

虚拟天文台与互联网格技术

中国科学院国家天文台 (100012)

崔辰州 赵永恒

摘要 虚拟天文台是建立在当代天文探测器制造技术以及计算机互联网技术高速发展的基础上的。它将把分散在世界各地的各种格式、各个波段的天文数据统一利用起来,形成一个有机的整体。同时,它还将提供各种数据检索、数据计算、数据可视化、数据挖掘等各种工具并可接受用户自己开发的程序,按照用户的要求对这些数据进行高效的处理,进而开展天文研究工作。互联网格技术是正在兴起的新一代互联网技术,它将把整个互联网整合成一台巨大的超级计算机,实现计算资源、存储资源、数据资源、信息资源、知识资源、专家资源的全面共享,为用户提供一步到位的服务。虚拟天文台与互联网格技术的结合将是客观发展的必然。网格技术为虚拟天文台的实现提供技术支持,虚拟天文台为网格技术提供试验场并对其发展不断提出新的要求。天文学界和信息产业界的合作必将是一种双赢的合作,将大大推动天文科学和信息科学的发展。

关键词 虚拟天文台 网格技术 数据共享

虚拟天文台(简称VO)基本工作原理

巡天(Sky Survey),就是对整个天区进行观测、普查。如果利用可视化工具和望远镜巡天得到的数据把所观测的天体再现出来,就是一个数字虚拟天空。如果综合利用 射线巡天、X射线巡天、紫外巡天、光学巡天、红外巡天和射电巡天所得到的观测数据,便可以构成一个全波段的数字虚拟天空。通过可视化软件按用户要求显示这些巡天望远镜所观测到的任何部分的天空,这就仿佛是一架虚拟的大型望远镜。如果在此基础上再开发出功能强大的计算软件、检索软件、统计分析软件等各种应用工具,这就是一个独一无二的虚拟天文台。到天文学家要进行科研工作便只需要打开电脑、登录网站、输入几个参数、点几下鼠标即可。

虚拟天文台产生的天文与技术背景

两百多年来,天文研究通常都是单个天文学家或者天文学家小组进行小数目天体的观测。在过去,天文学家花整个一生所获得的资料仅能勉强得出有统计意义的结论。加上威力大的设备观测时间非常有限,那些需要大量数据来解决的天体物理问题就无法进行研究。

现在情况正发生着戏剧性的迅速变化。这一变化是由近年来发生的前所未有的技术进步所推动的。这场天文学上革命性变化的主要领域集中在以下几个方面:望远镜的设计和制造、大尺寸探测器阵列的开发、计算能力的指数增长以及通信网络覆盖和容量的不断增长。

望远镜设计和制造的进步使得大型空基天文台成为可能,为 射线、X射线、光学和红外天文的发展开辟了新的前景。技术的进步也使得新一代的大口径地面光学和红外望远镜以及毫米波与厘米波单天线和多天线阵的建造成为可能。在光学和红外波段,这些进步已经与极灵敏的、高分辨率而尺寸不断增大的探测器阵列结合起来。随着这些技术进步不断成熟,功能更好、口径更大的空间和地面望远镜正在计划中。这些设备配有使用面积更大、像元更多的探测器的先进仪器。如同 Moore 定律反映计算能力随时间指数增长,近年来实测天文技

术的进步使得天文发展实际也要用 Moore 定律来描述。

更多的地面和空间大口径和复杂仪器的天文设备的出现必将产生一个关键而必然的结果：数据流极大地增加。比如我国将在 2004 年建成的 LAMOST 望远镜每夜观测得到的数据约为 2GB，目前哈勃空间望远镜(HST)每天大约产生 5GB 的数据，而美国正在计划中的大口径全景巡天望远镜（LSST）每天将产生 10 兆兆字节的数据！

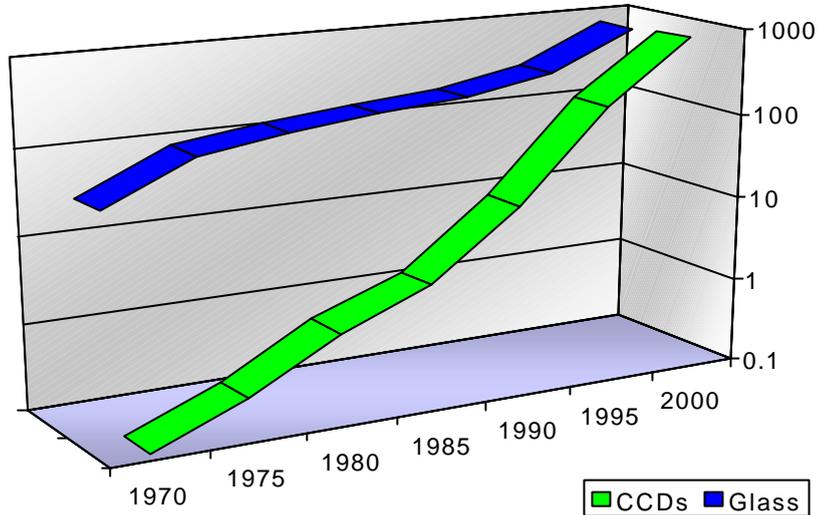


图 1 世界天文望远镜口径总面积与 CCD 总像素数的增长

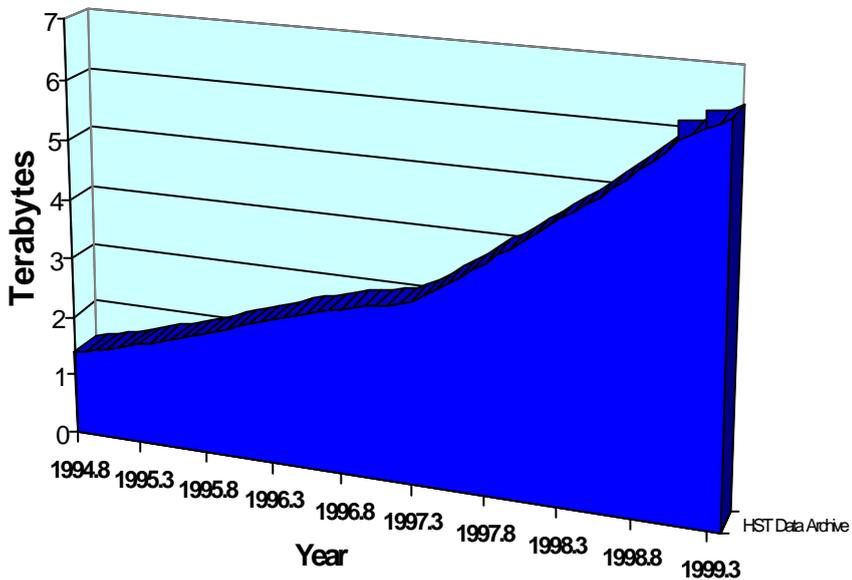


图 2 哈勃太空望远镜数据中心数据量的增长

除了数据增长速度外，进行观测的方式也有了变化。尽管新的地面和空间天文台还将继续分配给单观测者/单计划的观测模式的研究相当多的观测时间，但更多的时间是用来进行大规模的巡天。这些大的巡天计划将会产生大量质量均匀、标准统一的数据，通常会以兆兆字节来衡量。其内容的丰富和深度是空前的，因而能为广大用户提供独一无二的机会去应用这些数据从事各种科学问题的研究。此外，这些数据是使用昂贵的、而技术先进的设备获得的，而大大超额的对设备使用的申请，不允许重复已经做过的观测，所有这些都要求对数据

普遍存档。

使天文学研究的特征发生变化的另一个主要的技术进步是高速互联网技术，比如第二代互联网技术，特别是正在浮现的互联网格技术。高速的互联网传输速率加上地面和空间设备的高效率的数据采集，使得在不同地点间的大量数据交换成为可能。这将使得广大用户能够访问使用其中特定部分的数据，其潜在科学产出将是巨大的。

面对这样包含数亿个源的多波段数据档案，天文学界显然需要访问这些档案和分析其中数据的工具。数据挖掘、先进的模式识别、大规模统计交叉关联、罕见天体以及时变的发现等机会，显然都是存在的。此外，有了这样的数据集，将在天文学历史上第一次可以将复杂的数值模拟和统计上完备的多变量的数据进行有意义的对比。

近年来的这些技术进步汇合在一起，它将完全改变现行的大部分观测天文学的工作方式。这些变化是不可避免和不可逆转的，它们对天文学本身将产生巨大的影响。所有这些都表明，建立一种结构去有效地综合这些技术能力势在必行。

虚拟天文台就是这样一个机构，它将使天文学和天体物理学取得前所未有的进展。它将成为开创“天文学发现新时代”的关键性因素。虚拟天文台将是独一无二的，它将兆兆字节的数据档案、波长遍及从射线到射电波段的成百万个天体的图像库、高度复杂的数据挖掘和分析工具、可访问带有千兆字节存储容量的每秒运算次数达到万亿次的超级计算设备、以及在各主要天文中心间的极高速的互联，连成一体。它使成千上万的研究者可以快速查询各个达兆兆字节大小的档案；使埋藏在庞大星表和图像数据库中的多变量图案可视化；增加发现复杂图案和稀有现象的机会；鼓励多个研究小组间的实时合作；允许进行大型的统计研究，这些研究将首次使数据库的内容和复杂精密的数值模拟结果进行对比。虚拟天文台将促进我们对许多决定宇宙演化的天体物理过程的理解。它会用更经济的投资产生新的、更好的科学。虚拟天文台将作为一个协调性的和操作性的机构促进工具、协议和合作方面的发展，去充分实现未来天文数据库的科学潜能。

虚拟天文台的技术框架

虚拟天文台将成为“天文学发现”的一个发动机。虚拟天文台将是技术上可行的，是以科学来推动的。为了确保其持续的生命力，虚拟天文台必须包含几个主旨思想：

- 虚拟天文台必须是进化的。由于技术能力的不断提高，这种进化的特性将成为虚拟天文台自始至终不可或缺的组成部分。
- 虚拟天文台本质上必须是分布式的。不管是天文数据、各种应用工具软件还是虚拟天文台的组织管理机构都将在全球范围内分布。
- 虚拟天文台必须是集成的。它将向用户提供统一的、一体化的服务。资源的管理也许会通过全新的网络操作系统负责。研究中的网格操作系统将能满足这样的要求。
- 虚拟天文台必须为公众服务。通过虚拟天文台可以得到巨大的数据集和附带的分析工具，它们将给教育领域和公众领域提供一个前所未有的机会。
- 虚拟天文台必须面向全球需要。虚拟天文台使用的数据将来自世界各国的天文观测项目。同时，它提供的服务也要向全世界开放。
- 虚拟天文台必须提供一条通向未来的道路。它将是一个起催化和促进功能的机构，为天文学和技术科学的发展提供强大的动力。

虚拟天文台的实施涉及到许多重大的技术挑战。这包括现有的天文数据归档工作的协调以及新能力和新结构的发展。虚拟天文台的主要技术组成包括数据档案、元数据标准、数据

访问层、查询和计算服务，以及数据挖掘应用。这些能力的发展要求与信息技术界的密切合作。

实施虚拟天文台所面临的技术挑战是如何来解决相互对立的需求。一方面数据是广泛分布的，另一方面虚拟天文台的大型科学前沿研究需要巨大的计算资源和快速的数据本地访问；一方面要采用复杂的元数据标准和访问协议将分布的档案和网络服务连为一体，另一方面还要使小型档案接入虚拟天文台简便可行，以鼓励人们发表新收集的数据；一方面数据收集和计算服务是广泛分布的，另一方面还需要尽可能统一的系统接口以使数据和服务尽可能地透明。

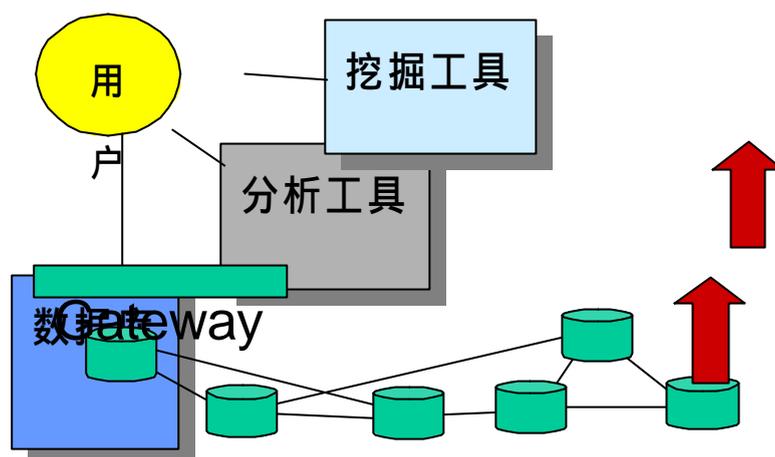


图 3 用户对虚拟天文台的使用

数据档案

数据档案里存储着数据集（如星表、图像和光谱），组织成逻辑相关的数据集合，也存储着描述数据档案及其数据保存的元数据。用户可以通过各种方式进行访问，如 WWW、FTP，或其它访问协议。

元数据标准

元数据是描述虚拟天文台的一些要素的结构化信息。元数据用来描述档案、档案提供的服务、其中的数据集合、每个数据集合的结构和语义、以及数据集合中每个数据集的结构和语义。典型的天文数据集是星表、图像或光谱等数据对象。

数据访问层

数据访问层（DAL）将提供统一的界面来访问虚拟天文台中所有的数据、元数据和计算服务。

查询和计算服务

当数据访问层和元数据标准允许虚拟天文台连接档案和访问数据时，就需要查询和计算服务来支持信息发现和提供数据挖掘所需要的统计相关和图像分析的能力。

数据挖掘应用程序

虚拟天文台应该提供这些工具，包括大型的多变量数据集的可视化和统计分析工具，让用户在极其丰富的数据海洋里产生新的成果。

信息系统研究

在存储技术、信息管理、数据处理、分布和并行计算、高速网络、数据可视化和数据挖

掘等各个领域,虚拟天文台将突破现有技术的限制。这就要求学术界和工业界携起手来共同研究和开发虚拟天文台所需要的信息系统技术,要求与其它科学分支和与超级计算机中心进行合作,来开发元数据处理、数据处理和分布式计算的标准。数据挖掘是一个需要天文学家、计算机科学家、数学家和软件专家一起合作才能解决的多学科问题。

教育和普及

虚拟天文台具有基于互联网的内禀性质,因此能够在前所未有的社会和地理范围内提供各式各样的高质量的科普和教育方法。职业教育和普及人士(教育家、天文馆和科学展览馆的工作人员、科普作家和科学记者等)将广泛地使用这些资源创建科普网站,编写课程教材和制作精美的演示等等。

国际 V0 的发展现状

虚拟天文台的概念设想首先是在美国名为“新千年中的天文学与天体物理学”科学展望中以首要位置提出的。概念已经提出便得到了世界各国天文学界乃至信息产业界的高度重视,纷纷提出自己的虚拟天文台计划。目前,有些虚拟天文台计划已经得到不同程度的经费支持并开始启动。当前在世界上比较有代表性的虚拟天文台计划主要有美国的国家虚拟天文台计划、欧洲的天体物理虚拟天文台计划、英国的天文网格计划、加拿大天文数据中心发展数据挖掘技术的计划。

美国国家虚拟天文台(NVO)

国家虚拟天文台是完全建筑于天文数据库和网络信息技术之上,需要强大的计算机与网络的软硬件支持,以及开发新的数据分析技术和知识发现工具。其利用最新的计算机技术、数据存储和分析技术,将来自地面和空间观测站的天文观测数据有机地统一在一起,目标是最大化来自这些观测数据的科学潜力,将数据以统一的标准的格式提供给专业研究者、天文爱好者和学生。该项目已通过美国国家科学基金立项,获得一千万美元资金资助,历时五年。

主要科学目标包括:多观测参数空间的探索、稀有天体与新型天体的发现、多波段天文学和统计天文学的崛起、数据挖掘技术的发展。

欧洲天体物理虚拟天文台(AVO)

天体物理虚拟天文台项目是由欧洲南方天文台(ESO),法国斯特拉斯堡天文数据中心(CDS),空间望远镜欧洲协调机构ST-ECF,巴黎天体物理研究所,英国曼切斯特大学等联合提出的。他们将历时三年来进行虚拟天文台的设计和应用的研 究(A阶段)。该提议于2001年2月份提交,如今已批准并得到近三百万英镑的启动资金。工作组决定于2002正式启动,主要集中在科学要求,科学互动和新的技术领域,A阶段将为B阶段(预计需要六千万英镑)的完全可操作的虚拟天文台打下坚实的基础。

A阶段主要发展目标有:

1. 制定细致全套的科学要求文档,用以设计、应用、操作虚拟天文台,配套以分布的、逐步升级的网格计算设施。
2. 构造恰当的平台和界面,把来自地面和空间观测设备的天文数据库有机地统一到虚拟天文台的数据库中。
3. 推动天文学家与计算机软硬件工程师的合作。在虚拟天文台和网格技术的主要领域(例如计算机网络、数据库设计和存储管理等方面),发展与欧洲工业界的对话。
4. 与一些要建立虚拟天文台的国家如美国、加拿大、澳大利亚加强合作与交流,最终创建全球性虚拟天文台。

英国天文网格

在英国，2003 年和 2005 年两个新的数据爆炸点即将到来，即当 UKIRT WFCAM、VISTA 项目正式上马，每晚将产生成百上千 GB 的数据。英国在 X 射线、太阳、空间等离子体物理等方面处于领先地位。但是随着新的数据的到来，给英国天文界带来了机遇和挑战，要想保持自己的领先优势，他们提出了天文网格来应对这个挑战。这个项目是英国天文界组织的将历时三年的天文服务网格项目，PPARC e-science 项目分配资金两千六百万，其中五到六百万用于天文研究，大多数用于天文数据网格。

项目目标主要包括：

- 1：建立用以处理英国主要天文数据的数据网格。
- 2：开发高效的数据库查询，数据挖掘工具。
- 3：提供一个标准的数据库查询和数据挖掘的软件接口。
- 4：实现同时检索多个数据库的能力。
- 7：实现用户上传算法进行数据挖掘。

加拿大天文数据中心 (CADDC) 发展数据挖掘技术的计划

加拿大天文数据中心决定发展从大的科学数据库中进行数据挖掘的工具，于是提出发展数据挖掘技术的提案，该项目历时三年，直接费用一百五十万美元。CADDC 在过去十年在数据存储、加工、存档、分布等方面处于领先地位，是首家使用 WWW 界面服务器，同时又是第一个使用 CD-ROM 技术的场所，而且与国际上许多单位都有过成功的合作关系，在计算机软件方面成果斐然。在各国争相出笼虚拟天文台的形势下，加拿大天文学界为保持其在国际方面的领先优势，增加该国天文学家的工作效率，提出了在加拿大天文数据中心发展数据挖掘中心的提案，为逐步过渡到全球性虚拟天文台做准备。

计划的主要目标：(决定在 2003 年结题)

- 1：加强加拿大在发展 TERAPIX 数据管道方面的作用，包括图像处理、图像分析和数据库建立。
- 2：对 TERAPIX 和所有其它 MEGACAM 数据，建立 TB 量级的数据库。
- 3：提高对 TERAPIX 数据进行数据挖掘的能力，如对 5 千万个天体快速查询、提取、可视化、统计分析、与其他的数据库交叉认证，创建多波段数据库交叉认证的工具雏形。
- 4：提高利用用户提供的子程序对 TERAPIX 的图像进行再现的能力。

网格 (Grid), 第三代互联网技术

从上世纪 60 年代末开始研究计算机分组交换技术到今天，互联网已经走过两代历程。第一代是 20 世纪 70~80 年代，主要的成就是把分布在世界各地的计算机用 TCP/IP 协议连接起来，主要的应用是电子邮件。第二代是 20 世纪 90 年代，主要成就是把成千上万网站上的网页连接起来，主要的应用是 Web 信息浏览以及电子商务等信息服务。目前正处于从第二代互联网向第三代互联网过渡的转型期。第三代互联网可称为信息服务网格 (Information Service Grid)，其主要特点是不仅仅包括计算机和网页、而且包括各种信息资源，例如数据库、软件以及各种信息获取设备等，它们都连接成一个整体，整个网络如同一台巨大无比的计算机，向每个用户提供一体化的服务。简单地讲，传统互联网实现了计算机硬件的连通，Web 实现了网页的连通，而网格试图实现互联网上所有资源的全面连通，包括计算资源、存储资源、通信资源、软件资源、信息资源、知识资源等。

国外媒体常用“下一代互联网”、“Internet2”、“下一代 Web”等词语来称呼与网格

相关的技术。要注意的是，“下一代互联网”和“Internet2”又是美国的两个具体科研项目的名字，它们与网格研究目标相交，但研究内容和重点有很大不同。中国科学院计算所李国杰院士认为，网格实际上是继传统互联网、Web 之后的第三个大浪潮，可以称之为第三代互联网。

物质与能量原则上只能分享，一吨水、一度电你使用了我就不能使用。而信息的最大特点是可以共享，不会因使用同一信息资源的用户多而被耗尽。在过去几年的信息化建设中，我们往往忽视了“信息应该共享”这一最本质的应用要求，把信息当成物质与能量一样使用，这已造成极大的浪费。

第三代互联网要解决的信息共享不是一般的文件交换与信息浏览，而是要把所有个人与单位连接成一个虚拟的社会组织，实现在动态变化环境中具有灵活控制的协作式信息资源共享。信息服务网格与 Web 最大的区别是一体化，即用户看到的不是数不清的门类繁多的网站，而是单一的入口和单一系统映象。比如一个用户打算出去旅游，只要向网格系统一次性输入出游人数、出游地点和时间以及其他特殊要求，不必分别与航空公司、铁路、旅行社、旅馆、气象部门、商店等单位联系，信息服务网格将自动与各有关部门协调，给用户完整的旅游方案并做好全部预订手续。由于信息资源的需求与供给都在动态变化而且分布在全国甚至世界各地，完成用户要求的一项服务可能要调用北京的超级服务器、上海的数据库或安装在西安的某台计算机上的软件，因此对服务器的响应时间、网络的带宽、特别是网格管理软件的复杂性与灵活性以及网络上各种设备的互操作性都有很高的要求。

互连网格的发展目标

现有的 Web 信息服务器就好像 Internet 世界上一个个孤立的小岛。虽然这些“小岛”之间暂时还有充足的带宽资源可用，但大量的信息还是被“锁”在各个小岛的中央数据库里，各“孤岛”之间并不能按照用户的指令进行有意义的交流。解决这一问题的最佳途径是建立跨越 Web 的信息分布和集成应用程序逻辑——信息网格。

信息网格是要利用现有的网络基础设施、协议规范、Web 和数据库技术，为用户提供一体化的智能信息平台，其目标是创建一种架构在 OS 和 Web 之上的基于互联网的新一代信息平台 and 软件基础设施。在这个平台上，信息的处理是分布式、协作和智能化的，用户可以通过单一入口访问所有信息。信息网格追求的最终目标是能够做到服务点播和一步到位的服务。简单地讲，网格是把整个互联网整合成一台巨大的超级计算机，实现计算资源、存储资源、数据资源、信息资源、知识资源、专家资源的全面共享。

第三代互联网的兴起将改变传统的 Client/Server 和 Client/Cluster 结构，形成新的 Pervasive/Grid 体系结构。客户端是各种各样的上网设备，而连在网上的各种服务器将组成单一的逻辑上的网格。

第三代互联网的本质特征表现在应用上。信息服务网格的服务包括文件消息、计算、信息内容、事务处理和知识服务等，因此信息服务网格可大致分为数据网格、计算网格、信息网格与知识网格。

网格技术的产生、发展必须具备以下三个基本条件：计算资源的广域分布、网络技术（特别是互联网）以及不断增长的对资源共享的需求。在计算机技术发展的早期阶段，只有很少数量的大型计算机，它们通常被安装在相互独立的计算中心内，多个计算机用户通过使用终端来共享一台大型机的资源，但却不能同时共享多台大型机的计算资源。随着网络技术的发展，多台大型计算机可以在局域网内互连，用户通过网络便可以同时使用多台计算机的资源。而互联网的飞速发展和普及使得网格计算技术的产生成为可能。图 4 显示了计算资源共享的发展过程。

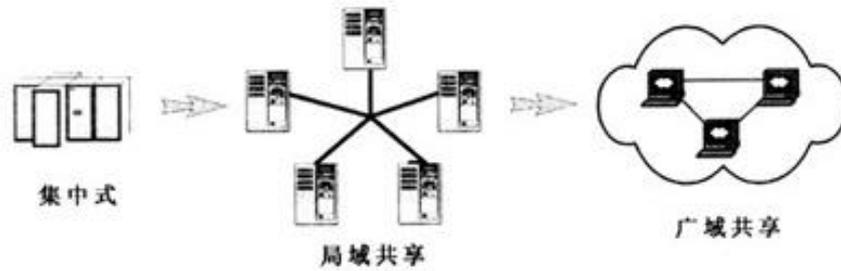


图 4 计算资源共享的发展过程

网络系统可以分为五个基本层次：构造层、连接层、资源层、汇集层和应用层。

构造层和连接层与传统互联网中的连接层、互联层和传输层的功能基本相同。

资源层是构成网络系统的硬件基础，它包括各种计算资源，如超级计算机、贵重仪器、可视化设备、现有应用软件等，这些计算资源通过网络设备连接起来。

汇集层是指一系列工具和协议软件，其功能是屏蔽网络资源层中计算资源的分布、异构特性，向网络应用层提供透明、一致的使用接口。汇集层也称为网络操作系统，它同时需要提供用户编程接口和相应的环境，以支持网络应用的开发。

应用层是用户需求的具体体现。在网络操作系统的支持下，网络用户可以使用其提供的工具或环境开发各种应用系统。

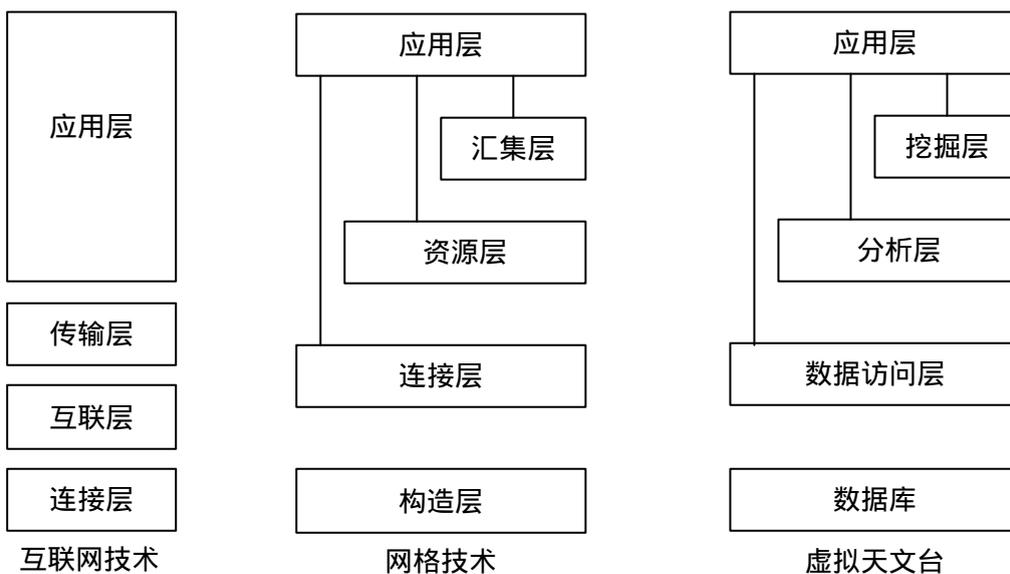


图 5 互联网技术、网络技术和虚拟天文台的对应关系

互连网的发展现状

网络作为互联网的第三次浪潮，在各国引起了前所未有的关注和重视。目前国外各大公司都在朝着这一方向努力。尽管各公司、各种学术团体采用的术语不同，但追求的目标都是攻克第三代互联网的关键技术。计算机方面的世界级大公司竞相推出了网络技术基础平台和相关的协议标准。

在国外，最著名的网络计算研究是 Globus 项目和 Legion 项目。他们现在都已经推出了

各自的网格软件。

微软的 .net 技术宣称是超越浏览器、超越网站的具有革命性的新技术，其基本理念是：不再关注单个网站和与互联网连接的单个设备，而是让所有的计算机、相关设备和服务商协同工作，提供更广泛和丰富的解决方案，使人们能够控制信息并让它在指定的时间以指定的方式传送给自己。据悉，除了 .Net 以外，微软内部正在开发全面支持网格的下一代视窗操作系统，计划在 2005 年完成。

IBM 公司倡导的 Web Services 是一种较成熟的商业计算服务共享解决方案，它可以使全球范围内的采购商、供应商和交易市场以低廉的价格共享商业服务。

Sun 公司利用其在跨平台语言方面的优势，以 Java 为核心推出了相应的 J2EE 平台规范。Oracle 在其最新版本的应用服务器 Oracle9iAS 中推出了全新的 Oracle Portal (Portlets) 技术。这些都是面向信息网格的新一代技术和解决方案。

同时，支持信息网格的关键协议如 XML、SOAP、UDDI、WSDL 都逐渐成熟并成为各种平台支持的基本协议。

我国在这方面也做了大量基础性和前瞻性的研究。从 1995 年开始，中国科学院计算所就建立了专门的网格研究队伍，开始研究与网格相关的技术，并开发成功了多个版本的信息一体化平台。

从 1999 年底到 2001 年初，中科院计算所联合国内十几家科研单位，共同承担了“863”重点项目——“国家高性能计算环境”的研发任务。该项目的目标是建立一个计算资源广域分布、支持异构特性的计算网格示范系统，它把我国的 8 个高性能计算中心通过互联网连接起来，进行统一的资源管理、信息管理和用户管理，并在此基础上开发了多个需要高性能计算能力的网格应用系统，取得了一系列研究成果。在今后五年内，我国的信息网格将初步建立。

建立在网格构架上的虚拟天文台

从上面的介绍，我们可以清楚的看出利用网格技术作为基础来发展虚拟天文台将是必然的发展趋势。

一方面，虚拟天文台的建立和实现需要网格技术的支持。

虚拟天文台的最终发展目标就是实现全球天文数据的高级共享，同时提供一整套的智能化工具。TB 量级甚至 PB 量级大型天文数据产出项目的不断涌现，对数据存储、数据管理、数据传输、数据检索等技术提出了更高的要求。在如此海量分布式数据的基础上进行科学研究，就必须有全新的数据共享、数据互操作、作业调度、数据可视化、数据统计分析、数据挖掘、数据安全管理等工具的支持。

虚拟天文台的这些需求正是网格技术要实现的目标。网格技术将实现把整个互联网整合成一台巨大的超级计算机，实现计算资源、存储资源、数据资源、信息资源、知识资源、专家资源的全面共享，为用户提供一步到位的服务。因此，虚拟天文台把网格技术作为自己的技术基础将是可行而明智的选择。

另一方面，虚拟天文台将为网格技术的发展提供最好的实验场。

天文数据有着其他学科数据无法比拟的特点：

1. 天文数据绝大部分是开放数据。国际上的许多大型天文观测项目的观测数据都会及时在互联网上公布，并可以免费下载或者向项目索取。这为数据共享提供了良好的基础，是其他学科无法比拟的。
2. 天文数据的数据量非常大。现有的以及即将实施的天文项目每天都会产生 GB 甚至 TB 量级的数据。天文数据中心的存贮容量已经达到 TB 量级，并开始向 PB 量级扩展。

3. 天文数据有比较好的归档。当前，世界上已经有多家天文数据中心在天文数据归档方面做了大量的工作，并取得了很好的应用，比如 ADS 和 CDS。这为数据库的建立、元数据标准的建立打下了良好的基础。
4. 天文数据的格式多种多样。在天文界，数据的内容主要有星表、星图、光谱等。数据的格式则各种各样。不管是以文本文件存贮还是以图像文件存贮，其内部格式都依不同的天文观测项目而变化。格式的多样化数据共享和互操作提出挑战。
5. 天文数据是全波段的数据。从伽玛射线、X 射线、紫外、光学、红外到射电波段都有观测项目在工作。多波段数据的证认为海量数据计算和挖掘提出很高的要求。

虚拟天文台要实现对这样数据的融合。这样的发展目标为网格技术提供了独一无二的试验场。从网格基础设施的构建，到网格操作系统的开发，最后到网格天文应用工具的实现，虚拟天文台为网格技术提供了一整套的应用需求。

虚拟天文台的实现不可能仅是天文学家的事情，必须与计算机、网络、数学等领域的专家共同努力。这无论对天文学还是对信息科学、数学都是双赢甚至多赢的合作。

参考文献

- [1] 天文学的新革命 - 虚拟天文台，崔辰州，天文爱好者，2001 年 5 期
- [2] Towards a National Virtual Observatory: Science Goals, Technical Challenges, and Implementation Plan, 2000
- [3] 超级服务器与信息网格，李国杰，科学新闻周刊，2001 年 42 期
- [4] 信息服务网格—第三代 Internet，李国杰，计算机世界，2001 年 40 期
- [5] 因特网涅槃—正在浮现的网格技术，徐志伟，计算机世界，2001 年 43 期
- [6] 万丈高楼平地起—浅谈网格计算基础，李伟，计算机世界，2001 年 43 期
- [7] 网格的称谓，徐志伟，计算机世界，2001 年 43 期
- [8] 信息网格—下一代信息服务平台，李晓林，计算机世界，2001 年 43 期