



YUNJISUAN
ZIYUAN DIAODU GUANLI

云计算

——资源调度管理

主编 田文洪 赵勇



国防工业出版社

National Defense Industry Press

云计算——资源调度管理

主编 田文洪 赵勇
编委 袁 敏 钟元棕 景 晨 董 旭
胡金安 王浩严 孙夏爽

国防工业出版社

· 北京 ·

内 容 简 介

本书围绕云数据中心基础资源调度管理的关键问题,介绍了云计算的发展背景和挑战性问题,以及主要云服务提供商的云数据中心解决方案、国内外研究现状,提供了对资源调度管理领域内的主要挑战性问题深入分析和探索的内容,希望为读者深入了解相关知识内容和为有兴趣的研究人员提供一些借鉴。

本书可作为相关领域研究人员的参考资料,同时也可作为相关专业高年级本科生和低年级研究生的教材。

图书在版编目(CIP)数据

云计算:资源调度管理 / 田文洪,赵勇主编. —北京:
国防工业出版社,2011.7
ISBN 978-7-118-07533-5

I. ①云... II. ①田... ②赵... III. ①计算机
网络 IV. ①TP393

中国版本图书馆 CIP 数据核字(2011)第 144019 号

※

国防工业出版社出版发行

(北京市海淀区紫竹院南路 23 号 邮政编码 100048)

北京奥鑫印刷厂印刷

新华书店经售

*

开本 787×1092 1/16 印张 9½ 字数 219 千字

2011 年 7 月第 1 版第 1 次印刷 印数 1—4000 册 定价 28.00 元

(本书如有印装错误,我社负责调换)

国防书店:(010)68428422

发行邮购:(010)68414474

发行传真:(010)68411535

发行业务:(010)68472764

前 言

云计算是一种商业计算模型和服务模式,它将计算任务分布在大量计算机构成的不同数据中心,使各种应用系统能够根据需要获取计算能力、存储空间和信息服务。提供资源的网络或数据中心被称为“云”。业界有研究者将云计算列为水、电、气、油之外的第五种公用资源(the 5th utility)。继个人计算机变革、互联网变革之后,云计算被看作是第三次 IT 浪潮,是世界战略性新兴产业的重要组成部分,它将带来生活、生产方式和商业模式的深刻改变,已成为当前全社会关注的热点。

云计算在网络搜索、科学计算、虚拟环境、能源和生物信息等领域都开始了应用和相关探索。IDC 预测,未来 4 年中国云计算将产生 1.1 万亿元的市场。赛迪顾问 2010 年底的《中国云计算产业发展白皮书》中预测未来 3 年,云计算应用将以政府、电信、教育、医疗、金融、石油石化和电力等行业为重点,在中国市场逐步被越来越多的企业和机构采用,市场规模也将从 2009 年的 92.23 亿元增长到 2012 年的 606.78 亿元,年均复合增长率达 87.4%。该报告预计中国云计算产业发展将分为准备阶段(2007—2010)、起飞阶段(2011—2015)和成熟阶段(2015 年以后)。

不少研究者预言“将来的核心竞争在云数据中心”。云数据中心是容纳计算设备资源的集中之地,同时负责对计算设备的能源提供和空调维护等。云数据中心可以是单独建设也可以置于其他建筑之内,可以是分布在不同地理位置的多个系统。云资源汇聚在一起,通过多租户模式服务多个消费者。在物理上,资源以分布式的共享方式存在,但最终在逻辑上以单一整体的形式呈现给用户。资源种类很多,分类角度也不一样,本书所涉及的资源主要包括:

(1) 物理服务器:是构成云数据中心的物理计算设备,每个物理服务器可以提供多个虚拟机,每个物理服务器可以由多个 CPU、内存、硬盘、网卡等构成。

(2) 物理集群:是由多个物理服务器、必要的网络和存储设施构成的物理服务器组。

(3) 虚拟机:是通过虚拟化软件在物理服务器上生成的虚拟计算平台,可以由多个虚拟 CPU、硬盘、网卡等构成。

(4) 虚拟集群:是由多个虚拟机、必要的网络和存储设施构成的虚拟机组。

(5) 共享存储:是为数据中心的计算资源提供大容量存储,可以被所有设备 and 应用共享。

以上资源本书统称为基础资源。

云数据中心的资源调度技术是云计算应用的核心,是云计算得以大规模应用和提高系统性能、兼顾节能减排的关键技术。先进的动态资源调度,对于提高学校、政府、研究机构和企业计算资源的利用效率、节约能源、提高资源共享和降低运营成本都具有极大意义,值得深入系统地学习和研究。

资源调度是将资源从资源提供方分配给用户的一个过程。对于资源过载(即需求大于系统容量)以及需求与容量随着时间的推移而有所差异的事实,通过资源调度管理,可以动态重新分配资源,以便更高效地使用可用资源。业界一般将调度分为作业级调度和设施级调度。作业级调度是指针对具体的运行程序,系统将作业具体分配到哪些资源上运行的问题,例如一些独立的需要较多计算资源和较长运行时间的程序,或者高性能并行处理程序,这些程序往往需要较大规模的高性能计算资源(如云计算)才能很快完成。设施级调度主要指将底层资源作为一种基础设施服务(Infrastructure as a Service, IaaS)提供给用户,用户依据实际情况使用这些资源,例如云数据中心的物理服务器(含 CPU、内存、网络带宽等),虚拟机(含虚拟 CPU、内存、网络带宽等)以及虚拟集群都属于底层基础资源。

本书侧重于设施级(也称应用级)调度,并介绍资源动态综合调度。如果把数据中心作为一个整体与人体比较,资源调度类似人的大脑,是最为核心的功能之一,具有极大的理论意义和实用价值。另外资源监控类似人的眼睛,资源部署类似人的四肢。目前的数据中心调度大多仅实现了简单的初级功能,还有很多问题亟需深入系统解决。针对基础资源调度,主要回答以下三大基本问题:

(1) 调度策略与算法的目标问题:将所需的虚拟机(或虚拟集群)配置在某个数据中心物理服务器(或物理集群)的依据或标准是什么?

(2) 分配问题:具体在哪个数据中心的哪个物理服务器上(或物理集群)分配所需的虚拟机(或虚拟集群)?

(3) 动态迁移问题:当过载,故障等告警发生时,如何将虚拟机迁移到其他物理服务器?

在解决基本问题的同时,实现动态调度而不是静态预先完全设定相关参数,同时兼顾资源的 CPU、存储、网络等的性能,热耗,利用率等特性,本书称为资源的动态综合调度技术问题。

云数据中心需要处理物理和虚拟资源的动态调度这一新问题,以实现高性能、节能减排及降低投资等目标。当前云数据中心的资源调度大多沿袭传统方法,较为简单化,难以满足以上目标的精细化要求。云数据中心调度面临的挑战性问题包括:在考虑配置动态可调虚拟机的分配和迁移以及物理机综合性能时,综合兼顾 CPU、存储、网络等资源因素而非单一因素情况下,如何解决用户需求不一致和资源规格不一致造成的系统性能、能耗以及成本问题。

本书围绕以上关键问题展开,介绍了主要云服务提供商的相关解决方案、国内外研究现状以及对以上挑战性问题的探索,希望为读者深入了解相关知识内容提供一些借鉴。

本书主要内容包括云计算概述(第1章),云数据中心概述(第2章),云资源定义与建模(第3章),云资源管理(第4章),云资源调度策略(第5章),云资源负载均衡调度算法的分析与实现(第6章),云数据中心资源调度模拟系统(第7章),然后是第8章(总结与展望),在附录部分准备了实验与应用的资料。本书适用相关专业高年级本科学生和低年级研究生作为教材使用,同时对于相关研究人员也可作为参考指南。

本书是编辑组成员对有关以上内容大量理论知识与实践经验的积累结果,因时间仓促,可能存在不妥之处,欢迎指正,请发邮件至 CloudSched@gmail.com。

目 录

第 1 章 云计算概述	1
1.1 云计算发展背景	1
1.2 云计算是集大成者	3
1.3 为什么需要云计算?	6
1.4 云计算发展现状和趋势	8
1.5 云计算应用初步分类	10
1.6 云计算的产业链中的不同角色	11
1.7 云计算主要特征和技术挑战	12
1.8 小结	14
思考题	14
参考文献	14
第 2 章 云数据中心概述	15
2.1 云数据中心概述	15
2.1.1 云数据中心介绍	15
2.1.2 云数据中心的需求和挑战	16
2.2 云计算数据中心资源调度需求分析	17
2.2.1 技术需求	17
2.2.2 技术目标	19
2.3 云计算数据中心资源调度研究进展	19
2.4 云计算数据中心资源调度方案分析	20
2.4.1 Google 解决方案	20
2.4.2 Amazon 解决方案	21
2.4.3 IBM 解决方案	21
2.4.4 HP 解决方案	23
2.4.5 VMware 解决方案	25
2.4.6 其他厂家解决方案	25
2.5 云计算数据中心资源调度标准进展	27
2.6 云计算数据中心资源调度关键技术及研究热点	29
2.7 小结	30
思考题	30

参考文献	31
第3章 云资源定义与建模	33
3.1 引言	33
3.2 云数据中心资源建模	34
3.2.1 云数据中心多级体系结构	34
3.2.2 云数据中心涉及的资源	34
3.3 云数据中心资源定义	36
3.4 资源管理	49
3.5 小结	53
思考题	54
参考文献	54
第4章 云资源管理	55
4.1 概述	55
4.1.1 面向基础设施搭建的管理软件	55
4.1.2 面向能效设备控制的管理软件	56
4.1.3 面向虚拟化的数据中心管理软件	57
4.2 云数据中心资源管理的内容	57
4.2.1 用户管理	57
4.2.2 任务管理	59
4.2.3 资源管理	59
4.3 资源管理的目标	60
4.3.1 自动化	60
4.3.2 资源优化	60
4.3.3 简洁管理	61
4.3.4 虚拟资源与物理资源的整合	62
4.4 资源管理的关键问题	62
4.4.1 动态多层次分布式资源监控	62
4.4.2 物理和虚拟资源动态调度	63
4.4.3 物理和虚拟资源动态快速部署与维护	63
4.5 数据中心管理系统案例分析	64
4.6 小结	68
思考题	68
参考文献	68
第5章 云资源调度策略	69
5.1 资源调度关键技术	69
5.2 云计算数据中心调度策略对比分析	70

5.2.1	Amazon 调度策略	70
5.2.2	IBM 调度策略	71
5.2.3	HP 调度策略	71
5.2.4	VMware 调度策略	71
5.2.5	其他方案	72
5.3	主要调度策略分类	73
5.3.1	性能优先	73
5.3.2	成本优先	74
5.4	调度策略约束条件	77
5.5	调度任务执行时间和触发条件	77
5.6	小结	78
	附:基本术语	78
	思考题	79
	参考文献	79
第 6 章	云资源负载均衡调度算法的分析与设计	81
6.1	云计算数据中心综合负载均衡调度策略概述	81
6.2	云计算数据中心负载均衡调度策略中主要调度算法分析	82
6.2.1	轮转调度算法(RR)	82
6.2.2	加权轮转调度算法(wRR)	83
6.2.3	目标地址哈希调度算法(DH)	83
6.2.4	源地址哈希调度算法(SH)	84
6.2.5	最小链接算法(LC)	84
6.2.6	加权最小链接算法(WLC)	85
6.3	几种动态综合负载均衡调度算法分析	86
6.3.1	综合利用率乘法	86
6.3.2	综合负载基准对比法	86
6.3.3	动态反馈综合负载均衡调度算法	87
6.4	负载均衡调度算法对比	87
6.5	动态反馈综合负载均衡调度算法详细设计	88
6.5.1	优化目标	88
6.5.2	动态反馈综合负载均衡调度算法	89
6.6	小结	91
	思考题	91
	参考文献	91
第 7 章	云数据中心资源调度模拟系统	92
7.1	简介	92
7.2	现有系统分析	93

7.2.1	CloudSim	93
7.2.2	CloudAnalyst	94
7.3	CloudSched 模拟系统设计框架	97
7.4	模拟系统分析——以考虑需求特性的调度算法为例	98
7.4.1	考虑需求特性调度算法解决的问题	98
7.4.2	考虑需求特性调度算法主要步骤	98
7.4.3	考虑需求特性调度算法流程图	99
7.4.4	考虑需求特性调度算法伪代码	99
7.4.5	算法类图	100
7.4.6	算法时序图	109
7.5	各种算法结果分析	112
7.5.1	其他算法思想简介	112
7.5.2	4 种算法数据中心不均衡度对比	113
7.5.3	4 种算法物理服务器不均衡度对比	113
7.5.4	4 种算法运行时间对比	114
7.6	小结	114
	思考题	115
	参考文献	115
第 8 章	总结与展望	116
附录 I	虚拟服务器管理系统 VMmanager	119
附 I.1	简介	119
附 I.2	体系结构	119
附 I.3	用户界面设计	121
附 I.4	安装与配置	121
附 I.5	VMmanager 使用说明	122
附录 II	云计算模拟分析器 CloudAnalyst	128
附 II.1	CloudAnalyst 的安装过程	128
附 II.2	CloudAnalyst 的运行过程及配量方法	133

第1章 云计算概述

本章主要介绍云计算发展背景、驱动因素、发展趋势、应用分类、主要特征与挑战等基本概念。

“云计算是一种新的商业计算模型和服务模式”。

本章主要内容：

云计算发展背景

发展云计算的主要技术积累

云计算发展现状与趋势

云计算应用分类

云计算主要特征与挑战

1.1 云计算发展背景

全球正在进入云计算的时代。云计算是一种新的商业计算模型和服务模式,其核心的概念是不再依赖本地计算机来做计算,而是依赖由第三方运营的集中的计算和存储资源。云计算的概念可以追溯到1961年,计算机界的先驱 John McCarthy 在麻省理工学院的百年纪念会上发言说:“计算某一天可能会像电话一样被作为公用资源(Public Utility),……计算机资源会成为一种新的重要的产业的基础。”1966年 D. F. Parkhill 则在其经典的著作《The Challenge of the Computer Utility》中大胆预测计算能力如同水和电一样被提供给大众。今天业界称云计算是继水、电、汽、油之后的第五种公用资源(“the 5th utility”)。

人们常用以下的两个经典故事解释云计算的应用^[9]。

故事一:Tom 是一家公司的员工,某天公司派 Tom 去伦敦出差。那么, Tom 想要了解他所乘坐的航班信息,从他住所到机场的最佳路线以及伦敦最新的天气以及住宿信息等。在未来,所有的这些信息都可以通过云计算来提供。云计算将与各种各样的终端(例如个人电脑、PDA、手机、电视等)进行连接,为用户提供广泛、主动、高度个性化的服务。

故事二:Bob 是这家公司的另一个员工,不过公司并没有派他出差,于是他照常到公司上班。到了公司,他打算管理一下自己最近的任务,于是他可以通过 Google Calendar 来管理自己最近的日程安排;整理完日程, Bob 可以通过 Gmail 收发邮件,通过 GTalk 来与同事朋友进行联系;如果他这时打算开始工作,可以通过 Google Docs 来编写在线文档,在这个过程中,如果他需要查阅相关论文,可以通过 Google Scholar 进行搜索,他也可以使用

Google Translate 翻译一些英文,他甚至可以使用 Google Charts 来绘制一些图表;如果 Bob 工作累了,他可以通过 Google Blogger 来分享日志,通过 Google 的 YouTube 来分享视频,通过 Google 的 Picasa 来编辑分享图片。

“云计算”是英文 Cloud Computing 的直译,“云计算”为何被称为“云”计算,一种流行的说法解释是: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,如图 1-1 所示。因为人们通过浏览网页访问互联网可能需要经过多个中间转接过程,而这些对于用户是透明的。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服务方式的时候就选择了“云计算”这个名词,意在不关心网络的转发过程,而去关注服务器端和客户端的服务与应用。这个解释非常有趣和新潮,但是却容易让人们陷入云里雾里,特别是中文的很多与云相关的词汇都是贬义词,使得很多人从“云计算”字面意思解释产生误会,这里有必要对云计算给予系统定义和特征分析。

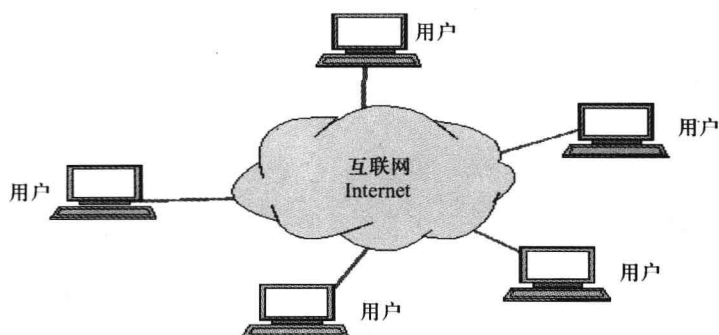


图 1-1 看似“云”的互联网

关于云计算的定义有很多种,维基百科的定义是:“云计算是一种商业计算模型和信息服务模式。它将计算任务分布在大量物理计算机服务器或虚拟服务器构成的不同数据中心,使各种应用能够根据需要获取计算能力、存储空间和信息服务^[19]”。伯克利云计算白皮书则定义为“云计算包括互联网上各种服务形式的应用以及数据中心提供这些服务的软硬件设施^[3]”。我们综合云计算的特点把它定义为:“一种由规模经济效应所驱动的大规模分布式计算模式,把抽象化的、虚拟化的、动态可扩展的、有效管理的计算、存储、平台和服务等资源池,通过互联网按需提供给外部用户^[2]。”它区别于传统的计算模式在于:①它的大规模性;②能被封装成一个抽象的实体,提供给用户不同级别的服务;③是基于规模经济效应的;④服务能够动态配置,按需供应。

云计算通过网络可以提供的计算与信息服务与应用如图 1-2 所示,包括计算、存储、网络、服务、软件等多种形式。

自 1966 年 D. F. Parkhill 在其经典著作《The Challenge of the Computer Utility》中大胆预测计算能力如同水和电一样被提供给大众之后,众多计算机科学家不断探索与创新去努力实现这一目标,然而始终未能找到一个为业界和用户所广泛接受的成功方案,很多方案被提出又推翻或者未能大规模应用^[8]。随着网络基础设施的不断改进以及互联网应用的飞速发展,云计算——这个集计算机领域的先进技术的大成者的出现,正在为越来越多的人所接受。今天业界称云计算是水、电、汽、油之后的“the 5th utility”——第五种公用资源。

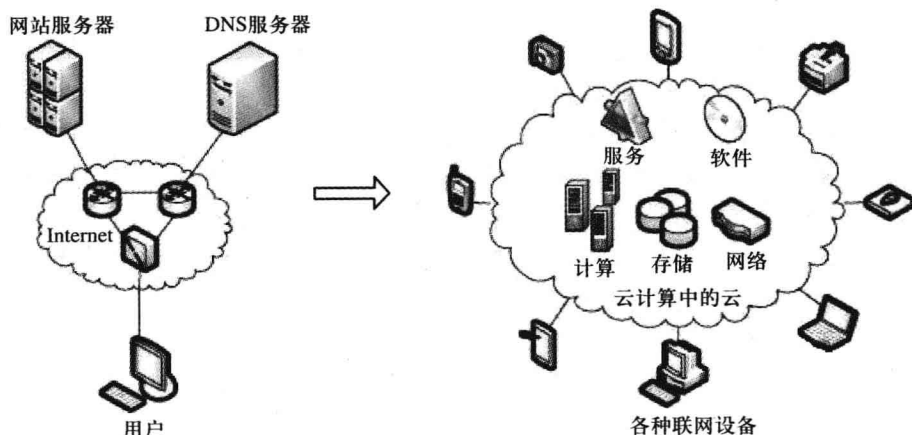


图 1-2 云计算提供的服务与应用

也有人称之为“The Poor Men’s Supercomputer”（穷人的超级计算机）。主要是因为用户不再需要购买和维护庞大的计算机集群,只需通过网络,按需消费计算资源。

1.2 云计算是集大成者

在计算机科学技术发展的历史上,经常出现一些里程碑式的技术。这些技术产生的时间或远或近,对当今世界的 IT 运用和服务模式产生了巨大的影响。这些技术包括并行计算、网格计算、效用计算、普适计算、软件作为一种服务和虚拟化计算等^[9]。云计算不是一蹴而就的,而是从这些技术中逐渐演进而来,既一脉相承,又有所不同。业界普遍认为云计算是各种先进计算和服务技术的集大成者,如图 1-3 所示。

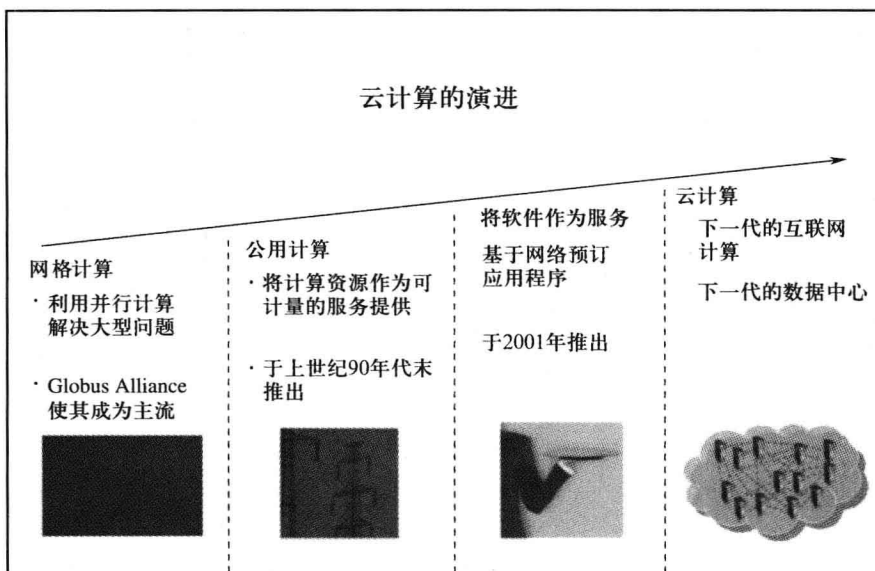


图 1-3 云计算演进过程

1. 并行计算

并行计算(Parallel Computing)是将一个科学计算问题分解为多个小的计算任务,并将这些小任务在并行计算机上同时执行,利用并行处理的方式达到快速解决复杂运算问题的目的。并行计算一般应用于诸如军事、能源勘探、生物、医疗等对计算性能要求极高的领域,因此也称为高性能计算(High Performance Computing)或超级计算(Super Computing)。并行计算机是一群同构处理单元的集合,这些处理单元通过通信和协作来更快地解决大规模计算问题。常见的并行计算机系统结构包括共享存储的对称多处理器(SMP)、分布式存储的大规模并行机(MPP)和松散耦合的分布式工作站集群(COW)等。解决计算问题的并行程序往往需要特殊的算法,编写并行程序需要考虑很多问题之外的因素,例如各个并发执行的进程之间如何协调运行、任务如何分配到各个进程上运行等。

并行计算机可以说是云环境的重要组成部分(IBM 云计算)。与云计算的思想相似,目前世界各国已经集中建立了若干超级计算中心来服务于该区域内有并行计算需求的用户,并采用分担成本的方式进行付费。但是,云计算与传统意义上的并行计算相比,又存在明显的区别。首先,并行计算需要采用特定的编程范例来执行单个大型计算任务或者运行某些特定应用,而云计算需要考虑的是如何为数以千万计的不同种类应用提供高质量的服务环境,以及如何提高这个环境对用户需求的响应从而加速业务创新。一般来说,云计算对用户的编程模型和应用类型等没有特殊限定,用户不再需要开发复杂的程序,就可以把他们的各类企业和个人应用迁移到云计算环境中。其次,云计算更加强调用户通过互联网使用云服务,而在云中利用虚拟化进行大规模的系统资源抽象和管理。在并行计算中,计算资源往往集中在单个数据中心的若干台机器或者是集群上。云计算中资源的分布更加广泛,正如上文所述,它已经不再局限于某个数据中心,而是扩展到了多个不同的地理位置。同时,由于采用了虚拟化技术,云计算中的资源利用率可以得到有效的提升。由此可见,云计算是互联网技术和信息产业蓬勃发展背景下的产物,完成了从传统的、面向任务的单一计算模式向现代的、面向服务的多元计算模式的转变。

2. 网格计算

网格计算(Grid Computing)是一种分布式计算模式。网格计算技术将分布在不同地理区域、不同管理域里的服务器、存储系统和网络连接在一起,形成一个整合系统,为用户提供功能强大的计算及存储能力来处理特定的任务。对于使用网格的最终用户或应用程序来说,网格看起来就像是一个拥有超强能力的虚拟计算机。网格计算的本质在于以高效的方式来管理各种加入了该分布式系统的异构松耦合资源,并通过任务调度来协调这些资源合作完成一项特定的计算任务。

可见,网格计算着重于管理通过网络连接起来的异构资源,并保证这些资源能够充分为计算任务服务。通常,用户需要基于某个网格的框架来构建自己的网格系统,并对其进行管理,在其上执行计算任务。云计算则不同,用户只需要使用云中的资源,而不需要关注系统资源的管理和整合。这一切都将由云提供者进行处理,用户看到的是一个逻辑上单一的整体。因此,在资源的所属关系上存在着较大差异,也可以说在网格计算中是多个零散资源为单个任务提供运行环境,而在云计算中是单个整合资源为多个用户提供服务。

Ian Foster 的文章^[2]对二者进行了更加深入详细的对比。

3. 效用计算

效用计算 (Utility Computing) 强调的是 IT 资源, 如计算和存储等, 能够根据用户的要求按需地提供给用户, 而且用户只需要按照其实际使用情况付费。效用计算的目标是 IT 资源能够像传统公共设施 (如水和电等) 一样的供应和收费。效用计算使得企业和个人不再需要一次性的巨额投入就可以拥有计算资源, 而且能够降低使用和管理这些资源的成本。效用计算追求的是提高资源的有效利用率, 最大程度地降低资源的使用成本和提高资源使用的灵活性。

效用计算所提倡的资源按需供应、用户按使用量付费的理念与云计算中的资源使用理念相符。云计算也可以按照用户的资源需求分配运算、存储、网络等各种基础资源。比效用计算更进一步的是, 云计算已经有了很多实际应用案例, 所涉及的技术和架构可行性更强。云计算所关注的是如何在互联网时代以其自身为平台开发、运行和管理不同的服务。云计算不但注重基础资源的提供, 而且注重服务的提供。在云计算环境中, 不但硬件等 IT 基础资源能够以服务的形式来提供, 应用的开发、运行和管理也是以服务的形式提供的, 应用本身也可以采用服务的形式来提供。因此, 云计算与效用计算相比, 技术和理念所涵盖的范围更广泛、更可行。

4. 普适计算

Mark Weiser 提出普适计算 (Ubiquitous (Pervasive) Computing) 的思想, 并预测计算模式会发展为普适计算。20 世纪 90 年代末, 普适计算这一概念得到广泛关注, 逐渐升温。1999 年, IBM 正式提出普适计算的概念。同年, 召开第一届 Ubicomp 国际会议。2000 年, 第一届 Pervasive Computing 国际会议。2002 年, IEEE Pervasive Computing 期刊创刊。

普适计算的促进者希望嵌入到环境或日常工具中去的计算能够使人更自然地和计算机交互。而普适计算的显著目标之一则是使得计算机设备可以感知周围的环境变化, 从而根据环境的变化做出自动的基于用户需要或者设定的行为。

普适运算技术运用无线网络科技, 让人们在不受时空限制的环境下享用资讯; 而与一般的行动计算 (Mobile Computing) 不同在于具有情境感知 (Context Aware) 的特性。普适运算技术可以透过感应个人所在的位置、所处的环境资讯、个人的情形及任务, 来提供最有效能的使用环境。

5. 软件作为一种服务 (Software as a Service)

SaaS (Software as a service) 的意思是软件即服务, SaaS 的中文名称为软营或软件运营。SaaS 是基于互联网提供软件服务的软件应用模式。作为一种在 21 世纪开始兴起的创新的软件应用模式, SaaS 是软件科技发展的最新趋势。SaaS 是一种软件布局模型, 其应用专为网络交付而设计, 便于用户通过互联网托管、部署及接入。SaaS 应用软件的价格通常为“全包”费用, 囊括了通常的应用软件许可证费、软件维护费以及技术支持费, 将其统一为每个用户的月度租用费。对于广大中小型企业来说, SaaS 是采用先进技术实施信息化的最好途径。但 SaaS 绝不仅仅适用于中小型企业, 所有规模的企业都可以从 SaaS 中获利。

2008 年前, IDC 将 SaaS 分为两大组成类别: 托管应用管理 (hosted AM) —— 以前称作

应用服务提供(ASP),以及“按需定制软件”,即 SaaS 的同义词。从 2009 年起,托管应用管理已作为 IDC 应用外包计划的一部分,而按需定制软件以及 SaaS 被视为相同的交付模式对待。

目前,SaaS 已成为软件产业的一个重要力量。只要 SaaS 的品质和可信度能继续得到证实,它的魅力就不会消退。

6. 虚拟化计算

虚拟化是一个广义的术语,在计算机方面通常是指计算元件在虚拟的基础上而不是真实的基础上运行。虚拟化技术可以扩大硬件的容量,简化软件的重新配置过程。CPU 的虚拟化技术可以单 CPU 模拟多 CPU 并行,允许一个平台同时运行多个操作系统,并且应用程序都可以在相互独立的空间内运行而互不影响,从而显著提高计算机的工作效率。

虚拟化(Virtualization)技术最早出现在 20 世纪 60 年代的 IBM 大型机系统,在 70 年代的 System 370 系列中逐渐流行起来,这些机器通过虚拟机监控器(Virtual Machine Monitor, VMM)的程序在物理硬件之上生成许多可以运行独立操作系统软件的虚拟机(Virtual Machine)实例。随着近年多核系统、集群、网络甚至云计算的广泛部署,虚拟化技术在商业应用上的优势日益体现,不仅降低了 IT 成本,而且还增强了系统安全性和可靠性,虚拟化的概念也逐渐深入到人们日常的工作与生活中。

虚拟化是一个广义的术语,对于不同的人来说可能意味着不同的东西,这要取决于其所处的环境。在计算机科学领域中,虚拟化代表着对计算资源的抽象,而不仅仅局限于虚拟机的概念。例如对物理内存的抽象,产生了虚拟内存技术,使得应用程序认为其自身拥有连续可用的地址空间(Address Space),而实际上,应用程序的代码和数据可能是被分隔成多个碎片页或段,甚至被交换到磁盘、闪存等外部存储器上,即使物理内存不足,应用程序也能顺利执行。

虚拟化技术与多任务以及超线程技术是完全不同的。多任务是指在一个操作系统中多个程序同时并行运行,而在虚拟化技术中,则可以同时运行多个操作系统,而且每一个操作系统中都有多个程序运行,每一个操作系统都运行在一个虚拟的 CPU 或者是虚拟主机上;而超线程技术只是单 CPU 模拟双 CPU 来平衡程序运行性能,这两个模拟出来的 CPU 是不能分离的,只能协同工作。

1.3 为什么需要云计算?

云计算是互联网和信息社会发展引起海量信息处理需求的必然结果。它的商业模式较之网格等先前的模式更加为全球企业和客户所接受,并被广泛应用。可以概括地讲,云计算是技术和社会需求发展的必然结果。云计算集成了计算机界先前的大规模数据中心、虚拟化技术、软件作为服务等先进技术。

基于互联网的信息爆炸是驱动云计算的主要因素:2006 年全球大约产生 161EB(1EB 相当于 10 亿 GB)的数据,印成书是地球到太阳距离的 10 倍;2007 年全球产生 280EB 数据,全世界平均每人 45GB;而人类历史 5000 年的文字记载只有 5EB 数据。数据的生成正以惊人的速度增长,据统计,2010 年全球有 50% 的企业,每日生成的数据量超过 1TB,有 5% 的企业每日生成 50TB 以上的数据。而数据是各行各业最关心、最有价值的东西,

而存储、处理这些数据需要大规模集中整合的资源,如图 1-4 所示。

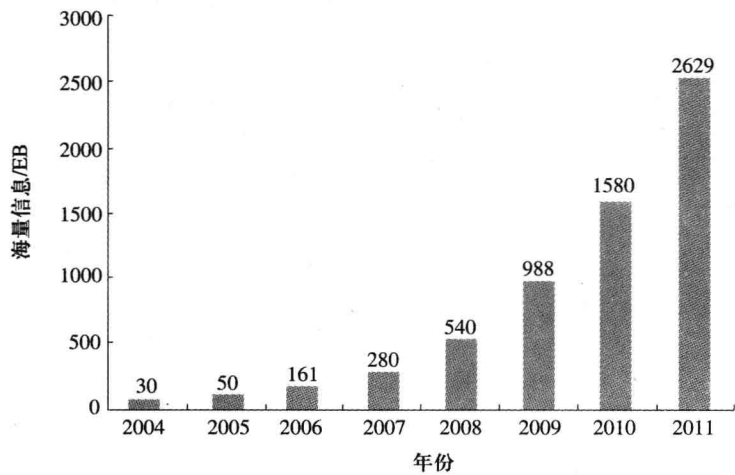


图 1-4 全球信息量的增长过程

计算机性能的增长和多核处理器的发展是云计算的另一个驱动因素。根据全球计算机 500 强 Top500.org 的统计,计算机的性能每年都在呈指数级的增长,目前世界上最快的天河 1 号 A 超级计算机,总核数达 186368 个,峰值性能达每秒 4.701 千万亿次浮点运算(PFlops)。如此强大的计算能力需要新的计算模式来管理和应用这些计算资源,如图 1-5 所示。

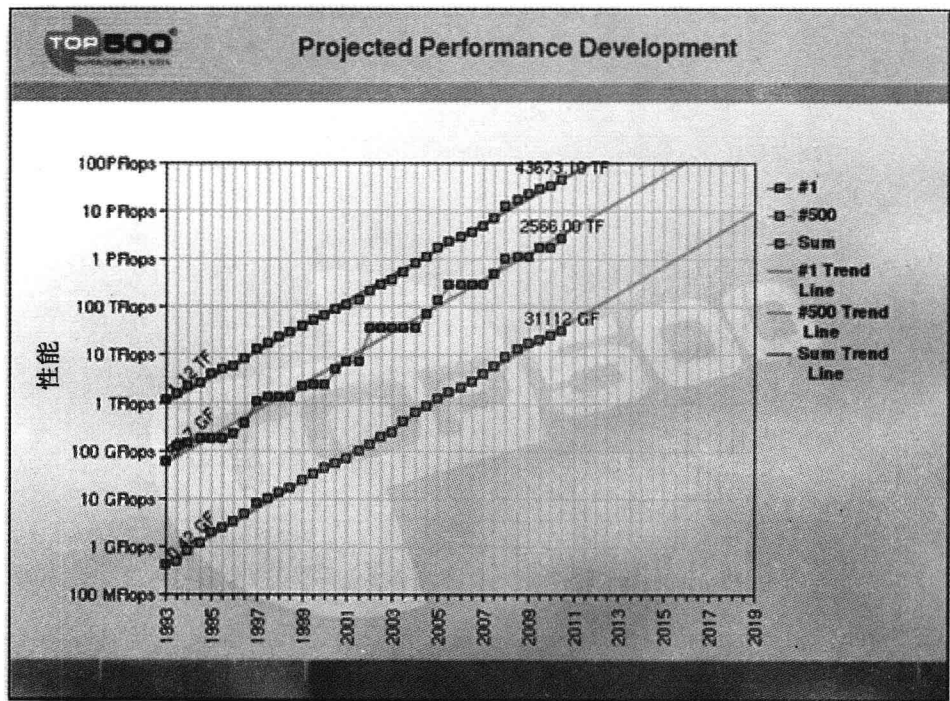


图 1-5 计算性能的增长

云计算普遍为人们所接受,还在于它提供集中统一的管理和维护,资源可以在多用户之间共享,总体上提高资源的利用效率。在传统的 IT 架构和集群管理模式下,各企业和机构需要购置维护自己的硬件软件资源,要么由于估算不足造成资源达不到企业进行大规模存储和计算的需求,要么由于要满足最大可能的容量,而造成资源在正常状况下闲置浪费。统计数据显示,大多数企业的数据中心和集群平均资源利用率不到 30%。而与此同时,企业用于管理维护自身的 IT 基础设施的开销则在逐年增长,能耗开销也在不断加大,如图 1-6 所示。

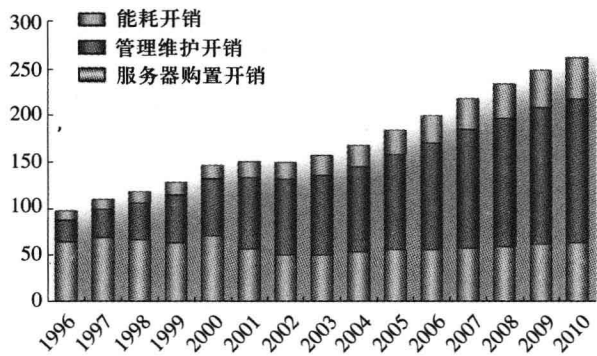


图 1-6 企业 IT 维护开销

1.4 云计算发展现状和趋势

云计算正逐渐被人们所关注,如图 1-7 所示。

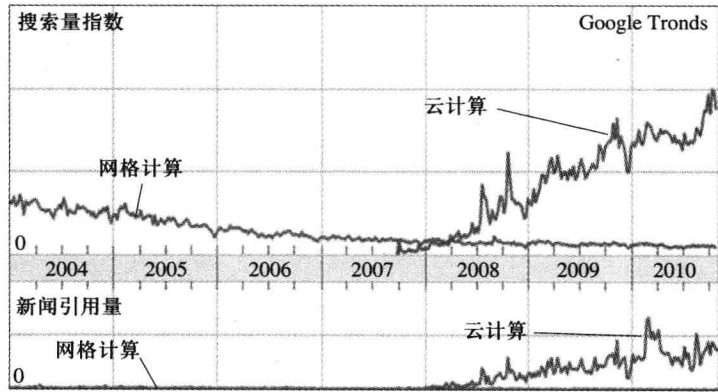


图 1-7 云计算关注度

IBM、Intel 公司等与美国高校 2005 年左右发起了云计算虚拟实验室项目。先在邻近 IBM 总部的北卡罗莱纳州立大学(North Carolina State University)开始实验,非常成功。IBM 又于 2007 年联合 Google 发起云计算,被称为挑战 Intel & Microsoft 传统计算模