

高性能计算的智慧城市机遇与挑战

冯圣中 樊建平

(中国科学院深圳先进技术研究院 深圳 518055)

摘 要 智慧城市、智慧地球是全球信息化浪潮的新发展阶段。作为支撑智慧城市深入发展和广泛应用的核心技术,高性能计算也面临着挑战和机遇。本文在分析智慧城市需求的基础上,把这些对高性能计算的挑战概括为大数计算、实时计算、效能计算和社会计算等四个方面,并简要介绍了中国科学院深圳先进技术研究院在这些方向上的布局及取得的进展。

关键词 智慧城市; 高性能计算; 先进技术研究院

The Challenge and Opportunity of Smart City to High Performance Computing

FENG Sheng-zhong FAN Jian-ping

(Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055)

Abstract Smart city and smart planet are the new stages of global information development. As the core technology for the development of smart city, high performance computing got new challenges and opportunities. The challenges of smart city to high performance computing include: massive data computing, realtime computing, social computing and effective computing. Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Sciences, is working on these areas; the latest development and infrastructure built at SIAT are summarized in this paper.

Keywords smart city; high performance computing; SIAT

1 引 言

智慧城市是城市生产、生活和社会服务信息化进程与前景的形象描述,其基本支撑技术包括互联网、移动互联网和物联网等。智慧城市将使得交通、教育、环境、生产等各类管理与服务更加智能、便捷和泛在。

高性能计算被誉为计算机科学皇冠上的明珠。2010年6月,曙光星云系统在全球高性能计算500强排名榜中位居第二;2010年11月,天河一号排名第一;2011年10月,具有完全自主知识产权的千万亿次高性能计算机神威蓝光系统问世。我国在高性能计算机研制方面已经取得重要进展,高性能计算应用也因此受到国内外各界人士的更多关注。

互联网、移动互联网、物联网应用日益广泛和深入,计算机已经成为我们生产、生活中必不可少的工具。今天的一部普通智能手机性能已经超过了十年前的一台笔记本电脑,就是一台性能不错的“计算机”。在以智慧城市为愿景的信息化建设快速推进的背景下,高性能计算能够发挥什么样的作用?未来发展方向在哪里?本文将探讨这些问题。

2 智慧城市发展对高性能计算技术提出新的挑战

应用需求是牵引高性能计算发展的主要力量。高性能计算在计算机技术和应用发展历史上,发挥着重要作用。计算机发展历史上的众多重大进展,往往是首先在高性能计算上取得突破,然后逐步下移到其

冯圣中,研究员,研究方向为高性能计算、网格计算、生物信息学等。樊建平,研究员,博士生导师,中国科学院深圳先进技术研究院院长,主要研究方向为体系结构、并行处理、操作系统等。

他计算设备。第一台电子计算机ENIAC1946年2月24日问世，占地170平方米、耗电150KW，计算能力5000次加法/秒或400次乘法/秒。2011年11月，我国自主研发的神威蓝光，几乎同样占地面积，耗电不超过2000KW，计算能力达到1千万亿次，比第一台计算机ENIAC计算能力提高了 10^{12} 倍。现在，就是一台普通的智能手机，计算能力通常都超过1GFLOPS（10亿次浮点计算/秒），大概是第一台电子计算机的1百万倍。图1是近20年来高性能计算机性能提升曲线。预计在未来10年，这种发展趋势仍将继续。

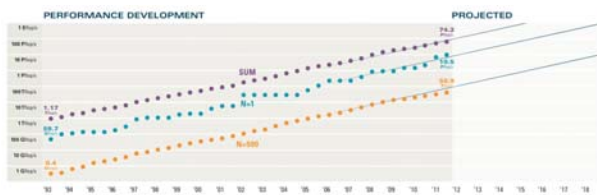


图1 高性能计算机发展趋势

现在常见的智能手机也支持的多用户或多任务，而最早实现这一特性的是1964年由贝尔实验室、麻省理工学院及美国通用电气公司所共同参与研发的安装在大型主机上Multics操作系统。苹果公司超大规模云计算平台支撑的Apple App Store上，应用软件就达到35万个之多，下载量已经超过250亿次。

智慧城市快速发展，必然伴随着数据增长更快、服务质量要求更高、降低计算能耗更迫切、虚拟-实体社会的交互更加复杂，高性能计算在这些领域大有作为，高性能计算的发展面临难得机遇。

2.1 大数计算

大数计算包括大规模数据管理与大规模数据处理。智慧城市背景下，各类传感器等数据采集设备普遍应用，数据呈现爆炸性增长。深圳市全市20-30万个交通、安全监控摄像头，每天产生PB量级的数据。据IDC统计，仅在2011年，全世界就有1.8ZB（即1.8万亿GB）的大数据产生。这相当于每位美国人每分钟写3条微博，而且还是不停地写2.6976万年。

大数据快速增长的主要原因是智能设备的普及，比如音视频采集设备、传感器、医疗设备等。非结构化信息如视频数据将占到未来10年新生数据的90%。

大规模数据的产生，对数据管理与处理技术提出严峻的挑战。过去6年，由于数据存储技术、数据压缩技术、数据管理技术等技术进步，单位数据的创建、存储、管理和搜索成本下降到1/6。6年来，企业在硬件、软件、管理和服务上的投资只增加了50%，达到4万亿美元。

现在面临的严峻挑战是：如何持续这种技术创新，以管理和处理更大规模的数据？

2.2 及时计算

及时计算，或者叫实时计算，在不同领域有不同的含义。智慧城市中，常用来指请求-响应等信息服务等感觉不到延迟，即小于1秒。“请求-响应”是智慧城市中“人-环境-信息”的基本交互方式。这种交互方式的重要需求是及时性。

智慧城市领域的及时计算一般都是针对海量数据流进行的。其主要特点是：数据源是实时的持续的，要求对用户请求的响应也是实时的，这种计算又称为实时流计算。数据流是在时间上持续延伸的系列记录。如，城市实时路况信息服务，部署在城市道路上的交通状况采集系统持续采集数据，并及时传送到中央服务器，中央服务器数据融合并提供实时路况服务。又如大型网站用户访问日志，用户访问了什么内容，访问内容的时序等。实时流计算可以动态实时分析每时刻流量和用户分布情况，这对于大型网站来说具有重要意义。

智慧城市数据流的实时计算过程可以划分为数据的产生、收集、传输、存储、分析、服务等过程。每一个过程都会有开销。对于高性能计算，研究重点在大规模数据流分析、服务等方面，即，如何更快速地处理更多的请求？如何获得更精准、更快捷的响应？

2.3 效能计算

智慧城市中的各类设备，包括大型服务器、智能终端、传感器和网络设备，都面临着降低能耗的问题。大多数终端设备，包括智能终端、传感器等，都采用锂电池、太阳能等供电方式，能耗大小直接决定其持续工作时间。

美国第一个IT节能计划是1992年环保署发起的“能源之星”，目标是降低所有硬件能耗。近十年来，无线传感网络环境下的效能计算受到广泛关注，如跟踪斑马的ZabraNet项目，从数据处理、通信协议等各个层次开展工作，以尽量延长传感网络工作时间，获取更多有效数据。

美国未来4年将关闭800个以上小型IDC（G1/G2），同时，正在巨资建造新一代节能低碳排的IDC（G3/G4），比如微软、谷歌、亚马逊、facebook、苹果都大力投入巨资，建造巨型数据中心。

大型数据中心能耗已经接近其采购成本，效能计算意味着大幅度降低其运营成本。图2是过去5年来，全球计算机最高计算效能的变化情况。可以看

到,过去5年,计算机效能有了显著的提升,达到5倍以上,这主要归功于低功耗芯片、制冷技术、先进集成技术的应用。但是,全球IT设备能耗仍以年增20%以上的速度在增长。而且计算机效能已经从快速提高过渡到平稳阶段,效能提升很难持续。如何进一步提高计算效能,实现可持续发展,是高性能计算面临的重大挑战。

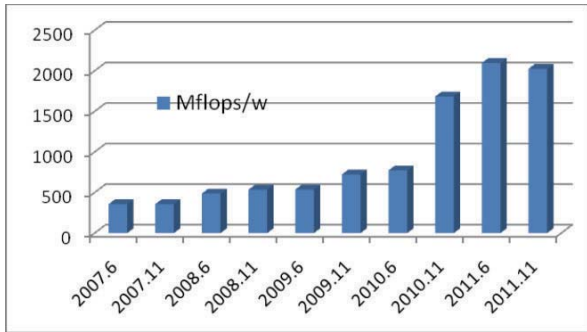


图2 过去5年计算机计算效能变化

2.4 社会计算

2009年2月,哈佛大学Lazer等15位美国学者在《Science》发表“Computational Social Science”,该文指出,计算社会科学这一研究领域正在兴起,人们将在前所未有的深度和广度上自动地收集和利用数据,增强社会问题研究的科学性。

智慧城市构建了一个虚拟社会,这个虚拟社会不只是实体社会的简单映射,相对地,虚拟社会对实体社会也会产生影响,两者之间形成交互。这种交互更值得关注。

计算社会科学,或者社会计算,是指运用计算技术,开展社会行为研究的科学。这里社会可能是虚拟社会,也可能是实体社会。社会行为包括社会群体行为、社会结构演化、社会运动趋势等。社会计算的研究成果,将帮助人们进行沟通与协作。社会计算研究包括社会网络分析、内容计算、社会模拟等。

社会网络是由节点(独立角色)和边(社会关系)构成的一张图,这张图往往非常复杂,节点之间的关系类型多种多样。社区结构发现,对网络信息搜索与挖掘、信息推荐以及网络演化与扩散预测具有重要价值。如何快速准确度量、发现和利用这些规律是大规模社会信息网络分析与处理的重要问题。通过对各种社会媒体,如新闻、论坛、博客、微博等的跟踪与挖掘,结合传统的舆论分析理论,可以有效地观察社会的状态,并辅助决策。如何快速准确捕捉热点舆情、预测其演化规律,甚至提出有效干预措施,是社会网络快速发展,虚拟-实体交互日益复杂背景下的

重要研究课题。对复杂社会问题建立计算模型,利用高性能计算机开展模拟试验,验证和预测社会经济政策的效果,对于提高公共政策制定的科学性具有十分重要的意义,目前,这一领域仍处于十分初级阶段。

社会网络分析、内容计算、社会模拟,既涉及大规模数据,又有较强的时效性,对高性能计算提出巨大挑战。

3 中国科学院深圳先进技术研究院的部署与进展

智慧城市已经越过了实体城市的疆域而拥有更广泛的内涵,其核心是发展和应用信息技术,为人类、为社会的运动、交流和协作,提供更智能、更便捷、更普惠的技术支撑。为支撑智慧城市发展,应对智慧城市提出的一系列挑战,中国科学院提出了面向“感知中国”的海云工程先导专项,中国科学院深圳先进技术研究院在高性能计算和智慧城市方面进行了部署,并获得了各类基金支持。

近5年来,中国科学院深圳先进技术研究院先后承担了863重点项目“华南高性能计算与数据模拟网格节点”、“深港创新圈网格节点”研制任务。深圳先进技术研究院领导敏锐观察到智慧城市快速发展背景下,大数据时代即将来临,2006年,首创性地部署了高性能数据密集型计算学科方向工作。这个工作早于IBM“云计算”战略(2007年底)、Gray提出第四范式“数据驱动的科学发现”(2009年)。在这一战略指导下,凝聚了一批国内外顶尖人才,完成了中国科学院深圳超算中心建设,并支撑了深圳市建市以来的最大科研基础设施国家超算深圳中心建设。

在效能计算方面,承担了中国科学院知识创新工程重点方向项目“数据超算中心智能绿色计算模型及系统研发”,前瞻性把握超大数据中心发展趋势,在绿色计算方面做了有益探索。

在大规模系统管理、大规模数据管理、能量管理、安全管理等方面,须成忠把握到管理技术在未来信息产业中的核心作用,以此为立足点,已经得到国家发改委云计算专项、深圳市产学研重大科技攻关等计划支持,并获得深圳市政府支持,筹建“高效能云技术工程实验室”,形成了先进云品牌。

在大规模数据处理等方面,已经获得中国科学院海云工程先导专项支持,并获得深圳市政府支持,筹

建“高性能数据挖掘重点实验室”。在健康数据挖掘、在线算法设计、图搜索、以及蛋白质结构预测和药物筛选等方面的工作受到国内外关注。

智慧城市方面，中国科学院深圳先进技术研究院也凝聚了一支高素质团队。在城市三维建模等方面，先后获得国家863、国家自然科学基金杰出青年基金、深圳市产学研重大科技攻关等计划支持，并获得深圳市政府支持，筹建“可视计算重点实验室”。在城市建筑群等三维自动建模方面进展显著，连续两年在国际顶级会议SIGGRAPH上发表多篇论文，受到瞩目。

在三维地理信息系统与遥感方面，先后承担了中亚国际合作重大科研计划项目、中国科学院先导专项课题、国家自然科学基金项目等，和联通集团、深圳市交委等建立了深入合作，成为我院横向合作和产业化工作的一面旗帜。

在城市实时感知与无线网络方面，先后承担了国家863重点项目、国家自然科学基金项目等支持，和深圳市交委等建立了广泛合作，在infoCOMM等发表系列成果。

在嵌入式软件系统、多媒体计算、科学计算等方面，也承担了国家自然科学基金、深圳市杰出青年基金等支持，并承担了一汽、深圳市交委等委托研发项目，关于智能电网安全的研究成果发表在国际学术期刊Proceedings of the IEEE。

数字所凝聚了包括新世纪百千万人才工程国家级人才入选者、国家千人计划入选者、国家杰出青年基金获得者、广东省领军人才入选者、中科院百人计划入选者在内的高端科技创新人才60余名，成为推动海云工程技术创新和产业创新的重要力量。

参 考 文 献

- 中国计算机学会通信, 2010,7.
- [7] 范伟, 李晓明. 物联网数据特性对建模和挖掘的挑战 [J]. 中国计算机学会通信, 2010, 9.
- [8] Lazer D, et al. Computational social science [J]. Science, 2009: 721-723.
- [9] Lynch C. Big data: how do your data grow [J]. Nature, 2008, 455: 28-29.
- [10] Steinbrook R. Personally controlled online health data-the next big thing in medical care? [J]. N. Engl. J. Med., 2008, 358:1653-1656.
- [11] Hey T, Tansley S, Tolle K. The fourth paradigm: data-intensive scientific discovery [J]. Microsoft Research, 2009.
- [1] Wayne W. The good news and the bad news(embedded computing column) [J]. IEEE Computer, 2007, 40 (11): 104.
- [2] Jones W. Forecasting traffic flow [J]. IEEE Spectrum, 2001, 38(1): 90-91.
- [3] Murugesan S. Harnessing green IT: principles and practices [J]. IEEE IT Professional, 2008, 1: 24-33.
- [4] Feng W C, Cameron K. The green500 list: encouraging sustainable supercomputing [J]. Computer,2007, 40 (12): 50-55.
- [5] Dongarra J J. Performance of various computers using standard linear equations software, (Linpack Benchmark Report) [J]. University of Tennessee Computer Science Technical Report, CS-89-85, 2011.
- [6] 孙凝晖, 徐志伟, 李国杰. 海计算: 物联网的新型计算模型 [J].