

云计算:概念、技术及应用研究综述

方巍^{1,2,3} 文学志^{1,2} 潘吴斌^{1,2} 薛胜军^{1,2}

摘要

近年来,云计算作为一种新的服务模式已成为计算机科学领域的一个研究热点.首先从定义、部署模式、特征和关键技术等方面对云计算进行了综述,然后剖析了当前主流云计算参与企业和应用实例,最后讨论了目前云计算发展存在问题与机遇,旨在为了解云计算当前发展状况、研究趋势以及科学地评估云计算的运营和使用提供参考.

关键词

云计算;虚拟化;服务;平台

中图分类号 TP393.4

文献标志码 A

0 引言

云计算(cloud computing)从当初提出时备受人们质疑发展到如今成为学术界和产业界争相研究的热点.根据美国市场调查公司 Gartner 评出的 2011 年 10 大战略技术中,云计算当仁不让地排名首位.云计算的快速发展预示着该技术可以带来美好的应用前景和更多的经济收益.所谓云计算,简单地说就是以虚拟化技术为基础,以网络为载体,以用户为主体为其提供基础架构、平台、软件等服务为形式,整合大规模可扩展的计算、存储、数据、应用等分布式计算资源进行协同工作的超级计算服务模式^[1].虚拟化为云计算实现提供了很好技术支撑,而云计算可以看作是虚拟化技术应用的成果.在过去的几年里,已经出现了众多云计算研究开发小组,如谷歌(Google)到 IBM、微软(Microsoft)、亚马逊(Amazon)、EMC、SUN、HP、VMware、Salesforce、Alisoft、华为、百度、阿里巴巴、中国电信等知名 IT 企业纷纷推出云计算解决方案.同时,国内外学术界也纷纷就云计算及其关键技术相关理论进行了深层次的研究.本文就云计算及其关键技术与主流产品做对比研究,并提出了一些在气象云计算领域的应用方向,希望能为云计算的进一步研究提供参考.

本文第 1—2 节阐述了云计算相关概念、特征以及云计算的主要关键技术,即 3 种交付模式、4 种部署模式、5 个基本特征;第 3 节重点讨论了云计算应用产品,分析比较了目前几个具有代表性的云计算平台、研究组织和研发方向;第 4 节提出了目前云计算研究存在的问题与发展机遇;第 5 节总结全文并对云计算研究做出展望.

1 云计算概述

1.1 云计算相关定义

目前有一种流行的说法来解释“云计算”为何被称为“云”计算:在互联网技术刚刚兴起的时候,人们画图时习惯用一朵云来表示互联网,因此在选择一个名词来表示这种基于互联网的新一代计算方式的时候就选择了“云计算”这个名词.虽然这个解释非常有趣和浪漫,但是却容易让人们陷入云里雾里,不得其正解.自 2007 年 IBM 正式提出云计算的概念以来,许多专家、研究组织以及相关厂家从不同的研究视角给出了云计算的定义.目前关于云计算的定义已有上百种.而维基百科对云计算的定义也在不断更新,前后版本的差别非常

收稿日期 2011-09-27

资助项目 国家自然科学基金(60970015);江苏省高校自然科学基金面上项目(11KJB520010);江苏高校优势学科建设工程资助项目.

作者简介

方巍,男,讲师,系统分析师,博士,主要研究方向为智能信息处理、数据挖掘和云计算.
fangwei@nuist.edu.cn

1 南京信息工程大学 江苏省网络监控中心,南京,210044

2 南京信息工程大学 计算机与软件学院,南京,210044

3 中创软件工程股份有限公司博士后工作站,济南,250014

大. 据 2011 年给出的最新定义^[2]:“云计算是一种能够将动态易扩展的虚拟化资源软件和数据通过互联网提供给用户的计算方式,如同电网用电一样,用户不需要知道云内部的细节,也不必具有管理那些支持云计算的基础设施。”

伯克利云计算白皮书的定义^[3]:云计算包括互联网上各种服务形式的应用以及数据中心中提供这些服务的软硬件设施. 应用服务即 SaaS (Software as a service, 软件即服务), 而数据中心的软硬件设施即所谓的云. 通过量入为出的方式提供给公众的云称为公共云, 如 Amazon S3 (Simple Storage Service)、Google App Engine 和 Microsoft Azure 等, 而不对公众开放的组织内部数据中心的云称为私有云.

美国标准化技术机构 NIST 定义^[4]:云计算是一种资源利用模式, 它能以方便、友好、按需访问的方式通过网络访问可配置的计算机资源池(例如网络、服务器、存储、应用程序和服务), 在这种模式中, 可以快速供应并以最小的管理代价提供服务.

Sun 公司^[5]认为, 云的类型有很多种, 而且有很多不同的应用程序可以使用云来构建. 由于云计算有助于提高应用程序部署速度, 有助于加快创新步伐, 因而云计算可能还会出现我们现在无法想象到的形式. 作为创造“网络就是计算机”(The network is the computer™) 这一短语的公司, Sun 公司认为云计算就是下一代的网络计算.

还有一些有关云计算的定义^[6-7]. 云计算的定义各有侧重, 众说纷纭. 笔者认为: 云计算是一种大规模资源共享模型, 它是以虚拟技术为核心技术, 以规模经济为驱动, 以 Internet 为载体, 以用户为主体, 按照用户需求动态地提供虚拟化的、可伸缩的商业计算模型. 更确切地说, 云计算是一种服务模式而不单纯是一种技术. 在云计算模式下, 不同种类的 IT 服务按照用户的需求规模和要求动态地构建、运营和维护, 用户一般以即用即付 (pay as you go) 的方式支付其利用资源的费用. 网络中的应用服务通常被称作 SaaS, 而数据中心的软硬件设施即资源池也就是云 (cloud). “云”是一些可以自我维护 and 管理的虚拟计算资源, 通常是一些大型服务器集群, 包括计算服务器、存储服务 and 宽带资源等^[6]. 云计算的核心思想是资源池, 池的规模可以动态扩展, 而使用资源池中的丰富的软硬件资源可以像煤气、水和电一样, 取用方便, 费用低廉. 当云以即用即付的方式提供给公众的时候, 称其为公共云 (public cloud), 这里出售的是

效用计算. 当前典型的效用计算有 Amazon Web Services (<http://aws.amazon.com/ec2/2009>)、Google AppEngine (<http://appengine.google.com>) 和微软 Azure (<http://www.microsoft.com/azure/>). 不对公众开放的企业或组织内部数据中心的资源称作私有云 (private cloud).

总之, 云计算是一种方便的使用方式和服务模式, 通过互联网按需访问资源池模型(例如网络、服务器、存储、应用程序和服务), 可以快速和最少的管理工作为用户提供服务. 云计算是并行计算 (parallel computing)、分布式计算 (distributed computing) 和网格计算 (grid computing) 等技术的发展. 云计算又是虚拟化 (virtualization)、效用计算 (utility computing) 的商业计算模型, 它由 3 种服务模式, 4 种部署模式和 5 点基本特征组成.

1.2 云计算的 3 种交付模式

云计算的服务层次可分为将基础设施作为服务层、将平台作为服务层以及将软件作为服务层, 市场进入条件也从高到低. 目前越来越多厂商可以提供不同层次的云计算服务, 部分厂商还可以同时提供设备、平台、软件等多层次的云计算服务^[6,8]. 图 1 为云计算服务类型.

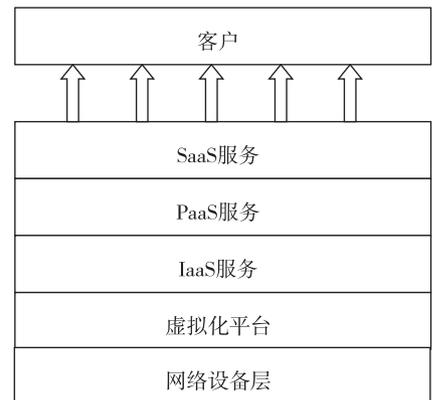


图 1 云计算服务类型

Fig. 1 Service types of cloud computing

1) 基础设施即服务 (Infrastructure as a Service, IaaS). 通过网络作为标准化服务提供按需付费的弹性基础设施服务, 其核心技术是虚拟化. 可以通过廉价计算机达到昂贵高性能计算机的大规模集群运算能力. 典型代表如亚马逊云计算 AWS (Amazon Web Services) 的弹性计算云 EC2 和简单存储服务 S3, IBM 蓝云等.

2) 平台即服务(Platform as a Service, PaaS). 提供给客户的是将客户用供应商提供的开发语言和工具(例如 Java, python, . Net)创建的应用程序部署到云计算基础设施上去. 其核心技术是分布式并行计算. PaaS 实际上指将软件研发的平台作为一种服务, 以 SaaS 的模式提交给用户. 典型代表如 Google App Engine(GAE)只允许使用 Python 和 Java 语言, 基于称为 Django 的 Web 应用框架调用 GAE 来开发在线应用服务.

3) 软件即服务(Software as a Service, SaaS). 它是一种通过 Internet 提供软件的模式, 用户无需购买软件, 而是租用服务商运行在云计算基础设施上的应用程序, 客户不需要管理或控制底层的云计算基础设施, 包括网络、服务器、操作系统、存储, 甚至单个应用程序的功能. 该软件系统各个模块可以由每个客户自己定制、配置、组装来得到满足自身需求的软件系统. 典型代表如 Salesforce 公司提供的在线客户关系管理 CRM(Client Relationship Management) 服务, Zoho Office, Webex, 常见的还有 Email 等.

1.3 云计算的 4 种部署模式

1) 私有云(private cloud). 云基础设施是为一个客户单独使用而构建的, 因而提供对数据、安全性和服务质量的最有效控制. 私有云可部署在企业数据中心中, 也可部署在一个主机托管场所, 被一个单一的组织拥有或租用.

2) 社区云(community cloud). 基础设施被一些组织共享, 并为一个有共同关注点的社区服务(例如任务、安全要求、政策和遵守的考虑).

3) 公共云(public cloud). 基础设施是被一个销售云计算服务的组织所拥有, 该组织将云计算服务销售给一般大众或广泛的工业群体, 公共云通常在远离客户建筑物的地方托管, 而且它们通过提供一种像企业基础设施进行的灵活甚至临时的扩展, 提供一种降低客户风险和成本的方法.

4) 混合云(hybrid cloud). 基础设施是由 2 种或 2 种以上的云(私有, 社区或公共)组成, 每种云仍然保持独立, 但用标准的或专有的技术将它们组合起来, 具有数据和应用程序的可移植性(例如, 可以用来处理突发负载), 混合云有助于提供按需和外部供应方面的扩展.

1.4 云计算的 5 个特征

无论是广义云计算还是狭义云计算, 对于最终用户而言, 均具有如下特征^[2].

1) 按需自助式服务(on-demand self-service). 用户可以根据自身实际需求扩展和使用云计算资源, 具有快速提供资源和服务的能力. 能通过网络方便地进行计算能力的申请、配置和调用, 服务商可以及时进行资源的分配和回收.

2) 广泛的网络访问(broad network access). 通过互联网提供自助式服务, 使用者不需要部署相关的复杂硬件设施和应用软件, 也不需要了解所使用资源的物理位置和配置等信息, 可以直接通过互联网或企业内部网透明访问即可获取云中的计算资源. 高性能计算能力可以通过网络访问.

3) 资源池(resource pooling). 供应商的计算资源汇集在一起, 通过使用多租户模式将不同的物理和虚拟资源动态分配多个消费者, 并根据消费者的需求重新分配资源. 各个客户分配有专门独立的资源, 客户通常不需要任何控制或知道所提供资源的确切位置, 就可以使用一个更高级别抽象的云计算资源.

4) 快速弹性使用(rapid elasticity). 快速部署资源或获得服务. 服务商的计算能力根据用户需求变化能够快速而弹性地实现资源供应. 云计算平台可以按客户需求快速部署和提供资源. 通常情况下资源和服务可以是无限的, 可以是任何购买数量或在任何时候. 云计算业务使用则按资源的使用量计费.

5) 可度量的服务(measured service). 云服务系统可以根据服务类型提供相应的计量方式, 云自动控制系统通过利用一些适当的抽象服务(如存储、处理、带宽和活动用户帐户)的计量能力来优化资源利用率, 还可以监测、控制和管理资源使用过程. 同时, 能为供应者和服务消费者之间提供透明服务.

2 云计算的关键技术

云计算作为一种新的超级计算方式和服务模式, 以数据为中心, 是一种数据密集型的超级计算^[9]. 它运用了多种计算机技术, 其中以编程模型、数据管理、数据存储、虚拟化和云计算平台管理等技术最为关键. 下面分别介绍云计算的一些关键技术.

2.1 编程模型

MapReduce 作为 Google 开发的 Java、Python、C++ 编程模型^[7], 是一种简化的分布式编程和高效的调度模型, 应用程序编写人员只需将精力放在应用程序本身, 使云计算环境下的编程十分简单. 而关于集群的处理问题, 包括可靠性和可扩展

性,则交由平台来处理^[10]. MapReduce 模式的思想是通过“Map(映射)”和“Reduce(化简)”这样 2 个简单的概念来构成运算基本单元,先通过 Map 程序将数据切割成不相关的区块,分配(调度)给大量计算机处理,达到分布式运算的效果,再通过 Reduce 程序将结果汇整输出,即可并行处理海量数据.图 2 介绍了一个用 MapReduce 编程模型处理大数据集的具体过程^[7].简单地讲,云计算是一种更加灵活、高效、低成本、节能的信息运作的全新方式,通过其编程模型可以发现云计算技术是通过网络将庞大的计算处理程序自动分拆成无数个较小的子程序,再由多部服务器所组成的庞大系统搜索、计算分析之后将处理结果回传给用户.通过这项技术,远程的服务供应商可以在数秒之内,达成处理数以千万计甚至亿计的信息,达到和“超级电脑”同样强大性能的网络服务.

以气象行业的中尺度气象预报为例,中尺度气象预报模式有着惊人的计算量,同时由于气象预报的特点决定了其极高的实时性,要求模式必须在指定的时间内完成运算.另一方面,人们对气象预报的精度提出了越来越高的要求,目前预报精度从几

百千米、几十千米提高到几千米,而这大幅度提高了模式的计算量.数值气象预报对计算的这一需求,靠单个 CPU 或普通的计算机根本不可能完成,必须利用并行计算.一方面,将模式预报软件通过消息传递或者共享存储的方式并行化,另一方面需要高性能并行计算机.目前绝大部分中尺度气象预报模式都已经完成了并行化,如 MM5、WRF、Grapes 既支持 MPI 消息传递并行,又支持 OpenMP 共享存储并行.采用云计算中的 MapReduce 编程模型可以使气象部门特别是市县级气象部门能够享受到超级计算机计算处理能力,而不需要购置大量的基础设施.

2.2 海量数据分布存储技术

云计算系统采用分布式存储的方式存储数据,用冗余存储的方式保证数据的可靠性.云计算系统中广泛使用的数据存储系统是 Google 的 GFS 和 Hadoop 团队开发的 GFS 的开源实现 HDFS^[6,8].GFS 即 Google 文件系统(Google File System),是一个可扩展的分布式文件系统,用于大型的、分布式的、对大量数据进行访问的应用.GFS 的设计思想不同于传统的文件系统,是针对大规模数据处理和

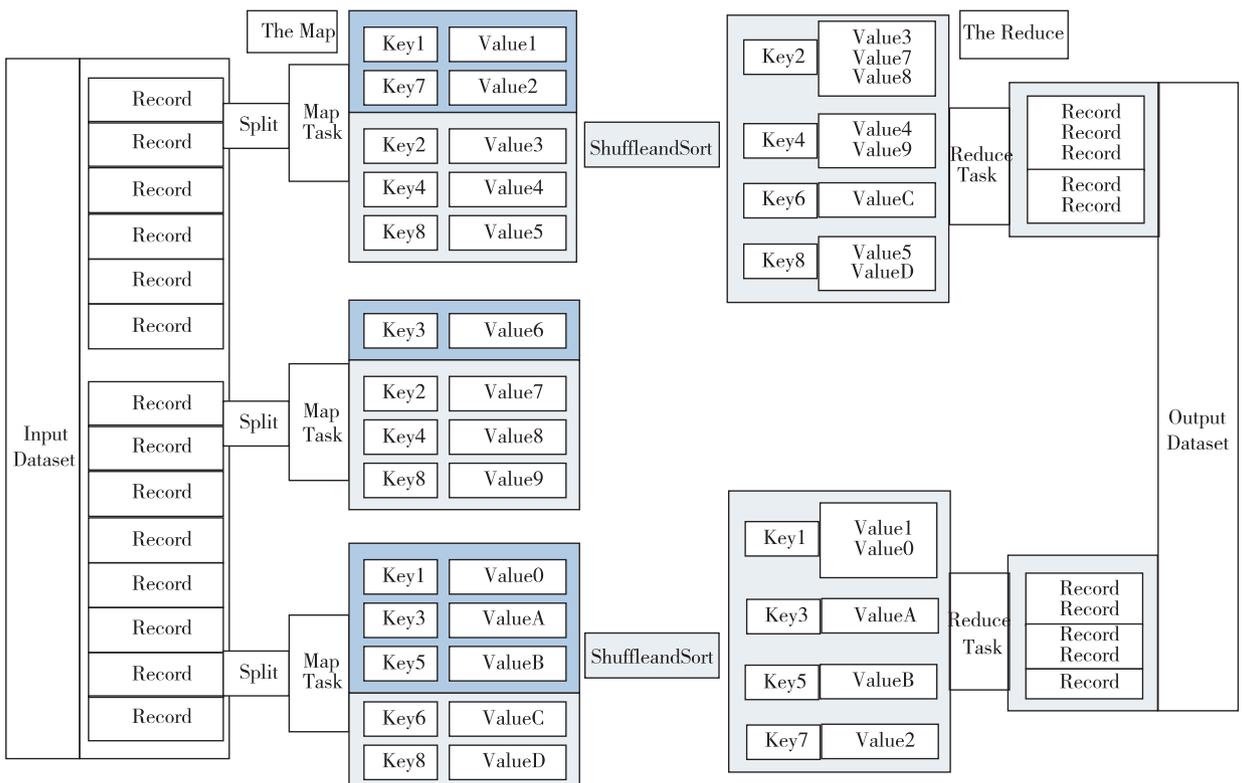


图 2 MapReduce 编程模型执行过程

Fig. 2 Executing process of MapReduce

Google 应用特性而设计的. 它虽然运行于廉价的普通硬件上,但可以提供容错功能. 它可以给大量的用户提供总体性能较高的服务. 一个 GFS 集群由一个主服务器 (master) 和大量的块服务器 (chunkserver) 构成,并被许多客户 (client) 访问^[11]. 主服务器存储文件系统所有的元数据,包括名字空间、访问控制信息、从文件到块的映射以及块的当前位置. 它还控制系统活动范围,如块租约 (lease) 管理,孤立块的垃圾收集,块服务器间的块迁移. 主服务器定期通过心跳 (HeartBeat) 消息与每一个块服务器通信,并收集它们的状态信息.

以气象云存储服务为例,目前随着自动站、雷达、雨量标校站、卫星站等的建设,气象资料数据也在与日俱增. 目前,数据存储仍以观测点和气象资料接收设备终端为主. 而云计算是由第三方服务商提供计算与存储等资源,并负责运行和维护,用户只需要通过终端工具接入系统,即可获得所需的服务. 这就是说,大家把气象资料存储在第三方提供的存储资源上,不需要因为存储资源不够而去购买设备,只需向服务提供商购买存储服务即可. 笔者设计的一个大规模气象云存储服务系统体系结构如图 3 所示. 客户与主服务器的交互只限于对元数据的操作,所有数据方面的通信操作都直接通过应用容器和云文件系统与主服务器联系,这样就可以大大提高系统的运行效率,防止主服务器负载过重. 为了使得整个结构更加清楚明了,部分非重要的模块(如访问控制和任务调度)并没有在图 3 中反映出来. 如其他云计算系统一样,本文的云存储服务系统是部署在计算机集群之上的.

2.3 海量数据管理技术

海量数据管理是指对大规模数据的计算、分析和处理,如各种搜索引擎. 以互联网为计算平台的云计算能够对分布的、海量的数据进行有效可靠地处理和分析. 因此,数据管理技术必需能够高效地管理大量的数据,通常数据规模达 TB 甚至 PB 级. 云计算系统中的数据管理技术主要是 Google 的 BT (BigTable) 数据管理技术,以及 Hadoop 团队开发的开源数据管理模块 HBase 和 Hive,作为基于 Hadoop 的开源数据工具 (<http://appengine.google.com>),主要用于存储和处理海量结构化数据. BT 是建立在 GFS, Scheduler, Lock Service 和 MapReduce 之上的一个大型的分布式数据库,与传统的关系数据库不同,它把所有数据都作为对象来处

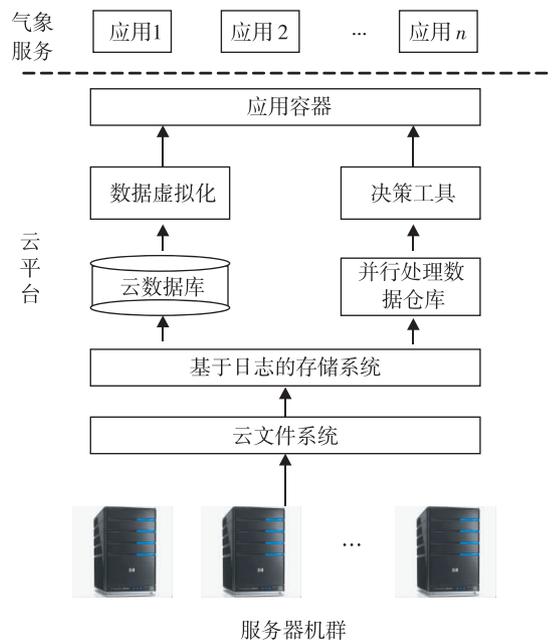


图 3 气象云存储服务系统体系结构
Fig. 3 System architecture of meteorological cloud storage service

理,形成一个巨大的表格,用来分布存储大规模结构化数据.

Google 的很多项目使用 BT 来存储数据,包括网页查询,Google Earth 和 Google 金融. 这些应用程序对 BT 的要求各不相同:数据大小(从 URL 到网页到卫星图像)不同,反应速度不同(从后端的大批处理到实时数据服务). 对于不同的要求,BT 都成功地提供了灵活高效的服务. 图 4 是整个 BigTable 的存储服务体系结构^[12].

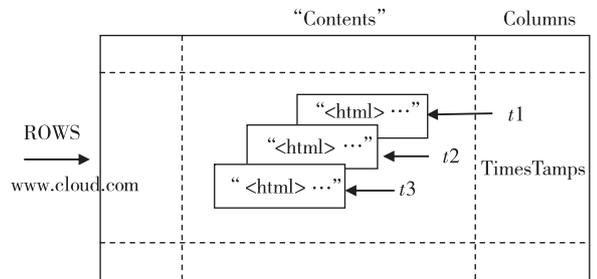


图 4 存储服务体系结构
Fig. 4 System architecture of storage service

同样以气象行业为例,采用云计算技术的气象业务云平台可以与既有的区域/省级共享平台和相关领域/行业共享平台耦合,实现云拓展. 区域/省级信息共享平台与气象业务云平台通过专用的宽带网络连接,遵循统一的元数据标准、数据分类标

准、共享分级标准和用户授权规则. 相关领域/行业信息共享平台主要布设于水利、海洋、国土等地球系统的相关业务部门以及相关科研院所和大学等,以吸纳和整合各圈层数据资源,为用户提供完整的信息共享服务. 该层次平台与气象业务云平台采取统一的元数据标准和数据分类标准,或者采用双方认可的数据交换规则与语义翻译表单,但可以各自拟定用户授权规则,以“松耦合”方式结合.

2.4 虚拟化技术

虚拟化(virtualization)技术是云计算系统的核心组成部分之一,是将各种计算及存储资源充分整合和高效利用的关键技术. 云计算的特征主要体现在虚拟化、分布式和动态可扩展,而虚拟化作为云计算最主要的特点,为云计算环境搭建起着决定性作用. 虚拟化技术是伴随着计算机技术的产生而出现的,作为云计算的核心技术,扮演着十分重要的角色,提供了全新的数据中心部署和管理方式,为数据中心管理员带来了高效和可靠的管理体验,还可以提高数据中心的资源利用率,低功耗绿色环保^[12]. 通过虚拟化技术,云计算中每一个应用部署的环境和物理平台是没有关系的,通过虚拟平台进行管理、扩展、迁移、备份,种种操作都通过虚拟化层次完成. 虚拟化技术实质是实现软件应用与底层硬件相隔离,把物理资源转变为逻辑可管理资源^[13-14]. 目前云计算中虚拟化技术主要包括将单个资源划分成多个虚拟资源的裂分模式,也包括将多个资源整合成一个虚拟资源的聚合模式. 虚拟化技术根据对象可分成存储虚拟化、计算虚拟化、网络虚拟化等,计算虚拟化又分为系统级虚拟化、应用级虚拟化和桌面虚拟化^[11,15].

气象业务云平台把云计算引入到气象业务中,将各种各样的物理计算资源组织在一个很大的资源池中,资源池被气象业务云平台管理之后,动态创立一个虚拟化资源池,把它变成新的气象数据处理中心. 用户只需向气象业务云平台发送指令即可动态上传添加新的资源,实现海量数据存储. 从另一个角度审视,气象业务云平台又是一个可靠的国家级气象信息存档中心和灾备中心. 一方面,气象业务云平台有丰富的存储资源,可以应对全国各级气象业务部门、研究机构和高校的气象信息存档任务;另一方面,云安全技术给数据灾备中心提供了强大的安全保障;再一方面,气象业务云平台落户地还须有优越的自然地理条件、完备的水电配套设施、稳定的社会周边环境、可控可接受的地域成本投入、积极的政策

支持环境以及丰富的高科技人才资源环境.

2.5 云计算平台管理技术

云计算资源规模庞大,一个系统服务器数量众多(可能高达10万台)、结构不同并且分布在不同物理地点的数据中心,同时还运行着成千上万种应用. 如何有效地管理云环境中的这些服务器,保证整个系统提供不间断服务必然是一个巨大的挑战. 云计算平台管理系统可以看作是云计算的“指挥中心”. 通过云计算系统的平台管理技术能够使大量的服务器协同工作,方便地进行业务部署和开通,快速发现和恢复系统故障,通过自动化、智能化的手段实现大规模系统的可靠运营和管理.

基于气象云计算的架构、软件和服务能够为全球各地的气象工作者提供一个有吸引力的合作平台. 例如可以将一些先进的处理遥感信息、卫星资料、雷达图像等的专业商业软件放在云平台中,供全国的气象工作者来使用,那将为我国气象事业节省巨大开支,并能提供廉价的高可靠和高性能的服务.

3 云计算应用现状

云计算平台是一个强大的“云”网络,连接了大量并发的网络计算和服务,可利用虚拟化技术扩展每一个服务器的能力,将各自的资源通过云计算平台结合起来,提供超级计算和存储能力. 下面就当前云计算相关主要研究团队和组织的研究进展情况进行比较分析,为云计算相关研究提供参考.

3.1 国际上相关研究组织

目前,国外已经有多个云计算的科学研究项目,非常有名是 Scientific Cloud^[16] 和 Open Nebula (<http://hive.apache.org/>) 项目. 产业界也在投入巨资部署各自的云计算系统,参与者主要有 Google、Amazon、IBM、Microsoft 等. 国内关于云计算的研究也已起步,并在计算机系统虚拟化基础理论与方法研究方面取得了阶段性成果. 国际上云计算主要研究组织及研发方向如表1所示^[17-20].

从表1可知:Google App Engine 是 Google 公司推出的云计算服务,允许用户使用 Python 编程语言编写 Web 应用程序在 Google 的基础架构上运行;IBM 的 Computing on demand 主要针对高性能计算,如汽车和航天工业模拟计算、气象数值天气预报等;Amazon 的弹性计算(EC2)和简单存储服务(S3)是 IaaS 的应用典范,提供廉价的计算和存储服务;Microsoft 提出云端的云计算构想,可以使应用在本地和云之间迁移;

EMC 优势在于其强大的虚拟化和存储技术,为企业和机构提供 PB 级的数据存储和信息管理解决方案;Salesforce 是 SaaS 的杰出代表,开发面向企业用户的客户关系管理(CRM)方案;Oracle 在软硬件产品方面非常出色,提供弹性虚拟云计算平台。

为进一步了解相关云计算产品,本文对国际上相关云计算产品性能比较如表 2 所示。

在程序开发方面,大多数平台支持 JAVA 语言,Microsoft 还支持 .Net、PHP、Python 和 Ruby, Salesforce 提供一种新的编程语言 Apex 和集成开发环境 Visualforce,能够降低应用开发的复杂度并缩短开发周期,Microsoft 和 Oracle 分别提供 Visual Studio 2010 和 NetBeans 集成开发环境,其余都在 Eclipse 集成开发环境上添加相应的工具集。在数据库方面,Google 的 MegaStore 是建立于 BigTable 之上的分布存储系统,可以将 NoSQL 和传统的关系型数据库融合在一起,Amazon 提供多种数据库支持,除自身的轻量级的数据库服务 SimpleDB 外,还支持 MySQL, SQL

server 和 Oracle 11g, Microsoft 提供 SQL Azure 数据服务,功能包括数据存储、查询和分析等,EMC 提供与其编程环境相适应的数据库服务 Spring Gemfire, Salesforce 数据库服务由 Oracle 提供,两者都是 Exadata 数据库,平台都实现了内置的自动负载均衡技术,使物理资源使用率得到有效提升。在内容分发网络方面,Amazon 提供内容推送服务 CloudFront, Microsoft 和 EMC 分别提供 Windows Azure CDN 和 DataDirect 服务,用来减轻源服务器的负担和改善整个网络流量分布不均匀的状况。Amazon 平台是开源的,在云计算发展中发挥了重要的作用,Oracle 的 Sun 部分也是开源的。

3.2 国内相关研究组织

国内学术界对云计算研究也非常热门,并做了大量的研究工作,如文献[6,12,15,21]等都相应提出了一些有意义的云计算研究方法。国内企业界主要云计算研究组织及研发方向如表 3 所示。

表 1 国际上云计算相关研究组织及研发方向

Table 1 International research direction and research organizations on cloud computing

团队组织	商业项目	技术特征	核心技术	适用范围
Google	Google App Engine, 包括 Google 搜索、Google Maps、Google Earth、Google AdSense、Gmail 等	储存及运算水平扩充能力	平行分散技术, MapReduce, BigTable, GFS	Google 各种日常互联网应用及开发者开发和发布各种应用程序
IBM	“蓝云” IBM 云环境管理解决方案(企业私有云), IBM LotusLive(会议服务、办公协作服务、电子邮件服务), IBM RC2 (IBM 8 大研究机构共同创建的私有云)	整合其所有软件和硬件服务	网格技术, 分布式存储, 动态负载	高性能计算(汽车和航天工业模拟计算、生命科学领域染色体组建模等)
Amazon	亚马逊网络服务(弹性计算云 EC2、简单存储服务 S3、简单数据库服务 simpleDB、简单队列服务 SQS、弹性 MapReduce 服务、内容推送服务、电子商务服务 DevPay 和 FPS)	弹性虚拟平台	虚拟技术 Xen	各类企业在其平台上搭建应用环境提供云计算服务(如在线照片存储共享网站 SmugMug、在线视频制作网站 Animoto)
Microsoft	Azure 平台提供 Microsoft .NET 服务、Microsoft SQL 服务、Live 服务	整合其所有软件及数据服务	大型应用软件开发技术	应用程序开发者在云端开发程序,且运行在微软云端的应用程序还是运行在本地的应用程序都可以使用云计算服务平台
EMC	云存储基础架构 EMC Atoms(PB 级的信息管理解决方案), Mozy 针对 Mac 用户的在线存储	信息存储系统和虚拟化	Vmware 的虚拟化技术, 一流的存储技术	向各种规模的企业和机构提供自动化网络存储解决方案
Salesforce	客户关系管理(CRM)包括 Sales Cloud、Service Cloud、Custom Cloud、Cloud Platform for CRM、Cloud Infrastructure for CRM	弹性可定制商务软件	应用平台整合技术	为企业客户提供客户关系管理(CRM)服务
Oracle	EC2 上的 Oracle 数据库, OracleVM, SunxVM	软硬件弹性虚拟平台	Oracle 的数据存储技术, Sun 开源技术	提供统一整合的界面来部署在云中任何操作系统上运行的应用软件, 操作系统包括 Open Solaris、Linux、Windows

表2 国外相关云计算产品性能比较

Table 2 Performance comparison of international cloud computing products

团队组织	开发语言	编程环境	数据库	负载均衡	内容分发网络 CDN	开源程度
Google	JAVA, Python	Eclipse 的 Google 插件	MegaStore	内置自动	有限的内置 CDN	不开源
Amazon	JAVA	Eclipse AWS 工具集	SimpleDB, MySQL, SQL server, Oracle 11g	内置自动	Amazon CloudFront	开源
Microsoft	JAVA, . Net, PHP, Python, Ruby	Visual Studio 2010, Eclipse Windows Azure 工具集	SQL Azure, SQL server 2008	内置自动	Windows Azure CDN	不开源
EMC	JAVA	Eclipse SpringSource 工具集	Spring GemFire	内置自动	EMC DataDirect	不开源
Salesforce	Apex	Visualforce	Oracle Exadata	内置自动	内置 CDN	不开源
Oracle	JAVA, Ruby	NetBeans	Oracle Exadata	内置自动	内置 CDN	部分开源

表3 国内云计算相关研究组织及研发方向

Table 3 Domestic research direction and research organizations on cloud computing

团队组织	商业项目	技术特征	核心技术	适用范围
中国移动	大云平台(大云数据挖掘系统 BC-PDM、海量结构化存储 Hugetable、大云弹性计算系统 BC-EC、大云弹性存储 BC-NAS 和大云并行计算系统 BC-MapReduce)	坚实的网络技术,丰富的宽带资源	底层集群部署技术,资源池虚拟技术,虚拟技术 Xen	中国移动内部的新型 IT 支撑系统(网管系统、业务支撑系统、管理信息系统), IDC, mSpaces、Email 等互联网应用
阿里巴巴	阿里云、软件互联平台、云电子商务	弹性可定制商务软件	应用平台整合技术	提供计算、存储、网络服务,还提供各种电子商务服务
世纪互联	CloudEx(互联网主机服务 CloudEx Computing Service、在线存储虚拟化 CloudEx Storage Service、互联网云端备份的数据保全服务)	弹性虚拟平台	虚拟技术 Xen, 集群虚拟化技术	提供计算、存储、备份服务
友友	友友云(数据平台、分布式虚拟存储服务、网络计算平台、网络虚拟机、友友企业地图)	互联网软件		高端的大型集团客户(如江苏省电力公司)
华为	云帆包括云数据中心解决方案(基础架构的平台解决方案包括计算、存储、网络、平台软件、工程设计以及基础设施)和云应用解决方案(电信业务云化方案包括业务云、支撑云、IDC 云、桌面云)	通讯设计及技术		与合作伙伴共建云计算(如盈世、普华、升腾、用友), 通讯行业(139 邮箱云存储、上海联通 OSS、客服系统), 医药行业(闸北区健康云)

中国移动大云平台包括数据挖掘、海量数据存储和弹性计算等,主要用于中国移动的业务支撑、信息管理和互联网应用;阿里云提供的计算、存储和网络服务主要用于提供各种电子商务服务;世纪互联的弹性云计算已经对外提供服务,支持多个操作系统、数据库和编程环境;友友云开发的数流平台(Bitsflow),分布式虚拟存储系统(DataCell),网络计算平台 GAP 和系统监控管理平台(NetVM)主要针对大型企业;华为的云帆计划主要包括云数据中心和针对电信的 IDC 云。

国内相关云计算产品性能比较如表 4 所示。

在平台的文件系统方面,中国移动,阿里巴巴和世纪互联都是基于 Hadoop 文件系统 HDFS,根据自身需求进行一些改进,而友友提出的 DataCell FS 采

用多租户、容错、负载均衡、虚拟化等技术;在数据库方面,中国移动的 HugeTable 适合存储海量结构化的数据,阿里巴巴和世纪互联都支持 MySQL,世纪互联还支持 SQL server 2005,而友友提供的 DataCell DB 主要针对结构化和半结构化数据,不完全支持关系数据库,但支持常用的关系操作(select、insert、delete、update);在开发环境方面,中国移动和阿里巴巴支持 JAVA 语言,世纪互联支持 PHP 和 .Net 开发环境,友友提供 C++ 和 JAVA 2 种开发语言.平台都支持 Linux 系统,世纪互联还提供 Windows 2003 环境,它们都内置了自动负载均衡技术,使资源利用率得到有效提升。

最后,总结一下目前云计算研究的相关公司与产业链分布情况,如表 5 所示。

表4 国内相关云计算产品性能比较

Table 4 Performance comparison of domestic cloud computing products

团队组织	文件系统(基于开源文件系统)	数据库	开发环境	基于操作系统	负载均衡
中国移动	HDFS	HugeTable	JAVA	Linux	内置,自动
阿里巴巴	HDFS	MySQL	JAVA	Linux	内置,自动
世纪互联	HDFS	MySQL, Windows 2005	PHP, .Net	Windows 2003, Linux	内置,自动
友友	DataCell FS	DataCell DB	C + + ,JAVA	Linux	内置,自动

表5 云计算研究相关公司与产业链分布

Table 5 Distribution of research companies and the industry chain on cloud computing

服务提供商	IT 基础设施				云基础设施			云平台			云软件	
	服务器	存储	网络设备	云终端	网络	在线存储	在线计算	开发环境	程序服务和工具	商业流程外包	在线企业应用	在线个人应用
IBM	●	●				IBM 蓝云						
HP	●	●			●	●	●					
EMC		●				●		SpringSource				
Cisco	●		●									
微软						●	●	Windows Azure			Office online	
谷歌								Google Apps Engine Android, Open Social			Apps Gmail	Google Apps Gmail
亚马逊						S3	EC2	Simple DB/SQS				
SAP								SAP Web App Librart		●	●	
浪潮信息	●											
中创软件	●										InforSuite Cloud	
阿里巴巴					●	●	●					
华为	●		●		●	●	●					DBank
中国电信					●	●	●					
中国移动					●	●	●					139 邮箱
中国联通					●	●	●					

国外公司提供云计算相关的多个产业, IBM 和 HP 主要依靠硬件和 IaaS 来抢占市场, 微软提供一整套的云计算服务, 谷歌主要提供 SaaS 为主, 包括 Google 搜索、Google Earth、Gmail 等, Amazon 以 PaaS 著称, 提供廉价的计算和存储资源, 而国内公司主要提供 IaaS, 包括计算、存储和网络资源, 还有少量的 SaaS, 如华为 DBank 数据存储和中国移动的邮箱服务。

4 云计算存在的挑战与机遇

目前, 尽管云计算如火如荼地开展着, 通过分析知道云计算具有许多优点, 让人们看到了 IT 服务将成为公共服务的曙光, 但是我们要清醒地认识到云计算也不是万能的灵丹妙药, 它仍存在着一些亟待解决的实际问题, 如服务可用性、数据主权与数据隐私问题、安全问题、软件许可证问题、网络传输问题和可伸缩性存储等。甚至还有一些专家和学者对

云计算持怀疑态度, 认为是在炒作概念。为此, 需要进一步加强宣传和用好云计算产品, 为科研生产服务。本文给出了云计算发展过程中主要存在的一些挑战问题与机遇如表 6 所示^[14], 其中每一个问题即是一种挑战, 同时也是一个机遇, 为未来云计算的研究发展指明了方向。

5 总结与展望

计算作为一项服务功能是人们长久以来的梦想。随着微软吸引人眼球的一句广告语“ We are all in cloud!”, 标志着云计算时代的到来, 人们长久以来的梦想终将变为现实。从云供应商的观点出发, 整合服务器平台技术能更好地推进云计算基础设施架构的搭建以及大规模数据中心管理能力的提升, 建立合作伙伴的生态链。利用商品化的计算、存储和网络低成本建立大型数据中心使得以低于许多中等规模的数据中心的价格“即用即付”的销售资源成为可

表6 云计算发展的挑战及可能解决方案

Table 6 Cloud computing development challenges and possible solutions

序号	挑战问题	主要内容	解决方案
1	服务的可用性	很多人担心效用计算服务是否能提供很好的可用性,对云计算不免有些怀疑.	选用多个云计算提供商;利用弹性来防范 DDOS 攻击.
2	数据锁定	云计算的 API 仍是私有的,或者说当前没有建立起统一的标准.因此,用户很难将他们的数据和程序从一个站点迁移到另一个,这也是很多用户不愿采用云计算的原因.	标准化的 API;将价格和服务质量绑定.
3	数据主权、安全性和可审计性	有些数据是企业的商业机密,数据的安全性关系到企业的生存和发展.云计算数据的安全性问题解决不了会影响云计算在企业中的应用.包括数据存储、传输安全,数据隐私、数据主权、身份认证等问题是用户非常关心的问题.	采用加密技术,VLANs 和防火墙;跨地域的数据存储.
4	网络数据传输瓶颈	当前应用发展越来越趋向于数据密集型,如果应用能够被拆分交由不同的云去处理,这将导致复杂的数据存放和传输.云计算服务依赖网络,目前网速低且不稳定,使云应用的性能不高.	快速硬盘;数据备份/获取;更低的广域网路由开销;更高带宽的 LAN 交换机.
5	性能不可预知	在云计算中多虚拟机能够很好地共享 CPU 和内存,但是 I/O 的共享却有明显的问题.	改进虚拟机支持;闪存;支持 HPC 应用的虚拟集群.
6	可伸缩的存储和快速伸缩性	短周期,无需提前花费成本,以及按需提供无限资源.如何申请一个持久的存储是不容忽视的问题.即买即用无疑很适合存储和网络带宽,这两者都可以用使用字节数来衡量.由于使用了虚拟机,计算稍稍有些不同.	发明可伸缩的存储;基于机器学习的计算自动伸缩;使用快照以节约资源.
7	数据隐私和法律危机	如何保证存放在云服务提供商的数据隐私不被非法利用,不仅需要技术的改进,也需要法律的进一步完善.	采用特定的服务和法律进行有效立法和保护.
8	软件许可	当前软件许可证通常限定在运行软件的机器上,用户购买软件并按年支付维护费用,许多云计算提供商从一开始就倾向于开源软件,部分原因正是因为商业软件许可证模式并不适合效用计算.	使用即用即付许可;批量销售.
9	用户使用习惯	如何改变用户的使用习惯,使用户适应网络化的软硬件应用是长期而艰巨的挑战.	方便用户使用,降低使用成本,扩大宣传力度.
10	技术标准和运营标准	缺乏统一的技术标准和运营标准,具体表现在数据接口、数据迁移、数据交换、测试评价等技术方面,以及 SLA、云计算治理和审计、运维规范、计费标准等运营方面,不利于用户的统一认识和云服务的规模化推广.	相关标准化组织统一制定行业技术标准和运营标准.

能,并且可以利用资源在大量用户间的复用来获取利润.从云用户的观点出发,云计算可以令一个初创的软件公司像初创的芯片厂商拥有为之服务的代工厂一样拥有自己的数据中心.

尤其值得关注的是,中国政府已经开始关注云计算的发展,大力发展云计算相关研发工作.2010年10月18日,国家发展改革委员会、工业和信息化部联合下发了《关于做好云计算服务创新发展试点示范工作的通知》,结合国家创新型城市建设,先行在北京、上海、杭州、无锡、深圳5个城市开展云计算试点工作^[22].云计算也是国家“十二五”规划中的重点关注项目,国家政策的扶植和地方政府的重视将会加快云计算产业的发展的步伐.所有这一切表明,云计算应用已经在中国市场上全面启动.通过本文

综述分析,云计算发展初期的重点受益公司主要来自IT基础设施、系统集成和云软件(SaaS)3个方面.

总之,国内外对云计算及其虚拟化关键技术方面的研究非常活跃,并开发了一些相关工具与产品,云计算作为新一代产业浪潮的重要驱动力,将对经济社会发展产生深远的影响.

参考文献

References

- [1] 吴吉义,平玲娣,潘雪增,等.云计算:从概念到平台[J].电信科学,2009,25(12):23-30
WU Jiyi, PING Lingdi, PAN Xuezheng, et al. Cloud computing: Concept and platform [J]. Telecommunications Science, 2009, 25(12): 23-30
- [2] Cloud computing [EB/OL] [2011-09-01]. <http://en.>

- wikipedia.org/wiki/Cloud_computing
- [3] Michael A, Armando F. Above the clouds: A Berkeley view of cloud computing[EB/OL][2011-09-02]. <http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf>
- [4] Mell P, Grance T. The NIST definition of cloud computing [R]. National Institute of Standard and Technology, U S Department of Commerce, 2010
- [5] Sun Corporation. The white paper on cloud computing architecture[R]. 2009
- [6] 刘鹏. 云计算技术基础[M]. 2版. 北京: 电子工业出版社, 2011
LIU Peng. Introduction to cloud computing[M]. 2nd Ed. Beijing: Publishing House of Electronics Industry, 2011
- [7] Dean J, Ghemawat S. MapReduce: Simplified data processing on large clusters [J]. Communications of the ACM, 2008, 51(1): 107-113
- [8] 王庆波, 金津, 何乐, 等. 虚拟化与云计算[M]. 北京: 电子工业出版社, 2010
WANG Qingbo, JIN Xing, HE Le, et al. Virtualization and cloud computing[M]. Beijing: Publishing House of Electronics Industry, 2010
- [9] VMware virtualization technology [EB/OL] [2011-09-02]. <http://www.vmware.com/virtualization/what-is-virtualization.html>
- [10] Goyal A, Dadizadeh S. A survey on cloud computing [R]. Technical Report for CS 508, 2009
- [11] 王鹏. 云计算的关键技术与应用实例[M]. 北京: 人民邮电出版社, 2010
WANG Peng. Key technology and application examples of cloud computing [M]. Beijing: Posts & Telecom Press, 2010
- [12] 陈康, 郑纬民. 云计算: 系统实例与研究现状[J]. 软件学报, 2009, 20(5): 1337-1348
CHEN Kang, ZHENG Weimin. Cloud computing: System instances and current research[J]. Journal of Software, 2009, 20(5): 1337-1348
- [13] Kamoun F. Virtualizing the datacenter without compromising server performance[J]. ACM Ubiquity 2009, 2009. Doi: 10.1145/1595422.1595424
- [14] Garfinkel T, Matthews J, Hoff C, et al. Virtual machine contracts for datacenter and cloud computing environments[C]// Proceedings of the 1st Workshop on Automated Control for Datacenters and Clouds, ACM, New York, 2009: 25-30
- [15] 石磊, 邹德清, 金海. Xen 虚拟化技术[M]. 武汉: 华中科技大学出版社, 2009
SHI Lei, ZOU Deqing, JIN Hai. Xen virtualization technology [J]. Wuhan: Huazhong University of Science & Technology Press, 2009
- [16] Keahey K, Figueiredo R, Fortes J, et al. Science clouds: Early experiences in cloud computing for scientific applications[C]// In Proc of Workshop on Cloud Computing and its Applications, 2008
- [17] Leavitt N. Is cloud computing really ready for prime time? [J]. Computer, 2009, 42(1): 15-20
- [18] Vaquero L M, Rodero-Merino L, Caceres J, et al. A break in the clouds: Toward a cloud definition[J]. ACM SIGCOMM Computer Communication Review, 2009, 39(1): 50-55
- [19] Lagar-Cavilla H A, Whitney J A, Scannel A, et al. Snow-Flock: Rapid virtual machine cloning for cloud computing [C]// Proceedings of the 4th ACM European Conference on Computer Systems, 2009: 1-12
- [20] 戴元顺. 云计算技术简述[J]. 信息通信技术, 2010(2): 29-35
DAI Yuanshun. The brief review of cloud computing technologies[J]. Information and Communications Technologies, 2010(2): 29-35
- [21] 李亚琼, 宋莹, 黄永兵. 一种面向虚拟化云计算平台的内存优化技术[J]. 计算机学报, 2011, 34(4): 684-693
LI Yaqiong, SONG Ying, HUANG Yongbing. A memory global optimization approach in virtualized cloud computing environments [J]. Chinese Journal of Computers, 2011, 34(4): 684-693
- [22] 赛迪网. 中国云计算产业发展白皮书[R]. 2011
Ccidnet. The white paper of China cloud computing industry[R]. 2011

Cloud computing: Conceptions, key technologies and application

FANG Wei^{1,2,3} WEN Xuezh^{1,2} PAN Wubin^{1,2} XUE Shengjun^{1,2}

1 Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science & Technology, Nanjing 210044

2 School of Computer & Software, Nanjing University of Information Science & Technology, Nanjing 210044

3 Postdoctoral Work Station of China Innovation Software Engineering Limited Corporation, Jinan 250014

Abstract In recent years, cloud computing as a new computing service model has become a research hotspot in computer science. This paper is to give a brief analysis and survey on the current cloud computing systems from the definition, deployment model, characteristics and key technologies. Then, the major international and domestic research enterprises and application products on cloud computing are compared and analyzed. Finally, the challenges and opportunities in current research of cloud computing are discussed, and the future directions are pointed out. So, it will help to provide a scientific analysis and references for use and operation of cloud computing.

Key words cloud computing; virtualization; service; platform