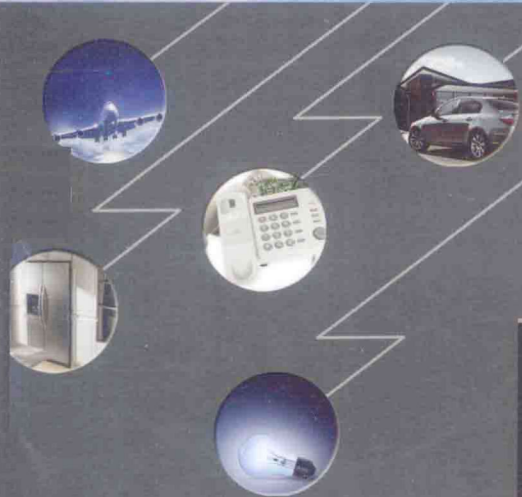




编著 张水平 张凤琴 等 |||||

YUNJISUAN YUANLI
JI YINGYONG JISHU

云计算原理 及应用技术



清华大学出版社
<http://www.tup.com.cn>



北京交通大学出版社
<http://www.bjtup.com.cn>

云计算原理及应用技术

编著 张水平 张凤琴
参著 张月玲 陈爱网 李小青 王 蓉
陈桂茸 管 桦 李卫华 严晓梅

清华大学出版社
北京交通大学出版社
· 北京 ·

内 容 简 介

本书是云数据中心课题组在多年理论研究和科研实践的基础上,参考了大量的研究文献编写而成。本书从广度和深度上系统地讲述了关于云计算、云数据中心的基本概念及其相关技术,包括虚拟化技术、基于 SOA 架构的服务开发方法与工具、云数据中心的规划设计、性能功能测试方法与工具、大数据存储与管理、数据的分析挖掘等。为加深对本书基本知识的理解和掌握,在理论讲述的基础上,还适当加入了实例,如在讲述云数据中心的测试方法时,加入了 Loadrunner 测试工具的使用方法;在讲述服务开发方法时加入了基于微软开发平台的 Web Service 开发实例;在讲解数据的分析挖掘时加入了两类典型的分析工具应用方法介绍等。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13501256678 13801310933

图书在版编目(CIP)数据

云计算原理及应用技术/张水平,张凤琴编著. —北京:北京交通大学出版社:
清华大学出版社,2013.10

ISBN 978-7-5121-1684-9

I. ①云… II. ①张… ②张… III. ①计算机网络 IV. ①TP393

中国版本图书馆 CIP 数据核字(2013)第 245601 号

责任编辑:郭东青 特邀编辑:张诗铭

出版发行:清华大学出版社 邮编:100084 电话:010-62776969 <http://www.tup.com.cn>

北京交通大学出版社 邮编:100044 电话:010-51686414 <http://www.bjtu.com.cn>

印刷者:北京艺堂印刷有限公司

经 销:全国新华书店

开 本:170×235 印张:20.25 字数:386 千字

版 次:2014 年 4 月第 1 版 2014 年 4 月第 1 次印刷

书 号:ISBN 978-7-5121-1684-9/TP·765

印 数:1~1500 册 定价:45.00 元

本书如有质量问题,请向北京交通大学出版社质监组反映。对您的意见和批评,我们表示欢迎和感谢。

投诉电话:010-51686043, 51686008; 传真:010-62225406; E-mail: press@bjtu.edu.cn。

前言 PREFACE



本书在介绍云计算基本概念的基础上，重点围绕云数据中心介绍了与其相关的主要技术，包括云数据中心的基本概念，体系架构，支撑云数据中心的虚拟化技术，云操作系统及实现原理，云数据中心大数据的存储与管理、数据分析技术及典型数据分析工具。同时通过具体应用实例，讲述了相关的实用性技术，包括基于 SOA 架构的服务开发、封装、发布技术及具体开发流程、开发实例，云数据中心规划设计方法，以及 LoadRunner 测试工具使用方法，并给出了通过设计的测试用例对云数据中心进行功能性能测试的思路，对测试结果进行分析的方法和分析实例。为了让读者深入了解并体会云操作系统在云计算中的地位和所扮演的角色，本书用一定的篇幅简要介绍了一个典型的云操作系统——TCloud Elaster Stack 及其功能，并重点介绍了虚拟机的创建及使用方法。

本书的特点一是概念新、系统性强。书中涉及的内容均是课题组近几年理论研究和科研实践的成果。二是深入浅出，易于理解。本书在结构设计上力求脉络清晰，概念叙述上力求通俗易懂，文字表达上力求简练准确。三是实践性强。在云数据中心规划设计、服务开发技术及云操作系统三个章节中，配合内容给出了具体的设计实例，有助于读者对相关技术的理解和实际学习操作。

本书的使用对象为计算机、通信方向等相关专业对云计算有兴趣的科研技术人员。

编者
2013 年 8 月

目录

CONTENTS



第1章 概述	(1)
1.1 云计算及发展	(1)
1.1.1 云计算概念	(1)
1.1.2 云计算特点	(4)
1.1.3 云计算发展	(5)
1.1.4 云计算相关技术	(7)
1.2 云计算体系架构	(10)
1.2.1 云计算参考架构	(10)
1.2.2 云计算技术体系	(17)
1.2.3 云计算分层服务模式	(20)
1.2.4 云服务布署	(21)
1.3 数据中心	(25)
1.3.1 数据中心概念	(26)
1.3.2 数据中心发展	(28)
1.3.3 传统数据中心特点	(31)
1.3.4 云数据中心	(33)
1.4 物联网与云计算	(35)
1.4.1 物联网概述	(35)
1.4.2 物联网与云计算关系	(37)
1.4.3 物联网与云计算集成应用	(38)
小结	(41)

第2章 虚拟化技术	(42)
2.1 虚拟化技术概述	(42)
2.1.1 虚拟化技术定义	(42)
2.1.2 虚拟化技术有关术语	(43)
2.1.3 虚拟化产品介绍	(44)
2.1.4 虚拟化技术优势	(45)
2.1.5 虚拟化的分类及关键特征	(47)
2.1.6 虚拟化技术应用	(48)
2.2 平台虚拟化	(50)
2.2.1 平台虚拟化技术分类	(50)
2.2.2 平台虚拟化技术架构	(52)
2.2.3 虚拟机监视器实现原理	(53)
2.3 资源虚拟化	(55)
2.3.1 网络资源虚拟化	(55)
2.3.2 计算资源虚拟化	(58)
2.3.3 存储资源虚拟化	(59)
2.3.4 软件资源虚拟化	(62)
2.4 虚拟机的动态迁移	(63)
2.4.1 虚拟机动态迁移简介	(63)
2.4.2 虚拟机动态迁移优势	(64)
2.4.3 虚拟机迁移相关技术	(65)
2.4.4 虚拟机迁移分类	(65)
2.4.5 虚拟机动态迁移实现	(66)
2.5 虚拟化资源管理	(68)
2.5.1 云操作系统管理功能	(68)
2.5.2 云操作系统管理实例	(70)
2.5.3 云操作系统开发的技术路线	(74)
小结	(78)
第3章 SOA 架构及开发技术	(79)
3.1 SOA 架构概述	(79)
3.1.1 SOA 定义	(79)
3.1.2 SOA 体系架构	(80)

3.1.3	SOA 特征	(82)
3.1.4	SOA 实现技术	(83)
3.2	Web Service 技术	(84)
3.2.1	Web Service 概述	(84)
3.2.2	Web Service 关键技术	(87)
3.2.3	Web Service 开发方式	(94)
3.3	Web Service 开发及应用实例	(95)
3.3.1	Web Service 服务开发流程	(95)
3.3.2	基于 .NET 的 Web Service 应用程序实例	(96)
3.3.3	基于 Java 的 Web Service 应用程序实例	(116)
小结		(127)
第 4 章	大数据的存储与管理	(128)
4.1	大数据概述	(128)
4.1.1	大数据的基本特性	(128)
4.1.2	大数据带来的挑战	(129)
4.2	大数据存储模型分类	(131)
4.2.1	键值存储模型	(132)
4.2.2	列存储模型	(134)
4.2.3	文档存储模型	(135)
4.2.4	图结构存储模型	(138)
4.2.5	典型 NoSQL 数据库系统	(139)
4.3	Hadoop 大数据存储管理技术	(147)
4.3.1	Hadoop 技术框架	(147)
4.3.2	HBase 数据存储技术	(149)
4.3.3	Hive 数据仓库技术	(154)
4.3.4	Pig 数据处理技术	(156)
4.4	Google 数据处理新技术	(158)
4.4.1	内容索引系统 Caffeine	(159)
4.4.2	数据交互分析系统 Dremel	(160)
4.4.3	大规模图计算方法 Pregel	(162)
4.5	大数据存储管理相关技术	(164)
4.5.1	大数据并行处理技术	(164)
4.5.2	大数据处理 Sharding 技术	(170)

4.5.3 分布式缓存与 Memcached 系统	(173)
4.5.4 数据容错与数据复制技术	(174)
小结	(176)
第5章 云数据中心设计与测试	(178)
5.1 概述	(178)
5.2 云数据中心需求分析	(180)
5.2.1 用户分析	(180)
5.2.2 应用分析	(181)
5.3 云数据中心规划与设计	(182)
5.3.1 规划设计原则	(182)
5.3.2 网络架构的规划设计	(185)
5.3.3 网络互连设备的选择	(187)
5.3.4 服务器规划设计	(188)
5.3.5 存储系统规划设计	(192)
5.4 云数据中心工程实施	(195)
5.5 云数据中心功能性能测试	(197)
5.5.1 LoadRunner 测试软件概述	(198)
5.5.2 网络架构性能测试	(199)
5.5.3 云管理平台功能测试	(210)
小结	(214)
第6章 数据分析技术	(215)
6.1 数据分析概述	(215)
6.1.1 术语和指标	(215)
6.1.2 数据分析体系结构	(219)
6.1.3 数据分析处理的步骤	(220)
6.2 数据预处理方法	(221)
6.2.1 主题分析	(222)
6.2.2 数据抽取	(222)
6.2.3 数据清洗	(223)
6.2.4 数据变换	(224)
6.2.5 数据聚合	(225)
6.2.6 属性约简	(225)

6.3 典型数据分析方法	(226)
6.3.1 描述式数据分析	(226)
6.3.2 关联挖掘分析	(231)
6.3.3 聚类挖掘分析	(239)
6.3.4 分类与预测分析	(247)
6.4 数据分析工具及应用	(257)
6.4.1 数据分析工具简介	(257)
6.4.2 Excel 数据分析	(260)
6.4.3 SPSS 软件数据分析	(268)
小结	(278)
第7章 云数据中心管理与维护实例	(279)
7.1 TCloud ElasterStack 概述	(279)
7.1.1 术语	(279)
7.1.2 系统布署架构	(280)
7.2 TCloud ElasterStack 布署和安装	(282)
7.2.1 系统布署	(282)
7.2.2 系统安装	(286)
7.3 用户管理	(290)
7.3.1 用户的建立和删除	(290)
7.3.2 账号的建立和删除	(292)
7.3.3 管理员的建立和删除	(293)
7.4 虚拟机管理	(295)
7.4.1 模板的创建与管理	(295)
7.4.2 虚拟机的创建与管理	(301)
7.4.3 虚拟机实时检测	(306)
小结	(308)
参考文献	(309)

第1章 概述

工业革命使水、电、交通、电信等实现了社会化、集约化和专业化，因此不需要人人挖井、家家发电，也不需要各部门铺设专门的航线、铁路和公路，这些都已成为全社会的公共基础设施。云计算将使信息技术和信息服务实现社会化、集约化和专业化，即云计算时代不需要家家买计算机、人人当软件工程师，也不需要各部门建设专门的信息系统，云计算将使信息技术和信息服务成为全社会的公共基础设施。

本章主要介绍云计算的概念、特点和发展，云计算的参考架构和技术体系，数据中心的发展和云数据中心的概念，云计算与物联网的应用集成。

1.1 云计算及发展

近年来，随着 Google、Amazon、Salesforce 等国际互联网巨头在云计算大规模应用上的成功，云计算作为新兴技术和商业模式的混合体，得到了国际信息技术（Information Technology, IT）界的重视，发展迅猛。云计算不绝于耳，云计算时代已经到来。

1.1.1 云计算概念

1. 云计算定义

云计算正处于成长初期，尚未形成统一的、公认的定义。美国国家标准与技术研究院（National Institute of Standards and Technology, NIST）给出的定义是：云计算（Cloud Computing）是一种新的资源使用模式，它使用户能通过网络随时、随地、快捷、按需地访问一个可快速部署和配置，仅需少量管理和交互，包括各种网络资源、服务器资源、存储资源、软件资源和服务的资源池。

中国工业与信息化部给出的定义是：云计算是一种通过网络统一组织和灵活调用各种 ICT（Information Communication Technology, ICT）资源，实现大规模计算的信息处理方式。它利用分布式计算和虚拟资源管理等技术，通过网络将分散的 ICT 资源（包括计算与存储、应用运行平台、软件等）集中起来形成共享的资

源池，并以动态按需和可度量的方式向用户提供服务。用户可以使用各种形式的终端（如个人电脑、平板电脑、智能手机甚至智能电视等）通过网络获取 ICT 资源服务。

本书给出一种定义，供读者参考。云计算是一种商业计算模式，它将计算任务分布在由大量计算资源、存储资源和网络资源等构成的资源池上，使用户能够按需获取计算能力、存储空间和信息服务。

云计算概念的直接起源是亚马逊的 EC2（Elastic Compute Cloud）产品和 Google-IBM 分布式计算项目。这两个项目直接使用到了“Cloud Computing”这个概念。之所以采用这样的表述，很大程度上是因为这两个项目与网络的关系十分密切，而“云”又常常被用来表示互联网。因此云计算的原始含义即为将计算能力放在互联网上。当然，云计算发展至今，早已超越了其原始的概念。

2. 云计算带来的改变

云计算正在引导 IT 产业进入一个全新的时代，并将为社会生活带来巨大改变。下面从软件使用方式、软件开发方式和思维方式三个方面进行阐述。

（1）云计算将带来软件使用方式的转变。在传统计算环境中，软件安装在用户的计算机上，当用户同时运行和管理多个软件时，需在不同窗口之间进行切换；在云计算环境中，软件将安装在云端，用户借助浏览器通过网络远程使用软件，对软件的全部操作都在浏览器中完成，当需要在不同软件之间切换时，仅需在浏览器的不同页面之间切换即可。用户不再需要购买昂贵的计算机设备、操作系统和应用软件，不再需要自己维护计算机软、硬件环境，不再需要关心数据的存放位置及应用程序是否需要升级等，也不会再被计算机病毒困扰，这些都由云计算中心负责解决。用户要做的就是选择自己喜爱的云计算服务提供商购买自己需要的服务，并为之付费。用户仅需一个端设备如计算机、笔记本、手机、智能电视等，通过网络就可以访问“云”中的各种资源，包括计算能力、存储空间、网络带宽和数据等。

（2）云计算将带来软件开发方式的转变。在传统计算环境中，企业开发一个软件，需要经过业务需求分析、系统设计、软件开发、测试、运行等步骤，工程量大，周期长；在云计算环境中，企业可直接租用“云”提供的标准软件，或使用“云”提供的标准组件开发自己的应用程序，并将其布署在“云”中，开发效率大大提高，开发周期大大缩短。

（3）云计算将带来思维方式的转变。在传统计算环境中，企业或个人用户主要考虑的是“购买什么样的设备和软件”；在云计算环境中，他们将只需考虑“购买什么样的服务”。用户关心的不再是计算机、设备、软件的功能如何，而是能用这些设备做什么。

总之，最大限度地利用计算、交互、存储乃至应用等资源，以最少的投入获得最优的产出，让最终用户实现绿色、高效、专业的信息服务，正是云计算带来的改变。

3. 云计算环境构成

云计算包括云端和客户端两部分。云端是各种硬件、软件、计算、存储、服务、网络等软硬件构成的资源池，它通过互联网对外提供服务。客户端通过网络访问云的各种资源，用户只需拥有可上网的终端设备，就能享受到想要的各种服务。

如图1-1所示，客户端通过互联网给云端发送请求，云端根据请求进行资源划分并提供相应服务。客户端可以灵活地获得和使用各种资源，并在使用完成后释放这些资源，而不必拥有、控制或了解提供这种资源的底层基础设施，甚至不必知道是谁提供的服务，只需关注自己需要什么样的资源或者想得到什么样的服务。

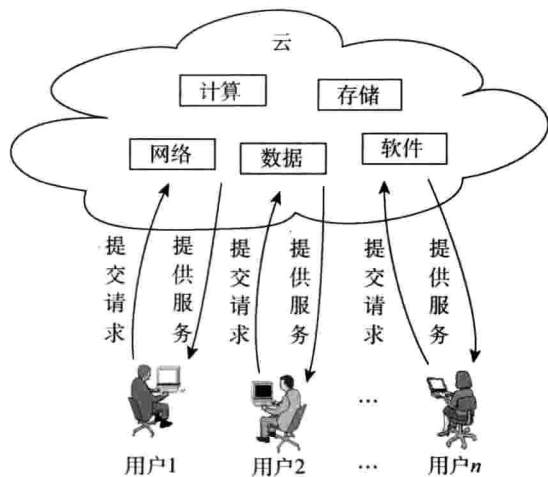


图 1-1 云计算示意图

通常，云端及其客户端使用“客户-服务器”模型，即用户通过端设备接入网络，向云提出需求，云接受请求后组织资源，根据请求内容执行相应操作，并通过网络将执行结果返回给用户，或者为用户提供服务。用户终端的功能将大大简化，诸多复杂的计算与处理过程都将转移到云上去完成。用户所需的应用程序将不需要运行在本地计算机、智能手机等终端设备上，而是运行在云数据中心；用户所处理的数据也无须存储在本地，而是保存在云数据中心；云服务提供商负责这些数据中心和服务器的管理和维护，并保证为用户提供足够强大的计算能力和足够的存储空间。在任何时间和任何地点，用户只要能够连接到互联网，

就可以访问云，实现按需随用。

图 1-2 是一个简单的云计算环境构成图，包括云、用户两大主体。如图所示，云具有动态性。云服务提供商可以根据业务需求动态地向云中添加新的硬件资源，删除旧的或失效的硬件资源。同时，为了提高资源利用率，云提供商会采取一系列的节能等优化措施，如在客户需求减少时关闭未使用硬件的电源，在需求增加时，重新开启被关闭硬件的电源。如图 1-2 中有 3 个服务器组件处于关闭状态。用户使用终端设备通过网络访问云，且该过程也是动态变化的，会有新的云用户不断加入，也会有已经完成访问的云用户不断离开，即每个时刻使用云的用户是不断变化的。

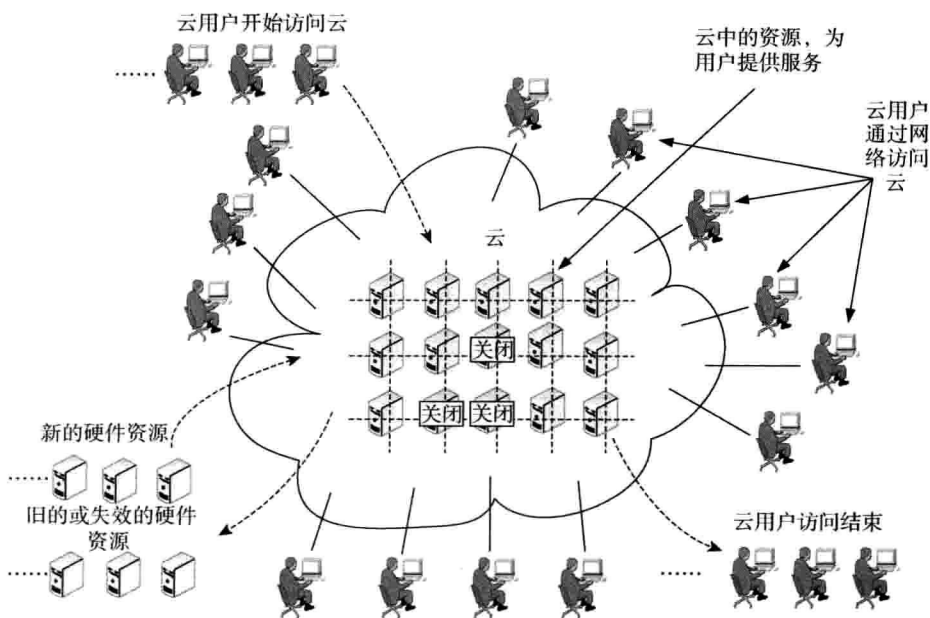


图 1-2 云计算环境构成图

1.1.2 云计算特点

云计算作为一种新的商业模式，具有以下特点。

(1) 一切皆服务。在云计算环境中，硬件、软件、存储、计算、网络等资源均以服务的形式提供和访问。

(2) 网络化访问。云计算环境采用分布式架构，用户可使用各种终端设备，如移动电话、台式机、笔记本、工作站等，通过网络访问云服务。

(3) 按需自助服务。用户可根据需求，通过人机交互自助请求和获取云服

务，而不需要和云服务提供商进行交互。因此，云用户只需具有基本的 IT 常识，经过简单的业务培训就可使用各种云服务，包括服务的申请、使用、管理、注销等，而无须经过专业的 IT 培训。

(4) 多人共享资源池。云服务提供商将各种物理资源和虚拟资源组织成资源池，根据用户需求动态地为多个用户分配资源，提供服务。资源池中的任何物理资源对云服务来说都是抽象的、可替换的，同一资源能够被不同的客户或服务共享。

(5) 快速布署。云计算中心可根据用户需求，自动、弹性地提供和释放各种资源。对云用户来说，其可以在任何时间获得需要的资源或服务，并在使用结束之后将其释放。

(6) 弹性扩展。服务使用的资源规模可随业务量动态扩展，且能保证在动态扩展过程中服务不会中断，服务质量不会下降，且这种扩展对服务使用者和提供者是透明的。

(7) 提供开放的服务访问和管理接口。云计算提供标准化的接口供其他服务调用，方便开发者利用开放接口开发和构建新服务，大大减少了二次开发的工作量。

(8) 持续的服务更新。云计算提供的各种服务能力可随使用者需求的变化不断演化和更新，同时这种改变可做到向下兼容，即保证原有使用者的持续使用。

(9) 自动化管理与快速交付。云计算能有效降低服务的运维成本，平均每百台服务器所需要的运维人员数量应该小于 1 人，且能够对云用户的服务申请进行快速响应，响应时间在分钟级。

(10) 服务可度量。在云计算环境中，资源和服务的使用可监测和控制，且该过程对用户和云提供商透明。云提供商可通过计量去判断每个服务的实际资源耗费，用于成本核算或计费，用户需要向云提供商缴纳一定的费用。

1.1.3 云计算发展

随着数字技术和互联网的迅猛发展，特别是 Web 2.0 技术的成熟和广泛应用，互联网上的数据量高速增长，导致了互联网数据处理能力的相对不足；但互联网上同样存在大量处于闲置状态的计算设备和存储资源，如果能够将其聚合起来统一调度提供服务，则可以大大提高其利用率，让更多的用户从中受益。

目前，用户往往通过购置更多数量或更高性能的终端或者服务器来增加计算能力和存储资源，花费极大。如果用户能够通过高速互联网租用计算能力和存储资源，就可以大大减少对自有硬件资源的依赖，而不必为一次性支付大笔费用而

烦恼。云计算应运而生,它通过实现资源和计算能力的分布式共享,能够很好地应对当前互联网数据量高速增长的势头。

但是,云计算的诞生并不是横空出世,而是经过了很长时间的孕育和发展。早在 20 世纪 60 年代,云计算的基本概念就已被提出。1966 年, Douglas Parkhill 在其 *The Challenge of the Computer Utility* 一书中,对云计算的各种模型进行了详尽的讨论。40 年后, Amazon 公司将其利用率不到 10% 的数据中心组织成云计算平台,面向企业内部和外部提供按需的计算服务,成为现代云计算的先驱。

2006 年 3 月, Amazon 推出弹性计算云 EC2 (Elastic Compute Cloud) 服务和简单存储服务 S3 (Simple Storage Service) 为企业提供计算和存储服务。收费的服务项目包括存储空间、CPU 资源及月租费。这里的月租费与电话月租费类似,存储空间、带宽按容量收费, CPU 根据运算量和时长收费。在诞生不到两年的时间内, Amazon 的注册用户就多达 44 万人,其中包括为数众多的企业级用户。

2006 年 8 月, Google 首席执行官埃里克·施密特在搜索引擎大会上首次提出云计算的概念,其源于 Google 工程师克里斯托弗·比希利亚所做的“Google 101”项目。

2007 年 10 月, Google 与 IBM 开始在美国包括卡内基梅隆大学、麻省理工学院、斯坦福大学、加州大学伯克利分校及马里兰大学等各大学校园推广云计算计划,这项计划希望降低分散式计算技术在学术研究方面的成本,并为这些大学提供相关的软硬件设备及技术支持,而学生则可以通过网络开发各项以大规模计算为基础的研究计划。

2007 年 11 月, IBM 推出了“改变游戏规则”的“蓝云”计算平台,为客户带来即买即用的云计算平台。它包括一系列自我管理和自我修复的虚拟化云计算软件,使来自全球的应用可以访问分布式的大型服务器池,使得数据中心能在类似于互联网的环境下运行各种计算。

2008 年 1 月, Google 宣布在台湾启动“云计算学术计划”,与台湾大学、新竹交通大学等学校合作,将这种先进的大规模、快速计算技术推广到校园。

2008 年 7 月, Yahoo、惠普和英特尔宣布了一项涵盖美国、德国和新加坡的联合研究计划,推出云计算研究测试床,推进云计算。该计划要与合作伙伴创建 6 个数据中心作为研究试验平台,每个数据中心配置 1 400 个至 4 000 个处理器。这些合作伙伴包括新加坡资讯通信发展管理局、德国卡尔斯鲁厄大学 Steinbuch 计算中心、美国伊利诺伊大学香槟分校、英特尔研究院、惠普实验室和 Yahoo 公司。

2008 年 8 月, IBM 宣布将投资约 4 亿美元用于其设在北卡罗来纳州和日本东

京的云计算数据中心改造，并计划 2009 年在 10 个国家投资 3 亿美元建设 13 个云计算中心。

2008 年 8 月，美国专利商标局网站信息显示，戴尔正在申请“云计算”商标。戴尔在申请文件中称，云计算是“在数据中心和巨型规模的计算环境中，为他人提供计算机硬件定制制造”。

2008 年 10 月，微软推出了 Windows Azure 操作系统。Azure（译为“蓝天”）是继 Windows 取代 DOS 之后，微软的又一次颠覆性的转型——通过在互联网架构上打造新云计算平台，让 Windows 真正由 PC 延伸到“蓝天”上。Azure 的底层是微软全球基础服务系统，由遍布全球的第四代数据中心构成。目前，微软已经配置了 220 个集装箱式数据中心，包括 44 万台服务器。

2010 年 3 月，Novell 与云安全联盟共同宣布一项供应商中立计划，名为“可信云计算计划”。

2010 年 7 月，美国国家航空航天局和包括 Rackspace、AMD、Intel、戴尔等支持厂商共同宣布“OpenStack”开放源代码计划；微软在 2010 年 10 月表示支持 OpenStack 与 Windows Server 2008 R2 的集成；而 Ubuntu 已把 OpenStack 加至 11.04 版本中。

2011 年 2 月，思科系统正式加入 OpenStack，重点研制 OpenStack 的网络服务。

1.1.4 云计算相关技术

从本质上来说，云计算并非一个全新的概念。如图 1-3 所示，云计算是计算技术、互联网技术和大规模资源管理等技术的融合与发展，是这些技术的商业实现。下面着重讨论和云计算密切相关且容易混淆的三种计算技术。

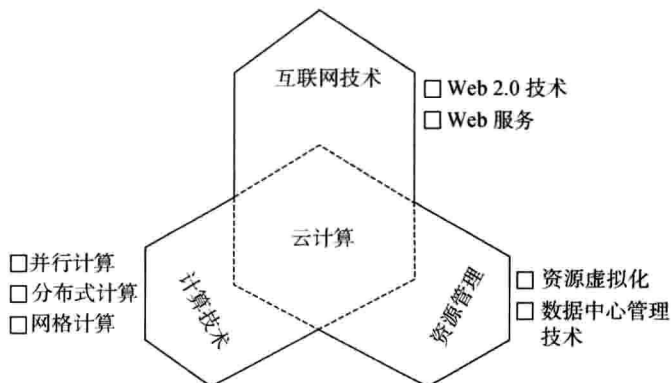


图 1-3 云计算与相关技术的关系

1. 并行计算

并行计算 (Parallel Computing) 是指, 在并行机上将一个应用分解成多个子任务, 分配给不同的处理器, 各个处理器之间相互协同, 并行地执行子任务, 从而达到加速求解速度或提高求解问题规模的目的。并行计算又称高性能计算或者超级计算, 主要用于快速解决大型且复杂的计算问题。

并行计算是相对串行计算而言的, 可分为时间上并行和空间上并行两种, 其中时间并行即流水线技术, 空间并行指多个处理器并发地执行计算任务。并行计算利用并行算法和并行编程语言能实现进程级并行和线程级并行。

为了成功开展并行计算, 必须具备三个基本条件。

(1) 并行机。并行计算需要在并行机上运行, 并行机应至少包含两台或两台以上处理机, 这些处理机通过互连网连接和通信。

(2) 并行度。并行计算要求应用问题必须具有并行度, 即应用问题必须可以分解为多个可并行执行的子任务。将一个应用分解为多个子任务的过程, 称为并行算法设计。

(3) 并行编程。并行计算要求在并行机提供的并行编程环境上, 编制并行运行程序, 从而达到并行求解应用问题的目的。

云计算是在并行计算之后产生的概念, 是由并行计算发展而来的, 两者在很多方面有着共性, 但并行计算不等于云计算, 两者的主要区别如下。

(1) 云计算萌芽于并行计算。云计算的萌芽是从计算机的并行化开始的, 并行机的出现是人们不满足于 CPU 摩尔定率的增长速度, 希望把多个计算机并联起来, 从而获得更快的计算速度。

(2) 并行计算只用于特定的科学领域和专业用户。并行计算的提出主要是为了满足科学技术领域的专业需要, 其应用基本限于科学研究领域。传统并行计算机的使用是一个相当专业的工作, 需要使用者具有较高的专业素质。云计算的用户群则非常广泛, 既包括企业或机构, 也包括普通的终端用户, 对用户的专业需求也比较低。

(3) 并行计算追求的是高速度。在并行计算时代, 人们极力追求的是计算速度。为此各国不惜代价, 采用昂贵的服务器, 不断刷新世界 TOP 500 高性能计算机的排名, 努力在计算速度上超越他国。而云计算时代, 云计算中心的计算能力和存储能力可根据需要逐步增加, 人们将不再追求使用昂贵的服务器, 也不用再去考虑 TOP 500 排名。

2. 分布式计算

分布式计算 (Distributed Computing) 是利用互联网上的计算机的闲置处理能力来解决大型计算问题的计算机科学, 它研究如何把一个需要非常巨大的计算能