



业务支撑系统云化平台资源调度方案

牟宏蕾, 杨波, 郝亚飞

(中国移动通信集团设计院有限公司黑龙江分公司, 黑龙江 哈尔滨 150080)

摘要: 4G 时代, IT 新技术的发展以及运营商商业模式的巨大变化, 推动业务支撑系统的云化演进。为提升云资源的运维效率, 对云资源的动态调度进行了探索。基于 4 种业务场景有针对性地提出资源调度模式。

关键词: 云计算; 动态调度; 虚拟机

中图分类号: TN919

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2015371

Resource scheduling scheme of cloud platform for business support system

MOU Honglei, YANG Bo, HAO Yafei

Heilongjiang Branch of China Mobile Group Design Institute Co., Ltd., Harbin 150080, China

Abstract: In 4G era, the development of new technologies and the great changes in the business model of telecom operators promote the progress of business support system cloudization. In order to improve the efficiency of the operation, the dynamic scheduling of cloud resources was explored. A resource scheduling model was proposed based on four business scenarios.

Key words: cloud computing, dynamic scheduling, virtual machine

1 引言

4G 时代, 运营商在从语音经营向流量经营和数字化经营的转型过程中, 原有的客户、渠道、产品、合作伙伴等商业要素的内涵、外延均已不同。新的商业模式和信息新技术的发展, 均推动着业务支撑系统的云化演进, 如图 1 所示。

而云平台中普遍存在着资源浪费和闲置的情况。以黑龙江 O 域为例, 为保证系统的稳定性, 云平台根据忙时业务需求进行虚拟机的配置部署。各应用独立占用预分配的计算、存储和网络资源, 从而在业务非忙时产生大量资源浪费现象。另外, 应用的上线、更新、扩容等资源调整依赖人为操作, 浪费了大量的人力。

为了提高资源利用率, 提升云资源的运维效率, 本文针对业务支撑系统不同场景的需求对云资源的动态调度进行了探索。

2 资源动态调度方案

2.1 基于负载性能的调度方案

2.1.1 适用场景

网络支撑客服系统 (以下简称客服系统) Web 应用有着显著的负载变化特点, 在客服业务流程中, 随着客户的投诉量等变化, 系统的负载变化明显。在平时, 客服量基本保持在一定水平; 在忙时或者重大事件发生时, 客服系统负荷显著上升; 在闲时 (如晚间话务量减少时), 会出现负

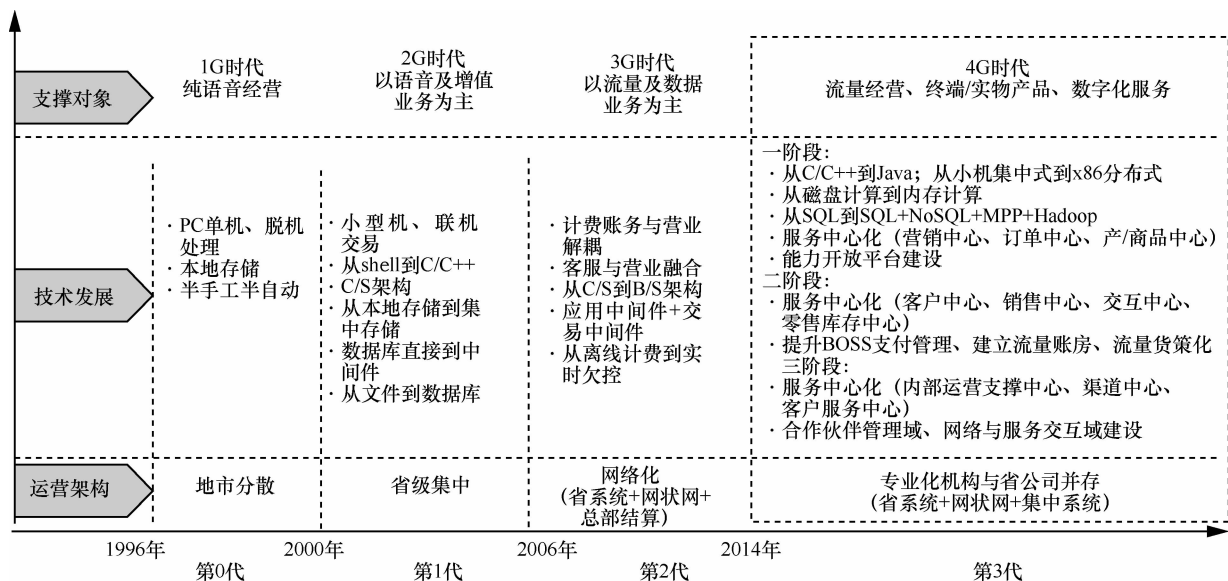


图1 业务支撑系统演进

载的大范围下降。

根据此类业务的负载变化特点,有针对性地动态调节资源供给量,以达到资源随着业务量变化联动的效果,保持资源使用率维持在较高水准,提高资源平均利用率。

2.1.2 调度方案

通过引入业务服务等级管理的概念,进行资源的动态调整。通过虚拟机服务等级的定义和各种选项的组合,影响虚拟机在创建和运行时的状态,从而区别对待不同的虚拟机。在业务申请时,可以根据自身业务的需求而选择不同的服务等级。

虚拟机在创建时如果指定了服务等级,则在资源调度系统中调度时,会将服务等级定义参数作为调度的参考因素之一。基础服务等级定义有如下种类。

- 计算类资源:包括资源池、集群、资源标签及内存预留。
- 网络类资源:包括设置网络、网络限速。
- 存储类资源:设置存储标签,对底层的存储设施进行标签化。在创建虚拟机时,通过不同存储标签的选择,影响虚拟机所使用的存储设施。根据存储设施的配置,可以设置如 SSD 存储、SCSI 存储、本地(local)存储等。
- 高可用性:虚拟机 HA。在创建 VM 时,可以选择启动 HA 功能。从而运行的 VM 在出现故障时,可以自动被云平台重建并迁移到其他 hypervisor 上面运行,避免关键性业务中断。

调度系统在调度虚拟机资源时,将根据服务等级的定义自动选择合适的 hypervisor 主机、网络以及存储类型。在分配 hypervisor 主机资源时,根据 SLA 的定义,预留主机 CPU 和内存资源,确保虚拟机运行性能。

在虚拟机运行过程中,引入了动态的资源分配预测和动态调整两种技术。云平台通过申请资源与实际使用率,获取实际的资源占用情况;再通过比较 hypervisor 主机当前的资源占用情况和上一个周期的资源占用情况,经过多点计算,获取 hypervisor 资源的使用趋势,实时评估 hypervisor 承载的虚拟机业务在未来几个周期内对资源的需求趋势。通过预留参数控制预留资源,确保 hypervisor 本身的性能稳定可靠。在动态资源分配预测中,如果 hypervisor 资源占用趋势趋于向上(即业务占用资源加大),则预留参数值较大,实时预留资源同时加大,保证业务新资源的申请可以即刻获得;如果资源占用趋势趋于向下(即业务释放资源),则预留参数值较小,实时预留资源减少。

2.1.3 调度流程

步骤 1 业务申请虚拟机:应用业务动态申请资源,实现业务直接与平台的对接,简化资源申请流程。

步骤 2 云管理平台根据预测算法计算分配量:通过应用在平台上设置的参数(业务等级、负载性能设定指标等),实现每个应用的资源个性化管理分配。

步骤 3 云管理平台选择适当的 hypervisor 运行业务虚拟机:根据应用设定的资源等级信息,对应开通相应的

资源,确保重点应用优质资源保障。

步骤 4 云管理平台在运行期监测业务虚拟机资源需求:平台根据应用的发展与负载性能的变化及时调整资源分配策略。

步骤 5 云管理平台动态调整虚拟机资源:平台根据调度策略及时更新应用的资源供给。

2.2 基于规划扩容调度方案

2.2.1 适用场景

客服系统 WAP 话单处理模块使用的是 Hadoop 架构,有着大数据应用的显著特点,即数据量不断扩大,需要资源的横向扩展能力很强,并且可以达到无限扩展能力。此类大数据类型应用需要资源的特点是资源在扩展后随时可加入,资源的使用情况是可预计的。基于这类应用的特点,给出基于规划扩容的动态调度机制。

2.2.2 调度方案

引入自动伸缩组的概念,将一组相关业务的虚拟机作为整体来看待。同时,通过监控和分析它们的性能指标以及各业务自动伸缩组配置的各种策略,实时调整伸缩组内的资源数量,如图 2 所示。

在使用横向扩展方式对应用性能进行管理时,应用在运行过程中主要有两种策略:扩容(scale-out)和收缩(scale-in)。

在配置扩容策略时,用户可以指定扩容的条件(规则引擎),配置扩容时进行 VM 创建或者进行物理机部署。同时,如果伸缩组配置了负载均衡器,还需要指定负载均衡器和配置负载均衡器的方式。

在配置收缩策略时,用户需要指定收缩时选取被回收

的节点的算法,如最新优先算法、最老优先算法或者随机算法。在回收了节点后,如果伸缩组配置了负载均衡器,也需要指定更新负载均衡的方式。

针对不同的业务类型,Web 类或者 Hadoop 分布式类需要在配置时选择不同的策略和后续处理方式。伸缩策略是一种插件机制,在有新类型的业务需要添加时,可以进行扩充。

2.2.3 调度流程

步骤 1 纳管 Hadoop 集群:通过应用监控采集 Hadoop 的运行数据。

步骤 2 以组为单位对 Hadoop 集群的整体运行情况进行分析和评估:如果 Hadoop 需要进行扩容,则利用物理机自动部署功能,从物理机资源池中分配合适的物理机资源,同时自动完成操作系统安装和网络配置等工作,并自动在新的物理机上完成 Hadoop 的各种业务服务的部署安装,使得 Hadoop 集群的资源实现自动扩展的功能。

2.3 基于业务申请调度方案

2.3.1 适用场景

有些业务系统具有周期性的特点,而且随着业务的增长,会不时地增加资源需求。这种业务不适用于负载性能调度和规划扩容调度方案,可通过 API 的调用和相应信息反馈的判定来解决这种业务系统的资源需求。

2.3.2 调度方案

业务系统根据自身相关任务的特点建立任务模型和长期有效的学习机制,通过对自身业务的不断学习,收集每个任务资源消耗,从而不断地更新任务模型,使得每个任务模型能持续优化,为每个任务建立一个比较准确的资源消耗值。

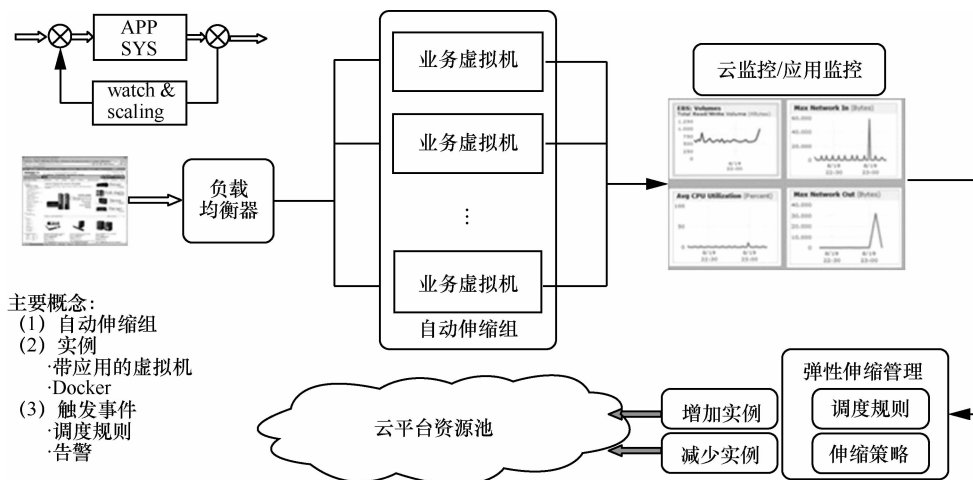


图 2 应用架构



通过对自身任务策略的分析和每个任务的资源消耗,对未来一段时间内的资源消耗进行预估,将预估值与系统的当前资源量进行对比,当预估值大于系统当前资源量时,将差值换算成虚拟机的数量,调用云平台的资源申请 API,实现资源的自动申请。如果预估值小于系统当前资源量时,将差值换算成虚拟机的数量,调用云平台的资源释放 API,实现资源的自动释放。系统架构如图 3 所示。

各模块功能如下。

任务管理:主要包括任务策略管理、任务生成、任务分发等功能。该模块根据任务策略生成相应的处理任务,并按照动态负载均衡策略把任务分发到虚拟机的任务执行(业务主机)节点;同时,对处理任务的执行状态进行监控。

模型管理:将各类处理任务抽象出任务特征值,对其进行建模,将任务学习管理模块监测、学习到的整个任务运行周期内资源消耗信息更新到任务模型管理的数据库中,实现对任务模型的管理,并支持向外扩展。

任务学习:根据任务调度管理模块中的任务策略以及任务基本信息,生成学习任务,下发至学习机,学习机针对学习任务定期上报采样数据,采样数据提供给任务模型管理模块进行任务模型的更新修正。

资源监控:主要实现虚拟机资源负载的实时监控功能、任务实时采样数据的收集。

资源决策:计算当前各虚拟机的运行负载、当前运行任务、剩余任务、预执行任务、任务模型等信息,通过综合评估判决是否进行 VM 资源的申请或释放。在实时监测各

任务执行节点的基础上,预测未来一段时间(时间可进行配置)的系统运行状况,实现虚拟机提前申请以及延后释放的功能。

资源调度:根据资源决策模块的分析结果,调用云平台提供的 API,实现虚拟机的申请和销毁,同时,将虚拟机的信息发送给任务调度管理模块,实现处理任务的动态调度,保证在申请和销毁虚拟机上的任务动态调整。

2.3.3 调度流程

步骤 1 任务生成:任务调度管理模块根据任务策略定时启动处理任务,并按照动态负载均衡策略把任务分发到云平台虚拟机的任务执行节点;对任务的执行状态进行监控,并通知任务学习模块进行任务的资源消耗监测。

步骤 2 任务建模:任务模型管理模块在每个应用任务第一次运行时,根据任务的特征建立任务模型,包括:任务标识、任务属性、采集时间点、资源 CPU 消耗、内存消耗、硬盘消耗等,并将相关信息通过数据表的形式保存到任务模型数据库中。首次执行的任务建立的新模型只有任务的相关信息,无资源消耗信息,资源消耗信息在任务学习模块中逐步采集更新。

步骤 3 任务学习:任务学习管理模块收到任务调度管理模块发送的任务启动消息后,根据任务标识,对在任务执行节点上的任务每 5 s(可配置调整)进行一次资源消耗监测,并记录监测到的数据,直至任务完成。

步骤 4 任务模型修正:任务学习管理模块记录每次监测到的资源消耗数据,计算出每次任务执行的时长、平

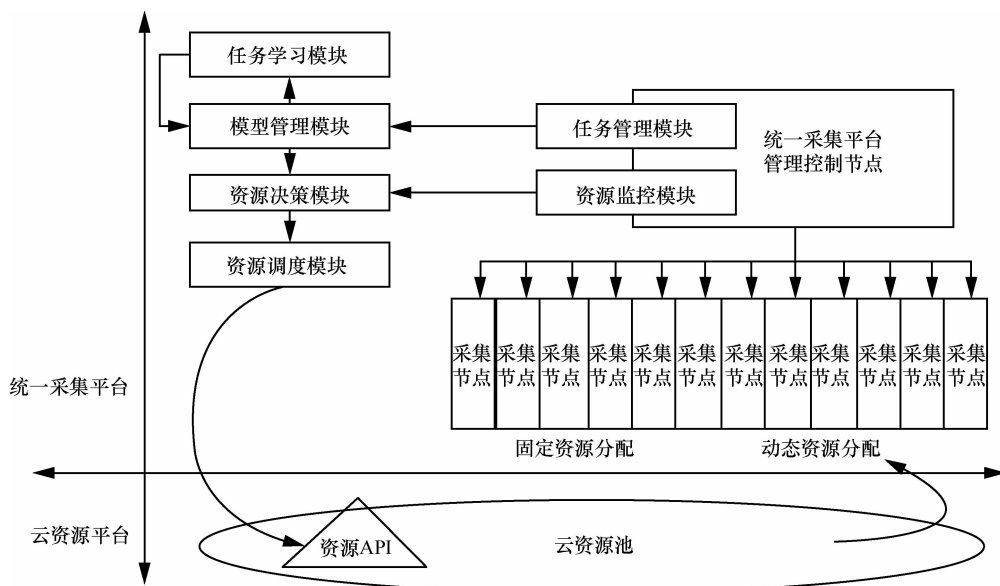


图 3 系统架构

均及最大 CPU 消耗、平均及最大内存消耗、硬盘消耗。通过任务模型修正的算法,通过计算可得到一个相对平稳正常的资源消耗预测数据,用于该任务的资源消耗估计,如果波动超过 10%(门限可配置调整),则作为学习异常的错误抛出,进行人工干预。

步骤 5 资源分析决策: 资源决策模块在任务到来时,决策模块根据任务模型管理模块中的任务资源二维表,按照任务资源消耗算法,得到任务资源消耗,将该任务资源消耗的数据预测和现存的资源进行对比,得到是否需要申请或释放的结果。资源占用指标暂定为 CPU 使用率、内存使用率、磁盘占用率,可后续进行扩展。

虚拟机资源申请: 未来每 5 min 的资源占用(连续 N 个 5 min 周期)均大于系统总资源(监控模块获取当前系统占用资源和空闲资源)时,则需要申请虚拟机的申请,超过当前系统资源的部分最终换算成需要申请的虚拟机台数。

虚拟机资源释放: 将 1 个大周期内(1 h)未来每 5 min 的峰值资源占用和系统总资源进行对比,如峰值资源按照任务资源消耗算法换算的虚拟机数量大于现有的虚拟机台数,则释放虚拟机。为了规避虚拟机释放过多导致的应用问题,系统设置一个虚拟机释放的门限 N , 设定 1 h 内最多减少 N 台虚拟机(N 可根据应用特点来动态设定)。

步骤 6 资源动态调度: 根据资源决策模块判断的申请/释放结果调用云平台的 API,实现虚拟机的申请和释放,同时,调用任务调度管理模块的消息,实现对申请/释放虚拟机的应用任务的动态分发调度,各应用服务的启动事先部署在云平台模板中,在申请虚拟机时,将任务需要的虚拟机模板发送到云平台上,这样,在云平台接收到虚拟机

申请后,会根据模板部署虚拟机,虚拟机使用模板附带的自动脚本来启动应用服务。

2.4 基于任务时间的调度方案

2.4.1 适用场景

有些业务有显著的时间特点,在不同的时间段有不同的资源需求,如网管网络支撑客服 Web 业务和桌面云业务,均在工作时间高发使用,而在晚上使用量会骤然下降。这类业务应用基于时间特点呈现的波峰波谷资源使用特点,需要使用时分复用的算法,将不同时间段内空闲出来的资源复用起来。

2.4.2 调度方案

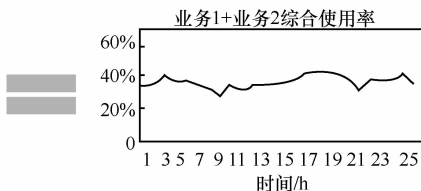
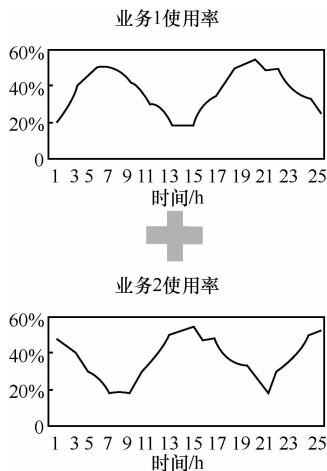
通过弹性伸缩组定义时间策略的方式进行时间维度的调度。在不同的时间段内,可以定义启动、停止、创建、删除虚拟机的策略,使得不同应用在不同时段使用相同的资源,达到资源高效利用的目的,如图 4 所示。

自动伸缩组是将逻辑上对等的一组机器资源组织成一个逻辑上的“大机器”。通过对资源、伸缩策略、负载评估算法、负载均衡等的配置和管理,提供横向的扩展或者收缩能力。

- 当业务配置的时间点到达时,则由时间策略管理模块发出策略触发事件。
- 伸缩组根据配置的方案,进行扩容。
- 伸缩组同时根据相关业务的时间重叠性,回收重叠业务已经分配的资源(如果业务的资源分配在扩容时间段内)。
- 当业务配置的回收策略时间点到达时,时间策略管理模块发出策略触发事件。
- 伸缩组根据配置的方案,进行资源回收。

业务特点:

- 业务负载有时间相关性
- 业务负载变化有规律
- 资源可根据时间进行计划性调整



云平台支撑:

- 业务时间策略配置
- 基于时间的资源分配
- 自动开关机、创建销毁 VM

图 4 基于任务时间的资源调度



- 伸缩组同时根据业务的时间重叠性,扩容重叠业务的资源分配(如果业务在扩容时间段内)。

2.4.3 调度流程

步骤 1 自动伸缩组配置时间策略。

步骤 2 应用 A 的时间策略触发:应用 A 根据业务需求,扩张资源,占据资源,提供服务。

步骤 3 应用 A 结束时间策略触发:应用 A 根据业务需求,减少资源占用。

步骤 4 应用 B 的时间策略触发:计算应用 A 资源池资源空闲量,创建应用 B 业务 VM,应用 B 占据资源运行。

步骤 5 应用 B 结束时间策略触发:应用 B 结束业务占用的 VM,释放占用的资源。

3 建设效果

3.1 基于负载性能的调度方案——网络支撑客服 Web 集群

实施基于负载性能调度后,Web 集群由从物理机迁移至虚拟机环境。Web 集群使用了基于负载性能调度对 Web 集群服务器 CPU、内存、每秒服务请求数量进行监控,当这些指标达到阈值上、下限时,自动调整 Web 集群节点数量。表 1 为实施前、实施后计算资源对比情况。

表 1 方案实施前后对比

阶段	服务器	分配量		使用率	
		CPU/核	内存/GB	CPU	内存
实施前	13	104	1 664	20%	14%
实施后	10	80	320	27%	56%

基于负载性能的弹性调度,只需要在被管理集群上部署监控模块,定时向云平台上报性能数据,由云平台根据弹性调度策略分析引擎进行评估,并进一步决定开通或回收虚拟机,完成弹性调度。应用系统只需要提供性能采集接口,无须增加应用系统额外的开发工作,适合在私有云大部分应用系统上进行推广。

3.2 基于规划扩容调度方案——统一采集集群

统一采集是网管系统的核心业务,因此,每个虚拟机 CPU、内存分配方式为 100%预留,以保证有足够的计算资源,承载突发业务流量。

实施基于业务申请调度时,创建了一组基于弹性调度的虚拟机,与原有统一采集集群并行运行。其资源动态调配比较复杂,需考虑到采集任务的队列长度、传输时延等业务运行指标。表 2 为实施前、实施后计算资源对比情况。

表 2 方案实施前后对比

阶段	虚拟机	分配量		使用率	
		CPU/核	内存/GB	CPU	内存
实施前	27	232	768	28%	23%
实施后	13	192	416	32%	59%

当应用系统计算资源调配不仅依赖于系统性能指标,而且依赖于业务运行指标,弹性调度算法相对复杂时,更适合使用基于业务申请的弹性调度。如果应用系统的业务运行指标可以方便地进行量化,可以将应用系统的业务运行 KPI 扩展至私有云平台,由私有云平台实现基于负载性能的弹性调度,以减少应用系统的开发、部署工作量。

3.3 基于业务申请调度方案——网络支撑客服 Hadoop 集群

在一般情况下,Hadoop 集群在运行一段时间之后,有资源扩充的需求,没有资源回收的需求。但在扩充 Hadoop 集群时,需要准备 IP 网络、存储网络环境,手工安装 Hadoop 节点,再将计算节点扩充至 Hadoop 集群,工作量大,操作复杂。

基于规划扩容的资源调度解决了这个问题,在 Hadoop 集群运行过程中,通过采集 Hadoop 运行性能信息,定期制作 Hadoop 集群节点需求趋势评估报告,由基础架构管理员根据评估报告制定扩容规划,扩充物理机计算节点至 Hadoop 集群,节省了大量的手工操作。

3.4 基于任务时间调度方案——桌面云调度与资源联动

在调度方案实施之前,桌面云系统使用详情见表 3。

表 3 方案实施前运行情况

时间	hypervisor	CPU/核	内存/GB	CPU 使用率	内存使用率
7:30-18:30	27	864	5 184	61%	63%
19:00-7:00	27	864	5 184	14%	59%

采用网管应用、桌面云联动方式:在桌面云系统业务闲时,运行网管 MR 补采、报表汇总等功能实时性要求不是非常高的应用。为保证二者之间的物理隔离,在同一个 hypervisor 上,网管业务、桌面云业务使用不同的物理网卡、FC 卡连接各自的 IP 网络、存储网络,并在夜间采取如下措施盘活计算资源:将空闲桌面休眠或关闭,释放 CPU、内存计算资源;集中迁移非空闲桌面应用至指定 hypervisor,释放其他 hypervisor CPU、内存计算资源;在业务人员上班前,恢复已休眠桌面或启动空闲桌面。表 4 为桌面云系统实施定时资源释放与回收之后的资源使用率性能。

表 4 方案实施后运行情况

时间	hypervisor	CPU/核	内存/GB	CPU 使用率	内存使用率
7:30-18:30	27	864	5 184	61%	63%
19:00-7:00	4	128	768	1%	8%

除桌面云系统外,基于任务时间调度可进一步推广至具有明显时间特性的资源池/应用系统,如业务支撑系统的 CRM、BOSS 等应用系统,主要集中在白天使用计算资源,而经营分析系统集中在夜间使用资源。这一类资源池/应用系统都适合采用基于任务时间的弹性调度,完成在时间上互补的应用系统相互填峰削谷,使计算资源错峰使用。

4 结束语

如何实现云资源使用的按需分配和动态管理是业务支撑系统云化中面临的重要问题,本文基于负载性能、规划扩容、业务申请及任务时间 4 个场景对资源动态调度方案进行了探索与实践,为今后业务支撑系统云化的资源调度提供了可借鉴的经验和思路。

参考文献:

[1] 宋维佳,马皓,肖臻,等. 虚拟化数据中心资源调度研究[J]. 广西大学学报, 2011(S1): 329-333.
SONG W J, MA H, XIAO Z, et al. A survey of resource scheduling for virtualized data center [J]. Journal of Guangxi University, 2011(S1): 329-333.

[2] 陈睦,黄黎明,李先锋. 云计算中虚拟机磁盘迁移时机优化策略[J]. 计算机工程与设计, 2014(2): 525-530.

CHEN M, HUANG L M, LI X F. Disk migration timing optimization mechanism in cloud computing [J]. Computer Engineering and Design, 2014(2): 525-530.

[3] 郭涛,刘菲军,杜垚,等. 云计算环境下虚拟机部署策略的优化[J]. 计算机应用研究, 2012(9): 3425-3431.

GUO T, LIU F J, DU Y, et al. Virtual machine deployment optimization in cloud [J]. Application Research of Computers, 2012(9): 3425-3431.

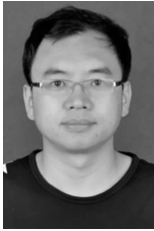
[作者简介]



牟宏蕾(1982-),女,中国移动通信集团设计院有限公司黑龙江分公司工程师,主要从事支撑网、数据网络方面的通信网络设计工作。



杨波(1975-),男,中国移动通信集团设计院有限公司黑龙江分公司黑龙江事业部技术质量负责人,主要从事数据、支撑、核心网专业方面的工程咨询设计工作。



郝亚飞(1984-),男,中国移动通信集团设计院有限公司黑龙江分公司工程师,主要从事业务支撑网方面的通信网络设计。