

超级计算中心核心应用的浅析

党 岗 程志全

(国防科技大学计算机学院 长沙 410073)

摘 要 目前,我国国家级超算中心大多采用“地方政府投资、以市场为导向开展应用”的建设思路,地方政府更关心涉及本地企事业单位的高性能计算应用和服务,超算中心常被用于普通的应用,很难充分发挥超级计算的战略作用。如何让超算中心这艘能力超强的航母生存下来,进而“攻城掠地”,推动技术创新,一直是业内人士研究的课题。初步探讨了国内超算中心核心应用所面临的挑战,提出了超算中心核心应用服务地方建设的几点建议。

关键词 超级计算机,超算中心,核心应用

中图法分类号 TP391 文献标识码 A

Brief Exploration on Technical Development of Key Applications at Supercomputing Center

DANG Gang CHENG Zhi-quan

(School of Computer, National University of Defense Technology, Changsha 410073, China)

Abstract National supercomputing centers at China work use building thought of local government investigation, and market-oriented application performing. Supercomputing resources are always applied at general applications, as the local government more focuses on the high-performance computing applications and services related to local business, rather than supercomputing working as strategical role in the traditional way. It is a long-term researching topic how to make the supercomputing powerful as a super-carrier active and applicable to benefit the technical innovation. Some challenging technical issues suiting for the supercomputing were discussed by taking domestic supercomputing center as the example, and some useful advises were addressed for applying international supercomputing center at local services.

Keywords Super computer, Supercomputing center, Key applications

1 引言

超级计算(简称超算)是解决国家社会、经济、科学研究、国防建设以及安全等领域中一系列具有挑战性问题的的重要手段,是国家综合国力的重要体现,也是国家创新体系的重要组成部分。超级计算已经成为世界各国特别是大国争夺的战略制高点^[1,2]。对于超级计算机的突出地位有很多比喻:从“更高”的视角,它被比喻为“计算机领域的珠穆朗玛峰”;从“更快”的视角,它可称为“高速公路”。本文从“更强”的角度,将其视作计算机技术海洋中的“航空母舰”。笔者认为较前两者,“超算”的“航母”形象更能显示超级计算机的国家战略定位和系统性功效。

超级计算能力和应用水平关乎一个国家或地区在信息技术领域国际竞争的成败,关乎工业设计、高技术产业、基础科研等的长期可持续发展。十二五期间,国家科技部设立了“高性能计算机及应用服务环境”重大项目,旨在面向重大挑战性应用需求,以自主技术为主,突破体系结构设计、高速互连、并行编程模型、功耗控制、可靠性等关键技术,研制新一代超级计算系统,保持和发展国际领先的超级计算研制能力,大力发展高端计算应用,支撑解决国家发展面临的重大战略性应用

问题。

在定位超算中心的“航母”的战略地位基础上,重点分析国内外知名超算中的核心应用现状,解析我国国家级超算中心发展所面临的挑战,进而提出“地方政府主导下”超算中心核心业务的发展建议,从而为我国超算中心的发展提供有益启发。

2 国内外超算中心核心应用的现状分析

2.1 核心业务

衡量一个国家在超级计算领域的实力,速度并非唯一指标,更重要的是看应用能力。国内的一些超算中心投资不菲,在成立时宣传得轰轰烈烈,可没过几年就销声匿迹了。还有一些超算中心虽然看似光鲜依然,却艰难度日^[4]。其原因在于这些超算中心的核心应用拓展不够,并没有产生预期功效。文献[3]用“高速公路上跑拖拉机”来形容目前我国超算的应用状况^[4,5]。若从“航母”的角度看,这种状况可视作“航母运载沙丁鱼”。造成此状况的原因在于国内超算中心建设主要依托地方政府,地方政府更关心涉及本地企事业单位的高性能计算应用和服务,因此超算中心常被用于简单的应用。这些应用虽然有重要的社会和经济意义^[6],但利用运算速度处

到稿日期:2012-10-19 返修日期:2012-12-28 本文受国家自然科学基金(61170284,61272334),“863”项目(2012BAH09B02)资助。

党 岗 男,博士,副研究员,主要研究方向为高性能计算、虚拟现实技术;程志全 男,博士,主要研究方向为计算机图形学、可视计算, E-mail: cheng_zhiquan@gmail.com(通信作者)。

于前列的超级计算机并非唯一的解决方案,很多情形下其他选择可能更具优势,超级计算的战略作用无法体现。如何让超算中心这艘能力超强的航母生存下来,进而“攻城掠地”,推动技术创新,一直是业内人士研究的课题。

美国三大部门,即承担核武器研制任务的美国能源部(the Department of Energy, DOE)、航空航天巨人——美国国家航空航天局(the National Aeronautics and Space Administration, NASA)和负责协调全美科研资源的美国国家科学基金会(the National Science Foundation, NSF)掌控了全美主要的超算力量。总体而言,美国能源部下属的国家实验室实力最强,知名的超级计算机“美洲虎”(Jaguar)就部署在橡树岭(Oak Ridge)国家实验室。目前世界排名第一的超级计算机“红杉”(Sequoia)被安装在美国能源部所属的劳伦斯·利弗莫尔(the Lawrence Livermore National Laboratory, LLNL)国家实验室。它被用来进行模拟核试验,以延长核武器的寿命。美国国家核安全局局长托马斯·达戈斯蒂诺(Thomas D'Agostino)称,“红杉”等超级计算机提供的潜在计算能力“提升了美国核威慑力量的信心”。美国国家航空航天局下属的艾姆斯(Ames)超算中心主要为 NASA 服务,与其他超算中心相比,少有公开成果发表,但偶一发布,就能让业界惊羡不已。美国国家科学基金会下属的超算中心则与高校紧密结合,常有新创的前沿技术成果。例如,最有名的圣地亚哥超算中心(the San Diego Supercomputer Center, SDSC)“寄居”于加州大学圣地亚哥分校,该中心在 2009 年全部采用最新的固态硬盘(Solid-State Drive, SSD)技术搭建集群,震动了整个业界。该中心也由此获得戈登·贝尔奖(Gordon Bell prize)。

在国内,针对气象等核心应用的几个特定领域的计算中心发展情况良好,但是并不面向社会提供高性能计算服务;而面向社会提供服务的超算中心大多依托地方政府筹建,但是地方又尚未完全拥有运用超级计算机的能力。只有天津、济南、长沙、深圳和上海等地的超算中心能提供良好的高性能计算服务。例如,国家超级计算天津中心自“天河一号”系统^[3]投入使用后,已在石油勘探、高端装备制造、生物医药、动漫设计等方面为几十家用户提供服务。

目前影响我国超算应用的最大问题是软件开发。这个问题不解决,将直接影响到我国超算中心的发展。我国的现状是,大型科学计算的应用软件基本上是依靠进口,我国的超算经费用于应用软件开发的不超过 10%,美国相应的投入资金约为中国的 6 倍^[7]。文献^[7]的观点认为,现阶段并行计算应用开发面临 4 大瓶颈:浮点性能、并行算法、运行稳定和并行协同。解决这些难题要靠两方面支持:一是高性能科学与工程计算的支撑软件架构,即中间件的开发;二是面向科学与工程计算的支撑语言,即并行编程模型和语言的创新。事实上,这些应用基础软件能否发挥功效,还取决于具体应用系统的开发工作能否做出具有国际影响力的创新成果。

2.2 技术创新

《华尔街日报》发表评论《中国超级计算机创新不足》(China's Not-So-Super Computers)^[8]指出:“尽管中国在研发方面的支出仍远远落后于美国,却已经取得了长足的进步。但是经过仔细观察就会发现,中国的超级计算机项目并没有像许多人认为的那样对美国在技术上的领先地位造成威胁。中国的超级计算机工程没有产生能够创造新产业的技术突

破。相反,政府部署这些工程的目的仅仅是想帮助国家从医疗保健、汽车设计到航空领域追上欧美。这样做对经济发展至关重要,但是也提醒人们,中国仍然是发展中国家,其主要目标仍然是缩小同更富裕国家在经济和技术上的差距。”客观地讲,该评论是相当公允和深刻的,非常值得我们深思。该报道称,中国的战略“从来不是引领,而是跟随。因为这是经济效益最高的发展方式”,以追随为目的并不能称得上是真正的科学或工程研究,经济效益最高并不意味着科技创新水平最高。迄今为止,还没有哪个以追随为目的的经济发展模式取得了最高的效益。尖端技术的研究风险很高,但可以带来巨大回报,使竞争对手望尘莫及。在该报道中,美国俄勒冈大学的中国科学政策专家理查德·萨特梅尔(Richard Suttmeier)认为,中国还没有弄清楚开拓新技术的正确方法,部分原因是研究人员的待遇取决于他们发表学术论文的数量,而不是他们工作的质量和创新。他指出,美国(对中国)的警惕性并非总是有根据的,对美国而言,“重要的一点是基于创新结果不断制定战略,以便使美国在竞争中遥遥领先。”这些看法应该促使我们对超级计算在技术创新中的作用进行认真的反思。

3 我国超算中心面临的挑战

在技术创新和专用人才储备都不足的局面下,如何实现超算中心的良性运营和可持续发展,是超算中心面临的挑战。实现重点领域的核心应用技术创新,对确保超算中心在短时期内聚集起高端用户群非常关键。

超级计算机系统的使用寿命大概为 5 年,根据 TOP500 的统计,对任何一套系统来说,从第一名落到第 500 名要经历 6~8 年,而再落到跟笔记本电脑一样的水平,只要经过 8~10 年。高性能计算系统技术更新周期短,这意味着高性能计算机用户会面临一个非常残酷的现实问题:如果在这 5 年里机器使用率不高,时常处于空闲状态,就无法收回投资!用户的需求永远是建设超算中心的出发点和落脚点,满足用户需求尤其是用户所涉及的重大行业需求是彰显超算中心实力,确保地方政府对中心巨大投资得到回报的必由之路。

超算中心核心应用的建设目标要合适。一方面,建设目标不能过大。例如,瑞士联邦理工大学科学家提出的“活地球模拟器”^[9]项目想通过搜集大量人类已经掌握的知识和数据,来模拟和预知地球上未来发生的事,并就可能发生的危急情况提前采取相应措施,以避免危机的发生或减小危机带来的灾害。这项技术需要循序渐进,也许还要等上 10 年,“活地球模拟器”的技术才会成熟。“活地球模拟器”这一设想遭到了一部分科学家的质疑。数据分析师瓦登认为这个设想根本就是空想。“首先,因为各种各样的原因,世界上有很多数据是很难获得的。要想让模拟器运行,除非承认只需要部分地球上的数据就能得到正确的结果;其次,仅有大量的数据是不足以建立一个可信的模拟地球的,经济学和社会学的很多理论不是数据能够解释得了的”;另一方面,建设目标也不能过于低端,否则会造成“航母运载沙丁鱼”的局面。

当前某些国内超算中心将云计算(cloud computing)这个被过度炒作的网络服务概念视作其核心建设目标,是否合理很值得探讨。云计算是一种基于互联网的计算方式,主要应用于网络、企业 IT 建设等领域,以服务密集型应用为主。而超算中心主要面向科学计算、工程模拟等领域,属于计算密集

型的应用。不少业内人士认为“云计算中心的建设与应用必将是未来高性能产业发展的重要趋势,是发展的主要方向”,进而建议国内超算巨头迎接云计算时代。但是,云计算并未进入国际知名超算中心(如美国劳伦斯·利弗莫尔国家实验室、阿贡国家实验室、橡树岭国家实验室)核心业务视野内。究其原因,一方面,超算中心的核心应用关注的是计算和存储密集型的大规模科学或工程计算问题,正如“航空母舰上运载的是舰载机”一样,运送“沙丁鱼”则失去了其战略意义和核心价值;另一方面,国际一流超算中心的主流业务不可能由超算系统供应商决定,正如上述国际知名超算中心的应用不是由IBM等计算和服务商所决定的一样。因此,以云计算中心作为超算中心的建设目标,其合理性存在疑点。

大型超级计算中心工程耗资巨大,运行开销昂贵。上海超算中心、深圳超算中心、天津超算中心等一年的电费均是以千万计算,加上软硬件的维护等开销,这些巨大的成本迫使我们必须考虑如何实现超级计算中心长期可持续发展,如何发挥最大效益。尤其是在市场经济环境下,采用什么样的技术机制、合作形式和保障措施,使处于产业领域最前沿的广大企业用户与超级计算中心形成合作共同体,将传统超级计算中心“请企业用户来”的不利局面扭转为“不仅企业用户主动来,而且企业用户与超算中心分享发展利好”,实现这一点最大的突破在于超算中心运营机制和模式创新。如何借鉴、探索和引入新的发展技术思路、超算市场策略,提高中心用户尤其是产业用户的参与度,甚至将部分中心用户在某种意义上转变为特定领域应用的运营商,达到中心和用户“共命运、谋发展”的效果,这是形成超算中心的后发优势和鲜明特色的运营模式基础。

综上所述,国内超算中心面临的是如何创建国际一流超算中心和服务地方经济建设的挑战,应探讨如何利用超算资源提高核心应用的科技创新水平,促进超算中心应用能力的提升,加强超算资源与科学研究、工程开发、政府投资、应用创新的综合集成和互相协调,推动超算中心计算与存储资源的深入开发和高效利用。超算中心只有在应用拓展、运营、人才、市场和产业等诸多方面实现创新,才有可能在服务地方经济建设和迈向国际一流的方向上有所突破。

4 超算中心核心应用的几点建议

深入开展重点领域应用先期开拓,对确保超算中心在短时期内聚集起高端用户群非常关键。用户需求永远是建设超级计算中心的出发点和落脚点,满足用户需求尤其是重点用户所涉及的重大行业需求是彰显超级计算中心实力,确保政府对中心巨大投资实现回报的必由之路。

超算中心核心应用除了传统的核模拟、气象模拟、宇宙天体探索、基因研究、飞行器设计之外,现在也越来越多地应用在自然灾害预报、生物医药研制、新型材料开发、汽车工业设计、动漫等领域。超算中心核心应用的建设,必须基于“地方政府投资、立足本地”的前提,追求国际一流的应用技术创新。我们觉得在以下几个方面,超级计算将有可能大显身手,为技术创新开拓新的疆域。

4.1 三维地理信息系统

数字三维城市系统利用计算机图形学与数据库技术来采集、存储、编辑、显示、转换、分析和输出地理图形及其属性数据,形成土地信息、灾害处理、管线成图与信息管理等模块。但因地理信息获取能力不足,又缺乏高效的海量空间信息处理与综合能力,加之三维地理模型数据出现爆炸性增长,甚至远远超出了当前的通用计算机的内存容量,使得运算时这些数据无法一次装载进内存。这类超大规模三维模型为海量(复杂)三维模型,给整个图形处理流水线带来了前所未有的挑战。而超级计算机可以提供这类海量三维模型的一次性载入或大规模载入的能力,更符合非一致性内存访问(Non-Uniform Memory Access, NUMA)体系结构的编程思想,为该问题的求解提供了硬件基础。

4.2 医疗诊断

案例对医疗诊断至关重要,可以借助超算,利用海量案例库来实现医疗“诊断”,而且置信度与可靠性很高,具备实用价值。今天的医学扫描技术在短短几秒内,就能从一个患者身上得到上千张透视图和几百万兆字节的数据。但是,医生如何从中分析出有用的信息呢?在TED^[1]哥德堡演讲上,科学可视化技术专家安德斯·因纳曼(Anders Ynnerman)^[11]向我们展示了一些用于分析海量数据的新技术。三维医学图像可视化技术作为有力的辅助手段,能够为用户提供具有真实感的三维医学图像,便于医生从多角度、多层次进行观察和分析,并且能够使医生有效地参与数据的处理分析过程,在辅助医生诊断、手术仿真、引导治疗等方面都可以发挥重要的作用。在这方面,无论海量数据库的管理和检索还是海量医学影像的可视化处理都为超算开辟了用武之地。

2011年的基因原始数据量已达到六万亿兆字节(60PB=6×10⁴TB)。自从人类基因组计划完成以来,个人基因组测序即将普及,更多物种的基因组数据即将被大面积采集,可以预期,一个更大规模的数据存储问题即将展现在我们面前。除了基因组之外的其它生物组学数据、生物物质和化学物质的分子谱学数据等对超算提出了更大的需求。

4.3 立体显示

立体显示技术是指能够提供符合立体视觉原理的具有深度信息画面的媒介技术^[12]。该技术的出现是视频显示技术从黑白到彩色、从模拟到数字、从标清到高清后的又一次重大变革。随着立体显示技术的不断发展与更新,立体显示产业于2010年开始腾飞,并进入快速发展轨道。据美国市场研究机构Display Search公司预测,立体显示产业的全球市场销售规模将由2008年的9.02亿美元飙升至2018年的220亿美元。从细分市场来看,立体电视、广播、电影和立体面板是立体显示领域中的4大市场。市场调研发现,立体显示是高投入、高产出产业。在所有投入中,最大的部分依旧是高性能计算环境,接近总投入的60%,其次需要投入的是高速信息编码和传输、显示介质。

4.4 智能监控

当有成千上万个摄像头遍布街头,昼夜不停地录制影像

(下转第166页)

^[1] TED是一个非营利组织,宗旨是“传播有价值的想法”,建立于1984年。TED通过会议将技术(Technology)、娱乐(Entertainment)和设计(Design)三方面的人士聚集在一起,故称为TED。其组织的项目称为TEDx。

评价最低且与合法节点有明显的差距,则其就会被发现并排除出网络。

图5表示网络中某网格内恶意节点数目占60%时,该网格内传感器节点感知到的事件情况,实验设计传感区域每个时间间隔同时发生一个温度为100的事件和一个温度为80的事件,恶意节点每次都广播自己感知到的温度是50,以提高相互间的数据一致性,实现联合攻击。由图看出,在前30s内,DC-TEMC方法没有排除恶意节点,反馈的信息是恶意节点广播的温度为50度的信息,一旦将恶意节点排除,每个时间间隔反馈的事件信息就比较准确了。而TEM方法在恶意节点联合攻击时,排除恶意节点的时间更慢,直到60s后才不受恶意节点的影响,因为恶意节点相互之间有较高的数据一致性,所以TEM方法中恶意节点的信任评价降低较慢,降低到低于阈值 T_{min} 的时间就比图4所用的时间久。由此看出,DC-TEMC方法在解决多个恶意节点联合攻击的情况时能准确迅速分辨出合法节点和恶意节点,并将恶意节点排除出网络。

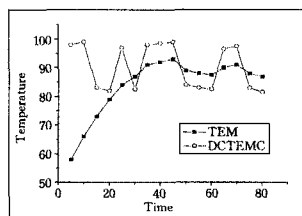


图4 单个恶意节点攻击时温度情况

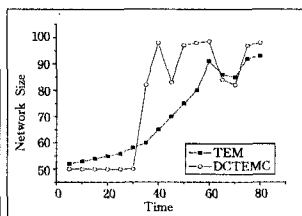


图5 多个恶意节点联合攻击时温度情况

(上接第135页)

产生出海量的视频数据时,你可想到这需要多大的监控管理平台,处理多少海量的数据吗?由于人手有限,大量的摄像视频无人监视,在规定期限无人查询就会被自动删除。理想情况是一旦有重要事件发生,系统就可以实时检测,警方可以快速响应,事后能快速查找到线索。如今,借助计算机强大的数据处理能力可以实现通过过滤掉视频画面中无用的或干扰信息,自动识别不同物体,分析抽取视频源中有用的关键信息,快速准确地定位事故现场,判断监控画面中的异常情况,并以最快和最佳的方式发出警报或触发其它动作,形成事前预警、事中处理、事后及时取证的全自动、全天候、实时监控的智能系统^[13]。

结束语 基于“更强”的视角,超级计算机被本文视作计算机技术海洋中的“航母”,是国家综合国力的重要体现,是国家创新体系的重要组成部分。由于投资模式、运作方式、应用导向等因素不同,我国国家级超算中心与国际知名超算中心具有很大的差距。为了实现在核心应用方面创国际一流超算中心的建设目标,需要将核心应用关注到计算密集型、数据密集型的“高、精、尖”方面,以免造成“航母运载沙丁鱼”的尴尬。

参考文献

- [1] Meuer H, Simon H, Strohmaier E, et al. TOP500 Supercomputer Site[OL]. <http://www.top500.org>
- [2] 张云泉,孙家永. 2011年中国高性能计算机发展现状分析与展

结束语 本文提出了一种适用于多事件并发的无线传感网数据一致性信任评价方法,该方法在多事件并发时不会把合法节点当作恶意节点处理,且在有恶意节点攻击时能准确区别合法节点和恶意节点,并及时将恶意节点排除出网络,另外本方法还能抵御恶意节点的联合攻击。

参考文献

- [1] Karhik S, Vanitha K, Radhamani D G. Trust management techniques in Wireless Sensor Networks; An evaluation[C]//International Conference on Communications and Signal Processing. 2011;328-330
- [2] Lopez J, Roman R, Agudo I, et al. Trust management systems for wireless sensor networks; Best practices[J]. Computer Communications, 2010, 33; 1086-1093
- [3] 荆琦,唐礼勇,陈钟. 无线传感器网络中的信任管理[J]. 软件学报, 2008, 19(7); 1716-1730
- [4] Hur J, Lee Y, Yoon H, et al. Trust evaluation model for wireless sensor networks[C]//The 7th International Conference on Advanced Communication Technology. 2005; 491-496
- [5] 刘方圆,严斌宇,张永齐,等. 无线传感器网络的信任模型研究[J]. 计算机测量与控制, 2011, 19(5)
- [6] Yan Z, Zhang P, Virtanen T. Trust evaluation based security solution in Ad Hoc Networks[C]//NordSec 2003 Proceedings of the Seventh Nordic Workshop on Secure IT Systems. October 2003
- [7] 望[J]. 高性能计算机发展与应用, 2012, 38(1); 2-8
- [8] 国家超级计算天津中心 [OL]. <http://www.nscctj.gov.cn>, 2012
- [9] 林新华. 应用和人才,非此即彼[J]. 中国教育网络, 2010, 6; 21-22
- [10] 赵毅,朱鹏,迟学斌,等. 浅析高性能计算应用的需求与发展[J]. 计算机研究与发展, 2007, 40(10); 1640-1646
- [11] 国家超级计算深圳中心 [OL]. <http://www.nscsz.gov.cn>, 2012
- [12] 莫则尧. 千万亿次并行计算研究与应用 [C]//全国高性能计算学术年会. 2011
- [13] Davis B. China's not-so-super computers [M]. The wall street journal, 2012-03-24
- [14] Living Earth Simulator Project. FuturICT homepage[OL]. http://www.futurict.eu/sites/default/files/docs/files/002-24%20PLATFORM%20Living%20Earth%20Simulator_0.pdf, 2012
- [15] 宋克福. 三维GIS的困境与出路[J]. 中国测绘, 2010(1)
- [16] Ynnerman A. Visualizing the medical data explosion[Z]. TED Speaking. 2012
- [17] 上海科技发展研究中心. 上海“十二五”科技规划重点领域的国内外发展跟踪研究之四: 国内外立体显示技术及产业发展态势分析[C]//上海科技发展研究. 2011
- [18] 计世资讯. 2010年中国智能视频监控市场报告[R]. 2010年中国国际信息通信展, 2010