

大数据时代数据总量增长的新摩尔定律辨析

——吉姆·格雷是否真的提出过新摩尔定律？

辛冬播, 李廷军

(江汉大学教育学院, 湖北武汉 430056)

摘要: 当前, 所谓的吉姆·格雷新摩尔定律在学界渐有流行趋势。大多数引用该定律的中文文献中未注明相关出处, 也没有结合统计数据加以验证。通过追溯溯源分析相关中英文文献表述, 逐步理清了国内有关新摩尔定律的各种来由, 并没有发现吉姆·格雷提出新摩尔定律的直接证据。结合逻辑分析, 以及全球数据量增长的相关统计, 认为所谓的吉姆·格雷新摩尔定律表述有明显漏洞, 据此推测吉姆·格雷并没有提出所谓的吉姆·格雷新摩尔定律, 应该是国内一些学者对吉姆·格雷某些观点的误读。经过研究认为, 新摩尔定律可以表述为全球数据总量每3年将翻一番。

关键词: 大数据; 数据总量; 新摩尔定律; 吉姆·格雷

DOI: 10.11907/rjdk.212268

中图分类号: TP301

文献标识码: A

开放科学(资源服务)标识码(OSID):

文章编号: 1672-7800(2022)009-0236-05



Probing into the "New Moore's Law" on Data Growth in the Age of Big Data

——Has Jim Gray Really Proposed the "New Moore's Law"?

XIN Dong-bo, LI Ting-jun

(School of Education, Jianghan University, Wuhan 430056, China)

Abstract: At present, the so-called Jim Gray's new Moore's law is becoming more and more popular in Chinese academic circles. This study found that most Chinese literature on this law failed to indicate the source of relevant references, nor did it be verified with statistical data. Secondly, this study has traced back to the source, analyzed various Chinese and English literature expressions, and gradually clarified various reasons for the new Moore's law in China, and found no direct evidence of Jim Gray's new Moore's law. Combined with logical analysis and relevant statistics on the growth of global data volume, the expression of the so-called Jim Gregory new Moore's law itself also has obvious loopholes. Therefore, this study speculates that Jim gray did not put forward the so-called Jim gray new Moore's law, which should be a misunderstanding of some views of Jim gray by some domestic scholars. Finally, this study holds that the new Moore's law of the exponential growth of the total global data can be expressed as: the total global data will double every three years.

Key Words: big data; amount of data; new Moore's law; Jim Gray

0 引言

1965年4月,《电子》杂志(Electronics)发表了一篇只有3页的论文——Cramming more components onto integrated circuits(《集成电路里正塞入更多的元件》)^[1]。文章作者戈登·摩尔(Gordon Moore)当时就职于仙童摄影器材公司的子

公司仙童半导体(Fairchild Semiconductor),其在文中指出:随着成本曲线的降低,集成电路上可以封装的元件数目迅速增加,并预测到1970年,一个集成电路上可以集成1000个元件,每个元件的生产成本将是1965年的1/10,并且集成电路的集成程度每年将以2倍速增长,以公式表示即为: $f(\Delta t_n) = Q2^n$ 。据此测算,1975年最低成本下单位集成电路封装的元件数目可以达到65000个,具体如图1所示。

收稿日期:2021-09-13

基金项目:湖北省教育厅人文社会科学研究一般项目(鄂教思政函[2018]2号,18G035)

作者简介:辛冬播(1980-),男,江汉大学教育学院硕士研究生,研究方向为数字化学习;李廷军(1970-),男,博士,江汉大学教育学院教授、硕士生导师,研究方向为数字化学习、媒体素养教育。

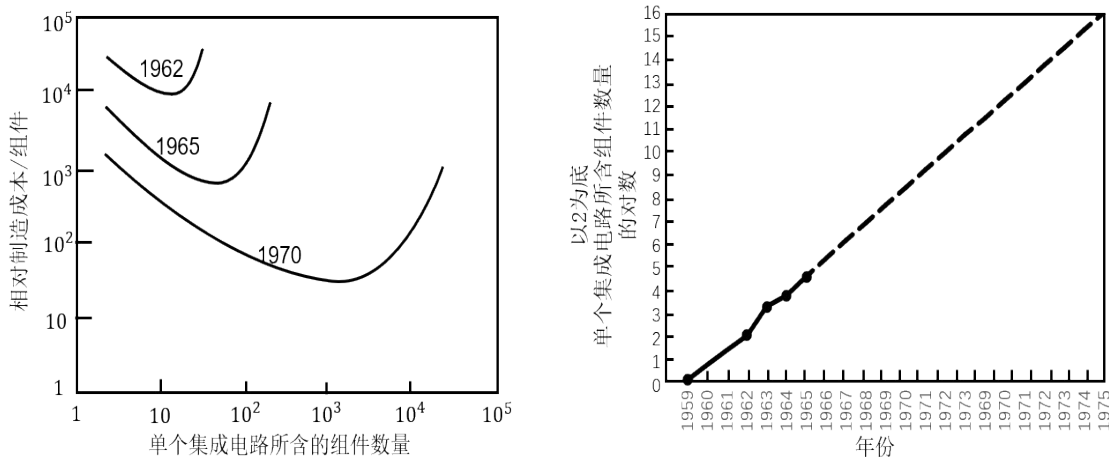


Fig. 1 Cost curve (left), component integration (right)

图 1 成本曲线(左)、元件集成度(右)

1968年,戈登·摩尔离开仙童公司并创立了英特尔(Intel),他继续将上述观察和推测作为芯片研发工作的指引,而后集成电路的不断发展印证了他的预测,于是有人将其命名为摩尔定律^[2]。计算机科学家卡弗·米德(Carver Mead)在很多场合谈到过摩尔定律,促进了其传播,曾被认为是其命名者,但实际情况已难以考证^[3]。1975年,戈登·摩尔认为集成电路增长速度会有所放缓,遂将摩尔定律中的速度修改为每2年翻一番。虽然此后实际状况基本上是按每18个月翻一番的速度发展的,但戈登·摩尔并没有再修改其定律^[4]。

摩尔定律原指单位集成电路封装的元件数目增加及单位成本相应下降的趋势随时间发展的一种规律,但在传播过程中出现了很多版本,主要有以下几种:①集成电路芯片上所集成电路的数目每隔18个月就翻一番;②微处理器的性能每隔18个月提高一倍,而价格下降一半^[5];③IC芯片上可容纳的晶体管数目每隔约18个月便会增加一倍,计算机性能也将提升一倍^[6]。严格来说,摩尔定律并不是定律,而只是一种观察和推测。经过50多年的发展,戈登·摩尔认为集成电路技术的发展速度会减慢,摩尔定律可能会逐步退出历史舞台。

随着信息技术的发展,人们开始将摩尔定律套用到各种指数式发展的技术和产品上,于是安放在吉姆·格雷(Jim Gray)名下的新摩尔定律便横空出世了,即每18个月全球新增信息量是计算机有史以来全部信息量的总和。吉姆·格雷是当代最杰出的数据库大师,是数据库和交换处理系统领域的领头人,是美国国家工程院院士、美国计算机协会(ACM)院士、微软公司高级研究员,曾获得1994年美国技术研究院大奖,以及号称计算机行业诺贝尔奖的图灵奖(1998年)^[7]。遗憾的是,他在2007年出海后失踪^[8]。与戈登·摩尔专注基础技术半导体工业不同,吉姆·格雷是数据库及处理系统方面的专家,如果他真的提出过关于数据量增长的新摩尔定律,以其背景而言似乎水到渠成,理所当然。然而,笔者认为所谓的吉姆·格雷新摩尔定

律的表述有些问题。计算机有史以来全部信息量总和是否为固定值并不明确。如果按固定值来看,以Q表示计算机有史以来全部信息量总和,那么再过18个月全球新增信息量 $f(\Delta t_1) = f(t_2) - f(t_1) = Q$,而36个月之后 $f(\Delta t_2) = f(t_3) - f(t_2) = Q$,其总量增速如图2所示,这是一种减速递增关系,简单计算能得出其倍数将依次为2、1.5、1.33、1.25……。而摩尔定律是一段周期翻一番的2倍指数递增关系($f(\Delta t_n) = Q2^n$),二者相差甚远。如果计算机有史以来全部信息量总和不是固定值,即也包括了每18个月的信息量,那么就会造成逻辑混乱。笔者深感困惑,于是追溯相关文献,探究吉姆·格雷是否真的提出过新摩尔定律。

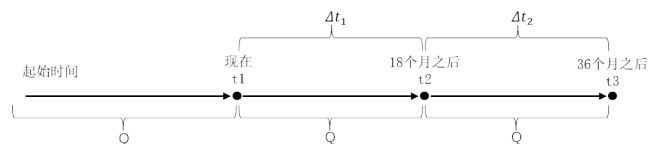


Fig. 2 The information quantity increment relationship

图 2 信息量增量关系

1 新摩尔定律文献统计与相关表述

1.1 新摩尔定律文献统计

在知网以“新摩尔定律”为关键词进行主题搜索,整理得到65篇中文文献,如图3所示。文献发表年度跨度为1999-2020年,其中以2009年发表的文献数量最多,2000年次之,2015年再次之。

采用Citespace软件分析关键词共现情况,结果如图4所示。可以看出,各关键词之间离散程度较大,但新摩尔定律与大数据、摩尔定律的关联性在图中仍有所体现。

关键词突现分析结果如图5所示,新摩尔定律作为关键词于2006年首次出现。以下将对相关文献表述进行具体分析。

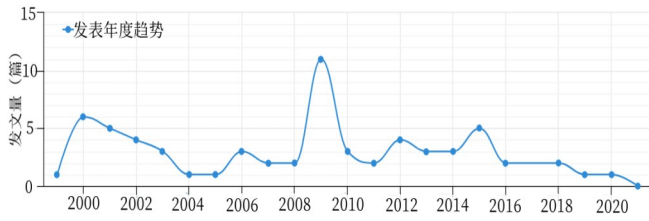


Fig. 3 Annual trends of published Chinese literatures about "New Moore's Law"

图3 新摩尔定律相关中文文献发表年度趋势



Fig. 4 Key words co-occurrence

图4 关键词共现



Fig. 5 Key words with the strongest citation bursts

图5 关键词突现

在 Web of Science 中以“data(数据)”和“Moore's Law(摩尔定律)”为关键词进行主题搜索,共查询到 171 篇相关英文文献。同时,在这 171 篇文献中再以“Gray”为关键词进行搜索,文献结果为 0,这说明相关英文文献的标题、摘要、关键词中没有关于吉姆·格雷新摩尔定律的表述。据此初步推测,吉姆·格雷并未提出过新摩尔定律,之所以中文文献中出现相关表述,应是国内学者对吉姆·格雷某些观点的误读。

1.2 新摩尔定律相关表述

检索到的中文文献中关于新摩尔定律的表述主要有以下几种:1999年,侯自强^[9]提出:“我们不妨称因特网流量每半年翻一番的估计为网络发展的新摩尔定律。”2006年,王克朝^[10]在其硕士学位论文的摘要中提到:“图灵奖获得者 Jim Gray 提出了一个新的经验定律:网络环境下每 18

个月产生的数据量等于有史以来数据量之和。”2007年,谢长生等^[11]提到:“信息时代的到来意味着数字信息量的不断增长。1998年图灵奖获得者 Jim Gray 曾说过:‘从现在起,每 18 个月新增存储量等于有史以来存储量之和。’”2008年,孟小峰^[12]提到:“图灵奖获得者 Jim Gray 曾在 1998 年的获奖演说中对未来数据量急剧增长的规律做过这样的预言:未来每 18 个月产生的数据量等于有史以来的数据量之和!”2011年,罗东健^[13]提到:“互联网环境下,每 18 个月新产生的数据量等于有史以来数据量之和。”2013年,杨寅^[14]提出:“图灵奖获得者 James Gray 总结并提出在当今网络应用环境下,每 18 个月新产生和增加的数据存储总量等于有史以来所有数据存储量之和的经验定律。”2013年及 2015 年,刘鹏等^[15-16]提到:“1998 年图灵奖获得者杰姆·格雷(Jim Gray)提出著名的新摩尔定律:每 18 个月全球新增信息量是计算机有史以来全部信息量的总和。”2018年,周兰^[17]指出:“图灵奖获得者 Jim Gray 更是提出新摩尔定律,即每 18 个月全球新增信息量是计算机有史以来全部信息量的总和。”2020年,董凯^[18]提到:“图灵奖获得者杰姆·格雷提出的著名的新摩尔定律:每 18 个月全球新增信息量是计算机有史以来全部信息量的总和。”

需要指出的是,相关学者引用所谓的新摩尔定律时并未辨析来源,也没有结合统计数据加以验证,这也正是令人困惑之处。

2 新摩尔定律的源流及辨析

追溯相关文献可以发现,1999 年侯自强首先提出了新摩尔定律的说法,但只是个人经验总结,并没有具名吉姆·格雷,其所提出的是网络流量每半年翻一番与 18 个月的时间间隔也不一样。最早完整提出吉姆·格雷新摩尔定律的应是 2006 年王克朝的硕士学位论文,其明确认定该定律是吉姆·格雷的经验定律,并提出“18 个月”和“数据量”等关键词。2007 年谢长生的表述中虽然用到的词是存储量,但其紧随在“数字信息量的不断增长”表述之后,极易使人将该存储量误解为数据量,之后相关文献中关键词采用的是数据量、信息量之类的表述。此外,孟小峰关于“图灵奖获得者 Jim Gray 曾在 1998 年的获奖演说中……”的表述是不严谨的。吉姆·格雷是为 1998 年度奖项作的报告,但是在 1999 年演讲的,与谢长石“1998 年图灵奖获得者 Jim Gray 曾说过”的表述相比,孟小峰的表述会使人误以为是在 1998 年发表的演讲。2013 年,刘鹏等将此类表述单独冠以新摩尔定律的称谓,之后其他文献的表述与其一致,但均未注明出处。

以上提到新摩尔定律的文献中,只有 2008 年孟小峰教授的引用注明了出处:What next? A dozen information-technology research goals, J. Gray, ACM Turing Award Lecture, June 1999, MS-TR-99-50^[19]。该文献发表于 2003

年,为吉姆·格雷图灵奖获奖演说的精简版,文中有3处提到摩尔定律,分别为:①Prices are falling faster than Moore's law——storage will likely be a hundred times cheaper in ten years(价格下跌速度比摩尔定律快——10年后存储器价格可能会便宜100倍);②Moore's law predicts a doubling every 18 months. This means that, in the next 18 months, there will be as much new storage as all storage ever built, as much new processing as all the processors ever built. The area under the curve in the next 18 months equals the area under the curve for all human history(摩尔定律预测每18个月就会翻一番。这意味着,18个月后新增的存储器(数量)将与现在已有的存储器一样多,18个月后新增的处理能力将与现在已有的处理器一样大。未来18个月曲线下面积等于现在为止所有人类历史曲线下的面积);③As I said at the beginning, progress appears to be accelerating; the base-technology progress, in the next 18 months, will equal all previous progress, if Moore's law holds(正如我在一开始所说的,进展似乎正在加速;如果摩尔定律成立,未来18个月基础技术的进展将等于到现在为止的所有进展)。

为进一步验证,根据2003年这篇文献的脚注搜索到1999年版本的原文^[20],文中对应的3处表述,与上文完全一致。吉姆·格雷在这份获奖发言中主要阐述了12个长期系统研究问题:①可扩展性(Scalability)。设计一个可扩展到 10^6 倍的软件和硬件体系结构,仅通过添加更多资源,应用程序的存储和处理容量可以自动增加100万倍,工作速度更快(加速 10^6 倍)或同时做 10^6 倍的工作;②图灵测试(The Turing Test)。建立一个计算机系统,模仿游戏的胜率至少达到30%;③语音文本(Speech to Text)。水平如同母语者;④文本语音(Text to Speech)。水平如同母语者;⑤视觉如人一样(See as Well as a Person)。能识别物体和运动;⑥个人记忆扩展器(Personal Memex)。记录一个人看到和听到的一切,并根据需要迅速检索到;⑦世界备忘录(World Memex)。建立一个给定文本语料库系统,可以回答有关文本的问题,并像该领域的人类专家一样准确、快速地总结文本,这个系统也能适用于音乐、图像、艺术和电影;⑧远程临场(Tele Presence)。模拟其他地方,使你作为观察者(远程观察),就如同一位在现场的人一样能听到和看到;模拟作为参与者出席另一地点的活动(远程临场)。与他人和环境互动,就好像你真的在那里一样;⑨无故障的系统(Trouble-free Systems)。建立一个每天被数百万人使用的系统,而它只需要一个兼职人员管理;⑩安全系统(Secure System)。确保问题⑨的系统仅服务授权用户,服务不能被未经授权者中断,信息也不能被窃取(需验证);⑪随时可用(Always Up)。确保系统每百年故障不到1s,即有99.999 999%的可靠性(需验证);⑫自动编程(Automatic Programmer)。设计一种规范语言或用户界面,可使人们更容易表述设计(更容易1 000倍),使计算机可直

接编译,并可描述所有应用程序(已完成)。系统应能质疑应用程序,对例外情况和规范完整性提出疑问,但使用起来不应很繁琐。

通过分析全文可以发现,吉姆·格雷阐述的增长是指存储(Storage)和处理器(Processor),而非数据量或信息量,其在文中提到的摩尔定律也只是旧摩尔定律,并未就数据量增长提出类似的说法。此外,吉姆·格雷没有将摩尔定律作为这篇文献的关键词之一,也间接证明了这一点。

进一步搜索关于吉姆·格雷的信息,在Web of Science核心合集发现其有3篇以“Gray J”署名的文献,包含在其ACM作者档案^[21]中所列的112篇出版文献中,其中就有What next? A dozen information-technology research goals(1999年及2003年)。此外,微软公司eScience研究组(Microsoft Research's eScience Group)在吉姆·格雷主页^[22]整理了其未完结的项目,在这些文献篇名、关键词、摘要等中均未出现Moore's Law(摩尔定律)这一关键词,因此无法证实吉姆·格雷提出过新摩尔定律。

针对渐有流行趋势的吉姆·格雷新摩尔定律,首先经文献分析发现大多数中文文献中未注明引用出处,也没有结合统计数据加以验证;其次进一步追根溯源,分析各文献表述,理清了国内有关新摩尔定律的各种来由;最后辨析各中文文献原文,查证相关英文文献,未找到吉姆·格雷提出新摩尔定律的直接证据。结合逻辑分析,以及全球数据量增长的相关统计,所谓的吉姆·格雷新摩尔定律表述本身也有明显漏洞。据此推测,吉姆·格雷并没有提出过新摩尔定律,该定律应是国内学者对吉姆·格雷某些观点的误读。当然,吉姆·格雷也许在其他地方确实提出过类似观点,受限于文献查找范围没有被发现。

3 大数据时代信息总量增长规律

所谓的吉姆·格雷新摩尔定律一直强调计算机有史以来全部信息量,而人类自公元前3000年开始书写后就在不断创造和保存数据信息,而且人类历史上技术进步促进数据和信息量大幅增长的情况也曾发生过,例如15世纪印刷术发明后以及19世纪后期出版成本下降时。计算机及互联网的相继问世使数据增长速度今非昔比,如今进入大数据时代,全球数据量更是飞速增长。根据Statista的统计和预测,2020年全球数据产生量预计达到47ZB(1ZB即 10^{12} GB),而到2035年,这一数字将达到2 142ZB。信息技术在未来较长时期内将保持渐进式发展态势,其中数据处理能力远落后于指数级增长的数据体量。图6为全球每年产生数据量估算图,这一指数级增长速率,即年均增速平均为1.3倍,也就是约每3年翻一番,这与国际数据公司(IDC)预测的2020-2024年数据增长的年复合增长率(CAGR)26%基本一致。

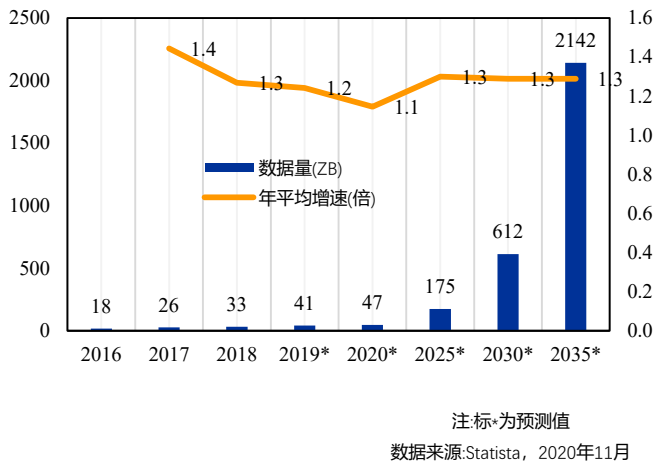


Fig. 6 Global annual data volume estimation

图6 全球每年产生数据量估算

4 结语

全球数据总量飞速增长的趋势已是共识,其增长规律预测的可信度与是否为吉姆·格雷所提出并无直接关系。将全球数据总量指数级的增长规律命名为新摩尔定律无可厚非,但也完全没有必要无中生有地拉吉姆·格雷来为其站台。根据前文提到的相关统计及预测,大数据时代数据总量增长的新摩尔定律应该表述为全球数据总量每3年将翻一番。同样的,这个新摩尔定律也只是一种观察和推测,并不是一个真正的定律,今后完全可以根据数据总量增长的实际情况进行修正。

参考文献:

- [1] MOORE G E. Craming more components onto integrated circuits[J]. Proceedings of the IEEE, 2002, 86(1):82-85.
- [2] ZHANG J. Gordon Moore's name means progressive——interviewing Moore on the 50th anniversary of Moore's law[J]. World Science, 2015(6):23-25.
张洁. 戈登·摩尔名字意味着进步——摩尔定律50周年之际专访摩尔[J]. 世界科学, 2015(6):23-25.
- [3] RACHEL C. The murky origins of "Moore's law"[EB/OL]. <https://spectrum.ieee.org/tech-talk/semiconductors/devices/the-murky-origins-of-moores-law>.
- [4] TONG Z Y. Moore's Law——extends to the 2010——interview with Intel founder Gordon Moore[J]. Equipment for Electronic Products Manufacturing, 1998(3):46,51-54.
童志义. 摩尔定律——向2010年延伸——对Intel公司创始人戈登·摩尔的专访[J]. 电子工业专用设备, 1998(3):46,51-54.
- [5] ZHOU S, WANG S P. Management information system in the era of big data[M]. Beijing: China Railway Press House, 2017.
周苏,王硕革. 大数据时代管理信息系统[M]. 北京:中国铁道出版社, 2017.
- [6] MING T. The spirit of innovation embodied in Moore's law lasts forever——marking the 40th anniversary of the publication of Moore's Law[J]. Semiconductor Technology, 2005(6):5-7.
明天. 摩尔定律体现的创新精神永存——纪念摩尔定律发表40周年[J]. 半导体技术, 2005(6):5-7.
- [7] LIN J. Jim Gray as one of Microsoft three musketeers: the giant behind the hero[J]. Software Engineering, 2000(7):57-62.
林君. 微软三剑客之吉姆·格雷:英雄背后的巨人[J]. 软件工程师, 2000(7):57-62.
- [8] LIU J. Superstar Jim Gray, who died with the waves[J]. Programmer, 2012(2):135-135.
刘江. 随波而逝的巨星Jim Gray[J]. 程序员, 2012(2):135-135.
- [9] HOU Z Q. New "Moore's law" and information construction for the development of the Internet[J]. Computer Security, 1999(2):7-9.
侯自强. 因特网发展的新“摩尔定律”和信息化建设[J]. 电子展望与决策, 1999(2):7-9.
- [10] WANG K C. Research of the reliability of network storage system based on redundancy mechanisms[D]. Wuhan: Huazhong University of Science & Technology, 2006.
王克朝. 基于冗余机制的网络存储系统可靠性研究[D]. 武汉:华中科技大学, 2006.
- [11] XIE C S, SHI Y J, WANG H W. Application of adaptive equilibrium in magnetic storage channels[J]. China Mediatech, 2007(2):53-54.
谢长生,石媛晶,王海卫. 自适应均衡在磁存储通道中的应用[J]. 记录媒体技术, 2007(2):53-54.
- [12] MENG X F. Annual report of WAMDM lab 2008 preface[EB/OL]. <http://idke.ruc.edu.cn/reports/report2008/Preface/Preface.pdf>.
孟小峰. 中国人民大学网络与移动数据管理实验室年报 WAMDM Report 2008 序[EB/OL]. <http://idke.ruc.edu.cn/reports/report2008/Preface/Preface.pdf>.
- [13] LUO D J. Research on key technology of high reliability in very large scale storage systems[D]. Wuhan: Huazhong University of Science & Technology, 2011.
罗东健. 大规模存储系统高可靠性关键技术研究[D]. 武汉:华中科技大学, 2011.
- [14] YANG Y. Research on key technology of reliability in storage systems[D]. Wuhan: Huazhong University of Science & Technology, 2013.
杨寅. 存储系统可靠性关键技术研究[D]. 武汉:华中科技大学, 2013.
- [15] LIU P, WU Z F, HU G Y. Big data——profound changes taking place[J]. ZTE Technology Journal, 2013, 19(4):2-7.
刘鹏,吴兆峰,胡谷雨. 大数据——正在发生的深刻变革[J]. 中兴通讯技术, 2013, 19(4):2-7.
- [16] LIU P, WU Z F, HU G Y. New Moore's law: in-depth analysis of the era of big data coming[J]. China Strategic Emerging Industry, 2015(7):76-77.
刘鹏,吴兆峰,胡谷雨. 新摩尔定律:深入分析大数据的时代来临[J]. 中国战略新兴产业, 2015(7):76-77.
- [17] ZHOU L. Computing chip: the cornerstone of big data and artificial intelligence[J]. China Radio, 2018(12):42-43.
周兰. 计算芯片:大数据和人工智能的基石[J]. 中国无线电, 2018(12):42-43.
- [18] DONG K. Exploring the operation and maintenance of cloud computing data center[J]. China New Telecommunications, 2020, 22(9):40.
董凯. 论云计算数据中心运行维护探索[J]. 中国新通信, 2020, 22(9):40.
- [19] GRAY J. What next? a dozen information-technology research goals[J]. Journal of the Association for Computing Machinery, 2003, 50(1):41-57.
- [20] GRAY J. What next? a dozen information-technology research goals[EB/OL]. <https://arxiv.org/ftp/cs/papers/9911/9911005.pdf>.
- [21] Digital LibraryACM. Author's profile/Jim Gray[EB/OL]. <https://dl.acm.org/profile/81100403088>.
- [22] Microsoft eScience Group. Jim Gray summary home page[EB/OL]. <http://jimgray.azurewebsites.net/>.

(责任编辑:尹晨茹)