

国家自然科学基金资助项目

《中国虚拟天文台试验系统研究与开发》

项目总结报告

负责人：赵永恒

承担单位：中国科学院国家天文台

执行年限：2004.6-2007.12

2008年2月

目 录

一、摘要.....	1
1、中文摘要.....	1
2、英文摘要.....	1
二、研究计划要点及执行情况概述.....	2
三、研究工作主要进展和所取得的成果.....	4
1、China-VO 系统平台建设.....	4
1)、中国虚拟天文台的体系结构.....	5
2)、银河系化学丰度梯度研究及其在Globus Toolkit环境下的实现.....	10
3)、利用TwoMASS星表搜寻OB星协研究银河系旋臂结构及其在Web服务环境下的实现.....	11
4)、利用SDSS DR5 测光数据寻找银河系子结构研究银河系结构及其在VO-DAS下的实现.....	15
2、国内外天文数据的统一访问.....	22
1)、ConeSearch初体验.....	23
2)、US-VO SkyNode的跟踪与部署.....	25
3)、星表数据自动化处理系统.....	29
4)、VO-DAS数据统一访问平台.....	33
5)、SDSS数据的国内镜像.....	35
6)、天文数据的整理与发布.....	36
3、现有天文工具的虚拟天文台集成.....	37
1)、应用程序的服务封装探索.....	37
2)、天文数据可视化软件的自主开发尝试.....	43
3)、PLASTIC粘结下的现有软件协同.....	44
4、应用工具开发.....	48
1)、VOFilter设计开发.....	48
2)、FitHAS设计开发.....	49
3)、SkyMouse设计开发.....	51
5、虚拟天文台基础上的公众教育.....	53
1)、真正的万年历.....	54
2)、SkyMouse智能信息系统.....	55
3)、在线天文词典和宇宙驿站.....	56
四、队伍建设与人才培养.....	58
五、学术会议、国际合作与交流.....	58
1、中国虚拟天文台年会及用户培训.....	60
2、IVOA会议与国际合作.....	60
3、与国内相关领域的交流与合作.....	63
六、发表文章目录.....	63
1、已发表：.....	63
1)、SCI索引.....	63
2)、EI收录.....	64
3)、核心期刊.....	64
4)、国际会议.....	65
2、待发表论文.....	66

1)、SCI索引	66
2)、核心期刊.....	66
3)、国际会议.....	66

一、摘要

1、中文摘要

本课题《中国虚拟天文台试验系统研究与开发》作为虚拟天文台领域在国内的首次尝试，从2004年开始到2007年这三年半的时间里，组织开展了大量而深入的研究开发工作：以“科学应用”为目标，通过“科学带动技术”的路线，确立了以VO-DAS为核心的中国虚拟天文台平台体系；通过VO-DAS基本实现了对星表、光谱、图像等不同类型天文数据的无缝访问；开发并发布了VOFilter、VO_IMPAT、FitHAS、SkyMouse等一系列应用工具；组织了虚拟天文台年会、IVOA互操作会议等多个国内国际学术会议；培养毕业研究生10多人；发表科学和技术论文、科普文章、新闻报道数十篇。本课题完成的工作证明了在国内建设虚拟天文台的可行性和科学价值，为一个生产性运行的虚拟天文台系统的建设奠定了基础。本报告全面回顾了项目执行期间的工作开展情况。

关键词：虚拟天文台，天文信息技术，试验系统，研发

2、英文摘要

The NSFC project, "Research and Development of Chinese Virtual Observatory Testbed", is the first exploration in the field of Virtual Observatory in China. During the period of the NSFC funding, since 2004 to 2007, various but deeply R&D projects were organized: focusing on "scientific application", following the route of "technology for science", the China-VO architecture with VO-DAS as kernel was determined; through VO-DAS platform, seamless access to catalogs, images and spectra located locally or remotely was achieved; a series of VO-compliant tools were developed and released including VOFilter, VO_IMPAT, FitHAS, SkyMouse, et al.; several domestic and international conferences including China-VO annual meeting series, IVOA interoperability meeting, IVOA small projects meeting, were organized and hosted; a dozen of graduated students brought up; tens of scientific and technological papers, popular articles and news reports were published in professional journals, educational magazines and websites. Work finished by the project team demonstrated the feasibility and scientific value of the construction of the China-VO. Furthermore, these achieves will act as important groundwork for future routinely operated VO system. This report is a comprehensive review of the project.

Keywords: Virtual Observatory, Astronomical Information Technology, Testbed, Research & Development

二、研究计划要点及执行情况概述

资助计划任务书中明确指出“根据国内的实际情况，China-VO 整个系统的研发重点将放在 IVOA 标准的应用和实用工具的开发方面，具体包括五个领域：China-VO 系统平台建设、国内外天文数据的统一访问、现有天文工具的虚拟天文台集成、天文设备的虚拟天文台集成、虚拟天文台基础上的公众教育。本课题，中国虚拟天文台试验系统研究与开发，的两大主要目标是：基于网络基础上的可运转的 China-VO 试验平台、按照 IVOA 的标准实现国内外主要天文数据资源的统一访问。”

从 2004 年下半年到 2007 年底，在项目执行的三年半的时间里，我们恪守任务书中明确的几大任务，同时密切关注国内外虚拟天文台和信息技术领域快速发展的事实，依据项目组的真实进展和人员情况及时调整研究开发计划和具体内容。经过三年半的努力，我们认为已经圆满甚至超额完成了资助计划任务书中规定的任务。对照上述任务书中列出的五大研究主题，我们的成果具体汇总如下。

China-VO 系统平台建设：在 China-VO 系统设计（见崔辰州论文）中提出的中国虚拟天文台体系结构的基础上，项目组采用了“以科学引领技术”的研发路线，先后经过“银河系化学丰度梯度研究及其在 Globus Toolkit 环境下的实现（见崔辰州、李长华论文）”、“利用 TwoMASS 星表搜寻 OB 星协研究银河系旋臂结构及其在 Web 服务环境下的实现（见刘波论文）”、“利用 SDSS DR5 测光数据寻找银河系子结构研究银河系结构及其在 VO-DAS 下的实现（见刘超论文）”三个科学范例的开展和研究，最终找到了一套适合中国虚拟天文台的技术路线，基本确定了 VO-DAS 在未来中国虚拟天文台系统中的基础平台地位。

国内外天文数据的统一访问：该主题上的研究同样也经历了数次探索。随着国际虚拟天文台联盟数据访问标准逐步的推出，我们的研究也先后经历了 ConeSearch、SkyNode、VO-DAS、SSA&SIA 等阶段。从最初对 ConeSearch 服务的学习和实践（见桑健论文），到对 US-VO SkyNode 的跟踪与部署（见刘波论文）以及星表数据自动化处理系统的开发（见路勇、高丹论文），再到 OGSA-DAI 基础上 VO-DAS 的设计与开发（见刘超、田海俊论文），以及刚刚完成的 SSA 在 VO-DAS 上的实现（见杨阳论文）。通过对 VO-DAS 在技术和科学上的论证和

测试,我们也基本确定它将作为未来中国虚拟天文台的数据访问环境,为国内外的虚拟天文台用户提供对星表、光谱和图像等不同格式数据的统一访问服务。

现有天文工具的虚拟天文台集成:这也是一个典型的技术探索的过程。在项目的早期,我们试图通过对已有工具的 Grid Service 封装来实现其与虚拟天文台的集成,比如尝试了对 DS9、ImageMagick 的封装(见王晓倩论文)。也试图参考现有的软件功能重新开发应用软件,比如参考 Aladin 开发 VOIMPAT 图像处理工具等(见王丹、邵慧娟论文)。英国虚拟天文台项目 AstroGrid 在自己开发需求基础上推出的 PLASTIC 为已有工具的集成提供了全新的解决方案:从单纯的追求技术接口上的相同到实现数据、消息上的互通。通过 PLASTIC,项目组实现了 VO-DAS 与 VOPlot、Aladin、Topcat 等 VO 工具的集成(见刘超、杨阳论文),还实现了 VO-DAS 与传统桌面科学软件 MATLAB 的互操作(见刘超论文)。下一步,China-VO 将通过 PLASTIC 把更多的应用工具集成到 China-VO 的系统中,为用户提供丰富多样的选择。

天文设备的虚拟天文台集成:坦白的说,天文硬件设备的集成不是非常符合虚拟天文台的宗旨,对硬件的集成并不会给虚拟天文台的用户带来太多用处,很大程度上是作为一项功能进行展示。虚拟天文台的宗旨是“充分挖掘现有数据的科学价值”。当然,我们在这个领域也与其他项目一起联合开展了一些工作,比如与国家天文台兴隆观测基地共同承担了中国数字博物馆公众天文台科学体验区的建设,实现了网络环境下对国家天文台兴隆基地三台望远镜的远程操作(见 <http://pubobs.bao.ac.cn>)。换个角度,虽然直接把硬件设备接入 VO 意义不大,但把与硬件系统相关联的数据处理软件,以及硬件产出的观测数据纳入到 VO 中是非常必要的。所以,China-VO 与 LAMOST 保持了紧密的合作,正在努力把 LAMOST 的 Pipeline 集成为 VO 服务,按照 VO 的数据访问标准发布 LAMOST 的科学产品。

虚拟天文台基础上的公众教育:丰富的资源和网络化的访问是 VO 的绝对优势,这注定了它会在天文科普教育上发挥作用,“让普通公众有机会和天文学家一样做出一流的科研成果”。在 IVOA 一系列标准协议基础上推出的 Google Sky 和微软 World Wide Telescope 定会在全球掀起一轮天文科普的高潮。China-VO 也在公众教育方面开展了许多工作,取得了很好的科学和社会效益。

这其中包括能够真正提供上下五千年的“万年历”——天文星历计算服务（见刘高潮、田海俊论文），网络探究式教学研究（见吴娟论文），通过鼠标取词实现对全球天文信息进行智能化访问的 SkyMouse（见孙华平论文），及时更新的中英文双向天文学词典数据库，以及国内最大、资源最丰富的天文科普平台——宇宙驿站等。

除了上面资助项目计划书中规定的动作，在项目执行期间，我们还完成了许多漂亮的“自选动作”，根据研究的需要设计开发并推出了多个应用工具和服务，比如多星表交叉证认服务（见高丹论文）、VOTable 数据格式转换工具 VOFilter、FITS 文件数据库化管理辅助工具 FitHAS（见李文论文）等。此外，项目组还在测光红移算法（见王丹论文）、多波段数据自动分类（见张彦霞、李丽丽论文）、数据挖掘技术（见张彦霞、刘超论文）等方面进行了深入的研究和探索。

综上所述，中国虚拟天文台试验系统研究与开发项目组经过三年多的努力和奋斗，圆满完成了计划书中规定的各项任务，从技术上和科学上对虚拟天文台的发展进行了全面和深入的探索，找到了一条适合国内网络化天文学研究的发展道路，为下一步中国虚拟天文台的全面建设和服务奠定了基础。

三、研究工作主要进展和所取得的成果

《中国虚拟天文台试验系统研究与开发》项目的执行时间为 2004 年 6 月至 2007 年 12 月，在为期三年半的时间里，项目组开展并完成了大量的技术探索、服务开发和科学研究工作。下面我们就采用“一事一议”的方法对项目执行期间完成的工作做一全面系统的回顾。

1、China-VO 系统平台建设

国际上虚拟天文台的研究方向大致可以分为两大类型：一个方向是面向应用，即利用现有成熟技术和标准开发适用于某项天文学研究特点的应用产品；另一个方向是全面制订互操作协议标准，从根本上建立虚拟天文台的规范和框架。两个方向并不矛盾，而是相辅相成地发展的。基于虚拟天文台的应用集中解决天

文学研究中遇到的一些具体的技术瓶颈，试图通过引入虚拟天文台的概念，通过互连网络实现天文数据资源的高效、无缝地访问，同时实现数据分析以及数据可视化。另一个方向是以 IVOA 组织为代表，致力于标准的制订，规范的确立。根据项目组以及国内在天文信息技术领域的基础积累情况，我们选择了面向应用的研究方向。

在中国虚拟天文台系统平台研究方面我们采用了“以科学引领技术”的研发路线。以完成具体的天文研究课题为目的，分析课题实施所需要的功能，在虚拟天文台理念指导下为这些功能实现寻求技术方案，开发相应的工具和服务。通过开发满足一个个科学需求的服务，不断积累和丰富应用工具和服务资源，逐步形成中国虚拟天文台平台体系。崔辰州的博士论文作为中国虚拟天文台研究开发的基础文献，给出了中国虚拟天文台的体系架构和路线图。以此为基础，项目执行期间先后经过“银河系化学丰度梯度研究及其在 Globus Toolkit 环境下的实现”、“利用 TwoMASS 星表搜寻 OB 星协研究银河系旋臂结构及其在 Web 服务环境下的实现”、“利用 SDSS DR5 测光数据寻找银河系子结构研究银河系结构及其在 VO-DAS 下的实现”三个科学范例的开展，我们最终找到了一套适合中国虚拟天文台的技术方案，基本确定了 VO-DAS 在未来中国虚拟天文台系统中的基础平台地位。

1)、中国虚拟天文台的体系结构

从基本的功能模块角度来分析，China-VO 系统结构组成如图 1 所示。整个体系结构分为四层，从下到上依次是构造层、资源层、汇集层和用户层。

构造层是整个虚拟天文台系统的资源基础，其中包括各种数据资源，计算资源，网络资源，存贮资源等。各种数据资源在虚拟天文台这样一个数据密集型在线研究平台中占有非常关键的作用，是 VO 成功运作的基础和前提。它主要包括星表、星图、光谱、时序数据、计数测量数据、模拟数据、多媒体数据、天文文献等。

资源层将以开放网格服务架构（简称 OGSA）为基础，配合其他网格系统服务工具，利用标准的数据模型和服务模型，通过抽象化实现统一的数据访问和统一的计算访问以及网格系统管理等功能。前面提到的数据访问层的功能在这部

分实现。这里，系统管理主要涉及作业管理、安全管理、资源状态管理、数据管理等。

汇集层包括最能体现天文特色的各种 VO 服务，比如数据处理、数据挖掘、统计分析、可视化等应用服务。当 OGSA 体系架构及其实现工具成熟以后，这些服务的开发和发布将是 VO 建设的重点。

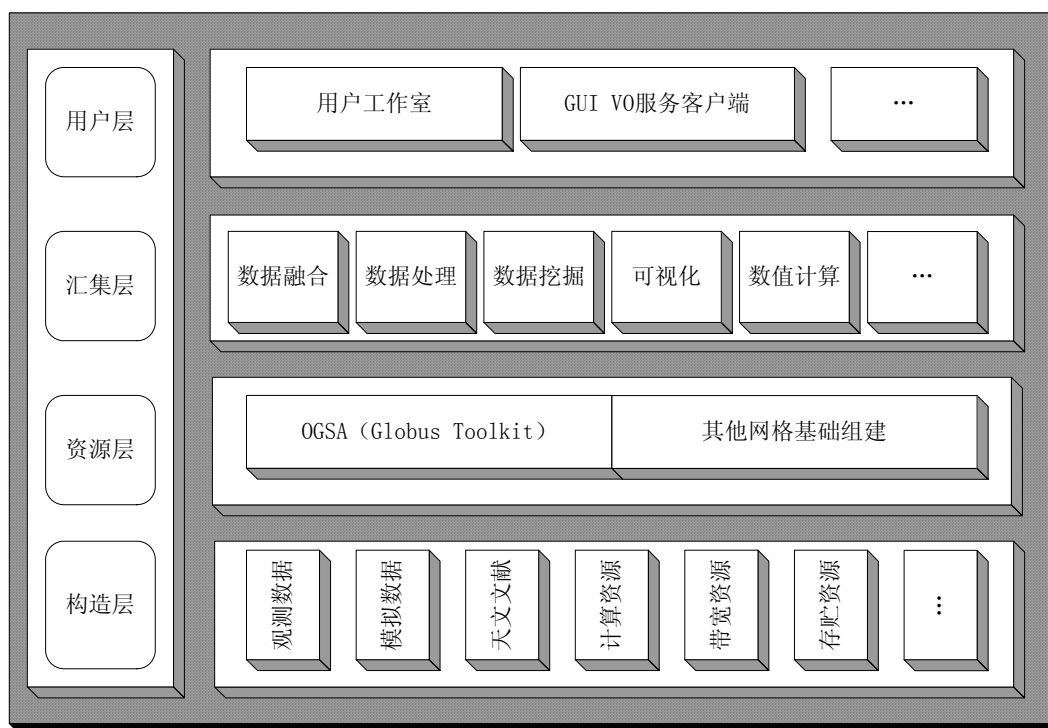


图 1 中国虚拟天文台系统结构

用户层，包括 VO 客户端服务和 VO 门户，是整个体系的最高层，直接与虚拟天文台用户接触。用户层的基本职能是用户任务提交和处理结果返回，主要功能包括用户登录、身份认证、VO 资源浏览、任务编制和提交、偏好设置等。

China-VO 的体系结构建立在 OGSA 的基础之上。物理上，整个系统是分布式的，在网络环境下实现的；逻辑上，通过网络操作系统的管理，它是一个统一的整体。

体系结构可以从多个方面来设计。现在，我们换一个角度来审视整个 VO 系统。其实在上节中我们已经提到了许多 VO 需要实现的功能，比如数据访问、数据处理、数据互操作、资源发现等。为了进一步明确 VO 的核心服务以及对网格平台的要求，下面我们按照 OGSA 面向服务的设计理念来重新分析 VO 系统。

OGSA 是一个面向服务的体系。在整个网格环境中所有组件都是以服务的

形式来体现的。服务可以是不依赖于其他服务而独立存在的原子服务，也可以是建立在其他服务之上的复合服务。不管是原子服务还是复合服务，它们最基本的共同点就是在网格环境中可以提供某种功能。

根据上面两节内容的阐述，我们将 VO 系统中需要用到的基本服务或功能综合为如下几个大类：

数据访问

- 数据库的简单访问：SQL92 直接支持的访问，比如简单的数据库插入、更新、检索；
- 数据库的高级访问：比如统计分析型访问、计算型访问、交叉认证等；
- 文件访问：单个文件的访问，文件系统的访问；
- VOQL 的生成、解析和优化；
- 数据整合：不同数据集查询结果的整合；
- 数据迁移：不同数据类型（数据、图像、流媒体）利用不同的迁移机制（GridFTP、可靠文件传输服务 RFT）在不同地点或不同服务间的传输；

数据挖掘

- 分类：目的是提出一个分类函数或分类模型，利用该模型把数据库中的数据项映射到给定类别中的某一个。
- 聚类：根据数据的不同特征，

将其划分为不同的数据类，使得属于同一类别的个体之间的距离尽可能的小，而不同类别的个体间的距离尽可能的大。

- 相关分析：目的是发现特征之间或数据之间的相互依赖关系。
- 偏差分析：发现观测结果与参照量之间有意义的差别。通过发现离群数据可以发现一些不同寻常的或奇异的天体。

计算服务

- pipeline 处理：海量数据的自动处理是 VO 应该提供的
- 光谱处理：包括光谱的自动分类和自动测量
- 动力学计算：比如星系演化、日月食、天体轨道的计算等
- 数值模拟计算：N 体模拟，宇宙大爆炸过程模拟等

可视化服务

- 二维和三维空间的散点图、

直方图、曲线图、轮廓图，

馅饼图

- 一维和二维光谱显示
- 常用天文图像的显示
- 动画模拟
- 流媒体播放

数据转换

- 数据编码、数据解码
- 物理单位换算
- 坐标换算
- 历元换算
- 图像格式转换
- 数据压缩与解压缩

注册与发现

- 数据集的注册与发现
- 应用服务的注册与发现
- 物理资源（计算资源、存储资源、网络资源等）的注册与发现
- 数据路由
- 服务路由

元数据服务

- 天文学本体
- 元数据目录
- 国际化多语种支持

安全服务

- 身份认证

- 访问授权

- 单点登陆

- 访问代理

- 安全策略

资源管理

- 状态管理

- 资源动态上下线管理

- 数据备份

- 数据缓存

作业调度

- 作业估计

- 负载均衡

- workflow 管理

- 断点管理

- 服务的重新获取

系统监测

- 系统日志

- 性能监控

- 日志分析

- 系统通告

- 用户通告

MyVO（我的 VO）

- 活动日志

- 任务编制

- 中间结果和常用数据存储

- 及时通告

- 偏好设置

这其中有些功能可以由符合 OGSA 标准的 Globus Toolkit 为代表的网格操作系统来提供，比如资源管理、作业管理、系统检测、安全服务等，其余则需要

由 China-VO 提供。不过由 China-VO 提供的服务还可以分为两种情况：某些功能服务是天文学研究所特有的，此类服务必须由 China-VO 开发者独立提供，比如元数据服务、数据交换格式编码、pipeline 计算等；而还有一些功能服务是许多科学研究领域甚至非科研领域所共同需要的，比如可视化、数据挖掘、统计分析等。对于这类服务，China-VO 将借鉴其他领域和行业的现有工具和实现方案，经过改造加工后融入到 China-VO 中。

China-VO 是一个数据密集型的在线研究平台。为了实现数据密集型在线研究这个目标，它必须实现三个方面的基本功能：数据访问、数据处理、数据互操作。这三个基本任务在 VO 系统中可由三个角色来承担：数据服务提供者（Data Service Provider, DSP），应用服务提供者（Application Service Provider, ASP）和注册（Registry）。把上面列出的 VO 系统的基本功能服务按照所处角色的不同将其分配到 DSP、ASP 和 Registry 中，便得到 VO 的服务模型。图 2 给出了 China-VO 的服务模型。

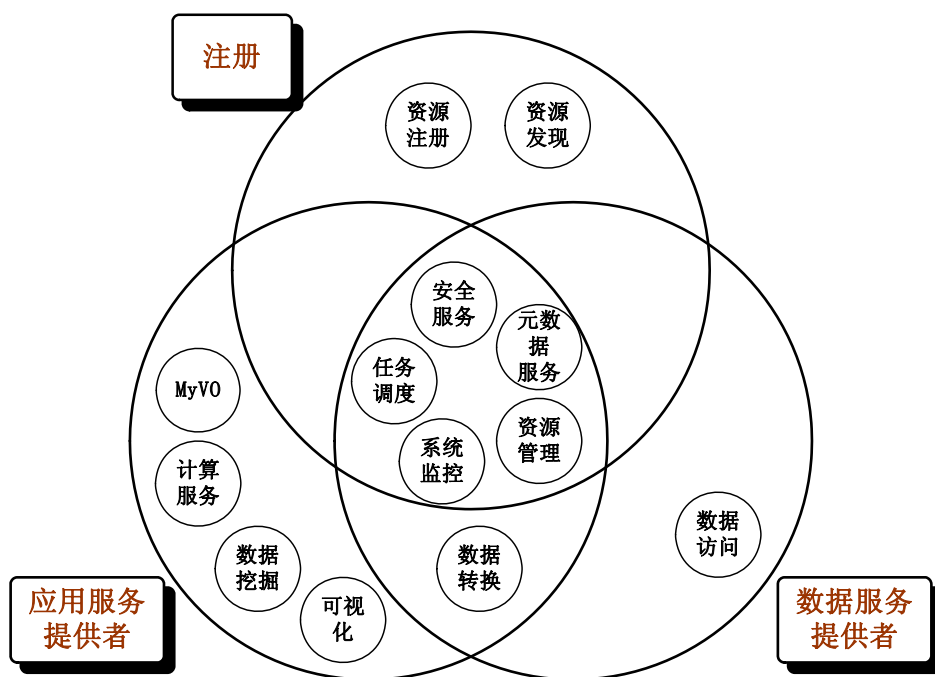


图 2 中国虚拟天文台服务模型

从模型中可以看出，资源管理、系统监控、安全服务、任务调度对所有角色都是必须的。但这部分对所有网格应用都是必须的，将由 OGSA 网格操作系统来提供。

虽然元数据服务对所有角色也是必须的，但由于这部分涉及许多天文学的学科特点和内容，所以 China-VO 将参考 OGSA 体系提供的元数据服务机制结合天文学的需求给出 China-VO 的元数据服务。资源注册与发现、数据访问、计算服务、数据挖掘、可视化、用户工作室等服务则将由 VO 社区独立开发或者借鉴其他学科领域的实践经验来实现。

除了上面这些 VO 服务，还有两个比较特殊的部件，即 VO 门户和客户端程序。它们可能不以网格服务的形式存在，而利用现有的 WWW 技术，比如 Java Servlet、Java Applet、PHP、ASP 等为基础开发。

2)、银河系化学丰度梯度研究及其在 Globus Toolkit 环境下的实现

银盘上化学元素的丰度梯度对研究银河系特别是银盘的形成与演化过程有着重要的意义。丰度梯度的时空变化趋势是银河系星际介质增丰历史以及内落、外流等过程的反映，是建立银河系化学演化模型的重要约束。按照银河系是从内到外形成的观点，银盘内部区域的密度比外部区域高，内部区域的恒星形成和化学演化速度都要比外部区域快。因此，内部区域的化学丰度要比外部区域高——即表现出丰度梯度。然而，银河系演化过程中有些效应（如银河系的动力学演化、恒星本身的运动、恒星与巨分子云的相互作用）会影响化学元素的分布，从而将这种丰度梯度抹平。总而言之，银河系或银盘上是否存在化学元素的丰度梯度，以及梯度随空间、时间的变化不仅为银河系化学演化模型提供重要限制，而且与银河系形成机制、银河系发生的动力学过程直接相关。

项目组利用中科院国家天文台陈玉琴等通过国家天文台 2.16m 望远镜对 90 颗 F、G 型星样本高分辨率、高信噪比光谱观测得到的样本的 Fe 丰度数据以及依巴谷星表提供的距离和动力学数据在 Globus Toolkit (GT) 网格环境下复现了崔辰州硕士论文的研究工作。在 Globus Toolkit 环境下实现了包括注册服务、数据服务、恒星轨道计算服务和可视化服务等网格服务，组成一个简单完整的网格应用系统，整个系统的拓扑结构如图 3 所示，系统分析结果显示如图 4 所示。

上述工作是在 2003—2004 年进行的，由于当时 GT 还相当的不成熟，GT4 刚刚推出，这为其上应用系统的开发带来很大困难。当时国内对 GT 的跟踪基本

上是学习和试用。我们的上述工作虽然实现了一个简单的研究过程，但也仅限于技术上的测试。限于 GT 本身的不成熟和多变，这项工作后来被 VO-DAS 的开发所取代。

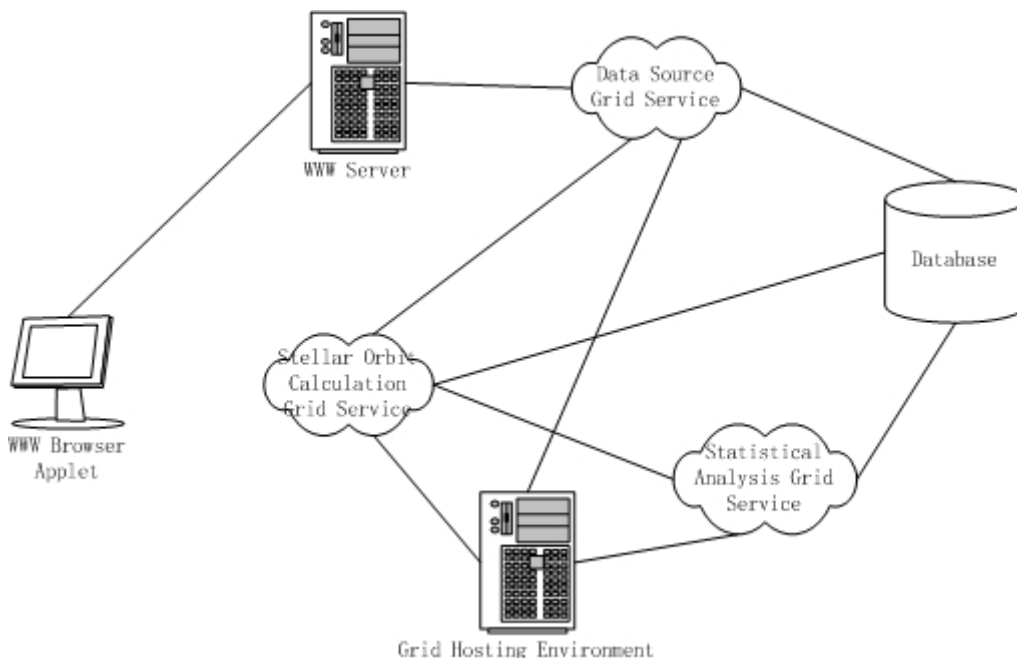


图 3 网格环境下银河系铁丰度梯度统计分析应用系统拓扑结构

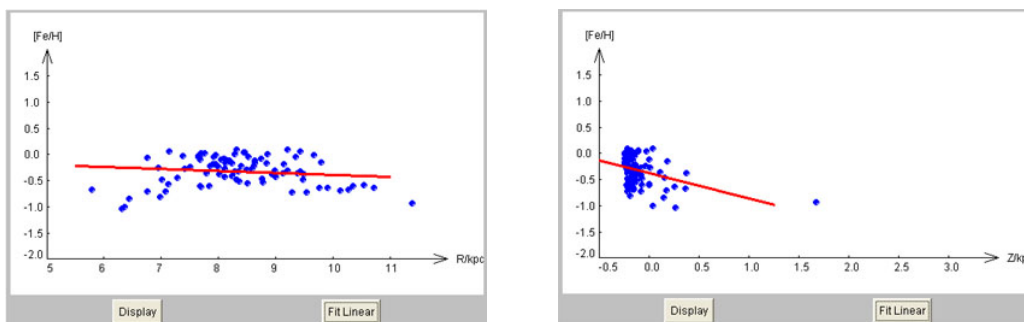


图 4a 银盘径向恒星样本铁元素丰度分布及线性拟合

图 4b 银盘法向恒星样本铁元素丰度分布及线性拟合

3)、利用TwoMASS星表搜寻OB星协研究银河系旋臂结构及其在Web服务环境下的实现

“利用 2MASS 星表寻找 OB 星协，研究银河系的旋臂结构”是项目组与国家天文台天文学家紧密合作开展的真实的研究课题。此科学范例的基本数据流程如图 5 所示，得到的银河系旋臂结构的初步结果如图 6 所示。

为了寻找隐藏在浓密气体和尘埃之后的 OB 星，我们需要使用 2MASS 星

表进行筛选。然而由于消光的影响，以及银盘上随着银经变化恒星分布的不均匀性，我们很难确立一个简单的 OB 星判断标准。在这个研究过程中，我们需要反复实验不同条件的效果。这就需要反复对 2MASS 星表进行海量数据查询，查询数据高达 2×10^8 之多。由于当时没有 VO-DAS，只有作为原型的 SkyPortal 作为数据查询工具，我们的数据访问工作非常费时，需要将数据分成很多部分，分别查询以后再进行合并。由于科学上的复杂性和技术上的准备不足，这个范例搁浅在了数据获取这个阶段。

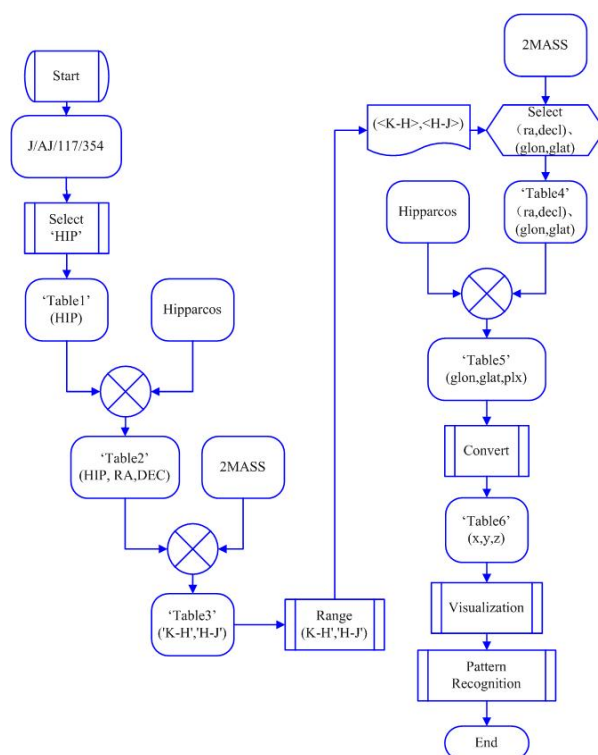


图 5 银河系旋臂结构研究科学范例基本数据流图

不过，通过此范例的研究开发，项目组给出了一个中国虚拟天文台数据挖掘系统的可行框架，如图 7 所示。提出了任务描述语言（JDL）和天文计算单元（CompuCell）的技术方案。JDL 是项目组根据虚拟天文台数据挖掘的特点设计的一种高级计算语言，它继承了著名的 MatLab 语言的基本语法，将数据挖掘计算环境扩展到分布式的网络环境中。它的结构见图 8。JDL 提供了一种统一的用户操作前端，借助 JDL，用户可以操作各种共享在网络中的数据查询服务和各种数据分析算法服务。在这个系统中，JDL 处于核心位置，任何一项数据操作都是由 JDL 描述的。JDL 把用户同复杂的网络服务隔离开，用户不需要知道数据存放在什么地方也不需要知道计算服务运行在哪里，只要提交自己的 JDL 程序给

系统，系统就会自动完成复杂的网络分布式数据查询和数据挖掘工作。最终，用户会直接得到一个结果数据文件，在其计算机上应用中国虚拟天文台提供的 Portal 软件进行多种类型的可视化展示。

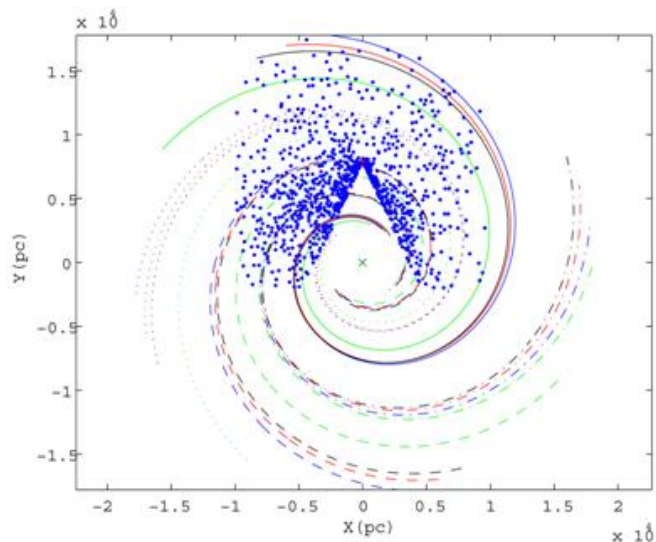


图 6 银河系旋臂结构研究范例初步结果

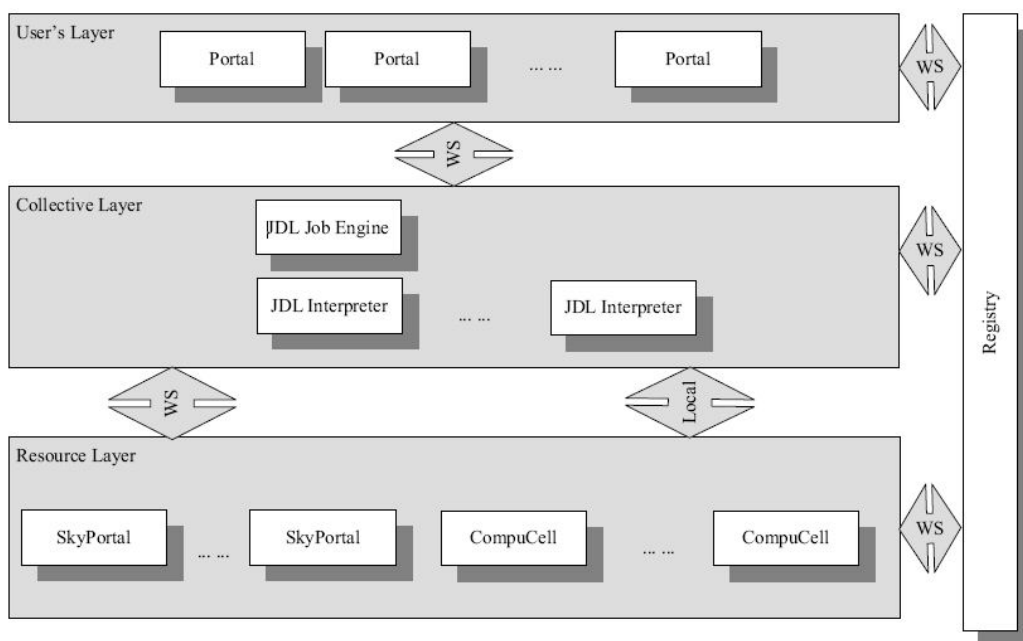


图 7 基于 JDL 的天文数据挖掘工具原型结构

这个范例给出的中国虚拟天文台数据挖掘系统如图 7 所示。其中，JDL 解释器负责对用户提交的 JDL 程序进行解释，如果发现需要查询数据，就会在网络上通过注册（Registry）服务找到合适的 SkyNode 服务，并要求它提供相应的数据查询服务。如果 JDL 解释器发现需要做一个复杂的数据分析计算，就会在

网络上通过 Registry 服务找到一个合适的 CompuCell 服务, 将这个计算任务转交给这个 CompuCell 服务完成。CompuCell 服务是项目组设计的一个通用数据计算服务组件。它和 JDL 解释器之间通过 CompuCell 接口进行通讯, 它封装了某种数据计算工具。该系统中, 一个 CompuCell 服务就会提供一种类型的数据分析算法, CompuCell 越多, 提供的数据分析算法也就越多。可以把 CompuCell 看成是 JDL 的功能扩展, 这种扩展能力是无限的。CompuCell 服务是一个 Webservice 服务, 一般不会和 JDL 解释器安装在同一台机器上, 而是可以安装在任何一台连通了网络的服务器上。

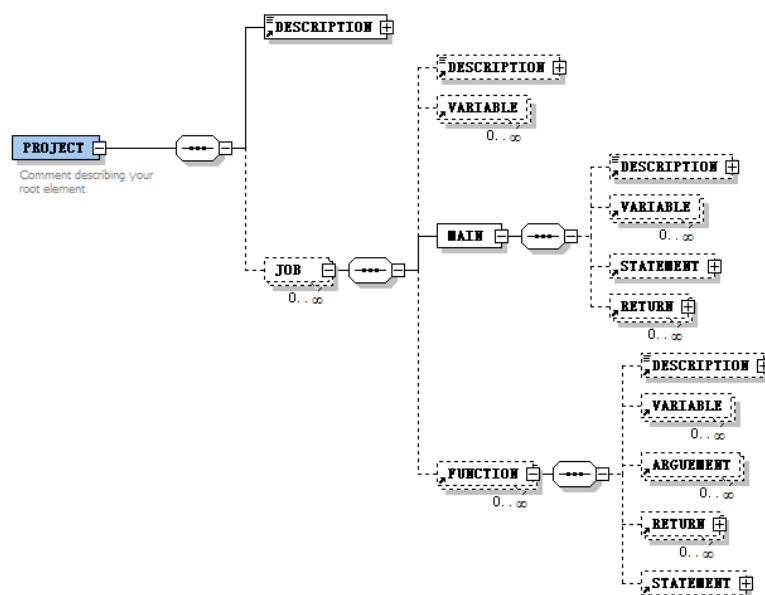


图 8 JDL 文档结构

经过项目组的调研分析, 在虚拟天文台数据和服务的注册和发现机制方面, 项目决定采用由美国虚拟天文台项目开发的开放源代码的解决方案: CARNIVORE。本项目按照自己的要求对这个系统进行了本地化和定制。

JDL 和 CompuCell 两项我们自主提出的技术一起构成了该系统的主框架。在这个主框架上, 项目组可以一方面不断完善作为客户端的 Portal, 让系统更加方便好用, 另一方面不断扩展 CompuCell 服务, 集成现有的数据挖掘和统计分析工具, 让系统的功能越来越强大, 能做的科学研究范围越来越宽。

通过这个 JDL 原型, 我们可以得出如下结论: 首先, JDL 的思路是新颖的, 它从功能上类似于 workflow 描述语言, 如 BPEL4W, 但是语法上更接近计算性的语言。这样有助于将基于网格的工作流式的处理方式同数据挖掘计算相结合。其

次，我们只使用了 Web Service 就成功让这样一个流程运转起来说明如果加上 WSRF，那么 JDL Job Engine 的结构设计会进一步简化，系统的性能会有更大提高。第三，我们也发现这样的运行结构执行效率非常低，特别是算法简单的情况下更加突出。此外，数据处理量也因为 JDL 的运行限制而不能很大。最后，我们认为 JDL 的思路是可行的，但是如果采纳这样的思路完成一个数据挖掘工具，带来的最大问题是工作量的激增，主要体现在 JDL 解释器的编写和 CompuCell 对各种算法的封装等。

截止到 2006 年春天，项目组完成了 JDL Alpha 阶段的开发工作，完成了 CompuCell 的接口规范定义。但是，受到项目经费、开发力量的限制，这方面的探索暂时停止在了这个层面。

4)、利用SDSS DR5 测光数据寻找银河系子结构研究银河系结构及其在VO-DAS下的实现

美国 SDSS 数据巡天项目释放的 DR5 的数据产品中提供了 u, g, r, i 和 z 五个波段的 8000 平方度的测光数据，可以通过这些数据利用其大样本的优势研究银河系银晕中的子结构，比如矮星系等。在这个范例中，我们选择其中的点源。这些点源包括恒星，类星体和非常暗的分类错误的星系。我们首先要从如此浩繁的天体中找出可能的有矮星系的位置。矮星系所在的区域恒星密度应该比周围大。但是如果矮星系非常暗，恒星密度很低，就不能从前景场星中区分出来。通过特定星等和颜色剪裁，突出矮星系的成员星的特征而压制场星的影响，就可能让矮星系从背景中突出出来。我们选择的星等范围是 $19 < i < 22$ ，选择的颜色范围是 $0 < g - i < 1$ 。这里的 g 和 i 星等都已经经过了消光改正。取这样的星等范围就可以有效避免厚盘恒星的干扰，而颜色选择在这个范围可以避开更红的薄盘晚型星和更蓝的类星体。在这个颜色范围内，晕星和矮星系成员星的主序星，亚巨星和红巨星都分布其中。削弱了薄盘和厚盘场星的影响以后，天空投影的恒星数密度主要贡献来自晕星，而晕星的密度较低，这就有可能让矮星系从背景中突出出来。为了避开 Sgr stream 和 Virgo overdensity 的影响，我们将考察的区域设定在 $120^\circ < \text{赤经} < 270^\circ$ ， $25^\circ < \text{赤纬} < 70^\circ$ 。

为了完成上述范例中数据访问以及后续数据处理和统计分析的需求，我们

设计开发了虚拟天文台数据访问服务 (VO-DAS), 结合 IVOA PLASTIC 协议, 集成 MATLAB 等应用环境, 给出了第三套中国虚拟天文台试验系统的技术方案。

(a) VO-DAS 框架

图 9 中描述的 VO-DAS 的总体结构, 它可以分成五个模块。图中上半部分最大的模块是 VO-DAS 服务器的结构, 最下面的模块是 DataNode 的结构。中间的两个模块分别是 VO Registry 和数据存储服务器 (图中以 VOspace 作为例子)。此外还有一个没有在此图中明确表示出来的客户端模块。

VO-DAS 服务器是 VO-DAS 系统的核心模块, 它起到承上启下和调度的功能。为了能够支持异步查询功能, 我们采用 WSRF 构架的 Web Service 方式。因而 VO-DAS 服务器必建立在支持 WSRF 的中间件上面。我们采用的中间件是 Globus Toolkit 4.0 Java WS Core。这是一个纯 Java 的服务平台, 和其他的网格服务有相对的独立性, 既可以独立运行, 也可以安装到一个网络服务的容器 (如 Tomcat) 下运行。为了能够提供安全的服务, VO-DAS 需要有用户身份认证的机制, 因此在 GT4 WS Core 之上是 Authorization 模块。认证功能也使用 GT4 自身的功能模块。由于 VO-DAS 会与 DataNode、Registry 以及 VOspace 通信, 因此需要在其内部有对应这三个部分的客户端模块, 分别是 OGSADAI Client、Registry Proxy 和 VOspace Client。OGSA-DAI Client 是一个标准的 OGSA-DAI 调用客户端模块, 负责将 SQL 查询和结果数据处理整理成 OGSADAI 支持的 Request 对象, 然后将请求送达指定的 OGSA-DAI 服务。Registry Proxy 负责向指定的 Registry 查询网络中有哪些 DataNode, 获得这些 DataNode 的元数据信息, 并保存在 DataResourceMap 对象中。DataResourceMap 是一个数据结构, 它是一个 DataResource 对象的数组。每个 DataResource 对象代表一个数据资源。DataNode 可以包含一个或多个 Resource, 具体数目由 DataNode 的元数据 XML 文件中 Resource 的数目决定。每个 Resource 元素对应 DataResourceMap 数组中的 Resource 对象。VOspace Client 目前是一个 FTP 客户端, 负责和存储服务器 VOspace 交换数据文件。

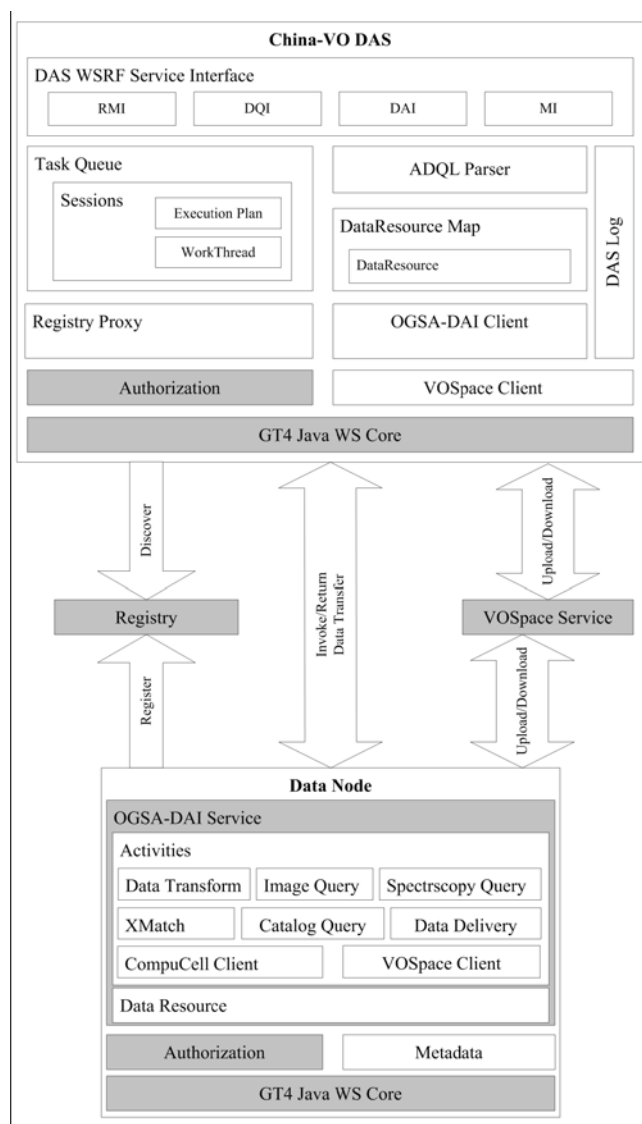


图9 VO-DAS 的设计结构

VO-DAS 服务器提供给客户端访问的接口有四类: 资源元数据接口(RMI)、数据查询接口(DQI)、数据存取接口(DAI)和管理接口(MI)。

在定义 VO-DAS 功能的时候,我们要求 VO-DAS 支持 ADQL 语言的查询。由于 DataNode 中的 OGSA-DAI 只支持 SQL 查询而不支持 ADQL, 我们需要将客户端传递来的 ADQL 转换成功能一致的 SQL。这个工作要在 VO-DAS 里面完成, 完成这项任务的模块是 ADQLParser。ADQLParser 支持 ADQL ver0.9 版本的语法, 同时支持我们所做的对 ADQL 语义的扩展以增加图像和光谱的查询。

VO-DAS 服务器具有任务调度功能, 这项功能的实现是在 Task Queue 模块中。Task Queue 为提交来的每个任务创建一个 session 对象, 所有的执行都是在 session 对象内完成的。Task Queue 管理将一个 session 对象组成的任务等待队

列和一个任务执行队列，以及一个任务完成队列。任务首先停留在等待队列中，当有空闲资源的时候，它会被调度到正在执行队列，并分配一个工作线程给它。在得到工作线程以后，session 会调用 ADQL Parser 解析 ADQL 语言并创建一个 Execution Plan 以向指定 DataNode 发送查询请求。当查询工作完成以后，session 会从正在执行队列中转移到完成队列。

由于各个模块之间的相互联系复杂，系统行为不确定性很大，因此需要有日志管理的功能模块用来记录系统的操作细节。这既便于查询系统历史上所作的操作，方便维护，也有利于开发周期内的调试和测试。这项功能由 DASLog 完成。

DataNode 基本沿用了 OGSA-DAI 的结构，但为了我们的一些特殊功能做了扩充。首先，Catalog Query 就使用 OGSA-DAI 提供的数据库查询功能实现。同时增加 XMatch，即交叉证认功能模块。其次，增加 Image Query 和 Spectrum Query 两个模块实现图像和光谱查询的功能扩充。Data Transform 会支持天文数据文件的特定格式，包括 VOTable 和常用的 ASCII 文件。Data Delivery 和 VOspace Client 联系 VOspace 将查询结果文件发送到指定的存储服务器上。CompuCell Client 是为数据挖掘功能预留的接口。

Registry 和存储服务器不需要我们实现，而是采用现成的产品。Registry 使用 AstroGrid 的注册服务在本地建立的一个拷贝，而存储服务器目前采用 FTP 服务器，最终实现 IVOA 的 MySpace 协议。

(b) VO-DAS 的客户端

VO-DAS 服务器提供给客户端四类访问接口，这是最基本的客户端形式，但是需要使用 Java 语言进行网络编程才能够实现。为了扩大 VODAS 的使用群，最大限度的发挥它的能力，我们定义多种不同形式的客户端。下面分别对这些客户端进行定义和说明。

GUI 客户端是采用 Java 实现的操作系统无关的图形界面客户端。它提供了简单易用的窗口界面，在连接上一个 VO-DAS 服务器以后，它会自动从服务器端下载所有数据资源的元数据，并采用树形结构显示出来。它提供了一个简单 ADQL 编辑环境，编辑好的 ADQL 只要按下一个按钮就可以将任务送到服务器端。在 GUI 客户端中，用户可以选择同步查询还是异步查询。如果是同步查询，

那么数据会直接返回客户端，需要用户将数据保存到本地文件中。用户也可以选择通过 PLASTIC 协议将数据传递到其他的 VO 工具软件中，例如 Aladin, Topcat 等。如果是异步查询，那么提交查询请求的同时，用户需要为这次任务起一个容易识别的名字。任务提交以后，在窗口的下方有一个监控窗，监控正在执行的查询任务当前的状态。客户端窗口可以关闭，并不影响任务的执行；重新打开窗口以后，仍然能够继续监控任务的进度。当异步查询的任务结束以后，可以从历史信息中找到异步查询的结果文件 URL。用户可以通过 FTP 客户端将数据下载到本地。

GUI 客户端非常适合数据访问频率不是很大，对计算机操作系统不是很熟悉的用户。

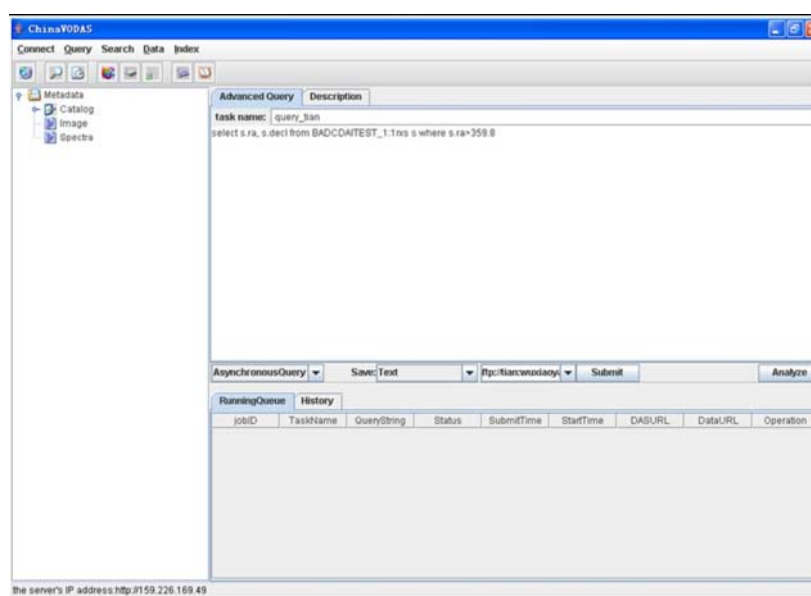


图 10 GUI 客户端的界面

命令行客户端提供了一组基本的操作命令，用户不需要图形终端就可以通过这些命令完成数据的查询。它们是运行在 linux 环境下的一组 shell 命令，在后台调用另外一组支持 Web Service 客户端接口的 Java 程序完成各个数据查询操作。表 1 中列出了这些命令的细节。诸如 VO-DAS 服务器地址等配置信息会存放在一个配置文件里而不需要每次执行查询命令的时候都输入。命令行方式可以集成到用户自己的程序里面，无论程序是 Python 还是 FORTRAN，都能轻易地通过这组命令完成频繁的数据访问。这种类型的客户端适合访问数据比较频繁或者需要使用自己编写的程序完成数据访问任务的用户。需要说明的是，经过一

层封装以后，用户自己编写调用这些命令的程序中并不需要包含 Web Service 服务支持库。这样，即使用户不了解什么是 Web Service 也能够在其程序中使用 VO-DAS 了。

表 1 VO-DAS 命令行客户端的命令集

命令	参数	说明
md		从VO-DAS服务器下载数据资源的元数据XML文件
asyn	ADQL文件, 文件格式, session文件名	向VO-DAS服务器提交一个异步查询任务, 同时这个任务相关的session会保存在参数中指定的session文件中
syn	ADQL文件, 文件格式	提交一个同步查询任务, 查询结果以字符串形式输出到标准输出上
jobmon	session文件名	查询一个已经提交的异步查询任务的执行状态, 返回结果输出到标准输出上
dataurl	session文件名	获得一个已经完成的异步查询的结果数据位置, 返回信息输出到标准输出上
close	session文件名	关闭一个session

由于 MATLAB 支持直接调用 Java 程序包，因此前面所定义的 Java 程序接口可以直接被 MATLAB 所支持。在此基础上，通过使用 M 语言进行一定的封装，我们又得到了 MATLAB 上面运行的客户端。这个客户端的内容和命令行客户端十分类似。所不同的是，shell 命令变成了 M 语言。此外，由于在 MATLAB 中不便于使用配置文件，因此配置参数也是通过一个 M 语言实现的。

由于 MATLAB 还同时支持 GUI，因此在提供了这些查询数据的命令函数以后，再将它们集成到一个 MATLAB 的 GUI 中。这样，在 MATLAB 中既支持了命令方式的查询，也支持了 GUI。在 MATLAB 中实现 VO-DAS 的客户端的最大优点就是它可以将数据查询操作和数据挖掘操作在一个软件之内无缝地连接起来。

另一类重要的客户端形式就是网页表单形式的客户端。这类客户端的用户群和第一类 GUI 客户端的很相似。网页客户端使用更加简单，不需要安装任何应用程序，只要用网络浏览器连接到服务器上就可以了。但是由于网络浏览器的

独特工作原理，这种类型的浏览器支持异步查询比较困难。

(c) VO-DAS 系统的测试与成果

目前项目组已经完成了 VO-DAS 核心功能的开发，并进行了一系列的测试工作。VO-DAS 的主要实现工作集中在 VO-DAS 服务器上，只有少量的工作是在 DataNode 之上，全部开发都是采用 Java 语言完成的。初期版本的开发完成了 ADQL 解析的基本语法部分、GUI 客户端和命令行客户端、VO-DAS 服务器、DataNode 的星表查询和交叉证认等功能，其他功能将在后续版本中实现。

实际测试环境中，VO-DAS 环境搭建在 6 台台式机之上。其中两台 VODAS 服务器，四台 DataNode（其中一台和一个 VO-DAS 服务器共用一台计算机）以及一台存储服务器。客户端分别运行在 Windows XP 和 Linux 环境中。测试内容包括 ADQL 解析器的正确性测试、VO-DAS 调度正确性测试、客户端接口测试、运行稳定性测试、大数据两访问测试等。ADQL 解析器经过测试发现，基本的 ADQL 语法解释正确，但是复杂的语法，如使用 TOP、GROUP BY、ORDER BY 等尚未实现，而 WHERE 条件非常复杂的时候 ADQL 解析还不稳定。对 VO-DAS 调度的测试证实 WSRF 上的任务调度基本达到要求。客户端的接口测试在不同操作系统下均达到设计要求，但 GUI 客户端的工作还不是很稳定。DataNode 的注册和发现工作尚未能达到设计要求，只能通过静态添加方式实现 VO-DAS 对 DataNode 的认知。

运行稳定性测试采用一条能够返回 1000 行左右的 ADQL 查询语句通过多个客户端连续不停向服务器发出请求，检查服务器的承受能力。实际测试发现，高负荷的不稳定来源于 OGSA-DAI 的不稳定性。大数据量访问受限于 DataNode 所在服务器的内存（2GB），一般能够达到 1M 行的数据访问。

测试结果表明，VO-DAS 的初步版本可以在内部小范围内使用，还需要在增加更多功能以后，并让系统更加稳定以后再正式发布。

(d) 中国虚拟天文台试验系统的基本方案

从数据访问到数据处理，再到数据分析，基于 VO-DAS 的中国虚拟天文台试验系统方案在经历了实际使用的考验后，基本上明确了下来。首先，它所使用的数据由 VO-DAS 在一个网格环境中从分布在各处的数据库里获取。数据获得以后，利用 IVOA PLASTIC 协议无缝地传给 MATLAB，利用 MATLAB 丰富的

数值计算工具箱完成数据预处理、数据挖掘和数据可视化工作。为了能够充分利用其他 VO 工具的独特功能，例如 Topcat 的交叉认证，Aladin 的数据联合显示，我们在 MATLAB 上开发了 PLASTIC 协议工具箱，支持 MATLAB 和其他 VO 工具的平滑数据交换。在这些数据交换过程中，除非用户显式地保存数据文件，否则感觉不到数据格式的任何问题。以往数据处理中占去很多时间的数据传输和格式转换工作已经隐藏到了背后，令操作者感觉不到了。当然仅仅依靠 MATLAB 自己携带的工具箱仍然很难满足天文数据挖掘的要求，这就需要在 MATLAB 之上进行功能上的二次开发，提供一个天文专用的工具箱。

在有了技术上的准备以后，天文数据挖掘就可以应用于实际的研究工作了。我们将 VO-DAS 和 MATLAB 整合起来作为一套天文数据挖掘环境应用到银河系晕结构的研究中，借助这套工具的帮助我们从 SDSS DR5 发布的数千万个天体中发现了 5 个矮星系 / 球状星团的候选体。这是我们在中国虚拟天文台的平台上取得的第一个科学结果。虽然这并不意味着中国虚拟天文台以及建在其上的数据挖掘工具已经成熟，可以立即投入使用，但是我们用这个过程很好地证明了方案的可行性。这是中国虚拟天文台建设的一个阶段性标志，因为它已经从早期的方案探索阶段逐渐向实用化方向演进了。

2、国内外天文数据的统一访问

就像对虚拟天文台平台研究的过程一样，该方向的研究同样也经历了数次探索。随着 IVOA 数据访问标准逐步的推出，我们的研究也先后经历了 ConeSearch、SkyNode、VO-DAS、SSA&SIA 等阶段。从最初对 ConeSearch 服务的学习和实践（见桑健论文），到对 US-VO SkyNode 的跟踪与部署（见刘波论文）以及星表数据自动化处理系统的开发（见路勇、高丹论文），再到 OGSA-DAI 基础上 VO-DAS 的设计与开发（见刘超、田海俊论文）。目前，经过对 VO-DAS 的论证和测试，我们基本确定把它作为未来中国虚拟天文台基本的数据访问环境，为国内外的用户提供对星表、光谱和图像等不同格式数据的统一访问服务。下面对上面提到的这些工作逐一简单介绍一下儿。

1)、ConeSearch初体验

桑健等利用基于开放网格服务架构（OGSA）的 Globus 工具包，结合现有的天文数据标准，搭建了多种数据访问服务。首先，设计了星表 Cone-Search 访问接口，并基于不同的星表数据库系统搭建了多个星表的 Cone-Search 服务，这是中国虚拟天文台实现的第一个网格服务，初步实现了不同本地环境下分布式天文星表数据的统一查询访问。证实了利用网格技术解决海量分布天文数据融合的可行性。其次，为了实现更加复杂的查询功能和融合现有的关系型数据库中的数据资源，又设计了星表查询访问接口，引入了 ADQL 作为描述查询要求的统一语言，初步搭建了基于 MySQL 数据库的星表查询访问服务。最后，在天文图像的统一访问上做了积极的探讨，设计了一个简单的天文图像访问接口并实现了 DSS-I 天图访问服务。

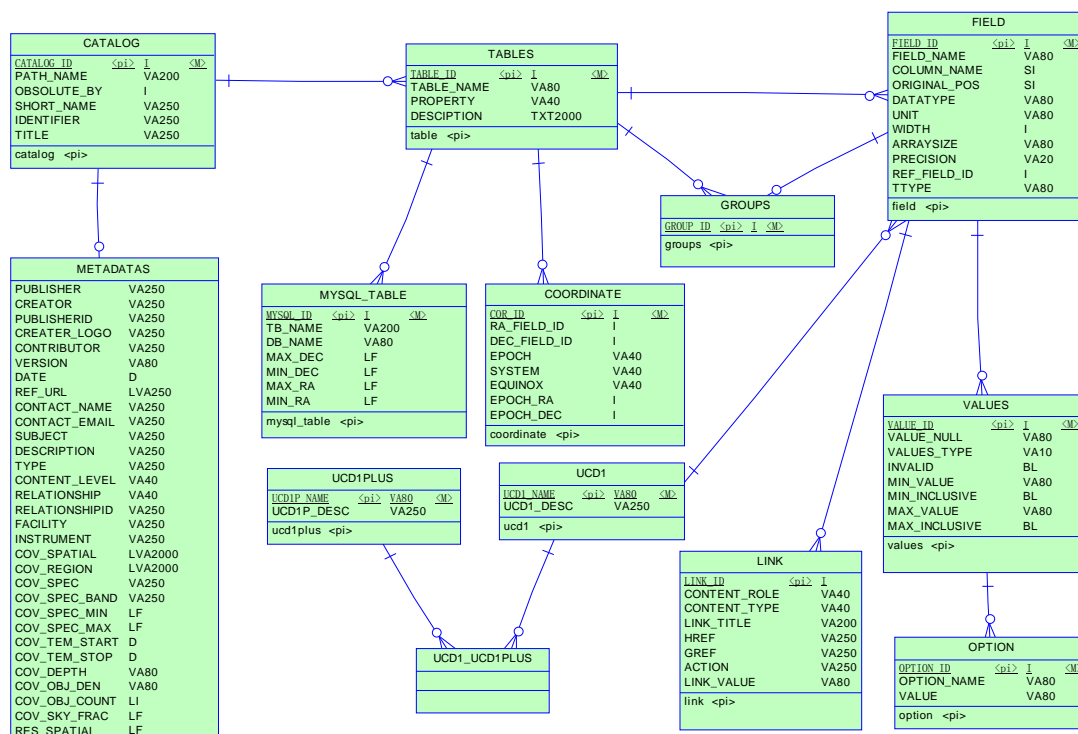


图 11 星表元数据库逻辑设计实体-关系图

在搭建星表锥形检索服务之前我们设计了一套星表数据库方案，包括“数据”库和“元数据”库两部分。“数据”库是由导入的星表数据组成，也称星表数据库。“元数据”库管理收集这些星表的有关元数据信息，也称星表元数据库。这两个数据库之间是数据与元数据的关系。星表数据库的各个表结构是由自动入库程序按照规定的入库规则形成的，其内容不断充实更新，而星表元数据

库结构是采用开发工具预先设计好的，结构不再做大的变动。

星表 Cone Search 服务的创建规则，将每一个数据资源例如星表按照定义的 Cone Search 服务接口封装成一个数据服务。不能将多个星表封装在一个服务里，即使它们存储在同一台主机的同一个数据库管理系统中。由于很多数据中心主机内存储多个数据集，所以一台主机可以提供多个数据服务。由于这些服务的地址名称不同，所以同一台主机上的数据服务间也不会发生冲突。同样，如果一个数据集分布星表在多台主机内，那么也只能基于多台主机部署一个数据服务。

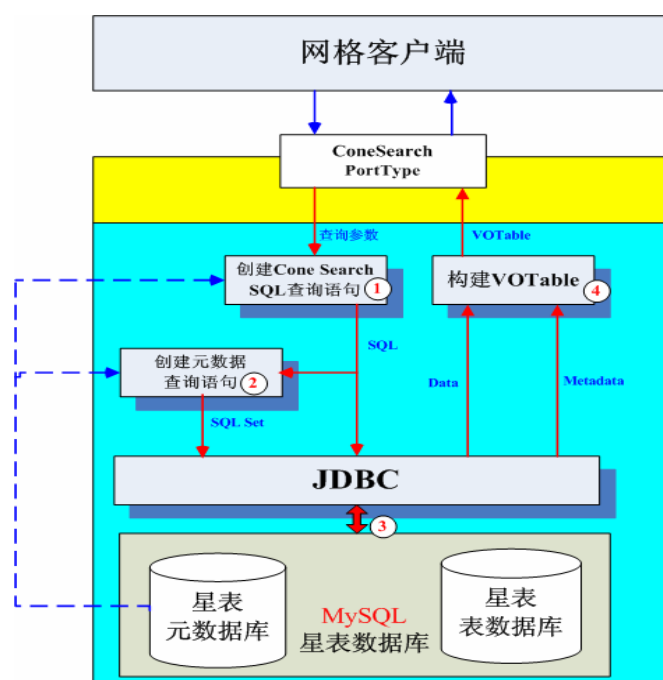


图 12 基于天文星表数据库的 Cone-Search 服务流程

管理星表这样的表列数据，传统的关系数据库软件提供了强大功能，所以现在大部分数据中心将表列数据存储存储在关系数据库管理系统中。我们基于前面提到的天文星表数据库，使用 Java 语言，实现了多个星表的 Cone-Search 服务。

基于天文星表数据库的 Cone-Search 服务流程如图 12 所示，流程具体说明如下：

- a) 将服务名称映射到相应要查询的星表数据库中的星表，然后根据查询元数据库返回的信息生成锥形查询的 SQL 语句；
- b) 根据生成的查询星表的 SQL 语句和元数据库的有关信息，生成查询元数据库的有关星表元数据和列元数据的 SQL 语句集；
- c) 上两步生成的 SQL 语句通过 JDBC 交给 MySQL 数据库去执行；

- d) 将查询得到的元数据信息和星表数据封装成 VOTable 格式文档，并将其转换成数组返回给客户端，在这里我们采用的是 DOM 方式生成 VOTable。

2)、US-VO SkyNode的跟踪与部署

SkyNode 天文数据结点是由美国虚拟天文台项目设计推出的星表访问服务，提供了对异地异构星表数据的 ADQL 接口的访问，支持多个星表间的交叉认证等操作。在本项目执行期间，我们尝试利用 Globus Toolkit 作为底层平台实现 US-VO 的 SkyNode 的功能，来解决中国天文数据中心与各国数据中心海量星表数据的访问、互操作与管理的问题。

SkyNode Interface 标准以 ADQL 标准和 VOTable 标准为基础，从概念上定义了天文数据结点需要实现的一些必要的接口。这些接口依据其功能大致分两类，一类接口用来获取元数据。这些元数据是描述一个查询中列的信息和表的信息，同时为了增强互操作和实现更高层面的数据处理，也获得了每一列的 UCD 信息以及每一列的单位。另一类接口则是用来接收 ADQL 查询语句并且返回 VOTable 格式的结果。

天文数据结点位于中国虚拟天文台架构的底层，它既可以构建在网格体系架构之上，也可以构建于 WEB 服务体系架构之上，支持一系列 IVOA 的数据访问与互操作标准。天文数据结点基于网格架构（GT 平台）或 WEB 服务架构（AXIS 平台），集成了网络环境下各种异构的天文数据资源并将它们统一组织起来，通过天文数据结点提供的数据访问与管理服务屏蔽底层数据资源的异构性，并在网格架构上为用户提供直观、方便、规范的访问和操作服务接口。

中国虚拟天文台的天文数据结点依据自顶向下的设计方法将它分为三层，分别是服务接口层、查询处理层和星表数据构造层，如图 13 所示。

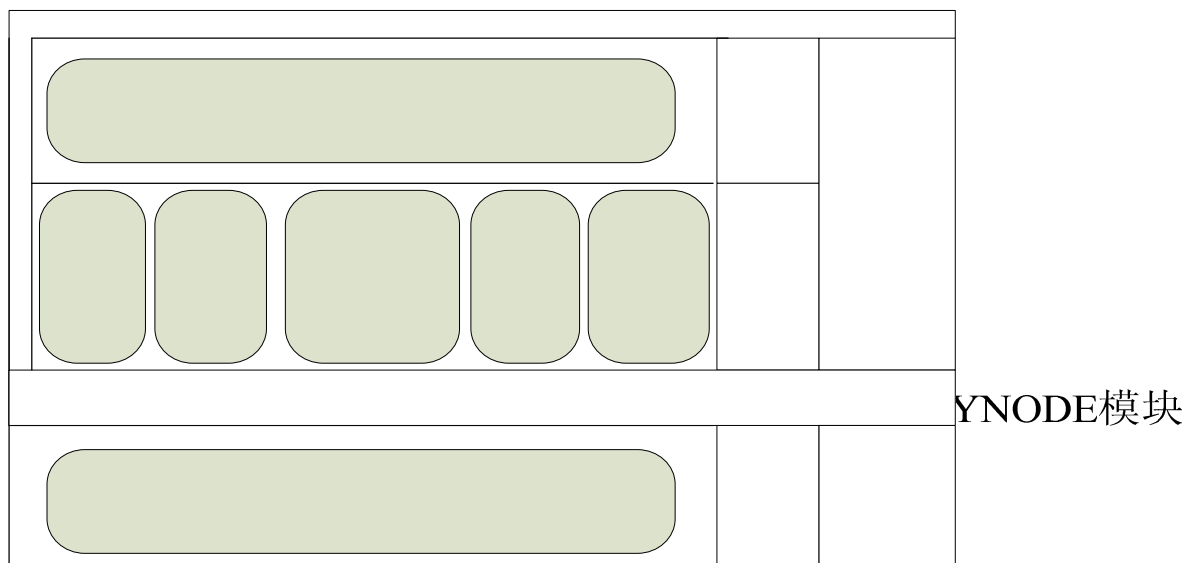


图 13 天文数据结点的体系结构

服务接口层和查询处理层：均位于中国虚拟天文台架构中的资源层，它们建立在 GT 平台或 AXIS 平台上，为上层数据分析和数据转换等高级服务提供整合后的异地、异构天文数据资源的网格服务或 WEB 服务接口。上层的应用服务、程序或者高级用户可以通过中国虚拟天文台的注册中心来动态的得到这些异地异构天文数据资源的 URL、状态信息以及元数据的信息，进而获取这些分布在世界各地的数据资源信息。

星表数据构造层：则是中国虚拟天文台架构中的构造层的一个子集，实现了中国虚拟天文台体系架构中的对星表资源的封装功能。星表数据构造层是分布在世界各地、存储于不同的数据库管理系统中不同格式、不同定义、不同类型的星表数据库的集合，即异地异构的天文数据资源。GT 平台或 AXIS 平台对这些异地异构的天文数据进行了统一的管理和集成。

天文数据结点的主要目的是让用户方便、快捷地获取异地异构的数据，同时将不同格式的数据转换为可互操作的、统一标准的数据，让用户利用这些数据在中国虚拟天文台的上层服务平台上处理并分析这些数据。天文数据结点在执行时有两种查询方式：简单的 ADQL 查询或带有 Cone Search 功能的锥形查询和带有交叉认证功能的复杂查询。

天文数据结点的简单查询适用于用户直接获取某一天区星表数据的情况，运行时仅仅调用天文数据结点的部分接口。用户执行简单查询时，天文数据访问门户或客户端程序会将 ADQL 查询语句传送至天文数据结点的 performQuery() 方

法，它可以将 ADQL 查询语句通过 ADQL 模块直接转换 SQL 或经过 REGION 模块和 COORDS 模块解析后再转换为数据库识别的 SQL 语句进行查询，查询结果经由 VOTABLE 模块转换成国际统一的标准 VOTABLE 格式的对象，以 VOData 对象的形式输出。这种查询方式高效、快捷，但缺点是没有星表的交叉认证功能，无法获取星表认证信息。天文数据结点简单查询的运行流程图如图 14 所示。

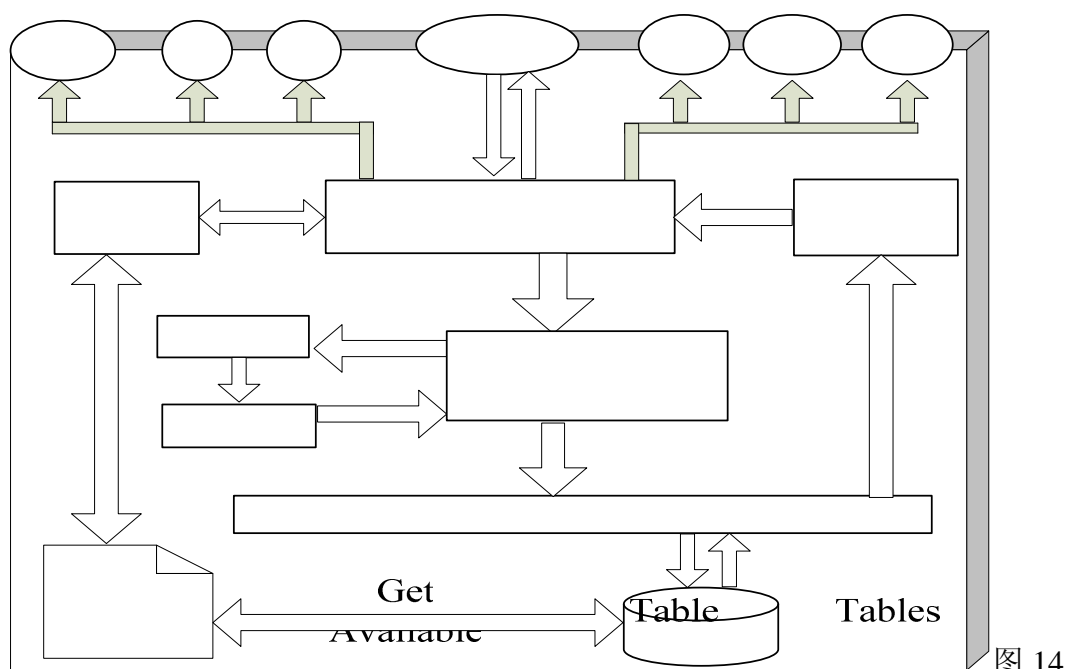


图 14 天文数据结点简单查询流程图

当执行复杂查询时，则需要调用 `executePlan()` 方法。天文数据结点复杂查询的运行流程图如图 15 所示。首先，天文数据结点的接口接收来自天文数据访问门户或者客户端程序提交给它的装载有查询条件、交叉认证的条件和锥形查询区域的 ADQL 语句、portalURL、执行计划、输出格式的 ExecPlan 的有序数组对象。执行计划中保存有主机名数组和 PlanID 的数组，并且定义了查询后的结果返回的目标结点及返回类型。主机名数组的顺序和 PlanID 的顺序相同，即查询代价最小的最先执行，依次类推。当然，这个顺序是由天文数据访问门户或者客户端程序调用 `queryCost()` 方法决定的。

接着，`executePlan()` 方法拿到这个 ExecPlan 对象后，对其中的执行计划进行判断，提取与自己主机名相同的 PlanID，并提取下一个要传送的主机名和 PlanID。最后一个结点没有下一个传送的主机名，只有一个 PortalURL。分配结束后开始查询。最终，将这个 VOData 结果对象返回给天文数

据门户或客户端程序。

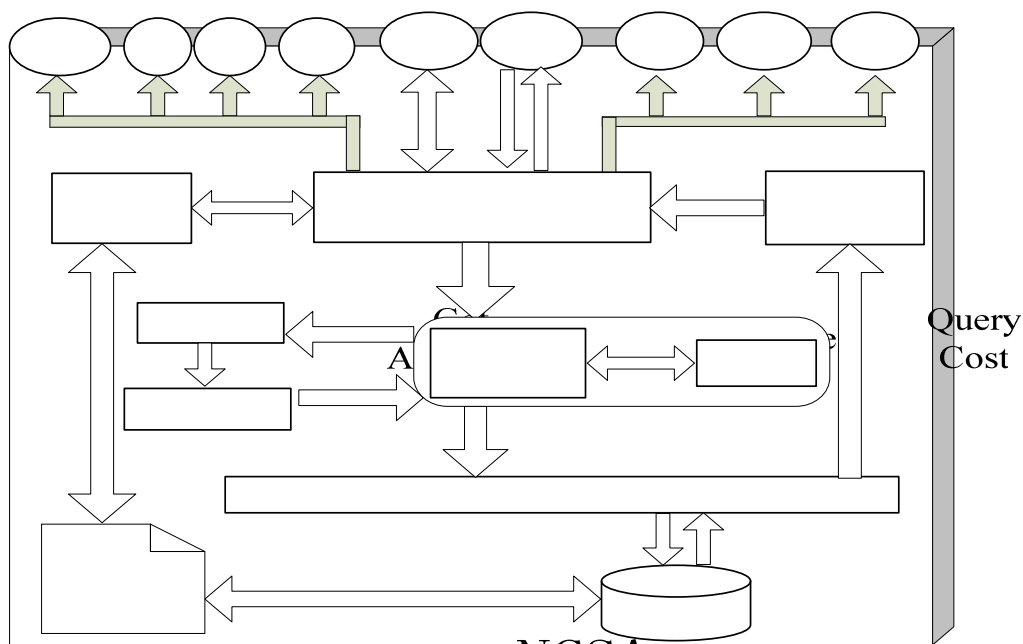


图 15 天文数据结点服务查询流程图

上述模型中，天文结点之间相互独立，由于网络访问接口的原子性和透明性，用户可以更灵活地使用这些数据集的服务，定义并开发适合自己的管理平台和用户界面，甚至还可以将结果直接返回给用户所编写的程序，进行下一步的天文数据处理。

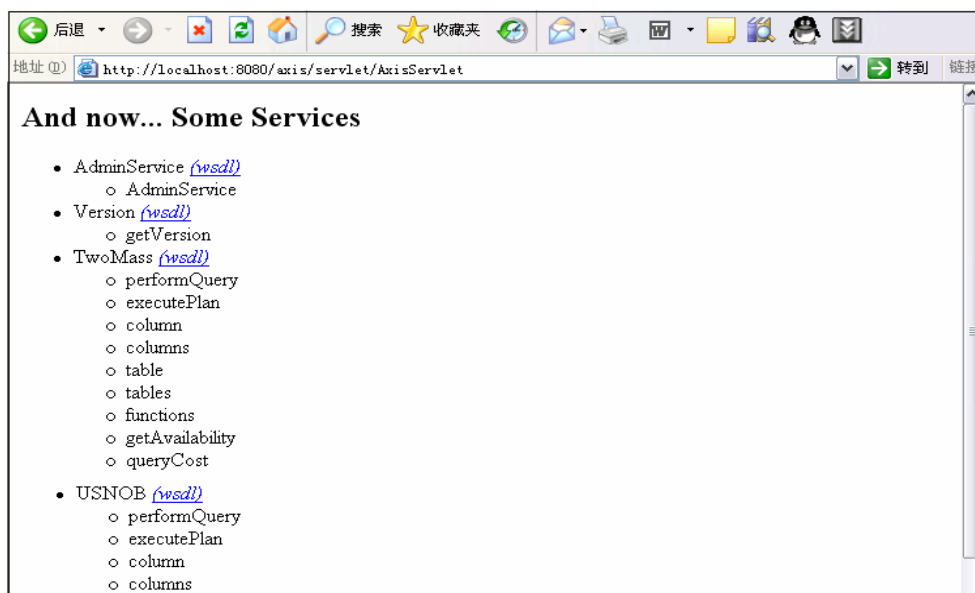


图 16 天文数据结点访问服务接口

具体实现过程中，我们选用了 JAVA 作为主要编程语言，它在封装性和扩展性几个方面都有很好的支持，同时它在不同软硬件平台上移植非常方便。JAVA 支持多线程，不必开辟多个进程，可以减轻系统负担符合项目的需求。天文结点

的软件开发工具采用了 Borland JBuilder 集成开发环境, 它提供了对 J2EE 应用环境的全面支持。虚拟天文台底层的搭建采用了网格支撑系统 Globus Toolkit, 向上层应用服务提供统一的网格接口。同时采用了 Apache Tomcat 作为 WEB 服务和应用服务的发布引擎, 支持各种服务的注册以及基于 B/S 结构的用户访问接口的发布。天文数据结点支持的接口如图 16 所示。

3)、星表数据自动化处理系统

天文数据主要包括星表、星图、光谱、文献资料等, 其中星表是包含天体信息的数据表格, 是天文学家最常用到的天文数据。将不同波段的星表、特别是巡天项目产生的大型星表进行交叉认证, 一直是进行很多天文学研究、尤其是数据挖掘、统计分析研究的瓶颈。

星表是包含天体信息的数据表格, 是天文学家最常用到的天文数据。星表大小差别悬殊, 小到几十个目标, 大到上亿甚至上十亿个目标, 而且近年来随着巡天项目增加和扩大得到的大星表越来越多。星表包括 ReadMe 文件和数据表文件。其中 ReadMe 文件以标准格式提供星表的有关信息, 如包含哪些文件, 并提供数据表文件中每一列的位置、名称、单位等描述(即表结构)。一个星表对应着一个数据库(即 Database), 一个星表可能包含多个数据表。星表的一个数据表对应着数据库的一个表(即 table)。

北京天文数据中心(BADC)作为中国目前最大的天文数据中心, 已经建立了自己的天文数据库和网络数据服务系统, 并通过国际交换和网络下载积累了大量适合我国天文研究需要的天文数据。它积累的天文数据库是虚拟天文台的重要要素之一, 也是其研究开发的基础和实验床。天文星表数据库作为国家天文数据中心数据库的重要部分, 收集和整理重要的和最新的星表数据集, 为国内外天文学家和其他用户提供星表查询服务和星表获得服务。

为了实现对星表的高效管理和查询, 我们设计实现一个功能完善的天文星表数据库管理系统。天文星表数据库的主要数据来源于斯特拉斯堡天文数据中心(CDS)负责收集整理天文星表, 也包括其他大型巡天项目和其它星表数据。高丹等基于北京天文数据中心实现了星表的自动入库工具, 把 CDS、SDSS、2MASS 等其他数据中心的数据自动入库到北京天文数据中心。此后, 又实现了多个具有

不同功能的天文数据处理工具，包括：自动入库工具、交叉证认工具等。

在这些工作的基础上，把各个工具进行整合，使得各个工具能够批处理地、多任务地进行工作，且将此工具发布成为一个能够基于任何支持 SQL 语言的数据库系统的、面向使用者的一项服务。这项工作很大程度地方便了交叉证认等工作的进行。

在进行这项工作之前，这些工具都是独立的程序，每个程序使用的时候都需要用户输入很多的参数，很容易出错，易用性方面不尽人意。而且当需要这些程序共同完成一个工作的时候：例如，先将两个星表入库，然后建立 HTM 索引，再进行交叉证认以及提取参数表，最后再将数据返回给用户，就需要用户分别进行多次参数输入，如星表入库、建立索引就分别需要两次，并且每次工作都必须等待上一步工作完成之后，用户来用命令行执行。这样，不仅需要用户反复查询当前工作是否完成，造成工作的低效，而且用户输入参数本身也是一项繁琐而易出错的工作。

现在各个步骤可以使用参数文件方便地进行编辑，原来需要输入的参数列表，现在可以在功能强大的编辑器中，根据我们提供的模版，编辑这些参数，可以很大的提高用户的效率，且减少了出错的可能性；而将各个步骤的灵活组合，更使得完成复杂功能对于用户而言，只需要一次编辑各步骤的参数文件，无论每一步需要进行多少次，都只是需要编辑此步骤对应的一个参数文件，然后运行即可。

此系统可以架构在任何支持SQL语言的数据库上。如图17 所示，此系统架构在 (BADC) 上，系统已经多次使用于数据挖掘工作的数据处理工作，运行正常。此系统也可以方便地移植到任何数据库系统上。当用户只是需要拿自己的一些小星表自身进行交叉证认，或者拿自己的小星表和此系统基于的数据库中的星表进行交叉证认的时候，可以上传自己的星表，使用本人的系统提供的服务。而当用户有着大量的星表需要进行交叉证认，或者需要上传大型星表的时候，如果使用别人提供的此服务，可能会在网络传输上花费较多的时间，更方便的办法是自己建立一个数据库，将这里描述的系统架构到自己的数据库上再使用。这对存储着大量数据的各大天文数据中心有着很大的意义。

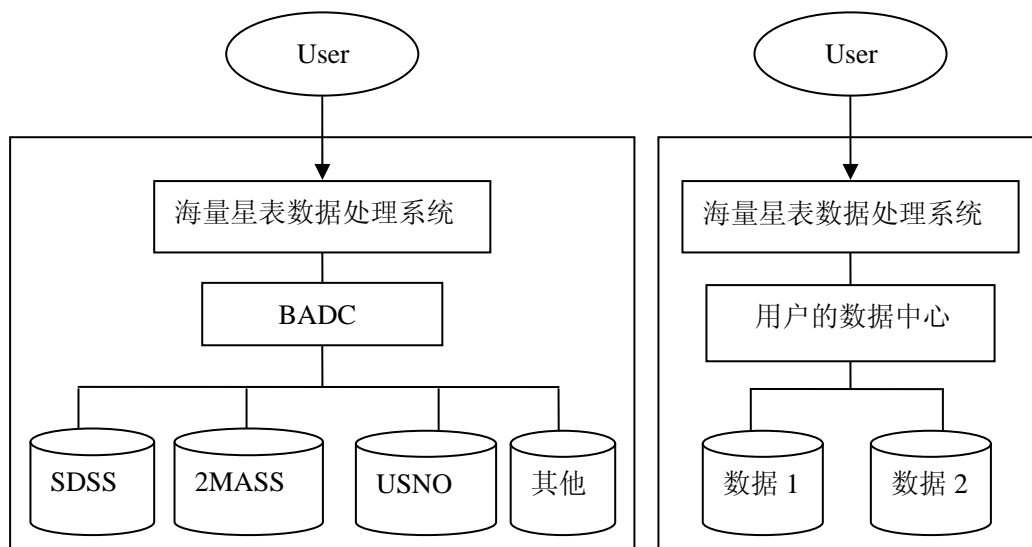


图 17 星表数据处理工具自动化架构

整合前和整合后的工作方式对比如下：

整合前

- 仅仅在 BADC 上运行
- 各个工具彼此独立
- 用户需要使用命令行，输入很长的参数表，易出错
- 每次只能执行单个操作
- 易用性差，效率低
- 用户上手难

整合后

- 可以架构在任何支持 SQL 的数据库
- 使用参数文件使得用户编辑参数更方便
- 提供了参数文件模版，用户更容易理解和使用
- 各个独立的步骤可以灵活组合
- 批处理地连续完成各个功能，不需要用户干预
- 用户学习很容易

整合这些工具的基本思路就是使用批处理的思想。整合这些工具使用了 Linux Shell 编程和 Vi 的自动处理功能，所以这些程序运行于 Linux 操作系统。首先用户根据需要进行的操作编辑批处理文件 steps_crosswork，编辑此文件的时候可以参考我们提供的模板 steps_crosswork_template。还是假设用户需要执行包括将两个星表都入库和建立索引，再将两个星表进行交叉认证，然后进行参数提取和数据传回。则 steps_crosswork 文件应类似如下：

```
./step1 step1_file_putdata
```

```

./step2 step2_file_HTMindex
./step3_7 step3_file_crossmath7
./step4 step4_file_sqldata
./step5 step5_file_getdata

```

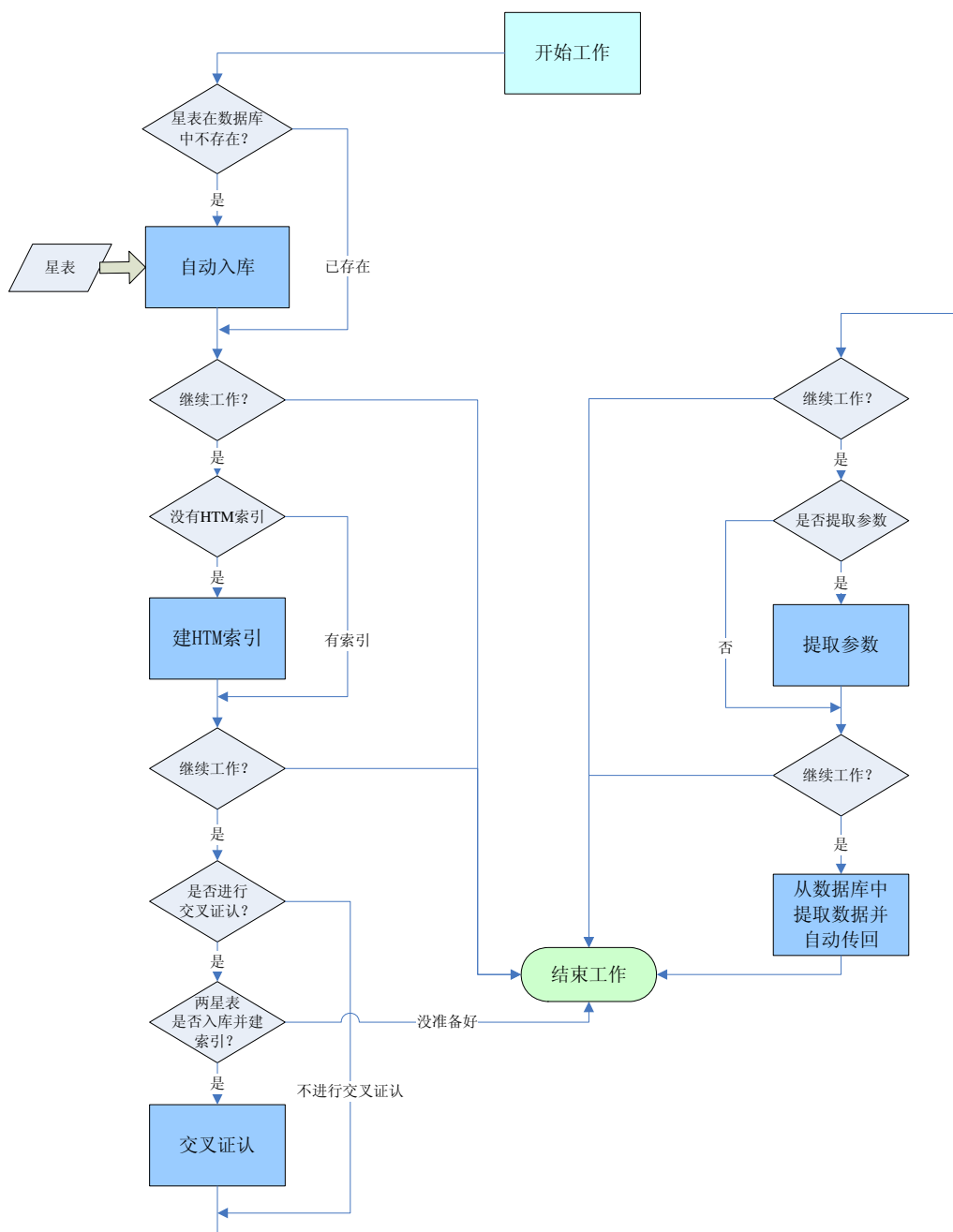


图 18 海量星表数据处理工具的流程图

五个步骤各自的功能如下：

- 1) 自动入库：根据星表 ReadMe 文件，自动建表、入库、将时分秒形式转换成度 (RAdeg、DEdeg)、建索引、建 id_htm 主键。
- 2) 自动建 HTM 索引：根据数据表的坐标数据 (度, J2000) 计算出对应 HTM 索引的 pcode 值，为星表建立 HTM 索引, 将 id_htm 主键和 pcode 值两

列新建 HTM 索引表。

- 3) 交叉证认：根据两个数据表及其 HTM 索引表、误差半径进行交叉证认，得到交叉证认表，并分别给交叉证认表中的两表 id 列建索引，方便进一步的数据参数提取。我们为不同的交叉证认需求提供了不同的程序。生成的交叉证认表中只有交叉证认两表的赤径、赤纬、id_htm 主键以及角距离 d 的信息。
- 4) 交叉证认参数提取：根据用户需要的参数，对数据库进行操作，生成交叉证认提取参数表。表中包括用户需要的所有参数信息。
- 5) 数据提取并返回：指定数据库名、表名等信息，自动从数据库中取出数据，并自动通过 FTP 传到用户自己的电脑上。

五个步骤的流程逻辑如图18 所示。

4)、VO-DAS数据统一访问平台

自从开始对天体进行照相观测以来，世界各国的天文观测资料不断增长，近年来由于先进的 CCD 技术、计算机处理技术和大型望远镜技术的快速发展，天文观测数据更是在以指数增长。天文信息爆炸式增长不仅给天文学家带来了天文发现的巨大机遇，也同时带来了数据访问和处理方面的巨大挑战。传统的借助人工来对数据进行分析处理的方式已经不能满足海量的天文数据了。另一方面，上百年的天体观测资料积累造成了这样一种情况：世界各国的天文资料保存和管理方式存在巨大差异，不同历史时期数据保存和管理方式也存在巨大差异。虽然世界各地的天文数据是自由共享的，但是数据结构的混乱和数据存储的多样性就像语言不通一样严重影响着天文学家对这些数据资料的使用效率。China-VO 的数据访问服务系统（VO-DAS）就是为了解决这样的困难而设计。

统一访问异地异构数据库是指数据访问者不必了解数据资源的具体物理位置以及数据资源的具体组织形式，通过统一的访问接口就可以取得他们希望得到的数据，并且可以在多个不同的数据源之间实现数据的联合查询或交叉证认。为了能够统一地访问异地异构数据资源，要求数据访问服务具有以下两种能力：从遍布世界各地的网络服务器上发现合适的数据资源；能够整合这些数据资源，当联合查询多个数据库的时候，能够提供一种机制把分散在不同地方的数据库从逻

辑上联系起来，实现联合查询，同时尽量减少网络带宽的占用。

通常我们经常接触的天文数据资源包括四类资源：星表数据、图像数据、光谱数据和文献，VO-DAS 将可以访问除文献外的其它几种数据资源。如果数据是星表数据，通常会按照关系型数据库的方式存放。这样，它的元数据描述应该包括这个星表的基本信息，具有多少个表结构，每个表结构的意义是什么，具有多少列（或用数据库的术语称为字段），每个列的名称、单位、精度、物理意义是什么。可以看出星表的元数据描述是一个树形结构。图像数据本身通常使用 FITS 文件来保存的。为了能够方便地索引到需要的图像，还要有一个对应的数据库的表，记录 FITS 文件的空间参考位置、波段、观测时间和地点，观测时候的天空视宁度、云量、投影参数、望远镜参数、CCD 或底版的参数等。光谱数据（一维光谱）也是以 FITS 文件形式保存的，通常每个 FITS 文件只有一个天体的光谱，因此 FITS 文件的名称应该能够和一个特定星表的一行一一对应。这样，要想找到一个天体的光谱，就首先在一个星表中找到这个天体，然后再根据星表行的关键字和 FITS 文件的一一映射关系找到光谱文件。

VO-DAS 要为以上三种天文数据设计一个统一的元数据描述方式。注意到 FITS 图像需要一个索引表检索，因此我们可以将这个索引表看成是和星表一样的。而光谱文件的搜索也一定要借助星表。这样，我们就可以设计一种能够描述表结构的元数据描述形式来描述所有的天文数据资源。有了资源注册机制，VO-DAS 就可以为使用者找到需要的天文数据资源。

有了统一的元数据描述方式，就可以让用户和数据访问服务理解访问的数据是如何组织的、类型是怎样的。这样，无论数据资源放在何处，也无论数据资源是怎么组织的，VO-DAS 都可以借助这两项功能找到数据资源并理解数据资源。统一的数据访问的另一层含义是对异地异构数据的联合访问，包括交叉认证。联合查询两个和两个以上的表，在关系型数据库中需要将两个表的相关行进行组合，然后从中挑出符合条件的行。如果两个表不在一个数据库系统中，甚至不在一个地方，那么必须要将一个表的部分信息传递到另一个表的数据库中，在本地数据库中完成联合查询。当数据量很大的时候，这样的数据传输是不经济的。一种可行的解决方法是把联合查询拆成两步查询，首先在一个表中完成查询，然后把查询结果传递到第二个表所在的数据库中完成第二次联合查询。这样的方法虽

然不能完全解决数据传输的问题，在某些最坏的情况下甚至还需要传递整张表，但是大多数情况下，特别是当用一个小天区作约束查询的时候是可行的。

为了支持对图像和光谱的查询，我们在 VO-DAS 中对 ADQL 的语义作了扩展，允许 FROM 分句中的表的类型是星表（等同于数据库中的表）、图像库（等同于文件集合）和光谱库（等同于文件集合）。当 FROM 分句列出的是星表的时候，整个 ADQL 的使用就是按照 IVOA 的标准进行；当 FROM 是图像和光谱的时候，SELECT 分句的列名实际是图像和光谱的输出参数名，其中当然包括文件位置 URL 信息；WHERE 分句的条件则是对要查询的图像和光谱文件的输入参数进行的约束。经过这种扩展，当 ADQL 查询图像或光谱的时候，返回的表中一定有一列是文件 URL，这样，再根据 URL 指示就可以取得图像或光谱的 FITS 文件了。通过这样的扩展，语法上 ADQL 没有任何改变，但是语义的扩展便让 ADQL 支持天文图像数据和光谱数据的查询了。

此外，DataNode 基本沿用了 OGSA-DAI 的结构，但为了支持图像和光谱数据，我们对它也做了扩充。首先，Catalog Query 就使用 OGSA-DAI 提供的数据库查询功能实现。同时增加 XMatch，即交叉证认功能模块。其次，增加 Image Query 和 SpectrumQuery 两个模块实现图像和光谱查询的功能扩充。Data Transform 会支持天文数据文件的特定格式，包括 VOTable 和常用的 ASCII 文件。Data Delivery 和 VOSpace Client 联系 VOSpace 将查询结果文件发送到指定的存储服务器上。

经过上述的处理，VO-DAS 便成为了一个能支持星表、图像、光谱等多种数据格式，能访问本地、异地数据资源的统一数据访问环境。关于 VO-DAS 的其他细节可以参考前面的内容以及刘超、田海俊、杨阳、高丹、路勇等人的论文。

5)、SDSS数据的国内镜像

美国斯隆数字巡天计划 (SDSS) 与虚拟天文台有着密不可分的关系，为 VO 提供了重要数据资源、科学牵引、技术实践。同时，SDSS 为 LAMOST 提供了宝贵的经验和资源。本项目通过与 LAMOST 大科学工程合作的方式，购置了一套 SGI 海量数据存储系统，有效存储空间达到 24TB，完全满足 SDSS 图像和星表的传输、存储、数据操作等工作的需求。

本项目通过与中科院网络中心合作，利用 GLORIAD 中美俄环球科教网络传输数据。目前已经完成 SDSS DR6 星表数据和星图数据共 10TB 数据的传输工作。这套宝贵的海量数据将为 LAMOST 大科学工程和虚拟天文台提供重要的数据资源。

6)、天文数据的整理与发布

中国虚拟天文台一个重要的目标和任务就是将国内的天文数据共享给国际天文界的同行，同时将国际丰富的数据资源推介、引进到国内。项目执行期间，项目成员与国家天文台 BATC 课题组紧密合作，将 BATC 十年来的观测成果 BATC 星表，一套含有数十万天体多波段测光结果的星表按照国际天文界的标准进行了共享，通过全球多个数据库系统都可以方便的查询检索。这项工作被列入国家天文台 2005 年十大天文进展候选项目。

大视场巡天 (BATC) 数据是利用我国国家天文台兴隆观测站施密特望远镜长期观测的数据整理的，是我国目前累积数据量最大 (700GB)、时间最长 (1995 年至今)、整理最完善的自主观测数据。大视场巡天 (BATC) 的原始数据和处理后的星表数据都在科研实践中发挥了相当重要的作用。近年来，基于 BATC 观测数据的研究论文每年约 10—15 篇进入 SCI；根据该数据发现了多颗小行星 (获得国际暂定编号的小行星 2707 颗，其中有 575 颗小行星 已经获得永久编号和命名权)。已经国家天文台的名义，对 BATC 发现的 30 余颗小行星的进行了命名工作；该数据还在星团、星系、类星体、超新星等一系列有特色的前沿研究领域发挥重要作用。

该数据目前除提供国内天文学家使用外，尚与国外研究单位开展合作研究，并在异地分别建有数据备份。2005 年本项目与国家天文台 BATC 课题以及法国 CDS 天文数据中心合作，将其引入 CDS 国际天文数据库，成为国际编号备案 (II/262/batc) 的星表数据库。目前 BATC 数据在 CDS 本部 (<http://cdsweb.u-strasbg.fr/viz-bin/Cat?II/262>) 和世界上多个数据库镜像点都可以检索，为该数据在国际天文界发挥更大的作用提供了有利的保障。

2007 年，项目组与国家天文台 BATC 巡天团组合作完成了 BATC 多波段测光图像的收集整理工作，目前所有的图像均已经提供了在线下载服务。网址为：

“<http://badc.lamost.org/archives/BATC/archive/>”。

在数据库建设与服务方面，本项目和世界数据中心天文中心合作收集整理 5000 多个天文星表，并提供了基于星表元数据检索的星表访问和查询服务。用户通过一个类似 Google 的简单的检索界面就可以从数千个星表中轻松的找到自己想要的星表。目前项目提供在线服务的总数据量达到 2TB。访问网址：“<http://badc.lamost.org>”。

另外，项目开发实现了虚拟天文台交叉认证服务：这是一个高级的星表交叉认证服务，提供了服务器端两星表交叉认证和用户自己上传小星表与服务器端星表交叉认证功能。还实现了对交叉认证结果的分类和参数的自由选择等功能。参见网站：“<http://badc.lamost.org:8080/xmatch/>”。

3、现有天文工具的虚拟天文台集成

虚拟天文台是随着天文学、天文仪器和计算机、信息技术的发展而提出的，通过先进的信息技术把全球范围内的研究资源无缝透明地联结起来，组成数据密集型在线天文研究环境。虽然虚拟天文台是一个全新的网络化天文在线研究环境，但它仍然会充分利用和容纳现有的天文软件资源和各种服务。传统的天文软件经过了几十年的开发和维护，功能强大，使用便捷，是天文界的一笔宝贵财富。现有工具如何与新开发的 VO 工具协同工作是 VO 开发者们从一开始就面临的课题。在项目的早期，我们试图通过对已有工具的 Grid Service 封装来实现其与虚拟天文台的集成，尝试了对 DS9、ImageMagick 的封装。也试图参考现有的软件功能重新开发应用软件，比如参考 Aladin 开发 VOIMPAT 图像处理工具等。IVOA PLASTIC 协议的推出为已有工具的集成提供了全新的解决方案：从单纯的追求技术接口上的相同到实现数据、消息上的互通。通过 PLASTIC，项目组实现了 VO-DAS 与 VOPlot、Aladin 等 VO 工具的集成，实现了 VO-DAS 与传统桌面科学软件 MATLAB 的互操作。下一步，China-VO 将通过 PLASTIC 把更多的应用工具集成到 China-VO 的系统中，为用户提供丰富多样的选择。

1)、应用程序的服务封装探索

由于中国虚拟天文台采用面向服务的架构 (SOA)，所以软件集成实质上是

将软件资源服务化。我们可以采用两种方法来实现软件的服务化。第一种方法是手工逐个集成，在软件的源代码的基础上加一些接口，来封装成一个服务，一般针对开源软件。这对规模小的工作技术组技术开发具备可行性。第二种方法是为软件添加代理，使软件与网格环境、用户之间进行交互，而并不对软件本身做改动。软件代理可以负责接收用户请求，启动软件，调用相关功能，返回执行结果等，结构如图 19 所示。

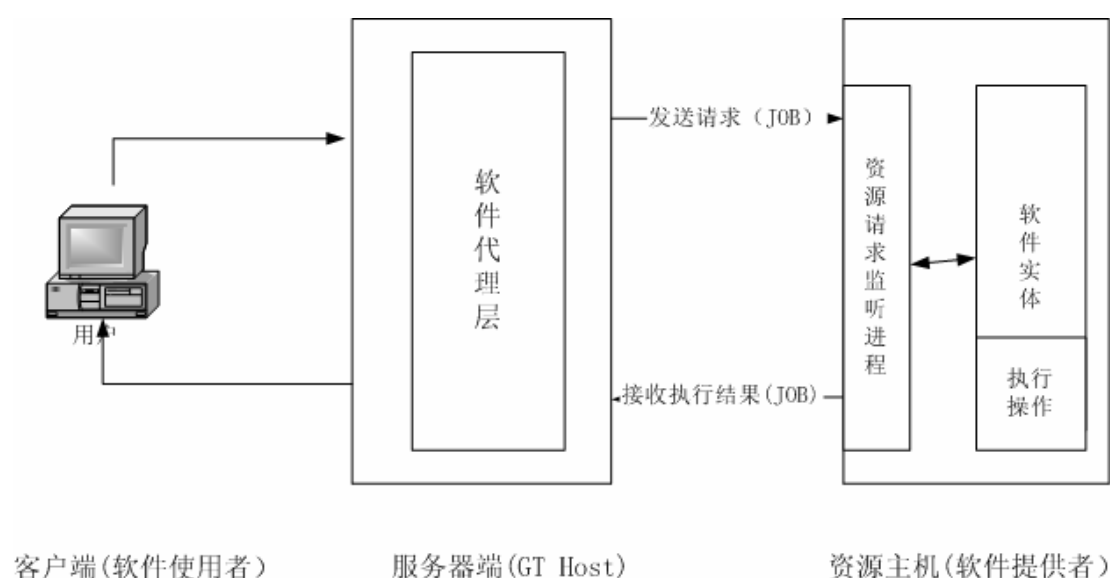


图 19 基于代理的服务化封装

软件的代理实际上也是一个网格服务，它的一个服务实例就对应到一个具体的资源主机，其上的所有软件都可共用此代理。当用户需要使用某种软件时，先寻找一个代理的实例，在这个代理中主要记录了软件所在的主机位置。然后，用户提交一个功能请求，代理首先要将这个功能请求转化为标准的请求头，在这里我们称它为一个工作(JOB)。请求头发送到资源主机上后，经过资源端监听进程解析后启动相应的进程，再将结果填入请求头中，然后，将此请求头重新返回。代理重新接受此请求头(此时已变成应答)，检索出结果字段，返回给服务调用者。不过，虽然这种方法具有通用性，很多软件可以共用一个代理，但是它并不适用于频繁交互和图形界面软件。

软件调用实现一般由用户界面开始。当用户浏览 Web 页面时，发现对其有用的操作，需要在页面上填写接口参数，以及上传需要处理的文件。这一部分的

工作是在 Web 服务的客户端完成的。用户所有操作都在此进行。

用户需求转发至 Web 服务的服务端, Web 服务根据请求启动相关网格服务。这里需要强调的是, Web 服务端其实就是网格服务的客户端。在分析应用程序源代码的基础上, 确定应用程序提供给外部调用的接口, 为它完整的服务描述及实现接口, 将其封装为网格服务。这一部分在网格服务端完成。其中网格服务的核心是提供操作的应用程序源码。网格服务端响应网格客户端请求, 应用程序执行相应的用户操作。服务调用过程如图 20 所示。

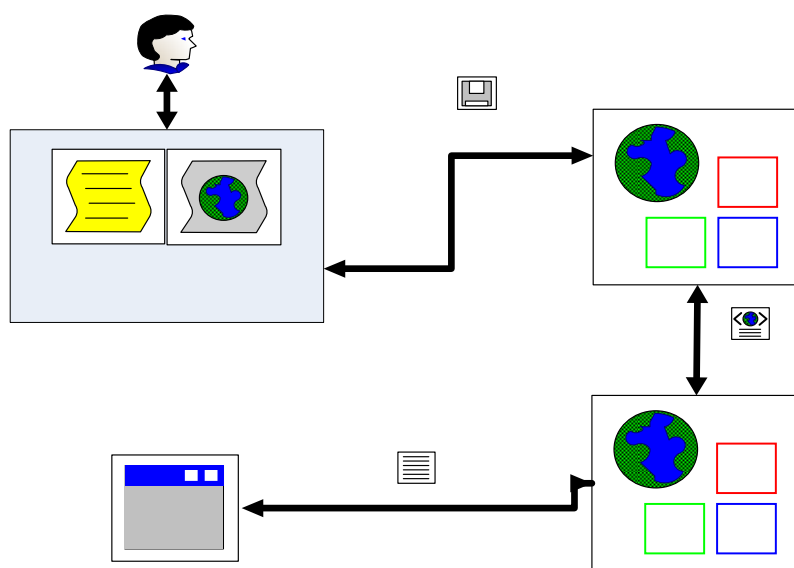


图 20 服务调用过程

图 20 中, Web 客户端使用 JSP 和 Servlet 开发。它有两种主要的接口: 一个用来接收终端用户的输入参数, 比如 ImageMagick 服务中, 需要填写源图像位置信息、目标格式和一些其他选项; 另一个接口用来从网格客户端接收处理结果, 并将最终结果展示给用户或下载。Grid Client 是一个网格客户端服务。同样也有两种主要的功能, 根据输入需求从客户浏览器端激活网格服务, 并返回结果给客户浏览器。Grid Server 是网格服务软件功能使能的关键部分, 它提供了标准的封装原始软件包的网格接口。有三个主要的功能, 从网格客户端接受激活需求; 调用本地软件执行程序; 返回给网格客户端结果。软件核心是各软件项目的原始包, 具有可执行命令和共享库。

ImageMagick 是一套稳定的、功能强大的、免费的工具集和开发包, 可以用

来读、写和处理超过 89 种基本格式的图片文件，包括流行的 TIFF、JPEG、GIF、PNG、PDF 以及 PhotoCD 等格式，也包括天文界经常使用的 FITS 格式。利用 ImageMagick，你可以根据 Web 应用程序的需要动态生成图片，还可以对一个图片或图片序列进行改变大小、旋转、锐化、减色或增加特效等操作，并将操作的结果以相同格式或其它格式保存。对图片的操作，可以通过命令行进行，也可以用 C、C++、Perl、Java、PHP、Python 或 Ruby 编程来完成。同时 ImageMagick 提供了一个高质量的 2D 工具包，部分地支持 SVG。现在，ImageMagick 的主要工作集中在性能、减少 bug 以及提供稳定的 API 和 ABI 上。

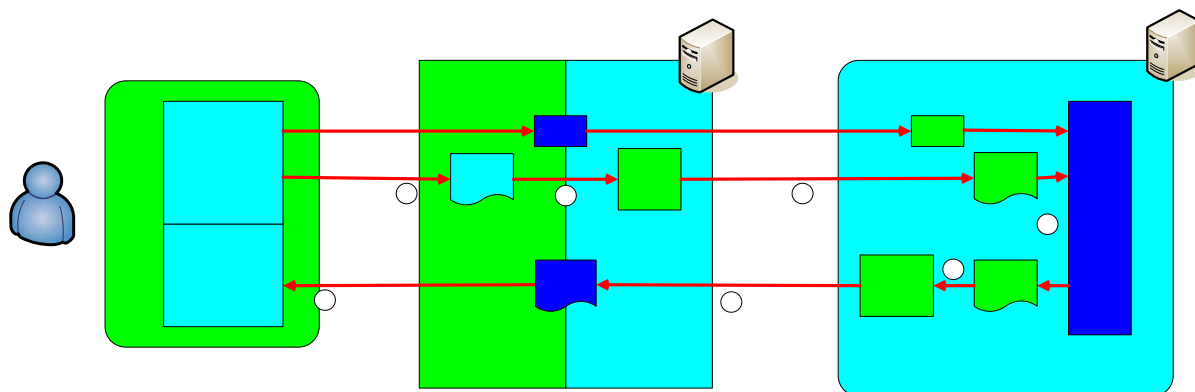


图 21 ImageMagick 工作流程图

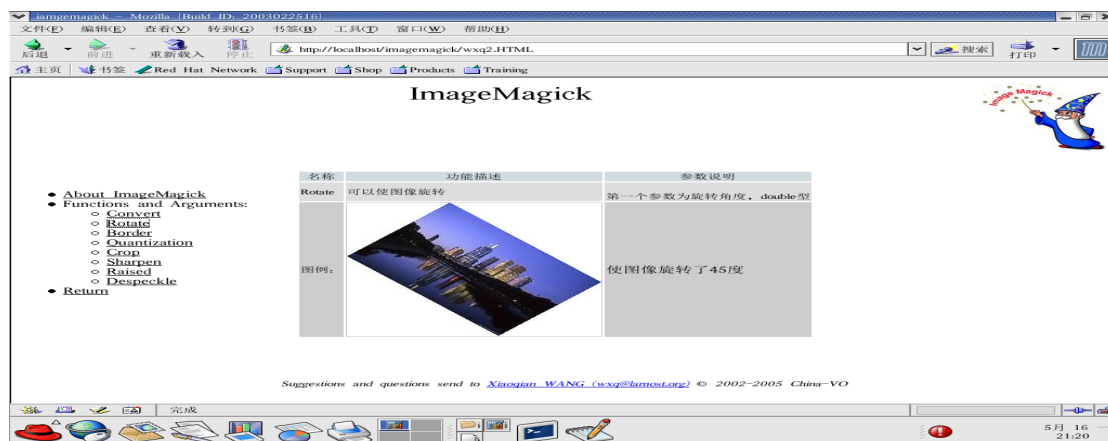


图 22 封装后的 ImageMagick 旋转函数示例

Jmagick 是一个开源 API，利用 JNI 技术实现了对 ImageMagick API 的 Java 访问接口。通过对 JMagick 的学习、研究，我们了解到通过调用 JMagick 访问接口，采用对软件源程序提供接口的方式，可以实现 ImageMagick 的功能调用。封装 ImageMagick 工作流程如图 21 所示。图 22 是封装后的 ImageMagick 旋转函数示例。

Web服务端

上传文件

参

有文

示例，允许用户通过网页来调用后台的 ImageMagick 服务。

对于源程序为 Java、C、C++应用程序来说，封装的过程是一样的。我们需要分析源代码，在源代码的基础上加一些接口，以封装成一个服务。我们知道，天文界的许多软件都是用 C、C++编写的，对其提供 java 接口的时候就需要使用 JNI(Java Native Interface)。JNI 是 Java 语言的本地编程接口，是 J2SDK 的一部分，提供了 Java 调用 C、C++编写的动态链接库的机制。实现这种方式需要编写生成和部署动态链接库(通过 JNI 调用本地方法，本地方法是以库文件的形式存放的，在 WINDOWS 平台上是 DLL 文件形式，在 UNIX 机器上是 SO 文件形式)，以及用 Java 编写调用链接库的类，开发量比较前一种方式要大些。我们可以把它想象为本地代码和 Java 代码的粘合剂。当然，对于应用程序也有所要求：函数必须是 C、C++的，同时接口必须简单；函数定义类型必须是基本类型。封装的具体步骤如下。

1) 首先我们需要编写自己的 Java 代码，在 Java 代码中我们会声明 native 方法，关键字 native，表明这个方法使用 Java 以外的语言实现。方法不包括实现，因为我们要调用实现它的 C、C++本地方法。在该类中，用 System. loadLibrary() 方法加载需要的动态链接库；

2) 其次是用 javah 编译生成头文件，生成相关 JNI 方法的头文件语句是 javah -jni *.class；

3) 然后在本地方法中引入此头文件，编译本地方法，生成动态链接库；

4)最后，将动态链接库复制到工作目录。当运行 Java 程序时，就可以调用本地方法了。这样，也就实现了源程序为 C、C++的应用程序的封装。如果在 Java 程序中直接调用 C、C++语言产生的机器码，该部分代码的安全性就由 Java 虚拟机控制。而且当本地链接库运行时，没有有效地防止数组越界错误、错误指针引用带来的间接错误等，所以必须保证本地代码的稳定性。

Stilts (Starlink Tables Infrastructure Library Tool Set) 是 StarLink 工作组开发的一个基于 STIL 的命令行工具集。STIL 是一个处理列表数据输入、输出的纯 Java 库，当读写表的时候，它不需要关心表的存储格式是 Fits 文件、VOTable 文件还是 ASCII 码文本文件就可以进行处理。STIL 支持的格式包括 VOTable、Fits、SQL、ASCII、CSV 等，非常适合于处理天文数据。所以，Stilts 用于处理表列数

据，是专门用来处理像星表这样的天文数据的。当然，Stilts 并不局限于天文数据处理。

有一些像 Stilts 这样的开源的天文软件提供了命令行的调用方式。不论是在 Windows 下 DOS 方式还是 Linux 下的 Shell 方式，用户只需输入操作关键词和参数，就可以执行相应的操作。所以我们可以采用封装命令行的方式，将其命令行封装成为网格服务，提供用户调用。

当然，应用程序能够行命令调用只是可以命令行封装的一个必要条件，并不是每个提供行命令的应用程序都能够进行如此封装，还需要有标准输出。每个应用程序可以有許多输出流，但我们最感兴趣的是标准输出，必须能够捕获流。因为对于某些天文所特有的格式，比如 fits，我们不能在屏幕上直接显示，在执行命令行调用时，必须将文件输入输出到流，这些数据流表示了 fits 文件中字符或者字节数组的流动序列，我们就可以通过捕获数据流的方式，把 fits 文件看作字节数组进行处理。如果输出方式是图形界面方式，我们就不能利用这种方式了。截获数据流的方式提供了处理这些特殊格式文件的方法。

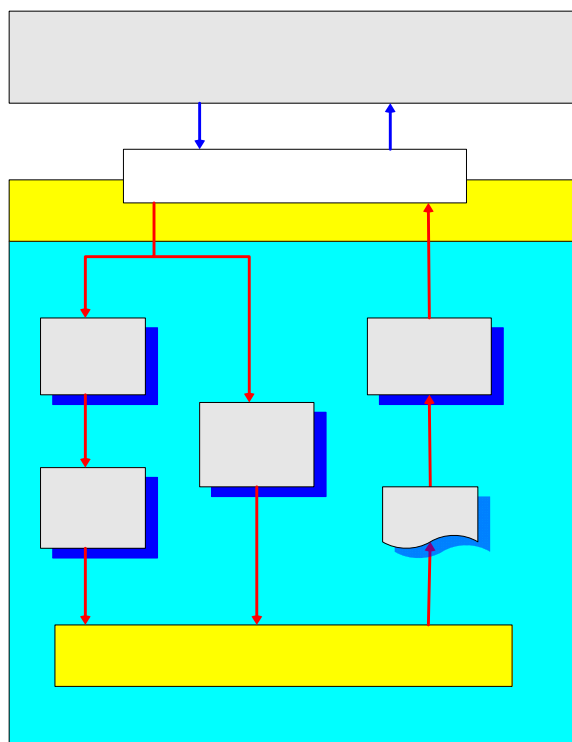


图 23 不同软件集成方案

此外，我们还尝试了对 DS9 和 Jsky 的封装工作。这样，以 ImageMagick、STILTS、DS9 和 Jsky 等几个天文学家经常使用的软件服务化为示例，项目组提

出了软件虚拟天文台封装的基本方案：对于提供了命令行的调用方式的软件，可以将其命令行封装为网格服务；对于没有命令行的软件，可以对其源程序提供接口进行网格封装，源程序为 C、C++ 的应用程序可通过 JNI 实现网格封装，见图 23 所示。

2)、天文数据可视化软件的自主开发尝试

这方面的研究和探索主要包括项目执行早期邵慧娟同学完成的数据可视化服务以及后来王丹博士开发完成的 VO_IMPAT 图象处理工具，下面进行分别介绍。

以 Globus Toolkit 为数据访问平台的基础，通过 Java Servlet 和 Applet 的结合，邵慧娟同学完成了一个基本的网格环境下数据访问和可视化系统，客户端界面如图 24 所示。这个系统实现了网格环境下对数据的访问，支持对 FITS 文件和 VOTable 文件的显示与简单操作。

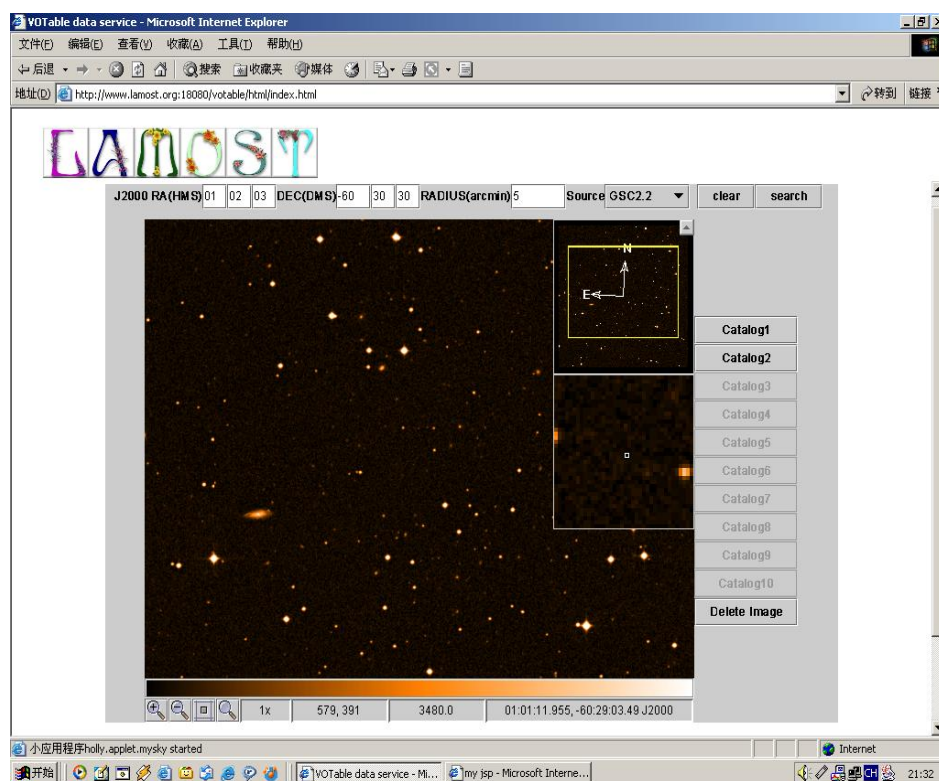


图 24 项目早期实现的数据可视化服务

2005 到 2006 年间，王丹博士以 CDS 的 Aladin 为参照对象开发了虚拟天文台图像处理工具 VO IMPAT (Image Processing and Analysis Toolkit for the Virtual

Observatory of China)。VO IMPAT 是个 JAVA 开发的交互式的图像处理工具，后台星表数据来自于北京天文数据中心(Beijing Astronomical Data Center, BADC)。其基本操作流程如图 25 所示：当使用 VO IMPAT 时，用户输入查找的区域和半径后，发送请求，通过 Web Service 传送到 BADC 服务器上，服务器分析查找请求后查询数据库，将查找到的数据返回给用户。

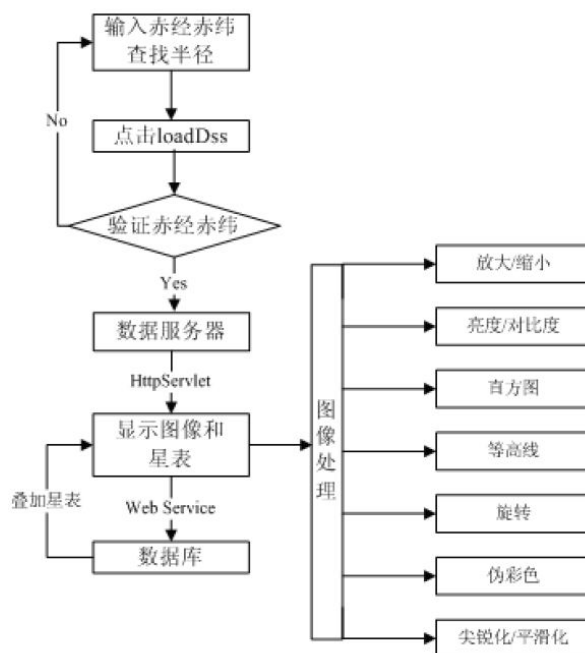


图 25 VO IMPAT 操作流程

VO IMPAT 允许用户可视化任何一部分天空，并将光学、红外、射电、X 射线波段的星表数据叠加在 DSS 底图上。它可以同时访问数字巡天图像 (DSS)、天文星表和其他的数据库。VO IMPAT 实现了多波段天文数据的融合，将不同波段的星表 (例如: USNO、2MASS、NVSS、RASS) 叠加在 DSS 图像上。VO IMPAT 还提供许多重要的图像处理功能，如放大缩小、改变颜色、等高线、画图工具、直方图、平滑化、尖锐化、旋转、标记等。

3)、PLASTIC 粘结下的现有软件协同

天文应用工具设计的关键是如何强化工具的功能特性来满足用户的需求。如果通过扩充工具本身功能的手段来实现，工具就会变得越来越庞大，它的相关资源信息也会不断地更新和扩展，这给工具的维护工作增加了复杂性，对用户的使用也带来很大的困难。另外，这类工具软件独立性地开发和使用的方

在工具之间功能的重复性问题。因此，开发出更多的专有工具就成为必要，但是如何实现各个专用工具交互式地协同工作来完成复杂的数据处理或分析，是天文应用研究的目标。

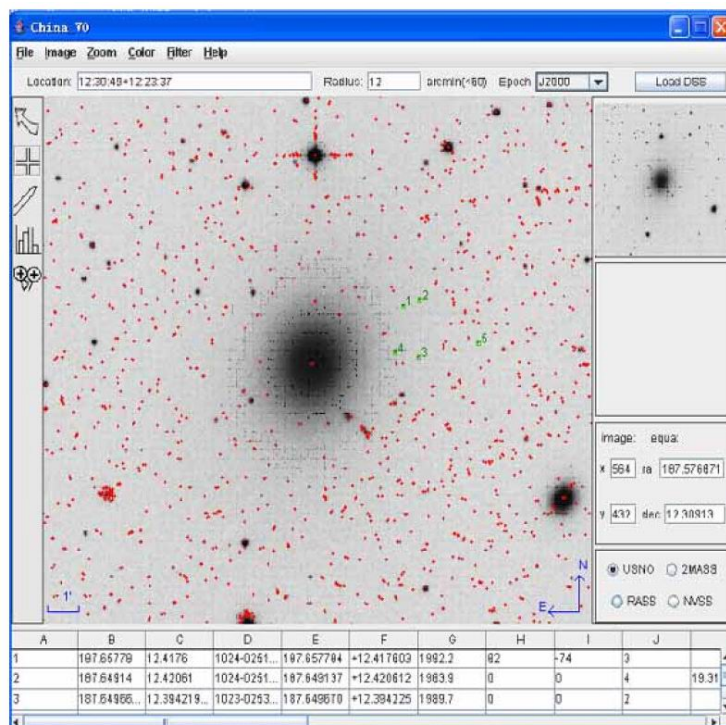


图 26 VO IMPAT 下 M87 与 USNO 星表叠加的效果

PLASTIC 的全称是 Platform for Astronomy Tool InterConnection，这是一个连接天文桌面应用程序的通讯协议。PLASTIC 解决了桌面应用程序的互操作性问题。它是由 AstroGrid 牵头提出的消息交换机制，得到了 IVOA 的认可，并已经成为天文应用客户端的标准通讯协议。目前已经支持了 PLASTIC 协议的桌面 VO 工具包括 TOPCAT、Aladin、AstroGrid AstroScope、VO-DAS GUI Client、VisiVO、VOPlot、VOSpec 等。这些工具的功能包括分布数据库的数据查询、星表显示、星表和图像的联合、数据可视化、光谱数据显示等。PLASTIC 的哲学是天文学家使用一群软件——而不是一个全能软件——完成研究工作，每个软件完成一项专门的功能。而 PLASTIC 则是这些软件的黏合剂，将这些小的工具软件结合在一起，使得它们之间可以进行数据交互，协同完成一项工作。这个设计哲学也是虚拟天文台的基本思想。

我们在 VO-DAS 的 GUI 客户端集成了 Plastic 协议，实现了与 Aladin 和 TOPCAT 等支持 PLASTIC 协议的应用程序之间的交互协作，如图 27 所示的序列

图说明了 VO-DAS 与 TOPCAT 的协同工作机制。

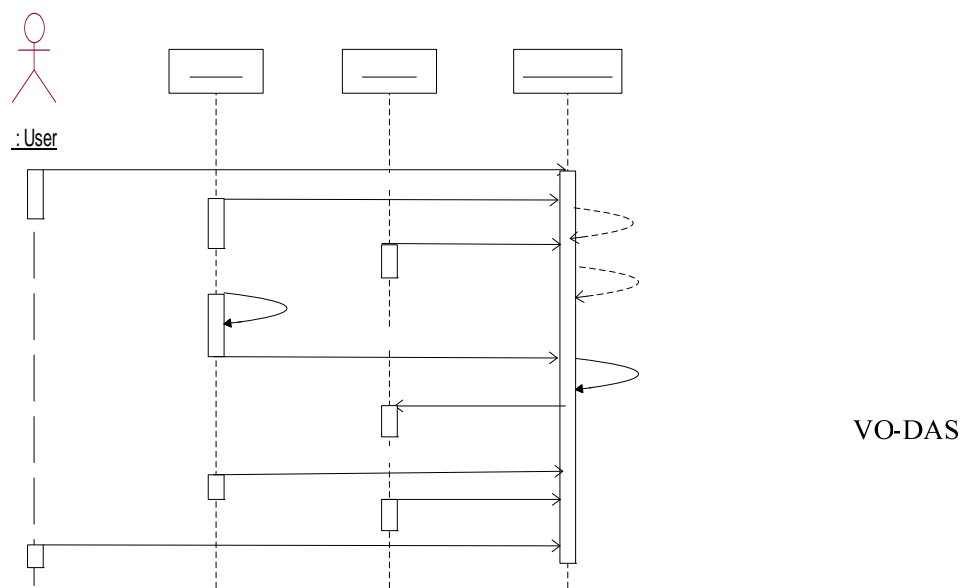


图 27 VO-DAS 和 TOPCAT 基于 PLASTIC 进行交互式操作示意图

PLASTIC 的应用场景通常是这样的：用户通过 VO-DAS 查询星表数据，得到的结果返回到 VO-DAS 的 GUI 客户端上，这时用户希望使用 TOPCAT 查看星表数据，画出简单的图形，这个时候可以使用 PLASTIC 协议自动将星表数据从 VO-DAS 的 GUI 客户端传送到 TOPCAT 上面。当用户希望检验这个星表对应的天区的图像的时候，可以使用 Aladin 搜索图像数据，然后再利用 PLASTIC 将星表数据从 TOPCAT 送到 Aladin。在 Aladin 显示叠加了星表的图像的时候，可以在 TOPCAT 的散点图（例如颜色—星等图）中选择某个感兴趣的数据点（例如可能的蓝离散星或水平分支星），这个时候可以利用 PLASTIC 即时在 Aladin 的图像上点亮对应的数据点。

为了扩种 VO-DAS 系统的数据挖掘和数据分析功能，我们为 MatLab 开发了 Plastic 插件。这样便可以充分利用利用现有的 VO 工具带来的丰富的功能，结合 MATLAB 的强劲的数值计算能力就可以组成非常有竞争力的天文数据挖掘工具集。

MATLAB 虽然可以作为服务器或计算机集群中使用的并行计算平台，但是大多数情况下用户还是喜欢在桌面应用它完成小规模但非常频繁的计算工作。在天文数据挖掘过程中，也并不是所有算法都需要使用并行计算才可以完成，一般规模的模型拟合，数据处理都可以在普通的台式机平台上完成。因此，有很多情

况下，MATLAB 是运行在桌面的。通过对 MATLAB 进行 PLASTIC 扩展，可以让它成为 VO 工具组中的一员，完成数值计算的专项功能。这样，天文学家桌面上的 VO 应用就会更加完整、功能更加强大，VO 工具集也就变得更加实用。

MATLAB 下加入 PLASTIC 协议支持主要需要两个方面的工作，一个是实现 PLASTIC 的消息收发，另一个是能够读写 VOTable 文件。我们利用 Java 程序实现了一个 PLASTIC 客户端，并将它接入到 MATLAB 中，同时创建了一个有 GUI 界面的 MATLAB 函数 plastic。只要在 MATLAB 中运行 plastic 就会启动一个 PLASTIC 客户端 GUI，使用这个 GUI 完成向 Hub 注册或反注册的工作。注册以后 MATLAB 就处于一种等待数据状态，这个时候 MATLAB 仍然可以进行别的工作。当其它 PLASTIC 应用程序向 MATLAB 发送消息的时候，MATLAB 窗口中就会产生提示信息，如果传送的是数据，那么可以利用 PLASTIC 客户端将数据保存到当前的工作空间(workspace)中。在保存过程中，数据已经从 VOTable 格式转换成 MATLAB 的 cell array 格式了。这个转换工作是通过调用 Starlink 提供的 STIL 软件包实现的，STIL 提供了 VOTable 到各种文件格式的转换方法。图 28 是我们开发的 MATLAB 的 PLASTIC 扩展功能的工作画面，图中展示了 MATLAB 通过 PLASTIC 协议从获得 VOTable 数据并转换成 MATLAB 数据类型的过程。

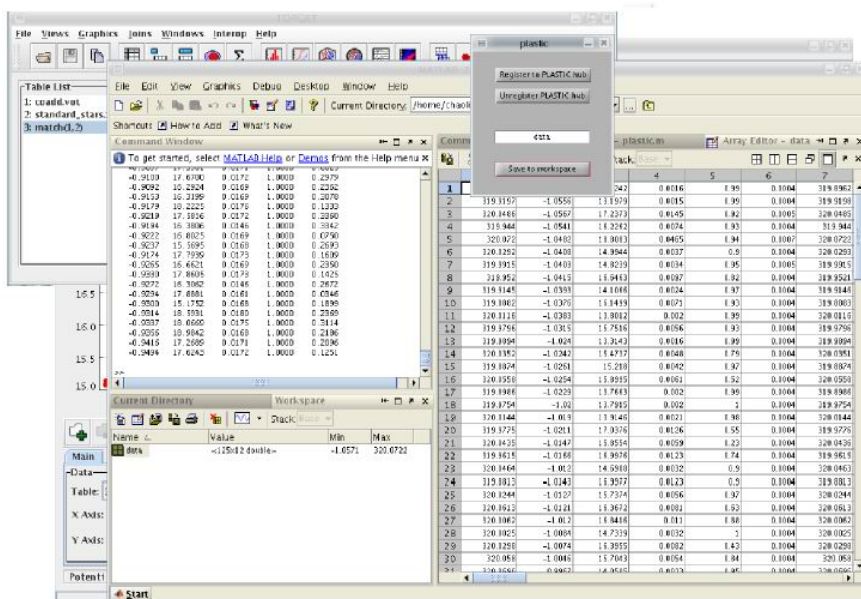


图 28 通过 PLASTIC 协议连接 VO 工具和 MATLAB

4、应用工具开发

在项目执行期间，我们还完成了许多漂亮的“自选动作”，根据研究的需要设计开发并推出了多个应用工具和服务，比如 VOTable 数据格式转换工具 VOFilter、FITS 文件数据库化管理辅助工具 FitHAS、智能化天文信息集成访问环境 SkyMouse 等。下面对这些工具进行简要的介绍。

1)、VOFilter设计开发

2005年8月，项目开发完成了 VOFilter 和 VOTable2XHTML 两个 XML 转换工具。前者是利用 XSLT 语言为开源办公软件 OpenOffice 开发的一个格式转换工具，实现了 OpenOffice Calc 对虚拟天文台数据文件 VOTable 的读取。后者是将 VOTable 格式的文件转换为 XHTML 格式，以便在网页浏览器上显示。相关工作的总结论文已经被 ChJAA 杂志接收。软件访问网址：“<http://services.china-vo.org/vofilter/>”。

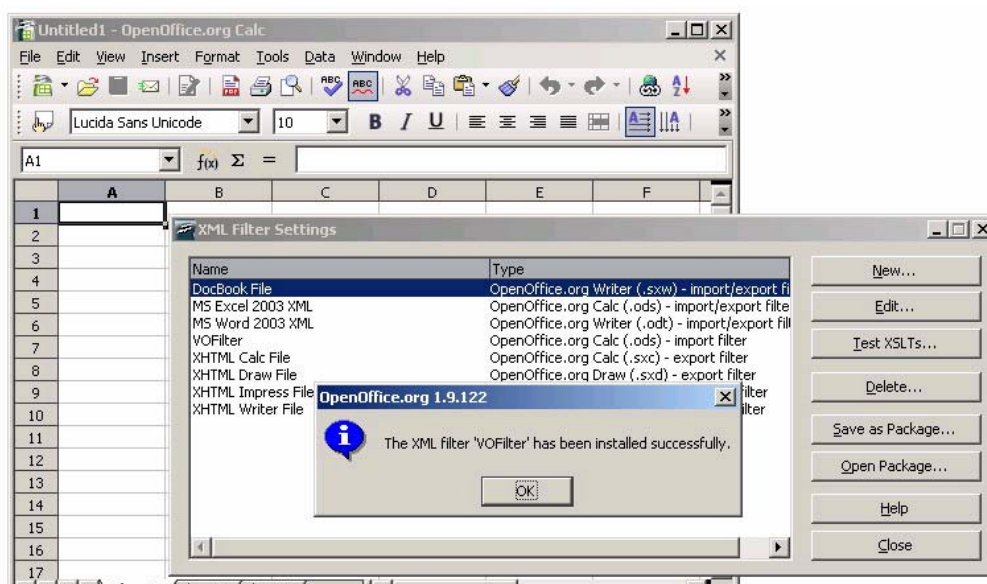


图 29 VOFilter 在 OpenOffice 系统上成功安装

VOFilter 是项目组或者说中国虚拟天文台项目向国际同行发布的第一款工具，已经被 IVOA 和多个虚拟天文台项目收录并使用。

2)、FitHAS设计开发

普适图像传输系统 (Flexible Image Transport System, FITS) 是天文学领域应用最广泛的数据格式。自从 20 世纪 80 年代 FITS 格式被国际天文联合会 (IAU) 正式公布为国际标准以来, 大部分的天文数据都是以 FITS 文件的形式在世界各地的数据中心和天文学家手中保存和交换。采用文件系统方式保存管理的 FITS 数据文件, 很难适应程序化的检索和定位需求。随着 FITS 文件数量的不断增加以及虚拟天文台技术的发展, 为这些以文件形式保存的天文数据提供检索功能的需求越来越迫切。FITS 头以关键词值对的形式确切定义了所属 FITS 文件的结构, 还提供了数据本身诸多特性的重要信息。FITS 头中保存的这些信息可以满足大部分 FITS 数据文件查询检索的需求。将 FITS 头中的信息利用数据库来保存和管理, 可以极大的方便对 FITS 文件的检索和定位。

The screenshot shows a spreadsheet with columns labeled RA, Dec, RAErr, DecErr, Epoch, Fmag, FmagErr, Jmag, JmagErr, and classical. The data rows contain numerical values for these parameters. A scatter plot titled 'Main Title' is overlaid on the spreadsheet, showing a distribution of points in a 2D space.

ID	RA	Dec	RAErr	DecErr	Epoch	Fmag	FmagErr	Jmag	JmagErr	classical
1	N1331213-4098	231.73	44.65	0.27	0.27	1994.51	18.44	0.43	19.13	0.22 Non star
2	N1331213-3972	231.63	44.63							90 Star
3	N1331213-3972	231.63	44.63							21 Star
4	N1331213-3673	231.62	44.58							90 Star
5	N1331213-3673	231.62	44.58	45.10						20 Star
6	N1331213-1118	231.60	44.65	45.00						90 Star
7	N1331213-3805	231.60	44.60	44.95						20 Star
8	N1331213-4055	231.56	44.54	44.90						20 Star
9	N1331213-4070	231.56	44.64	44.85						20 Star
10	N1331213-4099	231.64	44.65	44.80						20 Non star
11	N1331213-4105	231.55	44.60	44.75						19 Star
12	N1331213-3953	231.64	44.62	44.90						21 Non star
13	N1331213-4188	231.55	44.55	44.65						19 Star
14	N1331213-4126	231.83	44.65	44.60						21 Non star
15	N1331213-3473	231.62	44.54	44.50						22 Star
16	N1331213-3884	231.72	44.51	230.75	231.00	231.25	231.50	231.75	232.00	0.22 Star
17	N1331213-3934	231.72	44.61	0.27	0.27	1994.51	17.08	0.42	18.24	99.90 Star
18	N1331213-3934	231.72	44.61	0.27	0.27	1994.51	17.93	0.42	19.90	99.90 Star

图 30 VOTable 数据在 OpenOffice Calc 中显示和处理

FITS 头归档入库系统 (FitHAS) 是中国虚拟天文台项目开发的一个简单实用的 FITS 头信息归档入库工具, 能够方便的将单个或者多个 FITS 文件的主 FITS 头信息导入到数据库中, 为 FITS 数据文件的查询检索奠定基础, 进而为虚拟天文台等更高层次的数据访问服务部署创造条件。最新版的 FitHAS 可以从中国虚拟天文台 FitHAS 网站下载得到, 网址为 “<http://services.china-vo.org/fithas/>”。

FitHAS 的体系结构主要分为核心事务层和用户接口层, 如图 31 所示。系统的核心事务层从用户接口层接收一些必要的参数, 通过对文件系统和数据库管理系统的访问, 并执行相关的数据处理逻辑, 从而进行数据归档和管理的核心操作, 最终将生成操作报告传递给用户接口层。

用户接口层主要负责与用户的交互。除了提供可视化的视图和命令行方式

的控制台，还将从用户那里接收参数并将操作报告反馈给用户。系统的参数设置在用户接口层提供交互接口，允许配置文件的读取和更新。

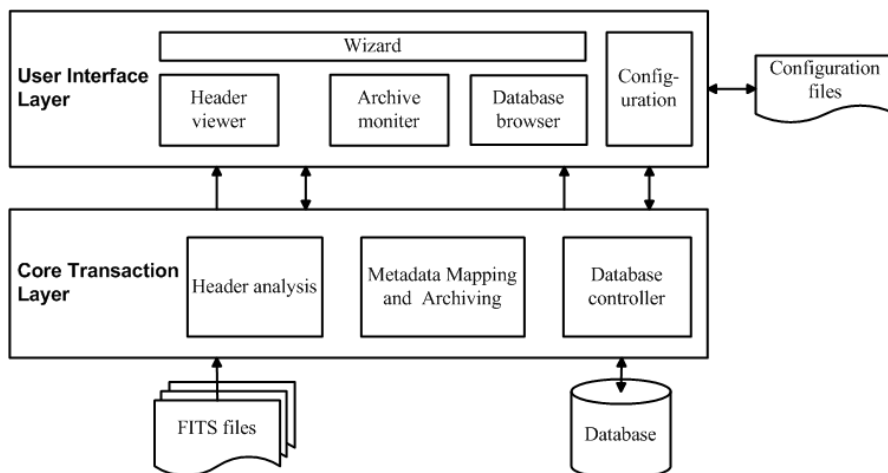


图 31 FitHAS 体系结构图

根据 FitHAS 的操作模式，用户界面层各功能按照操作向导的逻辑流程进行组织，将几个模块按照一定的顺序划归于 4 个操作向导中（如图 32）。这四个操作向导分别是：FITS 头浏览、单个 FITS 文件归档、成批 FITS 文件归档、数据库浏览。



图 32 FitHAS 主界面

FitHAS 最核心的功能是将 FITS 文件主 FITS 头中的关键词值对导入数据库，并为每个 FITS 文件记录添加文件保存路径等必要信息，以便利用数据库进行 FITS 文件的管理和检索。FitHAS 具体的功能包括如下几个方面：

- FITS 头信息查看，查看指定 FITS 文件的头信息；
- 根据 FITS 文件头中关键词的布局结构创建数据表；
- 将符合入库条件的 FITS 文件的主 FITS 头信息导入数据库；
- 查看被导入数据库的 FITS 头信息；

FitHAS 通过友好的图形界面向导方式实现 FITS 文件头信息浏览，将单个或者多个 FITS 文件的头信息导入数据库的功能。整个系统的研发宗旨是简单实用，用户不需要任何编程背景，只需要对 FITS 文件结构和数据库知识有基本了解。

随着国内许多大型观测项目（比如 LAMOST、FAST、HXMT、SVOM、丽江 2.4 米望远镜等）的建成和投入观测，我国天文数据的获取能力将大大增强。中国天文学将从单纯的数据引进向与国际同行共享观测数据转变。随着观测能力的增强，数据量的不断增加，科学的数据管理和维护工作必然越来越重要。当今国际天文数据已经达到 TB 量级，不久将突破 PB 量级。海量数据的管理必需依靠得力的软件工具作为协助。数据库管理系统以其天生的技术和性能优势必然会在海量数据管理过程中发挥重要甚至核心的作用。

FitHAS，将 FITS 数据文件的头信息导入数据库，提供了 FITS 文件数据库化管理的第一步，为 FITS 数据文件的在线检索和访问提供前提条件。相信随着虚拟天文台技术的不断成熟和推广，网络化数据管理、数据发布、数据访问需求的不断增加，越来越多的类似 FitHAS 的实用工具将被开发并得到广泛应用。

3)、SkyMouse设计开发

随着数据库、网络技术以及近年来 Web 服务、Grid 服务的发展和成熟，网络应用日益普及，在线的天文资源和服务越来越丰富，种类越来越多，形式越来越多样化个性化。天文研究已经开始步入 e-Science 时代。

但另一方面，网络资源使用的简易程度并没有随着网络资源的丰富而增加，反而使得用户越来越难以发现自己真正需要的内容和信息。例如，某个服务被开发出来但不为人知，或是服务比较复杂难以使用。同时，单个服务的功能往往局限性很大，只能通过多个服务协同工作才能完成更复杂的功能。在这样一种资源空前丰富、服务种类繁多的背景下，一个便捷高效的智能化天文信息集成访问平台，是天文学家和公众迫切的需求。

SkyMouse，这个智能化天文信息集成访问系统将改变这种局面。SkyMouse 以屏幕取词为入口，借鉴电子词典类软件中鼠标取词的方便性，结合后端服务系统个性化的服务定制环境，为用户提供一个便捷的、集成化的天文信息检索系统。

目前项目已经完成并推出了 SkyMouse 1.0 公开测试版。

SkyMouse 的系统架构如图 33 所示，整个系统由服务器端、客户端和后端服务资源库组成。

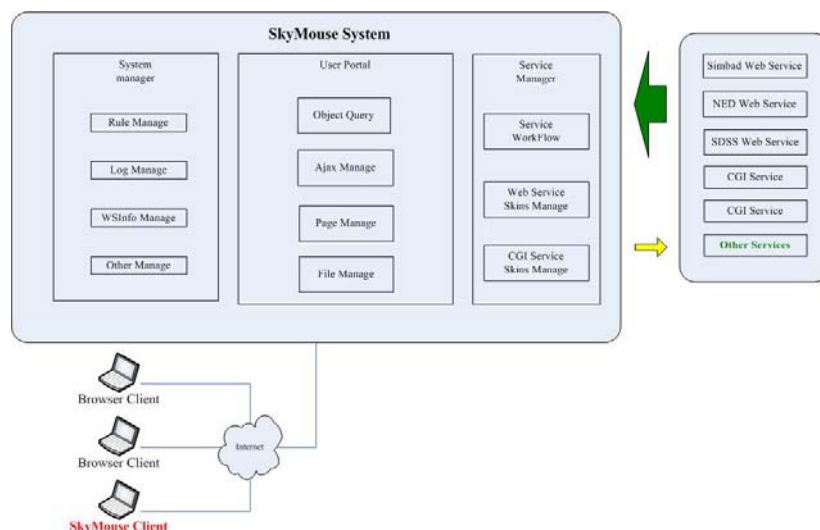


图 33 SkyMouse 系统架构图

SkyMouse 服务器端系统由三部分组成：系统管理模块、用户门户，服务管理模块。

- 系统管理模块：完成 SkyMouse 系统管理功能功能，包括四部分：
 - ✧ 规则管理子模块负责对预制用户的规则管理和其他规则管理功能；
 - ✧ 日志管理子模块负责网站日常的日志管理。包括查询关键字统计和服务调用日志的记录和分析功能；
 - ✧ Web 服务信息管理子模块负责 Web 服务的基本信息更新修改删除等功能，该模块负责提供调用 Web 服务所需参数，方法等信息；
 - ✧ 其他信息管理子模块负责以上三种模块外系统管理所需功能。
- 用户门户： SkyMouse 系统和用户交互的界面，包括四部分：
 - ✧ 关键字查询子模块负责对用户传递的关键字进行分析和查询的功能；
 - ✧ Web Service Skins 子模块负责将 Web Service 调用返回的的结果展现给用户。该模块将包括一个通用的基本展示页面和其他的定制页面。同时通用的 Web Service 调用引擎也包括在该模块中；
 - ✧ CGI Service Skins 子模块负责将 CGI 调用返回的的结果展现给用户。该模块将包括一个通用的基本展示页面和其他的定制页面。同

时通用的 CGI 调用引擎也包括在该模块中。

- 服务管理模块：该模块负责 SkyMouse 系统和互联网上的 Web Service 和 CGI 程序的交互，包括三部分：
 - ◇ workflow 引擎负责对多个 Web 服务进行协同调用的功能；
 - ◇ Ajax 管理子模块负责通过 Ajax 无刷新的调用 Page 管理子模块所获得调用 Web 服务结束后的展示页面，该模块可以提供更好的用户体验；
 - ◇ Page 管理子模块负责调用相应 Web 服务所需展示的页面；
 - ◇ 文件管理子模块负责对图片，视频等文件的远程获取功能，同时该模块还承载系统缓存的功能，对常用的关键字调用页面和结果进行缓存以加快下一次的调用速度。

客户端组成：SkyMouse 系统不需要客户端也可以访问，用户可以直接使用浏览器访问，支持的浏览器有 IE、FireFox、Netscape 等常用的浏览器。同时 SkyMouse 系统还提供一个基于屏幕取词的客户端。

SkyMouse 系统采用的软件开发技术与工具如下：

- ◇ 软件架构模式采用 B/S 模式(浏览器/服务器模式)；B/S 模式便于客户端的零维护管理；程序发布不受影响；需要的配置资源紧紧是浏览器，不需要其它而外资源。
- ◇ 开发语言采用 .NET；基于微软的 .Net 架构，利于构建基于 WEBSERVICE 的应用服务。
- ◇ 数据库采用目前采用 SQLSERVER2000；目前 SkyMouse 系统的数据库是基于微软的 SQLSERVER，SQLSERVER 与其开发语言拥有天然的结合力，便于开发。但是 SkyMouse 系统本身对于数据库并无特定要求，使用 Oracle，MySQL 甚至是 XML 作数据库都可以。

5、虚拟天文台基础上的公众教育

丰富的资源和网络化的访问是 VO 的绝对优势，这注定了它会在天文科普教育上发挥作用，“让普通公众有机会和天文学家一样做出一流的科研成果”。项目执行期间，China-VO 在公众教育方面开展了许多工作，取得了很好的科学和

社会效益。这其中包括能够真正提供上下五千年的“万年历”——天文星历计算服务（见刘高潮、田海俊论文），网络探究式教学研究（见吴娟论文），通过鼠标取词实现对全球天文信息进行智能化访问的 SkyMouse（见孙华平论文），及时更新的中英文双向天文学词典数据库，以及国内最大、资源最丰富的天文科普平台——宇宙驿站等。

1)、真正的万年历

很多公众、学生和天文爱好者对太阳月亮等天体的位置，日月食发生的时间以及万年历等很感兴趣，因此我们以 Swiss Ephemeris 星历表为基础开发了我国虚拟天文台星历计算服务。

星历的常用算法，比如行星和恒星的计算、食现象、行星现象以及黄道十二宫的计算等等，是星历研究人员或天文软件开发者必须掌握的基本计算能力，对于这些繁琐的计算，一般计算人员往往抛弃诸多实际因素，抽象成过于理想化的模型，比如忽略地球表面大气层的影响、相对论性偏差以及光在重力场中的弯曲等等，虽然简化了计算，但计算的精确度也受到一定的折扣；对于目前每一个星历研究人员或天文软件开发者来讲，每次开发都可能面临这些常用的但有繁琐的天文算法，显然这不符合“分工合作”的原则，效率必定不高。如果我们能够将一套较为成熟的、计算精确的星历算法搭建成一个服务“工厂”，仅需开发人员输入一定的参数，我们的“工厂”就可以为其计算出精确的“半成品”，利用这些“半成品”，开发人员就可以很方便地“组装”成自己所需的“成品”。这里所比喻的天文计算服务“工厂”，即为我们设计的星历计算平台——China-VO Ephemeris 计算平台。

Swiss Ephemeris 是由 Astrodienst 基于美国宇航局 JPL 实验室发布的 DE406 星表拓展而来的一个高精度的星历表。它不是一个面对终端用户的产品，而是面对天文软件程序开发者的一个工具集。它给我们提供了如下一些计算功能：行星和恒星的计算、食现象和行星现象的计算、日期和时间转换功能、黄道十二宫的计算以及一些辅助功能。

我们首先利用 JNI 技术将 Swiss Ephemeris 提供的星历计算的 C 程序接口封装成 JAVA 程序接口，然后利用 Apache Axis 将 JAVA 程序接口发布成一系列

Web Services 接口，最终形成一套基本天文计算服务平台方案，如图 34 所示。

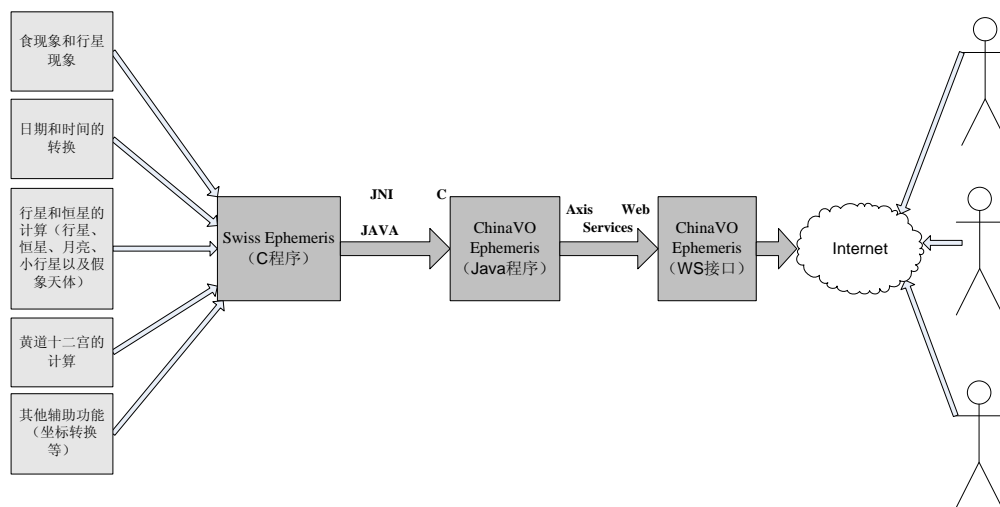


图 34 天文计算服务平台方案

这个平台提供了网页界面和 Web Service 界面，无论你是程序高手还是纯粹的天文爱好者都可以使用我们的服务。星象计算网站包括如下查询功能：利用技术将接口
行星以及太阳月亮的位置（赤道坐标系、黄道坐标系、直角坐标系）查询；全国主要城市的太阳月亮及九大行星的升落时间查询；全国主要城市的晨昏蒙影时间查询；1000 余颗基本恒星位置的查询；日月食时间查询；跨越一万年的万年历查询；年月日与儒略日之间的转换查询等。网站网址：“http://www.china-vo.org:8080/gcroot/new_first.htm”。

2)、SkyMouse 智能信息系统

一位天文学家正在认真的研读一篇重要文献。他对文中提到的一颗发现不久的新天体很感兴趣，希望能了解更多的信息，便将鼠标移到了这颗天体的名字上。天文学家计算机系统上的网络浏览器自动启动，很快在浏览器的页面上出现了关于这颗新天体的坐标、亮度、类型等基本信息，同时在不断刷新的页面上出现了这颗天体被世界各地天文望远镜以及空间望远镜观测的数据和图像；涉及这颗天体的已经公开发表的甚至刚刚投稿的论文目录也出现在了天文学家的浏览器上。通过浏览器上丰富的信息，天文学家很快了解了这颗新天体的情况。这就是 SkyMouse，一个智能化的天文信息集成访问平台。

这套系统的开发将改变当前天文学家访问网络资源的模式，将对世界范围

内日益丰富的天文数据和信息的检索查询过程简化到轻轻的“滑动鼠标”来完成。

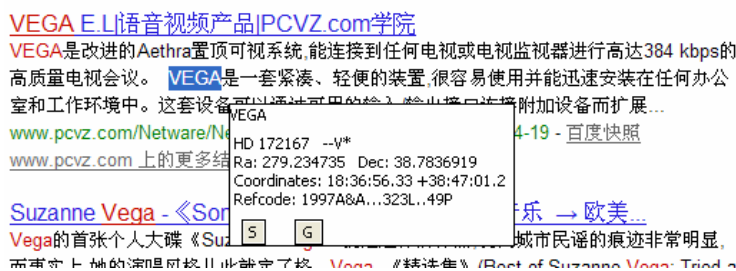


图 35 SkyMouse 屏幕取词后返回的织女星基本信息

2007年5月14日温家宝总理对同济大学师生讲话时借用一句哲言指出“一个民族多一些经常仰望天空的人，这个民族就大有希望；而一个民族总是看自己脚下的一点事情，那她很难有美好的未来。我们的民族是大有希望的民族，我希望我们的同学应当经常地仰望天空。”该项目将会吸引更多的人“经常的仰望星空”。



图 36 SkyMouse 服务器端返回的综合信息

3)、在线天文词典和宇宙驿站

项目组多年来与中国天文学会名词审定委员会（简称名词委）开展了全面的合作，协助名词委建立了名词数据库、开设了天文名词讨论组，为广大名词爱好者和用户提供了在线天文词典服务。这个数据库可以通过中国天文学会天文学名词审定委员会授权的“英汉双向天文学词典”网站进行检索，网址是

“<http://www.lamost.org/astrodict/index.html>”。通过这个网站还可以向名词审定委员会的专家推荐新词；订阅天文名词邮件列表，参与天文学名词相关的讨论等。

特别需要指出的是，在我们的不懈努力下将名词委 60 多年的工作成果“英汉天文学名词数据库”融入到了《金山词霸》系统中，受到业内专家和广大用户的极大欢迎。



图 37 金山公司收录天文数据库时的在线宣传图片

《英汉天文学名词数据库》是中国天文学会天文学名词审定委员会审定的英汉对照天文学名词汇编，共收录天文学名词 22300 余条。《数据库》是迄今收词最多的天文学名词库，内容涵盖 2000 年出版的《英汉天文学名词》（天文学名词审定委员会审定，李竞、许邦信主编，）以及 2001—2007 年《天文学进展》刊出的六批《天文学名词的推荐译名》；还包括 1998 年全国科学技术名词审定委员会公布的《天文学名词》。《数据库》收词及时，并将不断增补新词。《数据库》中既有全国科学技术名词审定委员会规范的天文学基本名词，又有近现代天文书刊和中国古代文献常见的天文学名词，还有恒星、行星、卫星、小行星、彗星、流星雨、陨星坑、月面环形山、星座、二十八宿、黄道十二宫、二十四节气、星云、星团、星系等的专用名，以及国际知名的天文台站、天文卫星和空间探测器的专用名。

金山软件公司在其包括《金山词霸》、“爱词霸”在线词典等在内的软、硬件产品中使用此数据库。这意味着今后广大天文爱好者和天文学家利用《金山词霸》这个深受国内外用户欢迎的电子词典工具就可以方便地使用《数据库》了。

本项目人员开发、维护和管理的国内最大规模的天文科普网络“宇宙驿站”在 2005 年由中国互联网协会主办的“第一届全国优秀科普网站与栏目评选活动”中获得“鼓励奖”。在这个天文网络科普平台上已经拥有一百多个天文科普网站

和栏目，是国内最丰富的、最受欢迎的天文科普网络，日访问量数万人。这为虚拟天文台环境下的科普教育研究提供了良好的资源基础和用户群体。

此外，项目组充分利用杂志、报纸、网络等媒体宣传虚拟天文台相关知识，发表科普文章和新闻报道数十篇，刊登和发表在《天文爱好者》、《科学》、《环球科学》、《航天报》、《科学画报》、《世界科学》、新浪科技、南方网等媒体上。

四、队伍建设与人才培养

项目自从获得资助以来非常重视队伍建设，目前已经形成了一支由项目负责人（研究员）、副研究员、博士研究生、硕士研究生、项目助手构成的相对稳定的、多层次的研究开发队伍，目前参加项目工作的固定工作人员有6人，其中研究员一名，副研究员两名，助理研究员三名。同时，项目组中还包括多名研究生和临时聘用人员。

自2004年至今，以本项目研究内容为论文题目毕业的研究生共18名，其中博士3名，硕士13名，学士2名，具体情况见表2。这些研究生的毕业论文可以从项目网站上全文下载。毕业的研究生中多人仍然在从事着天文相关的研究和教学工作。

为了增强项目人员的自身素质，2005年3月，项目组开展了一次为期2周的项目内部集中培训。所有项目组成员参加了培训，分工学习讲授了本项目计划和目标、虚拟天文台相关标准、工具使用、最新进展等内容。通过集中培训让项目成员对项目情况、国内外发展情况、技术背景、领域标准、工具服务的使用有了全面的了解。

培训的背景材料，项目成员所做报告文件全部公布到了项目的网站上。不但可以供项目成员参考，也是国内外其它对虚拟天文台领域感兴趣的人学习参考的好地方。培训资料网址：“<http://www.china-vo.org/cn/events/training05/>”。

五、学术会议、国际合作与交流

虚拟天文台是一个新兴的研究领域和方向，需要国内外相关人员的密切合

作。本项目非常重视与国内外天文学家和虚拟天文台同行的交流与合作。下面从三个方面介绍学术交流方面的情况。

表 2 历年虚拟天文台领域毕业学生一览表

序号	研究生姓名	论文题目	学位类型	毕业年份
1	崔辰州	中国虚拟天文台系统设计	博士	2003
2	张彦霞	多波段天体物理中的自动分类方法研究	博士	2003
3	桑健	中国虚拟天文台的数据库和数据服务	硕士	2004
4	邵慧娟	中国虚拟天文台可视化设计及实现	硕士	2004
5	王帆	中国虚拟天文台资源元数据管理系统原型设计	学士	2004
6	王晓倩	中国虚拟天文台中的软件集成	硕士	2005
7	刘高潮	虚拟天文台天文教育平台研究	硕士	2005
8	李长华	中国虚拟天文台资源管理系统的设计	硕士	2005
9	吴娟	网络探究教学研究	硕士	2005
10	徐祯	天文巡天数据管理发布系统的设计与实现	学士	2005
11	刘波	中国虚拟天文台的数据访问与互操作研究	硕士	2006
12	孙华平	SkyMouse 天文服务搜索整合系统的设计与实现	硕士	2007
13	路勇	虚拟天文台数据访问服务注册系统的设计与实现	硕士	2007
14	田海俊	虚拟天文台数据访问服务(VO-DAS)之任务调度研究及 VO-DAS 的应用	硕士	2007
15	李文	虚拟天文台环境下的海量数据存储与访问技术研究	硕士	2007
16	李丽丽	天文中的分类和回归方法初探	硕士	2007
17	王丹	基于大型巡天数据的测光红移算法研究—算法研究与工具开发	博士	2007
18	刘超	基于虚拟天文台的数据挖掘技术及其在银河系晕结构研究中的应用	博士	2008
19	高丹	海量天文数据融合系统的开发与数据挖掘算法的研究	博士	2008
20	杨阳	VO-DAS 客户端系统的设计和应用	硕士	2008

1、中国虚拟天文台年会及用户培训

从 2001 年 9 月在北京召开的第一次虚拟天文台研讨会开始，到 2007 年，中国虚拟天文台年会已经成功召开了 6 次，成为国内虚拟天文台研究领域乃至天文信息领域的标志性会议系列。每年秋末冬初召开，会议规模在 30—50 人之间。历届年会情况见表 3。其中三次会议有台湾地区代表参加，一次有英国虚拟天文台项目代表参见。每次会议都设有专门的网站，所有报告人的报告全部对外释放。相关网址请参考：“<http://www.china-vo.org/cn/>” 学术会议栏目。

虚拟天文台年会已经成为展示国内虚拟天文台领域研究成果、天文学家与技术开发人员交流的重要平台。今后该年会还将继续进行下去，为虚拟天文台的研究和天文信息化工作起到应有的推动作用。

表 3 历年虚拟天文台年会一览

届次	时间	地点	主题
1	2001.9	北京	
2	2003.9	北京	
3	2004.12	武汉	
4	2005.11	威海	从技术到科学
5	2006.11	桂林	天文学研究信息化
6	2007.11	广州	使用 VO 从现在开始

为了更深入的了解天文学家对 VO 的期望和需求，更好的向天文学家介绍和推广 VO，从 2006 年开始，我们在每年的中国天文学术年会期间都组织“虚拟天文台用户培训”。2006 年在上海，2007 年在广州，两次用户培训都吸引了不少青年天文学家和研究生参加。随着 VO 的逐步成熟和完善，相信这样的用户培训将会变的越来越重要。

2、IVOA会议与国际合作

国际虚拟天文台联盟每年召开两次互操作会议，春季、秋季各一次。同时，国际虚拟天文台联盟执委会会议一年举行 4 次，两次是面对面会议，两次电话会议。自从 2002 年中国虚拟天文台项目加入国际虚拟天文台联盟以来，我们参加

了大部分的 IVOA 执委会会议，并及时提交进展报告。

IVOA 互操作会议是 IVOA 主办的虚拟天文台技术领域专门会议，由各工作组汇报、交流、讨论、协商虚拟天文台各方面标准的相关工作。IVOA 互操作会议是 IVOA 最高规格的会议，每年举行两次。按照惯例，春季会议规模较大，单独举行。秋季会议规模较小，一般与 ADASS、IVOA 小项目会议等其它会议共同举行。

自 2002 年 IVOA 成立以来，历届互操作会议的基本情况如表 4 所示。

表 4 IVOA 互操作会议一览表

时间	承办 VO	承办单位	地点
2002.10.17	NVO	JHU	Baltimore (美国)
2003.5.12-16	AstroGrid	IoA	剑桥 (英国)
2003.10.16-17	VO-France	CDS	斯特拉斯堡 (法国)
2004.5.24-28	NVO	CfA	波士顿 (美国)
2004.9.27-29	VO-India	IUCAA	Pune (印度)
2005.5.16-20	VO-Japan	NAOJ	京都 (日本)
2005.10.6-7	ESAC-VO	ESAC	马德里 (西班牙)
2006.5.14-19	VO-Canada	CADC	维多利亚 (加拿大)
2006.9.20-22	VO-Russia	俄罗斯科学院天文所	莫斯科 (俄罗斯)
2007.5.14-18	China-VO	国家天文台	北京 (中国)
2007.9.27-28	AstroGrid	剑桥大学	剑桥 (英国)
2008.5.19-23	VO-Italy	INAF	的里雅斯特 (意大利)

我们不仅是被动参与国际虚拟天文台联盟的学术交流工作，还非常主动得组织承担相关的工作。2003 年夏季，本项目提出了召开国际虚拟天文台联盟小项目会议的提议。提议得到 IVOA 的积极响应和支持，并在 2003 年在北京成功举办了第一届 IVOA 小项目会议，得到国际虚拟天文台同行的高度评价。目前，由我们提出的国际虚拟天文台联盟小项目会议已经成为国际虚拟天文台联盟的系列会，第二届会议于 2004 年 10 月在印度召开，第三届于 2006 年 10 月在俄罗斯召开。由于我们在国际虚拟天文台联盟中对小型虚拟天文台项目的突出贡献，应印度虚拟天文台邀请与其联合组织了 2004 年十月在印度举办的 IVOA 大会。并应俄罗斯虚拟天文台邀请将与其联合组织 2006 年的 IVOA 互操作会议。

经过中国虚拟天文台项目的努力争取，IVOA 决定 2007 年春季的 IVOA 互操作会议由中国虚拟天文台项目承办，于 2007 年 5 月 14 日至 18 日在北京举行。2007 年的春季 IVOA 互操作会议是继 2003 年在北京举行的 IVOA 小项目会议后又一次由 China-VO 承办的国际虚拟天文台领域的学术会议。

2007年5月14日，国际虚拟天文台联盟（IVOA）2007年春季互操作会议在北京东方文化酒店开幕，来自15个国家的100多名天文学家、信息技术专家参加了此次会议。

此次会议可以说是在国际虚拟天文台联盟互操作会议历史上最成功的一次，成果显著，对IVOA的未来发展将会起到重要的推动作用，意义重大。在为期5天的会议期间共召开全体大会两场，分会28场，IVOA执委会2场。会议报告超过150多个，涉及到虚拟天文台科学、技术、规范、管理等方方面面，全面展示了各国虚拟天文台项目在最近一年时间里的最新进展。在此次会议期间十多个虚拟天文台的国际标准草案被深入讨论和被批准为IVOA标准。此次会议期间IVOA召开了两次执委会会议，认真商讨IVOA的未来发展问题。这是第一次一个IVOA互操作会议期间召开两次IVOA执委会，意味着此次会议必将在IVOA的历史上留下深远的影响。

我们充分利用这次展现China-VO数年来研究开发成果，与各国VO项目交流经验的好机会。在此次会议上，中国虚拟天文台项目作为东道主提交了三份口头报告，介绍了中国在虚拟天文台领域的工作和进展，得到与会各国同行的赞誉。同时，这些交流活动也为中国虚拟天文台项目与国际上的虚拟天文台项目开展合作和交流打下了非常好的基础。会议期间和会议结束后便有美国、印度、英国、日本等方面提出希望与中国虚拟天文台开展合作。

虚拟天文台作为天文学和信息技术融合的典型应用一直得到信息技术领域的高度重视。此次国际虚拟天文台联盟互操作会议同样也得到包括新浪科技在内的媒体的关注。会议开幕的消息在开幕当天便在新浪科技频道报道，然后被科学网、南方网等多家媒体转载，收到了很好的宣传和科学教育效果。

除了积极参与IVOA的学术、管理活动，本项目还与国际上多个虚拟天文台项目保持着紧密的合作关系。China-VO与美国NVO、英国AstroGrid、法国VO-France、印度VO-India、韩国KVO、日本JVO等都保持着密切的合作和人员往来。此外，本项目与台湾地区的同行一直保持着密切的合作。台湾同行先后四次赴大陆进行学习以及参加我们组织的学术会议。

此外，应国际天文学联合会（IAU）的邀请，本项目成员崔辰州于2007年3月以报告人的身份参加了在马来西亚举办的国际青年天文学家讲习班（IYAS）。

在讲习班上，崔辰州为来自马来西亚、印度、印尼、菲律宾、越南、泰国、尼泊尔、朝鲜、中国等不同国家的青年学生系统讲授了天文软件、天文数据库和虚拟天文台知识。

3、与国内相关领域的交流与合作

项目执行期间，本项目的成员充分利用各种机会和场合宣传虚拟天文台的知识和本项目的研究成果，与 IT 等相关学科领域的专家交流沟通。项目成员参加了很多相关领域的会议，比如：中美高级网络研讨会、网格世界（中国）大会、CODATA 会议、SPIE、ADASS、GLORIAD、亚太网络技术研讨会、两岸三院信息技术研讨会等。这些学术活动增进了 IT 等相关科学领域的人士对虚拟天文台乃至天文学的了解。

六、发表文章目录

1、已发表：

1)、SCI索引

- [1] C. Liu, J. Hu, H. Newberg, and Y. Zhao. Candidate Milky Way satellites in the Galactic halo. *A&A*, 2008, 477, 139-145
- [2] Chen-Zhou Cui, Markus Dolensky, Peter Quinn, Yong-Heng Zhao and Françoise Genova. VOFILTER: Bridging Virtual Observatory and Industrial Office Applications. *ChJAA*, 2006, 1: 379
- [3] Zhang Yanxia, Zhao Yongheng, “NNB, ADTree and MLP Comparison in Separating Quasars from Large Survey Catalogues”, 2007, *ChJAA*, 7, 289
- [4] Zhang Yanxia, & Zhao Yongheng. “Automated clustering algorithms for classification of astronomical objects”, 2004, *A&A*, 422, 1113
- [5] Zhang Yanxia. “Research on Automatic Classification Methods in Multiwavelength Astrophysics”, 2004, *PASP*, 116, 184
- [6] Li Lili, Zhang Yanxia, Zhao Yongheng, Yang Dawei, “Estimating Photometric Redshifts with Artificial Neural Networks and Multi-parameters”, 2007, *ChJAA*, 7, 448
- [7] Wang Dan, Zhang Yan-Xia et al., 2007, “Kernel Regression For Determining Photometric Redshifts From Sloan Broadband Photometry”, *MNRAS*, 382, 1601

2)、EI收录

- [8] Zhang Yanxia, Luo Ali & Zhao Yongheng. “Outlier detection in astronomical data”, 2004, *Astronomical Data Analysis II, Proc. of SPIE*, 5493, 483-490
- [9] Zhang Yanxia, Luo Ali & Zhao Yongheng. “An automated classification algorithm for multiwavelength data”, 2004, *Astronomical Data Analysis II, Proc. of SPIE*, 5493, 521-529
- [10] Luo Ali, Zhang Yanxia, Zhang Jiannan & Zhao Yongheng. “Mining the LAMOST spectra archive”, 2004, *Astronomical Data Analysis II, Proc. of SPIE*, 5493, 178-185
- [11] Chenzhou Cui, Yongheng Zhao, Xiaoqian Wang, Jian Sang, Ze Luo. Integrating Existing Software Toolkits into VO System. *Proceedings of SPIE*, 2004, Volume 5493, pp. 233-241
- [12] Luo Ali, Zhang Yanxia, & Zhao Yongheng. “Design and implementation of the spectra reduction and analysis software for LAMOST telescope”, 2004, *Astronomical Data Analysis II, Proc. Of SPIE*, 5496, 756-764
- [13] Gao Dan, Zhang Yanxia, Zhao Yongheng, “A system integrated with query, cross-matching and visualization”, 2006, *Advanced Software and Control for Astronomy*. Edited by Lewis, Hilton, Bridger, Alan. 2006, *Proceedings of the SPIE*, V. 6274, p. 6274:14
- [14] Wang Dan, Zhang Yanxia, Cui Chenzhou, Zhao Yongheng, “Software kits for measuring photometric redshifts”, 2006, *Advanced Software and Control for Astronomy*. Edited by Lewis, Hilton, Bridger, Alan. *Proceedings of the SPIE*, V. 6274, p. 6274:13
- [15] Liu, Chao; Wang, Dan; Liu, Bo; Gao, Dan; Cui, Chenzhou; Zhao, Yongheng. An astronomical data mining application framework for virtual observatory. *Proceedings of the SPIE*, 2006, Volume 6274, pp. 6274:15

3)、核心期刊

- [16] Zhang Yan-xia, Zhao Yong-heng, “Science, Tools and Applications of the Virtual Observatory”, 2006, *Progress in Astronomy*, 24(3), 189
- [17] Zhang Yanxia & Zhao Yongheng. “The Science of the Virtual Observatory”, 2004, *Progress in Astronomy*, 22 (4), 350
- [18] Zhang Yanxia, & Zhao Yongheng. “The outlier detection”, 2004, *Progress in Astronomy*, Vol. 22, No. 1, p. 1
- [19] Li Li-li, Zhang Yan-xia, Zhao Yong-heng, Yang Da-wei, “The Application of Artificial Neural Network in Astronomy”, 2006, *Progress in Astronomy*, 24 (4) ,285
- [20] Gao Dan, Zhang Yanxia, Zhao Yongheng, “Automatization of putting astronomical catalogues into database”, 2007, *PNAOC*, 4(1),53
- [21] Gao Dan, Zhang Yan-Xia, Zhao Yong-Heng, “The realization of cross-identification based on huge multi-wavelength catalog data”, 2005, *PNAOC*, 2, 186
- [22] Liu Bo, Cui Chen-Zhou, Zhao Yong-Heng. “Construction of the Sky Node system for Chinese Virtual Observatory”. *PNAOC*, 2006, 3(4), 355-364
- [23] TIAN Hai-Jun, ZHAO Yong-Heng, CUI Chen-Zhou & ZHENG Xiao-Ping. “Design and implementation of China-VO Ephemeris computing service based on web services”. *PNAOC*, 2007, 4 (3): 283-287
- [24] 王晓倩, 崔辰州, 赵永恒. 中国虚拟天文台软件集成. *天文研究与技术-国家天文台台刊*,

- 2005, 2 (4): 293-301
- [25] 崔辰州, 赵永恒. 中国虚拟天文台研发策略与重点. 天文研究与技术-国家天文台台刊, 2004, 1 (3): 203-209
- [26] 崔辰州, 赵永恒. 中国虚拟天文台体系结构. 天文研究与技术-国家天文台台刊, 2004, 1 (2): 140-151
- [27] 刘高潮, 崔辰州, 郑小平, 赵永恒. 虚拟天文台教育门户建设构想. 天文研究与技术-国家天文台台刊, 2004, 1 (2): 133-139
- [28] 邵惠娟, 赵永恒. 中国虚拟天文台可视化服务. 天文研究与技术-国家天文台台刊, 2004, 1(2): 152-159
- [29] 林刚华. 国内太阳数据 VO 化初步方案. 天文研究与技术-国家天文台台刊, 2004, 1 (3): 210-215
- [30] 桑健, 崔辰州, 赵永恒. 中国虚拟天文台数据访问服务. 天文研究与技术-国家天文台台刊, 2004, 1 (3): 216-228
- [31] 乔翠兰, 吴娟, 刘高潮, 郑小平. 虚拟天文台与科学教育. 天文研究与技术-国家天文台台刊, 2004, 1 (3): 229-233
- [32] 路勇, 刘超, 崔辰州, 赵永恒. “VO-DAS Registry 系统的设计与实现”, 天文研究与技术. 2007,4(4):355-359

4)、国际会议

- [33] Zhang Y., Zhao Y., “Data Mining the Multiwavelength Data”, 2006, Astronomical Data Analysis Software and Systems XIV ASP Conference Series, Vol. 351, Proceedings of the Conference Held 2-5 October 2005 in San Lorenzo de El Escorial, Spain. Edited by Carlos Gabriel, Christophe Arviset, Daniel Ponz, and Enrique Solano., p.173
- [34] Zhang Yan-xia, Zheng Hongwen, Zhao Yongheng. “Preselecting AGN candidates from multi-wavelength data by ADTree”, Probing Galaxies through Quasar Absorption Lines, IAU Colloquium Proceedings of the International Astronomical Union 199, ed. by P.R. Williams, C.-G. Shu and B. Menard. Cambridge: Cambridge University Press, 2005, pp.481-483
- [35] Zhang Yanxia, Zhao Yongheng, “Preselect Quasar Candidates by Automated Methods”, The Central Engine of Active Galactic Nuclei, ASP, Conference Series, vol.373 ,Proceedings of the conference held 16-21,October,2006,at Xian Jiaotong University, Xian China, 2007, ASPC, 373,734
- [36] Wang Dan, Zhang Yan-Xia et al., 2007, “Survey methods for photometric redshifts”, Astronomical Data Analysis Software and Systems XVI ASP conference series, 16-24,October, 2006. Vol.376,429 2007
- [37] Wang Dan, Zhang Yanxia and Zhao Yongheng, “Support Vector Machines for Photometry Redshift Estimation From Broadband Photometry”, Data Science Journal, Vol 6,2007,S474-S480
- [38] Zhao Yongheng, Zhang Yanxia, Cui Chenzhou. ”Classification of Active Objects in the Multiwavelength Parametric Space”, 2004, Toward an International Virtual Observatory, Proceedings of the ESO/ESA/NASA/NSF Conference held in Garching, Germany, 10-14 June 2002. Edited by P.J. Quinn, and K.M. Gorski. ESO Astrophysics Symposia. Berlin: Springer, 2004, p. 333

2、待发表论文

1)、SCI索引

- [1] Wang Dan, Zhang Yan-Xia, Liu Chao, ZhaoYong-Heng. Two novel approaches for photometric redshift estimation based on SDSS and 2MASS databases. 2008, CHJAA, in press
- [2] CUI ChenZhou, SUN HuaPing, ZHAO YongHeng, LUO Yu & QI DaZhi. SkyMouse: A smart interface for astronomical on-line resources and services. 2008, Chinese Science Bulletin, in press
- [3] Zhang Yanxia, Zhao Yongheng, Gao Dan, "Decision table for separating quasars from stars", 2007, accepted by Advances in Space Research
- [4] Li Lili, Zhang Yanxia, Zhao Yongheng, "k-Nearest Neighbors for Automated Classification of Celestial Objects", 2007, accepted by Science in China Press
- [5] Zhao Yongheng, Zhang Yanxia, "Comparison of decision tree methods for finding active objects", 2007, accepted by Advances in Space Research
- [6] Zheng Hongwen, Zhang Yanxia, "Feature selection for high dimensional data in astronomy", 2007, accepted by Advances in Space Research

2)、核心期刊

- [7] Gao Dan, Zhang Yanxia, Zhao Yongheng, "The development and application of the cross-match tool of China-VO", 2006, Acta Astrophys. Sinica, accepted
- [8] Wang Dan, Zhang Yan-xia, Zhao Yong-heng. "The Methods of Estimating Photometric Redshifts". 2007, Progress in Astronomy, submitted
- [9] 刘超, 田海俊, 高丹, 杨阳, 路勇, 崔辰州, 赵永恒. "异地异构天文数据资源的统一访问". 天文研究与技术, 已接收。
- [10] Chen-Zhou CUI, Wen LI, Ce YU, Zhen XU & Yong-Heng ZHAO. "Search and Location of FITS Data Files". PNAOC, 2007, accepted
- [11] LI Chang-Hua, CUI Chen-Zhou, LI Lian & ZHAO Yong-Heng. "Resource Information Management in Chinese Virtual Observatory". PNAOC, 2007, accepted
- [12] Gao Dan, Lu Yong, Zhang Yan-Xia, Zhao Yong-Heng, "The Design and Development of Large-scale Multi-wavelength Catalogue (XMaS_VO) Federation System", 2008, PNAOC, in press

3)、国际会议

- [13] Zhang Yanxia, Zhao, Yongheng, "Feature Selection/extraction in Multiwavelength Data", 2005, accepted by PPMO
- [14] Chenzhou Cui, Yongheng Zhao. "Worldwide R&D of Virtual Observatory". Proceeding of IAU Symposium No.248: "A Giant Step:from Milli- to Micro-arcsecond Astrometry". 2008,

submitted

- [15] Zhang Yanxia, Zhao Yongheng, Gao Dan, 2008, “Discrimination of Point Sources using Radial Basis Function Network and Random Forest”, Proceeding of ADASS XVII, in press
- [16] Gao Dan, Zhang Yanxia and Zhao Yongheng, 2008, “The Application of KD-tree in Astronomy”, Proceeding of ADASS XVII, in press
- [17] Wang Dan, Zhang Yanxia and Zhao Yongheng, 2008, “Estimating Photometric redshifts of Quasars”, Proceeding of ADASS XVII, in press