



(12) 发明专利申请

(10) 申请公布号 CN 104104551 A

(43) 申请公布日 2014. 10. 15

(21) 申请号 201310113359. 5

(22) 申请日 2013. 04. 02

(71) 申请人 中国移动通信集团公司
地址 100032 北京市西城区金融大街 29 号

(72) 发明人 杨雪梅 段翔 孙少陵

(74) 专利代理机构 北京派特恩知识产权代理有限公司 11270
代理人 张振伟 王黎延

(51) Int. Cl.

H04L 12/26 (2006. 01)

H04L 12/24 (2006. 01)

H04L 29/08 (2006. 01)

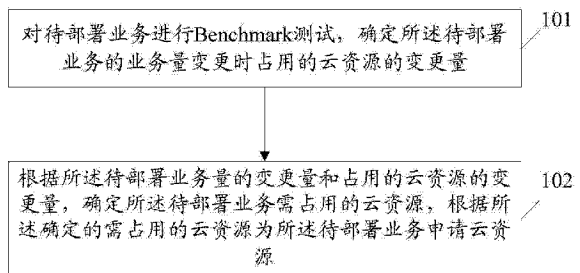
权利要求书2页 说明书7页 附图1页

(54) 发明名称

云资源需求评估方法及装置

(57) 摘要

本发明公开了一种云资源需求评估方法,包括:对待部署业务进行基准(Benchmark)测试,确定所述待部署业务的业务量变更时占用的云资源的变更量;根据所述待部署业务量的变更量和占用的云资源的变更量,确定所述待部署业务需占用的云资源。本发明同时还公开了一种云资源需求评估装置。本发明能够确定需为待部署业务申请的云资源,控制业务部署实施成本,提高云资源的利用率。



1. 一种云资源需求评估方法,其特征在于,所述方法包括:

对待部署业务进行基准 Benchmark 测试,确定所述待部署业务的业务量变更时占用的云资源的变更量;

根据所述待部署业务量的变更量和占用的云资源的变更量,确定所述待部署业务需占用的云资源,根据所述确定的需占用的云资源为所述待部署业务申请云资源。

2. 根据权利要求 1 所述的方法,其特征在于,所述根据待部署业务量的变更量和占用的云资源的变更量,确定所述待部署业务需占用的云资源,包括:

根据所述业务量的变更量 c_1 和占用云资源的变更量 z_1 ,确定待部署业务的业务量为 c 时需占用的云资源 X_1 为 $(z_1/c_1) \times c$ 。

3. 根据权利要求 1 所述的方法,其特征在于,所述方法还包括:

根据与所述待部署业务相同类型的历史部署业务的业务量、所述历史部署业务占用的云资源以及所述占用云资源的利用率,确定所述待部署业务需占用的云资源 X_2 ;确定需为所述待部署业务申请的云资源 X 为 $(a \times X_1) + (b \times X_2)$,其中, a 和 b 分别为 X_1 和 X_2 对应的权重。

4. 根据权利要求 3 所述的方法,其特征在于,所述根据与所述待部署业务相同类型的历史部署业务的业务量、所述历史部署业务占用的云资源以及所述占用云资源的利用率,确定所述待部署业务需占用的云资源,包括:

根据所述历史部署业务占用的云资源 T ,以及所述历史部署业务的业务量为 c_2 时占用的云资源 T 的利用率 η ,确定所述待部署业务单业务需占用云资源数量 z_2 为 $(T \times \eta) / c_2$,并根据所述待部署业务单业务需占用云资源 z_2 确定业务量为 c 的待部署业务需占用的云资源 X_2 为 $z_2 \times c$ 。

5. 根据权利要求 4 所述的方法,其特征在于,所述方法还包括:

待部署业务实际占用的云资源小于所述确定云资源 X 时,按照预设幅度增大所述确定的云资源 X_1 和 X_2 中较大值对应权重的值;

待部署业务实际占用的云资源数量大于所述确定占用云资源 X 时,按照预设幅度增大所述确定的云资源 X_1 和 X_2 中较小值对应权重的值。

6. 根据权利要求 1、2、3、4 或 5 所述的方法,其特征在于,所述方法还包括:

根据待部署业务运行 t 时间内的云资源的利用率的增长率 y ,确定所述业务预计支持运行时间 $(m \times t)$ 需申请的云资源 $X_{m \times t}$ 为 $X \times [(1+y)^t - 1]$ 。

7. 一种云资源需求评估装置,其特征在于,所述云资源需求评估装置包括:第一确定模块、第二确定模块和第三申请模块;其中,

所述第一确定模块,用于对待部署业务进行 Benchmark 测试,确定所述待部署业务的业务量变更时占用的云资源的变更量;

所述第二确定模块,用于根据所述待部署业务量的变更量和占用的云资源的变更量,确定所述待部署业务需占用的云资源;

所述第三申请模块,用于根据所述确定的需占用的云资源为所述待部署业务申请云资源。

8. 根据权利要求 7 所述的云资源需求评估装置,其特征在于,

所述第二确定模块,还用于根据所述业务量的变更量 c_1 和占用云资源的变更量 z_1 ,确

定待部署业务的业务量为 c 时需占用的云资源 X_1 为 $(z_1/c_1) \times c$ 。

9. 根据权利要求 7 所述的云资源需求评估装置,其特征在於,所述云资源需求评估装置还包括:

第四确定模块,用于根据与所述待部署业务相同类型的历史部署业务的业务量、所述历史部署业务占用的云资源以及所述占用云资源的利用率,确定所述待部署业务需占用的云资源 X_2 ;

所述第二确定模块,还用于确定需为所述待部署业务申请的云资源 X 为 $(a \times X_1) + (b \times X_2)$,其中, a 和 b 分别为 X_1 和 X_2 对应的权重。

10. 根据权利要求 9 所述的云资源需求评估装置,其特征在於,

所述第四确定模块,还用于根据所述历史部署业务占用的云资源 T ,以及所述历史部署业务的业务量 c_2 时占用的云资源 T 的利用率 η ,确定所述待部署业务单业务需占用云资源数量 $z_2 = (T \times \eta) / c_2$,并根据所述待部署业务单业务需占用云资源 z_2 确定业务量为 c 的待部署业务需占用云资源 X_2 为 $z_2 \times c$ 。

11. 根据权利要求 10 所述的云资源需求评估装置,其特征在於,

所述第四确定模块,还用于待部署业务实际占用的云资源小于所述确定云资源 X 时,按照预设幅度增大所述确定的云资源 X_1 和 X_2 中较大值对应权重的值;待部署业务实际占用的云资源数量大于所述确定云资源 X 时,按照预设幅度增大所述确定的云资源 X_1 和 X_2 中较小值对应权重的值。

12. 根据权利要求 9、10 或 11 所述的云资源需求评估装置,其特征在於,

所述第四确定模块,还用于根据待部署业务运行 t 时间的云资源的利用率的增长率 y ,确定所述业务预计支持运行时间 $(m \times t)$ 需申请的云资源 $X_{m \times t}$ 为 $X \times [(1+y)^t - 1]$ 。

云资源需求评估方法及装置

技术领域

[0001] 本发明涉及云计算技术,尤其涉及一种云资源需求评估方法及装置。

背景技术

[0002] 云计算技术应用日益广泛,云管理平台将信息技术(IT, Information Technology)资源封装为云资源,根据客户的申请为客户分配云资源并进行计费。

[0003] 当前,客户向云管理平台申请云资源之前,根据部署业务计算需要申请的云资源,对于缺乏IT实施经验的客户,往往无法进行准确计算,后续还需要根据业务实际需求对申请的云资源进行调整,如果前期申请云资源不足会降低业务运行性能,并且后期还需要重新申请云资源,增大了工作量;如果前期申请云资源过多则会增加业务部署成本,同时降低了云资源的利用率。

[0004] 如此,如何在部署业务时确定待部署业务需要占用的云资源,以合理申请云资源,控制实施成本,提高云资源的利用率成为亟待解决的问题。

发明内容

[0005] 有鉴于此,本发明的主要目的在于提供一种云资源需求评估方法及装置,以保证在部署业务时准确确定需要申请云资源,控制实施成本,提高云资源的利用率。

[0006] 为达到上述目的,本发明的技术方案是这样实现的:

[0007] 本发明提供了一种云资源需求评估方法,所述方法包括:

[0008] 对待部署业务进行基准(Benchmark)测试,确定所述待部署业务的业务量变更时占用的云资源的变更量;

[0009] 根据所述待部署业务量的变更量和占用的云资源的变更量,确定所述待部署业务需占用的云资源,根据所述确定的需占用的云资源为所述待部署业务申请云资源。

[0010] 优选的,所述根据待部署业务量的变更量和占用的云资源的变更量,确定所述待部署业务需占用的云资源,包括:

[0011] 根据所述业务量的变更量 c_1 和占用云资源的变更量 z_1 ,确定待部署业务的业务量为 c 时需占用的云资源 X_1 为 $(z_1/c_1) \times c$ 。

[0012] 优选的,所述方法还包括:

[0013] 根据与所述待部署业务相同类型的历史部署业务的业务量、所述历史部署业务占用的云资源以及所述占用云资源的利用率,确定所述待部署业务需占用的云资源 X_2 ;确定需为所述待部署业务申请的云资源 X 为 $(a \times X_1) + (b \times X_2)$,其中, a 和 b 分别为 X_1 和 X_2 对应的权重。

[0014] 优选的,所述根据与所述待部署业务相同类型的历史部署业务的业务量、所述历史部署业务占用的云资源以及所述占用云资源的利用率,确定所述待部署业务需占用的云资源,包括:

[0015] 根据所述历史部署业务占用的云资源 T ,以及所述历史部署业务的业务量为 c_2 时

占用的云资源 T 的利用率 η ，确定所述待部署业务单业务需占用云资源数量 z_2 为 $(T \times \eta) / c_2$ ，并根据所述待部署业务单业务需占用云资源 z_2 确定业务量为 c 的待部署业务需占用的云资源 X_2 为 $z_2 \times c$ 。

[0016] 优选的，所述方法还包括：

[0017] 待部署业务实际占用的云资源小于所述确定云资源 X 时，按照预设幅度增大所述确定的云资源 X_1 和 X_2 中较大值对应权重的值；

[0018] 待部署业务实际占用的云资源数量大于所述确定占用云资源 X 时，按照预设幅度增大所述确定的云资源 X_1 和 X_2 中较小值对应权重的值。

[0019] 优选的，所述方法还包括：

[0020] 根据待部署业务运行 t 时间内的云资源的利用率的增长率 y，确定所述业务预计支持运行时间 $(m \times t)$ 需申请的云资源 $X_{m \times t}$ 为 $X \times [(1+y)^t - 1]$ 。

[0021] 本发明还提供了一种云资源需求评估装置，所述云资源需求评估装置包括：第一确定模块、第二确定模块和第三申请模块；其中，

[0022] 所述第一确定模块，用于对待部署业务进行 Benchmark 测试，确定所述待部署业务的业务量变更时占用的云资源的变更量；

[0023] 所述第二确定模块，用于根据所述待部署业务量的变更量和占用的云资源的变更量，确定所述待部署业务需占用的云资源 X_1 ；

[0024] 所述第三申请模块，用于根据所述确定的需占用的云资源为所述待部署业务申请云资源。

[0025] 优选的，所述第二确定模块，还用于根据所述业务量的变更量 c_1 和占用云资源的变更量 z_1 ，确定待部署业务的业务量为 c 时需占用的云资源 X_1 为 $(z_1 / c_1) \times c$ 。

[0026] 优选的，所述云资源需求评估装置还包括：

[0027] 第四确定模块，用于根据与所述待部署业务相同类型的历史部署业务的业务量、所述历史部署业务占用的云资源以及所述占用云资源的利用率，确定所述待部署业务需占用的云资源 X_2 ；

[0028] 所述第二确定模块，还用于确定需为所述待部署业务申请的云资源 X 为 $(a \times X_1) + (b \times X_2)$ ，其中，a 和 b 分别为 X_1 和 X_2 对应的权重。

[0029] 优选的，所述第四确定模块，还用于根据所述历史部署业务占用的云资源 T，以及所述历史部署业务的业务量 c_2 时占用的云资源 T 的利用率 η ，确定所述待部署业务单业务需占用云资源数量 $z_2 = (T \times \eta) / c_2$ ，并根据所述待部署业务单业务需占用云资源 z_2 确定业务量为 c 的待部署业务需占用云资源 X_2 为 $z_2 \times c$ 。

[0030] 优选的，所述第四确定模块，还用于待部署业务实际占用的云资源小于所述确定云资源 X 时，按照预设幅度增大所述确定的云资源 X_1 和 X_2 中较大值对应权重的值；待部署业务实际占用的云资源数量大于所述确定云资源 X 时，按照预设幅度增大所述确定的云资源 X_1 和 X_2 中较小值对应权重的值。

[0031] 优选的，所述第四确定模块，还用于根据待部署业务运行 t 时间内的云资源的利用率的增长率 y，确定所述业务预计支持运行时间 $(m \times t)$ 需申请的云资源 $X_{m \times t}$ 为 $X \times [(1+y)^t - 1]$ 。

[0032] 本发明所提供的技术方案，分别根据 Benchmark 测试和待部署业务相同类型业务

的历史部署业务的信息,确定需为待部署业务申请的云资源评估值 X 为 $(a \times X_1) + (b \times X_2)$,并在待部署业务试运行占用云资源大于或小于所述评估值 X 时,按照预设幅度调整所述权重值 a 或 b 以对应减小或增大所述需占用的云资源评估值 X ,确保最终能够准确确定需要申请的云资源。部署业务时根据所述评估值 X 申请云资源,避免了初次部署业务时申请云资源不合理的情况,从而提高了云资源的利用率,有效控制了实施成本。

附图说明

[0033] 图 1 为本发明云资源需求评估方法的实现流程示意图;

[0034] 图 2 为本发明云资源需求评估装置的组成结构示意图。

具体实施方式

[0035] 下面结合附图及具体实施例对本发明再作进一步详细的说明。

[0036] 图 1 为本发明云资源需求评估方法的实现流程示意图,如图 1 所示,包括:

[0037] 步骤 101:对待部署业务进行 Benchmark 测试,确定所述待部署业务的业务量变更时占用的云资源的变更量;

[0038] 其中,所述云资源为云管理平台封装的云服务器的资源,包括但不限于:中央处理器(CPU, Central Processing Unit)资源、内存资源、存储资源和网络带宽资源。

[0039] 其中,所述 CPU 资源以云服务器的 CPU 的主频衡量,所述内存资源以所述云服务器的内存容量衡量,所述存储资源以所述云服务器的非易失性存储介质的存储容量衡量,所述网络带宽资源以所述云服务器所使用的网络带宽衡量,所述 CPU 资源、内存资源、存储资源和网络带宽资源的变更量,根据所述资源与所述资源利用率的变更量确定。

[0040] 其中,所述 Benchmark 测试为根据标准性能评估组织(SPEC, Standard Performance Evaluation Corporation)规范进行的测试。

[0041] 步骤 102:根据所述待部署业务量的变更量和占用的云资源的变更量,确定所述待部署业务需占用的云资源,根据所述确定的需占用的云资源为所述待部署业务申请云资源。

[0042] 优选的,所述根据待部署业务量的变更量和占用的云资源的变更量,确定所述待部署业务需占用的云资源,包括:

[0043] 根据所述业务量的变更量 c_1 和占用云资源的变更量 z_1 ,确定待部署业务的业务量为 c 时需占用的云资源 X_1 为 $(z_1/c_1) \times c$ 。

[0044] 例如,在待测试的承载待部署业务的 x 台云服务器上运行基准测试程序,所述基准测试程序在所述 x 台云服务器上加载预设负载的配置文件以模拟云服务器支持不同业务量的待部署业务的场景,可以确定如下数据:所述 x 台云服务器共同支持业务的业务量为 c_1 时,所述业务占用每台云服务器的 CPU 资源 z_{11i} 、内存资源 z_{12i} 、存储资源 z_{13i} 和网络带宽资源 z_{14i} ,其中, i 为正整数且 $1 \leq i \leq x$ 。如此,根据所述确定的业务量 c_1 和占用的云资源,确定业务量为 c 的业务需占用 CPU 资源 $X_{11} = (z_{11}/c_1) \times c$,需占用内存资源 $X_{12} = (z_{12}/c_1) \times c$,需占用存储资源 $X_{13} = (z_{13}/c_1) \times c$,需占用网络带宽资源 $X_{14} = (z_{14}/c_1) \times c$;其中,

$$z_{11} = \sum_{i=1}^x z_{11i}, z_{12} = \sum_{i=1}^x z_{12i}, z_{13} = \sum_{i=1}^x z_{13i}, z_{14} = \sum_{i=1}^x z_{14i}。$$

[0045] 以下对确定待部署业务需要占用的云资源的处理,进行举例说明:

[0046] 承载待部署业务的 x 台云服务器提供相同数量的云资源,通过 SPECWeb99 测试程序,确定一台云服务器支撑业务量为 100 的 Web 业务时,所述云服务器 6 核心主频 2 吉赫 (GHz, Gigabyte Hertz) 的 CPU 的利用率增长 5%,内存占用增长 1 吉字节 (Gigabyte),存储无变化,网络带宽占用增长 10 兆位/秒 (Mbps, Million bits per second),假设 x 台云服务器需支撑待部署业务的业务量为 c ,由于所述云服务提供相同数量的云资源供待部署业务占用,因此每台云服务器平均支撑业务量为 c/x 的业务,则待部署业务需要占用云资源如下:

[0047] CPU 资源 $X_{11} = (c/x) \times [(2 \times 6\text{GHz} \times 5\%) / 100] \times (x)$;

[0048] 内存资源 $X_{12} = (c/x) \times [(1\text{GB} \times 5\%) / 100] \times (x)$;

[0049] 存储资源 $X_{13} = (\text{Web 页面存储量} / x) \times (x)$;

[0050] 网络带宽资源 $X_{14} = (c/x) \times (10\text{Mbps} / 100) \times (x)$ 。

[0051] 其中,由于所述云服务提供相同数量的供待部署业务占用,因此近似认为每台云服务器的云资源占用情况相同,从而可以只测试一台云服务器业务量变更时被占用云资源的变更量。

[0052] 其中,通过对上述运算结果取上整数,可以避免为待部署业务分配云资源不足的情况。

[0053] 其中,由于云服务器业务量增长存储无变化,因此存储资源 X_{13} 为当前云服务器的 Web 页面存储量。

[0054] 优选的,所述方法还包括:

[0055] 根据与所述待部署业务相同类型的历史部署业务的业务量、所述历史部署业务占用的云资源以及所述占用云资源的利用率,确定所述待部署业务需占用的云资源 X_2 ;确定需为所述待部署业务申请的云资源 X 为 $(a \times X_1) + (b \times X_2)$,其中, a 和 b 分别为 X_1 和 X_2 对应的权重。

[0056] 优选的,所述根据与所述待部署业务相同类型的历史部署业务的业务量、所述历史部署业务占用的云资源以及所述占用云资源的利用率,确定所述待部署业务占用的云资源,包括:

[0057] 根据所述历史部署业务占用的云资源 T ,以及所述历史部署业务的业务量 c_2 时占用的云资源 T 的利用率 η ,确定所述待部署业务单业务需占用云资源 z_2 为 $(T \times \eta) / c_2$,并根据所述待部署业务单业务需占用云资源 z_2 确定业务量为 c 的待部署业务需占用的云资源 X_2 为 $z_2 \times c$ 。

[0058] 其中,云管理平台维护所述历史部署业务的信息,所述信息包括云服务器承载的所述历史部署业务的业务类型信息、所述历史部署业务占用所述云服务器的云资源信息和所述占用的云资源的利用率信息。

[0059] 具体的,所述历史部署信息包括以下数据时:云环境中云服务器支持 Web 业务时,所述 Web 业务占用数量 T_1 的 CPU 资源和且所述 CPU 资源的利用率为 η_1 、数量 T_2 的内存资源且所述内存资源利用率为 η_2 、数量 T_3 的存储资源且所述存储资源利用率为 η_3 、以及数量 T_4 的网络带宽资源量且所述网络带宽资源的利用率为 η_4 。根据所述数据确定业务量为 c 的待部署 Web 业务需要占用 CPU 资源 $X_{21} = (T_1 \times \eta_1 / c_2) \times c$,需占用内存资源 $X_{22} = (T_2 \times \eta_2 /$

$c_2) \times c$, 需占用存储资源 $X_{23} = (T_3 \times \eta_3 / c_2) \times c$, 需占用 CPU 资源 $X_{24} = (T_4 \times \eta_4 / c_2) \times c$ 。

[0060] 根据所述信息确定业务量为 c 的 Web 待部署业务需要占用云资源的处理, 举例如下:

[0061] 云环境中 x 台云服务器支持业务量 c_2 的历史 Web 业务, 且每台云服务器提供的 CPU 资源为 T_{1i} 、内存资源为 T_{2i} 、存储资源为 T_{3i} 和网络带宽为 T_{4i} , 每台云服务器 CPU 资源利用率为 η_{1i} , 内存利用率为 η_{2i} , 存储资源利用率为 η_{3i} , 网络带宽资源利用率分别为 η_{4i} , 其中, i 为正整数且 $1 \leq i \leq x$ 。则单历史 Web 业务在每台云服务器上占用的云资源平均为:

[0062] CPU 资源 $Z_{21i} = (T_{1i} \times \eta_{1i}) / c_2$;

[0063] 内存资源 $Z_{22i} = (T_{2i} \times \eta_{2i}) / c_2$;

[0064] 存储资源 $Z_{23i} = (T_{3i} \times \eta_{3i}) / c_2$;

[0065] 网络带宽资源 $Z_{24i} = (T_{4i} \times \eta_{4i}) / c_2$, 其中, i 为正整数且 $1 \leq i \leq x$ 。

[0066] 相应的, 单历史 Web 业务需占用 x 台服务器的云资源为:

[0067] CPU 资源 $Z_{21} = \sum_{i=1}^x Z_{21i}$;

[0068] 内存资源量 $Z_{22} = \sum_{i=1}^x Z_{22i}$;

[0069] 存储资源量 $Z_{23} = \sum_{i=1}^x Z_{23i}$;

[0070] 网络带宽资源量 $Z_{24} = \sum_{i=1}^x Z_{24i}$ 。其中, i 为正整数。

[0071] 待部署 Web 业务业务量 c 时, 需占用云资源为:

[0072] CPU 资源量 $X_{21} = c \times Z_{21} = c \times \sum_{i=1}^x Z_{21i}$;

[0073] 内存资源量的测试值 $X_{22} = c \times Z_{22} = c \times \sum_{i=1}^x Z_{22i}$;

[0074] 存储资源量 $X_{23} = c \times Z_{23} = c \times \sum_{i=1}^x Z_{23i}$;

[0075] 网络带宽资源量 $X_{24} = c \times Z_{24} = c \times \sum_{i=1}^x Z_{24i}$ 。

[0076] 其中, 分配所述权重的值的方法包括但不限于以下两种:

[0077] 权重 a 和 b 的取值空间满足且 $0 < a < 1$, $a+b = 1$;

[0078] 权重 $a = \frac{\sqrt{m^2 + n^2}}{2}$, $b = \frac{\sqrt{m^2 + n^2}}{2}$, 且 $0 \leq m \leq 1$, $m+n = 1$ 。

[0079] 优选的, 所述方法还包括:

[0080] 待部署业务实际占用的云资源小于所述确定占用云资源 X 时, 按照预设幅度增大所述确定的云资源 X_1 和 X_2 中较大值对应权重的值;

[0081] 待部署业务实际占用的云资源数量大于所述确定占用云资源 X 时, 按照预设幅度增大所述确定的云资源 X_1 和 X_2 中较小值对应权重的值。

[0082] 其中, 所述按照预设幅度增大所述确定的云资源 X_1 和 X_2 较大值或较中小值对应权重的值, 可以对应增大或减小所述确定需占用的云资源 X , 以保证待部署业务试运行, 可

以根据所述调整的确定需占用的云资源 X , 向云管理平台申请调整已申请的云资源。

[0083] 优选的, 所述方法还包括:

[0084] 根据待部署业务运行后 t 时间内的云资源的利用率的增长率 y , 确定所述业务预计支持运行时间 $(m \times t)$ 需申请的云资源 $X_{m \times t}$ 为 $X \times [(1+y)^t - 1]$ 。

[0085] 图 2 为本发明云资源需求评估装置的组成结构示意图, 如图 2 所示, 所述云资源需求评估装置包括: 第一确定模块 21、第二确定模块 22 和第三申请模块 23; 其中,

[0086] 所述第一确定模块 21, 用于对待部署业务进行 Benchmark 测试, 确定所述待部署业务的业务量变更时占用的云资源的变更量;

[0087] 所述第二确定模块 22, 用于根据所述待部署业务量的变更量和占用的云资源的变更量, 确定所述待部署业务需占用的云资源;

[0088] 所述第三申请模块 23, 用于根据所述确定的需占用的云资源为所述待部署业务申请云资源。

[0089] 优选的, 所述第二确定模块 22, 还用于根据所述业务量的变更量 c_1 和占用云资源的变更量 z_1 , 确定待部署业务的业务量为 c 时需占用的云资源 X_1 为 $(z_1/c_1) \times c$ 。

[0090] 优选的, 所述云资源需求评估装置还包括:

[0091] 第四确定模块 24, 用于根据与所述待部署业务相同类型的历史部署业务的业务量、所述历史部署业务占用的云资源以及所述占用云资源的利用率, 确定所述待部署业务需占用的云资源 X_2 ;

[0092] 所述第二确定模块 22, 还用于确定需为所述待部署业务申请的云资源 X 为 $(a \times X_1) + (b \times X_2)$, 其中, a 和 b 分别为 X_1 和 X_2 对应的权重。

[0093] 优选的, 所述第二确定模块 22, 还用于根据所述历史部署业务占用的云资源 T , 以及所述历史部署业务的业务量 c_2 时占用的云资源 T 的利用率 η , 确定所述待部署业务单业务运行占用云资源数量 $z_2 = (T \times \eta) / c_2$, 并根据所述待部署业务单业务量运行占用云资源 z_2 确定业务量为 c 的待部署业务运行需占用的云资源 X_2 为 $z_2 \times c$ 。

[0094] 优选的, 所述第四确定模块 24, 还用于待部署业务实际占用的云资源小于所述确定占用云资源 X 时, 按照预设幅度增大所述确定的云资源 X_1 和 X_2 中较大值对应权重的值; 待部署业务实际占用的云资源大于所述确定占用云资源 X 时, 按照预设幅度增大所述确定的云资源 X_1 和 X_2 中较小值对应权重的值。

[0095] 优选的, 所述第四确定模块 24, 还用于根据待部署业务运行后 t 时间的云资源的利用率的增长率 y , 确定所述业务预计支持运行时间 $(m \times t)$ 需申请的云资源 $X_{m \times t}$ 为 $X \times [(1+y)^t - 1]$ 。

[0096] 本发明实施例中所提供的云资源需求评估装置可以模块化嵌入云管理平台的服务器中, 也可以在云管理平台中作为独立服务器使用。

[0097] 实施例

[0098] 本实施例中, 以在云环境中的云服务器上部署邮件系统为例来对本发明方法做进一步详细说明。

[0099] 本实施例中, 通过在云管理平台嵌入云资源需求评估模块的方式, 来实现本发明对待部署业务需求的云资源进行评估的目的。

[0100] 所述云资源需求评估模块为不同类型的待部署业务维护相应的业务模型, 所述业

务模型包括根据 Benchmark 测试待部署业务,确定的待部署业务量 c_1 时占用云资源的变更量 z_1 的信息,以及根据待部署业务相同类型的历史部署业务的信息,确定与待部署业务相同类型单历史部署业务占用云资源 z_2 的信息。

[0101] 当用户登陆云管理平台下达部署邮件系统业务需求时,典型的,所述业务需求包括:最大支持并发用户数量、最大附件大小、最大响应时间和期望邮件服务软件。云资源需求评估模块根据所述业务需求信息确定对应的业务模型,将支持并发最大用户数量作为业务量 c ,根据对应模型,确定待部署邮件系统需占用云资源 $X_1 = (z_1/c_1) \times c$,确定待部署邮件系统需占用云资源 $X_2 = z_2 \times c$,为 Z_1 和 Z_2 对应分配权重 a 和 b ,从而确定需为所述待部署邮件系统申请的云资源 $X = (a \times X_1) + (b \times X_2)$;

[0102] 云资源需求评估模块将所述需占用的云资源评估值 X 通过云管理平台反馈给客户,客户还可以对上述评估值 X 进行修改,并根据修改后的评估值 X 向云管理平台进行云资源批量申请;

[0103] 相应的,在申请的云资源上部署邮件系统并试运行时,若根据所述评估值 X 已申请的云资源小于或大于邮件系统实际运行需要云资源,云资源需求评估模块将按照预设幅度调整所述权重 a 或 b 的值,以对应增大或减小需要占用的云资源的评估值 X ,并根据所述评估值向云管理平台申请更新已申请云资源;

[0104] 进一步的,在对所邮件系统进行扩容时,根据所述邮件系统运行 t 时间的云资源的利用率的增长率 y ,确定所述邮件系统扩容后预计支持运行时间 $(m \times t)$ 需占用的云资源 $X \times (1+y)^t$,相应的,需要向云管理平台再次申请云资源 $X_{m \times t}$ 为 $X \times [(1+y)^t - 1]$ 。

[0105] 以上所述,仅为本发明的较佳实施例而已,并非用于限定本发明的保护范围。

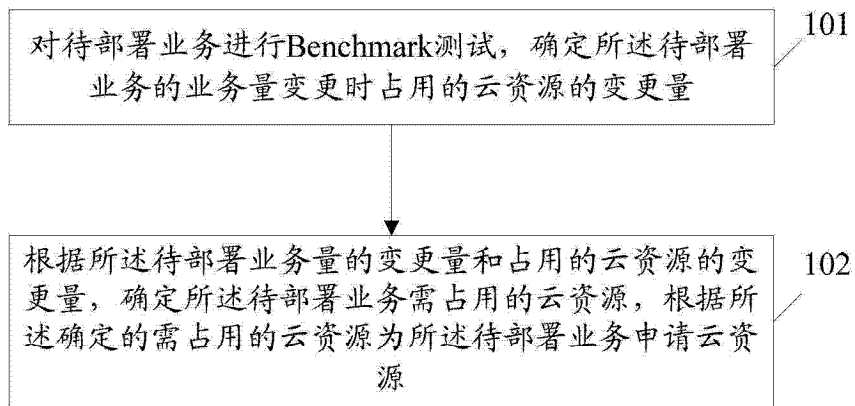


图 1

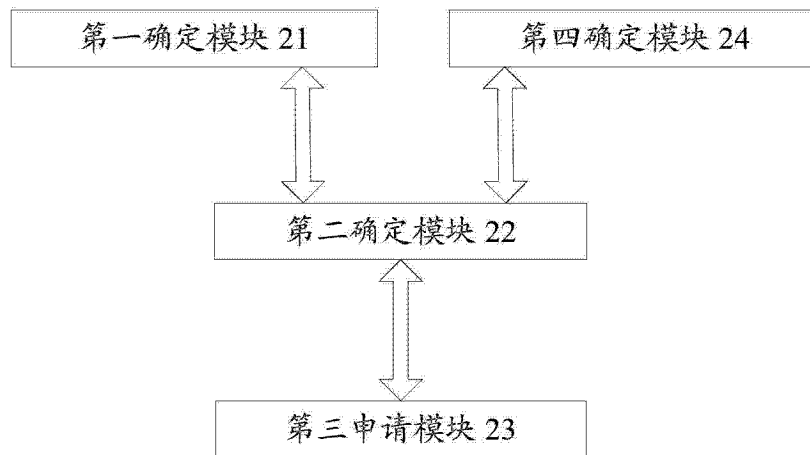


图 2