Hadoop

云计算一体机

实践指南

李宁　王东亮　等编著

机械工业出版社
CHINA MACHINE PRESS

Hadoop 云计算一体机实践指南

李宁　王东亮　等编著

机械工业出版社

全书分为3篇：第1篇（理论部分）对云计算、Hadoop及Linux操作系统进行了简单介绍；第2篇（基础实践部分）主要详细介绍了CentOS系统的安装和集群的搭建、Hadoop集群的常用命令及管理应用等；第3篇（项目实训部分）主要以实际项目开发为例，从易到难，对源程序进行了详细解释。

本书以详细的实践操作介绍为特色，可作为电子、通信、自动化、计算机等电类专业Hadoop云计算教学系统课程实验教材，也可供Hadoop系统相关工程技术人员参考。

图书在版编目（CIP）数据

Hadoop云计算一体机实践指南/李宁等编著. —北京：机械工业出版社，2013.8

ISBN 978-7-111-43496-2

Ⅰ.①H… Ⅱ.①李… Ⅲ.①数据处理软件 Ⅳ.①TP274

中国版本图书馆CIP数据核字（2013）第175911号

机械工业出版社（北京市百万庄大街22号 邮政编码100037）

策划编辑：顾 谦 责任编辑：顾 谦

版式设计：常天培 责任校对：陈立辉

封面设计：赵颖喆 责任印制：乔 宇

北京铭成印刷有限公司印刷

2013年9月第1版第1次印刷

184mm×260mm · 13.75印张 · 337千字

0001—3000册

标准书号：ISBN 978-7-111-43496-2

定价：39.90元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

电话服务 网络服务

社服务中心：（010）88361066 教材网：http://www.cmpedu.com

销售一部：（010）68326294 机工官网：http://www.cmpbook.com

销售二部：（010）88379649 机工官博：http://weibo.com/cmp1952

读者购书热线：（010）88379203 **封面无防伪标均为盗版**

前言

本书是基于北京斑步至伟科技有限公司 Hadoop 云计算一体机[搭配 PC(个人计算机)],由北京林业大学工学院与该公司共建的 Bamboo 云计算与物联网实验室共同编写而成,主要介绍如何搭建 Hadoop 平台并在其上面开发并部署云应用。本书主要由李宁、王东亮编著,北京斑步至伟科技有限公司的工程师对本书中的软件实例开发和验证进行了大量、细致的工作,北京林业大学工学院杨尊昊老师、即将留校参加工作的研究生吴健等人参与了一体机的开发与软件测试,实验室全体成员的共同努力确保了本书的质量。

近几年来云计算经过各大厂商的推广,很多读者对于 Hadoop 和云计算已经有了一定的了解,目前有相当一批人在从事这方面的工作,一些高校也开始有这方面的课程和讲座。

本书主要介绍在“裸机”的状态下,如何搭建 Hadoop 平台。通过由浅入深的介绍,引领学生步入 Hadoop 以及云计算的大门。全书从云计算、Hadoop 以及 Linux 的理论知识开始介绍,到集群的搭建、项目的设计和应用,贯穿了 Hadoop 学习的整个过程。通过学习本书,读者基本上可以熟悉市场上各种云平台开发过程和设计思想。

本书各章节内容主要安排如下:

第 1 篇　理论部分

第 1 章:云计算概念、发展现状和实现机制,目前各大厂商的云计算概述;

第 2 章:Hadoop 简介、架构和核心技术——HDFS 以及 MapReduce;

第 3 章:Linux 操作系统的介绍和命令的使用。

第 2 篇　基础实践部分

第 4 章:CentOS 系统的安装和集群的搭建过程;

第 5 章:熟悉 Hadoop 集群的常用命令并学会使用 Web 方式浏览集群;

第 6 章:Hadoop 内部的管理应用、系统体检权限管理;

第 7 章:基础开发实验:包含 HDFS 上传文件、浏览文件和目录、打开、下载和删除文件;

第 8 章:高级开发实验:包含数据去重、排序、平均成绩和单表关联,都是基于 MapReduce 的。

第 3 篇　项目实训部分

第 9 章:个人存储私有云,主要实现云文件的存储、浏览和删除;

第 10 章:气象数据分析云,主要实现对气象数据的分析和计算给出表格;

第 11 章:微信人物关系云,主要根据上传分析条件来对人物关系进行分析;

第 12 章:云图书馆,主要实现电子图书的上传索引,并且提供查询的功能;

第 13 章:物联网与云计算,主要实现智能分配快递的功能。

由于作者水平有限,本书难免存在一些错误和不足,欢迎大家指正。

编著者

2013 年 5 月

目录

第1篇

理论部分

第1章　云计算理论

1.1　云计算的概念

云计算(Cloud Computing)首先是基于互联网的，并且是对在互联网上的服务进行使用、增加和交付的，这个资源一般是虚拟化并且动态易扩展的。云其实就是对网络以及互联网资源的一种形象说法。在过去，很多人画图会使用云来表示电信网，后来出现了互联网还有对一些底层的基础架构的形象描述。狭义云计算是指一种 IT 基础设施的交付和使用模式，指通过互联网以按需、易扩展的方式获得所需资源；广义云计算指服务的交付和使用模式，指通过互联网以按需、易扩展的方式获得所需服务。这种服务可以是 IT 资源和应用软件、一切和互联网相关资源，也可是其他种类的服务。它意味着计算能力也可作为一种商品通过互联网进行流通。

什么是云计算？云计算是一种基于互联网的超级计算模式，在远程的数据中心，几万台甚至几千万台计算机和服务器连接成一片。因此，云计算甚至可以让人体验每秒超过 10 万亿次的运算能力，如此强大的运算能力几乎无所不能。用户通过计算机、便携式计算机、手机等方式接入数据中心，按各自的需求进行存储和运算。四款比较成熟而实用的云计算产品如下：IBM 公司的蓝云、亚马逊公司的 Amazon EC2、谷歌公司的 Google App Engine、微软公司的 Windows Azure。

1.2　云计算发展现状

1. 云计算发展现状

云计算作为业界热点，近年来世界各国对于它的研究和应用方兴未艾，许多政府部门和著名公司在研发与应用云计算的过程中作出了大量的工作和努力。

(1) 云计算在国外的发展

云计算与网络密不可分。云计算的原始含义是通过互联网提供计算能力。云计算的起源跟亚马逊和谷歌两个公司有十分密切的关系，它们最早使用到了“Cloud Computing”的表述方式。目前美国公开宣布进入或支持云计算技术开发的业界巨头包括微软、谷歌、IBM、亚马逊、Netsuite、NetApp、Adobe 等公司。

谷歌公司是云计算的提出者。2006 年，谷歌公司启动了“Google101”计划，引导大学生们进行“云”系统的编程开发。多年的搜索引擎技术的积累成果使谷歌公司在云计算技术上处于领先的地位，不仅提供在线应用，还希望发挥自身的数据库系统优势，成为在线应用的统一平台。谷歌公司以发表学术论文的形式公开了其云计算三大法宝：GFS、Map/Reduce 和 BigTable，并在美国和我国等国高校开设云计算编程课程。

微软公司于 2008 年 10 月推出了 Windows Azure 操作系统，这个系统作为微软公司云计

算计划的服务器端操作系统(Cloud OS)为广大开发者提供服务。微软公司拥有全世界数以亿计的 Windows 操作系统用户桌面和浏览器，Azure(蓝天)试图通过在互联网架构上打造新的云计算平台，让 Windows 操作系统由 PC(个人计算机)延伸到“蓝天”上。

IBM 公司从企业内部需求的逐渐上升出发，在2007 年11 月提出了“蓝云”计划，推出共有云和私有云的概念。IBM 公司提出私有云解决方案是为减少诸如数据、信息安全等共有云现存的问题，从而抢占企业云计算市场。依托 IBM 公司在服务器领域的传统优势，IBM 公司成为目前惟一一个提供从硬件、软件到服务全部自主生产的公司。

2008 年7 月，雅虎、惠普和英特尔公司联合宣布将建立全球性的开源云计算研究测试床，称为 Open Cirrus，鼓励开展云计算、服务和数据中心管理等领域中各方面的研究。

苹果公司是云计算领域的一位积极参与者。从近年来推出的 iTunes 服务，到 Mobile Me 服务，到收购在线音乐服务商 Lala，再到最近在美国北卡莱罗纳州投资 10 亿美元建立新数据中心的计划，无不显示其进军云计算领域的巨大决心。

这些国际知名大公司在全世界建造了庞大的云计算中心。譬如：谷歌公司的搜索引擎分布于 200 多个站点、超过 100 万台服务器支撑，而且设施数量正在迅猛增长。

(2) 云计算在国内的发展

目前我国云计算的讨论多数集中在早期云计算的概念、技术和模式上。早期的云计算是一种动态的、易扩展的、通过互联网提供虚拟化 IT(信息技术)资源和应用的一种计算模式。用户不需要了解云技术内部的细节，也不必具有云内部的专业知识，更不需要直接参与、投入、建设、维护和控制就能直接按需使用并按用量付费。

2008 年，IBM 公司在无锡建立了我国第一个云计算中心，在北京 IBM 中国创新中心建立了第二个云计算中心——IBM 大中华区云计算中心。2009 年初，在南京建立了国内首个“电子商务云计算中心”。世纪互联推出“CloudEx”产品线，包括完整的互联网主机服务“CloudEx Computing Service”、基于在线存储虚拟化的“CloudEx Storage Service”等云计算服务。

随着云计算的升温，国内的电信运营商也都积极投入到云计算的研究中，以期通过云计算技术促进网络结构的优化和整合，寻找到新的赢利机会和利润增长点，以实现向信息服务企业的转型。中国移动公司推出了“大云”(Big Cloud)云计算基础服务平台，中国电信公司推出了“e 云”云计算平台，中国联通公司则推出了“互联云”平台。

我国企业创造了“云安全”概念，通过网状的大量客户端对网络中软件行为的异常监测，获取互联网中木马、恶意程序的最新信息，在服务器端进行自动分析和处理，再把解决方案分发到客户端。瑞星、趋势等企业都推出了云安全解决方案。

随着云计算的发展，互联网的功能越来越强大，用户可以通过云计算在互联网上处理庞大的数据和获取所需的信息。从云计算的发展现状来看，未来云计算的发展会向构建大规模的能够与应用程序密切结合的底层基础设施的方向发展。不断创建新的云计算应用程序，为用户提供更多、更完善的互联网服务也可作为云计算的一个发展方向。

2. 云计算实现机制

由于云计算分为 IaaS(基础设施即服务)、PaaS(平台即服务)和 SaaS(软件即服务)三种类型，不同的厂家又提供了不同的解决方案，目前还没有一个统一的技术体系结构，对读者了解云计算的原理构成了障碍。为此，本书综合不同厂家的方案，构造了一个供商榷的云计

算体系结构。这个体系结构如图1-1所示，它概括了不同解决方案的主要特征，每一种方案或许只实现了其中部分功能，或许也还有部分相对次要功能尚未概括进来。

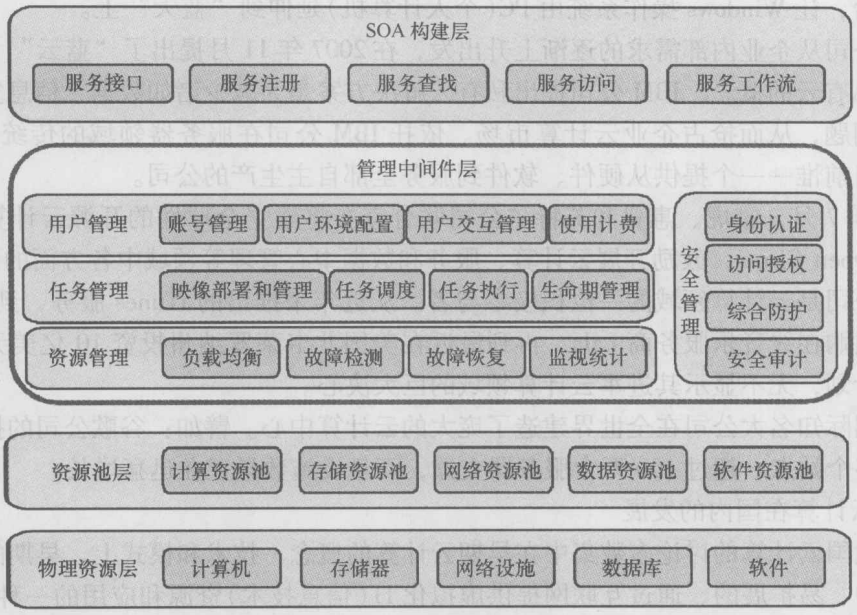

图1-1 云计算技术体系结构

云计算技术体系结构分为4层：物理资源层、资源池层、管理中间件层和SOA(面向服务的体系结构)构建层，如图1-1所示。物理资源层包括计算机、存储器、网络设施、数据库和软件等；资源池层是将大量相同类型的资源构成同构或接近同构的资源池，如计算资源池、数据资源池等。构建资源池层更多是物理资源的集成和管理工作，例如研究在一个标准集装箱的空间如何装下2000个服务器、解决散热和故障节点替换的问题并降低能耗；管理中间件层负责对云计算的资源进行管理，并对众多应用任务进行调度，使资源能够高效、安全地为应用提供服务；SOA构建层将云计算能力封装成标准的Web Services服务，并纳入到SOA体系进行管理和使用，包括服务注册、查找、访问和构建服务工作流等。管理中间件层和资源池层是云计算技术的最关键部分，SOA构建层的功能更多地依靠外部设施提供。

云计算的管理中间件层负责资源管理、任务管理、用户管理和安全管理等工作。资源管理负责均衡地使用云资源节点，检测节点的故障并试图恢复或屏蔽之，并对资源的使用情况进行监视统计；任务管理负责执行用户或应用提交的任务，包括完成用户任务映像(Image)的部署和管理、任务调度、任务执行、任务生命期管理等；用户管理是实现云计算商业模式的一个必不可少的环节，包括提供用户交互接口、管理和识别用户身份、创建用户程序的执行环境、对用户的使用进行计费等；安全管理保障云计算设施的整体安全，包括身份认证、访问授权、综合防护和安全审计等。

基于上述体系结构，本书以IaaS云计算为例，简述云计算的实现机制，如图1-2所示。

用户交互接口向应用以Web Services方式提供访问接口，获取用户需求。服务目录是用户可以访问的服务清单。系统管理模块负责管理和分配所有可用的资源，其核心是负载均衡。配置工具负责在分配的节点上准备任务运行环境。监视统计模块负责监视节点的运行状

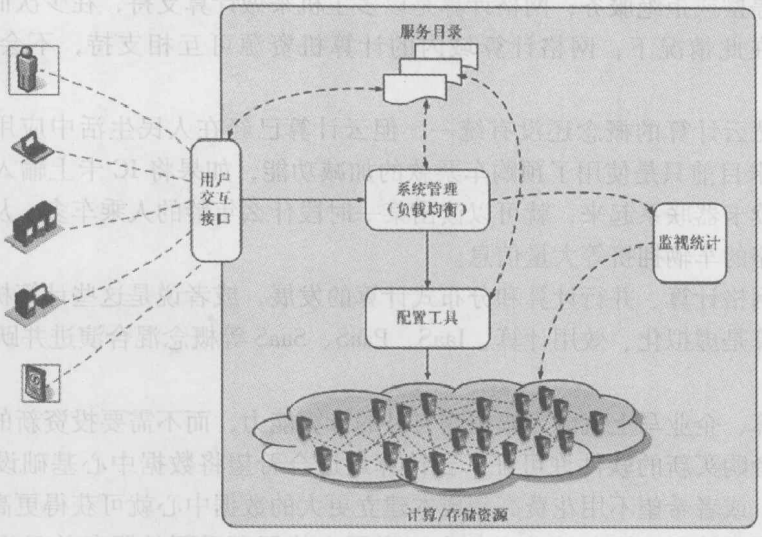

图 1-2 简化的 IaaS 实现机制

态，并完成用户使用节点情况的统计。执行过程并不复杂：用户交互接口允许用户从目录中选取并调用一个服务。该请求传递给系统管理模块后，它将为用户分配恰当的资源，然后调用配置工具来为用户准备运行环境。

1.3 网格计算与云计算

与云计算不同，网格计算已经是一个老词了。当云计算“大红大紫”的时候，人们很少提及网格计算，不过网格计算与云计算有着很深的渊源。

网格计算(Grid Computing)是通过利用大量异构计算机(通常为桌面)的未用资源[CPU(中央处理器)周期和磁盘存储]，将其作为嵌入在分布式电信基础设施中的一个虚拟的计算机集群，为解决大规模的计算问题提供了一个模型。网格计算的焦点放在支持跨管理域计算的能力上，运用平行运算，着重于企业间或跨企业的资源充分运用，共同解决困难的运算任务。这使它与传统的计算机集群或传统的分布式计算相区别。

云计算(Cloud Computing)是一种基于互联网的计算新方式，通过互联网上异构、自治的服务为个人和企业用户提供按需即取的计算。由于资源是在互联网上，而在计算机流程图中，互联网常以一个云状图案来表示，因此可以形象地类比为云，“云”同时也是对底层基础设施的一种抽象概念。

云计算的资源是动态易扩展而且虚拟化的，通过互联网提供。终端用户不需要了解“云”中基础设施的细节，不必具有相应的专业知识，也无需直接进行控制，只关注自己真正需要什么样的资源以及如何通过网络来得到相应的服务。虽然云计算源自平行运算的技术，不脱离网格计算的概念，但是云计算更专注于数据的处理。

云计算其实质还是与以往各类计算机运行的基本过程一样：由输入端输入数据，经数据处理后，再由输出端输出处理后的数据。云计算与网格计算的最大差异在于计算量，云计算大都以单一主机服务用户，主要偏向少量而多次的计算，少次而大量的计算易使资源用尽，

致使其他服务停摆或拒绝服务；网格计算是以多主机来做计算支持，在少次而大量的计算时较为有效率，在此情况下，网格计算域内的计算机资源可互相支持，不会有资源用尽的顾虑。

目前，虽然云计算的概念还没有统一，但云计算已经在人民生活中应用。比如，公交IC(集成电路)卡目前只是使用了预购车票款的加减功能，如果将IC卡上输入更多的持卡人的信息，再将读卡器联系起来，就可以读出某一时段什么年龄的人乘车多、从哪里上车哪里下车、什么线路的车辆拥挤等大量信息。

云计算是网格计算、并行计算和分布式计算的发展，或者说是这些计算机科学概念的商业实现。云计算是虚拟化、效用计算、IaaS、PaaS、SaaS等概念混合演进并跃升的结果。

1. 云计算

使用云计算，企业马上就能大幅提高自己的计算能力，而不需要投资新的基础设施、开展新的培训或者购买新的软件许可证。云计算最适合希望将数据中心基础设施全部外包的中、小型企业，或者希望不用花费高额成本建立更大的数据中心就可获得更高负荷能力的大型企业。不论哪种情况，服务消费者都在互联网上使用所需要的服务并只为所使用的服务付费。

服务消费者不用再守在PC旁边使用PC上的应用程序，或者购买针对特定智能手机、PDA(个人数字助理)及其他设备的版本。消费者不必拥有云中的基础设施、软件或平台，因此降低了前期成本、资本支出和运营成本。消费者也不用关心云中的服务器和网络怎么维护。消费者可以访问任何地方的多台服务器，不需要知道使用的是哪一台服务器以及它们的位置。

2. 网格计算

云计算是从网格计算演化来的，能够随需应变地提供资源。网格计算可以在云中，也可以不在，这取决于什么样的用户在使用它。如果用户是系统管理员或集成商，他们就会关心如何维护云。他们升级、安装和虚拟化服务器与应用程序。如果用户是消费者，就不必关心系统是如何运行的。

网格计算要求软件的使用可以分为多个部分，将程序的片段作为大的系统映像传递给几千个计算机中。网格的一个问题是如果某个节点上的软件片段失效，可能会影响到其他节点上的软件片段。如果这个片段在其他节点上可以使用故障转移组件，那么就可以缓解问题，但是如果软件片段依赖其他软件片段完成一项或多项网格计算任务，那么问题仍然得不到解决。大型系统镜像以及用于操作和维护的相关硬件可能造成很高的资本和运营支出。

3. 异同点

云计算和网格计算都是可伸缩的。可伸缩性是通过独立运行在通过Web服务连接的各种操作系统上的应用程序实例的负载平衡实现的。CPU和网络带宽根据需要分配和回收。系统存储能力根据特定时间的用户数量、实例的数量和传输的数据量进行调整。

两种计算类型都涉及多承租(Multitenancy)和多任务，即很多用户可以执行不同的任务，访问一个或多个应用程序实例。通过大型的用户池共享资源来降低基础设施成本，提高峰值负荷能力。云计算和网格计算都提供了服务水平协议(SLA)以保证可用性，比如99%。如果服务达不到承诺的正常运行时间，消费者将由于数据延迟而得到服务补偿。

亚马逊S3在云中提供了存储和数据检索Web服务。设置在S3中能够存储的对象数量

的最大上限，可以存储只有1B的对象，也能存储5GB甚至TB级的对象。S3对于对象的每个存储位置使用“桶(Bucket)”作为容器。这些数据采用和亚马逊电子商务网站相同的数据存储基础设施安全地实现存储。

虽然网格中的存储计算非常适合数据密集型存储，但是存储1B大小的对象从经济上来说不合适。在数据网格中，分布式数据的数量必须足够大才能发挥最大效益。

计算型网格关注的是计算量非常大的操作。云计算中的亚马逊Web Services提供了两种实例：标准和高CPU。

1.4 云计算的发展环境

1.4.1 云计算与3G移动通信

3G移动通信是第三代移动通信的缩略语。3G移动通信是指支持高速数据传输的蜂窝移动通信技术，是将无线通信与互联网相结合的新一代通信技术。目前国际电信联盟确定了三个3G移动通信标准制式：CDMA2000、WCDMA和TD-SCDMA。在我国，中国电信公司、中国联通公司、中国移动公司分别运营这三种不同制式的3G移动通信网络。3G移动通信的代表特征是具有高速数据传输能力，能够提供2Mbit/s以上的带宽。因此，3G移动通信可以支持语音、图像、音乐、视频、网页、电话会议等多种多媒体移动通信业务。

3G移动通信与云计算是相互依存、相互促进的关系。一方面，3G移动通信将为云计算带来数以亿计的移动宽带用户。到2009年7月，全球移动用户已达44亿，普及率达65%。3G移动通信用户已超过5亿，并以惊人的速度增长。2009年是我国的3G元年，当年用户数就超过一千万。这些用户的终端是手机、PDA、便携式计算机等，计算能力与存储空间有限，却有很强的联网能力，对云计算有着天然的需求，将实实在在地支持云计算取得商业成功；另一方面，云计算能给3G移动通信用户提供更好的用户体验。云计算有强大的计算能力、接近无限的存储空间，并支撑各种各样的软件和信息服务，能够为3G移动通信用户提供前所未有的服务体验。

1.4.2 云计算与物联网

物联网即“物物相连的互联网”。物联网通过大量分散的射频识别、传感器，GPS(全球定位系统)、激光扫描器等小型设备，将感知的信息通过互联网传输到指定的处理设施上进行智能化处理，完成识别、定位、跟踪、监控和管理等工作。笼统地看，物联网属于传感网的范畴。其实，传感器的应用历史悠久而且相当普及。那为什么还提物联网的概念呢？物联网是传感网的一个高级阶段，它通过大量信息感知节点采集信息，通过互联网传输和交换信息，通过强大的计算设施处理信息，然后再对实体世界发出反馈和控制信息。

物联网根据其实质用途可以归结为三种基本应用模式：对象的智能标签、环境监控和对象跟踪与对象的智能控制。物联网基于云计算平台和智能网络，可以依据传感器网络用获取的数据进行决策，改变对象的行为进行控制和反馈。

云计算服务物联网的驱动力有以下三个方面：

1）需求驱动：海量数据的处理在目前技术下有高成本压力。云计算充分利用并合理使

用资源，降低运营成本。

2）技术驱动：IT与CT(通信技术)技术融合，推动IT架构的升级。云计算的标准逐渐快速发展。

3）政策驱动：政府的低碳经济与节能减排的政策要求。政府高度关注物联网、云计算等基础设施自助发展战略。

物联网具有全面感知、可靠传递和智能处理三个特征，其中智能处理需要对海量的信息进行分析和处理，对物体实施智能化的控制，这就需要信息技术的支持。云计算具有超大规模、虚拟化、多用户、高可靠性、高扩展性等正式物联网规模化、智能化发展所需的技术。

云计算架构在互联网之上，而物联网主要依赖互联网来实现有效延伸，云计算模式可以支撑具有业务一致性的物联网集约运营。因此，很多研究提出了构建基于云计算的物联网运营平台，该平台主要包括云基础设施、云平台、云应用和云管理。依托公众通信网络，以数据中心为核心，通过多介入终端实现泛在接入，面向服务的端到端体系架构。基于云计算模式，实现资源共享和产业协作，提高效率，降低成本，提升服务。

有观点认为云计算是物联网“后端”支撑关键。所谓物联网的“后端”是实现物联网智能化管理目标和价值追求的关键所在。云计算协同信息处理与计算平台对信息处理与决策。实时感应、高度并发、自主协同和涌现效应等特征对物联网“后端”提出了新的挑战，需要有针对性的研究物联网特定的应用集成问题、体系结构以及标准规范，特别是大量高并发时间驱动的应用自动关联和智能协作问题。在互联网计算领域，将软件的实现与运维和用法相关部分(服务)相剥离，并纳入的互联网级基设中，这是大势所趋。而互联网级基设也是云计算、网格计算的本质所在。

物联网与云计算是交互辉映的关系。一方面，物联网的发展也离不开云计算的支撑。从量上看，物联网将使用数量惊人的传感器[如数以亿万计的RFID(射频识别)、智能尘埃和视频监控等]，采集到的数据量惊人。这些数据需要通过无线传感网、宽带互联网向某些存储和处理设施汇聚，而使用云计算来承载这些任务具有非常显著的性价比优势；从质上看，使用云计算设施对这些数据进行处理、分析、挖掘，可以更加迅速、准确地管理物质世界，从而达到“智慧”的状态，大幅度提高资源利用率和社会生产力水平。可以看出，云计算凭借其强大的处理能力、存储能力和极高的性价比，很自然就会成为物联网的后台支撑平台；另一方面，物联网将成为云计算最大的用户，将为云计算取得更大的商业成功奠定基石。

1.4.3 云计算与移动互联网

互联网与移动通信网是当今最具影响力的两个全球性的网络，移动互联网恰恰融合了两者的发展优势，被称作破坏性创新的云计算，在宽带移动互联网上将成为一种绕不开的趋势。市场调研公司认为，云计算将成为移动世界中的一股爆破理论，最终会成为移动应用的主要运行方式。掌握了云计算核心技术的企业无疑在移动互联网时代可以获得更强的主动性。

移动互联网和云计算是相辅相成的。通过云计算技术，软、硬件获得空前的集约化应用，人们完全可以通过手持一个终端，就能实现传统PC能达到的功能。两者在软、硬件设施成本上的极大节约为中、小企业带来了福音，为人们带来了舒适和便捷。

云计算和移动互联网似乎天生就是绝配。手机拥有便捷性和通信能力等众多天然优势，

而计算能力、存储能力弱，虽然各厂商推出的手机正逐渐向智能化演进，但受限于体积和便携性的要求，短时间内手机的处理能力难以和计算机相比。

从这点出发，云计算的特点更能在移动互联网上充分体现，将应用的计算与存储从终端转移到服务器的云端，从而弱化了对移动终端设备的处理需求，成为新业务的发展瓶颈，在云计算下，只要配备功能强大的浏览器就能应用各种新业务，在后台云计算的存储量和计算能力也解决了手机存储量有限和丢失信息等问题。同时，实现了手机移动和固定计算、便携式计算机的协同。对于追求个性化的移动互联网市场来说，云计算的力量十分关键。

移动互联网时代的来临，对用户来讲，最好的体验是淡化有线和无线的概念。在这样的理念下，云计算有望突破各种终端，包括手机、计算机、电视和视听设备等在存储及运算能力上的限制，显示的内容、应用都能保持一致性和同步性。各大 IT 厂商都是利用云计算制定如 IaaS、PaaS 和 SaaS 策略，希望通过互联网的力量，以软件为基准，将无缝的服务提供给移动终端用户。

云计算正从互联网逐渐过渡到移动互联网。目前社交网站越来越火爆，国外的 Facebook 以及国内的人人网、开心网等都是其典型的代表。社交网站运用云计算思维，实现了网站上各种信息的同步更新。沿着这个思路的移动云计算已经出现，如摩托罗拉公司推出的手机解决方案。

如今，随着一些典型的互联网云计算应用，互联网的云与端之间已经形成了平滑对接，而在移动互联网上，云与端之间还需要“管”来沟通它们之间的鸿沟。浏览器或许将成为重要的“管”角色。

云计算对于云与端的两侧都具有传统模式不可比拟的优势。在云一侧，为内部开发者和业务使用者提供更多的服务，提升基础设施的使用效率和资源部署的灵活性；在端一侧，能够迅速部署应用和服务，按需调整业务使用量。从目前云计算的成功案例中可以看出云计算极大地提高了互联网信息技术的性能，具有巨大的计算和成本优势。

1.4.4 云计算与三网融合

所谓的三网融合，是指广播电视网、电信网与互联网的融合，其中互联网是核心。据国务院三网融合领导小组专家组组长、中国工程院副院长邬贺铨估算，三网融合启动的相关产业市场规模达 6880 亿元人民币。其中电信宽带升级、广电双向网络改造、机顶盒产业发展以及基于音频、视频内容的信息服务系统建设的有效投资额为 2490 亿元人民币，可激发和释放的信息服务与终端消费额近 4390 亿元人民币。

三网融合被纳入“十二五”规划，并明确写入《国务院关于加快培育和发展战略性新兴产业的决定》。业内权威专家认为，三网融合的政策持续加码，将推动电信和广电业务相互进入、广电网络整合、网络运营商角色再定位等一系列革命性变化同步加速。仅中国电信一家运营商，其两年内用于宽带升级的投资将达到近 300 亿元人民币。

云计算使计算能力从分散终端向网络综合服务转变，使商业模式从网络设备基础设施向服务转变，从连续计算机资源向连接个人和设备转变。云计算的基础仍然是宽带，其服务手段和服务对象都需要宽带。社会的各种生活、娱乐和就业都对宽带发展提出了高要求，各国也加大了对宽带业务的投入，各厂商也都在加大对宽带业务的研发。

业内专家认为，随着三网融合政策的出台，以及下一代广电网络的出现，云计算不但会

为现有广电和电信产业带来新商机，还会大大拓展相关产业链，使更多企业受益，为云计算提供切实的应用机会。三网融合和下一代广电网项目是要为用户提供多样、便捷的服务。由于云计算能够大大降低数据存储、计算和分发成本，一些以前无法实现的应用，现在都有可能变成现实。云计算完成计算任务，加上物联网等终端应用和3G移动通信的数据信息传输，将三网整合形成一个系统的信息采集、接收和处理的整体。

三网融合和下一代广电网络的最终目标是构建全数据、全融合的国家骨干网络，借助云计算技术，下一代广电网络还会与传统行业相融合，实现诸如远程教育、网络医疗会诊、股票信息、交通查询、精确广告投放等更多应用。有了云计算技术，一些从事传统行业的企业也能搭上三网融合和下一代广电网的快车。例如，传统的GPS厂商只是生产商，而借助于云计算技术，他们可以成为服务性企业，通过增值业务获得更多收入。中国电子协会计算机委员会专家刘鹏认为，云计算在三网融合以及下一代广电网中的应用，涉及数据存储、数据计算、数据再处理、软件开发、数据传输、网络协同等多个方面，因此需要大量不同类型的企业参与其中。

1.5 各大IT厂商云计算平台特点概述

1. 谷歌公司

谷歌公司在互联网搜索方面建立了强大的商业模式，同时也是云计算领域的重要实践者。谷歌公司在其传统搜索引擎、Gmail、Google Web API等产品的基础上针对自己特定的网格应用程序开展起了众多云计算业务，现在不仅提供云服务给众多个人消费者，而且还涉足企业用户，所提供的服务形式包括应用托管服务和企业搜索等。为了支撑其云计算平台，谷歌公司在IT基础架构方面进行了巨大的投入，它在美国的爱荷华州、北卡罗莱纳州和南卡罗莱纳州等州近期已经完成或正在构建全新的数据中心，平均每个造价高达6亿美元。

这里将主要介绍Google App Engine(谷歌应用引擎,GAE)和Google Apps这两个云服务。

(1) GAE

2008年4月，谷歌公司推出了Google App Engine。这是一个可伸缩的Web应用程序云平台，使用户能够在谷歌公司基础设施上构建和托管Web应用程序。GAE提供了一个SDK(软件开发工具包)，使用户可以在本地使用Java或者Python开发和测试Web应用程序，然后部署在远程GAE的生产环境中进行运行、监控和管理。

GAE开始是免费使用，可提供超过500MB的存储空间，以及每月约500万页面浏览量的免费配额。用户可以创建账户，发布应用程序，而无需承担任何费用和风险。当应用程序启用付费后，配额将提高，但用户只需为使用的超过免费水平的资源付费。

GAE基于谷歌公司早就建立起来的底层平台。这个平台包括MapReduce分布式处理技术、GFS(Google File System,谷歌分布式文件系统)和分布式数据库BigTable。其中，MapReduce API提供Map(映射)和Reduce(化简)处理，GFS和BigTable提供数据存取。

(2) MapReduce分布式处理技术

为了简化分布式编程模式，谷歌公司设计了适合于大规模并行数据处理的编程模型——MapReduce，并将其用于自身的搜索引擎系统。该模型用于大规模数据集(大于1TB)的并行运算，使应用程序编写人员只需要将精力放在应用程序本身上，而关于计算机群的可靠性、