



## 中国科学院科研信息化 基础设施及应用\*

文 / 南 凯  
中国科学院计算机网络信息中心 北京 100190

**【摘要】** 经过多个“五年计划”的持续投入和建设,中科院在高速科研网络、超级计算和数据应用环境等科研信息化基础设施方面取得了长足的发展,有力地支持了各项科研创新工作。文章首先分析了国际科研信息化基础设施的发展趋势,阐述了中科院科研信息化基础设施建设及其在科研应用方面的进展情况,最后对所面临的机遇和挑战进行讨论,并提出下一步的发展设想。

**【关键词】** 科研信息化,基础设施,应用

DOI 10.3969/j.issn.1000-3045.2013.04.008

### 1 引言

科研信息化是当今时代科技活动的一个鲜明特征,是科研模式的重大变革,是迈向新一轮科技革命、提高科研水平和创新效能的必由之路,同时也是整个社会信息化的先导。具体而言,科研信息化是指充分利用网络信息基础设施与技术,促进科技资源交流、汇集与共享,变革科研组织与活动模式,推动科技发展转型的历史进程<sup>[1]</sup>。

近年来,科研活动对信息化基础设施的需求日益迫切并不断提高。高技术、基础研究、生态环境和生物技术等领域对高速网络、超级计算、数据存储及处理的需求都在快速增长,大科学工程对科学数据传输、处理、计算和存储需求急剧上升,

野外台站对信息化的需求也十分迫切。日益增加的国际国内科技合作也对信息化设施和服务提出了更高的要求。

中科院长期以来高度重视科研信息化工作,把科研信息化作为提升科技创新能力的战略任务,从“十五”到“十二五”连续部署了信息化建设专项。中科院计算机网络信息中心是中科院科研信息化的支撑服务机构,承担着信息化建设、技术支撑与服务等工作。经过10余年的不懈努力和持续发展,中科院科研信息化基础设施建设取得了显著的成效,科研信息化应用不断进步。

### 2 国际科研信息化基础设施

当前,欧美发达国家纷纷将科研信息化基础设施建设作为提升科技创新能力和未来国家竞争力的重要战略举措,持续部署新的科研信息化基

\* 修改稿收到日期:2013年4月17日

基础设施。

在高速科研专用网络方面,为应对大数据科研的需要,网络传输能力和传输速度大幅提升。在欧盟,泛欧科研教育网(GÉANT)<sup>[2]</sup>(图1)将部署先进的传输设备和转换平台,全面升级长达5万公里的GÉANT骨干网,支持GÉANT核心网络实现最高达2 Tbps的传输能力。在美国,政府投资6 200万美元启动“先进网络计划(Advanced Networking Initiative, ANI)”<sup>[3]</sup>,拟建设100 Gbps网络来支持大数据传输。2012年11月,在美国盐湖城举行的2012年全球超级计算大会上,由美国、加拿大的物理学家、计算机科学家和网络工程师领导的国际研究团队创造了339 Gbps的最新网络数据传输速度记录。

在高性能计算方面,研发具备百亿亿次计算能力的高性能计算机正成为各国及其科研机构的追求目标。“欧洲百亿亿次级软件计划(European Exascale Software Initiative, EESI)”将联合产业界和政府机构,帮助用户在未来10年内从千万亿次超级计算提升至百亿亿次超级计算。美国能源部的“先进科学计算研究(Advanced Scientific Computing Research, ASCR)”计划也致力于解决

数据密集型科学和百亿亿次计算面临的问题。欧洲“先进计算合作伙伴(Partnership for Advanced Computing in Europe, PRACE)计划”部署泛欧Peta-Scale生态系统,并在2020年之前达到Exa-Scale的运算性能。

在科学数据及应用方面,建设适应于海量科研数据产出的数据存储基础服务、探索和建立面向科学数据的保存和共享模式。美国国家科学基金会(NSF)于2007年发布了“科学数据可持续保存与共享网络伙伴计划(Sustainable Digital Data Preservation and Access Network Partners, DataNet)<sup>[4]</sup>”项目指南,旨在建立面向科学数据的保存及共享模式。NSF希望通过该项目建立一批具有显著示范作用的新型科学数据保存与共享组织架构,为长期的科学研究提供可靠的数据保存、访问、集成和分析能力。

### 3 中科院科研信息化基础设施

#### 3.1 高速科研网络环境

中科院高速科研网络环境以中国科技网(CSTNET)作为基础网络平台。中国科技网是中科院领导下的全国性计算机网络,是学术性、非赢利性、为科技服务的信息化基础设施。作为中国最早的互联网络单位,

历经10余年的发展,中国科技网已形成包括12个地区分中心的国内骨干网(图2)。近年来通过下一代互联网(CNGI)高速科研数据网络的建设,中国科技网在核心网、骨干网、研究所IPv6内网和网络互联等多层面显著提升了网络基础设施能力,全面支持IPv4和IPv6双栈。

在互联互通方面,中国科技网与中国电信、中国联



图1 泛欧科研教育网<sup>[2]</sup>



中国科学院

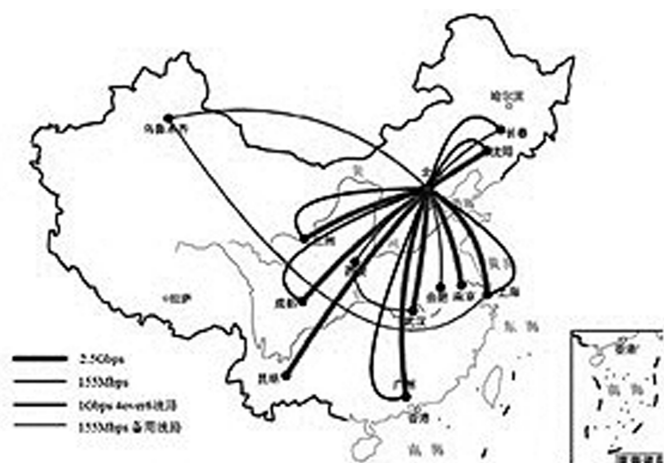


图2 中国科技网骨干网覆盖情况

通、中国移动、中国教育网等国内主要互联网运行商实现了高速互联,其中与中国教育网(CERNET)的互联互通带宽达到22 Gbps。与美国、欧洲、韩国、日本、香港、台湾等国家和地区的学术网络及主要互联网运行商实现了高速互联,其中中美、中欧之间均为10Gbps。

全球高速科研网络(Global Ring Network for Advanced Applications Development, GLORIAD)项目的参与国家和地区已由最初的中美俄三方扩大到加拿大、新加坡和韩国等15个国家或地区。中国科技网将在位于北美的国际科研网汇聚中心(PacificWave)建立科研网国际交换中心,借助NSF的支持,共享与印度、新加坡、越南和埃及等国家的高速网络连接,并由此带来更多的科研信息资源。

中欧科研网络连接(ORIENTPlus)项目实现了中国科技网与欧洲科研机构之间的高速互联,大力推进了中科院用户与欧洲各科研领域的全面合作,有力支持了中科院在高能物理、气象观测、生物信息、生命科学、天文联测、遥感观测和高性能网格计算等学科的前沿应用。为了更好地满足科研应用的需求,2013年1月,ORIENTPlus中欧之间的网络带宽从2.5Gbps提升至10Gbps。

### 3.2 超级计算环境

中科院在“优化布局、整合资源、强化服务、推

动应用”的建设原则指导下,利用自主研发的网格中间件有效地整合了院内超级计算资源,建成了具有3 000万亿次GPU和300万亿次CPU以上聚合计算能力的中科院超级计算环境<sup>[5,6]</sup>。该环境实现了统一运行管理、技术支持与服务。

中科院超级计算环境已形成由总中心、8家分中心、17家所级中心组成的三层架构式超级计算网格环境,应用领域涵盖了基础科学、地理资源与环境、生命科学和高技术等,对外提供机时服务4 126万多CPU小时,服务用户数近228个,已封装网格应用81个。超算环境实现了由单个计算节点向超级计算网格环境的转变,实现了资源整合与共享使用、提高资源利用率的目标。

### 3.3 数据应用环境

中科院统筹规划和实施了面向全院服务的数据应用环境建设,初步形成以科学数据中心为核心设施的存储备份环境,可提供云存储、云计算、云归档和云灾备等基础设施支撑服务,为海量科学数据的管理、共享和应用提供运行支撑环境<sup>[5,6]</sup>。

数据应用环境的云存储服务将分布在中科院12个分院区域存储节点的大量存储设备上,通过分布式文件系统集合起来协同工作,在存储资源虚拟化技术的基础上,形成统一的存储资源池,实现用户自助使用存储服务;云计算服务是建设可共享服务的虚拟化资源池环境,提供500个以上的虚拟机,为科研应用系统提供弹性、按需、自助的云计算虚拟机环境;云归档的系统归档能力大于20TB/天;云灾备环境旨在提供可在30分钟内恢复数据的灾备服务,达到国家灾备能力5级标准。

## 4 科研应用

中科院科研信息化工作坚持以应用需求为出发点,在应用学科领域开展信息化环境下新型科研方法的探索和实践。下面介绍几个典型应用案例。



#### 4.1 甚长基线观测网 e-VLBI (e-Very Long Baseline Interferometry)

e-VLBI 是甚长基线干涉技术与高速网络技术相结合的一种全新天文观测手段,射电天文望远镜将观测数据通过高速网络准实时地传送到相关处理中心进行处理,并实时获取相关干涉处理结果。中科院高速网络环境为 2007 年“嫦娥一号”测轨任务和 2010 年“嫦娥二号”测轨任务提供了网络保障,为中国、澳大利亚、日本三方首次 e-VLBI 实时观测提供了专用光通道及技术支持(图 3)。e-VLBI 使得嫦娥工程的测轨精度由几公里提升到几十米。



图 3 甚长基线观测网 e-VLBI

#### 4.2 北京及周边地区奥运大气环境监测

“北京及周边地区奥运大气环境监测和预警联合行动计划”为北京奥运会的空气质量保障提供及时准确的预测和风险评估数据。该联合行动由跨学科、跨地域的多个研究机构共同参与,包括环境光学监测、生态环境、大气、卫星遥感、数据分析与可视化等多个领域的研究机构和专家,项目组成员来自北京、合肥、上海等地,多个观测站分布在京津冀地区,是现代大范围协同的科研攻关活动。采用协同工作环境套件(Duckling)建立的奥运大气环境监测协同

平台为项目组提供了多维数据集成、项目协作和信息发布等服务,有效地支持了项目组成员之间的协作,实现了项目信息及时、准确、规范的发布和项目数据的集成与共享,并且能够系统、直观地展示监测结果。

#### 4.3 青海湖区域重要野生鸟类资源及疫病监测与风险评估

青海湖区域是众多候鸟的繁殖地或越冬地,而且还是多种候鸟迁徙路线的中转站。迅速而准确地掌握青海湖区域重要野生鸟类的迁徙路线及其迁徙过程中的信息对禽流感疾病的传播和防治等涉及公共安全的重大事件有重要意义。面向青海湖区

域重要野生鸟类集成研究的网络化科学研究平台是国内首个应用于国家级自然保护区的 e-Science 应用,也是国内首个在高原地区支持生态保护和研究的 e-Science 应用。该工作以中科院研究所为主,联合了多个单位和领域的科研人员,共同组成了一个虚拟科研组织,逐步形成了所所合作、院地合作、国际合作

等多种合作形式。

#### 4.4 高能物理数据密集型计算

中科院高能物理所联合相关单位建立了由高性能计算及存储组成的数据密集型计算平台。该平台通过中国科技网的中欧科研网络连接(ORIENTplus)与世界上最大的网格系统——全球大型强子对撞机计算网格(Worldwide LHC Computing Grid, WLCG)相连。该平台每年提供 2 000 万 CPU 小时的计算服务、处理超过 3 PB 的科学数据,为物理、生物医学、地质地理等领域



中国科学院

的科学计算提供了重要支撑,特别为大型强子对撞机实验发现类希格斯粒子做出了重要贡献<sup>[6]</sup>。

#### 4.5 子午工程空间天气研究信息化环境

子午工程是在我国境内沿东经120度子午线,部署23类监测设备、近50个监测台站,并与境外西经60度子午链线上的观测点合作,构成国际上第一个空间环境地基监测子午圈。子午工程项目中涉及大量数据的传输、储存和处理应用,有机融合了观测台站、网络通信系统、数据和高性能计算等多种要素,充分发挥科研活动信息化的技术支撑作用,为我国卫星、通信、导航、电力网、载人航天等高技术系统的安全运行提供重要保障<sup>[5]</sup>。

### 5 机遇与挑战

#### 5.1 中科院信息化基础设施发展的机遇

“十八大”提出实施创新驱动发展战略;国家“十二五”规划也明确提出,以科学发展为主题,把科技创新和进步作为重要支撑,把建设资源节约型、环境友好型社会作为重要着力点,把改革开放作为强大动力。这一系列战略部署对中科院信息化基础设施如何支撑科技创新跨越发展提出了更高的要求,这也带来了难得的发展机遇。

随着“创新2020”的启动实施,中科院的科技创新活动和组织管理方式将突破时空、学科和机构的限制,不断创新科研活动模式,最终实现“创新跨越、布局合理、四个一流、开放合作、和谐有序、持续发展”的目标,在我国科技事业发展中发挥服务全局、骨干引领和示范带动作用。为实现“创新2020”的宏伟目标,要求大幅提升科研信息化基础设施的能力和应用水平。

信息技术是当今最具活力的科技创新领域之一,不断孕育着重大突破。信息网络基础设施进入重大变革期,宽带、泛在、智能、融合的新一代信息基础设施建设步伐不断加快。信息技术及产业的加速创新和发展,为加快科研信息化基础设施的建设,强化基础设施与各学科领域发展的深度融合,提升科技创新能力和效率提供了良好的基

础和环境。

#### 5.2 中科院信息化基础设施面临的挑战

信息化基础设施的基本服务能力需要持续提升。当前,中科院的骨干网络带宽为2.5Gbps,超级计算网格聚合通用计算能力约为200Tflops,数据中心存储环境容量为6PB,使用率达到70%。根据对中科院15项重大科技基础设施在信息化方面的需求统计,到2015年前后,对骨干网络带宽的需求将超过10Gbps,通用计算能力(CPU)的新增需求超过1000Tflops,对数据存储备份容量的需求超过100PB。

信息化基础设施的综合服务能力需要着力加强。网络、数据与超算三大信息化基础设施环境建设取得了一定的成绩,但建设和服务还没有高度融合,整体服务效果尚不明显。在中科院新的发展战略、科技布局调整和实施先导科技专项的需求下,特别是国际科研信息化新的发展态势下,信息化基础设施的综合服务和集成能力需要进一步提高,建立融合性强、友好一致、灵活高效的科研信息化服务环境。

信息化基础设施的应用服务能力需要不断扩展。目前,缺乏能够最大限度地挖掘信息化基础设施潜力、具有自主知识产权的软件、算法和工具,主要表现在:部分科研现场缺乏有效的数据采集与方便易用的数据传输手段,缺乏面向应用需求的、可定制带宽的海量数据端到端高速传输工具,缺乏面向大科学装置的信息化基础设施的整体解决方案,缺乏面向大数据科研时代需要的高性能计算工具和算法,缺乏面向学科领域个性化数据处理和分析的数据挖掘和知识发现算法。

### 6 发展设想

“十二五”期间,中科院将以科研创新应用需求为牵引,不断完善科研信息化基础设施,持续提升其基本服务能力。中国科技网将根据需求不断提升骨干网络带宽,逐步建设高可靠性的骨干环网,持续推动与国内国际科研网络的高速互联,大

力加强创新性应用增值服务的开发和集成,使中国科技网成为重点支撑科技活动的综合性、一体化网络服务平台。中科院数据环境存储设施总容量将于2015年底达到50PB。中科院超级计算环境将建成具有云服务功能的超级计算平台,总体计算能力达10Pfllops,其中通用计算能力达1.5 Pfllops,逐步建设一个具有千万亿次计算能力的国家超级计算中心。

构建一批跨机构、跨地域和跨领域的科研信息化应用平台,强化信息化基础设施间的互操作性,形成可公开共享、无障碍、无缝使用的“科技云”环境<sup>[8]</sup>,通过“基础设施即服务(IaaS)”和“平台即服务(PaaS)”两个方面着力加强信息化基础设施的综合服务能力。

在信息化应用服务能力扩展方面,实现网络延伸到大科学装置、野外台站等数据获取现场,支持野外台站传感器、大科学装置等产生的数据从科研现场到用户端安全、可靠、实时传输,并联接高性能计算资源、海量数据中心、科学数字图书馆、大型专业软件等信息化基础设施与资源,提供海量数据端到端的高速传输服务,支持泛在的科研活动。建设支持自主信息技术和物联网技术试验的真实网络试验平台。发展新型并行计算方法,探索GPU/XPU等新型计算模式,在GPU/XPU等并行计算软件与应用等方面取得突破。研发数据挖掘和可视化分析方法与技术,增强研究人员对海量数据的

分析能力,支持科研人员通过交互式的可视化界面进行分析推理,通过集成多源的海量数据,从大量的、动态的、模糊的、甚至是相冲突的数据中发现新知识。

科研信息化工作方兴未艾,且任重道远。通过坚持不懈地努力把握科技活动对信息化的深层需求,敏锐捕捉和快速响应信息技术新发展带来的机遇和变革,踏实做好信息化服务的方方面面,重视科研用户关心的每个细节,中科院的科研信息化基础设施一定能在科技创新的大潮中发挥更大作用。

#### 参考文献

- 1 江绵恒. 大力发展科研信息化,服务国家科技创新. 科研信息化蓝皮书,北京:科学出版社,2011,1(1):1-2.
- 2 The European eInfrastructures Observatory-Pan-European Networks. <http://www.enventory.eu/pan-european-networks.html>.
- 3 100G Network to Support Big Science. Robert Gelber. [http://www.hpcwire.com/hpcwire/2012-05-02/100g\\_network\\_to\\_support\\_big\\_science.html](http://www.hpcwire.com/hpcwire/2012-05-02/100g_network_to_support_big_science.html).
- 4 孔丽华(编译). 美国国家科学基金会的DataNet计划. 科研信息化技术与应用,2009,2(1):73-74.
- 5 中国科学院十二五信息化发展规划. 2011,1(1):12-14.
- 6 中国科学院信息化工作领导小组办公室. 中国科学院信息化发展报告2013,2013,1(1):68-84.
- 7 中国科学院信息化工作领导小组办公室. 中国科学院信息化发展报告2011,2011,1(1):112-113.
- 8 科技云总体实施方案. 2012,1(1):15-18.

## The Cyberinfrastructure for Science and Innovation of Chinese Academy of Sciences and Its Applications

Nan Kai

(Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China)

**Abstract** The cyberinfrastructure for science and innovation of Chinese Academy of Sciences (CAS), including the high-speed network, supercomputing, and data storage facility, has made rapid progress in last more

(转至500页)



中国科学院