

虚拟天文台和网格技术

—网格战略研讨会

崔辰州

北京·中国

2002.4.4~5

各位专家、老师：

下午好！

今天的这个报告本来应该是由国家天文台 LAMOST 项目副总经理赵永恒博士来做，但由于他这两天出差不能前来参加讨论会，所以由我替他来做这个报告。我是赵永恒的博士研究生，崔辰州。报告中如有说得不对的地方，请各位老师指教！

在介绍虚拟天文台之前，让我先给大家介绍一下儿 LAMOST 项目。LAMOST 是“九五”国家重大科学工程项目，项目全称是“大天区面积多目标光纤光谱望远镜”。项目在 1997 年立项，计划 2004 年完成，为期 7 年。国家拨款共 2.35 亿人民币。这个望远镜在 2004 年建成后将安放在位于河北承德地区的国家天文台兴隆观测站。

LAMOST 是个很有特色的天文望远镜：

- 大视场。LAMOST 是个直径 4 米的望远镜，视场为 5 度，约为相同口径普通天文望远镜视场的 10 倍或更多。
- 高效率。在 LAMOST 望远镜的焦平面上将安插 4000 根光纤，也就是说它每观测一次可以同时得到 4000 个天体的光谱资料。目前世界上光纤数目最多的望远镜是美国的 SLOAN 望远镜，光纤数是 600 根。
- 大数据产出。LAMOST 每夜观测的原始数据大约为 3 到 5GB，每年可达 1 到 2TB。每个观测周期持续时间大约为 3 到 5 年。可以肯定，LAMOST 将成为未来我国天文界最重要的自产数据源。
- 专用。与通用的望远镜不同，LAMOST 望远镜是一个光谱巡天专用望远镜。它观测到的数据是 2 维光谱。观测方式为巡天方式，也就是对天区进行普查，而不是仅对单个天体目标进行观测。

在 2004 年 LAMOST 建成以后，它将成为世界上威力最大的光谱巡天望远镜，为天文学家及其他用户提供丰富的天文光谱数据。

在对 LAMOST 项目进行了简单的介绍之后，下面我给大家介绍被许多天文学家认为“将给天文学带来一场新革命”的虚拟天文台。

首先让我们了解一下儿什么是虚拟天文台，看看它的工作原理。刚才介绍 LAMOST 项目的时候我提到了“巡天”的概念。巡天就是对整个天区进行普查。把巡天观测得到的数据数字化（当然现代望远镜观测的数据大多本身已经是数字化的数据），便会得到一个虚拟数字天空。现在的天文学已经步入了全波段的时代，从 射线、X 射线、紫外、光学、红外，一直到射电波段都有仪器在观测。各个波段巡天数据的综合便构成了一个全波段的数字天空。利用数据检索以及可视化工具，按照用户需要把特定天区的信息显示出来，这就是一架威力巨大的虚拟天文望远镜。把世界各国的天文数据有机的结合起来，再加上相应的计算、统计、分析、数据挖掘等工具，这便是一个独一无二的虚拟天文台。天文学家的研究工作可以在鼠标的点击中完成。

虚拟天文台是个庞大的系统集成工程。从原始的观测数据、加工处理后的数据、数据传输协议、信息检索服务、各种工具软件以及高速互联网络都是虚拟天文台的成功实现所必需的。

今天我们在这里探讨网格技术的发展。我们认为虚拟天文台是网格技术最好的实验床。为什么这么说呢？这是因为天文数据有其他学科所不具备的特点。

- 天文数据绝大部分是开放数据。天文界的传统是所有的观测数据在一年后向公众开放。这既使得观测者有时间进行数据分析和发表早期结果，也使得其他天文学家可以有机会使用这些数据。国际上的许多大型天文观测项目的数据都及时在互联网上公布，这为数据共享提供了良好的基础，这在各类学科中是独一无二的。
- 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数 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。
- 天文数据有比较好的归档，并提供互联网服务。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。主要天文文献也都可以在线全文访问。
- 天文数据的格式多种多样。数据内容主要有星表、星图、光谱等，数据格式各种各样。不同的天文观测项目发布的数据格式往往也不一样。
- 天文数据是全波段的数据。各个波段都有观测项目在进行。重要的一点是，这些数据是高度相关的，在高维参数空间内进行研究具有非常重要的意义。
- 最后，新的观测数据总能带来全新的现象或规律的发现。

1998 年图灵奖获得者，微软研究院资深专家 Jim Gray 在不久前发表的一篇名为“适合虚拟天文台的计算机技术展望”的论文中就认为天文数据是计算机科学家研究数据分析、集成技术最好的实验场。Jim Gray 现在是美国国家虚拟天文台项目的顾问。

此外，虚拟天文台的实现也为网格发展提出了技术需求。首先，为了遍布在世界各地的不同格式的数据具有互操作性，需要新的数据归档方法。其次，虚拟天文台数据分析处理功能的实现也需要新的分析工具，比如快速检索工具、交叉证认工具、统计工具、可视化工具等。此外，海量数据的传输、运算也需要新的硬件设施、网络设施的支持。目前，国际上一些虚拟天文台项目已经考虑使用网格作为其网络基础设施。

与互联网一样，虚拟天文台的结构框架也是层次式的。主要包括数据库、分析工具和挖掘工具。一个完整的虚拟天文台将包括数据档案、元数据标准、数据访问层、查询计算服务、数据挖掘工具、信息系统研究、教育和普及等部分。

这是传统的互联网、网格、虚拟天文台层次结构的对照图。可以看出，虚拟天文台与网格的层次结构有异曲同工之妙，两者的结合将是发展的必然。以网格技术、XML 技术为代表的新一代互联网将为虚拟天文台的实现提供强大的技术支持。

虚拟天文台概念的提出不过两三年的时间，但已经引起了世界各国天文界的重视，纷纷提出了自己的虚拟天文台计划。目前，比较有影响的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（这个计划也是英国 e-Science 计划的一部分）。澳大利亚、加拿大、意大利等国也都提出了自己的虚拟天文台计划。

在我国，国家天文台、清华大学、北京大学也在积极推动虚拟天文台计划。但到目前为止，还没有形成一个完整、统一的行动计划。

虚拟天文台是一个雄心勃勃的计划，它将使天文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。但它的实现单靠天文学家的努力是不够的，它需要天文学家、计算机科学家、软件专家、数学家（特别是统计学家）以及其他学科领域专家的精诚合作。

我国虚拟天文台计划的提出和实现同样需要国内 IT 界、数学界的专家与天文学家的共同努力。

谢谢大家！