

---

中国科学院研究生院  
硕士学位论文

中国虚拟天文台的数据访问与互操作研究



作者 刘波

指导教师 崔辰州副研究员 赵永恒研究员

学科专业 天文技术与方法

申请学位 硕士

培养单位 中国科学院国家天文台

二〇〇六年六月

---

**National Astronomical Observatories  
Chinese Academy of Sciences**

**Research on Data Access and Cooperation  
for Chinese Virtual Observatory**



**Bo Liu**

**Advisor: Dr.Chen-Zhou CUI**

**Prof.Yong-Heng ZHAO**

**National Astronomical Observatories, Chinese Academy of Sciences**

**Beijing 100012, P.R. China**

**June 2006**

## 摘 要

近几年随着多波段天文学的发展,单一的天文数据中心在异地多波段数据访问、数据互操作等方面需求的局限性显得尤为突出。因此,各国虚拟天文台在国际虚拟天文台联盟的倡导下,正在计划建立统一数据访问与互操作标准的天文数据结点和数据访问平台,籍此来弥补各国数据中心在这些方面的不足。针对上述情况并结合中国虚拟天文台研发需要,本研究在数据访问、数据访问门户、数据分布式调度等几个方面进行了探索。

本文介绍了国际虚拟天文台数据访问与互操作现状,论述了中国虚拟天文台进行数据访问与互操作研究的可行性和必要性。详细地讲述了中国虚拟天文台的天文数据结点、天文数据访问门户的系统设计与实现方法,并在星表数据库封装粒度问题上提出了自己的解决方案。同时针对天文数据结点的分布式特征,设计了一个分布式的调度方法,对查询请求进行分割,用它来调度用户的查询和交叉认证等服务,以此来缓解数据中心的压力,从而解决实验海量数据的批量获取问题。

最后,在“银河系旋臂结构研究”的中国虚拟天文台范例结论中证实了天文数据结点以及天文数据访问门户的可行性、可靠性与稳定性。以上几个方面的开发和研究作为中国虚拟天文台提供了可靠的数据访问平台并且为其进一步发展打下了坚实的基础。

关键字: 中国虚拟天文台 数据访问与互操作 天文数据结点  
天文数据访问门户



## Abstract

With the rapid development of multi-wavelength astronomy in the last few years, single astronomy data center can hardly meet the needs of managing and accessing huge data sets in distributed nodes and different organizations. Under the lead of International Virtual Observatory Alliance (IVOA), Virtual Observatory (VO) projects from different countries plan to set up a supportable uniform data access and interoperation specification of astronomical data node and data query platform. Aiming at resolving these problems and realizing the goals of Chinese Virtual Observatory project (China-VO), we completed much research work on data access, portal and data access distributed dispatchment. In this thesis, we introduced the current status of global data access and cooperation, discussed the importance and feasibility of data access and cooperation in the China-VO system, demonstrated design and realization of China-VO SkyNode and SkyPortal in detail, and analyzed database granularity issue, which appears in the process of integration. At the same time, we developed a distributed dispatching method to solve huge size data access problem. This method disjoined the query request, which dispatched various services such as user queries and cross-match requests, in order to reduce the pressure of data access source. In the end, we tested the possibility, credibility and stability of SkyNode and SkyPortal using a scientific case of the China-VO, Galactic spiral arm study. All the work mentioned above provide a reliable data accessing platform and strong base for the further development of the China-VO.

**KeyWord:** China-VO, Data Access and Interoperability, SkyNode, SkyPortal



## 目 录

第一章 数据访问与互操作研究概述.....	1
1.1 中国虚拟天文台概述.....	1
1.1.1 虚拟天文台概述.....	1
1.1.2 中国虚拟天文台概述.....	2
1.2 虚拟天文台的数据访问与互操作.....	4
1.2.1 虚拟天文台的数据访问与互操作的必要性.....	4
1.2.2 国际虚拟天文台的数据访问与互操作的现状.....	5
1.2.3 中国虚拟天文台的数据访问与互操作的现状.....	7
第二章 数据访问与互操作相关技术概述.....	10
2.1 网络技术.....	10
2.1.1 网络的定义.....	10
2.1.2 网格体系结构.....	10
2.1.3 OGSA 体系结构与 WEB 服务.....	11
2.1.4 Globus Toolkit 4 平台和 WSRF.....	13
2.2 中国虚拟天文台体系结构.....	15
2.2.1 中国虚拟天文台体系结构.....	15
2.2.2 中国虚拟天文台的服务模型.....	16
2.3 IVOA 的数据访问与互操作标准.....	17
2.3.1 ADQL 标准.....	17
2.3.2 VOTable 标准.....	17
2.3.2 SkyNodeInterface 标准.....	18
第三章 中国虚拟天文台天文数据节点的设计与实现.....	20
3.1 天文数据节点产生的背景.....	20
3.2 天文数据节点的体系结构.....	20
3.2.1 天文数据节点的三层结构.....	21
3.2.2 天文数据节点各层功能.....	22
3.3 天文数据节点的粒度问题及命名规则.....	24
3.3.1 天文数据节点封装的粒度.....	24
3.3.2 天文数据节点接口封装的命名规则及配置.....	25
3.4 天文数据节点各模块的功能.....	28
3.5 天文数据节点的接口.....	29
3.6 天文数据节点的运行流程.....	34
3.6.1 天文数据节点的简单查询.....	34
3.6.2 天文数据节点的复杂查询.....	35
3.7 具体技术实现与实例.....	37
3.7.1 实现工具.....	37
3.7.2 天文数据节点锥形查询实例.....	38
第四章 中国虚拟天文台天文数据访问门户的设计与实现.....	42
4.1 天文数据访问门户产生的背景.....	42
4.2 天文数据访问门户的层次结构.....	42

---

4.2.1	天文数据访问门户的体系结构.....	42
4.2.2	天文数据访问门户各层的功能.....	43
4.3	天文数据访问门户各模块的功能.....	44
4.4	天文数据访问门户的分布式调度引擎.....	45
4.4.1	分布式调度引擎的理论依据.....	45
4.4.2	分布式调度引擎的调度方法.....	45
4.4.3	分布式调度引擎调度引擎接口.....	47
4.5	天文数据访问门户的接口.....	48
4.6	天文数据访问门户运行流程.....	52
4.7	天文数据访问门户实例与总结.....	54
第五章	天文数据访问门户与天文数据结点在“银河系旋臂结构研究”的中国 虚拟天文台范例中的应用.....	58
5.1	“银河系旋臂结构研究”的中国虚拟天文台范例的目的.....	58
5.2	“银河系旋臂结构研究”的中国虚拟天文台范例操作步骤.....	58
5.3	“银河系旋臂结构研究”的中国虚拟天文台范例体系结构及其主要模块功 能.....	58
5.3.1	“银河系旋臂结构研究”的中国虚拟天文台范例的门户.....	59
5.3.2	任务引擎.....	60
5.3.3	任务描述语言.....	60
5.3.4	天文数据访问门户与天文数据结点.....	60
5.3.5	计算单元.....	61
5.4	“银河系旋臂结构研究”的中国虚拟天文台范例的应用结论.....	61
5.5	结论.....	62
第六章	总结与展望.....	64



## 图表目录

图 1.1: 数据访问服务的沙漏模型.....	4
图 1.2: NASA 的天文数据访问门户.....	6
图 1.3: NVO 的天文数据访问门户.....	6
图 1.4: JVO 的天文数据访问门户.....	7
图 2.1: 网格沙漏体系结构.....	11
图 2.2: 网格技术与 WEB 技术融合.....	12
图 2.3: OGSA 体系结构.....	12
图 2.4 GT4 主要组件体系结构.....	15
图 2.5: 中国虚拟天文台的体系结构.....	16
图 2.6: 中国虚拟天文台的服务模型.....	17
图 3.1: 天文数据结点的体系结构.....	21
图 3.2: 中国虚拟天文台体系结构.....	22
图 3.3: 天文数据结点的星表数据构造层.....	23
图 3.4: 天文数据结点数据库封装粒度级别说明.....	25
图 3.5: 2MASS 星表访问服务接口及其方法.....	26
图 3.6: 天文数据结点接口类结构图.....	34
图 3.7: 天文数据结点简单查询流程图.....	35
图 3.8: 天文数据结点复杂查询流程图.....	36
图 3.9: 天文数据结点访问服务接口.....	38
图 3.10: 天文数据结点的锥形查询结果.....	40
图 4.1: 天文数据访问门户的体系结构.....	42
图 4.3: 天文数据访问门户的查询代价分析 SQL 语句.....	46
图 4.4: 分布式调度引擎的执行计划分配流程图.....	47
图 4.5: 分布式调度引擎的内部接口.....	48
图 4.6: 天文数据访问门户接口类结构图.....	52
图 4.7: 天文数据访问门户运行流程图.....	52
图 4.8: 天文数据访问门户测试页面.....	54
图 4.9: 分布式调度引擎实例的运行图.....	55
图 4.10 交叉认证结果和显示界面.....	56
图 5.1: “银河系旋臂结构研究”的中国虚拟天文台范例体系结构.....	59
图 5.2: “银河系旋臂结构研究”的中国虚拟天文台范例的门户.....	59
图 5.3: “银河系旋臂结构研究”的中国虚拟天文台范例的 SQL 查询语句.....	60
图 5.4: “银河系旋臂结构研究”的中国虚拟天文台范例的银河系旋臂模型.....	61
表 1.1: 国际各国虚拟天文台项目.....	2
表 2.1: GT4 中的新协议.....	14
表 3.1: 天文数据结点元数据配置文件.....	28
表 3.2: Cone Search 的输入参数.....	29
表 3.3: table()方法输入输出参数表.....	29
表 3.4: tables()方法输入输出参数表.....	30

---

表 3.5: column()方法输入输出参数表 .....	30
表 3.6: columns()方法输入输出参数表 .....	31
表 3.7: performQuery()方法输入输出参数表 .....	31
表 3.8: executePlan()方法输入输出参数表 .....	32
表 3.9: queryCost()方法输入输出参数表 .....	32
表 3.10: getAvailability()方法输入输出参数表 .....	33
表 3.11: functions()方法输入输出参数表 .....	33
表 3.12: 用户提交的 ADQL 锥形查询语句 .....	39
表 3.13: 经过 ADQL 模块转换后的 ADQL/x 锥形查询 .....	39
表 4.1: getSkyNodeTableInfo()方法输入输出参数表 .....	49
表 4.2: getSkyNodeTables()方法输入输出参数表 .....	49
表 4.3: getSkyNodeColumnInfo()方法输入输出参数表 .....	50
表 4.4: getSkyNodeColumns()方法输入输出参数表 .....	50
表 4.5: SubmitQuery()方法输入输出参数表 .....	51
表 4.6: SubmitDistributedQuery()方法输入输出参数表 .....	51
表 4.7: getSkyNodeAvailability()方法输入输出参数表 .....	51

## 第一章 数据访问与互操作研究概述

### 1.1 中国虚拟天文台概述

#### 1.1.1 虚拟天文台概述

传统天文学经过了几十年的观测积累，加上近几年新型的天文仪器不断投入运行，产出的观测数据已成指数增长。天文学家面对如此巨大的数据量，欣喜的同时也在为如何有效地处理和分析这些数据而感到忧虑。如何从这包含数十亿天体的多波段海量数据中探求科学发现，已成为天文学家亟待解决的问题。与此同时信息科学领域也在经历着快速的变革，伴随高性能计算和互联网技术的发展，网格计算应运而生。网格计算<sup>[1]</sup>是指通过将多台计算机组成网格状网络，“模拟实现高性能计算机”的技术，实现互联网上所有资源的互联互通，包括计算资源、存储资源、通信资源、软件资源、知识资源等。

近些年来伴随着网格技术、XML、WEB 服务等新兴 IT 技术日趋成熟，虚拟天文台应运而生。虚拟天文台成为基于网格技术解决海量复杂天文数据访问和管理的一种有效手段。虚拟天文台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<sup>[2]</sup>。它将 TB 甚至 PB 量级的数据、波长遍及从  $\gamma$  射线到射电波段的数十亿个天体的图像库、高度复杂的数据挖掘工具和统计分析工具、具有数千 PB 量级容量的存储设备和每秒运算次数达到万亿次的超级计算设备、以及各主要天文数据中心之间的高速网络连成一体；它能使世界各地的天文学家可以快速查询每个 PB 量级的数据库；使埋藏在庞大星表和图像数据库中的多变量模式可视化；增加了发现复杂规律和稀有天体的机会；可以进行大样本星表的统计研究及数据挖掘。

虚拟天文台的概念提出后各国相继出台了各自的虚拟天文台计划。为了相互借鉴、相互促进以及国际上各项目之间的协调与合作，2002 年成立了国际虚拟天文台联盟（International Virtual Observation Alliance, IVOA）

[3], 目前国际虚拟天文台联盟中成员已经增加到 16 个国家。如表 1.1 所示:

项目	地区	网址
中国虚拟天文台(China-VO)	中国	<a href="http://www.china-vo.org/">http://www.china-vo.org/</a>
国家虚拟天文台(NVO)	美国	<a href="http://www.us-vo.org/">http://www.us-vo.org/</a>
英国虚拟天文台(AstroGrid)	英国	<a href="http://www.astrogrid.org/">http://www.astrogrid.org/</a>
欧洲虚拟天文台(Euro-VO)	欧洲	<a href="http://www.euro-vo.org/">http://www.euro-vo.org/</a>
法国虚拟天文台(VO-France)	法国	<a href="http://www.france-vo.org/">http://www.france-vo.org/</a>
加拿大虚拟天文台(CVO)	加拿大	<a href="http://services.cadc-ccda.hia-ihp.nrc-cnrc.gc.ca/cvo/">http://services.cadc-ccda.hia-ihp.nrc-cnrc.gc.ca/cvo/</a>
印度虚拟天文台(VO-India)	印度	<a href="http://vo.iucaa.ernet.in/%7Evoi/">http://vo.iucaa.ernet.in/%7Evoi/</a>
德国虚拟天文台(GAVO)	德国	<a href="http://www.g-vo.org/portal/">http://www.g-vo.org/portal/</a>
澳大利亚虚拟天文台(Aus-VO)	澳大利亚	<a href="http://www.aus-vo.org/">http://www.aus-vo.org/</a>
日本虚拟天文台(JVO)	日本	<a href="http://jvo.nao.ac.jp/">http://jvo.nao.ac.jp/</a>
俄罗斯虚拟天文台(RVO)	俄罗斯	<a href="http://www.inasan.rssi.ru/eng/rvo/">http://www.inasan.rssi.ru/eng/rvo/</a>
意大利虚拟天文台(DRACO)	意大利	<a href="http://www.as.oat.ts.astro.it/wiki/bin/view/Draco/WebHome">http://www.as.oat.ts.astro.it/wiki/bin/view/Draco/W ebHome</a>
韩国虚拟天文台(KVO)	韩国	<a href="http://kvo.kao.re.kr/">http://kvo.kao.re.kr/</a>
匈牙利虚拟天文台(HVO)	匈牙利	<a href="http://hvo.elte.hu/en/">http://hvo.elte.hu/en/</a>
西班牙虚拟天文台(SVO)	西班牙	<a href="http://laeff.esa.es/svo/">http://laeff.esa.es/svo/</a>
亚美尼亚虚拟天文台(ArVO)	亚美尼亚	<a href="http://www.aras.am/arvo.htm">http://www.aras.am/arvo.htm</a>

表 1.1: 国际各国虚拟天文台项目

### 1.1.2 中国虚拟天文台概述

为了紧跟时代的发展, 缩小与发达国家的信息天文学方向的差距, 2002 年以中国科学院国家天文台为首在国内天文界首先提出建设“中国虚拟天文台<sup>[4]</sup>”项目。同年十月, 中国虚拟天文台加入国际虚拟天文台联盟。经过近几年的研究探索, 中国虚拟天文台目前的研究主要在以下几个领域深入开展:

- 1) 虚拟天文台系统中数据与服务的动态注册和发现机制

- 2) 国内外天文数据资源的无缝透明访问与互操作研究
- 3) 现有天文软件资源的 VO 集成
- 4) 天文仪器的 VO 集成
- 5) 数据挖掘平台的研究
- 6) 中国虚拟天文台基础上的天文科普教育

中国虚拟天文台是相互之间能进行互操作的数据集以及分析工具的集合，它通过互联网提供给天文学家一个可在线进行天文科学研究的平台。它的科学目的与意义是：

- 1) 中国虚拟天文台将采用与我国现在正在建造的大天区面积多目标光纤光谱望远镜<sup>[5]</sup>(LAMOST)项目紧密结合的方式,充分发挥LAMOST光学光谱数据中心的作用,建设“VO-oriented LAMOST”,使光谱数据及其相关处理技术成为中国虚拟天文台的核心与特色。
- 2) 中国虚拟天文台将利用先进的网格和WEB服务技术,方便、快捷、实时地把今后LAMOST产出的数据发布,以便能和国外的数据进行交互与融合。
- 3) 中国虚拟天文台将最大程度地集中国内各天文研究机构、IT 研究机构和数学等相关领域的人才资源,共同努力实现目标。
- 4) 中国虚拟天文台将采用开放的运作方式,与国际上各虚拟天文台计划开展充分的合作,在IVOA 中发挥积极作用。
- 5) 中国虚拟天文台将利用IVO丰富的资源加强教育和科学普及工作,提高我国公众的科学素质。

作为 e-Science 时代中国天文学研究的新平台和“网关”,中国虚拟天文台把国内的天文数据资源与国际共享,并把国际的资源输入到国内,逐步实现与国际各天文数据中心数据的互操作和透明访问。最终,中国虚拟天文台的用户将可以通过使用互联网的方式方便的获取国际范围内的天文数据,可以对分布在各地的数据进行大规模多波段的分析和挖掘,并利用中国虚拟天文台的可视化系统将结果以 2D、3D 图像的形式显示出来。作为国际虚拟天文台的成员,中国虚拟天文台正在与国际上各个虚拟天文台进行紧密合作,在 IVOA 中发挥积极作用。

## 1.2 虚拟天文台的数据访问与互操作

### 1.2.1 虚拟天文台的数据访问与互操作的必要性

在各国天文数据中心里存储了大量的天文数据资源,由于各国自身条件不同,这些数据大部分都存放在符合自己条件的数据库管理系统中,这些存储数据的资源相互孤立,存储方式、命名方式、单位、元数据标准、访问方式等等也有所不同,这就导致了各国数据中心数据的异构性和复杂性。如果将这些海量数据集中在一个地点存储,同时让世界所有天文台观测出的数据均使用同种的数据库管理系统来存储显然也是不现实的,为了使天文学家和计算机程序能方便的访问和处理这些异地异构的分布式数据,以便天文学家在这些数据的基础上进行多波段的数据分析与挖掘,这就必须建立一套国际统一的天文数据访问与互操作标准。

虚拟天文台的一个基本功能就是定义全球统一的互操作标准和分布式数据的访问机制,以此来屏蔽数据源在数据格式、存储格式、主机环境、访问形式等诸多方面的异构性和复杂性。数据访问与互操作的实现在虚拟天文台中体现为一系列的服务。这些服务存在于数据源和上层应用之间,为上层应用服务提供统一的数据访问服务,同时进行数据格式的转换,以此来增强数据的互操作能力。其原理好比沙漏<sup>[2]</sup>,如图 1.1 所示:

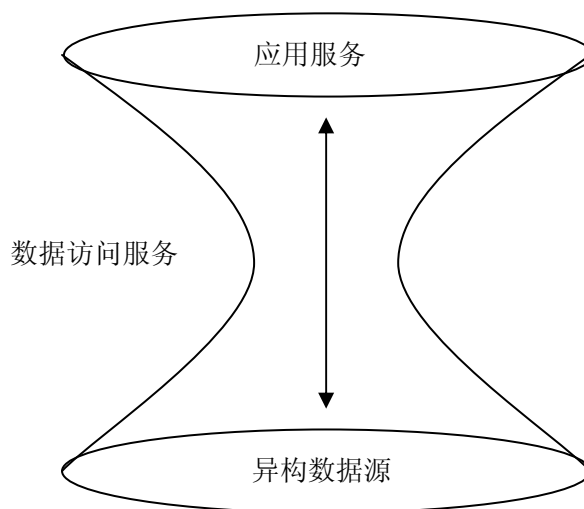


图 1.1 沙漏模型

这个访问机制不仅仅提供给天文学家获取数据和分析结果的便利,它还提供给用户的应用程序,使得程序也可以通过统一的访问接口来查询和获取数据,从而实现全球范围内主要天文研究资源,不同数据集、不同数据服务间的互操作,特别是异地数据资源统一、透明地访问与获取。因此,数据访问与互操作问题就成为了虚拟天文台的首要解决问题之一。

### 1.2.2 国际虚拟天文台的数据访问与互操作的现状

国际上几个大型的虚拟天文台项目开展了数据访问与互操作方向的研究工作主要集中在虚拟天文台查询语言 (Virtual Observatory Query Language, VOQL) 和天文数据结点接口 (SkyNodeInterface) 标准上 (这两个标准仅定义了建立天文数据结点所需要实现的接口和功能,但并没有给出具体的实现方式)。各国相继建设了符合 IVOA 标准的天文数据结点和天文数据访问门户,对外发布为全球可操作的 WEB 服务接口。但在具体实现方式上和所用数据查询语言上却各有特色。

美国宇航局 (National Aeronautics and Space Administration, NASA) 的 SkyQuery-SkyNode<sup>[6]</sup>: 是 NASA 资助建立的符合 IVOA 的 SkyNodeInterface 标准的天文数据结点。他们使用的查询语句是 SQL92 的一个子集再加上其特有的天文查询关键字组成的 SkyQL 查询语句,用来对天文数据结点进行访问,这种语句并没有完全按照 ADQL 的标准来设计。其天文数据结点中封装了美国多家天文研究机构的数据中心,底层数据库均为 SQL Server 2000,对外发布 WEB 服务接口。由于其数据数据库管理系统统一,因此它的运行效率较高并且不存在异构性和数据互操作的障碍。它的缺点是必须让用户使用同它完全相同的数据库管理系统和命名方式时才可以与之进行数据互操作。其天文数据查询门户如图 1.2 所示。



图 1.2 NASA 的天文数据访问门户

美国虚拟天文台（NVO）的 OpenSkyQuery-SkyNode<sup>[7]</sup>：是由美国虚拟天文台设计并建设的天文数据结点。它使用的 ADQL 是完全符合 VOQL 标准的查询语言。而 SkyNodeInterface 标准就是由美国虚拟天文台提出的。NVO 的 SkyNode 中封装了美国多家天文研究机构的数据中心，对外提供符合标准的、统一的 WEB 服务接口。它是标准的天文数据结点的原型。NVO 的天文数据访问门户如图 1.3 所示

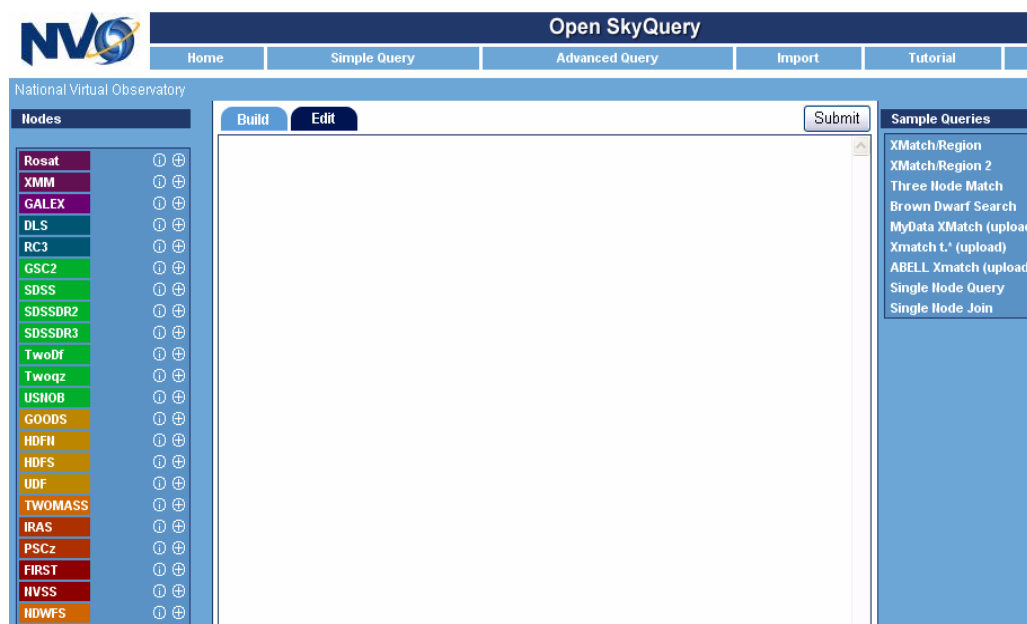


图 1.3 NVO 的天文数据访问门户



日本虚拟天文台（JVO）的 JVO-SkyNode<sup>[8]</sup>：所使用的查询语句是由符合 VOQL 的一个子集的 JVOQL 查询语句加上他们自定义的关键字组成。它的天文数据结点中封装了几家日本大型天文研究机构的星表数据中心，底层数据库为 PostgreSQL，同时也支持 SQL Server 2000 等其它类型的数据库管理系统的访问与操作。日本虚拟天文台数据访问门户如图 1.4 所示。

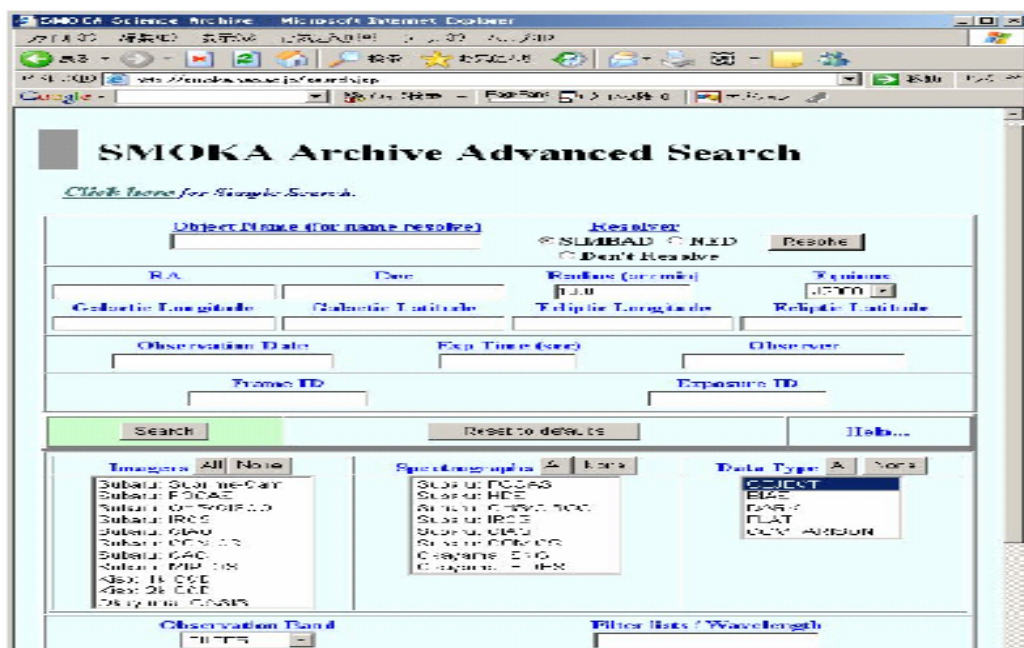


图 1.4 JVO 的天文数据访问门户

上述研究机构所设计的天文数据结点均符合 IVOA 的数据访问与互操作标准。它们的前台为网格服务或 Web 服务，可以使用标准 Java 服务发布工具和 Servlet 来实现。VOQL 和 SkyNodeInterface 标准就是融合了上述三家机构所设计出的天文数据结点后定义出来的。因此我们如果想要开展数据访问与互操作的研究，使我们的数据中心可以和国外的数据中心进行数据互操作，获取外国数据中心的资源，那么我们必须满足 IVOA 所推出的规范与标准，借鉴他们的研发经验。

### 1.2.3 中国虚拟天文台的数据访问与互操作的现状

作为未来中国天文学研究的基础平台，中国虚拟天文台系统实现的基础是提供给用户统一的数据访问资源。数据访问服务与互操作的研究直接关

系着整个中国虚拟天文台的成败，是当前阶段的工作重心。

中国科学院国家天文台经过了几年多的建设与设计、实施，已经成功的运行并发布了属于自己的数据中心（BADC），并且在数据中心中也做了很多国际大型星表数据库的镜像。但数据资源相对分散，数据存储结构、数据的定义方式、所使用的数据库管理系统、数据访问形式等等均和国际上的数据中心有差异。当然这种差异由于自身原因也是在所难免的。正因为这些差异，我国一直没有可供天文学家获取、处理和分析国际天文数据的平台，尤其是异地、海量多波段的天文数据分析平台。这已经对那些多波段研究领域的天文学家造成了很大的障碍，成为我国多波段天文学发展的瓶颈。

在中国虚拟天文台项目为了实现这个需求，需要在设计上符合 IVOA 的标准，在实现功能上又发挥出自己的特色。主要涉及的工作包括：

- 1) 中国虚拟天文台数据访问与互操作平台的设计与建立——天文数据访问结点（SkyNode）。
- 2) 中国虚拟天文台数据访问平台的门户设计与实现——天文数据访问门户（SkyPortal）。
- 3) 优化天文数据结点和天文数据访问门户，提高运行的效率，保证运行的稳定性。
- 4) 中国虚拟天文台数据互操作标准的设计与融合。

资源服务化是 China-VO 要实现的目标之一，这些资源包括数据资源、软件资源和应用服务资源，而海量数据资源的存储、访问与互操作则是中国虚拟天文台应用资源服务化实现的基础。传统的天文数据、巡天数据、个人观测和数值模拟数据对虚拟天文台都是非常重要的资源，只有将这些数据资源服务化、封装成为国际统一的数据标准格式，才能促进国际间的数据互访与互操作，从而融合这些数据资源、建立资源汇集平台，进一步开展数据挖掘和多波段天文数据研究。

## 参考文献

- [1] 都志辉,陈渝,刘鹏.网格计算.北京:清华大学出版社,2002
- [2] 崔辰州. 中国虚拟天文台系统设计: [学位论文].北京:中国科学院研究生院
- [3] [IVOA] <http://www.ivoa.net/>.
- [4] [China-VO] <http://www.china-vo.org/>.
- [5] [LAMOST] <http://www.lamost.org>
- [6] [SkyQuery] <http://www.skyquery.net/>
- [7] [OpenSkyQuery] <http://www.openskyquery.net/>
- [8] [SkyCom] <http://www.skycom.jp/>

## 第二章 数据访问与互操作相关技术概述

### 2.1 网格技术

#### 2.1.1 网格的定义

网格<sup>[1]</sup>概念最早于上世纪 90 年代中期提出，用于表述在高端科学和工程上分布式计算的一种基础构造形式。简单地讲，网格是把整个网络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。这样组织起来的“虚拟的超级计算机”有两大优势，一是数据处理能力超强；二是能充分利用网上的闲置处理能力。网格就是将收集网络上所有可用的分布式计算资源提供给最终用户或组建成一个巨大的计算系统。

潜藏在网格概念之中确切而特殊的问题就在于，实现对等的资源共享和解决动态的、分布式的虚拟组织所遇到的问题<sup>[2]</sup>。我们所关心的共享不仅仅是简单的文件交换，更强调直接对计算机、软件、数据以及其他资源的直接访问。共享联系能够开始于任何小组动态地处理新的节点，跨越不同的平台、语言和编程环境，显而易见，互操作性成为不可或缺的部分。在一个统一的网格环境下，互操作性意味着共同的协议。因此，在网格体系中最重要的是一个协议体系，该协议定义了基础机制，虚拟组织用户们通过这个机制来商议、建立、管理和开发分配关系。

#### 2.1.2 网格体系结构

网格的体系结构<sup>[3]</sup>是开放的、可扩张的体系结构。体系结构为整个系统提供了一个结构、行为和属性的高级抽象，由对构成系统的元素的描述、元素间的相互作用、元素集成的模式以及这些模式的约束组成。体系结构不仅指定了系统的组织结构和拓扑结构，并且显示了系统需求和构成系统元素之间的相互关系，提供了一些设计决策的基本原理。

图 2.1 为网络的沙漏体系结构图，将功能分散在不同的层次上。从底层到上层依次是：构造层、连接层、资源层、汇集层、应用层。

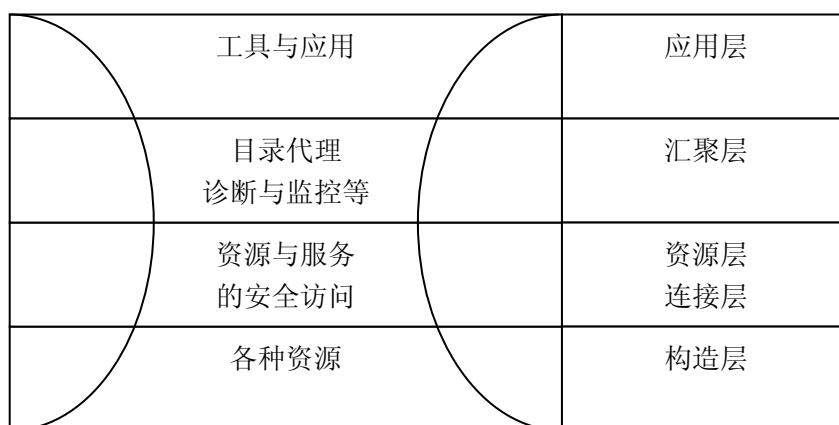


图 2.1：网格沙漏体系结构

- 1) 构造层：网格构造层是本地控制的接口，提供了共享的资源，执行本地与资源相关的操作。
- 2) 连接层：连接层定义了核心的网络事务处理所需要的通信和认证协议，以支持便利安全的通信。
- 3) 资源层：资源层建立在连接层的通信和认证协议之上，定义的协议包括安全的连接、初始化、监视与控制、计费等。
- 4) 汇集层：汇集层协调各种资源。如果说资源层着眼于和单个资源进行交互，那么汇集层就着眼于协调多种资源的共享。
- 5) 应用层：应用层是在虚拟组织环境中存在的，它是用户需求的具体体现。

### 2.1.3 OGSA 体系结构与 WEB 服务

开放网格服务架构(Open Grid Service Architecture, OGSA)<sup>[4]</sup>是在网络协议和 Web 服务技术融合的基础上提出的一套规范和标准。它最突出的思想就是以“服务”为中心。在 OGSA 框架中，将一切都抽象为服务，包括计算机、程序、数据资源、仪器设备等。这种观念有利于通过统一的标准接口来管理和使用网格。

OGSA 平台是旨在为各种网格系统所共同面临的基本问题定义标准的解决方案和机制。这些基本问题包括网格服务间的通信、身份确立、授权对话、服务发现、错误通告、服务集管理等。

随着近些年来 WEB 服务技术在商业上应用的不断扩大、功能不断完善，网格技术也借鉴了其优点，将 WEB 服务技术融入网格体系架构之中，弥补网格技术在使用现有资源上的不足。WSRF 的定义意味着网格和 Web 服务团体在一个共同的基础上前进。网格技术与 WEB 技术融合图例如图 2.2 所示。

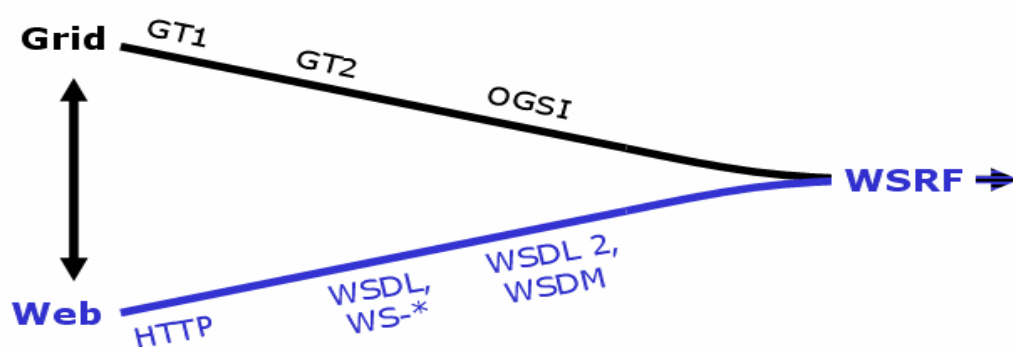


图 2.2 网格技术与 WEB 技术融合

OGSA 平台包括三个基本元素：OGSA 平台接口、OGSA 平台模型和 WEB 服务资源框架<sup>[5]</sup>（Web Service Resource Framework, WSRF）。OGSA 体系结构如图 2.3 所示。

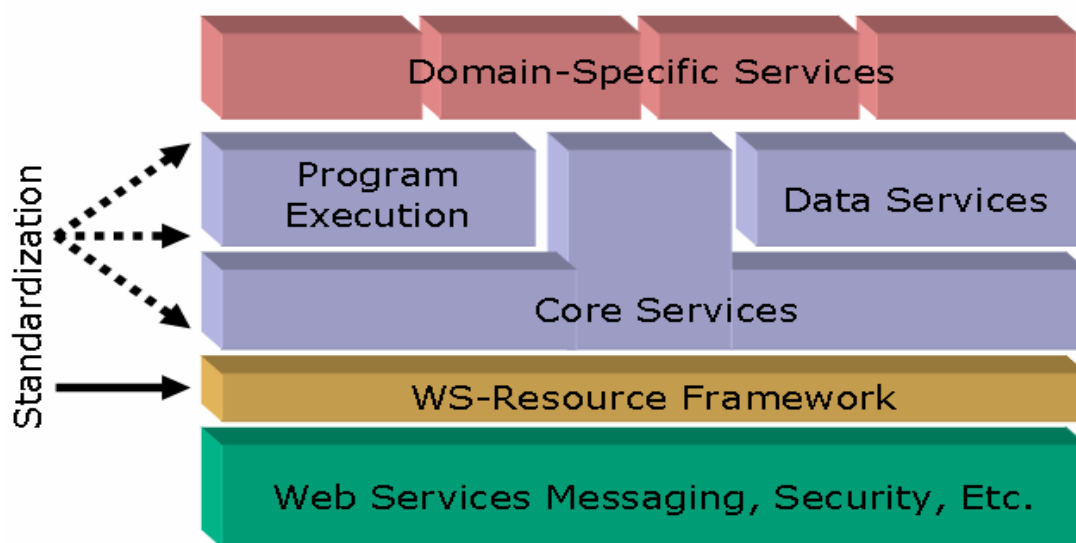


图 2.3 OGSA 体系结构

### 2.1.4 Globus Toolkit 4 平台和 WSRF

Globus 项目<sup>[6]</sup>是目前国际上最有影响的网格计算项目之一。Globus 对信息安全、资源管理、信息服务、数据管理以及应用开发环境等网格计算的关键理论和技术进行了广泛的研究,开发出能在多种平台上运行的网格计算工具包软件(Globus Toolkit),能够用来帮助规划和组建大型的网格试验和应用平台,开发适合大型网格系统运行的大型应用程序。Globus 项目组发布了 Globus Toolkit 4(GT4)最新版本。GT4 体系结构做了较大的变动,由开放网格服务基础架构(OGSI)迁移到了 Web 服务资源框架(WSRF),WSRF 是一套规范的集合,重新定义了网格服务的组成,提高了与 WEB 服务的兼容性。

WS-Resource 被定义为由 Web 服务和有状态的资源构成的实体。有状态的资源可以在 Web 服务消息交换中使用。可以创建和销毁 WS-Resources,而且可以通过消息交换查询或更改其状态。WS-Resource 有四个很重要的称为 ACID 特性的软件工程特性。在 Web Services Atomic Transaction 规范 [WS-AtomicTransaction] 中,对这些特性进行了介绍,如下所示:

- 1) 原子性 (Atomicity): 在事务单元中,要么对有状态资源全部进行更新,要么全都不进行更新。
- 2) 一致性 (Consistency): 有状态资源应该总能维护一致状态,即使失效之后也是如此。
- 3) 隔离性 (Isolation): 应该在给定的事务单元中隔离对有状态资源的更新。
- 4) 持久性 (Durability): 在事务单元中对有状态资源的更新是永久性的。

WSRF 的提出是为了避免将来两者不兼容,同时可以利用大量的现成 Web 服务的开发工具、已有成果;同时减少了太多的面向对象机制的引入,提高了系统性能。下表 2.1 总结了 Globus Toolkit 4 主要协议的一些基本特性和兼容性问题。

服务	协议	特性	向后兼容性
数据传输	可靠文件传输(RFT)	<ol style="list-style-type: none"> <li>1. 使用 GridFTP 控制和监视第三方文件传输。</li> <li>2. 指数补偿 (Exponential back-off)。</li> <li>3. 全部传输或者全部不传输。</li> <li>4. 并行传输。</li> <li>5. TCP 缓冲区大小。</li> <li>6. 递归目录传输。</li> </ol>	与 OGSi (GT3.2) 不存在向后兼容
资源管理	WS-GRAM	<ol style="list-style-type: none"> <li>1. 改进任务性能：并行性、吞吐量和等待时间等。</li> <li>2. 改进可靠性/可恢复性。</li> <li>3. 支持 mpich-g2 任务，包括： <ol style="list-style-type: none"> <li>1. 多任务提交。</li> <li>2. 在一个任务中协调处理。</li> <li>3. 在一个多任务的子任务之间进行协调。</li> </ol> </li> </ol>	该协议已经被修改成 WSRF 兼容的。该版本与以前的版本之间不存在向后兼容
信息服务	MDS4	<p><b>索引服务</b></p> <ol style="list-style-type: none"> <li>1. 基于 WSRF 而不是 OGSi。</li> <li>2. 已经删除 Xindice 支持。</li> <li>3. 已经重构聚合的永久性配置。</li> </ol> <p><b>全新的服务</b></p> <ol style="list-style-type: none"> <li>1. 触发器服务。</li> <li>2. 聚合器 (Aggregator)。</li> <li>3. 归档服务。</li> </ol>	与 GT3.2 的索引服务不兼容，因为该服务已经使用 WSRF 代替 OGSi，重新进行了建模

表 2.1 GT4 中的新协议

GT4 核心服务提供了程序开发支持的功能，包括开放式开发模式和访问网络服务的实现，像网格资源管理(Grid Resource Allocation Management, GRAM)。应用 GT4 强有力的理由就是因为它是建立在现有的 WEB 服务标准和技术的基礎上，像 SOAP 和 WSDL。网格服务提供的接口都是通过 WSDL 来描述的。GT4 提供了一个软件仓库，像安全支持，软件的探索，软件的资源管理，软件的调用，软件之间的通信，异常处理和数据管理等。图 2.4 主要描述了在服务器端的 GT4 中的主要组件体系结构。



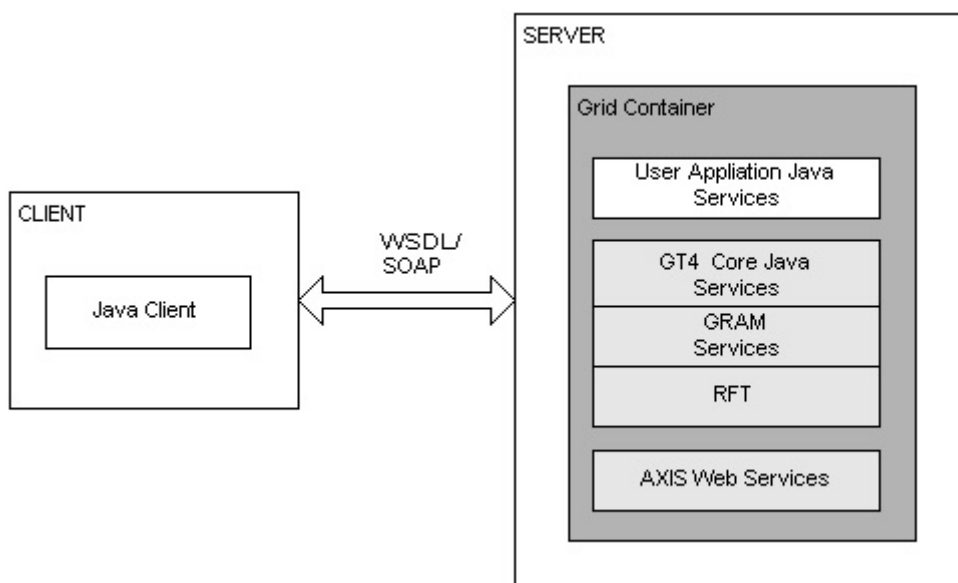


图 2.4 GT4 主要组件体系结构

GT4 结构由一个网格容器组成，网格容器主要用来管理所有部署的 WEB 服务，贯穿于每个 WEB 服务的运行周期。GT4 使用 apache 的 axis 作为它的 WEB 服务的引擎来处理所有的 SOAP 消息，JAX-RPC (Java API for XML-Based RPC)处理和 WEB 服务的配置。

## 2.2 中国虚拟天文台体系结构

### 2.2.1 中国虚拟天文台体系结构

根据中国虚拟天文台的科学目标以及采用的网格技术，从基本的功能模块角度来分析，中国虚拟天文台整个体系结构<sup>[7]</sup>分为四层：用户层、汇集层、资源层、构造层(如图 2.5 所示)。

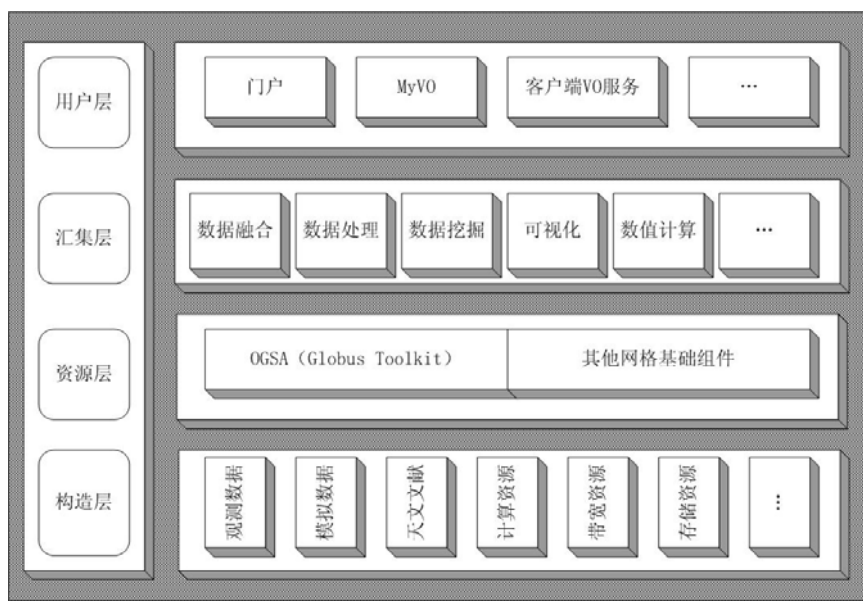


图 2.5: 中国虚拟天文台的体系结构

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、数据资源、存储资源等；资源层以 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；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；用户层包括客户端服务和门户，它直接与虚拟天文台用户接触，用于任务提交和处理结果返回。

我们所研究的数据访问与互操作问题，贯穿整个中国虚拟天文台体系结构，向下封装构造层的星表数据资源，在虚拟天文台的汇聚层提供天文数据资源汇聚接口，向上给用户提供一个方便直观的数据访问门户。

### 2.2.2 中国虚拟天文台的服务模型

从图 2.5 看出，资源层以开放网格服务架构 OGSA 为基础，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。概括起来，VO 系统中需要用到的服务或功能主要有以下一些：数据访问、文件访问、数据融合、数据迁移、数据挖掘、计算服务、可视化服务、数据转换、注册与发现、元数据服务等。中国虚拟天文台服务模型如

图 2.6 所示。

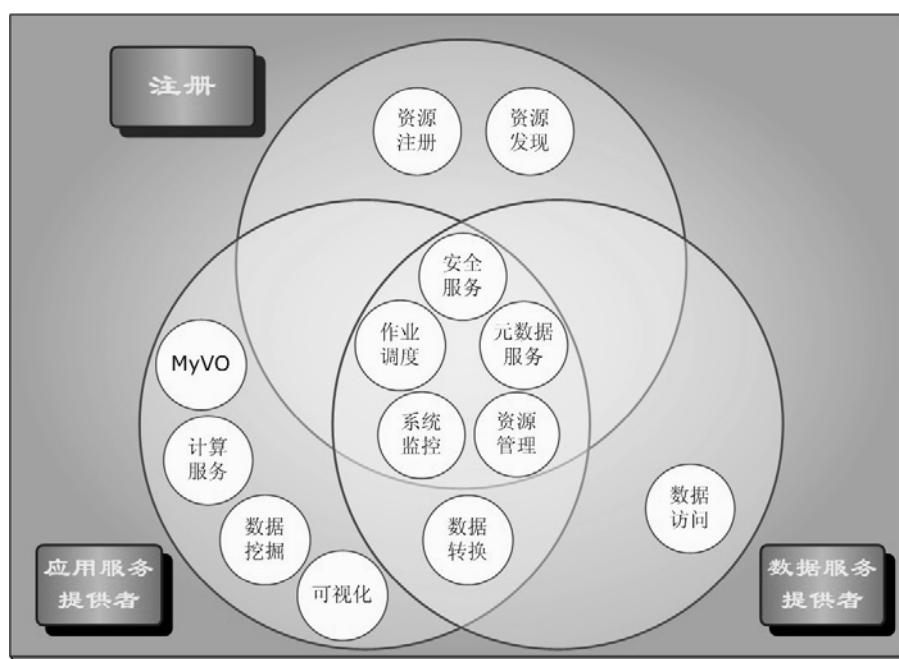


图 2.6: 中国虚拟天文台的服务模型

## 2.3 IVOA 的数据访问与互操作标准

### 2.3.1 ADQL 标准

天文数据查询语言(Astronomy Data Query Language, ADQL)<sup>[8]</sup>是一个 XML 格式的天文数据查询语言,它基于 SQL 子集并且加上 region 作为最小的支持,这个 region 实现了锥形检索(Cone Search)的功能(即 ADQL-circle)。天文数据访问门户或客户端程序发送 XML 文档形式的 ADQL 查询语句到天文数据结点的查询接口上,由其内部的 ADQL 解析器来将该查询语句转换为关系型数据库可识别的 SQL 标准语句。

### 2.3.2 VOTable 标准

VOTable<sup>[9]</sup>是 XML 格式的,用来存储和传输天文星表数据及其元数据的标准。在 VOTable 标准中,提供了对星表元数据的描述。元数据包括对星表自身的描述和对星表每一列的描述,如表名、使用的坐标框架、数据

类型、单位等等。同时它还引入了统一内容描述(Unified Content Descriptors, UCD), 使程序可以理解每一列的具体天文含义。

### 2.3.2 SkyNodeInterface 标准

SkyNodeInterface<sup>[10]</sup>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。

## 参考文献

- [1] Miguel L. Bote-Lorenzo, Yannis A. Dimitriadis.  
Grid Characteristics and Uses: a Grid Definition.  
[http://www.chinagrid.net/dvnews/upload/2005\\_03/05031923425425.pdf](http://www.chinagrid.net/dvnews/upload/2005_03/05031923425425.pdf)
- [2] Ian Foster Carl Kesselman Steven Tuecke. The Anatomy of the Grid.  
<http://www.globus.org/research/papers/anatomy.pdf>
- [3] 华宇. 浅析基于 WSRF 的网格体系结构.  
<http://www.chinagrid.net/grid/paperppt/GridArchi.ppt>
- [4] 肖侗. 网格体系结构之 OGSA.  
<http://www.chinagrid.net/grid/paperppt/bigfile/paperppt/XiaoN/Arch2.ppt>
- [5] 肖侗 网格体系结构之 WSRF  
<http://www.chinagrid.net/grid/paperppt/bigfile/paperppt/XiaoN/Arch3.ppt>
- [6] 陈渝. 网格平台 Globus 的核心技术.  
[http://www.chinagrid.net/dvnews/upload/2005\\_02/05022813117467.doc](http://www.chinagrid.net/dvnews/upload/2005_02/05022813117467.doc)
- [7] 崔辰州. 中国虚拟天文台系统设计: [学位论文].北京: 中国科学院研究生院
- [8] [ADQL] <http://www.ivoa.net/internal/IVOA/IvoaVOQL/ADQL-0.91.pdf>
- [9] [VOTABLE] <http://www.us-vo.org/VOTable/>
- [10] [SKYNODE]  
<http://www.ivoa.net/internal/IVOA/IvoaVOQL/SkyNodeInterface-0.9.pdf>

## 第三章 中国虚拟天文台天文数据结点的设计与实现

### 3.1 天文数据结点产生的背景

国际上提供给用户天文数据访问功能的数据提供中心有很多，我们经常使用的有：SDSS<sup>[1]</sup>、CDS<sup>[2]</sup>、Aladin<sup>[3]</sup>等。经过多年的积累这些数据中心相继开发了多种应用功能，它们在星表获取、天体定位、小星表交叉证认等方面功能都相当成熟了，而且查询方式也多样化，对于一般的天文学家来说它们的功能已经足以进行基本的天文研究工作了。

这些数据中心大都采用页面绑定后台数据库的传统方式来实现，但这种方式往往局限于数据中心自身所发布的数据，其他观测站的数据则需要通过邮寄光盘、网络下载、星表格式转换等方式将数据导入其底层的数据库中。这就导致了单一数据中心数据存储无限制扩大并且数据访问速度也随之变慢。这些天文数据中心往往采用在世界各地分布多个镜像的方法以缓解访问压力，提高工作效率，同时也采用查询限制的办法回避这个问题。

随着近几年多波段数据研究的发展，这些数据中心在异地多波段数据访问、数据融合、数据互操作等方面的技术缺陷显得尤为突出。因此各国在 IVOA 的组织下，纷纷建立了符合 IVOA 数据访问与互操作标准的天文数据结点来弥补数据中心在异地异构海量数据处理上的不足。

中国虚拟天文台利用新兴 IT 技术的同时严格遵循 IVOA 的数据访问与互操作标准，对现有体系架构的底层进行设计与实现，完成了其核心部分——中国虚拟天文台的天文数据结点（以下简称天文数据结点：SkyNode），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。本章详细描述了我国虚拟天文台天文数据结点的设计、发布、命名规则、功能、运行流程及其具体实现方法。

### 3.2 天文数据结点的体系结构

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。

### 3.2.1 天文数据结点的三层结构

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 3.1 所示。

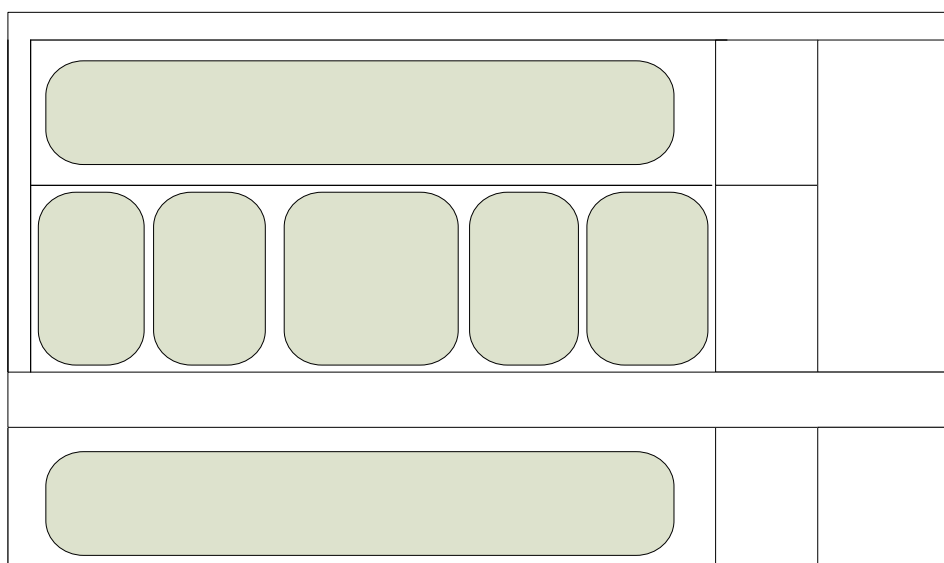
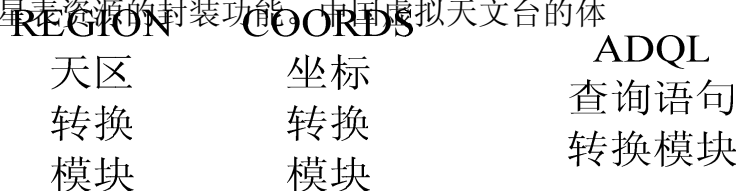


图 3.1 天文数据结点的体系结构

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，为上层的数据分析、数据挖掘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 SKYNODE 模块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。中国虚拟天文台的体系结构图<sup>[4]</sup>如图 3.2 所示。



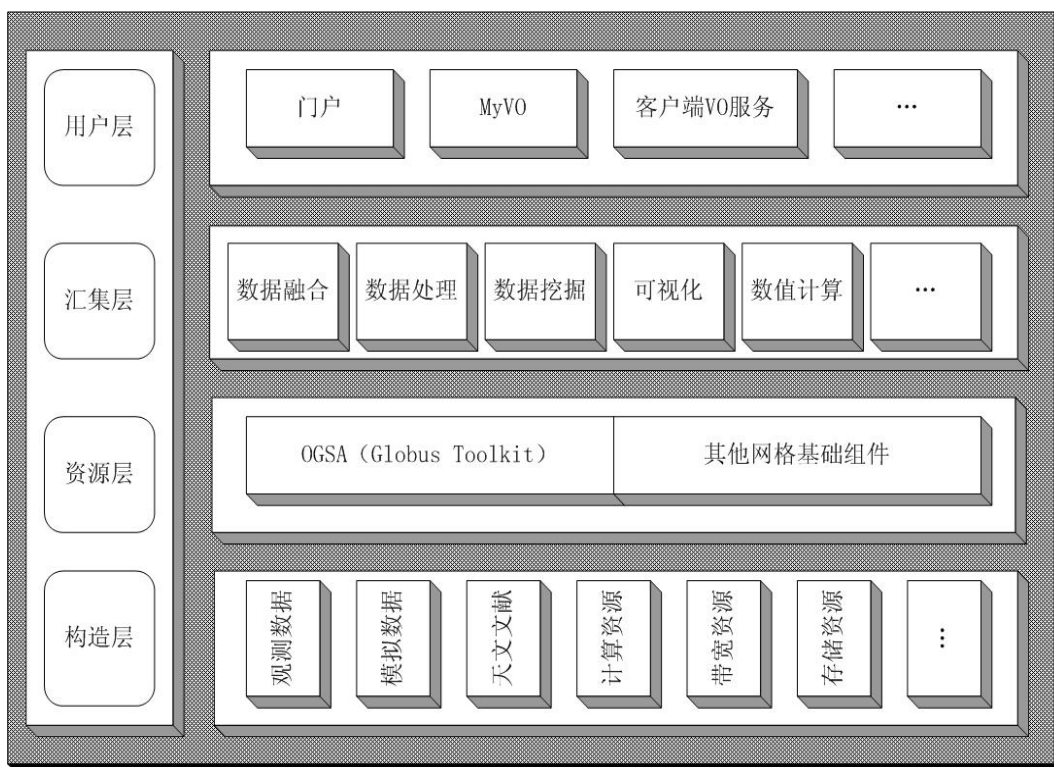


图 3.2 中国虚拟天文台体系结构

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。

### 3.2.2 天文数据结点各层功能

天文数据结点三个层次的功能如下：

- 1) **服务接口层**：主要是为中国虚拟天文台的上层应用服务（如数据挖掘服务、天文数据访问门户等）提供调用底层异地、异构天文数据资源的服务接口。数据资源服务接口设计的粒度为每个星表数据库的每个表作为一个单元，并对每个单元进行网格服务或 Web 服务接口的封装，使之成为原子化的访问接口，因此具有很好的互操作性。关于粒度问题 3.3 节中进行了详细的讨论。
- 2) **查询处理层**：主要是由一系列查询处理、解析、格式转换、存档模块组成。查询处理模块是用来处理中国虚拟天文台的上层应用服务



传递下来的简单的 ADQL 查询、锥形查询及带有多表交叉认证的复杂查询。查询结果最终由 VOTABLE 转换模块输出到客户端程序、数据挖掘平台或天文数据访问门户。解析模块、格式转换模块、存档输出模块则是用来解决异地异构数据库之间的星表数据相互之间存储方式、命名方式、单位、元数据标准定义不能互相转换、格式不匹配等一系列的互操作问题。

- 3) 星表数据构造层：是由异地异构星表数据库和少量 FITS<sup>[5]</sup>图像索引数据库构成。它依据各个国家自身情况来确定所使用的数据库管理系统，如 MYSQL、Oracle、SQLServer 等。它采用各个数据库管理系统支持的访问协议和驱动协议与方法，负责从数据集中获取属性信息，然后将结果返回给天文数据结点的查询处理层。中国虚拟天文台选用的是 MYSQL 数据库管理系统对星表进行管理。如图 3.3 所示。

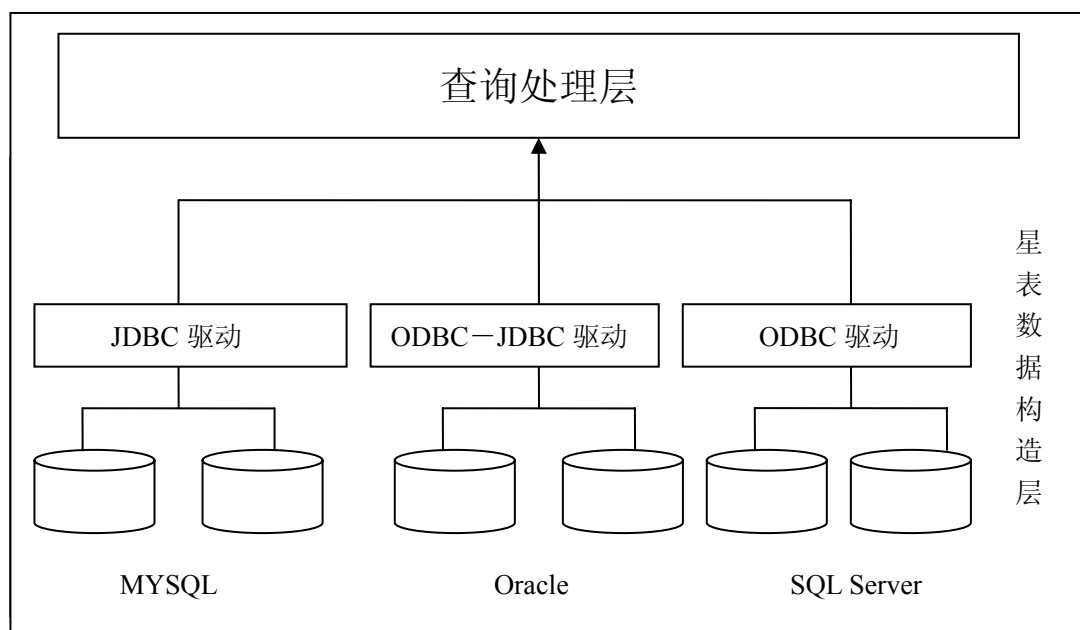


图 3.3 星表数据构造层

### 3.3 天文数据结点的粒度问题及命名规则

#### 3.3.1 天文数据结点封装的粒度

谈到数据资源的服务化封装就一定会设计到粒度问题<sup>[6]</sup>。对于中国虚拟天文台天文数据结点的服务也同样需要规划好粒度问题。粒度是指一个服务或一组服务、对象、组件能够提供多少功能。粒度的级别一般取决于服务实体的目的。服务的粒度级别通常会比对象的粒度或者组件的粒度级别更粗。

天文数据结点的服务有两种粒度，即细粒度和粗粒度。所谓天文数据结点的细粒度封装，是指将星表数据构造层中星表数据库中的每个表的每个字段都封装到中国虚拟天文台的数据访问结点中，并对外发布为服务接口。粒度为表级管理，字段级封装。细粒度封装所带来的好处是在字段级服务的基础上提供更多的数据融合功能，字段之间操作更为方便，天区之间的交叉认证工作可能会更灵活。但是细粒度封装的弱点是工作量大，并且用户操作复杂，集成、管理都非常困难，并且不适用。

天文数据结点的粗粒度封装则是指将星表数据构造层中的某一个星表数据库管理系统集成到数据访问结点中。粒度为数据库管理系统级管理，库级封装。比如说，将存有 2MASS、USNO-B、SDSS 等星表数据库的 MySQL 数据库管理系统进行粗粒度封装时，我们在天文数据结点的服务上仅仅能操作到 2MASS、USNO-B、SDSS 等的星表数据库，但其中存储数据的多个星表列表却很难操作。粗粒度封装的优势就是工作量小、服务的集成和管理方便，但是这些是以牺牲用户可操作性、可扩展性换来的。因此也并不适合作为中国虚拟天文台的天文数据结点的服务。

我们经过权衡做了一种折中的考虑，将星表数据构造层中数据库管理系统的某一星表数据库中的星表进行封装，集成到天文数据结点中。粒度为库级管理，表级封装。因为这样可以兼顾两者之优势，把数据资源管理工作的难度降低，并且可以提高服务的数据访问和互操作能力，更好的实现服务的原子性和扩展性。这样天文数据结点之间数据资源相互分离，用

户可以更灵活的使用这些数据访问服务，在天文数据结点上进一步开发适合自己的管理平台、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与分析工作。三种粒度的对比图如 3.4 图所示。

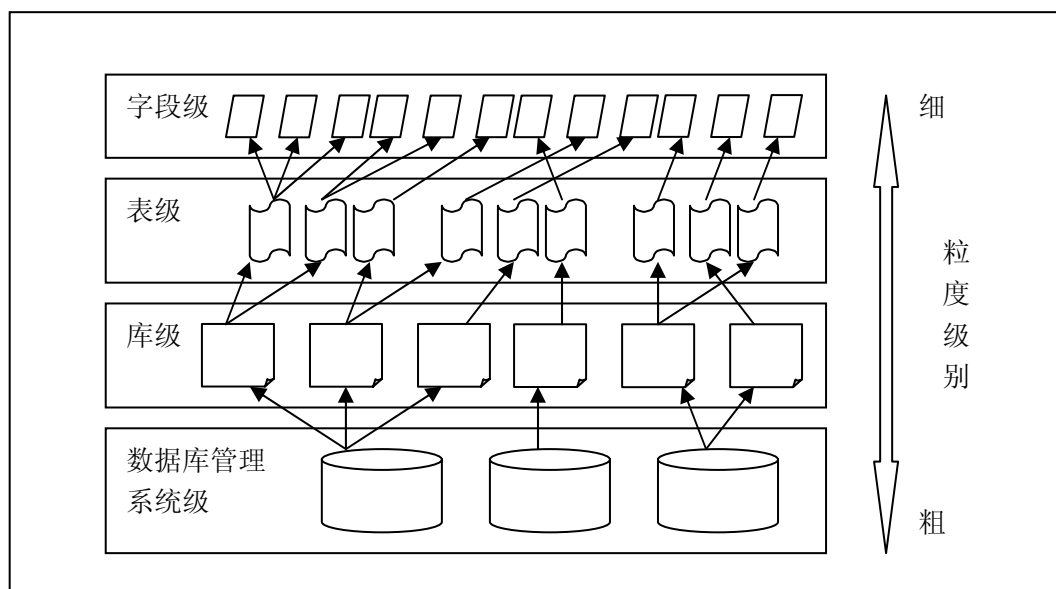


图 3.4 粒度级别说明

### 3.3.2 天文数据结点接口封装的命名规则及配置

#### 1) 数据访问服务接口的命名规则

天文数据结点的访问服务接口命名时遵循“同一原则”。当对一个星表数据库进行封装时，我们就以该数据库的名称来命名这个接口，该接口所实现的方法名和其他的接口相同。比如：将 2MASS 星表数据库进行封装，其里面有三个表：twomass\_psc、twomass\_scn、twomass\_xsn，用户只需要对这个库进行封装，就可以通过接口方便的管理其中的三个星表。接口的方法则按照天文数据结点数据访问服务接口所定义方法名来发布。接口发布命名规则如图 3.5 例图所示：

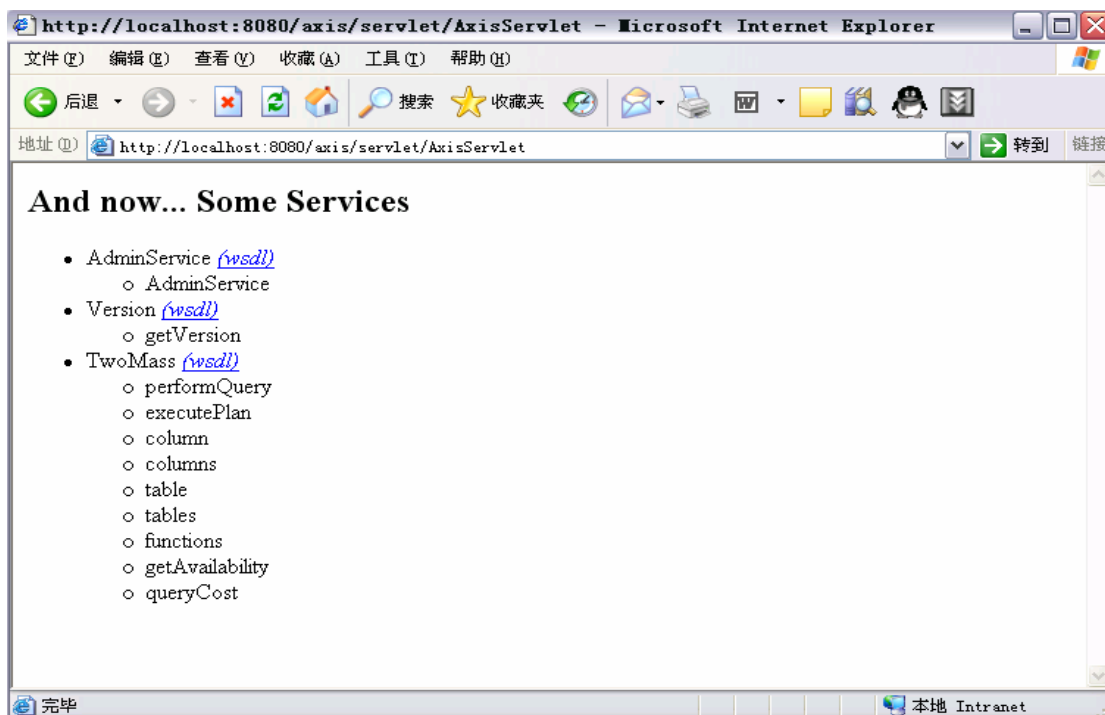


图 3.5 2MASS 星表访问服务接口及其方法

## 2) 数据访问服务元数据的配置规则

天文数据结点的访问服务接口对所封装的星表数据库以 XML 文件的形式进行配置管理，将数据库驱动信息存储在 `Configure.xml` 文件中。系统利用 NCSA 模块对配置好的文件进行访问，以此获取数据库驱动、数据库设置信息以及星表、字段等的元数据。服务器端配置文件

`Configure.xml` 内容如表 3.1 所示。

```
<?xml version="1.0" encoding="UTF-8"?>
<Configure xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns="http://jansky.nsa.uiuc.edu:8080/xml/SkyNode/v0.1"
  xsi:schemaLocation="http://jansky.nsa.uiuc.edu:8080/xml/SkyNode/v0.1Configure-v0.1.xsd">
  <ShortName>TwoMass</ShortName>
  <XSLTPath>c:\skynode</XSLTPath>
  <XSLTFile>adql2mysql_axis.xsl</XSLTFile>
  <JDBCDriver>com.mysql.jdbc.Driver</JDBCDriver>
  <JDBCurl>jdbc:mysql://badc.lamost.org/TwoMass</JDBCurl>
  <Username>boliu</Username>
  <Password>oblamost</Password>
  <ArchiveName>TWOMASS</ArchiveName>
  <MetaFinderClass>net.ivoa.node.ConfigMetaFinder</MetaFinderClass>
  <SkyNodeClass>net.ivoa.node.FullNode</SkyNodeClass>
```

```

<XML2SQLClass>net.ivoa.node.XSLTTranslator</XML2SQLClass>
<MetaSource>XML</MetaSource>
<Format>VOTABLE</Format>
<ColumnTypes>
  <DBColType>BIGINT</DBColType>
  <VOTColType>long</VOTColType>
  <DBColType>CHAR</DBColType>
  <VOTColType>char</VOTColType>
  <DBColType>DATE</DBColType>
  <VOTColType>char</VOTColType>
  <DBColType>DATESTAMP</DBColType>
  <VOTColType>char</VOTColType>
  <DBColType>DATETIME</DBColType>
  <VOTColType>char</VOTColType>
  <DBColType>DECIMAL</DBColType>
  <VOTColType>char</VOTColType>
  <DBColType>DOUBLE</DBColType>
  <VOTColType>double</VOTColType>
  <DBColType>FLOAT</DBColType>
  <VOTColType>float</VOTColType>
  <DBColType>INT</DBColType>
  <VOTColType>int</VOTColType>
  <DBColType>LONG</DBColType>
</ColumnTypes>
<IncludeTable>twomass_psc</IncludeTable>
<IncludeTable>twomass_scn</IncludeTable>
<IncludeTable>twomass_xsc</IncludeTable>
<MetaTable>
  <Table>
    <TableName>twomass_psc</TableName>
    <TableDescription>The TwoMass_PSC cluster Catalogue </TableDescription>
  </Table>
</MetaTable>
<MetaColumn>
  <Column>
    <ColumnID>twomass_psc.ra</ColumnID>
    <ColumnDescription> ra description</ColumnDescription>
    <ColumnUCD>TwoMass.twomass_psc.ra</ColumnUCD>
    <ColumnUnit>double</ColumnUnit>
  </Column>
</MetaColumn>
</Configure>

```

表 3.1 天文数据结点元数据配置文件

### 3.4 天文数据结点各模块的功能

天文数据结点的主体部分有六个模块,分别是由 VOTABLE 输出转换模块、ADQL 查询解析模块、COORDS 坐标解析模块、REGION 天区解析模块、SKYNODE 数据访问服务接口模块和 NCSA 元数据配置模块组成。其中 SKYNODE 模块是天文数据结点的核心,负责与其他模块之间的交互以及数据访问服务接口的发布;ADQL 模块、REGION 模块与 COORDS 模块负责数据格式解析与转换等数据互操作相关工作;NCSA 模块负责对天文星表数据库中元数据的描述及定义。VOTABLE 模块则对查询的结果进行输出格式转换。以下是各个模块的功能:

- 1) SKYNODE 数据访问服务接口模块:它负责发布数据访问服务接口,一共实现了九个方法。它通过方法 `performQuery()`和 `executePlan()` 进行查询,这两个接口都可以作为天文数据结点的入口和出口。虽然它们输入不同,运行机制也不同,但输出都是以 `VOData` 对象的形式返回给上层应用服务。其余的方法都是用来支持查询的辅助方法,比如获取天文数据结点中元数据列、元数据表、结点状态、查询代价、函数功能列表等信息。
- 2) NCSA 模块:它在提取上层服务对结点发送的 ADQL 查询信息中的关键字后,从服务器端配置的 `Configure.xml` 文件中获取相应的元数据、UCD、数据表的定义信息等等。该模块只有在获取星表元数据时才使用,查询时并不调用该模块。
- 3) ADQL 模块:它内部实现了 ADQL 解析、REGION 解析、XMatch 解析、COORDS 解析的功能,并将解析后的对象通过 `ADQLtoSQL` 翻译器转换为标准 SQL 查询语句送至星表数据构造层的异地的天文星表数据库进行查询,查询结果以 `ResultSet` 对象形式输出给 VOTABLE 输出转换模块。
- 4) REGION 模块:提供 Cone Search 的查询,返回 `CoordsType` 类型的对象。Cone Search<sup>[7]</sup>锥形天区查询依据 `SkyNodeInterface` 标准设计,即给出查询中心点的赤道坐标,即赤经、赤纬和半径三个约束条件,

查询出给定半径内所有目标的信息。具体意义描述如下表 3.2 所示:

参数名称	数据类型	参数意义	单位
ra	double	天区中心点赤经值	度
dec	double	天区中心点赤纬值	度
radius	double	搜索天区半径	度

表 3.2: Cone Search 的输入参数

- 5) COORDS 模块:接收 REGION 模块传递过来的对象,提供坐标转换。转换后的坐标对象送至 ADQL 模块。
- 6) VOTABLE 模块:根据天文数据的统一数据标准,把来自不同数据集的数据以 IVOA 标准的 VOTable 格式输出。它接收一个 ResultSet 对象,将结果转换为 VOTable 格式,以 VOData 对象的形式通过 HTTP 协议返回给上层服务。

### 3.5 天文数据结点的接口

根据 SkyNodeInterface 标准服务接口定义,天文数据结点封装星表数据库时对外发布 9 个可操作的服务接口。下面对每一个接口的输入输出进行详细的描述,在具体实现时开发者需要按照该接口描述完成编码。

#### 1) table()方法:

##### A) 方法原型:

```
public MetaTable table(String table)
```

##### B) 功能:

获取星表数据库中的一个指定表的元数据。

##### C) 参数:

参数名称	数据类型	参数意义
Input	String	输入所要查的表名(由注册中心获取)
Output	MetaTable	返回所查询表的元数据:表的名称、表的描述、表的行数。

表 3.3: table()方法输入输出参数表

## 2) tables()方法:

## A) 方法原型:

```
public ArrayOfMetaTable tables()
```

## B) 功能:

获取星表数据库中的所有表的元数据。

## C) 参数:

参数名称	数据类型	参数意义
Input	None	
Output	ArrayOfMetaTable	返回所查询库中所有表的元数据的数组: 表的名称、表的描述、表的行数。

表 3.4: tables()方法输入输出参数表

## 3) column()方法:

## A) 方法原型:

```
public MetaColumn column(String table, String column)
```

## B) 功能:

获取星表数据库中的指定表的指定列的元数据。

## C) 参数:

参数名称	数据类型	参数意义
Input	String	输入一个有效星表数据库的星表名
	String	输入该星表的列名
Output	MetaColumn	返回所查询库中指定星表的指定列的元数据: 列的名称、单位、描述、UCD信息

表 3.5: column()方法输入输出参数表

## 4) columns()方法:

## A) 方法原型:

```
public ArrayOfMetaColumn columns(String table)
```



B) 功能:

获取星表数据库中的指定表的所有列的元数据。

C) 参数:

参数名称	数据类型	参数意义
Input	String	输入星表数据库中的一个表名
Output	ArrayOfMetaColumn	返回所查询库中指定星表的所有列的元数据数组: 列的名称、单位、描述、UCD信息

表 3.6: columns()方法输入输出参数表

5) performQuery()方法:

A) 方法原型:

public VOData performQuery(SelectType select, String format)

B) 功能:

输入一个 ADQL 语句, 返回一个 VOData 对象。

C) 参数:

参数名称	数据类型	参数意义
Input	SelectType	解析上层服务传递来的 ADQL 语句
	String	指定输出格式, 默认为 VOTable 格式。
Output	VOData	返回查询结果的对象, 输出时可以转换为 VOTable 格式

表 3.7: performQuery()方法输入输出参数表

6) executePlan()方法:

A) 方法原型:

public VOData executePlan(ExecPlan plan)

## B) 功能:

输入经过解析后的 `ExecutePlan` 对象, 返回一个 `VOData` 对象。内部运行机制如下: 接收传送过来的 `PlanID`、查询语句、输出格式、返回地点的 `URL`、主机名等信息后, 将查询语句向 `ADQL` 模块传送, 传递给 `ExecutePlan` 对象进行执行。

## C) 参数:

参数名称	数据类型	参数意义
Input	ExecPlan	整合上层传递下来的对象, 然后将查询传递至数据库。
Output	VOData	返回查询结果的对象, 输出时可以转换为 <code>VOTable</code> 格式

表 3.8: `executePlan()`方法输入输出参数表7) `queryCost()`方法:

## A) 方法原型:

```
public float queryCost(long planid, SelectType select)
```

## B) 功能:

可以利用该方法返回一个查询代价, 以此来决定执行计划的顺序。它利用一些统计规则或启发式方法来判别每平方度中目标的密度, 由此来决定一个查询返回结果的大小。

## C) 参数:

参数名称	数据类型	参数意义
Input	SelectType	整合上层传递下来的对象, 经过 <code>SQL</code> 语句转换为执行代价, 然后将查询传递至数据库。
	Long	指定一个执行计划的 <code>id</code>
Output	Float	返回查询条件获得星标的行数。

表 3.9: `queryCost()`方法输入输出参数表

8) `getAvailability()`方法:

A) 方法原型:

```
public Availability getAvailability()
```

B) 功能: 由天文数据访问门户或者客户端程序来测试某个注册的结点是否还存在, 如果存在, 返回该天文数据结点的服务器时间、服务器名称、信息和位置。

C) 参数:

参数名称	数据类型	参数意义
Input	None	
Output	Availability	返回要测试的天文数据结点状态, 如果存在返回当前结点状态信息, 如时间戳、IP 地址、结点闲忙状态等。不存在返回 null。

表 3.10: `getAvailability()`方法输入输出参数表

9) `functions()`方法:

A) 方法原型:

```
public ArrayOfMetaFunction functions()
```

B) 功能: 获得一个天文数据结点的所包括函数名参数列表和这个天文数据结点支持的所有额外函数的注解。

C) 参数:

参数名称	数据类型	参数意义
Input	None	
Output	ArrayOfMetaFunction	返回要测试的天文数据结点所支持的所有函数方法描述的数组。

表 3.11: `functions()`方法输入输出参数表

根据以上发布的接口以及实现的方法，天文数据结点的类图如图 3.6 所示。

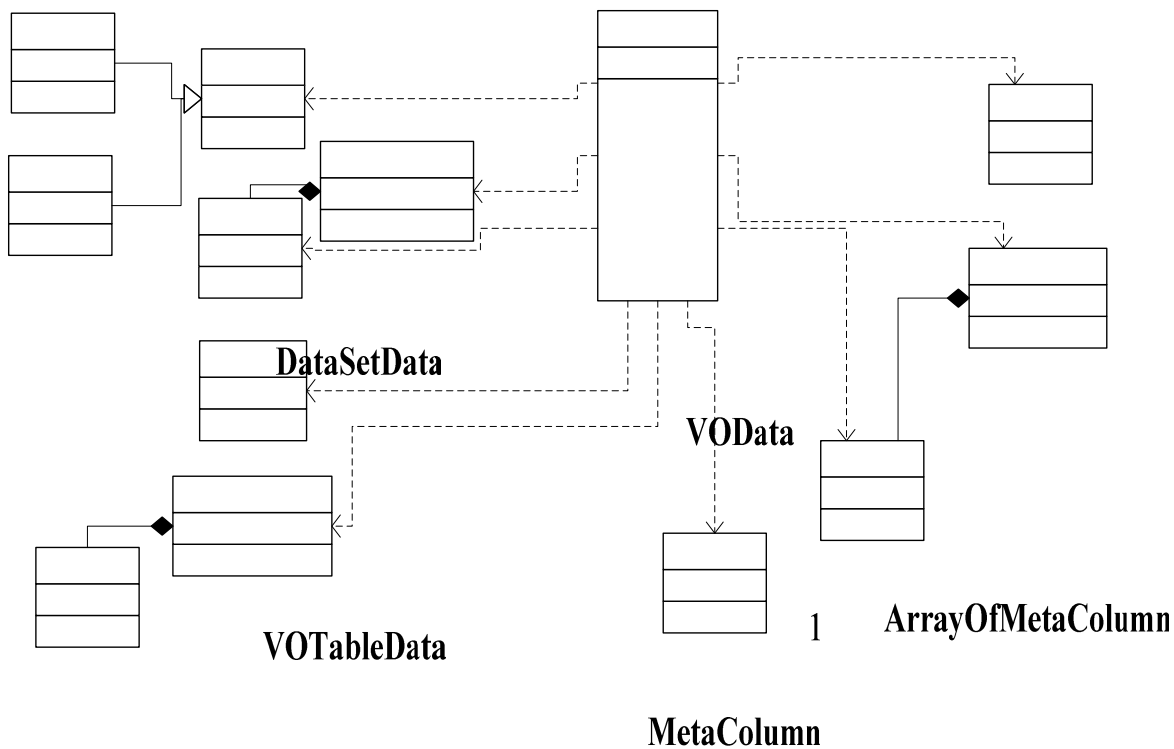


图 3.6 天文数据结点接口类结构图

### 3.6 天文数据结点的运行流程

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。

#### 3.6.1 天文数据结点的简单查询

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 模块直接转换为 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证认功能，无法获取星表证认信息。天文数据结点简单查询的运行流程图如图 3.7 所示。

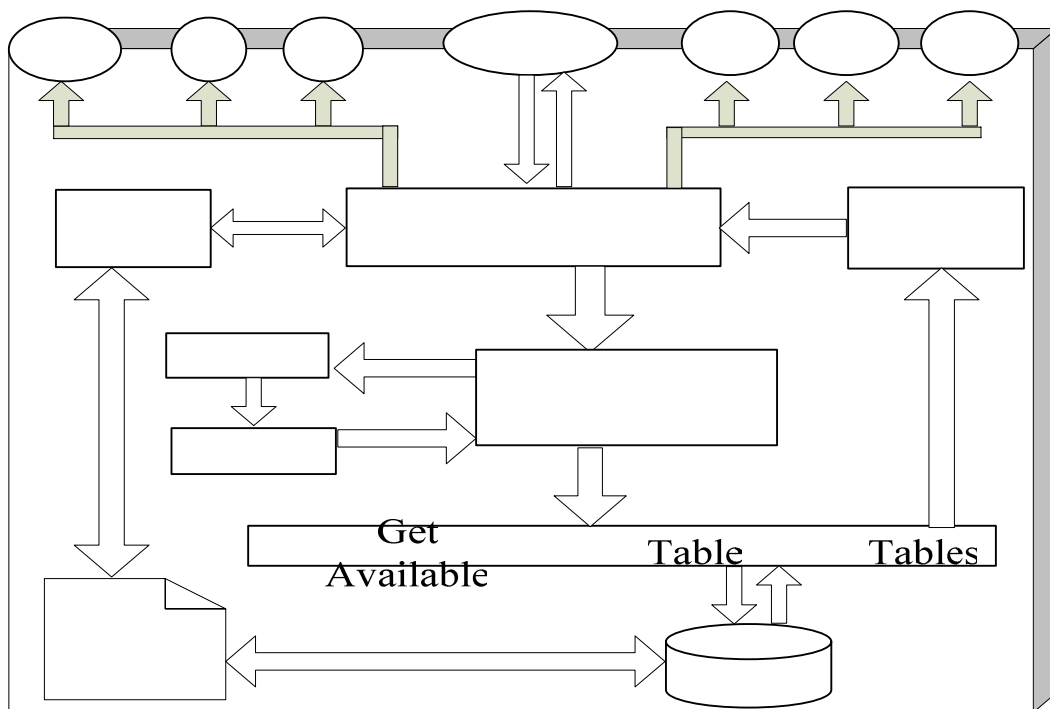


图 3.7 天文数据结点简单查询流程图

### 3.6.2 天文数据结点的复杂查询

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 3.8 所示。

NCSA

MetaInfo

SKY

MetaInfo

REGION

RegionType

COORDS

SelectType

Configure

MetaInfo

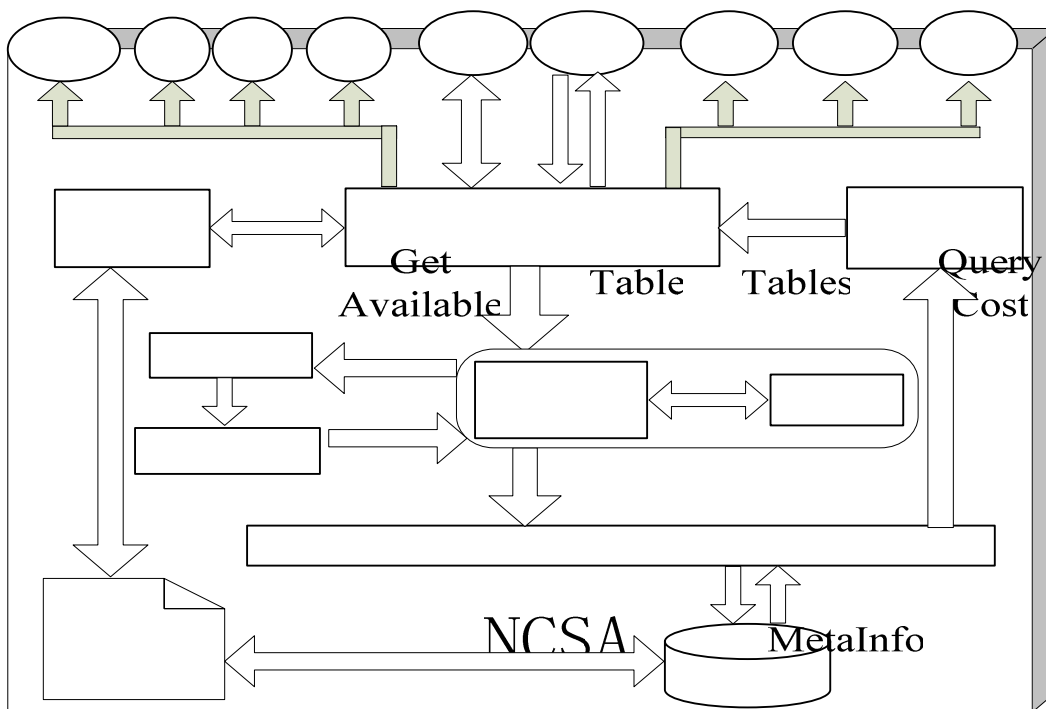


图 3.8 天文数据结点复杂查询流程图

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、交叉证认的条件和锥形查询区域的 `ADQLRegionType` 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()` 方法来决定的（具体的排序方法在 4.4 节进行详细的解释）。

接着，`executePlan()` 方法拿到了这个 `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`，并且保存下一个要传送的结点的主机名和 `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。具体查询步骤如下：

- 1) 提取 `ADQL` 查询语句中的 `Where` 子句中的查询条件，分别进入 `REGION` 模块和 `XMATCH` 模块进行解析。
- 2) `REGION` 模块收到 `RegionType` 的对象，进行区域查询的解析，然后再进入 `COORDS` 模块进行坐标解析，最后以 `SelectTypeList` 数组对

象返回给 ADQL 模块。

- 3) XMATCH 模块收到一个 XMatchType 的对象, 进行交叉证认条件的解析, 之后返回给 ADQL 一个 SelectTypeList 数组对象。
- 4) 由 ADQL 整理这些解析后的对象, 并且经过 ADQLtoSQL 的转换, 传递至数据库管理系统。
- 5) 查询结果以结果集的方式输出到 VOTABLE 模块中, 经由 VOTABLE 模块进行格式转换后, 保存为 VOData 对象由 executePlan 接口输出。
- 6) 输出的结果对象将依据开始保存的主机名和地址传送至下一个结点。
- 7) 天文数据结点会继续按照天文数据访问门户或客户端程序排列的顺序依次执行, 将第一个结点的查询结果传送到第二个结点并与它的查询结果证认完毕后, 保存为 VOData 对象, 并将这个结果对象传送到第三个结点。第三个结点接收第二个结点传来的 VOData 结果对象后, 与第三个结点的查询结果进行交叉证认, 完毕后将证认的结果保存为一个 VOData 对象, 继续向下一个结点传递。程序依次迭代, 直到与最后一个结点证认完毕后, 将多次证认的结果以 VOData 对象的形式保存。

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。

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。

## 3.7 具体技术实现与实例

### 3.7.1 实现工具

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<sup>[8]</sup> 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 3.9 所示。

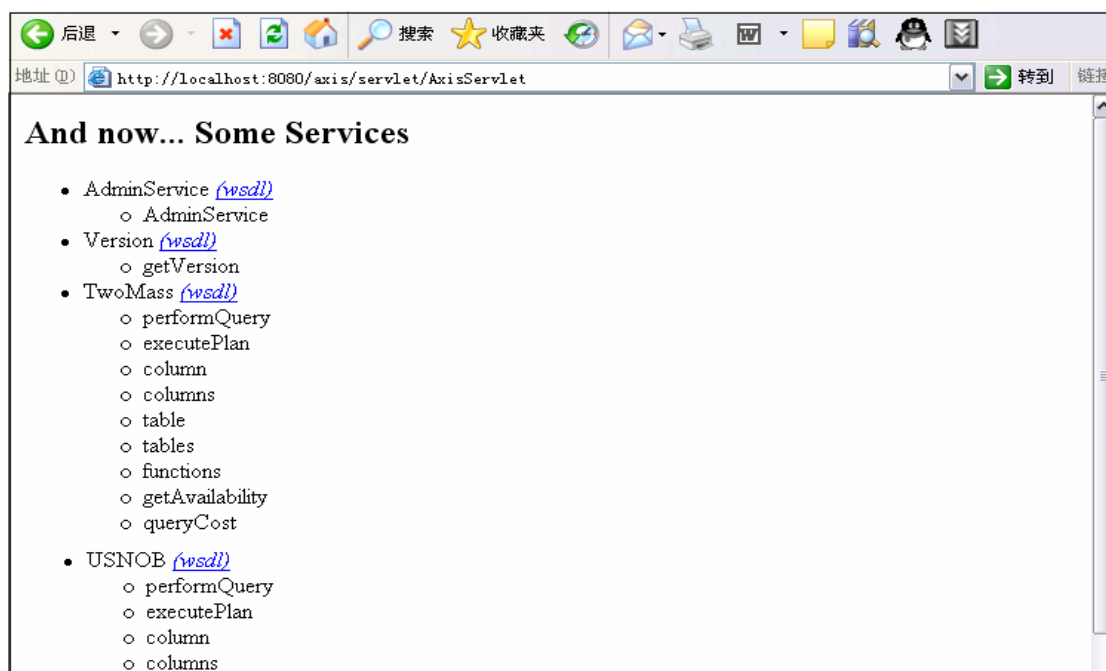


图 3.9 天文数据结点访问服务接口

### 3.7.2 天文数据结点锥形查询实例

锥形查询属于空间查询的范畴，很多大型的数据库如 Oracle、DB2 等提供了一些空间查询的函数，但并不太适用于天文研究。因此我们使用赤经、赤纬两列去实现锥形查询。

锥形查询的理论公式如下说明：

所有符合要求的天体与查询中心点之间的夹角小于等于给定的查询半径： $R$ ，天球上两个点赤经、赤纬值分别为 $(\alpha_1, \delta_1)$ 和 $(\alpha_2, \delta_2)$ ，即利用球面上余弦定理，天球上这两点间的角距离  $\theta$  为：

$$\theta = \arccos(\sin\delta_1 \sin\delta_2 + \cos\delta_1 \cos\delta_2 \cos(\alpha_1 - \alpha_2))$$



用户传送给天文数据结点锥形查询的 ADQL/s 查询语句如表 3.12 所示。

```
SELECT a.ra as twomass_ra ,a.dec as twomass_dec
FROM TwoMass:twomass_psc a
WHERE Region('CIRCLE J2000 181.3 -0.76 6.5')
```

表 3.12 用户提交的 ADQL 锥形查询语句

经过转换后用于传输的锥形查询实例的 ADQL/x 语句如下表 3.13 所示。

```
<Select xmlns:xsd=http://www.w3.org/2001/XMLSchema
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xmlns="http://www.ivoa.net/xml/ADQL/v0.7.4">
  <SelectionList>
    <From>
    <Where>
      <Condition xsi:type="regionSearchType">
        <Region xmlns:q1="urn:nvo-region" xsi:type="q1:circleType" coord_system_id="">
          <q1:Center ID="" coord_system_id="">
            <Pos2Vector xmlns="urn:nvo-coords">
              <Name>Radeg Dedeg</Name>
              <CoordValue>
                <Value>
                  <double>181.3</double>
                  <double>-0.76</double>
                </Value>
              </CoordValue>
            </Pos2Vector>
          </q1:Center>
          <q1:Radius>6.5</q1:Radius>
        </Region>
      </Condition>
    </Where>
  </Select>
```

表 3.13 经过 ADQL 模块转换后的 ADQL/x 锥形查询

由中国虚拟天文台门户的画图工具给出的可视化结论如图 3.10 所示。

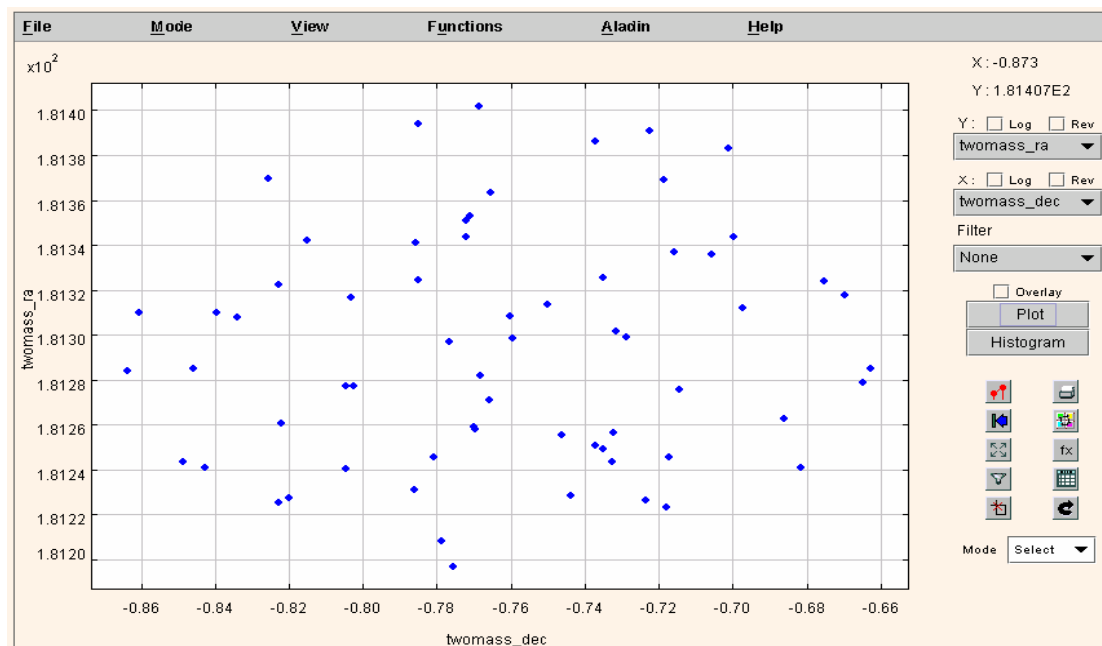


图 3.10 锥形查询结果

实现天文数据结点中国虚拟天文台数据中心的资源与国际上的数据资源可以方便地交互，加速了数据融合，从而有助于在海量数据上建立天文数据分析平台，为进一步的数据可视化工作、数据挖掘工作提供底层支持。这对推动中国虚拟天文台与各国数据中心进行大规模天文数据访问与分析具有很重要的现实意义。

## 参考文献

- [1] [SDSS] <http://www.sdss.org/>
- [2] [CDS] <http://cdsweb.u-strasbg.fr/>
- [3] [Aladin] <http://aladin.u-strasbg.fr/>
- [4] 崔辰州. 中国虚拟天文台系统设计: [博士学位论文].北京: 中国科学院研究生院
- [5] [FITS] <http://fits.gsfc.nasa.gov/>
- [6] 王晓倩. 中国虚拟天文台软件集成: [硕士学位论文].北京: 中国科学院研究生院
- [7] 桑键. 中国虚拟天文台数据库和数据服务: [硕士学位论文].北京: 中国科学院研究生院
- [8] [TOMCAT] <http://tomcat.apache.org/>

## 第四章 中国虚拟天文台天文数据访问门户的设计与实现

### 4.1 天文数据访问门户产生的背景

随着天文数据结点的发布，如何操作和管理这些天文数据结点为天文学家服务，以便进行全球范围内的数据访问和数据分析，为多波段数据分析和大样本的星表数据挖掘提供底层支持成为我们研究的重心。

我们对中国虚拟天文台的天文数据访问门户（以下简称天文数据访问门户，SkyPortal）进行设计与实现，并在其中尝试性的加入分布式调度引擎，对天文数据结点访问和管理进行优化。天文数据访问门户的分布式调度引擎为一直困扰着各国虚拟天文台的大数据量的获取问题提供一种解决方案。本章节阐述了天文数据访问门户的背景、体系结构、分布式调度引擎的设计与原理及门户的运行流程，并对相关具体实现技术进行了讨论。

### 4.2 天文数据访问门户的层次结构

#### 4.2.1 天文数据访问门户的体系结构

天文数据访问门户体系结构依据自顶向下的思想分为三层，分别是访问门户层、数据访问调度层和天文数据结点层。它的层次结构如图 4.1 所示：

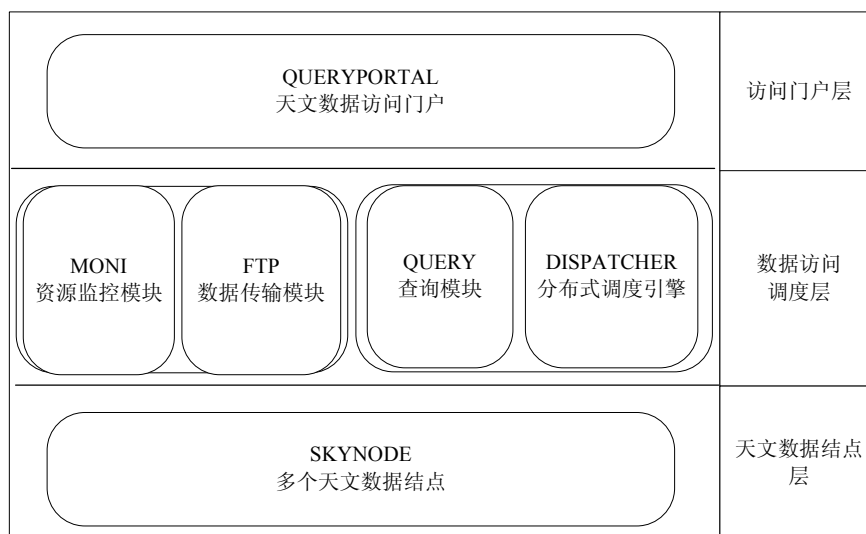


图 4.1 天文数据访问门户的体系结构

**访问门户层：**位于中国虚拟天文台的门户层，提供了用户与中国虚拟天文台的汇聚层交互的平台。它基于 B/S 结构以 JSP<sup>[1]</sup>页面的形式展现给用户。用户通过中国虚拟天文台的天文数据访问门户可以方便的获取到分布在世界各地的星表数据资源，同时也提供给用户星表交叉证认、锥形查询（Cone Search）等高级操作。

**数据访问调度层：**这是天文数据访问门户的核心，它位于中国虚拟天文台的汇聚层，是天文数据结点调度与查询的分配点。它为中国虚拟天文台的上层应用服务提供操作接口，通过内部的分布式调度引擎对分布式天文数据结点进行任务调度、查询分割，然后通过查询模块对数据资源进行获取。

**数据结点层：**位于中国虚拟天文台的数据资源层，由很多分布的天文数据结点组成，它们为数据访问调度层提供了统一的访问服务接口。这一层是上层一切服务的基础。

#### 4.2.2 天文数据访问门户各层的功能

天文数据访问门户三个层次的功能如下：

- 1) **访问门户层：**天文数据访问门户为普通用户提供星表数据查询功能以及星表的交叉证认功能。它用 Tomcat 作为发布引擎，以 JSP 页面的形式提供给用户可视化的界面，使用户能进行“一站式”获取自己想要的数。它的基本功能是用户查询任务的提交和处理返回结果。
- 2) **数据访问调度层：**是天文数据访问门户的核心，也是用户获取分布式天文数据结点的数据资源唯一的途径。它由两个子系统组成：资源管理子系统和查询调度子系统，主要是对分布的天文数据结点进行统一的调度、查询、监控并且将查询结果传输至中国虚拟天文台的存储服务中保存。
- 3) **天文数据结点层：**是基于 WEB 服务架构或网格体系架构，并且遵循了 IVOA 的标准，实现了 ADQL 规范、VOTable 规范、SkyNodeInterface 规范。它能够集成广域网环境下异构的各种星表数

据资源及星表数据库系统，并将它们统一组织起来，通过系统提供的数据库访问服务屏蔽底层数据资源异构性和多个管理域，为上层服务提供直观、方便、规范的访问和操作接口。

### 4.3 天文数据访问门户各模块的功能

天文数据访问门户按功能分为两个子系统，分别是由资源管理子系统和查询调度子系统组成。资源管理子系统中有两个模块：一个是 FTP 数据传输模块，另一个是 MONI 资源监控模块。查询调度子系统则由 QUERY 查询模块和 DISPATCHER 分布式调度引擎组成。其中资源管理子系统主要对已注册的天文数据结点进行监控管理，并将返回的查询结果传送到 MyVO 的 FTP 存储服务中保存。查询调度子系统则将用户提交的查询语句进行查询分割和执行计划分配，并执行数据检索。下面是各个模块的功能：

- 1) MONI 资源监控模块：它负责对中国虚拟天文台注册中心中已经注册的天文数据结点进行监控，定期的获取天文数据结点的激活状态和压力状况。当执行查询前，它调用天文数据结点的 `getAvailability()` 方法对某个指定的结点状态进行判断，如果该结点出于激活状态并且压力较小则开始执行查询。
- 2) FTP 数据传输模块：将查询后的结果通过 FTP 模块进行传输，按照调度引擎事先分配的编号以文本的形式保存到中国虚拟天文台的 MyVO 中。用户可以通过天文数据访问门户页面的链接对数据进行下载或画图操作。
- 3) QUERY 查询模块：它负责执行用户发送过来的查询请求，并且将结果以文本的方式返回给 FTP 模块。当用户执行简单查询时，系统调用天文数据结点 `performQuery()` 方法进行单结点查询。当用户执行带有多点交叉认证的查询时，系统经过 DISPATCHER 分布式调度引擎分配执行计划后，调用天文数据结点的 `executePlan()` 方法进行多点星表的交叉认证。
- 4) DISPATCHER 分布式调度引擎：依据用户提交的查询语句和天文数据结点的状态对查询进行代价评估、查询分割、分配执行计划，使

每个查询达到最优，然后将优化后的查询语句传送到压力较小的天文数据结点，下面的 4.4 节会详细的介绍分布式调度引擎的工作原理。

## 4.4 天文数据访问门户的分布式调度引擎

### 4.4.1 分布式调度引擎的理论依据

天文数据访问门户的分布式调度引擎利用天文数据特有的特点，对星表数据库的查询语句进行分割。依据星表中不同天区的赤经、赤纬，对一个大的查询进行合理的划分。这样消除了单一结点对大查询指令的处理，缓解了单一结点的计算机压力，用分布检索小查询片的方式实现了对大数据集的获取。

一般情况下，很难将 SQL 语言分成一个可以并行查询的指令集。但是考虑到具体的星表查询，这是比较容易做到的。几乎所有的星表中都含有目标的坐标，即赤经/赤纬或银经/银纬。我们根据天体的坐标对 SQL 查询语句进行一个划分。这个划分是对结果集的一个划分，即它们之间没有交集，而它们的并集就是原 SQL 查询语句的结果集。因此对任何一条针对星表的 SQL 查询语句，都可以按照天区分割成很多小块，在这个小块中查询。这些查询结果最后可以再合并在一起。

处理星表交叉认证时将交叉认证的工作从数据库转移到一个单独的程序（执行计划组件）中完成。上面描述的对 SQL 查询语句的分割在做了交叉认证相关的数据库查询语句调整以后仍然适用。

### 4.4.2 分布式调度引擎的调度方法

我们依据资源监控模块中获取的天文数据结点的状态信息，循环的监测每个结点的压力情况，选择压力小的结点进行查询分配。在调度方法上依次执行以下三个步骤：

- 1) 查询代价分析

它调用天文数据结点的 `queryCost()` 方法对用户提交的查询语句进行代价分析。该方法内部执行的是获取查询记录条数的 SQL 语句。依据查询代价的数据量大小来判断是否进行查询分割。查询代价分析执行的 SQL 语句如图 4.3 所示。

```
SELECT COUNT(*) FROM Tablename
```

图 4.3 查询代价分析 SQL 语句

## 2) 查询语句分割

将查询的数据量与预先设定好的阈值比较后，如果数据量没有超过设定的阈值则进行单语句查询，否则进行多点分布式、多语句查询。共有四种分配情况：

- A) 未超过阈值的单点查询不做任何调度分配处理。
- B) 超过阈值的单点查询仅进行查询分割。
- C) 未超过阈值的多点星表交叉证认查询，分布式调度引擎直接进行执行计划分配。
- D) 超过阈值的多点星表交叉证认查询，分布式调度引擎在执行查询分割后，再进行执行计划分配。

## 3) 执行计划分配

天文数据访问门户处理多点交叉证认<sup>[2]</sup>时由系统的执行计划组件完成。它会先将用户所查询的星表按照查询代价进行排序，然后依次把排序后的查询发送到不同的结点中。天文数据结点收到任务后，按照任务计划所指定的顺序，序号小的优先执行查询，并把结果依次向指定目标的结点传递，在最后一个结点把交叉证认的结果传递给虚拟天文台的存储设备保存。执行计划组件收到一个带有交叉证认的查询语句时，它的执行步骤如下：



- A) 评估每个证认星表的查询代价。向含有查询星表的结点发送查询，获得星表的数据量。
- B) 依据数据量的大小来排列结点的执行顺序，以 PlanID 来表示。并且把目标结点的 URL 与 PlanID 绑定在一起。
- C) 第一个结点查询完毕后，该结点会把结果向第二个结点传递，并与第二个结点的查询结果进行交叉证认。然后把证认后的结果向第三个结点传递，依次迭代执行。
- D) 最后一个结点证认完毕后，结果由 FTP 传输模块传送到虚拟天文台的存储设备中。

多点交叉证认方法实例的运行结构图如图 4.4 所示：

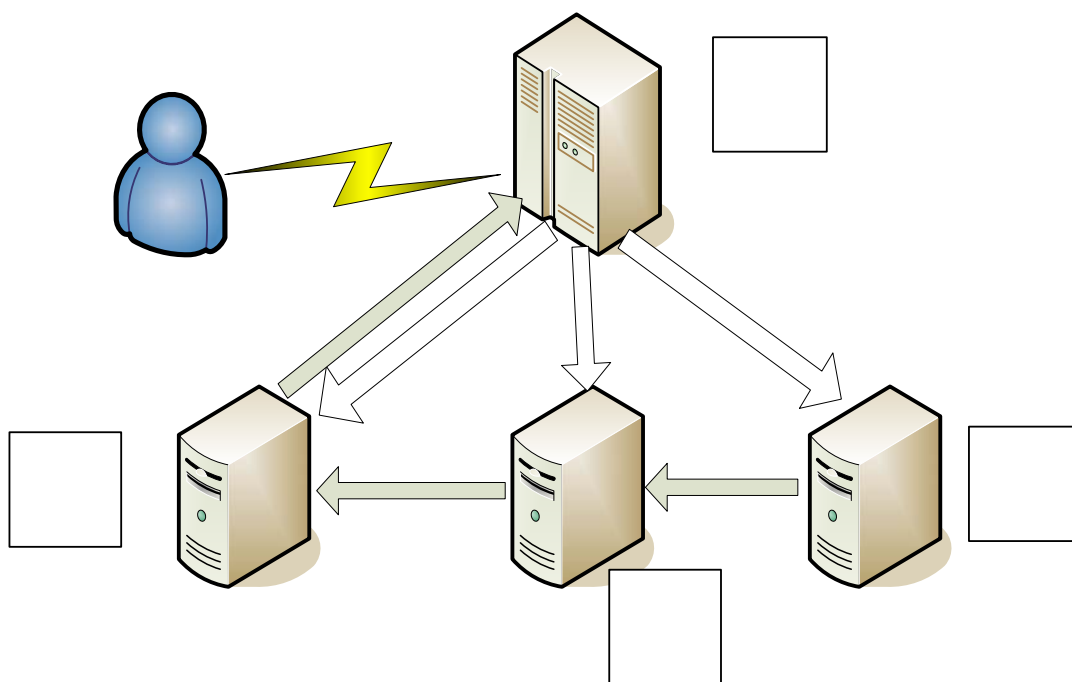


图 4.4 执行计划分配流程图

### 4.4.3 分布式调度引擎调度引擎接口

分布式调度引擎调度算法以 QueryDispatcher 接口为调度中心，接收上面传递下来的查询语句，并将其按规则分割并且对带有交叉证认的查询进

行执行计划分配后,分发至各个结点去查询。查询后的结果均传递到 MyVO 的 FTP 服务器中。分布式调度引擎的内部接口的定义用 Java 语言表示如图 4.5 所示。

```
public interface QueryDispatcher
{
    public void polling();
    public void querythread();
    public void dispatcher();
}
```

图 4.5 分布式调度引擎的内部接口

## 4.5 天文数据访问门户的接口

天文数据访问门户对外发布了 7 个可以操的天文数据结点接口,其中有 2 个查询接口,分别调用的是天文数据访问结点的 `performQuery()` 方法和 `executePlan()` 方法。其余 5 个都是元数据获取接口,分别调用的是天文数据结点的 `table()` 方法、`tables()` 方法、`column()` 方法、`columns()` 方法和 `getAvailability()` 方法。下面对每一个接口的输入输出进行详细的描述。

1) `getSkyNodeTableInfo()` 方法:

A) 方法原型:

```
public MetaTable getSkyNodeTableInfo(String node,
String tablename )
```

B) 功能:

获取指定 URL 的天文数据结点中的一个指定表的元数据。

C) 参数:

参数名称	数据类型	参数意义
Input	String	输入天文数据结点的 URL
	String	输入所要查询的星表表名

Output	MetaTable	返回所查询星表的元数据：表的名称、表的描述、表的行数。
--------	-----------	-----------------------------

表 4.1: getSkyNodeTableInfo()方法输入输出参数表

## 2) getSkyNodeTables()方法:

## A) 方法原型:

```
public MetaTable[] getSkyNodeTables(String node)
```

## B) 功能:

获取指定 URL 的天文数据结点中所有表的元数据。

## C) 参数:

参数名称	数据类型	参数意义
Input	String	输入天文数据结点的 URL
Output	MetaTable[]	返回所查询天文数据结点中所有星表的元数据的数组：表的名称、表的描述、表的行数。

表 4.2: getSkyNodeTables()方法输入输出参数表

## 3) getSkyNodeColumnInfo()方法:

## A) 方法原型:

```
public MetaColumn getSkyNodeColumnInfo(String node, String  
tablename, String columnname)
```

## B) 功能:

获取指定 URL 的天文数据结点中的一个指定星表的指定字段的元数据。

## C) 参数:

参数名称	数据类型	参数意义
Input	String	输入天文数据结点的 URL
	String	输入该天文数据结点中星表的名称
	String	输入该星表中字段的名称

Output	MetaColumn	返回所查询天文数据结点中指定星表的指定字段的元数据：字段的名称、单位、描述、UCD 信息
--------	------------	--

表 4.3: getSkyNodeColumnInfo()方法输入输出参数表

## 4) getSkyNodeColumns()方法:

## A) 方法原型:

```
public ArrayOfMetaColumn getSkyNodeColumns(String node, String
tablename)
```

## B) 功能:

获取指定 URL 的天文数据结点中的一个指定星表的所有字段的元数据。

## C) 参数:

参数名称	数据类型	参数意义
Input	String	输入天文数据结点的 URL
	String	输入该天文数据结点中星表的名称
Output	ArrayOfMetaColumn	返回所查询天文数据结点中指定星表的所有字段的元数据数组：字段的名称、单位、描述、UCD 信息

表 4.4: getSkyNodeColumns()方法输入输出参数表

## 5) submitQuery()方法:

## A) 方法原型:

```
public String submitQuery(String qry, String node)
```

## B) 功能:

执行天文数据结点的简单查询，返回 VOData 对象。

## C) 参数:

参数名称	数据类型	参数意义
Input	String	输入简单查询的 ADQL 语句（单结点）
	String	输入天文数据结点的 URL
Output	String	返回单结点的查询结果。

表 4.5: submitQuery()方法输入输出参数表

## 6) submitDistributedQuery()方法:

## A) 方法原型:

```
public String submitDistributedQuery(String qry, String node)
```

## B) 功能:

执行带交叉认证的多结点复杂查询，返回认证后的 VOData 对象。

## C) 参数:

参数名称	数据类型	参数意义
Input	String	输入带有交叉认证查询的 ADQL 语句（多结点）
	String	输入天文数据结点的 URL
Output	String	返回多结点认证后的结果

表 4.6: submitDistributedQuery()方法输入输出参数表

## 7) getSkyNodeAvailability()方法:

## A) 方法原型:

```
public Availability getSkyNodeAvailability(String node)
```

## B) 功能:

获取指定 URL 的天文数据结点的状态。

## C) 参数:

参数名称	数据类型	参数意义
Input	String	输入天文数据结点的 URL
Output	Availability	返回所查天文数据结点的状态信息。

表 4.7: getSkyNodeAvailability()方法输入输出参数表

天文数据访问门户接口的类结构如图 4.6 所示。

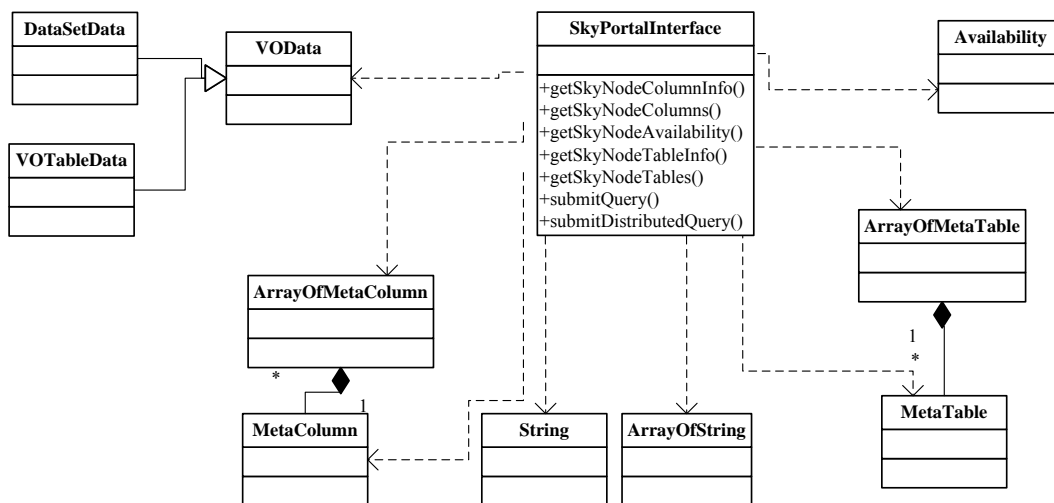


图 4.6 天文数据访问门户接口类结构图

#### 4.6 天文数据访问门户运行流程

天文数据访问门户运行流程图如 4.7 图所示：

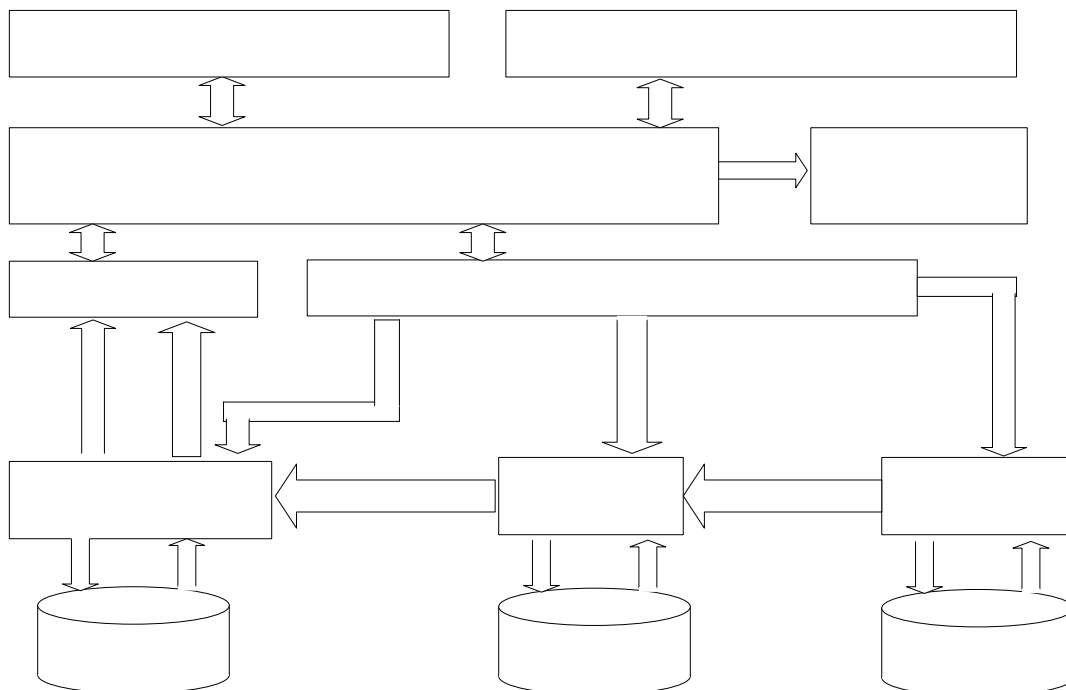


图 4.7 天文数据访问门户运行流程图

用户通过 JSP 页面向天文数据访问门户提交查询和交叉证认的条件后, 查询将向下传递至分布式调度引擎。分布式调度引擎对用户的条件进行分析并通过引擎中的资源组件来判断用户所需资源是否注册在平台之内。如果存在, 则由引擎中的调度组件执行以下的流程:

- 1) 在已经注册的天文数据结点列表中判断该查询是否为一个有效查询, 即用户所查询的星表是否已经注册在虚拟天文台的注册中心。
- 2) 调用天文数据结点的 `queryCost()`方法对用户提交的查询语句进行代价分析。
- 3) 与预先设定的阈值相比较, 判断是否进行查询分割。
- 4) 判断该查询是否是带有交叉证认的复杂查询, 如果是则产生一个执行计划, 由它来对后面的交叉证认进行管理。如果没有 `Xmatch()`字段, 则作为简单查询处理, 这个查询过程将不产生执行计划。
- 5) 系统依据用户所查询的不同的数据库进行查询分割, 当用户查询数据量很大时, 系统会将其按照每 0.1 度来进行分割, 以此来缓解每个结点上的压力。如果用户所查询的数据量本身就很小同时又是简单查询, 那么系统就不进行查询分割, 直接调用天文数据访问门户的 `SubmitQuery()`方法, 将查询语句送入空闲的天文数据结点进行查询。
- 6) 如果是带有交叉证认的复杂查询, 那么执行计划则依据所要查询的星表进行估测、排队和传输 :
  - A) 把查询语句中约束条件最多, 并且查询代价最小的天文数据结点作为第一个查询排队。依据数据量的大小来排列结点的执行顺序, 以 `PlanID` 来表示。并且把目标结点的 URL 与 `PlanID` 绑定在一起。依次排队直到执行计划分配完为止。
  - B) 由传输组件向天文数据结点分配执行计划, 同时产生一个等待查询结果的线程, 这个线程可能会等待几秒, 也可能几天。把执行计划传送到每个结点后, 结点依照排队循序依次执行。
  - C) 优先执行第一个天文结点查询, 将查询结果向第二个结点传递。
  - D) 第二个结点接收到的数据后生成临时表并与它自己的查询结果进行交叉证认后, 将证认结果向第三个结点传递。依次迭代执行。

- E) 最后一个结点认证完毕后,将结果存入虚拟天文台的存储设备,FTP 或者存入虚拟天文台的 MyVO 中,以便天文学家存取并保留数据。所有结果数据保存完毕后线程关闭。
- F) 用户可以通过天文数据访问门户的下载连接来获取查询结果,并利用画图按钮调用虚拟天文台的画图工具,对结果进行可视化。

## 4.7 天文数据访问门户实例与总结

天文数据访问门户实例利用天文台本地、中科院网络中心(CNIC)部署的多个结点,同时也利用法国 CDS 发布的数据服务接口,进行广域网范围内的数据访问、调度并将结果存储与可视化处理。本实例利用天文数据访问门户对天文数据结点中 2MASS 星表、USNO-B 星表进行交叉认证后,通过中国虚拟天文台的画图工具对结果进行显示。

### 1. 天文数据访问门户

用户对天文数据访问门户中的 2MASS 星表和 USNO-B 星表进行交叉认证。天文数据访问门户如图 4.8 所示。



图 4.8 天文数据访问门户测试页面



## 2. 天文数据访问门户内部调度处理

分布式调度引擎把查询分割后,把片段分配到四个不同的天文数据结点上进行处理。交叉认证星表时,将拆分后的查询进行结点查询代价评估,以此方式进行排序,遵从最小数据集最先执行的原则,对星表的交叉认证进行调度。交叉认证后的结果以文件的形式存放在 `badc.lamost.org` 的 FTP 服务器上。天文数据访问门户测试实例的运行图如 4.9 所示。

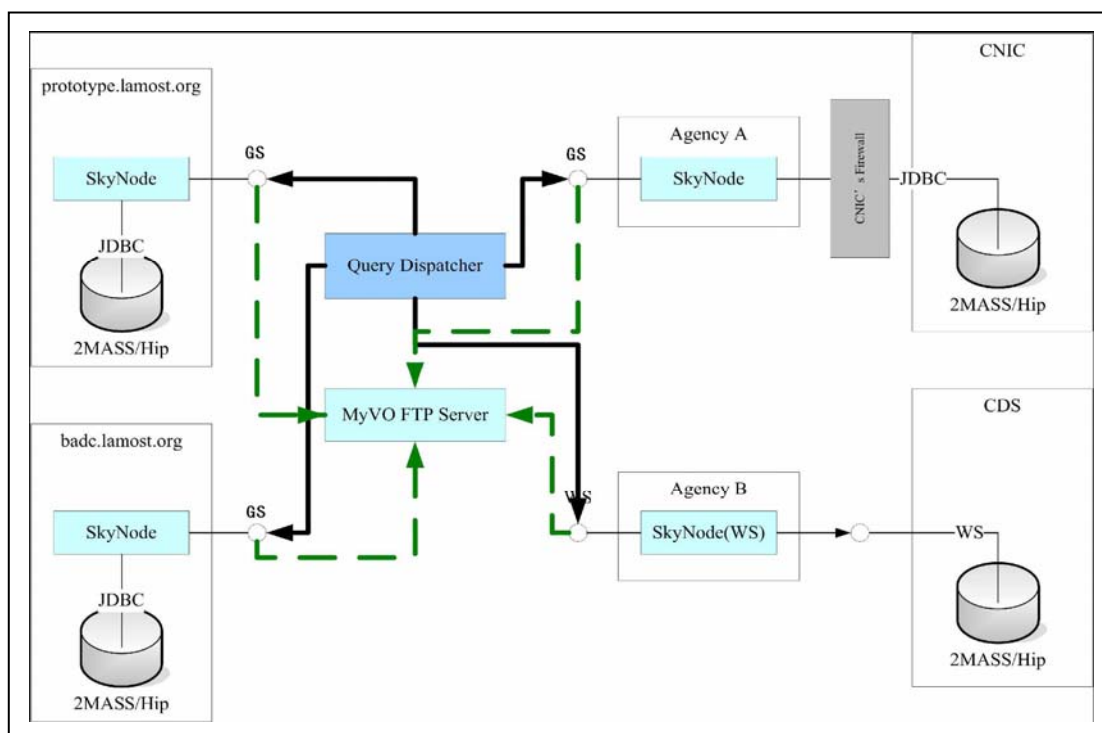


图 4.9 分布式调度引擎实例的运行图

## 3. 可视化显示

最终的结果由画图工具以可视化的方式显示给用户,给出一个直观的认证结果。交叉认证结果如图 4.10 所示。

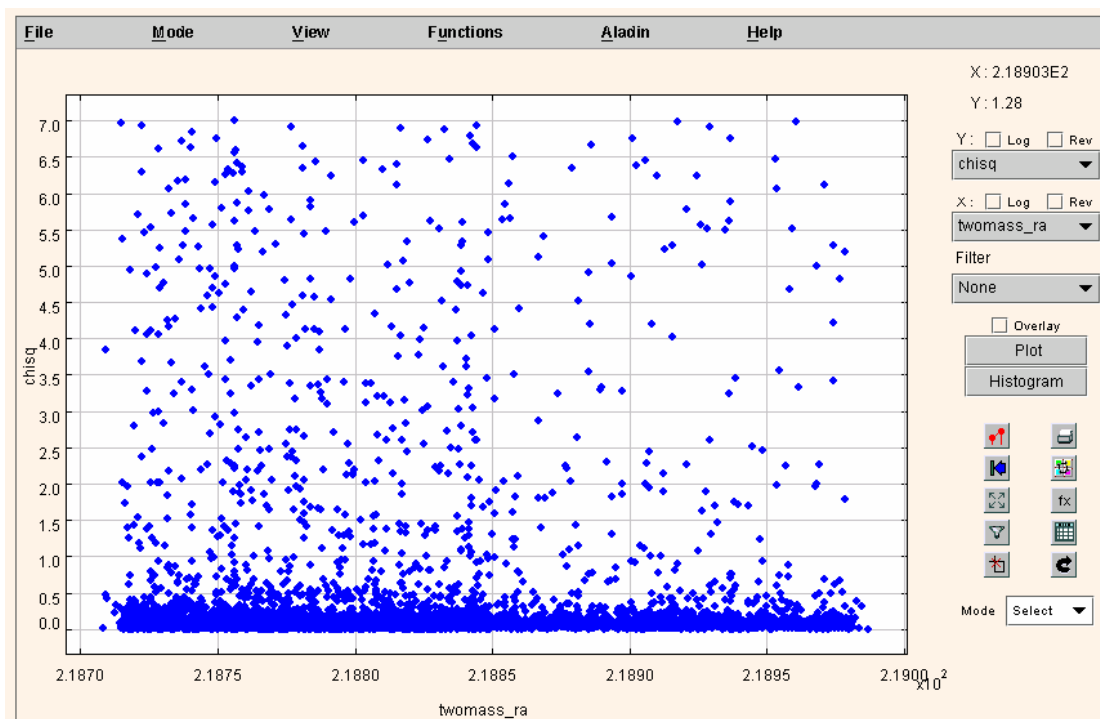


图 4.10 交叉认证结果和显示界面

实现天文数据访问门户使天文学家可以利用中国虚拟天文台对各国数据中心的观测数据直接进行访问、处理、分析。我们希望用这套系统作为中国虚拟天文台数据访问服务的门户，确保其能够提供可靠高效的大样本数据分析和研究所需的技术支持，为海量星表的数据挖掘工作提供底层支持。

## 参考文献

- [1] 飞思科技产品研发中心. JSP 应用开发详解.北京: 电子工业出版社.2003
- [2] [XMATCH] <http://www.openskyquery.net>

## 第五章 天文数据访问门户与天文数据结点在“银河系旋臂结构研究”的中国虚拟天文台范例中的应用

### 5.1 “银河系旋臂结构研究”的中国虚拟天文台范例的目的

“银河系旋臂结构研究”的中国虚拟天文台范例的目的是利用 2MASS 星表<sup>[1]</sup>红外波段的色指数证认 OB 星并估计星际消光<sup>[2][3]</sup>，然后利用绝对星等估计 OB 星的距离，以得到它们的空间分布。OB 星团或星协是银河系旋臂结构的有力示踪物<sup>[4][5][6]</sup>，借助它们的空间分布可以描绘出清晰的旋臂结构，并确定相关参数<sup>[7]</sup>，从而得出 OB 星和银河系的旋臂结构。将得到的结果与银河系旋臂的经典模型相比较，从而进一步分析银河系旋臂的构成。

### 5.2 “银河系旋臂结构研究”的中国虚拟天文台范例操作步骤

第一步：从天文数据访问门户中证认出 2MASS 星表中的 OB 星。

第二步：计算这些 OB 星的消光，估计距离。

第三步：在以银心为原点的三维直角坐标系下对 OB 星做聚类分析。

第四步：利用“银河系旋臂结构研究”的中国虚拟天文台范例门户的画图工具在 XY 平面上画出聚类后的结果。

第五步：对第四步结果应用 Hough 变换求出旋臂结构的参数

### 5.3 “银河系旋臂结构研究”的中国虚拟天文台范例的体系结构及其主要模块功能

“银河系旋臂结构研究”的中国虚拟天文台范例分三层，分别由客户端层、管理层和服务层组成。天文数据访问门户位于“银河系旋臂结构研究”的中国虚拟天文台范例的管理层，它为“银河系旋臂结构研究”的中国虚拟天文台范例的门户提供整合后的证认数据，同时为任务描述解析器提供计算数据。天文数据访问结点位于服务层，它提供异地的 2MASS 等星表资源服务。“银河系旋

臂结构研究”的中国虚拟天文台范例体系结构如图 5.1 所示。

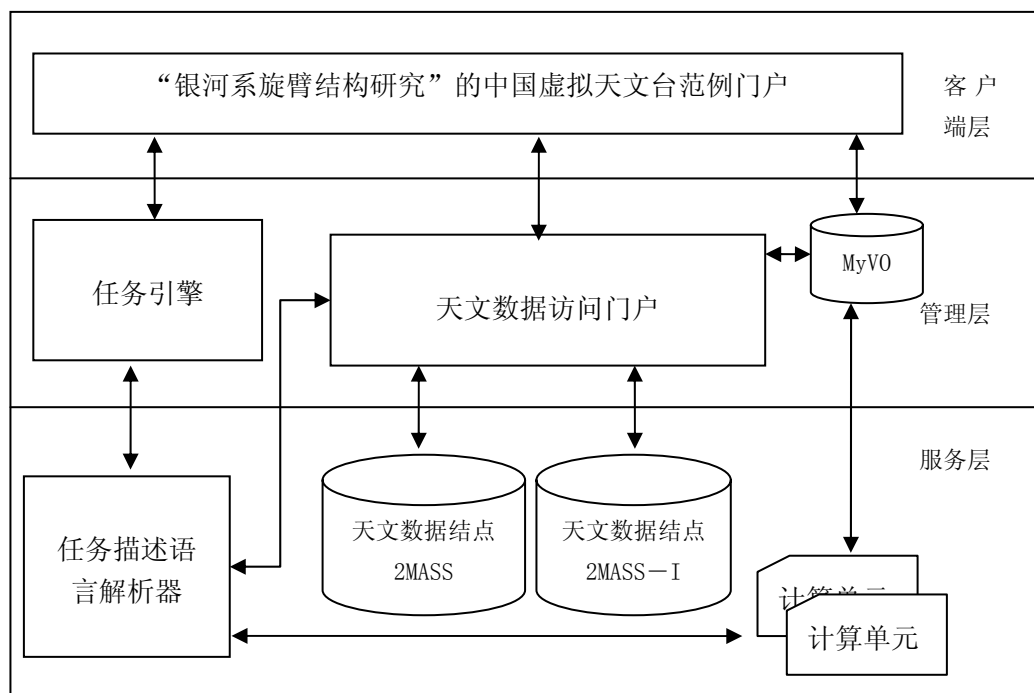


图 5.1 “银河系旋臂结构研究”的中国虚拟天文台范例的体系结构

### 5.3.1 “银河系旋臂结构研究”的中国虚拟天文台范例的门户

它提供整个“银河系旋臂结构研究”的中国虚拟天文台范例中的唯一图形用户界面，使用前需要由用户下载并安装在用户的计算机上，属于瘦客户端。它可以管理项目以及项目内的任务和数据文件，并且编写、检查 JDL 程序，运行任务。还可以监控系统的日志，下载或上传数据，提供可视化工具。

“银河系旋臂结构研究”的中国虚拟天文台范例的门户如图 5.2 所示。

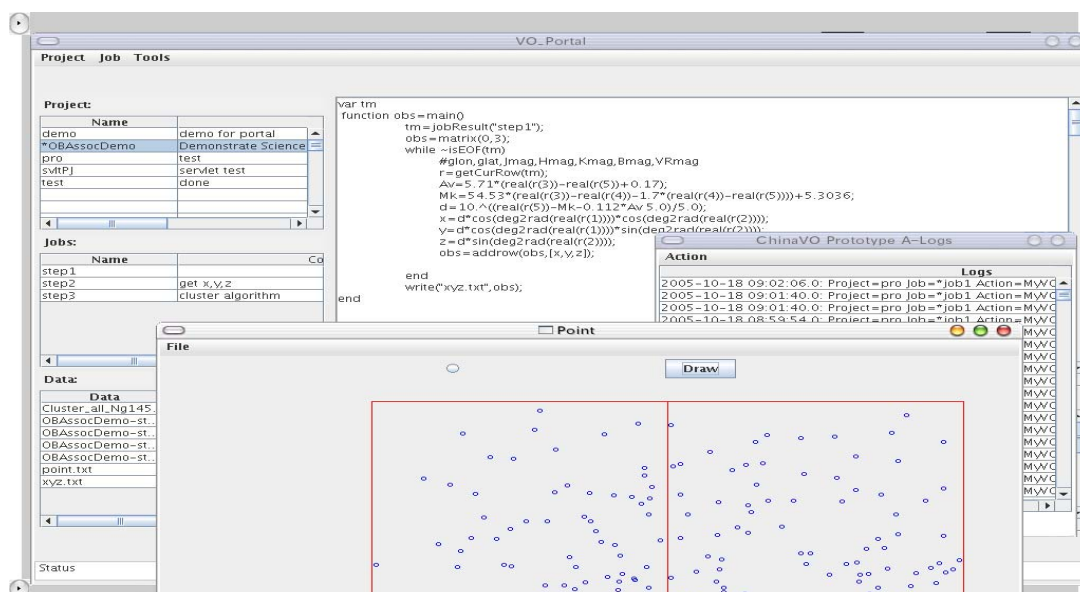


图 5.2 “银河系旋臂结构研究”的中国虚拟天文台范例门户

### 5.3.2 任务引擎

任务引擎<sup>[8]</sup>是一个可以保存状态的 WEB 服务，负责执行调度任务。它可以选择合适的任务描述语言解析器执行任务并监控任务的执行状态。任务引擎的状态则保存在后台数据库中。

### 5.3.3 任务描述语言

任务描述语言<sup>[8]</sup>是经过独立设计的、适应天文学的应用的语言。任务描述语言是独立设计的高级数值计算语言，它支持分布化的网格计算 (CompuCell)。它有两种表现形式：脚本、XML。脚本用于用户编写任务描述程序，而 XML 则用于程序自动生成和识别任务描述语言，它们二者等价可以相互转换。

### 5.3.4 天文数据访问门户与天文数据结点

中国虚拟天文台天文数据访问门户可同时管理多个分布的天文数据结点，是天文数据结点的汇聚点。为上层的任务引擎服务、可视化服务提供操作天文数据结点的接口，从而进行 2MASS 星表的交叉认证工作。

天文数据结点对中国天文数据中心 2MASS 等主要星表数据进行集成与管理，同时对外发布为国际上统一的数据访问标准接口，为用户屏蔽底层数据资源的异构性，提高分布、异构数据资源的访问性能，实现与各国数据中心资源的互联共享。

天文数据访问门户接收任务描述语言解析器解析后的查询语句，并将语句发送到相应的天文数据结点，查询后的结果返回给解析器，由解析器将结果送往计算单元进行处理。“银河系旋臂结构研究”的中国虚拟天文台范例的查询条件如下图 5.3 所示。

```
SELECT glon,glat,j_m,h_m,k_m,ra,decl,j_cm sig,h_cm sig,k_cm sig
FROM twomass_psc
WHERE glat>-10.0 AND glat<10.0 AND j_m-2.7*h_m+1.7*k_m<-
0.145 AND j_m-2.7*h_m+1.7*k_m> -0.19 AND k_m<=12
```

图 5.3 “银河系旋臂结构研究”的中国虚拟天文台范例 SQL 查询语句

### 5.3.5 计算单元

计算单元<sup>[8]</sup>支持任务描述语言调用各种数据挖掘工具箱，利用 WEB 服务支持远程调用，在有注册中心的支持下可以实现自动化的资源发现，可以用来集成现有的数据挖掘工具。

## 5.4 “银河系旋臂结构研究”的中国虚拟天文台范例的应用结论

按照上述步骤经过计算后得出银河系旋臂结构如下图 5.4 所示。图中虚线所标明的是经典银河系旋臂模型，虚线周围分布的黑点为聚类分析之后所得到的结果。

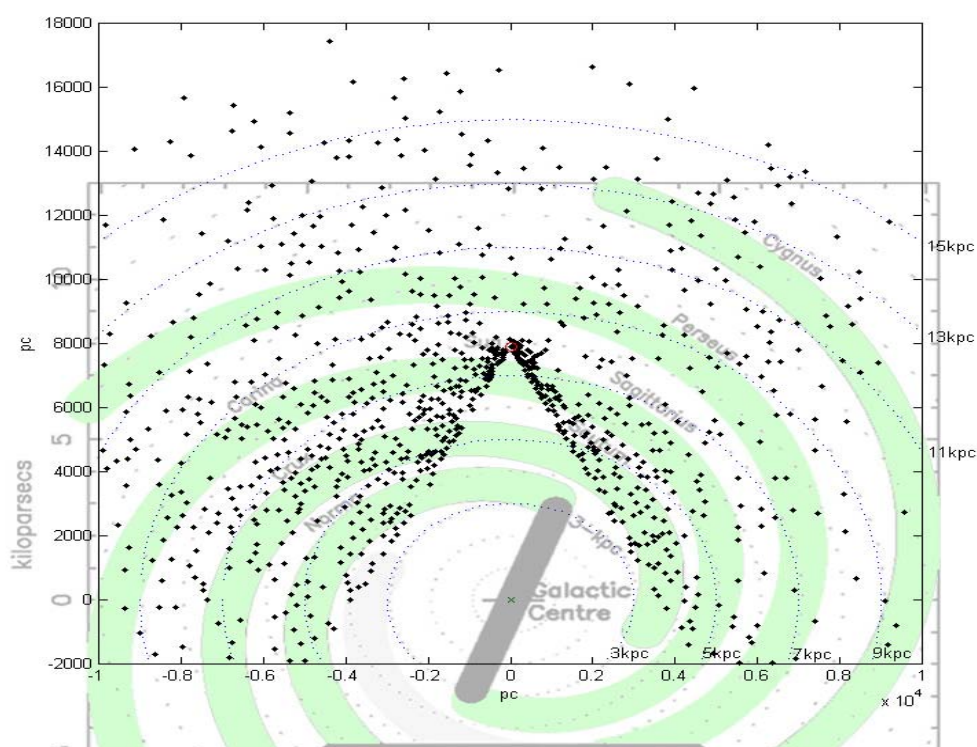


图 5.4 “银河系旋臂结构研究”的中国虚拟天文台范例的银河系旋臂模型

## 5.5 结论

“银河系旋臂结构研究”的中国虚拟天文台范例中天文数据访问门户管理着 2MASS 等多个天文数据结点，由“银河系旋臂结构研究”的中国虚拟天文台范例中上层的高级服务（任务引擎、任务解析器、数据挖掘服务）调用天文数据访问门户的接口来获取认证后的 OB 星。天文数据结点负责将异地、异构数据中心中的 2MASS 星表进行封装，为天文数据访问门户提供可供调用的服务接口。“银河系旋臂结构研究”的中国虚拟天文台范例充分的证明了中国虚拟天文台的天文数据访问门户和天文数据结点可以把数据资源以灵活、透明、标准的服务方式提供给国内外天文学家。使天文学家可以利用中国虚拟天文台对各国数据中心的观测数据直接进行访问、处理、分析。我们希望用这套系统作为中国虚拟天文台数据访问的门户，确保其能够为星表的数据挖掘、分布式计算工作提供底层支持，为多波段天文学研究提供坚实的底层平台。这对中国虚拟天文台进一步开展大样本天文数据分析、分布式计算等方面的研究具有重要的验证意义。



## 参考文献

- [1] [2MASS] <http://www.ipac.caltech.edu/2mass/>
- [2] Cardelli, J. et. al. The relationship between infrared, optical, and ultraviolet extinction. ApJ, 345, p245-256, 1989
- [3] Koornneef, J. Near-infrared photometry. II - Intrinsic colours and the absolute calibration from one to five micron. A&A, 128, p84-93, 1983
- [4] Blaauw, Adriaan. The O Associations in the Solar Neighborhood. A&A, 2, p213-246, 1964
- [5] Morgan, W. et. al. Studies in Galactic Structure. I. a Preliminary Determination of the Space Distribution of the Blue Giants. ApJ, 118, p318-322, 1953
- [6] Morgan, W. et. al. Some features of galactic structure in the neighborhood of the Sun. AJ, 57, p3, 1952
- [7] Vallée, Jacques P. The Spiral Arms and Interarm Separation of the Milky Way: An Updated Statistical Study. AJ 130,p569-575, 2005
- [8] ChaoLiu, DanWang, BoLiu. et.al. An Astronomical Data Mining Application Framework in China-VO SPIE accepted.

## 第六章 总结与展望

本论文从中国虚拟天文台的产生说起，分析了国际上天文数据访问与互操作领域的现状，介绍了 IVOA 标准和网格等相关技术，着重讲述了中国虚拟天文台的天文数据结点与天文数据访问门户的设计和实现方案。同时设计了查询分割方法和执行计划分配方法来解决大数据量星表获取以及交叉认证问题。最后用“银河系旋臂结构研究”的中国虚拟天文台范例来验证中国虚拟天文台的天文数据结点和天文数据访问门户的可行性与可靠性。论文的主要内容如下：

第一章介绍了虚拟天文台的产生背景，中国虚拟天文台目前主要的研究方向，并分析了数据访问与互操作在中国虚拟天文台建设中的重要性、必要性和可行性。同时也综述了各国在数据访问与互操作研究上的成果。

第二章概述了数据访问与互操作方面的相关技术。其中包括网格技术、中国虚拟天文台的体系结构和 IVOA 的数据访问与互操作标准。

第三章着重讲述了中国虚拟天文台天文数据结点的体系结构、模块功能、系统设计、运行流程等，同时针对数据库封装粒度问题展开了相关讨论，给出了一个较为合理的封装方案。

第四章阐述了中国虚拟天文台天文数据访问门户的体系结构、模块功能、分布式调度引擎的理论依据以及调度方法、运行流程等，最后利用测试平台检验了交叉认证功能的可行性。

第五章利用“银河系旋臂结构研究”的中国虚拟天文台范例来证明天文数据结点和天文数据访问门户的可靠性与稳定性。证实了天文数据结点和天文数据访问门户可以作为未来中国虚拟天文台的数据汇集与分析平台。

虽然我们按照 IVOA 的数据访问与互操作标准建立了天文数据结点，但它的功能还需要进一步完善。比如：我们目前还没有实现天文图像数据的访问，大样本数据的交叉认证工作也还需进一步研究。在数据访问门户方面，分布式调度引擎也需要尝试新的算法，不断的优化，提高效率。随着 IVOA 的数据访问与互操作标准的不断改进，我们也需要不断的调整我们的访问系统平台，以便使天文数据访问系统平台与国际接轨。

目前世界各国虚拟天文台都在从事天文数据结点和天文数据访问门户的建

设，如何能使它们的效率更高、运行更稳定、一次性地分析大样本数据成为今后几年这个研究方向关心的重点。流媒体算法、数据仓库、P2P 等相关技术都可能成为我们将来解决这些问题的实验方法，最终实现大样本数据的获取、分析和挖掘。

## 致 谢

三年时光转眼即逝，在此我要衷心地感谢我的导师赵永恒和指导我工作的崔辰州老师和张彦霞老师，感谢他们近年来对我无微不至的关怀和孜孜不倦的教诲。

首先，最为感谢的是我的指导老师赵永恒研究员。感谢他两年多来对我无微不至的关怀和孜孜不倦的教诲！两年来，无论在理论学习阶段，还是科研项目过程中，赵老师在学术上给予我启迪，拓宽我研究思路，引导我学术思维、教给我科学研究的方法。虽然赵老师日常工作非常繁忙，但他仍然在百忙之中对我不解的问题进行细致地指导和帮助，提供了优良的软硬件环境和学习环境，十分感谢赵老师对我研究能力的培养。导师的言传身教是我两年来最大的收获，其为人之谦和大度，其为学之严谨让我长久铭记。我将继续刻苦钻研、不断进取，只有这样才能不负老师之期望。

同样在这里我要深深地感谢我的指导老师崔辰州博士，我的每一步工作都是在他的指引下完成的。我从开始选题、资料收集、开题、研究和撰写文章的每一个环节，无不得到崔辰州博士的悉心指导和帮助，全文字里行间无不浸透了崔老师之心血。最让我难以忘记还是崔老师严谨的治学态度、对事业执著的追求以及平易近人的品质都是我学习的榜样。而且他还是一位十分尊重学生想法的老师，这一点更让我敬佩不已。

同时也感谢张彦霞老师，张老师为人心胸宽阔、治学严谨、待人热忱。虽然我跟张老师相处的时间不多，但是从相处的一段时间我已经可以深深地感受到张老师的为人。非常感谢她在工作上给予的帮助与支持。

感谢党支部书记孙盛慈老师。在研二的时候，在孙老师的指引下我的思想觉悟有了很大的提高，我一定时刻提醒自己不论身在何处，都要以党员的标准来要求自己。

感谢杜红荣老师和李冰。从我走进天文台开始，研究生期间的生活、学习到毕业分配，她们都给我给予了很大的帮助。

我还要感谢 LAMOST 项目组的袁辉老师和陈英老师在生活上的帮助。

感谢刘超博士在理论研究、算法实现和程序设计方面给我提供了很大的帮

助。感谢经常一起讨论的王丹、吴潮、尹红星、李会贤、高丹、孙华平、郑征、张旭、路勇同学。

最后，我要感谢我的父亲和我的爱人。正是他们默默无闻的奉献和一贯的支持和鼓励，才使我有信心去克服一切困难，有充足的时间和精力去完成学业。他们全心的付出和无言的关爱始终是我前进的巨大动力。

感谢所有关心过我的人。

## 发表文章目录

刘波,崔辰州,赵永恒. 构建中国虚拟天文台天文数据结点. 天文研究与技术, 2006 (4) In Press。

ChaoLiu, DanWang, BoLiu. et.al. An Astronomical Data Mining Application Framework in China-VO SPIE accepted.