

民航云计算环境下的SLA保证策略研究

沈强¹, 杨健¹, 王家亮²

(1. 中国民航信息网络股份有限公司, 北京 100105;

2. 中国民航大学 计算机科学与技术学院, 天津 300300)

摘要:为了营造民航业应用云计算的环境,让航空公司、机场等用户将关键的业务迁移到云计算平台,细化服务品质,实现服务协议的精细化愈加重要。云计算资源的池化对信息系统底层资源的管理提出了更高要求,基于云计算的高效负载、服务等级的资源调度策略将有助于交付更符合民航业云用户需求的平台。为此,设计并建立了以业务系统为核心,以业务系统SLA定义与监控、业务系统资源和服务使用、业务系统自动化容量管理和资源调配为主要功能的资源管理体系,设计了动态负载均衡策略和资源调度策略并实现了面向SLA的业务管理系统。实际生产运行结果表明,所建立的虚拟机动态迁移技术实现了负载均衡,保证了SLA的动态资源分配过程建模及所实现的业务系统自动化管理均具有较好的应用效果。

关键词:云计算;民航;服务;资源;服务等级协议;负载均衡;动态迁移

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2017)06-0156-04

doi:10.3969/j.issn.1673-629X.2017.06.032

Research on SLA Assurance Strategy for Cloud Computing Environment of Civil Aviation

SHEN Qiang¹, YANG Jian¹, WANG Jia-liang²

(1. China TravelSky Holding Company, Beijing 100105, China;

2. College of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China)

Abstract: In order to create the cloud computing environment for the aviation industry, making the key business of airlines, airports and other users moved to cloud computing platform, it becomes more and more important to implement the refinement of service agreements. The higher request has been proposed for cloud computing resource pooling to the underlying resource management of information system. A reasonable resource scheduling strategy based on cloud computation with efficient workloads and high service grade is helpful to deliver the platform more suitable to the requests of civil aviation cloud users. Therefore, a resource management system has been designed and established which considers business system as key and takes SLA definition of business system and its monitoring, automatic management of the system capacity and resource allocation as its functions. The dynamic load balancing strategy and resource scheduling strategy has been designed and thus the business management system for SLA has been implemented. The results of practical production operations show that balancing load has been realized with the virtual machine live migration technology to guarantee the better effects for both the modeling of dynamic resource allocation process and the realized automatic management of business system.

Key words: cloud computing; civil aviation; services; resources; SLA; load balancing; live migration

0 引言

随着社会经济的迅速发展和民航信息服务市场容量的扩大,民航服务产业不断拓展,民航系统也逐步向更集中、更统一、更简便的云计算平台转变。随着公有云服务的快速发展,云提供商日益增多,这也给用户选择云服务时带来诸如安全、隐私以及可靠等方面的挑

战。相比过去传统的集中紧耦合的大型主机系统,云平台在带来灵活性的同时也带来了系统高稳定性、大规模密集交易性能的挑战,研究云提供商所保证服务的质量、满足民航云用户的需求是关键问题^[1]。文中将针对云计算环境下航空业信息化服务水平的技术保证手段进行探讨。

收稿日期:2016-05-24

修回日期:2016-09-07

网络出版时间:2017-04-28

基金项目:国家发改委云计算工程(发改办高技[2014]1799)

作者简介:沈强(1974-),男,研究方向为新技术在民航信息系统的应用。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20170428.1702.014.html>

为了保障云消费者所获得服务的质量,同时提升云计算提供商的信誉,达到维护云用户和云服务提供商双方权益的目的,需要云消费者和云计算服务提供商签订关于服务质量的协议来约定细节,这个合同称为服务等级协议(Service Level Agreement, SLA)。目前云计算提供商主要依靠两种处理方式:一种是基于云用户需求优化配置^[2]现有资源;另一种是通过增加、扩展现有资源的方式来保障 SLA^[3]。其中,优化配置现有资源基础又存在两种形式:一是基于多租户服务请求的优先级调度;二是云服务的部署优化和应用迁移。

民航业信息系统向来以大规模、高性能、高稳定著称,面对云计算这种带有规模大、集群式架构特性的服务器部署模式,云计算服务器等异构的基础设施和服务类型的多样性容易造成系统中业务承载不均衡,重载节点响应速度急剧下降导致系统无法提供服务,容易造成 SLA 违约。同时在云计算环境下,数据中心的服务基础资源的位置,主机的资源可用性、响应能力等因素容易使云数据中心在不同位置的服务资源提供服务的成本有较大区别^[4-5]。因此部署高效的负载均衡策略和资源调度策略对于保障云计算 SLA 至关重要。

在动态资源管理框架及相应的调度策略基础上^[6-7],针对民航信息系统的特征,基于选择适合民航云计算服务的迁移集合及如何实施迁移两个维度,优化改进了最小迁移次数(Minimization of Migrations, MM)策略,保证了云计算平台服务运行过程中的 SLA,同时又降低了因迁移带来的额外能源消耗。

1 基于虚拟机动态迁移技术的负载均衡策略

云计算平台中往往利用虚拟化技术实现计算资源、网络资源等资源管理的便捷性。虚拟化技术的最新发展趋势是支持虚拟机的动态迁移,而不影响运行于其上的应用程序。因此,在服务繁忙时,将资源利用率超过上限阈值物理机上的部分虚拟机迁出,以确保不违反云用户和云计算服务提供商签订的 SLA。但是,关于虚拟机的具体迁移实施,需要合理的策略,否则,很容易产生额外的能耗,甚至会造成违反服务的 SLA。

动态迁移(Live Migration),是指在保证虚拟机中各服务正常运行的同时,将虚拟机系统从一台宿主物理主机移动到另一台宿主物理主机的过程^[8]。

虚拟机运行所需的镜像文件和相关配置文件都存放在共享存储中,如 SAN(Storage Area Network)或 NAS(Network Attached Storage)^[9]。在迁移过程中,只需将虚拟机数据执行状态迁移到宿主物理主机中即

可完成迁移。通过共享存储,可以确保不对云用户业务造成影响,基于虚拟机的动态迁移技术实现零宕机条件下的服务器维护。方便系统管理员能够动态、灵活地利用云计算平台的计算能力。在 HA 中支持网络及存储双路心跳,在发生 HA 时,支持被隔离服务器的自动重启提供服务。

设计虚拟机动态迁移策略的关键是处理好与虚拟机迁移相关的服务器内存、CPU 运行状态、网络连接和虚拟设备状态、存储设备等物理基础资源^[10-12]。目前比较常见的虚拟机动态迁移算法是内存预拷贝(pre-copy)迁移算法。最具代表性的是 VMWare 和 Xen-Source 分别推出的 VMotion^[13] 迁移以及 Xen Motion^[14] 迁移。这两种迁移方式都适用于局域网并且有相似的方案用于内存和网络连接的迁移。其中,预拷贝迁移的主要工作过程如下:虚拟机监控器将所有内存页标记为脏页面并拷贝到目的主机上;然后迭代拷贝。这里迭代是指采用几轮增量同步,上一轮传输过程中被修改过的页面在本次迁移。当需要传输的内存小于阈值或者迭代次数超过设置的最大迭代次数时停止迭代。最后,挂起源主机上的虚拟机,拷贝 CPU 状态和最后一轮产生的脏页面到目的主机,恢复目的主机上的虚拟机。民航公共云平台动态迁移方案里实现了预拷贝迁移方法。如果目的主机出现故障,则终止迁移过程,继续在源主机上运行虚拟机。

预拷贝迁移的主要目的在于实现了虚拟机的动态迁移,同时可以确保民航公共云服务平台的高可靠性,保证民航云用户的服务等级。

2 保证 SLA 资源调度的动态资源分配过程建模

云计算环境中的资源调度是指请求任务在多个数据中心寻找满足目标的资源。根据资源调度时间的不同分为在线调度和批调度。在线调度是指根据任务到达情况,实时寻找合适的可利用资源,如果无法找到,就等待或者拒绝。目前的云计算环境中,基于在线调度策略的研究主要集中在以满足用户的各种资源请求为目的,却较少关注云计算服务提供商的利益。需要从在线调度的方向研究有效的任务调度算法,使得在满足用户 SLA 的前提下,实现云计算资源提供商的收益最大化。对于 LaaS 层的调度,需要考虑的是,云计算环境中因服务器的位置、计算等资源的可用性等客观因素会造成不同资源提供的成本存在较大差异。面对接踵而至的任务,不同的任务 SLA 规定的 QoS 指标也不相同,Laas 层的调度策略既要考虑满足任务的 SLA,又要确保接收任务所获得的收益是最大的。

在云环境下,将云用户的动态需求表示成用户请

求,采用过程化模拟方法,将云计算资源提供服务的过程模拟成虚拟机的分配过程,确保云计算服务供应商在遵循 SLA 协议的前提下达到利润最大化,增强云计算服务提供商的竞争力。其中,云计算资源分配过程如下:

云环境中可以对虚拟化技术抽象后的计算、网络、存储等基础设施资源进行统一集中化管理。实现了逻辑计算资源、逻辑存储资源以及逻辑网络资源的监控功能、管理功能和调度功能,提高了云计算平台中逻辑资源的自动化管理水平。

鉴于用户对云计算平台中的计算资源和存储资源的需要差异大、变化大,所以需要预先建立标准的计算资源模板和存储资源模板。在云计算环境中,可以通过虚拟化技术,实时动态地分配、更改用户云主机的逻辑计算资源和逻辑存储资源。

云计算环境内,可以为不同的用户或者用户内的不同部门之间分配安全的逻辑隔离网络,实现逻辑网络资源分配的灵活性。通过云平台的网络组件 agent,实现对用户内部不同网络之间的策略控制和对外部的策略控制。

而资源分配与高可用模块分配的最大目的是要准确地计算和分布业务系统所需要的各种资源,包括硬件资源和软件资源,以保证业务系统所依赖的整个 IT 基础架构链条尽量冗余,且冗余组件间完全没有相互影响。其冗余设计如图 1 所示。在进行分布策略设计时,主要考虑如下:

(1)任何一个软硬件组件都是不可靠的,均可能会出现故障。

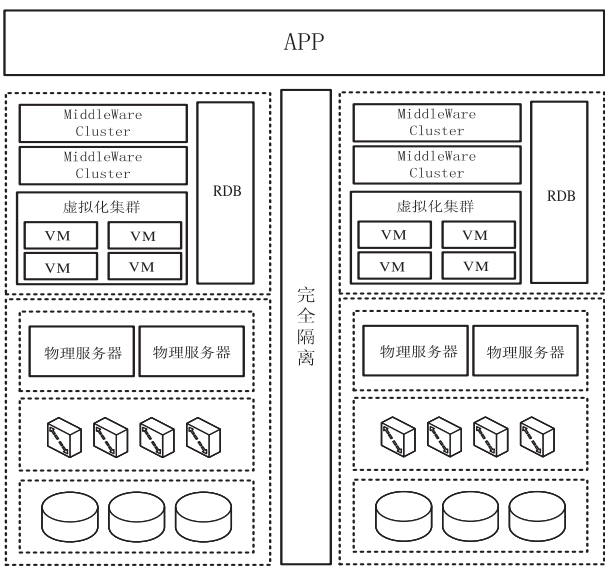


图 1 冗余设计图

(2)任何两个组件只要存在联系,即有可能发生故障传染。

(3)在特定业务场景下,如要求严格事物一致性的业务场景,对无法实现完全冗余的资源(如数据库集群组件本身)必须经过严格的可靠性评估,选择经过长时间运行检验,可靠性可以达到 99.99% 的组件;对其中涉及的组件,还需进行运行管理加强设计,即在不配运维资源时(如监控资源),运维资源也要做到冗余设计,以保证可以迅速发现和修复故障。

资源分配与高可用保证模块的一个理想分配结果如图 2 所示。一个业务系统将运行在两个完全隔离的 IT 基础架构之上,可见,任何一个架构中的任何一个组件的故障都不会影响业务系统的正常运行。

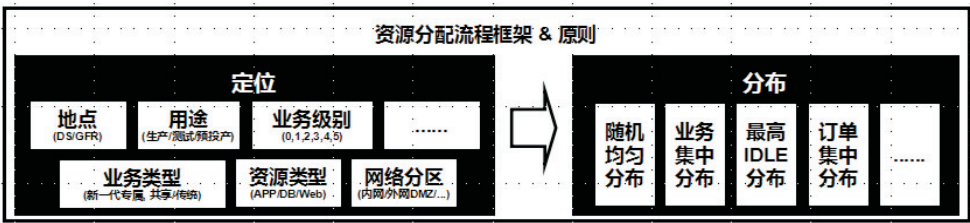


图 2 资源分配图

同时,资源分配与高可用保证模块还负责实现业务系统对资源分布的一些其他要求,如统一业务系统的资源尽可能集中等。因此,在具体实现上,资源分配模块如图 2 所示,主要由定位和分布两部分功能构成。定位是根据用户的各种输入,获得满足用户要求的各种资源,作为用户最终使用资源的备选;而分布则是根据多个分配策略,确定用户最终将使用哪几个具体的底层资源。但无论选择何种分布策略,资源分配均将首先保证用户所使用的一组资源必须分布在不同的底层资源上,从而保证了任何一个底层资源的失效不会对用户整体的数据质量产生影响。

对传统云计算中心的资源调度策略进行分析,提出了云平台管理系统的模型及系统实现,以保证 SLA 的服务水平。

3 面向 SLA 的业务管理系统实现

为更加灵活地管理业务系统,适应不同的业务系统分类场景,业务系统采用多级结构,分为“产品线”-“模块”-“业务系统”三个维护级别,并为各个级别的维护定义简单高效的维护流程,具体业务管理流程体系如图 3 所示。

流程的基本原则为:

(1)每个企业的超级管理员为整个业务系统管理体系负责,控制所有的业务系统架构及属性变更。

(2)每个级别的业务对象均需定义若干不同功能的管理角色,负责该业务对象不同方面的维护和管理工作,如:总体负责、维护工作等。同时,为支持简单的管理体系,不同的角色可以由同一个人来担任。

(3)上级业务结构的管理者可以定义下级业务的结构和属性,但其定义的变更需要超级管理员审批,以减少业务定义的随意性,保证业务系统的精简和相对稳定。

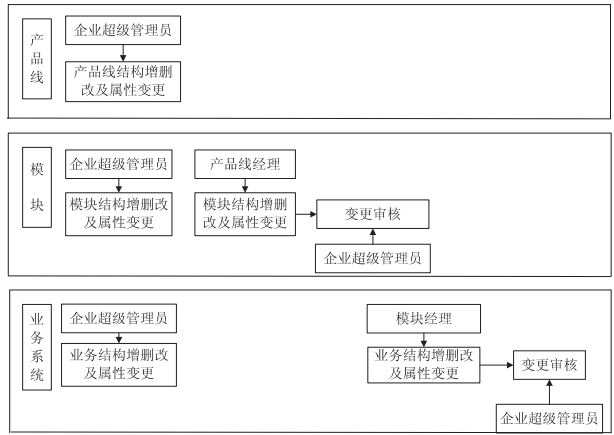


图3 在线用户管理

如图3所示,产品线中的企业超级管理员负责产品线结构的增删改以及属性的变更;模块中的企业超级管理员和产品线经理同时负责模块结构的增删改,同时企业超级管理员还负责变更的审核;业务系统中,企业超级管理员和产品线经理也共同负责模块结构的增删改,同时企业超级管理员还负责变更的审核。

通过虚拟机动态迁移技术实现负载均衡,以及对保证SLA资源调度的动态资源分配过程进行优化建模,实现面向SLA的业务管理系统。

4 结束语

民航云计算环境下的SLA保证策略非常重要,通过研究设计动态负载均衡策略和资源调度策略并实现面向SLA的业务管理系统,各个在线企业可以很容易地梳理自己的应用系统,提高了用户标准化管理各个业务系统的SLA水平,实现了对平台所提供的各种资源

的高效、有序管理,降低了管理和资源使用成本,提升了资源使用效率。

参考文献:

[1] 史玉良,王捷.一种多租户云数据存储缓存管理机制[J].计算机研究与发展,2014,51(11):2528-2537.

[2] 赵秀涛,张斌,张长胜.一种基于服务选取的SBS云资源优化分配方法[J].软件学报,2015,26(4):867-885.

[3] Castro P H P, Barreto V L, Corrêa S L, et al. A joint CPU-RAM energy efficient and SLA-compliant approach for cloud data centers[J]. Computer Networks, 2016, 94: 1-13.

[4] 罗军舟,金嘉晖,宋爱波,等.云计算:体系架构与关键技术[J].通信学报,2011,32(7):3-21.

[5] 谭一鸣,曾国荪,王伟.随机任务在云计算平台中能耗的优化管理方法[J].软件学报,2012,23(2):266-278.

[6] Antonescu A F, Braun T. Simulation of SLA-based VM-scaling algorithms for cloud-distributed applications[J]. Future Generation Computer Systems, 2016, 54: 260-273.

[7] Serrano D, Bouchenak S, Kouki Y, et al. SLA guarantees for cloud services[J]. Future Generation Computer Systems, 2016, 54: 233-246.

[8] Dastjerdi A V, Buyya R. An autonomous time-dependent SLA negotiation strategy for cloud computing[J]. Computer Journal, 2015, 58(11): 3202-3216.

[9] Marudhadevi D, Dhatchayani V N, Sriram V S S. A trust evaluation model for cloud computing using service level agreement[J]. Computer Journal, 2015, 58(10): 2225-2232.

[10] 刘鹏程,陈榕.面向云计算的虚拟机动态迁移框架[J].计算机工程,2010,36(5):37-39.

[11] 常德成,徐高潮.虚拟机动态迁移方法[J].计算机应用研究,2013,30(4):971-976.

[12] 熊安萍,徐晓龙.基于内存迭代拷贝的Xen虚拟机动态迁移机制研究[J].计算机科学,2013,40(8):63-65.

[13] Guan Haibing, Yao Jianguo, Qi Zhengwei. Energy-efficient SLA guarantees for virtualized GPU in cloud gaming[J]. IEEE Transactions on Parallel and Distributed Systems, 2015, 26(9): 2434-2443.

[14] Alrebeish F, Bahsoon R. Implementing design diversity using portfolio thinking to dynamically and adaptively manage the allocation of web services in the cloud[J]. IEEE Transactions on Cloud Computing, 2015, 3(3): 318-331.