http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0715

从网格到“东数西算”:构建国家算力基础设施

钱德沛*, 栾钟治, 刘轶
(北京航空航天大学 计算机学院, 北京 100083)

摘 要: 简要回顾了几十年来计算机使用方式的变迁,介绍了基于网络计算技术的国家高性能计算基础设施 CNGrid 的设计与实现。讨论了在“东数西算”战略工程背景下中国算力发展的新趋势,以及国家算力基础设施发展面临的新的技术挑战,并对中国未来超算应用生态和算力基础设施建设提出了展望。

关 键 词: 高性能计算; 基础设施; 网格计算; CNGrid; 东数西算
中图分类号: TP311; TP316; TP391
文献标志码: A **文章编号**: 1001-5965(2022)09-1561-14

1 计算机使用模式的演变

计算机是20世纪人类社会最伟大的发明之一,它的出现,彻底改变了人类生活、工作的面貌。计算与模拟和理论与分析、实验与观测一起,成为人类认识客观世界、开展科学研究的重要手段。对现代科学研究而言,计算的重要性不言而喻。利用计算,可以更清晰地揭示客观世界的发展规律,探索和预测未知的事物。例如,大数据处理分析、人工智能大模型训练和推理、新能源及其利用、新材料设计、工业产品创新设计、创新药物研发、精确天气预报、全球气候变化预测、社会治理和决策支持等,都依赖计算机的强大算力。因此,算力已经成为一个国家创新能力和综合国力的体现。

伴随算力的提高,如何能更容易地使用计算机,便捷地获得所需的算力,也是人们一直追求的目标。计算机诞生70多年来,其使用方式一直在不断变化。早期,人围着计算机转,用户要跑到专门的机房去上机,计算机被一个用户单道程序所独占。随着操作系统的进步,计算机的使用方式逐渐从单道程序的主机模式向多道程序的批作业模式[1]和分时交互模式[2]发展。在分时计算系统[3]中,众多用户可以通过终端,同时使用一台计算机,每个用户在分配给他的时间片内使用计算机,但感觉上似乎是在独占这台计算机。20世纪80年代,个人计算机的出现使计算机进入千家万户,但是个人计算机的性能有限,孤立的个人计算机难以胜任大的计算任务。直到20世纪另一个伟大发明计算机网络的出现和普及,带来网络计算的新变革,计算机的使用方式才发生了影响更深远的变化。所谓网络计算,就是通过网络连接网上分散的计算机,汇聚网络连接的各类硬件和软件资源,形成能力更为强大的计算系统。用户可以通过网络随时随地访问计算机,使用计算资源,完成自身任务,而无需关心计算资源的物理所在。

美国的超算中心联网是网络计算系统的早期范例。20世纪80年代中期,美国国防用途的Arpanet进入商用,Internet诞生。在美国国家科学基金会(NSF)支持下,建设了主干速率56 Kbps、运行TCP/IP协议的NSFNET网络。NSFNET将美国加利福尼亚大学的圣地亚哥超级计算机中心(SDSC)、伊利诺伊大学的国家超级计算应用中心

收稿日期: 2022-08-02; **录用日期**: 2022-08-18; **网络出版时间**: 2022-08-23 09:07
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20220822.1844.002.html
* **通信作者**. E-mail: depeiq@buaa.edu.cn

引用格式: 钱德沛, 栾钟治, 刘轶. 从网格到“东数西算”:构建国家算力基础设施[J]. 北京航空航天大学学报, 2022, 48(9): 1561-1574. QIAN D P, LUAN Z Z, LIU Y. From grid to “East-west Computing Transfer”: Constructing national computing infrastructure[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(9): 1561-1574(in Chinese).

(NCSA)、康奈尔大学的康奈尔国家超级计算机研究室(CNSF)、匹兹堡超级计算机中心(PBC)、冯·诺依曼国家超级计算机中心(JVNNSC)和美国国家大气研究中心(NCAR)的科学计算分部连接起来,对大学和科研用户提供可远程使用的计算资源[4]。

20 世纪 90 年代中期,网格计算(grid computing)的概念在美国兴起[5]。网格(grid)一词最初是指电力网(power grid),网格计算借用电力网的概念,提出要利用高速互联网把分布于不同地理位置的计算、数据、存储和软件等资源连为一体,通过调度、管理和安全保障机制,建立一个像电网一样的计算网格,把算力像电力那样输送给终端用户,支持共享使用和协同工作[6]。在美国 NSF 的支持下,分别由 NCSA 和 SDSC 牵头,实施了 2 个网格计算项目,初步建立了计算网格的雏形[7]。网格计算研究在 20 世纪末到 21 世纪最初 10 年达到高潮。在美国的倡导下,成立了全球网格论坛 GGF,与此对应,国际 IT 大公司联合成立了企业网格论坛 EGF。2006 年,GGF 和 EGF 合并,成为开放网格论坛 OGF。GGF 提出了开放网格服务基础设施 OGSI 和开放网格服务体系架构 OGSA 等标准[8-9],协调全球网格计算的研究和开发力量,研究资源管理,安全、信息服务及数据管理等网格计算基本理论和关键技术。在 Globus 项目[10]支持下研发了 Globus Toolkit 3.0 (GT3) 软件,GT3 作为 OGSI 的一个完整的参考实现,成为网格计算的事实标准。

在网格计算热潮中,美国、欧盟、日本、中国都实施了一批网格计算研究计划或项目。部分代表性项目如表 1 所示。

表 1 世界部分网格相关研究计划

Table 1 Part of grid-related programs in the world

国别	网格计算相关研究计划/项目
美国	TeraGrid, XSEDE
欧盟	EGEE, EGI
英国	UK e-Science
日本	NAREGI, HPCI
韩国	K* Grid
中国	CNGrid, ChinaGrid

美国是网格计算的发源地,该方向的研究计划持续时间最长,实施的项目数量最多。美国 NSF 专门设立 Cyberinfrastructure 部门,持续稳定支持网格计算方向的研究。美国的网格项目主要有2 类。第 1 类由美国 NSF 支持,在先进计算伙伴计划 PACI 之后,从 20 世纪末开始实施 TeraGrid 项目[11],其主要目标是用网格计算技术推动国家科技进步,保持美国的科技领先地位。2011 年,TeraGrid 的后继项目 XSEDE 项目(https://www.xsede.org/)正式启动,该项目旨在连接全球的计算机、数据和研究人员,建立可供科学家共享的计算环境。美国 NSF 资助的开放科学网格(OSG)在其基础软件 HTCondor(https://htcondor.org/)支持下,实现了众多大学与国家实验室的计算资源共享,为科学家提供了科学计算的环境。第 2 类网格项目由美国国防部、能源部等支持,其主要目标是更好地完成本部门的任务。2 类研究的应用目标有所不同,但共同点是要发展先进的基于网络的应用基础设施,实现应用层面的互联互通、资源共享、协同工作。

欧盟于 2000 年和 2001 年分别启动了欧洲网格计划(EuroGrid)(https://www.eurogrid.org/)和欧洲数据网格计划(European DataGrid)[12]。在欧洲数据网格计划的基础上,2004 年 3 月,欧盟框架研究计划启动了 EGEE 项目[13],其目标是基于网格技术开发欧洲的服务网格基础设施,供科学家全天候使用。2011 年,欧盟框架计划又启动了 EGEE 的后继项目 EGI(https://www.egi.eu/)。在这些项目支持下,研发了欧盟的网格中间件 gLite (http://glite.cern.ch/),建立了可持续运维的泛欧计算基础设施。

英国的网格研究计划是 UK e-Science[14],其目标是用网格技术改变科学研究的模式,推动科学技术的进步,长远目标是影响未来的信息技术基础设施。在 UK e-Science 计划支持下,英国在大学和研究机构建立了一批国家 e-Science 中心,依托 OMII-UK 项目研发了英国的开放网格中间件,开发了一批面向 e-Science 的网格应用系统。

日本文部科学省(MEXT)在 2003 年启动了“国家研究网格基础设施”项目 NAREGI[15]。NAREGI 构建在日本教育科研网 SuperSINET 之上,旨在研制并部署面向科学研究的网格基础设施,并参与全球开放网格组织 OGF 的工作,为网格的标准化活动提供支持。在 NAREGI 之后,日本政府又结合 E 级超级计算机的研制,启动了日本高性能计算基础设施项目 HPCI。HPCI 通过 SuperSINET 连接日本大学和研究机构中的 10 个大超算中心和 2 个大数据中心,形成日本的国家级计算基础设施。

中国的网格计算研究起步于 20 世纪 90 年代末,科学技术部(以下简称科技部)是支持网格计算研究的主要政府部门,从 1999 年起,中国在高

性能计算和网格方向连续实施了多个国家 863 重大项目和国家重点研发专项,表 2 列出了科技部在该方向支持的主要项目。在这些项目的持续支持下,研发了国家高性能计算环境系统软件 CNGrid GOS 和 CNGrid Suite,使用环境系统软件,聚合了分布在全国各地近 20 个超算中心和高性能计算中心的计算资源,实现了资源的互联互通与统一共享、作业的提交与全局调度、数据的全局管理和环境的安全管控,在此基础上,成功构建了基于网格/网络计算技术的国家高性能计算环境——中国国家网格服务环境 CNGrid(参见 http://www.cngrid.org)。CNGrid 历经 20 余年的发展,正从“可用”迈向“好用”,目前的聚合计算能力超过 50 亿亿次,存储容量近 500 PB,部署了 600 多个应用软件和工具软件,支撑了数千项国家科技计划项目和重要工程项目的研究工作,用户覆盖基础研究、工业设计、能源环境和信息服务等众多领域,极大促进了中国科技创新能力的提高,已经成为科学研究、技术创新、工程设计中不可或缺的新型信息基础设施。

表 2　中国科技部的网格和高性能计算项目

Table 2　Grid and high performance computing projects under the Ministry of Science and Technology of China

项目来源	项目名称	执行周期	项目成果
国家 863 重大课题	国家高性能计算环境	1999—2000 年	4 000 亿次曙光 3000;包含 5 个高性能计算中心的国家高性能计算环境原型
国家 863 重大专项	高性能计算机及核心软件	2002—2005 年	11.2 万亿次曙光 4000,5.36 万亿次的联想深腾 6800;国家高性能计算环境实验床“中国国家网格 CNGrid”,8 个结点,18 万亿次计算能力;一批网格应用
国家 863 重大项目	高效能计算机及网格服务环境	2006—2010 年	4 700 万亿次的天河 1A,3 000 万亿次的曙光 6000,1 071 万亿次的神威蓝光;具有服务特征的国家网格服务环境 CNGrid,11 个结点,8 000万亿次计算能力;一批网格和高性能计算应用
国家 863 重大项目	高效能计算机及应用服务环境	2011—2015 年	12.5 亿亿次的神威·太湖之光,10 亿亿次的天河 2A;以服务支持应用的国家高性能计算环境 CNGrid,14 个结点,20 亿亿次计算能力;一批高性能计算应用
国家重点研发专项	高性能计算	2016—2021 年	E 级计算机;初步具备基础设施形态的国家高性能计算环境 CNGrid,19 个结点,52 亿亿次计算能力;一批高性能计算应用

2006 年兴起的云计算是网络计算技术与应用模式的一次大变革。与以往由学术界主导的技术热潮不同,云计算从开始就是由 IT 公司提出并引领的。2006 年 3 月,亚马逊公司推出弹性计算云 EC2(http://aws.amazon.com/ec2),2006 年 8 月,谷歌公司首席执行官埃里克·施密特在搜索引擎大会首次提出“云计算”的概念。此后,微软、戴尔、IBM 等国际 IT 巨头和百度、阿里等中国互联网公司都纷纷跟进。在学术界,美国加利福尼亚大学伯克利分校的 Armbrust 等也专门发文,阐述云计算的学术问题[16]。几年之内,云计算已从新兴技术发展成为全球的热点技术。云的资源被虚拟化,可以动态升级,资源被所有云计算用户通过网络方便地使用。云计算的出现改变了 IT 应用系统部署运行的方式。在传统 IT 应用模式下,应用部门需要自行采购计算机硬件和软件,在私有的计算系统上安装部署自己的应用软件,运行和维护应用系统。在云计算模式下,用户无须自行采购维护计算机,而是从云服务商那里租赁所需的计算资源,在云中安装特定的应用软件,存放应用的数据,完成应用系统的部署,应用系统就能够运行在云端。应用部门本身不需要采购和维护私有的计算机,当应用需求变化时,可以根据需要增加或减少租赁的云计算资源。服务和按用付费是云计算的商业模式,是计算向基础设施形态迈出的一大步。根据所提供的服务内容,云计算可分为 IaaS(提供基础资源)、PaaS(提供平台服务)和 SaaS(提供应用软件)[17]。根据服务的范围和应用的性质,云计算又可分为公有云、私有云和混合云[18]。按照服务封装部署方式又可分为虚拟机、容器、裸金属服务器等[19]。今天,几乎所有大数据中心都在某种程度上使用云计算技术,提供云服务。云计算技术也被引入传统的高性能计算领域,出现了以云方式运行超级计算中心的“云超算”和提供高性能计算能力的“超算云”[20]。

中国的云计算和国际同步发展。国家 863 计划在 2010 年就启动了“中国云”重大项目,支持阿里、百度等互联网公司研发云计算系统。“十三五”期间实施了“云计算与大数据”重点专项,更加系统全面地推进云计算关键技术和系统的研发与应用。今天,阿里云、华为云、百度云、浪潮云等已经在国内市场举足轻重。

物联网(IoT)、基于移动互联网应用的蓬勃发展催生了边缘计算[21]。边缘计算的目的是使应用程序、数据和计算能力(服务)更加靠近端用户,而不是更靠近集中的云,这样就能减少数据的移动,降低数据传输的延迟,降低端系统和数据中心之间的传输带宽需求,达到更低的成本和更好的用户体验的效果。随着边缘计算技术的进步,云-边-端融合的 IT 应用模式也日趋流行,成为渗透更宽广应用领域的网络计算的新形态。

2 国家高性能计算基础设施 CNGrid

建设国家级高性能计算基础设施是创新型国家建设的战略需求。基于网络计算的计算基础设施具有如下特征:①动态性。系统的状态和行为动态变化,资源动态接入和退出、设备随时会出故障、网络可能拥塞甚至断开、用户的数量会不断变化等。②自治性。地理分散的资源在支持广泛共享的同时,仍能保持原有的隶属和管理属性。③开放性。硬件、软件和服务来自不同的厂商,由不同的团队开发,遵循不同的技术规范,兼容并蓄,形成自然生长演化的计算生态环境。这种动态、自治、开放的基础设施不同于资源集中拥有和控制的云计算环境。

在开放、动态的互联网环境下,聚合网上异构、自治的分散资源,构建在全国范围共享使用的国家高性能计算基础设施,面临重大技术挑战:①在动态环境下如何应对系统资源的不确定性,对用户提供稳定的高质量服务;②在不改变原有资源隶属关系和管理模式的条件下,如何实现受控共享;③在开放异构的环境下,如何高效开发和运行大规模分布并行应用,建立高性能计算应用的生态环境。国家高性能计算基础设施 CNGrid 通过体系结构、系统软件、应用模式、应用开发与优化技术等创新应对上述挑战,为在中国形成高性能计算资源提供、应用开发和运行服务的完整产业链奠定了技术基础。

2.1 非集中层次虚拟化体系结构及系统软件

针对开放动态环境下分布异构资源的统一管理与受控共享、系统安全及服务质量保障等重大技术难题,设计并实现了国家高性能计算基础设施“三横两纵”的非集中层次虚拟化体系结构,如图 1 所示。“三横”是指自底向上的内核系统层、系统服务层和应用层。内核系统层通过资源实体和虚拟组织等抽象,将地理分布、自治的高性能计算物理资源抽象和聚合为可动态划分、申请和调度的虚拟资源,通过运行时虚拟地址空间和自治安全策略,解决资源视图、资源发现及定位、异构资源统一访问等基础性问题。系统服务层通过访问虚拟资源,以服务化形式向上层应用提供作业管理、数据访问与传输、应用编程、用户映射等功能。应用层使用系统服务层提供的功能,实现应用的业务逻辑,服务最终用户。“两纵”是贯穿内核系统层、系统服务层、应用层 3 个层次的环境监控管理和安全机制,保障环境的可管理性和安全性。

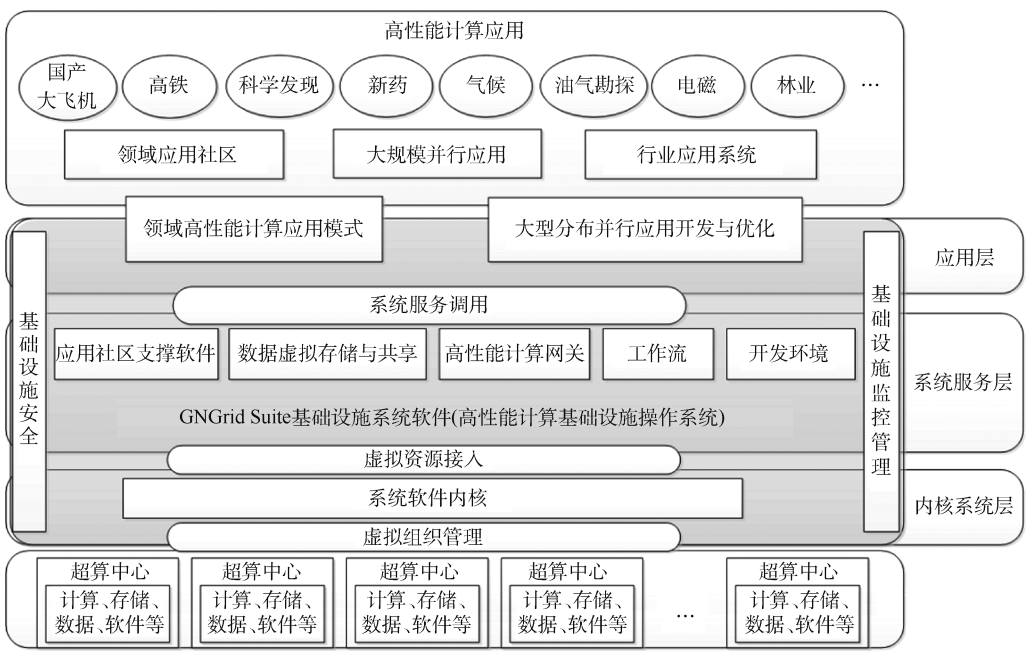

图 1 国家高性能计算基础设施的非集中层次虚拟化体系结构

Fig. 1 Decentralized hierarchical virtualization architecture for national high performance computing infrastructure

体系结构的非集中是指 CNGrid 的管控采用地理分布模式,即在每个 CNGrid 结点部署一台运行系统软件的服务器,通过覆盖网络将各个 CNGrid 结点动态组织成星型、网状或混合结构,以适应国家高性能计算基础设施对资源的分层分域管理的需求。

基于非集中层次虚拟化体系结构,研发了基础设施系统软件 CNGrid Suite,其系统架构如图 2 所示。CNGrid Suite 提出并实现了“资源实体”、“虚拟组织”和“运行时虚拟地址空间”等 3 个系统核心抽象,来表达系统中的各种资源要素、要素间的静态关系和运行时的动态关系,通过这些抽象,将分散、异构、无序的计算机硬件资源、软件资源和用户组织成逻辑有序、可受控共享的虚拟资源,支持资源的动态聚合、调度和安全访问。

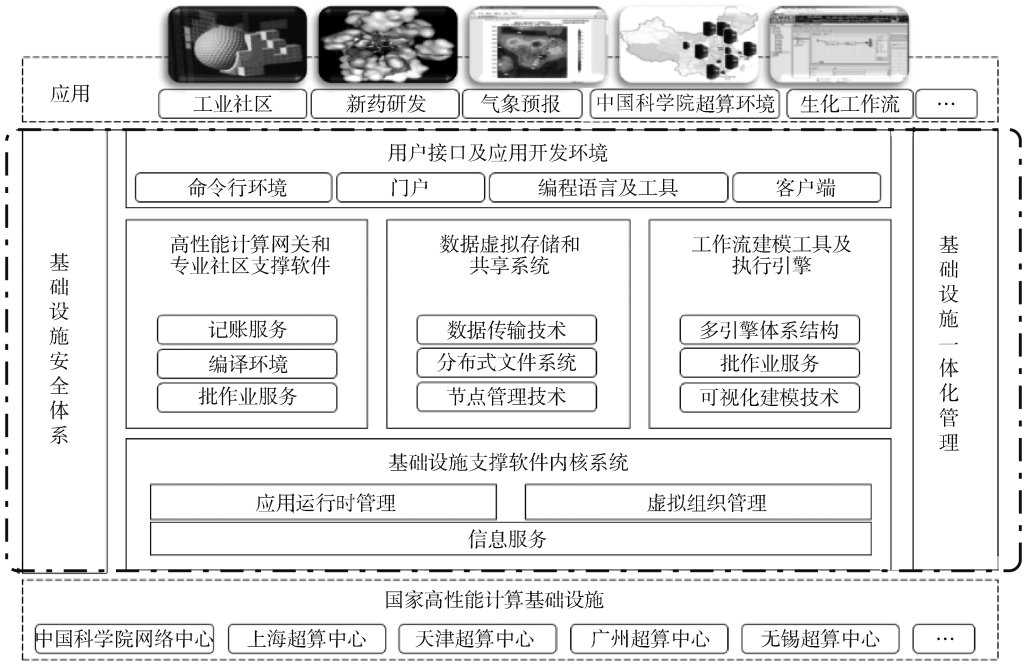

图 2　CNGrid Suite 系统架构

Fig. 2　CNGrid suite system architecture

2.2　资源组织与作业调度

针对国家高性能计算基础设施特点和应用需求,提出了“资源实体”、“虚拟组织”和“运行时虚拟地址空间”等系统软件创新概念,应对资源描述、组织和访问的挑战。

在 CNGrid 中,用户、资源和社区都被统一抽象为资源实体。每个资源实体由一个全局 id 来标识,代表一个可以访问其他实体,也可被其他实体所访问的全局资源要素。系统软件的全局命名管理模块对资源实体实施统一管控,完成资源实体创立与消除、资源定位、资源解耦等功能。

多个相关资源实体可构成一个虚拟组织。虚拟组织描述资源实体之间的静态关系,实现资源实体注册、资源实体接入与剔除、资源实体元信息管理、资源实体权限管理、资源实体访问控制等功能。通过虚拟组织把资源实体组织成可有效管控、相互协同的资源集合。

运行时虚拟地址空间描述了资源实体间的动态访问和调用关系,结合动态绑定的安全策略,解决了资源命名、资源视图、资源发现及定位、资源统一安全访问等基础性问题。与传统操作系统的进程概念相对应,CNGrid 提出了网程(Grip)概念。网程在运行时虚拟地址空间中代表资源实体动态访问其他资源,实施访问控制,分配、管理和回收资源,实现应用的可控启动和终止。

CNGrid 聚合了分布在不同地域、不同组织机构中的各种各样的高性能计算资源,面向用户提供统一的系统映像和透明的作业调度是基本需求。CNGrid 的作业调度由服务端、驱动器和客户端 3 部分组成[22]。面向用户提供统一的访问入口和使用方式。系统软件根据用户作业请求性质为其自动匹配适当的高性能计算资源。另外,也提供开放接口,为作业调度模型和调度策略的优化提供了可能。作业调度核心模块由资源收集器、资源匹配器和资源调度器构成,通过引入多种优先级作业队列,细化作业的系统状态,改善了作业调度策略的可配置性[23]。

2.3 监控管理和安全机制

针对资源的分布性、异构性和动态性,服务和应用的多样性及管理需求的各异性等特征,CNGrid提出了统一实体监控管理、管理功能动态构造、管理功能跨域动态部署及协同工作等创新概念与机制,设计了基于统一实体的监控管理体系架构,研发了一体化的监控管理系统(见图3),实现了CNGrid的资源监控和运行管理,为多层次资源的按需共享和自主协同提供了支撑。

与CNGrid系统软件中的资源实体抽象相对应,设计了基于实体的资源管理信息描述方法,采用统一的"被管对象"抽象建立全局信息模型,对各类资源信息进行有效的建模与表示,形成对不同层次、不同类别资源精确监控和管理的基础。

针对资源的动态性和多样性,提供了监控管理功能的动态生成、部署、运行的能力,支持管理功能的动态扩展和更新,实现监控管理系统的动态构造与演化。针对资源的跨域特点,提出了监控管理功能跨域动态部署的概念。基础设施的监控管理按分布层次式组织,监控管理功能分布在各个监控管理域中,各管理域既局部自治又相互协作。管理域设立自身的监控管理中心,形成多级监控管理中心的协同机制。同时,实现了基于复杂事件处理的监控信息高效获取、传输和控制的机制。这些措施有效减少了与监控管理有关的数据流量,降低了监控管理对基础设施正常应用业务的影响。

针对CNGrid环境下资源种类繁多、数目巨大的现状,设计实现了单维度、多维度及基于日志等多种故障扫描、识别和应对方法,能够准确定位故障,分析故障根因,及时通告故障事件并推荐应对的策略,为提高CNGrid的可用性、可靠性和可管理性提供了保障。

CNGrid的安全机制采用基于证书的身份认证和访问权限控制,系统软件基于代理证书实现用户认证和权限代理。首先定义访问控制的策略构建资源共享操作上下文,操作上下文包含用户在CNGrid中的身份信息(用户的代理证书)、用户所属虚拟组织及其所在组别,以及虚拟组织签发的资源访问令牌。资源提供者能够在虚拟组织中注册资源并对其进行持续的管理,通过向用户分配相应的权限,控制应用占用的资源并有效支持多个应用间的协同。在运行时,由网程维护用户身份并实施访问控制。当用户需要访问资源时,把自己的操作上下文从用户端传送到资源端,基于证书权限验证的结果控制资源的访问[24]。CNGrid中部署了证书的认证中心(CA),用户可通过CA的Web界面申请用户证书。CNGrid安全机制基于公钥基础设施(PKI),使用标准的X.509证书,提供用户和资源的双向认证。

CNGrid的安全机制在权限控制的前提下,尽可能地支持基础设施资源的共享。受控共享是CNGrid提出的一个重要概念。在受控共享机制下,只要访问控制权限允许,非属主用户和属主用户均可完成对资源的操作,此过程称为属主用户和非属主用户对资源的受控共享。

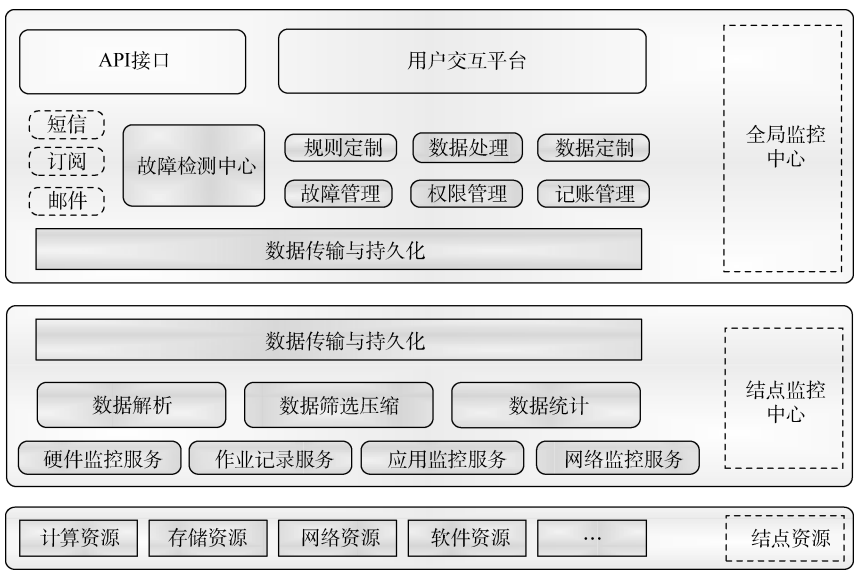

图3 一体化基础设施监控管理体系结构框架

Fig. 3 Architectural framework of integrated infrastructure monitoring and management

2.4　数据管理和高效传输

高效存储和访问分布、异构、自治的数据资源是 CNGrid 要解决的另一个关键问题。针对科学研究和行业应用的实际需求,设计并实现了基于虚拟数据空间的数据管理体系,有效集成了环境中的数据资源,构建了统一的数据管理空间,为用户提供了透明统一的数据存储、访问和管理能力。CNGrid 的数据资源主要包括文件系统和数据库系统。虚拟数据空间为文件系统的集成共享提供虚拟文件系统,为数据库系统的集成共享提供虚拟数据库系统。在这两者之上,虚拟数据空间提供数据基础服务,简化了存储和数据的使用逻辑,为用户或应用提供便利。

针对 CNGrid 中数据分布存储和自治管理等特点,虚拟数据空间采用面向服务的分布式层次结构进行构建。设计了基于分布域的联邦数据存储管理机制,在各数据域的自治管理基础上实现全局统一管理,保障数据管理的可扩展性。通过数据域之间的协作来满足应用的分布式存储需求,系统根据用户的访问位置等信息实现数据资源的就近存储和管理,以便提高用户对数据的访问效率。设计实现了异构数据库的整合机制,以统一的接口实现对不同数据库管理系统的数据访问,并且通过并行机制保障在大规模分布式环境下的访问效率。

影响 CNGrid 中数据传输效率的主要因素包括单次传输的数据量、网络带宽的利用率和传输引入的额外开销。因此,提高效率的关键在于减小单次传输的数据量,充分利用网络带宽,降低传输额外开销。数据传输不可靠的主要原因是网络链路及主机的不稳定,增强可靠性的关键在于克服不稳定因素,减少数据传输错误造成的损失。CNGrid 通过多个副本并行传输来提高带宽利用率,通过文件的分块传输来减少每次传输的数据量,通过就近传输来提高传输速度和可靠性,提供断点续传和三方传输来提高数据传输效率,减少额外开销。

CNGrid 环境的动态变化特性使数据存储资源难以保证持续的服务。CNGrid 引入数据副本管理机制来保证数据服务的可靠性。数据副本的引入带来数据一致性维护问题。为此设计实现了并行化的一致性有限状态机,有效降低了数据一致性维护的代价。数据按照其使用频度被定义为冷热数据,系统根据数据温度动态调整其副本数量,在提高访问效率的同时减少了不必要的开销。此外,还设计实现了位置及网络实时状态感知的数据副本放置策略,在保证一致性的同时提高了数据访问的效率。

CNGrid 的数据管理服务通过灵活和自适应的数据访问授权控制,解决了数据安全性与环境复杂动态性之间的矛盾。其采用细粒度的访问控制策略,为不同的资源拥有者和使用者对不同粒度的数据资源的访问,提供个性化的访问控制策略,满足了自治性和个性化的要求。

2.5　基于应用社区的应用新模式

针对以公共计算平台支撑个性化领域应用的需求,CNGrid 提出了体现领域应用特点的个性化领域应用社区概念。应用社区具有“批零”结合的资源管控与按需服务机制,既有网格聚合分散资源的能力,又有云计算集中管控、按用付费的特点,成为国家高性能计算基础设施的应用新模式。为了支撑应用社区的构建和运行,发展了领域应用中间件 Xfinity,实现了多层次的用户管理机制、按域划分的资源管理模式、基于模板的应用零开发热部署技术及资源动态绑定的工作流技术等体系架构和关键技术创新。

按需定制的服务模式体现在服务方式和内容的定制,可为不同用户定制满足其特定需求的专用社区。通过社区动态配置、资源动态绑定与复用、应用按需集成与动态部署等技术,实现了服务的按需定制。按需付费的交易模式贯穿服务交易全过程。资源拥有者在社区发布资源与价格信息,用户通过社区选择能满足其需求且价格合适的服务。社区监督服务交易过程和服务完成情况,保证交易各方的利益。

为实现按需调配的资源管理模式,提出了权属策略灵活配置的社区资源管理技术,将特定资源组织成资源子域授权给不同用户使用。不同资源子域的用户相互隔离,互不干扰。用户对其资源子域拥有完全的支配权,可做更精细的分级授权管理,实现社区内资源的有效调配和充分共享。

社区通过基于角色的权限访问控制、组管理和双层映射等技术实现了多层次、分角色的用户与资源的精细管理。实现了与企业业务系统相容的低开销安全机制,允许独立制定和修改国家高性能计算基础设施、社区、企业这 3 个管理域的安全机制,在管理域之间建立信任关系和映射机制,消解各管理域不同安全策略间的矛盾。

工业创新设计社区是该新型应用模式的一个实例,其系统框架如图 4 所示。工业社区将国家高性能计算基础设施的计算服务推送到汽车制造、

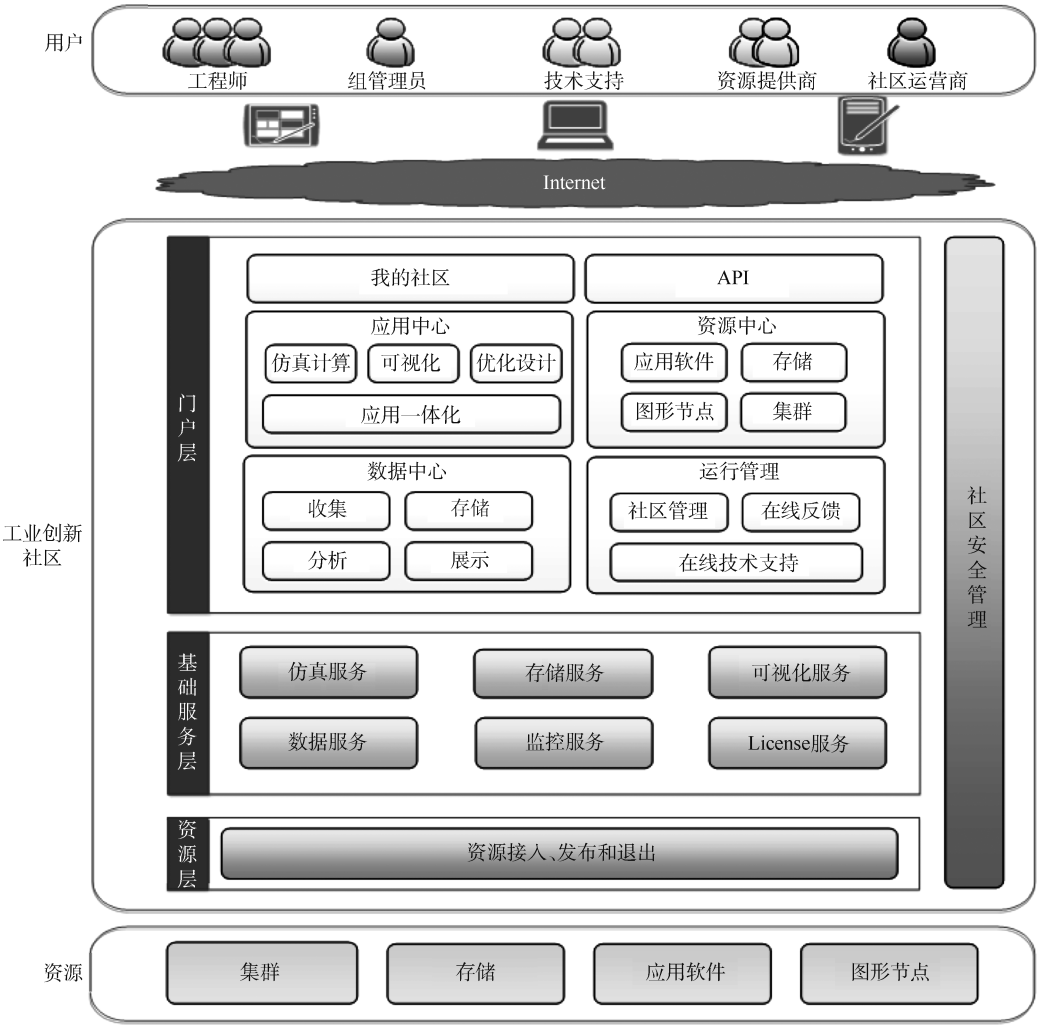

图4 工业社区系统框架

Fig. 4 Industrial community system framework

核电、飞机制造等行业的企业内部,加快了产品设计,降低了研发成本,提高了企业竞争力,取得很好的经济效益。

2.6 分布异构环境下应用软件的开发与优化

CNGrid的地理分布、资源异构的特征给大型应用的开发带来新的挑战。针对大规模应用的流程化与跨结点分布特征,CNGrid突破了构件与流程相结合的工作流编排、部署和运行技术,实现了流程在线组装、即时协作、即插即用的服务适配机制,支持分布资源的动态绑定、应用的快速开发和灵活部署执行,允许领域专业人员以低代码方式开发应用。

针对国产超级计算机多级并行和多种异构的特点,提出并实现了一系列并行程序优化方法和技术。例如,提出了节点间MPI、节点内OpenMP、处理器内多核并行的多级混合并行模式。提出了适用于不同异构平台的区域分解和动态负载平衡方法,通过动态可调的区域划分,实现加速器和通用处理器之间的负载均衡,隐藏加速器和通用处理器间的通信开销。提出了定制缓存及计算/访存重叠技术,充分发挥数据在片内核间的最佳重用。提出以定制DMA传输等方式实现计算和访存的最优化重叠,缓解内存带宽对应用整体性能的限制,大幅提升系统效率。研发了应用级断点保护技术,保证了大规模长时间作业的正确执行。发展了屏蔽硬件细节的并行算法库和编程接口,使不熟悉并行计算的应用领域专家能编写高效的并行应用软件。

在上述关键技术突破的基础上,研发了面向国家高性能计算基础设施的应用集成开发环境。集成开发环境包含基础算法库、应用模块库、程序模板库、优化工具库、拖拽式的工作流编排器、适配多种国产处理器的跨异构结点编译环境等,其系统架构如图5所示。开发人员可以使用集成开发环境中基于模板库的开发向导,自动生成程序代码框架,重用基本算法和模块库中的代码,快速构建应用程序,并在国家高性能计算基础设施中交互式地部署、优化和运行。

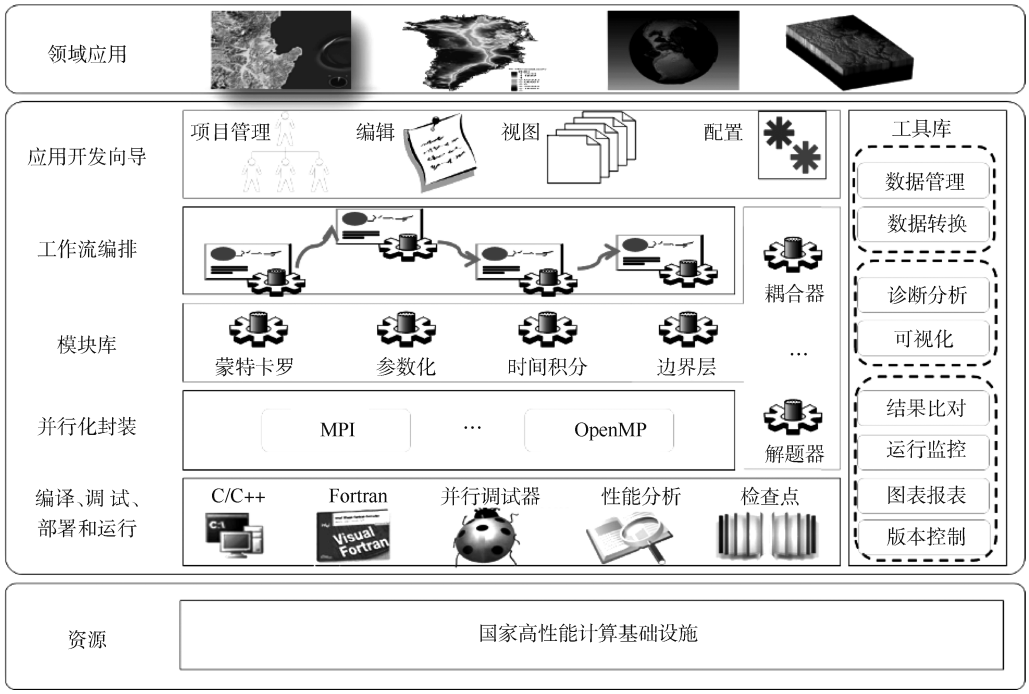

图 5　高性能计算应用集成开发环境系统架构

Fig. 5　System architecture of integrated development environment for high performance computing applications

3　算力基础设施发展趋势与展望

3.1　新兴应用和技术趋势

3.1.1　新兴应用及算力需求

近年来,一系列新兴技术与应用的快速发展对算力基础设施提出了更高的要求。其中,最具代表性的有人工智能、大数据和云计算等。

新一代人工智能的核心驱动力来自深度学习技术。通过对多层大规模人工神经网络进行训练并用于推理,促进了计算机视觉、自然语言处理等领域的突破性进展。虽然人工神经网络概念的出现和应用已有数十年,但之所以近年来才取得快速发展,离不开算力的支持。深度学习是一种计算和数据驱动的技术,由于深度神经网络规模庞大,且通常需要使用大量的数据进行训练,这带来了巨大的计算量。表 3[25] 给出了几种典型神经网络模型的训练计算量。庞大的计算量使得神经网络训练通常需要借助加速部件进行,即使这样,一次模型训练也需要花费数小时到数天时间。

表 3　典型深度神经网络模型的训练计算量[25]

Table 3　Training computations of typical deep neural networks[25]

神经网络模型	应用领域	训练计算量/Flops
AlexNet	图像分类	4.7×10^{17}
VGG16	图像分类	8.5×10^{18}
YOLOv3	图像目标检测	5.1×10^{19}
Transformer	自然语言处理	7.4×10^{18}
GPT-3	自然语言处理	3.1×10^{23}

注:数据来源于 https://docs.google.com/spreadsheets/d/1AAIebjNsnJj_uKALHbXNfn3_YsT6sHXtCU0q7OIPuc4。

大数据和云计算是另一种推动算力设施发展的新兴技术。在虚拟机和容器技术的支持下,人们可以在硬件平台上实现计算资源的灵活划分和隔离,以及软件环境的快速部署,这使得用户可以从云平台获得按需分配、可动态伸缩、易于部署且稳定可靠的硬软件平台和算力服务,这一特性吸引越来越多的用户将网站、业务平台和信息系统等迁移托管到云数据中心。由于企事业单位数量众多,此类应用的需求聚合到一起形成了对算力资源的庞大需求。

3.1.2　体系结构及算力设施的发展趋势

随着集成电路延续数十年的“摩尔定律”减缓并走向停滞,计算机体系结构进入了变革期,多样化的体系结构不断涌现。为了在集成电路规模和性能增长减缓的背景下持续提升应用性能,定制化(customization)成为近年来体系结构发展的一大特点,即通过设计面向不同应用的加速部件/处理器,持续提升应用性能,典型代表有 GPU 和深度学习处理器/加速器。在 GPU 方面,除了 Nvidia GPU 外,传统处理器厂商 AMD 和 Intel 也陆续推出自己的 GPU,国内也已研发出多款自主 GPU 芯片,为人工智能、科学计算、图形/图像处理等应用提供了高性能计算平台;在深度学习处理器/加速器方面,比较有代表性的有寒武纪、

Google TPU、华为昇腾等,这些处理器/加速器专为神经网络计算而设计,其性价比和能效均优于通用 CPU 和 GPU。

在多种新兴应用的推动下,各种处理器/加速器被应用于算力设施中,这带来了以下 2 方面的变化:

1) 算力中心内的异构化。在算力中心内部,异构已成为主流架构。表 4[26] 给出了 TOP500 超级计算机排行榜中前十位高性能计算机的体系结构。可以看出,10 台机器中仅有 1 台(富岳)采用同构架构,其他 9 台均为异构架构,除 CPU + GPU 结构外,中国的神威·太湖之光采用片内异构众核处理器,天河 2A 采用 CPU + 加速器结构。算力中心异构化的另一个体现是面向人工智能应用的异构体系结构,除了 CPU + GPU 结构外,CPU + 深度学习处理器/加速器结构也被智算中心广泛采用,如 CPU + Google TPU、CPU + 寒武纪、CPU + 华为昇腾等。

2) 算力中心的多样化(算力中心间的异构化)。传统的超算中心主要面向科学/工程计算,应用类型以并行数值模拟为主,主要特征是以双精度浮点运算为核心的计算密集型应用。与之相比,人工智能应用的计算类型主要是单精度/半精度浮点和定点运算,而大数据和云计算则以数据密集型应用为主。为了适应这些新兴应用的需求,算力中心的硬件配置也开始出现变化,出现了配置深度学习处理器/加速器、主要面向人工智能应用的智算中心,以及配置大容量内存和网络虚拟化设备、主要面向大数据和云计算应用的云算中心,同时,超算中心也开始支持人工智能和大数据应用。

表 4 TOP500 排名前十的高性能计算机(2022 年 6 月)[26]

Table 4 TOP10 in TOP500 high performance computing systems (June 2022)[26]

排名	系统	处理器/加速器	Linpack 性能/PFlops
1	Frontier	AMD 64C + AMD MI250X	1 102
2	Fugaku(富岳)	A64FX 48C	442.01
3	LUMI	AMD 64C + AMD MI250X	151.90
4	Summit(顶点)	IBM Power + Nvidia V100	148.60
5	Sierra(山脊)	IBM Power + Nvidia V100	94.64
6	Sunway TaihuLight(神威·太湖之光)	Sunway SW26010	93.01
7	Perlmutter	AMD 64C + Nvidia A100	70.87
8	Selene	AMD 64C + Nvidia A100	63.46
9	Tianhe-2A(天河 2A)	Intel Xeon + Matrix2000	61.44
10	Adastra	AMD 64C + AMD MI250X	46.10

注:数据来源于 http://www.top500.org。

3.2 算力基础设施面临的技术挑战

随着“东数西算”国家战略的实施,西部多个算力枢纽将建设算力中心,并面向东部经济发达地区提供算力服务。在这一背景下,如果由各个算力中心单打独斗,分散运营,则算力中心需投入人力物力自行发展用户,容易出现算力中心间的负载不均衡,导致算力碎片化和算力资源浪费;而在用户侧,由于各算力平台的硬件配置、软件资源、服务接口存在差异,也将给用户的软件开发和资源使用带来诸多不便。因此,通过将多个算力中心互联,向用户提供一站式、集成化的算力服务,形成覆盖全国的算力基础设施,对于提升算力资源利用效率和服务水平,促进国产软件和应用生态发展,支撑“东数西算”国家战略具有重要意义。

为了构建算力基础设施,需要解决算力中心异构化和多样化带来的诸多技术挑战,主要体现在以下几方面:

1) 计算任务在异构算力中心间的透明调度。算力基础设施必须具备的一项功能是:用户通过算力基础设施的服务平台提交一个计算任务后,可以直接得到计算结果,无需关心该计算任务在哪个算力中心上运行。这需要按照计算任务的类型和需求确定其所需资源,并根据各个算力中心的硬软件配置及可用资源数量进行任务分配和调度。例如,如果用户提交的是一个使用 CUDA 编写的可执行程序,就要在各算力中心寻找配置 Nvidia GPU 的节点并获取其当前使用状态,在此基础上进行调度;而如果用户只是希望在指定的数据集上完成深度神经网络模型训练,或者只是希望对所设计零件进行结构强度分析,那么计算任务的分派调度就不但要考虑硬件资源,还要考虑是否具备所需的软件资源。在以上基本调度功能的基础上,如果进一步考虑用户的服务质量和计费需求,如限定任务完成时间、限定资费水平等,任务调度要考虑的因素就更多。

2) 如何提供多层次、多样化的算力服务。算力服务的层次与云计算服务相似,也可分为基础设施(IaaS)、平台(PaaS)、应用软件(SaaS)3 个层次。在基础设施层,算力服务以处理器/加速器/计算节点的形式提供;在平台层,算力服务在基础设施之上还提供系统软件和支撑软件,如高性能计算用户需要的 MPI 环境和基础算法库,人工智能用户需要的深度学习框架,大数据用户需要的分布式处理框架等;在应用软件层次,用户则可以直接使用不同种类的应用软件和服务。在算力基础设施中,算力中心的异构化和多样化使得算力服务变得更加复杂。以平台层为例,即使是高性能计算所需的 MPI 环境和基础算法库,在不同的算力中心中也会有不小的差异,用户在不同算力中心中使用这些服务时往往需要进行适配和修改(如修改作业脚本等)。由于算力基础设施需要向用户提供透明、一致的算力服务,如何屏蔽这些差异,就是需要着力研究和解决的问题。

3) 多算力中心分布式协同计算与虚拟超级计算机。在过去,由于算力中心间的网络带宽有限,计算任务通常分配在单个算力中心上完成,即使进行跨中心计算,一般也仅限于工作流中的不同计算阶段。近年来,随着网络基础设施和算网融合技术的发展,算力中心间的网络互联带宽不断提升,传输延迟显著下降。国内已有超算中心间网络互联达到传输带宽10 Gbps、延迟接近 1 ms 的水平,这给多个算力中心进行分布式协同计算提供了可能性。这一方面需要算力中心基础支撑软件互联互通构成分布式计算环境,另一方面,还需要研究在这种环境中的任务划分、调度和迁移等技术。更进一步,在互联带宽和延迟满足要求的情况下,是否可能将多个超算中心互联成为统一管理、统一调度的单一计算系统,形成可完成数倍于原有规模并行计算的“虚拟超级计算机”,也是值得研究和探讨的问题。

4) 多样化算力中心和异构化体系结构的编程问题。异构体系结构显著增加了并行编程的复杂性,该问题在多样化算力中心场景下更加突出。为了支持异构处理器/加速器编程,厂商推出了相应的编程语言/接口,如用于 Nvidia GPU 的 CUDA、用于申威众核处理器的 Athread、用于 AMD GPU 的 ROCm/HIP 等,并基于这些编程接口开发/移植了各种基础算法库、求解器、深度学习框架等,但异构平台编程的复杂度仍然远高于传统的 CPU 平台。在多样化算力中心中,异构硬件平台的种类更多,为使程序具有更好的平台适应性,在软件编程模型和语言方面还需要进行更多的工作。虽然近年来已经出现了一些独立于厂商的加速器编程接口,如 OpenCL、OpenAcc、SYCL 等,但这些编程接口在不同硬件平台上的实现仍然有差异,为一种平台编写的程序通常难以不加修改地在另一种平台上编译和运行。有鉴于此,仍然有必要提出独立于硬件平台且可屏蔽硬件细节的编程模型/语言,与此同时,研究开发异构程序转换工具,实现并行程序在不同硬件平台间的透明转换和自动编译,进而支持并行程序在多样化算力中心的透明调度和运行,也将是一项很有价值的工作。

5) 数据在分布式算力中心间的放置问题。无论是科学工程计算,还是人工智能或大数据应用,其数据规模都较为庞大。由于算力中心间的数据传输和访问开销较大,在算力中心间进行计算任务调度和迁移时,数据放置就成为必须考虑的一个重要因素。

6) 公共算力中心的数据安全和隐私问题。在人工智能领域,为了满足数据隐私和安全需求,提出了联邦学习技术(federated learning)[27],通过多个数据拥有者协同完成训练,避免了数据向其他实体公开。这种数据隐私和安全性需求同样存在于大数据分析、科学与工程计算等领域。为了使用公共算力中心的计算服务,用户往往需要将数据上传至算力中心,虽然通过 VPN 等技术可以保证数据在网络中传输时的安全性,但在多用户共享的算力中心中,数据在外存中的存放,以及计算过程中数据在内存的存放,仍然存在数据外泄的可能性。为满足对数据安全性要求较高用户的需求,如何在大数据分析和科学工程计算领域提供类似于联邦学习的机制,或实现“可计算但不可读写”,是值得深入研究的问题。

3.3　中国超算应用生态存在问题和算力基础设施未来展望

3.3.1　中国超算应用生态存在的问题

近年来,中国高性能计算技术水平取得了长足进步,超级计算机研制水平已处于国际前列,拥有性能排名前列的超级计算机,生产和部署的高性能计算系统数量也世界领先。在高性能计算应用软件方面,面向国产超算系统研发了一批重点行业/领域应用软件,取得了众多的应用成果,大规模并行算法及应用也 2 次获得代表国际超算应用最高水平的 Gordon Bell 奖,但总体上,高性能计算软件与应用的发展相对不足,应用生态也不够丰富。产生这种现象的原因有多个方面:

1）软件和应用研发投入不足。中国科研领域和产业界长期存在着“重硬轻软”的现象,在高性能计算领域,国家投入经费的大部分都用于高性能计算机系统研制,软件和应用处于配合和支撑地位,与发达国家超算研究计划中硬软件投入接近 1:1 相比,中国的软件和应用研发投入明显不足。

2）软件和应用种类多、研发持续时间长。软件和应用种类繁多,应用生态的建立需要长期持续的努力。一种新型处理器/加速器推出后,初期往往只具备操作系统和编译等核心软件,而单靠硬件研发单位完成多种编程语言编译器、调试及性能分析工具、基础算法库、求解器、各种领域应用软件的研发非常困难,需要多方参与,经过若干年的持续努力,逐步研发和完善。

3）用户使用习惯不利于国产软件的推广。中国超算应用软件研发起步相对较晚,西方国家在很多行业和领域已推出了商业化软件,用户已形成了商业软件使用习惯和对商业软件的认知度,某些行业甚至只认可某种软件的仿真结果,这给国内应用软件的自主研发和推广应用带来了很大困难。一种软件研发完成后,需要通过推广应用来支撑驱动软件维护和持续升级,通过软件持续升级,不断增强功能并改善用户体验,进而吸引更多用户使用,然而,目前国产超算软件研发还未能形成这种良性循环。

软件和应用生态存在的不足使得中国高性能计算领域存在着“大而不强”的现象,与此同时,高性能基础和应用软件大量依赖国外软件,也存在“卡脖子”的风险。

3.3.2 中国算力基础设施的发展展望

“东数西算”战略的实施将形成算力中心建设的新高潮,为了构建国家算力基础设施,需要在研发突破关键技术的基础上,补足中国超算软件与应用的短板,并通过运营模式和机制创新,建立起丰富且自我发展的国产软件应用生态。为此,需要在以下方面开展重点工作:

1）研究突破关键技术,支撑算力基础设施发展。围绕算力中心异构化和多样化带来的技术挑战,解决算力基础设施面临的技术难题,研发核心软件和服务平台,实现多样化算力中心的互联互通、资源共享和服务提供,为国家算力基础设施的构建和发展提供技术支撑。

2）强化计算软件研发,补足国产软件与应用短板。以国产处理器/加速器的兴起为契机,加强基于国产硬件的工具链、算法库、求解器、领域应用等基础和应用软件研发,通过若干年的持续努力,建立较为完备的国产硬件支撑和应用软件栈,形成可自我发展的国产软件与应用生态。

3）改变单一机时服务方式,推动算力中心能力建设。国内已建成的超算中心为了弥补运行经费的不足,普遍以提供机时服务为主,即俗称的“卖机时”。这种低层次的算力服务消耗了较多的人力和精力,制约了算力中心向更高水平发展。通过建立国家算力基础设施,可以推动算力中心从机时提供者向应用研发者和解决方案提供者转变。同时,各超算中心通过研发建立领域应用平台,可以突出自己的技术特色,进而形成算力中心各有所长的态势,也可以避免算力中心发展同质化。

4）创新算力运营模式和机制,打造多方共赢的应用生态。算力中心涉及地方政府、投资方、设备提供方和运营方,在构建国家算力基础设施的过程中,需要以多方共赢为目标,通过运营模式和机制创新,鼓励多方参与,通过竞争促进技术进步和服务水平提升,与此同时,以应用商店(App store)等模式打造研发、服务、运营等多方共赢的软件和应用生态,进而推动国家算力基础设施做大做强。

4 结束语

CNGird 在国家科技计划支持下历经 20 余年发展,已经成为不可或缺的国家高性能计算基础设施,并为“东数西算”背景下国家算力基础设施的建设积累了宝贵经验,奠定了技术基础。温故而知新。面对“东数西算”的新任务,要认真总结 CNGrid 建设的历史经验,分析应用和技术发展的新趋势,定位亟待解决的瓶颈技术问题,探索新的应用模式和机制,更高效地构建新一代国家算力基础设施,实现“东数西算”的国家战略。这是国家创新发展赋予的历史使命。

参考文献(References)

[1] DENNIS J. Segmentation and the design of multiprogrammed computer systems[J] Journal of the ACM, 1965, 12(4): 589-602.

[2] SACKMAN H. Time-sharing versus batch processing: The experimental evidence[C]// Proceedings of the American Federation of Information Processing Societies. New York: ACM, 1968:1-10.

[3] SCHWARTZ J, COFFMAN E, WEISSMAN C. A general-purpose time-sharing system[C]// Proceedings of the American Federation of Information Processing Societies. New York:

ACM,1964:397-411.

[4] MILLS D L,BRAUN H. The NSFNET backbone network[C]// Proceedings of the ACM Workshop on Frontiers in Computer Communications Technology. New York:ACM,1987:191-196.

[5] FOSTER I T,KESSELMAN C. The grid:Blueprint for a new computing infrastructure [M]. San Francisco:Morgan Kaufman Publishers,1998.

[6] STEVENS R,WOODWARD P,DEFANTI T,et al. From the I-WAY to the national technology grid[J]. Communications of the ACM,1997,40(11):50-60.

[7] THOMAS M,BOISSEAU J,DAHAN M,et al. Development of NPACI grid application portals and portal Web services[J]. Cluster Computing,2003,6(3):177-188.

[8] FOSTER I,CZAJKOWSKI K,FERGUSON D,et al. Modeling and managing state in distributed systems:The role of OGSI and WSRF[J]. Proceedings of the IEEE,2005,93(3):604-612.

[9] TALIA D. The open grid services architecture:Where the grid meets the Web[J]. IEEE Internet Computing,2002,6(6):67-71.

[10] FOSTER I,KESSELMAN C. Globus:A metacomputing infrastructure toolkit[J]. International Journal of Supercomputer Application,1998,11(2):115-129.

[11] REED D. A. Grids,the TeraGrid,and beyond[J]. IEEE Computer,2003,36(1):62-68.

[12] KUNSZT P. European DataGrid project:Status and plans[J]. Nuclear Instruments and Methods in Physics Research Section A:Accelerators,Spectrometers,Detectors and Associated Equipment,2003,502(2-3):376-381.

[13] GAGLIARDI F,JONES B,GREY F,et al. Building an infrastructure for scientific grid computing:Status and goals of the EGEE project[J]. Philosophical Transactions of the Royal Society A:Mathematical,Physical and Engineering Sciences,2005,363(1833):1729-1742.

[14] HEY T,TREFETHEN A E. The UK e-Science core programme and the grid[J]. Future Generation Computer Systems,2002,18(8):1017-1031.

[15] MATSUOKA S,SHINJO S,AOYAGI M,et al. Japanese computational grid research project:NAREGI[J]. Proceedings of the IEEE,2005,93(3):522-533.

[16] ARMBRUST M,FOX A,GRIFFITH R,et al. Above the clouds:A Berkeley view of cloud computing:UCB/EECS-2009-28[R]. Berkeley:EECS Department University of California,Berkeley Technical Report,2009.

[17] SARASWAT M,TRIPATHI R C. Cloud computing:Analysis of top 5 CSPs in SaaS,PaaS and IaaS platforms[C]// 2020 9th International Conference on System Modeling and Advancement in Research Trends,2020:20421390.

[18] SOTOMAYOR B,MONTERO R,LLORENTE I,et al. Virtual infrastructure management in private and hybrid clouds[J]. IEEE Internet Computing,2009,13(5):14-22.

[19] BARIK R,LENKA R,RAO K,et al. Performance analysis of virtual machines and containers in cloud computing[C]// 2016 International Conference on Computing,Communication and Automation. Piscataway:IEEE Press,2016:16585534.

[20] SIMONS J. HPC cloud bad;HPC in the cloud good[C]// 2013 IEEE 27th International Symposium on Parallel and Distributed Processing. Piscataway:IEEE Press,2013:13683523.

[21] MOR N. Edge computing:Scaling resources within multiple administrative domains[J]. Queue,2018,16(6):106-116.

[22] 乔健,查礼. 中国国家网格作业管理设计与实现[J]. 计算机应用,2008,28(8):2003-2009.
QIAO J,ZHA L. Design and implementation of grid job management for China national grid[J]. Computer Applications,2008,28(8):2003-2009(in Chinese).

[23] 王小宁,肖海力,曹荣强. 面向高性能计算环境的作业优化调度模型的设计与实现[J]. 计算机工程与科学,2017,39(4):619-626.
WANG X N,XIAO H L,CAO R Q. Design and implementation of an optimal job scheduling model for the high performance computing environment[J]. Computer Engineering & Science,2017,39(4):619-626(in Chinese).

[24] 喻林,邹永强,查礼. CNGrid GOS 安全:设计与实现[J]. 华中科技大学学报(自然科学版),2010,38(S1):6-10.
YU L,ZOU Y Q,ZHA L. CNGrid GOS security:Design and implementation[J]. Journal of Huazhong University of Science & Technology (Natural Science Edition),2010,38(S1):6-10(in Chinese).

[25] SEVILLA J,VILLALOBOS P,C ERON J,et al. Parameter,compute and data trends in machine learning[EB/OL]. [2022-05-30]. https://docs.google.com/spreadsheets/d/1AAIebjNsnJj_uKALHbXNfn3_YsT6sHXtCU0q7OIPuc4.

[26] TOP500 list[EB/OL]. [2022-06-20]. https://top500.org/lists/top500/2022/06/.

[27] BONAWITZ K,EICHNER H,GRIESKAMP W,et al. Towards federated learning at scale:System design[C]// Proceedings of the Conference on Machine Learning and Systems. Piscataway:IEEE Press,2019:1-15.

From grid to "East-west Computing Transfer": Constructing national computing infrastructure

QIAN Depei*, LUAN Zhongzhi, LIU Yi

(School of Computer Science and Engineering, Beihang University, Beijing 100083, China)

Abstract: This article gives a review of the evolution of computer use-mode over the time since the invention of modern computers and presents the challenges and tasks in building the national computing infrastructure. The first section of this article provides a brief overview of how computer use-mode has evolved over the past several decades. Then the design and implementation of China's national high performance computing infrastructure CNGrid are introduced. Following that, the trends and new technical challenges in developing the computing infrastructure under the circumstance of the national strategic project of "East-west Computing Transfer" are discussed. Finally, the perspectives of building the supercomputing eco-system and constructing the new type of computing infrastructure in China are presented.

Keywords: high performance computing; infrastructure; grid computing; CNGrid; East-west Computing Transfer

Received: 2022-08-02; **Accepted**: 2022-08-18; **Published online**: 2022-08-23 09:07

URL: kns.cnki.net/kcms/detail/11.2625.V.20220822.1844.002.html

* **Corresponding author**. E-mail: depeiq@buaa.edu.cn