

# 面向大数据的绿色 IT 框架能效分类机制

王丽芳<sup>1</sup>, 齐勇<sup>2</sup>, 蒋泽军<sup>1</sup>, 彭成章<sup>1</sup>

(1.西北工业大学计算机学院, 西安 710072;

2.陕西科技大学电信学院, 西安 710129)

**摘要:** 与分散处理相比较, 大数据中心集中处理信息通信任务, 在能效上已有巨大的提高。但大数据中心包括数以千万计的服务器, 其能源消耗量甚至可以超过一座小型城镇。巨大的能源消耗、成吨的温室气体排放, 使大数据中心在能效与减排方面面临诸多挑战, 建立绿色高效的大数据中心势在必行。本文给出一个面向大数据的绿色 IT 框架, 重点研究了能效分类问题, 提出了一个基于度量模型的能效分类机制。根据工作量和能源消耗情况对设备和服务进行分类, 无缝地划分为不同的资源池, 将电力使用效率、数据中心工作效率和二氧化碳排放等综合计算衡量, 制定可以实现并遵循的能效标准, 使绘制大数据中心的碳足迹成为可能, 并提供服务能效评估方案。

**关键词:** 大数据; 绿色 IT; 能效分类; 能效度量模型; 能效评估

中图分类号: TP311.133.1 文献标识码: A 文章编号: 1674-4969(2014)03-0224-09

## 引言

近年来, 各类全球环境指标不断警示温室效应的危害, 而各大计算机软硬件企业与研究机构也都争相推出利于可持续发展的高能效节能技术、产品和服务, 绿色计算的倡议逐渐深入人心<sup>[1]</sup>。人们对气候变化和环境可持续发展的关注度空前重视, IT 企业从客户、股东、用户方面承受的压力越来越大, 国家政策对于能效优化的要求不断提高, 迫使绿色计算技术不断前行<sup>[2]</sup>。

信息与通信技术的进步在促进社会发展的同时, 也使政府、企业、个人的日常工作必须建立在大数据中心的 7 天 × 24 小时工作之上。目前, 大数据中心的骨干服务包括: web 托管、电子商务、社交网络, 以及更底层的服务, 如软件即服务 (SaaS)、平台即服务 (PaaS)、网络即服务 (IaaS)<sup>[3]</sup>。

## 1 IT 对环境的冲击

随着 IT 技术的蓬勃发展, IT 技术为各个领域的业务活动提供了极大的效率提升和方便性, 更带来了不可逆转的转型业务, 使世界进入了全球化的时代。但同时, 它已经造成了严重的环境问题。不幸的是, 大多数人, 包括许多 IT 专业技术人员却没有清晰地意识到这一点。IT 技术正在以多种不同的方式影响着我们的环境。

一台个人计算机从生产到报废的每个环节 (制造、使用、废弃) 都对环境提出了严峻的挑战。制造计算机以及相关的各种电子和非电子元器件需要消耗电力、原料、化学物质和水, 并且产生有毒的废气、废水, 所有这些因素都对环境问题造成严重影响。

在当今的大数据时代, 全球范围内, 数据中

收稿日期: 2013-10-10; 修回日期: 2013-11-25

基金项目: 国家自然科学基金资助项目(61373120), 航空科学基金资助项目(2012ZC53040)

作者简介: 王丽芳 (1964-), 女, 教授, 研究方向为电子商务安全、分布式网络存储、云存储等。E-mail: wanglf@nwpu.edu.cn

齐勇 (1982-), 男, 讲师, 研究方向为云计算、云存储、机器学习等。E-mail: qiyong@sust.edu.cn

心、服务器和计算机的总量和用电量大幅增加。而目前,电力主要来源于化石燃烧,国内更是主要使用基于煤炭的火力发电技术,因此用电量的提升就意味着大量二氧化碳的排放。同时,数据中心所使用的服务器、存储器和网络设备的使用寿命都只有数年时间,所产生的重金属污染难以消除,多数作为垃圾进行填埋。IT技术的广泛应用必然造成环境污染情况的进一步加剧。因此,有效降低IT污染,充分发挥大数据中心的工作效能和能效利用率,降低个人计算机的使用量和更新速度,已经成为各个领域专家的研究重点之一。

## 2 大数据中心的能效问题

大数据中心是指包含大量信息和通信技术的基础设施平台,由供电设备、冷却设备、网络设备、计算服务器和存储服务器组成。随着大数据技术的逐渐成熟,各类大数据应用应运而生,政府、企业、媒体、科研机构的信息管理、海量计算越来越多地转移到大数据中心。全球变暖和气候变化促使能源成本上升、二氧化碳排放量受到控制,要求大数据中心在能效方面不断优化。面对全球数字经济与可持续发展两方面的严峻挑战,必须通过资源的动态整合,凝聚其有限的可用性,探索近乎无限的可能性,建立新型的绿色大数据中心。

### 2.1 能源消耗现状

目前,大数据中心数以千万计的服务器消耗的能源甚至超过了小型城镇。为了满足日益增加的客户需求,微软公司每个月需要为他们的大数据中心新增两万台服务器<sup>[4]</sup>。这些服务器消耗了大量能源和电力,并以二氧化碳的形式影响全球气候。与此同时,这些服务器在日常工作中约有30%是不承担任何任务的,也就是说闲置的服务器只消耗能源,不产生价值。大数据中心的能源利用率普遍只有5%~10%<sup>[5]</sup><sup>[6]</sup>。因此,能够有效提高设备利用率和能效的软硬件框架成为业界研究的新热点<sup>[5]</sup><sup>[6]</sup>。

随着大数据中心设备规模的不断扩展,大数据中心正呈现出更大、更密、更昂贵的特点。而在能源消耗方面,以美国为例,2010年全美大数据中心消耗的电力占美国电力总消耗量的1.5%,约合4.5亿美元<sup>[7]</sup>;同时,大数据中心的电力消耗呈现出急速增长的态势,2011年,数据中心的耗电量超过100亿度,约合74亿美元<sup>[8]</sup>;随着大数据中心的不断扩张,大数据中心不仅消耗大量能源,而且中心将充斥着闲置的耗电服务器,如果数据中心的设计人员仍不考虑能效问题,他们将在未来五年内面对可能会增长16倍的能源成本<sup>[9]</sup>。美国政府针对能效提出了最新优化目标:到2030年,所有的新建建筑物均需达到“碳中和”标准,即碳排放量近似为零,全美将能效提高到50%(目前的能效值约为25%)<sup>[10]</sup>。欧盟制定了更加严格的标准,将在2020年实施,这将直接影响大数据中心的建设与运营,没有能效优化的大数据中心将难以生存。

全社会正在关注信息领域的能效问题,这是因为截至2007年,全球信息通信行业的能源消耗已经与航空业持平,且正以每5年翻一番的速度增加<sup>[11]</sup>。图1为2010年美国信息通信设备排放二氧化碳的分布情况<sup>[12]</sup>。而20世纪90年代制定的能源之星标准主要是针对个人计算机的用电量。如今大数据中心的服务器和通信设备所消耗的能

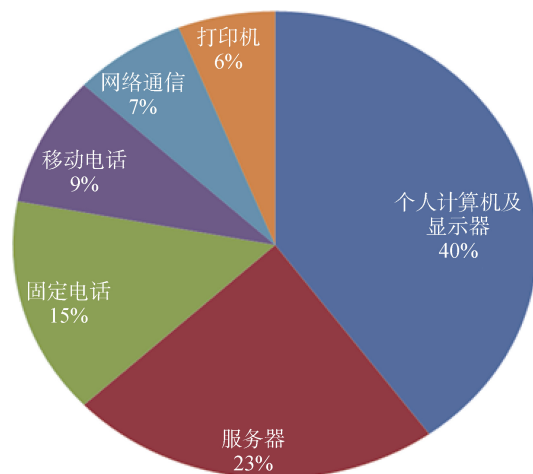


图1 美国信息通信设备的二氧化碳排放比例（2010年）  
数据来源：文献<sup>[12]</sup>。

源成为主要的增长点, 绿色计算在大数据时代需要新的绿色 IT 框架。

2.2 策略与趋势

能源价格居高不下, 已经开始影响许多大数据中心运营商的商业模式。Google 公司刚刚并购了一批风力发电企业, 为其大数据中心提供绿色能源支持; IBM 公司正将越来越多的大数据中心建立在电力价格较低的地区。能源安全与可用性已经成为大数据中心运营商的主要压力, 石化燃料的政策局限性使能源供应前景阴晴不定。在技术需求与可持续发展的两面夹击中, 大数据中心需要探索出一条新的应变之路, 而提高能效、挖掘设备使用潜力将成为主要的手段。大数据中心能效的提高涉及诸多领域, 包括绿色能源技术的发展、传统能源技术能效的提高等。以目前联合国公布的计算标准 计算二氧化碳排放量为每千瓦时电能产生 0.785 kg 二氧化碳<sup>[12]</sup>, 而现在的大数据中心一年消耗的电能约为 70 亿 kW · h<sup>[13]</sup>。目前温室气体的排放主要来自航空业、航运、交通、电信和制造业的快速发展, 而 IT 领域的排放总量增加的速度更快。这些排放量不断增加, 使全球信息通信设备的二氧化碳总排放量有趋势从 2007 年的 3% 增加到 2020 年的 6%, 其中大数据中心即使全面采用绿色新能源, 也将呈不断增长态势<sup>[14]</sup> (表 1)。

表 1 二氧化碳排放量的现状与趋势估计

类别	2007 年二氧化碳排放		2020 年的预估 二氧化碳排放	
	排放量/kg	比例/%	排放量/kg	比例/%
全球信息通信设备总和	830	100	1430	100
大数据中心	116	14	257	17
电信设备	307	37	358	25
个人计算机	407	49	815	58

数据来源: 文献[14]。

随着企业越来越依赖 IT 服务, 对服务的可用性和连续性的要求不断提高。确保安全的能源供应、环境保护和气候保护是当今世界面临的主要挑战。环保技术是可持续的经济活动的关键, 而目前一台不间断工作的服务器 3 年内使用的能源成本和环境成本已经超过了其购买成本<sup>[15]</sup>。

3 绿色 IT 框架

3.1 绿色 IT

绿色计算主要是指 IT 设备中的电源生产与使用, 而绿色 IT 是一个更加全面的概念, 是将虚拟化、云计算、IT 设备采购、回收、电源管理等作为关注中心, 进行全局能效优化的新目标。绿色 IT 的实现分为两部分:

1) 构建新的 IT 系统, 配置 IT 设备方面严格执行能效标准;

2) 改进现有的 IT 系统, 进行优化整合, 挖掘现有设备的使用潜力, 提高总体利用率。

这两部分可以概括为: 同样的工作使用更少的能源, 同样的能源完成更多的工作。

要在实施大数据中心使用节能技术和绿色指标必须克服上述涉及的诸多困难, 制定完整的绿色 IT 框架, 并进一步制定实施针对大数据中心的绿色 IT 解决方案。

绿色 IT 涉及许多重点领域和活动, 包括: 电源管理, 数据中心的设计、布局、选址; 生物可降解材料的使用; 能效指标和绿色标签的制定; 碳足迹评估工具方法及环境相关的风险缓解辅助决策系统。绿色 IT 将是近几年的热门话题<sup>[16]</sup>。越来越多的 IT 厂商和用户已经开始把注意力转向绿色 IT, 促使绿色税收、国家法规的制定和出台, 绿色 IT 产品和解决方案的需求将会大幅增加。

绿色 IT 的主要作用有:

见 <http://www.eia.gov/tools/faqs/faq.cfm?id=74&t=11>, 该表内的数据每月更新, 因此计算结果存在误差。本文根据各项能源的发电比例, 计算表内第三列 “Lbs CO<sub>2</sub> per kWh” 的值, 平均为 1.73 磅 (约为 0.785 kg)。

- 1) 提高能效,降低成本;
- 2) 改善公众因环境因素对大数据中心的顾虑;
- 3) 使大数据中心在未来可以符合环保政策及法规。

绿色 IT 框架是针对大数据中心设计的能效优化总体策略,其主要作用总结为如下十点:

- 1) 降低大数据中心的整体能源成本;
- 2) 延长或扩展现有的数据中心设备的寿命,提高智能温控设备的能效指标;
- 3) 降低大数据中心的设备维护频率及成本;
- 4) 减少有害废物的使用和排放;
- 5) 减少整体碳足迹;
- 6) 减少二氧化碳的排放量;
- 7) 减少可能带来的酸雨、雾霾等灾害天气发生的频率;
- 8) 降低对电网负载的影响;
- 9) 降低大数据中心的占地面积;
- 10) 符合未来的能源、环境法规。

### 3.2 绿色 IT 框架

绿色 IT 框架不仅为实现绿色 IT 提供具体的指导意见,也是一种绿色 IT 标准的评估工具,为绿色 IT 从概念转变为行动提供战略参考。

绿色 IT 框架一般遵循五个步骤(图 2):



图 2 实现绿色 IT 框架的五个步骤

- 1) 分析——从能效角度衡量大数据中心的设计方案;
- 2) 归类——将大数据中心的基础设备和软件服务根据能耗进行分类;
- 3) 实施——对基础设备和服务进行优化组合;
- 4) 回收——控制设备报废流程,降低碳足迹;

- 5) 维护——实时监控设备及服务状态,减少损耗,提高利用率。

## 4 能效评估机制

大数据中心有许多不同的组件和设备执行不同的任务,以满足最终用户的需求。从能效角度而言,应该将这些组件划分到可测量的资源池中,根据他们执行的工作量和绿色指标来衡量其性能和效率,然后汇总得到数据中心的整体能效。一些主要的数据中心组件有:服务器(主要指计算服务器)、存储设备、不间断供电电源(uninterruptible power supply, UPS)、冷却系统、外部机架配电装置、配电装置(power distribution unit, PDU)、照明设备等。

本文提出了一个面向大数据中心的基于度量模型能效分类机制,根据工作量和能源消耗情况对设备和服务进行分类,定义绿色能效指标以衡量大数据中的能源效率及二氧化碳排放量,为构建或改进大数据中心提供准确的定量依据和指导。

### 4.1 能效

节能是指使用较少的能量,而能效是指在使用较少的能源的前提下执行相同的功能。例如,LED 灯泡使用较少的能量,却产生了与白炽灯泡同样的光量。能源可持续发展的重点在于制定长期的能源战略和政策,确保任何时期都具备足够的能源,在满足今天需求的同时满足明天的期望。针对大数据中心能效优化,需要区分“功耗”和“能耗”两个概念。

功耗:单位时间内的系统峰值功率。

能耗:一段时间内的系统电能消耗总量。

之所以对功耗和能耗进行区分,是因为通常情况下超过 50%的能耗用于支持 UPS,真正用于计算的功耗小于 50%。目前的趋势是,不断增长的能耗及与其相关联的用电成本在未来几年内将超过 IT 设备本身的采购费用。因此,设计和开发高效节能的大数据中心资源管理策略势在必行。

主流的能效提高策略有以下四种：

1) 提高数据中心的能源使用策略之一是利用可再生能源。例如, 在具备丰富的可再生能源(太阳能、风能等)的地区建立大数据中心。

2) 避免重复的交流电(AC)和直流电(DC)切换。其基本思想是在大数据中心只做一次交直流电转换, 而不是在不同的服务器之间均做交直流电转换。目前交直流电一次切换的大数据中心解决方案已经完成<sup>[17]</sup>。

3) 考虑散热与热能再利用。集中使用大功率设备导致产生更多的热量。为了解决散热过程中的能效问题, 很多冷却策略正在研究和开发之中, 例如, Google 公司将其大数据中心建立在海上。Google 公司的海上数据中心将依靠海水和潮汐发电, 同时利用海水的流动, 对数据中心的机器进行冷却。电源和冷却效率的改进将继续。

4) 采用新型能效计算部件与存储设备。目前已经出现非对称多核心微处理器设计的硬件解决方案, 并尝试以高效的封装技术和能量成正比的硬件设计新型内存和磁盘系统, 纳米芯片的研制成功也为计算部件的能效提高带来了曙光<sup>[18]</sup>。

## 4.2 度量

为了对大数据中的基础设备和软件服务功耗情况设定准确的基准, 需要确定度量单位。由于大数据中心不同设备及服务的工作负载状态变化频繁, 度量单位粒度需足够小; 同时由于所需被观测值来自不同的系统(计算系统、网络系统、存储系统、软件服务系统等), 并且严重依赖各个系统的运行状态(负载情况、可靠性和安全性标准等), 所以该度量单位必须能够将不同的观测值融合为统一的标准量, 同时需要考虑系统和部件的运行状态。只有将不同的系统、不同的工作状态全盘综合考虑的情况下, 才可以认为该度量单位及其度量结果是真实体现系统能效的。为了实现度量单位之间的运算, 必须对数据进行分类, 继而转化为可测量的单位。

目前还没有适当的单一机制能够对能效指标进行衡量。这一过程目前只能由大数据中心的运行维护人员汇总后进行评估。本文提出了一个新的能效度量模型(图3), 是针对大数据中心的设备部署情况及服务性能综合评估过程设计的。

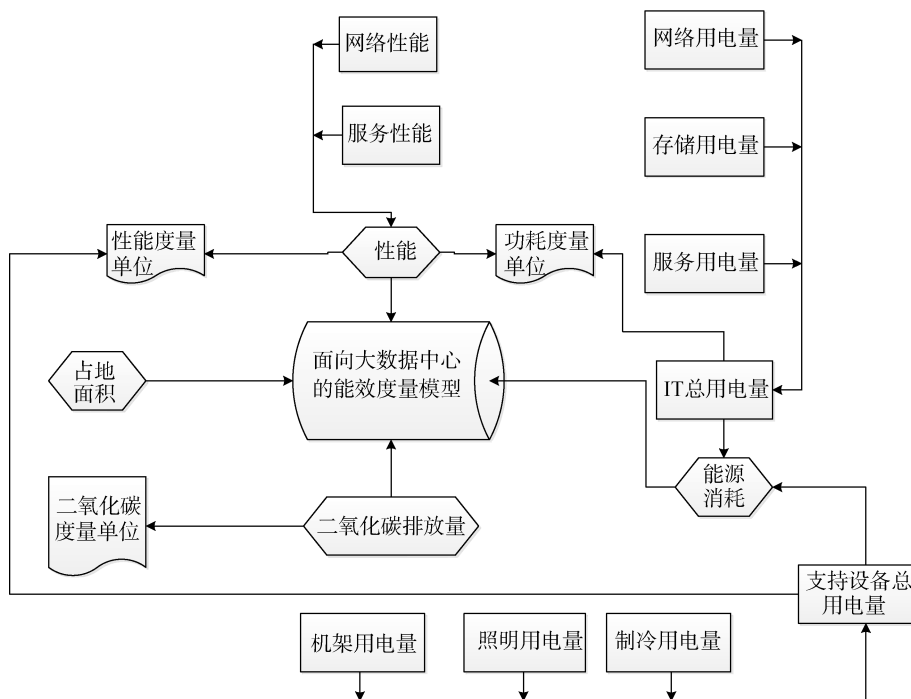


图3 面向大数据中的能效度量模型生成图

图3所示的能效度量模型的意义有三点:

1) 为建立节能数据中心提供了依据和标准,使大数据中心的能源使用更加透明,通过IT设备的能效优化,实现系统的最佳实践;

2) 建立模型后,可以通过知识获取和知识管理的方式逐步消除现有数据中心的能效瓶颈;

3) 为大数据中心更好地量化、理解能效回报提供数据支持。

### 4.3 设备能效

在建立度量单位和度量方法后,根据能效度量模型对大数据中心的所有设备及服务进行分类。总体上将硬件设备能耗分为两类:

1) 支持设备能耗;

2) IT设备能耗。

支持设备能耗又可以分为三类:机架用电、照明用电、制冷用电。

其中,机架部分包括:

1) 不间断电源;

2) 配电单元;

3) 电池;

4) 外部的IT设备的电力损失。

制冷部分包括:

1) 服务器内部冷却器;

2) 空调机组;

3) 换风处理;

4) 冷却塔。

IT设备能耗包括的组件有计算设备和辅助设备。

计算设备包括:

1) 服务器计算机;

2) 客户机;

3) 网络设备;

4) 存储设备。

辅助设备包括:

1) KVM切换器;

2) 显示器;

3) 用于监视或以其他方式控制数据中心的工作站或笔记本电脑。

假设大数据中心的所有设备均由已标准化组件分层次组成,并且全年无休运行。在此基础上,可以定义一个大数据中心的组件 $i$ 在制造、运输、销毁过程中的二氧化碳总产生量为 $c_i$ ,组件 $i$ 的平均适用年限为 $m_i$ ;同时组件 $i$ 在最低开机运行过程中的年平均耗电量为 $e_i$ ,根据适用电能转换为相应的二氧化碳排放量 $CE(e_i)$ 。因此,大数据中心的设备组件在最低运行状态下的年二氧化碳排放总量可以表示为:

$$CI_{total/year} = \sum_{i=1}^N [c_i / m_i + CE(e_i)]$$

其中 $N$ 表示大数据中心总设备组件数, $CE$ 函数用于汇总计算不同电能类型所对应的二氧化碳排放量,表示每一度电所消耗的二氧化碳量。这是由于目前国内的风电、潮汐发电最多只能占输送电网内电量的20%,即风力发电站必须同时配备4倍的火力发电或水利发电,以保证电能供给的稳定性。

根据设备能效情况可知,若提供的服务低于标准下限时,大数据中心将处于低能效状态,即较之分散布置的计算设备的能效综合情况更低。作为大数据中心的能效下界,低于此标准的大数据中心将缺乏能效建设意义。

### 4.4 服务能效

在进行硬件资源的能效分类后,还需要对软件服务按照能耗进行分类,提高设备工作负载是提高能效的有效方法。

在传统的数据中心,每个服务器都被指定完成一个特定的功能。例如,电子邮件服务器只用于管理电子邮件,工资服务器只处理工资等。但是这种传统的服务分类方式是低效的,因为电子邮件服务器只能运行65%的容量。例如,在营业时间内,数据中心被访问量总和接近甚至超过高峰的时候,使用单一功能的电子邮件服务器能效

很高;但在非营业时间,需求显著下降,数据中心能效随之下降,造成整体能效处于低下水平。另一方面,如工资服务器可能在营业时间内只有5%的时间有相关人员进行查询使用,而其存储能力往往大幅超过现有数据量。因此,在大数据中心使用虚拟化技术,使电子邮件服务器和工资服务器可以共享同一台计算机硬件,在营业时间内并行完成电子邮件处理和工资单处理,而在下班后将其租借给世界另一边的企业使用。使用这种方法,政府或组织购买和维护的设备减少。他们还可以进一步节省设备购置、场地建设、电力费用等成本。因此,大数据中心优于分散管理的小型数据中心,但由于其能耗集中,所以仍需要在能效方面进一步挖掘潜力。因此,除了使用节能设备以外,应该从提高软件服务的工作负载和工作效率等方面优化能效。

工作负载可以进一步从服务资源和应用资源的角度进行分类。

1) 服务资源分类。创建服务器软件服务工作负载资源池,其前提是避免任何安全和隐私数据泄露、损坏的服务器间资源共享。从而将服务器按照以下任务类型进行分类<sup>[19]</sup>:

- (1) 网络基础设施服务器;
- (2) 身份认证服务器;
- (3) 文件和打印服务器;
- (4) 应用程序服务器;
- (5) 协作服务器;
- (6) Web 服务器;
- (7) 数据库服务器。

2) 应用资源分类。当服务按类型分配到不同的资源池时,应用也可以被分为以下类型:

- (1) 自定义应用程序;
- (2) 传统与更新的应用程序;
- (3) 基础设施应用;
- (4) 支持业务应用;
- (5) 业务流程应用程序;
- (6) 关键业务应用程序。

在大数据环境下,服务主要以任务的形式提交给硬件资源(即设备组件)完成时,需要各类设备额外增加的电能消耗值为:

$$CS_{total/year} = \sum_{j=1}^M \left[ CE \left( \sum_{i=1}^N e_{ij/job} \right) Count_j \right]$$

其中, $M$ 表示服务类型总数, $e_{ij/job}$ 表示设备组件*i*每增加运行一个任务*j*所额外需要消耗的电能, $Count_j$ 表示任务*j*每年的平均调用次数。

因此,大数据中心的年二氧化碳排放总量为:

$$C_{total/year} = CI_{total/year} + CS_{total/year}$$

## 5 模型应用情况

根据在陕西省信息中心大数据云计算服务平台进行的调查结果,大数据中心相比传统个人PC时代的计算能效有了大幅度的提升,但仍有进一步提高能效的巨大潜力,同时大数据中心必须面对单位面积高密度能源消耗情况。

个人PC的硬件设备年碳足迹为0.543 t<sup>[14]</sup>,而该数据中心的设备组件年碳足迹为108.198 t,其中包括双核CPU 220枚、磁盘存储资源4 PB、网络资源2 Gb/s,可知该大数据中心的年碳足迹约为个人PC的200倍。

个人PC的年工作状态服务碳足迹为0.0272 t<sup>[14]</sup>,而该大数据中心共运行大型服务系统139套,实现365天×24小时全年无休运行,其 $CS_{total/year}$ 约等于2.563 t,是个人PC的约113倍。

若将该数据中心的所有服务部署在个人PC上,共需要约400台PC。由此可知,大数据中心在硬件设备资源方面节约碳足迹约50%,节约服务碳足迹约75%。通过评估可知,兴建任务量充足的大数据中心将成倍地降低能源开销。

## 6 总结与展望

在最近几年,能效问题已成为现代计算系统(如大数据中心)的最重要的设计要求之一。如果不对其进行能效优化,则大数据中心在电力管理方面就会首先出现成本问题。除了能源价格上涨



造成大数据中心营运成本高企之外,随着环保意识的不断增强,公众对于二氧化碳排放对环境危害的认识逐渐加深,大数据中心作为能源集中消耗单位将承受更多的社会压力和法律约束。目前的IT基础设施产生的碳足迹约为能源消耗总量的3%左右<sup>[12]</sup>。虽然IT基础设施目前还未作为全球能源的重点监控对象,但其增长速度正逐步加快,在可预见的未来,IT基础设施必将成为能源消耗的重要部分。因此,本文提出了一个基于度量模型的能效分类机制,从能效角度对硬件资源和软件服务的能效进行分类,为大数据中心的能效评估、能效监控打下了基础,并为未来大数据中心的建设提供了定量的能效优化模型。相同的计算能力、相同的计算任务下,设备组件利用率的不同、选择的电力来源不同,将大大影响数据中心的碳足迹和能效评估结果。未来的研究工作将集中在两个方面:

1) 修正基于度量模型的能效分类机制,提高大数据中心的能效评估结果的准确性;

2) 研究不同时间段、不同地域之间形成的大数据中心网络综合能效优化方案,为大规模可信计算提供绿色能效支持。

## 参考文献

- [1] Lamb J. The Greening of IT: How Companies Can Make A Difference for the Environment [M]. IBM Press, 2009: 31-32, 287-293.
- [2] Molla A, Cooper V A, Pittayachawan S. IT and Eco-sustainability: Developing and Validating A Green IT Readiness Model [C]// Proceedings of International Conference of Information Systems. 2009: 7-13.
- [3] Loper J, Parr S. Energy Efficiency in Data Centers: A New Policy Frontier [J]. Environmental Quality Management 2007, 16 (4): 83-97.
- [4] Microsoft. How Microsoft DESIGNs the Virtualization Host and Network Infrastructure [M]// Microsoft IT Showcase: Technical Case Study, 2009:1-9.
- [5] Uddin M, Rahman A A. Server Consolidation: an Approach to Make Data Centers Energy Efficient & Green [J]. International Journal of Scientific and Engineering Research, 2010, 1(1).
- [6] Talaber R. Using Virtualization to Improve Data Center Efficiency [R/OL]. Green Grid, 2009: 5-20. <http://www.thegreengrid.org/~media/WhitePapers/White%20Paper%2019%20-%20Using%20Virtualization%20to%20Improve%20Data%20Center%20Efficiency.pdf?lang=en>.
- [7] EPA. Report to Congress on Server and Data Center Energy Efficiency: Public Law 109-431 [R]. Washington, DC: Environmental Protection Agency, 2007: 60-78, 82.
- [8] Environmental Protection Agency. ENERGY STAR Program Requirements for Computers: Preliminary Draft Version 4.0 [S/OL]. Washington, DC: Environmental Protection Agency, 2009: 2-3. [http://www.energystar.gov/ia/partners/prod\\_development/revisions/downloads/computer/PrelimSpecDraftv4.pdf](http://www.energystar.gov/ia/partners/prod_development/revisions/downloads/computer/PrelimSpecDraftv4.pdf).
- [9] Singh B, Sharma D K, Kumar R, et al. Development of a New Controlled Pesticide Delivery System Based on Neem Leaf Powder [J]. J. Hazard Mater, 2010, 177(1-3): 290-299.
- [10] Obama B. Obama's Speech on Energy [EB/OL]. (2012-03-15)[2013-07-20]. [http://www.realclearpolitics.com/articles/2012/03/15/obamas\\_speech\\_on\\_energy\\_113511.html](http://www.realclearpolitics.com/articles/2012/03/15/obamas_speech_on_energy_113511.html).
- [11] Kumar R, Mieritz L. Conceptualizing Green IT and Data Center Power and Cooling Issues: Gartner Research Paper no. G00150322 [R/OL]. 2007: 1-2. <https://www.gartner.com/doc/519717/conceptualizing-green-it-data-center>.
- [12] U.S. Energy Information Administration, Independent Statistics and Analysis. Emissions from Energy Consumption at Conventional Power Plants and Combined-Heat- and-Power Plants [EB/OL]. (2010)[2013-07-20]. [http://www.eia.gov/electricity/annual/html/epa\\_09\\_01.html](http://www.eia.gov/electricity/annual/html/epa_09_01.html).
- [13] Koomey J G. Worldwide Electricity Used in Data Centers [J]. Environmental Research Letters, 2008(3): 4-6. [http://iopscience.iop.org/1748-9326/3/3/034008/pdf/1748-9326\\_3\\_3\\_034008.pdf](http://iopscience.iop.org/1748-9326/3/3/034008/pdf/1748-9326_3_3_034008.pdf).
- [14] The Climate Group. Smart2020: Enabling the Low Carbon Economy in Information Age [R/OL]. Global e-Sustainability Initiative, 2008. <http://gesi.org/files/Reports/Smart%202020%20report%20in%20English.pdf>.
- [15] Brill K G. The Invisible Crisis in the Data Center: the Economic Meltdown of Moore's law [R/OL]. Santa Fe, NM: Uptime Institute, 2007: 2-4. [http://www.mm4m.net/library/The\\_Invisible\\_Crisis\\_in\\_the\\_Data\\_Center.pdf](http://www.mm4m.net/library/The_Invisible_Crisis_in_the_Data_Center.pdf).
- [16] Murugesan S. Going Green with IT: Your Responsibility Toward Environmental Sustainability [J]. Cutter Business-IT Strategies Executive Report, 2007, 10(8): 1-25.
- [17] McKenna P. Can We Stop the Internet from Destroying Our Planet? [J]. New Scientist, 2008(2637): 1-3.
- [18] Paleologo G, Benini L, Bogliolo A, et al. Policy Optimiza-



tion for Dynamic Power Management [C]// DAC '98 Proceedings of the 35th Annual Design Automation Conference. New York: ACM, 1998: 182-187.

[19] Uddin M, Rahman A A. Prerequisites for Implementing Energy Efficient & Cost Effective Data Centers Using Virtualization [J]. Journal of Computing, 2010, 2(11): 95-101.

## Energy-Efficiency Classifying Mechanism for Green IT Framework of Big Data

Wang Lifang<sup>1</sup>, Qi Yong<sup>2</sup>, Jiang Zejun<sup>1</sup>, Peng Chengzhang<sup>1</sup>

(1. School of Computer Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China;

2. The School of Mechanical and Electrical Engineering, Shaanxi University of Science & Technology, Xi'an 710129, China)

**Abstract:** Compared with the distributed processing, large data center focuses on information and communication tasks and has already made huge improvement on energy efficiency. But large data center includes tens of millions of servers, and its energy consumption can be ever larger than one small town. Huge energy consumption and tons of greenhouse gas emission bring many challenges to large data center in terms of energy efficiency and emission reduction. Therefore, the establishment of green and efficient large data center is imperative. This paper provides a framework of green IT for large data, focusing on energy efficiency classification and proposes an energy efficiency classification mechanism based on measurement model. According to the workload and energy consumption, this paper classifies the equipment and services seamlessly into different resource pool. Power efficiency, data center efficiency and carbon dioxide emissions are calculated and measured. This paper also formulates energy efficiency standards which could be realized and followed. The standards make the drawing of the big data center carbon trace possible and provide scheme for energy efficiency evaluation.

**Keywords:** big data; green IT; energy-efficiency classifying; energy-efficiency metric model; energy efficiency evaluation