

基于云计算的高校信息化应用建设研究

朱铨, 赖宏图

福建工程学院 现代教育技术中心, 福建工程学院 福建省大数据挖掘与应用技术重点实验室

摘要: 本文基于高校信息化建设对云计算技术的具体需求, 分别从 IaaS 层、中间件层以及应用层探讨了实现云平台动态弹性扩展的关键技术以及仍有待解决的关键问题。基于云计算技术对教务管理系统进行了系统结构改造, 提出了云平台资源调度策略, 测试结果以及实际运行结果均表明云计算的动态弹性扩展特性极大地提高了教务管理系统的性能, 满足了系统在高并发情况下的应用需求。本文的研究, 可为具有类似特征的校园信息化应用的建设提供参考。

关键字: 云计算; 高校信息化; 信息化应用; 动态弹性扩展; 资源调度策略

引言

为解决海量数据的处理难题, Amazon、Google 等公司于 2006 年先后提出了云计算的概念^[1, 2], 目前已经得到工业界、学术界的广泛关注, 各国政府也纷纷将云计算列为国家战略, 投入了巨大的财力和物力用于云计算技术的研究以及云平台的部署和应用。

云计算所具有的弹性扩展、资源池化、按需服务等特点, 非常适合数据量弹性变化的应用环境, 比较典型的应用场景如淘宝的“双十一”大促销、12306 网站在春运期间的售票等。12306 网站其 2014 年的尖峰日 PV 值为 144 亿次, 2015 年春运期间峰值则达到了 297 亿次, 解决这个问题的关键是其构建了一个可扩展的云平台架构, 能够按业务需求随时部署所需要的资源, 从而确保了在高峰压力下业务的持续性^[3]。

目前, 各大高校陆续进入到了“智慧校园”的建设进程中, 旨在以各种应用服务系统为载体, 将教学、科研、管理和校园生活等进行充分融合, 从而为学校师生、校友等提供更为人性化的校园信息化服务。为此, 本文将深入研究云计算的弹性扩展等关键技术, 并以教务管理系统为例, 探讨在云计算条件下的高校校园信息化应用建设模式。

一、云计算与高校信息化

云计算是信息技术应用模式和服务模式创新的集中体现, 也是信息技术产业发展的重要方向, 能够推动经济社会的创新发展, 是世界各国积极布局、争相抢占的新一代信息技术战略制高点。为此, 国务院在 2015 年 1 月专门下发了《国务院关于促进云计算创新发展培育信息产业新业态的意见》。教育部发布的《教育信息化十年发展规划(2011-2020 年)》中, 则提出了具体的发展目标,

即: 通过充分整合现有资源, 采用云计算技术, 形成资源配置与服务的集约化发展途径, 构建新型的高校信息化发展模式。

目前, 随着我国高校信息化应用的增多以及信息化利用水平的不断提升, 高校的数据资源也在快速增长, 但由于普遍缺乏统一的管理规范、数据格式差异、业务部门自建应用系统等诸多问题, 使得大量的数据无法有效共享和利用, 各个信息化应用之间也存在诸多的“信息孤岛”^[4]。通过云计算技术搭建统一的数据处理平台, 实现数据信息的共享与协同, 并通过大数据挖掘技术实现对海量数据信息数据的动态、实时处理, 从而为高校管理部门以及师生用户提供各类信息服务。同时云计算平台通过虚拟化等技术, 整合服务器、存储、网络等硬件资源, 优化系统资源配置, 为大数据处理提供弹性可扩展的处理平台, 实现应用部署的灵活性; 能大大提升资源利用率, 降低总能耗以及运维成本^[5, 6]。

因此, 如何对高校信息化应用进行整合并对教育信息化大数据进行处理、分析、挖掘和利用, 将是未来高校信息化发展的关键, 而云计算技术以其自动化 IT 资源调度和快速部署以及优异的弹性扩展等优势, 将成为解决这一问题的重要技术手段^[7, 8]。

二、云平台动态弹性扩展关键技术

下图 1 所示的是我校正在规划和建设的校园信息化云平台逻辑体系结构图, 总体思路是分层架构, 服务为主, 基于教师、学生、校友、访客以及公众等不同角色的生命周期提供相应的校园信息化信息服务。云计算技术的应用将会带来诸多优势, 首先, 云计算的数据存储和维护能力能为逐年积累的校园大数据提供了存放和处理能力; 其次, 通过采用分布式计算等手段则可以实

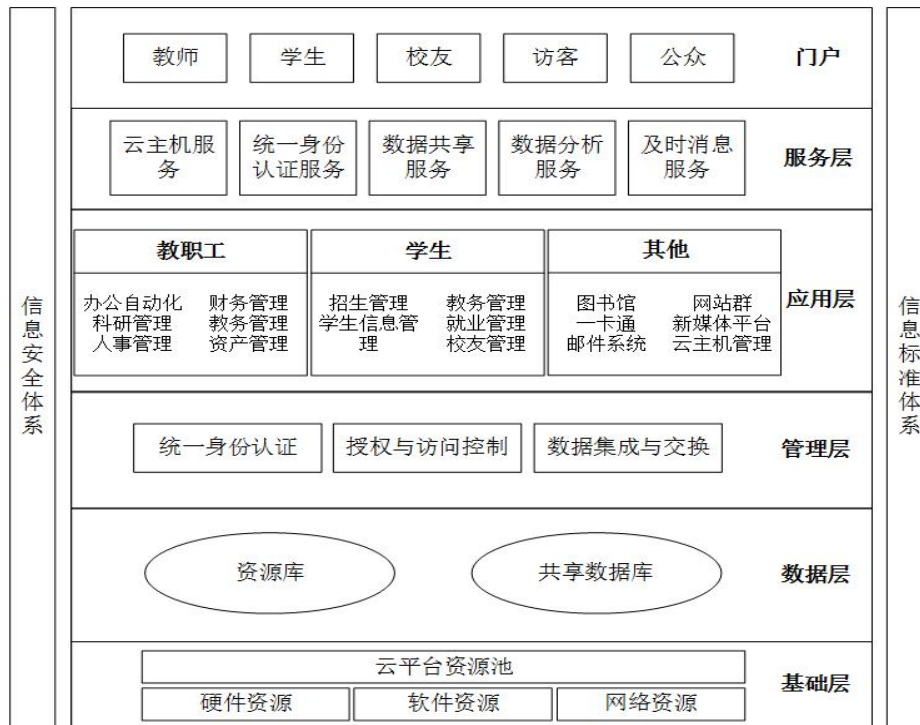


图 1 校园信息化云平台逻辑体系结构图

现高性能的并行处理能力，为高并发的校园信息化应用以及基于校园大数据挖掘的分析和云服务提供了条件保障；第三，云计算所提供的弹性扩展能力能够保证校园信息化应用的高可用性和扩展性，从而保证用户的访问和使用体验。

对于其中较为关键的弹性扩展技术，主要包括垂直扩展（节点 CPU、内存等升级）和水平扩展（节点新建和销毁）；而对于如何利用云计算技术的优势进行校园信息化应用建设，从而实现校园信息化应用的弹性扩展，则需要从校园信息化云平台的 IaaS 层、中间件层以及应用层进行全面考虑。

1. 基于虚拟化技术的 IaaS 层弹性扩展机制

虚拟化技术作为实现云计算平台的关键技术，主要分为两个层面：物理资源池化和资源池管理。其中物理资源池化是把物理设备（包括服务器、存储、网络以及安全设备等）由大化小，将一个物理设备虚拟为多个性能可配的最小资源单位，包括完全虚拟化、准虚拟化以及操作系统虚拟化三种模式，常见的相关技术包括 Xen、OpenVZ、KVM、VirtualBox、Hyper-V 以及 VMWare vSphere 等。资源池管理则是通过对集群中虚拟化后的最小资源单位进行管理，根据资源的使用情况和用户对资源的申请情况，按照一定

的策略对资源进行灵活分配和调度。目前较流行的技术包括 OpenStack、OpenNebula、CloudStack、Nimbus、Eucalyptus 以及 VMware vCloud 等，开源技术方面，OpenStack 呈现强劲的发展趋势，也比较适合于大数据处理等应用场景^[9]；商业软件方面 VMware vCloud 则是比较好的选择。

快速地实现资源池中虚拟机节点的弹性伸缩是实现云平台处理性能动态扩展的必要前提。目前虚拟机的资源分配（包括 CPU、内存和网络等）多采用静态的方式，即根据应用需要和管理员的经验对虚拟机的资源进行预估、划分，分配的资源都是固定的，这样对于单个虚拟机来说容易存在资源过度分配和分配不足的问题，而现有虚拟化技术对已分配资源（如 CPU、内存）进行调整则需要虚拟机重启后才能生效。因此，如何实现单个虚拟机所分配资源（主要是 CPU、内存）的动态伸缩是目前实现云计算弹性扩展的一个关键难题。

在实际应用中，校园信息化应用对软硬件资源的需求是不断变化的，资源池中虚拟机节点预留的多少都会造成资源的短缺或浪费。通过虚拟机复制、迁移等方法可以动态地调整集群的资源分配，而虚拟机节点的调度策略和算法，也是当前的一个研究热点。虚拟机复制技术是以现有有虚

拟机为模板（镜像）复制出多个具有相同物理资源、操作系统和应用环境的虚拟机，基于该技术，可以快速和动态地实现虚拟机节点的部署。

通过虚拟机迁移，即将一台虚拟机从一台主机移动到另外一台主机上，可以实现对服务器负载的均衡。常见的包括两种模式：虚拟机非在线迁移和在线迁移，前者一般采用“停止—拷贝”机制，迁移时间较短，但会造成停机（虚拟机挂起）；后者一般采用“预拷贝”机制，能降低迁移过程中的停机服务时间，可保证虚拟机服务的连续性和不间断性，有效地实现了动态的弹性扩展^[10, 11]。一些主流的虚拟化软件均提出了自己的虚拟机在线迁移技术，如 VMWare 公司的 VMotion、Xen 公司的 Live Migration 等。

2. 云平台中间件层弹性扩展机制

分布式提供了云平台环境下一一种可行的文件存储模式和计算模式，目前 Google 公司提出的 GFS（Google File System）分布式存储技术和 MapReduce 分布式开发框架已经成为了云计算的核心理念，Apache 软件基金会则根据 MapReduce 模式和 GFS 系统研发实现了开源的分布式并行编程框架 Hadoop 和 GFS 的开源实现 HDFS（Hadoop Distributed FileSystem）^[12]，已成为 Yahoo!、Facebook、百度、淘宝等国内外知名企业所采用的云计算核心数据处理技术，并进一步催生了众多相关项目，如能够支持任意存取的 NoSQL（Not only SQL）数据库 MongoDB、Redis、Aerospike 等；能够进行实时数据处理的 Storm、S4、Spark 等技术；能够克服关系数据库 I/O 瓶颈的分布式内存数据库 Gemfire 等^[7, 8, 13]。

对大数据进行处理，目前普遍采用的是类似 MapReduce 的分布式处理开发框架以解决海量数据处理问题的伸缩性需求^[14-16]；而针对大数据处理的时效性问题，传统的磁盘 I/O 成为影响处理性能的瓶颈，内存数据库通过将数据在内存中进行运算，避免了 I/O 操作，从而获得了优于传统磁盘数据库的处理性能。而在云计算环境下，分布式数据处理技术通过将海量数据分散在云平台的多个服务器节点上，并发执行数据运算任务，将能进一步提高云平台的处理性能。因此，分布式内存数据库非常适合海量数据的实时处理^[17]。

3. 基于容器引擎的应用层弹性扩展机制

LXC（Linux Container，Linux 容器）是一种内核虚拟化技术，可以提供轻量级的虚拟化。容

器技术通过将由单个操作系统管理的资源划分到孤立的组中，实现在孤立的组之间平衡有冲突的资源使用需求，常见容器技术包括 Linux-Vserver、OpenVZ 和 FreeVPS 等。相较于传统虚拟化技术，LXC 具有性能损耗更小、安全隔离、易于安装和移除等优势^[18, 19]。

Docker 是一个基于 LXC 的开源的应用容器引擎，可以实现把应用以及依赖包打包到一个可移植的容器中，并能在任何流行的 Linux 机器上运行，也可以实现虚拟化解决方案^[20]。其相比虚拟机最明显的特点就是启动和销毁的速度快、资源占用小，非常适用于构建隔离的标准化的运行环境，从而实现应用的弹性扩展^[21, 22]。Docker 革新了基于云计算平台的应用发布方式，从而吸引了 IBM、Google、RedHat、百度等业界知名公司的关注和技术支持。

基于 Docker 进行应用部署，将能有效地减少应用实例的部署时间，实现校园信息化云平台处理性能的弹性扩展，同时也方便进行应用、服务的部署和迁移，从而为部分实时性要求较高的校园信息化应用提供了一种可行的解决方案。

三、基于云计算技术的教务管理系统建设

在校园信息化应用中，教务管理系统的作用较为重要，它负责着学校排课、学生选课、教师成绩录入、学生成绩查询等一系列功能，是学校教学运行及数字化校园建设的基础，其运行的稳定与否直接影响到全校师生的教学活动。而教务管理系统的运行，也具有一定的特点：平时大部分时期内，师生主要是进行成绩、课表等的常规查询，并发连接数量不会太高，只需要较少的服务器资源即可支撑访问需求；而在学生选课期间，则会出现高并发的情况，常常会使教务管理系统运行出现宕机、拒绝服务等异常现象。

为此，本文将基于云计算技术对教务管理系统进行升级、改造，通过云计算的动态弹性扩展技术实现教务管理系统在高并发环境下的正常访问，而在平时的低峰时期，云平台的资源则可以回收并提供给其他校园信息化应用进行使用，从而提高整个校园信息化云平台的资源利用效率。

1. 教务管理系统架构

作者学校使用的教务管理系统的逻辑架构如图 2 所示，软件采用 B/S+C/S 架构，其中将 Web 层和 APP 层部署在 Windows IIS 服务上，实际上为两层架构，逻辑上为三层；功能上包括教师管理端、学生服务端和后台数据库。教务管理系统

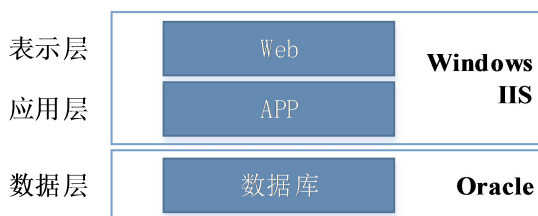


图 2 教务管理系统逻辑架构图

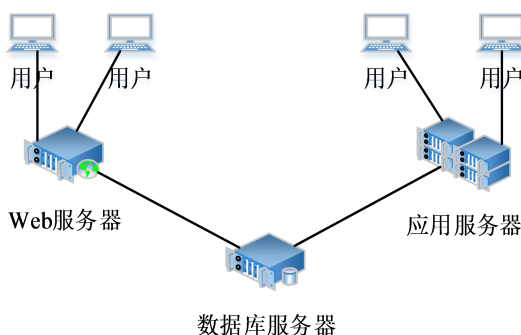


图 3 教务管理系统架构示意图

的实际部署架构如图 3 所示。

通过对教务管理系统的软件架构进行分析，可知其性能瓶颈主要会发生在前端 Web 表示层和数据库层。而通过实际测试发现，访问瓶颈主要在于 Web 服务器的并发连接数限制上。前端 Web 表示层提供了用户访问的接口，在单机的情况下由于系统的并发连接数限制，每逢选课时就会因为大量学生并发访问而导致系统进程锁死或者无法访问。因此，Web 层可以采用云计算技术，通过其动态弹性扩展能力，在系统访问量较大的时候，通过新增节点来提高系统的并发访问量；同时，通过负载均衡设备在各个节点服务器上進行负载均衡。基于此，改进后的基于云计算技术的教务管理系统架构如图 4 所示。

2. 运行测试

本文基于图 4 所示的教务管理系统架构，进行相关测试及运行，平台相关功能的产品选型如下所示。

其中，CAS 平台的组件可实现对操作系统、

中间件和应用程序等内容进行监控。本文设置的监控项包括：CPU 使用率（cpu，%）、内存使用率（memory，%）以及并发连接数（link_connections），监控频率设置为 30 秒。由于并发连接数是影响用户访问教务管理系统体验的主要因素，因此可在负载均衡设备上设置基于网络访问连接数对云平台进行负载均衡。

本文设计的云平台资源调度策略如下表 2 所示。

在实际情况下，由于 Step 2 中新建虚拟机节点所需时间较长，可预先复制一定数量的虚拟机节点供调用，Step 3 中的删除虚拟机节点也可以不执行。

针对在云平台上部署完毕的教务管理系统运行环境，本文首先通过 LoadRunner 压力测试软件，对云平台以及云平台资源调度策略进行测试。本文将单个 Web 服务器节点配置为 4 个 vCPU、4G 内存，通过 LoadRunner 模拟学生登录选课，测试每秒 16 个学生同时登录的场景，持续模拟

表 1 云平台系统功能选型

功能	产品
计算资源	刀片服务器
资源池化	CAS 平台
数据库中间件	Oracle 10g
负载均衡设备	L5000-C

表 2 云平台资源调度策略

云平台资源调度策略
Step 1: 初始化, 分别设置启动节点的监控阈值 cpu_A 、 $memory_A$ 、 $link_connections_A$, 以及关闭节点的阈值: cpu_B 、 $memory_B$ 、 $link_connections_B$; 资源池默认启动一个节点, 并设置此节点不加入到资源调度策略, 防止节点被关闭;
Step 2: if $cpu > cpu_A$ or $memory > memory_A$ or $link_connections > link_connections_A$ then 新建虚拟机节点; 启动虚拟机节点;
Step 3: else if $cpu < cpu_B$ or $memory < memory_B$ or $link_connections < link_connections_B$ then 关闭该虚拟机节点; 删除虚拟机节点;

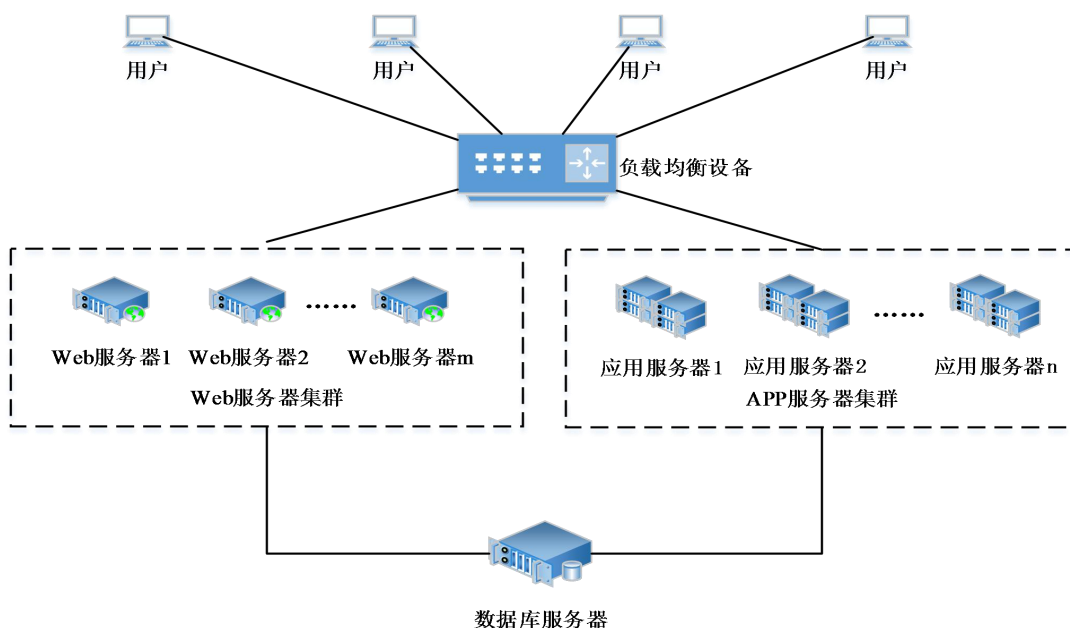


图 4 基于云计算技术的教务管理系统架构示意图

6000 学生登录。测试结果如图 5 所示。

从图中可以看出, 30 秒到 40 秒之间出现了一个响应时间的波峰, 表示响应时间随着登录学生数量的增加而增多。由于连接数达到资源扩展的阈值, 云平台资源调度策略启动了第二台虚拟机节点, 而在第二个节点启动后, 响应时间迅速下降并呈现平稳态势。在实际运行中, 该云平台通过 20 台 Web 服务器支撑了 7000 人同时在线选课。

改进后的基于云计算技术的教务管理系统, 可通过云平台资源调度策略实现实时、动态地扩展虚拟机节点数量, 从而保证了教务管理系统的

能的无缝扩容, 为全校师生提供了良好的使用体验。而且该云平台的资源可以实现自动回收, 在教务管理系统空闲时, 资源可以自动释放并提供给其他应用使用, 从而极大地提高了云平台的利用效率。

四、结语

本文基于高校信息化建设对云计算技术的具体需求, 分别从 IaaS 层、中间件层以及应用层探讨了实现云平台动态弹性扩展的关键技术, 并对国内外研究现状进行深入比较分析, 提出了尚需解决的 IaaS 层资源调度策略、基于容器引擎的应用层服务资源调度策略等关键技术问题。在

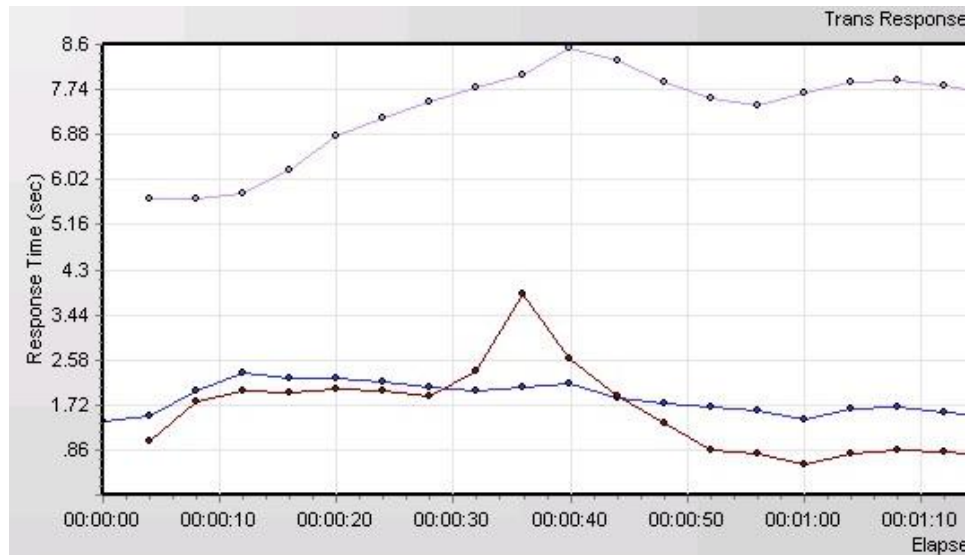


图 5 云平台压力测试结果

此基础上，基于云计算技术对教务管理系统进行了系统架构改造，提出了云平台资源调度策略，并进行了相关测试研究，测试结果以及实际运行结果均表明云计算的动态弹性扩展特性极大地提高了教务管理系统的性能，满足了系统在高并发情况下的应用需求。而除教务管理系统之外，还有一系列校园信息化应用具有类似特点，如网站群系统、招生系统等在平时的负载不会太高，而在一些特殊的时间（如招生、校庆等）其用户访问量会出现激增。因此，本文的研究，也可为相关校园信息化应用的建设提供参考。

参考文献

- [1] 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.
- [2] 邹复民, 蒋新华, 胡惠淳, 等. 云计算研究与应用现状综述[J]. 福建工程学院学报. 2013, 11(03): 231-242.
- [3] Csdn. 技术揭秘 12306 改造 [EB/OL].2015-2-18.http://www.csdn.net/article/2015-02-10/2823900.
- [4] 徐琦. 基于大数据的高校数据整合模式研究[J]. 中国教育信息化. 2015(15): 60-63.
- [5] 王意洁, 孙伟东, 周松, 等. 云计算环境下的分布存储关键技术[J]. 软件学报. 2012, 23(4): 962-986.
- [6] 殷波, 王颖, 孟洛明, 等. 综合迁移成本和通信成本的云计算节能策略[J]. 北京邮电大学学报. 2012, 35(1): 68-71.
- [7] 覃雄派, 王会举, 李芙蓉, 等. 数据管理技术的新格局[J]. 软件学报. 2013, 24(02): 175-197.
- [8] 申德荣, 于戈, 王习特, 等. 支持大数据管理的NoSQL系统研究综述[J]. 软件学报. 2013, 24(08): 1786-1803.
- [9] 李春艳, 张学杰. 基于高性能计算的开源云平台性能评估[J]. 计算机应用. 2013, 33(12): 3580-3585.
- [10] CHANCHIO K, THAENKAEW P. Time-Bound, Thread-Based Live Migration of Virtual Machines[C].// 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, CCGrid 2014. Chicago, IL, United states: IEEE Computer Society, 2014:364-373.
- [11] KEJIANG Y, XIAOHONG J, DAWEI H, et al. Live Migration of Multiple Virtual Machines with Resource Reservation in Cloud Computing Environments[C].// 2011 IEEE International Conference on Cloud Computing (CLOUD). Washington: IEEE, 2011:267-274.
- [12] SHVACHKO K, KUANG H, RADIA S, et al. The Hadoop Distributed File System[C].// MSST '10 Proceedings of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST). Washington, DC, USA: IEEE Computer Society, 2010:1-10.
- [13] 林子雨, 赖永炫, 林琛, 等. 云数据库研究[J]. 软件学报. 2012, 23(05): 1148-1166.

- [14] DEAN J, GHEMAWAT S. MapReduce: simplified data processing on large clusters[J]. Communications of the ACM. 2008, 51(1): 107-113.
- [15] 覃雄派, 王会举, 杜小勇, 等. 大数据分析——RDBMS与MapReduce的竞争与共生[J]. 软件学报. 2012, 23(1): 32-45.
- [16] 王珊, 王会举, 覃雄派, 等. 架构大数据:挑战、现状与展望[J]. 计算机学报. 2011, 34(10): 1741-1752.
- [17] 秦秀磊, 张文博, 魏峻, 等. 云计算环境下分布式缓存技术的现状与挑战[J]. 软件学报. 2012: 1787-1803.
- [18] DUA R, RAJA A R, KAKADIA D. Virtualization vs Containerization to Support PaaS[C].// 2nd IEEE International Conference on Cloud Engineering, IC2E 2014. Boston, MA, United states: Institute of Electrical and Electronics Engineers Inc., 2014:610-614.
- [19] XAVIER M G, NEVES M V, ROSSI F D, et al. Performance Evaluation of Container-Based Virtualization for High Performance Computing Environments[C].// 2013 21st Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, PDP 2013. Belfast, United kingdom: IEEE Computer Society, 2013:233-240.
- [20] DOCKERI.AboutDocker[EB/OL].2014-10-21.http://docs.docker.com/
- [21] BOETTIGER C. An introduction to Docker for reproducible research[J]. ACM SIGOPS Operating Systems Review. 2015, 49(1): 71-79.
- [22] SEO K, HWANG H, MOON I, et al. Performance Comparison Analysis of Linux Container and Virtual Machine for Building Cloud[J]. Advanced Science and Technology Letters. 2014, 66: 105-111.