# 云计算环境下
# 能耗感知模型与方法
# 进展研究

曲海平　著

清华大学出版社

# 云计算环境下
# 能耗感知模型与方法
# 进展研究

曲海平 著



清华大学出版社
北京

## 内 容 简 介

本书基于服务的负载与性能模型的研究完成服务能耗感知决策。在此基础上研究虑及能耗的全局资源优化布局算法，实现整个平台系统的高能低耗。

全书共分7章：第1章介绍在云计算中心开展能耗感知研究的背景和意义，进而阐述本书的研究内容和贡献；第2章从技术手段、模型构建和能耗管理三个方面介绍国内外研究现状；第3章提出能耗感知的云计算平台管理框架EADC，该框架以大数据集成应用平台为基础，构建了包括VEMS(服务集群内部)与DCMS(平台系统)两级的能耗优化管理；第4章研究统计回归分析方法，解决大范围变动下的负载特征提取，研究高阶MAP拟合的策略与机制，解决长时间相关负载性能分析中复杂度与时间开销过高的问题；第5章研究在性能分析基础上服务节点的按时定量切换的能耗感知决策机制，解决负载的有效整合与节点状态的按需切换；第6章研究兼顾性能与能耗的虑及服务迁移的全局资源优化调度算法，解决服务请求突发时因负载整合而造成服务质量严重下降的问题；第7章对本书的工作进行了总结，并阐述下一步的工作方向。本书的研究成果有助于提高云计算平台能耗有效性，具有重要的理论意义和实际应用价值。

本书的读者对象是云计算中心的信息主管、云计算平台的系统设计开发人员，以及对云计算平台有所了解或感兴趣的广大科技工作者和研究人员。本书也可以作为高等院校研究生与本科生云计算相关课程的参考教材。



**图书在版编目(CIP)数据**

云计算环境下能耗感知模型与方法进展研究/曲海平著. —北京：清华大学出版社，2018
ISBN 978-7-302-49276-4

Ⅰ. ①云…　Ⅱ. ①曲…　Ⅲ. ①云计算－应用－能耗计算－研究　Ⅳ. ①TK011-39

中国版本图书馆CIP数据核字(2018)第004501号

**责任编辑**：白立军
**封面设计**：傅瑞学
**责任校对**：徐俊伟
**责任印制**：李红英

**出版发行**：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦A座　　邮　　编：100084
社 总 机：010-62770175　　邮　　购：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
课件下载：http://www.tup.com.cn，010-62795954

**印 装 者**：三河市铭诚印务有限公司
**经　　销**：全国新华书店
**开　　本**：185mm×230mm　　**印　　张**：10.5　　**字　　数**：173千字
**版　　次**：2018年3月第1版　　**印　　次**：2018年3月第1次印刷
**印　　数**：1～800
**定　　价**：49.00元

产品编号：073668-01

# 前 言

云计算具有超大规模、虚拟化、高扩展性等特点，可以为企业和用户提供各种低成本的计算资源和IT服务，从而为实现高效信息化管理和海量数据服务提供强有力的计算支撑。但是管理这些云程序需要消耗大量电能并带来很大的操作开销，同时也会对环境造成巨大的负面影响。随着云计算平台规模的不断扩大，一方面数量不断增长、处理能力不断增强的服务器带来更多的能量消耗；另一方面每个服务器过低的利用率又造成巨大的电能浪费。国家能源局原局长张国宝2012年表示，他了解的中国联通云计算中心的能耗数据显示，该中心每年耗电99亿千瓦时，以中国目前标准煤的效能，需要消耗92万吨标准煤才能提供足够的电力供中国联通云计算中心的能耗需求；与中国联通云计算中心对应，中国电信云计算中心年耗电112亿千瓦时，总计年需消耗102.95万吨标准煤。所以，在云计算环境下开展能耗研究势在必行。

造成云计算平台能耗效率低下的一个重要原因是空闲服务器消耗的功耗没有随着其状态的空闲而线性降低，即使在诸如10% CPU利用率的情况下，服务器也要消耗超过50%的功耗。由此实现服务器状态的按需切换，进而缩减在线服务器规模是节省能耗的有效方式，最后的资源分配方案就是保证服务质量的前提下实现能耗尽可能的降低。由此可以得出如下两个结论。

(1) 云计算平台迫切需要进行能耗感知的研究，但是必须在服务性能优化与服务质量保证的前提下。这就需要深入研究服务能耗与性能的关系，寻求性能保证与能耗节省两者的和谐统一。

(2) 云计算平台能耗感知的实现是以其上所有服务的能耗感知为基础的。这就需要构建服务与平台两层的能耗感知系统框架，同时深入研究两者的交互关系，实现从局部到全局的高能低耗。

服务能耗感知研究是当前云计算能耗管理领域的研究热点与趋势，这些研究已形成

了较为成熟的研究模式，并积累了丰富的模型、方法与经验。但是当前相关研究没有很好地实现性能保证与能耗降低的优化权衡。

本书针对当前云计算平台存在的高性能保证与低能耗需求这一对矛盾的需求问题，从全面认识服务负载特性以及服务性能分析模型构建的角度出发，完成服务的能耗感知方案决策，在充分调动服务器集群的计算能力的同时完成服务能耗的有效降低，并在此基础上通过全局资源在服务间的有效调度与布局，实现整个高可扩展云平台的所有服务的高能低耗。具体研究内容包括以下五个部分。

**1. 大范围变动下的负载与性能模型**

对于多类方式描述的大范围变动请求服务时间，传统上采用多类服务时间定律（基于单类服务时间定律扩展得到）不适用于实际监控环境，因为实际中无法获取每类请求对应的利用率信息 $D_{i,j}$。本书采用统计回归的数学模型求解计算每类请求的平均服务时间。假定将请求划分为 $N$ 类，系统中有 $M$ 个资源，根据利用率定律（Utilization Law）$U_j=B_j/T=D_j\times X_0$，$j$ 代表第 $j$ 个资源，$X_0$ 是整个系统的吞吐量，$D_j$ 是请求在第 $j$ 个资源消耗的时间，$T$ 是监控窗口大小，$B_j$ 是第 $j$ 个资源忙碌期间（简称忙期），将利用率定律扩展为多类请求情况，得到第 $j$ 个资源在监控窗口 $k$ 的利用率公式为 $\sum_i N_{i,k}\times D_{i,j}=U_{j,k}\times T$，$1\leqslant i\leqslant N$，$1\leqslant j\leqslant M$。由此估算出 $D_{i,j}$ 的近似值 $C_{i,j}$，则第 $j$ 个资源在第 $k$ 个窗口 $T$ 期间利用率的估算值为 $U'_{j,k}=\dfrac{\sum_i N_{i,k}\times C_{i,j}}{T}$。

**2. 时间长相关负载下的性能模型**

长相关（Long-Range Dependent，LRD）作为云计算环境下 Internet 负载的一个重要特征，与不具有相关性的更新过程（如泊松分布）负载和短相关的非更新过程负载相比，其对性能的影响更大。基于 MAP（Markovian Arrival Process，马尔可夫到达过程）构建的性能模型可以通过矩阵分析法快速、有效地求解，且 MAP 模型描述负载的能力随着阶数增大而增大，所以以往研究常利用有限高阶 MAP 描述长相关负载。针对低阶 MAP 不准确而高阶 MAP 又存在复杂度高的问题，本书通过降低解决高阶 MAP 拟合问题复杂度最终达到利用非线性优化算法求解多个低阶组合 MAP(2)s（Markovian Arrival Process of

Second Order 的缩写，即 2 阶 MAP)的目的，然后借鉴 KPC 组合技术生成高阶 MAP，使得生成的 MAP 更好地匹配实际负载。

**3. 重载服务的准入控制规则**

当服务集群长期处于重载状态时，就要使用准入控制规则。学术界已有的研究大多利用简单、固定的策略：为请求到达速率、队列长度、服务器负载设定上限，一旦超过设定的阈值，则系统拒绝为新来的请求服务或拒绝建立新的连接/会话，并且现有的准入控制策略均未考虑负载的时间相关性这一因素。本书的关注点是研究针对时间长相关性 Internet 负载，何时开始丢弃以及如何丢弃才能最大化服务质量的同时尽可能减少请求丢弃的比例，采用了如下策略：在请求到达时检查队列长度，如果队列长度大于阈值，则依据到达过程或服务过程的自相关系数判断准入或者丢弃，称之为基于自相关性的准入控制策略。之所以采用自相关系数，是因为它是最常见的用来描述负载时间相关性的指标。

**4. 轻载服务的能耗感知决策**

当服务集群处于轻载时，需要做出如下决策：在当前的集群内的活跃节点能够保证服务质量的前提下何时可以将某些空闲节点的状态进行切换。对于当前的服务集群，就转化为一个局部优化问题，求解得到能够保证服务质量的活跃节点序列，将其余节点进行状态切换。本书引入了通信领域的爱尔朗公式，结合上述两个负载模型与性能模型，采用递推公式以迭代方式完成对不同状态服务的资源需求定量分析，最终求解得出该服务在满足服务质量前提下所需要的服务节点的最小序列。

**5. 兼顾性能与能耗的虑及服务迁移的全局资源调度算法**

已有的资源调度算法的整体目标是满足各个服务 SLA 需求的同时最大化全局效用，效用目标的定义各不相同，都有各自的考量，但基本上都关注解决共享式云计算平台资源闲置浪费的问题。针对传统全局资源布局中在资源动态调度时仅考虑性能指标并存在节点迁移开销过大的问题，加入性能与能耗的两维效用，采用改进的遗传算法确定节点在服务间的分配，最优解以满足重载服务的服务质量(QoS)为主要目标，通过提高能力高节点的利用率以实现全局能耗的降低。同时对于重载服务在扩容节点时加入迁移节点的时间开销惩罚，从而有效避免了节点在服务间的迁移抖动，在权衡性能与能耗的基础上达到绿色云计算中心的目标。

本书总结与归纳该领域已有的工作，首先完成服务模型构建、服务能耗感知决策以及全局资源调度算法等相关理论的学习和分析；然后对相关的关键问题逐一进行探讨，力求在模型与方法上有所创新，并通过实验验证其有效性；最后提出 EADC(Energy-Aware Data Center)系统框架，从系统的角度综合考虑各个环节，探讨整体的解决思路，避免对问题的孤立研究，并进一步对研究成果进行验证，EADC 框架包括服务(VEMS)与系统(DCMS)两层的能耗模块。

(1) 在服务集群内，实验数据表明构建的负载与性能模型的计算输出与观察数据基本匹配；以负载与性能模型为基础，VEMS 通过节点状态的切换决策，在服务质量保证的同时降低了服务能耗的 33%。

(2) 在数据中心级别，全局资源布局决策器 DCMS 通过资源的优化布局算法实现了请求突发服务的服务质量满足，同时降低了全局能耗的 24%；规模性测试表明算法具有良好的可扩展性，并没有因为云计算中心规模的扩大而降低算法的有效性。

本书为云计算平台的诸多关键问题的求解提供了一套系统的模型、方法与技术，推进了云计算平台服务能耗管理策略的进一步研究和发展，为解决云计算环境下的服务高能耗问题提供了系统的理论支撑。本书的研究成果对于云计算环境下实现精确、高效的服务能耗感知，构建高能低耗平台，从而为其上各类服务提供丰富强大的信息服务具有重要的应用价值，有助于为政府、高校、企业及个人用户提供高能低耗的可扩展的海量数据服务。

本书的出版得到国家自然科学基金面上项目(NO. 61472172)、山东省重点研发计划(NO. 2015GGX101014)、烟台市科技发展计划(NO. 2015ZH060)和鲁东大学博士基金(NO. 28090301)的资助。另外，本书的编写还得到鲁东大学王刚教授与岳峻教授、中国农业大学李振波副教授的大力支持，我的师姐王秀文副研究员对于本书的部分内容亦有一定的贡献。在此对这些教授的鼓励和帮助表示衷心的感谢。

特别感谢清华大学出版社，感谢责任编辑及其他参与此书编辑工作的各位老师为本书顺利出版而付出的辛勤劳动。

由于作者水平有限，书中不妥之处在所难免，诚恳希望同行和读者批评指正，以便以后改正和完善。

曲海平

2017 年 10 月于山东烟台

# 目 录

# 第1章 绪论

## 1.1 研究意义

**1. 能耗问题已成为制约当今云计算中心发展的关键瓶颈**

随着网络的高速发展，日益增长的网络需求通过大规模的数据中心处理、数千台服务器和网络系统存储，许多互联网公司如谷歌、亚马逊、脸书等在世界各地经营着庞大的数据处理中心，这种网络服务被称为云计算服务。云计算是一个高度可扩展和注重成本效益的基础设施，运行 HPC、企业和 Web 应用程序，然而云基础设施不断增长的需求大幅增加了云计算中心的能量消耗，这已成为一个关键的问题。高能耗不仅增加了运营成本，降低了云供应商的利润率，而且造成碳排放量增高、环境污染严重的现象。

工业和信息化部电信研究院的李洁表示，2011 年我国云计算中心总耗电量达 700 亿千瓦时，已经占到全社会用电量的 1.5%，相当于 2011 年天津市全年的总用电量，而随着云计算的快速增长，未来 5 年我国对云计算中心流量处理能力需求将增长 7～10 倍，机房面积需要翻一番才能满足云计算发展的需求。美国环境署报告指出，2011 年美国云计算中心全年耗电达 1000 亿千瓦时，占该国总量的 1.5%。而据英国布里斯托尔大学 2012 年发表的研究报告指出，现在每人每天云计算需求 3200MB，每兆字节耗电 4 瓦时，全球云计中心算每天耗电 11.75 亿千瓦时，相当于 100 个北京市的耗电量。2016 年 IDC 行业 OpEx 成本分析报告指出，当前云计算中心的能耗费用占整个经营成本的比例已接近 50%，提高能耗效率(PUE)已经成为当前云计算中心管理者的普遍共识[1]。我国政府在 2015 年 9 月的联大会议上明确提出：到 2020 年要实现 40%～50%节能减排的宏伟目标，因此构建能耗感知的云计算中心刻不容缓。

**2. 云计算环境下进行能耗感知研究的必要性和重要性**

云计算是在网格计算基础上提出的一种新型计算模型，具有安全可靠的数据存储、方

便快捷的互联网服务和强大的计算能力。云计算的超大规模、虚拟化、高扩展性等特点，可以为企业和用户提供各种低成本的计算资源和IT服务，从而为实现高效信息化管理和海量数据挖掘提供强有力的计算支撑。但是管理这些云程序需要消耗大量电能并带来很高的操作开销，同时也会对环境造成负面影响。能耗的持续升高会增加云计算基础设施的总拥有成本（Total Cost of Ownership，TCO），降低投资回报率（Return On Investment，ROI）。所以，云计算环境下进行能耗研究势在必行，迫切需要构建高能低耗的绿色云计算平台，以更高的性能/瓦比率提升云计算平台的竞争力。绿色云计算中心的目标是在高效处理和使用云基础设施的同时最小化能量消耗。但是当前的云计算基础设施很少提供支持能量感知的服务，无法在满足服务质量需求的同时最小化能耗开销，以获取最大化投资回报。在新的云计算环境下，需要新的理论方法以充分利用云计算平台其上各个服务的特性来完成准确、快捷的服务能耗感知，进而有效地降低平台的系统能耗。本书针对当前云计算平台存在的高性能需求与低能耗瓶颈这一对矛盾的问题，依托云计算技术，结合服务模型构建，开展云计算环境下的服务能耗感知策略与全局资源优化布局研究。本书的研究成果对于云计算环境下实现精确、高效的服务能耗感知，从而构建高能低耗平台为其上各类服务提供丰富强大的信息服务具有重要的理论意义和应用价值。

**3. 本书为解决云计算环境下的服务高能耗问题提供系统的理论支撑**

云计算环境下，各个应用相互独立并需要不同的SLA（Service Level Agreement）性能保证。随着运行于其上的服务越来越多，共享性云计算平台管理方面的研究也越来越受到重视。本书从应用需求出发，对Web、E-mail等为代表的Internet数据服务进行分析、归纳和总结，从Internet服务的大范围变动和时间长相关性等特征出发，提炼出此类系统需要解决的服务性能分析、服务能耗感知、服务资源调度和高可靠的系统平台构建等共性科学问题。通过系统、深入的研究，为云计算环境下各个服务性能保证与能耗降低这一对矛盾问题的解决提供新的理论支撑。可扩展性是共享云计算平台的主要考量点，而能耗瓶颈则是制约云计算环境下资源扩展的核心问题。从全面认识服务本质和服务模型特性的角度出发，构建各个服务的负载与性能模型，进而做出能耗感知方案决策，在充分调动服务器集群的计算能力的同时完成服务能耗的有效降低，并在此基础上通过服务器集群间资源的有效调度与布局实现整个高可扩展平台的所有服务的性能保证与能耗降

低，从而丰富和发展云计算环境下服务能耗的核心理论与技术。构建有效的服务负载与性能模型，并实现各个服务对全局资源的共享与调度，对整个系统的设计、管理和运行是至关重要的。本书提出基于大范围变动和时间长相关性负载特征的服务负载与性能模型的构建，满足在线动态负载可扩展的能耗感知方案决策和分布协同的自适应全局资源调度策略，为更好地动态使用全局资源，优化系统性能，降低服务与平台两级的能耗提供新的理论支撑。

**4. 能耗感知的云计算平台的研究与实现的应用价值**

本书以服务负载与性能模型构建为基础，发展服务集群与系统平台两级能耗感知的云计算系统框架，并用各个应用验证研究成果。针对服务的不同需求，通过大数据集成应用平台，提供所需的虚拟计算与存储资源，构建动态、透明、安全可靠的服务集群；构建服务负载与性能模型，并进而指导能耗感知决策实现在线服务集群内部的高能低耗；然后利用分层的思想，从服务节点、服务集群和系统平台三个层面监控系统动态行为与全局状态，进行全局资源的有效调度，最终建立能耗感知的全局资源布局决策系统平台。本书将基于该系统平台的诸多关键问题的求解提供一套系统的理论、方法与技术，推进云计算平台服务能耗有效性的进一步研究和发展，有助于为政府、高校、企业及个人用户提供高能低耗的可扩展的海量数据服务。

## 1.2 研究动机

过去，各种异构服务是由不同公司或者不同平台提供的，由于服务器请求到达的不同分布和随时间变化的不同资源需求，特别是服务请求突发时的资源激增的需求，以及对 QoS(Quality of Service)保证的需求，这样就必须为每个服务预留足够的资源，从而使得数据中心常常出现资源未被充分利用的情况。图 1.1 是 HP 公司给出的企业数据中心 CPU 的平均利用率情况[2]，从该图不难看出，存在 60%以上的服务器平均 CPU 利用率低于 50%，资源被严重浪费。

针对这种情况，共享式云计算平台应运而生，该平台提供了如下核心特性[2]：统一控制、自由的物理配置、资源共享和资源隔离等。这种资源共享与负载整合的架构既使得系

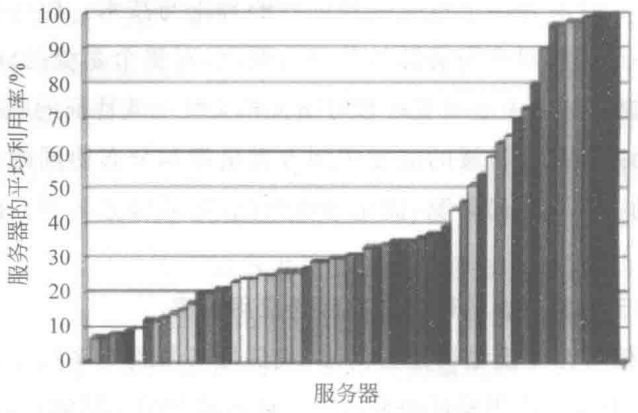


图 1.1 企业数据中心 CPU 的平均利用率

统规模扩大带来能耗需求激增，也提供了降低能耗的契机，因为并发异构服务请求到达的波动性和对于服务器资源需求的差异性为提升资源利用率和节省能耗提供了可能。

共享式云计算平台在满足各个应用对资源动态需求的同时也提高了资源的利用率，缓解了资源浪费的问题。随着云计算技术的快速发展，云计算系统规模快速扩大，系统部署需要投入越来越多的硬件，这无疑增加了能源消耗和环境负荷。如今，云计算的巨大能耗已成为制约其发展的关键问题，而大量的计算资源和存储资源都集中在云端，又给云计算中心的能耗管理带来巨大的挑战，以 Google 公司为例，在过去的十年时间里，其数据中心耗电量增加了 20 多倍，特别是随着云计算的到来，更多的资源集中到云端，给能耗的高效管理带来更大挑战[3]。所以，能耗已经成为影响云计算技术大面积推广的主要障碍和难题[4]。

随着当前云平台规模的激增，能耗已经成为制约其发展的瓶颈，如何能够在满足服务质量的同时实现更低的能耗需求以及更高的性能/瓦比率，构建绿色云计算平台，引起大家的广泛关注。所以，本书希望能够对这种共享式云计算平台进行能耗有效性的研究，其意义在于：①合理利用资源，降低能耗在 TCO 中的比重；②提高性能/瓦比率，实现系统能耗效率的最大化。

本书的研究目标就是在云计算平台实现这样一个满足能耗感知、自主调度的异构服务集群管理系统框架(Energy-Aware Data Center，EADC)，并在该框架下深入研究其中

的一些关键技术。本书通过对共享式云计算平台的服务模型构建、能耗感知方案决策及资源优化布局等基本问题的研究，构建服务的负载与性能模型，并以此为依据提供服务与平台两级的能耗感知的资源优化布局的决策方案与调度算法，从而形成一套完整的云计算环境下的能耗感知的模型和方法体系。该系统管理的云计算平台能够基于构建的服务能耗模型自适应地按照服务的请求负载变化进行服务节点状态的按需切换，在满足服务质量的前提下有效降低服务集群的能耗；同时面对请求突发的服务，通过资源优化布局算法的求解合理调度节点资源在服务集群间的分配，最终在云平台层面实现其上所有服务质量满足的同时显著降低云平台的整体能耗。

## 1.3　研究内容

在进行能耗研究时需要考虑服务的性能，因为服务性能的保证与系统能耗的降低是资源分配的两个相反方向。

(1) 节点/部件全时工作，获得的性能最佳，同时能耗也最高。

(2) 节点/部件降速或停止工作时，能耗降低，这以服务质量的降低为代价。

众所周知，造成云计算中心能耗效率低下的一个重要原因就是空闲服务器消耗的功耗没有随着其状态的空闲而线性降低，即使在诸如 10% CPU 利用率的情况下，服务器也要消耗超过 50%的功耗[5]。由此实现服务器状态的按需切换，进而缩减在线服务器规模是节省能耗的有效方式。最后的资源分配方案就是在保证服务质量的前提下实现能耗尽可能的降低，这实际上就是在性能保证与能耗降低之间实现一个优化权衡，这种权衡实际上提出了进行能耗研究的两个前提。

(1) 服务负载存在长时间的轻载：节点全部重载时进行节能没有效果，也没有意义，并且状态切换的能耗开销足以抵消短时间轻载的低能耗。

(2) 节点或者部件可以进行状态的切换：不同的状态提供不同的性能同时消耗不同的能耗，状态必须能够按需进行切换以达到能耗降低与性能保证的均衡。

当前共享式云计算中心已经成为 ISP(Internet Service Provider)的通用模式，其上运行多个独立、竞争的 Internet 服务。这里的所谓 Internet 服务，没有一个通用的标准定

义，在文献[6]中定义为泛指通过 Internet 给用户提供的服务，包括诸如 WWW、E-mail、FTP 等服务。Internet 服务请求呈现面向非连接特性与负载动态变化这两个特征。Internet 服务的非连接特性决定了服务请求可中断、可重连、可调度，由此可以由负载均衡器实现服务请求在节点间的合理分配。而负载的动态变化是指 Internet 服务存在服务请求突发，即大范围变动，其高峰流量是平均流量的 8～10 倍，如图 1.2 所示；同时还表现出统计的时间长相关性，即在不同时间尺度上都表现出突发性，如图 1.3 所示[7]。负载动态变化意味着系统需要能够自主、及时、准确地预测这种突发性，并通过调度更多的资源来平滑这样请求的突发。这些特性正好满足了本书的研究前提，所以选择 Internet 服务作为本书的研究对象。

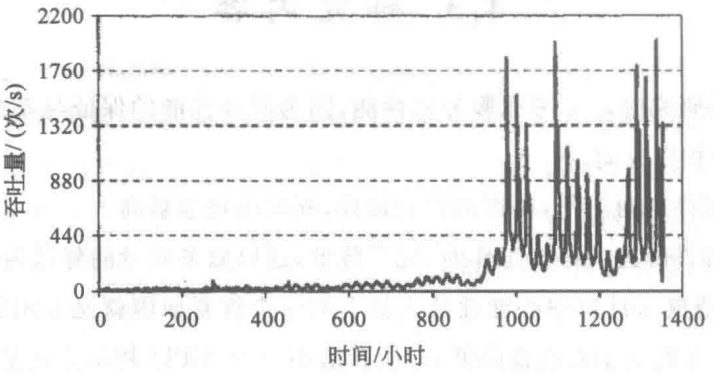


图 1.2　世界杯网站服务请求速率

面向 Internet 服务，本书通过服务集群的模型构建、服务节点状态的按需切换，以及服务节点在集群间动态调度的紧密结合提出了一个包括服务集群内部和全局服务系统的两层能耗管理框架。在该框架下主要关注两部分内容。

(1) 实现服务集群的高能低耗。这首先需要针对 Internet 服务的两个负载特性“大范围变动”与“时间长相关”进行模型的构建，在负载与性能模型的基础上做出服务感知决策，以保证服务质量为前提，以尽可能降低能耗为最终目标。主要的工作归结起来就是两个：完成服务请求与服务节点的映射以响应服务请求从而实现负载的有效聚集；根据负载变化进行服务节点状态的按需切换以完成节能的目的。

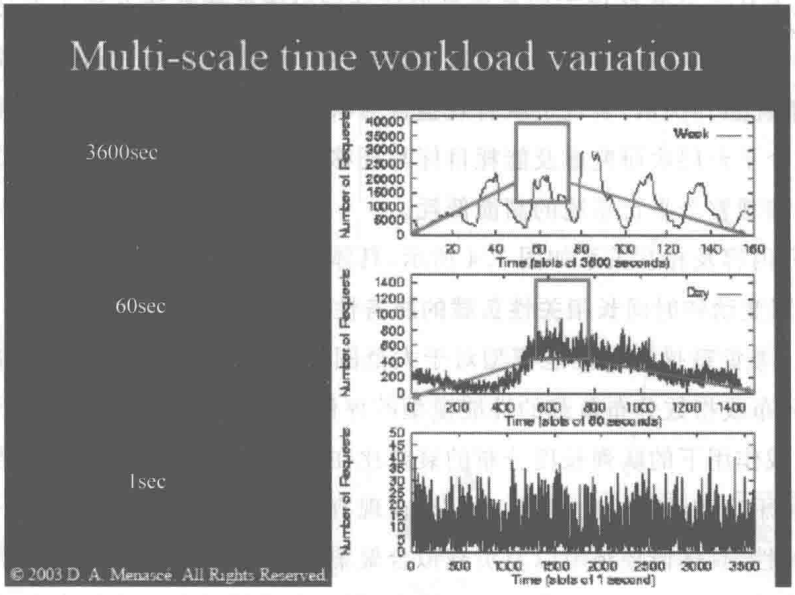


图 1.3　Internet 服务负载特征

(2) 云平台的各个服务集群为降低能耗而呈现负载聚集的趋势。当请求突发时服务性能会急剧下降,需要进行全局资源动态调度以保证该服务的服务质量。本书在各个服务集群自主运行的基础上进行全局资源的优化调度,以实现云计算中心所有服务集群的高性能与低能耗。主要的工作归纳起来有两个:对当前资源的配置进行量化以确定是否能够满足其上所有服务的性能需要;根据各个服务的状态效用,通过节点在服务集群间的有效动态调度实现全局资源的最优布局。

通过以上相关的介绍,可以得出如下结论。

(1) 云计算平台需要进行能耗有效性的研究,但是必须以服务性能优化与服务质量满足为前提。这就需要深入研究服务能耗与性能的关系,寻求性能保证与能耗节省两者的和谐统一。

(2) 云计算中心能耗的有效性是以其上服务的能耗有效性为基础的。所以,就需要构建服务与系统的两层能耗模型,同时深入研究两者的交互关系,实现从局部到全局的能耗有效。