、绘图仪、语音输出设备以及调制解调器。

[0114] 计算机系统1100可以包括包含软件元件的被示出为当前位于系统存储器1110内的存储子系统1118。系统存储器1110可以存储可加载并且可在处理单元1104上执行的程序指令,以及在这些程序的执行期间所生成的数据。

[0115] 取决于计算机系统1100的配置和类型,系统存储器1110可以是易失性的(诸如随机存取存储器(RAM))和/或非易失性的(诸如只读存储器(ROM)、闪存存储器等)。RAM通常包含可被处理单元1104立即访问和/或目前正在被处理单元1104操作和执行的数据和/或程序模块。在一些实施方式中,系统存储器1110可以包括多种不同类型的存储器,例如静态随机存取存储器(SRAM)或动态随机存取存储器(DRAM)。在一些实施方式中,诸如在启动期间,包含有助于在计算机系统1100内的元件之间传送信息的基本例程的基本输入/输出系统(BIOS)通常可以被存储在ROM中。作为示例但不是限制,系统存储器1110也示出了可以包括客户端应用、Web浏览器、中间层应用、关系数据库管理系统(RDBMS)等的应用程序1112,程

序数据1114、以及操作系统1116。作为示例，操作系统1116可以包括各种版本的Microsoft Windows®，Apple Macintosh® 和/或Linux操作系统、各种可商业获得的 UNIX® 或类UNIX操作系统(包括但不限于各种GNU/Linux操作系统、Google Chrome®OS等)和/或诸如iOS、Windows® Phone、Android® OS、BlackBerry® 10OS和Palm®OS操作系统的移动操作系统。

[0116] 存储子系统1118也可以提供用于存储提供一些实施例的功能的基本编程和数据结构的有形计算机可读存储介质。当被处理器执行时提供上述功能的软件(程序、代码模块、指令)可以被存储在存储子系统1118中。这些软件模块或指令可以被处理单元1104执行。存储子系统1118也可以提供用于存储根据本发明被使用的数据的储存库。

[0117] 存储子系统1100也可以包括可被进一步连接到计算机可读存储介质1122的计算机可读存储介质读取器1120。与系统存储器1110一起并且可选地与系统存储器1110相结合，计算机可读存储介质1122可以全面地表示用于临时和/或更持久地包含、存储、发送和检索计算机可读信息的远程、本地、固定和/或可移动的存储设备加存储介质。

[0118] 包含代码或代码的一部分的计算机可读存储介质1122也可以包括本领域已知或使用的任何适当的介质，包括存储介质和通信介质，诸如但不限于：以用于信息的存储和/或传输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。这可以包括有形的计算机可读存储介质，诸如RAM、ROM、电可擦除可编程ROM(EEPROM)、闪存存储器或其它存储器技术、CD-ROM、数字多功能盘(DVD)或其它光学储存器、磁带盒、磁带、磁盘储存装置或其它磁存储设备、或者其它有形计算机可读介质。这也可以包括非有形的计算机可读介质，诸如数据信号、数据传输，或者可以用于发送期望信息并且可以被计算系统1100访问的任何其它介质。

[0119] 作为示例，计算机可读存储介质1122可以包括从不可移动的非易失性磁介质读取或写入到不可移动的非易失性磁介质的硬盘驱动器、从可移动的非易失性磁盘读取或写入到可移动的非易失性磁盘的磁盘驱动器、以及从可移动的非易失性光盘(诸如CD ROM、DVD和Blu-Ray® (蓝光)盘或其它光学介质)读取或写入到可移动的非易失性光盘的光盘驱动器。计算机可读存储介质1122可以包括但不限于：Zip®驱动器、闪存卡、通用串行总线(USB)闪存驱动器、安全数字(SD)卡、DVD盘、数字音频带等。计算机可读存储介质1122也可以包括：基于非易失性存储器的固态驱动器(SSD)(诸如基于闪存存储器的SSD、企业闪存驱动器、固态ROM等)、基于易失性存储器的SSD(诸如固态RAM、动态RAM、静态RAM，基于DRAM的SSD、磁阻RAM(MRAM)SSD)、以及使用基于DRAM和闪存存储器的SSD的组合的混合SSD。盘驱动器及其关联的计算机可读介质可以为计算机系统1100提供计算机可读指令、数据结构、程序模块以及其它数据的非易失性存储。

[0120] 通信子系统1124提供到其它计算机系统和网络的接口。通信子系统1124用作用于从其它系统接收数据和从计算机系统1100向其它系统发送数据的接口。例如，通信子系统1124可以使计算机系统1100能够经由互联网连接到一个或多个设备。在一些实施例中，通信子系统1124可以包括用于访问无线语音和/或数据网络的射频(RF)收发器组件(例如，使用蜂窝电话技术、先进数据网络技术，诸如3G、4G或EDGE(用于全球演进的增强型数据速率)、WiFi(IEEE802.11系列标准)或其它移动通信技术、或者其任何组合)、全球定位系统

(GPS)接收器组件和/或其它组件。在一些实施例中,作为无线接口的附加或者替代,通信子系统1124可以提供有线网络连接(例如,以太网)。

[0121] 在一些实施例中,通信子系统1124也可以代表可以使用计算机系统1100的一个或多个用户来接收以结构化和/或非结构化数据馈送1126、事件流1128、事件更新1130等形式的输入通信。

[0122] 作为示例,通信子系统1124可以被配置为实时地从社交网络和/或其它通信服务的用户接收数据馈送1126,诸如Twitter®馈送、Facebook®更新、诸如丰富站点摘要(RSS)馈送的web馈送和/或来自一个或多个第三方信息源的实时更新。

[0123] 此外,通信子系统1124也可以被配置为接收以连续数据流形式的数据,该数据可以包括本质上可以是连续的或无界的没有明确终止的实时事件的事件流1128和/或事件更新1130。生成连续数据的应用的示例可以包括,例如,传感器数据应用、金融价格收报机、网络性能测量工具(例如,网络监视和流量管理应用)、点击流分析工具、汽车流量监视等。

[0124] 通信子系统1124也可以被配置为向一个或多个数据库输出结构化和/或非结构化的数据馈送1126、事件流1128、事件更新1130等,这一个或多个数据库可以与耦合到计算机系统1100的一个或多个流式数据源计算机进行通信。

[0125] 计算机系统1100可以是各种类型中的一种类型,包括手持便携式设备(例如,iPhone®蜂窝电话、iPad®计算平板电脑、PDA)、可穿戴设备(例如,Google Glass®头戴式显示器)、PC、工作站、大型机、信息站(kiosk)、服务器机架、或任何其它数据处理系统。

[0126] 由于计算机和网络的不断变化的本质,在图中绘出的计算机系统1100的描述仅旨在作为具体的示例。具有比图中绘出的系统更多或更少组件的许多其它配置是可能的。例如,定制的硬件也可以被使用和/或特定的元素可以在硬件、固件、软件(包括小程序应用(applets))或其组合中实现。另外,可以采用到诸如网络输入/输出设备的其它计算设备的连接。基于本文提供的公开内容和教导,本领域普通技术人员将认识到实现各种实施例的其它方式和/或方法。

[0127] 在前面的描述中,出于解释的目的,阐述了许多具体细节以便提供对本发明的各种实施例的透彻理解。但是,对于本领域技术人员将显而易见的是,可以在没有这些具体细节中的一些的情况下实践本发明的实施例。在其它情况下,以框图形式示出了众所周知的结构和设备。

[0128] 前面的描述仅提供示例性实施例,并且无意于限制本公开的范围、适用性或配置。更确切地说,示例性实施例的前述描述将向本领域技术人员提供用于实现示例性实施例的使能描述。应当理解的是,在不脱离所附权利要求书所阐述的本发明的精神和范围的情况下,可以对元件的功能和布置进行各种改变。

[0129] 在前面的描述中给出了具体细节以提供对实施例的透彻理解。但是,本领域普通技术人员将理解的是,可以在没有这些具体细节的情况下实践实施例。例如,电路、系统、网络、处理和其它组件可能已经被示为框图形式的组件,以便不以不必要的细节使实施例晦涩难懂。在其它情况下,可以已经示出了众所周知的电路、处理、算法、结构和技术而没有不必要的细节,以避免使实施例晦涩难懂。

[0130] 而且,要注意的是,各个实施例可以已经被描述为被描绘为流程图、流图、数据流图、结构图或框图的处理。虽然流程图可以已经将操作描述为顺序处理,但是许多操作可以

并行或并发地执行。此外，可以重新布置操作的次序。处理在器操作完成时终止，但可以具有图中未包括的附加步骤。处理可以与方法、函数、过程、子例程、子程序等对应。当处理与函数对应时，其终止可以与该函数返回到调用函数或主函数对应。

[0131] 术语“计算机可读介质”包括但不限于便携式或固定存储设备、光学存储设备、无线信道以及能够存储、包含或携带(一个或多个)指令和/或数据的各种其它介质。代码段或机器可执行指令可以表示过程、功能、子程序、程序、例程、子例程、模块、软件包、类或指令、数据结构或程序语句的任意组合。代码段可以通过传递和/或接收信息、数据、自变量、参数或存储器内容而耦合到另一个代码段或硬件电路。信息、自变量、参数、数据等可以经由任何合适的方式传递、转发或传输，包括存储器共享、消息传递、令牌传递、网络传输等。

[0132] 此外，实施例可以通过硬件、软件、固件、中间件、微代码、硬件描述语言或其任何组合来实现。当以软件、固件、中间件或微代码实现时，可以将执行必要任务的程序代码或代码段存储在机器可读介质中。(一个或多个)处理器可以执行必要的任务。

[0133] 在前述说明书中，参考本发明的各方面的具体实施例对本发明的各方面进行了描述，但是本领域技术人员将认识到的是，本发明不限于此。上述发明的各个特征和方面可以被单独使用或联合使用。另外，在不脱离本说明书的更广泛精神和范围的情况下，实施例可以在除本文所述的那些环境和应用之外的任何数目的环境和应用中被使用。相应地，本说明书和附图应当被认为是说明性的而不是限制性的。

[0134] 此外，出于说明的目的，以特定顺序描述了方法。应该认识到的是，在替代示例中，可以以与所描述的顺序不同的顺序执行方法。还应该认识到的是，上述方法可以由硬件组件执行，或者可以以机器可执行指令的序列被实施，机器可执行指令可以用于使机器(诸如编程有指令的通用或专用处理器或逻辑电路)执行方法。这些机器可执行指令可以被存储在一个或多个机器可读介质上，诸如CD-ROM或其它类型的光盘、软盘、ROM、RAM、EPROM、EEPROM、磁卡或光卡、闪存、或适合于存储电子指令的其它类型的机器可读介质。替代地，可以通过硬件和软件的组合来执行方法。

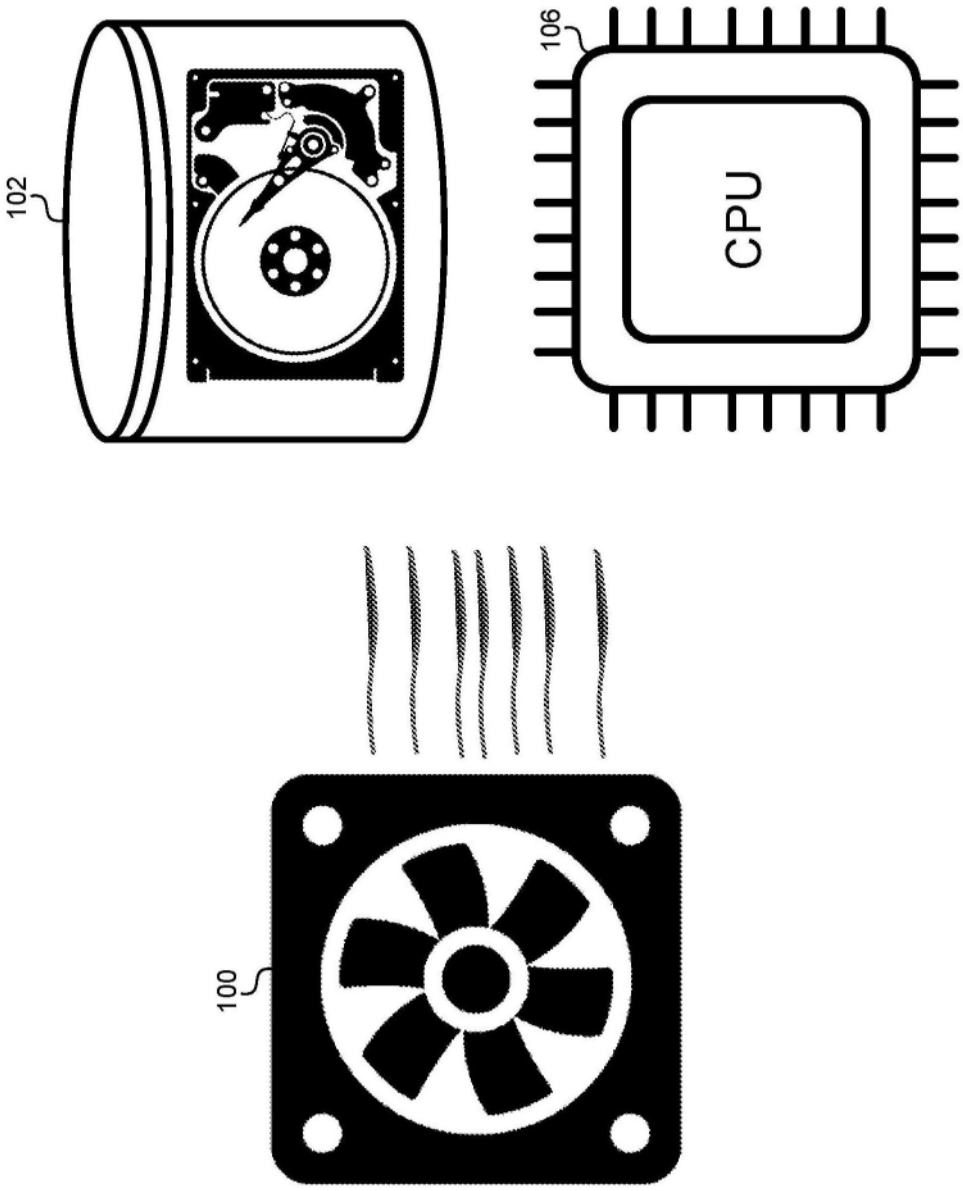


图1

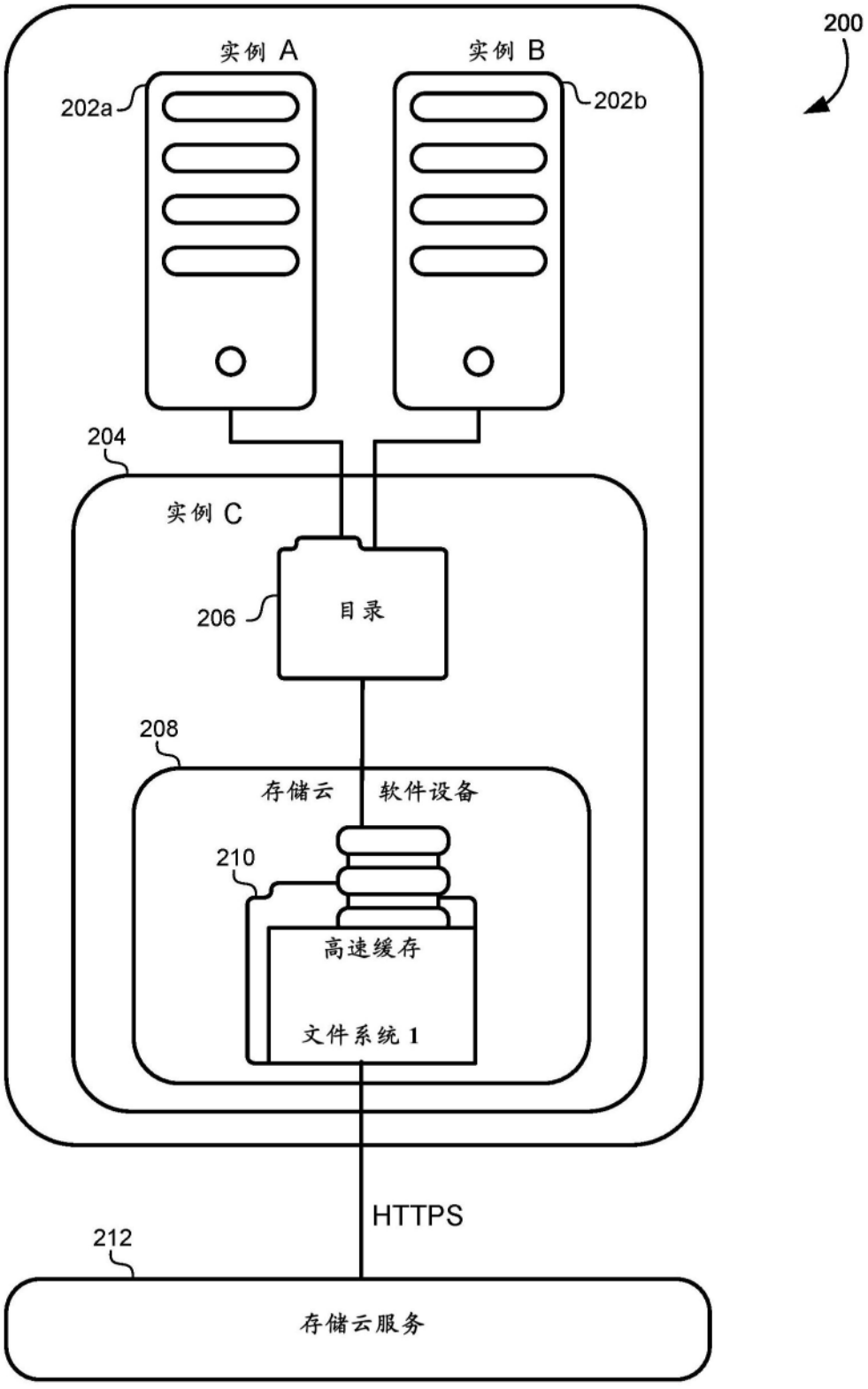

200
实例 A
实例 B
202a
202b
204
实例 C
206
目录
208
存储云
软件设备
210
高速缓存
文件系统 1
HTTPS
212
存储云服务

图2

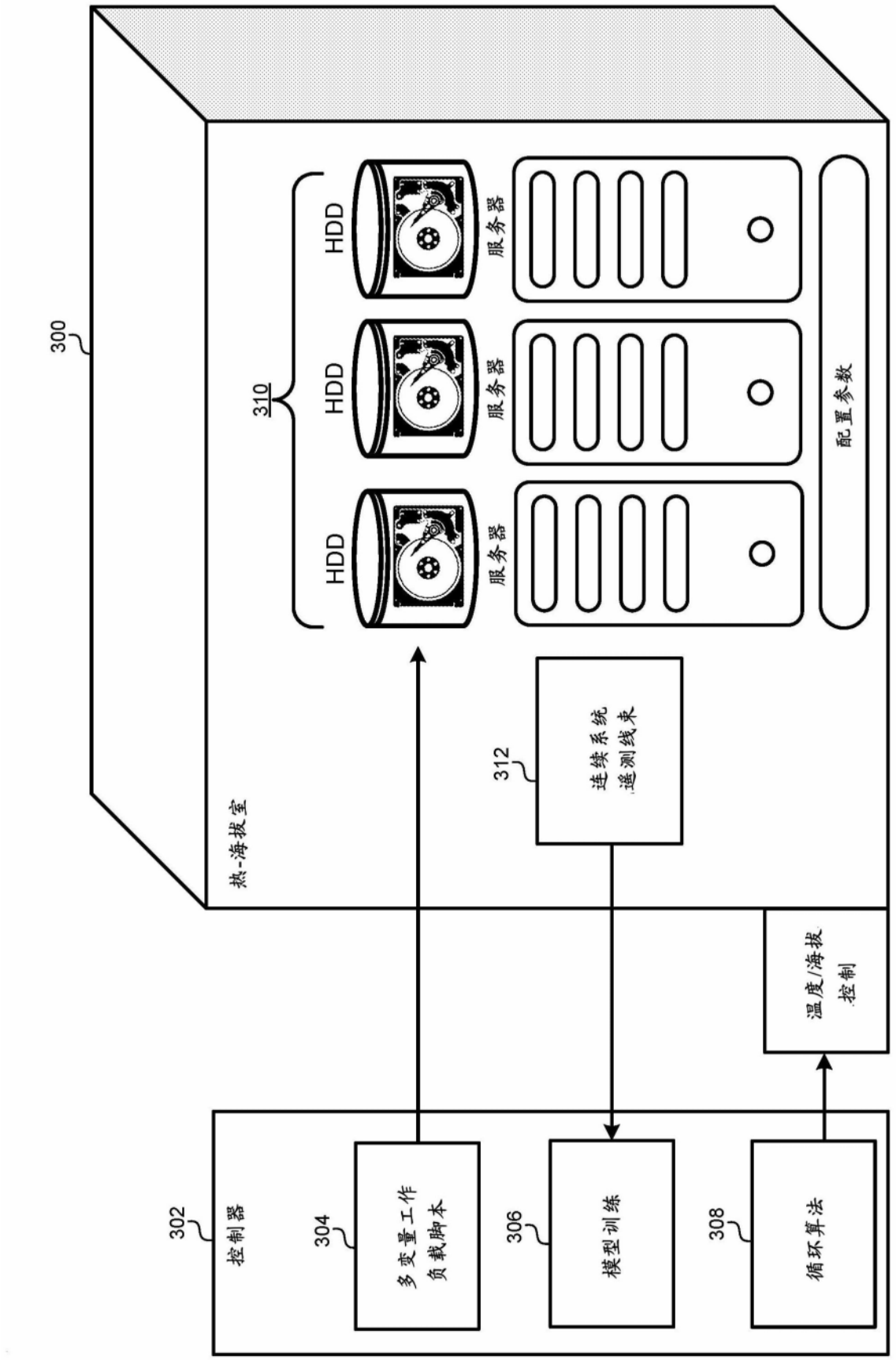

300
热-海拔室
310
HDD
服务器
HDD
服务器
HDD
服务器
配置参数
312
连续系统遥测线束
温度/海拔控制
302
控制器
304
多变量工作负载脚本
306
模型训练
308
循环算法

图3

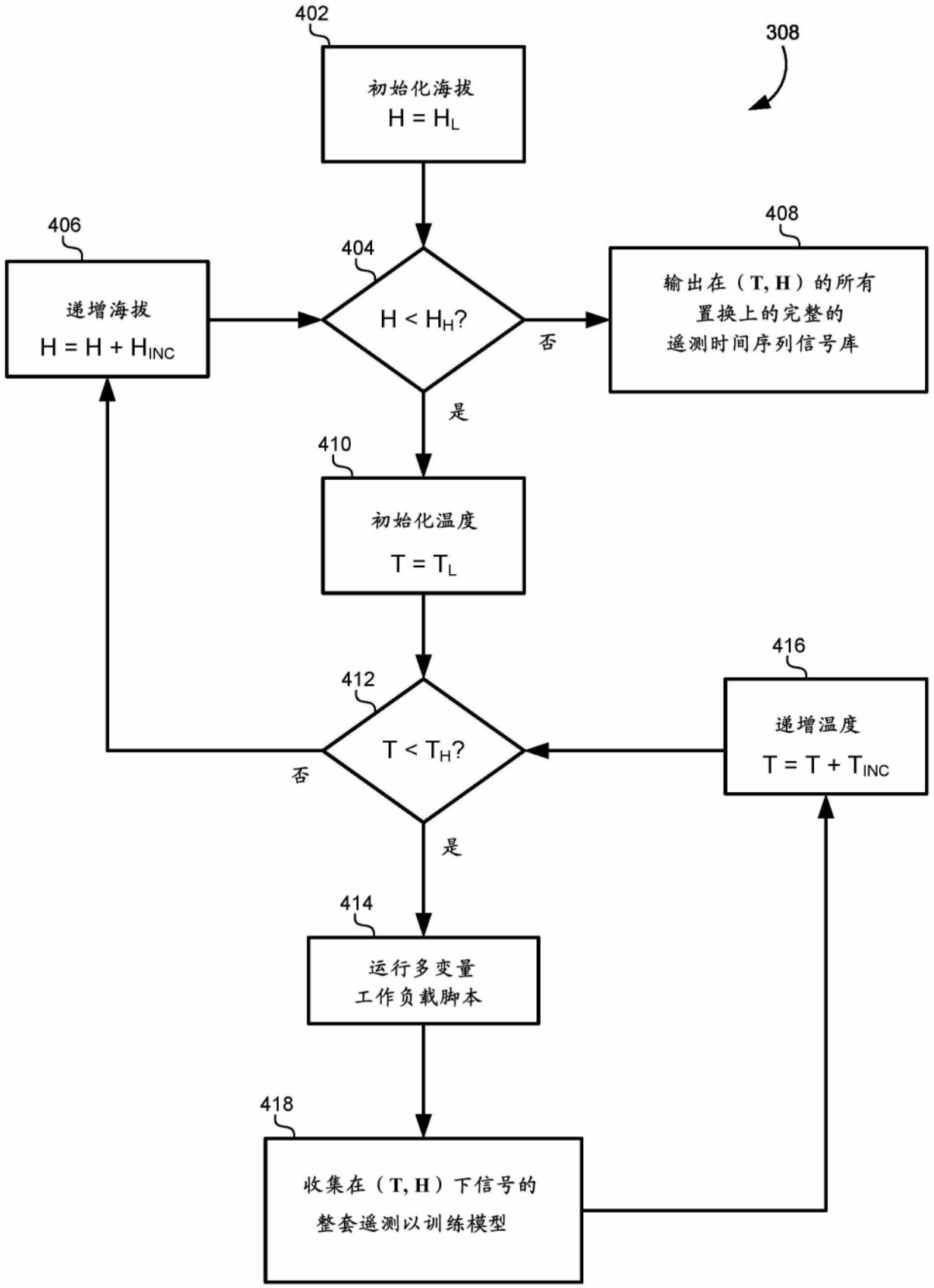

402
初始化海拔
H = H_L
308
406
递增海拔
H = H + H_INC
404
H < H_H?
否
408
输出在（T, H）的所有
置换上的完整的
遥测时间序列信号库
是
410
初始化温度
T = T_L
412
T < T_H?
否
416
递增温度
T = T + T_INC
是
414
运行多变量
工作负载脚本
418
收集在（T, H）下信号的
整套遥测以训练模型

图4

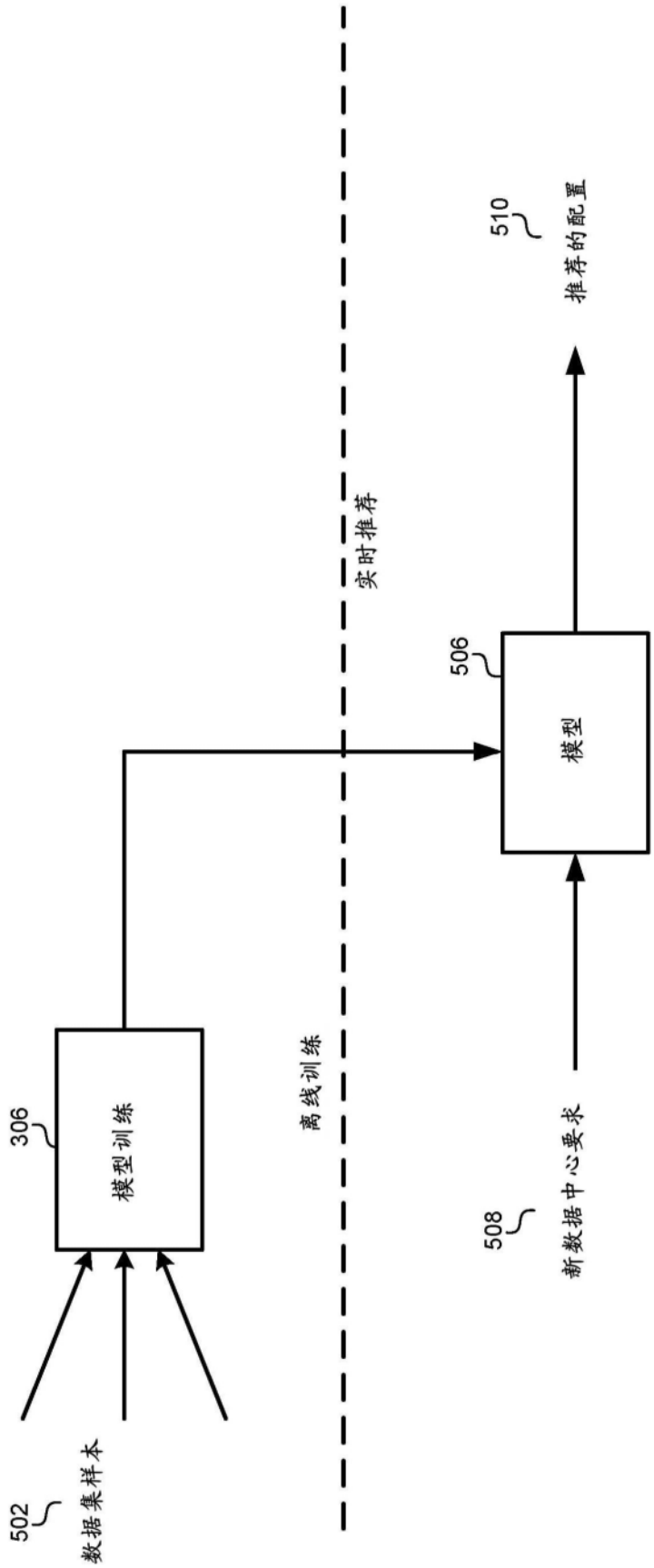


图5

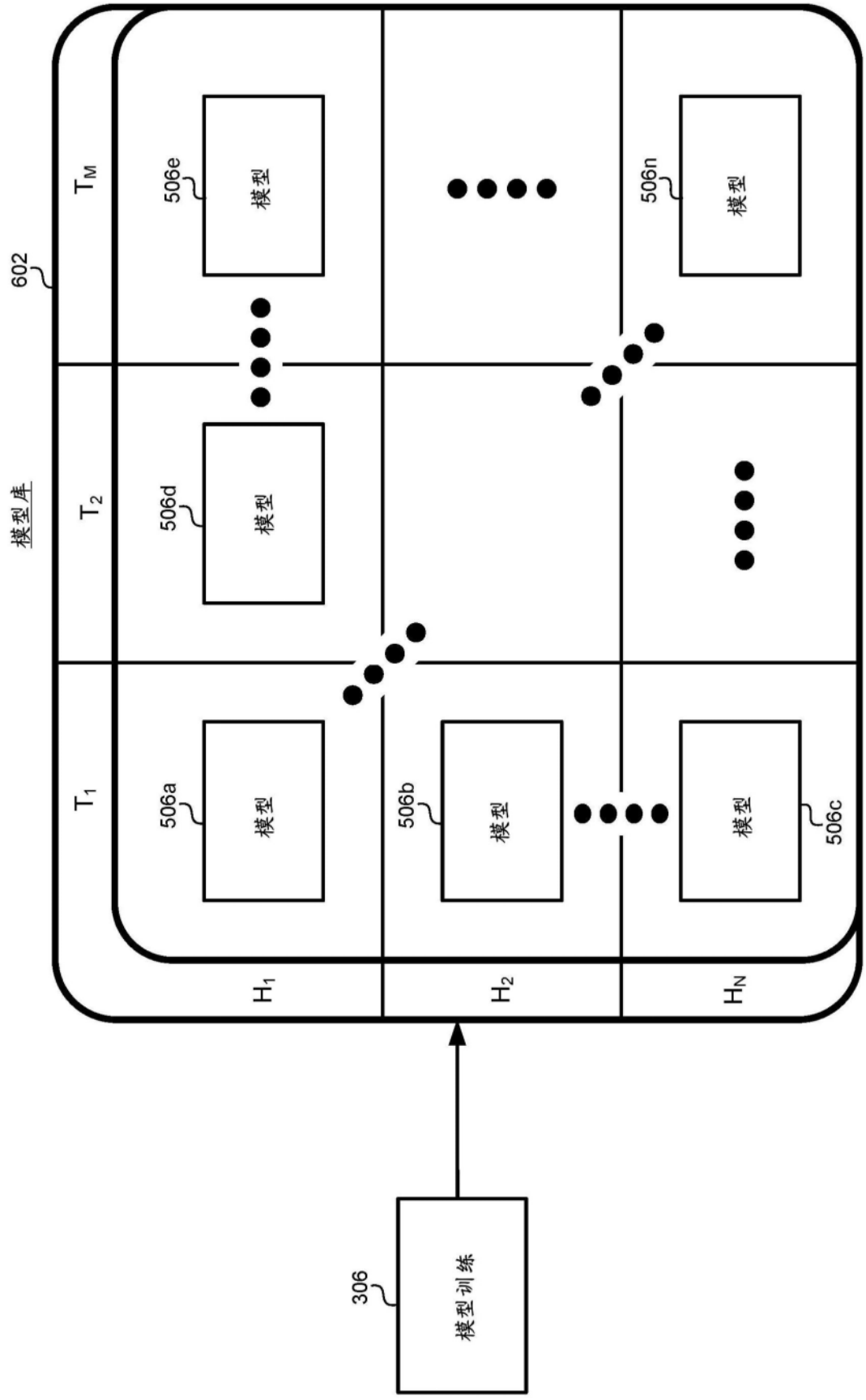

602
模型库
T_M
T_2
T_1
506e
模型
506d
模型
506a
模型
506b
模型
506c
模型
506n
模型
H_1
H_2
H_N
306
模型训练

图6

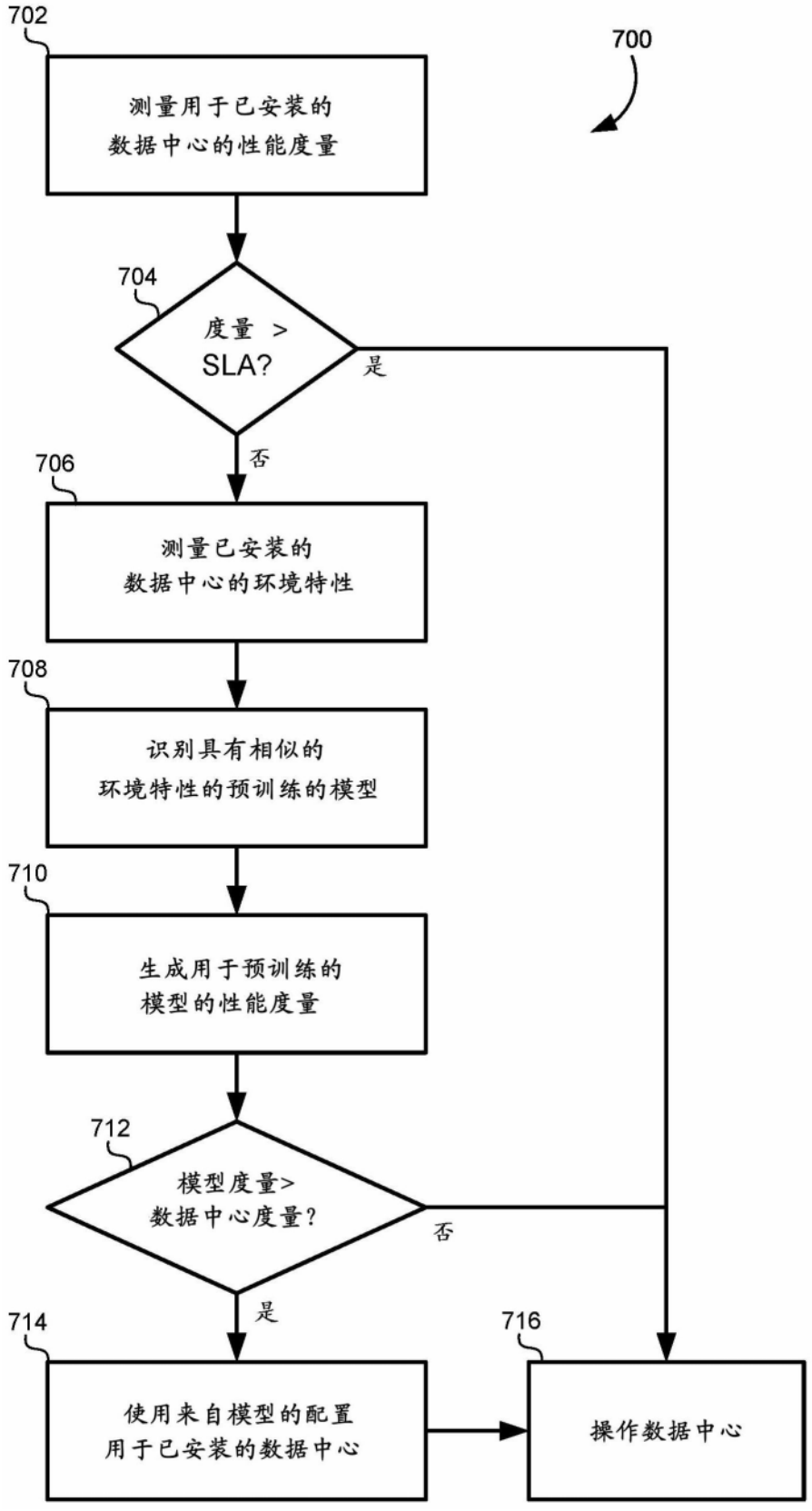

700
702
测量用于已安装的
数据中心的性能度量
704
度量 >
SLA?
是
否
706
测量已安装的
数据中心的环境特性
708
识别具有相似的
环境特性的预训练的模型
710
生成用于预训练的
模型的性能度量
712
模型度量>
数据中心度量?
否
是
714
使用来自模型的配置
用于已安装的数据中心
716
操作数据中心

图7

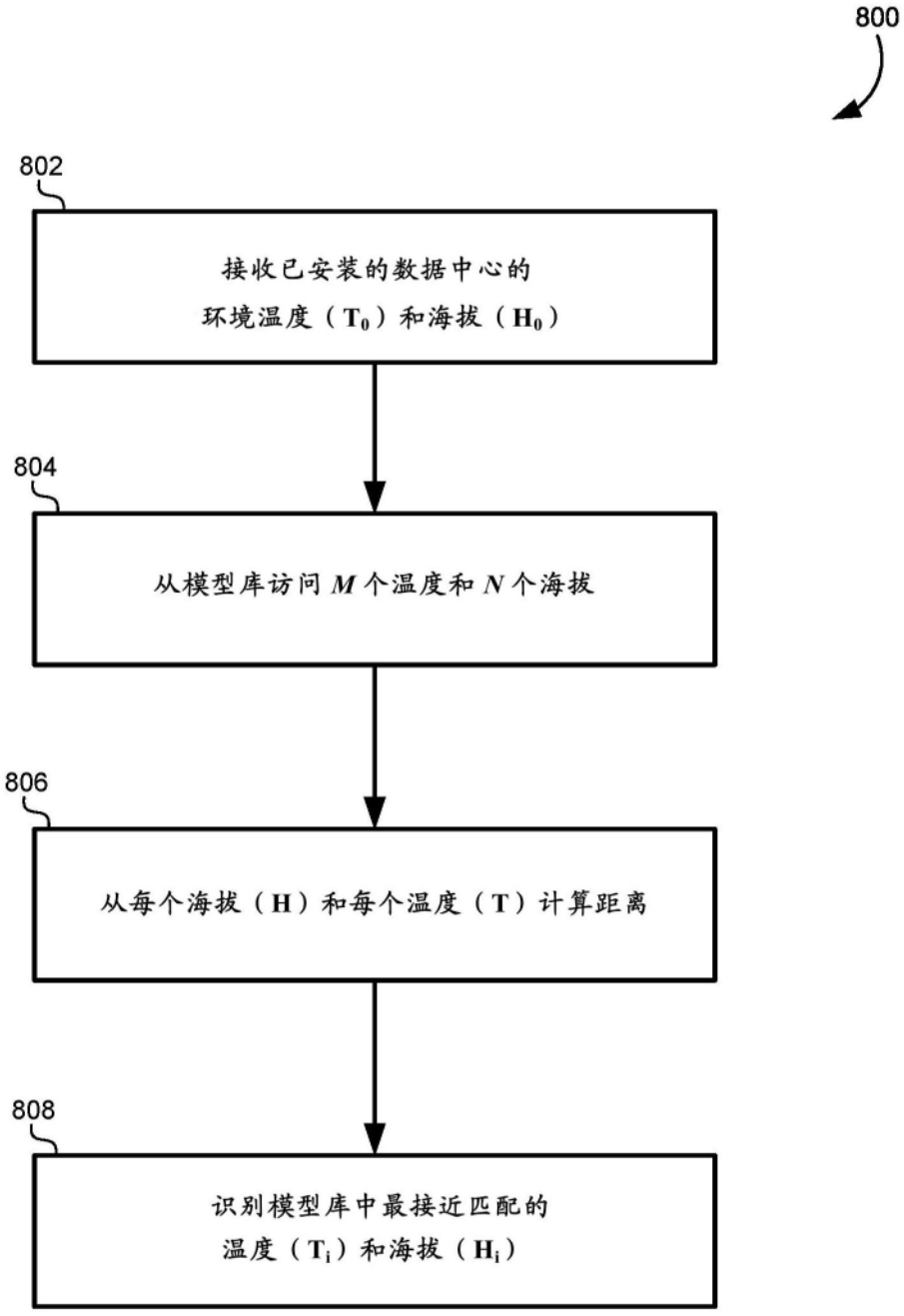

800
802
接收已安装的数据中心的
环境温度（T_0）和海拔（H_0）
804
从模型库访问 *M* 个温度和 *N* 个海拔
806
从每个海拔（H）和每个温度（T）计算距离
808
识别模型库中最接近匹配的
温度（T_i）和海拔（H_i）

图8

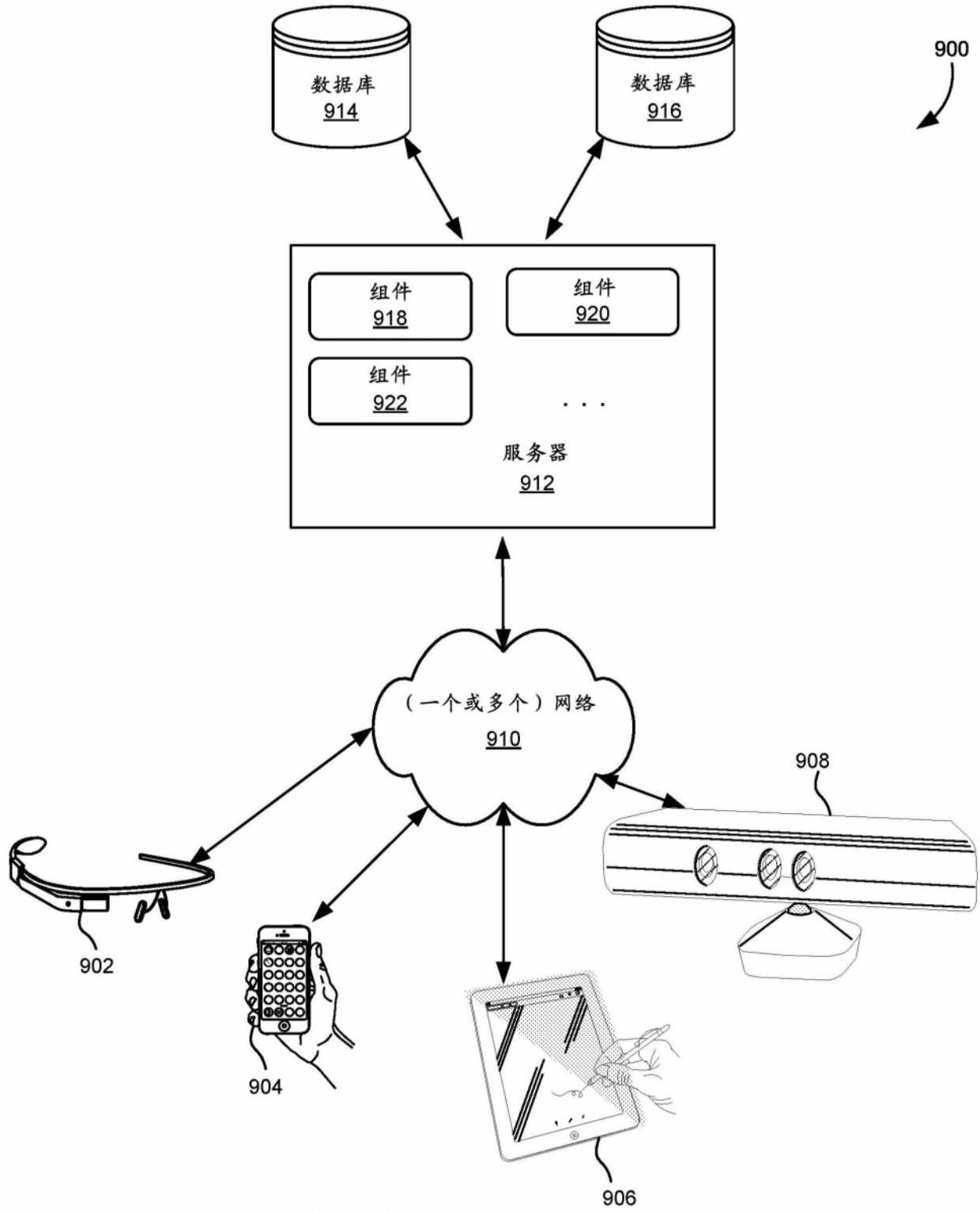

900
数据库
914
数据库
916
组件
918
组件
920
组件
922
. . .
服务器
912
（一个或多个）网络
910
908
902
904
906

图9

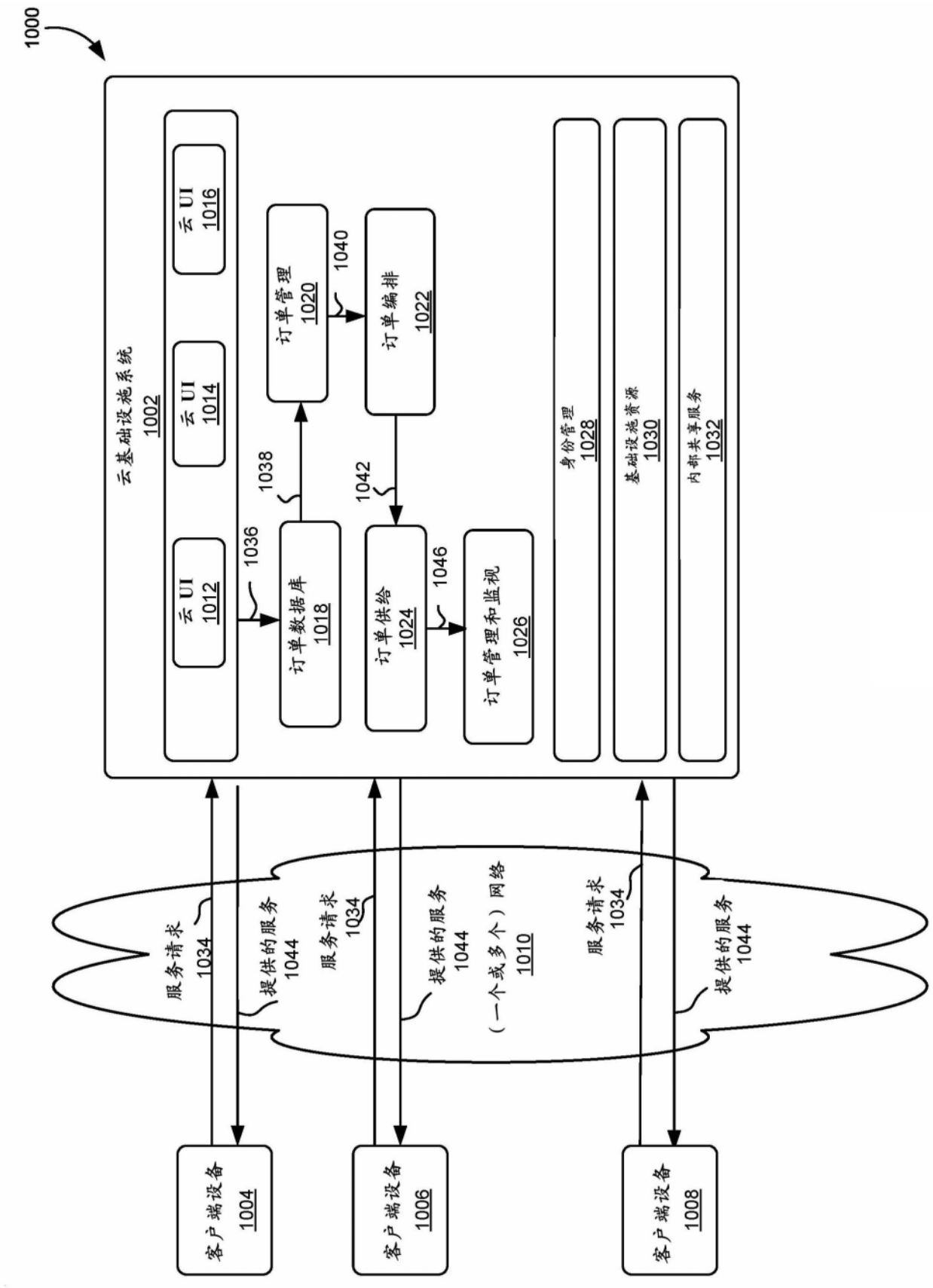

1000
云基础设施系统
1002
云 UI
1012
云 UI
1014
云 UI
1016
订单管理
1020
订单编排
1022
订单数据库
1018
订单供给
1024
订单管理和监视
1026
身份管理
1028
基础设施资源
1030
内部共享服务
1032
1036
1038
1040
1042
1046
服务请求
1034
提供的服务
1044
服务请求
1034
提供的服务
1044
（一个或多个）网络
1010
服务请求
1034
提供的服务
1044
客户端设备
1004
客户端设备
1006
客户端设备
1008

图10

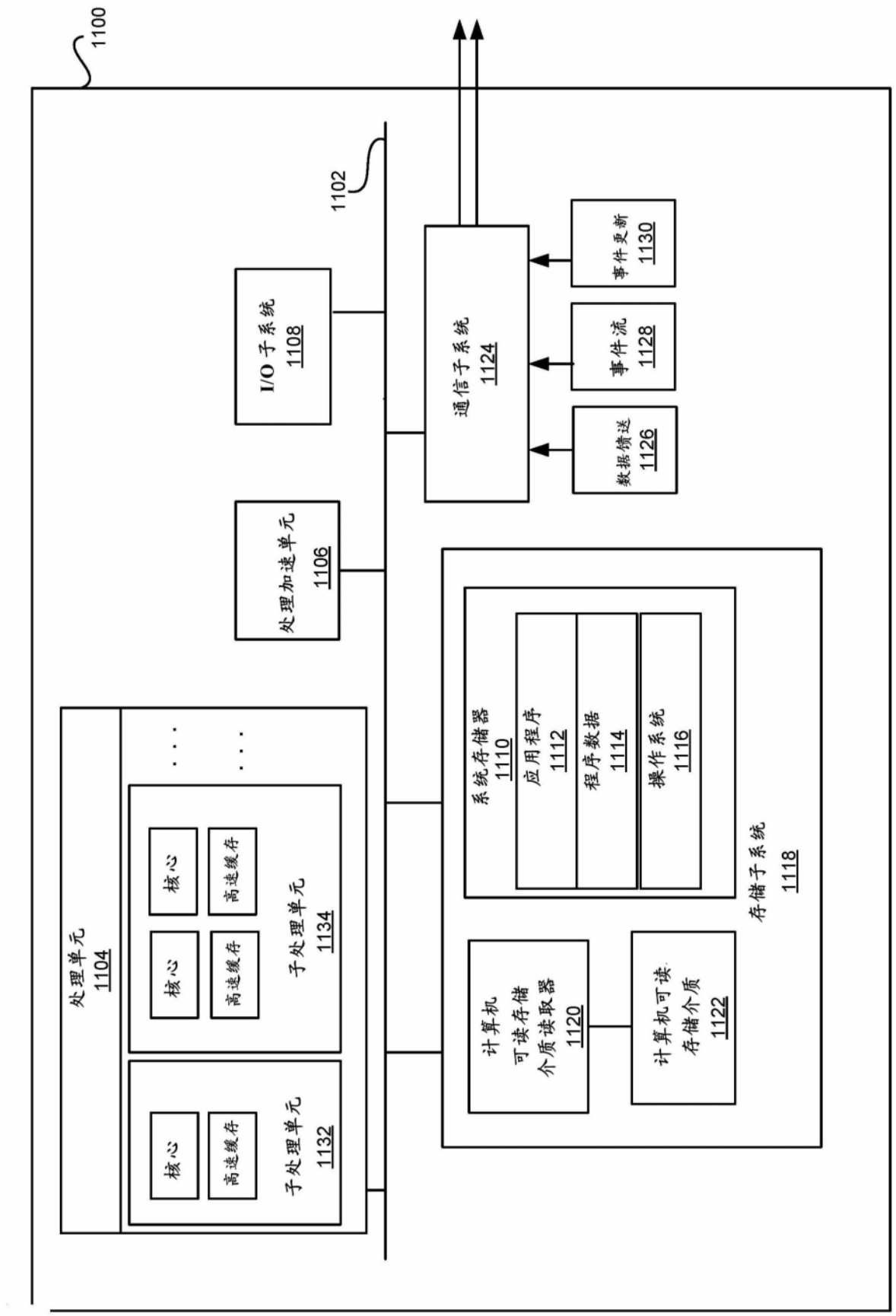

1100
1102
I/O 子系统
1108
通信子系统
1124
事件更新
1130
事件流
1128
数据馈送
1126
处理加速单元
1106
处理单元
1104
核心
高速缓存
子处理单元
1132
子处理单元
1134
系统存储器
1110
应用程序
1112
程序数据
1114
操作系统
1116
存储子系统
1118
计算机
可读存储
介质读取器
1120
计算机可读
存储介质
1122

图11